



International Encyclopedia of Education

3rd Edition



Editors-in-Chief:
Penelope Peterson, Eva Baker and Barry McGaw

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

How to go to your page

This eBook set contains 8 volumes. Each volume has its own page numbering scheme for front matter and body.

The pages are labeled with the Volume number and page separated by a colon.

For example, to go to page x of Volume 1, type '1:x' in the "page #" box at the top of the screen and click "Go".

To go to page 203 of Volume 6, type '6:203' in the "page #" box... and so forth.

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

EDITORS-IN-CHIEF

PENELOPE PETERSON

School of Education and Social Policy
Northwestern University
Evanston, IL
USA

EVA BAKER

UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

BARRY MCGAW

McGraw Group Pty Ltd
Mont Albert Nth, VIC
Australia



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
The Boulevard Langford Lane, Kidlington, Oxford OX5 1GB, UK

Copyright © 2010 Elsevier Ltd. All rights reserved

LATVIA
Copyright © 2010 A Joma

STUDENT DATA SYSTEMS AND THEIR USE FOR EDUCATIONAL IMPROVEMENT
Copyright © 2010 J C Wayman, V Cho, and M P Richards

QUALIFICATIONS FRAMEWORKS AND THEIR ROLE IN THE REFORM OF EDUCATION AND TRAINING
Copyright © 2010 M Young

LANGUAGE OF VOCATIONAL EDUCATION AND TRAINING
Copyright © 2010 J Misko, J MacKenzie, and R-A Polvere

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44)(0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at (<http://elsevier.com/locate/permissions>), and selecting *Obtaining permissions to use Elsevier material*

Notice

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Catalog Number
A catalogue record for this book is available from the Library of Congress

ISBN: 978-0-08-044893-0

For information on all Elsevier publications
visit our website at books.elsevier.com

PRINTED AND BOUND IN THE UNITED KINGDOM
09 10 11 12 13 10 9 8 7 6 5 4 3 2 1



PREFACE

A preface usually provides a brief introduction to a work, intended to set the stage, provide some background insight, and whet the appetite of the reader. In our case, however, this preface has to address a fundamental question, one that was in our minds at the time we were recruited as Editors-in-Chief for the International Encyclopedia of Education. The question was “Why do we need an Encyclopedia? Its subtext was inspired by the ever-growing, ever-popular Internet. We believe that *this* Encyclopedia is desperately needed and will become a valued resource in education and associated social sciences and arts. The reasons are intellectual and procedural. Anyone with a modicum of knowledge knows that finding and trusting information gleaned from the Internet are two separate actions. The reliance on browsers to help discover references and comments result in resources based on popularity not quality. Pithy titles catch the eye and references rise in the ranks of browser searchers. Related to this is the “editing” in the Internet realm of populist efforts at encyclopedia, references, and other compilations. Once again, after removing offensive material, the accuracy, completeness, lack of bias, and other provenance for entries simply do not exist. Experienced researchers in education can sort through and make intelligent choices. Novices and many journeyman, or practitioners, parents, and policy makers cannot. Contrast how this Encyclopedia was built. Key domains of educational research were identified, and a tentative list of sub-domains or useful applied areas was posited. Then the Editors-in-Chief (apologies for the awkwardness of the term) identified the leading researcher in a particular domain, and with surprisingly little effort, recruited them to participate. They in turn identified the two best researchers in a sub-domain, such as formative assessment or the training of pre-school teachers. The authors of the sections of the Encyclopedia do not represent a collective group of friends and acquaintances, although friendships have been made. Rather they embody a deep and broad scholarly community. The difference from compiled Internet resources is the built-expertise and intellectual engagement of the authors. The summary of the developments and futures in their personal areas of scholarship have been filtered through their years of experience, both as scholars and communicators. Quality, then, is endemic to each piece, developed through this top-down identification of expertise, and made indelible by the bottom-up application of high standards from people leading the sub-domains – the authors, and the domains themselves, the section editors.

On a procedural level, the publishers early committed to the notion that this Encyclopedia would also be an online resource, and access would be available through print, for those with strong bookcases and the persisting love of turning real pages. The Internet version will allow multiple prisms through which the reader may access articles and provide, as it were, an emulation of the Internet in our field, albeit bounded by expertise and high quality.

What must be underscored in the assessment of this effort are the Editors-in-Chief and the publishers’ commitment to find excellence worldwide. We tried very hard to persuade notable scholars from all parts of the world to make contributions. Less than to fulfill the title of “International,” we were on the hunt for perspectives that would enrich the scope and depth of the sections. Our section editors put in enormous time attempting to find the best in the field, wherever they resided. Yet, not everyone is in the volume. Some were overcommitted. Many were not fully confident of their English, and the automated translation software has not yet met standards for technical writing. We believe that such writing and editing tools will make the outreach to an even broader International group of scholars possible in future revisions, or online updates. Furthermore, the birth of the World Educational Research Association (in 2009) will provide a better set of interlocking networks to find and evaluate scholarship from any place on the globe.

Finally, the scope of the effort must be acknowledged: 28 section editors, 926 articles were commissioned, drafted, reviewed, redrafted, edited, and put together in the space of four years. The publishers underwent some internal changes, and alterations in management. We as Editors-in-Chief, changed roles, moved, and also had to keep our own research and development enterprises afloat. Deadlines wobbled; authors dropped from view and had to be replaced.

Yet, at times frustrating as all development is, we find the final product exhilarating. We are enthusiastic not simply because it came into being at all, but because the collective light of the minds that wrote have left a bright resource for the future, one that will impact the way our colleagues understand and experience the educational knowledge, improvement, and impact in the future.

Eva Baker
Penelope Peterson

EDITORS

EDITORS-IN-CHIEF

Penelope Peterson

School of Education and Social Policy
Northwestern University
Evanston, IL
USA

Eva Baker

UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

Barry McGaw

McGaw Group Pty Ltd
Mont Albert Nth
Melbourne, VIC
Australia

SECTION EDITORS

Walter Allen

HEOC Division
University of California, Los Angeles
Los Angeles, CA
USA

Vibeke Grøver Aukrust

Institute for Educational Research
University of Oslo
Oslo
Norway

Dominic J Brewer

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

Kai-Ming Cheng

Faculty of Education
The University of Hong Kong
Hong Kong

Gregory Chung

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Penny Enslin

Department of Educational Studies
University of Glasgow
Glasgow
UK

Noelle Griffin

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Dennis Hocevar

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

David Hopkins

Institute of Education
University of London
London
UK

Mary James

Faculty of Education
University of Cambridge
Cambridge
UK

Sanna Järvelä

Department of Educational Sciences and
Teacher Education
University of Oulu
Oulu
Finland

Thomas Karmel

National Centre for Vocational Education Research
Adelaide, SA
Australia

Harold G Levine

School of Education
University of California, Davis
Davis, CA
USA

Robert L Linn

University of Colorado at Boulder
Ouray, CO
USA

Allan Luke

Faculty of Education
Queensland University of Technology
Brisbane, QLD
Australia

Ann M Mastergeorge

Human Development and Family Studies
University of California, Davis
Davis, CA
USA

Patrick J McEwan

Department of Economics
Wellesley College
Wellesley, MA
USA

Barbara Means

Center for Technology in Learning
SRI International
Menlo Park, CA
USA

Ingrid Moses

University of Canberra
Canberra, ACT
Australia

Bill Mulford

Faculty of Education
University of Tasmania
Hobart, TAS
Australia

Yu Jing Ni

Faculty of Education
The Chinese University of Hong Kong
Hong Kong

Festus Obiakor

Department of Exceptional Education
University of Wisconsin-Milwaukee
Milwaukee, WI
USA

P David Pearson

Graduate School of Education
University of California, Berkeley
Berkeley, CA
USA

Colin Power

Graduate School of Education
University of Queensland
Brisbane, QLD
Australia

Kjell Rubenson

Faculty of Education
The University of British Columbia
Vancouver, BC
Canada

Sandip Sinharay

Education Testing Services
Princeton, NJ
USA

Collette Tayler

Queensland University of Technology
Brisbane, QLD
Australia

Nico Verloop

ICLON, Graduate School of Education
Leiden University
Leiden
Netherlands

HOW TO USE THE ENCYCLOPEDIA

The International Encyclopedia of Education is intended for use by students, research professionals, and interested others. Articles have been chosen to reflect major disciplines in the study of education and common topics of research by academics in this domain. Each article serves as a comprehensive overview of a given area, providing both breadth of coverage for students, and depth of coverage for research professionals. We have designed the encyclopedia with the following features for maximum accessibility for all readers.

The contents of the encyclopedia are arranged alphabetically by section, and within sections, alphabetically by article. The Subject Index is located in Volume 8. Some topics are covered in a multitude of articles from differing perspectives, while other topics may have only one entry. We encourage use of the index for access to a subject area, rather than use of the Contents list alone, so that a reader has a full notion of the coverage of that topic.

The articles include cross-references to other related encyclopedia articles, suggested further readings where applicable, and many contain relevant websites for additional information. We encourage readers to use the cross-references to locate other encyclopedia articles that will provide more detailed information about a subject.

The Further Reading sections include recent secondary sources to aid the reader in locating more detailed or technical information. Review articles and research articles that are considered of primary importance to the understanding of a given subject area are also listed. These suggested further readings are not intended to provide a full reference listing of all material covered in the context of a given article, but are provided as next steps for a reader looking for additional information.

This page intentionally left blank

CONTENTS

Preface	v–vi
Editors	vii–viii
How to use the Encyclopedia	ix

VOLUME 1

A

Adult Education

Adult Education Overview	<i>K Rubenson</i>	1
--------------------------	-------------------	---

Adult Education – Adult Learning, Instruction and Program Planning

Adult Learning	<i>S B Merriam</i>	12
Adult Learning and Instruction: Transformative Learning Perspectives	<i>P Cranton</i>	18
Adult Learning in a Biographic Perspective	<i>L West</i>	25
Adult Learning, Instruction and Programme Planning: Insights from Freire	<i>P Mayo</i>	31
Characteristics of Adult Learning	<i>K Illeris</i>	36
Informal Learning: A Contested Concept	<i>P Hodgkinson</i>	42
Organizational Learning	<i>P-E Ellström</i>	47
Program Planning	<i>A L Wilson and R M Cervero</i>	53
Workplace Learning Frameworks	<i>S Billett</i>	58

Adult Education – as a Field of Study

Class Analysis in Adult Education	<i>T Nesbit</i>	64
Gender Analysis	<i>J Stalker</i>	70
Race and Ethnicity in the Field of Adult Education	<i>J Johnson-Bailey and D Drake-Clark</i>	77
Trends in Workplace Learning Research	<i>T Fenwick</i>	83

Adult Education – Concepts

Lifelong Learning	<i>J Field</i>	89
Rewriting the History of Adult Education: The Search for Narrative Structures	<i>B J Hake</i>	96

Adult Education – Domains and Provision

Adult Basic Education: A Challenge for Vocational Based Learning	<i>R Höghelm</i>	102
Adult Literacy Education	<i>L Tett and R St.Clair</i>	107
Citizenship and Immigrant Education	<i>D Schugurensky</i>	113
Community Based Adult Education	<i>B Connolly</i>	120
Continuing Professional Education: Multiple Stakeholders and Agendas	<i>R M Cervero and B J Daley</i>	127
Financing of Adult and Lifelong Learning	<i>Å Sohlman</i>	133
Labor Education	<i>K Forrester</i>	139
Learning Cities and Regions	<i>C Duke</i>	144
Museums as Sites of Adult Learning	<i>R S Grenier</i>	150
Overview of Lifelong Learning Policies and Systems	<i>D-B Kwon and C H O Daeyeon</i>	156
Popular Adult Education	<i>S Tøsse</i>	162
Provision of Prior Learning Assessment	<i>P Andersson</i>	169
The Age of Learning: Seniors Learning	<i>P Jarvis</i>	175
University Adult Continuing Education: The Extra-Mural Tradition Revisited	<i>M Slowey</i>	183

Adult Education – Economy and Society

Adult Education and Civil Society	<i>J Crowther and I Martin</i>	191
Adult Education and Nation-Building	<i>S Walters</i>	198
Economic Outcomes of Adult Education and Training	<i>A Ferrer and W C Riddell</i>	204
Health and Adult Learning	<i>P Bélanger and M Robitaille</i>	211
Modernization Processes and the Changing Function of Adult Learning	<i>H S Olesen</i>	217
The Political Economy of Adult Education	<i>G Rees</i>	223
Wider Benefits of Adult Education	<i>R Desjardins and T Schuller</i>	229

Adult Education – Participation

Barriers to Participation in Adult Education	<i>K Rubenson</i>	234
Participation in Adult Learning	<i>R Desjardins</i>	240

C**Comparative Education**

Comparative and International Education: English-Language Scholarly Publications	<i>D Post and M Fedri</i>	251
Comparative Education: Philosophical Issues and Concepts	<i>M Mason and C W Evers</i>	257
Comparative Education: Societies and Associations	<i>M Bray</i>	266

Curriculum Development

Curriculum and Religion	<i>C Kapitzke and P Graham</i>	272
Gender and Curriculum	<i>J Marsh</i>	277
Local Teacher-Based Curriculum Development	<i>M D Osborne and H B Barker</i>	283

Curriculum Development – Contexts

Curriculum and Human Rights	<i>C Lohrenscheit</i>	287
Curriculum and the Education of Cultural and Linguistic Minorities	<i>S May</i>	293
Curriculum in Postcolonial Contexts	<i>A Hickling-Hudson</i>	299
Curriculum, Economic and Cultural Development	<i>A Rampal</i>	307
Globalization and Curriculum	<i>R Dale</i>	312
Popular Culture	<i>K Schultz and R Throop</i>	318

Curriculum Development – Evaluation and Research

An Overview of Research in Curriculum Inquiry	<i>F M Connelly and S J Xu</i>	324
Curriculum Evaluation: Approaches and Methodologies	<i>V Klenowski</i>	335
Curriculum Studies, Discourse Analysis, and the Construction of Historical Time	<i>J Laspina</i>	342
Curriculum Use in the Classroom	<i>M Ben-Peretz and B Eilam</i>	348
The Roots and Routes of Teacher-Based Action Research and Curriculum Inquiry: An Historical Perspective	<i>L G Putney and J L Green</i>	355

Curriculum Development – Planning and Development

Curriculum and Syllabus Design	<i>A Woods, A Luke, and K Weir</i>	362
Curriculum and the Publishing Industry	<i>J Larson, A-R Allen, and D Osborn</i>	368
Curriculum and Teacher Change	<i>J G Ladwig</i>	374
Curriculum Governance and Planning	<i>B Levin</i>	379
Curriculum Planning and Systems Change	<i>Z Deng</i>	384
Curriculum, Digital Resources and Delivery	<i>Y Zhao, G Zhang, and C Lai</i>	390
Textbook Development and Selection	<i>P Shannon</i>	397

Curriculum Development – Subject Matter

Curriculum and Globalization: Higher Education	<i>K H Mok</i>	403
Curriculum Development as Subject Matter: Social Studies	<i>J Brophy and J Alleman</i>	411
Curriculum Development in the Area of Reading	<i>C L Madda and T E Raphael</i>	415
Curriculum in the Arts	<i>C M Thompson, L Bresler, and T E Costantino</i>	423
Curriculum: Foreign Language Learning	<i>Angel M Y Lin</i>	428
First Language Learning	<i>T Rogers and L Moffatt</i>	435
Learning a Second Language in First Language Environments	<i>E Bernhardt and D J Tedick</i>	441
Literature	<i>S Hynds</i>	447
Mathematics	<i>B Kaur and C P Vistro-Yu</i>	453
Physical Education and Sports	<i>D Kirk</i>	459
Primary and Elementary/Middle Grades Reading	<i>D E Alvermann and C A Mallozzi</i>	464
Science	<i>Y Kali and M Linn</i>	468

Curriculum Development – Theory

Bilingual Learning (Learning L1 and L2 in an L1 and L2 Environment)	<i>D Schwinge</i>	475
Curriculum and Complex Systems Theory	<i>W-M Roth and J Thom</i>	481

Curriculum and Constructivism	<i>B Davis and D Sumara</i>	488
Curriculum and Critical Theory	<i>L Yates</i>	494
Curriculum and Poststructuralist Theory	<i>J L Miller</i>	499
Curriculum and Structuralist Sociology: The Theory of Codes and Knowledge Structures	<i>J Muller and J Gamble</i>	505
Curriculum Reform	<i>R Gilbert</i>	510
Early Childhood Curriculum and Developmental Theory	<i>S M-Y Lim and C Genishi</i>	514

D

Demography and Social Change

An Evaluation Framework for E-Learning Effectiveness in the Arab World	<i>W AbuSneineh and M Zairi</i>	521
Globalization and Social Justice in Higher Education	<i>E Kimura-Walsh</i>	536
Racial Inequality and the Economic Payoffs of Higher Education	<i>S A Lachtman</i>	542
School Health Education	<i>J S Yang</i>	547

Demography and Social Change – Demography

A Paradigm of Contradictions: Racism and Science Education	<i>B A Brown and J W Mutegi</i>	554
Africana Studies: Past, Present And Future	<i>B R Hare, A Young, J B Stewart, and D P Aldridge</i>	565
Beyond Black Nationalism to Black Internationalism: The Compelling Case for Examining Black Educational Challenges Globally	<i>K Freeman</i>	573
Beyond Racial, Ethnic, and Gender Bias in Education Statistics	<i>J T Lott and C Bennett</i>	581
Caste and Education in India	<i>G G Wankhede</i>	589
Demography, Community, and the Education of Asian American and Pacific Islander Students	<i>R T Teranishi</i>	595
Educational Attainment and Mortality Differentials	<i>H D Horton, L Hixson, and S M Monnat</i>	602
Educational Opportunity and Latino/Chicano College Choice: New Findings, and Theoretical Perspectives	<i>M A Ceja and P A Pérez</i>	610
Ethnic Minority Identity and Educational Outcomes in a Rising China	<i>G A Postiglione</i>	616
Ethnic Studies and Women's Studies: From the Past to the Present, Providing a Space in the Academy for Scholarship Focused on the Histories and Experiences of Marginalized People	<i>C A Brown</i>	623
Gender Equity in Higher Education: Challenges and Celebrations	<i>L Morley</i>	629
Locating Space and Place in the College Access Debate: New Tools for Mapping and Understanding Educational Inequity and Stratification	<i>B Cantwell and J F Milem</i>	636
Public Policy and Inequality in Postsecondary Opportunity: Educational Statistics and the Failure of Education Reform	<i>E P St. John, K Williams, and K Moronski</i>	649
The Education of Indigenous Students	<i>J W Tippeconnic and S C Faircloth</i>	661
The Health Advantages of Educational Attainment	<i>G Y Dinwiddie</i>	667
To Go or Not to Go: Access and Barriers to Ethnic Minority Participation in Higher Education within the United Kingdom	<i>P Brug and T L Oliver</i>	673
Universal Secondary Education and the Two-Tiered School System in the English-Speaking Caribbean	<i>P Anderson and J Devonish</i>	679
Women of Color in Science, Technology, Engineering and Mathematics (STEM): Refining the Concepts, Reframing the Issues	<i>C B Leggon</i>	686

Demography and Social Change – Social Change

Educational Diversity in U.S. Law Schools	<i>C E Daye, A T Panter, W R Allen, M E Deo, and L F Wightman</i>	691
Educational Reform	<i>F Lopez-Segrera</i>	698
Emergence of For-profit Higher Education	<i>X Li</i>	705
Engaged Education: Experiential Learning, Intensive Field Experiences, and Social Change	<i>T R Rone</i>	711
Higher Education and Social Change in Mexico	<i>A Mungaray, G Estrella, P Moctezuma, and M Ramírez-Urquidy</i>	719
The Changing Role of the State in Higher Education	<i>K Gürüz</i>	728
The Role of the State in School Reform: Responding to Increasing Ethnic Diversity in Britain	<i>P Ratcliffe</i>	737
Transforming Mindsets Through Education for Sustainable Development	<i>V M Segovia</i>	746

Demography and Social Change – Sociology

Achievement Motivation in Ethnic Minority Youth	<i>S Graham and E Bryant</i>	753
Affirmative Action and Higher Education in Brazil	<i>J T dos Santos and D M Queiroz</i>	760
Black in White: Black Students at White and Black Colleges	<i>L Evans and J R Feagin</i>	767
Bologna Process: On the way to a Common European Higher Education Area	<i>A V Rezaev</i>	772
Critical Race Theory in Education Research	<i>R Vue and C B Newman</i>	779
Cultural Diversification and Japanese Education: Social Constructions of the New Diversity	<i>R Tsuneyoshi</i>	787
Education in War Countries: A Disempowerment of Society	<i>J L Nkata</i>	793
Educational Attainments of U.S. Black Males and Females: 1971 to 2003	<i>F B Bonner, M-C Jipguep-Akhtar, and R J Harrison</i>	799
Educational Preparation: Fostering the Self-Efficacy and Resilience of Urban Adolescent Youth	<i>M B Spencer and B Tinsley</i>	806
HIV Education for Low-Literate People: Transforming Students and Communities through Paulo Freire's Praxis and the Pedagogy of Action	<i>N Z Haniff</i>	814
How Are Discussions of Interdisciplinary Studies Linked to Diversity Discourse?	<i>C N Spence</i>	823
Identifying Patterns of Doctoral Attrition Across Academic Fields	<i>A James, K J Fosnacht, E B Stolzenberg, and D Thirumalai</i>	827
Male Underachievement in Education Across the Globe: A Shift in Paradigm for Gender Disparities Regarding Academic Achievement	<i>J F L Jackson, J L Moore III, and R A Leon</i>	838
Noncognitive Measures for Higher Education Admissions	<i>W E Sedlacek</i>	845
Peace Education	<i>J Page</i>	850
Remote Control: Africa's Development and Outflow to Developed Nations of Technical Talent	<i>F Haniff</i>	855
Scholar-Baller®: Student Athlete Socialization, Motivation and Academic Performance in American Society	<i>C K Harrison, S Bukstein, J Mottley, E Comeaux, J Boyd, C Parks, and D Heikkinen</i>	860
Service-Learning and Diversity as Mediums to Foster Civic and Community Engagement in Undergraduates	<i>L Oseguera</i>	866
Success and Failure in Education and Criminal Justice: Identifying Common Mechanisms	<i>S Ramirez and A Harris</i>	874
The History of Education: Race and Education	<i>J O Jewell</i>	881

VOLUME 2**E****Early Childhood Care and Education Theory**

- Early Childhood Care and Education: The Family, The Market, and The State *S B Kamerman and M J Neuman* 1
- Gender Issues in Early Childhood Education and Care *B Pocock* 11

Early Childhood Education and Care Governance

- Early Childhood Risk, Protection and Abuse Prevention *E Khoo* 17
- Evaluating Early Childhood Education and Care Programs *M Janus and S Brinkman* 25

Early Childhood Education and Care Investment

- Investing in Early Childhood Education and Care: The Health and Wellbeing Case *T Moore and F Oberklaid* 32
- Cost-effective Early Childhood Programs from Preschool to Third Grade *A J Reynolds, J A Temple, and B White* 38
- Early Childhood in Post-Modern Cultures: Thoughts and Some Concerns *P Gammage* 49
- Investing in Early Childhood Education and Care: Some Policy Implications *J Bennett* 55
- Investing in Early Childhood Education and Care: The Economic Case *G Cleveland and M Krashinsky* 63

Early Childhood Education and Care Programs

- Early Care and Education Programs for Infants and Toddlers *B Nyland* 69
- Literacies in Early Childhood: The Preschool Period *M Flear and B Raban* 75
- Participation in Early Childhood Education and Care Programs: Equity, Diversity and Educational Disadvantage *M Vandenbroeck* 81
- Teaching in Early Childhood Centers Instructional Methods and Child Outcomes *I Siraj-Blatchford* 86

Early Childhood Education and Care Provision

- Early Childhood Education and Care in the People's Republic of China *Zeng Xiaodong* 93
- Home-Based and Institutional Early-Childhood Education and Care Services *R Fukkink and H Blok* 98
- Integrated Early Childhood Education and Care Services – Care, Upbringing, Education and Health *S Colley* 104
- Issues of Access and Program Quality *J T Hustedt and W S Barnett* 110
- Kindergarten Transitions and Linkages to Primary School – Readiness Re-conceptualized *A Petriwskyj* 120
- Parent Support in Early Childhood – Approaches and Outcomes *K S Howard, K A Beckmann, and J Brooks-Gunn* 126

Early Childhood Education and Care Research

- Development and Implementation of Early Learning Standards in the United States *C Scott-Little* 132
- Global Status of Early Learning and Development Standards *P R Britto and S L Kagan* 138
- Rethinking Early Childhood Education and Care: Implications for Research and Evaluation *K Thorpe, D Cloney, and C Tayler* 144

Early Childhood Education and Care Theory

- Child Rearing and Early Education: Parents and Professionals: Theoretical and Historical Influences *H Penn* 151

Early Childhood Education and Care Workforce

- Early Childhood Education and Care Professional Profiles in the United Kingdom: Multiagency Teamwork and Transdisciplinary Challenges *A Anning and P Calder* 157
- Early Childhood Education and Care Staff Support: Mentoring and Professional Development in Finland *P Lindberg* 162

Economics of Education

- Education and Health *P Muennig* 169
- Data *M Perez and M Socias* 178
- Empirical Research Methods in the Economics of Education *P J McEwan* 187
- Theoretical Concepts in the Economics of Education *D J Brewer, G C Hentschke, and E R Eide* 193

Economics of Education – Costs and Financing of Schooling

- Cost–Benefit Analysis and Cost–Effectiveness Analysis *C Belfield and H M Levin* 199
- Economic Approaches to Adequacy *J Imazeki* 204
- Economic Approaches to School Efficiency *L L Taylor* 210
- School Finance Reform *T Downes* 216
- School Finance: An Overview *J K Rice, D H Monk, and J Zhang* 222
- The Economics of Tuition and Fees in Higher Education *R G Ehrenberg* 229

Economics of Education – External Benefits of Schooling

- Education and Civic Engagement *T S Dee* 235
- Education and Crime *L Lochner* 239
- Education and Economic Growth *E A Hanushek and L Wößmann* 245
- Parental Socioeconomic Status, Child Health, and Human Capital *J Currie and J Goodman* 253
- The External Benefits of Education *W W McMahon* 260

Economics of Education – Labor Markets, Education and Earnings

- Education and Inequality *J Blanden and S Machin* 272
- Human Capital *E R Eide and M H Showalter* 282
- Race Earnings Differentials *M Carnoy* 288
- Returns to Education in Developed Countries *M Gunderson and P Oreopoulos* 298
- Returns to Education in Developing Countries *H A Patrinos and G Psacharopoulos* 305
- School Quality and Earnings *J R Betts* 313
- Signaling in the Labor Market *M E Page* 321
- The Economics of High School Dropouts *R W Rumberger* 325

Economics of Education – Markets, Incentives and School Choice

- Competition and Student Performance *A A Payne* 331
- Educational Privatization *C Belfield and H M Levin* 337

Student Incentives	<i>E R Eide and M H Showalter</i>	342
The Economic Role of the State in Education	<i>D N Plank and T E Davis</i>	347
The Economics of Catholic Schools	<i>W Sander and D Cohen-Zada</i>	354
The Economics of Charter Schools	<i>R Zimmer and R Buddin</i>	360
The Economics of Parental Choice	<i>R Chakrabarti and J Roy</i>	367
The Economics of School Accountability	<i>D N Figlio and H F Ladd</i>	374
The Efficacy of Educational Vouchers	<i>R Zimmer and E P Bettinger</i>	380
Tiebout Sorting and Competition	<i>T J Nechyba</i>	388

Economics of Education – Production of Schooling

Desegregation, Academic Achievement and Earnings	<i>S Rivkin</i>	394
Education Production Functions: Concepts	<i>D N Harris</i>	402
Education Production Functions: Developed Country Evidence	<i>E A Hanushek</i>	407
Education Production Functions: Evidence from Developing Countries	<i>P Glewwe and S Lambert</i>	412
Family Environment in the Production of Schooling	<i>R Rothstein</i>	423
Neighborhoods and Peers in the Production of Schooling	<i>J Vigdor and J Ludwig</i>	431
Teacher Quality in Education Production	<i>I Pelayo and D J Brewer</i>	438
The Economics of Class Size	<i>D W Schanzenbach</i>	443
The Economics of Early Childhood Interventions	<i>M Nores</i>	450

Economics of Education – Teaching Labor Markets

An Overview of Teacher Labor Markets	<i>H Lankford and J Wyckoff</i>	456
Compensating Differentials in Teacher Labor Markets	<i>J Chambers</i>	465
Economic Approaches to Teacher Recruitment and Retention	<i>S Loeb and J Myung</i>	473
Teacher Incentives	<i>L Santibañez</i>	481
Teacher Supply	<i>P Dolton</i>	489
Teacher Training and Preparation in the United States	<i>M Podgursky, M G Springer, and R Hutton</i>	498
Teachers in Developing Countries	<i>F H Rogers and E Vegas</i>	504
The Economics of Teachers' Unions in the United States	<i>K Strunk</i>	512

Education of Children with Special Needs

Adapted Physical Activity for Students with Special Needs	<i>S R Hodge</i>	519
Assessment Accommodations for Children with Special Needs	<i>R J Kettler and S N Elliott</i>	530
Assistive Technology and Educational Practice	<i>H P Parette and G R Peterson-Karlan</i>	537
Augmentative–Alternative Communication	<i>C Mason and K M Chinn</i>	544
Community Integration and Employment of Youth with Special Needs	<i>B A Ford and A W Daviso III</i>	550
Cooperative Learning for Children with Special Needs	<i>E A Dooley and K R Kossar</i>	555
Co-Teaching	<i>E D Cramer</i>	560
Cultural and Linguistic Diversity and Children with Special Needs	<i>J M Patton, N Day-Vines, and A McSwain</i>	565
Curriculum-Based Assessment and Students with Special Needs	<i>D J O'Shea and J D Katsafanas</i>	571

Diagnosis and Classification	<i>M Sapp</i>	577
Direct Instruction and the Education of Children with Special Needs	<i>T Mehring</i>	584
Disabilities and Gifted Education	<i>V I Daniels and M J McCollin</i>	588
Early Childhood Special Education: Birth to Eight Years Old	<i>T S Lovelace, R Gardner III, J B Axe, and D M Sainato</i>	594
Early Intervention for Students with Special Needs	<i>P A Grant</i>	608
Early Transition of Children with Special Needs	<i>S A Fowler</i>	613
Educating Children Who are Deaf-Blind	<i>J Wood and K M Chinn</i>	621
Educating Children with Deafness and Hearing Impairments	<i>G M Mukuria and C J Eleweke</i>	628
Educating Children with Other Health Impairments (ADHD)	<i>G M Mukuria and F E Obiakor</i>	634
Educating Children with Physical Disabilities	<i>Z P Nkabinde, F E Obiakor, M T Ofor, and D J Smith</i>	640
Educating Children with Specific Learning Disabilities	<i>P A Grant and P B Grant</i>	646
Educating Children with Speech and Language Disorders	<i>S L Green and C D Qualls</i>	654
Educating Students with Autism and Related Disorders	<i>T T Dyches</i>	661
Educating Students with Cognitive Disabilities	<i>J P Bakken</i>	669
Educating Students with Emotional and Behavioral Disorders	<i>S W Smith and G G Taylor</i>	678
Educating Students with Special Needs: An Overview	<i>F E Obiakor, M K Harris, P Harris-Obiakor, and G M Mukuria</i>	688
Educating Students with Traumatic Brain Injury	<i>S Burkhardt and A Rotatori</i>	695
English Language Learners with Special Needs	<i>R Rueda and G Ragusa</i>	701
Functional Behavioral Assessment	<i>J P Bakken</i>	708
History of Special Education	<i>L M Garrick Duhaney and S J Salend</i>	714
Inclusion of Students with Special Needs in General Education Classrooms	<i>M A Prater</i>	721
Instructional Accommodations for Children with Special Needs in Inclusive Settings	<i>A L Yurick, S Musti-Rao, and G Cartledge</i>	727
Interagency Agreements	<i>L A Owens</i>	733
Interdisciplinary Collaboration in Early Childhood	<i>N File</i>	739
Language Development in Children with Special Needs	<i>M S L Stewart</i>	745
Large Scale Assessment and Accountability and Students with Special Needs	<i>M L Thurlow</i>	752
Literacy Instruction for Students with Special Needs	<i>L H Mason</i>	759
Math Instruction for Children with Special Needs	<i>B A Bottge</i>	767
Overrepresentation of Culturally and Linguistically Diverse Students in Special Education	<i>S C Trent</i>	774
Paraprofessionals in Special Education	<i>B Y Ashbaker, S C Dunn, and J Morgan</i>	780
Parent and Family Involvement in the Education of Children with Special Needs	<i>L P Prater</i>	788
Peer Relations and Socialization of Children and Adolescents with Special Needs	<i>L M Bullock, M Wong-Lo, and R A Gable</i>	794
Peer-Tutoring	<i>L J O'Shea and D J O'Shea</i>	802
Postsecondary Participation of Students with Special Needs	<i>R A Stodden and K D Roberts</i>	808
Prereferral Strategies	<i>C G Simpson and J P Bakken</i>	815

Preventing Antisocial Behavior and Delinquency: A Comprehensive School-wide Approach <i>P R Hallam, K R Young, P Caldarella, D G Wall, and L Christensen</i>	820
Programs and Instructional Strategies for Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	829
Related Services for Children with Special Needs <i>S O Obi</i>	834
Risk and Resilience Frameworks in Understanding Special Education <i>W-Y Hsieh and M L Donahue</i>	840
School Dropout Prevention <i>F E Obiakor</i>	845
School-Based Services for Children with Special Needs <i>S O Obi</i>	850
Self-Determination <i>M L Wehmeyer</i>	857
Social Competence <i>L A Owens and S Johnston-Rodriguez</i>	865
The Identification of Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	870
Transition from School to Adult Life <i>M J McCollin and F E Obiakor</i>	876
Visual Impairment <i>W Sapp</i>	880
Workforce Issues in Special Education <i>E Carlson and B Billingsley</i>	886

VOLUME 3

Education of Professionals

An Overview of Accreditation, Certification, and Licensure Processes <i>J J Norcini and M van Zanten</i>	1
Educating America's Military Officers <i>M H Adamshick</i>	7
Education for Architecture in the United States and Canada <i>S Schreiber</i>	13
Education for Medicine <i>M B Anderson</i>	19
Education for the Legal Profession <i>J W Wegner</i>	25
Education of Clergy <i>D O Aleshire</i>	31
Framing Professions Education <i>P S O'Sullivan</i>	37
Issues in Accounting/Business Education <i>W H Gijsselaers and R G Milter</i>	39
Professions – Engineering <i>J Heywood</i>	45
Professions – Health-Related <i>G M Jensen and C B Royeen</i>	51
Professions – Pharmacy <i>D E Beck</i>	58
Teaching for the Complex Practice of Nursing <i>P Benner and M Sutphen</i>	65
Veterinary Education <i>K M Edmondson</i>	71

Education Research Methodology: Quantitative Methods and Research

Abductive Research Methods <i>B D Haig</i>	77
Field Experimentation in Education <i>M A Constat</i>	83
Internal Validity <i>M H Clark and S C Middleton</i>	90
Methods for Approximating Random Assignment <i>B Schneider and S-K McDonald</i>	97
Methods for Approximating Random Assignment: Regression Discontinuity and Propensity Scores <i>L M Scheier</i>	104

Multitrait–Multimethod Designs	<i>M Bong and S Hong</i>	111
Multivariate Longitudinal Data Analysis	<i>H Liu and L Chang</i>	120
Quasi-Experimentation: Two Group Design	<i>E Hong</i>	128
Regression Discontinuity Designs	<i>A Kelava, S Rohrmann, and V Hodapp</i>	134
Sampling	<i>R S Brown</i>	142
Statistical Conclusion Validity	<i>P L Busk</i>	147
Survey Research Methods	<i>A Fink</i>	152

Educational Assessment

An Overview of Educational Assessment	<i>M James</i>	161
Assessment and the Evaluation of Institutional Effectiveness	<i>S M Thomas</i>	172
Assessment in Schools – Dispositions	<i>R D Crick</i>	181
Assessment in Schools Related To Literacy: Reading	<i>K J Topping</i>	189
Impact of Assessment on Students' Learning Strategies and Implications for Judging Assessment Quality	<i>M Segers, F Dochy, and D Gijbels</i>	196

Educational Assessment – Assessment Activities

Alternative Assessment	<i>G Stobart and C Gipps</i>	202
Classroom Assessment Tasks and Tests	<i>J Phelan and J Phelan</i>	209
Dynamic Assessment	<i>J G Elliott, E L Grigorenko, and W C M Resing</i>	220
Instructional System Provided Feedback	<i>H F O'Neil, S Chuang, and T Huang</i>	226
Peer and Self-assessment	<i>C Harrison</i>	231
Portfolio Assessment	<i>V Klenowski</i>	236
Records of Achievement: Beyond Traditional Tests	<i>P Broadfoot</i>	243

Educational Assessment – Assessment in Domains

Assessment in Higher Education	<i>D R Sadler</i>	249
Assessment in Schools – Affective Domain	<i>M Ventura, R D Roberts, and M Zeidner</i>	256
Assessment in Schools – Creative Subjects	<i>T E Costantino and L Bresler</i>	262
Assessment in Schools – Learning to Learn	<i>A Hautamäki, J Hautamäki, and S Kupiainen</i>	268
Assessment in Schools – Literacy Writing (Extended)	<i>J M Parr and L Earl</i>	273
Assessment in Schools – Mathematics	<i>M Brown and J Hodgen</i>	279
Assessment in Schools – Oracy	<i>A L Bailey</i>	285
Assessment in Schools – Primary Science	<i>W Harlen</i>	293
Assessment in Schools – Secondary Science	<i>J Shemwell, A Fu, M Figueroa, R Davis, and R Shavelson</i>	300
Assessment in Schools – Technology Education and ICT	<i>A Jones, B Cowie, and J Moreland</i>	311
Assessment in the Early Years	<i>M J Drummond</i>	316
Assessment in the History, Civics and Social Studies Domains	<i>P G Avery</i>	323
Assessment in the Workplace of Performance, Developing Expertise and Competence	<i>A J Brown</i>	330
Assessment in Vocational Education	<i>K Ecclestone</i>	337

Educational Assessment – Concepts and Issues

Assessing Group Work	<i>M Galton</i>	342
Assessment and the Regulation of Learning	<i>L Allal</i>	348
Challenges of Developing and Implementing Formative Assessment Practices in Schools	<i>R Sutton</i>	353
Formative Assessment	<i>P Black</i>	359
Formative Assessment and Instructional Planning	<i>B Marshall</i>	365
Progression and Assessment: Developmental Assessment	<i>M Forster and G Masters</i>	369
Students' Rights in Assessment and Evaluation	<i>J Gardner and B Holmes</i>	378
Summative Assessment by Teachers	<i>R Daugherty</i>	384
The Multiple Purposes of Assessment	<i>P E Newton</i>	392
The Relationship between Assessment and the Organization and Practice of Teaching	<i>L B Resnick, D William, R Apodaca, and E S Rangel</i>	397

Educational Assessment – Enhancing Quality and Use

Assessment Conversations: Reading and Writing Conferences with Students and with Parents	<i>M Barrs</i>	403
Assessment: Pre-service and In-service Teacher Education	<i>A Davies</i>	410
Classroom Assessment in Policy Context (Australia)	<i>J J Cumming</i>	417
Classroom Assessment in Policy Context (England and France)	<i>M Osborn and M Raveaud</i>	425
Classroom Assessment in Policy Context (French Sources)	<i>D Laveault</i>	432
Classroom Assessment in Policy Context (Hong Kong)	<i>D Carless</i>	438
Classroom Assessment in Policy Context (New Zealand)	<i>T J Crooks</i>	443
Classroom Assessment in Policy Context (Sub-Saharan Africa)	<i>S J Howie, L Zimmerman, and K Draper</i>	449
Moderation of Student Work by Teachers	<i>G S Maxwell</i>	457
School Policies and Practices to Support Effective Classroom Assessment for Learning	<i>D Pedder</i>	464
Student Assessment: Policy and Practice in Eastern Europe	<i>G Bethell</i>	472

Educational Assessment – Key Relationships

Assessment Practice in Policy Context: Latin American Countries	<i>F M Rizo</i>	479
Impact of Assessment on Learner Groups (Boys/Girls)	<i>J Elwood</i>	486
Impact of Assessment on Learner Groups (Disabilities)	<i>M Rouse</i>	492
Impact of Assessment on Students' Test Anxiety	<i>P Buchwald and C Schwarzer</i>	498
Impact of Assessments on Classroom Practice	<i>J L Herman</i>	506
Social Practices in School Assessment and their Impact on Learner Identities	<i>A Pollard and A Filer</i>	512

Educational Evaluation – Concepts, Practice, and Future Directions

Educational Evaluation: Concepts, Practice, and Future Directions	<i>Y J Ni</i>	518
---	---------------	-----

Educational Evaluation – Contemporary Themes in Evaluation

Evaluating E-Learning	<i>D D Williams and C R Graham</i>	530
Evaluating Schools as Learning Communities	<i>S M Thomas</i>	539
Evaluation of Adult Education and Training Programs	<i>G McNamara, P Joyce, and J O'Hara</i>	548
Evaluation of Integrated Health Programs in School	<i>T M Franke</i>	555

Formative Assessment in Teacher Education and Teacher Professional Development	<i>H Tillema</i>	563
Informal Education and Evaluation	<i>T Jeffs</i>	572
Methods to Evaluate Technology	<i>W L Bewley</i>	578

Educational Evaluation – Contexts of Educational Evaluation

Cost Analysis in Evaluation Studies	<i>J D Fletcher</i>	585
Cultural Issues that Can Affect the Validity of Educational Evaluations	<i>R P Durán</i>	592
Evaluation and Accountability	<i>E Sato and S N Rabinowitz</i>	600
Evaluation Governance and Planning	<i>R Picciotto</i>	607
Evaluation Use	<i>V G Thomas</i>	614
Moral and Ethical Issues in Evaluation	<i>M Morris</i>	622
The Role of Stakeholders in Educational Evaluation	<i>S Taut and M C Alkin</i>	629

Educational Evaluation – Evaluation Domains

Curriculum Evaluation	<i>T Levin</i>	636
Evaluating Education in Three Policy Eras	<i>M A Peters, R Britez, and S M Weber</i>	645
Evaluation of Mathematics Education Programs	<i>J Cai</i>	653
Evaluation of Science Education Programs	<i>G Berberoğlu</i>	660
Evaluation of Social Studies and Civic Education Programs	<i>S G Grant</i>	667
Evaluation of Teacher Quality and Practice	<i>W Robinson and J Campbell</i>	674
International Evaluations	<i>A Schleicher</i>	681
National Assessment Programs: The Example of the U.S. National Assessment of Educational Progress	<i>M Schneider</i>	690
Principal Evaluation: Concepts, Needs and Realities	<i>N S Glasman and L D Glasman</i>	698
Program Evaluation	<i>C A Christie and L A Fierro</i>	706
School Based Evaluation: Purposes, Protocols and Processes	<i>J MacBeath</i>	713
School Inspection/External School Evaluation	<i>C Chapman and P Earley</i>	719
The International System for Teacher Observation and Feedback: A Theoretical Framework for Developing International Instruments	<i>L Kyriakides, B P M Creemers, C Teddlie, and D Muijs</i>	726

Educational Evaluation – Evaluation Methods

Evaluation Methodology	<i>M Chatterji</i>	735
Evaluation Reporting and Communicating	<i>R T Torres</i>	746
Longitudinal Evaluation Designs	<i>X Ma</i>	757
Meta-evaluation: Purpose, Prescription, and Practice	<i>L A Wingate</i>	765

Educational Evaluation – Purposes and Models of Educational Evaluation

Defining Quality in Evaluation	<i>P Dahler-Larsen</i>	775
Internal and External Evaluation	<i>D Nevo</i>	781
Needs Assessment in Education	<i>J W Altschuld and T L Lepicki</i>	786
The Purpose of Educational Evaluation	<i>S Mathison</i>	792
Understanding Approaches to Evaluation	<i>A J Love</i>	798

VOLUME 4**Educational Measurement**

Ability Testing	<i>P B Tymms</i>	1
Admissions Testing	<i>R Zwick</i>	7
Appropriateness Measurement: Person Fit	<i>R R Meijer and A M Weekers</i>	15
Automated Essay Scoring: Writing Assessment and Instruction	<i>M D Shermis, J Burstein, D Higgins, and K Zechner</i>	20
Classical Test Theory Reliability	<i>M D Miller</i>	27
Criterion-Referenced Measurement	<i>S A Shrock and W C Coscarelli</i>	31
Differential Item Functioning	<i>M D Hidalgo and J Gómez-Benito</i>	36
Educational Measurement: Overview	<i>R L Linn</i>	45
Equating and Scaling	<i>A A von Davier</i>	50
Fairness	<i>C Gipps and G Stobart</i>	56
Generalizability Theory	<i>R L Brennan</i>	61
Impact of Assessment on Classroom Practice	<i>J L Herman</i>	69
Intelligent Systems	<i>V J Shute and D Zapata-Rivera</i>	75
Item Response Theory	<i>W J van der Linden</i>	81
National Assessments	<i>T J Crooks</i>	89
Norm-Referenced Measurement	<i>H D Hoover</i>	95
Progress Monitoring	<i>L S Fuchs and D Fuchs</i>	102
Rasch Models	<i>D Andrich</i>	111
Scoring Rubrics	<i>J A Arter</i>	123
Second Language Assessment	<i>L F Bachman</i>	140
Self Adaptive Testing	<i>J Revuelta</i>	148
Student Test Results in School Accountability	<i>L S Hamilton</i>	153
Test Development	<i>S M Downing</i>	159
Test Translation and Adaptation	<i>A Allalouf and P Hanani</i>	166
Testing Creativity	<i>M A Runco</i>	170
Testing in History	<i>S Wineburg and B Fogo</i>	175
Validity	<i>R L Linn</i>	181
Validity of Achievement Gains on High-Stakes Tests	<i>D Koretz</i>	186
Validity of Educational Indicators	<i>B P M Creemers and L Kyriakides</i>	193

G**Globalization of Universities and Institutions**

Globalization, Universities and Medium of Instruction	<i>J Lo Bianco</i>	201
The Washington Accord on Engineering Curriculum	<i>E Ko</i>	209

H

Higher Education

Higher Education: An Overview	<i>P Scott</i>	217
-------------------------------	----------------	-----

Higher Education – Management, Leadership and Governance

Academic Managers – Heads of Departments/Schools	<i>S Marshall</i>	229
Leading and Managing the University – Presidents and their Senior Management Team	<i>R Middlehurst, T Kennie, and S Woodfield</i>	238
Management of and in Higher Education Institutions	<i>B Sporn</i>	245
Managerialism and Collegialism in Higher Education Institutions	<i>M Shattock</i>	251
Strategic Planning in Higher Education	<i>L Zechlin</i>	256
The Changing Role of University Presidents, Vice-Chancellors and Rectors	<i>M F Green and P D Eckel</i>	264
The Changing Roles of University Governing Boards and Councils	<i>J Fielden</i>	273

Higher Education – Research and Scholarship

Funding of University Research	<i>G Harman</i>	279
Mobility of PhD Students and Scientists	<i>S Avveduto</i>	286
National Science Policy and Universities	<i>M Henkel and U Kogan</i>	294
Research Commercialization	<i>S Slaughter</i>	300
Research Quality Assessment	<i>I McNay</i>	307
The Management of University Research	<i>J Kirkland</i>	316
The Training of New Scientists	<i>R Neumann</i>	322

Higher Education – Society

Academic Freedom	<i>S F Macintyre</i>	328
Access and Equity in Higher Education	<i>G Baldwin and R James</i>	334
Boundaries of Institutional Autonomy and their Impact on Higher Education	<i>V L Meek</i>	341
Diversity of Higher Education	<i>U Teichler</i>	347
Employability of University Graduates and Graduate Outcomes	<i>J Brennan</i>	354
Higher Education and the Knowledge Society	<i>J Välimaa</i>	360
Higher Education and the Labor Market	<i>J Enders</i>	365
Higher Education and the Transformation of Society	<i>P Scott</i>	370
Quality Assurance in Higher Education – Practices and Issues	<i>D Dill</i>	377
The Role of Accreditation of Higher Education Institutions	<i>J S Eaton</i>	384
Universities and Regional Development	<i>J Goddard and J Puukka</i>	390
Universities' Engagement with Society	<i>D Watson</i>	398

Higher Education – Teaching and Learning in Higher Education

Distance Education and Open Universities	<i>A Kanwar and J Daniel</i>	404
Education Development in Higher Education	<i>Y Ryan and K Fraser</i>	411
Effective Use of Technology in Teaching and Learning in HE	<i>D Laurillard</i>	419

Facilitating Students' Success – Student Support and Learning Resources in Universities	<i>E Woodberry and C Barnes</i>	427
Faculty Values and Expectations and Their Implications for Teaching, Learning and the Organization of Higher Education Institutions	<i>K Harman</i>	433
Flexible Learning in Higher Education	<i>S Alexander</i>	441
Innovation in Teaching and Curriculum Design	<i>J M Dearn</i>	448
Productivity of University Faculty Staff	<i>S Kyvik</i>	455
Promoting Effective Student Learning in Higher Education	<i>K Trigwell</i>	461
Student Attrition from Higher Education Institutions	<i>M Hayden</i>	467
University Students' Experiences of Higher Education	<i>C McInnis</i>	473
What Do University Students Learn?	<i>G D Kuh and B K Bridges</i>	479

Higher Education – Tertiary Education in a Globalized World

Associations of Universities	<i>G F Moodie</i>	485
Collective Bargaining in Higher Education	<i>C Allport, K McAlpine, and S Roberts</i>	491
Financing of Higher Education	<i>B Chapman and P Tulip</i>	499
Higher Education Crossing Borders	<i>J Knight</i>	507
Higher Education in Post-Conflict Societies	<i>M Daxner</i>	514
Higher Education in the New Economies	<i>J Fielden</i>	520
Human Resource Issues in Higher Education	<i>E El-Khawas</i>	527
International Accreditation in Higher Education	<i>A Stella and D Woodhouse</i>	533
Internationalization of Higher Education	<i>M van der Wende</i>	540
National and International Rankings of Higher Education	<i>S Marginson</i>	546
Steering of Higher Education Systems – The Role of the State	<i>K Yokoyama and V L Meek</i>	554
Student and Faculty Transnational Mobility in Higher Education	<i>U Lanzendorf and B M Kehm</i>	559
The Bologna Process in European Higher Education	<i>D Rich</i>	566
The Economics and Finance of Higher Education	<i>A Wolf</i>	573
The Role of the OECD in the Development of Higher Education in a Globalized World	<i>R Yelland</i>	584
Transforming Higher Education in Developing Countries: The Role of the World Bank	<i>J Salmi and R M Bassett</i>	590
UNESCO's Role in the Development of Higher Education in a Globalized World	<i>S Uvalić-Trumbić</i>	597

Higher Education – The Organization and Financing of Higher Education

Higher Education in Federal Systems	<i>M A M Cortés and U Teichler</i>	603
Pathways and Articulation into Higher Education	<i>G F Moodie</i>	609
Unitary, Binary and other National Systems of Higher Education	<i>L Goedegebuure</i>	616
Viewing Private Higher Education: How Much, Where, Why, and What?	<i>D C Levy</i>	622

I

Immigration

Children of Migrant Populations	<i>M M Suárez-Orozco and C Suárez-Orozco</i>	629
Third Culture Kids	<i>R E Van Reken</i>	636

International Comparisons

IEA Studies in ICT	<i>N Law</i>	644
IEA Studies in Mathematics and Science	<i>F K S Leung</i>	650
IEA Study in Civic Education	<i>J Torney-Purta, J-A Amadeo, and J Schwillie</i>	656
IEA: Globalization and Assessment	<i>H Wagemaker</i>	663

International Organizations in Education

Education and the European Union	<i>S Hegarty</i>	669
National University on Educational Planning and Administration, New Delhi	<i>J B G Tilak</i>	676
On School Management – International Council on School Effectiveness and Improvement, Commonwealth Council on Educational Administration and Management	<i>A Walker and T Townsend</i>	681
UNESCO's International Institute for Educational Planning	<i>S J Hite and M Bray</i>	688

L**Leadership and Management**

Leadership and Management Overview	<i>B Mulford</i>	695
------------------------------------	------------------	-----

Leadership and Management – as a Field of Study

Contexts for Leadership at the Beginning of the 21st Century	<i>M Bottery</i>	704
Educational Leadership: Philosophical Issues	<i>C W Evers</i>	710
Ethical/Moral Issues in Educational Leadership	<i>R J Starratt, L Langlois, and P Duignan</i>	717
History of Educational Leadership/Management	<i>R Bates</i>	724
Educational Leadership as a Field of Study: Professional Journals 1994–2006	<i>A R Thomas</i>	731
Research on Educational Leadership – Approaches/Promising Directions	<i>W A Firestone and V M J Robinson</i>	740

Leadership and Management – Leader Preparation

Leaders for Productive Schools	<i>J Murphy, S N Elliott, E B Goldring, and A C Porter</i>	746
New Approaches in Preparing School Leaders	<i>S G Huber</i>	752
Research for Leaders: The National College for School Leadership	<i>G Southworth</i>	762
School Leadership Standards	<i>L Ingvarson</i>	768

Leadership and Management – Leadership Issues

Effective Leadership in Challenging Schools	<i>A Harris</i>	777
Esthetics, Educational Leadership and Management	<i>R Bates</i>	783
Leader Mistreatment of Teachers	<i>J Blase and J Blase</i>	790
Leadership and Gender	<i>J Blackmore</i>	797
Leadership and Student Voice	<i>C Lanskey and J Rudduck</i>	803
Leadership and Urban Education	<i>K A Riley</i>	809
Leadership for Learning	<i>J MacBeath</i>	817
Leadership in Diverse Cultures	<i>A Walker and P K Yu-kwong</i>	824

Leading Diverse Schools	<i>J Ryan</i>	832
Supervision and Personnel Administration	<i>P D Tucker and D G Pounder</i>	839
The Big Challenge: Leadership for Inclusion	<i>M Ainscow and A Sandill</i>	846
The Educational Leader's New Work	<i>P Gronn</i>	852

VOLUME 5

Leadership and Management – Leadership Types

A Distributed Perspective on School Leadership and Management	<i>J P Spillane</i>	1
Leadership: Authentic	<i>P T Begley</i>	7
Leadership: Democratic	<i>J Møller</i>	12
Leadership: Instructional	<i>P Hallinger, K Leithwood, and R H Heck</i>	18
Leadership: Transformative	<i>C M Shields</i>	26
Strategic Leadership	<i>B Davies and B J Davies</i>	34
Teacher Leadership and Organizational Development	<i>A Harris</i>	40
Understanding How Leadership Influences Student Learning	<i>K Leithwood and B Levin</i>	45

Leadership and Management – Politics and Governance

Equity and Educational Effectiveness	<i>P Sammons</i>	51
Leading/Managing Schools in Communities Made Poor	<i>P Thomson</i>	58
Principals Role in Restructuring Schools	<i>G M Crow and K D Peterson</i>	66
School Reform and Restructuring: Self Managing School	<i>B J Caldwell</i>	72
System Leadership	<i>D Hopkins and R Higham</i>	78

Leadership and Management – School and Community

Educational Leadership and Social Capital	<i>J West-Burnham</i>	85
Networked Learning Communities School-to-School Collaboration as an Essential Component of a System Reform Strategy	<i>D Jackson</i>	92
Networks and Communities of Knowledge	<i>T Schuller and H Theisens</i>	100
School Community Relations	<i>R L Crowson and E B Goldring</i>	106
Social Capital, Educational Institutions and Leadership	<i>S Kilpatrick, S Johns, and B Mulford</i>	113

Leadership and Management – School Effectiveness and Improvement

Leadership and Technology	<i>C F Webber</i>	120
Leadership in the Implementation of Innovations	<i>I E Bogotch, T Townsend, and M Acker-Hocevar</i>	128
Leadership: School Improvement	<i>R H Heck and P Hallinger</i>	135
Organizational Learning in Schools	<i>B Mulford and H Silins</i>	143
Professional Learning Community	<i>L Stoll</i>	151
Transformational School Leadership	<i>K Leithwood</i>	158

Learning and Cognition – Issues and Concepts

Attention in Cognition and Early Learning	<i>M Heimann, T Tjus, and K Strid</i>	165
Concept Learning	<i>S J Unsworth and D L Medin</i>	172
Cooperative Learning	<i>R E Slavin</i>	177
Intelligence	<i>R J Sternberg</i>	184
Learning from Multiple Information Sources	<i>H I Strømsø and I Bråten</i>	191
Memory	<i>S Magnussen and T Brennen</i>	197
Metacognition	<i>L Baker</i>	204
Personal Epistemology in Education: Concepts, Issues, and Implications	<i>I Bråten</i>	211
Piaget: Recent Work	<i>L Smith</i>	218
The Adult Development of Cognition and Learning	<i>E A L Stine-Morrow and J M Parisi</i>	225
Vygotsky and Recent Developments	<i>J V Wertsch</i>	231

Learning and Cognition – Issues and Concepts – Focus on Cognition

Cognition and Emotion	<i>F Pons, M de Rosnay, and F Cuisinier</i>	237
Cognition: Overview and Recent Trends	<i>J Wiley and B D Jee</i>	245
Development of Creativity	<i>M A Runco</i>	251
Knowledge Domains and Domain Learning	<i>L Maggioni and P A Alexander</i>	255
Problem Solving and Human Expertise	<i>T J Nokes, C D Schunn, and M T H Chi</i>	265
Problem Solving and Reasoning	<i>R E Mayer</i>	273
Reasoning, Moral and Social	<i>N S Jampol, C B Richardson, and M Killen</i>	279
Remembering as Social Activity	<i>D Middleton</i>	285

Learning and Cognition – Issues, Concepts, Types

Computer-Supported Collaborative Learning: Basic Concepts, Multiple Perspectives, and Emerging Trends	<i>S R Ludvigsen and A I Mørch</i>	290
Learning as Inquiry	<i>S R Goldman, J Radinsky, S Tozer, and D Wink</i>	297
Learning Through Narrative Socialization	<i>A Fasulo</i>	303
Neuroscience Bases of Learning	<i>M H Immordino-Yang and K W Fischer</i>	310
The Neuroscience of Reading	<i>U Goswami</i>	317

Learning and Cognition – Issues, Concepts, Types – Focus on Learning

Learning Strategies	<i>C E Weinstein, J Jung, and T W Acee</i>	323
Learning Through Play	<i>K Aronsson</i>	330
Situated View of Learning	<i>A Collins and J G Greeno</i>	335
The Neuroscience of Aging and Cognition	<i>W J Hoyer and B Bessette-Symons</i>	340

Learning and Cognition – Language, Literacy and Subject-Related

An Overview of Language and Literacy in Educational Settings	<i>V G Aukrust</i>	345
Bilingualism and Learning	<i>L S Verplaetse and E Schmitt</i>	355
Classroom Discourse and Student Learning	<i>L Hemphill</i>	361
Early Writing	<i>B E Hagtvet</i>	367

First Language Acquisition	<i>N B Ratner</i>	375
Foreign Language Learning	<i>C Kinginger</i>	382
Learning Science	<i>J Leach and P Scott</i>	387
Learning to Read	<i>M S Burns and J K Kidd</i>	394
Mathematics Learning	<i>L Verschaffel, B Greer, and E De Corte</i>	401
Music Learning	<i>S Hallam</i>	407
Reading Comprehension: Reading for Learning	<i>C E Snow</i>	413
Second Language Learning	<i>J S Arnfast, J N Jørgensen, and A Holmen</i>	419
Writing: Advanced	<i>R Andrews</i>	426

Learning and Cognition – Other Issues

Gender and Learning	<i>L Yates and J McLeod</i>	432
Parent–Child Relationships in Early Learning	<i>C P Edwards, S M Sheridan, and L L Knoche</i>	438
Peer Interaction and Learning	<i>S Blum-Kulka and S Dvir-Gvirsman</i>	444

Learning and Cognition – Places of Learning

Learning in a Cross-Cultural Perspective	<i>L-F Zhang and R J Sternberg</i>	450
Learning Outside of School	<i>T Nunes</i>	457
Reasoning and Explanatory Talk: Learning in Everyday Settings	<i>D E Beals</i>	464

Learning and Cognition – Theoretical Perspectives – Learning

Apprenticeship Approach to Learning	<i>K Nielsen</i>	469
Connectionism and Learning	<i>T R Shultz</i>	476
Constructivism and Learning	<i>S Sjøberg</i>	485
Cultural–Historical Activity Theory	<i>A Havnes</i>	491
Learning in a Sociocultural Perspective	<i>R Säljö</i>	498
Self-Regulated Learning and Socio-Cognitive Theory	<i>P H Winne and A F Hadwin</i>	503

N

National Systems of Education

Australia	<i>C N Power</i>	510
Austria	<i>M J Gaertner</i>	517
Bangladesh	<i>R Z Dewan</i>	524
Barbados	<i>R Rudder</i>	534
Bulgaria	<i>A Chavdarova and D Vesselinov</i>	540
England	<i>C Whatford</i>	550
Estonia	<i>M Keränen and J Kõrgesaar</i>	555
Fiji	<i>I Senibulu</i>	563
Finland	<i>J E Tuovinen</i>	569
France	<i>M Sellier</i>	576

Germany	<i>A Hüfner and K Hüfner</i>	582
Ghana	<i>R Palmer</i>	590
Greece	<i>K Toura and A Karagiannis</i>	599
Iceland	<i>M Hardardottir and M Gunnlaugsdóttir</i>	605
Indonesia	<i>A C Alwasilah and Furqon</i>	611
Israel	<i>A Volansky</i>	618
Jamaica	<i>E Hannam</i>	627
Latvia	<i>A Joma</i>	630
Lebanon	<i>N A Mneimneh</i>	636
Lithuania	<i>J Navickaitė, L Paurienė, R Zablackė, A Tijūnelytė, and V Kožemiakina</i>	656
Malawi	<i>F R Mkandawire and D Mulera</i>	666
Malta	<i>C Farrugia</i>	673
Mongolia	<i>B Yembuu</i>	681
Namibia	<i>J Ellis</i>	687
New Zealand	<i>R Baker and C Wylie</i>	692
Niger	<i>A A I Daouda</i>	699
Norway	<i>H Farstad</i>	709
Oman	<i>S A Issan</i>	718
Pakistan	<i>N Mahmood and R Malik</i>	727
Poland	<i>A Janowski</i>	737
Qatar	<i>S R Salman</i>	744
Russia	<i>V M Filippov and G F Tkach</i>	750
Rwanda	<i>E Bahizi</i>	758
Scotland	<i>D Raffe</i>	770
Singapore	<i>P T Ng</i>	776
Slovenia	<i>T Plevnik and A B Lakota</i>	782
Solomon Islands	<i>D P Malasa</i>	790
South Africa	<i>P Kgobe and J Pampallis</i>	799
Spain	<i>A Tiana</i>	806
Sri Lanka	<i>B M J Balasooriya, M M Wehella, and S U Wijeratne</i>	813
Suriname	<i>H R Ori</i>	826
Taiwan	<i>C-M Lee</i>	834
Tanzania	<i>R Wedgwood</i>	841
Thailand	<i>C Rukspollmuang</i>	847
The Gambia	<i>Y A-M Jobe</i>	858
The Philippines	<i>E A P Valenzuela</i>	865
Uruguay	<i>B Guinovart and C Troncoso</i>	873
USA	<i>B Pusser and M E R Lloyd</i>	888
Zanzibar	<i>I M Hija</i>	896

VOLUME 6**P****Philosophy of Education**

Philosophy of Education: Overview	<i>P Enslin</i>	1
-----------------------------------	-----------------	---

Philosophy of Education – Contemporary Issues

Neoliberalism, the Market and Performativity	<i>M A Peters</i>	11
Parents and Children's Rights	<i>M Clayton</i>	17
Peace Education	<i>I Gur-Ze'ev</i>	22
Philosophy and Educational Research	<i>D Bridges</i>	29
Race, Critical Race Theory and Whiteness	<i>B Applebaum</i>	36
School Choice and the Common School	<i>G Haydon</i>	44

Philosophy of Education – Philosophical Perspective

African Philosophy of Education	<i>K Horsthemke</i>	50
Feminism	<i>L Stone</i>	56
Hermeneutics	<i>C Bingham</i>	63
Islamic Education	<i>Y Waghid</i>	69
Liberalism and Education	<i>W Feinberg</i>	74
Phenomenology and Existentialism	<i>B Levering</i>	80
Pragmatism	<i>B Thayer-Bacon</i>	86
Psychoanalysis and Education	<i>P Smeyers</i>	92
Social and Cultural Capital in Education	<i>M Print</i>	97
The Aims of Education	<i>R Marples</i>	103
The Analytical Tradition	<i>D Carr</i>	108

Philosophy of Education – Philosophical Themes

Adult Learning: Philosophical Issues	<i>D Beckett</i>	114
Autonomy	<i>C A Wringe</i>	121
Bildung	<i>J Masschelein and N Ricken</i>	127
Critical Theory and Pedagogy	<i>M Mason</i>	133
Critical Thinking	<i>H Siegel</i>	141
Environmental Education	<i>M Bonnett</i>	146
Identity	<i>M S Merry</i>	152
Inclusion	<i>R Cigman</i>	158
Justice and Care	<i>M Tjiattas</i>	164
Lifelong Learning	<i>J Halliday</i>	170
Post-Structuralism and Education	<i>M Mason and M Clarke</i>	175
Reason and Rationality	<i>P Hogan</i>	183

Religious and Spiritual Education	<i>J C Conroy and S J McKinney</i>	189
The Capabilities Approach	<i>L Terzi</i>	194

Primary and Secondary Education – Learning and Teaching in School Age Education

Class Size – Arguments and Evidence	<i>P Blatchford and K C Lai</i>	200
Continuing Professional Development of Teachers	<i>P Earley</i>	207
Homeschooling	<i>P Farenga</i>	214
Innovations and Early Intervention in the Teaching of Literacy	<i>R E Slavin</i>	221
Personalized Learning in School Age Education	<i>D Hopkins</i>	227
Perspectives on Schooling in the Middle Years	<i>V J Russell</i>	233
Reform of the School Workforce	<i>H M Gunter and G Forrester</i>	242
Secondary Education in Developing Countries	<i>P Hughes and A Walker</i>	249

Primary and Secondary Education – Social Justice Issues in School Age Education

Community Focused Schools	<i>A Dyson</i>	255
---------------------------	----------------	-----

Primary and Secondary Education – The Effectiveness of School Systems

Reforming the American School System	<i>M Barber</i>	261
School Effectiveness in Developed Societies	<i>D Reynolds</i>	269
School to School Collaboration: Innovation and Improvement	<i>R Higham</i>	274
Studies of School Improvement in Developing Countries	<i>S E Anderson</i>	281
Sustainable Educational Reform	<i>A Hargreaves</i>	289
The Role of the District in Tri Level Reform	<i>M Fullan</i>	295
Whole School Designs for Enhancing Student Achievement	<i>S Stringfield and J Nunnery</i>	303

Q

Qualitative Research

Action Research in Education	<i>E Stringer</i>	311
Classroom Ethnography	<i>F Erickson</i>	320
Computer Assisted Qualitative Data Analysis	<i>C Silver and A Lewins</i>	326
Conversational Analysis	<i>H G Levine</i>	335
Critical Ethnography	<i>D Foley</i>	337
Critical Race Theory	<i>D Gillborn and G Ladson-Billings</i>	341
Critical Theory	<i>J J Ryoo and P McLaren</i>	348
Cultural Studies	<i>R Gaztambide-Fernández and K Cairns</i>	354
Cultural–Historical Activity Theory	<i>Laboratory for Comparative Human Cognition</i>	360
Design Experiments	<i>A Collins</i>	367
Discourse Analysis	<i>P Nikander and S Vehviläinen</i>	373
Document Analysis	<i>T Rapley and K N Jenkins</i>	380
Ethnography	<i>M Hammersley</i>	386

Ethnomethodology in Education Research	<i>D Macbeth</i>	392
Focus Groups	<i>D L Morgan and K Hoffman</i>	401
Grounded Theory	<i>K Charmaz and A Bryant</i>	406
Hermeneutics	<i>J Dyer</i>	413
Interpretive Research	<i>E Rauscher and M E Graue</i>	419
Interviews and Interviewing	<i>A Marvasti</i>	424
Life History	<i>J A Hatch and S K Newsom</i>	430
Narrative Inquiry	<i>D J Clandinin and J Huber</i>	436
Participant Observation	<i>J Schostak</i>	442
Phenomenology	<i>M van Manen and C A Adams</i>	449
Qualitative Case Studies	<i>S B Merriam</i>	456
Qualitative Data Management	<i>C Grbich</i>	463
Semiotics	<i>M Danesi</i>	475
Symbolic Interactionism, Naturalistic Inquiry, and Education	<i>J A Forte</i>	481
The Quality of Evidence in Qualitative Research	<i>S K Biklen</i>	488
Visual Data in Education Research	<i>J Wagner</i>	498

Qualitative Research – Standards of Proof

Validity: Mapping Diverse Perspectives	<i>G B Rossman, S F Rallis, and A M Kuntz</i>	505
--	---	-----

Quantitative and Qualitative Approaches to Research – Integration

Mixed Methods	<i>A Tashakkori and I Newman</i>	514
---------------	----------------------------------	-----

Quantitative Methods and Research Design

External Validity	<i>G E Matt, A Brewer, and M Sklar</i>	521
-------------------	--	-----

R

Role of NGOs in Globalization

NGOs and Globalization of Education	<i>M Ahmed</i>	529
-------------------------------------	----------------	-----

S

Social and Emotional Aspects of Learning

Achievement Goal Theory: Definitions, Correlates, and Unresolved Questions	<i>T Urdan</i>	537
Affect, Mood and Emotions	<i>S Schnall</i>	544
Anxiety	<i>M Zeidner</i>	549
Attribution Theory	<i>B Weiner</i>	558
Children's Friendship	<i>S R Asher and K L McDonald</i>	564
Coping with Stressful Situations: An Important Aspect of Self-Regulation	<i>M Boekaerts</i>	570
Culture in Motivation Research: A Challenging and Enriching Contribution	<i>K Kimmel and S Volet</i>	576
Early Social Development and Schooling	<i>G M van der Aalsvoort</i>	585

Emotion in Educational Contexts	<i>P A Schutz, P D Quijada, S de Vries, and M Lynde</i>	591
Empathy	<i>R Arnold</i>	597
Flow in Education	<i>J A Schmidt</i>	605
Interest	<i>M Ainley</i>	612
Intrinsic and Extrinsic Motivation	<i>C Levesque, K J Copeland, M D Pattie, and E L Deci</i>	618
Motivating Students in Classrooms	<i>J Brophy</i>	624
Motivation Regulation	<i>C A Wolters and S A Mueller</i>	631
Peer Learning in the Classroom	<i>N M Webb</i>	636
Poverty, Effects of on Social and Emotional Development	<i>N Philipsen Hetzner, A D Johnson, and J Brooks-Gunn</i>	643
Seeking Help as an Adaptive Response to Learning Difficulties: Person, Situation, and Developmental Influences	<i>S A Karabenick and R S Newman</i>	653
Self-Concept in Learning: Reciprocal effects model between academic self-concept and academic achievement	<i>H W Marsh and L F Scalas</i>	660
Self-Efficacy Beliefs	<i>D H Schunk and F Pajares</i>	668
Social and Emotional Outcomes of Learning	<i>J Vorhaus</i>	673
Social Aspects of Collaborative Learning	<i>M Arvaja and P Häkkinen</i>	685
Social Development and Schooling	<i>G W Ladd</i>	691
Social Interaction and Learning	<i>K Littleton</i>	698
Social Networks and the Education of Children and Youth	<i>B Schneider, T G Ford, and L Perez-Felkner</i>	705
Sociocultural Issues in Motivation	<i>R A Walker</i>	712
Socioemotional Aspects of Technology-Supported Learning	<i>S Järvelä</i>	718
Volitional Control of Learning	<i>J Husman and L Corno</i>	724
Wellbeing	<i>T Hascher</i>	732

VOLUME 7

Statistics

An Overview of Statistics in Education	<i>S Sinharay</i>	2
Analysis and Interpretation of Multivariate Data	<i>D J Bartholomew</i>	12
Analysis of Covariance	<i>N T Longford</i>	18
Analysis of Extreme Values in Education	<i>J Beirlant, G Dierckx, K Luwel, and P Onghena</i>	25
Analysis of Variance	<i>B M King</i>	32
Bayesian Statistical Analysis	<i>G Kokolakis</i>	37
Bootstrap Method	<i>K Singh and M Xie</i>	46
Canonical Correlation	<i>A Basu and A Mandal</i>	52
Categorical Data Analysis	<i>S E Fienberg and B W Junker</i>	58
Causal Inference	<i>D B Rubin</i>	66
Cluster Analysis: Overview	<i>S Landau and I Chis Ster</i>	72

Cognitive Psychology and Educational Statistics	<i>H S Kim and S Embretson</i>	84
Computational Statistics	<i>J E Gentle</i>	93
Continuous Probability Distributions	<i>S Sinharay</i>	98
Correspondence Analysis	<i>M Greenacre</i>	103
Data Mining	<i>R S J d Baker</i>	112
Decision Theory	<i>M Mendoza and E Gutiérrez-Peña</i>	119
Design of Experiments	<i>C A Albers and T R Kratochwill</i>	125
Discrete Probability Distributions	<i>S Sinharay</i>	132
Discrimination and Classification	<i>R C Neath and M S Johnson</i>	135
Empirical Bayes Methods	<i>E Ip and G Molenberghs</i>	142
Evaluation Research	<i>T Kellaghan</i>	150
Exploratory Data Analysis	<i>A Unwin</i>	156
Factor Analysis: An Overview and Some Contemporary Advances	<i>F A N Alhija</i>	162
Generalized Linear Mixed Models	<i>S Rabe-Hesketh and A Skrondal</i>	171
Generalized Linear Models	<i>J J Faraway</i>	178
Generating Random Numbers	<i>P J Pashley and A Amodeo</i>	184
Goodness-of-Fit Testing	<i>A Maydeu-Olivares and C García-Forero</i>	190
Graphical Models	<i>R G Almond</i>	197
Growth Modeling	<i>I Plewis</i>	203
Hierarchical Linear Models	<i>D Rindskopf</i>	210
Hypothesis Testing and Confidence Intervals	<i>T R Konold and X Fan</i>	216
Instrumental Variables	<i>E P Bettinger</i>	223
Jackknife Methods	<i>S Sinharay</i>	229
Large-sample Statistical Methods	<i>J M Singer and D F Andrade</i>	232
Latent Class Models	<i>J K Vermunt</i>	238
Markov Chain Monte Carlo	<i>A M Johansen</i>	245
Matrix Algebra	<i>H Yanai and H Ishii</i>	253
Measure of Association	<i>G-H Huang</i>	260
Measures of Central Tendency	<i>D Clark-Carter</i>	264
Measures of Dispersion, Skewness and Kurtosis	<i>D Rindskopf and M Shiyko</i>	267
Meta Analysis	<i>J Sánchez-Meca and F Marín-Martínez</i>	274
Missing Data	<i>C A W Glas</i>	283
Model Selection	<i>S J Haberman</i>	289
Monte Carlo Methods	<i>A M Johansen</i>	296
Multidimensional Scaling	<i>Z Zhang and Y Takane</i>	304
Multiple Comparisons	<i>C Lewis</i>	312
Multivariate Analysis of Variance	<i>N T Longford</i>	319
Multivariate Linear Regression	<i>S Ganesh</i>	324
Multivariate Normal Distribution	<i>J Tacq</i>	332

Nonlinear Regression Analysis	<i>H-H Huang, C K Hsiao, and S-Y Huang</i>	339
Nonparametric Statistical Methods	<i>K Sijtsma and W H M Emons</i>	347
Observational Studies	<i>P R Rosenbaum</i>	354
Order Statistics	<i>I Bairamov</i>	360
Point Estimation Methods with Applications to Item Response Theory Models	<i>F Bartolucci and L Scrucca</i>	366
Principal Components Analysis	<i>D J Bartholomew</i>	374
Probability Theory	<i>T Rudas</i>	378
Recursive Partitioning	<i>S Landau and S Barthel</i>	383
Sampling	<i>J Qian</i>	390
Sequential Probability Ratio Test	<i>T J H M Eggen</i>	396
Sequential Testing	<i>H J Vos</i>	401
Signal Detection Theory	<i>H Abdi</i>	407
Small Area Estimation	<i>A Noble</i>	411
Statistical Analysis of Functional Data	<i>F Ferraty, P Sarda, and P Vieu</i>	417
Statistical Inequalities	<i>H M Mahmoud and J L Gastwirth</i>	426
Statistical Paradoxes	<i>Y-K Tu</i>	429
Statistical Power Analysis	<i>L V Hedges and C Rhoads</i>	436
Statistical Significance Versus Effect Size	<i>X Fan and T R Konold</i>	444
Stochastic Processes	<i>F Baudoin</i>	451
Structural Equation Models	<i>S-Y Lee and X-Y Song</i>	453
Survival Data Analysis	<i>A F De Champlain</i>	459
The Normal Distribution and its Applications	<i>J Tacq</i>	467
Time Series Analysis	<i>P J Brockwell</i>	474
Univariate Linear Regression	<i>M L Hazelton</i>	482
Value-Added Models	<i>D F McCaffrey</i>	489

T

Teacher Education – Evaluation in the Teacher Education Program

Accreditation and Standards in Teacher Education	<i>R J Yinger and K L Daniel</i>	495
Determining Long Term Effects of Teacher Education	<i>C N Brouwer</i>	503

Teacher Education – Evaluation of the Prospective Teacher

The Evaluation of Prospective Teachers in Teacher Education	<i>D E H Tigelaar and J van Tartwijk</i>	511
---	--	-----

Teacher Education – Macro Factors

Diversity in Teacher Education	<i>T Wubbels</i>	518
Social and Cultural Influences on Teacher Education	<i>M Ben-Peretz and R Lotan</i>	525
The Role of Governance in Teacher Education	<i>G E Karlsen</i>	532

Teacher Education – Post-Initial Professional Development

Action Research as a Tool for Teachers' Professional Development	<i>P Ponte</i>	540
Contemporary Approaches to Teacher Professional Development	<i>H Borko, J Jacobs, and K Koellner</i>	548
School Development for Teacher Learning and Change	<i>P Sleegers and K Leithwood</i>	557
Teacher Induction	<i>D Beijaard, J Buitink, and C Kessels</i>	563
Teacher Learning as Workplace Learning	<i>J Imants and K van Veen</i>	569
Teacher Learning with Lesson Study	<i>A Murata</i>	575
Teachers Career Stages and Professional Development	<i>R Fessler and E Rice</i>	582

Teacher Education – Preservice Teacher Education

A Pedagogy of Teacher Education	<i>J Loughran</i>	587
Case Methods in Teacher Education	<i>M Sato and C Rogers</i>	592
Conceptions of Teacher Education	<i>J A Whitcomb</i>	598
Inquiry-Oriented Teacher Education	<i>R Schulz</i>	604
Narratives and Biography in Teacher Education	<i>G Kelchtermans</i>	610
Supervision in Teacher Education	<i>J McCarthy and L F Quinn</i>	615
Taking Prospective Teachers' Beliefs into Account in Teacher Education	<i>M Valcke, G Sang, I Rots, and R Hermans</i>	622
Teacher Education and Models of Teacher Reflection	<i>V K LaBoskey</i>	629
Teacher Education as Teaching for Understanding with New Technologies	<i>M S Wiske and D E Spicer</i>	635

Teacher Education – Preservice: The Knowledge Base

Experienced Teachers' Craft Knowledge	<i>P C Meijer</i>	642
Moral Values in Teacher Education	<i>W Veugelers</i>	650
Pedagogical Content Knowledge	<i>J H Van Driel and A Berry</i>	656
Teacher Education and the Educational Foundations Knowledge Base	<i>F B Murray</i>	662
The Relationship Between Theory and Practice in Teacher Education	<i>F A J Korthagen</i>	669

Teacher Education – Teacher Educators

Characteristics, Scholarship and Research of Teacher Educators	<i>M Lunenberg</i>	676
Professional Development of Teacher Educators	<i>P K Smith</i>	681
Self-Study by Teacher Educators	<i>T Russell</i>	689

Teacher Education – Teaching Specific Domains

Arts Teacher Education	<i>L Beudert</i>	695
Mathematics Teacher Education	<i>K Krainer and S Llinares</i>	702
Music Teacher Education	<i>M W-y Wong</i>	706
Science Teacher Education	<i>J H Van Driel and S K Abell</i>	712
Social Studies Teacher Education	<i>A-L Halvorsen and S M Wilson</i>	719
Teacher Education for Elementary Education	<i>K D Brown and G N Carnes</i>	725

Teacher Education for Teaching the Gifted	<i>A Y Baldwin</i>	732
Teacher Preparation in Mother Tongue Education	<i>M Kooy and H Hulshof</i>	738
Vocational Teacher Education	<i>S Johansson</i>	744

Teacher Education – The Cooperating School

Mentoring in Teacher Education	<i>P D Tomlinson, A J Hobson, and A Malderez</i>	749
Partnerships Between Schools and Higher Education	<i>C Day and L Smethem</i>	757

VOLUME 8

Technology and Learning

An Overview of Technology and Learning	<i>B Means and J Roschelle</i>	1
Educational Data Modeling	<i>S Iksal</i>	11
Student Data Systems and Their Use for Educational Improvement	<i>J C Wayman, V Cho, and M P Richards</i>	14
Technology and Learning: Access in Schools Around the World	<i>R Anderson and J Ainley</i>	21

Technology and Learning – as a Subject of Instruction

Conceptions of Technology Literacy and Fluency	<i>O Erstad</i>	34
--	-----------------	----

Technology and Learning – Assessment

Technology and Formative Assessment	<i>J W Pellegrino</i>	42
Technology for Large-Scale Assessment	<i>R E Bennett</i>	48
Technology Supports for Assessment Design	<i>R J Mislevy, F I Winters, I I Bejar, R E Bennett, and G D Haertel</i>	56

Technology and Learning – Education Reform and Economic Development

National Strategies to Build a Technology Workforce	<i>H-M Cheah</i>	66
Public–Private Partnerships for Educational Reform	<i>V Quah, C P Lim, and C Brook</i>	73
Relating Technology, Education Reform and Economic Development	<i>R B Kozma</i>	81

Technology and Learning – Evaluating Technology

Examining the Effects of Educational Technology Programs: Challenges and Strategies	<i>M Russell</i>	88
---	------------------	----

Technology and Learning – Global Trends

Equity in Technology Access and Opportunities	<i>T Matuchniak and M Warschauer</i>	95
Internet-Based Education	<i>M Bakia</i>	102

Technology and Learning – Issues

Media Use and School Achievement	<i>K Roe</i>	109
Technology and Physical and Social Health	<i>K Subrahmanyam</i>	112

Technology and Learning – Meeting Special Students’ Needs

Universal Design for Learning *D H Rose and J W Gravel* 119

Technology and Learning – School Leadership for Technology Integration

Technology as a Support for School–Community Connections *K M Culp* 125

Technology and Learning – Supports for Managing Instruction

Classroom Uses of Technology to Manage Instruction *W R Penuel* 133

Technology and Learning – Supports for Skill Learning

New Media, Learning from *W Schnotz and H Horz* 140

Reading and Technology *N DeFrance, D Khasnabis, and A S Palincsar* 150

Technological Supports for Acquiring Twenty-First-Century Skills *C Dede* 158

Technology Supports for Acquiring Inquiry Skills *T de Jong* 167

Technology Supports for Acquiring Mathematics *M J Nathan* 172

Technology Supports for Lifelong Learning *A Collins and R Halverson* 184

Technology and Learning – Supports for Student Collaboration

Knowledge Communities in the Classroom *J D Slotta and H Najafi* 189

Technology and Learning – Supports for Subject Matter Learning

Technology Supports for Science Learning *C Delgado and J Krajcik* 197

Technology Supports for Second Language Learning *D Nunan* 204

Technology and Learning – Teacher Learning

Teacher Skills and Knowledge for Technology Integration *N Law* 211

Technology in Preservice Teacher Education *N Davis* 217

Technology Resources for Teacher Learning *J E Hinojosa, C Labbé, and L López* 222

V

Vocational Education and Training

An Overview of Vocational Education and Training *T Karmel* 229

Informal and Non-Formal Learning in Vocational Education and Training *F Beddie and S Halliday-Wynes* 240

School-Based Vocational Education and Training *B Knight and R Sweet* 247

The Economics of Vocational Education and Training *P Carneiro, L Dearden, and A Vignoles* 255

Vocational Education and Training – Individuals and Students

Age Equality in Education and Training *P Taylor* 262

Career Guidance in Vocational Education and Training and in the Workplace *P Plant* 269

Displaced Workers, Unemployed and Vocational Education and Training *B Kriechel* 275

Immigrant Investment in Host Country Vocational Education and Training *I Kogan* 280

Indigenous People and Vocational Education and Training	<i>R G Schwab</i>	289
Persons with a Disability and Vocational Education and Training	<i>F Ferrier and C S Smith</i>	294
Prisoner Education and Training	<i>S Dawe, S H Wynes, and J Misko</i>	300
Vocational Education and Training and the School-to-Work Transition	<i>G Hayward</i>	306

Vocational Education and Training – Industry and Employers

Apprenticeships	<i>E Smith</i>	312
Skills Shortages: Concepts, Measurement and Policy Responses	<i>C Shah and G Burke</i>	320
Training and Learning in the Workplace	<i>R L Jacobs</i>	328
Training in Small Businesses	<i>H Matlay</i>	337

Vocational Education and Training – Research on VET

Language of Vocational Education and Training	<i>J Misko, J MacKenzie, and R-A Polvere</i>	343
Research on Vocational Education and Training	<i>J Robson and B Bailey</i>	352
Vocational Education and Training Information Resources	<i>L-A Harris, R Naidu, and M Saunders</i>	358

Vocational Education and Training – Teaching and Learning

Employability Skills	<i>E Smith</i>	368
Learning Styles in Vocational Education and Training	<i>J Vorhaus</i>	376
Modularization in Vocational Education and Training	<i>H Ertl and G Hayward</i>	383
Personalized Learning and Vocational Education and Training	<i>I Duckett</i>	391
Recognition of Prior Learning and Vocational Education and Training	<i>J Keating</i>	397
The Competency Model	<i>A Gonczi and P Hager</i>	403

Vocational Education and Training – VET in Context

Globalization and Vocational Education and Training	<i>P Loveder</i>	411
Learning to Work for the Future	<i>J Fien, R Maclean, and D N Wilson</i>	416
The Relationship between Technical and Vocational Education and Training and Innovation	<i>K Ruth and L Deitmer</i>	423
The Status of Vocational Education and Training	<i>J Lasonen</i>	429
Vocational Education and Training in Post-Conflict Countries	<i>E Lyby</i>	441

Vocational Education and Training – VET System

Dual System	<i>T Deissinger</i>	448
Industry Involvement in the Vocational Education and Training System	<i>M Hughes and C Hughes</i>	455
Non-Completion in Vocational Education and Training: Issues, Patterns and Possibilities	<i>S Lamb and A Walstab</i>	463
Planning and Policy Development for Technical Vocational Education and Training Systems	<i>R Maclean and M Pavlova</i>	469
Qualifications Frameworks and their Role in the Reform of Education and Training	<i>M Young</i>	476
Quality Assurance in Vocational Education and Training	<i>D Gunning</i>	482

Training Finance	<i>A Ziderman</i>	489
Training Markets	<i>P Ryan</i>	496
Vocational Education and Training Teacher Training	<i>S Nielsen</i>	503
Vocational Education and Training Workforce	<i>D L Wallin</i>	513
Subject Index		519
Contributors		747

A

ADULT EDUCATION

Adult Education Overview

K Rubenson, University of British Columbia, Vancouver, BC, Canada

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In the introduction to the second edition of the *International Encyclopedia of Adult Education and Training*, the editor points out that adult education is heuristic, multidisciplinary, eclectic in orientation, and that its knowledge base is diffuse and incomplete (Tuijnman, 1996, xv). If anything, this is even truer a decade or so later. Other disciplines and fields of study (such as economics, political science, psychology, sociology, and organizational sciences) increasingly take an interest in adult learning and contribute crucial knowledge to its theoretical and empirical knowledge base. So, it increasingly becomes difficult to precisely define what constitutes the core of adult education and learning. With this in mind, the adult education section of this encyclopedia attempts to present an overview of the still-emerging field of adult education. The 39 articles that make up the section provide insight into the historical development of the field, its conceptual controversies, domains and provision, perspectives on adult learning, instruction and program planning, outcomes, relationship to economy and society, and its status as a field of scholarly study and practice.

Toward a Field of Adult Education

Although J. H. Hudson's 1851 *The History of Adult Education* identified it as a unique field, only in the last 60 years has adult education been recognized as a distinct field of practice and study. Organized forms of education serving very limited elite groups of adults have been noted since ancient time. Plato's *Republic* presents a well-laid-out structure of what we today would call lifelong learning for the ruling class and outlines in detail what kind of education would be required at various stages in life from childhood up to

middle age. Similar examples are available from early civilizations in Egypt, China, and India where the technological advancements and complex administrative systems relied on education and training of a small leading segment of the adult population (Kidd and Titmus, 1989: xxiv). A different and later form of adult education has its roots in the Protestant reformation in Northern and Central Europe during the latter part of the sixteenth century and the first-half of the seventeenth century. During this period, adult education was influenced by translations of the Bible and invention of the printing press that among other things supported Bible-study groups and a reading public. The modern forms of adult education and learning have their roots in the modernization of industrial processes and the resulting far-reaching changes to society taking place in the late nineteenth and early twentieth century. Thus, the historical evolution of educational movements, the rise and fall of celebrated adult-education institutions as well as policies and reforms have to be seen in the context of the broader dynamics of social change and conflict. Presently, we can observe how regionally different models of lifelong learning policies are emerging as a response to changes in relatively unique cultural, economic, and political contexts.

The second part of the nineteenth and first part of the twentieth century saw a marked increase in educational activities for adults in Europe and North America. However, as Kidd and Titmus (1989: xxiv), note, these were perceived as discrete activities rather than part of a coherent field of adult education. By the early twentieth century, adult education had become the fastest-growing educational sector in the USA. This highlighted the urgency to form national associations of adult education and saw a gradual move to a professionalization of the field. The first study of adult education in the United States, initiated by the Carnegie Corporation in 1924, resulted

in the formation of the American Association for Adult Education (AAAE), in 1926. The purpose of AAAE was to advance lifelong learning, serve as a central forum for a variety of adult-education interest groups, influence local, state, and regional adult-education efforts, monitor legislation, conduct special studies, and maintain a speakers' bureau. One of the first activities of the AAAE was to start the publication of a *Handbook of Adult Education* series, the first being released in 1928. Since then, nine handbooks have been published, including the latest edition that is to be published in 2010.

In Europe, the formation of adult-education associations originally had less to do with developing a professional orientation but was instead aimed at bringing adult education to the working class and others, for example, women previously denied access to education. The primary examples of this are the mechanics institutes and the Association for the Higher Education of Working Men formed in 1903, and renamed as the Workers' Educational Association (WEA) in 1905. The WEA movement quickly spread to Australia, Canada, and New Zealand where adult education became seen as an essential tool for encouraging new immigrants to contribute to their new society. Folk high schools and study associations tightly connected to the labor movement and other social movements appeared in the Nordic countries. Germany and some other continental and Eastern European countries saw the development of various forms of *Volksbildung*. However, it was not before the later part of the 1960s that a more discernable and cohesive academic field of adult education started to take shape.

One sign of a maturing field of adult education has been the mushrooming of local, national, regional, and world organizations. From its early beginnings, a defining character of the evolving field has been its strong international dimension built around shared values and aspirations. This has positioned adult education as an international movement promoting adult education as a way to combat inequalities, support democracy, and promote cultural and social, democratic development (Duke, 1994). The first United Nations Educational Scientific and Cultural Organization (UNESCO) International Conference of Adult Education (CONFITEA) took place in 1949, and was followed by additional conferences in 1960, 1972, 1985, 1997, and 2009. The growth of the field can be illustrated by attendance at these conferences. In 1949, at the conference in Montreal, there was attendance from 25 countries, and in 1997, in Hamburg, the figure had increased to 130 and is expected to reach 175 at the 2009 conference.

Concepts

Since adult education began to emerge as a field, there have been constant discussions about what terminology to

use to describe the enterprise as well as about its definition and boundaries. The understanding of what constitutes adult education has not only changed over time, but also varied depending on cultural and institutional context. Up to the 1920s, the tradition had been to talk about specific activities like literacy for immigrants, university extension, mechanic institutes, workers' education, etc., but not to group these under an overall umbrella labeled adult education. As part of the professionalization of adult education taking place in the USA, in the first two decades, the term adult education came to be commonly used as a composite term to denote different activities relating to the education of adults. However, in most other parts of the world, the professionalization process came later and it was not until the 1960s that the generic term adult education was more broadly used. In many European countries, it had up until then been customary not to refer specifically to the education of adults but more generally to popular enlightenment as in France or the Nordic countries or *Volksbildung* as in Germany.

While the term adult education steadily won ground, there remained a lack of agreement on its definition and boundaries. Selman and Dampier (1991: 2) note that different terms are used for the same thing and people doing similar work may refer to this as community education, continuing education, adult training, literacy, extension, or adult education. Similarly, the terminology differs between country and country depending on historical circumstances. In fact, the organization of articles in this encyclopedia reflects the idiosyncrasies of what to include or not to include under the term adult education. Thus, to many in the field, several entries that are now presented in the volume on vocational education and training ought to have been included in the adult-education section and similar argument could be made for moving entries from the adult-education section to the one on vocational education and training.

Over the years, different definitions of the term adult education have appeared (see e.g., Bryson, 1936; Houle, 1972; Verner, 1962). The most commonly used definition presently is the one adopted by UNESCO in 1976 (UNESCO, 1976: 2) and reads in abbreviated form:

The term "adult education" denotes the entire body of organized educational processes, whatever the content, level and method, whether formal or otherwise, whether they prolong or replace initial education in schools, colleges and universities as well as in apprenticeship, whereby persons regarded as adult by the society to which they belong develop their abilities, enrich their knowledge, improve their technical or professional qualifications or turn them in a new direction and bring about changes in their attitudes or behavior in the two fold perspective of full personal development and participation in balanced and independent social, economic and cultural development;

adult education, however, must not be considered as an entity in itself, it is a sub-division, and an integral part of, a global scheme for lifelong education and learning.

Critical in the UNESCO definition, like in most other attempts to outline adult education, is how the term adult is best understood. Building on (human-development) stage theory, Darkenwald and Merriam (1982: 9) state: “adult education is a process whereby persons whose major social roles are characteristic of adult status undertake systematic and sustained learning activities for the purpose of bringing about changes in knowledge, attitudes, values and skills.” Thus, adult education refers to “activities intentionally designed for the purpose of bringing about learning among those whose age, social roles, or self-perception define them as adults” (Merriam and Brockett (1997: 8). According to this view, it is not age as such that defines someone as an adult, but the social roles that the person is carrying out. The second key dimension in the UNESCO definition relates to the special educational organization of the education for adults. It is in this context that adult education refers to “its own peculiar organization, methods and curriculum, which distinguishes adult education from any other field of education.” In the words of Verner (1964: 1), “... the term adult education is used to designate all those educational activities that are designed specifically for adults.” However, it has become increasingly difficult to maintain this kind of definition. First, not all adults taking part in adult education are adults in the terms of social roles and functioning. Second, it is commonly pointed out that participation in adult education is not always a volunteer act but it is more commonly becoming something an adult has to do to keep her work or become eligible for certain benefits like unemployment insurance. Third, attempts to separate adult learners from first-time students attending regular school or university are also becoming more blurred. The traditional pattern of study has changed and with an increasing number of students moving in and out of the educational system and the labor market, it is difficult to identify who is in the first cycle of studies and who is a recurrent learner.

While recognizing the problems of defining who is an adult learner, various pragmatic solutions are being sought. So, for example, recent studies like the International Adult Literacy Survey and the Adult Literacy and Lifeskills Survey (OECD, 2003, 2005) allow for the exclusion of all regular, full-time students, except the following: full-time students subsidized by employers; full-time students over 19 enrolled in elementary or secondary programs; and full-time students over 24 enrolled in post-secondary programs. While this approach may be pragmatic, it is evident that in the emergent learning society, the traditional distinctions between initial education, particularly higher education, and adult education are becoming increasingly blurred.

The 1976 UNESCO definition reflects the strong and growing calls to see adult education as a segment within an overall principle of lifelong learning. Ideas of lifelong learning are today having strong influence over adult-education policy and different regional models are developing (ibid). In this perspective, it is of interest to note that the UNESCO Institute of Education changed its name to UNESCO Institute of Lifelong Learning. This use of the concept departs from the fundamental principle of lifelong learning as put forth by transnational organizations like the European Union (EU), the OECD, and the UNESCO which proclaim that the concept of lifelong learning is based on three fundamental attributes:

- it is lifelong and therefore concerns everything from cradle to grave;
- it is lifewide recognizing that learning occurs in many different settings; and
- it focuses on learning rather than limit itself to education.

In adult-education circles, particularly in the Anglo Saxon countries, lifelong learning is given a more restrictive understanding and has increasingly come to be used interchangeably with just one segment of lifelong learning, namely adult learning. This is, for example, the case in the UK where departments of adult education have been renamed departments of lifelong learning. Similarly, one of the leading scholarly journals in adult education is named *International Journal of Lifelong Education* focusing almost exclusively on adult learning.

The embracement of lifelong learning, in its broad or restrictive meaning, is resulting in a shift from the concept of adult education to adult learning resulting in a further proliferation of the terminology in use. So, for example, today a distinction is being made between three basic categories of settings where purposeful learning activity takes place (European Commission, 2000):

1. *Formal learning.* This learning typically takes place in an education or training institution, it is structured (in terms of learning objectives, learning time, or learning support) and leading to certification. Formal learning is intentional from the learner's perspective.
2. *Nonformal learning.* It is learning that is not provided by an education or training institution and typically does not lead to certification. It is, however, structured (in terms of learning objectives, learning time, or learning support). Nonformal learning may be provided in the workplace and through the activities of civil-society organizations and groups. It can also be provided by organizations or through services that have been set up to complement formal systems, for example, arts, music, and sports classes. Nonformal learning is intentional from the learner's perspective.

3. *Informal learning.* It is learning resulting from daily life activities related to work, family, or leisure. It is not structured (in terms of learning objectives, learning time, or learning support) and typically does not lead to certification. Informal learning may be intentional but in most cases it is nonintentional (or incidental/random).

While policy documents overwhelmingly subscribe to definitions of adult learning that broadly correspond to those presented by the European Communities' policy documents, the scholarly literature contains many different and competing definitions and questions the advisability of trying to seek clear definitional distinctions between the three concepts. Others warn that the tendency to substitute learning for education can de-politicize the field and move the focus away from broader issues like equity, the role of the state, policy, and resources which are central when addressing issues of democracy and equality (Duke, 1994; Rubenson, 2006). In this respect, it is of interest to note that the CONFINTEA VI meeting in 2009 uses the label adult learning and education (ALE), a concept that may become prevalent in the years to come.

As the debate over how to delimit education continues, many students of the field will likely agree with Duke (1994: 8) when he states:

The important point about the concept field and scope of adult education is that it is necessarily broad, diffuse, multi-locational in terms of research sites and academic identities. It reflects the diffuseness and weakly bounded nature of adult education (and learning) itself.

To some, the concerns over a breakdown of the traditional boundaries of adult education and adult education as a profession are misplaced. Instead, they recommend that the focus should be on practitioners working professionally with adults whoever and wherever they may be (Usher *et al.*, 1997: 27).

Adult Education Domains and Provision

While the structure and provision of adult education vary considerably from country to country, some main sectors are discernable:

- adult basic education including adult literacy and numeracy;
- immigrant and citizenship education;
- adult higher education;
- workplace education and training (see section on vocational education and training);
- community education;
- popular adult education; and
- museums, radio, and TV and libraries.

When the AAAE was formed in 1926, adult education was generally understood in a narrow sense and expected to fulfill two main purposes. First, it was to teach English to immigrants and to prepare them for American citizenship, and second, to provide a second chance for those that had been deprived of education in their earlier years. These two purposes have remained central to adult education. As immigration spread more and more, countries started to devote significant human and material resources to citizenship and immigrant education. From focusing almost exclusively on preparing the new immigrants for the labor market and assimilating into the prevailing culture, programs are increasingly having the dual purpose of stimulating unity and diversity. In the 1960s and 1970s, the expansion of educational opportunities for the young, in combination with rising skill demands in the economy, brought basic adult education and literacy onto the policy agenda. With the emerging knowledge economy, functional adult literacy and the national pool of human competencies have gained a renewed interest by policymakers (OECD, 2005). In the developing world, adult literacy is promoted as central to development as well as to address issues of human immunovirus/acquired immunodeficiency syndrome (HIV/AIDS). Adult basic education and particularly adult literacy education display a great variation in terms of structure, provider, and philosophy. It can be found in churches, schools, colleges, community settings, workplaces, and libraries and offered by professional or unqualified staff.

The structure of the current university adult and continuing-education programs reflect their historical roots. In the UK and the Commonwealth, these are to be found in the adult and extra-mural tradition with its special centers or departments dedicated to offering university courses to adults. In the USA, on the other hand, the tradition emanated from the Land Grant universities with their explicit mission to support the local community and rural development. Reflecting an acceptance of the principle of lifelong learning, adult learners now make up a large proportion of students in higher education – in some countries they even constitute a majority. A large number of these are enrolled in regular courses or programs while others take a continuing-education course. Among the latter, an increasing number is found in continuing professional education which has increased dramatically with rising demands for mandatory continuing education in several professions. As with any educational system, continuing professional education has many stakeholders with multiple agendas that are being shaped by professional and social, institutional, and educational considerations (OECD, 2005). A new growing clientele for the university are older adults who take advantage of courses offered by continuing-education departments or specialized programs like Elderhostel or the University of the Third Age.

Information on enrolment in different forms of adult education reveals a dramatic shift in provision over the last

three decades. In the OECD countries, this is primarily caused by a remarkable increase in employer-supported activities that have radically altered the landscape of adult education (Bélanger and Valdivieso, 1997; Desjardins *et al.*, 2006). The data reflect the broader changes that have occurred in the labor market, which, among other things, forces people to participate because they are ordered or feel pressurized to undergo some form of adult education and training linked to their work (Carré, 2000). This education and training is being offered by a multiplicity of providers from different sectors including: formal educational institutions, commercial schools/private training providers, and employers and suppliers of equipment. The range and diversity of providers of workplace-related education contributes to the uncertainties about the reach and impact of public policy decisions on the structure of adult education and training.

Another distinct sector involves popular adult education where we today can find two dominant traditions; one Nordic and one Latin American though each have some followers in other parts of the world. Folk high schools and adult-education associations with roots in the classical social movements are a vital part of the adult-education sector particularly in the Nordic countries. In the latter countries, it is free and voluntary, despite considerable state and municipal subsidies; it lies at the crossroads between civil society and the state and has three major roles: it acts as an agency of popular movements, it is an adult educator, and it is a supporter of culture. It does not primarily cater to individuals' careers and job needs, but broadly responds to their role as citizens, parents, and/or personal development. In many parts of the world, popular adult education, while less structured than in the Nordic countries and lacking state subsidy, plays an important role in the life of various social movements. The Latin American tradition is closely associated with the work and pedagogy of Paulo Freire. Working closely with oppressed groups, the focus is often on improving their literacy and consciousness raising with an aim to inspire solidarity and collective political awareness and action. This tradition of popular adult education draws heavily on popular culture traditions, dance, song, drama, and storytelling. Its vitality can be seen in other areas of educational activity as well. For example, labor education, which can take many different forms, is sometimes an important avenue for popular education. Community education is a form of locally funded educational and recreationally program that can be seen to occupy a space between the formal educational system and popular adult education. This form of provision is particularly well developed in the UK, Ireland, Australia, and parts of the USA and also in several African countries, for example, South Africa. Traditions and structure of community education vary between countries and at times the role of the state in setting direction is more direct while in other traditions, for example, the so-called Wisconsin model, it follows a hands-off approach.

However, a common denominator is the rootedness in the community and the focus on preparing the community and its inhabitants to respond to structural and educational disadvantage.

The section on domains and provision has addressed actual structures of adult education, but in addition to these, it is worth mentioning a recent idea that focuses not on specific providers but considers a city or region as a learning entity. The concept can be seen as a political and social utopia that expresses what a city or region wants to become in the emerging knowledge society. Another current development is the general idea of prior learning assessment (PLA), which is about acknowledging and giving formal recognition to prior learning, irrespective of when, where, and how learning has taken place. PLA can be seen as a natural consequence of the acceptance of the principle of lifelong learning and breaks with the tradition of defining adult-education activities in term of provider and instead recognizes learning in whatever form it takes.

Adult Learning, Instruction, and Program Planning

The literature on adult education has primarily focused on what and how adults learn, instructional methods appropriate when working with adults, and how to design and organize educational activities for adults.

A hotly debated issue has been whether or not adult learning should be understood as a distinctive process that significantly differs from how children learn. In *The Modern Practice of Adult Education: Andragogy versus Pedagogy*, Knowles (1970) argues that adult learning differs fundamentally from how children learn. The book, which contains a set of principles that should guide adult learning practices, quickly became somewhat of a Bible for practitioners. Facing extensive criticism for the sharp division he had drawn between how adults and children learn, Knowles changed his position and suggested that rather than see a dichotomy between andragogy and pedagogy, they constituted a continuum. Consequently, in the 1980 edition of *The Modern Practice of Adult Education*, the subtitle was changed to read *From Pedagogy to Andragogy*. To many, even this softening of the original position does not go far enough and they challenge the idea of trying to build an academic field focusing distinctly on adult learning. This is the case among scholars working in the biographic perspective on adult learning who stress the continuities between early life and adult experience (Knowles, 1970). Others, while recognizing that adults' and children's learning from the point of psychological functioning may not differ, maintain that this position only holds for some very basic features of learning. Their key point is that in the planning of adult-education practices, one has to recognize that adult learning, in

contrast to children's learning, is principally selective and self-directed.

Andragogy, like two other dominant perspectives on adult learning, self-directed learning and transformative learning, is informed by humanistic psychology and focuses on the individual adult learner. Following the general developments in learning research, scholars in adult education have become more skeptical to the traditional strong individual orientation and have increasingly adopted a situated-cognition orientation. This is particularly noticeable in writings on workplace learning, an area that is prominent in recent adult education learning literature. The research on the situated nature of learning at work has helped shed light on why there often is less of a transfer between what has been learned in a school situation and the application of this knowledge and experiences at the workplace. The focus on workplace learning has stimulated an interest to understand not only how individuals learn but also how organizations themselves might hinder or encourage learning. A learning organization is being promoted as an ideal form of organization that fosters continuous organizational renewal through encouraging the learning of its members (Knowles, 1970). With the interest in organizational learning and situated-cognition workplace learning, scholars have embraced a network and community of practice traditions, more recently a co-participation model that encourages combined micro- and macro-level analysis of workplaces.

Outcomes of Adult Education

With increased public and private investment in adult education, there is a growing interest in developing a more comprehensive understanding of how adult education and learning can contribute economically and socially to the well-being of the individual and society. The two central questions are: who is participating in what kind of adult education and learning, and what benefits does this learning have for the individual and society.

Existing research shows that participation rates vary substantially across countries and that there are marked differences between countries with quite similar economies too. Regardless of overall participation rate, participation patterns are very similar. In all countries, older adults, the poor, and unemployed, low-skilled workers, migrants, ethnic minorities, and rural inhabitants report a reduced tendency to participate (Desjardins, *in press*). Gender inequalities vary between countries with some reporting a higher rate of participation among women and the others a lower. However, despite similar patterns of exclusion, the degree of inequality varies substantially among countries. Comparative findings suggest that policy does matter and that political commitment to high

standards of equity, and willingness to address market failures through sustained public policy efforts affect participation patterns. Thus, government intervention can affect a person's capability to participate through fostering broad structural conditions relevant to participation and construct targeted policy measures that are aimed at overcoming both structurally and individually based barriers (Rubenson and Desjardins, 2009).

The main individual outcomes of participation in adult education that have been examined are labor-market outcomes such as wages, employment, and earnings. For employers, the key outcome of adult education and training is worker productivity (44). The literature suggests that while there may be unaccounted-for factors, returns to training, particularly employer training and off-the job training are generally positive.

Further, there seem to be interesting differences between countries (*ibid*). Most studies tend to focus solely on skills and competencies for economic outcomes and neglect the relationships between various forms of adult learning and quality of life as well as overlook the synergy between formal, nonformal, and informal learning. Naturally, the impacts of adult education and learning extend beyond various economic outcomes and during recent years, there is a growing interest in the wider benefits of adult learning, particularly health and civic engagement. The work carried out so far suggests positive impact on health civic and political participation, but it has also identified a range of difficult methodological issues. One general problem is that situating outcomes of adult learning in the broader context of lifelong learning requires measures that allow comparisons across formal, nonformal, and informal learning settings, so that substitution and complementarity between the various forms of adult learning can be assessed. While participation in organized learning is quite restricted, almost everyone seems to be engaged in purposeful informal learning activities – but so what? What consequences flow from being involved in particular forms of adult learning? The shift in focus from adult education to lifelong learning has moved the attention away from inputs toward outputs.

As part of the professionalization process of adult education, scholars and practitioners began to pay serious attention to designing and organizing adult educational activities. Over time, the program planning literature in adult education, predominantly from the USA, has come to provide a rich body of competing models and theories on decisions and actions about what is to be learned, how the learning is to take place, who the learners and the teachers are, and what the education is for. Despite the great number of models developed for specific programs over the years, there seems to be more uniformity than could be expected, which has made some scholars to suggest the existence of a generic model (*ibid*; Sork and Cafarella, 1989). These reviews have also identified a

sense of a growing gap between what is prescribed in the theories and what practitioners actually do when planning programs. The explanation for this is to be found in the fact that the proposed planning theories are limited in their understanding of what actual planning practice is required, and instead are rooted in a form of instrumental rationality that privileges normative prescriptions of what planners should do.

Adult Education, Economy, and Society

A full appreciation of adult education and learning requires that they be seen in their socioeconomic, political, and cultural contexts. In a historical perspective, the idea of modernization provides an insight into the changing institutional realities and conceptual meanings of adult education and learning. For example, literacy education was needed to enable modern societies and a condition for socioeconomic development while popular and community education have had a perspective of social, cultural, and political self-articulation as well as social and political struggle (ibid). Further, recent stress on continuing-skills upgrading reflects changes in the reproduction of labor. Traditionally, skills upgrading occurred when younger more skilled workers replaced retiring workers. In the new economy, this is no longer a sufficient process to secure economic survival. Thus, contemporary policies on and approaches to adult education are driven by the expected role of adult learning in human-capital formation, particularly for the so-called knowledge economy.

While considering how different societies are responding to the stress of a globalized knowledge economy, we should remember Martin Carnoy's point that there are crucial differences in what adult education attempts to do and can do in different social-political structures. He states (Carnoy, 1995: 3):

Ultimately, these differences depend heavily on the possibilities and limits of the state, since it is the state that defines adult education and is the principal beneficiary of its effective implementation. These possibilities and limits of the state are, then, a key issue understanding the form and content of adult education.

Carnoy's point helps understand the importance of situating the present discussion on skills in the political economy of adult education and to take account of the social construction of skills. It also foregrounds the political purpose of adult education which becomes evident when juxtaposing adult education and nation building. The region of the Southern African development community provides an illustrative example of how adult education and nation building is influenced by local regional and global developments (Carnoy, 1995).

The political will of government is reflected in the extent to which, as well as how, adult education gets funding, the balance between public and market-driven funding mechanisms, and how this reflects on the social justice and distributional aspects of adult education and learning. Concerned about the latter, adult educators have critiqued the dominance of a strong market approach and governments' reluctance to address market failures in the learning market. In response to the dominant political economy, driving adult education there has been a revival of adult education for civil society which further illustrates of how socioeconomic, political, and cultural context inform the formation of adult education. Adult education, for civil society, can in Gramscian terms (Adamson 1980; Boggs 1976), be understood as a struggle to build a counter hegemony to the dominant discourse on adult education and learning. In countries affected by the collapse of the totalitarian communist states, or being freed from other forms of dictatorship, it is also a response to the need for a new form of citizenship.

There are particularly three aspects of justice with regard to adult education that concern adult-education scholars, class, gender, and race and ethnicity. Understandings of class have been central in explaining inequalities of opportunities, standards of living, and widening disparities between rich and poor. From an adult-education perspective, two issues stand out. First, how are adult education and learning practices linked to the globalized nature of capitalism and second, the operation of a distinctive working-class learning style (ibid). Similar questions are asked by feminist scholars engaged with gender analysis. Gender analysis that can reflect different theoretical perspectives is shaped by its history and provides a strategy for research process and strategic initiatives to address inequalities (ibid). Like class and gender, race and ethnicity play a major role in structuring society and the opportunities it provides. Reviews of the literature reveal somewhat surprisingly that the way race has been socially constructed has remained stable over the last 50 years. Further, it is noted that it was not until the last 15 years that the issue of race was given an in-depth analysis in the adult-education literature (ibid).

Field of Study

Since adult education began to emerge as a field of study in the late 1920s, it has undergone three quite distinctive phases. These phases are most noticeable in the USA, which to a large extent has come to define the nature of the scholarly field, but they are also clearly discernible in Europe and to a lesser extent in some other parts of the world.

As the demand for trained instructors to teach the large number of immigrants grew, a few US universities

began offering specific courses on how to teach adults, starting with the University of Columbia in 1922 (Milton *et al.*, 2003). Quite soon, graduate programs began to emerge. The first began at Teachers College at Columbia University in 1930, and by 1964, 16 universities in the USA offered master and doctoral programs in adult education (Houle, 1991). A similar development took place in the UK, in departments of extra-mural studies offering university courses, where adults developed an interest in adult-education research and in 1926, the first chair in adult education was established at the University of Nottingham (Hake, 1994).

With a small but growing number of adult-education programs, faculty started to focus on how to generate a body of knowledge that would help in the growth of the evolving field. Guided by funding from the W. K. Kellogg Foundation, the Commission of Professors in the USA set out to define the conceptual foundations of adult education (Jensen *et al.*, 1964). Officially titled, *Adult Education: Outlines of an Emerging Field of University Study* and published in 1964, their work has come to be known as the Black Book. This publication can be seen as ushering in the second phase of adult education, characterized by a major expansion of graduate programs and the coming of age of adult education as a field of study. Between the release of the Black Book and the publication of its follow up, *Adult Education: Evolution and Achievements in a Developing Field of Study*, in 1991, the number of adult-education graduate programs in the USA increased from 16 to 124 (Houle, 1991). Another indication of the growing knowledge base of adult education was the launch in 1969 of the yearly American Adult Education Research Conference (AERC) and the development of *Adult Education Quarterly* – arguably, the preeminent journal in the field.

Similar developments took place in Canada and in parts of Europe. In the former, the first graduate program was established at the University of British Columbia in 1957 and by the late 1980s, there were ten graduate programs in the country (Selman and Dampier, 1991: 255). In the UK and Germany, the number of departments of adult education increased, first slowly in the 1950s and 1960s, and thereafter more frequently in the 1970s and 1980s. Sweden, with a very long tradition of adult education, introduced special funding for adult-education research at the end of the 1960s, and university programs and chairs sprang up in Finland, the Netherlands, Germany, France, Poland, the former Yugoslavia, and somewhat later in several other European countries. In 1991, the European Society for the study of adult education was formed.

In other parts of the world, particularly in the developing countries, the process begun later and many countries are in what can be seen as the first phase. This is the case in several African and Latin American countries. In some instances, like in Brazil, there is an

acceleration of programs and departments specializing in adult education. In China, the first MA program was launched at East China Normal University in 1993 and a PhD program in 2004. In 2008, China reports to have some 100 specialized institutions for adult-education research and is moving into what can be seen as the second phase.

The 1991 review of the evolution and achievements in the developing field of adult education (Peters and Jarvis, 1991) can be read as a summary of situation in North America and Northern Europe at the end of phase 2. The book paints a very positive picture noticing the significant expansion in the knowledge base, lessened dependency on related disciplines, broadening of research methodologies, exponential growth of graduate programs, and that the field has become more internationalized. Based on the achievements so far, the book ends on a positive outlook and with expectations of continuous growth and solidifying of the field of adult education over the coming 25 years. While there does not exist any comprehensive review of what has happened, since the 1991 book, there are several indications that the field of study has not progressed as anticipated and that it has entered into a new phase in its development. In North America and those parts of Europe where the field had expanded and matured during the second phase, the last two decades have not seen a continuing growth in specialized adult-education departments. Instead, the trend has been to amalgamate adult-education programs with other areas into larger departments or in some instances to close them down. These developments have been driven by a combination of forces external and internal to the field of adult education. First, the amalgamations have been part of a general restructuring of university departments into larger structures, which tends to hamper the building of a field of adult education as foreseen under phase 2. Second, the embracement of the principle of lifelong learning also weakens the ambition of building a separate field of adult education. Third, with workplace learning being by far the fastest-growing area of adult-education practice, adult-education researchers have increasingly come to be engaged in research on workplace learning. Over the last decade, workplace learning has increasingly taken on the shape of an alternative field with its own journals and scholarly conferences. While this development can weaken the field as adult educators break out of the traditional boundary of their discipline, it also provides an opportunity to strengthen the field as researchers from other disciplines are encouraged to break into adult education. The focus on the so-called knowledge economy has stimulated other disciplines, economics, sociology, organizational sciences, etc. to engage with issues that traditionally would have been seen to fall under adult education. Fourth, as in other fields of education, there is a general shift away from more pure fields and disciplines

and a move to organize the knowledge-generation process around cross-cutting themes like gender, immigration, etc. that are approached from an interdisciplinary or multidisciplinary stand. In this environment, adult-education scholars increasingly develop multiple alliances housed in different academic fields.

Tensions Within the Field of Study

Since adult education embarked on expanding its knowledge base, two issues have continued to cause tension: its relationship to other disciplines and link between practice and theory.

The Black Book was organized around the understanding that:

Adult education develops a unique body of knowledge suited to its purposes through two methods of procedures:

- (1) Experiences gained from coping with problems of practice lead to the formulation of principles or generalizations which provide guides for future practice.
- (2) Knowledge which has been developed by other disciplines is borrowed and reformulated for use in adult education. (Jensen, 1964: 105)

The Black Book allocates a chapter each to scrutinize how adult education can borrow and adopt knowledge, theory, and research traditions from the disciplines of sociology, social psychology, psychology, history, and organization and administration. As the field matured, the position presented in the Black Book began to be questioned. Instead of relying on other disciplines, the trend, particularly in the USA, was to primarily rely on adult-education literature (Boshier and Pickard, 1979). To build too closely on other disciplines was seen by some as a threat to the development of an adult-education knowledge base (Boyd and Apps, 1980; Kranjc, 1987). Some went as far as Plecas and Sork (1986) who argued against copying scholars in other disciplines and for adult-education scholars to remain focused on research that is closely related to well-established definitions of adult education. However, while warning adult educators to be cautious in their borrowing, it was generally recognized also within the US adult-education community that much could be learned from other disciplines (Peters, 1991). Outside the USA, the push toward a clearly defined field of adult education was, with a few exceptions like the former Yugoslavia, less strong. Here, the dominant position has been that other disciplines can provide the conceptual apparatus for a better understanding of the structure, the functioning, and problems of adult education as scholars worked to strengthening research on the education of adults rather than building a distinct field of study. A good example of this can be found in the German series *Handbuch der Erwachsenenbildung* where the first

volume appeared in 1974. Of special interest are vol. 3, *Anthropologie und Psychologie des Erwachsenen* (Zdarzil and Olechowski, 1976) and vol. 6, *Sociologie der Erwachsenenbildung* (Eggers and Steinbacher, 1977).

The second issue in the field of study stems from the fact that knowledge production in adult education has overwhelmingly been shaped by a stress on practicality. Consequently, the dominant view has been that theory is deemed useful only to the extent that it improves practice. Using this criterion, there has been a constant criticism of the limited relevance of the research enterprise for the practice of adult education. There are very varied views on the cause and remedy of the problem. One camp of adult-education scholars has argued that research has been influenced too much by the rituals of science and not enough by the needs of the learners, while others take a diametrically opposite position seeing the lack of influence being a consequence of research having concentrated too much on applied problems (Cevero, 1991: 26). Sork and Cafarella (1989) suggest that the gap, which has been present since the outset of adult education becoming a field of study, was widening rather than shrinking. This could be an outcome of the calls, during the late 1970s, for the field to become more theoretically sophisticated so that it might gain more respect in the scholarly world. In a response to this call, university departments of adult education began recruiting new faculty, into adult education, who often had less connection to the field of practice than the outgoing faculty.

Concern about the usefulness of adult education research is still flourishing today. However, not only practitioners in the field but also policymakers voice their disappointment. Thus, it is interesting to note that the criticism of adult-education research for a lack of usefulness is a dominant theme in the national reports from developing as well as developed countries prepared for the 2009 UNESCO CONFITEA meeting. Developing countries point to the need for research to more directly support initiatives focusing on reducing poverty, addressing HIV, and the strengthening of women's role, while developing countries talk about the need for knowledge to support evidence-informed policy agendas. The European summary report to CONFITEA speaks of the need "for a research interface to promote the use of research results in policy development and implementation" (Keogh, 2008: 10). The Arabic Summary report notes the lack of research dealing with literacy (Yousif, 2009) while the report from Latin America (Torres, 2009) speaks of the lack of dissemination of adult-education-research results outside academic circles. While we have to remember that the national reports have been produced by governments, in many cases in cooperation with the adult-education community, they reflect a special understanding of the relationship between research and practice and policy.

Concluding Comment

As evident from this wide-ranging overview of theories, policies, practices, institutions, and scholarship in adult education and learning, the field is very broad and does not allow itself to be neatly organized within strict boundaries. While scholars are engaged in heated debates on what constitutes the field of adult education and/or how the scholarly field should evolve, adult learning has become a way of life for one-third to 50% the population in the industrialized world and is increasingly spreading in the developing world. When responding to these challenges, practitioners, policymakers, as well social activists can fall back on a rich history of practice and scholarship in adult education.

See also: Adult Basic Education: A Challenge for Vocational Based Learning; Adult Education and Civil Society; Adult Education and Nation-Building; Adult Learning; Adult Learning and Instruction: Transformative Learning Perspectives; Adult Learning in a Biographic Perspective; Adult Learning, Instruction and Programme Planning: Insights from Freire; Adult Literacy Education; Barriers to Participation in Adult Education; Characteristics of Adult Learning; Citizenship and Immigrant Education; Class Analysis in Adult Education; Continuing Professional Education: Multiple Stakeholders and Agendas; Financing of Adult and Lifelong Learning; Gender Analysis; Informal Learning: A Contested Concept; Labor Education; Lifelong Learning; Modernization Processes and the Changing Function of Adult Learning; Museums as Sites of Adult Learning; Organizational Learning; Overview of Lifelong Learning Policies and Systems; Participation in Adult Learning; Popular Adult Education; Program Planning; Provision of Prior Learning Assessment; Race and Ethnicity in the Field of Adult Education; Rewriting the History of Adult Education: The Search for Narrative Structures; The Age of Learning: Seniors Learning; The Political Economy of Adult Education; Trends in Workplace Learning Research; University Adult Continuing Education: The Extra-Mural Tradition Revisited; Wider Benefits of Adult Education; Workplace Learning Frameworks.

Bibliography

- Adamson, W. L. (1980). *Hegemony and Revolution: A Study of Antonio Gramsci's Political and Cultural theory*. Berkeley, CA: University of California Press.
- Bélanger, P. and Valdivieso, S. (eds.) (1997). *The Emergence of a Learning Society. Who participates in Adult Learning?* Oxford: Pergamon Press.
- Boggs, C. (1976). *Gramsci's Marxism*. London: Pluto Press.
- Boshier, R. and Pickard, L. (1979). Citation patterns of articles published in adult education 1968–1977. *Adult Education* 30(1), 34–51.
- Boyd, R. D. and Apps, J. W. (1980). *Redefining the Discipline of Adult Education*. San Francisco, CA: Jossey-Bass.
- Bryson, L. (1936). *Adult Education*. New York: American Book.

- Carnoy, M. (1995). Foreword: How should we study adult education. In Torres, C. A. (ed.) *The Politics of Nonformal Education in Latin America*, pp 9–16. New York: Praeger.
- Carré, P. (2000). Motivation in adult education: From engagement to performance. *Proceedings of the 41st Annual Adult Education Research Conference*, Vancouver, University of British Columbia.
- Cervero, R. M. (1991). Changing relationships between theory and practice. In Peters, J. M. and Jarvis, P. (eds.) *Adult Education: Evolution and Achievements in a Developing Field of Study*, pp 19–41. San Francisco, CA: Jossey-Bass.
- Darkenwald, G. and Merriam, S. (1982). *Adult Education: Foundations of Practice*. New York: Harper and Row.
- Desjardins, R., Rubenson, K., and Milana, M. (2006). *Unequal Chances to Participate in Adult Learning: International Perspectives*. Paris: UNESCO IIEP Fundamentals of Educational Planning Series.
- Duke, C. (1994). Research in adult education – current trends and future agenda. In Mauch, W. (ed.) *World Trends in Adult Education Research*, pp 7–12. Hamburg: UNESCO Institute of Education.
- Eggers, P. E. and Steinbacher, F. J. (eds.) (1977). *Soziologie der Erwachsenenbildung*. Stuttgart: W Kohlhammer.
- European Commission (2000). *A Memorandum on Lifelong Learning*. Luxembourg: Office for Official Publications of the European Commission.
- Hake, B. (1994). Adult education research trends in the Western European countries. In Mauch, W. (ed.) *World Trends in Adult Education Research*, pp 124–153. Hamburg: UNESCO Institute of Education.
- Houle, C. O. (1972). *The Design of Education*. Madison: University of Wisconsin Press.
- Houle, C. O. (1991). Foreword. In Peters, J. M. and Jarvis, P. (eds.) *Adult Education: Evolution and Achievements in a Developing Field of Study*, pp 13–17. San Francisco, CA: Jossey-Bass.
- Jensen, G. (1964). How adult education borrows and reformulates knowledge of other disciplines. In Jensen, G., Liveright, A. A., and Hallebeck, W. (eds.) *Adult Education: Outlines of an Emerging Field of University Studies*, pp 105–111. Washington, DC: American Association for Adult and Continuing education.
- Jensen, G., Liveright, A. A., and Hallebeck, W. (1964) (eds.) *Adult Education: Outlines of an Emerging Field of University Studies*. Washington, DC: American Association for Adult and Continuing education.
- Keogh, H. (2009). The state and development of adult learning and education in Europe, North America and Israel. *Regional Synthesis Report*. [http://www.unesco.org/en/confinteavi/preparatory-conferences/europe-and-north-america/Helen Keogh](http://www.unesco.org/en/confinteavi/preparatory-conferences/europe-and-north-america/Helen%20Keogh) (accessed October 2009).
- Kidd, J. R. and Titmus, C. J. (1989). Introduction. In Titmus, C. J. (ed.) *Lifelong Education for Adults. An International Handbook*, pp 23–39. Oxford: Pergamon.
- Knowles, M. (1970). *The Modern Practice of Adult Education Andragogy Versus Pedagogy*. Chicago, IL: Association Press.
- Kranjc, A. (1987). Research in adult education: Major areas of theory and inquiry. In Duke, C. (ed.) *Adult Education: International Perspectives from China*, pp 86–99. London: Croom Helm.
- Merriam, S. and Brockett, R. (1997). *The Profession and Practice of Adult Education*. San Francisco, CA: Jossey-Bass.
- Milton, J., Watkins, K. E., Studdard, S. S., and Michele Burch, M. (2003). The ever widening gyre: Factors affecting change in adult education graduate programs in the United States. *Adult Education Quarterly* 54(1), 23–41.
- OECD (2003). *Beyond Rhetoric: Adult Learning Policies and Practices*. Paris: OECD.
- OECD (2005). *Promoting Adult Learning*. Paris: OECD.
- Peters, J. M. (1991). Advancing the study of adult education: A summary perspective. In Peters, J. M. and Jarvis, P. (eds.) *Adult Education: Evolution and Achievements in a Developing Field of Study*, pp 421–445. San Francisco, CA: Jossey-Bass.
- Peters, J. M. and Jarvis, P. (eds.) (1991). *Adult Education: Evolution and Achievements in a Developing Field of Study*. San Francisco, CA: Jossey-Bass.
- Plecas, D. B. and Sork, T. J. (1986). Adult education: Curing the ills of an undisciplined discipline. *Adult Education* 37(1), 48–62.

- Rubenson, K. (2006). The nordic model of lifelong learning. *Compare: A Journal of Comparative Education* 36(3), 327–341.
- Rubenson, K. and Desjardins, R. (2009). The impact of welfare state regimes on barriers to participation in adult education. A bounded agency model. *Adult Education Quarterly* 59(3), 187–207.
- Selman, G. and Dampier, P. (1991). *Adult Education in Canada*. Toronto: Thompson Educational.
- Sork, T. J. and Cafarella, R. S. (1989). Planning programs for adults. In Merriam, S. M. and Cunningham, P. M. (eds.) *Handbook of Adult and Continuing Education*, pp 233–245. San Francisco, CA: Jossey-Bass.
- Torres, R. M. (2009). Youth and adult education and learning in Latin America and the Caribbean: Trends, issues and challenges. *Regional Synthesis Report*. <http://www.unesco.org/en/confinteavi/righnavigation/download/> (accessed October 2009).
- Tuijnman, A. C. (1996). Preface. In Tuijnman, A. C. (ed.) *International Encyclopedia of Adult Education and Training*, pp 15–21. Oxford: Pergamon.
- UNESCO (1976). *Recommendations on the Development of Adult Education*. Paris: UNESCO.
- Usher, R., Bryant, I., and Johnston, R. (1997). *Adult Learning and The Postmodern Challenge: Learning Beyond Limits*. London: Routledge.
- Verner, C. (1962). *Adult Education Theory and Method: A Conceptual Scheme for the Identification and Classification of Processes*. Washington, DC: Adult Education Association of the USA.
- Verner, C. (1964). Definition of terms. In Jensen, G., Liveright, A. A., and Hallebeck, W. (eds.) *Adult Education: Outlines of an Emerging Field of University Studies*, pp 27–39. Washington, DC: American Association for Adult and Continuing Education.
- Yousif, A. A. (2009). The state and development of adult learning and education in the Arab States. *Regional Synthesis Report*. <http://www.unesco.org/en/confinteavi/righnavigation/download/> (accessed October 2009).
- Zdarzil, H. and Olechowski, R. (eds.) (1976). *Anthropologie und psychologie des Erwachsenen*. Stuttgart: W Kohlhammer.

Further Reading

- Brown, P., Green, A., and Lauder, H. (2001). *High Skills: Globalization, Competitiveness and Skill Formation*. Oxford: Oxford University Press.

ADULT EDUCATION – ADULT LEARNING, INSTRUCTION AND PROGRAM PLANNING

Contents

Adult Learning

Adult Learning and Instruction: Transformative Learning Perspectives

Adult Learning in a Biographic Perspective

Adult Learning, Instruction and Programme Planning: Insights from Freire

Characteristics of Adult Learning

Informal Learning: A Contested Concept

Organizational Learning

Program Planning

Workplace Learning Frameworks

Adult Learning

S B Merriam, University of Georgia, Athens, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

Adult learning is a phenomenon at once deceptively simple, yet enormously complex. It is simple because we know that learning “is of the essence of everyday living and of conscious experience; it is the process of transforming that experience into knowledge, skills, attitudes, values, and beliefs” (Jarvis, 1992: 11). However, it is also complex because there is no one definition, model, or theory that explains how adults learn, why adults learn, or how best to facilitate the process. Yet the learning of adults is the key theme that unites the otherwise widely disparate field of adult education. Whether in community-based literacy classes, training sessions in corporate settings, or continuing professional education seminars, practitioners share the common goal of facilitating adult learning. Rather than a single definition or description of adult learning, what we have is a colorful mosaic of theories, models, sets of principles, and explanations that, when combined, form the knowledge base of adult learning.

Until the mid-twentieth century, what we knew about adult learning was embedded in studies by behavioral and cognitive psychologists, studies that focused on problem solving, information processing, memory, intelligence, and motivation. Much of this research was conducted in laboratory settings, and if adults were included, what was of interest was how advancing age affected the learning activity. Thorndike *et al.*'s (1928) *Adult Learning* published in 1928 is an example of this early research. This book reports the results of adults being tested in a laboratory under timed conditions on various learning and memory

tasks. The authors concluded that adults aged between 25 and 45 could learn “at nearly the same rate” as 20-year-olds (p. 178). Research in the 1940s found that when time pressure is removed, adults up to age 70 did as well as younger people.

Adult learning from a psychological, and in particular a behaviorist perspective shaped adult learning research and theory building in North America until the late 1960s when other traditions and European influences broadened inquiry. We now know quite a bit about the individual adult learner, how context shapes adult learning, and how noncognitive factors play a role in adult learning. This article is thus organized into three sections, in a loosely chronological order, reflecting our growing understanding of adult learning. The first part of this article explores the foundational adult learning theories of andragogy, self-directed learning, and transformational learning. A second strand of theory building represents a shift in the focus of learning from the individual to the context in which learning takes place. The third section of this article presents the most recent additions to our understanding of adult learning. These perspectives go beyond the cognitive and include the role of emotions, body, and spirit in learning.

The Individual Adult Learner

By the mid-twentieth century, adult education was a recognized field of practice with its own professional

associations, journals, and conferences. Rather than extrapolating from research with children or research that placed adults under the same conditions as children, adult educators began to consider how learning in adulthood could be distinguished from learning in childhood. Humanistic psychology provided the philosophical underpinnings for three theories of adult learning, which have become foundational to the field of adult education – andragogy, self-directed learning, and transformational learning.

Andragogy

The European concept of andragogy was introduced by Knowles (1968) as “a new label and a new technology” distinguishing adult learning from children’s learning or pedagogy (p. 351). Probably the best-known set of principles or assumptions to guide adult learning practice, andragogy actually tells us more about the characteristics of adult learners than about the nature of learning itself. Knowles originally presented the following four characteristics or assumptions about adult learners:

1. As a person matures, his or her self-concept moves from that of a dependent personality toward one of a self-directing human being.
2. An adult accumulates a growing reservoir of experience, which is a rich resource for learning.
3. The readiness of an adult to learn is closely related to the developmental tasks of his or her social role.
4. There is a change in time perspective as people mature – from future application of knowledge to immediacy of application. Thus, an adult is more problem centered than subject centered in learning. (Knowles, 1980: 44–45).

In later publications, Knowles also suggested a fifth and sixth assumption:

5. The most potent motivations are internal rather than external (Knowles *et al.*, 1984: 12).
6. Adults need to know why they need to learn something (Knowles, 1984: 12).

Working from these assumptions, Knowles (1980) proposed a program-planning model for designing, implementing, and evaluating educational activities with adults. For example, with regard to the first assumption that as adults mature they become more independent and self-directing, Knowles suggested that the classroom environment be one of adulthood, both physically and psychologically. Adults who plan and direct their family, work, and community lives can also participate in their own learning by assisting in diagnosing their learning needs, planning and implementing learning activities, and evaluating those experiences.

At first heralded as the explanation of adult learning, andragogy underwent intense examination by educators

of both adults and children. It was recognized, for example, that some children and adolescents are independent, self-directed learners while some adults are highly dependent on a teacher for structure and guidance. Further, adults may be externally motivated to learn as when an employer requires attendance at a training program, and some children may be motivated by curiosity or the internal pleasure of learning. By 1980, Knowles had acknowledged that the dichotomy between andragogy and pedagogy was not as stark as originally drawn. A clear indication of his rethinking is represented in the subtitles of the 1970 and 1980 editions of *The Modern Practice of Adult Education*. The 1970 subtitle is *Andragogy Versus Pedagogy*, whereas the 1980 subtitle is *From Pedagogy to Andragogy*. He came to believe that there was a continuum ranging from teacher-directed (pedagogy) on the one end, to student-directed learning (andragogy) on the other, and that both approaches are appropriate with children and adults, depending on the situation.

Andragogy has been most severely critiqued for its assumption that the individual adult learner is autonomous and in control of his or her learning. Lacking is any recognition that both the learner and the learning that takes place are shaped by a person’s history and culture in conjunction with the institutional context where it occurs. Despite these critiques, practitioners who work with adult learners can intuitively connect with Knowles’ characteristics of adult learners and can see how they translate into concrete suggestions for program planning, instruction, and evaluation. More than 40 years after it was first proposed in North America, andragogy enjoys widespread recognition as one understanding of adult learning and a tested guide to working with adults in practice.

Self-Directed Learning

Appearing in North America about the same time that Knowles introduced andragogy, the concept of self-directed learning (SDL) also helped distinguish adult learners from children. The major impetus for this model of adult learning came from Tough’s (1971) research with Canadian adult learners. He found that 90% of the participants in his study had engaged in an average of 100 h of self-planned learning projects in the previous year and that this learning was deeply embedded in their everyday lives. The uncovering and documenting of SDL – learning that is widespread, that occurs as a part of adults’ everyday life, and that is systematic yet does not depend on an instructor or a classroom – has been a major contribution toward understanding and defining adult learning.

More than 35 years of research in North America and Europe on SDL has verified its widespread presence among adults, documented the process by which it occurs, and developed assessment tools to measure the extent of

individual self-directedness. Of these foci, the process of SDL speaks most directly to adult learning. How one actually moves through an SDL experience has generated a number of models of the process. The earliest models proposed by Tough (1971) and Knowles (1975) are the most linear, moving from diagnosing needs to identifying resources and instructional formats, to evaluating outcomes. Models developed in the late 1980s and 1990s are less linear and more interactive in which not only the learner, but the context of the learning and the nature of the learning itself are also considered. In the Danis (1992) model, for example, learning strategies, phases of the learning process, the content, the learner, and the environmental factors in the context must all be taken into account in mapping the process of SDL. The Spear and Mocker (1984) model considers the opportunities for learning found in one's environment, past or new knowledge, and chance occurrences. These opportunities cluster into the "organizing circumstance" which in turn, structures the SDL activity, and the "circumstances created during one episode become the circumstances for the next" (p. 5).

Yet another application of SDL is in instruction in formal educational settings. The most popular is the Grow (1991) model. He presents a matrix showing how four types of learners and four types of facilitators intersect with appropriate instructional methods. For example, a dependent learner (one who is not at all self-directed) is a good match with an authority or expert, and lecture and drill are appropriate instructional strategies. At the other extreme would be a highly self-directed learner, matched with a facilitator or delegator and instruction would be embedded in independent projects and discussions.

As with andragogy, SDL has proven to be a mainstay of adult learning theory. Recent applications of SDL include its role in lifelong learning and continuing professional education, how SDL can be acknowledged and incorporated into the workplace, and how being self-directed is one criterion for success in online learning environments (Merriam *et al.*, 2007).

Transformational Learning

The third contribution to adult learning that helped define what is different about learning in adulthood is transformational learning. Rather than focusing on the learner as andragogy and, to a large extent, SDL do, transformational learning is about the cognitive process of meaning making. It is particularly an adult learning theory because transformational learning is dependent on adult life experiences and a more mature level of cognitive functioning than found in childhood. The essence of transformational learning is that through sudden or dramatic experiences, people are changed in ways that they themselves and others can recognize.

Mezirow (2000) is considered the primary architect of transformational learning, although he readily acknowledges being influenced by the Brazilian educator, Paulo Freire. Freire (1970) emphasized the need for this type of learning to deal with oppression and to bring about social change. Mezirow (2000) focuses more on the process of individual transformation, a process that is personally empowering. Learning in adulthood is not just adding on to what we already know, although that is part of the story. It is also, according to Mezirow (1996: 162), "the process of using a prior interpretation to construe a new or a revised interpretation of the meaning of one's experience in order to guide future action." In short, learning is also making sense of our experiences. Learning can result in a change in one of our beliefs or attitudes, or it can be a change in our entire perspective. A perspective transformation is key to transformational learning.

Mezirow (2000) delineated a 10-step transformational learning process that is initiated by a disorienting dilemma – a life experience that cannot be accommodated by one's present worldview. This leads the adult to examine and critically reflect on the assumptions and beliefs that have guided meaning making in the past, but now are no longer adequate. From an examination of current beliefs, the learner moves to exploring new ways of dealing with the dilemma, often in conjunction with others confronting a similar crisis. It is in dialog with others that the learner tests out new assumptions, understandings, and perspectives. A plan of action is then formulated and put into motion. The new or transformed perspective is more inclusive and accommodating than the previous perspective.

Since the 1990s, transformational learning has moved center stage in terms of the volume of research and writing. Transformational learning conferences occurring every 2 years, with the most recent in 2007, have also contributed to the burgeoning knowledge base about this type of learning. Further, connections between transformational learning and adult development (Merriam and Clark, 2006), and transformational learning and spirituality (Tolliver and Tisdell, 2006), have expanded our understanding of adult learning and the meaning-making process.

In summary, andragogy, SDL, and transformational learning have come to define much of adult learning today. Both andragogy and SDL were instrumental in distinguishing adult learning from childhood learning at a time when the field of adult education was defining itself. They remain dominant in the real world of practice, perhaps because of their humanistic foundations and the fact that they capture what is popularly and intuitively understood about adult learning. Transformational learning, though powerful and emancipatory when it occurs, is more difficult to plan for, implement, and assess.

The Context of Adult Learning

Andragogy, SDL, and especially transformative learning theory focus on the individual learner; indeed, each has been critiqued for not recognizing how the context where this learning occurs also shapes the learning. Attention to context became prominent in the later decades of the twentieth century and remains central to understanding adult learning today. One perspective that attends to context draws from critical social science and related perspectives such as Marxism, critical theory, multiculturalism, critical race theory, queer theory, and feminist theory. What this literature has in common is the relentless questioning of power relations embedded in the structures of society. The focus is on the context of learning, not the individual learner. Through questioning and critique, the *status quo* is challenged, leading hopefully to social change.

More congruent with educational psychology is a second perspective that also shifts the focus from the individual to the context where learning takes place. Emerging from cognitive psychology and known loosely as situated cognition or contextual learning, learning cannot be understood as simply an individual, internal cognitive process; rather, learning is what is constructed by the interaction of people in a particular situation with particular tools or artifacts (including technology, language, signs, and symbols). Research (Lave, 1988; Lave and Wenger, 1991) has demonstrated that the context in which learning takes place is crucial to the nature of the learning, as are the tools in that setting, and the social interaction with others. Understanding human cognition means examining it in situations of authentic activity in which actual cognitive processes are required, rather than the simulated ones typical of school. Lave's (1988) experiments with grocery shoppers is a good example of the difference. Comparison pricing was found to be considerably more accurate in the activity of shopping (98% error-free) than in doing identical calculations on a paper-and-pencil test in the classroom (59% error-free).

The notion of situated cognition resonates well with what we already know about adult learning. Fenwick (2003) points out that one cannot separate the learning process from the situation in which it takes place. Knowledge is constructed in the context; it is "part of the very process of *participation* in the immediate situation" (Fenwick, 2003: 25; italics in the original). Learning takes place when people interact with the community (including its history and cultural values and assumptions), "the tools at hand," and the activity itself (Fenwick, 2003: 25). For example, in a recent study of Korean older adults in an intermediate computer literacy course, questions were asked about the cultural context of the course, the social-interaction patterns of the participants, the tools of the

setting, and how learning is constructed through the interaction of these elements (Kim, 2008).

Locating learning in the real-life experiences of adults has long been promoted as a good adult education practice. Schon (1987, 1996), for example, is noted for promoting contextually based reflective practice. Knowledge gained in school is not enough to make a reflective practitioner. One must also engage in the actual practice. Others recommend apprenticeships, internships, and practicums where one can learn through modeling, coaching, and trial and error.

An important component of situated cognition is entering into relationships with other learners, thereby becoming a member of a learning community. This learning community can be considered a community of practice (Lave and Wenger, 1991; Wenger, 1998). Communities of practice are groups of people who share insights and ideas and who help each other solve problems and develop a common practice. All people belong to communities of practice, whether through formal learning environments, civic organizations, or family structures. While most communities of practice do not have a name, they are quite familiar to us. We know who belongs. The concepts of practice are both explicit and tacit. It includes the language, documents, images, symbols, roles, procedures, regulations, subtle cues, rules of thumb, sensitivities, embodied understandings, underlying assumptions, and shared worldviews that are crucial to the success of the community. In a study of a community of practice of Wiccans, an earth-based faith group, it was discovered that learning was embedded in the rites and rituals of their practice; learning in practice was experiential, combined formal and intuitive knowledge, and was spread across the group (Merriam, *et al.*, 2003).

Communities of practice as a learning theory has extended the work on situated cognition. Situated cognition posits that learning is context bound, tool dependent, and socially interactive. These factors suggest looking at this type of learning from the perspective of a bounded system. A family, a classroom, a profession, an online community, a town, and a corporation can all be thought of as a community of practice, or a learning community. This approach contextualizes learning, uncoupling it from a preoccupation with the individual learner.

Emotions, Body, and Spirit in Adult Learning

The mind/body split so ingrained in Western notions of learning has dominated adult learning. In addition, learning has become so connected with schooling that the activity of learning is almost always framed from a rational, cognitive perspective. We learn through processing information in the brain. By the time we are adults, learning that is valued,

formal, and systematic is devoid of anything emotional or physical. However, some of the most recent research and theory building in adult learning are based on the premise that knowledge construction and learning can be through pathways other than those that depend on the mind. Scholars are now trying to explain and legitimize the role played by emotions, body, and spirit in learning.

Emotions and Somatic Knowing

Knowledge construction is more than a cognitive process of meaning making. In fact, there is little cognitive about this – rather, we know through our emotions and our physical body. Dirkx (2001) argues that learning itself is inherently an imaginative, emotional act and that significant learning is inconceivable without emotion and feelings. It is through emotions that deeply personal, meaningful connections are made so that really significant learning can take place. These connections are of two kinds. First, there is the connection to one's own inner experiences; "emotions are gateways to the unconscious and our emotional, feeling selves" (p. 69); second, "emotions and feelings can connect to the shared ideas within the world as well and are reflected in big words or concepts, such as Truth, Power, Justice, and Love" (p. 69). We learn to understand or make meaning of our experience through engagement with these emotions and the images they evoke.

Somatic or embodied learning is closely related to emotional responses in learning. In somatic knowing, we can learn through our bodies, as we do when we connect physical manifestations of stress to our psychological situation. Pert (1997) in fact argues that since receptors are found in the body's nerves of all kinds, it would then follow that emotions could be stored and mediated by parts of the body other than just the brain. "These recent discoveries are important for appreciating how memories are stored not only in the brain, but in a *psychosomatic network* extending into the body" (p. 141). In fact, the interconnectedness of body, brain, and emotions is itself receiving attention through the neuroscience of learning (Johnson and Taylor, 2006).

It is clear that a false dichotomy has been created by the Western philosophical bias that dissects the whole person into mind and body, limiting knowledge construction to what goes on in one's mind. Even physiologically, the mind, body, and emotions cannot be separated. Certainly, in our own real-life experiences of living and learning, we involve our emotions and our body at least as much as our intellect.

Spirituality and Learning

Part of the difficulty in considering spirituality in learning has been definitional. There is little consensus about the

boundaries of its meaning; the most writers can do is to define it as they are using the term. All agree that spirituality is not the same as religion, which is an organized community of faith; rather, spirituality is more about one's own beliefs and experience of a higher power or higher purpose. Spirituality is "about how we construct meaning, and what we individually and communally experience and attend to and honor as sacred in our lives" (Tisdell, 2003: 29).

Spirituality is connected to adult learning through the construct of meaning making. Aktouf (1992: 415) argues that "the human being is, by definition and necessity, a being whose destiny is meaning, intentions, and projects, a subject whose being is meaning and which has need of meaning." We are inveterate meaning makers. Tisdell (1999) makes several points about the relationship between spirituality, meaning making, and adult learning. First, educators should recognize that a search for or an acknowledgement of the spiritual in the lives of adult learners "is connected to how we create meaning in our relationships with others. It is in our living and loving. It is also connected with how we understand a higher power or a transcendent being" (p. 93). Second, adults come into our classroom with this agenda (meaning making), whether or not it is articulated. Third, meaning making is knowledge construction that uses images and symbols, "which often emanate from the deepest core of our being and can be accessed and manifested through art, music, or other creative work" (p. 93).

Those writing from this more holistic perspective on learning are not about promoting a particular form of embodied or spiritual learning. Rather, they are "committed to learning that makes a difference in learners' lives and increases their sense of knowing the content of the course in their heads, their hearts, their souls, and their entire being—that has meaning to them and makes a difference in the world" (Tolliver and Tisdell, 2006: 45). Journal writing, poetry, storytelling, myths, symbols, images, and even dreams can be used in an adult learning environment to foster a more holistic learning experience. Indeed, an entire volume of *New Directions for Adult and Continuing Education* is devoted to 'Artistic ways of knowing' (Lawrence, 2005). In this volume, authors speak to the use of art, music, poetry, photography, and drama to "extend the boundaries of how we come to know" (p. 3).

Summary

Learning in adulthood defies a simple explanation, and it is highly unlikely that there will ever be a single theory or explanation that encompasses all that we know now about adult learning. There is a substantial body of research and literature dating back to the early decades of the twentieth century where adult learning was conceived of as problem solving, memory, and information processing. From that foundation, adult educators began to differentiate adult

learning from pre-adult learning, a move that led to a focus on the adult learners themselves. Andragogy, SDL, and, more recently, transformational learning are the major distinguishing aspects of adult learning today.

In addition, our understanding of adult learning has been expanded to include consideration of the larger sociocultural and political context in which it takes place, and how the context itself both shapes and is an integral part of the learning transaction. The analytical tools of situated cognition and critical perspectives have allowed us to uncover how context shapes learning, as well as to critically assess and challenge the disparities in adult education and learning in particular social contexts. Finally, an even more holistic conception of adult learning acknowledges the role of emotions, body, and spirit in learning. These as well as other approaches to adult learning will continue to be investigated, contributing to our understanding of the complex nature of learning in adulthood.

Bibliography

- Aktouf, O. (1992). Management and theories of organizations in the 1990s: Toward a critical radical humanism? *Academy of Management Review* 17(3), 407–431.
- Danis, C. (1992). A unifying framework for data-based research into adult self-directed learning. In Long, H. B. and associates (ed.) *Self-Directed Learning: Applications and Research*, pp 47–72. Norman, OK: University of Oklahoma.
- Dirkx, J. M. (2001). The power of feelings: Emotion, imagination, and the construction of meaning in adult learning. In Merriam, S. B. (ed.) *The New Update on Adult Learning Theory*. New Directions for Adult and Continuing Education, No. 89, pp 63–72. San Francisco, CA: Jossey-Bass.
- Fenwick, T. (2003). *Learning through Experience: Troubling Orthodoxies and Intersecting Questions*. Malabar, FL: Krieger.
- Freire, P. (1970). *Pedagogy of the Oppressed*. New York: Seabury Press.
- Grow, G. (1991). Teaching learners to be self-directed: A stage approach. *Adult Education Quarterly* 41(3), 125–149.
- Jarvis, P. (1992). *Paradoxes of Learning: On Becoming an Individual in Society*. San Francisco, CA: Jossey-Bass.
- Johnson, S. and Taylor, K. (eds.) (2006). *The Neuroscience of Adult Learning*. San Francisco, CA: Jossey-Bass.
- Kim, Y. S. (2008). *Situated Learning in a Korean Older Adults' Computer Classroom*. Unpublished Doctoral Dissertation. University of Georgia, Athens, Georgia.
- Knowles, M. S. (1968). Andragogy, not pedagogy. *Adult Leadership* 16(10), 350–352, 386.
- Knowles, M. S. (1975). *Self-Directed Learning*. New York: Association Press.
- Knowles, M. S. (1980). *The Modern Practice of Adult Education: From Pedagogy to Andragogy*, 2nd edn. New York: Cambridge Books.
- Knowles, M. S. (1984). *The Adult Learner: A Neglected Species*, 3rd edn. Houston, TX: Gulf.
- Knowles, M. S. and associates (1984). *Andragogy in Action: Applying Modern Principles of Adult Learning*. San Francisco, CA: Jossey-Bass.
- Lave, J. (1988). *Cognition in Practice: Mind, Mathematics, and Culture in Everyday Life*. Cambridge: Cambridge University Press.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Lawrence, R. L. (ed.) (2005). *Artistic Ways of Knowing: Expanded Opportunities for Teaching and Learning*. San Francisco, CA: Jossey-Bass.
- Merriam, S. B. and Clark, M. C. (2006). Learning and development: The connection in adulthood. In Hoare, C. (ed.) *Handbook of Adult Development and Learning*, pp 27–51. London: Oxford University Press.
- Merriam, S. B., Caffarella, R. S., and Baumgartner, L. M. (2007). *Learning in Adulthood*, 3rd edn. San Francisco, CA: Jossey-Bass.
- Merriam, S. B., Courtenay, B., and Baumgartner, L. (2003). On becoming a witch: Learning in a marginalized community of practice. *Adult Education Quarterly* 53(3), 170–188.
- Mezirow, J. (1996). Contemporary paradigms of learning. *Adult Education Quarterly* 46(3), 158–172.
- Mezirow, J. (2000). Learning to think like an adult: Core concepts of transformation theory. In Mezirow, J. and associates (ed.) *Learning as Transformation*, pp 3–34. San Francisco, CA: Jossey-Bass.
- Pert, C. B. (1997). *Molecules of Emotion: Why You Feel the Way You Feel*. New York: Scribner.
- Schon, D. A. (1987). *Educating the Reflective Practitioner*. San Francisco, CA: Jossey-Bass.
- Schon, D. A. (1996). From technical rationality to reflection-in-action. In Edwards, R., Hanson, A., and Raggatt, P. (eds.) *Boundaries of Adult Learning*. London: Routledge.
- Spear, G. E. and Mocker, D. W. (1984). The organizing circumstance: Environmental determinants in self-directed learning. *Adult Education Quarterly* 35(1), 1–10.
- Thorndike, E. L., Bregman, E. O., Tilton, J. W., and Woodyard, E. (1928). *Adult Learning*. New York: Macmillan.
- Tisdell, E. J. (1999). The spiritual dimension of adult development. In Clark, M. C. and Caffarella, R. S. (eds.) *An Update on Adult Development Theory: New Ways of Thinking about the Life Course*. San Francisco, CA: Jossey-Bass.
- Tisdell, E. J. (2003). *Exploring Spirituality and Culture in Adult and Higher Education*. San Francisco, CA: Jossey-Bass.
- Tolliver, D. E. and Tisdell, E. J. (2006). Engaging spirituality in the transformative higher education classroom. In Taylor, E. W. (ed.) *Teaching for Change: Fostering Transformative Learning in the Classroom*. New Directions for Adult and Continuing Education, No 109, pp 37–48. San Francisco, CA: Jossey-Bass.
- Tough, A. (1971). *The Adult's Learning Projects: A Fresh Approach to Theory and Practice in Adult Learning*. Toronto, ON: Ontario Institute for Studies in Education.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge: Cambridge University Press.

Further Reading

- Brookfield, S. (1986). *Understanding and Facilitating Adult Learning*. San Francisco, CA: Jossey-Bass.
- English, L. M., Fenwick, T. J., and Parsons, J. (2003). *Spirituality of Adult Education and Training*. Malabar, FL: Krieger.
- Illeris, K. (2004). *Adult Education and Adult Learning*. Malabar, FL: Krieger.
- Jarvis, P. (2006). *Towards a Comprehensive Theory of Human Learning*. London/New York: Routledge/Falmer.
- Merriam, S. B. (ed.) (2008). *Third Update on Adult Learning Theory*. New Directions for Adult and Continuing Education, No. 119. San Francisco, CA: Jossey-Bass.
- Merriam, S. B. and associates (2007). *Non-Western Perspectives on Learning and Knowing*. Malabar, FL: Krieger.

Adult Learning and Instruction: Transformative Learning Perspectives

P Cranton, Penn State University at Harrisburg, Middletown, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Although adult education practice has a long history, the area only became recognized as a field of study and theoretical development in the 1930s. Lindemann's (1926) book was the first to have adult education in the title; he described adult education as cooperative, nonauthoritarian, and informal. Around the same time, the first adult education professional association was established, and we saw the beginning of the quest to understand adult learning as a distinctive process and efforts to develop teaching methods unique to working with adults. Transformative learning theory (first labeled in the mid-1970s and more fully developed by the 1990s) is described as the first comprehensive theory of adult learning.

Characteristics of Adult Learning

Throughout the 1970s and 1980s, considerable effort was put into describing adult learning as a distinctive process – different from children's learning in important ways. Tough (1979) conducted his classic survey in which he discovered that over 90% of adults engage in sustained independent learning projects. Cross (1981) and others worked to determine why adults participated (or not) in educational programs. Brundage and MacKeracher (1980) outlined the characteristics of adult learning and what those characteristics meant for program planning. Knowles (1975) presented his foundational work on adults as self-directed learners – a point of view which continues to influence our understanding of how adults learn.

At least seven broad themes can be drawn from the research and theory from those two decades – themes which are still relied upon today by adult educators in many and diverse settings: (1) Adult learning is often described as voluntary. Individuals choose to become involved in informal and formal activities in order to develop personally or respond to a professional or practical need. (2) Based on Knowles' influential work, adult learning is usually described as self-directed. Knowles (1975, 1980) saw self-directed learning as a process by which people identify their learning needs, set goals, choose how to learn, gather materials, and evaluate their progress; however, many other definitions and conceptualizations developed over time (e.g., see Candy, 1991). (3) Adult learning is seen as practical or experiential in

nature. This notion can be traced back to Dewey (1938) who, though not an adult educator himself, had a lasting and profound impact on the way we think about adult learning. (4) Adults are portrayed as preferring collaborative and participatory learning, largely due to the early influence of humanism on adult education practice (via the client-centered approach of Rogers (1961) and then Knowles who was a student of Rogers). (5) Adults bring rich experiences and resources to their learning – one of Knowles' (1980) defining characteristics of adult learning. (6) Since adults have often been away from formal schooling for some years (though this is less true today) and may have had negative early experiences with school, they are seen as reentering learning with anxiety and low self-esteem. (7) Adults have a variety of learning styles and preferences. One of the most influential theorists in this area has been Kolb (1984) who delineated the converger, diverger, assimilator, and accommodator learning styles.

In the 1990s and early 2000s, the trend in adult education toward critical theory, postmodernism, and poststructuralism has led researchers and theorists away from what is now seen to be overly simplistic and even stereotypical understandings of adult learning. Nevertheless, this foundational work permeates adult-education practice and continues to influence our theory development and research. Transformative learning is voluntary and under the direction of the learner, based on making meaning out of experience, collaborative (especially through dialog or discourse), and empowering.

Instruction for Adult Learning

Early adult education had social change as its goal (Lindemann, 1926). The Highlander Folk School in the United States and the Antigonish Movement in Canada are examples of approaches to working with adult learners to promote social action. Freire (1970), following in the radical tradition of adult education, proposed that dialogic and problem-posing strategies replace the banking model of education in which information was transmitted to the learner. However, in the 1970s and 1980s, humanistic approaches to teaching adults dominated the field and still maintain a strong influence on how we think about instruction for adult learning today.

Just as adult learning is seen to be distinct from children's learning, teaching adults is seen to be different from teaching children. Knowles (1980) distinguished between pedagogy and andragogy, first describing them as being diametrically opposed and later modifying his view to place them on a continuum. Essentially, pedagogy was seen to be more teacher-centered and andragogy more learner-centered. How-to books on teaching adults abound. Following Knowles' lead, most of these books take a humanist approach to teaching adults. Elias and Merriam (2005) outline six basic assumptions underlying humanism:

1. human nature is naturally good;
2. human beings strive for freedom and autonomy;
3. the individuality and uniqueness of each person is valued;
4. working toward self-actualization is an innate human goal;
5. each person perceives the world in his or her own way; and
6. people are responsible for developing their potential to the fullest.

Teaching from a humanist perspective puts the educator in the role of a facilitator rather than a provider of information. She creates the conditions in which learning can occur and trusts the learner to take responsibility for learning. This sets up a student-centered environment in which the growth of the whole person (the development of self-actualizing persons) is a goal. Learning is a personal process; motivation is intrinsic. Yet, self-development and learning do not occur in isolation. Humanist educators value collaboration and group work over competitive endeavors, or put another way, connected knowing over separate knowing (Belenky and Stanton, 2000).

Moving into the 1990s and 2000s, humanist approaches to teaching adults remain strong, especially in practice, but are being challenged primarily as a result of the increasing popularity of critical theory as a foundation of adult education. Brookfield (2005) calls on us to unmask power, challenge ideologies, and radicalize criticality. Although ideology critique originated in the 1970s in the Frankfurt School of Critical Social Theory, it is more recently that it has been brought to bear on teaching adults. Ideologies are values, beliefs, and assumptions that have been uncritically assimilated and believed to be true without question. They appear in our social norms and expectations and in the way we use language.

Teaching for transformation stands in both perspectives. It is humanist in its goals of self-development and freedom from constraints and oppression, and it is critical in its goal of ideology critique. Brookfield (2000) writes that an "act of learning can be called transformative only if it involves a fundamental questioning and reordering

of how one things or acts. If something is transformed, it is different from what it was before at a very basic level" (p. 139). Newman (2006) writes about teaching rebelliousness, defiance, and action. He describes critical learning (transformative learning) as both a personal endeavor and a political act. It helps us see "through ourselves" and "through others," as we become "less susceptible to hegemonic control" (p. 239). Newman rejects the teaching of what he calls domesticated critical thinking and proposes that we teach people how to resist (pp. 9–10).

Transformative Learning: Overview of the Theory

Transformative learning theory had its beginning in 1975 when Mezirow conducted a study of 83 women returning to college in 12 reentry programs. He described a process of personal-perspective transformation that included ten phases:

1. experiencing a disorienting dilemma,
2. undergoing self-examination,
3. conducting a critical assessment of internalized assumptions and feeling a sense of alienation from traditional social expectations,
4. relating discontent to the similar experiences of others,
5. exploring options for new ways of acting,
6. building competence and confidence in new roles,
7. planning a course of action,
8. acquiring the knowledge and skills for implementing a new course of action,
9. trying out new roles and assessing them, and
10. reintegrating into society with the new perspective.

It was Mezirow's 1991 book, *Transformative Dimensions of Adult Learning*, that brought the theory to the forefront of the adult-education literature. Since then, Mezirow and others in the field have continued to elaborate on and provide alternative explanations of transformative learning. Here, an overview of the basic theory from Mezirow's (1991, 2000, 2003) perspective is provided, and then in the following section, other theorists' points of view are presented. The author's own writing is drawn upon (Cranton, 2006) in this discussion.

Transformative learning is a process by which previously uncritically assimilated frames of reference (assumptions, expectations, and habits of mind) are questioned and revised to make them more open, permeable, and better justified. Experiences are seen through the lens of our frames of reference, which include distortions, prejudices, stereotypes, and unexamined beliefs. When we encounter a perspective that is different from the one we hold, we may

be provoked into critically questioning our current thinking. This can happen as a product of a single event or as a gradual cumulative process. The learning only becomes transformative when we make a deep shift in how we see ourselves and/or the world around us and act on the revised perspective.

Several types of meaning structures come into play in this process. A frame of reference is a meaning perspective, the structure of assumptions and beliefs that provide a lens through which we make meaning of our experiences. A frame of references has two dimensions: a habit of mind, which is composed of broad, generalized, predispositions for interpreting experience; and a point of view, which is comprised of sets of immediate, specific expectations and beliefs that shape a specific interpretation (Mezirow, 2000).

Mezirow (2003) sees discourse as central to transformative learning. Discourse is defined as a form of dialog that involves the assessment of beliefs, feelings, and values. The ideal conditions of discourse are that participants have accurate and complete information, are free from coercion, are able to weigh evidence, are open to alternative perspectives, are able to engage in critical reflection, have an equal opportunity to participate, and are able to accept informed consensus as valid.

Habits of minds can be of different types. Originally, Mezirow (1991) wrote about epistemic perspectives, those that have to do with knowledge and the way we acquire knowledge; sociolinguistic perspectives, related to social norms, cultural expectations, and the way we use language; and psychological perspectives, which have to do with how we see ourselves – self-concept, needs, inhibitions, anxieties, and fears. Later, he added three more types of habits of mind: philosophical, which can be based on a worldview, philosophy or religious doctrine; esthetic, including values, attitudes, tastes, and standards about beauty; and moral–ethical, which incorporates conscience and morality (Mezirow, 2000).

Beliefs, assumptions, values, and habits of mind can be undeveloped or unquestioned. We absorb the way we think about ourselves and the world around us from our family, community, peers, and from the social world we live in. In each of the types of habits of mind, unexamined perspectives are insidious and often do not even appear as questionable. We may become aware of unexamined perspectives through a disorienting dilemma – an experience which contradicts our assumptions. As discussed in the last section of this article, educators may consciously create learning experiences that are potentially transformative.

Critical reflection and critical self-reflection are central to transformative learning. Critical reflection involves objective reframing of the assumptions of others, and critical self-reflection involves the subjective reframing of our own assumptions. Reflection alone is not enough to label learning as transformative. It must lead to revised frames of reference upon which individuals act.

Perspectives on Transformative Learning

Mezirow's theory of transformative learning was criticized on several grounds: for its failure to address social change (Collard and Law, 1989), the neglect of power issues (Hart, 1990), the disregard for the cultural context of learning (Clark and Wilson, 1991), the overemphasis on rational thought (Dirkx, 1997), and the prominence of separate or autonomous learning (Belenky and Stanton, 2000). These critiques and others that followed, along with Mezirow's call for people to contribute to and elaborate on the theory, led to the development of a variety of perspectives on transformative learning.

Connected Knowing and Transformative Learning

Relational or connected learning and knowing are often associated with women's ways of learning (e.g., see Hayes and Flannery, 2000). Belenky and Stanton (2000) use connected knowing as a basis for adding to transformative learning theory. They describe six developmental stages of knowing for women: silenced, received knowers, subjective knowers, separate knowers, connected knowers, and constructivist knowers. They suggest that Mezirow's approach to transformative learning places separate knowing (following lines of reasoning and looking for flaws in logic) in a central role and argue that connected knowing also serves well to describe transformation. Connected knowers suspend judgment and struggle to understand others' points of view from their perspective. They look for strengths, not weaknesses in another person's point of view. The goal is to see holistically rather than analytically.

Social Change as Transformative Learning

Social reform has long been a goal of adult education, and those critics who see Mezirow as neglecting the social change aspect of transformation suggest that social reform needs to precede individual transformation. Mezirow (2000) distinguishes between educational tasks (helping people become aware of oppressive structures and learn how to change them) and political tasks (forcing economic change). It is his goal to help individuals learn how to create social change rather than to create social change himself. Others see this differently.

Brookfield (2003) proposes that the purpose of transformative learning is ideology critique, a process that helps "people uncover and challenge dominant ideology and then learn how to organize social relations according to noncapitalist logic" (p. 224). He holds that transformation includes not only the individual's structural change, but also structural change in the social world. Similarly,

Newman (1994) emphasizes that we should study not the oppressed, but oppression itself. Writers and theorists who advocate social change as a goal of transformative learning (and adult education as a whole) do not dismiss individual learning and transformation, but see it as the educator's goal to address the social context within which individuals live and learn.

Group and Organizational Transformation

The idea that groups and organizations could transform in a collective way began with the notion of a learning organization in which organizations are perceived as living entities that can learn. Watkins and Marsick (1993), in their now-classic work on learning organizations, made the link to transformative learning. Yorks and Marsick (2000) have continued with this line of thinking, basing their work on action learning and collaborative inquiry. Action learning involves teams working on real problems within the organization, and collaborative inquiry consists of repeated episodes of reflection and action in a group context. In this way, organizations transform in relation to the nature of the environment, the vision or mission of the organization, products and services of the organization, management styles and procedures, organizational structure, and individual organization members' perception of their roles.

Groups other than organizations are also seen to have the capability to transform. Kasl and Elias (2000) suggest that individuals, groups, and organizations all share common characteristics and that there can be a group mind. In this approach to transformation, frames of reference are transcended rather than analyzed through critical reflection, and transformative learning becomes an expansion of consciousness that is collective as well as individual. The notion of group transformation does not supplant individual transformation, but simply adds another possible dimension to it.

Intuition, Imagination, and Soul in Transformative Learning

One of the most popular elaborations on Mezirow's cognitive and rational approach to transformative learning is the addition of intuition, imagination, and nurturing soul. Many of the new writers in the field are drawn to this way of understanding transformation as can be seen in the large proportion of paper presentations and experiential sessions, at the International Transformative Learning Conference (Wiessner *et al.*, 2003), that are based on artistic, creative, and imaginative points of view.

This perspective began with the work of Boyd (Boyd, 1991; Boyd and Myers, 1988) who used Jungian psychology to explain transformative learning. Rather than reflection, they described discernment as the central

process in transformation. In this view, transformation is a personal inner journey of individuation – learning through the psychic structures that make up the self.

It is Dirkx's (1997, 2001) writing that has carried this approach forward into the current literature on transformative learning, providing a theoretical foundation for the many people who use drama, art, music, images, poetry, and symbols to promote transformation in their practice. In Dirkx's view, transformative learning involves personal, spiritual, emotional, and imaginative ways of knowing – the way of mythos rather than logos. Mythos is a facet of knowing that we see in symbols, images, stories, and myths. We experience soul through art, music, and film; it is that magic moment that transcends rationality and gives depth, power, mystery, and deep meaning to learning. In nurturing soul, we pay attention to the small, everyday occurrences in life, understand and appreciate images, and honor the complex, multifaceted nature of learning.

The extrarational perspective on transformative learning can exist side by side with the rational perspective. When Dirkx and Mezirow discussed their approaches at the 2005 International Transformative Learning Conference, they agreed that their perspectives are "similar with respect to [their] mutual concern for transforming frames of reference that have either lost their meaning or usefulness or have in some way become dysfunctional. [They] are both interested in fostering enhanced awareness and consciousness of one's being in the world" (Dirkx *et al.*, 2006: 137).

Ecological View

Some theorists broaden the scope of transformative learning to span individual, relational, group, institutional, societal, and global perspectives (O'sullivan, 2003). The Transformative Learning Centre at the Ontario Institute for Studies in Education promotes this point of view as does the Holma College of Integral Studies in Sweden (Gunnlaugson, 2003). Transformative learning is a deep, structural shift in the basic premises of thoughts, feelings, and actions. It dramatically and permanently alters our way of being in the world and involves not only our understanding of ourselves but also our relationship with all of humanity and the natural world. Through transformative learning, we strive for a planetary community, learn to love life in all forms, and move out to universal horizons. Gunnlaugson proposes that we need to consider "our collective evolutionary destiny from the vantage point of the history of planet Earth" (p. 324).

Teaching for Transformation

Less has been written about how to teach in such a way as to promote transformative learning than has been written about the learning process. It is proposed that there are

three fundamental aspects to teaching for transformation: empowering learners, fostering critical reflection and self-knowledge, and supporting learners (Cranton, 2006).

Empowering Learners

Empowerment is both a product of and a precondition of transformative learning. As Mezirow (2000) says, “Hungry, homeless, desperate, threatened, sick, or frightened adults are less likely to be able to participate in discourse to help us better understand the meaning of our own experience” (pp. 16–17). Those educators hoping to promote transformative learning need to be conscious of helping learners feel empowered. There are at least three aspects to consider: (1) exercising power responsibly, (2) encouraging discourse, and (3) involving learners in decision making.

As Brookfield (2006) reminds us, teachers have power, and we cannot deny its existence or think we can give it away. Exercising power responsibly and consciously serves to create learner empowerment – power is to be shared and acknowledged among those present in the teaching environment. A variety of practical strategies can be used, for example, avoiding being in the position of providing right answers, making sure that there is equal access to all resources, including self-evaluation in graded courses, involving students in managing the learning environment, and being open and explicit about what is happening and why.

If discourse or dialog is central to transformative learning, helping learners be empowered needs to include this element of the process, especially as equal participation in discourse may not occur naturally. Educators need to find provocative ways to stimulate dialog from different perspectives, encourage learners to take on different roles in the dialog, be careful not to regulate the discussion or dismiss learners’ contributions, and provide time for reflection.

Involving learners in decision making, a strategy which harks back to Knowles’ and others’ advocating self-directed learning, enhances feelings of empowerment. Educators can use participatory planning in which learners decide on some or all of the topics, provide choices of methods to be used in the learning, encourage self-evaluation, ask learners for their perceptions of the experience, and keep the decision-making process open and explicit.

Fostering Critical Self-Reflection and Self-Knowledge

The most an educator can do is to set up an environment and conditions in which learners are able to engage in critical reflection and critical self-reflection. Entering into this process is voluntary; to approach it otherwise is

ethically questionable. A variety of ways are suggested to create such an environment, all of which are based on the goal of opening up new perspectives, challenging existing assumptions, or presenting information from a different point of view (Cranton, 2006).

Asking questions is the most basic strategy for fostering reflection. Educators can ask learners to help them see what assumptions they are making and to challenge the premises underlying their assumptions. Questions can center on the content of individuals’ beliefs, how they came to hold those beliefs, and why they value what they value (content, process, and premise-reflection questions, as suggested by Mezirow (1991)).

Journals are often suggested as a means of promoting critical reflection, but guidelines need to be given to learners to prevent the journal from being a simple log of what happened. Proffitt’s (1992) extensive work on journal writing is helpful. He suggests a variety of formats including writing a life history; engaging in a dialog with a person in the writer’s life or a historical figure; incorporating metaphors, dreams, and images; and writing from the perspective of another person.

Experiential learning through practicums, field trips, service learning, job shadowing, and any other real-life experience that can be incorporated into a course can stimulate critical reflection if the learner encounters perspectives that are different from those he or she holds. To further the possibility of critical reflection, the educator can hold discussions before and after the experience, suggest students write about the experience, encourage critical questioning among students, and emphasize any discrepancies between learners’ prior experience and the new experience.

Art-based activities promote imaginative and intuitive transformative experiences. Learners can either engage in the creation of art or they can view art. Some strategies the educator can consider are: creating collages as a group; encouraging students to make art as a project in a course; using film, fiction, and photography to present alternative points of view on an issue; or going to an art gallery or concert as a group.

Supporting Transformative Learning

In transformative learning, people are letting go of assumptions, beliefs, and perspectives that they may have held for a lifetime. Scott (1997) writes about the grieving involved in letting go of our way of seeing the world. The educator who fosters transformative learning has a moral responsibility to provide and arrange for support. He or she needs to establish relationships with learners and be conscious of what is happening in their lives when they engage in transformative learning.

This is not to say that the educator is solely responsible. Encouraging learners to support each other by

helping to establish a cohesive group and setting up learner networks (formal or informal, with or without technology) are helpful strategies. However, it is also important to be there when an individual learner comes to the educator for advice and support. Even after a course is over, there are occasions when someone needs assistance in deciding how to act based on a transformative learning experience or simply needs to talk about his or her experience.

Summary

Since Mezirow (1975) first proposed his notion of perspective transformation more than three decades ago, transformative learning theory has developed into a comprehensive theory of adult learning. In recent years, with the addition of several alternative perspectives, it has become a holistic and integrated way of understanding how adults experience deep shifts in perspective. In this article, transformative learning theory has been set in the larger context of adult education. The original theory as presented by Mezirow (1991, 2000) is described and the major ways in which other theorists have elaborated on his work are presented. It concludes with a discussion of how educators can facilitate and promote transformative learning.

Bibliography

- Belenky, M. and Stanton, A. (2000). Inequality, development, and connected knowing. In Mezirow, J. and associates (eds.) *Learning as Transformation: Critical Perspectives on a Theory in Progress*, pp 71–102. San Francisco, CA: Jossey-Bass.
- Boyd, R. D. (1991). *Personal Transformation in Small Groups: A Jungian Perspective*. London: Routledge.
- Boyd, R. D. and Myers, J. B. (1988). Transformative education. *International Journal of Lifelong Education* 7, 261–284.
- Brookfield, S. (2000). Transformative learning as ideology critique. In Mezirow, J. and associates (eds.) *Learning as Transformation: Critical Perspectives on a Theory in Progress*, pp 125–150. San Francisco, CA: Jossey-Bass.
- Brookfield, S. (2003). The praxis of transformative education: African American feminist conceptualizations. *Journal of Transformative Education* 1(3), 212–226.
- Brookfield, S. (2005). *The Power of Critical Theory: Liberating Adult Learning and Teaching*. San Francisco, CA: Jossey-Bass.
- Brookfield, S. (2006). Authenticity and power. In Cranton, P. (ed.) *Authenticity in Teaching. New Directions for Adult and Continuing Education*, No. 111, pp 5–16. San Francisco, CA: Jossey-Bass.
- Brundage, D. and MacKeracher, D. (1980). *Adult Learning Principles and Their Application to Program Planning*. Toronto, ON: Ontario Institute for Studies in Education.
- Candy, P. (1991). *Self-Direction for Lifelong Learning*. San Francisco, CA: Jossey-Bass.
- Clark, M. C. and Wilson, A. L. (1991). Context and rationality in Mezirow's theory of transformational learning. *Adult Education Quarterly* 42(2), 75–91.
- Collard, S. and Law, M. (1989). The limits of perspective transformation: A critique of Mezirow's theory. *Adult Education Quarterly* 39, 99–107.
- Cranton, P. (2006). *Understanding and Promoting Transformative Learning: A Guide for Educators of Adults*, 2nd edn. San Francisco, CA: Jossey-Bass.
- Cross, P. (1981). *Adults as Learners*. San Francisco, CA: Jossey-Bass.
- Dewey, J. (1938). *Experience and Education*. New York: Collier Books.
- Dirkx, J. M. (1997). Nurturing soul in adult education. In Cranton, P. (ed.) *Transformative Learning in Action: Insights from Practice. New Directions for Adult and Continuing Education*, No. 74, pp 79–88. San Francisco, CA: Jossey-Bass.
- Dirkx, J. M. (2001). The power of feelings: Emotion, imagination, and the construction of meaning in adult learning. In Merriam, S. (ed.) *The New Update on Adult Learning Theory. New Directions for Adult and Continuing Education*, No. 89, pp 63–72. San Francisco, CA: Jossey-Bass.
- Dirkx, J. M., Mezirow, J., and Cranton, P. (2006). Musings and reflections on the meaning, context, and process of transformative learning: A dialogue between John M. Dirkx and Jack Mezirow. *Journal of Transformative Education* 4(2), 123–139.
- Elias, J. and Merriam, S. (2005). *Philosophical Foundations of Adult Education*, 3rd edn. Malabar, FL: Krieger.
- Freire, P. (1970). *Pedagogy of the Oppressed*. New York: Herder and Herder.
- Gunnlaugson, O. (2003). Toward an integral education for the ecozoic era: A case study in transforming the global learning community of Holma College of Integral Studies. *Journal of Transformative Education* 2(4), 313–335.
- Hart, M. (1990). Critical theory and beyond: Further perspectives on emancipatory education. *Adult Education Quarterly* 40, 125–138.
- Hayes, E. and Flannery, D. (2000). *Women as Learners: The Significance of Gender in Adult Learning*. San Francisco, CA: Jossey-Bass.
- Kasl, E. and Elias, D. (2000). Creating new habits of mind in small groups. In Mezirow, J. and associates (eds.) *Learning as Transformation: Critical Perspectives on a Theory in Progress*, pp 229–252. San Francisco, CA: Jossey-Bass.
- Knowles, M. (1975). *Self-Directed Learning: A Guide for Learners and Teachers*. Chicago, IL: Follett.
- Knowles, M. (1980). *The Modern Practice of Adult Education: From Pedagogy to Andragogy*. New York: Cambridge.
- Kolb, D. A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice Hall.
- Lindemann, E. C. (1926). *The Meaning of Adult Education*. New York: New Republic.
- Mezirow, J. (1975). *Education for Perspective Transformation: Women's Reentry Programs in Community Colleges*. New York: Center for Adult Education, Teachers College, Columbia University.
- Mezirow, J. (1991). *Transformative Dimensions of Adult Learning*. San Francisco, CA: Jossey-Bass.
- Mezirow, J. (2000). Learning to think like an adult. In Mezirow, J. and associates (ed.) *Learning as Transformation: Critical Perspectives on a Theory in Progress*, pp 3–34. San Francisco, CA: Jossey-Bass.
- Mezirow, J. (2003). Transformative learning as discourse. *Journal of Transformative Education* 1(1), 58–63.
- Newman, M. (1994). *Defining the Enemy: Adult Education in Social Action*. Sydney: Victor Stewart.
- Newman, M. (2006). *Teaching Defiance: Stories and Strategies for Activist Educators*. San Francisco, CA: Jossey-Bass.
- O'sullivan, E. (2003). The ecological terrain of transformative learning: A vision statement. In Wiessner, C. A., Meyer, S. R., Pfhal, N., and Neaman, P. (eds.) *Transformative Learning in Action: Building Bridges across Contexts and Disciplines. Proceedings of the Fifth International Conference on Transformative Learning*, pp 336–340. New York: Teachers College, Columbia University.
- Progoff, I. (1992). *At a Journal Workshop: Writing to Access the Power of the Unconscious and Evoke Creative Ability*. New York: Penguin Putnam.
- Rogers, C. R. (1961). *On Becoming a Person: A Therapist's View of Psychotherapy*. Boston: Houghton Mifflin.
- Scott, S. (1997). The grieving soul in the transformation process. In Cranton, P. (ed.) *Transformative Learning in Action: Insights from Practice. New Directions for Adult and Continuing Education*, No. 74, pp 41–50. San Francisco, CA: Jossey-Bass.

- Tough, A. (1979). *The Adult's Learning Projects: A Fresh Approach to Theory and Practice*. Toronto, ON: Ontario Institute for Studies in Education.
- Watkins, K. and Marsick, V. (1993). *Sculpting the Learning Organization: Lessons in the Art and Sciences of Systemic Change*. San Francisco, CA: Jossey-Bass.
- Wiessner, C. A., Meyer, S. R., Pfohl, N., and Neaman, P. (eds.) (2003). *Transformative Learning in Action: Building Bridges Across Contexts and Disciplines. Proceedings of the Fifth International Conference on Transformative Learning*. New York: Teachers College, Columbia University.
- Yorks, L. and Marsick, V. (2000). Organizational learning and transformation. In Mezirow, J. and associates (eds.) *Learning as Transformation: Critical Perspectives on a Theory in Progress*, pp 253–284. San Francisco, CA: Jossey-Bass.

Adult Learning in a Biographic Perspective

L West, Canterbury Christ Church University, Canterbury, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Auto/biographical – Being aware of the extent to which we use other's stories to make sense of our own biographies as well as how we use our own to make sense of others' lives and experiences.

Biographic perspectives – Using other people's lives as a basis for understanding processes of learning, and so on.

Lifewide and lifelong learning – The idea that learning, in and across all dimensions of experience, past and present, can interconnect. Learning, in these terms, may be conceived as a psychological orientation to experience, the tendency to be relatively open to new experience, or, at another extreme, to fear and resist it. This can be rooted in early experience and the quality of our interactions with others.

Narratives – These are important in a biographic perspective. We experience our lives through some conception of past, present, and future. Such frames help bring some coherence to the fragments of experience. Biographic researchers often pay great attention to the qualities of people's narratives.

Poststructuralist – The emphasis here is on what is seen to be the pervasive power of language and power-knowledge formations to shape how we think and make sense of the world, including of ourselves.

Psychosocial – An interdisciplinary perspective in studies of adult learning, which is not only sensitive to how society and culture can structure the way we think and feel about the world, but also how our psychologies have a life of their own. There is an interest in how psychoanalytic insights can contribute to an understanding of inner and outer dynamics in people's encounters with learning.

Introduction

This article focuses on the biographic perspective and how this can illuminate the nature and processes of adult learning in unique ways. It draws on a rich body of international research and scholarship, over two decades and more, which constitutes, in fact, a major turn to biography and life history in the study of adult learning.

This is part of a broader trend across the social sciences (West *et al.*, 2007). The trend is partly a reaction against forms of research (such as behaviorism) which tended to marginalize the perspectives and subjective experiences of learners themselves or reduced learning to overly abstract entities. It has also been a reaction to theories of learning that neglected the role of meaning making and the importance of agency in human development.

The label biographic is used, for present purposes, to encompass a family of research and scholarship, which can be known in varying ways; examples are narrative or auto/biographical research, which may have common yet distinct meanings. Terms like biography or life history can be used interchangeably and may also be understood in differing ways. In Denmark, for instance (West *et al.*, 2007), a distinction is made between biography as the told life, and a life history, in which the researcher brings his or her interpretations and theoretical insights into play. In this article, biographic perspective is used inclusively to encompass different members of the family, while remembering that each can have distinct as well as related characteristics.

Some History

Feminism and oral history have been important in the emergence of biographic perspectives on adult learning, especially in North America and parts of Europe. In continental Europe, however, influences vary: in Denmark, for instance, critical theory has played an important part in the development of biographical perspectives into learning processes, alongside psychoanalysis and feminism. For many biographical researchers, the biographic perspective, methodologically, reaches back to the Chicago School of sociologists and social psychologists of the 1920s. The Chicago School developed the notion of symbolic interactionism to capture the dynamic, malleable, constructed, and also learned quality of human identity and social formation. Symbolic interactionism treated the things that members of society do as being performed by them as actors, rather than as if done by something called the system itself. The social order, in short, is dynamically learned through and from the interactions of its members. This finds renewed resonance in contemporary ideas about the role of reflexivity or biographical learning being of central importance in the contemporary world.

Second-wave feminism in the 1960s was also important in the development of the biographic perspective in adult

learning, particularly in the United Kingdom, North America, and also in the French-speaking world (Dybbroe and Ollagnier, 2003). Feminism was concerned with the pervasive influence of gender in people's learning lives in positioning both women and men in particular ways. Feminism and feminist research were committed to the idea of giving voice to women previously hidden in educational research. An early biographic focus, unsurprisingly, was on women's experiences of education (Merrill and West, 2009). Writers and researchers were often mature women students themselves and brought gender and the personal directly into accounts of learning as part of a wider political project.

The biographical perspective has slowly increased in importance in the literature of adult learning and education. By the early 1990s, the proportion of papers at the Standing Conference of University Teachers and Researchers in the Education of Adults (SCUTREA, which is Britain's preeminent professional network for researchers in the field of adult and lifelong learning), concerned with life stories, narratives, transformations of selves, and struggles over identity, was increasing. In 1988, only two papers had used a biographical or narrative perspective – broadly defined – as a method; however, this was to become a mainstream preoccupation in the 1990s. There was evidence of a similar trend in the increased number of papers devoted to biographic perspectives in meetings such as the North American Adult Education Research Conference (AERC). The trend has continued (West *et al.*, 2007).

Notwithstanding, there has been resistance to the perspective too. Among researchers, for instance, who considered that such work lacked observational rigor and objectivity in its preoccupation with internal experience and meaning-making processes. For some historians, the perspective has been criticized for being overly obsessed with the detail of learners' lives, and insufficiently focused on big educational and social questions around, for instance, the nature of educational opportunity in a capitalist economic order. From a poststructuralist perspective, the preoccupation with learners' stories, as a basis for understanding adult learning, risks neglecting the way in which stories – of learning, but of lives more widely – are penetrated by powerful discourses that individuals may barely be aware of. The biographic perspective, in this view, carries the danger of neglecting how we are storied as well as storytellers in learning lives (Merrill and West, 2009).

Mapping the Field

Nonetheless, there is a rich range of biographic scholarship providing, as suggested, not only many unique insights into processes of adult learning, but also some

new theoretical developments and challenges to received wisdom. Researchers have focused not only on learning in traditional contexts, such as higher and adult education, but also on learning and identity building in professional and workplace settings as well as other diverse locations for informal learning. These include learning in families, community development settings, trade union engagements, relationships, and even cyberspace. Researchers are busy exploring the impact of learning on health and its place in therapeutic processes. The biographical perspective is being used to examine the relationship between learning, class, ethnicity, and gender. Researchers have also addressed many criticisms, which include the capacity of people in biographic research to be aware, reflexively, of how they may be positioned by dominant social structures and or discourses (Frosh *et al.*, 2005).

Biographic perspectives developed strongly in the Francophone and German-speaking as well as the Anglo-Saxon world. In French-speaking communities, the literature reaches back to the work of a preceding generation, such as Paulo Freire, Bertrand Schwarz, and Jack Mezirow. The biographic perspective here adopted critical approaches to recognize the adult learner on the basis of his or her life trajectory and personal story. He or she was conceptualized as an authentic knower, deserving respect for her/his learning efforts and her/his relationship to knowledge. There has been a fundamental challenge to the banking concept of learning and a celebration of human agency, not least in questioning a deeply patriarchal social order, through biographic forms of enquiry (Dybbroe and Ollagnier, 2003).

Pineau (2000) has explored the link between life history, autobiography, and self-directed learning. He has worked with homeless people and used the biographic perspective to build a fuller understanding of learning paths in such contexts. Guy de Villers, a psychoanalyst as well as an educator, has recently edited a book on the boundaries between life history and therapy, which is a source of major debate in the French-speaking life history networks in adult education and, for that matter, more widely (Niewiadomski and de Villers, 2002). For Pierre Dominicé, the biographic perspective has an educational function in its own right and can be used to analyze and better understand the experiences of learners. He stresses, through his concept of educational biography, how the biographic perspective potentially has great occupational importance in helping people understand their professional work and interactions with others: as adult educators and trainers, for instance (Dominicé, 2000).

Learning: Lifewide and Lifelong

At the heart of the biographic perspective, as noted, is a shift in focus from an exclusive preoccupation with

education in formal settings to learning in many and diverse places: formal, nonformal, and informal, and toward an engagement with the interplay between these different dimensions of learning. There has been movement from the idea of a distinct field of adult education, in defined settings, to a more diverse terrain of learning in many locations. Adult education as a field tended to be preoccupied with postcompulsory education and with particular categories of learners and learning. By contrast, the biographical perspective encompasses processes of knowing and skills acquisition, affective transformation, and experiential learning beyond the walls of schools, colleges, and universities. It includes informal, tacit, and emotional learning, alongside cognitive processes and encounters with more formal bodies of knowledge. It covers learning in intensely intimate as well as more public space, and the connections between these. Formal learning is but one part, if important, of a complex tapestry.

The biographic perspective has helped to reveal some of the complex interplay between formal, nonformal, or informal processes of learning. Conventional distinctions between public and private experience, or learning in schools and families, or between structured and everyday learning, tend to unravel. This has been illuminated, for instance, in studies of professional learning (West, 2001; Olesen, 2007). Linden West has focused on family physicians learning and working in difficult and demanding innercity contexts. Thus West is concerned to understand how informal processes of learning may interact with more formal aspects of a doctor's training as well as the culture and subcultures of the profession. He notes a continuing neglect, for instance, in medical culture of learning about self and emotionality in doctor's professional training and how doctors can come to feel emotionally on the edge when working in difficult social contexts. Disturbance in others – including mental illness – can come to disturb a doctor and incidents of stress, alcoholism, mental health problems, as well as suicide have been increasing among doctors. The sample of doctors in the biographic study, which lasted for 4 years, emphasized the importance of supportive relationships as well as of self-knowledge and cultural learning alongside scientific knowledge when working in multicultural, innercity environments. The study documented the subtle interplay of learning in personal life with professional practice, and how the particular subcultures of the doctor's surgery and interactions with colleagues can be vital in creating space to learn from difficult experiences with patients. Learning, in such a biographic perspective, is understood more holistically.

A New Paradigm of Learning?

The biographic perspective, in fact, may have helped forge a new paradigm of learning (Alheit and Dausien, 2007).

Learning, viewed biographically, seems conceptually close, for instance, to the idea of reflexivity, which has become a central preoccupation of mainstream social science. This is set within a postmodern awareness of the uncertainty and rapidity of social and technological change. Here is the territory in which the self becomes a reflexive project. The globalizing tendencies of the present have been regarded as ushering in profound changes in social life and personal experience, which require constant efforts to construct and sustain the self through narratives of self-identity, in a kind of perpetual process of learning.

Biographies themselves, from such perspectives, in their unpredictability, become essential, often profound sites for learning. The capacity to compose a biography and develop reflexivity (or what Peter Alheit calls biographicity; Alheit and Dausien, 2007) is considered to be a survival necessity in a more individualized, perpetually changing, paradoxical, risk-inducing culture. This is a culture, it is argued, of new opportunities, but those that coexist with biographical fragility: in a world that fuels the necessity and/or desire to compose ourselves in new and distinct ways while generating intense anxieties about our capacity to cope. Ours is a period that offers enticing opportunities to live more autonomous lives while generating doubt about the sustainability or even desirability of such a project.

People, it is suggested, are confronted with new knowledge and the need to make choices (e.g., about what we eat or in response to illness) almost everyday, without reference to undisputed sources of authority. Science, often considered the means by which we can seek guidance, is itself a site of dispute and contestation. This is a liquid, shifting world in which we have to learn to learn, in the sense of needing to be open to new and diverse possibilities, rather than rely on a past consumption of fixed bodies of knowledge, also remembering that many of the economic, social, and familial scripts which historically shaped people's lives have weakened. Existentially, we are more on our own and inherit a world of diverse competing versions of how we should live or be. Learning becomes, in these terms, a biographical necessity. On the other hand, we can retreat into a kind of antilearning: into fundamentalism in which we seek complete answers in a body of doctrine, yet close ourselves off to the complexities of experience and their potential as a source for learning, including from the other and otherness.

Peter Alheit and Bettina Dausien identify some distinct learning imperatives in such a world. First, the need to learn from biographical crises: the processes of individualism referred to above bring considerable risk of crises-ridden biographical development. The experience of the modern world, they suggest, can turn personal meaningfulness into a problem of principle. People may have to negotiate periods of unemployment alongside

employment; returning to school alongside participation in the labor market, as well as constant questions about who they are and might want to be in situations where relationships, at work or in private life, may be transient. Adult learning, in such a context, moves closer to therapeutic processes. Meaning, in Alheit and Dausien's perspective, is less and less guaranteed by an unquestioning participation of individuals in culture and society. Generating meaning, and meaningful lives, has become the province of subjects who may feel, structurally, thrown into the deep end. Learning, of necessity, takes on a quasitherapeutic dimension, as does the work of many educators.

Another dimension of adult learning is brought into sharper relief by the biographic perspective: the importance of composing life stories and of telling stories more generally. As intergenerational continuities weaken, it becomes more important to find ways to work out how we might want to live our lives, and on what terms, including which values to choose. Stories are critically important to this end. The biographic perspective, forged out of experiences in the women's movement, has constantly emphasized the importance of finding voice in struggles to learn, for agency and meaningfulness.

The biographic perspective may also problematize overly rigid notions of developmental stages in learning and human life. A well-known variant is the midlife crisis, which draws on Jungian ideas of how men, of certain ages, once competitive career-building pressures have abated, may seek more feminine, artistic, and expressive forms of learning. However, when predictable linear trajectories of school–work–retirement break down, the notion of quasifixed stages becomes questionable. Desire for and resistance to learning across the lifecycle may become much more complex and variable as we are forced, for instance, to relearn or radically change our identities in unexpected, undesired ways.

An Interdisciplinary Imperative

The biographical perspective has also engendered new forms of interdisciplinary, psychosocial understanding of learning and learners. These new developments seek to transcend old style disputes between structuralist sociology and essentialist psychology in the literature of adult learning. Adult learning was conceived, on the one hand, as a largely internal psychological process, dependent on dispositions, inherent motivations, and objectively measurable intelligence. The social and historical could largely be excluded from the frame (Tennant, 1997). On the other hand, there was great resistance to such ideas, especially in more radical and oppositional theories of adult learning. In the British tradition, for instance, there was criticism of what was seen to be a mainly North American tendency to overly psychologize the human

subject and learning processes, and to reduce sociocultural forms of oppression to matters of individual pathology (Tennant, 1997). On the other hand, more radical, structural perspectives on adult learning have lacked any convincing theory of how the sociocultural translates into diverse internal states and why some people, more so than others, in objectively similar situations, are able to transcend oppression and become life spacers.

There is a considerable momentum in the biographic research family to bridge the conceptual gap, as defined above. There are a number of psychosocial studies – often drawing on psychoanalytic ideas – which explore the interplay of inner and outer worlds, psyche, and society: of learning in families or in working contexts, for example (West, 2007; Weber, 2007). Psyche is perceived to be a product of our intersubjective experiences, which are shaped, in turn, by the structuring and discursive processes within a given society, such as of class, race, or gender. However, psyche is no longer reduced, as in an overly structuralist sociology, to a kind of epiphenomenal or determined status: the inner world has a dynamic and power all of its own. If culture and society shape even in the most intimate of spaces – through poverty or the gendered distribution of emotional labor – such processes are mediated through the intimate relationships in which we are embedded. They can encourage relative emotional openness or closedness to experience and others. If curiosity, desire, and wholehearted engagement with the world are repressed or profoundly inhibited, because of our most significant others, then learning, in a psychological sense, easily becomes something to be avoided rather than embraced.

Such processes have been mapped, biographically, in studies of young mothers in parenting projects, like Sure Start in the United Kingdom (similar to Head Start in the United States) (West, 2007). One project was designed to support young single mothers who lived on a run-down public housing estate, suffering badly from deindustrialization, demoralization, and poverty. The particular project provided the base for a university and a community arts collaboration to utilize the visual arts to stimulate creativity and build confidence among hard-pressed single young mothers. The project was located in a youth center and the disaffected young mothers were to be recruited through outreach. The arts, it was hoped, would boost participants' confidence, planning and parenting skills, as well as broaden horizons. The young mothers would be encouraged to progress toward structured educational achievement or into work.

One particular case study chronicles, in detail, how a young woman, Gina, could retreat defiantly to the edge of any group of learners. This was part of a pattern in a life riddled with abuse, hard drugs, and ridicule. Yet, providing a creative artistic space, in the context of the strong relationships that Gina forged with tutors and

youth workers over time, gradually enabled her to take some risks. Her messy feelings about her pregnancy, for instance, were projected into a sculpture, made of chicken wire and plaster of Paris, which became a narrative of her pregnancy. She was able to work on the story artistically and, to an extent, transform her relationship to the experience and to her life more widely, through new narrative understanding. Her progress depended on the quality of interactions between people (the social context) and the extent to which she felt encouraged and able to play, imagine, think, and perceive herself differently, as a learner, mother, and person, and to relate to her toddler in new ways. There was a shift in her internal psychological drama, as new characters and symbolic objects, in the language of psychoanalysis, entered the stage. We are witnesses, in work of this kind, to how experience can stifle the desire to learn, and can psychologically close us down to new possibilities. However, new transitional spaces for learning and people coming alongside and tolerating our ambivalence – with sufficient self-knowledge, patience, and love, in a non-narcissistic sense – can reinvigorate creativity and the capacity to learn anew.

Other biographical researchers have used psychosocial ideas to explore learning processes from a gender perspective. Weber (2007), for example, focuses on gender and the learning processes of adult men training for work in the caring professions. Drawing on what she terms critical psychodynamic theory, she reveals learning in the workplace as a gendered battlefield where learning subjects' basic orientations can manifest themselves along gendered lines. These processes are partly theorized with reference to classic psychoanalytic insights into male struggles with intimacy. However, this is not to neglect culture, language, or material conditions. In this view, gender is inherent in social structures, stemming from historic divisions of labor, and is reproduced or changed within the scope of the accessible choices that people can make. Yet, in understanding reproduction and change processes, Weber observes that girls tend to identify with and separate from a model of their own gender while boys' paths to autonomy involve separation from a first intimacy. These patterns of early interaction are reinforced by language as symbolic representation, all of which serves to define what she calls gendered subjectivity. (Weber, building on the work of others, distinguishes gender subjectivity – the processes by which a person becomes a psychological subject – and gender identity, which refers to sexuality and cultural conceptions of gender.) Gendered subjectivity can find expression in what can be the differing responses of men and women to new kinds of training opportunities in which learning a capacity to care for others is required. Men, mirroring earlier patterns, tend to achieve this by experiencing and demonstrating degrees of autonomy first, whereas women tend to seek intimacy as a prerequisite of autonomy.

Weber stresses that these distinctions are far from absolute, but there are patterns nonetheless. We are given glimpses, through such biographical research, of the defended as well as social subject at the heart of learning.

Adult and Lifelong Learning: A Reconfiguration?

Finally, the biographic perspective can challenge received wisdom on adult learning in another, quite basic sense. The work of Knowles (1984) and others helped build the case for an academic discipline focused, distinctly, on adult learning. Knowles was reacting to the derivation of theories of adult learning from an understanding of how animals (especially rodents and pigeons) behaved, followed by observations of children. Adults and their experiences were neglected. The attempt to refocus research and scholarship, and the emergence of the associated idea of andragogy, Knowles argued, was premised on notions of a more dynamic culture of learning than pedagogy allowed. Learners should be actively involved in diagnosing their learning needs, in planning their experiences, and in developing a suitable learning environment.

Yet, from a biographic perspective, the continuities between early and adult experience are also important. The study of learning, as noted above, can reveal how past and present often intertwine. Even very confident people can feel helpless and overcome in new and demanding situations, such as entering a new course in higher or adult education. Such moments may link back to earlier feelings of inadequacy or failure, as past and present elide. Melanie Klein, a psychoanalyst, termed this memory in feeling, expressed in bodily and emotional states, rather than conscious thought. Such embodied memory can be especially intense for those taught, from earliest times, that they are of little consequence, are inadequate, or authority cannot be trusted. A range of psychological defenses may come into play, including withdrawal and denial of needs (that something may be important or desirable). There can be constant glimpses of defensiveness in adult learner narratives: early experience is never simply transcended and past and present can constantly elide (West, 2007).

Conclusion: The Biographic Perspective

It is suggested that learning, from a biographic perspective, is seen as more of a piece: social and psychological; informal and formal; and lifewide and lifelong, forged in a dynamic interplay of past and present and even future (without a sense of a future, the past cannot really be embraced). The biographical perspective, in fact, encompasses an idea of learning that echoes the humanistic and imaginative spirit associated with Wright Mills (1970) in

his work on biography. Biography, he argued, was situated at a meeting point between historical imperatives, social structures, and the inner worlds of human beings. His interest was in notions of human agency and how people, even in the most oppressive of situations, can find the resources, individually and collectively, to build better worlds. We could add learn their way to better worlds: a process requiring an eclectic, interdisciplinary understanding of people in their resilience and also vulnerabilities. Moreover, the biographic perspective encourages us to engage with ourselves when engaging with others and their learning: in making sense of the other, we draw on our own experience, and in making sense of self, we can make more sense of another. This has been termed the auto/biographic perspective in adult learning. Turning to biography challenges many and varied boundaries.

See also: Adult Learning; Barriers to Participation in Adult Education; Characteristics of Adult Learning; Community Based Adult Education; Gender Analysis; Health and Adult Learning; Informal Learning: A Contested Concept; Lifelong Learning; Rewriting the History of Adult Education: The Search for Narrative Structures.

Bibliography

- Alheit, P. and Dausien, B. (2007). Lifelong learning and biography: A competitive dynamic. In West, L., Alheit, P., Andersen, A. S., and Merrill, B. (eds.) *Using Biographical and Life History Methods in the Study of Adult and Lifelong Learning: European Perspectives*, pp 57–70. Frankfurt: Peter Lang/ESREA.
- Dominicé, P. (2000). *Learning from Our Lives: Using Educational Biographies with Adults*. San Francisco, CA: Jossey-Bass.
- Dybbroe, B. and Ollagnier, E. (eds.) (2003). *Challenging Gender in Lifelong Learning: European Perspectives*. Roskilde: University Press.
- Frosh, S., Phoenix, A., and Pattman, R. (2005). Struggling towards manhood: Narratives of homophobia and fathering. *British Journal of Psychotherapy* 22(1), 37–56.
- Knowles, M. (1984). *The Adult Learner, a Neglected Species*. Houston, TX: Gulf.
- Merrill, B. and West, L. (2009). *Using Biographical Methods in Social Science*. London: Sage.
- Mills, C. W. (1970). *The Sociological Imagination*. London: Penguin.
- Niewiadomski, C. and de Villers, G. (eds.) (2002). *Souci et soin de soi: Liens et frontières entre histoire de vie, psychothérapie et psychanalyse*. Paris: L'Harmattan.
- Olesen, H. S. (2007). Professional identities, subjectivity and learning: Be(coming) a general practitioner. In West, L., Alheit, P., Anderson, A., and Merrill, B. (eds.) *Using Biographical and Life History Methods in the Study of Adult and Lifelong Learning: European Perspectives*, pp 125–141. Frankfurt: Peter Lang/ESREA.
- Pineau, G. (2000). *Temporalités en Formation*. Paris: Anthropos.
- Tennant, M. (1997). *Psychology and Adult Learning*. London: Routledge.
- Weber, K. (2007). Gender, the knowledge economy and every day life. In West, L., Alheit, P., Anderson, A., and Merrill, B. (eds.) *Using Biographical and Life History Methods in the Study of Adult and*

Lifelong Learning: European Perspectives. Frankfurt: Peter Lang/ESREA.

West, L. (2001). *Doctors on the Edge; General Practitioners, Health and Learning in the Inner City*. London: FABooks.

West, L. (2007). An auto/biographical imagination: The radical challenge of families and their learning. In West, L., Alheit, P., Anderson, A. S., and Merrill, B. (eds.) *Using Biographical and Life History Methods in the Study of Adult and Lifelong Learning: European Perspectives*, pp 221–240. New York: Peter Lang/ESREA.

West, L., Alheit, P., Anderson, A., and Merrill, B. (eds.) (2007). *Using Biographical and Life History Methods in the Study of Adult and Lifelong Learning: European Perspectives*. Frankfurt: Peter Lang/ESREA.

Further Reading

- Bauman, Z. (2000). *Liquid Modernity*. Bristol: Policy Press.
- Bron, A., Kurantowicz, E., Salling Olesen, H., and West, L. (eds.) (2005). *Old and New Worlds of Adult Learning*. Wrocław: Wydawnictwo Naukowe.
- Chamberlayne, P., Bornat, J., and Aritzsch, U. (2004). *Biographical Methods and Professional Practice*. Bristol: Policy Press.
- Chamberlayne, P., Bornat, J., and Wengraf, T. (2000). *The Turn to Biographical Methods in the Social Sciences*. London: Routledge.
- Field, J. (2000). *Lifelong Learning and the New Educational Order*. Stoke-on-Trent: Trentham Books.
- Goodson, I. and Sykes, P. (2001). *Life History Research in Education Settings: Learning from Lives*. Buckingham: Open University Press.
- Hodkinson, P., Hodkinson, H., Evans, K., et al. (2004). The significance of individual biography in workplace learning. *Studies in the Education of Adults* 36(1), 6–24.
- Hunt, C. and West, L. (2006). Learning in a border country: Using psychodynamic perspectives in teaching and research. *Studies in the Education of Adults* 38(2), 160–177.
- McClaren, A. (1985). *Ambitions and Realisations – Women in Adult Education*. London: Peter Owen.
- Merrill, B. (1999). *Gender, Change and Identity: Mature Women Students in Universities*. Aldershot: Ashgate.
- Plummer, K. (2001). *Documents of Life*. London: Sage.
- Schuller, T., Preston, J., and Hammond, C. (2007). Mixing methods to measure learning benefits. In West, L., Alheit, P., Anderson, A. S., and Merrill, B. (eds.) *The Uses of Biographical and Life History Approaches in the Study of Adult and Lifelong Learning: European Perspectives*, pp 251–273. New York: Peter Lang.
- Stanley, L. (1992). *The Auto/Biographical I*. Manchester: Manchester University Press.
- West, L. (1996). *Beyond Fragments; Adults, Motivation and Higher Education*. London: Taylor and Francis.
- West, L., Miller, N., O'Reilly, D., and Allen, R. (eds.) (2001). *Travellers' Tales: From Adult Education to Lifelong Learning ... and Beyond. Proceedings of the 31st SCUTREA Conference Proceedings*, London: UEL.

Relevant Websites

- <http://www.canterbury.ac.uk> – Canterbury Christ Church University.
- <http://www.esrea.org> – European Society for Research on the Education of Adults (ESTREA).
- <http://www.learninglives.org> – Learning Lives: Learning, Identity and Agency in the Life Course.
- <http://www.scutrea.ac.uk> – Standing Conference on University Teaching and Research in the Education of Adults (SCUTREA).

Adult Learning, Instruction and Programme Planning: Insights from Freire

P Mayo, University of Malta, Msida, Malta

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Paulo Freire (1921–97) was one of the most influential educationists of the twentieth century. It is from a Freirean perspective that this article on adult learning, instruction, and program planning is written.

Argentinean scholar Daniel Schugurensky says, with reference to adult education, that: “in Latin America, Paulo Freire constitutes a watershed. There is before and after Freire” (Schugurensky, 1998: 344). Several years earlier, another Argentinean scholar, Carlos Alberto Torres, remarked: “We can stay with Freire or against Freire, but not without Freire” (Torres, 1982: 94). In addition, although Freire was undoubtedly one of the most heralded educators of the twentieth century, who inevitably has his detractors, his influence extends beyond the field of education to be felt in a variety of areas, including sociology, political theory, development studies, theology, philosophy, cultural studies, anthropology, language studies, and communications.

Paulo Freire suffered imprisonment and exile for his efforts in planning what was perceived as being a subversive approach to literacy in Brazil in the early 1960s. Freire subsequently worked, as a person in exile, in Chile, Massachusetts, and Geneva. During his 16-year period of exile, he was frequently called upon by revolutionary governments to assist them in developing and evaluating educational projects. He also engaged in projects with a variety of groups in different parts of the world. After his return to Brazil from exile, which, on his own admission, he had to relearn, he entered the complex domain of municipal educational administration in São Paulo, one of the world's largest cities.

Freire was most prolific as a published writer, with many of his works having been translated into English and other languages. Freire's better-known work, *Pedagogy of the Oppressed*, is regarded by many to be exemplary in the way it provides reflections on his many worlds of social action in a process that also involves constant recourse to theory, with Freire drawing on many sources in this regard, including Marxism, phenomenology, Christian personalism, liberation theology, and postcolonialism.

Praxis

Knowledge of the Community

Freire's pedagogy emerged from the Latin American tradition of popular education which incorporates a strong

degree of nonformal education. Nonformal education is not *laissez faire* pedagogy, but includes a certain degree of planning and organization. In the classic Freirean approach, the entire process of planning involves an intimate knowledge of the community in which the learning is to take place. The team of educators and project organizers, and other project participants, were allowed to mix with community members in a variety of settings, including their most informal settings, listen to their speech patterns and concerns, as well as identify some of the thematic complexes of the community itself. This approach was repeated and reinvented by Freire within the context of public educational administration when he served as educational secretary in the municipal government of São Paulo in his native, Brazil (O'Cadiz *et al.*, 1998).

Codification

Once the information was gathered, the team worked together and consulted community members, besides other persons connected with the locality, to draw up a plan of action that focused on the reality gleaned from the research carried out in the locality. Important aspects of this reality were thus codified in the form of pictures, subjects for discussion, plays, generative themes, and other pedagogical approaches. The material connected with the participants' framework of relevance, but was codified in such a way that it allowed the participants to gain some critical distance from the matter being discussed. This process of gaining critical distance is referred to as praxis. Praxis is a key concept in the Freirean approach to education.

Praxis has a long history dating back to the time of the ancient Greeks and at least Aristotle. It involves reflection upon action for transformative learning and action. This is how Freire defines praxis in *Pedagogy of the Oppressed*:

But human activity consists of action and reflection: it is praxis, it is transformation of the world. (Freire, 1970a/1993: 125)

Exile as Praxis

Freire goes on to say that the whole process involved needs to be enlightened through theory. It is praxis that lies at the heart of Paulo Freire's notion of critical literacy. Freire and other intellectuals, with whom he has

conversed in talking books, conceive of different learning situations in their life as forms of praxis. This applies to adult learning in its broadest contexts including learning from life situations – informal learning. These situations are viewed as moments when people can gain critical distance from the context they know to perceive it in a more critical light. For instance, Freire and the Chilean Antonio Faundez considered exile a form of praxis (Freire and Faundez, 1989). Freire also makes statements to this effect in a book with Betto and Freire (1986). He refers to the period of exile as one that provided a profoundly pedagogical experience, thus echoing Frei Betto, who also presented, in the same discussion, his 4-year experience of imprisonment under the military dictatorship as one that had a strong and important pedagogical dimension. Freire's period of exile is presented as a time during which he gained distance from Brazil and began to understand himself and Brazil better. It was a case of obtaining distance from what he had carried out in Brazil to prepare himself better to continue being active outside his context.

Antithesis of Praxis: Empty Theorizing and Mindless Activism

Freire relates the whole process of action and reflection to theory and practice (ibid.). Freire's work underscores the point that action on its own, isolated from reflection, constitutes mindless activism. Likewise, reflection on its own, divorced from action, constitutes empty theorizing. It is for this reason that Freire, in keeping with the Marxist tradition, regards one's material surroundings as the basis for the development of one's consciousness. In the words of Marx and Engels, "Consciousness is, therefore, from the beginning a social product, and remains so as long as men (sic.) exist at all" (Marx and Engels, 1970: 51). The notion of praxis that lies at the heart of Freire's pedagogical approach and which informs learning contexts developed on Freirean lines is akin to Marx and Engels' notion of revolutionizing practice as expressed in the *Theses on Feuerbach*.

Dialectical Relations

The action–reflection–transformative action process is not sequential, but dialectical (Allman, 1999, 2001). In the introduction to the special issue of *Convergence* dedicated to Freire, Allman *et al.* (1998) state:

Dialectical thinkers understand the internal relations among all phenomena. In the case of human beings or groups, this is a social relation which could be harmonious but which, thus far in history, normally has been antagonistic, resulting in various social relations that Freire collectively refers to as the oppressor-oppressed relation (e.g., class relations, gender, race, colonial, etc.) The antagonism is often so great that nothing short of

abolishing the dialectical relation will improve the situation. When there are no longer the two opposing groups, the possibility emerges of human beings uniting in love, with a commitment to social justice and to care for all of our social and natural world (Allman *et al.*, 1998: 10).

Teacher Student and Student Teachers

Learners can be assisted in this process of praxis, of coming to understand their reality in a more critical light, through a process of what Freire calls authentic dialog and participatory learning, as well as collective learning. The educator learns from the learners in the same way that the latter learn from her or him, the roles of educator and learner becoming almost interchangeable. In what has become a classic formulation, Freire wrote about the teacher student and students teachers. The educator is therefore regarded as a person who, while engaging in dialog with the learners, is also being taught by them. The learners, for their part, are also teaching while being taught (Freire, 1970a/1993: 80). In a dialog with Ira Shor, Freire states that:

Liberatory education is fundamentally a situation where the teacher and the students *both* have to be learners, *both* have to be cognitive subjects, in spite of their being different. This for me is the first test of liberating education, for teachers and students both to be critical agents in the act of knowing. (Shor and Freire, 1987: 33)

Learner as Subject

The educator would therefore transcend the boundaries of his/her social location to understand and act in solidarity with the learners, no longer perceived as other. In adopting a Freirean approach, one would regard educators and learners as subjects in a humanizing relationship. Solidarity is the hallmark of this pedagogical relationship. The learner's reality constitutes an integral part of the subject matter that, therefore, becomes a mediator between the two subjects in question, that is, the educator and learner. Freire goes on to state that the dialogical process of education marks "the sealing together of the teacher and the students in the joint act of knowing and re-knowing the object" (Shor and Freire, 1987: 100). Borrowing from this conversation between Freire and Shor, one can argue that anything that the educator already knows is relearned when studied again with the learners, a point confirmed by Freire in the same conversation (ibid.).

Learners and Educators Not Equal

However, and here comes the apparent contradiction, a Freirean approach to learning based on dialog is one wherein educators and learners are not on an equal footing.

Obviously, we also have to underscore that while we recognize that we have to learn from our students . . . this does not mean that teachers and students are the same. . . there is a difference between the educator and the student. (Freire, 1985: 177)

Much depends on the specific situation in which the adult learning process occurs; however, it would be amiss to celebrate learner voices uncritically since they are never innocent (Aronowitz and Giroux, 1991: 130–131). They contain various manifestations of the oppressor consciousness which ought to be challenged. Dialog, as conceived by Freire, also involves educators allowing themselves to be challenged and to constantly undergo self-reflection and scrutiny to confront the oppressor consciousness within. In short, both educator and learner need to address their contradictions in an ongoing process of gaining greater coherence. The educator needs to help create the conditions whereby the learners develop the confidence necessary to challenge him or her where necessary in a situation of mutual respect and trust. This is part of the humility which, according to Freire, all critical educators must show.

Directive Approach

The directive nature of the educational process is affirmed (see, e.g., the discussion with Moacir Gadotti and Sergio Guimarães published in Brazil in 1989 – Gadotti *et al.*, 1995: 50). Guarding against the perceived danger of a *laissez faire* pedagogy, resulting from a misconception of his particular notion of dialog, Freire emphasizes this directivity in the conversation with Ira Shor and elsewhere: “At the moment the teacher begins the dialogue, he or she knows a great deal, first in terms of knowledge and second in terms of the horizon that he or she wants to get to. The starting point is what the teacher knows about the object and where the teacher wants to go with it” (Shor and Freire, 1987: 103).

Freire makes it clear that he believes that the educators’ pedagogical action is guided by a particular political vision and theoretical understanding. Freire, after all, considers education to be a political act, there being no such thing as a neutral education, with educators having to answer the question “for whom and on whose behalf they are working” (Freire, 1985: 180). Freire once stated that the learning experience entails a process of research and curiosity with all the elements involved – teacher, student, knowing object, methods, and techniques – providing direction (Fabbri and Gomes, 1995: 96). He argues that it is for this reason that every form of educational practice is directive, but not necessarily manipulative, and that every educational practice cannot be neutral; a directive practice cannot be neutral – no one is neutral when facing an objective to be reached (Fabbri and Gomes, 1995).

Authority and Authoritarianism

Educators therefore have a directive role; they need to exercise their authority, an authority derived from their competence as pedagogs. Freire, however, draws an important distinction between authority and authoritarianism. It is imperative that the authority derived from one’s pedagogical competence does not degenerate into authoritarianism: “. . . the democratic teacher never, never transforms authority into authoritarianism” (Shor and Freire, 1987: 91; on this, also see Horton and Freire, 1990: 181). This authoritarianism would render the difference that exists between educator and learner antagonistic (Gadotti *et al.*, 1995: 50). The educator exercises what Ira Shor calls democratic authority (Shor, 1992: 156–158).

What we have, therefore, in Freire’s nuanced concept of dialog is a paradox rather than a contradiction. Freire provides a complex notion of learning and instruction, based on dialog. Freire (1974) feels that the traditional educator regards the knowledge he or she possesses, often captured in the lesson plan, as complete. The Freirean-inspired educator regards knowledge as dynamic, an object of co-investigation and unveiling that necessitates the participation of co-knowing subjects – the learners. The process of knowing involved, with respect to the object of knowledge, is considered by both educator and learner as incomplete (see Allman, 2001).

Tact and Prudence

Freire has even advocated tact and prudence when engaging in a dialogical approach, conceding that people who have been conditioned by many years of exposure to banking education do not immediately do away with this conditioning to embrace dialog. They often resist attempts at dialog, perhaps even misconstruing a dialogical approach for lack of competence on the educator’s part. Freire concedes that some instruction is necessary at times. It is for this reason that he once stated that an educator can alternate between traditional and progressive teaching. It is as though he seems to be saying that, in such difficult circumstances, dialog should be introduced only gradually (see Horton and Freire, 1990: 160). Elements of the old pedagogy can coexist with the new in an overall context that, however, privileges democratic relations.

Given the strong relationship between knowledge and the learner’s existential situation in Freire’s approach, one assumes that the participant has a repository to draw on. This repository consists of one’s life experience. The participant is therefore encouraged to draw on this experience in order to arrive at new knowledge and at a new awareness. In drawing on this experience, one is able to relate to the codified material. The educator enables this process to occur not by depositing knowledge, but by

engaging the learner's critical faculties. Rather than being a dispenser of knowledge, the educator poses questions and problematizes issues. In this problem-posing education, the pedagogy applied is primarily not that of the answer, but that of the question (Bruss and Macedo, 1985).

Collective Dimensions of Learning

In adopting a democratic, dialogical approach, the circle or learning setting serves as a microcosm, indicating the potential that can exist within contexts characterized by democratic social relations. Furthermore, knowledge itself is democratized and is therefore not presented any longer as the preserve of a privileged minority. In addition, the knowledge disseminated is in itself democratic in that its starting point is the life experienced by the participants and it serves their interest. Finally, it is group knowledge that emerges from this experience that emphasizes the collective dimensions of learning and of action for social change. Freire argued that one engages in the task of becoming more fully human not on one's own (it is not an individualistic endeavor), but in solidarity with others (Freire, 1970a/1993: 85–86). This having been said, one eventually moves beyond the here and now to gain a greater level of awareness. "Educands' concrete localization is the point of departure for the knowledge they create of the world" (Freire, 1994: 85). It is just the point of departure; for the here and now represents only the starting point of an ongoing adult learning process and not the endpoint. Remaining within the here and now constitutes, according to Freire, a case of populism or *basismo*. In remaining there and not moving beyond (through co-investigation of the object of inquiry), one would be engaging in basism, the romanticization (or mythification) of the vernacular (see Freire, 1994: 84).

It is this aspect of a Freire-inspired theory of instruction, learning, and curriculum planning that renders it quite different from the more liberal notion of learning through dialog which, often, erroneously passes off as Freirean.

The insights we derive from Freire, with regard to program planning, learning, and instruction in adult education, are described in the following section.

Planning Together

One should not enter the community and impose a program on its members, but should, on the contrary, engage with a team of researchers, preferably including people with different disciplinary backgrounds and certainly including both educators and potential project participants (the adult learners), in studying the community, where the learning setting is to be developed, at close hand. This process of study or research comprises informal

meetings with community members. The planning of materials occurs on the basis of the insights and information gleaned from the research.

Learning and Instruction

Learning Based on Action and Reflection

The approach throughout is one based on praxis involving critical reflection on the area of action, which also involves recourse to theory, but which entails an authentic notion of dialog in which the subject of enquiry is the focus of collective co-investigation. The research leads to insights which are to form the basis of the codified learning material whereby the educator enables the learners to gain critical distance from the community they know to be able to perceive it in a different, hopefully more critical, light. The same applies to the adult educator herself or himself who also gains critical distance from the object of co-investigation and can come to perceive it in a more critical light. We have seen how even exile is viewed by Freire and co-authors, engaged in dialog with him, as a form of praxis, of gaining critical distance.

Adult educators working with migrants in this ever-growing context for adult education can take a leaf out of Freire's book. One of the challenges for critical pedagogical work with migrants, to emerge from this Freirean insight, is that of enabling the migrants to read not only the world they now inhabit as immigrants, but also the world they left through a process of obtaining critical distance from their context of origin. This can hopefully lead to a greater understanding of the politics of their own dislocation.

Dynamic Knowledge

Through this process of praxis, based on reflection on action, knowledge is conceived as dynamic rather than static. The approach to learning is directive since learning is conceived as a political act. The roles of adult educator and adult learner are almost interchangeable, as all learn from each other, but this is not to say that the adult learner and adult educator are on an equal footing. The latter must have a certain amount of authority which should not be allowed to degenerate into authoritarianism lest the spirit of genuine dialog be destroyed.

Starting with the Learners' Existential Realities

Only through dialog does the group collectively learn to unveil the contradictions that underlie the reality on which it is focusing. Adult educators are encouraged to show tact when promoting dialogical relations and there are moments when they temper dialog with a certain degree of instruction, especially on consideration that people exposed to

banking education for years do not embrace dialog easily. The starting point of co-investigation is the learner's existential reality which is, however, not the be all and end all of the learning process, lest one be guilty of populism or *basismo*. Adult educators must demonstrate the humility necessary to be disposed to relearn that which they think they already know through their dialogic interactions with the rest of the learning group or community.

Bibliography

- Allman, P. (1999). *Revolutionary Social Transformation: Democratic Hopes, Political Possibilities and Critical Education*. Westport, CT: Bergin and Garvey.
- Allman, P. (2001). *Critical Education against Global Capitalism. Karl Marx and Revolutionary Critical Education*. Westport, CT: Bergin and Garvey.
- Allman, P., Mayo, P., Cavanagh, C., Lean Heng, C., and Haddad, S. (1998). Introduction. "...the creation of a world in which it will be easier to love" *Convergence* **XXI**(1 and 2), 9–16.
- Aronowitz, S. and Giroux, H. (1991). *Postmodern Education*. Minneapolis, MN: University of Minnesota Press.
- Betto, F. and Freire, P. (1986). *Una Scuola Chiamata Vita*. Bologna: E.M.I.
- Bruss, N. and Macedo, D. (1985). Toward a pedagogy of the question: Conversations with Paulo Freire. *Journal of Education* **167**, 7–21.
- Fabbri, F. and Gomes, A. M. (1995). La mia pedagogia: Paulo Freire risponde a professori e studenti bolognesi. In Gadotti, M., Freire, P., Guimarães, S., Bellanova, B., and Telleri, F. (eds.) *Pedagogia: dialogo e conflitto*, pp 92–103. Torino: Società Editrice Internazionale.
- Freire, P. (1970a/1993). *Pedagogy of the Oppressed*, 30th anniversary edn. New York: Continuum.
- Freire, P. (1974). *Authority versus Authoritarianism. Audiotape, in Series: Thinking with Paulo Freire*. Sydney, NSW: Australian Council of Churches.
- Freire, P. (1985). *The Politics of Education*. South Hadley, MA: Bergin and Garvey.
- Freire, P. (1994). *Pedagogy of Hope*. New York: Continuum.
- Freire, P. and Faundez, A. (1989). *Learning to Question. A Pedagogy of Liberation*. Geneva: World Council of Churches.
- Gadotti, M., Freire, P., and Guimarães, S. (1995). *Pedagogia: dialogo e conflitto*. (Bellanova, B. and Telleri, F. (eds.)). Torino: Società Editrice Internazionale.
- Horton, M. and Freire, P. (1990). *We Make the Road by Walking. Conversations on Education and Social Change*. Philadelphia, PA: Temple University Press.
- Marx, K. and Engels, F. (1970). *The German Ideology*. Arthur, C. J. (ed.). London: Lawrence and Wishart.
- O'Cadiz, P., Wong, P. L., and Torres, C. A. (1997). *Education and Democracy. Paulo Freire, Social Movements and Educational Reform in São Paulo*. Boulder, CO: Westview Press.

- Schugurensky, D. (1996). Paulo Freire: From pedagogy of the oppressed to pedagogy of hope. In Reno, H. and Witte, M. (eds.) *37th Annual AERC Proceedings*. Tampa: University of Florida.
- Shor, I. (1992). *Empowering Education. Critical Teaching for Social Change*. Chicago, IL: The University of Chicago Press.
- Shor, I. and Freire, P. (1987). *Pedagogy for Liberation Dialogues on Transforming Education*. South Hadley, MA: Bergin and Garvey.
- Torres, C. A. (1982). From the pedagogy of the oppressed to a Luta Continua – the political pedagogy of Paulo Freire. *Education with Production* **2**, 76–97.

Further Reading

- Darder, A. (2002). *Reinventing Paulo Freire. A Pedagogy of Love*. Boulder, CO: Westview Press.
- Gadotti, M. (1994). *Reading Paulo Freire. His Life and Work*. Albany, NY: SUNY Press.
- Gadotti, M. (1996). *Pedagogy of Praxis. A Dialectical Philosophy of Education*. Albany, NY: SUNY Press.
- Kidd, R. and Kumar, K. (1981). Co-opting Freire: A critical analysis of pseudo-Freirean adult education. *Economic and Political Weekly* **XVI** (1–2), 27–36.
- Kirkwood, G. and Kirkwood, C. (1989). *Living Adult Education: Freire in Scotland*. Milton Keynes, UK: Open University Press.
- Marx, K. and Engels, F. (1978). *The Marx–Engels Reader*. (Tucker, R. (ed.)). New York: W.W. Norton.
- Mayo, P. (2004). *Liberating Praxis. Paulo Freire's Legacy for Radical Education and Politics*. Westport, CT: Praeger.
- Roberts, P. (2000). *Education, Literacy, and Humanization Exploring the Work of Paulo Freire*. Westport, CT: Bergin and Garvey.
- Saul, A. M. (1995). Municipal educational policy in the city of São Paulo, Brazil (1988–1991). In Torres, C. A. (ed.) *Education and Social Change in Latin America*, pp 155–162. Melbourne, VIC: James Nicholas Publishers.
- Shor, I. (1999). What is critical literacy? *Journal for Pedagogy, Pluralism and Practice* **4**(1). <http://www.lesley.edu/journals/jppp/intro.html> (accessed May 2009).
- Steiner, S., Krank, M., McLaren, P., and Bahruth, R. (eds.) (2000). *Freirean Pedagogy, Praxis, and Possibilities. Projects for the New Millennium*. New York: Falmer.
- Stromquist, N. (1997). *Literacy for Citizenship; Gender and Grassroots Dynamics in Brazil*. Albany, NY: SUNY Press.
- Taylor, P. V. (1993). *The Texts of Paulo Freire*. Buckingham, UK: Open University Press.
- Telleri, F. (ed.) (2002). *Il Metodo Paulo Freire. Nuove tecnologie e sviluppo sostenibile*. Bologna: CLUEB.
- Torres, C. A. (1993). From the pedagogy of the oppressed to a Luta Continua – the political pedagogy of Paulo Freire. In McLaren, P. and Leonard, P. (eds.) *Paulo Freire: A Critical Encounter*, pp 119–145. London: Routledge.
- Torres, C. A. (1994). Paulo Freire as secretary of education in the municipality of São Paulo. *Comparative Education Review* **38**, 181–214.

Characteristics of Adult Learning

K Illeris, Danish University of Education, Copenhagen, Denmark

© 2010 Elsevier Ltd. All rights reserved.

The slogan of lifelong learning basically involves the simple message that learning can and should be a lifelong occupation. This poses the fundamental question of whether the processes of learning are the same irrespective of age. In the traditional psychology of learning, there are no age-conditioned differences. Learning was considered to be a common phenomenon in which researchers endeavored to discover the decisive and basic learning mechanisms, and research and tests were often performed to observe animals and humans in constructed laboratory situations.

Many scholars and researchers have claimed that adult's learning, as a psychological function, is basically similar to children's learning. This was, for instance, the underlying assumption behind the massive resistance to American Malcolm Knowles' launching of a separate discipline of andragogy, dealing with adult education and learning, and at the same time limiting pedagogy to the area of children's upbringing and schooling (e.g., Knowles, 1970; Hartree, 1984; Davonport, 1993). More recently, Alan Rogers, from Britain, in connection with his description of adults' learning, has deliberately maintained, "that there is nothing distinctive about the kind of learning undertaken by adults" (Rogers, 2003: 7). However, this position is only valid in relation to some very basic features of learning. As soon as the question is examined a little deeper or related to concrete learning courses or events, there are obviously substantial life-age differences, partly because some biological capacities of learning only mature gradually during childhood and youth, and partly because learning is also a social and an emotional process, and people's social and emotional situations change with their age (Illeris, 2004, 2007).

In the following section, some of the most important fundamental features of learning that are independent of life age are outlined; the special features of adult learning are discussed subsequently.

Some Fundamental Features of Learning

Human learning is a very complex matter that has been understood and conceptualized in many ways. However, for a structural analysis, three kinds of fundamental features can be pointed out: (1) the basic processes and dimensions of learning, (2) the different types of learning, and (3) the types of barriers to learning (Illeris, 2007).

Basic Processes and Dimensions of Learning

A fundamental aspect of learning is that it always includes two integrated but very different processes: the external interaction process between the learner and the social, cultural, and material environment; and the internal psychological process of elaboration and acquisition in which new impulses are connected with the results of prior learning.

The criteria of the interaction process are of a social and societal character, that is, they are determined by time and place. The individual interacts with an environment that includes other people, a specific culture, technology, etc., which are characterized by their time and society. In the late-modern globalized world, this is blended into a giant and rapidly changing hodgepodge that offers almost unlimited possibilities for learning.

No matter how dominant and imperative the interaction process has become, in learning there is also always a process of individual acquisition in which the impulses from the interaction are incorporated. As discussed by scholars such as Piaget (1952) and Ausubel (1968), the core of this process is that the new impressions are connected with the results of prior learning in a way that influences both. Thus, the outcome of the individual acquisition process is always dependent on what has already been acquired, and ultimately the criteria of this process are of a biological nature and determined by the extensive, but not infinite, possibilities of the human brain and central nervous system to cope with, structure, retain, and create meaning out of impressions as perceived by our senses (cf. e.g., Solms and Turnbull, 2002).

Learning, thinking, remembering, understanding, and similar functions are not just cognitive or content matters, although they have generally been conceived of as such by traditional learning psychology. Whether the frame of reference is common sense, Freudian psychology, modern management, or brand new results of brain research, there is much evidence that all such functions are also inseparably connected with emotions and motivation. The Austrian-American psychologist Hans Furth (1987), by combining the findings and theories of Piaget and Freud, has unravelled how cognition and emotions during the preschool years gradually separate out as distinctive but never isolated functions. The Portuguese-American neurologist Antonio Damasio (1994) has explained how cognition and emotion work in our brain and what disastrous consequences there are when the connections between

the two are cut by damage to the brain, even when neither of the functions in themselves has been affected. Thus, the acquisition process necessarily always has both a cognitive and an emotional side or, more broadly speaking, a content and an incentive side.

Consequently, all learning always includes three dimensions: the content dimension of knowledge, understandings, skills, abilities, attitudes, and the like; the incentive dimension of emotions, feelings, motivation, and volition; and the social dimension of interaction, communication, and cooperation – all of which are embedded in a societally situated context. The learning processes and dimensions are illustrated in **Figure 1**.

Different Types of Learning

Another fundamental aspect of learning has to do with the different character and scope of different types of learning processes. A very basic distinction was made by the well-known Swiss biologist and psychologist Jean Piaget (1896–1980) when he distinguished between assimilation and accommodation as two essentially different ways of learning (Piaget, 1952; Flavell, 1963). Later, other researchers found that each of these can be further differentiated into two, so that altogether four basic learning types emerge (Illeris, 2007).

As already mentioned, the acquisition process implies a linking between new impulses and the results of prior learning. These results cannot be thought of as merely an unstructured mass of knowledge, emotions, abilities, etc. One of Piaget's most fundamental assumptions was that to learn something means to mentally structure something, that is, to incorporate it in a mental scheme, and it

is the difference in which this incorporation takes place that constitutes the learning types. From modern brain research we know that such schemes have the character of dispositions to reactivate specific electrochemical circuits between brain cells that represent the content and incentives in question (Damasio, 1994). When a scheme or pattern is established, it happens by the type of learning called cumulation, a kind of mechanical process, establishing an isolated formation characterized by a form of automation that means that it can only be recalled and applied in situations mentally similar to the learning context. This is, for instance, how learning by conditioning functions.

The most common type of learning is, assimilation, or learning by addition, meaning that the new impulse is linked to a scheme or pattern already established in such a manner that it is relatively easy to recall and apply when one is mentally oriented toward the field in question.

Sometimes situations occur where we receive impulses that are difficult to immediately relate to any existing scheme or pattern. This can then take place by accommodation or transcendent learning, implying that one breaks down (parts of) an existing scheme or pattern and transforms it so that the new situation can be linked in. Thus, one both relinquishes and reconstructs something, a process that can be experienced as something demanding and even painful. The result can be recalled and applied in many different, relevant contexts.

Finally, there is also a far-reaching type of learning, which the American adult educator Jack Mezirow termed as transformation, and involves simultaneous restructuring in all three learning dimensions. This very demanding type of learning implies changes in the organization of the learner's self or identity and typically occurs as the result of a crisis-like situation caused by challenges experienced as urgent and unavoidable (e.g., Rogers, 1951; Engeström, 1987; Mezirow, 1991).

It is important that none of the learning types can be said to be better or more valuable than the others, as the more complex types of learning always presuppose that other and more basic learning has provided the preconditions that make them possible.

Barriers Toward Learning

Finally, as the third basic area of learning, it is important to deal with the different types of barriers that can prevent, reduce, or distort possible learning (Illeris, 2007).

In the content dimension, barriers will typically be about something that is not acquired, grasped, or taken in as intended. This may be generally termed mislearning, implying that the impulse or message does not come through, for instance, because of insufficient involvement or concentration, a lack of necessary prior learning, or due

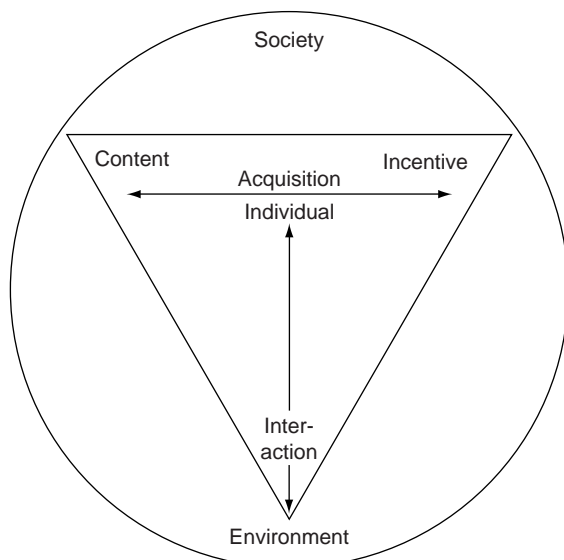


Figure 1 The processes and dimensions of learning. Adapted from Illeris, K. (2007) *How We Learn: An Introduction to Human Learning in Schools and Beyond*. London: Routledge).

to inadequate communication or teaching. The barriers to learning may also be rooted mainly in the incentive dimension. If so, it will typically be a case of some kind of mental defense. In our late-modern society, such defense mechanisms are no longer primarily the result of personal inhibitions as described by Freud a century ago, but are rather a general and necessary societally rooted defense against the overwhelming number and complexity of impulses and influences. Such defense either rejects or distorts the majority of impulses, preferably those that we dislike or are less interested in, but often those that we might profit from but do not immediately categorize as worth dealing with. Many also create learning defenses against the radically increased number of changes, the feeling of powerlessness experienced when authorities encroach on our life conditions, or the demands for change of identity, which the changing conditions impose on us.

Finally, nonlearning may be rooted mainly in the interaction dimension and have the nature of mental resistance. This can be inappropriate and annoying in many situations, but nevertheless constitutes symptoms of strong personal forces and engagement, which can lead to very important accommodative or transformative learning.

Adult Learning

In order to see what the characteristic of adult learning is, it may be useful to start by pointing out some basic features of children's learning.

Differences in Relation to Children's Way of Learning

In general, learning in childhood could be described as a continuous campaign to capture the world. The child is born into an unknown world and learning is about acquiring this world and learning to deal with it. In this connection, two learning-related features are prominent, especially for the small child: First, children's learning is comprehensive and uncensored. The child learns everything within her grasp, throws herself into everything, and is limited only by her biological development and the nature of her surroundings. Second, the child places utter confidence in the adults around her. She has only those adults and the ways in which they behave to refer to, without any possibility of evaluating or choosing what she is presented with. The child must, for example, learn the language these adults speak and practice the culture they practice.

Throughout childhood, the child's capturing of its surroundings is fundamentally uncensored and trusting as she endeavors, in an unlimited and indiscriminate way, to make use of the opportunities that present themselves. Of course, late-modern society has led to growing complexity and

even confusion of this situation as older children receive a lot of impressions from their pals and especially from the mass media, which go far beyond the borders of their own environment. But still the open and confident approach must be recognized as the starting point.

In contrast to this stands learning during adulthood. Being an adult essentially means that an individual is able and willing to assume responsibility for his or her own life and actions. Formally, our society ascribes such adulthood to individuals when they attain the age of 18 years. In reality, it is a gradual process that takes place throughout the period of youth, which as we see it today, may last well into the 20s or be entirely incomplete if the formation of a relatively stable identity is chosen as the criterion for its completion at the mental level (which is the classical description of this transition provided by Erikson, 1968).

As concerns learning, being an adult also means, in principle, that the individual accepts responsibility for his or her own learning, that is, more or less consciously sorts information and decides what he or she wants and does not want to learn. The situation in today's complicated modern society is after all such that the volume of what may be learned far exceeds the ability of any single individual, and this is true not only concerning content in a narrow sense, but it also applies to the views and attitudes, perceptions, communications options, behavioral patterns, lifestyle, etc. that may be chosen. Thus, a sorting of input must be made.

As a general conclusion it is, however, important to maintain that in contrast to children's uncensored and confident learning, adult learning is basically selective and self-directed, or to put it in more concrete terms:

- adults learn what they want to learn and what is meaningful for them to learn;
- adults draw on the resources they already have in their learning;
- adults take as much responsibility for their learning as they want to take (if they are allowed to); and
- adults are not very inclined to learn something they are not interested in, or in which they cannot see the meaning or importance. At any rate, typically, they only learn it partially, in a distorted way or with a lack of motivation that makes what is learned extremely vulnerable to oblivion and difficult to apply in situations not subjectively related to the learning context (Illeris, 2007).

In the following section, the nature of adult learning is discussed in relation to the three dimensions and other basic features of learning described earlier.

Adult Learning in the Content Dimension

For many years, it was a general understanding among learning theorists that humans acquire their full cognitive

learning capacity at about the age of 11–13, when they, as described in Jean Piaget's theory of learning stages, reach the so-called formal operational level, which makes logical–deductive thinking possible as a supplement to the forms of thinking and learning acquired at earlier stages (see, e.g., Flavell, 1963).

However, during the 1980s, this understanding began to be questioned from several quarters. On the one hand, it was pointed out that not all adults are actually able to think formally and operationally in the logical sense inherent in Piaget's definition. Empirical research pointed out that in England, it was actually less than 30%, but at the same time it was confirmed that at the beginning of puberty, a decisive development takes place in the possibilities for learning and thinking in abstract terms, so that, all in all, distinguishing a new cognitive phase was justified (Shayer and Adey, 1981). On the other hand, it has been maintained that, at a later age, significant new cognitive possibilities that extend beyond the formally operative may develop (e.g., Commons *et al.*, 1984). British-American adult education researcher Stephen Brookfield summarized this criticism by pointing out four possibilities for learning that are only developed in the course of adulthood: the capacity for dialectical thinking, the capacity for applying practical logic, the capacity for realizing how one may know what one knows (metacognition), and the capacity for critical reflection (Brookfield, 2000).

Recent brain research seems to indirectly support Brookfield's claims. Whereas it is a well-established understanding that the brain matures psychologically and neurologically for formal logical thinking in early puberty, evidence has been found that the brain centers of the frontal lobe that conduct functions such as rational planning, prioritization, and making well-founded choices, do not mature until the late teenage years (Gogtay *et al.*, 2004). This finding seems to provide some clarification of the differences between the capacity of formal logical and practical logical thinking and learning as well as between ordinary cognition and metacognition in adolescence and early adulthood.

At any rate, the general conclusion of all this must be that during puberty and youth, a physiological and neurological maturing process takes place that makes possible new forms of abstract and stringent thinking and learning, so that an individual becomes able to operate context-independently with coherent concept systems and manage a balanced and goal-directed behavior. Teenagers' determination to find out how things are structured and to use such understanding in relation to their own situation could be seen as a cognitive developmental bridge signifying the difference between children's and adults' ways of learning.

It is thus at one and the same time the longing for independence and the longing for coherent understanding of how they themselves and their environment function and why things are the way they are, which in the

content dimension separate adult learning from childhood learning. Up through the period of youth, individuals themselves increasingly assume responsibility for their own learning and nonlearning, make choices and rejections, and in this context understand what they are dealing with and their own roles and possibilities. However, all this has been enormously complicated by the duality of late modernity between, on the one hand, the apparently limitless degrees of freedom and reams of information, and on the other hand, far-reaching indirect pressure for control from parents, teachers, youth cultures, mass media, and formal conditions and possibilities. The transition from child to adult has thus, in the area of learning, become an extended, ambiguous, and complicated process, with blurred outlines and unclear conditions and goals.

Finally, it should be added that the possibility of transformative learning as well as the general defense systems in relation to learning seem to be developed along with the described cognitive development during the teenage years, and that the full development of the body and thereby the possibilities for bodily skills is also a process that is not fulfilled until about the age of 20 years.

Adult Learning in the Incentive Dimension

Learning in the incentive dimension is fundamentally aimed at helping the learner to control and direct emotions, feelings, motivation, and volition and thereby maintain an appropriate mental balance in relation to the complexity of life. In late-modern society, this process is concentrated on the development and maintenance of a self-understanding or identity that can secure the experience of being oneself, that is, of being the same person across the diversity of situations and challenges we meet and deal with. It is mainly in this dimension that the selective regulation and the defense mechanisms of adult learning are directed, which, in everyday life, is a mainly unconscious process and only taken up consciously in case of more important decisions. The identity, which, as mentioned, is developed during the teenage years, functions as the superior director of all this. On the more concrete level, adults typically have a range of life projects, subprojects, interests, and preferences to which the direction and regulation are related.

Most adults have a family project that concerns creating and being part of a family; a work project that concerns a personally and financially satisfying job; perhaps a leisure-time project concerning a special interest or a hobby; and sometimes a fulfillment or conviction project that may be religious or political in nature. All these projects result in many even more specific activities and attitudes that in this connection, serve as the measure for all the conscious as well as unconscious decisions about what to learn, with which kind of motivation and

investment to learn, and what not to learn or not to invest so much in that the learning process becomes demanding (which usually implies a weak and superficial kind of learning).

When dealing with adult learning, it is very important to realize that adults have and practice these conscious and unconscious ways of being selective in relation to their own learning. This is why and how adults, as stated before, “are not very inclined to learn something they are not interested in, or in which they cannot see the meaning or importance.” In contemporary adult education and lifelong learning, authorities and teachers very often do not sufficiently understand and respect these fundamental conditions of adult learning and many human and financial resources are wasted on programs and courses with limited or no possibilities of success (cf. Illeris, 2004).

Adult Learning in the Interaction Dimension

The general aim of learning in the interaction dimension is integration. Humans want and need to be integrated in their social environment, to communicate and collaborate with others, and this is a fundamental part of existence. But for adults, this side of life is highly selective: we have strong social preferences, and there are also communities which we do not want to be a part of. As to learning, this ambiguous attitude implies that the basic nature of our learning in the interaction dimension varies in relation to different learning spaces. In present society, five main types of spaces of adult learning can be pointed out.

Everyday life is the general and basic learning space in which a lot of learning takes place, as we are moving around and not participating in any specifically defined activities. This kind of learning is therefore mainly informal, multifarious, personal, and related to the cultures and subcultures in which the person is integrated, and it constitutes the general socialization and many patterns, norms, attitudes, general understandings, etc.

Workplaces have formed another very important learning space for adults ever since work was separated from everyday life. Here, of course, several professional qualifications are trained, but a lot of general learning takes place more or less incidentally as an inseparable part of work activities and communication (Marsick and Watkins, 1990). However, this learning is fundamentally different from general learning in everyday life, because it is marked by basic workplace conditions such as effectiveness, profit-orientation, and, not least, job stability. On the other hand, workplace learning may also include more formalized learning, which is usually accepted as relevant and meaningful, at least if the learner has a positive identification with the job. It is worth noting that even such goal-directed workplace learning is often restricted by the immediate needs of production or service and

therefore tends to lack theoretical understanding and overview (Illeris *et al.*, 2004).

Special interests or convictions are the basis of another type of learning spaces including, for example, communities, associations, hobbies, and grassroots movements. Learning here may be understood as a goal-directed kind of everyday learning in which incidental and informal features are replaced by a clear motivation and resolution, which generally make this type of learning space very effective.

Schools and other educational institutions form the type of learning space which society has established to secure a lot of learning that is today considered to be necessary for all of us to maintain the material and structural level we have reached. Originally, this type of learning space was mainly set up for children and youth, but it is significant for late modernity that such institutions are now also established for adults and the slogan of lifelong learning covers a massive upgrading of such learning. However, it is important to realize that school learning is by nature formal, rational, and externally directed. Although it is officially aimed at goals outside the educational system, it is usually experienced as directed by internal measures such as the school subjects and exams. Although adult schooling is not compulsory, many adults are brought into situations where they more or less have to undertake such formal learning. Low-skilled workers often oppose or are very ambivalent about going back to school as they have for so many years experienced that they are no good at school learning (Illeris, 2003, 2006).

Finally, it must be mentioned that computers and the Internet have opened a new learning space of rapidly growing importance, not least for adults. Computer learning has its own characteristics, advantages, and disadvantages. Net-based learning is very flexible, because it can be practiced independently of time and to some extent also of place. It also seems to have an advantage in that it forces the learner to express oneself in writing and thereby to make points, understandings, and opinions more clear than generally needed in face-to-face conversation. The disadvantage is the lack of direct social contact, but this can to some extent be eliminated by frequent classes or meetings, during a Net-based course, of some duration. Yet, so far we know very little about the transfer of Net-based learning onto the different spaces of reality.

As most adults today are involved in all or most of the main described types of learning spaces, the transitions and transfer of learning between them become increasingly essential and complex. The transition and transfer between school and education on the one side and workplaces and everyday learning on the other side forms a challenge to contemporary adult learning. Net-based learning may relate to either of these sides, whereas interest-based learning has a kind of independent position

from which transfer can often be made to any other learning space.

Conclusion

Learning in late-modern society is a very complex and at the same time very important issue, and the slogan of lifelong learning strongly indicates that this also includes adult learning. However, adult learning is very different from children's learning and much of the traditional learning theory and teaching practice does not apply very well to adults. Whereas childhood naturally is time for learning and development, adulthood traditionally has been the age for applying the acquired competencies. Today, this situation has changed, and learning in adulthood has become a necessary demand. This causes many challenges and expenses, but also new possibilities of a rich and expanding adult life.

It must be remembered and respected that in free and democratic societies, adults have the right to direct their own learning. Therefore, the typical patterns and processes of childhood learning cannot just be overtaken, and the idea of lifelong learning can only be practiced successfully if sustainable ways to meet the needs of the adult learners can be developed and practiced.

Bibliography

- Ausubel, D. P. (1968). *Educational Psychology: A Cognitive View*. New York: Holt, Rinehart and Winston.
- Brookfield, S. D. (2000). Adult cognition as a dimension of lifelong learning. In Field, J. and Leicester, M. (eds.) *Lifelong Learning: Education across the Lifespan*, pp 89–101. London: RoutledgeFalmer.
- Commons, M. L., Richards, F. A., and Armon, C. (eds.) (1984). *Beyond Formal Operations: Late Adolescent and Adult Cognitive Development*. New York: Praeger.
- Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Grosset/Putnam.
- Davonport, J. (1993). Is there any way out of the andragogy morass? In Thorpe, M., Edwards, R., and Hanson, A. (eds.) *Culture and Process*

- of Adult Learning*, pp 109–117. (first publ. in 1987). London: Routledge.
- Engeström, Y. (1987). *Learning by Expanding: An Activity-Theoretical Approach to Developmental Research*. Helsinki: Orienta-Konsultit.
- Erikson, E. H. (1968). *Identity, Youth and Crises*. New York: Norton.
- Flavell, J. H. (1963). *The Developmental Psychology of Jean Piaget*. New York: Van Nostrand.
- Furth, H. G. (1987). *Knowledge as Desire*. New York: Columbia University Press.
- Gogtay, N., Giedd, J. N., Lusk, L., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America* **101**(21), 8174–8179.
- Hartree, A. (1984). Malcolm Knowles' theory of andragogy: A critique. *International Journal of Lifelong Education* **3**(3), 203–210.
- Illeris, K. (2003). Adult education as experienced by the learners. *International Journal of Lifelong Education* **22**(1), 13–23.
- Illeris, K. (2004). *Adult Education and Adult Learning*. Malabar, FL: Krieger.
- Illeris, K. (2006). Lifelong learning and the low-skilled. *International Journal of Lifelong Education* **25**(1), 15–28.
- Illeris, K. (2007). *How We Learn: An Introduction to Human Learning in Schools and Beyond*. London: Routledge.
- Illeris, K., et al. (2004). *Learning in Working Life*. Copenhagen: Roskilde University Press.
- Knowles, M. S. (1970). *The Modern Practice of Adult Education: Andragogy versus Pedagogy*. New York: Associated Press.
- Marsick, V. J. and Watkins, K. E. (1990). *Informal and Incidental Learning in the Workplace*. London: Routledge.
- Piaget, J. (1952). *The Origins of Intelligence in Children* (first publ. in 1936). New York: International Universities Press.
- Rogers, A. (2003). *What Is the Difference? A New Critique of Adult Learning and Teaching*. Leicester: NIACE.
- Rogers, C. R. (1951). *Client-Centered Therapy*. Boston, MA: Houghton-Mifflin.
- Shayer, M. and Adey, P. (1981). *Towards a Science of Science Teaching*. London: Heinemann Educational.
- Solms, M. and Turnbull, O. (2002). *The Brain and the Inner World*. New York: Other Press.

Further Reading

- Jarvis, P. and Parker, S. (eds.) (2005). *Human Learning: An Holistic Approach*. Abingdon: Routledge.
- Merriam, S. B., Caffarella, R. S., and Baumgartner, L. M. (2007). *Learning in Adulthood: A Comprehensive Guide*, 3rd edn. San Francisco, CA: Jossey-Bass.
- Sutherland, P. and Crowther, J. (eds.) (2006). *Lifelong Learning: Concepts and Contexts*. Abingdon: Routledge.

Informal Learning: A Contested Concept

P Hodkinson, University of Leeds, Leeds, UK

© 2010 Elsevier Ltd. All rights reserved.

Although the term informal learning has been widely used for a long period of time, its precise meaning remains elusive and contested. The term is normally invoked in opposition to formal learning and/or formal education, and has, therefore, much in common with the terms nonformal learning and nonformal education. In its common sense usage, informal learning is that learning which takes place outside educational institutions, such as schools, colleges, and universities. The confusions around its meaning and usage arise from three interrelated problems. They are:

- differences between the research and political developments of the term;
- confusion and disagreement over the boundaries between formal and informal learning; and
- contrasted and contested theoretical ways of understanding all learning.

The Research Development of Informal Learning

Over the last 30 years, researchers have rediscovered the fact that much learning takes place in everyday life, without the interventions of teachers or educational institutions. The importance of everyday learning had been largely lost sight of, after the decline of behaviorism, as cognitive psychologists focused most attention on learning in schools. Reinforced by liberal humanist views of education as major root of civilization, schooling came to be seen as normally the best way for people to learn. The debates in the early 1970s focused mainly on how to make schooling better.

Scribner and Cole (1973) wrote one early seminal piece pointing out that much successful learning took place outside school. They explicitly used the term informal learning and juxtaposed it with formal learning, which was what schools were supposed to be good at. Since then, there has been much research showing informal learning in a wide range of contexts, including families, communities, and workplaces. It was not just the contexts of formal and informal learning that were different. Their processes and the nature of what was learned were also different. From this perspective, formal learning is planned and organized and assessed by teachers and others, while informal learning is incidental and ubiquitous in everyday practices. Formal learning is concerned with high status, mainly propositional knowledge, which, the argument goes, is generalizable, not

situation specific. Informal learning is concerned with low-status knowledge and skills, and much of it is assumed to be situation specific.

Within this strand of literature, most writers are either concerned with formal learning and education, or with informal learning. One result of this polarization is a sometimes vehement debate about whether formal or informal learning is more efficient. As schooling developed in the twentieth century, they were the places where Enlightenment ideals of science and rationality could be applied to learning. Schools were where learning was taken seriously, and were an improvement over simple and primitive everyday learning. On the other hand, researchers into informal learning argued that it often worked better. Thus, for example, children learn language very effectively in everyday life, compared with the ways in which at least British and American children struggle to learn a foreign language in school.

The Political Development of Nonformal Education and Nonformal Learning

Whereas recent usage of informal learning grew up in opposition to formal learning, usage of the term nonformal learning grew out of the older term, nonformal education. The first use of the term nonformal education was probably a United Nations Educational, Scientific and Cultural Organization (UNESCO) report on education in what was then termed the underdeveloped world (Hamadache, 1991). The essence of nonformal education was a central desire for empowerment, through democratization. In the 1940s, this was associated with post-World War II growth of anti- and postcolonialism. Tutors could be centrally involved, working with underprivileged groups, helping them achieve their own objectives. Thus, informal education was opposed to formal education, which was seen as elitist, and primarily concerned with social reproduction, rather than emancipation. However, for those writing about informal learning (see above), this struggle was between two different forms of formal learning.

Colley *et al.* (2003) argue that since 1947, a broadly nonformal educational movement has ebbed and flowed, growing in strength for short periods of time, interspersed with longer periods of the dominance of formal education. In what they term the second wave, in the 1970s, the term nonformal learning first appears. These second-wave movements were found, for example, in Brazil led by

Freire and others, and in Tanzania. This time, partly through Freire's (1972) work, the movement spread into the developed world, for example, in feminist, antiracist and working-class movements.

When nonformal learning surfaced again, from the late 1980s onward, it became more formalized, under the influence of free-market capitalism and human-capital theory. Funding in most parts of the world is increasingly focused on economic production. The formalization of much nonformal and informal learning can be seen, for example, in the growth of the competence movement, whereby the outcomes of work-related learning are specified, measured, and potentially accredited, regardless of where that learning took place. The radical and emancipatory agendas of the previous waves had been largely lost.

As these two quite different origins of the concept of informal learning converge, the usage and meaning of that concept becomes further confused. What both traditions share in common is a sense of informal learning being other than formal education. This leads directly to the next problem. Informal learning is seldom clearly defined in its own right. Almost always, it is formal learning/education which is defined, leaving informal learning as its opposite.

The Blurred Boundaries Between Formal and Informal Learning

Colley *et al.* (2003) analyzed a range of texts which had explicitly set out the differences between informal learning and formal learning. What their analysis showed was that, although there was significant overlap between them, each author defined the differences in significantly different ways. Colley and coworkers identified a list of 20 criteria, all of which had been used by at least one author. This meant that learning which would be classified as informal by one group of writers would be seen as formal by another group. More tellingly, there was no small group of criteria that were universally used by all writers. Thus for some, in the informal-education tradition, informal learning could be seen as part of, say, an adult-education course, provided that course was run upon democratic lines. For many others, informal learning only took place outside educational institutions. Similarly, many agreed that formal learning was assessed and accredited. However, others did not see, say, accredited workplace learning as formal, and much learning in schools, which most would describe as formal, is often unaccredited. Marsick and Watkins (1990) suggest that formal learning is deliberative and planned, while informal learning is incidental, which cuts right across divisions based on location and both types could be accredited, for example, within the competence movement. These boundary confusions are bad enough if the focus is

on formal learning, but when informal learning is the focus, there is a further problem. This is because it is often formal learning that is defined. Informal learning is all the rest.

Further confusions arose when some authors tried to distinguish between three categories: formal, nonformal, and informal learning. Traditionally, nonformal and informal had been used almost interchangeably. More recently, as we have seen, nonformal learning has been invented as a third category, possibly as a means of resolving some of the confusions in the distinctions between formal and informal learning. However, there is even less clarity and agreement between the boundaries of informal and nonformal learning than between formal and informal. Colley *et al.* (2003) concluded that the distinctions between informal and nonformal learning were largely meaningless. However, the threefold classification used by the European Union (European Commission, 2001) has become influential in relation to policy and practice in Europe. In this classification, nonformal learning was seen as a midway point between informal and formal: learning that is not provided by an education or training institution and typically does not lead to certification, but is structured (in terms of learning objectives, learning time, or learning support). Nonformal learning was defined as intentional from the learner's perspective. That is, nonformal learning is part of what was defined as formal learning by Marsick and Watkins (1990), but could equally be seen as an attempt to divide informal learning in workplaces and communities into the purposive (nonformal) and the incidental (formal).

Colley *et al.* (2003) argued that rather than distinct types of learning, termed formal and informal, there were many attributes of informality and formality in learning, and that most if not all learning contained a mixture of both. They suggest that these various attributes of formality and informality relate to four areas: the location of learning, the purposes of learning, the processes of learning, and the content of learning. Thus, the content of learning could be, say, mainly formal – perhaps academic history, the purpose relatively informal, in that the history is being learned from a TV program, watched for interest, in which case the process would also be largely informal, as would be the location, if the TV was being watched in the home, rather than as a planned part of an educational course. Multiple other combinations are possible.

Theoretical Differences

The third source of confusion around the term informal learning arises from deeper disagreements about the nature of learning itself. A full analysis of these debates lies beyond the scope of this article, although some relevant reading has been included, for those who want more detail. In essence, two radically different ways of understanding learning each

has very different implications for what we take the concept of informal learning to mean.

On the one hand, much of the research on learning from a cognitive psychological perspective implicitly adopts what Sfard (1998) terms an acquisition metaphor for learning. That is, learning is understood as the process whereby knowledge, skills, understanding, etc. are acquired by the learner. Within this intellectual tradition, there has always been a concentration on learning as a cognitive mental process, and on the nature of learning either in relation to early childhood, or within educational institutions. Thus, there is an implicit switch from what might be termed informal learning in childhood, to formal learning at school. However, even in relation to early childhood, a major emphasis has been on how tutoring or teaching (by the parent and others) can improve that learning. As tutor control is an often-cited criterion for formal learning, this way of understanding learning has traditionally neglected many informal attributes of learning.

Western policymakers often implicitly adopt an extreme and oversimplified version of learning as acquisition, which sees what is learned as a commodity, to be inserted into the heads of learners. Here, and in many of the more sophisticated versions of learning as acquisition, there is a tendency to see learning processes and learning outcomes as separate – with one leading to the other. Learning is also seen as separate from the context in which it takes place, so that a major objective of education (and also a major problem) is to facilitate the transfer of what was learned on one place (such as a school) to another (such as work). When informal learning enters the acquisition perspective, it is often as an alternative method to formal learning, whereby desired outcomes are/can be achieved. Often, it is seen as of secondary importance to formal learning.

Situated-learning theories, in their various guises, see learning differently. Here, as Sfard (1998) shows, the metaphor is not acquisition, but participation. That is, people learn through participating in practices and/or activities, including those of schooling, as well as, say, those in the family, at work, or in the local community. There has been work done on learning in education from this perspective, but arguably most of the major developments in situated learning have come from focusing on other places where people learn – perhaps, especially the workplace. Perhaps, the most radical and most significant claim of situated-learning theorists is that there is no separate process, called learning. Rather, learning happens as part of the practices of participating. Thus, in a workplace, work itself, and the various social processes involved in it, embrace learning. This sort of participatory learning is inherently social, and far from being separate from its location, is an integral part of the situation where it takes place. That is, the nature of learning varies significantly from one situation to another. Within any

situation, learning processes and learning outcomes are also integrated. Furthermore, for many writers, learning is embodied, involving the practical and affective as well as cognitive. Often such learning is tacit, in the sense that people are not really conscious of learning.

It follows that situated-learning theorists naturally focus on what is often termed informal learning. This is seen as the natural and therefore most common form of learning. If formal learning is thought of as different, then it is as a minor variation, not the major learning form. Thus, when researchers write about formal education, a common thrust is that we should make it more like informal learning.

If the boundary between formal and informal learning is contested and imprecise at best, from the situated-learning perspective, there is a different boundary problem with the concept. The logical conclusion from a situated-learning position is that learning is ubiquitous in life. Not only is learning ever-present, but as there is no separation between any social practices and learning, there is no clear division between learning and life. This definition problem leads to two contrary positions in the literature. On the one hand, some writers call for a return to seeing education as the main concept for analysis, rather than learning. For them, informal learning is just too vague and broad to be meaningful. The opposite view is that we have to accept that learning is ubiquitous, and it follows that learning is everywhere and that much of it cannot be measured. Rather than worry about this imprecision, we should ask different questions about learning (including informal learning), which relate to its value and its purposes, rather than its effectiveness. At the time of writing, situated-learning theories are in ascendance over cognitive theories, but a resolution to the implications of the ubiquity of learning is not in sight. More recent work is also seeking to build back an individual perspective on learning, for example, by finding ways of blending metaphors of participation with those of construction. Most of these approaches focus on learning as broadly undifferentiated into formal and informal, implicitly or explicitly adopting a similar stance to Colley *et al.* (2003), in this respect. However, there remain a significant number of cognitive researchers who are unconvinced.

Informal Learning in Educational Courses

If we adopt the situated-learning perspective when looking at learning in classrooms, schools, and colleges, it becomes apparent that there is much informal learning in such places too. As early as 1961, Jackson pointed out that while in school, the hidden curriculum was at least as important as the official curriculum. By that, he meant that pupils learned about how to behave in school, and more seriously, being in school reinforced a sense of

failure and inadequacy in many. Although he never used the term, informal learning, it can be applied to this still underestimated part of what is learned at school. His use of the term hidden is telling. Most of the literature on learning at school focuses on how well or badly the official curriculum is learned, ignoring the extensive informal learning that goes alongside it. Furthermore, the two are not separate. Some situated-learning theorists argue that learning to be a pupil in a school interferes with the learning of, say, physics. For example, pupils learn how to stay out of trouble, how to do well with minimal effort, how to pass tests rather than master a subject, and how to get on with their peers, for example, to avoid being labeled a teachers' pet. In relation to some recent research done on learning in an English college, Hodgkinson and Colley (2005) analyze the complex interrelationships between formal and informal learning in three very different courses. In each case, they claim that it makes little sense to see formal and informal learning as different. To do so, risks leaving a very partial understanding of the learning taking place. In arguing this way, they are taking a situated-learning view of what learning is.

Formal Learning at Work

Billett (2002), writing mainly about learning in the workplace, also argues that formal and informal learning are not distinct from each other. In fact, he argues, all learning is formal in several respects. Writing from a situated-learning perspective, he argues that the practices and activities through which learning takes place are themselves partly formalized, through formal structures of, say a workplace. Thus, work hierarchies, rules, and procedures interpenetrate all workplace learning, which is thus structured, controlled, and deliberative – that is, formalized. Whether his argument is accepted or not, it is clear that there are increasing amounts of planned, taught (or tutored), and assessed learning in modern workplaces, for example, through performance-management schemes, or managerial requirements for specific training – in health and safety, in using new technology, etc.

Is Informal Learning Becoming More Formalized?

As informal learning attracts more and more attention from researchers, policymakers, and others, those who still believe in the clear distinction between it and formal learning face a paradox. For, the more we discover about the importance of informal learning, the more we try to specify, measure, and control it. To use the terminology of Colley *et al.* (2003) we are attaching more and more of the

attributes of formal learning to it. That is, we are making informal learning more formal.

This increasing formalization can be seen in a number of places. Quantitative researchers are searching for better and more-encompassing measures of or proxies for informal learning, so that they can assess how widespread it is, who has access to it, etc. In many adult- and higher education contexts, there is a growing use of accreditation of prior experience and learning (APEL). That is, having recognized that people learn a lot informally, the drive is on to find formal evidence so that such learning can also be accredited, and can be counted toward a qualification. In the workplace-learning arena, there has been a widespread growth of competence-based qualifications. Their purpose is to record, verify, and ultimately accredit learning which has taken place during the everyday practices of work. Another example is the growth of performance-management schemes at work, where employees agree learning targets with a line manager.

For those whose tradition is the radical politics of non-formal adult education, this trend to formalize the informal is deeply worrying, even though some of these initiatives, such as APEL, may be of benefit to some disadvantaged learners. The trend brings dangers of increasing hierarchical control and increasing surveillance of workers and other relatively disadvantaged groups. Moreover, if and when it is discovered that even the most disadvantaged have lots of informal learning in their lives, this can provide an easy excuse for policymakers not to provide, say, additional access to adult-education classes, because these people have enough learning anyway.

Another set of concerns about this trend relates to whether or not it is inherently counterproductive. There are subtly different arguments here. First, if informal learning is different from formal learning, and if it is also beneficial, then there are dangers that in making it increasingly formal, some of its very advantages are lost. An extension of this line of argument is that this formalization is actually the reason for increased interest in informal learning, rather than it resulting from such interest. This raises suggestions that the focus on learning rather than education has political purposes, in focusing attention and responsibility on individuals, rather than the state. It also marginalizes older debates about the purposes of education and over what people should learn or should be encouraged to learn.

The second very similar argument comes from the line taken by Colley *et al.* (2003). They argue that the interrelationship between formal and informal attributes of learning is very important. It follows that when we choose to alter the balance between such attributes, it is necessary to think carefully about what any such changes do to the nature of the learning itself. The problem is that in most cases of increased formalization, this is not done. Rather, there are assumptions that (1) we are measuring what is

already there without changing it, or (2) we can simply improve what already happens, by managing it better.

The third argument links the two. If most learning is informal, and if learning is ubiquitous in everyday activity, attempts to measure it or to accredit it may miss a larger point. It is not whether people have learned or not, or how much they have learned that matters. What matters more are important questions about the value and purposes of what is being learned, and to whom. In other words, much of the growing formalization of informal learning has technical concerns, directly relating to the needs of those in power – employers, educational institutions, governments, etc. More important are questions of value – questions of why and for what purpose, rather than what or how. Of course, value questions are always contested and contestable, and raising this question about informal learning would place the contested nature of learning at the center of the debate.

Conclusion: The Value of Contestation

Ever since the first usages of the term, informal learning has been contested, and there are two dimensions to that contestation. There are theoretical and research-driven arguments about what the concept means, what its boundaries are, and how useful it is. There are political arguments about its purposes and about who can, should, or does control it.

At a practical level, all this makes the concept of informal learning difficult to use well. Too many writers are unclear about their own uses of the term, and appear unaware that many other writers use it in different ways to them. On the other hand, the concept has such widespread usage that, at least in English, it is almost impossible to avoid it. Some of its essence can be captured in alternatives, such as everyday learning, but each such alternative brings its own definitional problems and metaphorical baggage. Writers need to use informal learning and its alternatives with care, and should make clear the sense in which they are using the term.

However, if contestation and complexity bring problems, they also bring opportunities. The debates and arguments about informal learning lie at the heart of the ongoing search for better and clearer ways to understand learning *per se*. For our understanding of learning to advance, we need to keep these contestations firmly in mind, as we work to address them and find better ways forward. The political debates about the purpose of informal learning relate directly to arguably even more important concerns for

educational researchers and practitioners. These concerns are the extent of inequality and disadvantage in education and learning, related to social class, gender, ethnicity, poverty, etc. Despite the fears from some quarters that a focus on informal learning is part of a broader move to place responsibility and blame on individual learners, such issues of social and educational inequality concern all forms of learning in all aspects of people's lives. The trick is not to focus solely either on formal education or informal learning, but to see both as often related parts of a larger picture.

Bibliography

- Billett, S. (2002). Critiquing workplace learning discourses: Participation and continuity at work. *Studies in the Education of Adults* 34(1), 56–67.
- Colley, H., Hodkinson, P., and Malcolm, J. (2003). *Informality and Formality in Learning: A Report for the Learning and Skills Research Centre*. London: Learning and Skills Research Centre.
- European Commission (2001). *Communication: Making a European Area of Lifelong Learning a Reality*. http://europa.eu.int/comm/education/policies/life/communication/com_en.pdf (accessed June 2009).
- Freire, P. (1972). *Pedagogy of the Oppressed*. Harmondsworth: Penguin.
- Hamadache, A. (1991). Non-formal education: A definition of the concept and some examples. *Prospects* 21(1), 111–124.
- Hodkinson, P. and Colley, H. (2005). Formality and informality in college-based learning. In Kuenzel, K. (ed.) *International Yearbook of Adult Education 31/32, 2005: Informal Learning, Self-Education and Social Praxis*, pp 165–182. Cologne: Boehlau Verlag.
- Marsick, V. and Watkins, K. (1990). *Informal and Incidental Learning in the Workplace*. London: Routledge.
- Scribner, S. and Cole, M. (1973). Cognitive consequences of formal and informal education. *Science* 182, 553–559.
- Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher* 27(2), 4–13.

Further Reading

- Beckett, D. and Hager, P. (2002). *Life, Work and Learning: Practice in Postmodernity*. London: Routledge.
- Coffield, F. (ed.) (2000). *The Necessity of Informal Learning*. Bristol: Policy Press.
- Foley, G. (1999). *Learning in Social Action: A Contribution to Understanding Informal Education*. New York: St Martin's Press.
- Jackson, P. W. (1961). *Life in Classrooms*. New York: Holt, Rinehart and Wilson.
- Jeffs, T. and Smith, M. K. (1996). *Informal Education: Conversation, Democracy and Learning*. Derby: Education Now Books with the YMCA George Williams College.
- Lave, J. and Wenger, E. (1991). *Situated Learning*. Cambridge: Cambridge University Press.
- Livingstone, D. (2001). Adults' informal learning: Definitions, findings, gaps and future research. *NALL Working Paper No. 21*. Toronto, ON: Centre for the Study of Education and Work, OISE/UT.
- Torres, C. A. (1990). *The Politics of Nonformal Education in Latin America*. New York: Praeger.

Organizational Learning

P-E Ellström, Linköping University, Linköping, Sweden

© 2010 Elsevier Ltd. All rights reserved.

Although organizational learning has been a subject of research within management and organizational studies at least since the early 1960s, interest in this concept has grown considerably since the late 1980s. Important driving forces for this development have been the globalization and growing corporate competition, and thereby, an increasing interest in finding alternatives to established forms of organization. This development was mirrored not least in the movement to abandon traditional Tayloristic and bureaucratic models of work organization in favor of allegedly more flexible and integrated work systems, that is, what has been characterized as high-commitment work systems. The performance of these types of new work systems is assumed to critically depend on their capacity to create favorable conditions for organizational learning.

Today the concept of organizational learning has become established not only within economic and management disciplines, but also within behavioral and social science disciplines such as psychology, education, and sociology. Considering empirical research on organizational learning, it is clear that there is and has been a predominance of studies within private companies. This is the case in spite of the fact that many of the early studies by researchers such as Argyris, Cyert, and March were also carried out in governmental agencies and educational institutions. However, over the last decade there has been an increase in the number of studies of organizational learning in, for example, schools, military organizations, public administration, political organizations, and unions. Considering education specifically, the interest in organizational learning is presently clearly visible in at least three areas of educational research: school development and innovative schools, human resource development (HRD), and workplace learning.

The Field of Organizational Learning

The field of organizational learning is characterized by an increasing diversity and specialization. In order to map key topics within the field of organizational learning, Easterby-Smith and Lyles (2003) distinguish between four subfields within the general field of organizational learning: organizational learning, the learning organization, organizational knowledge, and knowledge management. The first of these four subfields, organizational learning, refers to descriptive and explanatory studies of learning processes of and within organizations. That is, studies of organizational learning largely from an

academic knowledge interest, aiming at an understanding and critique of what is.

In contrast to this, the learning organization subfield is characterized by more practical and normative knowledge interests. A learning organization is viewed as an ideal type of organization (or a vision of an organization), which has the capacity not only to facilitate the learning of its members, but also to transform this learning into continuous organizational renewal. The literature within this subfield focuses on how to create and improve the learning capacity of an organization through different types of intervention. One of the most well-known contributions to this subfield is the book on learning organizations by Peter Senge in 1990.

Turning to the remaining two subfields, the subfield called organizational knowledge refers to attempts to understand the nature of knowledge and processes of knowledge production within and between organizations. Within this subfield, there are studies of not only what has become known as knowledge work, but also studies of knowledge creation and the sharing and integration of organizational knowledge and competence. A key contribution to this subfield is the book by Nonaka and Takeuchi (1995) on knowledge creation through transformations of tacit and explicit knowledge. The subfield of knowledge management has a focus on issues related to the use of information and communication technology (ICT) to facilitate and support the acquisition, storage, sharing, retrieval, and utilization of knowledge in order to improve organizational performance.

As underlined by Easterby-Smith and Lyles (2003), it is not possible to make clearcut distinctions between these four subfields of organizational learning. For example, a critical study of processes of organizational learning within an alleged learning organization would belong to the subfield of organizational learning. Furthermore, studies of organizational learning interventions could belong both to the subfield of organizational learning and that of the learning organization depending on which knowledge interest predominates. As will be clear in the next section, it is also difficult to make clear-cut distinctions between the two subfields of organizational learning and organizational knowledge. In fact, in recent years, these two subfields have tended to merge.

Dimensions of Organizational Learning

In spite of the large number of studies on organizational learning that have been published during recent years,

there is little consensus on how to define the concept of organizational learning. It is also, to say the least, a very difficult task to try to formulate a definition that could adequately cover the many meanings of organizational learning that can be found in the literature. On a general level, however, it is possible to distinguish a number of dimensions along which many definitions vary. In the text below, three such dimensions are dealt with.

Levels of Analysis

A first dimension is what is considered as the proper level of analysis, and thus, as the locus of learning. Many definitions and studies of organizational learning focus on an individual level of analysis, that is, on the learning by individuals in an organizational context (Huber, 1991; Simon, 1991). A main argument behind this approach is that basic concepts in the literature on organizational learning, for example, the concept of memory, only apply to individual subjects. In line with this, it is assumed that organizational learning is mediated through individual learning, conceived as an interplay between processes of cognition and action. Furthermore, organizational learning is assumed to imply individual learning, but not vice versa. Thus, individual learning is viewed as a necessary, but not sufficient condition for organizational learning. In relation to this view, questions arise concerning what it means when an organization learns, and how to understand the links between individual and organizational learning.

Other definitions of organizational learning focus on a collective subject, that is, on the group (the team) or the organization as the locus of learning. The basic assumption is that the group or the organization is more than a collection of individuals, and thus, that learning at a collective level is different from and not only the sum of each individual's learning. For example, Senge (1990), in his book on the learning organization, argues that the team is the fundamental level of learning in an organization. It is typically argued that the learning process of a group closely parallels that of an individual and that it can be described as cycles of cognition, action, feedback, and reflection (Edmondson, 2002).

When it comes to studies that focus on the organization as the proper level of learning, the notion of learning is typically used in a more metaphorical sense. That is, groups or organizations are assumed to learn in a way analogous to individual learning, and concepts found in theories of individual learning are extended to the group or organizational level (Hedberg, 1981). This approach has raised criticism of anthropomorphism. However, some would argue that this criticism is irrelevant, and that the notion of learning should be understood as something qualitatively different when applied to a collective entity like an organization compared to individuals. More specifically, it has been argued that organizational

learning should be interpreted from a cultural rather than from a cognitive perspective, and that the notion of organizational culture would be useful for conceptualizing the collective aspects of organizational learning, and thereby, for understanding learning at an organizational level (Cook and Yanow, 1993). Accordingly, learning is assumed to be embedded in collective assumptions and interpretative systems, routines, technologies, and cultural practices. This view comes close to what below is called a sociocultural approach to organizational learning.

A fourth level of analysis is the level of inter-organizational learning. Much research on organizational learning has concerned learning within or of organizations, where the organization has often been treated as a self-contained system with fixed boundaries that operates to a large extent independently of other organizations in the environment. Today, there is a growing interest in new forms of organization at an inter-organizational level. Networks, clusters, innovation systems, and partnerships are a few such examples. Other examples include multinational corporations and joint ventures. A number of new issues of organizational learning are raised at this level of analysis, for example, issues concerning learning and innovation, or learning under conditions of competition and cooperation. At the same time, many of the issues dealt with above are also highly relevant to this level of analysis. Presently, there are few systematic studies of learning processes and outcomes at an inter-organizational level of learning.

Although there are a lot of studies that treat organizational learning as a process that should be analyzed and studied on one or the other of the four different system levels distinguished above, there are also studies that emphasize organizational learning as an interplay between different system levels or as a multilevel process. The latter position is taken, for example, by Crossan *et al.* (1999) in an article that emphasizes that a theory of organizational learning needs to consider the individual, group, and organizational levels. The latter authors also develop a framework including four subprocesses of organizational learning that link the individual, group, and organizational levels.

Organizational Learning as Change and/or Stability

According to many definitions, organizational learning is defined as an experienced-based process of change. It is sometimes also stated or implied that organizational learning means improvements of the actions or the performance of the learning subject, whether this is an individual, a group, or an organization. However, this kind of definition raises a number of issues that need to be addressed. First, observations of changes and adaptations to environmental events do not automatically mean that

these changes are the result of a learning process. On the contrary, changes in actions and improved performance may occur for a number of reasons that have little or nothing to do with learning, for example, situational factors that trigger certain changes in behavior. Second, learning processes are not always, for a number of reasons, mirrored in observable behavior. There may, for example, be situational factors such as a lack of sufficient resources or adequate tools that constrain behavior. In line with this, some scholars have defined organizational learning in terms of changes in potential behaviors rather than actual behaviors (Huber, 1991).

Third, observed changes in action may not be positively related to the performance of the individual, the group, or the organization. On the contrary, we may, in some instances, talk about negative learning in the sense that the learning outcomes may for some reason be undesirable, or in fact, negatively related to performance. That might happen if, for example, the members of a team acquire a form of learned helplessness and an accompanying lack of self-confidence. Thus, the learning process may under some conditions unintentionally result in a deterioration of individual or organizational performance. Furthermore, learning processes, whether they result in positive or negative outcomes from a performance perspective, are in many cases neither conscious nor intentional. Thus, processes of learning, change, and improvement in performance may be totally different processes that need to be clearly distinguished both conceptually and empirically.

While recognizing that learning and change may be two different processes, many researchers have defined levels of learning in terms of the character of the individual and/or organizational changes that are implied by the concept of learning. For this purpose, distinctions have been made between: (a) changes that occur within a given framework, for example, within a given set of beliefs or values or within a given organizational structure or situation, and (b) changes that represent a break with and something that goes beyond the given (Ellström, 2001). Perhaps the most well-known version of this distinction is the one made by Argyris and Schön (1978) between single-loop and double-loop learning. More recently, related and in some respects parallel distinctions have been proposed between first- and second-order learning; adaptive (reproductive) and developmental (innovative) learning; incremental and radical learning. While the former type of learning in each pair has a focus on improving or refining existing procedures or capabilities, the latter type of learning has a focus on the more radical change of institutionalized practices and the development of new capabilities.

Learning as refinement of existing structures and processes may be viewed as a way of reproducing or stabilizing an organization or a social system over time. In a sense, then, reproduction and change (transformation) of a social

system may be viewed as two sides of the same coin. In line with this, arguments have been raised to the effect that it is important to find a balance between radical and incremental change or between the exploration of new alternatives and the exploitation of existing knowledge and technologies, and that the returns to fast learning and change are not all positive (March, 1991). Others would take a further step, and argue – from a cultural perspective – that organizational learning is not just about change, but also about organizational stability and the maintenance and preservation of an organization's activities and cultural practices (Cook and Yanow, 1993). Thus, organizational learning could be seen as a means for cultural reproduction as well as a means for transformation.

Content and Processes of Organizational Learning

The third dimension distinguished here concerns the content and processes of organizational learning. While content refers to what is learned (e.g., knowledge), the process of learning refers to how learning takes place. For present purposes, the following three main approaches may be distinguished with respect to the content (outcomes) and processes of organizational learning: the cognitive-behavioral approach, the sociocultural approach, and the knowledge-creating approach (cf. Paavola *et al.*, 2004).

The cognitive-behavioral approach – an approach strongly anchored in the work of James March and his associates (e.g., March, 1991; March and Olsen, 1976) – has had a strong influence on theory and research on organizational learning for several decades. A basic assumption underlying this approach is that organizational learning is about the development and change of routines through the accumulation of experience. The key term routine includes rules, norms, procedures, strategies, and technologies that are assumed to guide actual behavior in and of organizations and their subunits. In some studies, the notion of routine is treated as explicitly formulated prescriptions. Other studies apply a broader definition and view routines as distributed procedural knowledge (knowing-how), skills, or habits. Still others broaden the view of the products of organizational learning to also include declarative knowledge (knowing-that) and mental models at the individual level, and at the organizational level interpretative systems or shared mental models and frameworks (Kim, 1993; Hedberg, 1981). Organizational learning is conceived of as a process where routines, beliefs, and actions adapt incrementally to past experience through feedback from organizational actions and their outcomes in relation to targets. Specifically, March and Olsen (1976) depict organizational learning as a cycle comprising four stages, including individual beliefs, individual action, organizational action, and environmental outcomes (responses).

The sociocultural (or situated) approach to organizational learning focuses on culture as a core concept. Rather than conceiving learning as a process of knowledge acquisition through experienced-based changes in cognition or action, there is a focus on learning as participation in work practices and activities. Furthermore, what is learned is assumed to be in a fundamental sense connected to the conditions under which it is learned, that is, it is in this sense situated. In line with this, a main tenet is that learning cannot be separated from working and other social practices where it is assumed to take place or be used (Brown and Duguid, 1991). On the contrary, learning is defined as a matter of participation in practices, and indeed, as an aspect of “legitimate peripheral participation” in “communities of practice” (Lave and Wenger, 1991). In line with this focus on participation, learning about practice is less central than learning to become a practitioner. Thus, processes of identity formation are viewed as important aspects of learning. Furthermore, learning and processes or activities (knowing) rather than content or products (knowledge) are emphasized. The sociocultural approach has received considerable attention in recent years as an alternative to the cognitive-behavioral approach, which has been criticized for a decontextualized view of knowledge, and for separating individual and organizational learning. However, critics of the sociocultural approach have, to some extent, reversed these arguments. Thus, it has been criticized for a tendency to reduce learning to an aspect of participation, thereby making it impossible to analytically separate learning and other organizational processes.

The knowledge-creating approach is based on the view that the production, transformation, and utilization of knowledge are fundamental for understanding organizational learning (Paavola *et al.*, 2004). Learning is viewed as an interplay between intra- and inter-individual (social) processes of knowledge creation. The content of learning is knowledge or competence and mechanisms (processes) of learning are typically conceptualized as cyclical processes of problem-solving and knowledge transformation. Perhaps the most well-known framework within this approach is the model proposed by Nonaka and Takeuchi (1995). This model assumes that knowledge creation could be understood as a cyclical process of knowledge conversion based on the interaction between tacit and explicit knowledge. Four modes of knowledge conversion are distinguished: socialization, externalization, combination, and internalization. Another example of a model within this approach is the model of expansive (innovative) learning proposed by Engeström (1999). This model, based on activity theory, views learning as a cycle of epistemic or learning actions starting with the questioning of prevailing practices in an organization. The learning cycle proceeds through an analysis of the existing situation and the creation, testing, and implementation of a

conceptual model of a new idea that is assumed to explain and provide a solution to the problematic situation that initiated the learning process. This model has also been used as an intervention method for facilitating innovative learning in organizations. A third example is the knowledge evolution cycle proposed by Zollo and Winter (2002) based on the evolutionary process of variation, selection, replication, and retention. Underlying this model is a distinction between three learning mechanisms called experience accumulation, knowledge articulation, and knowledge codification.

Conditions and Practices of Organizational Learning

Although organizational learning is sometimes viewed as a natural, continuous process of adaptation in response to internal or external events – processes that may not be conscious or intentional – there is much evidence that indicates that processes of learning are easily interrupted by different kinds of barriers or constraining factors. A conclusion drawn from such observations is that organizational learning needs to be consciously facilitated and supported by consciously planned interventions. Thus, it is assumed that organizations need to learn to learn. This could mean to learn to carry out processes of deliberative inquiry and reflection, and thereby, develop new routines, new knowledge, or new ways of handling a certain organizational problem or task. In accordance with such an interventionist view of organizational learning, a wide range of organizational learning interventions have been developed based on different conceptions of learning. Examples of organizational learning interventions include: project-based learning and action learning interventions based on notions of reflective practice; process consultation; dialogue meetings; open space technology; and different types of learning laboratories (Dierks *et al.*, 2001; Easterby-Smith and Lyles, 2003).

Most organizational learning interventions are, implicitly or explicitly, based on assumptions concerning factors that are likely to constrain or facilitate organizational learning. The purpose of the intervention is of course to attempt to create favorable conditions for learning. What then, does available research tell us about conditions for organizational learning? What factors are assumed to constrain or facilitate learning? One answer to this question is the influential theoretical model proposed by March and Olsen (1976). These authors distinguish between different types of interruptions or blockages that result in restricted or incomplete learning cycles. Others distinguish between conditions related to structural factors, subjective and cultural factors, and factors related to leadership. Among structural factors, many writers underline the importance of the characteristics of the tasks that the organization

is attempting to handle (e.g., task complexity and frequency). Other studies emphasize that centralized and hierarchical structures, as well as too high a degree of formalization and standardization of work processes, are likely to impede learning; in particular, when such structural conditions are combined with limited opportunities for organizational members to participate in organizational decision making. However, there are radically different views concerning the meaning and consequences of these factors for organizational learning.

Although many organizations are structured by gender, there are a few studies of gender aspects in relation to organizational learning. Rather, organizational learning has been studied as a gender-neutral process. This is the case in spite of the significance of gender in relation to other organizational processes. Considering this research, it is quite likely that organizational demographics with respect to gender – as well as in other respects – would be significant for processes of organizational learning. For example, organizations dominated by men could be expected to provide different conditions for learning compared to organizations dominated by women (Berthoin Antal *et al.*, 2001).

It is also possible to identify a number of factors with respect to subjective and cultural issues. In the analyses proposed by Argyris and Schön (1978), and in later studies, so-called defensive routines developed by individuals for protection from threatening situations are viewed as major obstacles to learning. Many writers also focus on anxiety and fear as barriers to organizational learning, and emphasize the need to create psychological safety and trust in order to counterbalance the feelings of threat and anxiety that may be provoked by organizational learning interventions. A factor that many assume is likely to facilitate learning is the extent to which the organizational culture encourages questioning and critical reflection on what is taking place in the organization. Other characteristics of an organizational culture that are considered by many as supportive of learning would include issue orientation, openness, trust, and norms that emphasize initiative and risk-taking, tolerance toward disparate views, and tolerance for admitting errors. Although conflict is assumed by some to be an obstacle to learning, many researchers would agree on the importance of conflicts as driving forces for organizational learning. An equally important factor is how conflicts are handled and resolved as part of an organizational learning process.

A recurrent issue in the literature on organizational learning is the importance of leadership as a condition for organizational learning and, conversely, the lack of adequate leadership support as an obstacle to organizational learning. Leadership support for organizational learning includes the design of an enabling learning environment in the organization, that is, an environment which has many of the structural and cultural features mentioned

above. Other important leadership tasks would include the provision of organizational resources (e.g., time) for learning and the facilitation of learning on the part of organizational members both individually and collectively, for example, by asking challenging questions, stimulating intellectual curiosity, and acting as a coach or mentor (Sadler, 2001).

Power and Politics

As observed by many writers on organizational learning, issues of power and politics in relation to organizational learning have, to a large extent, been neglected in the past. This is somewhat astonishing considering that power and politics for quite a long time have been recognized as important areas within the more general field of organizational studies. The relative neglect of these issues within the field of organizational learning is also astonishing, given that many writers in this field emphasize the importance of conflicts and contradictions as essential for learning in organizations.

However, the lack of emphasis on power and politics is not uniform across the field of organizational learning. In particular, many writers, from what was previously described as a sociocultural approach and a knowledge-creating approach to organizational learning, emphasize the importance of power relations for understanding organizational learning. In line with the practice orientation of these approaches, and the view of learning as integral to everyday work practices in organizations, it is emphasized that learning practices are embedded in and are enabled or constrained within relations of power (Contu and Willmott, 2003). Consistent with this position, it has also been argued that access (or lack of access) to specific learning practices, as well as the division of labor, and thereby, available opportunities for learning, is shaped by prevailing relations of power in the organization and in society at large. The issue, then, becomes one about who participates and who is not allowed to participate in specific learning practices. Another important issue from a political perspective concerns what is valued as knowledge, and whose knowledge and ideas are recognized in an organization. Power differences between different groups in an organization, for example, between different departments or between groups with different status may be assumed to influence whose learning is recognized and acted upon.

Other important issues that have been raised and debated from a political perspective on organizational learning relate to control and ideology in organizations. According to one position, organizational learning represents the ideology of particular power groups (e.g., the management of an organization, but also different groups of experts) and it is used to mask and legitimize the

interests of these groups. In addition, organizational learning is seen as a mechanism and a methodology of control that aims to discipline members of an organization (Gherardi and Nicolini, 2001). Thus, according to this view, organizational learning is a practice for the management of meaning, and a soft means for the subjugation of employees. More specifically, the latter could mean increasing the legitimacy of the organization in the eyes of the employees as regards its goals, fundamental ideology, and power structure, and thereby, contributing to increased employee loyalty with and support for the goals and values of the business. Although there are a number of articles and books that deal with these and related issues, there is a notable lack of empirically substantiated knowledge concerning these matters.

See also: School Development for Teacher Learning and Change; Workplace Learning Frameworks.

Bibliography

- Argyris, C. and Schön, D. A. (1978). *Organizational Learning: A Theory of Action Perspective*. Reading, MA: Addison-Wesley.
- Berthoin Antal, A., Dierkes, M., Child, J., and Nonaka, I. (2001). Organizational learning and knowledge: Reflections on the dynamics of the field and challenges for the future. In Dierkes, M., Berthoin Antal, A., Child, J., and Nonaka, I. (eds.) *Handbook of Organizational Learning and Knowledge*, pp 921–939. Oxford: Oxford University Press.
- Brown, J. S. and Duguid, P. (1991). Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organization Science* **2**, 40–57.
- Contu, A. and Willmott, H. (2003). Re-embedding situatedness: The importance of power relations in learning theory. *Organization Science* **14**, 283–296.
- Cook, S. D. N. and Yanow, D. (1993). Culture and organizational learning. *Journal of Management Inquiry* **2**, 373–390.
- Crossan, M., Lane, H., and White, R. (1999). An organizational learning framework: From intuition to institution. *Academy of Management Review* **24**, 522–537.
- Dierkes, M., Berthoin Antal, A., Child, J., and Nonaka, I. (eds.) (2001). *Handbook of Organizational Learning and Knowledge*. Oxford: Oxford University Press.
- Easterby-Smith, M. and Lyles, M. A. (eds.) (2003). Introduction: Watersheds of organizational learning and knowledge management. In *The Blackwell Handbook of Organizational Learning and Knowledge Management*, pp 1–15. Malden, MA: Blackwell.
- Easterby-Smith, M. and Lyles, M. A. (eds.) (2003). *The Blackwell Handbook of Organizational Learning and Knowledge Management*. Malden, MA: Blackwell.
- Edmondson, A. C. (2002). The local and variegated nature of learning in organizations: A group-level perspective. *Organization Science* **2**, 128–146.
- Ellström, P. E. (2001). Integrating learning and work: Problems and prospects. *Human Resource Development Quarterly*, **12**, 421–435.
- Engeström, Y. (1999). Innovative learning in work teams: Analyzing cycles of knowledge creation in practice. In Engeström, Y., Miettinen, R., and Punamäki, R. L. (eds.) *Perspectives on Activity Theory*, pp 19–38. New York: Cambridge University Press.
- Gherardi, S. and Nicolini, D. (2001). The sociological foundations of organizational learning. In Dierkes, M., Berthoin Antal, A., Child, J., and Nonaka, I. (eds.) *Handbook of Organizational Learning and Knowledge*, pp 35–60. Oxford: Oxford University Press.
- Hedberg, B. L. T. (1981). How organizations learn and unlearn. In Nystrom, P. C. and Starbuck, W. H. (eds.) *Handbook of Organizational Design. Vol 1. Adapting Organizations to Their Environments*, pp 3–27. Oxford: Oxford University Press.
- Huber, G. P. (1991). Organizational learning: The contributing processes and the literatures. *Organization Science* **2**, 88–115.
- Kim, D. H. (1993). The link between individual and organizational learning. *Sloan Management Review* **35**, 37–50.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science* **2**, 71–87.
- March, J. G. and Olsen, J. P. (1976). Organizational learning and the ambiguity of the past. In March, J. G. and Olsen, J. P. (eds.) *Ambiguity and Choice in Organizations*, pp 54–68. Oslo: Universitetsforlaget.
- Nonaka, I. and Takeuchi, H. (1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford: Oxford University Press.
- Paavola, S., Lippinen, L., and Hakkarinen, K. (2004). Models of innovative knowledge communities and three metaphors of learning. *Review of Educational Research* **74**, 557–576.
- Sadler, P. (2001). Leadership and organizational learning. In Dierkes, M., Berthoin Antal, A., Child, J., and Nonaka, I. (eds.) *Handbook of Organizational Learning and Knowledge*, pp 415–427. Oxford: Oxford University Press.
- Senge, P. M. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Doubleday.
- Simon, H. A. (1991). Bounded rationality and organizational learning. *Organization Science* **2**, 125–134.
- Zollo, M. and Winter, S. G. (2002). Deliberate learning and the evolution of dynamic capabilities. *Organization Science* **13**, 339–351.

Further Reading

- Antonacopoulou, E., Jarvis, P., Andersen, V., Elkjaer, B., and Hoyrup, S. (eds.) (2006). *Learning, Working, Living. Mapping the Terrain of Working Life Learning*. New York: Palgrave Macmillan.
- Cohen, M. D. and Sproull, L. S. (1996). *Organizational Learning*. Thousand Oaks, CA: Sage.
- Easterby-Smith, M., Araujo, L., and Burgoyne, J. (1999). *Organizational Learning and the Learning Organization. Developments in Theory and Practice*. Thousand Oaks, CA: Sage.
- Leithwood, K. and Seashore, K. L. (eds.) (1999). *Organizational Learning in Schools*. Lisse, PA: Taylor and Francis.
- March, J. G. (1999). *The Pursuit of Organizational Intelligence*. Malden, MA: Blackwell.
- Marsick, V. J. and Watkins, K. E. (1999). *Facilitating Learning Organizations*. Aldershot: Gower.
- Moingeon, B. and Edmondson, A. C. (eds.) (1996). *Organizational Learning and Competitive Advantage*. London: Sage.

Program Planning

A L Wilson, Cornell University, Ithaca, NY, USA

R M Cervero, University of Georgia, Athens, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

Those who study and practice adult education have tended to be more interested in questions about what and how adults learn than in questions of how to organize adult education. Yet adults intentionally decide to learn something in some sort of organized fashion that typically requires some individual or group to take on the responsibility of designing and organizing educational activities. Such educational responsibilities have been called different names – program planning, program development, curriculum and instructional design, program organizing, and program design – but regardless of the label such activities typically constitute decisions and actions about what is to be learned, how the learning is to take place, who the learners and the teachers are, and what the education is for. From the beginning of the formal organizing of the study and professional practice of adult education, there has been an interest in developing program planning theory as an aide to those organizing education for adults. As Sork and Newman (2004) note, this theoretical pursuit seems more peculiar to academics in North America than elsewhere. Nevertheless, there has been a continuing effort, both as a result of academic interest as well as practitioner inquiry, to develop principles and procedures – that is, theories and methods – for designing and implementing programs for the education of adults.

Beginning in the 1920s in the United States, such effort has produced literally hundreds of program planning theories in adult education (Sork and Buskey, 1986; Wilson and Cervero, 1997). Two observations have emerged as a result of this proliferation. First was the growing sense that amidst such variety there was actually more uniformity than was apparent which Sork and Caffarella (1989) demonstrated when they developed their generic model of program planning. Second was a sense of a growing gap between what theorists prescribed in planning theories and what practitioners actually did when planning programs (Cervero and Wilson, 1994; Sork and Caffarella, 1989). That gap is a function of many factors, chief of which is the failure of planning theorists to develop adequate theories of context and practical action (Cervero and Wilson, 2006). Our view is that theory should be able to both explain and enable practical action (Foley, 1999). Most adult education planning theory, until fairly recently, represented only one main tradition for understanding and enabling practical planning action. Yet most of that theory was limited in its understanding of what actual planning practice required and instead depended

upon normative prescriptions of what planners should do, which tended to be a highly idealized form of instrumental rationality (Cervero and Wilson, 2006).

The value and utility of program planning theory lies in its ability to provide plausible accounts of actual planning practice, interpretations of what is practically possible when planning programs for adults, and politically strategic practices that foster ethically informed goals (Bernstein, 1976; Forester, 1989). The purpose of this article is to review three traditions of planning theory – conventional, deliberative, and critical – in terms of how each answers the questions of what do planners do when planning programs, what really matters when planners plan programs, and what is good planning. Addressing these questions will enable us to examine the different understandings of context and action that each tradition depends upon in order to provide its theoretical accounts and practical prescriptions. We conclude with a discussion of the challenges facing adult education planning theory and practice.

Conventional Planning Theory

As is often the case with education more generally, planning theorists in adult education have tended to think of planning ahistorically (Wilson, 2005), yet it sometimes helps to see what is now by looking at what has been. There is one tradition that has dominated historically and theoretically, and continues to dominate planning discussions. Sork and Newman (2004) call this tradition the conventional which also has been described as the classical (Cervero and Wilson, 1994). Although clearly evident since near the beginning of the twentieth century (Pinar *et al.*, 1995), conventional planning theory is best landmarked by the appearance of Ralph Tyler's basic principles of curriculum and instruction in 1949. In his book, Tyler (1949) codified the essential conventional planning logic and method: "if an educational program is to be planned . . . it is very necessary to have some conception of the goals that are being aimed at. These educational objectives become the criteria by which materials are selected, content is outlined, instructional procedures are developed and tests and examinations are prepared" (p. 3). This is Tyler's rationale: that without objectives no other planning decisions can be made and that each decision logically induces the next, hence the numbering of his four famous questions (and the beginnings of decades of debate in

curriculum theory). This sequential logic driven by the definition of objectives is the underlying thought – theory – of all conventional planning theory, which Malcolm Knowles in his 1950 book *Informal Adult Education* made so clear and nearly all the adult education planning theory since has personified (Sork and Buskey, 1986; Wilson and Cervero, 1997). Conventional theorists in adult education have elaborated and refined conventional logic by adding steps such as analyzing planning contexts and administering programs but have not fundamentally altered the prescribed “sequentially and logically ordered set of tasks in which educational planners first assess learning needs, then develop learning objectives from assessed needs, next design learning content and instructional formats to meet learning objectives, and finally evaluate learning outcomes in terms of whether the objectives were achieved” (Wilson, 2005: 525). The proliferation of conventional planning theory that Sork and Buskey (1986) first identified and Sork and Caffarella (1989) codified has continued without abatement (Cervero and Wilson, 2006; Sork and Newman, 2004).

According to the conventional view, what do planners really do? Planners follow the steps. What really matters? Following the steps in order, for if they are not followed in order, then logic of the rationale breaks down and the resulting planning for education is not good. What is good program planning then? Following the sequence in the prescribed order? According to the conventional view, principally and practically the systematic process of planning always starts with objectives, proceeds through multiple design and implementations steps, and concludes with the evaluation of whether the objectives have been obtained. The efficacy of conventional theory depends upon following its internal logic. If the logic is violated, then consequences are moot. What really matters in conventional planning theory is the understanding of context and practical action that drive such theories or, perhaps more accurately, the lack thereof. These planning theories represent a form of professional scientific logic that promotes “instrumental problem solving made rigorous by the application of scientific theory and technique” (Schon, 1983: 21), or, as more commonly known today, technical rationality or instrumental reasoning. The critique of technical rationality as a limited basis for the performance of professional activities like education is now well developed. Its attempt to adopt rigorous, rational problem solving as the basis for professional practice dims in the face of the actual interactive complexities of humans and organizations that planners really have to address in order to plan. So a serious limitation of conventional planning theory is the assumption that rational planning procedures will work equally well in any circumstances and that practical action is defined by and limited to the execution of procedures. Rational planning, however, does not work the same in every situation and, indeed, can actually disguise the workings of power in certain circumstances (Cervero and

Wilson, 2006). The rational planning of conventional theory typically does not eliminate but routinely ignores the noise of context and people (Forester, 1989). So conventional planning theories provide neither a plausible account of practice nor ethical standards for practice, so they therefore provide no way of imagining the need for politically strategic planning action. Within its instrumental frame, has conventional theory provided a set of principles and procedures to which planners need to attend? Certainly it has, because planners who do not know the differences between needs assessments and instructional design, between evaluation and objectives, between administrative plans and curriculum plans will be of little use to their organizations. Not only do planners need to know these concepts and principles, they likewise need the wealth of technical methods so abundantly provided by conventional theory for conducting needs analyses and audience surveys, curricular organization and site management, evaluation and organizational analysis. Yet to rely solely on conventional planning principles and instrumental technique runs the risk of ignoring the exigencies of real people planning real educational programs in real organizational settings. To get a sense of efforts to understand and enable action in such conditions, we turn now to a related but distinct tradition called the deliberative.

Deliberative Planning Theory

If Ralph Tyler personified the logic of the conventional tradition, Joseph Schwab (1969) named its major limitations: it failed to understand the way in which people made planning decisions and the conditions in which such decisions were made. Practitioners for decades have routinely indicated that although the conventional principles and techniques are helpful, they rarely represent how actual planning decisions get made in the complex human and organizational contexts that planners work. In critiquing the procedural emphasis of conventional planning theory, particularly the requirement to always begin with objectives, proponents of the deliberative traditions (Schwab, 1969; Walker, 1971) argued that planning was best understood as a process of practical (as opposed to instrumental) reasoning. Practical reasoning as a deliberative process requires analyzing the context and then making the best judgments possible about what to do, given the restrictions and possibilities of the specific set of circumstances. Another major insight of this tradition is that planners may neither need to complete all the steps of the conventional tradition nor address them in the sequentially prescribed order. Houle’s (1972, 1996) insight was that planning was better understood as decision points (rather than steps) that could occur at any time and often simultaneously undermined the preferred sequential logic of the conventional view. Despite decades

of producing planning models, the only significant theoretical debate in the adult education planning literature has been about whether the steps have to be followed in order (Wilson and Cervero, 1997). Planning theory has moved steadily from a discounting of the preferred sequential enactment of planning tasks to an iterative understanding of planning human decision making within the constraints of specific contexts (see Caffarella, 2002; Sork, 2000; Sork and Newman, 2004). This theoretical move is believed to diminish the acknowledged theory–practice gap characterizing conventional theory. While promising in that the deliberative tradition does bring acting people into variable settings, “this tradition’s invocation of ‘deliberation’ . . . is actually similar to what the conventional view holds as rational problem solving . . . because deliberation essentially represents the decision-making aspects of implementing rationalist problem-solving steps” (Cervero and Wilson, 2006: 246).

According to this tradition, planning practice is best understood as people deliberating about the best course of action to follow given the constraints of the circumstances. This is what planners do. In addition, what matters in this view are the specifics of situational constraints, the values of the planners deliberating about what to do, and the judgments about courses of action that ultimately obtain, given the confluence of constraints and values. So good planning is defined by defensible judgments that justify action in terms of constraints and values (Schwab, 1969; Walker, 1971). Such an understanding of planning does begin to provide a plausible account of practice because it adds to the proceduralism of conventional theory the actual people who deliberate about and enact the decisions on planning as well as including the understanding that such decision making is always constrained and/or enabled by the context. Bringing people and settings into the equation is a promising advance on understanding practical action in context. Where this view is limited, however, in naming more precisely how people act in context, what defines context and how it works to constrain or enable, or whose values should matter. To see what more precision might look like in terms of practical action and context, we turn next to critical traditions.

Critical Planning Theory

Whenever people have sought to change oppressive social, political, and economic conditions, there have usually also been people who have organized education to facilitate such change. With antecedents in the nineteenth century and before, and throughout the twentieth century and beyond, there have been people who have organized adult education for settlement houses, in labor movements, at workers colleges, in civil rights movements, for anti-war movements, in anti-globalization efforts, and whenever and wherever

people have needed to learn in order to seek change. Places like Hull House, Highlander, and Antigonish, and people like Jane Addams, Myles Horton, and Moses Coady plus scores of lesser-known but no less involved people and places provide a lengthy history of educators working to redress social and political injustice. There has been very little planning theory in adult education to explain and enable this kind of educational work; indeed, it may well be that movement, union, and civil rights educators rarely even thought or think of themselves as educators. Nevertheless, there has evolved various critical theories of planning. Perhaps most prominent in terms of planning theory would be Paulo Freire. Sometimes implications for planning theory are noted in his *Pedagogy of the Oppressed* (Sork, 2000); others are explicit in naming it as planning theory (Boone *et al.*, 2002; Wilson and Cervero, 1997). Specific chapters in *Pedagogy* describe organizing and implementing strategies to foster critical transformation and social change. John Forester (1989) has developed a theory of planning based on and directed toward fostering Habermasian communicative ethics and participatory practices. Cervero and Wilson (1994, 2006) have been described as beyond the conventional (Sork and Newman, 2004: 112) because of their theory’s focus on fostering democratic planning by negotiating interests in relations of power.

Planners in the critical traditions typically understand the world to be structured with inequitable power relations and seek to organize possibilities for changing those relations by using education to enhance democratic participation in the planning, implementation, and consequences of educational actions. Education is thus a political and ethical activity whose purpose is to promote social justice and emancipation through a restructuring of society’s cultural, political, and economic relations – this is what really matters. Good critical planners anticipate the structural distortion caused by inequitable power relations and seek to counteract such distortion by fostering democratic participation. Critical traditions fully incorporate the insights of the conventional planning theories such as the need for a professional repertoire of technical procedures and deliberative insights concerning decision making and the influence of context. But in the critical traditions the context is described specifically as power relations and practical action as tactics to enhance participation. Thus, the value positions underlying critical planning are starkly articulated. While value positions of planners and organizations are allegedly neutralized by the deployment of rational problem solving in the conventional view (the faith in the efficacy of rationality is the value position), in the deliberative traditions values are clearly acknowledged to be a formative force in shaping decisions about planning, but no values are specified. The quest for social justice and democratic participation are the platform values for planning thought and action in the critical. Overly celebrated as well as routinely ignored, there is a wealth of historical accounts

and present-day examples that exemplify the plausibility of this tradition, which indicate that its interpretative stances are practically possible, and that the traditions represent ethically informed and politically strategic visions of what should be.

Conclusion

Sork and Newman (2004) have noted that most theoretical writing about program planning has emanated from North America, and we, as they, wonder what a genderized, feminized, colorized, or other-culturalized understanding of planning might be. Until the 1990s, almost all theoretical discussion represented just one tradition – the conventional (Sork and Buskey, 1986; Sork and Caffarella, 1989). That discussion tended to debate the details but not the underlying formative assumptions and structures of thought that enabled their proliferation. Because of the dominance of the conventional tradition (and probably because of who has been writing theory), there has been little awareness of possibilities of rethinking what has been the norm. Even so, as we have indicated here, there are at least three traditions of planning theory that purport to understand and enable practical planning thought and action: the conventional, the deliberative, and the critical (Cervero and Wilson, 1994, 2006; Sork, 2000; Wilson, 2005).

There are a number of challenges we could recount by way of conclusion. Knowles (1950) argued that it was time to get beyond trial and error and make adult education planning theory more scientific, by which he invoked the instrumental problem solving of the Tyler rationale. In one of the few studies of actual planning practice, Pennington and Green (1976) demonstrated empirically the deliberative tradition's understanding of planning as contextually bound decision making; Houle (1972) had already developed a theory of adult education planning defined by decisions points. Pennington and Green (1976) asked questions which we are still asking: what do planners really do and why is there a theory–practice gap. They concluded we need better planning theory because actual planning was “superficial at best” and planners often planned by “the seat of their pants” (p. 20). But even as their investigations illustrated the insights of the deliberative tradition, Pennington and Green advocated for the conventional in order to “strengthen the operational aspects of those decisions” (p. 23). Sork and Caffarella (1989) likewise sought a “firmer foundation” for planning theory because the planning literature promoted an “idealized process” that did not take “into account the exigencies of day-to-day responsibilities of practitioners” (p. 243). They concluded that the gap between theory and practice might be widening because theory was becoming “increasingly irrelevant to practice” (p. 243). Such irrelevancy might be because planners took “shortcuts” in their

planning work and because context does influence planning practice considerably, as Pennington and Green's investigations showed. Sork (2000), in noting the selective origins of most planning theory, asked whether we could develop planning theory that would “both explain the complexities of planning and guide those involved” (p. 179). Although he remained skeptical of success of accomplishing that, he nonetheless proposed his question-based approach to planning as a way to understand and guide thought and action in planning practice. Sork and Newman (2004) again questioned the narrowness of the male, North American view on planning theory and noted the perhaps not-so-coincidental dominance of technical rationality in planning theory that tries to make such theory more scientific. They invoke Malcolm Knowles's injunction about andragogy to argue that “it may be more accurate to say that program development can be thought of as *either* an art or a science, but might more productively be considered a bit of both” (p. 97; original emphasis).

The greatest challenge to planning theory remains the continuing “split between what theorists prescribe and practitioners do” (Wilson, 2005: 528). It is, no doubt, the case that there are those who are responsible for developing programs for adults do so by the seat of their pants, as Pennington and Green once characterized much adult education planning. Even so, there are many experienced practitioners who are well aware of the political, ethical, and technical dilemmas and opportunities that face them daily. Because of the dominance of the conventional theory, too many theorists continue to recommend rationalist problem-solving sequences as the only response set to the practical choices practitioners must routinely make. The consequence is the continuing theory–practice gap: “The practical, empirical, and theoretical question remains: how to integrate the technical, political and ethical dimensions of planning practice in order to better understand and act in the contexts planners actually work in” (Wilson, 2005: 528; see Caffarella, 2002; Cervero and Wilson, 1994, 2006; Sork, 2000, for integrating efforts). The gap continues because we have yet to develop the theoretical understandings of human action in context that represents plausible accounts of practice that provide politically strategic practices that foster ethically informed goals. We believe planning theories should explain and enable action and have described three related but different traditions by discussing them in terms of how they understand practical action in context.

In closing, we would like to make the following points. Each tradition contributes to creating the theory–practice gap by how they understand practical action in context while each also contributes to resolving that tension. The conventional creates the widest gap because of its rationalist assumptions and its contention that rationality neutralizes the effects of context. Yet its bountiful supply of

technical planning procedures is crucial to actual planning practice although procedures cannot be mistaken for practice itself as if often the case. The deliberative promisingly attempts to understand the unavoidable interaction of people and context yet its view of practical reasoning is but a variant on technical rationality. The critical gets closer to providing an understanding of practical action in context with its view of a world saturated with power relations and planners as politically strategic and ethically charged organizers. It is closer but not exhaustive. Clearly planning theory needs more empirical investigations of actual planning in many different settings if we are to further develop the insights and injunctions of the three traditions. While there has been a plethora of theoretical modeling and an abundance of technical prescriptions, there has been relatively few investigations of actual planning practice. While the discipline of adult education has emphasized adult learning, it has depended upon an assumed professional ethos of the planner as technical facilitator of planning processes. There is certainly merit to this conception but it is quite limited as the deliberative and critical traditions would suggest. Clearly planners represent and take sides when planning programs and we need to learn more about which sides we should take and how best to take them.

Bibliography

- Bernstein, R. (1976). *The Restructuring of Social and Political Theory*. Philadelphia, PA: University of Pennsylvania Press.
- Boone, E., Safrit, R., and Jones, J. (2002). *Developing Programs in Adult Education*. Prospect Heights, IL: Waveland.
- Caffarella, R. (2002). *Planning Programs for Adult Learners: A Practical Guide for Educators, Trainers, and Staff Developers*. San Francisco, CA: Jossey-Bass.
- Cervero, R. M. and Wilson, A. L. (1994). *Planning Responsibly for Adult Education: A Guide to Negotiating Power and Interests*. San Francisco, CA: Jossey-Bass.
- Cervero, R. M. and Wilson, A. L. (2006). *Working the Planning Table: Negotiating Democratically for Adult, Continuing, and Workplace Education*. San Francisco, CA: Jossey-Bass.
- Foley, G. (1999). *Learning in Social Action: A Contribution to Understanding Informal Education*. London: Zed.
- Forester, J. (1989). *Planning in the Face of Power*. Berkeley, CA: University of California Press.
- Houle, C. (1972). *The Design of Education*. San Francisco, CA: Jossey-Bass.
- Houle, C. (1996). *The Design of Education*, 2nd edn. San Francisco, CA: Jossey-Bass.
- Knowles, M. (1950). *Informal Adult Education*. New York: AAEE.
- Pennington, F. and Green, J. (1976). Comparative analysis of program development processes in six professions. *Adult Education Quarterly* **27**, 13–23.
- Pinar, W., Reynolds, W., Slatterly, P., and Taubman, P. (1995). *Understanding Curriculum: An Introduction to the Study of Historical and Contemporary Curriculum Discourses*. New York: Peter Lang.
- Schon, D. (1983). *The Reflective Practitioner*. New York: Basic Books.
- Schwab, J. (1969). The practical: A language for curriculum. *School Review* **78**(1), 1–24.
- Sork, T. (2000). Planning educational programs. In Wilson, A. L. and Hayes, E. R. (eds.) *Handbook of Adult and Continuing Education*, pp 171–190. San Francisco, CA: Jossey-Bass.
- Sork, T. and Buskey, J. (1986). A descriptive and evaluative analysis of planning literature, 1950–1983. *Adult Education Quarterly* **36**, 86–96.
- Sork, T. and Caffarella, R. (1989). Program planning for adults. In Merriam, S. and Cunningham, P. (eds.) *Handbook of Adult and Continuing Education*, pp 233–245. San Francisco, CA: Jossey-Bass.
- Sork, T. and Newman, M. (2004). Program development in adult education and training. In Foley, G. (ed.) *Dimensions of Adult Learning*, pp 96–117. Maidenhead: Open University Press.
- Tyler, R. (1949). *Basic Principles of Curriculum and Instruction*. Chicago, IL: University of Chicago Press.
- Walker, D. (1971). A naturalistic model for curriculum development. *School Review* **80**(1), 51–65.
- Wilson, A. L. (2005). Program planning. In English, L. M. (ed.) *International Encyclopedia of Adult Education*, pp 524–529. Houndsmill: Palgrave Macmillan.
- Wilson, A. L. and Cervero, R. M. (1997). The song remains the same: The selective tradition of technical rationality in adult education program planning. *International Journal of Lifelong Education* **16**(7), 84–108.

Workplace Learning Frameworks

S Billett, Griffith University, Brisbane, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Duality – Two related phenomena who enjoy some degree of interdependence (i.e., the opposite of a dualism).

Educational purposes – The kinds of goals, aims, and objectives to which educational efforts should be directed.

Integrating experiences – Drawing on and reconciling experiences from different settings to assist secure the contributions from both settings and thereby.

Learning environments – Social and/or physical spaces that afford experiences and activities, that by degree invite and are supportive of individuals' learning.

Personal epistemologies – The bases by which individuals initiate, direct, monitor, and evaluate their processes of thinking and acting, and learning, not the least through how they construe and construct what they experience.

Rich or robust learning – Learning that is not restricted in its applicability to the circumstances of its generation.

Socially rich environment – Social settings that provide close guidance, by expert or more experienced social partners as well as opportunities to observe and imitate the performances of more expert counterparts.

Emerging and Growing Interest in Workplace Learning

There is a growing interest in workplace learning. This seems set to continue as the securing of skills throughout working life becomes, globally, an important priority for governments, industry, and workers themselves. Currently, workplace experiences are acknowledged as playing important roles in preparing for working life. These include assisting individuals identify preferred occupations, experiencing and understanding the world of work, and most commonly, contributing to the initial preparation of their skills for paid work. Yet now, and in their own right, these experiences are seen to offer an effective means for maintaining skill currency right

across working lives, meeting enterprises' special needs, and maintaining the effectiveness of older workers. Together, these represent important social and economic purposes and in ways that are often tightly intertwined.

In effect, recently, workplaces have moved from being seen as providing experiences to support learning in educational institutions, to providing important social and economic purposes. These are associated with developing further and maintaining occupational competence, enterprise-specific skills, government economic imperatives for a competitive economy, including assisting workers resist redundancy, enhancing learning occurring in educational institutions and individuals' needs to develop and maintain skilfulness. Such an array of expanded purposes warrants adequate frameworks for understanding about how these kinds of learning might be provided for through workplace experiences.

There has also been a growth in scholarship about learning through work providing conceptual and procedural accounts of adults' learning in and throughout working life. These developments can inform both the kinds of pragmatic and strategic concerns outlined above, as well as those about learning in general, and how learning can be supported through activities in workplace settings. These include understanding the central role of the negotiations between the personal and social contributions to that learning. It is through a consideration of such purposes and these conceptual and procedural developments that frameworks for considering, organizing, promoting, and evaluating adults' learning through work can be best advanced.

Such frameworks need to be informed by a consideration of the purposes for this learning, conceptions of the processes of learning through work, and in what ways these experiences might be better ordered to secure particular kinds of learning for these purposes. This includes identifying what kinds of practices have demonstrated efficacy in supporting learning and the potential to guide workplace experiences toward particular outcomes. Moreover, there are perennial issues about the worth and legitimacy of the purposes, circumstances, and processes of workplaces as learning environments. Rather than accepting their worth, contributions and legitimacy in terms of a site for learning that predates educational institutions as we know them today, workplaces are seen as being inferior and subordinate to those institutions.

In advancing bases for a framework for workplace learning, this article first outlines some purposes for and

conceptions of learning through work. Then, it discusses procedural and conceptual developments that shape and reshape workplace-learning frameworks. It concludes by offering some parameters for such a framework principally comprising a duality between what the workplace affords learners in terms of opportunities and support, and how individuals engage with these affordances as they learn through their experiences.

Purposes of Learning Through Workplace Experiences

There are three broad categories of purposes for the provision of workplace-learning experiences. The first is contributing to educational processes about preparation for work and working life. This includes assisting individuals understand the requirements for working life and to identify which occupation they are suited to and interested in. For those in the structured program of occupational preparation in schools, universities, and college, workplace experiences are also used to practice, augment, and extend learning secured in educational institutions. Second, for many entering work or occupations, there is no available or accessible education provision. Hence, workplaces are the only source of learning experiences, to develop the capacities for performing that work. Third, as foreshadowed, beyond initial preparation, there are growing imperatives of governments, employers, and workers themselves to maintain skills throughout working life in order to resist redundancy, make effective work transitions as occupational requirements change, and contribute to the workplace's continuity and development and to the economic prosperity of their communities. Added to these are growing concerns about older workers and others who are marginalized groups (e.g., disabled, nonnative speakers, etc.) who often need to maintain their employment viability with limited assistance from their employers.

The prospect for realizing these purposes through learning in workplaces was granted legitimacy and buoyed by the situated-cognition movement of the late 1980s and early 1990s (Brown *et al.*, 1989; Collins *et al.*, 1989). This movement emphasized the particular contributions to learning from the settings in which practice occurred and evidence that strategic as well as situationally specific learning can be realized through such experiences (Raizen, 1991). Yet, the key motivation behind this research was to improve experiences in educational institutions (Resnick, 1987), by understanding processes of transfer from one setting or activity to another. Considerations of situationally distinct goals for practice and contributions for learning through practice offered an explanation for the paucity of the transfer of learning from educational institutions to circumstances and practices beyond educational

institutions (Newman *et al.*, 1984, 1989). Indeed, research into the situated nature of learning often looked to workplaces as sites of authentic practices and how learning arose through participation in work activities. However, more recent accounts have now focused on the learning potential of workplaces, and include not just a consideration of their physical and social settings, but also on those who engage in and learn through work.

Occupation preparation for both the professions and trades has long engaged students in extensive periods of occupational practice in workplaces. Concerns about postschool pathways and transitions are also now leading to schools providing workplace experiences to assist students understand the world of work beyond schools. Moreover, concerns about the transition from university and the applicability of what is learned there to practice in employment is motivating universities in a growing number of countries to engage their students in professional experiences, and finding ways of integrating these experiences with those within higher education programs. Although long exercised in teacher, nurse, and medical education, these kinds of experiences are now being extended across university programs preparing graduates for professional occupations. The cooperative education movement of North America is a longstanding response to this education concern. Yet, even so, genuine efforts to integrate the learning experiences, such as these, are until now the exception, yet are important for securing rich learning from the contributions of both settings.

Therefore, the purposes for considering workplaces as learning environments in their own right and as making distinct and legitimate contributions to educational institutions' programs and curriculum stand as important bases for understanding and appraising workplace learning.

Conceptions of Workplaces as Learning Environments

To understand workplaces as worthwhile places to learn in, it is necessary to capture their qualities, demonstrate their legitimacy, and advance ways in which their contributions can be fully engaged. Certainly, there has been long-held acceptance of the value of learning through practice and through workplaces. Plato describes the process of learning to become artisans and artists as that occurring through association, imitation, and practice, starting with play, within the family of artists and artisans and in the circumstances of practice. As Lodge (1947) notes "... at first, the imitation would be playful and childish, carried out with such toy tools as a child could handle. Later it would become more deliberately purposive. Practice produced technical proficiency in detail and

the growing boy would act first as his father's 'helper', then as his associate, and would eventually himself become the head of a family, and the centre from which further training in the family craft would radiate" (Lodge, 1947: 17). Yet, far earlier, similar processes were used in ancient Babylonia (Bennett, 1938). Subsequently, traditions of learning through practice for occupational purposes extended well into the medieval times, buoyed by the practices and power of guilds (Deissinger, 2002). Indeed, most of the buildings, structures, and artifacts from these times, and at which we now marvel, were the product of learning through work practice. The skills for constructing the great European cathedrals progressed in similar ways to those which Plato described in Hellenic times (Gimpel, 1961).

Therefore, long before education institutions were created for these purposes, generations of skilled artisans across Europe and elsewhere learned their skilled work through work-related learning experiences alone. Indeed, it was the impact of the decline of the guilds and the erosion of work-based experiences in some European countries that created the need for specific educational institutions. However, despite the growth of educational institutions for such purposes, experience of practice has been sustained to this day in the trades and the professions. Indeed, they are now being enacted more widely across countries and educational sectors. Yet, there is often a parsimonious acknowledgement of the quality of learning processes and outcomes that arise through these practice experiences. Typically, they are still seen as experiencing practice, rather than important sites for learning and as being less desirable and potent than learning through participation in educational institutions. Even in programs for the trades and professions, these experiences are often seen as primarily to apply or practice what has been learned in educational institutions. For instance, rarely are these components explicitly integrated into the curriculum or assessed, in ways that are commensurate with their contribution to student learning.

So, despite these traditions and now-emerging imperatives for workplace-learning experiences, their status still stands as being inferior, rather than different from what is provided and experienced in educational institutions. A significant barrier is legitimating learning through practice in an era where strong associations between teaching and learning have become embedded through universal compulsory and lengthy education, and in which teaching is seen as the way in which important learning is best mediated. While those who have learned occupational practice invariably report the importance of learning through practice, it still lacks the broader legitimacy, standing, and credibility of certified learning through participation in educational institutions. This seems to be the case across all education sectors, including those whose purpose is to develop specific vocational

knowledge (i.e., schools, colleges, and universities). Indeed, institutions express concern that their academic standing may be jeopardized through the inclusion of work-based experiences. In addition, there are fears that some educational norms, such as a liberal and critical education are being threatened by such inclusions (Boud *et al.*, 2001). Not surprisingly, in the discourses of educational institutions and practice, teaching is privileged over learning. Therefore, as (Marsick and Watkins, 1990) identified earlier, the absence of written curriculum, qualified teachers, and experiences purposely focused on individuals' learning, may lead to workplace experiences as being inevitably inferior.

However, such claims do not stand scrutiny. As studies from anthropology have found, robust (i.e., transferable) learning can arise as much through experiences outside of those organized and enacted in educational institutions, as in other settings (Rogoff and Gauvain, 1984). Indeed, anthropological literatures have provided helpful accounts of this kind of learning and evidence of their efficacy (Lave, 1990; Pelissier, 1991; Rogoff, 1990; Rogoff and Lave, 1984). So, experiences in educational institutions are not necessarily better at developing these kinds of knowledge. More likely, it is the kinds and combinations of activities and interactions that are afforded learners and how they engage with them that are central to the development of this knowledge, and not where these experiences occur. These then suggest a need to understand and appraise workplace learning through a consideration of both personal and social contributions to the processes and outcomes of that learning.

It seems that frameworks for understanding learning through work in contemporary times are well-served by a long tradition of learning through practice. These accounts also characterize workplaces as environments for learning, not just for experiencing and trialing and refining knowledge learned elsewhere. They refer to contributions such as activities, interactions, artifacts, and others as being pedagogic and purposive for learning, thereby offering frames for pedagogies and curriculum models for the workplace. Importantly, rather than didactic and school-like accounts of supporting learning, indirect forms of guidance and highly active roles for learners emerge as key qualities of these learning environments. The organization of learning through activities and interactions, and the active process of observation and rehearsal stand as central elements of the workplace curriculum and pedagogic practices as identified in these accounts.

Therefore, much of a first generation of recent research was directed at understanding how workplace experiences might improve learning experiences in educational institutions (i.e., schools, colleges, and universities). However, building on these accounts, a second generation of

research is now giving more attention to the particular attributes of sites and circumstances of practice (e.g., work and workplaces) as places to both participate and learn. For instance, theoretical and procedural considerations of the pedagogic qualities of different kinds of work (Colin, 2004; Nerland and Jensen, 2006), learning through errors (Bauer and Mulder, 2007), the active role of the learner (Billett, 2006b), their subjectivity and sense of self (Somerville and Abrahamsson, 2003), and the complex entanglements between personal interests and capacities and those of the workplace (Fenwick, 2004) have arisen from quite different traditions than mainstream education.

Conceptual and Procedural Advances

Conceptual advances associated with learning through work are aligned to those associated with the processes of learning more generally. Central among these are the greater emphasis on learning, instead of teaching, as is consistent with the growing purchase of constructivism, and the acknowledgement of situational contributions, in ways analogous to social constructivism. The former indicates acceptance of an expansive and active role for the learners, and positions them as being central mediators of what they experience and how and what they learn, thereby emphasizing personal epistemologies. Here, central to individuals' participation, mediation, and learning is their agency, capacity, and subjectivity. These are central to how they construe and construct their knowledge. The later acknowledges how the tasks, activities, and settings in which learners participate afford particular kinds of contributions, which in terms of workplace learning can be understood as pedagogic qualities of work settings. Both of these emphasize participation in work and learning as a dualistic conception comprising the contributions of social setting and the person. These advances are now discussed briefly in turn.

Self and Agency of Learners

The need for learners to be actively engaged in the process of learning is widely accepted within many, and probably most, contemporary theories and concepts of learning. This includes how they construe, construct, and interact with what they experience. However, this is particularly salient to considerations of learning through work and in workplaces. In provisions for learning offered through educational institutions, the teacher has a key mediating role in assisting the learner and guiding his/her development. However, while a range of guidance is available in work settings, including close guidance by more expert others, although usually on an intermittent basis, learners will necessarily mediate much of their own learning (Fuller and Unwin, 2003). This is also likely to be true of provisions that seek to integrate that learning in

both the educational and workplace setting, such as in apprenticeships, and now increasingly for professional preparation. This is because it is the learners who will have experiences in both kinds of settings and others' role are to assist them maximize their learning through integrating the contributions of both kinds of experiences.

So, the interests, intentions, and direction of learners come to the fore here and their agency will shape the potential to realize rich outcomes through these experiences. Moreover, the self and subjectivities as driving their learning in setting and securing rich learning outcomes will not be limited to workplace-learning experiences of this kind. A key basis for ongoing learning throughout working life will likely be directed by what motivates, directs, and focuses individuals' efforts at learning through work. For instance and in particular, it seems for older workers, semiskilled, those of color or with a disability, personal agency will be central in managing their learning through work, as the evidence suggests these workers are those least likely to be afforded workplace support for their learning.

To this end, there needs to be a greater decentering of the focus of learning through work from the physical and social attributes of the setting to accommodate the contributions of learners to workplaces as learning environments. Helpful accounts of personal epistemologies and learner agency have been generated in recent work (Hodkinson and Hodkinson, 2004). Many of these emphasize the relations and negotiations between individual workers' and their workplaces' (i.e., the personal and the situated) contributions to learning workplace practices, techniques, norms, etc. Such accounts illuminate how individuals likely elect to respond to governmental imperatives and employer requests to learn continually throughout working life. For instance, they advise governments and employers that their ambitions for lifelong workplace learning are unlikely to be realized unless more consideration is given to the needs, interests, and personal trajectories of those who they are exhorting to learn. Consequently, frameworks for workplace learning need to include considerations of the salience of personal epistemologies and the need to develop these for learning throughout working life, and in ways consistent with accommodating individuals' occupational trajectories. Yet, these epistemologies comprise more than personal strategies; they are central to and shaped by individuals' subjectivities and sense of self. This means that engaging the learners' interest is more than a nicety; it becomes an important imperative for efforts aligned with lifelong learning.

Pedagogic Qualities of Workplaces

Earlier socially orientated accounts of the contributions to learning of physical and social settings, such as

communities of practice (Lave and Wenger, 1991) and activity systems (Engestrom, 1993), advanced the pedagogic qualities of workplaces in terms of shared premises for participation and learning, and social system that shape participation and learning. However, recent accounts of the social contributions to learning through work have provided more elaborated accounts of the pedagogic qualities and potential of workplace settings. Some of these detail the pedagogic qualities of particular kinds of work and workplaces, while others provide more detailed considerations of the artifacts and practices in workplaces that contribute to learning, including the kinds and qualities of interactions, that lead authors to refer to their epistemic qualities. These are helpful as they lend finesse and refine earlier accounts. This then permits more comprehensive and detailed accounts of how the contributions of workplace settings can work to shape learning and how these might best be exploited to achieve particular kinds of learning outcomes.

Moreover, as considerations of workplace as learning environments have matured, the range of procedural responses has also grown. What is distinct about these provisions is that they center on the provision of workplace as learning experiences in their own right and not to augment or extend learning from other sources (i.e., educational programs). These include learning from work errors – how both qualities of workplace environments and workers' personal dispositions shape the prospect of learning purposively through workplace errors (Bauer and Mulder, 2007), learning projects (Poell, 2006), critical reflections through work experiences (van Woerkom, 2003), guided learning in the workplace (Billett, 2001), expanding learning opportunities (Fuller and Unwin, 2004), and the development of a workplace curriculum (Billett, 2006a). In different ways, these kinds of approaches open up considerations for the means by which workplace pedagogies can be developed, which stand as a key premise for framing workplace learning.

Yet, there are important conceptual and procedural goals still to be secured. For instance, much conceptual developments is premised on learners being in socially rich circumstances that provide models, guidance, etc. However, many workers are physically and socially isolated and therefore do not engage in or learn through these kinds of socially rich environments. Therefore, much needs to be known about learning in these relatively socially weak environments. Contemporary and emerging occupational and workplace practice is such that workers may be only afforded access to particular kinds of workplace experiences, those related to their occupation alone, and possibly not those that are sufficient for the comprehensive development of the workplace. So, understanding further how these kinds of knowledge can be developed stand as important goals. Then, there is the need to learn symbolic forms of knowledge and those that are not

opaque (Bresnahan *et al.*, 2002; Zuboff, 1988) or easy to learn, thereby requiring intentional pedagogic practices that can make this knowledge accessible and comprehensible. Moreover, a key concern about learning through practice is that uncritically accepting practices of the past and present may not assist the learning requirements for the future. More needs to be done to assist improving the equity of the affordances of opportunity, considering how workplace-learning opportunities are asymmetrically distributed on the basis of age, gender, language, and educational-achievement levels. Hence, there has to be a critical and questioning element to a framework that advances learning through work. Therefore, frameworks supporting and extending learning through work, need to constantly and critically examine the kinds and processes of learning that arise through participation in work and how these forms of learning can be aligned with the array of purposes to which they are directed (e.g., rich occupational knowledge).

Workplace-Learning Frameworks

At a macro level, frameworks for organizing adult learning in workplaces or pedagogies for the workplace have commonalities with those for other settings, albeit manifested in conceptions of epistemology, curriculum, and pedagogy in ways that reflect the particular kinds of social practices that comprise workplaces. These are premised on a set of practices that support learning that predates those of schooling. Overall, the commonalities include considerations of what is provided by the setting to support learning (i.e., its affordances) and how individuals elect to engage with, participate, and mediate in what is afforded them by the setting. These framing considerations are equally applicable for adults' learning across a range of social institutions (e.g., universities, colleges, communities, and workplaces). In curriculum terms, each of these has purposes (i.e., intentions for learning), enactments (e.g., means of supporting that learning), individuals who can support learning (e.g., teachers and coworkers), and learners (e.g., students and workers) that mediate how and what they learn through what they experience in these settings. Yet, there are some distinctions in the particular kinds of experiences, workplaces, and educational institutions afford, ways in which learners identify themselves and elect to participate in learning and for what purposes. The premises of workplace curriculum and pedagogy will be founded on practices being enacted in workplaces (the provisions of goods or services) as directed toward learning occupational knowledge as per the workplace's particular requirements. This may be distinct from the stated aim of educational institutions – intentional learning for an occupation. The identity and bases for participation by

workers, some of who will be guiding the work and learning of others (e.g., more experienced workers), is likely to be distinct from those who see themselves as teachers and students in educational institutions, and feature a strongly agentic role for learners. Yet, these are largely minor distinctions, thereby holding that learning through work can be informed, legitimated, and understood and appraised through orthodox curriculum concepts and pedagogic practices, and that there are also distinct features and characteristics that can be accounted for in these frameworks.

Therefore, in all, a framework for understanding learning through work can be founded on the dualities of what the workplace affords to those employed within them and how those individuals elect to engage with what is afforded them. Learning through practice, through errors, and by processes of observation and imitation, practice, and the direct guidance of experienced coworkers all stand as key elements of that framework. Besides, the kinds of work practices, and personal values and bases of identity, shape these dualities and the relations between them. In addition, the purposes of learning and the desired outcomes of that learning also stand as important foundations for making judgments about the value of that learning and how it might best be enhanced. These enhancements may well be realized through work activities or require particular and targeted intervention as perhaps increasingly, components of work-related knowledge become inaccessible and need to be made explicit.

Bibliography

- Bauer, J. and Mulder, R. H. (2007). Modelling learning from errors in daily work. *Learning in Health and Social Care* **6**, 121–133.
- Bennett, C. A. (1938). The ancestry of vocational education. In Lee, E. A. (ed.) *Objectives and Problems of Vocational Education*, 2nd edn., pp 1–19. New York: McGraw-Hill Book.
- Billett, S. (2001). *Learning in the Workplace: Strategies for Effective Practice*. Sydney: Allen and Unwin.
- Billett, S. (2006a). Constituting the workplace curriculum. *Journal of Curriculum Studies* **38**(1), 31–48.
- Billett, S. (2006b). Relational interdependence between social and individual agency in work and working life. *Mind, Culture and Activity* **13**(1), 53–69.
- Boud, D., Solomon, N., and Symes, C. (2001). New practices for new times. In Boud, D. and Solomon, N. (eds.) *Work-Based Learning: A New Higher Education?* pp 3–17. Buckingham: Open University Press.
- Bresnahan, T. F., Brynjolsson, E., and Hitt, L. (2002). Information technology, workplace organisation and the demand for labor: Firm-level evidence. *Quarterly Journal of Economics* **117**(1), 339–376.
- Brown, J. S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* **18**(1), 32–34.
- Colin, K. (2004). Workplace's learning and life. *International Journal of Lifelong Learning* **4**(1), 24–38.
- Collins, A., Brown, J. S., and Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics. In Resnick, L. B. (ed.) *Knowledge, Learning and Instruction, Essays in Honour of Robert Glaser*, pp 453–494. Hillsdale, NJ: Erlbaum.
- Deissinger, T. (2002). Apprenticeship systems in England and Germany: Decline and survival. *Paper Presented at the Towards a History of Vocational Education and Training (VET) in Europe in a Comparative Perspective*, Florence.
- Engestrom, Y. (1993). Development studies of work as a testbench of activity theory: The case of primary care medical practice. In Chaiklin, S. and Lave, J. (eds.) *Understanding Practice: Perspectives on Activity and Context*, pp 64–103. Cambridge, UK: Cambridge University Press.
- Fenwick, T. (2004). Learning in portfolio work: Anchored innovation in and mobile identity. *Studies in Continuing Education* **26**(2), 229–246.
- Fuller, A. and Unwin, A. (2003). Fostering workplace learning: Looking through the lens of apprenticeships. *European Educational Research Journal* **2**(1), 41–55.
- Fuller, A. and Unwin, L. (2004). Expansive learning environments: Integrating organisational and personal development. In Rainbird, H., Fuller, A., and Munroe, A. (eds.) *Workplace Learning in Context*, pp 126–144. London: Routledge.
- Gimpel, J. (1961). *The Cathedral Builders*. New York: Grove Press.
- Hodkinson, P. H. and Hodkinson, H. (2004). The significance of individuals' dispositions in the workplace learning: A case study of two teachers. *Journal of Education and Work* **17**(2), 167–182.
- Lave, J. (1990). The culture of acquisition and the practice of understanding. In Stigler, J. W., Shweder, R. A., and Herdt, G. (eds.) *Cultural Psychology*, pp 259–286. Cambridge, UK: Cambridge University Press.
- Lave, J. and Wenger, E. (1991). *Situated Learning – Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press.
- Lodge, R. C. (1947). *Plato's Theory of Education*. London: Kegan Paul, Trench, Trubner.
- Marsick, V. J. and Watkins, K. (1990). *Informal and Incidental Learning in the Workplace*. London: Routledge.
- Nerland, M. and Jensen, K. (2006). Insourcing the management of knowledge and occupational control: An analysis of computer engineers in Norway. *International Journal of Lifelong Education* **26**(3), 263–278.
- Newman, D., Griffin, P., and Cole, M. (1984). Social constraints in laboratory and classroom task. In Rogoff, B. and Lave, J. (eds.) *Everyday Cognition: Its Development in Social Context*, pp 172–193. Cambridge, MA: Harvard University Press.
- Newman, D., Griffin, P., and Cole, M. (1989). *The Construction Zone: Working for Cognitive Change in Schools*. Cambridge, UK: Cambridge University Press.
- Pelissier, C. (1991). The anthropology of teaching and learning. *Annual Review of Anthropology* **20**, 75–95.
- Poell, R. (2006). Organising learning projects whilst improving work: Strategies for employees, managers and HRD specialists. In Streumer, J. N. (ed.) *Work-Related Learning*, pp 71–94. Dordrecht: Springer.
- Raizen, S. A. (1991). *Learning and Work: The Research Base. Vocational Education and Training for Youth: Towards Coherent Policy and Practice*. Paris: OECD.
- Resnick, L. (1987). Learning in school and out. *Educational Researcher* **16**(9), 13–20.
- Rogoff, B. (1990). *Apprenticeship in Thinking – Cognitive Development in Social Context*. New York: Oxford University Press.
- Rogoff, B. and Gauvain, M. (1984). The cognitive consequences of specific experiences – weaving versus schooling among the Navajo. *Journal of Cross-Cultural Psychology* **15**(4), 453–475.
- Rogoff, B. and Lave, J. (eds.) (1984). *Everyday Cognition: Its Development in Social Context*. Cambridge, MA: Harvard University Press.
- Somerville, M. and Abrahamsson, L. (2003). Trainers and learners constructing a community of practice: Masculine work cultures and learning safety in the mining industry. *Studies in the Education of Adults* **35**(1), 19–34.
- van Woerkom, M. (2003). *Critical Reflection at Work*. Enschede: Twente University.
- Zuboff, S. (1988). *In the Age of the Smart Machine: The Future of Work and Power*. New York: Basic Books.

ADULT EDUCATION – AS A FIELD OF STUDY

Contents

Class Analysis in Adult Education

Gender Analysis

Race and Ethnicity in the Field of Adult Education

Trends in Workplace Learning Research

Class Analysis in Adult Education

T Nesbit, Simon Fraser University, Vancouver, BC, Canada

© 2010 Elsevier Ltd. All rights reserved.

The concept of class is indispensable but elusive. Notions of class are deeply embedded within broader beliefs about society and history and preserved by the need to explain continuing inequalities in opportunities and standards of living or the social divisions that they foster. However, it remains an ambiguous concept, challenged by the lack of a generally accepted definition, new forms of social stratification, and identity and other subjective politics. Also, the political and cultural changes of recent times – accelerating globalization, the rise of neoliberalism and free-market capitalism, a shift from manufacturing to service economies, the decline of an industrial working class, increased militarization, greater concern for the environment, the promotion of individualism and consumerism, and widening disparities between rich and poor – have witnessed diminishing public interest in the concept of class. Some commentators have even gone so far as to proclaim its demise or at least that it has outlived its relevance. Yet, class stubbornly fails to disappear and remains an important analytic concept for examining the structural causes and manifestations of inequality and power.

Class, however, means different things to different people: a theoretical device for analyzing the social world; shared social conditions; or a set of particular orientations, beliefs, and life practices. Popular understandings still describe class in terms of occupation, income, wealth, education, mobility, lifestyle, taste or ownership and the power that accrues from them. Others see class less as a possession than a dynamic – a relationship made between different people and groups divided along axes of power and privilege. As class differences play out in power relations, adult education can play a critical role in forming and mediating these relations.

Generally concerned with identity and personal and social change, adult education seeks to provide the knowledge, skills, and attitudes for people to engage more fully in and shape their individual and social worlds. Yet, such

worlds are not continually created anew but shaped by the past. So, adult education provides opportunities for personal mobility but within systems already marked by social inequality. Whatever its particular focus, approach, or clientele, adult education is essentially a sociopolitical endeavor with struggles for power – who has it, how they use it, and in whose interests – at its heart (Cervero *et al.*, 2001).

Adopting a class analysis of adult education does three things. First, it draws clear links between adult learners, educational institutions and processes, the social worlds of work, family, and community, and the economic systems that underpin them. Second, it makes visible the struggles of those who are variously poor, excluded, or dispossessed and the educational resources they can draw upon. Third, it highlights how educational institutions and processes function to inculcate and maintain dominant ideologies and values. As adult education is a moral and political endeavor as much as it is a technical practice, it is affected by its role in maintaining or challenging the social order. Do educational policies and practices reproduce existing relations of dominance and oppression? Alternatively, do they contribute to lasting social as well as personal change?

In general, adult education is intended to ameliorate the personal and social disadvantages created by circumstance and background. So, social divisions and the tensions they bring about are commonly seen as critical issues for adult educators. In fact, the educational approaches, structures, and activities that perpetuate the silence and invisibility of marginalized and disenfranchised groups are a regular feature of the adult education literature. This literature is permeated with regular exhortations to consider class, race, and gender as prime markers of social division. Yet, in comparison with its counterparts, the study of class has tended to be underexplored by adult education researchers. To examine why this might be so, several of the relationships between class and adult

education are outlined. First, some differing perspectives on class are provided before exploring how they inform and illuminate various areas of adult education practice.

Perspectives on Class

In capitalist societies, all aspects of peoples' lives and social relations are subjected to market requirements which are then normalized and made to seem natural. People's prestige and status are related to their productive ability; society values people by how much they earn or own. Basic aspects such as where we live, how we earn a living, who our friends are, and what access we have to healthcare and education are all dependent on our ability to produce wealth and other resources. Of course, these attributes are not fixed permanently; because the distribution of resources is unequal, people strive to maintain or enhance their own share. Thus, people's struggle for access to and control of resources is dynamic. Capitalist societies are stratified into classes, hierarchies of power and privilege related to the ownership and control of various forms of capital. Capitalist systems of structured inequality continue because society portrays them as normal or inevitable: the system encourages its victims to blame themselves for their failure to be successful. In this way, dominant groups are able to maintain the status quo and the hegemony of their own ideas without facing too strong a challenge from those less powerful.

Although ideas about class have been in use since Roman times, its current articulation stems from the Industrial Revolution. By the middle of nineteenth century, Marx used class as the foundational concept for explaining social organization in terms of understanding the ownership, means, and control of work processes. He claimed that societies consisted of two main classes: the bourgeoisie (who owned and controlled the mills, mines, and factories) and the proletariat (workers with only their labor power to sell). For Marx, the relationship between these two classes is essentially unequal and exploitative. The working class generates surplus wealth but does not profit from it as much as they might because the bourgeoisie disproportionately appropriates and accumulates it. He regarded all social life as marked by the struggles and conflicts over the generation and distribution of wealth and the status attached to it. Societal transformation was only possible when workers developed class consciousness about the sources of exploitation: a sense of a shared predicament, awareness of the capitalist class as their common enemy, and a realization of their common strength and destiny (Marx and Engels, 1845/1970).

Not everyone regards class in such materialistic terms. Weber, for instance, argued that class is better defined by also including notions of culture, politics, and lifestyle. People who fall within the same economic class may

nevertheless occupy different social class positions and have differing opportunities for work, income, developing skills, obtaining education, and owning property. For Weber, one's class is based more on these life chances, cultural background, status, and life outside of work than it is on one's relationship to the ownership and control of the means of production. Rather than see society as a two-class system, Weber posited a system of social stratification of many different classes that sometimes overlap. This less-deterministic approach can also be seen in the work of Bourdieu for whom a class is any grouping of individuals sharing similar conditions of existence and tendencies or dispositions. Equally important as one's location in an economic order is the possession of various forms of capital – economic, cultural, social, or symbolic – which can constellate differently in different societies. Bourdieu's concept of class thus takes into account other stratifying factors, such as gender, race, ethnicity, place of residence, and age. Finally, these class structures are not predetermined or imposed from without but subtly reproduced by people acting within preexisting contexts. Although both Weber and Bourdieu allowed more scope for human agency than did Marx, they still regarded external class structures as fundamental and quite constant. In other words, class relationships transcend the individuals who occupy the positions: people may move around (or stay put) but they still divide into exploiters and exploited.

These two broad views have shaped current understandings of class. Throughout the social upheavals of industrialization, definitions of class in Europe continued to be affected by older ideas of rank. The lower orders, laboring classes, and the middling ranks of society (such as merchants or teachers) existed alongside the aristocracy and the gentry. However, as the stratification of industrial society became more rigid, these definitions settled into the familiar classification of working, middle, and upper class. This depiction treats class as essentially static. Although it underlines the essentially economic nature of class, such a definition ignores the dynamic and shifting nature of the relationships between those who possess wealth and power and those who do not. More recently, class has come to be regarded as a relation that is constantly changing. As one major British historian puts it, "class is not a category. . . but rather an historical relationship between one group of people and another. . . . It is defined by men [*sic*] as they live their own history" (Thompson, 1971: 9–10).

Non-European countries – many still affected by the legacy of colonization – regard themselves as relatively free from the archaic categories of class distinction. For example, in the United States, one commonly hears either that class has ceased to exist or, alternatively, that everyone is middle class. Instead, the ethics of self-reliance and mobility and the ideologies of individualism,

egalitarianism, and meritocratic achievement have been more powerful forces than class solidarity. Nowadays, existential rather than social factors tend to influence who people think they are. For example, it is far more common for people to define themselves as black, gay, Jewish, Latino, lesbian, or age- or mobility-challenged than to refer to themselves in terms of class. In some countries, the historical legacy of slavery and its lingering consequences have shaped a more complicated system of stratification – one kept in place by ideological as well as economic factors (Levine, 1998). Older depictions of class are now regarded as too simplistic for heterogeneous countries such as the USA. This has been further confounded by the rise of movements specifically identified with racial, women's and gay equality, environmental concerns, and various manifestations of religious, national, and ethnic rights. Nowadays, constructs such as class, race, and gender are commonly seen as interrelated and overlapping (Collins, 1993; Wood, 2002); social class is recognized as both gendered and racialized and viewed as but one part of a wider and interlocking system of oppression and domination (Rothberg, 1998).

With these different and competing notions of, and perspectives on, class, discussing it can be difficult. Wright (1979) identifies four major approaches to understanding class: a functional differentiation of positions within a society, groups unified by their common position in a hierarchy of power or authority, groups with different market capacities that result in different life chances, and a shared location in the social organization of production. However, whatever one's orientation, an attention to class and class analysis reveals several general principles. First, a class analysis focuses on materialist concepts regarding the production and reproduction of social life and the importance of human activity in shaping both material subsistence and consciousness. Second, a class analysis highlights the fundamental and dynamic relationships between economic and social structures, the ideologies that frame our world, and the ways we experience, understand, and shape the world. Third, a class analysis suggests that we cannot explain social phenomena by their surface manifestations nor by the ways that individuals experience them, but as, instead, representations of external divisions of power. Fourth, a class analysis provides a basis for explaining why people organize themselves into collective forces to resist injustice and exploitation. Finally, for those with a commitment to social justice, a focus on class also raises several important questions: How do we negotiate or internalize dominant ideologies and relations of ruling? How might alternative ones develop? How can marginalized people, silenced by social, economic, and cultural relations of power, recover their voices and the right to be heard? These questions are often central to the practice of adult education, and I now turn to exploring the relationships between it and class.

Class and Adult Education

Various forms of adult and continuing education are now firmly established as central to the smooth functioning of economic systems and societies. As concepts such as life-long learning and the knowledge or learning society have gained prominence, various forms of adult education and training have become key vehicles for preparing people to be adaptable to economic and cultural changes in society. Prevailing educational practices tend to inculcate dominant values rather than confront them and, because most educational institutions are generally a middle-class domain, their policies and practices are weighted strongly in favor of middle-class values. So capitalist societies, in which class operates as a primary structuring of social inequality, usually ignore or bury class perspectives. As such, many adult educators are uncertain about how their work reflects underlying political structures, let alone economic systems. Observing the effects of power and privilege is far easier than determining their causes.

A number of studies explore how education can reproduce existing patterns of power. Economists Bowles and Gintis (1976) demonstrated how educational systems are part of a system of broader capitalist class relations. Their correspondence theory explains how, in general, schools reproduce the social relations that capitalist production requires. As Bowles and Gintis describe, capital requires two things: workers of specific types and relative social stability and ideological acceptance of class relations. The capitalist class thus has a broadly shared set of interests pertaining to educational systems and the capacity to promote such interests.

Some find the correspondence theory too mechanistic or reductive; it allows little agency for those involved. One less-deterministic approach suggests that education serves the interests of the privileged by structuring learners' access to and uses of various forms of social and cultural capital (Bourdieu and Passeron, 1977). Others have introduced notions of struggle and resistance into this process. For example, Willis (1977) showed how several working-class teenage boys consciously resisted and rebelled against school and classroom authority. Tellingly, however, this resistance worked better within school than outside it: when the boys left school, they remained unable to find anything but unskilled and unstimulating jobs. McLaren (1995) and Apple (1996) also show how individuals can resist and contest social and cultural oppression in educational settings. They document the complex relationships between cultural reproduction and economic reproduction and explore how class interrelates with the dynamics of race and gender in education.

These studies indicate the essential role of education in promoting and maintaining the social relations required for capitalist production. Adult education is far

from immune from these trends. Although we now recognize that the relationships between educational practices and political structures are much more complex than correspondence theory suggests, adult educators who work in areas such as adult basic education, literacy, vocational and workplace education, and welfare-to-work, and other social programs will recognize how often their work, the policies about it, and the textbooks and curricula they use are still much more closely tied to employers' needs than to their adult or working-class students' interests (Boughton, 2006; D'Amico, 2004; Kincheloe, 1999; Rose, 1989).

Much of the published research on class in adult education explores the consequences or experiences of class and examines issues such as the participation, access, and attainment of different groups. In documenting how social class affects participation in adult education programs, such studies consistently underscore how far social class remains a key determinant of adult participation in organized learning. To give just one recent example, Sargent and Aldridge (2002) indicate that upper- or middle-class adults are twice as likely to engage in some sort of learning activity than those from the working class.

However, although such studies detail how class remains a major factor affecting adult education participation, most do not really explore either the constitutive dynamics of class or how it operates in practice. From a conceptual perspective, they add little to London's classic study (London *et al.*, 1963) which explored the important contribution that adult education makes to larger society, specifically for those deemed less educated and less skilled. London and his colleagues found a strong connection between social class and people's abilities to prosper in a rapidly changing world. Class not only affected participation in adult education activities but was also closely related to other facets of social life such as jobs, vocations, and leisure pursuits. Anticipating subsequent debates about lifelong learning, London's report called for adult education and training to "become a continuing part of everyone's life" (p. 148), providing "both education for work and education for leisure" (p. 153).

Perhaps more seriously, too few published studies further class awareness or address what Allman *et al.* (2007) describe as the chasm between class as social inequality and class as a constitutive element of the world of struggle. Instead, the published studies linking class and adult education concentrate on the results rather than the causes or workings of class. This is understandable. Individuals tend to internalize the conflicts within hierarchical systems, especially those individuals without much power. Also, people usually closely experience class at the same time as other, more recognizable forms of oppression. These factors, when combined with the scarcity of class scrutiny, ensure that people do not always have readily available concepts to identify – let alone

analyze – the class aspects of their experiences. So scholars prefer to focus on more obvious markers of social division and overlook class in their theoretical lexicon. Class is also difficult to discern: we can only examine it through its consequences or outcomes. And, as it is not easy to identify or operationalize on an individual level, it is much better suited to macro- rather than micro- analyses. Even when it is acknowledged, class still tends to be regarded as an individual characteristic or an entity rather than a constituent social relationship. So, the tendency of many adult education researchers to assume a strong individual orientation tends to further divert attention away from class perspectives.

Yet, sophisticated and theoretical explorations of class and adult education do exist – generally emanating from countries with a more prominent appreciation of class relationships. For example, Allman (2001), Livingstone (1999), Stromquist (1997), Thompson (2000), Walters (1997), Welton (1995), and Youngman (2000) all provide rich empirical and theoretical examinations of how adult education practices are linked to social class and the increasingly globalized nature of capitalism. Appreciating that class is understood and experienced differently around the world, Nesbit (2005) gathers several adult educators from different countries to explore how class affects adult education and to examine how the complex relationships between class, gender, and race play out in its policies, practices, and discourses. Discussions of the relationships between adult education and social class are also more noticeable on the margins of the established adult education literature. To give a flavor of such work, I now explore how class has been used to analyze several different arenas of adult education practice: approaches to learning, its role in social movements, and the related fields of higher education and working-class studies.

One distinguishing feature of adult education is its close attention to learning. In their studies of working people's intelligence and learning, both Rose (2004) and Sawchuk (2003) provide two of the best recent analyses of the interrelationships of class and adult learning. Both identify a distinctive working-class learning style that operates independently of formal training and centers around informal workplace and community networks. This learning style is collective, mutual, and solidaristic: people exchange knowledge and skills, hardware and software, and draw upon each other's different expertise to develop group resources for anyone to use. So they develop an expanding learning network – a powerful working-class resource that stands opposed to the trajectory of dominant forms of workplace and institutionalized education that individualize and commodify learning. Although most research on adult education and class focuses on how working-class people are excluded and help to exclude themselves from formal education,

Rose and Sawchuk show that working-class adults can bring rich cultural resources to their learning.

Adult education deliberately seeks to link the personal with the social and much of its historical development and traditions are linked with movements for social justice and equity. A class perspective on adult education's role in social movements underscores that lasting social change comes about through people acting together to challenge inequality and oppression. For example, Altenbaugh (1990), Schied (1993), and Walters (2005) provide compelling accounts of the ways in which class affects the educational activities of those who are active in labor and workers' movements or who struggle for social justice in South Africa. As Walters shows, learning about and discussing earlier battles for social justice can provide resources and hope to those involved in current struggles. Other discussions of the role of adult learning and education in social movements are provided by Holst (2002) and Foley (1999). Basing their work on the ideas of Freire, Gramsci, and Marx, they each explore the idea that radical adult educators can help build civil society through social movements. Two aspects of their work are particularly important to a class-based appreciation of adult education. First, "learning in such situations is tacit, embedded in action and often not recognized as learning" (Foley, 1999: 3). Second, the notion of learning in and through struggle: "people's every day experience reproduces ways of thinking and acting support which support the, often oppressive, status quo, but [can also] enable people to critique and challenge the existing order" (ibid, pp. 3–4).

The related field of higher education also provides analyses of how universities treat adult learners from different class backgrounds and how such students experience university-level education (Ball, 2003; Tokarczyk, 2004). Other studies explore how curricula and pedagogy in different disciplines reflect class-based interests (Margolis, 2001), how the pedagogical status quo might be challenged and classroom practices democratized (Shor, 1996), and how class shapes the identity and teaching of educators (Malcolm, 2005). The emerging field of working-class studies has also reinvigorated discussion of the intersections of class, adult learning, and institutional practices (Roberts, 2007; Zandy, 2001). Acknowledging that working-class learners now constitute a significant proportion of students enrolled in institutions of higher education, this body of work incorporates a sensitivity to students' working-class roots while suggesting curricular and pedagogic innovations informed by an awareness of class culture.

Conclusion

Class analyses of adult education highlight the unexamined patterns of behavior through which society produces and reproduces social classes in the dynamics between

educational activities and the wider cultural politics of societies. They show how capitalism is not just an economic system but rather a totalizing system of social relations. They also expose the superficiality of a variety of currently prescribed educational reforms: instrumental outcomes, the individualizing of educational opportunities, increased competition, accreditation, and commercial involvement, privatization of post-secondary institutions, a return to so-called basics, the streaming of learners into cultural or functional literacies or core competencies, the increasing pressures to work harder and longer, and the tendency to disregard or downplay social and collective notions of identity and action.

Despite overwhelming evidence to the contrary, too many still claim that the idea of class has outlived its usefulness and that focusing on it and other forms of oppression is misguided and unnecessary. Instead, educational approaches seem generally more committed to the management of inequality rather than its elimination and often encourage people to accept rather than challenge their place in society. Yet, adult educators need not be bystanders to political contexts but vital and essential members of them. A class analysis of adult education raises important and challenging questions that build upon learners' lived experiences about how inequalities play out in communities, lives, and workplaces. It also confronts any approach to confine adult education to the production and maintenance of human capital or advances the political and economic interests of the powerful at the expense of the poor. Finally, it helps adult educators and their learners resist what has been called the postmodern condition of "skepticism, uncertainty, fragmentation, nihilism, and incoherence" (Allman, 2001: 209).

Bibliography

- Allman, P. (2001). *Critical Education against Global Capitalism*. New York: Bergin and Garvey.
- Allman, P., McLaren, P. L., and Rikowski, G. (2007). After the box people: The labour-capital relation as class constitution and its consequences for Marxist educational theory and human resistance. In Scott, A. and Freeman-Mohr, J. (eds.) *The Lost Dream of Equality: Critical Essays on Education and Social Class*, pp 125–150. Rotterdam: Sense Publishers.
- Altenbaugh, R. (1990). *Education for Struggle: The American Labor Colleges of the 1920s and 1930s*. Philadelphia, PA: Temple University Press.
- Apple, M. W. (1996). *Cultural Politics and Education*. New York: Teachers College Press.
- Ball, S. (2003). *Class Strategies and the Education Market*. London: Routledge Falmer.
- Boughton, B. (2006). Researching workplace learning and class. *Economic and Labour Relations Review* 17(2), 157–164.
- Bourdieu, P. and Passeron, J. C. (1977). *Reproduction in Education, Society, and Culture*. Thousand Oaks, CA: Sage.
- Bowles, S. and Gintis, H. (1976). *Schooling in Capitalist America*. New York: Basic Books.
- Cervero, R. M. and Wilson, A. L., and associates (2001). *Power in Practice*. San Francisco, CA: Jossey-Bass.

- Collins, P. H. (1993). Toward a new vision: Race, class and gender as categories of analysis and connection. *Race, Sex, and Class* **1**(1), 25–45.
- D'Amico, D. (2004). Race, class, gender, and sexual orientation in adult literacy: Power, pedagogy, and programs. In Comings, J., Garner, B., and Smith, C. (eds.) *Review of Adult Learning and Literacy: Connecting Research, Policy, and Practice*, pp 17–69. Hillsdale, NJ: Erlbaum.
- Foley, G. (1999). *Learning in Social Action*. London: Zed Books.
- Holst, J. D. (2002). *Social Movements, Civil Society, and Radical Adult Education*. New York: Bergin and Garvey.
- Kincheloe, J. L. (1999). *How Do We Tell the Workers? The Socioeconomic Foundations of Work and Vocational Education*. Boulder, CO: Westview.
- Levine, R. F. (ed.) (1998). *Social Class and Stratification*. Oxford: Rowman and Littlefield.
- Livingstone, D. W. (1999). *The Education–Jobs Gap*. Boulder, CO: Westview.
- London, J., Wenkert, R., and Hagstrom, W. O. (1963). *Adult Education and Social Class*. Berkeley, CA: University of California, Berkeley Survey Research Center.
- Malcolm, J. (2005). Class in the classroom. In Nesbit, T. (ed.) *Class Concerns: Adult Education and Social Class*, pp 45–52. San Francisco, CA: Jossey-Bass.
- Margolis, E. (ed.) (2001). *The Hidden Curriculum in Higher Education*. New York: Routledge.
- Marx, K. and Engels, F. (1845/1970). *The German Ideology*. Arthur, C. J. (trans. and ed.). London: Lawrence and Wishart.
- McLaren, P. L. (1995). *Critical Pedagogy and Predatory Culture*. New York: Routledge.
- Nesbit, T. (ed.) (2005). *Class Concerns: Adult Education and Social Class*. San Francisco, CA: Jossey-Bass.
- Roberts, I. (2007). Working-class studies: Ongoing and new directions. *Sociology Compass* **1**, 191–207.
- Rose, M. (1989). *Lives on the Boundary: The Struggles and Achievements of America's Underprepared*. New York: Free Press.
- Rose, M. (2004). *The Mind at Work*. New York: Viking Books.
- Rothberg, P. S. (ed.) (1998). *Race, Class and Gender in the United States: An Integrated Study*. New York: St. Martin's Press.
- Sargent, N. and Aldridge, F. (2002). *Adult Learning and Social Division*. Leicester: National Institute of Adult Continuing Education.
- Sawchuk, P. H. (2003). *Adult Learning and Technology in Working-Class Life*. New York: Cambridge University Press.
- Schied, F. M. (1993). *Learning in Social Context*. DeKalb, IL: LEPS Press.
- Shor, I. (1996). *When Students Have Power*. Chicago, IL: University of Chicago Press.
- Stromquist, N. P. (1997). *Literacy for Citizenship*. Albany, NY: SUNY Press.
- Thompson, E. P. (1971). *The Making of the English Working Class*. Oxford: Oxford University Press.
- Thompson, J. (2000). *Women, Class and Education*. New York: Routledge.
- Tokarczyk, M. M. (2004). Promises to keep: Working-class students and higher education. In Zweig, M. (ed.) *What's Class Got to Do With It?*, pp 161–167. Ithaca, NY: Cornell University Press.
- Walters, S. (ed.) (1997). *Globalization, Adult Education and Training*. London: Zed Books.
- Walters, S. (2005). Social movements, class, and adult education. In Nesbit, T. (ed.) *Class Concerns: Adult Education and Social Class*, pp 53–62. San Francisco: Jossey-Bass.
- Welton, M. R. (ed.) (1995). *In Defense of the Lifeworld: Critical Perspectives on Adult Learning*. Albany, NY: SUNY Press.
- Willis, P. (1977). *Learning to Labour*. Farnborough: Saxon House.
- Wood, E. M. (2002). Class, race and capitalism. *Political Power and Social Theory* **15**, 275–284.
- Wright, E. O. (1979). *Class Structure and Income Determination*. San Diego, CA: Academic Press.
- Youngman, F. (2000). *The Political Economy of Adult Education*. London: Zed Books.
- Zandy, J. (ed.) (2001). *What We Hold in Common: An Introduction to Working-Class Studies*. New York: Feminist Press at CUNY.

Further Reading

- Althusser, L. (1971). Ideology and ideological state apparatuses. In Brewster, B. (ed.) *Lenin and Philosophy*, pp 170–186. London: New Left Books.
- Gramsci, A. (1971). *Selections from the Prison Notebooks*. Hoare, Q. and Smith, G. N. (trans. and eds.). New York: International Publisher.
- Hooks, B. (2000). *Where We Stand: Class Matters*. New York: Routledge.
- Linkon, S. L. (ed.) (1999). *Teaching Working Class*. Amherst, MA: University of Massachusetts Press.
- Reay, D. (2001). Finding or losing yourself? Working-class relationships to education. *Journal of Education Policy* **16**(4), 333–346.
- Seabrook, J. (2002). *The No-Nonsense Guide to Class, Caste and Hierarchies*. Oxford: New Internationalist Publications.
- Sennett, R. and Cobb, J. (1972). *The Hidden Injuries of Class*. New York: Vintage.
- Tisdell, E. (1993). Interlocking systems of power, privilege, and oppression in adult higher education classrooms. *Adult Education Quarterly* **43**(4), 203–226.
- Weber, M. (1920/1968). *Economy and Society: An Outline of Interpretive Sociology*. New York: Bedminster Press.
- Wright, E. O. (ed.) (2005). *Approaches to Class Analysis*. New York: Cambridge University Press.
- Youngman, F. (1986). *Adult Education and Socialist Pedagogy*. London: Croom Helm.

Gender Analysis

J Stalker, University of Waikato, Hamilton, New Zealand

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Gender analysis is a useful tool which adult educators can use in their struggles for social justice. It is a complex activity which basically represents an optimistic attempt to end disparities and disadvantages among different genders. At its most expansive, it aims to create a paradigm shift to a world of an enduring, healthy environment and global peace.

Some might argue that the field of adult education has a strong track record of undertaking gender analysis. For decades theoreticians, researchers, and practitioners have explored women's participation and nonparticipation in various forms of adult education, the preferences and learning styles of women adult learners, the shifting roles and responsibilities of women, the different possibilities and consequences of development for women, formations of women's identities, and so on. Often, these were examined in relation to men's responses in the same contexts. However, although for many decades adult educators have theorized, researched, and practiced adult education in relation to both women and men, that is not the same as conducting the complex task of a gender analysis.

Given the above, the purpose of this article is to explore the complexities and possibilities of gender analysis for adult educators. First, this article presents the concept in terms of its interwoven theoretical positionings. Second, the article explores it as a research process of data collection, analysis, and strategic initiatives. Third, it briefly explores some benefits which accrue to gender analysis. Fourth, it examines the challenges which face adult educators who undertake gender analysis in the future. Finally, the article concludes with a summary and invitation to adult educators to make a commitment to gender equality.

Theoretical Positionings of Gender Analysis

In the industrialized world, our current understandings of gender analysis are shaped by its several theoretical positionings across the previous decades. Strictly speaking, gender analysis would encompass analyses of men, women, trans-genders, trans-sexuals, transvestites, homosexual lesbians, bisexuals, woman-born-woman, woman-born-man, man-born-woman, and inter-sexuals. Gender, after all,

is the culturally ascribed roles given to the sexual identities of women and men.

In reality, we ascribe the term gender analysis contemporarily and retrospectively to several kinds of analyses and these in turn guide our current approach to the task. Thus, below, the basic theoretical positionings, or strands, which shape adult educators' present understandings of the concept of gender analysis, are identified. Although it is tempting to present these as linear and developmental, these approaches are actually dynamic, unstable, interactive, and cyclical.

The first strand is driven by an oppositional approach and it has a long history. In this approach, analyses focus on women's situations in relation to men's situations. Such an approach rests on the evidence that women are missing from the data. As a result, its subsequent analysis and any resultant theory, research, or practice address only half the world's population. Solutions for this dilemma emphasize the notion of equality, that is, the same treatment of women and men in the hopes of creating equal outcomes.

This oppositional approach juxtaposes women with men, and positions them through their sexual identities as unique and distinctive. Those differences most often seem to be expressed dichotomously as women against men, feminism against patriarchy, powerless against the powerful.

This approach is not new, since for centuries the major religions of the world have analyzed and defined separate roles for women and men. In the wider context, as the industrialized world evolved, these definitions tended to be increasingly extrapolated and debated in terms of women's and men's relative and different economic roles as producers, reproducers, and consumers (Dimand *et al.*, 2004).

The second strand takes a diverse approach. Basically, it challenges the oppositional positioning of women and men, and values complexities and multiplicities. It argues that women, and men, do not belong to homogeneous groups. Rather, there are multi-faceted definitions of gender and many masculinities and femininities which theoreticians, researchers, and practitioners should acknowledge. The emphasis shifts to equitable treatment and outcomes within groups of women and of men, as well as between women and men, with a view to co-creating a new, more equitable, and sustainable world.

This strand has been strengthened by three forces. In the first instance, as the postmodern discourse in academe intensified, the oppositional positioning of women and

men came under close scrutiny. It became obvious that women and men were complex and not so tidily separated from each other into two groups. The rigid stereotypes and assumptions associated with each sexual identity were challenged. The clear, all embracing, distinctive divisions between the two groups were redefined as constructs with many overlaps and congruencies, influenced by context and culture. There was an awareness that women and men can make unique, yet also similar, complementary contributions. The emphasis changed to the need for both women and men to change patterns of the past.

In the second instance, women who did not see themselves in the oppositional discourse challenged it. Women from a wide variety of classes, colors, races, ethnicities, abilities, ages, and sexual orientations argued that the approach and analyses did not include the realities of their lived experiences. Rather, they suggested that the analyses were most often based in the lives of white, middle-class, heterosexual, able-bodied women. Ironically, the discourse they generated also has oppositional and diversity approaches.

Finally, it became clear that a dichotomous positioning of men against women could work to the disadvantage of women. Such a framework lent itself to straightforward comparisons between women and men. In patriarchal environments, this led too easily to a definition of men's values, abilities, and actions as the preferred norm and women's values, abilities, and actions as deficient relative to men.

In sum, the second strand has a stronger focus on gender and broader understandings of women and men. Notions of social construction, de-construction, identity formation, context, and culture merge to create a more complex, dynamic picture which is reflected in the term gender. The more static notions of reified, strongly categorized women and men are broadened to embrace wide diversity within those terms.

These two strands can be traced in the adult education literature in the publications of industrialized nations. As the selected exemplars below illustrate, the strands are simultaneously contemporary and historical. Publications in the first strand treat women as a unique phenomenon absent from our history (Hugo, 1990) and deserving of special study (Lewis, 1988; Stalker, 2005; Thompson, 1985). Often the latter publications make reference to the classic work by Belenky *et al.* (1986). Publications in the second strand reflect a view of 'woman' and 'man' as complex and diverse categories (e.g., Hayes and Colin, 1994; Graveline, 1994; Hill, 1995; Johnson-Bailey and Cervero, 1996; Luttrell, 1989; Tisdell, 1993; Walters and Manicom, 1996).

A third hybrid approach to gender analysis also exists. This is a complex strand which attempts to acknowledge the first strand's focus on women without losing the second strand's emphasis on complexity. It focuses on the

category of women yet also values the differences among and within the category. It argues that the stories, experiences, and knowledge of all women are needed in order to achieve their genuine empowerment and sometimes, it includes the role of men of multiple gender identities in the analysis and locates them as potential allies.

It is difficult to undertake gender analyses from within this strand. It requires a negotiation of space and power in new ways (Medel-Anonuevo, 1997). It is a conceptually very complex task to keep a useful balance between a dichotomous approach and a diverse approach. What often happens is that theoreticians, researchers, or practitioners give a brief acknowledgment to either approach and then proceed with a strong preference for one or the other. In other words, a truly hybrid, balanced approach to gender analysis is still in the process of evolution.

This is, in part, because the hybrid strand has struggled to survive in a sharpened global context in which those in developing, majority nations deal with cruel realities. Particularly in the late twentieth century, it seemed that their poverty flourished, war raged unabated, and environmental disasters increased in number. Even in more comfortable nations, the neo-liberal agenda thrived and the negative impacts of its capitalist focus on privatization, competition and globalization were keenly felt.

Against this backdrop, adult education activists around the world came and continue to come together at international conferences to demand increased participation and decision-making power for women. Although it is sometimes referred to as the Gender agenda, the essence of their arguments often is captured by the observation that "Women hold up half the sky." This exemplifies the hybrid positioning which hints at a more diverse approach yet has a strong, irrefutably oppositional stance. It illustrates that despite its critics, oppositional positioning continues to influence our theorization, research, and practice of gender analysis. This happens for several reasons.

First, the adult education agenda has become more globalized. It now includes the voices of many who come from contexts in which the situation of women is explicitly and irrefutably horrific. Adult educators in these contexts often argue that women's (and girls') continuing, persistent, and unrelenting oppression are the result of men and their patriarchal systems. These ensure that the effects of poverty, war, environmental disasters, and neo-liberalism are most severely experienced by women. They contend that disestablishing patriarchal systems is the real key to national progress, economic development, and self-reliance.

The voices and experiences of these educators seem to be particularly powerful and poignant. The media have helped to strengthen those voices through their increased exposure of appalling violences against women worldwide. In addition, adult educators in industrialized nations, perhaps valuing diversity, seem to have a will

to value the lived experiences of their majority nation colleagues and to accept their judgments about how to interpret and resolve gendered issues.

At the same time, the adult education global agenda has been shaped by the World Bank and International Monetary Fund. As the major global players for aid and development programs, they now incorporate women into many of their priorities. They and other international aid agencies see the inclusion of gender and in particular women as an efficient and effective way to increase national stability and improved, sustainable economic outputs. In other words, women are key to improving the lives of all citizens, men included. Significantly, education and adult education are identified as strategic devices to engage women and girls in these goals.

A second factor that encourages the oppositional approach to gender analysis concerns the criminalization and/or condemnation of sexual minorities. Numerous industrialized and majority nations reject the idea that multiple gender identities are possible. Many have explicit legislation that makes it illegal to either identify as, or practice as, any identity other than man or woman. Other nations are more subtle, but equally hostile to the idea. These forbidden identities can include trans-gender, trans-sexual, transvestite, homosexual lesbian, bisexual, woman-born-woman, woman-born-man, man-born-woman, and inter-sexual. Within these contexts, data collection is restricted to the categories of male and female. When international comparisons are planned, this then also impedes the research agenda of more liberal nations.

Finally, readily available historical, and also contemporary, databases are shaping the approach to gender analyses. Although these provide a logical foundation for further analyses and comparisons, they often used male–female categories. When these become the base for further research, the pattern of dichotomous positioning is entrenched and difficult to disrupt.

These databases have also kept alive the oppositional approach because, as some data have been collected for global projects, they have revealed boy's educational underachievement. Two arguments that revolve around these data further the oppositional discourse. Some argue that women's and girls' concerns have been largely solved and that boys and men must now become the priority. Others argue a preoccupation with establishing policies, practices, and funding for girls' and women's education has undermined boys' and men's achievement. Both arguments place women/girls and men/boys in competing opposition to each other.

Global publications reflected this persistent oppositional approach when women's nongovernment organizations and feminist activists organized internationally to put women, and increasingly girls, on the agenda and in their published documents (UNIFEM and UN/NGLS, 1995). The International Women's Decade (1976–85)

played an important role in this process as did policies generated at Women's World summits (e.g., 1975, Mexico City; 1985, Nairobi; and 1995, Beijing). Significant, clear resolutions and publications devoted to the advancement of women and girls were promulgated around the world (e.g., 1979, Convention for the Elimination of all Forms of Discriminatory Against Women, CEDAW) (REPEN, 1996).

Publications generated at summits not devoted entirely to women's issues also began to highlight the role of women (Scampini, 2001). Typical of their tone is a resolution included in the *1995 Copenhagen World Summit for Social Development* document. It stated specifically that social and economic developments were unlikely to be achieved in the long term if the full participation of women was not achieved.

Similarly, world summits focused specifically on education established global legislation, clear commitments, and specific goals related to women's and men's education. Key among these were *Education for All (EFA)* (launched in 1990), *CONFINTEA V* (1997 adult education global UNESCO conference), and the *Millennium Development Goals (MDG)* (signed by the United Nations in 2000). These summits committed, and continue to commit, signatory nations to many education-related goals and include the promotion of gender equality and empowerment of women and girls and more recently, men and boys. Significantly, the MDG's stated aim was to address these issues preferably by 2005, and in a worse case scenario by 2015. These tight and specific deadlines are propelling much of the current analysis.

In summary, this section has examined how gender analysis is shaped by its theoretical positionings. Clearly gender analysis is a complex concept – and made all the more muddy by the seemingly random use of the term gender. What we name concurrently and retrospectively as gender analysis can display oppositional, diverse, and/or hybrid approaches.

As the agenda of adult education becomes more globalized, sexual minorities continue to seek legal identities, and historical databases are used, the oppositional approach has re-emerged. However, simultaneously, there is a robust underlying discourse which acknowledges the complexity of gender and seeks equitable outcomes within groups of women and of men, as well as between women and men. In the end, the exposure of explicit differences between women and men, enhanced by this more complex view, creates the base for rich analyses. In the discussion which follows, the terms gender equality, women, and men will be used to represent that richer meaning.

A Research Process

It is against this vibrant, dynamic background that gender analysis is undertaken as a research process. Researchers

collect, measure, manipulate, and interpret data in systematic ways to provide evidence of the comparative situation between women and men. Gender analysis reveals, interprets, revises, monitors, and evaluates in a continuous, cyclical process which moves toward gender equality. The research process is the foundation of gender analysis and is composed of three interwoven elements: data collection, data analysis, and strategic initiatives.

Data Collection and Analysis

Data collection in gender analysis can be understood through an examination of the approach, focus, and technique. In terms of approach, data are collected systematically through both quantitative and qualitative approaches.

An increasingly popular and powerful quantitative approach used in gender analysis involves the use of indicators for monitoring. This is a kind of social watch which examines the extent to which nations have achieved the goals to which they committed at international conferences. Quantitative indicators related to the goals are identified so that "concrete aspects of reality can be observed and measured" (ICAE, 2003: 12). This allows the adult educator to collect rich data on a nation's real commitment to issues like: general policies on adult learning, adult literacy and basic education, promotion of active citizenship, awareness of discrimination, culture of peace and human rights, work-related adult learning, health education, involvement of civil society in environmental and development problems (ICAE, 2003). The data expose the increase, decrease, or lack of movement in these areas. They have the added advantage of being comparable across nations and this means that they can expose the relative commitments of nations.

A qualitative approach is particularly important at two levels. First, it is useful to supplement the limited representation of reality yielded by quantitative data. It provides a more multi-dimensional and complex account of the social facts and processes within which they develop (ICAE, 2003). Second, a qualitative approach is useful to uncover the subtle, murky yet powerful historical and cultural obstacles to equality. It can reveal these barriers through the study of peoples' views of gender disparities in a wide range of areas. It can reveal respondents' assumptions and allow for a more flexible probing of their beliefs to reveal their subjective interpretations of their lived experiences.

In both the qualitative and quantitative approaches, data may be collected from already-existing databases and may also be generated by new research. It can be a snapshot of a moment, frozen in time. To give more depth to the findings, those data can then be compared to previous, baseline data, or collected continuously over time to give a longitudinal picture of the situation.

The second important element in data collection is the focus. It is concerned with the domains of investigation and the way in which the data are framed. Data can be gathered from the microlevel of day-to-day household differential responsibilities and from more complex levels of policies, programs, service delivery, and political/economic/social/cultural elements of institutions and structures. It can be gathered on an endless number of variables such as employment, unemployment, violence, health, education, housing, sanitation, safe water, nutrition, property rights, literacy, social mobility, and so on.

Whatever the domain, a priority for gender analysis is that the approach is gender sensitive, and as we might expect from the theoretical discussion above, often it is framed from women's perspectives. However, gender analysis goes beyond the mere collection of data on women to men ratios. For example, in relation to participation, a key domain of gender analysis, one might begin by examining the women to men ratios in recruitment, involvement, and completion. Although those data are revealing in and of themselves, gender analysis would also include the study of the gendered nature of access to, and control and management of resources, production, reproduction, decision making, benefits, gendered roles, responsibilities, knowledge, skills, and activities. Gender analysis thus brings a greater depth of understanding to the issue of women's inequality because it emphasizes issues of women's relative power, control, and management.

Similarly, in relation to work, research traditionally focuses on paid work. Research based in gender analysis, however, acknowledges the reality of women's lives and reframes it as an issue of the gendered division of labor in both the public/paid and private/unpaid spheres. In the same way, traditional research often presented women as the victims and passive recipients of their situations. However, gender analysis research tends to present them as agents of resistance and change despite their oppressive contexts.

The technique of collection is the third important element in the data collection process. Adult educators undertaking gender analysis strive to ensure that the process is participatory and based in sound relationships. Ideally, the rights and needs of grassroots communities drive it and data are given back to those whose lives they effect so that they increase peoples' understandings of gender discrimination and foster their agency. These civil society relationships are strengthened by healthy partnerships with other key stakeholders. These are at the local, national, and international levels and include relationships with government departments, women's offices/leaders/organizations/networks, religious/educational/media organizations, think-tanks, corporations, businesses, and academics. This participatory, relationship-based approach to collection permeates gender analysis, including the process of data analysis.

Data analysis is not a summative activity which occurs only after data collection is completed. Rather, it is ongoing process which begins at the very first steps of the collection and measurement of data. Data analysis is a process of manipulating and interpreting the collected data so that it makes sense. It highlights and prioritizes patterns, trends, and relationships and is based in the search for differences and gaps, in the first instance, between women and men but also within those groups. It tells us what exists by providing a dense, information filled picture of the situation. It helps us to understand why it happens by exposing crucial links between components of the data.

In summary, data collection and analysis are dynamic, interactive, and ongoing elements of the research process. Both are done with rigor and are shaped by a focus on gender equality.

Strategic Initiatives

Gender analysis does not stop with the collection and analysis of data. It includes the creation of strategic initiatives which are interventions to foster equality between and within groups of women and men. It is important to note that data collection and analysis are strategic initiatives in and of themselves. The very process of beginning a gender analysis is a strategy which begins to reveal barriers to women's full participation in decision making. The initial activities of collection and analysis can provide us with new understandings of our worlds and, by placing women's and men's inequalities at the center of the research process, expose oppressive attitudes, systems, and structures.

The summative outputs of gender-sensitive data collection and analysis projects are equally important in shaping strategic initiatives. Like the initial collection and analysis, they can focus on inequalities embedded in a wide range of areas from day-to-day household differential responsibilities to the policies, programs, service delivery, and political/economic/social/cultural elements of institutions and structures. Strategies can be immediate and practical, as well as sustained and long term. As in the processes of data collection and analysis, strategic initiatives seek to be participatory. Ideally, civil society and key stakeholders together create collaborative strategies aimed at gender equality.

In brief, strategic initiatives are part of a sustained, ongoing, dynamic, cyclical process of planning, implementation, monitoring, and evaluation. They are neither tidy, final, nor definitive solutions and they need to be designed with a view to the particular contexts within which they will exist. Nonetheless, there are some important strategies which are worth mentioning.

First, as an adult educator might expect, gender training is a frequently used strategy. This training gives

communities, organizations, networks, and the people within them the time and space to engage with the concepts and issues of gender equality. It seeks to strengthen the capacity of women and men within the organization or network so that they can address gender inequalities. Ideally, the training is a process of both capability building and politicization. In the first instance, the objective is to build a solid information, skills and knowledge base and the self-confidence of the participants. In the second instance, the objective is to enable advocates to work collectively to claim equality at all levels.

A second example of a common strategic initiative is the reform of gender architecture, that is, the creation of structures within organizational structures to ensure that gender is a key consideration of every process, goal, and output. Often a gender unit is created to undertake a gender watch for the organization and acts in an advisory capacity to ensure that gender is mainstreamed. Gender training is frequently an important part of the unit.

A third initiative is gender-based budgeting. It uses a gender lens to assess and guide local, national, and international financial decisions about funds so that women and men can benefit equally. Such a lens reviews the allocation of funds for women-centric schemes, the incorporation of women into the decision-making processes of budget creation, the reflection of the concerns of women, and the assessment of the budget in relation to any commitments made to women and girls at international conferences.

In sum, there are many strategic initiatives which are part of the research processes of gender analysis. They must be responsive to their contexts, but there are some contemporary, exciting examples of global initiatives. It is worth noting too that essentially all initiatives seem to share the goal of gender mainstreaming. In other words, gender is irrefutably, irrevocably at the center of every consideration at every level. Women's views and their priorities are integral to every process, with the goal of achieving gender equality.

Benefits

Given the complexity of the concept and the intensity of the research process, one might well ask if the benefits which accrue to gender analysis are worth the effort. There are several responses to this. First, from an altruistic point of view, one might argue that the rights of women and men to lead full and satisfying lives is a worthwhile outcome for us all. It is a humane response to counter the asymmetric impact on women of fundamentalisms, wars, and neo-liberalism.

The second, and frequently presented, argument for gender analysis is an economic one. The basic line of reasoning is that, as 51% of the world's population,

women are a largely untapped reserve of labor which nations need to support their economic objectives. It is simply a matter of investing in people. Fuller participation of women, created by the insights of gender analysis and related strategic initiatives, can remove women from unproductive labor and engage them more fully in the paid labor market. Associated with this is the notion that women, once part of the paid labor market, can then become consumers – a role which also creates more demand for the production of goods.

An associated economic argument is that gender analysis allows for a more efficient allocation of resources. By exposing inequalities across a variety of variables, nations and funding aid agencies can see how to better utilize their funds. For example, gender analysis can reveal that participation of women in educational programs is low and that is because women travel miles to collect safe water. The problem can then be redefined as an infrastructure rather than educational problem and resources directed into drilling local wells to remediate the underlying problem.

The final economic argument is a very muted one, for it suggests that women have unique knowledge, skills, and experiences which will help achieve economic objectives. A gender analysis, for example, can reveal that women have longstanding practices which can better serve economic growth than ones introduced from other contexts.

Challenges for the Future

Gender analysis has a strong potential to address the oppression between and among groups of women/girls and men/boys. However, it faces several challenges. The first challenge is to create a global gender agenda to further the struggle for equality. It would be naive to believe that uncoordinated gender analysis projects, by themselves, are effective. They cannot effectively resist or indeed defeat the globalized neo-liberal agenda which impacts more harshly on women than on men. Lacking a global gender agenda, it is questionable if gender equality can ever be achieved.

The second challenge is to engage industrialized nations in the search for gender equality. Although many have ratified conventions like CEDAW, their operationalization of that commitment is often minimal. They too often seem to see gender analysis as an activity which needs to be undertaken in the nations to which they give aid rather than in their own backyards. They seem reluctant to expose the gender inequalities which so clearly still exist in their own industrialized nations. This reluctance also subverts the global gender agenda and is a major stumbling block to the realization of gender equality.

The third challenge is to extend the existing networks and alliances in the global struggle for gender equality. Currently, the movement is driven primarily by nongovernment organizations and women. Gender is related to

issues of the environment, peace, sustainability, and partnerships by gender activists with those issues are essential. In addition, men must be made accountable for women's ongoing disadvantage. This is particularly important since men continue to hold the majority of decision-making positions in the world. Although there are men who are allies in the struggle for women's equality, there are many more who, through their silences, allow women's oppression to continue.

Fortunately, adult educators are able to face these challenges with optimism. In the adult education sector, there are global organizations of women and men committed to gender analysis. They have been active for decades and have driven the global gender agenda with unfailing enthusiasm and commitment. These include the Gender Education Office (GEO) of the International Council for Adult Education, Development Alternatives with Women (DAWN) and Red de Educacion Popular entre Mujeres (REPEM). Our support for their work is a key to gender equality.

Conclusion

In conclusion, it is clear that gender analysis is neither a tidy concept nor exact research process. It is equally clear that there are individual and global benefits which accrue to gender equality. Despite the challenges, gender analysis is a useful tool to move us toward gender equality.

We know from history that the advancement toward women's equality is an uneven and precarious progress. However, the field of adult education is well located to play a key role in leading the struggle for gender equality. We have the expertise and experience to promote and undertake gender analysis. Adult education is, by definition, involved in the processes of teaching and learning so integral to that analysis. The final factor will be the extent to which the field makes an authentic, energized commitment to achieve gender equality. That is the ultimate challenge which confronts us all.

See also: Adult Education and Nation-Building; Community Based Adult Education.

Bibliography

- Belenky, M. F., Clinchy, B. M., Goldberger, N. R., and Tarule, J. M. (1986). *Women's Ways of Knowing*. New York: Basic Books.
- Dimand, R., Forget, E., and Nyland, C. (2004). Retrospectives: Gender in classical economics. *Journal of Economic Perspectives* 18, 229–241.
- Hayes, E. and Colin, S. A. (eds.) (1994). *Special Issue: Confronting Racism and Sexism. New Directions for Adult and Continuing Education: No. 61*. San Francisco, CA: Jossey-Bass.
- Hill, R. (1995). A critique of heterocentric discourse in adult education: A critical review. *Adult Education Quarterly* 45(3), 142–158.

- Hugo, J. M. (1990). Adult education history and the issue of gender: Toward a different history of adult education in America. *Adult Education Quarterly* **41**, 1–16.
- ICAE (2003). Agenda for the future six years later. *ICAE Report*. Uruguay: ICAE.
- Johnson-Bailey, J. and Cervero, R. (1996). An analysis of the educational narratives of reentry Black women. *Adult Education Quarterly* **46**(3), 142–157.
- Lewis, L. (ed.) (1988). *Special Issue: Addressing the Needs of Returning Women*. *New Directions for Adult and Continuing Education*. No. 39. San Francisco, CA: Jossey-Bass.
- Luttrell, W. (1989). Working-class women's ways of knowing: Effects of gender, race, and class. *Sociology of Education* **62**(1), 33–46.
- Medel-Anonuevo, C. (ed.) (1997). *Negotiating and Creating Spaces of Power*. Hamburg: UIE.
- REPEM (1996). *Education in Motion. Agreements on the World Conferences of the Nineties*. Uruguay: REPEM.
- Scampini, A. (2001). United Nations conferences: Commitments on education and gender justice. Education in motion. In Gender Education Office (ed.) *Women's Education. Chronicles of a Process*, pp 35–51. Uruguay: International Council for Adult Education.
- Stalker, J. (2005). Women's learning. In English, L. (ed.) *International Encyclopaedia of Adult Education*, pp 659–663. New York: Palgrave MacMillan.
- Thompson, J. (1985). *Learning Liberation. Women's Response to Men's Education*. London: Croom Helm.
- Tisdell, E. J. (1993). Interlocking systems of power, privilege, and oppression in adult higher education classes. *Adult Education Quarterly* **43**(4), 203–226.
- UNIFEM and UN/NGLS (1995). *Putting Gender on the Agenda. A Guide to Participating in UN World Conferences*. New York: UNIFEM.
- Veronica McGivney's recent critical survey. *Excluded Men*.
- Walters, S. and Manicom, L. (1996). *Gender in Popular Education: Methods for Empowerment*. London: Zed Books.

Further Reading

- Buttlerwick, S. (2005). Feminist pedagogy. In English, L. (ed.) *International Encyclopaedia of Adult Education*, pp 659–663. New York: Palgrave MacMillan.
- Hayes, E. and Flannery, D. D. (eds.) (2000). *Women as Learners: The Significance of Gender in Adult Learning*. San Francisco, CA: Jossey-Bass.
- Imel, S. (1995). Race and gender in adult education trends and issues. *ERIC Document ED382822*.
- Johnson-Bailey, J. and Cervero, R. (2000). The invisible politics of race in adult education. In Wilson, A. L. and Hayes, E. R. (eds.) *The Handbook of Adult and Continuing Education*, pp 147–160. San Francisco, CA: Jossey-Bass.
- Stalker, J. (1996). The New Right and adult educators: A feminist view. *Access: Critical Perspectives on Cultural and Policy Studies in Education* **15**(2), 67–79.

Relevant Websites

- <http://www.awid.org> – AWID.
- <http://www.icae.org.uy> – ICAE.
- <http://www.ilo.org> – International Labour Organization.
- <http://www.oecd.org> – Organisation for Economic Co-operation and Development.
- <http://www.unifem.org> – United Nations Development Fund for Women.
- <http://www.undp.org> – United Nations Development Programme.

Race and Ethnicity in the Field of Adult Education

J Johnson-Bailey and D Drake-Clark, The University of Georgia, Athens, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

Race and ethnicity are factors that act in major ways in determining how our society functions. Whether you divide the world into South and the North, East or West, or look within national cultures with diverse peoples, race can be tied to one's social and economic status. At present, the Northern Hemisphere is the site of unprecedented economic growth, while the Southern Hemisphere struggles with hunger, malnutrition, and preventable diseases. Within the North's multicultural nations, such as the United States, Canada, and the United Kingdom, race remains a more predictable marker of wealth, class, and poverty than educational level.

While ethnicity and other terms such as multiculturalism and diversity are identifiers and qualifiers often used more broadly than the term race, these descriptors not only denote racial groups, but the cultural aspects associated with racial groups and with groups that can be distinguished by religion, language, and traditions. There is no one correct way to define ethnicity since its usage varies across the world with different connotations specifying race, national origin, and tribal and clan membership. Over time, these terms came to be used interchangeably. Although race and ethnicity are socially constructed phenomena, herein, race and ethnicity will be used to refer to both genotypic and phenotypical differences.

Quite often societal customs, traditions, and norms conceal how one race is privileged over another or how a group of people is disadvantaged because of their racial or ethnic membership. The ordering of our society around race and ethnicity does not disappear at the classroom door; it continues to operate within our educational systems. Race is not the only positionality that categorizes or impacts the social order; however, race is an important factor in the rankings that regulate societal hierarchy. The effects of how race and ethnicity are viewed in the world are embedded in our educational fabric and the field of adult education is no exception.

There are historical and contemporary understandings of race in adult education and these views have framed the practical educational responses of adult educators. Race is a social construct that has no basis in biology (Frankenberg, 1993; Gregory and Sanjek, 1994; Winant, 1994) as anthropologists and biologists have long recognized that the human form cannot be examined through visual or scientific inspection to definitively determine a person's race. It is, at best, a fleeting notion established by an arguable set of physical characteristics. Although race

is a social construct, its effects are real in terms of social power and privilege (Giroux, 1997). Race has a major role in determining how our society functions. This ordering of the world along set queues occurs because, as people are categorized as belonging to a race, those persons are accorded all the rights, privileges, and baggage that accompany the classification (McIntosh, 1995). Therefore, to be of Asian ancestry has a different meaning than to be of White ancestry. And these meanings change depending on geography. So to be White in France carries a different set of connotations than does being Asian; and to be Asian in Malaysia carries yet another set of implications.

To discuss race and ethnicity in adult education is to confine the discussion largely to the West (as in North America, the United Kingdom, and Europe) and to limit the analysis to what is referred to as the North (highly industrialized nations located primarily in Europe and North America) simply because race and ethnicity are not widely acknowledged in adult education literature in the South. The adult education literature from the South tends to bound their discussions along the lines of nationality with national and regional parameters. Therefore, the discussion herein will focus unevenly on the North due to its availability of programs where an analysis of race and ethnicity are more possible. This article first presents a discussion of how race operates in the larger society and in the educational system. This introduction is followed by the discussions of adult education in the North and South.

A Historic Perspective of the West and the North

Adult education as a field was established in the early 1920s and its primary mission was the education and the continuing education of adults. From its inception, the field held that being an educated adult allowed one to participate fully as a citizen, to contribute materially to the well-being of society, and to develop as a person capable of striving toward self-actualization. Therefore, one of the primary purposes of adult education was to teach adults throughout their lifespan so that they could lead more fulfilling and productive lives. An American example of how the field worked toward the democratization of society through the education of adults was the Highlander Folk School (now called the Highlander Research and Education Center), an organization

established in 1932, whose adult education programs contributed significantly to two of the most important Western social movements of the twentieth century: the workers' rights and the Civil Rights movements.

While the stated goals of adult education have consistently been set forth as aspiring toward leveling the playing field for all adults, especially those lacking a basic education, and as desiring to empower learners so that they might engage in full citizenship, just the opposite often occurs. Since Western society is a place replete with hierarchical systems that privilege some and deny others, its academic systems are designed to reproduce the *status quo*. And as an academic discipline, adult education has also mirrored society by functioning like other areas of education. The major portion of adult education programs today in the United States, Europe, Australia, Japan, and in other heavily industrialized settings, involve continuing education, particularly continuing professional education, and human resource development programs in business and industry. Literacy, citizenship, and adult basic education programs in such settings have been largely relegated to federally and community-funded programs.

As a field, how we define race is reflected in how we write, research, and teach. For the most part, adult education literature has encompassed an implicit cultural hegemony, regardless of intentions, that centers the dominant White majority as the norm and empowers one group while disempowering others. In our major texts and publications, data on people of European descent is always present, making it therefore logical to assume that Whites have done the most or certainly have made the more significant contributions. The concealed message is that other groups have not done the same. In addition, in an educational setting, membership in a disenfranchised group translates into direct inequities: sub-standard education, tracking, and unequal opportunities for higher education. In society at large, such inequities have negative economic outcomes that construct barriers to the opportunities that are set forth as the ideals for all people in a democratic and prosperous society. Without naming it as race, or more accurately racism, in our educational arena, the underlying ideas of whiteness as superior and non-whiteness as inferior or deficient are ever present.

Other than the grassroots activist programs, which are bounded by their locale, the field of adult education in both practice and programming has been somewhat quiet on issues of race and ethnicity in that research by and about peoples of African, Asian, and Hispanic descent has been essentially absent from the adult education literature. Throughout the years, on the few occasions when race and ethnicity have been topics in the North American adult education literature, the discussion has been confined mostly to a dialog about Blacks or African Americans. Race and ethnicity in contemporary Western society are

used to refer primarily to people of color; Whites are treated as if they have no race.

In examining race in the West, the conversation in the United States has been more overtly framed in racialized terms, while European and British examination of difference have been framed by discussions of class. In the 1930s and 1940s in the United States, there was a strong conversation in the literature on Blacks in adult education that was led by Alain Locke, the first Black president of the American Association for Adult Education (an organization which exists today as the American Association for Adult and Continuing Education). It was under his direction that national conferences on African American adult education issues were held in 1938, 1940, 1941, and 1942, and it was his leadership that led to the chronicling of African American adult education national efforts in the 1948 yearbook issue of *The Journal of Negro Education*. However, these efforts waned after the publication of the 1948 yearbook.

The historical documents that define North American and European Adult Education are the nine *Handbooks* published from 1934 through 2000, which provide comprehensive overviews of the field's major issues and concerns. Given the importance of these texts in the field, the ways in which race is addressed or not addressed is informative. Although race is a central location for the negotiation of power and privilege in education and in society, this topic never formed the focal point of a single chapter until the 2000 handbook. While the 1934 (Rowden, 1934) and both handbooks of 1936 (Locke, 1936a, 1936b) had chapters on adult education for Negroes, and the 1948 handbook (McCurtain, 1948) had a chapter on adult education of American Indians that offered descriptions of programs for these groups, no chapter dealt with race and ethnicity as a topic by addressing the central role that race was playing in Western society during this period of legal segregation in the United States. This absence of discussion about race and ethnicity continued in subsequent handbooks, even in the face of major social upheavals such as the Civil Rights movement and the American desegregation of public schools and other public institutions that provided adult education. The 1989 handbook returned to the trend of the 1930s and 1940s by having a chapter on racial and ethnic minorities and adult education (Briscoe and Ross, 1989). In addition to this chapter on specific racial groups, a discussion of race in the context of the social setting for adult education appeared in two handbooks (London, 1970; Rachal, 1989).

The way race has been socially constructed in the literature of North American and European adult education over the past half-century has been remarkably stable. This view is that the White race is the norm against which all other races are to be compared. This perspective, which is the colorblind perspective (it is not seen and it is not discussed) is so deeply embedded in the social

fabric that there has been little discussion of adult education for Whites even though the White race has constituted the vast majority of the population for adult education. Whenever race is discussed in the handbooks, then, it is conceptualized as non-White. Of course, when one group is normative, then the others are viewed as abnormal. This leads to the obvious conclusion that separate chapters would be needed to discuss the specific educational efforts being made to address the needs of these special populations.

Starting in the 1980s, there was a renewed interest in issues of race and ethnicity in the adult education literature. These contemporary perspectives on race in adult education fall into two major categories: multiculturalism and social justice. The terms multiculturalism and cultural diversity are used interchangeably throughout adult education literature, with the latter being the more current term. However, multicultural scholars distinguish a difference between multiculturalism and cultural diversity. They explain that cultural diversity is a view of society as a collection of varied cultures whereas multiculturalism is a particular political and ethical stance toward cultural differences. Globalization, the process where goods move freely among nations without regard to national borders and national identity, is also included under the multiculturalism umbrella and routinely offers an analysis of the place of disenfranchised or indigenous cultures in the shrinking world economy. A central idea that is shared by all types of multiculturalism is that one culture is seen as dominant and therefore the educational need is to teach the importance of values and beliefs that are held by other cultures. Thus, from its inception, multicultural education has called for recognition and inclusion of the contributions of other cultures in the literature, research, and praxis. Multicultural education was first introduced in the adult education literature by Horace Kallen in 1915 and expanded on by Alain Locke in 1925. Locke, one of the few leading Black adult educators of the early 1900s, represents a segment of the field that champions the multicultural argument by making known the causes and worth of certain groups. He expressed a belief in the redemptive powers of multiculturalism and this view remains constant in the contemporary adult education literature.

The social justice perspective can be divided into two categories. One social justice outlook states that there is indeed a right and moral position that should direct our society. Proponents of this position, like contemporary noted adult educator Phyllis Cunningham, the Presidential Teaching Professor at Northern Illinois University, ask adult educators to remember and live by the mission of the field, which is to equip adults to be full participants in a democratic society. This message has remained constant since the field's inception in the early 1920s, when writings by Alain Locke suggested that adult education could

equalize the wrongs of society by providing an education to an adult populace that did not receive the basics through traditional education. This position continues in current writings by many adult education scholars.

The second trend in the social justice movement is not only to state the right and moral imperative of what should occur in the field but also to add an activist component, addressing the differences between groups and highlighting how power is exercised in favor of one group and to the detriment of others. The contemporary writings in this category use this approach not only for issues of race but also for all societal hierarchies constructed to serve one group while disenfranchising another. For example, scholars of critical race theory, such as Derrick Bell, explore how the systems and structures of White supremacy that subordinate people of color have been created and are maintained in American society as well as how to change the relationship between the law and racial power. An important part of fostering diversity and multiculturalism in adult education means discussing the elements of power and privilege accorded along the positionalities of race and ethnicity and the effects of this privilege on those from such marginalized positionalities. In the last decade, another important aspect of discussing race and ethnicity has been an examination of how Whites construct their racial identity and how the social construction of Whiteness affects educational systems.

When the discourse on Whiteness or privilege occurs in the adult education literature, it usually involves the following: recognition of privilege and of under-privilege, an examination of classroom practices, examples of curriculum and/or texts that reproduce privilege, and various anecdotal examples of how privilege operates in society. A large segment of the literature in social justice deals with the interlocking nature of race, gender, and class. Even though this area of research and writing acknowledges that the power lies in the hands of a dominant White majority, it stops short of saying that the concentration of power is deliberate and that the intent is to retain the present balance of interests. Instead, the social justice literature asks for a renegotiation of the balance of power.

Several factors contributed to the return of race and ethnicity in the 1980s as a focus in the adult education literature, research, and praxis. This specific focus was a direct outgrowth of two changes in perspective that were introduced in the 1970s: Freire's work around class and adult student empowerment, and the emergence of critical theory. In addition, the influx of people of color as students and instructors affected the attention given to issues around race and ethnicity. The impact of this new presence became evident in the 1980s as race and ethnicity emerged in the literature and at research conferences, as well as in curriculum and program planning.

The field's elite measure of acceptance, refereed journals, began to show an increase in the number of articles that could be classified as directly related to race and ethnicity. In 1991, a landmark article, *Needed: A multicultural perspective for adult education research*, by Ross-Gordon on the subject of race and ethnicity in adult education research, appeared in *Adult Education Quarterly*, one of the field's most prestigious journals. Ross-Gordon analyzed the status of racial and ethnic research and found 15 articles on race and ethnicity published between 1985 and 1990. A subsequent article in 2001 by Johnson-Bailey in the *International Journal of Lifelong Education*, examining the period between January 1990 and December 1999, revealed that 94 articles were published on issues, concerns, or prominent historical figures pertinent to ethnic or racial groups: 20 on African Americans; ten on Native Americans; ten on Asians; nine on Hispanics; and 45 on multiculturalism or cultural diversity. Interestingly, the articles pertaining to Asians only looked at Asian groups existing within their national homelands, and without exception, the works did not focus on 'Asian-ness' as otherness.

Research conferences, which provide a unique opportunity for the exchange of ideas among colleagues, have also demonstrated a surge of interest in race and ethnicity. The proceedings of major adult education conferences reveal a developing public discussion on race and ethnicity at the Adult Education Research Conference (AERC) and Standing Conference on University Teaching and Research in the Education of Adults (SCUTREA), two of the premier conferences in the field. With a small degree of variance, approximately 3–7% of the research presented at AERC and SCUTREA in the late 1980s through 2005 focused on ethnic and racial diversity and an increasing 2–3% on social justice issues such as gender and sexual orientation.

Only one adult education conference has research on race and ethnicity as its focus and unifying theme. The African Diaspora Adult Education Research Conference (formerly the African American Research Pre-Conference) was founded 1993 and has historically been attached as a pre-conference to the larger AERC. The African Diaspora Adult Education Research Conference debuted at Pennsylvania State University in May 1993, accompanying the 34th annual AERC. The theme of that first conference, *A Link for Community Development and Empowerment*, was suggestive of the conference's perspective and is now a consistent ideology connecting the subsequent conferences to the first.

The field of adult education in the North has enjoyed two distinct time periods relevant to focusing on race and ethnicity. During the discipline's formative years, beginning in the 1920s, issues of race and ethnicity were examined in a summative fashion through descriptive writings that related what was transpiring within various communities. The second and contemporary phase of

adult education began 60 years later and reflects both a more open, analytical, and a growing discussion of race and ethnicity with a definite focus on multiculturalism and social justice.

A Southern Focus of Adult Education

To speak of adult education in the Southern Hemisphere is to speak of nationality, indigenous peoples, poverty, globalization, and literacy, more so than to speak of race and ethnicity. However, race and ethnicity are tied to these issues and are especially linked to poverty and low levels of literacy in the South. Literacy and literacy initiatives dominate the Southern adult education literature, followed only by indigenous peoples' knowledge as a topic of discussion. The topic of indigenous knowledge includes honoring the oral traditions and languages of native populations amidst the emerging push to use Western languages to accommodate commerce.

The focus on literacy in the South seems driven by the global economic market as African nations, India, and China, the fastest growing and more populous world areas, are thought to lag behind industrialized nations due in part to the fact that large segments of their citizens are undereducated. So the literacy exchange is twofold, with the idea being that both sides win from the effort regardless of the impetus: for the purpose of advancing economic initiatives for the South and for the purpose of developing and advancing a labor pool for the North. However, it must be noted that when economics drives the market and dictates policies, individual nations and therefore groups of people are inevitably disenfranchised because the idea of equality and other democratic ideals are inextricably tied to governments and social movements and not to economies. The capitalistic process and the phenomenon of globalization disempower countries, their people, and any claims for rectifying any social ills that might have occurred within set national borders.

The countries that comprise the Southern Hemisphere are more homogenous populations than the North, thus minimizing the necessity to focus on race. Migration patterns show a movement of people from the South to the North. Much of the diversity in the North derives from voluntary migrations and involuntary relocation of peoples of color from the South. However, since many Southern Hemisphere countries have been subjected to colonization by Europeans and those of European descent, the presence of race-based privilege is still a factor in the South. This factor perhaps encouraged Bostwana scholar, Youngman, to ask about the role that adult education plays in the reproduction or contestation of race-ethnic domination (Wangoola and Youngman, 1996). In addition, globalization, which is driven by the Northern countries of the United States, Japan, Germany, and the United

Kingdom, is the new face of colonialism. The complex global-connectedness phenomenon, a result of globalization which deemphasizes borders, also weakens cultural ties to home and place. To this end, it is theorized that modern racism has been exported with capitalism (Wangoola and Youngman, 1996).

Although race and ethnicity are not as generally discussed in international forums as they are in the West, the United Nations Educational and Scientific Cultural Organization (UNESCO), which was established in 1945, adopted a *Declaration on Race and Racial Prejudice* in 1978. The preamble noted that racism, racial discrimination, colonialism, and apartheid continue to afflict the world in ever-changing forms and the United Nations declared a desire to play a vigorous and constructive part in implementing the program of a *Decade for Action to Combat Racism and Racial Discrimination*, as defined by the General Assembly of the United Nations at its 28th session. The ten articles of the declaration encouraged the countries of the world to eliminate discrimination, adopt policies to improve the lives of its citizens of color, and to respect cultural differences.

For the most part, any discussion of adult education in the South occurs against the backdrop of UNESCO, its conferences, and programs. For many nations in the South, adult education is a matter of redress as many Southern nations did not have free mandatory children's education and/or because of an agricultural base, experienced low participation in formal schooling as farming necessitated the involvement of all members of the family, including children. The United Nations' *Literacy Decade*, 2003–2012, emphasizes the importance of honoring indigenous knowledge as a means of promoting education. Important adult education enterprises spearheaded by UNESCO are Education for All (EFA) and International Adult Learners' Week. Adult Learners' Week was first held in the United Kingdom in 1992 and was formally endorsed by UNESCO in 1999. Since then, over 40 countries have joined in the annual participation of Adult Learners' Week. The Week's purpose is to impact policy, promote access to the existing adult education programs, and increase literacy. Another cornerstone program that addresses issues of redress is the EFA movement that began in 1990 at the World Conference in Jomtien, Thailand, setting as its universal goal, education for all children and adults and pledging a 50% reduction in the 1990 rate of illiteracy by 2000. By the 2000 World EFA Conference in Dakar, Senegal, 180 countries had signed on to promote basic education for all citizens.

The region of Southern Africa provides an opportunity to look at adult education in a Southern setting. Prior to the emancipation of many nations in Southern Africa, adult education was recognized by the 14 nations in the region as an essential and necessary part of liberation struggles. This is aptly illustrated by the underground educational plan that was laid by the Black South Africans

interned at Robben Island. The literate and educated prisoners worked in secret during their lunch breaks to tutor the less literate and undereducated inmates.

Southern African nations came together to form the Southern African Development Community (SADC) in 1992. This coalition of 14 countries, which includes Botswana, South Africa, Zambia, and Mozambique, expresses the importance of educating the citizenry as a means of encouraging full participation and exercising of rights, with most countries in the pact spending 20% of their budgets on education. Given the homogeneity of its countries and the legacy of colonization, SADC members focus on indigenous knowledge as a bridge between organic knowledge and formal knowledge. The concept of knowledge from the people is particularly important in health education. The low life expectancy of the region has been compounded by the HIV/AIDS pandemic. While championing the importance of local knowledge, the adult health educators are faced with the challenge that at times religious, gender, cultural, and tribal-based behavior, practices, and traditions may conflict with modern medical practices.

Summary

In conclusion, it is noted that the North has a perspective on race and ethnicity that is bound by genotypic and phenotypical differences. However, the South's viewpoint regarding difference is based more on nationality and culture. In addition, the adult education needs of the two regions are determined more by economic forces than by the races of its people. For example, the South's focus on literacy is driven by the North's need for a readily available and cheap workforce and the South's desire to become an economic competitor. Despite this commonality, within each of the two regions, the poor and disenfranchised are the colored peoples of the area, whether they are the majority Blacks of South Africa or the African Americans or Hispanics of the United States. Further, the breakdown of national borders by globalization is replacing national adult education initiatives with corporate initiatives. The North is experiencing a tremendous growth in the human resource development sector of adult education, while literacy efforts are relegated more to the non-profit sector and to being a low priority for the underfunded public sector.

Bibliography

- Briscoe, D. B. and Ross, J. M. (1989). Racial and ethnic minorities and adult education. In Merriam, S. B. and Cunningham, P. M. (eds.) *Handbook of Adult and Continuing Education*, pp 583–598. San Francisco, CA: Jossey-Bass.

- Frankenberg, R. (1993). *The Social Construction of Whiteness: White Women, Race Matters*. Minneapolis, MN: University of Minnesota Press.
- Giroux, H. A. (1997). Rewriting the discourse of racial identity: Towards a pedagogy and politics of whiteness. *Harvard Educational Review* 67(2), 285–320.
- Gregory, S. and Sanjek, R. (eds.) (1994). *Race*. New Brunswick, NJ: Rutgers University Press.
- Locke, A. (1936a). Adult education for Negroes. In Rowden, D. (ed.) *Handbook of Adult Education in the United States, 1936*, pp 126–131. New York: George Grady Press.
- Locke, A. (1936b). Lessons of Negro adult education. In Ely, M. (ed.) *Adult Education in Action*, pp 224–226. New York: George Grady Press.
- London, J. (1970). The social setting for adult education. In Smith, R. M., Aker, G. F., and Kidd, J. R. (eds.) *Handbook of Adult Education*, pp 3–23. London: Macmillan.
- McCurtain, R. H. (1948). Adult education of American Indians. In Ely, M. L. (ed.) *Handbook of Adult Education in the United States*, pp 65–69. New York: Teachers College Press.
- McIntosh, P. (1995). White privilege and male privilege: A personal accounting of coming to see correspondences through work in women's studies. In Anderson, M. L. and Collins, P. H. (eds.) *Race, Class, and Gender*, pp 76–87. Belmont, CA: Wadsworth.
- Rachal, J. R. (1989). The social context of adult and continuing education. In Merriam, S. B. and Cunningham, P. M. (eds.) *Handbook of Adult and Continuing Education*, pp 3–14. San Francisco, CA: Jossey-Bass.
- Rowden, D. (1934). Adult education for Negroes. In Rowden, D. (ed.) *Handbook of Adult Education in the United States*, pp 124–130. New York: American Association for Adult Education.
- Wangool, P. and Youngman, F. (eds.) (1996). *Towards a Transformative Political Economy of Adult Education: Theoretical and Practical Challenges*. DeKalb, IL: LEPS Press.
- Winant, H. (1994). *Racial Conditions: Politics, Theory, Comparisons*. Minneapolis, MN: University of Minnesota Press.

Further Reading

- Cassara, B. (1994). *Adult Education in Multicultural Society*. London: Routledge.
- Denton, V. L. (1993). *Booker T. Washington and the Adult Education Movement*. Gainesville, FL: University Press.
- Merriam, S. B., Courtenay, B. C., and Cervero, R. M. (2006). *Global Issues and Adult Education: Perspectives from Latin America, Southern Africa, and the United States*. San Francisco, CA: Jossey-Bass.
- Neufeldt, H. G. and McGee, L. (eds.) (1990). *Education of the African American Adult: An Historical Overview*. New York: Greenwood Press.
- Sheared, V. and Siskel, P. (2001). *Making Space: Merging Theory and Practice*. Westport, CT: Bergin and Garvey.

Trends in Workplace Learning Research

T Fenwick, University of Alberta, Edmonton, AB, Canada

© 2010 Elsevier Ltd. All rights reserved.

The nature and organization of work has changed so rapidly in the past decade with the introduction of new technologies and the effects of globalization that learning has become a lightning rod, attracting all sorts of new attention outside educational debates. Change in organizational structure and culture is now a learning issue. Innovation, now considered critical for competitive positions in the knowledge economy, is linked to learning. The increasing need to integrate migrant workers, along with attendant issues of inclusion, equity, gender, and race politics in the workplace, all have infused the workplace learning agenda. Therefore it is not surprising that research in workplace learning has accelerated since the mid-1990s, expanding into a wide assortment of fields including, besides adult education, organization and management studies, human resource development, sociology of work, economics, feminist studies, and industrial relations.

Definitions and Different Perspectives

The effect of this expansion has been to both enrich the existing perspectives and methods of understanding learning in work, and to blur categories. Across different fields, definitions of terms (such as learning, skill, empowerment, critical, and workplace) are multiple and often contradictory. Work can be paid or unpaid, based in organizational action or individual reflection. Work activity and learning varies widely depending on public, private, or not-for-profit sector, or on whether we are referring to a trades-worker or manager, self-employed professional, farmer, or domestic worker. Indeed, assertions about a generic workplace are inappropriate.

In some attempts to categorize different perspectives and definitions, the term workplace learning has been limited to individual change, with organizational learning reserved for groups. This is an arbitrary and rather unhelpful division, particularly when so many perspectives of learning in work refuse to separate the individual from the collective in examining learning processes. So in this article, workplace learning will refer to relations and dynamics among individual actors and collectives, along with group practices, tools, objects, cultural discourses, histories, and environmental forces that are all entwined with the behaviors of individuals and collectives. Workplace learning in this discussion refers not to formal planned training but to informal learning, that which is embedded in everyday material action and social interaction. Learning is defined not as

the product of change but as the process of generating new knowledge in activities and contexts of work; in particular, knowledge that expands human possibilities for flexible and creative action.

Overall, workplace learning research is a relatively new field characterized by multiplicity and vibrant debate, one that resists neat categories and typologies. This will become clear through the following discussion of three topics in workplace learning research, all prominent in current debates particularly in adult education: processes of learning, issues of identity and literacy, and power and politics in learning. These topics are not comprehensive, but illustrate key issues and point to future directions in workplace learning research.

Researching Processes of Learning in and Through Work

A review conducted of workplace learning literature published in nine scholarly journals between 1999 and 2005 identified various distinct perspectives of learning process (Fenwick, 2008). These can be distinguished according to the view of the individual relative to the collective, the role of reflection, and the role of action in learning. For example, some theorists see clear distinctions between the autonomous single individual and the environment, while others do not separate individual from collective or privilege the individual human over the complex mix of tools, cultural practices, and joint action of the system. Six perspectives are summarized here with the greatest attention accorded to the final theme of co-participation and co-emergence, as this appears to be a dominant direction emerging in workplace learning research.

Individual Knowledge Acquisition

Learning used to be understood primarily as acquisition: where individuals are believed to acquire and store new concepts and skills/behaviors as if knowledge were a package that didn't change as it transferred from its source to the learner's head. Learning tends to be valued according to the extent of its contribution to the organization's performance and productivity. Learning workers are therefore understood to be acquiring intellectual capital, increasing the organization's resources and return-on-investment on training. Research tends to focus on how to harness, draw out, and use the individual's acquired

knowledge. Preoccupations include transferring acquired knowledge to practice, measuring competency (reliable valid measures and competence definitions are identified as problematic), and narrowing the gap between training investment and results. While this perspective appears to have declined since about 2001, it still appears in certain human resource development and management/organization studies of learning.

Reflection and Human Development

Here the emphasis is on sense-making. Learning is viewed as individual and collective construction of (new or altered) meanings: to identify problems, emerge solutions, or engage in collective inquiry. The general base is constructivist learning, for example, through reflection, which is affected by both the individual's history and unique perspectives, and by the meanings and language shared among the collective. Two directions have emerged in the research. One focuses on developing knowledge and practice of individuals; the other focuses on developing the organization's processes and culture. In both directions, workplace learning is viewed as fundamentally driven by reflection-on-action, in the swamp of uncertain, ambiguous, and contradictory dilemmas of practice. Reflection during and after the doing supposedly transforms experience into knowledge, which can then be represented and generalized to new contexts.

In research focused on the individual, studies have examined how to promote self-directed learning capability of individuals, and illuminate the relation of work to individual developmental processes and learning styles. In research focused on the group, studies have explored the nature of group reflection, and what factors influence particular meaning constructions at work. However, researchers who are critical of sense-making ideas show the rarity in practice of group critical reflection, dialog, and inquiry. Individuals are disillusioned with such practices, and the notion fails to sufficiently account for power relations in workplaces and knowledge hierarchies – including those created by researchers. Some critics have maintained that the emphasis placed on reflection is simplistic and reductionist, overemphasizing rational thought, and understating the unpredictable social tangles of everyday practice in which people develop.

Network Utility Model

Here, learning is portrayed as individuals and teams sharing useful strategies through networks within and across organizations, often electronically enabled, primarily for purposes of improving others' performance. Learning is thus information access. The key research preoccupations are improving diffusion: capturing, managing and organizing content, removing network barriers, and generally

facilitating efficient, effective information flow or knowledge transmission (just-in-time) through a network. Learning networks are reported to take different shapes related to contexts, work characteristics, interactions, actor dynamics, and strategies; interorganizational networks are the most complex and take long time-periods to develop.

Most other research explores sociocultural issues, such as barriers and enhancements to knowledge sharing. For example, individuals and teams have been found to be willing to share if sharing is valued and supported and if the organization restructures payoffs for contributing, increases efficacy perceptions, and makes employees' sense of group identity and personal responsibility more salient. Overall, sociopolitical dynamics have been found to affect network effectiveness far more than technology.

Levels of Learning

Here the organization and individual (and team) are viewed as separate, distinct levels and forms of learning. This static layer-cake depiction is similar to the network utility model, but goes beyond linear transmission of information to acknowledge practices and politics. Research focuses on what happens at different levels, how different levels affect one another, how to link the levels in practice, and how/when to balance the exploratory (knowledge-creating) with the exploitive (knowledge-diffusion) dynamics. Learning levels might be depicted as units of people (individual, group, and organization) or as phases of learning (innovation, sharing, and routinizing). The link between levels is conceptualized rather mechanistically as cross-fertilization, diffusion, pipeline-sharing, and motoring.

Communities of Practice Model

This model, based on ideas first conceived by Jean Lave and Etienne Wenger, has been widely taken up since the publication of Wenger's (1998) book *Communities of Practice*. Learning is viewed as situated participation, embodied in the joint action evident in a community of practice (CoP). Individuals learn as they participate *in situ*: by interacting with a particular community (with its history, assumptions and cultural values, rules, and patterns of relationship), the tools at hand (including objects, technology, and language), and the moment's activity (its purposes, norms, and practical challenges). Knowing and learning are defined as engaging in changing processes of human participation in a particular CoP. A CoP is any group of individuals who work together for a period – such as a sports team, a workplace department or project group, a class, or club – developing particular ways of doing and talking about things that their members come to learn through action. The objective is to become a full participant in the CoP, not to learn about the practice. The community itself defines what constitutes legitimate practice.

Research seeks to explain the adaptation and reconfiguration of practices to meet changing pressures, and identify ways to facilitate these dynamics. Community learning is found to be affected by relational stability (trust), variety (new ideas, risk), and group structure (networks, competence). Learning is constrained by time pressure, deferral, and centralization within and across projects. Billett (2001) developed a useful model of guidance (direct, indirect, and environmental) that moves workers toward fuller participation in workplace activity, and hence to more comprehensive and critical knowledge as actors in their CoP. Critics of the CoP model of learning have pointed out its weak analysis of politics and solidarities within the community, including those which determine what counts as legitimate knowledge and expertise, and what knowledge (and identities) become marginalized. The CoP does not shed much light on how innovation occurs or how to develop specialized knowledge, especially during rapid change.

Co-Participation or Co-Emergence Model

In this orientation, individual and social processes are viewed as unique but enmeshed, and deserve examination at micro and macro levels of analysis. Learning is knowledge creation through social participation in everyday work. The conception is of mutual interaction and modification between individual actors, their histories, motivations and perspectives, and the collective (including social structures, cultural norms and histories, and other actors). Radical versions expand the collective to include environmental architecture, discourses, and objects, as in actor–network theory where knowledge circulates and is translated in each interaction of one agent mobilizing another. Cultural–historical activity theory views the individual and organization in dialectical relationship, where learning is occasioned by questioning practices or contradictions of the system, and is distributed among system elements: perspectives, activities, artifacts, affected by all contributors and clients. Complexity theory treats learning as inventive/adaptive activity produced continuously through action and relations of complex systems, occasioned in particular through disturbance. Most agree that learning is prompted by particular individuals (guides or mentors), events (conflict or disturbance), leaders (e.g., encouraging inquiry, supporting improvisation), or conditions (learning architecture).

Long popular in Nordic research of workplace learning but just recently emerged in North American research is cultural–historical activity theory (CHAT). Here, learning is viewed as change in a community's joint action. The community's activity is shaped by its rules and cultural norms, division of labor and power, and mediating artifacts (language, tools, and technologies) that

it uses to pursue the object – a problem at which activity is directed. Learning occurs as the collective construction and resolution of tensions or contradictions occurring within this activity system. Unlike other practice-based systemic perspectives of workplace learning, CHAT retains its Marxist influences in its recognition of the inherent contradictions in capitalist work systems based on labor exchange, and in its analysis of the historical emergence of particular practices and ideologies (see Chaiklin *et al.*, 2003).

In organizational studies and increasingly in educational study in Canada, complexity theory is also gaining acceptance as a useful way to understand how activity, knowledge, and communities emerge together in the process of workplace learning. Individual interactions and meanings form part of the workplace context itself: they are interconnected systems nested within the larger systems in which they act. As workers are influenced by symbols and actions in which they participate, they adapt and learn. As they do so, their behaviors, and thus their effects upon the systems connected with them, change. The focus is not on the components of experience (which other perspectives might describe in fragmented terms: person, experience, tools, and activity) but on the relationships binding them together. Workplace learning is thus cast as continuous invention and exploration in complex systems.

Critics suggest that such practice-based studies of workplace learning bypass questions of politics and power relations: who is excluded from the construction of knowledge in a CoP, what dysfunctional or exploitative practices are perpetuated in communities of practice, and what hierarchical relations in the workplace reproduce processes of privilege and prejudice. Issues raised include accreditation and assessment of learning when it's buried in co-participation, distinguishing desirable from undesirable knowledge development, accounting for changing notions of what is useful knowledge, and differentiating influences of particular groups in the co-participational flux (positional, generational, gendered, etc). At issue is the extent to which sociocultural learning theories including notions of communities of practice, complex adaptive systems, or even CHAT suppress or enable core questions about the politics and purposes of workplace learning.

Researching Identity and Literacy in Workplace Learning

Work communities are powerful sites of identity, practices, and knowledge systems in which individual workers' desires for recognition, competence, participation, and meaning are both generated and satisfied. Identity is ultimately a representation or mental conception that we ascribe to ourselves and others: our conception of who we are, our

identity, is constituted by the power of all of the discursive practices in which we speak, which in turn speak us, so to say (Chappell *et al.*, 2003: 41).

People's sense of their own knowledge in work, and the knowledge valued by the group to which they see themselves belonging, form a critical element of their sense of identity. Further, their participation in work practices is entwined with the identities they come to inhabit within a particular community. Identity work itself involves learning. Workers figure out how to position themselves in an organization, how to perform identities that are acceptable to their immediate peers but also that allow themselves freedom and some autonomy and control. In work environments of rapid change where people must transform their practices, they understand their knowledge as more mutable and fluid; they are adopting shifting identities: they literally learn to perform different selves and knowledge in different environments. Yet at the same time, people often employ deliberate strategies to anchor their identities.

Researchers have explored how particular identities are constituted among these varied coordinates, and how learning processes are implicated in individuals' subjections, negotiations, assertions, and shifts of subjectivity (see Billett *et al.*, 2007). One case studied miners compelled to transform their work from manual labor in heavy equipment operation to computerized manipulation of equipment using joysticks in an office. At issue was the men's macho-masculine identities, which were no longer relevant. Overall, adult education researchers are interested in how people come to recognize the limitations of their current work identities, how they recognize possibilities for new identities, and what strategies they learn to cope with repressive constraints on their work identities.

Language and literacy are closely related to identity and learning: people's sense of who they are and what they know and can do at work are embedded in the language and textual practices they use. One area of workplace research examines how learning is shaped by particular written texts in changing workplace environments such as documents, policies, record-keeping forms, or employee growth plans. Such texts standardize what counts as knowledge, thus controlling the work practices and working relations of people employed. As people are pressed to learn new literacies in their work, their sense of self shifts along with their ways of conceptualizing and doing their work. Globalizing forces, such as standardization, are introducing new texts and literacies into work such as the massive form-filling required to demonstrate compliance with ISO-9000 standards, or the accent training given in call centers to make Indian workers sound American on the telephone. New literacy practices have been engendered by the shift to post-Fordist work arrangements such as self-directed teams. Workers used to hierarchical communication pipelines have had to learn how to participate

productively in team meetings: how to set goals, analyze, and assess collective work through leaderless reflective team dialogs.

A concern for technological upskilling, or developing worker literacy in new technologies, has prompted many training programs as well as research to explore exactly how people learn to work with technologies. On this theme, Sawchuk (2003) studied technology learning practices of workers, integrated with everyday life and mediated by artifacts such as computer hardware and organizational settings. He presents a "working-class standpoint" that opposes even while it accepts managerial control, and often copes through subversion. Sawchuk found that for working-class people, computer technology is a "key signifier" for workers' deepest class-based desires and fears: the sense of losing control and being left behind in an inevitable techno-obsessive world, the frustrations of trying to figure out capricious computer processes, or the trials of purchasing the right computer on limited incomes. Yet working-class learning thrives in informal networks, what Sawchuk calls "solidaristic networks" (p. 123): where mutuality and group orientations within stable working-class communities produce knowledge in everyday computer learning. It is here that Sawchuk found an "enormous surplus" of knowledge production capacity as well as emancipatory potential for working-class people.

Researching Power and Politics in Workplace Learning

Calls for greater attention to research examining power relations in workplace learning have not resulted in much empirical research. In the literature review of nine journals 1999–2005 (Fenwick, 2008), only about 15% of published articles touched upon power or politics, and these were almost exclusively theoretical in nature. Five different perspectives of power appear to be represented among these. In the radical view, organizations are viewed as sites of central contradictions and ideological struggle between those who control the means of production and those whose labor and knowledge are exploited. In the discursive view power is viewed as circulating through regimes of knowledge and discursive practices. Power is not possessed by particular people or institutions, but is constantly created and readjusted through relations among people and practices, notions of what is normal and what is valuable. Workers participate in and help to sustain the very regimes that discipline and repress their identities and opportunities. In the identity politics view, power relations consolidate a dominant workplace culture whose practices and beliefs actively marginalize or even persecute individuals by virtue of their gender, race, religion, sexual orientation, or conformance to the ability-norm valued by the dominant. The micropolitics view analyses power

relations as confined to individual strategies to improve their own advantage, such as gamesmanship. Finally, the community view avoids a critical analysis of structures, knowledge politics or even interpersonal politics: power is viewed as benign energy, exercised mainly in mobilizing individuals around shared vision, mutual engagement, and sense of belonging.

Adult education analyses of work tend most to feature the radical and identity politics views of power, so further discussion of these two is warranted to show their links to learning. In the radical view, workplace learning is often envisioned as radical transformation among workers: empowerment purposed towards workplace reform. Radical or emancipatory learning involves workers first in critically analyzing existing repressive conditions of work, including mechanisms in place for controlling knowledge and the means of production. Then strategies for resistance and change are generated collectively, in a learning process that builds solidarity, individual and collective agency, and workers' capacity to defend their rights. This learning process of transformation is often positioned in opposition to reproduction, where workers learn to accept and even support exploitative, hierarchical structures that subjugate them and reproduce existing (inequitable) power relations. However, some have argued that this traditional dualism may be overly simplistic, that research needs to examine how reproductive and transformative learning are entwined in everyday work and with what workers themselves want to learn.

In critical workplace learning research, adopting an identity politics view of power, issues of race, disability, sexual orientation, and religion are almost completely absent despite their growing importance in other areas of adult education. Gender, on the other hand, has received substantial attention. Studies show that women continue to confront gendered work knowledge and training structures in organizations based on patriarchal values, male-oriented communication patterns, and family-unfriendly schedules. New expectations for continuous learning related to learning organization initiatives, self-directed teams, technological upskilling and development of new literacies, all create work overload that poses particular burdens for women who still carry the double-duty bulk of domestic and childcare work at home. Women in particular are often expected to nurture the close relationships and community that organizations want, to mentor others, and to display cheerfulness (Mojab and Gorman, 2003). Yet the learning valued and supported most in organizations tends to be related to leadership development, knowledge creation, and organizational growth, in professional/managerial jobs where women continue to be underrepresented. Meanwhile, women who are new immigrants and women of color are overrepresented in precarious, contingent employment such as call centers, food service, and home-based work where there are few learning

opportunities or communities offering the rich sort of participation and learning networks to help women obtain better-paying, more secure employment.

Future Directions

As is clear in this discussion of learning process, identity, and power in workplace learning research, perspectives range widely according to fundamental understandings of what constitutes knowledge, how it is constructed, how workers are connected to one another and to their environments, and how action and reflection are related. Furthermore, different researchers and educators propose very different purposes for workplace learning, and these tend to influence the way learning is understood. Some focus on individual human development, some on building solidarity and political consciousness among workers, while others are more interested in upskilling workers or changing organizational culture to increase productivity. Labor educators tend to focus on enabling workers to obtain control of their own knowledge. Management and human resource theorists tend to focus on increasing human performance. Other differences in perspective arise from the unit of analysis. Those who focus on the individual might explain processes in acquiring concepts, developing expertise or practical intelligence, or transforming beliefs. Those who focus on the system might examine social learning processes, construction of cultural narratives, or forms of knowledge production and change occurring in the group. These different perspectives are not necessarily irreconcilable, but neither do they nest neatly into one another. Clearly, workplace learning is contested terrain filled with fundamental tensions related to what knowledge counts most and who says so. No single model for workplace learning is acceptable in the face of such distinct positions.

Three emerging perspectives described in this article may become particularly important in guiding future research. One is what is here referred to as co-participation/co-emergent views of learning in work, including cultural-historical activity theory, actor network theory, and complexity theory. These approaches work at both micro and macro levels to help analyze the workplace mix of actors, objects, and culture and how forms of knowledge emerge and become adopted in this mix. Another is the area of new textual/literacy practices engendered by globalization, and their influence on people's work identities and knowledge. The third is analysis of power relations, particularly to address issues of disability, race, religion, and sexual orientation where significant workplace stigma and discrimination has been observed but surprisingly little research has been conducted, particularly connecting these issues to workplace learning. Overall, there continues to be great need for in-depth empirical research that traces

what people actually do and think in everyday work activity, and for research methods that can help illuminate the learning that unfolds in everyday work. Of all the ideas currently afloat in adult education literature addressing work issues, these trends seem most likely to influence future perspectives, program design, and pedagogical practice in workplace learning.

See also: Characteristics of Adult Learning; Economic Outcomes of Adult Education and Training; Labor Education; Workplace Learning Frameworks.

Bibliography

- Billett, S. (2001). *Learning in the Workplace: Strategies for Effective Practice*. Sydney: Allen and Unwin.
- Billett, S., Fenwick, T., and Somerville, M. (eds.) (2007). *Work, Learning and Subjectivity*. New York: Springer.
- Chaiklin, S., Hedegaard, M., and Jensen, U. J. (2003). *Activity Theory and Social Practice: Cultural Historical Approaches*. Langelandsgade: Aarhus Press.
- Chappell, C., Rhodes, C., Solomon, N., Tennant, M., and Yates, L. (2003). *Reconstructing the Lifelong Learner: Pedagogy and Identity in Individual, Organisational and Social Change*. London: RoutledgeFalmer.
- Fenwick, T. (2008). Understanding relations of individual-collective learning in work: A review of research. *Management Learning* 39(3), 227–243.
- Mojab, S. and Gorman, R. (2003). Women and consciousness in the learning organization: Emancipation or exploitation? *Adult Education Quarterly* 53(4), 228–241.
- Sawchuk, P. H. (2003). *Adult Learning and Technology in Working-Class Life*. New York: Cambridge University Press.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. New York: Cambridge University Press.
- Boud, D. and Garrick, J. (eds.) (1999). *Understanding Learning at Work*. London: Routledge.
- Bratton, J., Mills, J. H., Pyrch, T., and Sawchuk, P. (2003). *Workplace Learning: A Critical Introduction*. Aurora, ON: Garamond Press.
- Farrell, L. and Fenwick, T. (eds.) (2007). *Educating the Global Workforce*. London: Routledge.
- Fenwick, T. (ed.) (2001). *Socio-Cultural Understandings of Workplace Learning*. San Francisco, CA: Jossey-Bass/Wiley.
- Rainbird, H., Fuller, A., and Munro, A. (2004). *Workplace Learning in Context*. London: Taylor and Francis Group.
- Symes, C. and McIntyre, J. (2002). *Working Knowledge: The New Vocationalism and Higher Education*. New York: Palgrave Macmillan.
- Taylor, J. (2001). *Union Learning: Canadian Labour Education in the Twentieth Century*. Toronto: Thompson.

Further Reading

ADULT EDUCATION – CONCEPTS

Contents

Lifelong Learning

Rewriting the History of Adult Education: The Search for Narrative Structures

Lifelong Learning

J Field, University of Stirling, Stirling, UK

© 2010 Elsevier Ltd. All rights reserved.

Lifelong learning has become a dominant theme of education and training policies across the advanced industrial nations. Besides a wide range of national governments, it is endorsed by a wide range of intergovernmental policy actors, including the Organisation for Economic Co-operation and Development (OECD), the European Commission (EC), the United Nations Educational, Social and Cultural Organisation (UNESCO), the World Bank and the International Labour Organisation (ILO) (Schemmann, 2007). For governments, lifelong learning is an overarching policy framework which offers solutions to a number of common economic and social challenges; globalization and competitiveness often dominate the policy discourse, but promoting lifelong learning is also seen as relevant to social cohesion, demographic change, active citizenship, migrant assimilation, and public health.

Lifelong learning therefore has broad application across a variety of policy domains. It is also widely discussed by educational professionals and by academic researchers. Some claim that lifelong learning is such a broad concept that it has virtually no practical value (Gustavsson, 1995: 92). While its meanings are many and varied, they usually emphasize learning as a ubiquitous process, which takes place throughout the lifespan, and across a variety of life contexts. The recent focus among policymakers, educationalists, and researchers on the ability to learn continuously after the phase of initial education, and across a variety of contexts of which educational institutions are one among many, distinguishes the debate over lifelong learning from more conventional policy discussions of education and training as levers across a range of economic and social policy domains.

From Social Optimism to Economic Survival?

In recent years, lifelong learning has moved steadily toward the center of the policy stage. International governmental

bodies have played a particularly significant role in popularizing lifelong learning as a policy concept. While the term was in occasional use before the mid-1990s, it received huge impetus when the European Commission declared 1996 to be the European Year of Lifelong Learning, an idea first floated in the Commission's White Paper on competitiveness, employment, and growth (Commission of the European Communities, 1994). This context neatly exemplifies the way in which it is economic concerns that dominate policymakers' interest in lifelong learning. Particularly in the older industrial nations, policymakers argue that successful adjustment to a knowledge economy and society requires a highly skilled, knowledgeable, and flexible workforce as a key to sustained national and corporate competitive advantage; individuals equally need to invest continuously in their own competence in order to maintain their employability in an ever-changing labor market. This reflects and is expressed through a policy discourse that is centered on a human-capital approach to social inclusion and economic growth (Borg and Mayo, 2005; Coffield, 1999; Gustavsson, 1995).

This strong economic bias distinguishes the current debate over lifelong learning from earlier policy attempts to promote learning in adult life. Superficially, the idea of lifelong learning closely resembles notions of lifelong education, which were widely discussed in the 1970s. The idea of lifelong education was promoted particularly actively by UNESCO, who in 1972 published *Learning to Be*, a report by an international expert panel chaired by Edgar Faure, a former politician who had served in France as Minister of Education and Prime Minister (Faure, 1972). Faure's report was essentially humanistic in nature, arguing in favor of wider access to higher levels of education and greater support for and recognition of informal and non-formal learning in order to encourage personal fulfilment and development. Faure's report was enormously influential in stimulating debate, and in infusing that debate with an optimistic view of educational innovation and reform (Knoll, 1998). Its core ideas were taken up by the

OECD, which developed a parallel debate over recurrent education, the aim of which was intended to provide governments with practicable means of realizing the overarching goal of lifelong education (OECD, 1973).

In practical terms, the activities undertaken by UNESCO and OECD mainly helped focus policy attention on the educational needs of those who had benefited least from the front-loaded approach to initial education. In industrial nations, this often involved developing educational entitlements for workers, with laws on paid educational leave in a number of countries. In some, there was a broad entitlement to leave for general purposes (as in Sweden, and in state-level laws on *Bildungsurlaub* in Germany); in other cases, educational leave was guaranteed for specific purposes, such as vocational training under the French law on *congé de formation* or British laws on health and safety and workplace representation. Many more countries experienced a growth of adult basic education, with particularly impressive innovations in adult literacy provision and women's basic education.

By the 1990s, a concern with personal development, or worker participation as public policy goals, had not disappeared altogether, but was found far less frequently. Much more common has been a primary concern with lifelong learning as a means of underpinning economic competitiveness and growth. In a globalized economy, where material resources are more or less ubiquitous, skills and knowledge are said to be the only sustainable sources of competitive advantage (Commission of the European Communities, 1994; Reich, 1993; Thurow, 1994). Insofar as policymakers also share an interest in equity and social cohesion, lifelong learning's importance is often valued primarily as a means of re-insertion of vulnerable individuals or inactive workers back into the labor market, leading in turn to improved income and security for individuals from disadvantaged backgrounds.

The current debate over lifelong learning is therefore distinctive in a number of ways. It is characterized by the breadth, and sometimes vagueness, with which the concept is used; it is derived more from the policy domain than from the educational field; and its dominant usage tends to be primarily economic. Yet, more positively, the concept can also be taken to emphasize and recognize the many ways in which people build up new skills and capacities throughout the lifespan and across different life spheres, including workplaces, communities, homes, and voluntary associations. It gives a central place to people's learning, as opposed to education, teaching, and institutions. To use what has become a common abbreviation, current policy concerns are with education that is lifelong, and also life-wide. For these reasons, it has often become more or less synonymous with adult learning; however, it has powerful implications for all phases of the lifespan.

Initial Education as a Platform for Learning through Life

Initial education, including early-years development, is important in its own right. From a lifelong perspective, though, it is additionally important because it provides a platform for learning later in life. A number of commentators argue that family and neighborhood influences in the early years are particularly significant in determining patterns of learning across the lifespan. From this perspective, high-quality education during the earlier years is important primarily because of its role in providing the abilities and motivation to engage effectively in learning later in life (Hargreaves, 2004; Gillies, 2005; OECD, 2004). Sociologically, many of the factors that are associated with adult well-being are already present in the early years. One recent longitudinal analysis of adult learning in Wales demonstrated that most of the factors that affected the probabilities of participation in adult life were present by the time that the child entered primary school for the first time (Gorard *et al.*, 1999).

Education and well-being have often been associated. The idea that education can promote individual well-being indirectly, by improving earnings and promoting social mobility, is an old one; so are notions of education helping to promote the good society by contributing to economic growth and equality of opportunity. Recent debates about the wider benefits of learning have added a new dimension to the relationship, linking education to other facets of individual and collective well-being, such as health (including mental health), security from crime, and political tolerance (Schuller *et al.*, 2004). Through strengthening self-identity, learning is also said to help people develop a sense of authorship over their own biographies and take responsibility for their life choices (Côté, 2004).

Theoretical Perspectives

The 1970s debate over lifelong education was a broad one. While OECD's work on recurrent education was primarily concerned with the balance of resource distribution as between secondary and tertiary education, combined with an interest in worker participation in enterprise management, the work of UNESCO was profoundly influenced by the radical educational thinkers of the 1960s, along with the concerns of liberation theology and Third World development. The dominant voices in the 1990s debate, by contrast, came primarily from writers on globalization and economic change, and were almost entirely based in the economically advanced nations of the West. In a global knowledge economy, these new growth theorists argued, sustainable competitive advantage could only come from an ability to innovate continuously, and in turn this required a highly

skilled and flexible workforce (e.g. Porter, 1990; Reich, 1993). Neo-Schumpeterian concerns with innovation as a basis for economic growth came to be aligned with human-capital perspectives on skills development, as well as with an interest in regional and national innovation capacities. More organizationally focused analyses have tended to emphasize the importance of organizational learning and knowledge management as strategic responses to complexity and change (Smith and Sadler-Smith, 2006).

The dominant theories of lifelong learning, then, tend to be concerned with developing workers' abilities to innovate and respond to change, and therefore contribute to sustained economic growth. Many governments, particularly those led by social democratic or Christian democratic parties, also see lifelong learning as a means of promoting equity and inclusion. Again, this is associated with a strong focus on employability as an important active measure to promote social cohesion, and equity concerns are therefore closely related to economic goals. Finally, this dominant view takes a capitalist economic order as a given; lifelong learning is not seen as a way of changing society, but at most as a way of including the least advantaged in the existing order. Particularly in its most recent phase, which may be conveniently marked by the European Commission's *Memorandum on lifelong learning* of 2000, it is a highly pragmatic concept (Schreiber-Barsch and Zeuner, 2007: 693; Commission of the European Communities, 2000). However, there are also significant critical voices, albeit from a range of differing perspectives.

Some take a broadly radical, anti-globalization stance. Thus the Maltese writers Carmel Borg and Peter Mayo suggest that the primary economic focus of dominant theories is tied to a neoliberal agenda for welfare reform (Borg and Mayo, 2005). Others have asked whether the whole concept is not associated with Western interests, and question whether, at least in its current manifestations, lifelong learning presents opportunities for or is a distraction from adult basic education as a force for development and democratization in the majority of the world (Torres, 2003). Certainly, the current policy climate tends to assume that individual workers must assume at least partial responsibility for ensuring their own employability and invest in new skills in order to maintain their labor-market value. However, this is often accompanied by incentive regimes, which seek to encourage workers to invest in new competences and improve existing skills; in some cases, workers' own organizations have promoted skills improvements as a way of protecting collective security (Payne, 2005). It is also possible to see welfare regimes as themselves bureaucratic and unresponsive to diverse needs; even in adult education, devolution and autonomy may be viewed as a form of privatization, but some will also experience it as emancipatory. In other words, there is no necessary connection between an emphasis on continuous learning and the dismantling of the welfare state, but radical perspectives do

draw attention to both global and local inequalities that are material and structural, and which may be perpetuated by current lifelong learning policies.

Feminist writers have also made a significant contribution to critical debates over lifelong learning. From a feminist perspective, the radical expansion of post-compulsory education since the 1960s has brought rather ambivalent consequences. On the one hand, considerable growth in women's access to higher education has formed part of the remarkable transformation in the role of work in women's biographical trajectories (Spano, 2002). Like many radical writers, feminists tend to be sharply critical of policies and forms of provision that are driven primarily by market forces, though they go beyond the majority of radicals in identifying clear and practicable ways in which education and training might better meet the needs of women (Gouthro, 2005; Burke and Jackson, 2007).

From a feminist perspective, the invisibility of gender in a patriarchal society masks the fact that women face particular barriers to participation in learning, and much provision fails to address the diversities of women's identities; working-class women in particular are trapped in a cycle of low-paid and low-status jobs, whose skill content is barely acknowledged in public discourse about a learning society (Fenwick, 2004; Jackson, 2003). Gouthro goes rather far, suggesting that the language and ideas of lifelong learning represent a major incursion of public policy into the private sphere, as the identification of the homeplace as a site of learning is little more than a colonization of part of the lifeworld that has particular resonance for women (Gouthro, 2005).

A third alternative body of theory derives from post-structuralist and post-modernist writing. In particular, a number of writers have drawn on the thinking of Michel Foucault, the French philosopher/historian, to frame their analyses of power and knowledge and the construction of the learning citizen. Foucault's influence can be particularly seen in studies which treat knowledge as a social practice, governed by relations of power that may be expressed through various classificatory schema and their institutional manifestations. This might be seen as a relatively superficial reading of Foucault's work, and it has been supplemented more recently by studies that take Foucault's radical decentering of the human subject as their starting point. Here, instead of studying learners as agents, the focus is on studying the specific practices that constitute learning, the discourses produced by and producing these practices, and the different subject positions that are made available through these discourses and practices. These subject positions usually include the other, and discourses of nonparticipation and nonlearning are therefore analyzed as processes of othering, so that practices and discourses of lifelong learning always constitute subject positions that are excluded from the dominant framing (Fejes, 2006; Nicoll, 2006).

Finally, a number of writers have explored connections between lifelong learning and sociological theories of reflexive modernization. Ulrich Beck and Anthony Giddens both take human agency as the core of their accounts of late modernity (Beck, 1992; Giddens, 1991). There are distinct parallels between theories of reflexive modernization and core elements of the debate over lifelong learning. Beck and Giddens lead us to explore the socio-cultural forces that are shaping the demand for continuous learning, rather than seeing lifelong learning as an expression of economic forces alone. Their work also draws attention to learning and change in everyday life; people may well be confronting experiences of globalization and technological change, but they are also required to take an active approach to their own biographies, including the ways in which they negotiate intimate relationships and construct identity and social resources (Alheit, 1990; Field, 2006: 68–73).

Institutional Structures for Lifelong Learning

Lifelong learning is a highly complex area for policy, yet its current prominence is largely due to the interest of policy-makers. This paradoxical position reflects the challenges that current economic, social, cultural, and political changes pose to the policy community, particularly in the Western nations, which therefore require new approaches to governance (Field, 2006: 29–43). While policymakers are still able to resort to direct intervention of the traditional kinds, the most important actors in lifelong learning are usually non-governmental – primarily enterprises and individuals, but also trade unions, families, voluntary associations, and neighborhoods. Even within government, lifelong learning policies span the interests of a range of ministries, and a variety of layers from local and regional to national and supra-national. Lifelong learning therefore poses serious challenges of coordination of a range of actors of different kinds, besides bringing risks of unintended consequences. It also poses challenges to many of the existing institutions, particularly those providing opportunities in adult learning.

Although the current debate over lifelong learning has only been underway since the mid-1990s, governments have not had to write policies on a clean sheet of paper. Rather, they have sought to modernize and systematize existing patterns of provision of adult learning, and review existing institutional structures, with a view to raising levels of participation and attainment, usually right across the lifespan but with a strong concentration on learning in and for working life. Comparative researchers have identified a number of variations in post-compulsory education and training structures, in spite of the convergent pressures of globalizing economic forces and the modernization of

education systems. Particular attention has been paid to the roles of three distinct components of the lifelong learning systems:

- systems of transition between initial education and the labor market;
- higher education systems; and
- arrangements for adult education and training.

These components have attracted attention from policy analysts as well as academic researchers (see, e.g., the OECD's thematic reviews (OECD, 2005)).

The three institutional dimensions of national lifelong learning systems differ significantly from each other. The most complex, from both a policy and an analytical perspective, is the adult learning system, which involves a variety of actors and stakeholders, including a wide range of non-government organizations as well as individual citizens. Youth transition systems are only slightly less complex, as well as institutions, which may or may not be publicly funded; the key stakeholders generally include employers and sometimes trade unions, as well as varying degrees of state provision and regulation. Some national studies note that military service may also affect youth transition processes (e.g., Tsai, 1998), and schools are also often influential actors in their own right. Initial education systems at first seem relatively unproblematic from a policy perspective; the major players are usually publicly funded schools and the state itself (though policy implementation is often influenced by teachers, particularly where the latter are able to exercise a significant degree of professional autonomy). However, initial education is often less straightforward than it first appears; particularly in early years, nongovernmental providers are often involved in nursery-level education, and families and communities exercise significant influence over children's cultural capital and social capital.

Green *et al.* (2006) identify three distinct regional models of lifelong learning and the knowledge economy. Two of these – the Anglo-Saxon, neoliberal model, and the continental European, social market model – are relatively well established, and are clearly based on conventional social policy models of welfare regimes. Green and his colleagues add a third, Nordic model, which combines high levels of social cohesion with strong support for economic competitiveness. The Nordic model has recently been subjected to particular scrutiny because of its perceived relative success in combining comparatively equal participation with high overall participation in adult learning (Tuijnman, 2003; Rubenson, 2006; Milana and Desjardins, 2007). Overall participation rates in all the Nordic countries are consistently close to or over 50% of the population of working age (OECD, 2000). Further, Nordic participation rates are high both for job-related adult education and training and for non-job-related learning (Eurobarometer, 2003).

The roots of this pattern have been traced back to the 1960s, when governments and the social partners identified

adult education as a distinct and significant field of policy, linked closely to labor market policy, which itself was geared primarily to securing full employment and industrial consensus (Rubenson, 2006; Milana and Desjardins, 2007). Typically, the Nordic countries have a wide range of providing institutions, including well-established non-statutory providers (such as trade unions) and community-based providers accountable to local government. Rather than seeking to restructure the institutional system, public policy instruments since the 1960s in the Nordic nations have increasingly included targeted-funding measures aimed at engaging disadvantaged groups in the adult education system (Rubenson, 2006).

These measures have had some success in terms of overall participation. Nevertheless, despite high overall participation, and relatively high participation by disadvantaged groups, Milana and Desjardins (2007) note in a systematic review of international survey data that the same broad distribution is found in the Nordic countries as in other nations. The least likely to participate are older workers, those with lower skills levels, unemployed people, migrant workers, and those with weak initial educational qualifications. Nevertheless, on the basis of data from the International Adult Literacy Survey and a survey conducted by the EC in 2003, published by Eurobarometer, they conclude that the Nordic nations have created popular adult education systems that have led to “the attenuation of differences among these otherwise disadvantaged groups”, particularly older adults of working age and less-educated workers (Milana and Desjardins, 2007: 3). They further analyze Eurobarometer data to show that although adults in the Nordic countries reported similar constraints on participation as did respondents in other European Union (EU) member states, the average incidence of the constraints was generally lower in the Nordic countries, and adults in the Nordic countries were more likely to participate even if they faced these constraints (Milana and Desjardins, 2007: 6). Interestingly, this was true for dispositional barriers as well as for more material and institutional constraints.

Milana and Desjardins conclude that public policy has been particularly significant in producing high levels of overall participation, first by maintaining a strong public adult education system, and second by adopting special targeting measures to ensure that an open and broad system of provision is not simply colonized by the already well educated (Milana and Desjardins, 2007: 14–15). In addition, Nordic economies are typically characterized by forms of organizational networking that are likely to promote informal learning. Peter Maskell and his colleagues have demonstrated that high levels of informal exchange of information, techniques, and skills are critical to the competitiveness of Nordic enterprises, particularly those who are affected by high labor costs and low levels of technological development and must therefore compete on grounds of quality and added value (Maskell *et al.*, 1998).

Supplementing various studies of national policy, Michael Schemmann has conducted a detailed systematic analysis of the policies developed by inter and supranational government bodies such as the World Bank, UNESCO, EC, and OECD (Schemmann, 2007). Of these, the EC has been most influential in practice, since it is responsible for implementing policies directly, while UNESCO and the OECD exercise a more indirect influence. Nevertheless, Schemmann traces a number of common themes, as well as marked differences, across these four bodies; above all, he believes that they have established a global lifelong learning discourse with a number of shared reference points. In turn, of course, these common themes reflect the dominance of a neoliberal policy agenda at national level, with governments seeking similar solutions to similar problems. This includes a marked trend toward employer involvement with delivery, in order to promote responsiveness to economic demands, and the adoption of active approaches to labor-market training, particularly through welfare-to-work measures.

Schemmann also notes a pronounced tendency for international governmental bodies to seek to influence national policy by compiling comparative indicators and promoting policy borrowing and transfer, trends that he finds typical of the new governance that is being applied to complex policy areas like lifelong learning (Schemmann, 2007: 246). Both the OECD and the EC publish benchmarking data, compiled on the basis of selected indicators of educational activity; the OECD's publications usually attract high levels of media coverage. In 2003, the EC set its member states the target, by 2010, of at least 12.5% participation in learning by adults aged 25–64, though the Commission has few powers to enforce such targets other than by publicizing the results. The EC has also been charged by the European Parliament with developing a European-qualifications framework covering all areas of lifelong learning. Such developments have led some commentators to question whether there are tendencies toward an international standardization process in adult learning, particularly within the EU (Schreiber-Barsch and Zeuner, 2007: 699–700).

Conclusions

Since the mid-1990s, ideas of lifelong learning have been widely debated in policy and research circles. The idea itself rose to prominence in the mid-1990s when it was embraced by a number of international policy bodies and by several countries. While there were often exaggerated claims both for the novelty of the policies, and for the likely contributions they would make to a whole plethora of economic and social challenges, these policies did indeed mark a shift in policy focus, away from instruction toward learning and away from childhood and youth toward

learning through the life span. This shift reflected policy-makers' preoccupations with the consequences of globalization and rapid economic and technological change, as well as business leaders' recognition of the contribution of upskilling to competitive strategies. However, it also reflected wider sociocultural factors which were also leading to a new emphasis on continuous learning as a way of coping with the demands of everyday life in a risk society.

Conceptually, the idea of lifelong learning appears neatly to parallel influential sociological conceptions of institutionalized reflexivity and risk. The task of lifelong learning, it has been argued, is therefore to enable people to regain a degree of control over their existence, and develop a learning elective biography:

When flexibility constitutes the crucial capacity that work organizations and the unpredictability of life demands, having a stable identity can be a disadvantage. . . Questions such as 'Who am I?' and 'Whom do I want to be?' can become quite haunting existential questions (Glastra *et al.*, 2004: 294).

Others, however, view such a concern with promoting flexibility as potentially damaging and negative to the individual and community, and at worst as collusion with the excesses of globalized capitalism.

Policies for lifelong learning have tended to concentrate on learning in and for working life. Yet, particularly when compared with the innovations of the 1970s debate, most governments have notably shied away from the challenging and difficult issue of policies aimed at increasing the skills and knowledge content of jobs, especially in sectors and regions that rely on low relative labor costs as a basis for competition. Some attention has been paid to the implications for initial schooling as a preparation for learning in later life, as well as to the development of parenting skills, usually for mothers, with a view to raising their capacity for supporting their own children's learning (Gillies, 2005). Relatively little attention has so far been paid to support for learning in later life, even in countries like Scotland where population aging presents acute social and economic challenges. Patterns of participation in adult learning, even among people of working age, tend to mirror existing educational and socioeconomic inequalities. There is therefore a risk that market-led approaches to lifelong learning will simply accentuate and help to entrench the social hierarchy, as the knowledge-poor lag ever further behind in the shift to a knowledge economy. While no policy models have successfully combined uplifts in overall adult participation with a marked impact on inequality, the Nordic societies have been relatively successful in moderating the impact of existing patterns of disadvantage.

Lifelong learning is, then, a rather ambiguous concept which has been used for a range of policy purposes, mainly economic in nature. Yet, it is at heart extremely simple and – from a normative point of view – potentially rather

attractive. The vision of a society where people have broad opportunities to learn across and throughout their lives is an attractive one for many educationalists – particularly those with a background in adult education. More to the point, the broad social and economic trends that have brought lifelong learning to centerstage are not short-term ones. At least in the medium term, then, the debate is likely to continue.

Bibliography

- Alheit, P. (1990). *Alltag und Biographie: Studien zur gesellschaftlichen Konstitution biographischer Perspektiven*. Bremen: Universität Bremen.
- Beck, U. (1992). *Risk Society*. London: Sage.
- Borg, C. and Mayo, P. (2005). The EU memorandum on lifelong learning: Old wine in new bottles? *Globalisation, Societies and Education* 3, 257–278.
- Burke, P. J. and Jackson, S. (2007). *Reconceptualising Lifelong Learning: Feminist Interventions*. London: Routledge.
- Coffield, F. (1999). Introduction: Lifelong learning as a new form of social control? In Coffield, F. (ed.) *Why's the Beer Always Stronger up North? Studies of Lifelong Learning in Europe*, pp 1–12. Bristol: Policy Press.
- Commission of the European Communities (1994). *Competitiveness, Employment, Growth*. Luxembourg: Office for Official Publications.
- Commission of the European Communities (2000). *Commission Staff Working Paper: A Memorandum on Lifelong Learning*. Brussels: Commission of the European Communities.
- Côté, J. (2004). Identity capital, social capital and the wider benefits of learning. *Wider Benefits of Learning Research Centre Conference*. London, 15–16, 2004.
- Faure, E. (1972). *Learning to Be: The World of Education Today and Tomorrow*. Paris: UNESCO.
- Fejes, A. (2006). *Constructing the Adult Learner: A Governmentality Analysis*. Linköping: Linköpings Universitet.
- Fenwick, T. (2004). What happens to the girls? Gender, work and learning in Canada's 'new economy'. *Gender and Education* 16, 169–185.
- Field, J. (2006). *Lifelong Learning and the New Educational Order*, 2nd edn. Stoke on Trent: Trentham.
- Giddens, A. (1991). *Modernity and Self-Identity: Self and Society in the Late Modern Age*. Cambridge: Polity Press.
- Gillies, V. (2005). Meeting parents' needs? Discourses of 'support' and 'inclusion' in family policy. *Critical Social Policy* 25(1), 70–90.
- Glastra, F., Hake, B., and Schedler, P. (2004). Lifelong learning as transitional learning. *Adult Education Quarterly* 54(4), 291–307.
- Gorard, S., Rees, G., and Fevre, R. (1999). Two dimensions of time: The changing social context of lifelong learning. *Studies in the Education of Adults* 31(1), 35–48.
- Gouthro, P. A. (2005). A critical feminist analysis of the homeplace as a learning site: Expanding the discourse of lifelong learning to consider adult women learners. *International Journal of Lifelong Education* 24(1), 5–19.
- Green, A., Preston, J., and Janmaat, J. G. (2006). *Education, Equality and Social Cohesion: A Comparative Analysis*. Basingstoke: Palgrave Macmillan.
- Gustavsson, B. (1995). Lifelong learning reconsidered. In Klasson, M., Manninon, J., Tøsse, S., and Wahlgren, B. (eds.) *Social Change and Adult Education Research*, pp 89–110. Linköping: Linköping University.
- Hargreaves, D. (2004). *Learning for Life: The Foundations of Lifelong Learning*. Bristol: Policy Press.
- Jackson, S. (2003). Lifelong learning: Working-class women and lifelong learning. *Gender and Education* 15, 365–376.
- Knoll, J. (1998). 'Lebenslanges Lernen' und internationale Bildungspolitik – Zur Genese eines Begriffs und dessen nationale Operationalisierungen. In Brödel, R. (ed.) *Lebenslanges Lernen – Lebensbegleitende Bildung*, pp 35–50. Neuwied: Luchterhand.

- Maskell, P., Eskelinen, H., Hannibalsson, I., Malmberg, A., and Vatne, E. (1998). *Competitiveness, Localised Learning and Regional Development: Specialisation and Prosperity in Small Open Economies*. London: Routledge.
- Milana, M. and Desjardins, R. (2007). Enablers and constraints to participation: Has policy in Nordic countries reached its limit for raising participation in adult learning among certain groups? *Second Nordic Conference on Adult Learning*. Linköping University, 17–19, April 2007.
- Nicoll, K. (2006). *Flexibility and Lifelong Learning: Examining the Rhetoric of Education*. London: Routledge.
- OECD (Organisation for Economic Co-operation and Development) (1973). *Recurrent Education: A Strategy for Lifelong Learning*. Paris: OECD.
- OECD (Organisation for Economic Co-operation and Development) (2000). *Thematic Review on Adult Learning: Sweden*. Paris: OECD.
- OECD (Organisation for Economic Co-operation and Development) (2004). *Completing the Foundation for Lifelong Learning: An OECD Survey of Upper Secondary Schools*. Paris: OECD.
- OECD (Organisation for Economic Co-operation and Development) (2005). *Promoting Adult Learning*. Paris: OECD.
- Payne, J. (2005). What progress is Norway making with lifelong learning? A study of the Norwegian competence reform. *SKOPE Research Paper 55*. Oxford and Warwick Universities.
- Porter, M. (1990). *The Competitive Advantage of Nations*. New York: Free Press.
- Reich, R. (1993). *The Work of Nations: Preparing Ourselves for Twenty-First Century Capitalism*. London: Simon and Schuster.
- Rubenson, K. (2006). The Nordic model of lifelong learning. *Compare* **36**, 327–341.
- Schemmann, M. (2007). *Internationale Weiterbildungspolitik und Globalisierung*. Bielefeld: W. Bertelsmann Verlag.
- Schreiber-Barsch, S. and Zeuner, C. (2007). International–supranational–transnational? Lebenslanges Lernen im Spannungsfeld von Bildungsakteuren und Interessen. *Zeitschrift für Pädagogik* **53**, 686–703.
- Schuller, T., Preston, J., Hammond, C., Bassett-Grundy, A., and Bynner, J. (2004). *The Benefits of Learning: The Impact of Education on Health, Family Life and Social Capital*. London: RoutledgeFalmer.
- Smith, P. and Sadler-Smith, E. (2006). *Learning in Organizations: Complexities and Diversities*. Basingstoke: Palgrave Macmillan.
- Spano, A. (2002). Female identities in late modernity. In Chamberlayne, P., Rustin, M., and Wengraf, T. (eds.) *Biography and Social Exclusion in Europe: Experiences and Life Journeys*, pp 151–173. Bristol: Policy Press.
- Thurrow, L. (1994). New game, new rules, new strategies. *Journal of the Royal Society of Arts* **142**, 50–53.
- Torres, R. M. (2003). *Lifelong Learning: A New Momentum and a New Opportunity for Adult Basic Learning and Education (ABLE) in the South*. Bonn: Institute for International Co-operation of the German Adult Education Association.
- Tsai, S. L. (1988). The transition from school to work in Taiwan. In Shavit, Y. and Müller, W. (eds.) *From School to Work*, pp 443–470. Oxford: Clarendon Press.
- Tuijnman, A. C. (2003). The Nordic model of lifelong learning. *International Journal of Educational Research* **39**, 283–291.

Further Reading

- Walther, A. (2006). Regimes of youth transitions. *Young* **14**, 119–139.

Rewriting the History of Adult Education: The Search for Narrative Structures

B J Hake, University of Leiden, Leiden, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In collections of university libraries worldwide, one can locate relatively small sections of volumes pertinent to the history of adult education. Such volumes may include monographs and edited volumes with reference to the broad picture of the history of adult education and specific periods in the development of adult education, studies of particular institutions and organizations at national and local levels, biographies of important individuals, and the celebration of specific institutions in jubilee volumes. Essentially, these volumes will relate to the history of the country where the library is located. A much smaller number of volumes will relate to the history of adult education in other countries, while comparative histories will indeed be relatively scarce items. Even rarer will be volumes in foreign languages about the history of adult education in other countries.

Irrespective of their geographical scope or the period covered, the core question with regard to all histories of adult education involves the structure of the historical narratives which endeavor to tell the story of the development of the institutions and practices regarded as constituting adult education. There are two key issues with regard to the structure of historical narratives, namely: (1) the circumscription of the field of study in terms of the social and cultural phenomena recognized as comprising adult education, and, (2) the delineation of important eras or formative periods in the development of adult education.

Circumscription of the Area of Study

The definitions of adult education used in historical narratives raise fundamental questions with regard to the phenomena which are included in or excluded from historical studies of adult education. National histories tend to be constructed in terms of some sense of an adult education movement which is regarded as a clearly definable social system comprising the individuals, institutions, and associations which have been responsible for the development of adult education through time. This perspective views adult education as a field of readily recognizable activities which constitute the national system of institutional providers which emerged within a sequential process of historical development. These historical

narratives take three major forms. First, they may trace the emergence and development of organizations and practices in the form of institutional histories at national, regional, and local levels. Second, they may be organized around significant individuals who are regarded as the great innovators and reformers in adult education. Third, they may narrate the development of successive periods of philanthropic initiatives, intervention by governments, legislation, and public funding of the provision of adult education. Such narratives are constructed in terms of institutional success which tends to result in an unproblematic narrative of the development of well-known institutions or practices. In this manner, historical studies tend to comprise the selective construction of a lineage for the growth of remembered institutional forms of adult education. These selective historical narratives are often recorded in terms of those forms of adult education which constitute the national tradition of adult education practices. Consciously or unconsciously, these selective accounts of successful and enduring institutions actively exclude the unremembered, the inconvenient, and the historically embarrassing.

Historical narratives of adult education as the story of successful formal institutions, however, actively exclude a vast range of social and cultural phenomena in the more diverse spheres of nonformal and informal adult education. Although historical research must necessarily devote considerable attention to the detailed study of the institutions, significant historical actors, and the development of public policy, historical accounts of the development of adult education must necessarily lead out to the general history of society. Such an approach must lead away from the specific and historically bounded contexts of institutional history into the broader economic, political, social, and cultural history. One of the most striking features of the literature on the history of adult education, for example, is the widespread evidence of the significant contribution made by social, political, and cultural movements to the development of adult education practices. It is necessary, therefore, to identify historical studies which analyze the historical relationships between social movements and the development of adult education in terms of the broader patterns of economic, social, political, and cultural change beyond the institutional realms of adult education.

The broader and more diverse range of social and cultural phenomena identified here as the legitimate

object of study for the historian of adult education can be best understood in terms of the social organization of communication and learning. The term social organization refers here to the complex range of institutions, social movements, and groups which were involved in the historically specific development of adult education. This opens up the field of historical description and analysis of adult education practices in terms of the social organization of communication and learning in which adults were either organized by others or organized themselves for the purposes of disseminating and acquiring knowledge, skills, and sensitivities. Some of these institutions, movements, and groups will be recognized immediately as adult education, while others were embedded in the economic, political, or cultural dimensions of social life. Reconstruction of the ideas, institutions, and practices associated with adult education has, thus, to be pursued in terms of the social relationships involved in the social organization of communication and learning. The historical development of adult education institutions and practices is a socially structured process within societies, and this is a question of dominance and dependence in the history of social, political, economic, and cultural relationships. Innovation is sometimes undertaken by dominant social groups, but it may also be carried out by alternative or oppositional social groups and movements. This results in a more inclusive understanding of the range of nonformal and informal adult education at a distance from adult education institutions.

Periods in the History of Adult Education

Historical narratives are characterized by the marked consistency with which they relate the development of adult educational institutions and practices in terms of a number of specific eras or formative historical periods. The standard institutional histories tend to be formulated in terms of the notion of eras which describe periods of major changes in the institutional development of adult education. Such eras can also be regarded as formative periods of high levels of activity in the development of adult education institutions and practices. In terms of evidence-based empirical indices of activity, such formative periods are characterized, first, by high rates of innovation associated with the development of new institutions and practices; second, the significant expansion in the numbers of adults involved in organized learning activities; third, the opening up of participation in organized learning to new social groups or publics; and, fourth, significant levels of interest in developments taking place in other countries. The latter phenomenon was expressed in reports of visits, translations of foreign texts, and articles in contemporary journals. Formative periods were interspersed, however, with periods with a low conjuncture in

terms of innovation and change. This indicates the need to recognize the historical reality of breaks and shifts in historical development rather than the gradual unbroken line of the development of institutions and practices.

European and Anglo-Saxon literatures, for example, tend to reconstruct the history of adult education in terms of four significant formative periods.

The first of these periods was associated with the Protestant Reformation in northwestern and Central Europe during the second half of the sixteenth century and the first half of the seventeenth century. Development of the organization of adult learning activities in this period was influenced by the invention of the printing press, the translation of the Bible into the vernacular, Bible study groups, and high levels of adult literacy. A reading public emerged which gave rise to the demand for books, new literary forms such as devotional books, books of manners, and the first encyclopedias of knowledge. This was associated with new forms for the distribution of the printed word by colporteurs, booksellers, circulating libraries, and reading circles. In the Catholic areas of Southern Europe, however, the Counter Reformation was marked by the Baroque rejection of the promotion of literacy among the general population and the emphasis upon visual imagery rather than the written word. This geographical division between Northern and Southern Europe was expressed in very different dynamics in the development of adult education. These dynamics subsequently exerted their influence in the different strands of European expansion and colonialism in other continents. They continue to exert their influence in contemporary problems of illiteracy in many countries throughout the world.

From the mid-eighteenth century onward, a second formative period in the development of organized adult education throughout Europe and in the American colonies can be identified. The so-called Enlightenment movement, which marked the start of the European modernization process, was in effect a transnational social and cultural movement which gave priority to education, for both children and adults, in the improvement of society together with an emphasis upon virtue and individual moral behavior in the service of the common good. Historical research provides evidence of the activities of the state and voluntary societies and associations in the development of elementary education; the advancement and diffusion of knowledge; encouragement of the rational improvement of commerce, manufactures, and agriculture; the stimulation of literature, poetry, and drama; the organization of lectures and scientific demonstrations; circulating and lending libraries. The first adult schools were established in this period. The development of an active publishing trade and the further growth and diversification of the reading public was expressed in the periodical press, newspapers, and the growth of a political press.

One of the most manifest consequences of this period throughout Europe and the American colonies was the development of radical political movements, often organized in the form of corresponding societies. These movements demanded democratic rights and freedom of speech on behalf of both the commercial middle class and the artisan class in opposition to the closed oligarchies of dominant regimes. Such radical groups and their adult education activities were frequently repressed through the prohibition of the printing and selling of books and pamphlets in the vernacular. Banned books in French, such as the *Encyclopaedia* by Diderot and d'Alembert, were printed in The Netherlands and smuggled to France, while books in the Greek language were banned by the Ottoman authorities and were secretly imported from the printing presses of Budapest. This phenomenon of underground adult learning has remained a significant dimension of adult education in later periods of repression. The long-term repercussions of the French Revolution in 1789, which in itself fueled radical movements throughout Europe, resulted in the emergence of nation-states in the early nineteenth century. The cultivation of national identity led to the emphasis upon the development of national systems of elementary education and the organization of improving educational activities for adults. This was associated with the need to exert more rigorous control upon the self-organized learning undertaken by the common man, and the need to instruct adults in their rights and duties as responsible citizens of the new nation-states. In the longer term, this formative period, throughout Europe, experienced its nemesis in the revolutions of 1848 and their subsequent repression throughout Europe. This resulted in the emigration of many radicals together with the active export of many forms of self-organized adult education, from German in particular, to the other European countries and the United States.

The period between the 1870s and 1930s has been designated in the literature as a third important formative period. This period was characterized by industrialization and urbanization which contributed to the emergence of the organized working class, a militant women's movement, and the struggle for the right to vote. This resulted, on the one hand, in the development of independent adult educational activities organized by socialist, communist, and anarchist political parties, together with the trade unions and the women's movement. The period witnessed a significant expansion of independent working-class forms of provision such as the workmen's associations, Workers' Educational Association, workers' houses, workers' book clubs, workers' travel associations, Lenin and Marx houses, and the diverse range of educational initiatives associated with the Second Communist International. On the other hand, there was a range of educational responses to this challenge by conservative and liberal parties, together with the hierarchies of the Catholic and

Protestant churches. This resulted in the development of new institutional forms for the provision of adult education such as university extension, university settlements or co-called Toynbee work, the arts and crafts movement, folk houses, popular universities, public libraries, together with the folk high schools in Scandinavia, and other forms of residential education elsewhere. These forms of adult education provision were largely intended to provide educational solutions to the social question of the emergent working class, and they promoted reformist solutions to widespread concerns with urban housing, family life, working conditions, sanitation and health, prostitution, and alcohol abuse. Inherent to this conflict between adult education sponsored by independent working class and middle class was the issue of educating citizens to make use of the extension of the right to vote following World War I. Civic education became a key theme in adult education provision as was clearly demonstrated in the institutionalization of University Extension in the English-speaking world and the development of adult education institutions throughout Europe. These institutions placed the emphasis upon liberal adult education and also focused upon new didactic methods, such as reformist pedagogy in Weimar Germany and elsewhere.

This period also witnessed the development of concern of well-intentioned employers with new forms for the dissemination of scientific and technological knowledge to their employees. From the 1851 Great Exhibition onward, there had been efforts to bring employers and workers together in continuing education, learning in the workplace, and putative forms of vocational education and training. On the one hand, employers sought to establish industrial museums, perhaps badly named, which were intended to make new technological knowledge and production methods available to the working population by way of public demonstrations and short courses. Universities became involved in this process with the development of University Extension services devoted to the needs of agriculture and industry. On the other hand, elementary forms of vocational education and training developed with an emphasis upon technical drawing so that skilled workers could gain insights into the working of new machinery and production processes. At the same time, there was a growing concern with the changing employment patterns of women's participation in paid work. Initiatives by the women's movement in the 1890s were enhanced by the experience of World War I when women took the places in factories of men who were at the front. This often resulted in vigorous debates about the occupations which were appropriate to women and their needs for vocational training beyond the traditional domestic spheres of caring, cooking, sewing, and nursing. Following the Civil War in Finland, for example, there were attempts to retrain women as electricians and plumbers rather than as weavers and seamstresses.

Of particular significance later in this period were the consequences of the Russian Revolution in 1917, the end of World War I, the national independence movements following the peace treaties, and the rise of Fascism and National Socialism during the 1930s. The first All-Soviet conference on adult education was held in 1918 and the keynote address was given by Lenin, who did not fail to name the enemies of the revolution. The carnage caused by World War I resulted in significant interest worldwide in the role of adult education in the promotion of peace and international solidarity. The 1920s witnessed the establishment of numerous international associations, the first world conferences on adult education, and the establishment of international institutions such as the International Peoples' College in Ellsinore. The peace settlements of Versailles and Trianon broke up the territories of the Tsarist Russian, Austro-Hungarian, and Ottoman empires and granted national independence to many countries in the Baltic region, Central Europe, and the Balkans. In these newly independent nations, adult education became a battleground during the 1920s and 1930s between democratic and nationalist political factions. Fascist and National Socialist regimes were responsible during the 1920s and 1930s for the reorganization of adult education in the service of the state in Portugal, Germany, Italy, and Spain, together with other countries in Central Europe and the Balkans. This latter development had fundamental consequences for the organization of adult learning throughout Europe, which included the development of settlements and work camps for the ideological socialization of movement adherents, in particular, the emphasis upon youth movements, together with the so-called re-socialization of recalcitrant radicals such as intellectuals, social democrats, and communists. The first concentration camps established in Germany in 1933 were intended to re-socialize the recalcitrant through hard labor. At the same time, many refugees took their reformist pedagogy with them to other countries and became innovative forces within other adult education systems.

The period between the late 1950s and the present day can be regarded as the fourth and significantly complex formative period which has fundamentally reshaped the organization of adult educational institutions and practices, especially in the global context. On the one hand, the end of World War II led, in the longer term, to the emergence of national independence movements in the remnants of the British and French empires in Africa and Asia. This process of contested decolonization involved the recognition of a new role for adult education in nation building and economic development in the Third World which was driven by United Nations Educational, Scientific and Cultural Organization (UNESCO) and its institutes. A series of world conferences, namely Ellsinore (1949), Montreal (1960), Tokyo (1972), Paris (1985), and

Hamburg (1997) focused on the role of adult education in postcolonial nations. The emphasis upon the importance of literacy for development gave rise to the repression of emancipation movements in many countries in South and Central America. On the other hand, the end of World War II gave rise in Europe to the Soviet hegemony in the Baltic, Balkan, and Central European countries. Adult education was put to work there in the service of the communist revolution and priority, especially in terms of the access of adults to higher education, was given to party members, women, the military, workers, and farmers.

In Western Europe, and elsewhere in the English-speaking world, the late 1960s and the 1970s were marked by the development of compensatory educational opportunities for adults. This involved second-chance and second-way adult education, with an emphasis upon outreach work to the nonparticipants in adult education, which was associated with the development of nonformal and informal community-based forms of adult learning. In addition to the rapid expansion of evening and day institutes for adults during this period, there was a major expansion of distance learning for adults and in particular the establishment of open universities worldwide. Important policy concepts at international and national levels during this period referred to lifelong, permanent, and recurrent education in terms of the redistribution of educational opportunities throughout the life span.

From the mid-1980s onward, however, this largely social-democratic-driven reform agenda to expand educational opportunities for adults was displaced in Europe, indeed worldwide, by the resurgence of neoliberal ideologies. There emerged a renewal of interest in vocational education and training for adults as the core learning message of the global economy. The resurgent interest in lifelong learning has been largely informed by the need to ensure the competitiveness of national economies in the global market, employability of the workforce, the integration of immigrants, demographic change, and the graying of populations in postindustrial societies. This has resulted worldwide in policy narratives which talk in terms of a learning for earning ideology, which is focused on developing the competences required by individuals in order to survive in volatile economic markets. Traditional priorities in adult education, such as the promotion of citizenship and social capital, have been increasingly marginalized. The individualization of learning has increasingly replaced the collective acquisition of knowledge, skills, and attitudes via social movements, community education, and regional development. Individual survival, rather than collective learning to improve the quality of life of communities, now dominates lifelong learning policies. Key policy issues now emphasize individual learning rather than the political economy of promoting structures of opportunity for disadvantaged groups in society.

On the one hand, learning in the workplace has now become the dominant understanding of the development of adult education in the early twenty-first century. The retreat of the state, as the motor of the welfare state and the public responsibility for the redistribution of educational opportunities for adults, has led to the withdrawal of subsidies for many traditional forms of adult education and the privatization of many forms of provision, while the emphasis has shifted toward individual responsibility for investments in adult learning. This now often results in the negation of the very real educational needs of the indigenous unskilled proletariat who is left to fend for itself, or who is encouraged to participate in the commercial learning marketplace and Internet. On the other hand, global mass migration has more recently contributed to the core question of the integration of immigrants and the challenge of Islam as the educational issue in multicultural societies. A largely unresearched area of this multicultural context is the role of the mosque as a learning environment for immigrants in Western societies.

Conclusion

The larger question raised here is the degree to which histories written from Anglo-Saxon perspectives is relevant to the larger canvas of the global dimension of the development of adult education. This raises the almost un-investigated area of the dynamics of empires, colonialism, and postcolonialism in the worldwide development of adult education. The available standard works on the history of adult education have been largely written in terms of selective national histories, which in some small measure examine the colonial tradition of European expansion. Postcolonial understandings of the development of adult education are only now emerging as narratives of resistance in the old empires together with the processes of decolonization, national independence, and indigenous identities. There are many more such examples of adults learning in difficult circumstances, whether above ground or underground, whether in adult education institutes, on the barricades, or in prison. Indeed, the phenomenon of underground adult education, often organized by the learners themselves in the face of oppressive forces remains a recurrent, but inadequately researched, theme in the history of adult education in most countries.

Historical description and explanations of the development of the social phenomena commonly known as adult education need to be more firmly rooted in the conscious use of theories and concepts from the social sciences and cultural studies. This is not an argument for the deconstruction and marginalization of the reform discourses which have dominated historical narratives about the development of institutions and practices

associated with adult education in most countries. It is a critique of the strong element of celebration, the search for genealogies, and the construction of lineages in historical narratives about adult educational institutions and practices. The history of adult education is not real history when its narratives produce collections of the valued national antiques of institutionalized adult education from the past. It is necessary to recognize the complex levels at which critical historical analysis and interpretation can enter the debate and reconstruct dominant narratives. If the history of adult education is about the pioneers and their reputations, it is also about the forgotten and the defeated, even the uncomfortable and inconvenient reminders of the past. If it is about social reformers and formal institutional provision, it is also about social and cultural movements and their contributions to the social organization of nonformal and informal learning. If it is about the latter, it is also about ideologies and struggles between social groups to control communication and learning in the public sphere. If it is about ideologies and struggles in the public sphere, it is also about the formation of publics, popular expectations, and responses. If it is about the latter, it is also about the experiences of the autodidact and the learning biographies of resistance. All these aspects have to be provided with a theoretical perspective in order to achieve more insightful understandings of the construction of historical narratives. The history of adult education in its broader social and cultural contexts still has to be written. This will be the history of the individual and collective learning activities undertaken by adults in order to survive in difficult times and struggles to change society.

Further Reading

- Dobbs, A. E. (1919). *Education and Social Movements, 1700–1850*. London: Longmans, Green.
- Friedenthal-Hasse, M., Hake, B. J., and Marriott, S. (eds.) (1991). *British–Dutch–German Relationships in Adult Education 1880–1930: Studies in the Theory and History of Cross-Cultural Communication in adult education*. Leeds: Leeds Studies in Continuing Education.
- Hake, B. J. and Marriott, S. (eds.) (1992). *Adult Education between Cultures: Encounters and Identities in European Adult Education since 1890*. Leeds: Leeds Studies in Continuing Education.
- Hake, B. J. and Steele, T. (eds.) (1997). *Intellectuals, Activists and Reformers: Studies of Cultural, Social and Educational Reform Movements 1890–1930*. Leeds: Leeds Studies in Continuing Education.
- Hake, B. J., Steele, T., and Tiana, A. (eds.) (1996). *Masters, Missionaries and Militants: Studies of Social Movements and Popular Adult Education 1890–1939*. Leeds: Leeds Studies in Continuing Education.
- Hake, B. J., van Gent, B., and Katus, J. (eds.) (2004). *Adult Education and Globalisation: Past and Present*. Frankfurt: Peter Lang.
- Harrison, J. F. C. (1961). *Learning and Living 1790–1960: A Study in the History of the English Adult Education Movement*. London: Routledge/Kegan Paul.
- Knowles, M. S. (1977). *A History of the Adult Education Movement in the United States*. Huntington: Holt, Rinehart and Winston.

- Marriott, S. and Hake, B. J. (eds.) (1994). *Cultural and Intercultural Experiences in European Adult Education: Essays on Popular and Higher Education since 1890*. Leeds: Leeds Studies in Continuing Education.
- Morgan, W. J. (2003). *Communists on Education and Culture 1848–1948*. Basingstoke: Palgrave Macmillan.
- Paulston, R. G. (ed.) (1980). *Other Dreams, Other Schools: Folk Colleges in Social and Ethnic Movements*. Pittsburgh, PA: University of Pittsburgh.
- Pole, T. (1814). *A History of the Origin and Progress of Adult Schools*. Bristol: Mc Dowall Schools.
- Schama, S. (1977). *Patriots and Liberators: Revolution in The Netherlands, 1780–1813*. New York: Knopf.
- Simon, B. (ed.) (1990). *The Search for Enlightenment: The Working Class and Adult Education in the Twentieth Century*. London: Lawrence and Wishart.
- Steele, T. (2008). *Knowledge is Power! The Rise and Fall of European Popular Educational Movements*. Frankfurt: Peter Lang.
- Thompson, E. P. (1963). *The Making of the English Working-Class*. London: Gollancz.
- Williams, R. (1981). *Problems in Materialism and Culture*. Oxford: Oxford UP.

ADULT EDUCATION – DOMAINS AND PROVISION

Contents

Adult Basic Education: A Challenge for Vocational Based Learning
Adult Literacy Education
Citizenship and Immigrant Education
Community Based Adult Education
Continuing Professional Education: Multiple Stakeholders and Agendas
Financing of Adult and Lifelong Learning
Labor Education
Learning Cities and Regions
Museums as Sites of Adult Learning
Overview of Lifelong Learning Policies and Systems
Popular Adult Education
Provision of Prior Learning Assessment
The Age of Learning: Seniors Learning
University Adult Continuing Education: The Extra-Mural Tradition Revisited

Adult Basic Education: A Challenge for Vocational Based Learning

R Höghjelm, Stockholm University, Stockholm, Sweden

© 2010 Elsevier Ltd. All rights reserved.

Two recent European Community policy documents on adult education raise the need for a changing role of adult basic and vocational based learning (VBL) in the context of the evolving knowledge society and knowledge economy (EC, 2006, 2007). Further, the Council of Europe has, in 2009, adopted a new strategic framework for cooperation among European Union (EU) member states to reform their education and training systems so as to better prepare people to find jobs and to help businesses find the staff they need to succeed and innovate in the face of global competition. Similar discussions can be found, for example, in the USA, Canada, and Mexico, where one is looking for a closer connection between traditional adult basic education (ABE) and workplace learning. In this article, we analyze the new demands from society on ABE. While the discussion primarily draws on what is taking place in Sweden, the underlying issues and potential developments are of universal interest.

National Developments in ABE

During the 1960s, ABE was primarily an equality issue in developed countries. However, the emergence of a knowledge economy – with its demand of a more qualified workforce – has affected the role of ABE.

Looking at the development of ABE in the USA, it should be noted that it is an umbrella term used to describe a range of educational services for adults – from basic literacy (including English as a foreign language) and numeracy to high school equivalency (general educational development (GED))/adult diploma programs (ADPs). The program is federal – originating from the Economic Opportunity Act (20 August 1964), which created the first ABE program as a state grant. During the 1960s and 1970s, several studies focused on the ability of the program to fulfill its goals (see, e.g., Firoza, 1966 and the classical study by Mezirow *et al.*, 1975). Thereafter, the program dropped into the background. However, with changes taking place in the economy and the demand for a better-qualified workforce, it once again came into the focus of policy-makers and the Adult Education Act was repealed and replaced with the Workforce Investment Act, in 1998. In this change, the focus gradually moved from literacy to vocational education and training (VET).

In Australia, where ABE has its roots in the British-extension movement, the program has traditionally been quite broad – encompassing literacy, numeracy, communication skills, basic science, humanities, and social sciences up to the equivalent of year 10 of compulsory schooling. Further, it includes survival skills linked to personal health, social action, problem solving, and conflict resolution. With vocational training and skills

formation having become the Australian government's highest priority, there is now a strong focus on using ABE to upgrade the skills of the existing workforce. This has resulted in an emphasis on narrower vocational training over the traditionally broader ABE. Australian adult educators have strong reservations with regard to the narrow focus on job skills and the vocational/nonvocational distinction underlying recent policies.

In Finland, the 1975 Adult Education Committee argued for an ABE characterized by a close connection between liberal adult education and vocational education. It is worth noting that, at the time, over 60% of the workforce in Finland lacked a vocational education. During the following decades, a number of reforms were undertaken by the Finnish government to implement the Committee's intentions and preserve the tradition of a close connection between general adult education and vocational education.

Two things stand out from this brief overview. First, it is evident that changing skill requirements have raised demands for ABE to be more closely aligned with VET. Second, different models are being developed where one is characterized by a narrow focus on training, which – as with the Finnish example – tries to combine liberal and vocational traditions. To further explore this tension, I take a closer look at the development of ABE in Sweden with a focus on the link between labor-market demands and the nature of ABE.

The Link Between Education and Work in the Original Swedish Model of Komvux

While the early groundwork for establishing comprehensive schools was being laid, the issue of the older generation – who in their youth did not have the opportunity to continue beyond 6 or 7 years of education – was raised. The school reforms of the 1950s and early 1960s led to a rapid expansion of the Swedish educational system. The result of this expansion was an ever-widening gap between the older generation – who received a minimal education – and the new generation – who benefited from 9 years of compulsory education and increasingly chose to continue on to secondary school. As a consequence, the argument that the people who paid for the increase in primary and secondary education should have their share of the growing educational resources grew stronger. However, it was not only the rights and demands of the older generation but also contemporary human capital ideology that were behind the introduction of a municipal adult education (Komvux), which would offer education equivalent to that offered by primary and schools with the purpose of providing a platform for future studies or working life.

When looking at the target group of the new municipal ABE, it is important to note that the 1967 adult education

reform had its roots in an elitist concept of equality. The basic idea informing this strategy was that everyone should have an equal right to an education irrespective of social background, gender, or place of residence, and the mandate for ABE was to offer it to those who aspired to it and were able to benefit from such an education. Traditional evening-class students served as models for the target group of the newly introduced municipal adult education. These students belonged to the so-called pool of talent – they had a high level of aspiration, were motivated to study, and were often successful in their self-tuition. This was a very different target group than what has been traditionally associated with ABE. Looking at the link to labor markets, it is evident that what was being offered was an adapted form of what was regularly offered in comprehensive and secondary education and that no adjustments were made to make the program specifically vocation oriented. The economic benefit was seen coming from having more people with a solid basic education that would allow them to pursue postsecondary studies. The only direct reference to employment was that Komvux courses were mainly available as part-time studies so as to avoid production losses.

Komvux Economic Crises and Changing Demands on Competence Development

During the economic crises of the early 1990s, the debate on Komvux shifted from a focus on education discourse to its role in a labor-market strategy. The response was to introduce the Adult Education Initiative (Sweden Government Bill 1995/97: 222). It was not introduced as an educational bill, but rather as a cornerstone in a bill titled 'Special strategies introduced in order to half unemployment by year 2000.' The adult education initiative (AEI) was a massive 5-year program for adult education in which all municipalities participated. The project comprised some 110 000 new educational places per year for adults, mainly in municipal adult education. Over the 5 years, it aimed to reach 550 000 adults – roughly 15% of the labor force. The AEI signaled a fundamental broadening of the Swedish tradition of active labor-market policy. Instead of expanding traditional labor-market training programs, the AEI attempted to raise the general level of education in unemployed adults. Another goal of the AEI was that it would act as a vehicle for reforming the adult-education sector both in terms of content and working methods. Over the 5-year period, adult education was reformed and developed so as to better meet the challenges that the individual, working life, and society would face in the new millennium.

Two things stand out in the AEI strategy: the first is the use of general education as a labor-market strategy; and the second is that the content and teaching methods of

ABE in Sweden more directly began to be considered in the context of labor-market needs. Further, it is worth noting that, in view of the apparent danger of having a growing cadre of unemployed facing increasing obstacles in getting back into the labor market, the AEI contained several measures aimed at helping to reach adults who traditionally do not participate in adult education and training. A special education grant was introduced at the same time as the AEI, which was primarily intended for unemployed persons who had not completed a 3-year upper secondary program. Further, in order to reach persons with little or no experience of adult education and help them to start studying, more targeted recruitment and information activities were put in place. In a broader international perspective, the AEI that was replicated in other Nordic countries is of interest, as it illustrates an ABE strategy that combines an economic agenda with strong equality ambitions.

The new demand for continuing VET also set off a broad debate in Sweden – as in other countries – on how to develop new models of ABE that would be appropriate in the emergent knowledge economy.

New Forms of VBL in Formal Adult Education

Many providers of VET in EU countries are using a system in which theoretical and practical activities are mixed. A criticism has been that they are too often organized from a school perspective, focusing on how students' workplace activities match theoretical courses provided by the school. The transition from knowledge produced at school to the work situation has been recognized as a major problem. However, as discussed in several recent studies, this is a misconception. In this context, it is of interest to look a bit closer at the empirical findings from the so-called People Project funded by the European Social Fund, where unemployed adults participated in VET courses at the adult secondary school level. An alternative pedagogy was developed based on the idea that the foundation of a vocation is rooted in vocational culture, vocational praxis, and vocational knowledge. Traditional school-based VET does not recognize dimensions that are linked to each other like a Chinese box, see **Figure 1**.

The study reveals that when organizing VBL, it is preferable to consider this picture. This conclusion is

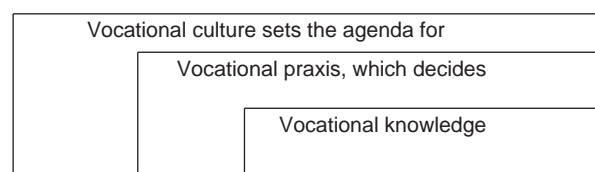


Figure 1 The Chinese Box of VET.

based on the fact that the People Project was very successful in two respects – first, because the unemployed participants managed to find jobs, and, second, because the training provided a foundation for lifelong learning (Höghelm, 2005).

The root of the problem of traditional vocational education within the school system is the existence of two different cultures, each with its own logic. Within school culture, subjects should be organized in a sequencing way – from the simple to the complex – and based on this logic, skills are supposed to be generated from basic to more specific. Working life has a different logic, with a number of occupations at the same workplace constituting a community of practice.

Figure 2 shows how the two different logics emanate from the two cultures and cause problems for the participant. In this context, the vocational adult teacher can be regarded as a broker, developing new connections across communities of practice, facilitating coordination and opening possibilities for new meanings. Vocational adult-educational students are in a unique position to act as mediators, bringing insights from work experience to school and vice versa. The vocational adult teacher has an important role in encouraging the development of this process.

The figure shows the sequential logic of school subjects illustrated as ladders, while working life has a more coherent structure – where acquiring vocational knowledge is characterized as a movement from the periphery toward the center. The concept of VBL relies on the idea of situated learning – a well-known concept since the beginning of the 1990s (Lave and Wenger, 1991; Wenger, 1998). Lave and Wenger have managed to phrase, in an attractive way, an old idea that has always been present within VET – going back to medieval times when the apprentice system was established in Europe. They also elucidated how modern educational systems and VET, in particular, have an ongoing discussion on the difference between practice and theory. Thus, the VBL concept has a socio-cultural approach going back to the ideas of Vygotsky (1987).

Traditional VET programs have a school perspective – implying students' workplace experiences can be organized to fit in school activities. VBL has a reversed perspective, that is, programs are organized in such a way that students' workplace experiences guide when theoretical school subjects are supposed to enter. This different way of organizing VET adult programs will also facilitate the vocational adult participant's boundary crossing between school and work (Tuomi-Grön and Engeström, 2003). Further on, vocational based adult learning can be a more powerful tool for organizing VET compared to a more traditional, organized, apprentice-based education. Applying the VBL concept implies that adult students must spend half of their time at a workplace, visited by teachers.

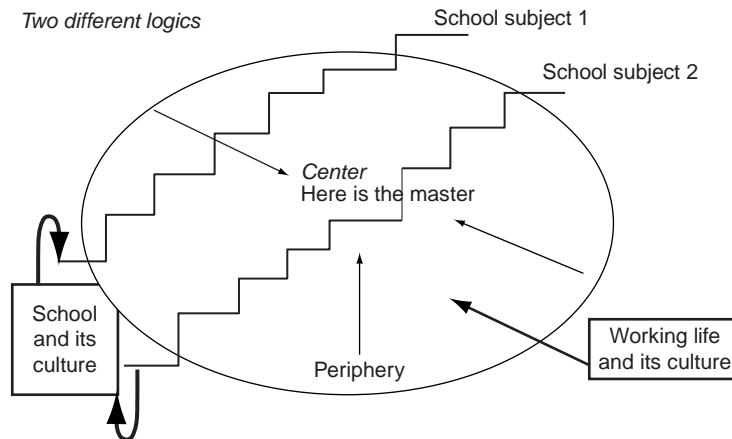


Figure 2 The relation between the two different logics emanated from school culture and working-life culture.

Concerning the demand on the European workforce to increase competitiveness and the ongoing demographic change within the EU, research of the kind discussed here suggests a need to develop a new way of organizing VET – which makes it necessary to relinquish the predominant, but obsolete, approach. By using a VBL approach, there will be a quicker and more obvious connection between practice and theory. The concept has similarities to the professional education of nurses and medical doctors practiced at the McMaster University medical school in Canada since 1967. Between 2005 and 2007, a major European Science Foundation (ESF) project in Sweden used the experiences from the municipality of Söderhamn, and VBL was applied in 29 municipalities. The study also involved 275 companies, 197 teachers, and a large number of principals (Höghi, 2005, 2009).

Future Directions of ABE

ABE – as it is being practiced and organized around the world today – is still heavily influenced by how it was originally set up in the 1960s. While there is an impetus in most countries for greater skills-oriented training, these programs have not adapted to the changing realities. First, demographics have changed and educational attainment has risen dramatically in the population. Second, educational attainment plays an increasingly important role for entering the labor market. Third, there is a rising demand for continuing VET. Fourth, there has been – as discussed above – a shift in the very understanding of VET and a movement to think of vocational education in terms of VBL. Consequently, there are growing demands for a major overhaul of ABE as we have known it.

Starting from the latter, one can think of a form of ABE that takes the individual and his/her vocational background as its point of departure. Traditionally, adult education theory has been considered as a basic guideline for

educational planners and trainers wanting to pursue a reflective approach to adult learning. Based on recent research on VBL, this seems to be too narrow a perspective if the ambition is to include the learner in the process of reflection. Instead of a general curriculum, there is a need for an individual study plan for each participant. In developing this plan, one has to take into account how the individual has been trained in order to manage both previous and new-found knowledge through a validating process that takes into account both general and personal skills from a VBL perspective. (For an interesting discussion of different models of validation see OECD (2005) report *Promoting Adult Learning*.) The purpose of this form of ABE is to allow participants to have the opportunity to develop specific skills linked to civil and working life as well as for pursuing their own educational/vocational project. The dialog with others makes the learner test out new assumptions, understandings, and perspectives (see Mezirow, 2000). This form of VBL would avoid the dilemma of the narrowness of the skills-driven training that is being promoted in some countries as well as the general education lacking vocational connections that can be found in other countries. A development in the direction of VBL can take place within a network organization. The location of such an activity would not be built around a traditional adult-education school such as Komvux, but could be a flexible learning center such as Centre for Flexible Learning (CFL) Söderhamn (Ekelöf, 2009). This form of structure has an opportunity to meet the dual and interconnected needs for a de-schooling version and a schooling-up version of ABE. Examples of this form of organization of ABE can be drawn from different regions in Sweden where municipalities have organized themselves into flexible learning centers. The slogan in these contexts is flexibility or flexible learning – where providers organize learning centers or networks and plan and implement teaching with the primary purpose of supporting student communication and learning (Holmberg, 2004).

Such a development could suggest a return in ABE to the andragogical ideals that were in the foreground of the initial ABE debate.

See also: Adult Learning and Instruction: Transformative Learning Perspectives; Adult Literacy Education; Organizational Learning; Provision of Prior Learning Assessment; Workplace Learning Frameworks.

Bibliography

- EC (European Commission) (2006). Communication from the Commission. Adult learning: It is never too late to learn. Brussels. <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2006:0614:FIN:EN:PDF> (accessed August 2009).
- EC (European Commission) (2007). Action Plan on Adult Learning. It is always a good time to learn. Brussels. http://ec.europa.eu/education/policies/adult/com558_en.pdf (accessed August 2009).
- Ekelöf, E. (ed.) (2009). *Yrkesutbildning – en framtida arena för vuxnas lärande. Anföranden och dokumentation från en konferens vid Stockholms universitet 28 – 29 oktober 2008* (Vocational education and training – a future arena for adult's learning. Contributions in Swedish and English at a conference October 28–29, Stockholm University, 2008.) Ljusdal: ab romi tryckeri.
- Firoza, A. (ed.) (1966). Adult Education Association of the U.S.A. Adult basic education study 1965–1966. ERIC # ED023032.
- Höghliem, R. (2005). *Yrkesbaserat lärande. Erfarenheter från PEOPLE delprojekt i Söderhamn 2002 till 2005*. (Vocational based learning. Experiences from PEOPLE subproject in Söderhamn 2002 to 2005.) CFL Söderhamn. Ljudal: ab romi tryck. <http://www.ntglar.se> (accessed August 2009).
- Höghliem, R. (2009). En ny yrkesutbildning i skärningspunkten mellan två kulturer. (Vocational education and training a crossroad of two cultures.) In Ekelöf, E. (ed.) *Yrkesutbildning – en framtida arena för vuxnas lärande. Anföranden och dokumentation från en konferens vid Stockholms universitet 28–29 oktober 2008*, pp 59–76. (Vocational education and training – a future arena for adult's learning. Contributions in Swedish and English at a conference October 28–29, Stockholm University, 2008.) Ljusdal: ab romi tryckeri.
- Holmberg, C. (2004). På den studerandes villkor – flexibelt lärande. Sid. i. *Utsikter och Insikter. En antologi om flexibelt lärande i vuxenutbildningen (2004)*, pp 67–81. (On the participant's conditions – flexible learning. In perspectives and insights. An Antologi about flexible learning in adult education, pp 67–81.) Nationellt centrum för flexibelt lärande; Rapport 5: 2004.
- Lave, J. and Wenger, E. (1991). *Situated Learning. Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Mezirow, J. (2000). Learning to think like an adult: Core concepts of transformation theory. In Mezirow, J., Kegan, R., Belenkey, M. F., et al. (eds.) *Learning as Transformation*, pp 3–33. San Francisco, CA: Jossey-Bass.
- Mezirow, J., Darkenwald, G. G., and Knox, A. (1975). *Last Gamble on Education: Dynamics of Adult Basic Education*. Washington, DC: Adult Association of the USA.

- OECD (2005). Promoting Adult Learning. <http://www.oecd.org/edu/adultlearning> (accessed August 2009).
- Sweden Government Bill (1995/97). 222.
- Tuomi-Grön, T. and Engeström, Y. (eds.) (2003). *Between School and Work. New Perspectives on Transfer and Boundary-Crossing*. Pergamon: Earli, European Association for Learning and Instruction.
- Vygotsky, L. S. (1987). *Mind in Society*. Cambridge, MA: Harvard University Press.
- Wenger, E. (1998). *Communities of Practice. Learning, Meaning and Identity*. Cambridge: Cambridge University Press.

Further Reading

- Bron, A. and Scemmann, M. (eds.) (2003). *Knowledge Society, Information Society and Adult Education. Buchum Studies in International Adult Education*, vol. 3. Munster: LitVerlag.
- Cole, M., Engström, Y., and Vasquez, O. (2004). *Mind, Culture and Activity: Seminal Papers from the Laboratory of Comparative Human Cognition*. Cambridge: Cambridge University Press.
- EAEA (European Association for Education of Adults) (2009). *Adult education in the world*. <http://www.eaea.org/se/index.php?k=12086> (accessed August 2009).
- Engeström, Y. (1999). Activity theory and individual and social transformation. In Engeström, Y., Miettinen, R., and Punamäki, R. -L. (eds.) *Perspectives on Activity Theory*, pp 19–38. Cambridge: Cambridge University Press.
- Engeström, Y. (2005). Cultural History Activity Theory. <http://www.edu.helsinki.fi/activity/pages/chatandddwr/chat> (accessed August 2009).
- Gordon, H. R. D. (2003). *The History and Growth of Vocational Education in America*. Long Grove, IL: Waveland Press.
- Lasonen, J. and Young, M. (eds.) (1997). *Strategies for Achieving Parity of Esteem in European Upper Secondary Education*. European Commission Leonardo da Vinci Programme: Surveys and Analyses. *Post-16 Strategies Project*. Jyväskylä: Institute for Educational Research, University of Jyväskylä.
- Manninen, J., Burman, A., Koivunen, A., et al. (2007). *Environments that Support Learning. An Introduction to Learning Environments Approach*. Finnish National Board of Education. Vammala: Vammalan Kirjapaino Oy.
- Nijhof, W. J. and Streumer, J. N. (eds.) (1998). *Key Qualification in Work and Education*. Dordrecht: Kluwer Academic Publisher.
- Schutze, H. G. and Istance, D. (1987). *Recurrent Education Revisited. Modes of Participation and Financing: Report Prepared for CERl and OECD*. Stockholm: Almqvist and Wiksell Internationell.
- Tuijman, A. and van der Kamp, M. (1992). *Learning Across the Lifespan. Theories, Research, Policies*. Oxford: Pergamon Press.

Relevant Website

- <http://www.esrea.org> – ESREA (European Society for the Research on the Education of Adults).

Adult Literacy Education

L Tett, University of Edinburgh, Edinburgh, UK

R St.Clair, University of Glasgow, Glasgow, UK

© 2010 Elsevier Ltd. All rights reserved.

The last few years have seen growing interest in the field of adult literacy education, also known as adult basic education and adult literacy and numeracy, with increased attention at national and international levels. This has been partly inspired by the International Adult Literacy Survey of the mid-1990s (and the less influential Adult Literacy and Life Skills Survey (ALL) of the following decade), which allowed adults' skills to be compared across countries for the first time. This coincided with the move toward an information economy, allegedly making information management skills such as literacy the bedrock for success. At the same time notions of human capital (where education is the fundamental key to prosperity) were gaining favor with agencies such as the World Bank and UNESCO (Wickens and Sandlin, 2007) so adult literacy education became seen as a central and critical educational sector.

Adult literacy education is marked by a high level of diversity in terms of structure, delivery, and philosophy and performs different roles in different parts of the world, whether industrialized Europe or a developing country. It can take place in church basements, further education colleges, universities, community settings, workplaces, and libraries. It can be delivered by professionally qualified staff, or staff who are qualified in other forms of teaching or unqualified volunteers. Learners can be employed or unemployed, men or women, refugees or indigenous people, full time students, or people who study part-time. Each of these presents a unique context for literacy education, and it is important not to generalize across settings without taking great care.

What Is Adult Literacy?

There are a number of ways of conceptualizing what is meant by adult literacy. These definitions contain assumptions that matter to the focus of education because they imply different understandings of learning (Papen, 2005). Three concepts have been particularly influential: the functional, critical, and liberal concepts of literacy (Table 1).

Functional Literacy

In this view literacy is seen as a skill that is required for a broad range of activities associated with the individual's

participation in society. There is an assumed correlation between individual skills and the overall performance of the nation in terms of modernization and economic productivity. This is particularly so in the OECD (1997, 2000), where a focus on improving literacy skills as the key to unlocking the benefits of globalization is dominant. Literacy is conceived as a set of neutral, technical skills with little to do with culture and society. Its assumed benefits are believed to include enabling access to information, developing thinking, and improving the individual's chances of finding employment and income. Reflecting this view, the ALL assessed skills against a suitable minimum for meeting the demands of daily work and life (UNESCO, 2005). The functional model emphasizes individual deficits and sees literacy as a set of discrete skills believed to be universal and transferable to all kinds of situations that require the use of written language (Barton, 1994).

Critical Literacy

The concept of critical literacy is associated with the Brazilian educator Paulo Freire and refers to the potential of literacy for not only reading the word but also reading the world (Freire and Macedo, 1987). It moves away from the functional model, toward a pedagogy intended to allow participants to understand their world in terms of justice and injustice, power and oppression, and how to transform it. Contrary to the functional model, primary purpose of critical literacy is not to help the individual to move up the existing social ladder, but to build a radical critique of the dominant culture and the existing power relationships between social groups (Shor, 1993). This model is often linked to democratic citizenship and the role that education plays in supporting people's participation in society (Crowther and Tett, 2001). People need the ability not only to decode the literal meanings of texts, but also to read between the lines and to engage in a critical discussion of the positions a text supports.

Liberal Tradition of Literacy

The third view of literacy is informed by a humanist view of education that emphasizes personal development and individual goals. It argues for the right of all citizens to education and goes beyond the functional-skills approach to include areas such as creative writing and access to

Table 1 Three views of literacy education

	<i>Functional</i>	<i>Critical</i>	<i>Liberal</i>
Reason for literacy education	Skills Survival/work	Understanding Empowerment	Tools Development of person
Participant group	(Potential) workers	Marginalized groups	Everybody

literature (Papen, 2005). Participants in programs are not limited to the working population but include older people or those who are not part of the workforce.

These different definitions present competing ideologies of literacy with associated assumptions, values, and standards that need to be questioned. However, in much of the world there is an unquestioning emphasis on the functional, vocational approach such as Welfare to Work in the US (Sandlin & Cervero, 2003), resulting in a discourse of literacy as a technical skill and vocational competence.

Social-Practice and Skills Models of Literacy

In addition to diverging perspectives on the purpose of adult literacy education, there are a number of theoretical positions on how people actually use literacies. A functional-skills-based approach focuses attention on the autonomy of the text and the meanings it carries. It searches for universal features of adult literacy and other semiotic sign systems. It leads to narrow definitions of reading, writing, and calculating, and ignores aspects of learning that cannot be dealt with at the individual or cognitive level. It excludes many issues that are important for understanding learner responses. All too often it can support a deficit view of literacy, where those with limited literacy engagement are seen to be lacking in some way, whether in ability or in education.

One approach has moved away from the individually focused cognitive skills model to include the social practices associated with number, reading, and writing (Hamilton *et al.*, 2006). In this view literacy is not seen as a purely individual activity – instead, it sees literacy and numeracy as being historically and socially situated and part of wider cultural and media engagement. The focus of the social-practices approach shifts away from literacy as something learners lack toward the many different ways that people engage with literacy. Social-practices approaches recognize difference and diversity, and challenge how these differences are valued within our society.

Street (1995) describes this as a shift from seeing literacy as an autonomous gift to be given to people to an ideological view of literacy that places it in the wider

context of institutional purposes and power relationships. From this perspective adult literacy is part of a range of social practices that are observable in events or moments and are patterned by social institutions and power relationships. Attention is focused on the cultural practices within which written and spoken words are embedded. Not just reading but also speaking and writing, as well as the use of new technologies, become central to the definition of literacy. The social-practices view requires that connections are made between the classroom and the community in which learners lead their lives; with a notion of situated learning; between learning and institutional power; and between print literacy and other media.

There is not just one social-practices theory of adult literacy, numeracy, and language, but a number of different versions. The social-practice approach that has characterized the new literacy studies (NLS) draws mainly on ideas and methodologies from sociology, sociolinguistics, and anthropology rather than the more psychological approach of active problem-solving theory rooted in the work of Vygotsky and others. The NLS involves looking beyond formal educational settings to informal learning, and to the other official settings in which literacies play a key role. Learning does not just take place in classrooms but in everyday life, with meanings, values, and purposes located within a broader literacy framework than the texts themselves.

There are two important principles underlying the implementation of a social-practice approach to literacy. First, a two-way dialog and movement between formal learning and the everyday world is essential. Everyday, situated cultures and practices cannot simply be acknowledged and imported into classroom settings. The boundaries between in and out of education must be blurred so that contexts become permeable.

Second, active learning is assumed by this approach. It characterizes the process of becoming literate as one of taking hold of the tools of writing and language. This has important implications for relationships within the learning process and for reflective and questioning activity on the part of both learners and teachers (Hamilton *et al.*, 2006). The ways in which teachers and learners participate in decision making and the governance of the organization in which learning takes place are crucial, whether through management committees, consultative bodies, and research and development activities. Citizenship is modeled and enacted within such arenas.

Reconciling the Skills and Social-Practices Perspectives

The social-practices approach recognizes the importance of learners' motivations, goals, and purposes; every literacy task is done for a reason and in specific contexts, hence the challenge to concepts of universal sets of literacy skills.

Skills and knowledge acquisition are, however, intrinsic to learners' purposes and enhance many different aspects of their lives. For example, improving skills for employment may not appear to serve social practices, but skills that are gained in the pursuit of employment or promotion can be applied in other domains of people's lives, such as helping children with homework, managing the household, or pursuing further learning. Both enhancing skills and recognizing their role within learners' lives are important and both aspects should be developed in good teaching.

How far might it be possible to reconcile the functional-skills approach and the social-practices approach within policy and practice? Could social practices be seen as encompassing and extending the narrower focus of skills? The idea of two opposing broad approaches is an oversimplification and there are other ways of characterizing the guiding philosophies people bring to literacy, particularly in everyday cultural settings (see Barton *et al.*, 2000). Freebody and Lo Bianco suggest (1997: 26) that effective literacy tuition draws on a repertoire of resources that allow learners to: break the code; actively interpret the meaning of the text; use texts functionally; analyse texts critically. This is a dynamic process as represented in **Figure 1** that is an attempt to acknowledge that both skills and critical practices are enmeshed in working with texts.

In the middle circle is the process of actually understanding the words as they are written on the page, and interpreting the meaning. The outer ring represents the social uses of that meaning, which can range from functional to critical. A literacy process that is missing any of these components can be considered as only a partial engagement with the text.

Research in the US is also providing new insights on the interrelations between skills and practices. The 5-year Longitudinal Study of Adult Learning (LSAL) in Portland, Oregon has revealed that both program participation and

self-study have positive, time-specific effects on literacy practices (Reder, 2008). The research showed that self-study was prevalent among adults of all literacy levels as a means of basic skills development, whether or not they also participated in classes. Self-study appears to act as a bridge between periods of program participation and to facilitate persistence. The mixed mode of learning identified by LSAL seems to bridge social-practices and skills-based approaches. It suggests that learners use a range of resources to enhance the social practices associated with literacy, and that programs are one resource, with the specific role of providing skills to underpin the practices. As we suggest with the diagram above, skills and practices form a self-reinforcing cycle of engagement with literacy and literacy education.

The broad mode of participation suggested by LSAL brings together social-practice and skills approaches. On the one hand, it recognizes that learning involves learners actively using resources as well as programs delivering services. On the other, it indicates that literacy programs appear to have the most direct and immediate impact on literacy practices, underlining the role of skills enhancement.

The Role of Adult Literacy Education

Throughout the world, adult literacy education fulfills a variety of roles. For those in industrialized countries, one common perception is that adult literacy learners are people who have not fully benefited from compulsory education. There are a number of possible reasons, ranging from sociological explanations concerning the tendency of schools to push out certain learners to psychological rationales involving learning difficulties. Overall, the common factor is the view that adult literacy education has an ameliorative role, improving literacy engagement and compensating, to some degree, for the failure of initial schooling (St. Clair and Priestman, 1997).

The ameliorative view assumes that learners have had an opportunity to learn literacy practices, and that this opportunity has not been effective. This can lead to a deficit view of learners, where they are assumed to have some kind of problem that has led to reduced literacy abilities.

One of the reasons that adult literacy education has experienced such variability in funding and policy interest is that it can be viewed as an optional form of provision within the ameliorative perspective. After all, if people have already had a chance to learn about literacy surely giving them a second chance is an act of generosity? It follows that the most effective argument for supporting literacy education is often a moral one. This can lead to a panic about literacy education (or more often illiteracy) with dramatically increased funding followed by gradual withdrawal of support until the next moral panic (Quigley, 1997). Ameliorative perspectives can be unhelpful for the general health and stability of the field.

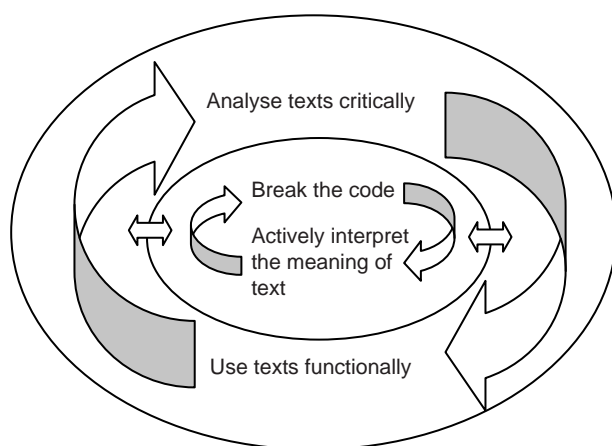


Figure 1 An approach to literacy instruction reconciling the skills and practices approaches.

Despite the dominance of some version of an ameliorative perspective in the industrialized world, there are two other roles that literacy education can play. Each of these roles is particularly relevant to groups with little access to mainstream education. The first is an adaptive role, where individuals from one language and literacy community enter another. This might include economic migrants or refugees, as well as people who have been displaced from one job or life situation into another with different demands, such as older adults attending college for the first time or farm workers following seasonal crops.

In the United States many literacy programs are going beyond people with English as their first language and working to develop initial language skills with speakers of other languages. This is particularly evident in Texas, where more than half the population are Spanish speaking. In the United Kingdom, some adult literacy agencies are working with substantial numbers of refugees to provide initial English language instruction.

Many of these learners will have well-developed literacy and numeracy practices in their first language – in the case of Texas many learners have an excellent Mexican secondary school education, and in the UK refugees are frequently academics, doctors, engineers, and other highly educated professionals. The provision of language education alongside literacy education in programs is generally not widely acknowledged. This can place unpredictable and occasionally unrealistic demands on instructors, resources, and learners themselves. The emphasis on the ameliorative role of literacy education can obscure the adaptive application of literacy learning, perhaps resulting in less appropriate services for this group of learners.

The final role of adult literacy education is foundational. Many countries throughout the world do not have the universally accessible, and generally compulsory, primary education found in the industrialized nations. While UNESCO (2005) has committed itself strongly to literacy as the core of education for all, many people around the world do not gain access to any form of literacy education until later in life. The gross enrolment rate in primary school is below 60% in some African countries, and the age of the child at enrolment may be considerably higher than usual for primary school (UNESCO, 2005). In addition, there is a degree of gender imbalance in school attendance in around 40% of countries, though this is generally reducing quite rapidly (UNESCO, 2005).

For learners in countries without universal access to schooling the ameliorative approach, with its assumption that the conventions and application of education are understood, is inappropriate. New adult learners in this context will be entering classes with little understanding of the nuances and expectations of education, and often will be motivated by economic or instrumental concerns about the care of their family.

There is also a danger of well-meaning aid agencies establishing projects that inadvertently create situations of neo-colonialism, where Western models of literacy education are applied to situations very different from the Western countries. This can easily come to be seen as the most valuable form of learning, displacing local approaches to text and traditional forms of numeracy. An example is drilling learners in rows, teaching literacy practices that then fall into disuse because of irrelevance (Wickens and Sandlin, 2007), or the use of English in post-Colonial settings (Robinson, 2007).

Given the different roles that adult literacy education can fill, some care must be taken when thinking about each situation. It is more complex than assuming that every literacy learner has somehow missed out on elementary schooling, and it is critical to avoid seeing learners as having some deficit.

Accountability and Assessment

There has been a general increase in the resources committed to adult literacy education throughout the world over the last two decades. It remains unclear how long this will last, or what the final results will be, but it has profoundly affected the conceptualization and delivery of literacy education. These changes have resulted in more attention being paid to the outcomes of literacy education. Historically, literacy programs for adults have rarely been strongly concerned with measuring the progress of learners, or indeed the efficiency and effectiveness of the agencies delivering the programs. This is no longer the case in the industrialized countries, resulting in profound transformations of the field.

In thinking about assessment and accountability, it helps to be clear about the two central ideas. Assessment is measurement of learner progress through standardized tests, individual progress reports, or some combination of these and other methods. Accountability is the requirement for programs to demonstrate that they are having a positive impact on the literacy use of learners. There is some confusion around these concepts because assessment data are often taken to be a straightforward measure of effectiveness. In this case, the best strategy for the program is to recruit only very competent learners, meaning they can easily show learners leaving the program with strong results. The issue of assessment leading recruitment and instruction has been tackled very rarely – all too often programs produce what they are asked to measure, potentially at the cost of meaningful learning (Merrifield, 1998).

One recent study (St. Clair and Belzer, 2007) looked at the accountability and assessment systems tied into adult literacy education in the United States, Scotland, and England (the latter have separate educational systems). The study suggested that there are two important dimensions to national accountability and assessment systems.

Table 2 Alignment and standardization in three national literacy systems

	<i>Weak alignment</i>	<i>Strong alignment</i>
Strong standardization	United States	England
Weak standardization		Scotland

The first is the degree of standardization in the system, meaning the extent to which tests, curricula, and methods are shared among all the programs surveyed. The second is the degree of alignment, meaning the extent to which philosophy, ideas, and approach to literacy are shared among the programs. It is possible to have one without the other, or to have both strong standardization and strong alignment (Table 2).

In the United States legislation of the late 1990s required the creation of a national reporting system, which collates results from across the country. To allow this to happen, a standard reporting approach has been developed, defining the preferred instruments and desired achievement while still allowing the state governments some latitude. This is a weakly aligned but strongly standardized system. In England, the adult literacy system has been standardized in outcomes in a way similar to the US, but in addition the curriculum and tests have been centralized to reflect a single approach to literacy education. This system is both standardized and aligned. In Scotland, there is strong alignment around the social-practices model of literacy discussed earlier, but very little standardization – programs are encouraged to develop their own approaches and resources.

Each of these approaches has its strengths and weaknesses. With high standardization, it is all too easy for programs and instructors to feel limited in responding to local circumstances, whereas low standardization can lead to uncertainty about the quality of the services learners are receiving. A single hierarchy of tests and exams and a requirement that learners demonstrate a certain amount of progress for a certain investment of time and money is not compatible with a social-practices view of literacy. High alignment is most effective where there is genuine commitment to a particular conception of literacy education, and that may be hard to maintain across a national system over any length of time. Despite recognition of the importance of locally tailored programs maintaining effective, learner-led practice may prove to be a significant challenge for literacy educators in the future.

Changes in the Literacy Education Workforce

The last few years have also seen a move toward professionalization of literacy instruction in the industrialized countries. This development, which is supported by many instructors and administrators, is generally accompanied

by pressure for adult literacy education to more closely resemble the established educational professions – school teaching in particular. This suggests that a specific qualification for teaching adult literacy may be developed, and that there would be some attention given to providing professional development for core staff. While professionalization would fit well with the agenda of accountability and quality control, it would also raise a number of problems.

One implication of any move to a professionalized workforce would be the loss of volunteers, who currently perform a central role in many systems. If they and part-time workers were expected to undergo substantial training before being able to work with learners, there is a danger that it would be more difficult to recruit. There is also the question of what workers should be taught – if they are only provided with a basic introduction to the field, it is unclear that they would be necessarily be able to deliver better-quality instruction. There is a real possibility that a straightforward, standardized curriculum would be developed for delivery by semi-trained staff, reducing the diversity of practices in the field.

Conclusion

Adult literacy education takes place in a wide range of settings where learners engage in a variety of ways with texts of all kinds. It is critical for effective instruction that both the method and the content of instruction recognize this diversity and that deficit approaches, where the learner is assumed to have something wrong with them, are avoided. Instead, a variety of outcomes of literacy instruction should be valued.

The recent structural changes in the field throughout the world have been substantial, with issues to do with accountability and professionalization rising up the agenda. These changes have tended to move adult literacy education closer to school-based education. At the same time, the importance of having a system that is highly aligned around values and ideology is being more widely recognized, perhaps as a response to the trend for managerialism. Finally, there is real interest in bringing skills and social-practices perspectives together to create a more nuanced understanding of teaching and learning that enables literacy education to be more closely aligned to the practices used in people's everyday lives.

See also: Adult Basic Education: A Challenge for Vocational Based Learning; Lifelong Learning.

Bibliography

- Barton, D. (1994). *Literacy: An Introduction to the Ecology of Written Language*. Oxford, UK: Blackwell.
 Barton, D., Hamilton, M., and Ivanic, R. (eds.) (2000). *Situated Literacies: Reading and Writing in Context*. New York: Routledge.

- Crowther, J. and Tett, L. (2001). Democracy as a way of life: Literacy as citizenship. In Crowther, J., Hamilton, M., and Tett, L. (eds.) *Powerful Literacies*, pp 108–118. Leicester: NIACE.
- Freebody, P. and Lo Bianco, J. (1997). *Australian Literacies Part II: What a National Policy on Literacy Should Say*. Canberra: Language Australia.
- Friere, P. and Macedo, D. (1987). *Reading the Word and the World*. London: Routledge/Kegan Paul.
- Hamilton, M., Hillier, Y., and Tett, L. (2006). Introduction: Social practice of adult literacy, numeracy and language. In Tett, L., Hamilton, M., and Hillier, Y. (eds.) (2006) *Adult Literacy, Numeracy and Language: Policy, Practice and Research*, pp 1–18. Maidenhead: Open University Press.
- Merrifield, J. (1998). *Contested Ground: Performance Accountability in Adult Basic Education*. Cambridge, MA: National Center for the Study of Adult Learning and Literacy.
- Papen, U. (2005). *Adult Literacy as a Social Practice*. New York: Routledge.
- Quigley, B. A. (1997). *Rethinking Literacy Education: The Critical Need for Practice-Based Change*. San Francisco, CA: Jossey-Bass.
- Reder, S. (2008). *Dropping Out and Moving on: Life, Literacy and Development among School Dropouts*. Cambridge, MA: Harvard University Press.
- Robinson, C. (2006). Context of key language in four adult learning programs. *International Journal of Educational Development* 27, 541–551.
- Sandlin, J. A. and Cervero, R. M. (2003). Contradictions and compromise: The curriculum-in-use as negotiated ideology in two welfare-to-work classes. *International Journal of Lifelong Education* 22(3), 149–165.
- Shor, I. (1993). Education is politics: Paulo Freire's critical pedagogy. In McLaren, P. and Leonard, P. (eds.) *Paulo Freire: A Critical Encounter*, pp 169–176. New York: Routledge.
- St. Clair, R. and Belzer, A. (2007). National accountability systems. In Campbell, P. (ed.) *Accountability in Adult Basic Education*, pp 159–206. Edmonton, AB: Grass Roots Press.
- St. Clair, R. and Priestman, S. (1997). What's the good of adult education? *North West Philosophy of Education Conference*. Vancouver: December.
- Street, B. (1995). *Social Literacies: Critical Approaches to Literacy in Development, Ethnography and Education*. London: Longman.
- UNESCO (2005). *Education for All: Literacy for Life*. Paris: UNESCO.
- Wickens, C. M. and Sandlin, J. A. (2007). Literacy for what? Literacy for whom? The politics of literacy education and neocolonialism in

UNESCO and World Bank sponsored literacy programs. *Adult Education Quarterly* 57(4), 275–292.

Further Reading

- Belzer, A. and St. Clair, R. (2003). *Information Series Number 391: Opportunities and Limits: An Update on Adult Literacy Education*. Columbus, OH: The State University of Ohio Center on Education and Training for Employment.
- Binkley, M., Matheson, N., and Williams, T. (1997). *Working Paper No. 97-33: Adult Literacy: An International Perspective*. Washington, DC: US Department of Education, National Center for Educational Statistics.
- Freire, P. (1972). *Pedagogy of the Oppressed*. Harmondsworth: Penguin.
- Graff, H. J. (1995). *The Labyrinths of Literacy*. Pittsburgh, PA: University of Pittsburgh.
- OECD (1997). *Education Policy Analysis*. Paris: OECD.
- OECD (2000). *Literacy in the Information Age*. Paris: OECD.
- Stephens, S. (2000). A critical discussion of the 'new literacy studies'. *British Journal of Educational Studies* 48(1), 10–23.
- Tett, L., Hamilton, M., and Hillier, Y. (2006). *Adult Literacy, Numeracy and Language: Policy, Practice and Research*. Maidenhead: Open University Press.

Relevant Websites

- <http://www.literacy.lancs.ac.uk> – Lancaster Literacy Research Centre, Research and Practice in Adult Literacy (RaPAL).
- <http://www.ncsall.net> – National Center for the Study of Adult Literacy and Learning (United States).
- <http://www.literacytrust.org.uk> – National Literacy Trust (United Kingdom).
- <http://www.nrdc.org.uk> – National Research and Development Centre (United Kingdom).
- <http://www.unesco.org> – United Nations Educational, Scientific and Cultural Organization.
- <http://go.worldbank.org> – World Bank on Education for All.

Citizenship and Immigrant Education

D Schugurensky, University of Toronto, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Citizenship – The status of being a citizen; membership in a community; the quality of an individual's response to membership in a community.

CONFITEA (Conférence internationale sur l'éducation des adultes (International Conference on Adult Education)) – A series of international meetings, which takes place every 12 years since 1949, organized by United Nations Educational, Scientific and Cultural Organization (UNESCO).

Immigrate – To enter and usually become established; to come into a country of which one is not a native for permanent residence.

Jus sanguinis – Latin for right of blood, a social policy by which citizenship is determined by having an ancestor who is a national or citizen of the state.

Jus soli – Latin for right of the soil, a right by which citizenship can be recognized to any individual born in the territory of the related state.

Naturalization – The acquisition of citizenship by individuals who were not citizens of that country at the moment of birth.

UNESCO (United Nations Educational, Scientific and Cultural Organization) – A specialized United Nations agency founded in 1945.

Introduction

In the field of adult education, it is possible to identify two separate areas of research, theory, and practice that are often related: citizenship education and immigrant education. Generally speaking, citizenship education aims at preparing individuals to become citizens in a particular political community. Typically, citizenship-education initiatives focus on the teaching of citizens' rights and duties, as well as the legal system and the functioning of government institutions. In some cases, it also includes the nurturing of critical analysis of social reality and engagement in civic and political activities. Immigrant education aims at enhancing the integration of new immigrants in the economic, social, cultural, and political life of the receiving countries through a variety of services and programs. These include programs such as second-language acquisition, special language

courses for immigrant professionals, entry-level job-skills development, entrepreneurial training, financial literacy (from basic banking to investment classes), and information about local culture, laws, and regulations. Services include employment counseling and job-search strategies, settlement assistance, and access to computers.

In many countries, citizenship education and immigrant education are offered to children, youth, and adults by a variety of governmental agencies, nonprofit organizations, private institutions, and community groups. The intersection between citizenship education and immigrant education can be found in the area of citizenship education for immigrants. This area focuses on preparing newcomers to become citizens in the host country. By and large, the size and importance of this area in a given country is directly correlated to the size and history of immigration. In other words, those countries with higher immigration rates and with longer traditions as recipient societies are more likely to have established programs and pedagogical materials on citizenship education for immigrants and to devote significant human and material resources to this area. In the past, the main impetus behind those initiatives was an interest to quickly assimilate immigrants into the prevailing culture and values of the host society, and particularly to incorporate them as soon as possible into the labor market. More recently, some countries established immigration policies and programs aiming at balancing the integration of immigrants into the host society and the respect for the immigrants' cultural identity, that is, promoting simultaneously, both unity and diversity.

Before addressing general trends related to adult citizenship education and immigration education, it is useful to discuss some conceptual and historical issues related to citizenship and to immigration.

Citizenship

As noted above, citizenship education includes all those educational programs and initiatives that have as their main purpose preparing individuals to become citizens. Although at first glance this seems like a straightforward proposition, it is pertinent to ask: What does it mean to become a citizen? There is no easy answer to this question because citizenship is a dynamic, contextual, contested, and multidimensional concept. It is dynamic because its meaning, characteristics, and scope have changed throughout history. For instance, in a not-so-distant past, there were

many countries that excluded slaves, women, nonwhites, and illiterates from citizenship. It is contextual because, even in the same historical moment, citizenship is a term that has interpretations and applications in different societies. Indeed, at any historical moment, many countries have applied – and still do – different criteria to determine who is a citizen and who is not, who can become a citizens and who cannot, what the traits of a good citizen are, and so on. Moreover, citizenship is a contested term because, even in the same time and space, in the same nation-state, and there are profound disagreements about what citizenship is and what it should be. In each society, different groups have different perspectives on the criteria for inclusion and exclusion, and different ideas about the qualities of good citizenship.

The term citizenship is also multidimensional because it encompasses four different dimensions that are frequently conflated: status, identity, civic virtues, and agency. Status relates to issues of membership to a particular political community. Identity relates to issues of feelings of belonging and loyalty to that political community. Civic virtues refer to the dispositions, values, and behaviors that are expected from citizens, and agency refers to the capacity to act as a citizen, that is, to engage actively in civic and political life with the possibility of making a difference. Since these four dimensions influence the approaches adopted by adult-citizenship-education programs, it is pertinent to elaborate on them.

In the most common understanding of the term, citizenship is a status bestowed on those who are full members of a community, which means that those who possess the status of citizens are equal with respect to the rights and duties endowed by it (Marshall, 1949). In modern times, that community is the nation-state, to the extent that citizenship is often equated with nationality. Noncitizens are usually known as foreign aliens. However, at this moment it is also possible to be a citizen of a supranational political entity like the European Community. Citizenship as status distinguishes between citizens, who are full members of a particular political community, and thus can have a passport, are eligible to vote and to be elected, and so on, and noncitizens (or aliens), who have limited rights or no rights at all. Today, citizenship status is usually granted by birthplace (*jus soli*), descent (*jus sanguinis*), or naturalization, although in the past it was often granted and denied on the basis of factors like class, gender, or race. For instance, in many countries it was not until the first half of the twentieth century (in some even later) that women were recognized as full citizens with equal rights. Even in the twenty-first century, some population groups are denied certain rights. For instance, only a few countries allow gay couples the same right to marriage that is granted to the heterosexual population, and although the Universal Declaration of Human Rights states that everyone has a right to a nationality, today

there are more than 15 million stateless people who have no citizenship status in any state, either because they never acquired it in their birth country or because they lost it. In considering the notion of citizenship as status it is pertinent to distinguish between formal and real citizenship: formal equality is largely irrelevant if it is contradicted on a daily basis by economic, social, political, and cultural inequalities. For this reason, any discussion on citizenship as status that is also concerned with real membership in a particular political community must consider the dynamics of inclusion and exclusion of that community.

Identity relates to feelings of belonging that a person has toward a community. This feeling as a full member of a community is independent of the formal citizen status of that person. The distinction between status and identity can be found everywhere, but it is more clear in multiethnic, multilingual, multicultural, and multireligious states, and particularly in nation-states that are multination states (Dean, 2004; Kymlicka, 2001; Wilkinson and Hebert, 1999). In these cases, identity is rooted in factors like a common history, language, religion, values, traditions, and culture, which seldom coincide with the artificial territory of a nation-state. This should not be surprising, because nation-states are imagined communities whose boundaries change over time due to invasions, wars, annexations, conquests, and separations (Anderson, 1983). Hence, it is not surprising that some cultural and linguistic minorities as well as aboriginal groups may not feel part of the nation-state in which they are legal citizens. In the same vein, immigrant communities may hold a legal citizenship status in two or more nation-states but feel allegiance to only one of them. Moreover, there are cases of immigrants who are eligible for citizenship in their host country but do not take it in order to keep a feeling of belonging to their home country, even if they never return to it. Another example of a mismatch between status and identity can be observed among people with internationalist inclinations who are legal citizens of nation-states but define themselves as citizens of the world, even if planetary citizenship is not a legal condition.

Civic virtues refer to the values, attitudes, and behaviors that are expected of good citizens. However, there is no universal agreement on the ideal of a good citizen. For some groups, the main civic virtues of a good citizen are patriotism, obedience, honor, diligence, and religiosity. Others emphasize compassion, respect, tolerance, honesty, solidarity, and individual responsibility, and others relate civic virtues to a critical analysis of social reality, interest for the common good, community participation, and political engagement. All these and many other civic virtues could be chosen, organized, and ranked according to the moral preferences of those who create the list. Moreover, the ideal of good citizen promoted by the state varies according to historical, ideological, and political contexts. For instance, the model of good citizen and

the virtues cultivated by the Nazi regime were different from the model and the virtues promoted in Germany in the nineteenth century or today.

Agency refers to the state of being in action or affecting change; thus, the dimension of citizenship as agency invokes the idea of citizens as social actors. The exercise of citizenship, individual or collective, does not occur in a vacuum, but in concrete social relations mediated by power. Indeed, social circumstances determine to a large extent what citizens can do or feel allowed to do, but at the same time, citizens have some degree of agency to change those circumstances and make history. Hence, the notion of citizenship as agency recognizes that social action – which varies both in the intensity and orientation of citizen actions – occurs in a context marked by a constant interplay of limits and possibilities, domination and autonomy, and control structures and liberating forces. Although the literature often refers to passive and active citizenship, in real life these constitute end points of a continuum rather than dichotomous categories.

Immigration

While human migration has existed throughout human history, immigration is the movement of people from one nation-state to another, and involves long-term permanent residence (and eventually citizenship) in the host country. Although short-term visitors are not considered immigrants, seasonal labor migration is often considered as a special form of immigration.

According to estimates from the United Nations, in 2005, the total number of migrants amounted to 190 million people, to which can be added an undetermined number of undocumented or irregular migrants, currently estimated as 30 million people. Typically, migration is driven by two simultaneous dynamics. On the one hand, push factors such as economic, political, or social problems in the country of origin; on the other, pull factors such as economic, social, and cultural opportunities in the destination country. Globally, the large majority of migrants flow from South to North, that is, from Latin America, Asia, and Africa to Europe and North America, but in recent decades, there has been an increase of South–South migration (often due to political instability or generalized violence) and even North–South migration, usually older people who seek retirement in countries with lower costs of living and better climates than their own home countries.

Many immigrants, particularly those who migrate from South to North, often face a great variety of problems, which range from emotional issues arising from leaving relatives, friends, and a known environment to issues related to legal barriers, moving expenses, uncertainty about the future, lack of familiarity with the new culture and language, difficulties in finding decent employment,

and in many cases also exploitation, discrimination, and racism. The acquisition of legal citizenship status in the host country may not necessarily translate into the enjoyment of full citizenship on the same foot as other social groups. In many instances, the discrepancy between legal and real citizenship status has more to do with language, race, gender ethnicity, country of origin, and social class than with the formality of official citizenship papers. As Banks (2004: 5) noted, “becoming a legal citizen of a nation-state does not necessarily mean that an individual will attain structural inclusion into the mainstream society and its institutes or will be perceived as a citizen by most members of the dominant group within the nation-state.”

From the perspective of the host societies, particularly those that regularly receive large numbers of immigrants, challenges include providing services, educational programs, and infrastructure to the new immigrants, making an effort to understand and respect different cultures, beliefs, and practices, and creating an environment that nurtures dialog, tolerance, and fairness. Countries with high international demand for immigration tend to establish specific selection criteria and admission policies. Whereas in the past those criteria were often based (sometimes explicitly) on racial or ethnic prejudices, presently they are likely to be based on factors such as schooling, occupation, wealth, and language proficiency, usually expressed in a point system. Such system gives preference to professionals and businesspeople, and discriminates against unskilled workers. This situation creates problems in the labor market due to unmet demand for workers in certain areas of the economy (e.g., construction, agriculture, and janitorial services), which in turn provokes a growth in undocumented migration. Another issue that has increased significantly as a result of the emphasis on professional immigrants is the limited recognition of foreign credentials and work experiences by certain professional colleges. This, in turn, creates a lose–lose situation. The home country suffers brain drain after considerable investment in human-resource development, the host country misses the opportunity of brain gain by forcing immigrant professionals to earn a living delivering pizza or driving taxis, and the professional immigrants lose the possibility of developing their full potential in their chosen field by taking jobs unrelated to their training.

Moreover, countries with large and diverse immigration populations often face the double challenge of ensuring a successful social, economic, political, and cultural integration of newcomers and at the same time reducing tensions that may arise across ethnic, national, or racial lines. In the past, most of these countries (known as settlement countries) favored the strategy of forced assimilation. This strategy, usually referred to with the metaphor of a melting pot, had as its main goal that immigrant groups

adopt the language, traditions, attitudes, and values of the dominant culture as fast as possible. In recent times, some settlement countries have adopted a different integration strategy through multiculturalism. Whereas assimilation approaches perceive diversity as a threat and tries to suppress ethnic identities, multiculturalism understands diversity as a resource and seeks to accommodate differences. Hence, multiculturalism tends to promote the celebration of cultural diversity and the management of such diversity through specific policies, programs, and initiatives in different levels of government. Kymlicka and Banting (2006) identify eight frequent policies that are emblematic of a multicultural approach to immigrant integration: (1) constitutional, legislative, or parliamentary affirmation of multiculturalism; (2) adoption of multiculturalism in school curricula; (3) inclusion of ethnic representation and/or sensitivity in the mandate of public media; (4) exemptions from dress codes, Sunday-closing legislation, etc.; (5) allowance of dual citizenship; (6) funding of ethnic-group organizations to support cultural activities; (7) funding of bilingual education or mother tongue instruction; and (8) affirmative action for disadvantaged immigrant groups. They point out that the first three policies celebrate multiculturalism, the middle two reduce legal constraints on diversity, and the last three represent forms of active support for immigrant communities and individuals.

Multicultural approaches have advocates and detractors. Detractors argue that multiculturalism and liberal democracy are fundamentally incompatible, and that the emphasis on diversity and difference undermines the sense of common national identity and nurtures resentment between minority and majority cultures (Sniderman and Hagendoorn, 2007). In some instances, the tension between the values and practices of immigrant groups and the prevailing values and practices of the host society have been addressed through initiatives of reasonable accommodation that generated much public controversy and political and juridical debate. Advocates argue that multiculturalism is an effective strategy to facilitate the full integration of immigrants into the host society while preserving their identity, balancing the principles of cultural distinctiveness and equality. They also argue that multiculturalism helps immigrants to keep their culture and language, reduces discrimination, enhances cross-cultural understanding, and promotes institutional change aimed at equalizing opportunities.

Citizenship and Immigrant Adult Education: Main Orientations

Citizenship and immigrant-education programs tend to address simultaneously the four dimensions of citizenship discussed above, but frequently they emphasize just one.

Those programs that focus on citizenship as status often emphasize formal membership to a particular political community, usually a nation-state. Beyond introductory content dealing with settlement challenges (e.g., language acquisition, resumé writing, and job-searching strategies), these citizenship programs concentrate on facts about national history and geography, government institutions, and the law. In some countries, the content of these programs is codified in textbooks that immigrants seeking naturalization must memorize in order to pass a citizenship test. More often than not, these textbooks tend to promote the official story of a nation's development, with limited coverage of controversial issues (Joshee and Johnson, 2007; Derwing *et al.*, 1998). These programs also describe the rights and duties of citizens, although seldom distinguishing between formal and real membership, or encouraging a critical analysis of the *status quo*. In the margins, however, there are some critical programs – usually carried out by nongovernmental organizations – that contrast the official perspective with the views of peoples that suffered conquest or discrimination. These programs also question taken-for-granted rules of inclusion and exclusion, interpret the law in the context of social dynamics of power and emancipatory struggles, and emphasize the fulfillment of human rights.

Education programs emphasizing citizenship as identity have tended for a long time to stress nation building and the assimilation of minority groups to the dominant group. Often, this has meant the mainstreaming and the malestreaming of curriculum content, and the elevation of the hegemonic language, religion, and culture to the higher level of a hierarchy. The main purpose (sometimes explicit, sometimes implicit) of many of these programs was identity conversion, that is, reshaping the identity of the recent immigrants in a way that corresponds less with their old world and more with their new world. This certainly included attitudes of loyalty and allegiance to the host country, its religion, and its political leaders. For instance, one of the main textbooks for immigrant citizenship education used in Canada during the first part of the twentieth century stated that the good citizen should love God, the Empire, and Canada (Fitzpatrick, 1919). Today, in the twenty-first century, new immigrants to Canada must pass a civics test and pledge allegiance to the Queen of England in order to become Canadian citizens. Interestingly, Canadian-born people are not required to take this oath or to pass a test as a condition for citizenship, even though 60% of Canadians fail a citizenship exam similar to the one taken by immigrants (Dominion Institute, 2007). Some citizenship-education programs strive at developing a planetary consciousness and an identity as world citizens. These programs are usually known as global education, and are often connected to peace education and environmental-education approaches.

Programs that focus on civic virtues tend to emphasize the development of a set of values, attitudes, dispositions, and behaviors that are expected from good citizens. Paraphrasing the Archbishop of York, who said that the true purpose of education is to produce citizens, Eleanor Roosevelt argued that the purpose of education is to produce good citizens. However, there is no universal agreement about the values and dispositions of good citizenship. Certainly, some values have more universal appeal than others, but the set of values privileged by official-education programs change from country to country and from time to time, and range according to factors like the ideological orientation of the government in power and the prevailing beliefs in a given society. From a pedagogical perspective, citizenship-education programs that focus on civic virtues follow two main approaches. One approach consists of instilling a particular set of values and dispositions through exhortations and inducements. The other approach aims at helping learners develop their own values by examining ethical dilemmas and examining different perspectives in a democratic environment.

Finally, programs that focus on citizenship as agency tend to promote the development of an active citizenry. However, this could take different forms and directions. Westheimer and Khane (2004), for instance, identify three implicit models of active citizenship: responsible, participatory, and justice oriented. Responsible citizens are expected to avoid littering, pick up litter of others, donate blood, recycle, volunteer, pay taxes, exercise, stay out of debt, and the like. Participatory citizens are expected to take active part in the civic affairs and social life of the community, and assume leading roles in neighborhood associations, school councils, or political parties. Justice-oriented citizens are expected to be able to critically analyze structures of inequality, consider collective strategies to challenge injustice, and, whenever possible, address root causes of social problems. Although the three models are important for the development of agency among learners, justice-oriented programs arguably have more potential for building a democratic and just society, as this requires nurturing political subjects who have a critical understanding of social structures and the capacity (and willingness) to participate actively in society. This tradition of citizenship education for transformative social action is grounded in the work of many adult educators, including Jane Addams, N. F. S. Grundtvig, Moses Coady, José María Arizmendi, Paulo Freire, and Myles Horton.

Regardless of their particular orientation and foci, most adult-citizenship-education programs tend to emphasize – often in theory, sometimes also in practice – the importance of nurturing an informed, purposeful, critical, caring, and active citizenship. This requires the development of certain attributes that include knowledge on a variety of areas (from political ideologies and societal structures to government institutions and democratic governance), skills

(from information seeking and analysis to conflict resolution and public speaking), attitudes (from self-confidence to respecting other people's opinions and cooperation), and practices (from participating in local associations to voice ideas and voting in municipal, provincial, and federal elections). These attributes can be learned through different programs, although those programs that include an experiential component (e.g., allowing learners to experiment with participatory democracy) tend to be more successful (Merrifield, 2001; Benn, 2000; Annette and Mayo, 2008).

Concluding Remarks: Adult-Citizenship Education in Immigrant Societies

In recent times, some countries with large-scale, continuous, and heterogeneous immigration patterns have established a variety of educational programs to promote successful integration processes and mutual understanding. These programs are usually inspired by two main approaches. The first one, known as multicultural education, focuses on familiarizing learners with different cultures in order to reduce stereotypes and ethnocentric attitudes. Multicultural education emphasizes the teaching of history, traditions, and customs of different cultures. More often than not, this is done superficially and uncritically, focusing almost exclusively on the three Fs of folklore, food, and festivities. However, occasionally multicultural education programs also encourage a deeper social analysis that includes examining dynamics of inequality, discrimination, and racism.

The other approach, known as intercultural education, is about proactive interaction among different ethnic groups. Intercultural education is not so much about acquiring knowledge about other cultures but about promoting communication, cooperation, and regular relations among groups. Although both multicultural and intercultural education programs have similar aims (e.g., promoting understanding, tolerance, and respect among different cultures), the former emphasizes the cognitive dimension whereas the latter puts more emphasis on skills and attitudes. In theory, intercultural education is a two-pronged project that includes working with both the immigrant and the nonimmigrant population. Theoretically, intercultural education is expected to assist immigrant groups in their integration process and at the same time it should help the nonimmigrant society to accept immigrants as equals, with the overall purpose of building a more democratic, inclusive, and egalitarian society premised on an intercultural and active citizenship. In practice, however, many intercultural education initiatives tend to concentrate their efforts on compensatory pedagogical interventions with minority groups rather than on projects that involve all groups in society (Rodríguez Izquierdo, 2009).

Related to this are the debates on the most appropriate pedagogical strategies to cultivate the habits of democracy in diverse societies. The most common approach is to avoid open discussions about controversial topics among participants. Many instructors fear conflict, and hence tend to implement a safe curriculum that reduces the possibility of risk. Others, however, believe that the best way to nurture civic virtues is to welcome controversial topics and hard questions, as well as to encourage participants not to get along. This implies recognizing the plurality of viewpoints among participants and facilitating a respectful dialog among them (McLaughlin, 2004; Hughes and Sears, 2004). As Dewey (1916) pointed out, one of the most effective ways to learn democratic values is to practice democracy in a democratic community.

In closing, the potential of adult education to contribute to a more democratic society is reflected in the first theme of the agenda for the future approved at the fifth International Conference of Adult Education (CONFINTEA) in Hamburg, where it was noted that the challenges of the twenty-first century “require the creativity and competence of citizens of all ages in alleviating poverty, consolidating democratic processes, strengthening and protecting human rights, promoting a culture of peace, encouraging active citizenship, strengthening the role of civil society, ensuring gender equality and equity, enhancing the empowerment of women, recognizing cultural diversity (including the use of language, and promoting justice and equality for minorities and indigenous peoples) and a new partnership between state and civil society” (CONFINTEA, 1997: 1). The agenda adds that to reinforce democracy, it is essential to strengthen learning environments, encourage citizen participation, and create contexts where a culture of equity and peace can take root.

To achieve these goals, CONFINTEA asks adult educators to make four commitments: to create greater community participation; to raise awareness about prejudice and discrimination; to encourage greater recognition, participation, and accountability of nongovernmental organizations and local community groups; and to promote a culture of peace, intercultural dialog, and human rights. Adult citizenship and immigrant-education programs can assist these efforts by addressing issues of status, identity, civic virtues, and agency. This can be pursued through a variety of pedagogical/political strategies that adapt and reinvent the contributions of twentieth-century adult education to the realities of the twenty-first century. Adult education can play an important role in improving the conditions for learning, in increasing the sharing of this learning along the lines of intercultural education, and in equalizing opportunities for learning and for meaningful participation among all members of society.

See also: Adult Learning, Instruction and Programme Planning; Insights from Freire; Barriers to Participation in Adult Education; Characteristics of Adult Learning; Children of Migrant Populations; Participation in Adult Learning; Race and Ethnicity in the Field of Adult Education.

Bibliography

- Anderson, B. (1983). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.
- Annette, J. and Mayo, M. (eds.) (2008). *Adult Learning for Active Citizenship*. London: NIACE.
- Banks, J. (2004). Democratic citizenship education in multicultural societies. In Banks, J. (ed.) *Diversity and Citizenship Education*, pp 3–15. San Francisco, CA: Jossey-Bass.
- Benn, R. (2000). Including citizenship in the adult curriculum. *Paper Presented at SCUTREA, 30th Annual Conference*, 3–5 July, University of Nottingham.
- CONFINTEA (1997). The agenda for the future. *Fifth International Conference on Adult Education*, Paris: UNESCO.
- Dean, B. (2004). Pakistani conceptions of citizenship and their implications for democratic citizenship education. In Mundel, K. and Schugurensky, D. (eds.) *Lifelong Citizenship Learning, Participatory Democracy and Social Change*, pp 1–11. Toronto, ON: Transformative Learning Centre, OISE/UT.
- Derwing, T., Kama, J., and Murray, M. (1998). Citizenship education for adult immigrants: Changes over the last ten years. *Alberta Journal of Educational Research* 44, 383–396.
- Dewey, J. (1916). *Democracy and Education*. New York: Free Press.
- Dominion Institute (2007). *What Do Young Adults Know About Canadian History? 10 Year Benchmark Study*. Toronto, ON: Dominion Institute.
- Fitzpatrick, A. (1919). *Handbook for New Canadians*. Toronto, ON: Ryerson Press.
- Hughes, A. and Sears, A. (2004). Situated learning and anchored instruction as vehicles for social education. In Sears, A. and Wright, I. (eds.) *Challenges and Prospects for Canadian Social Studies*, pp 17–37. Vancouver: Pacific Educational Press.
- Joshee, R. and Johnson, J. (2007). *Multicultural Education Policies in Canada and the United States*. Vancouver: University of British Columbia Press.
- Kymlicka, W. (2001). *Politics in the Vernacular: Nationalism, Multiculturalism, and Citizenship*. New York: Oxford University Press.
- Kymlicka, W. and Banting, K. (2006). Immigration, multiculturalism, and the welfare state. *Ethics and International Affairs* 20(3), 281–304.
- Marshall, T. H. (1949). Citizenship and social class. In Marshall, T. H. and Bottomore, T. (eds.) *Citizenship and Social Class*, pp 1–51. London: Pluto Press.
- McLaughlin, D. (2004). Cultivating habits of democracy. Asking the hard questions. *Education Canada* 45(1), 33–35. Toronto, ON: Canadian Education Association.
- Merifield, J. (2001). *Learning Citizenship*. London: Learning from Experience Trust.
- Rodríguez Izquierdo, R. M. (2009). La investigación sobre la educación intercultural en España. *Archivos Analíticos de Políticas Educativas* 17(4). <http://epaa.asu.edu/epaa/v17n4>.
- Sniderman, P. and Hagendoorn, L. (2007). *When Ways of Life Collide: Multiculturalism and Its Discontents in the Netherlands*. Princeton, NJ: Princeton University Press.
- Westheimer, J. and Khane, J. (2004). What kind of citizen? The politics of education for democracy. In Mundel, K. and Schugurensky, D. (eds.) *Lifelong Citizenship Learning, Participatory Democracy and Social Change*, pp 67–79. Toronto, ON: Transformative Learning Centre, OISE/UT.
- Wilkinson, L. and Hebert, Y. (1999). *Citizenship Values: Towards an Analytical Framework*. Calgary, AB: Prairie Centre of Excellence for Research on Immigration and Integration, Metropolis.

Further Reading

- Boggs, D. (1991). Civic education: An adult education imperative. *Adult Education Quarterly* **42**, 46–55.
- Bron, A., Field, J., and Kurantowicz, E. (eds.) (1998). *Adult Education and Democratic Citizenship II*. Krakow: Impuls.
- Callan, E. (1997). *Creating Citizens: Political Education and Liberal Democracy*. Oxford: Clarendon Press.
- Citizenship and Immigration Canada (2005). *A Look at Canada*. Ottawa, NJ: CIC.
- Coady, M. (1939). *Masters of Their Own Destiny*. New York: Harper and Brothers.
- Cogan, J. and Derricott, R. (eds.) (2000). *Citizenship for the 21st Century: An International Perspective on Education*. London: Kogan Page.
- Foley, G. (1999). *Learning in Social Action: A Contribution to Understanding Informal Education*. London: Zed Books.
- Freire, P. (1970). *Pedagogy of the Oppressed*. New York: Continuum Publishing.
- Giroux, H. A. (1980). Critical theory and rationality in citizenship education. *Curriculum Inquiry* **10**(4), 329–366.
- Gutmann, A. (1987). *Democratic Education*. Princeton, NJ: Princeton University Press.
- Hahn, C. (1998). *Becoming Political. Comparative Perspectives on Citizenship Education*. Albany, NY: State University of New York Press.
- Ichilov, O. (ed.) (1998). *Citizenship and Citizenship Education in a Changing World*. London: Woburn Press.
- Korsgaard, O. (2001). From social citizenship to active citizenship; from adult education to adult learning. *Special Issue: Citizenship, Adult Education and Lifelong Learning. Journal of World Education* **31**(1).
- Krischke, P. J. (2001). *The Learning of Democracy in Latin America: Social Actors and Cultural Change*. Huntington, NY: Nova Science Publishers.
- Ladson-Billings, G. (2004). Culture versus citizenship: The challenge of racialized citizenship in the United States. In Banks, J. (ed.) *Diversity and Citizenship Education*, pp 99–126. San Francisco, CA: Jossey-Bass.
- McKay, S. (1993). *Agendas for Second Language Literacy*. Cambridge: Cambridge University Press.
- Nie, N., Junn, J., and Stehlik-Barry, K. (1996). *Education and Democratic Citizenship in America*. Chicago, IL: The University of Chicago Press.
- Oldenquist, A. (1996). *Can Democracy Be Taught? Perspectives on Education for Democracy in the United States, Central and Eastern Europe, Russia, South Africa and Japan*. Bloomington, IN: Phi Delta Kappa Educational Foundation.
- Oliver, L. (1987). Popular education and adult civic education: The Third World is a different place. *Convergence* **20**(1), 40–50.
- Richardson, G. and Blades, D. (eds.) (2006). *Troubling the Canon of Citizenship Education*. New York: Peter Lang.
- Roosevelt, E. (1930). Good citizenship: The purpose of education. *Pictorial Review* **April**(4), 94–97.
- Schudson, M. (1998). *The Good Citizen: A History of American Civic Life*. New York: Free Press.
- Schugurensky, D. (2006). Adult citizenship education: An overview of the field. In Fenwick, T., Nesbit, T., and Spencer, B. (eds.) *Contexts of Adult Education: Canadian Perspectives*, pp 68–80. Toronto, ON: Thompson Educational Publishing.
- Selman, G. (1998). The imaginative training for citizenship. In Scott, S., Spencer, B., and Thomas, A. (eds.) *Learning for Life: Canadian Readings in Adult Education*, pp 24–34. Toronto, ON: Thompson Educational Publishing.
- Torres, C. A. (1998). *Democracy, Education, and Multiculturalism: Dilemmas of Citizenship in a Global World*. Lanham, MD: Rowman and Littlefield.
- Wildemeersch, D., Stroobants, V., and Bron, M. (eds.) (2005). *Active Citizenship and Multiple Identities in Europe. A Learning Outlook*. Frankfurt: Peter Lang.

Relevant Websites

- <http://www.unesco.org> – United Nations Educational, Scientific and Cultural Organization, CONFINTIA VI.
- <http://www.un.org> – United Nations, International Migration.

Community Based Adult Education

B Connolly, National University of Ireland, Maynooth, Republic of Ireland

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Consciousness raising – An increasing of concerned awareness especially of some social or political issue.

Empowerment – The promotion of self-actualization or influence.

Praxis – In the context of community education, praxis connotes the synergy of activism and reflection, in order to bring about a more just and equal society.

Really useful knowledge – This concept was developed in the nineteenth century to critique dominant forms of knowledge, and to contribute to changing all forms of domination while simultaneously promoting democracy and social amelioration. Really useful knowledge underpins the reflection in praxis.

Really useful methods – These include ways of working with learners to enhance democracy, participation, and equality. These methods underpin the activism in praxis.

Introduction

Any discussion on community education must take into account that perspectives vary from context to place. Community education may be seen as an extension of a pragmatic education service designed to target hard-to-reach people, and integrate them into the mainstream, through employment, further education, or rehabilitation. It may be interpreted as a dimension of community development empowering powerless people to address their own educational and social needs. It also may be perceived as an adjunct to civil society, in which citizenship and participation are enhanced and strengthened. It may be named and understood in different ways too. Terms such as nonformal adult education, outreach, extra-mural, liberal adult education, locally based adult education, lifelong learning, training, and informal adult education are also used as synonyms for community education in different circumstances. Further, community education, positioned within the meanings of community, may be construed as a caring process, with a big emphasis on relationship and interpersonal connection, with the attendant focus on processes and

methods shaped to enhance these caring qualities. This article discusses these myriad dimensions, in the context of exploring differing perspectives and contexts. The historical foundations are discussed before looking at the different facets of community education. A wide breadth of concrete examples of community education is then considered in the context of inherent differences and ambiguities. Finally, the discussion is appraised to establish common ground for deepening critical democracy in these changing times.

Historical Foundations

The story of community education is somewhat elusive, but this section outlines that story to provide a backdrop to its scope and some concrete examples on the ground. Lynn Tett documented the development of community education in Scotland, identifying the recommendations of the Alexander Report as the milestones of the current practice of community education. These recommendations led to the Community Education Service, to examine nonvocational adult education. The recommendations were strongly influenced by the understanding that adult education should help address social disadvantage endured by several groups, such as lone parents, unemployed people, early school leavers, and minority ethnic groups. Further, the report recommended that community education ought to nurture a pluralist democracy, by managing the tensions between state policies and community politics. These landmarks were congruent with the historical sources of adult education for working-class people. Crowther (1999) traces the lineage of community education with the radical working-class organizations of the eighteenth and early nineteenth centuries. This source was hard fought, and wrested from the grip of the mainstream schooling and education. In Denmark, Grundtvig founded the folk high schools with the explicit objective of providing space for learning for citizenship. The legacy of the folk high schools included a strong linkage with social movements, particularly with the cooperative movement, contributing to the foundations for the welfare society in the Nordic regions. Further, the model is one of nonformal education, elevating the value of education *per se* (Smith, 2007).

In Northern Ireland, the part that community education played in the development of the peace process is yet to be assessed fully, but Lovett's earlier work in England

was influential at practical as well as theoretical levels when he moved back to Northern Ireland. In Lovett (1971) he reflected on his early community-based practice, maintaining that adult and community education must be much more than the provision of classes, in that it must also be integral to the whole community and must also be a communal activity.

In the Irish Republic, women's community education developed within a setting of difficulties, from the isolation of new suburbs, high unemployment, and emigration, coupled with the destructive war in Northern Ireland and the improvement of the status of women. The literature on the sphere is sparse due to the growth in a nonformal, nonacademic manner, probably unique in education initiatives in that it was created, fostered, and supported by noneducationalists (Connolly, 2003). Slowey noted the early days of adult education in the community, especially the adult education organized and conducted by women on the ground, (Slowey, 1985). Bassett, *et al.* (1989), made the case for adult education and emancipation, emanating from the work of Irish National Association of Adult Education (AONTAS), when it became more apparent that adult education was changing enormously through the phenomenon of the women's community-education growth and development. By the time the author did a small study looking at her own experience of women's community education, it was clear that women's community education was totally different from the traditional provision of night classes in empty-ish schools or universities, based on liberal studies and leisure courses, in three specific aspects: the processes and methods, which she named really useful methods, relating this concept to Thompson's really useful knowledge, which underpinned the content, in addition to the total learning environment. Women's community education offered programs of non-formal and formal adult education, but it also provided the space for informal learning, plus, vitally, childcare within caring, supportive settings (Connolly, 1989). That is, community education, whether supported by educational institutions, sponsored by communities, or arising from difficult life worlds, provides a way of responding to oppression and discrimination, particularly by building the capacity of the learners, in their own personal development and, crucially, in their communal and social development.

Thus, these historical foundations locates community education within the milieu of community concerns, including the very difficult, deep divisions in Northern Ireland, or the almost intractable issues around suburban isolation, early school leaving, drug addiction, environmental concerns, and unemployment.

While this historical overview is scant, and elides numerous milestones, it serves the purpose of illuminating the rootedness of community education in the social agenda of liberation, problem solving, and self-determination.

The next section looks more closely at the facets of community education that help to illuminate the flexibility and resourcefulness that enables it to respond to the breadth of issues.

Facets of Community Education

This section endeavors to explore the facets of community education and to discuss the implications. The understandings of the term community is discussed in greater depth later on, but as a marker, community in this context is used in the sense of the physical location in which people live, that is, place, or in the sense of the community of interest, the social network of people who share a common interest, such as the Irish in the USA. Or community may be applied to the group of practitioners, a quasi-functional meaning, such as the community of educators, social workers, doctors, or the like. The principles of community education, identified by AONTAS (2004), the national adult learning association, in Ireland, include the rootedness of community education in social justice and the process of empowerment, which aims to build the capacity of local people to respond to educational and structural disadvantage and to participate in decision making and policy creation (AONTAS 2004: 18). That is, AONTAS aligns its perspective with the ideological positioning of working toward addressing the causes of disadvantage and inequality. On the other hand, Smith (2009) draws a thumbnail definition – education for the community within the community. He views that community is not just the place, or context of the education program – building the community is also a central concern. Thus, the overlap between AONTAS and Smith is this core concern with community. Implied in building community is the process of working on relationships between members of the community, in order to enhance their lived experience.

Thompson delineates the ways of working in community education as specific and purposeful. She includes formal education, that is, adult education with credentials; informal adult education that occurs within social networks, such as the breaks between classes; and nonformal adult education, with no credentials. She holds that the processes work with people's life experience and it connects issues, ideas, and understanding to political action. These processes also foster practical skills, which help people to overcome alienation, as well as increase their capacity for employment and participation. Finally, she contends that community education helps to build self-esteem as well as enables people to understand others' perspectives (Thompson, 2002: 9–11). That is, Thompson builds the ground in community education between the person and the communal, the individual and the social. By including credentials, she sides conclusively with the

contention that community education can embrace the full range of education, from basic literacy and numeracy to certificates, diplomas, and degrees.

Community education may also straddle the divide between institutionally sourced adult education and the kind of education that arises from the people themselves. McGivney (2000) in her study of outreach adult education for marginal groups, for example, unemployed men, traces the factors which underpinned its success. These factors included the ways of working with the learners, relationships, and methods, the response to community issues and support for community groups. Outreach in this sense comprised the provision of programs from universities or other education providers, but the processes and methods are not those of institution-centered provision; rather they are responsive and relevant to the learners' lives. McGivney's research echoes closely with the experience in Ireland, with women's community education.

Figure 1 endeavors to capture the facets of community education, including the contradictions and ambiguities. It revolves around the centrality of the learners, a key component of community education, not just in terms of

the processes and methods, but also in terms of meaning, experience, and responsiveness.

These myriad facets demonstrate the breadth of community education, and these dimensions are reflected in concrete examples on the ground. The next section discusses cases of community education in specific exemplary contexts.

International Examples

A rudimentary review of community education reveals concrete examples in many contexts throughout the globe. This section provides a summary of these examples, in order to illuminate the ways in which societies engage with community education.

Community education is found in contexts as diverse as a middle-class, affluent city in the USA to a rural childbirth project in Kenya. These diverse examples include content-centered education, with the focus on information and resources; process-centered, with the focus on participative creative methods; or a combination of both,

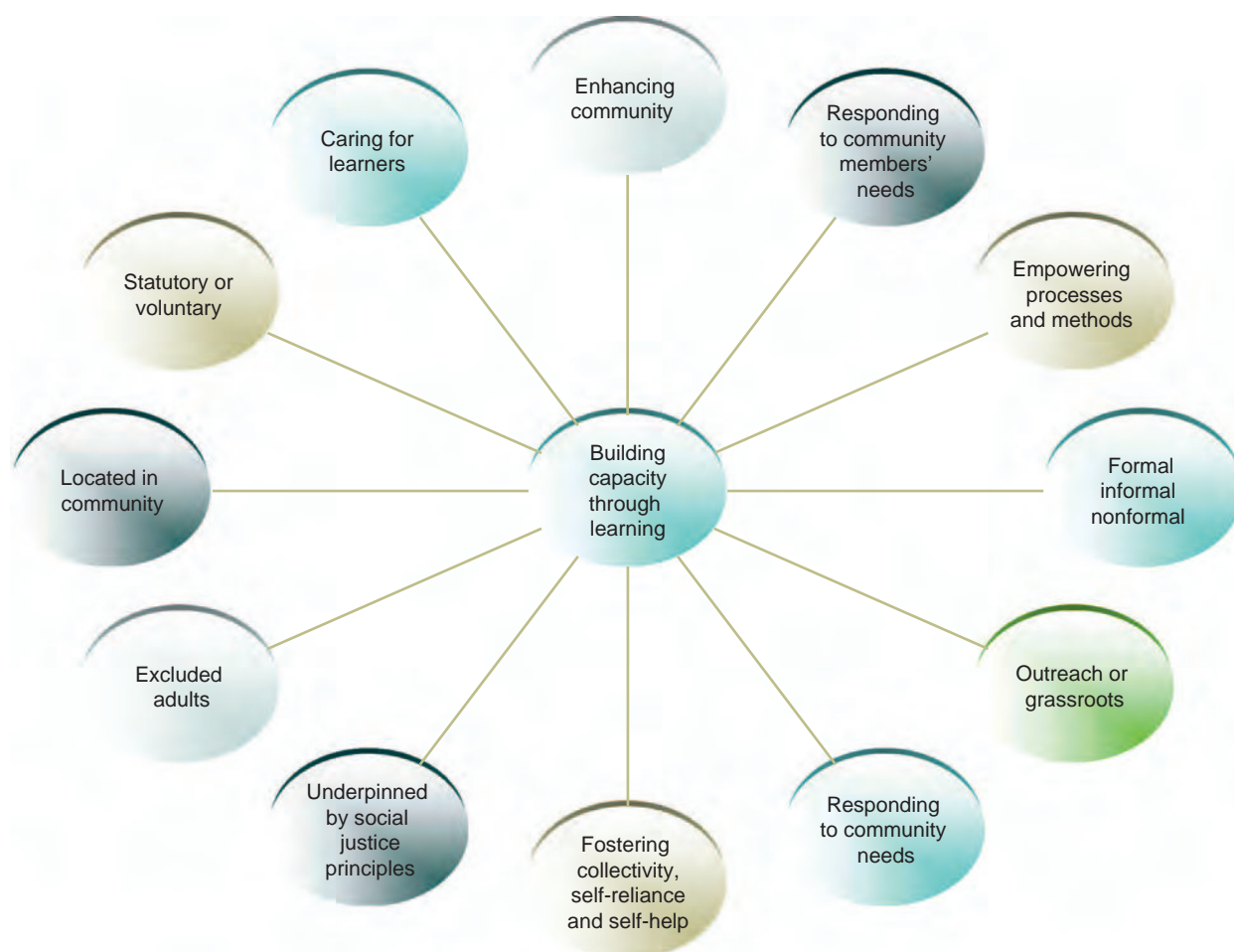


Figure 1 Facets of community education and the breadth of scope.

depending on the circumstances. They are also practiced in groups, in e-learning and in one-to-ones, employing basic resources on the one hand, and technology on the other, by a variety of educators, from peers to outsider professionals.

United Nations Educational, Scientific and Cultural Organization (UNESCO) has been at the forefront of the harnessing of community education for community, cultural, social, and personal development. The aim of meeting the learning needs of adults as well as children and youth, by 2015, is addressed through programs and targets on adult literacy, gender, and equality, human immunodeficiency virus/acquired immuno-deficiency syndrome (HIV/AIDS) education, and so on. This commitment to the use of community education as the means of addressing extremely difficult issues is evident in the examples outlined below.

In terms of the approach, the Australian agency, the Public Interest Advocacy Centre focuses on the use of information as the medium for community education. They supply a series of information leaflets, what Thompson (1996) terms really useful knowledge in their human rights community education. These information fact sheets are designed to reach out to people on the ground with background and strategies on civil, political, and human rights. The audience/learners include people from new communities, people who have come to Australia as refugee and asylum seekers, and other vulnerable people. They can access the information even from distant places, through the website.

In contrast, the indigenous community is the subject of another concrete example of community education, this time in Canada, in the Kehewin Community Education Centre. The overall objectives include the raising of awareness about the culture and practices of the Native Americans, in order to enhance the lives of both Native and non-Native communities. They undertake this through dance, theater as well as education. That is, the knowledge base is drawn from the traditional wisdom and practices within the Native community, and the learners are engaged in face-to-face creative processes.

The Grassroots Alliance for Community Education (GRACE), with its headquarters in the USA, works as a nongovernmental organization (NGO) in Africa, in a number of countries, with the objective of building the capacity of local people to take control over their own lives, by improving household health and welfare. Their processes include peer-educator training, which enables members of communities to relate to one another in an educational way, forming networks with local organizations. The agency trains local people, who use their knowledge to improve the health and welfare of families and households. Their interventions range from HIV/AIDS awareness, and consciousness raising about the illegality of female genital mutilation (FGM), to work with traditional birth attendants. That is, GRACE works in

specific African countries, at once removed from the grassroots, empowering local educators in order to gradually change traditional beliefs which endanger lives. Africa is also the source for the materials that many community educators have worked with for over 20 years, *Training for Transformation* (Hope and Timmel, 1985). These resources were inspired by the work of Freire, and are balanced between the really useful knowledge of social analysis on the one hand, with the creative methods and processes that have characterized community education ever since. These texts were located firmly in the grassroots of community participation, and this work parallels the educational initiatives in Latin America.

In Latin America, popular education fulfills the role of community education for adults, that is, education for grassroots, fundamental social change. Freire's work is evident in popular education and social movements, hinged on participative processes together with political and social analysis. For example, Gadotti (1992) shows the role of participation in popular consciousness raising, and also in the strengthening of community control over the state. This role for popular education/community education is a key conduit in bringing democracy to the continent with its troubled relationship with the USA. Interestingly, the GATEWAY project in the Washington DC area uses the community-education approach to enable recent immigrants from Latin America to the USA connect with people who have experienced multiple barriers such as lack of education and low income, using the full span of community-education-method processes to motivate and empower the learners on the issues around HIV/AIDS. That is, the purposeful intervention of community education is designed to sensitively address traditional beliefs in new communities in order to enable them to stay safe and healthy, through small group or one-to-one educational settings. The use of community education in the USA ranges from the life-and-death issue of health and welfare, to community problem solving.

With regard to problem solving, when community education was reviewed on the Internet, one of the first references to community education led the author to the Government of Wisconsin website. The Wisconsin Department of Public Instruction outlines its perspective, with a set of principles which underpins its version of community education. These principles include the principles of self-determination, self-help, social inclusiveness, and formal and informal learning, in order to meet the needs of local people, provided as close as possible to those who want it. This focus on self-determination and self-help also features in a community-education initiative in South East Asia, which straddles the Wisconsin principles, but with an environmental dimension that protects the earth's resources for the benefit of humanity, that is, macro-micro integration. Further, while many community-educational objectives include the sensitive challenge to

traditional beliefs, especially with regard to health, child-birth, and FGM, the Southeast Asia regional initiatives for community empowerment (SEARICE), promotes traditional knowledge of farming communities to protect them from the recent trends in agricultural production, with the attendant limiting of genetic diversity, and the copyrighting of seeds. The processes used in this community empowerment focus on information, lobbying, and advocacy. In contrast to the Wisconsin model, whereby the principles denote that community is a self-contained entity, a sub-set of the state, but autonomous and self-regulating rather than state controlled or supported, SEARICE sees the community as an integral dimension of the state – and the planet – but which needs to empower itself against the onslaught of the market. That is, the environmental concerns are held by the community, a huge burden in both the micro-sense of making a living, and also in the macro-sense of saving the planet.

On the other hand, the Scottish government locates the responsibility for community education firmly with the state, defining it as informal learning and social development, to strengthen communities by improving the capacity and knowledge of the community members. Further, it has prioritized community learning with the commitment not only to raise the educational attainments of adults, particularly literacy and numeracy, but also to support young people and to promote involvement with the planning and delivery of local services. Thus, the term community education is applied to the hands-on version of the Scottish government, to enhance democracy and participation, while the Wisconsin principles imply a hands-off approach, promoting self-sufficiency and independence. The Scottish Executive overtly locates community education within the social justice agenda, while Wisconsin government perceives community education as enhancing the community, improving services, and facilities, and applying not just to adults, and also schooling for children.

This overview of examples of community education on the ground, while superficial, illuminates the breadth of the field. It also shows the differences and contradictions, indicating the problem of delineation. The next section discusses these differences in order to establish the common ground in community education.

Contradictions and Ambiguities

Specific cases of community education have underlying contradictions, indicating the ideological understandings that underpin the practice. Tett (2002) discusses these inconsistencies with her discussion on the advantages and disadvantages of the hands-on state promotion of community education, with her reflections on the experience in Scotland. The key difficulty in the implementation of government policy is the high expectations of

community education to solve deep structural inequalities. This imposes unrealistic burdens on community educators and the communities they serve. Community educators are faced with a deep ambiguity. On the one hand, they are left with the responsibility for driving change at the most grass-roots levels, sometimes underfunded and undervalued, in comparison to other educational institutions. On the other hand, they remain quite invisible in the overall perception, seen as facilitators of change rather than actors and agents. Further, Tett contends, community educators cannot attain the same professional standards for their work that other educators can, because their role is diffuse, unspecified, and undefined (Tett, 2002: 12). This is all the more the case when the educators are members of the communities, subject to the barriers that the other members endure. The task of addressing social inequality is contingent on the re-distribution of power and resources, underpinned by the clarity of vision of the just society. Teaching egalitarianism also needs access to the body of knowledge that illuminates the causes of inequality, what Thompson (1996) calls really useful knowledge, that is, knowledge that enables people to understand the social forces that shape society. As the case studies show, sometimes the knowledge is traditional and honored, while at other times, traditional knowledge is dangerous, especially with regard to gendered health and welfare. Thus, access to really useful knowledge is problematic, and can be controlled according to the beliefs of the animators.

Moreover, with such high expectations of community education, the responsibility for fundamental change is also problematic. When the responsibility for justice-based change is left with the most powerless, it has very little prospect for success. For political leaders, advocating and supporting community education does convey the appearance of concern with addressing inequity, but in effect, the *status quo* prevails.

Further, in relation to the continuum between the local and the global, the globalized influences of economic and cultural capital impinge in an incalculable way on the individual, the community, and community educator. That is, the needs which communities identify, for example, drug misuse, unemployment, or environmental degradation or appropriation may have sources completely outside the control of the local level. The solutions to these issues reside at the global level; yet, there is an expectation that community education can armor people against them, say, by drug education, community enterprise, or ecological activism, on the one hand, or animate the people to act against them. Community education advocates indicate that local people, if they corral resources, if they network with like-minded groups, and if they develop leadership, can work to overcome the deep structural divisions of race and ethnicity. As a principle, the vision of the inclusive society is commendable; yet, the causes of divided communities are not just endemic

racism, ageism, and other discriminations, destructive as they are, but also globalized economic and cultural trends.

Thus, community education reaches difficult-to-reach people and communities, but not just that: it reaches difficult issues and trends, and as such, bears the burden of high expectations that it can actually resolve quite intractable social problems, as well as deal very effectively with other social issues. This raises the question about the nature of community, in which the community education operates. The next section discusses the connotations of community in this context.

Community Connotations

The term community is ambiguous and loose. Mayo (1994) contends that it is notorious for its shiftiness; yet, it is very useful for application to smaller sub-sets of society and to collectivities and communal dynamics. Williams (1976), in his discussion on the meanings and connotations of the term, traced the usage, originally referring to the common people, that is, peasants rather than people of rank. However, the connotation changed eventually to meanings closer to current usage, that is, the sharing of common characteristics and identity, and underpinned by relationship. Further, when the word is used, it conveys warm and persuasive sentiments. Thompson (2002) agrees with Williams that community has a feel-good factor associated with it that is difficult to undermine or challenge. She regards that community as a concept provides the space for security and common understandings. But she adds that community is frequently applied to others, that is, to poorer people, ethnic groups, and so on.

Toennies (1957) provided the discussion on *Gemeinschaft* and *Gesellschaft* which helped to develop insight into the characteristics of smaller social groups and looser associative relationships. These characteristics, though drawing on the contrasting traditional rural society versus urban industrialized society, nevertheless, enable the appreciation of experience of the postmodern city, and the large, isolated suburban enclaves in which huge populations endeavor to find common ground with the neighbors and common identities with like-minded people, to overcome alienation and loneliness.

The term community can also convey a more caring dimension than the larger institutional or social entity. Thus, in finding commonality, human relationships and constructions move into the central position, together with the sense of having some level of control over the immediate milieu. Harvey (1989) contends that different classes construct their sense of territory and community in radically different ways. He gives the example that middle classes can focus on tone, that is, control the status of the locality by ensuring that undesirable residents or developments are kept out. On the other hand, working

classes protect relationships, characterizing the quality of the community in terms of good, supportive, and present neighbors. Harvey is obviously concerned with class, primarily, but additional analysis shows that there are other differences including gender, race, and ethnicity. For example, Irish travelers refer to themselves as a community, with a strong emphasis on kinship, custom, and tradition, with a distinct culture that differentiates them from the rest of the populations. This is not static of course, but the influences from modern Irish society are mediated through the filter of the culture, rendering them encultured so that any new phenomenon takes on a distinct traveler flavor.

However, the term is very useful in community education, as it conveys its small-scale nature; the close relationships, including those of caring, inclusion, and supporting; its flexibility and shifting nature, particularly with regard to its responsiveness; the closeness of its provision to the learners and their contexts; and the ways in which it overcomes the estranging language that was more typical, such as outreach and extra-mural, liberal adult education, locally based adult education, or more critically orientated versions, such as emancipatory, popular, or empowering education.

Government of Ireland (2000) evaluates these characteristics, and endeavors to encompass the scope of community education by acknowledging that it reaches large numbers of people, often in disadvantaged settings. Community education also pioneers new ways of working with learning groups, in nonhierarchical processes. Finally, the lived experience of the learners provides the starting point for the learning. Thus, community education is framed as educational, in terms of processes and methods; communal, in terms of groups, both learning groups and community groups; and egalitarian, in terms of organization and responding to the needs of disadvantaged communities.

Conclusion

This article endeavored to capture the dimensions and facets of community education, not only delineating it, but also illuminating the demands and stresses that shape the practice on the ground. While the origins have a diffuse lineage, from the concern with workers' rights, to the desire to strengthen civil society, community education is subjected to a series of dialectical pulls and pushes, which ensures the dynamic, process-oriented development of the field.

Community education evolved with this complex, dynamic interaction of grass roots organic growth, and statutory or pioneering animation. However, regardless of the provision, the ownership of the process remains with

the participants. The community-centered approach ensures that learners participate freely, and the subjective experience of the participants is considered vital and transformative. Community education is located within the community and of the community.

Egalitarianism, which is very complex in itself, enables the learners to raise their consciousness about their own lives as well as the lives of other, engaged in the analysis of inequality. Freire's (1972) praxis connects the learning with activism, in a continuous cycle. The content of really useful knowledge (Thompson, 1996) contributes to the potential for societal transformation.

Bibliography

- AONTAS (2004). *Community Education*. Dublin: AONTAS.
- Bassett, M., Brady, B., Fleming, T., and Inglis, T. (1989). *For Adults Only: A Case for Adult Education in Ireland*. Dublin: AONTAS.
- Connolly, B. (2003). Women's community education: Listening to the voices. In *The Adult Learner, the Journal of Adult and Community Education in Ireland*. 2003 Dublin: AONTAS.
- Connolly, B. (1989). *Women's Studies in Lucan Daytime Classes 1985–1989: An Evaluation*. Unpublished Thesis.
- Crowther, J. (1999). Popular education and the struggle for democracy. In Crowther, J., Martin, I., and Shaw, M. (eds.) *Popular Education and Social Movements in Scotland Today*. Leicester: NIACE.
- Freire, P. (1972). *The Pedagogy of the Oppressed*. Harmondsworth: Penguin.
- Gadotti, M. (1992). Latin America: Popular education and the state. In Poster, C. and Zimmer, J. (eds.) *Community Education in the Third World*. Taylor and Francis.
- Government of Ireland (2000). *Learning for Life: White Paper in Adult Education*. Dublin: Government Publications.
- Harvey, D. (1989). *The Urban Experience*. Oxford: Blackwell.
- Hope, A. and Timmel, S. (1985). *Training for Transformation: A Handbook for Community Workers*. Gweru Zimbabwe: Mambo Press.
- Lovett, T. (1971). Community adult education. In Westwood, S. and Thomas, S. E. (eds.) *The Politics of Adult Education*. Leicester: NIACE.
- Mayo, M. (1994). *Communities and Caring: The Mixed Economy of Welfare*. Basingstoke: Macmillan.
- McGivney, V. (2000). *Working with Excluded Groups: Guidance on Good Practice for Providers and Policy Makers in Work*. Leicester: NIACE.
- Slowey, M. (1985). Education for domestication or liberation: Women's involvement in adult education. In Cullen, M. (ed.) *Girls Don't Do Honours*. Dublin: WEB.
- Smith, M. K. (2009). <http://www.infed.org/community/b-comed.htm>, accessed (6th March 2009).
- Smith, M. K. (1999, 2007). 'N. F. S. Grundtvig, folk high schools and popular education', *the Encyclopaedia of Informal Education*. <http://www.infed.org/thinkers/et-grund.htm>, accessed (6th March 2009).
- Tett, L. (2002). *Community Education, Lifelong Learning and Social Inclusion*. Edinburgh: Dunedin Academic Press.
- Thompson, J. (2002). *Community Education and Neighbourhood Renewal*. Leicester: NIACE.
- Thompson, J. (1996). Really useful knowledge: Linking theory and practice. In Connolly, B., Fleming, T., McCormack, D., and Ryan, A. (eds.) *Radical Learning for Liberation*. Maynooth: MACE.
- Toennies, F. (1957). *Community and Society*. Lansing: Michigan State University Press.
- Williams, R. (1976). *Keywords*. London: Croom Helm.

Further Reading

- Allman, P. (1999). *Revolutionary Social Transformation: Democratic Hopes, Political Possibilities and Critical Education*. Westport, CT: Bergin and Garvey.
- Apple, M. (1996). *Cultural Politics and Education*. New York: Teachers College Press.
- Barr, J. (2008). *The Stranger Within: On the Idea of an Educated Public*. Rotterdam: Sense Publishers.
- Coare, P. and Johnson, R. (eds.) (2003). *Adult Learning, Citizenship and Community Voices: Exploring Community-Based Practice*. Leicester: NIACE.
- Craig, G. and Mayo, M. (eds.) (1995). *Community Empowerment: A Reader in Participation and Development*. New Jersey: Zed Books.
- Crowther, J., Martin, I., and Shaw, M. (1999). *Popular Education and Social Movements in Scotland Today*. Leicester: NIACE.
- Duke, C. (2004). *Learning Communities: Signposts from International Experience*. Leicester: NIACE.
- Giroux, H. (1992). *Border Crossings: Cultural Workers and the Politics of Education*. London: Routledge.
- Habermas, J. (2004). *After Habermas: New Perspectives on the Public Sphere*. In Crossly, N. and Roberts, J. M. (ed.). Oxford: Blackwell.
- Horton, M. and Freire, P. (1990). *We Make the Road by Walking: Conversations on Education and Social Change*. In Bell, B., Gaventa, J., and Peters, J. (eds.). Philadelphia, PA: Temple University Press.
- Hooks, B. (2000). *Feminism is for Everyone: Passionate Politics*. London: Pluto Press.
- Hooks, B. (2003). *Teaching Community: A Pedagogy of Hope*. New York: Routledge.
- Mayo, P. (2004). *Liberating Praxis: Paulo Freire's Legacy for Radical Education and Politics*. Rotterdam: Sense Publishers.
- Mayo, P. (1999). *Gramsci, Freire, and Adult Education: Possibilities for Transformative Action*. London: Zed Book.
- Marx, K. and Engels, F. (1978). *The Marx-Engels Reader*. In Tucker, R. (ed.). New York: Norton.
- Poster, C. and Zimmer, J. (eds.) (1992). *Community Education in the Third World*. Taylor and Francis.
- Scottish Executive (2003). *Working and Learning together to Build Stronger Communities. Working Draft Community Learning and Development*. Scottish Executive, Edinburgh. http://www.infed.org/archives/gov_uk/working_together.htm, (accessed June 2009).
- SED (1975). *Adult Education: The Challenge of Change*. Edinburgh: HMSO.
- Tett, L. (2007). 'Working in and with Civil Society'. In Connolly, B., Fleming, T., McCormack, D., and Ryan, A. (eds.) *Radical Learning for Liberation 2*. Maynooth: MACE.
- Tight, M. (2002). *Key Concepts in Adult Education and Training*, 2nd edn. London: Routledge.
- Williams, S. (1994). *The Oxfam Gender Training Manual*. Oxford: Oxfam Publications.

Relevant Websites

- <http://www.infed.org> – Encyclopaedia of informal education.
- <http://www.graceusa.org> – Grassroots Alliance for Community Education.
- <http://www.telusplanet.net> – Kehewin Community Education Centre.
- <http://gateway.nlm.nih.gov> – NLM Gateway, A Service of US National Institutes of Health.
- <http://www.piac.asn.au> – Public Interest Advocacy Centre.
- <http://www.perfectfit.org> – Rage and Hope.
- <http://www.searice.org.ph> – Southeast Asia Regional Initiatives for Community Empowerment.
- <http://www.unesco.org> – UNESCO.
- <http://dpi.state.wi.us/fscp/ceprin.html> – Wisconsin Department of Public Instruction.

Continuing Professional Education: Multiple Stakeholders and Agendas

R M Cervero, The University of Georgia, Athens, GA, USA

B J Daley, University of Wisconsin – Milwaukee, Milwaukee, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

The leaders of professions and the public have always assumed that professionals would maintain their competence by continuing to learn throughout their careers through reading, discussions with colleagues, and educational programs. Since the 1960s, one of these educational forms, formal educational programs, has increased dramatically. Most professions now embrace the importance of lifelong professional education. In the rapid growth of continuing education (CE), educational leaders have generally relied for guidance and models on the distinctive knowledge base and structure of their own profession. However, many observers (Houle, 1980; Cervero, 1988; Brennan, 1990) noted the similarities of the CE efforts of individual professions in terms of goals, processes, and structures. Thus, the concept of continuing professional education (CPE) began to be used in the late 1960s. Various terms are used throughout the world referring to this concept, including continuing professional development, staff development, and professional learning. The rationale for this movement is that the understanding of similarities across the professions would yield a fresh exchange of ideas, practices, and solutions to common problems.

As with any educational system, CPE has many stakeholders with multiple agendas. The purpose of this article is to chart the expansion of CPE and three agendas that are shaping its present and future: professional and social, institutional, and educational agendas. Although the growth of CPE is a worldwide phenomenon, the most articulated systems exist in the nations of the Global North, such as Canada, Europe, Australia, and the United States. As such, CE and training have been addressed in global contexts such as the World Trade Organization and the General Agreement on Tariffs and Trade (GATT) (Lenn and Campos, 1997). This article accordingly focuses on the development of CE efforts in these regions of the world.

Systems of CE for the Professions

A central feature of societies in the twentieth century was the professionalization of their workforces. One estimate is that nearly 25% of the American workforce, for example, is classified as professionals (Cervero, 1988). These professionals include teachers, physicians, clergy, lawyers, social workers, nurses, business managers, psychologists, and accountants. Educational systems

have been a key feature of this professionalization project (Larson, 1977). An incredible amount of resources, financial and human, is used to support 3–6 years of professionals' initial education. Until recently, however, little systematic thought was given to what happens for the following 40 years of professional practice. Many leaders in the professions believed that these years of preservice professional education, along with some refreshers, were sufficient for a lifetime of work. However, with the rapid social changes, the explosion of research-based knowledge, the growing emphasis of evidenced-based practice, and spiraling technological innovations, many of these leaders now understand the need to continually prepare people for 40 years of professional practice through CE (Houle, 1980). Beginning in the 1960s, we started to see embryonic evidence for systems of CPE. Perhaps the first clear signal of this new view was the publication in 1962 of a conceptual scheme for the lifelong education of physicians (Dryer, 1962). The 1970s saw the beginning of what is now a widespread use of CPE as a basis for relicensure and recertification (Cervero and Azzaretto, 1990). By the 1980s, organized and comprehensive programs of CPE were developed in engineering, accounting, law, medicine, pharmacy, veterinary medicine, social work, librarianship, architecture, nursing home administration, nursing, management, public school education, and many other professions (Cervero, 1988). During this decade, many professions also developed their systems of accreditation for providers of CPE.

At the present time, the picture of an instructor updating large groups of professionals about the most recent theories and findings is easily recognizable as the predominant form of CE. We do not yet have a similarly recognizable picture of a system of CE in any profession. The major reason for this lack of a unifying picture is that the professions are in a transitional stage, experimenting with many different purposes, forms, and institutional locations for the delivery of CE. These systems, as such, are incredibly primitive and can be characterized as: devoted mainly to updating practitioners about the newest developments, transmitted in a didactic fashion by a pluralistic group of providers that does not work together in any coordinated fashion.

Relatively speaking, systems of CE are in their infancy. By way of analogy, CPE is in the same state of development as preservice education was at the beginning of the twentieth

century. It is unlikely that anyone in 1910 would have predicted the structure of medical education today. Likewise, systems of CPE are likely to grow through this transitional period to achieve an equivalent coherence, size, and stature as the preservice stage of professional education. While these systems of CPE are in a grand historical transition (Young, 1998), it is quite unclear what form they will take in the future.

Social and Professional Agendas

Within every profession, there are segments using knowledge for different social purposes. As Schon (1983: 345) argues, professionals' "special knowledge is embedded in evaluative frames which bear the stamp of human values and interests." It is not surprising that the professions have conflicting values about their role in society because professions are "loose amalgamations of segments pursuing different objectives in different manners and more or less held together under a common name at a particular period of history" (Bucher and Strauss, 1961: 326). For example, many social workers deliver services by means of individual casework. However, within the profession, some argue that casework is a form of conservative politics that reinforces institutions, processes, and ideologies that are destructive to human well-being (Galper, 1975). In this view, social work is considered conservative because its basic assumption is that nothing is wrong with societal arrangements, but rather with individuals who need to adjust to the status quo. The existence of internal dissension and value conflict is not new. For example, Perrucci (1973) found 18 radical movement organizations in 12 professions, including medicine, engineering, law, and psychology. These movements share the perspective that the professionals' role is to ask for whom and for what ends their expertise should be used. As professionals' knowledge can serve conflicting social purposes, CPE likewise can and does serve many different purposes.

CPE is being used more frequently to regulate professional practice. One of the major changes of the past 20 years has been the incorporation of CE into accountability systems for professional practice. As regulatory bodies struggled to develop accountability mechanisms, participation in CE was often the method of choice. However, these new requirements have been critiqued for "promoting the appearance of accountability but [doing] little or nothing to address the underlying issue of competence" (Queeney, 2000). In spite of this lack of demonstrable connection between CE and competent practice, the use of professionals' participation in CE to regulate their practice has not abated for the past two decades. Perhaps the most obvious example is the growth of state use of CE as a basis for relicensure. What started in the 1970s is now widespread such that every profession uses some form of

mandatory CE. The number of states requiring CE for relicensure has risen consistently for the past two decades. More professions are likely to follow the example of the Royal College of Physicians and Surgeons of Canada who have developed the Maintenance of Competence Program for recertification. This system allows physicians to use activities such as participation in audits of practice and a personal learning portfolio, which is a database of items of new learning recorded during the past year.

Institutional Agendas

Institutions develop the second part of the agenda, shaping present and future systems of CPE. Many different kinds of institutions have a stake in the growth of CPE systems, and each of these different institutions may have different goals and different methods by which CPE is provided. In discussing institutional agendas for CPE, four components are considered:

1. multiple goals,
2. multiple providers,
3. multiple modes of delivery, and
4. Collaboration among providers and professions.

Often the primary factor in determining the institutional agenda is the goals of the organization in which CPE is provided. Institutions tend to view CPE in different ways and most often these views are linked to the overall mission, vision, and values of the institution. For example, some institutions may view CPE as a method of employee development. Within these institutions it is believed that an educated and well-trained workforce is essential to the services provided by the institution. Healthcare, for example, most often focuses on employee development as a way to provide high-quality client care. The view here is that education assists employees to provide services that clients need. Thus, education or CPE is an essential link in the provision of institutional service.

Other institutions may see the provision of CPE as a method by which they can generate revenue. CPE is big business, especially in those states, countries, or territories that require CPE for recertification or relicensure of professionals. Higher education and professional schools, for example, are institutions that often rely on CPE as a way to generate revenue to support the overall mission of the institution. In these institutions, CPE is offered to large populations of professionals (often alumni) within the institutions' service area rather than to employees from one institution. As such, the revenue generated from day-long conference, multi-day conferences, or short courses returns to the more centrally managed CPE unit within the university or school.

The providers of CPE include the workplace, professional association, higher education, and for-profit companies. The workplace is by far the largest provider

of training and education for professionals and often this is accomplished through human resource development (HRD). Yet, each of these different providers offers unique CPE programs for specific audiences.

Dirkx and Austin (2005) help us understand the integrated nature of the multiple providers in CPE and the multiple goals of these providers. In their model of theoretical orientations in continuing professional development (Dirkx and Austin, 2005), they propose that the aims of professional development (based on Habermas, 1972) can range from technical, to practical, to emancipatory. Within the technical domain instrumental action, the scientific method and the hard sciences are addressed. In contrast, within the practical domain communicative action or human interest is most often the focus. Finally, the emancipatory domain focuses more on power and understanding our actions through self-reflection and self-knowledge.

In addition, Dirkx and Austin indicate that the goals of professional development are met in four primary contexts: HRD, CPE, faculty development, and staff development. Basically, these contexts represent the various areas in which professional development is most often provided. For example, HRD is usually provided as training in the workplace; CPE is most often provided through professional associations and higher education; faculty development is accomplished within colleges and universities; and staff development is associated with schools. What is evident in the Dirkx and Austin (2005) model is that the contexts of HRD, CPE, faculty development, and staff development may also overlap with different types of institutions. Finally, Dirkx and Austin (2005) indicate

that the focus of professional development may be individual or organizational. In their view, offerings are not exclusively individual or organizational but they tend to predominately focus on one or the other.

In analyzing developing systems of CPE, Dirkx and Austin (2005) offer a model that focuses on the multiple and overlapping providers and goals. Their model, depicted in **Figure 1**, demonstrates the complexity of currently developing systems of CPE. As Dirkx and Austin (2005) indicate, professional development can be conceptualized around “the overall aim of or purpose for professional development; the context of professional development; and the primary focus of the professional development activity” (p. 3). The model, depicted in **Figure 1**, promotes a new understanding of institutional agendas in CPE. Each cell in this model could be considered a different type of CPE offering. Dirkx and Austin indicate:

It is possible to take any particular context and identify different kinds of professional development that represent the different cells within that overall context. For example, one kind of faculty development program (context) might seek to develop specific skills (technical aim) in the use of a particular online software, so the professors can design and delivery more independently their own online courses (focus). This might be driven either by the organization’s need to develop additional revenue in budget-tight times (organizational focus), or individual professors may freely elect to develop this new expertise as part of a re-directing of their careers (individual focus) (Dirkx and Austin, 2005: 8).

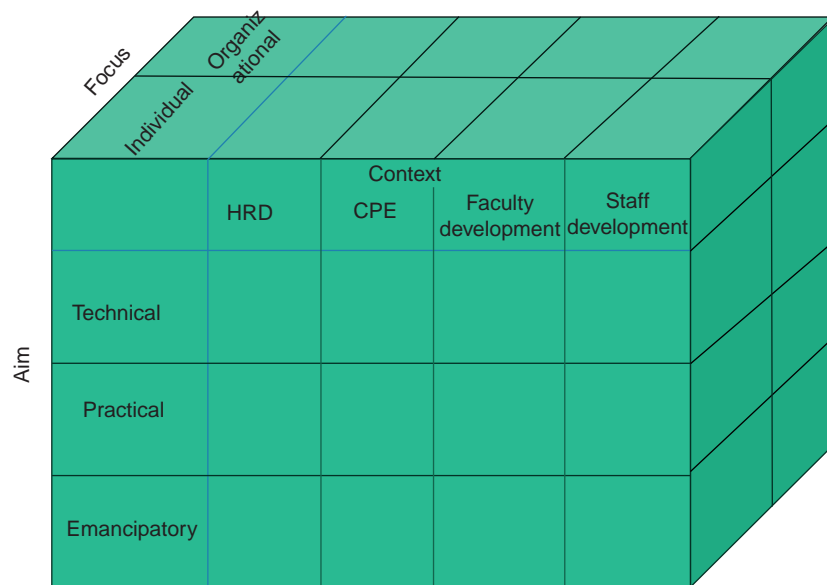


Figure 1 Dirkx/Austin model of theoretical orientations in continuing professional development. From Dirkx, J. and Austin, A. (2005). Making sense of continuing professional development: Toward an integrated vision of lifelong learning in the professions. Presented at the *Academy of Human Resource Development Continuing Professional Education Preconference*, Estes Park, CO, p. 8. 23 and 24 February, 2005, with kind permission from Dirkx and Austin.

Institutional agendas for CPE are also shaped by the multiple modes of delivery. CPE is predominately offered as limited, short-term educational programs presented in real time in a face-to-face format. These programs range from 1-day conferences, to multiday conferences, to short courses (i.e., 1–5 weeks). Most often, the mode of delivery for CPE focuses on an update of technical information so that the professional can return to their worksite with an increased understanding of new knowledge. However, increasingly distance education technologies are being incorporated into CPE programming. These technologies range from video-conferencing across multiple sites, to Internet-based courses, to Internet-based conferences, to online group discussion forums. Additionally, a relatively new form of distribution in CPE is learning objects. According to Lehman (2007) learning objects are “digital chunks of information” (p. 57). These digital chunks of information can be developed in the form of text, video, audio, or multimedia and they can be used to present case studies, new procedures, self-directed learning modules, or simulations. These smaller learning units can then be downloaded to hand-held devices making them easy to access and to share with other professionals.

Finally, collaboration among provider and professions is essential to present and future systems of CE. Collaborative arrangements among providers and across professions can not only help increase learning and understanding, but they can also help decrease the cost of providing CE. However, collaboration can raise issues within institutions and among providers. For example, providers of CPE at times feel that too much collaboration can demonstrate that their program is not needed. Once again, Dirkx and Austin (2005) offer us a way to consider collaboration. Within their model presented in **Figure 1**, areas of collaboration become evident. For example, if a healthcare employer decides that the aim of their educational programming will be technical, conducted within an HRD context at the organizational level, they may realize that they need to collaborate with universities and professional schools who can provide more emancipatory, CPE at the individual level. In this manner, different providers and different professions can analyze their own strengths and then collaborate with other providers offering different kinds of education using different methods. This type of analysis may draw on the strengths of organizations, prevent overlapping offerings, and streamline competition.

Educational Agendas

Building on professional/social agendas and institutional agendas, the educational agenda also impacts the continuing development of systems of CPE. However, for the educational agenda to be truly effective, we must include a model of learning (Cervero, 1988) at the center

of these educational agendas. As Eraut (1994) explained, behind “professional education lies a remarkable ignorance about professional learning” (p. 40).

Early models of professional learning have relied on the ideas of technical rationality (Houle, 1980), transfer of learning (Broad and Newstrom, 1992), and adoption of innovation (Rogers, 1995). In these views, knowledge for professional practice was created in one location, often a university setting, disseminated through CPE programs, and then transferred to or adopted in professional practice.

In the 1980s however, Cervero (1988) proposed a model for learning in the professions based on an understanding of how professionals “develop knowledge through practice” (p. 39). Cervero (1988) advocated that CPE providers develop a critical model of the learner that integrates the development of two forms of knowledge – technical and practical. Within this model, he incorporated components of cognitive psychology, reflective practice (Schon, 1987), and studies of expertise (Benner, 1984; Dreyfus and Dreyfus, 1985).

As systems of CPE continue to be developed, expanded models of learning will need to underlie educational offerings. Recent research (Daley, 2001a, 2001b) indicates that professionals construct a knowledge base for themselves in the context of their practice by linking concepts from new knowledge learned in education programs with their practice experiences. At this point, they actively make decisions on how to incorporate new knowledge into the context of practice based on their interpretations of the environment. What this newer research adds is an enhanced understanding of how learning in CPE occurs when professionals actively link new knowledge with the context in which they work and professional practice in which they engage.

As systems of CE continue to be developed, it is essential that education providers recognize this process of knowledge construction. Professionals do not simply take information from a learning situation and incorporate it in a practice context. They go through a much more complex process of taking in the new information, analyzing the context of their work site (including the politics and the people with whom they work), and then they decide if the new information is a good fit for their work based on the nature of their professional practice. This is a much more involved and sophisticated process than early professional learning models indicated. It requires that the educational agenda of CPE providers changes to incorporate a more active and integrated type of learning.

Additionally, an area that needs further development within professional learning is the question: How does the nature of professional practice connect to the educational agenda? We know that professionals construct a knowledge base for their practice and also know that the context of their work environment impacts the process. However, in addition, we need to understand how the nature of their

professional work impacts learning. For example, do social workers learn in a fundamentally different manner than lawyers? Recent work by Donaldson (2002) indicates that individuals within different professions fundamentally think in distinctive ways. Therefore, this raises the question: Does the nature of the professional's work drive the thinking and learning process? If this is the case, then how does that shape the educational agenda of CPE in the future?

Finally, the overall purpose of learning within the professions is designed to develop expert practitioners who provide high levels of service to clients. The study of professional expertise has grown from an understanding of serial problem solving (Newell and Simon, 1972) to an understanding of the stages of career development within specific professions such as physicists (Chi *et al.*, 1980), pilots (Dreyfus and Dreyfus, 1985), nurses (Benner, 1984), and physicians (Groen and Patel, 1988). These studies identified that as professionals gain experience they move from a novice to an expert practitioner. As professionals move from novice to expert, they develop the ability to think more abstractly and to see holistic patterns within their clients. They develop a sense of salience with the context that allows them to move from seeing individual issues to seeing systems and the integrated nature of professional problems. The study of expertise is now moving toward developing an understanding of the connections between learning and expertise. For example, Daley (1999) found that novice nurses tend to learn in a contingent manner, while expert nurses learn in a more constructivist manner. This type of research has implications for the educational agenda of CPE and necessitates that program planners consider the level of professional development of their audience. Finally, studies of expertise that focus on development of collective expertise across various contexts and disciplines are also needed. The idea of collective expertise within a profession or at a workplace is an area that will again shape the educational agenda of the future and will need more research.

In addition to professional learning models, evidence on the effectiveness of CPE is an important consideration in the educational agenda for CPE systems. When professionals attend CPE programs and learn new information, the question exists "Does it make a difference in the service provided to clients?" Early studies showed mixed results in response to this question, with some studies indicating little impact and other indicating great impact. To clarify this variability in the results of evaluation studies, Umble and Cervero (1996) conducted a meta-analysis of 16 studies that provided a synthesis of impact studies in CPE. A major purpose of this review was to analyze and critique the methodologies of the studies, but in addition they found that a first wave of syntheses studies existed that demonstrated a general causal connection between

CPE and impact on professional practice. Umble and Cervero (1996) found that CPE had an impact on knowledge, competence, performance and outcome.

In addition, Umble and Cervero (1996) found that a second wave of studies had begun to demonstrate a number of variables that moderate the impact. This second group of studies began identifying the types of programs that can promote performance change within the professions. Umble and Cervero (1996) indicate that as the development of CPE systems moves forward, more work is needed in understanding the impact of CPE on practice. They advocate for more experimental studies, more qualitative and mixed-methods studies, and the use of action research (Brooks and Watkins, 1994) as both a programming and evaluation methodology.

The educational agenda in current systems of CPE is thus shaped by the model of professional learning incorporated in education practice, by an understanding of the nature of expertise, and by the overall effectiveness of the impact of CPE programs. In the future, CPE systems will continue to be shaped by ongoing and expanding research in all three of these areas.

Conclusion

The task of building systems of CPE is fundamentally more complex than what faced leaders as they built the existing systems of preservice professional education. First, whereas preservice education takes place in a relatively short period of time, CE must help professionals for 30–40 years of professional practice, which is characterized not only by constant change, but also often by competing values. Second, while preservice education is predominately controlled by universities and professional schools, there are multiple institutions that offer CE, all of which stake a claim to being the most valid and effective provider. The leaders of workplaces, professional associations, universities, and governments have both a tremendous opportunity and a clear responsibility to further develop the systems of CE for the professions. As with any humanly constructed system, the building of a coordinated system of CE for any profession is a political process. While we cannot predict what these systems will look like, this process will be marked by fundamental struggles over the educational, institutional, and professional agendas and the competing interests of the multiple stakeholders for CE. As a political process, then, it is crucial that all of the stakeholders participate in a substantive way in negotiating these agendas for CE. The immediate and long-term negotiation of these struggles will define whether CE can make a demonstrable impact on the quality of professional practice.

Bibliography

- Benner, P. (1984). *From Novice to Expert: Excellence and Power in Clinical Nursing Practice*. Menlo Park, CA: Addison-Wesley.
- Brennan, B. (1990). *Continuing Education Promise and Performance*. Victoria, Australia: The Australian Council of Educational Research.
- Broad, M. L. and Newstrom, J. W. (1992). *Transfer of Training*. MA: Addison-Wesley.
- Brooks, A. and Watkins, K. (1994). The emerging power of action inquiry technologies. *New Directions in Adult and Continuing Education* **63**, 5–16.
- Bucher, R. and Strauss, A. (1961). Professions in process. *American Journal of Sociology* **66**, 325–334.
- Cervero, R. M. (1988). *Effective Continuing Education for Professionals*. San Francisco, CA: Jossey-Bass.
- Cervero, R. M. and Azzaretto, J. F. (1990). (Eds.), *Visions for the Future of Continuing Professional Education*. Athens, GA: The University of Georgia.
- Chi, M. T. H., Feltovich, P. J., and Glaser, R. (1980). Categorization and representation of physics problems by experts and novices. *Cognitive Science* **5**, 121–152.
- Daley, B. (1999). Novice to expert: An exploration of how professionals learn. *Adult Education Quarterly* **49**(4), 133–147.
- Daley, B. (2001a). Learning in clinical nursing practice. *Journal of Holistic Nursing* **16**(1), 43–54.
- Daley, B. (2001). Learning and professional practice: A study of four professions. *Adult Education Quarterly* **52**(1), 39–54.
- Dirtx, J. and Austin, A. (2005). Making sense of continuing professional development: Toward an integrated vision of lifelong learning in the professions. Presented at the *Academy of Human Resource Development Continuing Professional Education Preconference*, p. 8. Estes Park, CO, 23 and 24 February, 2005, with kind permission from Dirtx and Austin.
- Donaldson, J. (2002). *Learning to Think: Disciplinary Perspectives*. San Francisco, CA: Jossey-Bass.
- Dreyfus, H. and Dreyfus, S. (1985). *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.
- Dryer, B. V. (1962). Lifetime learning for physicians: Principles, practices and proposals. *Journal of Medical Education* **37**(6), 1–134.
- Eraut, M. (1994). *Developing Professional Knowledge and Competence*. Washington, DC: Falmer Press.
- Galper, J. H. (1975). *The Politics of Social Services*. Englewood Cliffs, NJ: Prentice-Hall.
- Groen, G. and Patel, V. (1988). The relationship between comprehension and reasoning in medical expertise. In Chi, M. T. H., Glaser, R., and Farr, M. J. (eds.) *The Nature of Expertise*, pp 287–312. Hillsdale, NJ: Lawrence Erlbaum.
- Habermas, J. (1972). *Knowledge and Human Interests*. Boston, MA: Beacon Press.
- Houle, C. O. (1980). *Continuing Learning in the Professions*. San Francisco, CA: Jossey-Bass.
- Larson, M. S. (1977). *The Rise of Professionalism: A Sociological Analysis*. Berkeley, CA: University of California Press.
- Lehman, R. (2007). Learning object repositories. In Conceição, S. (ed.) *Teaching Strategies in the Online Environment*. *New Directions for Adult and Continuing Education*, vol. 113, pp 57–66. San Francisco, CA: Jossey-Bass.
- Lenn, M. P. and Campos, L. (1997). *Globalization of the Professions and the Quality Imperative: Professional Accreditation, Certification, and Licensure*. Madison, WI: Magna Publication.
- Newell, A. and Simon, H. A. (1972). *Human Problems Solving*. Englewood Cliffs, NJ: Prentice Hall.
- Perrucci, R. (1973). In the service of man: Radical movements in the professions. *Sociological Review Monograph* **20**, 179–194.
- Queeney, D. S. (2000). Continuing professional education. In Wilson, A. L. and Hayes, E. R. (eds.) *Handbook of Adult and Continuing Education: New Edition*, pp 375–391. San Francisco, CA: Jossey-Bass.
- Rogers, E. M. (1995). *Diffusion of Innovations*. New York: MacMillan/Free Press.
- Schon, D. A. (1983). *The Reflective Practitioner*. New York: Basic Books.
- Schon, D. A. (1987). *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*. San Francisco, CA: Jossey-Bass.
- Umble, K. and Cervero, R. (1996). Impact studies in continuing education for health professionals: Critique of the research syntheses. *Evaluation and the Health Professions* **19**(2), 148–174.
- Young, W. H. (1998). *Continuing Professional Education in Transition: Visions for the Professions and New Strategies for Lifelong Learning*. Malabar, FL: Krieger.

Further Reading

- Bereiter, C. and Scardamalia, M. (1993). *Surpassing Ourselves: An Inquiry into the Nature and Implications of Expertise*. Chicago, IL: Open Court.
- Cervero, R. M. (2001). Continuing professional education in transition, 1981–2000. *International Journal of Lifelong Education* **20**(1–2), 16–30.
- Cervero, R. M. and Scanlan, C. L. (eds.) (1985). Problems and prospects in continuing professional education. *New Directions for Adult and Continuing Education*, 27. San Francisco, CA: Jossey-Bass.
- Daley, B. (2000). Learning in professional practice. In Mott, V. and Daley, B. (eds.) *Charting a course for continuing professional education: Reframing professional practice*. *New Directions in Adult and Continuing Education*, vol. 86, pp 33–42. San Francisco, CA: Jossey Bass.
- Daley, B. and Jeris, L. (eds.) (2004). Boundary spanning: Expanding frames of reference for human resource development and continuing professional education. *Advances in Human Resource Development* 6(1). Thousand Oaks, CA: Sage Publications.
- Moon, J. (2003). *Reflection in Learning and Professional Development: Theory and Practice*. London: Kogan.
- Nowlen, P. M. (1988). *A New Approach to Continuing Education for Business and the Professions: The Performance Model*. New York: Macmillan.

Financing of Adult and Lifelong Learning

Å Sohlman, Ministry of Employment, Stockholm, Sweden

© 2010 Elsevier Ltd. All rights reserved.

Adult learning is a wide-ranging phenomenon and accordingly its financing. It engages individuals, organizations, communities, and nations. It is pursued for a wide variety of purposes and in a wide variety of ways.

Theoretical foundations for principles of financing adult learning, as well as other types of activities, will be found in the literature of economics. In that literature, market failures are often highlighted as a motive for government funding of learning. However, a more fundamental concern has been social justice and distributional aspects. The market approach has been criticized for being too restrictive.

While economic principles may remain unchanged over time, political practice has to focus on the varying problem areas as they turn up, more recent adaptations to globalization, the knowledge economy, new information and communication technologies, and global warming. Therefore, in empirical practice and studies, different financing mechanisms will eventually be brought to the fore.

This article offers a few introductory comments on financiers and funding statistics. Financing principles and mechanisms are reviewed. In addition, the development of practical financing experiences in the Organization for Economic Co-Operation and Development (OECD) countries is described as an illustration to general funding problems in many countries.

Financiers and Funding Statistics

In most countries, the government is an important financier of adult learning. However, in federal states, not only will there be national governments, but also regional and local governments, each with its special financial responsibilities. Furthermore, within a government often several ministries are involved in the financing of adult learning, typically the Ministry of Education (formal learning including literacy), the Ministry of Labor (training for the unemployed), and the Ministry of Social Affairs (training for other disadvantaged groups).

Most of the public funds have traditionally been channeled to public learning institutions or sometimes to nongovernmental organizations (NGOs). The latter often receive funding from many different sources, public and private, including fees from participants. This sector includes much nonformal learning as exemplified by the folk high schools and study circles in the Scandinavian countries.

The social partners influence adult learning in different ways. An important factor can be collective agreements which regulate employee training possibilities, as in Germany and Austria. Besides that, employers, of course, engage in training their employees often purchased from private providers.

Needless to say, fundamentally, individuals remain responsible for their own education and training, not least the self-directed learning.

Some Statistics

Comparative statistics of expenditure on adult learning are not available. Not even for individual countries do complete statistics exist. (For participation in adult learning comparative statistics do exist (cf. e.g., Kailis and Pilos, 2005 for European Union (EU)-25 participation in both formal, nonformal and informal learning, and the OECD 2006 for nonformal job-related education and training). The most developed financing statistics concern labor market training and private enterprise training (OECD, 2006; EU-RA, 2004). The EU is preparing an adult education survey for 2008, including private costs for formal and nonformal adult education (SCB, 2007). Still, two financial aspects will be illustrated in this section: to what extent are individuals interested in financing their learning, and do governments fund adult training for the labor market.

Private financing

In 2003, the lifelong learning Eurobarometer charted the attitudes of individuals toward financing their own learning. The survey was directed to individuals aged 15 plus in the old EU-15, Iceland, and Norway (Ramprakash *et al.*, 2005).

Overall, 43% of the respondents were willing to pay for their own education and training. This average varied from 33% to 68%. The awareness of the importance and utility of adult learning thus greatly differs even between these European countries. Respondents in Denmark and Iceland were most prepared to invest in learning. Other countries with high figures were Germany, Sweden, and the United Kingdom. In contrast, respondents in Belgium, France, Spain, and Portugal were less willing to pay for their training.

Making a distinction between education for work-related and nonwork-related purposes, on average, there were no great differences. The same countries came out with high figures for work and nonwork training: Iceland,

Norway, and Luxemburg. In case of nonwork, Sweden and Denmark should also be added to this group. For both purposes, a middle group included Germany and Ireland (and Sweden for work purposes). Low investors in both cases were again Belgium, Spain, and Portugal.

Judging from these financial statistics, much more of a learning culture seems to exist in some countries, such as the Nordic countries of Iceland, Denmark, Norway, and Sweden, than in other countries, such as Belgium, Spain, and Portugal, with Germany, the United Kingdom, and Ireland in an intermediate position.

Public financing

For public financing of adult learning, there are OECD statistics for employment training. In 2001, the OECD (2004) concluded that the funds for adult education would, in view of their comparable small share of total education enrolments, at best equal labor market training expenditure. In the late 1990s and 2000, expenditure in support of labor market training ranged from 0.20% to nearly 1.00% of gross domestic product (GDP) in 9 of 25 countries for which data were available.

The 2007 edition of *OECD Employment Outlook* contains financial statistics for as many as 28 countries for the years 2003–05. The figures for public expenditure on training as a percentage of GDP are now spread over:

a low range 0.01–0.05;
a middle range 0.10–0.25; and
a high range 0.30–0.54.

Overall, the financing is now lower than in the 1990s. Moreover, if there are any changes over the 3 years covered in these data, they point in the direction of a lowering of expenditure.

In the low range, we find not only the Central and Eastern European countries such as Poland, the Slovak Republic, the Czech Republic, and Hungary, but also Mexico, Greece, Australia, Japan, Korea, and the United States. In the middle range are not only a number of European countries such as Luxemburg, the United Kingdom, the Netherlands, Spain, Belgium, Italy, and Ireland, but also Canada and New Zealand. On top of the list are the Nordic countries: Finland, Sweden, Norway, and Denmark. Other countries with rather high expenditures are Switzerland, Germany, Austria, France, and Portugal.

Many people seem to be interested in financing their adult learning – in some countries even a majority. Whether they do so or could be stimulated to do so we do not know. However, private interest in financing adult learning and actual government spending on training seem to go hand in hand (cf. the Nordic countries), but at a rather low expenditure level. Only a very small share of GDP is devoted to public spending on learning for the labor market.

Few aspects of the funding of adult learning are covered by statistics. This lack of financial statistics must be a drawback for the development of rational adult learning policies.

Efficiency Principles and Financing Mechanisms

In economic literature, there are two main reasons for government funding of adult learning: distribution and efficiency motives (Barr, 2001; Bohm, 1987). Arguments for governments to support a social infrastructure, including one for adult learning, have also been added (North, 1981; Rosenberg and Birdzell, 1986).

Efficiency

The efficiency, or allocation and economic growth motive, for public funding of adult learning, through individuals, institutions/organizations, or employers, is based on the idea of market failures. For a number of reasons – positive external effects, uncertainty, lack of information, and competition – individuals and firms acting on their own on markets would not achieve enough learning from a socioeconomic point of view. These market failures are often supposed to be more difficult when it comes to basic skills and less so for higher-level competences.

The government may consider financing socially profitable learning investments, that is, investments that are profitable when also social benefits are included. Privately profitable investment might be left to private investors. For tertiary education and certain adult training for the labor market, the arguments in favor of public funding are thus not so strong. Individuals and employers, to a large extent, reap the benefits from such training and should therefore bear at least part of the costs. For basic skills, the opposite is true; the main benefits of having a basically trained and trainable population and labor force are largely social.

Education and training for the unemployed, disabled, and other groups that have problems getting a foothold on the labor market can also be motivated by efficiency reasons. It may be better for society to help unemployed people to find a job. In this case, there are no opportunity costs in terms of foregone production during training.

Some arguments have also been put forward to motivate government funding of training in small- and medium-sized enterprises (SMEs). Much similar to individuals, they may not be aware of the benefits of training and existing learning opportunities and have problems financing training.

Sometimes governments also have to handle market distortions they have created themselves. Taxes introduced for financial reasons may, for example, inadvertently hamper investments in learning. Besides, the government has to

carefully choose financing mechanisms not to introduce new distortions, efficiently targeting disadvantaged groups, not crowding out private investments, and avoiding fraud. Selective measures may target specific groups and be cost efficient, while general, more all-inclusive measures are more costly but leave no one out.

Distribution

The distribution motive, for one thing, is also associated with basic learning. In a democratic society, all children should have access to at least primary schooling and in modern societies secondary education is also more or less a prerequisite for citizenship. In the same vein, it can be argued that public funding of second-chance education for adults at these levels is motivated.

In addition, a democratic society should arguably aim at making cultural and humanistic values accessible to all citizens.

From a viewpoint of social justice, investing in young and old people would be valuable. From an efficiency perspective, the government had better focus funding on the learning of young people with a longer time period for reaping the benefits of training.

Infrastructure

The importance of the infrastructure for economic transactions has been highlighted, generally and also in case of learning. It includes basic laws and regulations as well as attitudes and institutions. If one is to rely on markets, there has to be market regulations and surveillance of competition fostering economic security and efficiency in adult learning markets. Information about learning opportunities has to be available. To be able to develop certificates, a national qualification framework may be necessary. Teacher training as well as research and development are also important.

This is not to say that the government itself has to provide all these services. The duty of the government is to ensure that they are available and adequately provided.

The economic literature has produced arguments for certain learning activities to be financed by the government but no exact amounts are prescribed. How to balance investments motivated by social justice, distributional, efficiency, and infrastructure effects against each other and against other types of public expenditure is a fundamentally political or even philosophical question.

Financing Mechanisms

Whatever the purpose of the spending, financing mechanisms have to be carefully chosen according to efficiency properties (OECD, 2000, 2001, 2004, 2005).

Increasing efficiency in direct public funding of institutions

Traditional public service relies on bureaucratic hierarchies, confidence, and consensus. The mechanisms that have recently been introduced in public education and training services to increase efficiency belong:

- different types of incentives, for example, in the form of output-based funding to gain more value for money;
- decentralization of financial decisions, giving more autonomy to education and training institutions that are supposed to have better knowledge about local conditions; and
- competition for government contracts or stimulation of public institutions to rely on tenders – the requirement to buy education and training services from different providers may be linked to infrastructure in terms of lists of publicly accredited institutions and programs.

Education and training institutions may also be permitted to use fees to finance their provision. This can be a way of testing individual demand for learning. For distributional reasons, however, some types of courses may be exempted from fees and some groups of individuals may have a right to tuition-free courses.

Efficiency through partnership and co-financing

Financial co-sharing arrangements can be an efficient way to investigate whether there is any private demand and willingness to pay for certain services. At the same time, the public sector can benefit from private information as to production and distribution alternatives for these activities. Examples are requirements of co-financing from individuals, employers, or municipalities for subsidies to be paid to education and training projects, and formal public private partnerships (PPPs) to finance investments.

Voluntary co-financing schemes can also be introduced mainly to raise private funds for learning, then normally requiring a subsidy element. There may be three-party versions – government/employer/individual – or two-party versions – government/employer or government/individual. Examples are individual development accounts (IDAs) and individual learning accounts (ILAs) with matched contributions by individuals, employers, and the government and publicly supported train-or-pay systems for employers.

Employers can also be stimulated to create sector- or branch-specific funds to finance common training. Such pooling of employer resources for training can be of special interest to SMEs and also an efficient way to handle poaching (in addition, or as an alternative, to pay back clauses in employment contracts).

Parafiscal funds are compulsory measures to raise funds and may, for example, be created by payroll levies on employers, eventually combined with public funding; they are an alternative to taxes. Some part of the funds

may be used by the employer himself and some collectively (such as the levy/grant schemes). An additional purpose for such funds may be to raise awareness as well as to stimulate employer interest in training.

Individuals

For individuals, two types of learning costs are involved: direct costs and costs of living. Public incentives for individuals to train – to make up for market failures and handle distributional problems – take many different forms in addition to co-financing schemes:

- *Individual drawing rights, entitlements, or training vouchers.* In these cases, both direct and indirect costs may be covered. These incentives can be targeted and constructed so as to stimulate competition among education and training providers. Such arrangements are often combined with an infrastructure of information, advice, and counseling services sometimes at learning centers where the individual also can have access to accreditation of prior learning, a library, and education at a distance.
- For direct training costs, tax exemption and tax brakes can be used to stimulate participation though, of course, only individuals with a taxable income can profit from such measures.
- For indirect costs, grants and loans, mixed loan–grant models, or income-contingent loans (ICLs) are often available. The ICLs mean that the repayments of the loans are related to future incomes, which reduce individual risks.
- A legal right to training leave where the costs (direct and/or indirect) are supported by the government and/or the employer (with or without tax brakes for employers) also exists in some countries. Training leave without at least (partial) funding may not be very successful in raising participation.

Training for unemployed people is often free of charge while the individual can live off his or her unemployment benefits. Such training may be compulsory though the individual can have some choice as to actual training.

Employers

Public incentives for employers to undertake training resemble those for individuals. Vouchers, tax exemptions, tax credits, profit tax deductions, and subsidies have been used to stimulate employers to train their employees or make them benefit from training leaves. The idea is that this type of government-supported education and training should have wider purposes than the training the employer would normally be responsible for – basic skills training, non-work-related training, and general rather employer-specific training. The government may pay both training cost and wage compensation to the employers.

Aspects of the learning infrastructure that may interest the employers are easy access to information about training alternatives and flexible training providers.

Historically, government funding of adult learning has been supply oriented, that is, directed toward learning providers in various segments (e.g., further education, labor market training, and culture). Aspects of demand-led, market-oriented learning have been introduced to make the public providers more efficient and involve private providers in traditionally public education and training.

The market-oriented measures are supposed to increase efficiency through a better match between demand and supply and lower costs. Such measures, however, risk introducing segmentation and creaming in training markets reduce access to learning. In tendering, it may be difficult to balance equity, quality, and efficiency.

However, private co-financing has also been introduced to stimulate an overall growth in adult learning investments in light of shrinking public funding.

Learning markets rely on individual choice and competition. On the one hand, this means empowerment for individual actors. On the other hand, it means an alienation from an idealistic search for learning based on a public learning infrastructure.

Policy and Practice in OECD Countries

In 1996, the OECD launched the concept referred to as lifelong learning for all (OECD, 1996). It was motivated by major trends such as globalization, trade liberalization, the aging of the population, the growing ethnic and cultural diversity, the impact of information and communication technologies, and the changing nature of work. It was to address the fundamental objectives: personal development, social cohesion, and economic growth.

Closing the Gaps

A first analytical step was to try to estimate the existing gaps and the costs involved in closing them (OECD, 1996). For adults, this proved to be very difficult but a few examples were given. Scenarios for extending lifelong learning to adults with low literary proficiency, based on the literacy surveys of International Adult Literacy Survey (IALS) and available figures for training costs, were used. For Sweden, the costs were estimated to be around 4% of GDP, in the Netherlands 5%, and in Germany between 6% and 12%.

More cost calculations were produced later on (OECD, 2000). This time they were based on country reports and participation targets. The Netherlands came up with a well-documented assessment. Now closing the participation gaps for adults was estimated annually to require public costs corresponding to 0.3% of GDP.

Hungry made calculations both for public and private costs. For public costs, they were estimated annually to reach the level of 2.8% of GDP and, for private cost, 7.6% of GDP.

There seemed to be something like a natural progression. When targets for initial, primary or pre-primary, secondary, and tertiary education had been reached, lifelong learning for adults might be a reasonable target. For countries that had not achieved that stage, the challenges were daunting. For the central and eastern countries, these were also newly (re)discovered needs in response to the transition to a market economy.

Financing Lifelong Learning

Later OECD (2001) publications concentrated on how to make strategies for lifelong learning for adults affordable. Different ways of reducing costs, increasing benefits, and putting incentives to work were scrutinized. Still, the OECD ministers were convinced that the targets could be reached and in 2001 they declared that more resources had to be raised (OECD, 2004).

A project Financing Lifelong Learning was started. Participating countries submitted background reports, were visited by OECD review teams that wrote country reports, and produced material for further discussion.

However, by the mid-1990s it was evident that, although formal education systems were expanding, they were not able to ensure lifelong learning for all. One important shortfall, besides provision for young children, concerned adults, in particular those with low levels of qualifications.

The recurrent education of the 1960s and 1970s had never become a widespread practice. Training markets had not developed adequately in spite of government initiatives to stimulate supply and demand for adult learning. Neither training levies or collective agreements had spurred supply nor had vouchers for individuals or levy-exemption schemes for employers, noticeably stimulated demand.

Co-Financing

Now the OECD narrowed down the perspective. Lifelong learning for adults should not be financed but co-financed (OECD, 2003b, 2004).

Fiscal pressure in the era of global competition made for less public resources and they alone could not provide the necessary funding for lifelong learning. A consequent search for new models of financing investment revolved around the issue of co-financing by individuals, governments, and employers. Various co-financing mechanisms are available for different aspects:

1. *For reducing direct costs to individuals.* ILA and IDA (with matched contributions by individuals, employers, and the government), loans and vouchers to individuals and/or employers (subsidized by the government), and tax policy (deductions, credits, tax-sheltered savings supported by the government).
2. *For reducing individual foregone earnings.* Direct income support (from the government), ILA and IDA (with matched contributions), loans to individuals (subsidized by the government), collective agreements, and time accounts (with or without subsidies by the government).
3. *For sharing risks.* ICLs to individuals (subsidized by the government).

Much of the co-financing discussion centered around ILA and IDA (OECD, 2004). A special network, the European Learning Account Partners (ELAP), prepared a catalog of recent lifelong learning co-financing initiatives (OECD, 2004). The ILA was to be used only for learning purposes, while the IDA could be used for different purposes, including education and training. There were two main arguments in favor of ILA and IDA. On the one hand, they made for three-party co-financing, and thus the possibilities to raise resources might improve. On the other hand, they fitted a desired development away from supply-driven toward demand-driven learning, a strategy with individual needs at center that might motivate poorly qualified and disadvantaged learners to participate in learning.

Besides co-sharing arrangements, another concept was also launched and much discussed at this time, a whole-of-government approach. If resources were to be used efficiently, adult learning policy had to be developed in a coordinate way and government ministries cooperate. The Swedish Adult Education Initiative was mentioned as an example of both co-financing and cooperation as these important learning projects were co-financed by different ministries (OECD, 2004).

However, the ILA never left the preparatory stadium. In the United Kingdom and the Netherlands, they were tried but abandoned. In Sweden, the government had set aside funds for their financing, but in the end no decision to introduce them was taken. In the United States of America, some states have introduced IDA, but they have not become general government policy. (A private insurance company tried to make a product out of competence insurance (OECD, 2004) but it failed.)

There were many concurring problems attached to the ILA. As they generally were supposed to have a wide coverage, they were expensive to governments and at the same time only contained marginal subsidies to individuals. Employers also seemed reluctant to invest in savings accounts that they did not fully control; and as regards less qualified individuals, would they really be interested in these accounts and stimulated by them to undertake training?

Promoting Adult Learning

Parallel to the upcoming failures of co-sharing arrangements, the OECD initiated the Thematic Review of Adult Learning. It ran over the years 1999–2004 and resulted in two publications (OECD, 2003a, 2005) with the telling titles, *Beyond Rhetoric* and *Promoting Adult Learning*. Here, we are back to basics: how to stimulate adult learning especially for the least qualified. This now seemed to be the most pressing challenge in response to globalization and new technologies, given the restrictions to public spending. Equal distribution of skills has a strong impact on overall economic performance, making the labor forced more productive.

The focal point is on:

- financial incentive mechanisms and on policies to increase the participation of low-skilled adults;
- financial constraints to participation by low-skilled adults; and
- institutional arrangements that are conducive to investment by firms and individuals.

Structural preconditions for learning – visibility of rewards, recognition of prior learning, national qualification systems, information, advice and counseling, flexible providers, and one-stop centers – are important to increase the participation of low-skilled adults.

Still, the context of policy fragmentation and the need for an integrated approach to adult learning policies, of course, have to be kept in mind. However, a more realistic aim might be advisory bodies and coordinators suggesting priorities rather than actual whole-of-government policy-making bodies.

The grand vision of lifelong learning for all, where adult learning was identified as an important missing link, was an overarching government approach to achieve everything from social cohesion and inclusion of marginalized groups, to jobs for the unemployed and continuous upgrading of the workforce, to competitiveness in response to globalization and cultural and democratic development, including a responsible reaction to global warming. For adult learning, it has turned into a more realistic, piecemeal approach. Reforms cannot advance ahead of institutional prerequisites, political support, and coherent strategies.

The lessons learned by the OECD countries have proved to be relevant to the post-communist countries in Central, Eastern and Southeast Europe (Gunny and Viertel, 2006) and are certainly also relevant to countries such as Brazil, India, China, and middle-income developing countries. Adult learning has to be tackled to reduce underemployment and informal employment (OECD, 2007).

See also: Adult Literacy Education; Evaluation and Accountability.

Bibliography

- Barr, N. (2001). *The Welfare State as the Piggy Bank: Information, Risk, Uncertainty and the Role of the State*. Oxford: Oxford University Press.
- Bohm, P. (1987). *A Concise Introduction to Welfare Economics*, 2nd edn. London: Macmillan.
- EU-RA, European Research Associates (2004). *Exploring the Sources on Funding for Lifelong Learning*. Brussels: European Commission, Education and Culture DG.
- Gunny, M. and Viertel, E. (2006). *Designing Adult Learning Strategies*. Turin: European Training Foundation.
- Kailis, E. and Pilos, S. (2005). Lifelong learning in Europe. *Statistics in Focus 2005/8*. Brussels: European Communities, Eurostat.
- North, D. C. (1981). *Structure and Change in Economic History*. New York: W.W Norton.
- OECD (1996). *Lifelong Learning for All*. Paris: OECD.
- OECD (2000). *Where Are the Resources for Lifelong Learning?* Paris: OECD.
- OECD (2001). *Economics and Finance of Lifelong Learning*. Paris: OECD.
- OECD (2003a). *Beyond Rhetoric*. Paris: OECD.
- OECD (2003b). *Strategies for Sustainable Investment in Adult Lifelong Learning in Education Policy Analysis 2003*. Paris: OECD.
- OECD (2004). *Co-Financing Lifelong Learning: Towards a Systematic Approach*. Paris: OECD.
- OECD (2005). *Promoting Adult Learning*. Paris: OECD.
- OECD (2006). *Education at a Glance*. Paris: OECD.
- OECD (2007). *Employment Outlook 2007*. Paris: OECD.
- Ramprakash, D. and Wadolowska, E. EU-RA, European Research Associates (2005). *Private Household Spending on Education and Training*. Brussels: European Commission, Education and Culture DG.
- Rosenberg, N. and Birdzell, L. E., Jr. (1986). *How the West Grew Rich: The Economic Transformation of the Industrial World*. New York: Basic Books.
- SCB (2007). *Befolkning och välfärd. Tema: utbildning. Vuxnas deltagande i utbildning*. Rapport 2007 nr 2.

Further Reading

- Axelsson, R., Westerlund, O., Bjelkeby, Å., et al. (2005). *Vuxenutbildningens betydelse för inkomster, rörlighet och övergång till högskolestudier*. Stockholm: ITPS.
- Ikei (2004). *A catalogue of recent lifelong learning initiatives 2004*. Prepared in cooperation with the OECD. <http://www.oecd.org> (accessed May 2009).
- Kunskapslyfskommittén (2000). *Kunskapsbygget 2000 – det livslånga lärandet, SOU 2000:28*. Stockholm: Ministry of Education.
- OECD (1999). *Education Policy Analysis 1999*. Paris: OECD.
- OECD (2001). *Education Policy Analysis 2001*. Paris: OECD.
- OECD (2006). *Boosting Jobs and Incomes. Policy Lessons from Reassessing the Jobs Strategy*. Paris: OECD.

Relevant Websites

- <http://www.ec.europa.eu> – European Commission, Eurostat.
- <http://www.eu.org> – EU.org, free domain names.
- <http://www.etf.europa.eu> – ETF, Sharing Expertise in Training.
- <http://www.ilo.org> – International Labour Organization.
- <http://www.oecd.org> – Organization for Economic Co-operation and Development, Financing Adult learning, Home Page; Thematic Review of Adult Learning, Main Page.
- <http://www.worldbank.org> – The World Bank.
- <http://www.uis.unesco.org> – UNESCO Institute for Statistics.
- <http://www.unesco.org/uis> – UNESCO, UIE.
- <http://www.unesco.org> – United Nations Educational Scientific and Cultural Organization.

Labor Education

K Forrester, University of Leeds, Leeds, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

For many hundreds of thousands of people around the world, learning through their trade union has been an important, and occasionally, pivotal experience in their development and understanding as workers and citizens. Whether as a form of occupational socialization, self-improvement, or as a vehicle of social mobility, union education has been and remains, an important albeit uneven feature of trade union activity around the world. While most studies and reports focus on worker education's engagement with the knowledge, skills, and capabilities seen as necessary to survive within late capitalist economies, other (less-documented) recent experiences can be found of the union learning necessary in the challenge to apartheid regimes (South Africa) and forms of political injustice, authoritarianism, and dictatorship (Central Africa or South Korea). In the case of post-communist countries, union members are learning once again, to create organizations that defend worker interests within the workplace.

Paradoxically, the nature and organization of this educational activity remains an under-researched and documented area, barely visible within the voluminous literature centered on labor organizations. Few people would disagree with the Croucher (2004) view that "Education is often viewed as a key instrument for effecting change in trade unions" (p. 90). Despite such sentiments, labor education remains as a minority research concern and a marginal policy focus within labor organisations themselves (see, e.g., the recent country studies in the volume *Organised Labour in the 21st Century* (Jose, 2002) by the International Labour Office). Other commentators interested in union engagement with neoliberal globalization, such as Novelli (2004: 165), report that while "education and learning appear crucial . . . they are largely absent in the literature". To some extent, this absence is understandable (and prudent) given the formidable conceptual, organizational, and research complexities inherent in any overview of union learning activities. Labor, union, and education, for example, all remain the focus of considerable debate and contestation. Empirically and analytically capturing this education is fraught with difficulties especially in the rapidly changing nature of this provision as unions make and remake themselves in different historical settings and in response to changes in the employment relationship and wider societal context.

Despite these substantial methodological caveats, this article provides a selective flavor of the conceptual

schemas and organizational variety of union education in different parts of the world. It is suggested that recent labor education initiatives can be seen as being informed by one or more of three interrelated concerns, namely, the possible contribution of education toward union renewal and revitalization strategies, the emergence of lifelong learning as a key policy focus and third, changes in regulatory regimes resulting from more intensive capital-accumulation strategies.

The first section situates the peculiarities of worker education. Unlike most other forms of knowing, union education is characterized by its contradictory nature – its simultaneous promotion of resistance and of cooperation. Analytically grasping this tension at the center of worker education has proved difficult and controversial. A brief review of the formulations and conceptual frameworks used commonly in discussions of union education is provided. In contrast to these formulations, it is suggested that in recognizing the nature and extent of worker learning, union education needs to be situated within the wider societal context. This is the focus of the second section in the article. This wider societal approach provides the analytical focus for contextualizing particular initiatives at particular times in particular circumstances. Examples of concerns within union education are used to illustrate the argument developed in this section. The final part briefly considers the educational implications for unions within at least rhetorically, the information society, knowledge economy, or the post-industrial society.

The Peculiarities of Union Education

Trade unions are best understood as institutions that primarily mediate between their members and opposing social interests such as the state and employers. As a condition of their survival, they are required to navigate the material structural contradictions between capital and labor. In seeking to defend and further the interests of their membership, union relationships and practices are characterized simultaneously not only by struggle and conflict but also by cooperation in the maintenance of an orderly employment environment and favorable societal contexts. The effectiveness of worker "resistance to capitalism . . . also makes this resistance more manageable and predictable and can even serve to suppress struggle" (Hyman, 1989: 230). It is this key contradictory characteristic of unionism that not only unlocks understandings of

unionism but also provides a rich interpretative basis for framing union education. As such, it provides a start to overcoming static institutional categories commonly used in discussion of education in general. Situating union learning within this wider societal context provides the critical dimension intrinsic to this type of education.

Scholarly attempts to situate analytically, trade union education or more broadly, worker education within the general literature acknowledges the difficulties inherent in such an ambition; there is wisely a strong cautionary element. Hopkins in his international survey of workers' education, for example, warns of the historical variety and culturally contested understandings of education, workers, and trade unions: there "are grave dangers of over-simplification and confusion" (Hopkins, 1985: 8). Newman agrees and notes that "the world of union affairs, of workplaces, of industrial relations is a complicated and confusing one. No two unions, no two workplaces are the same. No issues are identical" (Newman, 1993: 11). The International Labour Office (ILO) recognizes the complexities and offers a broad understanding. "Worker's education", it suggests, "is designed to develop the workers' understanding of 'labour problems' in the broadest sense of these words" (ILO, 1976: 10). It "should always be regarded as a means to useful action. In many cases the education will make clear both the need for action and the best forms the action can take" (ILO, 1976: 10). Spencer (2002: 17–18) in his collection on labor education provided by trade unions goes further when distinguishing between tools or role courses (those preparing members for active roles in the union), issue or awareness courses (linking workplace to societal issues such as racism, union campaigns, or new management techniques) and labour studies courses (examining union contexts through historical, economic, and political perspectives). Such a taxonomy – common in the Anglo-Saxon literature – approximates to Hopkins' longer fivefold classification of the major components of the curricula of worker' education. These he sees as basic general skills; role skills; economic, social, and political background studies; technical and vocational training; and finally, cultural, scientific, and general education (Hopkins, 1985: 43). The situating of trade union education or workers' education within perspectives that focus on curricula, program design, and pedagogy (worker-centered) has contributed toward framing important discussions about the particularities of this type of learning. Implicit within many of the discussions are a number of distinguishing features that shape not only the content but also the language, values, pedagogy, relationships, objectives, and nature of the learning experienced in the education. Although not always made explicit in the written learning materials, the examination and linkages between union experience, discussion, knowing, and action in programs occur within a number of sensibilities that are powerfully

shaped by notions of social justice. More recent commentators on union education have attempted to move beyond the classificatory approach and explicitly incorporate wider societal concerns within definitions and understandings of union education. Burke and her Canadian union educators, for example, identify a number of threads which "hold together the fabric of our work: community, democracy, equity, class-consciousness, organisation-building, and the greater good" (Burke *et al.*, 2002: 3). Other commentators such as Salt *et al.* (2000) distinguish between transformatory and accomodatory education in their discussion of worker education and neoliberal globalization (p. 9).

These more recent attempts to adequately capture the complexities of union education have been accompanied by the exploration of understandings of education and learning. In doing so, it has been recognized that becoming a trade unionist involves much more than a consideration of formal educational categories, as any historical account of trade unionism demonstrates. An emphasis solely on curricula considerations risks privileging institutionally driven categories at the expense of cultural and material considerations. An unproblematic view of education emphasizes education as a product to be passively acquired or consumed and transferred. For some trade unions, education assumes this quality of technical rationality, as Schon (1983) puts it. By contrast, more recent approaches to understanding education such as sociocultural approaches (Lave and Wenger, 1991) and activity theory (Engestrom, 2001) stress the contextualized, mediated, and participative processes of, in this instance, becoming a trade unionist. Livingstone and Sawchuk for example, recognize the extensive and varied circumstances and complex processes within which union learning occurs. "Studies of adult education", they argue "have often ignored the actual array of learning activities of working people and generally implied inferior learning capacity" (Livingstone and Sawchuk, 2003b: 111). Studies by these authors have focused on union activity at the local workplace in different occupational settings in Canada in the exploration and documentation of "distinctive working-class learning practices across multiple spheres of activity, including paid employment, housework, both union-based and community volunteer work as well as private learning concerned with general interests". Using a conception of learning informed by a cultural-historical perspective, the authors analytically distinguish between formal education, continuing education, and informal learning (see also Sawchuk, 2003; Livingstone and Sawchuk, 2003b).

Novelli's (2004) study of social movement learning shares a similar concern with exploring and developing expanded understandings of worker education and learning. His analysis of the public service trade union SINTRAECALI in the south west of Colombia outlines the recent successful campaigns against privatization of

public utilities. Employing the notion of strategic learning, Novelli portrays the transformation of the union from a narrow corporate trade union focused on the defense of members' particular interests, to a social movement union that linked workers and local communities in the defence of public services (Novelli, 2004: 161). The union's occupation of the company's 17th-floor administrative building in Cali is conceptualized as an educational outcome which resulted from several years of "strategic learning through social action" (Novelli, 2004: 162–163). The study illustrates the benefits of questioning dominant understandings of education and of situating these understandings within a particular material and societal context.

Situating Union Education

Uncovering the significance of the learning experienced by union members or educational participants then, can be helped through the contextualizing of the learning within the particular political, cultural, and societal circumstances at a particular time and in particular parts of the world. Stirling (2002) uses such an approach in his survey of union education in Europe. In the Nordic countries, Germany and the UK for example, the settlements with social democratic governments provide the basis for historically understanding the comparatively extensive provision, institutional and financial support for union education. In other European countries such as Belgium and Netherlands, these settlements and support have been riven by cultural, linguistic, and religious divisions resulting in marked ideological differences to the education of the rival centers. In these instances where rival confederations are characterized by their socialist, catholic, or communist perspectives, "education programs are needed to reinforce the identity of particular Confederations as against others and to transmit an ideology to leading cadres" (Stirling, 2002: 27).

Situating union education and the unions themselves within this wider political context helps in addressing issues of significance, relevance, and purpose of the learning. Ost (2002) uses the example of unions in Poland to rebuke gently, Western commentators for reaching generalizations about union development and education based on their own particular circumstances, concerns, and frameworks. As he points out, in a number of countries, the new eastern-European unions energetically campaigned for the creation of a capitalist system and played an important role in educating workers about their responsibilities and obligations as wage employees. The early years of Solidarity in Poland in the 1980s, for example, were almost exclusively concerned with large successful mobilizations around societal issues to the exclusion of workplace concerns. The success of these Solidarity activities in Poland led to other unions, for

example, in Bulgaria and Hungary, seeking to replicate this model. However, the lack of attention given to the workplace by Solidarity resulted in dramatic membership losses in the 1990s. Withdrawal by Solidarity from the governing coalition in 2001 together with a greater emphasis on old-fashioned workplace unionism (recruitment, education courses, employer agreements, and workplace representation) is expected to reverse the period of decline experienced in the 1990s.

Unions, where possible, have always engaged educationally with employability concerns such as occupational training and apprenticeships. Although such role courses have not figured significantly in Anglo-American discussions of union education due to the frequent involvement of employers and state agencies, Lopez's account of the Brazil unions' struggle and development for the Programa Integrar (the integration program) is illustrative. Occupational training is used to include wider cultural and societal concerns by the national unions in the struggle against neoliberal initiatives. Civil society involvement and alternative economic plans are pursued through the establishment of local development forums (Lopez, 2002). In other instances such as Eastern Europe, mainstream educational courses have been designed and used to effect fundamental qualitative change (Croucher, 2004: 92) within the unions themselves. Focusing on particular trade unions in Moldova, Ukraine, and Belarus, Croucher reports on educational initiatives to assist union development in the move from Soviet-style organizations to organizations able to meet new demands in new circumstances.

The continuous search by unions through negotiation, struggle, and industrial and civil conflict for a possible and politically acceptable settlement with the state and employer agencies provides a constant tension not only within unions themselves but also in the available educational opportunities. Cooper (2002) illustrates these ambiguities in her evaluation of the Development Institute for Training, Support and Education for Labour (Ditsela) from its launch in 1996 in post-apartheid South Africa. Ditsela, as she reports (Cooper, 2002: 37), was created by the congress of the South African Trade Unions (COSATU) as a solution to "rebuilding the labour movement's capacity to respond to the major changes underway in the country There was a strong belief that a major educational initiative was required to build COSATU's capacity to play a proactive role in the new, democratic South Africa".

In many other countries, political accommodation is not possible. In many countries, the absence of pluralistic political cultures or the existence of repressive state regimes excludes often violently, recognition and legitimization of union practices. In parts of Africa and South America, such conditions characterize a number of countries; being a trade unionist can be very dangerous. Alexander's personal evaluation of "the courage and dedication of so many educators" in confronting "the brutal application of neo-liberalism and

imposed structural adjustment programmes” in Zambia and Zimbabwe over a number of decades, provides a stark reminder of circumstances in much of the world (Alexander, 2006: 595). In such volatile political contexts, union education has been continued and aided through external sympathetic agencies. The Commonwealth Trade Union Council, for example, in 2002, was involved in supporting Sierra Leone trade unionists in a civil society program with a strong focus on HIV/AIDS, campaigning against child labor in collaboration with unions in Botswana, Mozambique, Namibia, and Tanzania, and in encouraging women participation in Bangladesh. Similarly, financial support was provided to the Ghanaian trade unions educational work from the Netherlands, the Commonwealth TUC, and from the International Confederation of Free Trade Unions (ICFTU). The Trades Union Congress in Ghana has prioritized the education of members as an important part of its difficult struggle against structural adjustment policies. Training at the Labour College covers three broad areas: union organizational issues including health and safety, trade union history, and finally, special programs. In Niger, the ILO since 1999 has been supporting the two national union confederations in the provision of worker’s education in the informal sector and the French trade unions are involved in the development of a health insurance project. Maruatona’s study of adult education and trade union development in Botswana is situated against a background of significant constraints imposed by the state. Education he suggests, is an important vehicle in “helping address the worker’s problems” (Maruatona, 1999: 476).

Whether in Africa or elsewhere, labor organizations endeavor to ensure the initiation and reproduction of union leadership together with the necessary membership skills and capacities necessary to enter or maintain participation in the labor market. The employability agenda historically remains an important educational focus. The contexts and possibilities within which such agendas are pursued, however, differ significantly. In those regions of the world with strong repressive regimes, small formal economic sectors, and small union membership, worker learning is more likely to occur through forms of popular education (Kane, 2001), social action (Folley, 1999), or through participation in nongovernmental activity as Prieto and Quinteros’ (2004) study of women’s organizations and trade unions collaboration in the free-trade zones of Central America illustrates.

Union Learning in the Knowledge Economy

Today, it seems trade unions everywhere are in crisis. Whether in the Northern or Southern countries, falling membership, loss of bargaining power, new management

strategies, or the need for new workers in the new workplaces, labor organizations are seen as less relevant or significant than previously. Post-Fordism, post-industrial, new-information age, the knowledge economy, or post-Taylorist is but a selection of the formulations used to distinguish the new from the old. Globalization, however, is the dominant framework that usually uncritically is used theoretically to legitimate these formulations. Simply put, globalization has resulted in the international crisis of labor organizations. This was and is not a universal experience as recent struggles for basic democratic rights in, for example, South Africa, Brazil, and Korea illustrate (Kelly, 2002). However, while such deeply contested perspectives are beyond the focus of this article, there is a widespread consensus that particular worker organizations are struggling to successfully engage with the implications of the massively increased international mobility of capital.

From a trade union perspective, an important dimension to recent union revival or renewal discussion is the increased importance given to the education function. An extensive body of evidence, for example, is beginning to emerge in the UK around recent attempts to refocus trade unions as centers of learning (Healy and Engel, 2003). Learning to organize through the training and networked organization of workplace, union learning representatives (ULRs), and the creation of workplace learning centers are seen for British unions, as an important feature of engagement with the new employment agenda and with membership growth. Similarly, training for worker representation on the European Works Councils or the creation of new training academies or institutes in Australia, the USA, and Britain (Spencer, 2002) is illustrative of the increased importance being given by unions to the training function.

The rise rhetorically at least, of human resource strategies with the accompanying focus on knowledge management then, has resulted in the exploration of fresh educational developments as a means of accommodating to, or contesting with, the new circumstances. The increasing influence in employer (and international agencies such as the World Bank and International Monetary Funds) thinking and practices of human capital theory has resulted in a greater emphasis on employer–union partnership educational initiatives. Time, financial resources, and encouragement from employers are available in joint union educational initiatives that focus on strengthening employee corporate commitment, the promotion of soft social skills, and enhanced vocational training opportunities. At a regional level, there is usually an emphasis on union education that addresses more macro-, policy-informed subject areas. The educational provision of the European Trade Union Confederation, for example, has courses on foreign languages, freedom of movement of labor and work regeneration as well as more traditional union organization, recruitment, and leadership courses.

From a different perspective, there have been important recent international discussions and developments around what has been termed the new internationalism and social movement unionism (Moody, 1997; Munck, 2002). As Novelli (2004: 166) notes, "education is a key process for contestation in the struggle to develop an (alternative) movement and incorporate new allies." There is a need, she continues, to move beyond instrumental understandings of education and training toward conceptions and practices that situate research, investigation, and learning in the formulation of alternative strategies that are seen as part of a wider political project. Such sentiments historically have formed the basis of a consistent critical perspective in the understanding, discussion, and evaluation of union education and learning. The current defensiveness of trade unions internationally in the face of the neoliberal offensive not only reinforces the search for alternatives and new allies, but also increases the pressure for more instrumental educational solutions. It is likely that this contradictory dynamic will continue to shape union learning in the period ahead.

Conclusions

As most commentators recognize, analysis and understandings of union learning need to be contextualized within the changing nature of the unions themselves and the wider societal circumstances within which unions make and remake themselves in differing historical settings and in different parts of the world. However, irrespective of the particular place or forms that this learning might take, it remains true that for many people this learning provides an important (and sometime the only) means for an understanding of, and participation in, collective activities against social, economic, and political injustices. Despite the current difficult political circumstances, the increased attention in recent years to conceptualizing and documenting the ubiquitous nature of everyday learning incurred as a trade unionist, woman worker, or street trader has enhanced the importance and resources given to education by many union organizations.

Bibliography

- Burke, B., Geronuno, J., Martin, D., Thomas, B., and Wall, C. (2002). *Education for Changing Unions*. Toronto, ON: Between the Lines.
- Cooper, L. (2002). Union education in the new South Africa democracy. In Spencer, B. (ed.) *Unions and Learning in a Global Economy: International and Comparative Perspectives*, pp 37–49. Toronto, ON: Thompson Educational Publishing.
- Croucher, R. (2004). The importance of trade union education: A study in three countries in Eastern Europe. *European Journal of Industrial Relations* 10(1), 90–109.

- Engestrom, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualisation. *Journal of Education and Work* 14(1): 133–156.
- Folley, G. (1999). *Learning in Social Action*. London: Zed.
- Healy, J. and Engel, N. (2003). *Learning to Organise*. London: Congress House, TUC.
- Hopkins, P. G. H. (1985). *Workers' Education: An International Perspective*. Milton Keynes: Open University Press.
- Hyman, R. (1989). *The Political Economy of Industrial Relations: Theory and Practice in a Cold Climate*. Basingstoke: Macmillan.
- Jose, A. V. (2002). *Organised Labour in the 21st Century*. Geneva: International Institute for Labour Studies, International Labour Office.
- Kane, L. (2001). *Popular Education and Social Change in Latin America*. London: Latin America Bureau.
- Kelly, J. (2002). *Union Revival – Organising around the World*. London: Congress House, TUC.
- Lave, J. and Wenger, E. (1991). *Situated Learning. Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Livingstone, D. and Sawchuk, P. (2003a). Hidden Knowledge: Organised Labor in the Information Age. Washington, DC: Rowman and Littlefield.
- Livingstone, D. and Sawchuk, P. (2003b). Hidden knowledge: Working-class capacity in the "knowledge-based economy" *Studies in the Education of Adults* 37(2), 110–122.
- Lopez, A. M. (2002). *Programa Integrar in Brazil: Union intervention in employment, development and education*. In Spencer, B. (ed.) *Unions and Learning in a Global Economy: International and Comparative Perspectives*, pp 120–129. Toronto, ON: Thompson Educational Publishing.
- Maruaton, T. (1999). Adult education and the empowerment of civil society: The case of trade unions in Botswana. *International Journal of Lifelong Education* 18(6), 476–491.
- Moody, K. (1997). *Workers in a Lean World: Unions in an International Economy*. London: Verso.
- Munck, R. (2002). *Globalisation and Labour*. London: Zed Books.
- Newman, M. (1993). *The Third Contract: Theory and Practice in Trade Union Training*. Sydney: NSW: Stuart Victor Publishing.
- Novelli, M. (2004). Globalisations, social movement unionism and new internationalisms: The role of strategic learning in the transformation of the Municipal Workers Union of EMCALI. *Globalisation, Societies and Education* 2(2), 161–190.
- Ost, D. (2002). The weakness of strong social movements: Models of unionism in the eastern European context. *European Journal of Industrial Relations* 8(1), 33–51.
- Prieto, M. and Quinteros, C. (2004). Never the twain shall meet? Women's organisations and trade unions in the *maquila* industry in Central America. *Development in Practice* 14(1 and 2), 149–157.
- Salt, B., Cervero, R. M., and Herod, A. (2000). Workers' education and neoliberal globalization: An adequate response to transitional corporations? *Adult Education Quarterly* 51(1), 9–31.
- Sawchuk, P. (2003). *Adult Learning and Technology in Working-Class Life*. Cambridge: Cambridge University Press.
- Schon, D. A. (1983). *The Reflective Practitioner*. New York: Basic Books.
- Spencer, B. (ed.) (2002). *Unions and Learning in a Global Economy: International and Comparative Perspectives*. Toronto, ON: Thompson Educational Publishing.
- Stirling, J. (2002). Trade union education in Europe: Emerging from the gloom. In Spencer, B. (ed.) *Unions and Learning in a Global Economy: International and Comparative Perspectives*, pp 26–36. Toronto: Thompson Educational Press.

Further Reading

- Holland, C. and Castleton, G. (2002). Basic skills and union activity in the UK and Australia. In Spencer, B. (ed.) *Unions and Learning in a Global Economy: International and Comparative Perspectives*, pp 89–99. Toronto: Thompson Educational Press.
- Stard, A. (1998). On two metaphors for learning and the dangers of choosing the first one. *Educational Researcher* 28(7), 4–14.

Learning Cities and Regions

C Duke, University of Leicester, Leicester, UK; University of Stirling, Scotland, UK; RMIT University, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Engaged university – A university which plans and connects its teaching and research work with the needs and communities of the locality – city or region – in which it is located.

Learning city/learning region – An administrative region that has developed the capacity to collect, analyze, and use data and experience to enhance the quality of its understanding, leading to improved practice.

Place management – Application of the recognition of the importance of the concept and reality of place for effective, integrated, and acceptable governance.

Social construct – The understanding, concept, and definition developed by a community or network of policymakers, scholars, and practitioners about a concept and/or activity which draws them together.

The Learning City and the City-Region as a Social Construct

The learning city is an ideal, rather than a description of any actual place or places. It is an aspiration for the way the city might be better managed and manage itself in a complex world characterized by terms such as global and knowledge economy. The terms learning region and city-region are used also as a way of thinking about the management of a city area or region with an evident identity, boundaries, and some form of government. It reflects a sense that city and regional governance could and should be different, and better than they presently are. Local administrations in many parts of the world have in recent years expressed the intention to be learning cities; a number of them have announced that this is what they are, will be, or are becoming. In the sense that the learning city is a social and political aspiration rather than a well-defined and understood condition with agreed characteristic measures, it is best understood as a social construct.

The learning city is a new and evolving notion. Although urban and regional studies have existed for decades, the term learning city, or city-region, emerged only in the later years of the twentieth century. There is no established definition of what it means. In terms of

literature, it falls between several recognized areas of scholarship to each of which it is peripheral and ambiguous. It is best understood as heuristic, a metaphor which expresses a set of values and purposes, not an established concept in academic social science. The term is used mainly in policy arenas and among communities of practitioners who are concerned with how things can be done better.

This article explains the different ways the term is used. It then explains how it is related to issues of government rather than education. Various learning-city initiatives are referred to before considering research in the area and issues related to implementation. After a brief consideration of the important link with higher education, the article concludes by assessing the utility and the possible future of the concept.

Two Different Meanings

Although not hotly contested, the term carries different meanings and is used in different senses. Despite the word learning which locates it in the mindset of education, the richer and more fundamental sense has more to do with governance and politics than with education. Where a government authority is charged with the subject it may be a ministry of education, where it is seen as mainly related to schooling, education, and training.

In principle, there are two distinct levels of meanings to the concept of learning city or city-region. The easier-to-understand term implies an urban or regional authority that makes good provision for many people to learn. It sees learning, essentially in the forms supported by high-quality and widely accessible education and training, as crucial to economic productivity, competitiveness, and so to civic success. A UK report commissioned by the Department for Education and Employment in the late 1990s had the concept of the Educating or Learning City originating in an international conference convened by the Barcelona City Council in 1990, then focused in a report to the OECD 2 years later (Hirsch, 1992).

The policy agenda originating from this approach concerns increasing the volume and raising the quality of education, training, and learning opportunities for individuals. This may extend beyond the formal education sector, together with the less-formalized or nonformal adult and lifelong learning arena, to foster a culture of

lifelong learning. Occasionally it includes museums, libraries, and other arts and cultural facilities and venues seen as part of a wider civic learning environment. It is unlikely to go beyond this although learning in and through work (work-based and workplace learning) is recognized within the education profession and policy community as a significant dimension of education and training relevant to living and working.

The second, deeper and richer, as well as more historically accurate meaning concerns the capacity of a city or region itself to learn almost as an individual person or other organism can learn, understand, and adapt its behavior. There may be a difficulty in a Western democratic and individualistic tradition about accepting the notion of learning outside the individual. The recognition has been the strongest and longer established with respect to the learning organization as an informing principle for the study of organization behavior, for consulting to management, and in management education. Subsequently, learning has been affixed to many other institutions and phenomena, from schools and universities themselves to other kinds of organizations such as hospitals and banks to events such as festivals. In a geographical and political sense, it has been applied at levels from the nation to a small locality such as a village.

A widely used term with more conceptual underpinning than many of these is the learning community, used mainly with a sense of geographical locality and also sometimes of virtual communities. Linked to this are the terms and notions of communities of interest and communities of practice (Wenger, 1998). However, these concepts lack the component of place integral to the learning city-region: the recognition that an area with its inhabitants or citizens constitutes an entity or reality, cultural, social, and historical as well as economic, political, and geographical. The term place management has acquired currency in recent years to emphasize the importance of shared location, environment, and experience as important elements in how we organize, manage, and govern ourselves.

The terms learning city, region, and city-region have stronger validity and utility in this second and larger sense. The reduction in scope and meaning in relation to education and training may appeal to local and regional authorities wishing to be part of a new wave of thinking about governance; and it may enable or assist them to reinvigorate their education sectors by using a large and ambitious term.

The larger concept has more to do with the nature of governance and the capacity of political systems at whatever level to learn from their own and others' experience and to adopt new behaviors, possibly using concepts such as double-loop learning and triple helix to focus recognition of what this means. Insofar as the concept demands introspection and reflexivity, it may be radical and unsettling,

connected as it is to ideas about participation, devolution of authority, and empowerment. Perceived in this way, the learning city is about politics and government rather than about education.

Cities and Regions, Politics and Governance

In terms of fields of study, the learning city region belongs more to politics and government, geography, and urban and regional studies than to education. It is unavoidably political, having to do with the holding, sharing, and use of power rather than purely technical and organizational. Some of the terms with which it is associated have a heuristic, reformist, radical, or an ideological flavor and intent. In this sense, the idea of a learning city represents a challenge to current methods of governing and exercising authority and power. Looking at the administrative region as a place inhabited and used by different groups and communities of interest, and wishing to improve the quality of government so as to enhance the level of and capacity for learning, raises questions about participation as well as consultation in the process of government, and about the devolution of different kinds of decisions and control. The city-region may be seen as a contested space where interest groups vie for benefit and control. The style and nature of government, as well as the amount and nature of devolved authority, are called into question.

The concept of a learning region or territory can apply at all levels of government, although the main focus tends to be on the city. The term city-region is commonly used to refer not just to the metropolis, metropolitan area, or town but also to its physically, economically, and/or culturally natural territory or catchment area. A problem arises when these natural regions do not correspond with the local and regional authority boundaries, and also when the local and regional government are unstable and subject to central government intervention, with frequently changing boundaries, powers, and dispensations.

Here the idea of a learning city-region becomes entangled with another agenda: regionalization as a means of decentralizing and devolving power from the larger central state. Devolution is common in many countries; but the consequential required transfer of powers and resources does not always occur. Then a city or region may be unable to manage its resources and affairs so as to act on what it learns.

Problems occur for schools, colleges, and education systems when control and responsibility are divided among different levels of jurisdiction. They are more acute for a whole city or region. A major problem for effective learning applied to better governance is the separation and compartmentalization that is common among various functional departments, sometimes known

as silos, at whatever level. This is exacerbated when there is division, and the silo walls are strong between the parts of the administration, both vertically and horizontally.

Learning City Initiatives

There have been many recent initiatives to promote and disseminate the idea and practice of a learning community and region, for example, through projects funded by the European Union (EU). Some have gone beyond discussion of the ideas and processes to practical workshops and manuals on how to go about it, and how to equip local authority staff to enable local-level community learning as a part of the process of better – more participatory, more responsive and learningful – government (Longworth, 2006; Longworth and Allwinkle, 2005). A recent example is the 2005–07 project funded by the EU through its Grundtvig program and Pascal, on Learning in Local and Regional Authorities, in which partners in six European countries identify the training that the local authority staffs need to implement the learning city concept and design a training program and workshops to meet this need.

Many countries have seen learning city and learning community initiatives over the past decade. In Australia, the State of Victoria has sponsored an initiative for several years in which eight nominated towns were enabled and supported to use this title; they undertook and publicized the initiatives mainly related to education, training, or community learning. Several of these featured in conference reports and specialized in mainly local literature. A little later the State of Victoria created the Department for Victorian Communities, renamed as the Planning and Community Development in 2007. The central purpose was to enable local communities to develop confidence and expertise as place-based learning communities and to play an active role in managing their affairs, with better integration of services between and across government portfolios.

In the United Kingdom, the Labour Administration elected in 1997 initiated devolution to Scotland and Wales and strengthened the English regions with nine regional development authorities championed by the Office of the Deputy Prime Minister (ODPM). As part of this initiative, several learning towns and learning community projects led to over 20 towns and cities defining themselves as learning cities and undertaking various initiatives; a loose community of interest and practice grew up around and anchored in these. The activity peaked early in the current decade; there is a sense that interest and energy many have slackened, perhaps from lack of clarity on what the concept and label really mean and how to go about the complex process of implementing it.

A similar sense of uncertainty affects some initiatives in Canada. In Victoria, for example, pronouncement of

being a learning city was followed by hesitancy as to how to go forward, and a learning fair which proved unsuccessful, bringing the initiative to a hiatus. In Vancouver, the city resolved and pronounced similarly, setting up a working group to find ways to give expression to the concept. A steering group was led by the Superintendent of Schools and the City Librarian, reflecting a common focus in or near to the education system.

The ease of communication enabled by new information technologies globally, a tendency to compare and compete internationally, and being watchful for new initiatives mean that many countries now have local-level authorities claiming to be learning cities and regions, albeit with different emphases and meanings in different countries depending on the local traditions, culture, and conditions.

Research and Implementation

There is a lack of academic research and little strictly academic literature in the social science fields directly related to learning cities. The subject is touched upon tangentially in a number of discipline areas especially related to government, urban and area studies, and special interests within education, innovation, and organizational behavior. The main scholarly interest comes from a policy development perspective. Much of the intellectual endeavor concerns trying to improve governance and enhance practice.

The Organisation for Economic Cooperation and Development (OECD), as an economically oriented intergovernmental organization, has been prominent in developing and disseminating an understanding of the issues and dynamics. A key monograph on learning regions in the new economy was derived from five studies of city regions in Europe. It was published in 2001 (OECD, 2001) and followed by another round of case studies and an international conference in Melbourne in 2002, from which grew an international observatory on learning regions, place management, and social capital called Pascal. Pascal mirrors the learning city-region approach in being a means of dialog and exchange between policymakers, practitioners, and scholars interested as a policy community and community of practice in putting knowledge into practice, making knowledge work in this applied arena (Duke *et al.*, 2005, 2006).

The Pascal virtual community reflects the character of the learning city as a construct, in that it includes practitioners as well as academic scholars. The latter come from many different disciplinary backgrounds. It is emphatically applied in its orientation, with the explicit purpose of fostering a dialog across professional and disciplinary boundaries. In its commitment to breaking down silos, it echoes a central issue for the learning city region about

specialization and compartmentalization as an often ineffectual way of managing complexity.

Another intergovernmental organization that also has a direct governmental remit, the EU, also promotes consideration of the learning region, as well as of lifelong learning, as a means to manage and succeed economically in a competitive global environment. There is an emphasis on capacity for innovation conceived in terms of broader social issues as well as immediate economic indicators of growth. The EU has supported various activities related to learning cities and regions, some mainly of a research nature with a dissemination component, others more concerned with the development and exchange of good practice by means of networks of places and people. In thus seeking to give effect to the concept in terms of governmental and related practices, the EU mirrors some of the ideas of the learning city, trying deliberately and systematically to collect, analyze, learn from, and put into practice what is being done.

One project supported by the EU illustrates the kind of research needed to understand how cities function, and what might be implied to enhance their capacity to operate more effectively in terms of an ideal type. A 3-year research project which studied four European cities and one Australian city took the title CRITICAL as an acronym for city regions as intelligent territories, innovation, competition, and learning. Research groups in each place examined the same eight domains in each city to see how these functioned and how they compared. Domains included, for example, small and medium enterprises, cultural activities, neighborhood renewal, and city administration.

Based on these studies of different areas of city life and cities' communities, the research suggested a number of core principles on the basis of which a city functions in terms of understanding, learning, and developing in the chosen ways and directions. Crucial to the success of such a learning process appeared to be the way that governance as expressed through the city or city region administration is connected, giving and taking knowledge, and sharing reflection and planning with the different communities and arenas that comprised its complexity. The concept of communities of practice (Wenger, 1998) proved useful, as did work on the creative classes and what made cities attractive and successful places to be, attracting lively and innovative people as well as capital (Florida, 2002). Sustainability, now a significant concept for governance and management as well as in terms of the environment, was seen as an important element of learning and its wider applications.

A chronic problem at the heart of the notion of the learning city region is related to geographical scope and jurisdiction in terms of what powers are exercised at local and regional level. Thus Vancouver, a leading city within a federal system, accounts to both the national federal and the provincial government for different purposes, and is in turn divided between several city administrations. The reach of

Vancouver as a learning city is restricted to just one of these administrations, although as a learning region it makes sense to think of two other levels, the greater Vancouver metropolitan area and a larger economic region, which takes in the catchment of the river valley system on which Vancouver sits. Dublin, the fast-growing capital city of the vigorous Irish Celtic Tiger economy, confronts transport and social challenges which call for a city region response for which there is no local–regional government authority. The same is true for Melbourne and other Australian cities, where the metropolitan area is divided between many small city authorities. Greater Melbourne planning falls by default to the state administration, which has competing responsibilities, and regional development bodies tend to be weak, short-lived, and are subject to the vagaries of distant federal politics.

Higher Education and the City

Given the emphasis on knowledge, including research, evaluation, and innovation that is central to the learning city, it is natural that universities in particular, and education systems more broadly, should be a subject of relevant interest. The OECD in particular has sponsored studies related to the role of higher education in regional development and therefore as a part of the learning city-region concept. (OECD, 1999, 2007). A project by the OECD launched in 2004 and completed in 2007 studied the Supporting the Contribution of Higher Education to Regional Development by means of 14 case studies in 12 countries across five continents, with a weighting in favor of Northern Europe. The project was managed jointly by two distinct parts of the OECD, the Territorial Government and the Education Divisions, thus attempting to bring together two different approaches and remits.

From the perspective of a region, higher education is important as a reservoir of expert advice and research capacity; eminent universities, like international airports, are recognized as a significant element among the characteristics of successful competitive world cities and regions. For a university, the resources, interest, and support of the city and region where it is located may be important to its prosperity as well as in terms of comfortable and productive local relations. The idea of university engagement carries with it the idea of co-production of knowledge and shared benefit. However, the idea of local–regional responsibility and accountability also discomfits some universities, which fear the loss of academic freedom and too local or parochial an identity, at the expense of their standing in the unbounded worlds of disciplinary scholarship. In this sense, the idea of the learning city region, and its active promotion by governments at different levels, has become significant to a dialog about the nature of the university in an era of mass higher education.

An Assessment of the Utility and the Likely Future of the Concept

The learning city is one of a number of terms and concepts that has been developed in an attempt to explain, understand, and in some cases to influence and shape the situation in which we live and work, in an era called global and characterized by rapid change and complexity. As a concept, it has acquired meaning, identity, and a life of its own, expressed through the pronouncements and practices of local, city, and region-level governments in many parts of the world. It exists by means of the exchange of words and ideas, and the development of language and practices, that attempt to improve the understanding and practice of administration at these different local levels more effectively and better fitted for what is seen as contemporary need and purpose.

At least in the reduced form of a place and administrative authority which provides opportunities for education, training, and different kinds of nonformal and contextualized learning, the learning city and region seems likely to retain its currency and to exercise significant influence for a considerable time. Its connection with lifelong learning, another widely used though also an ambiguous concept, now firmly established among intergovernmental and national authorities as essential at least for innovation and economic success in a competitive global economy, strengthens that probability. Even in this narrower sense, the relevance to policy and practice of the learning city may be considerable for tertiary and higher education and training, since lifelong learning is associated with the economic and social effects of changing demography, especially longevity, and the need for people to go on learning and changing.

The more fully the larger concept of learning city region is adopted, the greater may be its impact on higher education policy through the idea and practice of the engaged university, and more broadly on technical and other tertiary education. Institutions can be thought of as suppliers of highly qualified skills, and knowledge to the labor force of a regional innovation system. They may also be seen as developing and applying knowledge vital to the economic and social well-being of a region. This perception affects the content and processes of the regular curriculum; students' connection with working and community life; and the character and outcomes of the university's research activity. In these and other ways it may be seen as an integral part of a city or a region as well as of a regional innovation system. A further point along a spectrum of engagement takes the university close to the heart of planning for the region, so that its and the region's or city's long-term development are woven together in a web of interdependency.

As an idea and an agenda for regional management and development, the social construct of the learning city region connects with a complex policy agenda for mass higher education and for whole education systems. It is

central to questions about the best configuration of government, devolution to subnational levels, and relations between different levels of administration in the multi-level systems. These may be federal, with significant powers belonging to the state or province, as in Australia, Canada, Germany, and the United States, or centralized but with increasing devolution, such as Spain has seen and the United Kingdom is developing.

The learning city, in the more complete sense, encourages recognition of different communities and zones of activity that cohabit the city or region. It emphasizes a need to connect and involve these, terms such as participation and empowerment as well as consultation, in order to involve and draw on their knowledge and experience in nurturing a prosperous, optimistic, competitive, and perhaps also harmonious, attractive, and sustainable social and physical environment. The concept thus calls to attention the nature of government, and the costs of dissociation, apathy, and alienation. In this stronger sense, it is about abiding issues of governance that have acquired new significance in current times.

See also: Lifelong Learning; Overview of Lifelong Learning Policies and Systems.

Bibliography

- Duke, C., Osborne, M., and Wilson, B. (eds.) (2005). *Rebalancing the Social and Economic. Learning, Partnership and Place*. Leicester, UK: NIACE.
- Duke, C., Doyle, L., and Wilson, B. (eds.) (2006). *Making Knowledge Work. Sustaining Learning Communities and Regions*. Leicester, UK: NIACE.
- Florida, R. (2002). *The Rise of the Creative Class and How It's Transforming Work, Leisure, Community and Everyday Life*. New York: Basic Books.
- Hirsch, D. (1992). *City Strategies for Lifelong Learning*. Paris: OECD.
- Longworth, N. (2006). *Learning Cities, Learning Regions, Learning Communities: Lifelong Learning and Local Government*. Abingdon: Taylor and Francis.
- Longworth, N. and Allwinkle, S. (2005). *The PALLACE Project: Linking Learning Cities and Regions in Europe, North America, Australasia and China*. (final report to the European Commission), Edinburgh: Napier University on line. www.pallace.net.
- OECD (1999). *The Response of Higher Education Institutions to Regional Needs*. Paris: OECD.
- OECD (2000). *Learning Cities and Regions in the New Economy*. Paris: OECD.
- OECD (2007). *Globally Competitive, Locally Engaged – Higher Education and Regions*. Paris: OECD.
- Wenger, E. (1998). *Communities of Practice. Learning, Meaning and Identity*. Cambridge: Cambridge University Press.

Further Reading

- Best Value Victoria (2002). *Community Consultation Resource Guide*. Melbourne, VIC: Victorian Local Government Association (VLGA).
- Bjarnason, S. and Coldstream, P. (2003). *The Idea of Engagement. Universities in Society*. London: Association of Commonwealth Universities.

- Commission of the European Union (2001). *The Local and Regional Dimension of Lifelong Learning: Creating Learning Cities, Towns and Regions* (European policy paper). Brussels: EU DG Education and Culture.
- Commission of the European Union (2002). *Realising a European Area of Lifelong Learning*. Brussels: EU DG Education and Culture.
- Duke, C. (2004). *Learning Communities: Signposts from International Experience*. Leicester, UK: NIACE.
- Faris, R. (2002). *Learning Community by Community: Preparing for a Knowledge-Based Society*. Victoria, BC: Education Canada.
- Field, J. (2006). *Lifelong Learning and the New Educational Order*. Trentham Book.
- Keating, J., Badenhorst, A., and Szlachetko, T. (2002). *Victoria as a Learning Region*. Melbourne, VIC: RMIT.
- Longworth, N. (2003). *Lifelong Learning in Action: Transforming Education in the 21st Century*. London: Taylor and Francis.
- Nyhan, B., Attwell, G., and Deitmer, L. (eds.) (2002). *Towards the Learning Region: Education and Regional Innovation in the European Union and the United States*. Thessaloniki: CEDEFOP.
- Pope, J. (2006). *Indicators of Community Strength*. Melbourne, VIC: Department for Victorian Communities.
- Schuller, T. (1998). Three steps towards a learning society. *Studies in the Education of Adults* 30(1), 11–20.
- Unesco (1972). *Learning to Be* (The Faure Report). Paris: Unesco.
- Wain, K. (2000). The learning society: Post-modern politics. *International Journal of Lifelong Education* 19(1), 36–53.
- Yarnit, M. (2000). *Towns, Cities and Regions in the Learning Age – A Survey of Learning Communities*. London: UK Local Government Association.

Relevant Websites

info@creative.communities.org.uk – Centre for Creative Communities.
<http://www.lifelonglearning.co.uk> – Learning Towns and Cities UK.
<http://www.ncl.ac.uk/critical/knowledgeattracts> – Newcastle University; CRITICAL.
<http://www.obs-pascal.com> – Pascal International Observatory.

Museums as Sites of Adult Learning

R S Grenier, University of Connecticut, Storrs, CT, USA

© 2010 Elsevier Ltd. All rights reserved.

Hundreds of millions of people throughout the world visit museums, galleries, zoos, aquariums, nature and science centers, and historic sites. This article discusses museums as sites of adult learning. These institutions are increasingly defined by their ability to act as dynamic agents of cultural dissemination and have the capacity to expand the range of learning opportunities for adults. Adult learning theory and research in the area of museum education transform the visitor or user's experiences into learning opportunities that occur in relation to sociocultural surroundings and stimulate active and reflexive learning. Through both free-choice and organized nonformal learning visits, adults have the opportunity to experience the unknown, revisit the familiar, stimulate their curiosity, and challenge their existing beliefs.

The first section of the article explores the educational role of museums by reviewing the variety of cultural institutions defined as museums, as well as the educational missions and purposes of these organizations. The next section discusses the intersection of adult education and museum education theories and how these influence adult learning in museums. The final section reviews adult learning in museums, and attention is paid to the concepts of nonformal and free-choice learning.

Educational Role of Museums

As lifelong learning increasingly becomes more important in society, greater demands have been placed on museums to offer learning opportunities for visitors throughout their lives. The significance of museums as sites of adult learning is reflected in the definitions and missions of these institutions. Museums were once merely private collections held by aristocrats and universities; however, today, they constitute much more. The field of museum studies focuses on an array of institutions including, but not limited to, historic homes and sites, science and technology centers, aquariums, zoos, and botanical gardens, as well as the traditional art, history, and natural history museums. In general, the term museum includes both institutions with collections of exhibits, and locations without their own collections or permanent exhibitions, such as ancient remains or historic sites. These institutions function as a place of memory, where heritage is conserved, collections are maintained, and the best possible conditions are created for visitors to experience these collections. The International Council of Museums

(ICOM) defines a museum as a permanent, nonprofit institution in the service to society and the development of society. It is open to the public, and works to acquire, conserve, research, communicate, and exhibit, for purposes of study, education, and enjoyment, the material evidence of people and their environment. The American Association of Museums (AAM) offers another definition based on seven characteristics. Museums must:

1. be a legally organized not-for-profit institution or part of a not-for-profit institution or government entity;
2. be essentially educational in nature;
3. have a formally stated mission;
4. have one full-time paid professional staff member who has museum knowledge and experience and is delegated authority and allocated financial resources sufficient to operate the museum effectively;
5. present regularly scheduled programs and exhibits that use and interpret objects for the public according to accepted standards;
6. have a formal and appropriate program of documentation, care, and use of collections and/or tangible objects; and
7. have a formal and appropriate program of presentation and maintenance of exhibits.

Definitions also exist to clarify the role of education in museums. The AAM's Task Force on Museum Education, in their 1992 landmark report, offers a broad notion of museum education, which sees museums as fostering visitors' ability to be a productive member of a pluralistic society, and contributing to solutions to address the challenges of global citizenship. Hooper-Greenhill (1994) stresses the educational role of museums as central to a museum's mission and defines museum education as the creation of open relationships between museums visitors in order to increase enjoyment, motivation, and knowledge. Falk and Dierking (1995) offer further clarification by suggesting that museums provide visitors with accessible content and facilitate connections between contrasting facts and ideas. Moreover, they stress the museum's affect on visitor values and attitudes, while also promoting culture, community, and familial identity. Lastly, museum education is foundational to encouraging a visitor's confidence, interest, curiosity, and motivation to gain new knowledge, as well as to affect a visitor's thinking and worldview.

The depth and breadth of these definitions have been recognized by museum professionals and, in order to meet

these demands, many museums employ educational staff or curators whose role is to direct the education, learning, and outreach functions of the museum. Several times, this is achieved within a museum education department. A museum's education department frequently works in the development of signage, exhibits, texts, community outreach, visitor services, guided and self-guided tours, visitor workshops, lectures, seminars, and speakers' bureaus. Additionally, these departments may include volunteers, guides, interpreters, and docents who work with the public to devise and deliver purposeful programming and free-choice learning experiences.

Depending on the size, expertise, and mission of the institution, museum programming may be comprised of lectures, guided tours, field trips, gallery demonstrations, costumed interpretations, teacher and public workshops, seminars and symposia, film series, classes, theater, the loaning of objects, and the development and dissemination of teaching kits and packs related to exhibitions and collections. Moreover, innovative and creative programming has emerged, including learning opportunities geared to personal and cultural development, community outreach and activism, programs for special needs visitors, collaborative programs between museums and other organizations, such as universities, libraries, and other cultural institutions, as well as virtual and mobile museums. With the advent of new forms of technology and the Internet, this last type of programming is increasing in popularity since it provides anytime, anywhere learning for visitors.

Museums have moved beyond static homepages to the creation of extensive databases for research purposes and sites and portals that provide the public with detailed content, resources, and educational material. Examples of how museums are utilizing technology for adult learning include virtual exhibitions that provide information about a thematic area determined by the museum, interactive sites that offer educational activities to support learning, and on-demand exhibits that let the user control the subjects being viewed. These multimedia forms expand the learning experience and educational programming beyond the doors of the museum.

In addition to the growing role of technology in shaping adult learning in museums, institutions have also emerged to specifically draw attention to and address the social, political, and cultural conditions around the world. Such museums extend adult learning to include the promotion of humanitarian and democratic values. For example, the International Coalition of Historic Site Museums of Conscience is a network of historic site museums throughout the world presenting and interpreting a variety of historic issues, events, and people. Its mission is to assist the public in making links between the history of the site and its contemporary implications. For example, the network includes the District Six Museum in South Africa whose mission entails ensuring

the memory of forced removals in South Africa and brings visitors' attention to confronting all forms of social oppression. By focusing on social and cultural conditions, sites such as this strive to not only facilitate adult learning, but also empower visitors to challenge their understanding of history and the human condition.

Using research and scholarship to inform practice, museums are increasingly positioning themselves as places for rich learning experiences. With mission statements that highlight their key role in public learning and education, an array of physical and virtual learning opportunities, and the increase in addressing sociopolitical and cultural conditions throughout the world, the growth of adult learning in museums is evident. From highly structured programs that are identified, designed, and delivered by the institution to less formal opportunities, including incidental or free-choice learning, museums are taking a more sociocultural perspective that emphasizes visitor experiences in relation to the objects, museum context, and society.

Educational Theories

Museums have utilized research and theory in order to intentionally design and facilitate learning opportunities for adults. By introducing people to cultural, historical, and social artifacts, as well as nature and science, museums have the ability to support visitors as they become engaged with novel ideas. To accomplish this, museum educators draw from traditional educational theory, adult education principles, and museums studies research to enhance the learning experience for adult visitors. A variety of empirical and theoretical literature for examining the learning needs of adults in museums is available and the two bodies of research parallel one another, emphasizing the underlying need for a deeper understanding and commitment to adult museum visitors.

Theoretical approaches to museum education have a foundation in the broader field of education. For example, some museums support a constructivist approach to learning in their institution. Constructivism is a theory of learning focusing on the learner and the personal meanings they make based on their prior experience, knowledge, and interests. Museums utilizing a constructivist approach argue that visitors learn the most when knowledge is constructed mentally, in contexts that are physically, socially, and intellectually accessible. Thus, the needs and motivation of the visitor guide the structure of the exhibits and are essential when museum educators conceive of and facilitate learning opportunities. Other institutions utilize a sociocultural approach in museum learning research. Museums using this theoretical framework recognize that visitors' meanings are made within a social context, rather than from facts learned.

The sociocultural approach to learning in museums focuses on the interplay between individuals acting in social contexts and mediators such as tools, talk, activity structures, signs, and symbol systems found in cultural institutions.

These theoretical perspectives influence a number of museum education models. Falk and Dierking (2000) argue for an interactive experience model based on the concept of free-choice learning. Similar to situated learning, which supports encouraging the use of knowledge and strategies in a variety of settings, the Interactive Experience Model addresses the personal, social, and physical contexts of learning in museums. The personal context recognizes that: motivation and emotional cues initiate learning; personal interest facilitates learning; knowledge is constructed from prior knowledge and experience; and learning is expressed in an appropriate context. The second context, sociocultural, emphasizes learning as an individual as well as a group endeavor and the physical context addresses the notion that learning is dependent on a person's ability to place prior experiences within the context of their physical setting. These three contexts combine in an attempt to explain how, why, where, what, and with whom people learn in museums. This model aids in framing museum services, and highlights the role of educational theory in the process of exhibit development, interpretation, and programming.

Additionally, a focus on the role of prior experiences and learning is central to both the museum education and adult education fields. Experience influences adults' approaches to learning, their ability to integrate new information, and the ways in which they build concepts around new knowledge. Museums researchers are emphasizing the same ideas by addressing the prior experiences of visitors and how those experiences establish what visitors will do, talk about, and take away from their visits. Thus, museum educators are facilitating learning opportunities that recognize and underscore the role of emotions, memories, background, and personal understandings in the process of assigning reference and meaning to museum content.

The significance of understanding the adult visitor is emphasized by Sachatello-Sawyer *et al.* (2002), who categorized adult learning in museums to include six dimensions: life-changing experiences, transformed perspectives, changes in attitude, increased appreciation, exploration of relationships, and acquisition of knowledge. These levels serve to explain the forms of personal change, long-term change resulting from the learning, and the integration of new learning. They also found that adult learners in museums fell into four distinct categories: knowledge seekers, socializers, skill builders, and museum lovers. Both the forms of adult learning and the categorization of adult learners in museums mirror the work of Mezirow (1991) and Houle (1961) in the adult education literature. Mezirow describes four processes of learning: elaborating an existing point of view, establishing a new point of view, transforming

one's point of view, and, lastly, becoming aware and critically reflecting on our generalized bias of how we view groups other than our own; and Houle's typology of adult learners includes: (1) the goal oriented, (2) the activity oriented, and (3) the learning oriented.

While terminology used in the two fields may differ, this section demonstrates that the underlying theoretical frames and the parallels in how both fields characterize learning and adult learners are similar. Regardless of the scholarship, the significance of lifelong learning in museums is receiving considerable attention from researchers and is providing opportunities for furthering our understanding of adult learning theory. Moreover, the research in the two fields, as well as from visitor studies, interpretation research, and leisure and recreation studies is informing museum practice, thus increasing the likelihood of presenting museums as sites of adult learning experiences.

Adult Learning in Museums

In general, adult learners frequent museums as visitors or users of museum services that include the availability of reference and resources, cultural and community programming, and virtual offerings. These learners are diverse and create a challenge for institutions seeking ways to provide adults with education and learning opportunities. Due to their broader life experiences, established identity, abstract thinking ability, understanding of the world they live in, and unique learning expectations, it is necessary to specifically address adult learning in museums. First, the motivations and agendas that adults take to museums are important in influencing the way they experience a visit, as well as their behavior and learning. Those facilitating educational opportunities in museums take into account such motivations in order to create engaging visits. Several factors influence the decision of whether or not to visit a museum, including an opportunity to socially interact with others and with family, the sense that they are doing something worthwhile, the challenge of new experiences, feeling comfortable with their surroundings, having the chance to actively participate, and having an opportunity to learn. The combination of these factors requires museums to develop programming that incorporates the breadth and depth of visitor characteristics ranging from purposeful learning to incidental learning resulting from a leisure visit.

Although learning is not a deliberate intention of many museum visitors, they often seek out or are unconsciously drawn into an experience that encompasses learning. When visiting museums, adults tend to look for opportunities to learn more about themselves, their culture, and their heritage, and gravitate to those places where they feel most comfortable. Overall, research in the area of visitor motivation suggests that adult visitors are not only motivated to

learn, but also perceive museums as sources of important information, are willing to commit to learning activities, and find such endeavors satisfying, all of which create both opportunities and challenges for the institutions.

Regardless of the motivation or intention, visits by adults to museums present both social and educational opportunities and create the possibility of purposeful and incidental learning that is the focus of much study in the fields of museum and adult education. Museums are distinct learning environments, and as such have received considerable attention by scholars. Within adult education, museums are categorized as nonformal education, while in museum studies, these learning environments are often referred to in relation to free-choice learning. Both terms provide a means for framing adult learning experiences in museums and offer similar perspectives on the role of the learner and the resources necessary to support the learning process.

Nonformal Education

Nonformal education is characterized as intentional and organized with the purpose of promoting learning to enhance an adult's quality of life (Heimlich, 1993). Ideally, it is learner centered, maintains a balance of power between the learner and the facilitator, is present-time focused, and is geared to meeting localized needs. In this same vein, learning in a nonformal context is often distinguished by activities outside the formal learning setting, with voluntary participation as opposed to mandatory participation. Within a taxonomy of adult learning, nonformal learning is identified by learners holding the objectives for learning with the means controlled by the educator or organization. In this way, nonformal education creates learning events in museums that can expand the range of opportunities for adults with practical applications to an individual's profession, personal interests, and community. Owing to the unique learning environment and resources museums offer, these institutions are constructing nonformal learning offerings ranging from senior citizen programs, programs for law enforcement officers, and restoration and preservation opportunities, to training for volunteers interested in working as docents, interpreters, and oral historians. One group of learners garnering significant attention is classroom teachers. Museums create learning opportunities for classroom teachers through in-service, continuing education and intensive, residential summer institutes. These experiences are designed and initiated by museums to provide educators with opportunities to explore museum resources, co-create curriculum with peers, and experience museum exhibits under the direction of museum staff and museum consultants specializing in k-12 curriculum and instruction.

Nonformal programming serves not only organized groups, but also those seeking personal development and

educational opportunities, and individuals serving as volunteers for museums. Adults involved in museum programs are often seeking out lifelong learning opportunities and look for ways to bring together their personal interests, professional expertise, and social consciousness. Programs like the Smithsonian's Resident Associate Program are one example of museums providing a range of experiences to adult learners including structured lectures, tours, and performances that offer cultural and educational opportunities for lifelong learners.

Despite the fact that nonformal learning includes museum-derived programming, nonformal education also provides for more self-directed and informal learning opportunities. For instance, many museums are emphasizing active involvement of visitors through the use of engaging questions, involved discussion, and co-constructed workshop methods. Such experiences encourage adults to handle objects, investigate the meanings and relationships between objects and exhibits, and address their own reactions to the museum content. Although museums serve the role of nonformal education for adult learners, a good deal of visitors are informal users that include individuals, friends, and families visiting museums casually as tourists or for entertainment and social interaction. These informal experiences are at the root of another form of learning in museums, free-choice.

Free-Choice Learning

Free-choice learning (Falk and Dierking, 2000) is a non-linear process that looks to visitors to bring their own awareness and interests to the museum experience in order to create a variety of learning outcomes. That being said, such learning opportunities are not completely unstructured since the nature and design of the exhibits form some structure to the learning experience. It is this choice and control over the learning that is central. In free-choice learning, the learner is intrinsically motivated by their desire to discover more about the world, gain information, and enhance their current understanding. More specifically, adults tend to engage in free-choice learning in museums because of one or more factors. The first factor is friendships and organizational relationships including the social connections established in civic associations, schools, and community and religious groups. These interactions strengthen a commitment to developing social networks through invitations to take part in the museum experience. The second is the visitor's family who communicates information about learning opportunities or emphasizes an interest in museums. The last factor is an adult's business and professional connections. These relationships can create the expectation of visiting a museum for professional networking, or to gain new knowledge to transfer to the work environment.

Museums are responding to free-choice learning by layering the experiences. By doing so, museums are able to present visitors with smaller segments that are more easily processed and integrated into prior experiences and learning. This limits visitors' sense of being overwhelmed and provides a chance for visitors to linger on specifics that are of most interest. Technology is also aiding museums in their facilitation of free-choice learning. Radio-frequency identification (RFID) is a method of automatic identification that relies on storing and remotely retrieving data using RFID tags. Using RFID, the Tech Museum of Innovation in San Jose, CA, began experimenting with Tech-Tags that work to link one exhibit to another within the museum, to personalize and customize the experience, and extend the museum experience beyond the visit. An RFID chip is embedded in a visitor bracelet enabling visitors to use their bracelet to activate exhibits as well as create a customized Internet record of the visit that can be explored after the leaving the museum.

In general, visitors in a free-choice environment decide whether to visit a museum, what will be viewed or done at the museum, and for how long. Essentially, free-choice is closely linked with educational leisure in that the experience is learner centered through exploration, connections to prior experiences, and control of the learning environment. Although museums are a significant source of free-choice learning for adults, it is important to note that people engage in free-choice learning through other venues including libraries, parks, radio, television and film, print media, and the Internet. When these sources are partnered with the learning opportunities of museums, what is produced is the opportunity for a robust and multidimensional form of understanding. Entities such as the Public Broadcasting Service (PBS) have used such partnerships to develop free-choice learning opportunities that bring together print media, technology, and museums to feature content and exhibits in relation to a PBS series.

Whether it is nonformal education or free-choice learning, the outcomes for adults' experiences in museums are diverse. Participation in museum-related activities contributes to the shaping of families, and other social groups, as well as expanding the personal and professional perspectives of individual adult visitors. By increasing knowledge and understanding, developing new skills and abilities, inspiring new learning and change, and stimulating lifelong learning, museums not only impact an adult's knowledge, but, often, also their attitudes, values, and beliefs.

Conclusion

Today, adult education stresses self-direction, critical reflection, experiential learning, learning to learn, distance learning, and collaborative learning. The same elements can also be associated with adult learning in museums, as is

evident in this article. Museums as sites of adult learning are stimulating and offer a place where ideas originating from the media or peers can be tested, confirmed, or modified. Museums also help visitors reformulate old pieces of understanding that have lost relevance or meaning. By purposefully taking part in educational activities in museums and engaging in free-choice learning, adults have the opportunity to share in conversations, discussions, debates, and social interaction, all of which are foundational to the work of museums and the expanding role of lifelong learning in society.

See also: Informal Learning: A Contested Concept.

Bibliography

- Falk, J. H. and Dierking, L. D. (eds.) (1995). *Public Institutions for Personal Learning: Establishing a Research Agenda*. Washington, DC: American Association of Museums.
- Falk, J. H. and Dierking, L. D. (2000). *Learning from Museums: Visitor Experiences and the Making of Meaning*. Walnut Creek, CA: Altamira.
- Hooper-Greenhill, E. (1994). *The Educational Role of Museums*. London: Routledge.
- Houle, C. O. (1961). *The Inquiring Mind*. Madison, WI: University of Wisconsin Press.
- Mezirow, J. (1991). *Transformative Dimensions of Adult Learning*. San Francisco, CA: Jossey-Bass.
- Sachatello-Sawyer, B., Fellenz, R. A., Burton, H., et al. (2002). *Adult Museum Programs*. Walnut Creek, CA: Altamira.
- Carr, D. (2003). *The Promise of Cultural Institutions*. Walnut Creek, CA: AltaMira.
- Dufresne-Tassé, C., Lapointe, T., Morelli, C., and Chamberland, E. (1991). L'apprentissage de l'adulte au musée et l'instrument pour l'étudier. *Canadian Journal of Education* **16**, 280–292.
- Falk, J. H. and Dierking, L. D. (eds.) (1995). *Public Institutions for Personal Learning: Establishing a Research Agenda*. Washington, DC: American Association of Museums.
- Falk, J. H. and Dierking, L. D. (1997). *The Museum Experience*. Washington, DC: Whalesback Books.
- Falk, J. H., Moussouri, T., and Coulson, D. (1998). The effects of visitors' agendas on museum learning. *Curator* **41**, 107–120.
- Heimlich, J. E. (1993). *Nonformal Environmental Education: Toward a Working Definition* (ERIC Document Reproduction Service No. ED 360 154). Columbus, OH: Educational Resources Information Center.
- Hein, G. E. (1998). *Learning in the Museum*. London: Routledge.
- Hein, G. E. and Alexander, M. (1998). *Museums Places of Learning*. Washington, DC: American Association of Museums.
- Hooper-Greenhill, E. (1994). *The Educational Role of Museums*. London: Routledge.
- Leinhardt, G., Crowley, K., and Knutson, K. (eds.) (2002). *Learning Conversations in Museums*. Mahwah, NJ: Erlbaum.
- Prentice, R., Davies, A., and Beeho, A. (1997). Seeking generic motivations for visiting and not visiting museums and like cultural attractions. *Museum Management and Curatorship* **16**, 45–70.
- Roberts, L. C. (1997). *From Knowledge to Narrative*. Washington, DC: Smithsonian Institution Press.
- Storr, A. V. F. (1995). *Current Practice and Potential: Research and Adult Education in Museums*. Report No. IR 017 333. (ERIC Document Reproduction Service No. ED385255) Washington, DC:

- National Institute on Postsecondary Education, Libraries, and Lifelong Learning (ED/OERI).
- Svedlow, A. (1997). Lifelong learning in museums: In pursuit of andragogy. *PAACE Journal of Lifelong Learning* **6**, 29–39.
- Taylor, E. W. (2006). Making meaning of local nonformal education: Practitioner's perspective. *Adult Education Quarterly* **56**(4), 291–307.

Relevant Websites

- <http://www.aam-us.org> – American Association of Museums.
- <http://www.gem.org.uk> – Group for Education in Museums.
- <http://www.sitesofconscience.org> – International Coalition of Sites of Conscience.
- <http://icom.museum> – International Council of Museums.
- <http://www.definitionsproject.com> – National Association for Interpretation.
- <http://www.sil.si.edu> – Smithsonian Institution Libraries, Museum Studies and Reference Library.
- <http://www.clmg.org.uk> – The Campaign for Learning through Museums and Galleries.
- <http://www.thetech.org> – The Tech Museum of Innovation.

Overview of Lifelong Learning Policies and Systems

D-B Kwon and C H O Daeyeon, Sookmyung Women's University, Seoul, Republic of Korea

© 2010 Elsevier Ltd. All rights reserved.

Since lifelong learning systems embodying various policies are emerging in a variety of unique contexts, they represent very different perspectives from country to country. This article reviews lifelong learning policy models proposed by Green (2000), including market-led, state-led, and social-partnership-led approaches. Different policy initiatives and/or systems can be deemed to belong to one of three different models, although they may not correspond precisely. This article depicts seven well-known policy examples according to the three models.

Three Models of Lifelong Learning Policy

During the past two decades, the significance of lifelong learning has emerged and many influential policy documents have been released to promote it throughout the world. However, policies implemented in particular contexts according to different rationales yield substantially different results. Consequently, lifelong learning policies initiated by various governmental bodies can be deemed to belong to different types; have different characteristics, emphases, and objectives; and require different kinds of interventions. To sketch out a number of different lifelong learning policies, analytical models need to be selected carefully.

Green (2000) proposes three hypothetical models of lifelong learning policy including market-led, state-led, and social-partnership-led approaches: "some policies stress the role of the market and the responsibility of the individual; some advocate the central role of the state in orchestrating and managing the learning society; others emphasize social partnership among multiple stakeholding" (p. 35). It is acknowledged that the models can be regarded as positions along a continuum.

In the market-led model, there is a belief that lifelong learning is an individual project, and, therefore, the burden of lifelong learning tends to fall on the shoulders of the individual learner. Today's global market forces individuals to take responsibility for their own learning for personal growth and development. Both employers and individuals have to make decisions about what kinds of skills and knowledge should be acquired. In this sense, learning supply through education and training can be decided based on the needs of the market, represented by employers and individuals. In the labor market, employers play a central role in providing lifelong learning for individuals at work (PARN, 2002). As the market has a

more privileged position in this model, the role of the state and civil society is relatively neglected (Rubenson, 2006). On the other hand, there are some weaknesses in this approach: market-led lifelong learning policies can lead to underinvestment, inequality, and low quality (Green, 2000).

The state-led model allows governments to invent lifelong learning systems through legislation, controlling bodies, and related policy activities. The state plays a key role as a planner; at the same time, it can be the primary source of funds and quality control for lifelong learning practice. To accomplish these, the state creates a blueprint that must meet the various long-term needs of individuals, in addition to reconciling the demands of a variety of interest groups. The state regards lifelong learning as a matter of public responsibility. Thus, it tries to create the required structural preconditions, promote well-designed arrangements between organizations and entities, and ensure policy coordination and coherence in lifelong learning (Rubenson, 2006). The state-led approach addresses equity issues for young, old, and low-skilled adults who are at risk of being routinely excluded from society. Such policies reduce social inequality and promote personal development for all. Additionally, a strong link between lifelong learning and labor market policies can be built on this model. The state has strong power to use labor markets and policies to promote human capital through lifelong learning. The state-led policies also have some disadvantages, including "a slow pace of change, less diversity and responsiveness to particular needs, misjudged plans, and bureaucratic inefficiency" (OECD, 1996, recited from Green, 2000: 39).

Another possible model is based on social partnership. This model emphasizes not only individual responsibility, but also the building of partnerships with multiple agencies involving diverse stakeholders. This implies a strengthened public-private cooperation. However, state agencies' participation and coordination are not always required in this model (Schuetze and Casey, 2006). Regardless of the state's participation, agreements or cooperation among social partners can have an impact on lifelong learning and training policies (OECD, 2003). In the 1970s, civil society developed lifelong learning systems for the purpose of reducing educational gaps in society. At that time, the volunteer sector, including nongovernmental organizations (NGOs), played an important role in promoting lifelong learning (Rubenson, 2006). As there is no doubt that civil society cannot exist alone, social partnerships with other institutions are emphasized in policy development (OECD, 2003).

The three models can illustrate the different lifelong learning policies with distinctive features. Nevertheless, it must be taken into consideration that the three models have common characteristics, which can be seen in the state-led model to some degree. The distinguishing characteristics and some policy examples of each model are summarized in **Table 1**.

More recently, Schuetze and Casey (2006: 282–283) introduced four different lifelong learning policy models: (1) an emancipatory or social justice model, which put an emphasis on equality of learning opportunity (lifelong learning for all); (2) a cultural model focusing on each individual's life; (3) an open society model in which lifelong learning is seen as an appropriate learning system (lifelong learning for all who want, and are able, to participate); and (4) a human capital model, which implies continuous work-related training and skills development for a qualified workforce (lifelong learning for employment). There exist many commonalities between Schuetze and Casey's models and those of Green. However, Schuetze and Casey classify policies according to their aims for what the policy is, while Green distinguishes the approaches according to who has the power and who is the main actor with a more holistic viewpoint toward the policy.

Policy Applications of the Three Lifelong Learning Models

This section provides some policy examples with reference to Green's three models from developed countries well known for their lifelong learning strategies.

The State-Led Model

The state-led model has many advantages insofar as it provides a certain degree of consistency and coherence in conducting policy, securing qualifications, and ensuring more equitable opportunities for individuals. Those advantages are found in the countries' initiatives that are introduced below. They vary in the details, but each shares the goal of lifelong learning for all.

Sharing financial responsibility in Japan: Kameoka Lifelong Learning Foundation

In Japan, the central and local governments have legislated strong initiatives to support lifelong learning. For a long time, governmental agencies have promoted lifelong learning by installing facilities and equipment, and by subsidizing learning events. The main responsibility for administrative support of lifelong learning tends to fall to the municipalities. Since the financial base of the municipalities is not sufficient, the National Council on Lifelong Learning proposed establishing foundations to obtain private funds at the level of the local government (Shiraishi, 1998). With this purpose, there are some foundations such as the Kameoka Lifelong Learning Foundation (KLLF).

In 1990, Kameoka City established the KLLF, giving it initial funding. This foundation operates with financial support from the municipality and the management of its initial funds. Although Kameoka City provides financial support, the KLLF is an independent institution beyond the purview of the municipal and other governmental bodies (Shiraishi, 1998). The organization is autonomous in how it conducts the bulk of events related

Table 1 Distinguishing characteristics of respective policy model

<i>Model</i>	<i>Characteristics</i>	<i>Policy examples</i>
State led	Law and regulation Long-term / slow change Equality / accessibility Coherence / consistency Transparency / bureaucracy Equity / control high quality Inefficiency	<ul style="list-style-type: none"> • Kameoka Lifelong Learning Foundation (Japan) • Ensuring equity for immigrants (Finland) • Continuing vocational education and training (France)
Market led	Market driven / demand led Rapid / short term Individual needs Individual responsibility Employability Flexibility / diversity Inequality / underinvest	<ul style="list-style-type: none"> • Q-Card at Fraport AG (Germany) • The Credit Bank System (South Korea)
Social partner led	Balance / partnership Stakeholders in different levels Soft regulation / mixture Negotiation Inefficiency Difficulty of cohesion	<ul style="list-style-type: none"> • Dual system (Germany) • The new apprenticeship program (Australia)

to lifelong learning for citizens, thus enabling it to more easily obtain financial support from other sectors.

In the case of Japan, the government played a key role in developing and funding the foundation at the initial stage, preparing the foundation to develop as an independent organization later by widening its base of financial support from private donors. This project was a part of the government's long-term plan to establish private funding for lifelong learning enterprises at the local level.

Ensuring equity for immigrants in Finland

To ensure equity for immigrants, the Finnish government has implemented many policy initiatives over the past years. As more countries join the European Union, Finland is expected to have more immigrants, and thus shares the necessity of having sound policies that seek to integrate its newcomers. In this context, the Finnish government has launched a variety of programs available to immigrants. For adult immigrants who seek employment, special programs such as language and vocational ones are provided.

Since May 1999, many special programs, for example a 6-month preparatory vocational training program, have been offered for low-skilled immigrants. The program incorporates Finnish language and customs; guidance and some courses in a student's native language; remedial instruction if necessary; support groups of students; tutors; and a personal study plan (OECD, 2005a). This effort aims to reduce unemployment among immigrants by promoting their equality within Finnish society.

The Finnish immigrant policy does have some challenges to overcome, such as inconsistencies in policy implementation, and a lack of programs specifically for low-skilled immigrants. To develop more effective policies for immigrants, the Finnish government has set out a 5-year plan for 2003–08 (Ministry of Education, 2004). The Finnish government puts much effort into pursuing equity and consistency in lifelong learning policy for immigrants. Consequently, the case of Finland focused more on immigrants as an underprivileged group. The Finnish government program examined tries to provide equity and equitable opportunities for individuals, especially for those who could be otherwise excluded from the benefits of lifelong learning.

Continuing vocational education and training in France: Social partnership regulation

France's lifelong education on continuing vocational education and training (CVT) has been in constant change and expansion for almost 40 years, thanks to the 1971 law and subsequent reforms characterized by a state-led system of social partnership regulation. This law insisted on the roles and responsibilities of the various partners in training by instituting a collective financing obligation

between social partners (Green, 2000). This regulation took into consideration the promotion of equitable distribution of training costs and achieving the skills needed between social partners. In short, the creation of a huge training market could be achieved under the 1971 law in France (Colardyn, 2004).

For enterprises, the financial contribution was differed from the amount of payroll based on the number of employees. It was a very useful way to obtain equitable sharing of training costs from different enterprises. Such an equitable cost distribution encouraged greater investment in training by employers. At the same time, the state invested much effort on deciding training priorities by forming clubs of providers as well (Green, 2000). Government subsidy for training played a crucial role to overcome market failure in training. As a result, strong links were established between private and public sectors led by the regulation and public authorities.

Under the 1971 law, training became a legal right for all employees, and the responsibility for training was shared by enterprises beyond the government. In so doing, the government had better chance to look at individuals such as the young, women, and older workers. Clearly, many individuals have accessed training; a large training market has been created; and the training supply has improved (Colardyn, 2004). For instance, in the year 2000, there were nearly 42 000 training bodies, as compared to only 25 000 in 1990. The number of trainees has practically doubled in these 10 years, surpassing 12 million in 2000. Employer investment in training continues to increase (Green, 2000) and enterprises have been the major contributors since 1999. In spite of many positive results, several major critical issues can be highlighted (Colardyn, 2004): people who have had higher levels of initial education are more likely to participate in CVT; CVT chances have hardly translated into a higher level of educational attainment such as a diploma.

The Market-Led Model

There is no doubt that lifelong learning is a huge enterprise beyond a government's capacities. In addition, lifelong learning in most countries is more likely to be decentralized. Furthermore, the change of perspective from lifelong education to lifelong learning entails shifting the responsibilities for learning from the providers to the learners. This shift emphasizes the individual's role in the process of learning and deemphasizes the governmental dimension (Schuetze and Casey, 2006). Consequently, the role of employers and individuals in decision making and financing is strongly addressed. The responsibility of individuals for their own learning is emphasized to a greater degree as well. In particular, companies play a more direct and active role in promoting employees' learning in order to obtain greater global

competitiveness. Workplaces that support and promote lifelong learning for employees function as good examples of the market-led approach.

Investment on employee training by Fraport AG in Germany: Qualification card

Fraport AG operates Germany's largest commercial airport with 13 000 employees. It regards personnel development as an important factor in corporate success, and has adopted a variety of strategies relating to education and training. Fraport AG introduced a new educational approach in 2000. It developed the Fraport Qualification Card (Q-Card). The Q-Card is a bonus card that the corporation loads with a virtual credit of 600 euros each year. With this card, employees can take courses provided by Fraport College and Fraport Academy. The courses at Fraport AG are not directly required for the performance of their current jobs.

Employees in Fraport AG must invest their flexitime credits from their working time accounts. The program encompasses information technology, media skills, work techniques, and business administration. Approximately 85 different training courses were offered in the first half of 2002 (Wilfried, 2003). As a cost- and time-sharing instrument for lifelong learning, the Q-Card program earned Fraport one of the initiative awards for training and further education given by the Otto Wolff Foundation (Wilfried, 2003). Responding to the demands of employees and the employer, Fraport AG created a motivational system to promote employee learning and provided a wide range of programs to meet the learning needs of employees.

The Credit Bank System in South Korea: Nonformal educational institutions

The South Korean government initiated the Credit Bank System (CBS) in 1998. This essentially allows individuals to accumulate credits from diverse institutions including colleges, universities, nonformal educational institutions, and cyber open universities. An adult student can obtain an associate or bachelor degree depending on the amount of the necessary CBS-approved credits (OECD, 2005b).

The CBS can be regarded as a precondition to realize a society of lifelong learning and open education. It has some unique features that can be discussed in the state-led model. For instance, this system has been run on a strong statutory foundation. The CBS contributes to guarantee educational equity, particularly for the under-educated. The Korean government, together with the National Center for Lifelong Education, screens the curriculum provided by all nonformal educational institutions twice a year and earned credits and/or learning experiences and activities of individuals. These efforts are to maintain and control the quality of the CBS.

On the other hand, the CBS has qualities of the market-led model as well. The CBS policies are more likely to

focus on giving individuals more choices from different institutions and diverse learning programs. All types of educational institutions under the CBS need to study whether they should supply market-driven educational programs in order to meet adult learners' demands. Educational institutions that have marketing capability need to promote their credit courses by creating a variety of programs that are required to be endorsed by the government. The CBS should be utilized to sharpen adult learners' competency and employability regardless of their employment status as a way of enhancing the quality of their lives. Consequently, some features, including regulation, equity, and accessibility, which are addressed in the state-led model, are shown in CBS. However, this system still runs on the strong foundation of the market-led model. Compared with the state-led model, this case is more responsive to individual needs and tends to have more flexibility in the modes of provision.

The Social-Partnership-Led model

As most stakeholders in lifelong learning exist outside of governmental agencies, the coordination between various entities involved in policy development and implementation is crucial for overall success (OECD, 2004). However, such coordination does not fall exclusively into the domain of governmental agencies. Nevertheless, there is general agreement that using the potential of existing cooperative mechanisms is a way to build partnerships (Dace, 2003). In current lifelong learning policy and practice, cooperation among various partners has become a prominent international feature. Both cases described below created frameworks for lifelong learning in terms of encouraging cooperation among stakeholders and emphasizing the responsibilities of different partners. In this framework, various stakeholders participated in the whole process of policy development, implementation, and assessment.

Vocational education and training in Germany: The dual system

With the advent of a knowledge-based economy, Germany has recognized the importance of lifelong learning in order to pursue social cohesiveness. In particular, different types of social partnerships have been utilized to share the responsibility for lifelong learning. Among them, the dual system has been recognized as a representative example of vocational training on the social-partnership-led model. The system is called dual because vocational training takes place both in the company and in the vocational school.

Under the Vocational Training Act of 1969, the federal state sets the training regulations by which the roles, rights, and responsibilities of the different partners are determined. Representatives of employers, employees,

and educators together with federal and regional state officials are positively involved in an elaborate system of social partnerships (Green, 2000). The system is financed principally by employers, a feature that distinguishes the German system from the Australian and other European models that rely heavily on government funding as well as on wage subsidies. Although the operation of vocational training schools is regulated by the state, a substantial involvement on the part of employers through very active chambers of commerce is taken into consideration as the essential feature of the social-partnership-led model (NCVER, 2001).

The German dual system of vocational training is renowned as an example of social partnership. Within such a system, companies and their social partners play a prominent role as innovators. This system emphasizes collaboration among various partners and suggests co-financing between them. As the state's full financial responsibility in lifelong learning is perceived to be limited, co-financing among stakeholders should be strongly considered. Meanwhile, the dual system itself does exhibit weaknesses, such as inefficiency, because it may take much time to negotiate among the key social partners through a very complex process (NCVER, 2001). To keep the dual system effective and efficient, the government must continue to become more flexible in how it deals with its partners.

Embracing older workers in Australia: The new apprenticeship program

The Vocational Education and Training (VET) system in Australia was created to boost the country's economy, increase employability, and develop a more skilled workforce. Under the direction of the VET, the apprenticeship program has been newly updated to combine training and employment so that people entering an occupation can receive appropriate instruction in the specific skills needed on the job (OECD, 2002). Compared to the previous apprenticeship program, the new apprenticeship program focuses more on supporting and recruiting older workers, recognizing that it is crucial to have skilled workers in the labor market and that older people still remain as key players.

The new apprenticeship program was introduced in 1998, following many reforms of the traditional apprenticeship system. The Australian government created the new apprenticeship center to promote the new program and to provide support services with employers. The government developed specific programs to ensure a more effective implementation of the new apprenticeship. These include the new apprenticeships incentives program to develop a more skilled Australian workforce and the new apprentice support services, a national network that provides support for employers and individuals across the nation. There has been a huge increase in the number of workers in the new apprenticeship program

since it was introduced, from 135 000 in 1995 (OECD, 2002) to 367 100 at the end of 2003 (NCVER, 2003).

Through the new apprenticeship program, the Australian government tries to embrace not only the need of young people to obtain skills, but also the need of older workers to be given opportunities to upgrade their skills. Moreover, the new apprenticeship program makes cooperation among various partners possible so that these partners can meet the demands of individuals and employers. On the other hand, some limitations exist, such as the lack of sufficient training plans and inadequate arrangements for monitoring the quality of training in the new apprenticeship program (NCVER, 2001).

Conclusions

In this article, several examples categorized according to a threefold system of lifelong learning policy models were examined. The KLLF in Japan, immigrant policy in Finland, and CVE in France were introduced and explored as examples of the state-led approach. The examples of the market-led approach included Fraport AG in Germany and the CBS in Korea. Germany's dual system and Australia's new apprenticeship program served as examples of the social-partner-led model. Each country examined has very different lifelong learning policies in its own unique context. In addition, every country was found to have taken various initiatives at different levels including individual, enterprise, and local and central government. Despite the diversity evident in the above examples, it is interesting to note that some clear messages regarding current and future lifelong learning policies can be discussed.

First, an economically oriented mindset has penetrated the policies of almost every country. With the advent of the knowledge-based economy, all countries seem to agree that a shortage of workers with high-level skills or a lack of adequate learning systems puts national economic competitiveness at risk. Lifelong learning systems tend to be a credible way to achieve national prosperity. In looking at the examples illustrating each model, it is found they are strongly associated with basic and upgraded skill-development strategies for individuals. At the same time, policy reports related to lifelong learning are also more likely to express concern for social inclusion, democracy, social equity, quality of life, and personal well-being (Martin, 2000).

Second, cooperation among many stakeholders seems to be an inevitable feature in terms of developing and implementing lifelong learning policies. State-led initiatives are basically the strongest and most effective way toward pursuing lifelong learning opportunities for all. While it provides relatively less accessibility of learning opportunities to individuals than the state-led approach, the market-led approach is also necessary. However, neither state-led nor market-led approaches can fully

succeed without collaboration with other partners. In this sense, it is safe to say that countries try to involve relevant stakeholders and boost social partnerships between them. All the exemplary initiatives introduced show that many different parties are more likely to be involved and cooperate for the purpose of building better lifelong learning systems, regardless of the models they belong to.

The third point is related to government intervention. Today's governments face major challenges in extending lifelong learning. The government, even in the social-partner-led model, is found to play a vital role, and its intervention is unavoidable in the field of lifelong learning. Government intervention in most cases involves providing financial support, operating a qualifications system, or organizing networks for the various partners. As seen in the KLLF of Japan, municipal funding was an absolute necessity at the initial stage, allowing it to become an independent organization. The dual system in Germany and the new apprenticeship program in Australia are also evidence of the necessity of government involvement in funding and assessing evaluation. In the case of CBS, the Korean government plays a greater role in monitoring curriculum, earned credits, and learning experience of individuals to control the quality of the system.

Bibliography

- Colardyn, D. (2004). Lifelong learning policies in France. *International Journal of Lifelong Education* 23(6), 545–558.
- Dace, N. (2003). Towards EFA goals: Situation in Latvia and cooperation in the Baltic Sea Region. *A Paper Presented at Lifelong Learning Discourses in Europe*. Hamburg: UNESCO Institute for Education.
- Green, A. (2000). Lifelong learning and the learning society: Different European models of organization. In Hodgson, A. (ed.) *Policies, Politics and the Future of Lifelong Learning*, pp 35–50. London: Kogan Page.
- Martin, I. (2000). *Reconstituting the Agora: Towards an Alternative Politics of Lifelong Learning*. In Sork, T. J., Chapman, V. L., and St. Clair, R. (eds.) *Proceedings of the 41st Annual Adult Education Research Conference*, pp 255–260. Vancouver, BC: University of British Columbia.
- NCVER (2001). Australian apprenticeship: Facts, fiction and future. <http://www.ncver.edu.au/research/prcj2/mk0006.pdf> (accessed May 2009).
- NCVER (2003). Australian vocational education and training statistics: Apprentices and trainees 2003. <http://www.ncver.edu.au/statistics/aats/ann03/03annsum.pdf> (accessed May 2009).
- OECD (2002). *Mechanisms for the Co-Finance of Lifelong Learning, Presented at the Second International Seminar: Taking Stock of Experience with Co-finance Mechanisms*. London: OECD. <http://www.oecd.org/dataoecd/51/27/2501342.pdf> (accessed May 2009).
- OECD (2003). *Beyond Rhetoric: Adult Learning Policies and Practices*. Paris: OECD.
- OECD (2004). The thematic review on adult learning: United Kingdom (England). *Background Report*. Paris: OECD.
- OECD (2005a). *Equity in Education, Thematic Review: Finland Country Note*. Paris: OECD.
- OECD (2005b). *Promoting Adult Learning*. Paris: OECD.
- PARN (The Professional Associations Research Network) (2002). Lifelong learning: Past, present and future. *Continuing Professional Development Spotlight*, 27.
- Rubenson, K. (2006). The Nordic model of lifelong learning. *Compare* 36(3), 327–341.
- Schuetze, H. G. and Casey, C. (2006). Models and meanings of lifelong learning: Progress and barriers on the road to a learning society. *Compare* 36(3), 279–287.
- Shiraishi, Y. (1998). Alternative approaches to financing life-long learning. *Country Report*. Japan: OECD.
- The Finland Ministry of Education (2004). *Education and Research 2003–2008 Development Plan*. Helsinki: Ministry of Education.
- Wilfried, K. (2003). *Lifelong Learning in Germany Financing and Innovation: Report of the Federal Ministry of Education and Research (BMBF) for OECD*. http://www.bmbf.de/pub/lifelong_learning_oecd_2003.pdf (accessed May 2009).

Popular Adult Education

S Tøsse, Norwegian University of Science and Technology, Trondheim, Norway

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Folkbildning – Swedish, meaning popular adult education.

Folkeop(p)lysning – Norwegian and Danish, meaning popular adult education.

Op(p)lysning – Norwegian and Danish, meaning enlightenment.

Introduction

Popular adult education is one of those slippery concepts within the field of education that is used to denote different educational aims, ideas, approaches, activities, and programs around the world. It is used more or less synonymously with other concepts; in addition, we also have to bear in mind that it is an English translation of specific national concepts with distinctive characteristics and differences of meanings. As such, popular adult education and its national linguistic synonyms are forms of educational phenomena that are socially constructed in a process of accommodation and transformation, cultural and hegemonic strife, and social movement learning in specific historical contexts. The concept is founded and shaped by the influence of individual educators and movements that have linked popular education to different aims, target groups, educational practices, and curriculum. Taking a worldwide perspective, the term popular adult education covers different trends and philosophies and is applied to so many different practices that it is hard to tell exactly what it means (Kane, 2001: 229–230; 2005: 135). The concept cannot be framed in a single and universal definition but, rather, has to be understood as an open-ended process that is continuously adapted to the concrete situations of local needs to expand people's participation into their learning and living (Han, 1995). The article starts with the different meanings and historical traditions, more deeply explores some Nordic, Latin American, and North American practices and initiatives, presents the current trends, and finally concludes with some common dimensions characterizing the popular adult education approach.

Different Meanings and Historical Traditions

In the European context, the word popular originates from the words people or folk, and popular adult education is associated with a long tradition of people's struggle for enlightenment, access to culture and knowledge, and the development of democracy. The popular concept is historically linked to the role of popular and radical movements as educative forces, for instance, the labor movement, which did not separate education from politics, but regarded education as a guide to social and political action for a better world. Knowledge was power and instrumental in the political struggle, and this struggle itself was educative. In many countries, it has been used more or less synonymously with education of the working class and loosely associated with the interests, aspirations, and struggles of ordinary people for democracy and social change (Martin, 2007). In this broad meaning, popular adult education has been the forerunner for, and almost used synonymously with, adult education (Jarvis, 1999).

In some countries, the popular education concept is also linked to the science-popularizing movement, which can be described as a charitable form of public service by university professors and industrialists in the Enlightenment tradition (Steele, 2007a, 2007b). It led to a multitude of associations for the popularization of science, as exemplified in the British and American mechanics institutes and in the Nordic countries' working men's institutes (from the 1880s) and their Norwegian successors' folk academies. Ideologically, this popular education signified political subversion, a means of class harmonization, or democratic emancipation. In the Nordic countries, these efforts of disseminating knowledge and information to the public (e.g., related to health, social questions, and policy matters) became state funded.

A specific form of the science-popularizing movement was the university extension movement spreading from England at the end of the nineteenth century to many European countries and, later, even to South America (Jarvis, 1992; Wallin, 2000), and resulted in so-called popular or folk universities devoted to the aim of bringing both university culture to the people and developing people's own culture. Van Gent (1992) has suggested the term sociocultural education for similar professionally focused training initiatives such as the people's colleges in Hungary, folk high schools in Germany and the Netherlands, folk

houses and cultural centers in the Netherlands, and the German centers for *Volkshbildung*. A primary aim for some of them was to provide training for people involved in popular education; however, many of these efforts failed to reach the people (Jarvis, 1992).

Although some of the above-mentioned forms were the result of personal initiatives to educate people, the concept of popular adult education is primarily linked to the social mobilization and the educational activities within social movements. Indeed, it is the social movements themselves who have coined the term and adapted it to their practices. In the literature, we even find popular adult education referred to as a separate social movement that is held together by regional or international associations and networks. It has been a world for a nonformal, out-of-school education organized by a variety of groups that have used the popular concept to defend education as a right for all people, to be designed for the people and by the people, and essentially associated to a form of pedagogical praxis that is controlled by participants. In Latin America, Africa, and Eastern Europe, the terms popular or folk education may be used instead of, or in conjunction with, community education (Tight, 1996: 65). Hamilton and Cunningham (1989: 440) regard "community development and popular education . . . as compatible." They do, however, find that subtle but important distinctions can be made, indicating that these two approaches are contradictory. Community development seeks reformation, while popular education seeks transformation. Indeed, a powerful idea, especially in Latin America, has been to make the concept identical to empower oppressed people to take part in the struggle for social and political change. Accordingly, it also shares many similarities with other terms expressing the aims of education such as resistance, liberating, and emancipatory education.

A problem for the demarcation of a popular education territory is that initiatives and practices might well fall within common definitions, but are not always considered as such. This is apparent in political liberation movements in Eastern Europe, for instance, Charter 77 in the former Czechoslovakia, which demonstrated a high activity of seminars, lectures, study groups, and free theater groups (Rubenson, 1995). The fall of communism and the drive for democratization in the former Eastern block coincided with a rise of civil society activities that have a close resemblance to popular education elsewhere. Some countries such as Slovenia have implemented new forms of study circle activities based on Nordic examples (Gougoulakis and Bogataj, 2007).

Regional Differences

The article cannot pay due attention to all forms of popular initiatives, and practices in all parts of the

world, but gives some examples of regional differences. One main form has originated from the ideas of the Enlightenment and Romanticism and has a strong basis in the Nordic countries. A second trend has emerged from the popular revolt against severe oppression in former colony countries or dictatorships. This form has its strongest basis in Latin America, which is the prime example here, but is spread to all parts of the world where people have been restricted the freedom to learn. A third example is popular education initiatives, which are less integrated in social movements and popular forces, but are more a result of personal initiatives and efforts of creating new mass organizations and institutions.

Nordic Popular Adult Education

The Nordic popular adult education is often referred to as a historical tradition that has emerged from the twin influence of the Enlightenment and Romanticism. The Norwegian and Danish concept *folkeop(p)lysning* is, in fact, a composition of enlightenment (*opplysning*), meaning education from above, and the word folk, which is a romantic invention. Sweden has adopted the German concept of *Bildung* and has constructed the word *folkbildning*, which indicates that enlightenment might also come from within (Korsgaard, 1997). A third, and increasingly stronger, influence through the nineteenth century, was the rise of social and democratic movements, which further developed the idea of folk as a political subject and transformed the meaning of popular education to be not only education to the people from above, but also education of the people and the creation of knowledge from below. Since the end of the nineteenth century, popular education in the Nordic countries has been identified as social movement learning within the civil society.

A characteristic of the Nordic situation is the organization of special study associations for popular adult education within movements. It can thus be understood as a popular movement itself and a specific organized activity within movements. In the institutional meaning, especially in Sweden, popular adult education embraces the study associations and the folk high schools. In charge of these, the meaning of popular adult education multiplied into a diversity of meanings which reflected the different ideologies, aims, and activities of the movements. Some made education a part of a distinctive countercultural struggle; however, this element has faded away in the course of time. Another characteristic is the ideal of education as a collective self-education based on the belief that enlightenment and culture can emerge from below, the common people, and that knowledge can be created and disseminated through dialog within groups of equal persons. A third

element is the emphasis on the aim of personal development. A unique character of Nordic popular adult education is also considered to lie in the pedagogy and methodology that is put into practice in study circles (Rubenson, 1995). These are built on the principle of independence of external requirements and are characterized by self-directed learning. In principle, all studies should not be examined and should provide a zone of freedom in which participants can choose subjects according to their real interests and needs (Andersson and Tøsse, 2006). Moreover, the study circles build on the common work of the participants and their experiences, and the active strive for knowledge, and work toward a collective goal.

Popular adult education in the Nordic countries includes a wide range of topics on every degree of difficulty, but is mainly understood as nonformal and out-of-school education, spare-time and hobby-related activities, and, by tradition, often nonvocational. The study circles provide an arena for deliberation and learning in an informal way. As the level of participation is high, especially in Sweden, peaking at 2.8 million participants in the 1990s – that is, a third of the total population – popular adult education may be considered as a specific public sphere that has contributed to establishing and enforcing basic elements in the Nordic model of democracy. The concept has also been used normatively as an instrument in the construction of a Nordic dimension in relation to adult education (Ehlers, 2006). Participation is therefore largely state funded on the ground that the study associations, in addition to being agents of popular movement, contribute to diversity, support culture, and create informed, active, and committed citizens (Rubenson, 1995; Larsson, 2001). The popular adult education field has, especially in Norway, been acknowledged as a part of the educational system and encouraged to focus on the needs and aspirations of people who traditionally do not participate in adult education or are not interested in formal qualifications, for instance, the elderly (Rubenson, 2006). In the recent years, however, the state subsidies have been decreasing; however, popular adult education in the Nordic countries is, as distinct from most other countries, still a state-funded as well as a self-governing area. As stated by Swedish researchers, popular adult education in Sweden is “part of the corporate state lying at the crossroad between civil society and the state ... at the intersection between the system world and the life world” (Gustavsson *et al.*, 1997).

Latin America

Outside the Nordic countries, popular adult education has almost exclusively been associated with the work and pedagogy of Freire, and has been described as “one of the most original and refreshing contributions that Latin America has made to universal pedagogical thinking” (Mera, 2005). Indeed, popular education is almost

conceived as a Latin American invention and mainly associated with the endeavors of social transformation in Third World countries. In the Latin American context, the term popular education, as Kane (2001: 247) observes, has been more sharply defined and the commitment to side with the oppressed is more openly spelled out. Building on Freire’s work, popular education has been advocated as a political, social, and educational process with the overall aim of counteracting the dominant worldview and creating an antihegemonic culture (Kane, 2001: 8–13). Its foundation is the convictions that grassroots people can collectively achieve critical consciousness and, from this awareness, act to challenge unjust uses of power that affects their social realities.

A pedagogical principle that Freire outlined in order to actuate the process of consciousness raising or conscientization was acquiring knowledge by way of problematizing the natural, cultural, and historical reality in which one is immersed. Another basic element is his thinking about language as inherently linked to culture which may thus convey a certain culturally transmitted worldview. Accordingly, language can both question and strengthen culture (Finger, 2005). The popular education approach is therefore making sense of the world to the ordinary people by uncovering and decoding languages (words) and meanings (themes) of their own. This must come through dialog, which in the Freireian sense is a process of generating and sharing the true word and actively naming and transforming the reality of the world (Han, 1995). This fundamental view of the teaching and learning transaction as a dialog is shared by Grundtvig and many other popular education theorists. However, far stronger than most others, Freire maintained that reflection through dialog means action, and emphasized that all education is political. A major aim of popular adult education is therefore to help participants put knowledge into practice. Under dictatorship in Latin America, popular education became a way of doing political work and naturally linked to the radical left-oriented policy and participatory research in cooperation with local groups. A basic principle politically as well as pedagogically is that it must be education with the people, not on their behalf, based on people’s experiences, and aiming toward empowering the ordinary people to become subjects of change.

In Latin America and Third World countries, educators inspired by the pedagogy of the oppressed often had to work among illiterate people; therefore, popular adult education is often identified with literacy and elementary education. Moreover, education with illiterate people also means drawing on popular culture and starting from the concrete, using drama, song, dance, art, and storytelling. Popular culture is here to be distinguished from elitist culture institutions and from the mass culture. One concrete example is Augusto Boal’s experiments with the People’s Theatre in Peru in the 1970s (Boal, 1979). Similar

efforts of using theater, dance, etc., in making the subaltern classes articulators of social and political change and waking up peoples' culture of silence, have also been, more or less successfully, executed in developing countries in Africa and Asia. The popular theater has best served the popular education approach when it has managed to not only make people aware of, but also activate participants (transforming Spectators into Spect-Actors!) in, the critical analysis of what is presented and mobilize them for taking action in their own development. At worst, it has been based on a developmentalist approach whereby popular theater forms are used to communicate government policies or to impose outsiders' views of what people need (Mluma, 1991; Lange, 1995; Bates, 1996).

Basically, popular education is defined by initiatives from below and on the grassroots level, and, in Latin America, has traditionally acted against the (oppressive) state. This was not always true in Europe, and, even in Latin America, could be a public- and state-initiated project. One example is the popular education in Nicaragua. After the revolution in 1979, the Sandinistas immediately started a literacy crusade that they followed up with a popular basic education campaign in which they copied Freireian methodology and principles. Both can be characterized as a massive training of the common people and presupposed an intensive training of teacher as well. Radio and newspapers were also put to use. However, it was a state-driven and-funded project, and some of the popular educators such as Fernando Cardinal occupied posts in the ministries of education in the 1980s (Flores-Moreno, 2005). The Nicaraguan example as well as similar educational initiatives of revolutionary movements in Latin America and elsewhere do, however, question whether education that is totally absorbed by the state or a movement in order to implement a specific policy or ideology can be termed popular. It might rather be characterized as a pedagogicalization of politics (Han, 1995).

North American examples

Popular adult education is primarily a specific form of social movement learning; however, initiatives and activities can often be traced back to individual educators. In North America, the Highlander movement is a prominent example of personal initiative and influence. It all started in 1932 when Myles Horton together with Don West founded the Highlander Folk High School in one of the poorest areas in Tennessee. Highlander did not provide mass education, but worked with leaders from several organizations or communities. Horton's assumption was that the leaders would take what they learned at Highlander and work with other actual or potential leaders and, thus, let the influence of Highlander multiply over larger numbers of people (Peters and Bell, 2001). His early contribution to popular education was then to combine the aims of social movement building and leadership training.

The creation of citizenship schools in the 1950s became, however, a mass activity. By 1970, these schools had helped 100 000 African Americans to read and write (Peters and Bell, 2001). The teachers followed the same principles and ideas as Freire and, in this way, made the citizenship education similar to popular education elsewhere. With the reorganization of Highlander to the Highlander Research and Education Center, its mission became to be an important alliance of the civil rights movement; thus, it retained its popular educational approach.

Another North America example of personal initiative is the Antigonish movement in Nova Scotia led by priest, philosopher, educator, and social reformer Moses Coady (1882–1959). His program was to organize shore fishermen so that they might be able to assist in formulating policies for the industry, promote scientific and technical education, and utilize the methods of producer and consumer cooperation. His work was founded on a strong belief in education as the solution of the economic problems of the region. Moses declared that education must begin with economics and proposed cooperation as an alternative third way between capitalism and state socialism. One of his original contributions to popular adult education was the organizing of mass meetings, which inspired people to join together into small groups called study clubs or discussion circles. As with many Nordic study circles, these groups were part of the self-help movement and did not involve any teacher (Crane, 2001). Development after the war, however, was not in favor of the cooperative ideas based on self-reliance, and, as an example of popular adult education, withered away.

Current Trends

Following Kane (2001), we may identify three currents of thought. One is concentrating on democracy, citizenship, new spaces, and social actors. This trend is intertwined with mainstream pedagogy and makes popular adult education part of the social capital formation that is initiated by social movements and may be seen as occupying and expanding the public space within civil society. This is evidently the case in Northern Europe, which has witnessed a process of convergence between the popular and the general education. In addition, in Latin America, we may see a pronounced permeability and interchange between the discursive configurations of popular and adult education. There is a feeling among some educators that the historical relevance of popular education has been lost and there is a growing willingness to make links with public and traditional schooling. As a consequence, the political aspect is de-emphasized and more importance is given to the quality of education, access, participation, citizenship training, and new social issues (Kane, 2005; Ruiz, 2006). A second trend is focusing on

class and structural change while remaining sensitive to issues of identity and difference. This trend tries to maintain its political vision and commitment. An example is the Popular Education Forum for Scotland, which calls attention to the need for reinvigorating the questions of equality and justice and linking popular education to the political efforts of stimulating a democratic renewal (Martin, 2007). A third trend is less concerned with theory and ideology and more with throwing itself into the struggle and grassroots work. This trend is linked to new social movements that have transformed the popular education discourse from class and structure to ethnicity, race, gender, ecology, and environment. It has adopted a community, practical, and participatory democracy approach of which the participatory budget of Porto Alegre in Brazil serves as a prime example and inspiration (Myers, 2007). New forms of popular education now seem to take place on a global scale with the help of the Internet as is demonstrated by the work of new social movements, the World Social Forum, and Attac which use the World Wide Web in organizing meetings and activities, spreading information, and educating the people. The Internet not only opens up access to expert knowledge for everyone, but also gives the popular forces new possibilities to take part in the oppositional social movement debates of a global call for action against poverty and injustice (Preece, 2006).

In recent years, popular adult education has developed areas of specialism, that is, for specific groups or for specific matters. From the early 1980s, feminists from different countries have gathered under the umbrella term of feminist popular education and formed an international network. A central aim of this is to support the struggles of women in oppressed communities, rather than women in general, and the work is therefore closely related to other nonformal, community, or radical educational practices. The feminists have also linked social justice issues more broadly to the whole person, that is, to embodied learning, which refers to practices that engage the body, mind, spirit, and emotion (Walters and Manicom, 1996; Stromquist, 2004).

Common Dimensions

Despite the regional differences and the different currents of thought, there are at least three dimensions, which, taken together, can give a core meaning of the term. These dimensions are all bounded to a normative philosophy, but the strength of the concept is just that it is based on some few, simple yet powerful ideas: of political commitment in favor of the oppressed, of enabling ordinary people to become subjects of change, of recognizing many types of knowledge, and a methodology that allows these different knowledge to be shared (Kane, 2001: 230).

First, popular adult education has a political–ideological dimension that is linked to a social commitment. The critical approach seems to be at the heart of the popular education approach, and, in dictatorships and oppressive countries, popular education has a specific role of unmasking and challenging unequal relations of power in order to change them. Change is defined as progressive in the interests of a fairer and more egalitarian society, and, in this way, it works in the interest of democracy. However, education and politics are not two sides of the same coin as Myles Horton reminded us. He drew a line between the two, which corresponded to the distinction between education and organizing. If popular education is subordinated to the command of a political goal and ideology, it will turn into propaganda and indoctrination and take the form of political instrumentalism and pseudoschooling (Han, 1995).

Second, popular adult education has a pedagogical–participatory dimension since it tries to recruit and organize people for collective learning and action while, at the same time, maintaining education as a voluntary activity and free from external demands. This requires an educational methodology that radically calls into question the authoritarian teacher-centered practice and transmission of knowledge characteristic of traditional pedagogy (Han, 1995). The popular pedagogy is based on active participation of the learners, dialogical teaching, and sharing of experiences between the learners in such a way that all participants acquire knowledge and take part in a collective learning process. Indeed, the knowledge base of popular education comes primarily from the experiences of the people involved. The method is critical and problematizing, rather than one that produces answers. This education believes in and trusts ordinary people, and is based on the belief that people are capable of understanding and working to improve the conditions of their lives (St. Clair, 2005: 50). Ideally, it is education of equals characterized by a horizontal relationship between teacher and learner.

Third, popular adult education has an action-oriented dimension. Action usually means concrete forms of grassroots activities, which may change power relations. However, critical learning, which corresponds to this education, is also a political act in itself (Newman, 2005) and a critical consciousness is a precondition for action and social change. According to Freire, reflection cannot be separated from action. In Nordic countries, study circles will hardly lead to social and political action; however, it is supposed that they contribute to foster the prerequisite for democratic action (Larsson, 2001). The transformative potential of Nordic popular adult education is not associated with its affiliation to political radicalism, as is assumed in oppressive and developing countries; rather, it comes from the democratic and participatory characteristics of its pedagogy.

Conclusion

Popular adult education is a nonformal type of education that has emerged from the educational work of social movements and may be seen as a paradigm shift to escape from the instrumentalization, systematization, and institutionalization of traditional schooling. It also questions the system of examination in the traditional mode of education, which incessantly serves to reproduce social injustice and its legitimization. The idea of popular education has been formulated as an educational movement to challenge the monopoly of the dominant school mode of education and is capable of presenting alternatives (Han, 1995). Defined in terms of aims, popular adult education may be seen as a specific response in the pursuit of social transformation and change as well as contributing to personal development. Defined in terms of pedagogy and methodology, it is education founded on the principles of voluntarism, institutional independence, dialogical teaching, active participation and the sharing of experiences, and working toward a collective goal. In the Latin American context, popular education represents a critical approach to adult education, which has both a pedagogical aspect of consciousness raising intended to foster a critical awareness of reality, and a political aspect of organizing people for collective action and putting them together for real social change (Han, 1995). In the Northern European context, popular adult education is more about fostering people's participation in order to stimulate personal as well as social development. The meaning of the term will, however, vary with the social context as the concept is an open-ended process that continues to adapt to individual interests and needs, to the popular and collective struggle to expand one's freedom, to the pursuit of popular values and cultures, and to the sharing and acquirement of knowledge and culture.

See also: Adult Education and Civil Society; Community Based Adult Education.

Bibliography

- Andersson, E. and Tøsse, S. (2006). The study associations in Sweden and Norway. A comparative investigation. *Paper to the NFPF's 34 Congress. Örebro, Sweden, 9–11 March.*
- Bates, R. A. (1996). Popular theater: A useful process for adult educators. *Adult Education Quarterly* 46(4), 224–236.
- Boal, A. (1979). *Theater of the Oppressed*. London: Pluto Press.
- Crane, J. M. (2001). Moses Coady and Antigonish. In Jarvis, P. (ed.) *Twentieth Century Thinkers in Adult and Continuing Education*, pp 223–241. London: Kogan Page.
- Ehlers, S. (2006). The decline and fall of "folkeoplysning and adult education": Nordic policy-making during the transition from adult education to adult learning. In Sprogøe, J. and Wither-Jensen, T. (eds.) *Identity, Education and Citizenship – Multiple Interrelations*, pp 339–354. Frankfurt am Main: Peter Lang.
- Finger, M. (2005). Conscientization. In English, L. M. (ed.) *International Encyclopedia of Adult Education*, pp 145–148. Basingstoke: Palgrave Macmillan.
- Flores-Moreno, C. (2005). Popular education and popular schools in Latin America. In Crowther, J., Galloway, V., and Martin, I. (eds.) *Popular Education: Engaging the Academy*, pp 169–180. Leicester: NIACE.
- Gougoulakis, P. and Bogataj, N. (2007). Study circles in Sweden and Slovenia – learning for civic participation. In Adam, F. (ed.) *Social Capital and Governance: Old and New Members of EU in Comparison*, pp 203–236. Münster: Lit Verlag.
- Gustavsson, B., Larsson, S., and Rubenson, K. (1997). Civil society and Swedish popular adult education. *SCUTREA Conference Proceedings*. Leeds: SCUTREA.
- Hamilton, E. and Cunningham, P. M. (1989). Community-based adult education. In Merriam, S. B. and Cunningham, P. M. (eds.) *Handbook of Adult and Continuing Education*, pp 439–451. San Francisco, CA: Jossey-Bass.
- Han, S. (1995). *An Exploratory Study of the Ideas and Ideologies of Popular Adult Education: Implications for Understanding Korean Minjung Education Movement*. Unpublished Doctoral Dissertation. State University of New York at Buffalo.
- Jarvis, P. (ed.) (1992). *Perspectives on Adult Education and Training in Europe*. Leicester: NIACE.
- Jarvis, P. (1999). *International Dictionary of Adult and Continuing Education*. London: Routledge.
- Kane, L. (2001). *Popular Education and Social Change in Latin America*. London: Latin America Bureau.
- Kane, L. (2005). Popular education seen "from afar" *Adult Education and Development* 64, 131–140.
- Korsgaard, O. (1997). *Kampen om Lyset (The Struggle for Enlightenment)*. København: Gyldendal.
- Lange, S. (1995). *From Nation-Building to Popular Culture: The Modernization of Performance in Tanzania. CMI Report 1995: 1* Bergen: Chr. Michelsen Institute.
- Larsson, S. (2001). Seven aspects of democracy as related to study circles. *International Journal of Lifelong Education* 20(3), 199–217.
- Martin, I. (2007). Introductory essay. In Crowther, J., Martin, I., and Shaw, M. (eds.) *Popular Education and Social Movements in Scotland Today*, pp 1–25. Leicester: NIACE.
- Mera, C. Z. (2005). Latin America and adult education. In English, L. M. (ed.) *International Encyclopedia of Adult Education*, pp 345–348. Basingstoke: Palgrave Macmillan.
- Mlama, P. M. (1991). *Culture and Development. The Popular Theatre Approach in Africa*. Uppsala: The Scandinavian Institute of African Studies.
- Myers, J. P. (2007). Citizenship education practices of politically active teachers in Porto Alegre, Brazil and Toronto, Canada. *Comparative Education Review* 51(1), 1–24.
- Newman, M. (2005). Popular teaching, popular learning and popular action. In Crowther, J., Galloway, V., and Martin, I. (eds.) *Popular Education: Engaging the Academy*, pp 22–31. Leicester: NIACE.
- Peters, J. M. and Bell, B. (2001). Horton of highlanders. In Jarvis, P. (ed.) *Twentieth Century Thinkers in Adult and Continuing Education*, pp 242–257. London: Kogan Page.
- Preece, J. (2006). Beyond the learning society: The learning world? *International Journal of Lifelong Education* 25(3), 307–320.
- Rubenson, K. (1995). *Swedish Popular Adult Education: Some Comparative Remarks*. (English translation of article in Bergstedt, B. and Larsson, S. (eds.) *Om folkbildningens innebörder*). Linköping: Mimer.
- Rubenson, K. (2006). The Nordic model of lifelong learning. *Compare* 36(3), 327–341.
- Ruiz, M. (2006). Research in popular and adult education in Latin America. In Merriam, S. H., Courtenay, B. C., and Cervero, R. (eds.) *Global Issues and Adult Education*, pp 412–421. San Francisco, CA: Jossey-Bass.
- St.Clair, R. (2005). University faculty and popular education in the United States. In Crowther, J., Galloway, V., and Martin, I. (eds.) *Popular Education: Engaging the Academy*, pp 43–53. Leicester: NIACE.

- Steele, T. (2007a). With "real feeling and just sense". Rehistoricising popular education. In Crowther, J., Martin, I., and Shaw, M. (eds.) *Popular Education and Social Movements in Scotland Today*, pp 95–105. Leicester: NIACE.
- Steele, T. (2007b). *Knowledge is Power! The Rise and Fall of European Popular Educational Movements. 1848–1939*. Oxford: Peter Lang AG International Academic Publishers.
- Stromquist, N. P. (2004). The educational nature of feminist action. In Foley, G. (ed.) *Dimensions of Adult Learning*, pp 35–52. Milton Keynes: Open University Press.
- Tight, M. (1996). *Key Concepts in Adult Education and Training*. London: Routledge.
- Van Gent, B. (1992). Netherlands. In Jarvis, P. (ed.) *Perspectives on Adult Education and Training in Europe*, pp 19–33. Leicester: NIACE.
- Wallin, K. E. (2000). *Folkbildning på export? (Popular Adult Education on Exportation)*. Stockholm: Pedagogiska institutionen, Stockholms universitet.
- Walters, S. and Manicom, L. (eds.) (1996). *Gender in Popular Education*. London: Zed Books.
- Hurtado, C. N. (2005). Contribution to the Latin America debate on the present and future relevance of popular education. *Adult Education and Development* **64**, 101–112.
- Kane, L. (2000). Popular education and the landless people's movement in Brazil (MST). *Studies in the Education of Adults* **32**(1), 36–50.
- Leis, P. (2005). Reflections on popular education. *Adult Education and Development* **64**, 113–125.
- Mignon, J.-M. (2008). Popular education in France. *Lifelong Learning in Europe* **XIII**(2), 103–110.
- Peters, J. M. and Bell, B. (2001). Horton of highlanders. In Jarvis, P. (ed.) *Twentieth Century Thinkers in Adult and Continuing Education*, pp 242–257. London: Kogan Page.
- Rivero, J. (2005). Key contributions of the popular education movement. *Adult Education and Development* **64**, 93–100.
- Welton, M. R. (1995). Amateurs out to change the world: A retrospective on community development. *Convergence* **xxviii**(2), 49–61.
- Yoo, Sung-Sang (2006). *Popular Education in Asia: A Comparative Study of Freirian Legacies in Popular Education of the Philippines and South Korea*. Unpublished Doctoral Dissertation, University of California, Los Angeles.

Further Reading

- Åberg, P. (2008). *Translating Popular Education. Civil Society Cooperation between Sweden and Estonia. Stockholm Studies in Politics 124*. Doctoral Dissertations 24. <http://su.diva-portal.org/smash/record.jsf?pid=diva2:198245> (accessed May 2009).
- Carrillo, A. T. (2007). Paulo Freire and Educación popular. *Adult Education and Development* **69**, 27–42.
- Foley, G. (1998). Clearing the theoretical ground: Elements in a theory of popular education. *International Review of Education* **44**(1–2), 139–153.
- Grace, A. P. (1995). The Gospel according to father Jimmy. *Convergence* **xxviii**(2), 63–77.

Relevant Websites

- <http://www.cpepr.info> – Center for Popular Education and Participatory Research.
- <http://www.highlandercenter.org> – Highlander Research and Education Center.
- <http://www.peopleseducation.org> – Institute for People's Education and Action.
- <http://www.poped.org> – Open Source Web Development for Social Change.
- <http://www.pepe.org> – pepe.org.
- <http://www.infed.org> – The Encyclopaedia of Informal Education.
- <http://www.popednews.org> – The Popular Education News.
- <http://www.trapeze.org> – Trapeze: Popular Education Collective.
- <http://www.cpe.uts.edu.au> – University of Technology, Sydney.

Provision of Prior Learning Assessment

P Andersson, Linköping University, Linköping, Sweden

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Prior learning assessment (PLA) is a topic that has emerged as a central aspect of the policy and practice of lifelong learning around the world. The topic is of interest in adult education as well as in higher education. It could be not only a matter of defining eligibility and widening of access to education, but also a matter of assessment for credit so that people do not study what they already know. Further, PLA is of interest in relation to work life, where it could facilitate mobility, competence development, and competence utilization. It is not necessarily always a formal assessment; more informal methods for recognition of and making learning visible (Bjørnåvold, 2000) could be seen as expressions of the idea of PLA.

The provision of PLA differs around the world. This article will not give a description of the provision of PLA in different countries. Rather, the aim of the article is to provide some general perspectives on PLA provision, independent of national context.

Different Terms and Acronyms

There are different terms used around the world to express the idea of giving recognition to prior learning: PLA is used in the United States; prior learning assessment and recognition (PLAR) in Canada; recognition of prior learning (RPL) in South Africa and Australia; accreditation of prior (experiential) learning (AP[E]L) in the UK; validation of prior learning (VPL) in the Netherlands; and *la validation des acquis* (recognition of experiential learning) in France, for example. These different terms will not be elaborated in detail here. Instead, the article starts from the general idea of PLAR to provide a discussion that is relevant for all readers, independent of national background.

The General Idea of PLAR

The general idea of PLAR covers a number of different aspects, of which the most central aspects will be discussed here.

PLA is about giving recognition to prior learning, irrespective of when, where, and how learning has taken place. The prior of prior learning does not prescribe a certain time gap between learning and assessment. It could be a matter of recent learning, but the time gap could also be wide – the basic condition here is the assumption that the individual has present knowledge or competence as a

result of prior learning. Where the prior learning was acquired could have taken place in people's daily life, in the work place, or in a foreign country for example, and learning could have taken place in the context of formal education – if that learning has not received relevant recognition – or more commonly in a nonformal setting, thereby giving this learning a more formal recognition.

One limitation in the irrespectivity of where prior learning took place is that the assessment within a course, of what has been learned in that particular course, is not normally defined as PLA. Rather, a second central aspect of the idea of PLA is that it is strongly related to transfer and mobility. PLA should facilitate transfer and mobility of knowledge and of individuals. What has been learned before – prior learning – is assessed to get present recognition. In addition to this transfer in time, PLA is also often a matter of transfer in space. The knowledge that is assessed is not normally developed in the same context as where it is assessed, which lies in stark contrast to more traditional educational assessments where the main object of assessment is the knowledge developed within a certain course or educational program. In PLA, informal and nonformal learning has taken place in everyday life, work life, voluntary organizations, study circles, etc., and is (often) assessed in an educational context. Another example of transfer through PLA is that prior learning could have been formally assessed and certified in one context, but if this certification is not valid in a new context a renewed assessment might be needed, like when professional immigrants come to a new country. Of course, there are exceptions from this transfer in space. Some processes of PLA are, for example, situated in a work place, assessing knowledge that has been developed before, informally, in the same workplace context.

The History of PLA

The idea of giving recognition to prior learning is not new. It has been present in different contexts for a long time, even if these practices have not been given a name.

Although its origins are commonly traced to the post-World War II USA (Weil and McGill, 1989), when returning veterans wanted their skills recognized by universities, RPL is not a totally new phenomenon [...]. Rather, it is the formalization and (re)naming of pre-existing practices concerning alternative access and admissions, mature age entry, and so on (Harris, 2006: 3).

Thus, the more explicit ideas of PLA have roots in the post-World War II era in the USA, in the context of giving recognition to the experiences of the war veterans, to help them settle back into a normal life. There are also roots in France, where “[a] law passed in 1934 concerning engineering allowed people over 35 years old with no higher education qualifications, who had worked for at least five years on activities which are normally those of engineers, to gain the official title of engineer through the preparation and presentation of a dissertation based on their work experience” (Feutrie, 2000). In the 1970s, the interest in PLA was growing in the USA, with the idea of social justice through widening access to higher education. In the 1980s the ideas from the USA had been transferred to the UK, where PLA moved on to another area, the labor market and vocational training. Aiming at economic development through better competence utilization, methods for recognition of vocational competencies were developed. In France VAP was developed. Of course, the idea spread to other countries too, but the full history of PLA is beyond the scope of this article (for further descriptions see, e.g., Evans, 2000). However, two more emerging aims of PLA should be mentioned: (1) from South Africa, the utilization of PLA in a process of social change (in South Africa, RPL) as a possible tool to give recognition to prior learning and experiences among those with limited access to education during the apartheid era; (2) the identification of possibilities of PLA for developing self-knowledge and strengthening self-esteem (see, e.g., Cleary *et al.*, 2002; Andersson, 2006).

Differing Aims and Approaches

As we see in the history described above, different aims of PLA are discerned: social justice, economic development, social change, and individual development. It should be pointed out that these different aims are parallel to aims promoted in relation to the provision of adult education. PLA should promote social justice, for example, in relation to the admission to higher education – like availability of adult education is a matter of social justice. PLA should also promote economic development, through the process of recognizing knowledge and competence, which results in better opportunities to make use of these in the work life. This is parallel to the provision of vocational education/training for adults. Further, the idea of PLA as a tool for social change, like in South Africa, is similar to ideas of popular adult education, related to social movements with the idea of changing society through literacy and mobilization, for example. Finally, PLA for individual development is a matter of strengthening self-esteem and self-confidence, which is an important aspect of some adult education as well.

Different approaches to PLA have been typified in different ways. One categorization is between PLA adapted to the system, as compared to PLA changing (or intending to change) the system (Andersson *et al.*, 2003, 2004). Here, system should be understood in a generative way – it could be contributions to changes of the social system as a whole, like in South Africa, or a change of the system of education or of a single course. Among other things, this puts focus on the question of power in PLA provision. If PLA is adapted to a certain system, this means that PLA is based on an established distribution of power, where the assessment is often based on norms and criteria that are rooted in this system. For example, the assessment could be based on the pre-assumption that learning takes place in organized educational processes, rather than through informal processes, which could result in the exclusion of certain knowledge from the process or recognition. The Procrustean RPL (Jones and Martin, 1997) is an example of this type of approach, assessing people as if they were all cast in the same mold.

According to Procrustes, a ruler in Greek mythology, everyone could fit into his bed regardless of their size and shape. If anyone was too short, he placed them on the rack and stretched them. If they were too long, he would chop off their feet. (Jones and Martin, 1997: 16).

On the other hand, there are approaches that are, more or less explicitly, aimed at changing the system. Trojan horse RPL (Harris, 1999) is a category of approaches that bring new groups into a system, and by taking their experience and knowledge seriously a change of the system from within becomes possible.

There are also other ways to categorize different approaches. For example, the difference between credit-exchange and developmental models is identified (Butterworth, 1992), and another categorization consists of the technical/market, the liberal/humanist, and the critical/radical perspectives (Breier, 2005). These differing aims and approaches have turned up in different contexts during the history of PLA.

A Broad Spectrum of PLA Methods

As mentioned, PLA in a general sense could include a broad spectrum of activities, not only formal assessments but also other types of processes to give recognition to prior learning. On the one hand, some of these activities are formal and standardized, with multiple-choice tests – like scholastic aptitude tests. On the other hand, there are PLA activities that focus on the individual and particular aspects of knowledge, to give recognition to a variety of competencies – for example, in a portfolio. This broad

approach to the idea of assessment and recognition of prior learning, focusing not only the assessment *per se* but also recognition in a wider sense, means that the idea covers a spectrum from divergent (exploring and descriptive) assessment to convergent (controlling and examining) assessment methods (cf. Torrance and Pryor, 1998). Another definition of divergent and convergent assessments is that divergent methods focus on what the individual knows, while convergent methods focus on if the individual knows certain (predefined) things. It should be noted that this division between convergent and divergent methods is not a dichotomy. It is rather, as in other types of assessments, a matter of a continuum, where a certain method to some extent is more or less convergent and less or more divergent. However, putting forth this divergent–convergent dimension in the context of PLA highlights the problems in terms of validity that might be the result if the irrespectivity of when, where, and how is taken seriously, unless more divergent methods than in traditional educational assessment are considered. That is to say there will be problems making valid assessments with convergent methods, if the degree of validity refers to the extent to which PLA provides a fair assessment of what has been learnt in different times, contexts, and ways.

Different Methods in Practice

The ambitions of PLA could be reached in a number of ways. Different methods are used in different contexts, and some of these methods will be exemplified here, namely portfolios, interviews, standardized tests, traditional tests, and authentic tests. These methods are, of course, not only used in PLA, but they are also often used in different PLA approaches and illustrate the broad spectrum of PLA.

The portfolio represents a group of methods used in somewhat different ways. Challis (1993) has, for example, identified the difference between the outcome-related and the self-oriented portfolios. In the portfolio, the PLA candidate brings together documentation or proofs of his/her prior learning. These could be formal documents such as grades and certificates as well as less formal documents such as a testimonial from a former employer or an NGO (nongovernmental organization) where the candidate has done voluntary work. Another part of the contents of a PLA portfolio could be examples from your prior production of goods or texts. The proofs could be brought together and assessed in relation to certain convergent learning outcomes, but they could also be more divergent and self-oriented, that is, the task of the candidate in the latter case is to prove the personal competence, and it is the assessor who might relate the proofs to goals or criteria.

Interviews are used in PLA, for example, as a more divergent starting point to find out the potential of the candidate, particularly when the area of knowledge to be assessed is not given. The interview could result in a mapping of (possible) knowledge and competence, which might be further assessed with other methods that provide more valid proof of the knowledge in question. Depending on what knowledge area the assessment covers, interviews could also be used in more convergent PLA. For example, an oral examination could give a better picture than a written test of what an individual knows in a theoretical subject area, if the knowledge is developed in untraditional ways and without learning the traditional way of presenting it with paper and pencil.

Standardized tests are another method used in PLA. A typical example is a scholastic aptitude test. A multiple-choice aptitude test can be used in the selection of applicants, for example, for higher education. The predictive test assesses knowledge independent of where the learning has taken place. Thus, any prior learning could be important when taking such a test, even if the convergent character of the method means a limitation of what is actually made visible.

Traditional tests – tests (or other assessment methods) used in the school system – could be used to assess prior learning too. However, a problematic issue could be that these tests are adapted to an organized learning in formal education, rather than to more divergent learning experiences from other contexts, which might reduce their usefulness in PLA. It should also be noted that all educational assessments to some extent could be a matter of PLA – as an assessor you cannot know if/what the students or pupils have learned in the class or somewhere else.

Authentic tests are still another group of methods in PLA, aiming at assessing competencies in an authentic or simulated environment. The rationality of using authentic tests in PLA is that the results of prior learning that has taken place outside school are assessed there too. These authentic tests are, for example, used in the assessment of vocational competencies developed in the work place, competencies that in this way are assessed in the original context of learning.

Theories of Learning Underpinning PLA

We can see how different theories of learning explicitly or implicitly underpin some of the different approaches to PLA that are discussed above. Here the portfolio method and the authentic assessment exemplify this and will be related to different theories of learning (these theories will not be described in detail here, but will only be mentioned briefly).

Experiential learning (Kolb, 1984) is the basis of many portfolio methods in PLA. The basic idea is the learning cycle, where concrete experience, observation, abstract conceptualization, and active experimentation, etc., form a process of learning from experience. However, there is a discussion concerning how these methods value individual experiences and knowledge as compared to collective experiences and learning processes. Individual(istic) portfolio methods are thus questioned based on the assumption that informal learning in collective contexts is not valued in a fair way in these approaches.

The idea of situated learning (see, e.g., Lave and Wenger, 1991) is (at least implicitly) the basis of authentic assessment methods, where the ambition is to assess knowledge in the context of where it has been developed and thus is situated. The questions concern if these methods really become authentic and holistic, or if the result is a behavioristic and atomistic approach (the latter seems to be the case when detailed lists of required competencies are the basis of, for example, certain vocational qualifications). These (or this group of) situated learning theories also help us understand problems with PLA in relation to transfer and mobility. From a situated perspective, it is not evident that prior learning from one context could be transferred and having the same value in another context, where the demands are actually different. The differences are related to tools (language as well as material tools) and to the social systems.

The Value of PLA

What is the value of developing different PLA methods? The main value is that PLA identifies the value of knowledge (stemming from prior learning). On the one hand, the knowledge could have an exchange value. When the PLA process results in a formal documentation of knowledge or competence, the exchange value means that the candidate through this documentation, for example, could be admitted to an education/training program, could get a job and/or a higher salary. More informal assessment or recognition processes could of course have similar results as well, where prior learning is given an exchange value. Further, the exchange value of knowledge, identified in PLA, is not limited to the individual level. For example, it might have an exchange value for a company to know, and be able to show potential customers, that there is a high level of competence among its employees. On the other hand, knowledge often has a use value. Thus, PLA is identifying knowledge that is useful. When we are aware of this knowledge, it is easier to use it, to put it into action. This value is also relevant on an individual as well as on an organizational level – the value for the company is not only the exchange value but also the possible utilization of the employees' prior learning.

Critical Perspectives on PLA

The development of PLA has been discussed and criticized from different perspectives. Some of these discussions will be covered here, concerning the situation of prior learning/knowledge, the authenticity of authentic assessments, and the governing character of PLA. For a more extensive discussion, drawing on perspectives from assessment theory, the sociology of education, poststructuralism and situated knowledge/learning theory, activity, actor-network and complexity theory, and symbolic interactionism, see Andersson and Harris (2006).

PLA has been developed in a Western context, with an underpinning perspective on learning as an individual process. As mentioned briefly above in relation to the portfolio, this individualistic approach has been questioned, and not only in relation to the portfolio. The main point in this critique is that (particularly) informal learning is taking place in a collective process, which is better understood in terms of situated learning or situated knowledge (see, e.g., Michelson, 2006). A focus on assessing individuals, and individual learning outcomes, thus means that a lot of experiential learning remains invisible, in spite of the opposite ambition of PLA. The argument is against the Enlightenment epistemology and for a perspective of situated knowledge, or as Michelson (2006) puts it:

Reinscribed within a theory of situated knowledge, RPL can become a venue for examining how each of us moves back and forth between our own particular stories and the social production that is knowledge and for challenging oppressive taxonomies of knowledge and the power relationships they enact. It can grant visibility to knowledge that is valuable for its divergence from academic ways of knowing, not only its similarity, and affirm knowledge produced outside epistemologically-sanctioned locations, through dialogue within (and, when we are lucky, between) historically-situated communities. (Michelson, 2006: 157)

Are authentic assessments really authentic? As mentioned, some PLA approaches lean on assessment methods that are situated in an authentic or simulated context. The ambition is to assess in the situation where knowledge or competence was developed and is enacted. However, it could be argued that the assessment situation to some extent is artificial rather than authentic, even if it is situated in a context where informal learning takes place, outside the school context. It is not a part of the daily practice to be formally assessed when you use your knowledge (cf. Bowden and Marton, 1998). Thus the assessment results are not necessarily more valid in this case, even if they are viewed as authentic.

Do assessments adapted to the system provide a valid measure of prior learning that has taken place outside this (school) system? This exclusion of knowledge in, for

example, the Procrustean RPL has already been discussed above, and is a matter of governing of what knowledge should count. However, PLA is not only governing what knowledge counts. In a discourse analysis of policy on adult education and lifelong learning, PLA is identified as a technique for constructing and governing the adult learner (Andersson and Fejes, 2005). When learning is assessed more broadly, new knowledge about the learners is produced, which could be the basis of a more extensive governing:

Assessment, including validation, is a technique that colonizes the human as a knowledgeable subject; he is created as a subject by being an object of knowledge production. One way of reasoning about this colonization of the entire subject is the objective of knowledge. Formal knowledge has been, and is, a way of controlling the subject. The documentation in itself is an objectification of the subject and is the starting point from where techniques of governing are set in motion. Knowledge about the subject to be governed is the basis of all governances (Foucault, 1991) and, therefore, informal knowledge has not been given the same attention. What we now see is a trend where the informal and non-formal competence/knowledge should be transformed into formal knowledge. Consequently, this knowledge will also be the foundation of governing and control. Everything you do, lifelong and lifewide, constitutes experiences that are part of the construction of the competent adult. The subject to be governed is constructed as a different subject than was previously the case. (Andersson and Fejes 2005: 610)

Conclusion

We have seen how the provision of PLA has developed in a divergent way – with differing aims, approaches, and methods, more or less adapted to, or potentially changing, the present system of education, and more or less convergent/divergent. A particular approach also means that a certain epistemological position is taken – explicitly or implicitly. Independently of approach, a process of valuing knowledge is enacted through PLA, a process that identifies the exchange and/or use value of knowledge and that includes knowledge stemming from informal and nonformal learning to a higher extent than other educational assessments. However, there are also critical perspectives on PLA. Even if new types of learning/knowledge are included, there are also processes of exclusion present in PLA. For example, PLA has been criticized for its individualistic approach, which possibly excludes knowledge that is developed and situated in collective processes. This way of defining what knowledge that counts is one side of the coin when it comes to critique of the governing aspect of PLA – the other side is the

governing of the adult learner, where PLA means that the process of assessing a broader scope of learning also means that it is governing in a more extensive way as compared to traditional educational assessment.

See also: Access and Equity in Higher Education; Educational Measurement: Overview; Qualifications Frameworks and their Role in the Reform of Education and Training; Situated View of Learning.

Bibliography

- Andersson, P. (2006). 'Validation': Mobilisation and disciplination. *Journal of Adult and Continuing Education* 12, 139–155.
- Andersson, P. and Fejes, A. (2005). Recognition of prior learning as a technique for fabricating the adult learner: A genealogical analysis on Swedish adult education policy. *Journal of Education Policy* 20, 595–613.
- Andersson, P., Fejes, A., and Ahn, S-e. (2004). Recognition of prior vocational learning in Sweden. *Studies in the Education of Adults* 36, 57–71.
- Andersson, P. and Harris, J. (eds.) (2006). *Re-Theorising the Recognition of Prior Learning*. Leicester: NIACE.
- Andersson, P., Sjösten, N.-Å., and Ahn, S-e. (2003). *Att värdera kunskap, erfarenhet och kompetens. Perspektiv på validering*. Stockholm: Myndigheten för skolutveckling, Forskning i fokus, nr. 9.
- Bjørnåvold, J. (2000). *Making Learning Visible. Identification, Assessment and Recognition of Non-Formal Learning in Europe*. Tessaloniki: Cedefop – European Centre for the Development of Vocational Training.
- Bowden, J. and Marton, F. (1998). *The University of Learning*. London: Kogan Page.
- Breier, M. (2005). A disciplinary-specific approach to the recognition of prior informal experience in adult pedagogy: "rpl" as opposed to "RPL." *Studies in Continuing Education* 27, 51–65.
- Butterworth, C. (1992). More than one bite at the APEL: Contrasting models of accrediting prior learning. *Journal of Further and Higher Education* 16, 39–51.
- Challis, M. (1993). *Introducing APEL*. London: Routledge.
- Cleary, P., Whittaker, R., Gallacher, J., et al. (2002). *Social Inclusion through APEL: The Learners' Perspective, Comparative Report*. Glasgow: Glasgow Caledonian University.
- Evans, N. (ed.) (2000). *Experiential Learning around the World: Employability and the Global Economy*. London: Jessica Kingsley.
- Feutrie, M. (2000). France: The story of *La Validation des Acquis* (Recognition of Experiential Learning). In Evans, N. (ed.) *Experiential Learning around the World: Employability and the Global Economy*, pp 103–107. London: Jessica Kingsley.
- Foucault, M. (1991). *Discipline and Punish: The Birth of the Prison*. Harmondsworth, UK: Penguin.
- Harris, J. (1999). Ways of seeing the recognition of prior learning (RPL). What contribution can such practices make to social inclusion? *Studies in the Education of Adults* 31, 24–39.
- Harris, J. (2006). Introduction and overview of chapters. In Andersson, P. and Harris, J. (eds.) *Re-Theorising the Recognition of Prior Learning*, pp 1–29. Leicester: NIACE.
- Jones, M. and Martin, J. (1997). A new paradigm for recognition of prior learning (RPL). In Fleet, W. (ed.) *Issues in Recognition of Prior Learning: A Collection of Papers*, pp 11–19. Melbourne: Victoria RPL Network.
- Kolb, D. A. (1984). *Experiential Learning: Experiences as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice-Hall.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.

- Michelson, E. (2006). Beyond Galileo's telescope: Situated knowledge and the recognition of prior learning. In Andersson, P. and Harris, J. (eds.) *Re-Theorising the Recognition of Prior Learning*, pp 141–162. Leicester: NIACE.
- Torrance, H. and Pryor, J. (1998). *Investigating Formative Assessment*. Buckingham: Open University Press.
- Weil, S. and McGill, I. (1989). *Making Sense of Experiential Learning: Diversity in Theory and Practice*. Buckingham: SRHE/Open University Press.

Further Reading

- Fraser, W. (1995). *Learning from Experience: Empowerment or Incorporation?* Leicester: NIACE.
- Harris, J. (2000). *Recognition of Prior Learning: Power, Pedagogy and Possibility*. Pretoria: HSRC.
- Mandell, A. and Michelson, E. (1990). *Portfolio Development and Adult Learning: Purposes and Strategies*. Chicago, IL: CAEL.

The Age of Learning: Seniors Learning

P Jarvis, University of Surrey, Guildford, UK

© 2010 Elsevier Ltd. All rights reserved.

Older learning has become a major focus in international and national government's educational and social policies. Indeed, in support of the International Year of Older Persons, United Nations Scientific, Educational and Cultural Organization (UNESCO) and the National Institute for Adult Continuing Education in UK combined to issue a small pamphlet calling for older adults to have opportunities: to manage themselves and their quality of life; to have access to adult learning without there being an age limit to learning; for new learning; to contribute to the development of both themselves and their societies (UNESCO, nd). This article explores so many of the developments that have occurred in recent years that indicate how societies are changing in this respect. It has three main parts: the age of learning, retirement and retirement education, and organizations that promote older learning. Finally, there is a brief conclusion which makes reference to the wider benefits of learning.

The Age of Learning

The global economy has also been called the knowledge economy: products and processes have to be developed and marketed by the most efficient and effective methods. This has called for a greater investment in all forms of knowledge production and for a greater proportion of knowledge workers than ever before in humankind's history. Education has become more widely accepted as a lifelong process than ever before, despite the fact that the first book on lifelong education was written as early as 1929 (Yeaxlee, 1929). But now, it is accepted that this is an age of learning, rather than of education, and lifelong learning is a frequently employed term: it is recognized that seniors can and do continue their learning throughout the lifespan. However, it is important to recognize that they have been the driving forces of global capitalism that have generated this changed approach to learning (Jarvis, 2007; *inter alia*). It is significant to note, however, that as Stehr (1994) points out the new learning age emphasizes scientific, technological, and social scientific knowledge rather than the humanities, as Kerr *et al.* (1973) predicted. This restricted approach to knowledge meant that the broad educational program of the educated person began to appear out of date as competent professionals were sought. Now, however, a new model might be appearing within the context of lifelong learning whereby younger adults concentrate upon the sciences, technology, and the work-based subjects while liberal education

becomes a focus in the latter years of life when seniors have the time and inclination to reflect upon life itself and its meaning – if it has one. The education of elders, therefore, has assumed a totally different approach as a leisure-time pursuit which includes academic learning but which is not exclusive to it. Indeed, one other effect of globalization on the lifelong learning of elders has been the growth in educational tourism – an activity in which healthy seniors have been able to participate fully, as is seen below.

While globalization is the major cause for the development of this learning age, the focus on older learners may be seen as much as a result of demographics since we are living longer in an aging world. This is certainly the outcome of the improved standard of living that many people have enjoyed over the past century. Globalization and demographics, then, underlie the development of educational gerontology.

Education and Learning

In order to explore the idea of seniors' learning, it is necessary to unravel the complexities of definition underlying lifelong learning. The European Commission, for instance, defines it thus:

all learning activity undertaken throughout life, with the aim of improving knowledge, skills and competences within a personal, civic, social and/or employment-related perspective. (EC, 2001: 9)

However, this definition actually conceptually confuses education and learning: learning, however, is personal and existential and transcends education. While there has been considerable debate within education about the nature of learning and a variety of different approaches and definitions have been suggested, combining them all within an all-embracing definition, learning, which is a lifelong process, may be defined as:

the combination of processes throughout a life time whereby the whole person – body (genetic, physical and biological) and mind (knowledge, skills, attitudes, values, emotions, beliefs and senses) – experiences social situations, the perceived content of which is then transformed cognitively, emotively or practically (or through any combination) and integrated into the individual person's biography resulting in a continually changing (or more experienced) person (Jarvis, 2006: 134).

This definition emphasizes the fact that learning is not just a cognitive exercise and that our whole life experience is learned. In contrast to the existential, lifelong learning may also be seen as a social institution which is as education, as opposed to personal learning, and may be defined as:

every opportunity made available by any social institution for, and every process by which, an individual can acquire knowledge, skills, attitudes, values, emotions, beliefs and senses within global society (Jarvis, 2007: 99).

Lifelong learning, then, is personal existential learning and recurrent education. In a sense, these two definitions also combine the more traditional ideas of informal, non-formal, and formal learning: formal learning being that which is organized within the traditional educational institution, nonformal is organized learning but does not occur within the traditional classroom setting and informal learning occurs during the processes of everyday interaction. As people age, there are fewer opportunities for them to engage in formal learning in many countries of the world, although where open universities have been founded, such as the British Open University and the Greek Open University, age is no barrier to entry to formal education and in many other countries, we are beginning to see universities and colleges open classes to older adults, as we shall see below. Additionally, there are more opportunities provided within society for nonformal learning. It is shown throughout this article that nonformal and even nonacademic learning is where the growth in education for seniors is to be found. This is to be expected since in later life, people follow their interests and attend to activities that are relevant to them – a point that Knowles (1980) stressed when he discussed the practice of andragogy.

Definitions of an Older Person

Traditionally, at the age of 65 years, a person has been deemed to have reached old age, but now we are living longer and so the question of who is old becomes a much more contentious question. Indeed, in 2004, in the UK, a 65-year-old male could expect to live a further 16.7 years, an increase of 4.4 years since 1971 whereas a female could expect to live a further 19.6 years, and increase of 3.3 years since 1971 (HMSO, 2006: 100–101). This has meant that the concept of old age is undergoing something of a re-definition with some educational gerontologists suggesting that middle age now finishes at 70 years. (This was a suggestion made by Dr Jim Fisher of Milwaukee Wisconsin University.) Nevertheless, it is still possible to treat the start of the third age in life as something between 60 and 65 years and that of the fourth age as

75 years, or as when a person loses the capacity to live independently. However, it is possible in the UK to join the University of the Third Age from the time when a person is no longer in full-time paid employment. What is clear is that people are living longer and that the span of old age is increasing quite considerably, which has resulted in some research projects now being government funded into the very old in order to prepare policies for future eventualities.

Statistics

It is impossible in a brief article like this to review the population statistics for the whole world but it is commonly accepted that the world's population is aging and Western Europe is a graying population (HMSO, 2006: 11) – it is the oldest region in the world (UNESCO, nd) – and the figures for the United Kingdom tend to confirm this. By the year 2021, it is estimated that UK will have a population of 64 729 000 people, of whom 19.67% will be 65 years and older while 9.47% will be 75 years and older. In contrast, in 1971, the population was 55 928 000 of whom 13.25% were 65 years and over and only 1.54% was 75 years and over, but in that year there were 25.49% of the population under 16 years of age compared with an estimated 17.61% in 2021 (HMSO, 2006: 10). Similar statistics can be discovered for the USA which will have an estimated population of 70 million people, 65 years and older by 2030 (McGuire *et al.*, 2005: 443), which is why these authors advocate the introduction of aging education as a national imperative. McGuire *et al.* (2005) now suggest that there are about 16 000 senior citizens centers in the USA.

However, by 2020, it is anticipated by the World Health Organization that 1 billion people worldwide will be classed as older persons, with 700 million living in developing countries: China will have 230 million, India 142 million, Indonesia 29 million, Brazil 27 million, and Pakistan 18 million. These five countries will be among the ten countries with the largest number of older people in the world. Additionally, by 2020, Japan will have 31% of its population classified as elderly and many countries will have a high proportion of its population over the age of 80 years (UNESCO, nd).

Retirement and Preretirement Education

Traditionally, the age of retirement in UK, and elsewhere in the world, has been 65 years for men and 60 years of age for women. As work has become more knowledge orientated, more women have been able to find employment and so their working lives have in some cases become similar to men's. In addition, as society has become more consumer orientated, it has become incumbent upon husbands and

wives to work so that the family can enjoy the wealthy consumer lifestyle presented in the mass media and so we find that in Europe, in 2004, 70.9% of men and 55.7% of women work (HMSO, 2006: 53). However, it is now being mooted that the statutory retirement age should be later – perhaps 68 years. There are at least two reasons for this: first, pensions were planned and financed on a shorter life expectancy and age discrimination has meant that it is more difficult for employers to terminate work when an employee reaches retirement age. Indeed, it has been discovered that many people continue to work long after retirement from their job at retirement age, even to the extent of being re-trained for new forms of employment.

In addition, in the UK, 23% of men and 11% of women worked over 48 h a week in 2005 (HMSO, 2006: 59). This has meant that for many people, retirement is a momentous change in lifestyle and this has been exacerbated by the fact that society has done very little to ease this status transition. Consequently, some enlightened employers have run preretirement education courses or have released their staff to attend such courses a few months or so before retirement in order to prepare for the change. Sometimes, such courses have been run by the human resources or welfare departments of employing organizations but often they have been run by colleges and other educational institutions on behalf of employers. Fundamentally, these courses are designed to help retirees consider their futures in terms of identity, health, wealth, and leisure-time pursuits. In more recent times, however, the idea of preretirement education has spread to developing countries and Ogunbameru and Bamiwuye (2004) have showed how older employees in Nigeria recognized advantages in preretirement education even though it was not widely practiced there – which indicates that preretirement education is now taking the place of instruction during the liminal period of the *rite de passage*.

Retirement is a time of major status change. In earlier times and in more primitive societies, *rites de passage* were very prevalent in social living and had three stages (van Gennep, 1960) – a ritual of separation from the former status, a period of liminality or transition, and finally a ritual of inclusion into the new status. The three stages facilitated the status change and also served to prepare the small community for the transition. It was during the transition period that those changing status were instructed into the ways of their new status before they were ritually incorporated into the new society. While the place of work has remained a more personal place, the wider society has become less personal and much more associational and flexible which means that it can adjust to minor changes, such as people retiring from work, without a great deal of tension so that it now has no need of the ritual. Hence, the ritual of inclusion has disappeared from contemporary society although the ritual of separation has still been continued and people are separated from

their place of employment and go into liminality but now retirement is an unfinished ritual because there is no rite of inclusion (Jarvis, 2001).

Stereotypes of Old Age

Old age clearly spreads over a considerable age range – from about 60 years to over 100 in some instances, so that there are certain stages of aging although studies like Sheehy's (1995) *New Passages* fail to explore them. However with the different phases, there are different images of the elderly and many of these are based on the idea of the aging body although some present pictures of the aging and poorly functioning brain through conditions such as dementia. While the study of stereotypes is beyond the scope of this article (see Featherstone and Wernick, 1995), it is important to recognize the interconnection of body and mind which is closely related to older adult learning and that they can continue to learn and benefit greatly from the process is now beyond question. However, many people in their third age do not like being associated with the image of the fourth age which reflects the ever-increasing life span. But many advertisements, often seeking to get people to invest in retirement insurance policies, depicting happy and contented retirement are now presenting a different, younger stereotype of older persons who are able to get engaged in the world during their extended leisure time.

By contrast, in societies where old age is much more respected, for example, Islamic societies like Turkey, it might be expected that old age would be seen in a much more positive manner but this is not necessarily the case. McConatha *et al.* report that:

Even in Turkey, a collectivist society where intergenerational contact is likely to be more common than in the United States, attitudes towards aging and old age tend to be negative. As the population of older adults grows in Turkey, it becomes increasingly important to address concerns regarding aging and the needs of older adults in order to avoid an increase in ageism. McConatha *et al.* (2004:180)

It is perhaps significant that as society becomes more individuated, younger people have fewer contacts with seniors than they did when the extended family was a more common phenomenon and so there has been an increase not only in family learning, but intergenerational learning and courses for young people to learn how to interact with older people. Lynott and Merola (2007), for instance, report that after a 5-month program of intergenerational learning in a school each year for 4 years, younger people's attitudes toward aging improved considerably. Additionally, such contacts also help older people understand more about the younger generation.

Learning and Intelligence

It was traditionally thought that intelligence increased during the early years of life but by early adulthood it had reached its peak and, thereafter slowly declined so that by old age it was impossible to learn new things: phrases such as “You can’t teach an old dog new tricks” reflects this generally held traditional belief. However, learning is both an existential and an experiential phenomenon as we saw from the above definition so that for as long as we have experiences we are capable of learning and the idea of people’s intelligence declining has been disputed many times since the distinction between fluid and crystallized intelligence was highlighted – the former is biologically based while the latter is experiential. Fluid intelligence is based, to some extent on memory stored in the brain and as the body slowly declines, especially as the synapses in the brain are destroyed, then the mechanism of the brain is affected and it has been suggested that fluid intelligence decreases. In contrast, crystallized intelligence is based on lifelong experience and for as long as individuals continue to remain involved in social life, through social interaction as well as through all other ways of learning, it increases. This has certainly given more credence to the idea that learning is experiential and some learning theorists (Jarvis, 1987, 2006; Kolb 1984; *inter alia*) have emphasized this in their own work. There have been many experiments in more recent times to demonstrate how seniors are still sufficiently intelligent to continue learning at advanced levels: in UK, for instance, there is a national award in adult learner’s week for the senior learner of the year, and so on – the winner recently was a 98-year-old man who had just completed a masters degree! Indeed, Kliegel and Altgassen (2006: 122) have concluded from their comparative study of 45 young adults and 45 old adults that from a developmental perspective, chronological age was not a significant factor in the explanation of individual differences in learning performance, and the existence of many organizations that promote older people’s learning supports this viewpoint. More recently, it has also been noted that many academics continue their teaching long past statutory retirement age even when their universities fail to support them. As Geoffrey Cantor says:

I have joined the army of retired academic in the arts and humanities who, out of commitment to their subjects have remained research-active. But I have been disappointed how little financial support there is (Reisz, 2008: 40).

Perhaps this prolongation of academic work calls for universities to rethink their retirement policies, although it has to be conceded that they still have to prepare for the future by assisting younger academics to develop their own knowledge and skills. This, then, is an employment problem for all employing organizations. At the same time, the fact that academics can continue to do this

demonstrates the fact that age is not necessarily a factor in intellectual decline. Indeed, two recent phenomena have been beginning to appear: the recognition that employers might want to retain their older workers and have to make work attractive to them, and they have to be prepared to adapt to a changing workplace (Yeatts *et al.*, 2000) and retired workers returning to new employments and even being trained for their new role.

Organizations That Promote Older Learning

As early as 1962, the Institute for Retired Professionals was founded in the USA by Hy Hirsch and sponsored by the New York School for Social Research: this was 10 years before the University of the Third Age (U3A) was established by Pierre Vellas at the University of Toulouse in France and it took another 10 years for the latter idea to spread to the United Kingdom. Despite the fact that the founder in UK, Peter Laslett was a Cambridge don, the U3A in UK has little connection with the university world. Each U3A in UK is an independent non-governmental organization (NGO), although there is a third-age trust which, in some ways, acts as a coordinating body. The diversity of U3As in UK is tremendous having 628 separate organizations affiliated to the trust and some 168 628 members as of March 2007 (as per information supplied by the third age trust). Significantly, the difference in the way these two types of U3A were founded reflects something of the difference in their approach to their activities: the ones that follow the pattern of the European continental ones are attached to universities and are much more academically orientated while those in UK tend to emphasize leisure as much as learning and it is generally recognized that only through cooperation can these diverse and independent U3As run more sophisticated academic groups, except the larger group who have sufficient members to run large and diverse programs. A small U3A, with a membership of just over 200, runs about 20 different regular-interest groups, has about eight open lectures a year, and engages in a number of social activities. There is another factor that affects the UK groups and that is the presence of the British Open University; since its foundation at about the same time as the U3As in France, it has offered formal education at a distance at undergraduate and postgraduate level to anybody and many of its students have been seniors. Indeed, for a period there was actually an older-learners-research group based at the UK Open University (Clennell, 1994: 39–46). Older people in UK seeking formal education and academic qualifications would naturally turn to the Open University and so U3As offer a wide intellectual and leisure time program with perhaps more focus of the latter than

their counterparts on the continent of Europe. At the same time, this democratic and local approach to seniors' education has not been without its critics. For instance, Huang (2006) has suggested that the standard of teaching and learning in UK U3As could be improved by utilizing more university-trained teachers since its present mode of operation cannot control the level of education offered and, at the same time, the locally based democratic system might be more efficiently organized. While the accusations have some justification, there is a sense in which the present U3A activities are performing two separate roles – those of leisure and learning – roles which are undertaken in the USA by the two separate arms of the Elderhostel Institute Network. Since the organizations in UK are growing, it is clear that they are responding to local demands and so no incentive appears to exist in many local organizations to change – even though it might be beneficial for them to anticipate future pressures for change at this time. In addition, it is clear that while the organization might be judged as inefficient in one sense, it is certainly democratic and is organized locally by autonomous local committees. Nevertheless, the U3A movement has spread more widely throughout Europe and beyond and has its own International Association of Universities of the Third Age (AIUTA) which has not been extremely effective in recent years although it has been in existence for many years now.

The need for research into older learners was demonstrated by the creation of Third Age Learning International Studies (TALIS) in 1990, founded by Jean Costa who was involved in Toulouse with the original founding of the University of the Third Age: TALIS was something of a breakaway from AIUTA because it wanted to concentrate on the academic study of third-age learning. From its outset, TALIS attracted scholars from all over the world, thus demonstrating the prevalence of older-adult learning.

In the USA, the development of the Institute for Retired Professionals took a totally different route: the Institute had begun under the sponsorship of the New School for Social Research in New York City but spread slowly. In 1976, however, a conference of interested parties led to these becoming known as institutes for learning in retirement. At roughly the same time (1975), in New Hampshire, another movement was born – Elderhostel and this grew extremely rapidly – offering educational travel. By 2006, it offered some 8000 programs throughout the world to about 160 000 members – its success once again reflects the significance of globalization and the wealth of the current retirees. However, it was in 1988 that 24 institutes for learning in retirement joined with Elderhostel to form the Elderhostel Institute Network. At about the same time, locally, they adopted the name Lifelong Learning Institute. The network is a voluntary association of lifelong learning institutes that are funded by Elderhostel. Lifelong learning

institutes run a wide variety of teaching and learning programs and they are often sponsored by their local universities, so that they approach the type of provision made by the universities of the third age on the continent of Europe. It is significant that the cognitive-interest-motivation factor is dominant among its members, if the small-scale study conducted by Kim and Merriam (2004) is to be taken as representative. However, not all centers are sponsored in this way and some are quite independent: for instance, in a relatively large retirement community (15 000 population) it has been possible to run a college for 30 years catering for a wide variety of learning interests of the residents (Streib and Folts, 2003). In contrast, Elderhostel clearly caters for the many seniors who want to travel and learn local knowledge at the same time. It is also not insignificant that the United Nations should run a network of university departments throughout the world which is involved in educational tourism and which seeks to preserve local knowledge. Elderhostel has also spread to Canada where it is now known as Routes to Learning.

Another model that has emerged in Germany and Spain, among other places, is where universities open their classes to seniors. This certainly occurred in Germany in the 1980s and in Spain, the so-called third-age classrooms, as the Spanish government would not permit the use of the term university, began to function as early as 1978 (Socias *et al.*, 2004) and by 1993 it was decided to open universities to seniors with the University of the Balearic Islands initiating an Open University for Seniors. Nine years after the universities opened classes to seniors, 50 universities had begun to offer programs specifically for seniors and Socias *et al.* suggest that these programs will become more institutionalized during the twenty-first century. In Japan, there has been a similar movement but Shirasha (1995) comments on the fact that many of the academic staff working in this area are untrained.

However, the Chinese who have always had specialist universities – for example, University for Banking – have also had schools and universities for seniors – the first began in 1993 and the number grew very rapidly. (Liu Pengsheng, 1994; Li Herzhong, 1997) For instance, the TALIS conference in 2002 was held in the Wuhan University for the Aged People. It is significant that the growth of these universities reflects the fact that in the Chinese 7-year development plan of work on the aged, all cities and counties should run schools or universities for the ages (Li Zhi *et al.*, 1997).

With this tremendous growth in elders' learning, it is significant to ask whether a new sub-discipline of adult education is not emerging and this is certainly something that concerned adult-education scholars about the time that Knowles introduced the term andragogy to American adult education. In the terminological debate of that time, humanagogy, gerogogy, and educational gerontology were

among the terms being discussed. Ultimately, the idea of educational gerontology has been the one that has come to the fore since it rightly locates the study of the education of elders within the context of education *per se* and as a major element in lifelong learning, although on the continent of Europe – where the concept of andragogy has more to do with the academic study of the education of adults than the process of teaching and learning, as presented by Knowles (1980) – andragogical terminology is still utilized.

Accreditation

It is very clear that most of the above educational institutes do not offer accreditation for their courses and neither do many seniors seek it. In the UK, for instance, if they do want it, there is nothing to prevent seniors from enrolling in the Open University or in some of the university degree-course programs. This is also true for many countries where universities have an open access and a general acceptance of part-time higher education but the idea of the continental European U3As offering accreditation for some of their courses has been discussed at various times. This has not occurred everywhere in world; however, such as in many parts of the old Eastern Europe and the so-called Third World, but as the development of part-time education is almost certain to continue, the idea of older (but working) persons attending lifelong educational programs will become more common place: this might well be a precursor to the development of a more extensive senior's education movement in these countries. However, accreditation for vocational education is important and this has meant that in some countries, such as UK, government has been prepared to offer financial support for part-time vocational education but not for part-time leisure education and therefore not for the greater majority of senior's education. This reflects the emphasis placed upon the relationship between education and work in neo-liberal societies. Consequently, poorer older people are still disadvantaged in the world of education because their education is not seen to make a contribution to the neo-liberal economy.

In contrast to this approach, it is worth noting that there are other ways of measuring the value of learning: the Chinese have a scale of the value of learning (Leung *et al.*, 2006) based on Confucius' aphorisms which has five items: continuing learning, no boundaries to learning, keeping fresh in one's mind what has already been learned, people will decline if they do not continue to learn, and the harder one learns the better. There are elements here that reflect ideas in gerontology, such as the discontinuity hypothesis – that for so long as people continue to stay involved in society and continue to learn, then they will continue to develop; but, once they disconnect from the wider society the chances of decline

increase. While this Chinese approach might not be the answer to accreditation in older learning, it does suggest other ways of measuring its value.

This scale reflects in some ways the work of Cusack *et al.* (2003: 398) who have emphasized the idea of mental fitness in which they suggest that there are nine items: confidence in mental abilities, ability to set and achieve goals, willingness to take risks, optimism, creativity, mental flexibility, ability to learn new things, flexibility and ability to speak one's own mind. To enhance these in later life, is clearly a major benefit of learning.

Cusack *et al.* review three recent research programs; they note that they all have a common benefits to learning – a more positive attitude to life and increased hope. They make the point that:

Hope plays a powerful role in life. More than a sunny disposition and a belief that everything will be fine, it means believing that you have a normal life, without fear of “losing it”, and that inside you have the resources to accomplish personal goals and influence the course of your life. Hope means believing in a better future; hope means not giving in to anxiety and depression. (Cusack *et al.*, 2003: 395)

While this research as a whole is tangential to our major concern here, its findings on older learners are significant for educational gerontology and it is perhaps significant that in a survey of publications, of 4 years, of the journal *Educational Gerontology*, the number of publications on health education and there were nine papers that were specifically concerned about dementia and Alzheimer's disease. There were also a couple of papers that focused on physical fitness and one which looked at mental fitness – as the papers show, these two cannot be separated: one paper only was concerned with depression. Two papers were specifically about counseling.

Conclusion

To these studies, there is an indication that perhaps “lifelong learning may prevent or delay the symptoms of dementia, notably Alzheimer's Disease” (Cusack *et al.*, 2003: 395). In their own studies, they demonstrate that intensive courses of study improve the level of mental fitness – but it also contributes to the improvement of the more general level of health. From their own work, they regard continuing learning as health-promoting behavior. Similar research in the United Kingdom reaches similar conclusion (Schuller *et al.*, 2002).

However, we should not assume that all older adults are longing to continue their education or that there are not barriers to it: not only do these barriers include untrained teachers but they include untrained

bureaucrats who fail to assist in the developments of learning for older adults, lack of financial support for those who cannot afford to pay for their own education, lack of confidence in their own ability and so a fear of failure, and a lack of mobility and a disengagement with the wider world. Indeed, one of the major problems which advocates of older learning are confronted with is that many older people currently have negative images of their schooling and so they do not avail themselves to the benefits of nonschool learning, but as those people who have worked in the knowledge economy retire, the significance of older learning will continue to increase at every level of society.

Globalization and demographic factors have combined to make learning across the whole life span a reality and educational gerontology is not an established sub-discipline within education. The significance of its findings, such as those related to physical and mental health, have not been fully utilized by governments in education, health, or welfare programs although with the continuing aging population, it is clear that elders' learning will be seen as a major contributing factor in governmental policy. Seniors' learning has a major effect on the social capital of local towns and regions and finally, its personal benefits are beyond question.

See also: Health and Adult Learning; Life History; Participation in Adult Learning; The Health Advantages of Educational Attainment.

Bibliography

- Clennell, S. (1994). Distance learning: Creativity and activity for older students. *Third Age Learning International Studies* 1994, 39–46.
- Cusack, S., Thompson, W., and Rogers, M. (2003). Mental fitness for life: Assessing the impact of an 8-Week Mental Fitness Program on healthy aging. *Educational Gerontology* 29(5), 381–391.
- EC (European Commission) (2001). *Making a European Area of Lifelong Learning a Reality*. Brussels: European Commission COM(2001) 678final.
- Featherstone, M. and Wernick, A. (eds.) (1995). *Images of Aging*. London: Routledge.
- HMSO (HM Stationary Office) (2006). *Social Trends*. Basingstoke: Palgrave Macmillan.
- Huang, C.-S. (2006). The University of the Third Age in the UK: An interpretive critical study. *Educational Gerontology* 32(10), 825–842.
- Jarvis, P. (2001). *Learning in Later Life*. London: Kogan Page.
- Jarvis, P. (2006). *Towards a Comprehensive Theory of Human Learning*. London: Routledge.
- Jarvis, P. (2007). *Globalisation, Lifelong Learning and the Learning Society: Sociological Perspectives*. London: Routledge.
- Kerr, C., Dunlop, J., Harbison, F., and Myers, C. (1973). *Industrialism and Industrial Man*, 2nd edn. Harmondsworth: Penguin.
- Kim, A. and Merriam, S. (2004). Motivations for learning among older adults in a learning in Retirement Institute. *Educational Gerontology* 30(6), 441–455.
- Kliegel, M. and Altgassen, M. (2006). Individual differences in learning performance: The effects of age, intelligence, and strategic task approach. *Educational Gerontology* 32(2), 111–124.
- Knowles, M. (1980). *The Modern Practice of Adult Education: From Pedagogy to Andragogy* (rev. and updated). Chicago, IL: Association Press.
- Kolb, D. (1984). *Experiential Learning*. Englewood Cliffs, NJ: Prentice Hall.
- Leung, A., Chi, I., Chow, N., Chan, K. -S., and Chou, K. -L. (2006). Construction and validation of a Chinese value of learning scale. *Educational Gerontology* 32(10), 907–920.
- Li Herzhong (1997). The University for Old People in China. *Third Age Learning International Studies* 4; *Third Age Learning International Studies* 7, 91–95.
- Liu Pingsheng (1994). Creative Learning in the Chinese Schools for the Elderly. 4, 55–61.
- Li Zhi, Du Zicai, and Chan Yuan (1997). The Chinese characteristics of and prospects for China's education for the aged in Wuhan. *Third Age Learning International Studies* 7, 71–77.
- Lynott, P. and Merola, P. (2007). Improving the attitudes of 4th graders toward older people through a Multidimensional Intergenerational Program. *Educational Gerontology* 33(1), 63–74.
- McConatha, J., Hayta, V., Rieser-Danna, L., McConatha, D., and Polat, T. (2004). Turkish and US attitudes towards aging. *Educational Gerontology* 30(3), 169–183.
- McGuire, S., Klein, D., and Couper, D. (2005). Aging education: A national imperative. *Educational Gerontology* 31(6), 443–460.
- Ogunbameru, O. and Bamiwuye, S. (2004). Attitudes toward retirement and pre-retirement education among Nigerian Bank workers. *Educational Gerontology* 30(5), 391–402.
- Reisz, M. (2008). Senior service. *Times Higher Education*, Feb. 21–27, pp 38–41.
- Schuller, T., Brassett-Grundy, A., Green, A., Hammond, C., and Preston, J. (2002). *Learning Continuity and Change in Adult Life*. London: Institute of Education – Centre For Research on the Wider Benefits of Learning Report No 3.
- Sheey, G. (1995). *New Passages: Mapping Your Life Across Time*. Toronto, ON: Random House of Canada.
- Shirasha, I. (1995). Japan, a developing country in the field of lifelong learning as well as ageing: Its new perspectives and barriers. *Third Age Learning International Studies* 5, 101–104.
- Socias, C., Brage, L., and Garna, C. (2004). University programs for seniors in Spain: Analysis and perspectives. *Educational Gerontology* 30(4), 315–328.
- Stehr, N. (1994). *Knowledge Societies*. London: Sage.
- Streib, G. and Folts, E. (2003). A college in a retirement community. *Educational Gerontology* 29(10), 801–808.
- UNESCO (nd). *Creative Learning and Active Ageing*. Hamburg: UNESCO and NIACE.
- Van Gennap, A. (1960). *The Rites of Passage*. London: Routledge and Kegan Paul.
- Yeatts, D., Folts, E., and Knapp, J. (2000). Older workers' adaptation to changing work place: Employment issues for the 21st century. *Educational Gerontology* 26(6), 565–582.
- Yeaxlee, B. (1929). *Lifelong Education*. London: Cassell.

Further Reading

- Beisgen, B. and Kraitichman, M. (2003). *Senior Centers*. New York: Springer.
- Blakemore, S.-J. and Frith, U. (2005). *The Learning Brain*. Oxford: Blackwell.
- Educational Gerontology* (2004–2007) – Philadelphia: BrunnerRoutledge, 30–33 (1–10).
- Hudson, F. (1999). *The Adult Years*. San Francisco, CA: Jossey-Bass.
- Jarvis, P. (2009). *Learning to be a Person in Society*. London: Routledge.
- Jewell, A. (ed.) (1999). *Spirituality and Ageing*. London: Jessica Kingsley.
- Kitwood, T. (1997). *Dementia Reconsidered*. Buckingham: Open University Press.

Manheimer, R. (1999). *A Map to the End of Time*. New York: WW Norton.

Moody, H. (1998). *Aging: Concepts and Controversies*, 2nd edn. Thousand Oaks, CA: Pine Forge Press.

Peterson, D. (1983). *Facilitating Education for Older Learners*. San Francisco, CA: Jossey-Bass.

Peterson, D., Thornton, J., and Birren, J. (eds.) (1986). *Education and Aging*. Englewood Cliffs, NJ: Prentice Hall.

Schuller, T., Preston, J., Hammond, C., Brassett-Grundy, A., and Brynner, J. (2004). *The Benefits of Learning*. London: RoutledgeFalmer.

The Learning Revolution (2009). Department for Innovation, Universities and Skills Her Majesty's Government: The Stationary Office Cm 7555.

Waters, E. and Goodman, J. (1990). *Empowering Older Adults*. San Francisco, CA: Jossey-Bass.

Relevant Website

<http://www.elderhostel.org> – Elderhostel (downloaded 01.04.07).

University Adult Continuing Education: The Extra-Mural Tradition Revisited

M Slowe, Dublin City University, Dublin, Republic of Ireland

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The expansion of higher education across Organisation for Economic Co-Operation and Development (OECD) countries at the end of the twentieth century represented one of the most significant aspects of change in higher education in recent times. Taking 1995 as a baseline, student enrolments increased dramatically across the higher education sectors of 21 member states (OECD, 2005). The growth was largely driven by young school leavers where, on average, participation rates of the population aged 18–24 years increased by 70%, and young adults, aged 25–29 years, where levels of participation increased by almost 50% (OECD, 2005). However, in some countries, significant numbers of older students also entered higher education (OECD, 2005) – a pattern reflecting what some analysts considered to be the basis of a lifelong learning model (OECD, 1999).

While the trends in relation to participation in higher education in the developing countries outside the OECD vary, they also reflect expansion in general. The scale is enormous – a study of middle-income countries involved in the United Nations Educational, Scientific and Cultural Organization (UNESCO) World Education Indicators (WEI) program showed that far more students entered and graduated from these countries than in the 30 OECD member states combined (UNESCO, 2007). (Countries participating in the UNESCO WEI program are Argentina, Brazil, China, Egypt, India, Indonesia, Jamaica, Jordan, Malaysia, Paraguay, Peru, Philippines, Russian Federation, Sri Lanka, Thailand, Tunisia, Uruguay, and Zimbabwe.)

In 2006, for example, China had more tertiary graduates (2.4 million) than the top three OECD countries combined – United States (1.4 million), Japan (0.6 million), and France (0.3 million). Overall, however, the proportion of young people gaining a higher educational qualification in the WEI countries (19.7%) was about half that of the OECD group.

In exploring the topic of university adult continuing education (ACE) and looking behind these headline participation figures, three factors need to be taken into consideration. First, one of the key comparative indicators of the levels of participation is the age participation rate (APR). This measures the proportion of the population of the typical school-leaving age which progress to higher education. By definition, this refers to young people, and only rarely are such measures used to explore levels of participation by

adults in higher education. Second, much of the expansion in higher education over the last two decades took place in non-university institutions such as polytechnics, community colleges, further education colleges, and the like. The statistics, therefore, generally refer not only to universities, but also, more generally, to participation across all types of higher education institutions. Third, international statistics usually refer to full-time undergraduate entrants, whereas mature students are more likely to be found on part-time, distance, post-experience, and non-credit programs.

The term university ACE is open to different interpretations reflected, since the 1980s or so, in the shifting policy discourse at national and international levels (Duke, 1992; Bourgeois *et al.*, 1999; Davies, 1995; Osborne and Thomas, 2003). A benchmark international study of adult participation in higher education undertaken by the OECD 20 years ago distinguished between four categories of the adult student in higher education (OECD, 1987):

1. students who enter or re-enter higher education as adults in order to pursue mainstream studies leading to a full first degree or diploma (delayers, deferrers, or second chancers, i.e., those who are admitted on credentials gained by means of work experience or second-chance educational routes);
2. adults who reenter to update their professional knowledge, or seek to acquire additional qualifications, in order to change occupation or advance in their career (refreshers and recyclers);
3. those without previous experience in higher education who enroll for professional purposes, especially in courses of short duration; and
4. adults with, or without, previous experience in higher education, who enroll for courses with the explicit purpose of personal fulfillment (personal developers).

In a follow-up study nearly 15 years later, another categorization, based on a combination of three criteria associated with (a) entry routes, (b) modes of study, and (c) nature of program attended, was developed (Wolter, 2000).

This resulted in a six-fold classification.

1. those who take a vocational route to the achievement of the high school leaving examination (the second educational route);
2. special admission examinations (the third educational route);

3. students with high school leaving qualifications who undertake vocational training or work experience before moving on to higher education (the double qualification route);
4. students studying on a part-time or distance-mode basis;
5. graduates who return to higher education for continuing professional development courses (often of short duration); and
6. older adults taking continuing education courses largely for personal development purposes.

Building upon this work, and developing it further for analytic purposes, five different conceptions of university ACE can be identified in the international literature at the end of the first decade of the twenty-first century:

- The first conceptualization focuses on the age of the learner – namely, mature students, defined in most countries as those aged somewhere between 21 and 25 years above when entering higher education) as opposed to young, recently matriculated high school graduates.
- The second focuses on the mode of study – predominantly part-time or distance, as opposed to full-time or on campus.
- The third focuses on the type of program undertaken – for example, professional updating or retraining and noncredit or community courses, as opposed to undergraduate degree qualifications.
- The fourth focuses on the life course stage, or predominant motivation of the learner – for example, second chance or postexperience, as opposed to direct progression from initial school education.
- The fifth conceptualization focuses on the mode of organization of provision for adult learners – through specialist institutions (such as open universities) or centers with a dedicated mission to meet the needs of adults (of which the extramural tradition, considered below, is the classic example) as opposed to widening access for adults to mainstream provision.

Global Trends in Lifelong Learning and University Adult Education

The widely debated concept of lifelong learning (Coffield, 2000; Bagnall, 2000) figured prominently in national and international educational policy discourse for three decades, with some more limited resonances in higher education (Field, 2006; Scott, 1995, 1997; Tavistock Institute, 2002; Watson and Taylor, 1998; Rubenson, 2006). In Europe, in a research exercise conducted as part of the major higher education reform agenda across all the

member states of the European Union – termed the Bologna process – 54% of respondents indicated that their university had a lifelong learning strategy; a further 25% said that one was in preparation, while 19% indicated that they did not have one; and 2% did not reply (Davies and Feutrie, 2008). Within this shifting discourse, one focus of attention relates to access by adults to higher education – and the statistics do point to generally increasing levels of participation by those aged over 21 years in higher education. This trend, for example, was highlighted in a joint report from the American Council of Education and the European Universities Association (Green *et al.*, 2002). In the United States, on a head-count basis (which includes part-time students), approximately 40% of all undergraduates were over 25 years of age, while, in Canada, 31% were over 24 years of age.

The reasons behind this growth in the numbers of adults in higher education in some countries are complex – ranging from demographic trends through to the emphasis on the development of the knowledge society (Schuetze and Slowey, 2003). In many developed countries, demographic projections of a substantial decline in the numbers of young people coming through from schools fuelled concerns about skill shortages and recruitment problems for higher education. Human capital theories had also become highly influential among policymakers and politicians, sparking a renewed emphasis on the continuous need for updating and retraining of the workforce and focusing attention on the extent to which university systems, structures, and staff were, in fact, prepared to meet the projected demand from employers and the professions. Additionally, for institutions of higher education, continuing professional/vocational education was being viewed as a potential source of new revenue – increasingly important as a result of funding pressures from the neo-libero public policy agenda.

Over and above all of these factors, and interacting with them in complex ways, is the issue of equality of opportunity and access to higher education. In most countries, the social class composition of higher education remains unrepresentative of the population at large; women students have grown as a proportion, but are concentrated in particular disciplinary areas; and the limited information available on participation by those from ethnic minority groups shows variable patterns in different countries.

Also important are internal changes of a qualitative nature, such as more flexible provision, new approaches to teaching and learning, and an increasing emphasis upon quality assurance and the evaluation and certification of learning. The forms of knowledge creation and dissemination, the greater access to information sources, the use of new media and new channels of communication, the development of complex partnerships with employers,

and the growing marketization of education are all factors that have a major impact on where, what, how, and why tertiary/post-secondary students learn.

In fact, many of the features associated with the policy emphasis on lifelong learning in higher education can be traced back to a wide-ranging reform agenda of Western higher education systems. For the better part of half a century, this agenda has included an emphasis on the student experience, especially at the undergraduate level; enhancing teaching quality; developing partnerships with employers and other external agencies; strategic planning at an institutional level; and expanding interdisciplinary teaching and research.

International Patterns of Adult Participation in Higher Education

While higher education systems in developed societies are subject to common economic, social, and demographic pressures, resulting in increasing levels of participation by adults, evidence from an international comparative study of ten countries suggests that the differentials between countries appear remarkably persistent over time.

Schuetze and Slowey (2000) developed a threefold typology along a continuum relating to the levels of adult participation in higher education.

- First, countries with relatively high levels of participation by adult learners and demonstrating a relatively high degree of flexibility in relation to entry criteria and study patterns; this category included Sweden and the United States.
- Second, countries with significant, but lower, proportions of adult learners across the system as a whole, and in which adult students were frequently located in open universities or dedicated centers of adult or continuing education within mainstream institutions; this category included Australia, Canada, New Zealand, and the UK.
- Third, countries at the other end of the continuum with very low levels of adult participation in higher education; this category included Austria, Germany, Ireland, and Japan.

Analyses of the same ten countries over a 10-year period indicated that, while overall levels of adult participation in higher education had increased, the relative levels of participation between countries remained largely the same.

This analysis suggests that, while trends toward convergence can be discerned as a result of globalization, the forms taken by contemporary university ACE continue to be largely determined by the higher education traditions from which they spring.

Historical Patterns of University ACE

In order to understand current patterns of university ACE, it is necessary to have a broad appreciation of the historical traditions from which they emerge.

In relation to university ACE, three major traditions can be identified.

1. First, the extramural model of bringing university education to the adult population. The origins of this model can be formally traced back to the 1870s in the ancient British universities, and led to the development of dedicated adult education departments. These departments were unusual in being multidisciplinary, and devoted to the delivery of university programs which were specifically tailored to meet the needs of adult students. This extramural approach provided a conceptual model which, for complex historical reasons, has had a widespread international influence, in particular in Commonwealth countries.
2. Second, traditions of university ACE which are connected to the ideal of service by the university to local and regional social and economic communities. A significant example of this tradition arises from the land-grant universities in the USA, which included a particular mission to support community and rural development from the start.
3. Third, traditions where educational provision to adults, as a distinctive group, was not seen as a significant function for universities. This is particularly evident where the university traditions emphasized research, and ACE allocated to other agencies and specialist institutions, such as open universities or vocational organizations (e.g., Germany).

The complexity and variability of patterns is demonstrated in the findings of a comparative study of continuing education in universities across 30 countries. This showed that:

...activity with ostensibly different purposes, including continuing professional development, second chance education, education for leisure and social development, U3A and technology transfer, are all within the remit of continuing education. Increasingly, continuing education within universities has become blurred with other aspects of flexibility including part-time education, summer universities, open and distance education, accreditation of prior learning and work-based learning. (Osborne and Thomas, 2003: 20)

In terms of theoretical writing on university adult education, and the practice in many countries throughout the world (particularly Commonwealth members), the extramural tradition was, for much of the twentieth century, a particularly influential model. In order to better

understand contemporary variations, therefore, it is useful to have a familiarity with the origins of the values and objectives underpinning this approach to university adult education.

Background to the Emergence of the Extramural Model of University ACE

The movement for university extension, such as the *folk-universitet* in some Scandinavian countries, formed part of the widespread growth of interest in university ACE in the nineteenth century. The extramural model which emerged from the ancient universities of Oxford and Cambridge in England is of particular interest as it was influential both in terms of time scale (several centuries) and geography (throughout the Commonwealth and in many English-speaking countries).

Some of the earliest university lectures for the general adult population can be traced to those given by Francis Hutcheson, Professor of moral philosophy at the University of Glasgow in the 1720s (Hamilton and Slowey, 2005). It was James Stuart – another academic who also had spent time at Glasgow University – who is credited with articulating the idea of university extramural (beyond the walls) provision in Cambridge University (Fieldhouse, 1996). In 1873, the University of Cambridge agreed to organize, formally, programs and lectures in various centers, and Oxford also undertook to provide a number of programs through extension centers. Shortly afterwards, the University of London established the London Society for the Extension of University Teaching (Jones, 2008). The work grew in fits and starts, but quite rapidly overall; thus, by 1902, Oxford and Cambridge had established well over 900 centers with more than 20 000 adult student learners attending.

However, from the outset, the activists were concerned to attract a broader cross section of the population and, in particular, students from working-class backgrounds – an objective which remained strong throughout much of the twentieth century. The challenges included low levels of literacy, relatively high costs, and, perhaps most notable of all, the cultural gulf between establishment, upper-class culture, as embodied in universities, and working-class culture.

What would be termed today quality assurance was also a persistent problem. How could the university ensure that its extension programs were of a standard appropriate for university-level provision? In an attempt to facilitate and sustain more focused study, both Oxford and Cambridge introduced residential summer meetings for adult learners. Despite all the problems, university extension was a significant force in the later years of nineteenth century in Britain and, consequently, also shaped aspects of the higher education systems of many countries.

Another aspect of this tradition related to the strengthening of links between universities and broader social movements – for example, the Workers' Educational Association in the UK, which reflected a broadly liberal ideological stance, exemplified the influential early leaders of university ACE such as Temple, Tawney, Lindsay, and Cole. Nottingham University was the first formally to establish a department of extramural studies, in 1920–21. Within a decade of World War II, most of the major British universities had developed such departments.

In 1947, the extramural departments came together to form the Universities Council for Adult Education (UCAE). This continues as a significant policy body in university ACE through to the present time (2009), albeit with changes in title, to the University Association for ACE (UACE) and, recently, to the Universities Association for Lifelong Learning (UALL). Many other countries developed equivalent bodies, notably the University Continuing Education Association (UCEA) in the United States, which grew out of the National UCEA – one of the oldest college and university associations in the USA, founded in 1915.

From the beginning, UCAE asserted that liberal studies, although the core concern of extramural departments, should be complemented by vocational, professional, and technical course provisions to address the changing needs of the postwar society. The resulting pattern of provision was a combination of shorter liberal studies programs (sometimes indistinct from WEA provision), attracting predominately middle-class students, and professionally orientated provision, much of it for the growing numbers of employees in the welfare state.

Internationally, in many university systems based around the Anglo-American model, large centers or departments developed over this period. Such departments often held unique positions in the university: on the one hand, they were part of the mainstream university structure, with representation on Senates, professorial appointments, and the like; on the other hand, however, they were also, in many important respects, separate and different, and more connected to their local communities.

Structural Tensions – Inside or Outside the Walls?

Structural tension – being partly inside, and partly outside, mainstream university structures and processes – is a distinguishing feature of the centers, institutes, and departments which have grown out of the extramural tradition.

To begin with, they are often separately funded, and, importantly, have the opportunity to generate their own direct fee income; their students are adults and part time, not standard age and full time; their teaching is

often off-campus and at nonstandard times; and they are multidisciplinary rather than having a single-subject focus as conventional mainstream academic departments do.

At a rather different level, it is also interesting to note that not only did they include, in their number, leading public intellectuals – such as E.P. Thompson, Richard Hoggart, and Raymond Williams – but also gave birth to, or were the key catalysts for, the development of new interdisciplinary areas of research and teaching, later to become major parts of mainstream university provision – for example, cultural studies, women's studies, industrial studies, and applied social studies.

Typical examples of the forms of provision of extra-mural departments and centers for adult education over the last three decades or so include:

- liberal studies of the traditional kind, characterized by intellectual effort on the part of the student, and the attainment of a standard appropriate for university study;
- continuing education programs designed to support a change of career or life course;
- the development of part-time courses for credit leading to university awards;
- role education for groups whose common element was their role in society – often including work with the public sector and nongovernmental organizations (NGOs);
- industrial education at all levels from management to the shop floor;
- social purpose, community education, and outreach;
- project research work;
- training for those engaged in the education of adults – increasingly at the postgraduate level such as masters and doctoral programs;
- research in adult education as an academic discipline; and
- acting as a seedbed for innovative programs – frequently regarded, for one reason or another, as high risk by the parent university.

In relation to the last point, there has been a growing recognition in many countries that universities should undertake more innovative and pioneering work to open up new fields of study and attract new learners where the provision, once established, might be handed over to other agencies. This, for example, includes access provision and third-arm activities such as work with local communities, employers, or other groups.

With the rapid expansion of higher education over recent decades, the overall shape changed as many countries developed some form of binary structure. In the UK, this distinction was abolished in the early 1990s. One consequence was that financial support for adult university education was mainstreamed across the entire system, so that part-time students, and the courses on

which they were enrolled, had to be accredited as equivalent to at least the first year of undergraduate study.

This created considerable turbulence in the system, with many adult learners withdrawing. On the one hand, it had the potential to bring adult learners into the mainstream of the university system; for the first time, adults could study for and, if successful, gain university awards through university ACE. In one sense, this could be seen as something of a coming of age for ACE. On the other hand, and to quite an extent, this eroded the autonomy and flexibility of university ACE, and was regarded by some as undermining the student-led, democratic culture of the ACE tradition. It also reflected an international trend, also seen, for example, in the shifting focus of the folk universities of the Nordic countries which, while originally not aimed at qualifications, "...are now increasingly involved in preparing students for professional life or upgrading previous qualifications" (Fagerlind and Stromqvist, 2004: 258).

Global Trends in University ACE – Convergence or Divergence?

As has been shown here, the forms which university ACE takes vary considerably over time, are integrally connected with the traditions from which they spring, and are subject to the broader global changes impacting national university systems. The case of the extramural tradition considered above illustrates common challenges faced. In essence, these boil down to the issue of the mainstream versus the dedicated provision for adult learners. How this balance tips is influenced by broad social and economic factors, mediated through university strategies at particular times.

In reviewing the current situation of university ACE, five broad international trends can be identified.

First, much of the recent growth in higher education has taken place in institutions other than universities. In some countries, adult students are predominantly located in these institutions which, typically, have a particular mission to work in collaboration with local and regional communities, and to strengthen links with employers. While these trends can be seen more generally (typically characterized as moving from knowing what to knowing how (Barnett, 1992, 2005)), their impact on newer and nonuniversity institutions tends to be strongest. On the plus side, it is very often the elements of flexibility and relevance that are of greatest interest to many adults. On the other hand, however, the growing emphasis on international university league tables – such as the Shanghai and the Times Higher Education rankings – serves to reinforce differentiation and hierarchical positioning of research-intensive, elite universities compared to the broader spectrum of higher education institutions.

In practice, in many countries, adult students tend to find it easier to gain access to nonuniversity institutions, and to be less well represented in the highest-status universities. Whether this matters or not – as long as learning needs are met and successful outcomes achieved – is the subject of ongoing debate.

Second, the use of new learning technologies and new modes of delivery are widespread in the broader field of continuing education for adults. However, their impact on mainstream university provision remains variable. Therefore, the utilization of new learning technologies does not (yet) appear to have resulted in an anticipated blurring of boundaries between full-time and part-time students in most countries. At the undergraduate level, the resourcing and organizational model of most university systems remains orientated toward the traditional, full-time student. However, perhaps this might change in the future as the patterns of study of young students alter, as they increasingly combine study with significant employment commitments, and as higher education comes to be viewed as an activity to be pursued over the life course (Slowey and Watson, 2003). At the professional masters' level, the pattern may be somewhat different. Some countries – for example, many member states of the European Union, through the Bologna process – have seen the growth of modular programs at the postgraduate level, which are designed to connect with the world of work and continuing professional development.

Third, core values associated with the extramural tradition of seeking to widen access to university learning to the wider public do indeed appear somewhat beleaguered along with being somewhat marginalized in relation to the broader field of research on higher education. Aspects of university ACE are also increasingly connected with the growing commercialization of higher education as evidenced, for example, through the inclusion of higher education under the World Trade Organization General Agreement on Trade and Services (GATS). As some universities in the richer countries seek to develop provision in the continuing education market as a global economic initiative, this undoubtedly carries potential implications for capacity and sustainability of universities in developing countries (e.g., UNESCO, 2005, 2008). Furthermore, within and between nation states, the arena of lifelong learning is one which is particularly subject to the growth of the private sector and both profit and nonprofit higher education providers (Peters, 2001; Marginson, 2007).

Fourth, in contrast to the structural and ideological issues which arise in relation to universities engaging in direct provision of adult education, activities such as the training of adult education professionals, doctoral education, and research in the field of adult education are all areas that sit relatively comfortably within the university environment – frequently as a special section or unit within faculties of education (Hinzen and Przybylska,

2004). Given the continuing international policy emphasis on lifelong learning, and interdisciplinary research, these are all areas which have the potential to form an important part of the future development of university ACE, and specialist research journals with a focus on adult and lifelong learning represent a flourishing area of publication (Jarvis, 2006).

Fifth, in organizational terms, long-standing structural tensions between university ACE and parent institutions continue to be observed in the twenty-first century – albeit in rather different forms. It is possible to detect on-going debates about the role of units, centers, or departments which operate at the interface between the higher education institution and external stakeholders such as local communities, employers, trade unions, and the like. This inside/outside – sometimes marginal – positioning continues to offer independence and creative space (Taylor *et al.*, 2002). The price to be paid for this independence, however, may involve exposure to institutional policies which can vary over time.

Whether or not contemporary developments will see a new flowering of university ACE and lifelong learning and will usher in a new era in which universities engage dynamically and imaginatively with the whole community remains to be seen. Without doubt, the environment is a volatile and unstable one, and there is much to strive for in the context of university ACE and the broader area of educational and training policies relating to adults.

Bibliography

- Bagnall, R. G. (2000). Lifelong education and the limitations of economic determinism. *International Journal of Lifelong Education* **19**, 20–35.
- Barnett, R. (1992). *The Idea of Higher Education*. Buckingham: SRHE/Open University Press.
- Barnett, R. (2005). *Reshaping the University: New Relationships between Research, Scholarship and Teaching*. Maidenhead/New York: SRHE/Open University Press.
- Bourgeois, E., Duke, C., Guyot, J. L., and Merrill, B. (1999). *The Adult University*. Buckingham: SRHE/Open University Press.
- Coffield, F. (ed.) (2000). *Differing Visions of the Learning Society*. Bristol: Policy Press.
- Davies, P. (ed.) (1995). *Adults in Higher Education: International Perspectives in Access and Participation*. London: Jessica Kingsley.
- Davies, P. and Feutrie, M. (2008). University lifelong learning to lifelong learning universities. *Bologna Handbook: No 8, June 2008*. Berlin: Raabe and the European Universities Association.
- Duke, C. (1992). *The Learning University*. Buckingham: SRHE/Open University Press.
- Fagerlind, I. and Stromqvist, G. (eds.) (2004). *Reforming Higher Education in the Nordic Countries*. Paris: International Institute for Educational Planning.
- Field, J. (2006). *Lifelong Learning and the New Educational Order*, 2nd edition, Stock on Trent: Trentham Books.
- Fieldhouse, R. (1996). *A History of Modern British Adult Education*. Leicester: National Institute of Adult Continuing Education.
- Green, M., Eckel, P., and Barblan, A. (2002). *The Brave New (and Smaller) World of Higher Education: A Transatlantic View*. Washington, DC: The American Council on Education and the European Universities Association.

- Hamilton, R. and Slowey, M. (2005). *The Story of DACE: The Department of Adult Continuing Education of Glasgow University*. Glasgow: DACE University of Glasgow.
- Hinzen, H. and Przybylska, E. (eds.) (2004). *Training of Adult Educators in Institutions of Higher Education: A Focus on Central, Eastern and South Eastern Europe*. Bonn: Institute for International Cooperation of the German Adult Education Association.
- Jarvis, P. (ed.) (2006). *From Adult Education to the Learning Society: 21 Years from the International Journal of Lifelong Education (Education Heritage Series)*. London: Routledge.
- Jones, C. K. (2008). *The People's University: 150 years of the University of London and Its External Students*. London: University of London.
- Marginson, S. (2007). Going Global: trends in higher education and research in the APEC region. *APEC HRDWG Symposium on Education Policy Challenges*, Brisbane. http://www.cshe.unimelb.edu.au/people/staff_pages/Marginson/Marginson.html (accessed May 2009).
- OECD (1987). *Adults in Higher Education*. Paris: Centre for Educational Research and Innovation/OECD.
- OECD (1999). *Education Policy Analysis*. Paris: OECD.
- OECD (2005). *OECD Thematic Review of Tertiary Education: Comparative Indicators on Tertiary Education* <http://www.oecd.org/dataoecd/56/63/35940816> (accessed May 2009).
- Osborne, M. and Thomas, E. (eds.) (2003). *Lifelong Learning in a Changing Continent*. Leicester: National Institute of Adult Continuing Education.
- Peters, M. (2001). *Neoliberalism, Postmodernity and the Reform of Education in New Zealand*. Auckland: University of Auckland Mcmillan Brown Lecture Series.
- Rubenson, K. (2006). The Nordic model of lifelong learning. *Compare – A Journal of Comparative Education* **36**, 327–341.
- Schuetze, H. G. and Slowey, M. (eds.) (2000). *Higher Education and Lifelong Learners: International Perspectives on Change*. London: RoutledgeFalmer.
- Schuetze, H. G. and Slowey, M. (2003). Participation and Exclusion – a comparative analysis of non-traditional and lifelong learners in systems of mass higher education. *Higher Education* **44**, 309–327.
- Scott, P. (1995). *The Meanings of Mass Higher Education*. Buckingham: Society for Research into Higher Education/Open University Press.
- Scott, P. (1997). The Postmodern University? In Smith, A. and Webster, F. (eds.) *The Postmodern University? Contested Visions of Higher Education in Society*, pp 36–47. Buckingham: SRHE/Open University Press.
- Slowey, M. and Watson, D. (eds.) (2003). *Higher Education and the Lifecourse*. Maidenhead: SRHE/ Open University Press.
- Tavistock Institute (2002). *Review of Current Pedagogic Research and Practice in the Fields of Post-Compulsory Education and Lifelong Learning*. Exeter: Economic and Social Research Council Teaching and Learning Research Programme.
- Taylor, R., Barr, J., and Steele, T. (2002). *For a Radical Higher Education: After Postmodernism*. Buckingham: SRHE/Open University Press.
- UNESCO (2005). *Implications of WTO/GATS on Higher Education in Asia and the Pacific, UNESCO Forum Occasional Paper Series Paper No.8*. Paris: UNESCO.
- UNESCO (2007). *Education Counts: Benchmarking Progress in 19 WEI Countries*. Paris: UNESCO Institute for Statistics.
- UNESCO (2008). *UNESCO Forum on Higher Education Research and Knowledge*. www.unesco.org/education/researchforum.
- Watson, D. and Taylor, R. (1998). *Lifelong Learning and the University: A Post-Dearing Agenda*. London: Falmer.
- Wolter, A. (2000). Germany: Non-traditional students in German Higher Education: Situation, profiles, policies and perspectives. In Schuetze, H. G. and Slowey, M. (eds.) *Higher Education and Lifelong Learners*, pp 48–66. New York: Routledge Falmer.
- Cropley, A. J. and Knapper, C. K. (1983). Higher education and the promotion of lifelong learning. *Studies in Higher Education* **8**(1), 15–21.
- Duke, C. (2004). Values, discourse and politics: An Australian Comparative Perspective. In Tapper, T. and Palfreyman, D. (eds.) *Understanding Mass Higher Education: Comparative Perspectives on Access*. London: RoutledgeFalmer.
- El-Khawas, E. (1999). The “new” competition: Serving the learning society in an electronic age. *Higher Education Management* **11**(2), 7–17.
- European Commission (1995). *Teaching and Learning: Towards the Learning Society*. Luxembourg: Office for Official Publications of the European Union.
- Fuller, A. (2007). Mid-life ‘transitions’ to higher education: Developing a multi-level explanation of increasing participation. *Special Issue: Studies in the Education of Adults* **39**(2), 217–235.
- Goldman, L. (1995). *Dons and Workers: Oxford Adult Education since 1850*. Oxford: Oxford University Press.
- Gorard, S. and Rees, G. (2002). *Creating a Learning Society? Learning Careers and Their Relevance for Policies of Lifelong Learning*. Bristol: Policy Press.
- Halsey, A. H. (1992). An International Comparison of Access to Higher Education. In Phillips, D. (ed.) *Lessons of Cross-National Comparison in Education*, pp 11–36. Wallingford: Triangle Books.
- Horn, L. J. and Carroll, C. D. (1996). *Nontraditional Undergraduates: Trends in Enrolment from 1986 to 1992 and Persistence and Attainment Among 1989–90 Beginning Postsecondary Students (Report No. NCES 97-578)*. Washington, DC: NCES.
- International Journal of Lifelong Learning (2000). *Special issue: Lifelong Learning* **19**(1).
- McLean, S. (2008). Extending resources, fostering progress, or meeting needs? University extension and continuing education in Western Canada. *British Journal of Sociology* **29**(1), 91–103.
- OECD (1996). *Lifelong Learning for All*. Paris: OECD.
- OECD (1997). *Literacy Skills for the Knowledge Society*. Paris: OECD.
- OECD (1998). *Redefining Tertiary Education*. Paris: OECD.
- OECD (2005). *Promoting Adult Learning*. Paris: OECD.
- Ordorika, I. (2006). *Commitment to Society: Contemporary Challenges for Public Research Universities*. Key Note Paper. Paris: UNESCO Global Forum of Higher Education.
- Osborne, M., Gallacher, J., and Crossan, B. (eds.) (2007). *Researching Widening Access to Lifelong Learning: Issues and Approaches in International Research*. London: Routledge.
- Rubenson, K. and Schuetze, H. G. (2000). Lifelong learning for the knowledge society: Demand, supply and policy dilemmas. In Rubenson, K. and Schuetze, H. G. (eds.) *Transition to the Knowledge Society: Policies and Strategies for Individual Participation and Learning*, pp 355–376. Vancouver: UBC (Institute for European Studies).
- Schuetze, H. G. and Istance, D. (1987). *Recurrent Education Revisited – Modes of Participation and Financing*. Stockholm: Almquist and Wicksel International.
- Schuller, T. and Burns, A. (1999). Using ‘social capital’ to compare performance in continuing education. In Coffield, F. (ed.) *Why's the Beer Always Stronger Up North? Studies in Lifelong Learning in Europe*, pp 53–61. Bristol: Economic and Social Research Council/Policy Press.
- Tapper, T. Palfreyman, D. (eds.) (2004). *Understanding Mass Higher Education: Comparative Perspectives on Access*. London: RoutledgeFalmer.
- UNESCO (1992). *Learning to Be: The World of Education Today and Tomorrow: The 'Faure' Report*. Paris: UNESCO.
- UNESCO (1996). *Learning: The Treasure Within: The "Delors" Report*. Paris: UNESCO.
- Wagner, A. (1999). Lifelong Learning in the University: A New Imperative. In Hirsch, W. Z. and Weber, L. E. (eds.) *Challenges Facing Higher Education at the Millennium*, pp 134–152. New York: American Council on Education/Oryx Press.
- Walters, S. (2005). South Africa's Learning Cape Aspirations: The idea of a learning region and the use of indicators in a middle income country. In Duke, C., Osbourne, M., and Wilson, B. (eds.) *Rebalancing the Social and Economic. Learning, Partnership and Place*. Leicester: National Institute of Adult Continuing Education.

Further Reading

- Bron, A. and Agell, K. (2000). Non-traditional students in higher education in Sweden: From recurrent education to lifelong learning. In Schuetze, H. G. and Slowey, M. (eds.) *Higher Education and Lifelong Learners: International Perspectives on Change*, pp 83–100. New York: Routledge Falmer.

Relevant Websites

<http://www.acheinc.org> – ACHE (Association for Continuing Higher Education).
<http://www.bc.edu> – Boston College, CIHE (Center for International Higher Education).
<http://cauce-aepuc.ca> – Canadian Association for University Continuing Education (CAUCE).
<http://www.esrea.org> – ESREA (European Society for Research on the Education of Adults).
http://ec.europa.eu/education/index_en.htm – European Commission Education & training.
<http://www.EUCEN.org> – European Universities Continuing Education Network.
<http://www.icae.org> – ICAE (International Council for Adult Education).

<http://www.oise.utoronto.ca> – OISE University of Toronto, Canadian Association for the Study of Adult Education.
<http://www.oecd.org> – Organisation for Economic Co-operation and Development.
<http://www.srhe.ac.uk> – SRHE (Society for Research into Higher Education).
<http://www.scutrea.ac.uk> – Standing Conference on University Teaching and Research in the Education of Adults.
<http://www.adulterc.org> – The Adult Education Research Conference (AERC).
<http://www.unesco.org> – UNESCO.org, IAU Scholarly Publications.
<http://portal.unesco.org> – UNESCO.org: Scientific and Cultural Organization, United Nations Educational.
<http://www.uall.ac.uk> – Universities Association for Lifelong Learning (UALL).
<http://www.ucea.edu> – University Continuing Education Association.

ADULT EDUCATION – ECONOMY AND SOCIETY

Contents

Adult Education and Civil Society

Adult Education and Nation-Building

Economic Outcomes of Adult Education and Training

Health and Adult Learning

Modernization Processes and the Changing Function of Adult Learning

The Political Economy of Adult Education

Wider Benefits of Adult Education

Adult Education and Civil Society

J Crowther and I Martin, University of Edinburgh, Edinburgh, UK

© 2010 Elsevier Ltd. All rights reserved.

Civil Society: The Problem of Definition

It is important to understand that the idea of civil society has been defined and deployed in different ways at different times, depending on what kind of social and political struggles were being fought, and won or lost. The meaning of civil society is part of the particular arguments and struggles of particular times. Consequently, for example, contemporary interpretations in adult education range from (literally) conservative formulations such as Robert Putnam's influential account of civil society as the source of social capital to Gramsci's radical notion of civil society as a site of struggle for moral and ideological hegemony.

Walzer (1996: 7) identifies civil society as the "sphere of uncoerced human association and also the set of relational networks . . . that fill this space." Essentially, civil society, in its most widely accepted current usage, is distinct from the state (formal/representative politics) and the market (economic production, consumption, and exchange). It is generally understood to refer to those aspects of informal social and political life in which citizens come together in voluntary groups, associations, and movements to pursue their own collective interests and projects in freely chosen and relatively autonomous ways. Modern civil society organizations would therefore include, for example, trades unions and employer associations, churches and voluntary agencies, nongovernmental organizations (NGOs) and not-for-profit social enterprises, cultural and ethnic groups, and a whole range of popular campaigns, struggles, and movements. It is widely recognized that the vitality and relative autonomy of such informal and non-institutionalized political activity outside the state is an important precondition for the health of democratic politics within the state.

In reality, civil society is frequently in danger of invasion, colonization, and control by the state or the market. For instance, third way policies often seek to incorporate citizens and civic associations into particular kinds of partnership with state and market interests. Similarly, in their concern to roll back the frontiers of the state, neo-liberal programs of privatization and structural adjustment may turn civil society into little more than an extension of the market or a manufactured instrument of state policy (Hodgson, 2004).

Finally, civil society is an ideal as well as an idea. Essentially, it may be said not only to embody a vision of democratic associational life, but also to represent the site of the kind of prefigurative work that is required to show that another world is possible. There is an important sense, then, in which civil society should be understood as an "intellectual space, where people in a myriad of different groups and associations can freely debate and discuss how to build the kind of world in which they want to live . . . a realm of emancipations, of alternative imaginations of economic and social relations and of ideological contest" (Howell and Pearce, 2002: 2–8).

Adult Education and Civil Society Today

There are a number of distinct, but related, reasons for the remarkable revival of interest in the idea of civil society in recent years. These include: the failure and collapse of the totalitarian communist state and its command economy in Eastern Europe; fiscal crisis in social democratic welfare states and the effects of neoliberal retrenchment on public service provision; growing disillusionment with the politics of representative democracy and its characteristic

democratic deficit; evidence of the increasingly important role of NGOs and other third sector agencies in both service delivery and governance; and the growth of new social movements and their demands for radical political and economic change. Edwards (2004: 13–15) distinguishes between the developmental roles of civil society in economic, social, and political terms: it compensates for market underdevelopment or failure; generates social capital and informal welfare; and is the site for the development of grassroots people power. In the context of what he calls peripheral capitalism, Youngman (2000: 209/210) sees the functions of civil society in wholly positive terms as promoting democratization and constructing alternative models of development – in both respects giving power back to ordinary people. At a more theoretical level, the work of Antonio Gramsci has been of singular importance in presenting civil society as the key site of struggle for the radical left; in addition, in terms of democratic thinking, the ideas of Jurgen Habermas have been particularly fruitful and influential. The impact of civil society on real political change is evidenced in the recent history of varied examples such as South Africa, Latin America, Poland, and Scotland.

As Welton (2001) emphasizes, there is nothing necessarily progressive, or even civil, about the idea of civil society. On the other hand, it does tend to be deployed in its current usage to focus attention on struggles for greater democracy and freedom and against injustice and oppression. Essentially, it has been important in stimulating new thinking. For instance, on the post-Marxist left, the focus of interest on new social movements suggests that the traditional working class can no longer be regarded as the privileged agent of the emancipatory project (Mouffe, 1992). Indeed, some writers who take an interest in civil society adopt a self-consciously radical position. For instance, Powell and Geoghegan (2004: 7) argue that what they call the new face of civil society denotes “a force for democratisation in liberal democratic societies, where the poor and oppressed have found a voice through associational activities”, thus making common cause in the struggle against exploitation and oppression.

There is an important sense in which civil society can be conceived as the distinctive space in which citizens come together to make democracy work (Cohen and Arato, 1992). On the one hand, this learning process may be ideologically constructed in different ways for different purposes. It is quite clear that much adult education today, particularly in the guise of lifelong learning, is primarily an instrument of state economic policy. The kind of technical rationality which drives it has nothing to do with democratic deliberation or social purpose. On the other hand, it is also important to recognize that learning in social movements has always been part of a particular kind of adult education. Indeed, some writers would argue that civil society is the natural territory of

radical adult education. Welton (1995, 2001, 2005) suggests that learning, particularly the kind consistent with the emancipatory interest which lies at the heart of radical adult education, develops directly out of the communicative interaction that characterizes the groupings, associations, and movements of civil society.

Three Perspectives on Adult Education and Civil Society

The Communitarian Perspective: Learning for Membership

The key characteristic of the communitarian perspective is the focus on a decline in moral standards in civil society caused by a variety of interconnected social changes. These include permissiveness in the 1960s, the break-up of traditional family structures, a decline in social trust, the growth of a dependency culture nurtured by welfare provision, increases in violent crime, a lack of responsibility, particularly among the young and the unemployed, and the proliferation of minority rights (Etzioni, 1995). These trends were exacerbated in the 1980s with the thrust toward the unbridled pursuit of self-interest associated with neo-liberal politics and policies. Communitarianism aims to shore up the moral foundations of society by remoralizing those groups which cause social problems. Educating people to be more responsible and active members of civil society is posed as the remedy for this situation.

The conservatism of communitarianism might have had very little appeal to many adult educators. However, its message that human relationships needed repairing and rebuilding chimed well with the emergence of social capital as a key policy concern, which has been influential in social policy and adult education. The dominant tradition of social capital refers to the norms of reciprocity and trust that appear to have been undermined by the breakdown in human relationships. Putnam's (2000) metaphor of Americans bowling alone (rather than in clubs) captured the sense of the loss of trust, cooperation, and shared activities that bind people together. Whereas communitarianism located the problem in a moral deficit in individuals, social capital relocates the deficit in the structure of people's relationships. Putnam's research (see also Putnam, 1993; Coleman, 1994) spurred the growth of a veritable social capital industry among academics. Distinctions have been made between bridging social capital (making ties between different groups), bonding social capital (developing contacts between like minded people), and linking social capital (connecting different levels of power or social status) (Kay, 2005).

The importance of trust, reciprocity, and participation in civic affairs has informed programs of community development which seek to create, enhance, and consolidate social capital in communities. In the UK, the Centre

for the Wider Benefits of Learning has documented various positive social, political, and health outcomes related to the impact of adult education on social capital (Schuller *et al.*, 2004). Participation in post-compulsory education – including adult literacy – is identified as one way of developing trust and building civic responsibility (Field, 2005; Tett *et al.*, 2006). Furthermore, social capital can increase the likelihood of enhanced economic activity as confidence levels increase and networks of connections ease the path into work (Falk and Kilpatrick, 2000). However, too much bonding social capital among groups with negative dispositions toward education has also been put forward as an explanation for non-participation in adult education (McGivney, 2001). The bonds that tie people can, in some cases, bind them. Social capital may also contribute toward civic involvement through volunteering and people becoming more active members of their community (Schuller *et al.*, 2004).

Arguably, social capital is an idea which is both politically expedient and highly normative. It is expedient in the sense that focusing on social relationships allows inequalities in social structure and resources to be ignored. Etzioni and Putnam take little interest in poverty and inequality as causal factors in the decline of communities and their social capital. The separation of civil society from the economy therefore leads away from politically controversial issues such as the distribution of wealth and power. It is normative in that only some types of social capital are valued, whereas others are not. For example, communities such as New Age Travellers may exhibit high levels of social capital, but are not the type of communities that governments want to support.

The focus on civil society, rather than the role of the state in reviving civil society, can let governments off the hook by transferring responsibility from statutory provision to voluntary effort (Ehrenberg, 2002). Ironically, the state's interest in encouraging participation in civic activity seems to be increasing at the same time as democratic spaces for learning seem to be diminishing. The distinction between the invited spaces of policy and the demanded spaces of communities are two distinct ways of thinking about democratic participation (Gaventa, 2006). In the former, participation is structured around the interests of top-down policy imperatives, whereas, in the latter, the spaces for participation emerge because of popular demands from below. Both present opportunities for adult education engagement, however, the context, purpose, and focus of the work provide very different challenges and prospects.

The Habermasian Perspective: Learning for Deliberation

From an adult education perspective, it could be argued that Jurgen Habermas is concerned about rescuing

democracy, and the learning process embedded in it, from both the instrumentalism of capitalism (in which it becomes simply a political means to an economic end) and the relativism of postmodernism (in which universal values and purposes are deemed no longer to matter). For Habermas, who “steadfastly refuses to ditch modernity's dream of using human reason to create a more humane world” (Brookfield, 2005: 25 and 26), the central task of critical theory is to encourage people to think constructively and creatively for themselves and to enable them to follow the agreed rules of democratic discourse. These are intended to ensure that “no one may be excluded; anything may be said, questioned, or challenged; and no force may be used” (Chambers, 1996: 197).

Habermas emphasizes the democratic work that is done in civil society when he describes it as a “network of associations that institutionalizes problem-solving discourses on questions of general interest inside the framework of organised public spheres” (Habermas, 1996: 367). According to John Keane, an eminent contemporary theorist of civil society, the key distinction for Habermas is between the “logics of the political and economic systems, regulated respectively by administrative power and money, and the life-world of self-organized public spheres based on solidarity and communication” (Keane, 1988: 18). In this sense, the state and the economy are complementary domains based on distinctive systems of power, and civil society is defined, partly, in contradistinction to them. The boundaries are often shifting and there may be a high degree of overlap and interpenetration; however, it is, nevertheless, important to maintain the distinction.

Democratic deliberation is essentially about communication, and how we use language to communicate with each other. This kind of discourse, including the rules of argumentation on which it is based, is something we have to learn. The appeal to reason in this process removes the distortions of power which would otherwise saturate and corrupt it. Democracy is constituted and sustained through a process of rational deliberation, which is conducted according to agreed rules and procedures, among equal citizens. In other words, adult learning is absolutely central to the Habermasian idea of building a democratic culture and defending the lifeworld, where people are authentically themselves, against invasion of and colonization by the steering mechanisms of the systems world of state or market. Habermas is concerned about stipulating the discursive conditions and procedures for conducting the democratic argument in the public sphere, understood as the space for deliberative democracy which civil society must keep open if it is to protect democracy. This is necessarily a continuing process, always unfinished.

There are many examples of adult education approaches which reflect this concern for democratic deliberation and organized public spheres. The Workers Educational Association in the UK is, at its best, a bearer of this tradition of

social purpose and political engagement (Fieldhouse, 1996). Perhaps the obvious example of this tradition is the study circle. The term itself suggests an egalitarian relationship between members in which participation and democratic discussion focused on issues of common interest or concern are fundamental. This form of organized democratic learning is strongly associated with the Scandinavian and Nordic countries. It also has important parallels in the USA, and resonates with the principles of mutuality which informed learning in the co-operative and labor movements in the nineteenth century (Bjerkaker and Summers, 2006). Moreover, as Welton (2005) points out, the social learning that occurs in progressive movements constitutes a communicative public sphere in which new meanings can be created and debated. These movements assert the importance of the lifeworld and seek to defend it against the corrosive power of instrumental rationality associated with the political and economic systems.

What the Habermasian perspective suggests is that the task of adult education is to create the pretexts and the contexts for this kind of learning to take place within the uncoerced associations and affiliations of civil society. In order to think through what this means, Habermas (1978) distinguishes between three domains of learning, each informed by particular interests: technical learning, which concerns the manipulation of environment and control of the world we live in; practical learning, which concerns the development of interpersonal understanding and relationships; and emancipatory learning, which concerns self-understanding and critical consciousness. Each has its part to play in learning for democracy; however, clearly, Habermas's main interest lies in the emancipatory project.

It is important to emphasize that good adult learning models the process of deliberative or discursive democracy. In this sense, adult education can be said to do the prefigurative work of democracy in which key ideas and arguments are tested out and interrogated in discussion. The purpose of this kind of learning is not necessarily to reach agreement: in the end, it may well be about agreeing to disagree, and learning to live democratically with the consequences.

The Gramscian Perspective: Learning for Activism

Radical adult education from the 1970s onwards drew inspiration from the Italian Marxist Antonio Gramsci because of his work on cultural politics and the role of intellectuals in social change. While Gramsci was never precise in his definition of civil society, its relationship with the state in his work is unique. He makes the distinction between two aspects of the superstructure of society (in contrast to the economic base) in terms of civil society and the state/political society. These correspond to the

exercise of two forms of power which reinforce class domination: hegemonic power (the leading and directive ideas and values in society) and the state's monopoly on legitimate coercive power. Civil society is made up of so-called private organizations like churches, trades unions, and voluntary bodies which are characterized by social relations based on autonomy and free association, whereas the state is primarily defined in relation to its coercive potential exercised through the activities of the army, judiciary, and courts. Gramsci was aware, of course, that the state was not merely coercive. More importantly, the boundaries between civil society and state are permeable, and organizations and practices can embody social relations belonging to both spheres. The connections between state and civil society in reproducing class rule are reflected in Gramsci's (1971) expanded view of the state as "political society + civil society, in other words, hegemony protected by the armour of coercion" (p. 263).

However, civil society's apparent distance from the state means it can be a powerful medium for the diffusion of the dominant hegemony. Education, for example, is provided in many countries by the state and influenced by its policies and economic priorities; however, Gramsci locates education firmly in civil society, which seems strange. The school is an important institution of social reproduction. In addition, schooling may be coercive, although adult education still has a voluntary status. Nevertheless, the core of educational relationships is primarily open, rather than closed; critical education can serve to engage the dominant hegemony, rather than simply reproduce it. This ambivalence (produced by social relations that are partly a product of the state and partly of civil society) provides opportunities and spaces for resistance against dominant ideas, values, and priorities.

The voluntary nature of adult education means that there is potentially more space for radical education because of its freer status. It has a crucial role to play in weaning the working class away from their dependency on traditional intellectuals, by creating their own organic intellectuals, who are able to articulate their interests and galvanize class action as a necessary first step toward social transformation. In this perspective, the focus is on winning hearts and minds before a frontal assault on the state – or even to remove the necessity for it. Education is therefore the foundation of revolutionary activity (Holst, 2002).

The radical nature of this type of education means that it is seldom located in the formal adult education provision. In the UK, there is the important historical example of the Labour College movement which focused on educating revolutionary militants (Simon, 1992). In some respects, this development prefigured Gramsci's analysis of the role of education in class struggle. Holst (2004) has made visible the largely unacknowledged educational role of revolutionary political organizations in the USA, and Boughton (2005) has documented the role of the

Communist Party in educating generations of militants in Australia. Allman (2001) addresses the centrality of alliance building under proletarian hegemony involving different social movements and popular forces as a key issue in critical and progressive educational work. The struggle against apartheid in South Africa is an example of the importance of alliance building across civil society in which radical educators played a significant role (von Kotze, 2005).

However, the emergence of new social movements from the 1960s onwards challenged the focus on social class as the privileged agent of change. The growth of neo-Marxist accounts of civil society and the development of postmodernist thinking also influenced various strands of radical adult education. Localized narratives of change (Usher *et al.*, 1997) and selective change around particular issues such as the environment (Welton, 1995), rather than generalized class struggle, have either used or dispensed with Gramsci to explain the dynamics of power in civil society and what can be done.

Toward a Global Civil Society? Roles for Adult Education

The discussion so far has considered three perspectives on adult education and civil society in the context of the nation state. It is now becoming increasingly important, however, to think in trans-national and supra-national terms. Each perspective suggests a distinctive way of thinking about the possible contributions of adult education in the context of a global civil society.

It might seem premature to talk of a global civil society, but it is important to recognize the growth of non-state actors such as trans-national social movements with a global interest (Mundy and Murphy, 2006). The spread of neo-liberal globalization, global systems of production, geo-political wars, nuclear armament, environmental degradation, and the role of powerful unelected international organizations in economic affairs have led to an increasing number of civil society groups, NGOs, and social movements contesting and challenging these trends. This globalization from below has been aided by the development of information and communication technologies which have made the coordination of action and dissemination of counter-information possible on a global scale. While the influence of these groups on policy development is debatable, they do have a highly visible international presence.

In the communitarian perspective, the focus is on developing membership among nation states at an international level to promote adult education and to develop reciprocal relations between different countries. There have been various adult education initiatives which might broadly be conceived as reflecting this perspective. The International Council of Adult Education (2007), for example, is a trans-national advocacy organization which is campaigning for

the right to learn and wider recognition of the role adult education can play in combating poverty, discrimination, and exclusion. In addition, United Nations Educational, Scientific, and Cultural Organization (UNESCO)'s International Conference on Adult Education (CONFINTEA) conferences target policy makers, academics, adult educators, civil society organizations, trade unions, and other interested parties to facilitate linking social capital, among other things. These events – held every 12 years – seek to promote adult education on a global scale and to embed it in various development goals and resourcing priorities of nation states and international organizations.

In the Habermasian perspective, learning in the public sphere is another arena where adult educators can make a contribution to the development of a global civil society. The World Social Forum, for example, actively contributes to what Barr (2007) calls “undiscovered public knowledge.” By this, she means independent sources of knowledge rooted in people's experiences and directed at the collective social, political, and human problems they seek to address. Newman (2007) argues that a deliberative democratic process is essential for making national politicians accountable. This process demands a sustained dialog between citizens and electors, and requires particular skills: “how to critically appraise the statements of others; how to think clearly for ourselves; how to think inventively; how to participate actively in the affairs of state; and how to participate wisely” (p. 10).

In the Gramscian perspective, globalization raises issues about the growth of a trans-national capitalist class, its hegemony, and how it can be opposed (Robinson, 2005). Gramsci's own work focused on civil society in the context of nation states; therefore, we need to extrapolate and develop his concepts for the new global context. This raises questions about the type of strategy and the type of social forces necessary to create an alternative hegemony. For example, the confrontational action against the World Trade Organisation in Seattle in 1999 and against the Group of Eight (G8) in Genoa in 2001 is, in Gramsci's terms, a “war of movement” (direct frontal assaults on the state) in contrast to his view about the necessity for a “war of position” (a systematic struggle for ideological hegemony). The latter points to the need for detailed and systematic critique of the intellectual and moral authority of international institutions, such as the International Monetary Fund, World Bank, World Trade Organisation, G8/9, and the Organisation for Economic Cooperation and Development, which legitimate the hegemony of an increasingly global capitalist class.

Conclusion

The three perspectives on adult education and civil society, constructed in terms of learning for membership, deliberation, and activism, demonstrate competing meanings and

distinctive practices. These are reflected in different adult education purposes, contexts, and constituencies. In the era of globalization, the struggle to define and control what civil society means and who it is for is increasingly likely to occur at an international level. Adult education has a distinctive contribution to make to the social and political contestation this will entail.

See also: Adult Learning, Instruction and Programme Planning; Insights from Freire; Class Analysis in Adult Education; Community Based Adult Education; Popular Adult Education; Wider Benefits of Adult Education.

Bibliography

- Allman, P. (2001). *Critical Education against Global Capitalism*. London: Bergin and Garvey.
- Barr, J. (2007). Educational research and undiscovered public knowledge. *Studies in the Education of Adults* **39**(1), 22–37.
- Bjerkaker, S. and Summers, J. (2006). *Learning Democratically: Using Study Circles*. Leicester: National Institute of Adult Continuing Education.
- Boughton, B. (2005). 'The workers university: Australia's Marx schools. In Crowther, J., Galloway, V., and Martin, I. (eds.) *Popular Education: Engaging the Academy – International Perspectives*, pp 100–109. Leicester: National Institute of Adult Continuing Education.
- Brookfield, S. (2005). *The Power of Critical Theory: Liberating Adult Learning and Teaching*. San Francisco, CA: Jossey-Bass.
- Chambers, C. (1996). *Reasonable Democracy: Jurgen Habermas and the Politics of Discourse*. Ithaca: Cornell University Press.
- Cohen, J. and Arato, A. (1992). *Civil Society and Political Theory*. Cambridge, MA: MIT Press.
- Coleman, J. S. (1994). *Foundations of Social Theory*. Cambridge, MA: Belknap Press.
- Edwards, M. (2004). *Civil Society*. Cambridge, UK: Polity Press.
- Ehrenberg, J. (2002). Equality, democracy, and community from Tocqueville to Putnam. In McLean, S., Schultz, D., and Steger, M. (eds.) *Social Capital*, pp 50–73. London and New York: New York University Press.
- Etzioni, A. (1995). *The Spirit of Community*. London: Fontana Press.
- Falk, I. and Kilpatrick, S. (2000). What is social capital? A study of a rural community. *Sociologia Ruralis* **1**(40), 87–110.
- Field, J. (2005). *Social Capital and Lifelong Learning*. Bristol: Policy Press.
- Fieldhouse, R. and Associates (1996). *A History of Modern British Adult Education*. Leicester: National Institute of Adult Continuing Education.
- Gaventa, J. (2006). Finding the spaces for change: A power analysis, exploring power for change. *IDS Bulletin* **37**(6), 23–33.
- Gramsci, A. (1971). *Selections from the Prison Notebooks* (translated by Hoare, Q. and Smith, G. N.). London: Lawrence and Wishart.
- Habermas, J. (1978). *Knowledge and Human Interest*. London: Heinemann Educational.
- Habermas, J. (1996). *Between Fact and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge, MA: MIT Press.
- Hodgson, L. (2004). Manufactured civil society: Counting the cost. *Critical Social Policy* **79**(24), 139–164.
- Holst, J. D. (2002). *Social Movements, Civil Society and Radical Adult Education*. London: Bergin and Garvey.
- Holst, J. D. (2004). Globalization and education within two revolutionary organisations in the United States of America: A Gramscian analysis. *Adult Education Quarterly* **55**(1), 23–40.
- Howell, J. and Pearce, J. (2002). *Civil Society and Development: A Critical Exploration*. Boulder, CO: Lynne Rienner.
- International Council for Adult Education (2007). Adults' right to learn: Convergence, solidarity and action. *Adult Education and Development* **67**, 65–68.
- Kay, A. (2005). Social capital, the social economy and community development. *Community Development Journal* **41**(2), 160–173.
- Keane, J. (ed.) (1988). *Civil Society and the State: New European Perspectives*. London: Verso.
- McGivney, V. (2001). *Fixing or Changing the Pattern?* Leicester: National Institute of Adult Continuing Education.
- Mouffe, M. (ed.) (1992). *Dimensions of Radical Democracy: Pluralism, Citizenship, Democracy*. London: Verso.
- Mundy, K. and Murphy, L. (2006). Transnational advocacy, global civil society? Emerging evidence from the field of education. In Lauder, H., Brown, P., Dillabough, J.-A., and Halsey, A.H (eds.) *Education, Globalization and Social Change*, pp 991–1015. Oxford: Oxford University Press.
- Newman, M. (2007). Adult education and the home front. *Adults Learning* **18**(7), 9–11.
- Powell, F. and Geoghegan, M. (2004). *Reclaiming Civil Society or Reinventing Governance?* Dublin: A and A Farmar.
- Putnam, R. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press.
- Putnam, R. (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon Schuster.
- Robinson, W. I. (2005). Gramsci and globalisation: From nation-state to transnational hegemony. *Critical Review of International Social and Political Philosophy* **8**(4), 1–16.
- Schuller, T., Brasset-Grundy, A., Green, A., Hammond, C., and Preston, J. (2004). *Wider Benefits of Learning*. London: Routledge Falmer.
- Simon, B. (1992). *The Search for Enlightenment*. Leicester: National Institute of Adult Continuing Education.
- Tett, L., Hall, S., MacLachlan, K., et al. (2006). *Evaluation of the Scottish Adult Literacy and Numeracy [ALN] Strategy*. Edinburgh: Scottish Executive Social Research.
- Usher, R., Bryant, I., and Johnson, R. (1997). *Adult Education and the Postmodern Challenge*. London: Routledge.
- von Kotze, A. (2005). People's education and the academy: An experience from South Africa. In Crowther, J., Galloway, V., and Martin, I. (eds.) *Popular Education: Engaging the Academy – International Perspectives*, pp 11–21. Leicester: National Institute of Adult Continuing Education.
- Walzer, M. (1996). The civil society argument. In Mouffe, M. (ed.) *Dimensions of Radical Democracy: Pluralism, Citizenship, Democracy*. London: Verso.
- Welton, M. (1995). *In Defense of the Lifeworld: Critical Perspectives on Adult Learning*. Albany, NY: State University of New York Press.
- Welton, M. (2001). Civil society and the public sphere: Habermas's recent learning theory. *Studies in the Education of Adults* **33**(1), 20–34.
- Welton, M. (2005). *Designing the Just Learning Society: A Critical Inquiry*. Leicester: National Institute of Adult Continuing Education.
- Youngman, F. (2000). *The Political Economy of Adult Education and Development*. Leicester: National Institute of Adult Continuing Education/Zed Books.

Further Reading

- Baron, S., Field, J., and Schuller, T. (eds.) (2000). *Social Capital: Critical Perspectives*. Oxford: Oxford University Press.
- Bauman, Z. (1999). *In Search of Politics*. Cambridge: Polity Press.
- Crowther, J., Martin, I., and Shaw, M. (eds.) (1999). *Popular Education and Social Movements in Scotland Today*. Leicester: National Institute of Adult Continuing Education.
- Delanty, G. (2003). *Community*. London: Routledge.
- Finger, M. and Asun, J. (2001). *Adult Education at the Crossroads: Learning Our Way Out*. Leicester: National Institute of Adult Continuing Education.
- Held, D. (1996). *Models of Democracy*. Cambridge: Polity Press.

- Hobson, B. (ed.) (2003). *Recognition Struggles and Social Movements*. Cambridge: Cambridge University Press.
- Kane, L. (2001). *Popular Education and Social Change in Latin America*. London: Latin America Bureau.
- Little, A. (2002). *The Politics of Community*. Edinburgh: Edinburgh University Press.
- Welton, M. (2005). Civil society. In English, M. (ed.) *International Encyclopedia of Adult Education*, pp 100–106. Basingstoke, UK: Palgrave Macmillan.
- Wildemeersch, D., Finger, M., and Jansen, T. (eds.) (1998). *Adult Education and Social Responsibility*. Frankfurt am Main: Peter Lang.
- Wildemeersch, D., Stroobants, V., and Bron, M. (eds.) (2005). *Active Citizenship and Multiple Identities in Europe: A Learning Outlook*. Frankfurt am Main: Peter Lang.
- Williamson, B. (1998). *Lifeworlds and Learning*. Leicester: National Institute of Adult Continuing Education.

Adult Education and Nation-Building

S Walters, University of Western Cape, Bellville, South Africa

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The late President of Tanzania said in 1976 that, “Adult education is a highly political activity. Politicians are sometimes more aware of this fact than educators, and therefore they do not always welcome real adult education” (Nyerere, 1976). By juxtaposing adult education and nation building, the political dimensions of adult education are foregrounded.

This article privileges experiences particularly in the Southern African Development Community (SADC) concerning adult education and nation building, to act as a lens to illuminate key issues which are also relevant to other regions of the world. The discussion is framed utilizing the following three social purposes of adult education:

1. Education that enhances strategies which enable women and men to survive the harsh conditions in which they live. Examples of this include literacy, primary healthcare, and some home-craft skills.
2. Education and training geared to developing skills for people in the formal and informal sectors that describe education for economic purposes.
3. Cultural and political education which aims to encourage women and men to participate actively in society through networks of cultural organizations, social movements, political parties, and trade unions.

The article begins with a discussion of various understandings of nation building, then moves to a discussion of adult education and nation building within SADC, highlighting universally important themes. This is followed by concluding remarks and suggestions for further readings.

Nation Building

“Nation building depends upon winning popular support for rapid change,” commented Lowe (1971). In the early 1970s, 20 years after the United Nations had been established, and many countries had recently gained independence from the colonial powers, there was much interest in the role of education in national development. Developing the nation-state was the primary focus of attention. Over 30 years later, a fundamental political question is whether existing plural states, which make up 90% of the current 180 or so nation-states in the world, would be able

to withstand the dual onslaught of ethnic nationalism and global economic integration (Phadnis and Ganguly, 2001).

The nation-state is a legal concept describing social groups who occupy a defined territory and are organized under common political institutions and an effective government. The state exercises sovereign powers within its borders and is recognized as sovereign by other states in the international system. A nation is therefore different from an ethnic group which can be transnational. (Phadnis and Ganguly 2001: 20) A key social construct within the nation-state is the citizen; therefore literature on adult education and citizenship (e.g., Bron and Schemmann, 2001; Korsgaard *et al.* 2001), can be a useful resource.

Mamdani (1996) provides a very rich analysis of the complexity of how the citizen and subject have been constructed in postcolonial states in Africa, which highlights the bifurcated state between urban and rural, between modern and customary, which is highly gendered, within one hegemonic state apparatus. Oga and Okwori (2005) illustrate the complexities of citizenship in Nigeria where people have stronger affiliation to ethnicity or religion than an imagined community of a nation-state. How citizenship is understood is therefore essential to the discussion on nation building.

Gaventa (2007), as editor of a series of texts on Claiming Citizenship: Rights, Participation and Accountability, points out that there is a growing crisis of legitimacy in the relationship between citizens and the institutions that affect their lives. In countries, both in the North and South, citizens speak of mounting disillusionment with governments, based on concerns about corruption, lack of responsiveness to the needs of the poor, and the absence of a sense of connection with elected representatives and bureaucrats. The rights and responsibilities of corporations and other global actors are being challenged, as global inequalities persist and deepen. Nation building is therefore deeply connected to notions of democracy, development, and globalization (Korsgaard, 1997; Finger, 2005). It is undoubtedly complex and contested.

The question arises then as to how far people's struggles for power, through liberation or social movements, can also be seen as part of a nation-building project Or is it a process that can only be led by a legitimate government in power? This question harks back to understandings of development and nation building (Youngman 2000). In this article, the position taken is that there are a range of actors in a society which contribute to nation building and, as Lowe suggests, winning popular support

for rapid change. There can be nation building from below, through social movements, and nation building through government interventions. As Finger (2005) points out, formal education for children, is often more tied to the state machinery than adult education which has most often stayed outside of government.

We look at what adult education and nation building means in southern Africa and in doing so elaborate on some of the above points, particularly relating to struggles for democracy and development.

Adult Education and Nation Building in the SADC

Southern African Development Community (SADC)

This section is adapted from an article entitled 'Adult education in lifelong learning in southern Africa', by Walters and Watters (2001).

Political and Socioeconomic Picture of the Region

The present global economy is pushing national economies and local industries to compete in the world market. The SADC was formed in 1992 and in 2000 the protocol on trade came into force which is moving SADC toward a free-trade area. SADC grew out of the Southern African Development Co-ordination Conference (SADCC), which had had South Africa as its common political enemy while sharing with it complex historical and economic dependence. The SADCC provided the structure for the countries to organize themselves in geopolitical terms in order to maximize their political clout and minimize their economic dependence on South Africa.

SADC is comprised of 14 countries, which vary in population from the Democratic Republic of the Congo (DRC), which has 62 million, to Swaziland which has just over a million inhabitants. Another four of the countries have populations of 2 million or less: Botswana, Lesotho, Mauritius, and Namibia. Many SADC countries still rely heavily on agriculture. Only Angola, Botswana, and South Africa have less than 10% of their production coming from agriculture. Five countries, Malawi, Mozambique, DRC, Tanzania, and Zambia, obtain more than 20% of their production from agriculture. It is one of the least urbanized parts of the world and only Angola, Botswana, and South Africa have urban dwellers in the majority. Economic performance is dominated by that of South Africa, which represents more than 70% of the combined sub-regional gross domestic product (GDP).

Commonly used indicators of poverty reveal that SADC members are among the poorest countries in the world. In 2005, the average human-development index for

the region's countries was 0.54, while the SADC website reports a 2003 GDP per capita of US\$1062. This compares unfavorably with the world average in 2003 of US\$5822 and the average for Western Europe, for example, of US\$30 449. The per capita income per person in the DRC in 2006 was only US\$128 per year, which makes it one of the poorest in the world.

There are a few countries which have a relatively high GDP per capita for developing countries. The 2003 SADC figures put Mauritius at the top with US\$4522 with South Africa and Botswana having above US\$3000 per capita per annum. But even in these relatively well-off countries, there is major inequity between the rich and poor. The SADC Regional Human Development Report of 2000 states that 30% of SADC population live in abject poverty while 30–40% of the labor force is unemployed or ekes out a living as subsistence farmers.

The life expectancy is very low in most SADC countries, for example, Swaziland where it is just 31 years and Botswana where it is 35 years (2004 figures). The terrible human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) pandemic is pushing the life expectancy even lower. The infant-mortality rate is as high as 260 in Angola and 205 in DRC per 1000 live births. Again, this compares unfavorably with, for example, Japan at 3.2 or the United Kingdom at 5.3.

Although still experiencing difficult economic situations, most SADC countries have adopted macroeconomic and social policies aimed at improving regional human-development performance. In order to achieve this, several governments spend around 20% of their budgets on education and almost 5% on health development.

This sketch provides a picture of a region that has a very wide spread of developmental needs, including adult education. The countries of SADC are peripheral capitalist economies and the development of adult education has been shaped very directly by this, including the macro-policies of international development agencies and the socioeconomic realities within each of the countries. (See, e.g., Torres, 2004) These in turn have sometimes spurred local people on to finding alternative approaches to development.

Social Purposes of Adult Education in the SADC Region

Most of the countries of the region have experienced major political and economic upheavals in the last 50 years. During this time, all of them went through, more or less, traumatic processes of decolonization. The last five countries to gain independence were Mozambique in 1976, Angola in 1975, Zimbabwe in 1980, Namibia in 1990, and South Africa in 1994. All five of these countries experienced extended liberation struggles. All of the countries in the region went through processes of reconstruction and

development toward building new nations in about the last 50 years. The approaches adopted by the different countries were shaped strongly by dominant development theories of the time which reflect particular ideologies and material interests (see, e.g., Youngman, 2000).

In order to give a sense of the fragmented and contested nature of the role of adult education in nation building, illustrations are provided, falling within the three major social purposes: survival; economic development; and political and cultural development.

Adult education for political and cultural development

As adult education is integral to social processes, it is not surprising that it gains in prominence at heightened political or economic moments through actions within the state, civil society, or the private sector. In the past, most governments have invested minimally in adult education with their emphasis on schooling. At the point of political independence, adult education became significant. For example, after the historic transition in South Africa to democratic governance in 1994, adult education was highlighted particularly in relation to redress for black people, economic development, and the growth of a democratic culture. Adult basic education was declared a presidential lead project but was to be dependent on international donors.

Another example is taken from Mozambique. The watchwords of Samora Machel, the late president of the newly independent Mozambique and head of the liberation movement, Frelimo, were “study, produce, and fight.” These were taken as serious marching orders in preparation for independence and a broad, popular literacy movement emerged throughout the country (Marshall, 1990: 83). Popular education involved mobilizing human resources for creating the new Mozambique and literacy was part of the dynamizing groups set up in most villages and towns. A national directorate of literacy and adult education was set up in 1976. Since those heady days with the systematic destructive power of the apartheid state in South Africa, state-sponsored adult education all but disappeared in Mozambique between the mid-1980s to mid-1990s. Since then, according to Mario and Nandja (2006: 192), there has been a process of rediscovery and rescue of adult literacy and education, with there now being an average rate of illiteracy of about 54%, with rates in rural areas and among women being disproportionately high.

The fact that most countries in the region have undergone radical political change in the last 50 years implies that there have been high degrees of political activism and education at different times. From the 1960s, Namibians and South Africans built powerful democratic movements of citizens, both inside and outside the countries, which created alliances of socialist revolutionaries, social democrats, and social reformers. Creative informal and

nonformal education was integral to these movements which involved thousands of community-based organizations which were forged into the democratic movements. (Walters 1989) There was rich learning through social and political action as people strove to build a new nation from below.

The struggles for people’s democracy in the mid-1970s in places like Angola and Mozambique had been overshadowed in the following 15–20 years by civil wars spurred on by the apartheid regime and other international interests. Many of the gains hoped for in education, health, social welfare, and economic development had been decimated. In Angola alone it is reported that over 500 000 people had been killed since 1989 and 3 million people became refugees (Oduaran 2000). The DRC too has recently emerged from a devastating civil war. In the region, while people’s movements for democratic social change have engendered innovative responses in some countries, in others, many millions of people are confronted with the struggle for survival from poverty, war, and disease under the most trying conditions. Not all people’s movements are necessarily positive; some are mobilizing children and adults for participation in violent crime and war. Literature is not available on what educational efforts are occurring within these situations but there are no doubt activists working to bring about peace and reconciliation on the one hand and activists training new recruits for violence and war on the other (see, e.g., Thompson, 2000).

Adult education for survival

One of the greatest educational challenges facing SADC presently is the devastating HIV/AIDS pandemic. The region has the highest incidence in the world. Thousands of people are dying of the disease. It is having substantial effects on the economies as it is the working adult population which is most vulnerable. The estimation is that there were 12 million AIDS orphans in Africa at the end of 2007 and estimates are that currently the higher education population in South Africa is 22% HIV/AIDS positive.

There have been various strategies developed over the last 15 or so years to counter the pandemic. It is primarily a sexually transmitted disease which is exacerbated by poverty. Educational processes are called on which challenge deep seated cultural, religious, ethnic, gender, or class attitudes and behaviors. In many societies, there are cultural practices that propagate the spread of the virus through promiscuity. Women are most at risk as often it is men who have multiple partners. The power relations between women and men make it nearly impossible for many women to insist on safe sexual practices. Some people predict that until women are empowered and gender relations are more equal, it will be extremely difficult to stem this tide.

Educational programs are being orchestrated in some countries through the health ministries, but this is seen to be inadequate. In South Africa, departments of education, labor, welfare, and health are working together. There are over 600 nongovernmental organizations (NGOs) working to counter HIV/AIDS. Some workplaces have begun running education and counseling services for workers. There is a growing awareness that all sectors of society, working with people of all ages, must join together to educate about HIV/AIDS. At the World Aids Conference held in Durban in July 2000, 13 000 scientists, activists, educators, development workers, government officials, and health workers, all came together to share research, information, methodologies, and policies. There were discussions, debates, information, and papers disseminated on a daily basis through community and national radio, television, and newspapers. It was a massive and impressive public educational process.

The growing campaign is being interpreted and taken forward by a very wide range of interest groups with different values, for example, rurally based indigenous healers, rural and urban women's groups, youth groups, religious and community organizations, and educational institutions. They are using different approaches, from ethnocultural, to feminist, to popular, to spiritual, among others, to organize awareness-raising, skills-training, and organizational-development strategies.

The responses to HIV/AIDS provide excellent contemporary examples of adult education for survival which involve most sectors of society and which draw on multiple pedagogical, organizational, and developmental frameworks simultaneously.

Adult education for economic development

In a region with such high unemployment and levels of poverty, economic development is of paramount importance. The United Nations Department of Economic and Social Affairs states that unemployment in the region varies from Namibia at 34% (2000) to Madagascar at 4.5% (2003). An independent assessment puts the rate in Zimbabwe in 2007 at 80%. A 1998 SADC report estimates that 30–40% of the labor force of SADC are either completely unemployed or are eking out a living as subsistence farmers. The same report indicates that less than 50% of the labor force is women. In the last 20 years, numerous adult-education programs focused on skill development have been embarked on by SADC countries in both the formal and informal economies.

Within the context of globalized economies, economic development and adult education, or adult learning, have become more urgent and complex. Within the debates on globalization are debates about the importance and the role of information communication technologies (ICTs) in economic development. Africa is the most poorly

served continent in ICT. Africa, with 14.2% of the world's population, has only 3.4% of its Internet users. South Africa has the most users in SADC, but is fourth on the continent after Nigeria, Morocco, and Egypt. This is seen as another major barrier to Africa's development and one which will lead to even greater inequality both within the continent and between Africa and other regions.

The type of adult education for economic development that has occurred in the last 20–30 years within SADC can be differentiated again in terms of competing interests. For instance, the economic-development projects for women have often been within a modernizing frame which has not challenged the sexual division of labor or attempted to transform women's subordinate positions.

Adult education for economic development occurs most frequently in large companies in the formal sector and for employees at the middle and upper levels. This is a worldwide trend which most often favors educated men and is reflected in the SADC region as well. This is likely to continue as, within the dominant neo-liberal framework, the globalizing economies require flexible, well-educated workers. As Stromquist (1998) suggests, it also requires uneducated workers to service the professional classes and hence the low priority given to literacy in the region despite government rhetoric.

The SADC is trying to position itself in the global economy. The discussions and debates about adult and lifelong learning are shaped directly by this. A key question is what is the primary objective of economic development? Is it to be globally competitive? Which is the predominant view? Or are there alternatives as argued by Klein (2007) and others? Within SADC, these are very pertinent and hotly debated issues within organizations of state, civil society, and business. Adult education is implicated in these debates and political contestations.

Adult education within lifelong learning

The discourse of adult education is being challenged by that of lifelong learning. Lifelong learning has entered the education and development debates of the region, as elsewhere. Education policy documents in Botswana, Namibia, South Africa, for example, all refer to lifelong learning as a goal. The contestation over lifelong learning for what among and within the different sectors is ongoing (see, e.g., Walters, 2006).

The shift in the discourse of adult education to lifelong learning may signal shifts in understandings of relationships between individuals and the nation-state; between individuals and their identities as national, regional, or global citizens. Lifelong learning most commonly relates to the need for continuing education and training for global competitiveness. It may reflect the weakening of the nation-state and the increasing requirements for

transnational and global relationships between the local and the global.

Conclusion

Experiences in SADC have shown how adult education gained in significance both during the struggle for and at the time of independence, with the formation of the new nation-state. Adult education was identified by social movements struggling for liberation in order to mobilize and prepare citizens for rapid change. It was also identified by new governments as necessary to mobilize and (re) shape citizens for their roles and responsibilities in the new society. The utilization of informal and nonformal adult education, within movements, helped to prefigure the new state. It was intimately connected to contestations concerning social and economic development, including notions of democracy and citizenship. Most of the SADC countries had leftist movements (i.e., Marxist, socialist, or social democratic) involved in the national revolutions. South Africa's apartheid government, in turn, played a catalytic role in the 1970s and 1980s in providing a common enemy, against which regional governments and liberation movements struggled. Solidarity among countries against apartheid was therefore an important element of citizenship of the region.

At independence, each of the governments projected adult education as an important part of their reconstruction and development of their nation state. However, few resources were invested and adult education and training has been dispersed among civil-society organizations, workplaces, and various departments in government. At the time of independence, adult education had a powerful symbolic purpose but it has not necessarily translated into systematic programs for the adult population. After the novelty and excitement of the birth of the new nation, adult education has settled back to being barely visible, but is integral to new social movements who are continuing to contest power relations in the society, and notions of citizenship, whether shaped by gender, race, class, geography, and physical ability; or specific cross-cutting issues, like health or criminality.

The discussions on adult education and nation building show that they cannot be divorced from understandings of democracy and development, within local, national, regional, and global contexts. This means that nation building is intimately linked to notions of citizenship, which, on one hand, are determined through the legal frameworks of a country. On the other hand, citizenship is often contested by social movements, political parties, or other social structures, as its interpretation reflects the power relations in the society. Nation building can therefore be driven by state structures or it can be

driven from below. Adult education, whether informal, nonformal, or formal is interwoven into political, economic, and social processes, which comprise the nation-building project at particular historical moments, in specific contexts. In the twenty-first century, with nation-states losing their predominance as defining economic entities, it is hard to hold the question of the relationship between adult education and nation building for long, before questions of the specificities of the local, regional, or global shout for attention.

Bibliography

- Bron, A. and Schemmann, M. (eds.) (2001). *Civil Society, Citizenship and Learning*. Germany: Lit Verlag.
- Finger, M. (2005). *Globalization*. In English, L. (ed.) *International Encyclopedia of Adult Education*, 1st edn., pp 269–273. New York: Palgrave Macmillan.
- Gaventa, J. (2007). Foreword. In Cornwall, A. and Schattan Coelho, V. (eds.) *Spaces for Change? The Politics of Citizen Participation in New Democratic Arenas*. Zed Books: London.
- Klein, N. (2007). *The Shock Doctrine. The Rise of Disaster Capitalism*. London: Penguin Books.
- Korsgaard, O. (1997). *The Impact of Globalization on Adult Education*. In Walters, S. (ed.) *Globalization, Adult Education and Training: Impacts and Issues*, pp 15–26. London: Zed Books.
- Korsgaard, O., Walters, S., and Anderson, R. (eds.) (2001). *Learning for Democratic Citizenship*. Copenhagen: Association for World Education and Danish University of Education.
- Lowe, J. (1971). *The Role of Adult Education*. In Lowe, J., Grant, N., and Williams, T. (eds.) *Education and Nation Building in the Third World*, pp 146–159. Edinburgh: Scottish Academic Press.
- Mamdani, M. (1996). *Citizen and Subject. Contemporary Africa and the Legacy of Late Colonialism*. Princeton, NJ: Princeton University Press.
- Mario, M. and Nandja, D. (2006). *Literacy in Mozambique: Education for All Challenges*. In *Adult Education and Development No 67*, pp 189–203. Germany: DVV.
- Marshall, J. (1990). *Literacy, State Formation and People's Power*. Bellville: University of the Western.
- Nyerere, J. K. (1976). *Adult Education and Development*. In *Adult Education and Development Journal*, No 67, 2006, pp 77–88. Germany: DVV.
- Oduaran, A. (2000). Order out of chaos: Repositioning Africa for globalization through lifelong learning.
- Oga, S. and Okwori, J. (2005). *A Nation in Search of Citizens: Problems of Citizenship in the Nigerian Context*. In Kabeer, N. (ed.) *Inclusive Citizenship*, pp 71–83. London: Zed Books.
- Phadnis, U. and Ganguly, R. (2001). *Ethnicity and Nation-Building in South Asia*. New Delhi: Sage.
- Stromquist, N. (1998). *The Confluence of Literacy, Gender and Citizenship in the Democratic Construction of South Africa. Occasional Paper*. Bellville: CACE Publications, University of Western Cape.
- Thompson, J. (2000). *When Active Citizenship Becomes 'Mob Rule'*. In *Adult Learning*. England: NIACE.
- Torres, R. (2004). *Lifelong Learning in the South*. Stockholm: Sida Studies No. 11.
- Walters, S. (1989). *Education for Democratic Participation*. Bellville: CACE Publications, University of Western Cape.
- Walters, S. (2006). Adult learning within lifelong learning: A different lens, a different light. *Journal of Education* **39**, 7–26.
- Walters, S. and Watters, K. (2001). Adult education in lifelong learning in southern Africa. *International Journal for Lifelong Education* **20**(1–2).
- Youngman, F. (2000). *The Political Economy of Adult Education and Development*. London: Zed Books.

Relevant Websites

<http://www.internetworldstats.com> – Internet World Stats.
<http://www.sarua.org> – Southern African Regional Universities Association (SARUA).

<http://www.sadc.int> – South African Development Community.
<http://unstats.un.org> – UN Department of Economic and Social Affairs.
<http://www.weforum.org> – World Economic Forum.

Economic Outcomes of Adult Education and Training

A Ferrer, University of Calgary, Calgary, AB, Canada

W C Riddell, University of British Columbia, Vancouver, BC, Canada

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Active labor market policies – Policies aimed to facilitate adjustment to job loss, assist the jobless to find work, and encourage labor force participation. In addition to training and skills development, these policies include public employment services (counseling, job-search assistance, information, and placement), youth measures to aid the transition from school to work, subsidized employment, and measures for disabled workers.

On-the-job training (OJT) – Training taking place in the firm to learn procedures that are job related, but where learning is an activity separate from regular job duties.

Portability (of training) – The capacity to utilize skills and obtain returns on human capital investments obtained with a previous employer. This concept is linked to that of the specificity of training.

Selection bias – Error committed when estimating the effect of an intervention such as a training program, due to the fact that treatment and comparison groups are intrinsically different, and would have experienced different outcomes even in the absence of the intervention.

Specific training – Training that provides skills that are exclusive to a particular firm, occupation, or industry and cannot be applied elsewhere. The antonym is general training.

and professions, rapid changes imply a greater need for maintaining and updating skills and knowledge about current best practice. In addition, because of demographic changes and reduced fertility in many countries, the labor force is both aging and growing more slowly than in the past. In an environment in which labor shortages may become more common, there is greater emphasis on the skills of the existing workforce. As the workforce ages, there is a growing risk that older workers who lose their jobs may not have the skills needed to become reemployed.

Skill acquisition of adult workers takes many forms. Learning by doing (acquiring skills during the course of one's regular duties), for instance, is an important source of growth in workers' skills. Here, we focus instead on the acquisition of abilities that are job related, but where learning is a separate activity from regular job activities. Within these, we distinguish between formal (learning leading to the completion of diploma or degree) and informal (seminars, interdepartmental visits, etc.) training. A second important distinction relates to who pays for training. From the institutional point of view, individuals, employers, and governments generally differ in their goals when purchasing or providing adult education and training. Government-sponsored training is often concerned with enhancing the employability of displaced or unemployed workers, disadvantaged workers with limited skills, and those at high risk of social exclusion. Recipients of publicly provided training are likely to be unemployed and to come from the bottom of the skill distribution curve. In contrast, employer-sponsored training is provided to those currently employed, as is much of the adult education and training funded by individuals. Furthermore, as many studies have shown, those who are already well educated and highly skilled are more likely to receive work-related training. Employer-sponsored training is also more likely to provide firm-specific skills, which do not raise the value of the trainee to other employers. Governments, on the other hand, prefer general skills that increase individual employability in a variety of workplaces. Finally, these types of training generally differ in their timing. Employer-sponsored training will likely be spread out over the worker's career but the duration of any particular spell may be short, whereas government-sponsored training is concentrated during spells of unemployment and is usually discontinued when the worker becomes reemployed. Thus, there

Introduction

Adult education and training have been focal points of labor market studies in developed economies since the 1960s. Increased attention to this topic is part of a shift toward greater emphasis on human capital in both economic and social policy in societies that are increasingly knowledge based. Several factors account for the growing interest in these issues. Technological change – especially advances in information and computer technologies – and globalization of production have resulted in a growing demand for highly skilled workers and changes in workplace skills requirements. Within specific occupations

are numerous reasons to distinguish between publicly financed and employer-sponsored training.

Human capital theory has long emphasized the importance of on-the-job training (OJT) and work experience for productivity and earnings. Recently, additional empirical findings have given rise to a new training literature (Acemoglu and Pischke, 1999). Leuven (2005) provides a useful survey of theoretical developments. Evaluating the impacts of adult education and training on outcomes has generated an extensive literature and a wide range of estimated impacts. In this article, we provide an overview of the methods commonly used to evaluate the outcomes of adult education and training and a summary of the results of this literature.

The principal outcomes that have been examined – especially for government-sponsored training – are individual labor market outcomes such as wages, employment, and earnings. Earnings is a useful summary measure because it captures both the price dimension – the wage rate – and the quantity dimension – hours, weeks, and years of work. For employers, the key outcome is worker productivity and thus firm profitability, although the provision of training opportunities may also influence other dimensions of performance such as innovative activity as well as recruitment and retention of employees. The impacts on earnings and productivity are key ingredients in any assessment of the costs and benefits of adult education and training.

Methodological Issues in Evaluating Outcomes

Obtaining credible estimates of the impacts of adult education and training programs is a difficult problem. In modern approaches to evaluation, the nature of the challenge is usually illustrated using the potential outcomes model (Angrist and Krueger, 1999). Each individual in the population of interest – which in this case consists of individuals who may or may not receive training – has two potential outcomes: (1) the outcome that would be experienced if he/she receives training, and (2) the outcome that would be experienced if he/she does not receive training. One of these two potential outcomes is inherently unobservable. For each participant, we can never know what outcomes that individual would have experienced had they not participated. Similarly, for each person who does not participate we cannot know what outcomes would have been experienced had they participated. The problem is thus one of obtaining credible estimates of the counterfactual outcome – the potential outcome that cannot be observed. For participants, this means obtaining an estimate of the outcomes they would have experienced had they not participated in adult education or training.

The gold-standard approach to the evaluation problem is to use random assignment to treatment and control groups. Random assignment ensures that the treatment group that receives training and the control group that does not receive training are statistically equivalent in their pretreatment characteristics; that is, that there are no statistically significant differences in the characteristics of the two groups. Since the two groups are statistically equivalent prior to treatment, any differences in outcomes experienced by the two groups can be attributed to the causal impact of the treatment.

When random assignment is not feasible or desirable, there are a number of nonexperimental methods that can be employed, including regression-based methods, natural or quasi-experiments, instrumental variables, matching estimation, and panel data methods. Each of these methods has advantages and disadvantages, and has different data requirements. Furthermore, no one method dominates other methods in all circumstances. Determining the methodological approach to employ in a particular setting requires a detailed understanding of institutional features relating to program operation and the nature of available data. Simply put, there is no magic bullet in nonexperimental methods.

All nonexperimental methods for estimating the impacts of education and training programs use a comparison group of nonparticipants to estimate the outcomes that participants would have experienced had they not participated in the program. However, in the absence of random assignment, participants and nonparticipants will generally differ along several dimensions, some that are observed by the researcher and some that are unobserved. Because of these differences between the two groups, the outcomes experienced by the comparison group – in the absence of statistical or econometric modeling – are unlikely to provide a reliable estimate of the counterfactual. In particular, since individuals choose whether or not to apply for and/or enroll in a training program, the individuals who receive training constitute a self-selected sample and may differ in various ways from the individuals who do not apply for and/or enroll.

Employers also choose which employees receive training and which do not. Similarly, for government-sponsored training, there may be some selection by those who administer or operate the programs. Simple comparisons of participants and nonparticipants are thus likely to be subject to selection bias and do not provide credible estimates of program impacts. The various nonexperimental methods provide different ways of dealing with the selection bias problem.

In the past two decades, there have been significant advances in nonexperimental evaluation methods. Equally important have been improvements in the quality of the data available for analysis. As a consequence, recent studies using state-of-the-art techniques to assess the impacts

of adult education and training generally provide more credible evidence than earlier studies.

Outcomes of Adult Education and Training

As noted previously, there are several reasons for distinguishing between government-sponsored adult education and training and private-sector-sponsored training. Although the expenditure on private-sector-sponsored training far exceeds that by governments – even in countries such as Sweden that devote considerable public expenditure to training – more is known about the outcomes of publicly supported training. This situation reflects a greater desire by governments (and taxpayers) to evaluate the impacts of their programs, as well as data challenges in assessing the impacts of employer-based training.

Government-Sponsored Training

Government-sponsored training is part of the broader category of active labor market policies (ALMPs) that facilitate adjustment to change, assist the jobless to find work, and encourage labor force participation. These policies include public employment services (counseling, job search assistance, information, and placement), youth measures to aid the transition from school to work, subsidized employment, and measures for the disabled. Our focus here is on training programs, which are the most important active measure in many Organization for Economic Co-operation and Development (OECD) countries.

There is a significant amount of research assessing the labor market outcomes of government-sponsored training and other active programs. The comprehensive survey by Heckman *et al.* (1999) summarizes more than 40 studies from the US and Europe. Other useful surveys of this literature include LaLonde (1995), US Department of Labor (1995), Martin and Grubb (2001), and Kluve and Schmidt (2002). Our treatment is brief because several excellent surveys are available and there are space constraints.

Relative to other OECD countries, the US has traditionally displayed a stronger commitment to rigorous evaluation of government programs, including substantial use of randomized trials. As a consequence, the state of knowledge about the effectiveness of these programs is dominated by US evidence. However, as the recent surveys by Martin and Grubb (2001) and Kluve and Schmidt (2002) indicate, there is growing use of serious evaluations in several European countries, resulting in a rapidly expanding body of evidence on the impacts of active programs.

Several salient features emerge from reviews of the US evidence (LaLonde, 1995; US Department of Labor, 1995; Heckman *et al.*, 1999). First, the impacts on earnings of disadvantaged workers are mixed, being positive for some groups and zero or even negative for others. In this regard,

it is worth noticing that training programs for displaced workers (who typically have considerable work experience and are highly motivated) have in general, a better track record than those for the disadvantaged (Jacobson *et al.*, 2005b). Other generalizations include:

- government-sponsored training significantly raises the earnings of economically disadvantaged adult women;
- the effects of these programs on disadvantaged adult men are often smaller than on women, and not always positive;
- the impacts of training on out-of-school youth are generally zero or negative;
- classroom training can be effective for adult women, but has limited success for adult men, especially those with low education; and
- best results are obtained when classroom training or OJT have a strong work experience component linked to local employers.

Greenberg *et al.* (2004) offer evidence from meta-analytic techniques on the persistence of these effects. Impacts on earnings seem to diminish over time for adult males and youth, but remain stable over time for females.

A second feature of the US evidence is that estimated impacts, even when they are positive, are generally modest in size. As noted by LaLonde, this outcome is not surprising, given the limited duration and cost of these programs:

The best summary of the evidence about the impact of past programs is that we got what we paid for. Public sector investments in training are exceedingly modest compared to the magnitude of the skill deficiencies that policymakers are trying to address. Not surprisingly, modest investments usually yield modest gains... (LaLonde, 1995: 149)

A third salient feature of the US evidence is significant heterogeneity in estimated effects. Evaluation of experimental programs shows earnings effects that range from negative impacts for individuals who never worked before, to large positive impacts on economically disadvantaged female household heads or economically disadvantaged male household heads. Substantial variation in program impacts across sites is common in training programs using random assignment. Variability of estimated impacts on earnings is also evident in nonexperimental programs. However, when studies most susceptible to selection bias are removed, the qualitative evidence from nonexperimental programs is similar to that of experimental programs (Heckman *et al.*, 1999).

In Canada, the published literature on government training programs is considerably sparser – likely due to the fact that program evaluations are carried out internally by government agencies and the findings are not published in academic or policy journals (Riddell, 1991). Riddell

(1995) reviews employment and training programs that operated in the 1980s and early 1990s. Programs targeted on the economically disadvantaged had modest effects similar to those found in US studies, but impacts of programs serving clients with fewer barriers were larger. Subsequent analysis by Park *et al.* (1996) found significant gains from training programs for some groups, including women reentering the workforce and individuals trained in areas with identified skills shortages.

European programs are more focused on speeding the transition to work and reducing unemployment, especially among youths. The difference in emphasis reflects the facts that European youth are generally less economically disadvantaged than US youth, but are more likely to be unemployed for extended periods of time. A common finding of European studies is that training results in substantial gains in employment but has little impact on wages (Heckman *et al.*, 1999; Kluve and Schmidt, 2002).

In summary, government-sponsored training has a mixed record, with rising earnings of some groups (adult women and displaced workers) but having little or no impact on earnings of others (disadvantaged adult men and youths). Even when they are positive, the impacts on earnings are modest and are usually not large enough to substantially reduce poverty rates. Training does, however, have a better track record in improving the transition to employment.

Two general responses have emerged in response to the mixed performance of public training programs. One, illustrated by Carneiro and Heckman (2003), is to argue that problems associated with low skills should be addressed much earlier in the life cycle. The other, illustrated by Martin and Grubb (2001), argues that public training programs can be made more effective by improving their design. Design features that they regard as crucial include careful targeting on participants most likely to benefit from training, keeping programs small in scale, including a strong work experience component to establish links with employers, and having programs that produce a certification that is recognized in the labor market.

Private Sector Training

Private-sector training refers to forms of skill acquisition that are job related and where learning is an activity separate from regular job duties. This type of skill acquisition may include both formal and informal learning, be general or specific and either employer- or individually-sponsored. In general, employer-financed training tends to be firm specific and informal in nature. However, firms may finance the acquisition of (formal) vocational training, such as apprenticeships, that is more general in nature. Studies of work-related training focus on one or the other type of training, depending on data availability.

Methodologically, the literature surveyed here relies on nonexperimental data. The selection problems are

acute, as employers are likely to select for training the most promising individuals, who may have higher-than-average earnings regardless of training. Similarly, individuals who choose to undertake training are likely to be more productive or to have unobserved characteristics, such as motivation, associated with higher wages. The papers we consider here use different methods to resolve this problem.

Interest in the training policies of firms and the skill acquisition choices of workers originated in the US during the 1980s when slow productivity growth (relative to Japan) was at the forefront of the economic agenda. However, there was little information about training in available data (Barron *et al.*, 1997). During the next decade, research produced many studies on the determinants of training and on the effects of training on wages, thanks to the availability of the National Longitudinal Survey of Youth (NLSY), which follows a representative sample of youth over the period since 1978, gathering exhaustive information about their labor market experiences. The survey asks about any type of training, other than schooling, military training, and government-sponsored training, and further categorizes it as: (1) company training or OJT, (2) apprenticeship programs, and (3) off-the-job training (OFT). The richness of information regarding skill-related characteristics contained in the survey allows for adequate measurement of the variables involved, including usually unobservable variables such as those measuring ability. In addition, the longitudinal nature of the data allows researchers to address the methodological issues regarding selection bias.

Most of the US literature is hence based on the NLSY. One of the earlier studies by Lynch (1992) focuses on a subsample of non-college-educated youth to estimate the effect on wages of weeks spent in company training, OFT, or apprenticeship training. The estimated coefficients translate into approximately 25% higher wages for an average apprenticeship of 63 weeks and 15% for an average spell of company training of 31 weeks. Using all workers in the NLSY sample, Veum (1995) finds smaller returns of 7.5% for an episode of company training, 6% for an apprenticeship, and 11% for an occurrence of OFT. In addition, he estimates that company training and OFT increase wage growth by similar magnitudes. Both these studies are confined to episodes of training lasting more than 1 month. Later, Parent (1999), using a more sophisticated methodology and not restricting the sample to long training episodes, estimates that 1 year of OFT or OJT training increases earnings between 12% and 16%.

Further issues arise within this literature. One relates to the portability of training. Lynch (1992) separates training into that received with a current employer and that received with a previous employer, finding high returns to apprenticeships and OFT received during a

previous job, but no returns to company training received during a previous job. The results in Parent (1999), on the other hand, suggest that employer-based training and OFT are the most portable forms of training. Given the small fraction of individuals undertaking apprenticeships (around 1%), results regarding this form of training should be treated with caution. A second issue is who pays for private-sector training. Parent (1999) finds that workers partially pay for OJT with lower initial wages. Barron *et al.* (1999) use a different data set and find that although the initial wages seem to be smaller for workers undergoing OJT, the difference is small. The general perception in this matter is that firms pay for OJT training and reap the benefits through higher productivity of the firm.

In Canada, private-sector training has been much less analyzed and the results are nonconclusive. Parent (2003) uses the Follow-Up of the School Leavers Survey (FSLs) and finds that participation in employer-provided training raises male earnings by 10% but not significantly so for females. Results from this study also suggest that training may increase employment. Hui and Smith (2003), using the Adult Education and Training Survey (AETS), find positive effects of self/employer-financed training on employment for women. In addition, small positive effects on weekly earnings from employer-financed training are apparent for both genders, while self-financed training seems to have small negative impact on female wages. These findings seem to be contradicted by Havet's (2006) study, which uses matched employer–employee data and finds a positive impact of firm-provided training on the wages of women, but not of men.

Studies from Europe also tend to find positive effects of training on labor outcomes. In Great Britain, Greenhalgh and Stewart (1987) find positive effects of vocational training on occupational status. Booth (1993) estimates that 1 week of training raises the earnings of graduates by 1%. Both studies find significant gender differences. More recently, Blundell *et al.* (1999), use the National Child Development Study (NCDS) to study the impact of work-related training on the earnings of UK workers. They find positive returns (5–6%) for employer-provided training and also that this type of training seems transferable across employers. These findings are corroborated in a later study by Arulampalam and Booth (2001) that uses the same data.

For countries other than North America or the UK, evidence on the returns to training is less abundant. Goux and Maurin (2000) estimate wage returns to employer-provided training for France. They report around 7% higher wages among those who received training, but the estimates become insignificant when they account for firm selection of workers on the basis of posttraining mobility. Similarly, Pischke (2001) finds that, in Germany, 1 year of full-time work-related training (training during leisure time) increases wages between 3% and 4%. When he

accounts for selection based on wage growth, the returns are higher in magnitude, from 3% (males) to 6% (females), but not statistically significant. Schøene (2002) finds that employer-financed training participation is associated with 1% higher wages in Norway when controlling for selection based on seniority and job complexity (down from a 5% return in the absence of selection controls).

Overall, the returns to private-sector training seem quite high – higher, for instance, than the returns to schooling. In contrast, the incidence of training is not large, which is puzzling if there are such high returns. A leading explanation for the high returns to training is selection bias – that is, high-productivity workers are more likely to receive training and to have higher wages and/or wage growth. All papers find that addressing this selection problem is important. Estimates produced with a selection correction procedure are generally smaller than those obtained without, corroborating the heterogeneity of the returns to training. The effect of accounting for selection varies depending on the method used. In general, estimates are reduced by 40–50%. In some cases, the results are quite extreme. For instance, when selection is based on worker's mobility, since firms are less likely to train workers at high risk of leaving in the near future, the effect of training practically disappears (Goux and Maurin, 2000). A potential explanation could be that most of the benefits of training accrue to job movers and not to job stayers. Gerfin (2003) provides some evidence for this hypothesis in his study of Swiss work-related training. He shows that training undertaken during the last year generates the highest returns to individuals who changed jobs (around 5% increases in monthly earnings) versus those who stayed at the firm (1.2%). Other studies find no returns to training once selection is taken into account. For instance, Leuven and Oosterbeek (2007) find that when comparing the returns to training of Dutch workers who took training with the returns of workers who intended to undertake training but did not do so because of a random event (sickness or family circumstances), earnings differences are not statistically significant. There is also cross-country evidence from the European Community Household Panel indicating that the effect of training is significantly smaller when individual heterogeneity is accounted for (Bassanini *et al.*, 2007).

Finally, other factors may explain the correlation between high wages and training. Promotions, for instance, are a factor generally overlooked. These positively influence both pay and training. In addition, the returns to training are likely to differ by type of job, with more productive jobs (managerial/professional) requiring more training and higher pay than blue-collar jobs. Failure to account for these factors is also likely to overestimate the effect of private-sector training (Frazis and Lowenstein, 2005).

In summary, studies of informal private sector training seem to agree on the following:

- returns to training are generally positive and significant, particularly employer training and OFT;
- there are gender and ethnic differences in the returns to training, although the sign of the difference seems to be country specific; and
- there is substantial evidence that unaccounted-for factors lead to overestimating the returns to training. In general, those who undergo training seem to be those whose productivity, occupation, or other characteristics would lead them to have high wages regardless of training.

The papers examined so far in this section do not formally distinguish between formal and informal education, although it is expected that some OFT may be conducive to a degree or diploma. However, there are studies analyzing the impact of adult education and training that focus specifically on the effect of acquiring formal schooling later in life. Although the notion of adult education is country specific, it is generally accepted that formal schooling is associated with the completion of courses toward the achievement of a degree or diploma. The skills acquired with formal schooling are also more likely to be general rather than trade or profession specific.

The North American literature finds substantial returns to formal certification for older workers. Leigh and Gill (1997) and Jacobson *et al.* (2005a) find returns to community college in the US for workers over age 28 that are around 8% (9%) for males (females). In Canada, Zhang and Palameta (2006) use longitudinal information from the Survey of Labor and Income and Dynamics to evaluate the impact of formal schooling on earnings for individuals who have been out of school for more than a year and then enrolled in formal education. They find substantial earnings gains for those individuals who obtained a certificate (7% increase in earnings for men, 10% for females). In this study, younger workers who switch firms after obtaining a certificate gain the most from their studies, whereas workers over 35 reap higher benefits staying with the same firm. Ferrer and Menendez (2007) also find evidence for Canada of large earning effects from acquiring formal education later in life.

However, the European findings about the impact of adult formal education show that formal qualifications obtained later in life seem to have little impact on earnings. With the exception of the Blundell *et al.* (1999) study, which finds that all forms of work-related training leading to formal qualifications enhance earnings by 5–10%; other studies do not find such positive returns. Egerton (2000) uses 10 years of the British General Household Survey, which incorporates a rich data set of covariates including father's social class, to examine this issue. Her results show no significant difference in earnings between

full-time mature students and early graduates, although she does not address selection bias. Jerkins *et al.* (2003) use panel data from the NCDS to analyze the impact of adult education on employment and earnings. Their findings reveal that episodes of adult education, particularly in occupational training, have positive effects on employment but a limited effect on wages, except for the least-qualified individuals. Part of the difference in results with Blundell *et al.* (1999) is likely due to differences in the goals of study, which lead them to consider different control groups. In Sweden, Ekström (2003) analyzed the impact of adult secondary education on annual earnings during the early 1990s, finding a negative effect for males and only weakly significant positive effects for females. Later, Albrecht *et al.* (2004) follow the large expansion of Swedish adult education programs during 1997 through 2002, called the Knowledge Lift, to estimate the impact on annual earnings and employment of increasing formal schooling for the low skilled. Their results show no effect of Knowledge Lift programs on earnings or employment, with the exception of an increase in the employability (but not earnings) of young men.

In summary, the literature on employer-based training finds that this type of human capital investment generally has positive effects on labor market outcomes such as wages. The magnitude of the returns varies depending on the type of training undertaken and the methodology employed. As is the case with the literature on the returns to education, careful econometric analysis is needed to take into account the possibility of selection bias arising because training is undertaken by individuals who are more productive and would have commanded higher wages regardless of training.

See also: Apprenticeships; Evaluation of Adult Education and Training Programs; Lifelong Learning; Participation in Adult Learning; Training and Learning in the Workplace.

Bibliography

- Acemoglu, D. and Pischke, J.-S. (1999). The structure of wages and investment in general training. *Journal of Political Economy* **107**(3), 539–572.
- Albrecht, J., van den Berg, G. J., and Vroman, S. (2004). The knowledge lift: The Swedish adult education program that aimed to eliminate low worker skill levels. *IZA Discussion Paper No. 1503*. Institute for the Study of Labor (IZA), Bonn, Germany.
- Arulampalam, W. and Booth, A. (1998). Labour market flexibility and skills acquisition: Is there a trade-off? *British Journal of Industrial Relations* **36**(4), 521–536.
- Barron, J., Berger, M., and Black, D. (1997). How well do we measure training? *Journal of Labor Economics* **15**(3), 507–528.
- Barron, J., Berger, M., and Black, D. (1999). Do workers pay for on-the-job training? *Journal of Human Resources* **34**(2), 235–252.
- Blundell, R., Dearden, L., and Meghir, C. (1999). *Work-Related Training and Earnings*. London: Institute for Fiscal Studies.
- Booth, A. (1993). Private sector training and graduate earnings. *Review of Economics and Statistics* **75**(1), 164–170.

- Carneiro, P. and Heckman, J. (2003). Human capital policy. *NBER Working Paper No. 9495*. Institute for the Study of Labor (IZA), Bonn, Germany.
- Egerton, M. (2000). Pay differentials between early and mature graduate men: The role of state employment. *Journal of Education and Work* **13**(3), 289–305.
- Ekström, E. (2003). Earnings effects of adult secondary education in Sweden. *Working Paper Series No. 16*. IFAU – Institute for Labour Market Policy Evaluation, Uppsala, Sweden.
- Ferrer, A. and Menendez, A. (2007). The returns to flexible post-secondary education: The effect of delaying school. *CLSRN Working Paper No. 7*. CLSRN, Vancouver, BC, Canada.
- Frazis, H. and Lowenstein, M. (2005). Reexamining the returns to training functional form, magnitude and interpretation. *Journal of Human Resources* **XL**(2), 453–476.
- Gerfin, M. (2003). Work-related training and wages – an empirical analysis for male workers in Switzerland. Mimeo, Volkswirtschaftliches Institut, Universität Bern.
- Goux, D. and Maurin, E. (2000). Returns to firm provided training: Evidence from French worker–firm matched data. *Labor Economics* **7**, 1–19.
- Greenberg, D. H., Michalopoulos, C., and Robins, P. K. (2004). What happens to the effects of government-funded training programs over time? *Journal of Human Resources* **39**(1), 277–293.
- Greenhalgh, C. and Stewart, M. (1987). The effects and determinants of training. *Oxford Bulletin of Economics and Statistics* **49**(2), 171–190.
- Havet, N. (2006). La valorisation salariale et professionnelle de la formation en entreprise diffère-t-elle selon le sexe? l'exemple canadien. *Working Paper 06-02 Groupe d'Analyse et de Théorie Économique (GATE)*.
- Heckman, J. J., Lalonde, R., and Smith, J. (1999). The economics and econometrics of active labor market programs. In Card, D. and Ashenfelter, O. (eds.) *Handbook of Labor Economics*, ch. 31, pp 1865–2085. Amsterdam: North Holland.
- Hui, S. and Smith, J. (2003). *The Labour Market Impacts of Adult Education and Training in Canada*. Ottawa, ON: Statistics Canada.
- Jacobson, L., LaLonde, R., and Sullivan, D. G. (2005a). Estimating the returns to community college schooling for displaced workers. *Journal of Econometrics* **25**(1–2), 271–304.
- Jacobson, L., LaLonde, R., and Sullivan, D. G. (2005b). Is retraining displaced workers a good investment? *Economic Perspectives* **XXIX** (2), 47–66. Federal Reserve Bank of Chicago.
- Jenkins, A., Vignoles, A., Wolf, A., and Galindo-Rueda, F. (2003). The determinants and labor market effects of lifelong learning. *Applied Economics* **35**, 1711–1721.
- Kluve, J. and Schmidt, C. (2002). Can training and employment subsidies combat European unemployment? *Economic Policy* **35**, 409–448.
- Lalonde, R. (1995). The promise of public sector-sponsored training programs. *Journal of Economic Perspectives* **9** (Spring), 149–168.
- Leigh, D. and Gill, A. M. (1997). Labor market returns to community colleges: Evidence for returning adults. *Journal of Human Resources* **32**(2), 334–353.
- Leuven, E. (2005). The economics of private sector training: A survey of the literature. *Journal of Economic Surveys* **19**(1), 91–111.
- Leuven, E. and Oosterbeek, H. (2007). A new approach to estimate the wage returns to work-related training. *IZA Discussion Paper No. 526*. Institute for the Study of Labor (IZA), Bonn, Germany.
- Lynch, L. (1992). Private-sector training and the earnings of young workers. *American Economic Review* **82**(1), 299–312.
- Martin, J. P. and Grubb, D. (2001). What works and for whom: A review of OECD countries' experiences with active labour market policies. *Working Paper Series 2001: 14*. IFAU – Institute for Labour Market Policy Evaluation, Uppsala, Sweden.
- Parent, D. (1999). Wages and mobility: The impact of employer-provided training. *Journal of Labor Economics* **17**(2), 298–317.
- Parent, D. (2003). Employer-supported training in Canada and its impact on mobility and wages. *Empirical Economics* **28**, 431–459.
- Park, N., Power, B., Riddell, W. C., and Wong, G. (1996). Special Issue (Part 1): An Assessment of the Impact of Government-Sponsored Training. *Canadian Journal of Economics* **29**(1), S93–S98.
- Pischke, J. (2001). Continuous training in Germany. *Journal of Population Economics* **14**, 523–548.
- Riddell, W. C. (1991). Evaluation of manpower and training programs: The North American experience. In OECD (ed). *Evaluating Labour Market and Social Programs*, pp 43–72. Paris: OECD.
- Riddell, W. C. (1995). Human capital formation in Canada: Recent developments and policy responses. In Keith, G. B. and Beach, C. M. (eds.) *Labour Market Polarization and Social Policy Reform*, pp 125–172. Kingston, ON: School of Policy Studies, Queen's University.
- Schøene, P. (2002). Why is the return to training so high? Mimeo, Institute for Social Research, Oslo, Norway.
- US Department of Labor (1995). *What's Working (and What's Not): A Summary of Research on the Economic Impacts of Employment and Training Programs*. Washington, DC: US Department of Labor.
- Veum, J. (1995). Sources of training and their impact on wages. *Industrial and Labor Relations Review* **48**(4), 812–826.
- Zhang, X. and Palameta, B. (2006). Participation in adult schooling and its earnings impact in Canada. *Statistics Canada, Analytical Studies Branch Series 2006. Research Paper No. 276*. Business and Labour Market Analysis Division, Ottawa, ON, Canada.

Further Reading

- Angrist, J. and Krueger, A. (1999). Empirical strategies in labor economics. In Card, D. and Ashenfelter, O. (eds.) *Handbook of Labor Economics*, pp 1277–1357. Amsterdam: North Holland.
- Bassanini, A., Booth, A. L., Brunello, G., De Paola, M., and Leuven, E. (2007). Workspace training in Europe. In Brunello, G., Garibaldi, P., and Wasmer, E. (eds.) *Education and Training in Europe*, pp 141–157. Oxford: Oxford University Press.
- Blundell, R., Dearden, L., Meghir, C., and Sianesi, B. (1999). Human capital investment: The returns from education and training to the individual, the firm and the economy. *Fiscal Studies* **20**(1), 1–23.
- OECD (1994). *The OECD Jobs Study – Facts, Analysis, Strategies*. Paris: OECD.
- OECD (2006). *Boosting Jobs and Incomes*. Paris: OECD.

Health and Adult Learning

P Bélanger, UNESCO Institute for Education, Hamburg, Germany

M Robitaille, University of Quebec in Montreal, Montreal, QC, Canada

© 2010 Elsevier Ltd. All rights reserved.

Generally absent until now from the *International Encyclopedia Education*, the theme adult learning and health is taking momentum in many countries as well as in international multilateral networks (UNESCO, 1997; US Dept. of Health, 2000; ICAE, 2001; WHO, 2005b). The growing recognition of “health as a basic human right” and the appeal for “relevant, equitable and sustainable access to health knowledge” (UNESCO, 1997), as well as the crisis of the welfare state, particularly with regard to health systems, is driving up the social demand for lifelong health education throughout life.

A new policy context is emerging linking health and lifelong learning policies. The call for Health For All, heard increasingly over the last two decades in multilateral organizations (European Commission, 2004; WHO, 2005a) finds growing legitimacy in positioning health as a key aspect of human development. In 2000, for example, the United Nations Economic and Social Council adopted a resolution recognizing the right to health, a state of physical and mental well-being, as “essential to the exercise of other rights.”

However, the attainment of this overall objective and of the more specific health-related Millennium Development Goals (to reduce infant mortality, improve maternal health, and fight world pandemic such as HIV/AIDS, malaria, and tuberculosis) could be severely hampered if the increasing cost of current curative policy cannot be controlled and the productivity of the health system is not improved. Researchers and policymakers are exploring ways to solve this predicament by pursuing the Health For All objectives while paradoxically searching to curb the increasing demand for curative services. One of the key responses to this contradiction has to do with adult learning.

Two major trends are emerging linking health and adult learning. First, health promotion is increasingly acknowledged as a critical component of health policies. It refers to the documented impact that general levels of health competency and knowledge have on both the social demand for health and the cost of curative services among different strata of a population. Health literacy and health promotion thus becomes mounting issues in national health policies. A second significant tendency relates to continuing medical education of health personnel. Intensive research and development in this domain is transforming the long-existing perspective and practices within medical and paramedical professions and, as a result, may turn medical and nursing continuing education into a prototype of a much larger

trend in continuing professional development (CPD) among other professions and occupations.

Health Literacy

In a time when an ever-growing part of public budgets is allocated to health services, the rapidly increasing need for healthcare tends to change the relations between health professionals and the public. In the context of a welfare state, already under pressure from neoliberal forces and now confronted by a continuously growing demand for healthcare services, the former pattern of interrelation between health professionals and patients is increasingly under time and economic pressure. Medical doctors, nurses, and therapists have less time available to spend with each of their clients who at the same time tend to question their traditional identity as patients. The tendency is, on the one hand, to reduce direct contact time by passing on more information through printed or electronic communication (pamphlet of information about the preparation of a coming surgery or treatment, instructions on direction for use of prescribed drugs, public vaccination campaign, information on preventative measures) and, on the other hand, to acknowledge ambiguity in past modes of unilateral relation between professionals and the public.

The trend toward written communication has the unexpected consequence of relying on people's capacity for initiative to be informed sufficiently and properly, and it does so without fully recognizing the impact of uneven distribution of health cultural capital. A mismatch is thus being created in these new healthcare systems between the growing complicated demand with which national health systems are confronted, and the average basic competency of the public for whom more self-gathering and treatment of information is required (Rudd *et al.*, 2004).

This unspoken transition in the mode of interaction between health professionals and people is bound to produce inequalities and puts at risk the attainment of the Health For All objective. In such a context, equitable and sustainable development of health literacy and health-related adult learning in general may become a key element of future public health and lifelong learning policy.

Of course, the association between formal education and health status is known in both developed and developing countries (Kickbusch, 2001), as well as between adult learning and socioeconomic benefit. What is new is the impact of various levels of health skills and knowledge on health condition. Health literacy represents the cognitive and social skills that determine the curiosity and ability of individuals to gain access to understand and use information in a way that will promote and maintain good health. It involves the ability to judge, sift, and act in the context of one's own life on the information provided (Kickbusch, 2001).

New inquiries on health literacy and more precisely on the direct assessment of people's health-related basic skills and knowledge are revealing a relation between levels of such competency and levels of education and income (Murray *et al.*, 2007). According to these surveys, between a third and half of all adults in postindustrial societies struggle with low levels of health literacy, a percentage much higher among the aging population. They have difficulty understanding and acting on currently available health information.

Health literacy levels tend, in many ways, to influence people's ability to benefit from healthcare and to prevent illness and avoid pandemic diseases. It affects the capacity to discriminate information printed on drugs or food labels, to search and use accessible information in order to prevent sickness. Health literacy also has an influence on the ability to navigate into the labyrinth of current healthcare services. This has to do, of course, with the capacity to read written dietary or medical advice, complete open entry forms, understand complementary professional advice given on paper, grasp the importance of danger notices found on domestic product packaging, access critical information on health and safety at home or at the workplace, and to read about disease prevention.

However, health literacy is more complex. It means not only to read messages, but also to interpret them and proceed accordingly in one's specific life context. It means to be able to draw consequences from consent forms, deal with health alerts conveyed in the media, and detect in one's own immediate environment early signs of emerging sickness. It means to have skills to interpret symptoms and be able to tell a professional one's health story or the story of a family member in order to inform him appropriately. Health competency is needed also to benefit proactively from preventive services operating increasingly through distance communication. Thus, health literacy impacts directly on the accessibility of the health system through people's capacity to get, screen, and mobilize information now mostly delivered not only in print, but also in a more diffuse way, through broader consequences of cultural and educational advantages or disadvantages throughout the life course.

Medical journals are beginning to address this issue. Uneven distribution of health competency is challenging

not only the functioning of current healthcare systems, but also its universal accessibility (Sentell and Halpin, 2006; Somnath, 2006; Paasche-Orlow *et al.*, 2006). The practice of including an adult literacy variable in health disparity research is increasing; it reduces the explanatory power of the already known variables (Nutbeam, 2000). Because health literacy significantly affects people's health and the ability of a system to provide effective quality healthcare (Institute of Medicine of the National Academies, 2004), it is of no surprise that growing awareness on the impact of health literacy is giving new impetus to the already well-known domain of public health. Health promotion comprises efforts to enhance positive health and prevent ill health, through the overlapping spheres of health education, prevention, and health protection (Downie *et al.*, 1990), and it is often centered on lifestyle diseases. However, following the horizontal perspective developed in health literacy, health promotion tends now to be extended to the full continuum of health-related activities and tends to rely more on interactive communication and adult learning approaches (see, e.g., Davis *et al.*, 2003b, 1999).

We observe, in the last few decades, a diffuse but growing demand for health-related adult learning, either through structured activities or through supported and unsupported self- and informal learning. Associated with shifts in attitudes and behaviors during adulthood, participation in adult learning is seen as an important element of health prevention policies (Feinstein and Hammond, 2004). A growing number of study circles in Nordic countries (Swedish National Council for Adult Education, 2005: 24–26) and already nearly one-fourth of night courses in the German adult education centers are related to health issues (Reichert and Huntemann, 2006; Nuissl and Pohl, 2004). A similar emerging trend is observed in adult literacy where some education ministries are introducing health and hygiene modules in their adult literacy programs (ICAE, 2001). Studies on informal learning in Canada indicate similar content interest (Livingstone, 1999). National health departments are developing evidence-based health education programs (Bartholomew *et al.*, 2001). They are setting up nutritional education programs, environmental sanitation training, prenatal courses, etc. Studies are made on patterns of health information handling in order to improve health education interventions (Zanchetta *et al.*, 2007; Kok *et al.*, 2004; Zorn *et al.*, 2004).

However, health literacy is more than functional; it means more than transmitting information and developing skills to be able to read pamphlets and successfully make significant appointments with physicians. It is also to be interpreted in the emerging expectation of people to participate in decision making related to their own health and to follow through on these decisions. In discourse and research on literacy in general, and also in health literacy

in particular, we observe “a new introduction of humans as active agents in the construction, negotiation over, and transformation of their social worlds” (Barton *et al.*, 2000: 5). Some refer to this as a paradigm shift from pathology to empowerment (Shernoff, 1997). Hohn (2002), a well-known author in this field of health literacy, refers to “empowerment health education.” The reality of health literacy is complex in another dimension. Current health literacy skills assessments tend often to ignore the multi-cultural dimension of health reality, the many social health literacies (Street, 1995), the popular knowledge, “the different medical traditions” and complementary “local ways of healing” (UNESCO, 1997: 6). In that sense, interactive and critical health literacy does not focus only on compliance, it relates to the autonomy of the subject, with the capacity of a local community to act on their health conditions, with people better equipped to overcome structural barriers to health (Nutbeam, 2000) and modify the relation between professionals and the subjects.

Aware of the important gap between the complexity of current health materials and the basic skills of the intended public (Rudd, 2004), public health agencies are looking not only on people’s capacity to participate in this evolving system, but also on communication and practice of institutions and health personnel. Public health agencies tend to revise accordingly their information–education–communication (IEC) plan. Communication strategies are developed using plain-language or clear-communication approaches. In the same perspective, new continuing education programs are created to help physicians, dentists, and nurses better interact and communicate with their various patients. Some journals of continuing medical education (e.g., the *American Journal of General Internal Medicine*) are even proposing proxy measures or screening items to help professionals identify patients with limited health literacy skills.

The issue of the various levels of health literacy and more broadly of its consequence on equity, quality, and productivity of health services is making a big push for stronger investment in health-related education and adult learning, particularly in low-income communities (Rudd *et al.*, 2004). Paradoxically, the recommendation of WHO to allocate 5% of national health budgets to health promotion, prevention, and education may put a strain on the already tight public finances, but in the long term may be the best strategy to bend demand for curative services as well as reduce the cost of those services precisely among people at risk. In this new context, prescription of learning (Institute of Medicine of the National Academies, 2004; James, 2001) may well become an integral part of national health policy: prescription of lifelong learning not only among the public, but also among the professionals. Already in 1986, WHO insisted, in its Ottawa Declaration, on the necessity of a two-pronged approach: individual participation and structural change. People cannot

assume more responsibility for more aspects of their health without more protection and better opportunities to improve their health competency (Gruman, 2003).

Professional Continuing Education in the Health Sector: In Transition

Professional continuing education is in transition (Cervero, 2001), particularly in the health domain. It is one of the fastest growing areas of adult learning. The number of hours spent on continuing education activities in the course of both medical doctors and nurses’ occupational careers, after their certification and licensure, tends to exceed the duration of their initial education and training (Davis *et al.*, 2003a). In many countries already, taking part in continuing education has been made obligatory for professionals and conditional to keeping their certification. Continuing medical education (CME) has become an essential effector arm in complex healthcare systems (Davis *et al.*, 1999). The demand for continuing education among health professionals is expected to grow even further. The accelerating pace of clinical and biopharmaceutical or medical research as well as of epidemiological studies requires doctors and other professionals to update their knowledge on indicators of disease predictability and on new practices or medication. It explains in part the growing demand for continuing education. Indeed, more than 80% of physicians as well as nurses of the coming decade have already left universities or colleges and have terminated their initial education. In fact, CME has already become in the United States alone an industry producing more than \$3 billion of activities every year. CME has even become a recognized international discipline (Davis, 1998) with its scientific journals like the *Journal of Continuing Education of Health Professions*, the *Canadian Journal of Continuing Medical Education*, and the *Journal of Advanced Nursing*.

What is of particular interest in the continuing education of health professions is not its expansion, which constitutes a trend that one could observe, though at different pace, in other professional fields, together with the typical diversity of the education agencies involved.

More significant is the shift in orientation of activities within the health sector. Up to the mid-1990s, CME activities tended to take the form of obligatory or voluntary information and formal education sessions aiming at updating knowledge among field practitioners. The logic of action tended to be one of passive dissemination of information and knowledge transfer in order to keep professionals up to date with the recent developments in their field of practice (Bero *et al.*, 1998).

Particularly because CME, together with the continuing education of nurses, became an intensive area of activities involving a growing amount of financial resources, many

studies have been requested to assess the acceptability and effectiveness of prevailing approaches (Thomas *et al.*, 2006; Davis *et al.*, 2003b, 2003a, 1999, 1995; Elwyn and Hocking, 2000; Brigley *et al.*, 1997; Kok *et al.*, 1997). Various aspects were scrutinized: continuing education practices, knowledge transfer, impact of various formal and nonformal strategies, contexts and conditions differently conducive to efficient linking between CME and practice, uneven circulation of clinical and scientific knowledge, as well as required alteration of initial education. Most of these assessments came to the same conclusion: formal and didactic transfer of information tends to have low educational value. They have little effect on professional practice. Attendance at passive educational events, even when reaching most practitioners through incentive, tends to have limited impact on individual practice and on the activity of health services involved.

Since then, a shift of orientation is taking place with new approaches being assessed for their impact not only on upgrading the knowledge of practitioners, but also on their daily practice as well as the operation of medical units and clinics. The aim is to make interventions relevant to the identified individual and organizational needs (Elwyn and Hocking, 2000). Some researchers have translated this change as a transition from CME to CPD (Du Boulay, 2000; Peck *et al.*, 2000). Essentially, what is going on is a shift in logic of action through which priority is given to supporting individual professionals in the ongoing development of their capacity to act, observe systematically, gather evidence, and co-produce research or at least be able to discriminate in one's context new knowledge relevant for one's practice. Then, validation and certification could be done but *a posteriori*, separated from the situated learning process.

Such approaches tend to be more tailor-made, more situated in their content and in their process, referring to both the perceived needs of professionals and local organization's demand. The input of recent scientific evidence produced by research networks remains central, but the question raised concerns the mobilization or appropriation of such external knowledge and its transfer into new capacity for action. However, such transition takes time and, even though more effective approaches have been proven, use of least-effective approaches still continues (Bloom, 2005).

A recent trend has emerged trying to create organic links between scientific institutional research, clinical observation, and practice. The aim is to connect more closely formal research agendas with questions emerging from practice, while simultaneously training field professionals on evidence-based medicine (Cusick and McCluskey, 2000), and offer advice and guidance to practitioners interested in research (Bateman and Kinmonth, 2001). This shift of orientation, bound to take place eventually among other professions, is happening earlier

among the health professions for many contextual factors. The first one, already mentioned, is the huge investment in time and money that has been made in CME. This considerable effort has urged agencies to assess its impact. The result was to look for CPD strategies that could be less focused on logic of education supply than on logic of individual and organization demand.

The very specific nature of practice in health professions and of its recent developments has also played an important catalytic role in its shift of orientation. First, the context of health service is changing. The multidisciplinary profile of today's health personnel (doctors, nurses, therapists, epidemiologists and sociologists, physical educators, dieticians, social workers, learning advisers, etc.) in clinics or hospitals, together with the increasing cooperation required from different departments (health, education, industry, sport, social affairs, etc.) raise questions about the monodisciplinary orientation of traditional CME and, more importantly, have put the focus on cooperation between these different content and agency inputs. Second, the relationships between health professionals and among health professionals and the public are changing. The health profession is relational in character; it is a profession of contact. At the core of their practice, health professionals have to relate directly with people coming from all social strata; they are working in close proximity with the general public. As such, this relational character is not new; the relation with patients has always characterized the practice of these professions. What has changed is the nature of these relations through external pressure to request more autonomy from the public, and through rising aspirations of patients to have their say and negotiate their life course, more so when it encounters health problems.

The reduction of time available in the relation between professionals and patients has tended to integrate health professions in a wider web of specialties. It has also created a demand for more capacity of initiative among a public having various levels of health literacy, with important consequences in daily practice of health professions, and therefore in CPD: necessity to acquire new relational skills and to practice new forms of cooperation with various agencies and professions. Relations between professionals are also changing to take into account social demand coming either from new social movements or simply from genuine aspirations of more literate patients eager to pilot their life, body, and soul. An interesting new phenomenon in this area is the emergence around typical sicknesses or pandemics of networks of patients and patients' relatives or peers who look for new information and review research reports in order to validate diagnoses, explore and suggest new therapies, and even discriminate, in given diagnoses, between scientific evidence and cultural bias. A good example is the role of gay movement supporting their members in their relations with health professionals more so at a time when research on AIDS

was proposing various curative procedures and some interpretations of AIDS occurrence amounted to discrimination (Epstein, 1995, 2000). The same can be said about the role of the women's movement questioning the handling of breast cancer (Lantz and Booth, 1998). Becoming credible participants in the process of knowledge construction, these movements, entering a realm traditionally restricted to the medical experts (Kleinman, 2000), bring about changes in the practices of biomedical research and of health professionals; they thus become complementary agencies of CPD.

The continuing education of health professionals, having to practice in such a changing social and scientific environment and being in daily intimate relations with a public also in transition, could not remain an updating unidirectional process. Hence, the shift in orientation of CME toward contextualized CPD was somewhat predictable. Of course, CPD of physicians and other health professionals is not without ambiguity. The interest of the biopharmaceutical industry to promote their brevetted drugs may explain their quasi-predominant role in nonobligatory CME (Relman, 2001). The vast resources involved in this new market tend also to create a tension between a logic of supply in the provision of courses and self-learning kits and the emerging logic, of a more reflexive nature, aiming to integrate more closely ongoing research with the specific concern of each practitioner, clinic, or hospital department.

Conclusion

A shift is emerging in national health policies that tend to balance their historically dominant remedial and prophylactic orientation with new concerns for health promotion and consideration of the health literacy dimension in public health strategies. Similarly, professional continuing development within health professions is reframing professional practices of physicians and transforming their initial and further education career. A timid but steady transition, not unrelated to the tension between functional and empowerment-driven health literacy, is thus taking place from the prevalent health welfare organization to a health participative and learnfare system.

More rapidly developed than in many other professional fields, CME is currently undergoing important changes that may be prototypal of similar developments taking place in other professions and occupations.

These two trends in health-related adult learning are indicative of the enlarged vision of adult learning policy that is currently taking place. Adult learning policy environment has indeed to be rethought and reconstructed; it can no more exclude indirect lifelong learning policies (Bélanger and Federighi, 2000), like the learning component of health policies described in this article.

The demand for health-related learning throughout one's life course both within health professions and among the general public may very well be, after work-related adult education and training, the strongest forces driving up the demand for adult learning in decades to come.

See also: Wider Benefits of Adult Education.

Bibliography

- Bartholomew, L., Parce, G., Kok, G., and Gottlieb, M. (2001). *Intervention Mapping: Designing Theory- and Evidence-Based Health Promotion Programs*. Toronto, ON: McGraw-Hill.
- Barton, D., Hamilton, M., and Ivanič, R. (2000). *Situated Literacies: Reading and Writing in Context*. London: Routledge.
- Bateman, H. and Kinmonth, A. L. (2001). Journeys and pathways: Exploring the role of professional development advice and educational guidance for practitioners expressing interest in research. *Medical Education* **35**(1), 49–55.
- Bélanger, P. and Federighi, P. (2000). *Unlocking People's Creative Forces: A Transnational Study of Adult Learning Policies*, 274p. Hamburg: UNESCO Institute for Education.
- Bero, L. A., Grilli, R., Grimshaw, J. M., et al. (1998). Closing the gap between research and practice: An overview of systematic reviews of interventions to promote the implementation of research findings. *British Medical Journal* **317**, 465–468.
- Bloom, B. S. (2005). Effects of continuing medical education on improving physician clinical care and patient health: A review of systematic reviews. *International Journal of Technology Assessment in Health Care* **21**, 380–385.
- Cervero, R. M. (2001). Continuing professional education in transition 1981–2000. *International Journal of Lifelong Education* **20**(1–2), 16–30.
- Cusick, A. and McCluskey, A. (2000). Becoming an evidence-based practitioner through professional development. *Australian Occupational Therapy Journal* **47**(4), 159–170.
- Davis, D., Barnes, B. E., and Fox, R. (eds.) (2003b). *Continuing Professional Development of Physicians: From Research to Practice*. Chicago, IL: American Medical Association.
- Davis, D., Evans, M., Jadad, A., Perrier, L., and Rath, D. (2003a). The case for knowledge translation: Shortening the journey from evidence to effect. *British Medical Journal* **327**(7405), 33–35.
- Davis, D., Thomson O'Brien, M., Freemantle, N., et al. (1999). Do conferences, workshops, rounds, and other traditional continuing education activities change physician behavior or health care outcomes? *Journal of the American Medical Association* **282**, 867–874.
- Davis, D., Thompson, M. A., Oxman, A. D., and Haynes, B. (1995). Changing physician performance: A systematic review of the effect of continuing medical education strategies. *Journal of the American Medical Association* **274**, 700–705.
- Downie, R. S., Fyfe, C., and Tannahill, A. (1999). *Health Promotion Models and Values*. Oxford: Oxford University Press.
- Du Boulay, C. (2000). From CME to CPD: Getting better at getting better? *British Medical Journal* **320**(7232), 393–394.
- Elwyn, G. and Hocking, P. (2000). Organisational development in general practice: Lessons from practice and professional development plans (PPDPs). *BMC Family Practice* **1**, doi:10.1186/1471-2296-1-2.
- Epstein, S. (1995). The construction of lay expertise: AIDS activism and the forging of credibility in the reform of clinical trials. *Science, Technology and Human Values* **20**(4), 408–437.
- European Commission (2004). *Enabling Good Health for All*. Brussels: European Commission.
- Feinstein, L. and Hammond, C. (2004). The contribution of adult learning to health and social capital. *Oxford Review of Education* **30**(2), 199–223.

- Gruman, J. C. (2003). The fragility of evidence: Why study health behavior now? *Paper Presented at the Bloomberg School of Public Health at John Hopkins University*. Johns Hopkins Bloomberg School of Public Health: Baltimore, MD, USA.
- Hohn, M. (2002). Literacy, health, and health literacy: State policy considerations. *NCSALL/Focus on Basics* 5(C), 20–25.
- Kleinman, D. L. (ed.) (2000). *Science, Technology, and Democracy*. Albany, NY: State University of New York Press.
- Kok, G., Schaalma, H., Ruiter, R., Van Empelen, P., and Brug, J. (2004). Intervention mapping: A protocol for applying health psychology theory to prevention programmes. *Journal of Health Psychology* 9(1), 85–98.
- Kok, G., van den Borne, B., and Mullen, P. D. (1997). Effectiveness of health education and health promotion: Meta-analyses of effect studies and determinants of effectiveness. *Patient Education and Counselling* 30, 19–27.
- ICAE (2001). *Agenda for the Future. Six Years Later ICAE Report*. Montevideo: ICAE.
- Institute of Medicine of the National Academies (2004). *Health Literacy: A Prescription to End Confusion*. Washington, DC: The National Academy Press.
- Kickbusch, I. S. (2001). Health literacy: Addressing the health and education divide. *Health Promotion International* 16(3), 289–297.
- James, K. (2001). *Prescription for Learning, Evaluation Reports*. Leicester: NIACE.
- Lantz, P. M. and Booth, K. M. (1998). The social construction of the breast cancer epidemic. *Social Science and Medicine* 46(7), 907–918.
- Livingstone, D. W. (1999). Exploring the icebergs of adult learning: Findings of the First Canadian Survey of informal learning practices. *Canadian Journal for the Study of Adult Education* 13(2), 49–72.
- Murray, S., Rudd, R., Kirsh, I., Yamamoto, K., and Grenier, S. (2007). *Health Literacy in Canada. Initial Results from NALS*. Ottawa, ON: Canadian Council on Lifelong Learning.
- Nuissl, E. and Pohl, A. (2004). *Portrait Continuing Education Germany*. Bonn: DIE.
- Nutbeam, D. (2000). Health literacy as a public health goal: A challenge for contemporary health education and communication strategies into the 21st century health. *Health Promotion International* 15, 259–267.
- Paasche-Orlow, M. K., Schillinger, D., Greene, S. M., and Wagner, E. H. (2006). How healthcare systems can begin to address the challenge of limited literacy. *Journal of General Internal Medicine* 21(8), 884–887.
- Peck, C., McCall, M., McLaren, B., and Rotem, T. (2000). Continuing medical education and continuing professional development: International comparisons. *British Medical Journal* 320, 432–435.
- Reichart, E. and Huntemann, H. (2006). *Volkshochschul-Statistik 2006*. Bonn: German Institute for Adult Education (DIE).
- Rudd, R., Kirsh, I., and Yamamoto, K. (2004). *Literacy and Health in America. Public Information Report*. Princeton, NJ: Educational Testing Service.
- Relman, A. S. (2001). Separating continuing medical education from pharmaceutical marketing. *Journal of the American Medical Association* 285, 2009–2012.
- Swedish National Council for Adult Education (2005). *Folkbildning of the Future, Its Role and Objectives*. Stockholm: SNCAE.
- Sentell, T. L. and Halpin, H. A. (2006). Importance of adult literacy in understanding health disparities. *Journal of General Internal Medicine* 21(8), 862–866.
- Sherhoff, M. (1997). Individual practice with gay men. In Mallon, G. P. (eds.) *Foundations of Social Work Practice with Lesbian and Gay Persons*, pp 77–103. New York: Haworth Press.
- Somnath, S. (2006). Improving literacy as a means to reducing health disparities. *Journal of General Internal Medicine* 21(8), 63–98.
- Thomas, D. C., Johnston, B., Dunn, K., et al. (2006). Continuing medical education, continuing professional development, and knowledge translation: Improving care of older patients by practicing physicians. *Journal of the American Geriatrics Society* 54(10), 1610–1618.
- UNESCO (1997). Health promotion and health education for adults. *Booklets 6b of the CONFINTEA V International Conference on Adult Education in Hamburg 1997*. Hamburg: UNESCO Institute for Education.
- WHO (World Health Organization) (2005b). Global health promotion scaling up for 2015 – a brief review of major impacts and developments over the past 20 years and challenges for 2015. *Unedited Working Paper from WHO Secretariat for the 6th Global Conference on Health Promotion in Bangkok, Thailand*.
- WHO (World Health Organization) (2005b). *Health for All: The Policy Framework for the WHO European Region*. Copenhagen: WHO.
- Zanchetta, M., Perreault, M., Kaszap, M., and Viens, C. (2007). Patterns in information strategies used by older men to understand and deal with prostate cancer: An application of the modelisation qualitative research design. *International Journal of Nursing Studies* 44(6), 961–972.
- Zorn, M., Allen, M. P., and Horowitz, A. (2004). *Understanding Health Literacy and Its Barriers. Bibliography*. Washington, DC: National Institute of Health.

Further Reading

- Ad Hoc Committee on Health Literacy. (1999). Health literacy: Report of the council on scientific affairs, American medical association. *Journal of the American Medical Association* 281, 552–557.
- Epstein, S. (1996). *Impure Science: AIDS, Activism, and the Politics of Knowledge*. Berkeley, CA: University of California Press.
- Hiramanek, N. (2005). Self directed learning and continuing medical education. *Australian Family Physician* 34(10), 879–880.
- Rootman, I. and Ronson, B. (2005). Literacy and health research in Canada: Where have we been and where should we go? *Revue canadienne de santé publique* 96(Supplement 2), S62–S77.
- Spiros, C. (2005). Health literacy: Concept analysis. *Journal of Advanced Nursing* 50(6), 633–640.
- Tang, K.-C., Beaglehole, R., and O'Byrne, D. (2005). Policy and partnership for health promotion – addressing the determinants of health. *Bulletin of the World Health Organization* 83(2), 884.
- WHO (World Health Organization) (2006). *Core Health Indicators: The Latest Data from Multiple WHO Sources*. <http://www.who.int> (accessed May 2009).

Modernization Processes and the Changing Function of Adult Learning

H S Olesen, Roskilde University, Roskilde, Denmark

© 2010 Elsevier Ltd. All rights reserved.

Education increasingly appears to be the natural form of learning in modern life within a contemporary Western mindset. However, learning is more than that. It is an integral aspect of human life, taking place all the time, necessarily and unnoticed. The societal functions of adult learning are accordingly multiple. The theme of this article is to look at the societal nature of adult learning and hence, the societal functions of adult education. We want to emphasize the historical dimension in the sense of linking adult education to the socioeconomic, political, and cultural context.

Adult Education and Modernization

We can distinguish three main types of adult education defined by their main content. These have developed as educational traditions in their own right, related to particular areas of learning. As such they have been described and critically analyzed elsewhere in the encyclopedia:

- Basic literacy education, such as reading and writing, with the aspect of cultural integration in the nation and state.
- Community and popular education.
- Education for work, such as continuing education and training.

Much of the recent discussion in adult education is a clash between educational cultures. On the one hand, there is a humanistic focus on personal and political self-articulation, which seems to be inherited from the traditional functions of community learning and liberal adult education. On the other hand, there is the instrumental perspective on lifelong learning for work, theoretically underpinned by human capital theory and similar frameworks of understanding. These ideological struggles refer to historical experiences, and so an understanding of the societal functions of adult learning in their contexts may bring them to bear on wider issues of contemporary society. The theoretical notion of modernization seems to be a productive backbone in understanding the multiple institutional realities, conceptual meanings, and historical changes of adult education and learning. Modernization, here, refers to the economic, social, and cultural changes which have taken place in the last 300–500 years, comprising the inclusion of feudal dynasties and independent city

republics in the melting pot of European nation-state building, as well as the imperial inclusion of cultures and countries in the Third and Fourth World that had been living separate from dynamic centers right until the great discoveries or later.

Capitalist economy has been the main motor in this modernization process, where traditional, self-sustaining local communities were included in larger societies, affecting all aspects of political, social, and cultural relations. The development of institutional (formal) education, replacing informal education and learning, is just one of these effects.

Adult education develops complementarily to this greater history of modernization and (formal) education, enabling individuals to deal with new societal realities. The very notions of adulthood and individuality result from this history, as a gradual and complex process creating the individual as a conscious agent in society has replaced the definition of adulthood by ceremonial inauguration.

There is a built-in risk in using the theoretical perspective of modernization framework. Seen from the dominant center of a global, modernized world, it may seem that adult education and learning is on hand to enable modernization, harmonize the levels of learning between generations, and live up to the accelerating needs for individuals to change. This may be a local truth of occidental modernization, where the efficiency and speed of knowledge transmission seems to be secured by institutional education. In order to avoid a narrow functionalist perspective, we must study adult education history with the perspective of discovering the multiple and infinite nature of this modernization process.

Literacy Education: Enabling Modern Societies

The original and most widespread understanding of literacy is related to reading and writing. Prototype literacy education has been engaged in making developing societies literate or in compensating for lack of adequate schooling in modern societies. If we look at it from a societal perspective, literacy is a precondition for citizenship and socioeconomic participation. In Europe, literacy has been closely related to the building of nation-states. The development of secular education activities and the emergence of literature in national languages were instrumental in the building of nation-states.

Today, literacy has become a political issue in the multi-cultural societies emerging in modern Europe, and in some cases, instrumental for minorities to establish home rule. However, the cultural integration of rural and urban working classes is also a condition for socioeconomic development. Today's industrial worker must be literate in order to fulfill simple work tasks because logistics and communication are built into every single task on the shop floor. Consequently, concern has moved from formal education to functional literacy – the competence to actually read and write in everyday life. New literacies are added, for example, numbers and mathematical modeling (Johansen and Wedege, 2002). Information and communication technology necessitates new reading and writing skills in order to be a competent member of society, but may also relativize the importance of traditional written language skills.

Literacy is empowering, but the acquisition of cultural techniques – whichever they are historically – also entails submission under cultural dominance. As a result of colonial modernization we can see derivatives of the great colonial empires, some united by the colonial language, and all to some extent influenced and shaped by colonial rule. Brazilian Paulo Freire answered this contradiction with a notion of political literacy he called conscientization, or learning to reflect the social reality and power relations involved in it. You may see Freire's ideas as congenial with mainstream modernist pedagogy, only related to the political learning process of those who are the victims and beneficiaries of a modernization coming to them from outside. In a more analytical key, questions have been raised about peripheral modernization, that is, a specific version of modernization in countries and regions that have had a modernization process pushed and influenced, but not entirely determined by colonial rule (e.g., Brazil).

Language is a medium for the individual sensual experience, as well as, for elaborating cultural experiences historically in a wider society. Written language has contributed significantly to societalization by enabling communication and knowledge transfer across time and space, creating a fundamental tension between immediate (local, situated) and mediated human experience. When printing technology created the industrial basis of the literate modernity, it reinforced this duality. The modern experience of the world is a language-mediated experience. Digital technology is probably proclaiming a new and more radical version of the same – or possibly a qualitative new relation to reality, an extended version of sensual access.

Community Education and Popular Education: Struggles about the Societal Atmosphere

All the different types of community and popular education are based in a community of people defined by

location, religion, cultural values, or political assumption, and often have a perspective of social, cultural, and political self-articulation. Most people engaged in such education probably perceive it as a free space for learning that is relatively independent of societal conditions and constraints. In a societal perspective however, adult learning is a substantial aspect of modernization itself. Community interests are very often defined by and responsive to societal change. Education preserves cultures that are threatened and overcome – and these pockets of social and cultural life that are not entirely penetrated by societal dynamics provide a productive space for self-articulation. Most typical community and popular education is probably based on resistance against some of the influences of modernization, for example, minority communities that also happened to be marginal and/or impoverished by capitalist modernization and centralization.

Independent of whether these communities see learning as explicitly political or not, learning is part of an attempt to create a public sphere of their own or set the cultural framework of understanding on a societal level. Out of closed communities or specific resistance grows – in a number of cases – a structural characteristic of modernity: the existence of civil society. Community and popular education based on specific histories of socioeconomic and cultural circumstances may – in very different and paradoxical ways – produce a general societal development. The history of independent Danish folk high schools provides a historical example. The folk high schools were based on and also contributed to a particular development of political, cultural, and social movement among Danish self-owning peasants, which in turn played a decisive role in establishing a modern democracy. In spite of the fact that the folk high school education was based on an antimodern, romantic ideology, it produced the experience of popular self-regulation and self-organization, which also in later developments contributed strongly to the specific modernization process in Denmark (Olesen, 1989). Its cultural-class compromise, uniting egalitarian and liberal principles and a basically anti-academic, informal notion of education, anticipated some of the developments of lifelong learning around the beginning of the twenty-first century.

The dialectic of local community learning and societal development is, more than anything, related to the role of work in culture and economy. Since the main driving force of modernization has been capitalist industrialization, the most important popular education activity in Europe in the previous century was of the labor movement and trade unions. Industry formed the life conditions in (urban) communities, and the labor movement in most countries organized working-class culture and its learning institutions, first as a resistance solidarity movement, then gradually as a more proactive cultural self-articulation and political movement. In some cases, community education also entails alternatives to the dominant capitalist economy – in the

form of cooperative economies – prominently in the Basque country and also at a smaller scale in most developed capitalist countries.

It is obvious that labor movement education activities are in one sense a product of an active resistance against some of the effects of modernization, similar to the culture of many communities that have been marginalized or impoverished by modernization. It is a clearly partial culture defined by political and trade union action, or at least from a general class perspective. Unlike many local and minority communities, however, labor movement education activities developed a universalistic perspective, challenging individualistic liberalism with ideas about equality and solidarity. The political struggles between different types of Socialism can be seen as different versions of a universalistic aspiration – communist, social democratic, anarchist, and syndicalist – that each carry in them more or less ambitious aspirations of justice and a new level of democracy beyond capitalism. The real histories of different countries have provided a variety of working-class cultures and not just labor movement experiences, as well as some in which universalist aspirations turned into totalitarian power.

Seen in a societal perspective, popular education comes out of premodern communities, as well as urban communities generated by modernization itself. They are formed by the specific histories of modernization and they take advantage of one of the effects of modernization, namely the existence of a space for noncoerced social organization, what we in modernization theory call civil society. The societalization – from *Gemeinschaftswesen* to *Gesellschaft* in the notions of the classical sociologists – eradicates or restructures communities at the same time as modernization extends the space of relatively free cultural activity. The juridical basic rights and economic affluence in modernized parts of the world create the societal basis of this civil society. A contemporary form of popular education must have its potential base in the organization of citizenship to deal with societal issues on a level and with an outlook adequate for contemporary society as a whole. Many one-cause movements and actions have a similar profile as community action of resistance or opposition though not necessarily founded in a community.

The bourgeois public sphere as the communicative framework of a fully developed civil society corresponds largely with the nation-state and ideas of formal (state) democracy. At the same time, global capitalism has in several ways bypassed this structure. Structurally, by its international operation and concentrations of power and capital in organizations larger than many states, creating a democratic deficit. Culturally, by the media and consumer cultures which take active part in the shaping of desires, fantasies, and preferences. One may see international labor organizations and forms of organizations like World Social Forum as – very fragile – civil society responses to this situation.

Learning for Work, or Human Resource Development

In the last couple of decades, the need for work competence building has tended to prevail over traditional forms and rationales of adult education and training in developed Western societies. In traditional societies, intergenerational transfer of knowledge and competences enables the reproduction of the labor force. Modernization has brought basic school education and some specialized institutions for academic and professional education – generally serving as basic, lifelong qualifications in the initial career. In advanced capitalist societies, this mode of transfer has increasingly come under pressure. When changes in work and labor market happen faster than the generational turnover in the labor force, adult education and training come in as a mediating instrument to secure the adaptation of labor to the requirements of work life.

It has mostly been left to employers and the individual worker to take care of retraining and adult education. The consequences have typically been a general market failure (i.e., underinvestment) and a very unequal distribution of training resources. Big industrial employers have in some cases been able to secure up-skilling of their own employees, but not mobility across sectors and competence levels. In Nordic countries, training of workers became part of welfare state policies. In Denmark, with its late development of industrial economy, urban industries needed skilled workers, and rural workers needed skills and socialization for industry. A separate new strand in adult education developed to support a rapid migration from rural to urban life, from agriculture to industry. What became later known as the flexicurity model (Jørgensen and Madsen, 2007) consisted mainly of adult education and training together with relatively good unemployment benefits. This combination enabled a more proactive policy from trade unions and employees than in a number of European countries in periods of economic growth. In the period of crisis and stagflation in the 1970s, continuing education was redirected/enhanced to take care of more long-term competence development for the more vulnerable segments of the labor force (e.g., women, young people without vocational qualification, and others).

Recently, adult learning seems to have assumed a more universal or all-embracing nature in all the advanced capitalist countries. As long as the development of work takes the form of strong division of labor based on mass unskilled wage labor, societal needs remain limited to training and retraining specialists and highly skilled craftspeople. However, with the development of postindustrial forms of work organization, a need for broader adult education emerges. The societal demand for knowledge economy has changed to include not only what were mostly called soft skills (e.g., communicative and collaborative skills, quality consciousness, professional attitudes,

and self-confidence) but also traditional literacy, as well as new literacies (e.g., numeracy and mathematic understanding, and computer literacy). Work-related learning seems to become broader and deeper and increasingly interferes with personal needs and identity (Olesen, 2005). Nevertheless, it is obvious in the rhetoric of lifelong learning that economic concerns and the focus on employment and work are determining factors. This can be seen as a very local view on global development. The position of most developed economies can hope to maintain their relative competitive advantage in a division of labor where they take care of knowledge-based, complex work and the service work for themselves, whereas developing countries deliver raw materials and build up low-tech industrial production.

The political consensus about lifelong learning of competences may not be so easy to maintain in this narrow key. Rather, the focus on work and human resource development may raise issues of control and the quality of work. The ideas of a knowledge-based economy have been criticized from several perspectives. One applies a wider, ecological perspective on work and learning, questioning the inward colonialism of human life without boundaries (Hochschild, 1997) and its cultural consequences (Sennett, 1998; Negt, 1984). The requirements on human flexibility and adaptation may erode the conditions of socialization and subjectivity, that is, the human resources on the whole. Another perspective emphasizes the direct political aspect of learning in which labor movements should take the opportunity to advance a politicization of work, including environmental questions, ownership, and use value of production, drawing on vanguard experiences of cooperative enterprises (e.g., The Mondragon cooperative – Antoni and Campbell, 1983), projects for conversion of production (Lucas Aerospace and others), and a vision of self-regulated work (Forrester, 2007). The dramatic emergence of the climate crisis and the fragility of the capitalist world economy underscore the need for more comprehensive perspectives on work and learning.

It seems most plausible to outline a neoliberal scenario of an individualized competence market, which will be subsumed into a global labor market. However, it also seems likely that this competence market will show an unprecedented example of market failure – and it will definitely have extreme effects in terms of inequality and the colonization of human labor. The question is whether there is another scenario in which the significance of the labor force as a subjective factor in the economy can be turned into individual and collective self-regulation of work and learning. This seems to be the open question that places the discussion about learning for work and the workers' role in the development of work as a central issue in global politics.

The resources for any alternative to neoliberal global capitalism must to some extent be found in institutional

practices, embodied experiences of the past, social organizations, and experiences of trade unions and other cultural organizations. They are present in the forms and levels of education, expectations, and preferences of young people as well as adults, but they do not form a simple and coherent alternative. While the new discourses of lifelong learning are international, Anglophone, and relatively homogenous, adult-education traditions have many names: popular education, community education, *educacao popular*, *politische bildung*, liberal education, *folkeoplysning*, *folkbildning*, *formation des adultes*, *formazione popolare*, *volksbildung* and citizenship education to name a few. In adult education discussions, these many names give rise to translation problems – although the names in different languages cover more or less corresponding phenomena, they do not have the same meaning because meaning is related to societal and cultural context.

Functions of Adult Learning in Historical Context

The exploration of the historical function of adult learning may lead to an open discussion of present-day modernization. The main types of adult education form strands of historical functionality over long periods of time. Analyzing them in terms of modernization may present a simplistic scheme: literacy enables modernization by integrating ordinary people in cultural communities that are independent of time and space; popular education elevates the ability of communities to reproduce themselves to a level of cultural construction and self-articulation, eventually a level of collective self-assertion and political activism; and continuing education and training for work aligns individual competence-building cycles with an accelerated and distributed organization of societal work in global capitalism. Obviously this complex societal evolution is not a unified process – it is asynchronous and very diverse across the world. So the societal function of adult education must be studied in concrete contexts, regarding the interplay with socioeconomic, political, and cultural history. Socioeconomic modernization and the type of learning needed and enabled are mutually interrelated, but institutional developments also set general conditions of this process. When and what type of adult education contributes to modernization (the level of general schooling and the influence on school by church and class movements), and political circumstances, may form very specific conditions and challenges (e.g., the *entnazifizierung* as a project for political education in Germany after World War II, or the modernization in Spain under a long-lasting dictatorship).

Today, on a global scale, we may ask whether modernization is just one process. The discussion about

peripheral modernization has been touched upon in this article. In postcolonial theory and political discussion, the emphasis on difference and multiple histories serves to demonstrate the overcoming of the modernist tale or vision of a rational evolution toward a better society. Sometimes the argument that modernization should not be seen as a continuous progress gets confused with the assumption that there was and is no modernization process at all. In a generalized discussion of global capitalism one may at the same time observe a pessimistic view of one culturally homogenized world (McDonaldization) and much more relativistic postcolonial theory of a general dissolution of the modernization process in a firework of difference. Both fail to grasp the complexity of durable social changes. Having said that, what is the role of learning in modernization processes? It can be argued and underpinned by examples that broad processes of learning and democratic participation lead to durable societal changes, whereas change processes that fail to engage the broader population remain unstable and produce conflicts in the form of imperialism and/or violent repression on macro- and micro-levels. The postcolonial discussion of education may, in the era of global capitalism, oscillate between the critical perspective of one culturally homogenized world at the end of the story (McDonaldization), and the postmodernist theory of a general dissolution of the modernization process in a firework of difference. There is a key to dynamic mediation between the two perspectives in an open-minded exploration of specific experiences and learning resources in their historical context.

It is essential to maintain that the possible futures of late- or postmodernity is a matter of social agency and hence also of learning processes taking place now. Lifelong learning implies a new discourse that brings learning beyond institutional education and into social reality. However, the lifelong learning discourse has been heavily influenced by neoliberal politics and human capital theory, and the world is relatively short of alternative ideas that can embrace the critique of educational institutions without accepting the neoliberal economic rationale. The options available are societal and subjectively relevant conditioned by experiences and resources of the past. The author has mentioned a long-lasting influence of the popular education tradition in his own country, which formed the ideological basis for a relatively democratic and liberal provision of formal and higher education, as well as adult education. Others are to be found in Latin America, South Africa, and other particular pathways of modernization.

Modernization is still an uncompleted development even in its original centers. While it seems that globalization brings forward further homogeneity, there are also factors that tend to enable a multicentered and polyphonic global world. The fact that China has had its

own almost independent cultural and social pathway, which is now – forcefully – joining global capitalism, forms an exciting experiment for the relation between human socialization and societal development. Oskar Negt calls it “the greatest social experiment in our time” (Negt, 1988/2007) in his discussion of the modernization(s) in China in the perspective of European modernization since the Renaissance. We may most productively see modernization as an infinite process that is still dependent on human efforts and choices on individual, as well as global level.

See also: Adult Education and Civil Society; Adult Literacy Education; Community Based Adult Education; Lifelong Learning; Popular Adult Education; The Political Economy of Adult Education.

Bibliography

- Antoni, A. and Campbell, A. (1983). *The Co-Operative Way: Worker Co-Ops in France, Spain and Eastern Europe* vol. viii, 75 pp. London: ICOM Co-Publications.
- Forrester, K. (2007). Becoming visible? Notes on work, union learning and education. In Shelly, S. and Calveley, M. (eds.) *Learning with Trade Unions: A Contemporary Agenda in Employment Relations*. Aldershot: Ashgate Publishing.
- Hochschild, A. R. (1997). *The Time Bind: When Work Becomes Home and Home Becomes Work*. New York: Metropolitan.
- Johansen, L. Ø. and Wedege, T. (eds.) (2002). *Numeracy for Empowerment and Democracy? Adults Learning Mathematics* Research Forum ISBN 87-7349-540-9.
- Jorgensen, H. and Madsen, P. K. (eds.) (2007). *Flexicurity and Beyond: Finding a New Agenda for the European Social Model*. Copenhagen: DJØF Publishing.
- Negt, O. (1984). *Lebendige Arbeit – enteignete Zeit*. New York: Campus Verlag.
- Negt, O. (1988). *Modernisierung im Zeichen des Drachens. China und der Europäischen Mythos der Moderne*, 2nd edn. Frankfurt a.M. Fischer: Steidl.
- Olesen, H. S. (1989). *Adult Education and Everyday Life*. Roskilde: Roskilde University.
- Olesen, H. S. (2005). Work related learning, identities, and culture: New work – new genders? New genders – new work?. In Bron, A., Kurantowicz, E., Salling Olesen, H., and West, L. (eds.) *'Old' and 'New' Worlds of Adult Learning*, pp 27–41. Wrocław: Wydawnictwo Naukowe.
- Sennett, R. (1998). *The Corrosion of Character*. New York/London: Norton.

Further Reading

- Elias, N. (1976). *Über den Prozess der Zivilisation* (orig. publ. (1939); 1994, *The Civilizing Process*. Oxford: Blackwell). Frankfurt am Main: Suhrkamp.
- Feidel-Mertz, H. (1975). *Erwachsenenbildung seit 1945: Ausgangsbedingungen und Entwicklungstendenzen in der Bundesrepublik*. Köln: Kiepenheuer und Witsch.
- Freire, P. (1970). *Pedagogy of the Oppressed*. New York: Seabury Press.
- Negt, O. and Kluge, A. (1981). *Geschichte und Eigensinn*. Frankfurt am Main: Zweitausendeins.

Olesen, H. S. (2006). Beyond the abstractions? Adult Education research from idealism to critical social science. *International Journal of Lifelong Education* 25(3), 241–256.

Olesen, H. S. (2006). Experience and Learning. In Billett, S., Fenwick, T., and Somerville, M. (eds.) *Work, Subjectivity and Learning*. Zürich: Springer.

Steven (2006). The fight for useful work at Lucas Aerospace. <http://libcom.org/history/1976-the-fight-for-useful-work-at-lucas-aerospace> (accessed June 2009).

Wallerstein, I. (2000). *The Essential Wallerstein*. New York: New Press.

The Political Economy of Adult Education

G Rees, Cardiff University, Cardiff, UK

© 2010 Elsevier Ltd. All rights reserved.

The expansion of post-compulsory education and training has been one of the most striking recent changes in the education systems of more developed societies. Much of this expansion is accounted for by rising levels of participation among adults. Of particular interest here has been the growing prominence attached to the provision of learning opportunities for individuals throughout their working lives and, indeed, into retirement.

In part, this reflects the demographic shifts which imply an aging population – and workforce – in most of the more developed countries. Equally, there has been a fundamental political re-evaluation of adult education's role. Governments in these societies (and beyond) have been remarkably consistent in prioritizing policies for what has come to be termed lifelong learning; that is, learning throughout the life course, from preschool to old age. These policies, in turn, are based on a robust conventional wisdom – what Grubb and Lazerson (2004) have dubbed the education gospel – about the essential role of education and training in generating the high levels of skills necessary for economic competitiveness and growth in the globalized economy.

This conventional wisdom partly derives from the work of academic economists in developing human capital theory (HCT). This theory has become one of the most influential economic doctrines of the age, especially in its contemporary manifestation in theories of the knowledge-based economy (KBE). However, the central characteristic of the political economy approach is that it interrogates such economic doctrines in order to uncover their political and sociological premises. Rather than simply treating economic theories as frameworks of analysis in their own terms, political economy aims to explore the – often implicit – social assumptions on which they are based. In this case, for example, approaches to adult education derived from HCT assume that individuals are able to calculate the instrumental benefits of human capital formation; and that individuals' skills are rewarded in ways which reflect their productivity. Alternative approaches are clearly possible.

Equally, the political economy approach seeks to explain the influence which economic theories exert on policy in terms of the wider social and political context in which they are located. Here, the increasing dominance of neoliberalism during the 1980s and 1990s, orchestrated by international agencies, such as the World Bank and the Organization for Economic Co-Operation and Development (OECD), created a fruitful ideological environment

for the development of the individualized and instrumental approaches to adult education which are embodied in lifelong learning strategies. While governments adopted policies which reflected national circumstances, the policy repertoire available to them was limited by the wider ideological climate.

The next section presents an account of how the current conventional wisdom views the relationships between adult education and HCT, paying particular attention to theories of the KBE. Then, the critique from political economy of this position is explained. In the final section, the significance of the ideological context of neoliberalism for state policies on adult education is reviewed.

Theories of Adult Education as Human Capital Formation

The analytical basis for the view that, in general, education has a key role in determining economic development is provided by HCT, although the versions adopted by governments to underpin their policies have involved highly simplified accounts of the academic research. HCT has a very long pedigree, but its influence on thinking about the relationships between education and economic life became especially marked during the 1960s. Here, the macroeconomic arguments that investment in education and technological change constituted the basis for national economic growth provided a rationale for the expansion of educational provision by governments, especially at the upper secondary and tertiary levels.

In the harsher economic climate of the 1970s and 1980s, the emphasis shifted to the returns (e.g., from enhanced future earnings) which would flow to the individual from his/her investment in increased human capital (e.g., through undertaking extra education and training); and the productivity gains which would accrue to employers from the resulting enhanced skills levels. State intervention became limited to ensuring that the optimal conditions for these private individual investments prevailed. By the 1990s, the intellectual terrain had shifted again, strongly influenced by the development of theories of the KBE.

Human Capital in the KBE

Theories of the KBE adapt HCT to what are construed as the essential characteristics of contemporary economic life.

On this view, the globalization of trade and investment has intensified competitive pressures, especially for those national economies which are unable to compete on the basis of low-cost production. In these economies, competitiveness can be maintained only by enhancing the productivity of capital and labor. This implies continuous innovation, developing new goods and services and more effective processes for producing, marketing and distributing them, and applying the most efficient patterns of work organization. Hence, the creation and – critically – the application in production of new forms of knowledge become the key factor in shaping the prosperity of individual enterprises and economic development more widely.

This valorization of knowledge production and innovation also has critical implications for the operation of labor markets. The KBE offers expanding opportunities for the so-called knowledge workers who have the very high-level skills required to operate in professional, managerial, scientific, and creative jobs in the upper echelons of the new occupational hierarchy. Equally, even workers who do not attain the very highest occupational levels require, at a minimum, high-quality general education and cutting-edge intermediate skills, so that they can engage actively in driving up productivity.

However, the corollary of these new occupational opportunities is the sharp decrease in unskilled or partly skilled occupations, which formerly comprised significant segments of the workforce. Some of these occupations have been relocated to the low-cost economies, reflecting patterns of globalization more widely. Others have simply disappeared because of changes in the industrial structure and transformations in product and process technologies. The implication is, therefore, that, in the KBE, employment opportunities for those individuals who do not have significant amounts of human capital are bleak.

Adult Education and the KBE

Theories of the KBE have significant consequences for how the production of human capital is conceived. Effective systems need to be in place to ensure that individuals with the requisite high-level skills are produced in sufficient numbers to meet the growing demand for knowledge workers. This implies the expansion of tertiary education, increasing the numbers of people with degrees and postgraduate qualifications. Equally, through the compulsory phases of education, everyone needs to attain high levels of general education – especially in respect of the essentials of literacy, numeracy, and science.

Moreover, theories of the KBE have specific implications for human capital formation through adult education. In an economy that is characterized by continuous innovation and the adoption of new technologies and modes of working, individuals need the forms of human

capital which ensure that they are able to respond flexibly to these changing demands.

This requires the provision of opportunities for adults already in the workforce to compensate for any shortcomings in their previous education, in terms, for example, of the essential basic skills of literacy and numeracy or general vocational skills, such as communication, team working, and information and communication technology (ICT) competences. It also implies that workers are enabled to develop new knowledge and skills, as and when these become necessary. Not only are individuals expected to transfer from one job to another much more frequently than hitherto, but also new demands arise as existing occupations change in character. In these circumstances, workers need to be competent and flexible learners; learning to learn is thus an essential prerequisite.

Adult education is required to deliver effective opportunities for individual workers to acquire new knowledge and skills not only through participation in formal learning within institutions dedicated to the provision of education and training, but also through nonformal learning effected, in particular, within the workplace itself and largely instigated by employers. The workplace is also increasingly recognized as a crucial setting in which informal learning, especially through interaction with colleagues, should be facilitated.

It is, of course, important to emphasize that these theories of the KBE – as with most theories in the social sciences – are not without their critics. In the next section, some of these opposing views, drawing on political economy approaches, are explored.

Beyond Theories of Adult Education and Human Capital

Adult Education, the Social Construction of Skills and Learner Identities

In general terms, the link between human capital formation and economic development is well attested empirically. At the macroeconomic level, innumerable econometric studies demonstrate the strong positive relationship between education and national economic growth in the long term (Barro, 1997). There are also many microeconomic studies which show the returns which accrue to individuals from their investments in human capital formation.

Much of this evidence relates to school-based and tertiary education and qualifications. However, the available studies indicate that work-related adult education (and basic skills development in particular) does have positive associations with the earnings of individuals and with productivity increases (although the returns to other forms of adult education are less certain) (OECD, 1999). Accordingly, on the basis of this evidence – wholly consistent with theories of the KBE and the consequent understanding of

adult education's role – both individual workers and their employers benefit from greater participation in work-related adult education (Blöndal *et al.*, 2002).

Approaches to adult education derived from HCT assume that individuals and employers recognize that increased human capital will increase productivity and that this will be rewarded in higher wages. However, as many commentators have argued, skills and the returns that accrue to them cannot be understood simply in terms of productivity. How skills are defined is the product of complex social and political processes, which are frequently as much about excluding social groups from the rewards which derive from what are defined as highly skilled occupations, as the technical requirements of production. Most notoriously, the skills attached to jobs which are characteristically done by women have not been recognized as such by employers and, indeed, male workers and their trade unions. While the knowledge and competences embodied in occupations may have changed, skills remain socially constructed. Moreover, once the notion of skill is detached from its technical content, it becomes clear that human capital may best be understood in terms of regulating access to occupations of different kinds, and, in the case of work-related adult education, to the kinds of screening which employers operate to determine what kinds of people enter more desirable jobs.

Equally, it cannot be inferred from the evidence on returns to human capital formation alone that individuals will actually participate in work-related adult education. Hence, approaches derived from HCT assume that individuals accrue human capital (e.g., by participating in work-related adult education) on the basis that future returns (e.g., through increased earnings) will outweigh the costs incurred. In reality, however, people's participation even in work-related adult education reflects more than this narrowly constructed, instrumental valuation. For many forms of work-related training, people choose to take part only in a very limited sense. Participation is experienced as compulsion by the employer; and may, in consequence, result in very little by way of new learning. Even where participation is a matter of choice, individuals' decisions are constrained by their material conditions: their financial resources, family commitments, and even where they live. Moreover, people's choices are shaped by the expectations and norms – deriving from their experiences in their families, communities, and previous educational institutions – which define their identities as learners and, therefore, potential participants in adult education of all kinds (Rees *et al.*, 2006).

Adult Education and the Demand for Skills

Theories of the KBE predict increasing demand for skills across the economy as a whole, largely through the growth

of high-skills jobs. However, while there has been a significant increase in the professional, managerial, scientific, and creative occupations filled by highly skilled knowledge workers, this has not been matched by the demise of low-skilled, low-waged employment. Many more developed economies have experienced an increase in jobs at both the top and the bottom of the occupational hierarchy, with a hollowing out of occupations at the intermediate levels. To this extent, therefore, a substantial and even growing proportion of occupations do not require high levels of skill, either to enter them or to carry them out effectively.

While there may be persistent shortages of specific skills, there is no convincing evidence that there are general shortfalls in skills across the more developed economies as a whole. This is the case even in those national economies in which human capital formation has been regarded as problematic. For example, recent studies in the UK suggest that, if occupations are categorized by broad level of required skills, the only category for which there are more job openings than people qualified at the appropriate level is that where no formal qualifications are required. Conversely, the number of jobs requiring tertiary-level qualifications is far exceeded by the number of people having such qualifications (Felstead *et al.*, 2007). Evidence of this kind has led to the claim that, in many of the more developed economies, demand for skills, especially at higher levels, is being outstripped by supply.

Work-related adult education is not provided across the board in the ways prescribed in theories of the KBE. Access to such learning opportunities is not distributed evenly between different groups of workers. In general, the higher someone is in the occupational hierarchy and the greater his/her previous educational attainment, the more likely they are to benefit from the provision of work-related adult education. As **Table 1** indicates, while there are important national differences in rates of participation, this is the pattern across the more developed economies; and it reinforces the notion that significant parts of the labor market remain dominated by low-skill occupations.

Indeed, it has been argued that some economies are characterized by what Finegold and Soskice (1988) called a low-skills equilibrium. Here, employers adopt strategies geared to the production of low-cost, low-quality goods and services. These strategies are sustainable because production is primarily or wholly for local or domestic markets with little threat from cheap imports. Employers rely upon labor-intensive production systems and the mass production of low value-added goods. Therefore, they do not require workers with high skills levels. While there is little evidence to suggest that a low-skills equilibrium exists across any of the more developed economies as a whole, it is apparent that particular sectors and even regions can be dominated by low-skills, low-cost

Table 1 Participation in work-related adult education and previous education^a

<i>Level of education</i>	<i>Lower secondary</i>	<i>Upper secondary</i>	<i>Tertiary</i>	<i>All levels</i>	<i>Ratio (%) hours in training to annual hours worked</i>
Canada	6	20	35	25	30
Denmark	22	36	54	39	63
France	9	19	33	19	49
Germany	3	10	24	12	28
Italy	1	6	12	4	5
Netherlands	5	11	13	6	21
Sweden	24	37	57	40	40
Switzerland	8	27	44	29	46
United Kingdom	7	26	46	27	19
United States	12	32	56	37	26
OECD	7	17	31	18	25

^aAnnual participation rate in nonformal, job-related education and training for 25–64-year olds (2003). From OECD (2007). *Education at a Glance 2007*. Paris: OECD.

production, often reflecting a residue of past economic activity (OECD, 2001).

The development and application of skills are situated within the wider strategies which firms and other organizations adopt with respect to the production, marketing, and distribution of goods and services. Moreover, there is not a single strategy that provides employers with a pathway to competitiveness and profitability. Hence, simply expanding the supply of skills is insufficient, unless employers' production strategies are of a kind which ensures that extra skills can actually be applied. This has important implications for the kinds of skills development strategies adopted by governments.

In spite of these well-founded critiques, the KBE continues to provide the foundation for the very robust conventional wisdom as to how governments across the more developed countries should deal with adult education. In the next section, the role of an ideological context dominated by neoliberalism in explaining this seeming paradox is addressed.

The State and the KBE

In the official discourses of governments and international organizations, theories of the KBE provide the key justification for policies aimed at promoting higher levels of skills through lifelong learning. While the latter encompasses learning throughout the life course, adult education has been identified as an especially important mechanism for enhancing human capital formation. Hence, many national governments in the more developed countries have shifted priorities for adult education toward directly vocational provision, thereby displacing its contributions to traditional concerns such as personal fulfillment or cultural development.

So robust is this consensus on the need to gear state policies to promote high skills levels that it has become

the common sense of policy on adult education across the more developed economies. Accordingly, it is important to emphasize that this current conventional wisdom is situated within a historically specific ideological and institutional context.

Neoliberalism and Lifelong Learning

For much of the twentieth century, the governments of the developed economies operated Keynesian national settlements between employers, trade unions, and government. These were based on the mass production of standardized goods and services, high wage levels, and, consequently, rising living standards as well as an emergent consumer culture. The state played a key role in regulating economic activity and in providing key welfare services. While approaches to skills development varied significantly between different national economies, HCT was interpreted as providing the rationale for increased government investment in education, especially at the upper secondary and tertiary levels, in order to keep pace with the growing technological complexity of goods and services; but adult education was not given special attention.

This Keynesian regime broke down during the 1970s, to be replaced by one based broadly on neoliberalism. On this view, in the face of new patterns of globalized economic activity and competitive pressures, markets are deemed to provide the most effective mechanisms for the organization of economic activity. The role of government is limited to ensuring the conditions within which markets can function effectively, through maintaining macro-economic stability and providing public goods (physical infrastructure, basic education, environmental protection, etc.). Only where markets manifestly fail to deliver effective outcomes should state provision play a major role, as, for example, in meeting the basic needs of the poorest members of society through income redistribution.

Therefore, although the development of high skills levels through lifelong learning is acknowledged as central to meeting the demands of global competition and the KBE, the state's role in skills formation is a restricted one. It should ensure that the supply of basic education through schools, colleges, and universities is maximally effective. However, employers can be relied upon to encourage the development of vocational skills in response to the changing patterns of demand, and to reflect the true value of such skills in wage levels.

Most importantly, the principal responsibility for human capital formation is seen to lie with the individual worker. Intervention by government should be restricted to ensuring the availability of information, removing barriers to participation, and, where the market fails, subsidizing individuals' engagement in work-related adult education. The very language of lifelong learning implies that its effectiveness in creating and preserving human capital depends upon the capacities and motivations of individuals themselves. Failure in the labor market is interpreted in terms of the decisions that people make about improving their employability through increased human capital; and increasingly, welfare payments to the unemployed and other economically inactive people are conditional on their participation in education and training.

In this way, neoliberal approaches emphasize the key importance of human capital formation. However, they prescribe particular means of achieving this, emphasizing individualized and instrumental orientations to lifelong learning and work-related adult education, more specifically.

Government Strategies for Adult Education

The work of international organizations, such as the World Bank and the OECD, has been a key influence in shaping this neoliberal agenda on lifelong learning. More specifically, such organizations have been crucial in ensuring their significance in shaping the actual roles undertaken by governments. The work of the OECD has been especially important here and illustrates this agenda very clearly (Henry *et al.*, 2001).

From the late 1980s onward, the OECD has produced highly influential analyses of the essential role of human capital in the development of the KBE (e.g., OECD, 1989). It has also been an enthusiastic advocate of the necessary benefits of lifelong learning and, more specifically, of enhancing the human capital of adults. Moreover, it has been crucial in specifying the exact roles to be played by governments in ensuring effective provision of adult education. Workers require access to ongoing opportunities for learning of all kinds. Where firms and other organizations are unwilling to provide these opportunities, the state needs to intervene to ensure that such market failure is rectified, through, for example, providing subsidies for

employer-led provision, specifying mandatory spending on training by enterprises or granting the right to be trained to employees (OECD, 2002).

However, despite the almost hegemonic influence of neoliberalism, there remain significant differences in the strategies adopted by national governments toward adult learning. At the most fundamental level, this reflects the choices made by governments. For example, as has been seen earlier, the high-skills strategy is not the only possible response to the competitive pressures of a globalized economy. One clear alternative is, in effect, to adapt the strategies of the Keynesian era to changed circumstances, through an intensification of Fordist production of goods and services, emphasizing low-cost and, to a considerable extent, low-skill methods. This kind of low-skills strategy has not been adopted explicitly by any of the governments of the more developed economies, reflecting both competitive pressures from lower-cost economies as well as the ideological influence of the neoliberal consensus on the KBE. Nevertheless, it remains the case that, in reality, a low-skills equilibrium is characteristic of substantial parts of all of these economies.

The choices which governments make, in turn, are shaped by the skills development strategies which have been characteristic of national societies in the past. For example, the enthusiasm with which the Anglo-Saxon economies (the USA, the UK, Australia, and New Zealand) have embraced market-driven approaches reflects their voluntaristic traditions, as well as an ideological willingness to accept sharp inequalities between enclaves of highly skilled knowledge work and large numbers of low-skilled and low-waged jobs. In contrast, many of the northern European governments (e.g., in Germany, the Netherlands, and the Scandinavian countries) have based their adult education strategies much more on consensus between employers, employees, and government. This reflects both their historical attachment to this form of political economy and a rooted ideological commitment to avoid the worst social inequalities of a market-driven system through state support for greater employment security and access to learning opportunities among the socially disadvantaged. In developmental states, such as Japan, Korea, and Singapore, circumstances have been different again. Here, the state has continued to play a significant role in skills development, simply to try to ensure that the supply of skills matches the demands of economies undergoing strong growth (Brown *et al.*, 2001).

The Value of Adult Education

Despite these important differences in approach, it remains the case that the hegemonic discourse on adult education is currently dominated by instrumental orientations. Its value is seen to lie overwhelmingly in the

contribution which it makes to economic life. In distributional terms, investment in human capital formation through lifelong learning is argued to benefit not only the individual workers and employers making the investments, but also the wider economy through increased productivity, thereby warranting – limited – state support. Moreover, it is the recognition of these individual and social benefits that are viewed as underpinning individuals' motivations to invest in human capital formation in the first place.

Adopting a wider frame of reference, however, suggests alternative frameworks through which the value of adult education may be judged. In the right circumstances, engagement in lifelong learning certainly offers the possibility of significant economic returns, both to the individual and to the economy more widely. However, as many commentators have argued, this by no means exhausts adult education's potential total public value. Ironically, however, realizing this total public value may require governments to reassert priorities for adult education which have declined in significance over recent decades, reviving their role as provider of a liberal general education, especially to those social groups who need an educational second chance.

See also: Economic Outcomes of Adult Education and Training; Human Capital; Lifelong Learning; Neoliberalism, the Market and Performativity; Participation in Adult Learning; Wider Benefits of Adult Education.

Bibliography

- Barro, R. J. (1997). *Determinants of Economic Growth: A Cross-Country Empirical Study*. Cambridge, MA: MIT Press.
- Blöndal, S., Field, S., and Girouard, N. (2002). *Investment in Human Capital through Post-Compulsory Education and Training: Selected Efficiency and Equity Aspects*. Economics Working Paper No. 333. Paris: OECD.
- Brown, P., Green, A., and Lauder, H. (2001). *High Skills: Globalization, Competitiveness, and Skill Formation*. Oxford: Oxford University Press.
- Felstead, A., Gallie, D., Green, F., and Zhou, Y. (2007). *Skills at Work, 1986–2006*. Cardiff: SKOPE.
- Finegold, D. and Soskice, D. (1988). The failure of training in Britain: Analysis and prescription. *Oxford Review of Economic Policy* 4, 21–53.
- Grubb, W. N. and Lazerson, M. (2004). *The Education Gospel: The Economic Power of Schooling*. Cambridge, MA: Harvard University Press.
- Henry, M., Lingard, B., Rizvi, F., and Taylor, S. (2001). *The OECD, Globalisation and Education Policy*. Oxford: Pergamon.
- OECD (1989). *Education and the Economy in a Changing Society*. Paris: OECD.
- OECD (1999). *OECD Employment Outlook*. OECD: Paris.
- OECD (2001). *Cities and Regions in the New Learning Economy*. Paris: OECD.
- OECD (2002). *Beyond Rhetoric: Adult Learning Policies and Practices*. Paris: OECD.
- OECD (2007). *Education at a Glance 2007*. Paris: OECD.
- Rees, G., Fevre, R., Furlong, J., and Gorard, S. (2006). History, biography and place in the learning society: Towards a sociology of life-long learning. In Lauder, H., Brown, P., Dillabough, J.-A., and Halsey, A.H. (eds.) *Education, Globalization and Social Change*, pp 926–935. Oxford: Oxford University Press.

Further Reading

- Becker, G. S. (1993). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. Chicago, IL: University of Chicago Press.
- Bélanger, P. and Valdivielso, S. (eds.) (1997). *The Emergence of Learning Societies: Who Participates in Adult Learning?* Oxford: Pergamon.
- Bell, D. (1973). *The Coming of Post-Industrial Society*. New York: Basic Books.
- Crouch, C., Finegold, D., and Sako, M. (1999). *Are Skills the Answer? The Political Economy of Skill Creation in Advanced Industrial Countries*. Oxford: Oxford University Press.
- Dore, R. (1997). *The Diploma Disease: Education, Qualification and Development*, 2nd edn. London: Institute of Education.
- Field, J. (2000). *Lifelong Learning and the New Educational Order*. Stoke-on-Trent: Trentham.
- Lazear, E. P. (ed.) (2002). *Education in the Twenty-First Century*. Washington, DC: Hoover Press.
- Lundvall, B.-A. (1992). *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. London: Pinter.
- Machin, S. and Vignoles, A. (eds.) (2005). *What's the Good of Education? The Economics of Education in the UK*. Princeton, NJ: Princeton University Press.
- Marginson, S. (1997). *Markets in Education*. St. Leonards, NSW: Allen and Unwin.
- OECD (1998). *Human Capital Investment: An International Comparison*. Paris: OECD.
- OECD (2001). *The Well-Being of Nations: The Role of Human and Social Capital*. Paris: OECD.
- Prais, S. J. (1995). *Productivity, Education and Training: An International Perspective*. Cambridge: Cambridge University Press.
- Reich, R. (1991). *The Work of Nations: A Blueprint for the Future*. London: Simon and Schuster.
- Woolf, A. (2002). *Does Education Matter? Myths About Education and Economic Growth*. London: Penguin.

Relevant Websites

- <http://www.skope.ox.ac.uk> – Centre for Research on Skills, Knowledge and Organizational Performance (SKOPE) (UK).
- <http://ioewebserver.ioe.ac.uk> – Institute of Education, Centre for Research on the Wider Benefits of Learning (UK).
- <http://www.oecd.org> – Organization for Economic Co-operation and Development, Statistics Portal.
- <http://ec.europa.eu> – The European Commission, Eurostat.
- <http://www.uis.unesco.org> – United Nations Educational, Scientific and Cultural Organization: Education.

Wider Benefits of Adult Education

T Schuller, OECD, Paris, France

R Desjardins, Danish University of Education, Copenhagen, Denmark

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This article focuses on the wider benefits of adult learning, covering a mix of personal and social (external) effects. The emphasis is on the effects of adult learning on health and civic engagement. Less attention has been given to this area than to the economic benefits of adult learning, although there is a growing body of research in this area. Some reference is made to the human and personal development aspects of adult learning as these can also be considered wider benefits; however, this topic has been discussed more elaborately elsewhere in the encyclopedia.

Defining the Wider Benefits

A basic issue is the question of what changes are affected by learning interventions. The changes are not limited to the individual. Adult learners can in turn initiate changes in the wider sense, by affecting the home/family, work, and community contexts that they engage in. There is a deep-rooted belief that adult learning has the potential to create personal, economic, and social value. This value can accrue to a variety of actors: to the learner, other private interests such as a current or future employer, and/or to society at large. In short, individuals, employers, and governments invest in adult learning with an expectation that there are benefits of different kinds to be realized.

Benefits can be categorized in a number of ways: cognitive or affective; psychological or behavioral; job or leisure related; and expected or unexpected. The notion of wider benefits usually refers to a mix of external, public, and nonmonetary benefits. The term social benefits is often used to refer to a similar range of effects (Behrman and Stacey, 1997; McMahon, 1999; Wolfe and Haveman, 2001; Psacharopoulos, 2006), although it does not include private nonmonetary benefits as in the notion of wider benefits as used by some authors (e.g., Schuller *et al.*, 2001, 2004).

Benefits that do not necessarily accrue to individuals (or other private interests such as firms) who make the decisions to invest are referred to as external. They are external because they are not taken into account when deciding to invest, even if they might be expected. The presence of external benefits provides justification for public policies that foster adult learning (Desjardins *et al.*, 2006). Otherwise there may be underinvestment from the standpoint of public policy. External effects are

commonly associated with public or societal effects because they tend to be nonexclusive (Lucas, 1988).

The notion of wider benefits can also encompass non-monetary benefits which occur at both the private or public levels. Private individual-level nonmonetary benefits are valued by the individuals who take up the adult learning but these are not directly exchangeable on markets and hence have no direct monetary value. Examples include the entertainment or consumption value of learning, health and life satisfaction, and improved family life. In contrast, public benefits are those benefits to members of society other than the learners themselves. Nonmonetary examples of public benefits include crime reduction, trust, social cohesion, political stability, and a well-functioning democracy. **Table 1** highlights examples of each private–public and monetary–nonmonetary combinations. Public monetary examples include reduced health and social transfer costs. These various benefits are not independent of each other (McMahon, 1999). For example, a private monetary return can in turn lead to the private nonmonetary return of improved individuals and family health – which in turn can lead to the public monetary return of reduced public expenditures on healthcare.

Different Types and Purposes of Adult Learning

The incentive to invest (time, effort, money, and other resources) varies among the different actors, depending on the type of value and who it is that benefits; in principle, this has implications for who should (and does) pay (Becker, 1964). It also has implications for the demand structure of different types of adult learning interventions. Majority of adults participate in adult learning for job-related reasons (OECD, 2003). Nonetheless, participation for personal and social-related reasons also plays an important role, and depending on the country these can form a substantial component of overall participation rates in adult learning (Desjardins *et al.*, 2006).

The reasons for participating have an impact on the nature and supply of learning interventions, including the form and content, and this in turn has implications for the potential wider benefits. However, specific evidence on the effects of different stages and types of schooling, or different curricula and pedagogical approaches is sparse. Most studies focus on the number of years of schooling or

Table 1 Possible economic and social outcomes of learning

	<i>Private</i>	<i>Public</i>
Monetary	Earnings, income, and wealth Productivity	Tax revenues Social transfer costs Healthcare costs
Nonmonetary	Health status Life satisfaction	Social cohesion Trust Well-functioning democracy Political stability

From OECD (2007). *Understanding the Social Outcomes of Learning*. Paris: OECD.

highest level of educational attainment, primarily because these data are relatively cheap to collect and hence readily available in social surveys. Few studies focus on adult learning and what it is about these learning experiences that matter for wider benefits. In essence, much of the discussion of the benefits of learning at various levels has viewed education as a black box. Findings suggest that adult learning has positive effects on a wide spectrum of health and social outcomes and that these depend on individual life histories, social context, and the type of learning experience (e.g., academic, vocational, and leisure – see Feinstein *et al.* 2003).

Channels through which Adult Learning Can Affect Outcomes

Effects on Economic Positions and Resources

Adult learning can improve employability and income, which is a key pathway to realizing a range of other benefits. Adults with a record of adult education and training are less likely to be unemployed and more likely to experience wage growth (OECD, 2005: 35). This can translate into improved personal satisfaction and autonomy; personal health and security; and quality of child rearing. It can generate resources (e.g., time and money) for engaging in social, civic, and political activities, which in turn are key elements for democratic processes; solidarity and social cohesion; human rights and peace; equity, equality, and the absence of discrimination; and ecological sustainability – all important dimensions of a well-functioning society (Gilomen, 2003). Further, it can reduce inclinations toward criminal and antisocial behavior by meeting basic needs of subsistence and improving the chances for a successful life (Feinstein, 2002). The instrumental effect of education on income and wealth therefore feeds through into intrinsic benefits such as better health.

Effects on the Self

Learning can develop skill and cognition and can modify the traits and behavior patterns of adults (OECD, 2007).

Educational experiences can also serve an enlightenment function (Lauglo and Øia, 2006). They can promote tolerance of and respect for other groups (Turner, 1991), and in turn social cohesion. Many learning experiences make people aware of others around them and the complex processes involved in society (Pring, 1999), creating an interest to take part in the processes of social change. This also promotes an awareness of the value of investing in the future as well as an awareness of risks by providing an insight into the trade-offs among costs and benefits occurring at different points in time, which in turn influences a range of choices and behaviors, for example, regarding healthy lifestyles (Feinstein *et al.*, 2006). It can also develop psycho-social capabilities such as resilience that help to cope with adversity (Schoon and Bynner, 2003). Resilience has been empirically linked to a set of internal attributes (i.e., autonomy, problem-solving skills, a sense of purpose and future, and social competence), all of which are plausibly affected by continued learning (Howard *et al.*, 1999). Effects on other psychological characteristics include self-efficacy (the belief that the self can influence the world around them) and external efficacy (the belief and trust that others will respond to one's own actions).

In summary, learning experiences can lead to wider benefits by directly: shaping what people know; developing competencies, which help people use their knowledge to yield benefits for themselves and society; and cultivating values, attitudes, beliefs, and motivations that foster the potential for generating wider benefits. There is also a potential for negative effects, particularly where access to learning opportunities is unequal and where provision is injurious to self-concepts, learning, and development.

Effects on Social Position

The channels described above share the assumption that education affects outcomes by directly changing the self. A distinctly different mechanism suggests that education's impact is indirect, and operates by changing the position of the individual in the hierarchy of social relations (Verba *et al.*, 1978; Campbell, 2006; OECD, 2007). The main premise is that the relative position of individuals in a social hierarchy is largely a function of access to learning opportunities, and education's primary effect derives from its ability to locate people in this hierarchy. This can be called the positional or relative effect. For example, Campbell (2006) finds that certain social outcomes fit this model best, namely participating in politics: belonging to a party, or seeking to influence politics through lobbying. His findings imply that an across-the-board increase in adult learning participation preserving overall inequality may do little to increase political participation; or education may help people to better health by enabling

them to secure access to health treatment ahead of those below them in the pecking order.

Thus, learning may benefit individuals by increasing (or preserving) their social status. However, the benefit to the wider community may be nil or even negative, so that the benefits to some are achieved at the expense of others. This is far from hypothetical: to the extent that education accentuates rather than mitigates inequality, its overall net impact on health and other outcomes may well be negative.

Effects via Context

Adult learning can also have an impact on the role of contexts. Contexts refer to family, work, or community settings, where individuals have varying but always limited or bounded agency. Adult learning may influence the structural conditions of choice and opportunity as well as the distribution of resources (especially through collective agency). Therefore, it enables people, to some extent, to choose and shape the contexts within which they live and work, and the peers they associate with. Peer effects are potentially extremely strong, to positive and negative ends.

Effects on Health

Recent studies highlight the significance of the relationship between education and health outcomes (OECD, 2007; Feinstein *et al.*, 2006). Typically, health professionals have interpreted the association more narrowly as a marker of socioeconomic status. Findings now indicate that there are sizeable differences in health for those with different levels of education and that these are partly due to the effects of education and not solely due to differences that precede or explain education, such as socioeconomic status. For example, Ross and Mirowsky's (1999) findings suggest that education has health effects at all levels of income. Using rigorous methods, Spasojevic (2003) suggests that the effect of education on health is at least as great as the effect of income. An extensive review of the evidence on the direct effects of education concluded that independent of economic position, those with more years of schooling are substantially associated with better health, well-being, and health behaviors (see Feinstein *et al.*, 2006). In some cases, the evidence is particularly robust and suggests causality.

By combining findings from the National Child Development Study in the UK with a series of insights from biographical case studies collected by the Centre for Research on the Wider Benefits of Learning, Feinstein *et al.* (2003) (also see Bynner and Hammond, 2004; Feinstein and Hammond, 2004; Bynner, 2001) provide a rare analysis of the extent and nature of the wider benefits of adult learning. They find that adults who took at least one course between the age of 33 and 42 are more likely to have given

up smoking, increased their level of exercise, and increased their life satisfaction. According to the authors, such improvements can in turn lead to economic return, by reducing pressure on health services, and thus offer a return to the taxpayer and the economy more generally.

All types of courses were linked to an increase in exercise, but the marginal effects were larger for academic and leisure-related courses than vocationally oriented courses. Further, leisure courses appear to have a more important effect among adults who did not complete secondary schooling qualifications. One possible rationale for this latter observation is that the path to an increased sense of self-value and empowerment (psychological attributes that help people lead healthier lives) through learning, depends partly on previous learning experiences. Many adults, who have not completed secondary schooling, will have experienced academic difficulties and even failure; therefore, for some adults, an alternate sequencing of different types of learning may be necessary to build up a positive attitude toward learning and to avoid negative overall effects to psychological well-being. The biographical data, complemented with statistical results, strongly suggest that adult learning features as an important element in positive cycles of development and progression, and that there are cumulative effects associated with learning that occur in reinforcing sequences.

In a separate study using the British Household Panel Survey, Sabates and Feinstein (2006) estimate the effects of adult learning on the take-up of preventative health services. Using a model to predict changes in the levels of uptake of screening, they simulate the impact of whether 100 000 women were enrolled in adult learning on cervical cancer prevention. Adult learning is associated with a 2.2% point increase in the probability of utilizing screening. Using statistics on the smear tests analyzed in 2002 in the United Kingdom and the claim by health officials that cervical screening can prevent 80–90% of cancer cases in women who attend regularly, the authors estimate that about 116–134 cervical cancers would be prevented for every 100 000 women in adult learning.

Effects on Civic and Social Engagement

Adult learning can be instrumental for many in providing aptitudes that are useful for civic living and contribution. Svensson (1996: 62) found that the majority of participants in study circles (a form of adult popular education in Sweden) believes that they develop useful knowledge from participating. Civic skills acquired through nonpolitical channels, including on the job and in voluntary associations, are an important predictor of whether someone is politically engaged (Verba *et al.*, 1995; SOU, 1996: 47). Having skills motivates people by instilling a sense of agency – skills make people feel like they have something

to offer in the civic and social realm. Further, adult learning that takes place in the civic realm has been linked as an important contributor to the sustenance of democracy (Larsson, 1999).

The same study that was introduced above, Feinstein *et al.* (2003), also explored the effect of adult learning on a range of civic and social outcomes. They found that adult learning has a pervasive impact on social and political attitudes, especially among adults who participate in courses that are academically oriented. The suggestion is that these types of courses are most suited for opening minds and challenging previously held beliefs. Effect sizes of academically oriented adult learning on racism and political cynicism are on the order of about -0.07 to -0.10 and -0.03 to -0.065 , respectively (Preston and Feinstein, 2004: 25). Some respondents to the biographical field work indicated that learning experiences led to a greater understanding of people from different backgrounds. Even though social and political attitudes are thought to be fairly stable by mid-adulthood, adult learning was found to have normatively beneficial effects on most of the attitudes considered on a magnitude of up to 5% points, representing up to a 34% change from the baseline level predicted for those who did not participate. Overall, their findings link adult learning to increased racial tolerance, a reduction in political cynicism, a higher inclination toward democratic attitudes, and a higher level of political interest.

Bynner and Hammond (2004) report findings which suggest that participation in adult education courses is linked to higher levels of civic and political participation, including increased membership in groups and voter participation. Those who participate in one or two courses are about 34% more likely to become a member of an association and 13% more likely to begin voting compared to all those who abstained in the previous election. In contrast to its effects on attitudes, it is leisure-oriented courses that have the most significant impact on civic and social participation, especially among adults who have not completed secondary schooling. From the biographical accounts, it is those adults who are most initially isolated and lacking confidence that ascribed significant changes in their social activity to adult learning participation. In Sweden, Svensson (1996) found that at least 33% indicate that their motive for participating in study circles was to meet others.

Despite the link to increased social activity, less has been said about the effects of adult learning on the intents and purposes, or other qualitative aspects of different groups or networks (Emler and Frazer, 1999). Using the same biographical data from the Centre for Research on the Wider Benefits of Learning, Preston (2004) demonstrates that learning experiences may not only lead to an expansion of social networks, but can also cause their relocation and dissolution, albeit in such a way as to

maintain improvements in well-being. He linked adult learning to increases in self-confidence and self worth, which help to motivate individuals in removing themselves from unhealthy or even dangerous relationships.

Conclusions

The impacts of learning extend beyond those measured by increased productivity in the production of goods and services exchanged on markets. Learning in adulthood is linked to diverse items such as reductions of criminal activity, increases in social cohesion, changes in income distribution, savings in welfare and medical costs, and voter participation. Strong theoretical expectations about the wider benefits of adult learning exist, and although less abundant and generally less rigorous than the evidence on the impact of initial education (compulsory and post-compulsory schooling), there is some empirical research to support these expectations.

The measurement of noneconomic outcomes is attaining greater significance in many Organization for Economic Co-Operation and Development (OECD) countries. This trend brings with it a range of difficult methodological issues, notably in establishing robust techniques for assessing causality. A key issue is how far adult education is brought into this debate. Quantitative data on adult learning are universally weaker than that on initial schooling, and effects are therefore harder to measure using conventional approaches. Yet, extending the rationale for public and private investment in education to cover outcomes, such as improved health and stronger civic democracies, is a challenge which has much to offer.

See also: Adult Education and Civil Society; Economic Outcomes of Adult Education and Training.

Bibliography

- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special References to Education* (revised in 1975/1993). Chicago, IL: University of Chicago Press.
- Behrman, J. R. and Stacey, N. (eds.) (1997). *The Social Benefits of Education*. Ann Arbor, MA: University of Michigan Press.
- Bynner, J. and Hammond, C. (2004). The benefits of adult learning: Quantitative insights. In Schuller, T., Preston, J., Hammond, C., Brasset-Grundy, A., and Bynner, J. (eds.) *The Benefits of Learning: The Impact of Education on Health, Family Life and Social Capital*, pp 161–178. London: RoutledgeFalmer.
- Bynner, J. (2001). The wider benefits gained by adults returning to higher education. In Lourenço, E. (ed.) *Wider Benefits of Learning: Understanding and Monitoring the Consequences of Adult Learning*, pp 142–149. Lisbon: For ESREA by Xàtiva.
- Campbell, D. E. (2006). What is education's impact on civic and social engagement? In Desjardins, R. and Schuller, T. (eds.) *Measuring the Effects of Education on Health and Civic/Social Engagement*, pp 25–126. Paris: OECD/CERI.

- Desjardins, R., Rubenson, K., and Milana, M. (2006). *Unequal Chances to Participate in Adult Learning*. Paris: UNESCO.
- Emler, N. and Frazer, E. (1999). Politics: The education effect. *Oxford Review of Education* 25(1/2), 251–273.
- Feinstein, L. (2002). *Quantitative Estimates of the Social Benefits of Learning, 1: Crime*. London: Centre for Research on the Wider Benefits of Learning.
- Feinstein, L. (2002). *Quantitative Estimates of the Social Benefits of Learning, 2: Health (Depression and Obesity)*. London: Centre for Research on the Wider Benefits of Learning.
- Feinstein, L. and Hammond, C. (2004). The contribution of adult learning to health and social capital. *Oxford Review of Education* 30(2), 199–221.
- Feinstein, L., Hammond, C., Woods, L., Preston, J., and Bynner, J. (2003). *The Contribution of Adult Learning to Health and Social Capital*. London: Centre for Research on the Wider Benefits of Learning.
- Feinstein, L., Sabates, R., Anderson, T. M., Sorhaindo, A., and Hammond, C. (2006). What are the effects of education on health? In Desjardins, R. and Schuller, T. (eds.) *Measuring the Effects of Education on Health and Civic/Social Engagement*, pp 171–354. Paris: CERI/OECD. <http://www.oecd.org/edu/socialoutcomes/symposium> (accessed May 2009).
- Gilomen, H. (2003). Desired outcomes: A successful life and a well-functioning society. In Rychen, D. S. and Salganik, L. H. (eds.) *Key Competencies: For a Successful Life and a Well-Functioning Society*, pp 122–147. Cambridge, MA: Hogefe and Huber.
- Howard, S., Dryden, J., and Johnson, B. (1999). Childhood resilience: Review and critique of literature. *Oxford Review of Education* 25(3), 307–323.
- Larsson, S. (1999). Studiecirkeldemokratin. In Arnå, E. (ed.) *Civilsamhället*, vol. 84, pp 243–279. Stockholm: SOU.
- Lauglo, J. and Øia, T. (2006). *Education and Civic Engagement among Norwegian Youths*. Oslo: Norwegian Social Research.
- Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics* 22(1), 3–42.
- McMahon, W. (1998). Conceptual framework for the analysis of the social benefits of lifelong learning. *Education Economics* 6(3), 309–346.
- McMahon, W. (1999). *Education and Development: Measuring the Social Benefits*. Oxford: Oxford University Press.
- OECD (2003). *Beyond Rhetoric: Adult Learning Policies and Practices*. Paris: OECD.
- OECD (2005). *Promoting Adult Learning*. Paris: OECD.
- OECD (2007). *Understanding the Social Outcomes of Learning*. Paris: OECD.
- Preston, J. (2004). A continuous effort of sociability: Learning and social capital in adult life. In Schuller, T., Preston, J., Hammond, C., Brasset-Grundy, A., and Bynner, J. (eds.) *The Benefits of Learning: The Impact of Education on Health, Family Life and Social Capital*, pp 119–136. London: RoutledgeFalmer.
- Preston, J. and Feinstein, L. (2004). *Adult Education and Attitude Change*. London: Centre for Research on the Wider Benefits of Learning.
- Pring, R. (1999). Politics: Relevance of the humanities. *Oxford Review of Education* 25(1/2), 71–87.
- Psacharopoulos, G. (2006). The value of investment in education: Theory, evidence and policy. *Journal of Education Finance* 32(2), 113–126.
- Ross, C. E. and Mirowsky, J. (1999). Refining the association between education and health: The effects of quantity, credential and selectivity. *Demography* 36(4), 445–460.
- Sabates, R. and Feinstein, L. (2006). Education and the take-up of preventative health care. *Social Science and Medicine* 62, 2998–3010.
- Schoon, I. and Bynner, J. (2003). Risk and resilience in the life course: Implications for interventions and social policies. *Journal of Youth Studies* 6(1), 21–31.
- Schuller, T., Bynner, J., Green, A., et al. (2001). *Modelling and Measuring the Wider Benefits of Learning: A Synthesis*. London: Centre for Research on the Wider Benefits of Learning.
- Schuller, T., Preston, J., Hammond, C., Brasset-Grundy, A., and Bynner, J. (2004). *The Benefits of Learning: The Impact of Education on Health, Family Life and Social Capital*. London: RoutledgeFalmer.
- SOU (1996). *Cirkelsamhället: Studiecirkelarnas Betydelse för Individ och Lokalsamhälle*. Stockholm: Fritzes/Regeringskansliet.
- Spasojevic, J. (2003). *Effects of Education on Adult Health in Sweden: Results from a Natural Experiment*. New York: City University of New York.
- Svensson, A. (1996). *Vilka deltar i studiecirkel? II. Rapport från en undersökning av deltagare i studieförbundens cirkelverksamhet hösten 1995*. Stockholm: Folkbildningsrådet.
- Turner, J. H. (1991). *The Structure of Sociological Theory*, 5th edn. Belmont, CA: Wadsworth.
- Verba, S., Nie, N. H., and Kim, J. (1978). *Participation and Political Equality: A Seven-Nation Comparison*. Chicago, IL: University of Chicago Press.
- Verba, S., Schlozman, K. L., and Brady, H. E. (1995). *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge, MA: Harvard University Press.
- Wolfe, B. and Haveman, R. (2001). Accounting for the social and non-market benefits of education. In Helliwell, J. F. (ed.) *The Contribution of Human and Social Capital to Sustained Economic Growth and Well-Being*, pp 221–250. ch. 11, Ottawa, ON: Human Resources Development Canada and Organisation for Economic Co-operation and Development.

Further Reading

- Bourdieu, P. (1990). *The Logic of Practice*. Cambridge: Polity Press.
- Pryor, B. W. (1990). Predicting and explaining intentions to participate in continuing education: An application of the theory of reasoned action. *Adult Education Quarterly* 40(3), 146–157.
- Rubenson, K. (1975). *Participation in Recurrent Education*. Paris: CERI/OECD.

ADULT EDUCATION – PARTICIPATION

Contents

Barriers to Participation in Adult Education

Participation in Adult Learning

Barriers to Participation in Adult Education

K Rubenson, University of British Columbia, Vancouver, BC, Canada

© 2010 Elsevier Ltd. All rights reserved.

Introduction

A quick review of international policy documents reveals the importance of adult learning in supporting the well-being of nations and individuals. Success in realizing lifelong learning is seen as vital in promoting employment, economic development, democracy, and social cohesion. This has resulted in an urgency to develop a better understanding of why some adults participate in lifelong learning and others do not. There are two closely integrated bodies of knowledge focusing on this question, research on determinants of participation and studies of barriers to participation, that is being addressed in this article.

This article is organized into three sections starting with a brief discussion of conceptual and methodological issues. This is followed by a presentation of empirical findings, and then finally a review of theoretical.

Conceptual and Methodological Issues

Attempts to assess barriers to lifelong learning have traditionally used a classification developed by Cross (1981: 98), who sorts obstacles to participation under three headings:

- situational barriers (those arising from one's situation in life e.g., lack of time because of work, family responsibility, etc.);
- institutional barriers (practices and procedures that hinder participation e.g., fees, lack of evening courses, entrance requirements, limited course offerings, etc.); and
- dispositional barriers (attitudes and dispositions toward learning).

A fourth category has recently been added to the list. Informational barriers refer to a lack of information on education and learning offers and benefits (OECD, 2005).

While most surveys employ some version of Cross' classification, there are fundamental differences in how barriers are conceptualized, affecting who is being asked questions on barriers. One common view is that barriers are obstacles that prevent certain groups from participating. If these deterrents could be overcome, these people would participate in lifelong learning. According to this view, questions about barriers would only be addressed to nonparticipants. A different conceptualization is that barriers are factors that lower the extent of participation but may not entirely prohibit participation. By accepting this position, it is of interest to ask participants about possible barriers that may have caused them to lower the extent of their learning activities.

Another related issue concerns whether or not those who have indicated no interest in participating should be asked about barriers to participation. Large-scale national or international surveys like the US National Higher Education Survey (NHES), the Canadian Adult Education and Training Survey, or the Organization for Economic Cooperation and Development (OECD)-led International Adult Literacy Survey pose a battery of questions on barriers only to persons who indicate that they failed to take courses/programs they wanted to take and/or had felt a need for. The logic is that barriers come into existence only when an expressed wish to participate is thwarted; the role of research then is to discover the impediment. It seems irrelevant to ask individuals not interested in participating about barriers, because without an expressed interest there can be no barriers. Consequently, surveys using this approach tend to concentrate almost exclusively on situational and institutional barriers and pay little attention to psychological impediments.

As revealed in longitudinal research on participation, there are problems with using expressed interest to decide who should answer questions on barriers. In fact, large numbers of those who at one time had indicated no interest actually came to participate while a substantial number of people who stated that they were interested

never showed up. This finding reflects the broader changes that have occurred in the labor market, which among other things, forces people to participate because they are ordered to or feel pressured to undergo some form of adult education and training linked to their work. Thus, contrary to a commonly held position in research on participation and barriers, participation is not always a voluntary act.

The other major approach in large-scale surveys of barriers, found, for example, in national adult-education surveys in the Nordic countries, the UK, and in the European Union (EU) barometer on lifelong learning, treats the lack of interest as part of the cluster of barriers. Following this approach, the questions are asked by three groups of respondents: persons who have not participated, those who have considered but not participated, and those who have participated. This assures that more interest is given to dispositional barriers. Thus, in the 2001 UK National Adult Learning Survey (Fitzgerald *et al.*, 2002) respondents were asked to sort a deck of cards describing barriers containing the following dispositional impediments:

- I prefer to spend my free time doing things other than learning.
- I do not need to do any learning for the sort of work I do.
- I am not interested in doing any learning, training, or education.
- I have difficulties reading and or/writing.
- I would be worried about keeping up with other people.
- I feel I am too old to learn.
- I do not see a point in learning or education.
- I would be nervous about going back to school.

Similar lists of dispositional-oriented reasons for not participating can be found in Swedish and Norwegian national surveys.

Either approach raises ethical dilemmas. In arguing that everyone should be asked about barriers – including those without a declared interest in participating – one runs the risk of turning participation into a moral issue where it is seen as bad not to want to participate. In fact, it may be highly rational from a personal point of view not to want to engage in organized forms of adult learning. While this line of reasoning has many advocates, others would claim that it is not that straightforward. The counter argument is that if the system of adult education too strongly assumes that the adult is a conscious, self-directed individual who possesses the instruments necessary to make use of available adult education possibilities, it will rely on self-selection to recruit participants. This will, by necessity, widen rather than narrow the educational and cultural gaps in society.

In this respect, the design issue around barriers raises crucial questions about the relationship between the state and its citizens, and what understanding should inform

national surveys into barriers. In this context, Sen's (1982) concept of basic capability equality: the need to take into account, among other things, differences in those abilities that are crucial for citizens to function in society, is informative. Thus, as Nussbaum (1990) discusses, people living under difficult conditions tend to accept their fate because they cannot imagine any reasonable alternative. Consequently, it would be important to recognize that some citizens may not be sufficiently able to judge the value of and conditions for participation in adult education.

Findings

The overwhelming amount of information stems from large-scale surveys but there are also some in-depth qualitative studies that provide valuable insights into the reasoning of nonparticipants.

Findings from Large-Scale Surveys

Findings may vary between studies because of differences in methodology, target group, list of barriers, and so on, but nonetheless, some groups of barriers are consistently found to be of crucial importance. Looking across various surveys, situational and institutional barriers strongly dominate. In many countries, the main reason for not participating is commonly a lack of time. According to the International Literacy Survey (OECD, 2000), around 60% identified this as the major reason for not having started a non-work-related education one had needed or wanted to take. The EU-barometer on lifelong learning, looking at situational constraints across the 15 EU states, found that family-related obstacles like “my family commitments take up too much energy” were mentioned somewhat more frequently than job-related hindrances but the pattern differed from country to country (CEDEFOP, 2003). Not surprisingly, women are more prone to refer to family responsibilities than men.

Institutional barriers are also of major importance although to a somewhat lesser degree than situational barriers. In the *International Adult Literacy Survey* (IALS), on the average, around 45% of the adult population mentioned at least one institutional constraint that prevented participation in some form of work-related studies. The figure was lower for nonwork-related studies, around 30% (OECD, 2000). Generally, among institutional barriers, financial reasons (too expensive/no money) are by far the most prevalent hindrance, particularly in North America. However, as evident from the Euro-barometer on lifelong learning, cost is also a major restraining factor in Europe. Only between 12% and 21%, depending on purpose, were willing to pay for the full cost of studying, while close to 50% would pay none of the cost (CEDEFOP, 2003: 86). Another institutional barrier that is frequently identified is a lack of appropriate courses and the scheduling of them.

It is problematic to interpret the implications of the findings on situational and institutional barriers for policy. Time is not an endless resource and people have to make choices regarding how they want to spend their spare time. This is not to deny that some, because of work and family, may have very little time left over which they can freely decide to spend. For many, mentioning lack of time is as much a statement of the value they ascribe to education and training as to the expected outcome of such an activity. Thus, it is of interest to note that several studies have shown that participants mention situational barriers to the same extent, and in some studies, even more often than nonparticipants. This is also the case with institutional barriers where in fact there is a tendency for participants to report this slightly more often as the reason for not having taken other courses they were interested in. Similarly, Jonsson and Gähler (1996) found that there were as many people with objective barriers in terms of handicaps, young children, working hours, etc., that participated in adult education as did not participate. Based on this finding, the authors conclude (p.38): “Instead of barriers, that might have to do with cost, lack of time, it is probably differences in expected rewards that can explain why some choose to participate while others remain outside.”

While many would agree with Jonsson and Gähler’s interpretation of situational- and institutional-barrier data, it is important not to deny that some people face major hindrances like childcare and cost that makes it very hard for them to participate. So, for example, the fact that fewer people in upper-income brackets mentioned financial reasons is an indication that the answers not only reflect the willingness to pay but also the ability to do so.

The general findings on institutional and situational barriers are more or less the same regardless of survey approach. This is not the situation with dispositional barriers. Naturally, studies that approached only those who indicated an interest in participating pay rather scant attention to psychologically oriented barriers. Consequently, it is primarily surveys that have also approached those who indicated no interest that are of interest here. Findings from studies that have employed this strategy reveal two things. First, in almost all countries, a substantial share of the population refers to dispositional barriers. In the 11 European countries that took part in the EU-barometer the percentage raising psychological barriers varied from a low of 14% in Denmark to a high of 31% in the UK (CEDEFOP, 2003). Similarly, Livingstone (1999) found that psychological factors have a major impact on Canadians’ readiness to enrol in organized learning activities. Thirty-five percent stated that they did not need more education and one in five saw studying as boring. These dispositional barriers refer to perceptions like, little to gain by participating, concerns about own ability to succeed, belief that one is too old to go back to study, and in some cases bad previous experiences of schooling.

Second, comparing participants with nonparticipants, several studies indicate that the negative attitudes and dispositions toward organized adult education and training (dispositional barriers) are by far the most deterring factor (Rubenson, 2007). This would suggest that the main obstacle to participation is not situational or institutional but rather a lack of interest.

Results from Qualitative Studies

The few existing qualitative studies of barriers tend to concentrate on indifference to participating and are commonly conducted with a small group of nonparticipants. The research provides an in-depth insight into the subjective rationale for actively declining to engage in organized forms of adult education. The overall finding is that the lack of interest often reflects a subjective rationality that is constructed around the person’s life context. Several studies have pointed to how a lack of stimulating employment opportunities – either in the form of unemployment with small opportunities to become employed and/or a monotonous job – discourages participation (Paldanius, 2007). For this group, nonparticipation becomes a highly rational act. It is first when participation in adult education results in better and higher paying work that it is meaningful. The following quotes from Paldanius interviews with a group of nonparticipants can exemplify the subjective rationality of this group.

Learning for the sake of learning never, I have much more important stuff to do, for instance I can plant onions and then know that it will take so and so long time until I see the results of my actions, I have actually made something, manufactured something (Paldanius, 2007: 472).

Carlén (1999) found that automobile workers viewed work and education as separate praxis related to class identity. As wage earners, they should produce and not enter into other spheres. Forms of adult education that were unrelated to their work challenged their routine and were perceived to encompass a threat of change. Similarly Paldanius (2007) reports that a dominant view among the nonparticipants was that education was something that had to be done while waiting for the real life that begins in adulthood. By not participating in adult education, a person can avoid boredom. Further, unemployment did not seem to stimulate an interest in participating in adult education, but instead it seemed to inhibit their readiness for action.

Theoretical Perspectives

Large-scale national surveys on participation and barriers are rarely constructed on the basis of theory but are mostly descriptive. However, there are economic as well as

adult-education frameworks that could guide questionnaire construction.

Economic Theories

The human-capital perspective is prominent in education and training literature, although mostly absent in adult-education-participation frameworks. The underlying assumption is that individuals maximize welfare as they conceive it. Human capital analysis has, as a starting point, that individuals decide on their education by weighing the benefits and costs of this investment (Becker, 1964, 1993). Every action has a price tag in the market and every human act can be reduced to some kind of rational economic calculus of cost and benefit. The probability of participation increases as a function of the benefit/cost ratio. Common cost variables include: tuition, materials, and transportation as well as the less-tangible value of the time invested in studying. Benefits mostly focus on future monetary gains in the form of higher salaries but might also address job security, work conditions, and in some rare cases, cultural and other nonmonetary gains (US Department of Education, 1998: 13). As evident in the findings section above, some of the results can be interpreted in the context of a human-capital perspective. Although beyond the scope of this article, it should be noted that the homo-economicus framework has been severely criticized for its strong assumption of rationality. As Dow (1998: 13) states: if we see social structures as being organic and evolutionary, with creative, nondeterministic behaviors alongside behavior conditioned by habits and institutions, then individuals cannot be modeled according to deterministic rational principles.

An interesting variation on the cost-benefit framework is case-based decision theory (Gilboa and Schmeidler, 1995; cited in US Department of Education, 1998: 14). The idea is that people remember past problems, how they resolved them, and the outcome of action. When they meet a new problem, past experiences of similar problems direct their decisions. The framework does not assume that individuals have beliefs in the absence of data (recalled cases) and therefore does not list all possible costs and benefits, as only those in the memory can be used in reaching the decision.

It is worth noting that most large-scale surveys on participation and barriers already collect many demographic and social-background variables of interest to economic cost-benefit or expected-utility frameworks (US Department of Education, 1998: 66). However, they lack measures of relative expected utility and they are relatively meager on social psychological variables like intentionality and normativity with regard to adult education. Further, they are weak on external context, for example, the situation at work or in civil society, and mainly ignore key past experiences.

Adult Education

In adult education, it is mostly theories on participation, particularly motivation for engaging in adult education, that provide the framework for understanding barriers. Cross (1980: 122–124) found many common elements in existing adult education theories on participation and barriers. According to Cross, all:

- are interactions,
- build on Kurt Lewin's field-force analysis,
- are cognitivist,
- refer to reference-group theory,
- apply the concepts of incongruence and dissonance, and
- directly or indirectly build on Maslow's model of needs hierarchy.

On the basis of her review, Cross presents the so-called chain-response model which incorporates work on learning orientations, need press theory, and expectancy-valence theory. The model takes the individual as the starting point and begins by identifying two main constructs: self-evaluation and attitude toward education. These internal factors are seen to influence the value of goals and the expectation that participation will meet goals. Valence and expectations are also affected by life transition and development tasks that confront the individual in various life-cycle phases. Opportunities and barriers and available information will then modify whether or not an individual will come to participate. This model, like almost all others reviewed by Cross, employs psychological concepts to develop an explanation of why some adults participate while others do not. Cross (1981) argues that this does not mean that societal aspects are ignored; on the contrary, all theories are interactionist, that is, they understand participation in terms of interaction between an individual and his or her environment. However, theories tend to neglect the individual's life history. Further, they do not directly address how the main constructs in the model are related to and interact with the broader structural and cultural context. However, comparative data on participation reveal some interesting national differences that bring into question the usefulness of trying to understand barriers by focusing solely on how the individual interprets the world, which most theories on barriers and participation tend to do. First, the participation research shows that there are substantial differences in participation between countries at comparable stages in the modernization process and with quite similar economies (Desjardins *et al.*, 2006). Second, it reveals that while age, family background, educational attainment, and work-related factors are linked to inequality in participation in all countries, the level of inequality varies substantially between countries (OECD, 2000). Third, the findings suggest that patterns of inequality in adult

learning mirror broader structural inequalities in society, like inequalities in income. Participation patterns in a country thus seem to reflect its particular welfare-state regime. These empirical findings suggest that we have to consider broader structural conditions and targeted policy measures, and analyze the interaction between these and the individual's conceptual apparatus. Based on these observations Rubenson and Desjardins (2009) developed, the so-called bounded-agency model that aims to take account of the interaction between structurally and individually based barriers to participation.

The Bounded-Agency Model

The bounded-agency model is premised on the assumption that the nature of welfare-state regimes can affect a person's capability to participate. In particular, the state can foster broad structural conditions relevant to participation and construct targeted policy measures that are aimed at overcoming both structurally and individually based barriers. Structural conditions play a substantial role in forming the circumstances faced by individuals and limit the feasible alternatives to choose from, and therefore they can bound individual agency. According to the model, a particular welfare-state regime can be found not only to be implicated in social structures, adult-education systems, and life chances, but also in individual consciousness. This assumption builds on Sen's (1999) concept of human capability and functioning, which stresses the importance not only of having resources available – internal (i.e., knowledge or skills such as literacy) or external (i.e., money) – but also in terms of individuals knowing about the range of possibilities of how these resources can be employed to realize things that matter to them, and knowing how to do so. In this sense, dispositional barriers can be seen as factors that restrict a person's capability and hence freedom to participate. Further, dispositional barriers can be affected and even caused by structural barriers, such as institutional and situational ones.

To illustrate this point, the authors analyze the barrier section in the EU-barometer, which suggests that adults in Nordic and non-Nordic countries experience similar barriers to participation and nearly to the same extent. With regard to the bounded-agency model, the key question becomes the extent to which structural conditions and individual dispositions afford the individual the capability and freedom necessary to overcome barriers. In nearly all cases, adults from Nordic countries were more likely to participate in adult education even though they may perceive the same barriers as their counterparts in other countries. However, as reflected in high participation rates and relatively lower levels of inequality, the Nordic welfare state seems to be comparatively effective at resolving barriers. The Nordic welfare states feature structural

conditions under which a larger group of adults, as compared to non-Nordic countries, seem to value participation and hence see an expected reward. These conditions include a labor market structured around a high-skill strategy and a civil society that fosters learning for both social and personal development. In the bounded-agency model, the impact of these conditions on a person's capabilities and consciousness with regard to the beginning of adult education is referred to as the 'conditioning of values and perspective on opportunity structure'.

Conclusion

Three findings stand out in the review of barriers. First, while the policy community's concern for adult learning has resulted in the development of major national and supranational policy-driven surveys, there does not seem to be much of an interest in the scholarly community to engage with barriers. This is in sharp contrast to the 1970s and first half of the 1980s, when a considerable body of conceptually oriented work on participation was produced. However, at that time it was more an unease regarding a lack of scholarly progress in adult education than efforts to contribute to evidence-informed policy that drove the interest. Second, the findings from national and international surveys as well as a review of the scholarly literature point to some limitations in the present design of the major comparative surveys that are informing the policy discourse on barriers and suggest that the existing design might result in too simplistic an account of the factors behind barriers. A main limitation with many of the present surveys is that they concentrate almost exclusively on situational and institutional barriers. Therefore, consideration needs to be given to how to strengthen assessment of dispositional barriers. To address the lack of attention to dispositional barriers and questions on future interest, some existing conceptual frameworks can provide a fruitful starting point for getting better measures on general attitude as well as intrinsic and extrinsic values of learning. The data suggest that it would be particularly important to construct better measures of factors related to the work context and link these to expected utility and intentionality. Third, there is a serious lack of comprehensive qualitative studies. It would be particularly useful for the development of evidence-informed policy as well as theory generation to have access to comparative qualitative data. This would allow researchers an in-depth insight into the subjective rational for citizens' reasons to decline to engage in organized forms of adult education and also to explore how individual dispositions are directly and indirectly affected by broader structures and specific adult-learning policies.

See also: Participation in Adult Learning.

Bibliography

- Becker, G. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. New York: Columbia University.
- Carlén, M. (1999). *Kunskapslyft eller avbytarbänk?* Dissertation. Acta universitatis Gothoburgensis: Göteborg.
- CEDEFOP (2003). *Lifelong Learning Citizens' Views*. Luxembourg: Office for Official Publications of the European Communities.
- Cross, K. P. (1981). *Adults as Learners: Increasing Participation and Facilitating Learning*. San Francisco, CA: Jossey-Bass.
- Desjardins, R., Rubenson, K., and Milana, M. (2006). *Unequal Chances to Participate in Adult Learning: International Perspectives*. Paris: UNESCO.
- Dow, S. C. (1998). Rationalisation in economics: Theory, methodology and action. In Dennis, K. (ed.) *Rationality in Economics: Alternative Perspectives*, pp 5–15. Boston, MA: Kluwer Academic Publishers.
- Fitzgerald, R., Taylor, R., and LaValle, I. (2002). *National Adult Learning Survey*. London: Department for Education Skills.
- Jonsson, J. and Gähler, M. (1996). *Folkbildning och vuxenstudier. Rekrytering, omfattning, erfarenheter-sammanfattning* (SOU1996: 159). Fritzes: Stockholm.
- Livingstone, D. W. (1999). Exploring the icebergs of adult learning: Findings of the first Canadian survey of informal learning practices. *Canadian Journal for the Study of Adult Education* **13**(2), 49–72.
- Nussbaum, M. (1990). Aristotelian social democracy. In Douglass, R. B., Mara, G. M., and Richardson, H. S. (eds.) *Liberalism and the Good*, pp 203–252. New York: Routledge.
- OECD (2000). *Literacy in the Information Age*. Paris: OECD.
- OECD (2005). *Promoting Adult Learning*. Paris: OECD.
- Paldanius, S. (2007). The rationality of reluctance and indifference toward adult education. *Proceedings of 48th Annual American Adult Education Research Conference*, pp 471–476. Minneapolis, MN: University of Minnesota.
- Rubenson, K. and Desjardins, R. (2009). The impact of welfare state regimes on barriers to participation in adult education: A bounded agency model. *Adult Education Quarterly* **59**(3), 187–207.
- Rubenson, K. (2007). *Determinants of Formal and Informal Canadian Adult Learning*. Ottawa: Human Resources and Skills Development Canada.
- Sen, A. (1982). *Choice, Welfare and Measurement*. Cambridge, MA: MIT Press.
- Sen, A. K. (1999). *Development as Freedom*. Oxford: Oxford University Press.

Further Reading

- Bélanger, P. and Valdivieso, S. (eds.) (1997). *The Emergence of a Learning Society. Who Participates in Adult Learning?* Oxford: Pergamon Press.
- Desjardins, R. (2004). *Learning for Well Being. Studies Using the International Adult Literacy Survey*. Stockholm: Institute of International Education, Stockholm University.
- Jonstone, J. W. C. and Rivera, R. J. (1965). *Volunteers for Learning. A Study of the Educational Pursuits of American Adults*. Chicago, IL: Aldine.
- MacKeracher, D., Suart, T., and Potter, J. (2006). State of the field review: Barriers to participation in adult learning. <http://library.nald.ca/research/item/6105> (accessed June 2009).
- Tuijnman, A. and Hellström, Z. (2001). *Curious Minds – Nordic Adult Education Compared*. Köpenhamn: Nordic Council.
- US Department of Education (1998). Adult education participation decisions and barriers: Review of conceptual frameworks and empirical studies. *Working Paper No. 98-10*. National Center for Education Statistics: Office of Educational Research and Improvement.

Participation in Adult Learning

R Desjardins, Danish University of Education, Copenhagen, Denmark

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Participation research is concerned primarily with three overarching questions: What is the extent of participation? Who is participating? Why are certain people or groups participating either more or less, or not at all? Traditionally, the focus has been on adult education and training (AET), rather than a broader notion of adult learning. In this article, a distinction is made between AET and informal learning where possible.

What Is the Extent of Participation?

Adult Education and Training

Participation rates based on the International Adult Literacy Survey (IALS; 1994–98) are reported in **Table 1**. **Tables 2(a)** and **2(b)** supplement with other international comparative data sources, and although the data from the alternative sources are not strictly comparable, the overall patterns are fairly consistent. In general, it can be inferred that rates vary substantially across countries, falling into four broad groups:

- close to or exceeding 50% – the Nordic countries include Denmark, Finland, Iceland, Norway, and Sweden;
- between 35% and 50% – countries of Anglo-Saxon origin such as Australia, Canada, New Zealand, the United Kingdom, and the United States as well as a few of the smaller Northern European countries such as Luxembourg, the Netherlands, and Switzerland;
- between 20% and 35% – this group features the remaining Northern European countries such as Austria, Belgium (Flanders), and Germany, some Eastern European countries such as Czech Republic and Slovenia, and some Southern European countries such as France, Italy, and Spain; and
- consistently below 20% – some Southern European countries such as Greece and Portugal, some additional Eastern European countries such as Hungary and Poland, and the only South American country where comparable data are available, Chile.

However, according to the 2003 data from the Adult Literacy and Lifeskills Survey (ALLS), Canada, Switzerland, and the United States appear to have climbed into the exceeding 50% category.

The mean number of hours per adult is also reported in **Table 1**. This combines the incidence and volume of

AET and thus offers a more comprehensive measure of the total effort. Denmark, Finland, and New Zealand report an average of over 100 h of AET per adult over a 12-month period – this is equivalent to every adult aged 16–65 years spending over 2.5 working weeks in AET per year. Countries featuring high participation but low average volume display comparatively lower adult learning per capita. Switzerland, the United Kingdom, and the United States have participation rates around 35–50% but, after adjusting for low volume, countries in the 20–35% range, such as Ireland and Slovenia, surpass them in their total AET effort. The former are considered to follow an extensive model, in which a fairly low volume is provided to a large number of adults, whereas the latter are considered to follow an intensive model, where provision is concentrated on fewer people (OECD, 2003a).

Changes in AET

Many national data sources point toward a general increasing trend in AET participation over the last 25 years. This is mostly attributed to the rising concern for human capital over the last decades since increases in AET for job-related reasons account for much of the rise since the early 1980s (Boudard and Rubenson, 2003: 267). More recent trend data from the European Union Labour Force Survey (ELFS) reveal a mixed pattern – participation rates appear to have generally increased between 1995 and 2000; however, this does not hold for all countries (OECD, 2003a: 39).

Types of AET

It can be seen from the IALS data in **Table 1** that job-related AET is dominant. Personal- and social-related AET can also play a substantial role. Comparatively high rates of participation in nonjob related AET are reported in Finland, the Netherlands, and Switzerland. It is, however, difficult to distinguish between different types of AET. Often, this is assessed on the basis of individuals' reasons for participating. However, Rubenson (2001) showed that there are many reasons for participation and that these are interrelated. Further, Desjardins *et al.* (2006) demonstrated that the way in which questions are phrased has implications for interpreting the complex motivations associated with participation. Ideally, surveys should not only permit respondents to state a number of different reasons for participating, but also ask them to rank them according to importance.

Table 1 Participation in adult education and training and average number of hours of participation in the previous year, by type of training, population aged 16–65 years^a

	<i>Total AET</i>			<i>Job-related AET</i>			<i>Nonjob-related AET</i>		
	<i>Participation rate</i>	<i>Mean number of hours per participant</i>	<i>Mean number of hours per adult</i>	<i>Participation rate</i>	<i>Mean number of hours per participant</i>	<i>Mean number of hours per adult</i>	<i>Participation rate</i>	<i>Mean number of hours per participant</i>	<i>Mean number of hours per adult</i>
<i>International Adult Literacy Survey (IALS), 1994–98</i>									
Australia	36.4	179.4	65.4	31.2	165.4	51.6	8.2	229.5	18.8
Belgium (Flanders)	20.9	125.4	26.2	12.5	103.7	13.0	8.1	90.7	7.3
Canada	36.9	230.8	85.2	30.0	236.1	70.9	9.9	179.9	17.8
Chile	19.5	235.7	45.9	11.2	157.0	17.5	9.2	330.1	30.3
Czech Republic	26.4	23.4	6.2	19.8	45.5	9.0	8.0	83.2	6.6
Denmark	56.9	219.3	124.7	48.6	213.6	103.9	13.0	148.5	19.2
Finland	58.4	207.8	121.3	39.6	205.5	81.3	28.6	195.9	55.9
Hungary	19.8	177.4	35.2	13.3	156.5	20.7	7.3	153.6	11.2
Ireland	23.5	262.1	61.6	17.3	262.1	45.3	6.8	98.3	6.7
Italy	22.6	192.8	43.5	16.1	133.6	21.5	7.9	282.6	22.4
Netherlands	37.9	239.8	91.0	24.9	266.0	66.3	16.7	176.8	29.6
New Zealand	47.9	222.9	106.7	39.4	223.2	87.9	12.8	176.9	22.6
Norway	47.0	190.3	89.4	42.7	167.5	71.4	6.7	322.3	21.7
Poland	14.7	151.3	22.3	9.7	119.6	11.6	5.5	165.0	9.0
Portugal	14.2	–	–	–	–	–	–	–	–
Slovenia	33.6	201.9	67.8	25.3	172.9	43.8	11.1	216.6	23.9
Sweden	50.8	–	–	–	–	–	–	–	–
Switzerland	42.1	134.7	56.7	26.6	139.9	37.3	19.7	111.2	21.9
United Kingdom	45.0	157.0	70.6	39.6	140.9	55.9	9.8	162.8	15.9
United States	40.7	131.1	53.3	36.4	128.3	46.6	6.3	114.6	7.3

^aAdults aged 16–19 years participating in full-time studies (4 or more days per week) toward ISCED 0–3, and who are not financially supported by an employer or union are excluded. Similarly, adults aged 16–24 years in full-time studies (4 or more days per week) toward ISCED 4–7, and who are not financially supported by an employer or union are excluded.

^bSweden and Portugal did not ask about job-related and nonjob-related training in a comparable way, nor did they ask about training durations. Germany is excluded because the survey did not ask about adult education and training in a comparable way.

^cMean number of hours per adult = Mean number of hours per participant* Participation rate/100.

^dDuring the past 12 months, that is since ..., did you receive any training or education including courses, private lessons, correspondence courses, workshops, on-the-job training, apprenticeship training, arts, crafts, recreation courses, or any other training or education?

–, indicates that data are not available.

Source: International Adult Literacy Survey, 1994–98; reprinted from Desjardins, R., Rubenson, K., and Milana, M. (2006). *Unequal Chances to Participate in Adult Learning: International Perspectives*. Paris: UNESCO.

Table 2 Participation in adult education and training in the previous year, population aged 16–65^a

<i>Participation rate</i>			
<i>(a) Adult Literacy and Lifeskills Survey (ALLS), 2003^b</i>		<i>(b) EU Barometer, 2003^c</i>	
Bermuda	47.0	Austria	39.3
Canada	49.3	Belgium	31.5
Italy	19.0	Denmark	58.5
Norway	53.3	Finland	55.9
Switzerland	56.9	France	22.8
United States	54.6	Germany	35.0
		Greece	16.7
		Iceland	69.1
		Ireland	33.6
		Italy	25.3
		Luxembourg	38.1
		Netherlands	41.0
		Norway	46.0
		Portugal	12.7
		Spain	26.5
		Sweden	53.5
		United Kingdom	41.4

^aAdults aged 16–19 years participating in full-time studies (4 or more days per week) toward ISCED 0–3, and who are not financially supported by an employer or union are excluded. Similarly, adults aged 16–24 years in full-time studies (4 or more days per week) toward ISCED 4–7, and who are not financially supported by an employer or union are excluded.

^bDuring the last 12 months... did you take any education or training? This education or training would include programs, courses, private lessons, correspondence courses, workshops, on-the-job-training, apprenticeship training, arts, crafts, recreation courses, or any other training or education.

^cHave you done any studies or training in the past 12 months? Please choose the three answers that best describe your own situation: yes, to meet new people; yes, to be less likely to lose my job/to be less likely to be forced into retirement; yes, to better enjoy my free time/retirement; yes, to be able to do my job better; yes, to obtain a certificate, diploma or qualification; yes, to be able to take greater responsibilities/increase my chances of promotion; yes, to better manage my everyday life; yes, to change the type of work I do altogether, including starting my own business (for retraining, etc.); yes, to achieve more personal satisfaction; yes, to get a job; yes, to improve my chance of getting another job, including one which would suit me more; yes, to increase my general knowledge; yes, for other reasons (spontaneous); no, I have not, but I would like to; no, I am not particularly interested, no, for other reasons (spontaneous); and don't know. Source: Adult Literacy and Lifeskills Survey, 2003; EU Barometer, 2003; reprinted from Desjardins, R., Rubenson, K., and Milana, M. (2006). *Unequal Chances to Participate in Adult Learning: International Perspectives*. Paris: UNESCO.

Definitions and Measurement

Most AET data sources focus on measuring formal provision; however, increasingly, there is an interest in nonformal and informal activities. This poses some challenges, especially in defining what counts as adult learning. Increasingly, it is difficult to distinguish adult learners

from first-time students attending regular school or university. Pragmatic solutions are to consider all the learning activities of the adult population aged 25–65 years or, if possible, to consider the population aged 16–65 years, but exclude full-time students aged 16–24 years. A more sophisticated definition, if the data allow for it, is to count the studies of the following groups as participation: full-time students aged 16–24 years who are sponsored by an employer or union/association; full-time students over the age of 19 years who are enrolled in primary or secondary programs; and full-time students older than 24 years who are enrolled in postsecondary programs. The last group is significant because there can be substantial overlap between what is considered adult learning and higher education, especially in countries where the higher education system features a high degree of openness to non-traditional adult students.

Another issue is the reference period for which participation rates are based on. For example, the IALS uses a 12-month reference period, whereas the ELFS uses a 4-week period. Shorter reference periods are adequate for reporting participation rates of populations, but inadequate for an in-depth understanding of adult learning pathways, both in terms of understanding the takeup of adult learning and its interaction to the provision of opportunities.

Informal Learning

Informal learning encompasses a broad range of learning activities (Livingstone, 1999). Patterns of engagement thus depend on what is being measured. Of the nine informal learning measures in the ALLS (2003), two dominate. Learning by doing is mentioned by around 90%, while learning by watching ranges from a high of 87% in Switzerland to a low of 77% in Canada. These former measures cover a broad range of nonspecific experiences; therefore, it is difficult to interpret their significance or value. More specific informal learning activities, related to work and culture, are particularly prevalent in Switzerland and less common in Canada and the United States with Norway somewhere in between. A vast majority of the Swiss (86%) report that they read manuals or other materials compared to 55% for Canada and the United States. The Swiss more often note that they learn by being sent around their organization or attending special talks. Further, they more frequently (44%) go on guided tours at museums or galleries than Canadians, Americans, or Norwegians (30%). Comparative measures of learning through the interactive use of information technology reveal only small differences in the use of computers or the Internet, while learning with the help of video, television, and tapes varies from a high of 52% in the United States to a low of 35% in Switzerland.

Who Is Participating?

Observed differences in participation have been linked to inequalities of opportunity and living conditions; therefore, the demographic and social make-up of participation becomes an important issue.

Adult Education and Training

For a range of countries (18 Organisation for Economic Co-operation and Development (OECD) ones and two non-OECD ones), IALS data show that those who are women, older, from low socioeconomic backgrounds (as reflected in their parents' level of education), low educated, low skilled, in low-skill jobs, unemployed, or immigrants are the least likely to participate in adult learning (see **Table 3**). In many cases, people belong to more than one group at the same time, which exacerbates observed differences (see Desjardins *et al.*, 2006: 74).

Informal Learning

Some research has stressed that the law of inequality does not apply to informal learning (e.g., Livingstone, 1999). However, this depends on the measure of informal learning used. Measures that are all inclusive or very general and refer to nonspecific situations of learning by doing or learning by watching show that informal learning is more or less a universal activity (see above). In contrast, measures which are context specific and reflect learning that is likely to enable the creation of or access to resources tend to reveal a clear pattern of inequality. Rubenson *et al.* (2007) find that groups with low levels of educational attainment report a substantially lower engagement in reading or using computers to learn. There is limited research on the extent to which different forms of informal learning contribute to strengthening resources that have economic and social value.

Why Are Certain People or Groups Participating More than Others?

Determinants

Different characteristics can be used to explore and reveal patterns of participation in greater detail. As described above, this can be done by distinguishing among salient groups such as those delineated by: age, sex, social class, level of education, level of skill, and occupational, employment, and minority status. This broadens our empirical understanding of who participates. However, a determinants analysis goes a step further by including a range of factors that are thought to be relevant in explaining the observed patterns. This may include the same

characteristics used to define a group; however, the difference in a determinants analysis is that there are underlying theories that link the characteristics to the observed pattern, and thus have explanatory value. Participation research has explored various explanatory factors.

Disciplinary Perspectives

Different explanations have been put forth, ranging from those rooted in psychological to sociological to economic perspectives (Tuijnman and Fägerlind, 1989). The most appealing ones have an interdisciplinary character since they are more useful for building a comprehensive understanding of adult learning participation. It is rare, however, that different disciplinary perspectives are brought together. There is no unified or comprehensive theoretical perspective guiding participation research. One downside to complex inclusive models is that they can inhibit empirical testing and limit the usefulness of results, especially with regard to specific situations or needs. For example, practitioners may specifically be interested in knowing what information is best suited for altering adult beliefs about the likely outcomes of participating. In this situation, a psychological perspective may provide the necessary depth. Ideally, a portfolio of models, which can be applicable in different contexts for different purposes, needs to be built up.

Further research requires explanatory models or theories that are disciplinary, multidisciplinary, and multi-level. Rubenson (1987) mentions three approaches to model building. The first focuses on the individual's psychological factors – the micro-level. The second emphasizes external factors and their structural conditions which influence the individual – the macro-level. The third approach looks into the interaction between individual and social forces. Each type of explanation is essential, but neither is sufficient in isolation from each other.

Explanations Based on Psychological Perspectives

While motivation can equally be a social phenomenon that is driven by external expectations which are placed on individuals – such as family, workplace, and community demands for competencies – many explanations focus solely on individual psychological factors. The psychological perspective focuses primarily on personality traits, intellectual abilities, and other behavioral dispositions that center on attitudes, expectations, intentions, and other motivational attributes.

Personality traits and abilities

Individuals have a degree of agency; therefore, cognitions, beliefs, and psychosocial capabilities feature as crucial

Table 3 Percent of adults participating in AET and adjusted odds ratios^{a,b} showing the likelihood of participating in AET during the year preceding the interview, by various classification variables, 1994–98

	<i>Australia</i>			<i>Canada</i>			<i>Chile</i>			<i>Czech Republic</i>			<i>Denmark</i>			<i>Finland</i>			<i>Hungary</i>			<i>Ireland</i>			<i>Italy</i>		
	%	Odds		%	Odds		%	Odds		%	Odds		%	Odds		%	Odds		%	Odds		%	Odds		%	Odds	
<i>Age</i>																											
16–25	45.5	3.6	***	43.6	4.8	***	24.9	3.6	***	24.4	1.8	**	68.0	2.4	***	69.8	1.8	**	27.7	3.7	***	29.0	3.5	***	34.5	1.4	
26–35	41.2	1.6	***	41.9	2.2		24.6	3.5	***	34.0	2.1	**	63.1	1.5	*	70.4	1.0		27.6	3.4	***	27.2	2.1	***	28.3	1.0	
36–45	40.4	1.5	***	41.8	2.3		20.7	3.3	***	29.1	1.6	***	64.3	1.8	***	64.9	1.0		19.6	1.8		25.3	2.1	*	25.0	1.1	
46–55	30.6	1.2		33.4	2.0		12.0	1.9		29.9	2.2		55.4	1.5		55.1	0.9		15.7	1.7		18.4	1.6		18.0	1.1	
56–65	17.9	1.0		14.7	1.0		6.9	1.0		8.8	1.0		32.0	1.0		29.7	1.0		3.4	1.0		9.1	1.0		9.3	1.0	
<i>Gender</i>																											
Women	35.1	1.0		36.3	1.0		19.9	1.0		21.7	1.0	***	59.2	1.0	**	62.3	1.0	**	20.5	1.0		25.0	1.0		19.1	1.0	
Men	37.8	1.1		37.6	1.3		19.1	0.8		31.2	1.7	***	54.7	0.8	**	54.5	0.7	**	19.1	0.9		22.1	0.8		26.2	1.1	
<i>Parent's education</i>																											
Less than upper secondary	34.2	1.0	**	28.4	1.0		15.5	1.0		22.2	1.0	*	49.2	1.0		50.0	1.0	*	11.6	1.0		21.3	1.0		18.5	1.0	
Upper secondary	41.3	1.2	***	42.4	1.3	***	33.1	1.2		34.2	1.3	*	59.5	1.0		65.7	1.2	**	25.1	1.2	*	33.4	1.0		39.2	1.0	
Higher than upper secondary	47.5	1.2		55.6	2.0		46.5	1.5		31.4	1.0		69.5	1.1		77.2	1.5		43.2	1.6		38.1	1.3		51.9	1.5	
<i>Education</i>																											
Less than upper secondary	24.0	1.0	***	20.7	1.0		9.8	1.0	**	17.7	1.0	***	41.2	1.0		35.9	1.0	***	7.1	1.0		14.1	1.0		9.2	1.0	
Upper secondary	38.0	1.3	***	32.2	1.0	*	25.5	1.7	***	35.8	1.6	***	56.2	1.2	***	63.5	1.5	**	18.2	1.3	***	27.7	1.2	***	37.5	2.0	***
Higher than upper secondary	56.1	1.8		55.7	1.9		45.2	2.5		47.0	1.7		74.6	1.8		80.0	1.9		47.7	4.0		44.9	1.7		51.3	2.8	***
<i>Prose literacy skill level</i>																											
Level 1	14.2	1.0	***	17.3	1.0		10.6	1.0		12.0	1.0	*	22.7	1.0	***	21.2	1.0		8.4	1.0		8.9	1.0		8.4	1.0	
Level 2	26.6	1.7	***	28.3	1.3		25.3	1.2		22.4	1.6		47.3	1.9		41.7	1.3		20.3	1.3		17.8	1.4		21.9	1.5	***
Level 3	42.6	2.8	***	43.1	2.0	***	40.2	1.4		33.1	2.0	**	69.2	2.9	***	68.2	2.5	***	34.6	1.6	***	30.0	2.2	***	38.6	2.1	***
Level 4/5	60.6	4.4	***	52.9	2.2		47.8	1.2		41.5	2.3	**	77.3	2.8	***	82.8	3.8	***	46.0	1.5		48.2	3.5	***	44.5	2.1	***
<i>Employment status</i>																											
Unemployed	28.8	1.0		29.5	1.0		19.3	1.0		19.5	1.0		53.4	1.0		30.4	1.0	***	12.6	1.0	**	9.7	1.0		16.3	1.0	
Employed	43.2	1.2	***	41.7	1.2		22.5	0.9		33.4	1.0		60.4	1.0	***	69.6	2.8	*	28.7	2.1	***	30.2	1.3		29.1	3.0	***
Retired	9.0	0.4	*	10.7	0.5		15.3	2.0	***	5.1	0.3		18.0	0.3	***	16.5	0.6		1.4	0.2	***	3.5	0.4		5.4	0.4	**
Student	96.2	69.5	***	66.1	3.9		100.0	>99	***	21.3	0.7	***	89.4	16.0	*	91.8	93.3	**	29.5	>99	***	50.2	27.2	***	100.0	>99	***
Homemaker	13.0	0.3		23.1	0.6		8.0	0.5	**	0.0	0.0	***	22.0	0.3		29.1	0.5	**	0.0	0.0	***	11.3	1.1		3.3	0.2	***
<i>Occupation</i>																											
Blue collar low skill	25.6	1.0		25.5	1.0		11.5	1.0		18.7	1.0		42.3	1.0		47.8	1.0		13.3	1.0		16.9	1.0		13.1	1.0	
Blue collar high skill	32.6	1.0	***	29.4	1.2	***	12.8	1.1	**	29.1	1.4		47.1	1.2	***	50.1	1.2	**	14.5	0.9		15.8	0.9	*	14.6	1.0	
White collar low skill	43.7	1.6	***	39.8	2.0	***	29.8	2.0	***	22.5	1.3	***	63.0	1.8	***	68.3	1.6	***	25.4	1.3	**	37.6	1.8	*	26.7	1.4	
White collar high skill	55.4	1.7	***	53.4	1.8	***	49.8	3.2	***	45.0	2.2	***	73.2	2.3	***	82.0	2.9	***	45.4	1.8	**	45.6	2.2	**	49.8	2.5	***
<i>Immigration status</i>																											
Foreign born	30.9	1.0	**	33.4	1.0		31.5	1.0		13.2	1.0	**	54.9	1.0		61.2	1.0		24.2	1.0		26.5	1.0		28.7	1.0	
Native born	38.5	1.3		37.9	1.3		19.4	4.5		26.5	2.3		56.9	0.6		58.3	2.0		19.8	1.7		23.3	1.0		22.4	0.9	

Continued

Table 3 Continued

	Netherlands			New Zealand			Norway			Poland			Slovenia			Sweden			Switzerland			United Kingdom			United States			International average	
	%	Odds		%	Odds		%	Odds		%	Odds		%	Odds		%	Odds		%	Odds		%	Odds		%	Odds		%	Odds
Age																													
16–25	48.3	2.5	***	60.4	4.4	***	42.8	2.8	***	17.7	3.8	***	43.3	2.1	**	41.0	1.1	*	48.8	2.3	***	51.5	2.7	***	33.4	1.8	*	41.9	2.8
26–35	45.9	2.0	***	51.6	2.1	***	56.2	2.1	***	17.4	2.5	**	45.5	1.8	**	57.0	1.4	*	49.6	1.9	**	51.2	1.5	***	46.5	1.5	**	43.3	1.9
36–45	40.4	1.9	*	50.5	1.7	**	52.5	2.0	***	18.4	2.7	***	38.9	1.8		60.7	1.6	**	43.7	1.7		54.3	1.7		44.9	1.3	*	40.9	1.8
46–55	30.6	1.4		42.7	1.4		47.3	1.8		11.3	1.8		27.6	1.3		57.1	1.5		38.6	1.5		41.0	1.1		43.8	1.4		33.8	1.5
56–65	17.0	1.0		27.6	1.0		23.8	1.0		2.8	1.0		9.5	1.0		34.7	1.0		24.6	1.0		21.1	1.0		26.9	1.0		17.8	1.0
Gender																													
Women	36.4	1.0		46.9	1.0		46.1	1.0		13.4	1.0	**	32.5	1.0	**	52.3	1.0		40.7	1.0	*	44.0	1.0		40.6	1.0		36.2	1.0
Men	39.4	1.0		49.0	1.1		47.9	1.0		16.1	1.4		34.7	1.3	**	49.4	1.0		43.5	0.8	*	45.9	0.9		40.8	1.0		37.2	1.0
Parent's education																													
Less than upper secondary	33.8	1.0		44.5	1.0		38.6	1.0		11.5	1.0		23.3	1.0		47.9	1.0		30.3	1.0		44.0	1.0		26.5	1.0		30.6	1.0
Upper secondary	49.5	1.4	***	51.7	0.9		49.5	1.2		21.4	1.0		44.8	1.2	**	53.6	1.2	*	47.7	1.2	*	62.8	1.5		44.7	1.1		44.4	1.2
Higher than upper secondary	46.2	1.1		62.3	1.3	*	58.3	1.4	**	35.7	1.3		66.1	1.4		57.4	1.1		50.1	1.1		67.7	1.3		55.3	1.3	**	53.3	1.3
Education																													
Less than upper secondary	26.5	1.0		37.8	1.0		1.0			8.0	1.0		12.1	1.0		35.0	1.0		20.4	1.0		34.1	1.0		14.0	1.0		21.8	1.0
Upper secondary	43.7	1.5	**	51.0	1.1		44.3	1.3		21.2	2.3	**	36.4	1.7	**	51.8	1.4	***	45.3	2.1	***	53.3	1.4	**	32.1	1.2		39.6	1.5
Higher than upper secondary	52.8	1.9	***	64.7	1.6	***	65.0	1.8	**	34.3	3.0	***	73.6	4.6	***	67.0	1.8	***	56.5	2.9	***	71.0	2.0	***	62.7	2.4	***	58.3	2.3

Continued

Table 3 Continued

	Netherlands		New Zealand		Norway		Poland		Slovenia		Sweden		Switzerland		United Kingdom		United States		International average										
Prose literacy skill level																													
Level 1	21.9	1.0		29.9	1.0		14.6	1.0	***	8.3	1.0		15.3	1.0	***	29.4	1.0		22.4	1.0		20.8	1.0		14.1	1.0	**	16.7	1.0
Level 2	29.6	1.4	**	36.2	1.0		38.7	2.5	***	15.4	1.0		40.7	1.8	***	41.1	1.1	*	35.8	1.2	*	34.5	1.3	***	30.6	1.9	***	30.8	1.5
Level 3	42.2	1.7	***	55.7	1.8	***	51.5	3.0	***	23.4	1.2		61.6	2.1	*	52.2	1.4	**	52.0	1.8	***	58.0	2.6	***	48.3	3.0	***	46.9	2.1
Level 4/5	54.4	2.5		68.3	2.1		64.2	3.9		40.8	1.5		75.0	2.5	*	60.1	1.5		63.7	2.5		74.5	4.4		64.2	3.8		59.3	2.7
Employment status																													
Unemployed	37.8	1.0		33.5	1.0		37.6	1.0		9.4	1.0	***	17.1	1.0	***	44.0	1.0		33.0	1.0		30.1	1.0	*	26.3	1.0		27.1	1.0
Employed	43.4	1.0		54.2	1.4		53.4	1.1		20.5	1.5	***	42.2	2.5	**	59.5	1.2	***	46.2	1.4	***	56.8	1.4	**	48.0	1.3		43.5	1.5
Retired	13.1	0.6	*	17.7	0.7		6.7	0.2	***	2.0	0.4		5.8	0.4		16.1	0.4		20.5	1.0	*	8.6	0.3		12.5	0.5	***	10.4	0.5
Student	65.6	6.5		91.8	>99	*	44.5	4.8	***	18.7	3.1		87.7	15.7		39.7	1.3		49.1	3.8		55.8	23.7	***	42.0	41.7	*	63.3	22.2
Homemaker	23.2	0.7		23.4	0.5		14.0	0.3		3.7	0.5		11.0	1.3		25.5	0.4	**	24.8	0.8		13.6	0.4		13.7	0.6		14.6	0.5
Occupation																													
Blue collar low skill	29.1	1.0		35.9	1.0		36.6	1.0		11.5	1.0		17.9	1.0		39.3	1.0		28.3	1.0		36.6	1.0		22.4	1.0		26.2	1.0
Blue collar high skill	39.5	1.5		40.5	1.0		43.3	1.3		10.2	1.0		24.9	1.4		41.9	1.0		35.6	1.2		36.9	0.9		29.4	1.3		30.5	1.2
White collar low skill	42.8	1.5	*	54.0	1.7	***	47.4	1.4	*	18.0	1.3		47.4	2.7	***	52.0	1.6	***	43.6	1.3		58.0	1.6	***	44.2	1.9	**	42.5	1.6
White collar high skill	47.8	1.5	*	69.2	2.6	***	63.2	1.9	***	37.6	2.3	**	68.0	3.5	***	68.2	2.5	***	56.2	1.8	**	68.9	1.7	***	64.9	2.3	**	58.0	2.3
Immigration status																													
Foreign born	45.0	1.0		46.4	1.0		41.1	1.0		3.3	1.0		23.4	1.0		37.9	1.0		27.4	1.0	***	45.0	1.0		30.1	1.0		33.6	1.0
Native born	37.4	0.7		48.2	1.3		47.4	1.5		14.9	2.8		34.8	1.6		52.1	1.2		45.6	1.8		45.0	1.0		42.3	1.5		37.3	1.6

^aOdds ratios reflect the relative likelihood of an event occurring for a particular group compared to a reference group. An odds ratio of 1 represents equal chances of an event occurring for a particular group vis-à-vis the reference group. Coefficients with a value below 1 indicate that there is less chance of the event occurring for a particular group compared to the reference group, and coefficients greater than 1 represent increased chances. From Hosmer, D. W. and Lemeshow, S. (1989). *Applied Logistic Regression*. New York: Wiley.

^bOdds are adjusted for: age, gender, parents' education, education, prose literacy skill level, employment status, occupation, immigrant status, minority language status, and size of community.

*p < 0.10, statistically significant at the 10% level.

**p < 0.05, statistically significant at the 5% level.

***p < 0.01, statistically significant at the 1% level.

Source: International Adult Literacy Survey, 1994–98; reprinted from Desjardins, R., Rubenson, K., and Milana, M. (2006). *Unequal Chances to Participate in Adult Learning: International Perspectives*. Paris: UNESCO.

elements that can explain participation. Participatory behavior is the result of diverse interactions between individuals' beliefs, skills, capabilities, and values. Rubenson (1987) stresses the importance of self-concepts such as self-esteem and self-efficacy in predicting participation. He suggests that adults who feel good about themselves are more likely to succeed in achievement-oriented situations. Conversely, an important benefit of adult learning is improved beliefs about self-efficacy (Hammond, 2003), pointing to a cycle of recurrent learning.

Previous learning experiences are key factors predicting further learning (Tuijnman, 1989; Boudard, 2001; Desjardins, 2004). Independent of educational attainment, Tuijnman (1989) finds that cognitive ability also exerts a positive influence on the accumulation of learning experiences over the lifespan. Educational attainment reflects accumulated knowledge, skills, and other traits that are associated with the probability of continued learning.

A common trait shared by early school leavers is a lack of self-confidence with regard to learning because of bad pedagogical experiences (Illeris, 2004a). Adults with low levels of education are less likely to participate because they lack readiness both in terms of knowledge and skills as well as their motivation to learn. A low readiness to learn is a substantial dispositional barrier to participation.

Motivational orientations and reasons

Motivational orientations toward learning can be important predictors of participation. Houle (1961) proposed a typology which suggests that participants are either goal oriented (use learning to accomplish objectives), activity oriented (find meaning in the circumstances of learning), or learning oriented (seek knowledge for its own sake). Boshier (1971, 1982) and Boshier and Collins (1985) developed the Education Participation Scale (EPS) to assign motivational orientation scores to individuals which allowed for in-depth investigation into the relationships among orientations and various demographic variables and other characteristic variables. Although the amount of variance explained by sociodemographic variables was small, research by Boshier and colleagues revealed that those with low education, low occupational status, and low income were most likely to participate for social contact, social stimulation, community service, and external expectation reasons. In contrast, those with higher levels of education, occupational status, and income were more often enrolled for professional advancement and cognitive interest reasons.

These findings partly corroborate with Maslow's (1954) hierarchy of needs theory which suggested that the motivating forces behind participation are conditional on whether subordinate needs are satisfied. Individuals with lower-order needs for survival which remain to be satisfied are deficiency motivated, whereas those working toward higher-order needs are growth motivated.

Arguably, obtaining secure employment is a lower-order need and, thus, the implication by Boshier's finding that professional advancement is associated with a higher-order need is puzzling. A more broadly based measure of job-related AET reveals that most adults participate for job-related reasons, whether associated with low socioeconomic status or not (e.g., OECD, 2003a; Desjardins *et al.*, 2006).

Learning for job-related reasons is linked to goals of finding a job, finding a better job, being promoted at work, keeping a job, and/or becoming more efficient in one's current job. It is a dominant reason reported (at least 60%) in recent surveys such as IALS, ALLS, and the 2000 European Union Barometer, and reaches as high as 90% in Australia, Denmark, Norway, the United Kingdom, and the United States. The divide between job- and non-job-related reasons, however, is not so clear-cut (Courtney, 1992: 50), which may be a reason why attempts to link distinct motivational orientations to participation explain little variation. Further, adults may find their reasons for participating difficult to articulate and they might not always be aware of them all (Darkenwald and Merriam, 1982: 136; Rubenson, 2001). Even a temporary lack of a specific reason may be seen as a reason for participating since the activity itself can be a way to obtain or rediscover new goals that could be pursued (Courtney, 1992: 87).

Attitudes and Intentions

Research has also pointed to the importance of attitudes toward learning. Houle (1961) claimed that every adult has an underlying conviction about the nature and value of learning which influences their opinion and, hence, the decision to participate. Darkenwald and Hayes (1988) constructed the Adult Attitudes toward Continuing Education Scale (AACES) to investigate the relationships among attitudes and participation. They found that the importance attributed to learning appears to be the most decisive factor in predicting participation; however, importance in relation to what needs further attention (see below).

Explanations Based on Social Perspectives

Individualistic perspectives have had serious consequences for how inequalities in participation are understood and what measures are deemed adequate to support lifelong learning for all. Individuals face several constraints in acting independently and making their own free choices. Shortcomings to the individualistic psychological and economic perspectives can be addressed by turning to the various external (and structural) influences on participation, such as social and economic institutions (government policy, organizations, industries, markets, and classes) at a macro level, and work structures at a micro level.

Separately, life history approaches to studying participation have broadened the structuralist approach by embracing much of the criticism of the individualistically oriented theories. This is done by situating the role of individual subjective experiences and actions as well as collective ones in their wider social and cultural contexts. A discussion regarding the relationships between life situation and participation as well as institutional barriers and participation, which are an elaboration of explanations based on social perspectives, is given elsewhere in the encyclopedia.

Explanations Based on Economic Perspectives

Most, but not all, explanations based on economic perspectives have tended to be dominated by individualistic approaches to the decision to participate. A common assumption is that individuals make a rational choice to participate or not, and that this decision is based on the information they have regarding the costs and benefits of participating. However, individuals are substantially limited by the imperfect information they have regarding the costs and benefits. There are also risks inherent with realizing the benefits which rational agents may not want to undertake.

Cost limitations and credit constraints

Desjardins *et al.* (2006) underscored the point that a lack of external economic support (from employers and governments) and credit constraints, especially for disadvantaged groups, is a significant barrier to participation. Even if individuals would like to participate because the future benefits outweigh the immediate costs, they may not have the financial means to do so because of credit constraints and imperfect capital markets. When individuals are not capable of borrowing money to invest in learning, because they do not have any collateral, which would otherwise be profitable, then market failure occurs.

Financial support and incentive-to-invest factor

The long arm of the job is becoming longer and stronger in terms of financing as well as motivation. IALS data show that about two-thirds of participants receive employer support (OECD and Statistics Canada, 2000). There are indications that participation may not always be voluntary and there may be increasing pressure to participate in job-related AET (Hight, 1998; Carré, 2000). Employer-supported AET is often suggested by employers, although a large portion that is suggested by employees is also supported by the employer (Tuijnman and Hellström, 2001). Many adults also participate for job-related reasons even though they do not receive financial support from their employer, which reveals the strength of the long arm of the job. Self-financing is, on average, the second most common

source of financial support, and, in some countries, the dominant source.

Firms represent a large portion of the training market (OECD, 2003a: 51–53). The Second Continuing Vocational Training Survey (CVTS2) data show that over 70% of firms in the majority of European countries provide support for AET (European Commission, 2002). Adults who work in large firms, especially those that compete on global markets and undergo significant technological change and/or changing work practices, appear to receive more AET.

The supply of employer-supported opportunities appears to be primarily targeted at prime-age employees who are highly educated and skilled. Plausibly, employers consider adults with higher levels of education more trainable. Older workers, women, and immigrants tend to face reduced opportunities for employer-supported AET (Desjardins *et al.*, 2006). These tendencies are consistent with the overall patterns of participation (see Table 3).

There is much debate about whether current levels of investment in AET are adequate. The potential for underinvestment arises due to several market failures, and can be linked to both employer and employee behavior (see Desjardins *et al.*, 2006). Overall, evidence shows that an underprovision of AET is likely to occur in all OECD countries (see OECD, 2003b: 248).

Government support, the least common source of financial support, tends to benefit those who already display high rates of participation, namely younger adults, the higher educated, and those who are in white-collar, high-skill occupations, rather than vulnerable groups. This is likely due to pressures for government policies to seek increases in efficiency through the adoption of a more market-oriented approach and outcomes-based funding. This increases the likelihood that AET initiatives/programs will target those easiest to recruit and most likely to succeed. Initiatives to reach disadvantaged groups often correspond better to the demands of the advantaged (Rubenson, 1999: 116). Few countries have effective public policies and structures in place to help those who are hard to reach. The Nordic countries are among the few and, accordingly, they tend to show comparatively higher rates of participation among the low educated (see Table 3).

Nature of work and skill-requirements factor

Recent research suggests that industrial and occupational structures of countries are instrumental in structuring participation. This perspective moves beyond the narrow individualistic one and encompasses a social and structural perspective. The world of work places substantial demands on individuals which necessitates continuous learning and periodic upgrades of competencies, acting as a substantial motivating force for many adults.

Recent research suggests that literacy practices at work – in particular the frequency and variety of reading practices – is one of the most significant determinants of participation in job-related AET (Desjardins *et al.*, 2006). More generally, the workplace is a learning space (Illeris, 2004b: 77–89); however, opportunities to learn new things on the job vary with the characteristic and position of the job, which exacerbates inequalities in adult learning (Åberg, 2002). Workers who are already better positioned in the labor market have more opportunities and incentives to acquire and develop competencies. Further, the structure of occupations and production in a particular country are likely to bear a strong influence on the distribution of work-related adult learning.

Explanations Based on Individuals' Interactions with Social Influences

Explanatory models of participation that include social and individual forces as well as their interaction have been proposed by a number of researchers. Some of the above explanations build on these, but a brief review of the more well-known ones is provided here. These models include most of the proximal and distal variables that are hypothesized as relevant to participation and are thus useful for building a comprehensive understanding of participation. The underlying reasons why adults participate (or not) are complex, featuring dynamic and interactive feedback effects that occur at multiple levels.

McClusky (1963) presented an individual–social interactive model called the power–load–margin model. Load represents the internal and external demands placed on the individual; power represents the agency of the individual to carry the load; and margin is the ratio of load to power which signifies the likelihood to participate. By linking Maslow's (1954) hierarchy of needs model with Lewin's (1947) force field analysis model, Miller (1967) suggested that when individual needs and social forces both point to a commonly perceived demand for participation, then participation will be high. Similarly, Boshier's (1971) congruence model suggested that when actual and perceived notions of self, others, and educational environment diverge, it is less likely that adults will participate.

Rubenson's (1975) expectancy valence model made a link between an individual's expectations about the value of participating, their attitude toward participating, and the likelihood of actual participation. According to this theory, participation will occur and persist if the learning activity is consistent with the learner's needs and expectations. According to Rubenson, the outcomes will depend on class since attitude and readiness are conditioned by structural and cultural factors. A model by Pryor (1990) and Pryor and Pryor (2005), which applied Ajzen and Fishbein's theory of reasoned action to participation,

focuses on more proximal variables. He suggested that participation is determined by the intention to participate and that the intention is determined by the attitude toward learning and perceptions of social pressures to participate. The latter is a subjective norm that is based on inferences about behavioral expectations of others. According to Pryor (1990), attitude tends to dominate over the subjective norm, and attitudes toward learning are primarily driven by a set of beliefs regarding the outcomes associated with learning.

Cross (1981) developed a psychosocial interaction model called the chain-of-response model, which suggested that participation relates to a complex chain of responses made by the individual vis à vis social circumstances. Beginning with a self-evaluation and a formation of attitudes toward learning, the importance of learning and the expectations associated with it are evaluated in relation to current needs and, in turn, this is influenced by available information, available opportunities, and institutional barriers.

Variants of these and other models exist and have been tested empirically. Based on a Swedish longitudinal study, Tuijnman (1989) puts an advanced set of structural models, which include psychological, sociological, and economic variables, to rigorous empirical testing. He reported that collectively, social origins, socioeconomic status, cognitive ability, initial levels of education, attitudes toward education, and specific interests in adult learning explain no more than 10–26% of the variance in the participation variable. Boudard (2003) more or less confirmed these patterns for a range of IALS countries.

Further Research

Major research questions that need to be addressed deal with the outcomes of adult learning and how this is linked with motivations to participate. Which mechanisms are more relevant for early development compared to the possible impact of later interventions? Is later intervention merely compensatory, with little chance of making a difference? Or could there be possibility for good timing later on? A better understanding of several relationships is needed:

- Substitutes and complements to more traditional schooling contexts.
- Formal education structures and adult learning, such as the degree of stratification and vocational specificity of pathways, and the extent and distribution of AET among different countries. This requires comparative data at both the system and individual levels.
- Occupational/industrial structures and adult learning. The structure of occupations and production in a particular country are likely to bear a strong influence on the distribution of job-related adult learning.

- Market failures and inequalities in participation. Market imperfections that may relate to learning outcomes are likely to have complex implications that spill over into other policy sectors; thus, reforms should not be undertaken without careful consideration of the relevant trade-offs. More specific information is necessary to devise viable strategies to overcome market failures and hence optimize the allocation and distribution of resources invested in the total learning effort.
- Government policy, governance, and adult learning. How are government policies regarding adult learning formed and coordinated at the intersection of various stakeholders and how does this shape the provision, purpose, and content of adult learning as well as participation?

See also: Barriers to Participation in Adult Education.

Bibliography

- Åberg, R. (2002). Överutbildning – ett arbetsmarknadspolitiskt problem. In Abrahamson, K., Abrahamsson, L., Björkman, T., Ellström, P.-E., and Johansson, J. (eds.) *Utbildning kompetens och arbete*, pp 41–62. Lund: Studentlitteratur.
- Boshier, R. W. (1971). Motivational orientations of adult education participants: A factor analytic exploration of Houle's typology. *Adult Education* 21(2), 3–26.
- Boshier, R. W. (1982). *Education Participation Scale*. Vancouver, BC: Learning Press.
- Boshier, R. W. and Collins, J. B. (1985). The Houle typology after twenty-two years: A large-scale empirical test. *Adult Education Quarterly* 35(3), 113–130.
- Boudard, E. (2001). *Literacy Proficiency, Earnings and Recurrent Training: A Ten Country Comparative Study*. Stockholm: Institute of International Education.
- Boudard, E. and Rubenson, K. (2003). Revisiting major determinants of participation in adult education with a direct measure of literacy skills. *International Journal of Educational Research* 39(3), 265–280.
- Carré, P. (2000). In Motivation in adult education: From engagement to performance. In *Proceedings of the 41st Annual Adult Education Research Conference*, 66–70. Vancouver, BC: University of British Columbia.
- Courtney, S. (1992). *Why Adults Learn: Towards a Theory of Participation in Adult Education*. New York: Routledge.
- Cross, K. P. (1981). *Adults as Learners: Increasing Participation and Facilitating Learning*. San Francisco, CA: Jossey-Bass.
- Darkenwald, G. G. and Hayes, E. R. (1988). Assessment of adult attitudes towards continuing education. *International Journal of Lifelong Education* 7(3), 197–204.
- Darkenwald, G. G. and Merriam, S. B. (1982). *Adult Education: Foundations of Practice*. New York: Harper and Row.
- Desjardins, R. (2004). *Learning for Well Being: Studies Using the International Adult Literacy Survey*. Stockholm: Stockholm University.
- Desjardins, R., Rubenson, K., and Milana, M. (2006). *Unequal Chances to Participate in Adult Learning: International Perspectives*. Paris: UNESCO.
- European Commission (2002). *European Social Statistics: Continuing Vocational Training Survey Detailed Tables (CVTS2 – Data 1999)*. Luxembourg: European Communities.
- Hammond, C. (2003). How education makes us healthy. *London Review of Education* 1(1), 61–78.
- Hight, J. E. (1998). Young worker participation in post-school education and training. *Monthly Labor Review* 121(6), 14–21.
- Hosmer, D. W. and Lemeshow, S. (1989). *Applied Logistic Regression*. New York: Wiley.
- Houle, C. O. (1961). *The Inquiring Mind*. Madison, WI: University of Wisconsin Press.
- Illeris, K. (2004a). *Adult Education and Adult Learning*. Copenhagen: Roskilde University Press.
- Illeris, K. (2004b). *Learning in Working Life*. Copenhagen: Roskilde University Press.
- Lewin, K. (1947). Frontiers in group dynamics: Concept, method and reality in social science. *Human Relations* 1, 5–41.
- Livingstone, D. W. (1999). Exploring the icebergs of adult learning: Findings of the first Canadian survey of informal learning practices. *Canadian Journal for the Study of Adult Education* 13(2), 49–72.
- Maslow, A. H. (1954). *Motivation and Personality*. New York: Harper and Row.
- McClusky, H. Y. (1963). The course of the adult life span. In Hallenbeck, W. C. (ed.) *Psychology of Adults*, pp 10–20. Washington, DC: Adult Education Association.
- Miller, H. L. (1967). *Participation of Adults in Education: A Force Field Analysis*. Boston, MA: Center for the Study of Liberal Education of Adults.
- OECD (2003a). *Beyond Rhetoric: Adult Learning Policies and Practices*. Paris: OECD.
- OECD (2003b). *Employment Outlook*. Paris: OECD.
- OECD and Statistics Canada (2000). *Literacy in the Information Age: Final Report of the International Adult Literacy Survey*. Paris/Ottawa, ON: OECD/Statistics Canada.
- Pryor, B. W. (1990). Predicting and explaining intentions to participate in continuing education: An application of the theory of reasoned action. *Adult Education Quarterly* 40(3), 146–157.
- Pryor, B. W. and Pryor, C. R. (2005). *The School Leader's Guide to Understanding Attitude and Influencing Behaviour: Working with Teachers, Parents, Students, and the Community*. Thousand Oaks, CA: Corwin.
- Rubenson, K. (1975). *Participation in Recurrent Education*. Paris: CERI/OECD.
- Rubenson, K. (1987). Participation in recurrent education: A research review. In Schuetze, H. G. and Istance, D. (eds.) *Recurrent Education Revisited: Modes of Participation and Financing*, pp 39–67. Paris: OECD/CERI.
- Rubenson, K. (1999). Supply of lifelong-learning opportunities. In Tuijnman, A. and Schuller, T. (eds.) *Lifelong Learning Policy and Research: Proceedings of an International Symposium*, pp 109–120. London: Portland Press.
- Rubenson, K. (2001). *Measuring Motivation and Barriers in the Adult Education and Training Survey: A Critical Review*. Hull, QC: Applied Research Branch, Human Resources Development Canada.
- Rubenson, K., Desjardins, R., and Yoon, E. (2007). *Adult Learning in Canada: A Comparative Perspective*. Ottawa, ON: Statistics Canada.
- Tuijnman, A. C. (1989). *Recurrent Education, Earnings, and Well-Being: A Fifty-Year Longitudinal Study of a Cohort of Swedish Men*. Stockholm: Almqvist and Wiksell.
- Tuijnman, A. and Fägerlind, I. (1989). Measuring and predicting participation in lifelong education using longitudinal data. *Scandinavian Journal of Educational Research* 33(1), 47–66.
- Tuijnman, A. C. and Hellström, Z. (2001). *Curious Minds: Nordic Adult Education Compared*. Copenhagen: Nordic Council of Ministers.

C

COMPARATIVE EDUCATION

Contents

Comparative and International Education: English-Language Scholarly Publications

Comparative Education: Philosophical Issues and Concepts

Comparative Education: Societies and Associations

Comparative and International Education: English-Language Scholarly Publications

D Post and M Fedri, The Pennsylvania State University, University Park, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Many national periodicals, dating back at least to the French Enlightenment, attended to developments in the schools and school systems of other nations. However, the first serial publications that were established explicitly to review education globally are more recent. This article discusses common trends and dilemmas facing several of the primary English-language periodicals whose purpose is to publish critical analysis, commentary, and research in education comparatively across nations. The article does not attempt a history of the field itself, its methods, findings, or the issues debated within specific journals. It does not offer an institutional history of these periodicals, since comprehensive accounts of their origins are available elsewhere. Instead, this article discusses the changing demands for, and demands on, journal publishing for readers who are concerned with comparative and international education.

Although reference is made to journals published in languages other than English, part of the dilemma facing authors and readers over the last 50 years is precisely that both have migrated to English-language publications, regardless of their language backgrounds. The identification of these publications and of comparative research with the English language – but not with any specific Anglophone country – reveals an important aspect of their construction. Another part of the recent trend is that these publications have increasingly tended to conceive themselves and be identified as scholarly journals. Although there remain publications with nonacademic

identities, the majority of journal articles in the field today are academic publications, that is, they are written and read by faculty of postsecondary educational institutions as opposed to school personnel or the educated public. For these reasons, English-language scholarly publications are the focus of this article.

The purpose of this short article is to describe the missions, administration, and the larger environmental context for publication about education comparatively across nations. The field of comparative and international education has evolved in response to new demands and opportunities. In addition to the field's evolution – but not independent of those changes – the media for propagating the field has also changed. New electronic technologies, gatekeeping functions, commercial relationships with readers, and library policies have all affected the direction of scholarly publications. In this article, such general changes are highlighted. More specific examples of the dilemmas facing authors and editors are drawn mainly from one journal, *Comparative Education Review* (CER).

Emergence of Scholarly Periodicals in Comparative and International Education

The longest-standing journal publishing today in the field is *International Review of Education* (IRE), which traces its origin to Cologne in 1931, when it was edited by Friedrich Schneider and co-edited by the American scholar Paul

Monroe, who himself had previously created the first edition of this *International Encyclopedia of Education*. After its interruption by World War II, IRE was reinvented as a periodical of the United Nations Educational and Scientific Organization (UNESCO), with articles appearing in German, French, and English. In its war-weary context, IRE's mission centered on mutual international understanding. The preamble to the UNESCO Constitution famously stated that, "since wars begin in the minds of men it is in the minds of men that the defenses of peace must be constructed." As the introductory editorial of the newly reconstituted IRE stated, "National educational systems and national educational ideas can no longer safely be developed without contact with educational thought and action elsewhere in the world. The maintenance of cultural life has ceased to be something to be guided by ideals purely national in scope." In addition to promoting international understanding and bridging political gaps, the IRE was intended to bridge professional/practitioner gaps. Thus, the introduction further stated the journal's intention to "provide a meeting-place for men and women from every country whose thoughts and actions deserve the attention of educationists throughout the world. Such men and women will usually, no doubt, themselves be professional educational thinkers and doers, but this will not always be the case: because education is today a matter of universal cultural concern contributions will be sought from leading minds in many fields."

The next journal in the field emerged shortly after the reconstituted IRE as a bulletin of a new association of US professors of education. The CER was created, along with the Comparative Education Society, in 1957 by US educators who saw scholarship and international cooperation not only as compatible but essential for a peaceful world. The tenor of the times was reflected in the international education work of Senator J. William Fulbright. In a recollection of his 1945 international education bill, Fulbright stated, "It is the task of education, more than any other instrument of foreign policy, to help close the dangerous gap between the economic and technological interdependence of the people of the world and their psychological, political and spiritual alienation." The journal's first editor, George Bereday, editorialized in the bulletin's premier issue that CER was designed "primarily for teachers of comparative education, who combine work in this subject with the teaching of other subjects and who cannot devote to the discipline their full-time research energies. . . . The *Comparative Education Review* is conceived as a service organ of the discipline and of the profession." The distinctions that are drawn today in universities between research and practice, or between service and scholarship, were invisible to Bereday and others concerned with education in a comparative light.

A third journal, *Comparative Education*, was launched in 1964 in the United Kingdom. It was designed to appeal to

a readership of general-education students and those who valued the role of schooling in the development of the British Commonwealth's new nation-members. Edmund King, writing in the inaugural issue of the journal, identified scholars and practitioners as equally important to the strength of the emerging field. King wrote:

The whole future of Comparative Education turns on whether we are trying to sanctify a cult or extending and strengthening the business of applying to the decisions of education a world perspective with the insights of many disciplines. Some decisions concern policy, all ultimately concern practice. Therefore we must make distinct and effective provision for students, teachers, and researchers in Comparative Education. We must also provide for those who (through this Journal) can apply the implications of Comparative Education in practice to their daily responsibilities in schools, homes, and education offices. The insights of these latter groups, if in line with world perspectives, are potentially as relevant as any others. They, and other pertinent information, must all be taken count of. In education today there is no one-way traffic.

Fundamental debates over the contours of comparative education flourished in the United Kingdom in the 1960s. These were reflected in the founding conference of the British Association for International and Comparative Education (BAICE) (under an earlier name). Once established, the association's bulletin – aimed only at keeping members informed of news – quickly evolved to one of the field's leading scholarly journals, *Compare*. One of the journal's early aims was to reach teachers of comparative education in many countries, and this aim remains today. The journal's broad scope, encompassing discourse, empirical study, and policy-centered analyses, directly reflect its parent association's mission. The journal has encouraged research into less-studied topics and contexts.

The most recently founded English-language scholarly journal in this field is the *International Journal of Educational Development*. The journal was conceived of in 1979 by two British management consultants, as a specific reaction to two troubling trends. First, they perceived a marginalization of education in development discourse and practice, which were dominated by economics. Second, they were dismayed by experts who felt able to issue policy recommendations and definitive scholarly analyses with little practical experience in or knowledge of specific contexts. This issue-driven motivation for the journal's founding, as opposed to service to a scholarly society, accounts for its relatively focused mission. It aims to highlight as well as interrogate the role of education in development, and to bring academic analyses to bear on practice and policy debates. Its enduring interest in the developing world is reflected in the number of submissions and published articles by a wide array of non-British

authors. In 2008, for example, its lead authors represented 24 countries from across the globe.

Trends and Challenges

Two of the English-language publications under review are sponsored by membership organizations. The BAICE sponsors *Compare*, and CER is sponsored by the Comparative and International Education Society (CIES). Although these associations originated in specific national contexts and were once dominated almost exclusively by UK and US citizens, it is significant that both have become increasingly international. For example, by 2008, approximately one third of CER's subscriber-members lived outside the United States, and the same fraction of libraries subscribing to CER are located beyond US borders. Many other readers are non-Americans who reside in the United States. Americans authored fewer than half the articles published by CER in recent years. The range of stakeholders in an originally US endeavor is striking, because the organizational membership of the World Council of Comparative Education Societies has expanded, and there is a growing number of non-English-language scholarly publication alternatives.

There are advantages to journal sponsorship by a membership organization that gives a journal subscription as a benefit of membership. A characteristic of emerging professions is greater identification with the calling of the profession than with the particular institution employing the professional. In an increasingly mobile profession, those who identify themselves with comparative education seek out a community beyond their institutions. By attending regular meetings of the BAICE and CIES, professionals obtain intrinsic benefits and collegiality that they might not otherwise experience. The economy of scale also lowers the cost of journal distribution, making individual subscriptions more numerous than institutional library subscriptions.

The three other journals of special interest in the article are not sponsored by membership organizations. As a consequence, individual subscribers are far outnumbered by institutional subscribers. Since there are fewer individual subscriptions, the publishers of these journals must charge higher rates to institutions. Furthermore, because these publishers are for profit, they pay particular attention to the financial bottomline. In recent times, subscription costs across the academic publishing spectrum have increased by about 13% per year, and for-profit publishers have typically charged six times more than nonprofit publishers. Another driver of cost has been the formation of for-profit conglomerates across all publishing fields, especially in the 1990s. For example, *International Journal of Educational Development* (IJED), published

from its early history by Pergamon Press, was enfolded by Elsevier. *Compare*, which was published by Carfax Publishers from 1976 to 2003, was incorporated into Taylor and Francis. The dilemma of pricing for institutional and membership subscriptions will only intensify as libraries scale back their purchases. In response to selective cuts, publishers now bundle subscriptions to their journals in packages. If a library decides to subscribe to one prominent journal, it is most cost effective to accept a package, which includes not only the desired journal but others that are in less demand. This strategy appears successful if success is measured by the number of citations to articles published in nonmembership journals. Despite the fact that *Compare* and CER (sponsored by the scholarly organizations) have greater numbers of individual subscribers than, for example, *Comparative Education*, the number of cites, downloads, and submissions are not so different as one might suppose. Electronic access through libraries, or access to the paper copies there, appears sufficient to engage scholars with the articles published in the non-membership journals.

The Role of the Academy

Except for IRE, which is edited by the staff of UNESCO, the publications under review have always been edited by full-time instructors and professors of higher education institutions. In the British context, this relationship has not entailed co-sponsorship by the institutions whose staff edit the IJED, *Compare*, or *Comparative Education*. The institutions housing those publications receive a comparatively small amount of support from the publishers and, in the case of *Compare*, from the sponsoring organization. The American context differs in that most university presses have traditionally been prepared to operate at a loss, and many US universities are willing to subsidize journals. Although this situation has changed somewhat over the years in the face of chronic financial concerns across higher education, it remains the case that many comparative-education journals based in the United States receive institutional support.

Significantly, CER has always been cosponsored by publicly funded universities with public service as part of their missions. A unique characteristic of American public higher education institutions is their commitment to open library and recreational access, as well as agricultural extension education, adult learning, and professional service. As professional dissemination is viewed as central to institutional missions, a partnership relation is more likely in the United States than in the United Kingdom. For example, CER drew more than half of its operating budget through contributions by its co-sponsoring institutions in recent years – the

universities of California, Pittsburgh, and Penn State. This arrangement is much less common outside the United States. Editorships of the other journals mentioned require greater personal sacrifice and volunteerism than does the editorship of CER. It is also likely that less personal time is available for each accepted manuscript in non-US publications, because of less-generous support provided by host institutions.

Manuscript Submission and Selectivity

In all the journals mentioned, there have been increases in the number of submissions. For example, in 2008, CER received 182 manuscripts, while IJED received 300. The editorial boards of journals read and give suggestions in response to every submission, if only to recommend a more appropriate journal to the author when there is an obvious mismatch. A smaller number of submissions are then sent for an external review by experts, some of whom may be chosen from the manuscript references themselves, and by other specialists in the region discussed or methods used by the author. Editors try to include at least one general reader, since the articles they publish need to attract a broad scholarly readership in addition to serve specialist researchers.

Publication in a peer-reviewed journal is time consuming both for authors and editors. For example, CER has rarely published an article without at least one major revision, and usually two or more. Even for the 15% of accepted articles, there are two further stylistic revisions, first by the CER editors and then again by University of Chicago Press copyeditors. CER editors generally invite revisions for about half the articles that are sent for an external review and, based on lengthy comments from reviewers and the editors' own reading of each submission, they reject the other half. In either case, the authors receive extensive suggestions. On several occasions, the reports submitted by external reviewers, together with a 2–3-page response from the editors, exceed 4000 words. Very rarely are fewer than six pages of comments returned to the author. This process can continue for as many as four rounds, over a 1–2-year period, becoming a friendly but critical three-way discussion between authors, editors, and reviewers.

Incentives for Reading and for Publication in Scholarly Journals

A fair question is why, given the large time investment needed to publish in a peer-reviewed journal, authors are willing to submit their research. One answer is that authors seek the engagement of this process at least as much as they seek formal publication or another line on

their curriculum vitae (CV). Consequently, even authors whose papers are rejected in the first round usually resubmit a revision if they are encouraged to do so by the editors. In this sense, a journal's function is as much about developing high-quality articles as about screening-out underdeveloped manuscripts. A more contention-filled reason is that journals have come to serve as gatekeepers for entry and advancement in higher education institutions, a function they did not originally intend but which affects the editing process nonetheless. Some editors feel that, since beginning scholars depend heavily on publication in their journals for advancement, it would be unfair to devote any part of their page budget to features other than research articles. Since double-blind, peer-reviewed research articles were originally only one part of these journals, the credentialing function they have been made to fill has clearly influenced their direction.

In the context of the gatekeeping function played by journals, it is not surprising that they have all experienced increasing numbers of submissions. However, the fact that authors are willing to submit revisions in such a lengthy and intensive process is remarkable in light of several other important contextual changes and challenges facing the journals in this field. One such change is the opportunity for scholars to publish in journals that are exclusively electronic, such as the now-mature and highly respected *Research in Comparative and International Education*, *Current Issues in Comparative Education*, and *Education Policy Analysis Archives*, the latter published in English, Spanish, and Portuguese.

Another change in the publishing context of comparative education, as much as other education fields, is the growing number of edited volumes that serve as an alternative to the increasingly lengthy process of publication in refereed journals. This marks a generational shift. In the 1950s, 1960s, 1970s, and even 1980s, periodicals served to disseminate quickly the findings of topical research. Books, however, took more time to produce and were expected to have longer shelf lives. Today, it has become far easier to publish an invited chapter in an edited anthology than to publish in a journal having a 15% acceptance rate, a 1-year backlog of accepted articles, and the expectation of two or more revisions. The result is an exploding number of edited, nonserial topical volumes from a wide range of commercial publishers. During its first decades, most journals in this field identified and reviewed new books of interest to comparativists. These books were almost always single authored or co-authored monographs representing a unified, coherent, sustained synthesis or presentation of original research. In recent years, most books received and/or reviewed were anthologies of shorter reports or investigations. In 2008, for example, out of 27 books reviewed in CER, only three were monographs, while the others were edited anthologies or conference proceedings.

Balancing Permanent and Current Scholarship

Many refereed journals in quickly moving fields seek ways to compensate for the lengthy review process so as keep their scholarship vital and current. No editor wants articles to become dated before they appear in print. Journal editors can help authors make a more immediate contribution with their articles by publishing electronic versions prior to paper versions. Most journals in international fields like comparative education now publish electronically, saving what otherwise would be many months from the time of acceptance to appearance online. Subscribers and members of subscribing libraries can sometimes download finished, edited pieces that have been scheduled for release in forthcoming numbers. Even so, the 6–9 months saved by web publication does not eliminate the time lag between submission and the possible acceptance of an article. Another approach used by editors is to anticipate proactively the themes and methods they believe will interest readers. Editors invite contributions along specific lines, using some combination of blind reviews or other selection methods. In the specific case of CER, the editors choose topics they believe are important to their readers and those who they target in efforts to expand readership. Next, they search for guest editors of the selected topic, who draft a detailed description of the issue and a call for submissions. CER circulates these calls, receives submissions and, after asking for reviewers' names from guest editors, places submissions in double-blind review as usual with at least three outside referees. Moderated discussions are the third way that some journal engage readers and create a more lively relationship with contemporary policy and research. For example, CER has hosted debates over the role of the World Bank in higher education development, patriotic education, the place-of-area studies centers, and reform within UNESCO.

Language of Publication

The irony of monolingualistic publications for comparative educators is apparent, but monolingualism is increasingly evident in all the publications discussed here. Polymath, polyglot comparativists of nineteenth-century education had no fear that their efforts would be lost in translation, since there was little need to translate their writings. However, after World War II, English became dominant in scholarly journals, such that even authors with mother tongues other than English began to use English in their scholarship. For example, during its initial period and even after it was reconstituted, IRE published equally in German, French, and English. Yet, recent contributions,

even by authors working in languages other than English, have been published almost exclusively in English. For example, of the 42 articles published in IRE during 2008, all but three were in English.

The dominance of English has not occurred for any lack of alternatives. The number of national associations for the study of comparative education has proliferated, many with journals, annual bulletins, or reports of congress proceedings. For example, *Education Comparée* was created for the French association in 1973, the Japanese Comparative Education Society created its own bulletin in 1975, and *Revista Española de Educación Comparada* (REEC) was first issued in 1995 as an instrument of expression for the members of *Sociedad Española de Educación Comparada* (Spanish Society of Comparative Education). An important exception to these examples of scholarly publications, which all originated from the initiative of nongovernmental organizations, is the mainland Chinese journal CER. This publication receives support from the Ministry of Education, and before that the propaganda department of the Central Committee of Communist Party of China. It was created in 1992 out of the former journal, *Foreign Education Trends*, and the mainland journal still has 'making foreign things serve China' as its official mission.

Several explanations have been offered for the dominance of English in comparative education despite the emergence of periodicals in other languages. One explanation is the legacy of colonialism. Intellectuals from the so-called periphery and former colonies continue to receive their education in English. Still, this does not explain the emergence of English in publishing among Francophone former colonies. Another explanation is that the leaders of comparative education have become members of English-based professional associations. An even more compelling interpretation, suggested by A. Suresh Canagarajah, is that journals in languages other than English do not appear in the libraries of peripheral countries. For example, because of donations from charitable foundations or university presses, a Sri Lankan library might not receive a Tamil-language publication from India, but nonetheless receive an English publication from the United States as a charitable donation. A library in the state of Oaxaca, Mexico, is more likely to receive a UNESCO publication than to receive *Revista Española de Educación Comparada* (REEC); and a Vietnamese library is more likely to receive CER than *Education Comparée*. This affects the orientation of scholars toward publication in English.

Evidence for the dominance of English can also be seen in the long-term trends in *Compare* and CER. These journals were founded by UK and US educators who sought to publish their contributions to an emerging field, especially articles coming from their annual conferences. Unsurprisingly, most of the contributors to these journals during their first years had English as their mother tongue. In the case of *Compare*, during its early years, nearly all

submissions came from the United Kingdom and few were from authors with non-English-language backgrounds. In recent years, the majority of submissions came from outside the United Kingdom and many were from authors whose mother tongue was not English. Similarly, in 2008, IJED received over 80% of approximately 300 submissions from outside the United Kingdom. The trends are likewise pronounced for CER. In its first 3 years, only five out of 47 published articles had non-English-background authors. By contrast, over the past 3 years, 39 out of 64 published articles were by non-English-background authors, although many had completed their formal training in English.

The Future of Publications in Comparative and International Education

Due to increased printing and international mailing costs, paper publications may well disappear over the next several years. However, regardless of the forms journals take in the future, their scholarly, community-building, and gatekeeping functions will persist. Despite the fact that membership societies continue to proliferate, authors in this field will increasingly publish for readers outside their own nations. Costs to institutions will rise as the numbers of individual subscriptions dwindle, especially for journals without a membership base. More readers will access journals through their institutions, and will download or read specific articles. This will increase authors' access to readers who are interested in the specific topics of their research. At the same time, selective reading could hinder the ability of journals to encourage cross reading. Maintaining a general readership could become increasingly difficult. The challenge for journal editors will be to enliven their publications in order to capture the greatest number of nonspecialist readers, while simultaneously adhering to high standards for scholarship.

See also: Comparative Education: Societies and Associations.

Further Reading

- Borgman, C. L. (2007). *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA: MIT Press.
- Brickman, W. W. (1956). The theoretical foundations of comparative education. *Journal of Educational Sociology* 30(4), 116–125.
- Brickman W. W. and Bereday, G. Z. F. (1957). A new journal in comparative education. *Comparative Education Review* 1(1), 1–2.
- Canagarajah, A. S. (2002). *A Geopolitics of Academic Writing*. Pittsburgh, PA: University of Pittsburgh Press.
- Cook, B. J., Hite, S. J., and Epstein, E. H. (2004). Discerning trends, contours, and boundaries in comparative education: A survey of comparativists and their literature. *Comparative Education Review* 48(2), 123–149.
- Fulbright, J. W. (1963). *Prospects for the West*. Cambridge, MA: Harvard University Press.
- Hess, C. and Ostrom, E. (eds.) (2007). *Understanding Knowledge as a Commons*. Cambridge, MA: MIT Press.
- Higginson, J. H. (1999). The development of a discipline: Some reflections on the development of comparative education as seen through the pages of compare. *Compare* 29(3), 341–350.
- Kazamias, A. M. and Schwartz, K. (1977). Editorial: The state of the art. *Comparative Education Review* 21(2–3), 151–152.
- King, E. (1964). Book review of comparative method in education, by G.Z.F. Bereday. *Comparative Education* 1, 37–38.
- Kranich, N. (2007). Countering enclosure: Reclaiming the knowledge commons. In Hess, C. and Ostrom, E. (eds.) *Understanding Knowledge as a Commons*, pp 85–122. Cambridge, MA: MIT Press.
- Lougee, W. P. (2007). Scholarly communication and libraries unbound: The opportunity of the commons. In Hess, C. and Ostrom, E. (eds.) *Understanding Knowledge as a Commons*, pp 311–322. Cambridge, MA: MIT Press.
- McIntosh, C. (2002). International review of education: A journal of many incarnations. *International Review of Education / Internationale Zeitschrift für Erziehungswissenschaft / Revue Internationale de l'Education* 48, 1–20.
- Raby, R. L. (2007). Fifty years of comparative education review bibliographies: Reflections on the field. *Comparative Education Review* 51(3), 379–398.
- Raby, R. L. (2008). The globalization of journals: A review of the 2007 comparative education review bibliography. *Comparative Education Review* 52(3), 461–475.
- Sutherland, M. B., Watson, K., and Crossley, M. (2008). The British Association for International and Comparative Education (BAICE). In Mason, M. (vol. eds) Masemann, V., Bray, M., and Manzoni, M. (eds.) *Common Interests, Uncommon Goals: Histories of the World Council of Comparative Education Societies and Its Members: Volume 21. CERC Studies in Comparative Education*, pp 155–169. London: Springer.
- Thompson, J. B. (2005). *Books in the Digital Age: The Transformation of Academic and Higher Education Publishing in Britain and the United States*. Cambridge, UK: Polity Press.

Comparative Education: Philosophical Issues and Concepts

M Mason, Hong Kong Institute of Education, Hong Kong, People's Republic of China

C W Evers, University of New South Wales, Sydney, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

In a special issue of *Comparative Education* on 'Philosophy, education and comparative education,' McLaughlin (2004: 475) suggests that comparison in education invites philosophical attention to at least the following four dimensions:

1. what is being compared with what (e.g., teachers, schools, teaching methods, and educational systems in differing cultural, national, and regional contexts);
2. the evaluative basis of comparison (e.g., the norms and principles being invoked in making comparisons);
3. the reasons and motives underlying the comparisons being made (e.g., disinterested scholarly enquiry, a search for insights, etc., to be applied from one context to another); and
4. the methods used in making comparisons (e.g., methods based on natural science, social science, hermeneutic traditions, etc.).

Limitless concepts, issues, themes, contexts, norms, principles, aims, purposes, and other ontological, epistemological, and axiological questions that constitute these four dimensions raise philosophical questions in comparative education, and this article, in providing an overview of and a brief introduction to some of those questions, is necessarily selective in which of these it considers, and brief in the extent to which it does. Following and expanding on Martínez (2003), the article considers the constitution and definition of the field of comparative education in terms of its substantive content or object of study, its method, and its aims and purposes; the differentiation of comparative education from related fields such as international education, educational development, and area studies; and conceptual issues associated with the role of cultures and countries, or nation-states, as units of analysis in comparative education.

The Constitution and Definition of the Field

The definition and distinction of the field of comparative education is perhaps the most intractable philosophical and conceptual debate in the field. It should, at the outset, be stated explicitly that comparative education is clearly not a discipline. If, following Hirst (1974), a discipline is epistemologically characterized by a common object of study, uniformly accepted criteria of truth and methods by

which truth might be approached, a particular conceptual structure, theoretical integration and internal consistency, and common products of knowledge, and a field is unified only by common phenomena of study or common material or practical pursuits, then comparative education is at best characterized in terms of the latter. Epstein (1994: 918) understands the field of comparative education as an interdisciplinary pursuit that applies historical, philosophical, and social science theories and methods to international problems in education. For Broadfoot, comparative education might be constituted yet more widely, as a context, a conceptualization which "allows for the interaction of perspectives arising out of a number of social science disciplines and from a wide range of national backgrounds" (Broadfoot, 1977: 133); it is, thus, a social science perspective or a context for critical cultural analysis (Broadfoot, 1999: 29). For Crossley and Watson (2003: 6), however, such a conceptualization of the field serves as an important reminder that context matters: the study of comparative education warns education policy-makers that major problems lie in any simplistic transfer of educational policy and practice from one sociocultural context to another.

Comparative education is an eclectically constituted field with flexible, porous, and, at times, indiscernible boundaries. Among many others, Kelly *et al.* (1982) have identified the central question of the identity of the field in asking what comparative education is, whether it is a distinct field of study, and whether it is best defined in terms of its method, or substantively in terms of its content or object of study. Garrido (1986) expands the last question as to how comparative education might be defined in terms of its object. Most definitions of comparative education understand its object in terms of the unit of analysis or comparison (Manzon, 2009), and most common among the latter is the comparative study of educational systems. This might commonly be generalized, following Martínez (2003), to the comparative study of educational policy, problems, and processes. Epstein (1994: 918) defines the object of comparative education simply as the trans-societal study of education. For Anderson (1961: 4), comparative education's core substantive content, its central object of study, is the relationship between school and society: comparative education is concerned with the cross-cultural comparison of the structure, operation, aims, methods, and achievements of various educational systems, and the societal correlates of these educational systems

and their elements. Noah and Eckstein (1969) also situate the relationship between school and society as the central object of study in comparative education. Olivera (1988) identifies the proper object of comparative education as the abstract patterns of relationships between two or more educational configurations, to which social scientific comparative methods are applied in order to understand educational phenomena beyond the differences associated with particular societal conditions.

Phillips and Schweisfurth (2007) suggest that:

What distinguishes the work of comparativists from that of other educationists – and what therefore distinguishes the approaches they use to research – is the obvious fact that they are concerned essentially with other cultures/countries. Aside from this concern with education ‘elsewhere’ and the question as to how comparisons might realistically be made, comparativists use all the research methods that other investigators of aspects of education employ in their research . . . , [utilizing additionally] approaches specific to the particular task of comparison. The essential questions are: What methods and conditions are appropriate to ensure adequate understanding of other cultures from a vantage point outside of those cultures? What does the act of comparison consist in and how might it best be undertaken?

Philosophical and conceptual issues associated with cultures and countries, or nation-states, lie, thus, at the heart of comparative education, and are considered briefly in the final section of the article.

For Halls, comparative education involves the study of “education by its various levels, and also systematically researches the historical, social, cultural, political, religious, economic and philosophical forces that partly determine and are partly determined by the character of education systems, and compares the resultant outcomes in two or more systems, areas or even globally” (Halls, 1990: 24).

In one of the most generous and inclusive conceptualizations of the field, Bray and Thomas (1995) suggest, on the other hand, that comparative education need not be limited to comparisons across countries, systems, or cultures. Their *Framework for Comparative Education Analyses* (Bray and Thomas, 1995: 475), well known in the field as the cube (see **Figure 1**), sets out to identify, in order to enable researchers to set down clearly their unit of analysis and comparison, what is being compared, who is being compared, and where are these phenomena being compared. What is being compared includes aspects of education and society that might include the financing of education, curriculum, or teaching methods. Who is being compared includes nonlocational demographic groups that could be defined in terms of gender, age, ethnicity, religion, and the like. Where, or at what geographical or locational level these phenomena are being compared, may be at the level of world regions, countries, systems, schools, classrooms, or even at the individual level.

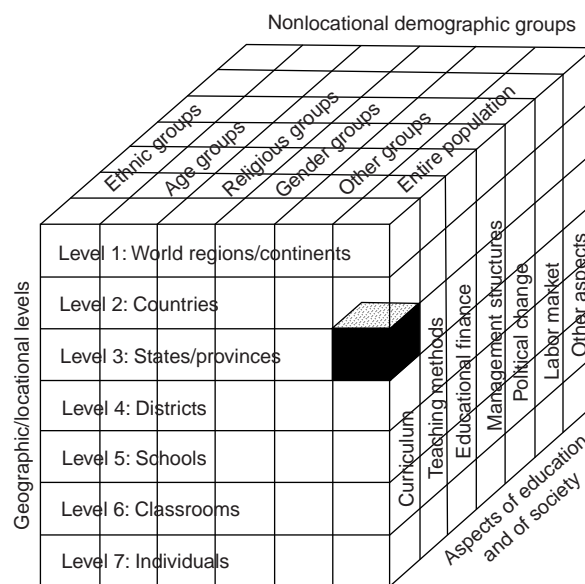


Figure 1 A framework for comparative analysis. From Bray, M. and Thomas, R. M. (1995). Levels of comparison in educational studies: Different insights from different literatures and the value of multilevel analysis. *Harvard Educational Review* 65(3), 475.

In their generosity, Bray and Thomas may, however, have expanded the borders of comparative education to include almost all research in education, perhaps even a single case study that relies, at best, on only implicit comparison, on the type of comparison, Schriewer's (2000: 9) “universal mental operation embedded in everyday social life,” without which it is impossible to distinguish anything from that which it is not. By not limiting the object of comparative education to the study of educational systems, such as Garrido (1986) would have researchers in the field do, the generosity of Bray and Thomas risks the dissolution of the field until it is indistinguishable from the broader field of education studies.

This substantive expansion of the borders of the field, and the associated risk of its dissolution, raises the question whether the field of comparative education can actually be defined substantively. Kazamias and Massialas (1965: 14) have, indeed, suggested that the subject matter of comparative education is “coterminous with the subject matter of education itself,” but intimate that what distinguishes it from the field of education studies is that it “goes beyond the confines of education in one nation, society or group, and uses cross-national or cross-cultural methods and techniques.” This lessening by Bray and Thomas of the principal emphasis of comparative education on comparisons among educational systems opens the way for an argument that comparative education should be defined in terms of its method or of its purpose, rather than in terms of its substantive object, since for them, the object becomes education in general, or at least almost any aspect thereof, rather than educational systems in particular.

Mason (2008) argues that it is not possible to isolate substantively a field of comparative education within or overlapping only in part with the field of education studies, given that the content of the former, as generously described by Bray and Thomas, is not able to be identified by any significant differences from the content of the latter. He concludes accordingly that it is the method specific to comparative education that identifies it as such: "Where comparison is used as an explicit research method, more than the implicit comparison that probably underlies most ways in which we make sense of the world (see Schriewer (2000) above), comparative education is a methodologically distinct aspect of education studies" (Mason, 2008: 1). For him, comparative education is a methodologically defined sub-field of education studies. Most well known in the field for the view that comparative education is defined by its method is Bereday (1964). More recently, Green has claimed that "comparative education needs to compare, and to do this systematically, if it is to avoid the accusation that it too often degenerates into a catalogue of traveller's tales, policy advocacy and opportunistic rationalization of unscientific policy borrowing" (Green, 2003: 95). Garrido (1986), however, cautions against defining comparative education in terms of method alone, since this could lead to an unacceptably wide interpretation of the scope of comparative education (such as is implicitly countenanced by Bray and Thomas (1995)) and its consequent equivalence in substantive terms to the field of education studies (as Mason (2008) has suggested is the case both conceptually and empirically) – and hence to its being indistinguishable from other sub-fields of education studies. For Garrido (1986), definition of the field should be in terms of both method and substantive object, with education systems being the appropriate object of study. Jones (1971), however, suggests that it is in terms of method and purpose that comparative-education research should be defined.

Debates about the identity of the field are not limited, however, to whether it is best defined substantively in terms of its content or object of study, or in terms of its method. The field has also been definitionally constituted in terms of its purposes. These purposes fall mainly into three dimensions: theoretical, pragmatic, and normative. Arnone (2007) describes a scientific (or theoretical) dimension that aims to formulate generalizable propositions about school systems and the economic, political, social, and cultural contexts in which they are situated; a pragmatic dimension associated with learning from the education systems of other societies to the end of improving educational policy and practice in the researcher's own society; and a normative dimension frequently oriented toward international understanding and peace. For Martínez (2003), these normative purposes are critical and emancipatory. (These critical and emancipatory purposes have had much to do with the interests of many comparative-education researchers working in the fields of international education

and educational development, closely related fields considered in the section titled 'The distinction of comparative education from other, related fields'.) It is no coincidence that the three dimensions identified by Arnone, among others, reflect the three knowledge-constitutive cognitive interests of Habermas (1971: 308): the technical cognitive interest that constitutes the empirical-analytic sciences; the practical cognitive interest that constitutes the historical-hermeneutic sciences; and the emancipatory cognitive interest that constitutes the critically oriented sciences.

Kandel (1933: xix) places the theoretical purposes of comparative education in the service of the practical, seeing the value of a comparative approach to educational problems "in an analysis of the causes which have produced them, in a comparison of the differences between the various systems and the reasons underlying them, and . . . in a study of the solutions attempted." Comparative education's purpose is "to discover the differences in the forces and causes that produce differences in educational systems" (Kandel, 1936: 406). For Bereday, the principal justification for comparative-education research is intellectual. "Knowledge for its own sake is the sole ground upon which comparative education needs to make a stand" (Bereday, 1964: 5). For Holmes, similarly, theoretical purposes in comparative-education research are prior to any pragmatic purposes: "One aim of comparative education is theoretical. It is to improve our understanding of education as such; and in particular of our own national problems in education" (Holmes, 1971: x–xi). A frequently articulated position in the field holds that it is concerned with a better understanding of one's own system of education by comparing it with others. Lauwerys (1959: 292) has noted "how hard it is for a teacher anywhere to explain his own school system to foreigners: often he considers as too obvious to need explaining precisely the most curious and puzzling features."

The identification of pragmatic purposes has been at the heart of comparative education since the earliest days when such a field could be identified. Both Jullien (1817) and Sadler (1900/1964) place the international transfer of policies and practices as a central, if not the central, purpose of comparative education, if from different methodological perspectives: positivist in Jullien's case and historical or relativist in the case of Sadler. King (1965: 148) identifies the implicit purpose of research in the field as the improvement of school systems. Gu has moved from a strongly ameliorative and reformist view of the purposes of the field to the view that its purposes lie in the identification of "general and specific principles in educational development [that might] act as a reference point for educational reform in our own country or region" (Gu, 2001: 240).

Several scholars have also placed normative purposes at the heart of comparative education, whether these purposes be ameliorative or critical. Masemann's (2003) position with respect to which theoretical perspective

most appropriately situates comparative ethnographic research in its wider context is located in the paradigm of conflict theory. She calls for a critical ethnography (an anthropological methodology informed by critical theory) that avoids the assumptions of neutrality and objectivity of functionalist and positivist approaches: “a critical or neo-Marxist approach is necessary to delineate the connections between the microlevel of the local school experience and the macrolevel of structural forces at the global level that are shaping the ‘delivery’ and experience of education in every country” (Masemann, 2003: 115). Mason argues that “comparative education is best conceptualized as a critical social science, incorporating an emancipatory interest focused on the distribution of power and its associated attributes: economic wealth, political influence, cultural capital, social prestige and privilege, and the like” (Mason, 2007: 196). He suggests that comparative-education researchers working in the domain of educational development identify the axes along which educational and other goods are differentially distributed (Mason, 2007: 196) and compare along and across one or more of these axes to the end of exposing educational inequities, inequalities, and injustices.

With respect to these critical purposes, it is worth nothing that Broadfoot describes a deep methodological divide between qualitative and quantitative methodological approaches to comparative-education research (Broadfoot, 2000: 360). As it is difficult to escape the value-laden nature of problems, methods, and conclusions in much comparative-education research, she accordingly calls (Broadfoot, 2000) for a more critical, theoretically informed, social-science perspective in the field, one steeped in a greater self-critical awareness. She suggests that “comparative educationists . . . need themselves to be willing to engage in fundamental debates about values; about the nature ‘of the good life’ and about the role of education and learning in relation to this” (Broadfoot, 2000: 370). For Cowen (2000), these questions imply a reading of the world by comparative-education researchers that involves wide-ranging cultural, historical, and political interpretations.

The Distinction of Comparative Education from Other Related Fields

Comparative Education and International Education

The fact that there is little consensus among researchers in the field about these distinctions notwithstanding, Bray (2007: 51–53) has identified four different definitions of international education that distinguish it from comparative education. Among these are Epstein’s view that, while comparative education is “a field of study that applies historical, philosophical and social science theories

and methods to international problems in education,” international education “tends to be less scientific than comparative education,” and is concerned less with analysis and study and more with policy implementation and practice (Epstein, 1968: 376–377). Postlethwaite (1988: xvii) has classified within international education those studies that “do not compare, but rather describe, analyse or make proposals for a particular aspect of education in one country other than the author’s own country.” For Epstein, international education is concerned with descriptive information about different educational systems and engaged in policymaking in the area of international exchange and understanding (Epstein, 1994: 918), while comparative education is concerned with explaining and analyzing the variance in educational systems and processes across societal contexts, and how those systems and processes relate to their wider contexts.

Ferrer (2002: 197) casts this distinction as the one between the theoretical and the applied, the analytical and the descriptive, and the explicative and the informative.

Wilson (1994: 452) has challenged Epstein’s relegation of international educators to a merely descriptive role. For him, “international educators originated – and continue to practice – the melioristic trend more prominently associated with comparative education; that is, the improvement of national educational systems by the addition of models, practices, innovations, and the like borrowed or transferred from other national educational systems.” Wilson locates this melioristic purpose in the work of “personnel in bilateral, multilateral, and non-governmental organizations engaged in national studies, usually related to a development project” (Wilson, 1994: 455). Many in the field see this purpose as generating the links among comparative, international, and development education, the latter frequently understood as international development aid (e.g., Crossley, 1999; Watson, 1999). This reflects Anweiler’s (1977: 109–110) contextualization of development education as a sub-field of international education. International education, in this sense, includes the study of the work of international educational organizations, such as the United Nations Educational, Scientific and Cultural Organization (UNESCO), that is typically focused on educational development in poorer countries and regions.

The term international education has also been identified with the work of international schools, typically in the research of Hayden and Thompson (see Cambridge and Thompson, 2004; Hayden and Thompson, 2001). Cambridge and Thompson describe two interpretations of international education: the first informed by educational purposes associated with the development of international attitudes, international awareness, international-mindedness, and international understanding; and the second associated with the global portability, recognition, and certification of educational qualifications (Cambridge and Thompson, 2004: 173).

Comparative Education and Educational Development

There is a marked discrepancy between the term comparative education and the number of articles published in the field's journals that are hardly or, at best, only implicitly comparative. Among many others, Halls (1990: 26) has identified this as a problem in the field. For Epstein (1994), however, this is not a problem, given his view that implicit comparison refers to a type of inquiry that may be focused on one particular society or historical period, but whose analytical scheme can be extended to problems in other societies and periods. Lauwerys has noted that, frequently, comparative studies produce such general statements, with the comparative element downplayed to an implicit level. "Studies can use the comparative method without being constantly and monotonously explicitly comparative. Comparative method can be the ladder by which one rises to a height from which a general view is gained; the scaffolding needed to build a house. However, there is no need to display the scaffolding ostentatiously, when the house stands solid and is provided with a staircase" (Lauwerys, 1959: 292).

That one of comparative education's frequently cited purposes is the development of education in poorer regions and countries of the world is a major reason for this discrepancy: many articles in the field's journals are less explicitly comparative and more concerned with issues of international educational development. Notwithstanding the fact that much of the most visible comparative-education research today has to do with large cross-national studies of student learning such as those undertaken by the International Association for the Evaluation of Educational Achievement (IEA) and the Organization for Economic Cooperation and Development (OECD)'s Program for International Student Assessment (PISA), Mason has suggested that "comparative education research . . . has its most worthwhile contribution to make [to the end of educational equity] in the domain of educational development" (Mason, 2007: 196).

For Parkyn (1977: 88–89), the distinction between comparative education and development education is one of purpose: while the purpose of comparative education is to increase our understanding of the relationship between education and development through a comparative study of trans-societal factors, the latter is concerned with the economic, political, and development needs of developing countries. He emphasizes that development education must rest on a foundation of comparative education (Parkyn, 1977: 90), not least because of the former's dependence on a holistic understanding of educational systems within their economic, political, social, cultural, and development contexts. One purpose of comparative-education researchers working in the domain of educational development is the "production of information

and plans to assist policymakers, the development of appropriate educational methods and techniques, and the training of personnel to implement programmes" (Halls, 1990: 24).

Comparative Education and Foreign Education, or Area Studies

Foreign education, a term used less frequently than previously, refers to the study of education (systems) in countries other than that of the researcher but perhaps not for comparative purposes – merely for description, analysis, and explanation. These days it is more frequently referred to as area studies, within one country or a particular region. Successive editors of the influential *Comparative Education Review*, such as Epstein (see Epstein, 1994: 922) and Rust, have argued that such studies could also be construed as comparative if they permit "readers outside that educational system to understand how its problems might illuminate problems elsewhere" or if "they rely on or test more general theories" (Rust, 2001: iv).

Cultures and Countries, or Nation-States, as Units of Analysis

(This explication of the conceptual issues associated with comparing education across cultures, countries, and nation-states draws substantially on the first-half of Mason's *Comparing Cultures* (Mason, 2007). For a consideration of the related methodological issues, which is beyond the scope of this article, the interested reader may consult the latter part of Mason (2007).)

McLaughlin (2004: 475) identifies a number of themes, recently and frequently considered by researchers in comparative education that are rich in their philosophical implications: globalization, postcolonialism, indigenous education, democracy, and citizenship. He also identifies (McLaughlin, 2004: 476) a number of educational contexts in which such comparison is situated, which are similarly rich in their philosophical implications: cultural, anthropological, political, and religious. Among the most common in the field, and also among the most complex, is the question of comparing education across cultures. "Culture, in comparative analysis and understanding, and certainly in national systems of education," claims Alexander (2000: 30), "is all." Limitations of space forbid philosophical explication of the other contexts and themes listed by McLaughlin, but this article would not be complete without at least a brief consideration of culture, of national culture, and of national culture under conditions of increased rates of globalization, in comparative-education research.

The summary claim of Phillips and Schweisfurth (2007) that comparative education is essentially concerned with

other cultures/countries raises at least two further conceptual questions, about culture, and about the nation-state. With respect, briefly, to the latter, Arnove and Torres (1999), Dale (2000), Ferrer (2002), and Green (2003) are among many scholars who have suggested that increasing rates of globalization have placed in question the role of the nation-state as a primary unit of analysis in educational comparisons. Green, for example, has suggested that the processes associated with globalization have created new “educational spaces which belong exclusively to neither nations nor systems” (Green, 2003: 95). Khôi (1981) describes supranational, international, and intranational classifications of education systems for comparison. Halls (1990) describes comparisons among nation-states, world systems, ideological systems, world regions; comparisons in terms of levels of economic prosperity; and local or intranational comparisons. As has been described above, the cube of Bray and Thomas (1995) lessens most appropriately, at least in the context of the processes associated with increasing rates of globalization, the emphasis on comparisons among education systems at the level of the nation-state by expanding the range of comparisons down to micro-social units such as schools, classrooms, teachers, students, and even individual lessons.

The Concept of Culture in Comparative-Education Research

Few comparative-education researchers would deny that cultural factors influence many aspects of education, but most would flinch from asserting precisely what these factors are. Such factors are notoriously difficult to isolate, and such assertions are often tenuous at best, given how easy it is not only to overstate the influence of a particular cultural factor in a complex world, but also to get it wrong. Perhaps worse than this, researchers who attempt to describe the influence of cultural factors on education face possible accusations of stereotyping, of treating culture as monolithic, and of overstating its influence in a hybrid world characterized by complex interactions and influences.

Comparisons of education across cultures are common, with the difficulty of generalization among the most intractable questions in this domain. Two well-known examples are the cross-national studies of educational achievement conducted under the auspices of the IEA and PISA. Secondary analysis of these results frequently involves a search for cultural factors associated with educational achievement – the immediately obvious first slippage being the one from country to culture (and indeed, if the adjective cross-national is used, from nation to country). The assumption that nation, country, and culture are synonymous is, of course, simply wrong. To assume that culture is a monolithic and discrete entity is equally wrong. Culture is not a fixed entity that shapes the lives of individuals. It is more accurate to speak of

a dialectical process between people and their social environments which also involves the shaping of the culture by those people as they manipulate its conventional symbols to create new meanings. Culture is, in other words, not a club, along with membership of which go certain attributes of membership. Culture functions more as a productive force constituted by a relatively amorphous aggregation of loosely bounded factors that both influences the lives of the individuals who share in it and is, in turn, influenced by those individuals.

There are at least two definitions of culture that are of interest to comparative-education researchers, the first underlying a research interest in material production, the second in symbolic systems, each of which overlaps the other substantially in much comparative-education research. With respect to the former, Williams (1985: 91) notes that “in archaeology and in cultural anthropology the reference to culture or a culture is primarily to *material* production,” while with respect to the latter, “in history and cultural studies the reference is primarily to *signifying* or *symbolic* systems” Williams (1985: 91).

The first, commonly understood as the anthropological definition, indicates “a particular way of life, whether of a people, a period, a group, or humanity in general” (Williams, 1985: 90). This way of life would include the shared values and meanings common to members of the group. Drawing on Keesing’s position that culture is “concerned with actions, ideas and artefacts which individuals in the tradition concerned learn, share and value” (Keesing, 1960: 25), Masemann’s anthropological approach to culture (Masemann, 2003: 116–117) assumes that:

culture refers to all aspects of life, including the mental, social, linguistic and physical forms of culture. It refers to ideas people have, the relationships they have with others in their families and with larger social institutions, the languages they speak, and the symbolic forms they share, such as written language or art/music forms. It refers to their relationship with their physical surroundings as well as the technology that is used in any society, [and] ... it expresses the value system(s) of a particular society or group.

The second definition of culture derives from its anthropological definition, and also refers to shared meanings within groups, but differs in emphasis from the former by focusing more on “the symbolic dimension, and on what culture *does* rather than what culture *is*” (Bocock, 1992: 232). Here, in cultural studies (more than in cultural anthropology), culture is less importantly a distinctive way of life as understood, for example, by its material artifacts, and more importantly “the set of practices by which meanings are produced and exchanged within a group” (Bocock, 1992: 233). At the heart of these practices lies language, because the sharing of a common language system enables people to communicate

meaningfully with one another. Language is here understood broadly, to include all sign and symbol systems through which meaning is produced: “any system of communication which uses signs as a way of referencing objects in the real world; it is this process of *symbolisation* which enables us to communicate meaningfully about the world” (Bocock, 1992: 233). These sign-and-symbol systems are most commonly understood as the words of a language, but they also include material objects. It is not least in the interpretation of the significance of the material object that this symbolic understanding of culture differs from, or at least extends, the anthropological understanding of culture.

In cultural anthropology, then, culture is understood as shared meanings and ways of life; in cultural studies and its associated fields, culture is understood as the practices which produce meaning (Bocock, 1992: 234). Again, the second draws on the first, and the first is also interested in the concerns of the second. It is more a matter of difference in emphasis: in the first, on the substantive contents of culture as a whole way of life; in the second, on the ways in which cultural practices produce meaning for those who share those practices. The approach to the analysis of culture typical of the second looks for the ways in which meaning is produced by the arrangement, the pattern, and the symbolic structure of an event (Bocock, 1992: 235).

National Culture in Modern Societies

Perhaps the most common expression of cultural identity in modernity is found in what is widely understood as national culture. While in premodern societies, cultural identity might typically have been constructed in terms of one's tribe, ethnicity, religion, or region, with the nation-state, the dominant political entity in modernity, these identities have gradually given way to a national cultural identity. Nation (as in national, associated with a country) and culture are, after all, often conflated in comparative-education research that attempts to identify the cultural factors associated with, say, educational achievement. The question then turns to the meaning of national culture. Here, we follow Hall (1994: 292), for whom a national culture is a discourse – “a way of constructing meanings which influences and organizes both our actions and our conception of ourselves.” National identity, argues Anderson (1983), is no more than an imagined community. That does not mean that national identity and culture are without real and substantial consequences; but before comparative-education researchers undertake comparisons across cultures, they should consider not only the ways in which the discourse of national culture is represented, but also the power of those representations to win national allegiance and to define cultural identity.

National cultural identity is best understood with reference to Hall, who points out that

national identities are not things we are born with, but are formed and transformed within and in relation to *representation*. We only know what it is to be ‘English’ because of the way ‘Englishness’ has come to be represented, as a set of meanings, by English national culture. It follows that a nation is not only a political entity but something which produces meanings – *a system of cultural representation*. People are not only legal citizens of a nation; they participate in the *idea* of the nation as represented by its national culture . . . National cultures construct identities by producing meanings about ‘the nation’ with which we can *identify*; these are contained in the stories which are told about it, memories which connect its present with its past, and images which are constructed of it. (Hall, 1994: 292–293, *emphases original*)

National culture emerged with and helped to shape modernity by gradually displacing (but of course, not entirely) the premodern discourses of identity mentioned earlier: tribal, ethnic, religious, and regional.

The origins of these representations, these narratives that constitute and reflect the discourse of national culture, lie in and may be constructed through:

- “the narratives of the nation, as it is told and retold in national histories, literatures, the media and popular culture,” which “provide a set of stories, images, landscapes, scenarios, historical events, national symbols and rituals which stand for, or represent, the shared experiences, sorrows, and triumphs and disasters which give meaning to the nation” (Hall, 1994: 293), and which “make up the threads which bind us invisibly to the past” (Schwarz, 1986: 155);
- the emphasis on “origins, continuity, tradition and timelessness” (Hall, 1994: 294), which represents national identity as primordial, “in the very nature of things” (Gellner, 1983: 48);
- the invention of tradition: as Hobsbawm and Ranger (1983: 1) point out, traditions which appear or claim to be old are often quite recent and sometimes invented;
- the creation of a foundational myth, one which “locates the origin of the nation, the people and their national character so early that they are lost in the mists of, not ‘real’, but ‘mythic’ time” (Hall, 1994: 295; Hobsbawm and Ranger, 1983: 1); and
- the symbolic grounding of national identity on the idea of a “pure, original people or ‘folk’” (Hall, 1994: 295; Gellner, 1983: 61).

Comparative-education researchers would be well cautioned about the shallowness and the arbitrariness of the foundations of cultural identity by understanding national cultural identity as more constructed than natural, more discursive than material.

Beyond these, questions about the rather arbitrary history and radically constructed nature of national cultural identity

are a further problem: whether national identities really are as unified, coherent, consistent, and homogeneous as appears in these representations of national culture. They are obviously not. As Hall (1994: 297) points out, "modern nations are all cultural hybrids." Most modern nations have, after all, been born out of violent conquest of one or more groups by another. National cultural identity is often constructed on a specious notion of race, marking as different those of different racial groups. National identity is also often strongly gendered, excluding women from its patriarchal norms. Class is another powerful divider, and it is almost without exception the cultural capital of the elite groups in a society that represents the norm, that constitutes what is to be emulated and sought by all. This generalization of the cultural norms of a society's elite groups to the level of national cultural identity does, what Bourdieu calls, symbolic violence to the representations espoused in the cultural identity of other groups in society, and all too frequently contributes to weak comparative-education research that misses such marginalized representations. Differences in language, geographical region, tradition, religion, customs, and the like constitute further lines marking difference and exclusion. While national cultural mythology works to draw together the different identities and local communities of which a nation-state is constituted, to make culture and polity congruent under the same political roof (Gellner, 1983: 43), and to paper over the cracks that divide those who identify with Anderson's imagined community from those who are not subsumed under the state's hegemony, it is a good comparative researcher who would be able to represent these subtleties and nuances to an extent that might be called true and fair.

The Influence of Increasing Rates of Globalization on National Culture

We have argued that national culture is somewhat arbitrary, probably best understood as myth, and not particularly successful at masking deep and cross-cutting social divisions. Increasing rates of globalization have complicated matters further. We turn in conclusion briefly to the consequences of globalization and its associated processes for national cultural identity. What might have been sedimented and layered into the accepted truths of national cultural identity has been marbled by processes associated with accelerated globalization. The cultural hybridity of the modern nation-state, masked as a homogeneous unity by the myths of national culture, is being exacerbated almost to the point of the displacement of the national culture by the processes of globalization. Of most relevance here are three such processes: first, national cultural identities are being rendered yet more tenuous than they already are; second, local and particular identities are being strengthened as a consequence of resistance to the processes of globalization; and third, these new hybrid

identities are becoming, at the expense of national cultural identities, increasingly visible. Perhaps the main conclusion to be drawn from the discussion in this section is that the anthropological definition of culture starts to look methodologically suspect in all but the most homogeneous and isolated of cultures, if indeed any exists anymore. It is perhaps to cultural studies and to sociological more than anthropological understandings of culture in contemporary society that researchers need to turn for making a comparison of education across cultures.

Bibliography

- Alexander, R. J. (2000). *Culture and Pedagogy: International Comparisons in Primary Education*. Oxford: Blackwell Publishers.
- Anderson, B. (1983). *Imagined Communities: Reflections on the Origins and Spread of Nationalism*. London: Verso.
- Anderson, C. A. (1961). Methodology of comparative education. *International Review of Education* 7(1), 1–23.
- Anweiler, O. (1977). Comparative education and the internationalization of education. *Comparative Education* 13(2), 109–114.
- Arnone, R. F. (2007). Introduction: Reframing comparative education. In Arnone, R. F. and Torres, C. A. (eds.) *Comparative Education: The Dialectic of the Global and the Local*, pp 1–20. Lanham, MD: Rowman & Littlefield Publishers.
- Arnone, R. F. and Torres, C. A. (eds.) (1999). *Comparative Education: The Dialectic of the Global and the Local*, 1st edn. Lanham, MD: Rowman and Littlefield.
- Bereday, G. Z. F. (1964). *Comparative Method in Education*. New York: Holt, Rinehart and Winston.
- Bocock, R. (1992). The cultural formations of modern society. In Hall, S. and Gieben, B. (eds.) *Formations of Modernity*, pp 229–274. Cambridge and Oxford: Polity Press, Basil Blackwell and The Open University.
- Bray, M. (2007). International and comparative education: Boundaries, ambiguities and synergies. In Hayden, M. C. and Thompson, J. J. (eds.) *The SAGE Handbook of Research in International Education*, pp 51–56. London: Sage.
- Bray, M. and Thomas, R. M. (1995). Levels of comparison in educational studies: Different insights from different literatures and the value of multilevel analysis. *Harvard Educational Review* 65(3), 472–490.
- Broadfoot, P. (1977). The comparative contribution-research perspective. *Comparative Education* 13(2), 133–137.
- Broadfoot, P. (1999). Not so much a context, more a way of life? Comparative education in the 1990s. In Alexander, R., Broadfoot, P., and Phillips, D. (eds.) *Learning from Comparing: New Directions in Comparative Educational Research*, vol. 1, pp 21–31. Oxford: Symposium Books.
- Broadfoot, P. (2000). Comparative education for the 21st century: Retrospect and prospect. *Comparative Education* 36(3), 357–371.
- Cambridge, J. and Thompson, J. (2004). Internationalism and globalization as contexts for international education. *Compare: A Journal of Comparative Education* 34(2), 161–175.
- Cowen, R. (2000). Comparing futures or comparing pasts? *Comparative Education* 36(3), 333–342.
- Crossley, M. (1999). Reconceptualising comparative and international education. *Compare: A Journal of Comparative Education* 29(3), 249–267.
- Crossley, M. and Watson, K. (eds.) (2003). *Comparative and International Research in Education: Globalisation, Context and Differences*. London: Routledge/Falmer.
- Dale, R. (2000). Globalization: A new world for comparative education? In Schriewer, J. (ed.) *Discourse Formation in Comparative Education*, pp 87–109. Frankfurt: Peter Lang.
- Epstein, E. H. (1968). Letter to the editor. *Comparative Education Review* 12(3), 376–378.

- Epstein, E. H. (1994). Comparative and international education: Overview and historical development. In Husén, T. P. and Neville, T. (eds.) *The International Encyclopedia of Education*, pp 918–923. Oxford: Pergamon Press.
- Ferrer, F. (2002). *La educación comparada actual*. Barcelona: Editorial Ariel, S.A.
- Garrido, J. L. G. (1986). *Fundamentos de Educación Comparada*, 3rd edn. Madrid: Editorial Dykinson S.L.
- Gellner, E. (1983). *Nations and Nationalism*. Oxford: Blackwell.
- Green, A. (2003). Education, globalisation and the role of comparative research. *London Review of Education* 1(2), 83–97.
- Gu, M. (2001). Comparative education in China: Name and reality. In *Education in China and Abroad: Perspectives from a Lifetime in Comparative Education*. Hong Kong: Comparative Education Research Centre.
- Habermas, J. (1971). *Knowledge and Human Interests*, Shapiro, J. (Transl.). Boston, MA: Beacon Press.
- Hall, S. (1994). The question of cultural identity. In Hall, S., Held, D., and McGrew, T. (eds.) *Modernity and Its Futures*, pp 273–325. Cambridge and Oxford: Polity Press, Basil Blackwell and The Open University.
- Halls, W. D. (ed.) (1990). *Comparative Education: Contemporary Issues and Trends*. London: Jessica Kingsley Publishers/UNESCO.
- Hayden, M. C. and Thompson, J. J. (eds.) (2001). *International Education: Principles and Practice*. London: Kogan Page.
- Hirst, P. H. (1974). Liberal education and the nature of knowledge. In Hirst, P. H. (ed.) *Knowledge and the Curriculum: A Collection of Philosophical Papers*, pp 30–53. London: Routledge & Kegan Paul.
- Hobsbawm, E. and Ranger, T. (eds.) (1983). *The Invention of Tradition*. Cambridge: Cambridge University Press.
- Holmes, B. (1971). Foreword. In *Comparative Education: Purpose and Method*, pp 0–11. St. Lucia: University of Queensland Press.
- Jones, P. E. (1971). *Comparative Education: Purpose and Method*. St. Lucia: University of Queensland Press.
- Jullien, M.-A. (1817). *Esquisse et Vues Préliminaires d'un Ouvrage sur l'Éducation Comparée*. Paris: Société Établie à Paris pour l'Amélioration de l'Enseignement Élémentaire (Reprinted 1962). Genève: Bureau International d'Éducation.
- Kandel, I. L. (1933). *Studies in Comparative Education*. Boston: Houghton Mifflin.
- Kandel, I. L. (1936). Comparative education. *Review of Educational Research* 6(4), 400–416.
- Kazamias, A. M. and Massialas, B. G. (1965). *Tradition and Change in Education: A Comparative Study*. Englewood Cliffs, NJ: Prentice-Hall.
- Keessing, F. M. (1960). *Cultural Anthropology: The Science of Custom*. New York: Rinehart.
- Kelly, G. P., Altbach, P. G., and Arnove, R. F. (1982). Trends in comparative education: A critical analysis. In Altbach, P. G., Arnove, R. F., and Kelly, G. P. (eds.) *Comparative Education*, pp 505–533. New York: Collier Macmillan.
- Khôi, L. T. (1981). *L'éducation comparée*. Paris: Armand, Colin.
- King, E. J. (1965). The purpose of comparative education. *Comparative Education* 1(3), 147–159.
- Lauwerys, J. A. (1959). The philosophical approach to comparative education. *International Review of Education* 3, 281–298.
- Manzon (2009). *Comparative Education: The Construction of a Field*. Unpublished doctoral thesis, The University of Hong Kong.
- Martínez, M. J. (2003). *Educación Comparada: nuevos retos, renovados desafíos*. Madrid: Editorial La Muralla, SA.
- Masemann, V. L. (2003). Culture and education. In Arnove, R. F. and Torres, C. A. (eds.) *Comparative Education: The Dialectic of the Global and the Local*, 2nd edn. Lanham: Rowman & Littlefield Publishers.
- Mason, M. (2007). Comparing cultures. In Bray, M., Adamson, B., and Mason, M. (eds.) *Comparative Education Research: Approaches and Methods*. Hong Kong: Comparative Education Research Centre and Springer.
- Mason, M. (2008). What is comparative education, and what values might best inform its research? *Presidential Address to the Annual Conference of the Comparative Education Society of Hong Kong*.
- McLaughlin, T. H. (2004). Education, philosophy and the comparative perspective. *Comparative Education* 40(4), 471–483.
- Noah, H. J. and Eckstein, M. A. (1969). *Toward a Science of Comparative Education*. New York: Macmillan.
- Olivera, C. E. (1988). Comparative education: Towards a basic theory. *Prospects* 18(2), 167–185.
- Parkyn, G. W. (1977). Comparative education research and development education. *Comparative Education* 13(2), 87–93.
- Phillips, D. and Schweisfurth, M. (2006). *Comparative and International Education. An Introduction to Theory, Method and Practice*. London: Continuum International Publishing Group.
- Rust, V. D. (2001). Editorial. *Comparative Education Review* 45(3), 3–5.
- Sadler, S. M. (1900/1964). How far can we learn anything of practical value from the study of foreign systems of education? *Comparative Education Review* 7(3), 307–314.
- Schriewer, J. (2000). Comparative education methodology in transition: Towards a science of complexity? In Schriewer, J. (ed.) *Discourse Formation in Comparative Education*, pp 3–52. Frankfurt am Main: Peter Lang.
- Schwarz, B. (1986). Conservatism, nationalism and imperialism. In Donald, J. and Hall, S. (eds.) *Politics and Ideology*. Milton Keynes: Open University Press.
- Watson, K. (1999). Comparative educational research: The need for reconceptualisation and fresh insights. *Compare: A Journal of Comparative Education* 29(3), 233–248.
- Williams, R. (1985). *Keywords: A Vocabulary of Culture and Society*. New York: Oxford University Press.
- Wilson, D. N. (1994). Comparative and international education: Fraternal or Siamese twins? A preliminary genealogy of our twin fields. *Comparative Education Review* 38(4), 449–486.

Further Reading

- Bray, M., Adamson, B., and Mason, M. (eds.) (2007). *Comparative Education Research: Approaches and Methods*. Hong Kong: Comparative Education Research Centre & Springer.

Comparative Education: Societies and Associations

M Bray, The University of Hong Kong, China

© 2010 Elsevier Ltd. All rights reserved.

This article focuses on professional societies and associations in the field of comparative education. Most of these bodies attract members as individuals, although some permit institutional membership. The societies and associations provide opportunities for networking through conferences and other activities, and many publish academic journals. Most of the world's national, regional, and language-based professional societies for comparative education are members of the umbrella World Council of Comparative Education Societies (WCCES).

The Oldest Societies

The first society in the field was established in 1956 in the USA. At that time, it was called the Comparative Education Society (CES), but in 1968 it became the Comparative and International Education Society (CIES). It is the largest body in the field, with over 2000 members and a highly-reputed journal titled *Comparative Education Review*. Although the society is registered in the USA, it has many members from other countries. The CIES holds annual conferences, usually in the USA, although occasionally in neighboring countries such as Canada and Mexico.

The second-oldest society is a regional body, the Comparative Education Society in Europe (CESE). This society was created in 1961 as an organization of individual membership, open to practitioners and scholars of comparative education from Europe and beyond. In due course, CESE created sections in Great Britain and Germany which became independent societies. The society organizes biennial conferences, and aims to do so each time in a different European location.

Creation of these two societies was followed by the establishment of three more during the 1960s. The Japan Comparative Education Society (JCES) founded in 1965, and, with a membership of approximately 900, is one of the largest academic societies in Japan in the field of education. It organizes an annual conference and publishes a journal (in Japanese) titled *Comparative Education Studies*.

The Comparative and International Education Society of Canada (CIESC), also known as the *Société canadienne d'éducation comparée et internationale* (SCECI), was founded 2 years after the JCES. As a society in a country which is officially bilingual in English and French, from the outset it has been a bilingual organization. The society organizes an annual conference and operates a journal titled *Canadian*

and International Education or *Éducation canadienne et internationale*.

The Korean Comparative Education Society (KCES) was founded in 1968 as an independent professional society. Two years later, it became part of the Korean Society for the Study of Education (KSSE) as its research section for comparative education, but it retained its original name to maintain a distinct identity. The KCES has organized multiple conferences and seminars, and publishes a journal (in Korean) titled *Korean Journal of Comparative Education*.

The World Council of Comparative Education Societies

In 1970, representatives of the five societies mentioned above came together in Ottawa, Canada, for the First World Congress of Comparative Education Societies. During that event, the WCCES was created as an umbrella body to bring together the various national and regional societies.

The leading figure in the WCCES was Joseph Katz, of Canada's University of British Columbia. Katz had been president of the CES in the USA, and had been the principal architect of the CIESC in Canada. The United Nations Educational, Scientific and Cultural Organization (UNESCO) played a significant role in the preparations, and one of the seven resolutions at the 1970 Congress was that the WCCES should seek consultative status with UNESCO.

Over time, the WCCES gained more members as additional societies were formed and sought affiliation. In 2009, the WCCES had 37 member societies (**Table 1**), of which some were national (e.g., for Cuba and Greece), some were subnational (e.g., for Hong Kong), some were regional (e.g., for the Mediterranean and Southern Africa), and two were language based (for speakers of French and Dutch). The number of member societies at the time of preparation of this table was the largest that it had ever been during the history of the WCCES. However, growth had not been entirely linear, and some societies which had existed at previous points in history had become defunct. Societies in this category included the ones in Colombia, Nigeria, and Portugal.

The statutes of the WCCES state that its broad goals are:

Table 1 Member societies of the WCCES (2009)

<i>Asociación de Pedagogos de Cuba (Sección de Educación Comparada)</i> (APC-SEC)
<i>Association française pour le développement de l'éducation comparée et des échanges</i> (AFDECE)
<i>Association francophone d'éducation comparée</i> (AFEC)
Australian and New Zealand Comparative and International Education Society (ANZCIES)
British Association for International and Comparative Education (BAICE)
Bulgarian Comparative Education Society (BCES)
Chinese Comparative Education Society (CCES)
Chinese Comparative Education Society-Taipei (CCES-T)
Comparative Education Section of the Czech Pedagogical Society (CES-CPS)
Comparative Education Society of Asia (CESA)
Comparative Education Society in Europe (CESE)
Comparative Education Society of Hong Kong (CESHK)
Comparative Education Society of India (CESI)
Comparative Education Society of the Philippines (CESP)
Comparative and International Education Society (CIES)
Comparative and International Education Society of Canada (CIESC)
Council on Comparative Education of Kazakhstan (CCEK)
Egyptian Comparative Education and Educational Administration Society (ECEEAS)
Greek Comparative Education Society (GCES)
Hungarian Pedagogical Society (Comparative Education Section) (HPS-CES)
Israel Comparative Education Society (ICES)
Japan Comparative Education Society (JCES)
Korean Comparative Education Society (KCES)
Mediterranean Society of Comparative Education (MESCE)
<i>Nederlandstalig Genootschap voor Vergelijkende Studie van Opvoeding en Onderwijs</i> (NGVO)
Nordic Comparative and International Education Society (NOCIES)
Polish Comparative Education Society (PCES)
Russian Council of Comparative Education (RCCE)
<i>Sektion International und Interkulturell Vergleichende Erziehungswissenschaft in der Deutschen Gesellschaft für Erziehungswissenschaft</i> (SIIVEDGE)
<i>Sezione Italiana della CESE</i> (SICESE)
<i>Sociedad Argentina de Estudios Comparados en Educación</i> (SAECE)
<i>Sociedad Española de Educación Comparada</i> (SEEC)
<i>Sociedad Mexicana de Educación Comparada</i> (SOMECE)
<i>Sociedade Brasileira de Educação Comparada</i> (SBEC)
Southern African Comparative and History of Education Society (SACHES)
Turkish Comparative Education Society (TUKED)
Ukraine Comparative Education Society (UCES)

- to advance education for international understanding in the interests of peace, intercultural cooperation, mutual respect among peoples, and observance of human rights; and
- to improve education systems so that the right of all to education may be more fully realized.

The statutes then specify two professional aims, namely:

- to promote the study of comparative and international education throughout the world and enhance the academic status of this field; and
- to bring comparative education to bear on the major educational problems of the day by fostering cooperative action by specialists from different parts of the world.

Then the statutes indicate that to achieve these aims, the WCCES will:

- encourage the teaching of, and research in, comparative education;
- promote interdisciplinary collaboration in the development of comparative approaches to the study of educational problems;
- facilitate cooperation between comparative educationists of different countries and regions, and foster the establishment of professional associations and groups of comparative educationists;
- support international programs in education and the agencies responsible for them, by focusing the attention of comparative research workers on the major problems encountered in these programs;
- organize research projects for which there is a particular need; and
- improve the exchange of information about research and methodological developments in comparative education.

The most visible WCCES activities are the world congresses held approximately every 3 years (**Table 2**). The Council's officers seek to hold the congresses in different parts of the world in order to promote balance in geographic and other emphases. The WCCES has also sponsored research and publication. The Council does not operate its own journal, although it has provided financial and other support to the journals of member societies.

Names and Approaches

Among the 37 bodies listed in **Table 1**, all had some variant of comparative education in their names, but some linked that field with another. Six societies included the related field of international education in their names, one linked comparative education with educational administration (Egypt), another linked it with history of education (Southern Africa), and another included intercultural education with comparative and international education (Germany). Further, among the 37 WCCES members:

- 27 were societies;
- three were associations (AFDECE, AFEC, and BAICE);
- two were councils (Russia and Kazakhstan);

Table 2 World congresses of comparative education societies

No.	Year	Place	Theme
1.	1970	Ottawa, Canada	Education and the formation of the teaching profession; educational aid to developing countries
2.	1974	Geneva, Switzerland	Efficiencies and inefficiencies in secondary schools
3.	1977	London, United Kingdom	Unity and diversity in education
4.	1980	Tokyo, Japan	Tradition and innovation in education
5.	1984	Paris, France	Dependence and interdependence in education: The role of comparative education
6.	1987	Rio de Janeiro, Brazil	Education, crisis, and change
7.	1989	Montreal, Canada	Development, communication, and language
8.	1992	Prague, Czechoslovakia	Education and democracy
9.	1996	Sydney, Australia	Tradition, modernity, and postmodernity
10.	1998	Cape Town, South Africa	Education, equity, and transformation
11.	2001	Chungbuk, Republic of Korea	New challenges, new paradigms: Moving education into the twenty-first century
12.	2004	Havana, Cuba	Education and social justice
13.	2007	Sarajevo, Bosnia and Herzegovina	Living together: Education and intercultural dialog
14.	2010	Istanbul, Turkey	Bordering, re-bordering, and new possibilities in education and society

- six were specialized comparative education sections in their respective national educational societies (China, Czech Republic, Cuba, Germany, Korea, and Hungary); and
- one (SICESE) operated as a quasi-national body but in constitutional terms was a section of a regional comparative education society.

The distinction between societies and associations was not of great significance. In French, the word association is more commonly used than society (*Société*) for academic bodies, with society being in widespread use in the commercial sector. It was chiefly for that reason that the French translation of WCCES, which was originally *Conseil mondial des sociétés d'éducation comparée* (CMSEC), was changed during the 1980s to *Conseil mondial des associations d'éducation comparée* (CMAEC). In the United Kingdom, the word association had been used in the London Association of Comparative Educationists (LACE), which was formed during the late 1970s but became defunct during the 1980s. When the British Comparative and International Education Society (BCIES) merged with the British Association of Teachers and Researchers in Overseas Education (BATROE) to form the British Association for International and Comparative Education (BAICE), the new body used the word association partly because it took the first two letters from BATROE and partly because the letters formed a good acronym.

In other dimensions, more important principles underlay the naming. For example, the Kazakh body chose to call itself a council rather than a society or an association because its members desired the body to have authority to grant doctoral degrees. In Kazakhstan, such power was given only to academic councils, and not to academic societies or associations. Similar factors related to the Russian Council.

Also significant have been changes over time, which have reflected changes in the nature and mission of some

bodies. Two categories of changes deserving more extended comment are a shift from comparative education to comparative and international education, and a shift from foreign to comparative education.

From Comparative to Comparative and International Education

International education can be defined in various ways, but when paired with comparative education it is commonly taken to imply a more practical and developmental approach to education which may focus on a single country (usually one which is not the original home of the person(s) presenting research findings or other information) and not necessarily based on a strong foundation of comparison. The six WCCES member societies listed in **Table 1** that had international education in their names were:

- the Comparative and International Education Society (CIES);
- the *Sektion International und Interkulturell Vergleichende Erziehungswissenschaft in der Deutschen Gesellschaft für Erziehungswissenschaft* (SHIVEDGE);
- the British Association for International and Comparative Education (BAICE);
- the Comparative and International Education Society of Canada (CIESC);
- the Australian and New Zealand Comparative and International Education Society (ANZCIES); and
- the Nordic Comparative and International Education Society (NOCIES).

In the CIESC, the international component was acknowledged from the outset. The only other member society that had CIES in its name from the beginning was the Nordic society, established in 1992. The historical evolution of

the names of the four older bodies – American, German, British, and Australian – reveals interesting patterns. They are discussed below in chronological order of the year in which they underwent the name change.

The CES existed for 12 years from 1956 without international in its name, and made the change a year after the founding of the CIESC. The name change was chiefly motivated by practical circumstances: the leadership felt that the society would attract funding if it added the word international. The society also had a growing constituency of members working in the applied field of international education, as distinguished from the original dominance of academics concerned with the theoretical and explanatory focus of comparative education.

The British society started as the British Section of CESE in 1966. It gradually diverged from its CESE parent, taking a more international and development orientation, and in 1979 became an independent national society called the British Comparative Education Society (BCES). Like its US counterpart, the British society witnessed a bifurcation of research interests between comparativists more interested in conceptual work and consultants and others more interested in practical international work. To reflect the dual nature of its constituency in a better manner, in 1983, the BCES became the BCIES. Fourteen years later, in response to a changed educational environment, it merged to create BAICE but retained both international and comparative in its title.

The British phenomenon of separation and then consolidation in response to wider interactions was paralleled in Germany. First was the German Section of CESE, formed in 1966 with a concurrent identity as a commission of comparative education in the *Deutschen Gesellschaft für Erziehungswissenschaft* (DGfE – German Society for Education). This created the *Kommission für Vergleichende Erziehungswissenschaft in der Deutschen Gesellschaft für Erziehungswissenschaft* (KVEDGE). In 1978, a Commission for Education with the Third World was formed alongside the comparative education body in the DGfE, echoing the emergence of an international education community in the United States and Great Britain. An intercultural education unit was also formed within the DGfE in 1992, reflecting the growing research interest in migration and its impact on German universities. In 1998, these three bodies were consolidated into one umbrella section, called the Section for International and Intercultural Comparative Education (SIIVE), also known as SIIVEDGE. These transformations were catalyzed by changing research interests and opportunities during the period. The initial fragmentation and splintering of research fields, as embodied in the proliferation of commissions and units, was later reversed with consolidation into one section.

ANZCIES offers another interesting case. It began in 1973 as the Australian Comparative Education Society (ACES). Within a decade, it underwent three name

changes: in 1975 to the Australian and International Comparative Education Society (AICES); then the Australian Comparative and International Education Society (ACIES) in 1976; and, finally in 1983, it became binational (and regional) as the Australian and New Zealand Comparative and International Education Society (ANZCIES).

The debate about names and aims of societies (and of the field) remained vigorous. The evolution in nomenclatures reflected evolution in the field's objects of study, its research interests, and its attractiveness to academics with different specializations. These academics have in turn responded to wider forces in international politics, institutional demands, funding policies, intellectual shifts, migration, etc. However, the patterns have not all been consistent. Thus, for example, the Spanish society in 1984 considered modeling its name after the CIES, but the proposal was not implemented. Similarly, the JCES, whose host country was actively engaged in international development assistance, considered adding international to its name in the mid-1990s. The move was stalled by awareness that renaming the society would cause confusion and overlap with the existing Japan International Education Society. At the same time, the fact that none of the other 31 WCCES constituent societies had the word international in their titles demonstrated that international education was not universally paired with comparative education – and indeed it was not universally recognized as a strong field in its own right in tandem with comparative education.

From Foreign Education to Comparative Education

The German and Chinese cases offer further interesting points for reflection on the difference between foreign education and comparative education. In the late 1950s, institutions in the German Democratic Republic (East Germany) distinguished between foreign countries and West Germany, since West Germany was not considered a foreign country. However, the term comparative education was not used at that time since most work was descriptive or consisted of translated articles of foreign authors. From 1963 to 1974, a department for comparative education existed in East Berlin, and comparative work between East and West flourished. This phase came to a close in 1974, and the term comparative education was replaced by education abroad, since comparisons were once more viewed by the authorities as ideologically risky. Only in 1990 did comparisons become visible again, although initially under the name comparative pedagogy rather than comparative education.

China offers a related story. The professional society in China started as a Foreign Education Research Sub-commission of the Chinese Society of Education (CSE), which was China's largest learned society in the

field of education. In 1983, the subcommission's name was changed from foreign education to comparative education. The name change reflected a change of understanding about the nature of comparative education as a subdiscipline of educational science, and a desire to provide reference for educational reform at home by identifying general laws in the development of education through comparative analyses. This new understanding of the nature and purpose of comparative education took place against the backdrop of China's Open Door policy, which aimed to accelerate national development by importing knowledge from all pertinent locations. A partial offshoot of these new thrusts was the renaming of several institutes of foreign education as institutes of comparative education.

Other Comparative Education Societies

Although the majority of societies in the field of comparative education are members of the WCCES, some are not. For example, during the 1980s and 1990s societies were formed in Albania, Romania, and Venezuela. They did not join the WCCES, either because the necessary connections were never made or because the societies had insufficient organizational capacity and later became dormant or defunct. More positive was the creation in 2005 of the Thailand Comparative and International Education Society (TCIES), with which the WCCES has had some contact.

Alongside these bodies, which were formed as national bodies with a geographic remit, have been two bodies with specialist foci within the field of comparative education:

- The International Society for Physical Education and Sport (ISCPEs) was formed in 1978, has held biennial conferences, and has produced some significant publications.
- The International Society of Comparative Adult Education (ISCAE) began as an informal grouping during the early 1990s. It had a loose membership without a formal constitution or membership fees, but nevertheless organized several significant conferences and publications.

One might expect similar professional societies with specialist foci to be established, although the majority of broadly-focused national bodies would be glad to provide a vehicle for such specialist work. For example, the CIES has created Special Interest Groups (SIGs), some of which focus on specific regions in the world while others focus on such themes as citizenship, early-childhood development, higher education, indigenous knowledge, and language issues.

Conclusion

In the field of comparative education, as in other fields, professional societies and associations have played an

important role in disciplinary institutionalization and scholarly networking. The expansion in the number of societies since the mid-twentieth century has significantly added to the weight of the field. Each society operates within its own sphere, but most societies have been glad to contribute to the global arena through the WCCES. The WCCES acts as a forum through which national, subnational, regional, and language-based societies can interact and collaborate on common agendas.

The fact that some societies have become defunct alongside the ones that have flourished shows that organizational structures may be fragile. Much depends on the leadership and enthusiasm of individuals, as well as on the institutional encouragement provided by universities and other bodies. Nevertheless, the societies have shown themselves to be valuable instruments for development of the field, acting as focal points for exchange of ideas and for publication of research. From a wider perspective, the development and expansion of these societies reflect developments within academia. Some societies also promote links with international organizations such as UNESCO and the World Bank.

Further Reading

- Bray, M. (2003). Tradition, change and the role of the World Council of Comparative Education Societies. *International Review of Education* **49**, 1–13.
- Bray, M. (ed.) (2003). *Comparative Education: Continuing Traditions, New Challenges, and New Paradigms*. Dordrecht: Kluwer.
- Bray, M. (2008). The WCCES and intercultural dialogue: Historical perspectives and continuing challenges. *International Review of Education* **54**, 299–317.
- Carry, A., Frenay, M., Pérez, S., and Gorga, A. (2005). L'éducation comparée dans l'espace francophone: tendances et perspectives. *Revista Española de Educación Comparada* **11**, 135–159.
- Cowen, R. (1990). The national and international impact of comparative education infrastructures. In Halls, W. D. (ed.) *Comparative Education: Contemporary Issues and Trends*, pp 321–353. Paris/London: UNESCO/Jessica Kingsley.
- Epstein, E. (2008). Crucial benchmarks in the professionalization of comparative education. In Wolhuter, C., Popov, N., Manzón, M., and Leutwyler, B. (eds.) *Comparative Education at Universities Worldwide*, pp 9–24. Sophia: Bureau for Educational Services.
- Luzón, A. (2005). La Sociedad Europea de Educación Comparada (CESE) en una época de cambios. *Revista Española de Educación Comparada* **11**, 215–240.
- Manzón, M. and Bray, M. (2006). The Comparative and International Education Society (CIES) and the World Council of Comparative Education Societies (WCCES): Leadership, ambiguities and synergies. *Current Issues in Comparative Education* **8**, 69–83.
- Manzón, M. and Bray, M. (2007). Comparing the comparers: Patterns, themes and interpretations. In Masemann, V., Bray, M., and Manzón, M. (eds.) *Common Interests, Uncommon Goals: Histories of the World Council of Comparative Education Societies and Its Members*, pp 336–363. Hong Kong/Dordrecht: Comparative Education Research Centre, The University of Hong Kong/Springer.
- Masemann, V., Bray, M., and Manzón, M. (eds.) (2007). *Common Interests, Uncommon Goals: Histories of the World Council of Comparative Education Societies and Its Members*.

- Hong Kong/Dordrecht: Comparative Education Research Centre,
The University of Hong Kong/Springer.
- Reischmann, J. and Bron, M. (eds.) (2008). *Comparative Adult
Education 2008: Experiences and Examples*. Frankfurt Am Main:
Peter Lang.
- Saito, Y. (ed.) (2004). *Forty Years of the Japan Comparative Education
Society* (in Japanese). Tokyo: Japan Comparative Education Society.

Relevant Websites

- <http://www.hku.hk/cerc> – Comparative Education Research Centre,
University of Hong Kong.
- <http://www.wcces.net> – World Council of Comparative Education
Societies.

CURRICULUM DEVELOPMENT

Contents

Curriculum and Religion

Gender and Curriculum

Local Teacher-Based Curriculum Development

Curriculum and Religion

C Kapitzke and P Graham, Queensland University of Technology, Kelvin Grove, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

Across continents and cultures and periods of history, religious beliefs and practices have underpinned curriculum in institutions of education. More recently, the so-called culture wars and terrorism have moved religion to the center stage. In both state and independent education sectors, deep-seated assumptions about the nature of reality, spirituality, religion, ethics, and knowledge, converge and clash in the curricular documents of science, history, literacy education, and the like. Prayer in school, intelligent design, and the expression of cultural and religious identity through dress codes, are issues around which otherwise disparate advocacy groups collaborate for policy activism in support of theological traditions.

Schooling and religion in the United States, the United Kingdom, as well as in Australia and Canada, have proven to be a divisive issue for governments and educators alike. Indeed, the classroom has become the locus of broader social conflict over religion's more general role in public life. In essence, government policies and public opinion, in public and faith-based schools, can be seen as an indicator of how each of these nations is handling the complex social issue of religion in the wake of recent world events.

Recent Trends and Issues

In the United Kingdom, for example, the Office for Standards in Education recently published a 5-year study into religious education in state schools (Ofsted, 2007). The report calls for an overhaul of the curriculum and recommends the need to address religion's changing role in a modern world facing an increase in religious fundamentalisms. The recommendations, including a push for a standardized, national religious education curriculum, marked a significant move toward addressing the political and social changes exacerbated by the so-called war on terror. This move toward an inclusive national approach

to broadening religious education in public schools, is in sharp contrast to the current public debate over whether faith-based schools pose a barrier to community cohesion.

By contrast, a recent United States report on the legal and constitutional battles over religion's place in public schools, (Lupu *et al.*, 2007), makes no mention of religions other than Christianity except under the heading multiculturalism. Recent conflicts over prayer in schools, the pledge of allegiance, creationism and evolution, and the rights of students and parents, are given in-depth examination, but there is no discussion on whether, post-September 11 public schools might have a legal obligation to provide students with a broader view of religion's place in the classroom as part of liberal education in a global context. In this respect, it seems that the United Kingdom's approach to religion in public education is grounded in a broader view of the sociopolitical prominence that religion holds outside of the classroom. Indeed, the report by Lupu *et al.* (2007: 1) found that 69% of Americans agreed with the notion that "liberals have gone too far in trying to keep religion out of the schools and the government." Yet, given the fact that in the same survey, almost two-thirds (58%) favored teaching biblical creationism along with evolution in public schools, it seems that in this instance, the word religion merely serves as a euphemism for Christianity.

In Australia, the support for faith-based schools is a well-established tenet of education policy. The deep-seated spirituality of Australia's Indigenous peoples and its centrality to successful learning and well-being is gradually being recognized (cf., Charlesworth *et al.*, 2005). There are perceptions, nonetheless, that some of the nation's Muslim schools are exposing students to extremist views pushed by visiting imams and religious scholars (see, e.g., Skelton, 2005). Skelton levels criticism at the Howard government for taking a big-spending *laissez-faire* attitude to faith-based schools' funding without implementing the necessary checks and balances to monitor curriculum, especially in primary schools.

A small but growing international literature on Muslim education indicates that these questions remain poorly understood for other cultures as well, and yet they have implications for curricular and pedagogical practice in educational contexts with increasingly diverse student populations. Other issues for Muslim education internationally are the role of *madrasabs* within a context of modernization and globalization, the educational needs of Muslim girls, and the interrelation of secular and traditional knowledge (Zia, 2006).

While Western nations are grappling to define religion's role in modern education, in China, Confucianism is making a comeback in both education and the national consciousness. Viewed by many as a means of returning to traditional values, Confucianism reentered the national consciousness last year after a television series based on modernized Confucian lectures proved to be a surprise hit, paving the way for a series-based book that is described as China's biggest bestseller in recent memory (see Yu Dan's *Insights on the Analects*). Chinese authorities are embracing Confucius even though it has been less than a generation since communist Red Guards denounced him as a symbol of a feudal past. Confucianism, both in education and in the home, is now being touted as an antidote to the sense of entitlement endemic in only children born under China's one-child policy.

The literature on the interrelation of education and religion typically begins with the Protestant Reformation and Luther's call for mass public schooling. Indeed, most theoretical and empirical work from the United States commences with the creation of common schooling in the early nineteenth century (cf., Slattery, 2006). By contrast, long before the breakthrough to alphabetic script and print literacy, religious beliefs in oral cultures were fundamental to the educational experiences of the adult and child alike. A turn to the past to rethink long-standing principles embedded within current curricular and pedagogical practice provides a useful context for understanding current issues and contexts. Mainstream scholarship, however, is heading in the other direction. Stark (2005), for example, argues that rather than being a barrier to progress, Christianity laid the foundation for it by enabling the West – specifically Protestant England, Holland, and later north America – to develop and pull ahead of other cultures economically and politically. However, this claim is neither benign nor well warranted.

An alternative argument is that Western beliefs about knowledge and knowing, teaching and learning, derive in large part from Greek culture, and the three dominant religions (Judaism, Islam, and Christianity) that privileged the book as a dominant cultural artifact. The epistemological legacy of this genesis is its implications for institutional education today. Within a context of intense cultural integration, economic globalization, and inequities in knowledge governance, the time has come for the West to engage

with the rest of the worldviews about education's role in society and culture in order to mitigate the epistemological ethnocentrism of Western education, curriculum development, and knowledge policy and practice.

Antiquity: Philosophical Ground

Educational practice, in any given cultural context, is heavily influenced by dominant social institutions; for example, the family in premodern China, theocracy for the Hebrews, and Socratic dialog in Greek pedagogy.

The Greek ideal of dialog through dialectics, logic, and rationality, which privileged the mind over the body, continues today in educational systems of the West. Plato asserted that the pedagogical technique of dialectic is “that strain which is of the intellect only” and that “when a person starts on the discovery of the absolute by the light of reason only, and without any assistance of sense, if he [sic] perseveres by pure intelligence, he attains at last to the idea of good, and finds himself at the end of the intellectual world” (cited in Monroe 1929: 211). The body plays little part here in the work of learning and knowing. Plato's later work emphasized science and mathematics more than poetry and philosophy, but this was a science that remained associated with metaphysics and religion. “Arithmetic and geometry culminate in astronomy and astrology, which forms the basis of social life” (cited in Monroe, 1929: 137).

Plutarch maintained an emphasis on reason as demonstrated in his treatise on learning in *Selections from the Discourse Touching the Training of Children*:

... what we are wont to say of arts and sciences may be said also concerning virtue: That there is a concurrence of three things requisite to the completing thereof in practice—which are nature, reason, and use. Now by reason here I would be understood to mean learning, and by use, exercise. (cited in Monroe, 1929: 307)

The manifestation of reason, believed to be language, continued as the focus of learning in Roman education. In the 12 books of *The Institutes of Oratory*, the Roman educator Quintilian developed his exposition on education, proposing that philosophy was divided into three parts: physics, ethics, and dialectics. As part of the continuing art of dialectics, the orator's business was “to know the exact significations of terms, to clear ambiguities, to disentangle perplexities, to distinguish falsehood from truth, and to establish or refute what he [sic] may desire” (cited in Monroe, 1929: 505). The disputations and the art of logic in the forum required “definitions and deductions, in marking differences and in explaining ambiguities, in distinguishing and dividing, in perplexing and entangling . . .”

The obligation for semantic precision and accuracy here constitutes the correspondence theory of truth, which

defines truth as requiring a correspondence between the thing (e.g., belief, proposition, or statement) and what renders it true (e.g., a fact, event, or state of affairs). It assumes an external, objective, and knowable world. These presuppositions of disembodied reason, ontological certainty, direct representation, the prohibition of contradiction, textual literalness, and thence ideological conviction, provided the philosophical ground upon which monotheistic religion subsequently thrived.

In many ways, the principles of modernity and corporate capitalism that are globally dominant today, were instituted by the Roman emperor Constantine in the third century CE (Steinberg, 2005). It is well known that Constantine's withdrawal of state support from pagan temples, the closing of non-Christian schools, and the exclusion of pagans from teaching following his conversion, Christianized the known world at the time (Lenski, 2006). It is less well known that Christianity was aristocratized by the patrician culture of pagan Rome. Finely grained statistical analyses of the social origins and conversions of the senatorial aristocracy of Rome during the third and fourth centuries have shown how the Christian message was molded and modified to appeal to this powerful aristocratic class (see Salzman, 2002). Their cultural values of honor, office, wealth, literary discernment, and *nobilitas* changed Christianity irrevocably as pagan Rome evolved into papal Rome, replete with the principles and practices of a hierarchized status culture and educational system.

Premodernity: Institutional and Textual Ground

After the collapse of Roman influence in Western Europe, literacy went into decline (Ganshof, 1949). Charlemagne issued his *admonitio generalis*, a main function of which was to establish "schools near the cathedrals or in the monasteries" (Ganshof, 1949: 522). The main purposes of the edict were to enable the careful transcription, correction, and conservation of biblical and liturgical texts, and to ensure minimal literacy standards among the clergy. Charlemagne's admonition was to be the rock upon which our current school, university, legal, and financial systems were built.

By the thirteenth century, especially in France, the system set up by Charlemagne had developed into a flourishing and formalized curriculum designed to produce ecclesiastical, bureaucratic, political, and professional expertise. According to Haskins (1904):

In the intellectual life of the middle ages the University of Paris occupies a place of preëminent importance. "The Italians have the Papacy, the Germans have the Empire, and the French have Learning", ran the old saying; and the chosen abode of Learning was Paris. The University of Paris was generally recognized as "the parent of the

sciences" and the first school of the church, and its supremacy was manifest not only in its position as the centre of scholasticism and the bulwark of orthodoxy, but also in the large number and wide distribution of its students, in its influence upon the establishment and the constitutions of other universities, and in its large share in the political and ecclesiastical movements of the later middle ages. (Haskins, 1904: 1)

Medieval curriculum fused form, content, and function according to its theological principles. Since most people could not read at that time, instructional communication between sacred and secular populations was oral. The *forma sermonis* (or sermon form) was the most common genre of public pedagogy, and was deployed by whichever branch of scholastic expertise for the purpose of directing public consciousness and action (Haskins, 1904).

As public discourse, the sermon had a hortatory function: it was deployed to get people to do things, just as contemporary government policy does. In any sustained period of political organization, genres of instruction are closely guarded. Conflicts over the right to use them characterize the most vigorous periods of political conflict. This is especially the case between and within institutions of power. Little wonder that once the sermon form became clearly defined, recognizable, and functionally effective, it proliferated as a genre of public discourse (Haskins, 1904). The right to use it publicly was contested for centuries.

The explosion in the use of the sermon for populist, political, and ecclesiastical purposes, led to a proliferation of *exempla*: the stories with which the sermons are embellished to exemplify the *thema* and elaborate subdivisions into which the thirteenth-century sermon was split (Haskins, 1904: 3). These stories came from all kinds of sources – fables and folklore, bestiaries, lives of saints, historical manuals, and personal experiences – and comprise the greatest variety of legends and miracles and contemporary anecdotes. In other words, common knowledge (consisting of vulgar history, myth, and everyday life experience) was appropriated in sermon form to change how people behaved.

A further effect of formally ritualizing public discourse was that popular preachers had to work hard to get attention:

In order to hold the attention of the people the preachers found it necessary to be entertaining, as well as simple and direct, and to make abundant use of marvels, anecdotes, and pointed illustrations from everyday life. If his audience showed signs of nodding, the speaker would begin, "There was once a king named Arthur", or shout suddenly, "That fellow asleep will not give away my secrets". (Haskins, 1904: 3)

The recently coined attention economy is merely a new expression to describe what happens when people

develop an indifference to time-worn rituals of public exhortation, when common experience and tradition are appropriated, recontextualized, and redirected at their popular source in order to meet the pedagogic demands of power. Medieval university chancellors were typically distinguished preachers (Haskins, 1904). No surprise, then, that around this time, knowledge begins to get confused with power in a formal sense. The illusion persists to this day in Bacon's widely cited dictum: "Knowledge is power." However, there is no such identity. Knowledge is not power. The ability to convert one's words directly into the actions of others is power, hence, the etymology of dictator and authoritarianism. What counts as knowledge has, since recorded history, been defined by those in power. For the largest part of human history, power has been claimed by a religious elite. Curriculum has been developed accordingly.

Although the Roman church had formed a recognizable hierarchy within a generation of its inception, its earliest writings already evidence a rebellion against the elders concerning the gifts of knowledge supposedly handed down by Christ (Bernier, 1992: 54–55). From its earliest stages, the institutional hierarchy of the church developed powerful forms for the control and dissemination of knowledge, including liturgies, sermons, and various sacramental rituals which endure to the present day (Bernier, 1992: 55–56). The scholastic method of presenting knowledge reached its apotheosis in the *Summa Theologiae* of Thomas Aquinas in the thirteenth century (Makdisi, 1974: 642). However, as Makdisi points out, "one must distinguish between the outer, external technique of presentation," which was organized into parts, questions, and articles, and the inner spirit of scholasticism, "of which the technical schema is merely the vehicle" (Makdisi, 1974: 643–645).

Which is to say: genres of official knowledge are also kinds of media. The more ephemeral texts deployed within them, and the more enduring institutional discourses that are produced, reproduced, and expressed by them, are the mutually constraining boundaries within which official curriculum develops and is disseminated. It is within institutions of power that the media of power are developed, in all their generic forms. What is a far more important and perhaps interesting aspect of the scholastic genres is that "it was not philosophy, not theology, but *law*, which supplied the most basic constitutive element" of a mode of expression that was "both a method of presentation and a way of thought" (Makdisi, 1974: 642). Scholasticism had both discursive and generic aspects; it was various mixtures of institutionally developed meanings, as well as their genres of expression. But the methods of law – formulating prescriptions for behavior and organization – were at the root of its development. Religion, politics, education, and the law became mutually defining institutions. From the institutional complex that

emerged from medieval theology came the rise of scientific method, an ever-uncomfortable Darwinism, the eugenics of Spencer, Galton, and Lamarck, and the disciplines of modernity which continue to multiply, mitigating against complete knowledge at every step.

Industrialism and Beyond: The Secular Theology of Price

As we have shown, formal curriculum has been shaped by religion since its historical emergence. The enlightenment was supposed to have changed all that. But many works, such as Callahan's *Cult of Efficiency* (Callahan, 1962), outline the secular mysticism that has transformed Western organs of education over the last century along the lines of business values – namely, the price system. The price system provides its own suite of moral imperatives, particular approaches to reason and representation, truth and text, as well as liturgy and law, generating a philosophical and institutional matrix for teaching and learning that serves a dominant belief system which reduces everything – beginning with human lives – to money. The current mysticism displays the same sense of theological certainty as was given through God's Holy Word to Muslim, Christian, and Jew, maintaining the schism of a rationally conceived world separate from the experience of teachers and learners alike.

Descartes' philosophical distinction between mind and body embedded a dualism between a knowable world and knowing learners, reifying a foundational space upon which superstitions have continued to underpin educational and scientific practice. Descartes' dualism privileged superstition, rendering embodied experience – literally common sense – unknowing and problematic. All Western curricular frameworks, sacred and secular, embody this principle: that only specific forms of knowledge which support the official mysticism *du jour* are considered valid and eternal. The phenomenology of quotidian experience, morality, and sociality, has no official validity, except as data to be recontextualized within the official framework of the day.

There are at least two reasons for the need to rethink the basis upon which curriculum is theorized and developed, given its almost universally religious foundations. The first is that recent empirical work in a number of disciplinary areas – neuroscience, cognitive psychology, linguistic and cognitive science, and quantum physics – has shown that any ontological separation between knowing and being is no longer tenable. Following this, Lakoff and Johnson (1999) argue for empirical responsibility in philosophy (p. 552). In turn, we argue for empirical and ethical responsibility in curricular theory. This call for ethical responsibility derives from the second reason for the need to rethink curricular precepts: the increasing global dominance of Western technocratic epistemologies

and the concomitant misalignment of knowledge cultures within a global knowledge economy dominated by the superstitions of the price system. It is a truism that occidental assumptions about people and their metaphysical and material relations to reality differ from oriental ones (Shen and Van Doan, 1995). The problematic place of indigenous beliefs and knowledges within new intellectual property regimes similarly indicates a need for genuine and deep engagement on the part of curriculum specialists with different modes of knowing, different belief systems, and different value systems (Reagan, 2005).

The ethical imperative for curriculum theory in a globalized knowledge economy also signifies a historic opportunity for the field: an opportunity to define the meaning of education as its own end and to define what a system thus oriented would look like; a chance to move from curriculum *occludere* to curriculum *vita*, the potential for education dominated by the needs of the people rather than by the dictates of price, prince, prophet, or Pope; to a system dedicated to the valorization of human potentials, understanding, and flourishing, rather than the principles of some occult agency as interpreted by a privileged elite. Such ideals are a distinct possibility given the emergence of a global knowledge economy, a system that seeks to embrace and commodify every aspect of a human being. But to do so, it must first render the full scope of those potentials visible. As a consequence, and however accidentally, glimpses of the fullness and diversity of human knowledge are coming into clearer focus every day. In this accidental view lies the potential to truly understand and learn from each other, and it is only in full view of humanity, we argue, that the ethics of truly enlightened curriculum development can be enacted.

Bibliography

- Bernier, P. (1992). *Ministry in the Church: A Historical and Pastoral Approach*. Mystic, CT: Twenty-Third Publications.
- Callahan, R. E. (1962). *Education and the Cult of Efficiency*. Chicago, IL: University of Chicago Press.
- Charlesworth, M., Dussart, F., and Morphy, H. (eds.) (2005). *Aboriginal Religions in Australia: An Anthology of Recent Writings*. Burlington, VT: Ashgate.
- Haskins, C. H. (1904). The University of Paris in the sermons of the thirteenth century. *American Historical Review* 10, 1–27.
- Lakoff, G. and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.
- Lenski, N. (ed.) (2006). *Cambridge Companion to the Age of Constantine*. New York: Cambridge University Press.
- Lupu, I. C., Elwood, F., Davis, E., et al. (2007). *Religion in the Public Schools*. Washington, DC: The Pew Forum on Religion and Public Life. <http://pewforum.org/docs/?DocID=206> (accessed October 2009).
- Makdissi, G. (1974). The scholastic method in mediaeval education: An inquiry into its origins, law, and theology. *Speculum* 49, 640–661.
- Monroe, P. (1929). *Source Book of the History of Education for the Greek and Roman Period*. New York: Macmillan.
- Ofsted (Office for Standards in Education) (2007). *Making Sense of Religion: A Report on Religious Education in Schools and the Impact of Locally Agreed Syllabuses*. London.
- Reagan, T. G. (2005). *Non-Western Educational Traditions: Indigenous Approaches to Educational Thought and Practice*, 3rd edn. Mahwah, NJ: Erlbaum.
- Salzman, M. R. (2002). *The Making of a Christian Aristocracy: Social and Religious Change in the Western Roman Empire*. Cambridge, MA: Harvard University Press.
- Shen, V. and Van Doan, T. (eds.) (1995). *Philosophy of Science and Education: Chinese and European Views*. Washington, DC: Council for Research in Values and Philosophy.
- Skelton, R. (2005). Muslims sound alarm over schools. *The Sunday Age*. <http://www.theage.com.au/news/war-on-terror/muslims-sound-alarm-over-schools/2005/07/30/1122144058379.html> (accessed October 2009).
- Slattery, P. (2006). *Curriculum Development in the Postmodern Era*, 2nd edn. New York: Routledge.
- Stark, R. (2006). *The Victory of Reason: How Christianity Led to Freedom, Capitalism, and Western Success*. New York: Random House Trade Paperbacks.
- Steinberg, M. (2005). *The Fiction of a Thinkable World: Body, Meaning, and the Culture of Capitalism*. New York: Monthly Review Press.
- Zia, R. (ed.) (2006). *Globalization, Modernization, and Education in Muslim Countries*. New York: Nova Science.

Further Reading

- Bohm, D. (1993). *The Undivided Universe: An Ontological Interpretation of Quantum Theory*. London: Routledge.
- Brown, R. A. (1972). *The Origins of Modern Europe: The Medieval Heritage of Western Europe*. Suffolk: Boydell.
- Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G.P. Putnam.
- Fraser, J. W. (1999). *Between Church and State: Religion and Public Education in a Multicultural America*. New York: St. Martin's Press.
- Greenawalt, K. (2005). *Does God Belong in Public Schools?* Princeton, NJ: Princeton University Press.
- Haynes, C. (1994). *Finding Common Ground: The First Amendment Guide to Religion and Public Education*. Nashville, TN: Freedom Forum First Amendment Center.
- Martínez López-Muniz, J. L., de Groof, J., and Lauwers, G. (eds.) (2006). *Religious Education in Public Schools: Study of Comparative Law*. Dordrecht: Springer. Religion in the public schools – Law and legislation – Europe.
- Maturana, H. R. and Varela, F. J. (1987). *The Tree of Knowledge: The Biological Roots of Human Understanding*, 1st edn. Boston, MA: New Science Library.
- Nord, W. (1995). *Religion and American Education*. Chapel Hill, NC: University of North Carolina Press.
- Scott, E. C. (2004). *Evolution vs. Creationism: An Introduction*. Westport, CN: Greenwood Press.
- Sears, J. T. and Carper, J. C. (eds.) (1998). *Curriculum, Religion, and Public Education: Conversations for an Enlarging Public Square*. New York: Teachers College Press.

Gender and Curriculum

J Marsh, University of Sheffield, Sheffield, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Hegemonic masculinity – An idealized form of masculinity that reflects normative ideologies regarding male power and dominance.

Heteronormativity – The reinforcement of ideologies, norms, and practices relating to the belief that heterosexual relationships between men and women are normal and natural and that homosexual and bisexual relationships are, therefore, aberrant.

In September 2000, the United Nations, World Bank, and the Organisation for Economic Co-operation and Development (OECD) developed a set of eight Millennium Development Goals (MDGs), which all 191 United Nation Member States have pledged to meet by 2015. The third of these MDGs aims to promote gender equity and empower women; the targets include improving the ratio of girls to boys in primary, secondary, and tertiary education and extending the ratio of literate females to males. Substantive work needs to be undertaken to meet these targets, and experience from those countries that have enjoyed gender equity in relation to access for education for some years (primarily in the developed world) suggests that even when equal numbers of girls and boys are attending formal education, there are still significant disparities in relation to issues of gender and curriculum that have implications for employment and standards of living.

Curriculum, which in etymological terms means a course to be run, is a concept that refers to both the taught elements of a program and the hidden values and messages that underpin that curriculum. In this article, the focus is on the taught elements of the program in schools and how they are predicated on implicit assumptions and hidden values regarding gender. The selection of knowledge that takes place in any construction of a curriculum is shaped by the wider social, cultural, political, and historical context. A particular curriculum does not simply mirror society's values, but also transmits those values and ideologies that are located within the domain of the dominant classes. In relation to gender, there are two key areas that need to be considered. The first is the relationship between gender, power, and knowledge in the construction of the curriculum and how this shapes the formation of school subjects. The second is that of gender and access to the curriculum and the way in which choice operates within curriculum delivery

and take-up. Both of these areas are addressed, in addition to more recent work on gender, sexualities, and curriculum.

This article is written from a theoretical standpoint that recognizes the relational nature of gendered identities. That is, it is acknowledged that gender is socially constructed and the categories of masculine and feminine are developed in relation to one another. As Francis (2000) suggests, "there is one (notional) masculinity and one (notional) femininity constructed as oppositional to one another, and consequently shifting, but flexible, and incorporating contradictions" (p. 15). The shifts, contradictions, and flexibility, therefore, can account for the diverse range of hegemonic and nonhegemonic masculinities and femininities that are constructed and performed by individuals. It is also important to note that gender does not work in isolation from other vectors of identity. Race, socioeconomic status, physical ability, sexuality, and other aspects of identity interact in complex ways and, therefore, any analysis of the curriculum that focuses on gender should acknowledge the interrelatedness of these elements.

Gender and the Construction of the Curriculum

The distinctions and disparities between the schooling of males and females have been apparent since historical records of education began to be kept. Rousseau's (1911) account of education, *Emile*, in which Sophie's education is marginal to that of her fiancé's, characterized, aptly, the fact that the schooling of females has been relegated as second in importance to that of males for hundreds of years. However, as this article is concerned with explicit curricula as defined by educational institutions, the analysis will be confined to the twentieth and twenty-first centuries. It is acknowledged that formal schooling took place in various forms prior to this; however, it was only at the advent of the twentieth century that curriculum studies, as a subset of the field of education, began to take shape and, thus, a study of gender and the curriculum emerged.

Curriculum theorists in the early twentieth century were not overly concerned with issues of gender and it was only with the advent of second-wave feminism in the 1970s and 1980s that the way in which the curriculum was inherently biased toward male interests began to be systematically documented. Texts such as those of de Beauvoir (1972) and Oakley (1972) were central to feminist moves to challenge the patriarchal nature of school

curricula. The development of formal schooling since the Enlightenment had been built upon the assumption that the highest forms of knowledge were decontextualized, rational, and not based on emotional or personal experience. Thus, subjects such as science and mathematics that were perceived to be based on reason and objectivity became equated with hegemonic masculinity. A Cartesian dualism, which purported a disjuncture between mind and body and disparity between technical rationality and the uncertainty of emotions, began to take hold in relation to school subjects. A hierarchy of subjects emerged, with science and mathematics, equated with masculine virtues and values, placed at the top of the pinnacle. From the earliest days of universal schooling in developed countries, there existed a sharp dichotomy between perceived interests of men and women and their psychological orientations to learning, which shaped the school curriculum in profound ways.

Much of the work on gender and curriculum in the second half of the twentieth century was concerned with identifying and analyzing these conflicting and competing epistemologies. Harding (1987) suggested that in the initial stages of critiquing andocentric forms of knowledge, feminist empiricists worked on the belief that experience is the basis for knowledge and that sex bias can be removed from research tools and processes in order to produce nonsexist knowledge. In contrast, in work that originated from a feminist standpoint epistemology, the issue was not about whether biased research methods were responsible for sexist assumptions and knowledge, but about acknowledging the way in which one's everyday experience informed one's viewpoint. Thus, the everyday experience of women meant that they had an understanding of the way in which sexism operated, and these experiences shaped the construction of knowledge. In this model, people who face oppression of any kind, including oppression due to gender, race, sexuality, and/or disability, are able to identify how such oppression works because of their experiences. Standpoint theory was thus also adopted in relation to other ways of knowing such as black feminist standpoints. Feminist standpoint epistemologies therefore countered andocentric knowledge by celebrating women's ways of knowing (Belenky *et al.*, 1986) in terms of curriculum development. Following this work, which indicated that there were differently gendered ways of viewing the world, research developed that looked, in greater detail, at the relationship between schooling and society and promoted critical approaches to concerns of gender equity. A large number of initiatives took place during this period in an attempt to change the construction of school subjects in order to produce antibias curricula.

In the later decades of the twentieth century, feminist poststructuralists challenged the rationalist Enlightenment view, which suggested that there was one solid version of reality that could be examined and understood, and

purported that there are multiple realities that exist alongside each other, with the implication that knowledge is therefore perspectival. In relation to curriculum studies, feminist poststructuralism led to an emphasis on the multiple positionings that learners might have and explored the interrelationship between gender and other aspects of identity in the experience of schooling. The relationship between power and knowledge was critiqued in the light of work on multiple gendered identities (Paechter, 1998, 2000). Postcolonial feminists challenged the othering of black and ethnic minority girls and women in the curriculum and pointed to the colonialist nature of much of the knowledge valorized in schools (Asher, 2005). This work, in turn, pointed to the necessity for critical self-reflection by both teachers and learners if curricula were to offer space for the deconstruction of ideologies and actively engage learners in the Freirean process of conscientization in which the processes of political, economic, and social oppression are laid bare.

Much of the scholarly work on curriculum and gender in the latter half of the twentieth century centered on the nature of the subjects such as science, mathematics, and technology that were perceived to be masculinist in nature. The lack of attention to the work of female scientists and mathematicians was pointed out, in addition to the ways in which these subjects focused on male concerns and interests. For example, practical examples and illustrations in science often started from typical male experiences; mathematical calculations were usually made in relation to objects, not people. It was thus argued that the relationship between power, knowledge, and gender was one that privileged masculine ways of knowing so that the various curricula reflected individualism, competition, depersonalization, and atomization of knowledge.

Attention was paid to the resources used in science, mathematics, and technology curricula in order to trace the delegitimization of female interests and practices. Textbooks are key educational artifacts in the selection, organization, and transmission of official knowledge; mathematics and science textbooks were identified in this period as being inherently sexist in their portrayal, or the lack of it, of females. Texts relating to other subjects were also identified as regurgitating stereotypical representations of gender. For example, early reading primers were found to be replete with images of women as homemakers and men as breadwinners, with the former portrayed as active in literacy practices and the latter less frequently engaged with written texts (Luke *et al.*, 2003). History textbooks marginalized and misrepresented the history of women and fostered colonial ideologies. Books used in the geography curriculum were both racist in their portrayal of different cultures and sexist in the ways in which they marginalized women, with the emphasis on physical rather than social geography exacerbating this tendency. The language of textbooks was also

analyzed in order to determine the ways in which sexism was inscribed in choices concerning vocabulary and syntax. The work conducted over this period on textbooks, artifacts, and other resources was central to establishing the material culture of gender stereotyping and identifying the marginalization of girls and women in the fabric of schooling.

The analysis of curricula during this period also led to an understanding of the way in which the hierarchies between subjects were both gendered and classed. For example, subjects with the highest status, such as science and mathematics, developed more coherent curricula that were adopted across schools, including boys', girls', and coeducational schools. Moves to introduce what were perceived to be feminine forms of these subjects, such as domestic science, were resisted by girls' schools that served the middle classes because it was recognized that this would disadvantage female school leavers, given that the masculine forms of the subjects would always be viewed as enjoying higher status. However, lower-status subjects were permitted to adapt their curriculum to cater to the perceived differences in the needs and interests of boys and girls (Paechter, 1998). For example, physical education developed distinct sports that were targeted at the different genders, such as gymnasium for girls and competitive outdoor sports for boys. Working-class children were channeled into more vocational subjects such as beauty therapy and woodwork that offered highly gendered curricula. During the later decades of the twentieth century, therefore, awareness was raised about the ways in which the curriculum both reinforced sexist stereotypes and (re)produced gendered subjects.

This was an intensive period for such analysis, influenced in some ways by the formation of gender studies as an academic discipline within higher education, a move which sought to further dissolve the boundaries between disciplines to challenge male-dominated traditions. The work during these decades was, by and large, a modernist project that sought to identify the extent of the problem and redress the balance in schooling in order to develop equitable curriculum practices. Toward the final years of the twentieth century, however, an alternative, transformative agenda was set by feminist poststructuralists and others who stressed the need to challenge the very foundations upon which curriculum studies were built. The advent of progressive reconceptualist curriculum theories led to the questioning of normative assumptions regarding the construction of knowledge and began to shift the focus toward the development of insider epistemologies and narrative inquiry as foundation for the curricula.

Poststructuralist feminist moves to challenge the gendered constructions of the curriculum were met, during the later years of the twentieth century, with a plethora of texts that created a moral panic regarding boys and underachievement, texts that Mills (2003) has labeled backlash

blockbusters. While the underachievement of boys in languages and literacy was not a new phenomenon, as it has been traced as far back as standardized assessment will allow, it became a prevalent discourse in the educational sphere and attention turned to the perceived feminine nature of the arts and humanities curricula. It was argued that language and literature curricula primarily focus upon narrative genres, and the emphasis is on producing personal responses to literature, neither of which is in line with the interests of boys (Millard, 1997). Nevertheless, both the evidence used to identify the underachievement of boys and to define language and literacy curricula as feminine in nature have been contested, with emphasis placed on the fact that socioeconomic status is a more significant factor than gender in underachievement. However, these exhortations to engage critically with the boys and the underachievement debate have been overlooked in the move to remasculinize education in neoliberal times, which has seen an increasing emphasis on masculinized discourses of technical rationality and a marginalization of gender equity issues (Arnot and Mac An Ghaill, 2006).

Gendered Access to the Curriculum

In addition to the content of the curriculum being highly gendered in nature, the way in which the curriculum is organized in formal education can affect how females and males take up or reject subjects. This is the case from the early years of schooling. In many nurseries and elementary classes, girls and boys have taken up different areas of the curriculum with more or less enthusiasm, with boys frequently orientating themselves to particular activities such as block play or outdoor play and girls electing to undertake activities such as sociodramatic role-play or art. Children in the early years of schooling have been found to have already-constructed normative discourses about gendered identities (Davies, 1989), which need to be challenged by educators in the development of curricula that offer children spaces in which to explore their multifaceted identities.

In countries that have introduced national curricula with specified subjects that are mandatory up to a certain level, there is greater gender equity in relation to subject participation. As soon as an element of choice is introduced, females and males begin to choose different options. Girls are less likely to choose science and mathematics when choices are available and, when they do opt for these subjects and succeed in them, their achievement is often viewed as the result of hard work rather than natural aptitude. Boys have traditionally been less likely to opt for subjects related to literature and languages. The way in which schools offer subject options can impact gendered curriculum choices. For example, if the

timetabling of options juxtaposes traditionally male-dominated subjects with traditionally female-dominated ones, there is likely to be a differently gendered take-up of subjects along stereotypical lines. A key difference in the study of subjects that lie outside of normative choices appears to occur in single-sex schools. While single-sex schooling appears to make little difference in terms of attainment, pupils who attend single-sex schools are more likely than their co-ed counterparts to choose subjects not traditionally associated with their gender.

It is not just the range of subjects offered that can shape differently gendered patterns of engagement, but also the ways in which educational institutions privilege masculinized subjects in terms of time and space allocation which affect the level of importance attributed to them. Mathematics, science, and technology, for example, all traditionally conceived as masculine subjects as suggested previously, are allocated both extended time within the curriculum and additional space in schools in order to accommodate laboratories and workshops. In contrast, subjects such as modern foreign languages and dance that have conventionally been perceived as feminine are marginalized both in terms of curriculum time and physical provision (cf. Paechter, 2000).

The way in which the construction of physical education curricula determines gendered participation in that subject has been widely documented. For example, gender as a physical embodiment becomes paramount in a subject in which males and females have, traditionally, been guided toward different sports depending upon assumptions about strength and stamina. Traditionally, more boys than girls have chosen to undertake outdoor sports, with the higher levels of competition, discipline, and regimentation apparent in this arena of male-dominated activity. Physical education has thus become a site for the performance of the gendered body and display of hegemonic masculinities and femininities.

It is, perhaps, in the area of vocational education and training (VET) that we see the most distinct forms of cultural, gendered reproduction occurring. The way in which vocational education is structured normally leads to distinct patterns in curriculum choices of girls and working-class boys, with patterns in many countries also reflecting ethnically defined choices (Rodgers and Boyer, 2006). Patterns in gender segregation in labor markets remain largely unchanged over the last 30 years and this is reflected in the choices males and females make in VET. The curricula of VET perpetuate gender dichotomies in that courses focused on the caring professions (nursing and childcare) emphasize affect and the importance of self-sacrifice, while male-dominated training (construction and manufacturing) focuses on rationality, objectivity, and instrumentalism (Colley *et al.*, 2003).

Intervention projects that have sought to address some of the barriers to curriculum participation have achieved varying degrees of success. Single-sex groupings can lead

to enhanced confidence in girls and enable them to engage in nontraditional curriculum activities. Mixed-sex activities can provide opportunities for male pupils to listen to and learn from the experiences of their female peers. Interventions are more likely to be successful when they are integrated into whole-school approaches that mainstream a commitment to gender equity. Success is also more attainable when curriculum development is linked to pedagogical changes so that more inclusive and democratic practices are promoted. The political shift in neoliberal times to a focus on boys and underachievement has seen a concomitant change in approaches to intervention projects. While the developments of the 1970s and 1980s that sought to entice girls into studying mathematics and science were conducted in an environment that focused on gender equity and paid attention to the needs of boys, more recent projects designed to orientate boys to languages and literacy have often marginalized girls and privileged masculine discourses, reinforcing stereotyped assumptions and essentializing gendered identities in the process.

Masculinities, Femininities, Sexualities, and Curriculum

In recent years, issues of multiple identities have risen to the fore in considerations of gender and curriculum. This work has sought to engage with the variations within gender categories, stressing the fluidity of category boundaries and the multiple positioning of individuals along the male–female continuum. This has necessitated feminist scholars engaging with conceptualizations of masculinities in an attempt to analyze the way in which gender as a category impacts educational experience. As the examination outcomes of girls have improved, there has been increasing attention paid to the way in which constructions of hegemonic masculinities work in opposition to notions of achievement and attainment in schools. The notion of laddishness (Jackson, 2006), the deliberate construction of resistant masculinities, has been advanced as a part-explanation for this male conflict with normative educational discourses. The work on masculinities has been valuable in further strengthening the understanding about the complexities of identity work and the relationship between structure and agency in the formation of gendered identities. Alongside this work, consideration has been given to multiple femininities, with acknowledgement that the relationship of girls with traditional gendered identities is complex and can be both compliant and transgressive. This has challenged stereotypes surrounding the compliance of girls in classrooms and led to a range of work that has illuminated how some girls work actively to destabilize academic identities, leading to alienation, underachievement, and even exclusion.

Drawing primarily from Foucauldian concepts of the relationship between power and knowledge, gender studies in recent years have sought to identify ways in which the embodiment and performance of gender can be traced in classroom discourses. Butler's (1990) insistence on gender as a citational practice indicates that the ritualized performances of what it means to be inscribed as male or female can be identified in the daily transactions of classroom life. This relates to the curriculum in a number of ways. If students are not challenged to question these norms, through the development of curricula that enable them to deconstruct normative gendered practices, then such practices produce idealized notions of gender that serve to further reinforce partial knowledge. In addition, a curriculum that fails to acknowledge the nature of multiple, fluid, and performed gendered identities is likely to silence those pupils who sit on the margins of classroom life and lead to pedagogical practices that promote alienation and the othering of difference. Recent work in gender studies and curriculum has, therefore, further problematized the relationship between curriculum and pedagogy and illuminated the way in which they work together to create the illusion of stabilized gendered identities.

There have been other developments in curriculum theory that relate to a consideration of issues of gender within the curriculum. Developing out of the work of a number of progressive reconceptualist curriculum theorists, the queer studies movement in education has challenged the heteronormative construction of the curriculum in which gender and sexuality are correlated in unproblematic ways (Pinar, 1998). This work has sought to disrupt the notion of heterosexuality as a secure category and indicated that the curriculum needs to incorporate work on constructions and performance of sexual identities. Attention needs to be paid to the way in which heterosexual practices and values become naturalized within curricula. Particular subjects have received more detailed analysis in this work than others. For example, physical education has been identified as an intensive site for heteronormative work, given the relationship between sexuality, gender, and the body. Compulsory heterosexuality is reinforced in a subject in which physical prowess is related to hegemonic masculinity. The study of literature offers more potential for the exploration of queer identities, although this is achieved in spite of, not because of, the official curriculum as gay, lesbian, bisexual, and transgendered readers reinscribe themselves into texts (Vicars, 2007). What is needed is the inclusion of texts that speak about the lives of marginalized pupils in the curriculum. In addition, the role of sex education in challenging prejudices and enabling self-exploration in relation to sexuality has been considered, and Britzman (1998) has argued that the subject should not simply reflect on issues relating to sex, but should also consider how knowledge about sexualities is constructed. There is still work to be done in relation to other curriculum subject areas, but the emergent

work on queer theory and curriculum has highlighted the need to trace, in close detail, the normative discourses regarding sexuality that are embedded within the development of curricula.

Conclusion

The place of gender within curriculum studies has developed considerably since the early years of the twentieth century. However, it continues to be an aspect of curriculum theory that takes place on the margins and needs to be central to the field as it moves forward. There are a number of areas that need attention in the years ahead; among them are issues relating to gender and curriculum in the developing world. As countries within the United Nations move toward the MDGs, it will be important to monitor how questions relating to gender and curriculum are addressed. In addition, while the progressive reconceptualist curriculum movement has challenged heteronormative constructions of the curriculum in recent years, there is still work to be done in further exploring the intersections of gender, sexuality, race, and class as they are inscribed in and by the curriculum. Further, there are aspects of gendered identities that are still to be examined in any depth in relation to curriculum studies. For example, the exploration of gender within the field of disability studies has not extensively explored the implications of that work for curriculum theory, neither has there been sufficient attention paid to the experiences of other marginalized groups such as women refugees, asylum seekers, and travelers. Finally, as the field of gender studies within social sciences develops, there will be further iterations of theories relating to the construction, embodiment, and performance of gendered identities that will have relevance for schooling. The future of gender and curriculum studies is one that will continue to be significant to the field as a whole, for it is in the reconceptualization of the curriculum in the light of multiple gendered and sexualized identities that questions regarding relevance, equity, and the nature of schooled knowledge itself become most salient.

Bibliography

- Arnot, M. and Mac An Ghaill, M. (2006). (Re)contextualising gender studies in education: Schooling in late modernity. In Arnot, M. and Mac An Ghaill, M. (eds.) *The RoutledgeFalmer Reader in Gender and Education*, pp 1–14. London: Routledge.
- Asher, N. (2005). At the interstices: Engaging postcolonial and feminist perspectives for a multicultural education pedagogy in the South. *Teachers College Record* **107**, 1079–1106.
- de Beauvoir, S. (1972). In Parshley, H. M. (ed.) *The Second Sex*. Harmondsworth: Penguin.

- Belenky, M. F., Clinchy, N., Goldberger, L., and Tarule, J. M. (1986). *Women's Ways of Knowing: The Development of Self, Voice, and Mind*. New York: Basic Books.
- Britzman, D. P. (1998). *Lost Subjects, Contested Objects: Toward a Psychoanalytic Inquiry of Learning*. Albany, NY: State University of New York Press.
- Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.
- Colley, H., James, D., Tedder, M., and Diment, K. (2003). Learning as becoming in vocational education and training: Class, gender and the role of vocational habitus. *Journal of Vocational Education and Training* 55, 471–498.
- Davies, B. (1989). *Frogs and Snails and Feminist Tales*. Sydney: Allen and Unwin.
- Francis, B. (2000). *Boys, Girls and Achievement: Addressing the Classroom Issues*. London: RoutledgeFalmer.
- Harding, S. (1987). *Feminism and Methodology*. Milton Keynes: Open University Press.
- Jackson, C. (2006). 'Lads' and 'Ladettes' in School: Gender and a Fear of Failure. Maidenhead: Open University Press.
- Luke, A., Carrington, V., and Kapitze, C. (2003). Textbooks and early childhood literacy. In Hall, N., Larson, J., and Marsh, J. (eds.) *Handbook of Early Childhood Literacy*, pp 249–257. London: Sage.
- Millard, E. (1997). *Differently Literate: Boys, Girls and the Schooling of Literacy*. London: Falmer.
- Mills, M. (2003). Shaping the boys' agenda: The backlash blockbusters. *International Journal of Inclusive Education* 7, 57–73.
- Oakley, A. (1972). *Sex, Gender and Society*. London: Temple Smith.
- Paechter, C. (1998). *Educating the Other: Gender, Power and Schooling*. London: Falmer.
- Paechter, C. (2000). *Changing School Subjects: Power, Gender and Curriculum*. Buckingham, PA: Open University Press.
- Pinar, W. F. (1998). *Queer Theory in Education*. Mahwah, NJ: Erlbaum.
- Rodgers, Y. V. and Boyer, T. (2006). Gender and racial differences in vocational education: An international perspective. *International Journal of Manpower* 27, 308–320.
- Rousseau, J. J. (1911). In Foxley, B. (ed.) *Emile*. London: Dent.
- Vicars, M. (2007). "I sort of read to be middle class and then I sort of read to be gay": The reading practices and identities of gay men. In Beame, E. and Marsh, J. (eds.) *Literacy and Social Justice: Closing the Gap*, pp 57–75. Stoke-on-Trent: Trentham.

Further Reading

- Arnot, M. and Mac An Ghaill, M. (eds.) (2006) *The RoutledgeFalmer Reader in Gender and Education*. London: Routledge.
- Collins, P. H. (1990). *Black Feminist Thought: Knowledge, Consciousness and the Politics of Empowerment*. New York: Routledge/Chapman and Hall.
- Epstein, D. and Johnson, R. (1998). *Schooling Sexualities*. Buckingham: Open University Press.
- Francis, B. (2000). *Boys, Girls and Achievement: Addressing the Classroom Issues*. London: RoutledgeFalmer.
- Noddings, N. (1999). Gender and the curriculum. In Jackson, P. W. (ed.) *Handbook of Research on Curriculum*, pp 659–684. New York: Macmillan.
- Paechter, C. (1998). *Educating the Other: Gender, Power and Schooling*. London: Falmer.
- Paechter, C. (2000). *Changing School Subjects: Power, Gender and Curriculum*. Buckingham, PA: Open University Press.
- Pinar, W. F. (1998). *Queer Theory in Education*. Mahwah, NJ: Erlbaum.
- Renold, E. (2005). *Girls, Boys and Junior Sexualities: Exploring Children's Gender and Sexual Relations in the Primary School*. Abingdon: RoutledgeFalmer.
- Skelton, C. (2001). *Schooling the Boys: Masculinities and Primary Education*. Buckingham: Open University Press.
- Walkerdine, V. (1988). *The Mastery of Reason: Cognitive Development and the Production of Rationality*. London: Routledge.

Relevant Websites

- <http://www.lancs.ac.uk> – ESRC Seminar Series: Girls and Education 3–16: Continuing Concerns, New Agendas
- <http://www.genderandeducation.com> – Gender and Education Association
- <http://www.oxfam.org.uk> – Oxfam (Gender and Education Series): Oxfam in Action – Issues we work on
- <http://www.ungei.org> – The United Nations Girls' Education Initiative (UNGI)

Local Teacher-Based Curriculum Development

M D Osborne, University of Illinois at Urbana–Champaign, Champaign, USA

H B Barker, Regis University, Denver, CO, USA

© 2010 Elsevier Ltd. All rights reserved.

The practice of organized, sanctioned, local teacher-based curriculum development – at least in the USA – is relatively new in public education. While the present view of teaching, especially at the elementary level, is that it is an interpretive act in which the teacher creates curriculum with each enactment; this is not the same as saying that teachers purposely and deliberately develop curricula – lay out the scope and sequence of the things that will go on during the school year in their schools and classrooms. In addition and most likely, in this age of national standards and outcomes followed by high-stakes achievement tests, teachers still are not asked to do this. They are, however, more and more frequently expected to develop the local, concrete, and connected expressions of national or state expectations. This is important, for it is what makes the abstract list of national standards or outcomes relevant and connected. It is this which makes the mandated curriculum teachable and actualizes educational reform – described at the policy level – in the classroom.

Philosophically, local teacher-based curriculum development is a practical restatement of a Deweyian educational principle – that to teach the abstract goals and values at the core of curriculum, it needs to be translated to the immediate and local, an activity that only the classroom teacher can carry out. This is based on the premise that learning is local and particular and learners construct knowledge in a much more holistic and experiential fashion than can be assumed by the preconstructed texts, worksheets, and basal readers typical of purchased curricula. To foster meaningful learning, teachers connect the differing experiences of individual learners and the abstract curriculum goals by reconstructing these goals into local curriculum. Understanding how their students think as well as what they know, teachers use an array of techniques as well as their own experientially shaped professional wisdom to build on and with the knowledge, ways of knowing and worldviews of their students. This more complex approach to teaching requires that teachers combine deep knowledge of subject matter and a wide repertoire of teaching strategies with intimate knowledge of students' growth, experience, and development. The move to acknowledge this and organize this process began in the USA with the postmortems conducted around the failures of the teacher-proofed curriculum of the 1960s and 1970s. In Great Britain, it crystallized in the work of Lawrence Stenhouse, in 1980.

History and the Professionalization of Teaching

The 1960s and 1970s were years in which the greatest efforts in the USA, and abroad, were made to reform education through curriculum materials. These curricular materials were produced by national organizations with unprecedented levels of funding (Darling-Hammond, 1996). The curriculum committees were comprised of experts from the disciplinary fields (as well as the occasional teacher) (Bruner, 1960/1977). Holding to the curriculum model of researchers such as Tyler (1949), educational experiences were selected and developed to attain objectives agreed upon by the expert curriculum committees. The effects of teachers translating the curriculum – with the assumption that this meant diminishing the curriculum – was purposely sidestepped through policy statements warning administrators to watch for and correct this tendency, extensive teacher manuals on implementation, workshops aimed at inducting teachers into correct ways of teaching the materials, and even scripted lessons. The results – whether disciplinary or equity driven – were not very positive, especially measured relative to the goals of the curriculum developers (e.g., Stake and Easley, 1978). In retrospect, this failure was attributed to the lack of acknowledgment of the role of the teacher both in forming the curriculum and in actualizing it. This set the stage for the present environment of organized, local teacher-based curriculum development.

The situation of education in the 1980s in the USA was shaped by two important sets of policy reports – the publication of *A Nation at Risk* in 1983 and the publications by the Holmes Group, the Carnegie Task Force of Teaching as a Profession, as well as other policy groups and teacher unions arguing for an acknowledgment of teaching as a unique professional activity with consequent public professionalization of teaching as a career. *A Nation at Risk*, in the USA, and similar studies and documents, in other countries, argued for an increase in academic standards and rigor in public schools. This initiated the development and implementation both of national and state disciplinary standards. Unlike *A Nation at Risk* – that called for greater external determinates of school curriculum – the Holmes Group and other reports on professionalization of teaching argued that the improvement of schools would come through increasing the status and power of classroom

teachers and by decentralizing school decision making. From then on until the recent policies such as *No Child Left Behind Act* (2001) came into play, the call was for the empowerment of teachers to participate in a central way in the determination of school goals and policies, and to exercise their professional judgment with regard to the content of the curriculum and means of instruction.

To reconstruct teaching as a formal profession analogous to other people-helping professions, such as social work or nursing (Stronach *et al.*, 2002), meant codifying the responsibilities of the teacher and constructing a framework for enabling and rewarding these professional activities. Most efforts in this regard focused upon pedagogy, and the resulting reward structures and professional certifications such as National Board Certification in the USA have endured. Similarly, the recent manifestations of local teacher-based curriculum development are linked to arguments around the professionalization of teachers. In other words, defining curriculum making as a teacher province has political overtones. The groundbreaking book *Teachers as Curriculum Planners: Narratives of Experience* by F. Michael Connelly and Jean Clandinin published in 1988, argued that curriculum development and planning were fundamental activities for teachers (Connelly and Clandinin, 1988). Since then, images of local teachers gathered together to write curriculum has become central to most current visions of the implementation of reforms of education in school – whether driven by disciplinary questions or those of social justice.

Discussion and Issues

This has fallen out in a number of ways, however, and, as an example, we describe two contrasting school districts we have worked with during the late 1980s through the 1990s as part of a larger set of studies on teacher-based curriculum development. In one, town X – with a heritage of libertarianism – until very recently had no curriculum, or rather there were no books or district-wide programmed course of study. Instead, there were lists of things each teacher was expected to cover and, finally, there were the state achievement tests to make sure they did. Each teacher was responsible for coming up with materials, activities, lesson plans, approaches, etc. – whatever was needed to achieve the goals. In town Y – which is next door to town X – there was a much more top-down approach. Here, there were the same lists of goals and outcomes but the administration organized teams of teachers to either evaluate commercial curriculum units and purchase them or to develop curriculum themselves. For example, in science in the late 1980s, the teachers decided to develop their own curriculum. To do this, they applied for grants to fund an infrastructure to support curriculum writing. They visited other school districts engaged in

similar endeavors. They developed a model of what they thought the curriculum should work like and a plan to actualize it. They applied for more grants to fund the teachers writing the curriculum and to do in-service work with the other teachers in the district so they would know how to teach it.

Arguably, what we have described are two models of the profession. In town X, the teacher is assumed the professional and the definition of the professional is that they are able to construct curriculum. In the second example, there is a systemic tie of curriculum development to professional development. In other words, the professional definition of the teacher is not that they know how to construct curriculum. Rather, they get smart and figure out ways to get smarter and they do this together; and this becomes school and professional development.

The distressing thing is that, often and in their varied ways, these processes duplicated the development and implementation years of the 1950s to the 1970s. For example, town Y, in recent years – when professional staff development funds from the state ran low – ended up using their science curriculum similar to the older teacher-proof materials of the 1970s. In town X, enterprising teachers formed affinity groups; so teachers with an interest in social studies, as an example, got together and began study groups developing curricular materials and discussing pedagogical issues. These were wonderful where they existed but often teachers who did not or could not join these groups ended up secretly hoarding ancient textbooks and guiltily teaching from those. The net effect was to reinforce the professionalism of some, but to disempower and isolate others.

Both descriptions illustrate the good and the bad of local school-based curriculum development. In the process of curriculum development, teachers translated the abstractions of policy and standards to the particular and teachable. Moreover, this development role required teachers to reinvent themselves professionally. They became, in practical terms, individuals that reinterpreted curricular ideas – adapting them to suit local circumstances and becoming curriculum developers (Schwab in Westbury and Wilkoff, 1978). Skilbeck's (1984) model of school-based curriculum development describes a form of situational analysis as an important element in the process whereby teachers redesign the curricula they teach. This is central in current manifestations of local school-based curriculum development and how it is placed in school-reform enactment. The net outcome of these processes in the 1990s was to assert the role of teachers in the development of distinctive school profiles and the institutionalization of collectively driven school-improvement plans – both of which are still customs within American schools today.

This professional role in which teachers are engaging in interpretative acts of curriculum construction and are asking themselves and their peers to recreate their

practice is full of conflicts, however. These roles often involve assuming grass roots leadership functions frequently in tension with existing hierarchies and relationships within a school. Teachers assume agency in constructing their interpretations of a reform, and this agency is responsive to, and shaped by, their interactions with others. In the process, reform is interpretative with teachers not as sole reform actors, but relationally positioned in the process of enacting reform. As an example, we describe another school district we have had a long-term relationship with, Z. This vignette is from a series of larger studies conducted by the authors (Barker, 2000; Leander and Osborne, 2008).

In district Z, the teachers in the science committee came to us to help them rewrite their science curriculum. After they outlined their needs, we formulated a threefold approach – ongoing support while they developed and chose curricular materials, a summer science camp in which teachers worked out pedagogical issues, and classroom support while the teachers piloted the units.

The members of the science committee took the lead. The first time that grade-level groups met in the second year of the project, they spent at least half of an in-service day discussing and going through the national and state standards applicable to each grade level. All of the sets of standards state that the learning of science should be a process-oriented experience, promoting science content through inquiry skills, and that the focus should be on the doing and discussing rather than on memorizing specific science content and facts. However, instead of seeing that fact as a powerful reason to make reforms, many of the teachers questioned whether these entities (the standards and the curriculum and pedagogy) were aligned and whether or not curriculum based on them would achieve good test scores.

David was not an original member of the science committee and he did not participate in the summer science camp, but he did come to the first of the grade-level meetings for the third grade. Part of the reason for a change in curriculum and the discussion of the science curriculum, according to David, was a lack of communication between grade levels and within grade levels about what science was taught. The following exchange took place between David and other teachers at the grade-level meeting.

David: So, there are only four unit topics to be covered the entire year? Do you see them lasting a full nine weeks each?

Jennie: Well, there is so much material and the kids are loving it. I'm hoping that I can get it all done before I have to move on.

Gordon (who was the administrator in charge of curriculum for the district): That's the mentality that we have to get away from. Now it's time for Science. Now it's time

for Social Studies. We need to start thinking of the learning as a whole.

David: That's easy for you to say, but I switch with another teacher and am locked into a schedule. I suppose we could combine classes and I teach science every day for a half hour. Will there be enough to last a whole nine weeks?

Liz: It's a different approach. Instead of a smattering of subjects, we are going in depth on a few.

David: So it's like if you teach multiplication in third grade then you don't get it again until it comes up in high school? Once the subjects are taught they aren't covered again.

Liz: But now we are teaching things that aren't appropriate for some grade levels. This scope and sequence looks at the grade levels and it isn't the only time a subject is taught.

The teachers who answered David had already been piloting some of the units in their classrooms or had been a part of the summer science camp where they had tried out the teaching. David had had no experience with the new curriculum, but was concerned about how it would impact the way his classroom and the science curriculum he had been teaching worked. The new program directly challenged the way that he perceived his success in teaching was achieved.

David wanted to make sure that his goals for teaching met the expectations of his community, the state, his administrators, and, most importantly, that they delivered what he believed to be good science teaching. He wanted to know that his students came away from his classroom ready to go to fourth grade, ready to pass the standardized tests, and ready to understand some basic science facts and concepts. David was not convinced that the new curriculum would meet these expectations. At one point, he made a list of the topics covered on the tests to show us. "Think about it. If the tests are in early spring that means my class will have only covered two maybe three topics. That's not what the tests expect that we've covered." David was worried about the test scores and concerned about the amount of content, vocabulary, and time spent on each unit. He was also concerned about the bigger picture of the topics taught and the depth of the topics taught across the grade levels.

The curriculum development and dissemination as well as the distribution of leadership roles among the teachers in district Z did not occur smoothly or without conflict. To the inherent difficulties grass roots leaders in schools face, we would add the question of dissemination and enactment. There are personal and professional tensions between teachers and those actually engaged in curriculum development – hierarchies exist within the teaching community and school and often curriculum developers come into conflict with these and, even when

this is not so, there is a tradition of individuality in classroom teaching that school-based curriculum development comes to confront. Sarason's (1982) mapping of school cultures – including his important analyses of the sociopolitical positions of teachers – critically foregrounds the tensions teachers face within schools, including the ones we are describing. Sarason and others have also noted how it is just these tensions which impede the evolution and change of curriculum and pedagogy.

Summary

This story and the previous highlight many of the conflicts and concerns teachers feel while engaged in local teacher-based curriculum development. Key issues include obtaining the time, space, and funding to support the teachers, obtain materials, hire substitute teachers, and professional support. When engaged in the experimental (Stenhouse, 1980) portion of curriculum development in which new units and their pedagogies are tried out, teachers need classroom support and spaces of relaxed accountability so that they can become risk-takers. School-wide dissemination of teacher-developed curricula is also full of tensions and role shifts. Support structures are rarely in place for this portion of the process. Finally, professional development for teachers involved in curriculum development currently takes the form of programs such as *Learning by Design* (Wiggins and McTighe, 1998). This might possibly include a component of presentation of research into making the 'practical' (Schwab, 1978) practical that goes beyond the action research programs currently in place.

Bibliography

- Barker, H. B. (2000). *Teachers and the Reform of Elementary Science: Stories of Conversation and Personal Process*. Charlotte, NC: Information Age Publishing.
- Bruner, J. (1960/1977). *The Process of Education*. Cambridge, MA: Harvard University Press.
- Connelly, M. and Clandinin, J. (1988). *Teachers as Curriculum Planners: Narratives of Experience*. New York: Teachers College Press.
- Darling-Hammond, L. (1996). The right to learn and the advancement of teaching: Research, policy, and practice for democratic education. *Educational Researcher* **25**, 5–17.
- Leander, K. M. and Osborne, M. D. (2008). Complex positioning: Teachers as agents of curricular and pedagogical reform. *Journal of Curriculum Studies* **40**, 23–46.
- No Child Left Behind Act (2001). No Child Left Behind Act on the United States Department of Education. <http://www.ed.gov/policy/elsec/leg/esea02/index.html> (accessed August 2009).
- Sarason, S. (1982). *The Culture of School and the Problem of Change*. Boston, MA: Allyn and Bacon.
- Skilbeck, M. (1984). *School-Based Curriculum Development*, pp 232–234. London: Harper and Row.
- Stake, R. and Easley, J. (eds.) (1978). *Case Studies in Science Education*, vols. 1 and 2. Urbana, IL: Center for Instructional Research and Curriculum Evaluation.
- Stenhouse, L. (1980). Curriculum research and the art of the teacher. *Curriculum* **1**, 40–44.
- Stronach, I., Corbin, B., McNamara, O., Stark, S., and Warne, T. (2002). Towards an uncertain politics of professionalism: Teacher and nurse identities in flux. *Journal of Education Policy* **17**, 109–138.
- Tyler, R. W. (1949). *Basic Principles of Curriculum and Instruction*. Chicago, IL: University of Chicago Press.
- Westbury, I. and Wilkoff, N. J. (eds.) (1978). *Joseph J. Schwab, Science, Liberal Education, and Curriculum: Selected Essays*. Chicago, IL: University of Chicago Press.
- Wiggins, G. P. and McTighe, J. (1998). *Understanding by Design*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Carnegie Task Force on Teaching as a Profession (1986). *Teaching as a Profession, a Nation Prepared: Teachers for the 21st Century*. New York: Carnegie Forum on Education and the Economy.
- Darling-Hammond, L. and Bransford, J. (eds.) (2005). *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do*. San Francisco, CA: Jossey-Bass.
- National Commission on Excellence in Education (1983). *A Nation at Risk: The Imperative for Educational Reform*. Washington, DC: US Government Printing Office.
- Ruddock, J. and Hopkins, D. (eds.) (1985). *Research as a Basis for Teaching: Readings from the Work of Lawrence Stenhouse*. Portsmouth, NH: Heineman.
- The Holmes Group (1986). *Tomorrow's Teachers*. East Lansing, MI: Holmes Group.

Further Reading

CURRICULUM DEVELOPMENT – CONTEXTS

Contents

Curriculum and Human Rights

Curriculum and the Education of Cultural and Linguistic Minorities

Curriculum in Postcolonial Contexts

Curriculum, Economic and Cultural Development

Globalization and Curriculum

Popular Culture

Curriculum and Human Rights

C Lohrenscheit, German Institute for Human Rights, Berlin, Germany

© 2010 Elsevier Ltd. All rights reserved.

Introduction: Linking Education and Human Rights

Education is a global fact. Education systems, and especially school systems, worldwide follow similar patterns, for example, in the choice of school subjects, in processes of curriculum planning and development, in the defined aims and goals, as well as the forms and styles of teaching and learning. At the same time, of course, every education system has its own historical, economic, social, and cultural context with specific problems and challenges that must be taken into account in curriculum development processes. Nevertheless, there are also common goals and principles in education which are shared by a growing number of pedagogs and experts worldwide. Such common goals and principles can be found in the documents and declarations of world education conferences, for example, the two World Education Forums at Jomtien (1990) and then Dakar (2000), the Global Campaign for the Right to Education which was launched in November 1999, or the Decade for Human Rights Education of the United Nations (1995–2004). They continue to develop in the struggle for freedom and equality in education and for a world where human beings share a life in peace. It is in this context where human rights become an important tool because they define basic norms for all actors involved in education: teachers, learners, and parents, as well as all actors involved in administration and planning processes.

For more than a decade, human rights are becoming more and more powerful in the international educational discourse. To link education with human rights poses questions to the existing forms and structures of education systems, in curriculum planning and development, as well as teacher training and the review of learning materials and text books. This article focuses on the three core

dimensions which must be taken into account when rethinking curriculum development within the context of human rights and through the lens of human-rights-based approach:

- the right to education as it is enshrined in international human rights covenants,
- rights in education and in the processes of curriculum planning and development, and
- human rights education as a guiding principle for curriculum development.

All three dimensions share important concepts which also constitute our modern understanding of human rights. At the core lies the understanding of individuals – learners as well as teachers – as human beings free and equal in dignity and rights. Freedom rights involved in education are, for example, the right to freedom of expression and participation, religious freedom, or the freedom of parents to choose for the education of their child according to their own beliefs and cultural values. Equality rights call for an inclusive approach in education and curriculum development. They try to embrace each learner, respecting the principal equality of each individual as well as everyone's right to be different. In pragmatic terms, this means that all forms and stages of education (curriculum development, text books and learning materials, teaching and learning styles, languages of instruction, teacher education, to list some examples) must be adapted to the core principle of nondiscrimination – which in turn means inclusion.

To make curriculum development a powerful tool for the stimulation of these core concepts, it has to create a framework in which every learner is seen as a subject to human rights (and especially the right to education) while education serves the aim of developing the full potential of the personality, as well as respect and competences for

realizing human rights and the values enshrined therein. A curriculum (curriculum is defined as a running, course, career, and is derived from Latin *currere*) entails more than just the aims and goals of a learning process or a study course. In the following, curriculum development is therefore defined in a broad sense, not only spelling out contents of education but also focusing on the process of learning, the broader framework in which learning takes place (e.g., the context of the school, teachers, learners, and parents) and the aims and goals of the learning processes which underlie continuing changes worldwide.

Creating a Common Language: What Do We Understand by the Human Right to Education?

The right to education is an empowerment right. It is not only a right, protected through international covenants, but also a tool for claiming one's own rights as well as – in solidarity – the rights of others. The best basis for understanding the right to education is found in the context of the United Nations (UN) specialized organizations for the promotion and protection of human rights. Therefore, in the following, the main sources for explaining what it means to fully realize the right to education originate from the human rights bodies of the UN, especially the reports of the UN Special Rapporteur on the right to education.

Historically, education has developed from being a privilege for some into being a basic human right for all. In the framework of the modern development of human rights, after the end of World War II, education was defined as a human right for the first time in the Universal Declaration of Human Rights (UDHR) of the UN. Here, it is crucial to also understand the historical background of the time: the UN were founded in 1945 and declared the UDHR 3 years later in 1948. It was in the spirit of never again that the community of nations decided for a human rights project. Never again should the world experience as massive violations of human dignity and rights as during the time of the Nazi regime and World War II. Concerning education, never again should education be misused to penetrate values into learner's minds that propagate hatred, racism, and anti-Semitism. To the contrary, education should be a catalyst to promote the spirit of humanity, human rights values, and the goals of the UN for the promotion of peace and justice in the world. Consequently, curriculum development should encompass the notion of human rights.

At the time, when the UN declared the UDHR, it was not clear that this document would develop into a powerful tool. However, in the following years, "to the astonishment of many, human rights would become a political

factor that not even the most hard-shelled realist could ignore. The UDHR would become an instrument, as well as the most prominent symbol, of changes that would amplify the voices of the weak in the corridors of power" (Glendon, 2001: xvi). The Article 26 of the UDHR defines the human right to education and has hence been the basis for all other formulations and definitions to follow. It founds the universal basis for education, including curriculum development, and it defines standards for quantitative as well as qualitative aspects of education.

- (1) Everyone has the right to education. Education shall be free, at least in the elementary and fundamental stages. Elementary education shall be compulsory. Technical and professional education shall be made generally available and higher education shall be equally accessible to all on the basis of merit.
- (2) Education shall be directed to the full development of the human personality and to the strengthening of respect for human rights and fundamental freedoms. It shall promote understanding, tolerance and friendship among all nations, racial or religious groups, and shall further the activities of the United Nations for the maintenance of peace.
- (3) Parents have a prior right to choose the kind of education that shall be given to their children. (Franklin and Eleanor Roosevelt Institute, 2001).

The quantitative notion of the right to education is simply that everyone is entitled to have access to this right. Most importantly, this is directed not only to children but also to adults, who, in their earlier life, did not have the chance to receive education, to have the right to a basic education to be able to participate in society. The state must ensure that every child – no matter if girl or boy, poor or rich, black or white, enabled or disabled – is included in education. For curriculum development processes, this means that they should start with asking: Who is the target of the curriculum, which learning groups are included or excluded, and why?

The qualitative notion of the right to education concerns the learning contents as well as the methods and didactics of teaching and learning, which again bears consequences for curriculum development. To further explain the essential features of the right to education and to create a common language, basic criteria are established to lead the way to accountability and commitment in a global context. The first UN Special Rapporteur on the right to education, Katarina Tomaševski, developed the 4-A scheme which defines the core obligations to fully realize qualitative as well as quantitative dimensions of this right, namely availability, accessibility, acceptability, and adaptability (Tomaševski, 1999). She explains that "education operates as a multiplier, enhancing the enjoyment of all individual rights and freedoms where the right to

education is effectively guaranteed, while depriving people of the enjoyment of many rights and freedoms where the right to education is denied or violated” (Tomaševski, 2001: 7).

Availability

For education to be inclusive and equally available for everybody, functioning structures and institutions must be in place. Format and shape of these institutions are of course bound to the context, but some basic requirements have to be met like, for example, buildings, where education can take place, with sanitary facilities and hygienic drinking water. In addition, teachers should be able to receive high-quality training before they start teaching in schools or other learning institutions. The teaching profession should be valued and wages shall be granted accordingly. This also means that the status of professional educators must correspond to their international recognized rights, including their right to form trade unions. Finally, availability must recognize not only the parental choice of education for their children but also their responsibility to ensure that they go to school.

Access

Access to education is bound to its compulsory nature. All school-aged children must be granted access free from any form of discrimination. This entails at least two dimensions, namely that compulsory education must be affordable and that exclusion of children from education on the grounds of sexist, racist or homophobic criteria, gender, physical or mental ability, age, social or marital status, and family background is prohibited. Here, the right to education pays attention especially to disadvantaged groups of learners who often need the chance to compensate what they might not get from their families or communities. Some children might need special attention in education because they have learning difficulties which must be taken into account. They too have the right to be treated equally and not in special institutions. Consequently, all obstacles that reject children's access must be eliminated, such as costs for education (school fees, books, transportation, uniforms, and school meals) or conditions which keep children out of school because families are too poor or boys and especially girls have to work to sustain the household of the family instead of going to school.

Adaptability

Education must be adapted to the needs of the learners. Therefore, curriculum development must be a flexible, continuing process because requirements and challenges of societies are constantly changing. Every curriculum

must take the diversity of learners as well as their diverse backgrounds and needs into account. This may also include, for example, different cultural or religious backgrounds or the home languages of the learners which may be different from the dominant language or *lingua franca* of a respective country. For the analysis of the adaptability of education, criteria are required which also include out-of-school children. Here, government's reports to the human rights bodies of the UN revealed that more than 30 categories of children are likely to be excluded from and in education (e.g., children who need to work, child mothers or pregnant girls, disabled children, HIV-infected children, homeless children, street children, or refugee children). Furthermore, curriculum development must be evaluated on the basis of a sound analysis of the impact of education on human rights and human rights values. Curriculums must be designed to assist children (and their teachers) to fully understand their rights and freedoms and to be able to apply them.

Acceptability

Education must be acceptable in terms of its form and content. The acceptability of education includes important norms concerning the contents and methods of learning, which must be relevant, suitable, and of a high quality. Curriculum development, therefore, shall be in line with the foundations of the right to education, as cited above in the article 26 of the UDHR: “Education shall be directed to the full development of the human personality and the promotion of respect for human rights and freedoms” (Franklin and Eleanor Roosevelt Institute, 2001). Minimum standards are required for the teaching process as well as the learning process. Additionally, curriculum developers should strive to integrate human rights as a content as well as guiding principle for the learning process. Both elements – adaptability as well as acceptability of education – are often hindered through curriculum development if it is concentrating solely on the preparation of children for the next level of education, to which some children maybe unable to proceed in any case.

Human Rights in Education

The second dimension of the human-rights-based approach to curriculum development is linked to the rights of all actors involved (teachers, learners, and parents) and offers safeguards for the prevention of human rights abuses in education. The full realization of the right to education involves three key actors: (1) the child as the principal bearer of the right, who also has the duty to participate since education entails an interactive dimension between the teacher and the learner, (2) the parents who shall stimulate

learning in the home environment and who must ensure that a child has access to schooling, and (3) the government as the provider and funder of public schooling – including education and training of teachers.

Children's Rights

Children's rights cover the life span between the ages of 0 and 18 years. They are protected through a special convention in international law since children are the easiest victims of abuses of power by their parents, teachers, and/or governments. Special human rights norms for children are necessary because of young persons' vulnerability, since history has proven that parents and teachers often do not have the best interest of the child in their hearts and minds. Children's rights include, for example, health rights, their freedom of expression, their right to participate in all matters that directly influence their well-being or their right to be protected from any form of violence or discrimination. At the core of the UN Convention on the Rights of the Child (1989) lays the best interest of the child, which consequently also constitutes the cornerstone for all curriculum development processes. The child itself is the subject to rights, and this must be given priority in any educational planning and practice. To put the best interest of the child at the core of curriculum development requires considering children not as passive recipients of education but as actively involved. Children, in turn, have the duty to participate in education since only through the participation of both, teachers and learners, the interactive notion of the right to education can be put in practice. It still is a great challenge to make children legally the subjects of the right to education instead of objects of agreements between parents and schools. Schoolchildren often express that they feel it is their right to be consulted with regard to their own education. They could also be involved in curriculum planning and development processes.

Parent's Rights

As cited in the article 26 of the UDHR above, respect for parental freedom indicates that they are able to choose for an education for their child in accordance with their religious, moral, or philosophical convictions. Nevertheless, human rights law also provides prohibitions against indoctrination, abuse, or exploitation of children. The rationale behind parental choice is not to recognize their denial of their children's rights, but to assure that they can choose for an education which is relevant, suitable, and of a high quality. Curriculum development must therefore allow for a diversification of learning contents as well as for teaching styles to include the diverse perspectives and backgrounds of children and their families. Freedom of choice, in this way, prevents state monopoly over education and protects pluralism.

Teacher's Rights

Education cannot take place without teachers. Although recent developments have shown that technological support for education could offer a lot of learning possibilities like, for example, webucation or distance learning, the teacher makes all the difference; even more if we take into account that children (but also adults) to a great extent learn through example, imitation, and role models. The problems that teachers are facing range from their low status (in some countries), discrimination, low salaries, or the denial of trade union rights including their right to strike, to the protection of their professional and academic freedoms (e.g., freedom of expression, freedom of assembly, religious freedom, and freedom from discrimination). To make sure that educators are in a position to actively promote human rights, their profession must be valued. Like no other profession, the teaching profession supports the intergenerational dimension between children and adults, which is a great challenge. Teachers shall not only equip children with the competences and skills needed for societies today, but they must also bear in mind that as future adults they may need other skills than the ones currently defined. To put this intergenerational dimension in the words of the UN Special Rapporteur on the right to education: "The knowledge, skills and values that the generation of future adults will need in their lifetime is not only unknown but unknowable" (Tomaševski, 1999: 24).

Prevention of Human Rights abuses in Education

Human rights, like the ancient Greek figure of Janus, are double faced. While one face turns to the full realization of rights, the another points to massive violations worldwide, and the unfulfilled promises which human rights stand for in many parts of the world. Every right can be violated, and the right to education as well as the rights in education are no exception. Violations of fundamental rights range from the denial of equality, or the rejection of privacy and free development of the personality, to freedom from fear and violence. Schools must be a safe environment for both, teachers as well as learners. But although in the past decades physical punishment of children has been abolished in most countries of the world, humiliating treatment of children – whether through other children or their teachers – still remains an enormous challenge for the creation of a safe learning environment. The first and utmost priority of every education should be that children learn to enjoy learning and that they like to go to school without fears of being a target of bullying or discrimination. Human rights abuses often may lead to the opposite. The challenge for curriculum development is therefore to provide a sound basis for the prevention of human rights abuses and the provision of tools to enable learners as well as

teachers to realize human rights values in their interactions. Contradictory practices in learning contents may occur compared to the styles of teaching. In addition, school discipline may conflict with any human rights messages verbally conveyed. Curricula and school textbooks repeatedly reproduce sexist or racist stereotypes, for example, if women are portrayed as staying at home, caring for the children while men go out to work in the public; or if black people are portrayed in low-status jobs rather than in leading positions.

Human rights law puts utmost priority to the elimination of inequalities. Education can be used as a means to both maintain and eradicate stigmatization, discrimination, and exclusion. Separate educational facilities that promote selection and segregation of children have proven to be inherently unequal. This insight goes back to the decision of the US Supreme Court in the famous case of *Brown versus Board of Education of Topeka* (Judgement of 17 May 1954, see E/CN.4/2000/5, page 28). To the contrary, inclusive education provides examples of how all children can benefit if they are in a position to learn from one another, embracing the differences and the diversity of all learners.

Human Rights Education as Guiding Principle for Curriculum Development

The third feature of the human-rights-based approach to curriculum development is human rights education itself as a guiding principle. Since the beginning of the 1990s, an immense amount of literature and learning material has been developed – especially since the UN has launched the Decade for Human Rights Education (1995–2004). Since education operates as a multiplier and is a passkey for unlocking other human rights, knowledge of human rights should become a priority in education policies. Through this conceptual shift, education will be a keystone for self-sustaining livelihoods as well as an instrument for enhancing the enjoyment of all human rights and freedoms.

There are three core elements which constitute human rights education and which can guide the curriculum development process, namely learning about human rights, learning through human rights, and learning for human rights. All three dimensions can be portrayed in a human rights education learning cycle where one element builds upon the others (Figure 1). Every stage, of course, must be adapted to the particular learning groups, concerning their ages, their previous knowledge and experience of human rights issues, or their cultural or religious backgrounds.

Learning about Human Rights

Learning about human rights covers explicit acquisition of knowledge. Too often curriculum development reduces human rights knowledge to an implicit approach,

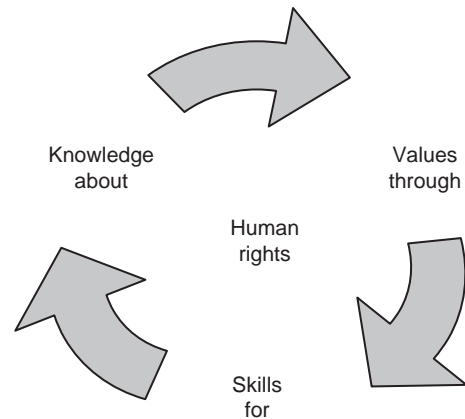


Figure 1 The cycle of human rights education – Learning about, for, and through human rights.

assuming that teaching about democratic values or civic education automatically includes a human rights dimension. This is rarely the case. Learning about human rights encompasses the history and development of human rights around the world. It includes knowledge about human rights conventions (civil and political rights, economic, social and cultural rights, children's rights, women's rights, protection against racism or degrading, and humiliating treatment), specific instruments and the various actors in the human rights field on the domestic as well as on the international level (government actors, intergovernmental actors such as the UN or the Council of Europe, nongovernmental organizations (NGOs), human rights defenders, and social movements). Besides internationally accepted human rights norms, knowledge about human rights also needs to be contextualized to the reality of teachers and learners. Here, it is important not only to reduce the teaching about human rights on a Eurocentric perspective but also to include important regional or local developments.

Learning through Human Rights

Learning through human rights incorporates human rights values and principles into the learning process itself. It enables children to experience the values of a community when the learning environment supports dignity of, and respect for, all individuals. As already stated above, methods and styles of teaching and learning may conflict with any human rights messages verbally conveyed. Therefore, it is of utmost priority that the learners are in a position to develop attitudes that help them to develop morally and prepare them for active participation in society.

Learning for Human Rights

Most probably, learning for human rights is the most complicated part of the learning cycle since it includes

the acquisition of skills and competences to respect and realize human rights and the values enshrined therein. Learning for human rights enables the learners to play an active role in shaping the community and/or society on the grounds of liberating, antioppressive strategies. Basic competences include communication skills, analytical skills, critical reflection skills, and problem-solving and cooperation skills. This part of a human rights curriculum may lead to children wishing to actively engage in human rights groups or to the stimulation of immediate changes in the learning environment, which sometimes poses great challenges to teachers as well as school headmasters or parents. Here, it is important not to frustrate children's expectations but to carefully interact with them and support them with their new ideas of transforming their direct environment to a world where human rights can become real.

To conclude, one could ask, if all curriculum development shall promote human rights education or if all education inherently shall encompass a human-rights-based approach. The answer is yes. There is a fundamental link between human rights and education. The right to education not only includes general guidelines on what has to be learned but also how to do it: with respect and in dignity. It provides core elements, relevant for all curriculum development processes, and it could inspire collective efforts for education as a tool for liberation.

See also: Gender and Curriculum; Peace Education; Peace Education.

Bibliography

Franklin and Eleanor Roosevelt Institute (2001). *Universal Declaration of Human Rights: Preamble*. <http://www.udhr.org/UDHR/default.htm> (accessed April 2009).

Glendon, M. A. (2001). *A World Made New. Eleanor Roosevelt and the Universal Declaration of Human Rights*. New York: Random House.

Tomaševski, K. (1999). *Preliminary Report of the Special Rapporteur on the Right to Education*, Report No. E/CN.4/1999/4913. New York: United Nations: Economic and Social Council.

Tomaševski, K. (2001). *Preliminary Report of the Special Rapporteur on the Right to Education*, Report No. E/CN.4/2001/52. New York: United Nations: Economic and Social Council.

Further Reading

Andreopoulos, G. J. and Claude, R. P. (1997). *Human Rights Education for the Twenty-First Century*. Philadelphia, PA: University of Pennsylvania Press.

Tomaševski, K. (2002). *The Right to Education*, Report No. E/CN.4/2002/60. New York: United Nations Economic and Social Council.

Tomaševski, K. (2003). *Education Denied: Costs and Remedies*. London: Zed Books.

Tomaševski, K. (2003). *The Right to Education*, Report No. E/CN.4/2003/9. New York: United Nations Economic and Social Council.

Tomaševski, K. (2004). *The Right to Education*, Report No. E/CN.4/2004/45. New York: United Nations Economic and Social Council.

Relevant Websites

<http://www.amnesty.org> – Amnesty International.

<http://www.campaignforeducation.org> – Global Campaign for the Right to Education.

<http://www.hrea.org> – Human Rights Education Associates.

<http://www.pdhre.org> – People's Decade for Human Rights Learning.

<http://www.right-to-education.org> – Right to Education (Website of the first Special Rapporteur on the Right to Education of the UN, Katarina Tomaševski).

<http://www.save-the-children.org> – Save the Children.

<http://www.unhcr.org> – United Nations High Commissioner for Human Rights.

<http://www.udhr.org> – Universal Declaration of Human Rights.

Curriculum and the Education of Cultural and Linguistic Minorities

S May, University of Waikato, Hamilton, New Zealand

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Additive bilingualism – A context in which both languages are valued, encouraged, and used in the teaching and learning process.

Bilingual education – An educational approach that uses two languages as media of instruction (i.e., subjects are taught in and through two languages).

Biliteracy – The ability to read and write (as well as speak) in two languages.

Immersion education – A form of bilingual education, where L1 majority language students are taught predominantly through an L2, usually a minority language, in order to become bilingual and biliterate in that language as well as the majority language.

L1 – The first language a person learns and uses in the family.

L2 – The second language a person learns, either in childhood or adulthood.

Language interdependence – The widely accepted view that the level of competence that a child attains in their L2 is dependent to a large extent on the level of competence already achieved in their L1.

Monolingualism – Using or knowing only one language.

Submersion programs – Where L1 minority language students are taught solely through a majority language in a society as an L2. Though often confused with immersion education, a form of bilingual education, it is clearly not, as eventual monolingualism in the majority language is its key aim and students' L1 is excluded from the teaching and learning process.

Subtractive bilingualism – A negative or deficit conception of students' bilingualism, where the students' L1 is seen as interfering with the learning of another language, rather than supporting it.

This article explores the curricular and pedagogical implications of teaching cultural and linguistic minority students. While a number of substantive fields – most notably, multicultural education and bilingual education – address this question in detail, they often do so from different directions and in relation to different audiences. It is not possible to address the multiplicity of debates engendered in and by these different educational fields. Accordingly,

this article is limited to an exploration of the bilingual education literature, and focuses, in particular, on the question of which educational approaches are most effective in relation to teaching bilingual students.

Before doing so, however, one needs first to clarify the term bilingual education. This is important because commentators employ widely different understandings of what such an education actually constitutes. At one end of the continuum are those who would classify as bilingual any educational approach adopted for, or directed at bilingual students, irrespective of their educational aims (fostering bilingualism or monolingualism) or the role (if any) of first language (L1) and second language (L2) as languages of instruction. In other words, simply the presence of bilingual students in the classroom is deemed sufficient to classify a program as bilingual. At the other end of the continuum are those who distinguish clearly between nonbilingual, weak, and strong bilingual programs. The latter analysis – now widely accepted – forms the basis of this article.

Defining Bilingual/Immersion Education

There are a plethora of existing typologies with respect to bilingual/immersion education in the research literature, although, as one might expect, they do not always correspond or overlap, depending on the initial starting point, and position of the researcher. Some of the most accessible and informed are summarized in Baker (2006).

Before unpacking the characteristics of bilingual/immersion education further in light of these typologies, however, it is useful to begin with a classic definition of bilingual education, first posited by Andersson and Boyer:

Bilingual education is instruction in *two languages* and the use of those two languages as mediums of instruction for any part or, or all, of the school curriculum. (Andersson and Boyer, 1970: 12, original emphasis)

Put simply, bilingual education involves instruction in two languages (see also Baker and Prys Jones, 1998). This immediately excludes programs that include bilingual students but do not involve bilingual instruction, most notably submersion majority language programs, where students are taught only in the majority language, irrespective of their language background. It also excludes programs where an L2 is taught as a subject only. English as a second language (ESL) classes, which include the

Sheltered Instruction approach that is increasingly popular in the US, are examples of this, as are foreign language classes. Along with submersion programs, they can also clearly be described as nonbilingual programs.

For a program to be deemed to be bilingual, the key is that both languages must be used as media of instruction and thus to deliver curriculum content. As Baker and Prys Jones (1998: 466) conclude: “If there is a useful demarcation, then bilingual education may be said to start when more than one language is used to teach content (for example Science, Mathematics, Social Sciences, or Humanities) rather than just being taught as a subject by itself.” On this basis, immersion models that teach majority language students predominantly through a minority language, such as French-immersion programs in Canada or Maori immersion programs in New Zealand, are also clearly bilingual programs, since some curricular instruction in the majority language (English, in both cases) almost always occurs at some point prior to the end of the program, even in those programs with very high levels of immersion in the minority language.

An additional key point addressed by many commentators in defining bilingual education relates to the philosophy and related educational goals of any given program. In short, does the program in question aim to achieve, foster, and/or maintain longer-term student bilingualism and biliteracy, adding another language to the student’s existing language repertoire, which has come to be termed in the research literature as an additive approach to bilingualism? Or does it eventually aim to shift students from bilingualism to monolingualism in the dominant language, losing or replacing one language with another, a process that has been described as subtractive bilingualism?

First postulated by Lambert in Canada in 1974, the additive–subtractive distinction is also useful for another reason. Research over the last 30 years has consistently demonstrated that those programs which are most likely to achieve bilingualism and biliteracy for their students – that is, additive bilingual programs – are also the most likely to see those students succeed educationally. In contrast, subtractive programs not only atrophy their students’ existing bilingualism, but also exhibit far lower levels of educational success for these students, particularly over time (see Baker, 2006; Cummins, 2000, see also below).

With this broad distinction between additive and subtractive bilingualism clearly outlined, the next level of classification of bilingual programs can now be made in terms of the specific linguistic and/or educational aims of particular bilingual education models. According to Freeman (1998: 3), models are defined in terms of “their language-planning goals and ideological orientations toward linguistic and cultural diversity in society.” They can be understood as broad categories that help us to understand on a very general level what bilingual education

means, although there is inevitably a degree of arbitrariness in distinguishing among them.

Despite the welter of different classifications of bilingual education in the research literature, there are three broad models of bilingualism that are consistently included in these various typologies. These are transitional models, maintenance models, and enrichment models of bilingual education. In addition to these three broad models, there are also what have come to be known as heritage models. These are most often associated with indigenous language education initiatives, such as Maori-medium education in New Zealand, Navajo language education in the US, Quechua language education programs in Peru, and Sa’mi language education in Norway, among many.

A transitional model of bilingual education uses the L1 of minority language students in the early stages of schooling but aims to shift students away from the use of their L1 as quickly as possible toward the greater use of the dominant language, in order to cope academically in mainstream or general education (Freeman, 1998). In other words, the L1 is used only to the extent that it facilitates the transition of the minority language (L1) speaker to the majority language (L2). Accordingly, most transitional programs are also early-exit programs, where the L1 is used for only 1–2 years, before being replaced by the L2, and can thus be regarded as both a subtractive and weak bilingual model. In assuming that the (minority) L1 will eventually be replaced by a (majority) L2, bilingualism is not in itself regarded as necessarily beneficial, either to the individual or to society as a whole. This in turn suggests that the eventual atrophy of minority languages, or the aim of moving eventually from bilingualism to monolingualism in the majority language, remains a central objective of transitional bilingualism programs. Until their recent demise in a political climate, largely antithetical to any kind of bilingual program, a transitional bilingual approach was most prominent in the US, where transitional bilingual programs were developed widely for Spanish (L1) speakers from the 1970s onward.

A maintenance approach to bilingual education, on the other hand, differs fundamentally from a transitional approach because it aims to maintain the minority language of students, strengthen students’ sense of cultural and linguistic identity, and affirm their individual and collective ethnolinguistic rights. As such, it is clearly an additive and strong bilingual model. There are many types of bilingual program that can be said to fit into this model and these will be discussed in detail later. However, the typical participant in a maintenance bilingual program will be a national minority group member (e.g., Welsh in Britain, Catalan in Spain, French Canadian in Canada, Latinos in the US), whose L1 is already developed to an age-appropriate level (although they do not need to be literate yet in the language). The language of

instruction of the program will either be predominantly in the L1 or, if both L1 and L2 are used as mediums of instruction, at least 50% in the L1. This is because the aim of such programs, as their designation suggests, is to maintain the L1 for a sufficient amount of time so that academic language proficiency in the L1 is achieved. This in turn facilitates the acquisition of literacy in an L2, on the basis of the developmental interdependence principle (see Cummins, 2000). Consequently, the most common programs in a maintenance bilingual model are late-exit programs – that is, the use of L1 as an instructional language continues for at least 4 years, often longer.

Closely related to maintenance bilingual programs are enrichment programs. If the former are geared toward maintaining the L1 of minority language students, the latter are generally (but not exclusively) associated with teaching majority language students (such as L1 English speakers) through a minority target language. French immersion in Canada, where many of the students come from middle-class L1 English-speaking homes, is perhaps the most often cited example of an enrichment bilingual program. Welsh-medium schools, which also include many middle-class L1 English speakers, are another example. Elite bilingual programs such as the European Schools movement are also widely regarded as enrichment programs.

As with maintenance programs, the emphasis in enrichment programs is not just on achieving bilingualism and biliteracy for individual students but also on the ongoing maintenance of the minority language(s) in the wider community. As Hornberger (1991) argues, the enrichment model “encompasses all those bilingual education program types which aim toward not only maintenance but development and extension of the minority languages, cultural pluralism, and an integrated national society based on autonomy of cultural groups” (p. 222). Accordingly, she asserts that this type of program has the greatest potential to educate students successfully, given its strong additive bilingual basis. It is also the program most likely to reduce the educational and wider social and linguistic inequalities experienced by minority language speakers.

This broad L1/L2 distinction between maintenance and enrichment approaches is a useful one, or at least a useful form of shorthand, in the research literature. However, it does not necessarily help us to identify clearly where a heritage language model of bilingual education might fit in. As indicated above, this model is most commonly associated with indigenous language revitalization efforts, along with a wide range of other indigenous language education initiatives, although in its wider sense, it can also include other established and immigrant groups (Wiley, 2001). The latter tend to be focused on the reclamation of a heritage language no longer spoken as an L1, that is, the students are second language learners of the heritage language. The former include a combination of

student language backgrounds. Some indigenous language programs (e.g., Navajo; Hualapai in the US; Inuit in Nunavut, Canada; Sa’mi in Finnmark, Norway) are aimed at students who still speak the indigenous language as an L1 and may therefore be regarded as L1 maintenance bilingual programs. But many also cater to students with a mix of L1/L2 speakers of the language (Maori in New Zealand, Hawaiian), and some have only L2 speakers (or rather, learners) of the language (the Master/Apprentice program developed for the now largely moribund indigenous languages of California) and are therefore closer to the enrichment end of the continuum.

As such, heritage programs can also clearly be regarded as an additive and strong bilingual approach, but tend to be situated somewhere in between maintenance and enrichment models in terms of the L1/L2 status of their students (May and Hill, 2005). That said, increasingly, the majority of students in such programs tend to be second language speakers of the target language, the result in turn of previous patterns of language shift and loss of the heritage language. For example, McCarty (2002) notes that in the Navajo heritage language program at Rough Rock in Arizona – one of the strongest and longest established in the USA – only 50% of Navajo now speak their own language and their numbers are declining each year. In Maori-medium education in New Zealand, the overwhelming majority of students are first language English speakers (May and Hill, 2005).

The final level at which bilingual/immersion education can be examined is the program level, which is also, necessarily, the most complex and diffuse. According to Hornberger (1991), bilingual programs are more concrete categorizations than models, and can be differentiated from one another by an analysis of specific contextual and structural characteristics. For Hornberger, contextual characteristics include characteristics of the student population (numbers, stability/mobility in the school, SES, minority status, and language background) and characteristics of the teacher population (ethnic background, degree of bilingualism, training, roles). Structural characteristics include programs in school (whether school-wide or targeted), languages in curriculum (sequencing, oral/literate development, and subject allocation of the languages), and classroom language use (patterns and functions).

There is no space here to discuss the complexity of programs involved (for an exemplary extended analysis, see Baker, 2006), except to highlight – in light of the preceding discussion – the most common types of programs. Nonbilingual programs include Submersion, ESL, and Sheltered Instruction programs (all subtractive programs). Bilingual programs include weak (and subtractive) bilingual programs, such as Transitional Bilingual Education, where use of the students L1 is limited usually only to the first years of schooling. Strong (and additive) bilingual programs include L1 Maintenance Bilingual

programs, Immersion and Heritage programs. These programs, which have also been termed One-way programs (Thomas and Collier, 2002), tend to vary in terms of both their level of immersion in the minority or target language and the related timing and balance of instruction in the majority language. However, most of these programs will use the minority or target language as the medium of instruction for between 50% and 90% of the time. For example, the program may begin as a 90:10 program in the early years (with 90% in the minority or target language) and change gradually to a 50:50 program by year 4 of the student's schooling.

A variation of this Oneway approach, increasingly popular in the US most often with respect to Spanish-English bilingual instruction, has come to be termed Two Way Immersion or Dual Language immersion. The aims of Two Way immersion are the same as other strong, additive programs – bilingualism and biliteracy for their

students. However, unlike other forms of immersion, Two Way programs include native speakers as well as non-native speakers of the target or minority language in the same classroom, wherever possible, in roughly equal proportion. These programs thus specifically integrate English L1 students and target-language (e.g., Spanish) L1 students with the goals of developing the bilingual and biliterate skills of both groups (Lindholm-Leary, 2001).

The discussion thus far can be summarized, albeit somewhat simplistically (**Figure 1**), where the left-hand side can be equated with subtractive approaches and the right-hand side with additive approaches to bilingual students. As we shall see, addressing these various dimensions of bilingual/immersion education is a necessary prerequisite for understanding what research has subsequently found in relation to the relative efficacy of the various approaches just described. This research is discussed in the following section.

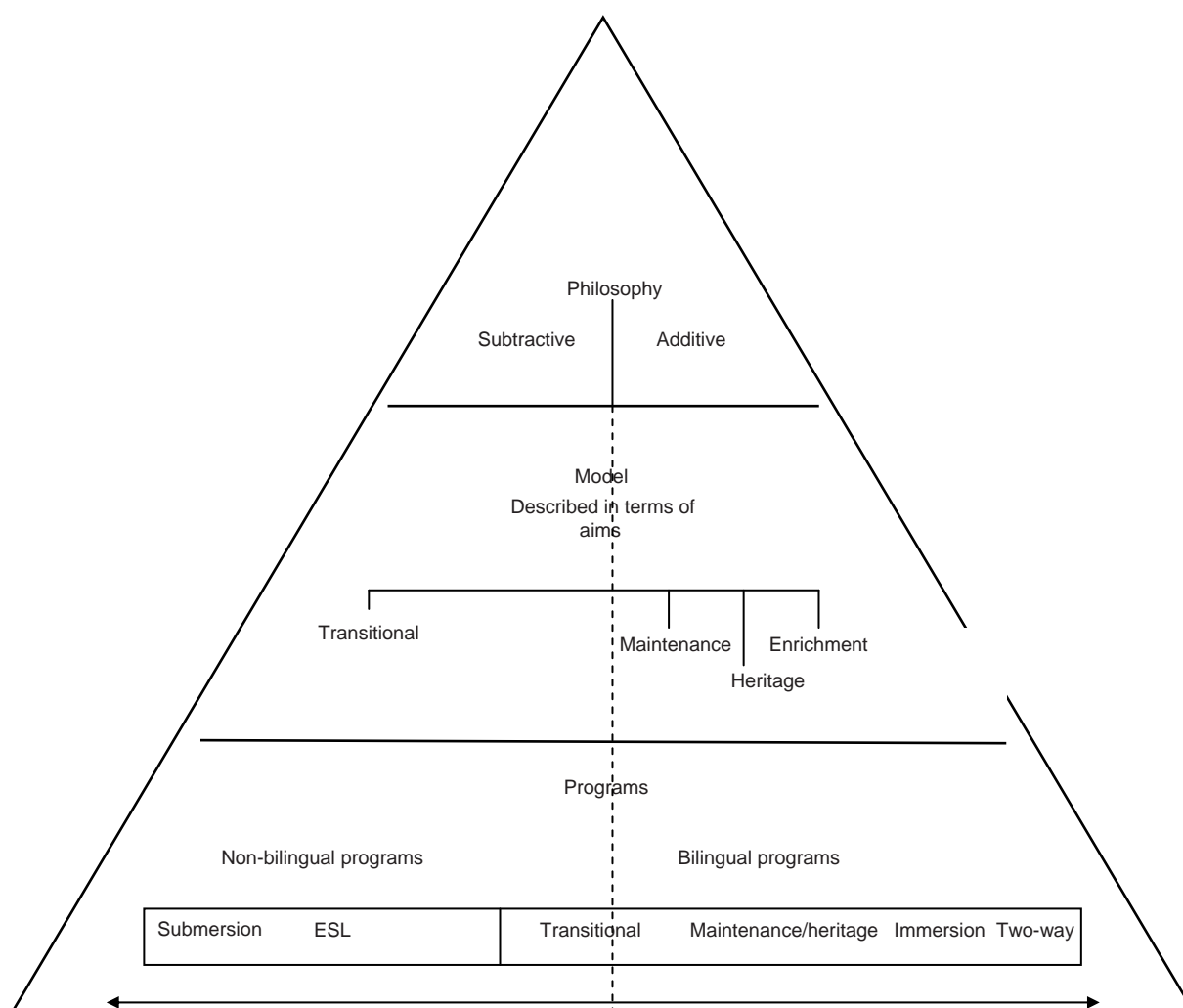


Figure 1 Principal axes of bilingual/immersion education. From May, S. (2008). Bilingual/immersion education: What the research tells us. In Cummins, J. and Hornberger, N. (eds.) *Bilingual Education. The Encyclopedia of Language and Education*, 2nd edn, vol. 5, pp19–34. New York: Springer. <http://www.springerlink.com/content/pj85556182584420/>, with kind permission by John Benjamins Publishing Company.

Educational Effectiveness and Bilingual Students

Bilingual education, as an educational approach, continues to generate controversy in the wider sociopolitical realm – particularly in those contexts, such as the US, where a subtractive view of bilingualism remains widespread, and where English-only approaches to schooling predominate in the US (see Cummins, 2000). What is interesting about these wider debates is their consistent tendency to ignore a compelling consensus in the research literature on the educational benefits, and efficacy, of strong, additive approaches to teaching bilingual students (see Ramírez *et al.*, 1991; Thomas and Collier, 2002).

Ramírez, Yuen, and Ramey (1991), for example, compared English-only programs with early-exit (1–2 years) and late-exit (4–6 years) bilingual programs, following 2352 Spanish-speaking students in US schools over 4 years. Their findings clearly demonstrated that the greatest growth in mathematics, English language skills, and English reading was among students in late-exit bilingual programs where students had been taught predominantly in Spanish (the students' L1) – equivalent to One Way Maintenance bilingual programs. These same students also demonstrated significantly better academic progress than those who were transferred early into all-English instruction. Ramírez *et al.* conclude that:

Students who were provided with a substantial and consistent primary language development program learned mathematics, English language, and English reading skills as fast or faster than the norming population in this study. As their growth in these academic skills is atypical of disadvantaged youth, it provides support for the efficacy of primary language development facilitating the acquisition on English language skills. (Ramírez *et al.*, 1991: 38, 39)

In contrast, the Ramírez study also confirmed that minority language students who receive most of their education in English rather than their first language are more likely to fall behind and drop out of school. In fact, it is important to note here that the English-only programs used for comparison in the Ramírez study were not typical to the extent that while the teachers taught in English, they nonetheless understood Spanish. This suggests that in the far more common situation where the teacher does not understand the students' L1, the trends described here are likely to be further accentuated.

In the largest and most recent study conducted to date, Thomas and Collier (2002) came to broadly the same conclusions. Thomas and Collier analyzed the education services provided for over 210 000 language minority students in US public schools and the resulting long-term academic achievement of these students. They did so by examining in depth five urban and rural sites from throughout the US over 5 years, from 1996 to 2001.

The school bilingual program types examined within these contexts varied widely – they included full immersion programs in a minority language, dual-medium or two-way programs, where both a minority and majority language (usually, Spanish and English) were used as mediums of instruction, transitional bilingual education programs, ESL programs, and mainstream submersion (English-only) programs.

As with the Ramírez study, one of Thomas and Collier's principal research findings was that the most effective programs resulted in achievement gains for bilingual students that were above the level of their monolingual peers in mainstream classes. Another key conclusion was that these gains, in both L1 and L2, were most evident in those programs where the child's L1 was a language of instruction for an extended period of time. In other words, Thomas and Collier (2002) found that the strongest predictor of student achievement in L2 was the amount of formal L1 schooling they experienced. As they state, "the strongest predictor of L2 student achievement is the amount of formal L1 schooling. The more L1 grade-level schooling, the higher L2 achievement" (p. 7). Only One Way and Two Way or Dual Immersion programs – strong bilingual/immersion programs in effect – achieved these results. As Thomas and Collier conclude:

[These] are the only programs we have found to date that assist students to fully reach the 50th percentile in both L1 and L2 in all subjects and to maintain that level of high achievement, or reach even higher levels through the end of schooling. The fewest dropouts come from these programs. (Thomas and Collier, 2002: 7)

Similar also to Ramírez, Yuen, and Ramey, Thomas and Collier found that students in English submersion classes performed far less well than their peers in strong bilingual programs, as well as dropping out of school in greater numbers. Students in transitional bilingual programs demonstrated better academic performance over time, but not to the extent of strong bilingual programs. In both these major large-scale studies, then, length of L1 education turned out to be more influential than any other factor in predicting the educational success of bilingual students, including socioeconomic status.

There are a wide range of other studies, both from the US and from other national contexts that broadly corroborate these findings. Of the wider, book-length, research-based literature, García (2008), Baker (2006) and Baker and Prys Jones (1998) provide magisterial overviews of the field of bilingual/immersion education. Cummins (2000) provides a useful overview of the key research findings with respect to the academic success of students in bilingual programs. Johnson and Swain (1997) and Tollefson and Tsui (2004) provide a wide range of international examples of effective bilingual and immersion education programs. All these contributions add to the

growing research literature confirming the efficacy of strong forms of bilingual/immersion education.

These various research studies and overviews are also complemented by ethnographic studies of effective bilingual/immersion education in various national and regional contexts. In the South American context, Hornberger (1988) explores Quechua language education programs in Peru, and King (2001) examines Quichua programs in the Ecuadorian Andes. In the US, McCarty (2002) provides a fascinating ethnographic account of Navajo language education, while Freeman (1998) discusses her school-based ethnography of Oyster Bilingual School in Washington DC. May's (1994) critical ethnography of Richmond Road School in Auckland, New Zealand is another key school-based ethnographic account of bilingual education. Such ethnographic accounts are important not only for highlighting the core curriculum and pedagogical principles required to teach cultural and linguistic minority students successfully, but also in unpacking the complex, lived experiences of all those involved in them.

Conclusion

This article has outlined the most commonly employed bilingual educational approaches to the teaching of cultural and linguistic minority students. Drawing on extensive research on bilingual education models and programs over the last 30 years, it highlights how strong, additive forms of bilingual education – where the students' bilingualism is both a basis and central aim of the teaching and learning process – are the most effective educational models for bilingual students. Not only do such models achieve and/or maintain the bilingualism and biliteracy of their students, they also provide their students with the greatest opportunity for long-term educational success. Given the ongoing relative lack of achievement of cultural and linguistic minority students in mainstream, often English-only programs, the linguistic and wider educational benefits of additive bilingual programs bear further serious consideration from not only educational researchers, but also crucially, educational policymakers, administrators, and practitioners.

Bibliography

- Andersson, T. and Boyer, M. (1970). *Bilingual Schooling in the United States: Vol. 1*. Austin, TX: Southwest Educational Development Laboratory.
- Baker, C. (2006). *Foundations of Bilingual Education and Bilingualism*, 4th edn. Clevedon: Multilingual Matters.
- Baker, C. and Prys Jones, S. (1998). *Encyclopedia of Bilingualism and Bilingual Education*. Clevedon, UK: Multilingual Matters.

- Cummins, J. (2000). *Language, Power and Pedagogy: Bilingual Children in the Crossfire*. Clevedon: Multilingual Matters.
- Freeman, R. (1998). *Bilingual Education and Social Change*. Clevedon: Multilingual Matters.
- García, O. (2008). *Bilingual Education in the 21st century: A global perspective*. New York: Wiley-Blackwell.
- Hornberger, N. (1988). *Bilingual Education and Language Maintenance: A Southern Peruvian Quechua Case*. Berlin: Mouton.
- Hornberger, N. (1991). Extending enrichment bilingual education: Revisiting typologies and redirecting policy. In García, O. (ed.) *Bilingual Education: Focusschrift in Honor of Joshua A. Fishman*, pp 215–234. Amsterdam: John Benjamins.
- Johnson, R. and Swain, M. (eds.) (1997). *Immersion Education: International Perspectives*. Cambridge: Cambridge Applied Linguistics.
- King, K. (2001). *Language Revitalization Processes and Prospects: Quichua in the Ecuadorian Andes*. Clevedon: Multilingual Matters.
- Lindholm-Leary, K. (2001). *Dual Language Education*. Clevedon: Multilingual Matters.
- May, S. (1994). *Making Multicultural Education Work*. Clevedon: Multilingual Matters.
- May, S. and Hill, R. (2005). Maori-medium education: Current issues and challenges. *International Journal of Bilingual Education and Bilingualism* 8, 377–403.
- McCarty, T. (2002). *A Place to be Navajo: Rough Rock and the Struggle for Self-Determination in Indigenous Schooling*. Mahwah, NJ: Erlbaum.
- Ramírez, J., Yuen, S., and Ramey, D. (1991). *Final Report: Longitudinal Study of Structured English Immersion Strategy, Early-Exit and Late-Exit Transitional Bilingual Education Programs for Language-Minority Children*. San Mateo, CA: Aguirre International.
- Thomas, W. and Collier, V. (2002). *A National Study of School Effectiveness for Language Minority Students' Long-Term Academic Achievement*. Santa Cruz, CA: Center for Research on Education, Diversity and Excellence (CREDE).
- Tollefson, J. and Tsui, A. (eds.) (2004). *Medium of Instruction Policies: Which Agenda? Whose Agenda?*. Mahwah, NJ: Erlbaum.
- Wiley, T. (2001). On defining heritage language education and their speakers. In Peyton, J., Ranard, D., and McGinnis, S. (eds.) *Heritage Languages in America: Preserving a National Resource*, pp 29–36. McHenry, IL: Delta Systems.

Further Reading

- Beardmore, H. B. (ed.) (1993). *European Typologies of Bilingual Education*. Clevedon: Multilingual Matters.
- Cloud, N., Genesee, F., and Hamayan, E. (2000). *Dual Language Instruction: A Handbook for Enriched Education*. Boston, MA: Thomson Heine.
- Cummins, J. (2003). Bilingual education. In Bourne, J. and Reid, E. (eds.) *Language Education: World Yearbook of Education*, pp 3–19. London: Evans.
- Genesee, F., Lindholm-Leary, K., Saunders, B., and Christian, D. (2006). *Educating English Language Learners: A Synthesis of Research Evidence*. New York: Cambridge University Press.
- May, S. (2008). Bilingual/immersion education: What the research tells us. In Cummins, J. and Hornberger, N. (eds.) *Bilingual Education. The Encyclopedia of Language and Education*, 2nd edn, vol. 5, pp 19–34. New York: Springer. <http://www.springerlink.com/content/pj85556182584420/>.
- May, S., Hill, R., and Tiakiwai, S. (2004). *Bilingual/Immersion Education: Indicators of Good Practice*. Wellington, New Zealand: Ministry of Education. http://www.educationcounts.govt.nz/publications/maori_education/5075.
- Thomas, W. and Collier, V. (1997). *School Effectiveness for Language Minority Students*. Washington, DC: National Clearinghouse for Bilingual Education.

Curriculum in Postcolonial Contexts

A Hickling-Hudson, Queensland University of Technology, Brisbane, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Benev(i)olence – A play on the two concepts of benevolent and violent, implying that one can merge with the other. This mixture might characterize concepts such as the white man's burden and the civilizing mission of the West in relation to non-Western cultures.

Cultural interface – As explained by Nakata, it is a blend of traditional indigenous and Western ways of knowing that inform and constitute the daily lives and knowledge systems of indigenous people, causing them to traverse intersecting discourses, take positions, make decisions, and, in the process, remake cultures, "a place of tension that requires constant negotiation" (Nakata, 2004: 27).

Eduscapes – A term coined to reflect the possibility of imagining new global ways of organizing and practicing education which would combine ideas of global flows, landscapes, and the social imagination. The concept is influenced by the idea of the social imaginary articulated by Arjun Appadurai. This imaginary serves as an analytical tool, being composed of five dimensions of global cultural flows: ethnoscapescapes; mediascapescapes; technoscapescapes; finanscapescapes; and ideoscapescapes.

Eurocentrism – The ideology of Western cultural dominance underpinned by beliefs in race hierarchies and by racist assumptions of European sociocultural superiority.

Interpellation – The process which persuades a person to recognize himself or herself as belonging to a particular identity, for example, young, white, male, or liberal, and to identify with the particular ideologies and texts that imbue such identities with emotional appeal. As a powerful strategy of domination, the ruling group surveys and interpolates the inferior subject in a way that fixes his/her identity in relation to the surveyor.

Modernity – A period of the emergence, in Europe, of social forms and processes including the nation-state, the organization of large-scale economic modalities of capitalism and communism advocated as universalist doctrines, human domination over nature, formal schooling, the secularization of values and norms, and the rise of science, rationality, and government conceptualized in a Western manner. Interwoven with modernity was the spread of

European imperialism and the conviction that it was synonymous with civilized ways of life and should displace premodern formations which were deemed static, irrational, barbaric, and thus in need of the civilizing mission of the West.

Orientalizing – The process stemming from the concept articulated by Edward Said in his groundbreaking book *Orientalism*. Said shows how Western scholarship, while establishing the academic discipline of Oriental studies, intertwined the field with prejudicial and distorted racist and religious presumptions. From this perspective, colonialism is not understood as a process of domination for political and economic gain, but as a selfless mission to rescue Orientals from their own cultural and political backwardness and economic mismanagement, to save them from themselves.

Self-orientalising – A process in which non-European peoples are socialized into believing that they themselves are backward, weak, and inferior as characterized in Eurocentric portrayals.

Postcolonial theory provides tools with which we can interpret the powerful and lingering effects of European and (later) North American global dominance. For nearly 500 years, colonialism and imperialism gradually extended economic, military, political, and sociocultural power over most of the globe. Using postcolonial perspectives to interrogate education illuminates relations of power in the curriculum that are deeply influenced by this spread of European society and norms. The insights gained help to address questions of why so many curriculum practices and educational programs still appear far away from consensus on the goals, contours, and implementation of equity. A postcolonial analysis problematizes these issues in a unique way, identifying the contradictions embedded in neocolonial norms, working out strategies of change, and identifying the complexities and contestations which accompany change. This article provides an introduction to how various scholars apply postcolonial theory to understand the curriculum. It starts with definitions and overviews and subsequently applies postcolonial perspectives to disrupt conventional views of the various aspects of the curriculum.

The prefix post in postcolonialism is both temporal and epistemological. It does not denote a simplistic after-colonialism in terms of describing historical events but,

rather, is interested in the complex and contradictory experiences underlying the processes of colonialism and decolonization or, in the case of countries which were not overtly colonized, such as China and Thailand, what it means to live with the global consequences of the European empires and their successors. It is also interested in the effects of empire on those in the centers of power – Europe and its diasporas. The postcolonial theory explores these contexts by means of an epistemological shift that not only utilizes, but also goes beyond, the post of postmodernity, seeking equitable alternatives to overcome the sociopolitical and economic problems of the exploitative development paradigm that colonialism so deeply entrenched. Working at the nexus of identifying commonalities as well as distinctive situatedness in education, it probes issues of domination, resistance, subversion, contradiction, ambiguity, and change in a much more complex way than is achieved by dependency theory (see Hickling-Hudson *et al.*, 2004; Tikly, 2004; McLaughlin and Hickling-Hudson, 2005).

Historically and culturally grounded ideas from a range of scholars helped to formulate major strands of postcolonial thinking. (Widely cited as the inspiration for the postcolonial theory are key ideas of scholars such as Fanon (1967, 1968), Said (1978), Spivak (1988), Bhabha (1994), and Hall (1996).) One strand addresses colonial discourse, seeing it as a complex of statements, texts, and practices within which society came to be known, organized, and reproduced by both the colonizer and the colonized. From this flows the concept of the knowledge/power nexus between colonization as a system of power and exploitation, and colonization as a system of knowledge and representation which deeply affects the identity formation of the colonizer and the colonized. The conflicts, ambiguities, and hybridities of identity in postcolonial contexts are of concern to many scholars. Since the postcolonial theory identifies the colonial roots of many of the problems of our globalizing era and deconstructs the aftermath, it throws light on how the curriculum can be problematized and challenged within this historicized context.

Systemic Educational Inequity and Its Challengers

The postcolonial theory has not been adequately applied to explore how global educational inequity might be challenged and overcome at local, national, and global levels. Various impoverished countries have improved education with new approaches, including community involvement, adult basic education, child-centered pedagogies, health reform, and other strategies which have provided some opportunities for the disadvantaged. However, for most learners across the globe, education is

dominated by the contradictory characteristics of Western modernity with inequities relating to social class, ethnicity, gender, religion, academic versus vocational knowledge, and the rural–urban divide, built into a system that proclaims equal opportunity.

An overview of some of the features of global educational inequity is provided by Hickling-Hudson *et al.* (2006). They argue that in a planet where vast proportions of people are impoverished and marginalized in terms of social justice, it would not be possible, without fundamental socioeducational change, to overcome the global division of labor whereby the large, poorly educated populations in the new nations still provide cheap and low-skilled labor for the globalized industries of the wealthy.

A postcolonial analysis of the curriculum within a system of global inequity would discuss the complexities of reforming it and the significance of this for global change. This kind of approach is taken by Borg and Mayo (2002) in a thought-provoking analysis of the educational system of Malta. In a postcolonial manner, they explore how concepts elaborated by the great Brazilian educator Paulo Freire apply as much to understanding the features of society and education in Malta as they do to those of Brazil. They discuss colonialism as a parasitic relationship between the colonizer and the colonized, religion as a site of struggle, cultural domination from without and within, racism as a dialectic rather than a binary relationship, and the contours of banking or transmission education in comparison to those of praxis as transformative reflection and action for educational change, citing the struggle over the new high school syllabus as an example.

Decades ago, President Julius Nyerere of Tanzania critiqued the education system implanted by the British in Africa as inculcating elitism, abstract, bookish knowledge divorced from local needs, a disregard of nonprint knowledge, and as a parasitic relationship to the community (Nyerere, 1968). In most new nations emerging from colonialism, this system of education expanded rather than changed over the second half of the twentieth century: there was very little fundamental redesign (Bishop, 1989). Since its revolution of 1959, Cuba is one of the few nations which have deliberately tackled the dysfunctionality of neocolonial education. The revolution's radical educational changes have significantly advanced educational equity and effectiveness, and have been recognized by the United Nations Educational, Scientific and Cultural Organization (UNESCO) for bringing about outstanding educational outcomes for students compared to other countries in Latin America (see Carnoy and Marshall, 2005). A postcolonial analysis would do more than document how this has been done. It would also explore the epistemologies underpinning the revolution's education system, and would wrestle with questions of how far

it has been able to supersede the constraining knowledges of Eurocentric modernity.

Assessment and the European Connection

The perpetuation of particular European approaches to organizing school and university examination systems is entrenched in new nations. This is bound up with questions of sociocultural, national identity, as is pointed out by Koh (2004) in his analysis of Singapore's education system. As one of the more economically successful new nations, Singapore is using its wealth to improve its entire education system to a world-class level. Identifying dilemmas from a postcolonial perspective points to contradictions in this process.

Koh (2004) argues that Singapore's sense of nationhood is contradictory; combining a nostalgia constructed out of a romanticized British colonial history with a strident promotion of Confucian/Chinese values in a process of self-Orientalizing and othering Western culture. Singapore's examination system is one structure illustrating this kind of contradiction and ambivalence. It still uses the Cambridge General Certificate of Education (GCE) syllabus and school-leaving examinations, organized by the UK's Cambridge Examination Syndicate, which charges any country taking the GCE high fees for administering the exam. While relying on the Western discourses and knowledge embodied in these examinations to determine scholastic pathways, the government tries to compensate for the risk of Westernization by means of curriculum intervention that promotes Asian values and a distinct Asian cultural identity (Koh, 2004: 163).

As part of the decolonization process, some countries have moved away from Cambridge examinations to design their own system. The postcolonial gaze seeks to ascertain to what extent colonial legacies live on in the newly-designed systems, and with what impact. For example, the new Caribbean Council examination system, independent of Britain since the 1970s, has neocolonial elements that echo the traditional British GCE in several ways, perpetuating a range of sociocultural tensions and dilemmas. One dilemma is the continuation of an age-graded, lockstepped approach to organizing the curriculum and an emphasis on print technologies, both of which are becoming increasingly archaic in an era of information technology. Another major dilemma is that the Caribbean Examinations Council (CXC) examinations have continued to be the examinations of the Caribbean elite. Adequate preparation for these exams is best provided by preferentially resourced schools modeled after the British grammar school, and distinct from the lower-status schools of the majority modeled after the old vocational secondary or secondary modern school,

and the resource-poor all-age school which takes young people up to a senior primary level at age 15.

The status divisions in new nations such as those of the decolonizing Caribbean are thus symbolized by and entrenched in their examination systems. Difficult questions remain: How are inequitable exam systems to be overcome? Which discourse is to drive the design of alternative systems? These questions demand a postcolonial lens, for they cannot be answered without reference to a historicized understanding of deep sociocultural attachment to neocolonial examinations, as well as of their continued global economic power.

The Curriculum and Cultural Identity

Colonial and neocolonial education afforded strong mechanisms of interpellation – calling people through their experience of targeted curriculum and pedagogy and their dreams of educational achievement, into identifying with the culture of the colonial metropole. The educated saw themselves as inheritors of a superior European culture. Tiffin (2004), in her analysis of what she terms the benev(i)olence of imperial education, argues that the curriculum of empire exerted a powerful emotional as well as intellectual impact on its pupils. In particular, learning in the European language of education and absorbing parts of literary texts and memory gems by heart became part of the emotional core of one's nature, the goal of these curriculum techniques being to produce people who were, in the language of Macaulay's (in)famous *Education Minute*, "Indian in blood and colour, but English in taste, in opinions, in morals and in intellect" (Tiffin, 2004: 148).

Tiffin (2004) also discusses the ambivalence of colonial education in its interpellation of people into European epistemes. This facilitated colonial rule by the establishment of a hegemony in which European culture was promoted as unproblematically ideal, while local culture was repressed, devalued, and/or distorted. A significant part of the colonial curriculum was the inculcation of racist stereotypes, myths, and the inferiority complex of the cultural cringe, which defers to the intellectual and cultural authority of the West, as Koh (2004) pointed out with his Singapore example. Yet, the ambivalence of benev(i)olence was that postcolonial engagement, with its contradictions, produced, in new nations and among indigenous peoples, outstanding intellectuals who have forged brilliant new ways of thinking and problem solving as well as "some of the most spectacularly significant creative writing of the twentieth century" (Tiffin, 2004: 153). Educated locals in the colonies, trained in Western education systems, often formed a colonial intelligentsia dedicated to independence and anticolonial movements

which ultimately displaced the European supremacy (Willinsky, 1998: 109).

The intellectual violence of imposing a culturally distorted Eurocentric curriculum on students is still, in the twenty-first century, a strong part of the tradition of Western education – the benev(i)olent gift of imperialism. The contradictions continue, in that the effects of this curriculum can, at the same time, position the non-European learner in new nations to be a player in the globally dominant field of Eurocentric knowledge, assault the learner's cultural identity with a sense of otherness, promote misplaced shame and self-doubt, and beckon the learner into the delights of pursuing the educated imagination based on the Western canon, expertise, and the pleasures of globally respected learning.

Whose Language? Whose Knowledge?

At the heart of a postcolonial understanding of the curriculum is recognition of the central importance of the interlinking of language and knowledge. The language of instruction shapes how the learner absorbs knowledge and culture, with powerful material effects. Thiong'o (1986), in his lyrical and penetrating work *Decolonising the Mind: The Politics of Language in African Literature*, uncovers the significance of immersing children in the colonial language for nearly the whole of their education except for the very earliest years. This enabled colonialism to dominate the mental universe of its subjects. Ngugi stresses that this was one of the most powerful means by which the West secured (and still secures) the domination of resources and wealth. "Behind the cannon was the new school. The new school had the nature of both the cannon and the magnet. . . The cannon forces the body and the school fascinates the soul" (Thiong'o, 1986: 9). The resistance of some of the colonized to retrieving their native tongue, teaching in it, and writing in it was an indication of how far imperialism had distorted the view of African realities, making the abnormal seem normal, and vice versa. For example: "Africa actually enriches Europe; but Africa is made to believe that it needs Europe to rescue it from poverty. . . is made to feel grateful for aid" (Thiong'o, 1986: 28).

From a postcolonial view, foreign language imposition can have devastatingly negative effects on learning. Brock-Utne (2005, 2006) discusses the continuing ambiguous attachment of many Africans to teaching and studying in the colonial language, and shows how Western donors, through financial aid, exert an enormous impact on the curriculum. The regressive policies of the privatization and cost sharing imposed on African education by international agencies such as the World Bank and the International Monetary Fund (IMF) often include the insistence that the colonial language must continue as the language

of instruction, rather than as an additional foreign language. Owing to strategies such as this, many Africans continue to equate education with Eurocentric learning. The Western description of African countries as Anglophone, Francophone, and Lusophone has never been effectively challenged, there being no united push to reinstate in education the Afrophone character of the continent (Brock-Utne, 2005, 2006). It can be added that there is apparently no sense of irony in the African adoption of the term renaissance to symbolize the goal of an African cultural renewal, or among Caribbean people who continue to call themselves West Indian, thus entrenching the geographical errors of sixteenth-century Europeans.

Discarding Colonially Tainted Understandings

For students in Europe and its diaspora, learning is still often dominated by the Eurocentric bias nourished by ideologies of empire. As Coulby and Jones (1996) point out, school curricula frequently construct and inculcate, in the young, a stereotype of a European citizen who is tolerant, rational, and a model of democracy, ignoring the prevalence of xenophobia, narrow nationalism, and racism, which means that for many citizens of the European Union, life is far removed from these virtues. If a postmodern point of view emphasizes that constructed boundaries and plural identities are the reality for most Europeans, a postcolonial perspective would go further, uncovering for learners the central role of the aftermath of empire in their current lives.

Perhaps the most systematic analysis of the Eurocentric curriculum is that of Willinsky (1998) in his path-breaking book *Learning to Divide the World: Education at Empire's End*. Willinsky frames his analysis by problematizing the educational mission: How are we to educate the young about the seemingly naturalistic and inexorable divisions in the world between wealth, poverty, race, gender, ethnicity, culture, and learning? His answer is that teachers need to help students learn how to turn the tables on education by showing how the curriculum has schooled us in understanding difference from the viewpoint of the European conquest. This insight would be an essential step in discarding colonially tainted assumptions about why the colonized should be "subservient to a born-to-rule civilization" (Willinsky, 1998: 1, 2 and 4).

Willinsky advocates not just that we teach students about the intellectual legacy of imperialism, but also that we teach about the conflicts over this legacy, the contest of ideas that continually challenged it. Such teaching is difficult enough at the university level, where it is highly contested, but it is almost unheard of at the level of the public school with its entrenched tradition of excluding dissenting voices. His book provides groundwork for

launching the preparation of educators in carrying out this task. The book is perhaps the only detailed study of how the subjects that became the major curriculum areas of history, geography, science, language, and literature were shaped by empire, how they continue to be influenced by the colonial imagination in today's classrooms, and how educators and textbook writers are, at last, starting to challenge these imperial formations.

Race, Representation, and Power

A postcolonial perspective links racism to its colonial roots and deconstructs the many ways in which it operates in order to challenge it. An increasing number of scholars of education have applied this perspective to analyzing and challenging the continuation of racism in school curricula of the late twentieth and twenty-first centuries. For example, Hickling-Hudson (2003, 2004a) discuss the racism shown in the Eurocentric curriculum provided in indigenous and multicultural primary schools in Australia and the USA, and describe the challenges that indigenous communities have mounted to this curriculum. They argue that "A Eurocentric curriculum teaches no critical view of culture which would enable students to see that all cultures have strengths and weaknesses and that they operate within particular epistemologies. Lacking this critical approach, it unabashedly asserts the superiority of European culture, turning non-European cultures into the inferior 'Other' [...] It disrespects and devalues other cultures and other learning styles by making them invisible or distorting them. This absence or distortion is not a simple oversight – it is an example of institutional racism. Many teachers and teacher educators are intensely socialized into its norms. It can have a devastating effect on non-European peoples who are required to submit to this episteme through compulsory schooling" (Hickling-Hudson and Ahlquist, 2004b: 41).

Hickling-Hudson (2006) investigates further complexities of educational decolonization in the differently racialized settings of Australia and the Caribbean, showing how colonial legacies of race and social power continue to shape the inequitable ways in which curriculum is provided for different sociocultural groups. Pedagogical complexities face white Australians when their Anglocentric curriculum is confronted with indigenous ways of knowing, Jamaican teachers when they have to struggle with how poverty issues distort schooling, Grenadian educators in their fashioning of revolutionary experiments, and aboriginal elders when they demonstrate the necessity of holistic education starting from very different premises to those of Western systems. The analysis illustrates how educators can benefit from using postcolonial and comparative thinking to integrate their understanding of cultural complexity and equity in

educational change. Addressing the legacies of neocolonial constructions of race and ethnicity in the Australian context, Hickling-Hudson (2005) discusses a project in which student teachers begin to apply postcolonial perspectives to analyze the implications of these racial discourses for school learning.

Feminism and Education: Postcolonial Views

Deep contradictions were embedded in women's struggles for feminism in decolonizing nations. As Young (2003) points out, the winning of national sovereignty was, for women, not the end. Women's sociopolitical objectives for equity constituted another type of liberation struggle, given that national independence almost always involved a transfer of power to local elites who continued many of the repressions of colonial, and sometimes precolonial, systems against women and the poor. In contexts where the interests of women are routinely othered and subordinated, their participation in education is often much lower than that of men, as Fox (1999) shows in her discussion of educational issues of women in Papua New Guinea. However, even where female participation and achievements in education exceed those of males, such as in some Caribbean countries, this has not enabled women to break the glass ceiling which maintains a situation of male dominance in most areas of public life, pointing to the conclusion that successful female participation in neocolonial education systems will not necessarily move a society toward gender equity.

The experience of formal Eurocentric education imposed on traditional indigenous communities is well known as a process that devastates and ruins traditional cultures and environments with its ideology of individualist economic maximization leading to a putative progress. How this schooling particularly devalues the position of women is powerfully analyzed by Aikman (1999) in her study of the indigenous Arakmbut people of the southeastern Peruvian Amazon. Aikman shows how such teaching "not only contributes to the destruction of the Amazon environment but erodes and undermines women's position as guardians and custodians of biodiversity" (p. 78), leading to new male/female relations that further devalue women's knowledge and skills.

What Would a Decolonized Curriculum Look Like?

Following Said (1978), the postcolonial theory analyzes the self-serving construction of knowledge under imperialism. It is interested in the effects of this kind of analysis during the process of decolonization, looking at how it can

both catalyze a subversion of the old ways and stumble in face of barriers to genuine transformation. Postcolonial perspectives on the curriculum would necessarily move to envisioning curriculum renewal. They would ask the question of all curriculum areas asked by South African scholar Crush (1994) with respect to geography: “what would a de-colonized, de-whitened, post-colonial geography look like?”

Crush points to the work of the subaltern studies group in India as suggesting a way forward in its complex project to seek the recovery and foregrounding of the authentic voice of those who were silenced and marginalized by systems of oppression. He observes that geography in South Africa traditionally exhibited the cultural cringe of defining itself by, and seeking validation from, the metropole, and asks: “Does the decolonization of the discipline require a rupture with the knowledge industry of the western heartlands of geographical enterprise, or is there room for a productive, post-colonial interface?” (p. 338–339). He answers by outlining how South African geographical scholars seek to delink the discipline from any slavish adherence to Euro-American thinking, develop an indigenous and eclectic synthesis, and “shatter the boundaries between campus and community” (Crush, 1994: 342).

The challenge to Eurocentrism in the curriculum is full of its own contradictions and pitfalls. The authors of new history texts, for example, can easily fall into the contradictory position of pointing out the Eurocentric biases of historians and geographers, yet themselves showing these biases, as Willinsky (1998: 129–130) illustrates. A contrast to this kind of curriculum confusion is the blunt approach taken by Gill and Levidow (1987), educators whose edited textbook promotes an antiracist science curriculum, drawing on the work of a community of science teachers actively working in the UK for social change. The text assumes responsibility for teaching science not abstractly, but within the context of its role in the global economic and ideological system. For example, students’ attention is drawn toward the political causes and consequences of the putative concept of race and its use in science, and toward the point that “Science teaching masks the real political and economic priorities of science; hides its appropriation of non-Western scientific traditions; [and] often attributes people’s subordination to nature. . . . rather than to the way science and nature itself have been subordinated to political priorities” (Willinsky, 1998: 184). In the field of literature, Bean (2004) explores the postcolonial theory and multicultural literature as potential frames of reference for challenging the narrow vision and hegemonic structure of the traditional literary canon. He shows how issues of institutional racism, stereotyping, and socially deterministic views of indigenous groups are confronted in recent fiction for young adults, and discusses how strategies such as body

biographies, reading against the grain, and discussion questions can challenge the preconceptions of adolescent students.

The Cultural Interface

What Nakata (2004) calls the cultural interface between Western and indigenous knowledge systems would be the logical site of curriculum change that protects the interests of indigenous and other marginalized peoples. In this interface, the change directed by indigenous people should be informed by meta-knowledge: knowledge about knowledges. It must be “change that incorporates into our own knowledge all the ongoing developments brought about by the convergence of other systems of understanding, so that our own corpus of knowledge. . . keeps expanding. . .” (Nakata, 2004: 28).

Nakata discusses the web as a site of new opportunities for indigenous peoples to control the presentation of knowledge and, thus, take charge of their own scholarly agendas. He is part of a working group of indigenous academics across six universities striving to develop an online degree in Australian indigenous studies. Alternative theoretical platforms from indigenous standpoints will be encouraged in a process that engages with both Western and indigenous knowledge domains, thus facilitating not only the continued development of indigenous knowledge traditions, but also the opportunity for nonindigenous people to interrogate their own systems of thought (see Nakata, 2004: 30–33). Just ten items selected from the list of 25 put forward by Tuhiwai Smith (1999) as concepts and projects important to indigenous research suggest the depth of the challenge to conventional Western assumptions about research: “celebrating survival, testimonies, revitalising, connecting, negotiating, gendering, protecting, creating, representing, sharing.”

For curriculum renewal at the cultural interface, then, it is essential to offer students an introduction to the study of knowledges: an exploration of indigenous and Western knowledges and philosophies, as well as of Eurocentric views of the other and the use of these to justify domination and exploitation. To work toward a new, shared, postcolonial, and global space, it is essential for all students to understand how they have been socialized into Eurocentric ways of thinking, study the consequences, appreciate formerly suppressed knowledges, and to be part of forging alternatives.

Conceptualizing Curriculum in Post-Postcolonial Contexts

We end with provocative questions by Luke (2005) that attempt to conceptualize post-postcolonial contexts and

their implications for curriculum. Luke asks: “What would it mean to speak of a post-postcolonial moment in education theory and practice? ... Where do the systems of Asia, the Americas, Africa and the South Pacific go from here, in the face of rapid forces of cultural and economic globalisation, shifting and recentring of capital, and emergent new blends of language and discourse? What kind of ‘eduscapes’ can and should be constructed in response to these conditions?”

Luke’s (2005: xvii) sketches of possible eduscapes which might meet new challenges of globalization include those which would bring together, for example, identity politics with basic technical/vocational skills training, bilingual education with scientific training for new economies, and training in local school leadership with preparation for radical pedagogies. Although these scenarios are not new and do not answer the epistemological question posed above of what a decolonized curriculum might look like, they draw attention to the necessity of imagining educational structures, approaches, and content which would tackle the unprecedented crises of our time. As long as neocolonial attachments, patterns, and desires continue to constitute elements of these crises, educators will arguably need to do the postcolonial work of imagining “an education system that does not restrict access to schooling: a system that does not discriminate, stratify or exclude groups on the basis of class, ethnicity, gender, sexuality, disability or location ... a postnational system that reaches beyond the boundaries and identities of nation states... that disrupts preconceptions about knowledges and power relations, and about its own ability to establish final and forever appropriate structures and solutions” (Hickling-Hudson *et al.*, 2004: 1).

Bibliography

- Aikman, S. (1999). Schooling and development: Eroding Amazon women’s knowledge and diversity. In Heward, C. and Bunwaree, S. (eds.) *Gender, Education and Development: Beyond Access to Empowerment*, pp 65–82. London: Zed Books.
- Bean, T. (2004). The role of multicultural literature as a counterforce to the literary canon. In Hickling-Hudson, A., Matthews, J., and Woods, A. (eds.) *Disrupting Preconceptions: Postcolonialism and Education*, pp 57–75. Flaxton: QLD: Post Pressed.
- Bishop, G. (1989). *Alternative Strategies for Education*. New York: St Martin’s Press.
- Borg, C. and Mayo, P. (2002). Making sense of a postcolonial context: A Freirean perspective. *Journal of Postcolonial Education* **1**, 101–125.
- Brock-Utne, B. (2005). Language-in-education policies and practices in Africa with a special focus on Tanzania and South Africa – insights from research in progress. In Lin, A. and Martin, P. (eds.) *Decolonisation and Globalisation: Language-in-Education Policy and Practice*, pp 173–193. Clevedon, UK: Multilingual Matters.
- Brock-Utne, B. (2006). Development cooperation in the field of education – between neo-liberal economics and alternative educational models. *Journal für Entwicklungspolitik (Journal for Development Policy)* **22**, 27–51.
- Camoy, M. and Marshall, J. (2005). Cuba’s academic performance in comparative perspective. *Comparative Education Review* **49**, 230–261.
- Coulby, D. and Jones, C. (1996). Post-modernity, education and European identities. *Comparative Education* **32**, 171–184.
- Crush, J. (1994). Post-colonialism, de-colonization and geography. In Godlewska, A. and Smith, N. (eds.) *Geography and Empire*, pp 333–350. Oxford: Blackwell.
- Fanon, F. (1967). *Black Skin, White Masks*. New York: Grove Press.
- Fanon, F. (1968). *The Wretched of the Earth*. New York: Grove Press.
- Fox, C. (1999). Girls and women in education and training in Papua New Guinea. In Heward, C. and Bunwaree, S. (eds.) *Gender, Education and Development: Beyond Access to Empowerment*, pp 33–48. London: Zed Books.
- Gill, D. and Levidow, L. (eds.) (1987). *Anti-Racist Science Teaching*. London: Free Association Books.
- Hall, S. (1996). When was the ‘post-colonial’? Thinking at the limit. In Chambers, I. and Curti, L. (eds.) *The Post-Colonial Question: Common Skies, Divided Horizons*, pp 242–260. London: Routledge.
- Hickling-Hudson, A. (2005). “White”, “ethnic” and “indigenous”: Pre-service teachers reflect on discourses of ethnicity in Australian culture. *Policy Futures in Education* **3**, 340–358.
- Hickling-Hudson, A. (2006). Integrating cultural complexity, postcolonial perspectives, and educational change: Challenges for comparative educators. *International Review of Education* **52**, 201–218.
- Hickling-Hudson, A. and Ahlquist, R. (2003). Contesting the curriculum in the schooling of indigenous children in Australia and the USA: From Eurocentrism to culturally powerful pedagogies. *Comparative Education Review* **47**(1), 64–89.
- Hickling-Hudson, A. and Ahlquist, R. (2004a). Teachers as “two-year tourists” in an Australian state school for aboriginal children: Dilemmas of curriculum, agency and teacher preparation. *Journal of Postcolonial Education* **3**(1), 67–88.
- Hickling-Hudson, A. and Ahlquist, R. (2004b). The challenge to deculturalisation: Discourses of ethnicity in the schooling of indigenous children in Australia and the USA. In Hickling-Hudson, A., Matthews, J., and Woods, A. (eds.) *Disrupting Preconceptions: Postcolonialism and Education*, pp 39–56. Brisbane, QLD: Post Pressed Publishers.
- Hickling-Hudson, A., Gonzalez, J., and Sabina, E. (2006). Education in newly independent countries: Problematic models and the significance of the Cuban alternative. *Journal für Entwicklungspolitik (Journal for Development Policy)* **22**, 96–125.
- Hickling-Hudson, A., Matthews, J., and Woods, A. (eds.) (2004). In Education, postcolonialism and disruptions, *Disrupting Preconceptions: Postcolonialism and Education*, pp 1–18. Flaxton, QLD: Post Pressed.
- Koh, A. (2004). The postcolonial education system: Postcolonial encounter of the Singaporean kind. In Hickling-Hudson, A., Matthews, J., and Woods, A. (eds.) *Disrupting Preconceptions: Postcolonialism and Education*, pp 155–174. Flaxton, QLD: Post Pressed.
- Luke, A. (2005). On possibilities of a post-postcolonial language education. In Lin, A. and Martin, P. (eds.) *Decolonisation and Globalisation: Language-in-Education Policy and Practice*, pp 13–20. Clevedon, UK: Multilingual Matters.
- McLaughlin, J. M. and Hickling-Hudson, A. (2005). Beyond dependency theory: A postcolonial perspective on educating Papua New Guinea students in Australian high schools. *Asia Pacific Journal of Education* **25**, 193–208.
- Nakata, M. (2004). Indigenous knowledge and the cultural interface: Underlying issues at the intersection of knowledge and information systems. In Hickling-Hudson, A., Matthews, J., and Woods, A. (eds.) *Disrupting Preconceptions: Postcolonialism and Education*, pp 1–16. Flaxton, QLD: Post Pressed.
- Nyerere, J. (1968). *Ujamaa: Essays on Socialism*, pp 53–58. Dar es Salaam: Oxford University Press.
- Said, E. (1978). *Orientalism: Western Conceptions of the Orient*. New York: Pantheon.
- Thiong’o, N. (1986). *Decolonising the Mind: The Politics of Language in African Literature*. Portsmouth, UK: Heinemann.
- Tiffin, H. (2004). The benev(i)olence of imperial education. In Hickling-Hudson, A., Matthews, J., and Woods, A. (eds.) *Disrupting*

- Preconceptions: Postcolonialism and Education*, pp 143–154. Flaxton, QLD: Post Pressed.
- Tikly, L. (2004). Education and the new imperialism. *Comparative Education* **40**, 173–198.
- Tuhiwai Smith, L. (1999). *Decolonising Methodologies: Research and Indigenous Peoples*. London: Zed Books.
- Willinsky, J. (1998). *Learning to Divide the World: Education at Empire's End*. Minneapolis, MN: University of Minnesota Press.
- Young, R. (2003). *Postcolonialism: A Very Short Introduction*. Oxford: Oxford University Press.

Further Reading

- Ashcroft, B., Griffiths, G., and Tiffin, H. (1998). *Key Concepts in Post-Colonial Studies*. London: Routledge.
- Bhabha, H. (2004). *The Location of Culture*. London: Routledge.
- Bray, M. (1994). Education and the vestiges of colonialism: New paradigms for the remnants of empire. *Comparative Education* **24**, 333–348.
- Bray, M. and Steward, L. (eds.) (1998). *Examination Systems in Small States: Comparative Perspectives on Policies, Models and Operations*. London: Commonwealth Secretariat.
- Breidlid, A. (2007). Education in Cuba – an alternative educational discourse: Lessons to be learned? *Compare* **37**, 617–634.
- Harber, C. (2004). *Schooling as Violence: How Schools Harm Pupils and Societies*. Oxford: RoutledgeFalmer.
- Hickling-Hudson, A. (2003). Multicultural education and the postcolonial turn. *Policy Futures in Education* **1**, 381–401.
- Hickling-Hudson, A. (2004). Caribbean “knowledge societies”: Dismantling neo-colonial barriers in the age of globalisation. *Compare* **34**, 293–300.
- Hickling-Hudson, A. (2007). Beyond schooling: The role of adult and community education in postcolonial change. In Arno, R. and Torres, C. (eds.) *Comparative Education: The Dialectic of the Global and the Local*, pp 197–216. Lanham, MD: Rowman and Littlefield.
- Ninnes, P. and Burnett, G. (2004). Postcolonial theory in and for comparative education. In Ninnes, P. and Mehta, S. (eds.) *Re-Imagining Comparative Education: Postfoundational Ideas and Applications for Critical Times*, pp 181–200. New York: RoutledgeFalmer.
- Papastergiadis, N. (1997). Tracing hybridity in theory. In Werbner, P. and Modood, T. (eds.) *Debating Cultural Hybridity: Multicultural Identities and the Politics of Anti-Racism*, pp 257–281. London: Zed Books.
- Pennycook, A. (1998). *English and the Discourses of Colonialism*. London: Routledge.
- Philippson, R. (1992). *Linguistic Imperialism*. Oxford: Oxford University Press.
- Quayson, A. (2000). *Postcolonialism: Theory, Practice or Process?* Cambridge: Polity Press.
- Rizvi, F., Lingard, R., and Lavia, J. (2006). Postcolonialism and education: Negotiating a contested terrain. *Pedagogy, Culture and Society* **14**, 249–262.
- Said, E. (1993). *Culture and Imperialism*. London: Chatto and Windus.
- Spivak, G. (1988). *In Other Worlds: Essays in Cultural Politics*. New York: Routledge.
- Stasiulis, D. and Yuval-Davis, N. (1995). *Unsettling Settler Societies. Articulations of Gender, Race, Ethnicity and Class*. London: Sage.
- Tikly, L. (2004). Globalisation and education in Sub-Saharan Africa: A postcolonial analysis. In Hickling-Hudson, A., Matthews, J., and Woods, A. (eds.) *Disrupting Preconceptions: Postcolonialism and Education*, pp 109–126. Flaxton, QLD: Post Pressed.
- Young, R. (2001). *Postcolonialism: An Historical Introduction*. Oxford: Blackwell.

Curriculum, Economic and Cultural Development

A Rampal, Delhi University, Delhi, India

© 2010 Elsevier Ltd. All rights reserved.

The Convention on the Rights of the Child and the aims of education it recommends require that the curriculum must be of direct relevance to the child's social, cultural, environmental, and economic context. This may seem ubiquitously unproblematic and, yet, encompasses some of the most contentious issues and challenges of contemporary curriculum development. In low-income countries with deeply differentiated and widely pluralistic societies, still unable to provide elementary education to large sections of children, struggling with local issues of human development – of sustenance and survival – while also aspiring for a place in the global market, notions of curricular relevance and context are becoming increasingly contested.

This article begins by looking at models of curriculum development within different educational traditions, with a focus on the critical indigenous approach. It locates the dialectical search for a national curriculum in India, within its anticolonial struggle, where contested notions of modernization and Indianization were etched onto its developmental track. It relates these historical tensions to contemporary struggles for an equalizing curriculum – so that national does not imply a common centralized curriculum; modern is not constrained to serve only the knowledge economies, by alienating most children from their lives and knowledges of production; and indigenization reckons with plural and multicultural identities, not appropriated by a dominant and fundamental religiosity.

Curriculum in Different Educational Traditions

The 'Education for all' (EFA) report (UNESCO, 2005) looks at the curriculum in different traditions, including the aims, processes, and outcomes of education. The humanist tradition rejects "standardized, externally defined or controlled curricula, [which are] seen as undermining the possibilities for learners to construct their own meanings and for educational programmes to remain responsive to individual learners' circumstances and needs" (p. 32). The social constructivist approach builds on the humanist traditions to lay emphasis on learning as a social, rather than an individual process. On the contrary, the behaviorist tradition, which continues to wield significant influence on the practice of curriculum development across the world, more so in the present neoliberal era, endorses standardized, externally

controlled curricula, based on rigidly prescribed and nationally tested competences often defined independently of the learner.

The critical approach to redress social inequality and injustice "is emancipatory in the sense that it lets students find their own voices, frees them from externally defined needs and helps them to explore alternative ways of thinking that may have been buried under dominant norms" (UNESCO, 2005: 34). It stands for transformative education designed through participatory processes to encourage critical analysis of social power relations and of ways in which formal knowledge is produced and transmitted. By incorporating popular culture and situated and everyday knowledge, the critical curriculum further attempts to engage with the cultural politics of identity. However, there is a danger of tokenism where multicultural curricula may profess knowledge as socially constructed, but, with policies pushing for standardized testing, may end up with conventionally packaged information, and also ignore indigenous knowledge traditions that have nurtured ecologically sustainable cultures as being backward or irrational.

According to the EFA report (UNESCO, 2005) the indigenous tradition emerged out of alternative educational ideas rooted in the realities of low-income countries that challenged imported knowledge, images, values, and beliefs. "Prominent examples include the approaches of Mahatma Gandhi and Julius Nyerere, both of whom proposed new and alternative education systems with culturally relevant emphases on self-reliance, equity and rural employment." The indigenous tradition:

- reasserts the importance of education's relevance to the sociocultural circumstances of the nation and learner;
- assures that relevance implies local design of curriculum content, pedagogies, and assessment, using learners' rich sources of prior knowledge; and
- moves beyond the boundaries of the classroom/school through nonformal and lifelong learning activities (UNESCO, 2005: 34).

The Gandhian model of basic education (Hindustani Talimi Sangh, 1938) called for education for life, through life, and used a productive craft – weaving, carpentry, agriculture, pottery, etc. – as the medium of interdisciplinary hands-on learning in the primary curriculum, with the mother tongue as the medium of instruction. At the upper primary stage, the distinction between traditional academic and vocational streams was sought to be

reduced through flexible tracks, where science, home science, or agricultural science could be placed at par. This radical move matched the agenda of the anticolonial freedom struggle – for inclusive schools independent of government funding to also interrogate the traditional caste system that stigmatized the low castes and their vocations. Basic education schools continued to run in the 1950s, after India became independent, but did not receive sustained support from the government and the elites aspiring for white-collar employment through modern education.

The National within the Local–Global Dilemma

Analyses of indigenous curricula that forge new local trajectories against the grain of dominant global tracks reveal a legacy of philosophical and dialectical thought in specific sociohistorical contexts. Gandhi and Tagore were major Indian thinkers whose respect for each other refined their differences and also enriched the decolonizing discourse on wide-ranging issues such as development, nationalism, education, and science and its domination. Their letters (Bhattacharya, 1997) show passionate visions of national education located within the freedom struggle led by Gandhi through the movement of noncooperation and support to *swadeshi* (the indigenous) with the boycott of British goods, including the schools run by them. Tagore ran his own indigenous school, but felt that the “struggle to alienate our heart and mind from those of the West is an attempt at spiritual suicide. . . [as indeed] for a long time we have been out of touch with our own culture” (ibid. p. 62). Gandhi maintained: “It is unbearable for me that the vernaculars should be crushed and starved as they have been. I hope I am as great a believer in the free air as the great Poet. . . I want the cultures of all the lands to be blown about my house as freely as possible. But I refuse to be blown off my feet by any” (ibid. p. 64). Both developed distinct models of indigenous curricula, while Tagore questioned the utilitarian centrality of manual work at the cost of art and esthetics, and wondered if a differentiated education “should be doled out in insufficient rations to the poor” (ibid. p. 34), who thus get assigned to a limited place and vocation.

Their historic debate reconnects with several contemporary curricular discourses. Pinar (2004) sees the curriculum as an ongoing complicated conversation – and a social and subjective reconstruction, that allows an analysis of “one’s experience of the past and fantasies of the future in order to understand more fully, with more complexity and subtlety, one’s submergence in the present” (p. 4). The modernization–indigenization dilemma of the 1920s indeed continues to reconfigure around differently nuanced curricular dimensions, ranging from the

developmental–ecological crisis, rational–moral values, academic–everyday knowledge, intellectual–manual work, and English or mother tongue as the medium, to the material–cultural politics of identity.

The Indian Education Commission (Government of India, 1966) had advocated nonviolent science for development so that India could engage in reinterpretations and reevaluations of its deep fissures of inequality and injustice, using its own cultural resources of compassion, tolerance, and spirituality while it drew upon the new liberalizing forces that had emerged in the West. However, attempts to incorporate cultural or civilizational resources in school curricula have been part of major political contestations, especially in multicultural countries with complex colonial histories. Through its National Curriculum Framework (NCF, 2000), a right-wing party in power at the center promoted a form of cultural imperialism of the dominant religious identity, while its related textbooks presented a distorted and divisive view of history (Safdar Hashmi Memorial Trust (SAHMAT) and *sabrang.com*, 2001), even introducing a spiritual quotient on the lines of the intelligence quotient. A change in government recovered the secular space in the subsequent NCF 2005; however, since actual curricula and textbooks are prepared by state governments, the right-wing ideology prevails in some states. Lall (forthcoming) analyzes this curricular contestation over national identity to argue that “fundamentalisation in general, and curricular fundamentalisation in particular, are state-controlled discursive mechanisms through which to contain and deflect potential dysfunctions produced by the effects of globalization in societies.”

The Education Commission (Government of India, 1966) policy recommendation for a common school system for social cohesion and equity had been problematically translated by the first NCF (Government of India, 1975) as justification for a centralized common curriculum. It had said: “For a vast country like ours with its diversity of languages, social customs . . . and uneven economic development, the needs and demands of individuals and society will have differential pulls on the school curriculum, varying from one region to the other. For the sake of uniformity of standards and of national identity, therefore it is necessary to develop a common curriculum within a broad framework of acceptable principles and values” (Government of India, 1975: Section 2.1). Policy commitments are seen to have been circumvented by most nation states since “modern education, despite its egalitarian rhetoric, was never designed to provide equal or even appropriate education for all,” having been created by and for the center to consolidate its power through peripheralism of the majority (Cummings, 2003: 277).

Indeed, the system’s inherent inability to define the role of the curriculum, inevitably reduced to a fixed list of contents, variously called the syllabus or standards,

deliberately distorts the notion of diversity to instead construct slow and fast learners, and flexibility to imply addressing “the special needs of the talented, the backward, and those coming from non-formal channels” (Government of India, 1975: Section 2.15). However, the latest NCF (Government of India, 2005) desists from the standards discourse and, acknowledging alarming inequalities, focuses on democratic participation of learners through the construction of knowledge in local cultural contexts. It has promoted participatory development of syllabi and textbooks in a broad critical-indigenous tradition, and a departure from the behaviorist approach is most evident in textbooks at the primary level and in history and political science at higher levels.

Issues of Relevance, Quality, or Efficiency

The traditional high-/low-status divide continues to inform debates on criteria of relevance versus quality – which is differentially defined to offer exclusive skills to privileged groups, as opposed to parents of less-privileged children who want the curriculum to foster more equitable learning opportunities (Weiler, 1993). Indeed, the lack of relevance for urban children and dilution of academic standards were reasons cited by the middle classes for resisting a transformative curriculum that had made a difference to school performance, especially of the poor. Ironically, this was in Kerala, an exceptional state of India, where there is no sharp rural-urban divide, but where, despite near-universal enrolment and high social development indicators, the quality of learning had remained dismal.

Questionable attempts to conflate quality with efficiency, by an increasingly technocratic managerial agenda, often result in problematic quick fixes. Enrolment drives to get children or computers into schools (often devoid of essential blackboards or libraries), without a transformative vision and no examination of the structural inequity perpetuated by the system, result in doling out the same somewhat more efficiently. These programs, aggressively supported by the corporate sector, often profess a minimalist vision of schooling for the poor, with basic literacy and numeracy, and even problematically claim that they can achieve reading first and reading with meaning only later.

A particularly neglected area is curriculum development for out-of-school children and youth, especially girls, which constitutes nonformal education and also the increasingly popular option of open schooling. These demand much more concerted effort through academic-activist partnerships to bring insights from theory and praxis to address the situated knowledge and fragile literacies of mature learners, and also undertake public action to ensure appropriate accreditation. Such effort was evident in a few programs on women's empowerment

and the national literacy campaigns, as can be seen in the work on numeracy (Rampal, 2003), incorporating learners' indigenous knowledge of folk and street mathematics, or health education conducted by nongovernmental organizations (NGOs) such as Bharat Gyan Vigyan Samiti (BGVS) and Nirantar. Interestingly, learnings from the nonformal arena are appearing in school curricula; for instance, insights from unschooled math now find a place in *Math-Magic* – the new National Council of Education Research and Training (NCERT) textbooks developed for primary schools.

Schools and nonformal centers run by NGOs continue to be relegated to the margins of unrecognized institutions, without accreditation or certification. Neoliberal reform has not encouraged diversity in curriculum but, instead, has devalued alternatives and promoted a return to traditionalism (Gewirtz *et al.*, 1995). For instance, the Madhya Pradesh State Government in India aggressively promoted its own minimalist education guarantee schools, while closing down the Hoshangabad Science Teaching Programme, which ran a critical participatory curriculum in over 1500 rural government schools for three decades (Rampal, 2000). Indeed, the struggle for equality and equivalence of nonformal and nongovernmental curricula, meant for those left out of the formal system, has to contend with the same power equations that marginalize them in the first place.

Work and Education

Indigenous curricula to promote education for production were developed in several countries such as Ghana, Botswana, Cuba, and Vietnam. The EFA report (UNESCO, 2005) notes that Cuba, one of the few countries with high-quality education, has long laid emphasis on developing the whole individual while linking education with life and work. Its curricular processes promote solidarity and collaborative competition in the form of emulation for self-improvement – among students, teachers, and schools. The *colectivo pedagogico*, or teachers' collectives, enhance motivation and ensure participatory development of curricula and materials. “Giving children productive responsibilities, a typical educational feature of pre-literate agrarian societies, is usually lost in western schools, which traditionally deposit all authority and responsibility with teachers and encourage passive attitudes on the part of students” (Gasperini, 2000: 13). However, Cuba's residential secondary schools in the countryside, which had participated in the agricultural development plan of each region, with students working a few hours a day on coffee or citrus plantations, have now been subject to the pressure of changing social aspirations.

A study in South Korea warns that curriculum reform to morally engineer aspirations in the name of national

development without addressing people's work subjectivities can result in a hegemonic project to mechanically impose manual work without the state making consonant academic and financial investments in vocational and technical education (Cho and Apple, 2003). In the 1990s, the state attempted to curb labor strikes and augment labor shortage through a program for career education by its Department of Moral Instruction as a way to keep people from pursuing irrational aspirations for college education, and retain them as industrial workers. However, flexibility in terms of basic and common knowledge skills meant to help students adapt to rapidly changing work structures, as opposed to the earlier vocational education model of specific job skills, was seen to append a further disqualification to a curriculum that was popularly perceived as the 3 Ds – difficult, dangerous, and dirty. Curriculum development therefore calls for a continuous dialogic negotiation between selection of knowledge worth knowing and public validation of such knowledge.

In low-income countries like India, vocational education curricula remain the least sought after, perceived as meant for the nonacademic backward learners, even while working-class families despair that schools alienate their children from their own vocations and livelihoods. More often, institutes or polytechnics that offer such courses are not creatively or academically engaged with education or curriculum development, and, in some cases, are even placed under the Labour Department. In the renewed globalizing discourse of brain versus body skills, where creative twenty-first-century skills are competitively sought for curricula in industrialized countries, almost justifying the outsourcing of low-skill jobs to low-income countries, there lies an urgent challenge to design indigenous vocational curricula with an innovative and academic high-skill edge for the majority. Moreover, as Brown *et al.*, (forthcoming) argue, the dominant discourse on education and globalization needs to be challenged to show “that Britain and the US are not knowledge economies, where the value of knowledge continues to rise, but are characterised by an economy of knowledge that is transforming the relationship between education, jobs and rewards.”

National Aims and Changing Contexts

Despite stated national aims, the sociopolitical environment of a country plays a significant role in how the curriculum translates into educational outcomes, as seen in the sharply contrasting situations of Cuba, Chile, and Brazil (Carnoy *et al.*, 2007). Even though, in all three societies, education is ideologically viewed as the great social equalizer – for transforming class structure into meritocracy, and providing pupils from different social classes a common experience of national education – their markedly different social structures produce different

results. Brazil and Chile display much more economic and social inequality than Cuba. Thus, in Brazil, education does not play its professedly equalizing role; children in low-income regions have far less access to schools, with far fewer resources than the public schools in Chile or Cuba, while the latter shows consistently outstanding achievements.

Despite the call for education for all with an expanded vision – for the empowerment of children, youth, and adults – by the 1990 Jomtien World Declaration, curricular aims are seen to be shifting to facilitate, rather than interrogate, the global neoliberal agenda. An EFA review (Amadio *et al.*, 2005) of the trends in national aims between the 1980s and 2000s shows an increased call for expanding human capabilities and choices with a curriculum shift toward a student-centered approach. Within the goal of strengthening human societies, notions of democracy and citizenship find more emphasis in countries in Central and Eastern Europe (having prior ties with Europe and current aspirations for integration into the European Union), while the accent on national identity is more marked in Central Asia. Though citizenship could ambivalently mean ensuring democratic participation in elections or simply pledging allegiance to the ruling party, national identity has remained one of the most important aims even in the 2000s, with positive recognition of linguistic and cultural diversity – in the past seen as major obstacles in nation building. The Latin American region stands out with at least 17 countries having developed bilingual education programs, and 11 constitutions amended to recognize the right of indigenous populations to education in their own languages. The EFA review notes that, unlike in the 1980s, only a few references to socialism are seen in the 2000s. Most significantly, the role of education in helping individuals to transform society – a notion expressed in several country statements of the 1980s – now lies abandoned in favor of “facilitating successful adaptation to an ever-changing world” (Amadio *et al.*, 2005).

Curriculum development in countries with large populations marginalized by the globalized agenda of development acquires a greater sense of urgency as inequalities have been seen to increase in their rapidly changing local contexts. This only underscores the role of indigenous critical agency to reiterate the relationship between education, equity, and social transformation.

Bibliography

- Amadio, M., Gross, S., Ressler, P., and Truong, N. (2005). *Quality Education for All? World Trends in Educational Aims and Goals between the 1980s and the 2000s*. Thematic paper for the UNESCO Global Monitoring Report. Paris: UNESCO.
- Bhattacharya, S. (ed.) (1997). *The Mahatma and the Poet: Letters and Debates between Gandhi and Tagore 1915–1949*. New Delhi: National Book Trust.

- Brown, P., Lauder, H., and Ashton, D. (forthcoming). Towards a high skills economy: Higher education and the new realities of global capitalism. In Boden, R., Deem, R., Epstein, D., Rizvi, F., and Wright, S. (eds.) *World Year Book of Education 2008. Geographies of Knowledge, Geometries of Power: Higher Education in the 21st Century*. London: Routledge.
- Carnoy, M., Gove, A., and Marshall, J. H. (2007). *Cuba's Academic Advantage: Why Students in Cuba Do Better in School*. Stanford, CA: Stanford University Press.
- Cho, M. K. and Apple, M. W. (2003). Schooling, work and subjectivity. In Apple, M. W. (ed.) *The State and the Politics of Knowledge*. New York: RoutledgeFalmer.
- Cummings, W. K. (2003). The limits of modern education. In Beauchamp, E. R. (ed.) *Comparative Education Reader*. London: RoutledgeFalmer.
- Gasperini, L. (2000). The Cuban education system: Lessons and dilemmas. *Country Studies: Education Reform and Management Publication Series* 1(5), 21752.
- Gewirtz, S., Ball, S., and Bowe, R. (1995). *Markets, Choice and Equity in Education*. London: Open University Press.
- Government of India (1966). *Report of the Education Commission 1964-66: Education and National Development*. New Delhi: Author.
- Government of India (1975). *National Curriculum Framework*. New Delhi: National Council for Educational Research and Training.
- Government of India (2005). *National Curriculum Framework*. New Delhi: National Council for Educational Research and Training.
- Hindustani Talimi Sangh (All-India Education Board) (1938). *Report of the Zakir Hussain Committee on Basic Education*. Wardha, Maharashtra: Author.
- Lall, M. (forthcoming) Globalisation and the fundamentalisation of curricula: Lessons from India. In Hopson, R. K., Yeakey, C. C., and Boakari, F. M. (eds.) *Power, Voice, and the Public Good: Schooling and Education in Globalized Societies*. Oxford: Elsevier.
- Pinar, W. F. (2004). *What is Curriculum Theory?* Mahwah, NJ: Lawrence Erlbaum Associates.
- Rampal, A. (2000). Education for human development in South Asia. *Economic and Political Weekly* 35(30), 2623-2631.
- Rampal, A. (2003). Counting on everyday mathematics. In Saraswathi, T. S. (ed.) *Cross Cultural Perspectives in Human Development: Theory, Research and Applications*. Thousand Oaks, CA: Sage.
- SAHMAT and Sabrang.Com (2001). *Against Communalisation of Education*. New Delhi: Author.
- United Nations Educational, Scientific and Cultural Organization (UNESCO) (2005). *Education for All Global Monitoring Report 2005 – The Quality Imperative*. Paris: Author.
- Weiler, H. N. (1993). Control versus legitimization: The politics of ambivalence. In Hannaway, J. and Carnoy, M. (eds.) *Decentralisation and School Improvement: Can We Fulfill the Promise?* San Francisco, CA: Jossey-Bass.

Further Reading

- Combleth, C. (2000). National standards and curriculum as cultural containment. In Combleth, C. (ed.) *Curriculum Politics, Policy, Practice*. Albany, NY: State University of New York Press.
- Freire, P. (1985). *The Politics of Education: Culture, Power and Liberation*, Macedo, D. (trans.). London: Macmillan.
- Goodson, I. F. (1993). *School Subjects and Curriculum Change*. London: Falmer.
- Green, A., Preston, J., and Janmat, J. G. (2006). *Education, Equality and Social Cohesion: A Comparative Analysis*. New York: Palgrave Macmillan.
- Kenway, J. (1990). Education and the right's discursive politics. In Ball, S. J. (ed.) *Foucault and Education: Disciplines and Knowledge*. London: Routledge.
- McLaren, P. (1994). *Life in Schools: An Introduction to Critical Pedagogy in the Foundations of Education*, 2nd edn., New York: Longman.
- PROBE Team (1999). *Public Report on Basic Education in India*. New Delhi: Oxford University Press.
- Young, M. F. D. (ed.) (1971). *Knowledge and Control*. London: Collier-Macmillan.

Relevant Websites

- <http://www.ncert.nic.in> – Government of India (2005). National Curriculum Framework National Council for Educational Research and Training, New Delhi.
- <http://www.ncert.nic.in> – NCERT (2006). Syllabus for Elementary School, National Council for Educational Research and Training, New Delhi.
- <http://www.ncert.nic.in/testing/> – NCERT (2007). Math-Magic: Textbooks for Class II-IV, National Council for Educational Research and Training, New Delhi; all textbooks.
- <http://www.rethinkingschools.org> – Rethinking Schools Online.
- <http://www.unesco.org> – UNESCO Global Monitoring Report.

Globalization and Curriculum

R Dale, University of Bristol, Bristol, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Governance – Used often to contrast with government, it implies a form or means of coordination of policy activity that is not wholly state run, but that which involves combinations of actors, such as state, market and community, and activities of governing, such as funding, provision, and regulation of services. It may be seen as the coordination of coordination, referring to the processes of combining forms of coordination, such as the state and market. This is the sense in which it is used in this article. As used in good governance in the discourse of the World Bank and other international organizations, it is seen as involving respect for the rule of law and private property in order to assure the safety of investment, and it is used to refer to changes in how universities are organized, managed, and controlled.

Mandate(for education) – What is considered desirable for education to achieve.

Modernity – “that historically unprecedented amalgam of new practices and institutional forms (science, technology, industrial production), new ways of living (individualism, secularism, instrumental rationality), and new forms of malaise (alienation, meaninglessness, a sense of impending social dissolution)” (Taylor, 2004).

PISA – The Programme for International Student Assessment, created by the Organisation for Economic Cooperation and Development (OECD), is an internationally standardized assessment that was jointly developed by participating countries and administered to 15-year-olds in schools.

Introduction

Globalization connotes a notoriously complex set of issues and problems that present particular problems in an article like this. The best way to address that complexity is to embrace it, rather than seek to simplify and possibly distort it. That involves recognizing that there are at least five distinct, but nonmutually exclusive, and overlapping perspectives on globalization, and that it is necessary to take each of them into account to get an effective sense of the

relationship between globalization and curriculum. These see globalization as an ontology; a discourse; a process; a condition, or state of affairs; and as a political project (of different kinds) (see Mittelman, 2000). This article first specifies its broad (ontological) approach. From this, key propositions will be derived, and addressed through each of the lenses set out above.

This is an auspicious moment to discuss the relationship between globalization and school curriculum, because their relationship is developing in the early stages of what will undoubtedly be a prolonged paradigm shift in the development of capitalism. This fundamental argument draws on the work of Boaventura de Sousa Santos in suggesting that it is crucial to the understanding of the current global era to distinguish between the trajectories of capitalism (as found currently in the form of neoliberal globalization), and modernity, and to examine the relationships between them, as they begin to unravel under neoliberal globalization. His fundamental argument is the “‘sociocultural paradigm of modernity’ will disappear before capitalism ceases to be dominant,” partly because it will be both superseded as it fulfills some of its promises, and become obsolescent as it proves incapable of meeting others (Santos, 2002: 1–2).

The basis of the argument in this article, then, is that education, in the form of schooling, with curriculum at its heart, has been perhaps the most significant modern institution, and that what we now observe and experience are symptoms of what may be seen as its obsolescence vis-à-vis the current, neoliberal phase of capitalism. One example of this obsolescence is the curriculum as it was framed, assumed, and experienced over most of the second half of the last century. The key consequence of this is that the relationship between globalization and curriculum is becoming not so much one of updating the contents of the curriculum as container, but one where the very shape of the container, and its place in the processes of education, are undergoing significant change.

This is because the shape of the curriculum container was determined in modernity by a common grammar of schooling, and perceived national needs, while the place of the curriculum was to embed the aims and objectives of education through determining the content of what was to be taught, and how it was to be taught, in schools. Now both of those things are under strain from the changes just mentioned, as the position of the national in both the mandate and the governance of education changes, and the demand of schools in what is represented as a global knowledge economy (GKE) shifts from content knowledge

to competences (and as we shall see below, it also comes under pressure from antiglobalization forces). One important complicating factor is that these changes are not reflected in the terminology used to describe what is happening. As the anthropologist Gavin Smith warns “a whole series of key concepts for the understanding of society derive their power from appearing to be just what they always were and derive their instrumentality from taking on quite different forms” (Smith, 2006: 628). Thus, the curriculum, the national, and the state that are discussed today are not the same curriculum, national, and state that were assumed 20 years ago. So, it is crucial in looking at the relationships between globalization and curriculum not only to take into account the different forms of globalization, but also to take a critical perspective on what the term curriculum is actually referring to, as well as recognizing that the relationships taking varying, not identical, forms across time and space.

A Very Brief Ontology and Epistemology: A Common World Education Curriculum, or a Globally Structured Agenda for Education?

Any theory of the relationship between globalization and curriculum has to have three components: an account of globalization, an account of curriculum, and an account of how the two are linked. Two different accounts that meet these criteria have been outlined and discussed elsewhere (Dale, 2000); these are referred to as the Common World Education Culture (CWEC) and the Globally Structured Agenda for Education (GSAE).

The labels themselves are, of course, deliberately contrasted. Both World and Global imply an extranational focus. While the former sees nation-states that organize education as part of an international system of nation-states, and as embedded in a world society (Meyer and Ramirez, 2000: 113), Global implies social and economic forces operating supranationally and transnationally, rather than internationally, to elude, break down, or override national boundaries, while reconstructing the relations between nations. Structured agenda is contrasted with culture; the latter implies a shared, and equally available, set of resources at a high level of generality, the former, a systematic set of unavoidable issues for nation-states that is framed by their relation to globalization.

In CWEC, education is largely equated with curriculum, particularly the increasing isomorphism of curriculum categories at a transnational level. The striking degree of homogeneity across the societies of the world, irrespective of their location, level of development, or religious, and other traditions derive from the very model of modern stateness that has spread since 1945. As the old empires broke down, they were replaced by new states whose legitimacy as real states

was based on their adherence to the same values and models of Western modernity as those of their former colonial masters (see Wong, 1992: 126).

Education is central to this modernizing mission; the curriculum, too, is based on world, rather than national level curricular scripts (Cha, 1992), so that the curricular commonality that we observe is not the product of globalization as currently conceived, but has a much longer history, associated with the spread of the values and assumptions of modernity, such as scientific rationality, individualism, and progress.

In the GSAE, what counts as education is seen as framed by questions of practice (who is taught what, how, under what circumstances), policy (how and by whom are these things determined), political economy, and outcomes (public, private, personal) (see Dale, 2000). For GSAE, globalization is a set of political-economic arrangements for the organization of the global economy, driven by the need to maintain the capitalist system rather than by any set of values. No nation-state, not even the United States or China, is now able to control or direct the global economy. Rather, it is the interests of the global economy, as represented through the plethora of international organizations created for the purpose such as Group of 8 (leading industrial countries; G8), World Economic Forum (WEF), Organization for Economic Cooperation and Development (OECD), World Bank, World Trade Organization (WTO), and regional organizations such as the European Union (EU) and North American Free Trade Association (NAFTA) to which individual nation-states willingly cede some part of their national power in the interest of collective control of the global economy in their joint interests (which means that rather than a tension between globalization and nation-states, certainly the rich ones, states are themselves the most active agents and promoters of globalization), that ultimately sets the parameters for education and curriculum.

In summary, then, the proponents of the CWEC approach highlight the presence of international discourses, institutions, and practices based on the ideas, norms, and values of Western modernity that inform – even script – national responses to a range of issues, including, especially, curricula. In the GSAE approach, neoliberal globalization is seen as eroding the national, state-centered, and curriculum-as-content assumptions of the CWEC approach (see Dale and Robertson, 2007).

On this basis, this article is structured around the following propositions:

- that the curriculum as it has been known and practiced takes its form and purposes from the discourses, institutions, and practices of modernity;
- that globalization represents a challenge to each of those elements, and thus to the nature of the curriculum that was shaped by them; and

- that it is possible to discern the outlines of a new container for the processes and products of education that involves a shift from content to competences and from teaching to learning.

The relationships between these propositions and each of the different conceptions of globalization noted above will now be briefly considered.

Globalization as Discourse

The CWEC account of the relationship between curriculum and the extranational has seen the discourses of modernity – science, progress, individualism, and the nation-state – as the key framework of curriculum debates. The place and role of curriculum were essentially meliorist and progressive, an essential means through which nation-states achieved their goals through education, and in particular the basis of nation-building. Education was essentially equated with schooling, and in particular with the grammar of schooling, which includes such conceptions as the school year and the school day and the division of that day into periods; classes and classrooms; and a curriculum based on subjects. This has developed over a considerable period, to become itself an independent element in schooling and curriculum.

Recognizing that significant changes have occurred, John Meyer has argued that curricular models that assume a national base are now being replaced by those based on global models of what national societies and states should be like and of what they should do (Meyer, 2001). However, this presents globalization as a process without an agent (see next section) and assumes the same shape and place of the curriculum, which is not perceived as requiring anything more than adjustment to changed circumstances.

For the major international organizations, the major discursive shift has been to see globalization as entailing a new GKE of which all nation-states are part, and to which education is to contribute in a major way. The GKE is seen to require a radical modification of the curriculum, with changes in its shape and place. In particular, competences are to take priority over content knowledge, and universalism is to give way to tailored offerings, right up to the point of personalized learning, which is to be achieved through any place, any time, any provider life-long learning. The shift to a GKE representation of globalization requires a transformation of thinking from curriculum as a body of knowledge justified by its basis in and contribution to the values of modernity, and to be taught in a planned, sequenced way, to competences to be learned *ad hoc*, as and when necessary.

Globalization as a Condition, or State of Affairs

The main curricular responses to the altered condition, or state of affairs in which we live as a result of globalization

have centered on a re-emergence (Marshall, 2003) of the field of global education. Marshall, writing mainly from a UK perspective, and Kirkwood (2001) from a US perspective, both suggest that global education originated essentially with the UN Declaration of Human Rights, addressing the need for a place in the curriculum for world studies, human rights, peace education, and so on. Here, global education seems tightly related to discourses of modernity. The main curriculum pattern is one of nationally based responses, as nation-states perceive a need to respond to the changed condition of globalization and to shape their responses to the way it may impact on them. This has been largely based on what might be called global education for, and as part of, national citizenship. This involves “bringing the world into the classroom, where teachers teach from a world-centric rather than an ethno-specific or nation-state perspective” (Kirkwood, 2001), or what Marshall (2005) calls a “global gaze.” Kirkwood commends Merryfield’s framework for global education, which comprises: human beliefs and values; global systems; global issue and problems; cross-cultural understanding; awareness of human choices; global history; acquisition of indigenous knowledge; and development of analytic, evaluative, and participatory skills (Merryfield, 1997). Marshall’s analysis produces rather more critical frameworks, with a more explicit emphasis on social justice and antiracism (Marshall, 2005), while Davies’ focus on the consequences of the security and conflict agendas leads her to take this point further, suggesting that this leads to a more direct concern with matters of social justice rather than “international awareness or ‘global citizenship’” (Davies, 2006: 6). Davies also very usefully sets out the range of possible combinations of global citizenship and education, as follows:

1. global citizenship + education (definitions of the global citizen, and the implied educational framework to provide or promote this);
2. global + citizenship education (making citizenship education more globally or internationally relevant; think global, act local);
3. global education + citizenship (international awareness plus rights and responsibilities); and
4. education + citizenship + global (introducing dimensions of citizenship and of international understanding into the school curriculum, but not necessarily connected) (Davies, 2006: 13, 14).

These can be seen as moves from reactions to banal (McDonaldization, global village) conceptions of globalization to more analytically based responses to the condition of globalization. Moving qualitatively beyond these response are approaches best exemplified in the work of Noel Gough on transnational curriculum inquiry. These explicitly recognize the need for problematizing and reimagining, rather than seeking to modify, however radically, the approaches to curriculum inherited from the

discourses of modernity, and the need for a shift from the national to the transnational as the basis for curriculum work. What these entail are, respectively, the need to move beyond representations of curriculum that see it as a fixed bundle of necessary elements, such as contents, processes, objectives, outcomes, and so on, that are mutually related; and international curriculum scholars creating transnational spaces to reframe and decenter existing knowledge traditions as a basis for collective work (Gough, 2004).

Globalization as Process

As noted above, it is crucial to recognize that while globalization can be seen as a process, it is never a process without agents, and that the main agents of globalization are nation-states themselves.

For CWEC, the main agents of the processes of curricular isomorphism are epistemic communities, international groups of experts, within and outside international organizations who collectively shape what is to count as curriculum, as effectively “carriers” of the principles of the world polity and culture (McNeeley and Cha, 1994: 2, 3). The picture is of modern discourses, carried by modern institutions with the authority of science, that bring about the isomorphism of national systems that is at the center of the world polity theorists’ claims about the curriculum.

By contrast, GSAE sees new agents and new activities, operating at multiple scales, emerging around the governance of education, including the shape and place of the curriculum. These include the development of education as a global commodity for sale on a world market (see Robertson *et al.*, 2002). This has profound effects on curriculum. Because the product has to be globally marketed, it must shed nationally, or regionally, defining features. Other new, or transformed, agents of globalization are international textbook publishers, and especially purveyors of electronic curricular materials. The two agents who are most altered by globalization are the state and international organizations. The role of the state has become smaller and more focused. In terms of objectives, it has become predominantly a competition state, whose priority is increasing the competitiveness of firms located within its borders. Its scope and *modus operandi* are also changed, in direct response to neoliberal globalization, particularly through the widespread application of the New Public Management (NPM) which calls for the separation of policy advice and implementation and of provision and funding of public services. Both of these, and especially the former, have direct implications for curriculum. However, the most significant impact of the NPM on curriculum comes in the form of the widespread use of techniques of audit, benchmarking, and indicators set in place as accountability mechanisms under the NPM. These involve a crucial shift from *ex ante* to *ex post* monitoring and evaluation of public services. In the case of the curriculum, this involves

a shift from monitoring and evaluation of curriculum content, prior to its being implemented, to monitoring and evaluating its outcomes. This gives much greater power to the monitors and evaluators. Setting the targets to be achieved, and monitoring (and publishing the results of) their achievement, can control what is taught in schools much more effectively, and curriculum becomes more and more what the test/evaluation requires.

The most powerful and wide-ranging example of this is the OECD’s Programme of International Student Assessment (PISA) program. PISA assesses the competences – the ability to use what they have learned, rather than the content of what they have learned – of 15-year-olds in a large number of countries, including many that are not members of OECD. The crucial point here is that the competence focus of the program, together with the publication of its results in league table fashion, has led to great hand-wringing and significant changes in the secondary education programs of a number of countries, most notably Germany. PISA may be the most effective mechanism of bringing about fundamental changes in secondary school education across the world that we have yet seen.

Globalization as a Project

Finally, the question of how the idea of globalization as a project to be pursued, rather than a process without agents, or as relatively passively received, might affect the curriculum is addressed. At least three quite distinct and contrasting projects of globalization can be identified. The first is the neoliberal form of globalization, which is itself a political-economic project, driven largely by finance capital and those who would benefit from minimizing barriers to free trade, which we have already discussed briefly; its focus on audit provides a common currency of educational exchange, suitable to a GKE, whereas the other projects, in different ways, aim to provide a common language of educational exchange, suitable to a more humane world.

The other two projects are conceived as respectively alternative forms of globalization, which might be loosely characterized as seeking good globalization, and oppositional forms of globalization, or antiglobalization movements, as they are generally represented.

Much of the alternative approach may be seen as radical versions of global education, such as Gough’s; the difference is summed up by Davies’ distinction between the “empathy” of the one, and the “outrage” of the other (Davies, 2007: 7). Several diverse strands come together under the alternative heading. It assumes Touraine’s argument that the problems created by globalization are not so much economic as social. There are connections with conceptions of cosmopolitanism, which welcomes the break up of the nation-state system generated by globalization but looks for a more humane version of globalization than the

currently economically dominated one. It also draws on the existence and nature of what are often called planetary problems, such as environmental and climate issues.

These alter-globalizations are seen to be susceptible to realization through education, particularly through appropriate curricula. They are not fundamentally opposed to either capitalism (except the worst excesses of neoliberal capitalism) or modernity (except in the form of an exclusive Western canon), but seek to work with both, while seeking to bring about changes in the societies, and the relations between them, that have been produced by the excesses of both.

By contrast, the oppositional approach is strongly opposed to both modernity and capitalism, both of which it sees as implicated in the subjugation of two-thirds of the world. While most alter-globalization comes from the West, and while there is certainly significant antiglobalization in the West (especially around the WTO), the most radical curriculum-relevant forms come from the (metaphoric) South. One form may be seen as reformist, emphasizing the importance of the local and an end to the domination of the West. Torres (2000), for instance, argues for a reversal of the (Western) supply-driven education, with donor-controlled agendas and conceptions of what counts as knowledge, and the conditions within which it is to be realized (such as the prescription of minimum class sizes of 40, and a limit on teachers' salaries), all of which impact on the nature and value of curriculum. The more radical views are represented in the very substantial and important postcolonialist and subaltern literatures. What differentiates these approaches from those we have discussed so far (with the exception of Gough) is that they represent critiques of globalization from outside rather than from within, using the language, epistemology, and cognitive understandings of modernity as the basis for the critique of modernity and globalization themselves. What the antiglobalization project emphasizes is that the Western globalization project is not only economically and politically, but also culturally, cognitively, and epistemologically driven.

Santos (2004) makes this point most clearly with his argument is that "there can be no social justice without global cognitive justice." Of particular significance is his "sociology of absences" which "consists of an inquiry that aims to explain that what does not exist is, in fact, actively produced as non-existent" (p. 238). He distinguishes five logics through which absences are created: "The *monocultures of knowledge, of linear time, of classification, of the universal and the global, and of criteria of capitalist productivity and efficiency*. These produce (respectively)... five principal social forms of non-existence: the ignorant, the residual, the inferior, the local and the non-productive" (Santos, 2004: 238, 239). The basis and the target for curriculum must be the acknowledgement of these absences and the means of their creation as a necessary condition of any antiglobalization project.

Conclusion

Globalization has generated two fundamental critiques of the curriculum as it has been recognized over the past half-century. The first is that it is increasingly unfit for purpose; it is unable to supply the requirements of the GKE. The second is that it has exposed starkly the obstacles to global social justice that curriculum as we have known it constitutes. Both these have been to a degree recognized by curriculum workers, but this has been in the form largely of reactions that draw on, while trying to modify, the existing paradigm. The more basic responses called for by both critiques, respectively, the transformation of the system of education as schooling and the abandonment of the current conception of curriculum in both its economic and cultural/epistemological/cognitive assumptions remain on the agenda, and may increasingly come to constitute the spaces that have until now been specified and occupied by conceptions of curriculum.

Bibliography

- Cha, Y.-K. (1992). The origins and expansion of primary school curricula: 1800–1920. In Meyer, J., Kamens, W., David, H., and Benavot, A. (eds.) *School Knowledge for the Masses: World Models and Curricular Categories in the Twentieth Century*, pp 63–73. London: Falmer.
- Dale, R. (2000). Globalisation and education: A common world education culture, or a globally structured agenda for education? *Educational Theory* 50, 427–448.
- Dale, R. and Robertson, S. (2009). Beyond methodological 'isms in comparative education in an era of globalisation. In Kazamias, A. and Cowen, R. (eds.) *Handbook on Comparative Education*. Dordrecht: Springer.
- Davies, L. (2006). Global citizenship: Abstraction, or framework for action? *Educational Review* 58, 5–25.
- Gough, N. (2004). A vision for transnational curriculum inquiry. *Transnational Curriculum Inquiry* 1. Retrieved 19th June, 2007 from <http://nitinat.library.ubc.ca/ojs/index.php/tci> (accessed April 2009).
- Kirkwood, T. F. (2001). Our global age requires global education: Clarifying definitional ambiguities. *Social Studies* 92, 10–17.
- Marshall, H. (2003). Review essay: Global education: A re-emerging field. *British Journal of Sociology of Education* 24, 397–405.
- Marshall, H. (2005). Developing the global gaze in citizenship education: Exploring the perspectives of global education NGO workers in England. *International Journal of Citizenship and Teacher Education* 1, 76–92.
- McNeeley, C. and Cha, Y.-K. (1994). Worldwide educational convergence through international organizations: Avenues for research. *Education Policy Analysis Archives* 2. Retrieved 28th July, 2007 from <http://epaa.asu.edu/epaa/v2n14.html> (accessed April 2009).
- Merryfield, M. M. (1997). A framework for teacher education: Preparing teachers to teach global perspectives. In Merryfield, M. M., Jarchow, E., and Pickert, S. (eds.) *A Handbook for Teacher Educators*, pp 1–24. Thousand Oaks, CA: Corwin.
- Meyer, J. W. (2001). *Globalization, National Culture and the Future of World Polity*. Wei Lun Lecture, The Chinese University of Hong Kong, November 28.
- Meyer, J. W. and Ramirez, F. O. (2000). The world institutionalization of education. In Schriewer, J. (ed.) *Discourse and Comparative Education*, pp 111–132. Berlin: Peter Lang.
- Mittelman, J. (2000). *The Globalization Syndrome: Transformation and Resistance*. Princeton, NJ: Princeton University Press.

- Santos, B. de S. (2002). *Toward a New Legal Common Sense*. London: Butterworth.
- Santos, B. de S. (2004). The world social forum: Towards counter-hegemonic globalization (Part 1). In Sen, J., Anand, A., Escobar, A., and Waterman, P. (eds.) *The World Social Forum: Challenging Empires*, pp 235–245. New Delhi: Viveka Foundation.
- Smith, G. (2002). When “the Logic of Capital Is the Real Which Lurks in the Background” Programme and Practice in European “Regional Economies”. *Current Anthropology* 47(4), 621–639.
- Taylor, C. (2004). *Modern Social Imaginaries*. Durham, NC: Duke University Press.
- Torres, R. M. (2000). *One Decade of Education for All: The challenge Ahead*. Paris/Buenos Aires: IIEP/UNESCO.
- Wong, S.-Y. (1992). The evolution and organization of the social science curriculum. In Meyer, J. W., Kamens, D. H., and Benavot, A. (eds.) *School Knowledge for the Masses: World Models and Curricular Categories in the Twentieth Century*, pp 124–138. London: Falmer.
- Davies, L. (2005). *Education and Conflict: Complexity and Chaos*. London: Routledge Falmer.
- Gough, N. (2000). Locating curriculum studies in the global village. *Journal of Curriculum Studies* 32, 329–342.
- Luke, A. (2002). Curriculum, ethics, metanarrative: Teaching and learning beyond the nation. *Curriculum Perspectives* 22, 49–55.
- Muller, J. and Subotzky, G. (2001). What knowledge is needed in the new millennium? *Organization* 8, 163–182.
- Priestley, M. (2002). Global discourses and national reconstruction: The impact of globalization on curriculum policy. *Curriculum Journal* 13, 21–138.
- Robertson, S., Bonal, X., and Dale, R. (2002). ‘GATS and the Education Service Industry’. *Comparative Education Review* 46(4), 472–496.
- Robertson, S., Novelli, M., Dale, R., and Tikly, L. (2007). *Globalisation, Development and Education*. London: Department for International Development.
- Tawil, S. (2003). Curricular change: A global perspective: Introduction to the open file. *Prospects* 33, 11–18.

Further Reading

- Benavot, A. and Braslavsky, C. (eds.) (2006). *School Knowledge in Comparative and Historical Perspective: Changing Curricula in Primary and Secondary Education*, CERC Studies in Comparative Education, No. 18. Hong Kong: Comparative Education Research Center, University of Hong Kong.

Relevant Website

- <http://www.iipe-buenosaires.org.ar> – Instituto Internacional de Planeamiento de la educación, Sede Regional Buenos Aires.

Popular Culture

K Schultz and R Throop, University of Pennsylvania, Philadelphia, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Popular culture influences and is influenced by our desires, language, identities, and moment-to-moment interactions. Daily, youth experience and (re)create popular culture through music transmitted onto iPods, phrases appropriated from television programs, political statements crafted on YouTube or podcasts, portraits created on MySpace, and poetry performances. Despite the possibilities engagement with popular culture can offer in classroom settings, its relationship to curriculum remains contested. Youth engage with new forms of media and digital technologies in a manner that is often perplexing to adults, including their teachers. What should be the relationship between popular or everyday culture and the content of education? Rather than explicitly answering this question, this article provides an overview of the literature that informs this debate.

Recently, Pennycook (2005) has argued that music, such as hip-hop, and languages, such as English, are both fixed and fluid. For instance, hip-hop texts and English language are fixed to the extent that they are situated in local contexts. On the other hand, they are also always fluid, as they change across space and time through performance. This same distinction is useful for understanding how popular culture has been conceptualized in relationship to curriculum. As fixed, the study of popular culture focuses on the critical analysis and consumption of texts. When understood as fluid, an examination of popular culture emphasizes youth productions across multiple media and modalities. These different analytic perspectives have practical consequences for how popular culture is understood in relation to school. An analysis of popular culture and curriculum as fixed leads to a discussion of what to include, or add, to the curriculum to reflect youth's knowledge, interests, and activities outside of school. An understanding of popular culture and curriculum as fluid takes the influence of popular culture on curriculum and pedagogical practices seriously, and focuses instead on the changing and global nature of youth cultural production both in and out of school.

This article begins with a historical look at the definitions of popular culture. Various conceptions of the relationships between popular culture and curriculum are explored. Next, the article reviews current studies that elaborate these relationships. Finally, future directions and lingering questions for the field are discussed.

Framing the Study of Popular Culture and Curriculum

Popular culture is often defined as texts and lived experiences received and acted on by people as a part of everyday life. Dolby (2001) elaborates, "Popular culture... is not a solid fixed object, but instead an ever-changing network of movement, which is structured by and through apparatuses of power and is itself a result of struggle" (p. 14). The study of popular culture seeks to explore the relationship between everyday life and the construction of cultural forms (McRobbie, 1994). Perhaps because we live in an age of identity politics, current research on popular culture tends to highlight the role it plays in the formation of identities, situating the popular as an important site of political struggle. Understandings of what popular culture is and how it operates, however, have evolved over time and should be situated historically.

Initially, popular culture developed as a construct to distinguish the culture of the bourgeoisie from that of the masses, carrying with it a disparaging connotation. In Marxist theory, and particularly in the work of Frankfurt School theorists, popular culture is positioned as mere ideology: beliefs and materials imposed on the masses by the mass media. More recently, scholars on the Left have argued that this stance ignores the way people act politically and show resistance through popular culture. While this stance validates the cultural forms of subordinate groups, such as the working class, it is predicated on false assumptions that these groups have a culture uncontaminated by the dominant ideology, producing an essentialist view of popular culture (Giroux and Simon, 1989). Further, it locates agency solely in resistance, ignoring the way social actors may exercise agency through conformity or submission.

Current work on popular culture aims at moving past these shortcomings. Popular culture has evolved into a more fluid construct that encompasses both high and low forms. Since World War II, popular culture has become the dominant mode of communicating knowledge (Kellner, 1995). As Willis (1990) explains, popular culture has "...helped to produce an historical present from which we cannot now escape and in which there are many more materials – no matter what we think of them – available for symbolic work than ever there were in the past" (p. 19). Rather than positioning popular

culture as an imposed ideology, or celebrating it uncritically, scholars contend that oppression and liberation are continuous social projects that exist simultaneously within the popular, depending on the ways in which it is taken up and employed by social actors.

The Relationship between Popular Culture and Curriculum

How are, and how should, popular culture and curriculum be related? What roles do popular culture and curriculum play in social organization? Just as understandings of what popular culture is have changed over time, so have ideas around the relationship between popular culture and school curriculum. Willis' (1977) *Learning to Labor*, a study of working-class high school boys, addresses this set of issues. Willis' text emphasizes the pedagogical influence of popular culture; he demonstrates the way youth use popular culture to respond to and, in some ways recreate, the formal curriculum of schooling. The lads in his study create a counterculture in opposition to the official school curriculum and, in so doing, participate dynamically in the (re)production of their social positions. Willis argues that, on some level, the lads sense the structural fact that school will not produce the promised rewards for them; the lads thus create a counterculture in school to protect their dignity. Ironically, this disrupts the intended consequences of the school curriculum and, in effect, disqualifies them from the opportunity to enter into middle-class jobs.

While Willis focuses on the role of popular culture in processes of schooling, Apple (1979) focuses instead on the effects of official school curriculum. In his influential text *Ideology and Curriculum*, Apple describes the role that the content and organization of school curriculum play in maintaining social inequalities. He argues that external control of curriculum distances learning from teachers' and students' lives. As he explains, "the language of learning tends to be apolitical and ahistorical, thus hiding the complex nexus of political and economic power and resources that lies behind a considerable amount of curriculum organization and selection" (p. 28). Apple's critique suggests that, in contrast to the distanced curriculum of school, youth have a vested interest, a voice and a stake in their production and consumption of popular culture. Drawing on Pennycook's (2005) distinction, a fluid rather than fixed notion of curriculum could lead away from the conception of curriculum as texts and materials contained within the school day and walls to a broader understanding of curriculum that extends beyond the boundaries of school (Hull and Schultz, 2002). In this framework, youth take up and recreate popular culture as they construct their identities across the boundaries of home, school, and community.

Today, the trend in many educational systems, however, is to move away from these ideas. Luke (2006) describes the current technocratic/industrial model of education which has led to the deskilling of teachers who are fed curriculum materials tied to large-scale assessments. As a result, curriculum and pedagogy have migrated far away from the popular everyday culture to decontextualized bits of knowledge; teachers have become both the objects and the subjects of the market-produced curriculum. Luke argues that in our global world, curriculum is closely tied to individual nation-states. Schooling is defined and regulated at the local level. He suggests that we need transcultural and cosmopolitan teachers who can move seamlessly between the local and the global. We would add that we need conceptions of curriculum that recognize the ways in which popular culture engages with global ideas while unfolding in distinct ways in local circumstances, crossing the boundaries of school, community, and family.

Whether one focuses on the pedagogical influence of popular culture or the content of official school curriculum, concepts of fixity and fluidity can help in rethinking possible relationships between curriculum and popular culture. For example, if curriculum is defined in a more fluid way and education is conceptualized as occurring across the boundaries of school and community, the fixed dichotomy of learning either in or out of school can be eliminated. This dichotomy suggests that youth engage in popular culture out of school and follow a standard curriculum inside school walls, with occasional moments of engagement in work that reflects their interests and knowledge. Understanding curriculum more broadly as the content and form of learning across time and space, however, allows for a larger picture of what youth learn across a school year. Further, a more fluid understanding of popular culture suggests that we pay attention to the pedagogical influence of the (re)making of popular culture that occurs in both in and out of school spaces. Willis (1990) suggests that schools are losing their influence as popular culture becomes the primary material youth use in constructing their identities. He explains, "Common culture will, increasingly, undertake, in its own ways, the roles that education has vacated" (p. 147). For some, Willis's work raises questions about whether youth are victims of mass marketing (or ideological curriculum) or producers of new cultural forms. Rather than entering this debate, Willis emphasizes that in order to address the relationships between popular culture and curriculum, educators must look beyond the products of popular culture to their uses in context.

Ethnographic research, through the close studies of youth's lives, has provided insight into the way young people relate to, and with, popular culture. School-based

work highlights the way teachers engage with popular culture through their pedagogy and curriculum. As Pennycook (2005) explains, “If we believe that education needs to proceed by taking account of student knowledge, identity and desire, we need to engage with multiple ways of speaking, being, and learning, with multilayered modes of identity at global, regional, national and local levels” (p. 40). The following section highlights current work that in some way extends our understandings of popular culture and school curriculum.

Current Work

It is useful to divide current work on popular culture and curriculum into two categories: research which positions popular culture *as* curriculum and research which demonstrates the inextricable relationship between popular culture *and* curriculum. The first body of work investigates what happens, or what might happen, when popular culture is added to a scripted or more traditional curriculum. Current scholarship along these lines argues for starting with the texts of students’ lives in schools, while concomitantly acknowledging that adding popular culture to the curriculum is not in and of itself sufficient to guarantee student success, nor is it without its own problems. Acknowledging these constraints, scholars thoughtfully explore what happens, or what might happen, when popular culture is incorporated into curriculum, and often detail the relationship between broadening traditional definitions of curriculum and school success.

Popular Culture as Curriculum

Elaborating the concept of circulating literacy practices, Vasudevan *et al.* (under review) describe a longitudinal research project in which they sought to understand the possibilities for increasing student engagement in learning by adding digital modalities and opportunities to bridge home and school worlds in literacy classrooms. By providing a detailed analysis of the composing practices of three students during this year-long multimodal storytelling project, and focusing on the way practices circulate across the boundaries of home, community, and school, the authors show how engagement with various media allowed students to draw on their lives outside of school to produce and rework rich in-school texts. The project demonstrates “. . . the need to reconceptualize composing practices as occurring beyond the printed page and outside the boundaries of the classroom” (p. 1), and addresses the possibilities inherent in broadening our conceptions of curriculum to draw on the popular culture of youth. Significantly, this work offers starting points for considering the affordances and constraints of integrating digital composing practices into the classroom. It points to the

possibilities for increasing student engagement in learning by bridging the popular cultures of youth with the more formalized school curriculum in innovative ways.

Hill (in press) locates his research in a classroom at the nexus of the interpretation and production of texts. Together with a colleague, Hill developed a high school English curriculum around hip-hop texts. He analyzes a particular kind of talk around the text that he calls “wounded healing,” which allows students in the class to produce new knowledge through critical analysis of dominant narratives. He emphasizes the role of texts, interpretation, collective storytelling, and a curriculum based on the lived experiences of youth, in forming a cohesive classroom community. The study reveals the complexity, and indeed danger, in bringing a curriculum based on popular culture into the classroom, warning against romanticizing this move. He illustrates how connections to the popular or everyday lives of students can too easily unearth deep personal pain and social inequality that teachers must be ready to address.

In their text, *Popular Culture in the Classroom*, Alvermann *et al.* (1999) focus on the praxis of teaching and researching critical media literacy. Each chapter of the text details and deconstructs a classroom teaching experience where a teacher taught a lesson on media or popular cultural forms, advocating a self-reflexive approach to media pedagogy. Their analyses demonstrate the way that youth cultures shape, and are shaped by, what happens in the classroom, and points to both the importance of and inherent complications in working with youth to become critical consumers of popular culture as part of the formal curriculum. Their work supports themselves become critical consumers of the media.

Other scholars have started outside of the classroom, considering the pedagogical influence new technologies have on youth today, and wondering about the ways these technologies might be appropriated as part of school curriculum (e.g., Knobel and Lankshear, 2007). Gee (2007), for example, considers the relationship between video games, pleasure, and learning. He argues that good video games act as technologies for cultivating learning by tapping into a source of pleasure in students. He claims that since these games allow students to have meaningful learning experiences in a visual and embodied way, they might provide clues for ways to learning in the future.

Interactions between Popular Culture and Curriculum

A second body of work explores, in the tradition of Willis (1977), how popular culture is taken up and used by youth in (re)writing the formal and informal curriculum of school. Over the years, Dyson has argued for the permeability of the classroom writing curriculum to include the linguistic and symbolic tools that children appropriate

from popular culture and through their social relations with peers. She claims that young children come to school with the experience and ability to manipulate a broad range of popular or media texts. They incorporate these texts into their play; this experience shapes their participation in school activities (Dyson, 2003). Young children take up compelling images and ideas from popular culture texts and transform the school curriculum through their writing, talk, and play. An important contribution Dyson makes is her documentation of the way that youth use media to support their official literacy practices through fostering an interplay between the unofficial and official worlds of school.

While Dyson's work demonstrates the way that youth's knowledge and use of popular culture can (re)shape the official school curriculum, Finders (1996) illustrates how a group of adolescent girls use the acquisition and reading of teen magazines, or zines, as a cultural practice to mark their status on the path to adulthood, distinguishing them from their peers. The girls further use zines to establish the boundaries of their friendship networks and to ascribe status within the group. The study demonstrates the power of popular culture in shaping youth identities, illustrating how the girls appropriated the words and experiences in the zines as their own. Finders argues for the importance of helping youth to develop skills to examine and critique the social scripts that undergird popular culture texts.

Dimitriadis (2001: 29) points out the "complex, emergent and messy relationships" young people often have with popular culture, detailing how two teenage boys use popular culture texts to construct a sense of place. Southern US rap music provides the boys, living in a Midwestern town, with the ability to index and feel connected to their Southern heritage. Dimitriadis's account provides a strong critique of textual analysis, highlighting the importance of locating popular cultural forms within their social context. Although textual analytic approaches might suggest that rap represents a degradation of values or tradition, these boys employ rap music texts to engender a sense of connection to their heritage and to the strong values of their elders. The boys use the rap music to forge connections that are meaningful to them, thus shaping their relationship to learning and the curriculum.

In her ethnography of a multiracial high school in South Africa, Dolby (2001) details the interrelationships between race, popular culture, and the public sphere of school. She demonstrates how students construct racial identities and boundaries through a discourse of taste. Popular culture acts as an alternative curriculum among the youth in her study, proving to be a rich space for identity formation. Youth use time in school to select and employ popular practices that reflect their taste, particularly with respect to clothes, music, and clubs, to both reproduce their positions within racial hierarchies, and simultaneously challenge those positions. Focusing on

the social curriculum, Dolby advocates exploring the ways in which difference is constructed and maintained in schools, critically considering the way popular cultural forms shape these processes.

In recent years, countries around the world have adopted core or mandated curriculum, often connected to high-stakes testing. From different angles, both these groups of studies raise questions about the decision to move in this direction. The first group suggests integrating popular culture into the school curriculum. This research argues that rather than conceptualizing curriculum as a fixed body of material, it can be understood as a fluid construct that reflects and provides opportunities to build on youth's knowledge and experience to further school success. The second group of studies suggests that the core curriculum is always and inevitably (re)shaped by youth's common culture, emphasizing the social nature of learning. In other words, whatever the official curriculum might be, there are no standard or predetermined effects given students' social interactions, and their particular interests, knowledge, and desires.

Future Directions

Moss (2001) argues that school knowledge and curriculum belong to a different age, especially when compared with the worlds inhabited by youth in which vast quantities of information and materials are available for symbolic work (Willis, 1990). She explains that school contexts tend to be vertical discourses that are ordered, explicit, and hierarchically organized, while everyday discourses are often local, context dependent, and multilayered. Moss suggests makes the transfer from one context to another difficult, at best. She warns against thinking that literacy practices from outside of school, such as viewing horror movies, can be simply transposed into school knowledge. Pennycook (2005) argues, further, that youth are engaged in learning across space and time, no longer bounded by the classroom and the school day. He suggests that youth across the world are involved in a wide range of transcultural, technological practices. Borrowing from Appadurai (1996), he suggests that teaching hip-hop with the flow "...suggests not so much an incorporation of hip-hop texts into the curriculum but an opening up of possible languages and identities, an engagement with multiple ways of speaking, being and learning" (p. 40). New conceptions of learning and engagement made possible by digital media and transnational networks suggest that current understandings and uses of curriculum are outdated. We need new ways of thinking about what curriculum is, and what it might be. Conceptions about the relationship between popular culture and curriculum must be recalibrated to reflect these understandings.

When curriculum is conceptualized as the content and form of learning in and out of school, erasing those boundaries which are often crossed in a digital age, the question of whether popular culture should be incorporated into school curriculum is transposed into the following questions, among others: What is the curriculum that youth are currently engaged in across a range of contexts? How is popular culture interwoven into this curriculum? How do these conceptions of curriculum and learning shift our understandings about the purposes of school? What are the opportunities and dangers of reconceptualizing curriculum, learning, and schooling in this manner?

Much of the current research is geographically and temporally located, either in schools and classrooms or out of school and in the community, and either during or after the school day. We suggest that researchers locate their study of popular culture and curriculum at the intersection of these locations so that constructs such as learning, engagement, authorship, identity(ies), and agency(ies) can be interrogated across space and time. Appadurai (1996) has introduced the notion of global flows to suggest that people, ideologies, images, and texts are always in constant motion. They are simultaneously local and global. This notion of the flow of information and people highlights the importance of considering the role of popular culture in shaping school curriculum. As Pennycook (2005) explains, if we locate classrooms within global transcultural flows, then schools are no longer seen as bounded sites, students do not enter from fixed positions carrying identities tied to local contexts, and curriculum is no longer seen as a static body of knowledge.

Buckingham (2003) suggests that contrary to past images of children as vulnerable targets of the media, today they can be more accurately understood as savvy consumers and producers of knowledge that may not be understood by adults. A focus on the shifting practices, relations, and identities of youth pushes us to rethink the functions of school curriculum in learning. Much of the research on popular culture and curriculum has focused on youth as consumers and producers of popular culture (e.g., Buckingham, 2003; Dimitriadis, 2001; Dolby, 2001; Dyson, 2003; Willis, 1990), teachers as critical pedagogues who incorporate popular culture into their teaching (e.g., Hill, in press), or texts (Giroux and Simon, 1989). Future research on popular culture and curriculum should be located at the intersections of youth practices, pedagogy, and texts or media. Recent work has examined youth productions outside of school pointing to new possibilities for the construction of identities through spoken word performances, written narratives, visual media, musical compositions, and animations, with a focus on fostering agency through a range of semiotic resources (Hull and James, 2007). These productions, with their emphasis on performativity, suggest new possibilities for the creation of spaces where youth can take on or perform new identities.

Recently, on the first day of school in an under-resourced district, one of us heard a US teacher ask her elementary students to write ideas on two note cards. On one note card they were asked to write the reasons why they would make Annual Yearly Progress (AYP) this year and on the other why they would not meet AYP. (AYP is the metric in the US established by the *No Child Left Behind* legislation that measures progress against local and national targets.) In a single assignment, the goals of the curriculum, and ultimately the purposes of schooling for the year, were reduced to meeting specific testing goals. Versions of these conversations are enacted each day across the US and throughout much of the world. In this climate, where AYP is held up as the standard for student achievement, it is difficult to envision a role for popular culture in the school curriculum. On the other hand, if one starts from the assumption that learning and education cannot be contained within school walls, as suggested by the research reviewed in this article, the questions become how – rather than whether – popular culture acts as curriculum in youth's lives, and how learning can be conceptualized as extending beyond the school walls and the school day. Understanding the changing and evolving conceptions of popular culture and curriculum is a first step in this direction

Bibliography

- Alvermann, D. E., Moon, J. S., and Hagoood, M. C. (1999). *Popular Culture in the Classroom: Teaching and Researching Critical Media Literacy*. Newark, DE: International Reading Association.
- Appadurai, A. (1996). *Modernity at Large: Cultural Dimensions of Globalization*. Minneapolis, MN: University of Minnesota Press.
- Apple, M. (1979). *Ideology and Curriculum*. New York: Routledge.
- Buckingham, D. (2003). *Media Education: Literacy, Learning and Contemporary Culture*. Malden, MA: Blackwell.
- Dimitriadis, G. (2001). In the clique: Popular culture, constructions of place, and the everyday lives of urban youth. *Anthropology and Education Quarterly* 32, 29–51.
- Dolby, N. (2001). *Constructing Race: Youth, Identity, and Popular Culture in South Africa*. Albany, NY: SUNY Press.
- Dyson, A. H. (2003). "Welcome to the Jam": Popular culture, school literacy, and the making of childhoods. *Harvard Educational Review* 73, 328–361.
- Finders, M. J. (1996). Queens and teen zines: Early adolescent females reading their way toward adulthood. *Anthropology and Education Quarterly* 27, 71–89.
- Gee, J. P. (2007). Pleasure, learning, video games, and life: The projective stance. In Knobel, M. and Lankshear, C. (eds.) *A New Literacy Sampler*, pp 95–113. New York: Peter Lang.
- Giroux, H. and Simon, R. (1989). *Popular Culture: Schooling and Everyday Life*. Granby, MA: Bergin and Garvey.
- Hill, M. L. (2009). *Beats, Rhymes, and Classroom Life: Hip-Hop Pedagogy and the Politics of Identity*. New York: Teachers College Press.
- Hull, G. and James, M. (2007). Geographies of hope: A study of urban landscapes and a university–community collaborative. In O'Neill, P. (ed.) *Blurring Boundaries: Developing Writers, Researchers, and Teachers. A Tribute to William L. Smith*; pp 255–89. Cresskill, NJ: Hampton.
- Hull, G. and Schultz, K. (2002). *School's Out: Bridging Out of School Literacies with Classroom Practice*. New York: Teacher's College Press.

- Kellner, D. (1995). *Media Culture: Cultural Studies, Identity and Politics between the Modern and the Postmodern*. New York: Routledge.
- Knobel, M. and Lankshear, C. (eds.) (2007). *A New Literacies Sampler*. New York: Peter Lang.
- Luke, A. (2006). Teaching after the market: From commodity to cosmopolitan. In Weis, L., McCarthy, C., and Dimitriadis, G. (eds.) *Ideology, Curriculum, and the New Sociology of Education: Revisiting the Work of Michael Apple*, pp 115–141. New York: Routledge.
- McRobbie, A. (1994). *Postmodernism and Popular Culture*. New York: Routledge.
- Moss, G. (2001). On literacy and the social organization of knowledge inside and outside of school. *Language and Education* **15**, 146–161.
- Pennycook, A. (2005). Teaching with the flow: Fixity and fluidity in education. *Asia Pacific Journal of Education* **25**, 29–43.
- Vasudevan, L., Schultz, K., and Bateman, J. (under review). *Beyond the Printed Page: Multimodal Storytelling with Urban Adolescents*.
- Willis, P. (1977). *Learning to Labor: How Working Class Kids Get Working Class Jobs*. New York: Columbia University Press.
- Willis, P. (1990). *Common Culture: Symbolic Work at Play in the Everyday Cultures of the Young*. Boulder, CO: Westview.
- Fisher-Keller, J. (1997). Everyday learning about identities among young adolescents in television culture. *Anthropology and Education Quarterly* **28**, 467–492.
- Gee, J. P. (2003). *What Video Games Have to Teach Us About Learning and Literacy*. New York: Palgrave Macmillan.
- Giroux, H. (1994). Doing cultural studies: Youth and the challenge of pedagogy. *Harvard Educational Review* **64**, 278–308.
- Grossberg, L., Nelson, C., and Treichler, P. (eds.) (1992). *Cultural Studies*. New York: Routledge.
- Hall, S. (1981). Notes on deconstructing the popular. In Samuel, R. (ed.) *People's History and Socialist Theory*, pp 227–240. London: Routledge.
- Hebdige, R. (1981). *Subculture: The Meaning of Style*. New York: Routledge.
- Walkerdine, V. (1997). *Daddy's Girl: Young Girls and Popular Culture*. Cambridge, MA: Harvard University Press.

Relevant Websites

- <http://oaklanddusty.org> – DUSTY (Digital Underground Storytelling for Youth), University of California, Berkeley, LA.
- <http://www.mediaawareness.ca> – Media Awareness Network, Ottawa, Canada.
- <http://www.mediaed.org> – Media Education Foundation, Northampton, MA.
- <http://mypopstudio.com> – My Pop Studio, Temple University.
- <http://www.thebeatwithin.org> – The Beat Within: A Weekly Publication of Writing and Art from the Inside, San Francisco, CA.
- <http://www.youthradio.org> – Youth Radio, Oakland, CA.
- Best, A. L. (2000). *Prom Night: Youth, Schools, and Popular Culture*. New York: Routledge.
- Buckingham, D. and Sefton-Green, J. (1994). *Cultural Studies Goes to School: Reading and Teaching Popular Media*. Bristol, PA: Taylor and Francis.
- Das, P. T. and Weaver, J. A. (1999). *Popular Culture and Critical Pedagogy: Reading, Constructing, Connecting*. New York: Garland.

CURRICULUM DEVELOPMENT – EVALUATION AND RESEARCH

Contents

An Overview of Research in Curriculum Inquiry

Curriculum Evaluation: Approaches and Methodologies

Curriculum Studies, Discourse Analysis, and the Construction of Historical Time

Curriculum Use in the Classroom

The Roots and Routes of Teacher-Based Action Research and Curriculum Inquiry: An Historical Perspective

An Overview of Research in Curriculum Inquiry

F M Connelly, University of Toronto, Toronto, ON, Canada

S J Xu, University of Windsor, Windsor, ON, Canada; National Research Center for Foreign Language Education, Beijing Foreign Studies University, China

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This is an opportune time for an overview of research in curriculum inquiry. The last two major assessments of the field took place in 1992 with the publication of the American Educational Research Association's *Handbook of Research on Curriculum* (Jackson, 1992), and in 1994 with the second edition of this encyclopedia. The follow-up to the Jackson handbook is due for publication in late 2007 with the *Sage Handbook of Curriculum and Instruction* (HCI) (Connelly et al., 2008), and this edition of the encyclopedia will appear not long thereafter continuing the symmetry between the handbook and the encyclopedia.

Many of the trends and developments seen in these earlier documents continued in the interim and are more strongly in evidence as this encyclopedia is written. Some of these trends have been important enough to convey altered views and structures in the field. In important respects research in curriculum inquiry, though continuous with its past, looks very different now than it did 15 years ago. New curriculum subject matters, and ways of thinking about them, have emerged; there has been a (1) massive expansion in the range of curriculum topics and preoccupations studied and (2) methodological and theoretical expansion paralleling postmodern philosophical developments. Moreover, there is an increased awareness of curriculum inquiry in international contexts and some see the field as being on the cusp of a more international and cross-cultural approach to curriculum studies which tends to be parochially North American in outlook (Hameyers, 1991). This article sums up these developments, raises some of the ongoing large questions in the

field, and relates curriculum inquiry to current views of the structure of the field. It begins with an introduction to definitions and meanings, and general observations on the field that make a difference to inquiry.

Meanings/Definition of Terms Internationally

Curriculum, and curriculum inquiry, have been commonplace educational terms in North America for over a century. Dewey's (1902) *The Child and the Curriculum* is probably the single most important document in the term's history, though Hamilton (1989) traces the origins to nineteenth-century Europe. Curriculum and instruction departments have been common in North American teachers' colleges and faculties of education for many years. These departments are often responsible for, or associated with, teacher education with the result that curriculum inquiry has strong overlapping links with instruction and teaching methods inquiry.

The definition of curriculum, and therefore of curriculum inquiry, has long been contentious, closely linked to philosophical, conceptual, and ideological views on the purpose of education. Multiple definitions were shown to be possible based on Schwab's (1960) commonplaces of curriculum. For Schwab, depending on the times and the circumstances, the appropriate focus for curriculum concern is one of four factors or commonplaces – student, teacher, subject matter, and society. With each reshuffling of the relationship of these commonplaces, a different definition of curriculum emerges.

For instance definitions of curriculum, and of curriculum inquiry, derived from taking the learner as the starting point differ from definitions, and from curriculum inquiry, which start with social purposes. The most comprehensive account of the definitional issue was provided by Jackson. Following a survey of the variability of definitions, Jackson concluded that an argument about what was worthwhile, and a definition, went hand in hand. For him, a change in the purpose or argument would result in a change in the definition, and hence in the forms of inquiry. Connelly and Xu (2008) adopt this notion for the HCI, adding to it the idea of practice. In that publication we argue that definitions, arguments, and practices are interwoven. The implication of adding a notion of practice is that definitions are never fully determined theoretically, logically, or linguistically, but are driven by practical school-based problems under investigation in inquiry. Thus, what is studied affects the inquiry.

Beyond these complications for inquiry arising out of the multiplicity of possible definitions of curriculum, and hence curriculum inquiry, is the fact that the term curriculum is not used universally throughout the world. Some countries use pedagogy, others didaktik, others have no counterpart to curriculum, and still others such as China appear to be borrowing North American curriculum terms. Comparing curriculum is not simply a question of translating terms such as pedagogy into curriculum because when a term like pedagogy is used it tends to imply a different configuration and range of considerations than does the term curriculum, similarly with the notion of didaktik. One of the issues, discussed below and foreshadowed here, that emerges from this is that as educational thought and curriculum thought become more international, there is, as Anderson-Levitt documents in her HCI chapter, much room for miscommunication. Moreover, as she goes on to show, there is the potential for a kind of North American curriculum colonization: the borrowing and using of North American curriculum terms and their uses in international contexts. Dialog across cultures is important, but dialog is not enough if it is carried out in only one dominant cultural voice and set of terms. As education, and curriculum inquiry develops internationally, more cross-culturally aware frameworks for reviewing research than exist now will be required.

Characteristics of the Field and of Curriculum Inquiry

The overall nature of curriculum inquiry reflects the place of curriculum studies in public discourse about education and in the relationship of curriculum to the broader field of education. Curriculum discourse cannot be cleanly separated from educational discourse more generally. Almost all public discussion of education either

hinges on curriculum or is essentially a curriculum topic. For example, discussions over the purposes of education, achievement, and accountability, while broad and educational in scope, are also fundamentally curricular. This interdependence of curriculum with the broader field of education and with political/social discourse complicates the question of what is and what is not curriculum inquiry. A study conducted in psychology, for example, might be a purely psychological matter or it might be a curricular matter if, for example, questions of subject matter, content, and/or methods of teaching and learning were included. This ambiguity has led to a situation where a great deal of curriculum research is carried out under a different name. Both the Jackson (1992) handbook and the upcoming Connelly (in press) handbook contain a wide range of curriculum inquiry research into topics and preoccupations by people who are not primarily known as curriculum writers and yet who do curriculum inquiry. Thus, if one were to review only the curriculum inquiry journals, one would miss much curriculum inquiry which often appears in other places. Part of the reason for this situation is that, in North America, curriculum theory has turned away from curriculum practice to textual analysis and to theoretical sources in the arts, humanities, and social sciences (Westbury, 2007). North American curriculum theorists tend not to be involved with practical, curricular research. People doing such research have tended to migrate to other subfields defined by newly emerging curriculum topics and preoccupations, and into other academic departments.

Structure(s) of the Field and Curriculum Inquiry

The definitional variability, and the various curriculum starting points that link curriculum inquiry with educational inquiry more generally, result in different working structures within the field. These structures, in turn, provide different frames for organizing curriculum inquiry. In the following two main ways of thinking about the structure of the field are described: (1) practice–context–theory, and (2) subject matter, topics and preoccupations, and curriculum theory/general curriculum. Subsequently, five developments emerging in the last 10–20 years related to these structures, as a basis for providing an overview picture of curriculum inquiry, are described.

Practice–Context–Theory

Jackson's review of curriculum inquiry led him to the view that the field was in structural disarray. His essay remains important reading because it documents the division between theoretical inquiry and practical inquiry

in curriculum studies. Though this dispute and separation has not disappeared, its impact has been mitigated (Connelly and Xu, 2008). Ironically, perhaps, context is essentially omitted in the theory/practice inquiry dispute even though context is central to all curricular. It is one thing to bridge the theory/practice argument, and gaps between theory and practice, by ensuring that practical curriculum inquiry be embedded in, or develop, appropriate theory; and conversely that theoretical inquiry be grounded in the real world of curriculum practice. However, context adds a new dimension to each. The contextual idea is that nothing practical happens in isolation but always happens in relation to people, places, and things; in relation to organizational structures, policies, political discourse, and ideological frames. Taking these into account makes for good curriculum inquiry. Likewise, curriculum theory needs to account for curriculum in context, not merely curriculum in practice. Moreover, curriculum theory is not derived merely from theoretical, postmodern abstract thought but, itself, takes place in the same contextual frames that surround the practice of curriculum. Theories and theoretical thinking in curriculum are contextual and dependent. They are not, as some might have it, abstract and universal.

Context is also the starting point that is, the problem or topic, of much curriculum inquiry. This point is discussed later in the section on developments in the last 10–20 years. For now it is sufficient to note that context is not merely a surrounding frame for practice and for theory but is itself a source of inquiry. Thus, within this structure of the field there are practical, contextual, and theoretical curriculum inquiries, all of which, depending on the inquiry, are related in different ways to one another.

One further point relevant to the overall frame provided by this encyclopedia is that while subject-matter inquiries are not part of the assigned topic for this section, subject matters are an important part of the practice of curriculum and, as Deng and Luke in their HCI chapter show, curriculum subject matter is currently being studied in terms of changing political and economic demands. In other words, subject matter is being studied in context.

Subject Matter, Topics and Preoccupations, and Curriculum Theory/General Curriculum

Readers will readily see the relationship between the just-discussed structure of practice–context–theory, and subject matters–topics and preoccupations–general curriculum. The former is a more formal structure while that discussed in this section, though closely related, refers directly to the way curriculum and instruction departments are organized, offer their courses, and conduct their professional business through conferences, journals, and the like.

Subject Matter

The most obvious feature of curriculum inquiry for a casual viewer of the field is subject matter. Parental and public concern is with matters such as reading, language, science, mathematics, the arts, and so forth. School curriculum is basically structured around curriculum subject matters with corresponding school board, province, and national policies in the form of guidelines, courses of study and policy memoranda. Though no quantitative accounting is available to our knowledge, there can be little doubt that the amount of research in this area equals, and likely exceeds, the total of research in the other two areas. Moreover, the most high-profile public/political discourse is most often about subject matter, for example, national literacy, mathematics achievement, reading capability, and so forth. This discussion is often associated with national policies as, for example, the American No Child Left Behind policy and, these, in turn, are related to a host of curriculum inquiry studies ranging from what the policy means, to how it may be implemented, what kinds of achievements are accomplished, and critical/theoretical inquiries into the social/political nature of the policies and programs.

It is worth noting that modern curriculum inquiry in subject matter goes beyond the methods research common in the past and takes on contextual and philosophical qualities. Chang and Rosiek's (2003) *Anti-Colonial Antinomies in a Biology Lesson: A Sonata-Form Case Study of Cultural Conflict in a Science Classroom* is illustrative. The study explores the cultural, philosophical, and subject-matter content tensions confronted by a Hmong immigrant teacher and his Hmong students in an American science classroom. The study of subject matter is conducted in cultural context and is a contribution to the contextual topic of culturally relevant pedagogy. Moreover, as the title indicates, postmodern narrative and arts-based research methodologies are used.

Topics and Preoccupations

The rapid growth of contextually aware curriculum inquiry in the past few decades has created an unrecognizable (at least for some) landscape of curriculum inquiry. However, a review of some of the topics and preoccupations in the HCI reveals a vast body of work. The problem is that some writers conflate general curriculum/curriculum theory with the broader field of curriculum inquiry. Thus, what is said of curriculum theory is often incorrectly thought to apply to curriculum inquiry as a whole.

Curriculum topics were not mentioned in earlier encyclopedic compilations. It would be a mistake to think that because they were not mentioned there has been a major turning point in the last 10–15 years. Everything in educational inquiry is continuous with its past and, although

the present, may, at times, appear startlingly new, careful thought reveals historical narratives. Multicultural education, or at least that part of it that might be properly called curriculum inquiry, is perhaps the most widely developed and researched curriculum topic. Though it might appear that this is new because of the increased attention given to it worldwide, for example, ethnic studies in Asia and Europe, this topic has a long history. Banks and Banks (2004) trace the origins of multicultural education to African-American scholarship in the nineteenth-century and to the curriculum reforms of the 1930s in North America. The *Handbook of Research on Multicultural Education* (Banks and Banks, 2004) has 49 chapters and is over 1000 pages in length. Not all of the titles are explicitly curricular, but many are, and those that are not are directly relevant to curriculum.

The curriculum topic, multicultural education, reveals many of the developing characteristics and qualities of curriculum inquiry found in other topics and preoccupations such as textbook analysis, anti-racism, gender, achievement, and equity. Foremost among these qualities is the fact that research in the topics and preoccupations explicitly begins in practice. For some, saying that research is grounded in practice may appear to be stating the obvious but, as Westbury (2007) notes, North American general curriculum studies have taken a turn toward texts. By this he means that curriculum writing and curriculum theory has a hermeneutic quality involved with the analysis of written texts and with the abstract application of theoretical positions to practical topics. Starting points for inquiry tend not to be real world practices but, rather, other texts and theory often originate outside of education.

Literature reviews of curriculum inquiry in the HCI illustrate the notion of contextual curriculum inquiry in the topics and pre-organization. For instance, Ladson-Billings and Brown's chapter on the topic of curriculum and cultural diversity begins with the practical problem of racism. Context is central to their review of curriculum inquiry and they discuss the relationship of selected social developments and their relationship to racism. Moreover, their work is strongly theoretical as they explore critical race theory to help them understand their topic. Their review of the literature demonstrates the in-between (Connelly and Xu, 2008) quality of curriculum inquiry in the topics and preoccupations. The literature on which they draw – and the way in which they review it – is in-between practice and theory. Practice is the starting point: theory is drawn upon to help account for the practices in question: and context is an important interpretive frame for understanding the topic in both its practical and theoretical aspects.

We draw attention to this in-between quality of curriculum inquiry in the topics and preoccupations because, while it may seem obvious to some, Jackson demonstrated

that general curriculum inquiry was split on the question of theory and practice. On the one hand, people who did practical work were often judged by their more theoretically oriented colleagues as pursuing merely practical, status quo research, while those pursuing theory for the sake of theoretical understanding were, according to Schwab (1983), doing the wrong thing in curriculum. Jackson's assessment at that time was that curriculum inquiry was splitting up in two directions, researchers aiming toward practice and those aiming toward theory. The current view as expressed in the HCI is that there is a vast in-between literature of topics and preoccupations in which practice and theory are brought into dialectical relationships in contextual frames.

To give one more example, He, Phillion, Chan, and Xu's HCI chapter on immigrant students' experience begins with a contextual, demographic, account of the world's changing multicultural and multilingual landscape and connects this to the everyday experience of education by immigrant students. In their summary of curriculum inquiry, context is not only used for justification of the significance of the topic for inquiry, and it is also the part of the topic as defined. This is so because culture is one of the central terms in their topic, and culture is part of the language of context. As their account develops, the interdependence of practice and context is shown as immigrant students' experiential problems in the school system are interpreted in contextual terms. Cultural content is seen as both the source of, and solution to, those experiential issues. In order to provide theoretical insight, the authors use a variety of theoretical frames, a practical process Schwab called the eclectic. They draw on key terms related to language, culture, identity, and power and bring forward theoretical resources in each; and they describe two main lines of curriculum inquiry, ethnography and narrative inquiry.

Their account conveys our opening remarks on the intimate relationship between curriculum inquiry and educational inquiry more generally. The theoretical resources for understanding and improving the educational lives of immigrant children are broad and are found throughout the field of education. Furthermore, the topic is linked to transmigration social patterns and associated educational matters.

To sum up this section on topics and preoccupations, it is important to note that it has not summarized the list of possible topics and preoccupations. The purpose here is to show the significance of this broad area in curriculum inquiry for understanding the nature of curriculum inquiry and its current directions. This work is closely linked to new directions for curriculum inquiry into subject matter. There are similar curriculum inquiry qualities between context-oriented subject-matter studies, and the in-between topics and preoccupations. Context and practice are increasingly interwoven, and research in these areas is increasingly far reaching and eclectic.

Curriculum Theory/General Curriculum

Curriculum theory is one of the more contested areas in curriculum inquiry. Schubert's (1986) scan of the field remains a useful overall view of curriculum studies. Slattery (2006) recently wrote that "curriculum theory has many hostile and competitive factions at odds with one another" (p. 193). Unlike the case in the physical sciences, for example, where competing theories such as the wave and particle theories of light are debated in terms of their explanatory power relative to observable phenomena, disagreements over appropriate curriculum theories tend to relate to ideological positions and to be settled politically as followers of one or the other ideology gain credence. Westbury (2007) traces developments in what might be called the politics of curriculum inquiry for the United States and Europe.

Furthermore, the purpose of the act of theorizing – the value of theory itself – is in question in curriculum inquiry in a way that it is not in the physical sciences. There is little dispute in the physical sciences over the ultimate purpose of theory, namely, to help account for the phenomenal world. But in curriculum inquiry, the very purpose of theorizing itself is in question. In a series of articles culminating in an agenda of work for curriculum professors, Schwab (1983) criticized the enterprise of curriculum theorizing arguing that describing and critiquing what went on in schools is the proper task of curriculum people. This was influential work. Levine (2006) and Reid (1999) believe that curriculum theorists would nominate Schwab to a list of the five most influential scholars in their field in the twentieth century. Schwab criticized curriculum theory's purpose rather than contesting any one of its claims. In Schwab's view curriculum theorists had abandoned curriculum phenomena as their starting point in favor of texts and theories external to curriculum and to education. He labeled the enterprise of curriculum theory as exhibiting a "flight from the field," meaning that it was talk, and talk about talk, not talk about curriculum as practiced in schools, and other places.

Schwab proposed that the principal purpose of theorizing in curriculum inquiry should be to describe and account for curriculum phenomena as experienced. He also argued for a critical role for curriculum inquiry but his idea of criticism was one founded in observation more so than on theoretical application. Again, it is a question of starting point. The basis for Schwab's notion of criticism was description of what was the case, rather than imagined scenarios derived from theory. Wraga and Hlebowitsh (2003) reapplied Schwab's analysis of theory in curriculum inquiry with similar results to Schwab's. In the debate that followed the central point about the proposed practical purpose of curriculum theory was mostly in the background. According to Doyle (1977), Schwab's argument and purpose lost out in the politics of curriculum

inquiry in North America to the reconceptualists (Pinar *et al.*, 1995). In the second edition of this Encyclopedia Goodlad (1994) remarked "the critical theorists have a tendency to stop with criticism; the reconceptualists continued their dialogue; there was little movement . . . into the study of ongoing practices recommended by Schwab" (pp. 1264–1265).

Connelly and Xu (2007) have argued that though the intensity of debate over the purpose of curriculum theory has waned, its significance has not. Westbury (2007) writes that the move to curriculum theory as text is a wrong turn. He contrasts the purpose in this turn with European work (e.g., Dahllöf, 1971; Lundgren, 1972) which is more closely akin to Schwab's proposed purpose. There were, of course, elements of both purposes of curriculum inquiry in North America and Europe. Basically, since curriculum theory has been primarily a North American enterprise, summations of it have been mostly from an American perspective. There is important curriculum inquiry work to be done in examining the nature of curriculum and curriculum theory internationally and cross-culturally, reaching beyond America and Europe.

The dichotomous picture of the enterprise of theorizing in curriculum inquiry found in Jackson, Doyle, Westbury, and Pinar overstates the case. To begin with, too little is known about the role of theory in curriculum theory outside North America and parts of Europe. It is possible, but seems unlikely, that the characteristics and history of theorizing in curriculum inquiry offered by American writers characterizes the rest of the world. But even in North America the dichotomous characterization and the question of which purpose is winning politically is not as clear as suggested by some. Westbury, for instance, modifies his own position by pointing to a list of works that fit Schwab's purpose for curriculum inquiry, that of providing descriptions of schools and school learning. Some, but not all, of these writers cite Schwab. More important, perhaps, as we have argued (Connelly and Xu, 2007) is the fact that in addition to Schwab's flights from the field, many curriculum scholars appear to have moved to the sidelines, and sometimes even out of curriculum. Thus, their work went unnoticed by those writing and debating curriculum theory.

However, they were working in Schwab's practical tradition. Writing in the HCI, Craig and Ross, in the most definitive current statement on the influence of Schwab on curriculum inquiry, catalog a large body of research explicitly drawing on Schwab, almost none of it cited or noted in the curriculum inquiry theory literature. In effect, these are writers who adopt the practical purpose for theory in curriculum inquiry but who, in a move to the sidelines, avoid the disputes, debates, and politicizing that takes place at the level of arguments over the correct theoretical path to follow. Craig and Ross bring

forward an extensive literature and present, in detail, programs of research associated with the arts (Eisner), leadership in higher education (Fox, Scheffler, and Maron), pedagogical content knowledge (Shulman), and teachers as curriculum makers (Clandinin and Connelly); and they describe six broad areas where Schwab's work has been significant: study of teacher education and teaching practices, action research, portfolio development, narrative practices, interdisciplinary efforts, teacher groups, and teachers helping teachers. In addition to the body of work explicitly drawing on Schwab's argument on the purpose of theorizing in curriculum inquiry, much, if not most of the work described in this essay on the topics and preoccupations, while not necessarily drawing on Schwab, fulfills a practical purpose for theorizing in curriculum inquiry.

The overall picture that emerges is that most curriculum inquiry is, in fact, practical in the sense proposed by Schwab. There are lines of research explicitly drawing on Schwab and there are large bodies of literature in the topics and preoccupations which begin in practice and where theory plays a practical role. The practical potential of theorizing in curriculum inquiry is mainly unrealized and is generally agreed to have had little impact (Doll, 2002). Rethinking the directions for theorizing in curriculum theory is taking place, a point evident in HCI essays by Enns, Hansen, Levin, Pinar, Schubert, and Short. For instance, Pinar references several theorizing strands aimed at practice – multicultural curriculum theory, curriculum theory and gender, phenomenological curriculum inquiry, esthetic curriculum inquiry, and theological curriculum inquiry. Moreover, Pinar's own attentions have recently been given over to international matters in curriculum inquiry. In a similar vein, Schubert refers to the grievous error of separating form and substance in curriculum inquiry. Adopting Schwab's practical term eclecticism, Schubert describes the integration of form and substance in nine curriculum inquiry areas: practical inquiry, curriculum evaluation, existentialist perspectives, hidden curriculum, critical theory, counter-culture teachers, teacher action research, reconceptualist theorizing, and curriculum history. This list of curriculum inquiry areas represents the first of two categories presented by Schubert: emergent eclecticism, which he calls neo-Deweyan inquiry, and contemporary venues. Eight contemporary venues are listed: intended curriculum, taught curriculum, experienced curriculum, embodied curriculum, hidden curriculum, tested curriculum, null curriculum, and outside curriculum.

To sum up this section, theorists in curriculum inquiry appear to be making attempts to minimize the abstract, textual, and distant-from-practice purpose of studying theory in curriculum inquiry. Several of Pinar's examples fit the area we call topics and preoccupations, and Schubert forthrightly argues against mere theorizing, arguing for, and illustrating, inquiries integrated in theory and practice.

Though context as such does not loom large in either of these two comprehensive overviews of theory in curriculum inquiry, context is playing an increasingly significant role in theorizing in curriculum inquiry.

Curriculum Inquiry Developments during the Last 10–20 Years

Our discussion of the structures of the field and of curriculum inquiry has touched upon ongoing trends and developments in curriculum inquiry since the publication of the last edition of this encyclopedia. The following picks up on these developments and briefly discusses five: expression/expansion of context in curriculum inquiry; curriculum subject-matter developments; movements to international curriculum inquiry; development of a theoretical curriculum inquiry literature; and a focus on practice and a move toward integration of theory and practice in curriculum inquiry. What is described herein is not so much new as it is an overt expression of factors, forces, and work long underway. There is a tendency in the general curriculum literature, to view the field cataclysmically as if there were reconstructions, new moments, paradigmatic shifts, renaissances, movements, and so forth. These ways of viewing the field are partially political, aimed at highlighting a pattern of inquiry of ideological orientation over another. In these cataclysmic accounts of the field, writers, and lines of work, that may have at one time been important and visible, may be obscured and lost. The HCI is built on the idea of continuity such that people and ideas are thought of as part of a narrative history. Such an interpretation is conveyed here. The following five developments are intended to be read in the spirit of continuity. They are not brand new ideas created *de novo*, conceptualized out of nothing. They are expressions of work well underway in the last edition of this encyclopedia, and before.

Expression/Expansion of Context in Curriculum Inquiry

One of the curiosities of the curriculum inquiry literature is the extent to which it is often labeled as being dysfunctional, while there has been steady growth of work on the curriculum in context. Jackson's answer to his question "How do curriculum matters look at a distance?" (p. 3) was "confusion." He cited 17 authors who variously described the field as in conflict, elusive, in disrepair, moribund, in disarray, suffering from severe disorientation, chaotic state of curriculum terminology, ill-defined epistemology, and great disenchantment with curriculum inquiry. Since that time the general assessment has more or less remained the same though the intensity of concern over it seems to have declined. Wraga and Hlebowitsh (2003) refer to

“a chronic state of disarray, even crisis” (p. 425). But some say that the expansion of theoretical writing overshadows the supposed chaos and is one of the principal achievements of curriculum inquiry. Those that take this view champion the idea of multiplicity of theoretical resource, seeing it as a sign of vitality (Connelly and Xu, 2008).

The attention to curriculum context, particularly in the topics and preoccupations, seems mostly to have gone unnoticed in these debates. Though unnoticed in the curriculum inquiry literature, curriculum inquiry carried out in context is even more diverse than curriculum inquiry in theory. This follows from the fact that there is not only theoretical variation but also practical and contextual starting point variation in curriculum inquiry in context. A second point to note is that though the contextual literature is more variable than the theoretical literature, there is nothing like the theoretical disputes in the contextual curriculum inquiry literature. The reasons for this are reasonably clear. First, because of the practical focus of contextual work, writers tend to have a more pragmatic approach. The question of whether or not work in the topic or preoccupation makes a difference to the practical world is what is principally at stake. Second, because curriculum inquiry in context is driven by practice, people working in different topics and preoccupations often have little collegial relationship and may not even be aware of and/or cite one another's work. Even topics that may appear similar to an outsider are often populated by people of different professional persuasions. For example, the Editorial Board for the HCI attempted to commission a chapter on ‘cultural and linguistic diversity’. Potential authors when contacted wrote that they could write on one or the other of the topics but not on both. What this means is that there tends to be little sense of community among scholars across topics and preoccupations. People may be in different departments, attend different professional meetings, write in different journals, and so forth. The advantage of this is that their preoccupation with practical matters and improving the world permits the growth of knowledge on important educational curricular problems. The downside is the lack of a sense of community. It is here that curriculum theory could play a role as it might knit together the main themes and curricular insights of the various practical and contextual curriculum inquiries.

But what is context? The simple answer is that it depends. It depends on the topic or preoccupation and the particular way questions are asked. The preoccupation with the topic of curriculum development and educational reform is illustrative. Westbury's HCI essay began by noting the international trend toward national, government directed, curriculum development. He asked why states develop curricula, why doing so is so attractive at this time in our history, and how effective can states be. For Westbury, the principal context for examining this question is the set of reasons state planners and politicians

offer for chosen curriculum directions and how state decisions are organized and conveyed publicly.

Westbury's analysis is different from that of Fullan who treats the same topic. Fullan casts the problem in terms of forces and factors for and against the state's successful implementation of policy goals. Thus, for Westbury, the context for this topic is the logic and rationale offered by policymakers, and for Fullan it is the tactics and strategies of policy implementation. Both Westbury and Fullan tackle the same broad topic with a similar sense of the overall problem, namely, that governments have taken on curriculum development responsibilities and have been notoriously ineffective in terms of impact. Both offer highly contextualized accounts. Westbury's inquiry has a somewhat abstract, analytic quality, while Fullan's inquiry tends to be more pragmatic and operational. Their contextual analyses are so different that writers in the Westbury mode tend not to cite those in the Fullan mode and vice versa, though both agree to the main problem, and both agree with the main effects.

Working in the same topic in the handbook, Apple, drawing on his influential writings on the distribution of power relations in education (e.g., Apple, 1995), tackles the same problem by starting with context; that is, he begins with power relations and works from there. For him, context is the matrix of power relations that exist in educational institutions and practices. He sees educational change in terms of shifts in power relations which result in social movements. Thus, for him, the appropriate context for understanding curriculum development is the social movements within which governments, policymakers, and curriculum developers operate. It might be said that a move through the work of Fullan, Westbury, and Apple is a move through different contextual levels such that Westbury's state-wide rationales and logical forms are the context for Fullan's tactics and strategies; and Apple's social movements and shifting power relations are the context for Westbury's analysis.

Contexts have this quality of being dependent on the way people ask questions and, somewhat like an ancient Chinese pagoda, of one strata unfolding to reveal yet another. For instance, when Anderson-Levitt asks questions about curriculum development and educational reform she does so by drawing on her own and others' (e.g., Meyer and Ramirez, 2000) work which takes educational systems in the world as context. She understands the problem of curriculum change and reform as one of globalization and internationalizing of the curriculum and she discusses processes of cross-cultural curriculum transition. Again, Anderson's topic is closely related to that of Fullan, Westbury, and Apple, but she comes at it with a different notion of context, one which might be seen as context for Apple's context.

Things, of course, are never quite as simple as suggested by the unfolding pagoda's notion of boxes within

boxes, or contexts within contexts, since contexts interact and intermingle both in the world of reality and in curriculum inquiry. The separation of contexts as presented above tends to overstate contextual independence. All deal in some fashion with the contextual matters most important to others. For example, Fullan, Westbury, and Apple acknowledge Anderson-Levitt's global context. Luke's handbook idea of eduscapes lets him name and symbolize the blurring and blending, and intermixing of contexts that marks this expanding and important broad area in curriculum inquiry.

Curriculum Subject-Matter Development

This brief section is not aimed at covering inquiry in the various curriculum subject matters. That is dealt with elsewhere in the encyclopedia. There are, however, developments in the idea of curriculum subject matter, its content, and how it is conceived. This section provides a brief sketch of these developments.

A comprehensive review of disciplinary, practical, and experiential conceptions and philosophies for the organization of curriculum subject matter is found in Deng and Luke's handbook chapter. Though much of the ground they cover has a recognizable status in curriculum inquiry, this ground has renewed significance given public and political discourse, primarily in the United States, Britain, and Australia, over achievement and the foundational role that basic disciplines may play in curriculum subject matter. Their analysis is particularly relevant for states where the main educational questions revolve around what knowledge should be taught, what level of achievement is desired, and how it is measured.

Curriculum subject matter is increasingly studied in contexts other than disciplinary origins. There are several developing contexts: cultural, classroom, pedagogical knowledge, and personal practical knowledge. School subject matter is traditionally conceptualized in terms of logical, epistemological psychological, pedagogical, and sociological matters bearing on its use in a classroom environment. Curriculum subject-matter inquiry is different from traditional ways of relating school subject matter to disciplinary knowledge, and it is different again from a cultural context approach and from pedagogical and personal practical knowledge approaches. The German *didaktik* is related to all of these contextual approaches in that its focus is on subject matter in classrooms, and teachers' and students' relationship to it.

The emphasis on disciplinary knowledge in the United States, the UK, and Australia is, in the overall scheme of things, not so much new as a recycling of intellectual traditions that come and go with the times. Mixed in with these moves, however, is the influence of technology. In her HCI chapter, Means shows how development and

research have moved from the early notions of technology as an efficient device for teaching subject matter and minimizing or even replacing teachers to what she calls a more humble approach. Not many years ago one of the mantras of curriculum inquiry and subject matters was the knowledge explosion. Now, the main impact on subject matter appears to be a tidal wave of access to information and knowledge through the Internet. Means notes that this results in access to approved standard-based curriculum materials, access to transformative curriculum materials, and access to curriculum processes such as assessment and curriculum decision making, and access to knowledge building tools. These developments have the potential to overwhelm the traditional textbook writing process which, regardless of prevailing ideologies and philosophies of knowledge, tended, at least at the secondary school level, to reduce to a process in which curriculum writers modified advanced college and university texts in various disciplines. That process is coming under siege.

One of the ways this transformation is occurring is through the production of student-oriented Wiki texts (Wikibooks, 2007). This process is only now underway. The old Dominion University approach, described in this citation, has a carefully planned research and development program over a series of years. There is an interaction between overall guidelines provided by course instructors, and texts produced by students. These student-produced texts are then used in subsequent courses, modified, and reused. There are numerous philosophical, cultural, ethical, and practical questions associated with this process. However, Wiki texts appear to hold promise and will increasingly be the topic of curriculum inquiry in the subject matters.

Movement to International Curriculum Inquiry

Curriculum inquiry has always been characterized primarily in local terms: how to influence this curriculum, in this situation in this classroom, and so forth. Local school districts were at the heart of the well known Eight Year Study. The tendency to localism in curriculum inquiry is understandable in terms of the fact that curriculum is at the payoff point of teaching and learning in educational thinking. But as shown above, one of the current marks of curriculum inquiry is its attention to context. Increasingly, one of those contexts is globalization and the internationalizing of curriculum (e.g., Meil, 2003). The HCI has a section on internationalizing the curriculum; people like Pinar are shifting their attentions from theory to international matters (which, of course, constitute theoretical issues); and the writings of anthropologists and world culture sociologists such as Anderson, Anderson-Levitt, Baker, and Meyer are filtering into curriculum studies.

Modern technology, communication, and transportation are speeding up the process as the world is increasingly thought of in terms of a global village. Historically, as Deyhle in her HCI chapter shows, curriculum has always been influenced by transmigration and by colonizing cultures. Anderson-Levitt describes anthropological and sociological world culture theorist approaches which are aimed not only at understanding curriculum in specific international cultural contexts but also at understanding how curriculum travels and is adopted and influenced as world culture changes. Westernization of curriculum is at issue in the world culture literature. Similar inquiries may come to the forefront relative to Chinese and other Asian cultures. Following her review of the literature, Anderson-Levitt concluded that there is a move to curriculum uniformity around the world. She also pointed out, however, that commonalities were most noticeable at the level of policy and curriculum materials. At the actual classroom instructional level, local adaptations and modifications tend to maintain diversity.

It appears that international curriculum inquiry may have a leavening effect on the intensity of ideological debates that frequent the literature. Anderson-Levitt, citing others, suggests that international curriculum inquiry reveals different ideological systems coming to somewhat similar practical educational ends. The opposite, of course, may also occur. Xu's (in press) work on the narrative origins of immigrant ways of knowing and being sheds light on the profound influence of historical contextual/cultural knowledge structures on how education takes place as cultures interact.

Development of a Theoretical Curriculum Inquiry Literature

As noted above, one of the main debates in curriculum inquiry has to do with the development and role of a theoretical literature. Pinar *et al.* (1995: 6) wrote that the reconceptualists moved away from practice and were "preoccupied with understanding." "Understanding," as Schwab showed, is the appropriate end in the natural sciences and other theoretical disciplines. In the HCI Pinar summarizes the reconceptualist's current purpose as: "a multi-discursive academic effort to understand curriculum: historically, politically, racially, autobiographically/biographically, esthetically, theologically, institutionally and internationally, as well as in terms of gender, phenomenology, post-modernism, and post-structuralism." The theoretical literature aimed at understanding is too massive to summarize here. In addition to Pinar's comprehensive overview, there are other summaries.

This rich and intellectually exciting literature is distant from practice and has little effect on it. Though theoretical possibilities are hardly exhausted, one might

imagine that the attraction this literature has for the self-development of its writers (Kridel, 2006) may tend, over time, to be replaced by more practically oriented inquiry. The following section explores this possibility.

A Focus on Practice and a Move Toward the Integration of Practice and Theory in Curriculum Inquiry

Schwab's argument that the aims of curriculum inquiry should be practical doings rather than theoretical understandings has been widely misinterpreted in the literature as meaning that Schwab, and those who worked in his tradition, wanted simplistic practical work on methods, curriculum development, and so forth. It is sometimes said that this practical line of work was dominated by the disciplines in such a way that conceptual, intellectual, and contextual matters were relegated to minor factors under the auspices of foundational disciplinary thinking. Ironically, however, and notwithstanding its rich, nuanced, postmodern thought, the understanding literature is more akin to the disciplinary movement than was that of Schwab and his followers. Schwab was a dialectician and proposed the interweaving of theory, practice, context, and criticism within a practical doing starting point framework. To take a simple example, his methodology for bringing practical, public, and theoretical factors to bear on curriculum situations was what he called deliberation and the eclectic application of theory. Fox (1985), a student of Schwab's, illustrated these ideas in the workings of curriculum planning teams. Aimed at practical solutions and considerations, the work of these teams was complex, requiring assessments of the practical situation involved, representation of participant voices, evaluation of possibilities among teachers, learners, subject matter and milieu, and the assessment of an eclectic array of theoretical resources. Craig and Ross in their HCI chapter document several programs of research following in the practical tradition.

Connelly and Xu's (2008) assessment of this matter based on their reading and interpretation of current reviews of the field was that when the field of curriculum inquiry is taken in its entirety, that is, subject matter, topics and preoccupations, and general curriculum/curriculum theory, there is evidence of a theory/practice dialectic taking place, primarily within context. Practical issues and problems are increasingly thought through in terms of a mix of theories, and some theoretical writings are carried out as explorations within the topics and preoccupations. Moreover, while theoretical understanding is still a goal of a small body of theoretical curriculum inquiry, the bulk of curriculum inquiry defined by subject matter and the topics and preoccupations originates in practice and in context and is aimed at improving curriculum situations.

Ongoing Puzzles for Curriculum Inquiry

Solutions and deliberative resolutions in curriculum inquiry always remain open. The same is true, but in a somewhat different sense of course, for the theoretical fields where understandings tend to pass the test of time and require major conceptual revolutions to modify. However, in practical fields such as curriculum inquiry a reconfiguration of the many factors and forces at work will catalyze an apparently settled matter. Accordingly, a very long list of ongoing curriculum inquiry questions might be drafted. Instead, named, with little discussion, are five such puzzles that, for this time, are more than ordinarily noticeable.

1. *What happens to curriculum inquiry in state policy development and implementation?* As states increasingly take on comprehensive policy formulation roles, there is concern that curriculum scholars will be marginalized. Policy formulation and implementation at the state level can be dominated by political and technical qualities. This puzzle of what role curriculum inquiry should or can play differs from state to state and, judging from the literature, is of most concern to curriculum scholars in the United States. The balance of public opinion, political policy, and curriculum inquiry in curriculum reform needs to be carefully studied worldwide.
2. *Can school reform make a difference?* The potential effects of policy reform, whether by states or by professional, academic, and other associations, shares disappointing results. Some believe that more adequate implementation technology will improve the practical impact of policy, others think that broad social effects overwhelm what might be accomplished within education, and still others believe that the purpose of state policy formulation is primarily a political act designed to assuage public opinion. This puzzle needs to be studied in different political contexts worldwide.
3. *Diversity: What is it and how to think about it in curricular terms?* The idea of diversity was, at one time, perhaps, most noticeable in curriculum under the rubric of individual differences. Diversity is now a broad postmodern notion pointing to variation and alternative possibilities in curriculum inquiry. Comparatively new diversity constructs have evolved into the broad notions of race, class, and gender. But the concept has taken on the intellectual role of a way of thinking about curriculum matters. This way of thinking has profound consequences for states, and local settings, where singular, perhaps certain, views of curriculum matters exist. Research is needed on how diversity and cultural norms interact in a world of accessible knowledge and cultural mobility.
4. *How to think about different ways of knowing and being as cultures interact?* Transmigration is reshaping the curriculum-for-immigrant-education question from

one of cultural adaptation to one of reciprocal, mutual reeducation. Curriculum inquiry is needed on cultural reciprocity and mutual learning as cultures intersect in different ways, in different places, throughout the world.

5. *How to maintain diversity and cultural equality in globalization?* This puzzle is related to the last. The puzzle has tended to be conceptualized in terms of Western curriculum colonization, but with shifting global patterns it appears the issue will take on a different complexion. Curriculum inquiry is needed on the conditions that affect diversity and equality as cultures intersect; and inquiry is needed on a global scale to track international value systems bearing on the curriculum.

Bibliography

- Apple, M. W. (1995). *Education and Power*, 2nd edn. New York: Routledge.
- Banks, J. A. and Banks, C. A. M. (eds.) (2004). *Handbook of Research on Multicultural Education*. San Francisco, CA: Jossey-Bass.
- Chang, P. J. and Rosiek, J. (2003). Anti-colonialist antinomies in a biology lesson: A sonata-form case study of cultural conflict in a science classroom. *Curriculum Inquiry* **33**, 251–290.
- Connelly, F. M., He, M. F., and Phillion, J. (eds.) (2008). *Handbook of Curriculum and Instruction*. Thousand Oaks, CA: Sage.
- Connelly, F. M. and Xu, S.-J. (2007). On the state of curriculum studies: A personal practical inquiry. *Curriculum and Teaching Dialogue* **9** (1&2), 3–19.
- Connelly, F. M. and Xu, S.-J. (2008). The landscape of curriculum and instruction: Diversity and continuity. In Connelly, F. M., He, M. F., and Phillion (eds.) *Handbook of Curriculum and Instruction*, pp 514–533. Thousand Oaks, CA: Sage.
- Dahllöf, U. (1971). *Ability Grouping, Content Validity, and Curriculum Process*. New York: Teachers College Press.
- Dewey, J. (1902). *The Child and the Curriculum*. Chicago, IL: University of Chicago Press.
- Doll, W. E. (2002). Ghosts and the curriculum. In Doll, W. E. and Gough, N. (eds.) *Curriculum Visions*, pp 23–70. New York: Peter Lang.
- Doyle, W. (1977). Paradigms for research on teacher effectiveness. *Review of Research in Education* **5**, 163–198.
- Fox, S. (1985). The vitality of theory in Schwab's conception of the practical. *Curriculum Inquiry* **15**, 63–89.
- Goodlad, J. K. (1994). Curriculum as a field of study. In Husen, T. and Postlethwaite, T. N. (eds.) *The International Encyclopedia of Education*, 2nd edn., vol. 3, pp 1262–1267. Oxford: Pergamon.
- Hamers, U. (1991). Curriculum theory. In Husen, T. and Postlethwaite, T. N. (eds.) *The International Encyclopedia of Education*, 2nd edn., vol. 3, pp 1348–1355. Oxford: Pergamon.
- Hamilton, D. (1989). *Curriculum History*. Geelong, VIC: Deakin University.
- Jackson, P. W. (1992). *Handbook of Research on Curriculum: A Project of the American Educational Research Association*. New York: Macmillan.
- Kridel, C. (2006). Sweet thoughts of reverie: Rereading AERA conference programs, 1974–2006. *Journal of Curriculum Theorizing* **22**, 95–103.
- Levine, D. W. (2006). Joseph Schwab's assault on facile teaching. In Levine, D. W. (ed.) *Power of the Mind: The Reinvention of Liberal Learning in America*, pp 114–145. Chicago, IL: University of Chicago Press.
- Lundgren, U. P. (1972). *Frame Factors and the Teaching Process: A Contribution to Curriculum Theory and Theory on Teaching*. Stockholm: Almqvist and Wiksell.

- Meil, A. (2003). History of WCCI. http://www.wcci-international.org/index.plp?option=com_content&task=view&id=24&Itemid=44.
- Meyer, J. W. and Ramirez, F. O. (2000). The world institutionalization of education. In Schriewer, J. (ed.) *Discourse Formation in Comparative Education*, pp 111–132. New York: Peter Lang.
- Pinar, W. F., Reynolds, W. M., Slattery, P., and Taubman, P. M. (1995). *Understanding Curriculum: An Introduction to the Study of Historical and Contemporary Curriculum Discourses*. New York: Peter Lang.
- Reid, W. A. (1999). The voice of the practical: Schwab as correspondent. *Journal of Curriculum Studies* **31**, 385–397.
- Schubert, W. H. (1986). *Curriculum: Perspective, Paradigm and Possibility*. New York: Macmillan.
- Schwab, J. J. (1960). The teaching of science as enquiry. In Schwab, J. J. and Brandwein, P. F. (eds.) *The Teaching of Science*, pp 3–103. Cambridge, MA: Harvard University Press.
- Schwab, J. J. (1983). The practical 4: Curriculum inquiry. *Curriculum Inquiry* **13**, 239–265.
- Slattery, P. (2006). *Curriculum Development in the Postmodern Era*. New York: Routledge.
- Westbury, I. (2007). Theory and theorizing in curriculum studies. In Forsberg, E. (ed.) *Research Reports 7: Curriculum Theory Revisited*, pp 1–19. Uppsala: Uppsala University, Studies in Educational Policy and Educational Philosophy.
- Wikibooks (2007). Social and cultural foundations of American education. The WikiText Development Process in EC1301 at Old Dominion University, Darden College of Education, Fall 2006/Spring 2007 2. http://en.wikibooks.org/wiki/Social_and_Cultural_Foundations_of_American_Education/Second_Edition/WikiText_Development_process (accessed April 2009).
- Wraga, W. G. and Hlebowitsh, P. S. (2003). Commentary: Conversation, collaboration, and community in the U.S. curriculum field. *Journal of Curriculum Studies* **35**, 453–457.
- Xu, S. J. (in press). *Narratives of Newcomer Families' Cross-Cultural Schooling Experience*. Beijing: Foreign Language Teaching and Research Press.

Further Reading

- Ayers, W. and Ford, P. (1996). *City Kids City Teachers: Reports from the Front Row*. New York: The New Press.
- Clandinin, D. J. and Connelly, F. M. (1992). Teacher as curriculum maker. In Jackson, P. W. (ed.) *Handbook of Curriculum*, pp 363–461. New York: Macmillan.
- Connelly, F. M. and Clandinin, D. J. (1988). *Teachers as Curriculum Planners: Narratives of Experience*. New York: Teachers College Press.
- Kliebard, H. (1992). *Forging the American Curriculum*. New York: Routledge.
- No Child Left Behind Act of 2001. Pub. L. No. 107-110 115 Stat. 1425 (2002). <http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf> (accessed April 2009).
- Schwab, W. H., Schubert, A. L., Thomas, T. P., and Carroll, W. M. (2002). *Curriculum Books*. New York: Peter Lang.
- Short, E. C. (ed.) (1991). *Forms of Curriculum Inquiry*. Albany, NY: State University of New York Press.
- Short, E. C. (2005). Curriculum Inquiry and Related Scholarship: A Searchable Bibliography of Selected Studies. <http://edcollege.ucf.edu/esdepart/cirs/main.cfm> (accessed April 2009).
- Westbury, I. and Wilkof, N. J. (eds.) (1978). *Joseph Schwab: Science, Curriculum, and Liberal Education: Selected Essays*. Chicago, IL: The University of Chicago Press.

Curriculum Evaluation: Approaches and Methodologies

V Klenowski, Queensland University of Technology, Brisbane, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Illuminative evaluation – This uses naturalistic methods to identify underlying factors and issues important to participants.

Naturalistic evaluation – This is responsive and involves the use of qualitative methods for the collection and analysis of data generally to complement the evaluation design.

Portrayal evaluation – This uses data collected using qualitative methods to provide description of the context and participants of the evaluation.

Transactional evaluation – This provides data about the processes and exchanges between the evaluator and the participant.

Curriculum evaluation, as a field of study, is dynamic. For over the years it has been responsive to the developments in the conceptualization of curriculum and the associated processes of curriculum change. The concept of curriculum is integral to curriculum evaluation and can be defined in terms of what can and shall be taught to whom, when, where, how, and why. Much of the decision making relates to what knowledge is to be selected for inclusion in the curriculum. The term curriculum is often used to refer to the programs of teaching and learning with some theorists emphasizing the content, standards, or objectives while others give more attention to didactic choices such as the teaching methods and pedagogical strategies used. In this article, a consideration of terminology, that is central to an understanding of the nature and development of curriculum evaluation, is given first. A historical overview and an outline of the approaches and the developments in the methodology of curriculum evaluation follow with some recent trends in the context of test-based accountability and the use of assessment data included. How these developments are seen to drive educational evaluation, in general, and curriculum evaluation, in particular, is taken into account.

Curriculum in terms of policy is best understood in the context of levels, from the broadest level such as the education system through to the level of the classroom. The emphasis in the curriculum in terms of the knowledge selected to be taught will vary depending on the context. At the level of the school, the curriculum is classified in terms of the subject such as the mathematics or the science curriculum. Such selections of knowledge are further

described in terms of: scope and sequence; syllabus; content; standards; textbooks; course of study; and planned experiences. The decision making involved from the origins of the curriculum through to implementation include choices regarding the range of subjects, disciplines, activities, subject matter, sequence, settings for learning, assessment, teacher development, and the like. In this context, there are competing viewpoints and interest groups and therefore curriculum development has been considered a political activity that involves political decisions.

Evaluation has been defined as the processes of description, analysis, and judgment of educational programs, practices, institutions, and policies for a range of purposes. Assessment is the term used most often in relation to student learning or performance and relates to the activity or task designed to show what a person knows or can do. Assessment constitutes only a part of evaluation in connection with a concern for the outcomes of a course, curriculum, or program.

Any evaluation project, will either explicitly or more often implicitly, contain within it a constellation of answers to certain key questions about its purpose (Why is this evaluation being undertaken? Who is sponsoring the evaluation? Who is conducting the evaluation?); ethics (Whose judgment counts? What rights do those being evaluated have?); political stance (Who is the audience? How is the evaluation reported? Who will take action as a result of evaluation?); and procedural choices (How is the evaluation to be carried out? Whose criteria of value are being applied?) (Harland, 1994).

Evaluation of curriculum involves determining the value or worth of particular products or processes that can include learning objectives, documents, or experiences for the purposes of informing decision making about the curriculum. Individuals, organizations, communities, and governments engage in curriculum evaluation to maintain, develop, or modify a particular curriculum program, policy, or practice with a view to implementation, development, or change. Curriculum evaluation can be conducted for formative or summative purposes. Formative evaluation of the curriculum fulfills a role in improvement, development, or modification and is conducted as curriculum is evolving. Often, it is carried out concurrently providing feedback into those processes of development and implementation. Summative evaluation on the other hand takes place generally to assess the outcomes, the merit, or worth of a particular curriculum.

Summative evaluation also informs decision making about whether or not a curriculum is considered worthy of continued support or adoption.

Curriculum evaluation serves different purposes ranging from providing understanding about how to improve a particular practice or program to providing proof that productivity and efficiency gains have been achieved by the implementation of a particular curriculum practice or program. Chelimsky's (1997) categorization of evaluation purposes is helpful. She has identified the purposes of accountability (measurement of results or efficiency), development (to develop or improve), and knowledge (deeper understanding of practice or policy). The purpose of the evaluation of the curriculum will impact on the choice of methods used. A range of procedures and processes are used to judge the value or merits and the shortcomings of both current and emergent practices of curriculum in the education system for varying purposes, including regulation and policy development. Initially, curriculum evaluation was carried out to support the development and implementation of curriculum innovation; the maintenance of curriculum through curriculum evaluation came later with the move toward control of the curriculum becoming more evident in recent times.

Historical Overview

For many years, teachers and others have informally evaluated curricula using school-level strategies. Such curriculum evaluation has been situational or problem specific and has been conducted to meet the information needs of particular audiences. In conducting such evaluations, teachers have used methods such as observations, interviews, tests, and questionnaires to collect data for judging the quality of certain aims or learning intentions, materials, processes, and outcomes. Informal curriculum evaluation conducted by teachers has helped raise questions about the quality of choices made and the choice-making process itself in connection with the curriculum and the teaching and learning programs.

In more formal settings, the strong interactive relationship between the concepts of innovation (planned educational and/or curriculum change) and evaluation can be traced back to developments in the United States. For instance, in the early 1920s, curriculum theorists such as Bobbitt and Charters were strongly influenced by the forces of industrialization, and accompanying concepts of division of labor. Their analyses of the social role of the school curriculum led them to identify the need for specialized training and the adoption of the procedures of job analysis and aspects of scientific management theory for the construction of a differentiated curriculum based on educational objectives in preparation for adult life. In terms of curriculum evaluation, a further development

was the identification of definite qualitative and quantitative standards (Kliebard, 1992) and as is evident today, the efficacy of standardized curriculum via the use of standardized assessment has survived into this century.

By the mid-twentieth century, it was the launch of Sputnik by the Russians that inspired further curriculum development. In particular, it was the science and mathematics projects which helped foster a curriculum development movement that continued into the 1960s and 1970s both in the US and the United Kingdom. Growth in funded curriculum projects was of a scale never before experienced which helped generate evaluation activity. Increased government funding to update curricula in the 1950s and 1960s led to particular strategies for curriculum development and enhancement of particular subjects such as science and mathematics. With such increased spending on curriculum development, evaluation became mandatory. The focus on value for money led to curriculum evaluation that included quantification in terms of inputs and outputs.

These funded projects were accompanied by systematic evaluation activities which stimulated interest in the processes of change and innovation. Evaluators were asking not only how successful a given program was proving, but also why some programs succeeded and some did not, and why some contexts and implementation strategies were conducive to effective implementation and others were not. Sometimes the overall value of the planned change was questioned. Curriculum evaluation thus emerged as a response to the need to establish the worth of innovative curriculum projects. An academic and professional discipline of evaluation (Stake, 2004) was now required.

New methodologies emerged in order to understand the implementation of curriculum projects and the nature of the complexity of the curriculum change process. With changed methodologies, the awareness of ethics became more obvious and with a greater awareness of the intrinsic political nature of evaluation, the need to provide information for a range of decision makers became important. The increasing sophistication of evaluation theory and practice in the context of curriculum was significant in alerting both policymakers and practitioners to the political and cultural issues surrounding curriculum innovation and change. The theorizing about curriculum evaluation that followed led to various models and approaches to evaluation.

Although formal curriculum evaluation was originally associated with innovation and development, since the 1980s it has increasingly come to be used as a management and, indeed, as a control strategy. Evaluation is a political activity and the process of valuing existing curriculum practice and current curriculum policy developments is a potent influence on educational reform. Curriculum evaluation can be seen as a vehicle for giving

priority to particular values and for determining who gets what, when, and how. Today, it is a mainstream activity, part of many policy initiatives, often having an impact beyond its apparent function and still capable of providing an important window on the processes of implementing curriculum change.

Curriculum evaluation is concerned with evaluating a particular curriculum project or program which can lead to the evolution of a theory of how to change the curriculum. Curriculum evaluation can also use theory (such as organization theory) to interpret and explain what is being evaluated or it can look to theory at the methodological level in the analysis of data. Curriculum evaluation differs from theory in its direct concern with a particular evaluation or that which is being evaluated (Stake, 2004).

Approaches to Curriculum Evaluation

Approaches to curriculum evaluation have a bearing on practice and have paralleled developments in curriculum. Some early curriculum evaluation was characterized by measurement and the collection of factual information. A psychometric approach was dominant and geared particularly to assessing student attainment rather than policies and programs. Tests were used to evaluate the efficiency of teachers and schools in delivering the curriculum.

The objectives achievement curriculum model, popularized by Tyler (1949), was a reaction to the reliance on traditional content curriculum. This development coincided with the modernization of industry. Tyler's objectives model saw the curriculum behaviourally; the curriculum was evaluated in terms of student achievements and outcomes. This approach to curriculum evaluation concentrated on inputs and outputs, involved pre- and post-testing, experimental design, and had common sense appeal because it was direct and simple. The limitations of this model were soon identified. These included the practice of teaching to the test and a lack of consideration of the quality of the objectives. Issues related to curriculum and pedagogy were seldom pursued in such evaluations, and questions of why certain outcomes were not achieved were not taken into account. In this approach to evaluation, the curriculum, in terms of the teaching and learning programs, remained the black box.

Critics of this objectives achievement model of evaluation soon emerged with Scriven (1967) refocusing evaluation toward judging the merit or worth of the goals and whether the program was indeed fulfilling the client's need. He introduced the concepts of formative and summative evaluation and distinguished between the roles (function) and goals (what is to be achieved) of evaluation. Evaluation involved perceptions of value and estimations of goal achievement according to Scriven. Cronbach (1981) added to the debate and emphasized the importance of

usability over accuracy in evaluative criteria for improvement particularly in the examination of possible causes and effects of teaching quality. Cronbach was one of the first to use evaluation data for course improvement or to fulfill a formative function. Stufflebeam (1971) highlighted how evaluation could feedback into the decision-making process at the beginning, during the planning, and implementation phases of curriculum development classifying data as context, input, process, and product (CIPP). However, this approach focused mainly on administrative decisions.

Stake (1967) introduced countenance evaluation which distinguished between describing and judging in the process of evaluation. Many useful data for program evaluation were proposed by way of the countenance matrix. This matrix provides the data collection structure for the evaluator to attend to the intents and observations for description on the one hand and the standards and decision making for judgment on the other. Stake maintained that to evaluate a program it needed to be both described and judged. In this model of evaluation, capturing the notion of time was included by differentiating the antecedent, transaction, and outcome data. This was a major shift from a focus on the outcomes of education to the inclusion of the conditions and factors that impact on teaching and learning. Responsive evaluation which Stake (2004) describes as an attitude more than a model is a perspective that responds to the teaching and learning program to be evaluated and constitutes a search for quality. This approach is more responsive to issues in the program identified by different people and represents a further development, particularly methodologically, toward a more process-oriented approach. The teaching and learning program in the context of curriculum evaluation provides the stimulus and focus. There is acknowledgment of the differing information needs of the audience. A teacher will have very different information needs from the curriculum evaluation to those of a policy officer. In responsive evaluation, the focus is on the issues, the conceptual organizers (Stake, 2004) of the particular curriculum development program, rather than its intents. Different people's views will be engaged in the reporting of its success or failure. In this sense, responsive evaluation recognizes different sources and different grounds for valuing particular teaching and learning programs.

In the UK, for example, there was no such tradition of curriculum evaluation as in the US. There was no consistent model of curriculum design. A climate of accountability at that corresponding time was not pervasive. The model of curriculum evaluation that did emerge was one of informed professional judgment. Schools and classrooms became the sites of implementation and evaluation and the teacher-researcher movement in UK emerged as a response to teachers as agencies of their own improved pedagogy. The curriculum development movement in the UK was an opportunity to develop innovative curricula

which started with the Nuffield Foundation funded projects in the sciences, followed by the development of projects in the humanities sponsored by the Schools Council. Some of these curriculum projects were evaluated formatively by curriculum development teams while others were evaluated summatively by independent, external evaluators. A cadre of external evaluators emerged and a profession of evaluators began to be established. Teachers and local authorities rather than politicians were the audience for these evaluations.

The adoption of the research and development approach to one's teaching, either with peers or alone, became prominent with the take-up of the notion of teacher as researcher and the view that there could be no curriculum development without teacher development (Stenhouse, 1975). This approach which arose in response to the limitations of the curriculum objectives model defined the classroom process in terms of what the teacher should teach at the level of principles and what should be learned by whom, when, where, how, and why. This model of curriculum was more process oriented. It constituted a reaction against the focus on objectives, products, and knowledge to a focus on principles (the best criteria for the selection of content and pedagogy), processes (including both teaching and learning processes), and development (from expanding the amount of students' knowledge to developing their minds or improving the quality of their learning). Similar to reflective practice, this approach to evaluation developed into an action-research movement. School self-evaluation (Simons, 1987) also developed in this professional context to take account of the need to change the institutional context if one was to change the teacher pedagogy or curriculum within it.

With such curriculum-inspired development, approaches to evaluation emerged that enabled evaluators to more closely examine the curriculum development program, in terms of the processes and interactions at the local professional level of the classroom (Simons, 1987). These interactions included those between teachers, students, subject matter, and context. Attention to processes and experiences in the analysis of curriculum programs also helped reveal the complexities and the change potential of a curriculum program. Responsive evaluation, illuminative evaluation, and democratic evaluation (MacDonald, 1976) provided the necessary strategies for exploration inside the black box of the teaching and learning programs. The naturalistic evaluation paradigm which drew on qualitative methods for data collection and analysis and on human experience as the data source grew in significance. These changes evolved as the basis for evaluation broadened to include issues related to the implementation of curriculum. Different information was required by different audiences as evaluation was used to describe the quality and value of particular curriculum and was used to help explain why

some curriculum practices had shortcomings. Teachers too were developing insights into their pedagogic and assessment practices as they grew professionally from their active involvement in the evaluation process.

It became recognized that there was a need for different evaluation information for political and professional accountability purposes at the various levels from the center to the local dimension. The typology of bureaucratic, autocratic, and democratic evaluation (MacDonald, 1976) helped heighten awareness of the power relationships that exist among the levels of accountability and evaluation. MacDonald's (1976) typology of evaluation provided three categories of enquiry: bureaucratic, autocratic, and democratic. Bureaucratic evaluation involves service to government agencies. The evaluation is conducted within the value system of the agency and the information provided addresses the agency's policy objectives. The methods chosen to conduct the evaluation also have to be acceptable by the bureaucracy whose interests dominate. The evaluator does not control how the information is used and the evaluation report is owned by the bureaucracy.

Autocratic evaluation is conducted as a conditional service to government. The evaluator provides validation of a particular policy with recommendations that are to be met by the agency. The value system within which the evaluation is conducted is derived from that which the evaluator perceives to be the responsibilities of the bureaucracy. The evaluator is an expert advisor and must provide proof using methods acceptable to academic peers as the report remains with the bureaucracy, and yet can also be reported in academic journals.

Democratic evaluation, a form of process evaluation, is the type that MacDonald (1976) supported and developed. The evaluator provides an information service to the community about a particular curriculum or educational program. The sponsor does not have privileged rights over the evaluation. Different audiences and values are acknowledged. To achieve this requires methods and reporting that are accessible to nonspecialists. The evaluator represents the range of interests in identifying the issues and the various responses to the program. This approach emphasizes the rights to know about the program and the responsibilities of the evaluator to offer confidentiality to the informants. Such democratic forms of evaluation represent a wider range of perspectives and interests, involve more stakeholders in the evaluation process, and provide increased opportunities for dialog and deliberation (House and Howe, 2003). Evaluation that provides information for a wider range of decision makers requires an approach that includes teachers, pupils, administrators, and curriculum developers.

Process evaluation that operates within a participatory inquiry paradigm and a democratic ethic engages the practitioner in understanding the demands of the curriculum program, policy, or practice. In recent years, the use

of more participatory and democratic forms of evaluation that value dialog, engagement, relationships of trust, and respect for multiple values has become more evident.

Developments in the Methodology of Curriculum Evaluation

Closer analyses of curriculum programs, policies, or practices were made possible by more democratic and process forms of curriculum and evaluation. They provided greater understanding of the potential for change of the particular curriculum development. No longer was curriculum change and innovation considered only from a technological or professional perspective; political and cultural points of view were equally valued (House, 1978).

A technological perspective is permeated by concepts of rationality, stage, and sequence, and is often strongly normative in quality. From this view, a curriculum innovation is implemented and explicated using a rational, planned process with staged, sequenced procedures. Techno-rational discourse has tended to dominate such evaluations. In the latter part of the twentieth century, some education management, leadership, and school effectiveness studies of curriculum change were conducted using a classical, quantitative methodology to capture this rational approach and to explain causality. The evaluation methods that tended to dominate were psychometric incorporating quantitative data analysis of student assessment data to measure teaching and learning outcomes.

Evaluation and analysis of planned curriculum change, from a political perspective, documents issues and problems associated with the irrational, the self-interested, and the awkward in the process of curriculum change. The key concepts of power, ideology, resources, resistance, and negotiation are identified and the territory in which they operate is the local, or the specific institution; this is micro- not macro-politics. To capture the micro-political nature in curriculum evaluation, it is important for the methodology to include qualitative and illuminative methods such as interviews and classroom observations. The evaluator in this context has to build and sustain relationships to gain and maintain access. There is a need for the evaluator to observe the interactions closely to explore contingencies and the relationships between people and between people and events. The evaluator works closely and more interactively with those involved in the evaluation enabling a richer understanding of the complexity of curriculum change. However, to evaluate curriculum change as only a political process and to ignore problems such as understanding, identity, competence, or management is overly reductionist; it is important to consider the change from a cultural perspective.

Curriculum innovation from a cultural perspective not only acknowledges the importance of the political, but also involves a more complex understanding of change, involving not only political negotiation between the self-interested and differentiating powerful groups, but a recognition of the importance of change and pluralism in terms of values, beliefs, understanding, practices, relationships, and structures. The whole culture of the institution is important to consider from this perspective. This means to understand curriculum change fully requires an understanding of the general culture of the institution and the cultures of the groups that are embedded within. In a planning context, this implies that the management of curriculum change involves not only the development of structures and procedures, but also the development of a cultural sensitivity and insight, plus an ability to understand the meaning of change from a range of perspectives other than one's own. This is a significant departure from the rationalist classification of models and strategies developed within a technological perspective. To capture this level of complexity and understanding of curriculum innovation and change the evaluation methodology is more comprehensive and holistic engaging both quantitative and qualitative methods. In this methodological approach, there is collaboration between the evaluator and the teacher or respondent; the methods are more interactive and iterative, with greater collaborative efforts to understand and explain action. In this context, the teacher is both informant and judge.

The case study in curriculum evaluation has been described as the methodological counterpart of democratic evaluation (Kushner, 2002). In democratic evaluation, the purpose and function is to understand, to describe, and to judge the characteristics of the curriculum change, policy, or practice from multiple viewpoints; the evaluator is sensitive to the different values of participants. Case study is an appropriate method to achieve this purpose and requires the use of more extensive, naturalistic, adaptable, empirical approaches such as ethnographic fieldwork. Interview, observation, portrayal, vignettes, and narrative accounts are often used to provide thick description of experiences and information for analysis. Different models of evaluation such as responsive evaluation, portrayal evaluation, transactional evaluation, illuminative evaluation, and holistic evaluation, that can incorporate case study, give a richer understanding of the cultural, the political, the ethical, and the methodological.

Recent Trends in Curriculum Evaluation

Since the early 1990s, an international trend of growing dissatisfaction with existing educational practices and outcomes has led to increased educational reform.

At the same time, there has also been a worldwide shift in control of education away from teachers toward the state for the purposes of restructuring economies. More bureaucratic forms of curriculum have resulted, with a return to the use of more techno-rational discourse in evaluation for purposes of efficiency, accountability, impact, and performance management.

Today, what appears to have survived from the work of Bobbitt and Charters is the view by some curriculum developers of the identification of the skills, knowledge, capabilities, or competencies that will benefit the child in the future. These predictions of what one will need have become the bases for planning of the curriculum (Kliebard, 1992) with the use of standardized testing in curriculum evaluation to check the efficacy of standardized curriculum.

To illustrate the bureaucratization of curriculum, an example is given from the UK when in 1988 a national curriculum that was preoccupied with achievement in terms of student results was adopted. Evaluation took on primarily an accountability function. The national curriculum identified the content of programs, and the objectives and processes, in terms of targets or standards. The publication of league tables became a device for judgments about school performance. Evaluation was divided into two trajectories. First, evaluation directed from the center was embedded in the bureaucratic process of centralized curriculum. The use of inspection and punitive accountability systems emerged. Second, the tradition of school self-evaluation (Simons, 1987) went underground as the dominant focus reemerged of testing and achieving targets. There was a groundswell of school self-evaluation activity which was eventually overtaken by other forms of evaluation that focused on assessment and testing.

Internationally, policymakers and others have shown increased interest in international measures of educational attainment such as the results from the Program for International Student Assessment (PISA), developed by the Organisation for Economic Co-operation and Development (OECD) or the Third International Mathematics and Science Study (TIMSS) of the International Association for the Evaluation of Educational Achievement (IEA). These comparisons have influenced policy development, yet important questions of whether we are comparing like with like have not always been considered. When such comparisons are made, it is important that a common set of criteria is used for measuring performance, that there is comparability between samples and the reporting of the results, that there is a match in terms of the content of the curriculum and the approach used, and that regard is also given to context.

The results from such international comparisons have been used by governments to justify the introduction of ongoing curriculum change. For example, in the UK, the then Department for Education and Employment (DfEE) commenced a National Numeracy Project to address

perceived weaknesses, particularly in the teaching of mathematics at primary school level, after the publication of the results of TIMSS in 1996. This project was followed by the adoption of the National Numeracy Strategy, the National Literacy Strategy, and the Key Stage 3 Strategy by the Department for Education and Skills (DfES).

For curriculum evaluation, the consequence of such developments has been that – far from representing a relatively independent and/or predominantly professional activity – evaluation has been incorporated into the processes of policy development and system management. Effectively this has meant that evaluation is little concerned with debating fundamental value issues in the curriculum program or curriculum strategy itself, but is now incorporated into implementation. Some organizations and government agencies are reporting evidence from curriculum evaluation using a what-works approach. In an accountability context, the need to demonstrate performance is heightened and explicating what works is pursued. Schwandt (2003: 362) has described such knowledge as a commodity that is bought, sold, and applied. The value of dialog and deliberation with practitioners in evaluation to facilitate understanding of the challenges of diverse values in the context of practice has been recognized.

However, the combination of a what-works approach and evidence-based decision making has reinvigorated concerns relating to measurement, validity, and reliability of quantitative measurement. Some government agencies are demanding evidence-based policy and practice (EBPP), derived from randomized experimental designs, as a basis for intervention and pursuit of policy agendas. In this drive for efficient use of resources with the development of guidelines and frameworks to regulate and assess evaluation practice, caution must again be taken to ensure that we are comparing like with like. It is important that in the synthesis of evaluations for EBPP that outcomes have not been simplified and contexts have not been ignored.

Conclusion

As time passes, the purposes of curriculum change as do the purposes of its evaluation. The evaluator of a particular curriculum program or practice will identify the purpose of the evaluation first, then the most appropriate methodology to serve that function is chosen. Important methodological insights have been gained by evaluators over time as the purposes of evaluation have shifted from proving a particular outcome to gaining a deeper understanding of the practice or policy and how it might be improved by engaging with practitioners. The refocusing of approaches and methodologies has been accompanied with a growing understanding of the complexity of the interactions among the levels from the political center to the local professional level of the classroom.

However, in spite of the generation of democratic, responsive, and deliberative forms and purposes of evaluation it would appear that evaluation for accountability and control continues to impact on current practice as is evident in the reemergence of bureaucratic forms of curriculum and the return of quantitative, reductionist approaches as evident in No Child Left Behind. What also becomes apparent in the name of efficiency is a return to technological and behavioristic refinements of curriculum evaluation and a possible trivialization that threatens the richness of the intellectual activity of those involved in the discipline of curriculum evaluation.

Bibliography

- Chelmsky, E. (1997). Thoughts for a new evaluation society. *Evaluation* **3**, 97–118.
- Cronbach, L. J. (1981). *Towards Reform of Evaluation*. San Francisco, CA: Jossey-Bass.
- Harland, J. (1994). *A Catechism for Evaluation*. London: Institute of Education, University of London.
- House, E. R. (1978). Assumptions underlying evaluation models. *Educational Researcher* **7**, 4–12.
- House, E. R. and Howe, K. R. (2003). Deliberative democratic evaluation. In Kellaghan, T. and Stufflebeam, D. L. (eds.) *International Handbook of Educational Evaluation*, pp 79–100. Dordrecht: Kluwer Academic.
- Kliebard, H. M. (1992). *Forging the American Curriculum. Essays in Curriculum History and Theory*. New York: Routledge.
- Kushner, S. (2002). The object of one's passion: Engagement and community in democratic evaluation. *Evaluation Journal of Australasia* **2**, 16–22.
- MacDonald, B. (1976). Evaluation and the control of education. In Tawney, D. A. (ed.) *Curriculum Evaluation Today: Trends and Implications*, pp 125–136. London: Macmillan.
- Schwandt, T. A. (2003). 'Back to the rough ground!' Beyond theory to practice in evaluation. *Evaluation* **9**, 353–364.
- Scriven, M. (1967). The methodology of evaluation. In Tyler, R. W., Gagne, R. M., and Scriven, M. (eds.) *Perspectives of Curriculum Evaluation*, vol. 1, pp 39–83. Chicago, IL: Rand McNally.
- Simons, H. (1987). *Getting to Know Schools in a Democracy: The Politics and Process of Evaluation*. Lewes: Falmer Press.
- Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record* **68**, 523–540.
- Stake, R. E. (2004). *Standards-Based and Responsive Evaluation*. Thousand Oaks, CA: Sage.
- Stenhouse, L. (1975). *An Introduction to Curriculum Research and Development*. Oxford: Heinemann Educational.
- Stufflebeam, D. L. (1971). The relevance of the CIPP evaluation model for educational accountability. *Journal of Research and Development in Education* **5**, 19–25.
- Tyler, R. (1949). *Basic Principles of Curriculum and Instruction*. Chicago, IL: University of Chicago Press.

Further Reading

- Apple, M. (2004). *Ideology and Curriculum*, 3rd edn. London: RoutledgeFalmer.
- Eisner, E. W. (1979). *The Educational Imagination: On the Design and Evaluation of School Programs*. New York: Macmillan.
- Greene, J. (2001). Dialogue in evaluation a relational perspective. *Evaluation* **7**, 181–187.
- Greene, J. (2002). Towards evaluation as a 'public craft' and evaluators as stewards of the public good. *Keynote Address at the Annual Meeting of the Australasian Evaluation Society*. Wollongong, Australia.
- House, E. R. (1993). *Professional Evaluation: Social Impact and Political Consequences*. London: Sage.
- House, E. R. (2005). Promising practices: The many forms of democratic evaluation. *Evaluation Exchange* **XI**, 1–2.
- House, E. R. (2006). Democracy and evaluation. *Evaluation* **12**, 119–127.
- House, E. R. and Howe, K. R. (1999). *Values in Education and Social Research*. Thousand Oaks, CA: Sage.
- MacDonald, B. and Kushner, S. (2004). Democratic evaluation. In Mathison, S. (ed.) *The Encyclopedia of Evaluation*, pp 108–113. Thousand Oaks, CA: Sage.
- Simons, H. (2004). Utilizing evaluation evidence to enhance professional practice. *Evaluation* **10**, 410–429.
- Madaus, G. and Kellaghan, T. (1992). Curriculum evaluation and assessment. In Jackson, P. W. (ed.) *Handbook of Research on Curriculum*, pp 119–154. New York: Macmillan.
- Stake, R. E. (1995). *The Art of Case Study Research*. Thousand Oaks, CA: Sage.
- Stern, E. (yearly). *The International Journal of Theory, Research and Practice*. <http://www.sagepub.co.uk/journalsProdDesc.nav?prodId=Journal200757> (accessed May 2009).
- Weiss, C. (1983). The stakeholder approach to evaluation: Origins and promise. In Bryk, A. S. (ed.) *New Directions for Program Evaluation*, No 17, pp 3–14. London: Jossey-Bass.
- Weiss, C. (1999). The interface between evaluation and public policy. *Evaluation* **5**, 468–487.

Relevant Websites

- <http://oerl.sri.com> – Online Evaluation Resource Library.
- <http://www.aes.asn.au> – The Australasian Evaluation Society (AES) Publications.
- <http://www.uwex.edu> – The University of Wisconsin – Extension, Evaluation Publications.
- <http://www.evaluation.org.uk> – UK Evaluation Society.
- <http://gsociology.icaap.org> – Understanding the World Today, Resources for Program Evaluation and Social Research Methods.
- <http://www.uea.ac.uk> – University of East Anglia, Centre for Research and Evaluation in Education.
- <http://www.ed.uiuc.edu> – University of Illinois at Urbana-Champaign, Circe Publications.
- <http://www.wmich.edu> – Western Michigan University, The Evaluation Center.

Curriculum Studies, Discourse Analysis, and the Construction of Historical Time

J Laspina, Independent Scholar, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

In every society the production of discourse is at once controlled, selected, organized and re-distributed by a certain number of procedures, the role of which is to conjure its powers and dangers. Michel Foucault *The Order of Discourse* (Paras, 2006: 49)

As the current literature attests, the practice of discourse analysis has entered the educational research mainstream (Rogers, 2004, ix). The relationship of discourse analysis to the curriculum and the field of curriculum studies has always been strong. It developed in several phases internationally since the 1970s, starting with the sociology of education movement. A theoretical base may be found in the seminal work of Basil Bernstein, Pierre Bourdieu, and Louis Althusser on cultural transmission (Karabel and Halsey, 1977: 451–455). However, the theory of transmission articulated in their work is not a simple one-way process of communication in which a sender conveys a message directly to a receiver. Rather, in Bernstein's work, knowledge is produced through a complex linguistic interaction of regulative and instructional discourses. In Bourdieu, the school tends to reproduce the dominant cultural and social discourse of society. For Althusser, the institutional process that reproduces that dominant discourse takes the social form of hegemony in which a privileged class dominates through a sociopolitical process of conflict and contestation in which social resistance is neutralized and a hegemonic order is reproduced. One of the principal ideological mediums for this transmission is the curriculum (Apple, 1979).

Power, dominance, and the institutional structures that produce these discourses are the key themes of discourse analysis (Van Dijk, 2001: 301–02). In general, discourse analysis considers three social science domains, and its practice is shaped by at least six distinct traditions (Wetherell *et al.*, 2001: 5–6):

Discourse Domains

- the study of social interaction
- the study of minds, selves, and sense making, and
- the study of culture and social relations

Discourse Traditions

- Conversation analysis and ethnomethodology;
- Interactional sociolinguistics and the ethnography of communication;
- Discursive psychology;
- Critical discourse analysis and critical linguistics;

- Bakhtinian research;
- Foucauldian research.

Within these traditions two methodologies by Fairclough (1992) and Gee (1999) who have been influenced by the work of Bernstein, Bourdieu, and Althusser are presently widely used in the field of educational research (Rogers, 2004: 6–7). However, the purpose of this article is not to present a general exposition of these prominent methodologies of discourse analysis, but to present a historical overview of the turn to discourse in the field of curriculum studies. The method of discourse analysis applied herein is primarily historical, but with a Foucauldian emphasis on the way discourse is understood as fundamentally historicized, that is having meaning only in terms of a specific historical context (Hall, 2001: 74).

This condition reflects the most elemental feature or condition of discourse – its temporality (i.e., production occurs in time). To an inevitable extent, the historical representation of the postmodern turn to discourse in the field of curriculum studies is conventional as an interpretative representation of the seminal texts which are reviewed over a period of the last two decades. Consideration of these texts over time tends to highlight not so much how their ideas may change, as how discourse is constructed. In Foucauldian terms, this historical account of major developments in the field of curriculum studies may be analyzed as it coexists with other accounts. In such a historical analysis, the purpose is not to “treat . . . the representations that might exist behind the texts defining this development, but rather the discourses themselves as regular and distinct series of events” (Paras, 2006: 50). For that distinctiveness or variability is but a reflection of how discourse is produced. For if they are considered successively, the historical rendering of these accounts may better serve to expose the linguistic procedures and institutional resources supporting their ideality and ideological power.

In *The Production of Reason and Power: Curriculum History and Intellectual Traditions*, Thomas S. Popkewitz, states that

Curricula are historically formed within systems of ideas that inscribe styles of reasoning, standards, and conceptual distinctions in school practices and subjects [and that these] systems of reasoning embodied in schooling are the effects of power.

In a word, Popkewitz is calling for a curriculum history that acknowledges its power as discourse (Popkewitz, 2001: 151). In light of his ambitious project, this article has a more modest objective: To present a historical overview and analysis of developments in the field of curriculum studies that disposed it to understanding curriculum as discourse. These developments are represented in three phases beginning with the postmodern turn to discourse in the field. In a Foucauldian sense this account is intended to present curriculum history as a historicized discourse.

Postmodernism and the Emergence of Curriculum as Discourse

A postmodern view of the field of curriculum studies tends to accentuate its temporal and ideological discontinuity. This discontinuity may be illustrated by three temporal and theoretical dichotomies:

- modern/postmodern;
- traditional/reconceptualist; and
- theorists/practitioners

The apparent binary is stressed because it can be used to illustrate the principal tenet of discourse analysis. According to Potter and Wetherell (2001), “people . . . us[e] their language to *construct* versions of the social world.” Being constructed, these “versions” or “accounts of events,” they note, “are built out of a variety of pre-existing linguistic resources,” and involve “active selection, some resources are included, some omitted.” This suggests that every account is a linguistic construct or version of reality, which in effect excludes other versions (Potter and Wetherell, 2001: 199).

The turn to postmodernism in the field of curriculum studies begins in the 1970s with the reconceptualist movement (Doll, 1993). Over the next three decades, the American reconceptualists introduced the European continental philosophical theories of phenomenology, deconstruction, and hermeneutics to the field. These disciplines require the conceptual movement of framing curriculum as text and the production of texts (Pinar and Reynolds, 1992; Slattery, 1995). For the reconceptualists, viewing curriculum as [text] signals the field’s entry into discourse.

But how do the reconceptualists conceive of discourse? With the publication of *Understanding Curriculum: An Introduction to the Study of Historical and Contemporary Curriculum Discourses*, they attempt to provide a comprehensive statement, a multivocal discourse that weighs in at over a 1000 pages, references included (Pinar *et al.*, 1995). In four sections and 15 chapters, the authors distinguish between discourses that are historical and those that are contemporary. For the authors, this distinction between past and

present, the historical and the contemporaneous, is meant to frame a paradigmatic shift in the field, which they describe as a conceptual movement “From Curriculum Development to Understanding Curriculum” (Pinar *et al.*, 1995: 3–11).

Invoking Kuhn (1962), the authors localize the production of discourse within fields which “are comprised of people with ideas working on problems – theoretical and practical . . . within institutional constraints” (Pinar *et al.*, 1995: 4). They stress that a field of study is a field of study, a study of language or discourse; that is:

To understand the contemporary field it is necessary to understand the curriculum field as discourse, as text, and most simply but profoundly as words and ideas. By discourse we mean a particular discursive practice, or a form of articulation that follows certain rules and which constructs the very objects it studies. Any discipline or field of study can be treated and analyzed as such. To do so, requires studying *the language of the field*. (Pinar *et al.*, 1995: 7)

The notion of text as the source of discourse can be viewed in the broadest of terms as: words and ideas that are spoken, written, or represented in multiple media. Three conventional examples are noted:

- enunciation as in speech acts (e.g., Fairclough and Gee’s methods of discourse analysis in large part describe and analyze the discursive relations and social practices embedded in ordinary language and discourse, see Rogers, 2004: 4–6);
- any or all written official documents involving the formal institutional aspects of schools and their organization, for example, state-approved curriculum frameworks; and
- any text representation of knowledge, for example, instructional textbooks, which figure prominently in *Understanding Curriculum* (Pinar *et al.*, 1995: 11–24).

Much hinges upon this from/to turn of phrase. For the reconceptualists, it serves as a conceptual crux to distinguish postmodern theory from modernist curriculum practice. The conceptual movement to an understanding perspective versus viewing curriculum development as action, that is, a prescribed set of industrial procedures in which a course of study is produced, represents the still dominant modernist paradigm of curriculum represented by Ralph Tyler (Slattery, 1995: 245–246). In effect, their pronouncement is meant to be taken as the ideological point of divergence. To make this point, they use a generic view of curriculum development by Decker Walker to highlight the difference between theory and practice. Walker states that

The one term “curriculum development” covers at least three distinguishable enterprises: *curriculum policy making*, the establishment of limits, criteria, guidelines, and the

like with which curriculum must comply, without developing actual plans and material for use by students and teachers; *generic curriculum development*, the preparation of curriculum plans and material for use potentially by any students or teachers of a given description; and *site-specific curriculum development*, the many measures taken in a particular school or district to bring about curriculum change there. (Pinar *et al.*, 1995, 665)

The reconceptualist advance reveals a long-standing debate in the field, between theorists and practitioners. Walker and Soltis (2004) consider this debate over curriculum development by going to its source: Schwab's (1970) landmark essay, *The Practical: A Language for Curriculum*. Walker and Soltis observe that

Schwab states flatly that the curriculum field is "unable, by its present methods and principles, to continue its work and contribute significantly to the advancement of education. . . . [It] has reached this unhappy state by inveterate, unexamined, and mistaken reliance on theory.' The field will only recover . . . "if curriculum energies are in large part diverted from theoretic pursuits . . . to three other modes of operation . . . which differ radically from the theoretic . . . the practical, the quasi-practical, and the eclectic" (2004: 59).

These three alternatives, Walker and Soltis observe, are practical paths of approach that reorient the field back to practice, deliberative paths of action that may be taken in "specific situations" (2004: 59). Theories, they argue, have merit only for their "applicability and usefulness in a particular case" (Walker and Soltis, 2004: 59).

Null (2006) places Schwab's *Practical* within a "deliberative tradition" that has both a classical and contemporary pedigree. In Null's introduction to a new edition of William A. Reid's *The Pursuit of Curriculum*, he places Schwab and Reid in a philosophical lineage that reaches back to Aristotle's *Ethics*. He notes that the distinction Schwab makes between the theoretic and the practical is simply a restatement of Aristotle's distinction in the *Ethics* between intellectual and moral virtue, how each play a complementary role in shaping and informing the deliberative acts of everyday life (Null, 2006: xiv). Null, however, adds little to easing the debate in the field that Schwab's challenge poses. In a direct reference to the 1995 opus, *Understanding Curriculum*, Null coolly states that "instead of the development of understanding, the purpose of practical inquiry is to arrive at a decision about what we should do at a particular time and within a particular context" (Null, 2006: xiv).

Another alternative path of curriculum discourse analysis that moves beyond this narrowly defined debate is suggested by Posner's (1995) *Analyzing the Curriculum*. Posner transcends the political dichotomy between theory and practice, by advancing an approach to the curriculum

that he calls: "Reflective Eclecticism," which wisely makes the two elements complementary. In the terms of a discourse perspective: the theoretic and the practical are recursive aspects of Null's deliberative intellectual process, but in a more open transformative field of discursive production.

Posner accepts the power of theory, but argues that critical analysis of curriculum requires an eclectic approach to theory (Posner, 1995: 255). Posner cites five "perspectives: the traditional, experiential, structure of the disciplines, behavioral, and cognitive," but is careful to note that these only "summarize, many, but certainly not all, approaches curricula may take" (1995: 46). Posner's conception of a perspective and the reconceptualists idea of viewing curriculum as text may simply be an academic one. Within Posner's alternative framework of reflective eclecticism, the discourse of modernism and the post-modern perspective represented in *Understanding Curriculum* (1995) appear to comfortably coexist. In the reconceptualized view of the field, curriculum as discursive text is mapped in multiple perspectives Pinar *et al.*, 1998: vi–x):

- Historical Text
- Racial Text
- Gender Text
- Phenomenological Text
- Post-structural/Deconstructed/Postmodern Text
- Autobiographical/Biographical Text
- Aesthetic Text
- Theological Text
- Institutionalized Text
- International Text.

Curriculum as International Discourse

The historical account rendered up to this point is largely an account of pivotal developments in the American field. The narrative arc may be characterized as one of extension and consolidation. The intent of the reconceptualists was to move their understanding of curriculum to an international stage and forum. Curriculum comes of age as discourse within the international field in the decade of the 1990s. A major expression of this development is the 1992 publication of the magisterial *Handbook of Research on Curriculum* edited by Phillip W. Jackson and produced by the American Educational Research Association (AERA). Pinar *et al.* (1995) rely heavily on Jackson (1992) in the opening chapters of *Understanding Curriculum*, which frame curriculum as a historical text (see Section II, chapters 3 and 4, pp. 67–240). If the *Handbook of Research on Curriculum* represents the paradigmatic consolidation of curriculum as a national discourse, its exhaustive structural mapping of the discipline still situates the field

firmly within the modernist confines of the nation-state system and the research and development (R&D) complex of the American university.

The AERA project on curriculum provided a model that the reconceptualists would reproduce, albeit in international terms a decade later. The *International Handbook of Curriculum Research* (2003) “represent[ed] the first move in postulating an architecture of a worldwide field of curriculum studies” (p. 1). Another publication appearing that year, which also heralds the consolidation of curriculum as discourse is *The Internationalization of Curriculum Studies* (2003), which is a volume of “selected proceedings from the LSU Conference 2000.” The conference sponsors traced the idea for the project to a conference held at the University of Oslo, Norway, in August 1995. The intent of the conference sponsors is indicated in their conference publication, *Didaktik and/or Curriculum: An International Dialogue* (Gundem and Hopmann, 1998).

Like the later Louisiana State University (LSU) conference, *Didaktik* was an early articulation by the field that considers the advent of globalization and its impact on education systems worldwide. It views these developments as an opportunity for dialog “across national borders” on “curriculum research and development” (1998: 1). The conference sponsors saw a definite “obstacle” to international “cooperation” and “inter cultural communication” because of “differing attitudes” to curriculum “planning and implementation,” more generally, development (1998: 1). The conference explored the complex national and cultural differences between just two “basic models” and “sets of attitudes” for teacher education. The two models are the North American “Anglo-Saxon tradition of *curriculum studies*,” and the “Central and North European tradition of *Didaktik*” (1998: 1). The term *Didaktik* has no literal equivalent in English and its precise meaning is refracted through regional linguistic and cultural idioms of Scandinavian and German pedagogical institutions. In its elemental sense, the term *Didaktik* is meant to convey a “comprehensive intertwining of action, reflection, practice and theory” (1998: 2).

By contrast, the 2003 *International Handbook of Curriculum Research* frames the problem of international dialog in terms of power. The problem is compounded not just by linguistic and cultural diversity, but by the political structures that tend to undercut the possibility of a democratic dialog (Trueit, 2003: xii). In his opening remarks to the LSU conference, William Pinar, the leading reconceptualist in the field, addressed the attendees by referring to America as the narcissistic center of the field. If American dominance is a threat to a level playing field of curricular discourse, so too, in Pinar’s opinion, is globalization in its perceived rush to reduce curriculum studies to a new market (Pinar, 2003: 5).

Nevertheless, the LSU Conference, which was held at Louisiana State University in April, 2000, was used as a platform to launch the first international association of

curriculum scholars (Pinar, 2003: 1–2). On 30 April 2000, the first organizational meeting of the International Association for the Advancement of Curriculum Studies (IAACS) was held. At that meeting, the mission of the association was articulated. The mission of the IAACS is:

To support a worldwide—but not uniform—field of curriculum studies [that does not mirror] the standardization and uniformity the larger phenomenon of globalization threatens. . . . Our hope, in establishing this organization, is to provide support for scholarly conversations within and across national and regional borders about the content, context, and process of education, the organizational and intellectual center of which is the curriculum.

The discourse of academic research presented in the *International Handbook of Curriculum Research* is intended to stand as a bulwark against economic globalization and the standardization of curriculum studies. A major force for this standardization is the international educational publishing industry. The historical account of the reconceptualists view that advances an understanding of curriculum as discursive text, in hindsight may be understood as a major development in the field.

However, in a Foucauldian sense, as discourse, it is already radically historicized, bearing in its version of the field, the same internal contradictions of the modernist paradigm. Up to this point, two successive phases describing major developments in the field of curriculum studies have been represented: the national advance of the reconceptualist paradigm and its transnational dissemination and diffusion. This is less discourse analysis in a classical sense (e.g., as practiced by Fairclough and Gee), than it is a particular reconstruction of these developments through a selective examination of seminal documents and publications. The method is basic historiography, using narrative (a form of discourse) to represent the theoretical turn to discourse in temporal terms. The emphasis on narrative form is deliberate because it highlights its discursive construction.

The emphasis placed on narrative form highlights the complex interaction of form and content in textual representation. The concluding section of this article focuses on the denouement of postmodern trends in curriculum discourse by considering new formal conditions for narrative production that may be changing the order of curriculum discourse that the reconceptualists have labored to produce.

Curriculum as Digital Discourse: The Virtual Materiality of Textbooks

Since the advent of the public Internet in May 1995, knowledge can no longer adequately serve as an organizer

for curriculum. There is no means by which 125 000 new books or 300 million websites each year can be assessed, shaped, or organized. There is no way that the instant retrieval of all information can be filtered or controlled. The boundaries of our understanding, in all fields, have been overrun; there is neither scope nor sequence to the new information age. (Wiles and Bondi, 2007: 298)

The print version of the *International Handbook of Curriculum Research* (2003) where the IAASC mission statement appears, concludes not with a citation to a printed source, but with a URI: a Universal Resource Identifier which begins with the string of letters: <http://> (the Hypertext Transfer Protocol), and the now ubiquitous code: WWW (W3), representing the World Wide Web or more colloquially the Web (Berners-Lee (1999); W3C-World Wide Web Consortium.). The URI identifies the database source on the Web and <http://> code the computer protocol for transferring and connecting with that information source.

Between the publication dates of *Understanding Curriculum* (1995) and the *International Handbook* (2003) is a mere 8 years, but its paradigmatic significance for this account of curriculum as discourse may be portentous. With the advent of the Internet, this string of coded letters signals a vast fast-moving transition away from the cultural conditions that made the reconceptualists account of curriculum possible. An additional sign of this change may be discerned in the curricular order of things found in the print version of the *International Handbook* in *Part II: Thirty-Four Essays on Curriculum Studies in 28 Nations*. The organization of the 2003 *International Handbook* arranges the essays alphabetically starting with two on Argentina, followed by Australia, Botswana, Brazil (for which there are three entries), Canada, China, etc.

Other than A–Z, no discernable relation between the nations (presented in alpha order) as exemplars of curriculum discourse is apparent. Several conventions for ranking are obvious but inadequate: By region/continent, First World/Fourth, developing/postindustrial. The absence of a compelling form of exposition to present an international perspective on curriculum studies suggests that the medium adequate to its representation may only be in a protean stage. If the subject index of the *International Handbook* is correct, the use of the term Internet only appears once in 665 pages of text.

The synoptic textbook figures prominently in *Understanding Curriculum* (1995), but its authors have little to say that is positive about the second edition (1984) of the Wiles and Bondi textbook (quoted above in its 2007 edition). For them it represents the standardization and uniformity of the global educational publishing industry, the antithesis of curriculum as discourse. Nevertheless, Wiles and Bondi share a common perception that the Internet is transforming the textbook. Is it also a sign that the linear

narrative form of the 2003 *International Handbook* is inadequate to a new order of discourse emerging within digital culture? As the educational publishing industry is grappling firsthand with the transformation of print media by digital technology, it is one site where these conditions are prevalent. As a consequence, it provides a crucial site for curriculum research and possibly the terms of an emerging curriculum discourse.

In a series of studies tracking changes in the educational publishing industry that parallel the rise of the Internet, LaSpina (1998, 2001, 2007) examines the impact of digital technology on the design of social studies textbooks. Rather than focus on the instructional nexus of computers and textbooks, LaSpina begins with the textbook publishing design process and examines what happens when computers are introduced into that process. LaSpina observes that when computer-based graphic design is used in development of instructional textbooks, it transformed the linear narrative conventions of the expository textbook. When the narrative form changed, so too did the textual organization and structure of content in the textbook. LaSpina argues that changes to the formal order of text representation portend a larger transformation, the transition from page to screen.

The dematerialization of printed text and its virtual reconstitution is a sign of a new curriculum settlement, unapparent within the present order of curriculum discourse. The historical narrative of the transition to a postmodern paradigm of curriculum as discourse presented in this article, is meant to emphasize the historical method as a discursive practice to complement other approaches to discourse analysis, and to present new ways of seeing curriculum that are only understandable in the context of historical time.

See also: An Overview of Research in Curriculum Inquiry; Curriculum and Poststructuralist Theory; Curriculum and the Publishing Industry; Curriculum, Digital Resources and Delivery; Globalization and Curriculum; Textbook Development and Selection.

Bibliography

- Apple, M. W. (1979). *Ideology and Curriculum*. New York: Routledge.
- Doll, W. E. (1993). *A Post-Modern Perspective on Curriculum*. New York: Teachers College Press.
- Fairclough, N. (1992). *Discourse and Social Change*. Cambridge: Polity Press.
- Gee, J. P. (1999). *An Introduction to Discourse Analysis*. London: Routledge.
- Gundem, B. B. and Hopmann, S. (1998). *Didaktik and/or Curriculum*. New York: Peter Lang.
- Hall, S. (2001). Foucault: Power, knowledge and discourse. In Wetherell, M., Taylor, S., and Yates, S. J. (eds.) *Discourse Theory and Practice: A Reader*, pp 72–81. London: Sage.
- Jackson, P. W. (ed.) (1992). *Handbook of Research on Curriculum* (AERA). New York: Macmillan.

- Karabel, J. and Halsey, A. H. (eds.) (1977). *Power and Ideology in Education*. New York: Oxford University Press.
- Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.
- LaSpina, J. (1998). *The Visual Turn and the Transformation of the Textbook*. Mahwah, NJ: Erlbaum.
- LaSpina, J. (2001). The locus of language in digital space. *Language Arts, Journal for the National Council for Teachers of English (NCTE)* **78**, 245–254.
- LaSpina, J. (2007). The digital page: Crossing boundaries of print in California social studies. In Openshaw, R. and Soler, J. (eds.) *Reading across International Boundaries: History, Policy and Politics, International Perspectives on Curriculum Series*, pp 131–154. Charlotte, NC: Information Age Publishing (IAP).
- Null, J. W. (2006). Introduction and postscript to Reid, W. A. *The Pursuit of Curriculum: Schooling in the Public Interest*. Charlotte, NC: Information Age Publishing (IAP).
- Paras, E. (2006). *Foucault 2.0: Beyond Power and Knowledge*. New York: Other Press.
- Pinar, W. F. (2003). Introduction: Toward the internationalization of curriculum studies. In Trueit, D., Doll Jr., W.E., Wang, H., and Pinar, W. F. (eds.) *The Internationalization of Curriculum Studies: Selected Proceedings from the LSU Conference 2000*, pp 1–13. New York: Peter Lang.
- Pinar, W. F., Reynolds, W. M., Slattery, P., and Taubman, P. M. (1995). *Understanding Curriculum*. New York: Peter Lang.
- Popkewitz, T. S. (2001). The production of reason and power: Curriculum history and intellectual traditions. In Popkewitz, T. S., Franklin, B. M., and Pereyra, M. A. (eds.) *Cultural History and Education: Critical Essays on Knowledge and Schooling*, pp 151–183. New York: Routledge Falmer.
- Posner, G. F. (1995). *Analyzing the Curriculum*, 2nd edn. New York: McGraw-Hill.
- Potter, J. and Wetherell, M. (2001). Unfolding discourse analysis. In Wetherell, M., Taylor, S., and Yates, S. J. (eds.) *Discourse Theory and Practice: A Reader*, pp 198–209. London: Sage.
- Reid, W. A. (2006). *The Pursuit of Curriculum: Schooling in the Public Interest* (Edited with an Introduction and Postscript by Null, J. W.). Charlotte, NC: Information Age Publishing (IAP).
- Rogers, R. (ed.) (2004). *An Introduction to Critical Discourse Analysis in Education*. Mahwah, NJ: Erlbaum.
- Schwab, J. J. (1970). *The Practical: A Language for Curriculum*. Washington, DC: NEA.
- Slattery, P. (1995). *Curriculum Development in a Postmodern Era*. New York: Garland.
- Trueit, D., Doll Jr., W. E., Wang, H., and Pinar, W. F. (eds.) (2003). *The Internationalization of Curriculum Studies: Selected Proceedings from the LSU Conference 2000*. New York: Peter Lang.
- Van Dijk, T. (2001). Principles of critical discourse analysis. In Wetherell, M., Taylor, S., and Yates, S. J. (eds.) *Discourse Theory and Practice: A Reader*, pp 300–317. London: Sage.
- Walker, D. F. and Soltis, J. F. (2004). *Curriculum and Aims*, 4th edn. New York: Teachers College Press.
- Wetherell, M., Taylor, S., and Yates, S. J. (2001). *Discourse Theory and Practice: A Reader*. London: Sage.
- Wiles, J. and Bondi, J. (2007). *Curriculum Development: A Guide to Practice*, 7th edn. New York: Pearson Merrill-Prentice Hall.

Relevant Websites

- <http://www.iaacs.org> – The International Association for the Advancement of Curriculum Studies.
- <http://www.w3.org> – World Wide Web Consortium: Web Standards.

Curriculum Use in the Classroom

M Ben-Peretz and B Eilam, University of Haifa, Haifa, Israel

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Commonplaces – They refer to the main issues and topics treated by curriculum developers in the development process: teacher, learner, subject matter, milieu (Schwab, 1973).

Curriculum use may be defined as the ways in which curriculum materials are employed, applied, treated, and even manipulated by several stakeholders in the educational world. Curricular materials mentioned are: syllabi; teacher handbooks or guidelines; textbooks; additional instructional materials, such as worksheets, films, tapes, and so forth; and materials for student assessment, tests, and other kinds of examination modes, such as questionnaires measuring attitudes toward specific subject-matter areas.

In this article, stakeholders are conceived of as representatives of three of the four commonplaces that Schwab (1973) suggested as the three main realms of curriculum, namely, learner, teacher, and milieu. The milieu is represented in this article by members of society. The mode of curriculum use by these stakeholders has far-reaching consequences. The article begins with the presentation of each stakeholder's mode/s of curriculum use based on relevant literature.

The modes of use in each case are analyzed on a continuum from low to intensive engagement with the curriculum. Low engagement refers to little active involvement, either as learners, teachers, or members of society. Intensive engagement means active involvement in decision making concerning the nature of curriculum materials. After presenting the uses of each stakeholder, an integrative, comprehensive model for analyzing curriculum use is presented, accounting for the nature of users and intensity of engagement with curriculum materials. In order to focus explicitly on the separate modes of use of each stakeholder, the interactions and linkages among the three are related throughout the text. Such linkages may have an impact on curriculum use. For instance, there may be a very strong demand by societal agents such as a Ministry of Education that teachers use National Curriculum as it is planned; or teachers' use may be shaped by learners' nature, for instance, in the case of new immigrants at schools.

Teachers as Curriculum Users

The following are possible distinct curriculum uses by teachers (Ben-Peretz, 1990). From research conducted in the 1980s on the way in which teachers used curriculum materials, the following modes were found to articulate different uses.

Teachers as Users of Teacher-Proof Curriculum

In this mode, teachers are expected to use the planned curriculum explicitly as it is prescribed. In order to achieve this goal, the curriculum presents teachers with minute details of the intended use.

Teachers as Curriculum Choice Makers

In this mode, teachers apply their own principles, criteria, and awareness of concrete circumstances to make decisions about the specific curriculum content, such as part of the textbook, to use in their teaching. Oftentimes, constraints constitute a strong component in this process. Another issue is the pedagogical tension between breadth versus depth in teaching any subject-matter area. Teachers may also be influenced in their choices by their understanding of the nature of students, their abilities, interest, or prior knowledge.

Teachers as Curriculum Adaptors

Teachers may act as curriculum adaptors who change the curriculum materials in response to the perceived needs of their students, or in line with their own interest and knowledge base. Fullan (1982) mentions the tremendous power teachers have to introduce any changes they see appropriate in the process of curriculum implementation. Teachers' adaptation of curriculum, and their attempts to introduce changes in existing texts, raises the problem of adherence to curricular guidelines or to curriculum materials. How far may teachers go in their adaptations without destroying the spirit and meaning of the planned curriculum they implement in their classes? Curriculum analysis may help teachers recognize the special characteristics of curriculum materials. Such insights may aid teachers in their efforts to interpret materials and to plan their lessons on the basis of this interpretation. The term curriculum

envelope has been suggested for describing the special characteristics of the planned curriculum. Inside this envelope, teachers may plan their own mode of use.

Teachers as Creators of Curriculum

Finally, teachers as creators of curriculum may create their own curriculum materials, based on specific general guidelines or syllabi, or completely independently. Teacher-based curriculum development may be called forth for a number of reasons. Special needs of student groups may require individualized curriculum materials which are not to be found commercially. The faculty of a school may decide to try to react to some societal problems, such as a high incidence of teenage pregnancies, through the local development of a course on family life and family planning. Because of the great sensitivity of such issues, it is necessary to bear in mind the specific beliefs and values of students, parents, and community while constructing such a course. Such school-based curriculum development (SBCD) may cover a whole school curriculum, or be confined to individual subjects or themes.

According to Marsh *et al.* (1990), “the term ‘school based curriculum development’ (SBCD) is used in various ways in the literature but typically as a slogan for devolution of control, for ‘grass-roots’ decision-making, and as a representation of the polar opposite of centralized education” (p. ix). Though there is a tendency nowadays to move toward more centralistic school systems, for instance, through enforcing standards, or the national curriculum in England, the practice of SBCD still seems to be occurring, and sometimes even flourishing. Marsh *et al.* claim that it is unlikely that schools will retreat to an isolated existence where parents and community have no say in curriculum matters.

At the site-level of SBCD, both personal and social processes have a strong impact on teachers’ use of curriculum. Fullan (1982) points out that beliefs guide and are informed by teaching strategies and activities and Miller and Seller (1985) go on to claim that personal factors affect implementation and use of curriculum materials. Miller and Seller (1985) argue that too often, implementation is centered on things, such as textbooks, teaching aids, and explanatory booklets. However, implementation is not as one-sided as it is often portrayed; rather, it is an interactive process during which the teacher adapts the program to his/her educational context.

The Notion of Curriculum Potential

One of the ways in which teachers gain ownership of the curriculum is through their imaginative use of curriculum potential (Ben-Peretz, 1990). Although curriculum materials are often perceived as the expression of their

developers’ intentions, Schwab (1973) warns us that these intentions convey the values of the developers only imperfectly and merely suggest ways of constructing teaching activities. Actual classroom experiences of the curriculum might serve to reduce the ambiguity of the stated intentions as well as modify them. Implementation of ideas and activities, which are proposed in the curriculum, turns these into concrete experiences. These experiences may correspond to the perceived intentions of the developers, but they may also serve different ends. It is important to remember that the notion of curriculum potential is dependent on the interaction between teachers and materials. Materials offer starting points, and teachers use their curricular insights, their pedagogical knowledge, and their professional imagination to develop their own curricular ideas on the basis of existing materials. The scope, variety, and richness of the curriculum potential embodied in materials are determined by the wealth of their content and the flexibility or rigidity of their structure. Good curriculum materials have many different potentials for diverse educational situations.

In terms of the continuum from low to intensive engagement with the curriculum, here the movement is from being users of teacher-proof curriculum to teachers as curriculum planners.

Learners as Curriculum Users

One of the most critical phases of the curriculum transformation model, suggested by Goodlad *et al.* (1979), is curriculum as experienced by students. Usually, the experienced curriculum is evaluated through measuring students’ achievement. The notion of learners’ experiences is expanded to include the process of actual interaction between students and curriculum materials such as understanding of text or manipulation of tasks. Students may play a role in evaluating curriculum materials. Overall, it is deemed important to give students a voice in curriculum matters. Frequently, classroom context determines to a large extent the role of students as curriculum users. The interaction of students with textbooks cannot be viewed as an entirely passive process. There is an implicit assumption that meaning is determined by the text itself. This approach neglects the reality of interpretation of textbooks by students. “Depending on social context and social psychological variables such as ethnicity, gender, social class, personality and psycho-pathogenesis, individual pupils may reach vastly different interpretations or decodings of one and the same text” (Kalmus, 2004: 470–471). These interpretations by students may be conceived as an important component of classroom curriculum use.

One of the modes of students’ interacting with the curriculum is in the form of a community of discourse.

Interpretive community is one in which constructive discussion, questioning, and criticism are the rule rather than the exception. Students and teachers alike have ownership of knowledge and experience but no one has expertise in all areas. Members of the community share their expertise through reciprocal teaching and collaborative learning activities. In such an environment, students may act as teachers and discussion leaders, thus having an active role in curriculum use.

In a constructivist classroom, students are viewed as thinkers with emergent theories about the world and teachers seek the students' points of view in order to understand students' conceptions for use in teaching. In this sense, students become part of the planners of curriculum use. This active role of students is opposed to their role in traditional classrooms, where students are often viewed as blank slates onto which information is etched by the teacher. Typical constructivist instruction asks learners to play more of the task-management role than in conventional instruction.

Self-directed learning represents an independent mode of curriculum use, through assignment, theme, and topics for students to learn by themselves (e.g., problem-based learning, inquiry, projects). Independent and self-regulated learning usually takes place in the frame of the mandated school curricula. This kind of learning may focus on several key self-directed learning processes: defining what should be learned, developing learning goals, identifying a learning plan, successfully implementing it, and self-evaluating the effectiveness of learning. Thus, students almost develop a curriculum according to the Tylerian rationale.

Among the salient approaches to involving students in curriculum implementation is, for instance, design-based science curriculum, learning and teaching through inquiry, or project-based classroom learning. Open-inquiry instruction is beneficial in engaging students at all levels of achievement in the learning process (Yerrick, 2000). "The 'just the basic' mentality and compromising science by watering down the curriculum for lower track students continue to perpetuate a learned helplessness for these students regarding scientific knowledge" (Yerrick, 2000: 231). In this case, allowing students to play a highly active role in the curriculum is a prerequisite of successful curriculum implementation.

There is an important message in these findings for the development of educative curriculum materials. In order to help students learn in the most productive ways, and be successful curriculum users, curriculum materials need to be designed to support teachers in engaging in particular teaching-learning processes and in addressing particular learning challenges. Teachers have to become aware of the positive impact of giving students an active role in the implementation of curriculum.

Homework may be conceived of as an extension of students' curriculum use outside the frame of school.

Almost all children at school all over the world are asked to do some kind of homework. In this sense, the curriculum comes home, where it is used, and not only in the classroom. There is generally consistent evidence for a positive influence of homework on achievement, especially in grades 7–12.

The movement here in terms of the continuum from low to intensified engagement with the curriculum is from students reading prescribed texts to students as independent learners in the inquiry mode.

Society as a Curriculum User

Society constitutes an important stakeholder. As stated above, society, the milieu commonplace includes members of society such as diverse societal groups, policy-makers, and parents. However, unlike the direct use of curriculum by teachers and students in the classroom, curricula serve to mediate between society and its goals and are used for achieving diverse societal goals and agenda. Such goals are constantly changing in accordance with the various contexts, local or global, as they are perceived in different times. Society and culture affect classroom use of curriculum as far as related to curriculum content and proposed goals. "Increased literacy demands and diversity in student populations are only two of several social changes affecting the scope and complexity of curriculum use in classrooms" (Sowell, 2000: 110). An example for curriculum fostering societal goals is the response of educators in the USA to the competition in space technology with the Soviet Union.

As long ago as the 1940s, Tyler regarded society as one of the sources of curriculum objectives, viewing the curriculum as serving societal needs. Curriculum scholars from the 1960s onward challenge the perception of society as a single entity, and called for a re-conceptualization of curriculum theories, which "always carry a privileging dimension that serves some groups and individuals, while simultaneously harming others" (Hlebowitsh, 1999: 345). They call for an increased sensitivity to social contexts and issues of race, ethnicity, gender, or social status, moving away from the imposing of control of one societal group over others (Hlebowitsh, 1999).

Social Class as Stakeholder

Curriculum dialog was enriched, becoming also more controversial, by the deepening and widening of perspectives regarding its use in schools and the revelation of the various mechanisms working behind curriculum use. Already, in 1981, in her well-known study, Jean Anyon called attention to school reproduction of unequal class structure (Anyon, 1981). Anyon demonstrated subtle

differences in the curriculum in use among schools situated in contrasting social-class settings. What counts as knowledge in these schools was found to be different along dimensions of structure and content. Curricula in the working-class schools were only partially used in comparison to elite schools, where high cognitive demands were placed on learners. By situating school knowledge in the particular social setting, Anyon showed how it contributes to contradictory social processes of conservation and transformation. Schools play a central role in both changing and reproducing social and cultural inequalities along generations.

Curriculum as Political Text

School was not perceived anymore as detached from ideologies or politics; rather, the term of ideological hegemony was coined together with the resulting belief that only a critical approach would undermine the reproductive forces and present individuals with alternatives (McLaren, 1989). Curriculum was perceived as a political text, aiming at the perpetuation of domination of the rich and strong over the weak and poor in society. The concept of the hidden curriculum has been used to describe all the implicit and tacit assumptions and ways in which knowledge influences behaviors and thoughts (Jackson, 1968). It is well understood that books, images, and media used in schools perpetuate negative racial and gender-related attitudes.

A recent example of a society's use of curriculum to respond to political needs is post-apartheid South Africa (Fataar, 1997). Placement of two million children in schools was viewed as central to rebuilding the country. The curriculum had to be changed from a colonial white curriculum to a post-colonial indigenous one. A policy of quantitative expansion of schooling should not ignore the quality of schools; otherwise, education becomes a factor contributing to social inequality rather than to social reconstruction.

Curriculum as Economic Text

The social dimension of curriculum use, societal utilization of curriculum in the classroom to achieve its own goals, may be best exemplified by literature reports concerning the preparation of students for labor in the current globalization and technological age. Although it is impossible to predict the future, one may recognize some core trends against which societal agendas may be evaluated. It is often indicated that characteristics of work are greatly changing especially concerning the demands for skills and technology, due to the rapid pace of changing economy and technology. Standards of efficiency, productivity, and quality are raised due to global competition

(Bishop, 1990; Smith, 1999). Various countries formed committees and initiatives to deal with employment-related key competencies while reevaluating and adapting national educational goals to face these changes (Smith, 1999). In the USA, national economic/educational goals were proposed in response to the skill gap. However, the debate concerning issues dealing with the need for preparing students for higher demands of workplace, and with the ability to transfer school-acquired procedural knowledge to the workplace, or the relations between education and worker productivity, continues.

In the past, some researchers suggested no necessary connections between education and workplace productivity (Bishop, 1990) or argued that such relationships between workplace characteristics and cognitive demands of workers were more complex and less direct (Smith, 1999). Kress (2000) suggested that demands of the coming era require education for instability rather than being directed for a certain disposition and cultural reproduction. He claims that present curricula in most Western states were developed to needs of nineteenth-century school, with its desire for "a homogeneously conceived citizen for that state" (p. 134). He calls for the dissolution of former frames and the emergence of new framings (Kress, 2000).

Parents as Stakeholders

Among the societal stakeholders in the curriculum domain, parents tend to play a central role (Klicka, 1998; OECD, 1997). Parents send their children to school so that they attain the necessary knowledge to function in the societal context and the labor market. The school curriculum and its uses are vehicles for achieving these goals. How much, if any, impact do parents have on the development and use of curriculum? Involvement of parents in curriculum development is usually lacking except in special cases of SBCE. On the other hand, parents' involvement in curriculum uses may be rather significant. Parents may play a role in assisting their children in school tasks, such as homework or projects. Usually their involvement reflects the provision of home support to their children in homework performance or learning for exams. Parents may also act as supporters, or obstacles, in the path of curriculum innovation.

Parents' involvement in the curriculum process, especially in reform movements, requires explaining the new approach of teaching and learning involved in the innovation. The most active and involved role of parents in the curriculum domain is in the context of home schooling (Klicka, 1998). Home schooling means that parents adopt the role of teachers including the planning of curriculum materials. Parents rely on numerous educational curricula and resources that are available to them. Independent

studies were found that, on an average, home-schooled children score above average on standardized achievement tests. This outcome reflects the skillful use of existing curricula adapting them to specific teaching situations.

The relationship between parental involvement and educational outcomes exists regardless of students' socioeconomic or race/ethnic background and regardless of whether parental practices occur in the middle grades or in high school (Catsambis, 2001). Parents as stakeholders come from different social, religious, ethnic, racial, or economic groups, drawing on different value systems, and relating differently to their children or to schools. It is certain that they all want the best for their children, but cannot agree on what best is, making it difficult to uniformly conceptualize parents' involvement in the curriculum domain.

A number of research projects looked at the relations between schools and parents in different social groups – immigrant parents and low socioeconomic status showed that active involvement of parents in school activities promotes students' learning and achievements. School activities may be in the form of classroom volunteers, being an integral part of the decision-making process at school, for instance, in curriculum matters. Parents, as educational partners, no longer feel intimidated by teachers but voice their opinions and suggestions for the school curriculum.

The term parent power refers to increased parental activism, expressed in increased numbers of parents' participation in various organizations and associations. It reflects parents' ability to pressure educational systems

to enact a curriculum that responds to their group needs. A current example is parents' involvement in the curricular debate concerning the teaching of evolution versus teaching creation (intelligent design). Present-day governments tend to support and legislate for parents' collective voice, expressed in parents' representation on policymaking bodies at the various levels and parents associations. An interesting survey has shown that one in three students learn in a school where his/her parents may influence the curriculum (OECD, 1996). Hoover-Dempsey and Sandler (1997) argued that three major constructs are center to parents' involvement in curricular decisions: (a) teachers' beliefs about parents' roles in their children's education, (b) parents' sense of efficacy for helping their children succeed in school, and (c) the perception that both child and school want parents to be involved.

In the case of society as a curriculum user, the move from low to high intensity of engagement has to be analyzed separately for each of the sub-groups mentioned above.

An Integrative Model for Analyzing Curriculum Use

Based on the analysis of each of the stakeholders' interactions with the planned curriculum, a model for viewing the complexities of curriculum use and the transformation of the planned into the enacted curriculum is presented (Figure 1). This model is presented in a visual

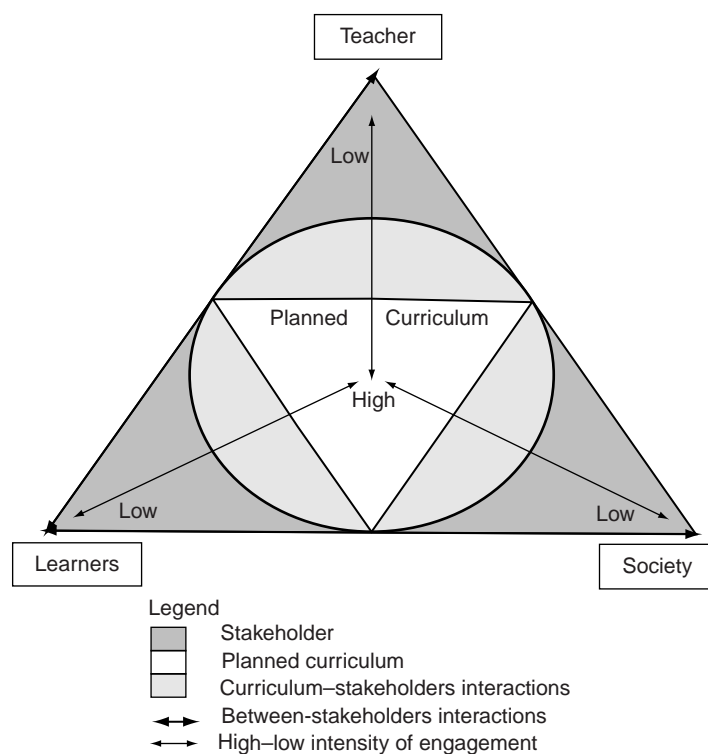


Figure 1 A model for analyzing curriculum use.

form, because the process of curriculum use is conceived as simultaneously involving three different stakeholders.

The model consists of a triangle, each apex representing one of the three stakeholders: teacher, learner, society. The teacher is placed at the top, because teachers are considered as main decision makers in curriculum implementation. A circle within the middle part of the triangle represents the planned curriculum. The three shaded parts of the model are formed by the interaction between the stakeholders and the curriculum, and represent the enacted curriculum. In order to remind readers of interactions among stakeholders themselves that potentially have impact on curriculum use (as mentioned before in the text but are not the focus of this article), two-way directional arrows are marked on the three sides of the triangle. The continuum from low to high intensity of engagement is marked by three arrows going from each stakeholder toward the planned curriculum. The closer a point on this continuum (arrow) is to the curriculum, the higher the intensity of engagement.

The novelty of this model presented here relates to two aspects of curriculum use: (1) viewing three different users whereas usually teachers alone are considered, and presenting their use as simultaneously enacted, and (2) conceiving of curriculum use as a continuum reflecting intensity of engagement – from low to high intensity.

This model has the potential to be used in studies of curriculum implementation and in the process of introducing teachers to the curriculum domain. The distinction between low and high level of intensity of users' engagement raises research questions that can be studied in relation to each of the subgroups of society, for instance, in relation to immigrant parents. Such a view of engagement allows curriculum researchers to go beyond the measuring of students' achievement as major factors in curriculum use. The continuum from high to low intensity of engagement provides refined insights into the understanding of stakeholders' impact on curriculum use and may enhance their possible involvement.

Bibliography

- Anyon, J. (1981). Social class and school knowledge. *Curriculum Inquiry* **11**, 3–42.
- Ben-Peretz, M. (1990). *The Teacher–Curriculum Encounter: Freeing Teachers from the Tyranny of Texts*. Albany, NY: State University Press.
- Bishop, J. H. (1990). The productivity consequences of what is learnt in high school. *Journal of Curriculum Studies* **22**, 101–126.
- Catsambis, S. (2001). Expanding knowledge of parental involvement in children's secondary education: Connections with high school seniors' academic success. *Social Psychology of Education* **5**, 149–177.
- Fataar, A. (1997). Access to schooling in a post-apartheid South Africa: Linking concepts to context. *International Review of Education* **43**, 331–348.

- Fullan, M. (1982). *The Meaning of Educational Change*. New York: Teachers College Press.
- Goodlad, J. I., Klein, M., and Tye, K. (1979). The domains of curriculum and their study. In Goodlad, J. I., Ammons, M. P., Buchanan, E. A., et al. (eds.) *Curriculum Inquiry: The Study of Curriculum Practice*, pp 43–76. New York: McGraw-Hill.
- Hlebowitsh, P. (1999). The burdens of the new curricularist. *Curriculum Inquiry* **29**, 343–354.
- Hoover-Dempsey, K. V. and Sandler, H. M. (1997). Why do parents become involved in their children's education? *Review of Educational Research* **67**, 3–42.
- Jackson, P. (1968). *Life in Classrooms*. New York: Holt, Rinehart and Winston.
- Kalmus, V. (2004). What do pupils and textbooks do with each other? Methodological problems of research on socialization through educational media. *Journal of Curriculum Studies* **36**, 469–485.
- Klicka, C. J. (1998). *The Right to Home School: A Guide to the Law on Parents' Rights in Education*, 2nd edn. Durham, NC: Carolina Academic Press.
- Kress, G. (2000). A curriculum for the future. *Cambridge Journal of Education* **30**, 133–145.
- Marsh, C., Day, C., Hannay, L., and McCutcheon, G. (1990). *Reconceptualizing School Based Curriculum Development*. Hampshire, UK: Falmer.
- McLaren, P. (1989). *Life in Schools: An Introduction to Critical Pedagogy in the Foundations of Education*. New York: Longman.
- Miller, J. P. and Seller, W. (1985). *Curriculum: Perspectives and Practice*. New York: Longman.
- OECD (Organization for Economic Co-Operation and Development) (1996). *Education at a Glance*. Paris: The Center of Educational Research and Innovation.
- OECD (Organization for Economic Co-Operation and Development) (1997). *Parents as Partners in Schooling*. Paris: The Center of Educational Research and Innovation.
- Schwab, J. J. (1973). The practical: Translation into curriculum. *School Review* **81**, 501–522.
- Smith, J. P., III (1999). Tracking the mathematic of automobile production: Are schools failing to prepare students for work? *American Educational Research Journal* **36**, 835–878.
- Sowell, E. J. (2000). *Curriculum: An Integrative Introduction*, 2nd edn. Upper Saddle River, NJ: Merrill.
- Yerrick, R. K. (2000). Lower track science students' argumentation and open inquiry instruction. *Journal of Research in Science Teaching* **37**, 807–838.

Further Reading

- Brown, A. and Campione, J. C. (1998). Designing a community of young learners: Theoretical and practical lessons. In Lambert, N. M. and McCombs, B. L. (eds.) *How Students Learn: Reforming Schools through Learner-Centered Education*, pp 153–186. Washington, DC: American Psychological Association.
- Carreon, G. P., Drake, C., and Barton, A. C. (2005). The importance of presence: Immigrant parents' school engagement experiences. *American Educational Research Journal* **42**, 465–498.
- Connelly, M. F. and Clandinin, D. J. (1988). *Teachers as Curriculum Planners: Narratives of Experience*. New York: Teachers College Press.
- Cooper, H., Robinson, J. C., and Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research 1987–2003. *Review of Educational Research* **76**, 1–62.
- Dalton, D., Dingess, D., Dingess, C., et al. (1996). The parents as educational partners: Programs at Atenville elementary school. *Journal for Education for Students Placed at Risk* **1**, 233–247.
- Gellert, U. (2005). Parents: Support or obstacle for curriculum innovation. *Journal of Curriculum Studies* **37**, 313–328.
- Grant, C. and Sleeter, C. (1986). *After the School Bell Rings*. Philadelphia, PA: Falmer.

- Harker, R. (1990). Bourdieu – education and reproduction. In Harker, R., Mahar, C., and Wilkes, C. (eds.) *An Introduction to the Work Pierre Bourdieu: The Practice of Theory*, pp 86–108. London: Macmillan.
- Kelly, A. V. (1999). *The Curriculum: Theory and Practice*, 4th edn. London: Paul Chapman.
- Ortiz-Franco, L. and Flores, W. V. (2001). Sociocultural considerations and Latino mathematics achievement. In Atweh, B., Forgasz, H., and Nebres, B. (eds.) *Socio-Cultural Research on Mathematics Education: An International Perspective*, pp 233–254. Mahwah, NJ: Erlbaum.
- Rudduck, J. (1999). Teacher practice and the student voice. In Lang, M., Olson, J., Hansen, H., and Bunder, W. (eds.) *Changing Schools/ Changing Practices: Perspectives on Educational and Teacher Professionalism*, pp 41–54. Louvain: Garant.
- Tyler, R. (1949). *Basic Principles of Curriculum and Instruction*. Chicago, IL: University of Chicago Press.
- Zimmerman, B. J. and Lebeau, R. B. (2000). A commentary on self-directed learning. In Evensen, D. H. and Hmelo, C. E. (eds.) *Problem-Based Learning: A Research Perspective on Learning Interactions*, pp 299–313. Mahwah, NJ: Erlbaum.

The Roots and Routes of Teacher-Based Action Research and Curriculum Inquiry: An Historical Perspective

L G Putney, University of Nevada, Las Vegas, Las Vegas, NV, USA

J L Green, University of California, Santa Barbara, Santa Barbara, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

This article provides a historical perspective on developments underlying the relationship of teacher-based action research and curriculum inquiry. The exploration of the relationship begins with a historical journey tracing the roots and routes of different approaches to teacher-based or practitioner research across the past 100-plus years and moves on to factors related to their development. The historical analysis will show that the evolutionary path (Noffke, 1997) of teacher-based action research, curriculum inquiry, and practitioner inquiry is not an unbroken chain of developments, but rather a set of routes taken by researchers, educators, and practitioners at particular periods of time in response to particular intellectual, social, political, and educational conditions.

Reviews over the past six decades have shown that the approaches identified are known by various names, including action inquiry, action research, collaborative action research, emancipatory action research, ethnographic action research, narrative inquiry, participatory research, practitioner inquiry, practitioner research, teacher research, and teacher action research, among others (e.g., Cochran-Smith and Donnell, 2006; McKernan, 1987; Noffke, 1997; Wann, 1953; Zeichner and Noffke, 2001). The question framing this article, therefore, is what counts as teacher-based action research and curriculum inquiry across different periods of education research.

An ethnographic perspective guided this historical analysis (Green *et al.*, 2003) in order to identify the roots and routes of different approaches to practitioner inquiry, action research, and curriculum inquiry. This perspective framed the following questions, which are designed to make visible what gave rise to the approach, and what each was seeking to accomplish: Who are the practitioners? With whom are they engaged? When and where does the inquiry take place? What types of inquiry are undertaken? What purposes does the research serve for practitioners (and others)? What knowledge is generated? How is the knowledge used in practice? How is it shared with others? What processes are central to research that can be defined within the overarching categories of action and practitioner research?

These questions unearthed a common goal across the varied approaches: “the study of a social situation with a view to improving the quality of action within it” (Elliott, 1991: 69). These questions also made visible that those

developing the forms of action research often had one or more purposes: to develop knowledge for practice by practitioners (e.g., Winter, 1987) and/or to develop knowledge for informed social action by members of a community (Carr and Kemmis, 1986; Collier, 1945). Furthermore, in exploring the historical development of teacher-based action research and its intersection with curriculum inquiry, the goal was to identify the social, political, and educational conditions that led researchers and practitioners, at different periods of time, to take up this common goal uncommonly.

Uncovering Early Roots: Locating a Place to Begin

Elliott’s definition proved useful in uncovering the early roots of teacher-based action research; it focused the search on teachers studying their social situation, either alone or in collaboration with others. This framing of the problem immediately led to several reviews of literature, which mentioned forms of curriculum inquiry and teacher-based or action research, and which were used to identify the roots and routes taken in particular periods of time. A key review identified was the one by Kenneth Wann in 1953. He argued that the adaptation of action research methodology had a long and evolutionary history, marking the 1920s and 1940s as key points for major curriculum inquiry work to improve secondary schools. Wann (1953) noted, “The development of action research in schools represents an attempt to provide a research methodology which is suitable for study and solution of school problems in relation to the total social situation and which can be conducted by teachers as a part of their teaching activity” (p. 337). In concluding his review, he pointed to the need to examine how teachers are part of curriculum inquiry, and what supports and/or constrains such work. Wann also pointed to a need to examine the role of university-based courses and collaborations with teachers, and to a school’s capacity to engage in such inquiry, in order to gain a robust picture of teacher-based action research and curriculum inquiry.

Wann (1953) served as an anchor in time, driving searches both backward and then forward in time on Google and Google Scholar to identify additional reviews

and references to teacher research that predated those cited in the reviews identified. This mapping approach located a broad range of work in the US and abroad and also located historical books and articles that were not readily available in a search of our university libraries. Although the focus in the current article is on the US as an illustrative case, some of the routes presented include ones developed in the UK and Australia that had impact on US developments.

Route 1 – John Dewey and the Teachers of the University of Chicago Laboratory School (1896–1904)

In the 1899 edition of *The School and Society*, Dewey argued that teacher questions, knowledge, and actions in the laboratory school that he and his wife, Alice Chipman Dewey, founded (1896–1904) were the basis for developing theories of knowledge and inquiry. Dewey (1899) argued that “[t]he teachers started out with question marks and if any answers have been reached, it is the teachers in the school who have supplied them” (p. 116). The teachers’ questions served as a focus for discussions about the developing goals of the school and the curriculum focused on personal growth rather than mastery of a discrete body of knowledge – the curriculum prevalent at that time. In describing the goal of the school, Dewey argued that “[w]e have attempted to find out by trying, by doing – not alone by discussion and theorizing – whether these problems may be worked out, and how they may be worked out” (p. 120).

Dewey described these elementary school teachers as engaged in processes and practices usually reserved for higher education: devising, investigating, discarding if necessary, and defending new approaches to reading, writing, arithmetic, and other areas of activity that constituted the developing curriculum. For him, collaborative inquiry with teachers was critical to developing new principles and forms of pedagogy and curriculum, as well as providing a basis for generating his theory of inquiry and knowledge development. The work of this collaborative community, while not yet named action research, clearly foreshadowed the collaborative nature of action and practitioner research, and the co-expertise of participants central to many of the approaches currently promoted, particularly in the area of curriculum inquiry.

Dewey’s approach makes visible the dialogic potential of collaborations and how participants can learn from practices in action. It also makes visible how theorizing can be a grounded process through which principles of practice are developed, challenged, discarded, and reformulated by the collective as well as by individuals within the collective engaged in research.

Route 2 – Laboratory Schools: Blurring the Boundaries between Research, Theory, and Practice

Dewey, along with other scholars of his time, was central to the progressive education movement, which was part of a larger US sociopolitical reform effort known as the Progressive Era (*c.* from the 1840s to the 1920s). Progressives sought political, educational, and social reforms that addressed corruption in politics, expanded high schools, enacted child labor laws, supported women’s suffrage, enacted minimum wage laws for women workers, and passed four constitutional amendments: women’s right to vote; income tax; direct election of senators; and manufacture and sale of alcoholic beverages. These reforms were undertaken in the face of growing populations in cities, waves of immigrants to the US from Europe, and the growing population that resulted from the admission of ten new states to the US.

Progressive educators drew on Dewey’s ideal that schools should be about personal growth and learning through experience in a school-as-community approach. To ensure that children from all classes received a basic, common education in elementary grades, progressive reforms increased the number of children in schools by establishing compulsory attendance laws and labor laws restricting children’s work. Many also embraced the need for early education proposed in Europe by Pestalozzi and Froebel, leading to the founding of kindergartens in the US. Along with new forms of curriculum, progressive educators also promoted a scientific approach to studying education.

During the Progressive Era, the need for greater numbers of teachers was met by the opening of laboratory schools in public and private universities. In these new institutions, as in Dewey’s school, the goal was to blur perceived gaps between theory, research, and practice, making practice a basis for theory development and integration. While many laboratory schools were established for teacher training, others were founded as places for exploring research, theory, and practice relationships as a form of school experimentation. In a tradition that continues today, laboratory schools served as officially sanctioned, dynamic contexts that involved teachers, administrators, students, and, at times, community members in collaborative research for developing and testing new forms of school organization, curriculum directions, and pedagogy.

The Progressive Era was also a time when research-oriented professional organizations in education were founded (e.g., National Council of Teachers of English in 1911, American Educational Research Association in 1916, and National Association for Research in Science Teaching in 1928). These professional organizations were important in promoting educational research and curriculum inquiry. They became forums for exploring and

disseminating educational research across university and education communities that continue today. These forums led to new understandings of what counted as scientific research and what counted as theoretically grounded approaches to the study of topics such as curriculum, learning, and administrative leadership.

Analysis of the larger historical contexts that gave rise to the progressive education movement indicated that the focus on scientific methods was not solely an educational research direction, but one embedded in the zeitgeist of the historical period. Furthermore, this analysis made visible the role that university laboratory schools play as sites for generating theory as well as practitioner inquiry. It also made visible ways that research, theory, and practice can play a dynamic part in educating teachers and developing innovative curriculum, and how scientific and grounded approaches to school and pedagogical experimentation are important parts of this process.

Route 3 – Curriculum Inquiry Research and Teacher-Involved Action Research

The period from the 1940s to 1960s was another dynamic period of developing approaches to practitioner inquiry. During this period, action research, a term often attributed to Kurt Lewin (1944) and/or to John Collier (1945), developed. At the same time, curriculum inquiry approaches to research begun in the Progressive Era were continued. These routes permit a close examination of how teacher action research and curriculum inquiry contributed to, and were (re)formulated to address, the growing interest in both community-based research and research on the individual–organizational relationships. These routes are described in this section.

The first route discussed is action research. Scholars who framed the approach called “action researcher” were focused on understanding the impact of social change on individuals or groups and on promoting social change at both the societal level and the organization or individual-within-group levels. This work includes two directions, one known as action research (Collier, 1945; Lewin, 1944), and a later direction known as action science (Argyris and Schön, 1989). These three directions influenced work in educational administration as well as work on reflective teaching in the post-1960s period of the development of teacher-based action research and practitioner research in the US.

The first form of action research identified focuses on action research as an approach to community development that was developed by Collier in 1917 for his work in the New York Peoples’ Institute. Later, he used action research as the director of the Bureau of Indian Affairs to support local inquiry by members of tribal councils as a way of transforming government policy and building

self-sufficiency. Central to Collier’s approach to action research was the identification of problems perceived by a local community, which then led to a process of gathering information and making decisions about how to take social action within the community based on the data and evidence collected. This form of action research was, in turn, a form of social activism implemented to bring about political, social, and educational reforms. Wann (1953) identified Collier’s multifaceted approach to inquiry as ethical actions as a historical root of action research in education.

The second form of action research was developed by Kurt Lewin (1944) to examine how changes among members of groups had an impact on behaviors, actions, and attitudes of individuals-within-the-group members. This work focused on group dynamics and was based on an understanding of the interdependence of fate among group members as well as task interdependence. For Lewin, action research was a way of engaging actors in explorations of the impact of change (often initiated by Lewin) in order to develop ways of achieving a more democratic form of group dynamics. This research, as well as his work on the impact of leadership and social organization, was later used in teacher education. Similar to Dewey, Lewin did not view democratic actions as static; rather, he argued that democracy had to be learned anew in each generation, that experiential learning was important to personal change and development, and that a scientific approach to the study of such learning was needed.

Lewin’s work was extended by Argyris and his colleagues (e.g., Argyris and Schön, 1991) over the next two decades to create what became known as action science. Argyris, Schön, and colleagues built on Lewin’s notion that “...causal inferences about the behavior of human beings are more likely to be valid and enactable when the human beings in question participate in building and testing them” (Argyris and Schön, 1991: 86). What distinguishes action science from action research is an emphasis on the differences between espoused theories and the actual, tacit, and spontaneous theories in use that participants bring to practice and research. Action science has influenced work on organizational leadership with administrators. Schön’s work on reflective practitioners and their theories in use influenced directions in teacher and practitioner research.

As action research was developing, curriculum inquiry in laboratory and related school was continuing as described by Wann (1953) in his review of curriculum inquiry and the intellectual on this approach. One influence he identified was the theoretical work and research of both Collier and Lewin, discussed previously. Another influence was the work of Stephen Corey and colleagues in the 1950s at the Horace Mann–Lincoln Institute for School Experimentation. Corey and his colleagues (such as Alice Miel) worked collaboratively with teachers

within the laboratory school context and related school sites in New York to explore issues of curriculum, social development, and early childhood development among other areas. Based on his review of this work, Wann argued that action research work in this period was an emerging methodology, and called for experimentation to “develop a way of working which will make it practical for teachers to carry on research of a high quality” (p. 342).

In a later review, Noffke (1997) also identified Corey and his colleagues as key developers of action research in the school context. She argued that they had the dual goals of legitimizing action research by placing “greater emphasis upon evidence, upon the action hypothesis, and upon the importance of cooperative research” (Corey, 1953: 40). Through this process, they sought to blur the boundaries between research, theory, and practice, while creating a basis for theorizing practice in action. Thus, Corey, similar to Lewin, Argyris, and Schön, reiterates Dewey’s earlier arguments about the need for high-quality research and warrantable knowledge grounded in the actions of teachers and students in classrooms (e.g., Dewey, 1938).

The picture emerging through these brief explorations of the roots of action inquiry is one in which educators, social science researchers, and university-based educators were seeking warrantable ways of understanding human experience by engaging with people in local environments. The developing picture also shows that some of the roots began in education, while others were created by researchers who used education as a site for exploring their interests. Additionally, analysis of the goals of these different researchers showed that some of them engaged in action research to develop theories of action and to understand change in social worlds, while others had a more transformative agenda of promoting democratic change. Regardless of the arena of action and research, what was common was that the theories and approaches developed grew out of collaborative work with educators and others, creating an approach to theorizing that meets Elliott’s (1991) goal presented earlier of “... improving the quality of action within [the social situation]” (p. 69).

Route 4 – Continuity, New Routes, and the Action Research Turn

The period from the 1970s to the present is one of continuity with historical directions, and great variation in approaches as described in the opening of this article. Given the variety of directions developed in this period, a discussion of each is not possible. Therefore, three routes representative of groups of developing traditions are presented. Two of these routes have their roots in curriculum inquiry and action research directions presented previously. The third represents a shift from a curriculum focus to one

currently referred to as practitioner research (Zeichner and Noffke, 2001) or practitioner inquiry (Cochran-Smith and Donnell, 2006). These three directions enable us to examine how the arena of action and purpose of study led to diverse sets of approaches (Noffke, 1997). The three routes identified represent conceptually and socially different ways of engaging in action research, curriculum inquiry, and the more recent iteration – practitioner inquiry.

The first route identified has continuity with the earlier traditions of the progressive education movement and with the curriculum inquiry approaches of the period from the 1940s to 1950s. One way to view directions in curriculum inquiry from 1970 to the present is to view them on a continuum from school- and teacher-based approaches to approaches designed to create a state, district, or national curriculum in which teachers had little input, but that claimed authority from certain forms of educational research – for example, large-scale research. Approaches at one end of spectrum have roots in the earlier work on curriculum studies. The most recent form of this work is referred to as arts-based research (Barone and Eisner, 2006), an approach that continues the progressive ideals, but argues that humanities based inquiries, not just social science approaches, are the appropriate ways of examining and representing the complex work of constructing curriculum with students. This approach seeks innovative ways of engaging teachers in exploring and constructing their theories in action as a way of enhancing curriculum. It is often undertaken by collaborative communities and has been influential in creating new understandings of curriculum as a dynamic process of meaning construction.

This situated and teacher-based approach contrasts directly with the dominant, reform-driven approaches referred to as curriculum alignment studies (Porter, 2006). In these studies, the question is a policy-related one – whether or not the teacher’s work is aligned with the prescribed goals and practices of the planned curriculum. The object of study is teaching behaviors, not teacher actions to improve the quality of curriculum. This movement is also visible in cross-national assessments and is grounded in twentieth-century views of scientific inquiry as searching for generalizable practices and knowledge. The alignment approach contrasts with the work of Dewey, Lewin, Argyris, and others. Although these scholars sought rigorous high-quality research, they also viewed teacher questions and knowledge as central to the research, and curriculum inquiry as a basis for theorizing practice in local contexts. Thus, the current focus on imposed curriculum ignores practice as a basis for new knowledge generation and for improving the quality of local actions.

The second route, one developed in Australia, has influenced critical, participatory, and collaborative action research directions in the US and elsewhere. Grounded in work by Stephen Kemmis, Robin McTaggart, and Shirley

Grundy, among others, this approach, similar to Collier's work, seeks to support local groups in constructing a more fair and equitable democratic society. Researchers within this tradition seek ways of becoming critical (Carr and Kemmis, 1986) as an ethical stance to engaging with communities in order to support members in researching locally identified problems. The research process is collaborative, locally situated, and recursive. Actions with community members include: identifying a problem, collecting information in local contexts, engaging in the analysis of information (data), interpreting the data, supporting the planning of actions based on the local study, and, at times, repeating the action research cycle until the community has achieved their desired goals. In education, this means that teachers conducting action research could be of value in affecting change of educational theory, policy, and practice when participants become agents of change in their local communities.

The third route focuses on practitioner research in a professional context and contrasts with the previous routes in particular ways. There are two interconnected directions for practitioner research in this tradition, one grounded in theories and research in the UK that have influenced US directions, and one in the US represented in recent reviews (e.g., Cochran-Smith and Donnell, 2006; Zeichner and Noffke, 2001). The first direction builds on theoretical traditions and perspectives from the UK (e.g., Day *et al.*, 2002), focusing on theory and practice of action research and situated views of learning. In this tradition, practitioners are viewed as cultural beings and as members of a professional community that shapes, and is shaped by, professional cultures in which they live and work. The boundaries of theory and practice are blurred and action research becomes a professional ideal realized not as a formal study, but as everyday work to enhance the quality of teaching, service, or care. The practitioners' questions and the challenges they face become the basis for action, and inquiry is ongoing, reflexive, and responsive to the needs in the local social situation. Practitioner inquiry becomes a means through which professionals are able to (re)discover, articulate, and communicate core values and take actions that promote the common good.

In the US, practitioner inquiry has become an umbrella term for a variety of approaches, many with different epistemological and theoretical frameworks (Cochran-Smith and Donnell, 2006). In contrast to earlier routes, the named approaches come from different theoretical orientations, purposes, methods, and sites or arenas of inquiry (Noffke, 1997; Zeichner and Noffke, 2001). Regardless of theoretical grounding or the method used, practitioner inquiry involves one or more practitioners in research about a problem or issue in their local context with or without external support or intervention. Such inquiry may take many directions. One direction involves teachers in research on a new curriculum or on a new area

of study initiated through participation in a university course. Another direction represents collaborative work in a professional development community. The third direction is one in which the individual teacher engages in reflective practice in his/her own classroom. These and other directions of practitioner inquiry, such as the practitioner action research of the UK, seek to develop reflexive action and inquiry as professional ideals, where inquiry is part of the professionals' repertoires for action, for (re)formulating practice, and for enhancing the quality of actions, service, or care within professional work settings.

Across all of the action research and practitioner inquiry traditions of this route is the understanding that practitioner-based action research and curriculum inquiry involves research designed to address questions of importance to practitioners and community members within a local context. This research often leads practitioners (or community members) to (re)theorize their work, (re)consider their assumptions about learning (or other related outcomes of practice), explore how they make decisions about their practice, develop plans for actions to resolve the problem identified, and monitor the impact of the planned actions on students and others in their classrooms (or other professional settings). This reflexive and recursive process is central to the different approaches within this route.

Possibilities and Challenges for the Twenty-First Century

Although many additional outcomes and goals of teacher-based action research, curriculum inquiry, and practitioner research exist, the ones highlighted in this article show a continuity of goals across the past 100-plus years – to frame sound educational (or locally responsive) processes and practices that represent the actions, work, knowledge, and understandings of teachers and community members in the contexts of professional practice or community life. In a recent review of different roots and routes, published between 2004 and 2006 in *Action Research*, Dick (2006) argues that the scope and depth of work indicates an action research turn, not just in education, but also across disciplines and fields of professional work. This turn, he argues, does not mean that all of the work of framing action research is complete, but rather that interest has been growing in this approach, one in which further theoretical and methodological work needs to be done, a point supported by Cochran-Smith and Donnell (2006) and Zeichner and Noffke (2001).

One further observation and a set of challenges to the continuing development of action research and practitioner inquiry provide a fitting close for this analysis of teacher action research. Bauman (2004) argues that the twenty-first century is a period of liquid modernity in which the theories of the nineteenth and twentieth

centuries will not be adequate to understand the fluid world of the digital age, with its new, often interdisciplinary, areas of study and its knowledge base that is changing with greater speed than in the twentieth century (i.e., every 2–4 years in some disciplines, in contrast to every 10 years). He also argues that educational systems and societies are now educating for jobs that do not exist yet. In a fluid world, action research holds a key to insuring equity of access to education, improving curriculum in changing fields, and enhancing community and student learning about the new challenges and directions. This value can be seen in the recent takeup of action research in fields such as informational technology, allied medical fields, and architecture.

Two unresolved challenges face those involved in developing this approach further. The first was articulated by Dewey more than 100 years ago, Wann almost six decades ago, Winter and Elliott in the UK, and Cochran-Smith, Zeichner, Noffke, and others in recent reviews – the need to build support, time, and policies for professionals to engage in action or practitioner research in their contexts of practice. The second is the challenge of building theories of action research and professional action and how this form of research can inform policymaking in new ways. In the twenty-first century, teacher-based action research and curriculum inquiry need to become professional ideals that support professional action in the fluid and complex world of work and professional activity.

See also: Early Childhood Curriculum and Developmental Theory.

Bibliography

- Argyris, C. and Schön, D. A. (1989). Participatory action research and action science compared: A commentary. *American Behavioral Scientist* **32**, 512–623.
- Argyris, C. and Schön, D. A. (1991). *Organizational Learning*. Reading, MA: Addison-Wesley.
- Barone, T. and Eisner, E. (2006). Arts-based educational research. In Green, J., Camilli, G., and Elmore, P. B. (eds.) *Handbook of Complementary Methods in Education Research*, pp 93–107. Mahwah, NJ: Erlbaum.
- Bauman, Z. (2004). *Identity: Conversations with Benedetto Vecchi*. Cambridge, MA: Polity.
- Carr, W. and Kemmis, S. (1986). *Becoming Critical: Education, Knowledge and Action Research*. London: Falmer.
- Cochran-Smith, M. and Donnell, K. (2006). Practitioner inquiry: Blurring the boundaries of research and practice. In Green, J., Camilli, G., and Elmore, P. B. (eds.) *Handbook of Complementary Methods in Education Research*, pp 503–518. Mahwah, NJ: Erlbaum.
- Collier, J. (1945). United States Indian administration as a laboratory of ethnic relations. *Social Research* **12**, 265–303.
- Corey, S. M. (1953). *Action Research to Improve School Practices*. New York: Teacher's College Press.
- Day, C., Elliott, J., Somekh, B., and Winter, R. (eds.) (2002). *Theory and Practice in Action Research: Some International Perspectives*. London: Symposium Books.
- Dewey, J. (1899). *The School and Society: Being Three Lectures*. Chicago, IL: University of Chicago Press.
- Dewey, J. (1938). *Logic: The Theory of Inquiry*. New York: Henry Holt.
- Dick, B. (2006). Action research literature 2004–2006: Themes and trends. *Action Research* **4**, 439–458.
- Elliott, J. (1991). *Action Research for Educational Change*. Buckingham, PA: Open University Press.
- Green, J., Dixon, C., and Zaharlick, A. (2003). Ethnography as a logic of inquiry. In Flood, D. L. J., Squire, J. R., and Jensen, J. M. (eds.) *Handbook of Research on Teaching the English Language Arts*, 2nd edn., pp 201–221. Mahwah, NJ: Erlbaum.
- Lewin, K. (1944). The dynamics of group action. *Educational Leadership* **1**, 195–200.
- McKernan, J. (1987). The potential and practice of action research. *Peabody Journal of Education* **64**, 6–19.
- Noffke, S. E. (1997). Professional, personal, and political dimensions of action research. *Review of Research in Education* **22**, 305–343.
- Porter, A. C. (2006). Curriculum assessment. In Green, J., Camilli, G., and Elmore, P. B. (eds.) *Handbook of Complementary Methods in Education Research*, pp 141–159. Mahwah, NJ: Erlbaum.
- Wann, K. D. (1953). Research on human relations. *Review of Educational Research* **23**, 337–345.
- Winter, R. (1987). *Action-Research and the Nature of Social Inquiry: Professional Innovation and Educational Work*. Aldershot: Gower.
- Zeichner, K. M. and Noffke, S. E. (2001). Practitioner research. In Richardson, V. (ed.) *Handbook of Research on Teaching*, 4th edn., pp 298–330. Washington, DC: American Educational Research Association.
- Bredo, E. (2006). Philosophies of educational research. Handbook of complementary methods in educational research. In Green, J., Camilli, G., and Elmore, P. B. (eds.) *Handbook of Complementary Methods in Education Research*, pp 3–31. Mahwah, NJ: Erlbaum.
- CARN (quarterly). *Educational Action Research*. <http://www.tandf.co.uk/journals/titles/09650792.asp> (accessed May 2009).
- Cochran-Smith, M. and Lytle, S. L. (1999). The teacher research movement: A decade later. *Educational Researcher* **28**, 15–25.
- Fullan, M. (1993). *Change Forces: Probing the Depths of Educational Reform*. Bristol, PA: Falmer.
- Gonzalez, N., Moll, L., Floyd-Tenery, M., et al. Teacher Research as Funds of Knowledge. <http://repositories.cdlib.org/crede/nrcds/leeducational/EPR06/>.
- Gore, J. and Zeichner, K. (1995). Connecting action research to genuine teacher development. In Smythe, J. (ed.) *Critical Discourses on Teacher Development*, pp 203–214. London: Cassell.
- Hollingsworth, S. (ed.) (1997). *International Action Research: A Casebook for Educational Reform*. New York: Falmer.
- Kemmis, S. and McTaggart, R. (eds.) (1992). *The Action Research Planner*. Melbourne, VIC: Deakin University Press.
- Reason, P. and Bradbury, H. (eds.) (2001). *Handbook of Action Research: Participative Inquiry and Practice*. Thousand Oaks, CA: Sage.
- Schön, D. A. (1992). The theory of inquiry: Dewey's legacy to education. *Curriculum Inquiry* **22**, 119–139.
- Winter, R., Burroughs, S., Crosson, D., Leontovitch, A., and Thorne, H. (1989). *Learning from Experience: Principles and Practice in Action-Research*. Philadelphia, PA: Falmer.

Further Reading

- <http://coe.westga.edu> – Action Research Special Interest Group, American Educational Research Association.
- <http://www.did.stu.mmu.ac.uk> – Manchester Metropolitan University, Collaborative Action Research Network (CARN).
- <http://www.nipissingu.ca> – Nipissing University, The Ontario Action Researcher.
- <http://www.scu.edu.au> – South Cross University, Action Research International.

<http://www.teacherasresearcher.org> – Teacher as Research Special Interest Group, American Educational Research Association (TAR).
<http://www.nwp.org> – Teacher Inquiry Networks, National Writing Project, Teacher Research.
<http://www.teacherresearch.net> – Teachers' Research: Local, National and International Work.

<http://people.bath.ac.uk> – University of Bath.
<http://www.uea.ac.uk> – University of East Anglia, Center for Applied Research in Education.
<http://www.scre.ac.uk> – University of Glasgow: Faculty of Education – The SCORE Centre.
<http://epaa.asu.edu/epaa/v1n1.html>.

CURRICULUM DEVELOPMENT – PLANNING AND DEVELOPMENT

Contents

Curriculum and Syllabus Design
Curriculum and the Publishing Industry
Curriculum and Teacher Change
Curriculum Governance and Planning
Curriculum Planning and Systems Change
Curriculum, Digital Resources and Delivery
Textbook Development and Selection

Curriculum and Syllabus Design

A Woods, A Luke and K Weir, Queensland University of Technology, Kelvin Grove, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

This article is an overview of syllabus design as part of the curriculum process, with a focus on design and the technical form of syllabus documents. The distinction between curriculum and syllabus has not been clearly delineated within either practice or policy in this field. Yet, the distinction is not a trivial one. The conflation of the design, structures, and functions of the syllabus, on the one hand, and the enacted curriculum, on the other, remains a practical problem for policymakers and practitioners. We review work in two fields of education that have a tradition of detailing issues related to syllabus design: English-language teaching and higher education, then turn to issues of syllabus design in the compulsory years of schooling. Models of curriculum have informed and constrained the content and technical form of school syllabus documents. However, current curriculum debates have tended to focus on ideological, cultural, and scientific curriculum content, with little or no discussion on or about the technical form of the curriculum or syllabus documents.

Defining Curriculum and Syllabus

Curriculum is the sum total of resources – intellectual, scientific, cognitive, and linguistic – that is brought to bear on the dialog and exchange of teaching and learning. It includes documents, textbook and adjunct resources and materials, both official and unofficial – that are brought together by teachers and students to structure teaching and learning in classrooms and other learning environments. Taken in its simplest terms, curriculum is simply what is taught and learned in schools. It is the constitutive cultural

and scientific content of education that is transmitted by the message systems of pedagogy and assessment.

In contrast, the syllabus is a bid to shape and set the parameters of the curriculum, usually prepared at a school-system level. It is a defensible map of what is valued as core skills, knowledge, competences, capacities, and strategies to be covered within a particular context at a particular time, usually with affiliated statements of standards, which are used for accountability. The *Oxford English Dictionary* tells us that the term syllabus has evolved to refer to a summary of what is to be taught and learned. The Greek etymology of the word refers to the verb ‘to collect’. Cicero used the term to refer to correct leader, suggesting an authoritative status. In the seventeenth century, the term referred to the finished curriculum or a course of study completed. By the nineteenth century, the term was used to refer to a statement, a table of contents, or heads of study. The syllabus was used in fields such as literature and law to refer to a map or outline of curriculum.

Current work on syllabus design in the field of higher education tends to represent and explain the syllabus and its functions through a series of metaphors. According to Parkes and Harris (2002), the three most common metaphors employed in discussions around the syllabus are to discuss the term as a contract, a permanent record, or as a resource for student learning. Many define the syllabus as a communication tool or device (e.g., Thompson, 2007; Slattery and Carlson, 2005). Another common metaphor utilized in the literature is the syllabus as a map. Green and Stortz (2006) take the idea of the syllabus as a journey to another level when they discuss the syllabus as a passport that enables a common culture of teaching and learning.

There is a paucity of discussion around the alignment of the curriculum and the syllabus. The default approach for those developing official syllabus documents has been the Tyler rationale for curriculum development. The end result is that the relationship between a syllabus and its curriculum is neither well understood nor discussed to any depth in contemporary literature and research except where its significance is integral to syllabus design. For example, Habanek (2005) views the syllabus as a permanent record of curriculum and “an agreement that specifies the accountability of the major players in the learning process” (p. 63). In his seminal work on curriculum development and design, Print (1993) presents a much broader description of the link between syllabus and curriculum, while acknowledging that the two entities are often confused. His idea is that the syllabus forms part of the overall curriculum and tends to be a list of content areas which will be assessed, whereas the curriculum is “all the planned learning opportunities offered by the organisation to learners and the experiences learners encounter when the curriculum is implemented” (p. 9). The syllabus has traditionally been defined at one level of technical generality from the actual content of the curriculum – forming an authoritative outline, schema, or structure for courses of study. However, because of the lack of conceptual clarity in defining the syllabus, these historical definitions are merged, often without clarity, in teachers’ and curriculum developers’ work.

Westbury (2008) defines the syllabus as a guide to the curriculum, while Schwartz (2006) describes the syllabus as a written curriculum that acts as an action-oriented guide or tool for teachers. The syllabus is an official map of a school subject – of the terrain to be covered (cf. Dewey, 1902). The syllabus is not an exhaustive view of the territory, but it sets the grounds for teachers’ and students’ actual educational journey through the terrain. Following Dewey, the syllabus is not, and cannot be, comprehensive or exhaustive. It prescribes pedagogic method, approach, style, and instructional interaction at the risk of constraining teacher professionalism in responding to student diversity. Therefore, the school syllabus can provide teachers with a rationale and outline of the school subject in question, an overview and specification of preferred expected content to be taught and learned, and a description of operational ways of appraising standards for gauging student performance.

Available Knowledge About Syllabus Design

There is little technical literature on syllabus design. The academic literature around syllabus design and implementation principally is found in two fields of study: English-language teaching and higher education.

An additional small body of work related to vocational and adult education focuses on module-based syllabus planning. Each of the two fields reviewed here views the syllabus from quite different perspectives and each differs in its depth of inquiry. However, the implied goal of the majority of research is methodological, with a focus on the ‘how to’ of planning and teaching a syllabus at a course or unit level in these areas.

Syllabus Design and Teaching English as a Second or Foreign Language

The syllabus literature that relates to research and scholarship in the field of English-language teaching originates from a corpus of general textbooks on the topic published during the 1980s. This reflects an era of renewed interest in teaching English as it became an essential global language, and coincides with a paradigm shift in the teaching of English toward a sociolinguistic model for understanding language acquisition that emphasizes communicative language teaching.

This work focuses on methodologies for teaching English in specified contexts. According to Nunan (1991), this reflects significant influence language teachers themselves have had on syllabus design and learning resources. The literature is therefore typified by the language-learning context in which the syllabus is to be implemented, and the syllabus appears to be represented in terms of a structured, linear pedagogical approach to teaching English. The approaches include English as a second or foreign language (ESL or EFL respectively); teaching English to speakers of other languages (TESOL); English-language teaching in its social context; and English for specific purposes (ESP), which incorporates English for academic purposes and occupational purposes.

In this context, the syllabus is mainly represented as a pedagogical tool that prescribes a particular method’ for teaching English and a specific sequence of incremental acquisition of language competence. The syllabus is described using a range of lexical descriptors that generally reflect the pedagogical approaches they represent. This representation of the syllabus appears to have been initiated by Wilkins (1976), who created a superordinate classification of notional syllabus types, labeling them analytic and synthetic: terms that relate to the role of the learner in acquiring a language. Other syllabus types described in the literature reviewed include the text-based syllabus; the task-based syllabus; the communicative syllabus; the topic-based syllabus; the negotiated syllabus; and the negotiation and process syllabus.

In sum, the English-language teaching field tends toward simplifying the syllabus as a guide for teachers wanting to implement a particular sequenced approach to teaching English (Breen, 1987).

Syllabus Design for Teaching in Higher Education: The USA Context

The other field of study that dominates the literature in syllabus design is the scholarship and practice of teaching in American colleges and universities, based in departments of instruction within faculties. The work tends to be presented as practical manuals or guides for writing syllabi to reflect institutional teaching and learning policies. However, some of the work at least identifies the underlying purposes and instructional principles of syllabus design, and outlines principal components of a syllabus.

Doolittle and Lusk (2007) indicate that syllabus components, functionality, and perceptions by faculty and students are three foci of the literature. Their work details not only a focus on the considerable variability in the construction and uses of syllabi (p. 65), but also that there is little in the higher education literature that provides ongoing dialog beyond the basic elements and utility of the syllabus as a material artifact of the curriculum. Cardozo's (2006) work further supports this as he claims that "syllabus construction itself remains a significantly under theorized professional activity" (p. 412). The lack of sustained inquiry around the syllabus seems incongruous considering its almost universal status as an educational tool. There remains very little research into syllabus design and implementation in relation to the compulsory years of schooling.

Curriculum Models and Syllabus Design for Schooling

A range of curriculum models have impacted upon the content included in syllabus documents and as such these models warrant discussion. Current curriculum models begin from stances on content. However, in doing so, these major curriculum models explicitly and implicitly define the technical form of the syllabus. The dominant models of curriculum underlying current practice and policy debates can be categorized into several models. It is important to note that we do not include commentary on the phenomenological model or other models of curriculum (e.g., Green, 2005) that have had less direct influence on actual school syllabus definition.

The Outcomes-Based Model

This model was developed in the United States by Spady (1982) and colleagues in the 1980s and 1990s. It argued that curriculum should focus less on the traditional selection and specification of knowledge content, and less upon pedagogic experience and process, and more on the specific visible and demonstrable outcomes in student performance. It marks out a technocratic model of education:

an industrial-era model of curriculum derived from Tyler (1949) that breaks subject areas into smaller constituent parts on the basis of claims about student needs. Its technical form is enumerated, categorical lists of outcomes for specific subjects and age/grades. These have tended, in the outcomes models, to be couched in behavioral terms, that is, as skills and behaviors that can be demonstrated and observed. This model is amenable to consensus-building, expert-committee, and professional consultation processes. Its focus on lists of outcomes also enables a dovetailing with systemic testing for accountability purposes. The model has been strongly criticized by critical theorists, traditionalists, and progressives. From radically different perspectives, they argue that it tends to atomize and narrow down the curriculum, reducing its richness, and that it excludes a range of valued contents, experiences, learner backgrounds, and pedagogical processes.

The Process-Based Model

The process-based model, affiliated with the cognitive developmental work of Bruner (1960) in the post-Sputnik era, tended to treat curriculum in terms of a developmental continuum of educational experiences and processes. The technical form of the curriculum when founded within these perspectives tends to be more strongly developmental, stressing students' engagement with and experience of particular repertoires affiliated with subject areas and content. However, skill and content narrowly defined are not taken as the central purpose of curriculum and instruction. This model is generally affiliated with traditions of progressivism and constructivism, with a strong focus on developmental experience and the active construction of new knowledge. The model has been strongly criticized by advocates of the technocratic/accountability model as soft on the specification and assessment of outcomes, and by classicists as failing to engage with cultural traditions, contents, and canonical texts and knowledge. It has been critiqued by critical theorists for overemphasizing individual development over social and cultural development.

The Critical Model

This model, affiliated with critical theory and cultural studies in the humanities and social sciences, strongly emphasizes the need for competing, revisionist descriptions and models of the world, and for critical, active, and agentive student engagement with knowledge. While it has influenced education and curriculum content in at least some systems, such as Australia, it has had little impact on the technical form of the curriculum. It directly addresses content issues and tends to stress higher-order or critical skills. The model has been strongly criticized by

technocratic advocates as failing to deliver demonstrable outcomes. It is critiqued by the classicists for being politically and ideologically biased, and leading to a diluted or skeptical engagement with key traditions and works. It is critiqued by developmentalists for overemphasizing social and cultural development at the expense of individual cognitive and psychological growth.

The Traditional Content Model

This model was based on a neoclassical model of curriculum (e.g., Bloom, 1956) – that neoclassical model was based on the identification of canonical knowledge and texts in fields. Its technical form entails the enumeration and prescription of content knowledge, prescribed reading and topics. Until recently, it has been the dominant model in university teaching, and is strongly affiliated with traditional curriculum examination systems. The model is criticized by technocratic/accountability advocates as failing to provide demonstrable outcomes other than content reproduction. It has been critiqued by developmentalists for being insensitive to individual and developmental diversity in background knowledge and approach to learning. It has been critiqued by critical theorists on the basis of the assumption that this model hides its biases and ignores the dynamics of economic, cultural, and social change.

These dominant curriculum models have influences on different school subjects and school phases across systems, mostly in blended forms. Different content claims about what should be taught and learned reflect particular paradigmatic positions on the curriculum. But each of these models offers different approaches to the taxonomic grids and descriptive categories for describing what should be taught and learned. Therefore, elements of major curriculum models are evident in the curriculum development and syllabus design of most school systems. However, the technical form of the syllabus has moved toward a technocratic/accountability model through the focus on measurable and observable outcomes within many Western systems of schooling. Within this technical form, different subject areas and phases have been shaped according to different curriculum models, contents, and theoretical assumptions – for instance, preschool and early-childhood curricula in many systems have taken a developmentalist philosophy – but these approaches currently often sit within the technical form of the outcomes-based structure, and in the context of system consensus around the technocratic/accountability model.

These approaches are responses to particular policy contexts, particular national histories and system contexts, and particular economic and social conditions. Further, they are adopted in relation to particular pedagogic patterns, school and classroom structures, and intellectual and cultural traditions. The curriculum settlements of the last

decade are in transition. The demands of the new economy, the multicultural, multilingual, and multimodal nature of new student populations, and changes in educational philosophy, funding, and governance have led to this transition period. There is an extensive published literature on curriculum theory, curriculum contents, curriculum development, curriculum management and implementation, systemic accountability and reporting, and teachers' and students' uses of the enacted curriculum in schools. However, there is very little written about the shape and structure – the technical form – of syllabus documents used to document or map the compulsory school syllabus.

The Technical Form of the Syllabus in High-Equity/High-Quality Systems

The technical form of the syllabus is a neglected area of current curriculum debates and of the field more generally. These debates have largely been preoccupied with questions of curriculum content – variously construed as cultural values, ideologies, specific skills sets, competences, and disciplinary knowledge. We have defined the syllabus as a map and descriptive overview of the curriculum, as a structured summary and outline of what should be taught and learned. The syllabus is not the curriculum – but what occurs in teaching and learning is shaped by a range of factors, with the official syllabus being one key factor. Other factors include, but are not limited to the background knowledge, cognitive and cultural resources that students bring to classrooms; teacher expertise gained through pre- and in-service teacher education and practical experience; textbook selection and content; availability of further training and professional resources; school leadership; system governance and accountability structures; high-stakes testing and examination; classroom assessment; available financial resources; and the physical site of the classroom. The syllabus may enable and constrain, but does not necessarily reflect or index what is taught and learned in classrooms. The principal way that national debates in many Western nations have dealt with this problem of control is to debate the political, cultural, and scientific values and truth claims of different stances on content in syllabus – and to augment this with criticism of teacher workforce capability and professionalism. This approach leads to a dual-policy approach: fix and mandate new (or old) content (change the prescription); and enforce this through increased accountability pressure, incentives, and disincentives for teachers (change the professionalism).

Schleicher (2007) refers to this as uninformed prescription that is linked to uninformed professionalism. Uninformed prescription, he argues, may entail strong centralized accountability without the resources or the opportunities for building strong knowledge-based and evidence-based teacher professionalism. He stresses the

need for an approach to documenting curriculum in syllabus documents that lays out informed prescription centrally (through the syllabus setting core learnings and specification of standards) but that also sets the conditions for local teacher professionalism, school and classroom-based developmental diagnostic use of evidence, and the exercise of local curriculum interpretation and translation, development, and implementation. This is part of a process of lifting expectations that all students will achieve, while encouraging a range of relevant pedagogical approaches. He refers to this as informed prescription.

In high-quality/high-equity systems, teachers use professional knowledge and evidence to make informed and relevant decisions about teaching and learning. In other words, informed prescription depends on teachers' professional capacity to locally interpret, adapt, and adjust curriculum content, pacing, presentation, interaction, and structure to particular institutional, community settings and student cohort characteristics. It includes a capacity to use evidence on student background, prior achievement, developmental and diagnostic progress, and school and classroom-based assessment to make curricular and instructional decisions.

An emphasis on centralized standards and curriculum mandates must be balanced against high levels of workforce curriculum professional decision making. Informed prescription requires well-resourced teacher professional capacity. The approach of high-quality/high-equity systems entails a balance of systemic standard setting and accountability with well-resourced, local school leadership with a strong focus on building teacher capacity at curriculum, pedagogy, and assessment (Schleicher, 2007).

Once the educational goals of a system are established, the syllabi – its contents and technical form – can be part of the solution toward realizing a high-quality and high-equity system, but syllabus documents cannot, in and of themselves, generate change. The syllabus has the important function of setting conditions for enhancing a knowledge-rich professionalism – but other policy settings also need to be in place. These include a clear and simplified message system about aims and priorities regarding quality and equity, professional infrastructure, workforce capacity, school governance, and management structures that likewise are geared to enable instructional quality.

Conclusion

We began this article with the claim that the syllabus is not the curriculum. The confusion of syllabus and curriculum as terms has created a number of issues for school systems. However, in relation to the field of curriculum and syllabus design generally it has resulted in the technical form of the syllabus being neglected in current curriculum debates and

literature of the field. For the past five decades, Western democratic education systems have attempted to strike a balance between the goals of economic development and competitiveness, on the one hand, and social and cultural development and cohesion, on the other. The purpose of democratic education is for all students to reach a democratic threshold (Guttman, 1987) of knowledge, skills, and dispositions for gainful and productive participation in democratic social institutions and economies. The syllabus cannot control the curriculum, but it can set enabling conditions for high-quality/high-equity outcomes. It cannot cause change and progress in any direct or simple way, but can be one of the key elements of an overall system strategy for enhancing teaching and learning.

How the syllabus is shaped, how it is used in the context of system accountability around standards, and how teacher's use of the syllabus is resourced and supported set the conditions for a balance of prescription and professionalism. Establishing that balance in ways that are conducive to high-quality with high-equity teaching and learning is the task facing policymakers and teachers currently. The syllabus is a map, and as such it should aim at low-definition, parsimonious, and economical statements. The syllabus can guide and enhance professional expertise – but it cannot act as a substitute for well-resourced and informed teacher pre- and in-service development and practice.

Bibliography

- Bloom, B. S. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals: Handbook I, Cognitive Domain*. New York: Longmans, Green.
- Breen, M. P. (1987). Contemporary paradigms in syllabus design (Parts I and II). *Language Teaching* 20(2 and 3), 81–91; 157–174.
- Bruner, J. (1960). *The Process of Education*. Cambridge, MA: Harvard University Press.
- Cardozo, K. (2006). At the museum of natural history: The experiential syllabus (or what happens when students act like professors). *Pedagogy* 6(3), 405–433.
- Dewey, J. (1902). The School as Social Center. *The Elementary School Teacher* 3(2), 73–86.
- Doolittle, P. E. and Lusk, D. L. (2007). The effects of institutional classification and gender on faculty inclusion of syllabus components. *Journal of the Scholarship of Teaching and Learning* 7(2), 62–78.
- Green, B. (2005). Curriculum inquiry in Australia. In Pinar, W. (ed.). *International Handbook of Curriculum Research*, pp 123–140. Mahwah, NJ: Lawrence Erlbaum.
- Green, B. and Stortz, M. (2006). The syllabus as a passport into a common culture of teaching and learning: Developing and assessing strategies for dealing with diversity. *Teaching Theology and Religion* 9(4), 221–228.
- Gutman, A. (1987). *Liberal Education*. Princeton, NJ: Princeton University Press.
- Habanek, D. V. (2005). An examination of the integrity of the syllabus. *College Teaching* 53(2), 62–64.
- Nunan, D. (1991). Communicative tasks and the language curriculum. *TESOL Quarterly* 25(2), 279–295.
- Parkes, J. and Harris, M. B. (2002). The purposes of a syllabus. *College Teaching* 50(2), 55–61.

- Print, M. (1993). *Curriculum Development and Design*. St Leonards, NSW: Allen and Unwin.
- Schleicher, A. (2007). Seeing school systems through the prism of PISA. In Luke, A., Weir, K., and Woods, A. (eds.) *Development of a Set of Principles to Guide a P-12 Syllabus Framework: A Report to the Queensland Studies Authority*. Brisbane, QLD: Queensland Studies Authority. http://www.qsa.qld.edu.au/downloads/publications/qa_p-12_principles_dev_ppr.pdf (accessed September 2009).
- Schwartz, M. (2006). For whom do we write curriculum? *Journal of Curriculum Studies* 38(4), 449–457.
- Slattery, J. M. and Carlson, J. F. (2005). Preparing an effective syllabus: Current best practices. *College Teaching* 53(4), 159–164.
- Spady, W. (1982). Outcome-based instructional management: A sociological perspective. *The Australian Journal of Education* 26(2), 123–143.
- Thompson, B. (2007). The syllabus as a communication document: Constructing and presenting the syllabus. *Communication Education* 56(1), 54–71.
- Tyler, R. W. (1949). *Basic Principles of Curriculum and Instruction*. Chicago: University of Chicago Press.
- Westbury, I. (2008). Making curricula: Why do states make curricula and how? In Connelly, F. M., He, M. F., and Phillion, J. (eds.) *The Sage Handbook of Curriculum and Instruction*, pp 45–65. Thousand Oaks, CA: Sage.
- Wilkins, D. A. (1976). *Notional Syllabuses: A Taxonomy and Its Relevance to Foreign Language Curriculum Development (ELT)*. Oxford: Oxford University Press.

Further Reading

- Apple, M. (1990). Is there a curriculum voice to reclaim? *Phi Delta Kappa* 71(7), 526–530.
- Bernstein, B. (1990). *The Structure of Pedagogic Discourse*. London: Routledge and Kegan Park.
- Connelly, F. M., He, M. F., and Phillion, J. (2008). (eds.) *The Sage Handbook of Curriculum and Instruction*. Thousand Oaks, CA: Sage.
- Gale, T. and Densmore, K. (2000). *Just Schooling*. Philadelphia, PA: Open University Press.
- Luke, A., Weir, K., and Woods, A. (2009). *Curriculum Syllabus Design and Equity: A Primer and a Model*. New York: Routledge.
- Nichols, S. and Berliner, D. (2007). *Collateral Damage*. Cambridge, MA: Harvard Educational Press.
- Shulman, L. S. (1987). Knowledge and teaching: Foundation in new reform. *Harvard Educational Review* 57(1), 1–22.
- Simola, H. (2005). The Finnish miracle of PISA: Historical and sociological remarks on teaching and teacher education. *Comparative Education* 41(4), 455–470.
- Snyder, I. (2008). *Literacy Wars*. Sydney: Allen and Unwin.
- Spady, W. (1994). *Outcomes-Based Education: Critical Issues and Answers*. Arlington, VA: American Association of School Administrators.
- Welner, K. G. and Oakes, J. (2008). Structuring curriculum: Technical, normative and political consideration. In Connelly, F. M., He, M. F., and Phillion, J. (eds.) *The Sage Handbook of Curriculum and Instruction*, pp 91–111. Thousand Oaks, CA: Sage.

Curriculum and the Publishing Industry

J Larson, A-R Allen and D Osborn, University of Rochester, Rochester, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Disciplinary technology – For Foucault (1979), the rise of capitalism involved a new form of “discipline” in which internalized surveillance rather than overt force works to create “docile bodies” necessary for institutions such as schools and the military. Discipline is a type of power and the modality, instruments, and techniques (technologies) through which power is exercised.

Hegemony – In educational studies, most associated with Antonio Gramsci (1971), hegemony is the domination of subordinated groups through ideological and cultural coercion and consent, such that the values and beliefs of the dominant group come to be taken as common sense, thus justifying status quo relations of power.

Literacy – A set of multi-modal, and rapidly changing social practices in which people use talk, texts, images, and other signs and symbols to accomplish specific purposes.

No Child Left Behind (NCLB) – The No Child Left Behind Act of 2001 is the reauthorization of the Elementary and Secondary Education act of 1965 by the United States Congress signed into law by George W. Bush in 2002. NCLB emphasizes accountability and scientifically based instruction, and requires public schools to administer state-wide standardized tests annually to all students in grades 3–8. NCLB mandates that 100% of students are proficient in grade level standards by 2014. Under NCLB, test scores for schools that receive Title I funding must demonstrate Adequate Yearly Progress toward proficiency.

Official knowledge – The title phrase of Apple’s (1993) text, official knowledge captures the inherently political struggle over what counts as the knowledge, history, and truth worthy of teaching and learning in schools.

Packaged Literacy Programs – Books, workbooks, teacher manuals, and related materials produced by a commercial publisher intended as a principal source of literacy instruction for a group of students. Rather than viewing literacy as a set of social practices, packaged literacy programs conceptualize literacy as a set of sequential skills to be explicitly taught and learned.

Textbooks – Books, workbooks, teacher manuals, and related materials produced by a commercial publisher, often intended as a principal source of study material for a given class or group of students. For parents and politicians, textbooks often represent the totality of school curriculum, and therefore often come to serve as de facto or surrogate curriculum. Given their synoptic nature, textbooks are inherently political materials that represent specific ways of thinking, knowing and viewing reality.

Educational textbooks are one of the most important components shaping educational practice in schools. As the primary tool of instruction internationally, textbooks represent the official knowledge considered worth teaching and learning, and often serve as the most significant resource – and the most significant constraint – for teachers. Because of the power of textbooks, publishers of educational materials have a tremendous influence on what counts as knowledge, and how particular constructions of knowledge and pedagogy dominate in schools. The question of who decides what counts as knowledge, and what knowledge students should learn in schools, has been contested since the beginning of public education. The production of textbooks for use in state-sponsored schools has historically been, and remains, a key site of struggle over whose knowledge, ideologies, histories, and values should be represented in curricula, and what kind of national identities should be promoted through education (Nozaki *et al.*, 2005). In different national contexts, textbooks are produced by different entities, including government agencies, multinational firms, and other private industry. While processes vary by national context, the production of school textbooks is an inherently cultural, economic, and political enterprise, with consequences for what and how teachers teach, and students learn. Textbooks are often deemed successful when guarantees are asserted that children will learn, no matter the quality of the teacher (e.g., teacher proofing). What students learn and the meaning and purpose of that content tends not to be examined.

While school textbooks are commonly identified with Western postwar educational theory and practice, the use of a common core text has a long history in many religious and educational traditions, including Confucian educational practices, and hermeneutic training for youth in

Muslim and Hebraic societies (Luke *et al.*, 2003). The contemporary context for textbook production, distribution, and use is complicated by tensions between international publishing concerns and national efforts to develop and use indigenous knowledge and publishing processes (Limage, 2005). This article examines the role of textbook publishers in the construction of curriculum, knowledge, and pedagogy. Then the article discusses the consequences of publisher-controlled knowledge production on teaching and learning using literacy curriculum in the US as a case example. The article concludes with the discussion of the problems associated with static definitions of knowledge and pedagogy that are assumed in textbooks and points to some possible rethinking of knowledge as commodity.

Deskilling, Control, and Discipline

Any discussion of the role of publishing industries in education has to take account of the now ubiquitous collaborations between corporations and governments in the production of knowledge, and the influence that profit-making has on the selection, organization and distribution of curricular knowledge. Due to a market-driven production system in the US and other nations (Larson, 2007), the publishing industry plays a disproportionately large role in determining what counts as knowledge in schools. Ostensibly, the goals of standardizing curricular texts are to promote consistency across schools, districts, and states; to establish a common knowledge base (national history, identity, language, and culture); and to ensure quality control. More often than not, however, corporate standardization and government regulation of textbook publishing impedes innovation and flexibility on the part of teachers and local schools. This results in the deskilling of teachers as professionals, and the production of texts that are characterized by superficial and biased treatment of topics, and include irrelevant materials and unchallenging tasks. Given the focus on competing in a global marketplace, collaborations between governments and corporate publishers usually result in textbooks aimed at covering as much material as possible (i.e., breadth over depth), and usually in ways that represent and reinforce the perspectives of dominant groups.

The power that textbooks have in shaping educational practice comes, in large part, from the dominant models of knowledge, teaching, and learning that so often accompany the use of textbooks. The heavy reliance on textbooks represents a traditional concept of knowledge as static and objective; a model that comes out of Enlightenment views of knowledge and rationality that form the basis of traditional common core and efficiency models of curriculum. This hegemonic view of knowledge is often paired with a static notion of teaching and learning as the transmission of objective knowledge. The model can be

summarized as follows: Knowledge resides in the textbook and learning is just a matter of getting it out of the textbook and into the heads of students. Studies of educators' beliefs about and uses of textbooks often confirm this view. For example, Shannon's (1992) work has shown that teachers and administrators often believe that the textbooks themselves can teach reading, irrespective of the role of skilled teachers. Thus, textbooks deliver the curriculum and shape the daily routines of classrooms. Powerful collaborations between governments and private publishing corporations determine what counts as quality in textbooks by focusing on standardized outcomes measured by tests produced by the same government/publisher collaboration. Specifically, the textbook has been transformed in a powerful "multinational phenomenon that can be adapted, translated, and niche marketed in a range of national and regional markets; a comprehensive suite of educational commodities with a pedagogic reach that extends far beyond children's narrative reading text; a scientifically 'tested' and 'proven' product" (Luke *et al.*, 2003: 251).

In what follows, the article moves to discuss further issues around control of textbook content and publishing in a global marketplace and some of the consequences of disciplinary practices associated with maintaining control over knowledge production (Foucault, 1977).

An Overview of Curriculum Publishing in a Global Context

A basic tension can be seen between national contexts in which decisions about texts are made locally, and contexts in which text production and/or adoption decisions are determined by state or national governments, in collaboration with publishing companies. Since colonial times, the production of textbooks has been imbalanced, with Western European countries, and then later the US and Japan, having the material and political capital to produce texts for dissemination to colonized states or developing nations. Altbach (1995) argues that textbooks are increasingly international commodities, given that multinational firms are increasingly the dominant producers of textbooks. However, the tension between global corporations and national publishing remains. During the opening up of textbook production in postwar Taiwan, for example, hegemonic struggles over school knowledge and textbook production focused on nationalizing Chinese culture, and the decentralization of curricular decision making forced the state to guide production of official knowledge from a distance (Chen, 2005).

Major text-publishing companies' domination of the international market and perpetuation of cultural domination remains a fact of life. Apple (2000) argues that this is partly because of the economic control of communication

by multinational firms and partly because of ideological control over official knowledge by new elites in former colonial states. Knowledge production is linked to new social and economic formations, cultural globalization, and geopolitical instabilities in the construction of what counts as official knowledge, text, discourse, and discipline (Nozaki *et al.*, 2005). Moreover, textbooks serve as a tool of bodily and ideological regulation that enforce a sense of duty, morality, nationalism, and cultural continuity in whatever society they are used and are vital in the construction of national identities (Chen, 2005).

In many developing countries, the state maintains controls over textbook form, content, production, and distribution; in the north and west, multinational publishing companies are the link between policy, assessment, and accountability systems in textbook production and consumption (Luke *et al.*, 2003). The reliance on multinational firms in the Third World is due to a variety of factors, including the shortage of paper in these contexts, as well as limited infrastructure for printing. Also at stake are disagreements over whether textbooks should be publicly or state produced.

In spite of a general agreement that multinational firms should not produce textbooks for Third World countries (Altbach, 1995), international corporations maintain a near monopoly on expertise, publishing infrastructure, and access to paper and capital (Limage, 2005). With the exception of some larger countries who can afford to maintain publishing industries, developing countries are dependent on centralized production and distribution controlled by others. One serious consequence of this inequity is international homogenization as profit-oriented companies standardize a curriculum that can be repeated year after year in order to maximize profit (Limage, 2005; Starnes, 2004). Limiting divergence of content is also applied in colonized African nations. Newly independent nations expanding their educational systems were prime markets for publishing companies who could use existing infrastructure to produce Africanized textbooks that adapted language and national curricular guidelines but were published by multinational corporations in Europe (Limage, 2005). Given their dominance in these markets, multinational corporations have little incentive for supporting the development of indigenous publishing infrastructures.

National curricula have been developed in countries such as England, France, and Cuba. In British schools, textbooks are the most widely used resource for teaching and learning, yet there remains a high level of negativism toward them (Marsden, 2001). The dominant view is that “textbooks undermine professionalism, typify an undesirable transmission model of teaching and learning, and are generally incompatible with progressive educational practice” (Marsden, 2001: 1).

In Asia, publishing has often not followed a Western model; however, textbooks quickly became a political tool

after World War II, when some countries insisted on government control of educational materials (Taylor, 1995). As in other contexts, textbook production has been bound up with the assertion of and resistance to certain representations of national interest. As mentioned earlier, Chen (2005) shows how textbook deregulation in postwar Taiwan contributed to a more democratic educational system and more diverse curricula in response to struggles over school knowledge and textbook production focused on nationalizing Chinese culture.

This article contextualizes the issues associated with knowledge production, curriculum, and the politics of curriculum publishing at the national level in the following discussion of literacy curriculum in the US.

The Political Economy of Literacy Curriculum in the Contemporary US: A Case Example

In the US, which has never officially produced national curricula, recent corporate consolidations have limited the number of publishing houses to four companies that now dominate the market: Pearson, McGraw-Hill, Reed Elsevier, and Houghton Mifflin (Sewall, 2005). This small circle of corporate publishers connects policy power with publishing in Western contexts, although the same corporations have expanded to Asian markets either within their current corporate structure or by taking over local publishing companies. Public and private zeal for reform has stimulated sales of textbooks, and the effect of fluctuating markets, corporate consolidation, and restrictive government policies about the content of curricula – and what counts as scientifically-based evidence of effectiveness – has led to a situation in which these four publishing houses produce 80% of the textbooks used in US schools (Sewall, 2005).

Further complicating matters in the US market is the dominance of the state adoption system in which state governments give an approved list of books for school districts to purchase. Given the relative lack of profitability in the textbook sector, publishers focus content on the largest markets (California, Texas, and Florida). Other states and local districts, even if they are not in a state with a mandated adoption policy, often purchase these same textbooks. In other words, publishers normalize content in order to make the largest profit. This normalization results in a “dumbing down” of the content as publishers vie to capture the largest market share (Sewall, 2005). US education policies, notably, the No Child Left Behind Act (NCLB), have made this much worse (Larson, 2007).

In the following section, we show how NCLB’s list of approved texts/companies amount to the restriction of the sources and kinds of educational materials used in

schools through a detailed description of one such corporate educational package, America's Choice. The rise in accountability and standardization associated with NCLB in the US has brought with it deeper, more reductionist calls for fixes to the problem of student performance indexed by annual yearly progress (AYP) on standardized tests. This pressure for rising scores has resulted in a plethora of curricular fixes that include Reading First, Success for All, and America's Choice. The lower performing schools in the US are overwhelmingly located in high-poverty areas in which non-dominant families struggle. The targets of these fixes, then, are pathologized as deficit and in need of repair (Larson, 2007). Under NCLB, schools are forced to adopt these packages to demonstrate AYP in particular ways or lose their funding. These packages include textbooks tightly coordinated to tests, and encourage highly scripted pedagogy.

Recent public hearings on the development and mandated use of Reading First in the US are more examples of the conflict inherent in private/state relationships that is coming to light (Manzo, 2007). The recent release of the Inspector General's audit of Reading First that reveals some of the inappropriate contracts and serious flaws in the research base for the program shows how deep the roots of corporate profit motive may lie. Furthermore, by limiting what counts as valid curricular materials to that which is deemed scientifically based, NCLB narrows local choices while improving corporate profit (Starnes, 2004). The case of America's Choice is a telling example of how a government provides regulations and financial resources while the private sector provides the materials.

America's forced choice

Paired with the high-stakes accountability system of rewards and punishments found in the NCLB and restrictive requirements for Reading First grants, America's Choice represents a forced choice in literacy instruction for many urban schools (Osborn, 2007). Furthermore, by focusing on low-performing and disadvantaged schools and students, America's Choice serves to perpetuate a parallel tracking system in which students from urban schools receive instruction that focuses on low-level skills in isolation of actual practice, while students in many suburban schools have opportunities for more authentic curricular and instructional practices.

America's Choice emphasizes systematic instruction and well-aligned instructional materials for higher rates of implementation and stronger effects (National Center on Education and the Economy, 2002: 1). The origins of America's Choice, a for-profit subsidiary of the National Center on Education and the Economy (NCEE), can be found in the scientific management and social efficiency movements of the early 1900s and extends to NCLB and

current federal mandates and initiatives, drawing on the discourse of business and industry with the influence of "A Nation at Risk" (1983) (National Commission on Excellence in Education, 2001), the National Business Council, and NCEE (Osborn, 2007). Like similar packaged literacy programs, America's Choice incorporates the language of federal standards and accountability mandates for scientifically based programs to facilitate adoption and ensure government funding.

America's Choice includes evaluations, curriculum that provides explicit and systematic instruction, a classroom organization system, routines and rituals for students that promote on-task behavior, and a data analysis tool. As such, America's Choice delivers what Luke *et al.* (2002) refer to as "a whole suite of 'teacher-proofed' curricular commodities" as well as materials to "assess teacher and system efficiency at delivering the whole package" (p. 251). This range of materials represents a trend toward total learning systems that include guides and workbooks, CDs, films, and technology-based tools. In addition, America's Choice provides professional development and on-site technical assistance, thereby exemplifying a recent development in the publishing industry to increase profits by providing support for a fee. In this way, money that otherwise would have been spent for professional development to deepen teachers' pedagogical knowledge, understanding of child development, and awareness of innovative practices is used to assure consistent and efficient delivery of the package. Adoption of America's Choice School Design costs between \$70 000 and \$105 000 per school.

America's Choice boasts that it "is built on the premise that teaching to explicit standards is the best strategy for helping disadvantaged and low-performing students — the same strategy adopted by the authors of NCLB" (National Center on Education and the Economy, 2006).

When urban schools choose America's Choice because it will facilitate Reading First grant money, or because they are deemed low performing, it presents these schools with a forced choice that dictates a certain kind of instruction that is supposed to be good for disadvantaged students and low-performing schools. From the perspective of America's Choice and similar packaged programs, improving literacy is just a matter of breaking skills down to more discrete units, and making instruction more systematic and explicit. Adoption of programs like America's Choice diverts attention from issues of race and poverty, and discussion of culturally relevant curriculum. At a time when literacy curriculum and instruction should encourage students to learn a range of discourses and modes of communication (Johnson and Kress, 2003), America's Choice focuses instruction on narrow, limiting conceptualization of literacy that is a mandate rather than a choice for many urban schools and contributes to the development of narrow and limiting identities for students.

The role of America's Choice in US school reform is illustrative of the deep and often hidden collaborations between educational policymakers and for-profit publishing companies. The strategic alignment of policies, curriculum, standardized tests, and school reform by government officials and corporate publishers is, in the current US context, generating profits for publishers, and dramatically inhibiting educators' decision-making about texts, teaching, and learning at the local level.

This coordination among educational policy, curriculum, and publishers constitutes a kind of disciplinary technology (Foucault, 1977) that has major consequences for what counts as knowledge in schools. At stake in these complex political and economic processes is not just the content that is produced for use in schools, but fundamentally, conflicting definitions of the common good and the purpose of schooling.

Conclusion

Recent years have been marked by a sharply increased alliance between textbook publishers and various levels of government through implementation of educational mandates and increasing reliance on high-stakes standardized tests, as well as the establishment of national curricula in some countries. A focus on accountability has resulted in narrow, selective, and minimalist goals that facilitate control over educational content and processes. An emphasis on efficiency and economic competitiveness characterized by standards and accountability movements, and reified through textbooks, has silenced discussion about other possible goals for education.

Textbooks promise discipline, standardization, and accountability in the mass delivery of instruction (Luke *et al.*, 2002). The seemingly unquestioned authority and privilege granted to the knowledge and pedagogy embodied in textbooks has become an invisible and pervasive part of the professional socialization of teachers, as well as the overall socialization of students and society at large. As a result of this domination, particular viewpoints about curriculum, instruction, and the very organization of schools are validated and maintained.

In part, textbooks have been accorded privilege because of presumed objectivity and neutrality in their content. Textbooks, however, are far from neutral. What counts as knowledge and pedagogy in textbooks is selected from a much larger universe. Certain principles, practices, and meanings are chosen for inclusion and emphasis while others are excluded, reinterpreted, or diluted. The exclusion of knowledge and alternative voices from textbooks represents a technique of disciplining power that pathologizes that which is outside of the norm (Foucault, 1977). In this way, students' identities are deformed by what is excluded from textbooks – absences, denials, and incompleteness.

If the examination embodied in high-stakes standardized tests provides “a normalizing gaze, a surveillance that makes it possible to qualify, to classify and to punish” (Foucault, 1977: 197), then traditional school curriculum provides the normalizing voice that accompanies and directs this gaze. Publishing companies codify this normalizing voice in the form of curriculum and programs like America's Choice. The new frontier in educational publishing – the production of full-service packages of educational materials, including textbooks, computer programs, test preparation supplies, and instructional materials (Osborn, 2007) – strengthens the power of these programs to mold teaching and learning in school, while at the same time remaining largely invisible. Like the normalizing gaze of the examination, the disciplinary power of curriculum is exercised through its invisibility; invisibility achieved through a focus on sameness, and supposed neutrality, which legitimates and may even disguise the exercise of power.

Current trends in curriculum publishing increasingly treat knowledge as a commodity that can be developed, produced, and marketed by private industry to a variety of national and local markets. While government and corporate publishers of curricula often assert their commitment to high-quality materials, a static notion of knowledge and learning undergirds the production of these educational commodities. Packages of coordinated educational products are often meant to inject standards into schools, and teacher proof the content to which students are introduced. A key question in this current context is what do teachers do with these powerful mediating forces. It seems that some follow textbooks closely, believing they can teach (Shannon, 1992) and, more recently, many follow them for the scope and sequence connected to mandated testing.

In summary, publishers of curricula have an enormous influence on the kinds of knowledge and texts that teachers and students have access to in schools. This influence is mediated by government policies; the extent to which decisions about textbook adoption are made at local, state, or national levels; and whether nations produce their own texts or, increasingly, contract with multinational firms who produce and adapt already-existing texts to particular national markets. Given that textbook production and distribution is an increasingly global enterprise, dominated by multinational firms with profit as their main motive, the relationship between curriculum and the publishing industry deserves serious scrutiny. The cases of NCLB and America's Choice in the US point to the key role that government policies and agendas play in endorsing particular curricula and publishers. Attention should be given to who makes consequential decisions about the production and codification of knowledge in the form of curricular materials, and who is benefiting – economically, politically, and culturally – from these globalized and unequal processes of production and consumption.

Bibliography

- Altbach, P. (1995). Publishing in the third world. In Altbach, P. and Hoshino, E. (eds.) *International Book Publishing: An Encyclopedia*, pp 271–278. New York: Garland.
- Apple, M. (2000). *Official Knowledge: Democratic Education in a Conservative Age*, 2nd edn. London: Routledge.
- Chen, J. (2005). Official knowledge and hegemony: The politics of textbook deregulation policy in Taiwan. In Nozaki, Y., Openshaw, R., and Luke, A. (eds.) *Struggles Over Difference: Curriculum, Texts, and Pedagogy in the Asia-Pacific*, pp 59–78. New York: SUNY.
- Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison* (trans. Sheridan, A.). Harmondsworth: Penguin.
- Johnson, D. and Kress, G. (2003). Globalization, literacy and society: Redesigning pedagogy and assessment. *Assessment in Education* 10, 5–14.
- Larson, J. (2007). *Literacy as Snake Oil: Beyond the Quick Fix*, 2nd edn. New York: Lang.
- Limage, L. (2005). Political economy of textbooks and literacy. *Paper Commissioned for the EFA Global Monitoring Report 2006, Literacy for Life*. Paris: UNESCO.
- Luke, A., Carrington, V., and Kapitzke, C. (2003). Textbooks and early childhood literacy. In Hall, N., Larson, J., and Marsh, J. (eds.) *Handbook of Early Childhood Literacy*, pp 249–257. London: Sage.
- Manzo, K. (2007). 'Reading First' information sent to justice department. *Education Week April 26*.
- Marsden, W. (2001). *The School Textbook: Geography, History, and Social Studies*. London: Routledge.
- National Center on Education and the Economy (2002). *America's Choice School Design: A Research-Based Model* (Brochure). Washington, DC: National Center on Education and the Economy.
- National Center on Education and the Economy (2006). *Ramp-Up Literacy: A Comprehensive System that Accelerates Reading and Writing Performance* (Brochure). Washington, DC: National Center on Education and the Economy.
- National Commission on Excellence in Education (2001). A nation at risk: The imperative for excellence in education. In Jossey-Bass Publishers (ed.) *Jossey-Bass Reader on School Reform*, pp 136–160. San Francisco, CA: Jossey-Bass.
- Nozaki, Y., Openshaw, R., and Luke, A. (eds.) (2005). *Struggles Over Difference: Curriculum, Texts, and Pedagogy in the Asia-Pacific*. New York: SUNY.
- Osborn, D. (2007). Digging up the family tree: America's forced choice. In Larson, J. (ed.) *Literacy as Snake Oil: Beyond the Quick Fix*, 2nd edn., pp 171–188. New York: Lang.
- Sewall, G. (2005). Textbook publishing. *Phi Delta Kappan* 86, 498–502.
- Shannon, P. (1992). *Becoming Political: Readings and Writings in the Politics of Literacy Education*. Portsmouth, NH: Heinemann.
- Starnes, B. (2004). Textbooks, school reform, and the silver lining. *Phi Delta Kappan* 82, 170–171.
- Taylor, S. (1995). An Asian overview. In Altbach, P. and Hoshino, E. (eds.) *International Book Publishing: An Encyclopedia*, pp 424–429. New York: Garland.

Further Reading

- Altbach, P. and Hoshino, E. (eds.) (1995). *International Book Publishing: An Encyclopedia*. New York: Garland.
- Apple, M. (1986). *Teachers and Texts*. New York: Routledge and Kegan Paul.
- Apple, M. (2003). Cultural politics and the text. In Ball, S. (ed.) *The RoutledgeFalmer Reader in Sociology of Education*, pp 179–195. London: RoutledgeFalmer.
- Chambliss, M. and Calfee, R. (1998). *Textbooks for Learning: Nurturing Children's Minds*. Oxford: Blackwell.
- De Alba, A., González-Gaudiano, E., Lankshear, C., and Peters, M. (2000). *Curriculum in the Postmodern Condition*. New York: Lang.
- Larson, J. (2007). *Literacy as Snake Oil: Beyond the Quick Fix*, 2nd edn. New York: Lang.
- Marsden, W. (2001). *The School Textbook: Geography, History, and Social Studies*. London: Routledge.
- Montagne, I. (2000). *Textbook and Learning Materials, 1990–1999. Thematic Studies for Work Education Forum. Education for All Assessment*. Paris: UNESCO.
- Nozaki, Y., Openshaw, R., and Luke, A. (eds.) (2005). *Struggles over Difference: Curriculum, Texts, and Pedagogy in the Asia-Pacific*. New York: SUNY.
- US Department of Education Office of Inspector General (2006). *The Reading First Program's Grant Application Process, Final Inspection Report*. <http://www.ed.gov/about/offices/list/oig/aireports/113f0017.pdf> (accessed April 2009).

Relevant Websites

- <http://www.sraonline.com> – The McGraw-Hill companies.
- <http://www.ed.gov> – US Department of Education Reading First Internal Audit.

Curriculum and Teacher Change

J G Ladwig, University of Newcastle, Newcastle, NSW, Australia

©© 2010 Elsevier Ltd. All rights reserved.

At least since its formation in the early twentieth century, curriculum has become almost synonymous with questions of educational reform and change. The very notion of curriculum development is fundamentally embedded in a modernist commitment to progress and reform: development implies change, and in modern times, education implies development (see, e.g., Popkewitz, 1991, 2004). When the focus of curriculum development is within schools, where teachers are the people who ultimately deliver curriculum, the commitment to curriculum reform has long implied a commitment to teacher change. Thus, for at least the past century, teacher change has been a major focus of attention in educational research and practice. Literally hundreds of thousands of references to teacher change are found in the archives of the twentieth- and the twenty-first-century educational literature in English-language databases alone. The commonality of calls for ways and means to change teachers is so prevalent that it has become part of the taken-for-granted hegemonic background of education globally.

One of the reasons a focus on teacher change has become so prominent comes from within the curriculum field itself. The connection between curriculum development and teacher change has been recognized throughout the twentieth and into the twenty-first centuries and is part of the mainstream canon of curriculum development. From Bobbitt's seminal push for efficiency and his adaptation of Taylorist principles of labor management in 1918, through Tyler's *Basic Rationale* in 1949, to Taba's expansion of Tyler in 1962, or from Smith, Stanley, and Shore's broader outline of curriculum development in 1950, up to the contemporary traditionalists of curriculum such as Tanner and Tanner, virtually all of the major mainstream addresses of curriculum development come to the question of teacher change in one form or another. The reason for this connection is clear once it is recognized that whatever the official curriculum may be, what students experience as curriculum fundamentally depends on what teaching they experience.

While several explanations can be, and have been, forwarded to understand this reality, there are two basic observations which help understand why the focus on teacher change has become so embedded in curriculum development. On the one hand, prior to some recent post-modern critiques of curriculum which decry development *per se*, all major paradigms of curriculum development shared a common commitment to teacher change no matter how fundamental the paradigmatic differences debated were. On the other hand, no matter how much

attention and overt research and action have been dedicated to teacher change, changing teaching and teachers in some consciously planned manner remains an unsolved dilemma for educational reformers (cf. Richardson and Placier, 2001). Add to this background one of the emerging understandings from research into instructional improvement – that substantial improvement in teaching requires significant changes in teachers. Taken together, it is clear that whoever wants improvements in the quality and nature of school curriculum simultaneously commits to both curriculum development and teacher change. Let us consider each of these points in turn.

The question of how curriculum development links to teacher change, no matter what the theoretical view is, can be demonstrated whether one draws together a very broad view of the curriculum field, or if more specific differentiations of curriculum perspective are developed. From a broad view, recent studies in curriculum that analyze the curriculum field as formed by multiple discourses include a wide range of curriculum perspectives. For example, Pinar and his colleagues address curriculum as different forms of texts: political, racial, gendered, phenomenological, post-modern, autobiographical/biographical, esthetic, theoretical, and institutionalized (Pinar *et al.*, 2002). In this view, curriculum development is identified specifically with the institutionalization of curriculum. However, even in this broad view, it is clear that no matter what the discursive perspective taken, the link between any notion of curriculum change and teacher change is pragmatically implicit. It is possible to analyze, theorize, and comment on curriculum without concerning oneself with teacher change, but it is not possible to change curriculum without also considering the people who will deliver it. From a more specific view of curriculum, Snyder *et al.* (1992) outline three perspectives in their review of curriculum implementation: the fidelity, mutual adaptation, and enactment perspectives. Each of these perspectives includes a role for the teacher as an agent of curriculum development, from the most proactive (the enactment perspective) to the least (the fidelity perspective).

From the perspective of curriculum development, the reason it is so important to recognize the link with teacher change is to confront a basic reality, that is, despite more than a century of attempts to change the educational experiences of students through curriculum reform, the realities of the classroom have not shifted all that much. This common finding in American research can also be noted globally. In Australia, for example, where there has

been a significant shift in the official, centralized, curricula of State authorities, there is little reason to believe that classrooms have changed considerably. In the words of one of Australia's most widely acclaimed curricularists, in his lament on the lack of pedagogical change in Australia after decades of dramatic curriculum changes, the late Garth Boomer noted, "It could well be that the last thirty years has been a matter of change on paper, rather than change in essential practice" (Boomer, 1999: 127). So, working simply from the question of curriculum development, the need to understand and find ways to enact teacher change is clear.

Another reason teacher change has become central in educational research and reform over the past decade or so has come from studies of educational reform and school effects. When focusing on what schools can do to improve student outcomes (typically, but not always, meaning student academic outcomes), research into what contributes the most to student outcomes has long pointed out the importance of differences between teachers and pedagogy. That is, when considering what levers of change a school can reasonably impact directly, the main focus has to be on the quality of teaching and teachers. (There are, of course, many other factors that schools cannot directly leverage.)

Sociologists of education have long spoken of variance within schools as a major contributor to the differential educational outcomes, at least since the 1966 Coleman Report (Coleman *et al.*, 1966). Picking up on this line of reasoning, sociological studies of streaming and tracking have also pinpointed differences in the nature of curriculum and instruction as key factors in understanding within-school effects. Any attempt to change the effects of streaming and tracking therefore implies changing curriculum and teaching. Many of the factors that contribute to differential effects of streaming and tracking are directly related to aspects of curriculum and teaching that vary significantly among teachers. In this way, even from the less-direct path of analyzing the social implications of schooling, the focus of school improvement eventually turns to the question of teacher change.

Levers and Dimensions of Teacher Change

With the focus on teacher change, research on curriculum development has identified several dimensions of teachers and teaching which need to be changed if pedagogical change is the desired result. Additionally, and consequently, to get these dimensions of teaching to change, research has attempted to identify levers which might lead to such changes. Around many of these levers and dimensions of teacher change, there is something of an uncontested consensus. Seven of these are discussed below.

Instructional Techniques

The historical legacy of curriculum development is built on what is now seen as a traditional split between curriculum and instruction. While it is now common to speak against this split as a simplistic dichotomy, it is equally true that many current reform efforts to improve teaching leverage a sharp distinction between the development of curriculum and the specific teaching techniques intended to deliver the goods of that curriculum. Perhaps the most obvious example of these can be found in the Success for All reform initiative in the United States (Slavin and Madden, 2001). Within this tradition, the question of teacher change is largely framed in terms of the implementation of pre-specified teaching techniques which can occur with varying degrees of implementation fidelity. This understanding of curriculum development and teacher change is consistent with and at least as old as Bobbit's Taylorist understanding of schooling.

Teacher Beliefs

Where studies in implementation fidelity have done a fine job of documenting failed or incomplete reform initiatives, they have often pointed out that the application of teaching techniques is likely to have a limited scope and range of success, without necessarily changing the beliefs teachers bring to their work. The focus for changing teacher beliefs can be a question of their beliefs about learning, teaching, students, subject areas, knowledge, and epistemology (Calderhead, 1996). While the specific focus of belief differs in each line of reasoning, the basic logic remains similar. If teacher behaviors are to change more than temporarily, across multiple contexts, then something more durable than a set of techniques is required to carry on along with the teacher herself. The idea of focusing curriculum development on the implementation of specific teaching techniques almost by definition precludes lasting change, to the extent that applications of technique are almost always context and time specific.

Content Knowledge

Since before the turn of the past century, one of the central questions of teacher change, whether from the point of view of specific pedagogical practices or larger curricular reform, has been a persistent questioning of teachers' content knowledge and expertise. At least from the publication of the first issue of *The School Review* in 1893 in the USA, in those heady days of secondary school curriculum reform advocated by The Committee of Ten, the question of the content expertise needed by teachers was central in the minds of those who sought to professionalize teaching (Hart, 1893). For example, later in the

same year, Aven Nelson (1893) decried the need for teachers to keep up with the growing knowledge demands of an ever increasing crowded science curriculum in high schools. Given the relatively stable nature of school practice and the continually growing nature of knowledge, it is perhaps not surprising to find similar calls for advancement of teachers' content knowledge in contemporary school reform literature. Current research has been decisive in measuring the ultimate effect of teachers' content knowledge on student achievement (Hill *et al.*, 2005). From this, it is not a far stretch to find policy implications drawn for any attempt to improve curriculum development that point out the need to improve teachers' content knowledge as a lever of change. Whereas early calls for teachers to improve content knowledge may have been fairly blunt and simple, contemporary focus on this area of teacher change sees the need to link teachers' content knowledge with the materials intended to be used in the reformed curriculum.

Pedagogical Content Knowledge

Since Shulman's 1980s call for a renewed reform focuses on teachers and teaching, knowledge about this link between specific content knowledge and instructional practices designed to maximize students' learning has been dubbed pedagogical content knowledge, and is seen to be the special province of teachers and teacher educators (Shulman, 1987). As perhaps the one form of knowledge unique to the profession, as noted by Hart (1893) above, the need to center curriculum development and teacher change on pedagogical content knowledge follows clearly from his position, and has been touchstone in the literature since Shulman's articulation of the idea. Around the time of Shulman's appeal, the question of changing teachers' pedagogical content knowledge figured in several areas of curriculum reform, notably in mathematics education (Peterson *et al.*, 1989). Since then, it has been a reference point for commentary on teacher education and professional development around the globe. The concern for developing teachers' pedagogical content knowledge in specific subjects continues in most, if not all, subject areas. For example, 124 years after Hart's 1893 lament about the needs of science teachers, similar calls can be found in contemporary educational research literature (e.g., Nahum *et al.*, 2007).

Dispositions to Their Own and Others' Learning

While the question of changing teachers' techniques, beliefs, and knowledge leaves much work for would-be curriculum developers, studies on the work of teachers sheds another light on the observation that teacher change is not altogether common. From a sociological view,

Lortie's (1975) seminal study of teachers and teaching has made it clear that any attempt to understand change in teaching must also recognize the durable dispositions that are developed while working as a teacher. Richardson and Placier (2001) pick up on this insight in their review of teacher change, pointing out that intentional attempts to change teachers are but some of the factors that lead teachers and teaching to be as they are and change as they do. In a sense, the literature on teachers' dispositions is linked to questions of beliefs, as is noted by some authors, but dispositions are seen to be less consciously accessible, being the individual equivalent of culture. And it is this link to the culture of schooling and teaching which drives several analyses of the lack of successful reform in curriculum development and teacher change (e.g., Sarason, 1993). The question of situating teachers into a broader social context that builds from and carries its own culture is the linchpin in the now-ubiquitous calls to see teachers as part of learning communities.

Professional Communities of Teacher Learning

In the wake of a two-decade push for teachers to adopt Vygotskian insights into the socially situated nature of learning, and pedagogical reforms that appeal to constructivist ideas, there is a consistency to the concurrent push to understand and develop teacher learning communities more broadly as well as in schools. From a focus on the collective nature of teachers' beliefs and sense of efficacy, to foci on school restructuring, to more skeptical theoretical examinations of the idea of teachers working in communities, it would be difficult to read any contemporary research on teacher change that does not in some way include the idea of a teacher professional community, or professional learning community. As Grossman *et al.* (2001) point out, the idea of community is now so overused that it has virtually become meaningless. Nevertheless, be it from a specific focus on teacher change (Richardson and Placier, 2001), or from a concern for teachers' professional development (Borko, 2004), the question of teacher change is clearly now recognized as not getting far without also taking their professional and social context into account. The logic behind this view carries the weight of common sense: however much teachers are effective individually, they are not likely to change their teaching on their own, since they do not work alone. This view, and the continued lack of successful intended teacher change, has led some researchers to focus on understanding how schools might support teacher change organizationally.

Organizational Change

The school restructuring research that was contemporaneous with Shulman's call to understand teachers' pedagogical content knowledge yielded a focus on schools'

organizational capacity long before the current interest in understanding organizational strategies to support teachers' learning. Newmann *et al.* (1996) began focusing on organizational supports for restructuring schools in the early 1990s. Following on from the work directed by Newmann, which focused on understanding how school restructuring could promote authentic pedagogy, researchers in this lineage have maintained a keen interest in understanding how the organizational efforts of schools and school systems might ultimately change teaching. This line of research has addressed questions of accountability, the relationship between professional development and school organizational capacity, teacher learning, school-wide inquiry, and collaborative teacher development, as part of a larger model of school organizational capacity building. Examining the role of leaders at school and system levels in attempts to steer curriculum development and teacher change follows the same analytical trajectory of this line of research.

Conclusion: Teacher Change as an Unintended Consequence of Policy Initiatives

The focus on teacher change in literature that links it with curriculum and pedagogical development has historically been ever hopeful, but the current literature related to the issue is frustrated by globally ubiquitous contemporary policies. While significant investments have been made in gaining implementation fidelity, there is little reason to believe that dominant policy initiatives akin to those deployed in the US are likely to leave any lasting improvement in teaching or learning. Elsewhere, the messages of developing professional communities of teachers, with foci on beliefs, knowledge, pedagogical content knowledge, and the like are pushing along. In the meantime, even as some US research has documented how policy in that country might best facilitate teacher change and curriculum development (Cohen and Hill, 2001), it seems the continued capacity to generate unintended consequences from policy initiatives will keep US researchers' attention, at least for a while longer. For example, while resource allocations to schools and school districts become ever more starkly distributed in the US, it appears that one unintended side effect of this has been in further institutionalizing the inequitable distribution of teachers. A growing production of international comparisons makes it increasingly clear that there are significant differences in how teacher change occurs within systems as part of an internally natural progression, but at the same time, within any one of those systems, attempts to purposively shift history's tides and consciously direct teacher change remain merely wishful aspirations.

Here, the critical call for understanding curriculum as part of the hegemonic apparatus of the school might perhaps be the most realistic call for teacher change. Where Apple's (1979) foundational analysis of curriculum in its larger social and political context leads him to call for all parties involved in curriculum development to become organic intellectuals, it is perhaps not all that idealistic to suggest that teacher change and curriculum development ultimately lie in the capacity of teachers to take up the challenge of being public intellectuals.

See also: Curriculum Planning and Systems Change; Curriculum Reform; Pedagogical Content Knowledge; School Reform and Restructuring: Self Managing School; Taking Prospective Teachers' Beliefs into Account in Teacher Education.

Bibliography

- Apple, M. W. (1979). *Ideology and Curriculum*. London: Routledge and Kegan Paul.
- Boomer, G. (1999). Curriculum and teaching in Australian schools, 1960–1990: A tale of two epistemologies. In Green, B. (ed.) *Designs on Learning: Essays on Curriculum and Teaching* by Garth Boomer, pp 127–146. Canberra: Australian Curriculum Studies Association.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher* **33**, 3–15.
- Calderhead, J. (1996). Teachers: Beliefs and knowledge. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 709–725. New York: Macmillan.
- Cohen, D. K. and Hill, H. C. (2001). *Learning Policy: When State Education Reform Works*. New Haven, CT: Yale University Press.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., *et al.* (1966). *Equality of Educational Opportunities*. Washington, DC: United States Government Printing Office.
- Grossman, P., Wineburg, S., and Woolworth, S. (2001). Toward a theory of teacher community. *Teachers College Record* **103**, 942–1012.
- Hill, H. C., Rowan, B., and Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal* **42**, 371–426.
- Lortie, D. C. (1975). *School Teacher: A Sociological Study*. Chicago, IL: University of Chicago Press.
- Nahum, T. L., Mamlok-Naaman, R., Hofstein, A., and Krajcik, J. (2007). Developing a new teaching approach for the chemical bonding concept aligned with current scientific and pedagogical knowledge. *Science Education* **91**, 579–603.
- Nelson, A. (1893). Science teaching in our public schools. *The School Review* **1**(8), 471–481.
- Newmann, F. M. and Associates (1996). *Authentic Achievement: Restructuring Schools for Intellectual Quality*. San Francisco, CA: Jossey-Bass.
- Peterson, P. L., Fennema, E., Carpenter, T. P., and Loef, M. (1989). Teachers' pedagogical content beliefs in mathematics. *Cognition and Instruction* **6**, 1–40.
- Pinar, W. F., Reynolds, W. M., Slattery, P., and Taubman, P. M. (2002). *Understanding Curriculum: An Introduction to the Study of Historical and Contemporary Curriculum Discourses*. New York: Peter Lang.
- Popkewitz, T. S. (1991). *A Political Sociology of Educational Reform: Power/Knowledge in Teaching, Teacher Education, and Research*. New York: Teachers College Press.

- Popkewitz, T. S. (2004). The alchemy of mathematics curriculum: Inscriptions and the fabrication of the child. *American Educational Research Journal* **41**, 3–34.
- Richardson, V. and Placier, P. (2001). Teacher change. In Richardson, V. (ed.) *Handbook of Research on Teaching*, pp 905–947. Washington, DC: American Educational Research Association.
- Sarason, S. B. (1993). *The Predictable Failure of Educational Reform: Can We Change Course before It's Too Late*. San Francisco, CA: Jossey-Bass.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the New Reform. *Harvard Educational Review* **57**, 1–19.
- Slavin, R. E. and Madden, N. A. (2001). *One Million Children: Success for All*. Thousand Oaks, CA: Corwin Press.
- Snyder, J., Bolin, F., and Zumwalt, K. (1992). Curriculum implementation. In Jackson, P. W. (ed.) *Handbook of Research on Curriculum*, pp 402–435. New York: Macmillan.

Further Reading

- Bobbitt, F. (1918). *The Curriculum*. Boston: Houghton Mifflin.
- Bobbitt, F. (1920). Mistakes often made by principals, Part 1. *Elementary School Journal* **20**, 337–346.
- Hart, A. B. (1893). The teacher as a professional expert. *School Review* **1**, 4–14.
- Smith, B. O., Stanley, W. O., and Shores, J. H. (1950). *Fundamentals of Curriculum Development*. Yonkers-on-Hudson: World Book.
- Taba, H. (1962). *Curriculum Development: Theory and Practice*. New York: Harcourt, Brace and World.
- Tyler, R. W. (1949). *Basic Principles of Curriculum Instruction*. Chicago, IL: Chicago University Press.
- Tanner, D. and Tanner, L. (2007). *Curriculum Development: Theory into Practice*, 4th edn. Upper Saddle River, NJ: Pearson Education.

Curriculum Governance and Planning

B Levin, University of Toronto, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The development of curriculum by governments or other sanctioned authorities for standard use in schools across a state, province, or country inevitably involves elements of both policy and politics, because it involves questions of public values (Manzer, 1994). The term politics here includes a wide range of informal influences and larger social processes in addition to the standard processes of elections and party politics.

Most frameworks for understanding politics and policy formation address similar topics, such as the institutional setting and structure around decisions (e.g., Crowson *et al.*, 1996), how issues come to be on the political agenda (e.g., Kingdon, 1994), and the processes through which decisions are made in the face of different interests and points of view (e.g., Ball, 1990).

The discussion in this article uses a heuristic framework drawn from Young *et al.* (2007), who propose analyzing political decisions in terms of five overlapping categories: issues, actors, processes, influences, and results.

In making decisions about curriculum governments are subject to the usual pressures and constraints that form the political process (e.g., Edelman, 1988). Citizen preferences are critical, since getting elected is a fundamental concern of every government. Citizens are not necessarily interested in or knowledgeable about every public issue, which is why there is pressure on politicians to communicate in very simple messages. At the same time, citizens can care deeply and feel strongly about issues about which they know very little. Where beliefs are very strongly held, political leaders challenge them at their peril. Issues that seem minor can take on great salience if they engage people's emotions, or are used by political opponents, or take on public salience through the media (Levin, 2005). The time available for any issue is very limited due to many simultaneous competing pressures and events, while unexpected events can have very large consequences; consider the impact on school curricula of the collapse of the Soviet Union and its satellite regimes.

Educators may believe that education policy should be based on their knowledge and experience, but from a political perspective evidence and experience are not enough to drive decisions, and they may be among the less important factors. Even when policy decisions in education are not of public concern, they may be controversial and highly political within the education sector,

engendering the same political dynamics as above within the system. A subfield of education politics, micropolitics, examines these small-scale political interactions.

Issues: Scope of the Politics of Curriculum

At one time curriculum was almost universally seen as central to education, in that it was assumed that students would learn what was in the course of study. Decades of experience with educational change have made it evident that the situation is much more complex. There is a substantial debate as to how important formal education as a whole is in shaping student outcomes, with some arguing that socioeconomic status and other nonschool factors are by far the most important influences on outcomes while others believe that what happens in schools – which in any case involves much more than the official curriculum – can play an important role (European Commission, 2006; Organisation for Economic Co-operation and Development, 2007). Ideas about the importance of curriculum have also been changing over time as a result of growing knowledge – for example, increasing awareness of the importance of teachers' assessment practices (Black and William, 1998).

As governments have attempted to make large-scale changes in schooling, curriculum has become less an activity in its own right and more one element in a comprehensive approach to education change. In many jurisdictions with centralized curriculum, review and renewal processes have been altered to be more consistent with wider education programs. In other settings curriculum has been a prime vehicle for realizing wider change (e.g., Luke, 2004). The United States is an interesting case in that there is a national education reform strategy – No Child Left Behind – that is based largely on demanding curriculum standards, yet curriculum is controlled mainly at the district or even school level, creating some substantial disjointedness.

The politics of curriculum can be thought of as involving two kinds of discussions. The first concerns the overall shape of school curricula: what subjects will be included (or excluded), how much of each, and at what stage of students' education. Examples would include debates over whether literacy or mathematics are getting a sufficient share of the school day and school year, whether sex

education or religion should be part of the curriculum, or when students should first study a foreign language.

The second area of debate is over the content of particular subjects: what should be included in each subject and at each age or stage of education. These debates take place not only in the most obvious areas, such as what is taught in history or studied in literature, but also in areas that might be thought to be more objective, such as science or mathematics.

In addition to the content of specific subjects, schools are seen as the place where children will be inoculated against all social ills or taught all the virtues. Schools are expected to prevent bullying, obesity, and drug abuse while also eliminating racism and promoting equity in all its forms. In many cases these topics cut across the formal school curriculum, so who provides what instruction and when can be an important curriculum issue.

Some of these discussions are a matter of pragmatics. There simply are not enough hours and days in 12 years of schooling to accommodate all the areas people want children to develop. So debates take place about the relative importance of various areas of study in the competition for scarce school time. These discussions can be intense in part because they are often led by people with a strong personal interest in the decision.

A second set of arguments, often much more vitriolic, occur when curriculum debates occur due to disagreements on important value questions. Because schooling is seen as so fundamental to the development of our children, it can turn into a battleground for wider social disputes. Significant philosophical or value disagreements are expressed in many areas of curriculum, such as the content of history and literature curricula or the role of first languages for immigrant students. Any issue that is politically contentious can also turn into a curriculum dispute. These debates are sharp because they embody deeply held views not only about the nature of education but about essential values. Edelman's (1988) concept of condensation symbols, in which even relatively small instances become highly symbolic as they seem to embody, or condense, a range of beliefs and values in a particular case, applies often to curriculum politics.

Because everyone has gone to school, just about everyone feels knowledgeable about and has a personal response to educational issues. The same would not be true of other policy fields such as environment or energy. One important result of the universal experience of schooling is that adults, like children, tend to see curriculum as a collection of subjects and topics without necessarily requiring coherence or integration across the curriculum.

Another significant feature of curriculum politics is that in many cases questions of content cannot be separated from teaching practice. Many of the most heated issues in curriculum – for example, whole language or constructivist mathematics – are as much about teaching

methods as they are about curriculum content. These debates often spill over into decisions on curriculum structure and content.

Actors: Who Is Involved?

Curriculum politics involve a wide range of participants. Education governance typically involves some combination of national, regional, and local input. The division of powers and responsibilities across these levels is quite variable from one country to another. In most jurisdictions final authority over curriculum rests with national or regional governments or with agencies they create for this purpose. In many federal systems it is provinces or states that control curriculum. In a few situations curriculum authority is largely located within local districts or even individual schools.

The central role of governments brings into play a range of both political and bureaucratic elements. An individual in a key position, such as a powerful minister or political advisor, can either shape or hold up decisions, including details of curriculum content, if determined enough.

An important element of governance structure is the role of elected lay persons as against civil servants or experts. Countries vary significantly in how much authority lay people have in shaping education policy, ranging from ministers to local school authorities to school councils, or governing bodies involving parents and others. Each of these forms will bring different dynamics to curriculum politics. Depending on national governance arrangements, local schools, or districts have varying degrees of control – from almost none to quite substantial – over the formal curriculum.

Whatever the formal control system is, in almost all settings, schools have some influence if only through the choice they are able to make as to which courses and programs are actually provided and the amount of attention that is given in the day-to-day life of the schools to particular subjects or topics. High schools, for example, may be able to choose which optional courses to offer. Elementary teachers and schools typically make important decisions about curriculum emphasis, especially where – as is often the case – the formal curriculum has more elements to it than can be taken up in most classrooms. Schools or districts also often decide on the areas of the curriculum will be the subject of professional development.

The main education stakeholder groups – teachers, principals, senior administrators, and elected local authorities where they exist – are almost always involved in curriculum reviews and decisions. Students, however, typically have a very small role, if any, in curriculum governance. Subject matter experts from schools and universities usually play a central role in the curriculum formation and review process and may also be very involved in public debates. Indeed, a

central dilemma in curriculum formation is the balance between subject matter expertise and larger perspectives on the role of a given topic in the overall school program. Not surprisingly, those associated with each subject or topic will advance its importance – which, it must be pointed out, is often linked with their own employment prospects and importance.

Tertiary institutions may have a powerful influence on school curriculum, especially in secondary schools, through the setting of entrance requirements to their institutions. Since secondary schools often see themselves – and are widely seen by students and parents – as preparing students for further study, schools may find themselves quite constrained by requirements set by tertiary institutions. Views about school curriculum expressed by the tertiary sector may not be particularly well grounded in evidence but may rest on the beliefs of individuals holding key roles. As elsewhere, tertiary experts tend to believe that all students need more of their subject and at a higher level.

Because schools are widely seen as playing a central role in the socialization of children and young people, a wide range of interest groups may be involved in curriculum politics depending on the issue. For example, business groups may have strong views about aspects of curriculum. Various industries may try to promote subjects and programs that support their labor market needs. In the wider society, many groups want the curriculum to reflect particular issues and perspectives – for example, the desire to include the language, history, and literature of various minorities and indigenous peoples. As discussed in the next section, curriculum processes do not necessarily provide very much direct opportunity for input from various interests. As usual in political processes, those bodies that are better organized and financed or whose concerns are more deeply felt will tend to be much more active and may have disproportionate influence (Lindblom, 1980).

Curriculum can also be influenced significantly by other policies. Student assessment policies may shape curriculum decisions, and are often important drivers of what is actually taught (Mehrens, 1998). The effect of assessment will be positive if schools and teachers understand the standards and if the tests are aligned to appropriate curricula and teaching methods, which is not necessarily easy to achieve (European Commission, 2006). Where assessment is disconnected from curriculum, the consequences may be problematic.

Processes: How Are Curriculum Policy Decisions Made

Jurisdictions normally have well-developed formal processes for creating and revising curricula. Typically, these processes involve bringing together groups of experts and sector representatives to draft the elements of a new or revised curriculum. Teachers of the subject will often be

in the majority, with representation from tertiary subject experts as well. The processes are often organized and to some degree directed by government officials from ministries of education. Typically, a curriculum review or renewal process would involve examining the existing curriculum, gathering data as to the strengths and weaknesses of current arrangements, considering various ideas for changes, and trying to arrive at consensus on recommendations for the new curriculum. Altogether curriculum processes can be quite extensive depending on the level of detail in the work, sometimes taking several years to be completed. However, if substantial authority is vested at the level of the local school, curriculum development or renewal is likely to be much more informal and ad hoc, as no school will have the resources or capacity to undertake these larger processes.

Expertise in a subject or in teaching that subject does not necessarily equate to expertise in constructing a curriculum. By definition, experts know more and care more about their subject than will most teachers. An expert-dominated curriculum development process may result in a product that can be used effectively only by people with high levels of expertise, while the reality of schools everywhere is that many teachers, especially in elementary schools, will have only a limited background in a given subject. One danger in curriculum development then is the production of curricula that are not readily usable by ordinary teachers. Interest groups can and do promote ever-higher standards for their subject and then use those standards to argue for more time in the school schedule as well as for more teachers, higher qualifications, and more resources.

A further consequence of expert input is the clash between experts with competing views – for example, those who favor new pedagogic approaches versus advocates of traditional content organization and teaching. Educators themselves are often divided over such questions, while the clash of expertise may be a subject of media interest as well.

The expert-dominated approach to curriculum development is changing as experts no longer have the political legitimacy they once did. An increasingly better-educated public is demanding greater input in all areas of public policy. Governments are therefore moving away in many areas from reliance on experts to greater involvement of average citizens through various forms of consultation and opinion gathering as part of policy formation. In curriculum decisions review groups are now more likely to include parents or students or noneducators such as business representatives. As discussed in the next section, these changes in composition can have significant implications for the ways in which curriculum processes unfold because they will bring different and more diverse interests to the table.

Curriculum review groups do not do their work in a vacuum. Sometimes these processes proceed relatively smoothly, but they can also be highly contentious.

Where important disagreements exist about curriculum there may be intense lobbying by various interests as to who should be named to a working group. Although these processes are not often documented, some interesting accounts exist of the politics of curriculum review. More such accounts would be valuable in understanding these processes more fully.

Curriculum formation may also be linked to more public political processes. Public debates around curriculum issues can occur simultaneously with official processes – as, for example, when public interest groups are aware of and try to influence the outcomes of the official process by lobbying for particular changes. Or public debate and concern can be the springboard for an official process as the system tries to respond to public concerns, as is evident in debates about issues such as global warming. Or public debate can erupt following a formal curriculum change proposal as various interests mobilize to change a draft curriculum they do not support.

Influences: What Shapes Decisions?

One essential tension in curriculum decisions, as already noted, is between expert opinion and concern of key interests or of the general public. Political leadership will take account of expert opinion, but will inevitably take much more interest in public opinion and particularly the views of opinion leaders in key sectors or constituencies.

Even where the curriculum process is dominated by experts there may be substantial disagreement on what to do. Teachers may see curriculum issues quite differently from tertiary experts. The latter may focus on the need for high-level skills in their own area, whereas teachers may be more concerned with a curriculum that will work for students with widely varying skills and interests. Advocates may want more topics at higher levels, whereas classroom teachers may push for exactly the opposite. In some fields, teachers of the subject are themselves deeply divided on key curriculum and teaching questions. They also have varying opinions on issues such as the relevance of interdisciplinary studies as opposed to more traditional disciplinary boundaries, or on the value of project work.

Research has increasing importance in influencing education policy generally as a more educated population is more inclined to want evidence about public policy issues. In school practice too there is growing interest in the use of evidence to guide decisions about teaching and learning practices (e.g., Marzano, 2003). The interest in research does not, however, mean that there will be a direct link between research results and policy choices; the latter will always remain the result of political processes. The commitment to evidence is shown in that all sides in the political debate do attempt to bring evidence to bear whenever possible and do use it to legitimate and support their own positions, which may increase public

confusion. Yet empirical evidence can also shape and change public and professional views on important questions, as it has around tracking in secondary schools, or the impact of socioeconomic status on school outcomes. Despite significant advances in education research evidence, it remains the case that in many – probably most – areas of curriculum there is not enough knowledge to guide policy or practice sufficiently; and quite often, existing knowledge is not available in a form that speaks effectively to the real problems and issues of policy and practice.

Still, curriculum review will generally take account of emerging knowledge in the field under discussion precisely because the process has so much expert participation. More recently, in keeping with a general trend in education toward using student assessment data, curriculum review may include data on student outcomes, both for the curriculum overall and for particular groups of schools or students. However, regular use of student outcome data to guide education policy is still not standard practice everywhere. Nor do curriculum working groups necessarily pay careful attention to research as part of the work. Even expert processes are susceptible to a preference for interest bargaining instead of evidence.

Results

It has been at least 20 years since it became evident that there is a large gap between producing a curriculum and the experience of students in the classroom. A substantial body of research (e.g., Cohen *et al.*, 2003; Hubbard *et al.*, 2006) shows how far classroom practice can be from new curricula and how little impact a change in curriculum can have on teaching practice. The more significant the proposed change, the more likely it is to have limited adoption.

Various efforts have been made to produce curriculum that would have a greater impact on students' real experiences. These efforts range from so-called teacher-proof curricula, in which teaching practice is built into the curriculum at a high level of specificity, to attempts to connect new curricula with extensive professional development so as to change teachers' practices. However given the nature of schooling and teaching, with very large numbers of teachers of quite varied backgrounds in highly varied contexts and with considerable autonomy in their daily practice, central attempts to circumvent these limitations are unlikely to be successful. Working practices of teachers are shaped primarily by day-to-day realities of their workplace, their habits, and their views about what is practical (Organisation for Economic Co-operation and Development, 2005). Insofar as curriculum changes do not pay attention to these realities, they further limit their chances of having an effect.

Research has helped develop understanding of the factors that do shape implementation. Fullan's (2007) work has been particularly influential. However, these

considerations may still not be well integrated into formal curriculum processes, as it is still common to find accounts of some new program that was carefully developed only to fall far short in the implementation stage for entirely predictable reasons, such as violating teachers' norms around their practice. Issues of implementation and results are often marginal to the curriculum development process.

Conclusion

Although curriculum is a fundamental part of the framework of schooling, curriculum decisions and choices are shaped in large measure by other considerations – ideology, personal values, issues in the public domain, and interests. Curriculum decisions are often part of a much larger public debate that often extends beyond education to larger questions of public goods. These dynamics tend to be poorly understood by most educators, who tend to believe that education policy choices can and should be made on the basis of educational expertise. When processes are put in place without adequate regard for the real drivers of decisions, the likelihood of poor decisions – that is, decisions that fail to produce the intended results – increases.

Bibliography

- Ball, S. (1990). *Politics and Policy Making in Education*. London: Routledge.
- Black, P. and William, D. (1998). Inside the black box. *Phi Delta Kappan* 80, 39–148.
- Cohen, D. K., Raudenbush, S. W., and Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis* 25, 119–142.
- Crowson, R. L., Boyd, W. L., and Mawhinney, H. B. (1996). *The Politics of Education and the New Institutionalism: Reinventing the American School*. London: Falmer.
- Edelman, M. (1988). *Constructing the Political Spectacle*. Chicago, IL: University of Chicago Press.
- European Commission (2006). *Efficiency and Equity in European Education and Training Systems*. COM(2006) 481 Final, Brussels: European Commission.
- Fullan, M. (2007). *The New Meaning of Educational Change*, 4th edn. New York: Teachers College Press.
- Hubbard, L., Stein, M., and Mehan, H. (2006). *Reform as Learning*. New York: Routledge.
- Kingdon, J. (1994). *Agendas, Alternatives and Public Policies*, 2nd edn. New York: HarperCollins.
- Levin, B. (2005). *Governing Education*. Toronto, ON: University of Toronto Press.
- Lindblom, C. (1980). *The Policy-Making Process*, 2nd edn. Englewood Cliffs, NJ: Prentice-Hall.
- Luke, A. (2004). After the marketplace: Educational research, accountability and evidence-based policy. *Australian Educational Researcher* 9, 43–73.
- Manzer, R. (1994). *Public Schools and Political Ideas*. Toronto, ON: University of Toronto Press.
- Marzano, R. (2003). *What Works in Schools: Translating Research into Action*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Mehrens, W. (1998). Consequences of assessment: What is the evidence? *Education Policy Analysis Archives* 6. <http://epaa.asu.edu/epaa/v6n13.html> (accessed May 2009).
- Organisation for Economic Co-operation and Development, OECD (2005). *Attracting, Developing and Retaining Effective Teachers – Final Report: Teachers Matter*. Paris: OECD.
- Organisation for Economic Co-operation and Development, OECD (2007). *No More Failure*. Report of the thematic review on equity in learning. Paris: OECD.
- Young, J., Levin, B., and Wallin, D. (2007). *Understanding Canadian Schools*, 4th edn. Toronto, ON: Nelson Thomson.

Curriculum Planning and Systems Change

Z Deng, Nanyang Technological University, Singapore

© 2010 Elsevier Ltd. All rights reserved.

Systems theory, as conceptualized by theorists after the 1950s, views the world in terms of a unified whole and self-organizing systems, with an emphasis on the interdependent nature of the relationships that exist among all components of the unified whole (Boguslaw, 2000). From this perspective, significant change requires systems change that entails impacting changes across all constituent systems. Such a theory is useful in analyzing the nature, character, and complexity of curriculum planning for significant change.

This article first describes three basic domains of curriculum planning which can be viewed as three constituent systems. It then discusses their interrelationships and the implications for curriculum planning for significant change. This is followed by an examination of three common models of curriculum planning which are intended to bring about significant curricular change. The article concludes with addressing what is entailed in curriculum planning for significant change.

Three Levels of Curriculum Planning

Curriculum planning refers to the decision-making process concerning the substance of schooling, that is, the knowledge, skills, and dispositions that constitute the experience and outcome of schooling. Broadly construed, it operates across three basic domains of curriculum; institutional, programmatic, and classroom (Doyle, 1992a, 1992b). Each of these curriculum domains constitutes a system or subsystem that, in varying degrees, has an impact on what is taught and learned in school.

Institutional Curriculum Planning

The institutional curriculum embodies a conception or paradigm of what public schooling should be with respect to a society. Curriculum planning at this level is characterized by discourse on curriculum policy at the intersection between schooling, culture, and society. It invokes images, metaphors, or narratives to “typify” what could happen in a school or school system (Westbury, 2000). For instance, thinking School is used by the Ministry of Education in Singapore to convey a vision of schooling for the twenty-first century – a vision that defines the development of critical thinking and creativity as a central purpose of schooling. Institutional curriculum planning

frames what should go on in a school or school system in terms of broad goals and general approaches to education. It serves as a means of drawing attention to educational ideals and expectations shared within a society and putting forward the forms and procedures of schooling as responses to those ideals and expectations (Doyle, 1992b: 70). Such curriculum links what is taught in schools to the social and cultural systems beyond schooling, and is always under pressure for change. Since social and cultural contexts often change rapidly, school systems always use the institutional curriculum as a convenient instrument to communicate responsiveness to the outside communities (Doyle, 1992a: 487).

Institutional curriculum planning is always a national or regional political undertaking. In countries with centralized education systems like France, Singapore, Malaysia, and China, the legal responsibility for institutional curriculum planning is the province of the central government. National educational bodies (ministries, departments) play a substantial role in the planning process. In countries with decentralized systems such as Canada, UK, and Australia, state or provincial governments are constitutionally responsible for making institutional curriculum decisions. Regional educational agencies are instrumental in the planning process. Regardless of it being in centralized or in decentralized systems, curriculum planning at the institutional level, more often than not, involves soliciting the opinions and suggestions from various representative groups – including policy advisory bodies, employment agencies, educational specialists, heads of schools, and various civic and special interest groups.

Programmatic Curriculum Planning

Programmatic curriculum planning is at the intermediate levels between institutional curriculum and classroom curriculum planning, with a focus on curriculum writing in the form of curriculum documents and materials (Doyle, 1992a). It translates the expectations and ideals embedded in the institutional curriculum into operational frameworks for schools, thereby bridging the gap between the abstract institutional curriculum and the (enacted) classroom curriculum (Westbury, 2000). In Singapore the notion of thinking schools becomes the introduction of thinking programs to be implemented in primary and secondary schools. The programmatic curriculum is characterized by an array of school subjects, programs, and courses of study provided to a school or a system of schools.

For those school subjects, programs, and courses of study, the programmatic curriculum also spells out instructional guidance in terms of content standards, instructional frameworks, criteria for textbook approval and adoption, and assessment criteria.

Programmatic curriculum planning refers to the formal process of constructing the curriculum which involves “framing a set of arguments that rationalize the selection and arrangement of content [knowledge, skills, and dispositions] and the transformation of that content into school subjects” (Doyle, 1992b: 71). The planning process, Tyler (1949) believes, entails addressing the questions concerning purposes, content, organization, and evaluation in a sequential fashion:

1. What educational purposes should the school seek to attain?
2. What educational experiences can be provided that are likely to attain these purposes?
3. How can these educational experiences be effectively organized?
4. How can we determine whether these purposes are being attained?

Tyler’s model prescribes what steps one needs to follow in planning a curriculum. However, it does not describe what curriculum developers actually do when they are engaged in curriculum planning (Westbury, 2008). In practice, programmatic curriculum planning is a highly sophisticated undertaking. It is a collaborative and cooperative endeavor that always involves committees made up of representatives from governments, educational agencies, schools, universities, business, industry, and the community at large. It occurs within “webs of societal and cultural ideologies and symbols, politics and organized interest groups, organizational and administrative structures and processes, and local understandings, beliefs and practices” (Westbury, 2008). As in institutional curriculum planning, national and regional educational agencies play a key role in programmatic curriculum planning.

Classroom Curriculum Planning

The classroom curriculum, also called curriculum as event or the enacted curriculum, is characterized by a cluster of events jointly developed by a teacher and a group of students within a particular classroom (Doyle, 1992a, 1992b). It is an evolving construction resulting from the interaction of the teacher and students. Curriculum planning at this level involves transforming the institutional and programmatic curriculum embodied in curriculum documents and materials into educative experiences for students. It requires further elaboration of the programmatic curriculum, making it connect with the experience, interests, and the capacities of students in a particular classroom (Westbury, 2000). Curriculum

planning can be the effort of an individual teacher or a team of teachers responsible for identifiable students, deciding alone or with students what shall occur in a specific educational setting. Such curriculum is shaped in a powerful way by a range of local factors, including teachers’ own classroom perspectives, students’ interests and experiences, school principals’ requirements, and parents’ expectations (Doyle, 1992b).

The three levels of curriculum planning together imply a hierarchy of decision-making activities from the institutional to the classroom arena. This, however, does not mean that curriculum planning must begin with a national ministry or state department of education in formulating curriculum policy, proceed to writing curriculum documents and related materials, and eventually end up with school teachers in implementing the documents and materials in classroom. In practice, curriculum planning could begin anywhere. It could, for example, start with an individual teacher or a group of teachers and a class of students. Teachers can participate in what is called school-based curriculum development (SBCD) in which they articulate their own visions and goals of teaching, develop their own curriculum materials in the light of their visions and goals, and put the materials into practice with students. However, this does not mean that teachers could completely ignore the existence of curriculum goals and instructional guidelines developed at the national or state level – a point that will be further addressed in sections that follow.

The Interdependence of Three Levels of Curriculum Planning

The three levels of curriculum planning are interrelated and interdependent; each does not function in isolation of others. Curriculum planning at the institutional level provides a departure point and a frame of reference for curriculum planning at the programmatic level. It serves to clarify the normative, ideological bases for curriculum decision makings, drawing attention to the public character of curriculum that goes beyond the private interests and concerns of individuals. By articulating the aims, purposes, ideals, and expectations for public schooling, institutional curriculum planning points to a desirable direction for programmatic curriculum planning. Without attending to those aims, purposes, ideals, and expectations, programmatic curriculum planning could lose sight of the responsibilities that schooling as an institution needs to bear in contributing to the common good of a society.

Classroom curriculum planning, although remote from institutional curriculum planning, has an indirect yet significant link with the institutional curriculum. The construction of the classroom curriculum often involves a teacher’s use of curriculum materials (e.g., textbooks

and teacher guides) adopted by a school or school system – materials that are normally developed with reference to what is desired and valued in the institutional curriculum. In planning instructional activities, a teacher interprets and transforms the materials in the light of his or her knowledge and beliefs about the purposes of schooling, about his or her students, about pedagogy, and about the school and classroom contexts (Wilson *et al.*, 1987). To a certain extent, the classroom curriculum thus reflects the teacher's interpretation of "what is desired by unseen, remote decision makers" (Goodlad *et al.*, 1979: 21). Furthermore, when the classroom curriculum becomes visible in various ways (such as public examinations, classroom surveys, and parent meetings) to administrators, parents, and others interested in education, questions concerning the interplay between schooling and society could be inevitable: for example, what are the purposes of schooling? How well does the curriculum prepare students to meet the current and future challenges of the social and political order? Social and political forces are brought to bear at the school or classroom level. The classroom curriculum thus is connected indirectly back to the institutional curriculum which, in one way or another, "serves as a normative framework for defining and managing the work of teachers" (Doyle, 1992a: 487).

Apparently, institutional curriculum planning seeks to affect classroom curriculum planning through the programmatic curriculum – curriculum that provides the medium in and through which the institutional curriculum operates. By translating the institutional curriculum into an array of school subjects, programs, and courses of study, programmatic curriculum planning creates the organizational and operational structure for classroom curriculum planning. School subjects, programs, or courses of study constitute the locus of classroom teaching; they frame the character of the classroom curriculum and the ways in which the curriculum might be seen within the school community (Grossman and Stodolsky, 1995). The ways of teaching a traditional school subject (e.g., history) and an integrated subject (e.g., social studies) in a classroom could be substantially different, so are the criteria of judging these two subjects. Furthermore, by developing content standards, instructional frameworks, syllabi, assessment criteria, and the like, for those school subjects, programs, and courses of study, programmatic curriculum planning furnishes the instruments that serve to regulate and control classroom curriculum planning, guiding what is to be taught and learned in the classroom (Doyle, 1992a). These instruments could "steer" classroom curriculum planning in the direction set out in the institutional curriculum (Cohen and Spillane, 1992). They could also be resources for teachers to learn the ideals and reform visions embedded in the institutional curriculum (Cohen and Hill, 2001).

The effect of programmatic curriculum planning on classroom curriculum planning, however, is complex and

in no sense straightforward. It has to do with the governance structure of curriculum planning. As already mentioned, in centralized education systems programmatic curriculum planning is carried out mostly by government educational agencies (e.g., national ministries of education). The programmatic curriculum, which more often than not is linked with high-stakes public examinations, tends to have high power and authority over the classroom curriculum. Classroom teachers are supposed to follow closely the standards, frameworks, and guidelines laid out in the programmatic curriculum; they have relatively less latitude in deciding what to teach and how to teach it (Cohen and Spillane, 1992). Programmatic curriculum planning thus could significantly influence classroom curriculum planning. However, in a decentralized education system where high-stakes public examinations do not exist and local or regional authorities make most programmatic curriculum decisions, classroom teachers have considerably high latitude in shaping the content and process of their curriculum. Programmatic curriculum planning tends to have much less impact on the classroom curriculum. The education system in the United States typifies this governance structure of curriculum planning in which state developed or mandated curriculum frameworks and materials do not have formal authority over classroom practice (Cohen and Spillane, 1992).

Furthermore, programmatic curriculum planning, whether in a centralized or a decentralized system, depends, for its effect, on those implementing the programmatic curriculum in school and classroom. While school leaders could make decisions concerning the adoption of a particular curriculum framework and related materials as the programmatic curriculum for the school, classroom teachers are the ones ultimately responsible for carrying out the programmatic curriculum in their classrooms. Programmatic curriculum planning affects classroom curriculum planning only if teachers understand and employ the adopted framework and materials (Synder *et al.*, 1992). Teachers are not "conduits" for the use of a curriculum framework and related materials; as mentioned earlier, they interpret and transform the framework and materials in the light of their experience, beliefs, and practice. This could further compound the impact of programmatic curriculum planning on the classroom curriculum. Teachers' entrenched beliefs, experience, and practice could override the educational ideals and innovations embedded in the programmatic curriculum.

Curriculum Planning for Systems Change

As highlighted in the above discussion, the three basic domains of curriculum are interrelated and interdependent, together constituting an organic whole. Significant curricular change thus cannot be achieved by just

tweaking one or two domains in isolation; it entails systems change that requires impacting change across all three domains of curriculum. To achieve this, all three levels of curriculum planning are necessary and need to work together in a way that ensures sustainable curricular change at the classroom level. A defensible model of curriculum planning for systems change, therefore, needs to take account of all three levels of planning and their relationships. None of the levels can be undermined without undermining a vital factor in curriculum planning and development. With this in mind, three common models of curriculum planning – namely top-down, bottom-up, and combination – that intend to bring about significant curricular change, or by implication, systems change – will be considered.

Top-Down Model

The top-down model has been widely used in countries with a centralized education system. For many centralized countries, curriculum reform has been for a long time part of national plans and development strategies. Usually, the central government initiates curricular change by putting forward new curricular visions and goals. These visions and goals are then translated into programmatic or curricular frameworks that specify course structure, content standards, pedagogy, and assessment. Using these frameworks as a point of reference, the national education body would implement a series of initiatives such as textbook revision, assessment modification, teacher preparation, and professional development restructuring. These initiatives are expected to steer teaching and learning in classroom toward the reform visions and goals, resulting in significant change in the classroom curriculum. Also termed the framework approach (see Skilbeck, 1994), the top-down model was adopted by many traditional decentralized systems over the last two decades. For example, in the 1990s following the lead of the federal government, virtually all American states developed their statewide curriculum frameworks. In Australia, Canada, and other federated countries, there was a growing collaboration between state and federal authorities concerning the construction of an overall framework for the school curriculum based upon new curricular visions and goals (Skilbeck, 1994). These curriculum frameworks were believed to play a crucial role in steering the classroom curriculum in the reform direction.

Such a top-down, centrally driven, and command-oriented model places an extremely high emphasis on institutional and programmatic curriculum planning for change. Curriculum planning virtually becomes a national or state enterprise under the authority of the central educational agency. This model, however, rarely works well at the school or classroom level. New curricular visions and frameworks often do not get implemented,

or are not implemented in the way they were intended due to various reasons. There could be problems in communication between the central agency and schools. Teachers might resist a new vision and a framework in the belief that they know better than those working at the central agencies (Chapman and Mählck, 1997). Furthermore, a reform vision and a curriculum framework can be read and understood by teachers in very different ways according to their varying beliefs, values, commitments, and experiences, thereby leading to different ways of implementing that vision and framework (Cohen and Ball, 1990). The importance of curriculum planning at the classroom level, and particularly the complex relationship between programmatic and classroom curriculum planning, is largely overlooked or ignored in the top-down model. In addition, the means and support for implementing curricular change at the classroom level is always inadequate in top-down approaches to curriculum planning for systems change.

Top-down, centralized planning for curricular change can be seen as an effective instrument for steering curriculum discourse at the institutional and programmatic levels; the discourse, however, often gets lost in school and classroom (Westbury, 2008). When coupled with high-stakes tests and accountability measures, this model could have a detrimental impact on the classroom curriculum. The current implementation of the No Child Left Behind (NCLB) Act in the United States, Apple (2008) argues, leads to an impoverished curriculum that is defined and driven by the measure of student ability and competency which is extremely narrowly defined, and that creates greater inequalities between minority and majority children in educational attainment.

Bottom-Up Model

In contrast to the top-down model, the bottom-up model holds that significant curricular change comes from inside out rather than the outside in or from the top down. Central educational agencies (e.g., ministries of education) can really do little to influence what happens in school and classroom. To bring about change at the classroom level, curriculum planning must be grounded in the deliberative knowing and practical action of school practitioners. By participating in bottom-up approaches to curriculum planning like SBCD and action research, teachers can become the central players in the curriculum reform endeavor (MacDonald, 2003). Furthermore, a school could become a learning organization by creating conditions for school leaders and teachers to continually develop new ideas and improve their quality of thinking and capacity for reflection. They can work with students and parents to form new curricular visions, translate their visions into operational frameworks, and decide how best to bring about change in the classroom curriculum

(Fullan, 1993). The assumption is that successful curricular change could be relatively easy to achieve in a local school or a cluster of local schools, and a significant number of such local changes can, over time, in an innovation diffusion process build from the bottom up into a major change in the overall education system (Farrell, 1997).

This bottom-up, decentralized model, while appearing promising, is fraught with problems and pitfalls. When the entire enterprise of curriculum planning is reduced to nothing but school and local business, the public character of the school curriculum would be seriously undermined, together with the public governance of curriculum planning. Personal and local interests and powers would prevail over those broader concerns for the society and the welfare of mankind. Furthermore, when the full task of curriculum planning is lobbed onto schools, at all three levels, with all its complexity, it creates tremendous demands on classroom teachers. Most teachers are neither prepared for, nor do they have the experience and expertise necessary to undertake, such a task. With respect to the implementation of SBCD, MacDonald (2003) observes that “what occurred in many Australian states and in the USA were less demanding, poorly resourced and loosely assessed curricula” (p. 141). The above innovation diffusion assumption is indeed questionable. When planning for curricular change is entirely left to the school, the possible consequences include: (1) not much change at school or classroom level; (2) changes in a problematic direction; and (3) changes that do not spread out or endure (Fullan, 2003). The basic problem of the bottom-up model lies in the failure of recognizing the need for institutional and programmatic curriculum planning that necessarily go beyond the realm of school localities.

Combination

There has been an increasing interest in strengthening the relationship between curriculum planning at the national or state level and at the school and classroom level. This is based on the realization that top-down guidance and bottom-up initiatives need each other. While central educational agencies are incapable of dictating or mandating change at the school and classroom level (Fullan, 1993), they still have a fundamental role to play in designing reform and translating reform into curricular frameworks, documents, and materials that could support and enable curricular change at the school or classroom level. While classroom teachers need to have sufficient freedom and autonomy in planning and carrying out curricular change, they need guidance and support provided by schools and external agencies as well.

A successful combination model needs to strike a balance between top-down and bottom-up approaches. It needs to acknowledge, on the one hand, the key role of classroom teachers as curricular change agents and, on the

other hand, the need for institutional and programmatic curriculum planning in guiding, supporting, and enabling curricular change at the classroom level. Three conditions are critical with respect to institutional and programmatic curriculum planning by external agencies. First, there needs to be coherence among new curricular visions, curriculum frameworks, assessments, and teacher professional development. Inconsistency would lead to different and divergent interpretations of curricular change. Second, curriculum frameworks and materials need to be developed in a way that supports teachers’ planning for significant change (Cohen and Hill, 2001). Curriculum frameworks and materials can be an effective agent that enables classroom teachers to plan for significant change in a particular classroom context, if they were designed to “place teachers in the center of curriculum construction and make teachers’ learning central to efforts to improved instruction” (Ball and Cohen, 1996: 7). Third, teachers need to have substantial opportunities for professional learning that are grounded in practice and in specific curricular changes (Cohen and Hill, 2001). In other words, substantial professional learning is a key element in curriculum planning for systems change.

Conclusion

Significant curricular change requires systems change that entails impacting change across the institutional, programmatic, and classroom curriculum. Curriculum planning for systems change is a highly complex and challenging endeavor. It entails a coordination of institutional, programmatic, and classroom curriculum planning, the absence of any of which would not result in significant change. It requires a collaboration of a multitude of participants and representatives from the government, educational agencies, universities, business, schools, and communities at large. It needs to provide teachers with sufficient support and resources. Furthermore, it is important to bear in mind that curriculum reform is part of a larger effort to reform the school system. Curriculum planning thus needs to be related to larger issues of school change and improvement, significantly influenced by other policies and factors.

See also: Curriculum and Teacher Change; Curriculum Governance and Planning; Curriculum Reform.

Bibliography

- Apple, M. (2008). Curriculum planning: Content, form and the politics of accountability. In Connelly, F. M., He, M. F., and Phillion, J. (eds.) *Sage Handbook of Curriculum and Instruction*. Thousand Oaks, CA: Sage.
- Ball, D. and Cohen, D. K. (1996). Reform by the book: What is – or might be – the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher* 25, 6–8.

- Boguslaw, R. (2000). Systems theory. In Borgatta, E. F. and Montgomery, R. J. V. (eds.) *Encyclopedia of Sociology*, 2nd edn, vol. 5, pp 3102–3106. New York: Macmillan.
- Chapman, D. W. and Mählick, L. O. (1997). Changing what happens in schools: Central level initiatives to improve school practice. In Chapman, D. W., Mählick, L. O., and Smulders, A. E. M. (eds.) *From Planning to Action: Government Initiatives for Improving School-Level Practice*, pp 1–31. Oxford: Pergamon.
- Cohen, D. K. and Ball, B. L. (1990). Policy and practice: A commentary. *Educational Evaluation and Policy Analysis* **12**, 331–338.
- Cohen, D. K. and Hill, H. C. (2001). *Learning Policy: When State Education Reform Works*. New Haven, CT: Yale University Press.
- Cohen, D. K. and Spillane, J. P. (1992). Policy and practice: The relations between governance and instruction. In Grant, C. (ed.) *Review of Research in Education*, vol. 18, pp 3–49. Washington, DC: American Educational Research Association.
- Doyle, W. (1992a). Curriculum and pedagogy. In Jackson, P. W. (ed.) *Handbook of Research on Curriculum*, pp 486–516. New York: Macmillan.
- Doyle, W. (1992b). Constructing curriculum in the classroom. In Oser, F. K., Dick, A., and Patry, J. (eds.) *Effective and Responsible Teaching: The New Syntheses*, pp 66–79. San Francisco, CA: Jossey-Bass.
- Goodlad, J. I., et al. (1979). *Curriculum Inquiry: The Study of Curriculum Practice*. New York: McGraw-Hill.
- Grossman, P. L. and Stodolsky, S. S. (1995). Content as context: The role of school subjects in secondary school teaching. *Educational Researcher* **24**, 5–11.
- Farrell, J. P. (1997). A retrospective on educational planning in comparative education. *Comparative Education Review* **41**, 277–313.
- Fullan, M. (1993). *Change Forces: Probing the Depths of Educational Reform*. London: Falmer.
- Fullan, M. (2003). *Change Forces with a Vengeance*. New York: RoutledgeFalmer.
- MacDonald, D. (2003). Curriculum change and the post-modern world: Is the school curriculum-reform movement an anachronism? *Journal of Curriculum Studies* **35**, 139–148.
- Skilbeck, M. (1994). Curriculum renewal. In Husén, T. and Postlethwaite, T. N. (eds.) *The International Encyclopedia of Education*, pp 1338–1343. Oxford: Pergamon.
- Snyder, J., Bolin, F., and Zumwalt, K. (1992). Curriculum implementation. In Jackson, P. W. (ed.) *Handbook of research on curriculum*, pp 402–435. New York: Macmillan.
- Tyler, R. (1949). *Basic Principles of Curriculum and Instruction*. Chicago, IL: University of Chicago Press.
- Westbury, I. (2000). Teaching as a reflective practice: What might didaktik teach curriculum. In Westbury, I., Hopmann, S., and Riquarts, K. (eds.) *Teaching as a Reflective Practice: The German didaktik Tradition*, pp 15–39. Mahwah, NJ: Erlbaum.
- Westbury, I. (2008). The making of formal curricula: Why do states make curricula, and how? In Connelly, F. M., He, M. F., and Phillion, J. (eds.) *Handbook of Curriculum and Instruction*. Thousand Oaks, CA: Sage.
- Wilson, S. M., Shulman, L. S., and Richert, A. E. (1987). “150 different ways” of knowing: Representations of knowledge in teaching. In Calderhead, J. (ed.) *Exploring Teachers’ Thinking*, pp 104–124. London: Cassell.

Further Reading

- Beane, J., Toepfer, C. F., and Alessi, S. J. (1986). *Curriculum Planning and Development*. Boston, MA: Allyn and Bacon.
- Eisner, E. W. (2000). Those who ignore the past. . . : 12 “easy” lessons for the next millennium. *Journal of Curriculum Studies* **32**, 343–357.
- Fullan, M. (2008). Curriculum implementation and sustainability. In Connelly, F. M., He, M. F., and Phillion, J. (eds.) *Sage Handbook of Curriculum and Instruction*. Thousand Oaks, CA: Sage.
- Gay, G. (1991). Curriculum development. In Lewy, A. (ed.) *The International Encyclopedia of Curriculum*, pp 293–302. Oxford: Pergamon.
- Levin, B. (2007). Curriculum policy and the politics of what should be learned in schools. In Connelly, F. M., He, M. F., and Phillion, J. (eds.) *Sage Handbook of Curriculum and Instruction*, pp 7–24. Thousand Oaks, CA: Sage.
- Posner, G. J. (1998). Models of curriculum planning. In Beyer, L. E. and Apple, M. W. (eds.) *The Curriculum: Problems, Politics, and Possibilities*, 2nd edn., pp 79–100. New York: State University of New York Press.
- Reid, W. A. (1992). *The Pursuit of Curriculum: Schooling and the Public Interest*. Norwood, NJ: Ablex.
- Schwab, J. J. (1973). The practical: Translation into curriculum. *School Review* **81**, 501–522.
- Westbury, I. (2003). Curriculum, school: Overview. In Guthrie, J. W. (ed.) *The Encyclopedia of Education*, 2nd edn, pp 529–535. New York: Macmillan.

Relevant Websites

- <http://www.ascd.org> – ASCD Learn. Teach. Lead.
- <http://www.curriculum.org> – Curriculum Services Canada.

Curriculum, Digital Resources and Delivery

Y Zhao, G Zhang and C Lai, Michigan State University, East Lansing, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

Over the past decade, digital technology has made significant inroads into education, largely due to the rapid spread of the Internet and its associated applications. Not only has digital technology become more accessible to students and teachers, but it also has expanded the scope of what teachers and students can do with digital technology. This article presents an overview of the interactions between digital technology and curriculum, with a focus on how digital technology affects the delivery of curriculum.

Recent Developments in Digital Technology

Expanded Access

Digital technology has become much more accessible in schools. In 2005, almost all public schools in the United States were connected to the Internet, among which 97% had broadband connections. On average, in 2005, every 3.8 students had access to a computer connected to the Internet in public schools, which was an increase from the 12 students per computer in 1998. In addition, 45% of schools nationwide enabled video streaming, 28 US states had digital whiteboards in over 35% of their public schools, 15% of all public school instructional rooms had wireless Internet connections, 19% of public schools provided hand-held computers to students or teachers for instructional purposes, and 10% of public schools lent laptop computers to students.

While the US may be ahead of many nations in equipping its schools with digital technology, it is certainly not the only one. Many developed countries, for example, Singapore, South Korea, and the United Kingdom, have all generously put technology into schools. Other nations, including many developing countries, have also been investing in digital technology. China, for instance, has been implementing a scheme to equip its rural schools with computers and connect them to the Internet. The MIT's One Laptop per Child initiative is developing laptops that would cost 100 dollars and has plans to distribute them to developing countries. It is expected that 40 million units will be shipped to some 14 countries in the next few years.

Expanded Capacity

Along with expanded availability of digital technology came expanded capacity. In the 1950s, the computer was still what its name suggested, a machine that did mathematical

calculations. But today, the computer is a television, a telephone, a telegram machine, a piano, a typewriter, a music box, a library, a community center, and much more. Thanks to digitization, the computer has integrated many media into one – enabling the mixing of contents that were once not mixable. Rapid development in storage technology has significantly lowered the cost of information distribution. High-speed computer networks, low-cost storage, increasingly sophisticated search engines, and digital format together enable the accumulation and sharing of unprecedented amounts of information among institutions and individuals. Today, virtually all organizations and many individuals in the world maintain a website in order to share and broadcast information. Google's attempt to digitize and make available millions of books to worldwide users is just one example of global information gathering and sharing.

Personalized Publishing

Perhaps the most fundamental change in information and communication technology is the capability it affords individuals to publish and broadcast their ideas to a broader audience. Low-cost digital tools and wider access to the Internet have made it possible for anyone with a connection to publicize his or her ideas, images, or any other personal information to a worldwide audience. Publishing and broadcasting are no longer controlled by corporations or organizations. They are now within the reach of the individual. Podcasting, Weblogs, and YouTube.com are just a few examples of personalized publishing and broadcasting.

Characteristics of Digital Technology and Impact on Curriculum

The expanded access provides the infrastructure and incentives to deliver curriculum and other educational resources to schools, while the expanded capacity of digital technology brings new possibilities for reshaping the format and nature of curriculum. Digital technology is different from traditional media in many ways, but three – multimedia, interactivity, and socialization – are among the most relevant to curriculum development and planning.

Multimedia

The defining quality of digital technology is that all information is in digital format. In other words, all forms

of information, be it video, audio, text, photographs, graphics, or mathematical symbols, can be encoded in zeros and ones, making it possible to mix different forms of information. While it is possible to combine these formats with traditional media technology, the effort and associated cost are considerably higher, if not prohibitive for most people, than with digital technology. Hence, one of most fundamental characteristics of curriculum delivered through digital technology is multimedia.

Multimedia presentation of curriculum affects how the content of curriculum is delivered. For example, CD-ROM books include multiple multimedia features, such as animation, illustrations, music with text, and online assistance features. Such CD-ROMs, digital library collections, and other computer-based digital sources are providing students and teachers with access to the world's knowledge in a form that can be readily searched and downloaded.

Multimedia presentation of curriculum not only has great influence on how the curriculum is delivered, but also redefines the content of curriculum itself. For example, the definition of literacy is expanded from traditional print-based text to computer-communication-based reading and writing. Text in the era of multiliteracies is intrinsically multimodal as nonverbal language (such as the visual, the audio, and the gesture) becomes more and more significant in the media. Being able to search for online information and publish writing online are considered the new skills required in this digital age (Leu, 2000; Reinking, 1998). As a result, the adoption of computer-based technology in literacy redefines literacy and transforms the literacy curriculum.

Multimedia curriculum resources are recognized to hold great potential for improving learning. Well-designed multimedia instruction materials have been shown to hold the potential to enable students to learn more deeply than from traditional single-mode communication with written words only (Baddeley, 1998; Mayer, 2001). Curriculum delivered in multimedia format can foster deeper learning in students than traditional print-only curriculum (Cognition and Technology Group at Vanderbilt, 1990).

Interactivity

Digital technology also affords increased interactivity. It is necessary to differentiate interactivity from two closely related concepts: interaction and transaction. While both interaction and interactivity include learner involvement, interaction is more often interpersonal and occurs within a social context and interactivity is often between a human and a computer. Compared with interactivity, transaction does not include any learner-initiated engagements. For example, simply reading the text would be considered learner-content transaction, while learner-content interaction would be something like creating new materials, seeking out additional Web materials and

posting content, or reorganizing the materials into a new presentation.

There are numerous instances of interactivity in digital-delivered curriculum, which is significantly less in traditional print-based curriculum. Hypertext or hypermedia reading, for example, involves learners to interact with the content. Learners are given the choices to access the embedded links in the text for more information or learning support. By clicking or not clicking these links, learners are making choices of how much to read, what to read, and in which sequence to read. Another example is e-textbooks, which enable students to learn at their own pace. Interacting with an animation-based simulation has become an increasingly popular practice for teaching complex scientific or mathematical concepts. Educational games require even more intense interaction between the learner and the content. Online tools, such as databases and search engines, also enable learners to seek information through user-generated interaction.

Learner-content interactivity is considered to be an integral component of digital-delivered curriculum. Students examine, consider, and process the learning material during the educational experience. It is through interacting with content that learners actively reconstruct knowledge and deepen their understanding. A highly interactive learning environment would be one in which learners have frequent opportunities to make a wide variety of significant choices. By doing so, learners feel a higher degree of involvement and engagement and as a result, become highly motivated.

Such learner-content interactivity reflects the trend of student-centered learning enabled by digital-delivered curriculum. Traditional curriculum is prescribed, teacher-centered, pre-packaged, and pre-sequenced. Interactivity of digital technology makes it possible for students to personalize their own learning and redesign their curriculum. The immediate access to online information makes it possible for students to seek knowledge of the areas of their interest instead of being fed with prescribed content. Technologies make curricula no longer a fixed learning plan to be implemented by the teacher and followed by the students. Instead, learners become the designers or co-designers of their own curriculum. In addition, it provides more opportunities for students to initiate their learning and bring their knowledge into learning. Students using digital-delivered curriculum are found to have the same or higher degrees of engagement compared to those in traditional instruction.

Socialization

While multimedia and interactivity affect how the learners approach and process content, technology, especially the vast computer networks and other communication devices, also expands the possibility for students to socialize. Thus, it

adds another dimension to digital-delivered curriculum. Researchers suggested that computer-mediated communication (CMC) would develop a potentially new type of learning community that would provide space for collective thinking and access to peers for socializing and communication. The Internet – with the choice of synchronous or asynchronous communication, communication connection at any time and anywhere, and embedded video and audio streaming – offers a wide range of possibilities for actively engaging students. Many different CMC applications have been successfully used to promote collaborative and cooperative learning.

CMC technology supports online social interaction on two levels. First, it supports instructor–learner(s) and learner–learner interaction. For example, course-management systems (CMS), such as Blackboard and Angel, enable instructors and students to easily communicate with each other. Second, it offers a variety of opportunities for learners to seek resources and interact with people outside the classroom. For example, communicating with native speakers from a distance has become an increasingly common practice in foreign-language teaching.

The socialization capacity of digital communication technologies has significant impact on curriculum because it has significant potential for student collaboration, cooperation, and co-construction of knowledge. Theories highlight the socially distributed nature of learning (Lave and Wenger, 1991; Pea, 1993; Vygotsky, 1978). It is now generally accepted that learners develop understanding through interacting with others. Research has confirmed the positive effects of collaborative learning on student achievement. Digital-delivered curriculum taps the power of distributed cognition by integrating a wide variety of tools to support online collaborative learning.

It is apparent that digital technology, as a new medium for curriculum delivery, has the potential to reshape curriculum in a number of significant ways:

- to enrich forms and sources of curriculum materials by mixing different media and bringing in content from different sources;
- to increase learners' control (thus decrease the developers' and teachers' control) of the curriculum;
- to make available an unprecedented amount of content to use; and
- to demand curriculum to be more fluid and flexible.

Uses of Digital Technology in Curriculum Planning, Implementation, and Delivery

The expanded access to and capacity of digital technology have resulted in an explosive growth in digital curriculum content, but not all digital content has fully taken

advantage of the capacities of the new medium, actually nor need they. Consequently, the nature of digital curriculum content varies on a continuum of complexity and comprehensiveness, ranging from simple electronic books (print books simply digitized without taking advantage of the new medium) to fully interactive comprehensive curriculum such as online courses. Correspondingly, the uses of digital technology in the delivery, implementation, and planning of curriculum vary.

Digital Curriculum Resources for Teachers

Online curriculum standards

Curriculum standards are one of the core components of curriculum planning and development since they define goals and expectations of curriculum. Making curriculum standards accessible to teachers and curriculum developers so that they can better incorporate these standards in their teaching has become popular. Today, virtually all curricular frameworks and content standards in the US and many other countries are available online. A search using Google or other search engines can point teachers and curriculum developers to these documents.

Curriculum planning tools

Digital tools have been developed to assist teachers plan their lessons. Teachers can find numerous lesson templates that help them create a new lesson plan or learning activity. These templates usually generate a step-by-step process and include major components in lesson planning, such as description of the activities, goals, and assessments. These tools can also be adjusted to better meet each individual teacher's need. For example, Florida Department of Education maintains a website for curriculum planning, which claims to provide a step-by-step instruction on lesson plan and learning activity design.

Lesson plans

One of the earliest efforts in the use of digital technology in curriculum planning and delivery is to make collections of exemplary lesson plans online so that teachers can use them as references. This is particularly useful for promoting new ways of teaching. There are numerous websites devoted to publishing lesson plans. For example, Scholastic offers K-12 lesson plans in reading, language arts, social studies, math, and science. From the Scholastic site, teachers can also find helpful information on thematic lessons to help them make lesson plans for more than one class session. Some other sites focus more on specific subjects. Teachers are also invited to share their own teaching experience and publish their lesson plans on the Scholastic site.

Curriculum resources

A variety of resources have been made available to teachers so that they can incorporate it in their delivery of

the curriculum. Such resources include audio and visual material, pictures, interactive material, global subjects, as well as local events. These resources have been provided by many different organizations and are often aligned with certain school curriculum. Some of them were created by education authorities and libraries, for example, the InforOhio website provides curriculum resources for all K-12 subjects. Some others were created by professional organizations or universities, for example, the MathForum at Drexel University has been providing curriculum resources for math teachers since 1992. And then there are those that were created by organizations and institutions outside the sphere of education, for example, the British Museum presents a great site of learning world culture.

Assessment tools

Assessment has been greatly affected by the introduction of digital technology into curriculum. The changes may be as simple as teachers using online spreadsheet to document students' attendances or as complicated as data-driven decision making (DDDM). DDDM's primary goal is to use students' assessment data and relevant background information to inform decisions related to planning and implementing curriculum at the district, school, classroom, and individual student levels. Systematically collected data are analyzed and used to measure student progress, measure program effectiveness, guide curriculum development, allocate resources, and promote accountability. By doing so, assessment becomes an integrated component of curriculum planning and development. Assessment is not the end of instruction, but a significant analytical tool for designing/redesigning curriculum. While DDDM is still a relatively new concept in education, there are some successful stories in adopting such performance systems into schools.

Online discussion forums

Online forums can help create learning communities that provide professional development for teachers to interact with both pre- and in-service teachers. Online discussion forums have been recognized as a potentially powerful tool for creating an online learning community for teachers. Online discussion can be synchronous or asynchronous. Teacher forums in Dave's ESL Café are examples of asynchronous online discussion forums for teachers. This forum is open to English as second language (ESL) and English as foreign language (EFL) teachers around the world. There are a total of 14 258 registered users of this forum and a total of 31 030 posted articles. The forum consists of about 20 relevant themes of ESL/EFL teaching. Upon entering the forum, users can choose the theme(s) that they are interested in and find more sub-topics under each theme. The live-chat room at Teachers.net is an

example of synchronous discussion. Meetings are usually scheduled in advance and teachers interact with others at the assigned meeting time.

Digital Curriculum Resources for Students

There is a continuum of digital curriculum resources for students with increasing complexity and comprehensiveness. The simple form of digital content is electronic books that are converted directly from printed materials. There are also digital contents in the form of supplementary CDs that come with print-based books. Following that, a series of digital materials – drill and practice programs, tutorials, multimedia learning materials, simulations, and educational games, can be integrated to support the curriculum or serve as stand-alone learning materials. Moreover, there are databases and encyclopedias such as Wikipedia, communication, and cooperative environments. Furthermore, there are comprehensive digital contents such as online courses and virtual learning environments that stand alone as digital curriculum.

E-textbooks

E-textbooks are some of the common digital curriculum resources for students. They vary in terms of degrees of interactivity. Some e-textbooks are simply electronic copies of the printed texts, some add advanced organizers and search functions into the text to ease the navigation of the texts, and others add multimedia components. e-textbooks enhance the accessibility and the mobility of as well as the ease of navigation through the textbooks and present multiple views of content, which are projected to make important contributions to a text's effectiveness and enhance motivation. Currently, there are at least 115 000 commercial English-speaking e-book editions offered in the US market. The production of e-texts has been increasing at an annual rate of 20% for the last 20 years. Despite the push for e-textbooks, researchers have noted limited use of e-textbooks in classrooms and suggested that the effectiveness of e-textbooks rests on the design of the e-textbooks and reading-technique training for students.

Educational software

Educational software is another form of curriculum delivered with digital technology. There are several categories of educational software: (1) productivity and presentation software (application software) that enables students to produce and present documents, spreadsheets, databases, and other products, such as Powerpoint, Hypercard, and The Writing Center; (2) drill and practice software that helps students practice on content they have already learned and provides immediate feedback, such as Number Munchers and Learn to Speak Spanish; (3) tutorial software that teaches a specific task, skill, or application, such as Success with Typing on keyboarding instruction

by Scholastic; (4) simulation software that gives students life-like experience to explore concepts in individual subject matters or integrated thematic units such as SimCity and Sky Lab; and, (5) educational games that embed problem-solving activities and enable experiential learning, such as the Carmen Sandiego series, the Oregon Trail series, and the JumpStart series.

Due to the large variation in the forms and types of educational software, research on the effectiveness of educational software is controversial. The most recent and perhaps the most comprehensive study on the effectiveness of educational software in the US found that educational software has no effect on student achievement. There is an abundance of literature lending support to the educational value of simulations and games in various fields such as medicine, language, and business. However, researchers also noted that their effectiveness depends a lot on the designs of the software and the characteristics of the users.

Online learning systems

Online learning systems have two basic forms: Online CMS and three-dimensional (3D) immersive online learning environments. A CMS typically mimics the traditional course organization and structure. It typically consists of user and content management components. Common online learning management systems include open source systems like Moodle and Sakai, and commercial systems like Angel and Blackboard. 3D virtual-learning systems provide interactive environments that attempt to provide a more immersive and social learning experience. For example, the River City Project is an interactive computer simulation that embeds standards-based content and learning activities into a 3D virtual world to teach middle-school students disease transmission and scientific inquiry. Slooodle combines Second Life, a 3D social game environment and Moodle, an open-source CMS, to provide well-managed immersive learning experiences to students. Online learning systems have been used to deliver completely online curriculum and used as a supplementary component to face-to-face instructions.

Although the research on 3D virtual worlds is still at initial stages, researchers have found these online learning systems extremely helpful in facilitating collaborative learning, peer mentoring and learning community building, and supporting situated learning in immersive, psychosocial contexts (Dieterle and Clarke, in press).

Online courses

Online courses have been gaining more and more popularity in all sectors of education. According to an NCES report in 2003, 56% of degree-granting postsecondary institutions offered distance education courses with around

three-million enrollments in the 2000–2001 academic year, and another 12% planned to start online education courses in the next three years. During the 2002–2003 school year, about one-third of public school districts with an estimated 8200 public schools had students enrolled in online courses – an estimate of 328 000 K–12 students enrolled in distance courses. Some online courses are offered in a blended manner with the majority of the learning contents and activities online, supplemented by some face-to-face meetings. Other online courses are offered completely online through either asynchronous or synchronous interaction or a combination of both. There is a great variation of online courses in terms of interactivity both among the students and with the instructors.

A series of meta-analyses on distance learning and online learning have concluded that they are at least as effective as face-to-face instruction (Bernard *et al.*, 2004; Zhao *et al.*, 2005). At the same time, research has also identified a variety of contextual and learner factors that affect the effectiveness of distance courses (Bernard *et al.*, 2004, 2007; Zhao *et al.*, 2005).

Issues and Concerns

Digital technology has been transforming both the content and format of curriculum. This transformation brings along a series of issues and concerns.

Ability to Use Digital Delivered Curriculum

Fast adoption of digital delivered curriculum does not necessarily guarantee real use (Cuban, 2001). Abundant evidence suggests that the best way to integrate technology into curriculum is to establish strong links between digital contents and curricula, that is, to align digital content with specific curriculum requirements. Thus educators need to develop the ability to make use of the available digital resources, which includes both the ability to conceptualize curriculum in a new way and the technical ability to use the necessary hardware and software (Mishra and Koehler, 2006; Zhao, 2003). Ability to use digital delivered curriculum also includes the ability to discern the authenticity and credibility of the digital resources. With the large amount of digital contents available, teachers must have the ability to evaluate their reliability and usefulness.

Power Relationship in Developing and Using Digital Delivered Curriculum

There are several layers of relationship that need to be dealt with in digital delivered curriculum: between technology and human, instructor and designer, and instructor

and student. First, technology and humans have their respective advantages in teaching and learning. Technologies' strengths lie in providing personalized learning, rich media content and representations, learning tools for basics, real-life simulation, and tremendous storage and computing power. However there are also activities and practices for which humans have a decided advantage: high-level communication practices, relevant feedback, social interaction, and instructions about learning strategies. Thus when developing and using digital delivered curriculum, educators need to capitalize on the respective strengths of technologies and human factors. In digital delivered curricula, teachers should play a facilitative and monitoring role, leaving technologies to the mechanical aspects of teaching and learning.

The relationship between designers and the instructors is also crucial. In the K-12 context, designers of the curriculum are often not the instructors. For any curriculum reform to be successful and sustainable, the involvement of the instructors is indispensable. The instructors' authoritative voice lies in their intimate knowledge of the content, their students, their colleagues, their school structures, and the resources available to them.

Technology can potentially change the power equation within the curriculum as well. Technologies that enable students to publish their work make students partners in generating knowledge and contributors to the curriculum. In a way, a digital delivered curriculum allows students to negotiate their individual learning. As such, the instructor or designer has less control over the curriculum. Teachers need to adjust their mindset to face the changed power equation within their terrain, and balance student involvement and teacher autonomy.

Attitudes toward the Effectiveness of Digital-Delivered Curriculum

There are two unproductive attitudes toward digital technology in curriculum planning and delivery. The first is the naive belief that multimedia technology automatically improves student learning. While digital technology indeed holds tremendous capacity for improving learning when used properly, poorly designed digital curriculum can be as ineffective as curriculum delivered using traditional media. Multimedia instruction must be designed in ways that are consistent with how people learn in order to promote learning (Bransford *et al.*, 1999; Bruner, 1993; Mayer, 2001). The second unproductive attitude is the belief that if you build it, they will come. While it is true that digital technology can potentially reach a broad audience, it does not guarantee eventual uses. Many digital curriculum resources have been underutilized and wasted. Thus it requires the developers to seriously consider how to ensure that their materials are adopted.

There is little doubt that both the capacity of and access to digital technology will continue to grow in the foreseeable future. It is likely to replace traditional print technology as the primary medium for curriculum delivery. However, how soon that can happen and to what degree the educational potentials of digital technology can be realized depends on a host of factors: institutional (how schools and educational systems respond to technological changes), pedagogical (how pedagogical theories and practices interact with technological capacities), and technological (how technology changes and how it reacts to school needs).

Bibliography

- Baddeley, A. (1998). *Human Memory*. Boston, MA: Allyn and Bacon.
- Bernard, R. M., Abrami, P. C., Borokhovski, E., *et al.* (2007). Beyond the obvious in the "what works in distance education" question: Preliminary results of a meta-analysis of DE vs. DE comparative studies. *Paper Presented at AERA*, Chicago, IL.
- Bernard, R. M., Abrami, P. C., Lou, Y., *et al.* (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research* **74**, 379–439.
- Bransford, J. D., Brown, A. L., and Cocking, R. R. (1999). *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academy Press.
- Bruner, J. T. (1993). *Schools for Thought: A Science of Learning in the Classroom*. Cambridge, MA: MIT Press.
- Cognition and Technology Group at Vanderbilt (CTGV) (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher* **19**, 2–10.
- Cuban, L. (2001). *Oversold and Underused: Computers in Education*. Cambridge, MA: Harvard University Press.
- Dieterle, E. and Clarke, J. (2008). Multi-user virtual environments for teaching and learning. In Pagani, M. (ed.) *Encyclopedia of Multimedia Technology and Networking*, 2nd edn., pp 1033–1041. Hershey, PA: Idea Group.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge, MA: Cambridge University Press.
- Leu, D. J., Jr. (2000). Literacy and technology: Deictic consequences for literacy education in an information age. In Kamil, M. L., Mosenthal, P., Pearson, P. D., and Barr, R. (eds.) *Handbook of Reading Research*, vol. 3, pp 734–770. Mahwah, NJ: Erlbaum.
- Mayer, R. (2001). *Multimedia Learning*. New York: Cambridge University Press.
- Mishra, P. and Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record* **108**, 1017–1054.
- Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In Salomon, G. (ed.) *Distributed Cognitions: Psychological and Educational Considerations*, pp 47–87. Cambridge: Cambridge University Press.
- Reinking, D. (1998). Introduction: Synthesizing technological transformations of literacy in a post-typographic world. In Reinking, D., McKenna, C. M., Labbo, L. D., and Kieffer, R. D. (eds.) *Handbook of Literacy and Technology: Transformations in a Post-Typographic World*, pp xi–xxx. Mahwah, NJ: Erlbaum.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Zhao, Y. (ed.) (2003). *What Teachers Should Know about Technology: Perspectives and Practices*. Greenwich, CT: Information Age.
- Zhao, Y., Lei, J., Yan, B., Tan, H. S., and Lai, C. (2005). What makes the difference? A practical analysis of research on the effectiveness of distance education. *Teachers College Record* **107**, 1836–1884.

Further Reading

- Bates, T. (1995). *Technology, Open Learning and Distance Education*. London: Routledge.
- Gee, J. P. (2003). *What Video Games Have to Teach Us about Learning and Literacy?* New York: Palgrave Macmillan.
- Gordon, D. T. (2003). *Better Teaching and Learning in the Digital Classroom*. Cambridge, MA: Harvard Education Press.
- Johnson, K. and Maguisin, E. (2005). *Exploring the Digital Library: A Guide for Online Teaching and Learning*, 1st edn. San Francisco, CA: Jossey-Bass.
- Tapscott, D. (1998). *Growing up Digital: The Rise of the Net Generation*. New York: McGraw-Hill.

Relevant Websites

- <http://www.angellearning.com> – Angel.
- <http://www.aolatschool.com> – AOL@SCHOOL.
- <http://www.atutor.ca> – ATutor.
- <http://online.union.edu> – Blackboard.
- <http://moodle.org> – Moodle.
- <http://teacher.scholastic.com> – Scholastic.
- <http://www.sloodle.com> – Sloodle.
- <http://www.thebritishmuseum.ac.uk> – The British Museum.

Textbook Development and Selection

P Shannon, Pennsylvania State University, University Park, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Textbooks are fixtures in schools, nearly as ubiquitous as teachers and tests. Traditionally, textbooks have been considered enduring repositories of knowledge that enable students to move past their daily experiences in order to grasp the possibilities encoded in the textbooks' articulation of history, the sciences, mathematics, language, the arts, and other topics deemed to be important. Educators have assumed that the relative extent to which students could understand those possibilities predicted their life paths throughout their further schooling and into the workplace. Attending to textbooks, then, is fundamental to understanding schooling and its role in society. As Michael Apple notes,

textbooks are simultaneously economic goods (they are often sold to students, parents, and schools), political objects (they are subject to state control and regulation and hence are the result of political and ideological tensions and compromises), and cultural representations (what is included and not included, and how such knowledge is organized, is a form of cultural politics). (Apple, 2003)

Although school textbooks have limited markets, publishers generate over 10-billion US dollars income annually through textbook sales worldwide. Textbooks are commodities – subject to market forces, even within countries with tight regulations or central controls. Following trends in other industries within a global economy, the number of independent school textbook publishers has decreased through mergers and purchases. For example in the United States, four corporations control over 85% of the school textbook market – two of these corporations are international firms and another was recently purchased by a consortium of private equity groups (Sewell, 2005). Consolidation within any market raises concerns for prices, product quality, and distribution (Baker, 2005).

School curricula are selected from all possible knowledge and ways of knowing. Who makes these curricular choices, how they are made, and by what criteria create politics that extends well beyond textbook pages. Groups vie to participate in these decisions in order to influence the knowledge that will be authorized as most important by the governments. As a primary means to translate the curriculum into instruction, textbook content and design must match the curricular choices in order to be considered viable – regardless of whether they are state or privately produced. In this way, textbooks encode the

power relationships among social groups within nations and states.

Despite the economic and political conditions of textbook production, the consequences of textbooks within school classrooms cannot be assumed. Allan Luke (1988) puts it simply, “For texts do not always mean or communicate what they say.” Rather, teachers and students construct meanings from textbooks employing values, beliefs, and ideas from the multiple groups of which they are members. This does not prevent their construction of meanings closely aligned with those that curricular designers and textbook publishers intend, but it is to suggest that those intended meanings are mediated rather than simply translated. Textbooks are a site and a result of economic, political, and cultural struggles and compromises.

A Global Focus

In its *Comprehensive Strategy for Textbooks and Learning Materials* (UNESCO, 2005), the United Nations Educational, Scientific, and Cultural Organization (UNESCO) affirms the central role of textbooks in schooling. Building on its 1995 textbook report, the comprehensive strategy articulates new challenges for textbooks in a world becoming increasingly connected economically, politically, and socially. Acknowledging that digital media and the Internet enable some populations to supplement textbook forms and information, UNESCO seeks policy, quality assurance, and greater access concerning printed textbooks that will promote equity in school outcomes internationally. Although cognizant of the potential of textbooks to help preserve local languages and cultures, the *Comprehensive Strategy* seeks international standards, and it poses and addresses six general questions for all concerned with schooling as a basic human right:

1. What is the purpose of textbooks?
2. How should the content of textbooks be decided?
3. What means of textbook production are possible?
4. How should textbooks be selected?
5. By what means should textbooks be distributed?
6. How should textbooks be used?

The *Comprehensive Strategy* treats the question of the purpose as a given: “The primary purpose of textbooks – to transmit knowledge, values, attitudes, skills and behaviors – is a constant”. Textbooks are expected to convey the accepted truths within academic disciplines,

enabling students to approximate the habits of mind and body of acknowledged experts in those fields. By sequencing the content of textbooks appropriately, schools can enable students to develop expertise in one or more academic disciplines according to their abilities and capacities. All other elements of students' lives being equal, the purpose of the textbook is to equalize access to that information.

The content of textbooks, then, becomes paramount. UNESCO has enough experience with textbook revisions since World War II to recognize the controversial nature of how textbook content is decided and by whom. "The question of which knowledge, which values, and which skills [textbooks] teach is of concern to localities, nations, and to the international community." Although the authors of the report entertain the possibility of conflicts among the local, national, and international levels of participation, they invoke the principle of rights-based education in order to suggest that human understanding and peace should guide all parties during their considerations of content in geography, history, literature, and even biology textbooks.

Textbook production includes issues of participation, design, publishing, and expense. Traditionally, countries with national school systems and the economic means have developed centralized systems to choose content, design, and publish textbooks. Although centralized practices afforded economies of scale and tight control of textbook content, the *Comprehensive Strategy* presents a market perspective for the production of textbooks, in which private publishers supplant the centralized organizations in order to enable more flexibility, innovation, and participation. Private companies, they argue, will build a sustainable context in which periodic review and modifications of textbooks can occur without taxing public funds. Moreover, the authors contend that privatization of textbook production will improve national and local economies, providing jobs and taxes.

A market economy means competition among publishers because publishers' customers must provide consumers with choices to produce sales. In the report, UNESCO considers the selection of textbooks within countries with the capacity to support multiple publishers of textbooks and for countries and regions that cannot support any textbook industry at all. At both ends of this spectrum, the authors of the report contend that decisions concerning textbook selection should be made as close as possible to the students who will use them. Toward this end, they offer policies intended to ensure that teachers and parents are able to participate in the selection among alternatives. Invoking a rights-based philosophy again, the authors reiterate their concern that minority populations are considered and involved in the selection process because their exclusion will diminish minority students learning from the selected texts.

The *Comprehensive Strategy* commits UNESCO to assist national and local school systems in order to ensure that each student has access to at least one textbook. The authors offer two barriers to equitable textbook distribution: cost of textbooks and a lack of appropriate infrastructure. In regions with few libraries or bookstores, the textbook may be the sole text available for students to read. Without mentioning the generosity of more affluent countries and publishers that send used and old textbooks to such regions, the UNESCO report states that the organization favors quality and up-to-date textbooks in the hands of students worldwide.

Making sure that every student has a textbook will have little effect unless the textbooks are part of what UNESCO labels quality education. Here the issue comes full circle, as the authors state that quality of use can only be measured against the purposes of textbooks. The *Comprehensive Strategy* advocates that teachers use textbooks to prompt, challenge, guide, and illustrate in ways that will accommodate the learning styles of all the students in their classes. The authors assume that the privatization of textbook production will bring innovations from publishers as they work to capture more of the market share. Such innovations would include charts, photographs, and illustrations that have proved too expensive for some national publishing operations. Kress (2003) argues that the printed page of textbooks have taken on the visual grammar of the computer screen affording readers more points of entry and nonlinear modes of addressing the text. More innovative textbooks provide teachers with more opportunities to use the texts creatively. UNESCO advocates that schools engage in research in order to determine the effectiveness of the textbooks, the innovations, and the new pedagogies, leading to new innovations and greater learning.

Research on School Textbooks

Despite their obvious importance in educational programs, Jason Nicholls (2005) laments the current state of research on school textbooks. He considers the extant research to be under-theorized and weak methodologically, and critiques three recent attempts to develop guidelines for further research: Falk Pingel's (1999) *UNESCO Guidebook on Textbook Research and Revision*, Robert Stradling's (2001) *Teaching 20th Century European History*, and Jaan Mikk's (2000) *Textbook: Research and Writing*. Nicholls organizes his critique loosely around Habermas' (1972) framework for understanding how human interests lead to different types of knowledge and research. Grounded in different aspects of social existence – work, interaction, and power – these interests produce different types of knowledge and different ways of determining whether knowledge claims are warranted.

Work refers to the way one controls and manipulates one's environment. Work interests about textbooks would include determining how textbooks are produced, selected, and distributed effectively and efficiently; how teachers use textbooks; and how effective the textbooks are in reaching intended goals. Work interests are concerned with getting things done well and seek knowledge about causes and effects of actions and consequences. Such knowledge is established through objective empirical investigations and is governed by technical rules.

Interaction interests consider the social aspects and meaning of human life. Social knowledge is directed by consensual norms, which define reciprocal expectations among and between individuals within specific contexts. Although social norms relate to empirical and analytical propositions, their validity is grounded only in the intersubjectivity of mutual understanding of intentions, and can never be completely determined. Rather than the empirical analytic methods to determine causality for work interests, the knowledge of norms, expectations, and intentions requires interpretive methods investigating negotiations and their meaning among participants in social events. Interactions with interest, knowledge, and research would consider the negotiation surrounding the purpose, content, production, selection, distribution and use of textbooks, and the ways that participants understand their actions and the entire process.

These negotiations, however, do not take place among equals. During each event, groups of participants exert different levels of force to accomplish their objectives. For example, government officials, often, influence textbook content more than classroom teachers; publishing companies typically, effect textbook design more than parents; and educational psychologists, usually, direct studies of textbook use more than school administrators because past social relations have framed institutional practices in ways that afforded them greater power. This power is not only coercive, but also discursive, leaving the participants with few ways to conceptualize alternative ways of acting during these events. Interests in this power would produce knowledge concerning the ways in which textbooks enable and/or limit the continuous development of members of differing groups within a social context. Production and validation of knowledge of power involves studies of the intersections of history, biography, and social structures with a goal of facilitating human agency to address unjust conditions.

Work Knowledge about Textbooks

Researchers and educators who pursue work knowledge about textbooks accept the purpose of textbooks without question. They seek to understand how that purpose can be realized most effectively within specific environments.

Toward that end, they establish empirical investigations of the relationships among the six questions reiterated by UNESCO's Comprehensive Strategy. For example, Erik Knain (2001) analyzed Norwegian middle grade science textbooks and found an overemphasis on science as an individual activity of discovering truth, projecting science as indisputable and directly from nature. Acknowledgment that science is conducted within a community with norms, values, interests, funding, priorities, and institutional organizations was largely omitted from the books. None conveys the ways in which the scientific community within those constraints transforms scientific findings into established facts that are to be trusted as presenting the world as it really is. He concludes that this imbalance poorly prepares students to understand and engage in science.

In a report on textbook adoption, the Thomas B. Fordham Institute presents a case for how reform in the K-12 textbook selection process in the United States would improve textbook content and effectiveness of textbooks on student learning (Finn and Ravitch, 2004). The report describes the process by which 21 states use centralized systems to broker the content and production quality of textbooks in all subjects, and then, make them available to school districts without charge. "This form of textbook adoption distorts the market, entices extremist groups to hijack the curriculum, enriches the textbook cartel, and papers the land with mediocre instructional materials that cannot fulfill their important educational mission." According to the report, the flawed system permits interest groups to insist on preferred content and design that result in "absurd sensitivity guidelines that have dumbed down textbook content in an attempt to render them inoffensive to every possible ethnic, religious, and political constituency." In the end, states with these centralized systems have the lowest achievement test scores in the United States. The report's recommendations echo those of the *Comprehensive Strategies* – encourage small private publishing companies to participate in publishing and make selections as close to the classroom as possible.

In a discussion paper from the Institute for the Study of Labor, Fragfich and Michaelowa (2005) use new statistical techniques in order to determine the effects of textbook distribution on students' learning in Francophone sub-Saharan Africa. Within these regions, textbooks are provided privately either through family purchase or communal ownership. By translating methods of labor research in which more variables can be factored into equations, the researchers make the case that textbook effects have been underestimated because previous studies have only examined the impact on textbook owners within these countries. By recognizing that more students than owners use and benefit from available textbooks, the researchers demonstrate the remarkably efficient power of textbooks to achieve their purpose of transmitting

knowledge, values, attitudes, skills, and behaviors. From these results, they recommend that funds should be allocated to the quality and distribution of textbooks before they are directed toward other innovations (e.g., reduction of class size, raising teachers' salaries, or other forms of instructional tools).

Interaction Knowledge

Researchers and educators interested in interaction knowledge about textbooks seek to describe how participants negotiate and understand the norms, values, and meaning of textbook purpose, content, production, selection, distribution, and use. In order to produce interaction knowledge, researchers and educators observe the interactions among participants and analyze their words and symbols as they work through the different events associated with textbooks and schooling. By making the intentions behind those actions and words more transparent, interaction knowledge promises the possibilities of more inclusive and productive outcomes in future negotiations surrounding textbooks. For example, in *School Science and Mathematics*, William Newman (2004) recounts his role as a parent on an elementary-level mathematics textbook adoption for a US school district. His efforts include analyses of the school district's criteria, the various ways in which teachers, parents, and administrators interpreted those criteria, an inability of the committee chair to elicit discussion among the selection committee members after validating a particular position early in the process, and an ultimate intervention from higher-level administrators and the school board. While Newman appears to describe dysfunction, he argues that each group acted appropriately by representing its intentions for the textbooks selection in explicit and subtle ways.

On a much larger scale, the Ateneo School of Government (2005) in the Philippines presents a report on the development of new procedures for monitoring and facilitating textbook production and distribution based on the coordination of private and public resources. Through surveys and interviews, they assess how the different groups understand their charge with the new system and field concerns for the greater coordination and articulation of efforts. Issues of public service collide with profit motives as the boy and girl scouts wait for, and then unload textbooks from trucks that local Coca Cola bottlers lent to the operation. Some parents objected to the company's participation in the textbook project as they felt it contradicted parents' efforts to improve the nutritional value of school lunches. The report ends with recommendations for how the various constituents might be blended, and blend themselves, into a culture of textbook quality and assessment through more open communication of intentions.

In *Globalization, Society, and Education*, McEneaney and Nieswandt (2006) describe the effects of the fall of the wall on an East German textbook publishing house. Prior

to that, the publishers worked within the highly centralized, tightly controlled, norms of government regulation within a planned economy. After being purchased by a Western publishing company in 1991, the company was plunged into a fiercely competitive textbook market with a broader focus on world cultures. The study examines the ways in which these new norms of publishing and textbook production were negotiated between the existing publishing house and the new owners. Complicating the transformation, the previous consumers of the textbooks remained closely aligned with a more nationalistic focus, despite attempted educational reforms. The study adds remarkable depth to the meaning of the tension between the global and local in textbooks and schools.

Power Knowledge

Power interests consider how the work and interaction of life position social groups and individuals within historical and institutional frameworks that distribute the benefits unequally among groups. The production of power knowledge is meant to inform groups concerning the mechanism through which the frameworks were and are constructed and maintained and to suggest how those poorly served do resist and reform the frameworks and distribution working toward equity. Power knowledge concerning textbooks identifies why and how groups struggle for control of content and flow of school knowledge.

For example, in *Pedagogy, Culture and Society*, Jyh-Jia Chen (2002) presents a power analysis of the textbook deregulation in Taiwan during the 1990s. Following Apple (1993), Chen argues that textbooks are not only commodities vying for a market share; they are also state-regulated products and cultural artifacts that reflect and create values and ideas. The historical forces of the 1970s and 1980s in Taiwan led both popular and government groups to call for a liberalization of the economic, social, and political norms and a repositioning of Taiwanese identity. Attempts to reform society provided opportunities to question the traditional Taiwanese curriculum and its conservative impact on popular cultural values and actions. Over time, the liberalization project, then, afforded the government and dominant social groups control in the creation of the new official knowledge to be encoded in school textbooks. As Chen concludes, "textbook deregulation was a hegemonic project through which the state and oppositional forces contended for control over textbook production, educational decision-making, and legitimate knowledge."

In *Argentine Educational Reform: Tensions between Democratic and Neoliberal Discourses*, Gabriela Mendez (2006) employs critical discourse analysis of the print and illustrations within social science textbooks in order to identify the government's official definition of democratic values and citizenship. Mendez demonstrates how the possibilities of democracy after the fall of the military

dictatorship in 1993 were/are narrowed further at each seceding level from law to policy to textbooks. In the passages and illustrations, she finds repeated appeals to neoliberal values in which the entrepreneurial self is created, seeking happiness, self-esteem, actualization, and fulfillment through individual projects and markets. By selecting certain topics and images and contrasting the limits of the past with the promise of the future, the textbooks normalize the neoliberal subjectivities and make it difficult for teachers and students to imagine other definitions of democracy and ways of being a citizen. Mendez considers the textbook renderings to be betrayals of the democratic promises of 1993 and offers teachers oppositional ways to use the textbooks in order to revitalize those promises.

Shannon (2007) reports on the role of commercial reading textbooks in the search for the one best method of teaching reading throughout the US during the twentieth century. From educational psychologists' first call for a real rationalization of reading instruction through to the scripted reading lessons promoted in the current federal education law, No Child Left Behind, educational publishers have offered textbooks and teachers' guides as the technological solutions to variability among teachers and students in order to reach universal literacy. Claiming scientific authority and backed by the state, the teachers' guidebooks direct the reading instruction in over 95% of the classroom, supplying the goals, planning, content, instruction, and tests for daily teaching; deskilling the classroom teacher; and limiting students' reading development to tested skills. Despite rhetoric concerning a free market of reading textbooks, only four companies control 90% of the reading instruction sales. Recently, the Department of Education Inspector General concluded that educational scientists while acting as federal officials violated conflict-of-interest laws when they forced states and school districts to purchase textbooks in which they had financial interests.

The State of Textbooks

The researchers and educators pursue different interests concerning textbooks, and thereby demonstrate that there are multiple ways of addressing the six questions raised in UNESCO's comprehensive strategy. Even the purpose of textbooks, which the report assumed to be fixed, is negotiated among participants who produce, select, and use them. While those pursuing work interests define the purpose of textbooks as furthering students' understanding in a straightforward manner, those working to produce power knowledge demonstrate how textbooks and textbook content are used to limit students' development in fundamental ways. Issues of content, production, selection, distribution, and use are equally complex and contested worldwide. Market-forces pressure centralized the production in Asia, Africa, and the United States, but did

not necessarily bring innovation; interest groups negotiate for recognition within textbook content in South America and Europe, yet treatments remain unequal; distribution problems, while more acute in some regions, are still apparent in wealthy nations; and use of textbooks varies greatly within and among schools at all levels, across all disciplines, and in all areas, and is not always an improvement on teaching without textbooks. Much of the research on textbooks makes the recommendations of the *Comprehensive Strategy* problematic. Local and particular knowledge and tactics challenge the universal and the strategic concerning all aspects of textbooks.

By definition, textbooks represent the sanctioned knowledge of the world to students. Throughout their history, schools have afforded textbooks sufficient authority to spawn social aphorisms of correctness – doing something by the book or an example is a textbook case. Such important artifacts warrant interest from all who are concerned with schooling and education.

Bibliography

- Apple, M. (1993). *Official Knowledge*. New York: Routledge.
- Apple, M. (2003). Prejudice and pride: School histories of the freedom struggle in India and Pakistan by Krishna Kumar [Book Review]. *Comparative Education Review* 47, 123–125.
- Ateneo School of Government (2005). *Report on the Civil Society Participation Textbook Count 3*. Makati: Author.
- Baker, D. (2005). Are copyrights a textbook scam? *CEPR Report*. Washington, DC: Center for Economic and Policy Research.
- Chen, J.-J. (2002). Reforming textbooks, reshaping school knowledge: Taiwan's textbook deregulation in the 1990s. *Pedagogy, Culture and Society* 10, 39–72.
- Finn, C. and Ravitch, D. (2004). *The Mad, Mad World of Textbook Adoption*. New York: Thomas Fordham Foundation.
- Fraglich, M. and Michaelowa, K. (2005). *Discussion Paper 1519: Peer Effects and Textbooks in Primary Education: Evidence from Francophone Sub-Saharan Africa*. Bonn: Institute for the Study of Labor.
- Habermas, J. (1972). *Knowledge and Human Interests*. Boston, MA: Beacon.
- Kress, G. (2003). *Literacy in the New Media Age*. New York: Routledge.
- Knain, E. (2001). Ideologies in school science textbook. *International Journal of Science* 23, 319–329.
- Luke, A. (1988). *Literacy, Textbooks and Ideology*. Philadelphia, PA: Falmer.
- McEneaney, E. and Nieswandt, M. (2006). Constructing school knowledge after the wall: The case of textbook publishing in the former German Democratic Republic. *Globalization, Society, and Education* 4, 337–365.
- Mendez, G. (2006). *Argentine Educational Reform: Tensions between Democratic and Neoliberal Discourses*. Unpublished dissertation, Pennsylvania State University, Pennsylvania.
- Mikk, J. (2000). *Textbook: Research and Writing*. New York: Peter Lang.
- Newman, R. (2004). Picking a textbook. *School Science and Mathematics* 104, 361–368.
- Nicholls, J. (2005). The philosophical underpinnings of school textbook research. *Paradigm* 3, 24–35.
- Pingel, F. (1999). *UNESCO Guidebook on Textbook Research and Textbook Revision*. Braunschweig: Georg Eckert Institute for International Textbook Research.
- Sewell, G. (2005). Textbook publishing. *Phi Delta Kappan* 86, 498–502.
- Shannon, P. (2007). *Reading against Democracy: The Broken Promises of Reading Instruction*. Portsmouth, NH: Heinemann.
- Stradling, R. (2001). *Teaching 20th Century European History*. Strasbourg: Council of Europe Publishing.

UNESCO (United Nations Educational, Scientific, and Cultural Organization) (2005). *Comprehensive Strategy for Textbooks and Learning Materials*. Paris: United Nations Educational.

Further Reading

Apple, M. (2000). *Official Knowledge. Democratic Education in a Conservative Age*. New York: Routledge.

Bourdieu, P. (1990). *Reproduction: Education, Society, and Culture*, 2nd edn. Thousand Oaks, CA: Sage.

Relevant Websites

<http://www.textbookreviews.org> – Education Research Analysts, an original contribution to the Christian Conservative Intellectual Renaissance.

<http://www.gei.de> – Georg Eckert Institute for International Textbook Research.

<http://www.rethinkingschools.org> – Rethinking Schools works to improve schools and reshape school reform in the United States.

<http://faculty.ed.uiuc.edu> – Westbury site, *Paradigm*, a journal of textbook colloquium.

CURRICULUM DEVELOPMENT – SUBJECT MATTER

Contents

Curriculum and Globalization: Higher Education
Curriculum Development as Subject Matter: Social Studies
Curriculum Development in the Area of Reading
Curriculum in the Arts
Curriculum: Foreign Language Learning
First Language Learning
Learning a Second Language in First Language Environments
Literature
Mathematics
Physical Education and Sports
Primary and Elementary/Middle Grades Reading
Science

Curriculum and Globalization: Higher Education

K H Mok, The Hong Kong Institute of Education, Hong Kong, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

No matter how we assess the impact of globalization on contemporary societies, no one can deny that globalization and the evolution of the knowledge-based economy have caused dramatic changes in the character and functions of education in most countries around the world. Recent research has demonstrated in a number of ways how the idea of the knowledge economy represents a qualitative shift from the assumptions and policy prescriptions of the 1980s and 1990s (Peters, 2001). A close scrutiny of agenda-setting documents published by the World Bank and the Organization for Economic Co-Operation and Development (OECD), as well as analyzing how nation-states in East Asia and in the West have responded to the challenges generating from the rise of the knowledge economy, has clearly indicated strong needs required by the knowledge economy to transform education systems and fundamental shifts from teaching-oriented to learning-emphasis reforms (Delanty, 2001). Comparative education analysts even believe that radical education reforms would be insufficient to bring about the shift from education in institutions to learning anywhere, any time and just for me (Dale, 2005). This article sets out in the context briefly outlined above to examine what major values underlying the design of new curricula in coping with the globalization challenges and what are the major curriculum changes. The article concludes with a critical reflection on globalization and curriculum changes.

Globalization: The Rise of the Knowledge Economy and University Education

Whether we agree or disagree to what extent the globalization discourse has affected curriculum design and educational governance, we cannot deny the fact that the rise of the knowledge-based economy has significantly affected the way knowledge is defined. Alvin Toffler, writing in 1990, before the knowledge economy was only a concept instead of being quantified and developed as a form of economy, argues, "The most important economic development of our lifetime has been the rise of a new system for creating wealth, based no longer on muscle but on mind" (Toffler, 1990: 9).

With the growing impact of the globalizing economy, Toffler was prescient in noting the now well-recognized changing patterns in employment internationally, moving from blue-collar, low-skilled jobs to white-collar, highly skilled employment. According to the definition adopted by the OECD, a country with a knowledge-based economy is one where "the production, diffusion and use of technology and information are key to economic activity and sustainable growth" (OECD, 1999: 7). In order to enhance the national capacity to compete in the globalizing world, a growing number of countries have begun to invest more in research and development (R&D) to generate new ideas and new knowledge. According to the recent report published by the World Bank, not only the

developed economies in the West but also the developing economies in Southeast Asia have increased their investments in new knowledge, its creation, dissemination, and adaptation to production (see Table 1).

In addition to the impact of the rise of the knowledge-based economy on educational transformations, another major drive for curriculum changes in higher education is the new definition of education as a tradable service. The Uruguay Round of the General Agreement on Tariffs and Trade (GATT) and the General Agreement on Trade in Services (GATS) classified education as a service and sought to ensure the gradual reduction of restrictions on educational services such as technology transfer, consultancy, and distance education. The eventual aim with the newly defined notion of education as a tradable service is to create an international marketplace in education (Kelsey, 1998; Knight, 2006). Abby Riddell has succinctly summarized the impacts of the rise of the knowledge-based economy and the new definition of education as a service to education:

Changing trade patterns influence the productive possibilities of the economy and thereby the demand for education, the uses put to education, and the demands made on education for tailoring the workforce to those demands. (Riddell, 1996: 1363)

In response to the shift from factory-based production to hi-tech and knowledge-intensive industries, the Fordist style of industry which was characterized by “narrowly specialized tasks on the factory floor and the fragmentation of skills and knowledge” is rendered inappropriate (George, 2006: 592). With industries of the future relying more on individuals who are able to adapt to rapidly changing environments, a number of countries have begun to reform their university curricula in order to nurture students with the flexibility to adapt to changing situations and emphasis has been placed upon work-based education. Having critical reflections on the role

of education in the knowledge-based economy, some scholars even argue for a shift from traditional economic models that regard education/training/knowledge formation/technology as exogenous factors in economic growth to newer human capital models developed by economists. Moving beyond the Fordist model of education in preparing students skills for the manufacturing sector, the post-industrial age needs students to be innovative, creative, flexible, and adaptive to their changing situations. In short, the knowledge-based economy has created a new set of expectations and demands for skills in education. Recognizing the changing dynamics of the global marketplace, it is believed that higher education institutions have to reform their curricula in general and teaching/learning strategies in particular in order to enhance students to become more creative, innovative, and adaptive to rapid social and economic changes.

Globalization and Curriculum Design: Some Major Underlying Values

There are many ideas/values underlying the debates on globalization and curriculum reforms. Some core values are related to the proposed shift from a teacher-oriented to a learner-oriented approach in learning, advocating the importance of lifelong learning, work-based learning, and whole-person development. Extending the market model, the World Bank has advocated lifelong learning for the global knowledge economy in recent years. In its 2003 report titled, *Lifelong Learning for the Global Knowledge Economy*, the World Bank states clearly that:

A knowledge-based economy relies primarily on the use of ideas rather than physical abilities and on the application of technology rather than the transformation of raw materials or the exploitation of cheap labour ... The global knowledge economy is transforming the demands of the labour market throughout the world. It is also placing new demands on citizens who need new skills and knowledge to be able to function in their day-to-day lives ... Equipping people to deal with these demands requires a new model of education and training, a model of lifelong learning. (World Bank, 2003: 1)

Believing that the knowledge-based economy is dependent upon a regime that is able to promote entrepreneurship, a skilled and educated population with a capacity to be creative and innovative is essential for the success of further social and economic development (Lloyd and Payne, 2003). What kind of education system that could nurture such a population? The answer given by the World Bank is straightforward: this education system must transform itself by “replacing the information based, teacher directed rote learning provided within a formal education system governed by directives with a new type of learning that

Table 1 Research and development expenditure levels, and as percent of GDP, 2002

Countries	<i>R & D spending 2002</i>	<i>R & D as percent of GDP^a</i>		
	US\$ Billions (PPP)	Percentage of world	1992	2002
SE Asia average	3.3	0.4	0.1	0.2
Developed world average	645.8	77.8	2.3	2.3

^aRegional data are sum of R&D divided by sum of PPP GDP. Source: World Bank (2003). *Lifelong Learning for a Global Knowledge Economy*, p 116. Washington, DC: Author. Copyright International Bank for Reconstruction and Development, The World Bank.

emphasizes creating, applying, analyzing and synthesizing knowledge and engaging in collaborative learning throughout the lifespan" (World Bank, 2003: xviii). **Table 2** compares and contrasts the characteristics between traditional and lifelong learning models.

The lifelong learning model set out by the World Bank moves beyond the traditional model which emphasizes "the weightiness of the past in the hands of teachers who prohibit progress through gate-keep learning and imposing a one-size-fit-all curriculum and pedagogy on the pupils" (Robertson, 2005: 161) to a new model that attaches weight to individualized learning plans. Unlike the traditional model, the lifelong learning model emphasizes the importance of developing individual learners' decision making and problem-solving skills as well as teaching them how to learn on their own and with others (World Bank, 2003: 3). In response to this fundamental shift, the UK government has started promoting a new learning approach which attaches weight to personalized learning, developing new strategies to help students learning to learn (DfES, 2004). In order to nurture students with a broader base of knowledge, some academics urge the development of an integrated and interdisciplinary curriculum for university education.

In addition, the globalization forces have not only resulted in developing a learner-oriented paradigm but also provoked debates and discussions about the importance of multiculturalism and internationalization in university education. In order to prepare students to take up the challenges of globalization, it is important to create a

conducive learning environment that could enhance students to appreciate the growing diversity of cultures and social differences, understanding and accepting divergences in social, economic, political, and religious systems across the globe. Recognizing that the globalizing world has become increasingly socially complex, culturally diverse, and politically contested, an ideal curriculum in global/multicultural world should provide equal learning opportunities to all students from different national, ethnic, cultural, and social backgrounds to promote the values of social inclusion, cultural pluralism, and world citizenship (Haigh, 2002). According to McFadden *et al.* (in Morey, 2000: 26), a curriculum addressing the complicated issues generated by globalization challenges should include: (1) the promotion of equity and of social justice; (2) the improvement of intergroup relations and the promotion of intercultural competencies; (3) the reduction of prejudice, stereotyping, and discrimination; (4) the acquisition and the imparting of knowledge of human diversity and commonality; (5) the acquisition of knowledge for cultural consciousness regarding one's own and other cultures; and (6) the development of skill in the critical understanding of the processes of knowledge construction.

Similarly, Leask (1999) points out the importance of the awareness of otherness, suggesting a new curriculum should be able to enhance students' (1) willingness to think globally and inclusively, (2) awareness of local implications for international communities, (3) self-awareness, (4) awareness of professionalism in both global and local perspectives, and (5) values of cultural diversity ability and other cultures, which is the indicator of graduate quality in a global environment. Moreover, Haigh (2002) points out that the progress of designing a curriculum for responding to understand multiculturalism/internationalism is to meet the needs of an international student body, by which no individual student from any particular social group or tradition would be granted an innate competitive advantage. In a deeper sense, internationalization and cultural inclusion are about universal suffrage (Haigh, 2002: 51).

Apart from helping students become more aware of the differences and diversity in cultural, social, economic, political, and religious aspects in the globalizing world, another globalization drive is to prepare students to be more competitive in the global marketplace. Therefore, universities and schools, in the age of globalization, are under pressure to enable their students to compete in an increasingly global world of work. Along with the multiculturalism, students should also be able to handle and cope with a multicultural working environment and to cooperate with people from different backgrounds. In this regard, education or higher education institutions are therefore expected to provide training for their students with transnational skills or competences such

Table 2 Characteristics of traditional and lifelong learning models

<i>Traditional learning</i>	<i>Lifelong learning</i>
<ul style="list-style-type: none"> • The teacher is the source of knowledge • Learner receives knowledge from the teacher • Learners work by themselves • All learners do the same thing • Tests are given to prevent progress until students have completely mastered a set of skills and to retain further learning • Teachers receive initial training plus <i>ad hoc</i> in service training • Good learners are identified as permitted to continue their education 	<ul style="list-style-type: none"> • Educators are guides to sources of knowledge • People learn by doing • People learn in groups and from one another • Assessment is used to guide learning and identify pathways for further learning • Educators develop individualized learning plans • Educators are lifelong learners; initial training and ongoing professional development are linked • People have access to learning opportunities over a lifetime

Source: World Bank (2003). *Lifelong Learning for a Global Knowledge Economy*, p 29. Washington, DC: Author. Copyright International Bank for Reconstruction and Development, The World Bank.

as the mastery of an international language (i.e., English), the mastery of information and communications technology, and the abilities of problem solving, creative and critical thinking, which are considered as major components in a new curriculum which would fit the global world (UNESCO, 2000). It is against such a wider context that a number of scholars have pointed out the importance to strengthen the work-based learning, enhancing students to integrate their academic learning with the real workplace. Work-based learning curriculum and work-integrated learning are becoming more popular, with attempts to promote the tri-production of specific and non-disciplinary knowledge (Boud and Symes, 2000). Having briefly discussed some major values underlying the paradigm shift in education toward a more learner-oriented, work-based, and problem-solving approach, let us now examine major curriculum changes.

Globalization and Major Curriculum Changes

Institutional Change in Promoting Curriculum Reforms

In order to promote the international and global outlook of the students, there has been significant structural change in the curriculum by making students better appreciate multiculturalism. Morey's framework for systemic change regarding multiculturalism at college level is considered as a relevant framework to shape the academic environment in the context of globalization. In the last few years, Morey has implemented her framework to assist nearly 100 American universities to make their students more aware of the cultural diversity and prepare them for a socially complex society. Central to Morey's framework, multiculturalism and internationalism are emphasized by:

1. increasing faculty expertise in multicultural and international education;
2. increasing the knowledge base on multiculturalism and internationalism through faculty research and other forms of scholarship;
3. infusing the curriculum with content and instructional strategies appropriate to the improvement of teaching and learning in multicultural/international contexts;
4. forming linkages and consortia with higher education institutions and other appropriate organizations to draw upon their expertise and resources; and
5. increasing the ethnic and international diversity of students (adapted from Morey, 2000: 27, 28).

With a significant weight attached to the multicultural and international elements in college education, Morey believes that the structural change at the college level is necessary because such changes would support changes in

the curriculum. With a strong conviction to connect the school and the outside world, Morey sees the importance of forming linkages and consortia, by which teachers can learn how to respond quickly to changing external environment. Moreover, the structural change can also be taken inside the school/university in order to break down the rigid hierarchies of the organizations, thereby empowering teachers and students to make educational choices and promoting curriculum changes. According to another study conducted by Vidovich (2004), the schools in Singapore have adopted the concept of consortia to create schools within a school in order to motivate teachers to take a more active role in curriculum reform. Similar to Morey, Vidovich (2004) found one of the schools in Australia has restructured its 22 departments into eight learning areas of curriculum framework. On the one hand, this restructuring exercise has flattened hierarchies in each school for empowering the frontline teaching staff. On the other hand, it also helps the school to develop a curriculum-oriented structure, which departs from the traditional curriculum dominated by mathematics and science to a more integrated and interdisciplinary curriculum (Vidovich, 2004).

Contents

In the context of globalization, education is expected to promote not only multiculturalism but also cultural inclusion. In accordance with Kitano's (1997) change model, the content of a cultural inclusive course should provide alternative perspectives through materials, readings, speakers, and the analysis of a diversity of sources. This multicultural component can be further transformed into a nondominant perspective by reconceptualizing the content through a shift in paradigm or standard (Kitano, 1997). To develop such a content of curriculum, educational institutions and educators may need to change their teaching and learning strategies. As for institutional changes, some institutions were founded to serve a particular social group, by which these institutions self-consciously conceive their role in national, religious, and/or gender-specific terms (Haigh, 2002). Embracing with internationalization, these institutions need not only to abandon their partisan link with those smaller geographical, cultural, or social units but also to reinvent their geographical and cultural identity with a broader view. For educators, they need to transform their styles of teaching from one based on the thought that defining student's role as learning to copy their way to a new way of allowing students to meet their own personally tailored learning needs (Haigh, 2002: 53). According to Haigh, the traditional British curriculum has been widely perceived as predominantly Eurocentric and most British universities have actually built upon an assumed knowledge base of English traditions, history, geography, art, music,

popular culture, and literature. To cope with the growing number of overseas students with diverse backgrounds, the British universities are needed to transform their curricula and teaching/learning strategies. For example, university faculty shall no longer assume that students are native British-English speakers with a thorough understanding of local accents and colloquialisms. Indeed, beyond the language used in its delivery, the pursuit of internationalism and multiculturalism in curriculum should also involve rethinking the presumption of course design and the character of course contents.

Nevertheless, alongside the tension between homogeneity and heterogeneity, globalization can mean the cultural imperialism and colonization that results in a transplantation of Western knowledge production system to other cultural sites and consequently causes the suppression or erosion of the local knowledge traditions (Gough, 2002). Language, for example, is a subject reflecting the tension between the demand for internationalization of curriculum and the resistance of universalizing imperialist discourse and practices. In Asia, many countries, such as Taiwan and Japan, treat adopting English as a medium of instruction on campus as a sign of internationalization of education. Indeed, these countries often have well-established local knowledge systems, which use local language for teaching. However, there is another story in Africa, especially in those former colonies, where the use of English is labeled as a sort of re-colonization or an instrument in effecting social stratification (Gough, 2002). Seen in this light, we must be culturally sensitive and politically aware of the potential tensions between the quest for the global trends and the local resistance.

Instructional Strategies and Activities

As discussed earlier, the change of power relation between teacher and student is another impact of globalization on curriculum. The core value of such a transformation is to involve students in constructing knowledge, thereby encouraging peer learning and critical analysis (Kitano, 1997). Some scholars also believe that even for launching transnational higher education programs, students enrolling in these programs should be given the opportunities to express their interests and expectations, while the offering institutions should take their views into serious consideration in order to meet the educational needs of the students. As Sadiki (1998) suggested, an internationally accessed course should also provide “a democratically responsive learning environment founded on conjunctive and consultative processes” (p. 5). Such a learning environment would leave greater scope for pluralism than conventional didactic instruction. The course should also be taught in a language that all students could understand. Instructor should not use dialects and local accent. The language requirement set by, and the language used in, the institution must be equivalents (Haigh, 2002).

Moreover, the inclusive form of instructional strategies and activities should adopt a method-centered and issue-oriented approach (Kitano, 1997). It is argued that beyond multicultural perspective, the inclusive instruction should also involve multidisciplinary approaches, which generalize different knowledge together based on an issue-oriented theme. Language in the globalization era is redefined as a sort of communication skill, which tends to integrate literacy with information and communication technologies. This simultaneously involves cross-cultural communication, of which students are expected to communicate with alien traditions by making use of various instruments of medium. Hence, many universities have heavily invested in developing multimedia courses. Cameron (2002) suggests that the evolution from language to communication skills is a result of the new work order, in which employer focuses on employees’ ability of interacting with others no matter in verbal and nonverbal (e.g., visual and oral) forms. It is against this context that there is an emergence of a new type of computer literacy, which usually consists of a large amount of visual materials but has very little literacy contents. The rise of nontext type of literacy has resulted in the language–literature divide, which requires a curriculum of language courses to handle different a set of concepts around linguistic features and linguistic analysis.

Assessment Methods

The changes discussed above in the curriculum have inevitably affected the assessment strategies. An inclusive course should be assessed upon multiple methods and alternatives to standard examinations and paper. It should also encourage self-assessment and focus on student growth. More practically, the assessment in an inclusive course should focus on its transferability and connectivity. The newly emerged computer literacy is an example of the new standard in globalized learning environment. As mentioned above, non-text-type materials are commonly used in current language education by highlighting language as a communication skill instead of knowledge. Then, the focus of assessment is no longer the strength of students in linguistic content but their ability to establish interpersonal relationship and being aware of others’ psychological and emotional needs (Cameron, 2002). In those courses set for linking language and cultural studies and technologies, good presentation, for instance, is to use less linguistic content to express more information. The decline of linguistic content in the language courses represents a transformation toward issue-oriented approach of curriculum in the age of globalization. In short, the learning objectives, the curriculum, and the teaching and learning strategies must be aligned with the assessment; hence, suggesting changes proposed in

the curricula should have resulted in changes in terms of assessment measures.

Although some scholars have advocated a diversification of assessment in response to multicultural context, centralization or universalization of assessment has been spread through the trend of globalization in parallel. Cambridge University O-level examination in English is widely adopted as a benchmarking system in many former British colonies (Gough, 2002). Another example is the worldwide adoption of the International Baccalaureate in school education. We can argue that there are different curricula and assessment frameworks, which are internationally adopted and are competing with each other. Therefore, this may not be a rigid homogenization of curriculum and assessment. The key issue here is that these internationally mandated curriculum and assessment frameworks, which usually originate from the West, have been replacing the conventional national framework in some countries. This brings us to reconsider the relationship between state and supranational forces and the extent to which globalization represents homogeneity and heterogeneity.

Classroom Dynamics

According to Morey (2000), classroom dynamics refer to “the human interactions that take place in the classroom between instructor and students, and among students” (p. 33). Sadker and Sadker (1992) argue that race, ethnicity, or sex can be factors affecting the interactions and relations between instructors and students in the classroom. In a globalized and inclusive learning environment, instructors are obligated to prevent any biased behaviors, as they may generate a hidden curriculum, which presents enforced rules about knowledge and behavior valued in the classroom (Morey, 2000). It is worth noting that the advance in technology has transformed the forms of instruction. Web-based conferencing has been used for tutorial and learning. This newly emerging form of teaching allows the private and interruption-free conversation between the instructor and students. The synchronous technology even allows the face-to-face and real-time interaction, which provides greater flexibility and enables collaborative learning through distance education and would eventually improve the quality of teaching as the traditional forms of distance learning tend to focus on content delivery but the synchronous interactions can facilitate problem solving and critical thinking. Nevertheless, there are queries raised about the extent to which a web-based conferencing can formulate a multicultural and international learning environment, which is regarded as an essential element for nurturing personnel who are able to work and learn in international and multicultural settings in the age of globalization. Our above discussions have briefly reviewed some major areas for curriculum changes,

the following will reflect on how far the globalization discourse has really offered more alternatives and diversity for or create potential constraints on nation-states in curriculum design.

Globalization and the Curriculum Reforms: Critical Reflections

Despite the fact that one of the principal values underlying the globalization discourse is to enhance students to appreciate the ideas and practices of multiculturalism and internationalism, it has been a dilemma between homogeneity and heterogeneity in the globalization discourse. Under the influences of globalization, it is believed that education should well prepare students to cope with “a world characterized by cultural diversity, inequity, interconnected-ness, co-operation, and conflict” by developing their cross-cultural skills and competencies (McFadden *et al.*, 1997: 10). Nonetheless, such an interpretation of globalization has simply adopted a practical approach which equates multiculturalism with globalization. In the real world, we are confronted with the tensions between homogeneity and heterogeneity, generating difficulties for formulating new curricula. Since universities/schools in general, and, teachers/academics in particular, have different interpretations of multiculturalism and internationalism, the different approaches that they adopt should have shaped the curriculum differently. In addition, the positioning of the institutions (as local or regional/international one) and the mission/vision in teaching (as locally oriented or cross-cultural oriented) should also affect the way curriculum is designed (Haigh, 2002). In this regard, to what extent would schools/colleges/universities adapt themselves to the changing world depends very much on staff’s and institutions’ willingness and readiness. However, one key question, which we are presently confronting is to what extent the curriculum design should be guided/driven by the globalization agendas/tides? From this point of departure, we must be aware of the tensions between developing a curriculum that can protect/promote local cultures and scholarship or going for the global one. More fundamentally, we should be aware of the potential costs of going along with the global trends: limiting choices instead of promoting diversity in curriculum design.

In addition, we should not simply understand that the process of curriculum formulation as an entirely autonomous enterprise beyond the control of the state or exempted from the constraints coming from the economic, social, cultural, and historical sources. Similar to other social services/public goods, education is affected by globalization with the same characteristics of the new state steering power and capacity. As Dale (2000) rightly points out, globalization does not represent a “new and distinct shift

in the relationship between state and supranational forces” (p. 90); conversely, there is the need for nation-states to respond to global force. Therefore, in the process of formation and implementation of curriculum policy, nation-states act as an agency to merge local traditions and influences with global trends through a process of glocalization (Green, 1999), by which local distinctive features are able to exist side by side in order to provide opportunities for heterogeneity of cultural tradition (Henry *et al.*, 1999: 86). Our above discussion regarding the structural change at the college level is seen to promote more diversity; the change of framework at the national level is developing toward a reversed direction. It is recognized that there are recent curricular trends of leading states to a central control over the curriculum and assessment.

According to Priestley (2002), the governments in the United Kingdom, the United States, Australia, and New Zealand have made attempts to introduce national curriculum since the late 1980s. These initiatives of national curriculum are characterized by a unitary competency-based system for all academic and work-based qualifications. From a national perspective, the centralization of curriculum and assessment is a response to economic globalization, by which a nation-state can enhance its competitiveness through the adoption of the technical-rational curriculum. Such a view is sponsored by the notion of developmental state, in which education is seen as an instrument for nation building. Then, a centralized national curriculum can effectively serve the national needs to respond to the challenges brought by the globalization of economic life and the advance in communication, information, and technology. Nevertheless, this form of centralized governance in education has indeed run contrary to the ethos of the globalization discourse that values diversity but the development of which has clearly demonstrated the tensions between the global forces and local drives and the dialectical push-and-pull nature of globalization.

Conclusion

In this article, we have discussed the wider policy context for driving changes in the university curricula across different parts of the globe. We have also examined the underlying values/guiding principles that have shaped the most recent curriculum reforms and also critically examined what major aspects of curriculum have changed. Before concluding the article, we have had deeper reflections on the role of the state in curriculum reforms, also examining whether the global trends of curriculum changes have promoted more diversity or limited choices in curriculum design. More importantly, we must be sensitive about the cultural and social differences embedded in different education systems and proper contextualization is needed during the processes of policy learning.

Bibliography

- Boud, D. and Symes, C. (2000). Learning for the real: Work-based education in universities. In Solomon, N. and McIntyre, J. (eds.) *Working Knowledge: The New Vocationalism and Higher Education*, pp 14–29. Buckingham: Open University Press.
- Cameron, D. (2002). Globalization and the teaching of ‘communication skills’ In Block, D. and Cameron, D. (eds.) *Globalization and Language Teaching*, pp 67–82. London: Routledge.
- Dale, R. (2000). Globalization: A new world for comparative education. In Schriewer, J. (ed.) *Discourse Formation in Comparative Education*, pp 155–188. Frankfurt: Peter Lang.
- Dale, R. (2005). Globalisation, knowledge economy and comparative education. *Comparative Education* **41**, 117–150.
- Delanty, G. (2001). *Challenging Knowledge: The University in the Knowledge Society*. Buckingham: Open University Press.
- DfES (2004). *Five Year Strategy for Children and Learners*. London: Author.
- George, E. (2006). Positioning higher education for the knowledge based economy. *Higher Education* **52**, 589–610.
- Gough, N. (2002). The long arm(s) of globalization: Transnational imaginaries in curriculum work. In Doll, W. E. and Gough, N. (eds.) *Curriculum Visions*, pp 167–178. New York: Peter Lang.
- Green, A. (1999). Education and globalization in Europe and East Asia: Convergent and divergent trends. *Journal of Education Policy* **14**, 55–71.
- Haigh, M. J. (2002). Internationalisation of the curriculum: Designing inclusive education for a small world. *Journal of Geography in Higher Education* **26**, 49–66.
- Henry, M., Lingard, B., Rizvi, F., and Taylor, S. (1999). Working with/against globalization in education. *Journal of Education Policy* **14**, 85–97.
- Kelsey, J. (1998). Privatising the universities. *Journal of Law and Society* **25**, 51–70.
- Kitano, M. K. (1997). What a course will look like after multicultural change. In Morey, A. I. and Kitano, M. K. (eds.) *Multicultural Course Transformation in Higher Education: A Broader Truth*, pp 18–34. Boston, MA: Allyn and Bacon.
- Knight, J. (2006). Crossborder education and GATS: Risks and opportunities. *Paper Presented at the International Forum on Vietnam Education, WTO Entry and Vietnam Higher Education Reform*, 11–12 December. Hanoi, Vietnam.
- Leask, B. (1999). Bridging the gap – internationalising university curricula. *Paper Presented at NLC National Education Conference, International Education: Bridging the Gap between Vision and Reality*. July, Sydney, Australia.
- Lloyd, C. and Payne, J. (2003). What is a ‘high skills society’? Some reflections on current academic and policy debates in the UK. *Policy Studies* **24**, 115–133.
- McFadden, J., Merryfield, M. M., and Barron, K. R. (1997). *Multicultural and Global Education: Guidelines for Programs in Teacher Education*. Washington, DC: American Association of Colleges for Teacher Education.
- Morey, A. I. (2000). Changing higher education curricula for a global and multicultural world. *Higher Education in Europe* **25**, 25–39.
- OECD (Organisation for Economic Co-Operation and Development) (1999). *The Knowledge-Based Economy: A Set of Facts and Figures*. Paris: Author.
- Peters, M. (2001). National education policy constructions of the ‘knowledge economy’: Towards a critique. *Journal of Educational Enquiry* **2**, 1–22.
- Priestley, M. (2002). Global discourses and national reconstruction: The impact of globalization on curriculum policy. *Curriculum Policy* **13**, 121–138.
- Riddell, A. (1996). Globalisation: Emasculation or opportunity for educational planning? *World Development* **24**, 1357–1372.
- Robertson, S. (2005). Re-imaging and rescripting the future of education: Global knowledge economy discourses and the challenge to education systems. *Comparative Education* **41**, 151–170.
- Sadiki, L. (1998). *Internationalising the Curriculum in the 21st Century*. Canberra: Australia National University.
- Sadker, M. and Sadker, D. (1992). Ensuring equitable participation in college classes. In Border, L. L. B. and Chism, N. V. V. (eds.)

- Teaching for Diversity. New Directions for Teaching and Learning*, No. 49, pp 49–55. San Francisco, CA: Jossey-Bass.
- Toffler, A. (1990). *Powershift: Knowledge, Wealth and Violence at the Edge of the 21st Century*. New York: Bantam Books.
- UNESCO (United Nations Educational, Scientific and Cultural Organization) (2000). *Globalization and Living Together: The Challenge for Educational Content in Asia*. Paris: Author.
- Vidovich, L. (2004). Towards internationalizing the curriculum in a context of globalization: Comparing policy processes in two settings. *Compare* **34**, 443–461.
- World Bank (2003). *Lifelong Learning for a Global Knowledge Economy*. Washington, DC: Author.
- Goodson, I. F. (1990). “Nations at risk” and “national curriculum”: Ideology and identity. In Fuhrman, S. H. and Malen, B. (eds.) *The Politics of Curriculum and Testing: The 1990 Yearbook of the Politics of Education Association*, pp 219–232. New York: Palmer.
- Kenway, J., Bullen, E., and Robb, S. (2004). The knowledge economy: The techno-preneur and the problematic future of the university. *Policy Futures in Education* **2**, 330–349.
- Ledwith, S. and Seymour, D. (2001). Home and away: Preparing students for multicultural management. *International Journal of Human Resource Management* **12**, 1292–1312.
- Peters, M. and May, T. (2004). Universities, regional policy and the knowledge economy. *Policy Futures in Education* **2**, 261–277.
- Stubbs, M. (2006). On teaching critical rationalism: Reconciling linguistic and literacy text analysis. In Gerbig, A. and Müller-Wood, A. (eds.) *How Globalization Affects the Teaching of English: Studying Culture through Texts*, pp 15–30. Lewiston, NY: Edwin Mellen.
- Turner, Y. (2006). Chinese students in UK business school: Hearing the student voice in reflective teaching and learning practice. *Higher Education Quarterly* **60**, 27–51.
- World Bank (2006). *An East Asian Renaissance: Ideas for Economic Growth*. Washington, DC: Author.

Further Reading

- Bates, T. (2005). *Technology, E-Learning and Distance Education*. London: Routledge.
- Fugate, D. L. and Jefferson, R. W. (2001). Preparing globalization – do we need structural change for our academic programs? *Journal of Education for Business* January/February, 160–166.

Curriculum Development as Subject Matter: Social Studies

J Brophy[†]

J Alleman, Michigan State University, East Lansing, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

The Social Studies Curriculum Strand

In addition to basic skills, schooling typically includes two broad content areas, addressing the physical world (science) and the social world. In some countries, the latter strand is organized as social studies; in others, it consists of separate courses in history, geography, the social sciences, cultural studies, religion, or philosophy. In the United States, the National Council for the Social Studies (NCSS) and most leaders in the field think of social studies (singular) as a coherent K-12 subject organized to prepare young people for citizenship. However, some discipline-based organizations and leaders view social studies (plural) as an umbrella term for courses in history, geography, and the social sciences.

The emergence of social studies in the United States is credited to a committee report issued by the National Education Association (1916), calling for incorporating content from history, geography, and civics within a social education strand to be called social studies. Its content would be selected based on its meaning and relevance to students and its value in preparing them for citizenship. This vision is still emphasized by social studies educators. For example, the NCSS defined social studies as “the integrated study of the social sciences and humanities to promote civic competence,” adding that its primary purpose is “to help young people develop the ability to make informed and reasoned decisions for the public good as citizens of a culturally diverse, democratic society in an interdependent world” (National Council for Social Studies, 1994: 3).

Elementary social studies (grades K-6) developed along the lines envisioned in the 1916 report. The curriculum drew from history, geography, civics, and economics, and later from sociology, anthropology, and psychology. The content was taught as integrated social studies, rather than as courses in the academic disciplines. Gradually, the expanding communities sequence became the dominant structuring framework. Also known as the expanding horizons or the expanding environments sequence, it begins with the self in kindergarten and then expands the purview to the family and school in first grade, the neighborhood in second grade, the community in third grade, the state and region in fourth grade, the nation in fifth grade, and the hemisphere or world in sixth grade.

Secondary courses (grades 7–12) are also taught within a social studies strand that includes responsibility for preparing students for citizenship. However, most secondary

courses are school-subject versions of history, geography, or one of the social sciences, in which content is addressed primarily within the single discipline rather than through integrated treatment of topics. The pattern for these grades is shown in Table 1.

There have been exceptions to this general trend: courses in law-related education, global education, peace and environmental studies, and conflict resolution; mini-courses on a variety of topics; and a 12th-grade problems of democracy course intended as a capstone for citizenship preparation. However, most secondary courses feature titles such as US history, economics, or American government.

Social studies educators hold contrasting definitions of citizen education and assumptions about how to accomplish it. Some differences are linked to curricular tensions observable in all subjects; some reflect issues especially salient to social studies; and some reflect competition for curriculum space among disciplinary and special-interest groups within social studies.

Curricular Tensions That Cut Across Subjects

Curriculum debates in all subjects reflect continuing struggles among supporters of four competing ideas about what should be the primary basis for K-12 education (Kliebard, 2004). The first group believes that schools should equip students with knowledge that is lasting and important. It looks to the academic disciplines as storehouses of knowledge and sources of authority on how to organize and teach it. The second group believes that the natural course of child/adolescent development should be the basis for curriculum planning. It would key content to the interests and learning needs associated with each grade level's ages and stages. The third group works backward from society's needs, designing schooling to prepare children and adolescents to fulfill adult roles in the society. The fourth group seeks to use the schools to combat injustice and promote social change, by focusing curriculum around social policy issues. Many curricular debates in social studies reflect ongoing competition among these four generic approaches to curriculum development.

Competing Approaches to Social Studies

Contrasting perspectives on citizen education lead to disagreements about how to prepare students for responsible

[†] Deceased

Table 1 Social studies courses commonly taught in grades 7–12

Grade	Courses
7	World or state history and geography
8	US history and government
9	Civics, world geography, or state history
10	World history or modern history
11	US history
12	Elective courses in sociology, government, economics, or psychology.

Table 2 Five approaches to social studies

<i>Social studies should be taught as</i>	<i>Citizenship education should consist of</i>
1. Transmission of the cultural heritage	Transmitting traditional knowledge and values as a framework for making decisions
2. Social science	Mastering social science concepts, generalizations, and processes to build a knowledge base for decision making
3. Reflective inquiry	Developing reflective thinking and inquiry dispositions through applications to social and civic problem solving and decision making
4. Informed social criticism	Providing opportunities for examination, critique, and revision of past traditions, existing social practices, and modes of problem solving
5. Personal development	Developing a positive self-concept and a strong sense of personal efficacy

citizenship. Five alternatives are discussed in **Table 2** (Martorella *et al.*, 2005; Vinson, 1998).

Transmission of the cultural heritage, the mainstream approach, emphasizes didactic teaching of content that features support for the *status quo*, focus on Western civilization, and uncritical celebration of and inculcation in American political values and traditions. Periodically, it is challenged by historians and social scientists who want preservation of the integrity of their disciplines in the form of separate courses, or by social reformers who want more inquiry into and discussion of social issues with emphasis on critical thinking, values analysis, and decision making.

Some social studies educators have called for additional emphases. Berman and LeFarge (1993) encouraged lessons in cooperation, conflict resolution, and respect for diversity. Hess (2001) suggested discussions that focus on divisive issues that students face. Wade (2001) suggested that citizenship is developed most powerfully through service learning.

To plan good social studies programs, teachers need to clarify their priorities concerning social education goals and their implications for curriculum. Most curricular models share commitment to citizen education goals, making it possible to construct a richer, yet still coherent, curriculum by incorporating multiple perspectives. A unit on government, for example, might incorporate all five: core democratic values (citizenship transmission), the three branches of government (political science), what it means to be an informed voter (reflective inquiry), rights and responsibilities of citizens in a democracy (social criticism component focused on whether there is justice for all in America), and the idea that government cannot do everything, so citizens should consider volunteering to contribute to society (personal efficacy).

Competition for Curriculum Space among Interest Groups within Social Studies

Organizations representing history, geography, and the (separate) social sciences have begun to issue policy statements concerning how their respective disciplines should be taught in K–12. These statements contain helpful summaries of powerful content and suggestions about learning activities. However, they also imply that each discipline ought to be taught much more extensively, usually without specifying what might be reduced to create the required curriculum space. Some advocates of history have gone to the extent of calling for a return to a social studies curriculum focused primarily on history. They argue that history is “the grand integrator” that allows for comprehensive coverage of each topic – not only its historical aspects but also its geographical, civic, cultural, economic, and social aspects (Bradley Commission on History in Schools, 1988; California State Department of Education, 1997). However, arguments for the other disciplines are just as compelling as those for history; the limited geographical and social science content that would be integrated within a history curriculum is only a small proportion of the content that is worth teaching; embedding it within history courses would not allow for systematic teaching of the other disciplines’ basic principles and processes; such integration efforts usually either advance the agenda of one subject at the expense of the other or result in content and activities that trivialize both subjects (Alleman and Brophy, 1993); knowledge about the past has limited application to the contemporary world (Engle and Ochoa, 1988; Evans, 2004); and replacing the pandisciplinary social studies curriculum with all history courses makes no more sense than replacing the science curriculum with all biology courses.

In addition, the history-centered proposals are associated with a politically conservative philosophy of schooling that emphasizes inculcation in a highly traditional

selection of knowledge and values (Cornbleth and Waugh, 1999; Nash *et al.*, 1997; Symcox, 2002). Most social educators prefer curricula that are more global and multicultural in purview, more critical of traditions, and more focused on current and future issues than on the past. Spillover from the culture wars that developed in the United States in recent decades has heightened the controversies surrounding these curricular issues in social studies.

NCSS Standards

The National Council for the Social Studies (1994) published standards for K-12 schools, after gathering input from social studies educators, scholars in the academic disciplines, and the general public. They provide a framework that includes ten thematic strands; performance expectations regarding essential knowledge, processes, and attitudes; and examples of classroom practice. The ten themes are:

1. culture;
2. time, continuity, and change;
3. people, places, and environments;
4. individual development and identity;
5. individuals, groups, and institutions;
6. power, authority, and governance;
7. production, distribution, and consumption;
8. science, technology, and society;
9. global connections; and
10. civic ideals and practices.

Basal Textbook Series

The curriculum enacted in most classrooms is influenced primarily by the basal textbook series produced by the major publishers. Some teachers use textbooks rarely if at all and others use them mostly as reference sources (Haas and Laughlin, 2001), but most use them extensively (Goodlad, 1984; Shaver *et al.*, 1979; Zhao and Hoge, 2005). The textbook publishers are highly responsive to state adoption guidelines and other market pressures. Consequently, their textbook series mainly reinforce the *status quo* by providing materials for teaching the traditional curriculum (Mehlinger, 1992). There are few alternatives available for educators who wish to experiment with different curricula or even vary the grade levels at which certain subjects are studied (Sewall, 2005).

Critiques of the textbook series commonly conclude that they feature bland, boring content that avoids controversy and seldom alerts students to multiple perspectives; try to cover too much, so the treatment of most topics is sketchy (the mile-wide, inch-deep problem); include a lot of trivia that occupies space that could

have been used to develop big ideas; isolate skills content from knowledge content instead of integrating them; are difficult for students to use as learning resources because of poor content coherence and lack of connections to prior knowledge; and offer suggested questions, activities, and test items that are mostly trite or superficial (Beck and McKeown, 1988; Beck *et al.*, 1989; Brophy, 1992; Brophy and Alleman, 1992/1993; Elliott and Woodward, 1990; Larkins *et al.*, 1987; Tyson-Bernstein, 1988).

Such textbooks lead to instruction that emphasizes memorizing of miscellaneous facts. As a result, students traditionally have rated social studies as their least favorite subject (Shaughnessy and Haladyna, 1985). However, it has become more popular lately (Chiodo and Byford, 2004; Pahl, 1995), especially with minority students, as it has become more multicultural and otherwise relevant. Overreliance on textbooks now is less necessary than ever, given a proliferation of good children's literature and Internet websites on social education themes (Berson, *et al.*, 2004; Shiveley and VanFossen, 2001).

Conclusion

Good curriculum development in social studies requires clarity about purposes and goals and systematic follow-up in constructing the program accordingly. To facilitate teaching for understanding, appreciation, and life application, the curriculum needs to be represented as networks of connected content structured around powerful ideas and taught with emphasis on their connections and applications. This implies reducing breadth of coverage to develop powerful ideas in depth, which is difficult to do using contemporary textbooks, especially given the recent imposition of high-stakes testing programs.

Bibliography

- Alleman, J. and Brophy, J. (1993). Is curriculum integration a boon or a threat to social studies? *Social Education* 57(6), 289–291.
- Beck, I. and McKeown, M. (1988). Toward meaningful accounts in history texts for young learners. *Educational Researcher* 17(6), 31–39.
- Beck, I., McKeown, M., and Gromoll, E. (1989). Learning from social studies texts. *Cognition and Instruction* 6, 99–158.
- Berman, S. and LeFarge, P. (eds.) (1993). *Promising Practices in Teaching Social Responsibility*. Albany, NY: State University of New York Press.
- Berson, M., Cruz, B., Duplass, J., and Johnston, H. (2004). *Social Studies on the Internet*, 2nd edn. Columbus, OH: Merrill-Prentice-Hall.
- Bradley Commission on History in Schools (1988). *Building a History Curriculum: Guidelines for Teaching History in Schools*. Washington, DC: Educational Excellence Network.
- Brophy, J. (1992). The *de facto* national curriculum in U.S. elementary social studies: Critique of a representative example. *Journal of Curriculum Studies* 24, 401–447.
- Brophy, J. and Alleman, J. (1992/1993). Elementary social studies textbooks. *Publishing Research Quarterly* 8(4), 12–22.

- California State Department of Education (1997). *History–Social Science Framework for California Public Schools: Kindergarten through Grade 12*. Sacramento, CA: California State Department of Education.
- Chiodo, J. and Byford, J. (2004). Do they really dislike social studies? A study of middle school and high school students. *Journal of Social Studies Research* **28**(1), 16–26.
- Combleth, C. and Waugh, D. (1999). *The Great Speckled Bird: Multicultural Politics and Education Policymaking*. Mahwah, NJ: Erlbaum.
- Elliott, D. and Woodward, A. (eds.) (1990). *Textbooks and Schooling in the United States (Eighty-Ninth Yearbook of the National Society for the Study of Education, Part 1)*. Chicago: University of Chicago Press.
- Engle, S. and Ochoa, A. (1988). *Education for Democratic Citizenship: Decision Making in the Social Studies*. New York: Teachers College Press.
- Evans, R. (2004). *The Social Studies Wars: What Should We Teach the Children?* New York: Teachers College Press.
- Goodlad, J. (1984). *A Place Called School*. New York: McGraw-Hill.
- Haas, M. and Laughlin, M. (2001). Elementary education: A profile of elementary social studies teachers and their classrooms. *Social Education* **65**, 122–126.
- Hess, D. (2001). *Teaching Students to Discuss Controversial Public Issues*. Bloomington, IN: ERIC Clearinghouse for Social Studies/Social Science Education.
- Kliebard, H. (2004). *The Struggle for the American Curriculum 1893–1958*, 3rd edn. New York: Routledge.
- Larkins, A., Hawkins, M., and Gilmore, A. (1987). Trivial and noninformative content of elementary social studies: A review of primary texts in four series. *Theory and Research in Social Education* **15**, 299–311.
- Martorella, P., Beal, C., and Bolick, C. (2005). *Teaching Social Studies in Middle and Secondary Schools*, 4th edn. Upper Saddle River, NJ: Prentice Hall.
- Mehlinger, H. (1992). The national commission on social studies in the schools: An example of the politics of curriculum reform in the United States. *Social Education* **56**, 149–153.
- Nash, G., Crabtree, C., and Dunn, R. (1997). *History on Trial: Culture Wars and the Teaching of the Past*. New York: Knopf.
- National Council for Social Studies (1994). *Expectations of Excellence: Curriculum Standards for Social Studies*. Washington, DC: National Council for Social Studies.
- National Education Association (1916). *Social Studies in Secondary Education: A Six-Year Program Adapted to the 6–3–3 and the 8–4 Plans of Organization* (Bulletin No. 28). Washington, DC: US Department of the Interior, Bureau of Education.
- Pahl, R. (1995). Do students prefer social studies, science or English? *International Journal of Social Education* **9**, 53–56.
- Sewall, G. (2005). Textbook publishing. *Phi Delta Kappan* **86**, 498–503.
- Shaughnessy, J. and Haladyna, T. (1985). Research on student attitude toward social studies. *Social Education* **49**, 692–695.
- Shaver, J., Davis, O. Jr., and Hepburn, S. (1979). The status of social studies education: Impressions from three NSF studies. *Social Education* **39**, 150–153.
- Shiveley, J. and VanFossen, P. (2001). *Using Internet Primary Sources to Teach Critical Thinking Skills in Government, Economics, and Contemporary World Issues*. Westport, CT: Greenwood Press.
- Symcox, L. (2002). *Whose History? The Struggle for National Standards in American Classrooms*. New York: Teachers College Press.
- Tyson-Bernstein, H. (1988). *A Conspiracy of Good Intentions: America's Textbook Fiasco*. Washington, DC: Council for Basic Education.
- Vinson, K. (1998). The “traditions” revisited: Instructional approach and high school social studies teachers. *Theory and Research in Social Education* **26**(1), 50–82.
- Wade, R. (2001). Social action in the social studies: From the ideal to the real. *Theory into Practice* **40**, 23–28.
- Zhao, Y. and Hoge, J. (2005). What elementary students and teachers say about social studies. *Social Studies* **96**, 216–221.

Further Reading

- Barton, K. and Levstik, L. (2004). *Teaching History for the Common Good*. Mahwah, NJ: Erlbaum.
- Brophy, J. and Alleman, J. (2006). A reconceptualized rationale for elementary social studies. *Theory and Research in Social Education* **34**, 428–454.
- Center for Civic Education (1994). *National Standards for Civics and Government*. Calabasas, CA: Center for Civic Education.
- Evans, R. and Saxe, D. (1996). *Handbook on Teaching Social Issues*. Washington, DC: National Council for the Social Studies.
- Jenness, D. (1990). *Making Sense of Social Studies*. New York: Macmillan.
- Levstik, L. and Tyson, C. (eds.) (2008). *Handbook of Research in Social Education*. New York: Routledge.
- National Center for History in the Schools (1996). *National Standards for History: Basic Education*. Los Angeles, CA: National Center for History in the Schools.
- National Council on Economic Education (1995). *A Framework for Teaching Basic Economic Concepts with Scope and Sequence Guidelines, K–12*. New York: National Council on Economic Education.
- People for the American Way (1986). *Looking at History: A Major Review of U.S. History Textbooks*. Washington, DC: People for the American Way.
- Shaver, J. (ed.) (1991). *Handbook of Research on Social Studies Teaching and Learning*. New York: Macmillan.
- The Geography Education Standards Project (1994). *Geography for Life: National Geography Standards*. Washington, DC: National Geographic Society Committee on Research and Exploration.
- Thornton, S. (2005). *Teaching Social Studies that Matters: Curriculum for Active Teaching*. New York: Teachers College Press.

Curriculum Development in the Area of Reading

C L Madda and T E Raphael, University of Illinois, Chicago, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Success in reading is accepted as a major determinant for students' achievement in school across disciplinary areas, and out-of-school in terms of personal, social, and economic success. Conventional descriptions of the reading curriculum content generally focus on basic skills and strategies readers need to make sense of words on a page (e.g., decoding skills, word knowledge, and comprehension strategies). In the United States, the reading curriculum, when defined in this way, is often equated to a commercially prepared program consisting of a teachers' guide that accompanies the students' texts, activity books, and assessments (such as a basal reading program). These programs prescribe what to teach, the order in which to teach it, texts for instruction and practice, contexts for applying what is taught, and assessments to measure students' progress. Often, both curricular content and instructional delivery are linked to national and state standards that establish generic end-of-year outcomes for students in each grade level and for related standardized tests.

It is important to expand conceptions of the reading curriculum beyond those of prepackaged program materials to understanding reading curriculum to be a developmental process that relies on the professional expertise of multiple stakeholders and is designed to address the diversity of students' needs. This article first describes traditional content central to a reading curriculum and typical performance standards. Second, it identifies issues that complicate traditional curriculum content in light of globalization, increased diversity across students and schools, and preparing students for the technological advances that affect living and working in the twenty-first century. Third, it describes principles underlying the process of reading curriculum construction designed to support today's linguistically, economically, and ethnically diverse learners.

Defining the Reading Curriculum: Traditional Content

Scholars have identified well-established reading curricular components with compelling research bases in key areas of comprehension, word knowledge, decoding, and genre study. Strategies and skills within these traditional components provide readers with the foundation to engage in higher-level thinking and meaning making, the long-standing goal of a quality reading curriculum.

To begin with, decades of research in comprehension and comprehension instruction detail both the processes good readers use and the ways in which teachers facilitate students' learning of these processes, through a model of instruction that moves from explicit teaching through guiding, coaching, and students' independent application (Pearson, 1985). Scholars have documented that good readers draw on prior reading and experiences to form hypotheses about upcoming text content, revise their ideas as they read, monitor their understanding, and use strategies as needed for meaning making (Pressley and Afflerbach, 1995).

Strategies taught within the reading curriculum support readers' predicting, identifying important information, summarizing, making inferences, questioning, and monitoring (Snow, 2003). Predicting is the readers' ability to draw from their background knowledge, text features, and structures to form hypotheses about upcoming information. Identifying important information requires locating the central idea in a text section and distinguishing it from supporting details. Summarizing is the ability to synthesize and condense the important information into new text. Making inferences involves reading beyond the words on the page – noticing authors' clues that require adding information not explicitly stated. Questioning requires that readers use their own knowledge in combination with information from the text to guide their reading. Monitoring refers to the reader's ability to evaluate their understanding of the text and apply strategies when comprehension breaks down. Each category is comprised of a variety of potentially useful strategies, depicted in **Table 1** (Raphael *et al.*, 2006).

Skilled readers apply these reading comprehension strategies in combination before, during, and after reading. Students' mastery of strategy use is then evidenced by a capacity to use them flexibly and, therefore, effectively based on different purposes for reading. Frameworks for instruction can promote students' fluid and efficient application of multiple strategies. Such frameworks serve as a vehicle for teaching this part of the reading curriculum, however, not as the curriculum itself.

Word knowledge, with its strong relationship to comprehension, is sometimes called the cornerstone of literacy (Beck and McKeown, 1985: 12) and is the second fundamental component of the reading curriculum. While scholars have attempted to identify a best method for vocabulary instruction, it has proved to be elusive. In many cases, studies that claim the superior effectiveness

Table 1 Comprehension and sample strategies

<i>Comprehension category</i>	<i>Selected sample strategies</i>
Predicting: creating a hypothesis based on background knowledge, text features, and text structure about upcoming information in a text.	<p>Hypothesizing about what the author might discuss next in the text</p> <p>Using relevant sources of information for predicting, such as</p> <ul style="list-style-type: none"> • Background knowledge the reader already possesses • Text features such as titles, headings, and embedded questions • Text structures – how the text is organized <p>Setting a purpose for reading: confirming content or disproving hypotheses (connects to monitoring if prediction is not confirmed – does it indicate comprehension block or is the author successfully surprising the reader?)</p> <p>Engaging in appropriate routines such as taking a picture walk or a book walk or using knowledge of parts of the text – index, title, table of contents, and illustrations – to predict</p>
Identifying important information: Identifying the superordinate, driving idea in a section of text; distinguishing this superordinate idea from the details that elaborate on it.	<ul style="list-style-type: none"> • Recognizing that some ideas are more important than others in a passage • Distinguishing between main ideas and supporting details • Identifying key story elements • Underlining or highlighting the sentence or phrase that captures the important information presented in a paragraph • Understanding the gist or overall topic • Identifying key words or phrases • Making use of text structure to identify key ideas • Distinguishing between author-determined versus reader-determined importance
Summarizing: creating a new, succinct text that encompasses the important information in the section of text being summarized	<ul style="list-style-type: none"> • Making use of story structure to highlight key points in a narrative • Composing a brief, new text based on important ideas already identified • Synthesizing text ideas in a succinct and coherent fashion • Integrating material into a coherent, accurate representation (such as a graphic organizer) • Distinguishing between writer-based and reader-based summaries • Selecting key information and deleting unnecessary or redundant material • Condensing some material • Substituting superordinate concepts (e.g., wild animals for lions, tigers, and bears) • Composing a topic sentence • Composing a thesis statement • Distinguishing between summarizing and retelling
Making Inferences: reading between the lines to add information not explicitly stated by the author but needed to make sense of the text.	<ul style="list-style-type: none"> • Reading more than the words, or between the lines • Accessing prior knowledge triggered by information or words provided by the author • Seeing connections among text ideas, when those connections have not been explicitly stated by the author • Making text-to-self, text-to-text, and text-to-world connections • Visualizing or creating mental images based on the text • Thinking of contexts in which the information learned might be useful • Imagining how a character might feel • Identifying with a character • Interpreting the author's message • Creating a theme or thesis for the text • Distinguishing between literal and figurative meanings
Questioning: using knowledge for answering and asking questions, including creating relevant questions to guide reading for explicit and implicit text information, drawing on both the reader's knowledge base and information presented by the text.	<ul style="list-style-type: none"> • Setting purposes or goals for reading • Clarifying confusions or confusing information • Determining the author's position or intention in writing the piece (this begins to get at critical thinking and evaluation) • Distinguishing among types of questions and knowing when to ask what kind (e.g., eliciting factual and explicit information; inferences that can be made with text information; critical and evaluative text interpretation; author's craft and style; and relationships among texts) • Understanding sources of information for creating questions and, thus, better understanding where appropriate information is found for answering questions
Monitoring: evaluating text understanding and using fix-up strategies appropriately in the face of comprehension difficulties.	<ul style="list-style-type: none"> • Evaluating understanding and appropriateness of the strategies used to learn from text • Establishing goals and consciously evaluating the degree to which the goals are being met • Modifying strategies when necessary to read goals (i.e., having a repertoire of strategies to draw on in appropriate circumstances) • Determining whether or not the text read makes sense based on expectations (can get at this by paraphrasing, checking predictions against current understanding of text information, and clarifying)

Continued

Table 1 Continued

<i>Comprehension category</i>	<i>Selected sample strategies</i>
	<ul style="list-style-type: none"> • Knowing reasons why text might not have made sense (e.g., unfamiliar vocabulary, awkward structure, unclear referent words, and idiomatic expressions) • Key strategies for teaching monitoring include asking questions before, during, and after reading; working collaboratively to answer questions because students must clarify their understandings when explaining to another student

Reprinted from Raphael, T. E., Highfield, K., and Au, K. H. (2006). *QAR Now*. New York: Scholastic. With kind permission, from READ.

of a particular approach are often contradicted by other studies with conflicting findings. It is likely that situation-dependent contexts exist for vocabulary instruction, varying in best method depending upon students' backgrounds, targeted words, goals for instruction, and both school and curricular contexts.

Supporting the components of higher-level thinking are lower-level processes. The third component of the reading curriculum contributes to students' ability to master language conventions. Several decades of research have established students' understanding of the linguistic base or symbol-sound relationships as necessary for early reading achievement in particular (e.g., National Reading Panel, 2000). These initial steps toward decoding and word recognition are part of the foundation for student's understanding of how language systems work. Phonological awareness (such as phonemic awareness, blending, and segmentation) serves as a building block for beginning readers working within an alphabetic system (Goswami, 2000; National Reading Panel, 2000). However, research findings indicate that the value of phonics instruction in the reading curriculum diminishes when applied in the intermediate and upper grades.

Research has demonstrated a range of effective methods for instruction in teaching this knowledge base of the language code. Ehri (1995) describes four approaches to helping young students learn to read words: sequential decoding, analogy, contextual analysis, and sight-word recognition. Sequential decoding consists of sounding words out letter by letter. Analogy teaches children to look for spelling patterns (known as phonograms or rimes) in words that are often used to characterize word families – or groups of words that look and sound similar. Contextual analysis applies to words not readily solved by methods of sequential decoding or analogy. In cases such as these, one must consider the morphology of the word in addition to its surrounding context. Sight-word recognition requires students' immediate retrieval of words from memory.

While overwhelming evidence from research has established the value of phonics or word-solving strategies in early reading achievement, lower-level emphases must not overshadow the more meaning-based emphases on word study and comprehension. This need to achieve balance often presents the greatest challenge to the reading curriculum. Just as reading must be balanced

with other school subjects in the curriculum, so must the wide body of knowledge be balanced within reading instruction.

The fourth component of the reading curriculum includes knowledge of various text forms or genre knowledge. Historically, the reading curriculum has emphasized learning to read through traditional literary genres such as narrative, contemporary realistic fiction, poetry, and so forth. As a consequence, there has been a dearth of informational text in the reading curriculum overall and, particularly, for young children (Duke, 2000). This stands to disadvantage readers given that the vast majority of mature reading involves informational text.

It is useful, though, to move beyond conventional genres for different underlying reasons. At a pragmatic level, in contexts such as the United States, with the stress on high-stakes testing, students benefit from being knowledgeable about the testing genre – knowing how to read and respond to test formats and various question types. However, beyond the pragmatic, issues of equity and access to a successful future require due consideration to genres of the twenty-first century where reading transcends traditional forms. A twenty-first-century reading curriculum must engage students with multimedia formats that may not only build on traditional curricular content, but also require a distinct set of literacy skills. In the following section, the article (re)considers reading as a part of the literate practices that shape our lives as twenty-first-century citizens, and explores implications for constructing curriculum.

Reading Curriculum in the Twenty-First Century

Life today represents a time of unprecedented diversity, complexity, and dynamic change that requires an ecological approach to literacy education (Luke and Elkins, 2000), and a redefining and broadening of the reading curriculum. Readers are not passive receivers of the printed word. Rather, skilled readers actively construct meaning, using many of the same processes that writers do as they create text. Further, readers today – even those reading conventional texts – must be facile at reading various images in addition to print. Reading today

involves meaning making from print and electronic media, through words as well as visual images. Schools must therefore prepare students to become literate citizens of the twenty-first century, using all the tools, texts, and systems that these citizens will encounter. School reading curricula must account for “the context of our culturally and linguistically diverse and increasingly globalized societies, the multifarious cultures that interrelate, and the plurality of texts that circulate” (New London Group, 1996: 61), that is the social, political, and economic contexts, texts, and readers of this century.

Changing Contexts

The implications of changing social, political, and economic contexts for the teaching and learning of students impact classrooms, pedagogy, and curriculum alike. For example, the workplace has been redefined by globalization and the increasing presence of new technologies. In many of today’s industrialized economies and institutions, individuals are “faced with the challenges of navigating community and academic discourses and of balancing forms of identity, new work practices, and the demands of new technologies and popular cultures” (Luke and Elkins, 2000: 396). In this “global, hypercompetitive, science-and-technology-driven capitalism, products and services are created, perfected, and changed at ever faster rates” (Gee, 2000: 414). What was once a stable, labor-driven economy has morphed into one that is fluid, market driven, knowledge based, and geographically dispersed (Bean and Readence, 2002). This new order of today’s workplace demands a labor force that is proficient in multiple and overlapping forms of literacies.

Changing Texts and Literacies

“In the emerging digital economic era, spurred by the recent proliferation of technology tools and resources in the form of affordable desktop computers, an accessible Internet, and user-friendly multitask and multimedia software, traditional concepts of what it means to be and to become literate are being challenged” (Labbo *et al.*, 1998: 273–274). Lankshear and Knobel (2003) describe new literacies as those “mediated by new information and communications technologies, as well as literacies that employ more conventional tools and forms of text” (p. 18). New literacies include weblogs, console games, instant messaging, and chatting online, to name a few. The presence of new literacies as well as shifting ideologies, economies, and demographics points toward a need to reconceptualize learners, the reading behaviors they engage in now and in the future, and the texts with which they interact.

Successful reading for the twenty-first-century global citizen requires two important added considerations to traditional reading curricula. First, the curriculum should

promote the reader’s ability to effectively navigate and construct meaning across a variety of text modes. For example, digital or multimedia text types often require navigating visual, spatial, audio, and printed forms. Such multimodal texts have implications for the reading strategies and skills our students need to be proficient in, and raise questions about the range of texts (or genres) students need to be able to control to achieve different purposes (Pearson *et al.*, 2007). Second, the curriculum must address the need for critical analyses of all texts, regardless of modality. The proliferation of texts and new text forms that characterizes today’s Information Age calls for the ability to navigate an abundance of resources and to identify those most credible or reliable. Critical literacy skills should also be promoted so that students can come to recognize the cultural and ideological assumptions embedded in texts that aim to position them in the world.

Changing Readers

The experiences of the twenty-first-century student are unique in that they have grown up in a world of unmatched technological growth, exposure to multimedia, and globalization. As a result of our changing global context and the range of text forms that now occupy our literate lives, it is necessary to reconceptualize the readers themselves that fill our classrooms. Diversity among today’s students includes the ethnic and racial background, family economic status, and linguistic resources that students bring to the task of reading. The curriculum must then account for this diversity in experience. A one-size-fits-all approach to curriculum is not only impractical, but would also be unjust. Curricular decisions must draw on the strengths and resources that students bring to the classroom in order to best support their specific needs. In addition, globalization and shifting world demographics bring new and increased access to education and corresponding issues of equity. In light of changing contexts, it is critical that curricula foreshadow the texts and tasks that students will be expected to engage in throughout their adult lives. To prepare students for anything less would be inequitable. Schooling and curriculum, in particular, need to do a better job of meeting the real needs of today’s students. This requires a process of curriculum development that involves multiple stakeholders and is rooted in the real needs of students. Prepackaged programs not only limit opportunities for constructing a curriculum that reflects the complex nature of reading today, but also lack the flexibility necessary to accommodate a range of student needs. In the following section, the article outlines the underlying principles of curriculum construction that stand to best serve students today.

(Re)Constructing the Reading Curriculum Today

Many general principles for curriculum development established over a half century ago (Tyler, 1949) continue to be relevant today and relevant to a myriad of contexts. At their core, high-quality curricula consist of clearly defined and appropriate outcomes, relevant learning experiences organized in a way that maximizes the impact of the instruction, and a system for evaluating the effectiveness of the curriculum and teaching. A curriculum is not simply a program, though programs can make useful contributions. Ultimately, the stakeholders closest to the clients being served – the teachers who serve the students in their school – must construct the curriculum, and do so in a way that insures they will meet high standards of achievement.

Since the mid-1990s, several successful school–university partnerships have attempted to raise school reading achievement levels through a combination of improving the reading curriculum and providing the related professional development (Taylor, Raphael, and Au, in press). Constructing and implementing the curriculum through the work of the school's professional learning community – a culture of collaboration focused on student learning and effecting change for school improvement (DuFour, 2004) – have shown to be key (e.g., Au, 2005). These approaches are consistent with both Tyler's (1949) long-standing and accepted guidelines for curriculum

construction and more recent research detailing factors that support successful school reform. These studies of school improvement and the theoretical base underlying them suggest that curriculum construction leading to increased student achievement privileges both the top-down voices, represented by high national and state standards, and bottom-up voices, the teachers and other school personnel with the closest knowledge of the students they serve. In what follows, the standards-based change (SBC) process (Au, 2005; Au and Raphael, 2007; Au *et al.*, 2008) is presented as illustrative of current models for principled construction of a customized reading curriculum.

Rather than implementing a solution or program designed by outsiders, the SBC process is adapted by insiders to meet the needs of their own school setting. Through professional development, the SBC process guides educators to customize curriculum designed to raise students' achievement. The SBC process is organized within four concepts illustrated in **Figure 1** that are consistent with Tyler's principles: (1) establishing curricular goals at the school level, (2) establishing grade-level end-of-year goals that contribute to the school graduation goals and making these goals visible to students, (3) determining where students are in relation to end-of-year goals, and (4) planning curriculum and instruction that helps students reach the target performance levels.

These components form an ongoing, iterative cycle of work in the professional learning community that leads to

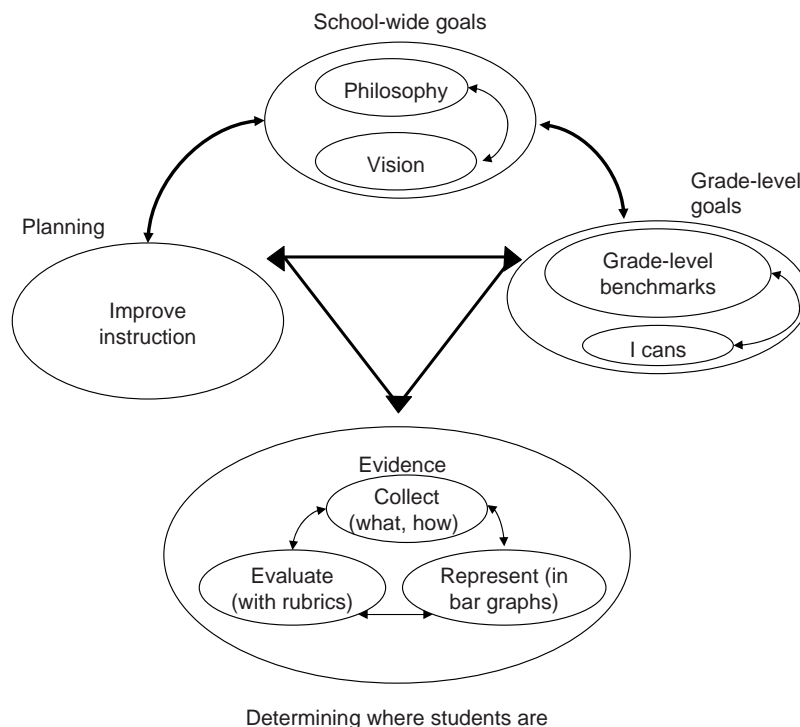


Figure 1 Four concepts of the SBC process. Reproduced from Au, K. H., Raphael, T. E., and Mooney, K. (2008). Improving reading achievement in elementary schools: Guiding change in a time of standards. In Wepner, S. B. and Strickland, D. S. (eds.) *Supervision of reading programs*, 4th ed., pp 71–89. New York: Teachers College Press.

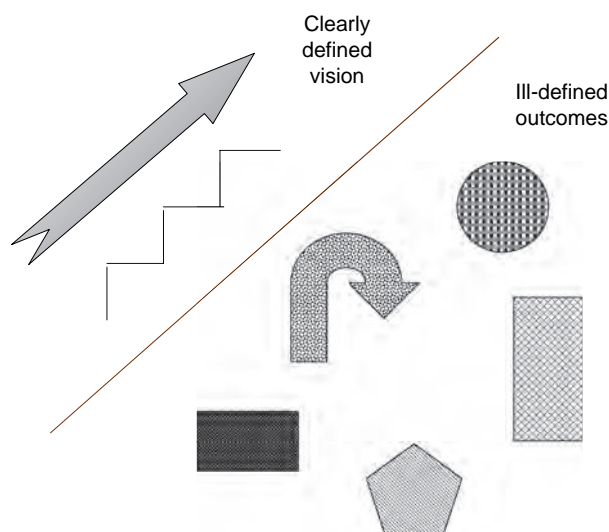


Figure 2 Staircase curriculum versus fragmented curriculum.

the construction of a staircase curriculum, represented in **Figure 2**.

The staircase curriculum (Au, 2005), illustrated on the left side of **Figure 2**, stands in contrast with what is often the case in schools struggling to achieve targeted levels of student performance.

First, while all schools have desired reading outcomes for their students, schools engaged in the SBC process have clearly articulated and defined outcomes that the staircase curriculum helps their students to attain. Second, the curriculum is defined by its coherence (Newmann *et al.*, 2001). In other words, grade-level teams consciously build upon specific instruction of the prior grade level, and work to insure students are prepared adequately for the subsequent one. The staircase curriculum contrasts with a fragmented curriculum, shown in the right side of **Figure 2**, in which teachers' work at different grade levels is not closely coordinated. The quality of the work in each grade level may be individually fine, but there is no schoolwide vision of their graduates' literacy knowledge and skills, thus reducing the possibilities of cumulative learning. The staircase metaphor underscores the importance of constructing curriculum so students receive strong, well-coordinated instruction as they progress through the grades – ultimately leading to rising expectations each year as students enter grade levels at higher levels of progress.

Some educators hope that a staircase curriculum can be achieved by adopting a packaged reading program, such as a basal reading program or a comprehensive reform model. Research suggests, however, that simply adopting a reading program does not lead to schoolwide curriculum coherence (Au, 2005; Newmann *et al.*, 2001) because teachers can choose not to follow the adopted program closely or they can interpret the program in different ways. Constructing a customized reading curriculum using an approach such as

the SBC process, however, involves the schools' professional learning community in four increasingly complex areas of focus: (1) learning the SBC process, (2) institutionalizing schoolwide reports, (3) constructing the school literacy curriculum, and (4) student and family ownership.

Learning the SBC Process

First, the professional learning community members work on learning to implement a series of steps to insure that the four concepts represented in **Figure 1** are in place at their school. To establish a common vision of the reading accomplishments of their graduates, teachers work across grade-level teams to articulate their philosophy of teaching, learning, and literacy, and then construct a vision statement that describes their graduate. A Chicago, Illinois, public school's vision states: the graduating students of South Loop Elementary School (name used with permission) will be lifelong readers who read for both information and enjoyment. They use reading effectively in their everyday lives (Au *et al.*, 2008).

With the school's vision of their reading graduate in place, each grade level determines their contribution to students' achieving the outcome drawing on state and national standards documents, school curriculum scope and sequence documents, and their own professional knowledge and experience. Benchmarks are constructed for each grade level and translated into "I can" statements for accessibility of students and their families, making curriculum goals at each grade level transparent to all stakeholders. The "I can" statements also make the progress on constructing the staircase curriculum easily visible to all professionals in the school and open for analysis and revision as needed. The "I can" statements of grades 1 through 3 constructed by South Loop Elementary School's grade-level teams reflect the increasing expectations as students move through the primary grades:

Grade 1 – I can understand a story I have read and retell it in order using my own words.

Grade 2 – I can understand, make connections to, and summarize stories I read.

Grade 3 – I can tell the difference between fiction and nonfiction books, what is important in a text, make inferences, and summarize what I have read.

With grade-level goals established, teachers then construct an evidence system that can be used at the beginning, middle, and end of the school year to determine where students are, the progress they have made, and success in achieving year-end outcomes, respectively. Performance levels at the beginning and middle of the year are central to the instructional decisions teachers make – the curriculum materials and the instructional approaches (including how to teach, grouping arrangements, and level of instructional materials) that will be used to insure

that students are engaged in appropriate learning experiences. Within the professional learning community, each grade-level team prepares a presentation following the beginning, middle, and end-of-year evidence gathering. The presentation is given at a schoolwide event in which each team describes their goals, how they evaluated students' progress toward their goals (the evidence gathered, the scoring rubric, and illustrative samples of students' work at each level), their analysis of their instructional response to the students' performance levels, and what they plan to do in the near future to address students' needs as indicated by the evidence.

Institutionalizing Schoolwide Reports

Second, once the process of the to-do list is learned, the professional learning community works to institutionalize 3-times-per-year reporting, using the schoolwide event as an opportunity to revisit and revise their staircase curriculum, evaluate the effectiveness of curriculum materials in place and instructional strategies used, and assure that the evidence system is capable of providing the information needed to insure high-quality teaching for high levels of thinking.

Constructing the School Literacy Curriculum

Third, with the institutionalization of the 3-times-a-year schoolwide reporting, the next focus for the professional learning community is to construct their school's reading curriculum more formally. Each grade-level team creates the grade-level curriculum guide to the content of their reading program that lays out the goals, evidence system, instructional materials, and instructional strategies. These guides provide the basis for the curriculum's sustainability over time and in the face of changes in school personnel.

Student and Family Ownership

The fourth focus of the professional learning community builds student ownership and engagement in their own learning processes, as student portfolios are developed to align with the constructed reading curriculum.

As a school engages in this reform process, learning within the professional community deepens and the reading curriculum is kept current as well as continues to be targeted to meet the needs of the students within the school. If maintained, this process can contribute to the upward progression of the school's overall performance levels (Au, 2005). Moreover, while illustrated here with relatively conventional reading goals and content, the process can be applied to any school subject area as well as to teaching twenty-first-century curricula.

Summary

Students' success in life is, in many ways, contingent upon their ability to construct meaning from a variety of texts. Research provides evidence to support the inclusion of several well-established curricular components in key areas of reading such as comprehension, word knowledge, decoding, and genre study. However, in addition, school reading curricula need to be responsive to changes in today's students, texts, and contexts. Twenty-first-century citizenry requires a broadened notion of the types of literate behaviors students will be expected to engage in, and the types of text forms they will be required to navigate with a critical eye. Furthermore, globalization and shifting world demographics contribute to diverse student populations where children often vary in the cultural and linguistic resources that they bring to the task of reading. Prepackaged commercial programs have limited flexibility and adaptability to differentiate the curriculum to meet the needs of the diverse range of students that fill our classrooms. What is more realistic is understanding curriculum construction as a process of development rooted in the professional collaboration of school colleagues and customized to the real literacy needs of real students.

Bibliography

- Au, K. H. (2005). Negotiating the slippery slope: School change and literacy achievement. *Journal of Literacy Research* **37**, 267–288.
- Au, K. H. and Raphael, T. E. (2007). Classroom assessment and standards-based change. In Paratore, J. and McCormack, R. (eds.) *Classroom Literacy Assessment: Making Sense of What Students Know and Do*, pp 306–322. Newark, DE: International Reading Association.
- Au, K. H., Raphael, T. E., and Mooney, K. (2008). Improving reading achievement in elementary schools: Guiding change in a time of standards. In Wepner, S. B. and Strickland, D. S. (eds.) *Supervision of Reading Programs*, 4th edn. New York: Teachers College Press.
- Bean, T. W. and Readence, J. E. (2002). Adolescent literacy: Charting a course for successful futures as lifelong learners. *Reading Research and Instruction* **41**, 203–210.
- Beck, I. and McKeown, M. (1985). Teaching vocabulary: Making the instruction fit the goal. *Educational Perspectives* **23**, 11–15.
- DuFour, R. (2004). What is a "professional learning community"? *Educational Leadership* **61**, 6–11.
- Duke, N. K. (2000). 306 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly* **35**, 202–224.
- Ehri, L. (1995). Phases of development in reading words. *Journal of Research in Reading* **18**, 116–125.
- Gee, J. P. (2000). Teenagers in new times: A new literacy studies perspective. *Journal of Adolescent and Adult Literacy* **43**, 412–420.
- Goswami, U. (2000). Phonological and lexical processes. In Kamil, M., Mosenthal, P., Pearson, P. D., and Barr, R. (eds.) *Handbook of Reading Research*, vol. 3, pp 251–267. Mahwah, NJ: Erlbaum.
- Labbo, L. D., Reinking, D., and McKenna, M. C. (1998). Technology and literacy education in the next century: Exploring the connection between work and schooling. *Peabody Journal of Education* **73**, 273–289.

- Lankshear, C. and Knobel, M. (2003). *New Literacies: Changing Knowledge and Classroom Learning*. Philadelphia, PA: Open University Press.
- Luke, A. and Elkins, J. (2000). Special themed issue: Re/mediating adolescent literacies. *Journal of Adolescent and Adult Literacy* **43**, 396–398.
- National Reading Panel (2000). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. (National Institute of Health Pub. No. 00-4769). Washington, DC: National Institute of Child Health and Human Development.
- New London Group (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review* **66**, 60–92.
- Newmann, F. M., Smith, B., Allensworth, E., and Bryk, A. S. (2001). *School Instructional Program Coherence: Benefits and Challenges. Improving Chicago's Schools [Report]*. Chicago, IL: Consortium on Chicago School Research.
- Pearson, P. D. (1985). Changing the face of reading comprehension instruction. *The Reading Teacher* **38**, 724–738.
- Pearson, P. D., Raphael, T. E., Benson, V., and Madda, C. (2007). Balance in comprehensive literacy instruction. In Gambrell, L., Morrow, L. M., and Pressley, M. (eds.) *Best Practices in Literacy Instruction*, 3rd edn., pp 30–54. New York: Guilford.
- Pressley, M. and Afflerbach, P. (1995). Verbal protocol analysis: The nature of constructively responsive reading. In Ruddell, R. B., Ruddell, M. R., and Singer, H. (eds.) *Theoretical Models and Processes of Reading*, 4th edn., pp 996–1056. Hillsdale, NJ: Erlbaum.
- Raphael, T. E., Highfield, K., and Au, K. H. (2006). *QAR Now*. New York: Scholastic.
- Snow, C. E. (2003). *Reading for Understanding: Toward a Research and Development Program in Reading Comprehension*. Santa Monica, CA: RAND.
- Taylor, B. M., Raphael, T. E., and Au, K. A. (in press). School reform in literacy. In Pearson, P. D., Kamil, M., Afflerbach, P., and Dutrow, E. (eds.) *Handbook of Reading Research*. New York: Routledge/ Taylor Frances Group.
- Tyler, R. (1949). *Basic Principles of Curriculum and Instruction*. Chicago, IL: University of Chicago Press.

Further Reading

- Aainsbury, M., Harrison, C., and Watts, A. (eds.) (2006). *Assessing Reading: From Theories to Classrooms*. Berkshire: National Foundation for Educational Research.
- Barr, R., Kamil, M. L., Mosenthal, P. B., and Pearson, P. D. (eds.) (1984). *Handbook of Reading Research*, vol. 1. Mahwah, NJ: Erlbaum.
- Barr, R., Kamil, M. L., Mosenthal, P. B., and Pearson, P. D. (eds.) (1991). *Handbook of reading research*, vol. 2. Mahwah, NJ: Erlbaum.
- Flood, J., Lapp, D., Squire, J. R., and Jensen, J. M. (eds.) (2003). *Handbook of Research on Teaching the English Language Arts*, 2nd edn. Mahwah, NJ: Erlbaum.
- Israel, S. E. and Duffy, G. G. (eds.) (2008). *Handbook of Research on Reading Comprehension*. New York: Routledge.
- Kress, G. (2001). *The Modes and Media of Contemporary Communication*. New York: Oxford University Press.
- Morrow, L. M., Rueda, R., and Lapp, D. (eds.) (2009). *Handbook of Research on Literacy and Diversity*. New York: Guilford.
- Mosenthal, J. H., Pearson, P. D., and Barr, R. (eds.) (2000). *Handbook of Reading Research*, vol 3. Mahwah, NJ: Erlbaum.
- Muspratt, S., Luke, A., and Freebody, P. (eds.) (2003). *Constructing Critical Literacies*. New York: Hampton Press.
- Pappas, C. and Pettegrew, B. S. (1998). The role of genre in the psycholinguistic guessing game of reading. *Language Arts* **75**, 36–44.
- Pearson, P. D. (2004). The reading wars. *Educational Policy* **18**, 216–252.
- Pinar, W. F. (ed.) (2003). *International Handbook of Curriculum Research*. Mahwah, NJ: Erlbaum.

Relevant Websites

- <http://www.ciera.org> – Book Club Plus: A Conceptual Framework to Organize Literacy Instruction.
- <http://www.ciera.org> – Center for Improvement of Early Reading Achievement.
- <http://www.planetbookclub.com> – Planet Book Club.
- <http://www.SchoolRiseUSA.com> – School Rise: Enlightened Teaching.

Curriculum in the Arts

C M Thompson, The Pennsylvania State University, University Park, PA, USA

L Bresler, University of Illinois at Urbana–Champaign, Champaign, IL, USA

T E Costantino, University of Georgia, Athens, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

Curriculum Development–Subject Matter: Curriculum in the Arts

Arts education in public schools has assumed radically different forms at different historical moments and in different contexts. It has served to cultivate the mind and the spirit, address utilitarian goals in the interests of industrial production, and provide children with emotional outlets and means of self-expression. The cognitive value of the arts has been recognized in the past 50 years, while the role of the arts in cultural experience dominates contemporary thought. Arts curricula are shaped by attitudes toward the arts, children, and education. They are influenced by developments in the informing arts disciplines, including the blurring of boundaries within and between art forms. Contemporary developments remind us of how dramatically arts curricula are shaped by external theoretical influences, political developments, and social conditions. The goals and functions generated in response to shifting sociohistorical circumstances, attitudes, and understandings suggest different rationales and practices for arts curricula. For example, as trends in the art world and conceptions of childhood shift, the values embedded in arts curricula respond in subtle but decisive ways (Wilson, 1997). Preservation of the lush energetic strokes of preschool painting motivated visual art curriculum practice in an era dominated by the philosophies of Lowenfeld (1957), the esthetic of abstract expressionism, and a postwar vision of idyllic childhood fed by Dr. Spock, Dick-and-Jane readers, and the televised utopia of *Leave it to Beaver*. In the sadder, but wiser, era of the 1970s, adolescent marginalia was recognized, valued, and cultivated as an artistic practice similarly in tune with its times.

Given these multiple, layered aspects of the arts, the fundamental nature of arts education in the school is perennially contested. Theories of esthetic education, as well as related theories and changing practices in individual arts disciplines, generate aspirations and justifications that scaffold practice. Excellent teaching in any curricular area is contextual, specific to individuals and situations. School arts invariably reflect an array of contexts, crucial to the design and implementation of curriculum. At the individual school level, these include practitioners' beliefs, goals, and perceptions; financial constraints and allocations; teachers' expertise; and the school's mission as part of the community in which it operates.

Educational institutions also operate within broader cultural contexts, introducing issues that shape education in general and arts education in particular. For example, new communication technologies alter our interactions with the arts in fundamental ways. Today's recording technologies eliminate barriers of time and geographical distance: The traditional listening and viewing spaces – churches, museums, and concert halls – give way to homes, hotel rooms, and cars. These changes make it difficult for school arts programs to compete with the pace and energy of the commercial art world. However, arts education provides complementary opportunities for direct immersion in creative interactions and performances, enhancing a sense of community in the school.

Technology plays an important role in bringing art to large and dispersed audiences, to the point where artwork is increasingly known by its reproductions. In educational settings, technology shapes contents and pedagogies, transforming experiences of creation and art appreciation. In music education, technology functions as a tool for the development of various skills and a novel medium for performance, and creates innovative possibilities for musical composition. In visual art, technology is used for studio/creative work, and provides access to images of many kinds. In dance and drama, videotapes are used to show masterful performances, provide material for pencil-and-paper assessment, and serve as a reliable and cost-effective means of recording student responses. Much of the impact of technology on drama and theater education is based on the increased communicative potential that interacting in a virtual space affords.

Similarly, globalization has meant the westernization of the world, generating increasing homogeneity between cultures and blurring distinctions between national, regional, and local communities. A closer look reveals that globalization produces diversity and heterogeneity through increased hybridization, highlighting historical, social, and cultural contexts. Arts education can play a central role in a globalized society as a frame through which various cultures are understood and appreciated and the students' concept of art is expanded.

Curricular areas are typically distinguished by their subject matter: visual art, music, dance, or drama and, in some countries, also poetry, literature, film, and media. They are also defined in other ways, by genre and types of instruction. Among the genres of arts used in the school

are: (1) child art, (2) child craft, (3) fine art, and (4) art for children (Bresler, 1998). Each genre is associated with different contents, pedagogies, and evaluation practices, as each is based on a separate set of ideologies and goals, related to different underlying assumptions about the nature of arts and arts learning.

Child art refers to original compositions created by children in dance, drama, visual art, and music. Child craft refers to products with prespecified outcomes, expected to be uniform. Fine art refers to classical works in the different genres created by established artists. Art for children refers to art created by adults specifically for children, often for didactic purposes.

Child art emerged with the child study movement more than a century ago. It is child produced and encourages self-expression. Fine art highlights the acquisition of cultural knowledge, typically through cognitive, factual, and critical approaches grounded in disciplines of arts history, arts criticism, and, less frequently, esthetics. Child craft and arts claim no scholarly framework, but serve practical needs in engaging children with materials meant to be developmentally appropriate, accessible, and relevant to children's lives. Though these types exist in all four arts disciplines, the emphasis on each differs depending on the discipline. Child art is prevalent in visual arts and, when taught, in dance and drama, but almost nonexistent in music instruction. Art for children is most prevalent in music in the form of songs or easy instrumental/choral works composed for young people.

An issue pertinent to all disciplines of arts education is the extent to which the curriculum is organized around masterpieces or exemplars of the fine arts, or focuses on children's processes of exploration, discovery, and self-expression. Similarly, the range and variety of images, objects, and performances that might be addressed in arts curricula vary widely, encompassing everyday phenomena such as advertising, music videos, toys, and other forms of commercial and vernacular culture, as well as images recognized as esthetically noteworthy. A related but different distinction arises between the arts as performance oriented, emphasizing the replication or interpretation of plays and established musical repertoire, and the arts as process oriented, focusing on the creation of original works of dance, drama, visual art, and music by children and youth. Schools across the world embrace curricula that display both orientations.

Two major orientations toward curriculum have shaped theory and practice in arts education. The first orientation highlights construction of knowledge through creation and engagement with artistic media, while the second highlights knowledge about the arts gleaned from the study of art history, criticism, and appreciation. These orientations are exemplified in the theories of John Dewey (1859–1952) and Harry Broudy (1905–1998), whose influence was decisive in the past 50 years of scholarship and practice in arts education.

Dewey (1934) recognized the essence of art in lived experiences rather than in the art works themselves. Art is a process, a conceptual and perceptual activity, sharing features of the methods of inquiry more commonly associated with science. Dewey's focus on art as experience supported the forms of exploratory, student-centered learning, and the focus on personal meaning that characterized art education in the United States for much of the twentieth century (Bresler, 2005; Lowenfeld, 1957).

The second, discipline-based orientation emerged in the late 1950s and 1960s, when increased attention to disciplinary learning was triggered by the launch of Sputnik and the American anxiety about being left behind in the technological race. As a result, individual disciplines were required to justify their inclusion in the curriculum in terms of their contribution to the total enterprise of education. Harry Broudy became a key voice in discussions about how the intellectual disciplines could be transformed into a program of general education in a democratic society. His rationale for the arts as part of general education was based on esthetic literacy, cultivated through esthetic experiences, involving the analysis of formal, sensory, and expressive qualities of artworks.

Broudy's ideas, along with those of Manuel Barkan, Laura Chapman, Elliot Eisner, Bennett Reimer, and others, provided a foundation for arts curricula that countered the exclusive focus on making art that prevailed in the first decades of the twentieth century. Esthetic perception entails reading images in much the same way as texts are read. Broudy (1980) emphasized the body of knowledge about the arts, specifically the history of the arts and their philosophical bases, and knowledge of the principles of criticism. He recommended a curriculum revolving around exemplary works, oriented toward students' acquisition of the knowledge of arts disciplines.

These ideas flourished in the next two decades through the Getty Center for Education in the Arts, established in 1982 and operating until the late 1990s as a strong advocate of discipline-based arts education (DBAE) (Dobbs, 2004). The Getty Institute played a major role in arts education, publishing papers by leading scholars of arts education, developing materials, and disseminating these ideas widely across the country. The promotion of academic respectability for the arts, based on a sophisticated body of knowledge and exemplary artworks, sat well with the Getty's commitment to fine arts and museums. The impact of these ideas at the level of policy was manifested in national and state goals and curricular guidelines developed during the 1990s in the US and elsewhere. DBAE represented a widespread reconceptualization of the aims of arts education at the level of theory and policy.

Critics of DBAE expressed concern that discussion and knowledge about art would replace more direct and intimate work in art. Searching for a balance between making and reflecting on the arts, North American

schools never tipped precipitously toward systematic reflection. The tradition of arts production in school arts programs remains strong. Current interest in more inclusive approaches to visual culture, multimodal, and vernacular practices in arts education emerged from the interest in the critical reading of images, objects, and events introduced by DBAE. This broadening of perspective emerged as a critique of DBAE's tendency to focus on the Western canon. This interest in criticality is motivated by social reconstructionist ideals and the desire to empower students to resist oppressive economic and political practices. Current approaches to visual culture in art education emphasize the pluralism of contemporary esthetics, focusing on artistic concept over technical mastery or craft, with art-making projects utilizing appropriation, collage, and digital technologies. The focus on artistic expression is less than it is on social critique. Social reconstructionist ideals are also evident in applied drama, which is significantly influenced by Augusto Boal's *Theatre of the Oppressed*.

Each discipline in arts education has its distinct historical context, and these are elaborated in excellent resources (see, e.g., Efland, 1990). The 1960s were times of intense activity in arts education. The esthetic education movement of the 1970s and the 1980s promoted an inclusive concept of arts education, uniting the individual disciplines of dance, drama, music, and visual art with their respective, well-established traditions. This movement to gather the four arts disciplines under one umbrella gained political power in the 1990s. As a result, dance and drama education found a place in national standards and many state goals. Still, of all school subjects, dance and drama receive the least attention in schooling, a fact reflected in the 1997 National Assessment of Educational Progress (NAEP) testing. In contrast to the widely assessed music and visual arts, the assessment in theater was confined to students who had completed at least 30 h of instruction in theater, and no assessment was conducted in dance. A survey in the Chicago public schools during 2000–01 exemplifies the same trends: The average student received 58 min per week of arts education from in-school arts specialists in the four arts disciplines: music was taught an average of 28 min per week, visual art for 27 min per week, theater for 2 min per week, and dance for 1 min per week. The relative proportion of minutes per week in each of the disciplines in the six regions in the city remained consistent (Costantino, 2003).

Even when the arts are consistently taught, arts instruction may be provided by arts specialists or by general classroom teachers. When taught by arts specialists, contents typically are conceptual, centering on elements of art such as color, shape, orchestration, melody, level, and plot. When taught by classroom teachers, contents typically are thematic, revolving around holidays, seasons, and academic content.

The Arts Education Partnership (AEP) database for 2003–04 indicates that 48 states have content standards in the arts, but only 20 states clearly mandate arts instruction (Chapman, 2005). In 1999–2000, 87% of public elementary schools offered some form of instruction in the visual arts (NCES, 2002). Reports suggest that there is strong support for arts education among administrators (67%), with less support from classroom teachers (47%; Chapman, 2005; NCES, 2002). In schools that offer arts instruction, 55% have at least one full-time teacher certified in the visual arts. About 15% of schools where arts instruction is offered schedule these classes less than once a week; this percentage increases to 36% in low socioeconomic status (SES) schools and to 42% in schools with a high proportion of students in minority groups.

The No Child Left Behind (NCLB) Act, introduced in 2001, carries complex implications for arts education. Arts educators long ago realized that standards-based reforms raise awareness of the arts as worthy domains for study. These hopes form the subtext of the national standards in the arts formulated by the Consortium of National Arts Education Associations in their 1994 document *National Standards for Arts Education*. Although NCLB includes the arts in a list of core academic subjects, the law does little to support education in the arts, foreign languages, or the humanities (Chapman, 2005). These subjects have been called the lost curriculum by the Council of Chief State School Officers (CCSSO, 2002) and cited in a discussion of the atrophied curriculum by the Council on Basic Education (2004).

Clearly, much has been achieved in arts education in the past 30 years, in terms of the range and complexity of arts curriculum, its visions, and contents. A central curricular challenge is how to cultivate qualities – depth, creativity, and originality – that are essential to the arts. These qualities have not been central values in schools. Creativity and originality in the arts always exist within a tradition, requiring mastery of skills and concepts. At the same time, they highlight the individual child's voice and interpretation. The need for mastery to create art raises the issue of depth versus breadth, within and across the arts. Mastery drives home the interconnectedness of skills, cognition, and affect, which support and enhance one other. A tension between the preservation of traditions and the push toward innovation, familiar to artists in all fields, pervades arts education.

A related challenge concerns the issue of accountability. Accountability is not a problem for arts educators (e.g., Donmoyer, 1999; Eisner, 2004): The results of arts instruction are often on public display, and arts programs receive more scrutiny from the public than programs in most other curricular areas, concerts, exhibitions, dance, and theatrical productions (Donmoyer, 1999). However, accountability in the lexicon of schooling entails standardized assessments rather than the informal and local assessments familiar to

arts programs. The impact of what is tested on what is taught is well established. Since it is easier in large-scale testing to assess the technical and factual, rather than the significant, addressing lower cognitive skills rather than those requiring synthesis and evaluation, narrow accountability (of the sort that is feasible economically) can trivialize arts instruction. Moreover, the arts are about breaking new ground and finding innovative solutions, aspects that are difficult to capture in standardized testing. When adopting tests in the arts, it is crucial to reflect on whether we are measuring what is important or what is easily measured.

An additional challenge concerns the role of the arts in schools dedicated to comprehensive education. Hope (1999) refers, tongue in cheek, to the wealth of approaches available – arts education, arts in education, arts as education, arts sprinkled on top of education, arts undermining education, arts transcending education, or arts saving education. One wonders, writes Hope, whether any other subject has the same problematic richness. For the most part, it seems that in other K-12 disciplines, the basics of each subject are more clearly defined. As the arts function so dexterously as processes, traditions of practice, methods of inquiry, and symbolic languages, they can be understood as intentional in nature: that is, a work of art is always about something, whether that something is a political position, a mood, a place, or an event.

The roots of arts integration can be traced to the ideals of progressive education and its child-centered curriculum at the beginning of the twentieth century. John Dewey, prominent in the formation of the ideals of progressive education, regarded experience and, in particular, esthetic experience as the basis around which education should revolve (Dewey, 1934). Integration penetrated from the scholarly world to the more practice-oriented circles of music and arts associations. Its earliest voices can be traced to the progressive era, for example, in the *Music Educators National Conference Yearbook* of 1933 and of 1935, which listed titles such as ‘Projects in the interrelation of music and other high school subjects’ and ‘Fusion of music with academic subjects’.

The notion of integration was revived in the 1960s and 1970s, a period of social upheaval when concern about students’ achievement yielded to concern for students’ experiences. Instead of regarding curriculum as a rigidly defined, given entity, educators focused attention on its meanings to students. This environment of innovation and experimentation with new educational goals, contents, and pedagogies promoted a fusion between the arts and academic subjects.

In the past decade, we seem to be witnessing a renewed interest in integration. Advocates for integrating the arts with academic disciplines reflect a variety of perspectives, interests, and goals. Since integration, similar to all concepts, is a construction, it can mean different things to

different people; yet, this multiplicity of meanings is not always explicit in the ways the term is used. Each constituency – principals, teachers of academic subjects, and arts specialists – has unique perspectives on contents and pedagogies in the arts, and, often, a different model in mind of what integration implies in terms of resources, planning, structures, and outcomes (Bresler, 2005). Eisner’s (1982) rationale for arts integration relates to the power of what he sees as forms of representation. Auditory, visual, and kinesthetic forms of representation develop our ability to interact with and comprehend the world around us and draw multiple meanings from it. By expanding forms of representation beyond the verbal and the numerical, our perception of the world is enriched immensely. This philosophy is embodied in the use of graphic languages and other means of symbolic representation in the early childhood schools of Reggio Emilia, where projects emerging from children’s conversations with teachers employ multiple symbolic languages as forms of exploration.

The cognitive merit of the arts and their ability to connect to other areas of human endeavors have made integration particularly appealing. Indeed, integration has been a hot issue all over the world. How is integration manifested in American curricula? Research studies suggest four ways in which the arts have been integrated into the school (Bresler, 1995). The most prevalent style in American public schools is the subservient style, in which the arts serve the basic academic curriculum in contents, pedagogies, and structures. A second, co-equal style, brings in the arts as equal partners, integrating the general curriculum with arts-specific contents, skills, and expressions. This expands perception and interpretation and typically addresses larger themes and issues. A third, affective style is typically manifested in the implicit or informal curriculum. An example is in exposure to music – as background or stimulus for movement, to soothe, calm, or to encourage free expression. It can also be manifested in exposure to the arts through concerts and museum visits that are not connected to the curriculum. The fourth, social style, rarely a part of the formal curriculum, but present in many American schools, revolves around the social function of the school and its role as a community. It is typified in school performances by children.

The subservience of arts education to other disciplines is common across cultures. In the past, writes Flolu (2000), drawing, weaving, sculpting, and the other art disciplines were taught in Ghana as creative training activities. Now, they are used as a means of enhancing understanding of other subjects. For example, after a lesson on ‘Building a healthy individual’, pupils may be asked to draw pictures of children at work (CRDD, 1998: 1) and use clay or paper to model the school building. In the US, the arts serve academic disciplines when classroom teachers ask music teachers to teach songs and dances to accompany their curriculum (e.g., to teach songs about the planets to

support study in science, or the Virginia reel for American history).

Similar to all curricular areas, arts education faces the persistent challenge of determining “What’s worth knowing, who decides, and in whose interests are decisions being made?” (Stinson, 2007: 144). These are exciting, if potentially perilous, times for arts education: times of rapid expansion of educational programming in museums, community-based settings, and other cultural institutions, balanced against the threat of constriction and dilution in schools dependent upon test scores in reading and math for their survival. These are times in which innovative curricular materials and proposals proliferate, and new media and technologies expand the possibilities for making and responding to works of art and popular culture, in which theoretical ideas from the informing disciplines of the arts and education stimulate innovation, and dialog across disciplines, within and beyond the arts, is possible. The influence of DBAE persists in recognition of the intellectual dimension of art experience and the importance of acknowledging multiple ways of creating, performing, and responding to works of art and to an expanded range of objects, images, and events, beyond the traditional Western canon of the fine arts. New materials, tools, and processes introduce new possibilities for art making, and new formats for viewing and responding to art forms. This is a time of rich multiplicity, in which the curricular orientations of the past coexist as possibilities and inclusive visions of arts education prevail.

Bibliography

- Bresler, L. (1995). The subservient, co-equal, affective and social integration style of the arts. *Arts Education Policy Review* **96**, 31–37.
- Bresler, L. (1998). “Child art,” “fine art,” and “art for children”: The shaping of school practice and implications for change. *Arts Education Policy Review* **100**, 3–10.
- Bresler, L. (2005). *Visual and Performing Arts: A Chapter of the Curriculum Handbook*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Broudy, H. (1980). On the third realm-aesthetic schooling. *Journal of Aesthetic Education* **14**, 5–9.
- CCSSO (Council of Chief State School Officers) (2002). *Key State Policies on PK-12 Education: 2002*. Washington, DC: Author.
- Chapman, L. (2005). Status of elementary art education: 1997–2004. *Studies in Art Education* **46**, 118–137.
- Costantino, T. E. (2003). The impact of philanthropy on arts education policy. *Arts Education Policy Review* **105**, 25–32.
- Council on Basic Education (CBE) (2004). *Academic Atrophy: The Condition of the Liberal Arts in America’s Public Schools*. <http://www.c.b-e.org> (accessed May 2009).
- CRDD (Curriculum Research Development Division) (1998). *Draft Syllabus for Environmental Studies: Basic Stages 1–6*. Accra: Ministry of Education.
- Dewey, J. (1934). *Art as Experience*. New York: Perigee Books.
- Dobbs, S. (2004). Discipline-based art education. In Eisner, E. and Day, M. (eds.) *Handbook of Research and Policy in Art Education*, pp 701–724. Mahwah, NJ: Erlbaum.
- Donmoyer, R. (1999). Rendering unto Caesar that which is Caesar’s: A critique of the NAEP in theater. *Arts Education Policy Review* **100**, 21–26.
- Efland, A. (1990). *A History of Art Education: Intellectual and Social Currents in Teaching the Visual Arts*. New York: Teachers College Press.
- Eisner, E. (1982). *Cognition and Curriculum*. New York: Macmillan.
- Eisner, E. W. (2004). What can education learn from the arts about the practice of education? *International Journal of Education and the Arts*. <http://ijea.asu.edu/v5n4>.
- Flolu, J. (2000). Re-thinking arts education in Ghana. *Arts Education Policy Review* **101**, 25–29.
- Hope, S. (1999). Time, distance, focus, excellence. *Arts Education Policy Review* **100**, 2–11.
- Lowenfeld, V. (1957). *Creative and Mental Growth*, 3rd edn. New York: Macmillan.
- NCES (National Center for Education Statistics) (2002). *Arts Education in Public Elementary Schools, 1999–2000, (NCES-131-2001)*. Washington, DC: US Department of Education.
- Stinson, S. (2007). Making sense of curriculum research in arts education. In Bresler, L. (ed.) *International Handbook of Research in Arts Education*, pp 143–146. Dordrecht: Springer.
- Wilson, B. (1997). Child art, multiple interpretations, and conflicts of interest. In Kindler, A. M. (ed.) *Child Development in Art*, pp 81–94. Reston, VA: National Art Education Association.
- Barrett, T. (2002). *Interpreting Art: Reflecting, Wondering, and Responding*. New York: McGraw Hill.
- Betts, D. (2003). Art + technology integration: Developing an after school curriculum. *After School Matters* **2**, 13–22.
- Bresler, L. (ed.) (2007). *International Handbook of Research in Arts Education*. Dordrecht: Springer.
- Bresler, L. and Ardichivili, A. (eds.) (2002). *International Research in Education: Experience, Theory and Practice*. New York: Peter Lang.
- Bresler, L. and Thompson, C. M. (eds.) (2002). *The Arts in Children’s Lives: Context, Culture, and Curriculum*. Dordrecht: Springer.
- Consortium of National Arts Education Associations (1994). *National Standards for Arts Education*. Reston, VA: Music Educators National Conference.
- Deasy, R. J. (ed.) (2002). *Critical Links: Learning in the Arts and Student Academic and Social Development*. Washington, DC: Council of Chief State School Officers.
- Eisner, E. and Day, M. (eds.) (2004). *Handbook of Research and Policy in Art Education*. Mahwah, NJ: Erlbaum.
- Hanna, J. L. (1999). *Partnering Dance and Education: Intelligent Moves for Changing Times*. Champaign, IL: Human Kinetics.
- Mark, M. L. and Gary, C. L. (1992). *A History of American Music Education*. New York: Schirmer Books.
- Reimer, B. (2003). *A Philosophy of Music Education: Advancing the Vision*. Boston, MA: Prentice Hall.
- Smith, R. (1987). The changing image of art education: Theoretical antecedents of discipline-based art education. *Journal of Aesthetic Education* **21**, 3–34.
- Stake, R., Bresler, L., and Mabry, L. (1991). *Custom and Cherishings: The Arts in Elementary Schools*. Urbana, IL: Council for Research in Music Education.
- Sullivan, T. and Willingham, L. (eds.) (2002). *Creativity and Music Education*. Edmonton, AB: Canadian Music Educators’ Association.
- Winner, E. and Hetland, L. (2000). The arts in education: Evaluating the evidence for a causal link. *Journal of Aesthetic Education* **34**, 11–76.

Further Reading

- Barrett, T. (2002). *Interpreting Art: Reflecting, Wondering, and Responding*. New York: McGraw Hill.
- Betts, D. (2003). Art + technology integration: Developing an after school curriculum. *After School Matters* **2**, 13–22.
- Bresler, L. (ed.) (2007). *International Handbook of Research in Arts Education*. Dordrecht: Springer.
- Bresler, L. and Ardichivili, A. (eds.) (2002). *International Research in Education: Experience, Theory and Practice*. New York: Peter Lang.
- Bresler, L. and Thompson, C. M. (eds.) (2002). *The Arts in Children’s Lives: Context, Culture, and Curriculum*. Dordrecht: Springer.
- Consortium of National Arts Education Associations (1994). *National Standards for Arts Education*. Reston, VA: Music Educators National Conference.
- Deasy, R. J. (ed.) (2002). *Critical Links: Learning in the Arts and Student Academic and Social Development*. Washington, DC: Council of Chief State School Officers.
- Eisner, E. and Day, M. (eds.) (2004). *Handbook of Research and Policy in Art Education*. Mahwah, NJ: Erlbaum.
- Hanna, J. L. (1999). *Partnering Dance and Education: Intelligent Moves for Changing Times*. Champaign, IL: Human Kinetics.
- Mark, M. L. and Gary, C. L. (1992). *A History of American Music Education*. New York: Schirmer Books.
- Reimer, B. (2003). *A Philosophy of Music Education: Advancing the Vision*. Boston, MA: Prentice Hall.
- Smith, R. (1987). The changing image of art education: Theoretical antecedents of discipline-based art education. *Journal of Aesthetic Education* **21**, 3–34.
- Stake, R., Bresler, L., and Mabry, L. (1991). *Custom and Cherishings: The Arts in Elementary Schools*. Urbana, IL: Council for Research in Music Education.
- Sullivan, T. and Willingham, L. (eds.) (2002). *Creativity and Music Education*. Edmonton, AB: Canadian Music Educators’ Association.
- Winner, E. and Hetland, L. (2000). The arts in education: Evaluating the evidence for a causal link. *Journal of Aesthetic Education* **34**, 11–76.

Relevant Websites

- <http://artsedge.kennedy-center.org> – ARTSEDGE, Lessons.
- <http://artscurriculum.guggenheim.org> – Guggenheim Museum, Arts Curriculum.
- <http://www.pbs.org> – Public Broadcasting Service (PBS), Art:21.
- <http://www.uic.edu> – University of Illinois, Spiral Art Education.

Curriculum: Foreign Language Learning

Angel M Y Lin, City University of Hong Kong, Hong Kong SAR, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

This article presents an overview of various issues related to curriculum in foreign language learning, and in particular focuses on learning English as a foreign language (EFL). Foreign language learning is taken to mean the learning of a language other than the learner's first language (L1), and this language is not ordinarily used in the learner's everyday life. Thus, foreign language learning contexts are very different from second language learning contexts, for in second language learning contexts, the language being learned is often used in the learner's larger social context (even though it might not be always used in the learner's immediate home or community). This distinction should not be seen as categorical, as some contexts lie in between the prototypical foreign language learning contexts and the prototypical second language learning contexts. However, as is discussed in this review, the importance of understanding the sociocultural, sociopolitical, and socio-economic situatedness of second and foreign language learning in diverse contexts of the world has, until recently, been underrepresented in the literature.

The Production and Circulation of Received Knowledge about Foreign Language Curricula

In the English language world, the key sites of production of knowledge about foreign language education lie in the English-speaking worlds (e.g., US, Britain) and Europe, where many applied linguists and educators work on developing (English as a) foreign language curricula and teacher education programs, and publish their findings and views in the English language academic world. Their publications have constituted the classic literature on foreign language curriculum (e.g., Wilkins, 1976). In the field of EFL education in particular, there is a lot of spilling over of knowledge from English as a second language (ESL) education, with knowledge derived from ESL research in English-speaking contexts often taken as largely pertinent to EFL education in diverse contexts of the world. Foreign language (and EFL) education knowledge paradigm shifts thus largely mirror second language (and ESL) education knowledge paradigm shifts. These paradigm shifts are discussed in the next section.

Paradigm Shifts in Second and Foreign Language Education

The major disciplines informing second and foreign language education have traditionally been linguistics and psychological theories of language acquisition. It follows that knowledge paradigm shifts in these disciplines also lead to corresponding shifts in knowledge and views on second and foreign language curricula and pedagogy. In designing any language curriculum (or curriculum in general), two central questions naturally arise: What should be included in the curriculum and how should the curriculum content be taught (e.g., in what sequence and with what kinds of teaching methods)?

Structural linguistics has long been the chief framework underlying the development of language curricula. Richards (2001) reviewed the historical background of vocabulary and grammar gradation/selection in developing language curriculums from the 1920s to the 1970s. The assumptions underlying early structuralist approaches to language syllabus design can be summarized as follows:

- the basic units of a language curriculum are vocabulary and grammar;
- learners everywhere have the same needs;
- learners' needs are identified exclusively in terms of language needs;
- the process of learning a language is largely determined by the textbook;
- the classroom and the textbook provide the primary input to the language learning process; and,
- the goal of the syllabus designer is to simplify and rationalize this input through selection and gradation (Richards, 2001: 15–16).

It can be seen that the basic assumptions of structural linguistics permeate the early approaches to language syllabus design. Mastering human language communication is seen as equivalent to mastering the structural units of the language system. The systematic, logical, sequencing and presentation of linguistic structural units become the central task for language syllabus designers. These assumptions had influenced the design of language syllabus and pedagogy until the 1970s, when functional linguistics became the strongest rival of traditional, structural linguistics.

From Structural/Grammatical to Notional/Functional Syllabuses

With the rise of functional linguistics in the 1970s and 1980s, there has been a movement toward the notional, semantic, and functional syllabi (Johnson, 1982; Wilkins, 1976). Instead of treating the structural linguistic elements as the units, the units are now functions and notions. Topics, settings, functions, and notions become the units to be rationally sequenced and organized in a language syllabus. So, whereas the traditional structural syllabus is basically an inventory of linguistic items, graded and sequenced in terms of their structural complexity, the new syllabus consists of an inventory of communicative functions (e.g., requesting services, seeking information, changing topics, expressing disagreement), notions (i.e., concepts such as distance, duration, quantity, quality, location, size), organized around different settings/situations or topics. However, unlike the task of the structural syllabus designer, the functional/notional syllabus designer has a much more difficult task. Whereas traditional structural linguistics provides a straightforward approach to describing, grading, sequencing, and organizing the different linguistic elements according to structural complexity (e.g., from simple to complex structures), the functions are much more diverse and messy to organize into a finite inventory. Thus, needs analysis was proposed as the main procedure informing the design of a functional/notional syllabus for a specific group of learners, addressing their communicative needs (e.g., ESL university students studying English to learn specific academic subjects). Needs analysis was thus introduced into language teaching mainly through the English for specific purposes (ESP) movement. Functional linguistics, especially register analysis, was also drawn upon to identify the linguistic features of disciplines such as medicine, engineering, or science. It can be seen that the ESP movement and the functional/notional syllabus movement have developed together and the student body served is usually adult learners of a second or foreign language, often for immediate, identifiable, specific purposes. The procedure of needs analysis, however, may not be easily applicable to general school foreign language learning contexts, where immediate needs for learning the foreign language might not be easily identifiable.

From Traditional Approaches to Communicative Language Teaching

While the what-question (i.e., what to include in a language curriculum) has been answered by notional/functional syllabus designers with the dual procedures of

needs analysis and register analysis, the how-question (i.e., how to teach the syllabus content) is answered by scholars working on the development of the communicative language teaching (CLT) approach. Richards (2001: 3) summarized the different teaching methods which characterized different periods in the past century:

1. grammar translation method (1800–1900);
2. direct method (1890–1930);
3. structural method (1930–60);
4. reading method (1920–50);
5. audio-lingual method (1950–70);
6. situational method (1950–70); and
7. communicative approach (1970–present).

Approaches before the 1970s are largely structure-drill-based. Behaviorist theories of language learning had emphasized the formation of accurate language behavior through habit formation based on practice and drills of structures. The rise of Krashen's (1981) SLA model and interactionist models leading to input studies of language acquisition have led to the pedagogical principles of providing comprehensive input in the classroom for students to develop their own L2 interlanguage, encouraging students' active negotiation of meaning, lowering learner anxiety by encouraging learners to speak up and take risks, and developing students' own monitoring ability to self-correct. These theories of language acquisition have converged with the rise of functional linguistics to focus on the learners' use of the foreign language for authentic, meaningful interaction and self-expression. CLT educators have since developed repertoires of techniques to promote communicative use of the target language in the classroom, for example, information-gap activities to promote authentic exchange of meaning. Task-based and project-based approaches (Nunan, 1989) to language learning have been proposed with the principle of promoting students' authentic use of the target language for meaningful communication. All these represent a departure from past approaches based on repetition and drills. However, some CLT educators also recognized the need for pre-communication activities or a pre-production phase, where practice/drill-based methods are used to help the learner acquire the necessary linguistic structures to be used later in the communicative production phase of the language lesson.

Tension between Local Approaches and CLT in Diverse Contexts of the World

By the late 1980s, CLT was widely accepted as the latest innovative teaching methodology in the second and foreign language education literature in the Anglo-speaking

world although there were still some unsolved questions, such as those summarized by Richards and Rodgers (1986: 83) in their review of CLT:

1. Can CLT be applied at all levels of a language program?
2. Is it equally suited to ESL and EFL situations?
3. Does it require existing grammar-based syllabuses to be abandoned or merely revised?
4. How can it be adopted in situations where students must continue to take grammar-based tests?

CLT has since been implemented and received with mixed responses by teachers and students in diverse contexts of foreign language learning, especially in contexts where English is taught as a foreign language. This has led to some language educators reflecting on the importance of considering the social context in determining what counts as appropriate methodology and the need to value indigenous knowledge, perspectives, and pedagogies. The 2000s further witnessed a new body of research literature, chiefly authored by language education scholars researching on the sociocultural and sociopolitical incompatibility between local approaches and CLT in diverse sociocultural and sociopolitical contexts of the world (e.g., India, China, and South Korea). These new scholarly voices (e.g., Ramanathan, Ouyang, Shin) show the inadequacies of the classic literature on second/foreign language education, especially those canons established in the English-speaking academic world. Below is summarized the major contribution of these new scholars in pointing to the need to pay attention to the sociocultural, socioeconomic, and sociopolitical situatedness of language learning and language curriculum (Ouyang, 2000), and the need for indigenous, nonuniversalist perspectives (Ramanathan, 2006; Shin, 2006) regarding how to develop appropriate foreign language curriculum and pedagogy in diverse contexts of the world. The major problematic aspects have been summarized by several scholars around a few recurring points:

1. CLT and ELT methodologies have been treated as a neutral, objective discipline or technology that can be exported from one country to another. However, this implementation of a West-based methodology has ignored many local constraints, communal needs, and values (Ouyang, 2000; Ramanathan, 2006; Shin, 2006).
2. In CLT, the role of the teacher is constructed in such a way as to mainly facilitate the acquisition of competence in performing communicative functions in the target language. However, the teacher in many cultures and societies is also expected to play other important roles; e.g., as an authoritative, moral leader to students (Ouyang, 2000).
3. CLT has been constructed in the West-based literature mainly as a value-free technology to teach a second or foreign language effectively. However, CLT actually

carries implicit values and ideologies such as utilitarianism, egalitarianism, and individualism. The kind of social relations encouraged in the CLT classroom might not be seen as acceptable in certain traditional societies and cultural milieus (Ouyang, 2000).

4. Certain local pedagogical practices disfavored by CLT might have their own rationality and usefulness in the local contexts of the students (e.g., use of L1, use of choral responses and translation techniques, drawing on students' familiar L1 lifeworld experiences to stimulate student interest and response, systematic and authoritative transmission of grammatical knowledge, etc.) (Ouyang, 2000; Ramanathan, 2006).
5. CLT and West-based TESOL implicitly assumes that the language teacher and learner naturally want to become like the native speaker. However, this might not be the case; for example, many Korean English language teachers and students might want to keep their Koreaness while they are teaching, learning, and speaking English. For instance, the teachers might want to keep their local identities as Korean teachers, or Chinese teachers, even when they are teaching English and do not want to adopt the teaching styles of foreign teachers (Ouyang, 2000; Shin, 2006).

To the above list is added Ramanathan's (2006) point about the need to pay attention to the socioeconomic and sociopolitical embeddedness of ELT in a society, especially the position of English when compared to other languages in society and how differential access to English and different pedagogies of English contribute to social stratification and inequalities (see also Lin, 1999). It is to a consideration of this point that we shall turn to in the next section.

Crisscrossing of Statist and Capitalist Desires in English Language Teaching

The list of critiques of CLT and West-based methodology in the above section can be summarized in the key observation that there is a fundamental problem with the traditional literature's key assumption and mission. This key assumption and mission can be summarized as follows: the language curriculum researcher's task is mainly that of the search for the most effective language curriculum and pedagogy as neutral technologies and procedures that are universally applicable and effective, without due attention to the sociocultural, sociopolitical, and socioeconomic embeddedness of language learning and language teaching in different societies.

Much of the critique of CLT and West-based ELT curriculum theories and pedagogies can thus be understood by situating the institution of EFL teaching in its

larger sociopolitical and global capitalist context, and by understanding the hidden but nonetheless inherent nature of (language) teaching and schooling as a socio-political and sociocultural process of (re)production of subjectivities and identities. With the processes of globalization and global capitalism growing fast at the turn of the century, the question of the global hegemony of English cannot be evaded and its impact on language-in-education policy and practice in diverse contexts of the world must be examined. Several major sources of tensions and dilemmas have arisen as scholars have examined the experience of teachers, students, parents, and language curriculum planners and policymakers in diverse contexts of the world.

Conflicts and Social Division

Under forces of globalization, many Asian states have begun to adopt a strong ELT policy in their schooling curriculum with the aim to produce a workforce communicative in English to feed the desire of global capitalism for English-conversant human resources. Shin (2006) described the South Korean government's lopsided emphasis on promoting the status of English in schools and parents' strong desire for English as related to the neocolonial role of the US in Korean history and not just to the status of English as a global language. And good English is often defined as the variety of English spoken by Anglophones from Western countries while Korean teachers of English are denigrated for their Korean accent (Shin, 2006). In Iran, middle-class parents who can afford it are sending their children to private English tutorial centers to learn to be communicative in English, in reaction to the drills-based, structure-oriented English curriculum taught in public schools. In China, a CLT-trained teacher moved to a private language school to teach English and make more money after being frustrated by the traditional curriculum culture and practice of her old school which resisted her CLT-oriented efforts in curriculum and pedagogical reform, although the Chinese government has spent the past decade reforming its national English assessment format to make it more communicative to produce a workforce more amenable to the English demands of the new market economy (Ouyang, 2000). In India, Ramanathan (2006) wrote about the social division created by two different sets of institutional curriculum goals and practices for the haves and the have-nots in the economy of English. In Hong Kong, the government's initiative to reform the O-level public exam in oral English into a school-based speaking test, with the aim to introduce more communicative, formative assessment formats to induce progressive pedagogies in the English curriculum, has created more workload to school

teachers and worries about how to ensure fairness of such a high-stake public exam.

The above examples of tension and conflicts revolving around English are illustrative of what is happening in many diverse contexts of the world. The sources of conflict and social division can be summarized as follows:

1. Under the global capitalist desire for an English-conversant workforce, the modern states, often under internal pressure of the business sector and the state's own desire to globalize, have, in the twenty-first century, begun to initiate reforms in their traditional structure-based English curriculum toward a more communicative curriculum.
2. However, conflicts arise due to other institutional and social demands: for example, the traditional societal and cultural emphasis on national, standardized tests for screening and credentialing purposes in many Asian states (e.g., communicative functions are difficult to assess in a national, standardized exam); and the traditional cultural milieu of de-emphasizing egalitarianism and individualist self-expressiveness in classroom interactions/teacher-student relationships, which have served good social control functions in maintaining the social order of many traditional societies.
3. The desire for English as a global commodity, the marketization of languages (Block and Cameron, 2002), and the global spread of the ELT/TESOL industry, combined, in some contexts, with the legacy of political colonialism of Western countries (Shin, 2006), have elevated the status of the foreign expert (Ouyang, 2000), that is, the native-English-speaker teacher, way above the local teachers of English, whose contribution in developing an indigenous, culturally compatible curriculum (Lin, 1999) is often under-estimated or unrecognized in the West-based CLT curriculum literature.
4. Progressive pedagogies such as the CLT curriculum and methodology often require ample linguistic and cultural capital in the schools to make them work (Lin, 1999). CLT English curriculum reforms have often created difficulties for the poorer sectors of the school communities; for instance: large class-size (e.g., 40 students); heavy teaching load; limited-English-proficiency students with little familial and community support to learn English; teachers lacking the training to teach and assess communicative functions, etc.

Thus, ELT curriculum development cannot be assumed to be a neutral technology but is embedded in, and constitutive of, sources of social division, stratification, and inequalities. ELT practitioners need to be aware of how their profession falls along socially stratifying lines and explore what they can do as teachers to sidestep these policies both inside and outside classrooms. The traditional ELT curriculum and research literature is,

thus, in dire need of a critical turn, after its functional/communicative paradigm shift in the 1970s and 1980s resulting in the CLT curriculum. In foreign language teacher education, especially EFL teacher preparation, a standard course in CLT curriculum might often leave EFL teachers ill-prepared for student resistance and classroom realities, because of a lack of attention paid to students' desires and identities.

Whose Desire Counts? Students' Resistance in EFL Classrooms

Unlike ESP curriculums tailor made for adult learners who have specific immediate needs for learning English, the general school English curriculum, especially in EFL contexts where English is not used in the students' everyday life, is often designed with the state's and the employers' desire to produce an English-conversant workforce in the future. Students themselves often have no immediate communicative needs to learn English except the extrinsic need to pass English examinations. In recent years, however, many postcolonial states (e.g., Malaysia) have re-installed their former colonizer's language (which is usually English) as the medium of instruction in schools and universities under the desire to produce an English-conversant workforce to participate in the global economy. Using immersion as an ambitious form of foreign language instruction has its origins in the often-cited French immersion programs started in Canada in the 1960s and 1970s. The rationale lies in creating in students an authentic communicative need for the second or foreign language by using it as the medium for teaching and learning other content subjects. However, this ambitious form of foreign language instruction has created difficulties in contexts where both students and teachers struggle to use an unfamiliar language to learn content subjects. In Kenya, the policy of installing English as the medium of instruction from Primary Four onward typifies enormous challenges to the majority of children and teachers, particularly those living in rural

and poor urban communities, where there is little access to English outside school.

The marketization of language teaching and the commodification of English have promoted approaches that lose sight of the desires of young learners themselves, who cannot be assumed to be ready and willing consumers of CLT curriculums infused with statist and global capitalist desires. What does learning English mean to young school learners, especially in contexts where English occupies a superior socioeconomic position in relation to the first language of students under increasing globalization forces? The converging desires of the state, the employers, and the larger society (including parents and principals) for students to acquire English for socioeconomic reasons has often led to the students' own resistance in the English classroom: students engage in creative verbal play by drawing on their indigenous language and youth pop cultural resources to subvert the English lesson agenda, to mock the English teacher, the English curriculum, or the language learning task being imposed on them, in a way to assert their own indigenous identities and youthful desires to engage in fun, transgressive verbal play of their own choice.

The Critical Desire in Foreign Language Curriculum

Desire is the motivating force behind a person's investment in the arduous task of language learning. The school foreign language curriculum is usually infused with the desire of the state, the employers, the parents, and increasingly the global capitalists. The 2000s have, however, witnessed the beginning of the critical and sociocultural turn in foreign language curriculum research. **Figure 1** from Osborn (2000) succinctly outlines what can be conceptualized as a desire and linguistic matrix that a critical foreign language curriculum researcher can present to teachers, students, and policymakers to help them critically reflect on and decide upon their own stance toward foreign language learning in their respective contexts.

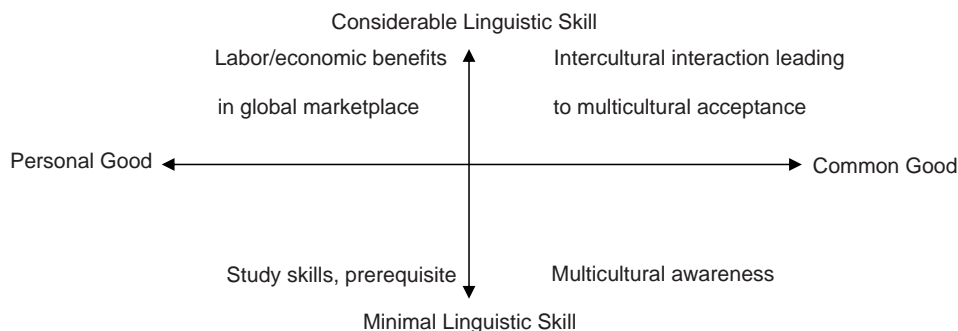


Figure 1 Proposed curricular advantages in terms of linguistic skill requirement. From Osborn, T. A. (2000). *Critical Reflection and the Foreign Language Classroom*. Westport, CN: Bergin and Garvey. With kind permission from osborn.

While CLT curriculum researchers and developers can still apply their needs analysis procedure and negotiate with their target student groups to chart out what might possibly count as their short-, mid- and long-term valuable learning goals (e.g., specific structures, functions, communicative, academic language, or examination needs) as perceived by the students, critical curriculum researchers can simultaneously work out and negotiate with their students the humanistic goals of intercultural understanding, multiculturalism, and social justice through learning foreign and world languages, including English as a lingua franca. However, for English to serve the critical desire as a language for common good (e.g., intercultural understanding and acceptance, respect for ethnic, linguistic, and cultural diversity) rather than global capitalism and social stratification, the kind of English to be taught and learnt has to take on different world accents and has to be multiculturalized and hybridized by speakers of other languages in the world, that is, to be interpenetrated and interilluminated by other voices, accents, meanings, and languages (Bakhtin, 1981).

Critical foreign language curriculums/educators can be as imposing and alienating as traditional or CLT curriculums/educators, depending on the context and the actual classroom interactions that transpire. However, if “education is ultimately the process of non-coercive rearrangement of desires,” as pointed out by the postcolonial scholar, Gayatri Spivak (personal communication, July 2002), critical foreign language educators need to work on exploring ways of noncoercively re-arranging and re-engaging students’ desires in the linguistic and critical project of learning world languages for social justice and intercultural understanding. Osborn (2006) has laid out some useful principles and practical resources that teachers can draw on in such a project. However, there will not, and should not, be any recipe books for delivering critical language curriculums, apart from some general principles and perspectives serving as guidance for the critical educator to adopt and develop their own indigenous, situated pedagogies for their students.

Conclusion: Toward Internally Persuasive Discourses in Foreign Language Curricula

The analyses of students’ resistance and verbal play in their attempts to subvert the imposed English lesson agenda underscore the resilience of human agency and creativity, the human need to go beyond monoglossia, that is, the types of social languages imposed on them in school and society, the drive to turn them into future worker commodities, disciplining them in the foreign languages expected of them in the adult worker world, forcing them to parrot service-worker languages, and constituting their

voices for them. Even in such a situation, some students did not fail to accentuate the parroted utterances with their own voice and accent, attaching to the prescribed utterances their own implicit social and political commentary and meanings.

Bakhtin differentiates between two kinds of discourses: authoritative discourse and internally persuasive discourse. Authoritative discourse is language or discourse imposed on one, but for one to really accept, acquire, and own a language or discourse, it has to become an internally persuasive discourse, hybridized and populated with one’s own voices, styles, meanings, and intentions:

... Both the authority of discourse and its internal persuasiveness may be united in a single word—one that is simultaneously authoritative and internally persuasive—despite the profound differences between these two categories of alien discourse. But such unity is rarely a given—it happens more frequently that an individual’s becoming, an ideological process, is characterized precisely by a sharp gap between these two categories: in one, the authoritative word (religious, political, moral; the word of a father, of adults and of teachers, etc.) that does not know internal persuasiveness, in the other internally persuasive word that is denied all privilege, backed up by no authority at all, and is frequently not even acknowledged in society (not by public opinion, not by scholarly norms, nor by criticism), not even in the legal code (Bakhtin, 1981: 342).

Bakhtin’s insights on the need for heteroglossia and local creativity even in the face of monoglossia suggest a way to resolve these tensions by co-creating heteroglossic, internally persuasive dialogs of interest to students so that English (or a foreign language) can become a language populated with students’ own voices and serve as a tool that students can draw on to construct their own preferred worlds, preferred identities, and preferred voices.

Future directions of research might further draw on cultural studies, postcolonial studies, critical pedagogy, and critical sociolinguistics to develop both theoretical and pedagogical projects. These projects can explore how teachers and students can cross and destabilize the socially constructed boundaries of codified languages, cultures, races, and so on, in exploring how different languages (e.g., the first language and the foreign language) and cultures can be brought together to interilluminate, interpenetrate, and mutually enrich and hybridize each other (Bakhtin, 1981), to create a heteroglossic, multicultural language classroom where both the teacher and students’ desires can be brought to critical consciousness, to be examined, negotiated, re-arranged, and re-engaged to work toward both personal good and common good. Research also needs to focus on how foreign language curricula can draw on new digital media, youth informal literacies, youth popular culture, and Internet practices

(Gee, 2007), areas all infused with young people's intense fantasies and desires, which might, perhaps, await a non-coercive process of re-arranging.

See also: Bilingual Learning (Learning L1 and L2 in an L1 and L2 Environment); Curriculum and the Education of Cultural and Linguistic Minorities; Learning a Second Language in First Language Environments.

Bibliography

- Bakhtin, M. M. (1981). *The Dialogic Imagination: Four Essays by M. M. Bakhtin*. In Emerson, C. and Holquist, M. (eds.). Austin, TX: University of Texas Press.
- Block, D. and Cameron, D. (eds.) (2002). *Globalization and Language Teaching*. London: Routledge.
- Gee, J. P. (2007). *Good Video Games + Good Learning: Collected Essays on Video Games, Learning and Literacy*. New York: Peter Lang.
- Johnson, K. (1982). *Communicative Syllabus Design and Methodology*. Oxford: Pergamon.
- Krashen, S. D. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Oxford University Press.
- Lin, A. M. Y. (1999). Doing-English-lessons in the reproduction or transformation of social worlds? *TESOL Quarterly* **33**, 393–412.
- Nunan, D. (1989). *Designing Tasks for the Communicative Classroom*. Cambridge: Cambridge University Press.
- Osborn, T. A. (2000). *Critical Reflection and the Foreign Language Classroom*. Westport, CN: Bergin and Garvey.
- Osborn, T. A. (2006). *Teaching World Languages for Social Justice: A Sourcebook of Principles and Practices*. Mahwah, NJ: Lawrence Erlbaum.
- Ouyang, H. (2000). One-way ticket: A story of an innovative teacher in Mainland China. *Anthropology and Education Quarterly* **31**, 397–425.
- Ramanathan, V. (2006). The vernacularization of English: Crossing global currents to re-dress West-based TESOL. *Critical Inquiry in Language Studies* **3**, 131–146.
- Richards, J. C. (2001). *Curriculum Development in Language Teaching*. Cambridge: Cambridge University Press.

Richards, J. C. and Rodgers, T. S. (1986). *Approaches and Methods in Language Teaching: A Description and Analysis*. Cambridge: Cambridge University Press.

Shin, H. (2006). Rethinking TESOL from a SOL's perspective:

Indigenous epistemology and decolonizing praxis in TESOL. *Critical Inquiry in Language Studies* **3**, 147–167.

Wilkins, D. A. (1976). *Notional Syllabuses: A Taxonomy and Its Relevance to Foreign Language Curriculum Development*. London: Oxford University Press.

Further Reading

- Canagarajah, A. S. (1999). *Resisting Linguistic Imperialism in English Teaching*. Oxford: Oxford University Press.
- Lam, W. S. E. (2000). L2 literacy and the design of the self: A case study of a teenage writing on the internet. *TESOL Quarterly* **34**, 457–482.
- Lin, A. M. Y. and Luk, J. C. M. (2005). Local creativity in the face of global domination: Insights of Bakhtin for teaching English for dialogic communication. In Hall, J. K., Vitanova, G., and Marchenkova, L. (eds.) *Dialogue with Bakhtin on Second and Foreign Language Learning: New Perspectives*, pp 77–98. Mahwah, NJ: Lawrence Erlbaum.
- Lin, A. M. Y. and Martin, P. (eds.) (2005). *Decolonisation, Globalization: Language-in-Education Policy and Practice*. Clevedon: Multilingual Matters.
- Lin, A. M. Y., Wang, W., Akamatsu, A., and Riazzi, M. (2005). Transnational TESOL professionals and teaching English for globalized communication. In Canagarajah, A. S. (ed.) *Negotiating the Global and Local in Language Policy and Practice*, pp 197–222. Mahwah, NJ: Lawrence Erlbaum.
- Luk, J. C. M. and Lin, A. M. Y. (2006). *Classroom Interactions as Cross-Cultural Encounters: Native Speakers in EFL Lessons*. Mahwah, NJ: Lawrence Erlbaum.
- Pennycook, A. (2007). *Global Englishes and Transcultural Flows*. London: Routledge.
- Ramanathan, V. (2005). *The English-Vernacular Divide: Postcolonial Language Politics and Practice*. Clevedon: Multilingual Matters.
- Smith, D. I. and Osborn, T. (2007). *Spirituality, Social Justice, and Language Learning*. Charlotte, NC: IAP – Information Age.

First Language Learning

T Rogers and L Moffatt, University of British Columbia, Vancouver, BC, Canada

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Critical literacy – Defining literacy as carrying power.

Diaspora – The movement of peoples.

Diglossic – Speaking two closely related languages.

Meta-cognitive strategies – The conscious control and awareness of one's own learning strategies.

Meta-language – The language about how written and spoken texts carry meaning.

Mother tongue – The first language spoken.

Multiliteracies – Literacy as consisting of multiple languages and modalities.

Phoneme – The smallest unit of speech that carries meaning.

Validity in interpretation – The meaning of a literary work that is ascertainable and related to authorial intent.

This article presents an overview of various issues related to curriculum in first language learning, including what is commonly referred to as language and literacy instruction from kindergarten through secondary school. It is important to point out that first language learning often takes place in students' second or third language, or in a language distinct from their mother tongue. Therefore, the term first language learning in this article refers to the language of schooling or the official language of education in any given country. The particular focus of this article is on language and literacy education in English.

When considering first language learning in English, it is important to recognize that in many contexts, although English is the official language of schooling, it may or may not be related to the mother tongue of many or, in some cases, most school-aged children or adolescents in a country or region. When English is used as an official language or as a language of schooling, it is generally the result of complex colonial, diasporic, and political histories. In some contexts, such as India and some countries in Africa, few students may actually speak English at home and yet, English is used as an official language in governments, businesses, and schools. In other countries, such as Singapore, English may be the official language of particular schools. However, students in those schools may be diglossic speakers using English as one of a number of languages with their friends

and family. In addition, there are contexts (e.g., Canada) where English is neither the only official language, nor the only language of schooling.

Factors Influencing Language and Literacy Curricula

Many factors influence the way teachers approach language and literacy teaching in their classrooms, ranging from the diversity of their students, the official and the tacit or hidden (Freire, 1972/2000) school curricula, current research and professional development, government policy and related means of accountability, publishing corporations, and the media. Language educators are also responsive to the interests of parents and communities, fellow educators (including teacher unions), researchers, and policymakers – both close and more distant. In other words, language curricula in all contexts are simultaneously influenced by local and global discourses of language, literacy, and learning in society. In examining current curricula and practices related to first language and literacy teaching, it is useful to remember that these instructional practices, similar to other cultural practices, travel across contexts and are often accompanied by ideological struggles (Brandt and Clinton, 2002).

In some contexts, language and literacy education is framed as part of larger political issues – such as the attempt to increase literacy among historically disenfranchised groups as a means of addressing their marginalization. In other contexts, language and literacy teaching is framed by more local, but nonetheless polarizing, debates, such as the focus on the best methods for teaching reading to young children. The myriad factors that affect first language curricula make an examination of first language learning in English a complex task. Therefore, this overview describes the central components of first language curricula in English, focusing on areas of tension and debate.

Curriculum: First Language Teaching across Primary, Intermediate, and Secondary Schooling

In the past 100 years or so, first language teaching has been known as the teaching of reading and writing, language arts, communication, English, and literacy or literacies.

Curricula for first language learning typically include instruction in understanding and producing alphabetic print and oral language as well as the integration of arts – such as drama and visual arts – and disciplinary subject matter. Students are generally asked to make sense of and produce a range of materials including informational texts, research reports, personal reflections, and fictional stories. In the past 10 years or more, first language and literacy instruction has also begun to be influenced by digital technologies and information literacies. First language classrooms may include the use of word-processing and digital-editing software, reading and researching using the Internet or other electronic databases, use of popular culture materials, and new media such as film and video consumption and production.

Issues and tensions in first language teaching shift as children move from the primary years of schooling into secondary schooling. This article, therefore, considers these two stages separately, while also recognizing continuities across the school years.

Primary Language Curricula

Early literacy curricula

Early childhood first language instruction in English often includes opportunities to learn the concepts and conventions of print (and, increasingly, of multimodal texts), shared reading and writing experiences, letter-sound instruction, and instruction in other relationships between spoken and written text. Over the past 40 years, work in emergent literacy (Clay, 1967) has influenced early literacy instruction in many schools by focusing attention on how young children construct their own understandings of print. This research asserts that children come to conventional literacy practices through extensive personal experimentation and that early efforts at representation can be seen as part of a literacy development continuum. These perspectives reject a distinct literate/nonliterate divide and, instead, view literacy learning as a process that begins before and continues after children produce or decode conventional print.

Work in primary literacy instruction has also been deeply influenced by studies of the relationships among home, community, and school language and literacy practices. Even when homes are filled with a range of rich language and literacy practices (Taylor and Dorsey-Gaines, 1988), some children experience a disjuncture between home and school language and literacy practices. While some children make a seamless transition, others struggle to appropriate school-based speaking, reading, and writing practices. This research suggests that children's ease or difficulty with school-based language practices depends on the alignment of literacy events and cultural practices in home and school contexts (Heath, 1983). Unfortunately, some educators and researchers

have interpreted this work to mean that homes must conform to school expectations, thus locating risk within families, rather than focusing on the ways that schools might be more sensitive and responsive to differences between home, community, and school language and literacy practices. Recent work in social literacy theories (Barton *et al.*, 2000) have been informative in illustrating the ways that the activities of reading and writing shape, and are shaped by, social contexts and structures.

A second area of contention in early childhood language education has centered on how official languages or the language of school should be positioned in relation to other languages. This is a particularly contentious issue in countries where English is the dominant or official language of schooling and yet, significant numbers of children enter school speaking other languages. In some regions, teachers are given the option of establishing a porous relationship between English/the official language of schooling and children's other languages. In these classrooms, teachers and students code-switch between official and other languages as needed, allowing students to scaffold their language learning. In other regions, policies and legislation dictate that teachers may only use the dominant or official language of instruction in school. In places where English is the language of schooling, these policies are known as English-only policies. However, because there is much evidence that becoming literate in one's mother tongue is a foundational step toward becoming literate in another language, and because of the important relationship between language and cultural identity, interest in preserving these policies are currently being challenged by newer understandings of the importance and value of mother tongue education (Alidou *et al.*, 2006).

Reading in the Primary Schools

One of the most contentious aspects of primary curricula in English has been how best to teach young children to learn to read, or how to teach them to break the code of printed text. How reading is taught has become a highly politicized topic in many countries, perhaps most notably in the United States. Beginning in the 1950s and extending into the present day, this struggle has often been referred to as the reading wars, and is characterized by dichotomies between phonics or code-based approaches and meaning-centered whole word or whole language perspectives on reading acquisition. The former are often associated with behaviorist or systematic learning theories, and support for this approach is generally drawn from traditional experimental studies of reading. These studies use standardized reading assessments or measures of phonological awareness. A measure of a child's ability to decode words is used to make predictions about how children learn to read. In contrast, proponents of whole language approaches to reading acquisition generally draw on constructivist

theories of learning, such as the work of Vygotsky, and on sociolinguistic theories of alphabetic print. Proponents of this approach draw parallels between learning to read and learning to speak, and emphasize the importance of using naturally occurring texts and authentic reasons to make meaning and to communicate over learning decontextualized phonemes.

In the past decade, governments in the United States and England have adopted national strategies to improve literacy that emphasize the teaching of phonics and other word-level skills. These mandates have thus defined what counts as literacy and have imposed code-oriented approaches to early reading instruction on a great number of public schools. In the United States, the authors of the *Report of the National Reading Panel* published in 2000 suggest that they used a particularly narrow definition of evidenced-based research in order to make a stronger case for their assertions concerning the need for direct instruction in phonics. However, recent initiatives based on these directives have come under considerable scrutiny. For instance, the New York Times (Schemo, 2007) publicized evidence that researchers and consultants hired to advise the US government on its Reading First Program had deep financial ties to publishers of specific reading programs, which has cast doubt on the solely scientific bases of their recommendations.

A less-contested area of reading education is the focus on the development of reading comprehension strategies among students in what is often referred to as the intermediate or middle years of schooling. There is a growing consensus that students need to be able to read a variety of text genres with understanding to be academically successful as they proceed through schooling. While the research base for this perspective is not new, there is a renewed focus on teaching reading comprehension strategies in the schools.

Literature Teaching in the Primary Schools

The history of books written for children is bound up with moral and instructional purposes. In the West, children often either read the same works as adults, such as religious texts, or folktales and fables intended to inculcate young readers with societal values (Zipes, 1983). The more recent notion that children might read for pleasure can be traced back to the 1800s with the arrival of books, such as *Alice in Wonderland* by Lewis Carroll and *The Wizard of Oz* by Frank Baum, in England and America. The use of children's literature in schools was often limited to religious texts and instructional readers until the middle of the last century when children's trade book publishing burgeoned, followed by the advent of literature-based approaches to teaching. Many primary classrooms and school libraries are now filled with trade books of many genres from fantasy to realism that are central to the

language arts curriculum; however, the didactic legacy of children's literature is still evident in many literary works for children, and in attempts to limit access to stories deemed inappropriate for young readers by private groups who attempt to block works of literature from being purchased, taught, or read in schools through censorship campaigns.

Writing Instruction in the Primary Years of Schooling

Writing instruction that emphasizes more than the products – penmanship, traditional grammar, and structured essays – is a relatively recent development in language and literacy education. Beginning in the 1970s, researchers and educators began to pay attention to the processes and contexts of composing texts (Graves, 1975). Research on younger children's writing processes gave rise to the recognition of children's use of recursive subprocesses, such as planning, drafting, revising, and editing. These process approaches to writing have been associated with progressive pedagogies that stress the importance of giving voice to students and valuing process over product. However, this approach has also been subject to various critiques. In particular, some educators and researchers have asserted that a process approach lacks attention to issues of power and difference in language use (e.g., Delpit, 1995). These educators and researchers have argued that if a school adopts a process approach and other progressive pedagogies, some students may not gain facility in dominant forms of literacy that they have not been exposed to outside of school.

A notable response to the inadequacy of process writing approaches to support disadvantaged students is the development of a genre-based approach to writing in several Australian states. This approach, developed in the 1980s (Martin and Rothery, 1986), was based on aspects of systemic functional linguistics (Halliday, 1975). The main argument for genre approaches is the importance of explicitly teaching key features of text structure to those students who do not otherwise have access to genres and registers of written language crucial to success in schooling and many workplaces. The genre approach, sometimes referred to as genres of power, is considered to be a form of critical literacy in which relationships between social context, functions, and structures of texts crucial to their functionality are made visible (Bernstein, 1990); a text-level meta-language is provided that allows students to access, deconstruct, and reconstruct texts, and to understand their own positions in relation to such texts. Critics of the genre approach to writing point to the way that this approach could actually reproduce, rather than transform, current social structures. It has been suggested that if an understanding of genres is not part of a larger critique of relationships among discourse, power, and

society (Luke, 1996), then students are not given an opportunity to critique and rework current unequal social relations. Those who support a genre approach argue that education cannot create equality unless all students are given an opportunity to master academic genres and participate in the production of knowledge (Hasan and Williams, 1996).

Secondary Language Curricula

The core components of the language curricula in secondary schools are writing (composition) and literature instruction. Often included in the curricula are courses in communication and media studies. Key debates and tensions related to secondary language and literacy education include what constitutes the literary canon and literature instruction, issues related to the literacy skills or practices that are most important for future citizens, and, more recently, the role of new technologies and media in the secondary curricula.

The content and structure of the language courses students experience in secondary school often depend on the ways and extent to which students are tracked or streamed – placements that are often linked to issues of race, class, and gender, rather than ability (Oakes *et al.*, 1992). Students who are streamed into academic tracks take courses in literature and composition, and are often taught a very different kind of curriculum than are students who are streamed into vocational tracks. Academically tracked students are more likely to be exposed to a range of titles considered to be the core of the Western literary canon that are often taught more traditionally. Students who are streamed into vocational tracks might take courses that are more remedial in nature, such as communications or courses that focus on functional or work-related literacy practices. These students are often taught through less traditional or more popular texts and a range of instructional methods. In fact, much of the current research and practice in new forms of literacy and multimedia with youth take place among marginalized students and in alternative or community settings.

The Literary Canon and Literature Instruction

One significant area of tension that has arisen over the past 40 years is what constitutes the role and content of literature teaching in the secondary language curriculum. Traditionally, literature teaching in the language of schooling curricula includes the reading and analysis of a range of works, including poetry, drama, fiction, and nonfiction. While the selection of works, beyond mandated textbooks, is often quite open in the primary years of schooling, secondary schools often privilege works that constitute the received literary canon (Applebee, 1993). In recognition of the literary canon as socially constructed

and representative of the interests and values of dominant cultures, some educators have worked to diversify the range of texts that students may study in class. Beyond traditional classics such as Shakespeare or the English Romantic poets, some educators have begun to include authors of color and women authors on their reading lists. In this way, moves to interrupt the canon and arguments for the inclusion of literature from a range of cultural groups seen in universities in the 1980s have been mirrored in primary and secondary schools.

The introduction of new works of literature in the secondary language curricula has been accompanied by new critical approaches to reading and analyzing texts. In the mid-twentieth century, the teaching of literature often focused on teaching students to recreate analyses of texts that were considered valid by renowned literary critics (Hirsch, 1967). Meaning was considered to be stable and students were expected to unlock authors' intended meanings through careful deliberation and a familiarity with other critics' ideas of such texts. More recent approaches include the influences of reader-response theory, cultural theory, and psychoanalytic, Marxist, feminist, and post-structural perspectives. These perspectives have opened up new ways of reading and responding to texts. In particular, these theories have made it possible to see literary works as cultural productions that both reflect and (re) produce specific kinds of social relations. While some secondary educators continue to teach as if there have been no changes in literary theory over the past four or five decades, others have begun experimenting with bringing new texts and new literary theories and perspectives into classrooms.

Adolescent Literacies and Social Futures

In addition to questions concerning the boundaries of the literary canon and how students should analyze a text, secondary school curricula have been influenced by work in content area reading, writing across the curriculum and, more recently, content literacies. This work focuses on the variety of comprehension, writing, and meta-cognitive strategies needed for academic success, given the diverse demands of negotiating texts across the disciplines. Reading skills in these paradigms are often defined as the ability of students to construct, expand, and reflect on the meaning of what they have read in a wide range of texts common both within and beyond school – skills that are assessed internationally by organizations such as the Office of Economic Co-operation and Development – and are seen as central to educating tomorrow's workers.

In the past decade, however, there has been increased attention paid to a broader range of literacy practices found in the lives and communities of adolescents in and out of schools. This area of research, often referred to as adolescent literacies, emphasizes the importance of identity,

culture, and multimedia on young adults' literacy practices (Moje *et al.*, 2000). This research has suggested that the new demands and new roles for youth coming of age in the twenty-first century may not be well served by curricula that focus on print literacy, or an emphasis on individual cognition with individual texts in literacy practices. The hybrid, multiple, and social texts associated with new technologies and media require new ways of reading, writing, viewing, and communicating, thus potentially transforming the epistemological foundations of language and literacy education (Luke, 2003).

While these new conceptions of literacy are exciting, there remain some tensions related to students' access to new media and new technologies, or what some have termed the digital divide. In some contexts, there is a significant gap between those who have access to informational and digital technologies and those who do not. This gap is often rooted in socioeconomic class and gender relations. Secondary school educators are often faced with the task of bridging this gap with few resources and little training. In many regions, students learn far more about using technology at home than they do in school. When educators choose to bring new technologies into schools, they are faced with further issues concerning commodity consumption associated with popular culture and media and how to take up such technologies in a way that allows new practices to emerge rather than just reproducing what students and teachers already know and do. Future educators and researchers are faced with the task of attempting to keep up with current changes in communication practices, integrating new and old language and literacy practices, and reflecting on the implications of their choices for the future lives of their students.

Conclusion

Approaches to teaching language and literacy have been affected by changing cultural and political perspectives and emphases (Lankshear, 1997). These approaches variously include skills emphases (e.g., phonics and basic skills), progressive pedagogies (e.g., whole language and look-say), sociocultural perspectives (e.g., New Literacy Studies and sociolinguistics), and postprogressive or critical approaches (e.g., genre based, critical literacy, and multiliteracies). The defining features of each of these emphases, respectively, represent shifts from a focus on language itself and its codes, to the construction of meaning, to social and cultural contexts of language and literacy learning, to the use of language and literacy to critique a range of texts, and to transform society. In turn, these approaches are often theorized by competing conceptualizations of literacy, learning, and schooling as cultural practices.

However, perhaps the most pressing consideration for first language researchers and educators in the twenty-first century stems from recent assertions that language and literacy policies and practices have a long history of perpetuating social inequality (Blackledge, 2001). While first language and literacy learning is often depicted as indispensable to social and economic participation, and as a means toward contributing to human development and poverty reduction by organizations such as UNESCO, others have argued that such constructions of language or literacy education need to be seen as part of what is referred to as a literacy myth (Graff, 1979). Recent research has begun to question how current discourses of language and literacy learning may obscure the role of economic, political, and social contexts in children's and adolescents' first language learning. As first language educators, we need to be attentive to the way debates about language curricula are discursively framed, the ideologies they carry, and their material consequences for children and youth, if we are to maintain a role for formal education in the creation of a just society.

See also: Assessment in Schools Related To Literacy; Reading; Curriculum Development in the Area of Reading; Early Writing; Learning a Second Language in First Language Environments; Learning to Read; Literature; Primary and Elementary/Middle Grades Reading; Reading Comprehension: Reading For Learning; Writing: Advanced.

Bibliography

- Alidou, H., Boly, A., Brock-Utne, B., *et al.* (2006). *Optimizing Learning and Education in Africa – the Language Factor*. Paris: Association for the Development of Education in Africa.
- Applebee, A. (1993). *Literature in the Secondary School: Studies of Curriculum and Instruction in the United States*. Urbana, IL: National Council of Teachers of English.
- Barton, D., Hamilton, M., and Ivanic, R. (2000). *Situated Literacies: Reading and Writing in Context*. London: Routledge.
- Bernstein, B. (1990). *The Structuring of Pedagogic Discourse: Class, Codes and Control*, vol. IV. London: Routledge.
- Blackledge, A. (2001). Literacy, schooling and ideology in a multilingual state. *Curriculum Journal* 12, 291–312.
- Brandt, D. and Clinton, K. (2002). Limits of the local: Expanding perspectives on literacy as social practice. *Journal of Literacy Research* 34, 337–356.
- Clay, M. (1967). The reading behaviour of five-year-old children: A research report. *New Zealand Journal of Educational Studies* 2, 11–31.
- Delpit, L. (1995). *Other People's Children: Cultural Conflict in the Classroom*. New York: The New Press.
- Freire, P. (1972/2000). *Pedagogy of the Oppressed*. New York: Continuum.
- Graff, H. (1979). *The Literacy Myth: Literacy and Social Structure in the Nineteenth Century City*. New York: Academic Press.
- Graves, D. (1975). An examination of the writing processes of seven-year-old children. *Research in the Teaching of English* 9, 227–241.
- Halliday, M. (1975). *Learning How to Mean: Explorations in the Development of Language*. London: Edward Arnold.

- Hasan, R. and Williams, G. (eds.) (1996). *Literacy in Society*. London: Longman.
- Heath, S. B. (1983). *Ways with Words: Language, Life and Work in Communities and Classrooms*. New York: Cambridge University Press.
- Hirsch, E. D. (1967). *Validity in Interpretation*. New Haven, CT: Yale University Press.
- Lankshear, C. (1997). *Changing Literacies*. Levittown, PA: Open University Press.
- Luke, A. (1996). Genres of power? Literacy education and the production of capital. In Hasan, R. and Williams, G. (eds.) *Literacy in Society*, pp 308–338. London: Longman.
- Luke, C. (2003). Pedagogy, connectivity multimodality, and interdisciplinarity. *Reading Research Quarterly* **36**, 399–403.
- Martin, P. and Rothery, J. (1986). What a functional approach to the writing task can show teachers about 'good writing'. In Couture, B. (ed.) *Functional Approaches to Writing. Research Perspectives*, pp 241–265. London: Pinter.
- Moje, E., Peyton-Young, J., Readence, J., and Moore, D. (2000). Reinventing adolescent literacy for new times: Perennial and millennial issues. *Journal of Adolescent and Adult Literacy* **43**, 400–410.
- Oakes, J., Gamoran, A., and Page, R. (1992). Curriculum differentiation: Opportunities, outcomes, and meanings. In Jackson, P. (ed.) *Handbook of Research on Curriculum*, pp 570–608. New York: Macmillan.
- Schemo, D. J. (2007). In war over teaching reading, a U.S.-local clash. *New York Times*, March 9, 16.
- Taylor, D. and Dorsey-Gaines, C. (1988). *Growing Up Literate: Learning from Inner City Families*. Portsmouth, NH: Heinemann.
- Zipes, J. (1983). *Fairy Tales and the Art of Subversion*. New York: Wildman.
- Ferreiro, E. and Teberosky, A. (1982). *Literacy Before Schooling*. Portsmouth, NH: Heinemann.
- Flood, J., Lapp, D., Squire, J. R., and Jensen, J. M. (2003). *Handbook of Research on Teaching the English Language Arts*. Mahwah, NJ: Erlbaum.
- Framework for Teaching (1998). The National Literacy Strategy. Department of Education and Employment. http://www.standards.dfes.gov.uk/primary/publications/literacy/nls_framework (accessed May 2009).
- Goodman, K. (1986). *What's Whole in Whole Language*. Portsmouth, NH: Heinemann.
- Hull, G. (2003). Youth culture and digital media: New literacies for new times. *Research in the Teaching of English* **38**, 229–233.
- Kamil, M. L., Mosenthal, P. B., Pearson, P. D., and Barr, R. (2000). *Handbook of Reading Research* vol. III. Mahwah, NJ: Erlbaum.
- National Reading Panel (2000). *Teaching children to read: An Evidence Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. Washington, DC: National Institute of Child Health and Human Development, National Institutes of Health.
- New London Group (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review* **66**, 60–92.
- Street, B. (1995). *Social Literacies: Critical Approaches to Literacy in Development, Ethnography and Education*. New York: Longman.
- Swadener, B. B. (1990). Children and families 'at risk': Etiology, critique, and alternative paradigms. *Educational Foundations* **4**, 17–39.
- UNESCO (2007). Literacy. http://portal.unesco.org/education/en/ev.php-URL_ID=40338&URL_DO=DO_TOPIC&URL_SECTION=201.html (accessed May 2009).
- Vygotsky, L. (1978). *Mind in Society*. Cambridge: Harvard University Press.

Further Reading

- Allington, R. (2002). *Big Brother and the National Reading Curriculum: How Ideology Trumped Evidence*. Portsmouth, NH: Heinemann.
- Brandt, D. (2001). *Literacy in American Lives*. New York: Cambridge University Press.
- Britton, J., Burgess, T., Martin, N., McLeod, A., and Rosen, H. (1975). *The Development of Writing Abilities 11–18*. London: MacMillan Education.
- Collins, J. and Blot, R. (2003). *Literacy and Literacies: Texts, Power, and Identity*. Cambridge, UK: Cambridge University Press.
- DeCastell, S. and Luke, A. (1983). Defining "literacy" in North American Schools: Social and historical conditions and consequences. *Journal of Curriculum Studies* **15**, 373–389.
- Eagleton, T. (1996). *Literary Theory: A Reader*, 2nd edn. Minneapolis, MN: University of Minnesota Press.

Relevant Websites

- <http://www.aate.org.au> – Australian Association for the Teaching of English.
- <http://www.alea.edu.au> – Australian Literacy Educators' Association.
- <http://www.csse.ca> – Canadian Society for the Study of Education (CSSE): Language and Literacy Researchers of Canada.
- <http://www.reading.org> – International Reading Association.
- <http://www.nate.org.uk> – National Association for the Teaching of English (UK).
- <http://www.ncte.org> – National Council of Teachers of English (US).
- <http://www.nrconline.org> – National Reading Conference.
- <http://www.nzra.org.nz> – New Zealand Reading Association.
- <http://www.peta.edu.au> – Primary English Teachers Association (Australia).
- <http://www.ukla.org> – United Kingdom Literacy Association.

Learning a Second Language in First Language Environments

E Bernhardt, Stanford University, Stanford, CA, USA

D J Tedick, University of Minnesota, Minneapolis, MN, USA

© 2010 Elsevier Ltd. All rights reserved.

This article uses the 1970s as an embarkation point from which to describe and discuss curriculum selection, development, and implementation in contexts in which second languages are learned in first-language environments. In the US and some European contexts this is referred to as foreign-language education, in other parts of Europe as language education, and in Australia as instruction in languages other than English (LOTE). The decade 1970s denotes a time of massive global movement in immigration and expatriation. Indeed, politics as well as natural and man-made disasters provided the impetus for much of this movement. Yet, at the same time, advances in technology enabled more and more individuals to consider living and working abroad. The broad-based need to prepare learners in traditional schooling contexts for using languages across the globe had a profound impact on language education: the forced march through a grammatical syllabus rooted in the virtue of understanding grammatical structures as grammatical structures had to be jettisoned – it was too costly, too time consuming, and too impractical. The concept of being able to do something with a foreign language took hold and has remained the mantra of language programs since that time.

The time frame also coincides with an awakening of research in sociolinguistics, discourse analysis, and second-language acquisition. Not only was doing a curricular philosophy, but also research indicated that learners did not learn and use language forms in the order in which they were presented in traditional textbooks; that is, they did not learn the present tense and then move on to learn the past. In like manner, learners did not absorb all of the prepositions, then move on to the adverbs, and so forth. In fact, research indicated that learners learn to use multiple grammatical functors and syntactic patterns over time, distributed across the learning of previous forms and often contingent on native-language background, sociolinguistic context, and communicative intention (Doughty and Long, 2003). These kinds of research findings and directions coupled with the need, desire, and ease to move about the globe mandated radically different curricular set-ups, dramatically diverse and varied materials, and profound modifications in teacher-preparation programs that conflate to characterize language teaching in the present day.

Key Concepts in Foreign-Language Teaching and Learning

The Notion of Doing

The intellectual shift from having learners know facts about language structure (such as: How is the periphrastic future formed in Spanish? Or, how is tense marked in Chinese?) to actively being able to use those structures in oral speech was profound. The driving intellectual force behind the shift is exemplified by Hymes' (1972) coinage communicative competence and codified in language teaching theory by Widdowson (1978) and in language teaching curricular models such as notional-functional (Wilkins, 1976). Subsequent to these important contributions, significant project work was conducted across the globe that attempted to describe how learners developed in their abilities to use oral speech. Importantly, it was critical to develop the descriptions apart from textbook, set of objectives, quality of teacher, or length of seat time devoted to language learning. In other words, the objective was to develop descriptors external to particular curricula that would enable a sense of what learners could or could not do with the language that they had learned at any given point in time not what they did or did not know about that language. In the United States this was referred to as the common yardstick and in Europe as common reference levels.

Proficiency Orientation in the US

Foreign-language teaching in the US was fundamentally altered by 'Strength through wisdom', the report of the President's Commission on Foreign Languages and International Studies (1979). This report, which called for a common yardstick, spurred US policy into the direction of developing descriptors for language proficiency. The American Council on the Teaching of Foreign Languages (ACTFL) took leadership in developing the descriptors across multiple languages and across the four skill areas of speaking, listening, reading, and writing. The original set of descriptors, known as the ACTFL Provisional Proficiency Guidelines (1983), were unveiled in 1982 and were adapted from the Foreign Service Institute scale. This scale made provision for 11 major ranges of proficiency, beginning with 0 (no functional ability in the language) to 5 (proficiency equivalent to an educated native speaker), using + designations between levels. The nominals on the

scale are novice (N), intermediate (I), advanced (A), and superior (S), and capture word-level (N), sentence-level (I), and paragraph-level (A) speech, through the upper reaches of highly educated native speech (S). In order to generate data for the scale, a structured interview – an oral proficiency interview (OPI) – that probes through questions and role plays, the function (what one can accomplish with language), content (themes and topics), and accuracy (grammar and discourse) of a speaker – is administered.

The proficiency orientation in the US enabled a national dialog among educators about, for example, how far a secondary school language program could bring learners (through the early intermediate level perhaps) and where post-secondary programs could pick up (at the intermediate level and into the advanced) with study abroad necessary for the development of the upper reaches of language on the scale. With the focus on doing, the concept of grammatical accuracy became only one dimension of language ability rather than the dimension of language ability. Over the years, the descriptors have been modified and refined (with most recent versions of the speaking and writing guidelines available at the ACTFL website), but the basic framework of a scale with ranges of language performances remains intact.

Common European Framework

The concepts embodied in the Common European Framework (CEF) are rooted in urgencies parallel to those identified by the proficiency movement in the US: a need to understand language development independent of particular curricula and teaching methods and a desire to have a common yardstick. The urgency of developing a common yardstick in Europe is/was, of course, much more complicated than that found in the US because development and consensus had to take place across speakers of multiple languages, speech communities, and national boundaries.

Similar to proficiency-based language teaching philosophies, the CEF examines language abilities in multidimensional spaces (Council of Europe, 2001). These spaces entail function, topic, purpose, and modality as well as micro-level grammatical precision. In these basic beliefs, proficiency and the CEF are absolutely identical. In like manner, the basic nominal scale is similar. CEF contains six levels from basic to proficient: basic (A1 and A2); independent (B1 and B2); and proficient (C1 and C2). Basic can be referenced as waystage; independent as threshold; and proficient as mastery. CEF descriptors contrast, however, with proficiency descriptors in that they are much more precise regarding the social contexts in which language is used and much less precise regarding

particular grammatical forms. CEF at A2, for example, notes: “Can communicate in simple and routine tasks requiring a simple and direct exchange of information.” US proficiency descriptors would note the use of the present tense. Importantly, proficiency descriptions often use the command of particular tenses as benchmarks. Ability to narrate in the past, as an example, is often noted as the main marker of crossing a major border, namely from intermediate to advanced. Such a definition is not found in the CEF descriptors. Rather, major boundaries such as moving from waystage to threshold are crossed by abilities to use greater quantities of language and greater social skill such as the ability to gain and hold the floor. A second significant feature that distinguishes the CEF from the American proficiency approach is substantial attention to self-assessment. Accompanying CEF are diagnostic statements of the “I can...” type that empower the learner to participate in a personal understanding of his/her language development and control.

Standards

The 1990s brought the advent of US national standards in all school-based subject-matter areas, including foreign language. The content standards define what students should know and be able to do in foreign-language education (National Standards in Foreign Language Education Project, 1996). They are intended to inform curriculum but do not in and of themselves serve as curriculum. Language-specific standards were unveiled in 1999 for eight modern languages and the classical languages.

The influence of the proficiency orientation in the US is clear in the standards, as the emphasis on doing remains. This influence is particularly prominent in the area of communication, one of five goal areas under which the 11 standards are organized. Yet, the standards reveal a departure from the separate skill areas of listening, speaking, reading, and writing that characterize the proficiency guidelines, to a framework of communicative modes, which emphasize the context and purpose of communication. This framework includes the interpersonal mode (direct oral or written language that reflects spontaneous, two-way communication between individuals in personal contact and involves negotiation of meaning), interpretive mode (interpretation of oral and written text when negotiation of meaning is not possible), and presentational mode (production of written or oral language for an audience without the opportunity for negotiation of meaning). In addition to the communication goal, the standards include goal areas of culture (understanding of the cultural contexts in which language use occurs), connections (links to other disciplines and knowledge that is unavailable to monolingual English speakers), comparisons (overt comparisons both linguistic

and cultural), and communities (culturally appropriate language use with communities outside of the school and the US). Together, these five goal areas are known as the five C's of foreign-language education (National Standards in Foreign Language Education Project, 1999: 31).

The Notion of Meaning

Similar to the shift from knowing to doing, there occurred a shift in language education from a focus on form to a focus on meaning. In a curricular sense, the focus on meaning is perhaps best captured by language-education programs that fall under the umbrella of content-based instruction (CBI; in the US and Canada), or content and language integrated learning (CLIL) (in Europe).

Content-Based Instruction

CBI refers to the integration of language and content in the language curriculum. The notion of what constitutes content in CBI varies, with some indicating that it refers to academic subject matter, while others argue that it need not be academic but can include any topic of interest to learners that is cognitively demanding and engaging for the learner. Met (1999) proposes a continuum of content and language integration with content-driven programs on one end of the continuum and language-driven programs on the other. In content-driven programs, content is given more importance than language, and students and teachers are held accountable for content learning. In language-driven programs, language learning is the priority and content is often used to contextualize the language and is often considered incidental.

CBI program models lie on the continuum depending upon whether they are more content or language driven. Met (1999) argues that immersion is “the most salient example of a content-driven language program” (p. 4). Indeed, in immersion programs the focus is on academic content, which children learn through the medium of the non-native language. More detailed information about immersion programs is discussed below. On the other end of the continuum, Met identifies content-related elementary school foreign-language programs that reinforce subject-matter concepts using the foreign language. For example, learners might review simple addition and subtraction problems as they practice the use of numbers in the foreign language. The emphasis in such programs is undeniably on language learning, and content is merely used as a vehicle for making language use more cognitively engaging.

A range of other CBI programs also lies on the continuum between the two extremes. Sheltered courses are content courses that incorporate a variety of teaching strategies to scaffold instruction and make the content

accessible to the learners, non-native speakers of the foreign language. Such courses might lie closer to the content-driven end of the continuum. The adjunct model lies at the center of the continuum, given its equal emphasis on language and content. In this model, students are enrolled concurrently in two linked courses, one focused on content and taught by a content instructor and the other focused on language and taught by a language instructor. Students are held accountable for content mastery and for the acquisition of academic language proficiency. Content instructors and language instructors share responsibility for student learning; the content instructor evaluates students' content mastery, while the language instructor assesses students' language skills. The adjunct model may combine students who are learning the language used for instruction as well as native speakers of the language. Only non-native speakers of the language are enrolled in the language course, whereas all students, native and non-native speakers alike, take the content course. Both sheltered courses and the adjunct model appear most often in post-secondary settings and are examples of Foreign Language Across the Curriculum (FLAC).

Yet another example of a CBI model is language for special purposes (LSPs) which is also commonly found at the post-secondary level. In this model, learners are introduced to specific linguistic features and discourse attributes related to a particular field, such as business or the health industry. Such courses may be categorized as being more language driven if the emphasis is on language or may lie at the center of the continuum if the course emphasizes both language and content learning.

Immersion

Language immersion education was pioneered in 1965 in Canada as a grassroots, parent-driven response to the perceived ineffectiveness of traditional French instruction. In immersion programs students receive at least 50% of their subject-matter instruction through the medium of a foreign language. The model was introduced in the US in 1971 in Culver City, California and since then has proliferated in the US and around the world.

The design of contemporary immersion programs varies depending upon the extent of instructional time in the immersion language (total vs. partial) and the grade at which the immersion program begins (early, mid, late, or late, late). In total/early immersion programs, the immersion language is used 100% of the time for subject-matter instruction in the early grades, K-1. English is not introduced until grade 2 or later, and is gradually increased so that by grade 6 both the immersion language and English are used approximately 20–50% of the time. In partial immersion programs, the immersion language and English are used for approximately 50% of subject-matter

instruction from the beginning of the program. In mid immersion programs, immersion begins around grade 4 and in late programs from grades 7 to 10, with the assumption that learners have studied the immersion language as a subject in school before the immersion experience begins. Late, late immersion begins at the post-secondary level. Mid, late, and late, late programs are found primarily in Canada, though late immersion programs in English are not uncommon in Hong Kong. Preschool immersion programs are found increasingly around the world and in the US. Irrespective of the model chosen, all immersion programs strive for additive bilingualism, referring to the acquisition of a second language at no expense to the first.

The early theory that drove the establishment of immersion programs argued that second-language acquisition would progress much like first-language acquisition. That is, if students were immersed in the second language, using it to communicate meaningfully about relevant content, they would acquire it. Research has since shown that although immersion students develop high levels of functional proficiency in the immersion language, achieving native-like proficiency in the receptive skills (listening and reading), they do not achieve native-like proficiency in their productive skills (speaking and writing). Immersion learners' language lacks grammatical accuracy, lexical specificity, and becomes increasingly Anglicized over time, particularly as instructional time in the immersion language decreases. Immersion scholars and teacher educators have therefore long advocated for a systematic and overt focus on language (form) in the context of content instruction (meaning). Thus, it is no longer a notion of meaning or form but rather meaning and form. Such seamless integration of language and content in immersion and other content-based instruction teaching environments is the ideal.

Supporting Language Teachers

Teacher Preparation

Teacher-education programs are designed for preservice foreign-language teachers and lead to an initial teaching license or certification. In the US most teacher-education programs exist at the undergraduate level, where prospective teachers pursue a major in the language they wish to teach and, often, a minor in education. They typically require a methods course and student teaching for a term, and these experiences constitute prospective teachers' foreign-language-specific teacher preparation. In this model, language departments are responsible for providing prospective teachers with requisite content knowledge and opportunities to develop proficiency in the target language, while departments of education are responsible for providing foundational coursework in

education and meeting state standards for teacher preparation, and the school-based practitioner is responsible for mentoring the student teacher in the context of the student-teaching experience (Schrier, 1994).

Some language-teacher-education programs in the US are at the postbaccalaureate level and require language proficiency as a prerequisite for admission to the program. Often housed in schools of education, these programs offer teacher candidates intense and extended coursework focused on foreign-language curriculum development and assessment, language pedagogy, culture integration, and general educational foundations along with concurrent, extensive school-based experiences. Such programs were developed in the late 1980s as a result of the Holmes Group report (Holmes Group Executive Board, 1986), which emphasized the need for more professionalism in teacher preparation.

Teacher preparation for immersion education takes a decidedly different direction. In Canada, programs designed specifically to prepare teachers for language immersion programs have long been available. In the US, immersion teachers are required to be licensed in elementary education or specific subject matters for the secondary level. The vast majority of states offers no specific programs to prepare immersion teachers even though it is established that immersion teaching requires a unique set of skills and a knowledge base that is different from that needed for regular classroom teaching or foreign-language teaching (Hartiala, 2000). The only state that offers a program to prepare immersion teachers is Hawaii, where an undergraduate program builds proficiency in the Hawaiian language while leading to an immersion teaching license. A similar program is offered in Australia, where teacher candidates increase their proficiency in Japanese while completing requirements to become teachers of Japanese either in an immersion or regular school setting. This Language and Culture Initial Teacher Education Program (LACITEP) was established in response to Australia's aggressive national policy agenda regarding the teaching of LOTE (Erben, 2005).

Just as there has been a shift in language education to emphasize the notions of doing and meaning, a parallel shift in language teacher education has occurred. There have been calls for moving beyond method to embedding a focus on understanding language teaching as it occurs within social, cultural, historical, and political contexts to developing reflective practitioners, and to striving for a critical pedagogy within both language teacher education as well as language teaching (Tedick and Walker, 1994).

Infrastructure for Language Teaching

Teacher development has become more centralized over the years. In the 1990s, the growth of National Language Centers as well as university language centers has

provided a dynamism and a research-based focus on language learning and teaching that did not exist in prior decades. In the US, a network of foreign language resource centers provides intensive workshop opportunities for teachers as well as for materials development and curriculum development assistance. In Canada, the Modern Language Centre has played an important role in fostering bilingual language policy as well as conducting research into different instructional models for the development of bilingualism. In the European Union, a significant multinational effort runs in parallel. It includes the important dimension of accrediting cross-national language programs. The National Centre for Languages in the United Kingdom is one example as are endeavors by the Federation of Modern Language Teachers Association in Australia. In like manner, individual universities have tended toward centralizing their language programs, making professional development and enrollment play more significant roles in the management of large language programs.

Ongoing Dilemmas in Language Teaching

The very theme of this article, learning a second language in a first-language environment, reveals the two most significant challenges encountered in language teaching: teacher language proficiency and student access to immediate, authentic, and copious native-language samples. Teachers who teach a language in a first-language environment are very often non-native speakers of that language. This demographic by and large makes for excellent, empathic teachers who are sensitive to grammatical and social contrasts between the native language and the second language. The demographic also means that these teachers often have no language example other than themselves, creating a constraining situation linguistically and culturally. The only way to break the constraints is by frequent and sustained interaction with native speakers; such interaction often implies lengthy stays in another country. The costs of such sustained interaction are not inconsiderable. Monetary costs involve living expenses, room and board, and sometimes tuition costs; these are inconsequential when compared with the personal costs of time spent apart from family, community, and work. Yet, without spending significant time in a setting where the language is spoken at fairly frequent intervals, language teachers face issues of credibility with their own students, a potential lack of confidence in their speaking abilities, and feelings of professional inferiority. As language teaching grows on a global scale, these issues will be exacerbated (Llurda, 2005).

The second major dilemma presented by language teaching is access to useful, appropriate, and plentiful language materials in all modalities. In the developed

world, almost all contemporary language-learning materials contain video components and digitized material for drills and exercises; yet, this material ages quickly. In this regard, a call to use the Internet seems to be a quick fix. Problematic is that the research base available from which to understand the impact of these technologies on the development of language proficiency is in its infancy. In addition, of course, confronting the issue of access to copious contemporary language examples in the underdeveloped world is a different dimension to the dilemma. Given that electricity is still not universally available and reliable, to speak of Internet- and other digitally-delivered materials to buttress language instruction borders on the absurd.

Caveats

This article does not presume to characterize every development in language teaching over the past decades. It provides the broad brush strokes of curricular trends and issues in language teaching. In fact, in spite of the important trends and issues in language teaching discussed above, many language programs around the world – whether they be in elementary or secondary and post-secondary institutions – remain decidedly grammar driven. This phenomenon occurs for a variety of reasons: lack of opportunities for professional development across the profession, lack of teaching materials that are not grammar driven, the tendency to teach the way one was taught, and the excessive teaching and student load placed on teachers, to name a few. Urging teachers to move from formal grammar instruction to operating with language contextually is demanding and exhausting. The fundamental belief is, however, that the language abilities of learners taught to use language in content or in context are substantial, sophisticated, and usable.

Bibliography

- ACTFL (American Council on the Teaching of Foreign Languages) (1983). *ACTFL Provisional Proficiency Guidelines*. Hastings-on-Hudson, NY: American Council on the Teaching of Foreign Languages.
- Council of Europe (2001). *Common European Framework of Reference for Language: Learning, Teaching, and Assessment*. Cambridge: Cambridge University Press.
- Doughty, C. J. and Long, M. H. (eds.) (2003). *Handbook of Second Language Acquisition*. Malden, MA: Blackwell.
- Erben, T. (2005). Teacher education through immersion and immersion teacher education: An Australian case. In Tedick, D. J. (ed.) *Second Language Teacher Education: International Perspectives*, pp 281–294. Mahwah, NJ: Erlbaum.
- Hartala, A.-K. (2000). *Acquisition of Teaching Expertise in Content and Language Integrated Learning*. Turku: Painosalama Oy.
- Holmes Group Executive Board (1986). *Tomorrow's Teachers: A Report of the Holmes Group*. East Lansing, MI: Holmes Group.

- Hymes, D. (1972). Models of the interaction of language and social life. In Gumperz, J. and Hymes, D. (eds.) *Directions in Sociolinguistics: The Ethnography of Communication*, pp 35–71. New York: Holt, Rinehart, Winston.
- Llurda, E. (ed.) (2005). *Non-Native Language Teachers: Perceptions, Challenges, Contributions to the Profession*. New York: Springer.
- Met, M. (1999). *Content-Based Instruction: Defining Terms, Making Decisions*. Washington, DC: The National Foreign Language Center.
- National Standards in Foreign Language Education Project (1996). *Standards for Foreign Language Learning in the 21st Century*. Yonkers, NY: ACTFL.
- National Standards in Foreign Language Education Project (1999). *Standards for Foreign Language Learning in the 21st Century*. Yonkers, NY: ACTFL.
- President's Commission on Foreign Language and International Studies (1979). *Strength through Wisdom: A Critique of U.S. Capability (report)*. Washington, DC: US Government Printing Office.
- Schrier, L. L. (1994). Understanding the foreign language teacher education process. *ADFL Bulletin* 24, 69–74.
- Tedick, D. J. and Walker, C. L. (1994). Second language teacher education: The problems that plague us. *Modern Language Journal* 78, 300–312.
- Widdowson, H. G. (1978). *Teaching Language as Communication*. Oxford: Oxford University Press.
- Wilkins, D. A. (1976). *Notional Syllabuses*. London: Oxford University Press.

Further Reading

- Björklund, S., Mård-Miettinen, K., Bergström, M., and Södergård, M. (eds.) (2006). *Exploring Dual-Focused Education: Integrating*

- Language and Content for Individual and Societal Needs*. Vaasa: Vaasan Yliopiston Julkaisuja.
- Cloud, N., Genesee, F., and Hamayan, E. (2000). *Dual Language Instruction: A Handbook for Enriched Education*. Boston, MA: Heinle and Heinle.
- Doughty, C. and Williams, J. (eds.) (1998). *Focus on Form in Classroom Second Language Acquisition*. NY: Cambridge University Press.
- Genesee, F. (1987). *Learning through Two Languages: Studies of Immersion and Bilingual Education*. Rowley, MA: Newbury House.
- Hall Haley, M. and Austin, T. Y. (2004). *Content-Based Second Language Teaching and Learning: An interactive Approach*. Boston, MA: Pearson Education.
- Hinkel, E. (ed.) (2005). *Handbook of Research in Second Language Teaching and Learning*. Mahwah, NJ: Erlbaum.
- Reagan, T. (2002). *Language, Education, and Ideology*. Westport, CT: Praeger.
- Shrum, J. L. and Glisan, E. (2005). *Teacher's Handbook: Contextualized Language Instruction*, 3rd edn. Boston, MA: Heinle.
- Ur, P. (2001). *A Course in Language Teaching*. New York: Cambridge University Press.

Relevant Websites

- <http://www.actfl.org> – American Council on the Teaching of Foreign Languages.
- <http://www.afmlta.asn.au> – Australian Federation of Modern Language Teachers Associations Inc.
- <http://www.cilt.org.uk> – CILT, The National Centre for Languages.
- <http://nflrc.msu.edu> – Foreign Language Resource Centers.
- <http://www.oise.utoronto.ca> – Ontario Institute for Studies in Education of the University of Toronto, Modern Language Centre.

Literature

S Hynds, Syracuse, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

This article argues that changes in literature curricula have typically reflected major historical and political developments within particular regions of the world. It begins with a brief overview of major shifts in literature curricula in the United States, followed by a comparative discussion of literature curricula internationally.

A Brief History of Literature and Its Teaching in the United States

In the United States, the purpose of reading in schools has shifted from the learning of religious and moral precepts in the seventeenth, eighteenth, and early nineteenth centuries, to the development of personal and social skills during the progressivist and neo-progressivist movements of the twentieth century, and finally, to the development of critical awareness and social transformation inspired by contemporary literary theory in the latter part of the twentieth and early twenty-first century. As the following discussion will reveal, this movement has been neither steady nor stable, but has reflected the ebb and flow of historical, economic, and political events.

From Primers to Anthologies: Literature Textbooks and the Purposes of Teaching

The early primers in colonial America were composed of precepts, prayers, and biblical excerpts designed to teach Christian ethics and ideals. Throughout the seventeenth and much of the eighteenth century, literary texts, as currently defined, were not part of the curriculum. These texts did not make their official entrance into American classrooms until the early nineteenth century, when various readers and textbooks from related fields such as rhetoric began to depart from the typical religious and moral fare to offer excerpts of poetry, prose, and other types of literature designed to improve the reading, memorization, and elocution skills of students. As the study of literature began to break away from its roots in reading, grammar, and rhetoric in the nineteenth century, a number of literature textbooks emerged. Most were historical and biographical discussions of literary periods and authors rather than collections of literary texts. Eventually, a few textbooks began to include excerpts from British, American, and even world literature. Nevertheless, throughout the mid-nineteenth and early twentieth centuries, literature textbooks remained narrowly confined to historical and biographical information with

an occasional smattering of classic British texts. Despite increasing racial and ethnic diversity in the population during that time, definitions of what counted as literature remained narrow and myopic. It was not until the latter part of the twentieth century that the literary canon expanded in the United States to include authors, protagonists, and content from beyond the English-speaking world. As the following sections will demonstrate, the teaching of literature in American schools, colleges, and universities during the twentieth century has undergone several shifts in focus and methodology.

Text-Based Approaches to Literature Teaching

From the 1920s to the 1960s, New Criticism emerged as a reaction to the preoccupation with the historical and biographical study of literature. Wimsatt and Beardsley (1946a, 1946b) popularized the notion that neither the author's intention (the intentional fallacy) nor the reader's emotional response (the affective fallacy) was relevant to the reading of literature. Borrowing from Richards' (1929) classic work, *Practical Criticism*, the New Critics believed that texts should be examined and analyzed apart from published biographical scholarship or contextual factors external to the work itself. As a result of New Criticism's powerful influence, literature curricula focused on the study of literary devices, genre characteristics, and close textual analysis. Even with the popularity of reader-response criticism in the early 1960s, New Criticism still exerts considerable influence on secondary and college/university literature curricula in the United States (Applebee, 1993).

Reader-Centered Approaches

A move from text-centered to reader-centered teaching began around the time of the Dartmouth Conference in the late 1960s, when a group of teachers and scholars from America and Britain were called together through the joint efforts of the Modern Language Association, The National Council of Teachers of English, and The National Association for the Teaching of English, for the purpose of defining English as a content area and discussing how it should be taught. The conference both reflected and promoted a revival of educational progressivism, an interest in cognitive developmental psychology, and what has come to be called constructivist education. A concern for the

cognitive development and esthetic/emotional needs of readers began to supplement, if not displace, the practice of close textual analysis and the study of literary forms in many literature classrooms. In describing the work of the Dartmouth conference, Squire (1968) noted that, as a result of Dartmouth, a new concern for “imaginative and emotional education” (p. 526) and the diminishing importance of British literature courses in college curricula had begun to displace discipline-centered approaches to English in the United States.

Under various umbrella terms, such as response-based, child-centered, learner-centered, or whole language instruction, learner-centered teachers assumed that students learn best when reading authentic unexcerpted literary texts that appeal to their unique learning practices, preferences, and developmental needs. The mass marketing of young adult and children’s literature also began to influence the available literature in schools. On the elementary level, teachers began to replace the skills-based basal readers with tradebooks or anthologies featuring authentic children’s literature. On the secondary level, the reader-response movement promoted the teaching of fiction written specifically for and about adolescents and privileged personal esthetic response over close textual analysis. During this time, the notion of multiculturalism emerged as a way of developing cultural understanding and tolerance among America’s youth. Today, this trend toward cultural inclusivity is reflected in the growing cultural, racial, and ethnic diversity, as well as the inclusion of female authors and protagonists, in the literary offerings of contemporary literature anthologies.

Critical Theory, Politics, and the Teaching of Literature

Toward the end of the 1980s, the proliferation of critical theories such as postmodernism, Marxism, cultural studies, feminism, and gender studies has spurred a redefinition of multiculturalism from a cultural tolerance to a social transformation perspective. Much has been written in the past two decades about literature and literacy as the cornerstone of a social justice, antiracist, antibias, or service learning curriculum. Teaching literature from this more critically and politically motivated perspective means redefining reading beyond a search for meaning to a way of acting upon and changing the world. Proponents of this approach argue for expanding the literary canon to include not only literature from a variety of racial, cultural, and ethnic perspectives, but also texts that challenge typical racist, sexist, homophobic, and ableist stereotypes. As with all curricular innovations, however, this movement has been neither orderly nor far reaching. Despite a growing interest in contemporary critical theory in secondary teacher education, curricular

reforms based on contemporary literary theory have been largely confined to the college and university level, and a great many institutions of higher education remain tied to genre-based or historically organized literature curricula. Similarly, teachers in America’s secondary schools have been slow to embrace contemporary critical perspectives, continuing instead to rely on some combination of reader response and New Criticism – approaches often in direct conflict with each other (Applebee, 1993).

Literature and Its Teaching in International Perspective

Just as the American literary canon has remained relatively ethnocentric throughout its history, prior to the 1970s, few scholarly publications focused on how literature was taught outside the American and British tradition or on cross-national analyses of literature curricula. In 1968, the United States Office of Education established the Institute for International Studies, and grant monies were allocated for studies on international and comparative education “in countries of interest to the United States and concerning which little useful material was readily available in English” (Leestma, 1981: 273). From the National Defense Education Act of 1958 until the early 1980s, much available research on comparative education has focused on foreign languages, particularly those uncommonly taught languages outside Western Europe (Leestma, 1981: 273).

In the early 1970s, as part of this initiative, a comparative analysis of international literature curricula was sponsored by the Institute of International Studies. In *Literature education in ten countries* (Purves, 1973), the report of this 9-year study, Alan Purves and colleagues concluded that schools have minimal effect on students’ interest in literature, but do impart a preferred style of response. The report concluded that American schools seem to do this earlier and with more impact than schools in other countries. In a later publication, Purves (1975) described three deep structures in the literature curricula of the countries in the International Association for the Evaluation of Educational Achievement (IEA) study: the imitative structure, the analytic structure, and the generative structure. Italy and Iran were closest to the imitative; Belgium and Chile, analytic; and England, generative. Purves noted the importance of a country’s curriculum in shaping the cognitive styles of students.

Although ambitious in scope, the IEA study was based on survey data. Since the time of the IEA study, no large-scale comparative analyses of literature education have been conducted. The extant literature tends to include comparisons of one or two countries (typically, those with a similar linguistic base) or, more typically, historical examinations of literature curricula within one country

or region (Applebee, 1993). Thus, it is difficult to discuss national and cultural differences in any comprehensive way. The remaining sections are an attempt to pull together a group of disparate studies and draw cross-national comparisons. The discussion centers on some key issues relevant to the political, social, historical, and economic influences on literature and its teaching.

Literature, Colonialism, and Revolution

In colonial and postcolonial periods, literature serves as a tool of colonizers and revolutionaries alike. In colonial times, literature was often censored when it reflected negative aspects of colonization. For example, D'Arcens (2000) has described attempts to repress or write over England's violent colonization of aboriginal people in Australia's university literature curricula during the nineteenth century. In post-revolutionary societies, literature often serves as a vehicle to promote political ideals. In describing the literature curriculum in Russia prior to the dissolution of the Soviet Union, Hopkins (1974) noted that literature chosen for classroom study was "oriented toward projecting an image – the dark pre-Revolutionary past vs. the glorious present" (p. 49). Under autocratic regimes, literature often serves an overt role in advancing a political or social agenda. Perhaps nowhere is this more apparent than during the African independence movements in the 1960s, many of which were followed by the institution of autocratic governments. For example, under the socialist regime of Sékou Touré in Guinea, writings by African authors in French were criticized for their precolonial associations, and students were expected to memorize Touré's poems and other writings in favor of Francophone literature by African authors.

Often, well past a country's independence, it is difficult for indigenous or national languages and literatures to make their way into the curriculum. Precolonial notions often prevail long after political revolutions have rendered them obsolete. As Bray (1993) notes, the term independence is often slippery, and relationships in post-colonial nations can range from full sovereign independence to sovereignty with strong ties to the former colonial power. For example, it was not until the latter part of the twentieth century that American literature, and later, the literatures of indigenous peoples, made their way into the curriculum of the United States. Although the *Journal of American Literature* was founded in 1929, it was not until mid-century that American literature began to enter high school and college curricula. In the same vein, Robinson (1987) has discussed the slow emergence of Canadian literature from the British tradition in Canada during the 1950s and 1960s.

Bray (1993) distinguishes between colonialism and neocolonialism, which he describes as "the control of

states by external powers despite the formal appearance of constitutional independence" (p. 334). Neocolonial influences have been exerted by countries such as Great Britain and the United States, which have no formal colonial relationships, but maintain economic superiority over smaller or developing nations. English, for example, is still a mandatory subject in many parts of the globe, and literature from the American and British tradition forms a significant core of literature curricula in a great many formerly colonized and noncolonized countries. For example, after Kenya's independence from Great Britain, the literature curriculum remained largely British due to its continued economic dependence and the lingering influence of the Cambridge Local Examinations Syndicate. Similarly, students in post-revolutionary Gabon continued to study a combination of Francophone and African literature due to the country's strong economic and social ties with France. In general, across Africa, few works by African or African diaspora writers have made it into classrooms, despite the achievement of independence in countries across the continent.

Thus, in many postcolonial societies, literature is a vehicle for establishing a national identity or promoting political ideologies. Notions of high culture are often tied to the languages and literatures of colonial or neocolonial powers and persist long after political revolutions should have diminished or negated their influence.

Political Stability and the Literature Curriculum

English-speaking countries such as the United States, Canada, Australia, New Zealand, and the United Kingdom have enjoyed relatively unchallenged political and economic stability. Thus, curricular debates in these countries often center on philosophical and epistemological questions such as enlarging the literary canon to meet the tastes and cultural backgrounds of students or the teaching of skills and strategies versus the development of esthetic sensibilities. In regions marked by political instability, large-scale curricular decisions often have dire consequences for citizens and schools alike. In such contexts, literature is often used as a vehicle for imparting political ideologies or suppressing antigovernment ideas, as, for example, the censorship campaigns in Nazi Germany prior to and during World War II.

In comparing the literature curricula of Greece and Great Britain, Hodolidou (1995) noted that the "political and constitutional stability of England is relatively unknown to Greece which has faced major political and constitutional changes during the last fifty years" (p. 86). The Greek education system is controlled by the Ministry of National Education and Religion, and strongly influenced by the Greek Orthodox Church. During the military dictatorship

prior to 1974, politically progressive ideas were censored and the prescribed literary canon served as a tool for the promotion of a unified Greek identity. By contrast, although the literary canons in Britain and the United States remain relatively narrow, politically progressive ideas have not been overtly censored, and teachers are relatively free to debate issues of inclusivity and exclusivity in the literary canon (Hodolidou, 1995).

Literature and the Economy

It is important to recognize that national identities are created as much by economic exigencies as they are by natural geographic boundaries. Movements toward national unity are often thinly veiled schemes to promote economic and political stability or supremacy. Often during challenging economic periods, the literary canon is restricted or narrowed, and concerns for personal relevance and artistic creativity in the teaching and writing of literature are stifled. In the recent history of the United States, political documents such as ‘A nation at risk’ (National Commission on Excellence in Education, 1983) and reform movements such as The Goals 2000: Educate America Act of 1994 and The No Child Left Behind Act of 2002 have stemmed from a desire to promote global competition, produce capable or compliant workers, and compete internationally on various measures of achievement. A similar move toward accountability and curricular reform occurred in Great Britain under the Education Reform Act of 1988 with the creation of a national curriculum and a mandatory testing program.

In other countries around the globe, large-scale attempts at curricular reform aimed at the production of a capable workforce have coincided with periods of economic insecurity or a perceived lack of global competitiveness. Under autocratic regimes, these reforms have had troubling consequences for schools and citizens alike. During the Chinese Cultural Revolution of the 1950s, for example, Mao Tse-Tung abolished traditional university and school programs, sending students and teachers from the cities to rural areas to be re-educated by peasant farmers. Often this re-education was as a response to perceived crises such as the need for steel production or the building of hydraulic power or irrigation systems. The Cultural Revolution resulted in the abolition of humanities programs in colleges and universities, the shortening of time spent in primary and secondary schools, and the persecution of writers who did not promote an overtly communist agenda. Similarly, inspired by China’s example, African socialist Julius Nyerere in Tanzania required all teachers to teach a curriculum based up socialist notions of practical action, specifically targeted toward the development of agricultural knowledge and farming skills. Cuba’s utilitarian education

movement in the mid-1960s resulted in a similar closing of schools and a migration of young people to rural areas to work and participate in a national literacy campaign. Due to the heavy emphasis on vocational training and applied science, enrollments in the humanities sharply declined in these countries. Similar radical movements were enacted in countries such as Brazil and Nicaragua. As socialist or communist reforms failed to exact their predicted economic effects, academic literature study slowly reentered the curriculum. In countries such as China and Cuba, basic literacy levels were raised among the population as a whole, but the neglect of more formal education exacted a heavy price in terms of intellectual development among all citizens and national economic competitiveness.

Recently, in response to global economic competition, the European Union movement has forced European countries with disparate linguistic, cultural, religious, and political traditions to seek common ground in a process similar to that experienced by American colonists prior to the creation of the United States and the nations of Africa during the pan-African movement of the 1960s. Hodolidou (1995) has suggested that literature curricula might play an important role in “promoting a new European identity” (p. 84).

Literature in Crisis: War and Other National Emergencies

Concerns for literature’s role in the transmission of national heritage as well as in the teaching and testing of literary analytic skills often arise due to fears over wars or terrorist acts, large or unusual immigration patterns, or changing racial and ethnic demographics. Learner-centered progressivist movements associated with the political left in the United States have often fallen out of favor as national insecurities have polarized public opinion toward a more skills-based accountability system in education. For example, in the United States, the launching of the Soviet satellite, *Sputnik* in the 1950s coincided with calls for intellectual rigor in the field of English through national programs such as Project English. The periods surrounding World War I, World War II, and the 2001 bombing of the World Trade Center all coincided with calls for national standards and testing programs, as well as with the need for some version of cultural literacy (Hirsch, 1987) as evidenced by familiarity with texts from an accepted literary canon.

Notably, however, one of the first efforts to broaden the American literary canon occurred in the period surrounding World War II, when literature began to be viewed as a vehicle for “sustaining peace, brotherhood, and democracy” (Leary, 1943: 88). This call for the inclusion of literature from non-English-speaking authors arose

more from a need to understand (and presumably compete with) foreign nations than a desire for peace, cultural tolerance, or social change; thus, it did not result in a significant expansion of the canon. In 1940, for example, the National Council for Teacher Education (NCTE) Committee on International Relations published two narrowly confined lists of foreign books in English, one encompassing Latin-American, and the other Russian literature (Leary, 1943). A glimmer of what has come to be called multiculturalism in the literature curriculum began in the early 1970s with arguments for the inclusion of world literature in the college and high school classroom (Dieterich, 1973). Citing the proliferation of new technologies, developing nations, and other signs of global dependence, Dieterich argued that when children are exposed only to British and American literature, "[T]hey learn that the rest of the world is incapable of writing literature which merits inclusion in their curriculum; that only the British and, to a lesser extent, Americans are capable of producing works of real literary value" (p. 482).

Literature and the Politics of Difference

Perhaps ironically, growing cultural, racial, or ethnic diversity within a nation has often yielded a reification, rather than a transformation of the literary canon. In the United States, the mass immigration of Chinese people in the mid-nineteenth century and the immigration of Italians, Slavs, and Jewish people from southern Europe in the early twentieth century spurred the first English-only legislation in American schools and a federal English-speaking requirement for citizenship. Citizenship and access to education were not granted to African American, Native American, Asian, or Mexican people until the mid-to-latter part of the nineteenth century. During the nineteenth century, the common schools movement was instituted as an attempt to Anglicize Native American and immigrant children. The teaching of literature during this period often focused on the creation of an American identity and culture, resulting from the growing fear that a rising immigrant population would displace American workers and enact a radical shift in political ideology among America's poorer classes.

More recently, the move toward high-stakes testing and basic skills instruction in the United States has been further intensified by the large numbers of children from non-English-speaking households entering American schools. Shortly after 'A nation at risk' was published with its conclusions about a rising tide of mediocrity (National Commission on Excellence in Education, 1983: 5) in American schools, the National Census Bureau reported a steep rise in the numbers of foreign-born residents in the United States. Around this time, bilingual education programs were dismantled in Florida and California.

On the other hand, what has been called the browning of America has resulted in a significant broadening of the literary canon in America's primary and secondary schools to include texts written in languages other than English, as well as a wider cultural and ethnic representation in contemporary literature anthologies. Teachers in Great Britain are caught in a similar struggle to include the literature of immigrant and other minorities into the British curriculum. It is not coincidental that, as the United States and the United Kingdom have become more dependent upon formerly disenfranchised Third World nations for economic success and political security, concerns for inclusion of the underrepresented have increased. Competing concerns for the narrowing or the widening of the literary canon have always existed, and such clashes and controversies will no doubt continue far into the future.

In sum, literature curricula are never politically neutral. Within the United States, the role of literature in schools has shifted from promoting social morality and a unified national identity to enhancing the esthetic, social, and cognitive sensibilities of individual learners. The mid-to-latter twentieth century has witnessed a move from literature as personal identification and a representation (or reification) of culture toward a view of literature as a vehicle for cultural critique and transformation. Within and beyond the English-speaking world, curricular changes are rarely bounded by distinct historical periods, but reflect the changing political, economic, and historic landscapes of particular countries. In periods of colonialism, literature tends to reify colonial values. Revolutions and political instability often result in governmental control of literature curricula for the purpose of advancing social reforms or political agendas. Economic uncertainty and national emergencies often lead to calls for national unity or increasing economic productivity through national literature curricula or standardized assessment programs. Large or unusual patterns of immigration or increasing cultural diversity have two very disparate effects. On the one hand, cultural changes often result in arguments for the creation of a national identity through familiarization with an accepted literary canon; on the other hand, increasing diversity often results in calls for the expansion of the literary canon to reflect a changing cultural, racial, and ethnic landscape.

Several recommendations can be drawn from the foregoing review. First is a pressing need for systematic study of literature teaching outside the American and British tradition and for cross-national studies of literature curricula from a variety of research methodologies. Without such large-scale analyses, it is difficult to understand the exact ways in which historic, geographic, political, and social exigencies shape literature curricula across cultural boundaries. It is also imperative to recognize that the politicization of literature and governmental censorship

of politically unpopular ideals in literary texts cannot be safely relegated to the past. In a time of great instability, evidenced by war, genocide, and international terrorism, nations must be eternally vigilant against repression in all of its forms. Even in the secular educational systems of multiparty states, researchers and scholars are often indirectly pushed to reify a political agenda when particular research perspectives and methodologies are kept outside the cycle of government funding. While scholars may exercise a certain degree of academic freedom within the walls of universities, it is notable that much academic discourse gets ignored by political administrations, school districts, and popular media. For example, no large-scale studies exist on the effects of high-stakes testing on the ways in which literary texts are read, taught, and evaluated in classrooms. Given the politically charged nature of the topic, it is questionable whether government funding would be available for such inquiry. Perhaps now, more than ever in the history of the world, members of the academy must bring the principles of cultural critique to their own work, redefining scholarship from a search for information to a form of political action – a way of transforming the world.

See also: Assessment in Schools – Creative Subjects; Curriculum in Postcolonial Contexts; Primary and Elementary/Middle Grades Reading.

Bibliography

- Applebee, A. N. (1993). *Literature in the Secondary School: Studies of Curriculum and Instruction in the United States*. Urbana, IL: National Council of Teachers of English.
- Bray, M. (1993). Education and the vestiges of colonialism: Self-determination, neocolonialism and dependency in the South Pacific. *Comparative Education* **29**, 333–348.
- D'Arcens, L. (2000). From holy war to border skirmish: The colonial chivalry of Sydney's first professors. *Journal of Medieval and Early Modern Studies* **30**, 519–545.
- Dieterich, D. J. (1973). ERIC/RCS report: The lessons of silence in the English classroom. *The English Journal* **62**, 482–488.
- Hirsch, E. D. Jr. (1987). *Cultural Literacy: What Every American Needs to Know*. Boston: Houghton Mifflin.
- Hodolidou, E. (1995). Literature in the English and Greek secondary curriculum: A comparative study. *Tertium Comparationis Journal für Internationale Bildungsforschung* **1**, 81–96.
- Hopkins, E. (1974). Literature in the schools of the Soviet Union. *Comparative Education* **10**, 25–34.
- Leary, B. E. (1943). Literature in school instruction. *Review of Educational Research* **13**, 88–101.
- Leestma, R. (1981). Comparative and international education in the U.S. Office of Education: A bibliography of studies and publications, 1968–1980. *Comparative Education Review* **25**, 272–288.
- National Commission on Excellence in Education (1983). *A Nation at Risk: A Report to the Nation*. Washington, DC: U.S. Government Printing Office.
- Purves, A. C. (1973). Literature education in ten countries, an empirical study. *International Studies in Evaluation*. New York: Wiley.
- Purves, A. C. (1975). Culture and the deep structure of the literature curriculum. *Curriculum Theory Network* **5**, 139–150.
- Richards, I. A. (1929). *Practical Criticism: A Study of Literary Judgment*. London: Routledge and Kegan Paul.
- Robinson, S. D. (1987). Canadian literature in Canadian schools: A response to Harker. *Canadian Journal of Education/Revue Canadienne de l'éducation* **12**, 433–435.
- Squire, J. R. (1968). The running water and the standing stone. *Publications of the Modern Language Association of America* **83**, 523–529.
- Wimsatt, W. K. and Beardsley, M. C. (1946). The affective fallacy. *Sewanee Review* **14**, 468–488.
- Wimsatt, W. K. and Beardsley, M. C. (1946). The intentional fallacy. *Sewanee Review* **54**, 468–488.

Further Reading

- Bradley, S. (1947). The preparation of the high-school teacher in American literature. *College English* **8**, 365–370.
- Harris, J. (1991). After Dartmouth: Growth and conflict in English. *College English* **53**, 631–646.
- Kaba, L. (1976). The cultural revolution, artistic creativity, and freedom of expression in Guinea. *The Journal of Modern African Studies* **14**, 201–218.
- Lillis, K. M. (1985). Processes of secondary curriculum innovation in Kenya. *Comparative Education Review* **29**, 80–96.
- Nietz, J. A. (1961). *The Evolution of American Secondary School Textbooks; Rhetoric and Literature, Algebra, Geometry, Natural History (Zoology), Botany, Natural Philosophy (Physics), Chemistry, Latin and Greek, French, German and World History as Taught in American Latin Grammar School Academies and Early High Schools before 1900*. Pittsburgh: University of Pittsburgh Press.
- Ogden, J. (1982). The Africanization of the curriculum in Gabon. *The French Review* **55**, 855–861.
- Piaget, J. (1954). *The Construction of Reality in the Child*. New York: Basic Books.
- Saunders, M. (1982). Productive activity in the curriculum: Changing the literate bias of secondary schools in Tanzania. *British Journal of Sociology of Education* **3**, 39–55.
- Shockley, M. S. (1946). American literature in American education. *College English* **8**, 23–30.
- Thron, E. M. (1987). The British Empire and the English curriculum. *ADE Bulletin* **87**, 4–7.
- Venezky, R. L. (1986). Steps toward a modern history of reading instruction. *Review of Research in Education* **13**, 129–167.

Mathematics

B Kaur, National Institute of Education, Singapore

C P Vistro-Yu, Ateneo de Manila University, Philippines

© 2010 Elsevier Ltd. All rights reserved.

Mathematics is an essential component of the school curriculum. It nurtures both the faculties of logic and reasoning among learners. In addition, its inclusion in the school curriculum as a compulsory subject until a certain level of schooling is justified by its utilitarian value. Hence, the content knowledge that comprises the school mathematics curriculum is shaped by the purposes it serves. The canonical school mathematics curriculum, a result of curriculum developers in one educational system simply copying what is being done in another, was developed in Western Europe in the aftermath of the Industrial Revolution. This curriculum was adopted practically in every country during the twentieth century (Howson and Wilson, 1986). The adoption was either voluntary or via colonization.

Over the span of the last two centuries although the content of the school curriculum has expanded to meet the changing needs of society, in the deeper structure with reform mostly confined to surface structure (Kilpatrick, 1996). The Trends in International Mathematics and Science Study (TIMSS) 2007 Assessment Frameworks (Mullis *et al.*, 2005) is an illustration of how core mathematical knowledge is globalizing. In the 2007 study 50 participant countries exhibited enough concurrence of curriculum that allowed the testing of the same basic mathematical (and science) content knowledge at certain grade levels.

Despite its structural stability, school mathematics curriculum has made two significant shifts in emphasis: one during the 1950s and 1960s known as the era of new mathematics and one from the 1980 onwards known as the era of problem solving. The era of new mathematics brought to many countries greater efforts to attract students to the study of mathematics by emphasizing its abstract structures with the hope and expectation that they would understand and appreciate mathematics more if they could see the simplicity and elegance of its laws (Kilpatrick, 1996). The impetus for the new mathematics era was the successful launching of Soviet Union's first artificial satellite into orbit around the Earth on 4 October 1957. The launch of Sputnik and the apparent Soviet lead in rocket technology disturbed many politicians and academics in the West. Consequently, political decisions translated into school curricula initiatives resulting in the new mathematics era. A major consequence of the political imperative was a desire to see the nurturing of budding scientists and mathematicians in the classroom.

The era of problem solving essentially emphasized the acquisition of mathematical knowledge to solve nonroutine

problems or doing realistic mathematics through applications, modeling, and mathematization (De Lange, 1996). This emphasis stemmed from the need to prepare students to be competent citizens for democratic life as opposed to qualifying for the future workforce (Keitel, 1993). A further development of the problem-solving era has been mathematical literacy as a goal of school mathematics curricula. Through the implementation of a differentiated curriculum for a diverse range of learners, intended so that all students could have access to the basic content knowledge deemed necessary for competent citizenship, attempts have been made to address equality and equity issues. Hence many of the issues related to the content knowledge of mathematics in curriculum development stem from the aims and objectives of teaching the subject rather than the body of content knowledge *per se*. For example, Wagner and Parker (1993) proposed the need to reconstitute the algebra curriculum by changing the emphasis from doing to thinking to meet the needs of knowledge-based economies. Similarly, two central issues that affect the geometry curriculum are; what geometry to teach at the respective year levels of schooling so that students are able to make sense of it, and, the debate about proofs in school mathematics. Advances in technology such as the advent of calculators (performing simple mathematical techniques to complex scientific and now graphic technologies) and the availability of technological tools such as computer graphics in the mathematics classroom have also had an impact on the emphasis of the content.

Curriculum development in mathematics may be addressed at three main levels: the intended, the implemented, and the attained. This article focuses on four aspects of the intended and implemented curriculum. They are curriculum selection – what guides the choice curriculum development and revision – how development may be done and what process of revision is followed, curriculum implementation – a consequence of how it was developed and what the different models of implementation are, and curriculum reform – some examples and their impact on mathematics education. Wherever necessary, relevant examples will be cited.

Curriculum Selection

Curriculum selection: What guides it? This phase usually precedes the initiation one. According to Howson *et al.*

(1981), curriculum development may be initiated for one or several of the following reasons:

- *Societal and political.* Economical and technological development, often result in reconsideration of education for larger or diverse groups and also education for specific desired outcomes which may be related to workforce requirements.
- *Mathematical.* Mathematics content knowledge is constantly evolving and, therefore, its implications are often addressed at the school level.
- *Educational.* The emergence of new theories of learning together with the advent of new technological aids and apparatus may result in reconsideration of some aspects of the curriculum.
- *The rewards of innovation.* A constant need to review and revise curriculum to keep abreast with world trends and also experiment may lead to attempts at innovation, thereby resulting in curriculum development.

The initiation phase would result in establishing guidelines for the need to develop curriculum, be it revision of a present curriculum or creation of a new curriculum. The selection of curriculum inevitably would be guided by a set of clear aims. Aims are expressions of intent, and intentions belong to groups or individuals. Therefore, who contributes to the aims would naturally be responsible for the direction of the intended curriculum.

Two dominant philosophical views influence what curriculum should look like: the traditional and the progressive views (Doll, 1996). The traditionalist view maintains that what has been done in the past has been done well and thus must be kept in the future. The progressivist view pleads to look at curricula more critically to see what can be done differently to enhance learning. These two views are, to a certain extent, manifested in mathematics curriculum development. Curricular reforms generally come slowly because of the resistance of traditionalists who do not wish to shake up old ways to welcome in new ideas or perhaps find the reforms difficult to implement. For example, in the Philippines, it took many years before mathematics curriculum specialists agreed to include probability in the elementary mathematics curriculum despite the numerous recommendations from studies to teach basic and intuitive concepts of probability.

How does one select curriculum? In several countries, such as Singapore, Malaysia, Hong Kong, and the Philippines, school mathematics curricula were adoptions of mathematics curricula of countries that ruled them at some point in their history (Wong *et al.*, 2001). In the case of Singapore, Malaysia, and Hong Kong, the curricula were adopted from Britain and in the Philippines it was adopted from America. Initially, these countries mimicked everything about the curricula; syllabi, textbooks, and assessment practices but gradually revised it in light

of their own needs and their changing positions in the increasingly globalized environments.

Much of the selection also has to do with the educational goals of a country. For many countries, these national goals are articulated in a document that government leaders and concerned educators prepare. The document helps mathematics teachers and educators focus on the educational goals. Culture also plays an important role in the selection of curriculum contents. This explains the different focal points in the mathematics curricula. For example, the Chinese-Confucian-influenced countries value public mathematics examinations and, as a result, examinations are a big part of their curricula.

Curriculum Development and Revision

There is no single way to develop and revise mathematics curricula. Countries follow their own curriculum development processes. In some, this includes a time frame. It may also either be process oriented where teachers take on the role of creators or product oriented where teachers take on the role of consumers. The role of the teacher in official curriculum development projects is shaped by beliefs as to who can and who should determine the curriculum.

Most countries set a period during which they study the current mathematics curriculum and its effectiveness according to the established national goals and standards for school mathematics. Asian countries, such as Japan, China, Singapore, and South Korea, are known to have established periods of curriculum reform. Japan, for example, follows a 10- to 12-year cycle during which their mathematics education curriculum leaders assess the goals of mathematics education and the implementation of the curriculum for that period (Nebres and Vistro-Yu, 1998). Until recently Singapore also followed a 10-years cycle, but in 2000 Singapore's Ministry of Education (MOE) decided that in order to remain relevant in a fast-changing world, the school curriculum should be reviewed every 7 years.

Following are two examples (or case studies), which compare the curriculum development and revision processes in Singapore and the United States. Exhibit A is an example of a curriculum development process which is engineered and carried out by a government body, the MOE, to produce a curriculum which is tightly aligned with the aims and goals of the nation's education system and also its socio-political goals. The curriculum developed is used in all schools and the implementation is also facilitated by the ministry. Exhibit B is an example of a curriculum development process which is undertaken by a professional body with inputs mainly from teachers, researchers, and mathematicians. In theory the curriculum development is only a guide, the adoption is left up to the different state systems to meet their specific needs.

Exhibit A: Curriculum Development and Revision in Singapore

Curriculum development in Singapore adopts the center → periphery model (Howson *et al.*, 1981) where a group of competent mathematics specialists, who are mathematics teachers seconded to the Ministry of Education's Curriculum Planning and Development Division (CPDD), engineer the entire process with inputs from all significant stakeholders. Being part of the school curriculum, the mathematics curriculum is naturally aligned with the aims and objectives of education in Singapore and also the desired goals of education for all students in the system. Mathematics is a compulsory subject at the primary (grades 1–6 corresponds to ages 7–12 years) and secondary (grades 7–10 corresponds to ages 13–16 years) levels. The curriculum is differentiated to meet the needs of students of differing abilities. The MOE issues the syllabus for the primary level and also for the lower secondary level (grades 7 and 8). Students take the Singapore-Cambridge Examinations at the end of their secondary schooling and therefore the syllabus for upper secondary levels (grades 9 and 10) are issued by the examining body in consultation with the MOE. The curriculum development and review process is the same for all the different courses of school mathematics.

The curriculum development process is set in motion by the setting up of a Mathematics Syllabus Review Committee comprising of a chair from the CPDD and members from the CPDD, Singapore Examinations and Assessment Board (SEAB), polytechnics, institutes of technical education, the National Institute of Education, and universities. The terms of reference guide the review and revision. Wong (1991) stated that in 1988 the terms of reference were, “To study the adequacy of the syllabuses

in meeting the needs of the pupils and to revise the syllabuses to reflect appropriate recent trends in mathematics education” (p 155).

Feedback gathered about the curriculum – be it anecdotal, or through research carried out by officers of the CPDD, or the Centre for Research in Pedagogy and Practice (CRPP) at the National Institute of Education – provides areas that require attention and revision. This feedback provides the committee with leads for verification and further exploration. At the start of the process, the committee would decide on the ways and means to gather information in a rigorous manner. Data collected provide the basis for decision making in due course. Some activities the committee may carry out to gather data include designing questionnaires and gathering feedback from heads of departments and teachers, organizing focus group discussions with teachers and pupils, and analyzing pupils' performance in national examinations and international studies such as TIMSS. The committee also studies reviews of mathematics curricula of other countries such as those in Europe, East Asia, the United Kingdom, as well as the United States and Australia. Feedback is also obtained from tertiary institutes such as polytechnics and the universities about the adequacy of the syllabi in preparation for future studies.

During the meetings, the data collected are reviewed and implications discussed. The aims and objectives of the curriculum under review are reexamined and aligned with educational policies and initiatives resulting in changes made to the mathematics content of the syllabus as well as the suggested pedagogy. Since 1990 the framework of Singapore's mathematics curriculum has mathematical problem solving as its primary goal, with five explicit contributory factors as shown in **Figure 1**. The framework

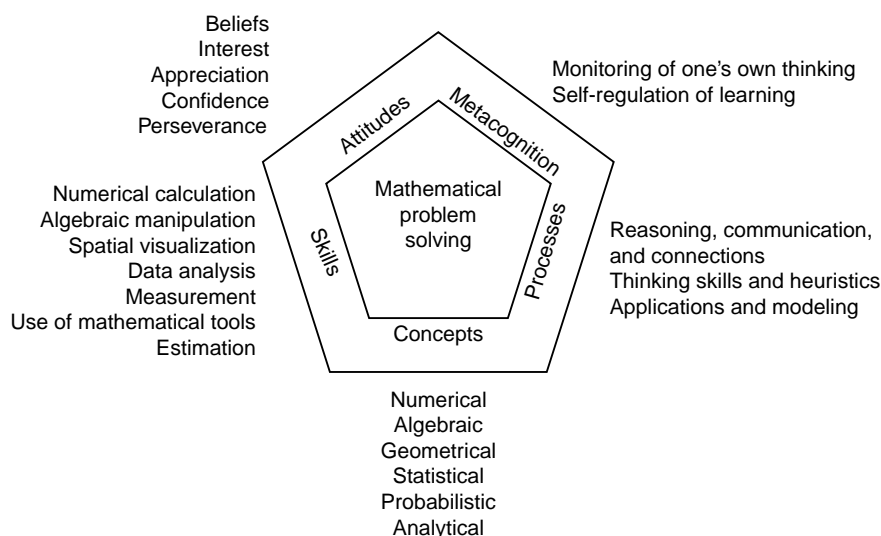


Figure 1 Framework of school mathematics curriculum (MOE, 2006).



Figure 2 Approval seal.

is also updated during the revision. The process of reviewing and revising the curriculum usually takes about a year. Once the revision is completed, the intended curriculum, which mainly is in the form of syllabus documents, is complete.

In Singapore, textbooks are produced commercially. When the revision of the syllabi is complete, the CPDD calls for expressions of interest to produce textbooks from the commercial sector of the economy. After an initial submission of sample chapters, the successful parties are guided very closely by the CPDD officers in the production of the textbooks. Only textbooks with the approval seal as shown in **Figure 2** may be used by schools.

Exhibit B: Curriculum Development and Revision in the United States

In the United States no single government agency controls grades K-12 public education at the national level. Authority for most educational decisions lies with education agencies in the 50 individual states, who in turn share decision making with the 14 883 individual districts contained within the United States. Hence, each district operating under its own authority and various state laws sets standards, designs delivery programs, and provides financial support for its own mathematics-education program. The guidance that states provide to their schools also varies. All states, with the exception of Iowa, provide schools with curricular frameworks for mathematics in required programs of study.

Two bodies, the National Council of Teachers of Mathematics (NCTM) and the United States National Commission on Mathematics Instruction (USNCMI), have significant influence on the school mathematics curriculum across the United States. NCTM is a professional body of teachers of mathematics founded in 1920 dedicated to improving mathematics teaching and learning from kindergarten to high school. It facilitates ongoing dialog and constructive discussion with all stakeholders about what is best for the nation's students. The USNCMI facilitates

the participation of the United States at the international level in mathematics activities and through the National Science Foundation (NSF), NCTM, and the Conference Board of the Mathematical Sciences advance mathematics education at the national level.

NCTM has been active in the curriculum development of school mathematics since its inception but in 1985 it began the process that led to three significant documents on curriculum and evaluation (NCTM, 1989), teaching (NCTM, 1991), and assessment (NCTM, 1995). Dossey and Usiskin (2000) claimed that although throughout the twentieth century various groups had made suggestions relative to what the US school mathematics curriculum should look like, none combined the scope and detail of these three documents. In 1989 NCTM produced the Curriculum and Evaluation Standards which provided a listing by grade level bands (K-4, 5-8, 9-12), of the mathematics that students should know about. These included problem solving, reasoning, communication, connections, and various content aspects of mathematics relevant to those grade levels. Similarly, no previous document had described teaching in as much detail as the 1991 document, the Professional Standards for Teaching Mathematics. The teaching standards asserted that concepts, procedures, and relationships were best developed in contexts that warranted students to develop knowledge themselves under the guidance of the teacher and to engage in discourse about the knowledge. Correspondingly, in 1995, the Assessment Standards for School Mathematics reflected a deeper understanding called for in the curriculum and its teaching through multidimensional assessment methods.

The visibility of the above standards led to a greater interest among groups within the mathematics and mathematics education in the revision that began in 1995 and culminated in the release of the Principles and Standards (NCTM, 2000) in April 2000. During the revision phase, NCTM solicited feedback from individuals and organizations by first developing a draft of the Principles and Standards and then holding a nationwide year of discussion on the draft. This process enabled them to consider all feedback and revise the Principles and Standards accordingly.

The Principles and Standards (NCTM, 2000) are guided by six principles. They are:

1. equity – excellence in mathematics education requires equity, that is, high expectation and strong support for all students;
2. curriculum – it must be coherent, focused on important mathematics, and well articulated across the grades;
3. teaching – effective mathematics teaching requires understanding what students know and need to learn and then challenging and supporting them to learn it well;
4. learning – students must learn mathematics with understanding, actively building new knowledge from experience and prior knowledge;

5. assessment – assessment should support the learning of important mathematics and furnish useful information to both teachers and students; and
6. technology – technology is essential in teaching and learning mathematics, it influences the mathematics that is taught and enhances students' learning.

Despite all that goes into developing a sound and rigorous school mathematics curriculum by NCTM, it remains as a resource and guide for the nation. At the state level, although all states set graduation guidelines, they differ in the extent to which they set curriculum guidelines for mathematics and the steps they take to ensure that the curriculum is implemented. Within states, local school districts make the decisions regarding which materials to use. Some states have textbook adoption practices, while others do not where the teacher is free to use any materials he or she deems suitable.

Curriculum Implementation

The teacher is the mediator between the curriculum and the child, and any attempt to change the curriculum must consider the teacher's role. Teachers can be implicated in curriculum development in two related ways; as participants in the process or as users of the product. In Japan, curriculum implementation uses both the top-down and bottom-up approaches (IFIC and JICA, 2004). The top-down implementation consists of three stages. The first stage is when the MOE publishes commentaries on the Courses of Study (COS). The second stage is when lectures are given by subject examiners to expound and explain the published commentaries to the users of the COS. The third stage is when guidelines for the teaching of the COS are disseminated to teachers through guidance documents. The bottom-up approach of curriculum implementation is essentially of three types: administrative leadership type, autonomy type, and the textbook publishing company intervention type. An example of an administrative-leadership-type approach is when research is conducted among designated schools and the results are made public to encourage and inform other schools of the effectiveness, perhaps, of the new curriculum. An example of the autonomy type is when collaborating persons such as teachers who understand fully the revised curriculum become key persons and leaders who would conduct research on teaching methodology and other topics that move the curriculum implementation forward. An example of the textbook publishing company intervention is when textbook companies ask teachers who have deep understandings of the revised curriculum to write textbooks or teaching guides that provide a fresh perspective of the methodology and contents of the revised curriculum.

The Fife Mathematics Project of Scotland in the 1960s as described by Howson *et al.* (1981). Approaches for curriculum implementation may be top-down, bottom-up or both. The roles of teachers in each of these may be distinct. The example from Japan illustrates the roles of the teachers in both a top-down and bottom-up approach. In contrast the Fife mathematics project highlights the role of teachers in a bottom-up approach.

Curriculum Reform

Curricular reforms may result from innovative projects or well-researched documents of school mathematics. Some of these curricular reforms have made tremendous impact. The Realistic Mathematics Education (RME) project is an example of a sustained reform in school mathematics curriculum. Developed by researchers at the Freudenthal Institute the theoretical framework of the project is guided by the didactical belief propagated by Hans Freudenthal. Three tenets of mathematics characterize the framework:

1. the starting points of instructional sequences should be applications experientially real to students;
2. the starting points should also be justifiable in terms of the potential endpoints; and 3. instructional sequences should involve activities in which students create and elaborate symbolic models of their informal activity (De Lange, 1996: 60–61). The success of this curricular reform can be seen not only in the Netherlands but also through the adoption of RME in many other countries around the world.

The open-ended approach of teaching mathematics (Becker and Shimada, 1997) that started in Japan in 1971 has also sustained its momentum in Japan as well as made its way into classrooms around the world. This method engages students in mathematical thinking through instructional activities that promote varied problem-solving approaches. The Agenda for Action, a position paper by NCTM released in 1980, is an example of a well-researched document that brought to the forefront the need to prepare students in mathematics classrooms to be problem solvers. This document initialized change around the world with one country after another emphasizing mathematical problem solving in their curriculum.

Curricular reforms may arise from global trends in mathematics education. The 1989 publication of the Curriculum and Evaluation Standards in mathematics by NCTM certainly led to many more countries articulating their intended curriculum in the form of standards. The years following the first publication of the Curriculum and Evaluation Standards documents saw how countries such as Australia, New Zealand, Canada, and Singapore followed suit. The Philippines recently put out the working

draft of its own Curriculum Framework and Teacher Education Framework in mathematics. A more recent mathematics reform has been created by the consistently high performance of Singapore's students in the TIMSS – 1995, 1999, and 2003. For example, Israel and parts of the United States have adopted Singapore's school mathematics curricula, in particular, textbooks used in schools, thereby engaging in reform of mathematics curriculum.

See also: An Overview of Research in Curriculum Inquiry; Curriculum and Constructivism; Curriculum and Critical Theory; Curriculum and Syllabus Design; Curriculum and Teacher Change; Curriculum and the Publishing Industry; Curriculum Evaluation: Approaches and Methodologies; Curriculum Governance and Planning; Curriculum Planning and Systems Change; Curriculum Reform; Curriculum Use in the Classroom; Curriculum, Digital Resources and Delivery; Textbook Development and Selection.

Bibliography

- Becker, J. P. and Shimada, S. (eds.) (1997). *The Open-Ended Approach: A New Proposal for Teaching Mathematics*. Reston, VA: NCTM.
- De Lange, J. (1996). Using and applying mathematics in education. In Bishop, A. J., Clements, K., Keitel, C., Kilpatrick, J., and Laborde, C. (eds.) *International Handbook of Mathematics Education*, Part 1, pp 49–97. Dordrecht: Kluwer.
- Doll, R. (1996). *Curriculum Improvement: Decision Making and Process*. Boston: Allyn and Bacon.
- Dossey, J. and Usiskin, Z. (2000). *Mathematics Education in the United States 2000*. Reston, VA: NCTM.
- Howson, G., Keitel, C., and Kilpatrick, J. (1981). *Curriculum Development in Mathematics*. Cambridge: Cambridge University Press.
- Howson, G. and Wilson, B. (1986). *School Mathematics in the 1990s*. Cambridge, UK: Cambridge University Press.
- IFIC and JICA (Institute for International Cooperation and Japan International Cooperation Agency). (2004). *The History of Japan's Educational Development*. Tokyo: Author.
- Keitel, C. (1993). Implicit mathematical models in social practice and explicit mathematics teaching by applications. In de Lange, J., Keitel, C., Huntley, I., and Niss, M. (eds.) *Innovation in Maths Education by Modeling and Applications*, pp 19–31. Chilter, UK: Ellis Horwood.
- Kilpatrick, J. (1996). Introduction to section 1. In Bishop, A. J., Clements, K., Keitel, C., Kilpatrick, J., and Laborde, C. (eds.) *International Handbook of Mathematics Education*, Part 1, pp 7–9. Dordrecht: Kluwer.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., et al. (2005). *TIMSS 2007 Assessment Frameworks*. Boston: International Study Center, Lynch School of Education, Boston College.
- NCTM (National Council of Teachers of Mathematics). (1980). *Agenda for Action*. Reston, VA: Author.
- NCTM (National Council of Teachers of Mathematics). (1989). *Curriculum and Evaluation standards for school mathematics*. Reston, VA: Author.
- NCTM (National Council of Teachers of Mathematics). (1991). *Professional Standards for Teaching Mathematics*. Reston, VA: Author.
- NCTM (National Council of Teachers of Mathematics). (1995). *Assessment Standards for School Mathematics*. Reston, VA: Author.
- NCTM (National Council of Teachers of Mathematics). (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.
- Nebres, B. F. and Vistro-Yu, C. P. (1998). A look at organizational structure for an effective delivery of science education. In Ogena, E. B. and Brawner, F. G. (eds.) *Science Education in the Philippines: Challenges for Development*, pp 148–166. Bicutan: NAST, SEI, UP-CIDS.
- Wagner, S. and Parker, S. (1993). Advancing algebra. In Wilson, P. and Wagner, S. (eds.) *Research ideas for the Classroom High School Mathematics*, pp 119–139. New York: Macmillan.
- Wong, K. Y. (1991). Curriculum development in Singapore. In Marsh, C. and Morris, P. (eds.) *Curriculum Development in East Asia*, pp 129–160. London: Falmer Press.
- Wong, K. Y., Taha, Z. B. H. M., and Veloo, P. (2001). Situated sociocultural mathematics education: Vignettes from Southeast Asian practices. In Atweh, B., Forgasz, H., and Nebres, B. (eds.) *Sociocultural Research on Mathematics Education: An International Perspective*, pp 113–134. Mahwah, NJ: Erlbaum.
- Battista, M. T. and Clements, D. H. (2000). *Mathematics Curriculum Development as a Scientific Endeavor*. Mahwah, NJ: Erlbaum.
- Black, P. and Atkin, J. M. (1996). *Changing the Subject: Innovations in Science, Mathematics and Technology Education*. London: Routledge.
- Blume, G. W. and Nicely, R. F., Jr. (eds.) (1993). *A Guide for Reviewing School Mathematics Programs*. Reston, VA: Association for Supervision and Curriculum Development and the National Council of Teachers of Mathematics.
- De Guzman, S. B. (2003). The dynamics of educational reforms in the Philippine: Basic and higher education sectors. *Asia Pacific Education Review* 4, 39–50.
- De Lange, J. (1987). *Mathematics Insight and Meaning*. Utrecht: OW and OC.
- Doll, R. C. (1989). *Curriculum Improvement*. Boston: Allyn and Bacon.
- Fujita, T. and Jones, K. (2003). Interpretations of national curricula: The case of geometry in popular textbooks in Japan and the UK. *Paper Presented at the British Educational Research Association Annual Conference 2003*. Heriot-Watt University, Edinburgh, 11–13 September.
- Goldsmith, L. T. and Mark, J. (1999). What is a standards-based mathematics curriculum? *Educational Leadership* 57, 40–44.
- Hoyle, C., Morgan, C., and Woodhouse, G. (1999). *Rethinking the Mathematics Curriculum*. London: Routledge.
- Manouchehri, A. and Goodman, T. (1998). Mathematics curriculum reform and teachers: Understanding the connections. *Journal of Educational Research* 92, 27–41.
- Ministry of Education (2006). *A Guide to Teaching and Learning of O-Level Mathematics 2007*. Singapore: Ministry of Education, Curriculum Planning and Development Division.
- UNESCO (2000). Science, technology and mathematics education: A global perspective. *Connect* 25, 3–4.

Further Reading

Physical Education and Sports

D Kirk, University of Bedfordshire, Bedford, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Models-based instruction – A means of aligning subject matter and teaching styles with learning outcomes. Examples of instructional models in physical education are sports education and teaching games for understanding.

The development of the curriculum for physical education has tended to be most strongly influenced by local forces within state and national educational systems. This feature has led to much unevenness in the forms of physical education that exist in countries around the world. Nevertheless, there have been some commonalities of curriculum development that can be grouped around three main trends: a trend toward national and state curricula, where physical education has been caught up in developments affecting the school curriculum as a whole; a trend toward examination and matriculation in physical education in the middle and later years of secondary schooling; and a trend toward the construction, implementation, and increasingly widespread dissemination of a models-based approach to instruction, incorporating a number of specific pedagogical or instructional models.

The initial impetus for each of these trends dates to the 1970s and a time when internationally physical education had begun to take shape as a multi-activity, sports-based curriculum. This multi-activity form of curriculum was the outcome of a dramatic shift that took place in the years immediately following World War II, a shift that overturned some 50 years or more of domination of the curriculum of physical education by various forms of gymnastics, at least within systems of state-provided education. In some countries, such as England, sports were played in schools serving the social elite from much earlier than the 1950s, although there is some dispute over whether these activities might properly be referred to as physical education.

This article considers the selection, development, and implementation of the curriculum of physical education. It focuses particularly on how each of the three trends – national and state curricula, examination in physical education, and models-based instruction (MBI) – provide contexts for the development of particular versions of physical education. In each case, the forms of physical

education curricula that result inevitably signal the perceived worth of some knowledge over other knowledge in, through, and about physical education.

National and State Curricula

The emergence since the mid-1970s of state curricula in a number of countries, including Australia, England and Wales, New Zealand, Scotland, Canada, and Ireland, has had important consequences for curriculum development in physical education. Although the original rationale for these developments is not always made clear, it would appear that most rest on a version of the core curriculum concept advocated by writers such as Dennis Lawton and Malcolm Skilbeck, which itself drew heavily on the educational philosophizing of Richard Peters and Paul Hirst. The core curriculum was intended to encapsulate the essential learning that all young people should experience during the compulsory years of schooling.

In the United States, Rink and Mitchell (2003) argue that physical education has most often been excluded from state educational reforms, damaging, in the process its credibility as an educational activity and further confirming its marginalization in the school curriculum. Elsewhere, physical education has been included in state curriculum reforms, with varying outcomes.

For example, Scottish educational reforms for the curriculum of years 3 and 4 of secondary schools, published in the report of the Munn Committee (Scottish Education Department, Consultative Committee on Curriculum, 1977), drew explicitly on the core curriculum concept, and provide a vivid example of the significance of such developments for physical education. The first of two rounds of national examinations (the ordinary grade) were typically taken in year 4 at age 16. Curriculum reform in years 3 and 4 which supported these high-stakes examinations was therefore controversial. In this context, as a nonexaminable, practical, and as the report of the Munn Committee described it, “non-cognitive” subject (Scottish Education Department, Consultative Committee on Curriculum, 1977), physical education’s place in the core curriculum was not guaranteed. Scottish physical educators of the time were much exercised over the prospect that students could cease to participate in their subject after year 2, at age 14 (Scottish Education Department, Consultative Committee on Curriculum, 1977).

The arguments presented to the Munn Committee during its review of the curriculum provided competing and alternative views of physical education, one humanities based and sponsored primarily by women physical educators, and another science based and supported primarily by men. The Committee was eventually persuaded by the men's arguments and a science-based physical education gained a place in the compulsory curriculum of years 3 and 4, an outcome that irrevocably damaged the humanities-based form.

In a review chapter, Penney (2006) argued that studies of state curriculum reforms, such as the National Curriculum developments in England and Wales and the development of a national framework of curriculum Statements and Profiles in Australia, have provided evidence of a number of common trends in the development of physical education. She notes that in the process of producing curriculum texts such as syllabi, a range of competing interests and multiple discourses is exposed. The reform of physical education in England was impacted by interest in sporting excellence, reducing childhood obesity, developing socially valued behaviors that could transfer to other settings, such as work, and combating antisocial behavior. In a similar vein, there were claims in New Zealand that reforms of physical education aspired to inoculate young people against the risk of drug use, alcohol abuse, low self-esteem and stress, sexual activity, and spiritual decline.

At the same time, Penney (2006) argues that these competing interests are shaped through the curriculum construction process to support broader government agendas, thereby revealing the power relations at play in the development of curriculum. A study of the writing of the National Statement and Profile for Health and Physical Education in Australia illustrated how the selection of the writing team, their remit, the processes of consultation, and timescales all worked to produce a particular curriculum. Penney and Evans (1999) suggest that the careful selection of pro-sports members of the first National Curriculum committee in England and Wales produced a form of physical education that sought to restore sport's place at the center of the subject.

As Penney (2006) notes, while these common trends are discernable across countries and educational systems, there are also important cultural specificities that militate against casual generalizations, for example, the enduring prominence of traditional team games in the England and Wales National Curriculum, and also, almost paradoxically, the assertion and inclusion of a distinctive Welsh culture as the curriculum is applied in Wales. In Australia, during the early to mid-1990s, subject areas were radically recast as key learning areas (KLAs), with physical education finding a home in the Health and Physical Education KLA alongside nutrition and home economics, health, ethics, and religious studies. As the National Framework of Statements and Profiles for KLAs were reworked within state

education systems, considerable differences appeared in terms of the selection of subject matter and its various combinations. Meanwhile, in New Zealand during the late 1990s, distinctively different Maori and Pakeha conceptions of health had to be reconciled in the construction of the Health and Physical Education curriculum.

Examination and Matriculation in Physical Education

Formal examinations first appeared in secondary school physical education in the 1970s in England and Australia, initially at the end of the compulsory stage of schooling at age 15 or 16, and later at the post-compulsory stage (up to age 19). The post-compulsory examinations in both England and Australia and later in Scotland and New Zealand could be counted toward university entrance or matriculation. The arrival of examinations in physical education can be attributed to a confluence of factors, key among them were state-level initiatives to modernize the secondary school curriculum, on the one hand, and on the other the growing academicization through the 1960s and 1970s of physical education in higher education which fueled the ambitions of physical educators to reduce their perceived marginality in schools and to protect their subject from exclusion (Carroll, 1994).

Formal examinations were not, necessarily, welcomed by all members of the physical education communities in these countries, and there was through the 1970s and 1980s, at times, a vociferous debate centered on their necessity and their likely impact on physical education programs at other stages of schooling. A particular concern was that participation in practical physical activities such as games, sports, swimming, and gymnastics – subject matter that was the mainstay of secondary school physical education programs from the 1950s to the 1980s – would be replaced by classroom-based, theoretical study of sport and human movement.

This debate did little to halt the spread of formal examinations which, by the early 1990s, were well established in England and Wales, in the Australian states of Queensland, New South Wales, and Victoria, and in New Zealand. By the end of the 1990s, Green (2005) reported that examination and matriculation in physical education in England and Wales was the second fastest growing subject in the secondary school curriculum at a rate for an A level of 57.4% between 1997–98 and 2002–03, compared with 6.9% for biological sciences over the same period. Similar growth was observed in Queensland and other Australian states, while examinable physical education appeared for the first time in the curriculum of Scottish secondary schools in the 1980s, followed by a matriculation-level syllabus in the early 1990s. More recently in Ireland, a syllabus for matriculation in physical

education has been developed and may soon be implemented as part of the Leaving Certificate.

This substantial growth can be attributed in large part to the continuing influence of the two factors that lead to the introduction of examination for physical education in the 1970s. The liberalization and broadening of the secondary school curriculum in these countries has occurred as governments have sought to entice young people to stay at school beyond the compulsory leaving age by offering interesting or attractive subjects for study. The continuing and powerful growth of the field of physical education/sport and exercise science/kinesiology has also been a significant driver of developments of the secondary school curriculum. It is around this latter source of influence in particular that debates within the physical education community have continued, despite the growth and popularity of the subject.

A major point of debate among physical educators has centered on the shift, predicted by early opponents of examination and matriculation in physical education in the 1970s away from engagement in practical physical activities and toward classroom-based theoretical study. During the 1990s, at least one examination board in England and Wales offered an advanced level (A level) version of physical education that required no participation in physical activity at all. At the same time, this trend away from practical knowledge has not been universal. In England and Wales, revisions to A-level specifications required all four examination boards to include engagement with practical physical activities as a necessary component of syllabi. In Queensland, during the mid- to late 1990s, the integration of theoretical and practical knowledge was a guiding principle in the construction of assessment tasks in the Senior Physical Education program.

A second and equally troublesome issue for some contributors to the debate has been an apparent scientization of the physical education curriculum. In the late 1980s, Fitzclarence and Tinning (1990) became involved as writers of the Victoria Certificate of Education physical education curriculum in introducing a socially critical dimension to the subject to counteract what they saw as its dominance by biophysical science knowledge. A study of the Scottish Higher Grade Physical Education also noted the absence of any social science material in the syllabus. In another study of Senior Physical Education in Queensland, researchers found that the curriculum required equal amounts of time to be spent on psychological, biomechanical, physiological, pedagogical, and socio-historical subject matter, while there was a wide range of sports, games, and other physical activities from which schools could choose to construct their programs.

Models-Based Instruction

Metzler (2000) identified an important trend in the selection, development, and implementation of curriculum of

physical education, starting in the 1970s, which he conceptualized as MBI. MBI involves the statement of outcomes for learning and then the selection of teaching styles and subject matter which will facilitate the realization of these outcomes. Central to this concept of MBI is the interdependence and alignment of learning with teaching styles and subject matter. In elaborating this concept, Metzler was building on over 20 years of development of specific forms of MBI, including sports education (Siedentop, 1994), Teaching Games for Understanding (TGfU; Bunker and Thorpe, 1982), personal and social responsibility (Hellison, 1995), and health-related exercise (Whitehead and Fox, 1983), among others.

Of these models, the development and implementation of Sports Education (Wallhead and O'sullivan, 2005) and TGfU (under various nomenclature) (Oslin and Mitchell, 2006) have been the most widely researched, although the development and implementation of the others listed above have begun to be informed by a substantial research literature. Each of the models uses some form of organized physical activity as its subject matter, although with different learning outcomes in mind facilitated by a variety of teaching styles, in various combinations.

Sports Education

Sports education originated in the work of Daryl Siedentop and colleagues affiliated with the Ohio State University (Siedentop, 1994). The impetus to develop sports education came from Siedentop's dissatisfaction with the ways in which multi-activity physical education had from the 1950s through the 1980s offered young people a generally superficial and fragmented experience of sport. The principal aim of sports education is to develop young people as competent, literate, and enthusiastic sportspeople (Siedentop, 1994). In order to achieve this outcome, sports education requires the physical education unit of the curriculum to be recast as a sports season, with the following selected features of sport:

- sports education operates in seasons of length considerably longer than most physical education units, between 8 and 16 lessons;
- players are on teams and stay on that same team as a persisting group for the entire season, in order to promote affiliation and membership;
- seasons are bounded by formal competition, which is interspersed with student-directed practice sessions;
- there is a culminating event to mark the conclusion of each season, which includes festivity;
- there is extensive record keeping and collection of information; and,
- in addition to experiencing the role of players, learners also take on additional roles such as team coach, referee, equipment officer, record keeper, timekeeper, team captain, warm-up coach, journalist, and sports panel member.

The earliest documented research on sports education in school physical education appeared in the early 1990s. In addition to the United States, sports education has been studied extensively in New Zealand and Australia, and has recently attracted some attention in England. Sports education has been examined at primary, secondary, and undergraduate levels and with specific attention to low-skilled pupils, high-skilled learners, and at-risk pupils.

While games have tended to provide the activity context, there are examples of sports education units using gymnastics, dance, volleyball, softball, and athletics. Kinchin's (2006) review suggests that sports education has promoted breadth and balance in the physical education curriculum, with more recent studies reporting the use of bicycle safety and outdoor adventure activities. He also notes evidence from research studies to suggest that sports education is popular with students since they develop an affiliation with their team over an extended period of time, that the additional roles promote ownership, responsibility, and accountability for learning, and that competence develops across the ability range. Some teachers are reported to have found some difficulty in adapting their teaching styles to suit sports education, while others have over time embraced the facilitation style required by this model.

Kinchin also notes the development of variations of sports education, including empowering sport which sought to incorporate elements of personal and social responsibility, cultural studies which aims to question assumptions and values in sport, Sport for Peace, developed for urban schools in the USA and aimed at conflict resolution and fostering personal and social responsibility, and sports education as situated learning, which has been concerned with the authenticity of the experience sports education provides, and possibilities for transfer of learning.

Teaching Games for Understanding

Most of the game-centered approaches to teaching physical education have their origins in the work of Bunker and Thorpe (1982) on TGfU at Loughborough University. Bunker and Thorpe (1982) argued that too much games teaching, at the time they were writing, aspired to develop techniques ahead of learning to play the game. They developed a model that placed the learner and the modified game at the center of the curriculum, with the modified game forming the context in which game appreciation, tactical awareness, decision making, and skill in execution could be learned. In addition to providing a basis for a new form of MBI, Bunker and Thorpe's original work inspired the construction of various systems of classifying games in the curriculum, and variations on their model such as the Tactical Games Model, Play Practice, Game Sense, and developments of the Bunker–Thorpe

model itself (Oslin and Mitchell, 2006). Research interest in new approaches to games teaching and coaching has also generated new models in France and Francophone Canada and in Germany.

The earliest studies of TGfU and its derivatives sought to compare the game-centered model with technique-led approaches to games teaching. Rink *et al.* (1996) noted that this research reported positive learning outcomes for students using TGfU. The most powerful finding across the studies reviewed by Rink *et al.* (1996) was that students taught from a TGfU perspective tended to perform better on tests of tactical knowledge than those taught from a technique-led perspective. Some studies suggested that a TGfU approach may be perceived by students to be more enjoyable than the technique-led approach, and so students may be more highly motivated to participate. Reviewing 13 studies carried out between 1995 and 2003, Oslin and Mitchell (2006) reached similar conclusions, noting that until 2001 all studies sought to measure skill development, game performance, and a variety of other variables such as decision making, motivation, and knowledge, and were informed by an information-processing theoretical perspective.

Rink *et al.* (1996) also noted that, despite some positive findings, the earlier studies reviewed could not provide conclusive support for TGfU over technique-led approaches, a complaint echoed by Oslin and Mitchell (2006). Both sets of authors argued that this was due to different research designs, which made comparisons difficult because studies varied according to the game chosen, the age of participants, the length and nature of the intervention, the variable chosen for investigation, and the ways in which these variables were measured.

Since the late 1990s, research on TGfU and its derivatives has eschewed comparative study designs and has instead adopted naturalistic designs to investigate how teaching experiments and other interventions impact on teaching and learning (Oslin and Mitchell, 2006). These more recent studies have adopted multiple theoretical perspectives, including cognitive, ecological, and situated learning approaches, and have tended to be concerned with cognition, knowledge, and game performance (Kirk and MacPhail 2002; Rovegno *et al.*, 2001).

Oslin and Mitchell (2006) report on two other lines of research relating to TGfU and its derivatives, one focusing on how teachers (predominantly) have responded to this model, and the attempts to measure game performance. In the former, they identify a trend in the reports of studies that found initial resistance to TGfU and other game-centered approaches from practicing teachers and student teachers, to growing acceptance once the model had become familiar. In the latter, they review the development of two instruments, the Game Performance Assessment Instrument (GPAI), developed by Oslin *et al.* (1998), and the Team Sport Performance Assessment

Instrument (TSPA), developed by Grehaigine *et al.* (1997), and conclude that the instruments have the potential to provide a better appreciation of the complexity of game performance, with application perhaps more appropriate to elite levels of performance.

Conclusion

This analysis of the selection, development, and implementation of the curriculum of physical education suggests the trends toward national and state curricula, examination in physical education, and MBI have resulted in some common features across countries and contexts, and some variation in what counts as physical education.

While the trend toward examination and matriculation in physical education would seem to be strongly supported by student interest, making the study of secondary school physical education in some countries one of the most popular school subjects, this form of physical education curriculum nonetheless presents some challenges to the notion that physical education is a form of practical knowledge requiring participation in physical activities. Developments of national and state curricula show least uniformity across countries and seem to be most affected by local forces within education systems and initiatives to reform the school curriculum. The trend toward MBI is by far the most fluid factor influencing curriculum development in physical education, existing as it does outside the jurisdiction of official curriculum authorities, and supported by a strong research base.

There is little evidence of rational curriculum planning in any of these trends, even those directed by government agencies. At the same time, there is clearly no shortage of ideas for developing physical education. We might conclude from this observation that the selection, development, and implementation of the curriculum is a process of socially constructing physical education, in which ideas of physical education and their practice reflect particular and somewhat different configurations of values.

Bibliography

- Bunker, D. and Thorpe, R. (1982). A model for the teaching of games in the secondary school. *Bulletin of Physical Education* **10**, 9–16.
- Carroll, R. (1994). *Assessment in Physical Education: A Teachers' Guide to the Issues*. London: Falmer.
- Fitzclarence, L. and Tinning, R. (1990). Challenging hegemonic physical education: Contextualising physical education as an examinable subject. In Kirk, D. and Tinning, R. (eds.) *Physical Education, Curriculum and Culture: Critical Issues in the Contemporary Crisis*, pp 169–193. London: Falmer Press.
- Green, K. (2005). Examinations: A 'new orthodoxy' in physical education. In Green, K. and Hardman, K. (eds.) *Physical Education: Essential Issues*, pp 143–160. London: Sage.
- Grehaigine, J. F., Godbout, P., and Bouthier, D. (1997). Performance assessment in team sports. *Journal of Teaching in Physical Education* **16**, 500–516.
- Hellison, D. (1995). *Teaching Responsibility through Physical Activity*. Champaign, IL: Human Kinetics.
- Kinchin, G. D. (2006). Sport education: A view of the research. In Kirk, D., Macdonald, D., and O'Sullivan, M. (eds.) *The Handbook of Physical Education*, pp 596–609. London: Sage.
- Kirk, D. and MacPhail, A. (2002). Teaching games for understanding and situated learning: Re-thinking the Bunker-Thorp Model. *Journal of Teaching in Physical Education* **21**, 177–192.
- Metzler, M. W. (2000). *Instructional Models for Physical Education*. London: Allyn and Bacon.
- Oslin, J. and Mitchell, S. (2006). Game-centred approaches to teaching physical education. In Kirk, D., Macdonald, D., and O'Sullivan, M. (eds.) *The Handbook of Physical Education*, pp 627–651. London: Sage.
- Oslin, J. L., Mitchell, S. A., and Griffin, L. L. (1998). The Game Performance Assessment Instrument (GPAI): Development and preliminary validation. *Journal of Teaching in Physical Education* **17**, 231–243.
- Penney, D. (2006). Curriculum construction and change. In Kirk, D., Macdonald, D., and O'Sullivan, M. (eds.) *The Handbook of Physical Education*, pp 565–579. London: Sage.
- Penney, D. and Evans, J. (1999). *Politics, Policy and Practice in Physical Education*. London: Spon.
- Rink, J. and Mitchell, M. (2003). Introduction: State level assessment in physical education. *Journal of Teaching in Physical Education* **22**, 471–472.
- Rink, J. E. R., French, K. E., and Tjeerdsmas, B. L. (1996). Foundations for the learning and instruction of sport and games. *Journal of Teaching in Physical Education* **15**, 399–417.
- Rovegno, I., Nevett, M., and Babiarz, M. (2001). A field-based project on learning and teaching of invasion game tactics in fourth grade: Introduction and theoretical perspective. *Journal of Teaching in Physical Education* **20**, 341–351.
- Scottish Education Department, Consultative Committee on Curriculum (1977). *The Structure of the Curriculum in Years Three and Four of the Scottish Secondary School*. Edinburgh: HMSO.
- Siedentop, D. (1994). Introduction to sports education. In Siedentop, D. (ed.) *Sport Education: Quality PE through Positive Sport Experiences*, pp 3–16. Human Kinetics: Champaign, IL.
- Wallhead, T. and O'Sullivan, M. (2005). Sport education: Physical education for the new millennium? *Physical Education and Sport Pedagogy* **10**, 181–210.
- Whitehead, J. and Fox, K. (1983). Student centred physical education. *Bulletin of Physical Education* **19**, 21–30.

Further Reading

- Green, K. and Hardman, K. (2005). *Physical Education: Essential Issues*. London: Sage.
- Penney, D. (2006). Curriculum construction and change. In Kirk, D., Macdonald, D., and O'Sullivan, M. (eds.) *The Handbook of Physical Education*, pp 565–579. London: Sage.
- Metzler, M. W. (2006). *Instructional Models for Physical Education*, 2nd edn. Holcomb Hathaway: Scottsdale, AZ.

Relevant Websites

- <http://hh-pub.com/book.php3> – Holcomb Hathaway Publishers: Investigative Reporting from Premise to Publication.
- <http://www.qca.org.uk> – Qualifications and Curriculum Development Agency.
- <http://www.qsa.qld.edu.au> – Queensland Studies Authority: Queensland Government.

Primary and Elementary/Middle Grades Reading

D E Alvermann, University of Georgia, Athens, GA, USA

C A Mallozzi, University of Kentucky, Athens, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Information communication technologies (ICTs) –

ICTs provide access to the Internet and its associated digital literacies, navigating the Web, multimedia, and so on.

Inoculation approach – The misinformed notion that provide reading instruction in the primary grades is sufficient for older readers' comprehension of subject-matter texts.

Subject-matter – The curricula and content areas, such as science, social studies, mathematics, and English language arts.

The reading curriculum for the primary-school-emergent readers and older elementary/middle-grade readers divides roughly at grade 3 (approximately age 9) in many countries (e.g., Canada, Chile, China, Malaysia, Nigeria, and the United States), though not all (e.g., Poland and Switzerland, where primary school ends much later). The rationale in selecting grade 3 as the dividing point between primary and elementary/middle grades reading rests largely on Barr's (2001) review of the research on teaching reading, which distinctly divided developmental beginning-reader issues from other issues in comprehending more complex texts in the upper grades. This divide, however, does not signal support for the cliché that first children learn to read and then they read to learn. Separating the act of reading from one of its functions – reading to learn something – makes little sense. Developmentally, emergent readers are different from skilled readers, but the difference lies more with the subject-matter materials than with the groups' purposes for reading.

Primary Reading

Curricular divisions in primary reading, especially where alphabetic languages dominate, tend to center on decoding and comprehension. This rather arbitrary divide between deciphering print and meaning-making masks two other aspects of reading. The first is concepts of print, a set of fundamental but complicated understandings about how print works in the world. An emergent reader demonstrates concepts of print by holding a book correctly, directionally

tracking print, knowing how to turn the page of a text, and other such behaviors that may seem obvious to more experienced readers but are basic to more advanced understandings of print. The second is fluency. A fluent reader comprehends texts of various types with speed, accuracy, and appropriate expression, similar to how a coherent speaker uses phrasing, tonal, rhythmic, and pacing conventions to communicate orally. When fluency is emphasized over meaning-making activities in the curriculum, students often become word callers rather than readers for comprehension.

Decoding typically points to two curricular areas of beginning reading: phonemic awareness and phonics. Phonemic awareness requires being mindful of the smallest units of sound that allow speakers of a language to differentiate among words (Snow *et al.*, 2005: 68). The primary reading curriculum includes activities involving alliteration, rhyming, and listening for sounds that make up a word (e.g., /b/ and /a/ and /t/). For the most part, these activities do not include matching sounds to letters or phonemes in print; they are usually restricted to the level of auditory and oral awareness. Phonics instruction, however, is the educational move from sound to letter-sound correspondences. These correspondences can be at the letter, phoneme, onset/rime, and word levels. Although phonics emphasizes sound and letter patterns, it is necessary to acknowledge the times when those patterns do not apply. For example, in the English language, approximately two-thirds of the words do not follow expected patterns.

Comprehending a text requires background knowledge and vocabulary. Although theoretical differences exist as to the source of meaning in a text, in the primary grades, reading comprehension is typically defined as “the process of simultaneously extracting and constructing meaning through interaction and involvement with written language” (RAND Reading Study Group, 2002: 11). One theory posits that the meaning of a text lies in the print that is on the page, and the reader is called to discover the meaning presented by the author. Another holds that a personal meaning of the text is constructed by the reader. Still another maintains that meaning is created as a transaction among the author, the reader, and their surrounding contexts. Underlying all of these theories is the assumption that a reader brings relevant background knowledge to the task of comprehending. When such knowledge is incomplete or when it departs radically from an author's intended purpose, a reader's comprehension will also be incomplete (Pearson *et al.*, 1979).

Curricular concerns related to vocabulary development at the primary grade level are well founded. When common words that most children know and use daily take on specialized or technical meanings in subject-matter classes, they can cause confusion and lead to problems in comprehending information. Consider for example, the common word *table* and its specialized meanings (a water table, a multiplication table, to table a motion, and so on). Children learn words through independent reading, listening to texts and oral language, direct instruction of specific words, context clues, and morphemic cues. Estimates of the number of words beginning readers need to know vary greatly. The range can be large (e.g., between 2562 and 26 000 words for first graders) and can also vary according to a family's socioeconomic status (White *et al.*, 1990).

Issues in Primary Reading

Policy has increasingly affected curriculum at the national, regional, and local levels. For example, the United Kingdom's National Literacy Strategy introduced in 1997–98 promoted a reading model that assumed a diverse range of instructional approaches to be preferred over a developmentally sequenced reading curriculum. Heavy critiques of this model (Rose, 2006) have led to a more defined reading curriculum being incorporated into the revised primary framework for literacy (House of Commons Education and Skills Committee, 2005). Similar trends toward more systemic reading instruction can be seen in countries such as the United States, with its legislated No Child Left Behind Act of 2001 (Pub. L. 107-110) and Singapore with its definitive national language syllabi (Curriculum Planning and Development Division, Ministry of Education, 2001). Curricular imbalances in favor of reading skills, such as decoding and comprehending, over reading for pleasure have potentially important implications for students' lifelong literacy (House of Commons Education and Skills Committee, 2005).

National policies in other countries have also created tensions by mandating instructional reading materials. For example, in Brazil, the official textbook for poetry instruction focuses young readers' attention on the structure of the genre over the deeper meaning that could be developed by making connections from the poem to themselves and to their world (Malloy and Botzakis, 2005). Although access to digital technology in literacy curricula may be dependent on region and socioeconomic status (Mallozzi and Malloy, 2007), reading domains require redescription for a clickerati generation who explore digital technology without fear (Turbill, 2004: 356). Concepts of print skills such as directionality of print and sequential page-turning are not the most effective mode of understanding how digital print works for

emergent readers exposed to more diverse sources of print; these young readers of a digital age may be better served by a more elastic reading curriculum. Devising and maintaining a reading curriculum is a challenge when the definitions and practices of reading are constantly changing in a politicized and globalized economy.

Elementary/Middle-Grades Reading

Beyond grade 3, the elementary/middle-grades reading curriculum increasingly focuses on the disciplinary or subject-matter aspects of acquiring information from a variety of academic materials (both print and nonprint). Typically, this focus is intended to deepen students' appreciation of reading for purposes of organizing and sharing ideas, revising one's thinking, inquiring more fully into a subject area, and so on. Key curricular questions at this level, which typically consists of grade 4–8 (or children approximately 9–13 years of age), include the following: (1) what role does self-efficacy play in elementary/middle-grade students' willingness to engage with reading in their subject-matter texts? and (2) to what degree is the new information communication technologies (ICTs) an influence on how the elementary/middle-grades reading curriculum is addressing the issue of student engagement with reading?

Elementary and middle graders' perceptions of how competent they are as readers, generally speaking, will affect the degree to which they are motivated to engage with learning in the core curricula (e.g., science, social studies, mathematics, and literature). In pre-adolescence as in earlier and later life, it is the belief in the self (or lack of such belief) that makes a difference in how competent an individual feels when faced with a challenging task, such as reading primary source materials in social studies class or comprehending a math word problem. Although the terms self-concept and self-efficacy are sometimes used interchangeably, they actually refer to different constructs. For example, a student may have a generally good self-concept of herself as a reader, but her answer "Not very" to the question "How confident are you that you can comprehend a primary source on the Battle of Gettysburg?" would indicate low self-efficacy for that particular task (Pajares, 1996).

In an extensive review of how curricular decisions affect elementary and middle-grade students' reading engagement and academic performance, Guthrie and Wigfield (2000) concluded that various reading strategies, while important, do not directly impact student outcomes (e.g., the time they spend reading independently, their performance on high-stakes assessments, and their beliefs about reading). Instead, it was the students' level of engagement in subject-matter reading that was the mediating factor in improved student outcomes. Guthrie and

Wigfield's conception of the engagement model of reading calls for a curriculum that fosters, among other things, self-efficacy and self-monitoring while reading subject-matter texts.

An emerging trend worldwide is a call for considering the multiple realities of ICTs on reading in the elementary/middle-grades curriculum. Based on the results of a survey sent to international literacy research correspondents covering fairly broad regions in South America, Europe, Africa, Pacific Asia, and North America, Botzakis and Malloy (2005) reported that the majority of respondents in all regions were positive in their descriptions of how new ICTs are influencing students' engagement with reading and hence the reading curriculum. At the same time, there were regional differences reflected in some of the respondents' reports. For example, in Argentina and Chile, most teachers who completed the survey indicated that they recognized the importance of electronic media and tried to use new technologies in making reading assignments. On the other hand, they also reported that "many computer labs in schools were scarcely used, except in instances when transcribing or searching for general information were the goals" (p. 113).

The extent to which ICTs are viewed as a positive influence on the elementary/middle grades-reading curriculum hinges in part on perceptions of their usefulness and issues of access. For example, some respondents from Estonia raised a question about whether or not print literacy and the literary curriculum were being diluted due to the influence of electronic media. A lack of access to ICTs was reported as a hardship in certain areas of South Africa, Malaysia, and Brunei. By way of contrast, respondents from Singapore pointed out the ministry of education's role in investing and promoting the use of new ICTs, while respondents in Hong Kong claimed that it was difficult to add anything new, including ICTs, to an already packed curriculum.

Not surprisingly, the key questions and issues that were identified as pertaining to the elementary/middle grades-reading curriculum apply to the secondary level as well where reading demands only increase in intensity and complexity. As Biancarosa and Snow (2004) have noted, "secondary school literacy skills are more complex, more embedded in subject matters, and more multiply determined" (p. 2). Adding to this complexity, particularly in the United States as a result of No Child Left Behind, is a culture of assessment that is essentially rewriting the face of the reading curriculum at the secondary level with little or no regard for input from local educators (Fecho *et al.*, 2007) or for learning through other than traditional print-based texts. On a global level, too, the push for higher standards and improved student achievement in reading is often grounded in print and reflected in a narrowing of the curriculum through overzealously adopting what Luke and Luke (2001) have referred to as an inoculation

approach to curriculum reform (but see the website of Literate Futures, for an exception to this critique).

International comparisons of reading achievement, such as the Program for International Student Assessment (PISA), administered by the Organization for Economic Cooperation and Development (OECD), indicate even more broadly the interest in measuring secondary students' performance levels and their implications for curriculum reform. A case in point is PISA's first report issued in the year 2000, which subsequently has had considerable influence on curriculum development in reading for 15-year-olds among the 32 participating nations, and particularly for those in Central and Eastern Europe (Meredith, 2004). In 2009, the date set for the next PISA survey of 15-year-olds' ability to analyze, reason, and communicate effectively about what they read, the assessment will involve reading electronic texts (OECD, 2007) – a stark contrast to the 2009 Reading Framework of the National Assessment of Educational Progress, a US assessment that will involve reading only traditional print texts. Perhaps advances in technology use may yet be perceived in some areas of the world as undermining long-standing assumptions about a reading curriculum.

Bibliography

- Barr, R. (2001). Research on the teaching of reading. In Richardson (ed.) *Handbook of Research on Teaching*, 4th edn, pp 390–415. Washington, DC: American Educational Research Association.
- Biancarosa, G. and Snow, C. E. (2004). *Reading Next: A Vision for Action and Research in Middle and High School Literacy* (A Report to Carnegie Corporation of New York). Washington, DC: Alliance for Excellent Education.
- Botzakis, S. and Malloy, J. A. (Compilers) (2005). International reports on literacy research: South America, Europe, South Africa, Pacific Asia, and North America. *Reading Research Quarterly* **40**, 112–118.
- Buehl, D. (2007). A professional development framework for embedding comprehension instruction into content classrooms. In Lewis, J. and Moorman, G. B. (eds.) *Adolescent Literacy Instruction: Policies and Promising Practices*. Newark, DE: International Reading Association.
- Curriculum Planning and Development Division, Ministry of Education (2001). *Subject Syllabuses, Language and Literature*. <http://www.moe.gov.sg/cpdd/syllabuses.htm> (accessed October 2009).
- Eakle, A. J. and Garber, A. (Compilers) (2004). International reports on literacy research: Australian, Central and Eastern Europe. *Reading Research Quarterly* **39**, 356–359.
- Fecho B., Mallozzi C. A., and Schultz, K. (2007). Policy and adolescent literacy. In Guzzetti B. G. (ed.) *Literacy for the New Millennium*, vol. 4, pp 37–52. Westport, CT: Greenwood Publishing Company.
- Guthrie, J. T. and Wigfield, A. (2000). Engagement and motivation in reading. In Kamil, M. L., Mosenthal, P. B., Pearson, P. D., and Barr, R. (eds.) *Handbook of Reading Research*, vol. 3, pp 403–422. Mahwah, NJ: Erlbaum.
- House of Commons Education and Skills Committee (2005). *Education and Skills, Eighth Report*. <http://www.publications.parliament.uk/pa/cm200405/cmselect/cmmeduski/121/12102.htm> (accessed October 2009).
- Luke, A. (n.d.) Evidence-based state literacy policy: A critical alternative. http://www.oise.utoronto.ca/projects/impactonpolicy/pdfs/Luke_chapter_may_10_04.pdf.
- Luke, A. and Luke, C. (2001). Adolescence lost/childhood regained: On early intervention and the emergence of the techno-subject. *Journal of Early Childhood Literacy* **1**, 145–180.

- Malloy, J. A. and Botzakis, S. (Compilers) (2005). International reports on literacy research: France, United Kingdom, Brazil. *Reading Research Quarterly* **40**, 514–518.
- Mallozzi, C. A. and Malloy, J. A. (Compilers) (2007). International reports on literacy research: Reading and writing connections. *Reading Research Quarterly* **42**, 161–166.
- Meredith, K. (2004). International reports on literacy research. *Reading Research Quarterly* **39**, 358–359.
- National Assessment Governing Board (2004). *Resolution on the NAEP 2009 Reading Framework*. Washington, DC: NAGB. http://www.nagb.org/release/resolution_09.html (accessed October 2009).
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. (2002). <http://www.ed.gov/policy/elsec/leg/esea02/index.html> (accessed October 2009).
- Organisation for Economic Co-Operation and Development (OECD) (2007). *PISA: The OECD Programme for International Student Assessment*, p 16. <http://www.pisa.oecd.org/dataoecd/51/27/37474503.pdf> (accessed October 2009).
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research* **66**, 543–578.
- Pearson, P. D., Hansen, J., and Gordon, C. (1979). The effect of background knowledge on young children's comprehension of explicit and implicit information. *Journal of Reading Behavior* **9**, 201–210.
- RAND Reading Study Group (2002). *Reading for Understanding: Toward an R&D Program in Reading Comprehension*. Santa Monica, CA: Science & Technology Policy Institute, RAND Education.
- Rose, J. (2006). *Independent Review of the Teaching of Early Reading, Final Report*. <http://www.standards.dfes.gov.uk/phonics/report.pdf> (accessed October 2009).
- Snow, C. E., Griffin, P., and Burns, M. S. (eds.) (2005). *Knowledge to Support the Teaching of Reading*. San Francisco, CA: Jossey-Bass.
- Turbill, J. (2004). International reports on literacy research. *Reading Research Quarterly* **39**, 356–358.
- White, T. G., Graves, M. F., and Slater, W. H. (1990). Growth of reading vocabulary in diverse elementary schools. *Journal of Educational Psychology* **82**, 281–290.

Relevant Website

<http://education.qld.gov.au> – Queensland Government, Department of Education and Training.

Science

Y Kali, Israel Institute of Technology, Haifa, Israel
M Linn, University of California, Berkeley, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Most researchers would agree that science learners engage in a process of making sense of diverse information by looking for patterns and building on their own ideas (Bransford *et al.*, 1999; Bruer 1993). Linn *et al.* (2004) describe this process as knowledge integration and show that students benefit from curricula that elicit ideas and encourage linking and connecting information, seeking and evaluating new ideas, as well as revising and reorganizing scientific ideas to make them more comprehensive and cohesive. They note that designing curriculum materials to support and guide the process of knowledge integration has required creative innovations along with many cycles of trial and refinement. Recent progress in synthesizing design knowledge (Brown 1992; Bransford *et al.*, 1999) has generated principles and patterns that give new designers a head start on creating materials that promote knowledge integration.

In spite of extensive research showing that students learn using knowledge integration processes, textbooks, lectures, and even hands-on activities reflect a view of learners as absorbing information rather than attempting to integrate new ideas with their existing knowledge. Too often students learn isolated facts from science courses, form conjectures based on experience or observation, and gather information from peers or media, but do not sort these ideas out (e.g., Roseman *et al.*, 2001). Students frequently fail to connect ideas learned in science class with observations made in other contexts. Designing curriculum materials that promote knowledge integration requires consideration of the many ideas that students develop. Effective designs encourage learners to compare their ideas, formulate criteria for promoting one idea over another, and link ideas into arguments. However, the translation of what we know about how people learn into curriculum materials is not obvious. Our theoretical knowledge is too abstract to help curriculum designers make important decisions (Kali, 2006). Recent efforts in the area of science curriculum design seek to bridge this gap.

Described here are two complementary approaches for synthesizing science curriculum design knowledge: the design principles approach and the design patterns approach. Both trajectories view learning as a knowledge integration process and aim to synthesize design knowledge from a variety of research projects to make it useful for curriculum designers. We show how the two approaches complement each other by analyzing the design principles

and design patterns in a thermodynamics curriculum unit. Then, we illustrate how the two approaches were used in a curriculum design and development course for graduate students. We also suggest ways in which they can be used in a curriculum design process. We conclude with a discussion of ways to make design knowledge useful.

Curriculum Development and Design-Based Research

Curriculum development is based on epistemological views of the designers. Designers have ideas about how learning takes place, such as a theory or a perspective on learning. They have epistemological assumptions about the nature of knowledge in a specific scientific domain. A good design process is sensitive to the needs and requirements of the users. A needs-analysis stage is therefore an important preliminary step in any curricular design process (Dick *et al.*, 2001). This stage is usually followed by a careful trial and refinement process, until the learning goals of the designed curriculum are met.

Design-based research is an emerging methodology for synthesizing the lessons learned from this trial and refinement process. To draw conclusions from these studies, researchers contrast multiple versions (Barab and Squire, 2004; Collins *et al.*, 2004; The Design-Based Research Collective, 2003). Design-based research studies contribute to what we know about how people learn. They also widen the body of design knowledge available for future curriculum development. To make this body of knowledge, which is sometimes referred to as design knowledge, become useful for curriculum designers, researchers have sought to abstract general guidelines that are based on large numbers of studies, and can assist curriculum designers to make important decisions (e.g., Brown, 1992; Quintana *et al.*, 2004; Merrill, 2002; Mor and Winters, 2007; van den Akker, 1999).

Design Principles and Design Patterns: Complementary Approaches

Researchers have begun to develop two approaches, which use different lenses to analyze promising curricular innovations, find common productive design elements, tie

them with theory, and articulate them as guidelines for future curriculum designers. Described below are two approaches which show how they can be used together in a curriculum design process.

The design principles approach

The design principles approach uses design principles as an organizational unit for synthesizing design knowledge (Merrill, 2002; Quintana *et al.*, 2004). Bell and co-workers describe design principles as:

...an intermediate step between scientific findings, which must be generalized and replicable, and local experiences or examples that come up in practice. Because of the need to interpret design principles, they are not as readily falsifiable as scientific laws. The principles are generated inductively from prior examples of success and are subject to refinement over time as others try to adapt them to their own experiences. (Bell *et al.*, 2004: 83)

Based on this approach, the Design Principles Database (Kali and Linn, 2008) was developed. This database is a mechanism to support researchers and curriculum designers to share their design knowledge in the form of design principles, exemplified by descriptions of features from learning environments. The database is an infrastructure for participants to publish, connect, discuss, and review design ideas, as well as to use these ideas to design new curricula. Thus, it serves as a collaborative knowledge-building tool as well as a mechanism to support design-based research (Kali, 2006). The current entries in the Design Principles Database represent the contributions of over 60 individual researchers. The database includes more than 100 features (mainly from physical, life, and earth sciences), connected with several dozen design principles.

For example, one feature in the database is labeled “Multiple representations in WorldWatcher”. This feature of WorldWatcher software provides scientific visualizations for the investigation of weather data. Students can explore weather data using visualizations that are similar to those found in environments that scientists use (Edelson *et al.*, 1999). The feature displays two-dimensional global data in the form of color maps with latitude and longitude markings and an optional continent outline overlay. A constantly updated readout follows the user’s mouse as it travels over an image, displaying the latitude, longitude, country or state/province, and temperature data value. Different types of representations of the data (such as textual representations of temperature, color schemes representing temperature, location of a mouse on the continent outline, and textual representations of the location as latitude and longitude) are linked, and changed synchronously as the mouse moves.

The feature “Multiple representations in WorldWatcher” is linked in the database to a design principle named “Use

multiple-linked representations”. This principle connects to about ten other features in the database. It provides a general rationale, theoretical underpinning, and important considerations, such as pitfalls, trade-offs, and limits of practical use, to help designers benefit from the many examples. This principle is applied in the thermodynamics curriculum unit described below.

The design patterns approach

The second approach for synthesizing design knowledge to support curriculum designers is the design patterns approach. A design pattern is a sequence of activities in a curriculum, followed by teachers and students in a classroom (Linn and Eylon, 2006). Design patterns based on knowledge integration involve four basic processes: (1) eliciting student ideas, (2) adding new, pivotal ideas, (3) developing criteria for distinguishing among ideas, and (4) sorting out ideas. Linn and Eylon (2006) illustrate how these four processes play out in ten design patterns, which research has shown, can promote knowledge integration (right column in **Table 1**).

For example, one of the ten patterns is “Experiment”. This pattern captures successful sequences of activities that support students in learning from experiments. It describes a recursive process of defining and refining an inquiry question, generating methods for investigating the question, carrying out the investigation, evaluating the results, and using the findings to sort out students’ own repertoire of ideas. When students use activities designed with the experiment pattern, they make decisions about what is a good experiment and what can be learned from an experiment. Students learn to select among varied data collection procedures, distinguish causal from correlational results, and link methods of investigation to the validity of findings (Linn and Eylon, 2006). The experiment pattern can help designers find ways to take advantage of hands-on learning.

This pattern appears in a curricular sequence of activities in the Learning by Design project (Kolodner *et al.*, 2003). Activities in this project interweave design and experimentation cycles. In the experimentation cycle, students begin by clarifying the question to be addressed and make a hypothesis. Then, they design the investigation by identifying conditions that need to be controlled, the variables that will be varied and their values, the steps to be carried out, the number of trials, and what to measure. Finally, students carry out the investigation, record data, analyze results, and present their findings in a poster session (Kolodner *et al.*, 2003).

To illustrate how the design principles and the design patterns approaches complement each other, in the following section we analyze a specific curricular unit in thermodynamics, using these two lenses. The juxtaposition of the two lenses highlights the different ways in which each of the approaches guides design of science curricula.

Table 1 Intuitions in sequences designed by students (left column) and design patterns that provide alternatives (right column)

<i>Intuitions in sequences designed by graduate students in the Kali and Ronen-Fuhrmann (2007) research</i>	<i>Design pattern providing an alternative along with relevant guidance</i>
Focus on flow of content rather than student engagement. Students' sequences focused on which content should be learned first and how to communicate a hierarchy of knowledge. They did not think enough about how to engage learners.	Pattern: <i>orient and elicit</i> . Instructors use the orient-and-elicited pattern to increase interest in a scientific phenomena, define the scope of the topic, connect the topic to personally relevant problems, link the new topic to prior instruction, gauge student interest, and identify the learner's entering ideas.
Providing instructions rather than diagnosing possible confusions. Students tended to start their sequences with instructions, instead of designing guidance that diagnoses learners' possible confusions, or enables learners to link to specific instructions when and if these are needed.	Pattern: <i>diagnose and guide</i> . Instructors use the diagnose-and-guide pattern to elicit the learner's ideas, identify confusions, and determine which ideas to add to stimulate knowledge integration. This pattern promotes knowledge integration by motivating instructors to analyze student thinking and select the more effective new idea.
Providing information rather than supporting inquiry. Students tended to provide information rather than support inquiry; providing learners with answers was often supported by visualizations, but these were usually designed to inform, rather than enable learner-exploration of phenomena.	Pattern: <i>predict, observe, explain</i> . The predict, observe, explain pattern involves providing a demonstration of a scientific phenomenon, eliciting predictions, running the demonstrations, and asking the learner to reconcile contradictions. Pattern: <i>explore a simulation</i> . The explore-a-simulation pattern enables learners to try out their ideas using simulations, virtual worlds, or scientific models. Simulations elicit ideas, support the testing of ideas with feedback, and encourage learners to monitor their performance, often in relationship to their peers or an ideal performer.
Seeking expected answers rather than encouraging negotiation of ideas. Students sought to provide a typical correct answer, or process, rather than illustrate a variety of ideas, and encourage negotiation of those ideas among peers.	Pattern: <i>illustrate ideas</i> . Using the diagnose-and-guide pattern described above, instructors illustrate their ideas by contrasting various perspectives – including those held by their audience – and discussing how a learner could go about distinguishing among them. Pattern: <i>collaborate</i> . In this pattern learners generate their own ideas, learn from the ideas of others, respond to group ideas, determine methods for distinguishing ideas, articulate warrants for their views, and reach consensus. Pattern: <i>critique</i> . In this pattern learners are asked to evaluate both established and potentially invalid, misleading, persuasive, or confusing information presented in Internet resources, textbooks, articles, models, experiments, arguments, or peer reports. Pattern: <i>reflect</i> . In this pattern learners are encouraged to analyze the connections they make and consider their prior ideas. The reflect pattern promotes knowledge integration by encouraging students to stop and analyze their own repertoire of ideas.
Recreations rather than enabling learners to design their own artifacts. Students tended to engage learners in predefined activities and rarely enabled learners to design their own artifacts	Pattern: <i>create an artifact</i> . The create-an-artifact pattern involves refining a question, selecting or using methods for creating an artifact, creating a draft artifact, evaluating the results, using the findings to improve the artifact, and connecting the results to views of the topic.

Design Principles and Patterns in the Thermodynamics Curriculum Unit

“Thermodynamics: probing your surroundings” (Clark and Sampson, 2007) is a curriculum unit for seventh- to ninth-grade students, which is part of the Technology-Enhanced Learning in Science program. The unit, which requires about 1 week of instruction, and uses the Web-based Inquiry Science (WISE) infrastructure, is designed to assist students to connect their observations (such as noting that metal feels cooler than wood at room temperature), and the explanations of these phenomena at the molecular level.

The design of this unit can be analyzed using design principles as well as design patterns. This dual analysis is described below and summarized in **Figure 1** (see below). The design principles and patterns are concisely summarized here. Information about each of the principles can be found in the Design Principles Database. The design patterns are explained in more detail in Linn and Eylon (2006).

The first activity in the unit, named “What do you think?” (**Figure 1**), introduces a driving question and elicits students' ideas about thermodynamics. (People talk about objects being naturally hot or naturally cold – what do they mean?) Students make predictions about the

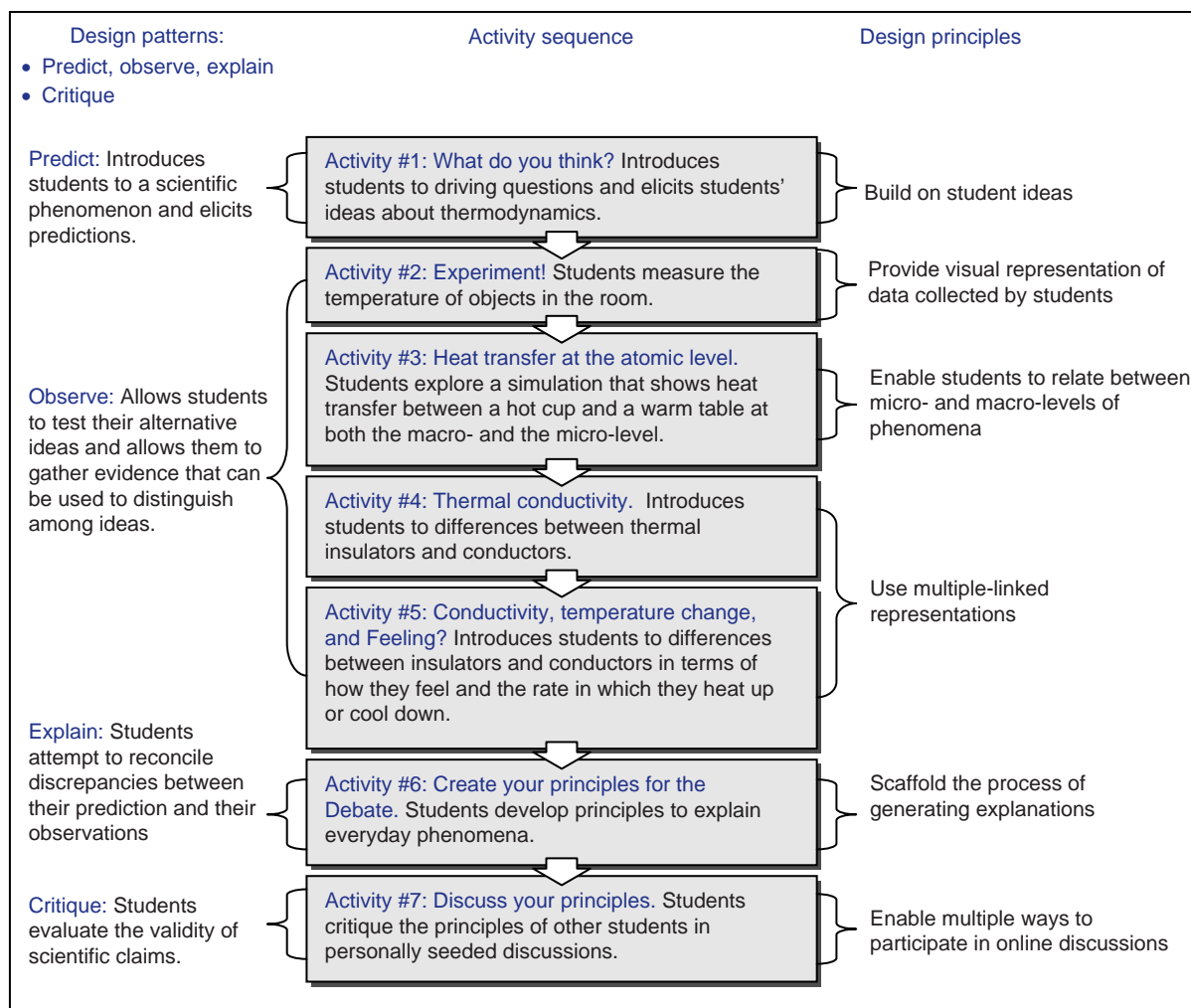


Figure 1 The sequence of activities in the thermodynamics curriculum unit described in terms of design patterns (left) and design principles (right). From Clark, D. B. and Sampson, V. D. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education* 29, 253–277.

temperature of various objects in the room, and record them online. If we analyze this feature with a design principles lens, we can see that it employs a most prominent principle for designing scientific curricula called “Build on student ideas”. This design principle calls for designing instruction that encourages students to build on their ideas to develop increasingly powerful and useful scientific ideas, rather than isolate new information from existing knowledge. This first activity (**Figure 1**) also illustrates the design patterns approach. The activity realizes the predict part of a design pattern named “Predict observe explain” (**Table 1**). The other segments (observe and explain) follow in activities as the unit progresses.

In the second activity of this unit, students measure the temperature of the objects and compare the measurements to their predictions using a feature that graphically presents predicted versus actual temperatures. From a design principles view, this feature follows the principle “Provide visual representation of data collected by

students”. The tool enables students to easily represent data they collect (temperature, time, and data from surveys) in interactive dynamic representations. Graphing data collected by students can promote their understanding of the phenomenon they are studying. Using tools that automatically graph data collected by students decreases the workload involved in graphing. From a design patterns view, this one, and the next three activities represent the observe section of the “Predict observe explain” pattern described above (**Figure 1**; **Table 1**).

Students’ exploration continues in the third activity with a simulation that shows heat transfer between a hot cup and a warm table at both the observable and molecular level. The simulation applies the design principle “Enable students to relate macro- and micro-levels of phenomena”. This principle calls for creating features that help students see the connection between their observations and molecular processes by animation that allows students to zoom into the phenomena until molecules

can be seen. Then, in the fourth and fifth activities, students learn about the differences between thermal insulators and conductors. They choose different objects with various conductivities and different starting temperatures for each object, and view animated graphs showing how their temperature changes as they heat up or cool down. This animation realizes the design principle described above – “Use multiple-linked representations”.

The sixth and seventh activities, which culminate the thermodynamics unit, enable students to develop their own principles for explaining everyday phenomena in terms of concepts in thermodynamics, and to discuss these principles with their peers. Using a tool called “Principle maker”, students create general principles that summarize their understanding of simulations and of data collected in previous stages of the project. The students use a series of pull-down menus in order to construct a principle. Each pull-down menu gives a list of possible phrases to choose from. Finally, the principles created in activity 6 are used as seeds for online discussions between students in activity 7.

Activities 6 and 7 represent curriculum features that employ two design principles: “Scaffold the process of generating explanations” and “Enable multiple ways to participate in online discussions”. From a design patterns view, when students use the “Principle maker”, they accomplish the explain section of the “Predict observe explain” pattern, followed throughout the sequence of activities 1–6. The online discussion, seeded with students’ principles in the final activity, represents a new pattern – the “Critique” pattern, which guides designers to create sequences of activities that support students in careful examination of any information they encounter, including ideas brought to the discussion by their peers (Table 1).

The above analysis is only one example of ways in which a curriculum unit can be analyzed using design principles and design patterns. Curricular features and sequences can realize more than one design principle or design pattern. Principles and patterns can be embedded into one another. For instance, activity 3 in the thermodynamics unit is part of the “Predict observe explain” sequence and also implements the “Explore a simulation” pattern as part of the observation. This pattern enables designers to create sequences of activities that specifically support students in making sense of new science ideas represented in scientific models and simulations (Table 1).

Using Principles and Patterns to Guide a Curriculum Design Process

Researchers are beginning to tap the potential use of frameworks, such as the design principles and design

patterns described above to improve the design process. Research highlights the affordances and challenges they pose for designers. To assess the use of the design principles approach, Kali and Ronen-Fuhrmann (2007) developed a course to guide graduate students in designing technology-based curricula using the Design Principles Database. Students use the database when they: (1) write a needs and analysis document for the curriculum they are designing, (2) brainstorm ideas for features, (3) create a flow of activities, (4) design the features in the activity flow, and (5) evaluate each others’ designs.

Initial enactments of this model in graduate courses showed that design principles were useful to students in several of these stages. For instance, the use of design principles at the brainstorming stage helped students to widen the scope of their design. Students were more likely to include socioconstructivist features than before they used the principles. Additionally, after students figured out a sequence of activities for the curricula, the Design Principles Database helped them design innovative features for each activity in the sequence and, thus, build on ideas tested and refined by other design-based research studies. However, designing the sequence of activities was a challenge for many of the students. They tended to rely on their intuitive absorption model of the learner and: (1) create a hierarchical flow of content rather than engaging students, (2) provide instructions rather than diagnose possible confusion, (3) provide information rather than support explanation, (4) seek predefined answers rather than encourage negotiation, and (5) create actions rather than enabling learners to design artifacts (Kali and Ronen-Fuhrmann, 2007; Table 1).

To remedy this problem, the next version of the course will employ design patterns. Table 1 illustrates how the patterns approach could help graduate students, in the Kali and Ronen-Fuhrmann (2007) study, consider alternative approaches and analyze their intuitive designs from another perspective.

The above analysis shows the benefit of using design patterns in the create-a-flow-of-activities stage of the design process. Recently, the design patterns have been added to the publicly accessible Design Principles Database, so that novice curriculum designers will be able to use design principles as well as design patterns throughout their design process.

Summary and Next Steps

Current curriculum materials, including textbooks (Roseman *et al.*, 2001) and lectures (Linn and Eylon, 2006), fail to take advantage of advances in understanding of the learner and of effective instruction. Since the 1990s, researchers have sought learning principles (Brown, 1992)

and design knowledge (diSessa, 1992) to remedy this situation. With the advent of technology-enhanced learning environments, many research partnerships have used embedded assessments to gather detailed accounts of student learning of complex science concepts (e.g., Clark and Sampson, 2007; Kolodner *et al.*, 2003). These research programs feature trial and refinement of curricular materials using evidence from student learning (The Design-Based Research Collective, 2003). Efforts to synthesize which refinements succeed have typically involved generating design principles (Collins *et al.*, 2004; Dick *et al.*, 2001; Linn *et al.*, 2004; van den Akker, 1999) or design patterns (Linn and Eylon, 2006).

Recently, researchers have begun to study ways to synthesize design knowledge so others can use it. Kali (2006) created the public Design Principles Database and engaged the community in contributing its design knowledge. Kali and Ronen-Fuhrmann (2007) created a curricular design course for graduate students and studied the role of principles in the design process. Based on analyses of student designs, they found that principles were not sufficient to help students take advantage of research on instruction. The most recent version of the course and the Design Principles Database now also includes design patterns. These first steps toward creating effective, proven designs suggest promising practices for all designers. Future research must apply this rigorous, evidence-based approach to the design of widely used materials, such as textbooks.

Acknowledgment

This material is based upon work supported by the National Science Foundation under grant No. ESI-0334199.

Bibliography

- Barab, S. A. and Squire, K. D. (2004). Design-based research: Putting our stake in the ground. *The Journal of the Learning Sciences* **13**, 1–14.
- Bransford, J. D., Brown, A. L., and Cocking, R. R. (eds.) (1999). *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Research Council.
- Bell, P., Hoadley, C. M., and Linn, M. C. (2004). Design-based research in education. In Linn, M. C., Davis, E. A., and Bell, P. (eds.) *Internet Environments for Science Education*, pp 73–85. Mahwah, NJ: Lawrence Erlbaum Associates.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences* **2**, 141–178.
- Bruer, J. T. (1993). *Schools for Thought: A Science of Learning in the Classroom*. Cambridge, MA: MIT Press.
- Clark, D. B. and Sampson, V. D. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education* **29**, 253–277.
- Collins, A., Joseph, D., and Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences* **13**, 15–42.
- Dick, W., Carey, L., and Carey, J. O. (2001). *The Systematic Design of Instruction* (5th edn.), pp 1–34. New York: Longman.
- diSessa, A. (1992). Images of learning. In De Corte, E., Linn, M. C., Mandl, H., and Verschaffel, L. (eds.) *Computer-Based Learning Environments and Problem Solving*. Berlin: Springer.
- Edelson, D., Gordin, D., and Pea, R. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *Journal of the Learning Sciences* **8**, 391–450.
- Kali, Y. (2006). Collaborative knowledge-building using the design principles database. *International Journal of Computer Support for Collaborative Learning* **1**, 187–201.
- Kali, Y. and Linn, M. C. (2007). Technology-enhanced support strategies for inquiry learning. In Spector, J. M., Merrill, M. D., van Merriënboer, J. J. G. V., and Driscoll, M. P. (eds.) *Handbook of Research on Educational Communications and Technology*, 3rd edn., pp 445–490. Mahwah, NJ: Erlbaum.
- Kali, Y. and Ronen-Fuhrmann, T. (2007). How can the design of educational technologies affect graduate students' epistemologies about learning? In Chinn, C. A., Erkens, G., and Puntambekar, S. (eds.) *Proceedings of the Computers Supported Collaborative Learning (CSCL), Mice, Minds and Society*, pp. 320–322. International Society of the Learning Sciences, Inc.
- Kolodner, J. L., Camp, P. J., Crismond, D., *et al.* (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting Learning by Design™ into practice. *Journal of the Learning Sciences* **12**, 495–547.
- Linn, M. C., Davis, E. A., and Bell, P. (eds.) (2004). *Internet Environments for Science Education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Linn, M. C. and Eylon, B.-S. (2006). Science education: Integrating views of learning and instruction. In Alexander, P. A. and Winne, P. H. (eds.) *Handbook of Educational Psychology*, 2nd edn., pp 511–544. Mahwah, NJ: Lawrence Erlbaum Associates.
- Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development* **50**, 43–59.
- Mor, Y. and Winters, N. (2007). Design approaches in technology enhanced learning. *Interactive Learning Environments* **15**, 61–75.
- Quintana, C., Reiser, B. J., Davis, *et al.* (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences* **13**, 337–386.
- Roseman, J. E., Kulm, G., and Shuttleworth, S. (2001). Putting textbooks to the test. *ENC Focus* **8**, 56–59.
- The Design-Based Research Collective (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher* **32**, 5–8.
- van den Akker, J. (1999). Principles and methods of development research. In van den Akker, J., Nieveen, N., Branch, R. M., Gustafson, K. L., and Plomp, T. (eds.) *Design Methodology and Developmental Research in Education and Training*, pp 1–14. The Netherlands: Kluwer Academic.

Further Reading

- Alexander, C., Ishikawa, S., and Silverstein, M. (1977). *A Pattern Language: Towns, Buildings, and Construction*. New York: Oxford University Press.
- Collins, A. (1992). Toward a design science of education. In Scanlon, E. and O'Shea, T. (eds.) *New Directions in Educational Technology*, pp 15–22. Berlin: Springer.
- diSessa, A. A. (2000). *Changing Minds: Computers, Learning and Literacy*. Cambridge, MA: MIT Press.
- Simon, H. A. (1969). *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

Relevant Websites

<http://www.design-principles.org> – Design principles database.
<http://www-static.cc.gatech.edu> – Learning by Design.

<http://telscenter.org> – Technology-Enhanced Learning in Science (TELS).
<http://wise.berkeley.edu> – Web-based inquiry science environment (WISE).
<http://www.worldwatcher.northwestern.edu> – World watcher.

CURRICULUM DEVELOPMENT – THEORY

Contents

Bilingual Learning (Learning L1 and L2 in an L1 and L2 Environment)
Curriculum and Complex Systems Theory
Curriculum and Constructivism
Curriculum and Critical Theory
Curriculum and Poststructuralist Theory
Curriculum and Structuralist Sociology: The Theory of Codes and Knowledge Structures
Curriculum Reform
Early Childhood Curriculum and Developmental Theory

Bilingual Learning (Learning L1 and L2 in an L1 and L2 Environment)

D Schwinge, Adelphi University, Garden City, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

With the growth of the global economy, more and more countries are recognizing that bilingualism is a national resource that can provide economic, social, and cultural advantages. Nations, local communities, and individuals are examining how the growth of bilingualism can be cultivated. Significant research has been conducted to examine how to help individuals to acquire multiple languages and how to improve bilingual learning in multilingual and multicultural environments. The opening section of this article discusses how multilingual environments for language learning can be created through the implementation of different models of bilingual education in public schools. Then, the article discusses how bilingual language acquisition can occur in homes, neighborhoods, and community-based heritage-language schools. In addition, since the acquisition of biliteracy is important for those who want to use their language abilities in formal contexts, such as higher education and the workplace, the article also examines theoretical models and practical techniques for developing biliteracy. While bilingual learning is embraced by many, there are still significant misconceptions about how the acquisition of two or more languages in a multicultural environment will influence overall linguistic and cognitive development. Thus, this article also examines language transfer, codeswitching, and cognitive development in bilingual individuals and shows how the research in these areas supports the idea that the acquisition of bilingualism is helpful for overall personal and academic growth.

Models of Bilingual Education

The bilingual learning of school-age students is strongly influenced by the organization of bilingual education in the formal educational contexts in which the students receive their elementary and secondary education. Bilingual learning is assisted by schools that provide extensive opportunities for students to achieve high linguistic proficiency through studying in two or more languages. Bilingual education can be provided in weak or strong forms. An example of a weak form of bilingual education is a program that provides a transitional bilingual education. The goal of these programs is to have students begin to study in the majority language as soon as possible. These programs are often seen as a track for low-achieving students or as a remedial program. Weak bilingual programs usually provide students with only a limited number of years of bilingual instruction (usually 2–3 years) before exiting the students into a monolingual educational environment.

In contrast, strong models of bilingual education, often referred to as enrichment bilingual programs, have a goal of helping students to achieve a high level of proficiency in two or more languages. These programs generally provide students with bilingual instruction throughout the students' time in the school, provide content-based instruction in multiple languages, and also seek to help students to become biliterate and bicultural. While there are many variations in how to organize bilingual education programs, there are three models of enrichment bilingual education that are frequently adopted. These

models include developmental bilingual education programs, immersion bilingual education programs, and two-way bilingual programs.

Developmental bilingual education programs, also known as maintenance or heritage-language bilingual education, provide students from language minority groups with a program that uses the minority language as the medium of instruction about half the time, and provides content-based instruction in both languages. This form of bilingual education is most often used where there are populations of recent immigrants, but it is also a popular model to encourage the maintenance of indigenous languages, such as Navajo in the United States (McCarty and Bia, 2002) and Maori in New Zealand (May, 2004).

Second, immersion bilingual education programs are designed to develop bilingualism in language majority students who are provided with intensive language-learning experiences and content-based instruction in a new language through multiple years of schooling. Most students enter immersion programs with little or no knowledge of the second language. Many of these programs offer early total immersion where students study in the new language for 100% of their first years in preschool and elementary school. While immersion programs are used around the world, this model of bilingual education has been particularly popular and successful in Canada (Johnson and Swain, 1997).

Third, two-way bilingual education programs, also sometimes referred to as dual-language or dual-immersion programs, are designed to educate balanced numbers of students from two language backgrounds. These programs allow students who are native speakers of different languages to study together so that all students develop high levels of proficiency in both languages as well as authentic cultural understanding. Research on these programs has shown that they not only aid students in becoming biliterate and bicultural, but they also often succeed in elevating the status of minority languages and their speakers in the community (Freeman, 1998). Dual-language classes are most effective when there are a balanced number of students from each language group in the class, and the languages are kept separate in the classroom. Language separation can occur in several ways. In some dual-language classes, two teachers can be used in the program and each teacher instructs the students using only one language. Other dual-language classes divide language use by time of day or by subject matter. There are different models of language use in dual-language programs. Some programs use each language 50% of the time in each grade. Other programs adopt a 90/10 model where the minority language is used for 90% of the time in kindergarten, but the use of the majority language increases in each grade so that the school day has a 50% language allocation for each language by the end of elementary school. Dual-language programs have increased rapidly in the last decade, and these programs have been shown to be the most effective of the types of bilingual programs.

The Importance of Families and Communities in Bilingual Learning

While schools may be one important site of bilingual learning, the home and community environments of bilingual learners also play a large part in the acquisition of competence in two or more languages. This section of the article reviews bilingual language acquisition at home and in the neighborhood, community heritage-language schools, and how local funds of community knowledge can be used to enhance bilingual language learning in formal educational settings.

Bilingual Language Acquisition at Home and in the Neighborhood

While the school can be a site of bilingual learning, the access to bilingual linguistic input in the home environment also plays an essential role for many learners by providing extensive exposure to multiple languages. Bilingualism in the home is especially important for the language acquisition of young children who often spend most of their time with their primary caretakers. While it is frequently reported that doctors or preschool teachers or other professionals have recommended to parents who are speakers of minority languages that they do not speak their native language with their young children, it is increasingly being recognized that early use of two or more languages in the home environment can provide children with advantages in language acquisition. Through using multiple languages for conversations in the home, sharing folklore or traditional stories, being exposed to multiple languages in religious settings, or taking part in ordinary daily activities such as letter writing to relatives, children can achieve a high level of bilingualism even if they are not exposed to a bilingual school environment.

One additional, and often-overlooked, setting for bilingual language acquisition is the interactions that learners have with their peers. As children get older, the peer group becomes a more important resource for bilingual learning. Older siblings often teach younger siblings vocabulary and other language forms that they can use to take part in games or sports with other children in the neighborhood, purchase needed items at local stores, and complete daily chores. One interesting example of research on how peers can positively influence bilingual language acquisition is described in Long (1997) who gives an account of how her daughter learned significant amounts of Icelandic from interactions with her peers after the family moved from the United States to Iceland for a year. She shows how her daughter acquired Icelandic through play experiences with her friends and from her participation with her peers in extracurricular activities after school. Certainly, a substantial amount of bilingual language learning can occur in the social environments of the home and community.

Heritage-Language Instruction

Heritage-language instruction in the community can be an important support for bilingual learning in the family and in the regular school classroom (Fishman, 2006). Heritage-language schools outside of the formal public school system often have classes that meet after school or on the weekend. These schools are organized by a number of different types of organizations, including religious groups, foreign governments, and local community ethnic associations. While heritage-language schools often concentrate on early childhood and elementary school children, some community heritage-language schools also teach teenagers and adults. One reason why community heritage-language schools are important is that they provide instruction in lesser-known and indigenous languages that are not included in the public school curriculum. They often also focus on transmitting cultural knowledge, a sense of cultural identity, and respect for the cultural traditions of the community. Thus, they often use activities such as dance classes or cooking programs as a vehicle for language instruction.

Local Funds of Knowledge

In recent years, substantial research has been done to demonstrate ways that the linguistic and cultural resources of the home and community environment can be drawn upon to improve students' educational experiences in public schools. It has been found that bilingual learning in schools and formal educational settings can be enhanced by using the local or household funds of knowledge that students have gained from participating in their families and communities. Funds of knowledge include the information, methods of thinking and learning, and the practical skills that are related to a community's everyday life. This knowledge can include linguistic and factual information about a variety of topics, including agriculture, economics, household management, medicine, scientific knowledge, and religion. Utilizing local funds of knowledge can allow parents to participate more fully in their children's education, and encourage students to place a higher value on the languages and knowledge that is taught to them in their homes and communities. Research also suggests that utilizing community knowledge and languages allows students to participate in activities and produce written products that they would be unable to produce if they were forced to utilize only traditional academic knowledge in completing their assignments (González *et al.*, 1995).

Funds-of-knowledge research has been carried out in a variety of different cultural and linguistic settings, and this type of research has added to our knowledge of the wide variety of vernacular and home literacy practices that can be linked to school learning. For example, Martin-Jones and Saxena (2003) researched how classroom bilingual teaching assistants in northwest England used their knowledge of the

students' home languages and cultures to create a variety of culturally appropriate literacy events. The teaching assistants used activities such as cooking chapattis and telling stories about the Sikh New Year to teach students academic content knowledge and literacy skills.

Biliteracy

For many learners, especially those who will use two or more languages in academic and work environments, developing biliteracy is an essential element of bilingual learning. Bilingual learners have opportunities to become biliterate when they are exposed to activities or literacy events where two languages are used for reading and writing. This section first discusses a major theoretical model of how bilingual learners acquire biliteracy and then discusses how an integrated approach to bilingual learning can aid learners in achieving competency in reading and writing in two languages.

Continua of Biliteracy

There are many important factors to consider when attempting to create a home or school environment that aids the acquisition of biliteracy. For example, decisions have to be made about whether to expose learners to the written forms of the two languages simultaneously or whether to teach reading and writing first in one language before introducing the writing system of the second language. In addition, bilingual learning can focus primarily on being able to read literary texts from the majority language or instruction can include vernacular texts and texts that include cultural content and information from the minority language culture as well. The model of the continua of biliteracy is a framework that suggests different possibilities for the acquisition of biliteracy by bilingual learners. The model suggests that the more a learning context allows learners to draw on the entire continuum of possible linguistic varieties, modes of expression, and types of knowledge in learning events, the greater is the chance of the learners to develop their bilingualism and biliteracy to the fullest extent possible. One important aspect of this model is that each continuum consists of weaker and more powerful ends, and the continua thus recognize that all modes of expression and types of knowledge are not viewed as equally powerful by society. This framework has been described in great detail and applied to many different educational situations in a recent edited volume (Hornberger, 2003).

Integrated Approach to Biliteracy

In addition to the major theoretical contributions that have added to our understanding of biliteracy, contributions

have also been made in creating an integrated approach to the teaching of biliteracy in bilingual classrooms. One text that gives a wealth of information about teaching reading and writing in two languages in a classroom context is *Learning in Two Worlds: An Integrated Spanish/English Biliteracy Approach* (Pérez and Torres-Guzmán, 1996). This text provides information about how to develop biliteracy in learners who have many different levels of proficiency, including emergent biliteracy, the development of basic reading and writing skills, and advanced proficiency including the reading and writing of content-area texts.

There is also research available that gives examples through case studies or samples of student work of how biliteracy can be developed and the high levels of biliteracy that can be achieved by some learners. Cox and Boyd-Batstone (1997) present three detailed longitudinal case studies of children's biliterate development between first and fifth grade, and also give a detailed account of how the classroom teachers used literature and various types of reading responses to increase the students' acquisition of English and Spanish in a school setting. In addition to content-area reading and writing, there has also been some research conducted in how biliterate individuals can use reading and writing for creative expression. Ada and Campoy (2003) describe how they have used bilingual children's literature and a variety of creative writing activities to help students draw on their cultural knowledge and their authentic life experiences. The text describes ten thematic units that are each based on a different genre of creative writing. One of the key features included in this text is the engaging examples of creative writing by teachers, parents, and students. Many of these are pictures of home-made books with illustrations or poems that show the high quality of writing that can be produced by young biliterate authors.

Common Misconceptions About Bilingual Learning

Among the general public, there are some misconceptions about the effects that bilingual learning will have on the linguistic and cognitive development of young bilingual learners. This section introduces three common concerns of parents and teachers who are considering whether to provide children with a bilingual education. People are often concerned that perhaps bilingual learning will cause learners to mix up the two languages that they are learning or will have negative overall cognitive effects. This section discusses how recent research findings demonstrate that bilingual learning is not detrimental in any way to language acquisition or overall cognitive development, but instead can confer an educational advantage for the bilingual learner.

Language Transfer

A common myth of bilingual learning is that the acquisition of multiple languages simultaneously will cause difficulties in language learning, especially in learning to read and write. However, researchers have shown that instead there is significant transfer of linguistic skills that actually gives bilingual learners an advantage in language acquisition. Many different language skills can be transferred, including a general understanding of how print works, reading comprehension strategies, spelling, a familiarity with the characteristics of specific genres, and an understanding of pragmatics (how language is used in context). The importance of language transfer was demonstrated in a foundational study by Edelsky (1986). Edelsky's research focused on how first through third-grade students in an English and Spanish bilingual program transferred the writing skills they had gained in one language to the second language that they were acquiring. This study also demonstrated how the students' bilingual competence gave them the ability to write for varying audiences in a larger range of genres.

While there is evidence of linguistic transfer in bilingual learning with many combinations of languages, recent research shows that there is a greater advantage in biliteracy acquisition when the two languages that are being learned share a similar writing system (Bialystok *et al.*, 2005). When languages have similar writing systems, it has even been shown that sometimes students can develop spontaneous biliteracy, that is, the ability to read in a second language even when they have received no substantial formal literacy instruction in the language. However, even when languages use varying writing systems, there is still the possibility for transfer to occur. For example, when researchers in England asked students who were in Saturday supplemental heritage-language programs in Chinese, Arabic, and Spanish to teach their classmates the additional writing systems that they were learning, it was observed that 5- and 6-year-old students understood the key differences between the writing systems they were acquiring, including the differences between alphabetic and logographic writing systems and the differing directionality in various writing systems. In addition, the heritage-language students were able to share this information clearly and in creative ways with other children (Kenner *et al.*, 2004).

Codeswitching

Codeswitching is mixing one language with another when speaking or writing, and it occurs frequently in environments where bilingual individuals share the same set of languages. Monolinguals often view codeswitching as something that indicates a lack of language proficiency. However, research has shown that codeswitching is a

complex use of language, and there are specific linguistic rules for what types of mixing can occur between languages in a sentence or conversation. Codeswitching can be utilized by bilingual learners for many purposes. While it is possible that bilingual learners may utilize codeswitching when they are unsure of a vocabulary word in a language, codeswitching is also used to show an affinity with others who speak the same language varieties or it may be used to emphasize a particular point in a conversation. One part of bilingual learning is acquiring an understanding of when it is appropriate to codeswitch and what type of codeswitching is appropriate in a particular linguistic and cultural context. One additional complexity in codeswitching is that in many cases, interlocutors utilize a range of dialects as they are codeswitching. For example, in Zentella's (1997) research of the language use of a low-income Puerto Rican community in New York City, she found that the children she was studying codeswitched between a whole range of linguistic codes, including standard Spanish, nonstandard Puerto Rican Spanish, standard English, Hispanized English, and African-American vernacular English.

Codeswitching has also been studied as part of a larger set of practices that can enhance bilingual learning. In some learning situations, it is possible for bilingual teachers and learners to purposefully draw upon multiple languages, semiotic modalities, and participation structures in a single lesson or learning experience in an organized and principled way. This mixing is sometimes referred to as hybridization or hybrid literacy practices. Gutiérrez *et al.* (1999) give an example of hybrid literacy practices in their analysis of an after-school program. Their research describes how elementary school students describe their activities and learning in daily e-mails to a fictional bilingual character, El Maga. The undergraduate students who answer these e-mails purposefully use both Spanish and English in a strategic way to help the bilingual students develop their literacy skills in both languages, and develop a strong bond with El Maga. This example shows how specific hybrid literacy practices can be used to reinforce learners' identity as bilinguals who are able to use codeswitching effectively.

Cognitive Effects of Bilingualism

Until the 1960s, it was commonly believed that bilingual learning had a negative effect on cognitive development in children. However, there are now many studies that show that children who acquire a second language may actually have cognitive advantages over their monolingual peers. These advantages occur when learners have reached a threshold level of proficiency, a level of basic proficiency, in both languages. It has been shown that children who are

bilingual tend to show a heightened metalinguistic awareness, especially during early childhood. Bilingual children are able to more easily understand certain linguistic concepts, such as phonological awareness or the ability to count the words in a sentence, earlier than monolingual learners. Research also suggests that bilingual language acquisition can improve creativity in children and increase their ability to complete tasks which require divergent thinking in order to solve problems.

Summary

With increasing levels of migration and internationalism, the number of multilingual communities is expanding and the need for bilingual individuals is increasing. Research has shown that creating opportunities for bilingual learning is possible and that bilingualism confers individual and societal advantages. Despite this, much work remains to be done to learn how to apply the best practices in organizing school bilingual programs, increase opportunities to use multiple languages in home and community settings, and develop teaching techniques to help students achieve full academic biliteracy. The potential to achieve societal bilingualism and protect indigenous languages is not fully realized, often because individuals have misconceptions that bilingualism might have negative effects on the development of children or on the solidarity of the nation. However, the large amount of research in bilingual learning environments from all around the world shows the incredible creative potential that exists and the high levels of bilingualism that can be achieved when schools, communities, and families work together to build and maintain environments that promote bilingualism.

Bibliography

- Ada, A. F. and Campoy, F. I. (2003). *Authors in the Classroom: A Transformational Education Process*. Boston, MA: Allyn and Bacon.
- Bialystok, E., Luk, G., and Kwan, E. (2005). Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems. *Scientific Studies of Reading* 9, 43–61.
- Cox, C. and Boyd-Batstone, P. (1997). *Crossroads: Literature and Language in Culturally and Linguistically Diverse Classrooms*. Upper Saddle River, NJ: Merrill.
- Edelsky, C. (1986). *Writing in a Bilingual Program: Había una vez*. Norwood, NJ: Ablex Publishing.
- Fishman, J. A. (2006). 300-plus years of heritage language education in the United States. In Valdés, G., Fishman, J. A., Chávez, R., and Pérez, W. (eds.) *Towards the Development of Minority Language Resources*, pp. Clevedon: Multilingual Matters.
- Freeman, R. (1998). *Bilingual Education and Social Change*. Clevedon: Multilingual Matters.
- González, N., Moll, L. C., Tenery, M. F., *et al.* (1995). Funds of knowledge for teaching in Latino households. *Urban Education* 29, 443–470.

- Gutiérrez, K. D., Baquedano-López, P., and Tejada, C. (1999). Rethinking diversity: Hybridity and hybrid language practices in the third space. *Mind, Culture, and Activity: An International Journal* 6, 286–303.
- Hornberger, N. H. (ed.) (2003). *Continuum of Biliteracy: An Ecological Framework for Educational Policy, Research and Practice in Multilingual Settings*. Clevedon: Multilingual Matters.
- Johnson, R. K. and Swain, M. (1997). *Immersion Education: International Perspectives*. Cambridge: Cambridge University Press.
- Kenner, C., Kress, G., Al-Khatib, H., Kam, R., and Tsai, K. (2004). Finding the keys to biliteracy: How young children interpret different writing systems. *Language and Education* 18, 124–144.
- Long, S. (1997). Friends as teachers: The impact of peer interaction on the acquisition of a new language. In Gregory, E. (ed.) *One Child, Many Worlds: Early Learning in Multicultural Communities*, pp. New York: Teachers College Press.
- Martin-Jones, M. and Saxena, M. (2003). Bilingual resources and ‘funds of knowledge’ for teaching and learning in multi-ethnic classrooms in Britain. *International Journal of Bilingual Education and Bilingualism* 6, 267–282.
- May, S. (2004). Maori-medium education in Aotearoa/ New Zealand. In Tollefson, J. W. and Tsui, A. B. M. (eds.) *Medium of Instructional Policies: Which Agenda? Whose Agenda?*, pp. Mahwah, NJ: Erlbaum.
- McCarty, T. L. and Bia, F. (2002). *A Place to Be Navajo: Rough Rock and the Struggle for Self-Determination in Indigenous Schooling*. Mahwah, NJ: Erlbaum.
- Pérez, B. and Torres-Guzmán, M. E. (1996). *Learning in Two Worlds: An Integrated Spanish/English Biliteracy Approach*, 2nd edn. White Plains, NY: Longman.
- Zentella, A. C. (1997). *Growing up Bilingual*. Malden, MA: Blackwell.
- Bialystok, E. (2001). *Bilingualism in Development: Language, Literacy, and Cognition*. Cambridge: Cambridge University Press.
- Cloud, N., Genesee, F., and Hamayan, E. (2000). *Dual Language Instruction: A Handbook for Enriched Education*. Boston, MA: Heinle.
- Crawford, J. (2004). *Educating English Learners: Language Diversity in the Classroom*, 5th edn. Los Angeles, CA: Bilingual Educational Services.
- Cummins, J. and Hornberger, N. H. (eds.) (2008). *Encyclopedia of Language and Education. Volume 5: Bilingual Education*, 2nd edn. New York: Springer Science/Business Media LLC.
- García, O. (2009). *Bilingual Education in the 21st Century: A Global Perspective*. Malden, MA: Wiley-Blackwell.
- Krashen, S. D. (1999). *Condemned Without a Trial: Bogus Arguments against Bilingual Education*. Portsmouth, NH: Heinemann.
- Kroll, J. F. and DeGroot, A. (2005). *Handbook of Bilingualism: Psycholinguistic Approaches*. New York: Oxford University Press.
- Perez, B. (2004). *Becoming Biliterate: A Study of Two-Way Bilingual Immersion Education*. Mahwah, NJ: Erlbaum.
- Valdés, G. (1996). *Con respeto: Bridging the Distances between Culturally Diverse Families and Schools*. New York: Teachers College Press.

Relevant Websites

- <http://www.cal.org> – Center for Applied Linguistics.
- www.usc.edu – Center for Multilingual Multicultural Research at the University of Southern California.
- www.iteachilearn.com – Dr. Cummins' ESL and Second Language Learning Website.
- <http://www.nabe.org> – National Association of Bilingual Education.
- <http://www.ncela.gwu.edu> – National Clearinghouse for English Language Acquisition.
- www.jan.ucc.nau.edu – Teaching Indigenous Languages.
- <http://www.terralingua.org> – Terralingua.

Further Reading

- Baker, C. (2006). *Foundations of Bilingual Education and Bilingualism*, 4th edn. Clevedon: Multilingual Matters.

Curriculum and Complex Systems Theory

W-M Roth and J Thom, University of Victoria, Victoria, BC, Canada

© 2010 Elsevier Ltd. All rights reserved.

I am inclined to think . . . that unlike the basic question of computability itself, the issues of complexity theory are not quite the central ones in relation to mental phenomena. (Penrose, 1989: 145).

Educators are increasingly engaging in a discourse that includes terms such as chaos, complexity, emergence, and fractals to understand, describe, and provoke teaching and learning situations (e.g., Davis, 2004). These terms are founded in scientific and mathematical fields where modeling phenomena such as weather, the shape of ice crystals, galaxies, or mountain ranges has contributed to the development of, among other things, film animation and video games (Mandelbrot, 1983). However, actual mathematical modeling of curriculum development, curriculum as a process, or curriculum as praxis rarely, if ever, occurs in terms of the complex systems theory, although the mathematics of nonlinear dynamical systems may be the most appropriate tool. (At the time of this writing, the ISI Web of Science database did not list any educational publication under this search term.) There may be two main reasons for this. First, sociocultural phenomena such as real classrooms are so complex that they exceed the capabilities of the mathematical models and computers that generate them. In addition, if classrooms are more complex than any other type of natural system or ecology (Costanza, 2003), then the developmental speeds of sociocultural phenomena exceed those of natural systems by orders of magnitude. Second, to make sense of and to subsequently model even the simplest of complex systems necessitates not only the conceptual mathematical competency of the educator, but also his or her procedural mathematical competency.

Given the enormous complexity of social systems, some may argue that concepts from the complex systems (complexity and chaos) theory be used primarily for analogical purposes, though the apparent simplicity underlying the models of complex systems appear to make it suitable for social systems as well (Guespin-Michel, 2005). However, as research on analogical reasoning shows, the user still has to understand the structures of both the source system and the target systems for any analogy to exist and be productive. For instance, if in the context of a fourth-grade lesson on electricity, the analogy linking a continuous train on the track to the electrons in the wire only proves conceptually effective when all the corresponding structural elements are precisely matched (track/train/people in station pushing/resistance and wire/current/battery/light bulb). Alternatively, one might

assume that the terms from the complex systems theory are to be used for metaphorical purposes, thereby drawing on the power of metaphors in understanding (Lakoff and Johnson, 1980). Even if this is so – and often, this is the case – using such terms for descriptive purposes does not eschew the need for assessing how these metaphors mediate one's conceptions and enactions of curricula. Furthermore, if the meaning of a metaphor is not communicated and comprehended appropriately, it does nothing but hide or obscure, rather than explicate the structural relations among sets of concepts. For instance, second-grade students learning about geometrical objects will find it difficult to understand and even memorize the word cylinder if they do not know that it comes from the ancient Greek term for roller. Having such knowledge would allow them to relate the term to their experience of rolling cylinders down an inclined plane or rolling it between their two hands.

The upshot of these considerations is that the use of concepts from the complex systems theory makes sense only if the speakers and their audiences know the source phenomena (the models related to them) and the target phenomenon (curriculum development and enacted curriculum). If this is not the case, one can use any imagery and does not have to refer to the complex systems theory at all. For example, it is possible to use the imagery of a branching tree and not have to refer to fractal geometry (e.g., Davis and Sumara, 2000) that allows producing a family resemblance between computer-generated images of trees and trees that can be touched and climbed. If we do not understand the underlying mathematical processes at least conceptually, it does not make sense to refer to the complex systems theory, fractals, or chaos, or to choose its associated images as metaphors. Thus, the addition of fractal does not make the metaphor anymore effective.

Modeling the Complexity of Social (Classroom) Interaction

Curriculum as a process (praxis) and curriculum development continue to be thought largely in terms of linear theories, whereby teachers translate a planned (written) curriculum into their classroom actions with the result of achieving specific learning outcomes in students. This is the case although every educator has had experiences where the planned curriculum and the curriculum that is actually lived out or enacted are not the same thing;

where plans or instructions – for example, recipe, assembly instructions, or modes of operation – led to very different sets and trajectories of action. Failing to implement one's own plans for making a dinner, or for operating a new gadget, points to the complexity of life. That said, one can only imagine how much more complex social actions are, such as a teacher bringing forth and carrying off a lesson in the company of a classroom of students. The linearity that underpins many accepted planning models does little to reveal the complexity of nature generally and, specifically, the many orders of greater complexity that is human life. This linearity, as should be clear from this article, underlies the fundamental inappropriateness that teachers experience in university classrooms or the irrelevance of the lesson plans they make to actual classroom events.

Complexity is or at least it should be assumed inherent in the term curriculum itself: one might be talking about a planned curriculum or an enacted one, with its different references to history, syllabi, classroom discourse, intended learning outcomes, and actual classroom events. Questions that arise include: which of these aspects is to be described and represented as a complex system and in terms of associated theories? How do the complex systems models for the different aspects of curriculum relate to one another? In which cases or for which phenomena do metaphors or analogies with complex systems hold true and under which conditions do they break down? In this article, two relatively simple examples – a parent–child exchange and the decision making concerning the (non) inclusion of a concept in a curriculum outline – are used to illustrate both modeling and complex-systems-theoretical thinking about curriculum-related issues. The article concludes with a critical interrogation of the limits and limitations of any analogy or metaphor derived from the complex systems theories.

As a way of introducing the readers to the modeling of complex systems, consider the following exchange that also serves as a model of a simple teacher–student transaction and, therefore, for curriculum as process (praxis):

Example 1

On the way from home to the vacation spot, a child (C) initiates an exchange with her parent (P).

Episode 1

- 1 C: When are we on vacation?
2a P: We are on vacation.

In this situation, we recognize the exchange as a question–answer pair. The parent (patiently) responds that they are on vacation, where the emphatic *are* indicates that the vacation has already started. Naïve analysts might think that the child asked a question. Such an analysis does not allow constructing a model for all possible next instances

that evolve from the child's utterance. That there are other ways of performing a next utterance leading to a different turn pair is evident from the parent's reply later during the 3-day voyage following the same utterance by the child.

Episode 2

- 1 C: When are we on vacation?
2b P: Stop bugging me.

This time, the parent interprets the child's utterance not as a question but as a form of annoyance. Of course, the parent could have responded in the same way as in episode 1, which is what happened throughout the trip between episodes 1 and 2. These instances illustrate that it is impossible to know in advance what the function of an utterance is until it actually occurs. In the first episode, the conversation unfolds as a question–answer pair; in the second instance, it is an annoyance–warning pair. Of course, many other forms of pair completion can be imagined. Any reader not yet convinced of this is encouraged to think of all those situations where an utterance was met with a response like “You hurt me!” and to which the initial speaker replies, “It was a joke” or “I didn't mean to hurt your feelings but . . .” How can this multiplicity of turn pairs emerging from the same utterance be modeled, and how might this diversity be realized in a model that takes into account the complexity of social relations?

A Model for Representing Social Processes

As a first explanation, one may draw on the catastrophe theory (Thom, 1979) as a framework that allows modeling the complexity articulated in the above episodes. The catastrophe theory is both a precursor and a parallel theory to the more popular chaos theory. It combines qualitative and quantitative mathematics to provide an easily intelligible way of modeling the two event categories.

Figure 1 presents a two-state system for completing the child's utterance as raising a genuine question versus aggravating the parent. To the right of point A, the system has but one state where the two turns together constitute

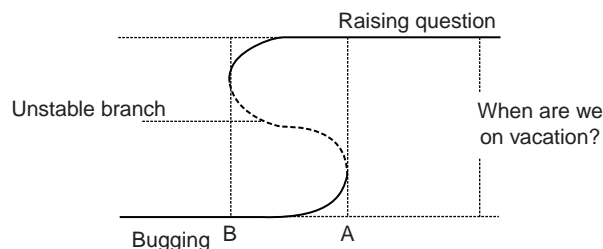


Figure 1 A simple catastrophe theoretic model for the complexity of social life.

one question–answer pair. To the left of point B, annoyance occurs as expressed in the singly possible annoyance–warning pair. Between A and B, two stable states are possible and simultaneous (the dotted part of the curve denotes unstable states of the system): genuine question and bugging. The interesting aspect of this situation is that we cannot know what response will follow the utterance of “when are we on vacation?”; that is, whether we are on the upper or lower branch until the event has taken place. The child’s utterance constitutes a bifurcation point where, for a moment prior to the parent’s reply, the system exists simultaneously on both branches and either response can occur; therefore, it is only when the parent replies that these possibilities collapse into one real event. Moreover, the system may jump from one branch to another in unpredictable ways, brought about by an infinitesimal change in conditions. We can see this as being the case when, at some point along the journey, the system flips and a different outcome results in episode 2 with the parent’s reply; or, equivalently, we can say that the reply flipped the system from one state into another.

What becomes clear when relating this simple example to everyday events is that social life is so complex that we never know whether we are to the left of point B or to the right of point A, though habits may make us think we are in either place. The unpredictability of soci(et)al situations, therefore, suggests that we are always between A and B; however, in relatively predictable (stable) cases, we are closer to A (on the upper branch) or B (on the lower branch). Here, only two actors and two possible states are considered, whereas in real classrooms, there are 20 or more individuals and the possibilities for next turns are large and, likely, infinite. Moreover, because interactions are contingent in the sense that one person never knows what another is up to, real (complex) systems such as classrooms are not only contingent, but also exist in states of instability, which makes them unpredictable even with respect to their immediate future. The best we can do as educators then, is to develop curricula that anticipate what could be as opposed to ones that prescribe what (we believe) is (e.g., Pirie and Thom, 2001).

A Second Model for Representing Social Processes

To model and better understand such complexity, one may use an analogy with another system (e.g., animal population) where complexity arises from recursion – a process of repetitive application to something, for example, a number. The simplest recursive system that leads to rather complex phenomena and can be used to model real events is given by the iterative equation

$$x_{t+1} = k \cdot x_t \cdot (1 - x_t) \quad [1]$$

where the value of some variable x_t at some point (in time) t is fed back into the equation to calculate its value x_{t+1} at the next instant ($t+1$). The end result of this iteration depends on the constant k value. When $k < 3$, x will always asymptotically approach one value (Figure 2). However, at $k = 3$, the system goes through a point of instability: to the left ($k < 3.0$), the animal population always stabilizes at one value, whereas to the right of the point ($k > 3.0$), the population oscillates between two values (i.e., attractors (Briggs and Peat, 1989)). Point $k = 3$ is then considered to be a bifurcation point and infinitesimal perturbations – which in computer simulations arise from rounding errors – are sufficient to change the observable behavior of the system: it is nondeterministic (Prigogine and Stengers, 1979). As k increases, there will be further bifurcations where there are 4, 8, 16, and so forth stable values of x until an infinite number of asymptotic $x_t = \infty$ are possible (Figure 2); here, periodicity ends and the population jumps unpredictably from one value to another. However, subsequently, with further increase of k , a small number of equilibrium states of x emerge from chaos. As a system moves through bifurcation points, it develops its own particular and contingent history, arising from the intertwining of deterministic laws of change (e.g., motion) and indeterminate (chance) fluctuations.

This model (i.e., eqn [1] and Figure 2) directly relates to the conversation between the child and the parent. It is, in fact, equivalent to the catastrophe-theoretical model presented earlier and depicted in Figure 1 for $3 < k < 3.45$, where x jumps between two values. The model allows us to understand the situation right after the child utters his/her sentence and prior to the parent’s response. At this point, the function of the child’s utterance in the collectively constituted conversation is not yet clear, and two possible responses are possible. For any k , therefore, the diagram provides the possibility space; as soon as the next person (parent) acts, the possibilities are reduced

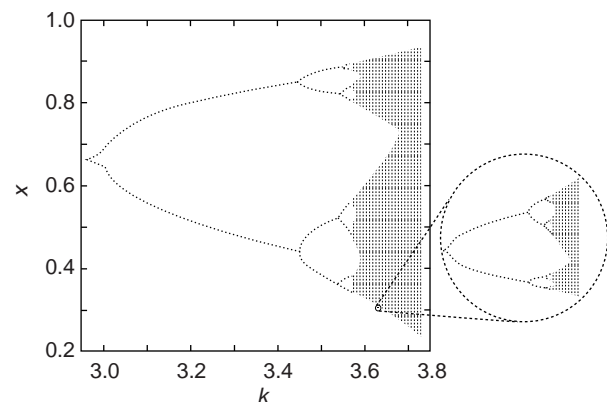


Figure 2 A bifurcation diagram for eqn [1]: depending on the value of k , there are 1, 2, 4, . . . values that $x_{t=\infty}$ takes.

to certainty and the value of x is certain. Readers will immediately realize that any moment in a real classroom situation corresponds to the situations with $k > 3.6$, where there are innumerable possibilities of the next answers setting up new situations and each will have innumerable possibilities of continuation. As k increases, windows (white panels (**Figure 2**)) open up, within which three stable numbers appear, followed by period doubling to 6, 12, 24, and so forth. These period doublings resemble the original image, but now occur at a new scale (insert in **Figure 2**). This is what is denoted by the terms self-similarity and scale invariance. In the classroom, these qualities could be observed where a whole-class discussion about a chaotic event mirrors the discussions that previously occurred in smaller groups where each group came up with its own explication of a phenomenon of interest (e.g., Roth and Duit, 2003). Self-similarity of content is evident across the different discussions and scale invariance exists as a result of the discussions emerging on both whole-class and small-group levels.

It is also important to note that real human life situations do not correspond to stable k values that ecologists, for example, use to model repeating size patterns of animal populations; rather, human life is so complex that it corresponds to (unpredictable) changing values of k , and because of this, it is impossible to be absolutely certain what the effect(s) of our (discursive and practical) actions will be. In more poetic words, Frost (1966) makes the same point in “The road not taken”; there are always more than two roads and we can never know which one we are placing our first step on. Each time, we immediately find ourselves at another bifurcation point with similar complexity. In addition, regardless of the scale (e.g., at a microlevel) at which we examine a human event, for instance, what a student focuses on during a science demonstration, the complexity is so great that we cannot be certain about how many different ways even the simplest of science demonstrations might be perceived by the students (e.g., Roth, 2006).

A Third Model for Representing Social Processes

Such perception phenomena can be modeled using morphodynamical (catastrophe-theoretical) models, which combine qualitative (typological) and quantitative (topological) dimensions. The discontinuities (e.g., in perception, seeing, hearing, and categorization) are the result of complex, continuous, dynamic, and, therefore, inherently unstable processes. Qualitative discontinuities are recovered mathematically as dynamic events of bifurcation (Petitot, 2003), where complex systems radically change their expression, here in a qualitative difference, for example, in the perception of a color, an edge, and so forth. These models therefore

synthesize objectivism and subjectivism into a dialectical identity of nonidentical entities.

Language and speech are other important nonlinear (complex) dynamic systems that are of importance in curriculum theory. What a second-grade student hears during a reading lesson, for example, when her teacher utters what linguists transcribe as “bəa” – bare, to bear, or bear (animal) – is a function of context. Each of the different possible hearing constitutes an attractor (Gaudin, 2005) similar to those that are observable in **Figures 1** and **2**.

A Fourth Model for Representing Social Processes

Educators (teachers) who come into contact with the chaos theory frequently liken curriculum and the process of developing it to strange attractors where the trajectories of learning cannot be predicted but, in the end, captured in the potential of one or the other attractor, only a small number of possible learning outcomes – one in the extreme – are achieved. To illustrate how this model is used as a form of scientific representation, let us examine a high school experiment involving a magnetic bob that moves back and forth across two magnets after it begins at a certain position in space and with some initial velocity (Roth and Duit, 2003). This constitutes a two-state system in motion in which its endpoint is unknown because even infinitesimally minute influences along its trajectory change not only the trajectory, but also the endpoint. Thus, the only certain prediction we can make is that the system will end up near one of the two attractors. What we cannot predict is where the bob ends up – because over many trials and everything else being equal, the probability of it ending up at one or the other magnets is the same, that is, $p = 0.5$.

To study this system’s behavior as it moves from left to right and right to left, changing its speed (i.e., magnitude of velocity) as it approaches and recedes from the magnets, requires scientists and mathematicians to plot its trajectory in phase space – a set of variables that completely describe the system. In this instance, there are two one-dimensional variables, position and velocity, and, together, they create a two-dimensional phase space (**Figure 3**). The figure shows how, beginning with the same position and velocity, the pendulum is caught at one point of time in the left attractor (A_1), whereas it is caught in the other attractor (A_2) at the end of another run.

Although such a model provides an entry point for high school students’ understanding of chaotic systems (e.g., Roth and Duit, 2003), it is a very limited analogy for understanding anything about curriculum planning or the enactment of it.

A Fifth Model for Representing Social Processes

A more realistic system consists of a simple model for how people make decisions, such as choosing one of two options or classifying an object into one of two categories. For example, let us imagine a group of curriculum developers attempting to decide whether to include a particular concept in a curriculum (i.e., phase space). A model that is based on parallel distributed processing, and used for artificial neural networks, allows us to represent such a decision as constraint satisfaction (McClelland and Rumelhart, 1988). Each statement made in the decision-making process describes an aspect of the curriculum as a node (**Figure 4**); three statements support the inclusion of a particular concept under discussion (**Figure 4**, see 1–3) and three statements are made not to include the concept. Some statements support each other explicitly and are modeled as reinforcing links between the nodes (+). Some pairs of statements are contradictory, each supporting a different decision, which is modeled as links between

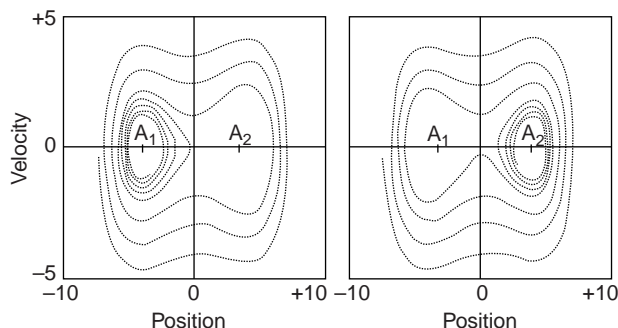


Figure 3 The trajectories of a simple chaotic system in phase space. Although the system begins at the same place and with the same velocity, its trajectory is unpredictable and, yet, the system will end up near one or the other attractor.

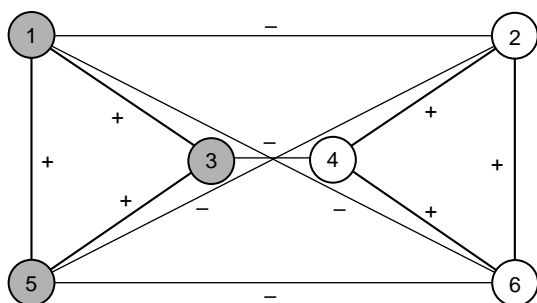


Figure 4 A simple model for decision making (categorization) that relates each statement supporting one of two decisions (gray, white statements). When a statement is supported by another, it is positively linked to it, whereas when another statement contradicts it, it is negatively linked. The strength of support or contradiction corresponds to how emphatic the statement is presented in the meeting.

nodes that weaken each other (–). Some statements do not relate to others and these are characterized in the model by the absence of a link. The final model consists of a network of statements, about a proposal, which mutually reinforce, weaken, or are neutral to one another. The outcome of the process is a solution with the least constraint in the network. Mathematically, the system finds itself somewhere up on a mountainside in a six-dimensional space and is moving (could we say flowing?) down the steepest gradients of the slope until it ends at some lowest place (lowest state). (Mathematically, the problem is the same as in eqn [1], the only difference being that there are now six mutually influencing variables rather than one.)

The dynamics of the system in six dimensions (its phase space) is impossible to envision; however, the (Euclidean) distance that the system has from the two end states (yes and no) can be calculated, yielding a representation of the systems dynamics in two dimensions, which constitute the interpretation or decision-making space (Hutchins, 1995). The left panel of **Figure 5** shows the trajectory of the decision-making process for different scenarios. In scenarios 1–3, the curriculum committee starts mildly favoring a decision to include the concept. However, because of different ways in which the same propositions are affecting others, the curriculum committee quickly comes to a supportive decision (trajectory 1), after initial consideration to the contrary (trajectory 2), or, during the considerations, shifts to support not including the concept (trajectory 3). The two end states – inclusion or noninclusion of the concept – constitute attractors where the system does end up.

There is a third attractor, which does not constitute an absolute minimum (relaxed state of the system) but a local minimum in which the system can be trapped. (Imagine a dip somewhere on our six-dimensional (fractal) surface that can trap the system in its course.) Scenarios 4–6 (**Figure 5**, right panel) model the different situations in which the curriculum committee is hung because both decisions (inclusion and noninclusion) are concurrently being supported and, therefore, a consensus cannot be reached. The different trajectories model the different ways in which the committee reaches the state of indecision. In scenario 6, despite the almost certain starting position of not including the concept, the curriculum committee moves toward a state where both funding and not funding are supported equally well.

Despite the complexity of the possible outcomes that this model can handle, it is still simplistic when we compare it to real-life social-psychological processes. Although it represents the changing strengths of the propositions 1 through 6, it cannot express the relation between them; it models all statements as if they arise simultaneously rather than unfolding over time, continuously modifying the context in which future statements are made; and it only models six statements as opposed to what is more realistic,

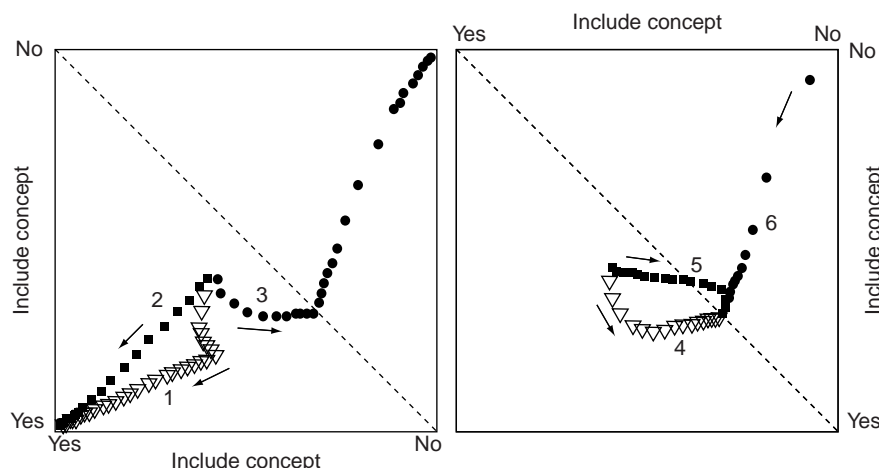


Figure 5 Modeling group decision making when some propositions favor including a concept in the curriculum (1, 3, and 5 (Figure 3)) and others favor not including the concept (2, 4, 6 (Figure 3)). Slight variations in emphasis with which arguments are made lead to the inclusion (1, 2) or noninclusion of the concept under consideration (3). Sometimes, even with extended discussion, the decision-making group ends in a stalemate (4–6).

a potentially infinite number of statements committee members might make in the process. Nevertheless, the strength of this particular model is that it underscores the notion that minor variations in initial conditions can lead to substantially different outcomes – one decision over another. This model also sufficiently captures the complexity of classroom events in that it demonstrates how, once all statements made in some situation are known, the possible ways in which concrete classroom conversations come to a conclusion (Roth, 2001).

Limitations of Analogies and Metaphors in Curriculum Theorizing

In this article, different models and situations are presented to exhibit the suitability of complex systems theories to curriculum theorists. It has been suggested that the mathematical models and graphical representations may be used in metaphorical and analogical ways. However, the effective use of metaphors and analogies requires knowledge not only of source and target domains, but also, and importantly so, the instances in and conditions under which the relationship breaks down. This is also the case for ideas, concepts, and theories from the domain of complex systems. All too often, concepts are taken from the complexity theory, applying them in loose ways that lack rigorous analysis but abound with generic advice (Fenwick, 2004).

From this articulation of different models, it is reasonable to conclude that the complex systems theory for actual physical, sociocultural, and cultural–historical contexts may be very different from the one that appears in the simple models presented here and elsewhere, and which figures prevalently in the educational literature. We do not know whether the phenomena of interest in the curriculum

show scale invariance, such as the one depicted in Figure 2. Do terms such as self-similarity and recursion allow us to expand possibilities for curriculum theorizing or restrict such actions? Current thinking across many disciplines, including landscape and physiological, paleo-, freshwater, and estuarine ecology, population interactions, meteorology and climatology, as well as global change, suggests that this is not precisely the case: to forecast the evolution of the complex system, one needs to take into account the particular scale chosen or involve the required scaling rules that, however, have not yet been developed (Costanza, 2003). The influence of scale, resolution, and hierarchy between different levels has not been investigated to any extent and remains a key question in developing coherent models (theories) of complex systems. Thus, although some curriculum theorists use the notion of scale invariance, even the inventor of the concept of fractal geometry takes a more differentiated view: “no one believes that the world is strictly homogeneous or scaling. . . . One should not be surprised that scaling fractals should be limited to providing first approximations of the natural shapes tackled” (Mandelbrot, 1983: 18 and 19). What, if so, are these limits? To what extent do models for simple constructed phenomena such as question–answer pairs scale up and capture truly complex and real events that arise moment to moment in classrooms, schools, and the field of education at large? In what manners do these theoretical frames allow us to understand events such as test taking, differential achievement, or the similarity in transnational curriculum comparisons?

In mathematics, the complexity theory is concerned with solving classes of problems (algorithms), some of which are vastly more difficult to solve (compute) than others. It is concerned not with the difficulty of solving the problem (algorithm), but with “infinite families of

problems where there would be a general algorithm for finding answers to all the problems of a single family” (Penrose, 1989: 141). In curriculum theorizing, we need a similar field concerned with tracking the solution(s) of infinite classes of problems. Fundamentally, the ideas that appear in much of the literature on complex systems are grounded in an epistemology of the same – recursion produced by some mathematical algorithm, the algorithm itself, self, similarity, self-similarity, and scale invariance are all terms that begin with the ideology of sameness and construct difference as the opposite, not the same, lying outside the boundaries of the same. The idea of the curriculum as a strange attractor, though implying the unpredictability of different trajectories, also implies one or more common endpoints that some enacted curriculum or curriculum committee might arrive at. Precisely here lies a problem that difference, which any real physical and cultural–historical system displays, is not modeled as a positive phenomenon, that is, as a phenomenon in and for itself. It is therefore necessary for the complex systems theories that have any value for describing and explaining real systems of interest to begin differently and recognize that any similarity is the outcome of a (decision making) process rather than *a priori*; sameness and similarity are human productions and not inherent characteristics of nature or culture (Deleuze, 1994). For example, although cooking pots and children’s cups that have a pair of handles appear to us as two different entities that are also substantially distinct on a practical level, from a mathematical perspective, they are considered to be topologically the same; that is, it is possible to morph one into the other without severing or reattaching any of its parts. Consequently, this raises an additional need for educators’ questioning and careful consideration of these forms of mathematics and the models they belong to as suitable structures to constitute the source of analogical thinking about curriculum.

The examples and models presented here – social interaction and decision-making processes – are complex and unpredictable. They serve to highlight just how non-simplistic the real world really is. At the very least, the complex systems theories provide us with ideas that are suggestive of how intractable the world is given that the simplest of phenomena already involve enormous and, perhaps, incalculable complexity.

Bibliography

Briggs, J. and Peat, F. D. (1989). *Turbulent Mirror: An Illustrated Guide to Chaos Theory and the Science of Wholeness*. New York: Harper and Row.

- Costanza, R. (2003). A vision of the future of science: Reintegrating the study of humans and the rest of nature. *Futures* **35**, 651–671.
- Davis, B. (2004). *Inventions of Teaching: A Genealogy*. Mahwah, NJ: Erlbaum.
- Davis, B. and Sumara, D. J. (2000). Curriculum forms: On the assumed shapes of knowing and knowledge. *Journal of Curriculum Studies* **32**, 821–845.
- Deleuze, G. (1994). *Difference and Repetition*, Patton, P. (trans.). New York: Columbia University Press.
- Fenwick, T. (2004). A review of chaos theory and higher education: Leadership, planning and policy. *Complicity: An International Journal of Complexity and Education* **1**, 113–116.
- Frost, R. (1966). The road not taken [1915; 1916]. In Herrig, L., Meller, H., and Sühnel, R. (eds.) *British and American Classical Poems*, p. 112. Braunschweig: Georg Westerman Verlag.
- Gaudin, F. (2005). Y a-t-il de la non-linéarité en sémantique. In Sève, L. (ed.) *Émergence, complexité et dialectique*, pp 280–288. Paris: Odile Jacob.
- Guespin-Michel, J. (2005). Réseau de regulation, boucles de retroaction et dialectique. In Sève, L. (ed.) *Émergence, complexité et dialectique*, pp 243–253. Paris: Odile Jacob.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Lakoff, G. and Johnson, M. (1980). *Metaphors We Live by*. Chicago, IL: University of Chicago.
- Mandelbrot, B. B. (1983). *The Fractal Geometry of Nature*. (updated and augmented edition). New York: W.H. Freeman.
- McClelland, J. L. and Rumelhart, D. E. (1988). *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*. Cambridge, MA: MIT.
- Penrose, R. (1989). *The Emperor’s New Mind*. Oxford: Oxford University Press.
- Petitot, J. (2003). Morphodynamical enaction: The case of color. *Biological Research* **36**, 107–112.
- Pirie, S. E. B. and Thom, J. S. (2001). Thinking through ecological metaphors in mathematics education. In Gunn, S. and Begg, A. (eds.) *Mind, Body, and Society: Emerging Understandings of Knowing and Learning*, pp 45–52. Melbourne, VIC: University of Melbourne.
- Prigogine, I. and Stengers, I. (1979). *La nouvelle alliance: Métamorphose de la science*. Paris: Éditions Gallimard.
- Roth, W. -M. (2001). Designing as distributed process. *Learning and Instruction* **11**, 211–239.
- Roth, W. -M. (2006). *Learning Science: A Singular Plural Perspective*. Rotterdam: Sense Publishers.
- Roth, W. -M. and Duit, R. (2003). Emergence, flexibility, and stabilization of language in a physics classroom. *Journal for Research in Science Teaching* **40**, 869–897.
- Thom, R. (1979). Worüber soll man sich wundern. In Maurin, K., Michalski, K., and Rudolph, E. (eds.) *Offene Systeme II: Logic und Zeit*, pp 41–107. Stuttgart: Klett-Cotta.

Further Reading

- Elman, J. L., Bates, E. A., Johnson, M. H., et al. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
- Emmeche, C. (1994). *The Garden in the Machine: The Emerging Science of Artificial Life*. Princeton, NJ: Princeton University Press.
- Gleick, J. (1988). *Chaos: Making a New Science*. New York: Penguin.
- Hayles, N. K. (ed.) (1991). *Chaos and Order: Complex Dynamics in Literature and Science*. Chicago, IL: University of Chicago Press.
- Raijmakers, M. E. J., van Koten, S., and Molenaar, P. C. M. (1996). On the validity of simulating stage wise development by means of PDP networks: Application of catastrophe analysis and an experimental test of rule-like network performance. *Cognitive Science* **20**, 101–136.

Curriculum and Constructivism

B Davis and D Sumara, University of British Columbia, Vancouver, BC, Canada

© 2010 Elsevier Ltd. All rights reserved.

Perhaps more than any other categories of learning theories used to inform educational practice, constructivist discourses challenge the commonplace separation of what one is taught from how one is taught. Indeed, an oft-cited quality of constructivism-informed practice is that learning is an elaboration of what is already known; hence, the suggestion that content or tasks might be selected independently of learners is regarded as antithetical to a constructivist mindset.

One might say that a constructivism-informed curriculum is more a path laid while walking than a preselected route, echoing the notion of *currence* as developed by Pinar and Grumet (1976). They noted that the root of the word curriculum has to do with the running of the course rather than the “course to be run, or the artifacts employed in the running of the course” (p. 18). As they develop, the ongoing process of meaning making entails a constant recovery and reformulation of one’s embodied history. Curriculum, in this sense, arises in the moment of engagement.

We elaborate this suggestion in this article, starting with a brief overview of some of the more prominent constructivist theories. We then foreground and contrast metaphors for curriculum employed by nonconstructivists and constructivists, highlighting that the title constructivism is often misappropriated in popular conceptions of learning and knowledge. Finally, we highlight some of the more prominent and more thoroughly researched suggestions for formal curricula that have been proposed by constructivist theorists.

Constructivist Theories

Over the past 40 years, there has been a proliferation of constructivist theories of learning that, on the surface, seem to have little in common beyond a rejection of behaviorist, cognitivist, and/or transmission-based emphases in education. The apparent disparities among these theories are rooted in the fact that they are concerned with such varied phenomena as individual understanding, collective process, cultural knowledge, and social justice. When one looks beneath these explicit interests, however, some shared convictions and deep compatibilities emerge, including the following:

- they all highlight the vibrancy of their particular interests, usually by invoking bodily and/or ecosystemic metaphors (e.g., body of knowledge, student body, and body politic);

- they tend to describe systemic dynamics in terms of adaptation and evolution (i.e., notions rooted in biology) rather than cause and effect (i.e., notions drawn from physics);
- they focus on fit (coherence) rather than match (correspondence) in their accounts of how agents or ideas arise and persist, arguing that the criterion for survival is not optimality/efficiency but adequacy/sufficiency (if something works, it will probably endure); and,
- they all embrace the notion of structure determinism, according to which learning is determined by the learner’s own complex history, not by experience or pedagogy (more colloquially, learning is dependent on but not determined by teaching).

More generally, these theories are focused on the manner in which agents in a system must cohere in order for the grander system to remain viable. The nature of an agent varies from one theory to the next, depending on the system under examination. For example, agents might comprise a set of ideas (as in the case of an individual’s understandings), customs and laws (for a society), or species (for an ecosystem).

The major point of departure of these frames from earlier theories of learning is that they are organized around the assumption that what really matters is internal coherence, not external correspondence. For a system to be viable, its parts must hold together. That is, the robustness of an understanding (or whatever the agent might be in the system under examination) is not measured according to how closely it matches, reflects, represents, models, or otherwise corresponds to a reality external to the system. Rather, the measure of an idea is the extent to which it fits with/in the knower’s network of understandings.

In popular parlance, the learners and knowers are most often understood to be individuals, and so it is perhaps unsurprising that the most prominent constructivist theories deal with the question of how individuals make sense of the world. Sometimes qualified with adjectives such as radical or subject-centered, these constructivisms are typically traced to the work of French biopsychologist, Jean Piaget. His project might be described as an effort to construe personal learning through the metaphor of emergent biological forms. He suggested that the development of understandings is analogous to the growth of living systems, and so his theory of personal learning foregrounds principles of contingency, adequacy, adaptation, and fitness. In brief, Piaget (1955) argued that

learning is a continuous process of updating one's sense of the world as prompted by new experiences. The learner is constantly construing and reconstruing in an effort to maintain a coherent system of interpretation.

In other words, for subject-centered constructivists, the biological body is not a structure through which one learns, but a structure that learns. Individual learning is not a brain-based phenomenon, but an ongoing process of embodying one's history. For this reason, bodily action is not seen as a demonstration of internalized understandings; rather bodily action is understanding.

Another strand of constructivist discourse is known as social constructivisms – a term that comprises constructionism, sociocultural theory, activity theory, situated learning, communities of practice, and actor-network theory (among others). The common interests across these frames are interpersonal dynamics and collective activity. They generally address the sorts of phenomena that subject-centered constructivisms set aside as context or circumstance, including language, artifacts, social status, cultural background, and disciplinary knowledge.

The major point of departure from these more collectivist theories is an extension of the notion of a cognitive unity. Social constructivists do not regard the individual as the locus of learning but as a learning system within a grander learning system. For the social constructivist, human cognition is diffuse, distributed, and collective. In this frame, mind is understood not as an individual possession but as a product of shared human interest that arises in an environment that is both social and physical. Much less preoccupied with individual sense-making, social constructivist theories focus on conversation patterns, relational dynamics, social habits, prevailing technologies, and other aspects of collective possibility. In this frame, cognition is always collective: embedded in, enabled by, and constrained by the social phenomenon of language; caught up in layers of history and tradition; confined by well-established boundaries of acceptability; defined by joint interests, shared assumptions, and common sense.

The most commonly cited influence across social constructivist theories is Russian sociopsychologist Vygotsky (1986), whose core metaphor (reflecting the Soviet context of his research) was shared labor rather than biological coherence. This emphasis is preserved in some branches of discussion, particularly those that are organized around notions of apprenticeship and communities of practice. However, biology-rooted notions of bodies, coherence, and fitness have also come to be pervasive across social constructivist theories.

For the most part, discussions of constructivism within education are bounded by the bookends of the individual's construal of the world and the self-maintenance of the social collective – which is perhaps to be expected. Through most of the twentieth century, the bulk of educational research was focused on individual learning and social contexts and

was oriented by theories drawn mainly from psychology and sociology. However, through the 1960s and 1970s, as researchers with backgrounds in anthropology and philosophy entered the field in larger numbers, increased attention began to be given to issues such as cultural context and the social implications of schooling. One might say that the conversation was broadened to include issues of the body politic. Furthermore, while this part of the conversation is not often associated with either subject-centered or social constructivist theories, it shares the tendency to invoke body-based metaphors and to describe events in terms of evolutionary dynamics. A major difference, however, is that cultural and critical theories tend to delve into issues of morality and ethics – seeking, for example, to uncover the ways that schooling contributes to (or, more hopefully, might interrupt) social stratification, gender roles, and other cultural habits of differentiation. In other words, cultural and critical constructivist theories are principally concerned with deeply entrenched habits of interpretation and implicit associations that support social constructions of gender, race, class, sexuality, ability, disability, opportunity, and so on. A parallel focus has to do with the manners in which certain domains, such as mathematics and science, have been given privileged voices in the contemporary academic world. One might say that cultural and critical constructivist discourses attempt to shift the topic of conversation from the individual's efforts to shape an understanding of the world to the manners in which the cultural world shapes the understandings of the individual.

Metaphors and Images of Curriculum

Over the history of the modern school, different images have been employed to describe curriculum. Most commonly, curriculum has been characterized in terms of progress along a linear path (always forward, usually upward) toward a prespecified, fully articulated goal.

The notion of the spiral curriculum rose to prominence, first proposed by Bruner (1960) (who is often identified alongside Piaget and Vygotsky as one of the principal authors of constructivist theory). The idea was that, rather than proceeding through topics sequentially, students should be introduced to concepts gradually, revisiting topics each year to elaborate understandings. The spiral image was intended to embody a sense of what might be referred to as dynamic sufficiency – a notion offered in opposition to assumptions of the static deficiency of learners. Rather than positioning learners as essentially fixed beings with inadequate understandings that must be completed by schooling, the spiral curriculum was prompted by the realization that human knowers can never be finished, yet are always strangely sufficient. The process of formal education, then, is not to complete

learners, but to present opportunities to expand horizons of possibility.

This point is perhaps better expressed in terms of underlying visual metaphors. In both the classical (sequential) and the more recent spiral models of curriculum, the images are line based and oriented by a desire to converge onto pre-specified objectives. Constructivist interpretations inject senses of contingency and emergent possibility, and thus lean more toward images that involve branches ramifying into new branches. Typical images include fractal trees (see **Figure 1**) and waterfalls, the latter of which highlights the importance of context in defining emergent possibility. In essence, learning and the production of knowledge are understood in terms of opening up spaces of possibility by exploring the spaces of the existing possible. These recursively elaborative processes do not completely reject notions of directionality and preselection of learning outcomes; however, they do reject the assumptions that paths of learning can be predetermined and that there are best methods for structuring curricula.

This point – that is, that there are no best, merely good-enough approaches to education – has been identified as one of the most troubling aspects of constructivist theories. In effect, these theories do not indicate what should, must, or can be done to affect learning systems; rather, they are more focused on what cannot be done or should not be attempted. In particular, they emphasize that teaching should not be construed as causing learning and that curriculum should not be construed as any sort of deliberate constructive process. This latter suggestion has the ring of a paradox: How can it be that constructivist theories reject curricula that are understood in terms of processes of construction?

The underlying issue here has to do with a problem of translation and multiple meanings. In English, the word construction most often calls to mind notions of foundations, platforms, rigid frameworks, hierarchies, planning, and so on – in brief, precisely the sorts of images that are associated with linear and spiral curricula. In French, the language of Piaget's theorizing (and, more significantly, the language of the structuralist philosophy that influenced his work; see Davis and Sumara (2002)), construction has two distinct meanings, and the more common of which is a biological sense of ever-evolving possibility that is both reasonably stable and subject to sudden

change. This meaning is better preserved in the English cognates *strew* and *construe* and is more readily associated with the notion of an emergent curriculum.

This issue is of vital importance in distinguishing current constructivist theories from trivial and naive constructivisms that are often encountered in contemporary discussions of educational practice. Mantras, such as everyone constructs their own understandings and learning is based on experience, are commonly deployed as indicators of constructivist sensibilities, yet these notions have actually been part of the educational literature for centuries. Locke (1960/1976), for example, asserted “No man's knowledge can go beyond his experience” (bk. 2, ch. 1). An upshot is that a constructivism-informed curriculum is not a straightforward matter of a rich learning environment, an attentive teacher, and open-ended outcomes.

Curriculum Informed by Constructivist Theories

It would be impossible and irresponsible to suggest that there is any sort of consensus around what might constitute a constructivism-informed curriculum. However, there are a number of commonly mentioned elements. Several of these are discussed below. In order to contextualize the discussion, and to be attentive to the core principles of situatedness and embodiment of learning, we illustrate the discussions through a specific example – namely, the topic of day and night, explained in terms of the Earth's relationship to the Sun.

Trans-Level Learning

Considered as a cluster of overlapping and intersecting frames, the family of subject-centered, social, cultural, and critical constructivisms call for attentiveness to the manners in which individual, social, and cultural forms unfold from and are enfolded in one another. Rather than getting trapped in dualistic arguments in which, for example, the interests of the individual are pitted against the interest of society, these frames are oriented by a more all-at-once mindset and thus gesture toward a curriculum that might be described as transphenomenal. A guiding question is

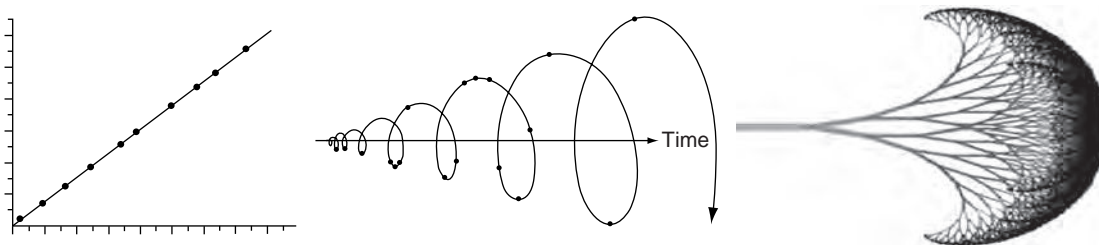


Figure 1 Common visual metaphors for curriculum: sequential, spiral, and emergent.

how the topic at hand reaches across various levels of learning/adaptation.

For example, working from the premise that personal knowledge is anchored to experience in the world, a subject-centered constructivist frame might prompt attention toward the various ways that individuals might interpret or explain day and night. Most popularly, as evidenced by phrase such as the rising sun, daybreak is described in terms of the sun transiting the Earth's sky (rather than the Earth spinning out of its own shadow). A curriculum that is sensitive to such interpretive tendencies might include suggestions for activities through which individuals physically engage with different models (e.g., involving flashlights, balls, and elliptical paths). Such activities would be selected and designed not only to address the topic of day and night, but also in a manner that sets up related topics such as phases of the Moon, the Sun-centered solar system, and seasons – in brief, aimed at the articulation of a coherent set of principles that provide the learner with some significant conceptual reach.

On the more collective level, prompted by social constructivist theories, a curriculum might incorporate topics around the production of scientific truth. The contested nature of early heliocentric models provides a good opportunity to address issues around the ways scientific theories are developed, presented, defended, accepted, or rejected. Conversely, the classroom collective might be organized as a community of inquiry in an attempt to reflect the exploratory and collaborative work of scientists. Whatever the case, the orienting principle is that science (or whatever domain) should not be reduced to a collection of pre-given facts to be mastered. Rather, it should be represented and experienced as a specific and very social orientation toward addressing questions of shared interest.

As for cultural and critical constructivist theories, the historical transition from a geocentric to a heliocentric model marked a dramatic shift in Western sensibilities. It occurred at the dawn of the scientific era and was accompanied by the rises of capitalism, industrialization, urbanization, and European imperialism – and these movements were likely influenced by the displacement of the Earth from the center of creation to the status of minor planet orbiting a typical star. Cultural and critical constructivist theories might highlight the need for curriculum topics that include interrogations of a male-dominated, Eurocentric science on religion, global politics, and cultural genocides. In brief, explanations of such simple phenomena as day and night are neither innocent, nor isolated.

Cognitive Conflict

Further to the matter of the need to provide individual learners with opportunities to engage physically with objects in their worlds, since the early 1980s there has been a pronounced emphasis among subject-centered

constructivists on cognitive conflict. Underscoring that curriculum should address students' suppositions, the suggestion that experiences should be organized in ways that compel learners to be attentive to (and, if necessary, to reconfigure) understandings. This emphasis is framed by the principle that learners will hang onto an interpretation, no matter how limiting it may be, until the effort required to maintain it exceeds the effort required to revise it.

In the 1980s, this principle was most often discussed in terms of learner misconceptions, which by the early 1990s was redubbed alternative conceptions (reflecting the conviction that knowledge is about internal coherence, not external correspondence). The advice is to develop sets of experiences that lead to surprising results, so that learners question their frames of interpretation and (hopefully) revise them. This emphasis has been described in various ways, including the introduction of cognitive obstacles, structuring of discrepant events, and deliberate attempts to prompt disequilibrium. By way of specific example, the sorts of activity suggested in the previous section (e.g., involving flashlights and spheres) might be structured in ways that anticipate and challenge common, but limiting interpretations (e.g., introducing the question of phases of the Moon alongside day and night to highlight the conceptual contortions needed to maintain a geocentric model).

A problem here is that learners often generate auxiliary explanations and strategies (e.g., "In this case, use this rule") rather than revising understandings. In more general terms, the strategy of introducing discrepant events has been tied to the topic of transference of understandings across contexts and experiences. Seeking to address the commonly noted problem of isolated learnings, curriculum developers are also urged to consider the ways concepts might be taken up across disciplinary realms and within the variety of contexts in which learners find themselves. By consequence, curriculum should not be conceived in terms of a collection of parallel strands, but as a set of interlacing discourses with multiple opportunities to address and elaborate topics introduced in other contexts.

Rethinking Errors and Evaluation

Linking two points mentioned in the preceding sections, constructivist theorists call for a reconsideration of the notion of error. Commonly, errors are understood as mismatches between internal and external worlds. For constructivists, all actions are rooted in a coherent interpretive system – and so an interpretation that works well for the individual (such as a belief that the Sun goes around the Earth) cannot be construed as a mistake on the level of the individual.

However, on the level of social action, it can and must be understood as an error. Otherwise, there is a risk of

interrupting or disabling collective action. Note that the point here is not that education must seek to ensure everyone holds the same set of ideas. On the contrary, as noted in the section titled “Collective learning,” tension among views held by individuals in a knowledge-producing system is important for its survival. However, while some idiosyncratic beliefs are valuable and others can be ignored, there are some that prevent collective action. When that happens, they are justly treated as errors.

Further as suggested in section titled “Cognitive conflict,” the teacher’s responsibility here is not to fix a student’s error, but to make sense of the web of associations that render the interpretation a sensible one to the student. Working from an appreciation that the student’s action was part of a coherent worldview, the teacher would attempt to involve the learner in new sets of experiences that might support the construal of more appropriate interpretations.

By implication, methods and structures of evaluation must be attentive to the particular contexts of student action. Thus, in constructivist terms, contemporary approaches to standardized examinations are rife with problems. Not only do they ignore the necessarily experiential and contextual dimensions of understandings, but they also introduce a new level of organization in which new categories of error might arise (i.e., actions that might not be treated as errors on the individual or classroom levels could potentially emerge as errors on the level of a standardized examination). Further, they promote a curriculum organized around mastery rather than engagement. Significantly, however, the point is not that standardized examinations are wrong. Rather, the issue is that they may not (and perhaps cannot) test what they are assumed to test, since they are necessarily inattentive to intermediate levels of educational organization and activity. In turn, this issue raises questions around what it might mean to standardize a curriculum.

Collective Learning

With the interest in collective knowledge production across social, cultural, and critical constructivist theories, it is not surprising that there tends (as in the point developed in the section titled “Trans-level learning”) to be a strong emphasis on incorporating topics such as standards of verification, political dimensions of disciplinary realms, and unanticipated consequences of privileged discourses. In recent years, an added dimension of collectivity has risen to some prominence, and it is an issue to highlight inseparability of pedagogy and curriculum.

Oriented by the conviction that the word learner can refer to any context-sensitive, adaptive, and self-maintaining phenomenon, one might argue that classroom groupings might properly be construed as learners – that is, not as collections of learners, but as collective learners – capable of producing and acting on their own knowledge.

For the most part, traditional classroom structures and emphases do not support the emergence of a knowledge-producing collective in any significant sense, in large part owing to the structures of activities that organize their engagements (cf. Surowiecki 2005). In particular, and in contrast to productive knowledge-producing communities, there is often an overemphasis on redundancy (e.g., preset objectives, determined independent of learners, based principally on age) and an underemphasis on diversity (e.g., opportunities to pursue idiosyncratic interest or to bring specialized knowledge to bear on a real issue). To address these matters, a variety of researchers and theorists from across subject areas have called for curricula that, simultaneously, enable the development of necessary redundancies among learners and provide opportunity for the articulation and elaboration of diverse interests – all within the context of shared projects.

In this frame, redundancy is understood to refer to those competencies that enable individuals to work together. These elements might be thought of as the basics – where one is mindful that this phrase only makes sense when considered in relation to era and location. Redundancy contributes to the robustness of a system as it facilitates communication, frames interests, and enables individuals to fill in for one another. The notion encompasses not only matters of conceptual knowledge, but also problem solving, communication, and other social competencies.

Its complement, diversity, might be interpreted as the basis of the collective’s intelligence. Varied interests and expertise provide the pools of possibility that a learner might draw on in order to address more complex problems. To this end, commonly identified aspects of curriculum tasks include variable entry points (i.e., that allow individuals to negotiate the difficulty to suit their interests and levels of conceptual development), authenticity (usually taken to mean relevant, or able to be made relevant), and possibilities for digression (i.e., opportunities to pose new problems).

Of course, such structures present an issue that represents a challenge to traditional curriculum structures: There exists the possibility of generating knowledge and pursuing topics that are not explicitly sanctioned by the official program of studies. As such, there is a concomitant need for a sort of decentralized control, not only in the classroom, but also at jurisdictional and ministerial levels.

Participation

In a strong sense, the notion of a participatory curriculum might be used to collect together the above points. Constructivist theories are, in effect, participatory epistemologies. They understand knowledge to inhere in interactions – that is, to be embodied or enacted in the ever-unfolding choreography of action within particular contexts at particular times. This orientation compels a certain mindful attendance not only to what one knows, but also to the

consequences of one's knowing. Education in this frame is more about becoming aware of emergent possibility and corresponding responsibility than about acquisition or mastery. From this it follows that a curriculum must not only allow for participation, but also be organized around participation.

As for the future of constructivist theories in curriculum development, there would seem to be both significant challenge to overcome and significant promise to be realized. On the side of challenge, it often seems that more effort is given to identify differences among particular constructivisms than to examine their deep complementarities. By consequence, the range of contemporary constructivist theories might be interpreted as too dispersed and conflicted to be of much practical use. On the side of potential contribution, most constructivist theories are explicit in their embrace of evolutionary dynamics and self-organizing structures – emphases that align these discourses with transdisciplinary discussions of complexity and ecology. They thus offer advice not only for the content of curriculum, but also for its enactment, its ongoing development, and its place in the grander contexts of cultural and biological worlds.

Bibliography

- Bruner, J. (1960). *The Process of Education*. Cambridge, MA: Harvard University Press.
- Davis, B. and Sumara, D. (2002). Constructivist discourses and the field of education: Problems and possibilities. *Educational Theory* 52, 409–428.
- Locke, J. (1960/1976). *Essay Concerning Human Understanding*. London: Dent.
- Piaget, J. (1955). *The Construction of Reality in the Child*. Cook, M. (trans.) London: Routledge and Kegan Paul.
- Pinar, W. F. and Grumet, M. R. (1976). *Toward a Poor Curriculum*. Dubuque, IA: Kendall/Hunt.
- Surowiecki, J. (2005). *The Wisdom of Crowds*. New York: Random House.
- Vygotsky, L. S. (1986). In Kozulin, A. (ed.) *Thought and Language*, rev. edn. Cambridge, MA: MIT Press.
- Cobb, P. (1994). Constructivism in mathematics and science education. *Educational Researcher* 23(7), 4–12.
- Davis, B., Sumara, D., and Luce-Kapler, R. (2008). *Engaging Minds: Changing Teaching in Complex Times*, 2nd edn. New York: Routledge.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Phillips, D. C. (1995). The good, the bad, and the ugly: The many faces of constructivism. *Educational Researcher* 24(7), 17–40.
- Spivey, N. (1997). *The Constructivist Metaphor: Reading, Writing, and the Making of Meaning*. San Diego, CA: Academic Press.
- Varela, F., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.
- von Glasersfeld, E. (1995). *Radical Constructivism: A Way of Knowing and Doing*. London: Falmer.
- Watzlawick, P. (ed.) (1984). *The Invented Reality*. New York: W.W. Norton.

Further Reading

Curriculum and Critical Theory

L Yates, University of Melbourne, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Capitalism – An economic and political system based on market competition, profit, and individualism.

Hegemony/hegemonic – The ruling ideas and social interests within a country.

Ideology/ideological – Beliefs and ideas that distort the truth, but represent the interests of a particular group.

Modernism – The belief that society and social institutions can be improved by the application of rationality.

Political economy – A focus on how institutions are shaped by capitalist economic interests and arrangements.

Postcolonial theories – The theories concerned with the effects of colonization or domination within a country on identities.

Reproduction theory – A theory that schooling and curriculum reflect the social structure and produce outcomes that continue that structure in its existing form.

How a society selects, classifies, distributes, transmits and evaluates the educational knowledge it considers to be public reflects both the distribution of power and the principles of social control. (Bernstein, 1971: 47)

Bernstein's much-quoted aphorism is a classic statement of a critical theory perspective on curriculum. In its historical and narrow sense, critical theory is the term given to a body of mid-twentieth-century, neo-Marxist European social theory (the Frankfurt School), whose writings analyze culture and hegemonic power. Education theorists drawing on this tradition were interested in the ways that curriculum functions as an ideological apparatus of the state by serving and maintaining the power of the dominant social groups and contributing to the subordination of others. Critical theorists analyze curriculum as an institutional set of arrangements that includes policies, contexts, identities, and pedagogy as well as explicit curriculum documents. They see curriculum as a social construction that works to produce ways of thinking and distributions of opportunity in which some social groups are powerful and others marginalized. They attempt to work toward new approaches and agendas for the

curriculum that might produce fairer and less subordinating social outcomes.

Over time, the range of activities and theories that might be considered to constitute critical theory has expanded, particularly in the prominence of feminist and critical race perspectives in the work, and in the taking up of some aspects of poststructural theory into the critical agendas. However, broadly, what distinguishes curriculum work bearing a critical tag is its ongoing concern with the relation between curriculum and broader social forms and its attempt to challenge social inequalities and differences and state hegemony as these function in and through curriculum.

Sociology of Knowledge and Curriculum as Ideological

In the critical theorizing of curriculum that flourished from the 1970s through the mid-1980s, two key ideas were prominent: that sociologists should investigate curriculum, and not simply school inputs and outputs, in order to understand how success and failure are distributed in schools; and an attack on the view that curriculum knowledge could be neutral. Curriculum was seen as serving the interests of particular social groups, and, more broadly, as serving an ideology of capitalism that represents different economic rewards as simply the outcome of individual effort and choice.

In the UK, the most influential text of this period was *Knowledge and Control: New Directions for the Sociology of Education* (Young, 1971). This signaled the beginning of a new prominence of specifically sociological perspectives on knowledge, and a fierce debate with philosophers of curriculum and with traditions of theorizing curriculum in terms of questions about the educated man, or the disciplinary forms of knowledge, or in terms of analytical distinctions about education versus indoctrination or socialization. It began an era of considering curriculum questions in terms of whose interests are being served by particular selections and forms of knowledge and assessment.

In the USA, influential works in the 1970s took a political economy approach: that is, they were concerned with economic determinants of education, curriculum as ideology, and education as a means of producing docile workers for the capitalist system. Bowles and Gintis (1976) argued that the hidden curriculum of schooling systems was designed to discipline populations for docile work in factories.

They were critical of the way differential school outcomes were attributed simply to internal factors of teaching and student effort, and, in a footnote, suggested that a more appropriate way to consider the achievement of schools would be to introduce a value-added measure that took account of their socially differentiated inputs – an idea that is now widely used in conservative education policy in many countries. Apple's (1979) work took up the economic interests of textbook publishers, and the ways these served the interests of capitalism by regulating what counts as curriculum knowledge in the USA.

Similar currents of work focusing on curriculum as a mechanism of social reproduction were being developed in many national settings in this period, including Sweden, France, Canada, and Australia. In Australia, for example, a major empirical study of working-class and ruling-class schools (Connell, *et al.*, 1982) identified the competitive academic curriculum as the lynchpin of the reproduction of class inequalities in education, and called for working-class schools to work with their communities to build different forms of knowledge and outcomes for them.

The critical theory approach is particularly associated with sociological work on curriculum; however, a few philosophers also took up a Marxist perspective on school knowledge, and from the 1980s and beyond, feminist philosophers played an important part in fundamentally critiquing perspectives of knowledge as objective or neutral. In the USA, some critical theorists would see a legacy from Dewey in their work, and also in the work of John Rawls. The use of the label critical theory is usually restricted to work that assumes an explicit critique of the current form of society as a foundation from which the analysis of curriculum begins.

In the 1970s, many radical books taking up particular elements of this broad critical approach to curriculum were widely circulating in general bookstores. Paulo Freire contrasted banking knowledge in which curriculum is conceived as a set of deposits to be transmitted to the learner, and which would inevitably reproduce power and advantage in its present form, with dialogical knowledge in which the teacher would work with the students to find key concepts for them, and begin to build these into more powerful (socially critical) ways of understanding their situation and the way to change it. Freire was later Minister of Education in Sao Paulo and his approach has been widely influential internationally, especially in literacy teaching.

Curriculum Theories: Bernstein and Bourdieu

Two important new theories of curriculum, by Bernstein in the UK and Bourdieu in France, were first developed in the 1970s and have had an enduring impact. Each has been widely used and elaborated by curriculum researchers in

many countries. Both bodies of work were built on a range of empirical studies of classrooms and school systems; however, both also developed more abstract formulations of curriculum and its relation to social forms and social inequalities, and the causal mechanisms that govern this relationship.

Basil Bernstein's work was interested in the relationship between social class-differentiated modes of language, identity and home socialization, experiences of curriculum in school, and forms of social structure being produced from school. In a series of important works, he used the concept of codes to develop a structural account of knowledge making and the ways of knowing developed by different social groups and curriculum arrangements. Bernstein's initial line of work focused on how the forms of language developed in different social class settings (restricted and elaborated codes) generated different forms of relationship and power in the classroom. Bernstein then turned his attention to the forms of school and university curriculum in different countries, and to the debates between advocates of integrated (or progressive) subject arrangements, and more traditional formal subject-based curriculum, which he named a collection code. By tracing what was open and what was closed in the different arrangements, Bernstein produced a highly generative schema for analyzing relationships and power of the teacher and student in different arrangements, effects of class difference, types of identities produced, and emerging new social forms (e.g., the knowledge and identity needs of the new (professional) middle class).

Pierre Bourdieu's work first became known to English-speaking readers primarily as a highly schematic account of schooling as a mode of social reproduction (Bourdieu and Passeron, 1977), and subsequently as a series of detailed investigations of the socially differentiated working of cultures and education institutions. His concepts of cultural capital, habitus, and intellectual field have been widely drawn on. Some key ideas in Bourdieu's work that have been important for curriculum are that family and community habitus (cultural, economic, and interpersonal environment) produces ways of learning to be in the world that are not only about language or what one knows intellectually, but also about physical and emotional presence and orientations to what one notices in the world (dispositions). For Bourdieu, the elite curriculum reproduces social elites because, through the curriculum and examination system, it asks of other groups what they cannot give, that is, it asks groups who are not of the elite to have already formed the tacit tastes and educational orientations of the elite.

Feminist and Critical Race Theories

From the 1970s on, feminist and critical race theories have been a core presence in critical curriculum theory, and

critical theory approaches have also been taken up by writers and activists concerned with disability and sexuality. In one sense, the work here expands and strengthens the critique of curriculum as representing the dominant interests of a capitalist/patriarchal/racist society. However, the new voices are each associated with social movements that do not simply emerge out of critical theory perspectives on capitalism and class that have been described to this point, though they do draw on these. They bring additional conceptual agendas and begin to challenge and reshape what is taken up as a critical theory perspective on curriculum.

In the most obvious sense, feminist theories expand the attention to a class-structured society by seeing society as structured by gender, and curriculum as a vehicle for transmitting inequalities and subordination related to gender; and race theories see society as structured by race, and curriculum as a vehicle for the reproduction of race and ethnic ideologies and outcomes. Both moves have been important in curriculum in mounting a critique of sexist and racist language and images in textbooks. At first, these concerns were often dismissed as a radical approach to curriculum; however, now, a modest form of that interest in representation is found in mainstream curriculum development and in instructions for textbook publishing, including in subjects such as math and science.

In practical terms, these movements expanded attention not just to what was in the curriculum (ideology of the mainstream or dominant class), but what was not in it as well. A feminist critical perspective was not just about adding women in to a curriculum story or an array of subjects that remained unchanged. It raised questions about what counts as public knowledge or history or science or academic achievement, and about the emphasis on a particular concept of rationality and individualism as the heart of education (Martin, 1994; Walkerdine, 1988). McCarthy and Crichlow (1993), Ladson-Billings and Tate (1995), and others developed similar challenges from the perspective of race issues.

Critical theory is overtly political in its orientation. Its concern is with processes and mechanisms that work through institutions, such as schooling, to covertly produce power and advantage, and discrimination and subordination for different social groups, and that work against democratic ideals. However, the development of feminist and critical race approaches has altered and challenged the way in which it is possible to think about the critical theory. They raise new ideas about what power or subordination looks like, and about the mechanisms or processes by which such social discriminations are reproduced.

Fraser (1997), for example, has named two different trajectories for understanding the politics of social justice: distribution and recognition. Distribution refers to the ways different groups gain or are excluded from economic goods in society. This is a formulation that has been

particularly associated with attention to class inequalities and poverty (though distribution concerns about curriculum are also part of the agenda for gender and race), and with the way school curriculum and assessment set up the conditions for success and failure of different social groups. Recognition is an issue that Fraser associates particularly with feminist and postcolonial movements, and it might be translated as the ways in which groups are, or are not, represented and acknowledged as part of the social polity, or as equally human. These are themes that have been particularly important in seeing school curriculum (or in Apple's phrase, official knowledge) as sexist or racist. Recognition (creating some groups as other, as lesser) is seen as an important discriminatory effect of curriculum, regardless of whether it has measurable outcomes in terms of immediate school success and failure.

Critical Theories and Poststructural Theories

In moving into these broader conceptualizations, the distinction between critical and poststructural theories of curriculum is not a sharp one, and the precise categorization a matter for debate (as is the question of whether feminist and postcolonial theories of curriculum are simply subcategories of other theories, or more appropriately seen as curriculum perspectives in their own right).

Traditionally, critical theory has been seen as a modernist form of theoretical and political activity. That is, it has been seen as an approach that is avowedly normative, one that believes truth (or at least more powerful or less-distorted understandings) can be obtained, and that wants to enhance the role of schooling as a form of social improvement. Poststructural theories (and postmodernism) are associated with a view that there is no single objective truth, that claims to truth are claims to power, and that the role of theoretical work is to unpack how truths and identities come to be constituted. In relation to curriculum, the critical theory has historically been more strongly associated with theories about the curriculum's distributional processes, and poststructural accounts with what kinds of identity formation and ways of being in the world are being produced by curriculum. However, especially with theories associated with feminist work, and with a widespread move to use the concept of discourse rather than ideology, the dividing line between these categorizations is not clear-cut.

One of the ways we might consider the distinction between these approaches for curriculum is to consider their form as theories. The early critical theory writing was clearly neo-Marxist in its foundations. That is, at least in its starting points, the work uses a causal form of explanatory theory grounded in showing a regularity of relationship between material practices (work and

family life) and ideas or ideology. The aim is to begin to reveal mechanisms for misrecognition and reproduction of social inequalities embedded in curriculum, and to find ways for nondistorted understandings of real social and individual interests to flourish. This is a political project as well as a set of theoretical perspectives, and one of its best-known forms in practice as well as theory, is the work of Paulo Freire.

With feminist work and work grounded in concerns about race, there is often less faith in a single explanatory theory as the potential solution to understanding curriculum effects, and less belief that a sphere of objective knowledge or undistorted communication can be attained once the distorting ideology is revealed. Rather, more diverse empirical and theoretical works are seen as needed to describe what is happening, and to take account of different conceptions of oppression. Much of the work here is critical of the single, big theory explanations and of neglect of the person, gender, and subjectivity in some other strands of critical theory work (Ellsworth, 1989; Luke and Gore, 1993).

Foucault's work is usually associated with a poststructural perspective on education, one that is interested in how people come to discipline themselves to serve the interests of a self-regulating polity, and that replaces the idea of a stable human subject who acts on the world with the concept of a subject positioned in discourse. However, much work that has been influenced by Foucault might also be seen as an extension of a critical theory approach, in that it develops from an interest in curriculum as a form of governance of the population that has particular disciplinary effects, both generally (e.g., in relation to neoliberalism) and for particular groups such as working-class children, or women. Feminist theorists of curriculum have had a particular interest in how consciousness and identity or subjectivity is constituted since the idea of the personal being the political has been a key conceptual agenda in feminist work.

The developments here are drawing poststructural perspectives into critical theory agendas in two ways: by developing a critical account of contemporary social forms and the role of curriculum as a mechanism of governing and disciplining identity; and by expanding attention to the interplay of desires and socially formed subjectivities with forms of curriculum. Subjectivity was traditionally a more central interest of feminist work; however, it is notable that many critical theorists, including Bernstein, Wexler, and others have also, in recent decades, focused more strongly on identities, and the role of curriculum in constituting them.

Power, Agency, and Action

One of the matters that has been of central concern for critical theory approaches to curriculum is how to

understand power and the scope for action by human subjects. The early phase of work that was labeled reproduction theory was subsequently much criticized for its overdeterminist picture and its apparent portrayal of human subjects as mere ciphers who reflect those mechanisms of schooling, and this led to some focus on resistance (i.e., studies of students from subordinated groups and the logic inherent in their refusal of the official knowledge schooling attempts to convey to them) and on collaborative action projects to produce change.

Critical action research has always been an important line of work within the critical theory (Carr and Kemmis, 1986). This work takes the position that moving to a more democratic or less ideological curriculum is not something to be worked out in the abstract by academics or bureaucrats, but requires collaborative work with teachers and students, and an ongoing cycle of identifying issues and working on and reviewing these. Other work (Giroux and McLaren, 1986) uses the term critical pedagogy and draws feminist and postcolonial agendas into its approach to change. In this approach, there is interest in the embodied person and in subjectivity as part of the terrain for critique and action.

As an avowedly political approach to curriculum, an intention of critical theorists has been not just to describe the world, but to change it as well. In practice, theorists who might come under this label take a wide range of positions in their work, with some operating almost entirely in terms of producing critical readings of curriculum, others working to develop new approaches to curriculum policy and teacher education, and yet others developing collaborative critical action research and critical pedagogy. A belief that the sources of curriculum and curriculum effects lie beyond the intention of individual subjects remains important in the work, but so also does attention to the subjective experiences and actions of groups who have been marginalized. In practice then, many different kinds of approaches to curriculum research may be found within the broad frame of critical theory: historical and demographic studies, textual and policy analyses, local qualitative small-scale studies, and action research.

Recent Directions

In the past few decades, critical theory in curriculum moved from an almost exclusive focus on class or poverty to a period when gender and race became central issues, and attention to language, identity, and discourse as the means by which human subjects were formed in curriculum assumed greater importance. The question of whose interests are being represented in the curriculum remained strong, but so did questions about how young people were being shaped to fit new forms of society.

Recently, three further developments have been notable. One is a reemergence of interest in class, this time incorporating some perspectives from feminist and critical race work, and particularly taking an interest in the power of middle class parents in shaping the curriculum (Dolby and Dimitriadis, 2004). Second, there are some signs of moves by sociologists to try to work beyond critique and somewhat relativist perspectives on the social construction of knowledge to a new critical realist position on knowledge and learning (Young, 1998). Third, there has been an upsurge of work post 2000 concerned with the socially discriminatory effects of the contemporary emphasis on evidence and scientifically based research in government policies, especially in the USA. An important part of this work is interested in global forces on curriculum and, especially, the sources and effects of international forms of educational measurement and ideology as a frame for curriculum work. (Dale, 2000) Another important component of this critical analysis of contemporary curriculum directions has been Apple's (2006) work on the conservative public constituency, including religion, which underpins these conservative directions.

Critical Theory and Curriculum Development

Since a critical theory perspective works from a critique of the social *status quo* (especially related to inequalities), and because theorists in this tradition do not see education as the only source and potential remedy of the inequalities and subordinations they critique, the contribution of critical theory to curriculum is not always or primarily about developing a new foundation for curriculum development. Rather, maintaining a critical watch on curriculum, revealing undesired social effects that are produced by it, is a central concern of this approach. However, contributing to social change and to better forms of education is also a concern associated with this approach. In both forms of contribution, a critical theory approach may work on the big picture of how a particular curriculum configuration is an effect of particular historical traditions and purposes and interests, or may work on a micro- and local level, concerned to identify the subjective making of curriculum by particular groups of students and teachers in particular circumstances.

See also: Curriculum and Human Rights; Curriculum and Poststructuralist Theory; Curriculum and the Education of Cultural and Linguistic Minorities; Curriculum in Postcolonial Contexts; Curriculum Studies, Discourse Analysis, and the Construction of Historical Time; Gender and Curriculum; Globalization and Curriculum.

Bibliography

- Apple, M. W. (1979). *Ideology and Curriculum*. Boston, MA: Routledge.
- Apple, M. W. (2006). *Educating the "Right" Way: Markets, Standards, God and Inequality*. New York: Routledge.
- Bernstein, B. (1971). On the classification and framing of educational knowledge. In Young, M. F. D. (ed.) *Knowledge and Control: New Directions for the Sociology of Education*, pp 47–69. London: Cassell/Collier MacMillan.
- Bourdieu, P. and Passeron, J.-C. (1977). *Reproduction in Education, Society and Culture*. London: Sage.
- Bowles, S. and Gintis, H. (1976). *Schooling in Capitalist America*. New York: Basic Books.
- Carr, W. and Kemmis, S. (1986). *Becoming Critical: Education, Knowledge and Action Research*. Warrnambool, VIC: Deakin University Press.
- Connell, R. W., Ashenden, D., Kessler, S., and Dowsett, G. (1982). *Making the Difference: Schools, Families and Social Division*. Sydney, NSW: Allen and Unwin.
- Dale, R. (2000). Globalization and education: Demonstrating a "common world educational culture" or locating a "globally structured educational agenda"? *Educational Theory* **50**, 427–448.
- Dolby, N. and Dimitriadis, G. (2004). *Learning to Labor in New Times*. New York: Routledge Falmer.
- Ellsworth, E. (1989). Why doesn't this feel empowering? Working through the repressive myths of critical pedagogy. *Harvard Educational Review* **59**, 297–324.
- Fraser, N. (1997). *Justice Interruptus: Critical Reflections on the "Postsocialist" Condition*. New York: Routledge.
- Giroux, H. A. and McLaren, P. L. (eds.) (1986). *Critical Pedagogy, the State, and Cultural Struggle*. Albany, NY: State University of New York Press.
- Ladson-Billings, G. and Tate, W. (1995). Towards a critical race theory of education. *Teachers College Record* **97**, 47–68.
- Luke, C. and Gore, J. (eds.) (1993). *Feminisms and Critical Pedagogy*. New York: Routledge.
- Martin, J. R. (1994). *Changing the Education Landscape: Philosophy, Women and Curriculum*. New York: Routledge.
- McCarthy, C. and Crichton, W. (eds.) (1993). *Race Identity and Representation in Education*. New York: Routledge.
- Walkerline, V. (1988). *The Mastery of Reason*. London: Routledge.
- Young, M. F. D. (ed.) (1971). *Knowledge and Control: New Directions for the Sociology of Education*. London: Collier-Macmillan.
- Young, M. F. D. (1998). *The Curriculum of the Future: From the "New Sociology of Education" to a Critical Theory of Learning*. London: Falmer.

Further Reading

- Arnot, M. (2002). *Reproducing Gender? Essays on Educational Theory and Feminist Politics*. London: RoutledgeFalmer.
- Bernstein, B. (1975). *Class, Codes and Control. Vol. 3: Towards a Theory of Educational Transmissions*. London: Routledge/Kegan Paul.
- Bourdieu, P. (1996). *The State Nobility: Elite Schools in the Field of Power*, de Saint Martin, M. (trans.). Cambridge: Polity Press.
- Freire, P. (1972). *Pedagogy of the Oppressed*. Harmondsworth: Penguin.
- Freire, P. (2000). *Pedagogy of the Oppressed*, 30th anniversary edn., Ramos, M. B. (trans.). New York: Continuum.
- Gewirtz, S. (1998). Conceptualising social justice in education: Mapping the territory. *Journal of Education Policy* **13**, 469–485.
- Popkewitz, T. S. and Fendler, L. (eds.) (1999). *Critical Theories in Education: Changing Terrains of Knowledge and Politics*. New York: Routledge.
- Teese, R. (2000). *Academic Success and Social Power: Examinations and Inequality*. Melbourne, VIC: Melbourne University Press.
- Weis, L. and Fine, M. (eds.) (1993). *Beyond Silenced Voices: Class, Race, and Gender in United States Schools*. Albany, NY: State University of New York Press.

Curriculum and Poststructuralist Theory

J L Miller, Columbia University, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Deconstruction – Often misunderstood as well as incorrectly attributed to Jacques Derrida as the originator of the word, deconstruction illustrates that language works not because there is a correspondence between a sign and a thing or because of presence, but because there is a difference, an absence. Deconstruction shows how a discursive system functions, including what it excludes or denies.

Differance – Central to Derrida's theory of deconstruction, this term combines the sense of the English verbs to differ and to defer. Thus, in the concept of differance, meaning is produced through the dual strategies of difference and deferral.

Discourse – In Foucault's conceptualization, it means the discursive practices that themselves form the objects of which they speak, consisting of written or spoken words that are grouped according to certain rules established within the particular discourse.

Many of those involved in curriculum studies and practices worldwide still consider Ralph Tyler's basic principles of curriculum and instruction, also known as the Tyler rationale, to be the model that primarily guides curriculum development and design. However, many curriculum critics decry various reductions of Tyler's intentions into a factory model of education, a model grounded in a behaviorist psychology that conceives curriculum as bits of information and universally agreed upon knowledge that supposedly can be arranged in a linear order and transmitted to the learner. In such a model, curricular development, design, and, most importantly, curricular meaning are determined by relationships among Tyler's sequential steps of developing educational purposes or objectives, as well as educational experiences, and then deciding how these can be organized and evaluated in terms of learners' attainment of the predetermined objectives.

In contrast, poststructuralist theory is utilized by curriculum scholars who wish to challenge forms of Tyler's schema that appear to them as singular, predictable, ahistorical, ideologically neutral, linear, sequential, and measurable versions of what and how one might acquire what is regarded as subject-area-specific content knowledge. That knowledge most often is considered to be stable

and universally accepted as the truth or best knowledge about a particular topic and is presented as such in classroom textbooks and teachers' curriculum guides. However, moving away from a technical-rational emphasis on curriculum development, where curriculum is only conceived as predetermined content to be covered in accumulative and additive ways, curriculum theorists and developers who work from poststructuralist perspectives assume curriculum development to be a political act, with incomplete, fractured, and deferred meanings constantly shifting and reconstructing versions of particular content knowledge. Poststructuralist curriculum theorists and practitioners, especially, are invested in developing and proliferating a multiplicity of views on any given field of knowledge.

Specifically addressing issues of power, poststructuralist curricularists further examine not only what and whose determinations and creations of knowledge count, but also how and under what conditions particular discourses come to shape what gets constructed as curriculum. They especially attend to what cultural and social practices as well as discourses constitute, reproduce, or call into question what is generally assumed to be the content – or what educators often generically refer to as the curriculum – as well as how and in what sequences it supposedly might be learned. From poststructuralist perspectives then, it is impossible to conceive of curriculum design, development, and evaluation as totally knowable, predictable, and universally applicable processes. Instead, poststructuralists argue that, given wildly varying social and cultural educational contexts, participants, and discourses, there is no one best way to proceed. For, to state any curriculum content, process, or activity as definitive is to insist on an ultimate meaning and a predetermined way.

Poststructuralist curriculum theorists also argue that educators' and students' subjectivities – the unconscious and conscious emotions and thoughts of individuals, their senses of themselves and their varying ways of understanding their relations to the world – are always in process, contradictory, produced historically, and reconstituted in discourse each time they think or speak. Indeed, poststructuralist theory posits that subjectivities, including perceptions, biases, and desires, rather than being considered inherent and part of a constant essence in humanist conceptions of the individual, are socially constructed in language, and thus can be considered sites of both struggle and potential change (Weedon, 1997).

Poststructuralist curriculum scholars argue that the subjects' perceptions, biases, desires, and identity constructions,

as well as their assumptions about what knowledge is of the most worth and what therefore should constitute the curriculum, are framed by and formed within particular educational discourses. The political elements of the processes of such formations are the focus of many poststructuralist curriculum scholars and practitioners. These educators often examine concomitant influences of psychosocial experiences in and out of schools as well as social and cultural contexts, including, for example, positionings in relation to normative constructions of gender, race, ethnicity, age, sexual orientation, and ability. Within these framings, the individual is a subject – subjected to dominate discourses that often impose predetermined meanings and thus guide educational processes, such as curriculum development, which are usually based on generally accepted assumptions about what knowledge is of the most worth. At the same time, in relation to content that might be studied as well as to others with whom the teacher or learner might interact, the individual student or teacher is a potential site for a wide range of possible forms of subjectivity, and thus might generate fresh and unanticipated versions of knowledges and identities that cannot be predicted or controlled.

Poststructuralist Theory and the Field of Curriculum Studies

In the late 1960s and early 1970s, a group of curriculum scholars in the United States began questioning positivist, social efficiency, and managerial assumptions that had undergirded the early-twentieth-century origins and development of the field of curriculum. These scholars working in the late 1960s and 1970s were loosely connected through their opposition to narrow conceptions of curriculum as able to be determined and sequenced by lists of behavioral objectives as well as students' absorption of that content as able to be measured by evidence that standardized tests supposedly provided. Working differently within the movement known as reconceptualization, from a variety of theoretical and ideological perspectives, these scholars attempted, early on, to expand conceptions of curriculum to include examinations of both political and autobiographical aspects and lived experiences of curriculum conceived as both content and process. Those involved in the reconceptualization were united only in their attempts to call attention to the interrelatedness of individuals who were differently situated in varying social, cultural, and political contexts and discourses, and they were determined to examine implications of the situated and discursive nature of any construction of curriculum and pedagogy.

Structuralism and Poststructuralist Theory

At approximately the same time, a new wave of philosophers primarily emanating from France, although differing in their

theoretical and disciplinary foci, became known collectively as poststructuralists because of their major and sustained critiques of structuralism. Structuralism is the intellectual movement and philosophical orientation often associated initially with the Western discourses of Levi-Strauss, Marx, and Althusser, for example, who claimed to analyze and explain invariant structures in and constitutive of nature, society, and the human psyche. Structuralism challenged the humanist and the Enlightenment projects, which regarded history as progress, placed humans at the center of creation, and privileged rational thought and the Western culture. The linguist Ferdinand de Saussure and the anthropologist Claude Levi-Strauss, among others, argued that structuralism could explain how certain cultural content (elements of kinship systems and stories told in varying cultures, such as myths and fairy tales) could be considered models of invariant structures.

Thus, those associated with structuralism claimed that cultural phenomenon could be examined according to the underlying formal systems out of which these phenomena emerged. It was argued that language and culture acquire meaning as they were intertwined in complex abstract relations. In addition, structuralism, in great part due to Saussurean linguistics and the subsequent development of language as a field of study, regarded language as a transparent and mirror-like medium through which societal regularities reveal themselves and are taken as constituting reality. Aiming for and acquiring the status of a scientific view of language and culture, structuralism thus assumed a systemic center that organized and sustained an entire society and its sets of relations.

Poststructuralist theories, although not amenable to full explanation, are more modes of cognition than sets of propositions that can be listed in a linear, rational manner. Poststructuralist thinking was an outgrowth of structuralism in that it incorporated structuralism's attack on humanism and the Enlightenment project. For example, many, although not all, of its theorists supported the psychoanalytic work of Lacan, especially his decentering of the humanist subject through the construing of the conscious and unconscious mind as products of language, of the symbolic. At the same time, poststructuralist theories offered sustained critique of structuralism. One of the major poststructuralist breaks with structuralism involved theories that highlighted how the underlying systems that structuralism analyzed were themselves caught up in language. Michel Foucault's work, in particular, conceptualizes discourse – discursive practices that themselves form the objects of which they speak – as consisting of written or spoken words that are grouped according to certain rules established within the particular discourse. Unlike structuralism's foundational sets of relations and systems, Foucault asserts that discourse is historically, socially, and culturally contingent, and that major analyses should focus not on what a particular discourse means, but rather, on investigations of how it works,

under what conditions, and how discursive formations and practices are part of nondiscursive practices. Discourses thus have both disciplinary and disciplining effects. Discourse, according to Foucault, defines fields of inquiry and knowledge as well as how rules within those fields govern what can be said, conceived, and acted upon (Foucault, 1975/1977, 1980).

Jacques Derrida, in his early work, also criticized structuralism's presumption that language could be described as a static set of rules. His work demonstrated how those rules could be examined for their contingency and their dependence on temporality, for example. Derrida further concerned himself not only with the regulatory nature of discourses, but also with cultural texts and the impossibility of their intrinsic authority as accounts of truth. In particular, he insisted that the act of reading extends from literary texts to films, popular culture, political scenarios, and to works of art. Derrida's notion of reading insists that our ability to understand relies on our capacity to interpret signs that come to signify in ways that no particular author or speaker can constrain in advance through intention.

Thus, in questioning Saussure's claim that signs have an already-fixed meaning recognized by the fully conscious and rational speaking individual, Derrida shifts from the Saussurean focus on speech to concerns with reading, writing, and textuality. He replaces the fixed signifieds of Saussure's chains of signs with a concept of *différance*. The concept of *différance* is central to Derrida's theory of deconstruction and combines the sense of the English verbs to differ and to defer. Thus, in the concept of *différance*, meaning is produced through the dual strategies of difference and deferral. In theorizing the problem of difference, Derrida wrote the term as *différance*, not only to mark the way that signification works – one term referring to another, always relying on a deferral of meaning between signifier and signified – but also to characterize an ethical relation, the relation of sexual difference, and the relation to the other. For Derrida, then, there are no fixed signifieds (concepts). Signifiers (sound or written images), which have identity only in their difference from one another, are subject to an endless process of deferral. Thus, any representation, in which meaning is apparently fixed, is only a temporary retrospective fixing. The name that we have for something, for ourselves, and for an other, is precisely what fails to capture the referent (as opposed to making or constructing it). Further, signifiers are always located in a discursive context, and any temporary fixing of meaning in a specific reading of a signifier depends on this discursive context (Butler, 2004; Weedon, 1997).

Conceptualizing any encounter with or construction of social and cultural worlds as discursive text with no original meaning, Derrida further argued that every structure – whether it be literary, political, religious, educational, or economic, for instance – that organizes our experiences within that particular structure, is constituted and

maintained through acts of exclusion. In the processes of developing curriculum, for example, by adding in certain presumed facts or content, something else inevitably gets left out. Such exclusions, often predicated on reified notions of difference, are the focus of Derrida's conceptualization of deconstruction. Deconstruction, often misunderstood as well as incorrectly attributed to Derrida as the originator of the word, illustrates that language works not because there is a correspondence between a sign and a thing or because of presence, but because there is a difference, an absence. Deconstruction shows how a discursive system functions, including what it excludes or denies (Derrida, 1998). At the same time, deconstruction cannot by definition be defined since it presupposes the undecidability of all conceptual or generalizing terms. Like any method of interpretation, it can only be exemplified, and the examples will all differ.

Deconstruction, then, according to Derrida, is only what happens if it happens because it is not a philosophy, a doctrine, a knowledge, a method, a discipline, and not even a determinate concept (Derrida, 2001). If it does happen, deconstruction enables one to critique structures that are held together by identity and presence, concepts that, in Western philosophy, represented transcendental order and permanence as manifested in beliefs and ideas such as the unified subject, the essence of an individual, and consciousness. Derrida utilized deconstruction not to dismantle, reject, or take things apart (as the term has come to be generically and incorrectly used), but rather to reinscribe them in another way. In particular, deconstruction allows one to challenge any notion of foundational center that creates binaries in which the first term of the binary most always indicates presence and power and, subsequently, to attempt to reconstitute that which has been previously inscribed. In addition, the reconstitution must then, in turn, be deconstructed. Derrida understood that social and political transformation is an incessant project, one that can not be relinquished, co-exists with encounters with the other, and insists on a reading of the rules by means of which a society constitutes itself through exclusion or effacement. Feminist theorists Gayatri Spivak and Judith Butler, in particular, consider deconstruction to be an affirmative political practice in that it enables us to rewrite the world and ourselves over and over again (Butler, 2004; Spivak, 1974).

Curriculum Reconceptualized in Relation to Poststructuralist Theory

In particular, the translation and dissemination of the work of Michel Foucault and Jacques Derrida in North America during the 1970s and 1980s enabled some curriculum scholars to take up major aspects of the French poststructuralist theory by addressing the central role of language, power, and discourse in any model or conception of curriculum

theory, development, and design. The further dissemination of this theory throughout the world, including the work of French feminists Helene Cixous, Luce Irigaray, and Julia Kristeva, encouraged curriculum scholars, within a variety of social and cultural contexts, to pursue poststructuralists' particular goal of troubling both discursive and material structures that limit or reify conceptions and enactments of curriculum.

Some of the earliest poststructuralist work in curriculum, especially theorizing in the United States that grew from the initial movement to reconceptualize the curriculum field, drew on the theories of Foucault and, secondarily, on the work of Derrida and Lacan in order to challenge essentialist notions of gender identity and examine various gender discourses that often were linked to the very same oppressive discursive systems they sought to dismantle. Explorations of the ways that discourse creates and is substantiated by the body and the unconscious followed, as did work that staged, in writing, deconstructive and poststructuralist performances of thinking about and enacting versions of curriculum that defied dominant positivist and behavioral metanarratives about what knowledge is of the most worth and how it should be presented in textbooks as well as in pedagogical strategies. Such critique became emblematic of contemporary contentions that currently characterize the field of curriculum studies, writ large, in terms its functions, philosophical and ideological commitments, and manifestations of those commitments for curriculum development and design in schools.

By the late 1980s and early 1990s, curriculum scholars were conducting studies of curriculum as phenomenological and deconstructed texts that emphasized multivocality, multiperspectivity, and the lived aspects of textbooks and classrooms. Some explained and analyzed poststructuralist theory in relation to structuralism. Others explored how curriculum as a field of study might now be characterized by ideas and metaphors from the new sciences of complexity, uncertainty, ambiguity, open systems, process, and transformations. A major poststructuralist line of inquiry in curriculum contrasted the behavioral orientation of the Tyler rationale and other positivist approaches to curriculum development with postmodern possibilities of a curriculum that is composed of complex and spontaneous interactions among students and teachers, and content both created from and with/in those interactions. Threaded throughout such poststructuralist curriculum work are constant critiques of any totality of representation that reduces learning to information transmission. Poststructuralist curricularists argue that such reductive narratives of education continue to foster the bifurcations that ignite racism, patriarchy, homophobia, colonialism, and classism. At the same time, others, working from still dominant neo-Marxist positions in curriculum theorizing, argue that the central tenet of poststructuralist theory and postmodernism, writ large – that there can only be

incredulity toward metanarratives (Lyotard, 1979) – is itself a grand narrative. Building upon these major influences on as well as upon critiques of the introduction of poststructuralist theory into curriculum theory, development, and design, a number of scholars have further explored poststructuralist theories from a variety of angles and for a number of educational purposes.

Women academics within the field of curriculum studies and practice began their poststructuralist work by attempting to conceptualize a multiple, indeterminate female subject. Multiple manifestations of this initial work have contributed to one aspect of what, some would argue, is now a major influence on the field of curriculum – the work of feminist poststructuralists.

Contributions of Feminist Poststructuralist Theorizing

Poststructuralist versions of feminism almost always advocate social change and, at the same time, mark a turn away from projects that promote humanist assumptions about progress, identity, and fully conscious selves who can attain, once and for all, freedom from oppression. Most poststructuralist feminists are concerned with gender power relations, in particular, and how those relations are constituted, reproduced, and, possibly, contested. A proliferation of work produced by feminists working in the field of curriculum as well as education, in general, has altered ways in which educators must now attend to issues of language, discourse, and power in all aspects of curriculum conception and development, as well as in the ways that curriculum is enacted, experienced, and created in the classroom.

A number of feminist scholars, especially in Australia, the UK, and the US, have drawn on the work of Foucault, in particular, as they work to understand and then to critique how modernist, humanist conceptions of a woman have been constructed through and by dominant discourses in societies in general, and in the field of education in particular (Baker and Heyning, 2004; Luke and Gore, 1992; Tamboukou, 2003). Many have relied on Foucault's work in their analyses of gender, the category woman, and constructions of woman and girl as subject and object in curriculum and pedagogy discourses. Some feminist scholars have studied how children, especially girls, are constituted as subjects within what are typically named child-centered curricula and forms of pedagogy (Walkerdine, 1990). Influential work, focused on Foucauldian issues of power, discourse, discursive practices, and the production of selves, has also centered on qualitative studies of preschool children and gender, as well as on children reading and writing beyond gendered identities (Davies, 2003).

Further, in investigating, from a variety of subject matter areas and disciplinary perspectives, a woman's subjugated positioning within educational discourses that focus on

binaries such as normal/abnormal or active/passive, many feminists utilize Foucault's insistence on historical analyses as well as attention to the ways in which attempts to assert legitimate claims to knowledge are often caught up in the very essentializing and patriarchal discourses that women wish to combat. Exploring additional ironies to which poststructuralist theory points, feminist curriculum scholars also have investigated how radical discourses in education, including feminist pedagogy, paradoxically operate as regimes of truth, to use Foucault's conceptualization. Some feminists, especially, have utilized aspects of Foucauldian poststructuralist theory in order to challenge essentialist and unitary notions of voice and dialog, two prominent components of critical pedagogy as well as of some versions of feminist pedagogy (Ellsworth, 1997; Orner, 1992).

Another major arena in which feminist poststructuralists have had and continue to exert a major influence on is qualitative curriculum research methodologies and practices. By addressing issues such as power relations with subjects in the field, the crisis in representation, and the stuck, unpredictable, and unknowable places and spaces of qualitative research, especially in relation to difference, feminist poststructuralist work has influenced curriculum scholars who wish to explore such issues within educative contexts. These feminists, for example, simultaneously, both use and immediately trouble typical categories of qualitative research, such as validity and generalizability, and move toward methodologies that foreground ambiguities, uncertainties, contradictions, and incoherences (Lather, 2007; Pierre and Pillow, 2000). In addition, some feminist poststructuralist curriculum theorists continue to grapple with how to conceptualize self, woman, teacher, researcher, or curriculum, not as permanently essentialized or naturalized through humanist and positivist educational discursive practices and regimes, but rather, as sites for cultural critique and social change (Miller, 2005).

Toward Further Ruptures

The wide-ranging areas of curriculum theorizing and inquiry that fall under the general heading of poststructuralist theory and curriculum have expanded rapidly in the late 1990s and into the twenty-first century. Major contributions of poststructuralist curriculum scholars, teachers, and curriculum developers have highlighted how poststructuralist theorizing challenges master narratives, postulates that all meaning is discursively constructed, and insists that we, as educators, examine how and under what conditions particular discourses come to shape what gets constructed as knowledge. At the same time, it is impossible to claim definitive and static manifestations of or future directions for poststructuralist theory in relation to curriculum theorizing and development.

What is possible to note here is the proliferation of those involved in the work of curriculum, writ large, who

are searching for ways in which to confront the reemergence of mechanistic, technical-rational approaches to teaching, curriculum conceptualizing, and development, and of standardized tests as the major tool for the evaluation of students' knowledge and teacher effectiveness. A number of curriculum scholars and practitioners are looking to the poststructuralist theory in order to call into question, for example, universalized versions of school reform efforts and educational policy mandates. Indeed, the poststructuralist theory offers perspectives that enable curriculum workers who wish to work toward creating ruptures, toward asking "... questions about necessary complicities, inadequate categories, dispersing rather than capturing meanings, and producing bafflement rather than solutions" (Lather, 2007: viii) to pose challenges to the educational *status quo*. The poststructuralist theory itself challenges all those involved in the curriculum field to interrogate our taken-for-granted assumptions as we investigate the discursive practices and relations of power that underlie any one answer to the classic curriculum question, what knowledge is of the most worth.

Bibliography

- Baker, B. and Heyning, K. E. (eds.) (2004). *Dangerous Coagulations? The Uses of Foucault in the Study of Education*. New York: Peter Lang.
- Butler, J. P. (2004). *Precarious Life: The Powers of Mourning and Violence*. New York: Verso.
- Davies, B. (2003). *Shards of Glass: Children Reading and Writing beyond Gendered Identities*, rev. edn. Cresskill, NJ: Hampton Press.
- Derrida, J. (1998). *Of Grammatology*, Spivak, G. C. (trans.) Baltimore, MD: The Johns Hopkins University Press.
- Derrida, J. (2001). *Writing and Difference*, 2nd edn. New York: Routledge.
- Ellsworth, E. (1997). *Teaching Positions: Difference, Pedagogy, and the Power of Address*. New York: Teachers College Press.
- Foucault, M. (1975/1977). *Discipline and Punish*, Sheridan, A. (trans.). New York: Pantheon.
- Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings*, pp 1972–1977. New York: Pantheon.
- Lather, P. (2007). *Getting Lost: Feminist Efforts Toward a Double(d) Science*. Albany, NY: State University of New York Press.
- Luke, C. and Gore, J. (eds.) (1992). *Feminisms and Critical Pedagogy*. New York: Routledge.
- Lyotard, J.-F. (1979). *The Postmodern Condition: A Report on Knowledge*. Paris: Minuit.
- Miller, J. L. (2005). *Sounds of Silence Breaking: Women, Autobiography, Curriculum*. New York: Peter Lang.
- Orner, M. (1992). Interrupting the calls for student voice in "liberatory" education: A feminist poststructuralist perspective. In Luke, C. and Gore, J. (eds.) *Feminisms and Critical Pedagogy*, pp 74–89. New York: Routledge.
- Pierre, E., St. and Pillow, W. (eds.) (2000). *Working the Ruins: Feminist Poststructural Practice and Theory in Education*. New York: Routledge.
- Spivak, G. C. (1974). Translator's preface. In Derrida, J. (ed.) *Of Grammatology*, Spivak, G. C. (trans.), pp ix–xc. Baltimore, MD: Johns Hopkins University Press.
- Tamboukou, M. (2003). *Women, Education, the Self: A Foucauldian Perspective*. New York: Palgrave Macmillan.
- Walkerline, V. (1990). *Schoolgirl Fictions*. London: Verso.
- Weedon, C. (1997). *Feminist Practice and Poststructuralist Theory*, 3rd edn. Oxford: Blackwell.

Further Reading

- Anyon, J. (1994). The retreat of Marxism and socialist feminism: Postmodern and poststructural theories in education. *Curriculum Inquiry* **24**, 115–134.
- Baker, B. (2001). In *Perpetual Motion: Theories of Power, Educational History, and the Child*. New York: Peter Lang.
- Britzman, D. (1995). The question of belief: Writing poststructural ethnography. *Qualitative Studies in Education* **8**, 233–242.
- Butler, J. (1995). Contingent foundations: Feminism and the question of 'postmodernism'. In Caruth, C. and Esch, D. (eds.) *Critical Encounters: Reference and Responsibility in Deconstructive Writing*, pp 213–232. New Brunswick, NJ: Rutgers University Press.
- Butler, J. P. (2005). *Giving an Account of Oneself*. New York: Fordham University Press.
- Cherryholmes, C. (1988). *Power and Criticism: Poststructural Investigations in Education*. New York: Teachers College Press.
- Davies, B. (2000). *A Body of Writing 1990–1999*. Walnut Creek, CA: AltaMira Press.
- Doll, W. E., Jr. (1993). *A Post-Modern Perspective on Curriculum*. New York: Teachers College Press.
- Martusewicz, R. A. (1992). Mapping the terrain of the post-modern subject: Post-structuralism and the educated woman. In Pinar, W. F. and Reynolds, W. M. (eds.) *Understanding Curriculum as Phenomenological and Deconstructed Text*, pp 131–158. New York: Teachers College Press.
- Peters, M. A. and Burbules, N. C. (2004). *Poststructuralism and Educational Research*. New York: Rowman and Littlefield.
- Pierre, E., St. (2000). Poststructural feminism in education: An overview. *Qualitative Studies in Education* **13**, 477–515.
- Pinar, W. F. (ed.) (1975). *Curriculum Theorizing: The Reconceptualists*. Berkeley, CA: McCutchan.
- Pinar, W. F. and Reynolds, W. M. (eds.) (1992). *Understanding Curriculum as Phenomenological and Deconstructed Text*. New York: Teachers College Press.
- Pinar, W. F., Reynolds, W. M., Slattery, P., and Taubman, P. M. (1995). *Understanding Curriculum: An Introduction to the Study of Historical and Contemporary Discourses*. New York: Peter Lang.
- Slattery, P. (2006). *Curriculum Development in the Postmodern Era*, 2nd edn. New York: Routledge.
- Tyler, R. (1949). *Basic Principles of Curriculum and Instruction*. Chicago, IL: University of Chicago Press.

Curriculum and Structuralist Sociology: The Theory of Codes and Knowledge Structures

J Muller and J Gamble, University of Cape Town, Cape Town, South Africa

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Classification – An analytical concept which refers to the power relations that result from the strength of boundaries between contexts.

Code – An orientation to meaning into which children are socialized at home and at school. Home codes may or may not be congruent with school codes that communicate knowledge at different levels of abstraction.

Framing – An analytical concept which refers to the social relations of control between teacher and pupils in a pedagogic practice.

Horizontal discourse – Meanings relating to a particular context, understood experientially and through the senses.

Realism – Realist social science is the understanding of society in terms of social mechanisms that structure but do not determine outcomes.

Structural functional linguistics – A socio-semiotic approach to language that analyzes how grammar realizes meanings through the functions language is required to serve. It is associated with M.A.K. Halliday and the Sydney School.

Structuralism – A theoretical approach that explains the social world in terms of law-like patterns of relations, rather than in terms of individual actions or individual consciousness.

Vertical discourse – Meanings that operate at a general principled level. Nonempirical meanings that are generated through deductive and inductive reasoning. Meanings not derived experientially from immediate contexts.

During the 1960s and into the 1970s, structuralism as a theoretical style became influential in a number of disciplines in the humanities and social sciences. While several lines of descent can be traced, this article focuses on the two most directly relevant to the structural turn in the sociology of curriculum. The first derives from Durkheim and his contention that society exists over and above the individual, nested in the relationship between the social base of society and its production of consciousness (Davies, 1993). The second derives from the structural linguistics of Ferdinand de Saussure, which viewed language as a system

of signs and investigated the underlying set of laws (langue) by which signs combine into meanings that make speech acts possible (parole). This led to the practice of treating the social world as a set of semiotic patterns with a logical law-like basis. Linguistic structuralism thus examines the underlying systems of relationships which order forms of social life, treated as though they are homologous with those of language. Taking these two lines of descent together, the curriculum is treated as an object produced by social relations which order consciousness in determinate ways, and which are related in turn to the social division of labor in society.

The 1960s was a time of rapid increase in public expenditure on schools and universities and increasing differentiation of educational systems. Concern arose as to whether educational investments were efficient or equitable, prompting government-funded research initiatives ranging from a political arithmetic approach that calibrated the chances of children of different class origins reaching various educational destinations, to causal models of attainment difference based on environmental and genetic/intelligence differences, and to psychometric testing. These studies were widely regarded as atheoretical black box studies.

Structuralist sociologies of education were one notable reaction against this policy-driven empiricism. A prominent variant was the work of Pierre Bourdieu and his group in France. In this article, we examine the structuralist sociology of curriculum of the British sociologist Basil Bernstein, and the theoretical and empirical work he has subsequently inspired. Bernstein went against the grain of conventional orthodoxy in that he viewed language (environment) and not intelligence (genetics) as making the difference to educational chances. While what came to be known in Britain in the 1970s as the new sociology of education is characterized by a number of sociological approaches (phenomenological, structuralist, and neo-Marxist) that shared a common concern about the relation of curriculum and pedagogy to social class and academic achievement, it is particularly the work of Bernstein and his longstanding dialog with systemic functional linguistics (SFL) in the work of Michael Halliday and Ruqaiya Hasan that best exemplifies the structuralist turn in sociology of education and in sociological curriculum theory in particular. Even though Bernstein identified himself as a Durkheimian rather than as a conventional structuralist, his work should arguably be located within the structuralist tradition and related to a French rather than an Anglo-American reading of Durkheim

(Atkinson, 1985). The remainder of the article is thus dedicated to a brief exposition of two major components of his theoretical work, namely the theory of codes and of knowledge structures and the fruitful avenues of research that these have opened up.

The Theory of Codes

For Bernstein, the basic unit of socio-educational analysis is not the individual but the communicative relation and its control (Bernstein, 1990). All communication has a pedagogic (or formative) effect on consciousness. The key regulator of communication is the code.

Bernstein's early work on communicative codes examined the production and reproduction of meaning systems in the home and the school, with class relations taken to refer to inequalities in the distribution of power (Bernstein, 1990). The communicative relation in the home is the primary socializing context, which creates for each child a primary orientation to meaning, or semantic style. This style is either in harmony with the style favored by the school code, which is an advantage; or it is not, which is a disadvantage. The school code in all formal schooling favors a semantic style that is elaborated, context independent, and capable of generating abstraction, generalization, and explanation, all of which are generic modes required by formal learning. The home code comes in one of two modal types: an elaborated code variant typically found in middle-class homes, and a restricted code variant typically found in working-class homes.

This appellation of restricted turned out to be a most unfortunate one, and early code theory was all but derailed by socio-linguists such as Labov and Rosen who took the designation as impugning the capacity and dignity of the working class. Two facts have seen it survive this early critique. The first is the enduring feature, discernible in virtually all formal schooling settings, of working-class under-attainment, for which code theory still offers one of the best socio-cognitive explanations. (Cuba may just be the exception that proves the rule.) The second is the subsequent empirical work of structural functional linguists such as Hasan, Cloran, Painter, and Williams (see Hasan, 2005), who have fleshed out in rich detail the semantic features of the communicative code of mother-child talk in the early years. This group of linguists has shown that:

Mothers from different social groups systematically vary the types of meanings they communicate to their children, which decisively affects the child's orientation to meaning/semantic style. Where the 'formative' maternal style leads to a communalized cultural style and personal orientation, the 'informative' style leads to an individualized orientation that fosters a far greater sense of individual control over the manipulation and extension of meanings; and,

This home-learned orientation to meaning has long lasting effects on the child's receptivity to types of meaning expected at school. It does this by creating certain semantic expectations (recognition rules) by which children understand what is required of them, and understand what constitutes an adequate performance (realization rules). These semantic styles can be adapted, but only under certain communicative / pedagogic conditions, as we shall show below. (Hasan, 2005)

The study by Holland (1981) provides a graphic display of these orientations. He asked two sets of 7-year-old children, one working class and another middle class, to group pictures of foodstuffs and to give reasons for their grouping. They were then asked to repeat the task. Holland found that working-class children invariably used context-dependent, everyday criteria for grouping the foods (e.g., 'I like those things'; 'That is what mother cooks for breakfast'), whereas the middle-class children, recognizing the request as pedagogic, classified first on the basis of a general principle (e.g., a food category like dairy or vegetable) but in the second round also tended to give local, context-dependent criteria. He concluded that the middle-class children had been taught at home to recognize and realize abstract classification, while the working-class children did not recognize this as the requirement and thus supplied only their everyday categories. In other words, while the middle-class children had already been introduced to the elaborated school code at home, the working-class children had to acquire it virtually from scratch when they got to school.

Anatomy of the School Code

Bernstein's further work explored the modalities of elaborated code transmitted in schools. While much of the work that was being done in the sociology of education in the 1970s and 1980s investigated pedagogic communication as a relay for power relations external to education (class, race, gender, and religion), his work focused on power and control relations within education itself. His question was: "If pedagogic communication is a relay for patterns of domination external to education, what makes the relay possible?" (Bernstein, 1990).

In order to answer this question, Bernstein developed a specialized language with which to distinguish more precisely between modalities of elaborated codes as relayed by pedagogy. The features of any communicative context are represented by two principles: a classificatory principle and a framing principle. The classificatory principle creates specific recognition rules whereby a context is distinguished and given its position in relation to other contexts. The degree of insulation determines the degree of specialization of a communicative context and of the knowledge

transmitted, with strong (C+) or weak (C-) classification indicating the strength of the boundary.

Framing is about who controls what at the level of pedagogic practice. At the instructional level, framing refers to the nature of control over:

- selection of the content of the communication;
- its sequencing (what comes first, what comes next);
- its pacing (the rate of transmission and expected acquisition); and,
- criteria for its evaluation.

Like classification, framing can be strong (F+) or weak (F-). Where framing is strong, the teacher explicitly controls features of the framing relation; where framing is weak, the pupils apparently have more control over the selection, sequencing, pacing, and evaluation of the knowledge transmitted. At the regulative level framing refers to whether the hierarchical relation between teachers and pupils is explicitly controlled by the teacher (strong framing over teacher-pupil relations), or implicitly and apparently controlled by the students (weak framing over teacher-pupil relations). The strength of both classification and framing can vary independently of each other. It is, for instance, possible to have weak classification and strong framing of certain or all of the features of the communicative context, just as it is possible to have strong classification and weak framing. Framing values can also vary independently of one another. Framing over evaluative criteria could, for instance, be strong while framing over pacing is weak.

For Bernstein codes are not only mechanisms for social reproduction; they are also sites of contradiction, challenge, and change (Bernstein, 1990). The re-description of the fundamental relation between the social division of labor (classification) and forms of pedagogic communication (framing) in code form provides researchers of classroom practice with the means to link empirical evidence to a theoretically generated network of related concepts to make visible how power and control translate into principles of communication and "how these principles of communication differentially regulate forms of consciousness with respect to their reproduction and the possibilities of change" (Bernstein, 2000: 4).

This opened up the way for a series of empirical studies of great significance because it marked a major theoretical advance that provided a tool with which to investigate what it was that rendered the expectations of the school code so peculiarly invisible to children from working class homes. In an earlier iteration of the theory, Bernstein had distinguished between two modal types of pedagogic communication – visible (strong classification and strong framing) and invisible (weak classification and weak framing) – and had shown that the invisible mode communicated well to the middle-class children; in other words, the requirements were not invisible to them but left the working-class children struggling to discern what was required.

The proto-concepts of visible and invisible pedagogies left the theory with a description of the consequences of different pedagogic modalities, rather than with a stipulation as to what exactly it was that was visible or invisible. Was it everything? (i.e., should everything be strongly classified and framed to render the invisible visible?) This was never going to be happily accepted by the pedagogic community at large, who were by the mid-1980s and 1990s deeply in the impress of Vygotskian constructivism and Deweyan progressivism, both interpreted as favoring an invisible pedagogy. However in any case, why should it be all or nothing? With this refinement of code theory, a series of studies, first by the ESSA Group (Sociological Studies of the Classroom) in Lisbon, later by others in Australia, South Africa, and the United Kingdom, explored the specific lineaments of a pedagogy that might successfully convey expectations equally to all children, but especially to those of the working class.

The ESSA group led by Morais was concerned mainly with primary school science acquisition (Morais and Neves, 2001). They developed the conceptual categories outlined above into an observational grid which captured values for each conceptual category on a four-point scale. They were able to use a variety of measures to show that a certain mixed pedagogic modality, together with a scientifically competent teacher who could retain the proper level of cognitive challenge, produced the fewest failures and the greatest degree of success across all classes. This modality combined strong and weak features in the same pedagogy.

As the systemic functional linguists have shown, the semantic orientation of working-class pupils coming to school is generally context specific, local, and communalized at the family level. This orientation ensures that the decontextual, abstract mode of much of school science and its requirements will be invisible to such pupils. Where should these requirements pre-eminently be made visible? The theory predicts that there are two places where visible signaling will be critical: first, in the curricular stipulations of the intended curriculum (i.e., strong framing over external selection); second, in the evaluation criteria which signal to pupils whether they have understood what was expected and have met the criteria for producing an adequate response (i.e., strong framing over the evaluation criteria). The first is the responsibility of the curriculum writers; the second, the responsibility of the teacher.

In order for the teacher to be able to communicate the criteria in a comprehensive way to pupils whose orientation to elaboration required careful nurturing, two other features of the pedagogic modality were necessary: time (weak framing over pacing) and a personalized attitude to the pupils (weak framing over teacher-pupil relations). The ESSA regression analyses showed that these four aspects (i.e., strong framing over external selection and evaluation criteria and weak framing over pacing and teacher-pupil relations) accounted for a significant part of the variance in

achievement outcomes. There are other features of the ESSA mixed pedagogy, but these four are the most significant: two traditionally seen as part of progressive pedagogy and two as part of a (supposedly opposed) more didactic mode. When the ESSA team trained teachers to teach to this mixed pedagogy, and compared their success rate with teachers teaching to other configurations, their expectations were confirmed.

In a number of subsequent studies this general pattern has been confirmed. Hoadley (2007), for example, looking at grade 6 mathematics learners in South Africa, found a very similar configuration linked to successful (in this case middle-class) pupil attainment; clear and explicit evaluation criteria, linked to weakened control over both pacing and teacher–pupil relations. Hoadley’s further finding is a little more sobering: she shows that not only the pupils but also the teachers from working-class homes come to the classroom with a communalizing orientation, which creates an additional barrier to the individualizing mode of teaching required for pedagogic voice specialization. The great virtue of this group of studies is to show that under the right socio-cognitive conditions, teachers can make a difference, and to point with some precision to the various dimensions of pedagogic modality that could bring this about.

Knowledge Structure

Code theory established the how of curriculum and pedagogy – the various modalities of transmission, but it remains silent on the what – the knowledge that is thereby transmitted. The theory maintains that curricular knowledge – the what of the classroom – is recontextualized from that produced in the laboratory and the library, real knowledge so to speak, but pursues the matter no further.

It was only late in his career that Bernstein turned to the question of what this knowledge was which was to be acquired, its structure, and its social base. Bernstein draws a strong distinction between two basic classes of knowledge: mundane or everyday knowledge and esoteric or universal, principled knowledge. These two classes of knowledge are intrinsic to language, and they exist in all societies, even though their content may vary historically and culturally. This is not to say that some knowledge is concrete as opposed to abstract but rather that there is a difference in the form that abstraction takes in relating an everyday, mundane world accessed through experience, to the abstraction of a transcendental world accessed through formal reasoning. The relation between these two worlds can be direct or indirect. A direct relation between meanings and a specific material base is termed horizontal discourse. In horizontal discourse meanings cannot transcend their immediate context and so always refer to everyday or mundane contexts. Vertical discourse, by contrast, requires systematic ordering principles for the generation of meaning.

The knowledge bits fit together in a time and space not given by a specific context.

Context-independent meanings take two distinct forms. One is through what Bernstein calls a hierarchical knowledge structure, where general propositions and theories integrate knowledge at lower levels into a higher level in the hierarchy (as in the natural sciences). A second is horizontal, taking the form of a series of specialized languages that lie next to one another (as in the social sciences and humanities). Through a principle of recontextualization, competent members of vertical discourse can give an explicit account of the way in which they have arrived at a specific position and they can retrace their steps and show how they have made the recontextualized objects hang together (Muller, 2000).

By taking disciplines such as science and history for example, the structural functional linguistic Sydney School has explored knowledge structure variation in terms of register and lexical specialization. They isolate grammatical metaphor as the “critical linguistic resource . . . for the construction of vertical discourse” (Martin, 2007: 60). Other advances in the theory of knowledge structure include the Moore (2007) work on the production of esthetic judgments that suggests conditions for progress in knowledge in the arts and humanities, and the Gamble (2004) study of the tacit base of craft knowledge as a horizontal knowledge structure.

Knowledge and Curriculum

The theory of knowledge structures brings to light the differential coherence relations of different knowledges: vertical coherence relations in hierarchical knowledge structures and connective coherence relations in horizontal knowledge structures. Curriculum theorists soon made the connection between these arrangements and the arrangements of the curriculum specified by code theory. In particular, what would be the consequences of, say, a C+/F+ or C-/F- pedagogical modality for the learning of school subjects that are either vertically or connectively coherent? The empirical research has so far concentrated on the consequences of a weakly classified and framed pedagogy for vertical coherence subjects such as mathematics and English. The theory would predict that such a modality would render invisible the sequential or necessary steps of the vertical structure to be acquired. If the evaluation criteria require competence in all the sequential steps, and are also weakly framed – that is, weakly specified – then learners might very easily get to the final examination only to find that they have not traversed the expected vertical steps. The discussion above would suggest that working-class learners are most at risk in such a modality.

Outcomes-based curricular systems like those found in Australia and South Africa have tended to favor weak

specification of content (weak intra-discursive classification), weak framing over sequence and pace, and weak framing of evaluative criteria – this latter a consequence of evaluation criteria specified in skill terms. In short, this modality disguises the what from both teachers and acquirers. When teachers are knowledgeable (Morais and Neves, 2001), this can be offset; when they are not, as is often the case in South African schools, the pacing slows down, and critical parts of the curriculum are not covered. In a vertical coherence subject such as mathematics, this freezes further learning (Reeves and Muller, 2005). In Australia, it has similarly been found that weak specification of the what has tended to disguise the vertical spine of school subjects. Christie and her Sydney School collaborators have rendered visible the verticality of subjects such as English by reconstructing a linguistic meta-language to specify the operations students should master to meet the invisible (weakly framed) evaluation criteria successfully (Christie and Macken-Horarik, 2007). What these and other studies show is that determinate structural forms of a particular knowledge field place distinct limits on effective curricular structure.

Conclusion

In the light of the discussion above, it may be useful to revisit in what way this approach to the sociology of curriculum should be regarded as structuralist. The Durkheimian and linguistic influences are clearly apparent. There is however a deeper sense in which this approach can be distinguished from the empiricism of current approaches to both school effectiveness and textual cultural critique. In a way that could more properly be called realist (Manicas, 2006), Bernstein and those using his theory are engaged in a search for explanatory mechanisms (structures) for reading the world. With the theory of codes, and in a less elaborated form with the theory of knowledge structures, researchers are able to create a powerful and precise language for modeling pedagogic modalities and explaining their effects. This search for explanatory mechanisms is what distinguishes this approach to the sociology of curriculum, whether we call it structuralist or realist.

Bibliography

- Atkinson, P. (1985). *Language, Structure and Reproduction: An Introduction to the Sociology of Basil Bernstein*. London: Methuen.
- Bernstein, B. (1990). *The Structuring of Pedagogic Discourse, Vol. IV: Class Codes and Control*. London: Routledge.
- Bernstein, B. (2000). *Pedagogy, Symbolic Control and Identity*. rev. edn. Lanham, MD: Rowman and Littlefield.
- Christie, F. and Macken-Horarik, M. (2007). Building verticality in subject English. In Christie, F. and Martin, J. R. (eds.) *Language, Knowledge and Pedagogy: Functional Linguistic and Sociological Perspectives*, pp 156–183. London: Continuum.

- Davies, B. (1993). Durkheim and the sociology of education in Britain. *British Journal of Sociology of Education* 15, 3–25.
- Gamble, J. (2004). Retrieving the general from the particular: The structure of craft knowledge. In Muller, J., Davies, B., and Morais, A. (eds.) *Reading Bernstein, Researching Bernstein*, pp 189–203. London: Routledge/Falmer.
- Hasan, R. (2005). *Language, Society and Consciousness, Vol. 1: The Collected Works of Ruqaiya Hasan*. London: Equinox.
- Hoadley, U. (2007). Reproduction of social class inequalities through mathematics pedagogies in South African primary schools. *Journal of Curriculum Studies* 39, 1–28.
- Holland, J. (1981). Social class and changes in the orientations to meanings. *Sociology* 15, 1–18.
- Manicas, P. (2006). *A Realist Philosophy of Social Science: Explanation and Understanding*. Cambridge: Cambridge University Press.
- Martin, J. R. (2007). Construing knowledge: A functional linguistic perspective. In Christie, F. and Martin, J. R. (eds.) *Language, Knowledge and Pedagogy: Functional Linguistic and Sociological Perspectives*, pp 34–64. London: Continuum.
- Moore, R. (2007). Hierarchical knowledge structures and the canon: A preference for judgements. In Christie, F. and Martin, J. R. (eds.) *Language, Knowledge and Pedagogy: Functional Linguistic and Sociological Perspectives*, pp 109–128. London: Continuum.
- Morais, A. and Neves, I. (2001). Pedagogic social contexts: Studies for a sociology of learning. In Morais, A., Neves, I. P., Davies, B., and Daniels, H. (eds.) *Towards a Sociology of Pedagogy: The Contribution of Basil Bernstein to Research*, pp 185–221. New York: Peter Lang.
- Muller, J. (2000). *Reclaiming Knowledge: Social Theory, Curriculum and Education Policy*. London: Routledge/Falmer.
- Reeves, C. and Muller, J. (2005). Picking up the pace: Variation in the structure and organisation of learning school mathematics. *Journal of Education* 37, 103–130.

Further Reading

- Cope, W. and Kalantzis, M. (eds.) (1993). *The Powers of Literacy: A Genre Approach to Teaching Literacy*. London: Falmer.
- Fitz, J., Davies, B., and Evans, J. (2006). *Educational Policy and Social Reproduction: Class Inscription and Symbolic Control*. London: Routledge.
- Hedstrom, P. and Swedberg, R. (2001). Social mechanisms. *Acta Sociologica* 39, 281–308.
- Karabel, J. and Halsey, A. H. (eds.) (1977). Educational research: A review and an interpretation. In *Power and Ideology in Education*, pp 1–85. New York: Oxford University Press.
- Moore, R. and Young, M. (2001). Knowledge and the curriculum in the sociology of education: Towards a reconceptualisation. *British Journal of Sociology of Education* 22, 445–461.
- Moore, R., Arnot, M., Beck, J., and Daniels, H. (eds.) (2005). *Knowledge, Power and Educational Reform: Applying the Sociology of Basil Bernstein*. London: Routledge.
- Moore, R. (2007). *The Sociology of Knowledge and Education*. London: Continuum.
- Morais, A., Fontinhas, F., and Neves, I. P. (1992). Recognition and realisation rules in acquiring school science. *British Journal of Sociology of Education* 15, 243–263.
- Muller, J. (2007). On splitting hairs: Hierarchy, knowledge and the school curriculum. In Christie, F. and Martin, J. R. (eds.) *Language, Knowledge and Pedagogy: Functional Linguistic and Sociological Perspectives*, pp 65–86. London: Continuum.
- Robbins, D. (1991). *The Work of Pierre Bourdieu*. Buckingham, UK: Open University Press.
- Sadovnik, A.R. (ed.) (2001). *Pedagogy and Knowledge: The Sociology of Basil Bernstein*. Norwood, NJ: Ablex.
- Williams, G. (2001). Literacy pedagogy prior to schooling: Relations between social positioning and semantic variation. In Morais, A., Neves, I. P., Davies, B., and Daniels, H. (eds.) *Towards a Sociology of Pedagogy: The Contribution of Basil Bernstein to Research*, pp 17–45. New York: Peter Lang.
- Young, M. F. D. (2003). Durkheim, Vygotsky and the curriculum of the future. *London Review of Education* 1, 100–117.

Curriculum Reform

R Gilbert, The University of Queensland, Brisbane, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

Curriculum reform is not a precise term. While it commonly refers to creating or revising curriculum documents, policies, or practices in order to improve what is taught and learned in educational programs and institutions, the term's limits are less clear. For instance, in his study of the history of curriculum reform in the twentieth century, Kliebard (2002) distinguishes the terms curriculum and pedagogy, yet nonetheless includes pedagogy and assessment in his history, implying either that curriculum is an all-inclusive term, or, more likely, that the three elements of the process of schooling are so intertwined that they must simply be discussed together. Similarly, curriculum reform, school reform, and educational reform are all discussed in Kliebard's history, despite the fact that the latter two terms include aspects of schooling, such as teacher characteristics or school resourcing, which lie outside any viable concept of curriculum.

The influences on curriculum reform are diverse, and its effects have consequences for students, teachers, and the community which go far beyond the curriculum's intended outcomes. In many respects, the broader elements of education reform, such as school organization and administration, teacher education and supply, facilities, technology, and class size, can be seen as means to curriculum reform, reinforcing the centrality of curriculum to the entire educational enterprise. The focus in this discussion is on the reform of the intended curriculum and its manifestations in institutional practice, that is, the specification of what is to be taught and learned in any educational program, and the processes by which these intentions are operationalized.

Curriculum reform occurs at a range of scales depending on the context and the type of reform under consideration. The term can refer to a change in a particular institution, where the case or unit of reform may be a university or school department; it can refer to improvements in the intended outcomes of a particular field, such as a school subject in science or citizenship; or it can encompass the educational programs provided for particular groups, such as differentiations by age level or phase of schooling in the middle years or early childhood curriculum.

However, curriculum reform is also and more commonly used to describe the general process of curriculum change or development at the system level. This general process of curriculum reform is the main focus of this discussion, and the concept of curriculum underlying it is well defined by Moreno:

Curriculum is a socio-historical construction which is expressed through general systems of knowledge characterization and hierarchy; these systems are in turn translated and transformed into legislative and administrative regulations, academic achievement standards, textbooks and teaching aids, and the practice of teaching and learning in classrooms and schools. (Moreno, 2006).

This somewhat linear view needs to be combined with the recognition that curriculum specifications are selections from the culture whose meanings are articulated in social practices. Curriculum knowledge is constantly interpreted and reinterpreted in contexts of conception, production, and implementation, and an understanding of curriculum reform must address this semiotic and indeterminate dimension of the curriculum. This in part accounts for the perennial controversies around curriculum reform, since curriculum decisions involve value judgments about the purposes of schooling and the most-valued forms of knowledge, and how they are related to the cultural practices of different social projects and interests. This recognition highlights the tendentious, or at least aspirational, connotation of the idea of reform, since curriculum reform, as distinct from curriculum change, refers to explicit attempts to improve the curriculum; but just what constitutes improvement is contentious. Curriculum reform in America has been compared with religion, with "periodic crusades, and occasional bouts of zealotry and apostasy" (Ravitch, 2004). Consequently, it is a contested area of educational policy and research.

Contexts and Elements of Curriculum Reform

Curriculum reform seems to be a popular cause, with interested parties from governments to academics to various community groups promoting change in the name of social, economic, or some other form of improvement. Yet, observers have a generally pessimistic view of the experience of curriculum reform, illustrated in McCulloch's claim that "longer term curriculum reform has generally failed to generate educational change of a fundamental kind" (McCulloch, 1998). Some commentators have pointed to curriculum change as an erratic or fortuitous process dominated by fads and pendulum-like swings from one ideology or theory to another (Good *et al.*, 1997; Ravitch, 2001).

Conversely, there is also a considerable and increasing degree of international consensus on the desired outcomes of education and the kinds of curricula which will achieve them. Reform in school systems is influenced by diverse sectoral, national, and international contexts, but governments are also subject to common economic and political influences which lead to convergence of educational policy decisions. Rosenmund analyzed curriculum elements of 102 national reports on education, and identified consistent and increasing emphases on a student-centered approach to learning, interdisciplinary knowledge, life skills and competencies, and values of rights, individualism, and identity. These emphases were seen to be driven by rationales based on concepts of the self-directed, competent, and rational individual, and societal development (Rosenmund, 2006). The desire for international competitiveness is often cited as a cause of this consensus, which is thought to be facilitated by the globalization of markets and information networks as influences on global curriculum change (Waks, 2003).

The importance of the teacher's role in interpreting and implementing the curriculum has been the rationale for reforms advocating a teacher-led approach to curriculum innovation and development, on the grounds that the curriculum that matters most is the one enacted by teachers and experienced by students (Elliott, 1994). Recent research has highlighted the ways in which teachers' involvement in innovation involves personal, social, and emotional responses to change which influence their commitment to reform (Zembylas and Barker, 2007). However, while the school-based curriculum movement and its associated action research approach to curriculum development provided important bases for such an approach, the early hopes for teacher-led curriculum reform underestimated the constraints of their environments (McCulloch, 1998). In the centralized, top-down approach of current standards-based approaches, this is an important issue, for the influence of teachers on the implementation of curriculum reform is well demonstrated.

Two Examples of Curriculum Reform

Many of the features of curriculum reform can be illustrated by reference to two recent examples: the outcomes-based education (OBE) movement, and the standards-based approach to curriculum and assessment. OBE was a reform driven by a desire to focus the curriculum development process more directly on the ultimate goals of schooling, rather than on inputs of course credits, class hours, or subjects taught. The belief was that the contribution of these inputs to the ultimate goals of education in students' later lives was taken for granted, and that school curricula were not succeeding in providing all students with the

learning required for contemporary society. This would be corrected through a process of clarifying the ultimate intended outcomes of schooling and designing the curriculum from them (Glatthorn, 1993; Schwarz and Cavener, 1994). This seems a logical process; however, in questioning established practice and traditional assumptions about curriculum and teaching, such as a normal distribution of student achievement, the universal grading of classes by age, or the subject-based curriculum, the OBE movement challenged convention. It also promoted an approach to curriculum and teaching which was not well established in either research or practice, and which therefore lacked exemplars or experience. After being enthusiastically promoted, OBE has stalled in the face of criticisms of its practicability as well as its ideology (the latter from both right and left).

Critics of OBE argued that its departure from traditional subjects and its focus on instrumental outcomes and contemporary relevance threatened the quality of what was learned. A countermovement developed which took the name standards-based curriculum (SBC), a largely rhetorical label typified by an emphasis on the subject-based curriculum, a closer specification of academic content, and an accountability process established through large-scale centralized assessment (McLaughlin and Shepard, 1995; Wixson *et al.*, 2003). This is an increasingly technical and bureaucratic return to the traditional means-end approach to curriculum.

Along with this development, the concept of alignment has become a key component of curriculum reform, where curriculum, teaching, and assessment are tightly controlled to increase their congruence with each other. If applied in a certain way, where the test specifies the outcomes and teaching programs, and strategies are aligned to it (known as the backloading approach to alignment), then the test effectively becomes the curriculum, and the process of curriculum development itself is undermined (Wraga, 1999).

Criticisms of the standards-based approach look at the emphasis on centralized standardization as a corruption of the curriculum process (Delandshere and Petrosky, 2004). On the other hand, the potential in the standards-based approach for clarifying curriculum emphases has been argued to improve teaching and learning, including for educationally disadvantaged groups. Schoenfeld (2002) points to problems with the approach if it is not introduced in a comprehensive and well-supported manner, but argues that it holds great promise for improving educational outcomes as a result of the clarity and clear guidance it provides. However, clarity of outcomes is a separate matter from the potential corruption of curriculum through an intrusive centralized testing regime. Au (2007) surveyed 49 studies of the effects of high-stakes testing on curriculum, and found that, with some exceptions for certain kinds of tests in some subject areas, the

primary effect of high-stakes testing was to narrow curricular content, to fragment subject knowledge into test-related pieces, and to increase teacher-centered pedagogy.

The examples of outcomes-based and standards-based approaches to curriculum well illustrate the conflicted nature of the curriculum reform process. They point to aspects of reform which, by virtue of their controversial nature, complicate the prospects of success of any reform.

The Efficacy of Curriculum Reform

The record of sustained success of curriculum reform is poor. This is often construed as resistance to change among relevant actors, resulting from cultural, social, organizational, or psychological barriers. Conflicts such as those illustrated above illustrate one aspect of the problem, where the significance of reform for long-held beliefs can be underestimated, or attempts at standardization provoke resistance from those wishing to protect local initiative. The sheer complexity of the curriculum process is another barrier to reform, and seems to have exacerbated the problem of unintended deleterious effects. The literature is replete with accounts of reform policies, which, whether or not they have achieved their specific goals, have caused new problems. The account by Fisher (2007) of post-16 reforms in the United Kingdom is a typical example. Fisher's study shows that reforms intended to broaden the curriculum by establishing parity of esteem for academic and vocational qualifications, and to promote participation and lifelong learning, led in implementation to more didactic approaches to teaching and an emphasis on knowing content at the expense of conceptual development.

McDonald (2003) reviewed three models of curriculum reform:

1. the top-down approach of the teacher-proof curricula of the 1960s and 1970s (which has reemerged in the standards-based approaches of recent times);
2. the bottom-up approach represented in school-based curriculum development and the action research movement; and
3. more recent approaches through collaborative partnerships of schools, professional associations, and other stakeholders.

McDonald suggests that these approaches have few prospects of success since they reflect modernist assumptions about knowledge and the role of students as consumers in a highly regulated institutional order, and that a more postmodern approach is needed. This would require more open and interactive learning contexts in which students have a more active role, with less bureaucratic control.

In similar vein, an OECD (OECD, the Organization for Economic Cooperation and Development) analysis of curriculum reform in a range of countries revealed "rather contradictory conclusions about the systemic reform endeavour," in which recognition of the organic and interconnected nature of the elements of reform contrasted with the "mechanistic, relatively static and linear" thinking of many reform proposals. The report concluded that the dominant metaphor of systemic curriculum reform 'owes too much to images of machines and of self-regulating systems drawn from the industrial production models. These characteristics mean that, despite its strengths, such proposals do not capture the complex nature of educational change' (Centre for Educational Research and Innovation 1998: 30).

The conflicts cited as explanations for the limited success of curriculum reform highlight its political dimensions, which range from internal institutional politics to the politics of national educational policy.

The Politics of Curriculum Reform

Rosenmund (2006) has pointed out that curriculum change cannot simply be seen as a planned technocratic reform to improve the productivity of the education system. It must also be understood as a political measure which reshapes relationships among individuals, institutions, and the nation-state through the selection and organization of school knowledge, and through the sorting and grading of students which result. It follows that the politics of curriculum reform can be seen as driven by broad ideological divisions, or by more specific political, economic, or other vested interests (Good *et al.*, 1997).

Curriculum reform is political not only in the conventional sense, but also in a deeply cultural way. Most visibly, reform will be driven by a motivating political source, with a corresponding agenda and implications for extant power relations (Hopmann, 2003). For instance, centralized initiatives may be resisted as undemocratic, lacking contextual authenticity, or representing a narrow ideological agenda. Local initiatives may be dismissed as parochial or irrelevant to broader central policy imperatives.

However, at its center, curriculum reform is a cultural process with a distinctly cultural politics. For instance, Paechter (2003) points to relationships among gender, power, and subject status as influences on the reform process. She shows that differential subject status is related to differential access to power and influence in the curriculum-change process, and that these relations are not gender neutral. From the perspective of cultural politics, curriculum reform will be successful only if it can link to significant beliefs about such values as inclusiveness, excellence, quality, and entitlement. Since these values are subject to different interpretations and emphasis, there is

potential for conflict – as when a new approach to knowledge is justified in terms of entitlement or progress, it may conflict with traditional interpretations of quality or excellence.

The idea of the educated person is founded in cultural traditions and values, and competing views of canonical knowledge. In contemporary curriculum reform, these cultural tensions are complicated by economic agendas, especially those driven by international competitiveness and global integration. In addition, curriculum reform is deeply implicated in issues of material well-being (involving the allocation of education as an economic good), social welfare (influencing cultural capital, status, and access to effective social participation), and personal identities (created through recognition of and incorporation in cultural patterns of knowledge and belief). The potential for disagreement in curriculum reform over priorities and programs runs deep, and its checkered history is likely to continue unless it is able to generate common understandings of and commitment to reform among its stakeholders. Consequently, to achieve its aspirational intent, curriculum reform must be part of a democratic project as both an end and a means. The democratic ideal must inform the intended outcomes of a curriculum for excellence for all which incorporates a wide range of community aspirations; however, it must be the basis of the reform process as well if it is to confront the cultural politics of the curriculum. The latter requirement is a key challenge to current large-scale approaches to curriculum reform.

Bibliography

- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher* **36**, 258–267.
- Centre for Educational Research and Innovation (1998). *Making the Curriculum Work*. Paris: Organisation for Economic Co-operation and Development.
- Delandshere, G. and Petrosky, A. (2004). Political rationales and ideological stances of the standards-based reform of teacher education in the US. *Teaching and Teacher Education* **20**, 1–15.
- Elliott, J. (1994). The teacher's role in curriculum development: An unresolved issue in English attempts at curriculum reform. *Curriculum Studies* **2**, 43–69.
- Fisher, L. (2007). Pedagogy and the curriculum 2000 reforms at post-16: The 'learn it, forget it' culture? *Curriculum Journal* **18**, 103–114.
- Glatthorn, A. (1993). Outcome-based education: Reform and the curriculum process. *Journal of Curriculum and Supervision* **8**, 354–363.
- Good, T., Clark, S., and Clark, D. (1997). Reform efforts in American schools: Will faddism continue to impede meaningful change? In Biddle, B., Good, T., and Goodson, I. (eds.) *International Handbook of Teachers and Teaching*, pp 1387–1427. Dordrecht: Kluwer.
- Hopmann, S. (2003). On the evaluation of curriculum reforms. *Journal of Curriculum Studies* **35**, 459–478.
- Kliebard, H. (2002). *Changing Course: American Curriculum Reform in the Twentieth Century*. New York: Teachers College Press.
- McCulloch, G. (1998). Curriculum reform, educational change and school improvement. In Hargreaves, A. (ed.) *International Handbook of Educational Change*, pp 1203–1215. Dordrecht: Kluwer.
- McDonald, D. (2003). Curriculum change and the post-modern world: Is the school curriculum-reform movement an anachronism? *Journal of Curriculum Studies* **35**(2), 139–149.
- McLaughlin, M. and Shepard, L. (1995). *Improving Education through Standards-Based Reform*. San Francisco, CA: National Academy of Education.
- Moreno, J. (2006). The dynamics of curriculum design and development: Scenarios for curriculum evolution. In Benavot, A., Braslavsky, C., and Truong, N. (eds.) *School Knowledge in Comparative and Historical Perspective: Changing Curricula in Primary and Secondary Education*, pp 195–209. Hong Kong: Comparative Education Research Centre.
- Paechter, C. (2003). Power/knowledge, gender and curriculum change. *Journal of Educational Change* **4**, 129–148.
- Ravitch, D. (2001). *Left Back: A Century of Battles over School Reforms*. New York: Simon and Schuster.
- Ravitch, D. (2004). Recycling reforms. *Education Next* **Winter**, 34–40.
- Rosenmund, M. (2006). The current discourse on curriculum change: A comparative analysis of national reports on education. In Benavot, A., Braslavsky, C., and Truong, N. (eds.) *School Knowledge in Comparative and Historical Perspective: Changing Curricula in Primary and Secondary Education*, pp 173–194. Hong Kong: Comparative Education Research Centre.
- Schoenfeld, A. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational Researcher* **31**, 13–25.
- Schwarz, G. and Cavener, L. A. (1994). Outcome-based education and curriculum change: Advocacy, practice, and critique. *Journal of Curriculum and Supervision* **9**, 326–338.
- Waks, L. J. (2003). How globalization can cause fundamental curriculum change: An American perspective. *Journal of Educational Change* **4**, 383–418.
- Wixson, K. K., Dutro, E., and Athan, R. G. (2003). The challenge of developing content standards. *Review of Research in Education* **27**, 69–107.
- Wraga, W. G. (1999). The educational and political implications of curriculum alignment and standards-based reform. *Journal of Curriculum and Supervision* **15**, 4–25.
- Zembylas, M. and Barker, H. (2007). Teachers' spaces for coping with change in the context of a reform effort. *Journal of Educational Change* **8**, 235–256.

Further Reading

- Craig, C. (2006). Why is dissemination so difficult? The nature of teacher knowledge and the spread of curriculum reform. *American Educational Research Journal* **43**, 257–293.
- Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record* **106**, 1047–1085.
- Fullan, M. (2009). Large scale reform comes of age. *Journal of Educational Change* **10**, 101–113.
- Gross, B., Booker, T., and Goldhaber, D. (2009). Boosting student achievement: The effect of comprehensive school reform on student achievement. *Educational Evaluation and Policy Analysis* **31**(2), 111–126.
- Leithwood, K., Jantzi, D., and Mascal, B. (2002). A framework for research on large-scale reform. *Journal of Educational Change* **3**, 7–33.
- Muller, J. (2000). *Reclaiming Knowledge: Social Theory, Curriculum and Education Policy*. London: Routledge Falmer.

Early Childhood Curriculum and Developmental Theory

S M-Y Lim, Nanyang Technological University, Singapore

C Genishi, Columbia University, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

The term curriculum refers largely to planned approaches to teaching and learning that are guided by theoretical and philosophical beliefs about the nature of learners and about the kinds of knowledge(s) that should be taught. From this perspective, a curriculum contains specific content knowledge, objectives and goals, teaching procedures, and assessment strategies. From another perspective, curriculum is less planned in advance and seen as developing through interaction between teacher and students. Regardless of the exact definition, it is generally acknowledged that no curriculum is ideologically neutral. Curriculum makers, like other educators, have philosophical, political, and pedagogical beliefs that underlie the curriculum they create (Apple, 2004).

Early Childhood Curriculum: Reflections of Images of Children

In the field of early childhood education (for children birth through age 8), there is a wide range of curricular approaches and models that are guided by different philosophical beliefs about young children as learners and beliefs about the kinds of knowledge that are worth teaching children. The field (as discussed below) may be characterized by multiple, sometimes opposing or overlapping, images of the child as learner:

- child as organism unfolding over time;
- child as scientific subject;
- child as creative player;
- child as active thinker; and
- child as relational and cultural agent.

These varied images of children have shaped curricular models for young children. Some models are closely guided by educational policies that serve specific groups of children, while others attempt to be more inclusively serving children with a full range of learning abilities and interests.

Child Unfolding: Early Conceptualizations of Curriculum

One of the earliest systematic conceptualizations of curriculum in the field was created by Johann Heinrich Pestalozzi, well-known for his school for mixed-age groups of orphans in Yverdon, Switzerland (established 1805).

Influenced by Plato's and Rousseau's images of the innately good child unfolding naturally over time, Pestalozzi founded the doctrine of *Anschauung* (direct, concrete observation), and capitalized on children's spontaneity and adult or peer-guided activity (Weber, 1984). Through object lessons that encouraged children to observe and explore objects collected from the natural environment, Pestalozzi's curriculum acquainted children with form, language, and number classifications and encouraged children to cultivate their own powers of seeing, judging, and reasoning. An underlying curricular principle was to introduce abstract concepts through concrete objects, always progressing gradually and cumulatively from the easier to the more difficult concepts.

It was Friedrich W. Froebel who documented a curriculum for young children in Germany in the 1800s, based on Pestalozzi's work. He also drew from the philosophical thinking of Plato, Rousseau, and Comenius, complementing his own religious beliefs and intuitive understanding of the child as a naturally creative and productive being (Weber, 1984). Froebel's enduring significance to early years of education was through his formulation of the kindergarten (a combination of two German words: *kind* [children] and *garten* [garden]) system with its emphasis on play and its use of gifts (didactic play materials) and occupations (hands-on activities). Singing, dancing, and gardening were included in order to develop the whole child, although much of the curriculum highlighted the learning of mathematical concepts and symbolic relationships through manipulative objects such as blocks, spheres, and cylinders. Froebel invented the kindergarten to prove that children under 7 years of age were capable of obtaining intellectual and emotional skills through structured forms of play and adult guidance.

The success of Froebel's model resulted in compulsory kindergarten education for all children in the Austro-Hungarian Empire by 1872. The gradual spread of the kindergarten movement can be traced across other parts of Europe and the USA in the second half of the 19th century. Its introduction into the USA is accredited to the efforts of educators such as Elizabeth Peabody and Susan Blow. With the establishment of free kindergartens in working-class neighborhoods in the 1870s, advocates of kindergartens in the USA began to suggest that the proper education of these children could eventually alleviate urban poverty.

Child as Subject: Scientific Views and Curricula as Early Intervention

At the turn of the twentieth century, the child study movement began in the USA. Seminal publications in psychology and philosophy that influenced early childhood education included G. Stanley Hall's *Contents of Children's Minds* (1883), Edward L. Thorndike's *Animal Intelligence* (1898), and John Dewey's *The School and Society* (1900). Influenced by scientific observation methods, especially those of Darwin in *The Origin of the Species* (1859), Hall advocated the use of systematic and direct observations of individual young children to inform curriculum. Arnold Gesell expanded on Hall's work by examining the correlation of ages and developmental stages and charting growth gradients of individual children. Such scientific studies of young children in the burgeoning field of developmental psychology appealed to many early childhood educators who subsequently propagated the generalized theoretical belief that there was an optimal period in the early years in which learning would be most effective.

The Froebelian kindergarten curriculum was to remain popular until the 10-year debate (1903–1913), which marked a turning point in early childhood curricular approaches. The debate was between those who supported Susan Blow and Elizabeth Peabody who argued for the continued relevance of the Froebelian kindergarten curriculum in the twentieth century, and progressive educators such as Patty Smith Hill who were influenced by John Dewey and Edward L. Thorndike arguing for curricular revision in order to keep up with the shifting sociopolitical context of society in the USA (Weber, 1984). Curriculum influenced by Dewey's educational philosophy placed social studies at the core of the curriculum since the purpose of education was for life in a democratic society. In contrast, Thorndike's behaviorist theory came to underlie the image of child as subject in the context of curriculum and of standardized testing.

Progressive educators increasingly influenced curriculum from the early 1900s and their work also drew upon a growing body of knowledge contributed by developmental psychologists. An example of one of the earliest eclectic curricula written by an educator of the progressive movement was the *Conduct Curriculum for Kindergarten and First Grade* (1923) by Patty Smith Hill at Teachers College, Columbia University. The curriculum was influenced by the theories of her contemporaries such as Thorndike and Dewey and was a culmination of her years of experimentation in loosening what Hill considered to be the rigidities of the Froebelian curriculum for children. The Conduct Curriculum combined elements of Dewey and Thorndike, respectively: it encouraged both social cooperation and democratic social responsibilities

(a response to the changing American society at the time), and it was a plan for systematic classroom instruction to modify children's thoughts, behaviors, and feelings (i.e., develop desirable habits).

Outside of the USA, in Italy around 1916, Maria Montessori (the first female doctor in her country) began advocating for her educational method for children who were then considered cognitively defective and those who lived in extreme poverty in Rome. Enduring initial dismissal by educators, including progressive educators in the USA, the Montessori approach is ironically still popular around the world among some middle- and upper-class communities, both as an academically focused approach in private schools and as a homeschooling method (Goffin and Wilson, 2001). It has also been modified to suit individual culture's values and needs while maintaining its emphasis on the training of young children's senses for intellectual development.

Child as Creative Player: A Psychodynamic and Developmental View

In contrast to child development study that focused on observable behaviors and outcomes, Freud's psychoanalytic theory encouraged early childhood educators to focus on the inner worlds of young children as biological and sexual beings. Over time however, his theory was modified and neo-Freudian psychodynamic classrooms featured curricula that encouraged children's self-expression and creativity through messy play with materials such as paints, clay, blocks, sand, and water. Lawrence Frank, a prominent American linked to the Laura Spelman Rockefeller Memorial Fund, built on the image of the child as an emotional being and highlighted the importance of free play in the curriculum for healthy socioemotional growth. Influenced by the theories of Erikson and Freud, Frank believed that emotional security and personality adjustment were keys to developing confident, social, yet independent learners. He advocated that teachers recognize individual differences among children, synthesizing psychodynamic work with the theory of Jean Piaget (Weber, 1984).

By the 1930s, in part because of Frank, many child development institutes and their corresponding lab schools had appeared in the USA. The research carried out at these institutes contributed to normative views of children's growth and abilities (e.g., in areas such as motor development, language, and adaptive behavior). Thus, such knowledge of average or normal child development became foundational in teacher education textbooks and guided the development of early childhood curriculum. Teachers planned curricular activities that would suit the assumed needs and abilities of the average child of a certain age group.

Child as Subject Redux: Concern for Economically Poor Children in the USA

In the 1960s in the USA, several scientific studies challenged the dominance of the fixed intelligence theory established earlier in the century by psychologist Alfred Binet, by showing how children's abilities could be enhanced by positive environmental influences (Goffin and Wilson, 2001). These studies eventually supported the inception of the federally funded Head Start program as part of the War on Poverty to provide academic skills and knowledge (e.g., language-based and numeracy skills) to children from low-income populations across the country. Interestingly, the motivation for the program was similar to Montessori's a half century earlier: to counter the impact of poverty through early education. Indeed, a curriculum based on Montessori's methods was one of a number of models that researchers and educators implemented as they tried to demonstrate positive effects of Head Start.

Another curriculum offered by Head Start was the Bereiter–Engelmann model (later developed into the direct instruction model) to help disadvantaged children catch up academically with their middle-class age mates. In this approach, work and play were on opposing ends of the curriculum as children were rewarded by food, tokens, or praise, when they accomplished academic tasks, that is, they got the right answers. Since the Bereiter–Engelmann model aimed to help Head Start children to attain some of the same skills and knowledge as their white, middle-class peers, it has been viewed as a deficit model that disregarded any of the children's own cultural knowledge and experiences. In the early twenty-first century, the direct instruction approach has been foregrounded again in the USA, due to the impact of the legislation known as No Child Left Behind. Funds associated with that legislation support scientifically based curricula. Thus, highly structured lessons that are clearly taught and whose outcomes are easily measured remain popular, especially in reading instruction, and in some special education approaches, for example, the reading curriculum Open Court and the applied behavioral analysis approach to special education.

Child as Active Thinker: Piagetian Theory and Its Curricular Offshoots

The 1960s also gave rise to curriculum models inspired by Jean Piaget's cognitive developmental theory that children's cognition develops through the sensorimotor, pre-operational, concrete, and formal operational stages. For those persuaded by psychodynamic and developmental approaches, Piaget's comprehensive theory was a better fit than behaviorist approaches to improving the education

of young children. A number of influential programs were based on Piagetian theory, the best known of which is the High/Scope cognitively oriented curriculum, initiated as an experimental program by David Weikart, offered as a Head Start model in the 1960s, and now incorporated into the Head Start curriculum. The curriculum encouraged children's active learning and independent knowledge construction. Other curricula, such as the Constance Kamii-Rheta DeVries constructivist approach and George Forman's constructivist approach, demonstrate different interpretations and applications of Piagetian theory. Constructivist activities capitalize on children's interests and provide physical experiences with objects, logico-mathematical experiences, and interactivity (e.g., experimentation with peers).

Developmentally Appropriate Practice and Its Critics

By the 1980s, another great debate occurred in the early childhood education field. This time, some leaders in the field spoke out against behaviorally oriented curriculum models such as direct instruction. Through the National Association of the Education of Young Children (NAEYC), these scholars published a position statement on developmentally appropriate practice (DAP) in 1986 (expanded in 1987) for standardizing the diverse range of practices found in an array of early childhood education provision (e.g., nursery schools, child care, Head Start, and kindergartens). The guidelines proved to be unexpectedly successful at the time as a much needed tool to provide uniformly high-quality services for young children and families. It was supported by a few longitudinal studies that verified the positive outcomes and benefits of High/Scope, a developmentally oriented curriculum. Such scientific support appealed to policymakers and public school leaders. The concept of DAP became a trendy phrase that was eventually taken up by some proponents of the direct instruction approach in the primary grades. The DAP guidelines were revised a decade after its first publication to accommodate the newer sociocultural perspective in developmental theory.

Just as there was a major debate that created a turning point in Froebelian curriculum in the early 1900s, a significant debate in the last decade of the twentieth century centered around DAP (see Charlesworth, 1998; Lubeck, 1998). Cross-cultural studies have also shown that adults in different cultures could have different views of what young children should learn or behave (see Tobin *et al.*, 1991; Tobin *et al.*, 2009). These studies point to the need to expand notions of what is meant by DAP, rather than to fixate on a search for the one best or most appropriate curriculum.

The Child as Relational and Cultural Agent

The Influence of Sociocultural Perspectives

Critics of Piagetian theory have argued that children do not merely construct knowledge in isolation from their social and cultural contexts. In the 1990s, proponents of the Russian psychologist, Lev Vygotsky, provoked educators to consider a sociocultural view of teaching and learning. Central to Vygotsky's theory is the belief that human development is inseparable from social and cultural activities. Thus, children's development of higher mental processes involves learning to use the tools of culture, such as language and symbols, through the guidance of other people who are more skilled in the use of these tools. Other American psychologists who became aligned with this view included Jerome Bruner and Barbara Rogoff. The sociocultural view led teachers to use tools and strategies to cultivate the latent abilities that lie within each child's zone of proximal development (ZPD). Such a curriculum includes extended blocks of time for free play, thereby encouraging children to advance their play behaviors to utilize symbolic representations and actions, multiple and complex themes and roles (e.g., doctor theme merged with a restaurant theme) through the use of language to create pretend scenarios (Bodrova and Leong, 2007). Props and play materials in the classroom are to be iconic and simple rather than realistic. The curriculum should also allow children to create and review their plans for play, to problem solve, and to learn from one another as a way to extend their individual ZPD. According to the Vygotskian approach, learning is believed to lead development, contrary to the Piagetian notion that development is a prerequisite for learning to occur. These differences in perspective have inevitably shaped curricular approaches.

In the early 1980s, some scholars from the USA began studying programs for children from infancy through age 6 in the small Italian city of Reggio Emilia. Although the Italian educators do not consider their curriculum to be a model, many US educators have studied and tried to implement or adapt the approach. Although the approach is often characterized as focusing on art, founders emphasize representational art as only one of many modes of symbolic representation, which they refer to as languages (Edwards *et al.*, 1998). Integrated learning through extensive project work is impelled by both teacher planning and children's interests in their environment. Teachers' respect for children's strengths, careful documentation of children's work, and collaborative relationships with families led some early educators to cite Reggio Emilia as an exemplar of blending developmental and sociocultural approaches, embedded in both Piagetian and Vygotskian theories. The approach maintains an international following, as educators visit, and are inspired by, the program's centers and schools.

The sociocultural emphasis of the Vygotskian approach has also been used to further the work of multicultural educators. Various multicultural curricula aim to foster an understanding and appreciation of cultural diversity, and to promote intercultural relations that may lead to constructive social action in communities. Inherent in multicultural education is the notion that there are diverse ways of constructing and acquiring knowledge and that one's cultural heritage, home language, and personal experiences are legitimate funds of knowledge (Ramsey and Williams, 2003). Thus, the notion of school knowledge is broadened beyond white, middle-class European-American cultures and histories, and children are viewed as active and interactive agents who not only learn but are capable of creating change. Examples of established multicultural curriculum in the USA include the Anti-bias curriculum (Derman-Sparks and ABC Taskforce, 1989). Outside the USA in New Zealand, the national Te Whariki curriculum (Ministry of Education, New Zealand, 1996) is an example of how policymakers and educators are determined to encourage bicultural early childhood curricular practices by empowering Maori, Pacific Islander, and white European-origin peoples.

Summing Up: Developmental Theories and Their Influence

Looking back historically, we have noted that educators in the field of early childhood see the roots of developmental psychology in the ideas of Pestalozzi and Froebel, who believed that children's abilities would unfold over time. Pestalozzi introduced careful observation, and both thinkers specified curricular choices attuned to children's interests and needs. Following the introduction of the kindergarten in Europe and the USA, the discourse of developmental psychology was built on the successes of nineteenth-century physical sciences inspired by the Darwinian revolution. Like their predecessors, developmental theorists of the twentieth century such as Piaget constructed universalist trajectories of human growth as dynamic, systematic, and irreversible processes.

Developmental psychology's prominence in early childhood curriculum, especially in the USA, was traced back to the 1920s and 1930s when a number of child development institutes were established around the country. These centers served to provide enrichment programs through nursery schools for young children from both middle- and low-income homes and gave developmental psychologists opportunities to further examine the role of the classroom environment, including curriculum, on affecting individual intelligence. After the inception of Head Start in the USA in 1964, part of the work of developmental psychologists was to assess the effectiveness of

early intervention programs for young children from disadvantaged backgrounds.

In the early twenty-first century, some theoretical and curricular stances seem to have cycled back to the 1960s as the discourse of developmental psychology gives way once again to a deficit orientation toward child learners. The legislation known as the No Child Left Behind Act has revived the child as subject, who learns best through scientifically based behaviorist curricula. This image of the child has persisted for about a century and is not likely to fade away. Interestingly, however, early childhood professionals such as those on both sides of the recent debate about DAP have resisted this image and consistently represented curricula as broad enough to include child choice and agency. In other words, they have opted for a definition of curriculum that is interactive, dependent on the knowledge of both children and teachers and not on published curricular scripts.

Postmodernist Critique: Limitations of Developmental Theory and a View toward the Future

Within an international context, many societies are heading toward global capitalism, and are faced with much tension created by the clash of global and local knowledge(s) and cultural values. For some scholars, educating citizens in an increasingly globalized age is not simply about training a skilled workforce capable of creating economic growth and technological advancement, it is also about educating citizens who are intellectually versatile, open minded, socially responsible, and who want to achieve greater global tolerance and understanding of differences in cultures and values (e.g., Burbules and Torres, 2000; Ryan and Grieshaber, 2005). Thus, many have argued for a reconsideration of early childhood educators' reliance on child development theories in curriculum design because they limit one's view of what young children ought to learn.

Within the field of developmental psychology itself, the assumption about the predetermined and cumulative nature of human development was critiqued as early as 1978 in the Presidential Address to the American Psychological Association's Division of Developmental Psychology, *The American Child and Other Cultural Inventions* (Kessen, 1983). Related critiques were taken up by educators and psychologists in the 1980s and 1990s (e.g., Burman, 1994). They urged a reconceptualization of early childhood curriculum so that there would be less focus on individual psychological differences and more on diversity within and across cultures (e.g., Kessler and Swadener, 1992; Mallory and New, 1994).

Postmodernist critics have continued to challenge taken-for-granted essentialist and scientifically biased assumptions about human development and learning and its view of the child as individual and self-contained

(Ryan and Grieshaber, 2005). Indeed, curricula that have been categorized as multicultural may offer the scope needed to encompass the diversity of views of children and curricula globally. Some scholars argue that the most flexible curricular approaches combine a constructivist theory of learning and development with a multicultural and activist orientation toward social justice. With a view of children as relational and cultural agents, adults teach about issues of equity and social justice and then encourage children to act to change communities (e.g., to create equitable rules for boys and girls within the classroom; start a recycling center in the school). The range of issues addressed could extend beyond race/ethnicity, class, and gender to include religion, spirituality, dis/ability, sexual orientation, technology and media, consumerism, environmental degradation, and economic exploitation in global communities. In short, there is an increasing need for more inclusive curriculum for all children.

Studies have shown that high-quality education for the early years is not entirely dependent on the curriculum model that is used, despite the continued preoccupation with such models. Instead, teacher quality is the most important factor (Goffin and Wilson, 2001). The way forward, then, is to stop searching for the one best model – to pause and evaluate the images of children that underlie value-laden teaching practices. From a postmodern perspective, socially just curricula (e.g., Grieshaber and Cannella, 2001; MacNaughton, 2004; Silin, 1995) require teachers to take the time to understand and situate individual children's cultures and the knowledge(s) to inform their curricular choices (i.e., content, perspectives, collaboration, strategies, and assessment). School leaders and policymakers, too, have the responsibility to encourage teachers to be knowledgeable enough to integrate a variety of teaching approaches and strategies instead of being implementers of fixed or scripted curriculum models.

In sum, early childhood curriculum will continue to be influenced by divergent theories of learning and development and contrasting images of young children. Thus, children and families around the globe will rely on the professional judgment of teachers and school administrators to choose curricula that are flexible enough for children's unique heritages and the funds of knowledge each child brings into the classroom. Maintaining such flexibility is no small task; but it is a challenge that must be met, as the increasing diversity within groups of children locally and globally make identification of one best curriculum for all an impossible and undesirable task.

See also: Development and Implementation of Early Learning Standards in the United States; Early Childhood in Post-Modern Cultures: Thoughts and Some Concerns; Gender Issues in Early Childhood Education and Care; Teaching in Early Childhood Centers Instructional Methods and Child Outcomes.

Bibliography

- Apple, M. (2004). *Ideology and Curriculum*, 3rd edn. New York: Routledge.
- Bodrova, E. and Leong, D. J. (2007). *Tools of the Mind: The Vygotskian Approach to Early Childhood Education*. Upper Saddle River, NJ: Pearson Education.
- Burbules, N. C. and Torres, C. A. (eds.) (2000). *Globalization and Education: Critical Perspectives*. New York: Routledge.
- Burman, E. (1994). *Deconstructing Developmental Psychology*. London: Routledge.
- Charlesworth, R. (1998). Developmentally appropriate practice is for everyone. *Childhood Education* **74**, 274–282.
- Derman-Sparks, L. and ABC Task Force. (1989). *Anti-Bias Curriculum: Tools for Empowering Young Children*. Washington, DC: National Association for the Education of Young Children.
- Dewey, J. (1900, 1902/1990). *The School and Society; and, The Child and the Curriculum: A Centennial Edition with a "Lost Essay" by John Dewey and a New Introduction by Philip W. Jackson*, (originally published 1900 and 1902 respectively). Chicago, IL: University of Chicago Press.
- Edwards, C., Gandini, L., and Forman, G. (eds.) (1998). *The Hundred Languages of Children: The Reggio Emilia approach – Advanced Reflections*, 2nd edn. Greenwich, CT: Ablex.
- Goffin, S. and Wilson, C. (2001). *Curriculum Models and Early Childhood Education: Appraising the Relationship*, 2nd edn. Englewood, NJ: Prentice-Hall.
- Grieshaber, S. and Cannella, G. S. (eds.) (2001). *Embracing Identities in Early Childhood Education: Diversity and Possibilities*. New York: Teachers College Press.
- Kessen, W. (1983). The child and other cultural inventions. In Kessel, F. S. and Siegel, A. W. (eds.) *The Child and Other Cultural Inventions*, pp 26–47. New York: Praeger.
- Kessler, S. and Swadener, B. B. (eds.) (1992). *Reconceptualizing the Early Childhood Curriculum: Beginning the Dialogue*. New York: Teachers College Press.
- Lubeck, S. (1998). Is developmentally appropriate practice for everyone? *Childhood Education* **74**, 283–298.
- MacNaughton, G. (2004). Exploring critical constructivist perspectives on children's learning. In Anning, A., Cullen, J., and Fleer, M. (eds.) *Early Childhood Education: Society and Culture*, pp 43–56. London: Sage.
- Mallory, B. L. and New, R. S. (eds.) (1994). *Diversity and Developmentally Appropriate Practices: Challenges for Early Childhood Education*. New York: Teachers College.
- Ministry of Education, New Zealand (1996). *Te Whariki: Early Childhood Curriculum*. Wellington, NZ: Ministry of Education, New Zealand.
- Ramsey, P. G. and Williams, L. R. (2003). *Multicultural Education: A Sourcebook*, 2nd edn. New York: Routledge.
- Ryan, S. and Grieshaber, S. (eds.) (2005). *Practical Transformations and Transformational Practices: Globalization, Postmodernism, and Early Childhood Education*. Oxford, UK: Elsevier.
- Silin, J. G. (1995). *Sex, Death, and the Education of Children: Our Passion for Ignorance in the Age of AIDS*. New York: Teachers College Press.
- Tobin, J. J., Wu, D. Y. H., and Davidson, D. H. (1991). *Preschool in Three Cultures: Japan, China, and the United States*. New Haven, CT: Yale University Press.
- Tobin, J., Hsueh, Y., and Karasawa, M. (2009). *Preschool in Three Cultures Revisited: China, Japan, and the United States*. USA: The University of Chicago Press.
- Weber, E. (1984). *Ideas Influencing Early Childhood Education: A Theoretical Analysis*. New York: Teachers College Press.

Further Reading

- Bereiter, C. and Engelman, S. (1966). *Teaching Disadvantaged Children in the Preschool*. Englewood, NJ: prentice-Hall.
- Brosterman, N. (1997). *Inventing Kindergarten*. New York: Harry N. Abrams.
- Consortium for Longitudinal Studies (1983). *As the Twig is Bent: Lasting Effects of Preschool Programs*. Hillsdale, NJ: Erlbaum.
- Cuffaro, H. K. (1995). *Experimenting with the World: John Dewey and the Early Childhood Classroom*. New York: Teachers College Press.
- Greenberg, P. (1987). Lucy Sprague Mitchell: A major missing link between early childhood education in the 1980s and progressive education in the 1890s–1930s. *Young Children* **42**, 70–84.
- Katz, L. G. and Chard, S. C. (2000). *Engaging Children's Minds: The Project Approach*, 2nd edn. Greenwich, CT: Ablex.
- Kessler, S. A. (1991). Early childhood education as development: Critique of the metaphor. *Early Education and Development* **2**, 137–152.
- Kohlberg, L. and Mayer, R. (1972). Development as the aim of education. *Harvard Educational Review* **42**, 449–496.
- Montessori, M. (1995). *The Absorbent Mind*. New York: Henry Holt and Company.
- Rogoff, B. (2003). *The Cultural Nature of Human Development*. Oxford, UK: Oxford University Press.
- Seefeldt, C. (ed.) (1999). *The Early Childhood Curriculum: Current Findings in Theory and Practice*, 3rd edn. New York: Teachers College Press.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Weikart, D. P. (1981). Effects of different curricula in early childhood intervention. *Educational Evaluation and Policy Analysis* **3**, 25–35.
- Zigler, E. and Valentine, E. (1979). *Project Head Start: A Legacy of the War on Poverty*. New York: Free Press.

Relevant Website

<http://www.ed.gov> – No Child Left Behind Act of 2001. Public Law. No. 107–110 115 Statistics. 1425 (2002).

This page intentionally left blank

D

DEMOGRAPHY AND SOCIAL CHANGE

Contents

An Evaluation Framework for E-Learning Effectiveness in the Arab World

Globalization and Social Justice in Higher Education

Racial Inequality and the Economic Payoffs of Higher Education

School Health Education

An Evaluation Framework for E-Learning Effectiveness in the Arab World

W AbuSneineh and M Zairi, University of Bradford, Bradford, UK

© 2010 Elsevier Ltd. All rights reserved.

Despite the valuable contribution of each of those studies, a holistic framework for evaluating the quality of e-learning could not be found (AbuSneineh and Zairi, 2007a, b). Developed frameworks either addressed the quality of e-learning programs with an emphasis on the classroom environment only (see, e.g., Cohen and Ellis, 2004) or overlooked some important aspects – such as the impact of the institution on the quality of e-learning programs (see, e.g., Cashion and Palmieri, 2002).

This article is based on a study which attempted to fill the gap, identified in the literature, by developing a comprehensive framework that addresses the critical factors for evaluating quality of e-learning programs from an all-encompassing perspective. The study explores critical factors for evaluating the quality of e-learning programs based on extensive literature review. As a result, 26 factors were identified and categorized under five dimensions, namely, pedagogy, faculty, technology, learning support, and institution. Each represents an important area in the e-learning environment. The 26 factors have been validated and tested at The eTQM College in Dubai – in an Arab cultural context.

Although the sample size of the surveyed learners and faculty is relatively small, the overall findings of this study conform – to a large extent – with what has been reported in the literature in this field and, therefore, support the appropriateness of the implementation of such framework in the e-TQM College.

Introduction

At present, higher education is experiencing new challenges driven by many emerging trends – among which is the introduction of means of online learning. Accordingly, many universities – all over the world – are adopting different moves to cope with the new-era requirements; some universities were established as purely online learning providers, other conventional universities launched new online programs, and some introduced online learning components to enhance some of their already existing programs.

e-Learning redefines the teaching/learning processes and the overall learning environment. It directs the old university to a new set of practices by introducing significant changes in the roles of its main key players – learners, faculty members, and the institution as an entity (Mendenhall, 2001).

The many changes e-learning brings to higher education – accompanied by its vast distribution as an accepted learning-delivery mode – has always been questioned by opponents who tie this type of learning to poor quality. Therefore, studying the different aspects of evaluating the quality of e-learning has been arising as an issue worth researching and enhancing (Zhao, 2003; Kistan, 2005; Wirth, 2005).

This article includes a thorough review for the literature to explore what has been recognized by the researchers and practitioners, of this field, as key factors that affect the quality of e-learning programs and institutions and,

therefore, should be used to evaluate this type of learning. To do so, many studies and researches were taken into account – whether they tackled the issue from a single aspect or went further to address multiple aspects.

Critical Factors

The evaluation model is based on a set of critical factors found to have significant impact on the quality of e-learning. Based on an intensive literature scanning and review, the authors constructed a framework that consists of 26 factors categorized under five dimensions; pedagogy, faculty, technology, learner support, and institution. The dimensions are highly interrelated and do not function independently. The following sections present a brief description of each of the framework dimensions.

Pedagogy

The reason why pedagogy always comes first when the quality of e-learning is discussed is that e-learning is about learning in the first place. Pedagogical issues are one of the first things educators address when they plan to teach.

To start with, pedagogy is a term which originated from Greece and means the art of teaching children (Ciavarelli, 2002). The definition – in the context of higher education – means, generally, the science or art of teaching. Sonwalker (2002) defines pedagogy as “the method by which educational content is exposed to learners.”

To understand the significance of this dimension, it is imperative to introduce the issues that program pedagogy usually addresses:

1. The learner and his characteristics.
2. The subject to be taught.
3. The organization of the content.
4. The strategy for content presentation.
5. The sequence of the topics in the course.
6. The delivery strategy of the content.
7. The assessment of the taught subject and its fulfillment of the teaching objectives (Ciavarelli, 2002).

Pedagogy, therefore, relates to four main components: faculty member, learner, tasks, or exercises that need to be completed with the help of faculty members, and, eventually, knowledge disseminated and gained by the learner (Boettcher, 1999).

Interestingly, the aforementioned pedagogical issues do not limit the learning and teaching processes to any specific time, place, or medium.

In order to understand, briefly, what is meant by learning styles – a major component of the pedagogy – the following section addresses this component and its most familiar theories and principles.

Learning Styles and Their Impact on the Quality of e-Learning

The behavioral, cognitive, and constructive learning theories are referred to as the main principal schools of learning styles (see, e.g., Ally, 2004; Ciavarelli, 2002; Boettcher, 1999).

The Behaviorist school of thought – influenced by Thorndike, Pavlov, and Skinner – hypothesizes that learning takes place as a change in behavior, caused by external stimuli in the environment (Skinner, 1974).

Ciavarelli (2002) explains that it is the change in behavior that bears on its ability to indicate whether or not learners have learned according to the Behaviorists. This is done through shaping learner's response to a set of stimuli – by focusing on skills development through practical means.

The behaviorists' learning theory was criticized by many educators who saw the learning process as being about more than a change in behavior, and that learning consists of elements that cannot be observed. As a result, there was an evolution of a new learning theory – the cognitive school of learning (Ally, 2004).

Craik and Tulving (1975) explain that the cognitive school of learning, on the other hand, relies on the processing capacity of learners and the depth and maturity of the processing. It claims that learning takes place through the use of learners' memory, motivation, thinking, and reflection.

The cognitive theorists believe the instructor is responsible for the information-input process. However, the learner is also an active participant in the learning process, where he/she is responsible to process information (input), store it, and retrieve it (Boettcher, 1999).

The third learning style, discussed in this section, is Constructivism. Cooper (1993) explains that, according to constructivists, learning takes place through learners' observation, processing, and interpretation. This theory is well-known for its argument that learners contextualize the information and personalize it into their personal knowledge and reality.

Learners – according to the constructivists – play a more active role, according to this theory, where they act as the creator and processor of the learning experience. This is why this theory emphasizes the learner-to-learner interaction which is claimed to enrich the learning process (Boettcher, 1999). The teacher – on the other hand – is the facilitator of the learning process, who guides learners in their learning experience, encouraging them to engage in active learning situations whether individually or in groups.

Appropriateness and Relevancy of Learning Styles to Course Goals and Learners' Needs

Although the aforementioned learning styles are the most popular so far, new learning styles are evolving, at present,

with the progress of research in this area. Educators, therefore, have to consider the appropriateness of the learning styles they use for course content to meeting learning goals of the e-learning course they teach (Ally, 2004).

At present, the question becomes: Are the traditionally used learning styles relevant to the pedagogical needs of faculty members and learners? Are they capable of delivering quality learning under today's new conditions where learning environment is more liberating and learner-centered and when more specialized resources are accessible to learners than ever before? (Beaumaster and Long, 2002).

The quality of e-learning is highly dependent on the course content. Learning styles cannot be taken without questioning anymore. Instead, faculty members have to adjust course content by selecting appropriate learning styles based on their conformance to achieve course goals and learners' and needs (Coffield *et al.*, 2004; Gillani, 2000).

Therefore, many researchers have highlighted that the continual review of programs pedagogy, courses' content, and learning styles is a very important tool that helps educators judge the relevancy of learning styles to achieve planned objectives. Bates and Poole (2003) even argued that the appropriateness of using learning content to program goals and objectives should be assessed by educators who were not involved in the design of the pedagogy to guarantee objectivity.

Howell *et al.* (2004) believed that the review of courses' content and learning styles does not only improve the quality of courses, but also helps faculty members to experiment with different approaches to the teaching/learning process and to establish faculty communities that can produce more relevant learning to learners' needs. This is why this factor is usually discussed in association with faculty innovation and engagement.

Clarity of Course Goals

Not surprisingly, there has been consensus in the literature of pedagogy design that, in order to assure the quality of the teaching/learning process, designing a course should start with explicit outcomes that identify the expectations from learners and faculty from the very beginning (e.g., Stone *et al.*, 2001; Bates and Poole, 2003). Learning outcomes – which usually drive the effectiveness assessment process of courses – should also be challenging and linked directly to the institution's overall goals and objectives (Mayer *et al.*, 2001; Watkins and Kaufman, 2003).

Diversity of Learning Styles to Meet Course Goals and Learners' Needs

Ally (2004) claims that with the overlap in existing learning styles and the progressing ones, no styles can be used in

isolation from others. Quality of pedagogy and course content depends, therefore, on using a combination of all appropriate styles – each one to serve a specific course goal or requirement and also learners' needs and wants. The same argument was also advocated by Sonwalker (2002) as discussed below.

In this context, Hughes (2004) argued that the comprehension of learners' needs – such as their readiness to e-learning, familiarity, accessibility and competency of technology, individualized learning styles, individualized career and educational goals, language, and cultural issues – are among the aspects that should be addressed when designing courses and learner-support services.

Pedagogy Is Not Only about Learning Styles

The development of e-learning, introduction of associated infrastructure, and changes hit the interface of education; all led to a reinvention in the pedagogy concepts.

At present, classrooms are being replaced or significantly supplemented with asynchronous and synchronous Web-based and Web-supported learning environments; therefore, methods by which educational content is disseminated to learners are also changing (Sonwalker, 2001).

In this study, components of the pedagogy dimension adopt Sonwalker's (2001) classification that considers pedagogy does not consist of learning styles only; it is a function of another two components – the multimedia and the interaction. The three components form a three-dimensional (3-D) cube – where each axis represents one of those dimensions.

The *x*-axis of the pedagogical learning cube represents the different learning styles, while the *y*-axis represents the multimedia elements used as part of the course content – such as the text, graphics, audio, video, animation, and simulation – and the *z*-axis represents the interaction component which refers to the interaction among faculty and learner, learners among themselves, and learners and content material using tools such as system feedback, adaptive remediation and revision, e-mail exchange, discussion groups, and bulletin boards (Sonwalker, 2002).

Sonwalker (2002) argued that the effectiveness of the pedagogy to meet certain goals and objectives increases with the diversity and the use of a wider range of learning styles, a wider range of multimedia elements, and interaction as well.

Many believe that engaging learners in active learning, allowing them to inquire about what they are learning, reflecting on it, and relating it to personal experiences and real-world issues is the only way to achieve learning outcomes (Laurillard, 1993). Diversity of learning styles, multimedia, and interaction will realize such learning experiences.

Similarly, Graham *et al.* (2001) argued that the use of learner-centric different styles of pedagogy has a direct impact on the achievement of learning outcomes and hence the quality of e-learning programs. Learner-centric pedagogy refers to the use of different learning styles and objects such as projects, challenging tasks, and sample cases to accommodate for the different needs and talents of learners.

Interaction

Interaction is one of the most widely mentioned factors in the literature, in relation to its impact on the quality of e-learning programs (Moore, 2002).

Diversity of Interaction

Simpson and Galbo (1986) defined interaction as

... behavior in which individuals and groups act upon each other. The essential characteristic is reciprocity in actions and responses in an infinite variety of relationships: verbal and nonverbal, conscious and non-conscious, enduring and casual. Interaction is seen as a continually emerging process, as communication in its most inclusive sense.

Interaction in the e-learning environment compensates for the loss of direct and face-to-face contact between learners and their faculty member or colleagues.

Moore (2002) has identified three types of interaction:

1. interaction between faculty and learners – which is considered to be the most vital component that creates an effective learning environment upon which the other two types depend and are facilitated through;
2. interaction between learners among themselves; and
3. interaction between the learner and content material.

Chickering and Gamson (1987) believe that learner motivation, involvement in the learning process, and, hence, the quality of e-learning are affected, primarily, by the learner–faculty interaction – whether inside or outside the class. Faculty's constant communication with learners keeps them engaged actively in learning. The authors also believe learning can be significantly enhanced when group interaction is taking place among the learners themselves – where involvement in active learning is increased, learners' thinking is improved, and knowledge understanding is deepened.

Prompt Responsiveness of Faculty Feedback to Learners

However, quality of interaction – rather than its quantity – is what really matters; Graham *et al.* (2001), for instance, has

identified the promptness of feedback from faculty members to learners as a critical factor that affects the quality of learning. Timely and prompt feedback helps the learner understand the learning content and also helps individualize the learning process (see also Rovai and Barnum, 2003; Roblyer and Ekhaml, 2000).

Vrasidas and McIsaac (2000) stipulate promptness of feedback from faculty members to learners – whether in the form of comments on assignments, homework, class contributions, or overall progress – as one of the critical factors that affect the quality of e-learning and compensate for the loss of the availability of projects, group discussion, or brainstorming activities linked to the desired outcomes of an e-learning program that has a positive impact on the effectiveness of learning.

Relevance of Interaction to Course Goals

Using an innovative rubric to evaluate the quality of interaction, Roblyer and Ekhaml (2000) also emphasized the importance of interaction relevancy to course goals and outcomes – whether during faculty–learner feedback or through learners' projects. The authors considered this link as one of the main factors affecting the quality of learning in the e-learning environment.

Flexibility of Interaction Time and Place

Offering flexibility for time and place of interaction has also been highlighted, by the literature, as a critical factor that affects learners' satisfaction and, therefore, the quality of the e-learning experience.

Through an empirical study conducted by Cashion and Palmieri (2002), it was found that flexibility was rated topmost among the factors affecting learners' satisfaction. This result is not surprising due to the fact that many e-learners usually have other obligations, such as full-time employment, families, etc.

Multimedia

Diversity of Multimedia Components

Similar to the other two components of the pedagogy – that is, learning styles and interaction – the diversity of multimedia components is one factor that affects the quality of the pedagogy, hence the quality of e-learning programs (Sonwalker, 2002).

Relevancy of Multimedia Components to Course Goals

Research on the quality of e-learning programs outlined that effective multimedia components are those which are highly

in conformance with – and explicitly linked to – course learning goals and objectives (Roblyer and Ekhaml, 2000).

Graham *et al.* (2000) have also identified four main technical components pertaining to how multimedia affects learners' and hence quality of e-learning programs; these are:

1. *Consistency of page layout and design.* This is necessary for allowing learners to recognize the graphic language of the interface, use and re-use their knowledge in different applications, and retrieve information easily.
2. *Clear organization and accessibility of information.* This allows learners to accomplish their tasks effectively when understanding the relationship among the media elements, without being distracted by irrelevant features.

In this context, accessibility is identified as "arrangements giving people the opportunity to study in an institution, in particular, the requirement to offer learners from unconventional backgrounds, those with traditional low take-up of Higher Education, or those with disabilities, the maximum opportunity to access education courses" (LTSN, 2004).

From the above definition, accessibility of multimedia content considers disabled learners and their rights for an equal use of content material (Foley and Regan, 2002).

3. *Navigation consistency and ease of use; if not consistent.* Navigation can place a heavy mental load on learners and, therefore, hinders their learning process.
4. *Attractiveness of design and graphics.* In this, learners – who are studying multimedia elements or using the learning interface – can be encouraged to spend longer hours by attractive design and graphics.

In conclusion, elements of the pedagogy act as a jigsaw – the quality of which can be evaluated by the factors outlined in **Figure 1**.



Figure 1 Critical factors for evaluating quality of e-learning programs.

Faculty

Faculty Competency

The amount of literature that addresses this aspect is enormous, where a broad consensus on the positive direct link between the competency of faculty members and the quality of e-learning programs seems to exist (Howell *et al.*, 2004; Caplan, 2004; Frydenberg, 2002; Bower, 2001; Graham *et al.*, 2000; Fredericksen, 2000; Hartman, 2000; Clay, 1999).

Although the competency of people is a general critical factor that has been emphasized heavily by different quality-management standards such as the ISO 9000 or excellence models such as the EFQM Business Excellence Model, yet – and perhaps – due to the rapid introduction of e-learning, institutions might assign online courses to faculty members without proper training, causing many painful experiences whether to the faculty, learners, or the institution itself.

Educators who seem to be very confident with their skills and abilities to deliver quality teaching are, sometimes, hindered by the new infrastructure, new skills required, and their new assumed role; therefore, development programs for faculty members are deemed to be necessary for quality e-learning programs.

According to a study conducted by SUNY (Fredericksen, 2000), faculty competency does not only impact the quality of the entire learning experience by affecting many aspects – starting from course designs to delivery practices. Faculty competency also affects the overall satisfaction of the faculty themselves. Faculty satisfaction has been identified as a critical measure that indicates the quality of e-learning programs (Bourne and Moore, 2003; Hartman *et al.*, 2000).

Faculty competency, according to Fredericksen's study, refers to many aspects, among which are:

1. Awareness of online teaching and learning, in general.
2. Awareness of the new role of faculty and also learners.
3. Required technical knowledge of the Internet and technology to work with and administer applications necessary during the teaching/learning process, during the design of courses, or following completion of courses.
4. Awareness of guidelines and resources related to courses development and planning.
5. Information of all support services and facilities offered to faculty members.
6. Familiarity with the program taught.
7. Familiarity with the policies and procedures pertaining to e-learning, such as copyrights, workload, etc.
8. Access to previous experiences of other faculty members.

Clay (1999) also considered the need to educate faculty members on how their involvement in e-learning programs fits with the overall strategy and mission of the institution. He also added that discussing faculty members' issues and concerns with regard to this type

of learning is also very important to increase levels of awareness.

The success of e-learning courses depends on the type of development offered to faculty members. The following three areas of development have been considered, by Caplan (2004), as the most crucial requirements due to their impact on the success of e-learning programs:

Pedagogical proficiency

This is where faculty members are introduced to – and experience – the classroom environment, requirements of course design, and delivery.

Administrative skills

An area in which methods of classroom management and identification of the support needed and offered by the institution are among the topics addressed in this area.

Technical skills

This includes learning basic and more advanced technical skills as needed and learning more about the technology, its benefits, and barriers.

In the same context, Bower (2001) suggested that faculty-development workshops should introduce faculty to varied e-learning pedagogical approaches, allowing them to explore the online environment and to take informed decisions during the design and the delivery of courses.

In addition to increasing faculty members' satisfaction, faculty development affects the quality of online learning by (Hartman *et al.*, 2000):

1. Encouraging experimental learning and sharing of techniques, tools, and best practices among faculty members.
2. Transforming faculty members to lifelong learners who understand the needs and wants of their learners.
3. Creating meaningful discussions with regard to the teaching/learning process among faculty communities, and allowing evaluation of successes and failures.

Improving competency of faculty members in e-learning institutions has high impact on the pedagogical critical factors discussed in the previous section. According to Hartman *et al.* (2000), faculty-development programs incite their innovativeness to create pedagogical styles and environments that embrace active learning suitable to learners needs.

It is important to mention that developing the competency of e-learning educators should be considered as an ongoing process, which should be evaluated periodically to judge its effectiveness. Surveying of faculty members can be used to generate the perception on a development program to enhance it (Clay, 1999).

Faculty Engagement

Faculty engagement refers to their involvement – in different academic and nonacademic endeavors – guided by the e-learning institution. Faculty engagement is argued to be a very good tool that minimizes faculty resistance, resulting from concerns about the quality of this type of learning. It also creates a cooperative culture that increases faculty motivation, awareness, and confidence (Bower, 2001).

A study – carried out at the University of Central Florida – showed there is a direct positive relationship between faculty engagement and faculty satisfaction, which proved to be a very vital measure that indicates the success of e-learning programs, and hence affects the quality of e-learning (Hartman *et al.*, 2000).

Faculty engagement is also seen as an important input for faculty innovation. Faculty who participate in more e-learning-related conferences and share experiences with colleagues develop high potential for innovating new teaching/learning methods and tools (Howell *et al.*, 2004).

Faculty Innovativeness

The emerging changes in higher education raise more and more questions, everyday, concerning the validity of the learning styles that have been used for decades and their ability to cope with the new educational interface. At present – with the association of technology to learning – faculty members are the best candidates to challenge and validate the existing learning/teaching methods and to innovate new styles that equip learners for such new environments.

As part of today's new educational frontier, the innovation opportunity for faculty has never been this wide; educators, therefore, have to respond with more courage to experiment with the learning environment and learning styles and to create and innovate. The literature on this subject has highlighted that e-learning institutions, at present, require more involvement and more innovation from faculty members to be effective and to produce quality learning (Laurillard, 1999; Conole *et al.*, 2000; Howell *et al.*, 2004).

According to Conole *et al.* (2000), the level of innovation in technology should be associated with a matching level of faculty innovation on aspects of teaching and learning to assure the quality of this type of learning. The authors argue innovation in teaching and learning should be the main purpose for advancement in information technology (IT).

A project, conducted in Pennsylvania State University, supported the argument that faculty innovation in teaching and learning has a direct and significant impact on the quality of e-learning. Through forming platforms of multidisciplinary teams of faculty and instructional

designers, faculty members were encouraged to re-examine their pedagogical approaches and come up with learning/teaching methods that enable learners of today (Oblinger and Maruyama, 1996).

Howell *et al.* (2004) argue effective learning – along with research promotion – can both be achieved if faculty members consider their e-learning experiences as an opportunity for experimenting and innovating new teaching and learning methods, as well as a part of their research efforts.

Faculty Recognition

Faculty satisfaction and motivation and, hence, quality of e-learning programs are also affected by the level of recognition received by faculty members (Clay, 1999; Srikanthan and Dalrymple, 2002).

Many argued that as e-learning demands faculty to develop new skills and assume different roles, setting appealing incentive structures for faculty members is, then, essential (Cashion and Palmieri, 2002; Graham *et al.*, 2000; Hartman *et al.*, 2000).

The argument of the importance of faculty recognition becomes more valid with the existence of research and studies, proving the fact that e-learning exerts extra workload on faculty (Moore, 2002). Therefore, effective and proper compensation should be made to maintain faculty motivation and satisfaction.

Institutions should demonstrate support to faculty members engaged in the design and delivery of e-learning program through a variety of recognition forms that acknowledge extra workload compared to, for example, traditional courses and intellectual contribution. The practice of existing universities indicated that recognition can take different forms – such as financial compensation, upgrade in office computer equipments, adjusted salary and workload, public recognition, notes of appreciation, special privileges (Bower, 2001), or even administrative support, privilege to attend related conferences, or appreciation letters and awards (Clay, 1999).

Researchers also demonstrated the necessity of higher education institutions to revisit their faculty incentive and promotion schemes to include stipends that accommodate for faculty innovativeness (Conole *et al.*, 2000; Oblinger and Maruyama, 1996).

It is, therefore, the reason why faculty compensation and recognition have been highlighted as a critical factor that affects faculty satisfaction and motivation and, hence, the quality of e-learning.

Technological Infrastructure

Appropriateness of Technology to the Pedagogical Requirements

e-Learning is about learning and not technology; the purpose of academic programs is learning and not offering

the most modern or impressive technological facilities. The success of e-learning programs, therefore, depends on how the technology is geared toward achieving the teaching/learning purpose of the program (Bates, 2000).

Bates (2000) added that, in order to provide quality learning, technology should not be planned or implemented in isolation from the operation in the institution, its structure and processes, and – most of all – the pedagogical needs and requirements of its programs.

The appropriateness and usefulness of technology to achieve its desired objectives have been emphasized, in general, and not just in the educational context. The ISO 9126 standard for selection and evaluation of software has put this factor first on the evaluation list and called it “functionality” – which refers to the ability of the technology to accomplish a specific task and provide the desired results, interact with other systems, adhere to necessary standards, and prevent unauthorized accidental or deliberate actions (ISO, 2002).

Similarly – in his well-known model for evaluation of technology – Davis (1989) highlighted the first and fundamental determinant for selection and acceptance of technology depends on its perceived usefulness, referring to its functionality and ability to achieve certain objectives.

Commencing by addressing the pedagogical needs of e-learning program when developing the technological infrastructure is very useful in achieving at least the following (Davis, 2004):

1. Building the suitable learning system.
2. Designing the learning assessment system.
3. Determining the degree of prior preparation needed before starting the course learning journey.
4. Assessing and evaluating the quality of programs and its component courses.

It is not surprising, therefore, to find an increasing amount of literature in the field of e-learning, stipulating that enhancing the e-learning experience and the achievements of its outcomes are also dependent on the appropriateness of technology to pedagogical requirements and needs.

Reliability

Reliability of technology refers to the capability of the technology to perform a required function and maintain a certain level of performance under stated conditions for a stated period of time (ISO, 2002).

Reliability of technology has been highlighted as one of the most important factors that essentially impact the quality of the learning experience and the success of online courses (Cashion and Palmieri, 2002; Moore, 2002).

Faculty members and learners can be easily distracted from the continuation of a teaching/learning task which usually causes high level of frustration due to unreliability or instability of technology (Hartman *et al.*, 2000).

Initially, many concerns with regard to the future and substantiality of online learning were raised – resulting from the doubts concerning technological reliability and its impact, not only on the quality of e-learning but also on the survival of the entire concept.

Technical shutdowns – faced by learners and faculty members – during virtual sessions, discussions, and even assignments were among those issues, and, therefore, reliability and stability of technology was rated topmost, in many studies, as a critical factor that affects the success of the e-learning experience (Fredericksen, 2000; Bates, 2000).

Accessibility

In this respect, accessibility refers to the ability of learners – of varying technical means and technical limitations related to computer power, bandwidth, or hardware limitations – to have equal opportunity of access to course material similar to other colleagues (Bates and Poole, 2003).

Schrum and Hong (2002) argued there is a substantial relationship between the number of learners dropping their online courses and the difficulties they face trying to access the technological tools and equipments. Accessible technologies, therefore, represent a distinct advantage for learners to study according to their convenient time and pace; otherwise learners will spend their time trying to solve technical problems, instead of engaging in the learning experience.

Therefore, an understanding of the technological and financial limitations of learners, their preparedness, and ability to participate equally in the learning experience is necessary for the success of e-learning. In addition, it is important to select appropriate technological infrastructure that enables learners to focus on their learning (Davis, 2004).

At present, for this reason, many e-learning institutions offer their learning content using online and offline tools (compact disks (CDs) and digital versatile disks (DVDs)) to accommodate those who have no access to high-speed connectivity.

Learner Support

The first audiences is of learner-support services are, usually, the prospective learners – through whom they learn about the program objectives, admission criteria, learning and teaching philosophy, their readiness for e-learning, required technology, and technical competence. Learner support then continues with learners throughout their learning journey, with services such as the academic advising, the library, and the technical support helpdesk until following graduation with alumni services.

Appropriateness of Support Services to Learners' Needs

There is no doubt with regard to the importance of learners' support services which, usually, complement the learning/teaching process and enhance their learners' experience. Many authors have highlighted the importance of learners' support and its impact on the quality of e-learning programs (Hughes, 2004; Goss and Prakash, 2003; Cashion and Palmieri, 2002; Frydenberg, 2002).

Although many higher education institutions claim awareness of the needs and wants of their learners, they actually do not. Cashion and Palmieri (2002) explained that many institutions go wrong by assuming learners' needs are only pedagogical and, therefore, learners' support is exclusively limited to pedagogical support.

For instance – and despite the assumption that many of today's learners are aware of the technological tools and facilities – the level of learners' awareness of educational technologies is still not up to the mark. Many of today's learners might be exposed to e-learning courses, for the first time, at the higher education level; therefore, the preparation of learners to cope with the requirements of study, to understand the new role they have to assume, and also to understand the technology they will deal with during their study is necessary for learners before they engage in their learning experience.

Schrum and Hong (2002) claimed that technical support provided to e-learners reduced the pressure on them and released their fear from the inability to use technology. Learners who feel technical support will be provided when necessary are more able to concentrate on their study and experience a more successful learning experience.

According to Hughes (2004), learner-support tools should exceed pedagogical needs, and should be extended to include:

1. information and administrative support;
2. technological support;
3. study-skills assistance – such as time management, balancing personal demands and study pursuits, information about usage of Web content, and plagiarism;
4. educational counseling; and
5. program advising and digital library.

Responsiveness of Learner Support

Responsiveness to learners' enquiries, suggestions, or complaints has been also highlighted as an important factor that affects learners' satisfaction and, therefore, impacts the quality of learning (Hughes, 2004; McNaught, 2002; Meyer, 2002; Srikanthan and Dalrymple, 2002).

According to Harvey (2002), responsiveness of learning-support tools and staff to learners' enquiries can increase their tolerance and satisfaction; hence, increasing the quality of their learning.

Most often, a high percentage of e-learners usually comprises busy senior professionals or full-time employees who would value and appreciate the timely response to their enquiries and the availability of support when needed. Smith (2004) outlines that the availability of responsive learners' support can be demonstrated in three forms:

1. The ability of learners to get in touch and to make contact with the person to whom they wish to speak.
2. The reassurance of learner support staff that, if a message is left or an e-mail is sent, this will at least be acknowledged.
3. The prompt-actioned response, wherein if the action cannot be taken immediately, the information of actioned support will be received promptly by the learner.

Institution

Leadership Commitment to Effective Learning

According to Barker *et al.* (1999), effective learning requires a whole-institution approach that pays enough attention to the teaching/learning processes, facilitates the work of faculty members, and provides necessary support to faculty whenever necessary – starting from its leadership.

e-Learning institutions – according to the author – cannot be successful without leadership commitment to set, maintain, and share a clear direction for the institution, with a focus on its teaching/learning processes.

Setting up institutional priorities in favor of the teaching/learning quality and providing support to develop competencies and infrastructure are some exemplary forms of leadership commitment for effective learning (Marcus, 2004).

Top management should, therefore, be involved in continuous monitoring, evaluation, and improvement of the institutions' academic performance; as part of that, e-learning institutions should continuously monitor learning effectiveness, learners' satisfaction, faculty satisfaction, efficiency, and cost-effectiveness (Bourne and Moore, 2003).

Pervasiveness of Commitment to Quality and Continuous Improvement

Quality commitment is an obvious factor for developing a quality culture in an organization and for assuring and maintaining continual improvement. This factor has been highlighted, generally, in many quality management and excellence standards or models – such as the ISO 9000 and the EFQM Excellence model.

Similarly, quality commitment at all levels of the e-learning institution – including top management, faculty, and staff – has been highlighted heavily in the literature as a precondition to the success of e-learning

institutions (Marcus, 2004; Frydenberg, 2002; Barker *et al.*, 1999; Mayer *et al.*, 2000).

Pervasiveness of quality commitment should be witnessed at all levels of the institution. Top management should create a shared vision for continuous improvement. They should also drive and support the institutional quality-assurance initiatives (Kistan, 2005). Faculty members and staff should demonstrate quality commitment by their involvement in the continual improvement directives, such as assessment of teaching/learning effectiveness, review, and improvement of courses and programs (Meyer, 2002; Moore, 2002).

Appropriateness of Processes to the e-Learning Environment

The changes that accompany the introduction of e-learning in a higher education institution should be associated with a new definition of processes which impact organizational structures, roles, and responsibilities of educators and staff.

According to Bates and Poole (2003), planning for e-learning programs by defining relevant and appropriate processes to the e-learning environment and changed roles and responsibilities of staff and faculty is considered to be a critical factor that impacts the quality of e-learning programs. Processes of e-learning are not quite similar to those of conventional learning; many additions have occurred to the day-work of institutions with the introduction of this type of learning. Instructional designing, dealing with intellectual properties, planning for and maintaining the technological infrastructure, program management, administrative support for faculty or learners, dealing with faculty workload and incentives, governance and decision-making processes, or pricing are only some examples of the processes that need redefinition or customization in the e-learning environment.

Among the changes e-learning introduces is the overall change of the needs and role of learners. In this context, Austin (2001) argues that the starting point for defining and shaping processes in an e-learning environment is an understanding of customers' needs (learners, families, employers, and society) who represent the input for the learning/teaching process. A change in the input will lead to a change in the definition of the process. He argues that appropriateness of processes to the e-learning environment and their integration to serve customers' needs is an important factor upon which the quality of e-learning programs depends as well as the successful implementation of processes and the potential improvement of them later.

Financial Feasibility and Sustainability

It is obvious that many elements in the e-learning type of education are costly and require more financial investments and support from higher education institutions.

Preparing learners to this type of learning through introductory and orientation programs, the necessity to offer different types of learning styles that accommodate the different learning needs and talents of learners – offering more flexible learning methods and timings, development of faculty competencies, compensation for faculty workload and copyrighted work, developing reliable, accessible and maintained IT equipments, infrastructure and multimedia components, the need for timely and responsive learners' support – are only few examples of what investment this type of learning would require.

For the aforementioned cost elements and many more, it is not surprising to know that cost structure and, consequently, the costing of e-learning programs is considerably different from conventional educational programs (Bartolic-Zlomislic and Bates, 1999).

Therefore, the financial management in e-learning institutions has been highlighted as an important factor that should be used to judge the quality of e-learning programs (Frydenberg, 2002; Moore, 2002; Bourne and Moore, 2004).

Financial sustainability which relates the cost of the program to its financial benefits (Bartolic-Zlomislic and Bates, 1999) is considered to be one of the factors that affect the quality of e-learning programs for the following main reasons:

1. The ability to support and sustain the program while it is running even if the number of enrolled learners did not match the expectation in order to protect the interests of learners.
2. The ability to re-invest in the program to enhance it and advance it.
3. The ability to offer affordable e-learning programs.

Financial planning and management of e-learning programs is a necessary activity that should be undertaken with a lot of consideration and responsibility to assure financial sustainability and feasibility of the program – maintaining sustainability of the institution and program quality (Moore, 2002).

Many practices have appeared in e-learning institutions to increase the financial benefits of a program or to decrease its costs such as offering for economies of scale and the re-use of learning objects and modules contents for different target groups or even different courses (Frydenberg, 2002).

Research Objectives and Methodology

The main purpose of this study was to develop a framework for the effective implementation of e-learning programs through investigating the 26 factors explained in the section above. It also intends to examine the relevancy and importance of such framework by investigating the experience of learners and faculty members of e-learning programs.

The research of this study adopts a methodological triangulation approach where a mixture of quantitative and qualitative methods is used. The study starts with a deductive approach where necessary literature review in the field of e-learning quality-implementation is conducted – to assure relevant work in the same subject is not overlooked and to generate a framework of factors critical for the effective implementation of e-learning programs. A case-study strategy was selected to explore the relevancy and importance of the framework factors, in which:

- Quantitative data is collected by surveying learners and faculty members' views using a questionnaire.
- Qualitative data is collected by interviewing a group of the questionnaire's participants to triangulate the data collected by the questionnaire means and to further investigate some of the findings from the questionnaire results.

The selected multi methods approach was beneficial for the study for at least the following reasons:

- helped generate a better understanding of the results collected;
- ruled out the effect of being dependant on one method; and, therefore,
- created greater confidence in the conclusion drawn.

Case-Study Analysis and Discussion

Average Comparative Ratings

Using the mean value for each factor, **Table 1** outlines the rank of each factor as perceived by faculty members and learners in addition to the study that supported it – which represents the average comparative rating of importance for each factor as viewed by both learners and faculty members.

As can be seen from **Table 1**, responses of the learners and faculty members do not follow a specific pattern of peaks and troughs; average comparative rates of importance (indicated by the mean value) do coincide in some of the factors, such as faculty–learner interaction and relevancy of interaction to learning goals; however, they also show some significant differences in importance perception.

Referring to Ehler's (2004) argument that the effective implementation of e-learning programs is correlated to the level at which higher education institutions are consistently satisfying their principle stakeholders – that is, educators and learners – the difference in perception of importance between faculty members and learners indicates both the impact of those factors on the quality of e-learning programs from a different perspective and the different priorities which would add more pressure on e-learning institutions to fulfill the different needs and demands both of faculty members and learners.

Table 1 Comparative rating of importance for the factors of the framework

Factor	Rank of importance		Related studies
	Learner	Faculty	
Pedagogy			
1. Clarity of learning goals and objectives that are challenging and linked directly to the institution goals and objectives	4th	9th	Stone <i>et al.</i> (2001); Bates and Poole (2003); Mayer <i>et al.</i> (2001) check; Watkins and Kaufman (2003)
2. Diversity of learning styles to meet learners different needs and talents	4th	3rd	Ally (2004); Sonwalker (2002); Boettcher (1999); Hughes (2004); Chickering and Gamson (1987); Laurillard (1993); Graham <i>et al.</i> (2001)
3. Relevancy of learning styles to the course learning goals	5th	4th	Ally (2004); Coffield <i>et al.</i> (2004); Gillani (2000); Freire (1997); Bates and Poole (2003); Howell <i>et al.</i> (2004)
4. Diversity in interaction methods to meet learners' different talents	8th	6th	Sonwalker (2002); Graham <i>et al.</i> (2001)
5. Prompt responsiveness of faculty feedback to learners	4th	11th	Graham <i>et al.</i> (2001); Vrasidas and McIsaac (2000); Rovai and Barnum (2003); Smith (2004); Roblyer and Ekhaml (2000)
6. Relevancy of interaction to course goals and outcomes	8th	9th	Roblyer and Ekhaml (2000)
7. Flexibility of learning time and location	7th	12th	Brennan <i>et al.</i> (2001); Cashion and Palmieri (2002)
8. Diversity of multimedia components to meet learners different talents	9th	10th	Sonwalker (2002)
9. Relevancy of multimedia to course learning goals and objectives	11th	8th	Roblyer and Ekhaml (2000)
10. Consistency of multimedia objects design	5th	11th	Graham <i>et al.</i> (2000)
11. Clear organization and accessibility of information	3rd	7th	Graham <i>et al.</i> (2000)
12. Navigation consistency and ease of use	5th	8th	Graham <i>et al.</i> (2000)
13. Attractiveness of design and graphics	9th	12th	Graham <i>et al.</i> (2000)
Faculty			
14. Faculty competency whether pedagogical or technological	4th	2nd	Howell <i>et al.</i> (2004); Caplan (2004); Frydenberg (2002); Bower (2001); Graham <i>et al.</i> (2000); Fredericksen (2000); Hartman <i>et al.</i> (2000); Clay (1999); Bourne and Moore (2003)
15. Faculty engagement in e-learning-related events and activities	8th	12th	Bower (2001); Hartman <i>et al.</i> (2000); Howell <i>et al.</i> (2004)
16. Faculty innovativeness and creation of new methods that enhance the e-learning environment	9th	1st	Laurillard (1999); Conole <i>et al.</i> (2000); Howell <i>et al.</i> (2004); Oblinger and Maruyama (1996)
17. Faculty recognition of extra load, intellectual contribution, innovativeness, etc.	8th	7th	Clay (1999); Srikanthan and Dalrymple (2002); Cashion and Palmieri (2002); Graham <i>et al.</i> (2000); Hartman <i>et al.</i> (2000); Smith (2004); Brennan <i>et al.</i> (2001); Moore (2002); Bower (2001); Conole <i>et al.</i> (2000); Oblinger and Maruyama (1996)
Technology			
18. Appropriateness of technology to the pedagogical content	6th	8th	Bates (2000); Davis (2004)
19. Reliability and the ability of the technology to perform whenever needed	2nd	1st	Cashion and Palmieri (2002); Moore (2002); Hartman <i>et al.</i> (2000); Fredericksen (2000); Bates (2000)
20. Accessibility by all learners, regardless of the connectivity and bandwidth options available	1st	1st	Bates and Poole (2003); Schrum and Hong (2002); Davis (2004)
Learner support			
21. Appropriateness of support services to learners' needs; pedagogical, technological, and others	6th	11th	Hughes (2004); Goss and Prakash (2003); Cashion and Palmieri (2002); Frydenberg (2002); Schrum and Hong (2002)
22. Responsiveness of learner support services and staff to learners' enquiries	5th	8th	Harvey (2002); Hughes (2004); Smith (2004); McNaught (2002); Meyer (2002); Srikanthan and Dalrymple (2002)
Institution			
23. Leadership commitment to effective learning	8th	1st	Barker <i>et al.</i> (1999); Marcus (2004); Bourne and Moore (2003)

Continued

Table 1 Continued

<i>Factor</i>	<i>Rank of importance</i>		<i>Related studies</i>
	<i>Learner</i>	<i>Faculty</i>	
24. Pervasiveness of quality commitment culture	5th	8th	Marcus (2004); Frydenberg (2002); Barker <i>et al.</i> (1999); Mayer <i>et al.</i> (2000); Austin (2001); Smith (2004); Kistan (2005); Meyer (2002); Moore (2002)
25. Appropriateness of processes to the e-learning environment and strategy	9th	12th	Bates and Poole (2003); Austin (2001)
26. Financial sustainability and feasibility	8th	5th	Frydenberg (2002); Moore (2002); Bourne and Moore (2004); Bartolic-Zlomislic and Bates (1999); Moore (2002); Rovai (2003)

Common Top-Rated Factors by Learners and Faculty Members

Among the top ten factors rated by faculty members and learners, six factors were common. Those factors are believed to be critical factors to both faculty members and the learner when implementing e-learning and, therefore, e-learning institutions should tackle with extreme importance,

1. Reliability of the technology.
2. Accessibility of the technology.
3. Faculty competency.
4. Use of diversified learning styles to suit learners' talents and needs.
5. Faculty–learners interaction.
6. Relevancy of learning styles to learning goals and objectives.

Dimensions Analysis

Pedagogy

Positive and close responses obtained by both faculty and learners in this dimension supports the work of Sonwalker (2002) and Ally (2004), who stipulated diversity of learning styles is a powerful solution to accommodate for learners different needs and courses different goals and objectives. In addition, the responses support the importance of clarity of course goals (Stone *et al.*, 2001; Bates and Poole, 2003; Mayer *et al.*, 2001; Watkins and Kaufman, 2003) and the importance of appropriateness and relevancy of learning styles factor (Ally, 2004; Coffield *et al.*, 2004; Gillani, 2000; Freire, 1997).

Faculty

Faculty competency is one of the top-rated factors at all dimensional levels as per the views of learners – and as per the views of faculty members. Those results come to support the work of many researchers in this field such as

Howell *et al.* (2004), Caplan (2004), and Graham *et al.* (2000). On the other hand, many studies have argued that faculty members' engagement in the institute's e-learning activities, ventures, and communities is positively linked to faculty members' innovativeness (Bower, 2001; Hartman *et al.*, 2000; Howell *et al.*, 2004). This direct relationship was not seen, however, by the surveyed faculty members who highly rated faculty innovativeness, yet the majority of the surveyed faculty members perceived faculty engagement as not important.

Similarly, an enormous amount of literature has emphasized that the quality of faculty–learner interaction is what matters – urging the need for prompt and timely response from faculty to learners when needed (Graham *et al.*, 2001; Rovai and Barnum, 2003; Smith, 2004; Roblyer and Ekhaml, 2000; Vrasidas and McIsaac, 2000). Despite this, 50% of surveyed faculty believed prompt responsiveness of faculty feedback to learner does not affect the quality of e-learning programs.

Given this fact, this factor was among the highest-rated factors by learners at all dimensional levels; this area needs more investigation where the results of faculty members can be due to their lack of readiness or awareness for the need to engage in extra workload, or other reasons.

Technology

Despite the differences in views and perception of surveyed faculty members and learners on the level of importance of many factors; they agreed that technology reliability and technology accessibility were among the two top-rated factors. Those results support the findings of Cashion and Palmieri (2002) and Fredericksen (2000), who also concluded that technology-related factors occupy the highest priorities when it comes to learners' and faculty satisfaction.

The results also support the work of Schrum and Hong (2002), who suggested that the importance of technology

accessibility is an undervalued issue and is not taken into consideration by many e-learning institutions despite its impact on withdrawal rates.

This is not surprising – simply because e-learning cannot work without the proper technological infrastructure. Assume a situation where the institution has the best and fully engaged faculty member, the right pedagogy in place, all necessary support is provided, yet have technological infrastructure which fails to operate when needed, or have unreliable infrastructure that shuts down in the middle of a virtual classroom. The level of frustration caused to faculty and learners by inaccessible or unreliable technology, reported in the literature (Moore, 2002; Hartman *et al.*, 2000; Bates, 2000), is, therefore, not exaggerated according to the views of surveyed faculty and learners.

Learner Support

Learners placed a lot of emphasis on prompt responsiveness to their enquiries whether from faculty members or learners' support staff. This again supports the reported literature which finds timely responsiveness from e-learning faculty and staff compensates for the loss of face-to-face and social interaction (Vrasidas and McLissac, 2000) and increases learners' tolerance and satisfaction (Harvey, 2002).

Prompt responsiveness of faculty feedback to learners was highly repeated in the literature for its impact on the quality of e-learning programs and also received a lot of enthusiasm by the surveyed learners, yet was believed to be not important by a majority of faculty members.

This result contradicts the increasing literature that emphasizes the new role of faculty members. Faculty have more expanded roles than lecturing at present; they are facilitators of the learning process – mentors who assure the engagement of learners at all levels (Cashion and Palmieri, 2003; Mendenhall, 2001; O'Neill *et al.*, 2003). This realization cannot be achieved without full dedication of the faculty who should be available to learners with timely responses whenever needed. The literature has reported how learners can lose interest in the course if they do not receive timely feedback on their queries and progress (Smith, 2004).

Institution

There is a general assumption among many educators that e-learning programs can operate under the same conventional environment and support system as any traditional education program (Bates and Poole, 2003; Howell *et al.*, 2004). This, perhaps, explains why 66.7% of the surveyed faculty members did not see the importance of having appropriate processes for the e-learning environment.

Conclusion

The main purpose of this study was to explore the factors that are considered critical for the effective implementation of e-learning programs, to develop a framework of such factors, and to test the relevancy and importance of the framework's components in an e-learning institution in the Arab world.

In the development of the critical factors framework, a thorough exploration and review of the existing literature was carried out. A framework of 26 factors – categorized under five main dimensions – was developed as a result of interpreting and analyzing the present literature. The study then continued by testing the relevancy and levels of importance of each factor using a case-study approach.

The empirical research was, mainly, based on gathering qualitative and quantitative data using distributed questionnaires to learners and faculty members, as well as conducting semi-structured interviews.

As a result of this study, all factors included in the evaluation framework were found relevant and important by either faculty members or learners or by both. The findings of this study support the work of researchers upon which the framework was built.

Generally, there was a difference in perception between learners and faculty members on the level of importance of many factors. Yet – and despite the differences in views and perception of surveyed faculty members and learners – they agreed that technology reliability and technology accessibility were among the two top-rated factors.

In addition to the technology reliability and technology accessibility factors, learners and faculty members shared the perception of importance of: faculty competency, use of diversified learning styles to suit learners' talents and needs, faculty-learners' interaction, and relevancy of learning styles to learning goals and objectives.

It is crucial that e-learning institutions keep searching for novel processes, methods, and support systems to cope with today's new educational frontier and to provide quality education. Measuring and evaluating if the institution is capable of doing this is yet another crucial matter.

What have been identified in this study as critical factors for the effective implementation of e-learning will need to be continually reviewed and enhanced. e-Learning faces challenges to stay updated in a learning/teaching world that is changing faster than ever, supported by powerful tools, tendered by different types of teachers, and offered to more aware and educated learners.

Bibliography

- AbuSnieneh, W. and Zairi, M. (2007a). Critical factors for the effective implementation of e-learning programs in higher education institutions – research gap. *Proceedings of Middle East Quality Association 1st Congress*. Dubai, UAE.

- AbuSnieneh, W. and Zairi, M. (2007b). Quality of e-learning– literature review. *Proceedings of e-Learning Excellence Conference*. Dubai, UAE.
- Ally, M. (2004). Foundations of educational theory for online learning. In Anderson, T. D. and Elloumi, F. (eds.) *Theory and Practice of Online Learning*, pp 3–31. Athabasca: Athabasca University.
- Barker, K., Wendel, T., and Richmond, M. (1999). Linking the literature: School effectiveness and virtual schools. *The society for the advancement of excellence in education*. Vancouver. <http://www.futured.com/pdf/Virtual.pdf> (accessed October 2009).
- Bartolic-Zlomislis, S. and Bates, T. (1999). Assessing the Costs and Benefits of Telelearning: A case study from the university of British Columbia. [online]. <http://research.cstudies.ubc.ca/nce/EDST565.pdf> (accessed October 2009).
- Bates, A. and Poole, G. (2003). *Effective Teaching with Technology in Higher Education: Foundations for Success*. San Francisco, CA: Jossey-Bass.
- Bates, T. (2000). *Managing Technological Change: Strategies for College and University Leaders*. San Francisco, CA: Jossey-Bass.
- Beaumaster, S. and Long, J. (2002). Pedagogy, technology and learning styles: Their effect on the online learner. *The 22nd Adult Higher Education Alliance Annual Conference: Creating New Meaning in Leading Learning*. Pittsburgh, PA.
- Boettcher, J. V. (1999). *Faculty Guide for Moving, Teaching and Learning to the Web*. Mission Viejo, CA: League for Innovation in the Community College.
- Bourne, J. and Moore, J. (2003). *Elements of Quality Online Education*. Needham, MA: Sloan Center for OnLine Education.
- Bower, B. (2001). Distance education: Facing the faculty challenge. *Online Journal of Distance Learning Administration* 4(2).
- Caplan, D. (2004). The development of online courses. In Anderson, T. D. and Elloumi, F. (eds.) *Theory and Practice of Online Learning*, pp 175–194. Athabasca: Athabasca University.
- Cashion, J. and Palmieri, P. (2003). The secret is the teacher: The learner's view of online learning. Australia: NCVER. <http://www.ncver.edu.au/research/proj/nrOF03a.pdf> (accessed October 2009).
- Chickering, A. W. and Gamson, Z. F. (1987). Seven principles of good practice in undergraduate education. *AAHE Bulletin* 39, 3–7.
- Ciavarella, A. P. (2002). Assessing the Quality of Online Instruction. Apr. NPS.
- Clay, M. (1999). Development of training and support programs for distance education instructors. <http://www.westga.edu/~distance/clay23.html> (accessed October 2009).
- Coffield, F., Moseley, D., Hall, E., and Ecclestone, K. (2004). Should we be using learning styles? *What research has to say to practice* [Online]. London: Learning and Skills Development Agency. http://www.itslifejimbutnotasweknowit.org.uk/files/LSRC_LearningStyles.pdf#search='should%20we%20be%20using%20learning%20styles' (accessed October 2009).
- Cohen, M. and Ellis, T. (2004). Validating a criteria set for an online learning environment. *34th ASEE/IEEE Frontiers in Education Conference*. Georgia Tech Savannah Campus, Savannah, GA, 20–23 October.
- Conole, G., O'Leary, R., Bayne, S., Cook, J., and Butland, G. (2000). Information and communications technologies: Issues for learning and teaching. *ASCILITE 2000 Conference*. Southern Cross University, Coffs Harbour. Australia, 9–14 December.
- Cooper, P. A. (1993). Paradigm shifts in designing instruction: From behaviorism to cognitivism to constructivism. *Educational Technology* 33(5), 12–19.
- Craik, F. I. M. and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General* 10(4), 268–294.
- Davis, A. (2004). Developing an infrastructure for online learning. In Anderson, T. D. and Elloumi, F. (eds.) *Theory and Practice of Online Learning*, pp 97–114. Athabasca: Athabasca University.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly* 13, 319–339.
- Foley, A. and Regan, B. (2002). Web design for accessibility: Policies and practice. *Educational Technology Review* 10(1). <http://www.aace.org/pubs/etr/issue2/foley.cfm> (accessed October 2009).
- Fredericksen, E., Pickett, A., Shea, P., Pelz, W., and Swan, K. (2000). Factors influencing faculty satisfaction with asynchronous teaching and learning in the SUNY learning network. *Journal of Asynchronous Learning Networks* 4(3).
- Frydenberg, J. (2002). Quality standards in e learning: A matrix of analysis. *International Review of Research in Open and Distance Learning* 3(2).
- Gillani, B. (2000). Using the web to create student-centered curriculum. In Cole, R. A. (ed.) *Issues in Web-Based Pedagogy*, pp 161–182. Westport: Greenwood Press.
- Goss, H. and Prakash, S. (2003). Enhancing the quality of support for online learning. *OLT 2003 Excellence Conference: Making the Connections*. 5 November. Brisbane Convention and Exhibition Centre.
- Graham, C., Cagiltay, K., Craner, J., Lim, B.-R., and Duffy, T. M. (2000). Teaching in a web-based distance learning environment: An evaluation summary based on four courses. <http://crlt.indiana.edu/publications/crlt00-13.pdf> (accessed October 2009).
- Graham, C., Cagiltay, K., Lim, B. R., Craner, J., and Duffy, T. M. (2001). Seven principles of effective teaching: A practical lens for evaluating online courses. *The Technology Source*. <http://ts.mivu.org/default.asp?show=article&id=1034> (accessed October 2009).
- Hartman, J., Dziuban, C., and Moskal, P. (2000). Faculty satisfaction in ALNs: A dependent or independent variable? *Journal of Asynchronous Learning Network* 4(3).
- Harvey (1995). Editorial. *Quality in Higher Education* 1(1), 5–12.
- Howell, S., Saba, F., Lindsay, N., and Williams, P. (2004). Seven strategies for enabling faculty success in distance education. *Internet and Higher Education* 7, 33–49.
- Hughes, J. (2004). Supporting the online learner. In Anderson, T. D. and Elloumi, F. (eds.) *Theory and Practice of Online Learning*, pp 367–384. Athabasca: Athabasca University.
- ISO (2000). ISO 9126. <http://www.issco.unige.ch/ewg95/node14.html> (accessed October 2009).
- Kistan, C. (2005). Improving the quality of e-learning through evaluation: Whose interest is served. In Smout, M. (ed.) *The Decade Ahead*. South Africa: South African Universities.
- Laurillard, D. (1993). *Rethinking University Teaching: A Framework for the Effective Use of Educational Technology*. London: Routledge.
- LTSN Generic Centre (2004). LTSN pedagogy and policy vocabularies. *Higher Education Academy*. http://www.heacademy.ac.uk/scoping_notes.doc (accessed October 2009).
- Marcus, S. (2004). Leadership in distance education: Is it a unique type of leadership – a literature review. *Online Journal of Distance Learning Administration* 7(1). <http://www.westga.edu/~distance/ojdl/spring71/marcus71.html> (accessed October 2009).
- Mayer, D. P., Mullens, J. E., and Moore, M. T. (2000). *Monitoring School Quality: An Indicators Report*. <http://nces.ed.gov/pubs2001/2001030.pdf> (accessed October 2009).
- McNaught, C. (2003). Supporting the global e-teacher. *International Journal of Training and Development* 7(4), 287.
- Mendenhall, B. (2001). Technology: Creating new models in higher education. *National governor's association*. <http://www.nga.org/cda/files/HIGHEREDTECH.pdf> (accessed October 2009).
- Meyer, K. A. (2002). Quality in distance education: Focus on online learning. *Higher Education Report* 29(4), 1–121.
- Moore, J. C. (2002). *Elements of Quality: The Sloan-CTM Framework*. Needham: Sloan Center for On-line Education.
- Oblinger, D. and Maruyama, M. (1996). Distributed learning. *CAUSE Professional Paper #14*, CAUSE.
- Schrump, L. and Hong, S. (2002). Dimensions and strategies for online success: Voices from experienced educators. *Journal of Asynchronous Learning Network* 6(1).
- Simpson, R. J. and Galbo, J. J. (1986). Interaction and learning: Theorizing on the art of teaching. *Interchange* 17(4), 37–51.
- Skinner, B. F. (1974). *About Behaviorism*. New York: Knopf.
- Sonwalkar, N. (2001). Changing the interface of education with revolutionary learning technologies. *Syllabus Magazine*. November.
- Sonwalkar, N. (2002). A new methodology for evaluation: The pedagogical rating of online courses. *Syllabus Magazine*, January.

- Srikanthan, G. and Dalrymple, J. F. (2002). Developing a holistic model for quality in higher education. *Quality in Higher Education* 8(3), 215–224.
- Stone, W. S., Schowalter, E. D., Orig, A., and Grover, M. (2001). An empirical study of course selection and divisional structure in distance education programs. *Online Journal of Distance Learning Administration* 4(1). <http://www.westga.edu/~distance/ojdla/spring41/stone41.html> (accessed October 2009).
- Vrasidas, C. and McIsaac, M. (2000). Principles of pedagogy and evaluation of web-based learning. *Educational Media International* 37(2), 105–111.
- Watkins, R. and Kaufman, R. (2003). Strategic planning for distance education. In Moore, M. G. and Anderson, W. G. (eds.) *Handbook of Distance Education* (QUESTIA), pp 507–517. Mahwah, NJ: Erlbaum.
- Wirth, M. (2005). Quality management in e-learning: Different paths, similar pursuits. *Proceedings of the 2nd International SCIL Congress*. <http://www.scil.ch/congress-2005/programme-10-11/docs/workshop-1-wirth-text.pdf> (accessed October 2009).
- Zhao, F. (2003). Enhancing the quality of online higher education through measurement. *Quality Assurance in Education* 11(4), 214–221.

Further Reading

- Craik, F. I. M. and Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11, 671–684.
- Phillips, J., Phillips, P., and Zuniga, L. (2000). Evaluating the effectiveness and the return on investment of e-learning. *What Works Online*. 2nd quarter.
- Rovai, A. (2002). Building sense of community at a distance. *International Review of Research in Open and Distance Learning* 3(1), [online]. <http://www.irrodl.org/content/v3.1/rovai.pdf> (accessed October 2009).
- Vrasidas, C. (2000). Constructivism versus objectivism: Implications for interaction, course design, and evaluation in distance education. *International Journal of Educational Telecommunications* 6(4), 339–362.
- Wager, J. (2005). Support services for the net generation, EDUCAUSE. <http://www.educause.edu/ir/library/pdf/pub7101j.pdf> (accessed October 2009).

Globalization and Social Justice in Higher Education

E Kimura-Walsh, Santa Clara University, Santa Clara, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Democratization – The shift from an authoritarian social or political system to one that provides democratic participation in which all people have legitimate representation.

Divest or divestment – A removal of funds from an area of investment, including higher education, corporations, etc.

Recuperated businesses – Businesses that were closed or under threat of being closed as a result of Argentina's 2001 economic collapse until workers took on collective ownership. They are part of a larger recuperated businesses movement.

Introduction

Scholars are developing increasingly complex – and, often, contradictory – understanding of globalization and its influences. Some scholars credit globalization with troubling results, such as exploiting the poor (McLaren and Farahmandpur, 2001) and westernizing the world (Altbach, 2005), whereas others note more progressive results including increases in foreign aid (Stiglitz, 2002) and opportunities for international scholarly collaboration (Altbach, 2005). Within the realm of higher education, scholars find globalization equally complex and contradictory. This article examines the current literature with regard to globalization's effects on higher education. Moreover, it suggests how higher education institutions can use the mechanisms of globalization to further social justice issues, locally and globally.

Globalization from Above, Globalization from Below

Kellner's (2002) critical theory of globalization is helpful in understanding the numerous and contradictory conceptions and consequences of globalization. Kellner (2000) notes that one must acknowledge that globalization has "progressive and regressive features, as well . . . as fundamental ambivalences" (p. 305). By encompassing both positive and negative characteristics, Kellner (2000, 2002) suggests a critical theory of globalization that

deconstructs and denounces oppressive and undemocratic features of globalization, while supporting and promoting the positive aspects of globalization that uphold justice and democratization. (Democratization refers to the shift from an authoritarian social or political system to one that provides democratic participation in which all people have legitimate representation.)

To further explore the contradictory nature of globalization, Kellner (2002) suggests that a distinction be made between globalization from above and globalization from below. Globalization from above – or neoliberal globalization – is the use of globalization to inhibit democracy, exploit and oppress the poor, and generate more power and wealth for the rich (Kellner, 2000, 2002). Neoliberals share the belief that liberalization of the market will create equality by removing human bias and promoting rationality and fairness in the allocation of goods and services (McChesney, 1999). As a result, they attempt to use the mechanism of globalization to create a hierarchical global structure that emphasizes capitalism, privatization, individualistic prosperity, free market values, and a weak state (Apple, 2000). These views are best known for guiding the policies of the World Trade Organization (WTO), World Bank, and International Monetary Fund (IMF) – the three primary institutions that govern global economy and trade.

Critics view these neoliberal ideals as advancing the exploitation of marginalized people and countries (Chomsky, 1999; Chua, 2004). They argue that, instead of providing for fair policies, the neoliberal agenda of the WTO, World Bank, and IMF has detrimental effects on developing countries and their citizens, while looking out for the interests of the G-7: the United States, Canada, Japan, France, Germany, Italy, and United Kingdom (Stiglitz, 2002). For example, these three organizations practice capital market liberalization, which requires poor countries seeking aid to liberalize their trade and financial markets as a condition for economic assistance. Similar policies contributed to Argentina's economic collapse when – as a result of opening their markets – their banking system became dominated by foreign banks that failed to fund domestic small- and medium-size businesses (Stiglitz, 2002).

In contrast, globalization from below describes anti-neoliberal global movements that are mainly concerned with challenging the liberalization of markets, establishing rules to govern the global economy, and ensuring the interests of all people are served (Torres and Rhoads, 2006). Moreover, globalization-from-below movements

resist neoliberal exploitation and oppression. They use the trends of globalization to empower marginalized people and counter global structures of domination and exploitation (Kellner, 2000). Globalization from below takes many different forms including environmental, human and labor rights, and identity movements (Brecher *et al.*, 2000; Torres and Rhoads, 2006). Examples of globalization-from-below activism include the Anti-Apartheid Movement (Altbach and Cohen, 1990); protests against the World Bank, IMF, and WTO (Smith, 2001); as well as the Group of Seventy-Seven (G-77) – a United Nations coalition of 131 developing countries that promotes the economic interests of developing countries and advocates for a shift in global economic decision making from the World Bank and IMF to the United Nations (Brecher *et al.*, 2000).

Higher Education and Globalization from Above

Kellner's (2002) concept of globalization from above and below is helpful in examining globalization's impact on post-secondary education. Globalization has mixed – often conflicting – effects on access and equity within and between national higher education systems, and individual institutions. The majority of literature exploring globalization's impact on higher education focuses on the consequences of globalization from above. Some of the effects most prominently noted in the literature include: increased disparities between post-secondary systems throughout the world, a trend of privatization within institutions, and the westernization of higher education research and teaching.

Hierarchies Intensified

Globalization has intensified long-standing hierarchies among various systems of higher education throughout the world (Altbach, 2005). The United States, Europe, and a few other wealthy countries have well-developed post-secondary systems with ample resources to fund teaching, research, and other activities. Developing countries, however, struggle to maintain and expand their colleges and universities. As a result of globalization, privileged higher education systems wield increasing international influence and have become the ideal to which other systems are compared. As developed countries become dominant centers of knowledge through research production, knowledge exchange (e.g., academic journals, conferences), and advanced training, it is more difficult for less developed higher education systems to catch up (Altbach, 2005). As a result of growing international competition, existing disparities have become more extreme, if not irreparable.

Moreover, the IMF and World Bank aggravate these disparities by promoting policies that force poor countries

to divest from higher education as a condition for further borrowing (Puplampu, 2006; Torres and Rhoads, 2006). (Divest or divestment refers to a removal of funds from an area of investment, including higher education, corporations, etc.) The IMF and World Bank promote these policies because they view higher education as having a low rate of return (Puplampu, 2006; Santos, 2006), and education as a private good that should be allocated through the free market. As a result, countries such as Mexico, Argentina, and Brazil have been forced to redirect funding from higher education to elementary and secondary education (Torres and Rhoads, 2006).

Similarly, in many African countries, the World Bank and IMF's structural adjustment packages resulted in state divestment from education, generally, and higher education, specifically (Puplampu, 2006; Santos, 2006; Shizha, 2006). Shizha (2006) argues the enactment of these neoliberal policies in Zimbabwe diminished access to education for the poor because tuition increased to supplement the state's declining financial investments. As a result of higher education's declining resources, institutions also face falling faculty salaries and scholars leaving to seeking more stable institutional environments. These trends have led to a shortage of academic personnel, and undermined Africa's knowledge production (Puplampu, 2006; Teferra and Altbach, 2004).

The WTO is also contemplating further imposition of neoliberal-influenced policies on higher education. As part of the General Agreement on Trade in Services (GATS), the WTO is considering including higher education in the open market. This agreement is still being developed; however, if it is carried out, developing countries would be forced to allow multinational educational providers (primarily those from G-7 countries) free access to their countries. Simultaneously, developing countries would be prevented from sufficiently supporting their own higher education systems, thus lessening their chances of competing with better-resourced, foreign educational suppliers (Altbach, 2005). Furthermore, without any stipulations to ensure mutually beneficial relationships between foreign providers and their host countries, GATS could weaken the social good purpose of higher education (Santos, 2006). This could occur in South Africa where affirmative action is utilized in higher education to integrate the society and advance opportunities for all. Foreign higher education institutions, that utilize selective admissions likely to favor whites, would undermine the country's use of education in the broader battle against racism (Santos, 2006).

Privatization of Higher Education

Universities are also engaging in more capitalist-oriented behaviors as a result of neoliberal divestment from higher education. Clark (2004) identifies this trend on an

international level, documenting entrepreneurial business-like activities at such diverse institutions as the University of Strathclyde in Scotland, Makerere University in Uganda, and Catholic University of Chile. In US institutions, the influence of market ideology has grown to the extent that Slaughter and Rhoades (2004) propose a major shift from a public good knowledge/learning regime to an academic capitalist knowledge/learning regime where higher education practices are increasingly driven by the market.

This neoliberal influence can be seen in the ever-increasing presence of private industry on college campuses. Higher education institutions are involved in patenting new knowledge and products, and their boards of trustees are largely made up of private industry executives (Slaughter and Rhoades, 2004). Private corporations also have a growing influence on academic research (Gumport, 1991). For example, the pharmaceutical company Novartis offered the University of California, Berkeley, \$25 million for the Department of Plant and Microbial Biology in exchange for two seats on the grants committee that determined which research to fund, and the ability to vet all inventions emerging from the department (Press and Washburn, 2000; Rosenzweig, 1999). (With the exception of research funded by other corporations.) Further, Novartis would also have the right of first refusal to license a fraction of the patents equivalent to the percentage of the department's funding provided by the company (Rosenzweig, 1999). Critics are concerned corporate intervention could taint academic research by potentially creating bias, and shifting its purpose to serve the private not public good (Press and Washburn, 2000).

Increasing Homogenization

At an international level, globalization has a homogenizing effect on higher education. As countries strive to develop world-class institutions in a global context, these institutions are likely to follow Western models. For example, Mohrman (2005) points out that China's leaders are failing to take advantage of their universities' strengths in the humanities and social sciences. Instead, they are attempting to catch up with Western universities' achievements in the natural and physical sciences – an ambition they have little chance of fulfilling.

Furthermore, research standards have also become biased toward Western values. The methodological practices of Western researchers are accepted as the norm, forcing non-Western scholars to adapt their research practices to be both accepted and understood in the international academic realm (Altbach, 2005). Additionally, the growing dominance of English in academic communication is also evidence of globalization's homogenizing effects (Altbach, 2005). As the most widely learned and used second language, English dominates academic journals,

cross-border degree programs, and international conferences (Altbach, 2005). This widespread use of English creates an advantage for English speakers and English-speaking countries in academe, while further marginalizing non-English speakers (Altbach, 2005; Puplampu, 2006).

Overall, scholars argue neoliberal globalization is having a detrimental effect on equity within higher education. These neoliberal policies engender a broad trend of public divestment from higher education which is contributing toward the privatization of institutions. Further, this divestment is increasing the divide between the haves and the have nots among higher education systems, and limiting access for the poor in developing countries to post-secondary education.

Higher Education as Globalization from Below

While scholars have identified the impact of neoliberal globalization on higher education, less has been written with regard to how higher education institutions can resist neoliberal globalization, and use the mechanisms of globalization to advance social justice. Higher education institutions in developed countries have a particularly pressing responsibility because of their privileged status (Altbach, 2005). As a result, there have been calls for universities – particularly in the United States – to engage in globalization from below by resisting neoliberal trends and addressing social inequality (Altbach, 2005; Kimura-Walsh and Allen, 2008). While literature on institutional participation in globalization from below has been limited, scholars have explored some ways institutions can advance the social good in the context of globalization.

International and Local Collaborations

Collaborations at the international and local levels offer an important means for higher education institutions to use mechanisms of globalization, in conjunction with existing resources, to benefit marginalized higher education institutions, countries, and people. Technological advances have enhanced worldwide communication and travel, thus making international research and scholarly collaborations more viable. Altbach (2005) argues US institutions and faculty are in an ideal position to pursue and promote global scholarly relationships. In this way, higher education institutions can “develop collaborative, mutually beneficial initiatives and policies, eschewing profit-making ventures that advance the hegemony of powerful academic institutions and systems” (Altbach, 2005, p. 64). If so motivated, US colleges and universities can avoid buying into and profiting from their dominance in the global higher education system. One natural link that deserves more exploration is potential collaboration between scholars

from developing countries who have migrated to developed countries, and higher education institutions or other organizations in their home countries. As technology expands, these scholars can maintain better contact with their countries of origin, and provide an important source of intellectual capital (Altbach, 2005; Teferra, 2000).

Higher education institutions and their constituents can also play an important role in local globalization-from-below activities. In addition to promoting a neoliberal agenda, the World Bank and IMF, often, fail to consider regional context (Kempner and Jurema, 2006; Slocum and Rhoads, 2009). As a result, local initiatives are necessary to address the unique economic struggles of developing countries because they take into account factors such as culture, values, community dynamics, and local economy (Kempner and Jurema, 2006). For example, Kempner and Jurema (2006) describe Integrar Institute – a collaboration initiated by Brazil's *Central Única de Trabalhadores* (CUT; Workers Central Union) with public universities in Rio de Janeiro, Campinas, and São Paulo. This program provides courses to union workers, managers, and unemployed citizens in areas such as basic education, management, planning, computing, and economics. Through these courses, participants improve their workplace skills and gain exposure to advanced educational opportunities. As seen in this example, universities in collaboration with grassroots organizations can address the economic challenges of their communities more effectively than global organizations while also undoing some of the negative ramifications of neoliberal policies.

In addition to institutional efforts, individual faculty and students can also play important roles in local movements opposing neoliberal globalization. Slocum and Rhoads (2009) examine ways faculty and students at Argentina's University of Buenos Aires (UBA) opposed neoliberal trends in the wake of the country's 2001 economic collapse resulting from IMF and World Bank policies. In addition to participating in social protest, faculty and students engaged in forms of "intellectual transference" (p. 99). They generated theory and research on the broad forces shaping the country which provided a new lens for marginalized groups to understand their experiences. Further, university constituents helped grassroots organizations, such as recuperated businesses with services such as accounting and technological assistance. (Recuperated businesses are businesses that were closed or under threat of being closed as a result of the economic collapse until workers took on collective ownership of them. They are part of a larger recuperated businesses movement in Argentina.)

Massification of Higher Education

In some cases, globalization has encouraged countries to increase access to higher education because of

the growing perception that post-secondary education increases competitiveness in the global market (Yang, 2002). While the increasingly market-oriented goals of higher education are questionable, the expansion of higher education offers a wider range of citizens more job opportunities, as well as social mobility, and improved lifestyles. For example, China expanded their higher education system to increase their economic competitiveness and stimulate the economy with families' educational savings. As a result, the proportion of the population attending college grew from 3.4%, in 1990, to 11%, in 2000 (Yang, 2002). Unfortunately, paired with this growth in access is a trend of increasing tuition which limits the opportunities of poorer citizens (Yang, 2002). As a result, we are seeing a global trend of massification – but not democratization – in higher education (Santos, 2006). As globalization heightens the priority placed on higher education, it is also important to consider how access can be more equally distributed among all citizens.

Engaging in Social Responsible Practices

Within a global context, university responsibilities extend beyond community and nation-state borders, and beyond the realm of education to include international human and labor rights issues. Using their global networks, universities can challenge neoliberal trends and promote the well-being of people throughout the world (Kimura-Walsh and Allen, 2008). For example, some US higher education institutions have signed on to the student-led anti-sweatshop and Sudan Divestment movements which promote socially responsible consumerism and investing, respectively, to advance specific social justice issues (Kimura-Walsh and Allen, 2008). It is important to point out, however, that often institutions are not eager participants. Many student activists have engaged in extended advocacy and protest activities to gain the recognition and support of their universities in furthering these causes. Once involved, higher education institutions engaged in globalization from below by using their global economic connections to advance both issues. In the case of the anti-sweatshop movement, institutions used their business connections with overseas factories to promote the welfare of apparel workers. They ensured their university apparel was manufactured in factories that upheld workers' basic rights. As a result, they challenged the liberalization of the market which undermines the well-being of developing countries and their citizens (Featherstone and United Students Against Sweatshops (USAS), 2002). In the Sudan divestment case, institutions used the mechanism of their international financial connections by engaging in targeted divestment from corporations that work in the Sudan. This strategy was undertaken to pressure the Khartoum regime to end the Darfur genocide (Sudan Divestment Task Force, 2004).

Conclusion

Literature examining the impact of globalization on higher education has, primarily, focused on the impact of neoliberal globalization on colleges and universities. Scholars have found that neoliberal trends potentially threaten institutions' ability to benefit the broader good. With a decline in public funding, countries as diverse as the United States and Zimbabwe are seeing tuition increases that prevent less privileged students from accessing advanced education (Shizha, 2006; Zusman, 2005). These declines in financial support also provoke more business-like behaviors within institutions, including trends toward entrepreneurialism (Clark, 2004), and collaborations with private industry (Slaughter and Rhoades, 2004). Lastly – in an increasingly global context – Western institutions have become the model to which everyone aspires and English the common language. This trend is contributing to a homogenizing of higher education institutions, and to disparities between Western and non-Western scholars.

An examination of the literature reveals that colleges and universities are doing little to advance globalization from below, despite being in an ideal position to promote social justice. These institutions are often in positions of privilege with a plethora of material and knowledge-based resources. Similarly, they are well situated to take advantage of the mechanisms of globalization – such as technology and international financial connections. They are, however, doing little with these resources to promote global social justice. One of the primary ways in which institutions could advance globalization from below is through collaboration. Altbach (2005) suggests that colleges and universities – particularly those in the United States – could benefit developing countries through international collaborations made more viable by increasing technology and travel.

Additionally, higher education institutions should engage in collaborations with local organizations to advance the betterment of community members disenfranchised by the global economy and neoliberal policies. The cases of Brazil's Integrar Institute and the University of Buenos Aires' (UBA) faculty and students in Argentina provide insight into the societal benefits that higher education-community collaborations can engender. However, in the case of Integrar Institute, the program was initiated by the unions, while – in the case of UBA – individual faculty and staff took action. It is time that universities take initiative to identify ways they can counter neoliberal globalization in their local communities. Moreover, Santos (2006) proposes institutions prioritize (1) action research in which participants play a central role in conducting research that addresses problems they face; (2) and the ecology of knowledge by reversing the knowledge exchange that has, historically, occurred within higher education. Instead, knowledge should be to travel from outside to inside the university.

Institutions must also identify ways to further democratize higher education by providing access to the diverse populations represented in their countries. While the growth of higher education in some countries is encouraging, in many cases, access continues to be granted to the privileged. Universities must consider how to enhance financial assistance to students, particularly in the form of scholarships (as opposed to loans); and develop affirmative action policies to address racial, ethnic, and class disparities. Further, universities must acknowledge and challenge their role in the colonial project which has excluded marginalized groups from educational opportunities and constructed knowledge about their inferiority (Santos, 2006).

Furthermore, higher education institutions must be conscientious members of society by engaging in socially responsible activities, and advocating for social justice throughout the world. The cases of the sweatshop and Sudan divestment movements offer insight into how institutions can use their global connections and prestige to advance critical social justice causes. However, these examples also reveal the limitations of colleges and universities; they failed to take the initiative in addressing these injustices. Instead, they followed student activists, who – in many cases – had to push, prod, and protest to convince these institutions to take a stand. As a result, the trend – at present – is that universities are followers in globalization from below movements. Higher education could learn much from students in becoming agents for social change.

With the increasing influence of colleges and universities in a global, knowledge-based society, it is essential that post-secondary institutions – particularly in developed countries – take an active role in advancing globalization from below. They are in an ideal position to use the mechanisms of globalization to help developing countries advance their higher education systems and increase access. Furthermore, they are equally well-situated to work for equity in their local communities and promote social justice throughout the world. Unfortunately, in most cases, post-secondary institutions have far to go in reaching these goals. Within the context of globalization, universities must broaden their social good purpose to reach out beyond nation-state borders and the boundaries of education.

See also: Emergence of For-profit Higher Education; Ethnic Minority Identity and Educational Outcomes in a Rising China; Globalization and Curriculum; Globalization and Vocational Education and Training; Higher Education in the New Economies; Neoliberalism, the Market and Performativity; Student and Faculty Transnational Mobility in Higher Education; The Changing Role of the State in Higher Education; The Role of the OECD in the Development of Higher Education in a Globalized

World; Transforming Higher Education in Developing Countries: The Role of the World Bank; UNESCO's Role in the Development of Higher Education in a Globalized World.

Bibliography

- Altbach, P. G. (2005). Globalization and the university: Myths and realities in an unequal world. In National Education Association (ed.) *The NEA 2005 almanac of higher education*, pp 63–74. Washington, DC: National Educational Association. <http://www2.nea.org/he/healma2k5/images/a05p63.pdf> (accessed August 2009).
- Altbach, P. G. and Cohen, R. (1990). American student activism: The post-sixties transformation. *Journal of Higher Education* 6(1), 32–49.
- Apple, M. W. (2000). Between neoliberalism and neoconservatism: Education and conservatism in a global context. In Burbules, N. C. and Torres, C. A. (eds.) *Globalization and Education: Critical Perspectives*, pp 57–77. New York: Routledge.
- Brecher, J., Costello, T., and Smith, B. (2000). *Globalization from Below: The Power of Solidarity*. Cambridge, MA: South End Press.
- Chomsky, N. (1999). *Profit Over People: Neoliberalism and the Global Order*. New York: Seven Stories Press.
- Chua, A. (2004). *World on Fire: How Exporting Free Market Democracy Breeds Ethnic Hatred and Global Instability*. New York: Anchor Books.
- Clark, B. R. (2004). *Sustaining Change in the University: Continuities in Case Studies and Concepts*. New York: Open University Press.
- Featherstone, L. and United Students Against Sweatshops (2002). *Students Against Sweatshops*. New York: Verso.
- Gumport, P. J. (1991). Built to serve: The enduring legacy of public higher education. In Altbach, P. G., Gumport, P. J., and Johnstone, D. B. (eds.) *In Defense of American Higher Education*, pp 85–109. Baltimore, MD: Johns Hopkins University Press.
- Kellner, D. (2000). Globalization and new social movements: Lessons for critical theory and pedagogy. In Burbules, N. C. and Torres, C. A. (eds.) *Globalization and education: Critical perspectives* pp 299–321. New York: Routledge.
- Kellner, D. (2002). Theorizing globalization. *Sociological Theory* 20(3), 285–305.
- Kempner, K. and Jurema, A. L. (2006). Brazil's local solutions to global problems. In Rhoads, R. A. and Torres, C. A. (eds.) *The University, State, and Market*, pp 221–249. Stanford, CA: Stanford University Press.
- Kimura-Walsh, E. and Allen, W. R. (2008). Globalization from above, globalization from below: Mechanisms for social disparity and social justice in higher education. In Hopson, R. K., Yeakey, C. C., and Boakari, F. (eds.) *Power, Voice, and the Public Good: School and Education in Global Societies*, pp 201–230. Oxford: Elsevier.
- McChesney, R. W. (1999). Introduction. In Chomsky, N. (ed.) *Profit Over People: Neoliberalism and the Global Order*, pp 7–16. New York: Seven Stories Press.
- McLaren, P. and Farahmandpur, R. (2001). Teaching against globalization and the new imperialism: Toward a revolutionary pedagogy. *Journal of Teacher Education* 52(2), 136–150.
- Mohrman, K. (2005). World-class universities and Chinese higher education reform. *International Higher Education* 39, 1–4.
- Press, E. and Washburn, J. (2000). The kept university. *The Atlantic Monthly March*, 39–54.
- Puplampu, K. P. (2006). Critical perspectives on higher education and globalization in Africa. In Abdi, A. A., Puplampu, K. P., and Dei, G. J. S. (eds.) *African Education and Globalization: Critical Perspectives*, pp 31–52. New York: Lexington Books.
- Rosenzweig, R. M. (1999). What's for sale these days in higher education: Two stories. *Paper Presented at a Seminar of the Center for Studies in Higher Education on Strategic Policy Issues in Higher Education*. Berkeley, CA: <http://cshe.berkeley.edu/publications/docs/PP.Rosenzweig.4.99.pdf> (accessed August 2009).
- Santos, B. D. S. (2006). The university in the 21st century: Toward a democratic and emancipatory university reform. In Rhoads, R. A. and Torres, C. A. (eds.) *The University, State and Market: The Political Economy of Globalization in the Americas*, pp 60–99. Stanford, CA: Stanford University Press.
- Shizha, E. (2006). Continuity or discontinuity in education equity: Contradictions in structural adjustment programs in Zimbabwe. In Abdi, A. A., Puplampu, K. P., and Dei, G. J. S. (eds.) *African Education and Globalization: Critical Perspectives*, pp 187–210. New York: Lexington Books.
- Slaughter, S. and Rhoades, G. (2004). *Academic Capitalism and the New Economy: Markets, State and Higher Education*. Baltimore, MD: Johns Hopkins University Press.
- Slocum, J. and Rhoads, R. A. (2009). Faculty and student engagement in the Argentine grassroots rebellion: Toward a democratic and emancipatory vision of the university. *Higher Education* 57, 85–105.
- Smith, J. (2001). Globalizing resistance: The battle of Seattle and the future of social movements. *Mobilization: An International Quarterly* 6(1), 1–19.
- Stiglitz, J. E. (2002). *Globalization and Its Discontents*. New York: W. W. Norton.
- Sudan Divestment Task Force (2004). Key divestment information. <http://sudandivestment.org/divestment.asp#q1> (accessed August 2009).
- Teferra, D. (2000). Revisiting the doctrine of human capital mobility in the information age. In Tapsoba, S., Kassoum, S., Houenou, P. V., et al. (eds.) *Brain Drain and Capacity Building in Africa*, pp 62–77. Dakar: ECA/IRDC/IOM.
- Teferra, D. and Altbach, P. G. (2004). African higher education: Challenges for the 21st century. *Higher Education* 47, 21–50.
- Torres, C. A. and Rhoads, R. A. (2006). Introduction: Globalization and higher education in the Americas. In Torres, C. A. and Rhoads, R. A. (eds.) *The University, State, and Market: The Political Economy of Globalization in the Americas*, pp 3–38. Stanford, CA: Stanford University Press.
- Yang, R. (2002). Lost opportunities in the massification of higher education in China. *International Higher Education* 28, 1–4.
- Zusman, A. (2005). Challenges facing higher education. In Altbach, P., Berdahl, R. O., and Gumport, P. J. (eds.) *American Higher Education in the Twenty First Century: Social Political and Economic Challenges*, 2nd edn., pp 115–160. Baltimore, MD: Johns Hopkins University Press.

Racial Inequality and the Economic Payoffs of Higher Education

S A Lachtman, University of Oxford, Oxford, UK

© 2010 Elsevier Ltd. All rights reserved.

In *The Republic*, Plato describes an ideal society where the elite receive educations that are complementary to their future roles and these leaders are recruited through both meritocratic and heredity criteria. In many modern industrial societies, including the United States, one concern has been that although ethnic minorities obtain elite educations and enter elite professions and industries, they do not move to the highest levels or positions in their societies and professions. The economic standing of college-educated Americans has been a concern of elite social scientists throughout the twentieth century, including W.E.B. Du Bois, Gunnar Myrdal, Peter Blau and Otis Dudley Duncan, and, more recently, Michael Hout and William Bowen. The social science literature highlights that national samples of college graduates have superior wealth, income, and labor-market outcomes than people with lesser levels of educational attainment. However, there is also general agreement that race-based income and employment inequalities are both historic and contemporary features of American society (Blau and Duncan, 1967; Bowen and Bok, 1998). In addition, social scientists have found the following mix of overlapping trends, to differing degrees:

1. some overall social mobility, but much less at the elite level;
2. higher-income returns for graduates of elite universities;
3. income penalties for ethnic minority alumni of elite universities compared to nonethnic minority alumni; and
4. substantial career penalties against ethnic minorities with both nonelite and elite educations.

Many intellectuals have described education as a key method for social uplift and as a key indicator of how meritocratic a society is. Social scientists throughout the world continue to ask: What are the relationships between the origins and destinations of citizens? How are advantages and disadvantages passed or not passed throughout the generations? Also, of what utility is education for the masses? In the United States, the query has most frequently been: Can education – or how much education does it take to – overcome low socioeconomic origins and racial inequality?

Scholars have articulated a theoretical and empirical link between higher education and substantial economic rewards for African Americans. A long line of social

science literature, beginning with W.E.B. Du Bois in the late nineteenth century, examines the comparative payoffs of education for different racial groups. Du Bois concludes that the fate of the black elite is tied to education and work and that their mobility is essential for social stability. Chronicling the black college alumni of the nineteenth and early twentieth centuries, he finds college-educated African Americans doing quite well, the vanguard of black social uplift and economically well-off property holders. While Du Bois describes education as a training field, he also notes how limiting ethnic stratification is for the life chances of educated blacks. He explains that college-educated blacks are social elites within the black community, holding both relatively high-prestige professions and substantial wealth as well as being occupationally diverse, including teachers, clergy, graduate students, lawyers, civil servants, businessmen, farmers or artisans, editors, secretaries, and clerks. To Du Bois, these statistics represent a watershed in occupational diversity, a change from the historic perception that clergy and teaching were the only choices for the educated black elite. However, Du Bois recognizes the stratification of the black elite in comparison to white America, and particularly the white elite.

In *An American Dilemma: The Negro Problem and Modern Democracy*, Myrdal (1944) equates African Americans to a caste because of their lack of social mobility and opportunity. Blau and Duncan (1968) provide a more nuanced, although very similar, picture in *The American Occupational Structure*, arguing that in the United States at the time, high status was less inherited than generations before and more dependent upon actual, normative achievement, with education serving as a central means for generational transmissions. However, their conclusion “that the American Occupational structure is largely governed by universalistic criteria of performance and achievement” is presented with the Myrdal-like caveat: “the notable exception of the influence of race.” They find African-American educational levels declining and stratification increasing with education. To them, this conclusion is especially surprising because the lower socioeconomic origins of nonwhites should make upward mobility, as measured through income and occupational prestige, more likely. They theorize that this indicates both the higher returns whites receive for education and the possibility of a widening attainment gap between whites and nonwhites.

In *Black Elite: The New Market for Highly Educated Black Americans: A Report Prepared for the Carnegie Commission on Higher Education*, Freeman (1976) finds dramatic changes in the labor market for black college graduates since Blau and Duncan's analysis. To analyze change over time, he notes that as recently as 1959, the income of white male college graduates was larger than their black counterparts by nearly 81%, far greater than differences between black and whites with lesser educational attainment. Moreover, almost 50% of black male college seniors at that time planned on teaching careers compared with one-seventh of whites. In contrast, Freeman finds that in the 1970s, starting salaries of black college graduates reached parity with whites and black male graduates had more increases in income than whites for the first time in modern American history. This ended a long historical trend of widening black-white income gaps with ascending education. Freeman connected the continuing income gaps among older college graduates to historical discrimination and was optimistic about future equality. He noted the then-recent increase of blacks in managerial positions as well as distribution of blacks and whites throughout the occupations were additional encouraging signs.

Michael Hout, in his examination of General Social Survey cumulative data files for 1972–85, finds that college degrees overcome disadvantaged origins. He not only describes a less-than-ideal situation where origin status affects workers without college education but also finds that education cancels out the effect of background status for college-degree holders. In *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions*, Bowen and Bok (2000) analyze the College and Beyond Database, which tracked 80 000 undergraduates who matriculated at 28 selective colleges and universities in fall 1951, 1976, and 1989. They find that among the alumni of these highly selective institutions, gender and racial inequalities persist, as women earn less than men and black graduates less than white graduates, with especially significant racial earning differentials between males.

Intraracial benefits of graduating from a selective institution instead of a nonselective institution are also interesting. They find that black women from the cohort earn 73% (\$27 200) more than all black women with BAs, black men earn 82% (\$38 200) more than the average earnings of all black holders of BA degrees, and that white graduates of these selective institutions also have significant intraracial income advantages, but slightly less so than their black counterparts. In addition, they find large fractions of both black and white graduates from elite schools earning elite salaries, even at relatively young ages. Furthermore, for graduates whose grades were in the bottom third of their class, the black-white differential was outstanding, \$68 500–83 200,

while the top- and middle-level grade earners were about even. Quite clearly, there are income benefits for the alumni of selective institutions compared to alumni of less-selective institutions, but these degrees do not generate income equality among alumni with similarly selective credentials.

In addition, the nation where education is received is crucial in understanding some nonwhite income differences. For example, Zeng and Xie (2004) find no earnings difference across American-born white and Asian Americans compared to US-educated Asian immigrants. In contrast, foreign-educated Asian immigrants earn roughly 16% less than the aforementioned groups. We can thus infer that degree nativity matters.

Turning to another example of variation in the benefits of higher education, in *The Game of Life: College Sports and Educational Values*, Shulman and Bowen (2001) find that for all three surveyed cohorts of the College and Beyond Database, college athletes from top schools earn more than their student peers from the same elite universities. Furthermore, they analyze that the positive earnings differential is surprising because male athletes typically have lower test scores, lower socioeconomic origins, are more likely to belong to minority groups, and have lower college grade point averages (GPAs), than the average surveyed student, all of which are negative indicators of future earnings. Shulman and Bowen explain that athletes' greater presence in the for-profit sector and self-employment may explain a significant amount of the higher-earnings outcomes.

In their analysis of athletes, Shulman and Bowen do not mention the existence of racialized wage differences even though this is a featured section in Bowen's other co-authored book, *The Shape of the River*, using the same College and Beyond data set. Despite Shulman and Bowen's reluctance to explore the issue, other historic and current analyses demonstrate the probability of racially unequal benefits for collegiate athletes. Throughout several publications in the 1960s, sociologist Edwards (2002) surveyed the relationship of black athletes to universities as a site of exploitation, discrimination, and an instrument for maintaining inequality. To Edwards and others, what had resulted from attempts at integration was the creation of opportunities that did not essentially change racially unequal outcomes. This debate rages on, with recent surveys consistently finding that white and black athletes have unequal academic outcomes and graduation rates. Even among white and black athletes on the same collegiate teams, the racial difference in graduation rates are often severe and consistently point to intense racial inequality even among students of the same institutions.

In summary, both the historical and current social science literature have found college-educated African

Americans to be substantially better off than less-educated members of their race, but that they receive lesser income and occupational benefits than their white peers for their degree attainment.

The Experiences of Nonwhites in Corporate America

As reviewed in the previous section, the unequal payoff of a collegiate credential has been a persistent problem throughout American history, but some evidence indicates that the inequalities have lessened. In the literature, income and occupational prestige are consistently used to measure inequality. How ethnic minorities experience their workplace environment is important to understand the possible connections between education, income, occupational prestige, and, ultimately, economic well-being. This section thus reviews the social science literature covering ethnic minority work experiences. Success, failure, rewards, and/or the lack thereof are ultimately translated into long-term economic consequences for individuals, with work and career serving as the primary (and well-studied) mechanism through which the payoff and benefits of higher education are distributed.

There is a rich American literature analyzing ethnic disadvantage in relation to the experiences of the highly educated, upper class, and managerial elite. This literature suggests that ethnic minorities are intensely marginalized and pushed away from the power elite, despite their efforts to obtain such status. However, many argue that elite occupations have become more open since the civil rights movement of the late 1950s and 1960s, and that being nonwhite no longer disqualifies one from membership. This section surveys descriptions of nonwhite experiences in white corporate America. Edwards' characterization that attempts to integrate sports has resulted in limited islands of black opportunity appropriately summarizes findings about the integration experiences of nonwhite college graduates in the workplace.

Some scholars counter with arguments and findings of how nonwhites can now succeed in the predominantly white corporate world. In *The Declining Significance of Race*, Wilson (1978) argues that education, socioeconomic status, talent, and merit are becoming increasingly important to the life and career chances of African Americans. He argues that these historic changes are related to both national economic expansion and increasing quantities of white-collar positions, as well as to public pressures and affirmative action policy programs. Wilson, much like Freeman (1976) in *The New Black Elite*, argues that educational credentials and professional opportunities dictate one's circumstances more than discrimination. Consequently, this scenario envisions that for those with adequate human and social capital, including credentials, the opportunity to

succeed is an accessible and open door. Wilson's argument also insinuates that education should translate into economic opportunities that ultimately lead to wealth.

Wilson's 1978 thesis has many virtues, including its descriptive utility for contextualizing historical change (Wilson, 1978). However, it has been at the center of a social science debate since its publishing, and much research has contradicted it by outlining limited progress during the 1980s and 1990s as well as documenting black exclusion from elite economic rewards. In *How Capitalism Underdeveloped Black America*, Marable (1983) argues that blacks in white business, and even black businesses, all operate at the subservient pleasure of white interests. The combined total value of the 1060 black businesses in America with more than 20 employees was less than just the cash assets of a major individual white-owned and dominated corporation. Thus, according to Marable, black elites are especially vulnerable and subordinate to white businesses because black businesses are predisposed to failure because of their comparative lack of capital and limited market access to servicing larger white markets. Marable's assertions that the black corporate elite exist at the behest and pleasure of whites, and that they are comprised of less than 200 people, are powerful statements about segregation and exclusion at the top of the economic system. Marable's description of access to elite positions and wealth as closed to African Americans, regardless of qualifications and credentials (with the exception of the symbolic 200 beneficiaries), challenges the thesis of Wilson's *The Declining Significance of Race*.

In 1999, Zweigenhaft and Domhoff conducted a C. Wright Mills-inspired survey of the American power elite, with a detailed focus on the corporate elite. Their findings support the islands of opportunity model, in contrast to Wilson's depiction of expanding white-collar positions, and largely compliment Marable's thesis. Zweigenhaft and Domhoff (1991) find staggeringly low numbers of nonwhites in high-level management positions and much higher – although not populationally proportionate – representation on corporate boards. Concerning their outlook on the future, they cite data from 1990 that indicate that blacks in high-level management positions had increased from 0.2% to 0.6% from the 1980s to the 1990s. Subsequently, they recommend appropriating the glass-ceiling metaphor, first used to describe limits to the mobility of white women, and now to describe the continuing barriers for black men and women in the corporate world.

In addition, they find that Hispanics are less than 1% of all corporate directors, despite composing 10% of the population, while Asian Americans only rose from 0.2% to 0.5% of all corporate directors from 1992 to 1995, despite being a higher proportion of the population. Blacks may be better represented as board members than Latinos and Asians, (this representation is still less than proportional in

terms of population and does not extend to management positions) because of several reasons:

1. a lack of full participation in all areas of corporate life;
2. African Americans are manipulated by white executives for symbolic purposes, as Marable argues;
3. the success of targeted programs to promote minorities; and
4. the effect that interlocking directorates facilitate repeated appearances by the same black elite on multiple boards so that, as Marable describes, the same black elites are members of multiple corporations while few other blacks are high-level employees.

Zweigenhaft and Domhoff (1991) also conducted a study of 38 participants in the A Better Chance program, which places disadvantaged black youth from all over the United States at elite private boarding schools. While 99% of the participants in the program went to colleges, which is far above the national average for black youth, the program is commonly judged successful and lauded because so many alumni attended the most elite universities and were more successful than both their siblings and peers from their home communities who had not participated in the program. However, many alumni had changed positions, moved to work at black-owned businesses, or started their own businesses in order to escape the racial job ceilings they found in mainstream corporate America. This trend in the career trajectories of these highly educated blacks led the authors to conclude that the power elite will likely remain the least integrated segment of American society. Despite establishing substantial lifelong networks with whites at the most elite high schools and colleges, gaining the same elite education credentials and even having high intermarriage rates, these African Americans unequivocally believed that they were marginalized in the workplace.

These experiences are consistent with other findings. For example, in *The Marginalization of Black Executives*, Sharon M. Collins explores “the careers of 76 of the top-ranking black executives employed in Chicago white-owned corporations.” She finds that black upward mobility in the otherwise white management has been restricted, with these workers largely filling human resource and public relations jobs that were recently created in response to public demands. In sum, this situation leads to the black executives being marginalized away from central decision making and thus to dead-end careers. This undermines the intentions of the social and political pressures that employment opportunities would enable upward mobility for the educated black elite. Instead, highly educated and capable black executives experienced new, less obvious, but nonetheless substantial employment inequality.

Although Wilson argues that class and educational credentials are more important than discrimination in getting African Americans access to the means of production, a line of literature, including Collins, has found that African

Americans are systematically all but excluded from reaching the top of the corporate hierarchy. Similarly, this literature finds that African Americans who achieved upward promotion in the corporate sectors were often channeled into dead end, racially oriented jobs that not only did not match their qualifications, but also derailed their career trajectory by excluding them from power and decision-making roles. Such findings have many implications about the comparative economic well-being of college alumni. Exclusion from elite, decision-making positions indicates that African Americans are likely to have dissimilar career trajectories, and thus dissimilar incomes and wealth outcomes, than white peers with the same educational credentials. As a result, African Americans have a much lower chance of gaining extreme income or ever accruing extreme wealth. This directly relates to the literature on intergenerational transmissions of wealth, indicating that historically unequal intergenerational transmissions of wealth will likely continue, even among the educated elite.

Wealth and wealth inequality are the culminating consequences of unequal payoffs to education, effectively shaping long-term economic well-being. The literature has consistently found that income inequality is the tip of an economic inequality iceberg, with wealth inequality significantly greater than wage inequality. As discussed previously, many scholars have articulated a theoretical and empirical link between higher education and substantial economic rewards for African Americans, but some have called for more investigations into wealth to gain a fuller understanding of economic life, well-being, status, and the payouts to educations. Wealth has emerged as a holistic measurement of economic well-being in the social sciences. Racial wealth inequalities have been found to be exponentially greater than income inequality, to the extent that the highest-income brackets of African Americans are found to have the same wealth as the lowest-income earning whites.

Oliver and Shapiro (1998, 2006) provide the most recent and comparable analyses regarding wealth, race, and education. They find that African Americans with college degrees possess just 23 cents for every dollar of wealth owned by similarly educated whites. While blacks’ net worth increases with education, whites gain far more wealth with almost every level of additional degree attainment in both absolute and relative measurements over similarly educated blacks.

We can hypothesize that the intensity and persistence of racial wealth inequality even among college graduates is likely associated with different mechanisms of cumulative disadvantage as follows:

1. a stratified history of property relations, including racialized historic and persisting political, legal, and economic obstacles to home ownership;
2. systems of neighborhood segregation that devalue black-owned property and translates neighborhood

segregation into various disadvantages, hazards, privileges, and/or opportunities to residents according to their race and location;

3. unequal payoffs for higher education that are associated with the marginalization of and discrimination against nonwhites at the workplace; and
4. wealth mobility differs by race.

These mechanisms help perpetuate the racial stratification in wealth observed in national samples, regardless of education. As evidenced by the continued wealth disparity, even between families with similar incomes, it is necessary to look beyond income to understand economic well-being, racial inequality, and the payoffs of education. While the American higher education system strives to provide equal access and opportunities for all, college-degree attainment does not make its alumni economic equals.

See also: Physical Education and Sports; Race and Ethnicity in the Field of Adult Education.

Bibliography

- Blau, P. and Duncan, O. D. (1967). *The American Occupational Structure*. John Wiley and Sons.
- Bowen, W. G. and Bok, D. (2000). *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions*. Princeton University Press.
- Edwards, H. (2002). The man who would be King in the sports world. Special to espn.com, *Black History Month* February 28. <http://espn.go.com/gen/s/2002/0226/1340982.html> (accessed October 2009).
- Freeman, R. (1976). *Black Elite: The New Market for Highly Educated Black Americans: A Report Prepared for the Carnegie Commission on Higher Education*. McGraw-Hill.
- Marable, M. (1983). *How Capitalism Underdeveloped Black America: Problems in Race, Political Economy, and Society*. South End Press.
- Myrdal, G. (1944). *An American Dilemma: The Negro Problem and Modern Democracy*. Harper & Row.
- Shulman, J. L. and Bowen, W. G. (2001). *The Game of Life: College Sports and Educational Values*. Princeton University Press.
- Wilson, W. J. (1978). *The Declining Significance of Race: Blacks and Changing American Institutions*. University of Chicago Press.
- Zhen, Z. and Xie, Y. (2004). Asian-Americans' earnings disadvantage reexamined: The role of place of education. *American Journal of Sociology* 109(5).
- Zweigenhaft, R. L. and Domhoff, G. W. (1991). *Blacks in the White Establishment? A Study of Race and Class in America*. Yale University Press.

Zweigenhaft, R. L. and Domhoff, G. W. (1999). *Diversity in the Power Elite: Have Women and Minorities Reached the Top?* Yale University Press.

Further Reading

- Allen, W. R. (1992). The color of success: African American college student outcomes at predominantly White and historically Black colleges and universities. *Harvard Educational Review* 62(1), 26–44.
- Allen, W. R., Epps, E. G., and Haniff, N. Z. (eds.) (1991). *College in Black and White: African American Students in Predominantly White and in Historically Black Public Universities*. State University of New York Press.
- Arrow, K. J., Bowles, S., et al. (2000). *Meritocracy and Economic Inequality*. Princeton University Press.
- Arum, R., Gamoran, A., et al. (2007). *Stratification in Higher Education: A Comparative Study*. Stanford: Stanford University Press.
- Bourdieu, P. (1979, 1996). *Distinction-A Social Critique of the Judgement of Taste*. Routledge.
- Collins, R. (1979). *The Credential Society: An Historical Sociology of Education and Stratification*. Academic Press.
- Du Bois, W. E. B. (1903). *The Negro Problem: A Series of Articles by Representative American Negroes of Today*. Kessinger Publishing.
- Frankenberg, E., Lee, C., and Orfield, G. (2003). A multicultural society with segregated schools: Are we losing the dream? *Harvard Civil Rights Project*.
- Gurin, P. (1999). *Expert Report of Patricia Gurin, in the Compelling Need for Diversity in Higher Education*. Gratz et al. v. Bollinger et al., No. 97-75321 (E.D. Mich.) Grutter et al. v. Bollinger et al., No. 97-75928 (E.D. Mich.). Ann Arbor, MI: University of Michigan.
- Ogbu, J. U. (1974). *The Next Generation: An Ethnography of Education in an Urban Neighborhood*. Academic Press.
- Oliver, M. and Shapiro, T. (1995). *Black Wealth/White Wealth: A New Perspective on Racial Inequality*. Routledge.
- Pattillo-McCoy, M. (1999). *Black Picket Fences: Privilege and Peril among the Black Middle Class*. University of Chicago Press.
- Sorokin, P. (1927). *Social Mobility*. Harper and Brothers.

Relevant Websites

- <http://www.choices.gseis.ucla.edu> – CHOICES Project: Access, Equity and Diversity in Higher Education.
- <http://www.demos.org> – inequality.org.
- <http://www.tcrecord.org> – Teachers College Record.
- <http://www.law.berkeley.edu> – The Chief Justice Earl Warren Institute on Race, Ethnicity and Diversity.
- <http://www.civilrightsproject.ucla.edu> – The Civil Rights Project.
- <http://www.tidesport.org> – The Institute for Diversity and Ethics in Sport.
- <http://www.usccr.gov> – US Commission on Civil Rights.
- <http://sociology.ucsc.edu/whorulesamerica/power/wealth.html> – Who Rules America, G. William Domhoff.

School Health Education

J S Yang, California State University, Fullerton, Fullerton, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Health and Education

Healthy People 2010, the US government's decennial statement of its national health goals, set a target for high school completion among persons aged 18–24 at 90% in the year 2010 (USDHHS, 2000), consistent with the national education goals. The inclusion of a high-school-completion goal in a government-agenda-setting publication for health underscores the well-established and enduring positive relationship between health and education. Dropping out of school is associated with increased risk of substance abuse, delinquency, intentional and unintentional injury, and unintended pregnancy. There is a dose–response relationship between health and education: the more years of education a person has obtained, the more likely they are to report being in excellent or good health. The relationship between health and education is positive independent of the effects of family income, age, or poverty status.

Although the relationship between health and education is enduring, evidence of a causal effect of one on the other is elusive. Health status is thought to affect educational attainment, while educational attainment has a clear impact on health. Instead of declaring either health or educational attainment as the causal agent, it is most accurate to describe health and educational attainment as intertwined.

There are three major hypotheses to explain the observed effect of child and adolescent health on educational attainment. The first is based on the status-attainment model that views both educational attainment (Sewell *et al.*, 1969) and child health (Aber *et al.*, 1997) as primarily dependent on parental occupation and position in the social hierarchy. This hypothesis argues that parental social status determines both adolescent health and educational attainment. The emerging field of life-course health development (Halfon and Hochstein, 2002) – which suggests experiences in critical life stages and cumulative life experiences affect adolescent and adult health – may hold some explanatory power for this hypothesis. Numerous analyses, however, have shown that the positive relationship between adolescent health and educational attainment endures after controlling for parental socioeconomic status.

A second hypothesis suggests that poor health has a negative effect on cognitive development and functioning, leading to lower educational attainment. The evidence to support this hypothesis, however, is weak and no systematic studies to examine the relationship between adolescent

health, cognition, and educational attainment have been conducted.

A third hypothesis posits that health limitations have detrimental psychological, behavioral, and social effects on children and adolescents, leading to lower educational achievement. Lower levels of self-esteem, attachment to school, commitment to schooling, and greater absenteeism have been associated with poorer health in adolescents. There is a growing literature on the psychosocial impact of poor health on adolescents and school performance, though the type and magnitude of effect varies by disease or behavioral risk factor.

Given the complex interaction among family context, school environment, and disease pathology, the effects of poor health on educational attainment take on any number of pathways. The health conditions and behaviors that adolescents deal with on a daily basis demonstrate how the general notion of poor health includes any number of specific impacts on educational attainment. Asthma, for example, has been linked to increased school absenteeism, with overweight and obese children more likely to have low self-esteem and higher rates of anxiety disorders, depression, and other mental health conditions. Insufficient or interrupted sleep is associated with reduced attention, memory, intelligence, and cognitive ability, and undernutrition is associated with poor concentration and greater irritability and apathy. There remains a paucity of research which looks directly at specific diseases and related morbidities as mediators between health and educational attainment. Emerging evidence, however, provides basis to believe the effect of health on educational attainment is mediated through diverse pathways.

School Health Education

With over 55 million and 95% of all children aged 5–17 years enrolled in American schools (Snyder *et al.*, 2009), school health education is a unique opportunity to improve the health and educational outcomes of children and adolescents. School health education is a subset of health education, a field concerned with educational interventions that can affect and facilitate adoption of behaviors conducive to health, which are focused on schools (Green and Iverson, 1982). Though in principle school health education is not limited to classrooms, didactic methods, and students, in practice, these are the settings, pedagogy, and audience that school health-education programs commonly employ and target.

Health education targets three categories of factors that affect health behavior (Green and Iverson, 1982): predisposing factors, enabling factors, and reinforcing factors. School health education typically focuses on affecting predisposing factors that orient adolescents toward a particular behavior. These include beliefs, knowledge, attitudes, and values. In, for example, school health-education programs to prevent tobacco use, predisposing factors such as knowledge of the health effects of smoking, beliefs about the acceptability of smoking among peers, and values toward exposing others to the dangers of second-hand smoke must be addressed. Changing predisposing factors like knowledge and attitudes is a necessary but insufficient component of school health education.

Enabling factors, such as skills acquisition, critical thinking, effective decision making, and management of peer influence, allow adolescents to carry out health-enhancing behaviors. In school smoking-prevention programs, skills such as reading cigarette-constituent labels, critical thinking about cigarette advertisements, and decision making on accepting cigarette-promotional items should be taught.

Reinforcing factors influence whether a behavior that has been tried will be repeated. They are material, emotional, and social in form, penalizing or rewarding individuals for engaging in particular behaviors. For example, peers can offer or withhold the benefits of social inclusion for smoking cigarettes. Reinforcing factors are considered more influential than predisposing and enabling factors on health behavior. Since they are the factors most beyond the ability of the school to control, school health-education programs are limited to helping students learn how to resist the array of factors that encourage adoption of unhealthy behaviors.

National Health Education Standards

In response to the absence of an explicit health goal in the national education goals, the Joint Committee on National Health Education Standards was formed to develop standards for health education. The committee was made up of representatives from the numerous health organizations and sponsored by the American Cancer Society. The resulting National Health Education Standards (NHES) provide a framework for health-education curricular development (Joint Committee on National Health Standards, 2007). Instead of approaching health education from the perspective of predisposing, enabling, and reinforcing factors, the NHES focuses on improving health literacy among students. Performance indicators for four age categories – pre-K through second grade, third through fifth grade, sixth through eighth grade, and ninth through twelfth grade – guide the development of age-appropriate health-education curricula.

School Health-Education Content Areas

The School Health Education Study, conducted from 1964 to 1972, introduced the idea of comprehensive health-education

programs. The results of the study identified ten content areas that provided the basis for health-education curricular development through the 1980s.

More recently, the US Centers for Disease Control and Prevention (CDC) has identified six health areas responsible for over 70% of illness, disability, and death among adolescents and young adults: intentional and unintentional injuries, tobacco use, alcohol and illicit drug use, sexual behaviors that cause unintended pregnancies and sexually transmitted diseases, dietary patterns that cause disease, and sedentary lifestyle (Kann *et al.*, 1995). As these areas are responsible for such a significant burden of disease among adolescents and young adults, they are priority-substantive areas for health-education programs.

Comprehensive Health Education Programs: Best Practice

Taking predisposing, enabling, and reinforcing factors; health literacy; and the core content areas together, the CDC has outlined a framework for developing a comprehensive school health-education program (Collins *et al.*, 1995):

1. A documented, planned, and sequential program of health education for students in grades K through 12.
2. A curriculum that addresses and integrates education about a range of categorical health problems and issues.
3. Activities to help young people develop the skills they will need to avoid behaviors that result in unintentional and intentional injuries; alcohol and other drug use; tobacco use; sexual behaviors that result in human immunodeficiency virus (HIV) infection, other sexually transmitted diseases (STDs), and unintended pregnancies; imprudent dietary patterns; and inadequate physical activity.
4. Instruction provided for a prescribed amount of time at each grade level.
5. Management and coordination in each school by an education professional trained to implement the program.
6. Instruction from teachers who have been trained to teach the subject.
7. Involvement of parents, health professionals, and other concerned community members.
8. Periodic evaluation, updating, and improvement.

Time allotted to school health education and appropriate training of health-education teachers are two critical areas for effective health-education programs. The exact amount of time needed for health-education programs to produce desired outcomes is unclear. But brief curricular exposure does not produce changes in behavior. Sustained exposure to a health-education program and follow-up sessions after program participation are needed for stable improvements in behavior. Estimates suggest that at least 1.8 h per week are needed for changes in student knowledge and attitudes. Health knowledge has been observed to

increase after 15 h of exposure to health-education programming. Behavior begins to change after 50 h of program exposure, and maximal change in knowledge, attitudes, and behavior occurs after 60 h of exposure.

As health-education teachers are not only expected to provide knowledge but also influence present and future behavior, well-trained teachers are integral to effective health-education programs. Training opportunities for health-education teachers are numerous. Professional associations are available for information regarding health teacher training. Professional development can be accessed through programs offered by universities and through conference attendance, professional journals, and online courses. Even with increasing access to opportunities for professional development, there is a shortage of health-education teachers with a background in health education.

Coordinated School Health Programs

Early health-education programs focused on hygienic behaviors that would help stem the transmission of communicable diseases. As the conditions that afflicted children and adolescents shifted to chronic, noncommunicable, multifactorial diseases and conditions, the limits of traditional health education became apparent. Didactic strategies alone could not meet the behavioral goals of school health programs. Over time, the economic, social, and political – as opposed to an individualistic and behavioral – determinants of health became more widely accepted as a part of health education. As they were, school health education became embedded within the broader framework of coordinated school health programs.

Coordinated school health programs approach child and adolescent health as a multisectoral enterprise to be shared by health, medical, educational, and social-service organizations. Working together, multiple agencies provide students with a comprehensive set of programs and services to prevent disease and promote health.

One model includes eight components in a coordinated school health program (**Figure 1**) (Fisher *et al.*, 2003). Access to health services through direct care or referral services allows for control and management of chronic diseases and provision of emergency services. Counseling, psychological, and social services provide resources to improve student mental and emotional health and ensure a healthy social environment. Health promotion for staff includes health education and activities that can improve health status, morale, and commitment to the school's overall health program. A healthy school environment ensures that the design and structure of the school environment is optimal, protects students from biological or chemical agents damaging to health, and enhances the psychological, emotional, and social culture of the school. Access to nutritious and appealing meals

that meet the nutritional needs of all students are part of appropriate nutritional services. A high-quality physical-education curriculum provides cognitive content and experiential activities in a variety of movements. Involvement of family and community actors helps build support for coordinated school health programs while helping to build reinforcing factors for healthy behavior outside of schools.

School health-education programs are most effective when they not only follow best practices but are located within the broader context of a coordinated school health program. The access to resources, school culture, organizational structure, and prioritization of health created by coordinated school health programs amplify the impact of school health education. Coordinated school health programs provide the context to effectively translate change among proximal targets of school health education – knowledge, beliefs, attitudes, skills, critical thinking, and decision making – into health-promoting behavior change.

Tobacco-Use Prevention in Schools: A Cautionary Tale

Consequences of Adolescent Smoking

As the leading cause of preventable death in the United States, tobacco use is a core content area for school health education. Every day, almost 5000 children in the United States try their first cigarette (Fisher *et al.*, 2003). Childhood experimentation often leads to adult addiction because of the presence of nicotine in tobacco. Concern over child adolescent tobacco use is threefold. First, tobacco use is associated with lower levels of educational achievement and school involvement (though the causal direction is unclear). Lower levels of achievement may be mediated by behavioral problems associated with smoking among adolescents, which include fighting, carrying weapons, attempting suicide, and engaging in high-risk sexual behavior. Second, the health effects of smoking include increased cough, higher risk for respiratory illness, increased likelihood of developing risk factors for cardiovascular disease, decrease in levels of physical fitness, and retardation of lung growth and optimal lung function. Use of smokeless (i.e., chewing) tobacco is associated with health concerns ranging from bad breath to oral cancer. Third, a vast majority of adult smokers begin before they leave high school. If students leave high school as nonsmokers, they are highly unlikely to begin smoking as adults. Thus smoking prevention among adolescents has enduring health and attainment implications into adulthood.

Tobacco-Use Prevention Programs

Educational interventions for tobacco-use prevention have evolved over the second half of the twentieth century (USDHHS, 2002). Early interventions relied on an

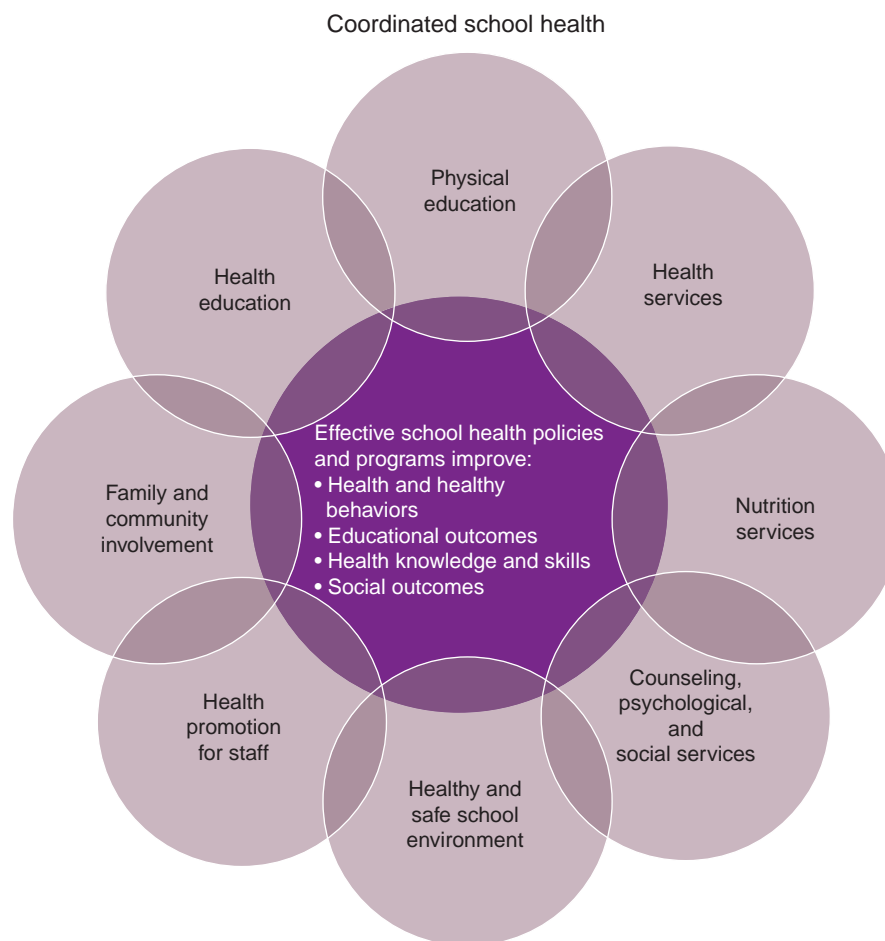


Figure 1 Coordinated school health programs. From US Centers for Disease Control and Prevention.

information-deficit model, assuming provision of information regarding the harms of tobacco use would prevent uptake and spur cessation. A second approach incorporated intrapersonal factors such as self-esteem and improving attitudes toward family, school, and community efforts to bridge information gaps. While programs based on both approaches were shown to increase knowledge, they were ineffective in preventing smoking initiation.

Over time, smoking initiation was found to be a social behavior influenced by peer smoking, smoking by others in the environment, and other social and psychological factors. These insights led to a third orientation to tobacco-use prevention programs called the social-influences model. Programs based on this model aim to equip adolescents with the skills and resources to help resist offers to take up tobacco use. Specific strategies include learning the short-term negative social and health consequences of tobacco use, reframing tobacco use as a behavior of a minority of students, recognizing the social influences of smoking, and developing skills to resist social pressures to use tobacco.

Although early social influences-based programs had limited effects, recent studies have shown more consistent

results. Contemporary programs using the social-influences mode begin during the sixth or seventh grade and utilize classroom activities over multiple sessions for a duration of between 1 and 3 years. The effect of these programs on smoking behavior has been larger and longer lasting than earlier studies, but sustaining program effects on behavior to the end of high school has been difficult.

Curricula focusing solely on skills training to resist peer influence may be ineffective. Modifying perceptions of peer and adult smoking and denormalizing smoking behavior have more powerful effects than skills training. Combined curricula result in the greatest impact. As a result, some tobacco-use prevention programs expand intense social-influence-based school programs to incorporate community education toward adults, mass media, parent education, and involvement in program activities and community organizing.

The CDC has issued seven recommendations for a school health programs to prevent tobacco use (CDC, 1994):

1. develop and enforce a strong school policy on tobacco use that is consistent with state and local laws;

2. provide instruction about the short- and long-term negative health effects and social consequences of tobacco use, social influences on tobacco use, peer norms regarding tobacco use, and refusal skills;
3. provide tobacco-use-prevention education in kindergarten through twelfth grade with more intensive education in junior high or middle school reinforced in high school;
4. provide program-specific training for teachers;
5. involve parents or families in school-based programs for tobacco-use prevention;
6. support cessation efforts among students and school staff; and
7. assess the tobacco-use-prevention program at regular intervals.

The Tobacco Industry and School-Based Programs

In 1997, the CDC concluded that the two curricula, Life Skills Training (LST) and Project Toward No Tobacco Use were effective in reducing adolescent tobacco use. This announcement was seized upon by the tobacco industry, in particular by Philip Morris (PM) and Brown and Williamson (B&W), to expand their youth smoking-prevention programs (Mandel *et al.*, 2006). The tobacco industry has a long-standing interest in promoting its youth tobacco-use prevention programs. They are public-relations tools to help ease public suspicion that the industry markets tobacco products to children. The industry also uses its voluntary youth smoking-prevention programs to stave off strong legislation to restrict activities of the industry with respect to adolescents, including advertising and promotions.

Tobacco industry youth smoking-prevention programs target youth, parents, tobacco retailers, and youth organizations, and are disseminated through multiple channels such as media campaigns, brochures, public gatherings, stickers and posters, and third-party organizations – including the National Association of State Boards of Education (Landman *et al.*, 2002). Tobacco industry documents reveal that youth smoking-prevention programs send messages consistent with their advertising campaigns and are used to displace more aggressive educational programs. They lack components, such as mass-media campaigns that de-normalize tobacco use, price increases for tobacco products, and smoke-free policies to decrease the social acceptability of smoking, which have proven to be effective in reducing youth smoking. Industry programs do not prevent, and may even encourage, youth smoking.

The promotion of the LST curriculum for youth smoking prevention followed the general pattern of industry youth smoking-prevention programs. It was promoted by PM and B&W because it steered attention away from programs focused on the manipulative behavior of the tobacco

industry, a theme of successful state programs in California and Florida. Instead, it wanted a program that stressed positive youth development and allowed the industry to be seen as a partner of health and education organizations, strengthen relationships with youth organizations, and build the industry's image as responsible and reputable. LST was particularly attractive because it had been endorsed by the CDC, a credible government agency, as an effective tobacco use-prevention program.

PM and B&W actively promoted LST among school superintendents, school board members, legislators, and educators using a public-affairs consulting firm. PM and B&W also set up a grant-matching mechanism for state departments of education or health to use money from the Master Settlement Agreement, a legal settlement between the tobacco industry and 46 state attorneys general, for LST.

As PM and B&W promoted LST, a second consulting firm provided data to the companies that showed that the LST programs they were sponsoring had a statistically significant effect on knowledge about the physiological effects of smoking, no impact on social acceptability of smoking, and reduced decision-making skills after 1 year. Results after 2 years of LST implementation were no better. PM and B&W were promoting a program they knew was ineffective.

On the face of it, PM and B&W's promotion of an LST, a CDC-approved tobacco-use-prevention curriculum, appeared to be a positive development to improve adolescent health. It was not, however, because the real-world effectiveness of LST was limited to improvements in knowledge compared to the highly controlled studies that the CDC used to deem LST an effective curriculum. Adoption of LST was also problematic because it supplanted the implementation of programs shown to be effective in real-world contexts. In the broader struggle between the tobacco industry and public health advocates, school health education was co-opted to serve the business goals of an industry with decidedly negative impacts on health.

The case of the tobacco industry and LST highlights three points about school health education:

1. *Careful attention must be paid to the development and implementation of school health education programs.* Curriculum development for health education differs markedly from other subjects because the aim is not only to impart knowledge but to change behavior. Yet, the low stature of health on the educational-research landscape results in a lack of proven and effective strategies to achieve health-education goals. Continued research must be conducted to improve school health-education curriculum. For example, anti-industry messages and de-normalization of tobacco use are effective approaches to tobacco-use prevention that fall outside of the standard social-influences model. Additionally, training of

schools and teachers to implement effective curricula must improve. In the case of tobacco and LST, highly controlled studies showed curricular effectiveness on behavior. But schools taking part in the PM and B&W program did not demonstrate equivalent results due to process factors.

2. *School health-education programs must be complemented with other activities outside of school.* School health education is useful for affecting knowledge, but is not sufficient to change adolescent health behavior. Health education must be embedded within a coordinated school health program to yield optimal results, with investment from family and community actors. In the case of tobacco, anti-industry mass-media campaigns have been effective tools for tobacco-use prevention as have tax increases on tobacco products and smoke-free policies in workplaces, schools, and other indoor areas. The knowledge impacts of tobacco-use-prevention programs can be expanded to behavior change when accompanied by other interventions.
3. *Schools must carefully select partners for health education.* When partnering with and finding support from individuals and organizations for health-education programs, schools must collaborate selectively. The case of the tobacco industry is an obvious case of conflict of interest and extreme deception. Though the industry portrayed itself as concerned about youth smoking, it did so for public relations and regulatory purposes with a goal of delaying, not preventing, smoking initiation. Schools must be aware of the competing interests of various organizations and develop school health education and coordinated school health programs with partners whose mission and purpose align with those of the school.

Summary

Good health is a necessary condition for educational achievement. The challenges to health for children and adolescents are complex and wide ranging, requiring coordinated effort on the part of health and educational systems to overcome. School health education is an integral part of this process. A solid conceptual foundation for effective school health education exists in national standards, prescribed best practices, and technical-assistance resources at the state and national level. When best practices are implemented, school health education can change predisposing and enabling factors needed to ensure adolescent health. But school health education alone is not sufficient to observe sustained behavioral change. A larger health-enhancing enterprise must be undertaken by schools to support the behavior-change goal of health education. As the case of school tobacco-use-prevention programs show, effective school health-education programs must be developed, correctly implemented, and complemented

with efforts in families, communities, and at the broader policy level to ensure that health does not act as a barrier to but enhances student-achievement potential.

See also: Education and Health; Educational Attainment and Mortality Differentials; Evaluation of Integrated Health Programs in School; Health and Adult Learning; Investing in Early Childhood Education and Care: The Health and Wellbeing Case; Technology and Physical and Social Health; The Health Advantages of Educational Attainment.

Bibliography

- Aber, J. L., Bennett, N. G., Conley, D. C., and Li, J. (1997). The effects of poverty on child health and development. *Annual Review of Public Health* **18**, 463–483.
- CDC (Centers for Diseases Control and Prevention) (1994). Guidelines for school health programs to prevent tobacco use and addiction. *MMWR* **43**, no. RR-2.
- Collins, J. L., Small, M. G., Kann, L., et al. (1995). School health education. *Journal of School Health* **65**, 302–311.
- Fisher, C., Hunt, P., Kann, L., et al. (2003). Building a healthier future through school health programs. In Dalmat, M. E. and Wheeler, F. C. (eds.) *Promising Practices in Chronic Disease Prevention and Control*. Atlanta, GA: Centers for Disease Control and Prevention.
- Green, L. W. and Iverson, D. C. (1982). School health education. *Annual Review of Public Health* **3**, 321–338.
- Halfon, N. and Hochstein, M. (2002). Life course health development: An integrated framework for developing health, policy, and research. *Milbank Quarterly* **80**, 433–479.
- Kann, L., Collins, J. L., Pateman, B. C., et al. (1995). The School Health Policies and Programs Study (SHPPS): Rationale for a nationwide status report on school health programs. *Journal of School Health* **65**, 291–294.
- Landman, A., Ling, P. M., and Glantz, S. A. (2002). Tobacco industry youth smoking prevention programs: Protecting the industry and hurting tobacco control. *American Journal of Public Health* **92**, 917–930.
- Mandel, L. L., Bialous, S. A., and Glantz, S. A. (2006). Avoiding “truth”: Tobacco industry promotion of life skills training. *Journal of Adolescent Health* **39**, 868–879.
- Sewell, W. H., Haller, A. O., and Portes, A. (1969). The educational attainment and early occupational attainment process. *American Sociological Review* **34**, 82–92.
- Snyder, T. D., Dillow, S. A., and Hoffman, C. M. (2009). *Digest of Education Statistics 2008*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education.
- USDHHS (US Department of Health and Human Services) (2000). *Healthy People 2010: Understanding and Improving Health*, 2nd edn. Washington, DC: US Government Printing Office.
- USDHHS (US Department of Health and Human Services) (2002). *Reducing Tobacco Use: A Report of the Surgeon General*. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Further Reading

- Allensworth, D., Lawson, E., Nicholson, L., and Wyche, J. (1997). *Schools & Health: Our Nation's Investment*. Washington, DC: National Academies Press.
- Birdthistle, I. and Health and Human Development Programs (eds.) *Improving Health through Schools: National and International Strategies*. Geneva: World Health Organization.
- Glanz, K., Rimer, B. K., and Viswanath, K. (eds.) (2008). *Health Behavior and Health Education*, 4th edn. San Francisco, CA: Jossey-Bass.

- Joint Committee on National Health Standards (2007). *National Health Education Standards PreK-12*, 2nd edn. Atlanta, GA: American Cancer Society.
- Symons, C. W. and Cinelli, B. (1997). Bridging student health risks and academic achievement through comprehensive school health programs. *Journal of School Health* **67**, 220–227.
- WHO Expert Committee on Comprehensive School Health Education and Promotion (1997). *Promoting Health through Schools: Report of a WHO Expert Committee on Comprehensive School Health Education and Promotion*. Geneva: World Health Organization.

Relevant Websites

- <http://www.aahperd.org> – American Association for Health Education.
- <http://www.cdc.gov> – Division of Adolescent and School Health, National Center for Chronic Disease Prevention and Health Promotion, Center for Disease Control and Prevention.
- http://www.who.int/school_youth_health – World Health Organization School Health and Youth Health Program.

DEMOGRAPHY AND SOCIAL CHANGE – DEMOGRAPHY

Contents

A Paradigm of Contradictions: Racism and Science Education
Africana Studies: Past, Present And Future
Beyond Black Nationalism to Black Internationalism: The Compelling Case for Examining Black Educational Challenges Globally
Beyond Racial, Ethnic, and Gender Bias in Education Statistics
Caste and Education in India
Demography, Community, and the Education of Asian American and Pacific Islander Students
Educational Attainment and Mortality Differentials
Educational Opportunity and Latino/Chicano College Choice: New Findings, and Theoretical Perspectives
Ethnic Minority Identity and Educational Outcomes in a Rising China
Ethnic Studies and Women's Studies: From the Past to the Present, Providing a Space in the Academy for Scholarship Focused on the Histories and Experiences of Marginalized People
Gender Equity in Higher Education: Challenges and Celebrations
Locating Space and Place in the College Access Debate: New Tools for Mapping and Understanding Educational Inequity and Stratification
Public Policy and Inequality in Postsecondary Opportunity: Educational Statistics and the Failure of Education Reform
The Education of Indigenous Students
The Health Advantages of Educational Attainment
To Go or Not to Go: Access and Barriers to Ethnic Minority Participation in Higher Education within the United Kingdom
Universal Secondary Education and the Two-Tiered School System in the English-Speaking Caribbean
Women of Color in Science, Technology, Engineering and Mathematics (STEM): Refining the Concepts, Reframing the Issues

A Paradigm of Contradictions: Racism and Science Education

B A Brown, Stanford University School of Education, Stanford, CA, USA

J W Mutegi, Indiana University, Indianapolis, IN, USA

© 2010 Elsevier Ltd. All rights reserved.

Race in Science versus Race in Science Education

An examination of the role of race and racism in science education must begin with an understanding of the importance of race to the ideology and practice of Western modern science. Bell (1992) asserts racism is an enduring attribute of American culture that has maintained a prevailing position in the cultural milieu. America has a history of sustained racism that influences contemporary interactions. In other words, because racism has been a pervasive component of society at large throughout modern history, then many institutions are likely to reflect those racist patterns to some degree. This perspective

could lead to the question of how modern science, whose foundational epistemology relies on objectivity, free itself from the influence of racism.

The answer, of course, is that science has not been able to invoke immunity to racism. The fact that science is an inherently human activity means it is fraught with all the biases and tendencies humans bring into it. As a result, science as experienced in schools is impacted by issues of race and racism, much as any other Western cultural practice (Eishenhart *et al.*, 1996; Norman, 1998). Contrary to the reality of scientific practice, science education curriculum (and science education as a whole) tends to confer upon scientific practice an unwarranted mantle of 'objectivity.' By portraying the scientific enterprise and its

practitioners as objective, they have enabled racism in scientific ideology, practice, and pedagogy to hide behind the cloak of scientific objectivity.

It is the inherently contradictory notions of scientific objectivity and science as a human activity that stand at the heart of this article. Despite being influenced by racism, the science presented in schools is often viewed as objective, culturally neutral, and uninfluenced by racism. This culturally neutral representation of science continues to bring us to the same end as more overtly racist representations did decades ago. The myth of scientific objectivity enables racist conceptions of science to merely retrofit their previously overt constructions for more subtle alternatives. Specifically, it has the effect of rendering the societal structures that work against the participation of minority students in science, technology, engineering, and mathematics fields as invisible. In this article, we first argue that presentation of scientific practice as completely objective is inaccurate. We then demonstrate how scientific practice has historically been influenced, to a substantial degree, by racism and is apparent in both scientific ideology and scientific practice. We will also illustrate that historical influence, which is readily visible to the keen observer, is only a pretext for the current ideology and practices we witness today. Time has yielded little change at the core of scientific practice and ideology. Instead, it has given us a theoretically retrofitted version of science, such that it looks different on its surface, but remains unvarying at its core. We conclude with a discussion of the implications this presentation of science has for the contemporary practice and research in science education.

A History of Racism in Science

Historically, science has provided a wealth of insights and technological projects to advance contemporary culture. However, it has also served a role in marginalizing people on the basis of race and ethnicity. Norman (1998) describes how science has a history of being a marginalizing discourse. He argues that science, conducted in the name of promoted sociopolitical ideologies, needs to be framed as an alternative to science if scientific objectivity is to remain central to the construction of scientific knowledge. Norman explains:

The discourses on race and gender raise the question of scientific objectivity in very compelling ways. In the context of scientific literacy, it is absolutely crucial that the questions of fraudulent science in the interest of political ideology be treated separately from the more fundamental question of how problematic it is to achieve complete objectivity in science. (p. 371)

Norman's argument discloses how scientific research avoids addressing the sociopolitical subjectivity of the

scientist's intentions as a means to maintain the reified status of science. As a result, much science research and pseudo-scientific research designed to promote and sustain racial ideologies has been able to continue under the cloak of scientific objectivity.

Norman's assessment can be extended to science education as the perception of an objective science makes the cultural components of science invisible. Although nearly invisible in most science texts and curricula, science has historically been influenced by race and racism while being used to support racist ideologies.

Historical records of race and science extend as far back as Grecian cultures (Norman, 1998). Although the culture of the early Grecian times was more philosophical and metaphysical than empirical science, Hippocrates was accused of asserting that people with darker skin were perceived as less courageous warriors (Norman, 1998). Although not supported by scientific evidence, the biological basis of racist ideologies served as a foundation to these early philosophical Grecian arguments.

This traditional assumption formed the foundational ideology of Europeans for most of the eighteenth and nineteenth centuries (Norman, 1998; Gosden, 2006). In fact, Linnaeus provided one of the earliest conceptions of racial inferiority in his *Systema Naturae*. His 1758 text attempted to construct a systematic taxonomy of human beings. Linnaeus categorized blacks as "Sapiens Afer (the Black African) as 'ruled by caprice,' while Homo Sapiens Europeans were said to be 'ruled by customs.'" (Norman, 1997: 367). This characterization of the black African as 'ruled by caprice' (or impulsive) started the attempt of the scientific community to construct people of color as intellectually inferior.

Racial Evolutionist

After these philosophical constructions of race were presented, a series of attempts to construct Africans as inferior emerged through racial evolution theories (Norman, 1998; Morris-Reich, 2006). Racial theories embedded in the philosophy of science served to define much of the research on race in the 1800s. Arthur Schopenhauer offered theoretical perspectives that became essential to frameworks on literacy (Norman, 1998; Morris-Reich, 2006). Early scholars argued the growth of civilization was based on the contributions of whites and that led to the development of a refined north (Hasian, 1996; Larson, 1995; Norman, 1998; Morris-Reich, 2006; Tucker, 2002). This perspective completely ignored the early scientific contributions of the entire continent of Africa. Others made the argument that racially advanced whites migrated to the north in an effort to survive. They suggested that whites had to develop higher intellectual powers as a part of their struggle to survive the misery and challenges

of northern climates (Dubow, 1995). These racial evolutionist perspectives were used by early scientists to make an argument that attempted to establish whites as superior. These initial studies led to the development of more empirical approaches to research designed to ‘prove’ the racial inferiority of people of color.

Much of the pseudo-scientific research conducted by racial evolutionists was executed in an attempt to validate institutions of slavery and colonization. As a result, it framed Africans as physically gifted and sexual predators in an attempt to justify the cultural practice of slavery. Dubow (1995) describes the work of Linnaeus as an example of such a perspective. Dubow explains Linnaeus’ work in the following:

To each [race] he attached character descriptions. For example, Europeans were ingenious, inventive and governed by law; whereas Africans were crafty, lazy, careless and governed by the arbitrary will of their masters. (p. 26)

This perspective is customary of early evolutionist work that attempted to taxonomically categorize the African as inferior by means of physiological and mental differences. In fact, the racial evolutionist attempted to use physical observations as a means to justify racial ideologies about the scientifically proven inferiority of African people. Civis (1993) explained this perspective by claiming African people were inherently sweaty and pungent:

The Negro is eminently a sweating animal. The remarkable sweating capacity of the negro renders him objectionable in the cars, in the jury box, in the halls of legislation, in the crowds that assemble on the court green, but wonderfully fits him for his proper functions as a laborer in tobacco and rice fields and on the great cotton and sugar plantations of low latitudes. (p. 5)

Such an example attempts to attribute a physical characteristic (that is not empirically true) to an ability to perform cognitive tasks such as serving on a jury or participating in legislative activity. Civis goes on to express his opposition to advocacy of public school systems of “the doctrine of negro equality,” which aimed to prepare people of color for the “highest functions of public life” (p. 17).

In present times, such an argument would be met with great political opposition. However, at the time this perspective was so pervasive that other scholars strongly supported it. Thomas Jefferson’s (Jefferson and Peden, 1781/1982) *Notes on the State of Virginia* (see also Elkins, 1968; Elliott, 1969; Mellon, 1969) includes detailed provisions of such positions. One would think such perspectives were a relic of centuries ago; however, How’s (1971) argument to the General Synod of the Reformed Protestant Dutch Church attempts to reify this argument using contemporary language.

What becomes challenging about these perspectives is the manner in which they tend to reinvent themselves. Although

the early perspectives like racial evolutionists and racial taxonomies were nonempirical and greatly refuted, they continued to be reinvented and proposed using alternative methodologies. Scientists from Alfred Wallace (Osborne, 1980) to the founder of our modern taxonomical system, Carolus Linnaeus, all attempted to use taxonomical means to declare the African as physiologically inferior (McLeod, 1974). In fact, contemporary scholars argue these taxonomical perspectives still find their way into new media outlets (Chideya, 1995; Means-Coleman, 1998).

Given the attempts of racial evolutionists to frame African inferiority as biological, much of the early science research attempted to use classification as a way to establish Africans as cognitively inferior. However, with a few exceptions (e.g., Hernstein and Murray, 1994; Jensen and Johnson, 1994), the notion of genetic inferiority has been dismissed.

What becomes intriguing about how racism impacts science is the ways racial evolutionist perspectives have been retrofitted to adopt contemporary methodological means (Hernstein and Murray, 1994; Jensen and Johnson, 1994). Modern perspectives tend to frame these same assumingly biological characteristics of inferiority as psychological, cultural, or cognitive (Hernstein and Murray, 1994; Jensen and Johnson, 1994). The legacy of this early portrayal exists in the image of Africans as less capable beings or as Smith (1993) says ‘inferior others.’ Moreover, this image of Africans (and more recently African-Americans) can be traced throughout recent history and continues in retrofitted form in contemporary scholarship.

It is this image of blacks as inferior others that, we believe, remains central to how contemporary research is designed. Subsequently, African-Americans who are somehow able to conquer their inherent characteristics are able to escape characteristics of race that limit academic progress. The conception that those blacks who can appropriate white culture find it easier to achieve because they come from ‘good’ homes and have ‘white’ values, finds its roots in racial evolutionist perspectives. By contrast, assuming that underachieving blacks are less capable academically and have behavior problems as a result of cognitive differences is the result of the roots of racist science. This assumption is most notably found in the University of Chicago’s National Opinion Research Center’s study on race (Smith, 1990). Smith reported that racist views of blacks were so pervasive that the vast majority of respondents thought they were fundamentally less intelligent, less patriotic, and lazier than whites; and 78% said that Blacks prefer to live on welfare. Although these perspectives prevail, there is little research concerning the impact of a majority of Americans holding these views about African-Americans. We maintain that the sustained prevalence of historically constructed racial evolutionists models leave educational researchers unable to understand how the sociohistorical implications of these negative frameworks affect African-American students.

Craniometry and Physical Anthropology

Another paradigm in the recurring retrofitting of racist scientific ideology involved the use of the more empirically based craniometry and physical anthropology studies. During the 1770s, Camper measured human skulls in an attempt to justify racial differences. He used facial angles acquired by drawing a horizontal line from the nostril to the ear and the other perpendicular one from jawbone to the forehead to measure the intelligence of various species. This framework, which would later be known as craniometry would essentially reform the racial evolutionist perspective with a more empirical support. Camper proposed hierarchical differences between Europeans of 80°, Africans 70°, and the orangutan of 58° also reflected intellectual hierarchies.

The paradigm of physical anthropology reshaped the empirically based racial evolutionist perspective by attempting to use the physical patterns of skulls to argue for racial inferiority. Morton's (1839) *Crania Americana* measured the volume of the interior skulls to demonstrate that larger skulls yielded larger brains (Bruce, 2002; Fredrickson, 1972). He used these measurements to construct an argument that Caucasians were intellectually superior to Africans (Bruce, 2002; Fredrickson, 1972). Morton eventually attempted to retrofit earlier racial evolutionist perspectives by maintaining that skulls were reduced in size because of evolution. Morton's work eventually concluded that Egyptians were not actually African and should be considered European. Ironically, Ripley (1899) countered his argument by suggesting that Morton actually deceived his readers by stuffing skulls in an effort to justify his racist opinions. Despite contradictory evidence and counter-arguments, the result of this work was the reification of previously nonempirical theories of racial superiority with erroneous empirical data.

Racism in Medical Research

Another paradigm of racism in science involves the inappropriate use of minorities as participants in scientific research (Gamble, 1997; Lyles, 1994; Steinbrook, 1989). Scholars have asserted that inappropriate research on minority students is a practice which has been around for centuries. Based on the scientific research on racial bias, studies were designed that used African-Americans as test subjects (Gamble, 1997; Lyles, 1994; Steinbrook, 1989). Beech and Goldman describe this historical racism in science in the following:

Gamble (1997) traced the blatant and racist violations of medical research among African-Americans to the 1800s when Blacks were used as test subjects to promote race-specific stereotypes. A priority goal of such medical research was to provide scientific evidence for widely

held beliefs that African-Americans were physically, genetically, and intellectually inferior to whites. (p. 127)

This perspective discusses the premise that medical research has the potential to be influenced by racism in the form of deciding who is being studied, what illnesses are being studied, and how the participants are being treated.

In 1932 the United States Public Health Service (PHS) under the direction of Dr. Taliaferro Clark initiated its Study of Syphilis in the Untreated Negro Male. This study has since come to be known as the Tuskegee Syphilis Study (The Tuskegee Experiment). One of the most well-known instances of human subject abuse, the study has been the subject of several books (Jones, 1981), at least two feature films, stage plays, several documentaries, as well as academic, newspaper, and magazine articles (Bowman *et al.*, 1999; Corbie-Smith, 1999; Thomas and Quinn, 1991).

With this broad exposure, there is a great deal known about The Tuskegee Experiment. For example, it is clear over 400 men of African descent were known to be infected with syphilis and were observed to determine the means by which syphilis would ravage their bodies. We also know these men were denied treatment by the United States PHS, the US military, and other agencies (Bowman *et al.*, 1999; Corbie-Smith, 1999; Thomas and Quinn, 1991). In addition to being denied treatment, the participants were never told that they were actually part of the study. This study lasted 40 years and was wholly supported by the US government. The horrific abuse of these men, their families, and their communities was finally brought to an end in 1972 (Bowman *et al.*, 1999; Corbie-Smith, 1999; Thomas and Quinn, 1991). This is 125 years after the American Medical Association first published its medical code of ethics, 26 years after the Nazi party was excoriated for human subject abuse at Nuremburg, 25 years after penicillin had become the standard treatment for syphilis, and 8 years after the Declaration of Helsinki (Rothman, 2000).

What becomes clear from a case like The Tuskegee Experiment is the fact that it exemplifies a central aspect of the nature of Western modern science. It reflects how racism was central to Western science and that people of African descent were often the subjects of abuse. Incidents such as The Tuskegee Experiment are characterized as aberrations or the result of a few troubled practitioners who stepped outside the bounds of true and acceptable scientific practice. The reality, however, is people have endured every conceivable desecration at the hands of research scientists. Their corpses have been stolen for unauthorized autopsies and dissection; their bodies (both living and dead) have been put on public display; they have been exposed to myriad pathogens; they have been test subjects for experimental drugs and other products; they have endured experimental surgery (at times

without anesthesia); and they have been used as host vessels for the farming of blood and tissue. Washington's (2006) landmark text is the most comprehensive book to date on the medical exploitation of Africans in America and interested readers would do well to examine the complete history that she offers.

Monogenism and Polygenism

Another of the retrofitted paradigms of racist science involved the notions of monogenesis and polygenic research. In an attempt to dispute Africans and Europeans were not of a common origin, scholars developed the notion of polygenism (Nott, 1854; Ellis, 1988). Nott's (1854) *Types of Mankind* attempted to construct the argument that races from different regions actually developed from different species. This perspective contradicted the more widely held migration frameworks, yet enabled scholars to differentiate between the biological features of Africans and Europeans based on biological grounds. Although Darwin's notion of monogenism, or common origin, prevailed in the long term, science researchers attempted to debate about biological constructions of race rooted in past pseudo-sciences such as phrenology and craniometry.

Eugenics

Perhaps the most consistent form of retrofitted scientific racism involves the eugenics paradigm of scientific racism. Eugenics is a framework that states that the human gene pool can be improved by the removal of deficient minority gene pools (Glad, 2006). This perspective served as the foundation of the Nazi holocaust, American slavery, the Ku Klux Klan, and more contemporary problems of ethnic cleansing. These ideologies operated on the common eugenic assumption that a particular race of humans comprised of inferior gene pools must be addressed through various forms of social disorder and regulations. What remains central in these positions is the fundamental argument and baseless assumption that people are biologically inferior. Such a view finds its roots in the previously refuted taxonomical, eugenic, and polygenic perspective.

Unlike many other forms of scientific racism, eugenic perspectives continue to reinvent themselves well into the twentieth and twenty-first centuries. In 1901, Galton founded a psychometric science journal entitled *Biotrika* that promoted eugenic perspectives of race (Bulmer, 2003). Davenport's (1970) *Race Crossing in Jamaica* used psychometric research to argue that cultural and intellectual inferiority resulted from interracial childbearing.

Not surprising, this eugenic perspective was used as justification by the Nazis in World War II.

Psychometrics

Despite changes in the social position of race in American culture, research continued to attempt to affirm racist ideologies through psychometric research. Despite the fact that the research has been proven to be statistically inaccurate and lacking methodological sophistication to make the claims it is making, most contemporary studies that attempt to support racist ideologies find their home in psychometric research. Much of the early physical anthropological studies on race are now discredited as based on the methodological inaccuracies. However, this early psychometric work continued to reemerge in the form of intelligence quotient (IQ) testing during World War I. Much of the eugenics research attempted to use IQ testing to identify how Africans were genetically inferior. Valencia and Susuki (2001) critiqued this work by examining the culturally based assessment devices and methodological shortcomings. Morris-Susuki (1998) noted how the post-World War II denunciation of Nazi-based scientific research essentially discounted scientific research that attempted to use psychometrics to reify racist ideologies.

Perhaps the most famous form of intelligence research is found in Herrnstein and Murray's (1994) *Bell Curve*. They attempted to redesign previously discredited research to debate the intellectual superiority of Europeans. Despite being discredited a half century ago, the bell curve was supported by other works including Lynn's (2002) *IQ and the Wealth of Nations*, and Jensen's *The g Factor: The Science of Mental Ability*. Ultimately, the eugenics psychometrics movement has been able to sustain itself despite being discredited in each of its forms. What becomes clear is the fact that a number of researchers have attempted to use psychometrics as a means to argue for a racist perspective that believes biological differences are based in genetic inferiority.

The scope and depth of research conducted to prove racial theories is too detailed to address in a manuscript of this nature. Therefore, we have attempted to outline the major traditions in science research that have been used to sustain racist ideologies. Given this review, of a diversity of ways in which race and racism have influenced science, we hope to examine how making issues of race invisible in science education has limited the growth of the science education community.

Theoretical Retrofitting Scientific Racism

The history of racism in science shows a patterned progression of racial theories in science. Initially, the theories

are used to reify racist ideologies; they are refuted and removed from the science landscape until they are revisited years later with new methodological support. The transition from racial evolutionist, to craniometry, physical anthropology, racism in medical research, monogenism, polygenism, eugenics, and psychometric research, all reflect this pattern of the theoretical retrofitting of scientific racism.

Although the early racial evolutionist perspectives were philosophical and lacked any empirical support, they established an ideology of individual difference as the root of scientific research on minorities. The attempt to classify African-Americans as scientifically inferior was revisited in the craniometry and physical anthropology research paradigms. As a result, the notion that African-Americans were physiologically inferior remained central to how science viewed differences between African-Americans and Europeans. Despite the fact that craniometry and physical anthropology perspectives were empirically refuted, these views allowed the fundamental assumption about the perceived biological differences between races to pervade. Imagine any other arena of scientific study. How would science research that disputed previously held beliefs impact the intellectual landscape? In most cases, scientists would alter their perspective in favor of the new empirical frameworks. In the case of race and science, however, the research merely used newly discovered methods and tools to retrofit the previously held assumptions about minorities.

The eugenics, monogenism, and polygenism perspectives are fruitful examples of such a retrofitting process. With the emergence of genetic research and the establishment of insights regarding the human genome, scientists attempted to reify their racist ideologies about race by using the newly discovered information about genetics. Ultimately, such research has continued to be refuted, but the long-term effects of that type of research are enduring. In fact, these assumptions have become so reified that intellectual leaders like James Watson, one of the founders of DNA research, continued to make racist comments about African-Americans till as late as 2007. A recent article about Watson's perspective explains his argument in the following:

The 79-year old geneticist said he was "inherently gloomy about the prospect of Africa" because all our social policies are based on the fact that their intelligence is the same as ours – whereas all the testing says not really. He said he hoped that everyone was equal, but countered that "people who have to deal with black employees find this not true."

Comments like these from people who are central in the world's understanding of science enable racist ideologies to continue to reemerge in different forms. Watson's perspective is the same as the argument proposed by

Linnaeus and other racial evolutionists in the 1700s and continues to pervade the public opinion about Africans and science. Ironically, Watson's comments are incorrect. The research on race and intelligence argues there are no biological differences that account for performance differences. In fact, Watson has completely ignored years of psychological and educational research that addresses how differential academic performance results from stereotype threat and other psychosocial limitations (Steele, 1997). So in fact, the continuous retrofitting of racist ideologies under the cloud of science do more to shape racial differences than any biological factor. As a result, the continual process of using science to support racist perspectives has a profound impact on how people use, experience, and understand science.

Implications for Contemporary Science Education

We can understand how the history of race and racism in science had a substantial impact on science education if one appropriates a critical perspective on how race and racism can translate into science education. Science education has developed norms designed to reflect the norms of science (Fang, 2004; Halliday and Martin, 1993; Wellington and Osborne, 2001). However, these science norms reflect the impact of race and racism in defining discourse systems, funding priorities, and the presentation of images of what counts as science, and cultural expectations for who can become a scientist. The establishment of these science norms are Eurocentric and androcentric, which has the potential to frame how science is experienced by students. If a critical lens on race and racism is applied to understanding how these norms impact science education, one could consider the diversity of ways in which the issues of race and racism have ultimately impacted how science education is experienced.

Determining the Norms of Communication and Valued Contributions

Research that challenges the normative patterns of scientific communication has maintained a consistent role in science education research (Hildebrand, 1998; Lee, 2001; Warren *et al.*, 1994). Research on the genre of science discourse has established that using the language of science serves as a site for cultural conflict for many science students (Brown, 2004, 2006). Hildebrand (1998) argued that discursive practices of science education are hegemonic because they do not incorporate cultural ways of communicating in attempting to prepare students to learn the content and language of science. In the broader education community, scholars propose the use of native ways of communicating as a means to providing bridges through

which students are able to learn the discourse practices of academic environments (Lee, 2001b; Hull and Shultz, 2001). These new literacy studies essentially debate discourses that are inherently political and teachers must become skilled at providing students the ability to see the continuity between their native ways of communication and those they will learn in their classroom (Lee, 2001b; Hull and Shultz, 2001).

More recently, research that analyzes the impact of the discursive norms of science environments has argued that discursive practices of science serve as symbols of cultural mismatch for many minority students (Brown, 2004, 2006; Gilbert and Yerrick, 2001; Brown, 2005; Reveles *et al.*, 2004). Brown (2004) noted how students often avoided using science as a means to maintain their public identity. As a result, he argued some students developed misconceptions due to their inability to appropriate the discourse practices of the classroom. Reveles, Kelly, and Cordova explored how teachers attempted to design instructional practices to build bridges between their cultural modes of communication and those valued by science. Without an ideology that takes a critical approach to the role of race and the culture of science classrooms, issues of language and identity remain invisible to contemporary framework in science education.

How does research implicating the discourse practices of science reflect the racial history of science? Given the functionality of science discourse, learning to use the language of science is a critical component of learning science. In fact, Lemke (1990) states that learning science is essentially the process of learning to talk science. If science presents a culture-free and objective image of itself, and science education does not incorporate cultural ways of communicating in its approaches to teaching, the science classrooms reflect the cultural and linguistic norms of the cultural mainstream. Said differently, the ways of communicating in science reflect the culture in which they are designed. The great irony is that science has a history of being culturally biased and Eurocentric, yet the science education presented in schools proposes an image of science as a culture-free entity. Subsequently, minority students who attempt to study the subject find it to be discursively conflicting and culturally disconcerting. If people assume that all students experience science from the same cultural perspective and make no allocation for people from different cultures, then the culture of science may serve as a marker of cultural difference.

Access to Funding and Training

Another essential challenge in contemporary science education involves the current crisis in encouraging minority students to participate in science-related fields. If examined from a critical race perspective, one can analyze how

a history of retrofitted racism in science has impacted funding and training aspects of science education. Currently, there are over 100 historically black colleges and universities in the United States whose existence and legacy reflect the role of race in science. The very need for these universities reflects America's history of racism as they were designed to educate those who were not provided access to traditionally white colleges and universities. As a result, early research and training of African-American scientists was done by under-funded researchers who developed science programs at historically black colleges and universities.

Given this historical backdrop, one might ask how such a history of segregation impacts contemporary science education. First, a critical perspective on the influence of race in science suggests that access to science careers and science education at the secondary level was limited to a small number of African-Americans. More specifically, true access to careers in education was first made readily available after the desegregation of schools in the 1950s.

Second, one can argue that research on issues of importance to African-American communities has been limited due to the limited numbers of African-Americans conducting research. If minority students are not afforded access to research careers, and there are a limited number of minority researchers to advocate and propose scientific research of importance on minority communities, a disconnect may continue to exist. How can we expect students to matriculate into a science community that does not value their presence or their medical research challenges? Taken from this perspective, issues of underrepresentation in science education and the lack of medical research geared specifically to address the needs of minorities can be seen as a direct product of how science research and training was influenced by the history of segregation in America.

Another effect of racism on contemporary science education involves the issue of student isolation. Many students who choose to pursue careers in science find themselves experiencing isolation when they major in science careers. Students experience feelings of cultural mismatch at being one of few minority students enrolled in a science major at historically white universities. Research in higher education has documented these feelings of cultural isolation in a variety of ways (Solórzano, 1998). Solórzano (1998) describes how students experienced 'microaggressions' during their college experiences as a result of their being considered as representatives of their entire race. His argument suggests that students experience subtle forms of racism as teachers and other students make comments that are negative commentaries about the students' culture.

When issues of marginalization and microaggressions are extended to an analysis of science classrooms, one

could argue that feelings of peripheral membership are heightened due to the small number of minority students pursuing careers in science at traditionally white colleges and universities. This can be considered a direct reflection of years of cultural oppression and racial stereotyping that were reified through retrofitted racist science research.

Given the racial history of science and the evolving paradigmatic attempts to retrofit racist ideologies through scientific research, we assert failure to recognize the mythical objectivity of science education has prevented science educators from attempting to consider the impact of sociohistorical and racial issues in the career participation and classroom instruction of minority students. Lewis' (2003) exploration of how African-Americans pursued careers in science suggested the underrepresentation of African-American students is a direct response to failure of science educators to incorporate a thorough understanding of the sociohistorical impact of racial oppression.

This failure can be seen as a by-product of the fundamental assumptions that tend to frame contemporary research on minority achievement. Many scholars design science education research on minority students that fundamentally assumes minority underrepresentation as a result of erroneous choices of minority students. This assumption is made in a variety of ways. For example, Griffin's (1990) study implicated students' choices as a primary factor limiting their career decisions. Griffin argued "the incidence of career choices among ABC students" served as a central limitation (Griffin, 1990: 428). These decisions were not analyzed as the byproduct of larger sociohistorical issues or as potentially being informed by the failure of the science community to effectively market its careers to minority students. Post *et al.* (1991) were of the view that self-efficacy was fundamental in limiting students' matriculation into science. Post *et al.* described "how self-efficacy and interests related to consideration of math/science careers among Black freshman" were key factors in limiting students' pursuit of science careers (Post *et al.*, 1991: 179). Their study did not address how hearing rumors of racial inferiority may have negatively impacted the students' self-efficacy. Maple and Stage (1991) offered a more sociohistorically informed position as they sought to understand "what background, ability, and high school experience factors are related to choice of a quantitative major" (p. 42). However, they supported the fundamental argument that the failure of students to achieve success in their careers could be directly attributed to choices.

Our goal here is not to critique the theoretical findings of the studies cited earlier. Rather, our belief is that truly gaining an understanding of the impact of students' matriculation and experiences with science requires researchers to gain a rich understanding of the context in which those decisions are made. We contend that one cannot adequately understand the context of interactions of students and decisions without gaining a rich understanding of how

those contexts are designed and impacted by a sociohistorical past full of retrofitted racist scientific research.

The inclusion of the sociohistorical context of race is prevalent in other educational domains (Nasir and Saxe, 2003). In fact, Nasir and Saxe (2003) maintained that one cannot adequately understand students' identity without thoroughly considering the sociohistorical context framing the possibilities for identity development. We are quite concerned that the notion of scientific objectivity inherent in science has limited how science education research addresses issues of race in contemporary research.

A second theoretical assumption that tends to frame research on minority students involves perceived differences in psychological dispositions of students. This assumption treats disparities in performance of students and choice as the result of disparate psychological orientations. Post *et al.* (1991) applied Bandura's (1982) theory of self-efficacy to examine science career decisions of African-American college freshmen. They opined that a key figure in students' underperformance involved freshmen students' lower self-efficacy, interest, and overall consideration of mathematics and science careers. They compared students' lower self-concepts and self-efficacy without adequately exploring the sociohistorical effect of how social issues, like a history of American racism, impact how students develop psychological models. This assumption is derived without considering the wealth of research from psychology that details how a history of racism shaped a sense of self and self-efficacy in students (Steele, 1997; Cross, 1971, 1978). Their study and others like it (Gilleylen, 1993; Griffin, 1990, 1993; Maple and Stage, 1991) frame disparity in psychological terms pointing to psychological antecedents such as self-confidence, interests, and attitudes.

An additional assumption limiting contemporary research perspectives involves a fundamental assumption that the culture of minority students is embedded with deficient cultural characteristics that limit their success. These frameworks tend to position minority and white cultures in contrast; framing minority culture as culturally disparate containing traits such as family background and community values (Griffin, 1990), parental influence (Maple and Stage, 1991), cultural awareness and social support (Gilleylen, 1993), and students' interest in social service fields (Hager and Elton, 1971; Thomas, 1984). The studies presented here and the types of antecedents outlined are neither exhaustive nor mutually exclusive.

These studies are, however, representative of the larger body of literature aimed at explaining the underrepresentation of African-Americans in science and are useful in illustrating how contemporary science education research has not explained why or how race is operative in science career attainment and classroom performance.

Alone, each of these perspectives offers valuable insights by revealing different sides of a complex social

problem. Taken as a whole, however, this body of literature could be vastly improved by incorporating a framework rich in its understanding of how race and racism have shaped the sociohistorical context in which students develop academic identities. Ultimately, without adequately incorporating a framework that includes an informed perspective on the sociohistorical impact of race and racism, science education research is simply unable to assess how the construction of race, which has been a defining feature of Western culture since the sixteenth century, may serve as the foundation of the academic achievement gaps we are currently experiencing in science.

Sociohistorical Construction of Race

The sociohistorical construction of race is the fundamental assumption that humans are physiologically equal. Given that physiological equity, differences in their ways of thinking, academic performance, and cognitive modes reflect the influence of the way social context and social interactions have shaped individual identity development. If race is not biologically determined, then scholars must develop an in-depth understanding of how the sociohistorical construction of race had resulted in how people see themselves and define the possibilities for their identity development. Hacking (1999) argues that the primary use of 'social construction' is as a conscious-raising measure.

Those who adopt a sociohistorical framework of the role of race in science education offer an analysis critical of education research which attempts to measure characteristics of minority students that are assumed to be biological or inherent. Social constructionists argue certain issues do not need to exist, or would not exist if not for the impact of racism in society. For example, Steele's (1997) notion of stereotype threat states that academic differences are best viewed as purely sociohistorical phenomena that would not have existed due to biological differences, but rather are purely social effects resulting from years of racism. In contrast, sociohistorical theorists critique those who do not adopt a sociohistorical framework by arguing that they merely attempt to change the conditions under which socially shaped behaviors are developed. Said differently, they suggest that much of our contemporary research simply offers different treatments to conditions that are not biological but are rather sociohistorical. Therefore, without addressing the sociohistorical impact, one could design interventions and research that merely continue to assess the impact of sociohistorical racial difference in American culture.

In this sense, the social construction of race typically refers to the idea that race is not biologically determined, but is a social construction that must be studied accordingly. Seen this way, the very notion of race is not a matter of inevitable physiology but is rather a result of the impact

of social decisions. Social constructionists make the essential argument that scholars should engage in research that attempts to transcend racial classification as fundamental and explore the nuances of the sociohistorical influences which shape cultural interaction. In this way we may be able to improve our understanding of research if it transcends the assumptions it makes about race and culture. Ferrante and Brown (2000) provide an example of such a perspective as they attempted to use research that explicitly assigned meaning to racial categories rather than to be critical of the use of those categories.

Although, exploring a theory of racial construction is essential to this manuscript, we are not attempting to craft an alternative perspective on the sociohistorical construction of race. Rather, we argue that just as science has been influenced by the subtext of racial ideologies, science education has also been impacted by assumptions about what counts as race. Over the years, these perspectives continue to be retrofitted, but the primary assumption about a construction of race as physiologically or cultural inexorable remains the same. Therefore, defining a theory of racial construction is beyond the scope of this article. Instead, we assert that appropriating a sociohistorical perspective of the role of race in science education and recognizing the history of retrofitted scientific racism would illuminate how minority students, as a result of centuries of racial oppression, experience the discourse, artifacts, and knowledge and practices of science.

Bibliography

- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist* **37**, 122–147.
- Bell, D. (1992). *Faces at the Bottom of the Well: The Permanence of Racism*. New York: Basic Books Publishers.
- Bowman, J. E., Corbie-Smith, G., Lurie, P., et al. (1999). Tuskegee as a Metaphor. *Science* **285**, 47.
- Brown, B. (2004). Discursive identity: Assimilation into the culture of science and its implication for minority students. *Journal of Research in Science Teaching* **41**, 810–834.
- Brown, B. (2005). The politics of public discourse: Discourse, identity, and African-Americans in science education. *The Negro Educational Review* **56**(2), 205–220.
- Brown, B. (2006). It isn't no slang that can be said about this stuff: Language, identity, and appropriating science discourse. *Journal of Research in Science Teaching* **43**(1), 96–126.
- Bruce, D. (2002). *A Hideous Monster of the Mind: American Race Theory in the Early Republic*. Cambridge, MA: Harvard University Press.
- Bulmer, M. (2003). *Francis Galton: Pioneer Of Heredity and Biometry*. Baltimore, MD: Johns Hopkins Press.
- Chideya, F. (1995). *Don't Believe the Hype: Fighting Cultural Misinformation about African-Americans*. New York: Penguin Group.
- Civis (1993). The public school in its relation to the Negro. In Smith, J. D. (ed.) *The Benefits of Slavery: The New Proslavery Argument*, vol. 4, pp 1–37. New York: Garland Publishers.
- Corbie-Smith, G. (1999). The continuing legacy of the Tuskegee Syphilis study: Consideration for clinical investigation. *American Journal of the Medical Sciences* **317**, 5–8.
- Cross, W. (1971). The Negro-to-Black conversation experience. *Black World* **20**, 13–27.

- Cross, W. (1978). The Thomas and cross models of psychological nigrescence: A review. *Journal of Black Psychology* **5**, 13–31.
- Davenport, C. (1970). *Race crossing in Jamaica*. Westport, CT: Negro Universities Press.
- Dubow, S. (1995). *Scientific Racism in Modern South Africa*. New York: Cambridge University Press.
- Eisenhart, M., Finkel, E., and Marion, S. F. (1996). Creating the conditions for scientific literacy: A re-examination. *American Educational Research Journal* **33**(2), 261–295.
- Elkins, S. M. (1968). *Slavery: A Problem in American Institutional and Intellectual Life*. Chicago, IL: University of Chicago Press.
- Elliott, E. N. (1969). *Cotton Is King and Pro-Slavery Arguments*. New York: Negro University Press.
- Ellis, H. (1988). *Boulainvilliers and the French Monarch: Aristocratic Politics in Early Eighteenth Century*. Ithaca, NY: Cornell University Press.
- Fang, A. (2004). Scientific literacy: A systemic functional linguistic perspective. *Science Education* **89**(2), 335–347.
- Ferrante, J. and Brown, P. (2000). *The Social Construction of Race and Ethnicity in the United States*, 2nd edn. Upper Saddle River, NJ: Prentice Hall.
- Fredrickson, G. (1972). *The Black Image in the White Mind: The Debate on African-American Character and Destiny*. New York: Harper Torchbooks.
- Gamble, V. (1997). Under the shadow of Tuskegee: African-Americans and health care. *American Journal of Public Health* **87**, 1773–1778.
- Gilbert, A. and Yerrick, R. (2001). Same school, separate worlds: A sociocultural study of identity, resistance, and negotiation in a rural, lower track science classroom. *Journal of Research in Science Teaching* **38**(5), 574–598.
- Gilleylen, C. E. (1993). *A Comparative Study of the Science-Related Attitudes and the Factors Associated with Persisting in Science of African American College Students in Science Majors and African American College Students in Non-Science Majors*. Unpublished Doctoral Dissertation, Indiana University of Pennsylvania.
- Glad, J. (2006). *Future Human Evolution: Eugenics in the Twenty-First Century*. Schuylkill Haven, PA: Hermitage Publishers.
- Gosden, C. (2006). Race and racism in archaeology: Introduction. *World Archaeology* **38**, 1–7.
- Griffin, J. B. (1990). Developing more minority mathematicians and scientists: A new approach. *Journal of Negro Education* **59**, 424–438.
- Hacking, I. (1999). *The Social Construction of What?* Cambridge, MA: Harvard University Press.
- Hager, P. C. and Elton, C. F. (1971). The vocational interests of Black males. *Journal of Vocational Behavior* **1**, 153–158.
- Halliday, M. and Martin, J. (1993). *Writing Science: Literacy and Discourse Power*. Pittsburgh, PA: University of Pittsburgh Press.
- Hasian, M. (1996). *The Rhetoric of Eugenics in Anglo-American Thought*. Athens, GA: The University of Georgia Press.
- Hernstein, R. and Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.
- Hildebrand, G. (1998). Disruptive hegemonic writing practices in school science: Contesting the right way to right. *Journal of Research in Science Teaching* **35**, 345–362.
- How, S. (1971). *Slaveholding Not Sinful: Slavery, the Punishment of Man's Sin, Its Remedy, the Gospel of Christ*. Freport, NY: Books for Libraries Press.
- Hull, G. and Shultz, K. (2001). Literacy and learning out of school: A review of theory and research. *Review of Educational Research* **71**, 575–611.
- Jefferson, T. and Peden, W. E. (1781/1982). *Notes on the State of Virginia*. New York: W.W. Norton.
- Jensen, A. R. and Johnson, F. W. (1994). Race and sex differences in head size and IQ. *Intelligence* **18**, 309–333.
- Johnson, H. M. (1993). Racism in science – replies. *Scientific American* **268**(6), 12.
- Jones, J. (1981). *Bad Blood: The Tuskegee Syphilis Experiment*. Glencoe, IL: Free Press.
- Larson, E. (1995). *Sex, Race, and Science*. Baltimore, MD: Johns Hopkins Press.
- Lee, C. (2001a). Is October brown Chinese? A cultural modeling activity system for underachieving students. *American Educational Research Journal* **38**, 97–141.
- Lee, O. (2001b). Culture and language in science education: What do we know and what do we need to know? *Journal of Research in Science Teaching* **38**(5), 499–501.
- Lemke, J. (1990). *Talking Science: Language, Learning and Values*. Norwood, NJ: Ablex.
- Lewis, B. F. (2003). A critique of literature on the underrepresentation of African Americans in science: Directions for future research. *Journal of Women and Minorities in Science and Engineering* **9**, 361–373.
- Lyles, C. (1994). Blacks hesitant to donate; Cultural beliefs, misinformation, mistrust make it a difficult decision. *The Virginia-Pilot* 15.
- Lynn, R. (2002). *IQ and the Wealth of Nations*. Westport, CT: Praeger Press.
- Maple, S. A. and Stage, F. K. (1991). Influences on the choice of math/science major by gender and ethnicity. *American Educational Research Journal* **28**, 37–60.
- McLeod, D. J. (1974). *Slavery, Race and the American Revolution*. London: Cambridge University Press.
- Means-Coleman, R. R. (1998). *African American Viewers and the Black Situation Comedy: Situating Racial Humor*. New York: Garland Publishers.
- Mellon, M. T. (1969). *Early American Views on Negro Slavery*. New York: Bergman Publishers.
- Morris-Reich, A. (2006). Race, ideas, and ideals: A comparison of Franz Boas and Hans F.K. Gunter. *History of European Ideas* **32**, 313–332.
- Morris-Susuki, T. (1998). Debating racial science in wartime Japan. *Osiris, 2nd Series* **13**, 354–375.
- Nasir, N. and Saxe, G. (2003). Ethnic and academic identities: A cultural practice perspective on emerging tensions and their management in the lives of minority students. *Educational Researcher* **32**, 14–18.
- Norman, O. (1998). Marginalized discourses and scientific literacy. *Journal of Research in Science Teaching* **35**, 365–374.
- Nott, J. (1854). *Types of Mankind*. Ann Arbor, MI: University of Michigan Library.
- Osborne, H. (1980). A great naturalist: Alfred Russeo Wallace, 1823–1913. *Natural History* **89**(4), 118–129.
- Post, P., Stewart, M. A., and Smith, P. L. (1991). Self-efficacy, interest, and consideration of math/science and non-math/science occupations among Black freshmen. *Journal of Vocational Behavior* **38**, 179–186.
- Reveles, J., Cordova, R., and Kelly, G. (2004). Science literacy and academic identity formulation. *Journal of Research in Science Teaching* **41**, 1111–1144.
- Ripley, W. (1899). *The Race of Europe: A Sociological Study*. London, UK: Kegan, Paul, Trench, Trubner.
- Rothman, K. (2000). Declaration of Helsinki should be strengthened **12**, 442–445.
- Tucker, W. (2002). *The Funding of Scientific Racism*. Urbana, IL: University of Illinois Press.
- Smith, J. (1993). *The Benefits of Slavery: The New Proslavery Argument* vol. 4. New York: Garland Publishers.
- Smith, T. W. (1990). *Ethnic Images*. Chicago, IL: University of Chicago, National Opinion Research Center.
- Solórzano, D. (1998). Critical race theory, race and gender microaggressions, and the experience of Chicana and Chicano scholars. *International Journal of Qualitative Studies of Education* **11**, 121–136.
- Steinbrook, R. (1989). AIDS trials shortchange minorities and drug users. *Los Angeles Times*, September 25.
- Steele, C. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist* **52**(6), 613–639.
- Thomas, G. E. (1984). *Black College Students and Factors Influencing Their Major Field Choice*. Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools.
- Thomas, S. and Quinn, S. (1991). Light of the shadow of the Syphilis study at Tuskegee. *Health Promotion Practice* **1**, 234–237.
- Valencia, R. and Susuki, L. (2001). *Intelligence Testing and Minority Students: Foundations, Performance Factors, and Assessment*. Thousand Oaks, CA: Sage.

- Warren, B., Roseberry, A., and Conant, F. (1994). Discourse and social practice: Learning science in language minority classrooms. In Spencer, D. (ed.) *Adult Biliteracy in the United States*, pp 191–210. Washington, DC: Center for Applied Linguistics and Delta Systems.
- Washington, H. (2006). *Medical Apartheid: The Dark History of Medical Experimentation on Black Americans from Colonial Times to the Present*. New York: Random House Press.
- Wellington, J. and Osborne, J. (2001). *Language and Literacy in Science Education*. Buckingham, UK: Open University Press.

Further Reading

- Betcherman, G. (2002). Racism in science. *New Scientist* **173**(2336), 52.
- Brattain, M. (2007). Race, racism, and antiracism: UNESCO and the politics of presenting science to the postwar public. *American Historical Review* **112**(5), 1386–1413.
- Czech, H. (2006). Race, racism, and science. Social impact and interaction. *History and Philosophy of the Life Sciences* **28**(2), 275–277.
- Fairchild, A. and Bayer, R. (1999). Uses and abuses of Tuskegee. *Science* **284**(5416), 919–921.
- Haas, F. (2008). German science and black racism – roots of the Nazi Holocaust. *FASEB Journal* **22**(2), 332–337.
- Horsley-Meacham, G. (1991). Bull of the Nile: Symbol, history, and racial myth in 'Bento Cereno'. *New England Quarterly* **64**, 225–242.
- Leslie, C. (1990). Scientific racism – reflections on peer-review, science and ideology. *Social Science and Medicine* **31**(8), 891–905.
- Lewis, B. F., Pitts, V. R., and Collins, A. C. (2002). A descriptive study of pre-service teachers' perceptions of African American students' ability to achieve in mathematics and science. *Negro Educational Review* **53**, 31–42.
- Madison, R. (1988). Nationalism, racism, and science. *American Scientist* **76**(3), 233–234.
- Moore, R. (1999). Science, objectivity and racism. *American Biology Teacher* **61**(4), 242–242.
- Musselman, E. G. (2006). Race, racism, and science: Social impact and interaction. *Isis* **97**(3), 597–598.
- Rennie, M. (2002). Racism in science. *New Scientist* **173**(2336), 52–52.
- Rose, S. (1983). Science and pseudo-science. *New Scientist* **100** (1388), 822.
- Sales, V. and Langaney, A. (1997). Racism put to the test by science. *Histoire* **214**, 50–51.
- Schaffer, G. (2007). 'Scientific' racism again?: Reginald Gates, the 'mankind quarterly' and the question of 'race' in science after the Second World War. *Journal of American Studies* **41**(2), 253–278.
- Uebel, T. E. (1990). Scientific racism in the philosophy of science – some of historical examples. *Philosophical Forum* **22**(1), 1–18.
- Zumo, L. A. (1993). Racism in Science. *Scientific American* **268**(6), 12.

Africana Studies: Past, Present And Future

B R Hare, Syracuse University, Syracuse, NY, USA

A Young, Georgia Southern University, Statesboro, GA, USA

J B Stewart, Pennsylvania State University, Pennsylvania, PA, USA

D P Aldridge, Emory University, Atlanta, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

Schools of Thought

An expanded version was published in the (Fall, 2002) *International Journal of Africana Studies*, 8(2). An earlier version was published in Diamond, R. and Adams, B (eds.) (2000). *The Disciplines Speak: A Continuing Conversation: Rewarding the Scholarly, Professional, and Creative Work of Faculty*, pp 125–151. Washington, DC: American Association of Higher Education.

While some schools of thought within the field may be more or less intent upon counterbalancing the traditional Eurocentric bias characteristic of traditional American and European scholarship in the field, the central tendency of Africana studies is to investigate and illuminate the African world experience, in its own right.

As in any dynamic field, there are schools of thought in Africana studies. It is commonly agreed that the diversity of perspectives in the field is critical to its vitality and creativity. For example, there is an ongoing debate as to whether Africana studies is a field of study, a discipline, or something else. Some scholars declare Africana studies a discipline in that, like sociology, scholars look at different areas of the human experience, be it history, politics, family, etc., through the same lens. Such advocates call the discipline Africology. Some scholars consider Africana studies to be an interdisciplinary field of focused study to which, like the field of education, scholars bring a variety of disciplinary lenses, be they anthropological, psychological, historical, etc., to study the African-world experience. Still others would argue for a multidisciplinary or nondisciplinary label (for a discussion of the implication of terms, see Stewart, 1992).

We choose to use the term postdisciplinary as useful for describing the developmental trajectory of the field. It expresses a view of Africana studies as more than a territory to which scholars bring disciplinary lenses. It rather describes an evolving intellectual enterprise attentioning traditional, traditionally ignored, contemporary, and future ways of knowing. For example, one characteristic of the field's approach is the recognition of indigenous cultural knowledge which lies outside the traditional boundaries of humanistic and scientific discourse, such as that attainable through visual and oral accounts.

Central Tendencies in the Field

There is widespread agreement that many forms of cultural knowledge are necessary to understand the complexity of the Africana experience. In the sense that thought precedes speech and orature precedes literature, indigenous knowledge sometimes precedes scientific acceptance. For example, many folk remedies for illness have subsequently been found to have a basis in science and only then declared acceptable. Africana studies is open to other ways of knowing.

Another central tendency in the field is the rejection of the notion of pure objectivity as an attainable neutral space from which humans might analyze the human experience without bias (see Ani, 1994; Asante, 1990). While recognizing the value of the scientific method as a tool useful for minimizing bias, Africana scholars see danger in pure objectivity claims. The symbolic danger of such an illusion is akin to looking through a pair of binoculars from the wrong side and getting the illusion of distance. In short, we believe the use of traditionally recognized disciplinary tools, along with other ways of knowing, such as oral communications, spirituality, intuition, and reflections may generate richer and fuller descriptions of many phenomena.

There exists a positive creative tension within the field, within the units, and between the units even as Africana-studies scholars share both the pursuit of scholarly knowledge and social responsibility as the dual goals of the Africana-studies enterprise. Despite recognizing our common cultural roots, the presence of multiple cultural traditions in the field presents the challenge of shared and unique cultural visions as a consequence of different cultural experiences. We recognize that culture and color are not the same thing. While we generally agree on the ends, there is great variation in the means by which these dual goals are pursued. Given the change agent mission characteristic of many Africana-studies units, this tension perhaps becomes inevitable. This positive paradox may have been best expressed by Martin Luther King Jr. when he declared that no change is possible without creative tension.

Titles and Foci

The variety of labels under which Africana studies is included generally reflects different geographic areas of concentration such as African, African American, and African Caribbean Studies. Diaspora studies is sometimes used to describe studies of populations of acknowledged African descent not located on the African continent. The term acknowledged African descent is deliberately chosen to avoid unnecessary controversy about lines of human descent, as there is solid anthropological and genetic evidence that the human roots are common and African.

The variety of different geographical locations attended by the field provides challenges and a wealth of opportunities to explore continuity and change in cultural patterns within and between peoples over time and over space. Such scholarship is also engaged in by scholars located in traditional departments such as English, history, sociology, and political science. Scholars in the field also engage in specialized-area studies such as women studies and education (Aldridge, 1984).

Africana studies, the global term used throughout this discourse, is the term generally recognized in the field and used by the National Council for Black Studies to describe the field as a whole. The term unit is also used in this discourse to include the various structural labels, be they programs, centers, or departments, which Africana studies operates under in the academy.

Applied Scholarship

Due to the unique cluster of methodological approaches and emphasis on service to students and community espoused by and expected of scholars in the field, Africana studies challenges the traditional elite model of scholarship that, in our view, too often advocates scholarship for scholarship's sake. Africana studies, in contrast, generally seeks praxis in advocating what might be called applied scholarship. Thus, performance evaluations for tenure and promotion, when controlled by the units themselves, reflect this centrality of service to students and community. This difference in performance expectations and subsequent method of faculty evaluation has been a source of friction between the units and higher university-evaluation bodies over the differing criteria, especially when the Africana-studies unit has not acquired the autonomous rights associated with departmental status.

Historical Background

While the roots of the study of the Africana experience can be traced back to ancient Africa and ancient African universities, the formal emergence of the field is anchored in

the world of our elders in the nineteenth century, such as Anna Julia Cooper, Maria Stewart, Ida B. Wells-Barnett, George Washington Williams, Henry McNeal Turner, and Martin Delany. In the early to mid-twentieth century, many of the pioneers in Africana studies functioned largely outside the academy including W. E. B. Du Bois, J. A. Rodgers, Arthur A. Schomburg, Zora Neal Hurston, Amy Jaques Garvey, Carter G. Woodson, Lorraine Hansberry, Jacob Lawrence, Charles S. Johnson, and Alain Locke. However, the formal institutionalization of the field of Africana studies within and across disciplines can be linked to the work of W. E. B. Du Bois at Atlanta University, Charles Johnson at Fisk University, and especially the work of Carter G. Woodson with the Association for the Study of Negro Life and History (presently ASALH), the *Journal of Negro History*, and the Annual Negro History Week Celebration (currently Annual Black History Month Celebration) (Young, 1997).

The activities of all three of these historical giants in the field of Africana studies demonstrate a commitment to the integration of scholarship with a commitment to fostering constructive social change; one of the fundamental values of contemporary Africana studies. Useful perspectives on the historical foundations of Africana studies and the symbiosis between academic inquiry and social change can be found in a number of writings including Aldridge (1994); Anderson (1990); Hall (1999); Karenga (1982, 1993); Semmes (1992); Stewart (1992); Turner (1984); Turner and McGann (1976); and Harris (1990).

Competing Conceptions

There was a high degree of overlap of visions among the early advocates of Africana studies. Representative early discussions of the nature, goals, and future of Africana studies were produced by Blassingame (1969, 1972); Jackson (1970); McClendon (1974); Robinson *et al.* (1969); Rose (1975); and Russell (1975). At the same time, however, as noted by Allen (1974), three distinct views about the nature of Africana studies were espoused:

1. an academic conception, whereby the mission of Black studies is to research Black history and illuminate the contributions of Blacks;
2. an ideological political conception, whereby Black studies is seen as an instrument of cultural nationalism;
3. an instrumental political conception, whereby Black studies is considered a vehicle for social change with a functional relationship to the Black community.

While Allen's categorizations are useful, they mask the importance of ideologies other than cultural nationalism important in the early development of Africana studies, such as Marxism (see Alkalimat *et al.*, 1986; People's College Press, 1977).

The evolving nature of Africana studies as an intellectual enterprise at this stage of its development is open to interpretation. For example, as previously indicated, there has been much debate about whether Africana studies is a discipline, in the traditional sense, or an area of inquiry to that one brings tools from selected disciplines and applies them to a particular subject matter. Stewart (1981) discusses the philosophical implications of various conceptions of Africana studies. The general consensus has been that the field has had, as its developmental goal, the creation of an academic unit with a discipline-like configuration and mechanisms to link scholarly inquiry to social action and social change.

Black Studies Defined

Karenga (1982: 35–36) defines Black studies as an interdisciplinary discipline that has seven basic subject areas. According to Karenga, these intradisciplinary foci that at first seem to be disciplines themselves are, in fact, separate disciplines when they are outside the discipline of Black studies; but inside, they become and are essentially subject areas that contribute to a holistic picture and approach to the Black experience. Karenga suggests that the qualifier Black, attached to each area in an explicit or implicit way, suggests a more specialized and delimited focus that of necessity transforms a broad discipline into a particular subject area of inquiry. Karenga delineates seven basic subject areas: Black history; Black religion; Black social organization; Black politics; Black economics; Black creative production (Black art, music, and literature); and Black psychology. As previously indicated, use of terms such as interdisciplinary and multidisciplinary must be viewed with caution. For some scholars, these terms simply convey the idea of casual discourse among scholars trained in traditional disciplines. However, as used by many Africana-studies specialists, the terms have much more of a transformative interpretation. To illustrate, Hall (1996: 14) observes, “In theory, a body of knowledge or field of study that is truly interdisciplinary does not just cross conventional disciplinary boundaries, it obliterates them. In practice, however, the trend remains that most African American studies programs represent multidisciplinary approaches in that — disciplinary boundaries intact — scholars trained in history, humanities, or social science fields apply their particular disciplinary approaches to some piece of knowledge in relative isolation from the whole.”

Aside from the general issue of disciplinary intersections, there is also the question of which disciplines are especially germane to the mission of Africana studies. Karenga (1982: 32) suggests that, “Black Studies, as both an investigative and applied social science, poses the paradigm

of theory and practice merging into active self-knowledge that leads to positive social change.” In Karenga’s view, “Black Studies is a discipline dedicated not only to understanding self, society, and the world but also to critique them in a positive developmental way in the interest of human history and advancement.”

Despite the diversity reflected by these multiple linkages, for the purpose of this assessment, we use the term discipline to characterize the field because it has all of the characteristics of Kuhn’s definition of a paradigm or disciplinary matrix (see Karenga, 1988; Stewart, 1992). We recognize, however, that future developments may eventually produce an enterprise that differs significantly from existing disciplines.

African-Centered Thought

As previously indicated, all academic disciplines have different schools of thought that contribute diverse perspectives to the collective enterprise. There is, however, always an overarching worldview and a set of values, part of a disciplinary matrix that includes theories, research methods, etc., that is shared among schools of thought (see Kuhn, 1970). One value that links different schools of thought within Africana studies is the commitment to producing a perspective that reflects the beliefs, values, culture, and interests of peoples of African descent. The terms African-centered and Afrocentric are used in contemporary discourse to convey this value. Critics of Africana studies have attempted to denigrate the field by attaching specific interpretations of this value expressed by some particular authors. Such specific formulations are confused with the generalized value that constitutes an undergirding value for the discipline. These terms are used in a general rather than limited sense as suggested by Semmes (1992: 18) “Afrocentric or African-centered thought is solidly embedded within a historical tradition that can be understood by recognizing internal dialectics and contradictions over historical time within the field and the society.” Thus, it may be argued that Afrocentric or African-centered thought exists regardless of whether or not it is called Afrocentric or African-centered.

Each school of thought within Africana studies has its own particular interpretation of what constitutes African-centered thought and how that thought should be linked to other intellectual traditions. One schema representing the range of thought in Africana studies is presented in **Table 1**.

Within each school of thought, there are subtle but important distinctions. At the same time, it is possible to compare the treatment of core theoretical constructs and approaches to social change across different schools of thought.

Table 1 Categorization of alternative conceptions of Africana studies

1. Africana studies as a disciplinary matrix-driven enterprise
Kawaida theory (Karenga)
Africology (Asante)
2. Africana studies as a discipline constituting syntheses of traditional disciplines
Multidisciplinists
Nondisciplinists
3. Africana studies as an adjunct to Eurocentric metatheories
Marxism-paradigm of unity (Alkalimat <i>et al.</i>)
Feminism
4. Africana studies as a subcomponent of individual disciplines
History
Literature
Sociology
5. Africana studies as a component of nondisciplinary aggregates
African studies
American studies
Ethnic studies
Multicultural studies
6. Africana studies as expressions of selective emphases
Kemetologists
Melanists
Generalized folk approaches

Table 1 (Stewart, 1992) presents a classification scheme that clarifies the conceptions of the field discussed to this point. The framework articulated in the introduction is most closely associated with the first category, but, as discussed previously, the current state of development is also strongly identified with the second category. External evaluators tend to identify with the fourth and fifth conceptions. There is no doubt that what can be described as popular Afrocentrism has permeated various popular culture media, most notably film and music. However, this phenomenon has virtually no relationship to formal efforts to develop Africana studies into a disciplinary matrix-driven enterprise. Popular Afrocentrism draws heavily on self-published works produced by persons largely outside the academy. Although, as discussed previously, there is a long history of African-centered thought generated outside the academy, during the current period, the locus of development has been and remains solidly inside the academy.

The marginalization of peoples of acknowledged African descent and other similarly situated groups spurred the development of a school of thought heavily influenced by Marxist thought and practice. The collective of scholars associated with this school of thought were, in fact, responsible for institutionalizing the commitment of the National Council for Black Studies to the linkage between scholarship and activism as expressed by the organizational adoption of the motto “academic excellence and social responsibility.”

The categories presented here are sometimes overlapping and thus not presented as exclusive.

Gender Studies

The treatment of gender within Africana studies has precipitated increasing examination and debate. There are two distinct tendencies within this movement and both have been instrumental in exposing and countering the sexist tendencies of some Black nationalist ideologies. Some scholars have called for the application of feminist constructs for the analysis of the experiences of Africana women and have critiqued cultural nationalism as a belief system that is inherently nonliberatory because of what are perceived as inescapable sexist tendencies (see Hull *et al.*, 1982).

In contrast, many Africana womanist scholars urge the redefinition of the partnership between Africana men and women in pursuit of intellectual and political objectives (see Aldridge 1989, 1992; Gordon, 1987; Hudson-Weems, 1989, 1993). Proponents of this view generally identify with and contribute to the overall effort to establish Africana studies as a discipline.

Afrocentricity

Given the emphasis that has been focused on Afrocentricity as a key construct associated with the efforts to develop Africana studies as a disciplinary matrix-driven enterprise, it is useful to examine its content in more detail to examine the credibility of claims of limited coherence. **Table 2**, (Stewart, 1992), offers a framework for comparing schools of thought. The framework is useful for comparing elements of the disciplinary matrix of different variants of a given school of thought (in this case Afrocentric thought), and comparing constructs associated with different schools of thought (here Afrocentric and Marxist paradigms).

The variants of Afrocentric thought examined in **Table 2** are those generated by two of the most widely acclaimed scholars in the field, Molefi Asante (systematic Afrocentricity) and Maulana Karenga (Kawaida theory). The principal point is that there is an acceptable level of coherence within Africological thought and that systematic comparisons can be used to identify desirable research thrusts that would accelerate the development of Africana studies.

The National Council for Black Studies

The National Council for Black Studies (NCBS) was founded in 1975 and has been the premier professional organization dedicated to the professional development of

Table 2 Comparison of systems of thought associated with Africana studies

<i>Comparative criteria</i>	<i>Kawaida theory</i>	<i>Systematic Africology</i>	<i>Paradigm of unity</i>
Concept of Afrocentricity	In the cultural image and human interests of 'African Americans'	Placing African ideals at the center of any analysis that involves African culture and behavior	Unspecified
Treatment of gender	In the context of male/female relationships	Unspecified	As a biological category
Theoretical focus	Theory of culture	Theory of inquiry	Theory of social change
Theoretical emphasis	Social organization	Authenticity of knowledge and culture	Social dynamics
Key constructs	Cultural authenticity Nguzo Saba	Afrocentricity Nommo	Class relations Social cohesion
Treatment of race	Cultural (emphasis on consciousness)	Cultural (emphasis on language and symbols)	Social disruption Biological
Observational language	Partially transformed	Completely transformed	Standard Marxist
Subject areas	History, religion, social organization, economic organization, political organization, creative production, ethos	Society, communication, history, culture, politics, economics, psychology	Consciousness, society, economy, biology
Periodization scheme	Unspecified	Unspecified	Two-stage cycle
African emphasis	Classical civilizations	Classical civilizations	Traditional pre-slave trade

Africana scholars and the development of programs in the United States and around the world. The annual national and occasional international conferences sponsored by NCBS are notable because of their ability to bring together established and emerging scholars to deliberate on the African world experience. The motto, "promoting academic excellence and social responsibility," captures the past, present, and future direction of seeking to bring about social change through scholarship that informs social action. The organization remains centrally involved in monitoring and assessing the field in light of this dual goal.

Publication

Aldridge (1994) identifies approximately 40 000 Black/Africana studies publications in the discipline in the forms of books, monographs, articles, and special publications, along with several dozen new journals and texts. Of course, much of this research has reflected traditional methods and has been published in traditional organs. But, as noted previously, a distinct body of knowledge has been generated that is associated uniquely with Africana studies as a self-standing intellectual enterprise. Some of this work has been published in professional journals that have been created to serve scholars in the field, such as *Journal of Black Studies*, *Western Journal of Black Studies*, and *International Journal of Africana Studies* (the official organ of the

NCBS). Special issues or sections of other periodicals have also served as publication outlets such as *The Black Scholar*, *Journal of Negro Education*, and *Journal of Black Psychology*.

Africana-studies research has also been published in a variety of monograph formats including introductory texts (e.g., Alkalimat *et al.*, 1986; Anderson, 1990; Karenga, 1982, 1993; People's College Press, 1977); singly authored examinations of broad issues facing the field (e.g., Asante, 1986, 1987, 1990); single-authored works that provide the foundations for current research thrusts (e.g., Diop, 1974, 1990); and collections of essays designed either to complement texts or cover a range of critical issues (e.g., Alkalimat, 1990; Anderson, 1990; Conyers, 1997; Hayes, 1992; Turner, 1984).

It is interesting to note that the character of the research published has changed as the academic training background of faculty has changed. The first generation of faculty staffing Africana-studies units was trained solely in traditional disciplines, and they tend to approach the study of the Africana experience through their various disciplinary lenses, those who have developed other Africana ways of seeing have had to do so through self-development. Such scholars tend to publish both in traditional and specialized Africana-studies publications. However, as more graduate programs have recently produced new faculty trained specifically in the field, alternative methodologies are being introduced both in the classrooms and the journals. Some of

this new scholarship introduces ways of knowing and methods of analysis that either fail to meet or challenge, or transcend the expectations of traditional journals.

Thus, while bringing fresh and new energy and vision, Africana-studies scholars run the risk of not meeting traditional standards when measured by traditional criteria. Both the journals that would publish their works and the units in which they reside in the university remain suspect. This dilemma speaks to the need for Africana-studies units to retain or attain, whether through the attainment of departmental status or another arrangement, the right to make tenure and promotion decisions independently.

Creativity

Africana studies also endeavors to embrace the more creative work of the professoriate advocated in Dr. Ernest Boyer's (1988), *Scholarship Reconsidered: Priorities of the Professoriate*. The work is a model of the scholarship of "discovery, integration, application, and teaching that so urgently needs to be acknowledged and rewarded." To this end, the pursuit of external funding is reflective of the creative work or applied scholarship.

Teaching

The best pedagogical approaches associated with Africana-studies instruction at the undergraduate level are designed to expose students to the ideas, philosophical orientations, and benefits and liabilities of the variety of schools of thought. Instructional strategies are designed to transmit critical analytical skills that allow students to make their own informed decisions about the world. Instruction is generally delivered through organized units, usually departments, programs, centers, or institutes using faculty with various types of affiliations within the academy. The range and orientation of instruction generally reflect the mission, faculty expertise, and general focus of the academic unit, for example, African American or diasporic studies.

Africana studies courses are increasingly included in general-education curricula, expanding the numbers and variety of students in classes. Among opportunities offered in Africana studies for students, in addition to classroom exposure to a variety of professors and perspectives, are stimulating learning activities, conducted on and away from campus. Study-abroad opportunities in Africa and the Caribbean, and field experiences in local community service agencies are offered. Unique teaching models and strategies have been employed to provide students in Africana studies with a wide range of intellectual and personal experiences at home and abroad.

Study Abroad

Africana-studies units, like their counterparts in international studies and individual departments and/or programs, have developed study-abroad programs. However, the major global foci of Africana studies, Africa, and the Caribbean area tend to be neglected in traditional international-studies offerings. The opportunity for students in Africana-studies-oriented study-abroad programs to travel, to learn, and to experience the language, people, and culture of Africa is, according to participants, a profound world-view-changing experience. Not only do the students report upon return that their view of Africa will never be the same, their view of their home nation is simultaneously altered as they become citizens of the world.

This central benefit of the Africana-studies experience comes through the encouragement and, at times, insistence of Africana-studies units that college and university offerings have to be expanded beyond Europe, Latin America, and Asia, to include Africa and the Caribbean. Such programs have also differed from the Africana-studies-program experience traditionally provided, because they present an African-centered rather than European-centered exposure. The Africana view may sometimes be through a different, though equally valid lens. It is a critique from the inside out rather than from the outside in.

Conclusion

The continuing advancement of Africana studies requires, in addition to a well-managed interface with traditional disciplines, a well-developed intellectual core that both differentiates the enterprise from and connects it to traditional disciplines. In Stewart (1992), seven useful developmental thrusts are identified:

1. Development of a theory of history.
2. Articulation of a theory of knowledge and social change.
3. Delineation of a theory of race and culture.
4. Expansion of the scope of inquiry encompassed by the disciplinary matrix.
5. Expanded examination of the historical precedents to modern Africana studies.
6. Increased emphasis on applications of theoretical work.
7. Strengthened linkages to interests outside the academy.

The dual missions of student centeredness and applied scholarship, in combination with the too frequently ambiguous status of the units within the academy, make Africana studies unique and vulnerable in its relationship to the academy. This relationship will continue to present as yet unseen challenges. The field, nevertheless, remains as

innovative, adaptive, and optimistic as its founders and the peoples it educates, studies, presents, services, and defends in the intellectual arena called the academy.

The type of synthesis of ways of knowing pursued through Africana studies is actually at the cutting edge of a broader de-emphasis of science (as distinct from technology) as the only source of socially useful knowledge. This erosion of scientific influence is signaled by several developments. Within the scientific establishment itself, there is increasing soul-searching among practitioners in various social and behavioral disciplines regarding the value of the current incrementalist approach to knowledge. The growing prevalence of meta-analyses is a manifestation of a desire to generate broader interpretations from potpourris of small-scale scientific studies. Popular culture, with its decided sensate emphasis, has supplanted science as the major purveyor of cultural imagery, values, and, most important, sensory stimulation. This shift in attitude of mind finds its intellectual expression in deconstruction and cultural-studies paradigms that challenge both the claims of universalism and linguistic assumptions associated with scientific explanation. The growing significance of popular culture media is, of course, linked directly to the increasing commodification of culture and the associated scramble for profits in the global cultural marketplace. From this vantage point, one of the challenges of Africana studies is to preserve culturally specific sources of innovation and resist market tendencies to commodify cultural forms.

In addition to the increasing role of sensate approaches to understand reality, there are also ideational challenges to scientific discourse, including various new-age philosophies and resurgent traditional religious movements. Not surprisingly, many of the new-age philosophies borrow heavily from non-Western cultures. This resurgence of idealistic ways of knowing constitutes an expanded search for spiritual guidance in a rapidly changing world. Africana-studies research examining the spiritual and religious foundations of classical African civilizations provides an important cross-cultural dimension to this growing approach to the search for meaning.

Finally, technology plays an increasingly critical role in propagating these shifts in attitude of mind. In particular, new technologies of information acquisition and dissemination, especially the Internet and other rapidly growing instant forms of worldwide communication, are accelerating shifts in patterns of knowledge production and information acquisition as well as changes in patterns of human-human and human-machine interaction. It is critical that Africana studies prioritize the mastery of information technologies, not just as consumers, but as designers of computer-human interfaces and the design of systems to ensure that cultural diversity is reflected and respected in the foundations of this new and powerful technological force. Thus, emphasis on the serious

scrutiny of science and technology is a necessary and logical extension of the ongoing quest by Black/Africana studies scholars/activists.

The central mission of Africana studies is to advance knowledge that simultaneously facilitates scholarship and activism, toward advancing human liberation.

Bibliography

- Aldridge, D. (1984). Toward a new role and function of black studies in historically white and black institutions. *Journal of Negro Education* 53, 354-367.
- Aldridge, D. (ed.) (1989). Black women in the American economy. *Special issue of the Journal of Black Studies* 20(2).
- Aldridge, D. (1992). Womanist issues in black studies: Towards integrating Africana women into Africana studies. *The Afrocentric Scholar* 1(1), 167-182.
- Aldridge, D. (1994). Status of Africana studies in higher education in the USA.
- Alkalimat, A. (ed.) (1990). *Paradigms in Black Studies*. Chicago: Twenty-First Century Books and Publications.
- Alkalimat, A., et al. (1986). *Introduction to Afro-American Studies: A Peoples College Primer*. Chicago: Twenty-First Century Books and Publications.
- Allen, R. (1974). Politics of the attack on black studies. *The Black Scholar* 6(1), 2-7.
- Anderson, T. (1990). *Black Studies, Theory, Method, and Cultural Perspectives*. Pullman, WA: Washington State University Press.
- Asante, M. (1986). A note on Nathan Huggins' report to the Ford Foundation on African-American studies. *Journal of Black Studies* 17, 255-262.
- Asante, M. (1987). *The Afrocentric Idea*. Philadelphia: Temple University Press.
- Asante, M. (1990). *Kemet, Afrocentricity and knowledge*. Trenton, NJ: Africa World Press.
- Blassingame, J. (1969). Black studies: An intellectual crisis. *The American Scholar* 38, 548-561.
- Blassingame, J. (1972). *The Slave Community, Plantation Life in the Antebellum South*. New York: Oxford University Press.
- Boyer, E. (1988). *Scholarship Reconsidered: Priorities of the Professoriate*. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.
- Conyers, J. (ed.) (1997). *Africana Studies, a Disciplinary Quest for Both Theory and Method*. Jefferson, NC: McFarland & Co.
- Diop, C. (1974). *African Origins of Civilization, Myth or Reality*. In Cook, M. (ed. and trans.). Westport, CT: Lawrence Hill.
- Diop, C. (1990). *The Cultural Unity of Black Africa, the Domains of Patriarchy and of Matriarchy in Classical Antiquity*. Chicago: Third World Press.
- Gordon, V. (1987). *Black Women, Feminism, Black Liberation: Which Way?* (Revised edn.). Chicago: Third World Press.
- Hall, P. (1996). Introducing African American studies: Systematic and thematic principles. *Journal of Black Studies* 26(6), (July), 688-712.
- Hall, P. (1999). *In the Vineyard*. Knoxville: University of Tennessee Press.
- Harris, R. (1990a). Coming of age: The transformation of Afro-American historiography. In Alkalimat, A. (ed.) *Paradigms in Black Studies*, pp 53-72. Chicago: Twenty-First Century Books and Publications.
- Harris, R. (1990b). The intellectual and institutional development of Africana studies. In Harris, R., Clark-Hine, D., and McKay, N. (eds.) *Black Studies in the United States, Three Essays*, pp 15-29. New York: The Ford Foundation.
- Hayes, F. (ed.) (1992). *A Turbulent Voyage: Readings in African American Studies*. San Diego: Collegiate Press.
- Hudson-Weems, C. (1989). Cultural and agenda conflicts in academia: Critical issues for Africana women's studies. *Western Journal of Black Studies* 13(4), 181-189.
- Hudson-Weems, C. (1993). *Africana Womanism: Reclaiming Ourselves*. Troy, MI: Bedford Publishers.

- Hull, G., Scott, P., and Smith, B. (1982). *All the Women Are White. All the Blacks Are Men. But Some of Us Are Brave: Black Women's Studies*. Old Westbury, NY: Feminist Press.
- Jackson, M. (1970). Toward a sociology of black studies. *Journal of Black Studies* 1(2), 131–140.
- Karenga, M. (1982). *Introduction to Black Studies*, 1st edn. Los Angeles, CA: University of Sankore Press.
- Karenga, M. (1988). Black studies and the problematic of paradigm. *Journal of Black Studies* 18(4), 395–414.
- Karenga, M. (1993). *Introduction to Black Studies*, 2nd edn. Los Angeles, CA: University of Sankore Press.
- Kuhn, T. (1970). *The Structure of Scientific Revolutions*, 2nd edn. Chicago: University of Chicago Press.
- McClendon, W. (1974). Black studies: Education for liberation. *The Black Scholar* 6(1), 15–20.
- People's College Press (1977). *Introduction to Afro-American Studies* vols. 1 and 2. Chicago: Peoples College Press.
- Robinson, A., Foster, C., and Ogilvie, D. (1969). *Black Studies in the University, a Symposium*. New York: Bantam Books.
- Rose, A. (1975). *Afro-American Studies in Higher Education*. Coral Gables, FL: Center for AfroAmerican Studies.
- Russell, J. (1975). Afro-American studies: From chaos to consolidation. *The Negro Education Review* 26(4), 181–189.
- Semmes, C. (1992). *Cultural Hegemony and African American Development*. Westport, CT: Praeger.
- Stewart, J. (1981). Alternative models of black studies. *UMOJA* 5, 17–39.
- Stewart, J. (1992). Reaching for higher ground: Toward an understanding of Black/African studies. *The Afrocentric Scholar* 1(1), 1–63.
- Turner, J. (1984a). Africana studies and epistemology: A discourse in the sociology of knowledge. In Turner, J. (ed.) *The Next Decade: Theoretical and Research Issues in Africana Studies*, pp v–xxv. Ithaca, NY: Africana Studies and Research Center.
- Turner, J. (ed.) (1984b). *The Next Decade: Theoretical and Research Issues in Africana Studies*. Ithaca, NY: Africana Studies and Research Center.
- Turner, J. and McCann, C. (1976). Black studies as an integral tradition in African-American intellectual history. *Issue* 73–78.
- Young, A. (1997). Why We Still Need to Celebrate Black History Month. *Homefront* (Winter), 27.

Further Reading

- Young, C. (ed.) (1972). *Black Experience: Analysis and Synthesis*. San Rafael, CA: Leswing Press.
- Young, C. (1984). *An Assessment of Black Studies Programs in American Higher Education*. Special issue of *Journal of Negro Education* 53.
- Young, C. and Martin, G. (1984). The paradox of separate and unequal: African studies and Afro-American studies. *Journal of Negro Education* 53, 257–267.

Beyond Black Nationalism to Black Internationalism: The Compelling Case for Examining Black Educational Challenges Globally

K Freeman, Southern University and A & M College System, Baton Rouge, LA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

- Black populations** – African descendants.
Cultural alienation – The process to isolate groups.
Cultural annihilation – The process to remove culture and eliminate groups.
Culture of exclusion – The process where groups are removed from settings – whether schools or other societal settings.
Internationalism – Collective, comparative, globalized group-focused.
Nationalism – A particularized country-specific focus, where groups experience challenges in isolation.

Introduction

With all the worldwide debates about the impact of globalization on every aspect of cultures, from the environment to technology to the economy and education, it is astonishing that – around the world – black populations (broadly defined in this article as African descendents) remain underrepresented at every level of schooling. Yet, the economics of education theorists have long ago (Schultz, 1962) documented the linkages between education participation, economic development, political participation, and environmental justice. With such documentation, it is surprising that there is such a dearth of research and/or focus on the comparative struggles that black populations share in their quest to participate in education.

Writers across disciplines (e.g., history and the humanities) such as Baldwin (2002) and Lorde (1992) have challenged the existing paradigm used to understand and examine the cultural link among blacks in the diaspora. However, educational researchers have been slow to see or move beyond the Black National model of examining educational underrepresentation. Consequently, the necessity of refocusing and/or shifting to a different paradigm to examine the comparative, common educational experiences black populations confront globally have been largely misunderstood and under-examined. This perspective – utilizing a broader, cross-disciplinary, and cross-boundaries approach – makes the case for moving

from a Black Nationalist model to a black internationalist model to examine black educational challenges globally.

A Black Nationalist model is a particularized model where African descendents in each country focus on their challenges in isolation – whether they are issues of education, economics, or environmental justice. A black internationalist model differs from a single-focus paradigm to one that builds on the collective, comparative experiences of black populations globally. Contrary to common perceptions that there are more differences than similarities between black populations, this research makes the case for an expanded paradigm to researching and developing an international agenda for better understanding educational challenges confronting black populations globally.

Different Countries, Similar Historical Experiences

As James Anderson has so rightly indicated in numerous writings and speeches (e.g., 1988 and 1999) – to understand the educational experiences of blacks, it is necessary to examine the historical context of their existence. Although, as Lorde has indicated, blacks in different countries have experienced particular histories – as it relates to their humanity and their pursuit of education – blacks across societies have had similar experiences. The common historical linkage among blacks in the diaspora is captured by Opitz *et al.* (1992) who wrote in reference to Afro-Germans: “In the course of colonial exploitation, enslavement, and domination ‘Negro’ (from Latin *niger*, i.e., black) became an especially negative epithet. The thinking underlying this label attempted to link physical characteristics with intellectual and cultural ones” (p. 7). That is, across cultures, being black has historically been thought of as being intellectually inferior and being without a culture or having a primitive or uncivilized culture.

While often not recognized, blacks have been a part of European societies – in most cases, longer than in America. For example, Fryer (1992) who wrote what would be considered the seminal work on blacks in Britain, indicated that African descendants have been in Britain for centuries, as a group they “have been living in Britain for close on 500 years. They have been born in Britain since about the year 1505” (p. ix). That would mean that, based on

American history, Black Britons – as a group – were in England more than a hundred years prior to the arrival of blacks in America in Jamestown, Virginia, in 1619 (Clarke, 1972). While Opitz *et al.* (1992) indicated that “there is no precise method of determining when the first Africans came to Germany and when the first Afro-Germans were born,” they indicated that “several paintings have survived from the twelfth century that depict Africans living in Germany” (p. 3). Although the exact arrival of blacks across cultures may be unclear, what is certain is that, by the mid-sixteenth and seventeenth centuries, blacks were settled in countries around the globe. For example, “in the mid-sixteenth century, one-tenth of the population in the Portuguese capital were Black, and, as in France and England, it was probably also true in Germany” (Opitz *et al.*, 1992: 3).

Blacks in different countries faced similar treatment in the portrayal of their culture. Blackness was associated with evilness, inferiority in every way, and subhumanism. In Britain, for example, “Africans were said to be inherently inferior, mentally, morally, culturally, and spiritually, to Europeans” (Fryer, 1992: 7). Likewise, in Germany, blacks were portrayed negatively. In general, “Africans were seen as the lowest human form, thought to be related to the highest animal form, the monkey” (Opitz *et al.*, 1992: 8). As well, “Most Portuguese seem to have thought that blacks as a people were innately inferior to whites in physical beauty and mental ability and moreover, that they were temperamentally suited to a life in slavery” (Saunders, 1982: 166). Although on a different continent, blacks in the United States were experiencing the same devaluation of their humanity. In the United States, this statement about the use of slave codes best captures how blacks were viewed: “There were variations from state to state, but the general point of view expressed in most of them [Slave Codes] was the same, that is: slaves are not persons but property, and laws should protect the ownership of such property, and should also protect the whites against any dangers that might arise from the presence of large numbers of Negroes” (Franklin and Moss, 1988: 114). This devaluation of the black culture by the nineteenth century, according to Clarke (1972), caused Africans the world over to begin to search for a definition of themselves.

Another similar link in the historical experiences of black populations was the exploitation of their labor. As Clarke (1972) indicated, “The story of the African slave trade is essentially the consequences of the second rise of Europe . . . They were searching for new markets, new materials, new manpower, and new land to exploit. The slave trade was created to accommodate this expansion” (p. xvii). Just as blacks in America were relegated to working the land and as servants to increase the wealth of this country, so were blacks in European countries. For example, according to Fryer (1992), “The majority of the 10,000 or so black people who lived in Britain in the

eighteenth century were household servants – pages, valets, footmen, coachmen, cooks, and maids – much as their predecessors had been in the previous century” (p. 73). Although working menial jobs, Fryer conceded that as a Liverpool writer declared in 1893, “it was the capital made in the African slave trade that built some of our docks’ and the price of human flesh and blood that gave us a start” (p. 66). Similarly in Germany, blacks were used for menial labor. In Germany, for example, blacks “were forced to cultivate export products or to work on the plantations and in the mines of whites” (Opitz *et al.*, 1992: 25). The same was the case in Portugal. According to Saunders (1982), “The nobility employed – or underemployed – large numbers of slaves solely as domestic servants” (p. 63).

Even when blacks were interested in working higher-status jobs, they were forbidden. For example, in London after 1731, blacks were not allowed to learn a trade. In fact, on 14 September of that year, the lord mayor of London issued the following proclamation prohibiting apprenticeships for black people:

It is Ordered by this Court, That for the future no Negroes or other Blacks be suffered to be bound Apprentices at any of the Companies of this City to any Freeman thereof; and that Copies of this Order be printed and sent to the Masters and Wardens of the several Companies of this City, who are required to see the same at all times hereafter duly observed. (cited in Fryer, 1992: 75)

Although there were some blacks in the United States who possessed some skills – especially the few slaves who lived in towns – the great majority of slaves’ responsibilities were divided between two distinct groups – the house servants and the field hands (Franklin and Moss, 1988). According to these researchers, slaves had little opportunity to develop initiative because their responsibilities were proscribed for them. Therefore, the idea that blacks did not want to work and thus played a role in having their skills underutilized was not the case.

Understandably, the exploitation of labor is – and has always been – intricately linked to lack of educational opportunities. To keep groups uneducated or undereducated has been a formula across societies for the underutilization of their talents. Nkrumah – the son of the late African leader – describes it as a sort of worldwide formula where African descendants everywhere have been relegated to the bottom educationally and economically. As such, the idea has been to prevent blacks from being empowered intellectually, culturally, and economically.

As Anderson (1988) has indicated, it is through education that individuals begin to feel empowered, and African-Americans were active agents in their right to be educated. From slavery until now, African-Americans have had to struggle to have the opportunity to participate in any form of education. According to Anderson,

"Blacks emerged from slavery with a strong belief in the desirability of learning to read and write" (p. 5). As an example of the intensity of anger that slaves held for keeping them illiterate, Anderson quoted a former slave: "There is one sin that slavery committed against me which I will never forgive. It robbed me of my education" (p. 5). Restrictive legislation was passed to prohibit slaves from learning to read and write (Fleming, 1981). According to Fleming, from 1850 to 1856, less than 5% of African-Americans out of a population of 4.5 million could read and write.

Just as blacks in America were forbidden to learn to read and write, the same was true in other countries where the controlling population was non-black. In Portugal, for example, Saunders (1982) indicated that very few blacks were able to read and write.

As Opitz *et al.* (1992) indicated as it related to Afro-Germans, "The limitations of educational opportunities concurrent with the favoring of some individuals led to hierarchical structures that undermined the solidarity of the community" (p. 33). In that sense, not only has lack of educational opportunities been utilized to limit the use of the talent and skills of blacks but education has also been used as a force to destabilize communities. That is, education as a commodity has been used as a means to favor some intra-group members over others as a way to undermine community relationships. In Portugal, for example, mulattoes were thought to be more conversant "with Portuguese customs, were supposed to be more gifted intellectually than were blacks from African" (Saunders, 1982: 172).

Therefore, the commonality of the black historical experience in countries where non-blacks are the dominant populations has been the underutilization of the potential of blacks by demeaning their humanity through the slavery experience, destabilizing their communities, exploiting their labor, and limiting their educational opportunities. However, even when blacks were allowed to participate in education, a process of cultural alienation and/or annihilation was implemented.

Cultural Alienation and Annihilation

According to Anderson (1994), one way that the use of power over a less powerful group takes form is "the group with the greater power annihilates the powerless group or drives them out of the territory" (p. 82). Thus, cultural alienation and annihilation can be defined as that process that controlling populations use to minimize or eradicate the culture of minority populations. Generally, this process is synonymous with assimilation, acculturation, or deracination (the term Mankiller (1993) used to describe the mission of boarding schools to annihilate the American Indian culture) – that is, the uprooting or destruction of a race and its culture. More specifically, she stated, "the

primary mission of Sequoyah and the other boarding schools was a full-scale attempt for the children to leave everything behind that related to their native culture, heritage, history, and language" (p. 8).

In order for blacks to assimilate into the dominant culture in different countries, they were often separated (alienated) from their own cultural group or an attempt was made to eliminate (annihilate) their culture altogether. Clear examples of cultural alienation and annihilation can be found across black populations. In Australia, for example, "Between 1910 and 1970 it [Australian Government] forcibly stole up to 100,000 aboriginal children from their families to live with whites in an attempt at forced integration – 'to breed the black out,' as politicians of the day expressed it" (Evenson, 1998: A10). As in other countries, particularly children who – through rape – had white fathers were taken from their families. Pilkington (1996), whose aboriginal name is Nugi Garimara, wrote an intriguing account of the assimilationist policy of Australia that took her mother and her mother's sisters – Daisy and Grace – away from their families. She wrote the following:

Patrol officers travelled far and wide removing Aboriginal children from their families and transported them hundreds of kilometres down south. Every mother of a part-Aboriginal child was aware that their offspring could be taken away from them at any time and they were powerless to stop the abductors. (p. 40)

The aboriginal experience was not unlike that of Black Britons and African-Americans. As it related to the experience of Black Britons, Fryer (1992) indicated that most black Londoners "had been torn from their parents and ethnic groups while still children. They were atomized in separate households, cut off from the cultural nourishment and reinforcement made possible by even the most inhumane plantation system" (p. 70). As a Black Briton recently stated in the popular press (15 August 1999): "I have done a lot of thinking about issues of assimilation and national identity. And I'm beginning to suspect that immigrants can only blend totally into their host environment if they are the same colour as the host or dominant population" (p. 58). Similarly, African-American families were divided. Franklin and Moss (1988) described the process of dividing black slave families in the United States: "Since the domestic slave trade and slave breeding were essentially economic and not humanitarian activities, it is not surprising to find that in the sale of slaves there was the persistent practice of dividing families. Husbands were separated from their wives, and mothers were separated from their children" (pp. 106–107). Although the dividing of families might have been justified for economic reasons, it also served the function of cultural annihilation and/or alienation. That is, when families were divided, they had to reconstruct their social institutions into new forms.

Education has been used as one of the primary channels through which cultural alienation and annihilation have occurred. As Pilkington (1996) noted about the Aboriginal girls who were taken from their family, the belief was that “part-Aboriginal children were more intelligent than their darker relations and should be isolated and trained to be domestic servants and labourers” (p. 40). In a like manner, in America, blacks were treated to educational opportunities differently by color. For example, according to Franklin and Moss (1988), mulattoes had more of a chance of schooling than others.

In addition to using education as a divisive tool based on color among blacks, cultural alienation and annihilation have occurred through the transmission of education. That is, the way in which education has been transmitted (teaching style) and the content of educational materials (curriculum) have discounted the social and cultural capital of black populations (consciously or subconsciously) and have, therefore, minimized the culture of black populations. Researchers such as DiMaggio and Mohr (1985) have suggested that cultural capital is typically specialized social behaviors that make one accepted at different levels of society. Whereas some theorists (e.g., Coleman, 1990) have indicated that while social capital is related to cultural capital, social capital is more related to relations among persons. For example, Coleman (1988) explains social capital as the networks that provide information, social norms, and achievement support.

In simplest terms, the concepts of cultural and social capital mean assets – in the form of behaviors – on which individuals and/or families can draw to meet a certain set of established values in a society (Freeman, 1997). As Freeman noted, these societal values are generally established by majority groups in society and encompass behaviors – such as the way individuals speak to the way they dress. The more individuals are able to meet these established standards, the more they are accepted by different institutions (e.g., schools) in society. There is no doubt, however, the cultural and social capital that students bring to the classroom have tremendous implications for how they will be accepted, treated, and provided necessary information. According to Cicourel and Mehan (1985), students are provided different educational opportunities because students arrive in school with different types of culture capital. Black students typically arrive in school with different cultural capital, and schools, therefore, attempt to eradicate their cultural values in order to make them assimilate.

Who has taught, what has been taught, and how it has been taught over time have severely eroded the cultural identity and educational opportunities of blacks. Although this has been the case historically, black educators and researchers – particularly in the United States – are currently extensively examining and discussing ways to undo the intellectual damage to black children by

demonstrating the importance of valuing the culture of blacks rather than eradicating their culture. For example, several researchers and educators have written about the impact of the influence of the curriculum (what is being taught) on the education of black children (e.g., Banks, 1988; Freeman, 1999; Hollins, 1996; King, 1995).

The school curriculum, as defined by Hollins (1996), is “in fact that package of knowledge, skills, and perspectives that prepares us to develop the attributes of thought and behavior that comply with the prescribed norms” (p. 82). When there are inconsistencies in the compliance of these norms by different cultural groups – in this case, black populations – this can lead to various group members questioning their identity, being turned off from learning, and/or under-performing academically.

The curriculum validates individuals’ culture, history, and sense of self – what is possible. Therefore, when black populations’ culture is not included in the very heart of school, this must create the feeling within students that something is missing. Banks (1988) says it best: “It is important for students to experience a curriculum that not only presents the experience of ethnic and cultural groups in accurate and sensitive ways, but that also enables them to see the experiences of both mainstream and minority groups from the perspectives of different cultural, racial, and ethnic groups” (p. 161). Searle (1994) – a Black British educator – mirrors Banks’ statement. Referring to the national curriculum of Britain, he stated:

The national curriculum, with its gradgrindian sequence of learning and testing, the narrow cultural chauvinism of its approach to knowledge and human experience and its blatantly racist exclusion of cultures, histories, languages and perspectives of Britain’s black people, is already creating a tedium and uniformity which will do nothing to spark the interest and motivation of young people to learn. (p. 26)

It is as though the fact that many black students have underachieved is completely divorced from the curriculum. Linkages have clearly been established between the curriculum and its effect on African-American students’ achievement (Hollins, 1996; King, 1995). As Hollins (1996) indicated, for African-American children the “discontinuity between the home-culture and school learning ultimately disrupts the learning process for many children and the resulting failure may lead them to reject the Euro-American culture and school learning as well” (p. 84).

Aside from what has been taught (the curriculum), as a way to stem the tide of alienation and annihilation of the black culture through education, Black educators have also more recently focused their research on who has taught and how black children have been taught. For example, researchers/educators such as Foster (1997), Irvine (1994), Ladson-Billings (1994), and Siddle-Walker (1996) have examined the role of culturally relevant pedagogy

and the relevance of the cultural perspective of black teachers in black students' achievement. More specifically, these researchers have suggested the importance of historically and currently understanding the role black teachers have played in helping black students achieve.

At the same time that black educators in the United States have focused their research on the inclusion of a black perspective and valuing black culture in the educational system, black researchers in other countries have also begun to address these same issues. Searle (1994) speaking on the British system stated: "The ignorance of teachers and the school system generally about the communities whom they serve is still a vital factor which promotes conflict and misunderstanding between teachers and students" (p. 25). Bridges (1994) – another British educator – supported Searle's assertion. He indicated that more black teachers were employed as a way "of counterbalancing the underrepresentation of black teachers in the borough and, through this, to attack the real problems of underachievement among black children" (p. 4).

Cultural alienation and annihilation have had a devastating effect on black students' participation in education. Over time, the process of trying to breed the black out (assimilation) – whether through the devaluation of the cultural capital of blacks or the what and who of the transmission of knowledge – has severely impacted on black students' sense of self and achievement. As such, there should be little doubt that cultural alienation/annihilation has led to a culture of exclusion for black populations globally.

Culture of Exclusion

Even when blacks have had the opportunity to participate in education, as it relates to black British, Searle (1994) indicated a culture of exclusion has existed. Searle describes culture of exclusion as it relates to Black Britons in this way:

There is much mystification surrounding the word 'exclusion' in education parlance. Schools do not refer to 'expulsions' now, even though almost all parents would know what that means. The preferred term is 'permanent exclusion', but it comes to the same thing. (p. 19)

Searle reported that secondary schools "in Nottingham, Reading, Bristol and the north London borough of Brent showed that black students were up to six times more likely to be suspended from school than their white peers" (p. 24). Drawing on the research of Searle and Bridges (1994), a culture of exclusion, then, can be defined as that process whereby black children are excluded from schooling, whether through suspension or expulsion or placement in the lower tracks of schooling – which would be referred to as "internal expulsion" (Bridges, p. 11).

Educational expulsion (suspension) is a phenomenon that is similarly faced by black populations globally. For example, the honorable Ka (1998) – in a report on the conditions of blacks in Portugal – indicated the following: "The black community in general – and the children and young people in particular – are victims of educational expulsion. The number of those that manage to complete compulsory education (up to the 9th grade) is frightfully low, and even worse if we consider the number of those who manage to complete secondary education (12th grade)" (p. 2).

In the United States – as in Britain and Portugal – a greater number of black children are subjected to suspension or expulsion. For example, a comprehensive study conducted by the Children's Defense Fund (1975) in the 1970s reported the following:

No one is immune from suspension, but black children were suspended at twice the rate of any other ethnic group. Nationally, if they had been suspended at the same rate as whites, nearly 50 percent or 188,479 of the black children suspended would have remained in school. Although black children accounted for 27.1 percent of the enrollment in the districts reporting to OCR, they constituted 42.3 percent of the racially identified suspensions. (p. 12)

In support of these earlier findings, in a more recent study, Morris and Goldring (1999) cited studies that concluded that "desegregation was often accompanied by an increase in the overall student suspension as well as a high disparity between black and white student suspension" (pp. 60–61).

However, the greatest numbers of African-American students are subjected to "internal expulsion." A recent report (Klenbort, 1999) in the Southern Regional Council periodical indicated, as an example, a statement from a high school sophomore who spent his schooling in the lower-level track: "You live in the basement, you die in the basement. You know what I mean?" Researchers, such as Oakes (1985), Wheelock (1992), and Braddock, II and Slavin (1995), support this student's description of tracking or what can be referred to as internal expulsion. Page and Page (1995) describe how tracking became the norm following desegregation: "Schools in the region [Southern region] became increasingly resegregated through the use of tracking, with the majority of African-American students assigned to lower tracks and the majority of Caucasian students assigned to higher tracks" (p. 73).

The way in which tracking can best be thought of as internal expulsion is best captured by Oakes (1985) and Braddock, II and Slavin (1995). For example, Braddock, II and Slavin indicated that the effects of tracking on students were striking. They found that students in lower tracks performed significantly less well than similar low achievers in untracked schools and were much more likely

to end up in noncollege-preparatory programs by tenth grade. This effect, they suggested, “being in the low track in eighth grade slams the gate on any possibility that a student can take the courses leading to college” (p. 8). Oakes suggested that “lower-track students are more alienated from school and have higher drop-out rates” (p. 9). These researchers also found that tracking hurts students’ self-esteem, causing them to feel inferior. Based on these researchers’ findings, then, the effects of tracking on students’ life chances is tantamount to being excluded. That is, black students who are in tracks are in school but because they are more likely to drop out or to have limited opportunities beyond secondary school, in a sense, the school has excluded them. Morris and Goldring (1999) in their study on disciplinary rates of African-American and white students in Cincinnati magnet and nonmagnet schools explained it in this way:

The overall effect of disciplining students, which involves removing them from the classroom, will drastically impact students’ acquisition of educational materials presented by the teacher. Other long term effects might include African-American students falling behind academically, or worse, dropping out of school altogether. (p. 64)

Whether internally excluded or suspended or expelled, black students globally share similar experiences. At least – as described in Britain, Portugal, and the United States – black students comprise the majority of students facing a culture of exclusion. This culture of exclusion has led to the underutilization of blacks in education which, in turn, has implications for societal and individual costs.

Blacks in Education: The Case for an Expanded Research Agenda

This research suggests that to more fully comprehend the educational dilemma of black populations in particular countries, it is important to examine their experiences individually yet collectively. As Lorde (1992) indicated, although particular histories have fashioned the experiences of blacks in different countries, similar historical experiences make the case for expanding the research agenda on black populations – intra-, inter-, and across cultures. This research suggests that a more in-depth comparison of the historical experiences and how the impact of these experiences have constructed the current educational phenomenon in which blacks find themselves, supporting James Anderson’s position to better understand the experiences of black populations. As pointed out in this research, blacks in different countries have had similar historical experiences, and it is important for these to be examined collectively in order to determine where similar programs and models might be instituted.

When cultural differences are taken in considerations, the wheel might not have to be reinvented.

Another agenda that this research suggests is that, not only is it important to expand the research agenda across borders to understand the black educational experience, but it supports the necessity of examining the educational experiences of blacks across disciplinary boundaries. That is better linkages between the different research agenda (whether K-12 and higher education or different aspects of educational phenomena) need to be established. This research demonstrates how each aspect of the black experience impacts on each experience and level of schooling.

To better understand the barriers constructed across cultures to underutilize the potential of blacks’ human potential in countries where non-blacks are the controlling populations, much more research needs to be conducted. In order to more fully understand this process, more research on each point where underutilization occurs has to be explored. This is to understand the costs – particularly the nonmonetary costs – as Bowen (1977) indicated, associated with the underutilization of black potential so that countries better understand how everyone is losing is an imperative. The quote from the Commission on Research in Black Education from Wendell Berry summarizes the imperative of countries to assess the underutilization of the human potential of Black populations: “If the white man has inflicted the wound of racism upon Black men, the cost has been that he would receive the mirror image of that wound into himself.”

Finally, although, it is not enough for countries to assess the costs associated with the underutilization of the potential of black populations, for, this research suggests that – at each point – blacks have been underutilized. At the beginning of the twenty-first century, countries will find it necessary to develop strategies to address the societal and individual costs associated with black populations’ underutilization while simultaneously increasing the utilization of the potential of blacks. Carnoy (1997) describes the way the process of increasing the spending on the underutilization of blacks should work: “The vicious cycle of increasing social costs will gradually break. Down the road, as early-childhood investment reduces spending on adult social problems, more public funds will become available for general education and other activities that improve worker productivity and growth rates” (p. 241). However, countries have not yet been able to develop a formula for assessing the individual and societal costs, therefore, to appropriately target their spending.

For globalization to be successful for individual countries and countries collectively, the potential of all of their citizens have to be utilized. In order to balance this imperative, different paradigms and players have to be a part of the research agenda and the agenda has to be expanded, particularly to address the educational challenges confronting

black populations. This must be done by utilizing the collective, comparative experiences of black populations globally, an internationalist model.

Bibliography

- Anderson, C. (1994). *Black Labor White Wealth: The Search for Power and Economic Justice*. Edgewood, MD: Duncan & Duncan.
- Anderson, J. D. (1988). *The Education of Blacks in the South, 1860–1935*. Chapel Hill, NC: The University of North Carolina Press.
- Banks, J. A. (1988). *Multithnic Education*, 2nd edn. Boston, MA: Allyn and Bacon.
- Braddock, J. M. II and Slavin, R. E. (1995). Why ability grouping must end: Achieving excellence and equity in American education. In Pool, H. and Page, J. A. (eds.) *Beyond Tracking: Finding Success in Inclusive Schools*, pp 7–19. Bloomington, IA: Phi Delta Kappa Educational Foundation.
- Bridges, L. (1994). Exclusions: How did we get here? In Bourne, J., Bridges, L., and Searle, C. (eds.) *In Outcast England: How Schools Exclude Black Children*, pp 1–16. London: Institute of Race Relations.
- Bowen, H. (1977). *Investment in Learning*. San Francisco, CA: Jossey-Bass.
- Children's Defense Fund (1975). School suspensions: Are they helping children? *A Report by the Children's Defense Fund of the Washington Research Project*.
- Cicourel, A. V. and Mehan, H. (1985). Universal development, stratifying practices, and status attainment. *Research in Social Stratification and Mobility* 4(5), 728–734.
- Clarke, J. H. (1972). Introduction. In Rogers, J. A. (ed.) *World's Great Men of Color*. Macmillan: New York.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology* 94, 95–120.
- Coleman, J. S. (1990). *Foundations of Social Theory*. Cambridge, MA: Belknap Press of Harvard University.
- DiMaggio, P. and Mohr, J. (1985). Cultural capital, educational attainment, and marital selection. *American Journal of Sociology* 90(6), 1231–1261.
- Evenson, B. (1998). Land crisis brewing down under: A 1996 court ruling means aboriginals could claim 80 per cent of Australia's land mass. *Ottawa Citizen* A10.
- Foster, M. (1997). *Black Teachers on Teaching*. New York: New Press.
- Franklin, J. H. and Moss, A. A. Jr. (1988). *From Slavery to Freedom: A History of Negro Americans*, 6th edn. New York: McGraw-Hill.
- Freeman, K. (1997). Increasing African Americans' participation in higher education: African American high school students' perspective. *Journal of Higher Education* 68(5), 523–550.
- Freeman, K. (1999). My soul is missing: African American students' perceptions of the curriculum and the influence on college choice. *Review of African American Education* 1(1).
- Freeman, K., Carnoy, M., Findlay, H., Joiner, B., and Magyari-Beck, I. (1999). Economic development and the utilization of human potential: Bridging the gap between higher education, economics, and culture. *Final Narrative Report to the Kellogg Foundation Grant*.
- Fryer, P. (1992). *Staying Power: The History of Black People in Britain*, 6th edn. London: Pluto Press.
- Hollins, E. R. (1996). *Culture in School: Revealing the Deeper Meaning*. Mahwah, NJ: Erlbaum.
- Irvine, J. J. (1990). *Black Students and School Failure: Policies, Practices, and Prescriptions*. Westport, CT: Praeger.
- Ka, F. (1998). Black people's situation in Portugal. *Paper Prepared for Fisk University Race Relations Institute Conference*. Nashville, TN.
- King, J. E. (1995). Culture-centered knowledge: Black studies, curriculum transformation, and social action. In Banks, J. and McGee Banks, C. A. (eds.) *Handbook of Research on Multicultural Education*, pp 349–380. New York: Macmillan.
- Ladson-Billings, G. (1994). *The Dreamkeepers: Successful Teachers of African American Children*. San Francisco: Jossey-Bass.
- Lorde, A. (1992). Foreword to the English language edition. In Opitz, M., Oguntoye, K., and Schultz, D. (eds.) *Showing our Colours: Afro-German Women Speak Out*. London: Open Letters. pp i–vii.
- Mankiller, W. (1993). *Mankiller: A Chief and her People*. New York: St. Martin's Press.
- Morris, J. E. and Goldring, E. (1999). Are magnet schools more equitable? An analysis of the disciplinary rates of African American and White students in Cincinnati magnet and nonmagnet school. *Equity and Excellence in Education* 32(3), 59–65.
- Oakes, J. (1985). *Keeping Track*. New Haven, CT: Yale University Press.
- Opitz, M., Oguntoye, K., and Schultz, D. (1992). *Showing our Colours: Afro-German Women Speak Out*. London: Open Letters.
- Page, J. A. and Page, F. M., Jr. (1995). Tracking and its effects on African-Americans in the field of education. In Pool, H. and Page, J. A. (eds.) *Beyond Tracking: Finding Success in Inclusive Schools*. Bloomington, IA: Phi Delta Kappa Educational Foundation.
- Pilkington, D. (1996). *Follow the Rabbit-Proof Fence*. Queensland: University of Queensland Press.
- Saunders, A. C. and de, C. M. (1982). *A History of Black Slaves and Freedmen in Portugal 1441–1555*. Cambridge, UK: Cambridge University Press.
- Schultz, T. W. (1961). Investment in human capital. *American Economic Review* 51, 1–17.
- Searle, C. (1994). The culture of exclusion. In Bourne, J., Bridges, L., and Searle, C. (eds.) *Outcast England: How Schools Exclude Black Children*. London: Institute of Race Relations.
- Walker, V. S (1996). *Their Highest Potential: An African American School Community in the Segregated South*. Chapel Hill, NC: The University of North Carolina Press.

Further Reading

- Carnoy, M. (1992). Education and the state: From Adam Smith to Perestroika. In Arno, R., Altbach, P., and Kelly, G. (eds.) *Emergent Issues in Education*. Albany, NY: State University of New York Press.
- Carnoy, M. (1994). *Faded Dreams: The Politics and Economics of Race in America*. Cambridge, UK: Cambridge University Press.
- Edwards, B. H. (2003). *The Practice of Diaspora: Literature, Translation, and the Rise of Black Internationalism*. Cambridge, MA: Harvard University Press.
- Farrell, J. (1992). Conceptualizing education and the drive for social equality. In Arno, R., Altbach, P., and Kelly, G. (eds.) *Emergent Issues in Education*. Albany, NY: State University of New York Press.
- Freeman, K. (2002). The underutilization of the human potential of individuals who are different: Nonmonetary costs to societies and individuals. *Society and Economy* 24(2), 219–230.
- Green, C. (ed.) (1997). *Globalization and Survival in the Black Diaspora: The New Urban Challenge*. Albany, NY: State University of New York Press.
- Hearn, J. C. (1991). Academic and nonacademic influences on the college destinations of 1980 high school graduates. *Sociology of Education* 64, 158–171.
- Hine, D. C. and McLeod, J. (eds.) (1999). *Crossing Boundaries: Comparative History of Black People in Diaspora*. Bloomington, IN: Indiana University Press.
- Hosler, D. and Gallagher, K. (1987). Studying student college choice: A three-phase model and the implications for policymakers. *College and University* 62(3), 207–221.
- Johns, R. L., Morphet, E. L., and Alexander, K. (1983). *The Economics and Financing of Education*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.
- Kogbara, D. (1999). A sense of elsewhere. *Mail on Sunday* 58–59.
- Leviton, S., Mangum, G., and Marshall, R. (1972). *Human Resources and Labor Markets: Labor and Manpower in the American Economy*. New York: Harper & Row.
- Merisotis, J. P (1998). Who benefits from education? An American perspective. *International Higher Education*. Chestnut Hill, MA: Boston College Higher Education.

- Okepewho, I., Davies, C. B., and Mazrui, A. A. (eds.) (1999). *African Diaspora: African Origins and New World Identities*. Bloomington, IN: Indiana University Press.
- Steele, C. M. (1999). Thin ice: "Stereotype threat" and Black college students. *Atlantic Monthly* **284**(2), 44–54.
- The United States Department of Labor (1999). *Labor Day 1999 Executive Summary: Future Work: Trends and Challenges for Work in the 21st Century*. Washington, DC: U.S. Department of Labor.
- Thurow, L. C. (1972). Education and economic equality. *Public Interest* **28**, 66–81.
- Tony Blair's war on poverty (1999). *The Economist* **352**(8138), 18–19.
- Working towards a solution (1999). *The Daily Telegraph*, 18.
- Wheelock, A. (1992). *Crossing the Tracks: How "Untracking" Can Save America's Schools*. New York: The New Press.

Beyond Racial, Ethnic, and Gender Bias in Education Statistics

J T Lott, Tamayo Lott Associates, Colesville, MD, USA

C Bennett, US Census Bureau, Suitland, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

In recent decades social statistics in federal and other statistical systems were used to measure the progress of the total population, selected racial and ethnic groups, and women. In societies composed of racial majority and minorities, measures are refined to comparing minority populations to the majority. (The term minority in the United States refers to historical racial minority groups (Black or African American, American Indian or Alaskan Native, Asian, Native Hawaiian, or Other Pacific Islander). The term minority also includes Hispanics who may be any race. It excludes non-Hispanic Whites.) When disproportionality or disparity exists and persists, the question of possible bias arises. Given the existence of equal opportunity policies and programs, including education ones, for several decades, it is time to consider the further utility of this disparity measure. This article suggests an alternative measure of productivity for the present and foreseeable future. A twenty-first-century perspective of education focuses on maximum productivity for each student which may include, but is not limited to, race, ethnicity, and gender. Twenty-first-century education will be multidimensional, engaging, and increasingly unbound by traditional paper media, requiring fluency in many languages. In today's global, knowledge-based economy, the economic growth and social well-being of nations around the world are increasingly dependent on a well-educated workforce and individual access to quality education. Shared knowledge, ongoing learning, and skilled intelligence are key factors defining education in the twenty-first century.

At the national level, the United States is a timely case to better understand relevant education and human capital for the twenty-first century and the role of bias. The evolution of racial, ethnic, and gender bias in education statistics is rooted in US legal and demographic frameworks. The US was founded as and continues to be a nation of immigrants and refugees. One-third of its population is minority whose share of the total population is projected to increase. The US joins other nations in addressing three major trends in twenty-first-century education: (1) the growing impact of electronic information, including statistical data, (2) continuing educational disparities despite greater choices, and (3) education beyond borders. At the international level, other federal statistical systems consider race, ethnicity, or gender as a proportion of the population and in measures of disparity and productivity. Globally, The Dakar Framework for

Action, Education for All: Meeting Our Collective Commitment of the World Education Forum reaffirms that education is a fundamental human right for all children and adults.

Toward a Twenty-First-Century Perspective of Education

Meaning and use of statistics and other data are based on context. The twentieth-century context for race, ethnicity, and gender in education statistics was the status of US minorities and females compared to white males. Traditional bias has focused on measures of inequity, specifically disparity between majority White men, who traditionally had greater access to education. This focus has become problematic as White men are a relative, not an absolute, standard of equality. In recent decades, access to education at all levels from preschool to postgraduate school is rising. In particular, minorities have increased their high-school graduation and college-enrolment rates. In the US and some majority White postindustrial nations, the addition of immigrants and foreign students/faculty form a sizable presence in higher education, particularly in the sciences, mathematics, statistics, and engineering. Regarding gender, the rising proportion of women in higher education has been accompanied by the decrease of men's enrolment in and graduation from selected degree programs. These shifts suggest an alternative twenty-first-century perspective of education on the role of racial and ethnic minorities and women that goes beyond closing disparities with White men. An alternative twenty-first-century perspective is to ensure education for all members of society that allows maximum productivity for an informed citizenry and competitive workforce beyond race, ethnicity, and gender.

We cite three trends that support such an alternative. First, ready access to electronic data means an increasing global capacity to collect more varieties and quantities of data. From a research perspective, this means going beyond historically labor-intensive longitudinal and cross-sectional survey data to embracing cutting-edge dynamic, real-time administrative, biometric, and geospatial data. From a societal perspective, the general public is inundated by statistics and other data 24 h, 7 days a week. Yet, higher expectations for accessible data continue. What are the implications for an educated society when individuals can

readily learn what they need, or think they need, online? What are the implications for the digital divide and thousands of years of fact-to-face and hands-on learning? These questions are addressed below in the section titled Growing Impact of Electronic Information, including Statistical Data. A second trend is that disparities across and within groups continue, despite greater educational choices and access. This occurs even as pluralism increases in the twenty-first century. In the United States, for example, traditional minorities were already becoming the majority in selected cities by 1990. At a global level more nations are becoming more pluralistic given greater migration. Are new disparities, biases, or other priorities arising from a change of relations? This question is addressed below in the section titled 'Continuing disparities despite greater choices and greater access'.

A third trend is that the concept of education in the twenty-first century has outgrown its twentieth-century meaning; in the twenty-first century, education is without borders. A strategic discussion in the twenty-first century is warranted on: (1) the idea of what is education; (2) what is the value and utility of education; and (3) who pays for it? These questions are addressed below in the section titled 'Education beyond borders'.

The Case of the United States: A Nation of Immigrants

For a better understanding of relevant future education and human capital, we use the case of the United States. From a demographic perspective, the US, has an over 200-year history of heterogeneous demography due to a variety of settlers from Europe, Africa, Asia and the Pacific, the Middle East, and the Americas. Except for Native Americans and African slaves, the US evolved from refugees, pioneers, and immigrants. Continued population growth and heterogeneity are due to open immigration and refugee policies based on family reunification, labor needs, and humanitarian concerns, and due to fertility rates at replacement levels or higher. Finally, as a nation with separation of powers between church and state, the US population tolerates various religions, languages, and cultures.

Local and state governments initiated universal public education in the nineteenth century for elementary and secondary education led by reformers such as Horace Mann. By 1870, every state provided free public education. Statistics on education was a natural response to the increase in the US population and the need for an educated labor force. The 1840 census was the first to include data on school enrolment. The 1870 census marked the first time that women were included in the labor-force count.

Civil rights statutes and educational/employment-opportunity policies also shaped data collection for a more diverse population and labor force. During World War II, women and minorities assumed positions traditionally reserved for White men in civilian and military occupations. Rosie the Riveter symbolized women's ability to do men's work. Minorities and women began their permanent expansion into nontraditional occupations. Integration of US military forces after World War II was the first postwar civil rights policy facilitating a multiracial labor force. This was followed by the 1954 Supreme Court decision *Brown versus the Board of Education* that struck down the policy of separate but equal public education between White and Black and other minority children. In 1965, the national-origins quota, which had maintained a White majority and non-White minority for decades, was abolished. With 1965, amendments to the McCarran-Walter Act, US born minority populations began to grow. Refugees of the Vietnam War further increased the growth of non-White population in the 1970s and the 1980s, especially in school enrolments from prekindergarten to high schools. These two processes of increased immigration, especially from Asia and Latin America and increased births of non-White populations were noted by careful observers at local, regional, and federal levels.

A steady increase in the proportion of females in higher education and in the labor force also occurred. The launch of Sputnik in 1958 awakened the US to the need to increase the pool of young Americans proficient in mathematics, science, and engineering. While young women in high school were encouraged to take higher level courses alongside their male classmates, they still comprised a minority in colleges in 1960. By the 1970s, affirmative action policies, notably Title IX of the Higher Education Act sponsored by Congresswoman Patsy Takemoto Mink, facilitated the admissions of women to higher education and professional schools traditionally limited to men.

Education was characterized by public funding, local and state control, separation of church and state, and an academic curriculum for an expanding industrial economy. Tuition was free or minimal in the public state universities. For example, through the 1970s, the University of California and state university systems did not charge tuition for undergraduates. University of California, Berkeley, the system's flagship, was and continues to be renowned as a world-class research center. American and foreign parents and their children actively sought, and many times sacrificed for, a US education. All the above, taken together, contributed to the rise of the US as an economic and intellectual world power.

Nevertheless, during the same period, two reports indicated that all was not well. In the 1970s, the US joined a growing list of nations including Canada, France, Germany,

Great Britain, Japan, the Netherlands, and the Philippines in producing social indicator reports (US Commission on Civil Rights, 1978). Such reports generally focused on indicators that measured production, consumption, and satisfaction. In its social indicators report, *Social Indicators of Equality for Minorities and Women*, the US Commission on Civil Rights (1978) focused on the degree of inequality in the distribution of resources in the society. The study analyzed data from the 1960 and the 1970 censuses and 1976 survey of income and education, a supplement to the current population survey. It focused on the underdevelopment of human skills through delayed enrolment, non-enrolment in secondary education; and nonparticipation in higher education. It examined rates of delayed education, being behind in school, high-school nonattendance rates, and educational attainment. In terms of labor-force participation, it looked at the lack of equivalent returns for educational achievement in terms of occupational opportunities and earnings, earnings for educational levels, and occupational mobility. The study found continuing and serious problems of inequality between women and minority men compared with majority men.

The second prescient report was the landmark *A Nation at Risk*, issued by the National Commission on Excellence in Education (1983). It identified major deficiencies in the overall quality of education in America and stated that “the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people” (1983). Further, it indicated that American students were no longer competitive with students in many other countries and “that individuals in our society who do not possess the levels of skill, literacy, and training essential to this new era will be effectively disenfranchised, not simply from the material rewards that accompany competent performance, but also from the chance to participate fully in our national life.” A high level of shared education is essential to a free, democratic society and to the fostering of a common culture, especially in a country that prides itself on pluralism and individual freedom. The report urged America to consider new approaches to public education while continuing to define public education as an institution financed and operated by the government.

In 2001, President George W. Bush addressed education reform with the No Child Left Behind Act. This Act called for education reform in public schools that ensures high standards and accountability to students with specific emphasis on mathematics and reading skills. Between 1972 and 2005, the percentage of racial and ethnic minority students enrolled in the nation’s public schools increased from 22% to 42%, primarily due to the growth in Hispanic enrolment (US Department of Education, 2007a). The number of children of ages 5–17 who spoke a language other than English at home more than doubled between 1979 and 2005. While Spanish is the predominant second

language in the US, many school systems have children from families speaking dozens of languages. For example, over 150 languages are spoken in the Miami-Dade School system. In Maryland, the second most frequently spoken language is Haitian Creole and in Iowa, it was Hmong. “It is clear that America must determine the manner in which we will manage this language diversity. But, in a real sense, we must also capitalize on these skills. National security concerns clearly call for a nation increasingly fluent in multiple world languages. With the diversity clearly present in our nation’s schools, these concerns can be addressed with our current population of students. America’s schools clearly face the challenge of including these children and to providing them with language acquisition skills necessary to function in an English-speaking culture. At the same time, we must understand that fluency in these languages is a cultural and economic benefit for the society as a whole” (National Association of State Boards of Education, 2008).

In 2007, the US population was estimated to be 301.6 million. The single-race White non-Hispanic population was 199.1 million or 66% of the total population and 34% or 102.5 million were racial and ethnic minorities (US Census Bureau, 2008a). According to data from the National Assessment of Educational Progress (NAEP), 33% of fourth graders and 26% of eighth graders scored below basic in reading. However, for both Black and Hispanic fourth graders, the percentages were greater than 50% and for eighth graders the figures were 53% for Blacks and 43% for Hispanics. At both the fourth and eighth grades, a higher proportion of males than females scored below basic in reading (US Department of Education, 2007b). A similar picture existed in mathematics. Eighteen percent of fourth graders and 29% of eighth graders scored below basic in mathematics. Similar to performance in reading, a larger proportion of Black and Hispanic students scored below basic in mathematics and males’ scores were less than that of females (US Department of Education, 2007c).

Have We Reached the Top? Educational Attainment Projections of the U.S. Population (Day and Bauman, 2000) illustrates some of the factors to be considered beyond race, ethnicity, and gender in educational attainment. These include age, cohort, immigration, and questionnaire effects. In answering the above question, Day and Bauman examined three major influences – the process of cohort succession, ethnic shifts in the US population, and immigration. These influences were also noted in an earlier report commissioned by the US Department of Labor, *Workforce 2020*. A major theme of both works is that minority education is directly tied to concerns about the quality of the American labor force. Day and Bauman found that the cohort succession rate remains powerful enough to raise overall education levels. They also noted, however, that if younger cohorts were to attain less education, then their

projections may be overly optimistic, and that any policy designed to address educational growth needs to be flexible enough to address the particular challenges facing various groups. This conclusion reinforces the complexity of understanding educational attainment, addressing disparity, and promoting productivity.

In summary, equal educational opportunities of women and historical minorities have expanded since the 1960s. Studies track growth in accession, retention, and graduation rates from high school and college. Of equal note is the growing presence of minorities and women as employees and managers in education professions, from school systems to flagship universities. Various indicators suggest that progress has occurred, especially among younger cohorts. Children of color and White girls have a range of educational and occupational opportunities open to them that were denied their parents and grandparents. Immigrants – both White and of color – continue to come to the United States so that their children can obtain a US education. At the same time, research suggests strengths and limits of educational data by race, ethnicity, and gender. Disparity gaps are closing in some areas but replaced by new ones. Progress is mixed. It is still important to be cautious concerning the explanatory power of data and possible misuse of statistics. They are limited given the paradox of diminishing competitive educational achievement amidst programs to close disparities. This is a crucial point as the United States shifts to greater proportions of minorities and women. At the global level, the parallel shift is migration of residents from the southern hemisphere to become immigrants in northern hemisphere nations, increasing their proportion.

Growing Impact of Electronic Information, Including Statistical Data

Past discussion on minorities and women in the US showed a dearth of data on these populations. That is not so in the twenty-first century with voluminous and various data sets. The challenge is to manage and make sense of data given the various research methods and formats and different definitions of these populations. Moreover, ready access to electronic data means an increasing global capacity and higher expectations to collect more varieties and greater quantities of statistical data 24/7.

Current debates on whether a digital world closes or widens disparities between haves and have-nots are not new. They mimic the debates on whether television was a medium for increasing public intelligence or shifting to the lowest common denominator of mediocrity. In *The Medium is the Massage*, McLuhan (1967), who also coined the term global village, illustrated that media are extensions of human senses, minds, and bodies. He noted

that worldviews and how humans process information are changed by the adoption of new media.

It is not yet clear what the implications are for an educated society when individuals can readily acquire what they need and want online particularly in the absence of universally adhered to standards and principles. What are the implications for the digital divide and thousands of years of face-to-face and hands-on learning? What is the role of wikis, blogs, and other nontraditional ways of creating, collecting, and verifying information and transferring knowledge that might suggest not just bias in educational statistics but also in what it means to be educated?

While technological literacy is crucial for an educated society and workforce, it does not replace the traditional human survival skills of critical thinking, processing of a wide range of knowledge and information, tacit knowledge, fluency in world languages, face-to-face group learning and communication, learning by doing, research and experimentation, and the apprentice-to-master learning tradition. Neither does it substitute for the tradition of scientific thought and method. Education is much more than quicker computation or easy information retrieval. Education entails interpretation, context, meaning, and reflection to appreciate and understand lag time between observation of phenomena and making sense of them.

Continuing Disparities Despite Greater Choices and Greater Access

The United States and other pluralistic societies are facing the paradox of continuing disparities despite greater choices for consumers, individuals, and groups. At a global level, more nations are becoming more pluralistic given greater migration. Much of this migration is primarily from Third World nations to First World nations. Nations such as China, Brazil, India, and Argentina comprise an intermediate group or second world emerging out of third-world poverty on the path to a sizable middle class. Are there new disparities and biases or perhaps other priorities arising when choices may not be limited to black and White, and to haves and have-nots?

For example, in the past 40 years, women have increased their proportion of undergraduate degrees. While they currently hold more than half of bachelor's degrees among adults 25 years and older in the US, men remain slightly more likely than women to hold at least a bachelor's degree (28% compared with 26%) (US Census Bureau, 2008b). About 33% of young women 25–29 had a bachelor's degree or more education in 2007, compared with 26% of their male counterparts. Yet these degrees are largely in female-dominant majors ensuring the continuing disproportion and segregation of women in traditional female occupations.

A second example is that, in 2006 about 49% of Asians aged 25 and older had at least a bachelor's degree compared with 30% of non-Hispanic Whites, 17% of Blacks and 12% of Hispanics (US Census Bureau, 2008b). The proportion of the foreign-born population with a bachelor's degree or more was 28% with 29% of the native population. However, the proportion of naturalized citizens with a college degree was 34% (US Census Bureau, 2008b). Other research indicates that Asian Americans equal, or surpass, Whites on standard achievement tests. Asian foreign nationals and other immigrants are disproportionately enrolled in and graduate from math, science, and engineering doctoral programs compared with Americans. On the other hand, other research raises the question of whether there is a quota on Asian Americans in admissions to first-tier schools.

Hirschman's life-course perspective and life-table research (Hirschman *et al.*, 2007) suggests that choices at various stages of an education do not occur in a vacuum. Individual educational behaviors are not isolated events. For example, dropping out of school is a significant milestone hindering access and ability to take advanced courses or graduating from high school let alone going to college. Tracing the enrolment status of cohorts of students through high school and the rigor of instruction they undertake in timely sequence provides new insights into the temporal dynamics of attrition, retention, exit, catching up, and reentry. Innovative work like that of Hirschman and his demographic colleagues provide integrated models that consider individual and systematic social background risk factors, actual school experiences, and cumulative decisions in the probability of educational success.

Students and their parents in the US find a wide array of competing choices. Superficially, such choices seem beneficial but they are not without consequences that may exacerbate disparities. For example, in recent decades, minority students have been recruited to attend first-tier universities so they can choose from a variety of colleges. One of the consequences has been creaming the historically Black colleges and universities and institutions traditionally serving Hispanic, American Indians, and Asian Pacific Islanders of top students and future leaders. Another choice is the heterogeneity of children in a classroom. This has gone from the legislative mandate to integrate White and non-White children in public schools to mainstreaming disabled children. Such heterogeneity can be overwhelming for students, teachers, and parents. Current consequences of such choice include placing gifted and talented magnet programs of primarily White students in primarily racial/ethnic minority schools or what some education critics call a school within a school. Other parents respond with counter-choices removing their children from public schools to private and charter schools or to home schooling. Along with public education, a host of private and corporate-education programs are now

available for a price. They provide certificates but not necessarily degrees. They may or may not be state certified. Another recent choice is that students have a range of faculty from tenured professors to lecturers. American universities more and more rely on teaching assistants and adjunct faculty. They may not have benefits and privileges or even an office in which to meet with students. This is a new disparity between first-tier and second-tier faculty, especially as the latter increase their proportion of total faculty. While more students can choose to go to college and earn a PhD, the reality is fewer numbers are being employed in tenure track positions. While there are more ways to fund an education, the reality is that undergraduate and graduate students even in public state universities incur sizable debt to attain their degrees.

These current practices and programs are being watched closely for new disparities, especially for students most at risk. Routine mechanical and clerical tasks are becoming obsolete and being replaced with computerized tasks. The new reality is that there may be more disparity as jobs in the twenty-first century will require critical thinking and problem-solving skills, not just multiple-choice or fill-in-the-blank mechanics. Employers in the twenty-first century require employees who possess an understanding of science, mathematics, and technology and the critical-thinking skills that underlie these topics. They look for employees with soft skills who possess diplomatic communication and team-building skills. They search globally for such employees.

Elimination of choices and access also occurs. Post-1996 research indicates that neighborhoods and schools are returning to levels of racial segregation witnessed in the 1960s. In California, in the aftermath of Proposition 209, the number of Black students on University of California or California State University campuses has declined or remained steady, while all other races increased enrolment over the decade. According to Allen and others, systematic reproduction of racial and ethnic disparities results in lower Black and Latino college enrolment; a negative climate for students of color in recruitment, retention, and achievement; the educational value of diversity not fully realized to benefit all students and faculty; and US competitive disadvantages in the global economy as producers and consumers (Allen *et al.*, 2004). Now, when the state most needs its public colleges and universities to educate the next generation of workers and to help boost economic prosperity, college opportunities are rapidly declining for African Americans and Latinos (Allen and Jaykuma, 2007).

In the twenty-first century, historical minorities and new immigrants from less-traditional areas including Latin America, Africa, the Middle East, and Asia provide a ready pipeline of human capital. This is imperative given the already high labor-force participation rates of American men and women. Racial and ethnic minorities comprise one-third of the population and are expected to

grow. Such experience may be instructive to other nations moving from a less-homogeneous society with a clear majority and minority in relationships of inequality. In the twenty-first century, postindustrial societies in a competitive global economy are dependent on a widely educated workforce that is composed primarily of women, traditional, racial, and ethnic minorities, and new immigrants from both very old and emerging cultures.

Education Beyond Borders

A third trend is that the concept of education in the twenty-first century has outgrown its twentieth-century meaning. It is time to move beyond measures of disparity by race, ethnicity, and gender to education beyond borders. A strategic discussion is warranted on: (1) the idea of what is education; (2) what is the value and utility of education; and (3) who pays for it?

The above are not academic questions. A parallel discussion is surfacing regarding bias/inequality in the labor force as employers shift from retirement systems with defined benefits to retirement systems with defined contributions. The notion of permanent jobs with a career ladder is giving way to serial, less-than-permanent jobs, and multiple careers with a growing contingent workforce; some of which is outsourced. In this new economic and global order, racial and ethnic minorities, and women are not only increasing their roles in education and the workforce, but they constitute the majority of intellectual and human capital required for a society's reproduction, productivity, and competitive capital as well. Traditional bias by race, ethnicity, and gender as depicted by statistics and other data based on twentieth-century theory and methods are insufficient to understand what is education and its value and costs to all.

Education takes place in conditions ranging from a one room, rural classroom with one teacher instructing students in many grades to individual students tapping into distance learning all over the world beyond physical borders. Degrees, credentials, and certifications connote a measure of formal education. For the twenty-first century, however, education is not necessarily synonymous with credits earned toward some finite degree or credential. Rather, ongoing learning and the sharing/creation of knowledge are lifelong processes for individuals and organizations. Increasingly, innovation, and creative productive capital demand a multidisciplinary approach, not specialization. That is, education is beyond borders.

This is not to say that there is no longer a role for formal education but to acknowledge the value of ongoing informal education and shared knowledge, especially performance-based knowledge and team work. For example, emergency and rescue workers and medical personnel continuously build upon formal education and training

by collectively adapting to real-life situations. They extend existing knowledge when faced with situations not covered in formal training. Formal and informal education contributes to shared institutional knowledge. Education in the twenty-first century builds upon a cumulative human heritage and legacy of shared creativity and innovation which requires multiple perspectives, interdisciplinary contributions, fluency in multiple languages, shared memories, and collective experience. As such, education is a fundamental public good for an informed society. It is inclusive not exclusive. It is accessible to all.

The case of California is instructive in this regard. Free public education for all was the prevailing philosophy of California for most of the twentieth century. For its historical agricultural and military industries and the emerging film/entertainment, aerospace, science, finance/banking, and silicon/electronic industries, investment in human capital defined the California educational system. Education was for all segments of the population. Access was through free public education from elementary and secondary levels to public higher education resulting in flagship public universities and world-renowned research centers. The state's exemplary system of public higher education provided a steady supply of college graduates for what was one of the country's most highly skilled labor pools. Educational opportunity was directly related to societal outcomes. A well-educated workforce was responsible for the economic success and competitive human capital of California. Nonetheless, with Proposition 13 in 1978, California reversed the earlier philosophy of education as a fundamental public good. Proposition 209 in 1996 went further with state voters amending the state constitution to prohibit state and local governments from discriminating against or granting preferential treatment to anyone on the basis of race or sex. In the first decade of the twenty-first century, not just in California, but in other public school systems and public universities, education is no longer free or readily affordable within many family budgets, except for the wealthy.

This is a serious dilemma for policymakers, educators, employers, and families. While education is not without costs, the lack of education is much more costly. Employers compete with one another for a finite global pool of skilled labor. Employers contend with employees who lack a work ethic. Teachers worry about teaching to the test and some students work only for a desired grade. Parents consider whether they will have to support their children in adulthood. Policymakers grapple with increasing children and adult-dependant populations, and a decreasing competitive workforce. Educational opportunity and access are related to societal outcomes. The cost of a poorly educated population in terms of workforce and citizenry can also go beyond borders. The challenge is to again view education as a public good and sound human-capital investment that appreciates.

The Case of the Global Village

The opportunities and possibilities of an educated world are boundless. From a data perspective, federal statistical systems are concerned with accurately describing human phenomena by demographic, geographic, social, and economic characteristics. They seek to address and measure in a better manner, the disparate and productive status of racial/cultural minorities and women through international conferences such as Challenges of Measuring an Ethnic World (Statistics Canada and the US Bureau of the Census, 1993). Such intellectual inquiry and civic accounting is part of the greater education of the global village. An electronically connected world with more rapid and frequent travel and residence abroad provides a greater proportion of the world's population to education beyond national or pedagogical boundaries. Study abroad, traditionally reserved for college student-exchange programs, is growing from elementary schools to corporations. Such exciting challenges and opportunities connect world citizens to different cultures, languages, and traditions of learning while sharing knowledge and affirming commonalities.

Migration across national boundaries, including the growing feminization of immigration, will continue due to needs of a global economy. Given the population imbalances by size, and age distribution between northern and southern hemispheres, new and younger workers from the latter will migrate to the former to attend to aging populations with growing dependency ratios.

Through individual nation's initiatives and collective ones, the global village is addressing the importance of education particularly for underrepresented and disenfranchised groups for the greater good. It still takes a global village to raise a child. *The Dakar Framework for Action-Education for All: Meeting Our Collective Commitment* was adopted by the World Education in Dakar, Senegal during 16–28 April 2000.

It is an example of a collective initiative drawn from education forums in sub-Saharan Africa, the Americas and the Caribbean, the Arab States and North Africa, Asia and the Pacific, and Europe and North America. The framework states that, "Starting from early childhood and extending throughout life, the learners of the twenty-first century will require access to high-quality educational opportunities that are responsive to their needs, equitable and gender-sensitive. These opportunities must neither exclude nor discriminate. Since the pace, style, language, and circumstances of learning will never be uniform for all, there should be room for diverse formal or less formal approaches, as long as they ensure sound learning and confer equivalent status" (UNESCO, 2000).

It further notes, "The right to education imposes an obligation upon states to ensure that all citizens have opportunities to meet their basic learning needs. Primary

education should be free, compulsory and of good quality. The education systems of tomorrow, however diversified they may be, will need to be transparent and accountable in how they are governed, managed, and financed. The indispensable role of the state in education must be supplemented and supported by bold and comprehensive educational partnerships at all levels of society. Education for 'All' implies the involvement and commitment of all to education." The cost and return on investment in education to individuals and society are being questioned nationally and globally. Productivity, rather than disparity, across populations nationally and internationally is a more realistic measure of returns on investment to education for current and future generations of a global village.

Summary

A twenty-first-century perspective of education focuses on maximum productivity for each student undemarcated by, race, ethnicity, and gender. It is inclusive of various learning styles complementing traditional time-tested ones with new innovations; greater access to, appreciation of, and facility with various cultures and languages; and teaching/learning through various media both low and high tech. Education will be multidimensional, engaging, and increasingly unbound by traditional paper media and traditional disciplines. Workers in the most productive and better paying jobs will be proficient in reading, writing, mathematics, and computer technology, while fluent in timeless soft skills of civility and world languages. Knowledge, learning, information, and skilled intelligence for the greater good will be the hallmark of education in the twenty-first century.

Bibliography

- Allen, W. R. and Jaykuma, U. (2007). Water, water everywhere and not a drop to drink: Higher education access and success for America's growing population of underrepresented students of Color. *International Joint Statistical Meetings*, Salt Lake City, UT.
- Allen, W. R., Suh, S. A., Gonzalez, G., and Yang, J. (2004). *Education Research: Explaining or Defending Who Wins and Who Loses in the Competition for Educational Achievement*. *International Joint Statistical Meetings*, Toronto, Canada.
- Day, J. C. and Bauman, K. J. (2000). Have we reached the top? Educational attainment projections of the U.S. population. *Population Division Working Paper No. 43*, US Census Bureau, Washington, DC.
- Hirschman, C., Pharris-Ciurej, N., and Willhoft, J. (2007). The process of High School attrition: An analysis of linked student records. *Working Paper*, Department of Sociology, University of Washington, Seattle, WA, USA.
- McLuhan, M. (1967). *The Medium is the Message: An Inventory of Effects*. New York: Bantam Books/Random House.
- National Commission on Excellence in Education (1983). *A Nation at Risk: The Imperative of Education Reform*. <http://www.ed.gov/pubs/NatAtRisk/risk.html> (accessed August 2009).

- Statistics Canada and the US Bureau of the Census (1993). *Challenges of Measuring an Ethnic World*. Washington, DC: US Government Printing Office.
- UNESCO (2000). *The Dakar Framework for Action-Education for All: Meeting Our Collective Commitment*. Dakar, Senegal.
- US Census Bureau (2008a). *Annual Estimates of the Population by Sex, Race, and Hispanic Origin for the United States*, April 1, 2000 to July 1, 2007 (NC-EST2007-03). <http://www.census.gov/popest/national/asrh/NC-EST2007-srh.html> (accessed August 2009).
- US Census Bureau (2008b). *2006 American Community Survey Data*. <http://www.census.gov/acs/www>. (accessed August 2009).
- US Commission on Civil Rights (1978). *Social Indicators of Equality for Minorities and Women*. Washington, DC: US Commission on Civil Rights.
- US Department of Education (2007a). *The Condition of Education 2007a*. http://nces.ed.gov/pubs2007/2007064_Commstat. (accessed August 2009).
- US Department of Education (2007b). Institute of Education Sciences, National Center for Education Statistics, National Assessment Center for Educational Progress. *Reading Report Card*, http://nationsreportcard.gov/reading_2007/r001.asp (accessed August 2009).

US Department of Education (2007c). Institute of Education Sciences, National Center for Education Statistics, National Assessment Center for Educational Progress. *Mathematics Report Card*, http://nationsreportcard.gov/math_2007/r001.asp (accessed August 2009).

Further Reading

- Federal Interagency Committee on Education (1975). *Report of the Ad Hoc Committee on Racial and Ethnic Definitions of the Federal Interagency Committee on Education*. Washington, DC.
- Hirschman, C. (2006). *The American Century*. Presidential Address, Population Association of America, April, San Francisco, CA.
- Judy, R. W. and D'Amico, C. (1997). *Workforce 2020: Work and Workers in the 21st Century*. Indianapolis, IN: Hudson Institute.
- Smith, R. (1979). *The Subtle Revolution: Women at Work*. Washington, DC: The Urban Institute.

Caste and Education in India

G G Wankhede, Tata Institute of Social Sciences, Mumbai, India

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Manu – A sixth-century Hindu scholar and philosopher. The laws of Manu consist of 2685 verses on topics interrelated in Hindu thought which is in sum an encompassing representation of life in the world, as it is, and how it should be lived.

Purusha – This finds mention in the Rig Veda in Hindu spiritual literature and in its most literal sense means man. This is not a generic for human, but includes the masculine sense and principle of the world. He was the source of all creation. It describes this form of his, as having countless heads, eyes, legs, manifested everywhere, and beyond the scope of any limited method of comprehension. All creation is but a part of him.

Varna – Taken from the Sanskrit word which means to choose from a group. Hindu society was divided into four *varnas*, the Brahmin, *Kshatriya*, *Vaishya*, and *Shudra* and these groups had social and occupational differences among them. Varna literally means color referring to distinctions between dark and fair.

Vedas – The Vedas are the ancient scriptures of the Hindu teachings which manifest the divine word. There are four Vedas: the Rig Veda, Sama Veda, Yajur Veda, and Atharva Veda. The Vedas contain hymns, incantations, and rituals from ancient India.

Introduction

India has attracted the attention of several scholars both national and international particularly in the field of sociology due to the intricacies woven into the country's social fabric. One among these has been the interest in a particular aspect of stratification, a term considered a unique feature of Indian society – caste. This stratification refers to a hierarchical division between people on various grounds like religion, caste, occupation, etc. Intrinsically woven into any system of stratification is the notion of superordination and subordination, one that divides society into different strata. In Indian social reality, the caste system emerged as a major base for stratification.

The multiplicity of definitions of the term caste suggests the wide interest in the concept and brings to light

different perceptions on the issue. While historically the effects of the caste system were seen in every sphere of life, today the same exists but mostly has subtle undertones. The word caste has Spanish and Portuguese origins and comes from the word *Casta* meaning lineage or race. It began to be used in the Indian context in the seventeenth century. Various definitions of caste brought forth a few similar key features:

- functional interdependence – economic, political, and ritual;
- rules that safeguard level of caste-group purity;
- hereditary occupation;
- no opportunity for members of lower groups to rise into the upper groups;
- backing of social institutions: as a subsystem and sub-subsystem;
- necessary justification provided for enforcement of caste duties and obligations by Hindu religion;
- ascribed status;
- closed system;
- endogamous;
- segmentary – divided into subcastes which are themselves endogamous;
- predicated on graded social inequality;
- purity–pollution; and
- hierarchy.

Origin of Caste

Many studies of caste have reached different conclusions about its origin. One theory was that of Divine Will indicating division of *Purusha* – the creator into the four main *Varnas*, the Brahman, *Kshatriya*, *Vaishya*, and *Shudra* (Table 1). Paralleling this theory and drawing from the Rig Vedic hymns was the law of Manu justifying social differences by laying out the tasks for each caste. Another explanation was the Aryan invasion of India wherein they defeated the dark-skinned indigenous Dravidians. Thus, it was believed that the Brahmins hailed from the Aryans whereas the *Shudras* hailed from the Dravidians.

One of the reasons why the caste system was able to exist was the functional interdependence in Hindu society, permitting a wide range of differentiation without hampering the social fabric. The religious basis given to the caste system made it easier for the upper castes to perpetuate the differences and enjoy the accruing privileges of suppressing the lower castes.

Table 1 Broad *varna* and caste categories with their occupational and social status and estimated population^a

<i>Varnas</i>	<i>Broad categories of castes</i>	<i>Social status and occupations</i>	<i>Percentage</i>
1. Brahmin	Brahmins (upper (high) castes)	Performing rituals and teaching	3–4
2. <i>Kshatriya</i>	Middle castes	Warriors and landlords dominant	20
3. <i>Vaishya</i>		Traders and moneylenders	
4. <i>Shudra</i>	Lower castes (OBCs)	Artisans Peasants	52
5. Scheduled castes (untouchables)	Lowest but out of <i>varna</i> system	Menial occupations Polluting status	16
5. Tribals		Primitive lifestyle Physical isolation	8

^aEach *varna* consists of several castes based on hierarchy among themselves and all of them practice untouchability with scheduled castes (SCs). SCs were constitutionally recognized after independence based on their polluting status, as untouchables with special provisions for their socioeconomic development. Even the SCs numbering 1092 have hierarchy and practice of untouchability among themselves. Untouchability was legally banned in 1955. Enumeration of SCs and scheduled tribes (STs) is done only by the Government.

The interpretation of caste saw heterogeneous conceptions. While some saw it as a division of labor, others such as Dr Ambedkar pointed out that it was in fact a division of laborers into water-tight compartments graded one above the other. The assertion of an individual's own freedom of expression is the way to social reform; if in fact this desire is thwarted and controlled by certain groups, it cannot lead to any kind of reformation. The caste system through its rigid control kept the lower castes from acquiring the intellectual and social inheritance that the upper castes were fortunate to have received.

Contemporary Indian society is also witness to social discrimination based on caste divisions. Disadvantaged groups in society today include the scheduled castes (SCs), scheduled tribes (STs), and the other backward classes (OBCs). Legally they have been referred to as follows, in the Constitution of India:

- Article 341: Scheduled Castes

1. The president may with respect to any state or union territory, and where it is a state, after consultation with the governor thereof, by public notification, specify the castes, races, or tribes or parts of or groups within castes, races, or tribes which shall for the purposes of the constitution be deemed to be scheduled castes in relation to that state or union territory, as the case may be.
2. The parliament may, by law, include in or exclude from the list of scheduled castes specified in a notification issued under clause (1), any caste, race, or tribe or part of or group within any caste, race, or tribe, but save as aforesaid a notification issued under the said clause shall not be varied by any subsequent notification.

- Article 342: Scheduled Tribes

1. The president may with respect to any state or union territory, and where it is a state, after consultation with the governor thereof, by public notification, specify the tribes or tribal communities or parts of or groups within tribes or tribal communities which shall for the purposes of the constitution be

deemed to be scheduled tribes in relation to that state or union territory, as the case may be.

2. Parliament may by law include in or exclude from the list of scheduled tribes specified in a notification issued under clause (1), any tribe or tribal community or part of or group within any tribe or tribal community, but save as aforesaid a notification issued under the said clause shall not be varied by any subsequent notification.

- Article 340(1): Other Backward Classes

The president may by order appoint a commission, consisting of such persons as he thinks fit to investigate the conditions of socially and educationally backward classes within the territory of India and the difficulties under which they labor and to make recommendations as to the steps that should be taken by the union or any state to remove such difficulties and as to improve their condition and as to the grants that should be made, and the order appointing such commission shall define the procedure to be followed by the commission.

- Article 340(2)

A commission so appointed shall investigate the matters referred to them and present to the president a report setting out the facts as found by them and making such recommendations as they think proper.

Adhering to Article 340, the first backward classes commission was set up by a presidential order on 29 January 1953 under the chairmanship of Kaka Kalelkar.

For identifying socially and educationally backward classes, the commission adopted the following criteria:

1. low social position in the traditional caste hierarchy of Hindu society;
2. lack of general educational advancement among the major section of a caste or community;
3. inadequate or no representation in government services; and
4. inadequate representation in the field of trade, commerce, and industry.

The commission submitted its report on 30 March 1955. It had prepared a list of 2399 backward castes or communities for the entire country and of which 837 had been classified as the most backward.

There was considerable divergence of opinion among the members of the commission as to what should be the criteria of backwardness. However, the commission in its final report recommended caste as the criteria to determine backwardness. However, this report was not accepted by the government as it feared that the backward classes excluded from the caste and communities selected by the commission may not be considered and the really needy would be swamped by the multitude and would hardly receive special attention. Thus, there was a need of second backward classes of commission.

The decision to set up a second backward classes commission was made official by the president and the commission was popularly known as the Mandal Commission, its chairman being B. P. Mandal.

The Mandal Commission Report of 1991 was in favor of reservations in higher education and government services for the other backward classes of India.

With a view to provide safeguards against the exploitation of SCs and STs and to promote and protect their social, educational, economic, and cultural interests, special provisions were made in the Constitution. Due to their social disability and economic backwardness, they were grossly handicapped in getting reasonable share in elected offices, government jobs, and educational institutions and, therefore, it was considered necessary to follow a policy of reservations in their favor to ensure their equitable participation in governance. Thus, a national commission for scheduled castes and scheduled tribes was set up in 1978. Consequent upon the Constitution (Eighty-Ninth Amendment) Act, 2003 (Annexure II of the handbook), the erstwhile national commission for scheduled castes and scheduled tribes has been replaced by (1) national commission for scheduled castes, and (2) national commission for scheduled tribes.

Castes and Education

India is a country with 1028.7 million as its population according to the 2001 census report, with 72.2% living in rural areas and 27.8% in urban areas. The sex ratio is 933 females per thousand males. The population of scheduled castes is 166 635 700 accounting for 16.2% and scheduled tribes is 84 326 240 making up 8.2% of the country's population. Exact figures are not available but the population of OBCs is estimated to be at 52% of the country's total population. Demographically a diverse population, India is sated with caste and class problems. While education began inducing the spirit of social transformation, inequalities in the system still in certain ways

strengthened the hold of the upper castes. The link between caste and education has been there since the time of inception of the caste system in India. Formal education was confined only to the study of the Vedas in schools which were established for this purpose. The state did not hold itself responsible for opening establishments for the study of the arts and sciences that concerned the life of merchants and artisans. In the absence of a formal education system, each caste managed to transmit to its progeny the ways of doing things it was traditionally involved in doing. Therefore, illiteracy became an inherent part of the caste system by a process that was indirect to Hinduism. The education provided was religious education in the classical language of Sanskrit and only the Brahmins could teach the same. It was a direct result of the rule relating to the rights to teach and study the Vedas. Brahmins, *Kshatriyas*, and *Vaishyas* could study the Vedas, Brahmins could teach, and *Shudras* could not study and were not even allowed to hear it. If *Shudras* disobeyed, the offence was punishable with direct penalties.

In the period from seventeenth to nineteenth century, the British East India Company took over the Moghul empire in India and exercised political power in order to manipulate and control trade in India to their advantage. While they had no interest in reforming the Indian society socially or religiously, their endeavors to introduce development strategies were for their own vested interests. Bearing witness to this was the emphasis on English education also considered as modern which was restricted to reserved sections of society like children of employees of the company (East India Company), Anglo-Indian families, and Indian upper-caste Christians. Introduction of these new forms of education in a way was perpetuating the caste system by restricting access to people from the upper castes.

Beginning of the twentieth century, there was a tendency of caste organizations and associations to enter into education to spread education through the establishment of institutes by building schools and hostels for their own children.

Postindependence, national progress was the aim of policymakers and in order to achieve this, formulation of national policies on education became the need of the hour.

The educational developments in the postindependence period can be divided into three main periods:

1. the first is the period between 1947 and 1965–1966 when there was no formal statement of a national policy on education;
2. the second is the period between 1965–1966 and 1977–1978 when the national policy on education (1968) was formulated and an attempt was made to implement it; and
3. with 1978–1979 begins the third period when an attempt was made to implement the draft on the national policy on education (1979).

Several policies and programs were continuously being developed since the time of independence and while they had their flaws, they contributed to the country being able to move ahead in the field of education. When India became independent, it committed itself to a socially just and equal social order. The constitution of the country, promulgated in 1950, recognized the scheduled castes and the scheduled tribes as the two most backward groups needing special protection. There are 1092 scheduled castes, 547 scheduled tribes and 150 million is the population. A number of provisions have been made in the constitution with a view to abolish all forms of discrimination and put these groups at par with others. According to the 2001 census, the literacy rate of schedule castes was 54.69% and that of schedule tribes was 47.10% as against the national average of 64.84%. In 2001, the Government of India introduced the 93rd amendment to make free and compulsory elementary education for children in the age group 6–14, a fundamental right. The reasons for illiteracy among various caste groups in the country including SCs/STs are various socioeconomic factors such as poverty, lack of awareness on the relevance of education, gender disparity, etc. To better understand the situation on caste discrimination with respect to education, one has to look at three levels of education – primary, secondary, and higher education.

Educational Problems

High dropout and lower literacy rates **Tables 2** and **4** among lower caste populations have rather simplistically been characterized as the natural consequences of poverty and underdevelopment. Although these rates are partly attributable to the need for low-caste children to supplement their family wages through labor, more insidious and less well documented is the discriminatory and abusive treatment faced by low-caste children who attempt to attend school, at the hands of their teachers and fellow students. Over 50 years since India's constitutional promise of free, compulsory primary education for all children up to the age of 14 – with special care and consideration to be given to promote the educational progress of scheduled castes and tribes – illiteracy still plagues a large percentage of the their population.

SCs and STs are considered the most backward sections of Indian society in terms of educational development **Table 3**. They constitute the statutorily recognized weaker sections of society and form a distinct target group under the existing pattern of planning. The Government of India has assumed a wide-ranging responsibility for the educational upliftment of the disadvantaged groups. This has included free primary education, reservation in certain centers of higher education, and financial assistance in a variety of educational programs, to name a few. All these policies and programs are essentially designed to allow the lower castes to avail of these opportunities and the assumption is that their rapid advancement will bring them at par with the rest of the population.

Indisputably, education holds the key to socioeconomic development; however, the system of education itself is colored by several biases of caste, language, economic gradations, and gender. These perpetuate the inequality that exists in society. The Government of India was aware of the hierarchical nature of Indian society and hence education was considered necessary for social change. Indian society appears to be still categorized as a pyramid with select strata of the population which form the tip of this pyramid, enjoying the benefits of education. Access to higher education is a seemingly fair practice with a universal entrance exam but on closer inspection, the very nature of this examination favors the upper castes of society to whom, the benefits of education have accrued since a long time. Students from vernacular backgrounds are most often if not always expected to answer an entrance examination in English. In a multicultural and multilingual country like India, a universal language as the general medium of instruction has not spread to all the corners of the country. With SCs and STs receiving their primary and secondary education in a medium of vernacular instruction, they are most often at a disadvantage at centers for higher education. Even though the Government of India has now made it an available option for students to answer their examinations in their own vernacular language, the medium of instruction in classrooms remains English and the reference material suggested is in English.

Language is not the only barrier to the access of education; there are also gender disparities that are characteristic of Indian education. Factors for low enrolment of girls include low socioeconomic status of the parents,

Table 2 Population and literacy in India according to the 2001 census

	<i>Total</i>			<i>Scheduled castes</i>			<i>Scheduled tribes</i>		
	<i>Male</i>	<i>Female</i>	<i>Total</i>	<i>Male</i>	<i>Female</i>	<i>Total</i>	<i>Male</i>	<i>Female</i>	<i>Total</i>
Population (in millions) 2001	532	496	1028.7	86	80.5	166.6	42.6	41.7	84.3
Literates (in millions) 2001	336.5	224.2	560.7	47.4	27.9	75.3	20.6	11.8	32.4

From www.censusindia.gov.in.

Table 3 Enrolment data for India covering primary, middle, and secondary education for the year 2004–2005

<i>Enrolment</i>	<i>Total (figures in millions)</i>			<i>Scheduled castes (in thousands)</i>			<i>Scheduled tribes (in thousands)</i>		
	<i>Male</i>	<i>Female</i>	<i>Total</i>	<i>Male</i>	<i>Female</i>	<i>Total</i>	<i>Male</i>	<i>Female</i>	<i>Total</i>
Primary (1–5)	69.7	61.1	130.8	13762	10995	24757	7367	6369	13737
Middle/upper primary (6–8)	28.5	22.7	51.2	5100	3597	3597	2395	1776	4171
Sec./Sr. Sec./Inter/Predegree (9–12)	21.7	15.4	37.1	3228	3228	5218	1290	795	2085

From Selected educational statistics 2004–2005, Government of India; Ministry of Human Resource Development, New Delhi.

Table 4 Dropout rate in India covering primary, elementary, and secondary education

<i>Dropout rate (%)</i>	<i>Total</i>			<i>Scheduled castes</i>			<i>Scheduled tribes</i>		
	<i>Male</i>	<i>Female</i>	<i>Total</i>	<i>Male</i>	<i>Female</i>	<i>Total</i>	<i>Male</i>	<i>Female</i>	<i>Total</i>
Primary (1–5)	31.81	25.42	29	32.7	36.1	34.2	42.6	42	42.3
Elementary (1–8)	50.49	51.28	50.84	55.2	60	57.3	65	67.1	65.9
Secondary (1–10)	60.41	63.88	61.92	69.1	74.2	71.3	77.8	80.7	79

From Selected educational statistics 2004–2005, Government of India; Ministry of Human Resource Development, New Delhi.

who are more interested in their daughter helping with earning an income for the family and dropout rates are generally high in secondary school with hardly a handful being able to avail of higher education opportunities. In the multicultural and multiethnic Indian society, the parameters of gender, caste, class, and region are crucial in determining access to higher education. Again, gender becomes the all-inclusive negative parameter conferring cumulative and competing disadvantages on women. The educational policies and programs are unable to encompass the complex social reality within a single framework and are, therefore, unable to bridge the gap between policy and practice.

The disadvantaged groups do not have the means to keep their children in school beyond the middle stage, and the small minority that manages to do so can only send them to government schools. None can afford private coaching. Despite these handicaps, a sizable number, much larger than is recognized of children from these groups, manage to do well, and some very well, in public exams. But their family means do not even allow them the luxury of even thinking of higher education. Most educational institutions have a certain number of seats set aside as reserved seats for these communities. A certain group of people especially those against the policy of reservations in the country believe that reservation in admissions to institutes of higher education discriminates against merit and will result in further deterioration of quality. Moreover, it is important to note that reservation in higher education as a whole and in professional courses in particular has focused exclusively on caste as the criterion of backwardness and created strong pressures for widening the scope of the OBCs. This has led to the

cornering of the benefits of reservation by the better-off segments, the so-called creamy layer, of these castes.

Summary

As education connects more closely with economics and development, it assigns certain status to individuals. The years of schooling and special training for career growth make some more advantaged than other. Thus equality of educational provisions and access across divisions of class, caste, and race becomes a very important social and educational problem.

Education cannot be thought of in isolation, it must be studied in the socioeconomic and cultural framework of the society that include the communities that have remained handicapped educationally. It remains beneficial for those who have traditionally been at that level for whom education was always available. Contemporary analysis of the Indian education system should look at access, affordability, performance, and utility as the indicators to measure success of planning initiatives. In this regard, it is important to note that while education is assumed to be a great leveler of society, in a way with its policies and frameworks it is in fact clutching the reins of inequality. Access to education seems the privilege of those who are assumed to benefit most from it and in India, the upper castes.

Development is the new catch word of the twenty-first century in the globalizing world and while every country strives to join this bandwagon, little is done to alleviate the situation of the country's less fortunate. High-caste positions have since time immemorial meant high-class

positions in society. Today, efforts are on to better the lives of people belonging to the SCs, STs, and OBCs who form the highest population proportion in India but for any of these endeavors to be a success, one needs the cooperation of the entire nation's people.

Further Reading

- Ambedkar, B. R. (1979). *Dr Babasaheb Ambedkar Writings and speeches-Vol. 1*. Maharashtra: The Education Department, Government of Maharashtra.
- Anitha, B. K. (2000). *Village, Caste and Education*. New Delhi: Rawat Publications.
- Broken People: Caste violence against India's untouchables (1999). *Human Rights Watch*. Washington: Books for change.
- Dutt, N. K. (1968). *Origin and Growth of Caste in India, Calcutta*. Calcutta: Asiatic Society of Bengal.
- Ghurye, G. S. (1996). *Caste and Race in India*. Bombay: Popular Prakashan.
- Godbole, M. (2001). Elementary education as a fundamental right: The issues. *Economic and Political Weekly* **15**, 4609–4613.
- Hutton, I. H. (1951). *Caste in India*. London: Oxford University Press.
- Jaffrelot, C. (2003). *India's Silent Revolution – The Rise of the Lower Castes in North India*. London: C. Hurst.
- Jaiswal, S. (1998). *Caste: Origin, Function and Dimensions of Change*. New Delhi: Manohar Publishers and Distributors.
- Ketkar, S. V. (1979). *History of Caste in India*. Jaipur: Rawat Publications.
- Mathur, M. L. (2005). *Caste and Educational Development: Before and After Independence*. Delhi: Kalpaz Publications.
- Rao, R. S. (1989). *Caste System in India: Myth and Reality*. New Delhi: India Publishers and Distributors.
- Rao, S. S. (2002). Dalits in education and workforce. *Economic and Political Weekly* **37**, 2998–3000.
- Sharma, K. L. (1994). *Social Stratification and Mobility*. Jaipur: Rawat Publications.
- Srinivas, M. N. (1996). *Caste: Its Twentieth Century Avatar*. New Delhi: Penguin Books.
- Srinivas, M. N. (1998). *Village, Caste, Gender and Method: Essays in Indian Social Anthropology*. Delhi: Oxford University Press.
- Teltumbe, A. (2005). *Anti-Imperialism and Annihilation of Castes*. Dombivli: Ramal Prakashan.
- Thorat, S. and Kumar, N. (eds.) (2008). *Perspectives on Social Exclusion and Inclusive Policies*. New Delhi: Oxford Press.
- Wankhede, G. G. (1999). *Social Mobility and Scheduled Castes*. Jaipur: Rawat Publications.
- Wankhede, G. G. (2002). *Dalits in India: A Critique of the Issues and Challenges – from Think India* vol. 5, No. 4, October–December 2002. New Delhi: Vichar Nyas.
- Wankhede, G. G. (2003). *Caste and Social discrimination*. In Sharma, M. C. and Sharma, A. K. (eds.) *A Handbook for Sensitizing Teachers and Teacher Educators*, pp 14–16 and 63–71. New Delhi: National Council for Teacher Education.

Relevant Websites

- <http://www.censusindia.gov.in> – Census of India.
- <http://www.indiastat.com> – Indiastat.com.
- <http://www.education.nic.in> – Ministry of Human Resource Development.
- <http://www.ncsc.nic.in> – National Commission for Scheduled Castes.
- <http://www.ncst.nic.in> – National Commission for Scheduled Tribes.
- <http://www.ncbc.nic.in> – National Commission for Backward Classes.

Demography, Community, and the Education of Asian American and Pacific Islander Students

R T Teranishi, New York University, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

There are tremendous challenges to addressing the educational needs of the Asian American and Pacific Islander (AAPI) student population. The challenge is not merely that the educational solutions are difficult to come by; rather, the root of the problem stems from the indelible idea that the AAPI population is a model minority – a racial minority population that succeeds despite the odds. Consider the December issue of *US News and World Report*, in 1966, which states bluntly, “At a time when Americans are awash in worry over the plight of racial minorities, one such minority is winning wealth and respect by dint of its own hard work – not from a welfare check.” (Success Story of One Minority Group in the U.S. (1966, December 1966). *U.S. News and World Report*.)

In 1971, *Newsweek* had a feature story that proclaimed that Asians are outwhiting Whites. *Time Magazine* ran an article declaring that Asians are The New Whiz Kids (August 1987). More recently, in May 2006, a *New York Times* column on The Model Students declared that “stellar academic achievement has an Asian face” and that others would be fools not to learn from these perfect students.

The perspectives with regard to the educational success of the AAPI population have led to the idea that they are a population that does not need attention in research and policy. Thus, unlike other minority populations in the US, the AAPI population faces the unusual obstacle of overcoming the stereotype that they are a successful minority and the idea that they are not educationally disadvantaged such as other minority groups. However, an informed discussion of access and equity requires an understanding of the AAPI educational experience within the context of the conditions that are affecting all students. Moreover, an understanding of the AAPI educational experience requires a contextual examination of the population itself. The AAPI educational experience is actually a conglomeration of varied experiences, opportunities, and outcomes driven by a wide distribution of demographic characteristics with regard to national origin, ethnicity, immigration histories, social class, and cultures. In fact, there are more differences between groups than similarities.

While many studies have attempted to examine the AAPI population, most have been comparative by design, measuring the outcomes of AAPIs against those of other

racial groups. These studies are often guided theoretically and conceptually by frameworks that have been applied to the study of race in the context of the black/white binary, which has been found to have several limitations for the study of the AAPI educational experience. In other words, a key to understanding the position of AAPIs within the equity agenda is to first understand that the exclusion and misrepresentation is predicated upon two key conceptual problems in the study of race. First, normative framing – a perspective that is comparative in nature – has historically been focused on the gap between blacks and whites, without much consideration for other racial populations. While there is certainly a place for the use of comparative research, it should not be the only basis for understanding race in America. In fact, scholars have noted that this perspective often leads to a system of deficit thinking – an idea with regard to marginalized groups that Valencia (1997) states is “tantamount to ‘blaming the victim’ . . . [and is] founded on imputation, not documentation” (pg xi). Allen (1999) cites this paradigm as commonly applied to educational research as an often “misguided and counterproductive game of ‘oppression sweepstakes’ ” (pp. 206–207), whereby various groups are pitted against one another in competition for the dubious status of “Year’s Most Oppressed.” Blacks and Latinos are viewed as a problem in comparative research, while AAPIs are viewed as a solution. This leads to the second problem with the conceptualization and positioning of AAPIs in America’s equity agenda, which are the deeply held beliefs and stereotypes that people hold about AAPIs as a model minority.

At present – in the world of higher education research and policy – some of the key assumptions that drive the treatment of AAPIs in higher education state that: (1) the AAPI population is a successful minority group with high achievement and no academic challenges, and (2) there is an Asian invasion in colleges and universities and the AAPI population is overrepresented in the US higher education. These ideas about AAPIs are not just restricted to research and policy. Rather, they can also be found in the public’s perception that AAPIs are a successful model minority in American society – a population that has lived the American dream, overcoming barriers, and succeeding despite their challenges. While some Asian Americans have achieved the American dream – similar to many other immigrants that have arrived in the United States – the lives of the

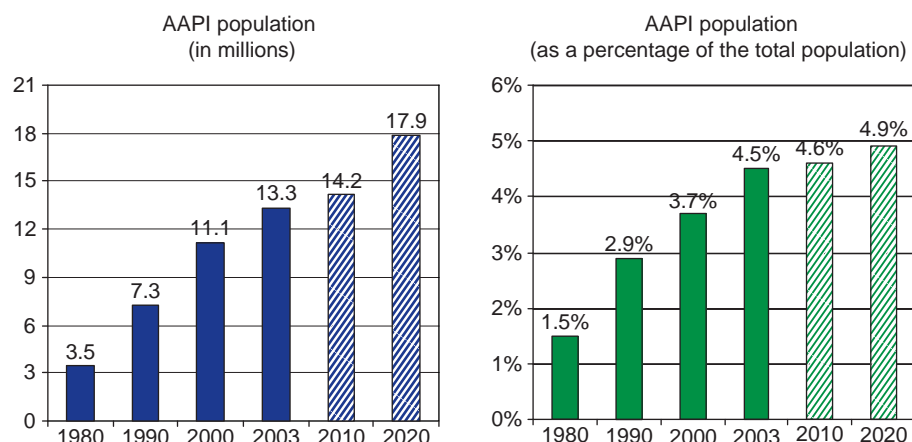


Figure 1 AAPI population in millions and as a percentage of the total US population, 1980–2020. 1980 to 2003 are actual data; 2010 and 2020 reflect projections. From US Census Bureau, decennial census of population, 1980–2000.

many individuals and communities that comprise Asian Americans cannot and should not be reduced to a single story (Zia, 2001). Thus, while there is tremendous interest in rectifying inequality in the educational opportunities and outcomes of different racial groups in America, the study of AAPIs, ironically, is a major obstacle for understanding the AAPI educational experience.

This article attempts to transcend these conceptual challenges both by systematically examining some basic baseline information concerning AAPI participation and outcomes in US higher education, and discussing the ways in which demographic distinctions of the AAPI population has implications for the educational experiences and outcomes of different AAPI subpopulations. In particular, this article describes the ways in which differences among the population – with regard to ethnicity, immigration histories, and social class – influence differences in post-secondary opportunities and outcomes that exist among the population. This article will demonstrate that a broader, more thoughtful consideration of the subpopulations that contribute to the larger AAPI racial category can lead to better research, policy, and practice.

Higher Education Participation and Attainment

AAPI participation in US higher education can be traced back to late in the nineteenth century. However, as late as 1940, the Asian American population was still considerably less than one percent of the US population (Daniels, 1997). Exceptions of this trend could be found in Hawai'i – which was then a territory and, therefore, excluded from US population figures. In Hawai'i, 58% of the people, in 1940, were of Asian descent (Tamura, 2001). The growth of the AAPI population in mainland US

	Number	10-year percentage change
1980	271 006	–
1990	570 873	110.7%
2000	903 360	58.2%

Figure 2 Total AAPI enrolment in US higher education, 1980–2000. From US Department of Education, IPEDS.

began with the passage of the 1965 Immigration Act, which created many new opportunities for immigrants that were formerly excluded. Between 1960 and 1994, the Asian American population grew from 1 million to nearly 8.8 million people. The sizable growth became more evident, beginning in 1980 when the population reached 3.5 million (see **Figure 1**).

Over the past three decades – from 1970 to the present – AAPIs have been the fastest growing population in the United States. At present, the AAPI population is large and continuing to grow. In 2003, there were over 13 million AAPIs in the United States, representing 4.5% of the total population. In 2020, the population is projected to be nearly 18 million. By 2050, the AAPI population is projected to be nearly 10% of the total US population (not in table).

The presence of AAPIs in US higher education coincides with the growth in the overall population – which also occurred following the 1965 Immigration Act. There were more than a quarter of a million AAPIs enrolled in US higher education in 1980, nearly a half million in 1990, and almost one million in 2000 (see **Figure 2**). These were similar trends to what occurred for the total AAPI population, with the total enrolment of AAPIs in US higher education doubling between 1980 and 1990, and nearly tripling between 1980 and 2000.

While the growth in AAPI participation seems drastic, it is typically overstated and assumed to be greater than

that of other racial minority groups. However, the growth of AAPI total enrolment is actually quite similar to the pace of the growth among African Americans and Latinos. For example – between 1993 and 2003 – Latino total enrolment in US higher education increased by 68.8%, compared to 43.5% for AAPIs, and 42.7% for blacks. In fact, the change in participation among AAPIs is tied mostly to two distinct demographic features. First, the sheer size and growth of the total AAPI population created a larger pool from which AAPI college students could be drawn. Second – during the past three decades – there has been a general shift in opportunity for higher education that has led to greater opportunities to attend college for all minority students in the United States.

While there has been tremendous growth in AAPI participation in US higher education overall, it is important to also consider the distribution of AAPIs in different regions of the country. Colleges and universities are distributed unevenly throughout the United States, with some states and regions having a larger student population than others. The disparate distribution of student populations across different states and regions is often correlated with the disparities in the distribution of the general US populations throughout the country. For example, analysis of population estimates, in 2003, shows that a little over half (51.2%) of the US population was located in just eight states. Twenty states comprised about three-quarters (76.7%) of the US population (See ‘Annual estimates of the population by race alone or in combination and Hispanic or Latino origin for the United States and States’, 1 July 2003. Population Division, US Census Bureau (US Census Bureau, 2003)).

The distribution of the total US population in different states and regions of the country is not necessarily the same as the distribution for individual racial groups. In 2003, AAPIs were more concentrated in fewer states than the general US population, with half of all AAPIs residing in three states (California, New York, and Hawai‘i). Ten states comprised approximately three-quarters (74.4%) of the total AAPI population. The residential patterns of AAPIs are tied – in large part – to patterns of immigration settlement and migration resettlement with concentrated populations of AAPIs in large urban communities such as New York, Los Angeles, Long Beach, San Francisco, Oakland, Honolulu, Houston, and Jersey City (see Teranishi, 2004).

Similar to the residential distribution of AAPIs nationally, the regional distribution of AAPIs in colleges and universities throughout US higher education is also unique. For instance, there is a large concentration of AAPIs in a small number of institutions. In 2000, two-thirds of all AAPI college students nationally were concentrated in 200 institutions out of more than 4200 Title IV institutions in the United States. These 200 institutions represent less than 5% of all Title IV institutions. Three-quarters of all AAPI college students nationally are concentrated in less than 300 institutions in the United States. The distribution of

	Total FTF to any college	Percent of national total	Cumulative percentage
California	323 351	35.8%	35.8%
New York	67 725	7.5%	43.3%
Texas	50 578	5.6%	48.9%
Illinois	43 140	4.8%	53.7%
Hawaii	35 489	3.9%	57.6%
Washington	26 340	2.9%	60.5%
New Jersey	24 041	2.7%	63.2%
Massachusetts	23 718	2.6%	65.8%

Figure 3 AAPI first-time freshmen attending any college by state, 2000. From US Department of Education, IPEDS.

AAPIs across a small number of institutions is nearly always overlooked in research and policy considerations.

AAPI enrolment is also highly concentrated in a small number of states. In 2000, nearly two-thirds of all AAPI college students attended college in eight states (**Figure 3**). Nearly half of all AAPI first-time freshmen nationally attended college in California, New York, and Texas. In California alone, over 300 000 first-time freshmen were enrolled in higher education in 2000 – comprising over one-third of all AAPI students who entered college nationally.

While the largest concentration of the national AAPI student population can be found in the West and North-east, the greatest growth in AAPI college participation occurred in the Midwest and the South between 1990 and 2000 – a rate of 75% and 86%, respectively. These states have historically had a very small AAPI population overall and very few AAPIs enrolled in higher education. However, the Midwest and the South – in recent decades – have experienced an increase in their AAPI population caused by immigration settlement directly to these regions along with migration and resettlement among AAPIs looking for better occupational opportunities and a lower cost of living (see Teranishi, 2004).

In addition to the trends in the overall AAPI college-participation rate and their changing regional representation, it is also important to consider their distribution in different institutional types. US higher education is diverse and highly stratified with colleges that range in type (2-year and 4-year), selectivity (selective and nonselective), and control (public and private). The idea that AAPIs are universally successful is tied to the assumption that they are all selected by the most selective universities in the country. However, when total AAPI enrolment in US higher education is disaggregated by institutional type and control, there is evidence of a wide distribution of AAPI participation.

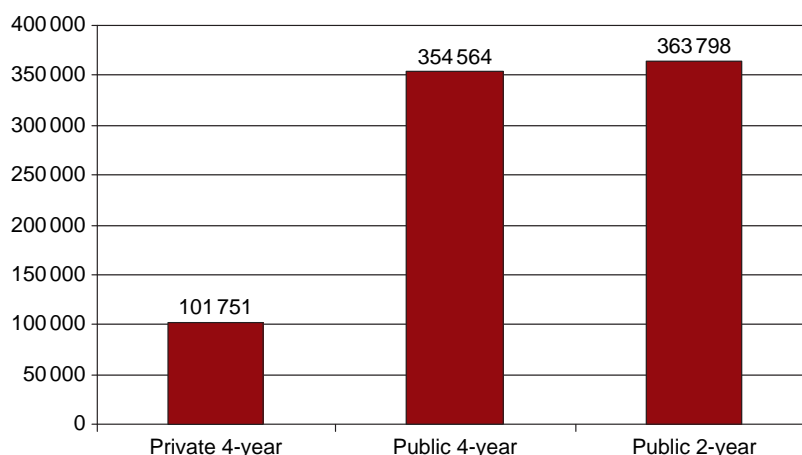


Figure 4 AAPI enrolment in US higher education by institutional type, 2000. Private 4-year does not include private for-profit institutions. From US Department of Education, IPEDS, 2006.

Most surprising in this analysis is the sizeable concentration of AAPIs in community colleges (**Figure 4**). In 2000, there were more AAPIs enrolled in public 2-year colleges (363 798 AAPIs) than in public 4-year colleges (354 564 AAPIs) or private 4-year colleges (101 751 AAPIs). AAPI college participation is rarely discussed in the context of the community college sector, and is, instead, mostly a discussion reserved for looking at the elite colleges, where it is assumed they are all located.

While a large number of AAPI students in higher education can be found in the community college sector, it is also important to recognize that the greatest increase in the number of AAPI college students between 1990 and 2000 also occurred in that sector. Between these time periods, the participation of AAPIs in public 2-year institutions increased by 73.3%, which was greater than their change in enrolment in public 4-year colleges (42.2%) and private 4-year colleges (53.4%).

These data demonstrate that AAPIs have a wide range of participation in US higher education, which varies by demographic region and institutional type. These varied outcomes among the population challenges the notion that AAPIs are all succeeding equally, and experiencing universal educational success. The next section considers more thoroughly the demographic distinctions within the population and the extent to which these differences factor into disparate outcomes within the population. Specifically, this author examines ethnicity, immigration status, social class, and language as predictive factors for educational outcomes.

Ethnicity, Social Class, and Disparities in Opportunities and Outcomes

Understanding the participation of AAPIs in US higher education requires consideration of a larger context of

AAPI demography and community. For the AAPI population, there is a great deal of diversity within the population that is masked by the larger racial category. The 1996/97 Minorities in Higher Education Report (Hune and Chan, 1997) indicated that aggregated data on Asian Americans homogenizes the experiences of Asians and depicts a distorted picture of the educational participation of subgroups within the population. This racial categorization hides the diversity within the population with regard to ethnicity, social class, and immigrant subgroups that encounter different social and institutional experiences. Unfortunately, lack of careful and critical approaches to the study of race in higher education results in misleading generalizations with regard to the population.

The limitations of existing research on AAPIs are not restricted to the examination of within-group distinctions; there is also evidence of confounding of data with other categories. Consider data from the Eighth Annual Status Report for Minorities in Higher Education (2001) which show that 32% of doctorates conferred in the US were to Asians; however, 86% of those degrees were actually conferred to international students from Asia, rather than to Asian Americans. A recent National Science Foundation (NSF) publication reported that, during the same time period, doctoral degrees awarded to Asians who were US citizens accounted for a mere 2% of all doctoral degrees awarded (Thurgood *et al.*, 2006). By reporting these racial categories and including international students, we use inflated numbers that exaggerate the achievement of Asian Americans. Such inflations prompt some scholars to describe Asian Americans as overrepresented in terms of earned doctorates; they are actually underrepresented.

In addition, it is also problematic when studies on race in higher education completely omit Asian Americans from the analysis, or even conveniently include them with the Others category – a practice that is also found

Asian American		Pacific Islander American	
Bangladeshi	Laotian	Carolinian	Papua New Guinean
Bhutanese	Malaysian	Chamorro	Pohnpeian
Burmese	Maldivian	Chuukese	Saipanese
Cambodian	Nepalese	Fijian	Samoan
Chinese	Okinawan	Guamanian	Solomon Islander
Filipino	Pakistani	I-Kiribati	Tahitian
Hmong	Singaporean	Kosraean	Tokelauan
Indian	Sri Lankan	Mariana Islander	Tongan
Indo Chinese	Taiwanese	Marshallese	Yapese
Iwo Jiman	Thai	Native Hawaiian	Polynesian
Japanese	Vietnamese	Ni-Vanuatu	Micronesian
Korean	Other Asian	Palauan	Melanesian

Figure 5 Asian American and Pacific Islander American ethnic categories. From US Census Bureau, 2000.

in the treatment of Native Americans. As a result, very little is known about the Asian American educational experience, including their participation, satisfaction, and outcomes in US higher education. This leads to the conclusion that there needs to be more research that understands the diversity of the AAPI population, including a better understanding of the differences across ethnic groups – with regard to social class, timing of migration to the United States, and population distribution across different states and regions within the United States.

Evaluating basic, baseline information pertaining to the AAPI population can yield important information to consider in educational research and policy. Consider one of the most obvious indications of the diversity among the umbrella racial category Asian American and Pacific Islander – ethnic diversity. The US Census Bureau (2000) included 48 different AAPI categories to make up the AAPI racial population.

Across racial groups, some populations have a greater likelihood of having a tighter distribution of education, income, and wealth compared to other groups. Consistently, the perception of AAPIs is that their educational attainment levels are exceedingly high compared to all other racial groups in the nation. Implicit in this assumption is the idea that both high educational achievement and attainment are universally attainable goals representative of the entire population. This is not only untrue, it is actually quite the opposite of what the trends show; AAPIs have a wider distribution of educational attainment than any other racial group in the United States. Consider that while AAPIs – in the aggregate – have a higher level of education than all other racial groups, a large proportion of Hmongs (59.6%), Cambodians (53.3%), Laotians (49.6%), and Vietnamese (38.1%) adults over the age of 25 do not have a high school diploma. The national averages for certain AAPI ethnic sub-populations are

even more striking when you isolate certain ethnic communities (see **Figure 5**). Consider the Hmong population in St. Paul, Minnesota – the largest of concentration of Hmong in the United States. Among Hmongs in the St. Paul community, nearly two-thirds (65%) of adults only have high school education or less. The implications of low educational attainment among adults have been linked to a lack of college guidance and support for children (McDonough, 1997; Teranishi, 2003) and limitations in the likelihood of upward mobility from one generation to the next.

Low educational attainment among certain communities is consistent with a high level of poverty – which is another class indicator that is seldom associated with the Asian American population. For example, according to the US Bureau of the Census (2000), Southeast Asians (e.g., Laotians, Hmong, Lu-Mien, Cambodians, Vietnamese) experience poverty rates that are 2–4 times greater than the national average. In 2000, 38% of Hmongs, 29% of Cambodians, and 19% of Laotians lived in poverty – compared to the national average of 12%.

For particular AAPI ethnic communities, poverty is especially pronounced. A group that has a particularly high prevalence is Southeast Asians – a population with the highest welfare-dependency rates of any ethnic or racial group in the United States. Southeast Asians constitute nearly 80% of all Asian Americans on welfare. There are similar trends in employment rates among Asian Americans. In a study of Southeast Asian high school seniors in California, 49% of fathers and 75% of mothers of Cambodian students were unemployed (Teranishi, 2004).

Among AAPIs, the challenges associated with low social class are not restricted to Southeast Asians. Consider the native Hawai'ian and Pacific Islander population. In the Wai'anae community in Hawai'i – which is the

largest concentration of native Hawaiians in the United States, 79% of the adults have achieved high school education or less and 32% of families with children are below poverty line – a rate that is three times the national average. The high level of poverty often leads to high levels of dependency on social services, including welfare. This wide distribution in economic capital has implications for the distribution of resources that are available to students as they pursue their educational aspirations.

Other areas for which AAPIs are often not recognized are immigration and language research and policy. For Asian Americans – which are the fast-growing racial group in the nation – much of their growth is fueled by immigration. A high degree of immigration implies that, among AAPIs, there is a larger proportion of the population that is foreign-born. Many of the individuals, families, and communities that represent this sector of the population face challenges associated with language. The likelihood that English is spoken at home in some groups is very low. As a result, students often feel they cannot leave home to attend college because their families need them to translate English. For Hmongs, it was emphasized more so than among Vietnamese because the language is so rare and different than the structures of other languages. For example, a written language does not exist for Hmongs.

Patterns can be found in the rate of AAPI foreign-born residents and the likelihood that the same community will speak English with fluency. The Chinese American community in Brooklyn, New York, is a good case in point. This Chinese American community has a high rate of residents who are foreign-born (77%) along with a high rate of children who speak English less than very well (64%). In Westminster – which is one of the largest Vietnamese American communities outside of Vietnam – 67% of the adults speak English less than very well.

While language is a challenge for most first-generation AAPI immigrants, there are important variations that also warrant consideration; among immigrants from Asia, there is great unevenness in their immigrant experiences. Consider the rate of individuals granted access to the US under employment preferences, which varies widely by country of origin. Following the Vietnam War, immigrants, refugees, and asylum seekers from Southeast Asia – including Vietnamese, Hmong, Cambodians, and Laotians – created a vibrant Southeast Asian population in America that is quite significant today. For those populations that arrived in the US from Laos, Cambodia, and Vietnam, a high rate arrived in the US via refugee status (88%, 52%, and 50%, respectively) between 1990 and 2000 (Office of Immigration Statistics, US Department of Homeland Security, 2005). Conversely, during the same time period, a large share of immigrants from Taiwan (37%), China (36%), and Korea (25%) were much more likely to be admitted under employment preferences. Therefore, it is clear that different sectors of the AAPI

population arrive in the United States with a wide distribution of resources and capital, including educational attainment prior to their arrival, economic capital, and other forms of capital that translate well into the United States opportunity structure.

Conclusion

Research and policy on race need to transcend the stereotype types and assumptions with regard to AAPIs that have been used to characterize their experiences and outcomes in education. To be truly inclusive of the population requires thoughtful consideration of the complexity and uniqueness of the population. Asian Americans are a unique racial group with needs, issues, and challenges that vary within the population. Even beyond the ethnic diversity of the AAPI population, there are also a wide range of indicators and outcomes related to educational attainment and poverty status. These demographic distinctions within the population lead to a wide range of opportunities and outcomes that can be found across subpopulations within the larger racial category. In other words, just as some Asian Americans are at the top of the curve, many struggle at the bottom of the curve. The differences among Asian American groups are often greater than their similarities, and, because of this, Asian American data must be disaggregated to have meaning.

See also: Educational Attainment and Mortality Differentials; Educational Attainments of U.S. Black Males and Females: 1971 to 2003; Educational Preparation: Fostering The Self-Efficacy And Resilience Of Urban Adolescent Youth; Locating Space and Place in the College Access Debate: New Tools for Mapping and Understanding Educational Inequity and Stratification.

Bibliography

- Allen, W. (1999). Missing in action: Race, gender, and Black students' educational opportunities. In Carabado, D. W. (ed.) *Black Men on Race, Gender, and Sexuality: A Critical Reader*, pp 194–211. New York: New York University Press.
- Daniels, R. (1997). No lamps were lit for them: Angel island and the historiography of Asian American immigration. *Journal of American Ethnic History* 17, 14.
- Hune, S. and Chan, K. S. (1997). Special focus: Asian Pacific American demographics and educational trends. In Carter, D. and Wilson, R. (eds.) *Minorities in Higher Education*, Vol. 15, pp 39–67. Washington, DC: American Council on Education.
- McDonough, P. M. (1997). *Choosing Colleges: How Social Class and Schools Structure Opportunity*. Albany, NY: State University of New York Press.
- Tamura, E. H. (2001). Asian Americans in the history of education: An historiographical essay. *History of Education Quarterly* 41, 1.
- Teranishi, R. (2003). 'Raced' perspectives on college opportunity: Examining Asian Americans through critical race theory. *Equity and Excellence in Education* 35(2), 144–154.

- Teranishi, R. (2004). Yellow and *Brown*: Residential segregation and emerging Asian American immigrant populations. *Equity and Excellence in Education* 37(3), 255–263.
- Teranishi, R. (2007). Comparative frameworks in higher education research: Implications for the study of race and ethnicity. In Stage, F. (ed.) *Using Quantitative Research to Answer Critical Questions*. New Directions for Institutional Research. San Francisco, CA: Jossey-Bass.
- Thurgood, L., Golloday, M., and Hill, S. (2006). *U.S. Doctorates in the 20th Century*. Washington, DC: National Science Foundation.
- U.S. Census Bureau (2000). Summary File 1. Washington, D.C.: U.S. Census Bureau.
- U.S. Census Bureau (2003). *Annual Estimates of the Population by Race Alone or in Combination and Hispanic or Latino Origin for the United States and States*. United States: Population Division.
- Valencia, R. R. (1997). *The Evolution of Deficit Thinking: Educational Thought and Practice*. London: Falmer.
- Zia, H. (2001). *Asian American Dreams: The Emergence of an American People*. New York, NY: Farrar, Straus and Giroux.

Educational Attainment and Mortality Differentials

H D Horton and L Hixson, University at Albany-SUNY, Albany, NY, USA

S M Monnat, University of Nevada, Las Vegas, Las Vegas, NV, USA; University at Albany-SUNY Albany, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction: Defining Mortality and the Causes of Mortality

Previous research has consistently found extensive and positive associations between education and health. Those with higher levels of education have lower risk of overall mortality, and the gap has increased over time. Defined by the number of deaths in a population, mortality is caused by a number of different factors, including: chronic diseases, HIV/AIDS, suicide, accidental death, infant and childhood conditions, and poor overall health quality.

This article summarizes the literature on education and mortality and assesses the impacts of social and economic stratification on mortality differentials in the United States and internationally. Although findings from various disciplines are included in this entry, this article is couched in a sociological perspective and focuses heavily on the social structural conditions that contribute to the link between education and mortality. Multiple aspects of stratification contributing to the link between education and mortality are discussed, including: race/ethnicity, gender, socioeconomic status, marital status, citizenship, and neighborhood context. The article concludes with a discussion of policy implications.

Chronic Diseases

Historically, infectious diseases were the major cause of mortality around the world. Malaria, the plague, and smallpox killed hundreds of thousands of people internationally. Today, the top four causes of mortality in the United States are degenerative diseases rather than infectious diseases. The major chronic diseases associated with degeneration – the biological deterioration of a body – are heart disease, malignant neoplasm (cancer), cerebrovascular disease (stroke), and chronic lower respiratory disease.

HIV/AIDS

Rates of HIV/AIDS are spreading rapidly around the world, especially in sub-Saharan Africa, Eastern Europe, and Asia. Yet, deaths from HIV/AIDS have slowed in North and South America and Western Europe. Rates continue to increase because the vast majority of people do not know they have the disease or do not acknowledge the infection. Education is directly linked to HIV/AIDS

as the disease is more prevalent among those with lower levels of education.

Suicide

Approximately 1 million people each year commit suicide and as many as 20 times that number attempt suicide but are not successful. Europe, especially Eastern Europe, has the highest rates of suicide; the United States and Canada have average rates; and Mexico is well below the world average. Suicide is caused by depression, hopelessness, and feelings of powerlessness, all of which are increased in individuals with lower educational attainment.

Infant and Childhood Mortality

Although the United States is among the wealthiest countries in the world, it continues to have infant mortality rates greater than any other industrialized nation. While the rates for the United States are extremely high, not all groups experience similar risk of infant mortality. Some communities experience very high infant mortality rates, while others experience much more modest rates. The reasons for these disparities are complex but have generally been thought to reflect the enduring social and economic inequality in the United States. For example, while infant mortality rates in the United States have consistently declined since the 1950s (largely as a result of technological innovation), the gap between whites and African-Americans has remained relatively stable since the early 1970s. Infant mortality rates are highest among those with the lowest levels of education.

Trends

Life expectancy in the United States has increased since the 1980s. Yet, life expectancy increases have been almost exclusively restricted to those with high levels of education. The mortality gap between those with low versus high levels of education in the United States is increasing. In 2000, the life expectancy for a 25-year-old with a high school diploma was 50 years, but for a person of the same age with some college education, the life expectancy was almost 57 years. What is more, the data show that the widening disparities in mortality are not explained by increased race or sex differences in mortality trends.

Socioeconomic differences in rates of mortality over the life course are greatest among those who are middle-aged. These differences are primarily a function of income at the low end of the socioeconomic continuum and education at the high end. In addition, inequality trends are more adverse for older persons than for younger

Table 1 Life expectancy (in years) at age 25, by race, sex, and educational status (1990 and 2000)

	1990	2000	Change
White men	49.6	51.0	1.4 ***
Low education	47.0	47.1	0.1 ***
High education	53.2	55.0	1.8 ***
White women	55.6	55.8	0.2 ***
Low education	54.3	53.4	-0.9 ***
High education	57.8	58.8	1.0 ***
Black men	42.4	44.9	2.5 ***
Low education	40.2	42.3	2.1 ***
High education	47.3	50.6	3.3 ***
Black women	50.6	51.2	0.6 ***
Low education	49.4	49.2	-0.2 *
High education	52.9	54.6	1.7 ***

Data come from the Multiple Cause of Death (MCD) files and the Integrated Public Use Micro Sample (IPUMS) of 2000 US Census.

* $p < 0.10$, *** $p < 0.001$

Reproduced from Meara, E.R., Richards, S., and Cutter, D.M. (2008). The gap gets bigger: Changes in mortality and life expectancy, by education, 1981–2000. *Health Affairs* 27, 350–360.

ones. The cumulative effect of education on mortality over the life span means that there is greater disparity in risk of premature mortality between older individuals with divergent levels of education. A comparison of the magnitude of educational mortality differentials in the United States and Europe, where healthcare is more equally distributed, shows that at middle age, the proportionate reductions in mortality for each additional year of schooling are similar. This suggests that healthcare access may not be as important a determinant of mortality as previously thought.

Not only does the effect of education on mortality vary by age, but also by gender. Research finds that educational differentials in mortality have increased for men over the past several years but not for women.

Table 1 displays life expectancy by race, gender, and educational status in the United States. While life expectancy increased from 1990 to 2000 among all groups with a high level of education, it actually decreased over the 10-year period for white and black women with low levels of education. Further, the life expectancy gap between those with low education and those with high education increased for each of the four race–gender groups from 1990 to 2000, indicating that educational attainment is becoming more important in reducing one's risk of mortality.

Figure 1 shows that certain conditions have contributed more to the increased gap in educational disparities in mortality than other conditions. Heart diseases and

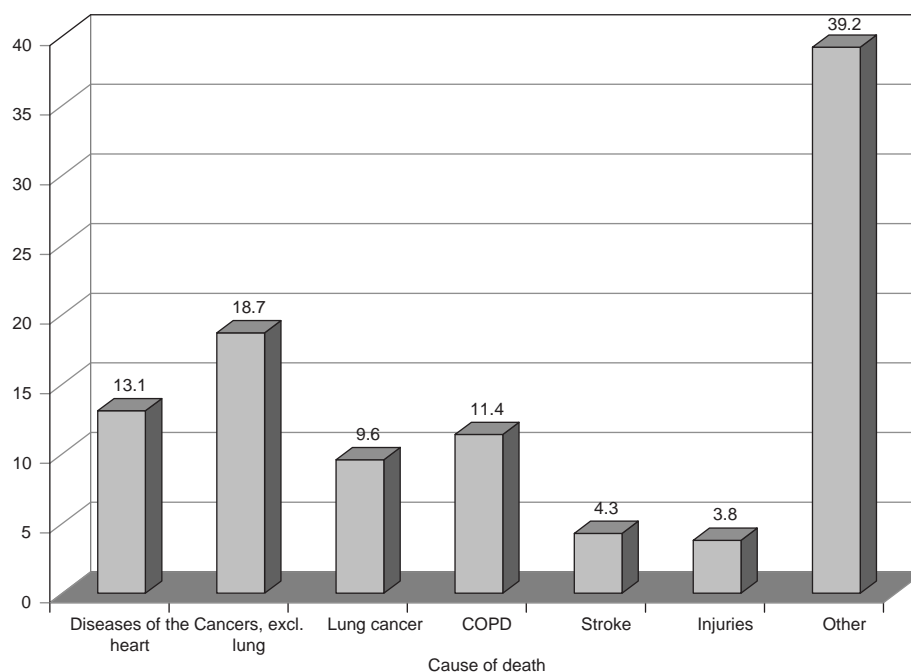


Figure 1 Contribution to growth in education-related disparities, ages 25–84 (1990–2000). Authors' calculations were based on data from non-Hispanic blacks and whites in the Multiple Causes of Death (MCD) files linked to census data. COPD, chronic obstructive pulmonary disease. Reproduced from Meara, E. R., Richards, S., and Cutler, D. M. (2008). The gap gets bigger: Changes in mortality and life expectancy, by education, 1981–2000. *Health Affairs* 27, 350–360.

cancers contributed most to the rising mortality disparities between those with low and high educational attainment from 1990–2000. Lung cancer and chronic obstructive pulmonary disease (COPD) largely attributable to smoking, accounted for 21% of the increase in mortality between those with high and low education.

Explanations for the Association between Education and Mortality

Social Structural Conditions

The sociological perspective emphasizes that social structural conditions affect the association between education and mortality. Two important social structural conditions affecting mortality are inequality and stratification and access to healthcare resources.

Inequality and stratification

Inequality refers to disparities in the distribution of income, wealth, social, economic, and political power in society. Stratification is the hierarchical arrangement of social classes within a society. Inequality is created and maintained through social structural forces that privilege certain groups of people and disadvantage others. Increasing social inequality is manifested, in part, by increasing geographic isolation. Community conditions affect individual mortality risk. Residential segregation, for example, reinforces social isolation and economic segregation. People in poor and disadvantaged places have significantly higher mortality rates. Residential segregation and political disenfranchisement are, in part, responsible for elevated infant mortality rates in US cities. Furthermore, decay in US housing stock, undesirable land use, and marginalization of places through withdrawal of municipal services has led to higher rates of homicide, suicide, and deviant behavior, all of which increase mortality risk.

Research demonstrates that the social and economic context of the neighborhood affects mortality risk net of individual characteristics. This pattern is especially salient in highly segregated neighborhoods. For example, net of individual socioeconomic characteristics and neighborhood characteristics are accountable for risk of death experienced by blacks. Further, this risk is not specific to blacks but is shared by anyone who lives in highly segregated black communities. Residential segregation perpetuates racial disparities in poverty, education, and economic opportunities that in turn lead to differences in health outcomes. Residents of economically deprived neighborhoods have higher body mass indices, larger waist circumferences, higher waist–hip ratios, and greater risk of diabetes than their peers in more economically advantaged residential areas. The higher cost of food in segregated neighborhoods leads to consumption of poor-quality food items which affect residents' overall health quality.

Working conditions

Work and economic conditions also directly and indirectly contribute to the link between education and mortality. The likelihood of being employed and the type of work an individual does is influenced by educational level. First, the well educated are more likely to be employed. Lack of education limits employment opportunities. The poorly educated are more likely to work only part time, and are at greatest risk of losing their jobs in times of economic downturn. Employment affects an individual's overall well-being; being unemployed is associated with poor health.

Second, low educational attainment translates into lower expected income. Poorly educated people experience more economic hardship than those with more education. Economic hardship and poverty negatively affect health. Ongoing strain and stress related to not being able to pay the bills, depression, and hopelessness all decrease an individual's resistance to certain diseases and makes them more susceptible to untimely death.

Third, those with lower levels of education work in occupations that put one at greater risk of mortality. Fatal injuries by industry are most represented in occupations entered by individuals with little or no education, including construction, transportation and warehousing, agriculture, forestry, fishing, hunting, and mining. Further, nearly half of all work-related fatalities in the United States involve vehicles. Occupations that require workers to travel include logging, refuse and recyclable material collection, farming and ranching, electric power-line installing, roofing, traveling sales, and truck-driving. Most of these high-risk occupations are entered by those with lower levels of education.

Finally, well-educated people are employed in occupations that allow them greater autonomy and less monotony. Increased independence on the job and nonroutine work increase both psychological functioning and job satisfaction. Work that is characterized by job security, variation in tasks, and inclusion in decision making is more subjectively rewarding, leading to better mental and physical health.

Health lifestyles

There are a number of ways in which a person's health lifestyle can play a role in risk of mortality. Compared to the poorly educated, those with higher levels of education are less likely to engage in lifestyle behaviors that would increase risk of untimely death, including smoking, drinking, and risky behaviors. Further, individuals with higher levels of education are more likely to engage in healthy lifestyle behaviors including exercise and health checkups. There is an interaction between lifestyle and educational level such that behavior posing a threat to good health has less impact among those with higher levels of education.

First, smoking has been linked to chronic diseases that often lead to death, including lung and other cancers,

coronary heart disease, stroke, atherosclerosis, aneurysms, emphysema, and chronic bronchitis. **Table 2** displays the percentage of smokers in the US population by gender and educational status. Statistics show that as one's educational level increases, the likelihood of smoking decreases. Prevalence of smoking is highest for those with a GED diploma and lowest for those with a graduate degree. This pattern holds for both men and women.

Second, those with lower levels of education are more likely to drink heavily, often as a strategy to cope with the stresses of poverty. While excess drinking is less ubiquitous in its health consequences than other unhealthy

behaviors, it is linked to death caused by cirrhosis of the liver, suicide, homicide, and car accidents. Conversely, those with high levels of education are more likely to engage in moderate drinking, which has been linked to positive health outcomes, including lower risk of coronary heart disease, stroke, and hypertension.

Education also affects risk of preventable death. The risk of dying from either a highly preventable or less preventable death decreases as years of education increases. **Figure 2** displays hazard scores representing risk of dying by either a highly preventable or low preventable cause in the United States. Preventable deaths are those wherein death can be prevented by means of medical treatment or other interventions and/or by preventing the incidence of the disease to start with through good hygiene, diet, vaccines, and other lifestyle factors. Among individuals in the United States aged 25–44, those with the fewest years of education have the greatest risk of highly preventable death.

Finally, those with higher levels of education are more likely to engage in healthy lifestyle behaviors, such as exercise and regular health checkups. Physical activity reduces mortality related to cardiovascular disease, atherosclerosis, colon cancer, obesity, high blood pressure, and diabetes. The well educated are also more likely to get preventive healthcare, including annual exams, immunizations, and screenings. This type of preventive healthcare helps detect early signs of illness, reducing the risk of premature mortality.

Table 2 Estimated percentage of persons aged 18 or older who are current smokers^a by Gender and Educational Level (2006)

	Men	Women	Total
0–12 (no diploma)	30.6	23.0	26.7
GED diploma	51.3	40.2	46.0
High school diploma	27.6	20.4	23.8
Some college	26.1	20.0	22.7
Associate degree	25.4	17.8	21.2
Four year degree	10.8	8.4	9.6
Graduate degree	7.3	5.8	6.6

^aPersons who reported smoking at least 100 cigarettes during their lifetime and who, at the time of the interview, reported smoking every day or some days.

Date retrieved from the Centers for Disease Control Morbidity and Mortality Weekly Report Vol. 56. No. 44 (November 9, 2007).

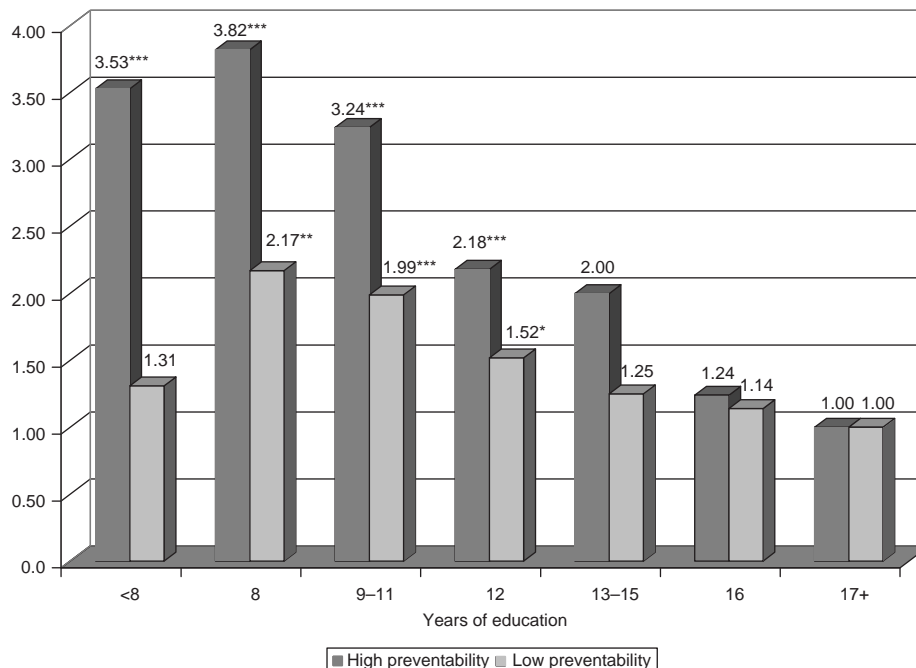


Figure 2 Cox proportional hazard scores predicting death by educational attainment for ages 25–44, US National Longitudinal Mortality Study (1995). Reproduced from Phelan, J. C., Link, B. G., Diez-Roux, A., Kawachi, I., and Levin, B. (2004). 'Fundamental causes' of social inequalities in mortality: A test of the theory. *Journal of Health and Social Behavior* 45, 265–285.

Social Psychological Resources

In addition to social structural conditions and individuals' own health lifestyles, social psychological resources, including a sense of personal control and social supports are linked to improved health and lower risk of mortality, while other social psychological conditions, such as perceived stigma and discrimination may lead to increased risk of mortality.

Sense of control

Education affects the sense of self-control one has over their own life. Perceived powerlessness and lack of control are linked to one's socioeconomic position within society. Those with more years of education are at higher socioeconomic levels and experience a greater sense of control, mastery, and ability to effectively alter their life chances. Education helps one develop self-direction, communication skills, analytical ability, and problem-solving skills. Those with lower levels of education and poor economic circumstances often learn to accept failure and perceive that others have control over their lives. This lack of personal control has deleterious affects on one's health, since those with less perceived control are less knowledgeable about their health, less likely to initiate preventive behaviors, and more likely to experience emotional stress leading to worse immune systems.

Social support

The well educated also have higher levels of social support than the poorly educated. Economic hardship, linked to low levels of education, decrease the number and quality of social support. Social support increases a person's sense of being loved, valued as a person, and being part of a community or social network. Social support decreases mortality through psychological and behavioral mechanisms. Psychological distress increases likelihood of premature death through depression, smoking, drug use, and heavy drinking.

Discrimination

Perceived or actual discrimination may increase one's level of stress, thereby negatively affecting health outcomes. Those with lower levels of education, particularly if they are also non white, experience greater perceived and actual stigma and discrimination by employers and co-workers. Both the economic and non-economic forms of discrimination relate to each other and combine with one's socioeconomic position and other risk factors to affect health. Socioeconomic inequalities in health are produced, in part, by differential exposure to stress and negative life events. Longitudinal survey data reveal that negative lifetime events are positively associated with mortality. Perceived discrimination – both daily exposure and lifetime exposure – is associated with poor mental health, putting one at greater risk of untimely mortality.

Demographic Characteristics Affecting Mortality

Several demographic characteristics interact with social structural conditions, health lifestyles, and social psychological resources to mediate the link between education and mortality. Such characteristics include race/ethnicity, gender, socioeconomic status, marital status, and context of residence.

Race/Ethnicity

Race/ethnicity confounds with education to affect mortality risk in several ways. Of all racial groups in the United States, blacks live the fewest years and spend a high proportion of those years with a chronic health problem. Socioeconomic conditions, not health behaviors, are the primary cause of racial health stratification. Poverty, residential segregation, racism, the lack of adequate social and economic resources, and cultural beliefs about health are some of the contributors of the observed racial differences in mortality risk.

Studies have demonstrated that in the United States, blacks are more likely than whites to die of cancer and heart disease, more likely to get diabetes and asthma, and less likely to seek preventive care. They are 3 times more likely to be hospitalized and 4 times more likely to die as a result of asthma than are white patients.

Early-life socioeconomic conditions, such as parental occupation and family structure, explain part of the race gap in mortality. While black men's higher rates of death are associated with lower socioeconomic standing in early life, these effects operate indirectly through adult socioeconomic achievement processes. Research suggests that education indirectly accounts for the racial gap in black men's mortality since lower levels of education lead to lower family income, wealth, and occupational standing.

Race and ethnicity also affect infant mortality. **Table 3** shows that in general, infant mortality rates are lowest for infants born to Chinese and Japanese mothers and highest for infants born to non-Hispanic black and Puerto Rican mothers. These same patterns hold even when controlling for educational attainment. Infant mortality rates generally decrease with increasing educational levels among all racial groups. Rates are highest for infants born to mothers with only 9–11 years of education and lowest for infants born to mothers with 16 years or more of education.

Gender

Gender also combines with education in several ways to affect risk of mortality. Social structural conditions, health lifestyle, and social psychological resources all play a role in the link between gender, education, and mortality.

One of the most striking recent trends in mortality has been the divergence of male and female death rates.

Table 3 Infant mortality rates^a by race and mother's education at attainment (1999)

	<i>White</i>	<i>Black</i>	<i>American</i>	<i>Asian/Pacific</i>		<i>All races</i>
			<i>Indian</i>	<i>Islander</i>	<i>Hispanic^b</i>	
0–8 years	6.8	16.5	13.9	6.7	5.7	7.5
9–11 years	8.0	15.1	11.0	5.4	6.1	9.5
12 years	6.1	13.9	8.5	5.6	5.4	7.6
13–15 years	4.8	12.0	8.5	4.4	5.1	5.9
16 years and over	3.9	10.2	*	3.8	4.1	4.4

^aInfant mortality rates per 1000 live births in specified group.

^bPersons of Hispanic origin may be of any race.

Data retrieved from the Centers for Disease Control. National Vital Statistics Reports Vol. 50 No. 4 (January 30, 2002).

There is evidence that biological differences between the sexes lead to lower female mortality. Various environmental factors also interact with gender to create differential mortality risk for men and women. Relative mortality differentials by educational attainment are greater for men than for women. Differences are more apparent at prime working ages than at older ages, with relative differentials narrowing as age advances. As socioeconomic inequalities have increased in recent years in the United States, these trends have been more adverse for men than women. Alternatively, some countries actually have higher life expectancy for men (i.e., Botswana, Kenya, Zambia, and Zimbabwe). Research suggests that this difference is due to cultural and socioeconomic factors, such as nutritional practices that discriminate against female children, food allocations that favor men, and higher quality medical treatment for men.

Evidence suggests that men are also more likely to engage in dangerous lifestyle behaviors, such as excess drinking, risky driving that may lead to accidents, drug abuse, and poor eating habits. Accidental deaths are among the top five causes of mortality in the United States, Canada, and Mexico. About 50% of all deaths in the United States are attributable to motor vehicles. These victims are disproportionately young males. Those with lower levels of education are most likely to engage in risky behaviors, including substance abuse, dangerous driving, and unhealthy eating. Thus, gender and educational level interact to increase risk of mortality for men with lower levels of education.

Finally, there is a well-documented literature examining men's reluctance to seek out preventive medical treatment. Largely due to their socialization, men are far less likely than women to get regular medical exams or go to the doctor when they are sick, increasing their risk of premature mortality.

Socioeconomic Status

There is a direct link between education and income. Those with greater educational attainment earn more income over their life. Studies show that higher income

provides increased access to healthcare, proper diet, quality housing, consistent employment, and a healthy lifestyle. Individuals with both high income and high education have lower mortality on average than individuals with high education and lower income or those with high income but low education. Both education and income have important independent associations with mortality. This association is not limited to the United States. In Great Britain, mortality differentials according to socioeconomic circumstances increased in tandem with increases in income inequality. Recent research has found that almost 25% of deaths in Great Britain are attributable to unfavorable socioeconomic circumstances.

Marital Status and Family Conditions

Marital status and childhood family conditions also combine with education to affect mortality. Education has a comparable effect on mortality for both married men and women, but the negative effect of being single and having a low education level has increased over time.

Research also suggests that there is a link between spouse's educational level and mortality, particularly for women with higher levels of education. While women with low levels of education are not affected by their husband's educational level, highly educated women have an almost twofold increase in risk of cardiovascular disease when married to less educated husbands.

Early childhood family conditions also impact mortality. Living conditions in the parental home contribute to social class differences in mortality. Parental home characteristics, such as social class, number of siblings, parents' marital status, and employment status, all affect risk of various types of mortality. Research has found, however, that much of the effect of early childhood family conditions is mediated by an individual's educational path.

The Role of Context on Mortality

Research has also been conducted on how context of residence impacts mortality. These studies, largely

conducted under the ecological perspective, have focused on the relationship between community socioeconomic status characteristics and overall mortality rates. In some cases, high mortality in a community may reflect its population composition rather than the characteristics of the place, while in other cases, high mortality risk may reflect the quality of the neighborhood and exposure to accidents, violence, and environmental hazards. Where one lives also affects one's access to quality education and the subsequent life opportunities that come from that education. Context of residence interacts with education to impact mortality in two ways: access to health-care services and environmental hazards and social disorder.

Individuals living in rural areas in the United States have historically had less access to a wide array of health-care services due to lower income, lack of private health insurance, and shortage of healthcare providers. Yet, despite these healthcare disadvantages, rural residents in the United States experience better health, including longer length of life. **Figure 3** presents life expectancy for the typical married white male at the age of 45 who is in good health for three different socioeconomic levels in urban, suburban, and rural areas in the United States. Among those living in both urban and suburban areas, those with high socioeconomic status (higher income and more years of education) have longer life expectancies. However, there

is no significant difference in life expectancy by socioeconomic status among those living in rural areas. This suggests a greater level of health parity among socioeconomic groups in rural areas in the United States.

That same equality is not found in highly segregated urban areas in the United States. Healthcare access for blacks living in residentially segregated urban areas is limited. While geographic access to physicians has continued to improve over the past two decades, blacks disproportionately lack regular sites of care and experience less continuity of healthcare. Physicians are significantly less likely to participate in Medicaid in areas where poor people are nonwhite and those that are racially segregated. Residential segregation also reduces access to physicians and hospitals for blacks and poor people and they must travel much farther on average than white patients. Patients visit hospitals more frequently when the distance traveled is minimized and the rate of use of hospitals declines with distance.

Policy Implications

In their landmark study on mortality, Kitagawa and Hauser (1977) state that, "mortality is presumed to have little relevance to policy because health and longevity are near-universal human values, and governments are expected to spend all that they can to promote the health and longevity

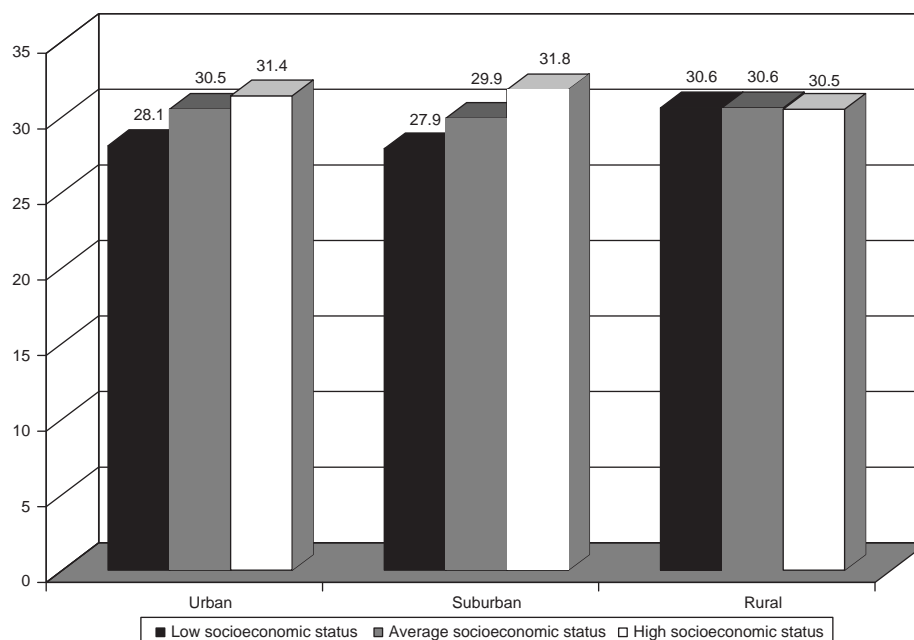


Figure 3 Life expectancy at age 45 by socioeconomic status and urban, suburban, and rural residence in the United States. The controlled characteristics for the life expectancies are: white, married, in good health, and currently retired. Low socioeconomic status men are defined as having less than 9 years of education and a family income of \$14 177 (in 1983 dollars); average socioeconomic status men are defined as having 9–12 years of education and a family income of \$24 567 (in 1983 dollars); and high socioeconomic status men are defined as having some college education and a family income of \$35 416 (in 1983 dollars). Reproduced from Hayward, M. D., Pienta, A. M., and McLaughlin, D. K. (1997). Inequality in men's mortality: The socioeconomic status gradient and geographic context. *Journal of Health and Social Behavior* 38, 313–330.

of their citizens” (p. 381). A public discussion about educational differentials in mortality is important because it points to the possibility of reducing mortality through improved socioeconomic conditions.

Some causes of death are more preventable than others and are therefore more amenable to policy interventions. The Healthy People 2010 initiative aims to entirely eliminate all health disparities in the United States by the end of this decade. Several interventions have been designed to reduce disparities in health and mortality, including the Medicare disproportionate-share hospital program, the State Children’s Health Insurance Program, and Medicaid expansions. Efforts to improve public health and disease prevention have intensified over the past several years, with particular focus on risk factors that are prevalent among those with lower levels of education: smoking, obesity, and under-use of prevention services. The very large socioeconomic differentials in mortality in the United States suggest that the most important next step in mortality reduction in the United States and internationally might be to improve social and economic conditions of disadvantaged groups rather than further advances in medical knowledge.

Further Reading

- Backlund, E., Sorlie, P. D., and Johnson, N. J. (1999). A comparison of the relationships of education and income with mortality: The national longitudinal mortality study. *Social Science and Medicine* **49**, 1373–1384.
- Blackwell, D. L., Hayward, M. D., and Crimmins, E. M. (2001). Does childhood health affect chronic morbidity in later life? *Social Science and Medicine* **52**, 1269–1284.
- Centers for Disease Control (2007). Cigarette smoking among adults – United States, 2006. *Morbidity and Mortality Weekly Report* **56**, 1157–1161.
- Collins, C. and Williams, D. R. (2001). Racial residential segregation: A fundamental cause of racial disparities in health. *Public Health Reports* **116**, 404–416.
- Elo, I. T. and Drevenstedt, G. L. (2002). Educational differences in cause-specific mortality in the United States. *Yearbook of Population Research in Finland* **38**, 37–54.
- Elo, I. T. and Preston, S. H. (1996). Educational differentials in mortality: United States, 1979–85. *Social Science and Medicine* **42**, 47–57.
- Gortmaker, S. L. and Wise, P. H. (1997). The first injustice: Socioeconomic disparities, health services technology, and infant mortality. *Annual Review of Sociology* **23**, 147–170.
- Hayward, M. D., Pienta, A. M., and McLaughlin, D. K. (1997). Inequality in men’s mortality: The socioeconomic status gradient and geographic context. *Journal of Health and Social Behavior* **38**, 313–330.
- Jaffe, D. H., Eisenbach, Z., Neumark, Y. D., and Manor, O. (2006). Effects of husbands’ and wives education on each other’s mortality. *Social Science and Medicine* **62**, 2014–2023.
- Kitagawa, E. M. and Hauser, P. M. (1973). *Differential Mortality in the US: A Study in Socioeconomic Epidemiology*. Cambridge, MA: Harvard University Press.
- Landale, N., Oropesa, R. S., and Gorman, B. K. (2000). Migration and infant death: Assimilation or selective migration among Puerto Ricans? *American Sociological Review* **65**, 888–909.
- LeClere, F. B., Rogers, R. G., and Peters, K. D. (1997). Ethnicity and mortality in the United States: Individual and community correlates. *Social Forces* **76**, 169–198.
- Magadi, M., Diamond, I., and Madise, N. (2001). Analysis of factors associated with maternal mortality in Kenyan hospitals. *Journal of Biosocial Science* **33**, 375–389.
- Matthews, T. J., MacDorman, M. F., and Menacker, F. (2002). Infant mortality statistics from the 1999 period linked birth/infant death data set. *National Vital Statistics Reports* **50**, 1–9.
- Meara, E. R., Richards, S., and Cutler, D. M. (2008). The gap gets bigger: Changes in mortality and life expectancy, by education, 1981–2000. *Health Affairs* **27**, 350–360.
- Pensola, T. and Martikainen, P. (2004). Life-course experiences and mortality by adult social class among young men. *Social Science and Medicine* **58**, 2149–2170.
- Phelan, J. C., Link, B. G., Diez-Roux, A., Kawachi, I., and Levin, B. (2004). ‘Fundamental causes’ of social inequalities in mortality: A test of the theory. *Journal of Health and Social Behavior* **45**, 265–285.
- Preston, S. H. and Elo, I. T. (1995). Are educational differentials in adult mortality increasing in the United States? *Journal of Aging and Health* **7**, 476–496.
- Rogers, R. G., Hummer, R. A., and Nam, C. B. (2000). *Living and Dying in the USA: Behavioral, Health, and Social Differentials of Adult Mortality*. San Diego, CA: Academic Press.
- Ross, C. E. and Wu, C. (1995). The links between education and health. *American Sociological Review* **60**, 719–745.
- Shaw, M., Gordon, D., Dorling, D., Mitchell, R., and Smith, G. D. (2000). Increasing mortality differentials by residential area level of poverty: Britain 1981–1997. *Social Science and Medicine* **51**, 151–153.
- US Department of Labor, Bureau of Labor Statistics (2006). *Census of Fatal Occupational Injuries*. <http://www.bls.gov/news.release/pdf/cfoi.pdf> (accessed September 2009).
- Warner, D. F. and Hayward, M. D. (2006). Early-life origins of the race gap in men’s mortality. *Journal of Health and Social Behavior* **47**, 209–226.
- Weeks, J. R. (2005). *Population: An Introduction to Concept and Issues*. Belmont, CA: Wadsworth/Thompson Learning.
- Williams, D. R. and Collins, C. (1995). U.S. socioeconomic and racial differences in health: Patterns and explanations. *American Review of Sociology* **21**, 349–386.

Educational Opportunity and Latino/Chicano College Choice: New Findings, and Theoretical Perspectives

M A Ceja, California State University, Northridge, Northridge, CA, USA

P A Pérez, California State University, Fullerton, Fullerton, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Increasing educational attainment and achievement among Latino students continues to be an important policy concern. Latinos, defined as individuals of Mexican, Central and/or South American descent, make up the largest racial/ethnic minority group in the United States and are one of the fastest growing groups, second only to Asians/Pacific Islanders (National Center for Education Statistics (NCES), 2007). Despite the growth experienced by the Latino population in the United States, accessing equitable educational opportunities and achieving educational success has proven to be a substantial challenge. For example, national data show a 60% high school completion rate among Latino students (Harvey, 2002). When compared to all other youth aged 16–19, Latinos are more likely to drop out from high school (Fry, 2003). These low rates illustrate an educational crisis for the Latino population at all levels of the educational system.

The direct impact of these trends on adult Latinos is clear, given that approximately 41.5% of Latinos aged 25 and over have attained less than a high-school diploma (NCES, 2007). Latinos also lag behind every other student population group in attaining college degrees (Fry, 2003). When Latino students do transition into postsecondary education, they are less likely to enroll in a 4-year institution immediately after graduation and more likely to be concentrated in the community college system (Kurlaender, 2006). These trends are particularly disturbing given recent findings suggesting that in California only 18% of community college Latino students with intentions to complete a program or to transfer into a 4-year institution managed to do so within 6 years (Shulock and Moore, 2007).

Understanding how and with what resources Latinos form and complete postsecondary plans will assist in enhancing their educational opportunities. Indeed, the college-choice decision process that students make impacts both the individual student and, collectively, society as a whole. That is, evidence proposes a more educated workforce provides better economic returns for our nation as well as a multitude of social benefits to the individual (Pascarella and Terenzini, 1991). Research suggests, however, not all individuals have access to the same resources to make college a reality and make an informed college-choice decision (McDonough *et al.*, 2003, 2004). Research highlights Latinos,

as well as other minority groups, experience the college-choice process differently than their white counterparts and that traditional models of college choice are not sufficient to explain their particular processes (Ceja, 2001; Kim, 2004; Perna, 2000; Talavera-Bustillos, 1998). To this end, we examine the theoretical and empirical research on educational opportunity for Latino students. In addition, this article explores how factors such as race, socioeconomic status, college generational status, and gender, influence access to resources and the college-selection decisions Latino students make. Finally, we propose a new model for understanding the college-choice process of Latino students.

Latino Educational Access and Opportunity

Across the educational pipeline, Latinos experience learning opportunities that are far from equitable (Orfield and Lee, 2007). It is well documented that Latinos start school less prepared than their white counterparts, giving way to an achievement gap that has Latinos and other ethnic/racial minorities succeeding at a lower rate than their white peers (Garcia and Gonzalez, 2006; NCES, 2007). Contributing to this is the fact that Latinos are the least likely of all major ethnic/racial groups to enroll in pre-kindergarten (pre-K) programs (NCES, 2007). This is important given research findings showing positive gains in school readiness for students who attend pre-K programs (Barnett *et al.*, 2005); and for Latino children in particular, research has shown substantial improvements in prereading, prewriting, and premath skills among those who attend (Gormley, 2007).

In addition, because almost 25% of all persons living in poverty are Latino (US Census, 2001), they are more likely to be ethnically segregated in neighborhoods with predominantly minority schools (Orfield, 1992). In 2005, an estimated 77% of Latinos attended schools that were predominantly minority, compared to only 12% of white students attending similar types of schools (NCES, 2007). In their work on school resegregation, Orfield and Lee (2007) lay out the relationship between the low socioeconomic conditions of minorities and the types of schools they attend. Orfield and Lee (2007) write:

Poverty has long been one of the central problems facing segregated schools. Segregation tends to be multidimensional. Few highly segregated minority schools have middle class student bodies. Typically students face double segregation by race/ethnicity and by poverty (p. 18).

Since Latinos are concentrated in resource-poor schools with less competitive instruction, they are more likely to have underprepared and fewer credentialed teachers, higher teacher turnover, less access to honors and advanced placement courses, be in classes with large teacher-student ratios, and more likely to enroll in schools that receive fewer per-pupil dollars than nonminority schools (Yosso and Solórzano, 2006). Oakes (1990) argues that in predominantly minority schools, ability grouping is more common. Oakes highlights ability grouping as increasing academic underachievement.

Furthermore, evidence suggests quality college academic preparatory programs are less likely to be found in predominantly minority schools versus majority schools (Solórzano and Tejeda, 1998). In predominantly Latino high schools, for example, curricula tend to focus on remediation rather than academic enrichment. As several scholars document, inadequate academic preparation is a major caveat to attending a postsecondary institution (Cabrera and La Nasa, 2001; McDonough, 2004; Perna, 2005). In particular, Perna (2005) identifies academic preparation as essential "because the groups of students who continue to be underrepresented in higher education are also the groups that [were] the least likely to be academically prepared" (p. 114). Consistent with Orfield's (1992) research, Perna (2005) argues disadvantaged groups are less likely to be academically prepared, attend schools with less-rigorous programs, and are less likely to be placed in challenging classes when available.

Latino academic underpreparation and tracking into non-college preparatory courses effectively steer them into non-competitive and less-selective higher education institutions (Gutiérrez *et al.*, 2000). McDonough (2004) reinforces that the role of counselors within this tracking dynamic is crucial considering, "counseling [is] often tied to the track placement of students. Those students who [are] not in the college track do not receive college information" (p. 10). The importance of receiving college information cannot be understated. For example, focusing on financial aid information, in a national survey study conducted by the Tomás Rivera Policy Institute (TRPI), results indicate 75% of young Latinos (aged 18–24) not currently enrolled in college would have been more likely to attend college if exposed to better financial aid information (TRPI, 2004). Consistent with these findings are those of Santiago and Cunningham (2005), suggesting that Latinos receive less financial aid for college.

These unbalanced educational environments have produced academic outcomes for Latino students that are substantially lower than their white student peers

(Yosso, 2005). Indeed, the most recent report by the NCES (2007) reveals that throughout K-12 (4th, 8th, and 12th grade), the percentage of Latinos performing below basic in reading and math is among the highest in the country. In fact, by the time Latinos complete high school, only 34% would have taken some type of advanced math course compared to the national average of 50% for all students. The report also notes that only 12% of Latinos, compared to 64% of white students, take advanced placement courses and of those who take the exam, only 47% achieve a passing score or three or higher. Furthermore, recent national data from the College Board shows that only 11% of Latinos took the Scholastic Aptitude Test (SAT), a requirement for admission at many 4-year colleges and universities (The College Board, 2007). Consequently, by the time Latinos move through the K-12 educational system, less than half will be academically prepared for college entry (NCES, 2006).

College Choice

College-choice research suggests that most students who plan to apply to college normally rely on high-school counselors for college guidance, with more affluent students experiencing greater access to such school resources (Bemark and Chung, 2005; Blumberg *et al.*, 2004; McDonough, 1997). In addition, students who have college-educated parents are more likely to participate in higher education and develop a college roadmap, conavigated by family who possess the expertise to aid successfully in their college preparation (Gándara, 2002; McDonough *et al.*, 2000). For these students, the home setting becomes a critical source of college information that children can access during the college-choice process (McDonough, 1997). These students are also provided with a large and sophisticated network of resources, strategic academic options, access to costly college test-taking courses, and perhaps equally important, parental testimonies of successful college experiences (Hossler *et al.*, 1999). This privileged wealth of information helps shape students' understanding of their role in higher education, perceiving college to be a natural progression and a rite of passage afforded to them by their home and schooling experiences (Bateman and Hossler, 1996; Yang, 1981).

Efforts to understand how these factors impact students' college choices have led various researchers to develop models that attempt to explain the stages in students' college decision-making process (Chapman, 1981; Litten, 1982; Hossler *et al.*, 1989). The prevailing model among these researchers identifies three general stages of college choice; predisposition, search, and choice. According to Hossler *et al.* (1989), the predisposition phase can begin very early when educational aspirations are first

developed, which can then lead to a potential search stage. Hossler *et al.* (1989) estimate the search stage runs from approximately the 11th grade through the beginning of the 12th grade. Finally, the choice phase, when the student selects which colleges to apply to and a final postsecondary institution to enroll in, typically takes place throughout the senior year (Hossler *et al.*, 1989). These models, as illustrated in **Figure 1**, have been helpful in allowing us to think about the decision to attend college as a “complex, multistage process during which an individual develops aspirations to continue formal education beyond high school, followed later by a decision to attend a specific college, university or institution of advanced vocational training” (Hossler *et al.*, 1989: 234).

The review of the literature suggests a number of important influences on the college-choice process for Latinos. The early stages of the college-choice process are confounded by several of the background characteristics common to many Latinos, including being from a lower income strata, first generation, older, married, and having more dependents. In addition, Latinos are more likely to postpone college enrollment and attend part time, and are less likely to maintain continuous enrollment once in college. In the predisposition phase, parents play a critical role in encouraging high academic performance and higher education through the development of what Gándara (1995) calls a “culture of possibility” (p. 112). This culture is transmitted from parents through stories and modeling of a hard work ethic. In fact, much of the college-choice literature cites parents as the most important reason for Chicanas, the largest subgroup within the Latino population, wanting to pursue a higher education (Ceja, 2001; Talavera-Bustillos, 1998). It is important to note, parental support in this case is emotional or monetary and does not come in the form of college knowledge. In addition to parents, siblings, teachers, counselors, outreach officers, peers, and caring adults, are known to play important roles for Chicanas with regard to their college-choice processes. These roles range from the initial development of aspirations to enrollment at a particular institution.

In the search phase of the college-choice process, access to a college preparatory curriculum, recruitment programs,

and college and financial aid knowledge are instrumental to the advancement of postsecondary plans. As noted earlier, however, Latino students’ access to these types of resources has been difficult, leading many Latino students to make college choices under less-than-ideal educational circumstances. Indeed, Latino college-choice literature indicates that students are attracted to less-selective institutions that are most likely 2-year institutions, public, less costly, have high dropout rates, and are close to home. In addition, financial aid offers seem to encourage attendance at a particular postsecondary institution that is not necessarily the students’ first-choice institution.

Moreover, Latinas (women) seem to be slightly more sensitive than Latinos (men) to financial issues in the selection of a college, opting to gravitate toward financial and need-based scholarship assistance more often. Latinas, who matriculate into 4-year institutions, tend to enroll in universities that are less selective, less research-intensive, and are primarily master’s-granting institutions. Meanwhile, research has found that Latinas who attend community colleges do so because they decided too late that they wanted to pursue a higher education, find it less costly, are not academically prepared, or opt for a 2-year institution as a last resort. Based on their quantitative examination of the Latino college-choice process, McDonough *et al.* (2004) conclude that “gender, in addition to race, is indeed a critical factor mediating the college choice process for Latinos and Latinas and merits further attention” (p. 35).

Toward a Latino College-Choice Model

Given the emphasis on quantitative methodologies to understand college-choice processes, alternate theoretical frameworks have employed qualitative methods and corresponding theoretical frameworks to highlight nuances inherent in this process. Although quantitative methods are helpful in creating a general picture for mainstream students, existing models fail to capture how race, class, and gender mediate the college-choice process for Latino students. As a result, scholars have utilized frameworks, such as social capital, chain migration, and resiliency, among others,

Predisposition	Search	Choice
Decision to attend college	Explore college information	Completes applications and selects a college
Kindergarten-9/10th	11th-beginning of 12th	12th

Figure 1 College-choice model. From Hossler, D. and Gallagher, K. S. (1987). *Studying Student College Choice. A Three Phase Model and the Implications for Policy Makers*. College and University 2(3), 207–221; and Hossler, D., Braxton, J., and Coopersmith, G. (1989). Understanding student college choice. In Smart, J. (ed.) *Higher Education: Handbook of Theory and Research*, vol. 4, pp 231–288. New York, NY: Agathon.

to bring to light the particular college-choice process for Latino students.

Social Capital Theory, Chain Migration, and College Choice

Coleman (1988) maintained that social networks provided the opportunity for the exchange of information. However, the quality of information exchanged was largely based on the functional components of social capital, trust, and norms. High levels of trust between individuals facilitated the exchange of more knowledge, while norms regulated and influenced behavior related to this exchange. Within education, social capital is the relationships between students, families, communities, and teachers available to support and motivate students toward academic success. Further, social capital theory captures the effects of the school, parents, and community on a students' learning environment (Croninger and Lee, 2001).

Social capital has been used on research that includes Latino students to look at the role of institutional agents on high-school students' educational and occupational expectations (Stanton-Salazar and Dornbusch, 1995); to examine the influence of familial and nonfamilial member social capital on the study habits of Mexican-American eighth graders in comparison to Vietnamese youth (Goyette and Conchas, 2002); and has also been used to understand how familial and school staff positively influence (or hinder) postsecondary access for Latinas (González *et al.*, 2003).

Meanwhile, Person and Rosenbaum (2006) used chain migration theory to examine access to college information and the enrollment decisions of Latino students at 14 2-year colleges. They argued "research on immigrant communities can inform the study of college enrollment, as it encourages the researcher to examine enrollment and persistence as part of a continuous process" (p. 52). Person and Rosenbaum (2006) also suggested that the structures that facilitated

migration to a new community were likely to influence an individual's future experiences. Future experiences may include career as well as academic endeavors.

Using social capital theory as a framework for studying college choice brings into focus those relationships that create and established networks and resources that facilitate Latinos through the educational pipeline from initial aspirations on to higher education. In addition, chain migration influences also are emphasized and interpreted within this social capital theoretical framework. Indeed, having the requisite social capital is necessary in order to establish chain migration contacts (Massey, 1998). More clearly, without having established relationships with certain individuals and developed networks with accessible information, those social structures necessary for migration are absent (Tilly and Brown, 1968). According to Person and Rosenbaum (2006), when applying the theoretical concept of chain migration to college choice, students might (1) apply or select colleges where someone they know has preceded them; (2) apply or choose a college along with someone they know; or (3) seek out contacts once at the college. Based on research, using the concept of chain migration and social capital theory, a new framework for understanding Latino and Chicano college choice is proposed (Figure 2).

This revised college-choice model allows us to understand that for Latinos, the decision to attend college cannot be conceptualized as a natural progression that occurs in their educational trajectory. As noted by Yosso and Solórzano (2006), Latinos experience significant challenges in the educational pipeline, with many of them leaking out before the completion of high school. The educational opportunities that Latino students envision for their future, as well as the subsequent choices they make about their postsecondary options, are shaped by their access to educational learning opportunities available at their schools as well as the resources and information

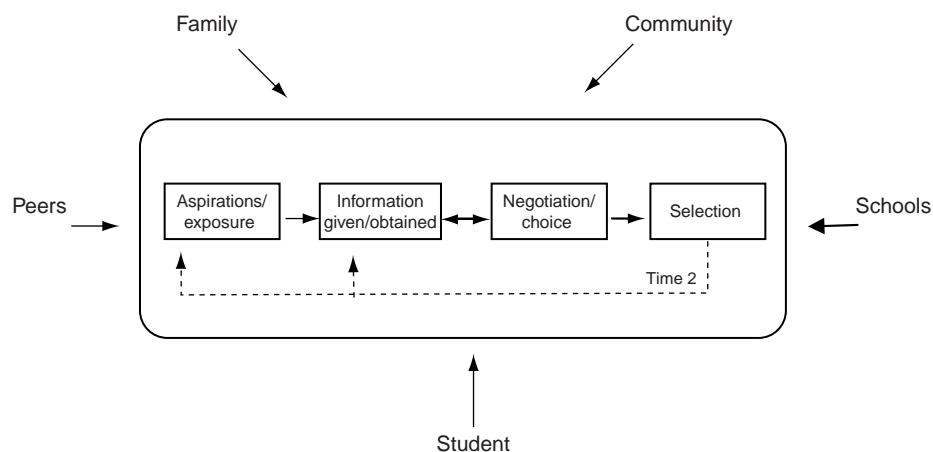


Figure 2 Latino/Chicano college-choice model. Pérez, P. A. (2007). *Social Capital and Chain Migration: The College Choice Process of Chicana and Chicano Community College, Transfer and University Students*. Unpublished Doctoral Dissertation, University of California, Los Angeles.

available to them at home. Latinos, like other ethnic/racial groups, who have parents with little or no college background often face the daunting task of navigating their academic endeavors and college-choice process alone (Ceja, 2006; González *et al.*, 2003). Many of these students are not well informed about accessing higher education and they also lack the instrumental knowledge, mentorship, and social support network necessary to negotiate their academic preparation and college options (e.g., Cooper *et al.*, 1998). In fact, research has shown that ethnic/racial minorities, including Latinos, receive poor and conflicting academic mentorship from institutional agents, such as high-school guidance counselors, resulting in gate-keeping of college resources (Stanton-Salazar, 2004).

This Latino/Chicano college-choice model takes the educational conditions of Latino students into account by placing issues of access to college information and the negotiation of such resources at the center of the college-choice process. These educational conditions are undoubtedly molded by external factors such as the surrounding community and/or school. The model further accounts for community college and transfer students, who may not necessarily follow an immediate or linear progression toward higher education. Should transfer students desire and are able to transfer, this model takes into consideration the second college-decision-making process (see **Figure 2** – time 2). Unlike previous college-choice models, the proposed model highlights student agency and the influence of family, peers, schools, and the community on the Latino/Chicano college-choice process (Pérez, 2007). A conceptual model that better reflects the college-choice decision-making process for Latino students can assist us in enhancing their postsecondary prospects by highlighting critical areas in need of resources. This model emphasizes the need to involve all constituents (e.g., families, peers, and schools) to truly improve equality of postsecondary opportunity for Latino and Chicano students. Finally, this type of model can be helpful in understanding the college-choice experiences of students outside of the United States. For example, the chain migration focus of this model provides an alternate framework for capturing Hamrick's (2007) finding that Pakistani students who enroll in US higher education institutions tend to choose schools that are near relatives or colleges where other Pakistanis are enrolled. Likewise, the applicability of this model can be tested to understand Kaufman's (2008) research on how perceived costs of higher education affects college attendance in Mexico, or how college choices in Portugal are influenced by proximity to home, parents, teachers, and access to college marketing resources (Raposo and Alves, 2007).

See also: Access and Equity in Higher Education; Diversity of Higher Education; Pathways and Articulation into Higher Education.

Bibliography

- Barnett, W. S., Lamy, C., and Jung, K. (2005). *The Effects of State Pre-K Programs on Young Children's School Readiness in Five States*. New Brunswick, NJ: The National Institute for Early Education Research. Rutgers University.
- Bateman, M. and Hossler, D. (1996). Exploring the development of postsecondary education plans among African American and white students. *College and University* 72(1), 2–9.
- Bemark, F. and Chung, R. (2005). Advocacy as a critical role for urban school counselors: Working towards equity and social justice. *Professional School Counseling* 8(3), 196.
- Blumberg, C. Z., Venegas, K. M., Oliveres, P. M., and Colyar, J. E. (2004). School counsel: How appropriate guidance affects educational equity. *Urban Education* 39(4), 442–457.
- Cabrera, A. F. and La Nasa, S. M. (2001). On the path to college: Three critical tasks facing America's disadvantaged. *Research in Higher Education* 42(2), 119–149.
- Ceja, M. (2001). *Applying, Choosing, and Enrolling in Higher Education: Understanding the College Choice Process of First-Generation Chicana Students*. Unpublished Doctoral Dissertation, University of California, Los Angeles.
- Ceja, M. (2006). Understanding the role of parents and siblings as information sources in the college choice process of Chicana students. *Journal of College Student Development* 47(1), 87–104.
- Chapman, D. W. (1981). A model of student college choice. *Journal of Higher Education* 52, 490–505.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology* 94, S95–S120.
- Cooper, R., Cr., Jackson, J. F., Azmitia, M., and Lopez, E. M. (1998). Multiple selves, multiple worlds: Three useful strategies for research with ethnic minority youth on identity, relationships, and opportunity structures. In McLoyd, V. C. and Steinberg, L. (eds.) *Studying Minority Adolescents: Conceptual, Methodological, and Theoretical Issues*, pp 111–125. Mahwah, NJ: Erlbaum.
- Croninger, R. G. and Lee, V. E. (2001). Social capital and dropping out of high school: Benefits to at-risk students of teachers' support and guidance. *Teachers College Record* 103(4), 548–581.
- Fry, R. (2003). *Hispanic Youth Dropping Out of U.S. schools: Measuring the Challenge*. Washington, DC: PEW Hispanic Center.
- Gándara, P. (1995). *Over the Ivy Walls: The Educational Mobility of Low Income Chicanos*. Albany, NY: State University of New York Press.
- Gándara, P. (2002). A study of high school Puente: What we have learned about preparing Latino youth for postsecondary education. *Educational Policy* 16(4), 474–495.
- Garcia, E. E. and Gonzalez, D. M. (2006). *Pre-k and Latinos: The foundation for America's future*.
- González, K. P., Stoner, C., and Jovel, J. E. (2003). Examining the role of social capital in access to college for Latinas: Toward a college opportunity framework. *Journal of Hispanic Higher Education* 2(1), 146–170.
- Gormley, W. T. (2007). The effects of Oklahoma's pre-k program on Hispanic children. *Paper Presented at the Annual Meeting of the Association for Public Policy and Management*. Washington, DC.
- Goyette, K. A. and Conchas, G. Q. (2002). Family and non-family roots of social capital among vietnamese and mexican American children. *Sociology of Education* 13, 41–72.
- Gutiérrez, K., Baquedano-López, P., and Álvarez, H. H. (2000). The crisis in Latino education: The norming of America. In Tejeda, C., Martinez, C., and Leonardo, Z. (eds.) *Charting New Terrains of Chicana(o)/Latina(o) Education*, pp 213–232. Cresskill, NJ: Hampton Press.
- Hamrick, J. F. (2007). *The College Choice Experience of Pakistani Undergraduates in the United States*. Unpublished Doctoral Dissertation, The University of Michigan.
- Harvey, W. B. (2002). *Minorities in Higher Education 2001-2002: Nineteenth Annual Status Report*. Washington, DC: American Council on Education.
- Hossler, D., Braxton, J., and Coopersmith, G. (1989). Understanding student college choice. In Smart, J. (ed.) *Higher Education: Handbook of Theory and Research*, vol. 4, pp 231–288. New York, NY: Agathon.

- Hossler, D. and Gallagher, K. S. (1987). *Studying Student College Choice. A Three Phase Model and the Implications for Policy Makers. College and University* 2(3), 207–221.
- Hossler, D., Schmit, J. S., and Vesper, N. (eds.) (1999). *Going to College: How Social, Economic, and Educational Factors Influence the Decisions Students Make*. Baltimore, MD: Johns Hopkins University Press.
- Kaufman, K. M. (2008). Understanding the income gradient in college attendance in Mexico: The role of heterogeneity in expected returns to college. Department of Economics, Stanford University.
- Kim, D. (2004). The effect of financial aid on students' college choice: Differences by racial groups. *Research in Higher Education* 45(1), 43–70.
- Kurlaender, M. (2006). Choosing community college: Factors affecting Latino college choice. *New Directions for Community Colleges* 133, 7–16.
- Litten, L. H. (1982). Different strokes in the applicant pool: Some refinements in a model of student college choice. *Journal of Higher Education* 53(4), 383–402.
- Massey, D. S. (1998). Contemporary theories of international migration. In Massey, D. S., Arango, J., Hugo, G., et al. (eds.) *Worlds in Motion: Understanding International Migration at the End of the Millennium*, pp 17–59. New York, NY: Oxford University Press.
- McDonough, P. (1997). Access, equity, and the privatization of college counseling. *Review of Higher Education* 20(3), 297–317.
- McDonough, P. (2004). *The School-to-College Transition: Challenges and Prospects*. Washington, DC: American Council on Education, Center for Policy Analysis.
- McDonough, P., Nuñez, A.-M., Ceja, M., and Solórzano, D. G. (2003). A model of Latino college choice. *Paper Presented at the Annual Meeting of the Association for the Study of Higher Education*. Portland, OR, USA.
- McDonough, P., Nuñez, A.-M., Ceja, M., and Solórzano, D. G. (2004). Building models of Latino/a's college choice. *Paper Presented at the Annual Meeting of the American Educational Research Association*. San Diego, CA, USA.
- McDonough, P., Perez, L., Fann, A., et al. (2000). *Parent Involvement Program in Education: Best Research and Practices*. Los Angeles, CA: University of California, Los Angeles, GEAR-UP State Support Systems for Families Implementation Committee.
- National Center for Education Statistics (NCES) (2006). *Digest of Education Statistics, 2005, United States Department of Education*. Washington, DC: US Government Printing Office.
- National Center for Education Statistics (NCES) (2007). Status and trends in the Education of racial and ethnic minorities. <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007039> (accessed June 2009).
- Oakes, J. (1990). *Multiple Inequalities: The Effects of Race, Social Class, and Tracking Opportunities to Learn Mathematics and Science*. Santa Monica, CA: The RAND Corporation.
- Orfield, G. (1992). Money, equity, and college access. *Harvard Educational Review* 62(3), 337–372.
- Orfield, G. and Lee, C. (2007). *Historic Reversals, Accelerating Resegregation, and the Need for New Integration Strategies*. Los Angeles: The Civil Rights Project.
- Pascarella, E. and Terenzini, P. (1991). *How College Affects Students*. San Francisco, CA: Jossey-Bass.
- Pérez, P. A. (2007). *Social Capital and Chain Migration: The College Choice Process of Chicana and Chicano Community College, Transfer and University Students*. Unpublished Doctoral Dissertation, University of California, Los Angeles.
- Perna, L. (2000). Differences in the decision to attend college among African Americans, Hispanics, and whites. *Journal of Higher Education* 71(2), 117–141.
- Perna, L. W. (2005). The key to college access: Rigorous academic preparation. In Tierney, W., Corwin, Z. B., and Colyar, J. E. (eds.) *Preparing for College: Nine Elements of Effective Outreach*, pp 113–134. Albany, NY: State University of New York Press.
- Person, A. E. and Rosenbaum, J. E. (2006). Chain enrollment and college enclaves: Benefits and drawbacks of Latino college students' enrollment decisions. *New Directions for Community Colleges* 133, 51–60.
- Raposo, M. and Alves, H. (2007). A model of university choice: An exploratory approach. <http://ideas.repec.org/p/pramprapa/5523.html> (accessed June 2009).
- Santiago, D. A. and Cunningham, A. (2005). *How Latinos Pay for College: Patterns of Financial Aid for 2003–04*. Washington, DC: Excelencia in Education and the Institute for Higher Education Policy.
- Shulock, N. and Moore, C. (2007). *Rules of the Game: How State Policy Creates Barriers to Degree Completion and Impedes Student Success in the California Community Colleges*. Sacramento, CA: Institute for Higher Education Leadership and Policy, California State University.
- Solórzano, D. G. and Tejeda, C. (1998). The social and educational conditions of Latinas and Latinos in Los Angeles. Unpublished manuscript, University of California, Los Angeles.
- Stanton-Salazar, R. D. (2004). *Manufacturing Hope and Despair: The School and Kin Support Networks of U.S.-Mexican Youth*. New York: Teachers College Press.
- Stanton-Salazar, R. D. and Dornbusch, S. M. (1995). Social capital and the reproduction of inequality: Information networks among Mexican-origin high school students. *Sociology of Education* 68(2), 116–135.
- Talavera-Bustillos, V. H. (1998). *Chicana College Choice and Resistance: An Exploratory Study of First-Generation Chicana College Students*. Unpublished Doctoral Dissertation, University of California, Los Angeles.
- The College Board (2007). *2007 College Board Seniors*. Washington, DC: The College Board.
- Tilly, C. and Brown, C. H. (1968). On uprooting, kinship, and the auspices of migration. *International Journal of Comparative Sociology* 8, 139–164.
- TRPI (2004). *Caught in the Financial Aid Divide: A National Survey of Latino Perspectives on Financial Aid*. Reston, VA: The Sallie Mae Fund.
- Yang, S. W. (1981). Rural youth's decision to attend college: Aspirations and realizations. *Paper Presented at the Annual Meeting of the Rural Sociological Association*. Guelph, ON, USA.
- Yosso, T. J. (2005). Whose culture has capital? A critical race theory discussion of community cultural wealth. *Race Ethnicity and Education* 8(1), 69–91.
- Yosso, T. J. and Solórzano, D. G. (2006). Leaks in the Chicana/o educational pipeline. *Latino Policy and Issues Brief No. 13*. Los Angeles, CA: Chicano Studies Research Center, University of California.

Ethnic Minority Identity and Educational Outcomes in a Rising China

G A Postiglione, The University of Hong Kong, Hong Kong, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Gini coefficient – A measure of statistical dispersion most prominently used as a measure of inequality of income distribution or inequality of wealth distribution.

Harmonious society – A political and educational campaign of the Chinese Communist Party to address indications of growing social instability and focused on the coordination of economic and social development, promote social equity and justice, and to strengthen social affairs management and services.

Nationality autonomous regions – The five provincial level areas of the country associated with an ethnic minority and recognized in the constitution as having a number of rights not accorded to other provinces.

Preferential treatment policies – The policies meant to provide advantages to ethnic minorities. In education, it might include financial subsidies for schools, bilingual education, and points added to university entrance-examination scores.

Introduction

Despite China's meteoric rise, little international scholarly attention has focused on its 110 million ethnic minority population who occupy half of China's land mass and 90% of its border regions (Gladney, 1991; Harrell, 2001; Mackerras, 1994, 1995). China's centuries of experience with ethnic intergroup processes and how it navigates its global economic integration in an increasingly multicultural world add up to a Chinese ethnicity on the move, and one with potentially far-reaching international implications. Already, there is an apparent clash of East–West cultural perspectives on the plight of Tibetans in China.

China presents itself as a multiethnic state but promotes cultural assimilation (Heberer, 1989). Educational institutions in ethnic minority regions become civilizing agencies to transmit mainstream Han Chinese cultural capital. The aspiration to become an economic power, and also to restore its status as one of the world's great civilizations – especially following the nineteenth-century humiliations at the hands of the Western powers – drives

national ethnicity. While its economic rise is a promising start for this restoration, China remains highly sensitive about the presentation of its twenty-first-century civilization to the larger global community.

Thus, reconciling ethnic minority cultures within the vision of its national civilization is a major undertaking. China is moving with trepidation toward embracing a Western-style multiculturalism. Yet, it is aware that many of its overseas families – still considered patriotic and Chinese by blood – have become integrated into Western multicultural countries. Unlike the United States with a large population of voluntary migrants, China is a land of indigenous minorities who are not likely to migrate and therefore more akin in this respect to Native Americans (Ogbu, 1978). While several of China's ethnic minority groups sustained their own slave societies into the early twentieth century, these differed in key respects from the slave experience of African Americans. The American government's concern with a Hawaii separatist movement pales in comparison with China's intense focus on its territorial unity – still a major theme in its ethnic minority discourse and foreign policy. For these and other reasons, the Chinese education makes national unity a central focus. With most of its overseas students and future leaders having studied in the United States, China has come to view the United States as a nation that has successfully used education to build a staunchly patriotic nation. (Like the United States in earlier times, China has used the term nationalities to refer to what is now rendered in English as ethnic groups (the term, in Chinese, for nationalities (*minzu*) and ethnic groups is the same), although there has been a growing popularity among Chinese anthropologists (Bilik, 2000) to use the word *zucun* for ethnic group.)

This article argues that China contributes to the global picture of ethnicity by having sustained the longest history of state-sponsored preferential treatment policies in education, and the manner in which it uses education to navigate its current phase of critical pluralism within its emergent global integration.

Ethnic Minority Education Policy

Most – though not all – of China's ethnic minorities occupy western regions of the country which are poorer and more remote than China's prosperous east-coast

urban areas. Many ethnic minority families have only just entered their first generation of literacy and schooling. Moreover, since most ethnic minorities live in concentrated communities called autonomous regions, their educational opportunities are far lower than what would be available to those in China's prosperous cities such as Shanghai, Beijing, and Guangzhou. In fact, China's pace of economic development has exacerbated, rather than alleviated, social inequality. According to the Gini coefficient – which is used to measure the degree of inequality within a country – China's Gini coefficient of 0.47 makes it the most unequal country in Asia following Nepal. Nevertheless, of the 55 officially designated ethnic minorities of China, several groups (i.e., Koreans, Tatars, Daur, Russian, Xibe, Bai, and others) have educational achievement levels near or above the national average. It is also notable that socialist China had much earlier instituted a set of comprehensive preferential treatment policies that are still in effect.

These preferential policies provide subsidies for minority schools, bilingual education, school textbooks in minority languages, curriculum reforms, ethnic minority teachers training, and boarding schools. In higher education, the most notable policies are those that provide preferential access to college and university through both admission quotas and points added to scores on the national entrance examinations, as well as remedial classes during the first year of college and university (Xia *et al.*, 1999).

These policies are widely believed to ensure educational opportunities for economically and educationally underserved ethnic minorities, especially under the increasing impact of marketization. While these policies have been successful in reducing inequalities by raising the numbers of ethnic minorities attending school and university, most ethnic minority groups still have lower rates of access and educational achievement, especially at university level.

The Sociology of Ethnic Minority Education in China

Research on ethnic minority-education underachievement in China has moved ahead in recent years. The focus has been on multiple factors – ranging from language and religion, cultural transmission and family background, migration, and politics (Iredale *et al.*, 2001; Hannum, 2002; Lee, 1986, 2001; Stites, 1999; Lam, 2005; Zhou and Sun, 2004; Mackerras, 1999; Gladney, 1999; Lin, 2005b; Harrell and Ma, 1999; Hansen, 1999; Tsung, 2003; Nam, 1996; Yu, 2009; Cheung, 2003; Chen, 2004). While there are several works that focus on university students, most focus on ethnic identity in provincial or special universities (Lee, 2001; Trueba and Zou, 1994; Clothey, 2003) rather than

ethnic minority access to and achievement at China's top universities (Sautman, 1999). There is also a growing literature on specific groups. For example, in the case of Tibetans, literature has increased in recent years (Nyima, 1997, 2000; Upton, 1999; Bass, 1998, 2008; Zhang *et al.*, 2008; Seeberg, 2006, 2008; Bangsbo, 2008; Wang, 2007; Yi, 2005; Wang and Zhou, 2003). While the literature on ethnic minority achievement in education is multidimensional, most of it is focused on school rather than university access (Chapman *et al.*, 2000).

Among mainland-based Chinese sociologists and anthropologists, the literature has also grown rapidly. For example, Teng Xing established a research institute for ethnic minority education at the Central University of Nationalities and edits its *Journal of Research on Ethnic Minority Education* (*Minzu jiaoyu yanjiu*) (Teng, 2002). More recently, in a special issue of *Chinese Education and Society*, Ma (2007) has examined one of the stickier issues in the sociology of ethnic minority education – bilingual education and sees an increase in support for Chinese as the medium of instruction in ethnic minority schools. While he favors giving ethnic minority parents a choice of either native language or Chinese as medium of instruction, a market economy prods parents to choose the language of the job market, irrespective of the educational benefit. Zhao (2007) has examined the ethnic minority cultural dimension and concludes that minority culture is marginalized at universities. Zhu (2007) explores how minority students in state boarding schools construct ethnic identities and the specific modes used by individual students. Qian (2007) examined how the hidden curriculum creates discontinuities within cultural traditions and ethnic identities. Ba (2007) – a member of the Yugur ethnic minority of Gansu province – studied how schools disseminate both ethnic minority and modern culture, and argues that there is an estrangement and disjuncture between the culture of the classroom and the community – a result of the choice to absorb the national plan of instruction. Finally, Wang (2008) examined how the children of rural migrants from minority regions who try to settle in Beijing experience double discrimination – as migrants and as ethnic minorities – as they strive for urban membership.

Case Studies

Recent case studies have improved understanding of the processes at work in China's minority education. Although the 1990s saw an increase in the amount of useful quantitative data available, it was unable to capture detailed processes at work as minority students construct meanings and identities from state schooling or to explain why minority children have higher dropout rates. The following section contains the case studies of the education of minorities like Tibetans, Uyghurs, Mongols, Naxis,

and Koreans in China. While the themes of these studies are about national integration, ethnic and national identity, and cultural recognition, a common concern in all of these studies is equal opportunity in educational access and achievement.

Zhu (2007) studied the challenge to Chinese education posed by Tibet. Tibetans have had a written language for over 1500 years – one that is in common use across a territory as large as the continental United States. Despite a rich cultural heritage, Tibet has the lowest educational levels of any provincial level entity in China. Doubtless, this has something to do with poverty and remote geographical location. Therefore, schools and classes for Tibetan students were relocated to Chinese cities beginning in 1985 where better school facilities and trained teachers could be used to foster talent to drive Tibet's economic development. In the first year, one-quarter of all primary school graduates were sent to relocated schools in China for 4 years of junior secondary education and 3 years of senior secondary or vocational education. Zhu provides a grounded view of what actually happens in these schools through the eyes of Tibetan students. He also illustrates the contestation over the meaning of Tibetan culture, who defines it, and how students innovate in constructing their identities around that definition. While the school attempted to assign a desired identity in accordance with the state ideologies, the Tibetan students were able to assert a Tibetan identity expressed through the representation of Tibetan culture, as well as influenced by their experiences on and off their school campus.

Chen (2008) examined how Uyghur students in the so-called Xinjiang Classes (Chinese boarding schools located outside of the home province of Xinjiang), respond to the school goal of ethnic integration. Guided by the theoretical framework of social capital analysis, the study's findings suggest that Uyghur students' response to the goal of ethnic integration can be viewed within a series of analytical levels, including the history of the Uyghurs within China, the Xinjiang boarding schools as a formal organization, Uyghur students' social networks, communal norms and sanctions, and Uyghur students' social actions in the Xinjiang Classes. The study finds that the Uyghur students have created bonding social capital within the social practice of their ethnic norms and sanctions. Their social practice draws ethnic boundaries and demonstrates resistance to the school goal of ethnic integration. The students have also created a linking social capital to tap into the resources of peers, as well as staff and teachers, that improves their academic performance and helps them adapt to life in the boarding school. Nonetheless, they lack bridging social capital which connects them with students and teachers of other ethnic groups, thus making the goal of ethnic integration more difficult to achieve. The study explains the Uyghur students' responses as a form of social recapitalization. While

boarding-school life limits the acquisition of social capital from their families and communities in Xinjiang, they develop new forms of social capital among ethnic peers on campus to help facilitate their academic success.

Zhao (2009) examined ethnic minority cultural recognition at universities through a multicase study focused on the Mongol undergraduate students' experiences and perceptions. She examines the institutional obstacles to cultural recognition in higher education – despite the state and university discourses of equal access to learning through preferential admission policies. Zhao reports on three particular institutions: the Inner Mongolia Normal University (governed by a Mongol nationality autonomous region), Beijing Normal University (governed by the State Ministry of Education), and the Central China University of Nationalities (governed by the State Ethnic Affairs Commission). Zhao expresses doubts on the extent to which the universities recognize Mongol culture. She asserts that Mongols are a “decorated culture,” marginalized within the context of university life and few Mongol students speak Mongol language on campus. She also notes the complex layering of identity between *minkaomin* and *minkaoban* students – the former who enter university by taking the examination in their native language and the latter who take it in Chinese. Not unexpectedly, Inner Mongolia Normal University gives more attention to Mongol culture than Beijing Normal University because the former is within the Inner Mongol Autonomous Region. However, even in the Central China University of Nationalities, ethnic culture is marginalized. Zhao acknowledges that preferential admission policies help Mongol students, but concludes that the lack of cultural recognition on the part of universities limits meaningful access to higher education in China for its ethnic minorities, thereby sustaining their patterns of underachievement.

Yu (2009) studied state schooling and Naxi minority-identity construction. She examined Naxi secondary school students' experience, as well as the role played by Naxi intellectuals', as an asset in student-identity resurgence since the 1980s. The changing roles of school, community, and family in the identity construction suggest that Naxi students retain a strong Naxi identity, by inheriting the knowledge, values, and worldview of their ethnic group, while also managing to fit into mainstream culture. Three forces affect identity construction of the Naxi students: the state and the school; Naxi intellectuals; and socialization in the family and community. As an institution of the state, the school conveys national ideology and instills a sense of ethnic unity and an understanding of the culture of the Chinese nation. While the school takes an active role in ethnic identity construction of the Naxi students, Naxi intellectuals – through their research publications – respond to policies and activities so as to revitalize Naxi culture. The Naxi process of identity construction is characterized by a relatively harmonious

and creative engagement with ethnic and national identity. Two factors contributed to this harmonious identity construction. First, since the late 1970s, the identity of ethnic minority groups has been gaining strength and recognition in China, while – at the same time – market forces have been creating assimilationist pressures. The minorities have taken steps to revive the use of their native languages, and to demand that their native cultures are taught in the public schools. Second, the Naxi already have a long historical tradition of integrating well into Han Chinese culture. Their traditional education is heavily influenced by Confucianism and interactions over several hundred years with Han Chinese. The study contributes to an understanding of why Naxi students adapt generally better than many other ethnic minority groups in China to state schooling.

Gao (2008) studied ethnic Koreans in China, a group widely recognized as a model minority primarily for their academic success rates which are above the national average in China. This research examines how Korean elementary-school students construct meaning out of the model minority stereotype in the context of their school and home experience, and how the meaning construction impacts their educational aspirations and strategies in peer networks. Through comparative analysis, Gao notes that, in a variety of cultural contexts, ethnic Koreans survive as a distinct group that participates in the mainstream without being completely assimilated. Koreans in China and the United States are believed to pull themselves up by their cultural predispositions. This research points to the continued need to modify the model minority stereotype that tends to essentialize ethnic Koreans as a homogeneous group with academic attitudes and success. Research results argue that the model minority stereotype may reinforce the cultural deficiency argument about the academic failure of backward minorities, silence the disadvantages suffered by ethnic Koreans, and lead to no active intervention to remedy them.

China's Ethnic Pluralism

Thus, pluralism is as important as harmony in conceptualizing ethnic intergroup processes in China, and has been the source of much cultural vitality throughout China's history. Yet, this pluralism has not been free of ethnic conflict as in imperial times during the Mongol and Manchu Eras when intergroup processes included both harmonious acculturation and conflict-prone impact integration (Dikotter, 1992).

It is helpful to understand the background themes that guide education of ethnic minorities in China (Dreyer, 1976; Fei, 1980). For much of its history, China was a highly pluralistic society in the world and guided by a culturalist tradition that assimilated many groups into

its cultural center. At about the time of the incursions of the Western powers into China in the nineteenth century, this began to change, and, by the twentieth century, China began to adopt the policies of former USSR. This amounted to a more politicized set of themes which led to the establishment of national autonomous regions. There are a number of scholars in China who now suggest that China should draw upon that characteristically culturalist position so as to strengthen national identity among its ethnic minorities (Ma, 2007a, 2007b). In fact, ethnic minority-education policies and practices since the founding of the People's Republic of China have paralleled the changing political climate.

Following the revolution in 1949, the government worked with ethnic minority elites to integrate diverse territories into the national fabric (Dreyer, 1976). Ethnic minority groups were identified and minority languages were recognized and supported. However, political campaigns that stressed class struggle resulted in less generous policies toward cultural vitality of the ethnic minority. The Cultural Revolution wrought havoc on cultural traditions of ethnic minorities. This was followed by a national effort to redress past wrongs, and accompanied by a resurgence of ethnicity. Since 1978, China's economic reforms and its opening to the outside world have greatly increased the interactions between different ethnic communities. This has contributed to a critical pluralism in education in which national patriotism and ethnic minority cultural autonomy have to keep pace with market forces and globalization.

The decision of the Chinese leader Deng Xiaoping in December of 1978 to launch economic reforms and open China to the outside world continues to have major implications for ethnic minority education. In China's new market economy, competition has become part of the national ethos – making preferential policies for minorities less popular among the majority ethnic group. Moreover, as restrictions on population movement are lifted to fuel the labor market for increased economic growth, both ethnic minorities and Han Chinese migrate to urban areas where they increasingly interact and compete. Many Han Chinese also migrate westward to ethnic minority autonomous areas for work and, often, out-compete local ethnic minorities for jobs. This is especially true for jobs in national infrastructure projects linked to the central government's western development project. The same is true for many small-scale businesses that serve the growing population of Han Chinese from other parts of the country.

Within the ethnic minority autonomous regions, inter-ethnic interaction and competition for jobs has increased. The resulting increase in inter-ethnic animosities is viewed as part of the reason for the government's nationwide harmonious society campaign. (While this campaign focuses on ethnic minority regions and minority-Han relations, it is also related to growing urban labor, urban

migration for jobs, rising costs in rural areas, relocation of households due to urban development, and other social problems.)

The underlying theory of ethnic relations in China since the launching of the economic reforms and the opening to the outside world is Fei Xiaotong's *duoyuan yiti geju*, also rendered in English as plurality within the unity of the Chinese nation (Fei, 1986, 1991). This assimilationist theory has floundered in the arena of globalization and market forces. Rather than move China's ethnic minorities toward cultural assimilation, economic globalization and state schooling have also made ethnicity more salient and intensified ethnic identities.

As the reform period unfolded, China's ethnicity entered a period of critical pluralism. This phase of increased interethnic contact, resurgence, and saliency of ethnic identity occurred along with discontinuance of the job-allocation system, market competition for jobs, compulsory mass schooling that stresses a unified national identity, and telecommunications that make it easier for remote ethnic communities to sustain and extend ethnic solidarity. The increasingly critical nature of ethnic pluralism has placed Chinese ethnicity at a crossroad. Ethnic intergroup antagonisms and misunderstandings can foster a nation of plural monoculturalisms in which ethnic groups emphasize their cultural identities above those of the nation and limit their potential to take on a multiple role in national development. The other direction for Chinese ethnicity is toward a harmonious multiculturalism. This would coincide more closely with the state's campaign for a harmonious society. However, the Chinese state has been unwilling to fully embrace multiculturalism. How educational institutions handle the current phase of critical pluralism may determine whether it generates a national society of plural monoculturalism or harmonious multiculturalism (Sen, 2006). If China is to head in the direction of harmonious multiculturalism – in which conflictual tendencies that characterize interethnic relations are viewed as positive opportunities for building increased mutual understanding, then education has a role in transmitting multiculturalism in three ways: first, by making the curriculum of state schools more relevant to the cultural vitality of ethnic communities; second, by stressing multiculturalism and critical thinking skills; and, third, not only by sustaining preferential treatment policies that increase access to higher education but also increasing ethnic minority cultural recognition on campus so as to foster more meaningful access and academic success.

Conclusion

Ethnic diversity in China rivals that of anywhere else in the world and is the source of much of the nation's vitality.

This not only includes diverse cultural traditions and practices, but also the differing social, economic, and political statuses. While the majority Han Chinese culture and Confucian education heritage exert a significant influence on East Asia and beyond, the cultural traditions and practices of China's ethnic minorities are not widely recognized in China's school curriculum. Nevertheless, they are in the family and community through linguistic, religious, and social practices. Religion is central to the lives of most of China's minorities and the Chinese constitution guarantees freedom of religion (Mackerras, 1994, 1995). However, there is a strict separation between schools and religious institutions. All but two of China's 55 officially designated ethnic minority groups have their own languages. Some have more than one, and at least 21 ethnic minority languages are used in schools, either taught as a subject or used as a medium of instruction. Since 90% of China's border regions are occupied by ethnic minorities, schooling is expected to ensure that national identity is strengthened enough to ensure social stability, especially in the case of Tibet and Xinjiang where the government has identified separatist movements. Therefore, educational policies are challenged to strike a balance between local and national interests so that the dual construction of ethnic and national identities can take place simultaneously.

See also: Affirmative Action and Higher Education in Brazil; The Education of Indigenous Students; The History of Education: Race and Education.

Bibliography

- Ba, Z. (2007). A socio-anthropological analysis of the function of Yugur-nationality schools. *Chinese Education and Society* 40(2), 77–88.
- Bangsbo, E. (2008). Schooling for knowledge and cultural survival: Tibetan community schools in nomadic herding areas. *Educational Review* 60(1), 69–84.
- Bass, C. (1998). *Education in Tibet: Policy and Practice Since 1950*. London: Zed Books.
- Bass, C. (2008). Tibetan primary curriculum and its role in nation building. *Educational Review* 60(1), 39–50.
- Bilik, N. (2000). *Xiandai beijingxia de zuqun jiangou (The Structuring of Contemporary Ethnicity)*. Kunming: Yunnan Education Press.
- Chapman, D., Chen, X., and Postiglione, G. A. (2000). Is pre-service teacher training worth the money: A study of ethnic minority regions in China. *Comparative Education Review* 44(3), 300–327.
- Chen, Y. (2004). *Uyghur Students in a Chinese Boarding School: Social Recapitalisation as a Response to Ethnic Integration*. Unpublished doctoral dissertation, University of Hong Kong.
- Chen, Y. (2008). *Muslim Uyghur Students in a Chinese Boarding School: Social Recapitalization as a Response to Ethnic Integration*. New York: Lexington Press.
- Cheung, W. C. (2003). Narrating ethnic identities: A comparative study of education and ethnic identity construction of a minority group. *Paper Presented at the Meeting on Chinese Education*, Chinese University of Hong Kong.
- Clothey, R. (2003). A study of the ethnic identity of students at the central university of nationalities. *Paper Presented at the Annual*

- Meeting of the Comparative and International Education Society, New Orleans, LA.
- Dikotter, F. (1992). *The Discourse on Race in Modern China*. Hong Kong: Hong Kong University Press.
- Dreyer, J. T. (1976). *China's Forty Millions: Minority Nationalities and National Integration in the PRC*. Cambridge, MA: Harvard University Press.
- Fei, X. (1980). Ethnic identification in China. *Social Sciences in China* 1, 97–107.
- Fei, X. (1986). Zhonghua minzu de duoyuan yiti geju (Plurality within the organic unity of the Chinese nation). *Beijing Daxue Xuebao (Journal of Beijing University)* 4, 1–19.
- Fei, X. (1991). *Zhonghua minzu yanjiu xin tance (New Explorations in China's Ethnic Studies)*. Beijing: Chinese Academy of Social Sciences.
- Gao, F. (2008). What it means to be a "model minority"? Voices of ethnic Koreans in Northeast China. *Asian Ethnicity* 9, 55–67.
- Gladney, D. C. (1991). *Muslim Chinese: Ethnic Nationalism in the People's Republic*. Cambridge, MA: The Council of East Asian Studies and Fellows of Harvard University.
- Gladney, D. C. (1999). Making Muslims in China, education: Islamicization and representation. In Postiglione, G. (ed.) *China's National Minority Education: Culture, Schooling and Development*, pp 55–94. New York: Falmer Press.
- Hannum, E. (2002). Educational stratification by ethnicity in China: Enrollment and attainment in the early reform years. *Demography* 39(1), 95–117.
- Hansen, M. (1999). *Lessons in Being Chinese: Minority Education and Ethnic Identity in Southwest China*. Seattle, WA: University of Washington Press.
- Harrell, S. (2001). *Ways of Being Ethnic in Southwest China*. Seattle, WA: University of Washington Press.
- Harrell, S. and Ma, E. (1999). Folk theories of success: Where Han aren't always the best. In Postiglione, G. (ed.) *China's National Minority Education: Culture, Schooling, and Development*, pp 213–242. New York: Falmer Press.
- Heberer, T. (1989). *China and Its National Minorities: Autonomy or Assimilation*. New York: M. E. Sharpe.
- Iredale, R., Bilik, N., Su, W., Guo, F., and Hoy, C. (eds.) (2001). *Contemporary Minority Migration, Education and Ethnicity*. Cheltenham: Edward Elgar.
- Lam, A. (2005). *Language Education in China*. Hong Kong: Hong Kong University Press.
- Lee, C. J. (1986). *China's Korean Minority: The Politics of Ethnic Education*. Boulder, CO: Westview Press.
- Lee, M. J. B. (2001). *Ethnicity, Education and Empowerment: How Minority Students in Southwest China Construct Identities*. Aldershot: Ashgate Press.
- Ma, R. (2007a). Bilingual education for China's ethnic minorities. *Chinese Education and Society* 40(2), 9–25.
- Ma, R. (2007b). *A New Perspective in Guiding Ethnic Relations in the 21st Century: "De-politicization" of Ethnicity*. Paper read at the Beijing Forum, Beijing, China.
- Mackerras, C. (1994). *China's Minorities: Integration and Modernisation in the 21st Century*. Hong Kong: Oxford University Press.
- Mackerras, C. (1995). *China's Minority Cultures: Identities and Integration Since 1912*. New York: St. Martin's Press.
- Mackerras, C. (1999). Religion and the education of China's minorities. In Postiglione, G. (ed.) *China's National Minority Education: Culture, Schooling and Development*, pp 23–54. New York: Falmer Press.
- Nam, Y. (1996). *A Comparative Study of Pai Yao and Han Chinese Junior Secondary School Dropouts in Liannan Yao Autonomous County, Guangdong Province, the People's Republic of China*. Unpublished doctoral dissertation, University of Hong Kong.
- Nyima, P. (1997). The way out for Tibetan education. *Chinese Education and Society* 30(4), 7–20.
- Nyima, P. (2000). *Wenming de kunhuo: Zangzu de jiaoyu zhilu [The puzzle of Civilization: The Way Out for Tibetan Education]*. Chengdu, China: Sichuan Education Press.
- Ogbu, J. (1978). *Minority Education and Caste: The American System in Cross-Cultural Perspective*. New York: Academic Press.
- Qian, M. (2007). Discontinuity and reconstruction: The hidden curriculum in schoolroom instruction in ethnic minority areas. *Chinese Education and Society* 40(2), 60–76.
- Sautman, B. (1999). Expanding access to higher education for China's national minorities: Policies of preferential admissions. In Postiglione, G. (ed.) *China's National Minority Education: Culture, Schooling and Development*, pp 173–211. New York: Falmer Press.
- Seeberg, V. (2006). Tibetan girls' education: Challenging prevailing theory. In Postiglione, G. (ed.) *Education and Social Change in China: Inequality in a Market Economy*, pp 75–110. New York: M. E. Sharpe.
- Seeberg, V. (2008). Girls first! Conditions for promoting education in Tibetan areas of China. *Educational Review* 60(1), 51–68.
- Sen, A. (2006). *Identity and Violence: The Illusion of Destiny*. New York: W. W. Norton.
- Stites, R. (1999). Writing cultural boundaries: National minority language policy, literacy planning, and bilingual education. In Postiglione, G. (ed.) *China's National Minority Education: Culture, Schooling and Development*, pp 95–130. New York: Falmer Press.
- Teng, X. (2002). *Zuqun, wenhua yu jiaoyu (Ethnicity, Culture, and Education)*. Beijing: Minzu Press.
- Trueba, H. and Zou, Y. (1994). *Power in Education: The Case of Miao University Students and Its Significance for American Culture*. Washington, DC: Falmer Press.
- Tsung, L. (2003). Language policy and minority education in China: The case of the Yi, Naxi, Dai and Tibetan schools in Yunnan. *Paper Presented at the Meeting on Chinese Educational Research*, Chinese University of Hong Kong.
- Upton, J. (1999). The development of modern school based language education in the PRC. In Postiglione, G. (ed.) *China's National Minority Education: Culture, Schooling, and Development*, pp 281–342. New York: Falmer Press.
- Wang, C. and Zhou, Q. (2003). Minority education in China: From state preferential policies to dislocated Tibetan schools. *Educational Studies* 29(1), 85–104.
- Wang, L. J. (2008). *Ethnic Migrants, Social Networks and Educational Access: Membership Capitalization in Beijing*. Doctoral dissertation, The University of Hong Kong.
- Wang, S. (2007). The failure of education in preparing Tibetans for market participation. *Asian Ethnicity* 8(2), 131–148.
- Xia, Z., Ha, J., and Wushuer, A. (1999). *Xizang Zizhiqu Minzu Jiaoyu 50 Nian (50 years of ethnic education in the Tibetan Autonomous Region), Zhongguo minzu jiaoyu 50 nian (50 Years of Ethnic Education in China)*. Beijing: Hongqi Chubanshe (Red Flag Press).
- Yi, L. (2005). Choosing between ethnic and Chinese citizenship: The educational trajectories of Tibetan minority children in northwest China. In Fong, V. L. and Murphy, R. (eds.) *Chinese Citizenship: Views from the Margins*, pp 41–67. London: Routledge.
- Yi, L. (2008). *Cultural Exclusion in China: State Education, Social Mobility and Cultural Difference*. London and New York: Routledge.
- Yu, H. (2008). *Schooling and Identity among the Naxi*. Unpublished doctoral dissertation, University of Hong Kong.
- Yu, H. B. (2009). *Identity and Schooling Among the Naxi: Becoming Chinese with Naxi Identity*. New York: Lexington Press.
- Zhang, L., Fu, H., and Jiao, B. (2008). Accounting for Tibetan university students' and teachers' intellectual styles. *Educational Review* 60(1), 21–37.
- Zhao, Z. (2007). Ethnic Mongol students and cultural recognition: Case studies of three Chinese universities. *Chinese Education and Society* 40(2), 26–37.
- Zhao, Z. (2009). *Am I privileged?: Minority Mongol Students and Cultural Recognition in Chinese Universities*. New York: Lexington Press.
- Zhao, Z. Z. (2009). *Minority Mongol Students and Cultural Recognition in Chinese Universities*. New York: Lexington.
- Zhou, M. and Sun, H. (eds.) (2004). *Language Policy in the People's Republic of China: Theory and Practice Since 1949*. Norwell, MA: Kluwer Academic Press.
- Zhu, Z. (2007). *State Schooling and Ethnic Identity: The Politics of a Tibetan Neidid School in China*. New York: Lexington Press.

Further Reading

- Clothey, R. (2005). China's policies for minority nationalities in higher education: Negotiating national values and ethnic identities. *Comparative Education Review* 49(3), 389–409.
- Harrell, S. and Ma, E. (1999). Folk theories of success: Where Han aren't always the best. In Postiglione, G. (ed.) *China's National Minority Education: Culture, Schooling and Development*, pp 213–242. New York: Falmer Press.
- Nyima, P. (2000). *Wenming de kunhuo: Zangzu de jiaoyu zhilu* (The puzzle of civilisation: the way out for Tibetan education). Chengdu: Sichuan Education Press.
- Postiglione, G. A. (1992a). China's national minorities and educational change. *Journal of Contemporary Asia* 22(1), 20–44.
- Postiglione, G. A. (1992b). The implications of modernization for the education of China's national minorities. In Hayhoe, R. (ed.) *Education and Modernization: The Chinese Experience*, pp 307–336. New York: Pergamon Press.
- Postiglione, G. A. (ed.) (1997a). *Special Issues: The Schooling of Tibetans. Chinese Education and Society* 30(4), 3–6.
- Postiglione, G. A. (ed.) (1997b). *Special Issues: The Schooling of Tibetans. Chinese Education and Society* 30(5), 3–6.
- Postiglione, G. A. (ed.) (1999). *China's National Minority Education: Culture, Schooling and Development*. New York: Falmer Press.
- Postiglione, G. A. (ed.) (2001). *Special Issue: Bilingual Education in China. Chinese Education and Society* 34(2).
- Postiglione, G. A. (ed.) (2006). *Education and Social Change in China: Inequality in a Market Economy*. New York: M. E. Sharpe.
- Postiglione, G. A. (2007). School access in rural Tibet. In Hannum, E. and Park, A. (eds.) *Education and Reform in China*, pp 93–116. New York: Routledge.
- Postiglione, G. A. (2008). Making Tibetans in China: The challenges of harmonious Multiculturalism. *Educational Review* 60(1), 1–20.
- Postiglione, G. A. (2009). Dislocated Education: The Case of Tibet. *Comparative Education Review* 53(4), 483–512.
- Postiglione, G. A. and Jiao, B. (2010). Tibet's Relocated Schools: Popularization Reconsidered, with Ben Jiao. *Asian Survey*.
- Postiglione, G. A., Jiao, B., and Gyatso, S. (2005). Education in rural Tibet: Development, problems, and adaptations. *China: An International Journal* 3(1), 1–23.
- Postiglione, G. A., Jiao, B., and Gyatso, S. (2006). Household perspectives on school attendance in rural Tibet. *Educational Review* 58(3), 317–337.
- Postiglione, G. A., Jiao, B., and Manlaji (2007). Language in Tibetan education: The case of the Neidiban. In Feng, A. W. (ed.) *Bilingual Education in China: Practices, Policies and Concepts*, pp 49–71. New York: Multilingual Matters.
- Postiglione, G. A., Teng, X., and Ai, Y. (1995). Basic education and school discontinuation in national minority border areas of China. In Postiglione, G. and Lee, W. O. (eds.) *Social Change and Educational Development: Mainland China, Taiwan and Hong Kong*, pp 186–206. Hong Kong: Centre of Asian Studies.
- Postiglione, G. A., Zhu, Z., and Jiao, B. (2004). From ethnic segregation to impact integration: State schooling and identity construction for rural Tibetans. *Asian Ethnicity* 5(2), 195–217.
- Yi, L. (2005a). Choosing between ethnic and Chinese citizenship: The educational trajectories of Tibetan minority children in northwest China. In Fong, V. L. and Murphy, R. (eds.) *Chinese Citizenship: Views from the Margins*, pp 41–67. London: Routledge.
- Zhou, E. (1983). Guanyu woguo minzu zhengce jige wenti (A few questions concerning our country's nationality policy). In Gansu Sheng Minzu Yanjiusuo (Gansu Province Institute of Minority Nationality Studies) (eds.) *Zhongguo minzu guanxishi lunwen xuanji* (The Historical Relations among China's National Minorities), pp 1–23. Gansu: Gansu Minzu Press.

Ethnic Studies and Women's Studies: From the Past to the Present, Providing a Space in the Academy for Scholarship Focused on the Histories and Experiences of Marginalized People

C A Brown

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Ethnic studies and Women's studies emerged on college campuses as academic disciplines in the late 1960s to the early 1970s (Blumhagen and Johnson, 1978; Karenga, 2002; Labelle and Ward, 1996; Wood, 1979; Yang, 2000). Students, faculty, staff, and community activists played a role in the development of these programs (Blumhagen and Johnson, 1978; Karenga, 2002; Labelle and Ward, 1996; Wood, 1979; Yang, 2000). Through much protest and debate over discrimination and the marginalization of women and people of color, these programs were ultimately approved and institutionalized (Blumhagen and Johnson, 1978; Karenga, 2002; Labelle and Ward, 1996; Wood, 1979; Yang, 2000). These programs are unique in their histories, contributions, and in their interdisciplinary nature. While thousands of universities have implemented programs, departments, and even schools on their campuses, there are still some colleges and universities who have yet to move beyond offering a few classes. However, recent data show that on many college campuses and universities, the initial reasons for the establishment of these disciplines persist in higher education and called for the continued scholarship with a focus on women and people of color.

History

The 1960s was a time of unrest and confrontation, and students (people of color and whites) were at the center of the struggles that ultimately engendered the legitimacy of ethnic studies and women's studies. The struggle against the racist structure and functioning of society began off college campuses. However, students began to see the university as playing a significant role in distributing power. The university was pictured as a microcosm of how society looked and functioned in terms of race, class, and power. It was perceived as racist and unresponsive to women or peoples of color. Students felt as though colleges were committed to the exploitation and oppression of Blacks, other people of color, and the poor, and to their exclusion from the social knowledge, wealth, and power in US society. The decision then was made to take up the struggle against society and the university, which

was seen as society's brain and its intellectual factory which produces leaders and followers as well as its social myths (Karenga, 2002; McEvoy and Miller, 1969; Robinson, *et al.*, 1969; Van Deburg, 1993). This perspective led Black-studies scholars and students to link knowledge and power, campus and community, student learning with student service and activism in and for community, and society and the world.

The social struggles in the 1960s served as both a context and encouragement for the emergence of a student movement which linked itself to these larger struggles for social change both on campus and off campus (Blumhagen and Johnson, 1978; Karenga, 2002; Labelle and Ward, 1996; Wood, 1979; Yang, 2000). There were four basic thrusts in the student movement, each of which aided in creating the context and support for the emergence of Black studies as a discipline. These are (1) the Civil Rights Movement; (2) the Free Speech Movement; (3) the Anti-Vietnam War Movement, and (4) the Black Power Movement (Karenga, 2002). The final thrust of the student movement which led directly to the establishment of Black studies began in 1965 with the emergence of the Black Power Movement.

The Black Power Movement stressed on self-determination, cultural grounding, relevant education, cultural pluralism, and student activism. Within this movement, Black Studies emerged as a movement and a discipline. It began in 1966 at San Francisco State College (SFSC), now San Francisco State University, and was initiated and led by Black students (Edwards, 1970; Karenga, 2002; Orrick, 1969; T'Shaka, 1982). By 1966, the Watts Riots and the Black Power Movement engendered in a more racially self-conscious and assertive activism among Black students at SFSC and on other campuses as they began to respond to this resurgence of national activism. In 1966, the Negro Students Association changed their name to the Black Student Union (BSU), indicating a new identity and direction (Karenga, 2002). That same year, the BSU produced a document demanding the first Department of Black Studies.

To begin working toward their ultimate goal, Black students established a Black arts and cultural series in the Experiment College, which was created in 1966. They also became involved in SFSC's tutorial program for surrounding communities. This and other community-service

activities demonstrated their social commitment to service, which Black Studies advocates placed at the center of the academic and social mission of Black Studies.

There was no resistance to the Experimental College because it was set up with student money, but the demand by the BSU for a legitimate Black Studies Department funded by the college and controlled by Black people brought stiff resistance (Karenga, 2002). By 1968, the situation had escalated to the point where the BSU launched a strike on 6 November around a series of demands connected to the struggle of Black people, including a Black Studies Department, special admissions, financial aid, and decisions on personnel.

In 1967, Black and other minority students at SFSC presented a proposal to the administration. This proposal pushed for more minority students and for institutionalizing a Department of Black Studies. On 8 November 1968, the Black Student Union and Third Liberation Front, a coalition including Black, Chicano, American Indian, and Asian American students, called for a general strike to reclaim their education (Karenga, 2002; LaBelle and Ward, 1996). The main justification for this strike was the historic and systemic exclusion of minorities in higher education, both in admissions and in the curriculum. They felt that the study of American history and society should not merely include minorities but must also derive from the perspective of people from those communities.

Their demands broadened to seek the establishment of a school of ethnic studies that included four minority groups (Blacks, Chicanos, American Indians, and Asian Americans) rather than a separate Black Studies Department. In response to the protest, SFSC decided that it was in the best interest of the institution to establish an ethnic studies program that would offer courses and conduct research from the perspective of minority groups. They argued that those who taught the classes and did the research must be participating members of ethnic minority communities. This criterion was thought to ensure a more authentic view of those communities.

By the fall of 1968, the experiences of SFSC were being duplicated on dozens of campuses throughout the country. That year, ethnic-studies initiatives, programs, centers, and departments at universities and colleges were established around the world. SFSC established the first college of ethnic studies in 1969. While ethnic studies was not expected to last, the growth has continued. In 1983, the University of California, Berkeley approved a doctoral program in ethnic studies.

Much like ethnic studies' parallel with the black power movement, women's studies paralleled the women's movement that occurred in the mid-1960s. The relationship between the two suggests that the principles upon which the movement flourished rest on a strong intellectual base. They felt that the need for knowledge about women and for

women should have a legitimate place within academia. Women's studies was seen as a vehicle for change and expression, as an integral part of the larger context of the feminist movement.

The first women's-studies courses appeared in the late 1960s and early 1970s and the first women's-studies program was founded at San Diego State University in 1970 (Butler *et al.*, 1991). By the end of 1970, the Modern Language Association's commission on the status of women listed 110 college and university courses dealing with women (Crowley, 1999). Moreover, in the early 1970s, journals supporting the idea of women's studies as a separate academic activity appeared.

By the 1990s, more than 2000 accredited colleges and universities offered women's studies as a program of study. Today, women's studies continues to grow as more and more universities and colleges offer majors and minors and new masters and doctorate programs are established.

Women's studies was built by scholars who dedicated themselves to claiming space in the curriculum for women's diverse experiences and societal contributions. They explored gender as one of several important and interdependent social and cultural aspects of identity, including race, class, age, ability, nationality, and sexual orientation. They practiced pedagogies that valued personal growth, reciprocity in relationships, and exchange of knowledge. Their work challenged disciplinary boundaries and critiqued masculine biases embedded in the academy.

Scholarship

Ethnic-studies and women's-studies programs are generally labeled interdisciplinary. Interdisciplinary connotes the idea of multidisciplinary, that is, combining courses from a number of different disciplines into one course of study. Nonetheless, although multiculturalism and feminism were part of a wider seismic shift of social movements, the fact that ethnic studies and women's studies established space either within the disciplines or on their margins was an act of consummate collective political and intellectual will. Scholars in these fields would have to confront research that typically only took into account the perspective and experiences of White men. They were charged with the responsibility of adding other histories, voices, and perspectives to the curriculum to increase the understanding of the human experience to more accurately reflect the diversity of experiences that exist for people.

Justifications for ethnic studies and women's studies moved beyond the allegation that there was need for knowledge about people of color and women to incorporate an additional need for the creation of knowledge by and for them. It was argued that women and people of color could bring new perspectives and values to scholarship

(Blumhagen and Johnson, 1978; Buker, 2003; Burghardt and Colbeck, 2005; Karenga, 2002; Labelle and Ward, 1996; Wood, 1979; Yang, 2000). Although not everyone was convinced that these new perspectives and values were unique to them, as a result of their own subordinate place in the academy as well as society, women and people of color began to pose questions about the prevalent ideas and ways of thinking. While knowledge about women and people of color is important, it is equally important to value the creation of knowledge by them.

Ethnic studies and women's studies have their own set of cohesive and persistent questions that encourage scholars to question the questions (Bartlett, 1990; Wishik, 1986). These scholars persistently question the distribution of power between genders, among ethnic groups, sexualities, and classes. Their questions attempt to recover the contributions of women and people of color in science, society, and the arts. They have worked to challenge and revise canons and standards, and by correcting historical omissions to include the outstanding human contributions rather than merely the most outstanding White male contributions.

Striving for More Autonomy

Despite being in the academy for over 35 years, most ethnic-studies and women's-studies courses cross disciplinary boundaries to foster integrative learning. On many campuses, ethnic and women's studies remain in marginalized positions as programs rather than departments. This can create challenges, as disciplinary departments hold the power associated with intellectual and administrative authority over curricular, hiring, tenure, and promotion decisions. As such, many ethnic-studies and women's-studies programs lack this power and therefore depend on traditional departments to hire and promote faculty trained and willing to do interdisciplinary work. Faculty holding these appointments are required to appease and offer a course that fits in two departments. They also have to rely heavily on their department for resources that impact their professional development.

Furthermore, individual faculty members who are teaching in these programs rather than ethnic-studies and women's-studies departments make scholarly decisions in the context of a patriarchal system. Many worry that power imbalances in colleges and universities favor White males and that much of the curriculum is male centered (Karenga, 2002). Nevertheless, there are increasing numbers of women- and ethnic-studies graduate programs across the country, and more talented interdisciplinary feminist and ethnic-studies scholars will be seeking opportunities to associate with and to teach courses within these disciplines. They are likely to expect

the freedom to work within and beyond departmental boundaries, so retaining them in the academy may depend on the flourishing these programs.

Power Dynamics within the Disciplines

Although ethnic studies and women's studies have similar missions and intent with regard to challenging and adding to previous scholarship, the relationship between these two areas have not been perfect. In both disciplines, there has been critique about the marginalization of the perspective of women in ethnic studies and women of color in women's studies.

Despite differences of sexuality and class, White women had an unrecognized unity which allowed them to be identified as women. Black feminists argued that this collective identity established a sense of White women sharing a reality as women, and of collectively belonging to the cultural traditions which were identified as constituting the realities of women in general (Burghardt and Colbeck, 2005; Crowley, 1999; Hooks, 1982). However, many academic feminists came to realize that their whiteness gave them access to a privilege that was entirely unselfconscious and so unacknowledged and unrecognized (Burghardt and Colbeck, 2005; Crowley, 1999).

This resonated closely with the criticism that women scholars had of their male colleagues. They argued that their maleness gave them access to a privilege that was entirely unselfconscious. For White feminists to concede that differences exist for Black women and women of color, would mean admitting that they had a role in the racist social order (Bhavnani, 1997; Burghardt and Colbeck, 2005). Women's studies was on the whole racially unselfconscious, engendering a critique that women of color were excluded from much of the rhetoric and practices of hegemonic feminism (Bhavnani, 1997; Burghardt and Colbeck, 2005). Their exclusion was compounded not only by the privilege of whiteness but also by the hard-won privilege that White feminists had gained through the development of women's studies (Bhavnani, 1997; Burghardt and Colbeck, 2005). Much in the way that multiculturalism allowed ethnic minorities to participate in the dominant culture provided they had no designs on either the arrangement or the dominant group which initiated it, Black women and women of color could be added-in and be incorporated by the women's studies (Bhavnani, 1997; Burghardt and Colbeck, 2005). Such an accommodation was unsatisfactory because it left the category of feminist itself as singular, universal, unquestioned, and therefore intact (Bhavnani, 1997).

In similar ways, women of color had to make sure their voices and histories were examined and taught in ethnic studies. Given their exclusion from women's studies scholarship, it became even more important for ethnic-studies

programs, their experiences, and histories. Ethnic-studies scholarship, particularly Black-studies scholarship was critiqued for its lack of adequate treatment on women (Karenga, 2002). Black women pushed for space for the teaching and research in Black studies and in women's studies. They challenged male-centered interpretations of African and human reality. Black women scholars produced and pushed for alternative visions (Karenga, 2002). The critique of the exclusion of women of color caused scholars to begin examining how race, class, gender, and sexuality intersect to create different realities and experiences. This inclusion calls for a richer understanding of the diversity that exists within the human experience.

Continued Need for Ethnic and Women's Studies

Recent research (Brown, 2006) that examined two African American studies programs revealed that students' desires to take classes or earn a degree in this field still continue to reflect a number of the reasons that the students fought for the development of such programs. Data were collected at two selective traditionally White institutions. Through document analyses of campuses articles, survey data from 198 student participants, hour-long interviews of 23 students and 11 faculty, Brown (2006) was able to document the contemporary importance of these programs.

The data from this study revealed that racism continues to exist in college campuses and that students take African American studies classes to deconstruct stereotypes and critique racism and learn about the history (which continues to be silenced or marginalized in many K-12 schools). The survey data revealed that the students who take these classes are not just Black; on both campuses over 30% of the students who responded were non-Black and that a large percent (32% at Calvin, 39% at Ford) were White.

The comparative analysis of the data from these two institutions revealed that Black students attending elite, traditionally White institutions have experiences that are similar in some ways and in other ways unique to their institution, personal background, previous education, and racial experiences. Students from both institutions felt that their university's value of the presence of Black students was communicated through a number of policies and practices. Students were particularly concerned with admission policies and practices, something that was of great concern 40 years ago when ethnic studies was developed. Students from the public university explained that the stance against affirmative action, particularly due to their decreased numbers on campus, made them feel devalued. Furthermore, students argued that the absence of affirmative action did not engender a new perspective that they had earned their admission to the institution through merit like other students.

This type of subtle racism has been termed aversive racism. Describing aversive racism Tatum (1997) states that, "Americans have internalized the espoused cultural awareness of fairness and justice for all at the same time that they have been breathing the smog of racial biases and stereotypes pervading popular culture" (p. 118). Students explained that the end of affirmative action did not affirm their admission or presence at Calvin. They still struggled with other's speculations about their inability to do college-level coursework or their level of intelligence. The debate about affirmative action references ideologies that suggests that to treat people equal you must treat them the same. However, despite the reality that Black students at Calvin were held to the same standards in admissions, they did not receive the same treatment as other students.

In addition to admissions, students were greatly affected by their campus culture and their experiences with micro-aggressions (subtle forms of racism) and marginalization. Students on both campuses discussed their need to feel validated. On both campuses, students felt somewhat marginalized either within the curricula, through the privileging of White culture and values, or more overtly through exclusionary practices. In contrast to the negative experiences they had across campus, they explained that they felt valued and validated in African American Studies classes.

Students at both institutions explained that race played a role in their adjustment to college. African American studies served as a safe place to address their issues around race and racism. Students from Calvin explained that African American studies helped them resist racism and cultural oppression. They explained that they were able to have agency in redefining their experiences on campus by seeking out this academic counterspace where their culture, history, and experiences were dominant and valued. For some students, particularly those who came to college from predominantly White high schools, African American studies gave them the opportunity to deconstruct their own stereotypes about Black people. They described these classes as space to shape a positive self-image. Other students who already had positive Black identities explained that they felt validated in African American studies because they could critique and resist racism. Students also expressed a need to know that their experiences with racism were real and that their anger was warranted (Feagin, 2006; Tatum, 1997).

Faculty played a major role in validating and creating validating spaces and experiences for Black students. Faculty on both campuses showed genuine concern for Black students in their adjustment to college and the development of positive racial and intellectual identities. They were described as approachable and caring. They saw their students as capable and provided meaningful feedback. These faculty members were agents who fostered

academic and personal development in and out of the classroom. The effects of their involvement with Black students served as a form of validation. The faculty members also served as transformational role models who through their research and classes modeled a critique and stance against racism while promoting social justice. They also served as transformational mentors, helping students learn to critique and resist racism, while working to disprove negative stereotypes of Blacks. Although it is important for students to find safe spaces where they feel comfortable in college, academic counterspaces are crucial for Black students. These spaces allow students to escape micro-aggressions and marginalization, critique racism, find validation, inclusion, and support. The data from this study reveal the continued relevance of scholarship that challenges the dominant culture and perspective. It reveals a continued need to have spaces for the voices of marginalized people. It also reveals two models, a department and a center, each of which work well for their environment.

Conclusion

Ethnic studies and women's studies brought the issues of marginalized people to the attention of our universities. Although this knowledge has and continues to be challenged every step of the way, ethnic studies and women's studies have accomplished a lot in the past 40 years and continue to have an impact on scholarship that affects the perspectives of students and informs the decisions made in our society at large. Although our society and our universities have made progress with regard to racism and sexism, ethnic-studies and women's-studies programs still have a very important role in the maintenance of marginalized voices through scholarship and courses. These courses and researchers not only impact marginalized groups but students from dominant groups who enrol in them, requiring them to include these voices, histories, and perspectives in their learning and hopefully their understanding of the world.

Bibliography

- Bartlett, K. T. (1990). Feminist legal methods. *Harvard Law Review* **103**, 829–888.
- Bhavnani, K. K. (1997). Women's studies and its interconnections with 'race', ethnicity and sexuality. In Robinson, V. and Richardson, D. (eds.) *Introducing Women's Studies*. Washington Square: New York University Press.
- Blumhagen, K. O and Johnson, W. D (eds.) (1978). *Women's Studies: Contributions in Women's Studies* vol. 2. Westport, CT: Greenwood Press.
- Brown, C. A. (2006). *African American Studies: Helping Traditionally White Universities Support Black students, Deconstruct Racism, and Promote Diversity*. Ann Arbor, Michigan: ProQuest Information & Learning Company.

- Buker, E. (2003). Is women's studies a disciplinary or an interdisciplinary field of inquiry? *NWSA Journal* **15**(1), 73–93.
- Burghardt, D. A. and Colbeck, C. L. (2005). Women's studies faculty at the intersection of institutional power and feminist values. *Journal of Higher Education* **76**(3), 301–330.
- Crowley, H. (1999). Women's studies: Between a rock and a hard place or just another cell in the beehive? *Feminist Review* **61**(1), 131–150.
- Edwards, H. (1970). *Black Students*. New York: Free Press.
- Feagin, J. R. (2006). *Systemic Racism: A Theory of Oppression*. New York, NY: Routledge.
- Hooks, B. (1982). *Ain't I a Woman: Black Women and Feminism*. London: Pluto.
- Karenga, M. (2002). *Introduction to Black Studies*, 3rd edn. Los Angeles, CA: The University of Sankore Press.
- Labelle, T. J. and Ward, C. R. (1996). *Ethnic Studies and Multiculturalism*. Albany, NY: State University of New York Press.
- McEvoy, J. and Miller, A. (1969). *Black Power and Student Rebellion*. Belmont, CA: Wadsworth Publishing Co.
- Orrick, W. H. (1969). *Shut It Down! A College in Crisis: San Francisco State College, October 1968-April 1969*. Washington, DC: U.S. Government Printing Office.
- Robinson, A., Foster, C., and Ogilvie, D. H. (eds.) (1969). *Black Studies in the University*. New Haven, CT: Yale University Press.
- Tatum, B. D. (1997). *Why Are All the Black Kids Sitting Together in the Cafeteria? And Other Conversations about Race*. New York: Basic Books.
- T'Shaka, O. (1982). The San Francisco State strike, a study of the first black student strike in the US. *Journal of Black Studies* **1**(1), 15–23.
- van Deburg, W. L. (1993). *New Day in Babylon: The Black Power Movement and African American Culture, 1965–1975*. Chicago: University of Chicago Press.
- Wishik, H. R. (1986). To Question Everything: The inquiries of feminist jurisprudence. *Berkeley Women's Law Journal* **1**, 64–77.
- Wood, D. J. (1979). *Women's Studies Programs in American Colleges and Universities: A Case of Organizational Innovation*. Ann Arbor, Michigan: University Microfilms International.
- Yang, P. Q. (2000). *Ethnic Studies: Issues and Approaches*. Albany, NY: State University of New York Press.

Further Reading

- Allen, W. and Haniff, N. (1991). Race, gender, and academic performance in U.S. higher education. In Epps, A. and Haniff (eds.) *College in Black and White: African American Students in Predominantly White and Historically Black Public Universities*, pp 95–110. Albany, NY: SUNY Press.
- Assensoh, A. B. (2003). Trouble in the Promised Land: African American studies programs and the challenges of incorporation. *Journal of Black Studies* **34**(1), 52–62.
- Barrett, M. and Phillips, A. (1992). *Destabilising Theory: Contemporary Feminist Debates*. Cambridge: Polity.
- Bowles, G. and Klein, R. D. (eds.) (1983). *Theories of Women's Studies*. Boston, MA: Routledge & Kegan Paul.
- Collins, P. H. (1990). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. Boston: Unwin Hyman.
- Fries-Britt, S. and Turner, B. Spring (2002). Uneven stories: Successful Black Collegians at a Black and a White campus. *The Review of Higher Education* **25**(3), 315–330.
- Harding, S. (1991). *Whose Science? Whose Knowledge? Thinking from Women's Lives*. Ithaca, NY: Cornell University Press.
- Hunter College Women's Studies Collective (1995). *Women's Realities, Women's Choices: An Introduction to Women's Studies*. New York: Oxford University Press.
- Hurtado, S. (1992). The campus racial climate: Contexts of conflict. *Journal of Higher Education* **63**(5), 539–569.
- Hurtado, S. and Carter, D. F. (1997). Effects of college transition and perceptions of the campus racial climate on Latino Students' sense of belonging. *Sociology of Education* **70**(4), 324–345.
- Kann, M. (1980). *Thinking about Politics: Two Political Sciences*. St. Paul, MN: West Publishing Co.

- Lorde, A. (1984). 'The Master's Tools Will Never Dismantle the Master's House' in Audre Lorde *Sister Outsider: Essays and Speeches*. New York: The Crossing Press.
- MacNabb, E. L., Cherryl, M. J., Popham, S. L., and Prys', R. P. (eds.) (2001). *Transforming the Disciplines: A Women's Studies Primer*. Binghamton, NY: The Haworth Press.
- Mohr, J. and Lee, H. (2000). From affirmative action to outreach: Discourse shifts at the University of California. *Poetics* **28**, 47–71.
- Pascarella, E. and Terenzini, P. (1991). *How College Affects Students: Findings and Insights from Twenty Years of Research*. San Francisco, CA: Jossey-Bass.
- Rendon, L. (1994). Validating culturally diverse students: Toward a new model of learning and student development. *Innovative Higher Education* **19**(1), 33–50.
- Robinson, V. (1997). Introducing women's studies. In Robinson, V. and Richardson, D. (eds.) *Introducing Women's Studies*. Washington Square: New York University Press.
- Solórzano, D. and Bernal, D. D. (2001). Examining transformational resistance through a critical race and LatCrit theory framework: Chicana and Chicano Students in an Urban Context. *Urban Education* **36**(3), 308–342.
- Solórzano, D., Ceja, M., and Yosso, T. (2000). Critical race theory, racial microaggressions, and campus racial climate: The experience of African American college students. *Journal of Negro Education* **69**(1/2), 60–73.
- Solórzano, D. and Villalpando, O. (1998). Critical race theory, marginality, and the experience of students of color in higher education. In Torres, C. A. and Mitchell, T. R. (eds.) *Sociology of Education: Emerging Perspectives*, pp 211–224. Albany: State University of New York Press.
- Tierney, W. G. (1999). Models of minority college going and retention: Cultural integrity versus cultural suicide. *Journal of Negro Education* **68**(1), 80–91.
- Turner, C. S. (2003). Incorporation and marginalization in the academy: From border toward center for faculty of color? *Journal of Black Studies* **34**(1), 112–125.
- Williamson, J. A. (2003). *Black Power on Campus: The University of Illinois 1965–1975*. Urbana and Chicago: University of Illinois Press.
- Yosso, T. (2000). *A Critical Race and LatCrit Approach to Media Literacy: Chicana/o Resistance to Visual Microaggressions*. Unpublished dissertation, University of California: Los Angeles.
- Young, C. (1984). An assessment of Black studies programs in American higher education. *Journal of Negro education* **53**(3), 368–378.
- Zinsser, J. P. (1993). *History and Feminism: A Glass Half Full*. NY: Twayne Publishers.

Gender Equity in Higher Education: Challenges and Celebrations

L Morley, University of Sussex, Falmer, Brighton, UK

© 2010 Elsevier Ltd. All rights reserved.

In higher education, gender is reproduced in processes of knowledge production and distribution, opportunity structures, and social relations that occur on a quotidian basis. Gender is a verb as well as a noun, and we constantly do gender. Drawing on recent empirical studies from three continents (Morley *et al.*, 2006, 2008) and international literature, this article investigates how gender is not a given, but is in continual production.

Gender Challenges and Gains

It would be easy to produce yet another catalog of challenges for gender equity in the academy. There has been extensive exposure of gender and power in higher education. Concerns include gender-insensitive pedagogy (Welch, 2006); sexual harassment (Townsend and Geist, 2000); gendered micropolitics (Morley, 1999); limited opportunities for promotion and professional development (Knights and Richards, 2003; Morley *et al.*, 2006); gendered curricula and subject choices; and the underrepresentation of women in senior academic and administrative positions (Blackmore and Sachs, 2001; Husu, 2000) or in high-status disciplines (Bebbington, 2002) and prestigious institutions (Dyhouse, 2003). Women's relation to knowledge itself has been theorized in terms of the way in which gender structures relations of production and reproduction and is linked to knowledge construction, research opportunities, and dissemination (Mama, 1996; Stanley, 1997; Spivak, 1999). Studies have revealed how liberal and strategic interventions for change such as equality policies (Bagilhole, 2002; Deem *et al.*, 2005) and gender mainstreaming are poorly conceptualized, understood, and implemented (Charlesworth, 2005; Morley, 2007a).

Without wishing to advocate an economy of sameness for women in different national locations, it seems that some gender inequalities are globalized (Morley *et al.*, 2005). For example, there is consistently low representation of women in positions of seniority in a range of countries in divergent cultural and geopolitical contexts (Brooks, 1997; Morley *et al.*, 2006; Singh, 2002). It is now almost a form of common-sense knowledge to berate the patriarchy of higher education and the victimhood of women who dare to enter and remain there. Yet, this binaried power relation does not explain why women are entering the academy in droves! The pleasure/pain juxtaposition has intrigued feminist scholars (Quinn, 2007). Women must be getting something out of their engagement with academia.

Women's academic identities can be precarious. Women are simultaneously constructed as winners and losers – winners because they are gaining access in significant numbers, and losers because of their lack of entitlement to leadership and prestigious disciplines. Women, at whatever stage in their professional careers, are expected to be the good girls of the academy (Evans, 2004). Yet, globally, they are remaining largely underrepresented in senior positions (Singh, 2002). So, what is changing? What are the feminist performance indicators for gender equity? Have there been any successes in the last two decades? Is the academy changing in the context of the concerns about feminization (Quinn, 2003)? (There are beliefs that academic work is rapidly becoming feminized insofar as the dispositions required for scholarship are now heavily perceived as more conducive with socially constructed female behavior, for example, discipline, diligence, perseverance, and commitment to self-improvement (Quinn, 2003). The feminization of learning is becoming an international concern, with increasing anxieties about boys' and men's academic failure and exclusion (Burke, 2007; Epstein *et al.*, 1998).

What exactly do we mean by women's increasing participation in higher education? As the numbers have increased, does this mean that higher education allows more discursive space for gender? Issues of voice, silence, and participation have long been a concern of feminist theorists (Gatenby and Humphries, 1999). Speaking as a woman can mean speaking as a gendered self which is often at odds with the putative ungendered or gender-neutral representations and assumptions of academia (Evans, 2008). However, voice is not just about women speaking, but about the inclusion of gender equality in policy, pedagogy, and planning. There has been a long and painful history of exclusion for women, and this is threatening to reemerge now; some quantitative changes have occurred in audit cultures that privilege measurement. In the UK, gender is a disqualified discourse in higher education policy. Quantitative change seems to imply to policymakers that gender is no longer an issue in a postfeminist era (Morley, 2007b). For example, the former Minister for Higher Education in the UK – John Denham – made a speech about priorities for the next 15 years, and these all relate to innovation and wealth creation, rather than to inclusion and equalities. Gender equality was not mentioned once (Denham, 2008). It is pertinent to ask, at this stage, if the ongoing role of the feminist scholar is to indicate gaps and silences in dominant

discourses and practices. Feminists have had to spend considerable amounts of time on critique. We have had few opportunities to celebrate our successes or indeed to envision a more creative future.

Climate Change and Gender Warming?

One major success is the increased numbers of women entering higher education as students. If we consider that women were barred in the UK until the late nineteenth century (Dyhouse, 1995), this represents quantitative progress. In the UK, in 1995 there were 2.5 times more women in the system than in 1970–71 (Abbott and Wallace, 1997). Participation rates for women in higher education have increased between 1999 and 2005 in all regions of the world, with a global gender parity index (GPI) of 1.05, suggesting that there are now more undergraduate women than men in higher education (UNESCO, 2007: 132). However, the increase in women's participation has been unevenly distributed across national and disciplinary boundaries. Women's participation rates are higher than those of men in North America and Europe, but lower in regions such as East Asia and the Pacific, South and West Asia, and sub-Saharan Africa.

Globally, women students are concentrated in non-science subjects. There is still a sense of what constitutes a gender-appropriate discipline in many high- and low-income countries, with worldwide concern about the underrepresentation of women in the science, technology, engineering, and mathematics (STEM) subjects. Men predominate in subjects related to engineering, manufacturing and construction, and math and computer science (OECD, 2007). In many countries, two-thirds to three-quarters of graduates in the fields of health, welfare, and education are women (UNESCO, 2006). Thus, women continue to be concentrated in subjects associated with low-wage sectors of the economy, in particular health and welfare, humanities, arts, and education.

What do all these facts about women's increased participation or exclusion add up to in terms of how women experience the academy? The author has conducted studies on the micropolitics of academic life and frequently found that the gendered relays of power that cause the most distress and discomfort are the everyday transactions and relations (Morley, 1999; 2006). Blending quantitative facts with interview data helps to reveal both the scale and the lived complexities that structure women's participation in higher education, as students and staff. Focusing on everyday micro-level incidents can provide important information about more macro-focused challenges for gender equality. Dealing with quotidian examples of sexism and discrimination can have a detrimental effect on women's self-confidence and career aspirations (Morley, 1999; Seymour and Hewitt, 1997). The personal is political,

or to use more contemporary vocabulary, the self can become an object of reflexive knowledge (Beck and Beck-Gernsheim, 2002; Hey and Leathwood, 2007).

The metaphor of the chilly climate has been used to symbolize and capture the discomforts and inhospitality of academic cultures for women. Sandler *et al.*'s (1996: 1) study in the USA found some 30 ways in which faculty members often treated women students differently in the classroom. This chilly climate impeded women's full participation in the learning process. It is pertinent to ask today if the temperature is rising, and the ecology, culture, and climate changing for women? There are empirical data that seem to suggest both yes and no. While there are evident structures of inequality, women's agency, as opposed to privilege, is certainly cause for celebration. The following sections will consider data from the author's recent research projects in Africa, Asia, and Europe.

Academic Identities and the Gendering of Ability

It is questionable whether the success of quantitative change this has been accompanied by a qualitative change. Has the nature of the higher education product changed in relation to the changing constituencies of students? A criticism of much of the scholarship on academic women is that it focuses on the experiences and voices of privileged white women in high-income countries (Twombly, 1999). Theorizing links between differently located practices can produce a sense of the patterns and scale of gender challenges. The author draws upon some of her research findings from Morley *et al.* (2006). This study explored gender equity in higher education in low-income countries. It aimed to go behind the statistics and explore women's everyday experiences of higher education in Nigeria, South Africa, Sri Lanka, Tanzania, and Uganda. Identifying key sites of gender-differentiated experiences of the academy was a purpose of the research.

The countries were selected for their varying national policies on gender equity and their commitment to international policies to end discrimination against women. South Africa has generated a range of race and gender equity and transformation policies in higher education. Uganda and Tanzania have specific gender-equity policies. Sri Lanka initiated social welfare policies in the post-independence period. Nigeria has a policy of widening participation in higher education, with emphasis on developing science and technology. A total of 209 interviews were conducted with students, academic staff, and managers. Observation of classrooms and boardrooms was conducted and statistics and policies were analyzed. A noticeable finding was how gender inequalities appear to be fairly globalized. While transnational feminism is

problematic in terms of the diversity of women's oppressions (Mills and Ssewakiryanga, 2002), observations from women in low-income countries can sound remarkably similar to women's voices in the West.

In the study, a strong sense of a hidden curriculum emerged (Margolis, 2001). During the 1970s, the hidden curriculum became one of the powerful concepts in the new (critical) sociology of education. Increasingly, the hidden curriculum is perceived as a vital part of more general curriculum transformation. The overt and the hidden curricula are not mutually exclusive but form a complex mechanism of production and reproduction (Apple, 1980). The hidden curriculum is irrational and contradictory. Negative attitudes to women's academic abilities do not correlate with their actual achievements.

One aspect of the hidden curriculum relates to the conjunction between gender and academic ability and authority. Studies have reported how discrimination against women can involve not taking them seriously and doubting their ability and motivation (Seymour and Hewitt, 1997). Discrimination due to perceived incompetence is based on descriptive gender stereotypes (Rudman and Glick, 2001). Difference is frequently expressed in terms of deficit and located within particular bodies rather than in the "invisible values and assumptions structuring curriculum and pedagogy" (Abu El-Haj, 2003: 411).

In Morley *et al.* (2006), there was widespread reporting of negating women's academic abilities and hostility from male students and staff, as a South African student illustrates:

And I mean the guys ... think we are absolutely useless. I mean we might score high marks you know in courses, but it is just the fact that they think we are stupid. And even our lecturers, I mean, I have a particular lecturer, who just thinks I am an idiot, and I have no reason, I have given him no reason to think that. But I am not interested in proving him wrong and if he wants to think that, then that's great.

Femaleness is repeatedly perceived as irreconcilable with intellectual authority (Shah, 2001). Another South African student reports negativity toward women staff from male students, thus reinforcing the notion that women are perceived as inauthentic academics:

I have also noticed how we've had maybe two or three female lecturers and how the guys in our class just do not listen to them, they do not respect them. And I mean these women are really good, they are brilliant, they know their stuff they worked hard they have their PhDs, but guys laugh at them, ridicule them.

In an observation evocative of Spender's (1982) early work in the UK, and Brooks' (1982) early work in the USA, the Sri Lankan team relate how their classroom observations revealed male students receiving more pedagogical attention:

Towards the end of the lecture when the lecturer was relating real-life examples to the theory he had been teaching, he mentioned a project the female student was involved in and briefly asked her a question related to it but he did not give her any time to answer, smiled and moved on to the next question very swiftly.

The Sri Lankan team also notes how generally male students were invited to comment and question more than females. Consequently, male students became more confident, more assertive, and relaxed than their female counterparts. This gendered interaction did not go unnoticed by students. A Sri Lankan student describes discriminatory behavior from some male lecturers:

There are some who try to put the women down by asking a question and then laughing at us when we can't answer it, or ask something just to put us down.

The gendering of pedagogical interactions poses questions about the full meaning of the concept of women's participation.

A Ugandan student points out that there is a horns effect toward women and a halo effect toward male students, and how this distorts academic assessment:

Male lecturers need to be gender sensitised so that they can avoid their patriarchal tendencies of thinking that male students are better than female students e.g. there was a situation when two students (a female and male) handed in the same piece of work, the lecturer awarded marks to the male student and cancelled the work of the female student on assumption that the female student had cheated. This in my view was not fair.

The notion of men as real academics and women as fakes is also noted by a Sri Lankan student:

[Men] hardly attend class. But get their notes from women. I know of several incidents where the boys have copied the tutorial and given it in and they've got higher marks for the same thing.

Anti-feminism was also widely reported. A Nigerian student describes:

...when you are tagged as a Women's Lib person they see you as way out, deviant person.

It appears that the social penalties are sometimes high for those individuals who set out to interrogate and challenge gender inequalities. They are easily othered, raising questions about whether women participate in higher education on their terms and to follow their own intellectual and political interests.

Women students in Morley *et al.* (2006) were also perceived as impeded by internalized oppression, that is, their interior worlds or psychic narratives that constantly played recordings of inferiority. Women's academic

self-worth was often presented as fragile and unstable. A Ugandan student states:

The problem most girls have is lack of confidence.

A Sri Lankan policymaker in Morley *et al.* (2006) also attributed the low level of women in management to women's reluctance to apply for the posts:

Managerial posts are not held by women in large numbers. In universities, if you take generally speaking how many heads of departments are females ... no not even 20 per cent are held by the females... That is because they don't come forward. That is the reason.

The problem with affective explanations and attributing problems to psychic narratives, such as lack of self-confidence, is that it suggests that women are deficit in and lack the personal attributes needed to succeed. Cognitive, rather than organizational restructuring, is seen as the solution. Problems that are largely collective and social are individuated. It represents the privatization of the public. The power relations that create structures and barriers, and indeed that undermine women's confidence in their abilities, are overlooked. By offering very conventional indicators of professional success, it also marks out women as losers who prefer not to occupy managerial roles.

The gendering of academic ability has been a theme that has emerged in the author's recent study on widening participation in higher education in Ghana and Tanzania (Morley *et al.*, 2008). This study utilizes statistical data, life-history interviews with 200 students and semi-structured interviews with 200 staff in two public and two private universities. It focuses on how gender and socioeconomic status intersect and constrain or facilitate participation in higher education. The data so far suggest that any activity that is perceived as difficult is seen to be inappropriate for women. A Tanzanian female student describes how female students frequently believe that they need to be academically rescued by male students:

You know that for example this question is tough and only boys can tackle it ... and a girl cannot, and we have to look for a boy, who we think can tackle it.

A male Tanzanian student suggests that academic ability is psychosocially constructed:

The problem is always being talked in the newspapers, that the girls they don't have the confidence ... that's all ... they have high capability ... on starting ... the Form Ones and Form Twos they are certain who are very bright, very bright ... but as the days were going on their capability was decreasing and decreasing ... and perishing.

Success criteria for gender equality frequently relate to women's increased participation in male-dominated areas. It is almost as if by working and studying with men, there will be disidentification with the inferior

world of women and a type of positive contagion of male values and behaviors.

Women are constructed as poor choosers when it comes to academic disciplines. Their entry into nontraditional disciplines is seen as a form of empowerment and hence a cause for celebration. In Tanzania especially, where affirmative action interactions have created access or preentry programs for women to enter science, quantitative change is widely acknowledged as success. A Tanzanian professor provides some statistics:

In the year 1994 for example, the female student population was about sixteen percent after this intervention, we are now in the sciences in particular in the Faculty of Science stands now at around thirty to thirty one percent which to me in ten years is quite an achievement. With Engineering that it was formally six percent less than six percent it has now gone to slightly above ten percent.

Women's academic identities are frequently constructed in terms of remediation or absence. A Tanzanian dean of a science faculty discusses women's underrepresentation in high-status disciplines.

When it comes to gender, I think it's the girls who are not well represented particularly in some disciplines. Sciences are less than fifteen percent. When it comes to Physics, Mathematics, Geology there is huge imbalance between the girls and boys ... In Mathematics it could be up to ... you know between eighty and twenty percent. Even in Geology you know twenty percent girls, eighty percent boys.

It is questionable why underrepresentation usually features in relation to certain subject areas, and why men's underrepresentation in female-dominated disciplines is rarely mentioned in policy terms. In Ghana, a female professor comments on gendered patterns of subject choice:

The Education Faculty has the highest proportion of females ... even though we are getting more women than men, they are still moving into those areas that are known as traditional ... The School of Agriculture ... has 143 students but the number admitted females applied which were qualified ... was 22.

The question of what and where women are accessing can also be related to the type of higher education institution (HEI). In Ghana, women comprise 35% of the overall university population (NCTE, 2006a,b), although they make up 41% of the students in private higher education (NCTE, 2006b). In Tanzania, 33% of the overall undergraduate population is female, with women comprising 38% of students in private higher education (MHEST, 2006). If rates of participation for women are higher in lower-status private higher education, this poses questions about core and periphery provision. Socially

disadvantaged groups could be getting diverted into peripheral higher education, thus reinforcing stratification of the sector and social differentiation. In this analysis, widening participation in higher education can be conceptualized as a process of diversion, that is, a re-routing of members of socially disadvantaged groups into lower-status institutions in order to reserve the higher-status universities for the elite (David, 2007). “*Buying* an education becomes a substitute for *getting* an education” (Kenway *et al.*, 1993: 116).

Vaporizing Gender

Policy commitments to gender can evaporate during implementation (Goetz, 1997; Longwe, 1995). In the UK, gender is no longer included in higher education policy. Elsewhere, there are policies that fail to be implemented. In Ghana, an academic relates how gender remains at the level of policy text, with no strategic implementation plans:

They are all making many noises about tertiary education. All they said was they recommended 50–50 and that was it that was it . . . The only one I can think of is with regard to women; it's a little bit more of lip service, if anything.

The implementation gap was discussed by another Ghanaian academic who believed that insufficient resources were allocated to ensure domestication of international policies:

I think Ghana Government has been very proactive as far as international policies of all types are concerned and in the case of education . . . but I must say that the trouble with African states, including Ghana, is that it is one thing saying ‘yes, we’ll do, we will put it into practice’ and another thing actually getting funds to get it really done, putting it into practice . . .

These observations about the implementation gap are evocative of findings from a UK-based study that the author conducted with Rosemary Deem and Anwar Tlili (Deem *et al.*, 2005). The research involved six case studies of HEIs across England, Scotland, and Wales. The project’s aims included exploring staff experiences of equity issues and institutional equity policies. A central finding was that although all six institutions studied had equal opportunity policies in place, not all the policies were comprehensive, completely up to date, or easy to understand. Policies were often communicated to staff via e-mail, which may not reach those with e-mail overload or staff with no computer access at work. Some policies gave the impression of often having been reactively rather than proactively constructed and with an eye to compliance to legislation rather than empowerment of the workforce

and enhancement of their working conditions. Staffs were wary of utilizing grievance procedures for fear of recrimination and professional suicide. The policies were not integrated into strategic management, and there was little action planning or proactivity. The problem of senior- and middle-management inactivity was observed by an academic trade-union representative from one university who felt that for some, it was sufficient simply to note the numbers rather than take any action to rectify underrepresentation:

Now on sex equality last year there was a round of promotions to principal lecturer and, it was noted that I think the proportion of women who applied, as compared to the proportion of women employed, and the proportion of women I think, was one out of six appointees. And the personnel office simply in their report, noted the numbers. But we tried to push them to think about what might they do about it but they were quite content to just note the disparity between the number of women employed in the academic role and the outcome of this round.

Many of the staff that we interviewed frequently noted how policies existed at a textual level – often to meet the requirements of audit and funding bodies – rather than working at the grassroots level of day-to-day work and felt that there was a major implementation gap.

Conclusion

Women’s exclusion from higher education is a historical injustice. Today, women are participating, in increasing numbers, in higher education, in a range of national locations. Yet, women’s academic identities are often forged in otherness, as strangers in opposition to (privileged) men’s belonging and entitlement. This means that gender in higher education is often encoded in a range of formal and informal signs, practices, and networks.

A number of contradictory arguments are emerging. One is that the increasing participation rates for women means that they are in danger of taking over the academy. However, this is unraveling against contexts and cultures that still position them as academically inferior and lacking confidence. Women are simultaneously accused of taking over and not coming forward. Another contradiction is that women are constructed as strategically engaged with higher education, recognizing the impact on their employability and life chances. However, they are choosing the wrong subjects to study, shying away from STEM disciplines that would provide a greater return for their investments. In terms of staff, women are frequently seen as the good girls of the academy (Evans, 2004); yet, they are not getting equally promoted to positions of seniority, or desiring them sufficiently (Singh, 2002).

The contradiction between policy as text and as lived experience is also a noticeable feature of gender architecture. Policy interventions including gender mainstreaming and gender equality suggest that now gender sensitivity and strategic actions should be everywhere – including the curriculum, management, and resourcing of academic life; yet, informants in the author's studies report that they are nowhere!

Feminist scholars and researchers will continue to critique, theorize, and audit power and privilege in higher education, as it is a major site of cultural practice, identity formation, and symbolic control. However, we also need to imagine or re-imagine a different future, asking what type of academy we wish to see. Whereas the UK minister for higher education has a wish list for the next 15 years that includes the expansion of technology, innovation and research-based wealth creation, we wish to invite an international feminist political imaginary to ask what women want, and what would the gender equitable academy of the future look and feel like?

Bibliography

- Abbott, P. and Wallace, C. (1997). *An introduction to Sociology: Feminist Perspectives*, 2nd edn. London: Routledge.
- Abu El-Haj, T. (2003). Challenging the inevitability of difference: Young women and discourses about gender equity in the classroom. *Curriculum Inquiry* **33**(4), 401–425.
- Apple, M. (1980). The other side of the hidden curriculum: Correspondence theories and the labor process. *Journal of Education* **162**, 47–66.
- Bagilhole, B. (2002). Challenging equal opportunities: Changing and adapting male hegemony in academia. *British Journal of Sociology of Education* **23**(1), 19–33.
- Bebbington, D. (2002). Women in science, engineering and technology: A review of the issues. *Higher Education Quarterly* **56**(4), 360–375.
- Beck, U. and Beck-Gernsheim, E. (2002). *Individualization*. Thousand Oaks, CA: Sage.
- Blackmore, J. and Sachs, J. (2001). Women leaders in the restructured and internationalised university. In Brooks, A. and McKinnon, A. (eds.) *Gender and the Restructured*, pp 45–66. Buckingham: Open University Press.
- Brooks, V. (1982). Sex differences in student dominance in female and male professors' classrooms. *Sex Roles* **8**, 683–690.
- Brooks, A. (1997). *Academic Women*. Buckingham: Open University Press.
- Burke, P. (2007). Men accessing higher education: Theorising continuity and change in relation to masculine identities. *Paper Presented at the SRHE Conference*. Brighton.
- Charlesworth, H. (2005). Not waving but drowning: Gender mainstreaming and human rights in the United Nations. *Harvard Human Rights Journal* **18**, 1–18.
- David, M. (2007). Equity and diversity: Towards sociology of higher education for the 21st century? *British Journal of Sociology of Education* **28**(5), 675–690.
- Deem, R., Morley, L., and Tlili, A. (2005). *Negotiating Equity in UK Universities*. London: HEFCE.
- Denham, J. (2008). Check against delivery. Wellcome collection conference centre: London http://www.dius.gov.uk/speeches/denham_hespeech_290208.html (accessed August 2009).
- Dyhouse, C. (1995). *No Distinction of Sex? Women in British Universities 1870–1939*. London: UCL Press.
- Dyhouse, C. (2003). Troubled identities: gender and status in the history of the mixed colleges in English universities since 1945. *Women's History Review* **12**(2), 169–194.
- Epstein, D. E., Hey, V., and Maw, J. (eds.) (1998). *Failing Boys? Issues in Gender and Achievement*. Buckingham: Open University Press.
- Evans, M. (2004). *Killing Thinking: Death of the Universities*. London: Continuum Books.
- Evans, M. (2008). The Silence of the Girls. *Paper Presented at the Centre for Higher Education and Equity Research (CHEER) Seminar*. University of Sussex, 21 January.
- Gatenby, B. and Humphries, M. (1999). Exploring gender, management education and careers: Speaking in the silences. *Gender and Education* **11**(3), 281–294.
- Goetz, A. (ed.) (1997). *Getting Institutions Right for Women*. London: Zed Press.
- Hey, V. and Leathwood, C. (2007). Gendered discourses and emotional sub-texts in UK higher education: Governing the demo(cra)t(c) self. *Paper Presented at the SRHE Conference*. Brighton, UK.
- Husu, L. (2000). Gender discrimination in the promised land of gender equality. *Higher Education in Europe* **XXV**(2), 221–228.
- Kenway, J., Bigum, C., and Fitzclarence, L. (1993). Marketing education in the post-modern age. *Journal of Education Policy* **8**(2), 105–122.
- Knights, D. and Richards, W. (2003). Sex discrimination in UK academia. *Gender, Work and Organisation* **10**(2), 213–238.
- Longwe, S. H. (1995). *The Evaporation of Policies for Women's Advancement, in a Commitment to the World's Women: Perspectives on Development for Beijing and Beyond*. New York: United Nations Development Fund for Women (UNIFEM).
- Mama, A. (1996). Women's studies and studies of women in Africa during the 1990s. *CODESRIA Working Paper Series*, May 1996.
- Margolis, E. (2001). *The Hidden Curriculum in Higher Education*. London: Routledge.
- Mills, D. and Ssewakiryanga, R. (2002). That Beijing thing: Challenging transnational feminisms in Kampala. *Gender, Place and Culture* **9**(4), 385–398.
- Ministry of Higher Education, Science and Technology (MHEST) (2006). *Basic Statistics on Higher Education, Science and Technology 2001/2002 – 2005/2006*. Dar es Salaam: MHEST.
- Morley, L. (1999). *Organising Feminisms: The Micropolitics of the Academy*. London: Macmillan.
- Morley, L. (2006). Hidden transcripts: The micropolitics of gender in Commonwealth universities. *Women's Studies International Forum* **29**(6), 543–551.
- Morley, L. (2007a). Sister-matic: Gender mainstreaming in higher education. *Teaching in Higher Education* **12**(5/6).
- Morley, L. (2007b). Gender and UK higher education: Post-feminism in a market economy. In Danowitz Sagaria, M. (ed.) *Women, Universities and Change: Gender Equality in the European Union and the United States*, pp 133–144. New York: Palgrave Macmillan.
- Morley, L., Gunawardena, C., Kwesiga, J., et al. (2006). *Gender Equity in Selected Commonwealth Universities. Research Report no. 65 to the Department of International Development*. London: DFID.
- Morley, L., Leach, F., and Lugg, R. (2009). Democratising higher education in Ghana and Tanzania: Opportunity structures and social inequalities. *International Journal of Educational Development* **29**(1), 56–64.
- Morley, L., Sorhaindo, A., and Burke, P. (2005). *Researching Women: An Annotated Bibliography on Gender Equity in Commonwealth Higher Education*. London: Institute of Education.
- NCTE (2006a). *Statistical Digest of Universities and IPS 2005/6*. Accra: NCTE Ghana Records.
- NCTE (2006b). *Student Enrolments for Private Tertiary Institutions from 1999 to 2006*. Accra: NCTE Ghana Records.
- OECD (2007). *Women and Men in OECD Countries*. Paris: OECD.
- Quinn, J. (2003). *Powerful Subjects: Are Women Really Taking over the University?* Stoke-on-Trent: Trentham.
- Quinn, J. (2007). Welcome to the pleasure dome: Women taking pleasure in the university. In Cotterill, P., Jackson, S., and Letherby, G. (eds.) *Challenges and Negotiations for Women in Higher Education*, pp 117–129. Dordrecht: Springer.
- Rudman, L. and Glick, P. (2001). Prescriptive gender stereotypes and backlash toward Agentic women. *Journal of Social Issues* **57**(4), 743–762.

- Sandler, B. R., Silverberg, L. A., and Hall, R. M. (1996). *The Chilly Classroom Climate: A Guide to Improve the Education of Women*. Washington, DC: National Association for Women in Education.
- Seymour, E. and Hewitt, N. M. (1997). *Talking about Leaving: Why Undergraduates Leave the Sciences*. Oxford: Westview Press.
- Shah, S. (2001). Tertiary colleges in Pakistan: Gender and equality. *The School Field* **X11**(3/4), 49–70.
- Singh, J. K. S. (2002). *Still a Single Sex Profession? Female Staff Numbers in Commonwealth Universities*. London: Association of Commonwealth Universities.
- Sponder, D. (1982). *Invisible Women: The Schooling Scandal*. London: Writers and Readers Co-operative.
- Spivak, G. (1999). *A Critique of Postcolonial Reason*. London: Harvard University Press.
- Stanley, L. (ed.) (1997). *Knowing Feminisms*. London: Sage.
- Townsend, N. and Geist, P. (2000). The discursive enactment of hegemony: Sexual harassment and academic organizing. *Western Journal of Communication* **64**(2), 190–217.
- Twombly, S. (1999). Review essay new scholarship on academic women: Beyond women's ways. *The Review of Higher Education* **22**(4), 441–454.
- UNESCO (2006). *Global Education Digest 2006: Comparing Education Statistics across the World*. Montreal: UNESCO Institute of Statistics.
- UNESCO (2007). *Global Education Digest 2007: Comparing Education Statistics across the World*. Montreal: UNESCO Institute of Statistics.
- Welch, P. (2006). Feminist pedagogy revisited. *LATISS-Learning and Teaching in the Social Sciences* **3**(3), 171–199.

Further Reading

- Mantyla, H. (2000). Dealing with shame at academic work – A literary introspection. *Psychiatria Fennica* **31**, 148–169.

Relevant Websites

- www.sussex.ac.uk/education/cheer – Centre for Higher Education and Equity Research (CHEER).
- www.sussex.ac.uk/education/wideningparticipation – ESRC/DFID funded research project 'Widening Participation in Higher Education in Ghana and Scorecard.'

Locating Space and Place in the College Access Debate: New Tools for Mapping and Understanding Educational Inequity and Stratification

B Cantwell and J F Milem, University of Arizona, Tucson, AZ, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Issues critical to college access are closely related to demographic, social, and economic trends. As society changes, so does the social role of higher education. Given the high social and individual returns that accrues from tertiary participation, as well as an interest in social justice, both researchers and policymakers have been focused on broadening participation and exploring ways in which access to higher education can become more equitable. Today, more students from all backgrounds participate in higher education than at previous times. Yet, a greater volume of participants has not resulted in increased equity. College access remains stratified along a number of dimensions including race and ethnicity and socioeconomic status. While evidence suggests that ability is an important factor in access to college, any notion of meritocracy is confronted by the fact that high-income/low-ability students attend college at a higher rate than do low-income/high-ability students.

Continued inequity in college access is particularly disturbing given two social trends. The first is related to demographic changes. While the population is growing older overall, youth populations are becoming significantly more diverse in terms of race and ethnicity, country of origin, first language, and religious and cultural backgrounds. Minority students, or students from nondominant groups, have lower levels of access to higher education, particularly at 4-year colleges and universities. As students from these groups have increased as a proportion of the total population, there have not been commensurate increases in their share of higher education access and attainment at all levels. In other words, access to higher education remains stratified and in many cases this stratification has increased. If this trend continues uninterrupted, larger and larger portions of the population will be excluded from higher education, even if participation increases overall.

The second trend is related to changes in the labor market. As the United States and other developed countries continue to transfer economic production into high-end services and high-tech industries, fewer and fewer jobs will be available to individuals without tertiary education. For individuals and communities with inadequate access to higher education, the effects of this trend could be devastating. Generations may be condemned to unemployment

or underemployment, poverty, and social alienation. The social implications are also troubling. Given the demographic changes underway, the United States and other countries face a shortage of skilled workers that will, in the long run, weaken their global economic competitiveness.

Given these social changes, it is imperative that the inequalities in access to higher education at all levels continue to be addressed through research and policy initiatives. Most often, college access is discussed either implicitly or explicitly using a pipeline metaphor, the idea being that there is a single, streamlined channel through which access to higher education flows. As more students proceed through the pipeline without leaking out, more students will arrive at the end-point of higher education. In this metaphor, colleges and universities are represented as isolated points of entry that stand at the end of a linear path to access. However, colleges and universities are not isolated points of entry but are rather institutions situated as part of a spatially segregated education system in which the structures of opportunity are not the same in all places. That is, the pipeline metaphor does not adequately account for the fact that not all places present clear entrance to a channel that leads to college and university access. Since large portions of many communities are unable to access the pipeline, it is important to reconceptualize college access in a way that allows for a better understating of how many communities are largely excluded.

This article explores how space and place may be related to access to higher education and discusses ways in which to study these issues. The next section reviews some of the most influential ideas in college-access research and discusses how they are related to space and place. The third section introduces a set of technology, data, and methods known as geographical information systems (GIS) that are used in geospatial research. The fourth section uses three examples to show how GIS can be used in college access research. The concluding section presents implications for college-access research.

College Access, Space, and Place

Over time, research on college access has developed into a corpus of knowledge that is sophisticated and nuanced. Distilling this scholarship inevitably obscures its richness; nevertheless, three main factors that contribute to college

access have emerged: (1) finance; (2), college knowledge; and (3) academic preparation.

Finance is one main factor relating to college access. This has to do with the ability to pay for the increasing cost of attending college, the availability of financial aid to subsidize students who cannot afford the cost of higher education, and the opportunity cost of attending college in the form of foregone earnings. During the last few decades, the cost of attending college has increased substantially in real terms. Coupled with increased costs, there has been a shift in the provision of federal financial aid from primarily grants to mostly interest bearing loans and a shift in institutional and state sources from need-based aid to merit scholarships. As a result, the cost of college has become out of reach for low-income students, and the aid available to meet this gap has become less responsive to the needs of low-income students.

A second important factor for college access has to do with possessing information about the importance of higher education, as well as social and cultural knowledge that is useful for gaining admission. Much of the access literature on this topic is informed by Bourdieu's concepts of cultural capital, social capital, and habitus. For the sake of conceptual parsimony, this family of concepts will be referred to here as college knowledge. Simply put, students whose family members and acquaintances have attended college are more likely to know the rules of the game relating to college access and social mobility in general. Many students with this background are prepared to attend college from a young age. Students who do not come from a community where college attendance is the norm are less likely to possess the tacit knowledge that is useful for gaining access to higher education. For these students, the college-admissions process can appear mysterious and overly complicated.

A third factor related to college access has to do with ability and academic preparation. Students who are well prepared for college at the primary and secondary levels of the educational system stand in good stead for both access to and success in the tertiary level of education. In practical terms, this means having a strong primary educational background, having taken a rigorous academic program in secondary school as well as having prepared for and taken college entrance examinations such as the Scholastic Aptitude Test (SAT) and American College Test (ACT). Students who are not adequately prepared for higher education not only have greater difficulty gaining access because of a weaker academic record but also because they are simply less prepared for the admissions process. If these students do enrol in higher education they are more likely to enter higher education at 2-year colleges and less likely to persist to earn baccalaureate degrees.

Research on college access has shown that these factors are interrelated. Students who come from middle and

high income families are likely to have parents or relatives who have attended college and pass on college knowledge, and are also more likely to be well prepared academically. Similarly, students who come from low-income families are likely to attend schools with fewer resources and are less likely to know the college-access rules of the game. Society is stratified in a vertical hierarchy by socioeconomic status and race/ethnicity. As a result, low-income and minority students are often disadvantaged in terms of access to higher education.

Scholars of college access have not fully explored how factors related to college access are distributed horizontally across space. American society has a long history of segregation (both racial and economic) and despite some efforts in the past to redress this, evidence suggests that US society is becoming more racially segregated. This means that people within cities and towns are not evenly dispersed but are rather isolated in highly concentrated communities of people of color and with low incomes. As a result of both vertical stratification and horizontal segregation, individuals face an uneven structure of opportunity depending on where they live. In some communities, or neighborhoods, the pathway (or pipeline) to college access is well known and straightforward. In others, the pathway to access is idiosyncratic and opaque. One way of exploring the differences in place-biased structures of opportunity is by using geospatial data and GIS.

Geographical Information Systems

GIS refer to a set of technologies that include the interface of hardware, software, data, procedures, and networks. GIS are used to store, display and represent, interpret, and analyze geospatial data. In other words, GIS are a group of information systems that store and represent data that are associated with specific places in the real world. GIS are frequently used for research in the natural sciences and are becoming more popular in fields like sociology and education. In the social science, GIS are particularly useful for social-science research that has a spatial dimension such as research that investigates the spatial dispersion of social characteristics and trans-spatial social networks.

It is estimated that 80% of all government policies are informed in some way by geographic data. Spatial data can also inform many of the activities of colleges and universities, which are located in actual social communities and recruit students from real places with notable social characteristics. Potential data sources include national census data, local and regional governmental data, and data already gathered and stored by colleges and universities. All educational data that can be anchored to a place, such as neighborhoods, addresses, and schools (and school service areas) can be use in GIS analysis. Institutional research

and enrolment-management offices at many colleges and universities are already using GIS for a variety of institutional research and marketing purposes.

GIS and Access Research

The value of using GIS for college access research is that they offer the possibility of exploring how social characteristics that are known to be important for access, and actual college-access data, are distributed horizontally across space. This, in turn, allows researchers to produce representations and conduct analyses to help better understand the spatial distribution of college-access opportunity structures. In other words, GIS are a set of tools that can be used to explore how horizontal segregation affects college access in addition to vertical stratification. The remainder of this section is organized as follows: (1) data sources and methods are explained; (2) examples of the sorts of representations and analyses that are possibly using GIS for access research are presented and discussed; and (3) suggestions are given for how GIS can be used to inform research and policy regarding college access.

Data sources and methods

All geospatial data are connected to real places in the world. Thus, in order to provide examples for the use of GIS in college access research, a real and identified place must be used. Tucson, Arizona, the city in which the authors reside, has been selected to demonstrate these ideas. We recognize that colleges and universities are not stand-alone points of entry but rather are parts of a system of education that includes K–12 schools and are situated in real communities in which people live, work, and learn. Hence, community, school, and college access data are used in our analyses. These data come from three sources.

The first data source is the United States Census Bureau. The decennial census gathers aggregated population and socioeconomic data, including race/ethnicity, educational attainment, and income, for all communities in the United States. Census data are aggregated by spatial units called census blocks, which, in urban areas, are roughly the size of a city block. There are limited data available for individual blocks because their small size creates the possibility of identifying individuals. However, data are publicly available for block groups and larger spatial units comprised of several block groups known as census tracts. Additionally, the Census Bureau provides spatial data for the census units so that demographic data can be plotted over space and represented visually.

The second source of data is the state of Arizona's Department of Education, which makes publicly available annual data for all public and charter schools in the state. These data include information on enrolments by race, school performance ratings, graduation rates, and so forth. As school addresses are provided, it is possible to georeference these

data. The third source of data is the University of Arizona Office of Enrolment Management. The Office of Enrolment Management keeps track of information like the number of applications, admissions, and enrolments by school and race/ethnicity.

These data were used to produce seven maps of Tucson, Arizona, that represent social and educational attributes distributed across its space. In producing these maps, census-tract data were layered with school data. Census tracts are areas represented by polygons, while schools are specific places represented by points. Attribute variables used for analysis included population demographics (race/ethnicity, income, and educational attainment) and school data (enrolment, school performance, and college applicants). These variables were categorized using Jenks' natural breaks, which is a technique that allows researchers to set a desired number of categories and then order data within these categories by minimizing within-group sum of squared differences:

$$SSD_{i,j} = \sum_{k=1}^j (A[k] - \text{mean}_{i,j})^2$$

Jenks' natural breaks-classification scheme determines the best arrangement of values into groups according to natural groupings within the data. Five categories were used for all map variables. Categories are represented using color codes and feature sizes. Visual analysis was then used to interpret the maps.

GIS examples

Using seven maps, this section provides some basic examples of how GIS data and analysis can be used for college access research. The first example (**Figures 1, 2, and 3**) shows the pattern of racial segregation in schools and neighborhoods in the city of Tucson. Access to higher education in the United States is stratified by race and ethnicity. Minority students, especially African Americans, Latinos, and Native Americans, are less likely to attend college than their White counterparts. When these students do attend college, they are more likely to attend 2-year community colleges and are less likely to earn a baccalaureate degree. Furthermore, research suggests that attending a racially diverse school can benefit all students but that American communities and schools are becoming more racially segregated.

Figures 1 through 3 show racial segregation by neighborhood using 2000 census-tract data. The lighter-colored areas represent neighborhoods with a lower percentage of White residents, while the darker-colored areas represent neighborhoods with a higher percentage of White residents. Census data are overlaid with school-enrolment data from the 2002/3 school year; the closest period to the census for which these data are publicly available. **Figure 1** shows elementary-school enrolment in the city of Tucson. Each circle represents the location of an elementary school.

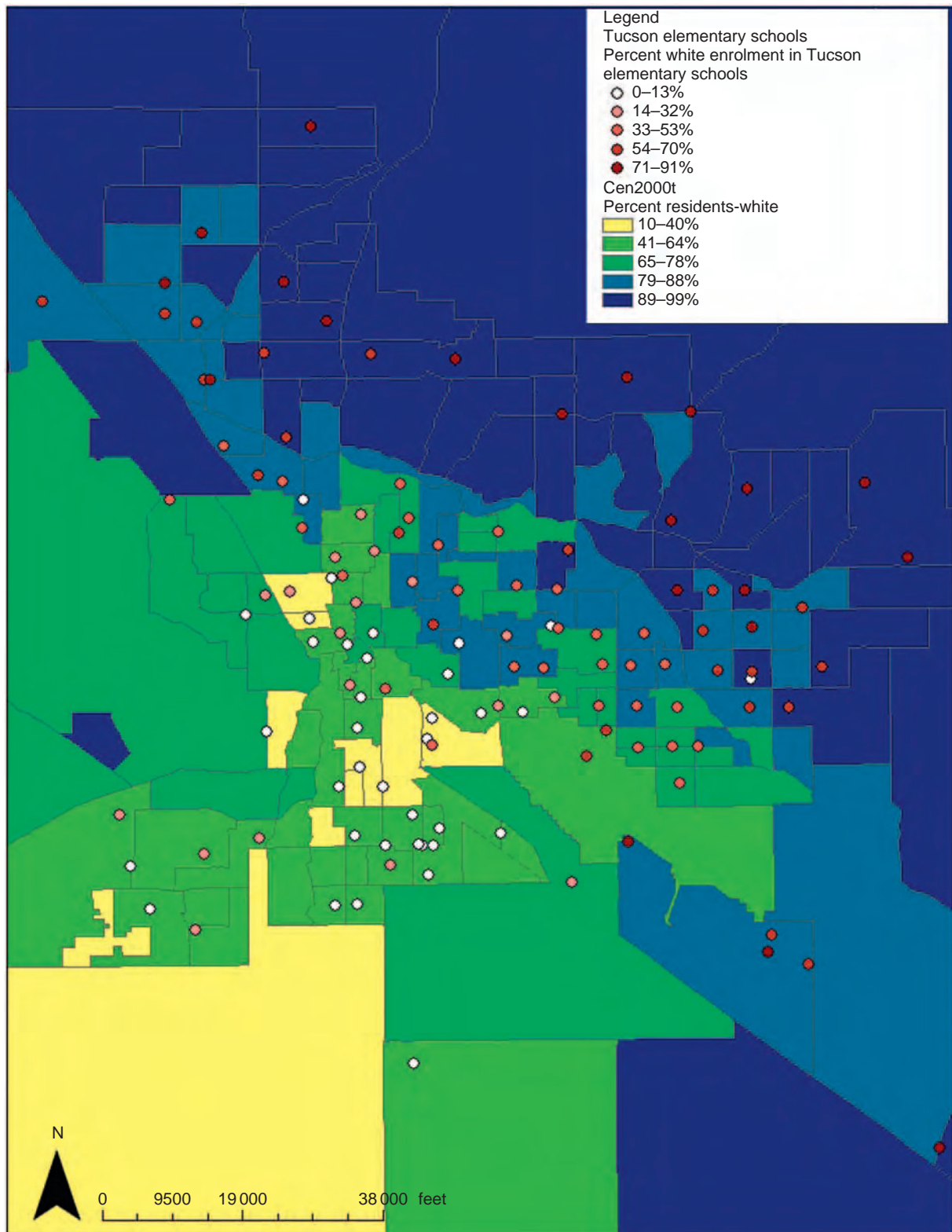


Figure 1 Tucson elementary schools.

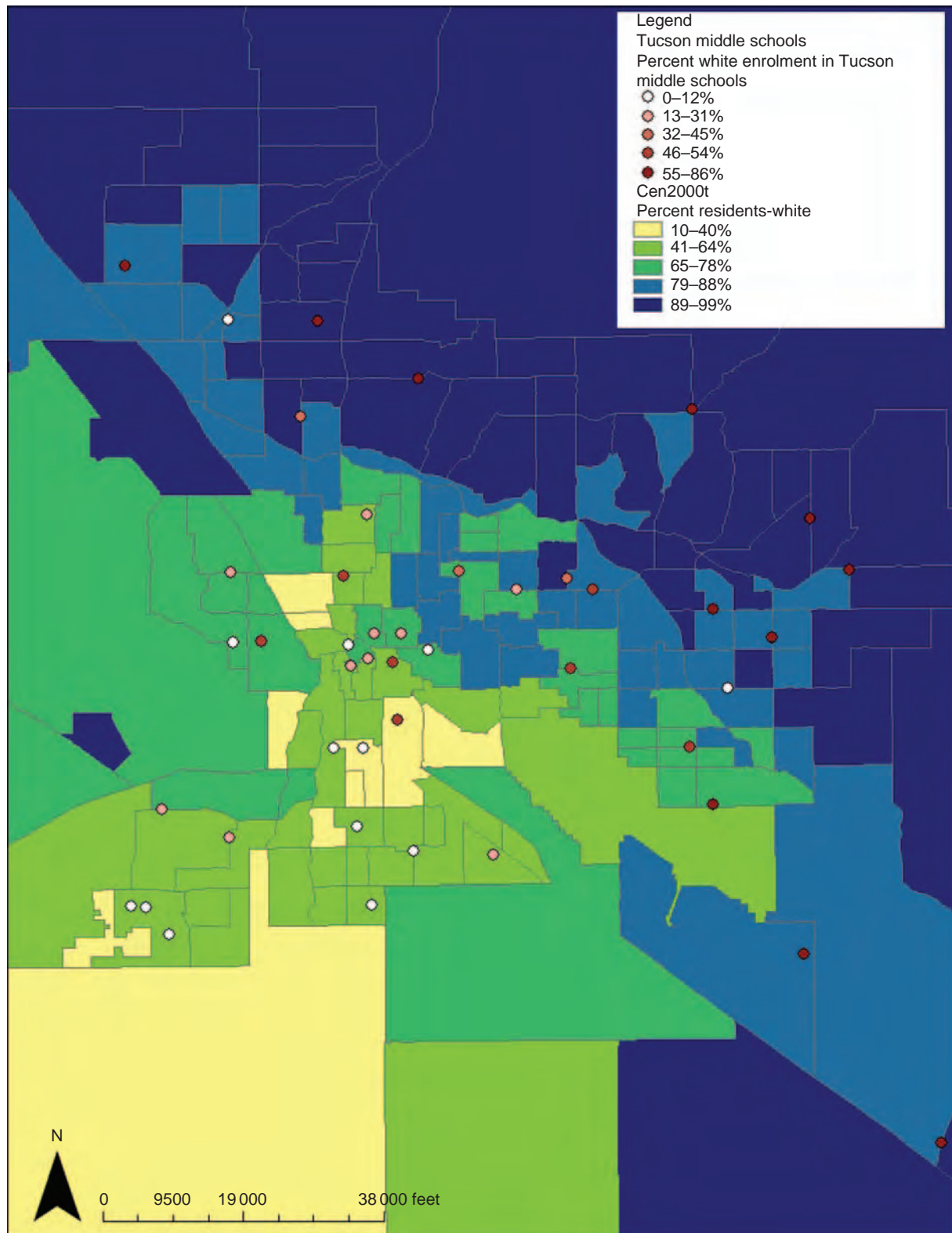


Figure 2 Tucson middle schools.

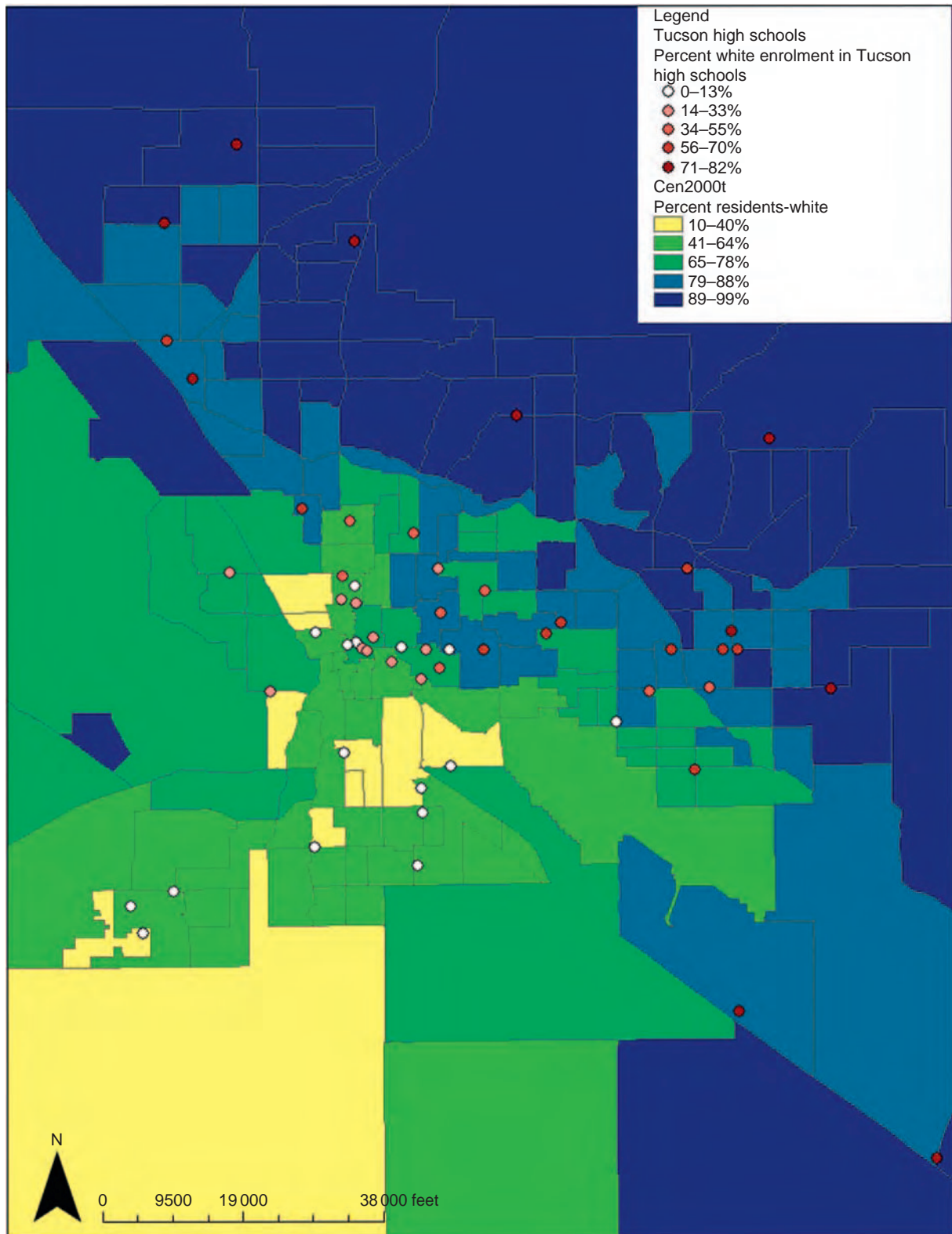


Figure 3 Tucson high schools.

The larger circles represent schools with a larger percentage of White enrollees and the smaller circles represent schools with a smaller percentage of White enrollees. **Figures 2** and **3** represent similar data for Tucson's middle schools and high schools, respectively.

Collectively, **Figures 1, 2, and 3** show that both Tucson's neighborhoods and schools are highly segregated by race. A lower percentage of residents in the south and east of the city are White (primarily Latino) and the cities' northern and eastern neighborhoods are occupied by a higher percentage of White residents. The same is true for schools. These maps show that high White-enrolment schools are located in the north and east while low White-enrolment schools are located in the south and west of the city. In central Tucson, where census tracts tend to be more racially diverse, there are schools with both a high percentage of White enrollees and a low percentage of White enrollees. This trend is most pronounced among elementary schools, which are more numerous and tend to have the smallest school-service areas. This suggests that even within diverse census tracts, communities are segregated by race, which results in segregated schools, especially at the elementary level.

Using the natural-breaks method, features (in this case census tracts and schools) are broken into categories where there are large jumps in values for the variable of interest. For the schools, categories are fairly evenly distributed across the range of the population. Take for example elementary schools, which range from 0% to 91% White enrolment. The natural-break categories for the schools cover about 15–20% points of White enrollees. Natural breaks for the census tracts, however, are clustered at the high end of the range (1%–99%) and the low category covers around 40% points; yet, the high category covers only about 10% points. This suggests that schools may be more segregated than neighborhoods. It also suggests that shifting demographic trends are coupled with increased spatial segregation in the southwest, which is the fastest-growing region of the country. Diverse groups are migrating into the area and their children are attending public schools. Moreover, while undocumented immigrants are unlikely to be captured in the census data, their children are likely to be captured in school-enrolment data. These enrolment data suggest that the overall population is becoming more diverse but that it is also more segregated across space. Thus, school enrolment may be a vanguard to large demographic trends of increased racial and ethnic diversity and increased racial segregation in southwestern cities like Tucson.

The second example shows the relationship between elementary school performance and neighborhood economic and racial composition (**Figures 4** and **5**). Academic preparation is an important factor related to college access and not only refers to individual student's performance but also to the educational resources available to students in their

schools and communities. Schools with more resources, on average, provide better opportunities for college access to their students. In this example, 2003 high school performance ratings, determined by the Arizona Department of Education, are used as a proxy for varying levels of school resources and academic preparation.

Figure 4 is a map of Tucson using 2000 census-tract data that shows median household income and elementary school performance levels. The darker census tracts represent neighborhoods with higher median household income while the lighter tracts show less-affluent areas. The schools represented by darker points represent excelling and highly performing schools and the lighter points represent performing and underperforming schools. The north and east parts of Tucson are more affluent than the central and southern parts of town. As discussed above, the northern parts of Tucson also have a higher percentage of White residents. Visual analysis suggests that there is at least a loose relationship between median household income and school performance. Ten of the 13 excelling schools are located in census tracts that are among the top-three income categories while 13 of the 15 underperforming schools are located in census tracts among the bottom-two income categories. None of the excelling schools are located in census tracts in the lowest-income category and no underperforming schools are located in the census tracts with the highest median incomes.

Figure 5 is a map showing the same school-performance data overlaying census tracts categorized by the percentage of residents who are White. This map shows that while household income is related to school-performance levels, race is an important factor as well. Most of the excelling schools are in census tracts with a high percentage of White residents and most underperforming schools are in census tracts with a lower percentage of White residents. Hence, comparing these two maps (**Figures 4** and **5**) shows that, with some exceptions, there is a strong relationship between the racial composition of the neighborhood and medium household income in that many of the non-White census tracts have a lower median household income. When underperforming schools are located in higher-income neighborhoods, these neighborhoods also tend to have a relatively low proportion of White residents. The three excelling schools in lower-income neighborhoods are in census tracts that are among the top three categories in terms of the percent of residents who are White. There are no excelling schools in census tracts in the bottom-two White resident categories. This means that White residents account for at least 64% of the total population in every census tract that contains an excelling school. These maps (**Figures 4** and **5**) show that families and students who live in low-income neighborhoods and neighborhoods with a significant population of people of color are more likely to attend elementary schools that are marginally performing or underperforming.

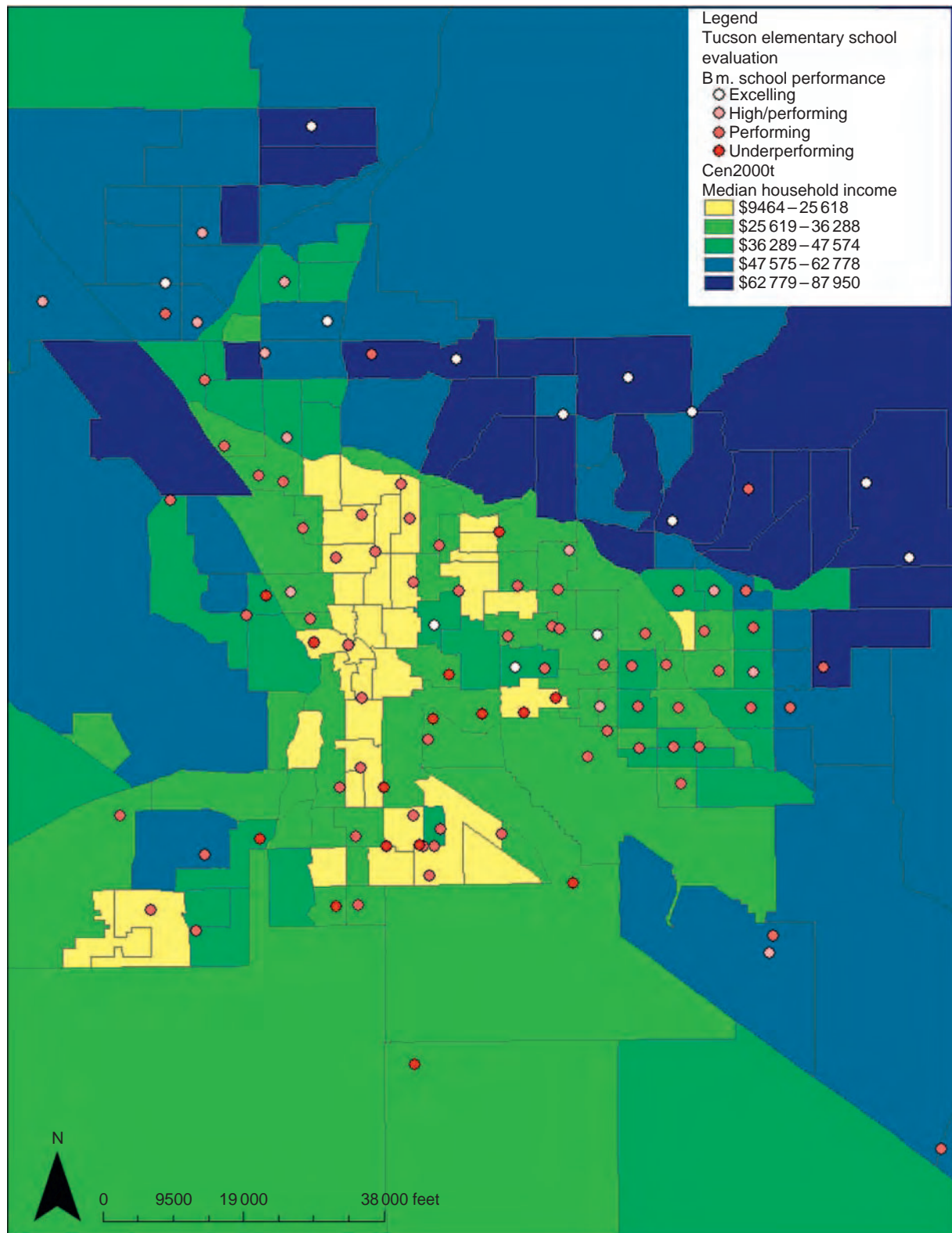


Figure 4 Tucson elementary school evaluation.

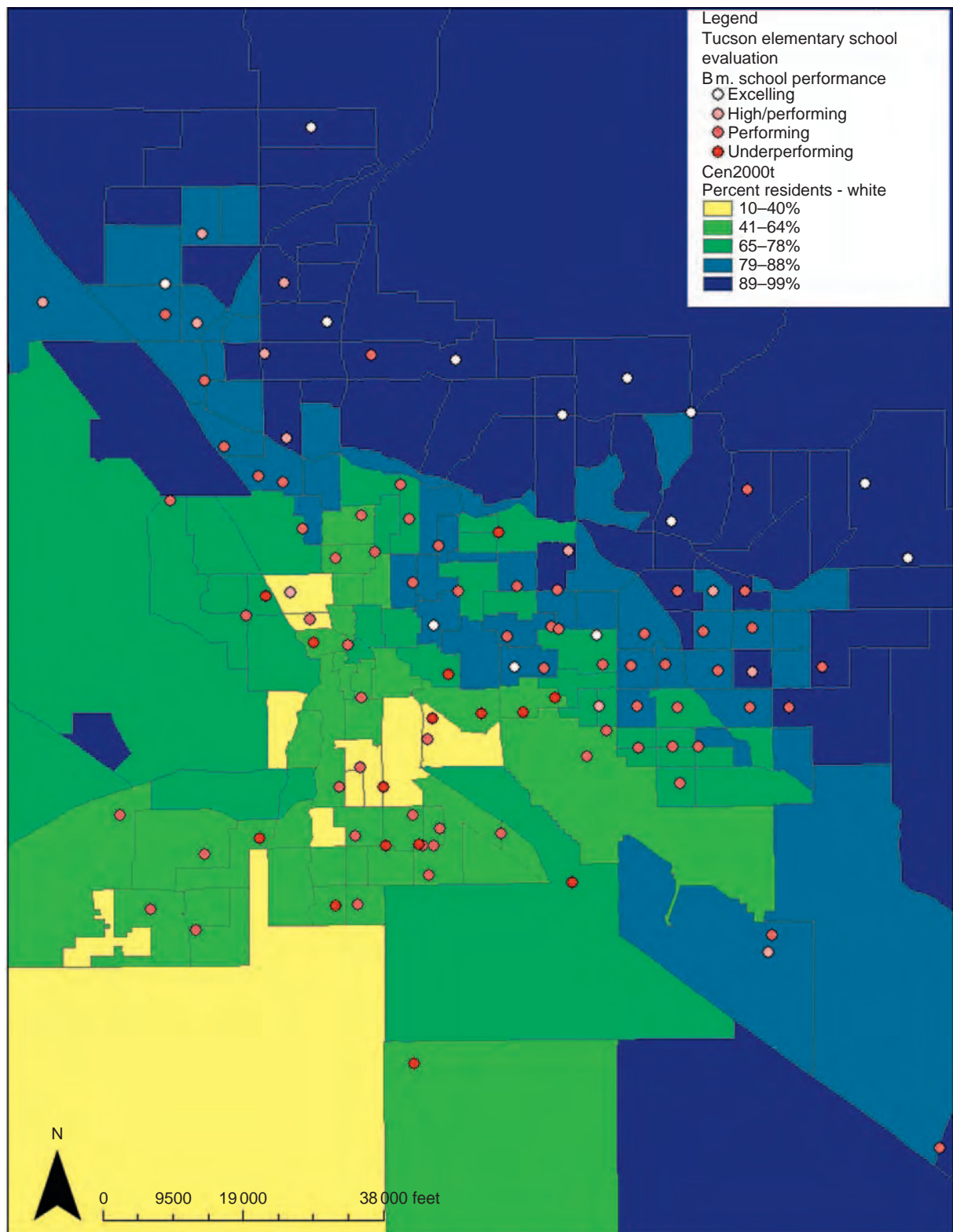


Figure 5 Tucson elementary school evaluation.

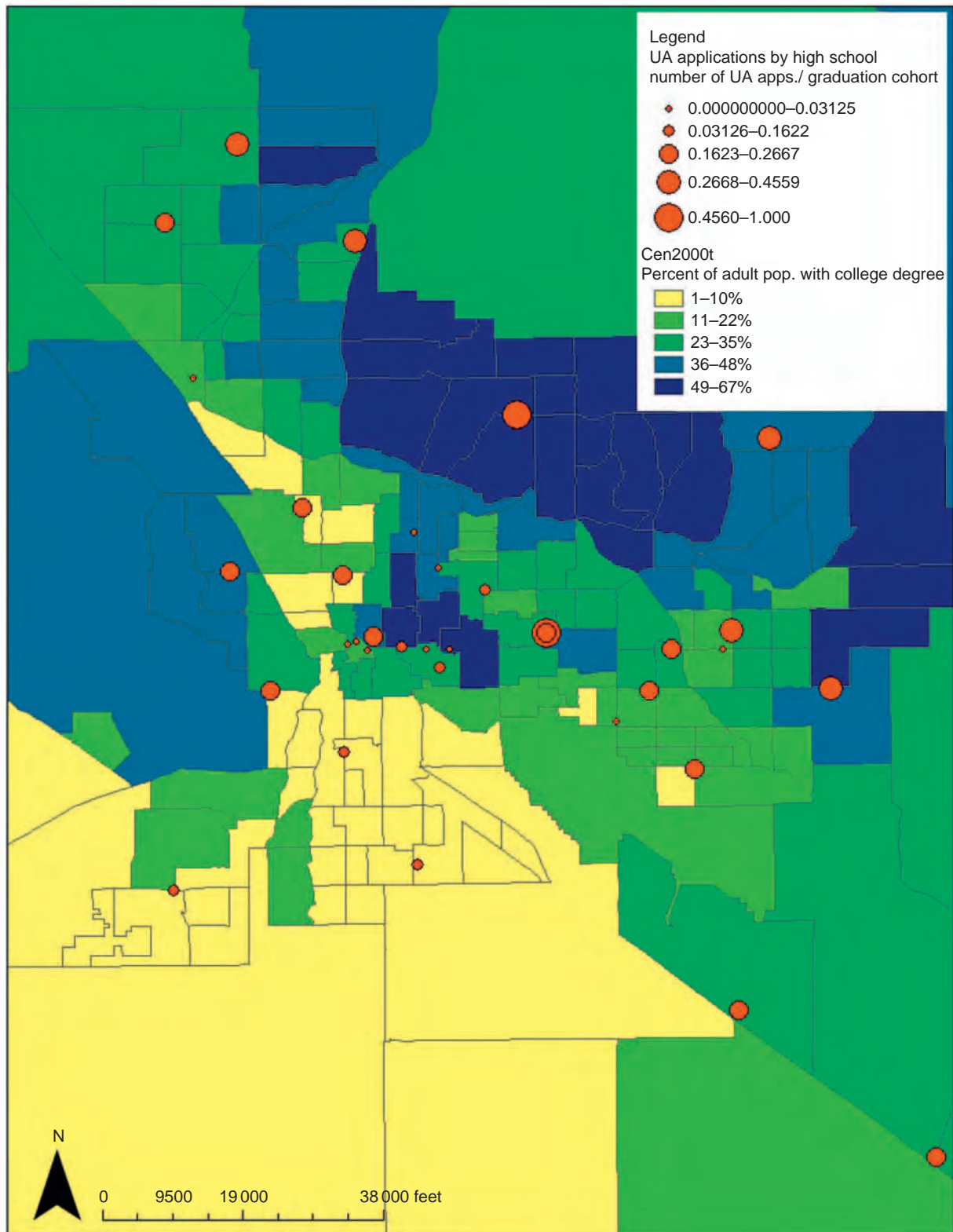


Figure 6 University of Arizona (UA) applications by high school – number of high school applications/graduation cohort.

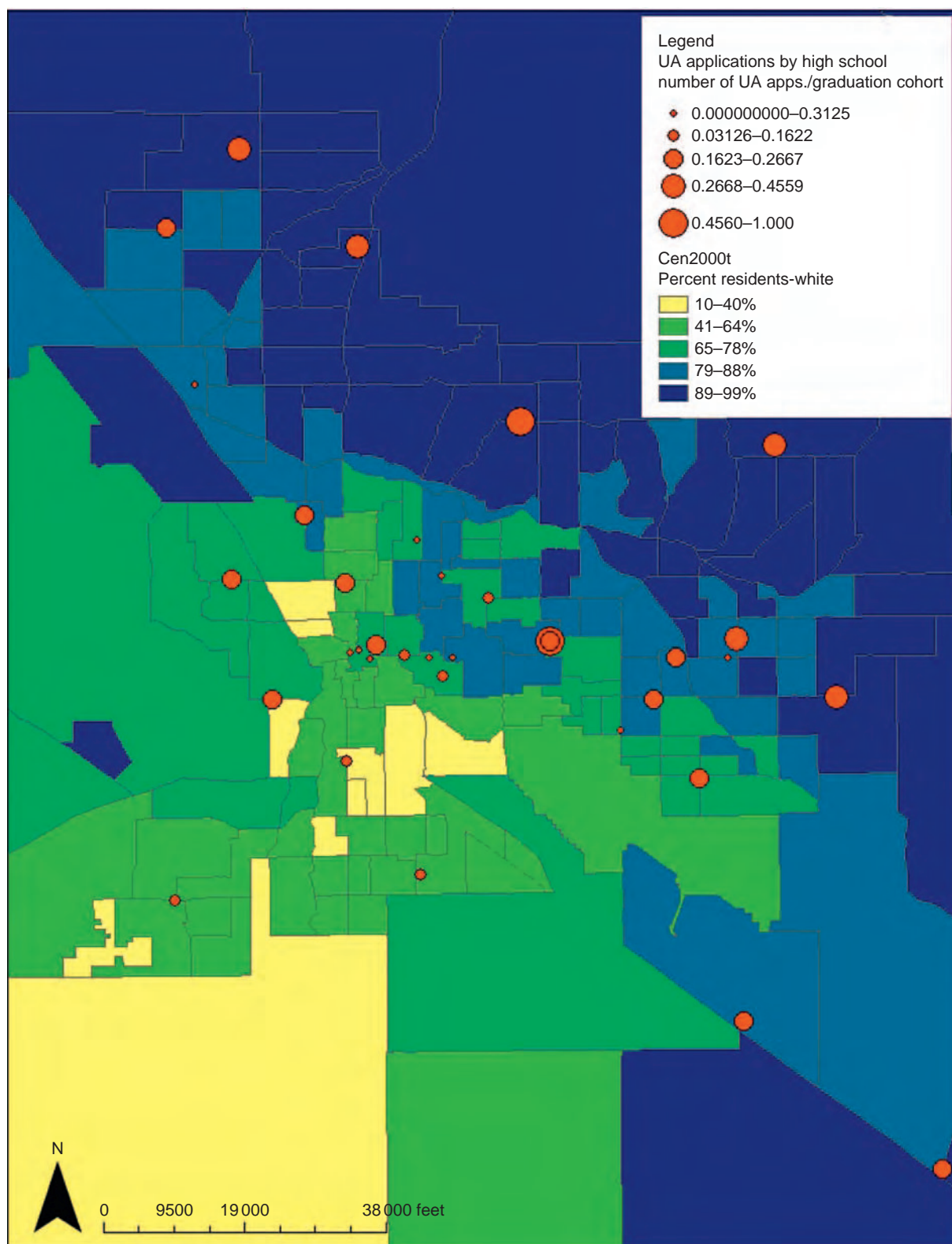


Figure 7 University of Arizona (UA) applications by high school – number of high school applications/graduation cohort.

The third example explores the relationship between neighborhood college knowledge, neighborhood racial/ethnic composition, and college-going practices (Figures 6 and 7). In Figure 6, college knowledge is represented using 2000 US census-tract data on the percentage of residents 25 years and older who have completed at least a baccalaureate degree. Darker census tracts represent neighborhoods that are comprised of a higher percentage of college-educated residents (assumed to have more college knowledge) and lighter census tracts represent neighborhoods that are comprised of a lower percentage of college-educated residents (assumed to have lower college knowledge). College-going practices are estimated by looking at the ratio of applications submitted to the University of Arizona with the size of the graduating cohort at each high school in Tucson for which there were data from the 2006/7 school year. Schools that exhibit high levels of college-going behavior are represented by larger points.

Figure 6 shows that neighborhoods in the northern parts of the city have higher levels of college knowledge and that schools in these neighborhoods exhibit higher levels of college-going behavior. Likewise, neighborhoods in the southern part of the city have lower levels of college knowledge and schools in these neighborhoods exhibit lower levels of college-going behavior. Figure 7 shows that in northern and southern neighborhoods, college knowledge and college-going behaviors tend to be closely associated with the racial composition of the neighborhoods. Census tracts with a high percentage of White residents in these neighborhoods tend to have relatively high levels of college knowledge and college-going behavior while non-White neighborhoods have lower levels of college knowledge and college-going behavior.

These patterns do not hold completely true in the central part of Tucson, which tends to be more diverse in terms of both race/ethnicity and socioeconomic status. This may be because in these more-mixed neighborhoods, college knowledge is more closely associated with the individual household than the community. It may also have to do with the way college-going behaviors are measured. Only applications to the University of Arizona are counted toward college-going behavior. Since census tracts in central Tucson include both higher income and predominantly White neighborhoods as well as lower income and more diverse neighborhoods, and because these different neighborhoods may feed in to the same high schools, the relationship between college knowledge and college-going behaviors in this area of the city is less apparent at the level of data granularity presented here.

Discussion of the three examples

Overall, the three examples and seven maps presented show that Tucson, Arizona is socially segregated along the lines of race/ethnicity and socioeconomic status and that this segregation has created unequal structures of opportunity

for college access. Students in affluent, White neighborhoods with higher levels of college knowledge have access to higher-performing schools and exhibit high levels of college-going behavior than students from less-affluent neighborhoods with a higher proportion of minority residents. Additionally, schools and neighborhoods in Tucson are racially segregated, depriving all students of the cognitive and social benefits of diversity that could lead to higher levels of college access for all. Given demographic and social trends underway, these examples show the importance of exploring place-biased opportunity structures in college-access literature rather than relying on pipeline modes for exploring educational opportunity.

Implications for educational research

Clearly, there are limitations to the analyses that we did in this study. Census-tract data do correspond directly to school-service areas which can provide imprecise representations of the relationships that we depict in this study. However, through the use of more advanced statistical techniques and methods of spatial analyses, it is possible to more accurately and appropriately represent school-service areas which can provide a much more precise picture of the nature of these important relationships.

We think that the use of GIS tools can make several important contributions to the study of college access and other educational phenomena related to important social and economic characteristics. As can be seen in the analyses described in this article, GIS tools can be used to create helpful visual representations that can be used as cognitive maps depicting important and often complex patterns of relationships. In these analyses, we used GIS tools to examine the relationships between segregation, stratification, and educational opportunity in one US city – Tucson, Arizona. Cognitive maps like the ones that we created can be used to further inform important public policy debates and improve educational decision making.

Further Reading

- Astin, A. W. and Osegura, L. (2004). The declining "equity" of American higher education. *Review of Higher Education* 27(3), 321–341.
- Cabrera, A. F. and La Nasa, S. M. (2000). Three critical tasks America's disadvantaged face on their path to college. In Cabrera, A. F. and La Nasa, S. M. (eds.) *Understanding the College Choice of Disadvantaged Students*, vol. 107, pp 23–29. San Francisco, CA: Jossey-Bass.
- Cabrera, A. F. and La Nasa, S. M. (2001). On the path to college: Three critical tasks facing America's disadvantaged. *Research in Higher Education* 42, 119–149.
- Cobb, C. D. (2003). Geographic methods and policy: Using geographic information systems to inform education policy. *Education research quarterly* 27(1), 28–39.
- Freeman, K. (1997). Increasing African Americans' participation in higher education: African American high-school students' perspectives. *Journal of Higher Education* 68(5), 523–550.
- Garson, D. G. and Briggs, R. S. (1992). *Analytical Mapping and Geographical Databases*. Thousand Oaks, CA: Sage.

- González, K. P., Stone, C., and Jovel, J. E. (2003). Examining the role of social capital in access to college for Latinas: Toward a college opportunity framework. *Journal of Hispanic Higher Education* **2**, 146–170.
- Hamrick, F. A. and Hossler, D. (1996). Diverse information-gathering methods in the postsecondary decision-making process. *Review of Higher Education* **19**(2), 179–198.
- Hamrick, F. A. and Stage, F. K. (2004). College predisposition at high-minority enrollment, low-income schools. *Review of Higher Education* **27**(2), 151–168.
- Hearn, J. C. (1984). The relative roles of academic, ascribed, and socioeconomic characteristics in college destinations. *Sociology of Education* **57**, 22–30.
- Hearn, J. C. (1988). Attendance at higher-cost colleges: Ascribed, socioeconomic, and academic influences on student enrollment patterns. *Economics of Education Review* **7**, 65–76.
- Hearn, J. C., Griswold, C. P., and Marine, G. M. (1996). Region, resources, and reason: A contextual analysis of state tuition and student aid policies. *Research in Higher Education* **37**(3), 241–278.
- Heller, D. E. (2004). NCES research on college participation: A critical analysis. In St. John, E. P. (ed.) *Public Policy and College Access: Investigating the Federal and State Roles in Equalizing Postsecondary Opportunity*, vol. 19, pp 29–64. New York: AMS Press.
- Henry, S. A. (1995). The Topeka experience: Uses of the SDDB and other census and GIS software in district planning. *Paper presented at the annual meeting of the American Educational Research Association*, San Francisco, CA, 18–22 April.
- Horvat, E. M. (2001). Understanding equity and access in higher education: The potential contribution of Pierre Bourdieu. In Smart, J. C. (ed.) *Higher Education: Handbook of Theory and Research*, vol. 16, pp 158–171. New York: Agathon Press.
- Hossler, D., Braxton, J., and Coopersmith, G. (1989). Understanding student college choice. In Smart, J. (ed.) *Higher Education: Handbook of Theory and Research*, vol. V, pp 231–288. New York: Agathon Press.
- Hossler, D. and Gallagher, K. S. (1987). Studying college choice: A three-phase model and the implications for policy-makers. *College and University* **2**, 207–221.
- Hossler, D., Schmit, J., and Vesper, N. (1999). *Going to College: How Social, Economic, and Educational Factors Influence the Decisions Students Make*. Baltimore, MD: Johns Hopkins University Press.
- Hossler, D. and Stage, F. K. (1992). Family and high school experience influences on the postsecondary educational plans of ninth-grade students. *American Educational Research Journal* **29**, 425–451.
- Hossler, D. and Vesper, N. (1993). An exploratory study of the factors associated with parental saving for postsecondary education. *Journal of Higher Education* **64**(2), 140–165.
- McDonough, P. M. (1994). Buying and selling a college education: The social construction of the college applicant. *Journal of Higher Education* **65**(4), 427–446.
- McDonough, P. M., Antonio, A. L., and Trent, J. W. (1997). Black students, Black colleges: An African American college choice model. *Journal for a Just and Caring Education* **3**, 9–36.
- Perna, L. W. (2000). Differences in the decision to enroll in college among African Americans, Hispanics, and Whites. *Journal of Higher Education* **71**(2), 117–141.
- Perna, L. W. (2006). Studying college access and choice: A proposed conceptual model. In Smart, J. C. (ed.) *Higher Education Handbook of Theory and Research*, vol. 21, pp 99–157. Amsterdam: Kluwer.
- Perna, L. W., Steele, P., Woda, S., and Hibbert, T. (2005). State public policies and the racial/ethnic stratification of college access and choice in the state of Maryland. *Review of Higher Education* **28**(2), 245–272.
- Perna, L. W. and Titus, M. (2004). Understanding differences in the choice of college attended: The role of state public policies. *Review of Higher Education* **27**(4), 501–525.
- Perna, L. W. and Titus, M. (2005). The relationship between parental involvement as social capital and college enrollment: An examination of racial/ethnic group differences. *Journal of Higher Education* **76**(5), 485–518.
- Teodorescu, D. (ed.) (2003). *Using Geographic Information Systems in Institutional Research: New Directions for Institutional Research*, pp 120. San Francisco, CA: Jossey-Bass.
- USGS (2005). Geographic information systems in education. http://rockyweb.cr.usgs.gov/outreach/gis_gps_rs_in_education_fs.pdf (accessed September 2009).

Public Policy and Inequality in Postsecondary Opportunity: Educational Statistics and the Failure of Education Reform

E P St. John, K Williams and K Moronski, University of Michigan, Ann Arbor, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Academic preparation – A high school curriculum that includes the advanced courses in math, science, and English necessary for enrolment in 4-year colleges.

Accountability – A government system of regulation of schools and/or colleges that requires compliance with specific mandates.

Basic Educational Opportunity Grants (BEOGs) – The original name of Pell grants when the program was created by the Education Amendments of 1972.

Campus-based aid – College work study, national direct loans and supplemental educational opportunity grants are awarded to colleges, based on specific criteria. Campuses award aid to students based on financial need. Campus-based programs were created by the Education Amendments of 1972, which reauthorized the Higher Education Act (HEA).

College access – The rate of enrolment in college by a population or subgroup. Usually measured for the traditional age group (18–24) or the percentage of high school graduates going on to college the year after high school.

College success – Students finishing college with majors consistent with their interests. Simple measure in college-completion rates, a measure sometimes used in state accountability systems for higher education, including some budgeting systems.

Cost of attendance – The sum of tuition, books, living costs, and other expenses for attending a college or university.

Educational Opportunity Grant (EOG) – The original federal, generally available need-based grant program, now named supplemental educational opportunity grants.

Education Amendments of 1972 – The reauthorized version of the HEA that created BEOG (now Pell) and reorganized supplemental educational opportunity grants, work study, and national direct loans into the campus-based programs.

Expected family contribution – The estimated amount a family can pay for college, taking into

account parental income, number of siblings in college, savings, and other assets.

Federal loans – The student loan programs administered by the federal government and authorized under Title IV of the HEA.

Financial aid – The grants, scholarships, loans, and work study awarded to college students based on financial need, measures of merit (e.g., high school grades or test scores), and/or special abilities (athletic ability, music talent, etc.).

Financial inequality – The inequality in opportunity attributable to income differences. In an equitable society, education should be equally available to all groups regardless of income or race.

Financial need – The financial need of a student, which equals the cost of college minus expected family contribution.

Great Society – The image of a fair and just America, used by President Lyndon B. Johnson as a rationale for major, new educational and social programs in the 1960s.

Higher Education Act (HEA) – Passed originally in 1965, the HEA all-student-aid programs under Title IV, developing institutions under Title III, and other federal programs under other titles. The act is reauthorized every 4–6 years.

High school (K-12) graduation requirements – The courses required for high school graduation. Historically, the standards for high school graduation were set by local school boards. Since the early 1980s, federal education policy has encouraged states to set graduation requirements.

Merit grants – Aid awards based on student achievement. Example: Georgia Hope.

Middle Income Student Assistance Act of 1978 – A law that reauthorized Title IV of the HEA and removed income limitations on Pell grants.

Need-based grants – Student awards of money that are based on the families' ability to pay for college; for example, Pell grants.

Pell grants – The major federal need-based grant program. A portable form of aid students can take to any institution in which they are accepted. The programs are named after Senator Pell.

Pell maximum – The maximum award for Pell grants in any given year. Pell eligibility is determined by financial need. The Pell maximum is set each year in the federal budget. Since the late 1970s, the actual budgeted Pell maximum has consistently been less than the limit authorized under Title IV of the HEA. Limiting the Pell maximum has become a method Congress uses to control the costs of Pell grants.

Supplemental Educational Opportunity Grants (SEOG) – A federal need-based grant program administered by colleges.

Tuition and fees – The charges colleges and universities collected from students for the education and services provided.

Unmet need – The cost of attendance minus the sum of expected family contribution and financial aid awarded.

Unsubsidized loans – The loans provided by lenders without any federal subsidy. Some programs are federally authorized, such as some family loans, while others are administered by private lenders. Federal subsidies reduce interest rates and insure loans.

Work–loan burden – The amount of cost of attendance after grants and family contributions that a student must cover through work and borrowing. Since some loans are considered need-based aid, the work–loan burden is usually higher than unmet need.

Introduction

For the past quarter century, the federal government has published reports advocating improvements in the nation's high schools. There have also been a plethora of statistical reports emphasizing correlations between high school courses completed and college-enrolment rates. Some have argued that the disparity in preparation explains differences in college-enrolment rates across income and racial/ethnic groups. Not only have publications on the tool box focused on high school math courses completed as a predictor of college completion, but reports by the National Center for Education Statistics (NCES) have argued that all students who take the right courses and who apply for college have the opportunity to enrol. These reports have been used by the National Governors Association to argue that states should raise their standards for high school graduation to a level equaling college preparation, including at least algebra II.

A counter argument advanced by the Advisory Committee of Student Financial Assistance (2002) is that increased funding of federal Pell grants, the primary federal program promoting equal opportunity for college

enrolment, is necessary to reduce inequalities. This counter-interview to the dominant rationale of academic preparation has its origins in an earlier period of public policy, the Great Society programs of President Johnson. The federal role in K–12 and higher education developed as part of the initiatives in the 1960s aimed at reducing inequality in educational opportunity. Interestingly, a few recent reports like that of the Spelling Commission have promoted increases in need-based grant aid along with higher educational standards and increased accountability for higher education.

This article examines changes over time in public policy on college access along with trends in two key indicators related to college access: academic preparation and enrolment rates. Our focus is on how policies relate to changes in college enrolment rates and differentials in these rates. Our approach differs fundamentally from the one used by analysts who identify correlates of college success, like high school math courses, coupled with building rationales for policies that require these correlates. Instead, our focus is on how these policies, based on flawed rationales, actually relate to changes in enrolment and equality in educational opportunity.

Public Policies Promoting College Access

Federal policies promoting equal opportunity in college access began in earnest with the passage of the Higher Education Act of 1965, which funded the first generally available grant and loan programs under Title IV, Educational Opportunity Grants. In 1972, the Pell Grant Program replaced the Educational Opportunity Grant (EOG) program with Basic Educational Opportunity Grants (now called Pell Grants). The grant programs of the 1960s and 1970s were implemented during a period of increasing access attributable to the growing size of the college-age cohort and college expansion to accommodate the baby-boom generation. The rise and decline of Pell Grants took place in a decentralized system of higher education in which states had the primary responsibility for academic preparation through K–12 graduation requirements as well as funding for higher education. As background before reviewing key enrolment trends, we examine trends in Pell grant funding relative to college costs, state policies on high school graduation, and coordination of state funding for grants and tuition.

Federal Pell Grants

Recently, several leading economists have argued that reinvesting in Pell grants provides a means of increasing social justice and promoting economic development, restating an argument that had a substantial influence on the early development of federal grant programs. In the

1960s and early 1970s, in the wake of *Brown versus Board of Education*, the inequality in educational opportunity was used to argue for federal student financial-aid programs. These earlier arguments about inequality were often coupled with arguments about economic development in advocacy of federal grant programs. The Higher Education Act (HEA) of 1965, which initiated the contemporary federal role in student financial aid, was one of Lyndon B. Johnson's Great Society programs aimed at remedying racial inequality. The reauthorization of the HEA in 1972 created Basic Educational Opportunity Grants (BEOGs) (now Pell Grants) as a portable grant program intended to equalize post-secondary educational opportunity and increase enrolment. (For a more complete discussion of the history of federal student aid see St. John (2003).)

Initially, the maximum Pell grant was higher than the costs of attending public 4-year colleges, but there was a half-cost provision in the original legislation which limited the maximum award to half of the total costs of attending. Students attending more expensive private colleges could get the full grant and, in these cases, Pell grants enabled low-income students to choose more expensive colleges. Initially, program eligibility was limited to low-income students, but in 1978 the Middle Income Student Assistance Act extended the eligibility to include middle-income students, removing income caps in favor of using unmet need as the indicator of eligibility. (Unmet need equals the cost of attendance (COA) minus expected family contribution (EFC). Early in the history of the Pell program, grants often met need after EFC, but this is no longer the case. Usually low-income families have substantial unmet need after total grants, causing excess borrowing and work.) By the early 1980s, the cap for the Pell program had also dropped.

Trends in the Pell award maximum compared to the average COA at public 4-year colleges and universities (Figure 1) reveal that college costs increased much faster than the Pell maximum award after 1975. The maximum Pell award actually declined in constant dollar value between the fall terms of 1975 and 1995, while the costs of attending public colleges rose. Although the dollar value of the maximum Pell award rose slightly after 1995, the increase did not keep up with the rising costs of public colleges. Consequently, the gap between the Pell grant maximum and the cost of attending 4-year colleges continued to increase. Other forms of federal aid did not increase, so while many colleges had other forms of federal grants to award low-income students, the total amount of federal aid declined in relation to tuition. (Supplemental Education Opportunity Grants (SEOGs) are awarded to only some public and private colleges as a result of historical grant applications. It is very difficult for a new or growing institution to increase its pool of money through SEOG.)

The gap between the Pell maximum and COA at public colleges and universities, the amount of unmet need for the highest-need students, is filled by a combination of loans, earnings from work, and other types of grants if available. Only a few states have sufficient need-based grant aid to substantially reduce the gap, so low-income students in public colleges and universities often had work/loan burdens in excess of US\$8000 in 2005. The growing cost of attending public 4-year colleges has added to inequality in financial opportunity for low-income students to enrol in 4-year colleges (Fitzgerald, 2004), a fact too often ignored in arguments which posit that improvements in high school curricula and additional academic requirements for graduation will solve the access challenge in the United States.

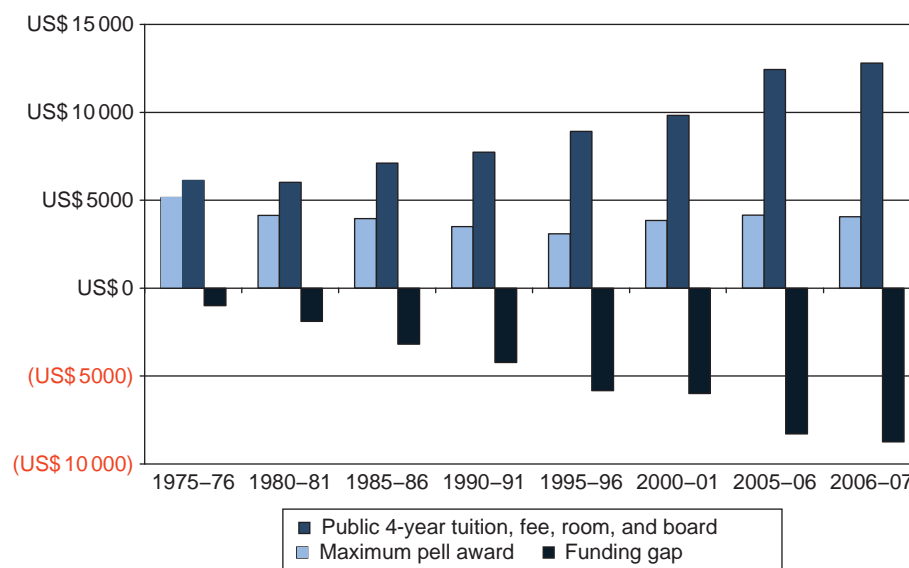


Figure 1 The Gap between the actual maximum Pell award and university attendance costs.

State Policies on Academic Preparation

The publication of *A Nation at Risk* (US Department of Education, 1983) emphasized improving high school graduation requirements in the United States and advocated other reforms related to standards, curriculum, and assessment. Since then, the US Department of Education has published and republished various papers on the correlation between completion of high school math and college-success indicators, from initial enrolment through college completion. These correlations have been used to advocate for adding requirements for high school graduation, echoing a rationale advanced in *A Nation at Risk*. Not only has there been nearly continuous advocacy for increasing requirements along with raising standards and using high-stakes tests, but many of these policies have been implemented in most states since 1990 (see Table 1). For example, more states now require students to pass exams to graduate from high school (rising from 15 states to 22).

The most visible indicator of change, however, has been in math requirements, a development that appears to have had an influence on math achievement. (This section summarizes a discussion of school reform in St. John (2006), updated by the Promoting Equity in Higher Education Project, part of the University of Michigan's National Center for Institutional Diversity (NCID).) The number of states requiring at least two math courses increased from 11 in 1990 to 28 in 2005. Further, no state required algebra I for high school graduation in 1990, about the time federal reports began to emphasize the correlation between high school math and college success,

but 22 states required this level of math (or higher) for graduation by 2005. In addition, the percentage of students taking the Scholastic Aptitude Test (SAT) and the average scores on the SAT math exam improved over the 15 years, providing a tangible sign of the policy's impact on achievement.

However, while math requirements and achievement rose, high school graduation rates decreased during this period, dropping to about 70%. Further, there is no evidence to support claims that improving graduation requirements is associated with improved enrolment opportunity, a situation that complicates the reality for recurrent arguments about academic preparation. It is necessary to consider the state's role in financing public higher education as background to this aspect of the problem.

State Coordination of Tuition and Need-Based Grants

While states collaborate with local school districts to provide free K–12 education, no such arrangement exists in higher education. As noted in Figure 1, the cost of attendance at public 4-year colleges has grown substantially in recent decades. While the decline in federal Pell grants has eroded access, states continue to have the responsibility for providing equal opportunity for their citizens to receive an education. In theory, raising the minimum standards for high school graduation could improve opportunities for low-income students to enrol in college, but only if adequate need-based student financial aid was available.

Table 1 State policy indicators for selected years, 1990–2005

	1990	1995	2000	2005
<i>Policy-related variables</i>				
State-established content standards in math	7	46	50	50
Requires three or more math courses for graduation	11	12	21	28
Requires one or two math courses for graduation	33	31	24	17
Requires at least algebra I or above	0	2	12	22
High school curriculum is locally controlled	6	7	5	5
Offers an honors diploma	15	17	19	22
Exam required for high school diploma ^a	15 ^b	12	14	19
Percentage of schools participating in AP ^{a,c}	45%	51%	58%	62%
Percentage of students taking SAT ^d	42% ^a	41%	44%	49%
Ninth-grade cohort size (millions)	3.2	3.32	3.79	3.96
<i>Outcomes of interest</i>				
SAT verbal mean	500	504	505	508
SAT math mean	501	506	514	520
SAT combined	1001	1010	1019	1028

^aBased upon numbers reported in 1991.

^bThis number is higher than anticipated but cannot be externally validated.

^cReflects the median percentage for AP and the median dollars per FTE for K–12 expenditures.

^dThese numbers reflect the national figures reported by Educational Testing Service.

^eDollars reported are unadjusted.

From the project on Promoting Equity in Higher Education Project, National Center for Institutional Diversity.

States also provide funding for need-based grants, although the level of state funding per student varies substantially across states, as illustrated in **Figure 2**. Only three states – New York, California, and Indiana – had an average of 20% or more per full-time equivalent enrolment (FTE) funding of grants. (This indicator used the FTE of students in public and private colleges in the numerator and total funding for state grants in the denominator. State funding for grants is not reported with a breakdown by different types of colleges.) These states provide substantial grant aid for students enrolling in private colleges and public colleges. Only 12 other states maintained this ratio at a rate above the national average. Thus, high-need students in the majority of the states have substantial unmet need.

The link between state funding of need-based grants and enrolment rates for college students should not be overlooked. The link is illustrated by the correspondence between grants and the inequality in enrolment across income groups: The years when the Pell maximum was higher than the average cost of attending public 4-year colleges, enrolment rates were nearly equal across racial/ethnic groups. **Figure 3** illustrates that trends in the ratio of state funding for grants and tuition corresponds with the continuation rate (the ratio of high school seniors to college freshmen enrolling in states), a further illustration of the link between funding for need-based grants and enrolment. This correlation between state financing and enrolment rates has received substantially less attention in national policy reports than has the correlation between math courses completed and eventual college completion. This oversight is highly problematic.

Public Policy and Post-Secondary Opportunity

Access to higher education is defined as the rate of enrolment in college. In the 1970s, as noted above, enrolment rates did not increase. Equal opportunity to enrol is defined as equality in access across income and racial/ethnic groups. It is possible to nearly equal opportunity for college enrolment across groups whether access increases or decreases during a specific period of time. At the present time, the extent of equal opportunity depends on whether equally prepared students have the same odds of enrolling in 4-year colleges regardless of their family income. Unfortunately, the work–loan burden after grants for the average low-income student in a 4-year college is excessive, so conditions do not favor equal opportunity for enrolment in 4-year colleges.

With the varied and often competing arguments about the role of public policy in promoting college access, it can be difficult to untangle how policy impacts educational

outcomes related to college access. Instead of arguing about correlations – like the ratio of high school math courses to college enrolment or completion – we compare trends in policy (above) to trends in critical outcomes related to college access. After more than two decades of arguments for making high school tougher, it is important to consider whether improvements in preparation (as evident in gains in test taking and test scores) actually result in increased opportunity or reduction in inequality across groups.

Our analyses focus on three key indicators of access to higher education:

- *Access*. The percentage of 18–24-year-olds who indicated enrolment in college (current or previous) and college continuation rates.
- *Inequality of opportunity*. The differentials in rates of college enrolment for African American and Hispanics compared to Whites.
- *College choice*. The percentage of a racial/ethnic group's FTE enrolment in public 2-year and public 4-year colleges, as ratios to the group's percentage in the population, as indicators of whether different groups choose different types of colleges.

These definitions are interrelated, but it is important to distinguish between these outcomes when equity is being considered. Equal access can be realized in a society that has equal rates of college-prepared students enrolling in 2- and 4-year colleges. However, there can still be unequal opportunity if low-income students who are prepared to enrol in 4-year colleges cannot afford to enrol in 4-year colleges. College choice extends the concept of equal opportunity to consider the range of choices, from open-access colleges through elite private colleges. Ideally, measures of choice would consider preparation; this is not only difficult to do, but there are prior inequalities in the opportunity to prepare for college in almost all US states (St. John, 2006). Therefore, we consider the representation of a group in different types of colleges (i.e., proprietary, public 2-year, public 4-year, and private non-profit colleges) as an indicator of whether equity in colleges choice has been realized. This measure implicitly considers the role of college cost and prior preparation. In a society with equal opportunity to prepare for college and fair and equitable student aid, all groups would have representation in each type of college equaling their representation in the population (nation, state, or locality) being considered.

Trends in College Access

While the college continuation rate – the ratio of high school seniors to college freshman enrolment (**Figure 3**) – has varied little since 1992, the percentage of 18–24-year-olds

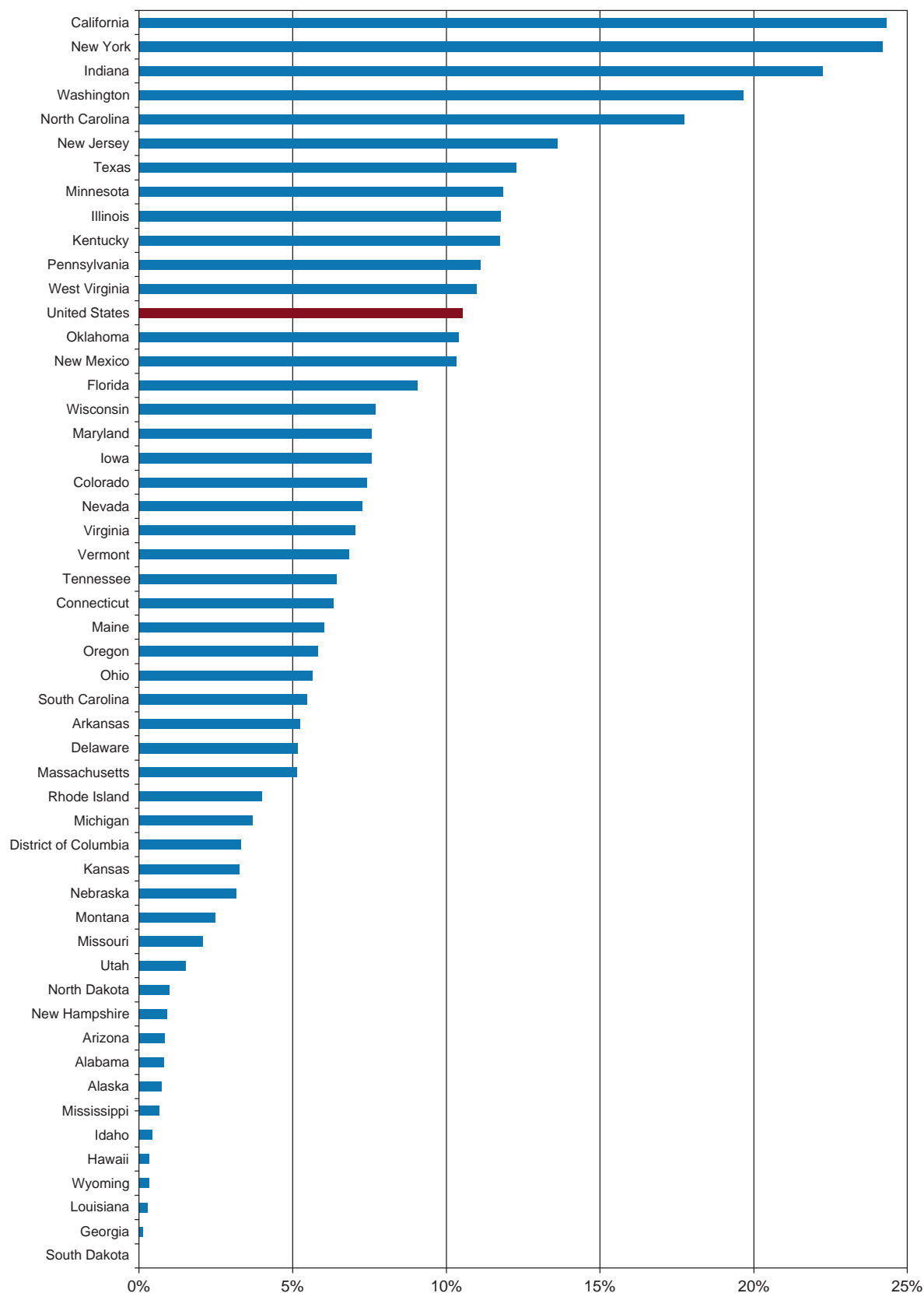


Figure 2 Ratio of average state funding per full-time equivalent (FTE) on need-based grants to average public tuition charge (2006). Data from NCES Integrated Postsecondary Education Data System and National Association of State Student Grant & Aid Programs.

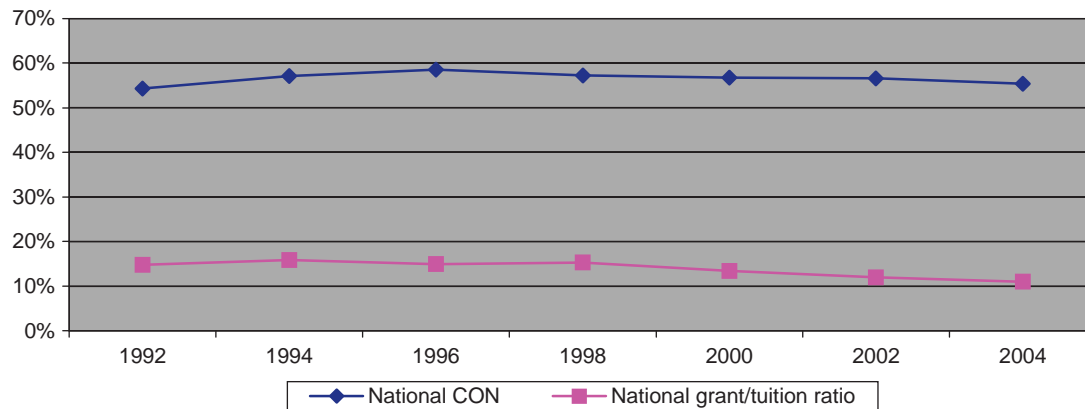


Figure 3 National trends in continuation rate and grant/tuition ratio.

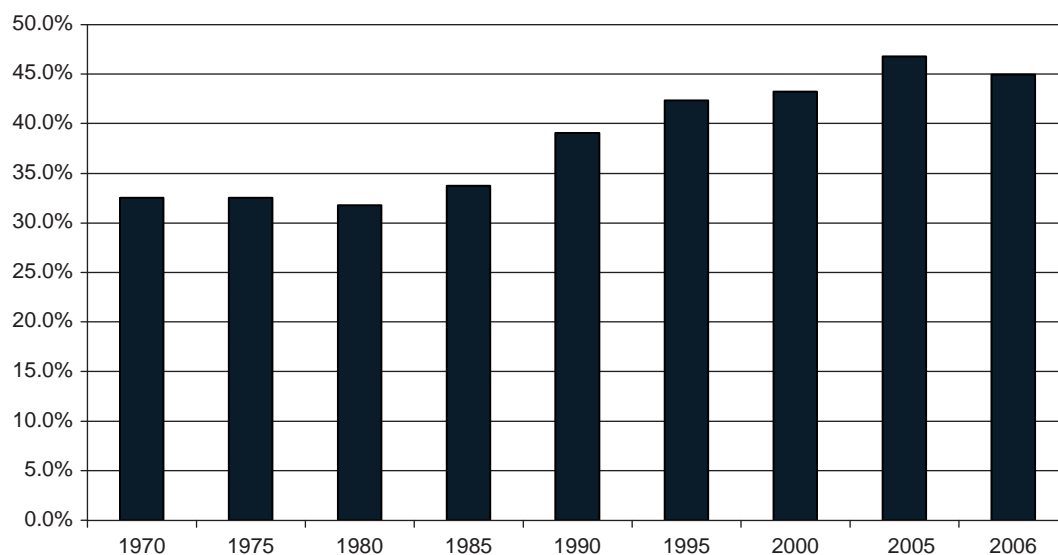


Figure 4 College enrolment rates for 18–24-year-old high school graduates. Data from *NCES Digest of Education Statistics 2007*, table 195.

(college-age students) who were currently or previously enrolled actually increased between 1995 and 2005 after having remained relatively flat between 1970 and 1985 (**Figure 4**). If we used this as our only indicator of access, we might conclude that college enrolment rates improved, possibly as a consequence of either improvement in preparation or availability of loans. Certainly this is possible, but we also need to ponder the discrepancy between percentage enrolled and the trend in the college continuation rate.

The fact that college-enrolment rates for freshmen did not increase substantially during a period when the percentage of traditional-age students enrolled did increase appears contradictory. However, it is essential to consider these trends in relation to college cost. The measure of currently and previously enrolled students includes students who have dropped out, while continuation rate is a measure of currently enrolled freshmen. People can have access to college – be able to enrol – and not be able to persist. Persistence can be deterred by academic, social, and financial

factors. The second measure of access – the percent of the population enrolled (used in **Figure 4**) – is a broader measure that includes students who have delayed enrolment and those who drop out. Low-income students are more likely to delay enrolment and to drop out than middle-income students (St. John, 2006), so the second measure captures more of the low-income, traditional college-age population that eventually enrolls.

The college-continuation rate (**Figure 3**) provides a ratio from institutional reports on freshman enrolment compared to school reports on students who graduate. The rate of current/previous enrolment provides a self-reported measure of students enrolled at any time over a 6-year period (**Figure 4**) as reported on the Current Populations Surveys (CPS). (The continuation rate, calculated every 2 years by Tom Mortensen, provides a ratio of the actual number of high school graduates to the actual number of college freshmen. In contrast, the self-reports on CPS (the indicator reported in **Figure 3**) cannot be as easily

verified.) There are two ways to interpret the discrepancy between these two measurements. It is possible that the CPS has consistently overreported enrolment after 1992. It is also possible that many students enrol in proprietary colleges or delay enrolment, as has long been reported. There is reason to expect that the number of students who eventually find a way into some type of college has increased over time, especially given the improvements in preparation noted above.

In combination, these two measures of college access appear to indicate that the rate of enrolment after high school has been relatively static (i.e., continuation rate), but that more students have eventually entered college since the educational reforms were implemented after 1990. Thus, there is mixed evidence of success on access. Educational reforms may be working, but they do not seem to be enabling growth in the traditional pipeline to college. This apparent discrepancy raises an important issue given that the government reports on access extolling the virtues of math preparation use a pipeline rationale which assumes continuous enrolment.

Unequal Opportunity

The CPS self-reports on college enrolment are also important because they provide a measure of equity in enrolment opportunity. In **Figure 5**, we compare the differentials over time in the percentages of Whites, Hispanics, and African Americans who reported enrolling in some form of post-secondary education. Trends in this indicator illustrate persisting inequality since a brief period of relative equality in the 1970s; instead of, for example, nearly equal African

American and White enrolment rates and higher enrolment rates for Hispanics than Whites in 1975.

In 1975, the Pell program was fully implemented and college costs were relatively stable. In addition, the Pell grant maximum had not yet been eroded. There is reason to speculate that relative equality in enrolment at this point was at least partially attributable to Pell grants. A comparison of trends in the Pell gap (**Figure 1**) with trends in the differential in enrolment rates (**Figure 4**) suggests that the growth in inequality since 1975 could be related to the decline in Pell grants coupled with increasing college costs. (As indicated in **Figure 4**, the data on Pell maximums are based on reports on college costs published annually by the College Board.)

All types of student aid increased over the last decade, adjusting for inflation: Pell grants increased by 73%, federal grants (Pell plus other) increased by 83%, and total federal aid (grants plus loans) increased by 77% (College Board, 2007). (This set of percentage rates was provided by the editor and included upon the request of the editor.) Given that all forms of student aid increased at rates similar to Pell and the gap between the Pell maximum and COA increased substantially, student aid did not keep up with tuition. All federal programs have maximum awards. The amounts of aid students could receive simply did not increase as fast as inflation.

The problem with the interpretation of trends is that Pell should have resulted in an increase in enrolment rates, according to arguments by economists, so there are lingering questions about the efficacy of Pell. In actual fact, total federal grant aid did not increase after Pell was implemented because of cuts in other grant programs, but

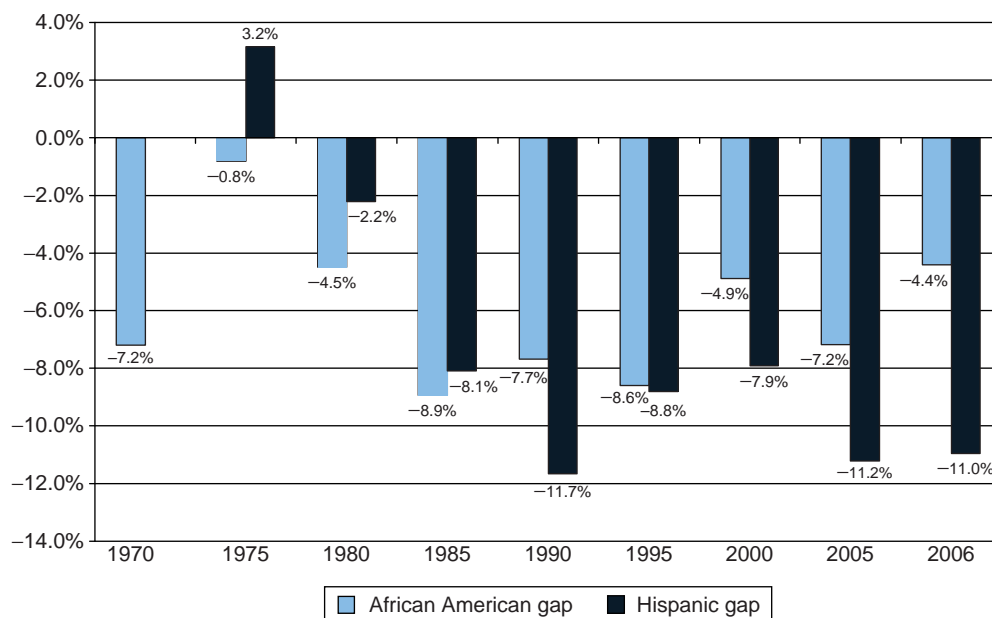


Figure 5 Difference in college enrolment for African Americans and Hispanics compared to high school graduates. Data from NCES *Digest of Education Statistics 2007*, table 1.

Pell did target aid to financial need. Economists have tended to focus on increases in Pell funding without considering the fact that total federal aid did not increase. With this additional background, it is apparent that targeting financial aid on need-based programs equalized opportunity and that additional funding may have been needed to increase enrolment rates.

Based on the review of access trends above, it may be tempting to argue that improvements in preparation improved enrolment rates better than Pell, but this argument does not solve the problem of unequal enrolment rates. The statistical analysis of the effects of state K–12 policies on preparation by students in the high school class of 2004 indicates that minorities benefited more substantially than majority students from raising standards, probably an artifact of improvements in preparation in high-poverty schools. The inequality in enrolment rates, however, does not suggest a link between improved preparation for minority students and improved enrolment rates. That is, if there were benefits from the new education policies on enrolment, they did not take the form of reducing inequality in enrolment rates. Quite the opposite, inequalities in enrolment opportunity persisted in the 2000s, even after educational reforms were widely implemented and modest gains were recorded in the rates of students taking the SAT and improved SAT math scores.

The persistent and growing inequality in post-secondary opportunity in the 2000s is perplexing because unequal access is often portrayed as a problem of academic preparation (Adelman, 2005; Choy, 2002). The inequality, however, appears to be linked to changes in student aid. In fact, graduation requirements increased during the period, a policy shift that favored minorities because it gave them access to college preparatory courses (Daun-Barnett, 2008; St. John, 2006; St. John and Moronski, 2008). In addition, test scores improved, an indicator that the new education policies worked, resulting in improved test scores. However, in spite of improved preparation and scores, inequality grew: The widespread implementation of the new math policies did not result in reduced inequality in access. Even if improved overall rates of enrolment were attributable to improvements in graduation requirements, a real possibility, this explanation would not hold up as a cause of the differentials, the inequality. To address this question further, we need to dig deeper into the sources of the inequality.

College Choice or Opportunity Denied?

As part of the Projects Promoting Equity in Urban and Higher Education, we have developed a set of indicators of representation: the percentage of FTE students by racial/ethnic group in a sector of higher education compared to that group's percentage of the population. (We generated

these statistics from fall enrolment data in the NCES Integrated Postsecondary Data System (IPEDS) and national statistics on population composition.) This measure has the ratio of the minority group to the college population. To judge the meaning of this, we needed to weight the percent of the group enrolled in a type of institution by that group's representation in the population as a whole. This approach is the best available measure of equal representation in higher education. We present trends in representation of students by group (African American, Hispanic, Asian, and White) for enrolment in public 2-year and public 4-year colleges.

Trends in enrolment composition at public 4-year colleges (**Figure 6**) illustrate a substantial and persistent differential in enrolment across groups. Hispanics and African Americans enrol at rates substantially below their representation in the population, while Whites enrol at a rate nearly equal their percentage of the population and Asian Americans enrol at a substantially higher rate. (The decline in the rate of enrolment of Asian Americans in the sum across groups after 2000 in **Figures 6** and **7** may be attributable to the growth in the percentage of students with unidentified racial/ethnicity.) This inequality is a serious problem in the United States. All groups subsidize public colleges through their taxes. In addition, state education policies have been implemented to improve preparation for all students, a policy that has benefited African Americans and Hispanics more than other groups (Daun-Barnett, 2008; St. John and Moronski, 2008); yet representation by these groups has not grown at a rate we would expect from the improvement in academic preparation. Thus, while there are indicators of growing access to 4-year colleges for some groups, it does not extend to all groups. Hispanics and African Americans clearly do not have equal opportunity to enrol in public 4-year colleges. This is sad and ironic given that the educational reforms implemented in the late 1990s and early 2000s (**Table 1**) were rationalized based on raising the level of preparation for and access to 4-year colleges for all students.

Instead, there has been a shift in enrolment of minorities into 2-year colleges at the same time that Whites have moved out of 2-year colleges (**Figure 7**). In 1992, both Whites and Hispanics were represented in 2-year colleges at a rate equaling their representation in the population, while African Americans were underrepresented. However, during the subsequent 14 years, African Americans increased enrolment in 2-year colleges relative to their proportion of the population. Latino representation in community colleges declined after the 2000 census. When interpreted in combination with other trends, it is apparent that: increased access by African Americans was limited to 2-year colleges, while Latin lost ground in access to both 2-year and 4-year colleges.

If we examine these trends in representation in comparison to patterns of high school reform and test scores,

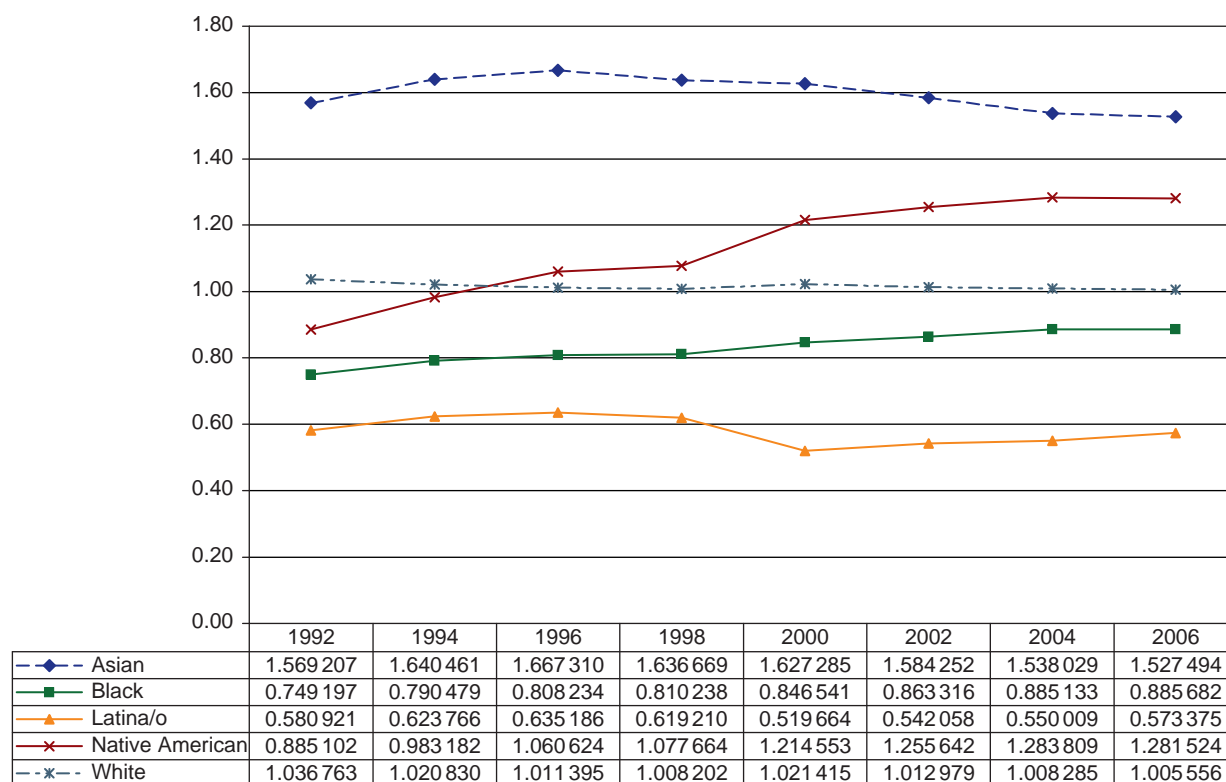


Figure 6 Ratio of group representation in public 4-year institutions to group representation in the United States. Data from NCES Integrated Postsecondary Education Data System and US Census Bureau.

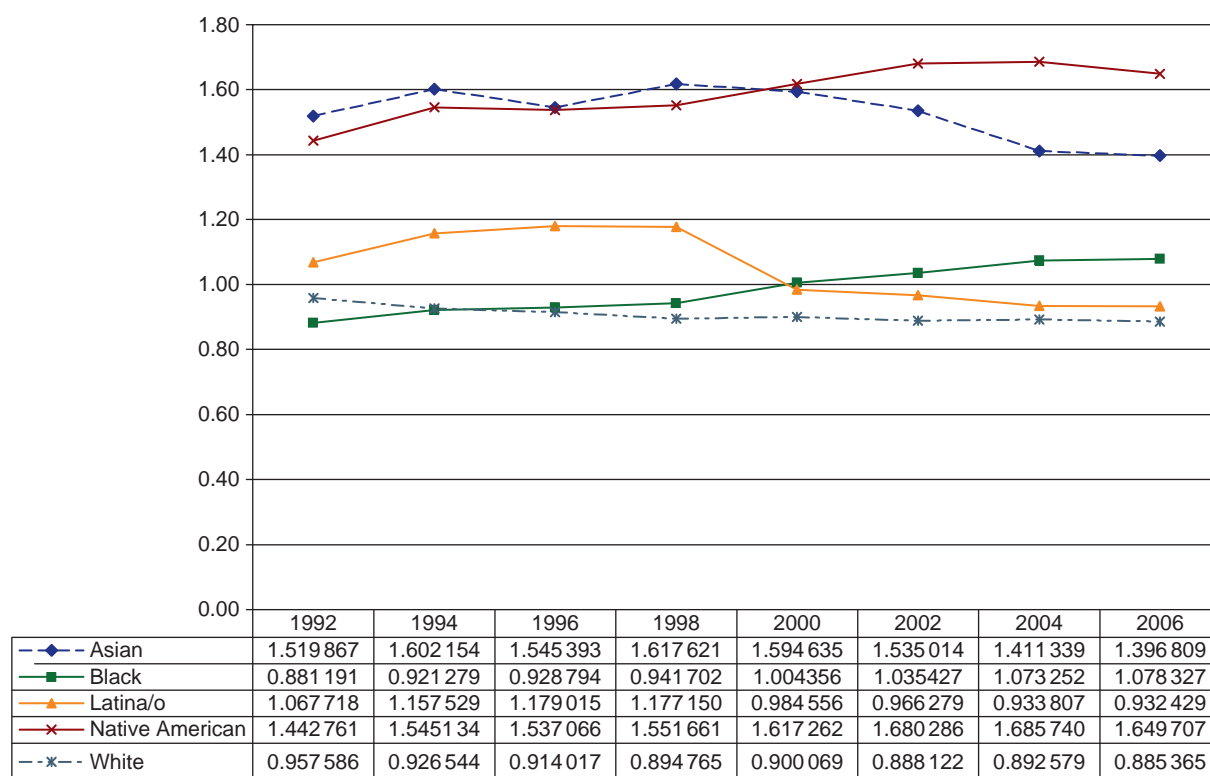


Figure 7 Ratio of group representation in public 2-year institutions to group representation in the United States. Data from NCES Integrated Postsecondary Education Data System and US Census Bureau.

it is apparent that the reforms have not had their intended consequences for African Americans and Hispanics. While reforms over the past two decades have been rationalized as a means of improving enrolment in and graduation from 4-year colleges, gains in opportunity have largely been denied to the groups in greatest need. If all students had benefited equally from the education reforms – which is not an unreasonable hypothesis since most reforms required improvements in preparation of all students – then inequalities in opportunity for higher education should have been reduced.

The best explanation for the growing inequality in opportunity to enrol in 4-year colleges is the failure of state and federal grants to keep up with college costs. The gap between the Pell maximum and costs at 4-year colleges grew after 2000 (Figure 1), when there was also a slight drop in state funding of need-based grants as a ratio of tuition (Figure 3). With the reduced ability to pay, the subpopulations with the highest percentages of low-income families – Hispanics and, to a somewhat lesser extent, African Americans – were priced out of 4-year colleges. The low representation of Hispanics and African Americans in public 4-year colleges, coupled with their overrepresentation in public 2-year colleges, reflects this outcome.

Conclusions

For more than two decades, correlations between college high school math courses and college outcomes have been used to argue that preparation is the most critical issue in expanding college access. Many of these studies used longitudinal data on the high school class of 1992 to build this argument. An examination of policy trends since 1992 reveals that all states have raised math standards since that time and increased the number of math courses required for high school graduation. In addition, the percentage of students taking the SAT and their average scores improved. Yet, inequality in post-secondary opportunity worsened. Since 2000, a period when this new regime of policies was well established, college continuation rates have decreased slightly, inequality in enrolment rates has increased, and gains in minority enrolment have been constrained to 2-year colleges. These conditions are serious and problematic.

What explains the growing inequality? Certainly improvement of opportunity to prepare for college should remain a high priority for state and federal officials. However, these trends reveal that we cannot rely on education reform alone as the remedy to inequality in opportunity. The growing gap between college costs and Pell grants, coupled with the decline in state funding for need-based grants relative to tuition charges, provides the best explanation for the growing inequality of the early twenty-first century. It is time to face up to the challenge by sustaining

and enhancing efforts to improve high schools while ensuring low-income students who meet the qualifications now being set as standards for high school graduation will have the financial means to pay for college.

Misreporting of statistics on academic preparation and hauling out statistics from studies of the high school class of 1992 no longer make sense. There have been extensive educational reforms since the class of 1992 enrolled in college, along with a very substantial rise in costs. The repetition of the message that deficits in preparation are the cause of inequalities may have been a motivation for raising requirements, but as recent history has shown, raising requirements does not guarantee equal opportunity. Leaving qualified students behind by limiting their opportunities to enrol in public 4-year colleges denies the promises of the new reform rationale. To fulfill this implied promise, it is necessary to ensure financial access for low-income students who take the steps to prepare for college.

Of course, social and cultural factors also influence enrolment behavior. No doubt interventions that encourage parents and children in low-income families to visit campuses and apply for student aid can lead to gains in enrolment rates, provided students are prepared academically and they have adequate student aid. Indeed, a comprehensive approach to improving access to higher education should include encouragement along with high school reform and guarantees of adequate student aid. The focus on academic preparation and financial aid in this article is not intended to overlook the role of encouragement. However, encouragement without the underpinnings of finances and academics could create false hope and cannot ensure equal opportunity.

Acknowledgment

This article reports statistics generated from national sources as part of the Projects Promoting Equity in Urban and Higher Education funded by the Ford Foundation at the National Center for Institutional Diversity (NCID), University of Michigan. The opinions expressed in this article represent the interpretations of the authors and do not necessarily reflect policy or positions of the Ford Foundation or NCID.

Bibliography

- Adelman, C. (2005). *Moving into Town and Moving on: The Community College in the Lives of Traditional-Age Students*. Washington, DC: US Department of Education.
- Advisory Committee on Student Financial Assistance (2002). *Empty Promises: The Myth of College Access in America*. Washington, DC: Author.
- Choy, S. P. (2002). *Access & Persistence: Findings from 10 Years of Longitudinal Research on Students*. Washington, DC: American Council on Education.

- College Board (2007). *Trends in Student Aid 2007*. New York: Author.
- Daun-Barrett, N. (2008). Preparation and access: A multi-level analysis of state policy influences on the academic antecedents to college enrollment. PhD Dissertation, University of Michigan.
- Fitzgerald, B. K. (2004). Federal financial aid and college access. In St. John, E. P. (ed.) *Public Policy and College Access: Investigating the Federal and State Roles in Equalizing Postsecondary Opportunity. Readings on Equal Education*, vol. 19, pp 1–28. New York: AMS Press.
- St. John, E. P. (2003). *Refinancing the College Dream: Access, Equal Opportunity, and Justice for Taxpayers*. Baltimore, MD: Johns Hopkins University Press.
- St. John, E. P. (2006). *Education and the Public Interest: School Reform, Public Finance, and Access to Higher Education*. Dordrecht: Springer.
- St. John, E. P. and Moronski, K. (2008). The impact of the Florida Bright Futures Scholarship Program on college preparation and access for low-income and minority students. *A Paper Prepared for Enlace Florida*. <http://enlacefl.usf.edu/research/Research%20Briefs/2009/The-Impact-of-FL-BrightFuturesScholarship-on-CollegePrep.pdf> (accessed September 2009).
- US Department of Education (1983). *A Nation at Risk*. Washington, DC: Author.

The Education of Indigenous Students

J W Tippeconnic and S C Faircloth, Pennsylvania State University, University Park, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

The awareness, knowledge, and bond of Indigenous peoples around the globe have increased in recent years. A greater understanding of their education conditions has been facilitated by actions of the United Nations; policy and practice by national and Indigenous governments; the growth of Indigenous schools, institutions, and organizations; and the collective research, teaching, and service of many individuals and organizations. Key leadership in this renewal and reconstruction of education has been provided by Indigenous peoples throughout the world. In keeping with this global movement, this article provides an overview of the historical and current factors that have served to shape the education of Indigenous children and youth. In doing so, we focus on Indigenous education from a global perspective.

According to the United Nations Permanent Forum on Indigenous Issues, (2006) there are more than 370 million Indigenous people residing in approximately 70 countries across the globe. Within this population, there are more than 5000 individual groups, representing approximately 4000 languages (IFAD). The term indigenous is often used synonymously with aboriginal, with different usages preferred in various regions of the world. Indigenous is used most often in international reports such as those published by the United Nations. However, terms such as native, tribal, First Nation, traditional, and Indian are also used to describe Indigenous people. Although these terms are used, there is a growing preference within many Indigenous communities to refer to themselves in a more precise and respectful manner that acknowledges the Indigenous name of the individual tribe or group (e.g., Marker, 2000). For example, in the United States, Indigenous people may refer to themselves as *Diné*, *Comanche*, *Lumbee*, or any of the more than 650 federally and state-recognized tribes across the nation.

In a study of the problem of discrimination against Indigenous populations, a report commissioned by the United Nations, (cited in Secretariat of the Permanent Forum on Indigenous Issues, 2004), Martínez Cobo described Indigenous communities, peoples, and nations as those with

historical continuity with pre-invasion and pre-colonial societies that developed on their territories, consider themselves distinct from other sectors of the societies now prevailing on those territories, or parts of them. They form at present non-dominant sectors of society and are determined to preserve, develop and transmit to

future generations their ancestral territories, and their ethnic identity, as the basis of their continued existence as peoples, in accordance with their own cultural patterns, social institutions and legal system (Introduction).

Indigenous peoples have also been characterized as having a long-term, continuous connection to a specific place, ancestry/lineage, culture, and language. Each of these groups generates and, to some extent, lives by its own unique core of Indigenous knowledge and ways of knowing.

Indigenous Knowledge and Its Relationship to Education

According to the United Nations Educational, Scientific and Cultural Organization (UNESCO), Indigenous knowledge is “culture- and context-specific; non-formal knowledge; orally transmitted, and generally not documented; dynamic and adaptive; holistic in nature; [and] closely related to survival and subsistence for many people worldwide.” Indigenous knowledge has also been described as “the local knowledge that is unique to a given culture or society ... the basis for agriculture, health care, food preparation, education, environmental conservation, and a host of other activities” (Sri Lanka Centre for Indigenous Knowledge, 1996). (For additional information, see the website of *Best Practices on Indigenous Knowledge*.) Transmission of Indigenous knowledge often occurs from generation to generation through use of oral traditions. (Adapted from *Indigenous Knowledge and Sustainable Development*, Sri Lanka Centre for Indigenous Knowledge, University of Sri Jayewardenapura, 1996, pp. vii–viii.)

Indigenous knowledge is important not only for Indigenous people, but the world as a whole. For example, Indigenous knowledge has much to say about the ecosystem and the ways in which communities can ensure the sustainability of natural resources.

Cajete (1994) relates Indigenous knowledge to education by identifying a number of elements of Indigenous education, including the recognition of Indigenous languages as a “sacred expression[s] of breath.” Indigenous education also recognizes that each person and each culture contains the key to one’s well-being and development. In addition, integration and interconnections are universal principles of Indigenous education. The purpose of Indigenous education is to teach a way of life that sustains

both the individual and the community by integrating “human individuality with communal needs” (p. 30).

Although a critical component in the successful education of Indigenous peoples, Indigenous knowledge lacks widespread understanding and use in most Western systems of education. According to Battiste (2008):

colonial and neocolonial [educational] models continue to offer publicly funded schools and their students a fragmented, negative, and distorted picture of Indigenous in history, textbooks, and curricula. These models characterize IK [Indigenous Knowledge] as primitive, backward, or superstitious, causing Indigenous peoples to be viewed as deficient and requiring remedies that destroy Indigenous peoples’ self-esteem and self-confidence. (p. 86)

The inclusion of Indigenous knowledge in national education systems has been a long-standing challenge, made difficult to realize because of the deficit approach mentioned above and because of the complex nature of the task. King and Schielmann (2004), in discussing the recognition and inclusion of Indigenous knowledge, suggest that indigenous communities must support and consent to what knowledge is included in school curricula. They go on to discuss the inclusion of Indigenous knowledge in Western education:

Quality indigenous education offers innovative solutions to the complex issue of incorporating indigenous knowledge, and values indigenous knowledge systems as equal and complementary to Western systems. This includes:

- Respect for, and recognition of ownership of, indigenous communities as holders of indigenous knowledge, and for their specific way of generating and transmitting knowledge.
- Identification and incorporation of relevant local, cultural knowledge with the participation and informed consent of indigenous communities and elders, in particular as regards the selection, transmission and documentation of that knowledge – in planning programmes, the selection of teaching methods, the design of curricula and the production of educational material.
- Inclusion of stories, diaries, textbooks, etc., as well as non verbal education materials produced by indigenous teachers [and]
- The active participation of students and community members who serve to develop a curriculum founded on indigenous people’s cultural identity and history (p. 37).

Indigenous Knowledge and Education from a Historical Perspective

Historically, educational policies and practices have excluded Indigenous knowledge in schools. Battiste (2008)

notes that a common experience among Indigenous peoples is colonization by foreign forces. This has resulted in assimilationist policies which have sought to abolish the use of Indigenous languages, knowledge, and cultures in the education of Indigenous children and youth. Although Indigenous knowledge has always been a guiding force in the lives of Indigenous peoples, historical accounts of Indigenous education routinely describe education as if it began only at the point of contact. In general, colonizers worldwide viewed Indigenous people as individual lacking knowledge, inferior in their ways of thinking, illogical, and intellectually weaker, and set to work to change and improve them. For instance, in Australia, Eckermann (1998) noted “since colonization in 1788 Australia’s history has followed patterns common to most colonial nations. Early contact between colonists and Indigenous groups was marked by conflicting attitudes and perceptions” (The living past section, para. 1).

Formal education in schools has been used as a means of assimilating, acculturating, civilizing, and Christianizing Indigenous people. These goals were carried out by sending Indigenous children to schools where they were prohibited from speaking their native languages or practicing native religions and other cultural manifestations. In many instances, particularly in the United States, Canada, and Australia, these practices were in direct violation of the provisions outlined in treaties between the colonizing governments and Indigenous nations – treaties in which many Indigenous nations ceded land to the governments in exchange for provisions including the education of Indigenous children (National Indian Education Association, n.d.; Eckermann, 1998; Barman, *et al.* 1986). By entering into treaties with Indigenous nations, governments were recognizing the sovereign status of these nations and their right to establish a government-to-government relationship with other nations – a right which has been routinely abrogated.

In effect, politics, power, and control were intertwined with education. Adams (2008) argues that in the United States, “the Protestant ideology, the civilization–savagism paradigm, and the White hunger for Indian land were all mutually reinforcing and hopelessly intertwined as factors influencing the educational campaign to assimilate the Indian” (p. 32). Over time, the Federal Government of the United States supported different religious groups in their efforts to Christianize and civilize Native Americans; established a federal system of schools for Native children; and later provided incentives to public schools to accept and educate these students (Lomawaima and McCarty, 2006; Reyhner and Eder, 2004). Boarding or residential schools were an educational approach used by national governments, not only in the United States, but elsewhere across the globe, and by religious groups to assimilate and change Indigenous peoples. Students, often young students, were taken from their families and communities and placed in boarding schools where they were not

allowed to speak their native languages or practice their cultures and indoctrinated in a course of study that aimed to assimilate them into mainstream society (Adams, 1995; Lomawaima, 1994; Carroll, 2000; Child, 2000; Trafzer *et al.* 2006).

In a discussion on education and culture, the United Nations Permanent Forum on Indigenous Issues provides a snapshot of the historical effects of schooling around the world for Indigenous students. In general, the UN observed:

that millions of children [continue] to be taught in languages they [do] not use or even understand ... the participation of indigenous peoples in designing curricula [is] still limited, and education still [falls] short of eliminating prejudice and discrimination targeted at indigenous peoples. The lack of indigenous education ... continue[s] to set indigenous youth apart from their own cultures. (UN Press Release HR 46/74)

Among the issues identified at the UN Forum were high drop out rates among Indigenous students; the need to integrate Indigenous languages and cultures into school curricula; preservation of human rights of Indigenous peoples; bilingual education; lack of funding; adequate facilities; the need for educational policies to show respect for cultural and linguistic identity; and the critical role of self-determination in improving the status of education among Indigenous students and schools (UN Press Release HR 46/74).

The findings reported above are similar to findings reported in a 1969 study commissioned by the United States' Congress. This study concluded that the education of American Indians and Alaska Natives in public and federal schools was a failure and a national tragedy and a national disgrace (US Senate, 1969: x). The failure of the educational system has "condemned him [American Indian] to a life of poverty and despair" (p. x). In addition to the need for adequate funding, the report recommended "maximum participation and control by Indians in establishing Indian education programs" (US Senate, 1969:106). A change in the national educational policy, including a return to Indian controlled education, was considered necessary to reform the educational system for American Indians and Alaska Natives in the United States. Similar calls soon began to resonate across the world.

Movement Toward Self-Determination

The movement toward increased Indian control and self-determination was in response to the historical policies and practices in the education of Indigenous peoples. Policy changes in the 1970s marked by the growth of tribal self-determination in the United States (Tippeconnic, 2007) and a movement toward Indian control of education

in Canada (Hare, 2007) resulted in increased parental and community involvement, integration of Indigenous knowledge into the curricula, and an increase in the number of Indigenous educators in teaching and leadership positions.

In Australia, self-determination became official policy in 1972, but only recently have concrete steps been taken to bring the policy to life. According to Eckermann (1998),

all states have initiated Aboriginal Education Consultative Groups (AECGs) which, at local, regional and state levels, have been actively advising Departments of Education. Most States have developed Aboriginal education policies in order to address issues related to Aboriginal Studies curricula, providing appropriate environments and teaching strategies to facilitate Aboriginal expectations of Aboriginal decision making within educational institution. (Current trends, para. 1)

In New Zealand, the Māori continue their quest for self-determination in the educational arena as they work to gain educational parity. They seek a system of education that is shaped and guided by Māoris for Māoris (Durie, 1999).

Effects of the Self-Determination Movement on Post-Secondary Education for Indigenous Peoples

An area in which Indigenous peoples have witnessed much growth and success is at the higher education level. In establishing institutions of higher education, Indigenous peoples asserted their right to shape the design and delivery of educational programs across the lifespan. As Jennings (2004) suggests, Indigenous people recognized that the factors that influence educational attainment are not the same for all individuals and communities. For successful education practices to be sustained, they must be desired, shaped, and controlled by local communities (Jennings, 2004).

In the United States, the tribal college movement began in the 1960s with the establishment of the Navajo Community College in 1968. Today, there are approximately 40 tribally controlled colleges and universities across the nation. (See the website of the American Indian Higher Education Consortium for additional information regarding tribally controlled colleges and universities in the United States.) Similar movements have taken shape in Canada and New Zealand. In 2002, the World Indigenous Nations Higher Education Consortium was established to recognize and promote the significance of Indigenous education through a global higher education network. Moreover, in the United States, there are approximately 80 Native-studies programs nationally that stand alone or offer minors in related academic areas. These programs

have experienced a renewed interest and growth in recent years (Tippeconnic and Swisher, 2000).

In addition to working within tribally controlled colleges and universities, Indigenous scholars, many of them working in institutions of higher education, are conducting research using native epistemologies (Cajete, 2008; Smith, 1999). Tribes are also reframing research by establishing protocols that control the conduct of research in their communities (Lomawaima, 2008).

Current Status of Indigenous Education

In spite of the progress Indigenous people have made in regaining self-determination and control of education, for many Indigenous peoples worldwide, education remains a sociocultural, as well as politically contested arena. As Eckermann (1998) writes about Indigenous people in Australia, Indigenous people around the world have been “subjected to systemic bias by being excluded from gaining the skills necessary to participate in and impact . . . mainstream systems, whether these be economic, educational, health or political” skills (The living past, para. 15). It is a fact that Indigenous education remains in a state of crisis across the globe (see **Table 1**).

In Canada, Aboriginal children experience high drop out rates (Australia Government Department of Education, 2007). Less than 50% of Aboriginal students in British Columbia graduate compared to more than 80% of non-aboriginal students. Graduation rates are even lower in many districts. Aboriginal children are also disproportionately represented in special-education programs and services; they perform lower on achievement tests and enter college at rates significantly lower than their peers. Many of these children also tend to experience high rates of mobility as they move on and off reserves. In the classroom, Aboriginal

students are rarely taught by Aboriginal teachers and the curricula and teaching methods often do not reflect Aboriginal needs and values (Council of Ministers of Education, Canada, 2004). Responding to the call for reform in Aboriginal education policies, in 2006, the Canadian government granted First Nations’ people the right to control education for First Nations students living on reserves in British Columbia. This was the first province within Canada to enact such legislation.

In Australia, Indigenous peoples are less likely to receive preschool education than their peers. They also lag behind their peers in literacy and numeracy skills. Access to education is another issue of concern as there is limited access to community-based secondary education for Indigenous students. Indigenous students also experience high drop out rates and school failure. Those who do remain in school are often placed on a vocational track (Australia Government Department of Education, 2007).

In New Zealand, approximately 90% of Māori children attend preschool/early childhood programs; however, they score significantly lower on tests of achievement and drop out of school at rates significantly higher than their peers (Ministry of Education, n.d.a). The situation is similar in Mexico, where fewer Indigenous girls than boys attend school and there is a high rate of illiteracy. To combat these problems, there has been a movement since the 1970s, toward bilingual education and the use of fluent native speakers as teachers. Although Indigenous students have experienced greater levels of success in Indigenous schools as compared to non-Indigenous schools, there are few secondary Indigenous schools in Mexico (Ministry of Education, n.d.b).

In the United States, American Indian and Alaska Native students represent less than 1% of the total school-age population. While the majority (90%) attends public schools, approximately 10% attends schools operated or

Table 1 Selected demographics

<i>Country</i>	<i>Total population</i>	<i>Indigenous population</i>	<i>Indigenous school-age population</i>
Australia	20.9 million	458 500 (2.2%)	135 668 (4% of total)
Canada	33.4 million	976 305 (3.3%)	324 500 (5.6% of total)
Mexico	100 million	10 million (10%)	1.3 million (16.4% do not attend school)
New Zealand	4.14 million	635,000 (15%)	162 685 (21.6% of total)
United States	281.4 million	4.1 million (1.5%)	1.35 million (33% of total) ^a

^aThis figure represents children and youth between the ages of birth and 18.

Adapted from Australia Government Department of Education. Employment and Workplace Relations. (2007). What works. The work program. Improving outcomes for Indigenous students. Retrieved May 11, 2008, from http://www.whatworks.edu.au/2_3.htm (accessed October 2009); and US Census Bureau. (2002, February). The American Indian and Alaska Native population: 2000. Census 2000 brief. Washington, DC: Author. Retrieved May 11, 2008, from <http://www.census.gov/prod/2002pubs/c2kbr01-15.pdf> (accessed October 2009).

funded by the Bureau of Indian Education and tribes. American Indian and Alaska Native students represent approximately 650 federally and state-recognized tribes, each with their own unique culture and many with their own languages still being spoken. American Indian and Alaska Native students experience high drop out rates and continue to perform significantly below their non-Native peers on achievement tests in reading and math (e.g., Rampey, *et al.* 2006).

As demonstrated above, Indigenous peoples continue to feel the effects of colonization and its assimilationist ideologies and practices, many of which have been inextricably linked to educational institutions and practices. The challenge for educators working in and with Indigenous schools and communities, as well as those working with Indigenous students displaced, either willingly or unwillingly, from their places of origin, is to develop and implement educational practices that honor and validate Indigenous knowledge and ways of knowing. Such a commitment does not negate the pursuit of high academic standards, but affirms the potential for such standards to coexist, and in many cases, to be cross-fertilized and nurtured by Indigenous languages, cultures, and practices.

Working to Strengthen and Improve Indigenous Education

Kind and Schielmann (2004) identified 16 Indigenous programs from different parts of the world that provide quality education to Indigenous peoples, including efforts in early childhood education; bilingual education; language preservation; training of Indigenous educators; educational reform; and capacity building. As educators work to improve and strengthen education for Indigenous children and youth, they are challenged to:

1. facilitate meaningful parent and community involvement in planning, delivering, and evaluating educational programs and services;
2. promote educational self-determination, the right to determine the purpose, mission, and ultimately the way in which education is designed, delivered, and evaluated;
3. diversify the teacher and administrative ranks to include Indigenous teachers and school leaders;
4. expand and improve early childhood education and care;
5. ensure children and youth have access to education across the life span;
6. provide high-quality education through secondary school such that Indigenous youth will graduate prepared for a full range of post-secondary options; and
7. expand Indigenous efforts in developing and operating tribally or Indigenous-controlled systems of higher education (or institutions of higher education) (Magga, 2005).

Conclusion

Although progress has been made in providing quality education for Indigenous students, many obstacles and challenges remain (King and Schielmann, 2004). According to the United Nations Permanent Forum on Indigenous Issues (n.d.), Indigenous children around the world face a number of challenges including lack of adequate healthcare; relocation to urban areas; suicide; illiteracy; high drop-out rates; lack of culturally appropriate education; poverty; and racial and ethnic discrimination. Lack of participation in higher education programs is also evident among Indigenous people in Canada, Australia, Mexico, and the United States (Australia Government, 2007).

Each of these factors impact Indigenous children's access to high-quality educational programs, which not only equip them to thrive in an increasingly globalized world economy, but also maintain and strengthen their individual and collective ties to their Indigenous languages, cultures, and practices. Language loss, in particular, poses a threat to the continued production of what many have deemed Indigenous knowledges and Native ways of knowing (see e.g., Barnhardt and Kawagley, 2005). Some have argued that loss of language leads to loss of culture. According to McCarty, *et al.* (2006), "when even a single language falls silent, the world loses an irredeemable repository of human knowledge" (p. 29).

The education of Indigenous children and youth is an area in need of continued research and, more importantly, of translation of this research into practice – a practice that recognizes, affirms, and respects the critical role of Indigenous knowledge, culture, and ways of knowing into the teaching and learning process for Indigenous children and youth. Indigenous peoples across the world have always engaged in and valued the praxis of teaching and learning; unfortunately, the Indigenous way of engaging in this praxis has often been overshadowed by the process of colonization and globalization. The key to improving the overall education of Indigenous children and youth is to build upon the strengths they possess – strengths grounded in Indigenous knowledge. In the long run, the strengths of Indigenous knowledge and education (Lomawaima and McCarty, 2006) will persevere and grow as Indigenous peoples seek greater participation and control over the education of their children and young adults.

Bibliography

- Adams, D. W. (1995). *Education for Extinction: American Indians and the Boarding School Experience 1875–1928*. Lawrence, KS: University Press of Kansas.
- Australia Government Department of Education (2007). Employment and Workplace Relations. What works. The work program. Improving outcomes for Indigenous students. http://www.whatworks.edu.au/2_3.htm (accessed October 2009).

- Barman, J., Hebert, Y., and McCaskill, D. (1986). *Indian Education in Canada*. Vancouver: University of British Columbia Press.
- Barnhardt, R. and Kawagley, A. O. (2005). Indigenous knowledge systems and Alaska Native ways of knowing. *Anthropology and Education Quarterly* **36**(1), 8–23.
- Battiste, M. (2008). The struggle and renaissance of Indigenous knowledge in Eurocentric education. In Villegas, M., Neugebauer, S. R., and Venegas, K. R. (eds.) *Indigenous Knowledge and Education: Sites of Struggle, Strength, and Survivance*, pp 85–91. Cambridge, MA: Harvard University Press.
- Carroll, J. T. (2000). *Seeds of Faith: Catholic Indian Boarding Schools*. New York: Garland Publishing, Inc.
- Cajete, G. (1994). *Look to the Mountain: An Ecology of Indigenous Education*. Durango, CO: Kivaki Press.
- Child, B. J. (2000). *Boarding School Seasons: American Indian Families 1900–1940*. Lincoln, NE: University of Nebraska Press.
- Council of Ministers of Education, Canada (2004). *Background on the CEMC Aboriginal education action plan*. <http://www.cmec.ca/publications/BackgroundunderAboriginalEducation.en.pdf> (accessed October 2009).
- Durie, A. (1999). Emancipatory Māori education: Speaking from the heart. In May, S. (ed.) *Indigenous Community-Based Education*, pp 67–78. Philadelphia, PA: Multilingual Matters Ltd.
- Eckermann, A. K. (1998). The economics of aboriginal education. *International Journal of Social Economics* **25**(2), 301–313.
- Hare, J. (2007). First Nations education policy in Canada: Building capacity for change and control. In Joshee, R. and Johnson, L. (eds.) *Multicultural Education Policies in Canada and the United States*, pp 51–68. Vancouver, Canada: UBC Press.
- Jennings, M. (2004). *Alaska Native Political Leadership and Higher Education: One University, Two Universes*. Walnut Creek, CA: Altamira Press.
- King, L. and Schielmann, S. (2004). *The Challenges of Indigenous education: Practice and Perspectives*. United Nations: UNESCO Publishing.
- Lomawaima, K. T. (1994). *They Called It Prairie Light: The Story of Chilocco Indian School*. Lincoln, NE: University of Nebraska Press.
- Lomawaima, K. T. and McCarty, T. L. (2006). *To Remain an Indian: Lessons in Democracy from a Century of Native American Education*. New York: Teachers College Press.
- Magga, O. (2005). Indigenous education. *Childhood Education* **81**(6), 319–320.
- Marker, M. (2000). Review essay: Ethnohistory and indigenous education: A moment of uncertainty. *History of Education* **29**(1), 79–85.
- McCarty, T. L., Romero, M. E., and Zepeda, O. (2006). Reclaiming the gift: Indigenous youth counter-narratives on Native language loss and revitalization. *American Indian Quarterly* **30**(1/2), 28–48.
- Ministry of Education. Te Tāhuhu o te Mātauranga (n.d.a). Indigenous education in New Zealand. Retrieved May 11, 2008, from http://www.dest.gov.au/NR/rdonlyres/18C85F88-2204-4875-A4EB-7EF22B97A33C/17799/Country_Report_NZ.pdf (accessed October 2009).
- Ministry of Education. Te Tāhuhu o te Mātauranga (n.d.b). Mexico fact sheet. Retrieved May 11, 2008, from http://www.dest.gov.au/NR/rdonlyres/FEAD46E0-08F2-457C-9684-4DAECFDF963C/17798/Country_Paper_Mexico.pdf (accessed October 2009).
- National Indian Education Association (n.d.). History of Indian Education. Education facts and history. Retrieved April 20, 2008, from <http://www.niea.org/history/educationhistory.php> (accessed October 2009).
- Rampey, B. D., Lutkus, A. D., and Weiner, A. W. (2006). *National Indian Education Study, Part I: The Performance of American Indian and Alaska Native Fourth- and Eighth-Grade Students on NAEP 2005 Reading and Mathematics (NCES 2006-463)*. Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Statistics.
- Reyhner, J. and Eder, J. (2004). *American Indian Education: A History*. Norman, OK: University of Oklahoma Press.
- Secretariat of the Permanent Forum on Indigenous Issues (2004). *The Concept of Indigenous Peoples*. United Nations: Department of Economic and Social Affairs, Workshop on data collection and disaggregation for Indigenous peoples, PFII/2004.WS.1/3.
- Smith, L. T. (1999). *Decolonizing Methodologies: Research and Indigenous peoples*. New York: Zed Publishing.
- Sri Lanka Centre for Indigenous Knowledge (1996). Indigenous knowledge. Sri Lanka: University of Sri Jayewardenapura Retrieved May 11, 2008, from <http://www.ens.gu.edu.au/ciree/LSE/MOD5.HTM#act5> (accessed October 2009).
- Tippeconnic, J. W., III (2007). Policy issues in the education of American Indians and Alaska Natives. In Joshee, R. and Johnson, L. (eds.) *Multicultural Education Policies in Canada and the United States*, pp 69–82. Vancouver, Canada: UBC Press.
- Tippeconnic, J. W., III and Swisher, K. G. (2000). Higher education: Our way. *Native Americas: Hemispheric Journal of Indigenous Issues* **17**(4), 64.
- Trafzer, J., Keller, J. A., and Sisquoc, L. (eds.) (2006). *Boarding School Blues: Revisiting American Indian Educational Experiences*. Lincoln, NE: University of Nebraska Press.
- United Nations Permanent Forum on Indigenous Issues (n.d.). Indigenous children and youth. Retrieved April 28, 2008, from <http://www.un.org/esa/socdev/unpfii/en/children.html> (accessed October 2009).
- United Nations Permanent Forum on Indigenous Issues (2006). Indigenous children and youth. About UNPFII and a brief history of indigenous peoples and the international system. <http://www.un.org/esa/socdev/unpfii/en/history.html> (accessed October 2009).
- US Senate (1969). *Indian education: A national tragedy-a national challenge*. Washington, DC: U.S. Government Printing Office.

Further Reading

- Almeida, D. (1998). Indigenous education: Survival for our children. *Equity and Excellence in Education* **31**(1), 6–10.
- Doige, L. A. C. (2001). Literacy in aboriginal education: An historical perspective. *Canadian Journal of Native Education* **25**(2), 117–128.
- May, S. and Aikman, S. (2003). Indigenous education: Addressing current issues and developments. *Comparative Education* **39**(2), 139–145.
- Rural Poverty Portal (n.d.). *Statistics and Key Facts about Indigenous Peoples*. Retrieved March 11, 2008, from <http://www.ruralpovertyportal.org/english/topics/indigenous/statistics.htm> (accessed October 2009).
- United Nations (n.d.) *Best practices on Indigenous Knowledge*. Retrieved April 2008, from <http://www.unesco.org/most/bpindi.htm> (accessed October 2009).
- United Nations Permanent Forum on Indigenous Issues (2003). Press Release HR 46/74. <http://www.un.org/news/Press/docs/2003/hr4674.doc.htm> (accessed October 2009).
- US Census Bureau (2002). The American Indian and Alaska Native population: 2000. Census 2000 brief. Washington, DC: Author. <http://www.census.gov/prod/2002pubs/c2kbr01-15.pdf> (accessed October 2009).
- Villegas, M., Neugebauer, S. R., and Venegas, K. R. (eds.) (2008). *Indigenous Knowledge and Education: Sites of Struggle, Strength, and Survivance*. Cambridge, MA: Harvard Press.

Relevant Websites

- <http://www.ankn.uaf.edu> – Alaska Native Knowledge Network.
- <http://aihec.org> – American Indian Higher Education Consortium.
- <http://www.ens.gu.edu.au> – Faculty of Environmental Sciences.
- <http://www.niea.org> – National Indian Education Association.
- <http://www.un.org> – United Nations Permanent Forum on Indigenous Issues.
- <http://www.unesco.org> – UNESCO.
- <http://www.win-hec.org> – World Indigenous Nations Higher Education Consortium.

The Health Advantages of Educational Attainment

G Y Dinwiddie, University of Maryland College Park, College Park, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Social status – The honor or prestige attached to one's position in society (one's social position). The stratification system, which is the system of distributing rewards to the members of society, determines social status. Social status – the position or rank of a person or group within the stratification system – can be determined in two ways. One can earn their social status by their own achievements, which is known as achieved status, or one can be placed in the stratification system by their inherited position, which is called ascribed status. Ascribed statuses include race/ethnicity, age, gender, and nativity.

Socioeconomic status (SES) – A combined measure of an individual or family's economic and social position relative to others, based on income, education, and occupation. When analyzing a family's SES, the mother and father's education and occupation are examined, as well as combined income, versus with an individual, when their own attributes are assessed. In addition, income, occupation, and education have been shown to be strong predictors of a range of physical and mental health problems, such as respiratory viruses, arthritis, coronary disease, and schizophrenia.

Introduction

Over the past several decades, the relationship between (socioeconomic status (SES)) and health has been well documented. In developed countries, such as the United States and United Kingdom, research has shown that individuals with higher levels of SES are less likely to suffer from chronic and acute illness, have healthier life-styles, and are able to prolong life compared to individuals with lower SES. Measures of SES, such as income, education, and occupation, are the indicators most commonly used to assess the mechanisms through which social factors influence morbidity and mortality. Although there is a strong relationship between high levels of SES and health advantages, there has been continuous debate concerning whether different measures of SES have similar buffering effects on health when examined independently. Income

and occupation provide some indication of exposure to unhealthy living and working conditions, as well as the ability to purchase medical and preventive care (Lynch, 2003). However, there is no clear consensus about whether the relationship between health advantages and educational attainment are strong predictors of health compared to income and occupation or how education operates to protect health. Some argue that education mediates the relationship between SES and health because it increases income potential and allows individuals to seek higher-paying occupations. Others claim that educational attainment works independently on health outcomes because the educated are less likely to practice health-risk behaviors and are more informed about new technologies that improve health. However, not much is known whether the advantages of educational attainment are similar for racial/ethnic or other marginalized groups in the United States and abroad. From a social science perspective, this article reviews the current research that has informed the relationship between education and various indicators of morbidity and mortality in the United States. We briefly discuss the direct and three popular mechanisms associated with educational attainment which scholars have identified as advantageous to health over the life course. Using demographic and sociological research that has informed our knowledge about educational attainment and health, we also discuss popular research in the international front which identifies how intergenerational processes are important for understanding how education mediates the relationship between SES and mortality. After reviewing the mechanisms of health advantages, we then comment on whether using levels of schooling as a mechanism for understanding health advantages is useful in global health.

Measuring Educational Attainment and Health Outcomes

The relationship between educational attainment and health advantages has been well documented. Beginning with Kitagawa and Hauser's (1960) study of socioeconomic variations in mortality in the United States, many scholars have since used educational attainment as an important measure of SES. Educational attainment, schooling, or educational status has been the preferred measure of SES for several reasons. First, in most industrial countries, basic levels of schooling are required and although some

individuals drop out before obtaining a high school diploma, most have had exposure to the fundamentals of reading, writing, and arithmetic. For example, a vast majority of citizens in the United States (with the exception of older adults born in the early twentieth century and some marginalized racial/ethnic groups) have had some level of schooling because prohibiting a child from attending primary school is against the law. In terms of measurement, in most developed countries, educational attainment is usually measured by the number of years respondents spent in school which can range from primary (kindergarten, middle, and high school) to secondary schooling (undergraduate, graduate, professional, and technical training). Levels of schooling are regarded as better indicators of SES compared to employment because individuals may or may not have a steady source of income but have had some exposure to school. In addition, one's educational attainment is unaffected by health impairments that may emerge in adulthood (Elo and Preston, 1996) so educational attainment is stable and does not change over time to the extent that income might. For these reasons, scholars contend that education is a good measure of SES when assessing the relationship between schooling and health advantages/disadvantages.

As regards indicators of health, it has been measured in various ways, including self-rated health, mortality, functional health, mental health, cognition, and psychological well-being, to name a few. Depending on the health indicator used, we can assess various states of the body that are physical, mental, or cognitive. For example, scholars contend that self-rated health is a good indicator of health status because it assesses general well-being and is a better predictor of mortality than physician-assessed health. On the other side, some scholars contend that measuring the presence of disease is appropriate for assessing how the body degenerates over time and can become at risk for comorbid conditions over the life course. Moreover, mortality is important for measuring endpoints of life and the antecedents that are implicated in death. Regardless of the health indicator used, the association between educational attainment and health outcomes has shown to be a strong predictor of good health over the life course with the most benefits occurring in later life particularly between the years of 45 and 60 or compromised health because of the lack of educational attainment at earlier years.

The Relationship between Education and Health

The literature focusing on how education operates to protect health relies on two prevailing explanations. The first measures a direct effect of education on health and mortality, whereas the second identifies several mediating mechanisms implicated for explaining the relationship

between knowledge acquisition and health outcomes. Typically in demography, the focus of studies has been to use mortality as the type of health measurement, whereas states of health such as morbidity are more popular in other areas of social science, particularly sociology.

In terms of the direct effect, several studies have provided empirical evidence that levels of schooling and their effects on health are substantial. Majority of studies that examine this association have provided strong evidence for education's protective effects on self-rated health and mortality. In one of the most extensive discussions and tests of the effects of schooling on health, Ross and Wu (1995) found that self-rated health and physical functioning are strongly and positively associated with education, both cross-sectionally and over time. When we follow the trajectory of education's impact on health, longitudinal studies have found that higher levels of schooling are mostly advantages in the earlier years, peak in midlife, and then do not show much of an effect in the later years. However, studies in this area are few, fairly new, and have not differentiated between age and cohort effects. There is one recent study which examined both age and cohort differences on the effect of education on self-rated health. Lynch (2003) found that the probability of reporting fair or poor health increases across age but decreases across birth cohorts, indicating that education's buffering effects on health are more apparent for individuals who were born later than earlier in the twentieth century. More recent research on education and health has also extended to mental health, physical impairments, and physical health outcomes or examines these effects concomitantly. For example, in a study that examined educational differentials in functional and mental health over the life course, Kim (2008) found that among those with less than a high school diploma had higher physical impairment. Moreover, individuals with a high school diploma and higher educational attainment measured by college degrees had more favorable trends in depression which varied by age over the life course.

As regards education's buffering effects on mortality, the inverse relationship between education and mortality has been well established since the famous Kitagawa and Hauser (1973) study which found significant differences in mortality rates across educational categories for both men and women. In one of the more recent studies that substantiated the Kitagawa and Hauser study, Elo and Preston (1996) found that college graduates tend to have lower mortality compared to high school graduates and that people who did not attend high school have higher mortality. They also found that the mortality of those who attended but did not complete college did not differ significantly from that of high school graduates, except for females aged 65–89. Interestingly, they discovered a large educational mortality differential between blacks and whites between the ages of 25 and 64. Most studies that research status variations for education and health find that

education may or may not have the same protective factors for marginalized groups. Further evidence that substantiates the varying relationship between education and mortality by social status contends that blacks are at an increased risk for all the causes of death that are atherosclerosis related (Howard *et al.*, 2000), leading to cardiovascular-related causes of death which are most prevalent in the population, particularly for individuals with less-than-average educational attainment. Other studies that have substantiated this conclusion have provided evidence for self-rated health and mortality. For example, in a study that examined the relationship between race, SES, and self-rated health, Farmer and Ferraro (2005) found that over time, the health of blacks declined faster than whites. For blacks and whites of average age and education, their health is similar; however, the process operates at higher levels of education – for every year past the average level of education (normally 13 years) Blacks experienced a health deficit that increased with age. Popular conclusions about the differential benefits of education between blacks and whites are not well explained or understood. It is assumed that the decreased ability of racial minorities at any given level of education to translate educational resources into good health includes the influence of racial discrimination on access to healthcare and the lack of accumulation of social capital over the life course. However, these conclusions are yet to be empirically challenged.

In the sociological and social science literatures, the important mediating mechanisms that are believed to drive the relationship between educational attainment and health are economic, social, psychological, and behavioral. The relationship between education and economics is strong because in the United States, the more education that is acquired allows of higher-paying jobs, more prestigious occupations, and the means to purchase the necessary goods to support a comfortable or privileged lifestyle. If we look solely at the advantages of having a steady income and its impact on health, it improves health through the ability not only to support ones self financially, but also to purchase health-enhancing goods – such as health-care services and nutritious food – live in cleaner environments, and reduce stress. Also in the United States, the more schooling a person acquires allows of higher-paying positions that offer better health insurance and other employee benefits. Health insurance coverage is basically fee for service, where a significant proportion of Americans use employee-sponsored healthcare to pay for health insurance. For those without employee-sponsored healthcare and meet the requirements necessary for subsidized care from the state or government, programs such as Medicare and Medicaid pay for doctor visits and hospitalization. However, individuals who rely on these two programs are more likely to be poor, elderly, reside in lower SES neighborhoods, or are concentrated in urban areas that do not provide similar quality of care compared

to hospitals and clinics in more affluent areas. Moreover, a significant proportion of individuals residing in low-SES areas can be considered the working poor, that is, they work full or part time but their employers do not offer health coverage or other benefits. Moreover, their salaries are too high to qualify for subsidized medical coverage by the state or government. Typically, the demographics of this population are low income, having less than a high school degree, being members of marginalized racial/ethnic groups, and being residentially segregated. Education's impact on health is significant because the well educated are less likely to be unemployed, more likely to work full time, have higher incomes, and experience less economic hardship, which are all protectors against adverse health outcomes (Sewell and Hauser, 1975).

Education's impact on health also has a social psychological component. Education is beneficial to health because it allows of an individual to develop a sense of personal control, social support, and coping mechanisms needed to minimize stress. Education develops one's sense of self-efficacy because it allows individuals to gather and interpret the necessary information to control events and situations in life. Lack of personal control affects health through physiological mechanisms because experiences of uncontrollability are associated with suppression of the immune system (Rodin and Timko, 1992). The body's immune system is designed to defend the body against foreign or dangerous substances that invade it and when individuals are bombarded with untoward stress, it suppresses the immune system, thus increasing the susceptibility to disease. Since psychological stress has proved to be associated with the breakdown of physiological systems which protect the individual against disease, the regulation of stress is imperative for overall good health. Research has shown that people with different levels of schooling have various coping resources and strategies and social support to avoid the deleterious effects of stress. For example, Ross and Wu (1995) found that individuals with higher levels of education were more likely to report having supportive relationships, less likely to practice adverse health behaviors, and had greater life satisfaction. Berkman and Glass (2000) found that social networks reduce not only physiological stress responses but also health-damaging behaviors and encourage health-promotion behavior. Education provides individuals with resources needed to make life easier which in turn reduces stress and operates to protect health. The better educated have a higher sense of control, and better life conditions allow them to adopt preventive lifestyles against diseases in cases where lifestyle is uncovered as important for health (Mirowsky and Ross, 2005b).

Educational attainment has a direct effect on health not only through income potential and psychological mechanisms but also through knowledge about behaviors that promote healthy lifestyles (Backlund and Keller, 1995),

and having access to information about technological advancements in health innovations. Adverse health behaviors account for nearly 10% of the variance in SES differences in morbidity and mortality; therefore, it is important to examine how lifestyle mediated through levels of schooling is associated with health outcomes. The most common health behaviors examined in the literature are smoking of tobacco and substance abuse, such as excessive drinking of alcohol. The relationship between adverse behaviors, education, and health contends that individuals with lower levels of education are more likely to practice these behaviors because of cultural influences on the acceptability of smoking, stress, community norms, and parental/peer influences. In a study that examined how area-level characteristics could affect smoking by serving as a source of stress to local residents, Tseng *et al.* (2001) found increased odds of smoking for black women with low, average education. Moreover, smoking has a significant impact on health because it compromises the operation of physiological systems that operate to protect the body against disease. Research has shown that the direct links to poor health practices, such as smoking, can increase the risk of coronary heart disease, stroke, atherosclerosis, aneurysms; lung and other cancers, including that in esophagus, pancreas, bladder, larynx, and cervix; emphysema, bronchitis, pneumonia, and other respiratory infections; as well as liver disease and burns (Ross and Wu, 1995). Public service announcements and programs to educate the community have been employed to reduce smoking behavior and resulting deaths from tobacco use. Research has shown that individuals with more years of schooling tend to have access to information which chronicles negative risk factors associated with adverse health behaviors, particularly from the health consequences associated with smoking tobacco. Majority of programs that operate to educate individuals about the negative consequences of smoking have made the most considerable impact on young adults. Recent evidence has suggested that college graduates have made gains in improving important health-related behaviors, such as smoking cessation.

The health consequences associated with heavy drinking are similar to those of smoking. Alcohol use and other drugs are major risk factors in the prevalence of acquired immune-deficiency syndrome (AIDS), violent crimes, child abuse and neglect, and unemployment (Hawkins *et al.*, 1992). Not only is excessive drinking detrimental to health at older ages, but also it interferes with educational trajectories if started in early years. Alcohol consumption's effect on educational attainment has been studied primarily at the K-12 level showing the negative consequences for life chances and well-being. Research has shown that alcohol abuse is most detrimental to health in adolescence and early adulthood since these are the major years of physiological, cognitive, and physical growth. Furthermore, drinking in early life is positively associated with alcohol

abuse over the life course. A recent report from a 25-year longitudinal study of inner-city youth found educational level and school dropout rates to be a predictor of later alcohol consumption (Crum *et al.*, 1998). The concluding consensus concerning alcohol use and education is that the effects are cyclical. Excessive drinking influences dropping out of school, which in turn leads to low self-esteem, which encourages more alcohol use and other behavioral problems at older ages. In addition, reduced educational achievement increases stressors such as financial or employment problems which in turn elevate alcohol use as a coping mechanism.

More educated people are better informed about health-related innovations (National Science Board, 2000). Several studies have documented that educational attainment substantially increases the propensity to seek care from medical specialists, even after controlling for health status, and in countries with national health insurance systems. For example, in a study that examined the association between education gradients, health-related innovation, and all-cause mortality, Gleid and Lleras-Muney (2008) found that the education gradient was steepest for those diseases that have seen the most health-related innovation. The effect was most substantial for mortality conditional on cancer diagnosis for both men and women. In another study, Lleras-Muney and Lichtenberg (2002) show that education is positively correlated with the use of new drugs recently approved by the Food and Drug Administration. The ability to access drugs to prolong life and continue taking medication is imperative when trying to treat or manage health conditions. Goldman and Smith (2005) found that among individuals with diabetes and those infected with human immunodeficiency virus (HIV), the best-educated patients were more likely to adhere to new complex treatment regimens.

Research has shown that the various mechanisms in which education operates to influence health are direct and/or mediated through social mechanisms. In the United States and developed countries, where SES stratification shapes life outcomes, educational attainment is important not only for economic advancement, conditioning social psychological impacts on health, and knowledge about the negative effects of adverse behaviors, but also for improving access to health innovations responsible for promoting health and well-being. Although literature has proved that levels of schooling are directly linked to good health, whether this association is strong for various health outcomes in various social contexts outside of the United States, especially in underdeveloped countries, remains understudied.

International Focus

An important question to ask is whether educational attainment can be an effective measure of SES when

examining the underlying causes of health stratification in the global community. Researchers have found health inequalities within every country studied and these inequalities vary considerably from country to country (Feinstein, 1993). Although different countries have stratification systems based on various criteria, such as skin color, income, occupation, and gender, can we use education as an appropriate tool for measuring change or stability in health status? Not all countries use educational attainment as a reliable indicator of SES because their stratification systems are different from those of the United States. For example, an overwhelming majority of studies that examine the root causes of inequalities in health in the United Kingdom and most European countries have found a link between occupational status and health outcomes with individuals on the lower end of the occupational hierarchy exhibiting worse health compared to the more privileged. In countries such as Brazil and South Africa under apartheid that base their stratification systems on subjective measures such as skin color, individuals who bear the racial marker, particularly with darker skin tones on average, have worse health outcomes compared to the rest of the population. Moreover, gender stratification exists around the world but have different meanings, depending on the severity and history of gender subjugation within a respective country or region. Of the few studies that have found a strong association between educational stratification and the health gradient over the past 30 years, intergenerational processes seem to be the most popular when assessing how SES impacts on health. Researchers have found that mothers' education plays a major role in reducing child and infant mortality. A considerable body of research suggests that maternal education is the single most important factor in explaining differentials in child health outcomes more important than paternal education, health service availability, and SES measured by income and occupation. The educational level of the mother is closely linked to both her own and her household's socioeconomic conditions as well as to other complex factors relating to her self-esteem, coping ability, and competence in mobilizing resources for herself and offspring (Bicego and Boerma, 1993). In the past 40 years, child mortality has declined with maternal literacy for Costa Rica, Mexico, and Bangladesh. Research has provided strong evidence that regardless of social context, educational attainment of mothers matters most for children's health and well-being.

Conclusion

Education is the most influential investment a country can make because it improves skills, wages, economic well-being, contraception, living conditions, health of children and adults, and life expectancy (Summers, 1992). At this

particular point in time, the direct and mediating mechanisms conditioned by educational attainment that impact on states of health and well-being speak volumes about how lived opportunities are important at all stages of life. What has been missing in the discourse on educational attainment and health is a clear consensus on whether for some status groups, particularly highly educated blacks in America or marginalized groups in underdeveloped countries, the degree to which schooling translates into positive returns to protect health over the life course. Moreover, besides intergenerational advantages, how does education operate to protect health for adult women in various social contexts or countries around the world? More life course studies of the returns of education on physical and mental health outcomes in the United States and abroad are needed to adequately assess the various mechanisms involved in promoting population health. An important task is to be critical of the meaning of measures of schooling since its interpretation can vary depending on social context. Further, future research should pay attention to how social context, mainly where people live, conditions educational opportunities which are important for health. More recent studies have found that place characteristics, such as noise, quality of neighborhoods, residential segregation, and other risk factors, are important for understanding segments of population health. As a global community, we should focus on ways to reduce educational inequality and create policy that aims to equalize educational opportunities since they are integral to good health over the life course.

Bibliography

- Backlund, S. and Keller, J. B. (1995). U.S. mortality by economic, demographic and social characteristics: The national longitudinal mortality study. *American Journal of Public Health* **85**(7), 903–905.
- Berkman, L. F. and Glass, T. (2000). From social integration to health: Durkheim in the new millennium. *Social Science and Medicine* **51**(6), 843.
- Bicego, G. T. and Boerma, J. T. (1993). Maternal education and child survival: A comparative study of survey data from 17 countries. *Social Science and Medicine* **36**(9), 1207.
- Crum, R. M., Ensinger, M. E., Ro, M. J., and McCord, J. (1998). The association of educational achievement and school dropout with risk of alcoholism. *Journal of Studies on Alcohol* **59**(3), 318.
- Elo, I. T. and Preston, S. H. (1996). Educational differentials in mortality: United States, 1979–85. *Social Science and Medicine* **42**(1), 47–57.
- Farmer, M. M. and Ferraro, K. F. (2005). Are racial disparities in health conditional on SES? *Social Science and Medicine* **60**, 191–204.
- Feinstein, J. S. (1993). The relationship between socioeconomic status and health: A review of the literature. *Milbank Quarterly* **71**(2), 279–322.
- Gleider, S. and Lleras-Muney, A. (2008). Technological innovation and inequality in health. *Demography* **45**(3), 741–761.
- Goldman, D. and Smith, J. P. (2005). Socioeconomic differences in the adoption of new medical technologies. *American Economic Review* **95**(2), 234–237.
- Hawkins, J. D., Catalano, R. F., and Miller, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. *Psychological Bulletin* **112**(1), 64–105.

- Howard, G., Anderson, R. T., Russell, G., Howard, V. J., and Burke, G. L. (2000). Race, socioeconomic status and cause-specific mortality. *Annals of Epidemiology* **10**, 214–223.
- Kim (2008). Intercohort trends in the relationship between education and health. Examining physical impairment and depressive symptomatology. *Journal of Aging and Health* **20**(6), 671–693.
- Kitagawa, E. M. and Hauser, P. M. (1973). *Differential Mortality in the United States: A Study in Socioeconomic Epidemiology*. Cambridge, MA: Harvard University Press.
- Kitagawa, E. M. and Hauser, P. N. (1960). *Differential Mortality in the United States: A Study of Socioeconomic Epidemiology*. Cambridge, MA: Harvard University Press.
- Lleras-Muney, A. and Lichtenberg, F. R. (2002). The effect of education on medical technology adoption: Are the more educated more likely to use new drugs? *NBER Working Paper No. 9185*. Cambridge, MA: National Bureau of Economic Research.
- Lynch, S. M. (2003). Cohort and life-course patterns in the relationship between education and health. *Demography* **40**(2), 309–331.
- Mirowsky, J. and Ross, C. E. (2005b). Education, learned effectiveness and health. *London Review of Education* **3**(3), 205–220.
- National Science Board (2000). *Science and Engineering Indicators–2000*. Arlington, VA: National Science Foundation. <http://www.nsf.gov/statistics/seind00/pdf/append/c8/at08.pdf> (accessed October 2009).
- Rodin, J. and Timko, C. (1992). Sense of control, aging and health. In Ory, M. G., Abeles, R. P., and Lipman, P. D. (eds.) *Aging, Health and Behavior*. Newbury Park, CA: Sage.
- Ross, C. E. and Wu, C. (1995). The links between education and health. *American Sociological Review* **60**, 719–745.
- Sewell, W. H. and Hauser, R. M. (1975). *Education, Occupation and Earnings*. New York: Academic Press.
- Summers, L. (1992). The most influential investment. *Scientific American* **270**, 132.
- Tseng, M., Yeats, K., Millikan, R., and Newman, B. (2001). Area-level characteristics and smoking in women. *American Journal of Public Health* **91**(11), 1847.
- Anderson, E. and Massey, D. S. (eds.) (2001). *Problem of the Century: Racial Stratification in the United States*. New York: Russell Sage Foundation.
- Berkman, L. F. and Kawachi, I. (2000). *Social Epidemiology*. New York: Oxford University Press.
- Charles, C. Z., Dinwiddie, G., and Massey, D. S. (2004). The continuing consequences of segregation: Family stress and college academic performance. *Social Science Quarterly* **85**(5), 1353.
- Collins, P. H. (1991). *Black feminist thought*. New York: Routledge.
- Fang, J., Madhavan, S., Bosworth, W., and Alderman, M. H. (1998). Residential segregation and mortality in New York city. *Social Science and Medicine* **47**(4), 469–476.
- Geronimus, A. T. (1992). The weathering hypothesis and the health of African American women and infants: Evidence and speculations. *Ethnicity and Disease* **2**, 207–221.
- Jeeves, A. (2001). Public health in the era of South Africa's syphilis epidemic of the 1930s and 1940s. *South African Historical Journal* **45**, 79–102.
- Jochelson, K. (2001). *The Colour of Disease: Syphilis and Racism in South Africa, 1880–1950*. New York: Palgrave Macmillan.
- LaVeist, T. A. (1996). Why we should continue to study race...but do a better job: An essay on race, racism and health. *Ethnicity and Disease* **6**, 21–29.
- Myer, L., Ehrlich, R. I., and Susser, E. S. (2004). Social epidemiology in South Africa. *Epidemiologic Reviews* **26**, 112–123.
- Nightingale, E. O., Hannibal, K., and Geiger, H. J. (1990). Apartheid medicine: Health and human rights in South Africa. *Journal of the American Medical Association* **267**, 2097–2102.
- Omi, M. and Winant, H. (1996). *Racial Formation in the United States: From the 1960s to the 1990s*. New York: Routledge.
- Williams, D. R. and Collins, C. (2001). Racial residential segregation: A fundamental cause of racial disparities in health. *Public Health Reports* **116**, 404–416.

Relevant Websites

- <http://www.gatesfoundation.org> – Bill and Melinda Gates Foundation.
- <http://www.cdc.gov> – Centers for Disease Control Office of Health Disparities.
- <http://www.globalhealth.org> – Global Health Council.
- <http://www.hhs.gov> – United States Department of Health and Human Services.

Further Reading

- Adler, N. E., Boyce, T., Chesney, M. A., et al. (1994). Socioeconomic status and health: The challenge of the gradient. *American Psychologist* **49**, 15–24.

To Go or Not to Go: Access and Barriers to Ethnic Minority Participation in Higher Education within the United Kingdom

P Brug and T L Oliver, St. Mary's University College, Twickenham, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

A-level – An A-level, short for advanced level, is a General Certificate of Education usually taken during further education. It is a noncompulsory qualification taken by students in England, Wales, and Northern Ireland (the equivalent to the final two years at high school in the United States).

BSc – Bachelor of science.

Oxbridge – Pertaining to a historically and traditionally elite university, most often with reference to students or graduates of Oxford and Cambridge University.

Post-1992 universities – The term has more recently been used to describe any of the former polytechnics, central institutions, or colleges of higher education that were given university status by the government in 1992 or subsequently.

UK degree classifications – 1st: equivalent to an A grade in the United States; 2i: equivalent to a B grade in the United States; 2ii: equivalent to a C grade in the United States; and 3rd: equivalent to a D grade in the United States.

Widening participation – The concept of encouraging and allowing the underrepresented to partake in higher education.

The 2008 US election, which saw Barack Obama elected as President of the United States, led to many discussions in the United Kingdom about the possibility of a black or Asian Prime Minister. Trevor Philips, head of the Equality and Human Rights Commission, noted that while President Obama's win was impressive, he was unsure that had Obama lived in the United Kingdom, he could have managed to become head of government. Noting in particular the roadblocks within the Labour Party, which has traditionally been more in line with the Democratic Party in the United States, Mr. Philips said he would be surprised, "if even somebody as brilliant as him (Obama) would have been able to break through the institutional stranglehold on power within the Labour Party" (Alice Thomson and Rachel Sylvester, *Times*, 8 November 2008). Politics in the United Kingdom has often been viewed analogously as a machine, not only involving race and

ethnicity but also class and gender; a machine thought to be fueled by an elite group of Oxbridge graduates. Access to these prestigious universities is often thought to be restrictive and highly selective, with the vast majority of these institutions' diversity coming from international students, who contribute more to the (financial) bottom line of these institutions. However, diversity as the result of permanent immigrant and ethnic groups already within the United Kingdom is not so well represented among all groups at Oxbridge and other top universities in the United Kingdom.

The United Kingdom has witnessed a steady increase in the number of people attending University to obtain a (first) degree qualification (i.e., bachelor's degree). In the West (e.g., United States, United Kingdom, and the Netherlands), this can be attributed, in part, to the extension of mandatory schooling until the age of 16, as well as to the decline in the manufacturing sector (see Billington *et al.*, 1998). In the years between 1994 and 2006, first-year enrolment in the United Kingdom increased by over 50%. While the number of students attending university in the United Kingdom (approximately one in three) is less than in the United States (approximately one in two), the growing trend in the United Kingdom is reflective of growing efforts to increase higher educational opportunity to more students. This includes students from underrepresented ethnic groups as well as those from lower-income and working-class families. To some extent, the growing rates of university entry are in line with trends following the 1980s that saw a sharp decline in manufacturing in the United Kingdom. Widening participation schemes that looked to increase access to higher education focused on more than traditional minority ethnic groups, which are the focus of this article.

It has been noted in previous research (see Barn, 2001) that children from ethnic minority groups that attend university tend to come from families that place a high value on education. This is evident from observing the breakdown of incoming first-year students. It indicates that while whites represent just over 92% of the United Kingdom population, they only represent about 76% of the UK student population in higher education, while black African and Asian Indian students are overrepresented in these figures. This may speak well of diversity in the United Kingdom, however it needs to be interpreted with caution. Minority students with lower qualifications are more likely to be unemployed than their white counterparts, which may

serve as a catalyst for continued education. In addition, the figures may represent regional factors. The majority of ethnic minorities in the United Kingdom live in and around London (over 60%); this is reflected in the fact that Greater London has an ethnic minority population of about 30%, far above the 7.9% overall representation. It is likely that students within this region have higher rates of educational attainment because of the economic situation (e.g., need for skilled labor) and the possible cultural emphasis on education that is part of the regional divisions in the United Kingdom. As such, the data may not depict an overrepresentation of certain ethnic groups, but an underrepresentation of the white population.

Further to the above-mentioned issues, in the period from 1994 to 2006, the increase in student numbers varied across the main ethnic groups. Although all groups saw increased numbers, the representation within higher education remained relatively unchanged for most groups; (see Table 2) dramatic increases were only seen for black Africans and to a lesser extent Asian Indians. The significance of these statistics is not clear, so any interpretation needs careful consideration. The increase number of Black African students could be the result of an increase in the overall numbers of Black Africans in the United Kingdom. In addition, the number of students entering university is not always representative of the number of students obtaining qualifications (e.g., BA or BSc) and the classification of those qualifications (e.g., 1st or 2ii).

Data indicate that degree attainment and classification vary considerably by ethnic group. Ethnic minorities are more likely to enter university with qualifications other than advanced levels (A-levels) and go on to have lower rates of obtaining higher degree classifications. In addition, many of the black ethnic minority population are more likely to be enrolled in post-1992 universities, considered to be less prestigious than traditional universities (e.g., University of Cambridge and St. Andrews University). This is not the case for Asian Indians and Chinese, who are well represented at the latter type of institutions. It is also more likely that black populations will enter higher education as mature students, which itself brings up concerns such as being employed while studying and supporting a family. The data further reflect that black Caribbean males are under-enrolled as are Bangladeshi females. These factors draw attention to the fact that while minority representation in higher education is relatively strong, in relation to their representation in the general population, their success (e.g., degree classification) does not mirror this pattern.

These facts reflect a clear need for ethnic groups to be discussed individually, as groups such as the Asian Indians differ in success, as compared to the black Caribbean group. Given these factors, this article does not attempt to cover all aspects on ethnic representation in higher

education, but instead focuses on highlighting general areas of concern. This may explain the previously mentioned disparities and offer some guidance for both practical consideration and further research.

Addressing the issue of diversity on university campuses in the United Kingdom requires, at least, a threefold approach. The first issue is qualifications and the restrictions that are created at the primary- and secondary-school level that make access to university difficult for certain sectors of the United Kingdom population. The second area relates to the support and encouragement from parents and peers. Here, the concern focuses on different levels of attainment for various ethnic groups. The third area of concern is the campus climate – how do particular groups make the transition to university life? This leads to concerns over retention rates and the level of academic and social support available to students. The following sections look at these areas and attempt to present a better idea of the reality of diversity on the university campus and possible ways of addressing key concerns.

Primary and Secondary Education

The reality of higher education is that it is part of a continuous educational system, requiring prior education in order to gain entry to institutions of higher learning. In the United Kingdom, this is most often achieved by studying at A-levels. While not all university students in the United Kingdom have studied A-levels, it is the best route to higher education. However, in recent years, there has been concern about the relative increase in the number of students getting top marks (e.g., four A grade A-levels), whereby universities have had to reject highly qualified students. To this extent, some universities (e.g., Imperial College, University of Oxford) have discussed the idea of separate exams for potential students or higher entry requirements (e.g., from A to A*). However, the reality for many of Britain's ethnic minorities is that they often do not achieve minimal standards. Therefore, an increase in entry requirements will further disadvantage those who are struggling to meet the current requirements. In 2000, while Asian Indians had some of the best A-level and GCSE results, other Asian groups, particular the Pakistani and Bangladeshi groups, did not perform as well. The results were similar to the latter for black Africans and Caribbeans.

The complexity of this divergence in academic progress based on ethnicity demonstrates why the various ethnic groups need to be studied independently. Discussing them as an autonomous homogenous entity does not give an accurate reflection of the different groups. Therefore, it should be noted that the following discussion focuses on issues surrounding the under performance of

many of the ethnic minority students and the research into the causes of the disparities keeping in mind that individual group differences may exist. The concern regarding underperformance often involves two key factors: relative socioeconomic depravity and institutional dynamics within the system.

Social Depravity

Underperformance among ethnic minority students is often seen to reflect a cycle of poverty, whereby children are born into poor families, live in socially deprived areas, go to poorer-quality schools, tend not to go to university, and get low-paying jobs, and so the cycle regenerates itself.

The discussion of social depravity with regard to education comes across at some level as ironic. Research has suggested that education, particularly higher education, is linked with improved future earnings. In many ways, the early notions of widening participation in the United Kingdom can be viewed as attempts to replace apprenticeships with higher education (e.g., Kettley, 2007), thereby creating the opportunity for higher earning potential among students from less-fortunate families, who would have been more likely to have taken up apprenticeships. For the most part, students themselves see (higher) education as a means of securing an improved financial status (e.g., Patiniotis and Holdsworth, 2005).

In the United Kingdom, like other places (e.g., the United States), there is often a strong correlation between ethnicity (and race) and poverty. Statistics show that white students are less likely (25%) to come from low-income households compared to minority ethnic group members (*c.* 45%). However, individual variation exists among the various minority ethnic groups. The high rates of poverty for the Pakistani/Bangladeshi group may in some part be a result of the fact that this group also has the largest household size (average 4.5 people) compared to other groups: Indian (3.3 people) and white and black households (2.3 people). Larger household sizes are often a reflection of multiple generations living in a single household. At the opposite end, black Caribbean families had the highest rate of single-parent families, another factor often associated with poverty.

The irony of the poverty–education connection is that while education is seen as a vehicle for improved financial circumstances and opportunity, educational outcomes are very much tied to the current economic situation. Reports and investigations into the poverty–education connection continue to show that a family's socioeconomic status influences members' educational participation. For instance, young people from middle-class families are more likely to remain in education longer and pursue higher education than their peers from skilled or unskilled families (e.g., Gayle *et al.*, 2002). This has important

implications for policymakers and society, as this suggests that individuals' social class can hinder them from developing their academic ability.

The British government is attempting to make university education more accessible to the lower classes, as evident in the suggestion by Egerton and Halsey's (1993) that it is now a requirement for many institutes to accommodate more students from the lower classes. Some scholars have argued this can reduce the credibility of university education and may be symbolic of the divide between the more historic traditional universities and the post-1992 universities. However, the concern over poverty is not limited to a notion of the poverty cycle but also that poverty has direct bearing on the institutional dynamics of the educational system. These dynamics include: (1) the quality of education students receive, (2) the conditions of schools, and (3) other dynamics, such as discrimination.

Institutional Dynamics

Recent discussion in the UK news media has highlighted concerns over the quality of teachers in poor areas, particularly in the inner city. Teachers were said to possess lower qualifications and have fewer years of experience. There appears to be a consensus in the literature with regard to the impact poverty has on students' educational experience. Students who attend schools that reflect the economic deficiency of the area have lower results than students in advantaged areas. These same students also appear to have lower results than similarly deprived children who are able to attend more privileged schools (Yu and Taylor, 1997). Furthermore, deprived students often have a less positive attitude toward education and schooling (Horgan, 2007). In addition, teacher expectations of students in deprived schools are often thought to be lower, whereby students are often not encouraged or meant to feel positive about their potential. This relates to another concern regarding institutional discrimination.

The research clearly shows there is a pattern within the educational system, either by coincidence or somehow created by the design of education (in the West): minority ethnic students often feel discriminated against, more so than their majority student counterparts. Whether this discrimination is real or grounded in misunderstanding, the impact it has on the students is often detrimental. A common outcome is academic disengagement which can have a negative impact on student performance. The other concern with regard to institutional discrimination is that minority students, particularly black male students, are more likely to be excluded from school. Several studies have shown this high rate of exclusion may in some part be due to overreactions by educators based on negative stereotypes of black males (see Barn, 2001).

Culture of Higher Education: Influence of Family and Friends

Common knowledge has often assumed that university graduates produce university graduates. This logic not only fits in with the idea of the poverty cycle but also underlies the influence of parents as role models. Within families where neither of the parents entered higher education, the lack of so-called role models could make it less likely that children in these homes would themselves go to study at university. Clearly, the influence of parents is quite strong if at least one of the parents has a university degree (e.g., Gayle *et al.*, 2002). Many of the UK's minority ethnic population come from households where neither parent went to university. In some cases, this is a result of the parent's native country, where education was either restrictive or poor in quality (Louie, 2005). As a result, many of these parents migrated to the United Kingdom in part to provide their children with better educational opportunity. As parents play an important role in their children's educational potential and the likelihood that they will remain in the education system longer (Urdan *et al.*, 2007), the importance placed on education by minority ethnic parents may go a long way to influence their children's choices. Even though ethnic minority students at times can be frustrated by the level of pressure put on them by their parents; they nonetheless come to appreciate it (see Louie, 2005).

The role of parent is also important in giving students information with regard to what to consider when it comes to figuring out what to do, applying for university, and considering funding support. Students whose parents do not have higher degree qualifications may choose not to seek out their parents for input, because they perceive them to be less knowledgeable about this matter. This can then place a greater reliance on the schools (and sixth form colleges) to inform students about the cost and benefits of higher education. The expectation that schools will fill this gap may be a bit presumptuous and misconceived as students are likely to identify more with their parents' situation (Trusty, 1991). In addition, concern could also be raised that better schools where families with higher education levels attend are more likely to provide better information about the benefits of further education.

Despite the role of parents, adolescents are also influenced by peers. Whether as a result of peer conformity or simple social comparison, friends can have a strong impact on the educational attainment of a student. This concern is often discussed with reference to potential detrimental impacts. For example, when it comes to black Caribbean students, particularly males, there has been discussion about the belief this cohort either directly or indirectly discourages one another from serious participation within the educational system – the idea being

education is a 'white thing.' However, this is not solely a black issue; negative peer influence can be found in various settings (e.g., urban and rural) and among various populations (e.g., white and Asian; Louie, 2005). It is not possible for this article to offer extensive coverage of this complex issue. However, what is worth highlighting is that the support and encouragements, by peers, to attend university may vary for some populations within Britain's ethnic communities. This is likely to become an issue if there is also no proper and sufficient information and encouragement come from within the schools (e.g., teachers) and/or from home (e.g., parents).

There is also the concern that peers can have the potential to undermine the influence of parents; nevertheless, it has been noted that while teens are spending more time with their peers, who tend to have more influence on an adolescent's day-to-day behavior, they are more likely to take on values, like the importance of education, from their parents (Noller *et al.*, 2001). Furthermore, it should not be assumed that peer influence is always negative (e.g., drinking or smoking); research has shown that peers can help improve student performance by setting challenges, such that an individual works harder to perform as well as their close friends (Azmitia and Cooper, 2001). In some cases, this may be reflected in the fact that students who go to university are more likely to have close friends who are also going to university.

The key concern here is that students from disadvantaged backgrounds may not have equal access to information about university studies or are not given enough opportunity to carefully consider university as an option. In the United Kingdom, there are growing attempts at opening up opportunity, primarily through widening participation schemes. These schemes often involve visits to campus by students or visits to schools by university representatives. These are noble schemes but their impact is yet unclear. It is likely any serious impact, in the most-needed areas/schools, will only occur if a more sustained and comprehensive approach is undertaken, as opposed to a few sporadic interventions. However, getting into university is half the battle, once there, students should want to remain and feel a part of the academic community. Therefore, attention has to be paid to the campus climate.

Campus Climate

The issue of diversity in higher education is complex. Figures may reflect an increase in the number of minority applicants, as well as an increase in the number of first-year students from minority ethnic backgrounds (particularly black African and Caribbean). However, lower retention rates and mediocre results among certain groups highlight the fact that entry into higher education is not where the

concern and attention should end. While Asian Indian students, for example, do well at university, black and some of the other Asian students, in general, tend to perform average or worse, with few able to obtain higher honor degrees (i.e., 1st). While some of this may relate to the previous sections (quality of secondary education) other aspects may be a result of campus life.

For many ethnic minority students the demographics of the university may be in sharp contrast to the demographics of their previous school and the neighborhoods where they grew up. This is an additional adjustment that these students have to cope with, as part of the normal transition to university. Furthermore, many students from ethnic minority as well as low-income families are choosing to live at home during higher education. This is in large part a way to reduce the cost of higher education. It is also likely that for some ethnic minority students it is a way parents seek to protect their children from the perceived darker side of university life (i.e., partying and sexual relations). The concern here is that students who live at home may not make use of the full support services on campus (e.g., library and tutors). In addition, these students may have less social contact with their cohort and less peer support. This can make the experience of higher education particularly isolating. Even if the students develop close friendships at university, they may not provide an upward social comparison. This type of comparison is thought to offer positive academic support and inspiration, thereby improving academic performance.

Students may also differ in their ability to take part in extracurricular activities, such as sports, student clubs, and student events (e.g., dances and trips). The inability to fully embrace university life may make it more likely that a student feels disconnected and thereby easier to leave, even if not for academic reasons. University student satisfaction among ethnic minority students does not appear to have been studied extensively, but is an area that deserves some consideration with regard to retention. Student satisfaction in many ways may be linked to academic performance in that it offers students a positive attitude about the institution and makes them more inclined to seek support, as opposed to feeling further frustrated by the university and what it has to offer.

Conclusion

It should be the case that if a university accepts a student, they should be in a position to support that student, so they have an equal chance of graduating, as any other student. However, some students may not be supported by higher educational institutions in a manner which best serves them. This article has dealt with some factors that impede the matriculation of large numbers of ethnic

minorities in the United Kingdom. However, what is presented here is a snapshot of a large and more complex situation that deserves more comprehensive attention. Still article should help to highlight and bring to the forefront some areas for consideration, if greater equity is to be established in higher education and increased academic performance.

In this race-income dynamics, one of the key factors will often be the family. While parents may have a strong desire that their children go to college, the attainment of that goal will require more than just desire. While the debate and the struggle to create true equality in primary and secondary education continue, often it will be the level of parental support and involvement which has a real impact on a child's progression through the educational system and the performance outcomes. This is not to put the blame on parents or shift responsibility away from the educational system, but to emphasize the significance of parents in helping breakthrough in the cycle and open up opportunity even if the financial resources are lacking. Neither does this resolve the complexity or eliminate the stress for many families with regard to funding higher education.

The challenges of primary and secondary school education are also vast. While this article has discussed several areas to be addressed (institutional dynamics), these factors need to be teased out and addressed in combination with other factors. They are also factors that are part of the political discourse, which not only complicates matters but also makes a timely solution unlikely. However, because of the connection between lower-level education and higher education, this point cannot be emphasized too much.

One of the first cautions noted in this article was the fact that while the focus would be on minority ethnic students as a whole, the diversity of these group warrants more individual focus on each group, even those deemed to be the model minority. Attempts should be made to assess what lessons can be learned from the individual groups, especially the groups that are having greater success. Another issue in that has not been discussed but that warrants mention is the fact that some researchers would argue that academic performance is tied to intelligence and that some minority ethnic students are going to perform better than others based on their race. However, the merit of this argument fails to be met when considering the Southeast Asian groups. For example, Asian Indians tend to have great success in school and at university, while the results for Bangladeshi students are very different. Therefore, there have to be other reasons that help explain disparity in academic outcomes for ethnic minority students in the United Kingdom. It is vital to look across the various groups and try and assess the differing success and failure in order to see what can be learned and possibly shared.

Bibliography

- Azmitia, M. and Cooper, C. R. (2001). Good or bad? Peer influences on Latino and European American adolescents' pathways through school. *Journal of Education for Students Placed at Risk* **6**, 45–71.
- Barn, R. (2001). *Black Youth on the Margins: A Research Review*. York, UK: Joseph Rowntree Foundation.
- Billington, R., Hockey, J., and Strawbridge, S. (1998). *Exploring Self and Society*. London: Palgrave Macmillan.
- Egerton, M. and Halsey, A. H. (1993). Trends by social class and gender in access to higher education in Britain. *Oxford Review of Education* **19**, 183–197.
- Gayle, V., Berridge, D., and Davies, R. (2002). Young people's entry into higher education: Quantifying influential factors. *Oxford Review of Education* **28**, 5–20.
- Horgan, G. (2007). *The Impact of Poverty on Young Children's Experience of School*. York, UK: Joseph Rowntree Foundation.
- Kettley, N. (2007). The past, present and future of widening participation research. *Journal of Sociology of Education* **28**, 333–347.
- Louie, V. (2005). Compelled to excel: Immigration, education, and opportunity among Chinese Americans. *International Migration Review* **39**, 760–761.
- Noller, P., Feeney, J. A., and Peterson, C. (2001). *Personal Relationships across the Lifespan*. Hove, UK: Psychology Press.
- Patiniotis, J. and Holdsworth, C. (2005). 'Seize that chance!' Leaving home and transitions to higher education. *Journal of Youth Studies* **8**, 81–95.
- Urdu, T., Solek, M., and Schoenfelder, E. (2007). Students' perceptions of family influences on their academic motivation: A qualitative analysis. *European Journal of Psychology of Education* **22**, 7–21.
- Yu, C. M. and Taylor, W. L. (eds.) (1997). *Do magnet schools serve children?* Washington, DC: Citizens' Commission on Civil Rights.

Further Reading

- Connell, R. W. (2004). Working-class parents' views of secondary education. *International Journal of Social Psychology* **8**, 227–239.
- Fan, X. (2001). Parental involvement and students' academic achievement: A growth modelling analysis. *Journal of Experimental Education* **70**, 27–61.
- Morgan, V. and Frazer, G. (1992). Parental involvement in education: How do parents want to become involved? *Educational Studies* **18**, 11–21.
- Rutter, M. and Tienda, M. (eds.) (2005). *Ethnicity and Causal Mechanisms*. New York: Cambridge University Press.
- Trusty, J. (1991). Family influences on educational expectations of late adolescents. *Journal of Educational Research* **5**, 260–270.
- Verkuyten, M. and Brug, P. (2002). Ethnic identity, achievement, self-esteem, and discrimination among Surinamese adolescents in the Netherlands. *Journal of Black Psychology* **28**, 122–141.
- Werner-Wilson, R. J. and Arbel, O. (2000). Assessment of interpersonal influences on adolescents: The parent and peer influence scale. *American Journal of Family Therapy* **28**, 265–274.

Relevant Websites

- <http://www.actiononaccess.org> – Action on Access:
The national coordination team for widening participation in higher education, funded by the Higher Education Funding Council for England.
- <http://www.hesa.ac.uk> – Higher Education Statistics Agency.
- <http://www.statistics.gov.uk> – UK National Statistics Publication Hub.

Universal Secondary Education and the Two-Tiered School System in the English-Speaking Caribbean

P Anderson and J Devonish, University of the West Indies, Kingston, Jamaica

© 2010 Elsevier Ltd. All rights reserved.

Overview

The education system of the English-speaking Caribbean bears both the imprint of its colonial origin, as well as the struggle to reverse this heritage of inequality. This continuing struggle has infused efforts across the Caribbean region to reshape educational provisions to ensure greater access and development for the mass of persons who had been disenfranchised through slavery. This section looks specifically at the system of secondary education – examining its structure and performance during the first decade of the current century. The Commonwealth Caribbean Region is generally defined to include the islands in and along the Caribbean Sea, as well as some of the countries located on the adjoining continental mass bordering the Caribbean Sea, which were originally part of the British overseas possessions. The focus is on three of the 17 territories which constitute the English-speaking Caribbean region. These are Barbados, Trinidad and Tobago, and Jamaica.

The three countries which are discussed here share a common history of exploitation, as the system of plantation agriculture was implemented by European powers using forced labor from Africa. However, the trajectories which were followed by their education systems varied significantly as a result of several factors. These included the size of the country and its economic resources, the size and composition of the planter class under slavery, and their relative needs for locally delivered schooling, given the pattern of sending children back to the metropole for their education. In the first period following emancipation (1834–1944), critical factors which fostered the expansion of primary schooling for the majority black population included the role of the missionary churches in providing instruction, as well as the impact of indigenous resistance and protest movements in modifying the outlook of the Planter class with regard to the dampening effects of socialization through education. The Negro Education Act – which was passed in 1834 – committed State support on a phased basis for the development of education for the masses, following the abolition of slavery in these territories. Accordingly – between the efforts of the state and the church – a broad-based system of primary education was established throughout the Caribbean. This system combined two types of dualism: that between the planter class and the dispossessed, and between the church and the State.

World War II ushered in a period in which many of the English-speaking Caribbean territories countries moved to self-government and subsequently achieved political independence, and in this vast restructuring of the society, education played a central role. Far-reaching changes were implemented with the objective of fostering national commitment and identity, removing the separate and highly unequal system of provision for secondary education, and increasing access for the less privileged. As noted by Miller, between the 1850s and the 1950s, secondary education in the region was limited to less than 3% of the school-age population, and depended on the ability to pay.

The 30 years from the 1950s through to the 1970s witnessed huge investments in education. However, the extent of this expansion again varied with the size of the country, its economic resources, and its philosophy with regard to the role of education in nation-building. Nearly 60 years later, Barbados – which is the smallest of these island-nations – holds the distinction of having almost universal secondary education. In 2007, Barbados had a Human Development Index which ranked first among developing countries, and was 31st globally. Trinidad and Tobago – which was a late participant in the system of plantation sugar production under slavery – has the benefit of rich petroleum resources, and was able to expand its education system rapidly during the period 1956–81. In addition, this twin-island state was governed by a single party for 30 years, with a major commitment to decolonization in the society and explicitly, through transforming education. Jamaica – which is the largest of the countries under review – had an extensive plantation system during the period of slavery, and faced the challenge of trying to build a secondary education system on the basis of an inadequate and unequal primary schooling foundation.

In reviewing this period of educational expansion in the Caribbean, Errol Miller observed:

Nationalist politicians gloried in the social transformation of educational opportunities. Each claimed to have offered greater opportunities to the middle and lower classes by building more schools and altering the rules that governed access to the most coveted segments of the school system. The result of this competition between political parties

was a remarkable continuity in educational policies from one government to the next. Each tried to outdo the other by offering more. (Miller, 1992: 37)

From the 1980s, these Caribbean countries had to come to terms with the economic recession which was engendered by the Organization of the Petroleum Exporting Countries (OPEC)-oil-price increases of the early 1970s, and all entered some form of structural adjustment program. Although Trinidad and Tobago had benefited from the influx of petro-dollars – and had used this to invest heavily in educational expansion – there were general cutbacks in the 1980s. The end of the century also saw an increasing concern with issues of educational quality in all of these countries – as evidenced both in low or unsatisfactory student performance in several sectors, and imbalances in resource distribution.

Common to all of these territories is the continuing reliance on external examinations to establish levels of competency at the end of the secondary cycle. This approach to assessment was derived from the integration of colonial education with that in the metropole, as it was essential for the children of the planter class to be able to access tertiary education in Europe, if they had not been schooled there. Success in external examinations continues to play a similar role for the brown and black majority in the Caribbean, and serves as the most important mechanism of credentialing – both for absorption into the upper tiers of local labor markets as well as for being the gateway to tertiary education. From as early as 1863 – led by Trinidad and Tobago – the Caribbean had started to participate in the overseas school-leaving examinations which were designed and administered by the Examination Syndicate of Cambridge University (the Senior Cambridge Exams). This was followed by the Oxford and Cambridge Higher School Examinations around 1917. Once these examinations were opened up to participation by students in private and denominational secondary schools, they provided the foundation for the entire system of accreditation for secondary school students.

Since 1979, these external examinations have been replaced by regional examinations which rely on meeting standards that are both internationally comparable, but also incorporate material that is of greater regional relevance. The regional examining body is known as the Caribbean Examination Council (CXC), which was established – in 1972 – as a fully managed Caribbean entity with the mandate to replace the overseas examinations with local secondary school examinations in the 16 participating territories. The expected performance level at the end of grade 11 is for students to obtain a minimum of five passes on the Caribbean Secondary Education Certificate (CSEC) in subjects that include English Language and, preferably, mathematics. The performance of students in these examinations, therefore,

establishes the overall success level of each territory. It also distinguishes the higher-performing schools in each country. Since these schools are in general the elite grammar schools – whose establishment predates the major expansion of government secondary schools in the postwar period – the external examinations continue to document and legitimate their continuing higher status.

In several Caribbean countries, the expansion of secondary education has been achieved in stages through the addition of years to the secondary cycle. These extra years of instruction may have been tacked on to junior secondary schools which started with only grades 7–9 forms. Alternatively, they may have been added to the All-Age Schools which were in essence an extended primary school that terminated at age 15, or at the equivalent of the ninth grade. While these additional forms represented a logical solution to the need to expand secondary education, they were being grafted on to a structure which was generally much weaker than the comparable secondary schools – whether public or private. The New Secondary schools – as they were labeled – therefore, represented a lower tier in the education system – and in some countries – they did not register their students for the external examinations. The move to upgrade the secondary system, which occurred in the 1990s, was consequently directed at offering a common curriculum in these schools, and to ensure these students could, in principle, appear for the external examinations. This was an essential step if all secondary students were to be positioned to access tertiary-level training.

As long as secondary education remained in short supply – or as long as there were perceived disparities in the quality of education in each tier of the secondary system – a mechanism was required for allocating students to each tier. This was achieved through the operation of the Common Entrance Examination (CEE) – which students appeared for at the end of the primary school period, or before they progressed to the seventh grade. The Common Entrance Examination has been described as a cruel mechanism which exacted harsh emotional costs on its participants and their parents. Nonetheless, despite its recognized limitations, this examination remains in operation in most Caribbean territories. The main exception is in Jamaica – which replaced the CEE with a Grade Six Achievement Test. This exam serves to distribute students to different secondary schools based on a combination of factors, including performance, preference, and location.

The assessment made by the World Bank, in 1993, with regard to access, quality, and efficiency in the Caribbean region still holds relevance. This report noted sharp differences existed between the secondary schools which served those who passed the CEE, and the junior secondary and all-age primary schools which are the destination for those who do not take – or who do not pass – this selection exam. The authors noted that, in many respects,

the large overall secondary enrolment ratios masked what, in most countries, were largely elitist educational systems.

Although the Caribbean territories embarked on their educational systems with the same gender biases which characterized the colonial societies of the eighteenth and nineteenth century, it is important to note that – by the end of the period under review – there were no restrictions against women in the school system, and girls – at present – outnumbered boys in many institutions. The underrepresentation of boys, and their relative underachievement, now became questions of increasing concern.

As the first decade of the twenty-first century draws to a close, issues of educational reform remain at the centre of education policy in the English-speaking Caribbean. The debates which are engendered still revolve around questions of social inequalities, that are not yet fully divorced from racial inheritance.

Barbados

Barbados is the most easterly of the Caribbean islands, occupying a land area of 166 mi² (approximately 432 km²). The census of 2000 recorded a population of 250 010 persons – which reflected an annual growth rate of 0.11% over the intercensal period 1990–2000. Early fertility declines in Barbados – as well as high levels of emigration – have resulted in declining proportions of persons younger than 15 years. This stood at 23% in 2000. The small size of the country, the fairly flat terrain, the early efforts to control population growth, and relatively high income levels have all facilitated the country's efforts to make education widely available.

This tiny island is distinguished by the fact that, when it achieved political independence in 1966, it had an unbroken history of 339 years as a British colony. During this time, the island was exploited for sugar and rum production. Its subsequent economic development since independence has expanded to include tourism, information services, and light manufacturing.

During the period of plantation slavery in Barbados, there was some limited instruction for the enslaved Africans, and this was mainly with the objective of being able to read the Bible. At this point, public education was established primarily for the poorer whites and the free coloreds who could not afford to send their children to England for schooling, or to hire private governesses. Following emancipation in 1834, public education was established for the general population, and, by 1846, Barbados was reported to have primary enrolment levels comparable to the highest in Western Europe. The subsequent development of the primary school system – and its effectiveness – are evident from the findings of a 1982 survey on reading achievement conducted by the Ministry of Education in the primary schools which showed that approximately 85% of

children were functionally literate by the end of the primary cycle.

While mass primary education was achieved by the beginning of the twentieth century, it was not until after the end of World War II that any attempt was made to reform the hierarchical structure which allocated primary education to the masses and secondary education to the elite. In describing educational policy in Barbados prior to this period, Layne observed it was a highly reactionary force that provided the white planter-merchant class and its allies with the skills needed to perpetuate their hegemony over the rest of the population, and to perpetuate the values upon which that hegemony rested.

After gaining full internal self-government in 1954, and adult suffrage in 1959, several reforms were introduced. Among the most important was the establishment of modern secondary schools as a more relevant counterpart to the traditional grammar schools – which were the bedrock of the colonial period – as well as the introduction of a CEE to provide a competitive mechanism for entry to the government grammar schools. Access to secondary education, therefore, no longer depended on the ability to pay. While it increased access to secondary education, the system which evolved was essentially based on two tiers with pronounced inequalities. The two types of secondary schools had different school-leaving ages, and differed markedly in resources and access to trained teachers.

Following 1966 when Barbados attained Independence, education policy shifted radically toward meeting the needs of modernization of the country – with an explicit focus on moving the society from an elitist to an egalitarian society. By 1979–80, the budget for the Ministry of Education accounted for 20% of the government's total capital and recurrent budget. There was a large expansion in secondary school places in order to meet the demand in densely populated areas. However, the country also fell victim to the economic crisis of the 1980s, and was forced to embark on a structural adjustment program. This program was designed jointly by the International Monetary Fund (IMF), the World Bank, and the Inter-American Bank – taking effect in 1990–92. Although expansion continued at a slower pace over this period as a result of the economic contraction, there was, nonetheless, a steady series of reforms which sought to increase access, improve teacher-training, and diversify the secondary curriculum.

In Barbados, basic education is legally defined to include education for children in the age range between 5 and 16 years, and thus it includes secondary education. The Modern Education Act, in 1981, also made education to age 16 compulsory. By the mid-1990s, Barbados could boast of having excellent coverage at the secondary level – with this being estimated at 95% of the age group from 11 to 16 years. At the secondary level there were 33 secondary

schools, and 23 of these were government-run. The most recent data from the 2000 Census shows that, among persons aged 15 or older, nearly four-fifths (79.4%) had achieved at least a secondary education, with 20% of this age-group reporting tertiary training.

This outstanding educational achievement was made possible both because of the country's universal access to primary schooling, and because of the decline in the school-age population, resulting from the earlier fertility declines of the 1960s. Nonetheless, as in all this region, the secondary school system displays lines of fracture based on the academic level of the students who are allocated to different schools through the mechanism of the CEE. The traditional elite schools – which are referred to as the older secondary schools or the government grammar schools – continue to receive the students who perform best on this qualifying entry exam, thus contributing to their continued superior performance on the exit exams (the CXC). The newer secondary schools receive those with significantly lower CEE scores. These differences in student aptitude are reinforced by differences in budgetary allocation, and in the proportion of trained teachers in individual schools within each secondary tier. The government of Barbados has recognized the distortions that result from the operation of the CEE, and has declared its intention to replace the CEE with a new system of transfer from primary to secondary schools, and to implement a common curriculum during the first 3 years of secondary education. However, these reforms are yet to be implemented.

Trinidad and Tobago

The twin-island state of Trinidad and Tobago lies off the north-east coast of South America and comprises the island of Trinidad – which covers 5130 km² (about 1980 mi²) – and Tobago – which is approximately 300 km² in area. Although both islands were seized by the Spanish toward the end of the fifteenth century, Trinidad remained under Spanish control until it was captured by the British in 1797. In contrast, Tobago was the subject of considerable rivalry among European nations, being finally settled by the British in 1802. The history of this period records that the Spanish did little to develop the island during the first three centuries of occupation, once they confirmed there was no gold to be extracted. However, close to the end of their period of occupation, they actively encouraged the large-scale immigration of French planters and their slaves from the neighboring French Islands. Accordingly, Trinidad and Tobago became distinguished by a mixture of European cultures. These were to be later reflected in the education system, and the respective influence of the Church in the provision of education.

The opening of the island to French settlers, in 1783, and the subsequent conquest by the British led to the plantation cultivation of sugar cane with slave labor, along the lines which were characteristic of the other colonies at this period. Following the abolition of slavery in 1834, the continuation of this relatively new plantation system depended on having access to other supplies of unfree labor. This was achieved through the indenture system, which – from 1845 – brought contract workers mainly from India and a few other countries. The large numbers of indentured workers, therefore, added a further dimension to the cultural complexity of this island state, as the Indian immigrants adhered to their Hindu and Muslim religions. After 6 years of the inflow of indentured workers, the Indian population was 5.8% of the entire population. However, by 1891, this community accounted for more than one-third (35.1%) of the total population, which was estimated at 200 000 persons. The most recent census, in 2001, enumerated a population of 1 114 700 persons – with 40% being of East Indian descent and 37.5% of African origin.

In explaining the limited development of local schools during the eighteenth and nineteenth centuries, historians have placed emphasis on the late development of the plantation system in Trinidad, and the small size of the settler class – which was itself internally divided between the French and the British, in addition to the Spanish descendants. In countries such as Barbados or Jamaica – where the plantation system flourished – local schools were usually established for the children of the poorer settlers, or for the free colored population. These often originated through a bequest from a wealthy planter or merchant. The abolition of slavery provided the first major opportunity for the development of schools for the masses, and – as was the case in other former British colonies – this was facilitated by the Negro Education Grant between 1835 and 1845. In the period immediately following emancipation, these schools were established by the churches, and these included the Roman Catholic Church – the stronghold of the French and Spanish settlers – and the Church of England. In addition, schools were founded by the Mico Charity, which was a nondenominational, but Protestant, organization, and which received the bulk of the funding from the Negro Education Grant.

The colonial government's concern to firmly establish the British culture in the island led to considerable unease with regard to the dominance of the church schools, as they continued to teach in their own language. Accordingly, from the nineteenth century, there were tensions between the church and the state in regard to the delivery of education, and this continued through to the twentieth century – with a significant influence on the secondary school system. Following 1868, the system of church-run schools was further elaborated by the work of the Canadian Presbyterian Mission – which concentrated its efforts on

converting the Indian population through a wide network of elementary schools. While it has been argued all of the elementary schools in this period served the traditional function of seeking to maintain a subservient laboring class, the government schools had the additional challenge of trying to integrate this diverse society on the basis of English culture.

At the secondary level, a handful of schools were established between 1836 and 1869 to meet the needs of the white upper class. Such schools were modeled on the public school system of Britain, charged fees, and were not linked with the primary school system. Among the black and colored population, it was only those young men of exceptional intellect who gained access to these prestigious institutions through scholarships. Even when the government established its own secondary school in 1859, this institution – then known as the Queen's Collegiate School and later renamed Queens Royal College – was conceived as meeting the need for superior education for white boys. It was not until after World War II that there was any major expansion or liberalization of secondary school provisions.

The war had been preceded by widespread economic depression and a series of labor uprisings throughout the Caribbean, emphasizing the need for social change. In the case of Trinidad and Tobago, the Butler Riots of 1937 served to force concessions from the ruling classes, and these included recognition of the importance of increasing access to education. Achievements of the postwar period also included the gaining of universal adult suffrage, and the establishment of a ministerial system of government.

The postwar period itself brought about a greater liberalization of ideas concerning the role of education in social development, and this was translated into a major expansion in the provision of primary school places in Trinidad and Tobago. By 1956, it was estimated the coverage of the elementary school population was in the vicinity of 95%. This provided the basis for the radical increase in the number of secondary school places which began from the period of the War. The continuing expansion of the school system – and its re-organization – were part of the lasting contributions of the 30-year administration of the People's National Movement, which commenced in 1956. This was largely due to the overriding influence of its leader Dr. Eric Williams – a fierce nationalist.

This period of Trinidad's education history was associated both with large-scale expansion of secondary education, and the continuing attempt to reduce the role and the voice of the denominational schools in educational provision. Starting from a base of only three governmental secondary schools in 1957, the government expanded their stock to 21 by 1967. This total compared with 23 denominational schools in that year. Although this period was characterized by a process of renegotiation between

church and state in regard to their spheres of influence, an agreement was reached, in 1960, in regard to the incorporation of the private schools into the public secondary system. This became known as the Concordat. Further efforts to curtail the power of the denominational sector were encapsulated in the Education Act of 1966, which brought the church-run schools under the inspection of the Ministry of Education. In addition, the government claimed the right to control the admission policy of these church schools.

The opposition of the government to the denominational sector was based on the strongly articulated view of Dr. Williams that it was the government's responsibility to build a nationalist education system which could provide social integration in this ethnically diverse country, and so to establish a foundation for economic development. The 1973–74 increase in oil prices catapulted Trinidad and Tobago into a position of wealth, and thus provided the government with the resources to advance its ambitious program of education reform.

Other significant changes – which accompanied the expansion of the governmental secondary school system – were the removal of tuition fees and the location of new schools in rural areas which had been previously underserved. The postindependence period – commencing in 1962 – saw the restructuring of the secondary system into two cycles. This comprised a junior secondary school (for children in the age range: 11–14 years) and a senior secondary school (for children in the age range: 15–18 years). It was estimated that, by 1986, nearly 80% of the eligible age-cohort had access to secondary education. In addition – in tandem with the country's decision to embark on a program of heavy industry – technical training became integrated into the secondary school curricula, instead of being concentrated in the technical secondary schools. By 2001, the United Nations Educational, Scientific and Cultural Organization (UNESCO) estimates showed a gross enrolment rate of 71% for the secondary school-age population, with the out-of-school population being estimated at 26%.

The incorporation of larger numbers of students into the public secondary school system was reflected in increasing failure rates, based on the external examinations. Some of these problems were traced to inadequacies at the primary level, as it was argued many students were moving on to the secondary level without having mastered basic literacy skills. Nonetheless, there has been steady improvement at both the primary and the secondary level. Statistics collated by the World Bank showed that, in 2001, among primary school students who were tested at the end of the sixth grade on the Secondary School Assessment Examination, 63% obtained passing grades in English Language and 68% obtained passes in mathematics. This examination was the equivalent of the CEE in other territories. On the CSEC exams which

assessed performance at the end of the secondary cycle, 68% of secondary students obtained passes in English while 44% passed mathematics. Concerns with improving the quality of education, therefore, continue to dominate recent approaches to education in Trinidad and Tobago.

Jamaica

Situated in the Northwest of the Caribbean, the island of Jamaica comprises one of the Greater Antilles – and covers a land area of 4243.6 mi² or 10 989.7 km². The population which was enumerated in the 2001 census comprised 2 607 600 persons – representing an intercensal increase of 0.88% annually.

Having been originally captured by the Spanish in 1494, Jamaica remained a Spanish colony until 1655 when it was seized by the British. Both the Spanish and the English colonial powers exploited the country through the cultivation of sugar cane with forced labor from Africa until 1834 – when slavery was abolished. Although Jamaica's economy, like Barbados, was dominated by sugar cane production, the island differed significantly both in size and in terrain. The mountainous nature of the country limited the land available for plantation production, and once emancipation was achieved, many of the former slaves moved up to the hills where they engaged in peasant farming. These dispersed patterns of settlement were to present challenges in the effort to incorporate the population within the ambit of a basic education system.

As in other Caribbean countries, the education system derived its origin and structure from the plural structure of the society, and educational policies came to reflect the struggles of the dispossessed to assert their rights to social advancement. The strong missionary presence in Jamaica meant these groups led the way in providing primary education for the masses. This set the basis for the dual system of church and state control of primary schools which has continued into the present century. The development of secondary education was also stimulated by the work of various denominations – with the first high school being established by the Roman Catholics in 1850. These early secondary schools were designed as grammar schools along the British model, and catered for the rising middle classes – since the children of the wealthiest classes were already receiving their education in Britain.

An important outcome of the 1938 Disturbances and the ensuing shift in political power to the black population was a re-examination of the education system. As documented by Errol Miller, two separate systems of education had developed during colonial times; these were the preparatory school/high school system and the elementary/teachers' college system. The first catered to the ruling elites and some segments of the middle

classes – particularly those living in urban centers – while the second catered mainly to the lower strata and members of the middle strata in rural areas. Although far-reaching recommendations were advanced in the Kandel report of 1943 for the reorganization and integration of the entire education system, the removal of barriers to secondary school access remained a goal for most of the remaining century. This ongoing process of educational reform which started in the postwar period has been largely directed toward addressing the intractable problem of separate and unequal provisions for different social classes.

Change was facilitated by the attainment of adult suffrage and the achievement of self-government in 1944. This set the stage for the introduction of two critical policy initiatives. The first was the 1957 Reform in education which introduced a CEE to select students for the secondary schools. This was the first major reform in secondary education since the grant-in-aid system was introduced in 1920, and it doubled the number of students who could receive free tuition. The second was the 1966 New Deal in Education as a result of a UNESCO-led planning mission. This program promoted the development of junior secondary schools for the 12–15 age group by providing schooling at the grades 7–9 level. This was conceived as the first cycle of secondary education. Both of these initiatives were accompanied by a large expansion in the number of schools, and so served to ensure greater access by a wider cross-section of the society.

The country achieved independence in 1962, and the main adjustment to the secondary school system over the 1970s was the addition of two additional years of schooling – so the junior secondary schools became upgraded to new secondary schools. However, the terminal school-leaving examination for the New Secondary schools (the locally administered Jamaica Secondary School Certificate) was not the same as for the high schools which relied on external examinations. Since passing in the external examinations was required for university admission, the new secondary schools did not constitute a route to higher education. In this two-tier structure of secondary education, a link was provided by the Grade 9 Achievement Test which acted as a conduit through which a handful of students could transfer to the traditional high schools, and so participate in the elite system which provided the main foundation for tertiary education.

Although the recession and the budgetary restrictions of the 1980s and 1990s curbed educational expenditure, the government instituted a far-reaching policy initiative, the World-Bank funded Reform of Secondary Education (ROSE) program (1993–2000). This program incorporated the recommendations of the 1983 UNESCO mission, and sought to ensure that a common curriculum would be delivered within the secondary grades of all institutions.

Under the ROSE program, several new secondary schools were upgraded to the level of secondary high schools and comprehensive schools. For the first time also, the All-Age schools were brought into this ambit. This meant those students who remained in the All-Age schools after the age of 11 years could participate in the same curriculum as those in the secondary schools. This was expected to be a transition stage until the All-Age schools were phased out. By offering only grades 1–6 schooling, they would be converted into primary schools. Over this period, the new secondary schools were reclassified as comprehensive high schools, and, by 2001, all were further reclassified as secondary high schools – being currently known as the upgraded high schools. The outcome of these efforts to expand the secondary school system was reflected in the findings of the 2001 census, which showed that among the population aged 15 and older, 69.4% had attained at least some secondary schooling. Data from UNESCO point to a gross enrolment rate of 86% for the secondary school-age population in 2001, while the out-of-school population was estimated at 18%.

In order to assess the extent to which the country has been able to meet the challenge of integrating secondary education, it is useful to compare the two types of secondary school in regard to their participation in the CSEC examinations and their relative success. Data collated by the National Council on Education showed that, in 2005, the cohort eligible to take the CSEC examination for English Language stood at 12 393 students in the 54 traditional high schools. The number who actually took the examination was 10 751 and, of these, 8336 or 78% passed with Grades 1, 2, or 3. In contrast, among the 80 upgraded high schools, the matching cohort stood at 19 930 students and, of these, only 6477 students were enrolled for the exam. This represented less than one-third of the group. The number who passed the English exam totaled 2515 students, or only 39% of those who enrolled. In summary, the students in the traditional high schools (the upper tier) were at least 2.5 times more likely than those in the upgraded high schools (the lower tier) to be entered for this critical exam, and they were twice as likely to pass. It is apparent, therefore, that – despite efforts at reform – there are marked differences in the experiences of children in each tier of the secondary school system. The conclusion of Derek Gordon appears to still hold true – namely the structure has largely remained a class-based dual-track system.

Summary

The secondary school system in the Caribbean is still engaged in the struggle to transform itself from being a mechanism of colonial patterns of social distribution to being a pathway toward equitable access, equality of opportunity, and justice. The situation which faces each Caribbean island differs as these challenges are shaped by history and culture, and constrained by economic resources. Equally important is the philosophy that guides this important set of institutional choices, and which will determine the future development and social stability of the Region.

Bibliography

Miller, E. (1992). *Education for All: Caribbean Perspectives and Imperatives*. Washington, DC: The Inter-American Development Bank.

Further Reading

- Campbell, C. (1992). *Colony and Nation: A Short History of Education in Trinidad and Tobago*. Kingston: Ian Randle Publishers.
- Campbell, C. (1997). *Endless Education: Main Currents in the Education System of Trinidad and Tobago 1939–1986*. Kingston: The University of the West Indies Press.
- Cogan, J. (1983). Jamaica: Education and the maintenance of the social-class system. In Thomas, E. (ed.) *Politics and Education: Cases from Eleven Nations*. Oxford: Pergamon.
- Di Gropella, E. (2003). *Monitoring Educational Performance in the Caribbean*. Washington, DC: The World Bank.
- Gordon, D. (1991). Access to high school education in post-war Jamaica. In Miller, E. (ed.) *Education and Society in the Commonwealth Caribbean*. Kingston: University of the West Indies, Institute of Social and Economic Research.
- King, R. (1999). Education in the British Caribbean: The legacy of the nineteenth century. In Miller, E. (ed.) *Educational Reform in the Commonwealth Caribbean*. Washington, DC: The Organization of American States.
- Layne, A. (1999). Educational reform in Barbados in the post-war period. In Miller, E. (ed.) *Educational Reform in the Commonwealth Caribbean*. Washington, DC: The Organization of American States.
- Miller, E. (1990). *Jamaican Society and High Schooling*. Kingston: The University of the West Indies, Institute of Social and Economic Research.
- Miller, E. (ed.) (1999). *Educational reform in the Commonwealth Caribbean*. Washington, DC: The Organization of American States.
- Task Force on Educational Reform Jamaica (2004). *Report of Task Force: A Transformed Education System 2004*. Kingston: Ministry of Education.
- The World Bank (1993). *Caribbean Region: Access, Quality and Efficiency in Education*. Washington, DC: The World Bank.
- The World Bank (1999). *Jamaica Secondary Education: Improving Quality and Extending Access*. Washington, DC: The World Bank.
- Williams, E. (1964). *History of the People of Trinidad and Tobago*. London: Andre Deutsch.

Women of Color in Science, Technology, Engineering and Mathematics (STEM): Refining the Concepts, Reframing the Issues

C B Leggon, Georgia Institute of Technology, Atlanta, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The fields of science, technology, engineering, and mathematics (STEM) are the drivers of economies worldwide. In such global knowledge economies, it is imperative that nations develop, enhance, and utilize human resources effectively and efficiently. The greater the diversity of backgrounds of scientists and engineers, the greater the variety of perspectives brought to bear on the science and engineering enterprise; this results in more creativity and ingenuity that enhances both science and engineering (Bassett-Jones, 2005; Malcom *et al.*, 2004). Several commissioned reports express concern that the relatively homogeneous science and engineering workforces in the US, puts at risk its competitive edge in the global knowledge economy.

Recognizing the link between science/engineering/technology and domestic development, many nations have incorporated increasing women's participation in these fields as an integral part of their domestic policy strategy. Why is it that despite policies and programs to increase women's participation in STEM, data from the United States (US) and the Organization for Economic Co-Operation and Development (OECD) indicate that in STEM fields, women remain unevenly distributed in both research and decision-making positions? This is counterproductive to maximizing economic development because there is a positive correlation between the diversity of perspectives among scientists and engineers, and the quality of science and engineering – and consequently, economic development. For example, lagging growth in sub-Saharan Africa, North Africa, and the Middle East has been attributed to gender inequality (Katseli, 2007).

Scientists and engineers affect the direction of science and engineering fields insofar as they determine the:

- problems selected for study;
- framing and formulation of research problems;
- populations/groups that should be studied – and how they should be studied;
- methods used to collect and analyze data; and
- ways in which results are reported and disseminated.

In addition, scientists function as gatekeepers who control certification of both scientists and the institutions that train them, including

- the curricula for institutions of higher education that train scientists and engineers and

- who will be recruited, educated, trained, and mentored for science careers in science and engineering and how this will be done.

Although the extent varies from one country to another, gatekeepers limit who and which categories of persons enter and advance in a career. Even though small groups of individuals might benefit from arbitrarily (and artificially) limiting participation in science and engineering, research indicates that the most productive societies have permeable boundaries between categories of people. Diversifying the science and engineering workforces brings a greater variety of perspectives to science and engineering which, in turn, can have beneficial results. Examples of these beneficial results include:

- assessing differential impacts of the products of science and technology on various segments of the US population
- increasing the potential for reducing health disparities between racial and ethnic groups
- increasing the extent to which more of the population benefits from science and technology.

In the US, discourses on diversifying the science and engineering labor forces are problematic in two major aspects: defining terms and conceptualizing issues. Language enables not only the expression of ideas and abstract concepts, but also puts limits on the forms those ideas and conceptualizations can take (Whorf *et al.*, 1998). It is problematic when a word and/or concept used to describe a phenomenon takes on a meaning of its own that differs from the phenomenon to which it refers – as is the case for research, policy, and programs concerning women's participation in science and engineering careers. This is best illustrated by the way in which under participation in the US science and engineering labor forces tends to be conceptualized in terms of two groups: women and racial/ethnic minorities. Not only does this dichotomous conceptualization imply that the categories are mutually exclusive, it also masks significant gender differences within racial and ethnic groups as well as significant racial/ethnic differences within gender groups – both impede the ability to target policy, programs, and practices designed to promote and enhance participation in science and engineering careers. This article argues that there is a need to both reframe the issues and refine the concepts concerning women's under participation in science and

engineering. Just as critical as it is to systematically disentangle concepts, it is also imperative to disaggregate data.

Data, Documentation and Indicators of Women's Participation in Science

The UNESCO Institute for Statistics (UIS) estimates that women comprise slightly more than 25% of the world's researchers (UIS, 2006). Women are under-represented in science worldwide, although this under-representation varies by region and within region by nation. For example, women comprise 15% of researchers in Asia overall, ranging from a high of 50% in Central Asia to a low of 12% in South Asia; the low percentage for South Asia is due to the fact that in India women comprise only 10% of the researchers. Women comprise 85% of all researchers in Myanmar, almost half of all researchers in Latin America, 42% in Europe, and 35.2% of researchers in China. The extent of women's under-representation in science is difficult to measure because of limited comparable cross-national data and statistics (UIS, 2006). Moreover, despite having a large number of researchers, Australia, China, and Mexico lack data on women's participation in science and engineering.

Women in both the US and European Union (EU) tend to be more highly represented in the humanities, biological and social sciences, and less well represented in engineering, mathematics, and physics. Despite similarities between the US and EU in terms of women's participation and uneven distribution in science and engineering fields, the EU is ahead of the US in developing indicators for tracking trends in women's participation in science and engineering careers. Austria, Germany, and the Netherlands have low percentages of female scientists and little data on them; countries formerly under communist rule have good statistical documentation of their high proportions of women scientists (UNESCO, 2007).

Increases in the numbers of women graduates in science and engineering fields have not been reflected in concomitant increases in the proportions of women in academic posts in these fields in the US, EU, and Canada. Unlike in the US, the EU has made issues concerning women in academic careers in science and engineering a top priority; consequently, the EU has a centralized mechanism to collect data, create, and disseminate indicators of women's participation in SET. Assessing women's participation in science and engineering careers requires collecting and analyzing longitudinal data. Without data that track careers of the same women over time (rather than the careers of different women at different points in time), it is impossible to accurately assess trends and identify factors that enhance career success as well as those that impede such success.

Refining the Concepts

Field Concepts: STEM and SET

Much of the research on under-representation in the science and engineering workforces in the US uses the acronym STEM to refer to science, technology, engineering and mathematics as if they were one homogeneous entity; similarly, researchers in Europe use the acronym SET to refer to the science, engineering and technology workforce. Both acronyms obscure important distinctions between the sciences and engineering/technology in terms of the structure of both the knowledge base and careers in these fields. Similarly, these aggregations obscure significant distinctions among fields within both categories (sciences and engineering).

Demographic Concepts: Race, Ethnicity, and Gender

By 2005, one in every three US residents was part of a group other than single-race non-Hispanic white. The two largest minority groups in the US are African Americans and Hispanics. The US Census Bureau uses the term Hispanic to refer to people "who identify their origin on the census form as Mexican, Puerto Rican, Cuban, Central or South American, or some other Hispanic origin" (US Census Bureau P20-353). Hispanics comprised the fastest growing portion of the population, growing by 3.3% between 2004 and 2005. Numbering 42.7 million, Hispanics are the largest minority group in the US. It is somewhat ironic that the racial and ethnic groups that comprise the fastest growing segments of the US population are the same groups that are under participating in science, technology, and engineering fields. Despite the growth among Hispanics and African Americans, the US science and engineering workforces remain overwhelmingly composed of non-Hispanic European American males (BEST, 2006; Leggon and Pearson, 2006).

Traditionally in the US, the science and engineering talent pool consisted largely of white non-Hispanic males; however, as this group's interest in the sciences and engineering declined over the last two decades, concern arose over increasing the participation in this talent pool of other groups. In the past, the US remedied any shortages in the science and engineering talent pool by importing talent from abroad rather than developing it among indigenous racial/ethnic groups that are under participating in science and engineering. Increasingly, this remedy is no longer viable for the US for many reasons. First, the EU and other nations are increasingly competing successfully for this talent. Second, the nations from which much of this imported talent came are becoming attractive enough in terms of infrastructure and jobs for them to return. Third, after 11 September 2001, heightened security

concerns have limited certain jobs in science and engineering fields to US citizens.

In theory, it is assumed that women from racial/ethnic minority groups – African Americans, Mexican Americans, Puerto Ricans, Native Americans/Native Pacific Islanders – are included in the categories of both women and racial/ethnic minorities; in practice, these women are frequently treated as marginal or invisible in both categories in terms of research, policy, and programs. There is a parallel situation in Europe: research on women in science, engineering and technical careers seldom discusses African-Caribbean women's under representation in science and engineering employment, nor how race and ethnic differences among women affect their experiences in those work forces. It is important to note that race and ethnicity are defined differently in different nations. Regardless of how race and ethnicity may be defined by a given nation, it would be useful to track women's participation in science, engineering and technical careers by disaggregating data by immigration status. For example, in the US, the category Black sometimes includes individuals born outside of the US; this gives a misleading picture insofar as it makes the extent of African Americans' participation in the science and engineering workforces seem greater than it actually is.

Disaggregating the category women into two groups – women and women of color is also problematic, primarily because women of color is an umbrella category that includes women from disparate race/ethnic groups – African American, Hispanic, Native American/Native Pacific Islander; this umbrella category obscures significant socio-historical and cultural differences among these women. Similarly, Hispanic is an umbrella category that obscures significant socio-historical and cultural differences among women who are Mexican American, Puerto Rican, Cuban, and from Central America, South America, and Spain.

Since 1992 in the US, there has been an increase each year in the numbers of bachelor's degrees in the natural sciences awarded to women who are US citizens or permanent residents. However, disaggregating the data reveals that the percentages of bachelor's degrees in the natural sciences to non-Hispanic white women declined from 79% to 70%, while the percentages awarded to non-Hispanic African American women and Hispanic women increased from 6% to 8% and from 5% to 6%, respectively (CPST 2004, table 2–74). Regardless of race and ethnicity, women are severely under-represented among tenured faculty at research universities. In terms of tenure, the gender gap has remained fairly constant over the past 30 years. Furthermore, among academics with doctorates in science and engineering fields, only 25% of women have been awarded tenure as compared to 50% of the men (Trower and Chait, 2002).

Race/ethnic data must be disaggregated by gender because how one experiences race and ethnicity is

mediated by gender. Similarly, gender data must be disaggregated by race and ethnicity because how one experiences gender is affected by race and ethnicity. Among African Americans and Hispanics, males as well as females are under-represented among tenured faculty at US research universities. Among both African Americans and Hispanics there are gender differences in the field distributions of PhDs and representation on academic science and engineering faculties. During the 1990s among both African Americans and Hispanics, women were responsible for the increases in the numbers of science and engineering doctorates earned by their respective race/ethnic group, but males still held more faculty positions and a greater number were tenured than females in their race/ethnic group. It is also significant that women of color – African American, Mexican American, and Puerto Rican – are less likely than either non-Hispanic white women or men of any racial or ethnic group to be awarded tenure. Masking crucial intra-gender and intra-racial distinctions results in policy, programs, and practices that are at best ineffective and at worst counterproductive for diversifying the science and engineering workforces.

Reframing the Issues

In the US, research on under participation in science and engineering fields has been framed in terms of two groups – women and racial/ethnic minorities. This dichotomous categorization is problematic because it implies that women are non-Hispanic and white, while racial/ethnic minorities are male. Even more problematic, this dichotomy suggests that women and race/ethnic minorities are mutually exclusive groups. Whether women of color are included in the category of women or a given racial or ethnic group, significant differences are obscured. Although women of color share characteristics with both men in their respective racial/ethnic group and women in different racial/ethnic groups, it is crucially important to systematically identify differences between women of color on the one hand, and men of color and non-Hispanic white women, on the other hand. The effects of the intersection of gender, race, and ethnicity are not additive – that is merely summative. Rather these effects are synergistic and bidirectional: that is, gender impacts how race and ethnicity are experienced; and race and ethnicity impact how gender is experienced.

In the US, analyses of participation in science and engineering tend to focus on quantitative indicators, such as numbers of degrees earned and labor force participation rates. This focus is problematic because numbers provide no information on the qualitative aspects of this participation. For example, focusing on the numbers of advanced degrees earned in science, engineering and technology does not adequately address the issue of the quality of the

institutions from which those degrees were earned. In the US some institutions that were formerly degree mills (and not accredited) now award the greatest numbers of degrees to indigenous racial/ethnic minorities – African Americans, Mexican Americans, and Native Americans. The absence of information on the quality of degrees earned makes it easier to make unwarranted inferences that differences in workforce participation among degree earners are due to deficiencies in individuals rather than differences in the quality of the institutions from which they were awarded their degrees. Although advanced degrees are entry credentials to the science and engineering workforces, much of the research on gender differences in participation in the science and engineering workforces stops with the award of the PhD or tertiary degree.

Qualitative indicators of women's participation in the sciences and engineering are just as important as quantitative indicators. In other words, equally important as the numbers of women in careers in science and engineering is the quality of those women's career experiences. Indicators of the qualitative aspects of women's participation in science and engineering include rank, honors, and prestige. Although women in China comprise more than one-third (35%) of all scientific researchers and 34% of those in technical service, they comprise only 4.6% of chief scientists, and receive only 12% of the top three government awards in science and technology. In 2000, female academicians comprised only 6% of the total membership in the Chinese Academy of Science. Data from Canada also reveal significant gender differences in the quality of participation in the sciences and engineering. In 2003, 88% of male academic staff either had tenure (70%) or held tenure track positions (18%); by contrast, only 40% of female academic staff was tenured and 25% were in tenure track positions (CCWESTT, 2006). Women had been appointed to 21% of the Canada Research Chair Positions as of 2006.

In the US, there is an inverse correlation between the number/percentage of women faculty and the quality of an institution of higher education: the higher the rank of an institution, the fewer the number of women faculty; moreover, this applies across fields. Regardless of the quality rating of a college or university, there is an inverse correlation between academic rank and the number of women: the higher the rank, the fewer women in that rank. Women in both the US and EU were more likely than males to have short-term contracts and/or to be on soft money. In the Glass Ceiling Index, an innovative indicator that compares women's and men's chances of reaching top academic positions, the higher the score, the greater the barrier to women's advancement. According to this index, the greatest barriers were in Malta, and the fewest barriers were in Romania and Turkey.

Another aspect of the quality of women's participation in science and engineering careers is the extent to which

they are in decision-making positions and can influence research agendas. One indicator of this is participation is the portion of scientific boards comprised by women. The 2000 Survey on Leadership in China conducted by the Association for Science and technology found that women comprise 15% of director generals, and 28.6% of directors. Women comprise 40% of science board members in Finland, and more than 30% in the UK and Denmark. In the US, women comprise 25% of the 24-member National Science Board. In Argentina, there are no female deans in private universities (UNESCO, 2007).

Finally, research on women's participation in science and engineering fields should conceptualize such participation in terms of careers, and progress over time. This would enhance the identification of factors that enhance as well as impede women's career at each stage. For example, in 2004 the Athena Project, a consortium of UK research funding organizations, universities, and government science departments, conducted a survey that found that although women had a slightly higher success rate than did men in applying for their first lectureship post, women were significantly under-represented both at senior levels and in older age groups. The UIS, Eurostat, and the OECD have developed a method of tracking aspects of the careers of doctorate degree holders across nations, such as the time it takes to secure employment and salaries.

Summary and Conclusions

Through focusing on women of color in science and engineering, this article points out the criticality of refining concepts and reframing issues concerning under participation in science and engineering. Refining concepts and reframing issues should be data driven and in turn should inform data collection. Data on participation in science and engineering careers must be longitudinal and disaggregated by race, ethnicity, and gender to help identify issues, track trends, and monitor the effectiveness of interventions.

Bibliography

- Bassett-Jones, N. (2005). The paradox of diversity management, creativity and innovation. *Creativity and Innovation Management* 14(2), 169–175.
- BEST (Building Engineering and Science Talent) (2006). *Minorities in Engineering and Related Fields – Diversity Analysis of Students Earning Bachelor's Degrees*. Menlo Park, CA: BEST.
- CCWESTT (Canadian Coalition of Women in Engineering, Science, Trades and Technology) (2006). University Academic Staff.
- CPST (Commission on the Status of Professional Women and Minorities) (2004). *Professional Women and Minorities: A Total Human Resources Compendium*. Washington, DC: CPST.
- Katseli, L. T. (2007). *Gender and Economic Development*. Oslo: Norway.
- Leggon, C. B. and Pearson, W., Jr. (2006). Race, gender and the American scientific enterprise: The ramifications of exclusion in a global market. *International Sociological Association*. Durban, South Africa.

- Malcom, S. M., Chubin, D. E., and Jesse, J. K. (2004). *Standing Our Ground: A Guidebook for STEM Educators in the Post-Michigan Age*. Washington, DC: American Association for the Advancement of Science.
- Trower, C. A. and Chait, R. P. (2002). Faculty diversity: Too little for too long. *Harvard Magazine* **104**(4), March–April 2002.
- UIS (2006). Women in science: Under-represented and under-measured. *UIS Bulletin on Science and Technology Statistics*, issue no. 3, November 2006.
- UNESCO (United Nations Educational, Scientific and Cultural Organization) (2007). Women in science: The missing links. *The UNESCO Courier* **2007**, No. 2.
- Whorf, B. L., Carroll, J. B., et al. (1998). Language, thought and reality: Selected writings of Benjamin Lee Whorf. <http://www.library.gatech.edu:2048/login?url=http://www.netlibrary.com/urlapi.asp?action=summary&v=1&bookid=49458> An electronic book accessible through GALILEO.

Further Reading

- Leggon, C. B. (2003). Women of color in Information Technology: Degree trends and policy implications. *IEEE Technology and Society Magazine* **Fall**, 36–42.
- Leggon, C. B. (2006). Women in science: Racial and ethnic differences and the differences they make. *Journal of Technology Transfer* **31**, 325–333.
- MacLachlan, A. J. (2006). The lives and careers of minority women scientists. *Conference Proceedings*. New Orleans, LA: NAWA.
- Pearson, W. and Fechter, A. (eds.) (2004). *Who Will Do Science? Educating the Next Generation*. Baltimore, MD: Johns Hopkins University Press.
- Trower, C. A. and Bleak, J. L. (2004). *The Study of New Scholars: Tenure-Track Faculty Job Satisfaction Survey*. Harvard University, Graduate School of Education.

DEMOGRAPHY AND SOCIAL CHANGE – SOCIAL CHANGE

Contents

Educational Diversity in U.S. Law Schools

Educational Reform

Emergence of For-profit Higher Education

Engaged Education: Experiential Learning, Intensive Field Experiences, and Social Change

Higher Education and Social Change in Mexico

The Changing Role of the State in Higher Education

The Role of the State in School Reform: Responding to Increasing Ethnic Diversity in Britain

Transforming Mindsets Through Education for Sustainable Development

Educational Diversity in U.S. Law Schools

C E Daye and A T Panter, University of North Carolina, Chapel Hill, Chapel Hill, NC, USA

W R Allen and M E Deo, University of California, Los Angeles, Los Angeles, CA, USA

L F Wightman, University of North Carolina, Greensboro, NC, USA

© 2010 Elsevier Ltd. All rights reserved.

It is useful to situate the modern-day concept of educational diversity in its historical, social, political, and legal contexts. The modern version of the term educational diversity is derived from the United States Supreme Court's usages in the Michigan law school affirmative action case – *Grutter v. Bollinger* in 2003 – in which the Court held that a state university has a compelling interest in seeking to derive the educational benefits that flow from a diverse student body and may pursue that interest through a narrowly tailored system and in a time-limited manner. Social scientists have identified institutional and student diversity, particularly racial and ethnic diversity, which is associated with a range of positive educational outcomes for the students, their institutions, and society. Research linking educational diversity to these outcomes ranges from experimental studies that involve different combinations of students who are instructed to interact on problem-solving tasks to large national studies of undergraduate and graduate/professional students that examine attitudes, cross-racial interaction, course taking, and academic performance during and after graduation.

The study of educational diversity within the law school setting is of particular interest because of legal case history as well as the nature of the academic experience and training for this highly accomplished set of students. Law students arrive at law school with a set of personal characteristics, family background, academic and interpersonal experiences, attitudes, and expectations about their future in law school and their careers. Depending on many

student-level, institution-level, and curricular factors, the composition and experiences of law students may frame their law school educational experiences, including how they interact with their professors and their peers, analyze cases, engage in activities while in law school, choose courses, select internships and interview opportunities, and apply their legal knowledge after graduation.

Legal Background of Educational Diversity

Thrusts to improve the political, economic, and social conditions of African Americans did not begin with the modern era of affirmative action programs. Indeed, the beginning point for a full discussion of America's relationship with its oppressed peoples of African descent would have to recount the justifications, deprivations, and continued effects of a legacy of slavery. But that is beyond the scope of this article. Affirmative action effectively began with the Emancipation Proclamation of 1 January 1863, which freed the slaves. Proponents of abolishing slavery and of uplifting former slaves recognized the need to protect the former slaves' new freedom and to remediate some of the past injustices they had suffered. Therefore, they pushed for antisubjugation provisions, political protection measures, and remedial measures. Some of these provisions were added as amendments to the Constitution of the United States:

1. The Thirteenth Amendment adopted in 1865 prohibited slavery and provides, in part, that “[n]either slavery nor involuntary servitude, except as a punishment for crime . . . shall exist within the United States, or any place subject to their jurisdiction.”
2. The Fourteenth Amendment adopted in 1868 made persons born or naturalized in the United States citizens and provides, in part, that “All persons born or naturalized in the United States, and subject to the jurisdiction thereof, are citizens of the United States and of the State wherein they reside.” To prohibit the defeated slave states from formally re-subjugating the newly freed slaves, the Fourteenth Amendment also provided that no state could deny “the equal protection of the laws”: “No State shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any state deprive any person of life, liberty or property without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws.”
3. The Fifteenth Amendment adopted in 1870 gave the newly freed slaves, declared by the Fourteenth Amendment to be citizens, the right to vote, and provides that: “The right of citizens of the United States to vote shall not be denied or abridged by the United States or by any State on account of race, color or previous condition of servitude.”

These constitutional measures empowered Congress to enact programs to protect the former slaves and to improve their political, economic, and social conditions. As the Civil War drew to a close, the US government created the Freedmen’s Bureau in 1865 (ended in 1871) to provide assistance (e.g., educational services, land, and other aid) to the newly freed slaves, but not to white refugees or others. The government promised to sell or lease to farmers parcels of unoccupied land and land that had been confiscated by the Union during the war; it also promised the loan of a federal government mule to plow that land (40 acres and a mule). Some African Americans took advantage of these programs and either bought or leased parcels of land. During reconstruction, however, President Andrew Johnson vetoed a bill to enlarge the powers and activities of the Freedmen’s Bureau, and he reversed many of the policies of the Bureau. Much of the promised land, which had been leased to African American farmers, was taken away and returned to Confederate loyalists. As the promise of significant, sustained aid and land was soon breached, for most African Americans, the promise of 40 acres and a mule was never kept.

After the Civil War legislation, and the aborting of much of the Reconstruction steps to protect and uplift the former slaves, a bleak era of Jim Crow descended upon the nation. Segregation and discrimination pervaded America. In 1896, the US Supreme Court’s decision in *Plessy v. Ferguson* in 1896, sanctioned discrimination against

blacks under the separate-but-equal rubric. Another half a century of struggle ensued. The issue of black subordination and exclusion would not go away. So again it reached the Supreme Court. This time, at last, the Court recognized, in its *Brown v. Board of Education* decision in 1954 that black citizens had a right not to be excluded from educational opportunities the states offered to whites, repudiating the separate-but-equal doctrine of *Plessy v. Ferguson*.

After *Brown*, institutions realized that they needed to take additional steps to make educational opportunities realistically available to blacks. Affirmative action provisions also included women and selected other minority groups who had been historically victimized and who were under-represented at elite educational institutions. Many institutions adopted various forms of affirmative action that considered the race and ethnicity of applicants. One prime reason for the need to consider race in an attempt to assure access for minorities and to attain diversity within educational institutions is the disparate performance by blacks and some other minorities on standardized tests as compared with white test takers. The reasons may derive from contemporaneous effects of historic racism, continued educational discrimination, and pervasive economic disadvantages.

A quarter-century after *Brown*, the Supreme Court’s *Bakke* decision in 1978 held that race could be a plus factor in university admission decisions. The Supreme Court, in *Grutter v. Bollinger*, again affirmed that race may be a plus factor in admission decisions. Yet, debate continues about consideration of race in university admissions. The issue of color, thus, remains controversial in the United States.

Grutter held that the Constitution permits educational institutions to make a narrowly tailored use of race in admission decisions to further a state’s compelling interest in attaining a diverse student body. The case sanctions the use of race as a factor that universities may consider as they seek the “educational benefits that flow from a diverse student body.” Justice O’Connor’s opinion for the Court in *Grutter* recognized specifically that an educational institution, in this case a law school, may, in good faith, define its institutional mission so as to derive the benefits of educational diversity because diversity “promotes learning outcomes,” “better prepares students for an increasingly diverse workforce and society,” and helps in “cultivating a set of leaders with legitimacy in the eyes of the citizenry.” Justice Scalia, one of the dissenting judges, set forth issues that subsequent suits may challenge including “whether . . . any educational benefits flow from racial diversity” or “the bona fides of the institution’s expressed commitment to the educational benefits of diversity.” Pointedly he says, “Other lawsuits may focus on whether, in the particular setting at issue, any educational benefits flow from racial diversity.”

After *Grutter*, therefore, major research tasks remain, including deriving a more articulated and analytical

conceptualization and operationalization of educational diversity, empirically examining whether race/ethnicity contributes to that diversity, and if it does, explicating how any contribution of race is manifested during an educational experience. A clearer understanding of these issues is essential to explain how considering an applicant's race/ethnicity supports permissible educational objectives that an institution may pursue consistently with constitutional limitations. It would not be permissible to consider race for the purposes of enrolling minorities to achieve a certain racial balance, correcting for societal discrimination, training professionals to serve underserved communities, or creating minority role models for minorities to emulate.

Diversity as a concept has been used in contemporary discussions of higher education since the mid-twentieth century. In 1950, *Sweatt v. Painter* held that segregated education provided at makeshift black law schools was not substantially equal to education available at the University of Texas School of Law, in part, because of a lack of diversity among the students enrolled in the all-black law school. The court made similar statements in 1951 about the important value of diversity in the case that ordered the end to state-sanctioned segregation of the law school at the University of North Carolina in *McKissick v. Carmichael*. Diversity has been linked to deeper historical antecedents and earlier judicial references, including Justice Powell's citation in the 1978 case of *Regents of University of California v. Bakke* to Harvard College's use of diversity for admission purposes dating back 30 years.

Social Science Research on Diversity

Our examination of previous research on race, education, and the value of diversity leads us to conclude that most of that research can be characterized as examining foundational questions or secondary questions. Foundational questions ask whether racial diversity is essential to educational diversity or what role race plays as an aspect of educational diversity in fostering educational goals. Secondary questions assume – implicitly or explicitly – that racial diversity is a given, that racial diversity is an independent goal, or that racial diversity equates to educational diversity. Having made one or more of these assumptions, researchers can analyze various secondary questions, such as, whether diversity affects educational outcomes or in what ways educational diversity affords benefits to institutions, individuals, or society.

Substantial and valuable social science research has addressed law, diversity, and race. More particularly, some examinations of diversity have assumed that racial or ethnic diversity (or some other aspect of diversity) is a given. From that starting point, researchers have made broadly theoretical examinations of diversity. Other works have assumed

either that racial diversity is the end to be sought or that achieving racial diversity assures some measure of educational diversity. From that vantage point, researchers have examined one or more educational diversity constructs, such as interactions with cross-race peers, choice of social justice-oriented courses, or ability to solve complex cognitive problems. Based largely on the assumption that racial diversity creates educational diversity or that the goal is to assure racial diversity, scholars have explored aspects of diversity from empirical perspectives, through theoretical discussion, as well as by offering speculative legal and doctrinal viewpoints. Other works have assumed the existence of diversity (based wholly or in part on having achieved some degree of racial diversity) and then examined the impact of diversity in one or more educational domains (e.g., the effects of a particular undergraduate course on students' educational experiences and interactions with peers both inside and outside the classroom), as well as the perceived benefits of diversity based on opinion surveys of faculty members in higher educational institutions and law schools. Still other research has examined how assumed educational diversity influences student learning, educationally related outcomes in educational domains, and preparedness for future interactions in a democratic society, again largely on the premise that racial and ethnic diversity equals educational diversity. Research has also focused on ways to foster diversity at an institutional level with attention to racial and ethnic diversity; recognizing structural diversity or achieving a particular percentage of minority representation is not a sufficient condition for influencing students' attitudes and orientation toward racial equity. Studies have examined the educational domain of minority graduates' achievement in rendering special service to society. Other research has focused on general attitudes and opinions about ethnic diversity as an aspect of academic life or in the nation, with some polls addressing attitudes toward several diversity constructs such as sociopolitical attitudes on equal opportunity, affirmative action policies, or symbolic racism. Finally, researchers have conducted empirical analyses of the extent to which minorities would be excluded from selective educational institutions and graduate/professional schools and law schools if racial diversity were not used as a consideration in making admission decisions.

To study diversity and race more fully, researchers might first examine whether students' race/ethnicity or other definable personal attribute (e.g., gender, family background, socioeconomic status, geographic origin, or political orientation) is associated with educational diversity. Such a study would identify educational diversity construct areas in which diversity might be relevant, such as:

- Diversity of student background, including a person's race/ethnicity, gender, age, sexual orientation,

highest level of education, religion, spirituality, political orientation, current debt levels, and other descriptors of a person's current state.

- Diversity of family background, including demographic and social family factors, such as family structure (one/two parent), family size, family socioeconomic status during the growing up years, culture, language spoken at home, customs, and traditions that may influence students' perceptions and interpretations of curricular material.
- Diversity of experience, including positive and negative life experiences, which each student brings to the classroom and the campus. These experiences might include exposure to customs, cultures, and perspectives as well as experiences of prejudice and disadvantage that might influence a student's perspective on the social order.
- Diversity of attitudes, including differences in values, beliefs, sociopolitical attitudes, and conceptions of the world and one's self. Students who hold different beliefs about what is important, worthy, beautiful, and good in life may be more likely than a group whose values are homogeneous to show flexibility in thought, have the ability take others' perspectives, and discover for themselves the depth and interminability of the disputes in which human beings find themselves entangled.
- Diversity of educational expectations, including predispositions that students bring to the educational setting, such as curricular interpretations, classroom interactions, and how learning will occur. These predispositions derive from prior classroom experience (e.g., from one's undergraduate experience) and are manifested in, for example, class participation rates, the way assignments and class projects are prepared and presented, and the extent to which students engage in study groups, class project groups, and other group-oriented academic behaviors. These predispositions may also be reflected in student expectations and plans to engage in specific law school academic activities both inside and outside of the classroom. For example, students may be differentially likely to choose certain courses related to social justice, participate in political groups as a means of advancing issues of social justice, engage in specific volunteer public service activities, or orient toward certain curricular activities.
- Diversity of career goals and aspirations including differences in reasons for pursuing higher education and variability in student expectations for their own professional careers beyond law school to different foci that students bring to issues under study. It explores the extent to which students foresee that their education will be beneficial to themselves or to their communities after they leave the formal educational setting.

Second, researchers might consider the multilevel influence, if any, that race plays as an aspect of educational diversity in educational domains, such as:

- The individual domain emphasizing enrichment of each individual student's educational experiences during law school by deepening his/her personal understanding about ideas through exposure to many different perspectives (with other students and faculty members) and by making educational encounters richer, livelier, and more interesting.
- The institutional domain emphasizing the presence of diverse groups and the opportunities for intergroup interaction within the educational settings, widening the scope of perspectives expressed in the classroom, in the hallways, and on campus. This perspective focuses on an increased range of discussions, activities, programs, and interests represented within institutions and on formal and informal institutional changes that may be fostered by these phenomena.
- The social domain emphasizing that the creation of opportunities for students to interact with others of different backgrounds, races, and cultures within the educational institution, where they are enrolled will, in turn, increase and enhance their ability to interact positively and effectively in a diverse and global society.

Recent Research Examining Educational Diversity in the Law School Setting

Much of the extant research examining how student body diversity contributes to educational outcomes has focused on college student samples. The undergraduate setting allows the study of a broad cross section of students and institutions and captures student experience of social interaction, attitudinal development, and academic growth during the formative years.

Law students reflect an academically successful subgroup, who have completed a very deliberate set of steps by the time they have matriculated at one of the 196 American Bar Association (ABA)-approved law schools (in 2007). Members of this select group of students have:

1. chosen to pursue a law-related career;
2. completed all of the necessary academic and preparatory behaviors and research during their undergraduate and in many cases, post-undergraduate years, to apply for law school admission;
3. been admitted to law school;
4. matriculated at a specific law school;
5. embarked on a very structured set of common courses during their first year and internship experiences thereafter; and
6. have chosen educational activities that would prepare themselves to enter the workforce and contribute meaningfully nationally and globally, in areas such as government, private practice, education, business, and public service.

The Educational Diversity Project (EDP) is a longitudinal, multiple method study of law students, assessing a broad set of experiential and psychological characteristics to understand (1) the individual differences that incoming students have upon entry to law school and how these differences do or do not relate to race/ethnicity (and other personal characteristics); and (2) how individual differences identified during the first weeks of law school may frame a student's legal training, interaction with classmates with diverse viewpoints different from their own, and personal decisions about how they will use their legal training in employment after graduation. The study also examines the extent to which student variability would be attenuated if there were no racial diversity. Thus, the EDP examined the research questions: Does race/ethnicity (and other personal factors) relate to educational diversity, as defined by individual differences in student background and history, attitudes, academic experiences, and career aspirations? Does a student's racial/ethnic diversity provide a context for learning, interacting, and changing future expectations during law school?

Over 8000 incoming law students were assessed at 64 American Bar Association (ABA)-approved law schools during the first weeks of law school. Students self-reported about six different diversity domains: demographic background, family background, past experiences, perspectives and attitudes, educational expectations, and career aspirations. Supplementing these data on the institutional characteristics of the students' degree-awarding undergraduate institution and their enrolling law school, we examined associations between race and individual differences in a multilevel context. Incoming law student data were supplemented with a wide range of ancillary, archival information, including the Law School Admission Test (LSAT) registration information (e.g., race, gender, LSAT scores, counts of the number of times they took the Law School Admission Council (LSAC), law school application outcomes, such as where did students apply and where were they accepted, rejected, or waitlisted?), and first-year law school averages. A second assessment was administered prior to graduation, which tapped school experiences, attitudes, and professional aspirations.

To capture more nuanced views of the law school experience (from the student perspective) and how, if at all, student body diversity affected student learning, a subset of law students participated in focus groups during each year of law school. Each year the general themes for the focus groups varied. Students in the first year were asked about their learning experience, the diversity of the incoming class, and how race and gender were treated in analyses of specific legal cases. In the second year, they were asked about their activities, support sources (financial, emotional, and intellectual), mentoring, relationships with faculty, course choices, and perspectives on student diversity. In their last year, they were asked about

support sources during law school (financial, emotional, and intellectual), mentoring, relationships with faculty, future plans, and perspectives on student diversity.

Controlling for law school attributes of the schools where students were enrolled, the first major finding was that substantial variation exists in each of the six diversity domains as a function of race/ethnicity. Differences due to race/ethnicity were obtained in most areas of student background (e.g., one's family structure, parental educational and occupational achievement, racial socialization while growing up, religion, neighborhood while growing up, academic behaviors during college, and presence of a mentor), as well as in reported discrimination history (everyday discrimination, chronic discrimination, and perceived future discrimination in professional settings), race-related attitudes (e.g., domestic federal spending, negative attitudes toward immigrants, entitlements for all American citizens, symbolic racism, race relations, work ethic, gay/lesbian rights, affirmative action, current levels of discrimination toward minorities, low-income groups and religious groups, and value of educational diversity), and academic expectations (expectations about classmates, learning style, intentions to join student groups, personality, reasons for attending law school, desired future employment settings, desired job attributes, expected area of law, and anticipated eminence in law).

Not all variables under study showed variability as a function of race/ethnicity, including political orientation, sexual orientation, high school type, intent to work during law school, personal attributes relative to classmates (e.g., academic ability, writing ability, agreeableness, conscientiousness, emotional intelligence, and honesty/integrity), or attitudes about the military or certain social issues (e.g., rehabilitation for drug offenses, abortion, general fairness of the college process, and perceived discrimination against religious groups). Race/ethnicity also did not strongly predict certain academic behaviors as an undergraduate (e.g., whether or not students missed classes due to work and family responsibilities, met with faculty outside of class, participated in a student organization or in student government, or worked on assignments in advance of deadlines). Finally, in general, views about what might occur during law school did not differ as a function of student race/ethnicity (e.g., expected likelihood of learning about courtroom procedures, conflict resolution, and legal problem solving).

Longitudinal findings from the baseline assessment during the first weeks of law school to graduation showed that a law school with high racial diversity (i.e., having a high minority student representation vs. a low minority student representation; institutional domain) was related to reduced prejudice and increased exposure to diverse ideas (individual domain), and these effects were mediated by more frequent interpersonal contact with diverse peers (social domain).

Focus group findings confirmed several prior findings from undergraduate student samples that minority law students experienced alienation, lack of support, and everyday discrimination during all years of law school, especially students at institutions in which student racial diversity was low. Reported classroom experiences differed considerably as a function of race/ethnicity, and there was an oft-noted classroom-level tension between a professor's desire to adhere to strict legal analysis versus the consideration of a case's sociocultural context. For example, despite law schools' fairly uniform first-year curriculum, focus group participants identified strong differences in the extent to which issues around race/ethnicity were introduced into the discussion of certain legal cases, and these differences varied by professor and by the representation of minorities in the class. As a result of these observations, a national interview study is underway with a random sample of law professors probing whether and, if so, how they present case law and course materials related to race/ethnicity. The faculty interviews also inquire about the extent to which student racial diversity in their classrooms (vs. other forms of student diversity) impacts how they conduct their courses, about the extent to which difficult dialogs involving racially sensitive topics take place in the classroom, and about their viewpoints of student and faculty racial diversity in legal education.

The EDP's longitudinal and multiple method design has helped capture the law student's entire educational context, rather than taking a snapshot of a moment in time. We were able to assess their self-reported background, attitudes, and expectations when starting school; their undergraduate education's attributes; their law school's attributes; their concerns, opportunities, and academic performance during school; and their perceptions about their learning experiences upon completing school. Archival data and reports from professors about their pedagogy and interactions with students add nuance to the student trajectories under empirical study.

Conclusions

The legal state of educational diversity relies on the accumulation of empirical evidence across a range of educational settings (K-12, undergraduate, and graduate/professional schools) to identify educational benefits resulting from the exchange of ideas permitted in the context of a diverse student body. Research has clearly shown that diversity of background, family, positive and negative life experiences, attitudes, academic expectations, and aspirations vary as a function of race/ethnicity. That is, diversity of history and thought is associated with student racial diversity and is a lens for future educational interactions and learning. The increasing body of evidence about the effects of these sources of variability on

learning outcomes, interactions with peers of different backgrounds, exchange of ideas related and unrelated to course material, and preparation for active and meaningful participation in a global and diverse workforce will benefit from research that captures multilevel features of the student *in situ*. Such research should focus on students within their institution, as affected by their personal characteristics, family, experiences, attitudes, course choices, curricular and extracurricular experiences, peers, professors, and aspirations for the future.

Further Reading

- Allen, W. R. and Solorzano, D. (2001). Affirmative action, educational equity and campus racial climate: A case study of the University of Michigan Law School. *Berkeley La Raza Law Journal* **12**, 327–363.
- American Council on Education and the American Association of University Professors (2000). Does diversity make a difference? Three research studies on diversity in college classrooms. *Academe* **86**, 54–57.
- Antonio, A. L., Chang, M. J., Hakuta, K., et al. (2004). Effects of racial diversity on complex thinking in colleges students. *Psychological Science* **15**, 507–510.
- Bowen, W. G. and Bok, D. (1998). *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions*. Princeton, NJ: Princeton University Press.
- Brown v. Board of Education* (1954). 347 U.S. 483.
- Chang, M. J., Denson, N., Saenz, V., and Misa, K. (2006). The educational benefits of sustaining cross-racial interaction among graduates. *Journal of Higher Education* **77**, 430–455.
- Chang, M. J., Witt, D., Jones, J., and Hakuta, K. (eds.) (2003). *Compelling Interest: Examining the Evidence on Racial Dynamics in Colleges and Universities*. Stanford, CA: Stanford University Press.
- Franklin, J. H. and Moss, A. A. (2000). *From Slavery to Freedom: A History of African Americans*. New York: McGraw-Hill.
- Freedmen's Bureau Act of 1865 (1865). Pub. L. No. 38–90, 13 Stat 507.
- Grutter v. Bollinger* (2003). 539 U.S. 306, 123 S. Ct. 2325.
- Gurin, P. (1999). Expert report of Patricia Gurin. *The Compelling Need for Diversity in Higher Education*, Gratz et al. v. Bollinger et al. No. 97-75231 (E.D. Mich.) and Grutter et al. v. Bollinger et al. No. 97-75928 (E.D. Mich.), pp 99–234. Ann Arbor, MI: University of Michigan.
- Hurtado, S., Milem, J. F., Clayton-Pedersen, A., and Allen, W. R. (1998). Enhancing campus climates for racial/ethnic diversity: Educational policy and practice. *Review of Higher Education* **21**, 279–302.
- Kronman, A. T. (2000). Is diversity a value in American higher education? *Florida Law Review* **52**, 861–895.
- Lempert, R. O., Chambers, D. L., and Adams, T. K. (2000). Michigan's minority graduates in practice: The river runs through law school. *Law and Social Inquiry* **25**, 395–505.
- McKissick v. Carmichael* (1951). 187 F.2d 949 (4th Cir.), cert. denied, 341 U.S. 951.
- McNeil, G. R. (1984). *Groundwork: Charles Hamilton Houston and the Struggle for Civil Rights*. Philadelphia, PA: University of Pennsylvania Press.
- Milem, J. F. (2003). The educational benefits of diversity: Evidence from multiple sectors. In Chang, M., Witt, D., Jones, J., and Hakuta, K. (eds.) *Compelling interest: Examining Evidence of Racial Dynamics in Higher Education*, pp 126–169. Palo Alto, CA: Stanford University Press.
- Orfield, G. and Kurlaender, M. (eds.) (2001). *Diversity Challenged: Evidence on the Impact of Affirmative Action*. Cambridge, MA: Harvard Education Publishing.
- Orfield, G. and Whitley, D. (2001). Diversity and legal education: Student experiences in leading law schools. In Orfield, G. and Kurlaender, M. (eds.) *Diversity Challenged: Evidence on the Impact of Affirmative Action*, pp 143–174. Cambridge, MA: Harvard Education Publishing.
- Pigford v. Glickman*, 185 F.R.D. 82, 85 (D.D.C. 1999), *Affirmed*, 206 F.3d 1212, 340 U.S. App. D.C. 420 (D.C. Cir. 2000).

- Plessy v. Ferguson* (1896). 163 U.S. 537.
- Regents of the University of California v. Bakke* (1978). 438 U.S. 265, 291.
- Schmidt, P. (2003). Foes of affirmative action in Michigan plan to take their battle to the ballot. *The Chronicle of Higher Education* 9 July.
- Schuck, P. H. (2002). The perceived value of diversity, then and now. *Cardozo Law Review* 22, 1915–1960.
- Shaw, E. J. (2005). *Researching the Educational Benefits of Diversity: Research Report*. New York: The College Board.
- Solorzano, D., Allen, W. R., and Carroll, G. (2002). Keeping race in place: Racial microaggressions and campus racial climate at the University of California, Berkeley. *Chicano-Latino Law Review* 23, 15–111.
- Sullivan, R. S. (2003). Multiple ironies: Brown at 50. *Howard Law Journal* 45, 29–57.
- Sweatt v. Painter* (1950). 339 U.S. 629.
- The United States Constitution, Amendment XIII, Sections 1 and 2.
- The United States Constitution, Amendment XIV, Sections 1 and 5.
- The United States Constitution, Amendment XV, Sections 1 and 2.
- Wightman, L. F. (1997). *The Threat to Diversity in Legal Education: An Empirical Analysis of the Consequences of Abandoning Race as a Factor in Law School Admissions Decisions*, 72 N.Y.U. L. Rev. 1.
- Wygant v. Jackson Board of Education* (1986). 476 U.S. 267.

Relevant Website

<http://www.unc.edu> – Educational Diversity Project, The University of North Carolina at Chapel Hill.

Educational Reform

F Lopez-Segrera, Technical University of Catalonia, Barcelona, Spain

© 2010 Elsevier Ltd. All rights reserved.

Educational reform is a transformation plan and movement, which tries to bring about a systematic change in educational theory and practice in the level of basic or higher education (HE) in a given community and society.

Educational reform has been pursued for a variety of specific reasons, but generally most reforms aim at redressing some social problems – poverty, gender, race, or class-based inequities – or eliminating perceived ineffectiveness and lack of quality and pertinence, and inadequate management of funds.

It is not only true that the current changes and reforms of HE systems cannot be reduced to the impact of globalization, but it is also true that globalization – in its neo-liberal dimension, a concept that combines a market ideology with specific practices drawn from the world of business – has had a profound influence on educational systems during the last two decades of the past century and the start of the twenty-first century.

Our concise analysis has the following objectives:

1. refer to the impact of globalization at the world level over the last 20 years;
2. highlight the fact that certain levels of social equity are necessary in order to attain ongoing lifelong education as well as successful educational reform;
3. analyze the quantitative and qualitative impact of globalization on HE at the world level; and
4. formulate several proposals that would not only mitigate the negative impact of globalization on the education sector, but would also produce the necessary transformation of the main trends.

Among the latest trends, we can include the permanent updating of professionals and educational content; introduction of electronic networks for learning; translation and adaptation of the main scientific contributions; modification of institutional management; *ad hoc* public policies; and mutual enhancement of different educational models (public–private) and between different educational institutions – formal, informal, and distance. In the documentation of the World Conference on Higher Education (1998), convened by the United Nations Educational, Scientific and Cultural Organization (UNESCO) and its follow-up meetings, a series of valuable suggestions were adopted for confronting these global and regional challenges and problems.

It is important to identify the main trends in HE – quantitative expansion; growth of privatization; institutional diversification; public spending restriction; and

brain drain (UNESCO, 1995, 1998) – to solve the basic problems – reduction of public spending; changes in government policies; bureaucracy; rigid structures; and relationship to enterprise and production – and the challenges facing them. In the documentation of the World Conference on Higher Education (1998), convened by UNESCO, a series of valuable suggestions were introduced for confronting these challenges and problems.

The impact of globalization on the higher education institutions (HEIs) in Latin America and the Caribbean (LAC) is shown in the dialectics between three contradictory key themes: the public versus the private university (or administration vs. market), autonomy versus regulation, and production of knowledge versus mere teaching universities. Conversely, in the different perceptions of education and globalization, some attribute the changes to a different origin, minimizing the theory of great effects (Brunner), while others consider globalization to be the independent variable of the main changes (Gonzalez Casanova, Tunnermann, Didriksson, Lopez Segrera, Tedesco, Filmus, Broveto, Dias Sobrinho, and Boron).

Other key elements related to the globalization impact are the modifications that the TIC generates in the context and the debate concerning the possibilities of radically transforming our higher educational systems along UNESCO's guidelines, as well as following the guidelines of different reports from international organizations (e.g., UNESCO and World Bank (WB)), or from the national commissions of developed countries (Attali, Dearing, Boyer, and Bricall) or less-developed countries (ANUIES, Mexico). These reports, as well as the work of the principal Latin-American, European (Burton Clark, John S. Daniel, and Guy Neave), and North American experts (Altbach), despite diversity in scope, seem to suggest as guidelines permanent lifelong education for all; distance education; and institutionalization of the networks.

Our main hypothesis is that a dialectical synthesis is on its way, but many consider that market forces will have the final victory due to the exponential growth of private HE. The WB is evolving in its position and now asserts that investment in HE is important to development. On the other side, the state – as well as HE leaders – is convinced of the necessity of evaluation, and more diversified sources of financing.

In our opinion, globalization is a new qualitative phenomenon that arises from three interdependent processes that coincide simultaneously, and which have their

own internal logic: the crisis and downfall of socialism itself, neoliberalism, and the accelerated development of TIC.

Globalization is having a major impact on education in five principal areas (Carnoy, 1999: 15–17):

1. In the organization of work and types of work that people generate, which demands a higher level of education in the labor force and permanent ongoing training through *ad hoc* courses.
2. The governments of the developing countries are under increasing pressure to invest more in all levels of education, in order to create a better-equipped labor force, capable of using sophisticated techniques for production, which is the only way to compete in a more-globalized world market.
3. The quality and level of the educational system are growing at the international level. Curricula are becoming more complex and education, especially HE, should train students to handle new technology in several languages. On the other hand, there is an increasing tendency to exact accountability regarding the management of resources and results achieved, with regard to predefined objectives.
4. The virtualization of education is developing at an incredible pace, seeking to expand education, at a lower cost, through distance education. Education on Internet will become the predominant means of education, particularly HE.
5. Globalized information networks indicate the transformation of world culture. Those excluded from this world order fight against the values of this world culture, which possesses overwhelming market forces – as can be seen in the protests held in Davos, Seattle, Prague, and all those places where the *maîtres du monde* meet, the leaders responsible for the neoliberal policies and their consequences. They themselves admit that it is necessary to mitigate these policies, due to their disastrous consequences on the prosperity of the North and the adversity that has prevailed in less-developed countries.

Brunner states that it is not strange that the principal problem or weakness of the thesis of great effects lies in attributing the immediate cause of a variety of consequences to globalization, in the field of education or education policies – as a background context; consequences that without a doubt have very different origins. To be in the background is not necessarily to cause something. For instance, “the decentralization of education could be due to many causes and follow different national tendencies in different societies” (Brunner, 2001: 3).

Although this text by Brunner restores the multicausal theory, it tends to breakdown a specific model into causes and consequences of different origins, when in our opinion the independent variable that explains diminishing investment in education is closely tied to the triumph of

financial capitalism in its speculative form – neoliberal globalization.

The neoliberal model for universities has the following connotations: the university negotiates with the state, enterprises, and fathers of families who are able to pay for their children’s education. They all demand excellence in the field of knowledge and learning, which would be marketable, and try not to overreach the demand with surplus offers that would lower salaries and jobs. This new neoliberal university has a humanitarian policy whereby public or private foundations offer scholarships or credits to the poorer students who are not able to afford studies. “If the concept of the neo-liberal university puts emphasis on the public as well as on the private domain, the concept of university capitalism emphasizes the transformation of the university’s activities into merchandise” (Gonzalez Casanova, 2001: 102). They are conceived as market terms and this leads to the creation of corporate universities, to meet the necessities of the corporations, such as the case of the pioneer (in 1950), General Motors.

UNESCO proposes lifelong education for all for the recuperation (or inauguration) of democracy in education. This proposal – if put into practice – is incompatible with the neoliberal project.

It is not only true that the HE systems have progressed from simple and homogenous systems to ones that are more complex and diverse in the last decade, but it is also true that as a result of the deregulation of the 1980s policies have been drawn up for quality control, through institutional evaluation. It is clear that these are recently introduced policies that do not make a substantial difference with regard to the transformation of the HEIs. Nevertheless, they constitute an important change and a prerequisite for attaining real quality change. The mere market orientation of many private universities on the one hand, and the traditional concept of corporate autonomy on the other, seem to be some of the main obstacles impeding this quality change. The appropriation by the university of corporate logic is the perversion of autonomy. The university ceases to be an institution of and for society, and becomes an institution of and for members of the university community.

From the 1990s, even though a new conceptual framework was adopted, we cannot claim that these policies have essentially modified HE at the world level. What prevails is tension between the positive and negative aspects of the market and the state, respectively.

The analysis and systematic study of HE, and its irruption as an autonomous discipline, have hardly seen four decades. This interest in HE policies, associated with its emergence as a discipline, constitutes a study field that arose at the time of the disappearance of the classic university of elites when HE was mass extended at the beginning of the 1960s, showing a remarkable quantitative increment in enrolment. (Neave, 2001: 15).

The enrolment projections of UNESCO provide a worldly increase from 65 million students in 1991 to 79 million in the year 2000, 97 million in 2015 and 100 million in 2025. In 2004, we already had 134 million students enrolled in HE. This means a challenge for the countries of the South in relation with pertinence in an increasingly globalized world, in which competition will be more sophisticated in the knowledge market and where the young people of the poor countries have 17 times less chances of continuing a HE than those of the developed countries (GUNI, 2006).

The World Declaration on Higher Education in the 21st Century: Vision and Action (Paris, 1998), as a result of the World Conference summoned by UNESCO, states that there exists an unprecedented demand for a greater diversification of HE, and also an increasing degree of awareness of its vital importance for economic and socio-cultural development. The guidelines of this Declaration and the documents resulting from this Conference imply a solid change of paradigm, concerning the recommendations of HE policies (UNESCO, 1998; Segrera, 2001: 189).

In the WB report – *Higher Education: The Lessons of Experience* (1994) – it is stated: “there exists evidence that the investments in higher education have lower social rates than those made in primary and secondary education”, hereby these “will continue to be priorities for the Bank’s loans” (World Bank, 1994: 84–85). However, in a later report – *Higher Education in Developing Countries: Peril and Promise* (2000) – made by a task force of the WB and UNESCO, we observe that the WB in this document has changed its traditional position substantially. In chapter 2 of this report, it is stated: “the task force supports that great investments should continue to be done in primary and secondary education, but believes that the traditional arguments of economic character are based in a limited comprehension of the contributions of higher education institutions” (World Bank, 2000: 39).

The HE systems, according to the UNESCO/WB report – *Higher Education in Developing Countries: Peril and Promise* – should have the following characteristics or desired features (World Bank, 2000: 50–53):

1. Stratified structure.
2. Stable and long-range financing.
3. Competitiveness.
4. Flexibility.
5. Clear objectives and results.
6. Autonomy in front of political manipulations.
7. Narrow and well-defined links with the other levels of the general education system.
8. A legal system as a base for innovations and achievements.
9. An adequate information and database systems.

Concerning the reports developed by national commissions, we should mention the proposal of ANUIES in

Mexico, which, starting with the recommendations of the UNESCO World Conference of Higher Education (1998) comprises 14 programs grouped in three levels.

The *University 2000 Report* of Spain, known as the Bricall Report, since the ex-rector of Barcelona’s University, Joseph Bricall, coordinated it, came out with a request from the Conference of Rectors of Spanish Universities (CRUE), that a work group make some recommendations to improve HE in Spain. The report submits the constitution of a new social pact between the university, society, and state.

In the United States a report sponsored by the Carnegie Foundation and worked out by a commission directed by Ernest L. Boyer, president of the Foundation, – in the same way as the other commissions – carried out a critical diagnosis of HE in that country. A very severe criticism is made of the undergraduate system and it states as an objective the reinvention of the degree in universities of research. “The University of Research should facilitate the inquiries in the libraries, computers and study, expecting that the more advanced students and the proper professors and researchers body be the fellows and guides of the student. The University of Research gives to each student an integral educative experience, in which the whole of it is more profound and more comprehensive than what can be measured by won credits”.

This report shows – the ones that come from international agencies as well as those of independent national commissions – the priority that is given to HE at the world level in a knowledge society. Its key role in the training of human capital and its strategic dimension to modernize and give competitiveness to national economies are recognized. It stresses its value in citizenship training and in the acquisition of new identity values in the framework of the globalization process. It insists on the need for greater linkages with the social environment and there are formulated viable recommendations of specific transformations through a HE of quality (Segrera and Maldonado, 2002).

In ‘*Creando universidades innovadoras. Estrategias organizacionales para la transformación*’ (2000), Burton R. Clark defines the innovative universities, comparing them with European case studies – Warwick University, England; Strathclyde, Scotland; Twente, Holland; Chalmers, Sweden; and Joensuu, Finland. These institutes have changed from merely forming professionals, to becoming modern institutions with important development in scientific and technological research and technology, having a practical and innovative attitude. The essence of these transformation strategies, according to this author, is comprised of five common elements in the case studies:

1. strengthening of a central focus;
2. widening the periphery of extended development, going beyond the limits of the university’s boundaries

- to unite with external groups and organizations, via a transdisciplinary definition of problems;
3. diversification of financial sources;
 4. converting teaching staff and departments into dynamic innovative units; and
 5. creating a innovative work culture that adopts and promotes change.

In the following paragraphs, we summarize the main hypothesis of Boaventura Sousa Santos concerning university reform. According to him, the financial crisis in universities is related to their institutional crisis. Public universities have lost priority in states' public policies. This is due, above all, to the general decrease in the priority of social policies (e.g., education, health, and social security), caused by a model of economic development known as neoliberalism or neoliberal globalization. This model has spread around the world since the 1980s. In public universities, the imposition of this model meant that existing institutional weaknesses – of which there were many – were not used to create a pedagogical, political program of university reform. Instead, the problems were judged insurmountable and used to justify opening up universities – a public good – to commercial exploitation. Likewise, these weaknesses were used to explain the recapitalization and restructuring of public universities in favor of an emerging university market, where human resource transfers sometimes involved a rudimentary type of accumulation by private universities to the detriment of the public sector (Sousa Santos in GUNI, 2008).

Another contradiction also emerged in this area: the rigidity of university education and the changeability of the qualifications required by the market. The university became a service that one could access as a consumer, by paying for it. Thus, a citizen's right to education was drastically diminished. The number of free university education systems dropped and most grants were replaced by loans. As a result, students were transformed from citizens to consumers. The other main factor in the neoliberal project for universities is the transnationalization of university services.

As mentioned above, this neoliberal project is linked to a reduction in public funding. However, other factors are equally significant, such as a general deregulation of commercial exchanges; the defense, or even imposition, of the commercial solution by multilateral financial agents; and the high concentration of the potential benefits of new information and communication technologies in the North.

Public universities – and the education system as a whole – have always been linked to the project of nation-building. This project is now in crisis due to the progress of neoliberalism.

The only effective, liberating way to face neoliberal globalization, according to Boaventura de Sousa Santos,

is to challenge it with an alternative, counter-hegemonic globalization. National public university reforms should reflect a project of nation-building. This project has to be focused on political preferences that mark the integration of the country into contexts of knowledge production and transfer that are increasingly transnationalized and polarized between contradictory transnationalization processes: neoliberal globalization and counter-hegemonic globalization. Nation-building should be the result of an extensive political and social contract, specified in several sectorial contracts. One of these is the educational contract, which includes the idea of universities as a public good. The main aim of the reform is to respond to social demands for the radical democratization of universities. This should end the exclusion of social groups and their access to knowledge.

Counter-hegemonic globalization of universities as a public good is therefore a demanding political project. To attain credibility, it must squash two deeply rooted, conflicting, preconceived ideas: (1) that universities can only be reformed by university staff, and (2) that universities never reform themselves. The main protagonists are public universities themselves, that is, they are the ones interested in alternative globalization. The second protagonist capable of responding to these challenges is the nation-state. Finally, the third protagonist of the reforms proposed in this article are citizens, organized either individually or collectively, as social groups, unions, social movements, nongovernmental organizations and their networks, and progressive local governments. Reforms should be based on the assumption that in the twenty-first century a university only exists when three factors are present: graduate and postgraduate courses, research, and extension activities. If any of these factors is missing, there will be HE but no university (Sousa Santos, GUNI, 2008).

Universities must overcome a three-fold crisis, which they have been facing since the 1990s: a crisis of hegemony (it no longer has a monopoly on research), a crisis of legitimacy (it is perceived as an institution that is inaccessible to underprivileged people), and an *institutional crisis* (due to the difficulties it has in maintaining its independence in the face of pressure from market demands and the view of universities as companies). Therefore, reform should be undertaken, in accordance with a viable national project that considers education to be a public good and trains its graduates to work on sustainable development and equity (Sousa Santos, GUNI, 2008).

The principles and objectives that Didriksson proposes for the new education policy, which would become a new education pact, are consonant with the paradigms that UNESCO proposes for the twenty-first century. They are:

1. educating should be a task shared by both government and society;
2. education should have the maximum public priority;

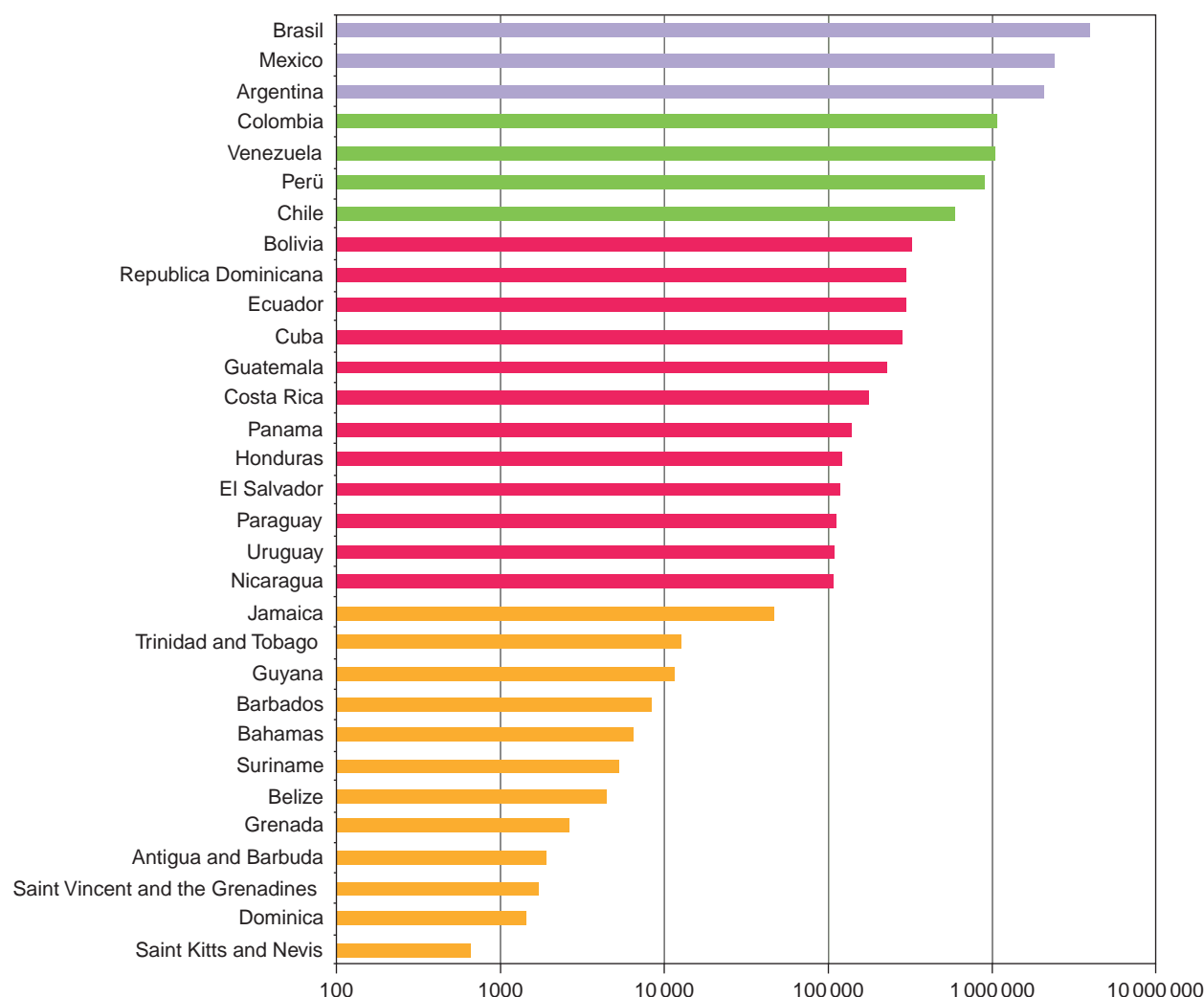


Figure 1 Diagnosis of higher education in Latin America and the Caribbean according to enrolment rates. Sixty percent of the enrolment of higher education in Latin America and the Caribbean (LAC) is concentrated in three countries: Brazil, Mexico, and Argentina. LAC had 13 991 517 students enrolled in universities in 2003, but in contrast, it still has 37 million illiterates. From IESALC/Informe de la ES 2000–05.

3. learning is the new teaching paradigm for change and innovation;
4. education is a key factor for poverty reduction;
5. quality is intrinsic to the processes, results, and social value of knowledge;
6. education should be pertinent and respond to criteria of equality and equity;
7. basic education needs to elevate the achievement average;
8. the role of HE needs to be redefined;
9. teaching staff need to be professionals;
10. formation for work should provide new abilities, capacities, and work competence;
11. HE needs to be urgently redefined, rationalized, and diversified; and
12. the objective should be to generalize the social capacity to create a knowledge society.

It is necessary to introduce new reforms in education (basic and well as higher levels) along UNESCO's guidelines, the Delors Report, Boaventura de Sousa recommendations, and Edgar Morin's ideas.

In conclusion, in spite of the advances made, we have still not been able to create a new university model where the production of knowledge predominates and not just the transmission and management of it. We are witnessing a crisis at the university, not only in its financial evaluation and curricula management but also in the concept of the university itself, which we should adjust to the circumstances, reflecting a growing identity crisis and basic suppositions. Some consider that this socialization deficit is linked to the social and teaching capacity crises of our traditional institutions such as the family and the school. The challenge consists of creating a new university in this climate of uncertainty, avoiding the victory of anomie and pessimism.

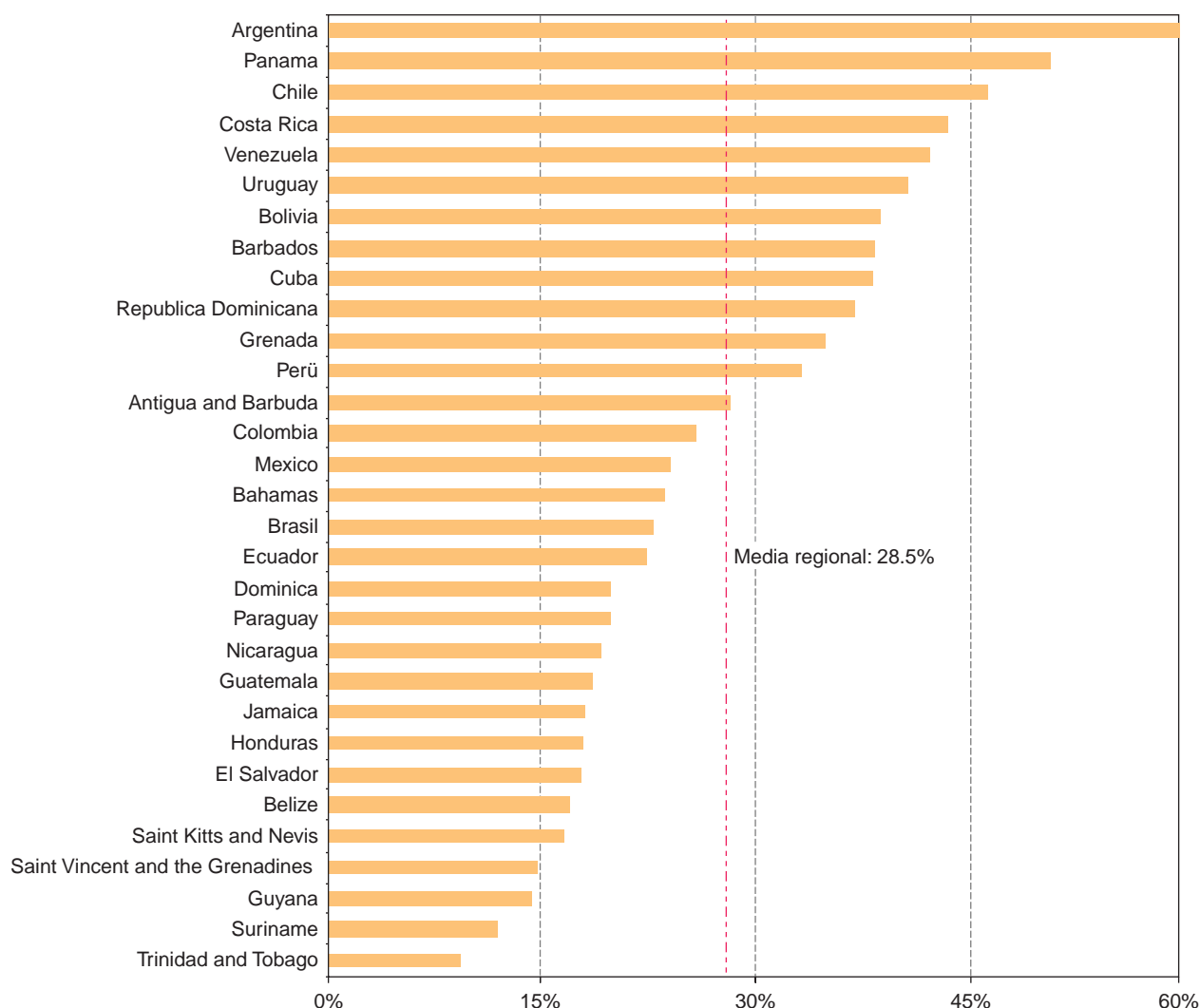


Figure 2 Regional average (28.5%) of higher education enrolment in LAC. Data from Series 2000–2004/UNESCO Follow Up Report 2006 1.

See also: Access and Equity in Higher Education; Higher Education and the Labor Market; Transforming Higher Education in Developing Countries: The Role of the World Bank; UNESCO's Role in the Development of Higher Education in a Globalized World.

Bibliography

- Brunner, J. J. (2001). *Globalización y el futuro de la educación. Tendencias, desafíos, estrategias. En Análisis de Prospectivas de Educación en América Latina y el Caribe*. Santiago: UNESCO.
- Carnoy, M. (1999). *Globalization and Educational Reform: What Planners Need to Know*. Paris: IIEP/UNESCO.
- Casanov, P. G. (ed.) (2001). *La universidad necesaria en el siglo XXI*. México: Era.
- GUNI (2006). *Higher Education in the World 2006. The Financing of Universities*. Basingstoke: Palgrave Macmillan.
- GUNI (2008). *Higher Education: New Challenges and Emerging Roles for Human and Social Development*. Basingstoke: Palgrave Macmillan.
- Neave, G. (2001). *Educación Superior: Historia y Política*. Barcelona: Editorial Gedisa.
- Segrera, F. L. and Maldonado, A. (2002). *Educación Superior Latinoamericana y organismos internacionales. Un análisis crítico*. Prefacio Philip G. Altbach. UNESCO: Boston College and Columbia: Universidad de San Buenaventura de Cali.
- UNESCO (1998). *La educación superior en el siglo XXI: Visión y acción. Conferencia Mundial sobre la Educación Superior. Documentos Varios, Octubre, París*.
- World Bank (2000). *Higher Education in Developing Countries: Peril and Promise*. Washington, DC: World Bank.

Further Reading

- Brunner, J. J. (2006). Documentos varios elaborados para el Curso impartido en el Centro Universitario Salesiano de Sao Paulo. <http://mt.educarchile.cl> (accessed August 2009).
- Didriksson, A. (2000). Tendencias de la educación superior al fin de siglo: Escenarios de cambio. En Tünnermann, C. y López Segrera, F. *La educación en el horizonte del siglo XXI*, pp 185–199. Colección Respuestas No. 12. Ediciones IESALC/UNESCO-Caracas.

- Didriksson, A. (2005). *Recent Tendencies in Higher Education Financing: The Case of Mexico*. Paris: IAU-Horizons.
- Forrest, J. F. and Altbach, P. G. (2006). *International Handbook of Higher Education*. Dordrecht: Springer.
- Gazzola, A. L. (2008). *Integración académica e internacionalización de la educación superior*. Conferencia en Universidad 2008, Habana.
- OECD (2007). *Education at Glance*. Paris: OECD.
- Seddoh, F. K. (2002). *The World Conference on Higher Education: Which Strategies and Which Actions to Follow?* Key Note Speaker at University 2002, La Habana.
- Segrera, F. L. (2006). *Escenarios Mundiales de la Educación Superior*. Buenos Aires: Ediciones CLACSO.
- UNESCO (1997). *Consolidated Declarations and Plans of Action of the Regional Conferences on Higher Education Held in Havana, Dakar, Tokyo and Palermo*. Retained Lessons. Division of Higher Education. Paris.
- UNESCO (2004). *Final Report of the Meeting of Higher Education Partners*. Paris: WCHE.
- UNESCO (2006). *EFA Global Monitoring Report*. Paris: UNESCO.
- UNESCO (2006). *Global digest. UIS*, Montreal.
- World Bank (2002). *Constructing Knowledge Societies: New Challenges for Tertiary Education*. Washington, DC: World Bank.

Relevant Websites

- <http://www.gatswatch.org> – Acuerdo General sobre el Comercio de Servicios (GATS), GATSwatch.
- <http://www.unesco.org> – Asociación Internacional de Universidades (AIU), UNESCO.
- <http://www.upc.es/cudu> – Cátedra UNESCO de Dirección Universitaria (CUDU), Universidad Politécnica de Cataluña (UPC).
- <http://www.bc.edu> – Center for International Higher Education (CIHE), Boston College.

- <http://www.cinda.cl> – Centro Interuniversitario de Desarrollo (CINDA).
- <http://www.crue.org> – Conferencia de Rectores de las Universidades españolas (CRUE).
- <http://www.clacso.org.ar> – Consejo Latinoamericano de Ciencias Sociales (CLACSO).
- <http://www.fiuc.org> – Federación Internacional de Universidades Católicas (International Federation of Catholic Universities (IFCU)).
- <http://www.educarchile.cl> – Fundación Chile, Educarchile.
- <http://www.guni-rmies.net> – Global University Network for Innovation (GUNI).
- <http://www.iesalc.unesco.org.ve> – IESALC.
- <http://www.iadb.org> – Inter-American Development Bank.
- <http://www.usp.br/nupes> – Núcleo de investigación sobre la educación superior de la universidad de Sao Paolo (NUPES), Universidade de São Paulo (USP).
- <http://www.orus-int.org> – Observatoire International des Reformes Universitaires (ORUS).
- <http://www.OECD.org> – Organisation for Economic Co-operation and Development.
- <http://www.cepal.org> – Portada de la CEPAL (Economic Commission for Latin America (ECLA)).
- <http://www.universia.net> – Proyecto universia.net, Universidades iberoamericanas. Universia.net.
- <http://www.unam.mx/coordhum/riseu> – Red de investigadores en educación superior (RISEU), Universidad Nacional Autónoma de México.
- <http://www.sida.se> – Swedish International Development Cooperation Agency (SIDA).
- <http://www.worldbank.org> – The World Bank.
- <http://portal.unesco.org> – UNESCO FORUM.
- <http://www.uis.unesco.org> – UNESCO Institute for Statistics (UIS).
- <http://unesdoc.unesco.org> – UNESDOC – UNESBIB.
- <http://www.undp.org> – United Nations Development Programme.
- <http://portal.unesco.org> – United Nations Educational, Scientific and Cultural Organization.
- <http://www.unu.edu> – Universidad de las Naciones Unidas (UNU).

Emergence of For-profit Higher Education

X Li, University of California, Los Angeles (UCLA), Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Growth of For-Profit Higher Education

For-profit higher education in the United States remains with a relatively small sector, when compared its traditional counterparts, public and private nonprofit colleges and universities. The total education industry is estimated at about \$826.6 billion in 2003. Higher education expenditures were generally estimated at about \$315.4 billion in 2003 for an enrolment of about 17.4 million students (US Department of Education, 2005). The corporate education and training market comprises another \$75 billion and is expected to grow more rapidly than any other segment. The overall scale of the degree-granting, for-profit post-secondary institutions currently remains a small sector. The four degree-granting institutions most cited – the University of Phoenix, DeVry Inc., Strayer Education Inc., and ITT Technical Institute – together account for more than 100,000 students and over \$1 billion (Breneman *et al.*, 2006). Thus, the overall scale of the degree-granting for-profit post-secondary institutions is currently quite modest.

Conversely, for-profit higher education is often considered an emerging phenomenon. This is particularly true if we look at its rapid and tremendous growth in the past 10 years. For example, the for-profit degree-granting enrolment increased by more than 200% between as compared to 1995 and 2006 public institutions 20% for and –20% for private nonprofit institutions (Table 1). In the 2005–06 academic year, the number of for-profit degree-granting institutions is twice the number in 1995–96 (from 345 to 936), while the numbers for public and private nonprofit institutions have remained almost unchanged (US Department of Education, 2008).

Furthermore, if nondegree-granting institutions are included, there are currently more for-profit post-secondary institutions than public and private nonprofit post-secondary institutions. The 2006–07 census of institutions by the National Center for Education Statistics documents 2679 for-profit institutions, 1848 private nonprofit institutions, and 2009 public nonprofit institutions (Table 2).

According to the literature, for-profit higher education is an understudied area even though this sector has received much attention in recent years, due in part to a rapid development, in ways that traditional higher education has not. This article examines the recent growth of for-profit higher education, focusing on degree-granting higher education institutions in the United States. It explores the driving forces for such a rapid growth.

Throughout the article, different voices for the development of for-profit institutions are listed and compared. Their similarities to and differences from traditional (public and private nonprofit) colleges and universities are also presented. The article concludes with a discussion of potential directions and developments for this sector.

For the purpose of this article, the theory of academic capitalism is utilized to examine influential current processes of for-profit higher education in a global context. Slaughter and Rhoades (2004) describe how higher education is linked through networks to the new economy. They move beyond the conception of students as consumers and conceive colleges and universities as marketers that compete intensely at different levels. The theory of academic capitalism explicitly encompasses processes such as commercialization and marketization, and goes beyond these in several aspects. Most importantly, it conceptualizes changes to the system as a general shift from a public good knowledge/learning regime to an academic capitalist knowledge/learning one. Whereas the former was characterized by valuing knowledge as a public good to which the citizenry can lay claim, the latter regime values knowledge privatization and profit taking in which institutions, inventor faculty, and corporations have claims that come before those of the public. This concept of knowledge economy is well observed in all aspects of for-profit higher education sector.

Driving Forces for the Growth of For-Profit Higher Education

Several reasons and forces have been identified as contributing in part to the growth of for-profit higher education in the United States. Among such forces, permeation of globalization, technology advancement, decreased federal and state funding, market factors, as well as for-profit institutions' unique methods of catering to the special needs of adult learners are widely accepted.

Globalization and Commercialization Serve as a Broader Context

One major force for change in higher education is the globalization of economic, sociocultural, political, and technological factors, along with the increasing interdependence of nations. When the promotion of a free market constitutes the core of globalization theories and practices, efficiency is believed to be the key to success both at

Table 1 Numbers of degree-granting institutions and enrollment by level and control (1995–96 vs. 2005–06 academic years)

	<i>Public</i>		<i>Private-non-profit</i>		<i>For-profit</i>	
	<i>Two year</i>	<i>Four year</i>	<i>Two year</i>	<i>Four year</i>	<i>Two year</i>	<i>Four year</i>
Number of institutions						
1995–96	1047	608	187	1,519	228	117
2005–06	1053	640	113	1,534	528	408
Enrollment						
1995–96	5 314 463	5 806 036	75 375	2 867 181	173 489	130 976
2005–06	6 184 229	6 837 605	43 522	3 411 170	260 304	750 645

From US Department of Education, National Center for Education Statistics, IPEDS enrollment and institutions (1995, 2006).

Table 2 Degree-granting and non-degree-granting institutions, by control and type of institution (2006–07)

<i>Number of institutions</i>	<i>Total</i>	<i>Public non-profit</i>	<i>Private non-profit</i>	<i>For-profit</i>
Degree-granting	4314	1688	1640	986
Non-degree-granting ^a	2222	321	208	1693
Total number	6536	2009	1848	2679

^aReferring to non-degree-granting Title IV institutions.

From US Department of Education, National Center for Education Statistics, IPEDS enrollment and institutions (2006–07).

the institutional and individual levels. As Altbach and co-workers claim, “the correct path to improvement and prosperity is that of deregulation, decentralization, privatization, information technology (IT)” (Altbach *et al.*, 2005: 1).

Public good or private good

With the rapid growth of the for-profit sector, the controversy concerning the legitimacy of for-profit colleges growing is also in academic, governing, and accreditation agencies. They are criticized for lack of qualified faculty, predominant use of distance-teaching and learning approaches, lack of library resources, and so on. However, the core of the controversy has to do with the nature of the for-profit institutions’ motive – profit—which leads to the ongoing debate whether higher education is a public good or a private good, a relationship that has been cause for great concern.

Higher education – public institutions in particular – was traditionally and mainly sponsored by federal and state funding. Therefore, one of its major missions was to provide public service or public good to society. Higher education as a public good was dedicated to the idea that US universities must be more responsive to the changing needs of American society and the development of human beings. That is why wordings, such as commitment,

engaged, service, and partners with their communities are often seen in mission statements of colleges and universities.

However, with the growing commercialization of higher education, driven in part by globalization, the values of the marketplace have entered the campus. One of the main factors is the change in society’s attitude toward higher education, which is now seen as a private good benefiting those who study or do research. Slaughter and Leslie (1997) suggest that the nonprofit university research enterprise has already moved significantly in the direction of commercial production at the expense of public good.

According to Slaughter and Rhoades (2004), as colleges and universities become more entrepreneurial in a postindustrial economy, knowledge is less seen as a public good and more as a commodity to be capitalized on in profit-oriented activities. By emphasizing the role of the market when defining academic capitalism, they describe and analyze the aggressive engagement of US higher education institutions in the knowledge-based economy and the efforts of colleges and universities to develop, market, and sell research products, educational services, and consumer goods in the private marketplace.

From this perspective, it seems reasonable that the consumers should pay for this service as they would for any other. The provision of knowledge becomes just another commercial transaction. The main provider of public funds, the state, is increasingly unwilling or unable to provide the resources needed for an expanding higher education sector. Universities and other post-secondary institutions are expected to generate more of their funding themselves. They have to think more like businesses and less like educational institutions (Slaughter and Rhoades, 2004).

In this environment, the privatization of public universities is unavoidable as they are selling knowledge products, partnering with corporations and increasing student fees. The proliferation of private academic institutions of all types, especially in the for-profit sector, is another byproduct of commercialization (Altbach, 2001). Education companies, some of which call themselves

universities, sell skills and training, awarding degrees or certificates to their customers (students).

The Institute for Higher Education Policy (IHEP), in examining the value of higher education, explicitly made the case for viewing personal and public, economic and social benefits as a whole (IHEP, 1998). They point out that public dialog has predominantly moved away from an appreciation of the multiple societal and personal benefits of education. The focus now, according to IHEP, is personal economic benefit.

In these changed circumstances, it is not surprising for the World Trade Organization (WTO) and its related General Agreement on Trade and Services (GATS) to propose that knowledge products are freely traded in the international marketplace. Many of these developments link academic institutions and systems globally. The use of English as the lingua franca for scientific communication and for teaching, especially when combined with the Internet, makes communication easier and quicker. The advent of multinational higher education institutions makes it possible to spread new curricular and other innovations quickly and to meet the immediate needs of students.

Conversely, while much of the concern over the growth of for-profit education points to its leading to private-good, it has also been argued that the expansion of non-profit education in the post-war period and the concurrent implementation in those institutions of public policies on affirmative action and gender equity have contributed significantly to increased access, diversity, and efficiency in higher education (Hurtado and Navia, 1997).

Funding Cut for Public Universities

During the mid-1990s, public educational institutions in many states faced increasing criticism and tough state legislatures as well as tighter-than-ever federal and state funding. State legislatures, increasingly sought to hold public colleges to higher standards by implementing accountability regulations that tied at least some funding to the performance of the institution (Burke *et al.*, 2000).

Many states increased the reporting requirements for their public colleges as a step to encourage greater efficiency and accountability. Influential analysts foresee a radically changing and much more competitive higher education landscape in which the traditional, established institutions are threatened by burgeoning educational providers and new forms of educational technologies. Thus, Frank Newman, the past president of the Education Commission of the States, in an article subtitled 'The end of the *status quo* and the rise of the market in higher education' argues that "competition is forcing a hard reexamination of the purpose and effectiveness of every activity – from how well and often faculty interact with students, to whether expenditures on student life actually create a learning community, to the issue of costs and wise use of resources"

(Newman, 2001: 9). As institutions seek to make up for declining federal and state funding, public colleges and universities are increasingly becoming privatized (Zusman, 2005).

On the other hand, it has been argued with regard to costs, that the different institutional norms of the for-profit institutions may enable these providers to reduce costs and achieve greater efficiencies than their peers in the nonprofit sector. The costs of higher education at public and private institutions and the increased burden on individual students are creating enormous pressure on the nonprofit institutions in the United States and around the world (Ehrenberg, 2000). This in turn suggests an increasing competitive leverage for the for-profit sector if those institutions are able to utilize capital investments and lower wage scales to maintain lower overall operating costs. According to the National Sciences for Education Statistics (NSES, 2008), at public institutions in 2004–05, the average total expenditure per full-time-equivalent student was \$23 353. At private not-for-profit institutions, the total expenditure per full-time-equivalent student rose by 14% between 1996–97 and 2004–05, after adjustment for inflation. In 2004–05, the total expenditure per full-time-equivalent student at private not-for-profit institutions was \$38 472. The expenditure per student at for-profit institutions was \$11 205 in 2004–05, which is significantly lower.

In contrast to the reduction in public state funding, growing interest in education by ventures and corporations has contributed to the growth in for-profit higher education. For the high rates of return based on the human capital theories proposed by Becker (1964, 1993), these investors provide crucial financing for new companies as they see high potential for rapid growth here. For similar reasons, these firms are more interested in investing in post-secondary education, including for-profit and technical schools, rather than K-12 (Breneman *et al.*, 2006).

Technology Development Makes Distance Learning Possible and Easy

It was widely believed that campus-based universities and colleges were an important, if not the only, transition platform for youth from high school to join the workforce and job market. The introduction and increasing use of technology has greatly expanded access to higher education and fundamentally changed the models of education. In particular, technology has changed education from teacher centered to learner centric, making education far more relevant to the needs of individuals. Technology has freed education from being time- and place-bound, as it allows for anytime, anyplace learning, which is particularly flexible and attractive to working adults. Moreover, it enables lifelong learning in the true sense.

As Green (1999) argued, information technology increases access to higher education and the growing numbers of adults engaging in lifelong learning are fostering an evolutionary event in American higher education: “Technology brings new, rich resources to the learning of content; creates new contexts for interaction between and among instructors and students; and can fundamentally change the way students and institutions approach assessment and certification” (p. 14). It is fair to say that emerging technologies are changing the way of teaching and learning in both nonprofit and for-profit sectors; however, they provide more opportunities to for-profit education and training organizations not only because they fit working adults better, but also because corporations who operate for-profit institutions respond to modern technologies faster.

The Internet, for example, is used as the preferred primary technology to improve instruction, increase access, and raise productivity in higher education. A great number of universities, public and private alike, offer some extension or degree credit courses over the Internet. However, number of these types of courses remain relatively small, while for for-profit institutions, University of Phoenix for instance, using the Internet to reach geographically dispersed students has become one of the major means of instruction in their curriculum.

For-Profit Institutions Meet the Demand for More Skillful Labor Forces

Among the reasons believed to be responsible for the growth of for-profit higher education, their clear and narrow focus on professional training as well as customer-first and convenience-centered services are commonly accepted.

A global job market emerges as markets become more globalized. More and more corporations tend to recruit the skilled workers needed for these markets, and there is a growing demand for accessible and career-oriented adult education at the post-secondary level to fill this need. The increase of the general demand for the training, skills, and credentials offered in higher education, the needs of adult learners, and the rising tuition at traditional colleges and universities have stimulated the emergence of for-profit higher education in the United States as a viable sector of post-secondary education (Breneman *et al.*, 2006).

According to Levine (2001), the demographic composition of college students in the United States has changed. Traditionally, full-time 18- to 24-year-old residential students comprise the majority of the student population. However, this is no longer the case. In fact, only 16% of college students today fit this traditional profile. As a growing trend, over 60% of students are attending school part time and working part time. About 45% of college students are 24 years of age or older

(Levine, 1997). This explains clearly why much of the growth in enrolment in the last two decades has occurred among older, nontraditional students.

This growing trend provides great opportunities for for-profit institutions, and for them the key to success is their ability to accommodate the special needs of nontraditional adult students, who need the right credentials to obtain good employment as their primary goal. Recognizing that they can not receive a degree from a reputable university, adult learners enroll in for-profit institutions, which focus on convenience and shorter period of study time. They prefer courses in the evening and on weekends. Most importantly, easier access and lower cost of attendance than afforded by the traditional private universities and colleges in their region matter to most for them.

Levine (1997) found in a study of undergraduate attitudes and experiences that older, part-time, and working students, especially those with families and children, want a different type of relationship with their college – a relationship that is more like those they have with their bank, supermarket, or gas company. They want their colleges to be convenient, accessible, high quality for low cost, open during the evenings and on weekends, and having helpful staff, sufficient parking, and no waiting in long lines, and so on. Many students in the study preferred distance education so that they could study at home or in the office. They viewed higher education as one of many routine daily activities. They valued convenience, expected high-quality service, and were cautious about cost. In brief, for-profit institutions focus on students as customers and provide services for them that minimize the amount of bureaucratic control through which a student very likely has to pass in a traditional public or private nonprofit college or university.

The success of these institutions also depends on their market-oriented programs, such as international business, business administration, criminal justice, information technology, and health. Instead of helping students become well-rounded persons by giving them a liberal education, as these institutions promise in their mission statements, the promise students that they can complete the program in a shorter period of time, and get a better job. According to Ruch, (2001), the US for-profits sector pays more attention to placement rates, which help with their advertisement and recruitment. For example, they range from 96% at De Vry to 76% at Strayer. Similarly, the return on educational investment (ROEI) is higher than for the average BA: 28% as compared to 18.7%.

Coexistence of Distinctions and Similarities

When people talk about for-profit institutions, they are more likely to do so by making comparisons to their

nonprofit counterparts. Along with the debate over higher education being public good or private good, the essential difference between private nonprofits (e.g., Harvard University) and private for-profits (e.g., University of Phoenix) is that for-profit institutions provide education to make profit, while nonprofit colleges and universities charge tuitions to provide education. This fundamental difference in mission makes a great difference in areas such as admissions (required tested scores and comprehensive evaluation when entering nonprofit institutions vs. 'come on in' welcoming at for-profit), instruction and curriculum (walled-classroom liberal education vs. market-oriented e-learning courses), governance (more corporation management style at for-profit), services (more customer-first services at for-profit), and so on.

The source of their revenues has constituted a fundamental distinction between nonprofit and for-profit organizations. Traditionally, the public nonprofit sector received significant direct funding from the federal and state governments. Private nonprofits received large endowments and benefited from tax treatment. On the opposite end of the spectrum, for-profits relied solely on tuition income and tax benefits that their affiliating corporations received from the government. Consequently, different sources of funds determine their specialization, with public nonprofit institutions dominating the highly subsidized 2-year sector, private nonprofit institutions dominating 4-year degree-granting institutions, and for-profit institutions dominating the 2-year certification and training institutions (Pusser, 2005).

Although great differences between for-profit and nonprofit sectors can be observed, researches on institutional forms, using the theory of institutional isomorphism (DiMaggio and Powell, 1983) suggest that there is a convergence of nonprofit and for-profit institutions because of the increasing adoption of commercial behavior in nonprofit higher education institutions. In the mean time, for-profit institutions intend to be like nonprofit by means of one or another for the purpose of advertisement, recruitment, or accreditation.

Similar sources of revenue are attempted by both nonprofit and for-profit higher education institutions. In response to the reduction of state funding, a great number of public and private nonprofit institutions have initiated innovative efforts to offer programs and degrees without direct state subsidies. They charge tuition at a similar, if not the same, level, as the for-profit institutions. We see a rapid growth of affiliated enterprises, continuing education programs, industry-university partnerships, and the creation of for-profit subsidiaries of nonprofit institutions (Pusser, 2005). As both nonprofit and for-profit face the lack of public funding, the share of revenue coming from tuition charges and commercialized service activities has risen for them all.

Conversely, in the past 35 years, due to the passage of the 1972 Higher Education Act, students in for-profit

institutions are becoming increasingly eligible to apply for federal financial aid, a service previously available only to the nonprofit sector. They also benefit significantly from tax subsidies in which some or all of their tuition is paid by their employers, who in turn receive tax deductions for those payments. Students can often fund their tuition payments through federal subsidies, including financial aid programs such as the Pell grant program, subsidized student loans, and tax credits.

Conclusion

The for-profits have proven successful in competing with the nonprofits. In a gradually globalized market, which demands more skillful workers, the needs of adult learners and the rising tuition at traditional colleges and universities have stimulated the emergence of for-profit higher education. Increased availability of investment capital, excellent job-placement records, freedom from traditional curricula, lower costs through increased productivity, economies of scale, and state-of-the-art technology are cited as key assets of the for-profits. Taken together, these factors are seen as enabling the for-profit degree-granting institutions to capture market share from the nonprofits, and their perceptions are in accordance with those expressed by investment industry analysts and for-profit providers (Ortmann, 1998). The for-profit sector has successfully managed to become an active, viable, and financially successful part of the post-secondary education landscape.

However, for-profit providers currently represent a tiny fraction of the total degree granting activity in American higher education. The various layers of regulation requirements for accreditation in different regions of the United States also create significant barriers for for-profit higher education. By and large, it is believed that for-profit institutions are not a significant competitive threat to most of traditional higher education, despite views to the contrary expressed by some observers (Breneman *et al.*, 2006). For-profit colleges and universities have so far remained on the periphery of higher education. It is less than likely that for-profits will be able to supply a product that can match the peer effects and signaling value of elite nonprofit higher education institutions.

As Bowen (1986) wisely puts, "The future of higher education is on the whole not predictable . . . [it is] inscrutable" (p. 8). It is even more so when higher education has undergone so many dramatic shifts the twenty-first century. As the United States, as well as in any other parts of the world, is undergoing the severest economic recession probably since the Great Depression, it is reasonable to foresee the continuing and, more likely, dramatic cut of public and government funding in higher education, which will very likely affect the development of the nonprofit sector, both public and private universities. However, this can be a very difficult time for the for-profit

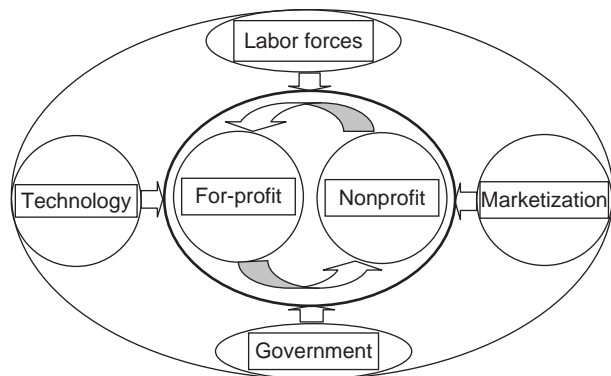


Figure 1

sector as well, given that many corporations have closed down, the credit market may affect student loan programs, and high unemployment ratios indicate a shrinking labor force. Thus, it is safe to say that, in the long run, the for-profit sector will continue to grow, as many of its major driving forces discussed in this article are growing, such as permeation of globalization, technology advancement, continued shrinkage of federal and state funding, and for-profit institutions' unique catering to special needs of adult learners.

Bibliography

- Altbach, P. G. (2001). Higher education and the WTO: Globalization run amok. *International Higher Education* 23, 2–4.
- Altbach, P. G., Berdahl, R. O., and Gumpert, P. J. (2005). *American Higher Education in the 21st Century*. Baltimore, MD: Johns Hopkins University Press.
- Becker, G. S. (1964, 1993). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, 3rd edn. Chicago, IL: University of Chicago Press.
- Bowen, H. R. (1986). *American Higher Education: Problems and Opportunities*. Howard R. Bowen Lecture. Iowa City, IA: University of Iowa.
- Breneman, D. W., Pusser, B., and Turner, S. E. (2006). *Earning from Learning: The Rise of for-Profit Universities*. New York: SUNY Press.
- Burke, J. C., Rosen, J., Minassians, H., and Lessard, T. (2000). Performance funding and budgeting: An emerging merger? *The Fourth Annual Survey (2000)*. Albany, NY: The Rockefeller Institute.
- DiMaggio, P. J. and Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review* 48, 147–160.
- Ehrenberg, R. G. (2000). *Tuition Rising: Why College Costs So Much*. Cambridge, MA: Harvard University Press.
- Green, K. (1999). When wishes come true: Colleges and the convergence of access, lifelong learning, and technology. *Change* 31(2), 11–15.
- Hurtado, S. and Navia, C. (1996). Reconciling college access and the affirmative action debate. In Turner, C., Garcia, M., Nora, A., and Rendon, L. (eds.) *Racial and Ethnic Diversity in Higher Education*. ASHE Reader Series. Needham, MA: Simon & Schuster.
- IHEP (Institute for Higher Education Policy) (1998). *Reaping the Benefits: Defining the Public and Private Value of Going to College*. Washington, DC.
- Levine, A. (1997). How the academic profession is changing. *Daedalus* 126, 1–20.
- Levine, A. (2001). The remaking of the American University. *Innovative Higher Education* 25, 253–267.
- Newman, F. (2001). *Interesting Times: The End of the Status Quo and the Rise of the Market in Higher Education*. Providence, RI: The Futures Project, Brown University.
- NSES (National Center for Education Statistics) (2008). *Digest of Education Statistics 2007: NCES 2008-022*. Institute of Education Statistics. Washington, DC: GPO.
- Ortmann, A. (1998). The emergence of the for-profit higher education sector: Recent developments and some prognostications. *Uncirculated Draft*. Bowdoin College Economics Department.
- Pusser, B. (2005). What are those data worth? The problems of conducting research on for-profit colleges and universities. *Unpublished Manuscript*. University of Virginia: Charlottesville, VA.
- Ruch, R. S. (2001). *Higher Ed, Inc.: The Rise of the for-Profit University*. Baltimore, MD: Johns Hopkins University Press.
- Slaughter, S. and Leslie, L. (1997). *Academic Capitalism: Politics, Policies, and the Entrepreneurial University*. Baltimore, MD: John Hopkins University Press.
- Slaughter, S. and Rhoades, G. (2004). *Academic Capitalism and the New Economy: Markets, State and Higher Education*. Baltimore, MD: Johns Hopkins University Press.
- Zusman, A. (2005). Challenges facing higher education. In Altbach, P., Berdahl, R. O., and Gumpert, P. J. (eds.) *American Higher Education in the 21st Century*, pp 115–160. Baltimore, MD: Johns Hopkins University Press.

Engaged Education: Experiential Learning, Intensive Field Experiences, and Social Change

T R Rone, Morgan State University, Baltimore, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Engaged education – A pedagogical tool, which promotes and fosters active, invested relationships between learners, educational institutions, and communities.

Experiential education – A philosophy of education, which emphasizes direct, active learning experiences.

Field experience – Similar to field trip in referring to a learning outside of the classroom. Distinguished from trip by a focus on extended, regularly occurring firsthand experiences. May be undertaken individually or as a group.

Field trip – Learning opportunities outside the classroom, generally as a group and limited in duration to a few hours.

Intensive field experience – Firsthand learning opportunities, which take place outside of the classroom.

Introduction

The concepts of engaged education, experiential learning, intensive field experiences, and social change are distinct, yet have the potential to be inextricably linked. In this article, it is understood that the four are indeed linked if one approaches the process of education as one leading, directly or indirectly to education as an active learning process and a product of relationship building. The article presents a discussion that: (1) describes engaged education; (2) defines and discusses the interrelatedness of this article's guiding concepts – experiential learning, intensive field experiences, and social change—as relevant to engaged education; (3) discusses the uniqueness of intensive field experiences as an effective pedagogical strategy; (4) illuminates the aims, components, and strategies of experiential education, and specifically, intensive field experiences, through a discussion of examples suitable for use in undergraduate and graduate social science courses; (5) describes a model for working with economic, transportation, housing, and social justice issues characteristic of urban areas through institutional strategies and resources that facilitate capacity building for community building and empowerment in local

communities; (6) outlines successful strategies for implementing intensive field experiences as an example of experiential learning; and (7) closes with a summary and overview. In an era when complex social issues of inequality, poverty, and injustice continue to markedly inform experiences of education, health, justice, housing and environmental protection, the construct of an engaged university, indeed, a globally engaged university, is increasingly becoming a norm rather than an anomaly.

The term engagement used in a more traditional sense denotes a betrothal, military battle, employment, or run of a performance. Since the 1960s, the term engagement is frequently used in educational contexts to describe learner engagement, student engagement, classroom engagement, school engagement, civic engagement, and political engagement. The concept of globalization has inspired the notion of global engagement. The concept of relationship is central to the meaning of the term engagement in more traditional and educational contexts. For example, in university settings, engagement marks the relationship between students as learners and faculty as teachers, between universities and communities, and between individuals and communities or society as a whole. These are not intended as diametrically opposed categories, but as attempts to highlight existing roles in universities and communities. The overall emphasis of this article is on the potential significance of strategies for dismantling the traditional barrier that exists between classrooms and communities. Engaged education addresses critical long-standing issues related to the extent to which universities and communities perceive their roles and responsibilities as unrelated. In contrast, engaged education calls for partnership, for merging resources and assets, as by the overlapping circles on the right side of **Figure 1**.

The concept of engagement, as embedded in the learning process, has been integral to human survival for centuries. The complete absence of engagement in learning carries substantial costs to human populations and society, as knowledge necessary for survival would fail to be successfully communicated. Broadly defined, engaged education can be characterized as a relationship between participants who in the roles of students, teachers, universities, and communities, are active co-participants in the instructional process. Engaged education refers to relationships characterized by learning via seeing, doing, and experiencing first hand, within and between what is generally perceived as a distant polar relationship between universities and

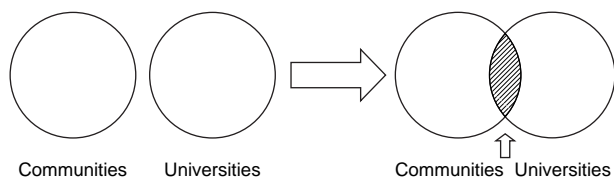


Figure 1 Universities and local communities: from disengagement to engagement.

communities. Since the 1980s, the measurement of the persistence of academic decline in middle and high school, and significant disparities in educational achievement and attainment between historically marginalized populations and majority populations, have contributed to increased attention on engagement, generally specified as learner, student, or classroom engagement in kindergarten through 12 grade education in the United States. However, studies of learner engagement have not been limited to the United States. Studies from Australia, Canada, Asia, and Europe exemplify the challenges of creating effective and civically responsible learning contexts, globally.

In the United States, precursors to today's attention to engaged education and its linkages to social change can be traced to institutional commitments beginning with the founding of American universities, whether dedicated to training ministers or land-grant colleges with public service missions. Since the 1800s, engaged education movements have been led most consistently by institutions with missions committed to social justice and equality, notably Historically Black Colleges and Universities (HBCU), Tribal Colleges and Universities Programs (TCUP), and Hispanic Serving Institutions (HIS). Increasingly, since the 1950s, universities in the private sector are becoming more socially conscious and creating linkages to local, regional, national, and international communities that engage, empower, and educate inside and outside of university settings. Several factors have contributed to substantial changes and revisioning of institutional commitment and strategic plans, including the surge of civil rights and social movements of the 1950s and 1960s. These changes include (1) postmodernist, multivocal, and dialogic approaches to interpretation and representation; (2) a marked demographic shift resulting in increased training of members of communities, from white elites to ethnic white and racial minorities, whose previous roles were limited to serving almost exclusively as the subject of study; and (3) technological advances have afforded global accessibility to information which has heightened awareness and institutional civic commitments and responsibility.

The term engagement has been used to frame discussions of civic, social, and political engagement, often in the context of debates about the role of higher education in society, or about the need for higher education to contribute to the development of university students as

productive citizens who, after graduation, contribute to society in more engaged, active, and productive ways. In this article, engaged education is understood as a process and a product. Engaged education is a process through which the method of instruction fosters conditions, which position and prompt learners to take on more active roles in education as listeners, seers, thinkers, and doers. The concept of engaged education is understood as a product of the interface between experiential education, intensive field experiences, and social change. Community transformation resulting from the process of engaged education is a product that takes many forms. Engaged education refers to the tangible changes communities are able to implement as a result of participating in the process of engaged education. As a product, engaged education includes the increased vision and visibility given to social issues of injustice, inequality, and human struggle. Also, engaged education, as a product, refers to the documents, papers, and video projects produced from engaged education.

Since the 1980s, interest in and commitment to the importance of engagement in the learning process is reflected in the development of major national surveys, journals, and changes in the criteria of university classifications. National efforts include the Center for Information & Research on Civic Learning and Engagement (CIRCLE), the National Center for Public Policy and Higher Education's *Measuring up 2000: A State-By-State Report Card on Higher Education* (2000), and national surveys such as the High School Survey of Student Engagement (HSSSE), the Survey of Entering Student Engagement (SENSE), the National Survey of Student Engagement (NSSE), and the Community College Survey of Student Engagement (CCSSE). These surveys consist of instruments, which measure a broad range of factors associated with engagement in and out of the classroom. Newsletters and journals include the *Community Works Journal*, *The Generator*, *Journal of Higher Education Outreach and Engagement*, *Journal for Civic Commitment*, *Metropolitan Universities*, *Michigan Journal of Community Service Learning*, *Journal of Experiential Education*, *National Service-Learning and Civic Engagement Research E-Newsletter*, *Partnership Matters*, *Partnerships: A Journal of Service-Learning and Civic Engagement*, *PRAGmatics: A Journal of Community Based Research*, and *Service Learning Advances*.

Since the 1990s, universities in major US cities, including Chicago, Detroit, and Philadelphia, have become actively involved with meshing undergraduate learning, outreach, and civic engagement. National and international initiatives for engaged education at the university level include Campus Compact and the Community-Campus Partnerships for Health (CCPH), as well as funding from private foundations, including Pew Charitable Trusts and the Kellogg Foundation, and federal agencies, including the US Department of Housing and Urban Development, the Department of Education, the Department of Justice, and the National Science Foundation. In 2005, the increase

in American universities' attention to community engagement prompted the Carnegie Foundation for the Advancement of Teaching to announce community engagement as a new elective classification available through a voluntary application process. The classification includes three categories: (1) Curricular Engagement, (2) Outreach & Partnerships, and (3) Curricular Engagement and Outreach & Partnerships.

The purpose, role, and responsibility of the university have been a point of rigorous philosophical and epistemological debate at various, even most times throughout the history of institutions of higher education. Further, in industrialized and postindustrialized nations, major social problems have resulted from historical legacies of social inequality, unequal access to rights, privileges and resources, as well as domination and oppression, and these are widely acknowledged. The university, as an institution, has been the target of criticisms about its ideological and social distance from the communities that often serve as the subject of study, and which are often located in close geographic proximity to the university. In some cases, universities and local communities may be in conflict over resources such as territory. More often, criticisms stem from hostile and contentious relationships resulting from histories of exploitation under the guise of research without reciprocal return of knowledge to the communities studied. Within a select group of universities, especially in urban US regions, there has been a concerted philosophical and strategic commitment to social change through a broader construct of the engaged university in local, regional, and global communities. A parallel movement to address issues of social justice and inequality has been initiated through social entrepreneurship efforts, in some cases, resulting in a blending of the two interrelated approaches.

Engaged education, as exemplified through experiential education and specifically illustrated through intensive field experiences, helps to create and foster linkages between colleges and universities and local, regional, national, and international communities. Engaged education has the potential to build extendable, capacity-building knowledge and sustainable relationships that contribute to social change. Optimally, engaged education, as a pedagogical tool, has a direct, sustainable reciprocal exchange. However, not all engaged education has a direct, reciprocal exchange with a specific community. In these cases, the impact of the learning experience can still have such a significant impact that it shapes how students directly and indirectly engage in the world community.

Key Concepts

Experiential education is a philosophy of education, which proposes the application of educator-guided strategies

intended to encourage learners' direct, active experience as a conduit to increased knowledge, awareness, and skill development. Experiential education is often used interchangeably with experiential learning, which focuses on the individual learning process through direct experience. Related examples of experiential education include active learning, outdoor education, theory practice learning (TPL), service learning, cooperative learning, cultural journalism projects, exchange programs, collaborative learning, adventure education, and field trips. Experiential education has been examined as a viable option for improving outcomes across the curriculum in the sciences, social sciences, and professions of law, medicine, business, public health and social welfares; as well as for community development. Experiential education is also related to the concepts of problem-based learning, place-based education, critical pedagogy, and community-based participatory research (CBPR).

Theoretical and conceptual origins of experiential education are often attributed to John Dewey, David A. Kolb's and Roger Fry, and Kurt Hahn. Dewey (1938) proposed a philosophy of education grounded in a Principle of Continuity, experience occurs along a continuum, and a principle of interaction, where the internal and objective aspects of experience inform a situation of learning. Kolb and Fry (1975) posited a four-point learning cycle, with learning beginning at any of its four points: concrete experience, observation and reflection, forming abstract concepts, and testing in new situations. Experiential education theory is also informed by the contributions of Paulo Freire, Jean Piaget, Maria Montessori, and Rudolf Steiner. Central to pedagogical strategies, which advance experiential education, is the importance of developing a sense of building knowledge through layers or stages of experiences culminating with firsthand experience. These strategies contribute to learning experiences that enhance educational outcomes for students, faculty, and communities. For example, course readings and lectures can inform each other as a foundational layer of knowledge. Film screenings, audio recordings, online resources, and class visits from related experts provide another layer of exposure to course material. Course assignments requiring students to research central concepts create learning opportunities for students to begin to synthesize course material and develop expert knowledge on their own. All of these stages or layers of activities are, to a certain degree active, engaged firsthand learning experiences. Firsthand learning experiences in which students encounter and synthesize readings, lectures, films, audio recordings, or web content, still lack the firsthand experience resulting from spending time in places that are the sites of study and meeting people who are the subjects of study, yet whose experiences are generally filtered through more traditional processes and products of scholarship which privileges the researcher's role and voice. **Figure 2** illustrates stages

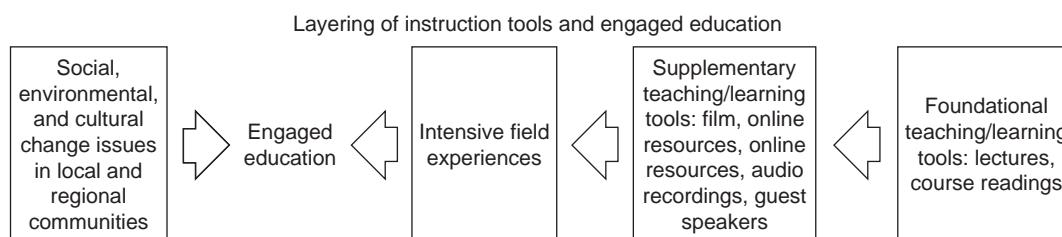


Figure 2 Layering of instructional tools and engaged education.

of experiential knowledge flowing from instruction through intensive field experiences as a process that, when partnered with local communities, results in engaged education.

Intensive Field Experiences

Broadly defined, intensive field experiences are firsthand learning opportunities, which take place in sites outside of the classroom. The term ‘field experience’ is related to the term ‘field trip’. Despite the proven success of the field trip in Kindergarten through corporate settings, field trips are still commonly associated with passive learning at the elementary school level. Contrary to popular belief, field trips have the potential to provide rich, intensive, and transformational experiences along the learning continuum of the life cycle.

Both concepts, field trip and field experience, call for a learning experience in a designated site outside of more traditional classroom instruction. The duration of time, as defined by the length of time and frequency, distinguishes the concepts of field experience from field trip. Field experience is broader than field trip and affords for extended, regularly occurring firsthand experiences with sites and communities. Field trips present an opportunity for field experiences, though at a minimum, less intensive field trips are less likely to result in the relationships that lead to sustainable, capacity-building social change. However, all field experiences are not field trips. Field trips are generally group or class excursions, while field experiences may be undertaken individually with guidance from instructors. However, both field experiences and field trips are examples of experiential education.

In the social and environmental sciences, specifically in disciplines, such as anthropology, sociology, education, environmental studies, and political science, field experiences are often used to introduce students to firsthand knowledge beyond textbooks and lectures. Field experiences may be useful in helping students develop skills as ethnographers, in participant observation, and collecting data through interviews. In turn, these skills can be used to help local communities with critical knowledge and resources. In the social sciences, field experiences have the capacity to illuminate lived experiences of social inequality

and blocked access. There are two defining, yet interrelated dimensions of field experiences: duration and depth of experience. Duration of time is a defining aspect of the field experience. Short-term field trips are a few hours to a full day, in duration. The intensive field experience may be characterized by student participation in an activity that recurs repeatedly over an extended period of time, or an immersion experience over an extended period of time. For example, some undergraduate and graduate courses in anthropology, education, psychology, or sociology may require students to incorporate weekly participation in a community setting into instruction in the context of examining a particular social issue or problem. The immersion field experience may take place over the course of a few days, a week, month, or semester. The types of activities and contexts of interaction in field sites determine the depth of experience. While the goal of any field experience should be to increase student engagement, the intensive field experience is distinguished by in-depth experiences over an extended period of time. Short-term immersion intensive field experiences may last a few days in length.

The intensive field experience is an especially appropriate vehicle for civic engagement and social change because it affords an opportunity for sustained partnership between universities and communities, provides a context for firsthand learning about complex social issues, and is supported by a defined pedagogical strategy. The successful implementation of intensive field experiences directed toward social change efforts requires instruction that provides information in layers or stages associated with experiential education pedagogies. The intensive field experience is a suitable strategy for simultaneously fostering student learning about social change and creating social change because it facilitates learning through firsthand experience in community settings.

Social change and cultural change are endemic to human societies and the passage of time. Rapid social change is especially the case in technologically intensive and highly stratified societies. Conceptual orientations of social change may frame change as resulting from shifts in population, technology, or environment. Other conceptualizations of social change may frame change as resulting from the impact of economic, social, and political forces.

Multiple conceptualizations of social change may serve to stimulate insight and understanding in industrialized and postindustrialized nations. Because societies are constantly changing, movements of social change, whether large or small, are simultaneously influenced by and emergent as history, conflicts, environmental events, and shifts in power and resources. Over time, this history becomes increasingly distant from students who currently fill university classrooms. It is for these reasons that engaged education is an especially effective pedagogical tool for undergraduate and graduate instructions. Engaged education, as illustrated through experiential learning and intensive field experiences, offers several potential advantages.

For over a century, the university has been criticized for its intellectual distance from the struggles and challenges of everyday life. Faculty and the academy have been criticized for exploiting communities for the good of professional advancement, without reciprocal investment in the communities, which serve to generate knowledge. Engaged education is intricately linked to experiential education, intensive field experiences, and social change in four fundamental ways. Firstly, engaged education can help strengthen and empower local communities by actively using underutilized resources, assets, and knowledge for the benefit of social change. Secondly, engaged education helps to break down barriers and form relationships between institutions, who have often been perceived as hoarding knowledge and resources, and local communities. Thirdly, engaged education affords real world, problem-based learning experiences which have the potential to yield sustainable solutions. Lastly, intensive field experiences, as a conduit of engaged education, extend learning beyond the classroom and enable a deeper sense of connection between course readings and lectures and people's everyday lived experiences, which may often be rife with injustice, unequal access, and inadequate resources. Ultimately, engaged education results in a multilevel reciprocal synergy of linkages across domains, between schools and communities.

The Intensive Field Experience

The intensive field experience has two primary aims. First, the intensive field experience calls for a collaborative partnership between universities and communities, which helps to elevate knowledge and awareness of social issues, as well as to foster intellectual and civic engagement through firsthand problem-solving, solution-seeking strategies. Second, the experience extends traditional pedagogical methods, which focus on lectures and discussion of course readings beyond the physical boundaries of the classroom to contexts that support learning outside of the traditional classroom. In best-case scenarios, universities have established comprehensive centralized offices, which actively partner faculty and students with community partners

to support student placement to meet the needs of the community. However, many faculty members incorporate engaged education pedagogical strategies into instruction without the support of formal university partnerships.

There are short- and long-term intensive field experiences. Short-term experiences include one-time experiences and immersion experiences. One-time experiences are characterized by participation in events that may last in duration from a few hours to 1 day. Examples include museum visits, real-estate tours which examine property value disparities as mapped upon the racial and class stratification of place and space, or historic preservation walking tours of neighborhoods, which can be examined through a critical lens of power and narrative to determine whose story, is being told, and whose story is being erased. While these experiences may be short in duration, they have the potential to spark immediate tremendously transformational experiences for students. For example, extending classroom instruction on the economic, social, and political inequality to museum exhibits on the history of lynching in the US, or Confederate currency, can be instrumental in the learning process of aiding the understanding of the political economy, history, and social organization of race in the United States. For many students who have grown up in suburban areas, even the journey of getting to and from sites utilizing public transportation is a transformative experience; and it helps students to better understand the struggles faced by marginalized people whose movement and access to housing, employment, schooling, child care, and health care is restricted by urban transport patterns in metropolitan terrains which often leaves some populations afforded greater access than others.

Short-term intensive field experiences are learning experiences that are part of a specialized, planned event, and journeys that may take students considerable distance from the home university; so much so that either a single trip or regular, recurring travel would be logistically improbable or impossible. Field experiences of this type may take place over the course of 3–7 days. In the United States, the legacy of racial inequality and oppression is integral to understanding American history, public policy and law, as well as persistent racial disparities in education, health, and wealth. Rone (2007) describes an intensive immersion field trip to sites of protest and resistance in the American South, which were pivotal in serving as a catalyst for the US Civil Rights legislation of the 1960s that dismantled legal barriers to voting rights, as well as fair housing, education and health care access, education access. The critical pedagogical strategy of layering information was achieved by course readings, including works by James Baldwin and Aldon Morris, and films, *Bridge to Freedom*, 1965, Episode 6 of the Public Broadcasting Service's (PBS) *Eyes on the Prize* series (1987), and *4 Little Girls* (Lee, 1997). Course lectures, readings, local speakers, and films help prepare students for an intensive weekend

field experience to the Birmingham Civil Rights Institute, Sixteenth Street Baptist Church, Kelly Ingram Park in Birmingham; Montgomery; and Selma for the fortieth anniversary commemorative march of Bloody Sunday, a historic event which led to the 6 August 1965 passage of one of the most significant acts of legislation to result from the US Civil Rights Movement. Similarly, a short-term intensive field experience to the Gullah Sea Island region of the southeastern coastal US was central to courses in linguistic anthropology, African-American studies, and religion. Again, course readings and films prepared students, for intensive field experience, which helped to give students who were enrolled in the course, a firsthand knowledge about the challenges of cultural preservation in the Gullah region, and created relationships with local Gullah leaders' social movement to educate the general public and influence national preservation legislation. Intensive field experiences, both short and long in duration, are appropriate for cultural and historical preservation efforts where students are involved in conducting research, and increasing public awareness and taking potential policy decisions about preservation.

Long-term intensive field experiences recur at regular intervals over the time period of a semester or quarter. Students may work individually or in pairs at different sites. Or, depending on the site, smaller seminar-type classes, with an enrollment of less than 15 students, may work with one community in one site. The latter option presents expanded opportunities for creating a synergy between multiple issues at the same site, as well as for more closer instructor guidance. For classes in institutions where students may face transportation barriers getting to local sites, it may be easier to arrange transportation for groups of students if all students are traveling to the same site at the same time. Long-term intensive field experiences are especially appropriate for experiential learning in local communities which are in close proximity to universities. Ongoing, regularly occurring activities, such as schools and community programs or organizations, are ideal because they afford for relationship and knowledge building over time. Examples of activities include tutoring programs with elementary schools and community centers for literacy development programs, mentoring assistance with after-school and Saturday school youth development programs, or providing support by assisting social studies or science teachers to prepare students for specific research projects. Projects may also involve research and include data collection, analysis, and report writing on topics such as housing regulations for populations displaced by gentrification practices, identification of neighborhood resources to support public school students' academic success, and the membership needs of a neighborhood food coop. Semester or quarter-long intensive field experiences are especially suitable for affecting social change for core social issues such as education, housing, health, and justice.

Curriculum Development Planning and Logistics

The successful implementation of intensive field experiences requires additional time and effort for planning and incorporating stages of learning into the course curriculum, as well as the careful arrangement and orchestration of logistics to include clearances, access, admission, transportation, and, where necessary, lodging and meals. Most importantly, reflection and monitoring must be part of the ongoing process for individuals in university and community roles. In **Table 1**, the tasks associated with intensive field experiences are categorized into curriculum development and logistics for three temporal stages: prefield experience, field experience, and postfield experience. These stages are especially relevant for immersion intensive field experiences which require participants to travel an extensive distance for a period of a few days, in contrast to recurring field experiences in local communities.

Despite the ideological shift in the role of and involvement of universities in communities as partners beyond the walls of the university, for the most part, public policy and faculty still largely define research agendas. Exceptions to this include reverse research days in which community organizations are given an opportunity to present proposals for research ideas to faculty research teams. Other examples include research institutes, such as the Institute for Urban Research at Morgan State University, located in the metropolitan area of Baltimore, Maryland, USA (Rone, 2008). This institute sees its mission in addition to focusing on urban issues, as defined in part by a commitment to servicing the needs of the local community through research and intervention. In this capacity, its research staff members are able to partner with local community organizations and address complex social issues, including environmental justice, education, transportation, housing inequality, and inequalities in the juvenile justice system through integrated research methods, such as interviews, surveys, and geographical information surveys (GIS).

Summary

The university has increasingly shifted its focus on knowledge for the sake of knowledge to knowledge for the sake of the common good. Engaged education through intensive field experiences has the potential to focus on educating students who are critically engaged, productive citizens of the broader community. Engaged education, which affords firsthand experience, allows students to observe and actively work in partnership with local communities to identify and propose solutions to problems associated with inequality and the common good. These collaborations reduce barriers, socialize advocacy and

Table 1 Intensive field experiences: Curriculum development and logistics

<i>Phase</i>	<i>Tasks list</i>
Pre-field experience	<p><i>Curriculum</i></p> <ul style="list-style-type: none"> ● Incorporate related readings, films, and Internet resources into syllabus as student preparation for field experience ● Obtain and prepare for distribution as necessary any instructional materials that originate from site of study ● Develop course assignments which will help students prepare for field experience and develop student expertise in specific areas ● Incorporate relevant guest speakers into course planning ● Have in-class discussion(s) to prepare students for trip (e.g., what to expect, what to wear, local cultural and respect preferences) ● For intensive field experiences requiring long distance travel, have students prepare entries for a compendium about the trip's main focal points <p><i>Logistics</i></p> <ul style="list-style-type: none"> ● Contact site and investigate options for transportation, accommodations, meals, activities, ticket purchases, admission fees ● Create a trip information log (contact person, telephone number, fax number, email, confirmation number, fee due dates) ● Make all arrangements and follow up with vendors ● Create, distribute, and collect trip release forms (and be aware of university requirements regarding parental signatories) ● For local trips with students driving their own cars, get directions to distribute to students ● Prepare a trip information sheet for students ● Administer any baseline or pre-test instruments
Field experience	<p><i>Curriculum</i></p> <ul style="list-style-type: none"> ● For ongoing field experiences, encourage participants to maintain handwritten or electronic field notes ● Make and distribute copies of the compendium to students to read and share ● During the journey of getting to the site and during the field experience, identify relevant points that extend class instruction ● Incorporate time for reflection and dialog <p><i>Logistics</i></p> <ul style="list-style-type: none"> ● Plan for possible delays due to unforeseen circumstances in scheduling departure time for an early arrival ● Bring food and water, directions, camera, first-aid kit, change for parking meters/lots, and copies of all trip confirmations and correspondence ● Make sure all students have a contact cell number and prepare a list of all student cell numbers ● Review local cultural preferences, behaviors, and norms with students ● Leave list of contact names and emergency contact information for all trip participants with departmental office
Post-field experience/ follow-up	<p><i>Curriculum</i></p> <ul style="list-style-type: none"> ● Use an in-class free-write exercise assignment to stimulate and frame discussion about the field experience ● Provide time and space for a reflective, honest, and critical dialog between university and community partners to reflect on the progress, review community and student impact, identify needs, shortfalls, and future directions ● Require students to complete a major post-experience assignment which affords students an opportunity to synthesize course material and firsthand learning experiences from the field experience ● Follow-up on any promised deliverables to participants at field site <p><i>Logistics</i></p> <ul style="list-style-type: none"> ● Ask students to complete a post-field experience evaluation form ● Send notes and/or if appropriate small gifts of appreciation ● Follow-up with transportation, and other vendors ● Administer any post-test instruments

agency, and promote civic participation. Engaged education through experiential education and intense field experiences simultaneously fosters student learning and social change in sustainable capacity-building ways. As global inequality along with globalization increase, the demand for institutions to respond and to produce well-educated, productive graduates will increase. Engaged

education as a pathway to social change, as exemplified through intensive field experience, has many potential benefits, including benefiting local, regional, national and international communities; enhancing student learning of course information and application of abstract concepts to real-world concepts; increasing instructor effectiveness; equipping local communities with capacity-building

knowledge and resources to create sustainable change; and functioning as a critical agent of restorative justice initiatives which transform contentious community relationships into healthy, viable productive relationships.

See also: Action Research in Education; Innovation in Teaching and Curriculum Design; Learning Outside of School; Peer Learning in the Classroom; Social Aspects of Collaborative Learning.

Bibliography

- Dewey, J. (1938). *Experience and Education*. New York: Macmillan.
- Kolb, D. A. and Fry, R. (1975). Toward an applied theory of experiential learning. In Cooper, C. (ed.) *Theories of Group Process*. London: Wiley.
- Lee, S. (1997). *4 Little Girls*. 102 min. HBO Documentary. New York.
- Levine, H. G., Gallimore, R., Weisner, T. S., and Turner, J. L. (1980). Teaching participant-observation research methods: A skills-building approach. *Anthropology & Education Quarterly* 11(1), 38–54.
- Rone, T. (2007). Giving spirit to out parents' stories: (Re)creating memory, intellectual, and social engagement beyond the classroom. *Transforming Anthropology* 15(2), 125–134.
- Rone, T. (2008). Culture from the outside in and the inside out: Experiential education and the continuum of theory, practice and policy. *College Teaching* 56(4), 237–245.
- Scarce, R. (1997). Field trips as short-term experiential education. *Teaching Sociology* 25, 219–226.
- Warren, K. (1998). Educating students for social justice in service learning. *Journal of Experiential Education* 21(3), 134–139.
- Wink, J. (2000). *Critical Pedagogy: Notes from the Real World*, 2nd edn. New York: Addison-Wesley Longman.

Further Reading

- Cantor, J. (1997). *Experiential Learning in Higher Education: Linking Classroom and Community*. Washington, DC: ERIC Clearinghouse on Higher Education.
- Crossley, C. and DeVinney, J. A. (1987). *Eyes on the Prize, Bridge to Freedom, 1965*. Alexandria, VA: PBS Video.
- Freire, P. (1970). *Pedagogy of the Oppressed*. New York: Herder and Herder.
- Freire, P. (1998). Education for critical consciousness. In Araju Freire, A. M. and Macedo, D. (eds.) *The Paulo Freire Reader*, pp 80–110. New York: Continuum.

- Jakubowski, L. M. (2003). Beyond book learning: Cultivating the pedagogy of experience through field trips. *Journal of Experiential Education* 26(1), 24–33.
- Jones-Jackson, P. (1987). *When Roots Die: Endangered Traditions on the Sea Islands*. Athens: University of Georgia Press.
- Kolb, D. A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice-Hall.
- McCurdy, D. W. (1997). The ethnographic approach to teaching cultural anthropology. In Kottak, C. P., White, J. J., Furow, R. H., and Rice, P. C. (eds.) *The Teaching of Anthropology: Problems, Issues and Decisions*, pp 62–69. Mountain View, CA: Mayfield.
- Montgomery, M. (ed.) (1994). *The Crucible of Carolina: Essays in the Development of Gullah Language and Culture*. Athens: University of Georgia Press.
- Nanda, S. (2000). Museum visits in cultural anthropology courses. In Rice, P. C. and McCurdy, D. W. (eds.) *Strategies in Teaching Anthropology*, pp 117–121. Upper Saddle River, NJ: Prentice Hall.
- Pollitzer, W. S. (1999). *The Gullah People and Their African Heritage*. Athens, GA: University of Georgia Press.
- Rodriguez, C. (1996). African American anthropology and the pedagogy of activist community research. *Anthropology & Education Quarterly* 27(3), 414–431.
- Toepke, A. and Serrano, A. (1998). *The Language You Cry In: Story of a Mende Song*. San Francisco, CA: California Newsreel.
- Twining, M. A. and Baird, K. E. (eds.) (1991). *Sea Island Roots: African Presence in the Carolinas and Georgia*. Trenton, NJ: African World Press.
- Williams, B. (1988). *Upscaling Downtown: Stalled Gentrification in Washington, D.C.* New York: Cornell University Press.
- Wright, M. C. (2000). Getting more out of less: The benefits of short-term experiential learning in undergraduate sociology courses. *Teaching Sociology* 28(2), 116–126.

Relevant Websites

- <http://www.ashoka.org> – Ashoka: Innovators for the Public.
- <http://www.communityservicelearning.ca> – Canadian Alliance for Community Service-Learning.
- <http://www.civicyouth.org> – Center for Information & Research on Civic Learning and Engagement (CIRCLE).
- <http://www.servicelearning.org> – Corporation for National & Community Service: Interdisciplinary models of Service-Learning Education.
- <http://www.umich.edu/~mjcs/> – Michigan Journal of Community Service Learning.
- http://www.nsee.org/about_us.htm – National Society of Experiential Education (NSEE).
- <http://www.partnershipsjournal.org/index.php/part> – Partnerships: A Journal of Service-Learning & Civic Engagement.
- <http://www.skoll.org> – Skoll Foundation: Connecting and Celebrating Social Entrepreneurs.
- <http://www.universitynetwork.org> – University Network for Social Entrepreneurship.

Higher Education and Social Change in Mexico

A Mungaray, G Estrella, P Moctezuma, and M Ramírez-Urquidy, Universidad Autónoma de Baja California, Tijuana, Mexico

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Economic development – Economic growth and improvement of social welfare indicators: education, health, housing, income, poverty rate, inequality.

Economic growth – Growth in the per-capita GDP which is the ratio GDP over population.

Entrepreneurship – The concept in the most general way refers to enterprise creation, although it is frequently associated to innovation and ideas that pull the enterprise start ups, where the individual acumen, talent, and human capital play an important role.

Flexible educational practices – Off-class and off-campus learning practices based on learning-by-doing, experiential learning and situated cognition such as service learning, internships projects in enterprises, and community-based programs promoting the students abilities in the workplace.

Gross domestic product (GDP) – Value of the product and services produced in a region over a period.

Innovation – New way of doing something or 'new stuff that is made useful.' It may refer to incremental and emergent or radical and revolutionary changes in thinking, products, processes, or organizations.

Productivity – Ratio output–inputs, such as the partial indexes with respect labor and capital inputs, as well as, the total factor productivity index, which measures the ratio between output and weighted inputs.

Research and Development (R&D) – According to the OECD, it is creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture, and society, and the use of this stock of knowledge to devise new applications.

Social change – Transformation of demographic, economic, and organization patterns.

Traditional educational practices – In-class teaching methods relying on professor lectures and board (chalk-talk), active learning practices in seminar form, and role playing and simulation in control environment, including the use of technologies.

Value added – Difference between value of the product and services produced at market prices and the cost factors and materials for production.

Introduction

A structural change has occurred in Mexico since the middle of the 1980s, following the implementation of pro-market policies such as trade and capital flow liberalization, privatization, and deregulation. In order to manage this new scenario after decades of economic protectionism, the connection between higher education institutions (HEIs) and diverse social and business sectors has intensified during the past two decades, guided by the argument that higher education (HE) and increased knowledge are necessary conditions to succeed in the new economic structure. These linkages between HE and society, particularly enterprise and industry, have traditionally focused on the mutual benefits derived from high-quality research activities that address the needs of the business sector. However, this justification for linking HEIs to industry has neglected the labor-market perspective, which has the potential to be a great opportunity to connect the developing of professional human capital suited to social and business needs.

Linking the professional training processes of students enrolled in HE with community and industrial needs continues to be one of the great challenges facing the Mexican HE system (Mungaray, 1992). Currently, the non-linking is particularly pronounced in segments of the industrial structure such as micro and small enterprises, whose dynamics have also been impacted by the emergence of more competitive markets, given the liberal trade tendencies. The traditional disregard of HEI–community linkages that are based on training students in real contexts and enterprises, combined with the stagnation of certain sectors of the economy, has led to a mismatch between society's structures and needs and HEIs in Mexico. This disconnect, in turn, limits the connection of social- and business-management-related disciplines to economic development, and the enhancement of business education for social change.

A revitalized conception of the linkage between HEIs and society, which promotes the labor-market perspective and entrepreneurship, would produce a new supply of

trained social actors to support Mexican innovation and promote the growth of businesses by encouraging graduates to look for innovative ways to utilize their education within a renovated entrepreneurial environment. This perspective requires a change in the Mexican assumption that students will participate in labor markets as public or private sector employees to a more entrepreneurial lens where we expect students to become proprietors of businesses. A reorganization of Mexican HEIs, which considers the incorporation of complementary, flexible educational practices that coexist with traditional practices, accompanied by a permanent quality evaluation process, would help to build stronger and more profitable linkages between institutions and both community and the industry sectors.

Economy and Labor Market for HEI Graduates

The economic and political development in Mexico since the 1940s has brought a wide transformation in society. The great economic growth during those years, together with welfare improvement, stimulated a social mobility process, cultural progress, and political stability. Those who were able to advance during these years because of education or business opportunities became the main actors producing the constant changes in Mexican society through the rest of the century.

The outstanding economic dynamics during the period was achieved in parallel with some social and economic inequalities. The industrial structure was characterized by unequal growth and level of prosperity in its different branches, and even great disparities among the different types of activity within a branch. This process generated the coexistence of modern and backward productive sectors.

Fast demographic growth and the industrial and urban concentration in the principal cities stimulated the demand for qualified professionals to fulfill the requirements of modern industrial sectors of the economy. In the middle of the 1970s, the student population spread among distinct fields, negatively impacting upon the quality of service of HEIs due to improvised institutional management and unbalanced student concurrency (Moctezuma, 2003).

The idea of industrial development at the beginning of the 1980s included policies to protect national corporations from international competition, major stimulus for technology imports, and an excess of regulations. These policies inhibited the links between academic centers and industrial development, the national spread of the learning ability to encourage private investment in research and development leading to technological innovations inside the enterprises, and the organization of HE as an important support for national entrepreneurship (Pallán, 1995). Consequently, the economy lost the capacity to generate

an entrepreneurial base and professional employment at the pace required by the alumni from HEIs.

The liberal economic and trade policies were aimed at transforming the different sectors of the economy by generating an environment of competition that would stimulate productivity and quality in order to be competitive. The development of Southeast Asian countries influenced this vision by demonstrating the possibility of building competitive, dynamic employment and growth through this path (Hobday, 2003). However, the business dispersion in the different economic sectors in Mexico limited the scope of these general policies of economic promotion, which have failed to provide differentiated incentives to different firms and industrial sectors.

From the perspective of the labor market, the underemployment of professionals and the consistent underestimation of schooling years has created the scenario to put HE, particularly the public system, at the center of a social discussion questioning its pertinence given its functions. Because of the structure of the Mexican HE system, which responded to the 1917 Mexican Constitution's promises of political stabilization, economic development, and social change, policymakers have insisted on the need for HEIs to respond to government development initiatives; provide the business sector with professionals and qualified technicians; and fulfill societal expectations, which consider university degrees indicative of social mobility (Lorey, 1993).

Demographic Structure and Employment Trends

Mexico is one of the most populated countries in the world, although the current demographic dynamics are substantially lower as compared to the 1960s and 1970s. During the decade between 1995 and 2005, the population in absolute values has gone from about 91 million to more than 103 million people, with a falling annual average rate of change of 1.7% and 1.2% in the period 1995–2000 and 2000–2005, respectively (**Table 1**). Nevertheless, the particular dynamics of some regions recall old patterns of demographic concentration in selected cities and poles, their vitality fundamentally associated with a particular economic vocation. Thus, some of the regions with above-average population growth are associated with the expansion of manufacturing and maquiladora industry derived from either their outward orientation or their vicinity to dynamic markets such as populated regions in central Mexico or the northern states bordering the United States. This is the case of some traditional industrial poles such as Nuevo Leon, whose growth rate increased even more than that in the first half of the decade, and the states of Mexico and Puebla, which had declining growth rates that were, nonetheless, above average. More recently,

Table 1 National and regional demographic structure 1995–2005

		1995		2000		AVG 95–00		2005		AVG 00–05		AVG 00–0	
<i>National/states</i>		91 158 290	100%	97 483 412	1.7%	100%		103 263 388	100%	1.2%		1.2%	
1	Aguascalientes	862 720	0.9%	944 285	1.8%	1.0%		1 065 416	1.0%	2.4%		2.1%	
2	Baja California	2 112 140	2.3%	2 487 367	3.3%	2.6%		2 844 469	2.8%	2.7%		3.0%	
3	Baja California Sur	375 494	0.4%	424 041	2.4%	0.4%		512 170	0.5%	3.8%		3.1%	
4	Campeche	642 516	0.7%	690 689	1.4%	0.7%		754 730	0.7%	1.8%		1.6%	
5	Coahuila	2 173 775	2.4%	2 298 070	1.1%	2.4%		2 495 200	2.4%	1.6%		1.4%	
6	Colima	488 028	0.5%	542 627	2.1%	0.6%		567 996	0.6%	0.9%		1.5%	
7	Chiapas	3 584 786	3.9%	3 920 892	1.8%	4.0%		4 293 459	4.2%	1.8%		1.8%	
8	Chihuahua	2 793 537	3.1%	3 052 907	1.8%	3.1%		3 241 444	3.1%	1.2%		1.5%	
9	Distrito Federal	8 489 007	9.3%	8 605 239	0.3%	8.8%		8 720 916	8.4%	0.3%		0.3%	
10	Durango	1 431 748	1.6%	1 448 661	0.2%	1.5%		1 509 117	1.5%	0.8%		0.5%	
11	Guanajuato	4 406 568	4.8%	4 663 032	1.1%	4.8%		4 893 812	4.7%	1.0%		1.0%	
12	Guerrero	2 916 567	3.2%	3 079 649	1.1%	3.2%		3 115 202	3.0%	0.2%		0.7%	
13	Hidalgo	2 112 473	2.3%	2 235 591	1.1%	2.3%		2 345 514	2.3%	1.0%		1.0%	
14	Jalisco	5 991 176	6.6%	6 322 002	1.1%	6.5%		6 752 113	6.5%	1.3%		1.2%	
15	México	11 707 964	12.8%	13 096 686	2.2%	13.4%		14 007 405	13.6%	1.3%		1.8%	
16	Michoacán	3 870 604	4.2%	3 985 667	0.6%	4.1%		3 966 073	3.8%	–0.1%		0.2%	
17	Morelos	1 442 662	1.6%	1 555 296	1.5%	1.6%		1 612 899	1.6%	0.7%		1.1%	
18	Nayarit	896 702	1.0%	920 185	0.5%	0.9%		949 684	0.9%	0.6%		0.6%	
19	Nuevo León	3 550 114	3.9%	3 834 141	1.5%	3.9%		4 199 292	4.1%	1.8%		1.7%	
20	Oaxaca	3 228 895	3.5%	3 438 765	1.3%	3.5%		3 506 821	3.4%	0.4%		0.8%	
21	Puebla	4 624 365	5.1%	5 076 686	1.9%	5.2%		5 383 133	5.2%	1.2%		1.5%	
22	Querétaro	1 250 476	1.4%	1 404 306	2.3%	1.4%		1 598 139	1.5%	2.6%		2.5%	
	Arteaga												
23	Quintana Roo	703 536	0.8%	874 963	4.4%	0.9%		1 135 309	1.1%	5.2%		4.8%	
24	San Luis Potosí;	2 200 763	2.4%	2 299 360	0.9%	2.4%		2 410 414	2.3%	0.9%		0.9%	
25	Sinaloa	2 425 675	2.7%	2 536 844	0.9%	2.6%		2 608 442	2.5%	0.6%		0.7%	
26	Sonora	2 085 536	2.3%	2 216 969	1.2%	2.3%		2 394 861	2.3%	1.5%		1.4%	
27	Tabasco	1 748 769	1.9%	1 891 829	1.6%	1.9%		1 989 969	1.9%	1.0%		1.3%	
28	Tamaulipas	2 527 328	2.8%	2 753 222	1.7%	2.8%		3 024 238	2.9%	1.9%		1.8%	
29	Tlaxcala	883 924	1.0%	962 646	1.7%	1.0%		1 068 207	1.0%	2.1%		1.9%	
30	Veracruz	6 737 324	7.4%	6 908 975	0.5%	7.1%		7 110 214	6.9%	0.6%		0.5%	
31	Yucatán	1 556 622	1.7%	1 658 210	1.3%	1.7%		1 818 948	1.8%	1.9%		1.6%	
32	Zacatecas	1 336 496	1.5%	1 353 610	0.3%	1.4%		1 367 692	1.3%	0.2%		0.2%	

Annual Average Growth (AAG)

Source: INEGI. II Conteo de población y Vivienda 2005.

INEGI. XII Censo General de Población y Vivienda 2000.

INEGI. I Conteo de Población y Vivienda 1995.

emergent industrial areas are included in this group, such as Baja California and Queretaro, which have the highest population expansions, although the former had a declining rate. The counterparts to regions with demographic expansion are those consolidating or emerging as service economies, perippled fundamentally by tourism activities. This is the case of the regions with highest expansion, Quintana Roo and Baja California Sur, and others with lower rates, though higher than average, such as Chiapas and Colima.

The population-expanding regions constitute two-fifth of the Mexican states, which are concentrated areas and are increasing their share in the demography. The other regions are either in a steady state or are declining in terms of their share of total population, including traditionally highly crowded regions such as Mexico City and

the states of Veracruz, Jalisco, Guanajuato, and Michoacán, among others.

The regional demographic dynamics illustrated above is reflected in employment patterns by sector, broadly classified into primary, secondary, and tertiary activities, which mainly include agriculture and mining in the first group; all kinds of manufacturing activities in the second group; and commerce and all types of services in the tertiary group.

If the analysis goes back to the 1970s, the figures then show an economy dedicated mostly to the primary activities, in which about 40% of the labor force was employed (Table 2). The period from 1970 through 1990 experienced a drastic change in the employment structure by sector, fundamentally favoring the tertiary and the secondary sectors, whose rate of increase was 4.9% and 4%,

Table 2 Employment by sector 1970–2005

Sector	1970	%	1990	%	AAG 70–90	2000	%	AAG 90–00	2005	%	AAG 00–05	AAG 70–05
Primary	5 103 519	39.4	5 300 114	22.6	0.2	5 482 720	16.3	0.7	6 047 360	14.9	2.0	0.5
Secondary	2 973 540	23.0	6 503 224	27.8	4	9 088 142	26.9	6.7	10 445 351	25.7	2.8	3.6
Tertiary	4 130 473	31.9	10 796 203	46.1	4.9	18 146 769	53.8	10.4	23 769 471	58.6	5.4	5.0
Unspecified	747 525	5.8	803 872	3.4	0.4	1 012 579	3.0	4.6	313 691	0.8	–23.4	–2.5
Total	12 955 057	100	23 403 413	100	3.0	33 730 210	100	7.31	40 575 873	100	3.7	3.3

Source: Average Annual Growth (AAG) in percentage. From INEGI, Censos Generales de Población y Vivienda, 1970, 1990 and 2000. INEGI, Encuesta Nacional de Ocupación y Empleo for 2005 data.

respectively, far higher than the primary sector, which experienced no growth. With this, the latter sector was transformed from the greatest employer into the smallest one (22.6%) and the shares of total employment of the secondary and tertiary sectors increased from 23% and 32% to 27% and 46%, respectively. These figures are associated with urbanization trends and the rise of populated cities acting as attracting poles even before 1970, as well as the abandonment of rural areas whose population migrated to either larger Mexican cities or the United States.

The period 1990–2000 was characterized by high growth in secondary and tertiary, and low growth in primary employment, with rates reaching up to 6.7%, 10.4%, and 0.7%, respectively. This caused a share increment of the latter as compared to the former, which experienced a slight reduction. The 2005 figures confirm the earlier pattern where primary and secondary sector shares stabilize to 15% and 26%, respectively, whereas the tertiary employment share jumped to almost 58%.

The demographic trends and data on employment by sector complement each other. The particular expansion of tertiary activities, the stability of secondary sector participation, and the decline of primary activities are associated regionally with those areas expanding in population and becoming concentrated. These facts illustrate the profound economic transformations of Mexico resulting from the increasingly outward orientation of the economy and its social impact on the employment structure at all levels. The effects of the North America Free Trade Agreement (NAFTA) since 1994 are reflected by the significant growth in both the secondary and tertiary sectors in the 2000 data, although the latter also expanded its share of employment, given its extraordinary dynamics. The transformation in the Mexican economy is to be expected when regions partake in deep transformation processes. This idea is explained by Fajnzylber's (1983) comments that as the industrial structure is modernized and expands with international dynamics, the employment in primary activities tends to decline and that of industry tends to expand. With growing industrial employment, commercial and service activities accelerate, and the

Table 3 Industrial classification in Mexico

	Manufacturing	Commerce	Services
MEs	1–10	1–10	1–10
Small	11–50	11–30	11–50
Médium	51–250	31–100	51–100
Large	More than 250	More than 100	More than 100

Source: *Diario Oficial de la Federación*, 2002.

employment associated with these sectors is higher and the consumption more dynamic because of the growth of distribution channels.

Industrial Structure, Employment, and Productivity Trends

The industrial sector in Mexico is officially divided into four main groups according to the number of employees and economic activity: micro, small, medium, and large enterprises (**Table 3**). The last census figures indicate that in 2003 there were 2.9 million enterprises in the country operating in manufacturing, commerce, and services, which employed more than 14 million workers (**Table 4**). During the 15-year period between 1988 and 2003, enterprise and job creation have steadily declined. In terms of enterprises, the growth rates fell from 4.46% in 1988–1993 to 1.92% in 1993–1998, and to only 0.57% in the 1998–2003 census periods, whereas in terms of job creation, these figures exhibited a decline in growth rate of 2.92%, 2.20%, and 1.46%, respectively. It seems that the overall dynamics of the economy and the strengthening of the economic opening impacted upon the performance of both indicators during the different periods: the period 1988–1993 had relative economic expansion; 1993–1998 experienced recession, implementation of trade reforms, and the start of NAFTA; and 1998–2003 was characterized by recovery, the consolidation of trade liberalization, and intensified market competition.

Three main trends in enterprise and jobs in this context have become well known. The first refers to the large

Table 4 Industrial structure by enterprise size 1988–2003

	<i>Enterprises (n)</i>		<i>AAG</i>	<i>Workers (L)</i>		<i>AAG</i>	<i>VA</i>	<i>AAG</i>	<i>VA/n</i>	<i>AAG</i>	<i>VA/L</i>	<i>AAG</i>
Total												
1988	1 308 327	100%		6 681 484	100%		1 353 636 928	100%	1 035		203	
1993	2 187 427	100%	4.46%	9 351 330	100%	2.92%	1 728 066 699	100%	2.12%	790	–2.34%	185
1998	2 729 133	100%	1.92%	12 046 601	100%	2.20%	2 006 849 299	100%	1.30%	735	–0.62%	167
2003	2 915 708	100%	0.57%	14 258 563	100%	1.46%	2 560 785 931	100%	2.12%	878	1.54%	180
AAG			2.32%			2.19%			1.85%		–0.47%	
Micro												
1988	1 240 442	94.8%		2 753 241	41%		207 924 212	15%	168		76	
1993	2 095 384	95.8%	4.55%	4 508 317	48%	4.28%	407 523 501	24%	5.84%	194	1.29%	90
1998	2 619 025	96.0%	1.94%	5 547 683	46%	1.80%	484 094 469	24%	1.50%	185	–0.44%	87
2003	2 770 843	95.0%	0.49%	6 285 904	44%	1.09%	375 129 037	15%	–2.21%	135	–2.70%	60
AAG			2.33%			2.39%			1.71%		–0.62%	
Small												
1988	49 527	3.8%		949 393	14%		151 808 584	11%	3 065		160	
1993	68 244	3.1%	2.78%	1 242 997	13%	2.34%	222 930 567	13%	3.34%	3 267	0.55%	179
1998	80 496	3.0%	1.43%	1 483 203	12%	1.53%	244 566 151	12%	0.80%	3 038	–0.63%	165
2003	107 080	3.7%	2.48%	1 725 777	12%	1.32%	254 413 819	10%	0.34%	2 376	–2.14%	147
AAG			2.23%			1.73%			1.49%		–0.74%	
Medium												
1988	14 808	1.1%		1 320 224	20%		306 672 938	23%	20 710		232	
1993	18 836	0.9%	2.09%	1 625 293	17%	1.81%	382 799 238	22%	1.93%	20 323	–0.16%	236
1998	22 800	0.8%	1.66%	1 985 228	16%	1.74%	428 875 342	21%	0.99%	18 810	–0.67%	216
2003	28 180	1.0%	1.84%	2 197 466	15%	0.88%	501 180 756	20%	1.35%	17 785	–0.49%	228
AAG			1.86%			1.48%			1.42%		–0.44%	
Large												
1988	3 550	0.3%		1 658 626	25%		687 231 194	51%	193 586		414	
1993	4 963	0.2%	2.91%	1 974 723	21%	1.52%	714 813 393	41%	0.34%	144 028	–2.57%	362
1998	6 812	0.3%	2.75%	3 030 487	25%	3.72%	849 313 337	42%	1.50%	124 679	–1.25%	280
2003	9 605	0.3%	2.98%	4 049 416	28%	2.52%	1 430 062 319	56%	4.53%	148 887	1.54%	353
AAG			2.88%			2.58%			2.12%		–0.76%	

Average Annual Growth (AAG) in the period

Enterprises by size according to number of employees: Micro, Small, Medium, and Large.

Value added (VA) in thousands of pesos in 2003, number of workers (n), number of enterprises (L).

Source: INEGI, Censos Economicos 1989, 1994, 1999, 2004.

participation and growth in total enterprises and jobs in the group of micro-enterprises. These enterprises have a stable share of total enterprises, fluctuating around 95%, and an increasing share of employment, from 41% in 1998 and 44% in 2003; both indicators growing at 2.33% and 2.39% on an annual average, respectively, overtaken only large enterprises. The second trend is the extraordinary performance of enterprise start-ups and job creation by large enterprises, where both variables grew 2.88% and 2.58%, respectively, during 1998–2003, well above the overall figures. Although the share in total enterprises remained stable during the period analyzed (0.3%), the jobs provided by this segment of large firms increased substantially from 25% and 21% in 1988 and 1993, respectively, to 28% in 2003, transforming this group of enterprises, which are mainly foreign, into the second largest job provider. The third trend is the relatively poor dynamics of small and medium enterprises in terms of firms and employment, whose rate of growth are below average (**Table 4**). After reducing its share of total businesses from 3.8% to 3.0% in the decade from 1988 to 1998, and recovering to 3.7% in 2003, the small enterprise segment nonetheless reduced its share of jobs from 14% in 1998 to 12% in 2003. During this period, the medium enterprise segment remained invariant in terms of participation in the overall industrial structure with a figure around 1%, but suffered a drastic decline in job creation, which reduced participation from 20% in 1998 to 15%. Hence, the segment of small and medium enterprises, which are primarily domestic, have seen their importance as job providers fall from 34% in 1998 to 27% in 2003.

The performance in terms of value-added generation has favored the large enterprise segment since it has increased unevenly during the period at 2.12% on average, increasing its share in total value added from 51% in 1988 to 56% in 2003. The reduction in the participation by other groups is variable; the micro-enterprise group exhibits the most pronounced changes from 15% in 1998 and 24% in 1993 and 1998, to 15% in 2003. The small and medium enterprises exhibited less drastic fluctuations, yet their share reduced to 10% (from 11% in 1998) and 20% (down from 23% in 1998), respectively, by 2003. With this, the productivity performance as measured by the value added per enterprise and per worker has reduced, on average, to 0.47% and 0.35%, respectively, for all enterprise sizes during the period of analysis (1988–2003; **Table 4**). It is noted that the medium and large groups of businesses, large enterprises in particular, reverted to this negative trend in productivity indicators by experiencing increments during 1998–2003, which may suggest a recovery from the 1994 crisis and adaptation to more competitive markets after NAFTA. The micro and small enterprises appear to have struggled in terms of value-added generation and productivity under these circumstances, since they exhibit either great negative changes

(micro-enterprises), or practically no growth (small enterprises), and the highest productivity losses.

The classification of industry by enterprise size during 1988–2003 illustrates the great disparities in terms of dynamics and potential among the enterprises. There is a huge concentration of income in medium and large enterprises, where no more than 1.4% of enterprises generate up to three-quarters of the value added and a disproportionate share of entrepreneurial activity, since millions of smaller enterprise owners and self-employed, who comprise more than 98% of the total enterprises, generate the remaining share. Hence, the data exhibit the importance of designing policies to support smaller enterprises in overcoming their constraints, and which consider the importance of linking small enterprises with larger enterprises as suppliers in building productive chains to optimize their great potential (Humphrey and Schmitz, 2002), and help large enterprises to take advantage of their flexibility.

In this context, a realistic educational policy would contribute by generating research and encouraged learning about smaller enterprises through suitable academic programs designed to support their competitive ability, as well as promote entrepreneurial training (Mungaray, 2002). This requires classification of businesses in government policies and promotion activities organized by development agencies. As Sylos (1993) has shown, industrial reorganization programs have limitations that are different from those of enterprises: while large enterprises encounter more challenges because of their size, small enterprise constraints are associated with the availability of qualified workers.

The Challenge of HEIs and Community Linkages in the Twenty-First Century

Linking HEIs and society has been considered a strategy for development. Consequently, education, science, and technology policies must collaborate closely with all the economic sectors: industrial, commercial, and service (Mungaray, 2002). It is apparent that the current educational and industrial policies work without coordination, on the one hand, and lack association with priority problems related to development, on the other. In the first case, the efforts of public HEIs seem to be disassociated from those of the government, because the organizations created to promote those linkages have been distant from scientific and technological development centers. In the second case, the concept of linkage has been mainly constructed on the premise of supporting large firms and their requirements to compete in international markets. Hence, the linkages between HEIs and government agencies, despite having different purposes, are oriented to serve the needs of larger

enterprises because of their great contribution to value added productivity and higher-paying jobs.

The scientific personnel in Mexico are considered excellent; they constitute a small community. Policies about researchers have been oriented to basically find differential mechanisms to solve their remuneration problem, to restore scientific equipment in order to avoid the brain drain from established groups, and increase the student population in scientific areas (Sarukhán and De la Fuente, 1993), and, more recently, to foster regional development activities (Rubio, 2006). The importance and capacity of these important research groups associated with the few large companies have fundamentally oriented the linkage practices to university research institutes. For instance, since the middle of the 1990s, in two of the largest Mexican public universities about 27% of the agreement or contracted linkages to enterprise have been research activities where the principal partners are large corporations such as Condumex, Resistol, Negromex, Celanese Mexicana, Bimbo, Bacardi, or Syntex, among others (Casas and Luna, 1994). The importance of research and development activities derived from modern science by large national and international corporations has also fostered private education institutions and public research centers to link research activities with large Mexican firms since the increased trade liberalization. These illustrate how linkages have been constructed among the government, large firms, and the most important HEIs of the country. In all the cases, it is obvious that the most important initiatives came from educational authorities, the Ministry of Public Education (SEP), and the National Council on Science and Technology (CONACYT), which supported many research projects that constituted an essential base for large companies (Foro Científico y Tecnológico, 2007; Conacyt-Foro Científico & Tecnológico, 2008; OECD, 2008).

During the 1990s, the university–industry linkages were beneficial because of the agreements and contracts to directly perform high-quality research, including more regional participation, similar to the practices of other countries, where important educational and industrial leaders have achieved a convergence of interests and are now working together (Shutter and Van Alsté, 1998). It is always the case that resources are scarce and the needs are numerous, so priority allocation supported the linkage between highly prestigious research groups and large enterprises.

Discussions of the need to reconceptualize academic linkages in Mexico go back to the mid-1980s when the International Conference of Higher Education Consultants of 1986 was dedicated to evaluating the mechanisms by which HEIs support, or could support, economic development in their regional or local dimensions. The conclusions emphasized the need for HEIs to develop flexible mechanisms that would allow them to better achieve

an academic environment, and most importantly, orient research and professional training to the technological support of small enterprises (Watson and McGinn, 1987). This emphasizes of the importance of education to industrial policy, because supporting local productive systems induces industrial activity to spread in the locality. Another academic event promoting changes in Mexican academic linkages was the Perspective 2000 Forum attended by Mexican and French researchers, where the latter group emphasized the industrial structure of Mexico, showing the predominance of micro and small enterprises with traditional processes, and on the contrary, the practice of HEIs in the country having linkages with large firms. The French experience in alternative education practices, promoting the involvement of students in real contexts, where the enterprises constitute a great attractions, was relevant to the Mexican context because it allows students to become familiar with labor environments even while studying (Rubio, 1994) and because it is possible to expect students to influence their community through the socially relevant work they do (Mungaray *et al.*, 2007).

It seems clear from the perspective of industrial development of a country, which has an important though small modernised sector, that the strategy of linkages should not only be multipolar and relevant to support modernization and the competitiveness of medium and large enterprises, but also be oriented toward the entrepreneurial learning of micro and small enterprises, which offer great potential for industrial and social development because of their contribution in the industrial structure and employment, and flexibility when linked with larger enterprises. Thus, coordination between education and industrial policies was important when NAFTA started and remains one of the main challenges in reinforcing linkages between university–business–society, not only through research associated with the larger enterprises, but also by defining institutionally a linkage pattern between university and the smaller enterprises through technical assistance, productive collaboration, and the promotion of networks with larger enterprises, as practiced in developed and rapidly growing countries.

The idea of reevaluating HE to transform it to be more practical and relevant for students and pertinent for society is presented in the conclusions of the Round Table about Education and Work in the Second Educational Research International Conference. This implies that besides research and continued education, students' education through internships in small, medium, and large enterprises is also very important for better training. In fact, it is acquiring priority, given the dominant trends of the industrial structure discussed above, and the need of the productive sectors to participate responsibly in the academic sectors in the instruction of future professional employees and businessmen (Estrella and Ponce, 2006). This requires

a reevaluation of the traditional concept of excellence regarding labor criteria and students' participation, as well as the reconsideration of the relevance of practice and social context in learning, which will infuse a social sense into the professional formation (Wenger, 1998).

Thinking the Future

At the beginning of this century, educational institutions confronted the challenge of incorporating research as one of their substantive functions; currently, they confront the challenge of linking research and academic activities with economic and social development as a legitimating factor of their existence. Success in confronting this challenge can be achieved by implementing regular evaluation of study plans and programs, faculty training, and the orientation of basic and applied research toward the country's development needs.

Because academic and administrative organization of disciplines and subjects in Mexico is extremely rigid, the pressures imposed by economic and social change are motivating an intense updating of study plans and programs toward content and study methods (Mungaray and Cuamea, 1992). This evidences a social tendency to get away from traditional academic reforms and instead consider the way new actors in the shaping of professionals learn and participate. Consequently, academic reforms suitable to the requirements of the productive sector have been important, including joint guardianship stays, professional practice through learning by research, and reforms oriented to mitigate the over-specialization of some disciplines. These elements evidence intentions toward an education linked with the needs of local and regional productive sectors (Mungaray and Moctezuma, 1997; Corona, 2005).

A new academic structure is slowly emerging, organized around the development of different professionals, sustained by academic flexibility, and the institutional mobility referred to in international parameters. This is a consequence of accreditation trends observed at an international level and the linkage of the educational process with the regional productive needs, illustrating that organizational innovation in education processes can be achieved from any academic-administrative organization model (Rubio, 2006). This flexible and socially linked academic organization also requires a basic formative level and a level of co-responsible internships or training in enterprises. When considering these levels, disciplinary and professional training help to achieve quality, coverage, and expansion of the services, which help to strengthen linkages between university and jobs. Collaboration is essential in the challenging issue of orienting HEIs to serve society, and must be part of a national and regional agreement concerted with government,

enterprises, and community, inspired by the fundamental idea of linking students to the social context, the prevailing industrial structure and to entrepreneurship, to transform them into real agents of change and development.

Bibliography

- Casas, R. and Luna, M. (1994). Condicionantes políticos de la nueva relación entre universidad e industria. In Campos, M. A. and Corona, L. (eds.) *Universidad y vinculación nuevos retos y viejos problemas*, pp 1–17. México: IIMAS-UNAM.
- Conacyt-Foro Científico & Tecnológico (2008). *Promoviendo la innovación y el desarrollo tecnológico. Programa de estímulo fiscal*. México.
- Corona, L. (2005). *México: el reto de crear ambientes regionales de innovación*. México: CIDE-FCE.
- Estrella, G. and Ponce, M. T. (2006). *Impacto laboral de egresados universitarios y opinión de empleadores*. México DF: M. A. Porrúa-UABC.
- Fajnzylber, F. (1983). *La industrialización trunca de América Latina*. México: Nueva Imagen.
- Foro Científico y Tecnológico (2007). *La tecnología mexicana al servicio de la industria*. México.
- Hobday, M. (2003). Innovation in Asian industrialization: A Gerschenkronian perspective. *Oxford Development Studies* **31**(3), 293–314.
- Humphrey, J. and Schmitz, H. (2002). How does insertion in global value chains affect upgrading in industrial clusters? *Regional Studies* **9**, 1017–1027.
- Lorey, D. (1993). *The University System and Economic Development in Mexico since 1929*. Palo Alto, CA: Stanford University Press.
- Moctezuma, P. (2003). *La Administración federalista del sistema de educación superior en México*. México: UABC – Plaza y Valdés.
- Mungaray, A. (1992). Requirement for skill, training and retraining in the Mexican food and drink industry: The perspective for the nineties. *Journal of International Food and Agribusiness Marketing* **4**(1), 71–93.
- Mungaray, A. (2002). Re-engineering Mexican Higher Education toward economic development and quality. The XXI century challenge. *Higher Education Policy* **15**(4), 391–399.
- Mungaray, A. and Cuamea, F. (1992). The Mexican–United States free trade agreement: Challenges and perspectives for Mexican higher education. In Ganster, P. and Valenciano, E. O. (eds.) *The Mexican–U.S. Border Region and the Free Trade Agreement*, pp 91–97. San Diego, CA: San Diego State University Press.
- Mungaray, A. and Moctezuma, P. (1997). Vinculación entre unidades productivas e instituciones de educación superior. *Revista Mexicana de Sociología* **LIX**(3), 109–130.
- Mungaray, A., Ramírez-Urquidí, M., Taxis, M., Ledezma, D., and Ramírez-Angulo, N. (2007). Promoting learning in small entrepreneurs and higher education students through service learning programs. *International Journal of Business Research* **VII**(3), 10–28.
- OECD (2008). Reviews of innovation policy: Mexico. *Overall Assessment and Recommendations*: Paris.
- Pallán, C. (1995). Crecimiento, evaluación e innovaciones en la educación superior (1988–1994). *Revista de Administración Pública* **89**, 1–35.
- Rubio, M. (1994). Conclusiones de la Mesa Educación y Trabajo. In Miklos, T. (ed.) *Memoria del Foro Prospectiva 2000. México y Francia ante los retos educativos del nuevo milenio*, pp 250–253. México.
- Rubio, J. (2006). *La política educativa y la educación superior en México 1995–2006. Un balance*. México: FCE.
- Sarukhán, J. and De la Fuente, J. R. (1993). Science in developing countries: The case of Mexico. In Boldú, J. L. and De la Fuente, J. R.

- (eds.) *Science Policy in Developing Countries: The Case of Mexico*, pp 11–16. México: UNAM-FCE.
- Shutter, F. and van Alsté, J. A. (1998). The regional impact of the University of Twente. In Van Alsté, J. A. and Van Der Sijde, P. C. (eds.) *The Role of the University in Regional Development*, pp 7–17. Enschede: University of Twenty Press.
- Sylos, P. (1993). *Nuevas tecnologías y desempleo*. México: FCE.
- Watson, J. A. and McGinn, M. J. (1987). Higher education in support of regional economic and industrial development. *International Journal of Institutional Management in Higher Education* **11**(3), 342–348.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. Cambridge: Cambridge University Press.

The Changing Role of the State in Higher Education

K Gürüz, Retired Professor, Middle East Technical University, Ankara, Turkey

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The university is one of the oldest institutions in the world. It precedes the nation-state, but the creation of the modern university and national higher education systems are generally attributed to the nation-state. Thus, rather than speaking of the changing role of the state in higher education over time, one must speak of the changing role of a higher external authority as well as of what that authority was in various periods of history.

Furthermore, the role of the external higher authority must be described relative to the other major groups of actors/stakeholders that are involved in higher education governance. The latter were identified by Burton Clark in the early 1980s as the state bureaucracy, the academic oligarchy, and the market. For the purpose of this work, the term higher external authority will be used instead of the term state bureaucracy. In what follows, the term market is expanded to include the term society so that all other stakeholders are included in the analysis. With these caveats, **Figure 1** shows a modified version the triangle of coordination proposed by Burton Clark to compare national higher education systems, which are used below as an analytical framework to describe the changing roles of the major groups of actors over time.

The Medieval University

Bologna University, recognized as the first university in the world, traces its origins to 1088, when the jurist Irnerius started teaching Roman law in Bologna for a fee. One can thus argue that initially the university was a demand-driven institution structured by market forces.

As this new institution spread throughout western Europe, the need was felt for the shelter of a higher external authority for several reasons: protection against the outside (townsmen, local authorities, etc.), a source of direct and indirect funding, and a source of legitimacy. For example, the *authentica habita* of the Emperor Frederick I in 1158 provided imperial protection to students and teachers, exempted them from local jurisdiction, and guaranteed their free movement in Europe.

The first seven universities, Bologna, Paris (f. 1160), Oxford (f. 1167), Cambridge (f. 1209), Montpellier (f. 1220), Padua (f. 1222), and Orleans (f. 1235) had come into being spontaneously. These universities, in time, acquired a prestige such that they could grant teaching licenses that

were respected throughout Christendom. Toulouse was the first university to be founded by a papal charter in 1229. Medieval Spanish universities, on the other hand, were founded by royal charters; they derived their authority from the king, rather than the pope. The papal bull of 1233, which stipulated that anyone admitted to be a teacher in Toulouse had the right to teach everywhere without further examinations (*ius ubique docendi*), in time, transformed this privilege into the single most important defining characteristic of the university and made it the symbol of its institutional autonomy. Authorizing universities to grant teaching licenses that would be valid throughout Christendom without further examinations subsequently became the basis for founding charters by popes, emperors, and kings. By the year 1292, even the two oldest universities, Bologna and Paris, had felt the need to seek similar bulls from Pope Nicholas IV. Many of the later medieval universities were founded by papal charters that gave them the title of *studium generale*, and authorized them to award degrees.

The medieval university thus evolved out of the interaction between the Catholic Church, the external higher authority, and the guild(s) of teachers and/or students, that is, the academia, or the academic oligarchy in **Figure 1**. Philip Altbach has defined the medieval university as an international institution, supervised by the international authority of the Catholic Church that used a common language, Latin, to provide training to students from many countries by internationally recruited teachers who used an international knowledge base comprising books translated from Arabic and Greek.

Students from different regions were organized under nations, which then only had a geographic connotation indicating birthplaces. Nations played significant roles in the administration of many of the medieval universities, most notably in Paris and Bologna. As it is rather difficult to distinguish between teachers and students in a typical medieval university, both groups are treated under academic oligarchy in the discussion ahead.

The Catholic Church was gradually replaced as the external higher authority by the temporal authorities as represented by the emperor, king, prince, or local government. Throughout the Middle Ages, the external higher authority, be it spiritual or temporal, acted as a source of legitimacy, standard-setter, and a protector. The various charters, bulls, and edicts it issued served as a licensing, accreditation, and quality assurance mechanism. Its role as a provider of financial resources became more important

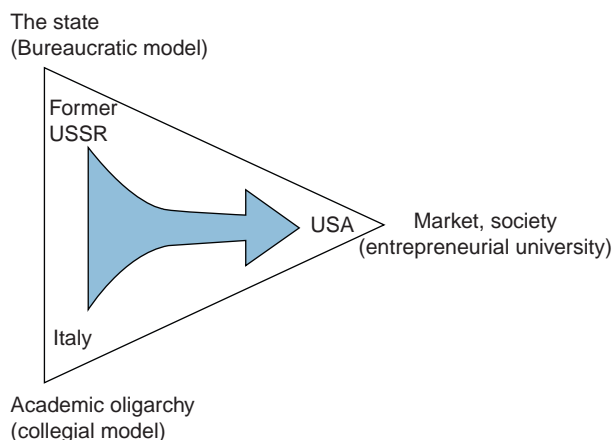


Figure 1 The triangle of coordination. The rise of market forces: lay governance and entrepreneurial university. The models indicated in the parentheses on the apexes are the author's interpretation. The original form of the triangle of coordination appeared in Clark (1983, p. 143). Its modified forms have been used by Kogan and Marton (2000) and Gürüz (2008).

as the influence of the Church waned, and the temporal authority became predominant.

The university emerged as an autonomous body corporate with the power to award degrees. By the year 1500, it had diffused to all parts of Europe, except Russia and the Balkans.

Seeking privileges from and protection by a higher authority came at a price. While all of these papal and royal charters addressed issues such as meals and lodging for students and teachers, libraries, and rights and responsibilities of teachers and students regarding matters of discipline, they also included requirements by the higher authority that pertained to academic matters, such as admission and graduation requirements, length of studies, and in some cases, even syllabi, course contents, and books that were allowed. Such charters protected the university as an institution against interference by local administrations and ecclesiastical authorities, and laid the foundations of university governance and autonomy.

From the governance point of view, the medieval university was configured on the academic oligarchy—higher external authority axis, with small roles for market and society. The rector, who was elected by peers from their own ranks, was *primus inter pares* with ceremonial, rather than executive, powers at the head of the institution. Exceptions were some universities where bodies made up of interested citizens had been put in charge by local governments in Italy and princes in Germany.

1500–1800: The Effects of the Protestant Reformation

The Protestant Reformation had three lasting effects on higher education governance:

- Universities in Protestant Europe started teaching in vernacular languages. They were soon linked with the emerging nation-states, eventually becoming absorbed within state bureaucracies. This led to the emergence of the continental European model of university governance.
- Both the Protestants and the Catholics saw opportunities in education to spread their faith. The Protestants, who viewed the university as a Catholic institution, founded new types of educational institutions of their own that became the cradles of many of today's universities and nonuniversity institutions in central and northern Europe. To counter the Reformation, the Catholics, led by the Jesuits, Franciscans, and Dominicans started opening new universities not only in Europe, but also throughout the world. The first university outside of Europe, Santo Domingo (f. 1538) in today's Dominican Republic, and the first university along the Asia-Pacific Rim, Santo Tomas (f. 1611) in the Philippines were founded by the Dominicans. Spanish colonizers, supported by the Catholic Church and the Spanish King founded universities in many parts of Latin America, where clergy and the administrators of colonial administrations were trained.
- The Calvinist Academy in Geneva (f. 1559) had a lay governing board in accordance with the Calvinist creed. This served as the prototype first for the governance of the Protestant universities in Holland and Scotland and Trinity College in Dublin (f. 1592), and later, for the American colleges of the colonial era. The dissenting academies, though short-lived, with lay governing boards and much closer ties to the community, rather than Oxford or Cambridge, influenced the structures and the governance systems of both the civic universities in England and American universities. From these historical roots, the Anglo-Saxon model emerged as a different approach to university governance, which was predicated upon the presence of lay members in the highest governing bodies of institutions.

By the end of the eighteenth century, the central state was emerging as the higher external authority, and the Church was no longer a significant actor in governance. In Russia, too, the University of St. Petersburg, which was part of the Akademia Nauk (f. 1725), and Moscow University (f. 1755) were creations of the czar.

The role of the state evolved differently in the two models of governance. In the continental European model, the state took on a regulatory role that involved issuing rules and regulations pertaining to nearly every aspect of university activities, including the appointment of professors, academic and administrative structures, curricula, graduation requirements, salaries of staff, and matters related to expenditures and budgets. On the other hand, the role of the state in the Anglo-Saxon

system evolved as steering from a distance through intermediary bodies and governing boards dominated by lay members rather than being regulatory.

1800–1950: The Emergence of the Modern University

The Université de France (f. 1806–1808), sometimes called the Napoleonic University, and Berlin University (f. 1810) were both placed directly under ministries of education. With the office of the Grand Maître at the head of the university, Université de France was, in fact, a national system of education rather than a university. Its basic features, such as the division of the country into educational zones called academies, with the rector of each *académie* acting as the chancellor of all educational institutions, including universities in that zone, exist even today in France.

Berlin University was established to revive academic life in Germany in the aftermath of the devastation inflicted by Napoleon. Initial preparations for its foundation had been carried out by von Humboldt. When it was finally established, it incorporated few of Humboldt's ideas. The kaiser appointed full professors, and the minister made the appointments at the associate professor level. Nevertheless, the principle of the unity of teaching and research advocated by von Humboldt had a lasting effect that shaped the modern university.

The ministry of education was established in Prussia in 1817. Similar ministries were founded all over continental Europe, and these soon took over both the financing and the administration of higher education. The process of secularization of higher education was essentially completed in the nineteenth century. Even church-affiliated private universities increasingly relied on the state for their finances, with only vestigial influence of religion left in their governance and curricula. In the Netherlands, for example, the law enacted in 1876 provided for public financing for all institutions, including private ones. These were the beginnings of today's government-dependent private universities.

Higher education in continental Europe came to be viewed as a purely public service to be financed exclusively from the public purse. In return, the state assumed full control of the administration of institutions by appointing a civil servant as the head of administration, such as the *kanzler* in Germany, who was not subordinate to the rector.

In this period, teachers were transformed from being members of the universitas, the guild, a collegial community of scholars, into civil servants. Higher education institutions became the main sources of manpower for the emerging bureaucracies of the nation-states. The examinations used in selecting public servants, such as the *staatsexamen* in Germany, affected curricula and assigned

to the state the role of a standard-setter. The period 1860–1930 especially saw the transformation, professionalization, expansion, and diversification of higher education itself, as also its role in nation-building and an agent for social change and mobility.

The continental European model of governance thus took shape along the state bureaucracy–academic oligarchy axis of the triangle of coordination. The state played two major roles as the major provider of funding and the regulator of processes. University autonomy naturally depended on a delicate balance between the state bureaucracy and the academia.

The evolution of the Anglo-Saxon system, especially its American variant, from the nineteenth century on, introduced the society and the market as a third group of actors in governance in addition to the state and academia. Lay members represented the former in institutional governing boards and intermediary bodies. This meant that autonomy was defined across the whole of the triangle of coordination rather than just along one of its sides.

The state at both the central and the local levels remained as a major source of funding throughout the Anglo-Saxon world, too.

The role of the state, both in the provision of part of the funds and in setting standards, however, was exercised indirectly through intermediary bodies, institutional governing boards, and external organizations. This has continued to date as a characteristic feature of university governance in the Anglo-Saxon countries.

By the second half of the nineteenth century, Germany had emerged as the academic center of the world. It was in Germany that research became an integral function of universities, and the academic profession took shape through the introduction of the research-based doctoral degree and procedures and requirements for progressing through academic ranks. It was also in Germany, where the various vocational and technical institutions of higher education that had been founded earlier were elevated to an equal status with universities, such as the *technische hochschulen* that hosted so many great scientists and served as a model for many countries.

The German research university had a profound influence on the modern American university, both in academic structure and governance. Between 1815 and 1914, there was an extraordinary migration of about 10 000 American students to Germany. As students returned from Germany and took up positions in universities, the idea of the unity of teaching and research and the German research university model permeated American higher education, effectively diminishing British influences.

In 1862, the Morrill Land Grant Act had been passed, which provided public land and federal funds for educational institutions that offered programs in liberal, agricultural, and mechanical arts. The Hatch Act of 1887 provided funds to the experimental research stations in

these institutions and set forth requirements for granting them university status. In this manner, the public state universities were born. The role of the state as a provider of funds was affirmed in American higher education, too. From then on, American higher education evolved as a market-responsive, diverse mixture of public and private institutions.

By the beginning of the twentieth century, American universities had transferred and adapted the German research university model to build the largest higher education system in the world. The German influence did not end there. The first list of accredited institutions was prepared by the American Association of Universities (AAU) in 1914, when Berlin University made it clear that it would accept students to its doctoral programs only from those institutions that were recognized as universities by the AAU. Regional accreditation boards were established by universities and colleges and their number increased rapidly. This was the beginning of institutional accreditation in the American system. State legislatures had earlier recognized the awarding of the license to practice a particular profession as the prerogative of the professional associations. This eventually led to discipline-specific, or specialized accreditation. The G I Bill of Rights of 1944 stipulated that to be eligible for financial support under that federal law, veterans had to enroll in accredited institutions of higher education. This provided another instrument for the state to steer higher education in the United States from a distance. Later federal legislation authorized the US Department of Education to supervise regional accreditation bodies, and state higher education boards to approve new programs in public institutions. Thus, another role of the state in the American system is meta-accreditation, that is, to recognize accreditors at the federal level, and a form of licensing at the state level. At the federal level, the state shares this role with the Council for Higher Education Accreditation, a private, nongovernmental national coordinating body for national, regional, and specialized accreditation.

Except for Oxford and Cambridge, the British variant of the Anglo-Saxon system differs only slightly from the American version in terms of the role of state in governance. In both cases, university heads are appointed and have strong executive powers. While governing boards of American universities in general comprise only lay members, those in British universities have internal as well as external members, with the latter category always constituting the majority. Board members are in general appointed by state governors in American public universities. Those in British universities are stipulated in their founding charters. Compared to the tightly regulated continental European institutions, British universities are considered as government-dependent private institutions.

Such charters authorize British universities to award degrees autonomously. The need to have a quality assurance

system for nonuniversity institutions was recognized early in the nineteenth century. The model for this was provided by Oxford and Cambridge universities, which gave examinations and awarded degrees, while instruction was carried out in their constituent colleges. Durham (f. 1832) and London (f. 1836) universities were founded as federal universities to function in a similar manner for the various institutions in their regions. Furthermore, professional bodies, rather than the state, have traditionally played significant roles in shaping curricula and degree requirements in regulated professions in a manner similar to specialized accreditation in the United States.

The period 1800–1950 was also a period when universities, nonuniversity higher education institutions, and the various institutional models spread to eastern and southeastern Europe, the Middle East, Africa, the Far East, Latin America, and the Oceania. The universities founded in eastern and southeastern Europe paralleled the foundation of the new nation-states in the region, and all were typically continental European in structure. Elsewhere, universities founded by colonial administrations and missionary groups in Australia, New Zealand, India, Hong Kong, Thailand, Vietnam, China, Singapore, Indonesia, South Africa, Latin America, North Africa, and the Middle East were modeled after institutions in the home country.

A point pertinent to the role of the state in higher education is that outside of Europe and North America, universities and other institutions of higher education were introduced by foreigners. Japan and Turkey were the two exceptions that willingly transplanted such uniquely Western institutions into their societies. The modern Japanese higher education system came into being with the foundation of the Tokyo University in 1887 as part of the Meiji reforms, and developed under both continental European and American influences. A university-like institution took more than half a century before it was firmly founded in 1901, which was replaced by İstanbul University in 1933 by the Republic of Turkey.

1950–1980: Massification of Higher Education

For centuries, access to higher education was a privilege for certain societal groups: first the clergy, then the children of the aristocracy, landed gentry, urban merchants, and industrialists. Enrolment ratios were barely above 1% by the turn of the twentieth century even in Europe and North America. Owing arguably to its particular features, higher education in the United States was massified – gross enrollment ratio above 15% in Martin Trow's classification – by the 1930s. This was achieved in Western Europe about 40 years later.

Following World War II, and especially from the 1970s on, demand for higher education took off worldwide as most jobs in the emerging global knowledge economy required tertiary-level qualifications. Global higher education enrolment, which was about 6 million in 1950, increased to 28 million in 1970, 68 million in 1990, and 88 million in 1997. The increase in the enrolment in developing countries was even more dramatic: from about 2 million in 1950, to 7 million in 1970, and 43 million in 1997, which accounted for more than half the global total in that year. According to 2006 the United Nations Educational, Scientific and Cultural Organization (UNESCO) statistics, the historic threshold of 100 million was passed in 2002. Global enrolment in 2004 was 132 million with 8.5 million academic staff, and the average gross enrolment ratio was 24%. This means that higher education in the world, taken as a whole, is now massified. China, with an enrolment of over 20 million students has recently taken over the United States as the largest national higher education system in the world. China and India currently account for about 25% of the global enrolment. Increasing enrolments in higher education is now an established global trend, and global enrolment is projected to double by 2025.

Peter Scott estimates that of the 1854 universities founded between 1200 and 1985, three-quarters were established since 1900, and 1101 were founded between 1950 and 1985. New types of tertiary-level institutions were founded in order to meet the demand from students with increasingly diverse backgrounds, motives, and career prospects in a cost-effective manner. These were generally more vocationally oriented and of shorter duration, such as the community colleges in the United States, the polytechnics in the United Kingdom, the *Fachhochschulen* in Germany, and the *Instituts Universitaires des Technologie* (IUTs) in France. Distance education institutions were founded in many countries, which were modeled after the British Open University (f. 1969). Thus, national higher education systems came into being, in which different types of higher education institutions with stratified missions and diversified structures coexisted. Private institutions, fully independent or government-dependent, are now parts of national systems in many countries; according to estimates by UNESCO, about a third of the students worldwide are currently enrolled in private institutions.

Scott's figures clearly show that the modern university is indeed a creation of the nation-state; this is even more so in the case of national systems of higher education. This led to increased regulation by ministries, and centralization of decision-making in many countries, as reflected in laws enacted by parliaments from the late 1960s on to the early 1980s in France, the Netherlands, Ireland, Germany, Austria, Sweden, and Spain. Guy Neave refers to this process of enacting legislation as a means of enforcing practice and implementing policy as the juridification of

higher education in Europe, whereby nearly all aspects of higher education, ranging from admission and access, curricular content, and internal governance to expenditures, were indeed regulated in detail.

1980 to the Present: Transformation to the Evaluative State and Higher Education as a Global Enterprise

Starting in the mid-1980s, patterns of higher education governance started to change radically. The main drivers of this transformation were the following:

- The changing view of higher education from a purely public service to be financed from the public purse to a semi-public service, the costs of which should be borne by all those who benefit from its outputs.
- A financial crisis resulting from increasing demand coupled with increasing costs.
- The political discourse on what the role of state in general should be in an advanced participant democracy.

Two developments in the 1980s, one in the United Kingdom and the other in the United States, had profound influences on higher education policy formulation throughout the world. The report issued by the Committee of Vice-Chancellors and Principals, known as the Jarrat Report, concluded that in order to be responsive to the needs of the market, managerial techniques had to be used in university governance, and that performance of universities had to be evaluated both qualitatively and quantitatively.

In 1980, the US Congress passed the Bayh–Dole Act, which allowed universities to patent and commercialize the results of federally funded research conducted within the university. Not only was a new source of income created for the universities, but also the ties with corporations grew stronger, and the university came to be viewed as a place that supplies commercially valuable initiatives. Perhaps as importantly, the traditional view of the products and outputs of the activities of the university as public goods began to change.

To the Jarrat Report and the Bayh–Dole Act must be added the literature that emerged, which portrayed higher education as a semi-public good with a private and a social return, rather than a purely public good.

The basic elements of the ensuing changes worldwide included the changing role of the state; new funding arrangements coupled with resource diversification, and increased managerialism in the administration of institutions. Neave refers to these changes as the transformation from the regulatory to the evaluative state. This transformation was accompanied by the introduction of the market as the supreme regulating principle of higher education.

The regulatory state prescribes the processes by which institutions function to produce outputs through an array of detailed legal instruments including laws, line-item budgets, guidelines, and rules. The evaluative state, on the other hand, sets forth institutional missions, qualitative and quantitative input, output targets, and confines itself to evaluating achievements while allowing institutions to determine their own ways of achieving those missions and targets. Among the major changes were lump-sum budgets, resource diversification through the introduction of or increase in tuition fees and provision of incentives for income generation, increased institutional powers, including professorial appointments, and discretion in financial matters.

There was little change in the United States when compared to the changes in continental Europe, Australia, and Japan. In the United Kingdom, too, the basic structure of governance remained essentially unchanged. Elsewhere, governance and funding mechanisms were radically restructured. These changes are collectively referred to as the rise of market forces in higher education, and can be depicted in Clark's triangle of coordination as a move from the state bureaucracy–academic oligarchy axis towards the

market–society apex as shown in **Figure 1**. This move manifests itself in the following ways:

- Resource diversification, with increased tuition fees, and a move from collegial structures to an entrepreneurial institutional culture.
- Spread of lay governance and strengthened institutional leadership.
- Emergence of new providers of higher education services and the blurring of the demarcation line between public and private institutions.

Tables 1 and **2**, adapted from the Organization for Economic Cooperation and Development (OECD) statistics, show the expenditure per student in higher education and the share of parents in those expenditures in the period 1990–2004 in OECD member and partner countries. There are wide cross-country variations in both indicators. In general, both the amount of money spent per student and the share of parents in expenditure have increased significantly. Tuition fees, though still controversial in many countries, are now established features of higher education governance even in continental Europe, except in Greece, the Czech Republic, and Scandinavia.

Table 1 Institutional Expenditure per Student, \$ (PPP)

Country	1990	1995	1998	2000	2001	2002	2003	2004
USA	ND	20 207	19 802	20 358	22 234	20 545	24 074	22 476
Germany	8459	9698	9481	10 898	10 504	10 999	11 594	12 255
Australia	9288	13 897	11 539	12 854	12 688	12 416	12 406	14 036
UK	9805	10 981	9699	9657	10 753	11 822	11 866	11 484
Brazil	ND	ND	14 618	11 946	ND	10 361	10 054	9019
Czech Rep.	ND	8785	5584	5431	5555	6236	6774	6752
Denmark	ND	11 499	9562	11 981	14 280	15 183	14 014	15 225
Finland	7070	10 900	7327	8244	10 981	11 768	12 047	12 505
France	6601	7801	7226	8373	8837	9276	10 704	10 668
Indonesia	ND	ND	6840	1799	1414	1296	ND	ND
India	ND	ND	ND	1831	2522	2486	ND	ND
Spain	3696	5654	5038	6666	7455	8020	8943	9378
Sweden	ND	ND	13 224	15 097	15 186	15 715	16 073	16 218
Switzerland	16 022	15 802	16 583	18 450	20 230	23 714	25 900	21 996
Israel	ND	ND	10 765	11 550	11 494	11 295	11 945	11 289
Italy	7300	5621	6295	8065	8347	8636	8764	7723
Iceland	ND	ND	ND	7994	7674	8251	8023	8881
Japan	ND	9691	9871	10 814	11 164	11 716	11 556	12 193
Canada	11 662	ND	14 579	14 983	ND	ND	16 937	ND
Korea	ND	ND	6356	6118	6618	6047	7089	7068
Hungary	ND	7767	5073	3223	7122	8205	8576	7723
Mexico	4463	4821	3800	4698	4341	6074	5774	5778
Portugal	ND	4664	ND	4766	5199	6960	7200	7741
Russia	ND	ND	ND	892	ND	ND	2451	2562
Chile	ND	ND	5897	7483	6901	7023	7011	6873
Turkey	2696	ND	ND	4121	3950	4267	4248	4231
New Zealand	ND	ND	ND	ND	ND	ND	8832	8866
Greece	ND	3264	4157	3402	4280	4731	4924	5593
OECD (CT)	ND	ND	11 720	11 109	12 319	13 343	14 530	14 027

ND, no data; CT, OECD country total average.

From OECD (1995–2007). *Education at a Glance*.

Table 2 Share of parents in higher education expenditures, %

Country	1995	2001	2002	2003	2004
USA	ND	33.9	38.9	36.7	35.1
Australia	20.0	31.0	33.7	34.6	35.6
UK	ND	17.3	16.6	18.5	19.4
Czech Rep.	3.3	7.7	7.4	7.3	9.2
Denmark	0.6	2.2	2.1	3.3	3.3
France	11.8	10.3	10.1	11.8	9.8
Indonesia	ND	49.4	48.4	ND	ND
India	ND	0.2	22.2	ND	ND
Spain	19.4	21.5	20.2	19.4	20.8
Israel	24.3	28.0	33.2	29.6	35.4
Italy	12.7	16.0	15.7	18.9	18.4
Iceland	ND	5.0	4.4	11.3	9.1
Japan	58.0	56.9	58.5	60.3	ND
Canada	16.7	22.9	ND	20.6	ND
Korea	ND	58.1	63.8	56.7	55.6
Hungary	4.8	6.1	5.4	5.3	6.6
Mexico	22.6	28.9	28.5	30.4	30.6
Portugal	3.5	ND	8.7	8.5	14.0
Chile	ND	77.8	81.4	83.3	83.7
Turkey	3.0	4.2	9.9	4.8	10.0
New Zealand	ND	ND	37.5	38.5	39.2
Greece	ND	0.4	0.4	0.4	0.4
OECD (CM)	ND	17.1	18.5	ND	ND

ND, no data; CM, OECD country mean average.

From OECD (1995–2007). *Education at a Glance*.

So are private institutions of all types, including for-profit providers, especially in countries along the Asia-Pacific Rim, Latin America, and Eastern Europe, where, in some cases they account for over half the national enrollment. All of these changes show that the role of the state as a provider of funds is diminishing, and in countries like Chile, Japan, Korea, and Indonesia private sources now account for most of the funding.

The accompanying changes in governance structures worldwide include a transfer of power to the university head and other university administrators and a loss of authority and decision-making power on the part of traditional participatory and collegial bodies. Although election of university heads continues in a number of countries, the trend seems to be moving towards appointment, often by a board with a majority of lay members. Sweden, Austria, Denmark, Norway, some of the Länder in Germany, such as Lower Saxony, and most recently Japan are countries where radical legislative changes have been introduced in that direction. In those countries where lay governance already existed, as in the United Kingdom, Australia, Sweden, and Holland, it has been strengthened.

Clark himself refers to these changes as the emergence of a new model of institutional behavior, which he calls the entrepreneurial university. As depicted in **Figure 1**, the new governance structures can also be characterized as a move away from the traditional collegial/bureaucratic continental

European model of governance to the entrepreneurial university model epitomized by American institutions.

A radical restructuring of traditional accountability structures and mechanisms accompanied these changes. From the mid-1980s on, starting in Europe, agencies and bodies with various mandates were established at the national and regional levels, which were charged with evaluating the performance of higher education institutions. Presently, there are over 100 such agencies in some 50 countries. Quality assurance and accreditation agencies and evaluation and assessment schemes are now established features of higher education governance worldwide. However, other than in the United States, and for part of the research funding in the United Kingdom, the outcomes of these practices are not yet directly tied to funding from public sources.

In summary, the emergence of national quality assessment agencies and the switch from line-item to lump-sum budgets, accompanied by strengthened role of the university head and the increased discretionary powers of the central institutional administration are basic features that characterize the transformation from the regulatory to the evaluative state.

The rise of market forces and the transformation of the role of state from regulatory to evaluative were no doubt accompanied and enhanced by the increased pace of globalization driven by rapid technological and political changes in the late 1980s and the early 1990s, and this is an ongoing process. Many view globalization as the biggest challenge faced by higher education in its history. Currently, the global higher education agenda is shaped by the following interdependent and mutually reinforcing phenomena: (1) accelerating demand, (2) demographic shift and nontraditional students, (3) the rise of market forces, (4) impact of technology, (5) new providers and increased competition, and (6) internationalization as a response to globalization.

The geographical scale of the competition for students, faculty members, and funds is now global. Competitors no longer comprise similar institutions, but include, in addition to traditional institutions, a whole array of new providers, most of which are for-profit and international, and use advanced technologies for delivery. The number of foreign students, which was 110 000 in 1950, is now over 2.7 million. Many different forms of cross-border delivery of higher education services are now available. In short, higher education is increasingly becoming a global enterprise.

The American university is emerging as the new referential model, and English is becoming the new *lingua franca*. Europe is countering this by the Bologna Process, which some regard as a more subtle and effective means of accomplishing change than legislation to make European institutions more competitive in the global higher education area/market by becoming more like American

institutions. The United States, supported by Australia, New Zealand, and Japan, much to the dismay of the global academic community, is pushing for educational services to be included in the presently stalled the General Agreement on Trade in Services (GATS) negotiations. Multilateral organizations such as UNESCO and OECD are responding by issuing codes of good practice in the cross-border provision of higher education.

Concluding Remarks: Toward a Global System with Universal Enrolment?

The new paradigms of higher education are meritocracy and entrepreneurialism rather than democracy and egalitarianism. At the national level, these changes are reflected in the form of new and more sophisticated control mechanisms. From the mid-1980s on, one-to-one control by the state and routine reporting and evaluations have been replaced by various types of quality assurance schemes, essentially adapted from the world of business. Many of the powers previously exercised by various ministries have been devolved to institutions themselves and external quality assurance agencies.

The role of the state in this increasingly global environment is becoming more complex. The international competitiveness imperative must also be recognized among the main drivers of this new approach to higher education governance, along with the transformation from the regulatory to the evaluative state and the rise of market forces globally.

According to UNESCO, a national gross enrolment ratio of 40–50% is needed for a country to be an active participant in the global knowledge economy and reap its benefits. Most Western countries have now achieved universal enrollment in higher education (gross enrolment ratio above 50%), with countries like Korea, Finland, United States, Sweden, Norway, New Zealand, and Australia now providing higher education to nearly the full age cohort. Many developing countries, on the other hand, are far from that stage, including China (19%) and India (11%), where enrolments have almost doubled in the past decade. Despite recent impressive growth rates, situation in many countries, especially in sub-Saharan Africa, remain bleak with dismal enrolment ratios of less than 5%.

Achieving universal enrolment levels worldwide is desirable for benefiting from globalization in an equitable manner. This requires more than the doubling of the current global average enrollment ratio, with the vast portion of the increase in student numbers occurring in China, India, and elsewhere in the developing world. States in the most developed countries may be required to assume additional, complex, and presently undefined roles in order to achieve a global higher education system

with universal enrolment, an international complementary role, as it were.

It appears that a fourth apex needs to be added to Clark's triangle of coordination that represents the increasingly complex international dimension of higher education.

Further Reading

- Altbach, P. (2006). Globalization and the university: Realities in an unequal world. In Forest, J. J. F. and Altbach, P. (eds.) *Global Themes and Contemporary Challenges, Part 1: International Handbook of Higher Education*, 2 Parts, pp 121–139. Dordrecht: Springer.
- Carnegie Foundation (1982). *The Control of the Campus*. Princeton, NJ: Princeton University.
- Clark, B. R. (1983). *The Higher Education System: Academic Organization in Cross-National Perspective*. Berkeley, CA: University of California Press.
- Clark, B. R. (1998). *Creating Entrepreneurial Universities: Organizational Pathways of Transformation* (Published for the IAU Press). London: Pergamon Press.
- de Groof, J., Neave, G., and Svec, J. (eds.) (1998). *Democracy and Governance in Higher Education, Vol. 2: Council of Europe Series on Legislating for Higher Education in Europe*. The Hague: Kluwer Law International.
- Guruz, K. (2008). *Higher Education and International Student Mobility in the Global Knowledge Economy*. Albany, NY: SUNY Press.
- Jarausch, K. H. (1983). Higher education and social change: Some comparative perspectives. In Jarausch, K. H. (ed.) *The Transformation of Higher Learning 1860–1930*, pp 9–36. Chicago, IL: University of Chicago Press.
- Jarausch, K. H. (1995). American students in Germany, 1815–1914: The structure of German and U.S. matriculants at Göttingen University. In Geitz, H., Heideking, J., and Herbst, J. (eds.) *German Influences on Education in the United States to 1917*, pp 195–211. Cambridge: Cambridge University Press.
- Kogan, M. and Marton, S. G. (2000). The state and higher education. In Kogan, M., Bauer, M., Bleikle, I., and Henkel, M. (eds.) *Transforming Higher Education: A Comparative Study*, ch. 4, pp 89–107. London: Jessica Kingsley Publishers.
- Neave, G. (1988). On the cultivation of quality, efficiency and enterprise: An overview of recent trends in higher education in Western Europe 1986–1988. *European Journal of Education* **23**, 7–23.
- Neave, G. (1998). The evaluative state reconsidered. *European Journal of Education* **33**(3), 265–283.
- OECD (1995–2007). *Education at a Glance*. <http://www.oecd.org/edu/eag> XXXX.
- OECD (2003). Changing patterns of governance in higher education. *Education Policy Analysis*, pp 59–75. Paris: OECD.
- Perkin, H. (2006). History of universities. In Forest, J. J. F. and Altbach, P. (eds.) *Global Themes and Contemporary Challenges, Part 1: International Handbook of Higher Education*, 2 Parts, pp 159–205. Dordrecht: Springer.
- Ramirez, F. O. and Riddle, P. (1991). The expansion of higher education. In Altbach, P. G. (ed.) *International Higher Education: An Encyclopedia*, vol. 1 (2 vols.), pp 91–106. New York: Garland.
- Ridder-Symoens, H. de (ed.) (1992). Universities in the middle ages. In Ruegg, W. (gen. ed.) *A History of the University in Europe*, vol. 1 (3 vols.). Cambridge: Cambridge University Press.
- Ridder-Symoens, H. de (ed.) (1996). Universities in early modern Europe (1500–1800). In Ruegg, W. (gen. ed.) *A History of the University in Europe*, vol. 2 (3 vols.). Cambridge: Cambridge University Press.
- Ruegg, W. (ed.) (2004). In Ruegg, W. (gen. ed.) *Universities in the Nineteenth and the Early Twentieth Centuries (1800–1945)*, Vol. 3:

- A History of the University in Europe*, 3 vols. Cambridge: Cambridge University Press.
- Scott, P. (1998). Massification, internationalization and globalization. In Scott, P. (ed.) *The Globalization of Higher Education*, pp 108–129. Buckingham, UK: Open University Press.
- Trow, M. (2006). Reflections on the transition from elite to mass to universal access: Forms and phases of higher education in modern societies since WWII. In Forest, J. J. F. and Altbach, P. (eds.) *Global Themes and Contemporary Challenges, Part 1: International Handbook of Higher Education*, 2 Parts, pp 243–280. Dordrecht: Springer.
- UNESCO (1963–1999). *Statistical Yearbook*. Paris: UNESCO.
- UNESCO (2004). *Global Education Digest 2004*. Montreal: UNESCO Institute for Statistics. <http://www.uis.unesco.org>.
- Veld, R., Fussel, H. P., and Neave, G. (eds.) (1996). *Relations between State and Higher Education, Vol. 1: Council of Europe Series on Legislating for Higher Education in Europe*, 2 vols. Dordrecht: Kluwer.

The Role of the State in School Reform: Responding to Increasing Ethnic Diversity in Britain*

P Ratcliffe, University of Warwick, Coventry, UK

© 2010 Elsevier Ltd. All rights reserved.

Education, especially in the early years of schooling, is universally acknowledged as crucial to future life chances. Arguably, this is even more the case for children from migrant communities whose command of the locally spoken language may be poor. This article focuses on the ways in which the British government has sought to address these challenges. A focus on the period since World War II (WWII) is appropriate since, unlike many other major Western countries such as the United States, large-scale immigration (at least in the modern era) has been largely confined to the last half century. Immigration up to the late 1980s was dominated by those from the New Commonwealth and Pakistan – NCWP (Layton-Henry, 1992) and most of today's settled minority communities can be traced back to these migrant cohorts. For this reason, the article focuses principally on changes in the educational system during this period. Over the next two decades, however, the influx of migrants was much more highly differentiated, with significant numbers of refugees and asylum-seekers and, since the expansion of the European Union (EU) post-2004, large numbers of workers from Eastern European accession countries. In the concluding section of the article, we therefore reflect on the implications of these new arrivals for the school system.

The dominant theme of the article is that most central government policies have been triggered by immediate political considerations and as a result have tended to adopt an overly narrow, doctrinal approach rather than taking a broader, historical perspective. This has invariably led to short-term, and short cut, solutions to long-term, deeply entrenched problems.

The Impact of Post-War Migration

In the wake of WWII, Britain witnessed a significant rise in immigration levels (Layton-Henry, 1992; Solomos, 2003). Much of this was in response to the needs of the economy which was desperately short of labor. As a consequence of the country's colonial exploits there was, it has been argued, a ready source waiting to be tapped (Castles and Kosack, 1973). However, following the first wave of

immigration in the late-1940s represented symbolically by the docking of the S.S. *Empire Windrush* in 1948 (Phillips and Phillips, 1998), the then British government consciously steered away from recruiting colonial labor. It turned instead to eastern and southern Europe, on the grounds that these workers were more likely to be white (Carter *et al.*, 1987). Skin color therefore counted for more in the government's calculations than the fact that workers from colonized countries were much more likely to speak English and had grown up in countries heavily influenced by British colonial regimes. When it, therefore, attempted to argue that the decision was taken on grounds of culture this argument had a somewhat hollow ring to it.

This source of labor did not satisfy the voracious appetite of the post-war economy, however, and the government therefore reluctantly turned to the Caribbean and the Indian subcontinent. The impact on the education system was immediate, provoking furious debates about how it could cope with the significant influx of children who not only often had a poor command of English but were from a very different cultural background, and were non-white. The latter is important, given a considerable degree of hostility on the part of the host community towards the new arrivals, a hostility that broke out into open conflict during serious race riots in the late-1950s (Rowe, 1998; Solomos, 2003; Ratcliffe, 2004). According to the government, the solution lay in a policy of assimilation.

From Assimilation to Multiculturalism

There had been little public debate about the implications for the education system of a multiethnic school population prior to the substantial influx of migrants from the disparate societies that constituted the New Commonwealth. Now, a range of issues were raised.

The first of these was language. Many children from the Indian subcontinent did not have English as a first language. Those from the Anglophone Caribbean were seen as presenting a somewhat different problem. Many of these children spoke various forms of *patois*. Viewed through an assimilationist lens, this was most commonly regarded as an unacceptable form of English. More significantly in many ways, though, was the fact that many teachers saw its use as undermining their authority and/or evidence that black children (especially males) had

* Adapted from Ratcliffe, P. (2004). Race, ethnicity and the educational system. In *Race, Ethnicity and Difference: Imagining an Inclusive Society*, pp. 73–88. Maidenhead: McGraw-Hill/Open University Press. With permission from McGraw-Hill.

a chip on their shoulder. The accusation, as Carby (1982) suggests, is that it was being used as a mode of resistance to the disciplinary code of the institution. In other words, the children stood accused of (consciously) opting for a form of English that was incomprehensible to their teachers. It was also, of course, a potent symbol of cultural difference, and therefore implicitly a threat to the goal of assimilation.

This is also interesting in that it reflects a common theme in the characterization of black as distinct from South Asian students. Whereas the latter were, and (with the possible exception of some Muslims) still are, usually viewed as conscientious and respectful, African-Caribbeans (especially males) have been characterized as more likely to be both confrontational and to lack the necessary application to their studies. Here, there are likely to be parallels with the US educational scene, especially with regard to comparisons between urban blacks and other, more overtly aspirant and upwardly mobile, minority groups.

The second issue links directly to the migration process and the settlement patterns which accompanied it. The majority of early migrants (irrespective of ethnicity) moved into inner urban areas. This meant in effect that residential segregation was accompanied by (ethnically) segregated schooling, although precise ethnic mix and segregation level might vary significantly from area to area.

As far as the government and key educationalists were concerned both represented major problems for the system. Rather than recognizing the potential for a richer teaching environment it was seen as undermining the dominant policy paradigm: the assimilation of migrant groups. Significantly, however, high numbers in particular schools of those whose first language was not English were seen as threatening the performance of local (white) students. It could therefore be argued that the main driver of state reforms at this point was as much to defend the interests of white pupils as it was to improve the performance levels of migrants.

Extra resources were directed toward the teaching of English and a local authority with a sizable New Commonwealth population could recruit additional teachers (part-funded directly by central government). As in the United States, bussing was seen as the solution to the imbalance in the ethnic composition of school populations (Carby, 1982). As in the United States also, the latter was both extremely unpopular (perhaps especially amongst the white population) and its application rather patchy. Bussing across the city also led to objections from migrants about the disruption to their children's teaching day and many white parents were unhappy for their children to study alongside (artificially) inflated numbers of African-Caribbean and South Asian pupils. To some, influenced by stories of disruptive pupils and underachievement on the part of some migrants, there were genuine fears that their children's education would be adversely affected. To rather more, however, their stance

was laced with hostility toward the other, based on a racist world view.

It was clear by the mid-1960s that the policy of assimilation was not working (Troyna and Carrington, 1990; Mason, 2000). There was understandable resistance on the part of migrant communities to the pressure to conform to an implicitly mono-cultural model imposed by the central government (Parekh, 2000). Coinciding with the rethink of immigration policy geared to the limiting of numbers was the idea that integration did not necessarily mean assimilation. The power balance at central government level had shifted to the Left with the Labour government of Harold Wilson replacing the former Conservative administration. Key parliamentarians, such as Roy Jenkins, felt that a more liberal policy of promoting cultural diversity in an atmosphere of mutual tolerance was far preferable to attempting to coerce migrants into shedding core elements of their ethnic and cultural heritage.

The upshot of this was that cultural pluralism became the watchword in theoretical terms, the idea being that migrants would have equal access to the polity and equal citizenship rights (including, of course, education). As far as the school system was concerned this meant the introduction of multicultural education (MCE).

MCE was seen as a solution to a problem referred to earlier, namely underachievement. The latter will be seen later to have taken a highly particularistic form during this period in the sense of being both racialized and gendered, that is, being applied to black, and not South Asian children, and to young males rather than females. There were also clearly identifiable links with the Race and IQ and Bell Curve debates flourishing in the United States (Jenson, 1969; Herrnstein and Murray 1994) and replicated in the United Kingdom by Eysenck (1971). Poor performance was also blamed on pathological familial forms (such as the absent male parent), evidence of a victim blaming model based on a narrow ethnocentrism (Lawrence, 1982).

Promoting Change: MCE and Anti-Racist Education (ARE)

The idea behind MCE was that if pupils of different heritages understood more about each other's ethnic, religious and cultural backgrounds this would bring two benefits. It would result in a healthier and more productive learning environment and would bring more long-term benefits to the society as a whole. Compared with what had preceded it, this was an extremely laudable attempt to democratize the school system. It suffered from some major defects, however. By identifying culture and cultural traditions as the key barriers to the integration of minority pupils and their performance potential, the approach failed to appreciate the broader social and political significance of race and racism. In other words,

it failed to recognize the bigger picture: that an imbalance of power stemming from a white hegemony did much more to limit the potential life chances of minorities than did the latter's cultural diversity.

It could also be viewed as the easy option. It is, after all, much easier to incorporate cultural diversity into teaching than to undertake a radical overhaul of the curriculum and to rethink core aspects of school culture. In practice, the authorities that made serious attempts to embrace the new reforms endorsed variants of a policy characterized by Troyna and Ball (1985) as the "three Ss – Saris, Samosas and Steel Bands." This banality of the approach is well illustrated by the following statement from the Select Committee on Race Relations, 1968–69 and quoted by Carby:

.....much can.....be done.....to create better understanding.....of the national and cultural background of immigrants..... (thus leading the Committee to recommend that).....specific teaching about the countries from which the immigrants in a particular town come. Here material direct from those countries can be displayed in the classroom by immigrant children. Children in primary schools in Hackney or Brixton [London], for example, could be taught West Indian songs, or children in Wolverhampton are taught Indian art, jewellery and costumes. This would help bring the immigrant children into the life of the school. (Carby, 1982: 194)

Although its adherents were for the most part well meaning, in practice, it did little more than essentialize cultural difference and generate crude caricatures of Asian or West Indian culture. Radical-left critiques often saw it as essentially an exercise in social control (Carby, 1982), an attempt to lure minority children into acquiescing in their own exploitation by appearing to respond to their interests and needs. Others (on the political Right) attacked the policy from a quite different perspective, suggesting that it represented unwarranted special treatment for minorities and an unwarranted waste of public funds. In the latter respect they were presented with ample ammunition. There was much anecdotal evidence of teachers being sent on fact-finding trips to the (former) colonies to observe cultural differences for themselves. For the critics, such profligacy merely compounded concerns about the extra expense incurred by changes to curricula.

On the positive side, MCE did at least constitute an official systemic response to the increasingly diverse school population. It made a conscious commitment to respecting the belief systems of faiths other than Christianity. It also helped to bring about a greater sense of relevance to classroom discussions about geography, history, arts, music, and the media. The problem, as with the Black Studies movement in the United States, is that new teaching material was essentially tacked onto existing curricula. It therefore tended to appear as supplementary, rather than

integral, to the existing syllabus. Crucially, it was often viewed as of more relevance to (or even only of relevance to) the migrant than to the indigenous child.

The main problem, though, was that it did not address, at least directly, the main issue for schools. It assumed that racism was simply a question of ignorance: that with a concerted attempt to educate young citizens about the nature and value of cultural diversity, it would wither away. It therefore conveniently ignored the fact that racism was far too deeply embedded in common sense knowledge for this approach to bear fruit. Racism is about power relations outside the specific context of the school and the educational system. However, its effects are also in evidence within schools, with instances of insensitive, ethnocentric, and even racist behavior on the part of teachers and other staff (Mason, 2000: 71; Ratcliffe, 2004). This, combined with the ongoing debate about underachievement on the part of minority children, provided the rationale for the introduction of anti-racist education (ARE).

A further problem with MCE was that it was never adopted universally in British schools. Authorities representing the white highlands were loathe to accept that they had a problem in their area (by which they meant children from minority ethnic groups). The argument that all children need to be prepared for entry into a multicultural adult society was rarely seen as persuasive, nor was the point that areas are not hermetically sealed (as white) in perpetuity. Its application was even patchy elsewhere. Given this, and in the absence of a strong lead from government, it was never likely that ARE would gain universal approval. After all it would, if properly applied, entail a radical form of intervention into the culture and ethos of the school (as well as the curriculum).

Few authorities took a robust approach to ARE. Those which did were prone to attack from neoliberal politicians. One such authority was Brent, in north-west London. Brent became the symbol of an extremist form of intervention in schools. Cross *et al.* (1991) provide a dispassionate account of the authority's wide-ranging review of curriculum and school management. Rather than 'doing good by stealth,' Troyna's characterization of the normative process of educational system evolution, Brent undertook a root-and-branch reappraisal of the education delivered to local children. They disposed of the remnants of MCE, a process symbolized most prominently by the suspension of a universally popular head-teacher, Maureen McGoldrick, and set up what was effectively a local inspectorate to oversee the installation of teachers who were committed to ARE. The degree of radicalism, and the fact that even many South Asian and African-Caribbean parents were extremely concerned by the apparent administrative chaos, left the authority vulnerable to attack even from the mainstream (and relatively pro-ARE) media. Most notable here was a program in the long-running flagship

documentary series, *Panorama*, from the BBC in 1986. Entitled *Brent Education rules OK*, this presented a savage indictment of a school system allegedly captured, and subverted, by a small clique of left-wing ideologists determined to impose their will irrespective of opposition from the ‘voices of reason’ (including Mrs. McGoldrick and other proponents of MCE).

Some local authorities claimed allegiance to the idea of ARE, and undertook a more limited review of pedagogic practice and management style (Troyna and Carrington, 1990). Others did little more than relabel and repackage existing MCE policies. This led many commentators to argue that although very different in theory, MCE and ARE tended to differ little in practice (Carby, 1982).

A further near-fatal blow to ARE came with a tragedy in Burnage High School in south Manchester in northern England. Here, an Asian pupil was murdered in the playground by a white fellow student (Macdonald *et al.*, 1986; Troyna and Hatcher, 1992). This provoked a wave of horror, given that this was a racist murder in a school that promoted its ARE policy assiduously. As this was seen as a national test case, both the miscreant and the school’s policy were effectively on trial. The public inquiry set up to look into the broader implications of the incident deemed that the latter was also, to a large extent, culpable. MacDonald *et al.* (1989) argued that poor, dispossessed white pupils who were already devalued within the normative ethos of the school were inadvertently made to feel even more worthless by the ARE approach. Though not providing the immediate catalyst, the resulting anger was regarded to have contributed in large part to the murder. It therefore shone a spotlight on ARE and deterred other schools that might have been considering going down this route.

The Development of a National Curriculum

The election in 1979 of the Conservative government led by Margaret Thatcher also had a significant impact on the national policy context within which schools operated. Thatcher had no sympathy for the attempts to rectify the problems faced by migrant pupils, and therefore the problems faced in Brent and Burnage simply played into her hands. She and her Cabinet colleagues set about transforming the school system into one which was based around a national curriculum of core subjects such as English, math, and science. According to Troyna and Carrington (1990), the first half of the 1980s was characterized by a government preparing the ideological terrain for a radical reconstruction of the educational landscape. Left-inspired initiatives such as ARE were attacked and ridiculed and local authority education departments in major towns and cities were vulnerable to attack if

they invested in them. Although the core curriculum introduced in the Education Reform Act 1988 was to leave some space for soft subjects designed to prepare youngsters more generally for their lives after school, the bulk of a pupil’s time would be devoted to what many educationalists saw as overly academic pursuits. The government also demanded a return to traditional (i.e., more formal) teaching methods, mocking the trendy methods introduced in the 1960s.

The fear, justified as it turned out, was that many working-class pupils would feel even more alienated from school life. Levels of truancy soared and those who attended often exhibited disenchantment to the point where they were excluded from school, either on a temporary or permanent basis. This effectively left a large number of young people with no formal education.

Schools were encouraged to secede from local authority control and run their own budgets. This policy, heralded (by central government) as promising an increase in local choice and self-determination, was actually a way of undermining and disempowering local authorities. It was effectively an exercise in centralization masquerading as empowerment (Ratcliffe, 2004). Although schools could decide through their boards of governors to institute, say, forms of MCE/ARE, this was always extremely unlikely, given resource implications and the ethnic makeup of most boards (Troyna and Carrington, 1990).

Recapping Core Issues: Ethnic Differentials in Schooling

Policies such as MCE and ARE were introduced as part of a conscious effort to generate a more inclusive educational system and, as an important byproduct, to address some key ethnic differentials in experience and outcome. We now summarize the major issues under three headings: underachievement and differential performance, suspensions and exclusions, and rationing. The issue of ethnic segregation and how this might be tackled dominates current debates and will be treated as a major topic in its own right (in the next section of the article).

Underachievement and Differential Performance

The thorny issue of differential outcomes remains at the top of the public policy agenda, featuring in a recent government report (Department for Communities and Local Government, 2007) and a major national study based on a sophisticated multivariate analysis of all available quantitative data on the subject (Strand, 2008). In the context of the current article, however, it is important to recognize that the debate has a long history stretching back at least half a century. Significantly, in 1979, the then

Labour government sponsored an inquiry chaired by Anthony (Lord) Rampton. This was tasked to investigate the roots of educational underachievement, the premise being the long-held fact that young males of African-Caribbean origin performed consistently worse than those from other minorities. This had become so much a part of common-sense knowledge that it had been incorporated in the syllabi of prospective teachers, and in the process becoming a self-fulfilling prophecy. Teachers had literally been trained to observe the problem, not question it (a point of great contemporary significance (see Strand, 2008)).

Rampton endorsed the orthodox view, but argued that institutional racism was a serious problem to be addressed. This had angered the incoming Thatcher administration, which sought to argue that racial/ethnic (and even class) differentials were no longer a significant issue. Unfortunately for them, however, the race issue would not go away. Serious outbreaks of urban violence in 1980 and 1981 were deemed race riots by politicians, media, and academia alike (Phillips and Phillips, 1998; Rowe, 1998; Solomos, 2003; Ratcliffe, 2004).

Rampton stepped down as Chair, allegedly because of disagreements with the government, and the agenda was subsequently taken up by Lord Swann (Troyna and Carrington, 1990). His Commission endorsed the position on institutional racism taken by Rampton, but failed to say how schools might address the problem (apart for a number of comments on the curriculum (Mason, 2000: 73). Worse than this, it individualized what was effectively a collective problem, by endorsing the use of racism awareness training (RAT). On the basis of the rotten apple theory of racism, offending teachers would be taken out of the classroom and subjected to retraining. This is, of course, much easier to do than to address a collective institutional failure.

On the positive side, Swann championed the cause of targeted recruitment drives so as to increase the number of teachers from the minority communities. Extra resources would be devoted to this end. It also focused in some depth on the value of creating an atmosphere which emphasizes the value of cultural diversity.

Like Rampton before him, the Swann Report (1985) reiterated the facts of underachievement. As Troyna (1988) rightly points out, however, the analysis was deeply flawed. The crude label Asian was used to represent a large number of culturally and ethnically distinct migrant groups and, as if to compound the error, birthplace was used as the basis for classification. The lack of a serious analysis of the influence of social class factors also made it difficult to say how much differential educational performance could be explained by class rather than ethnicity (or, for that matter, some other factor such as gender or school attended). As Mirza (1992) points out, global data on those of African-Caribbean background

conceals marked gender differences with young females performing consistently better than their male counterparts (as they continue to do).

In fact, class background has a major influence on performance, both directly and indirectly. The middle-class child (irrespective of ethnicity) is likely to benefit both from a relatively privileged socialization and a more generously funded school outside an inner urban area. This largely explains why those defined (in conventional census terms) as Indian tend to leave school with better results than their poorer Pakistani and Bangladeshi neighbors, irrespective of gender.

The real dangers in the erroneous analysis from Swann, however, lay in its impact on those policies which sought to redress such ethnic differentials: MCE and ARE. As Troyna (1988) argues, the results seemed to justify the arguments championed by proponents of cultural deficit theory. If Asians not only performed better than black pupils, but sometimes (even) outperformed whites, then surely this casts doubt on the need for ARE (as racism would presumably also impact upon Asians)? In addition, if those from different cultural and ethnic backgrounds achieved such different results, is it wise to treat all cultures as equally conducive to educational advancement (as MCE does)?

If we look at the contemporary data on school-leaving qualifications it is clear that despite some narrowing of ethnic differentials, some major deficits remain. The lowest-performing group is made up of children of Gypsy, Traveller, and Roma origin (DCLG, 2007), explaining the formation of a number of voluntary and charitable bodies seeking to improve their lot in this respect. Working-class pupils (especially boys) of Bangladeshi and Pakistani origin continue to perform badly, but so do those of black- or African-Caribbean heritage and whites. As in the United States, some (middle class) minority groups outperform whites of a similar background.

The picture is therefore extremely complicated, but in terms of debates about performance the focus of attention is shifting somewhat toward socioeconomic class and gender. In terms of the latter, there is increasing research and comment about the often problematic relationship between ethnic/class identity, masculinity, and the educational process. In particular the suggestion is that, among some groups of boys at least, it is simply not cool to do well at school. Controversy has raged over the suggestion that black street styles stemming from phenomena, such as Gangsta Rap and Yardie culture, have not only impacted on attitudes to formal education on the part of black male youth but have also had a knock-on effect on their white peers. In other words, black subcultural forms are now often being blamed for white (working-class male) underachievement! To the extent that sections of South Asian youth are also adopting black street styles which leads to intergenerational conflict within these

communities. The effect of this is to disturb existing ethnic stereotypes that associate Asian culture with commitment to education and status enhancement via a conventional aspirational orientation.

Suspensions and Exclusions

One of the major effects of alienation brought about by poor leadership within schools and insensitive policies in the face of a polyethnic student body has been a rise in school suspensions and, more especially, exclusions. All the available evidence suggests that males of African-Caribbean origin have long been far more likely to be permanently excluded than any other group (Bourne *et al.*, 1994; Gillborn and Gipps, 1996). But there is evidence that this is ultimately the outcome of a series of rather more insidious exclusionary processes. Teacher-training folklore, as was noted earlier, has embraced the notion that these children are to be expected to underperform. This expectation (rather than evidence) has often resulted in allocation to inappropriately low streams in key subjects (Commission for Racial Equality, 1988). Role ascription, that is, applying assumed group characteristics to individuals, then often associates poor performance with resistance to the aims of the institution (the children having chips on their shoulders – c.f. earlier). Ostensibly similar behavior on the part of white, South Asian, or (say) Chinese students is much less likely to be interpreted in the same way or regarded as justifying a similar punishment (Ratcliffe, 1999).

Rationing

In Britain, the legal minimum school-leaving age is 16 (though the current government has indicated its intention to raise this to 18). Prior to leaving full-time schooling most students sit a number of General Certificate of Secondary Education (GCSE) examinations. Although formal unseen examinations are still common, much of the assessment is undertaken through written work that is not examined under such strict supervision. Crucially, teachers decide the appropriate number of subjects for which a student is registered (a potential site, therefore, for discriminatory behavior). The passing at (top) grades A*–C of five or more subjects has become the gold standard for assessing achievement at the statutory minimum school-leaving age. Schools are then ranked in league tables on the basis of the proportion of students who achieve this standard.

The importance to individual schools of obtaining a high rank in terms of their popularity and hence future financial viability means that they value their high achievers and regard so-called problem students (the less academically talented as well as the disaffected) as a threat. The impact of this is that the temptation to suspend and exclude such students has become more likely. Many prominent educationalists have contested use of the

GCSE 5 A*–C economy as a core performance indicator, principally on the grounds that it does not convey any sense of value added. In other words, it is simply not valid to use this statistic to compare the performance of poorly funded and staffed urban schools with suburban schools that can recruit a more privileged student body. Gillborn and Youdell (2000) point to a rather more sinister unintended consequence, however. As implied earlier, those students who are seen by teachers as destined to let the school down (by failing to reach the standard) may receive less attention than those considered potential achievers. Even worse, there is now effectively an incentive to exclude those considered disruptive, especially if they are likely to impinge on the performance of others. This explains why the authors refer to the rationing of education. Once again, it appears that it is male students of African-Caribbean origin who suffer disproportionately.

Alternative Educational Futures: Separatism or Diversity?

Although we have largely focused in this paper on state policies to redress underachievement, this is not intended to suggest that migrants have been passive victims of an exclusionary educational system. To counteract the latter's effects many local minority communities have developed neighborhood support strategies (usually with a rather broader educational remit than available in mainstream provision). The Saturday school movement pioneered by African-Caribbean communities around the country was already in evidence even in the 1960s (Rex and Tomlinson, 1979). Mosques, temples, and community centers have served a vital parallel function in the case of South Asians.

These developments were almost universally welcomed as a highly positive form of community self-help. The fact that they were examples of ethnically segregated schooling was not seen as a problem because this would simply supplement integrated state schooling. There has been an increasing demand over the last few years largely, though not exclusively, from Muslims for separate schooling. This is justified on religious, cultural, and pedagogic grounds. As Islam relates to a whole way of life, it is not possible to separate the public from the private domain. It is also argued that the state school system is incompatible with Islam in a number of respects, not least in the area of gender relations. More recently arguments have often focused on the obvious fact that many of Britain's schools seem to be failing their children, generating a new cohort of underachievers. Increasing pressure from the Muslim Association in Bradford led in 2000 to the establishment of Britain's first Muslim state secondary school (Wainwright, 2000). (In fact, there have long been private Islamic schools in parts of the country where these communities are

concentrated. More recently, some of these have sought state support, exploiting the apparent popularity of faith schools in government circles.)

The more general question of whether this should be seen as setting a broader precedent presented government with an extremely tricky dilemma. There have been Church of England, Roman Catholic, and even Jewish state schools in Britain for many years. It was therefore difficult to argue against the principle of faith schools. They were also proven to perform rather better than average on the government's own performance indicators. Further, in the spirit of social inclusivity and of the more participatory democracy championed, in theory at least, by Tony Blair in his period as prime minister, this seemed to be a clear case in point. The subtext was that Blair was known (not least through the choice of schooling for his own children) to support the notion of selectivity, contrary to many of those on the political Left of his party.

The position held by most traditional Labour Party supporters has long been that selection by schools based on academic or social criteria was inherently unfair. It was seen, further, as a means by which social elites reproduced themselves over generations, allowing little opportunity for upward mobility on the part of those outside these more privileged groups. Private schools, in particular, were seen as anathema to the egalitarian ideals of the Party. However, New Labour under Prime Minister Blair shifted ground somewhat much to the chagrin of those on the Left.

There were also some extremely compelling reasons why this new form of separatism might not represent a positive development. Acute spatial segregation in a small number of urban areas in England was largely responsible for a phenomenon labeled in national policy discourse as parallel lives (Home Office, 2001a, b; Cantle, 2005). Alluded to earlier, this stems from the presence of urban segregation by ethnicity and faith. The fact that many children were seen to grow up in largely mono-ethnic communities, went to segregated schools and workplaces, and did not even mix with those from other ethnic/faith backgrounds at leisure or prayer was deemed to be a major cause of urban tensions. This became even more of an issue in the wake of the urban unrest in a number of northern English towns in summer 2001 and the events of 11 September later that year (Kalra, 2003). The problem then for policy makers was how to de-segregate schools, not how to sponsor the creation of more (along religious lines). Realizing that her views would be highly unpopular with many fellow Muslims, Alibhai-Brown (2000) added to the voices of those opposing segregated schooling. Her view was that this would simply increase levels of distrust between communities and was not in the long-term interests of a cohesive, polyethnic Britain.

The government's position gradually became more equivocal. From an initial position of support for more

faith schools, ministers began to suggest that conditions might be applied prior to approval. The key argument was that school governors would not be permitted to operate an exclusivist policy with respect to those of other ethnic or faith backgrounds. In fact, a certain proportion of places would effectively have to be reserved for (in our example) non-Muslims. As to the argument that there were already many faith schools in Britain, the response was that these did not formally restrict entry to members of a particular faith community. Government ministers also suggested that, with the benefit of hindsight, the formation of such schools may not have been very wise.

Since 9/11 and the wars in Afghanistan and Iraq, there seems to have been a general change of tack by the government. Initially welcoming the celebration of difference (as well as diversity), the policy focus has shifted to modes of integration and, some commentators have suggested, even assimilation (if only by the back door). Further exacerbated by continuing flows of refugees and asylum seekers, there is a renewed emphasis on universal English language acquisition and a commitment to an (as yet ill-defined) notion of Britishness.

Concluding Thoughts: British Education Policy in the Early Twenty-First Century

Whereas, from the mid-1970s, racial equality was the keyword in British social policy discourse (including education); the focus since 2001 has gradually shifted toward the idea of promoting 'community cohesion' (Cantle, 2005). Although seen by its proponents, most notably Cantle, as a new means to the same end – equality – this is highly debatable. Furthermore, despite many protestations to the contrary the agenda has been driven largely by the Muslim question: in other words, how can Britain respond to those (sections of the) Muslim population who appear disenchanted by contemporary Western cultural norms and the geopolitical strategies of most Western nations? More pejoratively, radical Islam is viewed as evidence of an enemy within, the archetypal other. Cohesion is therefore driven more by a concern for social order than by a commitment to generate a more inclusive and equitable society. In terms of schooling, there are two core strands to this. One relates to segregation by faith/ethnicity and how to minimize it. The other concerns how schools might modify their pedagogical approach and normative value systems to promote cohesion.

As noted earlier, residential segregation on the basis of faith and ethnicity, although a relatively minor issue in the United Kingdom relative to the United States and a number of other Western societies, can and does generate parallel demographic profiles in schools (i.e., high concentrations of certain groups in particular localities). This phenomenon is then at times magnified by what has

been characterized by some writers as ‘white flight’ – white parents opting to send their children to a school in which white pupils predominate.

Current cohesion policies rely largely on selective bussing to facilitate special interschool events based on sporting or other activities. With wider demographic changes across urban Britain, however, certain areas have been presented with a unique opportunity to tackle the problem by deploying a far more radical solution. This involves razing to the ground all or most existing schools in a local authority area, and building entirely new schools in localities that are known to appeal to those from all communities, irrespective of ethnic and faith background (Department for Communities and Local Government (DCLG), 2006). With this *tabula rasa* the theory is that a more inclusive education may be more easily achieved.

The problem is that, even where this approach is successful in generating a more balanced pupil roll, it may not in itself deal with the problem of segregation within schools. As the DCLG study referred to in the previous paragraph amply demonstrated, friendship and intraschool/college networks have thus far tended not to transcend ethnic/faith boundaries. Central government policy, however, is based on the idea that a greater level of communitarianism has a better chance of flourishing in such multiethnic environments. In other words, it might ultimately serve to improve overall societal cohesion.

The quest for cohesion is seen as yet more urgent following the rapid expansion of the European Union since 2004. With ten new Accession countries, inward migration from Eastern Europe has increased substantially. More worrying for local education authorities is the fact that the impact on localities varies enormously and is highly unpredictable (both in terms of numbers and migrants’ countries of origin). This has resulted, not surprisingly, in significant pressures on certain schools required, at short notice, to add to their rolls pupils whose native language is not English. In many cases, the number of languages spoken by pupils creates major problems for schools (given the absence, in the foreseeable future, of appropriate (and substantially increased) resources).

To summarize, current national policy aims to tackle two key challenges faced by Britain’s school system:

- The need to minimize differential educational outcomes by ethnicity, faith, gender, and socioeconomic status (Cabinet Office, 2007; Department for Communities and Local Government (DCLG), 2007).
- The need to generate an educational system that works effectively both to integrate a highly diverse and rapidly changing student population and to serve the much wider social goal of achieving a more cohesive society over the longer term (Commission on Integration and Cohesion, 2007).

Bibliography

- Alibhai-Brown, Y. (2001). *Who Do We Think We Are? Imagining the New Britain*. Harmondsworth: Penguin Books.
- Bourne, J., Bridges, L., and Searle, C. (1994). *Outcast England: How Schools Exclude Black Children*. London: Institute of Race Relations.
- Cabinet Office (2007). *Fairness and Freedom: The Final Report of the Equalities Review*. London: Cabinet Office.
- Cantle, T. (2005). *Community Cohesion: A New Framework for Race and Diversity*. Basingstoke: Palgrave Macmillan.
- Carby, H. (1982). *Schooling in Babylon*. In Centre for Contemporary Cultural Studies (ed.), *The Empire Strikes Back*. London: Hutchinson/Routledge.
- Carter, B., Harris, C., and Joshi, S. (1987). The 1951–55 Conservative government and the racialisation of Black immigration. *Immigrants and Minorities* 6(3), 335–347.
- Castles, S. and Kosack, G. (1973). *Immigrant Workers and Class Structure in Western Europe*. London: IRR/Oxford University Press.
- Commission for Racial Equality (1988). *Medical School Admissions: Report of a Formal Investigation into St. George’s Hospital Medical School*. London: CRE.
- Commission on Integration and Cohesion (2007). *Our Shared Future*. London: Commission on Integration and Cohesion/DCLG.
- Cross, M., Brar, H., and McLeod, M. (1991). *Racial equality and the local state: An evaluation of race policy in the London Borough of Brent*. Monographs in Ethnic Relations No. 1. Coventry: ESRC/Centre for Research in Ethnic Relations.
- Department for Communities and Local Government (DCLG) (2006). *Managing for Diversity: A Case Study of Four Local Authorities*. London: DCLG.
- Department for Communities and Local Government (DCLG) (2007). *Improving Opportunity, Strengthening Society: Two Years on, a Progress Report*. London: DCLG.
- Eyesenck, H. (1971). *Race, Intelligence and Education*. London: Temple Smith.
- Gillborn, D. and Gipps, C. (1996). *Recent Research on the Achievements of Ethnic Minority Pupils*. London: HMSO.
- Gillborn, D. and Youdell, D. (2000). *Rationing Education*. Buckingham: Open University Press.
- Herrnstein, J. and Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.
- Home Office (2001a). *Community Cohesion: A Report of the Independent Review Team*. London: Home Office.
- Home Office (2001b). *Building Cohesive Communities: A Report of the Ministerial Group on Public Order and Community Cohesion*. London: Home Office.
- Jenson, A. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review* 39(1), 1–123.
- Kalra, V. (2003). Police lore and community disorder: Diversity in the criminal justice system. In Mason, D. (ed.) *Explaining Ethnic Differences: Changing Patterns of Disadvantage in Britain* pp 139–152. Bristol: Policy Press.
- Lawrence, E. (1982). In the abundance of water the fool is thirsty: Sociology and Black ‘pathology’. In Centre for Contemporary Cultural Studies (ed.) *The Empire Strikes Back: Race and Racism in 70s Britain*, pp 95–142. London: Hutchinson/Routledge.
- Layton-Henry, Z. (1992). *The Politics of Immigration: Immigration, Race and Race Relations in Post-War Britain*. Oxford: Blackwell.
- Macdonald, I., Bhavnani, T., Khan, L., and John, G. (1986). *Murder in the playground: The Report of the Macdonald Inquiry into Racism and Racial Violence in Manchester Schools*. London: Longsight Press.
- Mason, D. (2000). *Race and Ethnicity in Modern Britain*. Oxford: Oxford University Press.
- Mirza, H. (1992). *Young, Female and Black*. London: Routledge.
- Parekh, B. (2000). *Beyond Multiculturalism: Cultural Diversity and Political Theory*. Basingstoke: Palgrave Macmillan.
- Phillips, M. and Phillips, T. (1998). *Windrush: The Irresistible Rise of Multi-Racial Britain*. London: Harper/Collins.

- Ratcliffe, P. (1999). 'Race', education and the discourse of exclusion: A critical research note. *Race, Ethnicity and Education* 2(1), 149–155.
- Ratcliffe, P. (2004). *'Race', Ethnicity and Difference: Imagining an Inclusive Society*. Maidenhead: McGraw-Hill/Open University Press.
- Rex, J. and Tomlinson, S. (1979). *Colonial Immigrants in a British City: A Class Analysis*. London: Routledge and Kegan Paul.
- Rowe, M. (1998). *The Racialisation of Disorder in Twentieth Century Britain*. Aldershot: Ashgate.
- Solomos, J. (2003). *Race and Racism in Britain*, 3rd edn. Basingstoke: Palgrave Macmillan.
- Strand, S. (2008). *Minority Ethnic Pupils in the Longitudinal Study of Young People in England*. London: Department for Children, Schools and Families.
- Troyna, B. (1988). Paradigm regained: A critique of "cultural deficit" perspectives in contemporary educational research. *Comparative Education* 24(3), 273–283.
- Troyna, B. and Ball, W. (1985). Views from the chalk face: School responses to an LEA's multicultural education policy. *Policy Papers in Ethnic Relations, No. 1, ESRC/Centre for Research in Ethnic Relations*.
- Troyna, B. and Carrington, B. (1990). *Education, Racism and Reform*. London: Routledge.
- Troyna, B. and Hatcher, R. (1992). *Racism in Children's Lives*. London: Routledge.
- Wainwright, M. (2000). Approval for first Muslim secondary state school. *Guardian*, 7/10/03.

Transforming Mindsets Through Education for Sustainable Development

V M Segovia, Miriam College – Environmental Studies Institute (ESI), Quezon City, Philippines

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Climate change – Any long-term significant change in the average weather that a given region experiences. Average weather may include average temperature, precipitation, and wind patterns. It involves changes in the variability or average state of the atmosphere over durations ranging from decades to millions of years.

Dark Green School – A school in which environment is taught, preached, and lived. This is an accreditation process initiated in the Philippines by the Environmental Education Network of the Philippines in which standards are set, and a review is undertaken to determine the level of greenness of an educational institution. The degree to which a school is committed reflects the degree it can be deemed green.

Dominion model – A model of man's way of relating to the environment characterized by the belief that man is above and apart from nature, and that his role in the world is to dominate or tame nature.

Ecosystems – The dynamic interactions between plants, animals, and microorganisms and their environment working together as one functional unit.

Enlightenment philosophy – A philosophical movement in the eighteenth century, identified with Western philosophy, which refers to the period when reason reigned supreme over mysticism and religious belief. This movement left a deep imprint in the intellectual, moral, and cultural life of the world, paving the way for a mindset of utilitarianism and human dominion over nature.

Environmental spirituality – This is also known as eco-spirituality. This is a spirituality that advocates kinship or closeness with the Earth and the natural world.

Gated suburban subdivisions – A social phenomenon in many urban and suburban places, where the rich sectors of society converge and build large and beautiful homes in exclusive subdivisions secured with high walls and a very tight security system.

Indigenous Peoples – A group of people or homogenous societies identified by self-ascription and

ascription by others, who have continuously lived as an organized community on communally bounded and defined territory, and who have, under claims of ownership since time immemorial, occupied, possessed, and utilized such territories, sharing common bonds of language, customs, traditions, and other distinctive cultural traits. Modern man can learn sustainable ways of living from the genuine Indigenous Peoples who have inhabited their territories for centuries without disturbing their ecosystems.

Judaean-Christian tradition – A term used to describe the body of concepts and values that are thought to be held in common by Judaism and Christianity, and considered by some as (often along with classical Greco-Roman civilization) a fundamental basis for Western legal codes and moral values.

Mindscape – Also worldview or paradigm, is a person's mental model of the world. It is a set of fundamental assumptions a person makes about how the world works, and his/her relationship with that world.

Paradigm shift – Sometimes known as extraordinary science or revolutionary science. It is the term first used by Thomas Kuhn in his influential 1962 book *The Structure of Scientific Revolutions* to describe a change in basic assumptions within the ruling theory of science. It is in contrast to his idea of normal science. In popular parlance, paradigm shift is a term used to denote a change in how a given society goes about organizing and understanding reality.

Sustainable development – A term coined by the group of Gro Harlem Brundtland in the book *Our Common Future*; it refers to development that meets the needs of the present generation without compromising the ability of future generations to meet their own needs.

Transformation – A marked change, as in appearance or character, usually for the better. In this article, educational transformation refers to changes in the educational systems, content, and processes that favor a change in the mindsets of students toward a sustainable future.

United Nations decade of education for sustainable development – The 10 years from 2005 to 2014 has been declared the United Nations decade on education for sustainable development – a time when peoples around the world are asked to think about an educational approach toward a more sustainable planet Earth.

Introduction

The illiterates of the 21st century will not be those who cannot read or write, but those who cannot learn, unlearn and re-learn. (Alvin Toffler)

The world is changing very fast. Technology that is being peddled this year becomes obsolete even before one has completed payments for it on one's credit card. The irony is that with all the advances we are making in technology, we are also given the capacity to destroy the world faster. Thus, today we see a planet beset with problems in the face of this fast-paced change; environmental crisis, in particular, climate change is threatening many countries with disasters and social dislocation. Indeed, in our present world, we can no longer do things the way we used to. We have to change our ways of thinking, doing, and the way we confront present-day challenges.

Given the present situation, how do we prepare our students to face a world different from that into which we stepped after our own graduation? Is there a need to review and revise knowledge management? In the face of present cultural and societal realities, how should education be rendered?

Surely, we do not intend to graduate more engineers who will hasten the destruction of our ecosystems! Why is it then that our present graduates seem to be more interested in seeking jobs that would enable them to live in gated suburban subdivisions rather than in service-oriented professions that would satisfy both the pocket-book and the spirit? In the Philippines, many young people seek entrance into nursing or allied health professions that would ensure them job placements in other countries, such as the United States, United Kingdom, Middle East, and other more developed countries in Asia where wages would be much higher than those they would get in their own country. The worst scenario is the case of graduates in education and/or business management who acquire jobs as domestic helpers abroad. Do we want this situation to continue or worsen? Given these phenomena, educators need to review the content and process of education. What new teaching pedagogy should we employ in order to form the kind of citizen we envision for the future? What should

be included in the future curriculum and what campus culture should we develop so that we can ensure a world where people are productive, educated citizens who are positive agents of change toward a better global community?

UNESCO declared 2005–14 as the United Nations decade of education for sustainable development (ESD) saying that, with the present planetary conditions, ESD is no longer a choice but an absolute priority. In 1987, sustainable development was defined by the World Council on Environment and Development (WCED) as “development that meets the needs of the present without compromising the ability of future generations to meet their own needs.” Can ESD be the key in the transformation of present societies toward a sustainable future?

Modern Society: Values and Culture

Our landscape, our seascape and our skyscape is a product of our mindscape.

What is mindscape? It is a person's worldview or paradigm; a mental model of the world. It is a set of fundamental assumptions a person makes about how the world works and his/her relationship to that world. This is usually a by-product of education, upbringing and culture – plus what the religions teach. This is called paradigm development, or the development of values that translate into culture, traditions, and mores.

In order to bring about ESD, there is a need to review the prevailing paradigms in our modern society and critically scrutinize how these paradigms somehow seep into what we teach our students, influencing their value systems, and even campus culture, thus unwittingly prodding them to a lifestyle that could threaten the Earth's ecosystems.

Thomas Berry, an American priest, cites two reasons for the shortsighted and misdirected way modern man has handled nature: (1) a lack of cosmology, that is, Father Berry says that modern people have lost a meaningful sense of identity due to the fact that they have lost their myths or their stories to account for the emergence of the world and their place in that world; and (2) the emphasis of Christian tradition on transcendence, where God is seen as separate from nature. The Christian tradition on transcendence is closely linked with the development of Western philosophy that has been dominated by the so-called Enlightenment mentality. In turn, the Enlightenment philosophy, with its proclivity toward science and industrialization, has brought about the unintended negative consequences that have been observed as detrimental to the environment. Some observers allege that the Judeo-Christian tradition beginning with the story of Genesis, in

which humankind is granted dominion over the Earth, has led civilization to its present dismal state. In its basic form, the charge is that this tradition assigns divine purpose to the exercise of virtually complete power to work human will over nature. This tenet of endowing human beings with a completely unique relationship to God, and then delegating God's authority over nature to human beings, sanctions as ethical all choices that put a higher priority on human needs and desires than on the rest of nature. Simply put, according to this view, it is ethical to make sure that whenever nature gets in the way of what humans want, nature loses.

Faced with environmental and ecological destruction, today's educators have started to reexamine this worldview. They say that the cure for the present crisis will come not only from a scientific or technological advance, but from a radical shift in philosophical outlook. This change includes rethinking metaphysical, epistemological, and political, as well as ethical concepts. It involves both personal and cultural transformations, and goes deep into the realm of values and beliefs, which others call the spiritual consciousness of humanity.

Albert Gore, one-time Vice-President of the United States stated: "The more deeply I search for the roots of the global environmental crisis, the more I am convinced that it is an outer manifestation of an inner crisis that is, for lack of a better word, spiritual. As a politician, I know full well the special hazards of using spiritual to describe a problem like this one. But what other word describes the collection of values and assumptions that determine our basic understanding of how we fit into the universe?" In his topic on environmentalism of the spirit, he adds,

the way we experience the world is governed by a kind of inner ecology that relates perceptions, emotions, thinking and choices to forces outside of ourselves. We interpret our experience through multiple lenses that focus – and distort – the information we receive through our senses. But this ecology threatens to fall badly out of balance because the cumulative impact of the changes brought by the scientific and technological revolution are potentially devastating to our sense of who we are and what our purpose in life might be. . . . To some, the global environmental crisis is primarily a crisis of values . . .

(Albert Gore, *Earth in the Balance: Ecology and the Human Spirit*, 241–243)

Given this worldview, Al Gore is calling for a new environmentalism of the spirit. Echoing these sentiments, the former Minister of the Environment of Brazil, Honorable Jose Lutzenberger, adds: "We humans find ourselves at war with the rest of creation. We must change our philosophy, our worldview; look at the planet and see it as something alive, that we humans are only part of an enormous symphony where millions of species participate. We are one

species among millions of others. Unless we realize that we are only part of a bigger system, we have no future."

In our quest for a new ethics, we look at a new metaphysics – an alternative view of reality. Viewing human beings as individuals with power over nature is the way that the dominant worldview has understood humans and broken up reality into a hierarchical pattern, but it is a dangerous and misleading metaphysics. What then, could be a reasonable model for understanding nature? It would be wise to review the different worldviews gleaned from other groups and derive from them a renewed model of conceptualizing the environment. For example, Native American spiritual teachers often tell of the importance of place to maintain one's spiritual connections to the earth. The Native American land ethic teaches that spiritual connections to the earth are born from the heightened awareness and identification that comes from staying in one place long enough to experience, learn from, and care for the spirit of the place one is in. That assumption of care and responsibility for the environment within which one lives thereby becomes natural and inevitable. The best illustration of this environmental spirit is manifested in the reply of Chief Seattle to the President of the United States of America in 1854 when he refused to sell their territory to the white colonizers:

How can you buy or sell the sky, the warmth of the land? The idea is strange to us. If we do not own the freshness of the air and the sparkle of the water, how can you buy them? . . . Every part of this earth is sacred to my people. . .

He adds:

the white man's dead forget the country of their birth when they go to walk among the stars. Our dead never forget this beautiful earth, for it is the mother of the red man . . . we are part of the earth and it is part of us. We know that the white man does not understand our ways. One portion of the land is the same to him as the next; he treats his mother the earth, and his brother, the sky as things to be bought, plundered, sold like sheep or bright bead.

The Kalingas of the Cordilleras echo the same sentiment expressed through the resistance of Makli-ing Dulag and his people when their lands were threatened by a development project of the government: "Such an arrogance to speak of owning the land, when you shall be owned by it. How can you own that which will outlive you. Only the race owns the land, because only the race lives forever."

In the Philippines, statistics report about 127 main groups of Indigenous Peoples (IPs) whose way of life manifests an environmental ethic that indicates their closeness to the environment. To them, ancestral domain is a sacred land area: God's gift to a tribe or tribal community, the source of their life where their ancestors live. Another tribe in Mindanao, the Lumads, claims that

“the reason for all this environmental degradation is that when foreigners came to the Philippines, they put God too high in the Sky.” We have suffered here because we can no longer see the sacred in the trees, flowers, and our fellowmen. It is sad to note that the IPs have been marginalized and disempowered, and are losing their lands to development aggression. Today, this is pungently observed in the brazen way some indigenous groups are forced to leave their lands to give way to mining concessions. In Southern Philippines, the tribal leader of the Subanons and his family are being persecuted due to their resistance to the mining activities of a foreign-owned company. On one occasion, he lamented,

X corporation has started its illegal operations, and has desecrated our altar, the tip of Mount Canatuan which is our most sacred place. They bulldozed the tip of the mountain, destroyed our most holy place, and in a matter of weeks, our community will be wiped out due to their mad drive for gold.

Thomas Berry put it very wisely when he stated that the “universal need at the present time is a re-orientation of the human venture toward such an intimate experience of the world around us.” He said that “a return to the mystique of the Earth is a primary requirement if we are ever going to establish a viable rapport between humans and the earth.”

So what new model should we adopt to bring about this new consciousness? This model need not be the creation of something new but a re-awakening of something very old. It involves the cultivation of an ecological consciousness – that is, an ecological, philosophical, and spiritual approach to the present crisis, that recognizes the unity of humans, plants, animals, and the Earth.

When the human subject is viewed as essentially one with the natural world, the rationale for clinging to a strict distinction between objective and subjective, between real and perceived, and between fact and value is weakened. The perception that the real world exists out there separate and apart from us would be transformed into one of we exist in the real world.” This is where an alternative metaphysics, one inspired by aboriginal spirituality, can offer a fresh insight into the meaning of man and his relationship to his environment. The following paragraphs will tackle indigenous spirituality discourse in the attempt to show how this attitude and value system has been respectful of the earth.

In ancient history, called the prehistoric hunter-gatherer and agrarian society, humans strove to live in harmony with nature; nature was all powerful. From that time to the scientific age, nature and its forces have been personified as deities to be worshipped and revered. In the horticultural societies of southeastern Europe in the period 7000–3500 BC, female deities represented the natural fertility and generative power in nature. In many aboriginal

and Native American societies, Mother Earth or the Great Mother was regarded as the beginning and end of all life on earth. The reverence for nature deities was present in the spiritual traditions of shamanism. For over 20 000 years, shamanic spiritual traditions have guided indigenous peoples throughout the world. In this holistic and integrative tradition, spirituality is not a religion with fixed set dogmas but rather, it pervades and infuses all forms of existence – human, animate, and inanimate. It informs the view that all that exists in nature is living and sacred, therefore deserving of respect and care. In sum, these nature-centered traditions emphasize human beings’ physical and spiritual connections with animate and sacred nature. As an ecological belief system, they teach the need for respect of all that exists in the natural environment as well as the need for transformational change to regain holistic balance and harmony in the cosmos. In addition to native spiritual traditions, there are also lessons that one can gain by studying other oriental-based religions that are holistic in the manner that they view man’s relationship with his/her environment.

Developing a sustainable relationship with the environment requires a deep awareness not only of the biophysical environment within which humans live but also of their own spirituality. Mankind needs to regain its sense of awe and interconnectedness with other creatures and the Earth.

Today, there is a worldwide call for a shift in paradigm by groups who believe that our present civilization could no longer afford to live with a business-as-usual attitude. They warn that doing so would lead twenty-first-century civilization to perish with the planet’s demise. In this call for a change of consciousness, educational institutions are challenged to take the lead through a review of educational contents and approaches. Educators are further urged to develop higher-order thinking skills in their students, so that graduates can wisely critique the basic structures and institutions, as well as philosophical underpinnings of the present-day order of things in society.

This call to lead in the task of transformation is indeed timely; it is about time education begins to serve as a catalyst in the change of the human mindset toward a new world order that promotes life, and not merely material or economic goods.

The Role of Education

In a world threatened by social collapse and environmental degradation, educational institutions are called upon to play a major role in the transformation of mindsets to develop citizens who are trained to think in terms of the common good. To the chagrin of some observers in Philippine society, some schools in the country have narrowed down the function of education from developing the dimensions of human potential, to the goal of simply

becoming competitive in the marketplace or bureaucratic defenders of one-sided economic policies. This concern is echoed in the call of educators to re-invent education to work toward participatory and sustainable development. This was the theme of the 11th UNESCO – Asia-Pacific Programme of Educational Innovation for Development (APEID) International Conference in Bangkok, Thailand in December 2007. APEID held a conference on Reinventing Higher Education – Toward Participatory and Sustainable Development. Beyond the Philippines and Asia Pacific, to what extent does education transcend the generation of wealth and the advancement of self-interest? It will be interesting to make a comparative analysis of this issue. Education is one of the most effective means that society possesses to shape the world of tomorrow. It plays a critical role in creating and disseminating knowledge, skills, and values for society. Colleges and universities prepare most of the professionals who develop, lead, manage, teach, work in, and influence society's institutions. Progress increasingly depends upon the products of educated minds: upon research, invention, innovation, and adaptation that schools impart to their students. As such, educational institutions have the privilege, the power, and the mandate to bring about the transformation of society.

Education therefore is called upon to guarantee that students are knowledgeable about the changes that are needed, capable of envisioning alternative futures, committed to ethical ways of achieving them, and sufficiently skilled and motivated to work actively for a better world. But for this to happen, it is necessary for educators and educational institutions to review, deconstruct, and re-construct their current mindsets. There is a saying that you cannot give what you do not possess. To become leaders in the movement toward a sustainable future, they need to understand, appreciate and own the concept of ESD.

Understanding ESD

What Is ESD?

ESD is a vision of education that seeks to balance human and economic well-being with cultural traditions and respect for the earth's natural resources.

ESD applies transdisciplinary educational methods and approaches to develop an ethic of lifelong learning; it integrates concepts and analytical tools from a variety of disciplines to help people better understand the world in which they live. It fosters respect for human needs that are compatible with the sustainable use of natural resources and the needs of the planet.

The four pillars of sustainable development have been identified as:

- society;
- culture;

- economy; and
- environment.

Sometimes only three are identified, with culture taken as part of society. The overall aim of ESD is to empower citizens to act for positive environmental and social change, implying a participatory and action-oriented approach.

Brief Background of ESD

ESD is a concept or paradigm that came about as part of an action plan formulated in the 1992 United Nations Conference on Environment and Development (UNCED) or the Earth Summit in Rio de Janeiro. Its aim was to provide a comprehensive set of principles to assist governments and other institutions in implementing sustainable development policies and programs.

The four major thrusts of ESD as articulated in Agenda 21, Chapter 36 are:

- The promotion and improvement of basic education – basic education must focus on imparting knowledge, skills, values, and perspectives that encourage and support citizens to lead sustainable lives;
- The reorientation of existing education at all levels to address sustainable development – that is, rethinking and revising education from nursery school through university to include more principles, skills, perspectives, and values related to sustainability in each of the three realms – social, environmental, and economic – as important to our current and future societies;
- The development of public understanding and awareness of sustainability, and making progress toward more sustainable societies requires a population that is aware of the goals of sustainability and has the knowledge and skills to contribute toward those objectives. Informed voting citizenry and knowledgeable consumers can help communities and governments enact sustainability measures and move toward more sustainable societies;
- Training – all sectors of the workforce can contribute to local, regional, and national sustainability. The development of specialized training programs to ensure that all sectors of the workforce have the knowledge and skills necessary to perform their work in a sustainable manner has been identified as a crucial component of ESD.

ESD means education that enables people to foresee, confront, and solve the problems that threaten life on the planet. Thus, education is that which produces professionals equipped with knowledge, competencies, and skills not only in their respective disciplines/professions, but also in valuing and taking responsibility for the need to balance economic and social progress with concern for the environment and the stewardship of natural resources.

There is, therefore, the need to reorient the curriculum. It should be reviewed, rethought, and reformed to be a vehicle of knowledge, thought patterns, and values required to build a sustainable world.

The ESD-integrated curriculum equips graduates with an understanding of the complexity and interrelationships of problems such as poverty, wasteful consumption, environmental degradation, urban decay, population growth, gender inequality, health, conflict, and the violation of human rights. ESD will look at the social transformation of the student: (1) on the personal level, ESD will be education for life; and (2) on the societal level, ESD will be education for the common good. The 2002 Johannesburg Summit stated that to promote the pillars of sustainable development, economic development, social development, and environmental safeguards should be seen as interdependent and mutually reinforcing.

Why ESD?

Modern society lives in an endangered planet. While information technology has grown very fast, the irony is that even with all our high technology, we live in a disturbed world – a violent world and we are destroying the world, our only habitat in the present. The signs of unsustainable growth are visible all around us. The economic growth rate is high, with environmental sustainability at risk. Yet, there are many unmet needs in developing countries of the South, like the Philippines, where there is evidence of high poverty levels, undernourishment, inadequate health services, high death rate, and natural disasters. There is also inadequate housing, energy, water, and sanitation. In developed countries, there is noticeable expansion of urban settlements, the proliferation of resource-intensive industries, with the accompanying hidden costs in health, increase in waste, and air pollution.

Meanwhile, the spread of the global consumer society has increased the demand for raw materials, energy, and water, taxing the finite resources of the planet Earth.

Education for the Future

Mankind needs to change the way they live, the way they consume, and the way they promote business if they want to ensure a sustainable world for the future generations. Green growth should be the policy in government, industry, and in school.

How should ESD be integrated in the school of the future?

1. Update educational practice with recent knowledge derived from research on sustainable development, while catalyzing changes in educational paradigms to

enable knowledge of sustainability to be naturally incorporated into the system and curriculum.

2. Conduct school-based longitudinal studies using student work samples to determine the impact of ESD curriculum on student-learning results.
3. Conduct research to establish and strengthen an open-ended research agenda to inform and strengthen key areas of ESD practice in teacher education, for example, in curriculum change; participatory action research; auditing of institutional resource management, and sustainability practices.
4. Conduct research to establish and strengthen a vibrant community of practice in education, which strengthens the teaching and research capabilities of educators involved in ESD.

One important component of ESD is the emphasis on the holistic approach to education. The three components of education, specifically higher education, which comprises of teaching, research, and extension, should be synergized to produce ESD. Systems thinking and systems analysis should be employed to develop a shared framework for understanding and addressing complex nonlinear systems that govern the natural world and society.

In the Philippines, the Environmental Education Network of the Philippines (EENP) is embarking on a project titled 'Institutionalizing the Dark-Green Schools Concept in the Accrediting System of Philippine Schools.' EENP is composed of 68 member institutions with the aim of promoting environmental education for sustainable development (EESD) through a strong and independent network of Philippine institutions able to provide, national and local governments, private firms, nongovernment organizations, people's organizations, and local communities, advice and expertise on environmental education.

The Dark Green School or DGS project sets standards and initiates a process to determine the level of greenness. Greening schools is a catch phrase that refers to a school committed to mainstream environmental learning and consciousness in all aspects of its educational function. The degree to which a school is committed reflects the degree it can be deemed green. A DGS, therefore, is one in which environment is taught, preached, and lived. The vision is to integrate the DGS concept in the accreditation system of the academic community of the country. The accreditation project looks into the following elements in schools: (1) school policy; (2) administration and finance; (3) academics (which includes research); and (4) outreach to see whether these contribute to the sound protection and sustainable management of the environment. This accreditation program will start with EENP schools but is expected to be adopted by the accrediting system of Philippine schools. The project is expected to encourage schools in strengthening the environmental base of all the aspects of their operation. Indirectly, it will gauge the

strengths and weaknesses of environmental education in the Philippines, and find out how these could be enhanced or mitigated.

Ultimately, the project aims at toward empowering the schools to effectively deliver EESD.

Imagine what the future could be if all educational institutions joined in this effort of transforming twenty-first-century society into sustainable communities? Imagine the kind of society that we would develop if all schools adopted ESD as their educational strategy? Students would learn the art of addressing issues as an interconnected system and not as fragmented problems with separate and often competing solutions. The curriculum would inculcate an appreciation of the human–environment interdependence, making biocentric values and ethics a central part in all the disciplines, rather than offering them as special subject only for certain courses. Education would be participative, democratic, and exploratory so as to facilitate real-world problem solving in the campus and in the larger community.

Education for a sustainable future involves a comprehensive approach to educational reform. May this generation bravely meet the challenge of transforming education toward a better world for everyone!

Further Reading

- Agenda 21 and the UNCED Proceedings (1992). New York, USA: Oceania Publications.
- Berry, T. (1988). *The Dream of the Earth*. San Francisco, CA: Sierra Club Books.
- Clover, D. E., Follen, S., and Hall, B. *The Nature of Transformation: Environmental Adult Education*, 2nd edn. Toronto, ON: York University Bookstore.
- Desjardins, J. R. (1992). *Environmental Ethics: An Introduction*. Belmont, CA: Wadsworth Publishing.
- Dryzek, J. S. (2005). *The Politics of the Earth, Environmental Discourses*, 2nd edn. New York: Oxford University Press.
- Earth Summit (1992). *The United Nations Conference on Environment and Development*. Rio de Janeiro, 1992. ABB Asea Brown Boveri Ltd., Reader Services Center, PO Box 822, CH-8021 Zurich, Switzerland.
- Education For Sustainability: An Agenda for Action (1994). *A Report on the National Forum on Partnerships Supporting Education About the Environment*, held at the Presidio, San Francisco, CA, President's Council on Sustainable Development.
- Engel, J. R. and Engel, J. G. (eds.) (1992). *Ethics of Environment and Development: Global Challenge, International Response*. London: Belhaven Press.
- Fischer, F. and Maarten, A. H. (eds.) (1999). *Living With Nature: Environmental Politics as Cultural Discourse*. New York: Oxford University Press, Inc.
- Fox, M. (1998). *The Coming of the Cosmic Christ: The Healing of Mother Earth and the Birth of a Global Renaissance*. San Francisco, CA: Harper Collins Publishers.
- Gore, A. (1992). *Earth in the Balance: Ecology and the Human Spirit*. New York: Houghton Mifflin Company.
- Greening the Ivory Tower*. The MIT Press. <http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=5269> (accessed October 2009).
- Haigh, M. J. (1995). Worldviews and environmental action: A practical exercise. In Yvonne, G., et al. (eds.) *Values and the Environment: A Social Science Perspective*. New York: Wiley.
- Hallman, D. G. (ed.) (1994). *Eco-Theology: Voices from the South and North*. Geneva: WCC Publications.
- Kalaw, M. T. (1997). *Exploring Soul and Society: Papers on Sustainable Development*. Pasig City: ANVIL Publishing.
- Kaufman, F. A. (2003). *Foundations of Environmental Philosophy: A Text with Readings*. New York: McGraw Hill.
- Mary, E. T. and John, G. (eds.) (1994). *Worldviews and Ecology. Religion, Philosophy and the Environment*. Maryknoll, NY: Orbis Books.
- McDaniel, J. B. (1995). *With Roots and Wings. Christianity in an Age of Ecology and Dialogue*. Maryknoll, NY: Orbis Books.
- Moore, T. (1994). *Care of the Soul: A Guide For Cultivating Depth and Sacredness in Everyday Life*. New York: Harper Collins.
- Perlas, N. (1998). *Philippine Agenda 21: Handbook*. Quezon City, Philippines: CADI.
- Philippine Agenda 21 (1997). *A National Agenda for Sustainable Development*. Quezon City, Philippines.
- Philippine Institute for Alternative Futures (1992). *Handbook On Sustainable Development*.
- Quisumbing, L. R. (2007). Education for Sustainable Development: Teaching and Learning for Sustainable Futures. *Lecture Given to Miriam College Faculty during the Conference on Implementing the UNESCO Decade for ESD*.
- Sessions, G. (ed.) (1995). *Deep Ecology for the 21st Century: Readings on the Philosophy and Practice of Environmentalism*. Boston, MA: Shambala Publications.
- UNESCO (1997). *Educating for A Sustainable Future*.
- UNESCO (2006). *Teaching and Learning for a Sustainable Future. A Multi-Media Teacher Education Programme, 2006*. <http://www.unesco.org/education/tlsf/> (accessed October 2009).
- 11th UNESCO-APEID International Conference (2007). *Reinventing Higher Education Toward Participatory and Sustainable Development*. Bangkok, Thailand, 12–14 December.
- Wright, N. and Kill, D. (1993). *Ecological Healing, A Christian Vision*. Maryknoll, NY: Orbis Books and CODEL.

Relevant Websites

- <http://www.urbanoptions.org> – Sustainability Education Handbook.
- <http://www.secondnature.org> – Second Nature – Education for Sustainability.
- <http://www.concord.org> – The Concord Consortium, Sustainable Development Education.

DEMOGRAPHY AND SOCIAL CHANGE – SOCIOLOGY

Contents

Achievement Motivation in Ethnic Minority Youth
Affirmative Action and Higher Education in Brazil
Black in White: Black Students at White and Black Colleges
Bologna Process: On the way to a Common European Higher Education Area
Critical Race Theory in Education Research
Cultural Diversification and Japanese Education: Social Constructions of the New Diversity
Education in War Countries: A Disempowerment of Society
Educational Attainments of U.S. Black Males and Females: 1971 to 2003
Educational Preparation: Fostering The Self-Efficacy And Resilience Of Urban Adolescent Youth
HIV Education for Low-Literate People: Transforming Students and Communities through Paulo Freire's Praxis and the Pedagogy of Action
How Are Discussions of Interdisciplinary Studies Linked to Diversity Discourse?
Identifying Patterns of Doctoral Attrition Across Academic Fields
Male Underachievement in Education Across the Globe: A Shift in Paradigm for Gender Disparities Regarding Academic Achievement
Noncognitive Measures for Higher Education Admissions
Peace Education
Remote Control: Africa's Development and Outflow to Developed Nations of Technical Talent
Scholar-Baller®: Student Athlete Socialization, Motivation and Academic Performance in American Society
Service-Learning and Diversity as Mediums to Foster Civic and Community Engagement in Undergraduates
Success and Failure in Education and Criminal Justice: Identifying Common Mechanisms
The History of Education: Race and Education

Achievement Motivation in Ethnic Minority Youth

S Graham and E Bryant, University of California, Los Angeles, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Disidentification – A term used by stereotype threat researchers to describe students who no longer view academic achievement as a domain that is either important to them or their self-definition; the absence of a relationship between academic performance and self-esteem that has been associated with declining achievement from middle school to high school.

Entity theory or intelligence – The belief that intelligence is basically fixed, each person has their allotted share, and there is little they can do to increase that allotment.

Incremental theory of intelligence – The belief that intelligence is modifiable through increased effort.

Microaggressions – A subtle kind of discrimination that many people of color encounter on an almost daily basis, such as being ignored or overlooked

while waiting in line, being suspected of cheating because one received a good grade on a test, being followed or observed while in public places, or being mistaken for someone who serves others.

Model minority stereotype – A belief that Asian Americans are hardworking and intellectually gifted high achievers who are especially competent in math and science; originated in the 1960s to account for the seemingly unprecedented successful entry of East Asian immigrants into mainstream American society.

Oppositional identity – A tendency of some adolescent of color to show relative indifference, or even disdain toward achievement behaviors that are valued by the larger society because they are perceived as threatening to one's social identity.

Stereotype threat – The awareness that individuals have about negative stereotypes associated with

their group; in the achievement domain, it has been applied to African American students' awareness of the cultural stereotype associating their race with intellectual inferiority.

Motivation is the study of why people think and behave as they do. A person would be concerned with motivation if he/she were to ask, for example, why some students choose tasks so difficult that they are bound to fail; or why some students persist to task completion despite enormous difficulty, while others give up at the slightest provocation. Note that this focus on the 'why' or the 'why not' of achievement is quite different from the study of achievement itself. Educators sometimes confuse the topics of researchers who study motivation with the topics of researchers who study achievement and learning.

Most contemporary approaches to motivation as defined above fit within an expectancy-value framework. Motivation (e.g., choice and persistence) is determined by both the likelihood of goal attainment, which is the expectancy component, and how much that goal is desired, which is the value component. The study of expectancy addresses beliefs about ability (Can I do it?), whereas the study of values addresses preferences and desires (Do I want it?). The notions of 'can' versus 'want' closely correspond to the distinctions between aptitude and effort expenditure, which are two of the most important constructs in motivation theory and research.

In this article, the review of research on achievement motivation in children and adolescents of color is organized around overarching questions about 'can' and 'want.' With the focus on race and ethnicity, this approach differs from the general motivation literature in two ways. First, we start with the larger sociocultural context and factors unique to the everyday lives of people of color in the United States. Some of those factors are historical and structural in nature. Thus, we recognize that many racial and ethnic minority groups in contemporary America are positioned at the bottom of a status hierarchy wherein barriers to opportunity often override personal strivings for achievement. Second, we approach the study of motivation in ethnic minority youth from a social contextual perspective. We believe students' social lives and their academic lives are interrelated and that one cannot fully understand an individual student's motivation to achieve without also understanding his/her relationships with other people – in particular, what he/she thinks and feels about others and how others think and feel about him/her. The social nature of motivation is particularly important during adolescence when peer approval takes on a heightened significance.

We begin with how beliefs about ability (Can I do it?) are shaped by racial stereotypes about intelligence, students' implicit theories about intelligence, and coping

with discrimination. Next, we turn to achievement values (Do I want it?) in which we focus on oppositional identity, the valued characteristics of peers, and self-affirmation. The main topics reviewed here encompass vast interdisciplinary literatures that we cannot do justice to in this article. Our goal is to use our knowledge of the topics as a framework for highlighting the unique challenges of racial and ethnic minority youth as they strive to achieve and master their environment.

Beliefs About Ability: Can I Do It?

Racial Stereotypes About Ability

Stereotypes are culturally shared beliefs, both positive and negative, about the characteristics and behaviors of particular groups. There is a great deal of research from social psychology showing that the cultural stereotypes about African Americans remain largely negative. Many studies show respondents associate being black (and male) with low intelligence, hostility, aggressiveness, and violence. The much smaller stereotype literature on other ethnic groups in the United States also portrays the more marginalized groups in a negative light. For example, cultural stereotypes of Latino youth also depict them as unintelligent and with little personal ambition (e.g., Kao, 2000).

African Americans and stereotype threat

Research on stereotype threat (Steele, 1997) has provided important insights into the negative consequences of racial stereotypes about intelligence. Stereotype threat is the awareness individuals have about negative stereotypes associated with their group. Although considered to be a general psychological state applicable to any negative group stereotype, the construct originated in the achievement domain and it has been applied to African-American students' awareness of the cultural stereotype associating their race with intellectual inferiority; this awareness can be quite debilitating, especially for those African-American students who are invested in doing well in school. For example, in a series of studies with black and white students attending Stanford University, Steele and colleagues found that black students performed more poorly than whites on standardized test items when they were told the test was of a diagnostic nature to test their abilities. When the studies told the test was a problem-solving activity unrelated to ability, there was no difference in the performance of the two racial groups. In ability-related contexts, therefore, what became threatening for black students was the fear that they might confirm the stereotype or that they might be judged by others based on that stereotype. Stereotype-threatened students are often dividing their attention between the task itself (e.g., taking a Scholastic Aptitude Test (SAT)) and ruminating about the meaning of their performance (e.g., what

does this say about me or about members of my racial group?). Such ruminations have been shown to arouse anxiety as well as deplete the executive resources needed to do well on academic tasks. Furthermore, it is not necessary that a student should endorse the stereotype – mere awareness of its existence is sufficient to activate threat.

Stereotype threat researchers have documented two motivational consequences of the anxiety associated with thinking about race and intelligence in highly evaluative achievement contexts (Steele, 1997). Some students may choose to work especially hard as a way of disconfirming the stereotype. Of course, heightened effort in the face of increasing academic challenge may be difficult to sustain and may even lead one to question the students' abilities. Stereotype threat can also have the opposite effect, causing students to minimize effort and downplay the importance of doing well in school. Steele coined the term 'academic disidentification' to describe the state of students who no longer view academic achievement as a domain that is important to them or their self-definition. Disidentification has been operationalized as the absence of a relationship between academic performance and self-esteem and it has been associated with declining achievement from middle school to high school, particularly among African-American boys. A similar process, labeled academic disengagement, occurs when students begin to discount the feedback they receive about their performance or to devalue achievement altogether. Thus, while disidentification and disengagement may be self-protecting mechanisms for coping with negative racial stereotypes, in the long run, their detrimental effects on achievement strivings will probably outweigh any short-term, self-enhancing effects.

Asian Americans and the model minority stereotype

There is also a literature on the motivational effects of racial stereotypes related to intelligence at the other end of the ability spectrum. Unlike African Americans, the cultural stereotype about Asian Americans is that they are hardworking and intellectually gifted high achievers who are especially competent in math and science. The term model minority was coined in the 1960s to capture those characteristics and to account for the seemingly unprecedented successful entry of East Asian immigrants into mainstream American society. Many studies have now documented that Asians and non-Asians alike are aware of the culturally shared association between high academic achievement strivings and being an Asian American (e.g., Kao, 2000).

While it may be more tolerable to know that one's racial group is viewed as smart and hard working rather than dumb and lazy, that stereotype also has its own set of challenges. Ethnography, survey, and experimental

research all point to psychological and emotional costs associated with living up to the model minority stereotype. Ethnographic studies, for example, detail the anxiety many Asian-American students feel when forced to cope with the perception of their group as academic superstars. Many report feeling frustrated and pressured to attain or maintain high academic achievement because of the expectations placed upon them. As one Asian-American student poignantly disclosed: "When you get bad grades, people look at you really strangely because you are sort of distorting the way they see an Asian. It makes you feel really awkward if you don't fit the stereotype" (Lee, 1994: 419).

Thus, while some African-American students worry their performance might confirm a negative stereotype, some Asian-American students worry their performance might not confirm a positive stereotype. These worries may be exacerbated among groups from Southeast Asia (e.g., Cambodia and Lao) whose immigrant experience have been quite different from those of their East Asian and South Asian counterparts. We believe coping with ability-related stereotypes in the academic domain, negative or positive, can lead to performance-avoidance goals (doing whatever it takes to conceal low ability), which have well-established negative motivational consequences.

Entity Versus Incremental Theories About Intelligence

Whether individuals are vulnerable to racial stereotypes or to other challenges to 'I can,' is in part determined by their beliefs about the meaning of intelligence. For example, does it make a difference if students view their abilities as stable and essentially beyond their control ("either you have it or you don't") or as changeable and largely determined by how much effort they exert? To address this question, Dweck proposed a distinction between individuals who hold entity versus incremental theories of intelligence (Dweck, 1999). Some individuals are entity theorists: they believe intelligence is basically fixed, each person has his/her allotted share, and there is little he/she can do to increase that allotment. In contrast, other individuals appear to be incremental theorists: they believe intelligence is modifiable through increased effort. These implicit theories have been related to a variety of achievement-related cognitions and behaviors. For example, students who subscribe to the incremental theory of intelligence prefer challenging tasks so that they can increase their ability as well as achieve greater mastery. Entity theorists, on the other hand, often avoid challenge because their primary concern is with the adequacy of their presumably fixed ability and what failure might convey about that ability.

It should be evident that the negative motivational consequences of an entity theory of intelligence resemble the

vulnerability to stereotype threat discussed above. Recognizing this overlap, a number of social psychologists have designed interventions for stereotype-threatened students that focus on changing beliefs about intelligence (Good *et al.*, 2003). In these interventions, African-American college students and middle school students, who were trained to view intelligence as malleable rather than fixed, showed significant improvements in their academic performance when compared with peers who did not receive the intervention.

Coping with discrimination

Like people who hold entity theories of intelligence, individuals of color who believe they are treated unfairly because of their race or ethnicity can lose confidence in themselves and in their ability to be self-efficacious. Perceived discrimination can occur in almost any arena. It can be blatant, intended, and obvious; or subtle, unintended, and not easy to detect. Some researchers have used the term microaggressions to capture a particularly subtle but pernicious kind of degradation many people of color encounter on an almost daily basis (Solórzano, 2000). Examples of microaggressions include being ignored or overlooked while waiting in line, being suspected of cheating because one has received a good grade on a test, being followed or observed while in public places, or being mistaken for someone who serves others.

Among the most prevalent kinds of unfair treatment reported by ethnic minority youth are those that take place in school. In their study with high school students, Fisher *et al.* (2000) created an educational discrimination index that included such items as “you were discouraged from joining an advanced level class,” “you were given a lower grade than you deserved,” and “you were wrongly disciplined or given after-school detention.” African-American and Latino students were significantly more likely to endorse these items than their white classmates. Such experiences have been linked to decreased perceptions of mastery among early adolescents of color and increased negative attitudes about school. Perceived discrimination can lead to mistrust of teachers, and to the general belief that the school rules and policies are unfair. A number of studies now document that personal experiences with discrimination in combination with racial mistrust can contribute to academic disengagement and other problem behaviors at school. The best antidote to these motivational challenges seems to be a combination of strong racial/ethnic identity and parental socialization about race. For example, when African-American youth report that their race is central to their self-concept and that parental messages communicate both racial pride and vigilance against racism, then the relations between experience with discrimination and academic disengagement are weakened (e.g., Neblett *et al.*, 2008).

To summarize, in this section we have focused on how beliefs about ability can be challenged for ethnic minority

youth. Such youth must cope with stereotypes linking race to low or high intelligence and with the widely shared perception that intelligence is a fixed entity that cannot be changed. These youth also confront discrimination, both overt and subtle, that can undermine achievement strivings. Recognizing race or ethnicity as central to one’s identity and receiving parental messages that instill pride and prepare one for racism can buffer some of the negative consequences of discrimination.

Beliefs About Effort: I Can, but Do I Want To?

We now turn to the second big construct in motivation research – achievement values. Unlike expectancy, which is largely centered around beliefs about ability, values have to do with desires and preferences and are more concerned with the perceived importance, attractiveness, or usefulness of achievement activities. Like expectancies, values have motivational significance because they guide thoughts, feelings, and behavior. For example, what we judge to be important, attractive, or useful can influence the activities we choose, how we evaluate other people and events, and our worldviews. Values are also rooted in the moral constructs of ‘ought’ and ‘should’ as illustrated by the belief that one should try hard in school regardless of one’s perceived abilities. Particularly during adolescence, effort in school is sometimes compromised because of the tension between the need to achieve and the need to belong, both of which are fundamental human motives. James Coleman, writing in *The Adolescent Society*, reminded us nearly 50 years ago that adolescents care more about being popular than about being perceived as smart and they even downplay effort in order to be liked by popular peers.

Oppositional Identity

The tension between having friends and being a high achiever can play out in particular ways for youth of color, African-American youth in particular, in the phenomenon known as oppositional identity. Oppositional identity is a construct that has emerged from anthropological analyses, which take into account the historical circumstances that have shaped the experiences of ethnic minorities in this country. African Americans are what anthropologist John Ogbu calls an involuntary minority – that is, a group that has become part of the American fabric not by choice, but as a result of slavery, conquest, or colonization (Fordham and Ogbu, 1986). One consequence of this history is that the acceptance of mainstream values of hard work and school success may be perceived as threatening to one’s social identity. Particularly during adolescence, African-American youngsters

may adopt oppositional identities whereby they show a relative indifference, or even disdain toward achievement behaviors that are valued by the larger society. Fordham and Ogbu (1986) coined the term 'acting white' to describe African-American high school students' perceptions of their same-race peers who work hard to do well in school. Case studies of youth of Mexican-descent and Asian-American youth (in part reacting to the model minority stereotype) document similar findings. While scholars disagree about the level of empirical support for the acting white construct (e.g., Tyson *et al.*, 2005), there is more consensus among researchers that many ethnic minority high-achieving adolescents experience a particular kind of conflict between their achievement strivings and their desire to be accepted by the general peer group.

The discourse surrounding oppositional identity during adolescence has remained lively, in part at least because it provides a motivational explanation for the achievement gap between black and white students. One would be hard pressed to find an article on academic motivation in African-American adolescents in the last 10 years that does not explicitly or implicitly make reference to oppositional identity. This construct has also been linked to other motivational phenomena discussed earlier in this article, such as stereotype threat and disidentification, as a way to fully capture the academic challenges that African-American students face (Steele, 1997). Note that the argument is not that talented black students, who have a strong racial identity, will devalue doing well in school. By every indication, a strong racial identity is a buffer that promotes academic engagement. Rather, the question is whether some talented students of color experience more tension as they negotiate their identities both as students and as members of their racial/ethnic group. The evidence suggests that, they do, indeed.

Characteristics of Valued Peers

Survey methods that directly ask students of color whether they value effort in school have sometimes yielded disappointing findings, in part because of the social desirability built into such questions (almost everyone agrees that effort is important). Recent motivation research has offered innovative ways that are less reactive to measure the degree to which students of color value working hard and doing well in school. For example, Taylor and Graham (2007) used peer-nomination procedures in which they asked elementary and middle school African-American and Latino youth to list the names of classmates whom they admire, respect, and desire to emulate. The rationale for these questions was that researchers can identify the characteristics of classmates whom a student admires, respects, and wants to be like, and this can tell us something about the characteristics that the student values. The researchers were particularly interested in how nominations as someone that the

nominator admired, respected, and wanted to be liked varied as a function of the achievement level of the person nominated. Taylor and Graham reported that girls of all grade levels overwhelmingly nominated high achievers as classmates they most valued, whereas boys were less likely to nominate high-achieving peers as they approached middle school. By the time African-American and Latino boys reached adolescence, they appeared to look more admiringly upon less academically inclined peers.

Such findings highlight the particular motivational challenges of ethnic minority males, a gender pattern that has not been adequately acknowledged. In mainstream gender research on motivation, a dominant theme is the heightened vulnerability of girls to motivational deficits. Some argue that gender role socialization and cultural stereotypes about women and achievement lead many girls to question their academic competence more, particularly in math to display more maladaptive reactions to failure, and to perceive more barriers to success. We believe gender analyses in motivation research may need to be reframed because it is evident that ethnic minority males (i.e., African American and Latino) are faring more poorly than females (e.g., Way and Chu, 2004). The ethnicity by gender differences increase over the school years and are particularly apparent when the measures are so-called markers of adolescent success (i.e., high school graduation) and young adult mobility (i.e., enrolment in and completion of college). Harsh disciplinary practices such as from school suspension also fall disproportionately on African-American males. These findings indicate that ethnic minority boys, more than girls, must cope in unique ways with the dual stressors of academic challenge and stereotypes about their group, and that these stressors can undermine the endorsement of achievement values.

Self-Affirmation

Self-affirmation theorists have examined how an opportunity to express one's core personal values can have positive motivational consequences. In one novel intervention, Cohen *et al.* (2009) had African-American and white middle school students choose from a set of values (including those related to achievement) the two or three that were the most important according to them and to write a paragraph about why they had selected those values. Students in the control condition wrote about the values that were least important to them. The rationale for the intervention was that reflecting on personally important values can be self-affirming and can therefore buffer some of the challenges associated with ethnic minority status. This simple intervention yielded powerful effects. African-American students in the values affirmation condition attained higher grades over the course of the school year and were actually able to reduce the achievement gap between black and white students.

Conclusions

This discussion of motivation in ethnic minority youth has focused on the challenges to expectancy and value, due in part to the structural barriers that often accompany minority status in the United States. However, it is also important to acknowledge the resilience in youth of color and their successful efforts to overcome those challenges. There are promising motivation interventions, such as those related to beliefs about intelligence and self-affirmation, which foster achievement strivings among youth at risk for school failure, just as personal characteristics such as strong racial/ethnic identity and family characteristics such as parental socialization about race, also function as protective factors.

At the institutional level, in light of challenges to perceptions of 'I can,' teachers need to be especially vigilant about not employing instructional practices that can indirectly and unintentionally communicate low-ability messages. For example, it has been documented that undifferentiated praise for success at easy tasks, unsolicited offers of help, and too much sympathy following failure can lead students to attribute their academic setbacks to low ability (Graham, 1991). Furthermore, altering pedagogical practices to be more effort oriented rather than ability oriented can have an immediate impact on students' motivation, even among those who are highly identified with the achievement domain. Cohen *et al.* (1999) found that African-American college students displayed more subsequent task motivation when feedback about poor performance was accompanied by criticism and communicated high expectations than when the same criticism was accompanied by general praise as a buffer. Such wise feedback can shift the attribution for failure away from low ability and toward those factors, such as lack of effort, that are under volitional control.

There is a need for alternative methods of assessing educational progress that go beyond traditional standardized testing. As revealed in research on stereotype threat, the anxiety elicited by fear of confirming negative stereotypes about race and intelligence can deplete the cognitive resources needed to do well on standardized tests. These same concerns can be raised about high-stakes testing in general, which has had deleterious effects on children of color.

Reconciling equity, assessment, and the necessity of meeting the needs of diverse learners is a complex problem with no simple solutions. As long as assessment procedures rely mainly on established tests of cognitive abilities, there will be racial/ethnic imbalances. Reform-minded educators need to develop creative ways that not only meet standards of scientific rigor and accountability but also support talented youth of all racial/ethnic groups and of all ages. A motivational approach, which focuses on the why of achievement-related behavior rather than behavior itself, can offer fresh insights into the educational challenges faced by students of color.

Bibliography

- Cohen, G., Garcia, J., Purdie-Vaughns, V., Apfel, N., and Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the achievement gap. *Science* **324**, 400–403.
- Cohen, G., Steele, C., and Ross, L. (1999). The mentor's dilemma: Providing critical feedback across the racial divide. *Personality and Social Psychology Bulletin* **25**, 1302–1318.
- Dweck, C. (1999). *Self-Theories: Their Role in Motivation, Personality, and Development*. Philadelphia, PA: Psychology Press.
- Fisher, C., Wallace, S., and Fenton, R. (2000). Discrimination distress during adolescence. *Journal of Youth and Adolescence* **29**, 679–694.
- Fordham, S. and Ogbu, J. (1986). Black students' school success: Coping with the "burden of 'acting White'." *Urban Review* **18**, 176–206.
- Good, C., Aronson, J., and Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology* **24**, 645–662.
- Graham, S. (1991). Communicating low ability in the classroom: Bad things good teachers sometimes do. In Graham, S. and Folkes, V. (eds.) *Attribution Theory: Applications to Achievement, Mental Health, and Interpersonal Conflict*, pp 17–36. Hillsdale, NJ: Erlbaum.
- Kao, G. (2000). Group images and possible selves among adolescents: Linking stereotypes to expectations by race and ethnicity. *Sociological Forum* **15**, 407–431.
- Lee, S. (1994). Behind the model-minority stereotype: Voices of high- and low-achieving Asian American students. *Anthropology & Education Quarterly* **25**, 413–429.
- Neblett, E. W., Jr., White, R. L., Ford, K. R., *et al.* (2008). Patterns of racial socialization and psychological adjustment: Can parental communications about race reduce the impact of racial discrimination? *Journal of Research on Adolescence* **18**(3), 477–515.
- Solórzano, D. (2000). Critical race theory, racial microaggressions, and campus racial climate: The experiences of African American college students. *Journal of Negro Education* **69**, 60–73.
- Steele, C. (1997). A threat in the air: How stereotypes shape the intellectual identities of women and African Americans. *American Psychologist* **52**, 613–629.
- Taylor, A. Z. and Graham, S. (2007). The development of relations between achievement values and perceptions of barriers in African American and Latino youth. *Journal of Educational Psychology* **99**, 52–64.
- Tyson, K., Darity, W., and Castellino, D. (2005). It's not "a black thing": Understanding the burden of acting white and other dilemmas of high achievement. *American Sociological Review* **70**, 582–605.
- Way, N. and Chu, J. (eds.) (2004). *Adolescent Boys: Exploring Diverse Cultures of Boyhood*. New York: New York University Press.

Further Reading

- Aronson, J. and Steele, C. (2005). Stereotypes and the fragility of academic competence, motivation, and self-concept. In Elliott, A. and Dweck, C. (eds.) *Handbook of Competence and Motivation*, pp 436–456. New York: Guilford Press.
- Bergin, D. and Cooks, H. (2002). High school students of color talk about accusations of "acting White." *Urban Review* **34**, 113–134.
- Eccles, J., Wigfield, A., and Schiefele, U. (1998). Motivation to succeed. Eisenberg, N. (ed.) *Handbook of Child Psychology*, 5th edn, vol. 3, pp 1017–1095. New York: Wiley.
- Eccleston, C. and Major, B. (2008). Responding to self-threats: Understanding stigma in the academic domain. In Hudley, C. and Gottfried, A. (eds.) *Academic Motivation and the Culture of School in Childhood and Adolescence*, pp 165–186. New York: Oxford University Press.
- Graham, S. and Hudley, C. (2005). Race and ethnicity in the study of motivation and competence. In Dweck, C. and Elliot, A. (eds.) *Handbook of Motivation and Competence*, pp 392–413. New York: Guilford Press.

- Graham, S. and Weiner, B. (1996). Principles and theories of motivation. In Berliner, D. and Calfee, R. (eds.) *Handbook of Educational Psychology*, pp 63–84. New York: MacMillan.
- Ogbu, J. (2003). *Black Students in an Affluent Suburb: A Study of Academic Disengagement*. Mahwah, NJ: Erlbaum.
- Osborne, J. and Walker, C. (2006). Stereotype threat, identification with academics, and withdrawal from school: Why the most successful students of color might be most likely to withdraw. *Educational Psychology* **26**, 563–577.
- Schuman, H., Steeh, C., Bobo, L., and Krysan, M. (1997). *Racial Attitudes in America: Trends and Interpretations*. Cambridge: Harvard University Press.
- Sellers, R. and Shelton, J. (2003). The role of racial identity in perceived racial discrimination. *Journal of Personality and Social Psychology* **84**, 1079–1092.
- Skiba, R. J. (2001). When is disproportionality discrimination? The overrepresentation of black students in school suspension. In Ayers, W., Dohrn, B., and Ayers, R. (eds.) *Zero Tolerance: Resisting the Drive for Punishment in Our Schools*, pp 176–187. New York: New Press.
- Spencer, M., Noll, E., Stoltzfus, J., and Harpalani, V. (2001). Identity and school adjustment: Revisiting the “acting White” assumption. *Educational Psychologist* **36**, 21–30.

Affirmative Action and Higher Education in Brazil

J T dos Santos and D M Queiroz, Federal University of Bahia, Salvador, Brazil

© 2010 Elsevier Ltd. All rights reserved.

In Brazil, over the last 5 years, 51% of the public state universities and 41.5% of the public federal universities have adopted some form of affirmative action. The majority of these institutions included some seat reservations in their courses. For this reason, the seat reservations are popularly known as quota systems for students who came from the public education system, blacks (blacks and mestizos), indigenous people, and people with special disability needs. Seven public universities adopted a system of points (bonus) without indicating, like the other institutions, a percentage of seat reservations.

This article examines the implementation of quota systems in the Federal University of Bahia (UFBA). Bahia is the state with the highest percentage of black people in Brazil, where more than 70% of the population are self-declared as black or mestizo. The quota system in the UFBA indicates reservation of 43% of the vacancies in the college for students from the public education system and black students. Eighty-five percent of those vacancies are reserved for black students from public schools, and 15% are reserved for white students from public schools.

If public school students do not reach the minimum grade to obtain a seat in the university through the system of quotas, the vacancies will be filled by black students from private schools. The benefit of the quota system to private-school black students is related to the perception that the number of black students in public universities is small, mainly in the courses considered as prestigious, such as medicine, engineering, dentistry, law, and architecture.

Initially, the article compares the results of two surveys conducted by the DataFolha Institute in 2005 and 2006, which reflect the changes in perception of Brazilian people regarding affirmative action for blacks. Second, the authors present the debate on quota systems between teachers in the UFBA. The article analyzes the performance of students who were admitted by the quota system during 2 years. The goal is to show how affirmative action affected the access of black students who came from public schools and its impact on Brazilian universities. At the same time, it is argued that affirmative action must be understood in different contexts, principally in relation to other countries such as the United States and India.

In the 23rd July 2006 edition of the *Folha de São Paulo* newspaper, the Instituto DataFolha (an institute of public opinion on various issues) published the result of a survey on the quota system. The scope of the research was about

6264 people, over 16 years old. The result of the research indicates that 65% of the Brazilian population do support the adoption of quotas for black people in Brazilian universities. The supporting percentage grows to 87% when the question refers to the adoption of quotas for poor people.

Such results point to a certain dynamics in Brazilian society concerning the perception of public policies. Toward this discussion, we propose to compare both the surveys. The first one refers to the results of the survey cited above; the second to the survey denominated *Racismo Cordial* (gentle racism), conducted in 1995, by the same institute (i.e., the DataFolha) at a historical moment when there was no adoption or discussion of the quota system in Brazilian public universities. The comparison between these two surveys shows the changes that occurred in the perception of the Brazilian people about public policies targeting underrepresented sectors of the population.

In the 1995 survey, a question was formulated on reservation of seats for blacks in the field of education and the labor market. The results showed that 49% of the interviewees disagreed and 48% agreed with the quota proposal. Among men, the percentage of blacks and mestizos who disagreed was different by one point percentage (47% and 48%). The proportion for whites was much higher – 53%. Among women, the rejection percentage was also high and it varied in racial terms, more than among men (40% among blacks, 46% among mestizos, and 58% among whites).

Education and family income are also variables that are important for the comparison to the 2006 survey. The disagreement toward the seat reservation increased with the interviewees' level of education, especially among those who had completed a high school level and those who had completed a degree. Among the latter, 67% of interviewees who were self-designated as white were against the proposal. Among blacks and mestizos, the proportion was even higher – 68% and 76%. When it came to the issue of family income, the rejection followed the pattern observed in the issue of education. For the individuals with higher income, the support was lower. Finally, among those who received more than minimum salaries – a value higher than \$2000 – the index of rejection was about 55% among blacks, 60% among the mestizos, and 58% among the whites (see *Racismo Cordial*, Turra and Venturi, 1995).

The results found in the 2006 survey reveal a shift in the perception of racial relations in the country: 65% of

the interviewees support the proposal. The support of quotas is observed in each and every racial group, and in the higher levels of education and income. Among those who reject the proposal, 55% have a university degree and 57% have a family income higher than ten minimum salaries. Although in a reduced proportion, the profile of those who show some resistance follows the same pattern observed in the research on *Racismo Cordial* (gentle racism): the higher the level of education and income, the higher is the rejection. In all of the racial segments, the support reaches a level over 60%; that is, the levels of rejection are low in all groups: 32% among whites and 23% among blacks and mestizos. It is important to highlight that in this dynamics of racial relations, the elements of a national and international conjuncture are present; and since the last decade, it has presented new demands, which come from the academic environment or from social movements, especially from the black awareness movement.

Since the 1990s, when proposals of affirmative action appeared on the Brazilian scene, there has been a tendency to update the history of privileges and perversions. The current rhetoric, a result of the end of military dictatorship, is necessary to establish citizenship, especially of the historically marginalized blacks, indigenous, and the socially excluded. The same rhetoric has a common denominator which appears in the media, for instance, through researches revealing the racial hierarchy among the Brazilian people. It is what is noticed when, starting from certain consensus, there is expressive racism in Brazil, not mere social prejudice of class. Some reactions even appeared against the proposals to create laws that would compensate for the countless debts owed to the Afro-Brazilian peoples.

It is important to consider the dilemma of race relations in Brazil and the United States and the Brazilian myth of racial democracy. In fact, the myth implies that the three races in Brazil – the black, the native, and the white – will live without conflicts.

Until the 1950s, racial democracy was thought of less as an ideology that is part of Brazilian society than as academic studies and official speeches; in fact this was true even in the 1960s for Afro-Brazilian and North American black intellectuals. In the 1970s, the reference takes on other colors and names: the racial democracy is referred to as a myth and identified as merely an ideology. In that process of denying racial democracy, the analytical rupture did not happen, but reflections about the racial inequalities in Brazilian society influenced political sense.

Since the 1980s, a growing number of journalists, historians, and North American social scientists have denounced the discrimination they claim to have suffered in Brazil. This was very curious since these individuals, who knew about the existence of discrimination against the blacks in the country, expressed surprise when confronted

with racism in Brazilian society. The dialogs between Brazil and the United States, therefore, will continue to be a dilemma. This is a discussion that will continue beyond the twentieth century. The official qualifying system in Brazil includes five categories for the definition of color: black, brown, white, yellow, and indigenous. The brown term refers to an individual who could be identified as mestizo in other contexts (the mixture of black with white). Black is a category that, if synonymous with black in the study of the racial inequalities, includes the sum of black and brown. It is necessary to point out that the use of the black category reflects an influence of the sociology of racial inequalities in the United States during the 1970s. (See Hasenbalg's studies, Hasenbalg (1997); on the use of the categories in Brazil, see Harris *et al.* (1993), and Telles (1995); and see also Fry (1995); Hanchard (1994); and Guimarães (1995).)

Conversely, the unequal access to Brazilian universities is one phenomenon that, only recently has attracted the attention of researchers. By the end of the 1980s, Ribeiro (1987) demonstrated that the majority of the students in a Brazilian universities were directly related to a certain family and educational heritage, mostly in the middle classes. Ivonne Maggie, at the time, pointed to this characteristic when she stated that poor students did not have access to this environment; she pointed out that to enter the university, it was necessary to have a certain level of income and cultural background (see Maggie). The research on the UFBA confirmed this analysis, revealing that traditionally, the students who had access to its courses were those whose parents had higher levels of scholarship and income, that is to say, a higher status (see Brito and Carvalho, 1978).

Until the end of the last decade, there was little information about the participation of several racial segments of the Brazilian population in higher education. From a survey done in the UFBA in 1997 and in other federal public universities in 2000, it was possible to have a first mapping of the racial inequalities that were present in universities considered to be prestigious in certain parts of the country (see **Tables 1, 2, and 3**) (see Queiroz, 2002).

These studies revealed that blacks and whites, especially those who come from the public educational system, were unequally represented in courses of great competition and high prestige, such as medicine, dentistry, law, architecture, electric engineering, computer sciences, psychology, and business administration.

By the end of the 1990s, besides a growing debate about the inclusion of black students in public universities, there were, in some federal institutions, actions surrounding affirmative action for black individuals. The Ministry of Justice determined that the directive and superior assessor posts would have a quota of 20% for Afro-descendent people, 20% for women, and 5% for physically disabled people. With resources from the

Table 1 Percentage of students in federal universities by race

Color	UFRJ	UFPR	UFMA	UFBA	UnB
Whites	76.8	86.5	47.0	50.8	63.7
Mulattos	17.1	7.7	32.4	34.6	29.8
Black	3.2	0.9	10.4	8.0	2.5
Yellow	1.6	4.1	5.9	3.0	2.9
Indigenous	1.3	0.8	4.3	3.6	1.1
Total	100.0	100.0	100.0	100.0	100.0

From Queiroz, D. M. (2002). *O negro na universidade*, Série Novos Toques: Salvador.

Table 2 Percentage of blacks in Brazilian states and the universities

States	Population of state	University	Population in university
Rio de Janeiro	38.2	UFRJ	20.3
Paraná	22.4	UFPR	8.6
Maranhão	75.1	UFMA	42.8
Bahia	77.5	UFBA	42.6
Distrito Federal	53.6	UnB	32.3

From Queiroz, D. M. (2002). *O negro na universidade*, Série Novos Toques: Salvador.

Table 3 Percentage of whites in the Brazilian states and the universities

	Population of state	University	Population in university
Rio de Janeiro	61.7	UFRJ	76.8
Paraná	76.2	UFPR	86.5
Maranhão	24.8	UFMA	47.0
Bahia	22.1	UFBA	50.8
Distrito Federal	45.9	UnB	63.7

From Queiroz, D. M. (2002). *O negro na universidade*, Série Novos Toques: Salvador.

Inter-American Bank of Development, the Ministry of Education has created a program called Diversity Program at University, which supports the preparatory courses for the black and indigenous people. All these actions initiated in a context marked by internal demands, which came from black activists and entities, and external demands from international organizations.

In August 2001, the 3rd International Conference on Racism, Xenophobia, and other forms of correlated intolerances, was held in Durban, South Africa. Promoted by the United Nations (UN), the conference had intense participation by entities from the Brazilian black movement, and in the months that preceded the conference, there were intense debates within the social movements about the formulation of the claims and the definitions toward the concepts of reparation.

In Brazil, the event that took place was a kind of follow-up of another event that occurred in Lagos in December 1990, when a group of intellectuals, government representatives, and leaders of other entities from all over the African continent, the United States, Great Britain, and the Caribbean, gathered to discuss the historical, legal, and moral dimensions of what reparation should be. In this meeting, an international committee was created for reparation, and in 1992, the African Union (AU), with the same objective, created a group made up of specialists and an executive secretary (see Araújo, 2001). In this period, the discussion on reparation for the blacks in Brazil reflected the proposals and the projects leading toward an indemnity to those of African descent in the country.

The Quotas in the Federal University of Bahia: Some Ethnographic Notes

Since 1998, some proposals have been introduced by the UFBA on the politics of affirmative action. In 2002, the Rectory constituted a task force for the elaboration of a proposal on strategies for social inclusion. Representatives of professors, staff, students, and the black movement participated in the task force. After a year of intense discussions, a report on the issue was presented and approved by the UFBA. In a period of 10 years, the objective is to increase the number of students from public schools, blacks and browns, and indigenous descendants in all of the courses because currently most of the students of the university come from the private system.

The proposal approved for UFBA in 2004 was marked by an initial absence of debate among professors, civil servants, and students. The task force attempted to hold several debates but was not successful. However, before the proposal could be approved by UFBA, there was a reaction via e-mail by a professor opposed to the system of quotas:

Dear friends,
the logic of the argument is impeccable. The discrimination really constitutes an element access trouble to the Higher Education. But why only solve the problem of the afro-descending ones? For the same logic, and since the socioeconomic position doesn't work as cut line, I propose immediately that others discriminated be mediated equally with: quotas of 51% for women; quotas of 4.7% to natives' descendants; quotas of 0.3% to people that suffered violence or sexual violence in the childhood; quotas of 0.8% to blind men or people with serious visual deficiencies; quotas of 2.1% to bearers of the Syndrome of Down; quotas (to be defined) for transvestites and transsexual; quotas of 1.2% to paraplegics or hemiplegics;

quotes of 0.1% to stutterers or affected people for other difficulties similar; quotas of 4.9% to vegetarians. The list certainly is not exhausting. But certainly it can be completed in the discussion process. Or will it be that it remains some common sense? Greetings,

This message provoked a great debate in the Web site of the university: The debate was marked by several opinions. Most of the professors were against the adoption of quotas for blacks and came up with the following arguments:

1. Brazil is a mestizo country;
2. the descent conception in Brazil does not resemble the United States's one-drop rule;
3. the race concept is no longer applied in the sciences;
4. the difficulty of access for blacks into universities is based on the class condition ("they are poor");
5. merit is the mark of access to the system of higher education;
6. the students in the quota system will find it difficult to study in the universities because the public schools are not good;
7. the system of quotas will result in a discriminatory system in the university;
8. quotas would be political opportunism and a demagoguery; and
9. the quota system is an imperialistic foreign proposal.

The teachers favorable to the system of quotas argued that

1. it would be a form of giving access to the black and indigenous populations to the university; and
2. it would change the color of the university in the courses considered prestigious.

An argument of a feminist was provocative:

Dear friends,

For terms certainty that it is not just treated of a defense of color privileges, it would be important that those emitting opinion against or in favor of the quotas identify themselves in color terms. I am white phenotypically and totally in favor of the politics of affirmative actions, among which is included our politics of social quotas now - of affirmative actions for no white. In fact, I should say that when reading the friends' positions (up to now, all men...) contrary the those politics, I am fearing for our conquest - the women's conquest - in relation to minimum quotas in the parties and unions. Until I can imagine the argument type that it will get up when us, women, we enter with our demands of 30% for women and black in the State legislatures, National Congress, etc. After all, the power was always in white men's hands; it is not easy to give up millenarian privileges.

The exercise of ethnography is a tradition in anthropology. How can we build an ethnography of colleagues?

That seems to be a fundamental question in attempting to understand the reactions of more than 100 professors. The anthropological exercise became difficult because the native point of view was too close for comfort. Those who were to be the subject of the study were in the same university and some were in the same departments.

The positions contrary to and in favour of the quota system in the universities do not imply there is a distinction in the scientific field, as it can be observed in Bourdieu (2000) when he argues about the meaning of scientific competition. It is not publish or perish. The scientific field is known for its constant tension, monopoly of ideas, maintenance of consolidated speeches, and disputes over current theories. It is fundamental to understand that these intellectual practices are really social in nature.

The Impact of Quota Policy

If one is to observe the origins from the schooling, the system of entering the universities with reservation has caused a great revolution in UFBA because it has provoked their most prestigious course students who come from public schools and who, at the same time, have been excluded from this space. The participation of students from the public schools, which was less than 27% in degree courses, such as medicine, architecture, law, media, dentistry, computing, civil engineering, and electric engineering, has increased extraordinarily, reaching more than the 43% of the vacancies reserved for them in the quota system.

The participation of students from public schools, at UFBA, which was around 38% before the introduction of the quota system, increased to 51% in 2005. Although it can be verified that in 2006 there was a reduction of this figure to 44.9 %, the level has stayed above the one that is intended by the quota system. In courses such as computing, civil engineering, electric engineering, and geophysics, for instance, there was a reduction of students from public schools compared to 2005, although, the quota remained the same. It is noteworthy to remember that the quota system adopted by UFBA does not imply a necessity to fill all the vacancies for, it relies heavily on student performance (**Table 4**).

Confirming the results of the previous research (Queiroz, 2003), these data show that the Brazilian public university is an extremely selective space, and that only through the adoption of specific policies of access, such as the affirmative actions directed to students from public schools, is it possible to assure them some chance to gain access. As shown in the data collected, it does not mean changes in the demands of the examination to get into the university - the vestibular - or the triviality of the teaching system as some may argue or suppose.

Table 4 Percentage of the distribution of the students selected according to the type of fundamental school attended (2003–06)

<i>School</i>	<i>Year</i>			
	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>
Public	38.3	33.8	51.0	44.9
Private	61.7	66.2	49.0	55.1
Total	100	100	100	100

From Serviço de Seleção e Orientação (SSOA) – UFBA.

Academic Performance of the Quota Students and Nonquota Students in the Vestibular

The data for student performance in the vestibular examination are other important indicators, not only for the evaluation of the impact of the reservation policy, but also to dissolve pessimists' expectations about the adopted measures. The information about the performance of the quota and nonquota students shows that the gap between the averages of both groups is not so high expressive in the majority of the courses of higher social prestige, as can be observed in **Table 5**. The method of evaluation for the disciplines is measured by marks between 0 (zero) and 10 (ten). In medicine, for instance, considered to be one of the toughest to get into, at UFBA, the gap between the two groups does not even reach 1.0 (one) point. The same occurs in the law course, also one of the most targeted by students. The largest gaps observed were in the electric engineering course (1.7) and the mechanic engineering (1.4), both in 2006.

For a better understanding of the difference between the averages of the quota and nonquota students, **Table 6** shows student performance in the 2005 vestibular, comparing the averages of the first and of the last student qualified in each of these groups, per chosen course. As previously observed, for the set of courses examined previously, in all courses at UFBA, a great difference is not observed between the average of the quota and nonquota students in the two positions evaluated, that is, in the first and in the last position for each group.

Performance in the Courses

The adoption of the quota system by public universities over the last 2 years allows analysis of the impact of the new system to enter the university. This fact is of extreme significance because there had only been data about students' entrance and performance in the vestibular. The hypothesis about course performance was based on facts relating to the other system, which was a classificatory one.

At UFBA, in 11 of the 18 courses of higher dispute (i.e., 61% of them), the quota students had a coefficient equal

Table 5 Average performance of quota and nonquota students in most disputed courses with social prestige (2005–06)

<i>Course</i>	<i>Quota students average grades</i>		<i>Nonquota students average grades</i>		<i>Difference between quota students and nonquota students</i>	
	<i>2005</i>	<i>2006</i>	<i>2005</i>	<i>2006</i>	<i>2005</i>	<i>2006</i>
Medicine	6.7	6.7	7.5	7.4	0.8	0.7
Law studies	6.2	6.2	7.0	7.0	0.8	0.8
Dentistry	5.2	4.9	6.2	6.2	1.0	1.3
Business	5.5	5.1	6.3	6.2	0.8	1.1
Computing sciences	5.8	5.6	6.6	6.7	0.8	1.1
Electric engineering	6.2	5.7	7.1	7.3	0.9	1.7
Psychology	5.7	5.6	6.3	6.5	0.6	0.9
Civil engineering	5.2	4.9	5.8	6.0	0.6	1.1
Mechanic engineering	5.5	5.4	6.5	6.8	1.0	1.4
Architecture	4.9	4.8	6.1	6.1	1.2	1.3
Media and journalism	6.1	5.7	6.8	6.9	0.7	1.2
Media and cultural production	5.4	5.4	6.2	6.2	0.8	0.8

From Serviço de Seleção e Orientação (SSOA) – UFBA.

or better than those who were not from the quota system (**Table 7**). Contrary to the expectations of those people who were resistant to the implementation of the quota system for fear of lowering the standard in education, the quota students' performance in the examination revealed very satisfactory results in courses from diverse areas.

This polarization is based on fragile arguments, principally the data concerning the student's performance in the vestibular and in the courses. In a recent paper, Barreto (2007) points out not only the good level of performance by students who receive scholarships in these programs but also their distinct perspective. Black students who entered the university through this system tend to positively affirm themselves as black as well as individuals (e.g., involvement in the black movement or non-black association, students' union, and political parties). Therefore, it is absolutely worth questioning: What is the meaning of racialization of Brazilian society, so feared and propelled by intellectuals against the quota system, if not the defense of points of view that translates the permanence of certain privileges and the exercise of power?

The result of this analysis characterized the quota policy, adopted by UFBA, as a very meaningful instrument to promote the democratization of access to higher education. The reservation of seats has shown to be not only capable of broadening the contingent of students who come from public schools, but it also permitted

Table 6 The averages for the performance in the vestibular of quota and nonquota students (2005)

<i>Course</i>	<i>First place quota student</i>	<i>First place nonquota student</i>	<i>Least place quota-student</i>	<i>Least place nonquota student</i>
Architecture	6.8	8.1	4.4	5.6
Computer sciences	7	7.9	4.7	6.1
Civil engineering	7.5	7.7	4.6	5.3
Mines engineering	5.9	6.6	4.6	5.1
Electric engineering	7.4	8.1	4.9	6.7
Mechanic engineering	6.6	8.1	4.7	6.2
Chemistry engineering	7.1	8.2	5.1	6
Environmental and sanitary engineering	6.1	7.6	4.7	6.1
Statistics	5.5	6.9	4.5	4.5
Physics	6.3	7.2	4.7	5.2
Geophysics	6	6.4	5.2	5.5
Geology	5.3	5.9	4.3	4.4
Mathematics	7	7.6	5.1	5.4
Chemistry	7	6.8	5	5.1
Agronomy	5.6	6.4	4.2	4.6
Biological sciences	6.5	8	4.9	5.6
Natural sciences	5.3	6.1	4.3	4.8
Medicine	7.8	8.3	4.7	7.3
Veterinary	6.2	7	4.7	5.1
Nutrition	5.7	6.6	4.6	5.3
Oceanography	6.2	7.3	5.1	5.9
Dentistry	6.3	7.3	4.2	5.8
Business	6.5	7.2	5	5.8
Archivology	5.6	5.5	4.5	4.5
Accounting	5.9	6.4	4.6	5.1
Economics	6.5	6.8	4.8	5.6
Social sciences	6.6	6.7	4.7	5.5
Media – journalism	7.3	7.3	5	6.4
Media – cultural production	6.7	7.1	5	5.8
Law studies	8.1	7.8	4.4	6.7
Physical education	5.7	6.1	4	5.1
Philosophy	6.1	6.5	4.5	5
Geography	6.1	6.2	4.5	5
History	6.5	7.2	4.5	5.8
Museology	6.1	6.3	4.6	5.1
Pedagogy	5.5	6.1	4.2	4.8
Psychology	6.5	6.8	4.4	6
Executive secretary	5.9	5.5	4.5	4.5
Language studies – vernacular	6.6	7.9	4.6	5.1
Language studies – vernacular and foreign language	7.1	6.4	4.6	5.2
Language studies – foreign languages	6.3	6.4	5	5.2
Drama – theater	5.9	5.5	5.1	5
Drama – direction	5.9	6.1	4.9	5.7
Drama – interpretation	5.7	6.2	5.3	5.7
Art	6.5	6.5	4.4	4.7
Decoration	5.8	6.1	5.1	4.9
Dance	6.1	6.2	4.7	5.4
Industrial design – visual rogrammer	6.4	6.4	4.9	5.5

From Serviço de Seleção e Orientação (SSOA) – UFBA.

students to enter the university to do courses of elevated competition so that they can have great social prestige. The racial segments (blacks and indigenous), who had shown the necessary performance, had earlier been systematically kept apart from this possibility, due to the reduced number of seats offered by the university as well as the classificatory system of their fulfillment.

Although it is recognized that the quota system has its own virtues, that its adoption is pertinent as a provisory policy within a context marked by the exclusion of a great number of social segments to the university, it is absolutely necessary not to lose sight of its limitations. The overall situation revealed by the present analysis signals an urgent educational policy

Table 7 Percentage the quota and nonquota student with an average between 5.1 and 10.0 in the courses of higher dispute in two semesters at UFBA in 2005

Course	Quota students	Nonquota students
Business administration	83.3	95.4
Architecture	85.6	81.3
Computing	66.6	53.7
Media – journalism	100.0	87.5
Media – cultural production	100.0	88.9
Law	95.2	88.9
Electric engineering	55.5	75.0
Mechanic engineering	75.0	100.0
Civil engineering	94.1	80.0
Pharmacy	92.3	82.3
Medicine	93.3	84.6
Veterinary	77.0	81.0
Nutrition	87.5	92.3
Oceanography	27.2	40.0
Dentistry	100.0	100.0
Psychology	77.8	100.0

that might be able to broaden the offer of higher education in the state.

The participation of students, who had their educational background in public schools, must not be blind to what such institutions may offer. The research on the reality of Brazilian public schools does not leave room for doubt about the long road ahead for the state to offer adequate basic education. On the other hand, we must highlight that if affirmative action in higher education is considered in central countries or in the peripheral ones (United States (see Bowen and Bok, 1998; Allen, 1988), United Kingdom, China, Macedonia, and South Africa) and with strong emphasis on the minorities, the Brazilian experience suggests affirmative action directed to groups of minorities (indigenous and people with special needs) as much as to groups of high percentage representation, such as blacks, because they are almost 50% of the Brazilian people.

Another important difference in relation to these countries is the fact that most universities that adopted the quota system focus on students from public schools. The Brazilian uniqueness is due to a reformulation in the educational system over the last three decades and the subsequent lack of investment from the government in the public system of education. There had been an

increasing access of students to higher education coming from the private system, the majority of them belonging to middle classes and self-declared as whites. Therefore, the challenge in the universities is how to combine affirmative action policies directed both at race as well as at class (poor ones).

Bibliography

- Allen, W. R. (1988). Black students in U.S. higher education: Toward improved access, adjustment, and achievement. *Urban Review* 20, 3.
- Araújo, U. C. D. (2001). Reparação moral, responsabilidade pública e direito à igualdade do cidadão negro no Brasil. In Sabóia, G. V. (ed.) (org.) *Anais de Seminários Regionais Preparatórios para a Conferência Mundial contra o Racismo, Discriminação Racial, Xenofobia e Intolerância correlatas*, pp 315–324. Brasília: Secretaria de Estado de Direitos Humanos, Ministério da Justiça.
- Barreto, P. C. D. S. (2007). Apoio à permanência de estudantes de escolas públicas e negros em universidades públicas brasileiras: as experiências dos projetos tutoria e Brasil AfroAtitude na Ufba. In Brandão, A. A. (ed.) *Cotas Raciais no Brasil: A primeira avaliação*, pp 145–162. Rio de Janeiro: André Augusto Brandão.
- Bourdieu, P. (2000). O campo científico. In Bourdieu, P. (ed.) *Grandes cientistas sociais*. São Paulo: Ed.Ática.
- Bowen, W. G. and Bok, D. (1998). *The Shape of the River: Long Term Consequences of Considering Race in College and University Admissions*. Princeton, NJ: Princeton University Press.
- Brito, L. N. and Carvalho, I. M. D. (1978). *Condições sócio-econômico dos estudantes da Universidade Federal da Bahia*. Salvador: CRH/UFBA.
- Fry, P. (1995). Why Brazil is different. *Times Literary Supplement* 8 December, 6–7.
- Guimarães, A. S. (1995). Racism and anti-racism in Brazil: A postmodern perspective. In Bowser, B. P. (ed.) *Racism and Anti-Racism in World Perspective*, pp 208–226. Thousand Oaks, CA: Sage.
- Hanchard, M. (1994). Black Cinderella? Race and the public sphere in Brazil. *Public Culture* 7, 165–185.
- Harris, M., Consorte, M., and Lang, J. (1993). “Who are the Whites”? Imposed census categories and the racial demography of Brazil. *Social Forces* 72(2), 451–462.
- Hasenbalg, C. (1979). *Discriminação e desigualdade racial no Brasil contemporâneo*. Rio de Janeiro: Edições Graal.
- Maggie, Y. Graduação e Pós-Graduação nas Ciências Humanas no Brasil: desafios e perspectivas (mimeo). IFCS – UFRJ. S/d.
- Queiroz, D. M. (2002). *O negro na universidade*. Série Novos Toques: Salvador.
- Queiroz, D. M. (2003). O ensino superior no Brasil e as ações afirmativas para negros. *Revista Universidade e Sociedade, ANDES, ano XII*, 29.
- Ribeiro, S. C. (1987). “A visão dos professores das IES hoje”. *Seminário vestibular hoje*. Brasília: MEC.
- Telles, E. (1995). *Who are the Morenas? Social Forces* 73(4), 1609–1611.
- Turra, C. and Venturi, G. (1995). *Racismo cordial: A mais completa análise sobre preconceito de cor no Brasil*. São Paulo: Ática.

Black in White: Black Students at White and Black Colleges

L Evans and J R Feagin, Texas A&M University, College Station, TX, USA

© 2010 Elsevier Ltd. All rights reserved.

Obtaining a degree from an institution of higher education is a major goal for a great many Americans. In the US economy and the global economy, a college degree has moved beyond the scope of a reasonable individual goal and of status attainment to one of economic necessity. However, in US society, obtaining an adequate education has been a reality for only some groups, while educational aspirations have historically, as well as presently, remained only an unfulfilled desire for others.

African Americans, as a group, have historically faced significant racial obstacles to gaining entry into major institutions of higher education. Moreover, after legal segregation in the 1970s, once most major educational institutions became receptive to African-American students, the racial environments often told a different story. Several social science disciplines have studied the consequences of blocked access to educational opportunities for African-American students and other students of color, and we draw on them here. We also extend this area of research by adding contemporary examples of black student experiences at institutions that have been historically white as well as those that have been historically black.

Consider these reasons for examining student experiences in these often racially hostile environments. We need to push for a more diverse and multiracial curriculum not only in the area in which the research we cite has taken place, but also throughout this country, especially as more countries across the globe are experiencing multiracial and multicultural population changes. Indeed, going beyond the scope of a multiracial curriculum means that colleges and universities must do a more adequate job of recruiting and retaining not only African-American students but also students of color generally.

This article examines the historical significance of race, racism, and racist ideology as well as the challenges confronting black students and other students of color in their pursuit of higher education. Next, we chronicle the educational challenges, psychological consequences, and the everyday realities of African-American students within these white and black institutions of higher education. Finally, we introduce the agonizing dilemma faced by the parents of these students: African-American parents routinely have to make the tough decision of where to send or encourage their child to attend school. Although this article primarily focuses on US institutions of higher education, the implications may be extended to provide an understanding of issues facing students of color in various other contexts. First, we focus on the history and role of

systemic racism and its meaning for African-American students in education.

Education and the Role of Systemic Racism

In order to understand the current realities of many students of color and their experiences in higher education, a look at the role of systemic racism is needed. From this perspective, racism is and has been both foundational and systemic; it involves a material, social, and ideological reality that is both historically and presently entrenched in major US institutions (Feagin, 2006). Because of the persistence of racism well into the twentieth century, African Americans were often prevented from obtaining an education, or at best were allowed access to a very limited education. For instance, the 1890 census states that of the nearly 10 million students enrolled in public common schools in the United States, less than 1% of those students were then colored (US Census, 1890). Through the practices of institutional racism, not only were African-American students and other students of color prevented from obtaining an education at any significance, but also those whites in power used this lack of schooling as a means of continuing the vicious cycle of employment and other racial discrimination at most societal levels.

Understanding the historical background is of importance in considering the connection between past racial atrocities and current discrimination realities. Discrimination in educational opportunities throughout US history has not been limited to specific regions or sections of the country. Indeed, many educational institutions have institutionalized a variety of discriminatory practices, North and South (Feagin *et al.*, 1996). It was not until well into the twentieth century that African Americans were allowed to attend historically white universities in any significant numbers. In some cases, African-American students were only allowed to attend schools that were separate and unequal. For instance, in the state of Texas, the first African-American student admitted to a historically white college did not enter until 1954 due to segregation laws in place, restricting undergraduate admission of African-American students at such colleges. Moreover, when historically white institutions did finally allow entry of people of color, the experiences of these students were often very discriminatory and disheartening.

With institutional racism foundationally in place, the role of historically black college (HBCUs) and universities

early became paramount as a means of educating young black men and women. Because of the inability of students of color to attend historically white institutions, predominantly black universities became an intellectual haven for these students. Even today, these historically black institutions educate and provide degrees to nearly 25% of the African-American student population (Borden, 2005). Indeed, HBCUs have become more than an educational experience for students of color; they have become a safe haven from a racialized society that often marginalizes and discriminates against them. In the next section, we examine African-American students attending predominantly white universities, their relationship with faculty and peers, and their perceptions of the general campus climate.

Students Attending Predominantly White Universities

Presently, some African-American students find themselves attending a university that has been and remains a historically white institution. In this setting, African-American students often report experiences of white racism on and off campus. These experiences can range from blatant racist acts to instances of racial joking and subtle exclusion. Most of this behavior is indeed taken for granted by most white administrators, faculty, staff, and students, who view that the campus as “a ‘white’ place in which blacks are admitted, at best, as guests” (Feagin *et al.*, 1996: 13). In addition, African-American students in these majority-white settings must be able to deal with racial slurs, avoid becoming lost and isolated, and protect themselves from harm while enduring or confronting an often negative racial climate (Feagin, 1996; Stewart, 2008). In these predominantly white settings, African-American students often “have little choice but to accept the racist epithets as part of the heavy personal and psychological price they pay for securing certain societal opportunities that they consider important” (Picca and Feagin, 2007).

In a recent examination of African-American student experiences at predominantly white universities in the state of Texas, we questioned students about the significance of race and racism in their everyday lives. These black students often mentioned feeling as if they are not wanted, or as though they do not belong on these historically white campuses. This feeling can often carry over into the classroom, impacting the grades of the students as well as their interactions with their classmates. For instance, Sandy a senior at a predominantly white university stated, “I don’t have a relationship with the other [white] students in most of my classes. None of them really talk to me and if someone asks a question and I know the answer, they always turn their heads and ask someone else” (Evans, 2006). In many of the interviews, other African-American students attending this university

frequently echoed her sentiments. In cases such as these, we likely observe the long-standing belief in the minds of many whites about the inferior intellectual competence of African-American students. Indeed, considerable evidence exists showing that African-American students continue to have race-related experiences in classroom interactions with faculty and peers that can strongly influence their academic development and willingness to participate in class (Chavous *et al.*, 2004).

Important educational relationships in predominantly white settings between white faculty and students of color are in some senses lacking. For instance, when asked about her relationship with faculty at her university, one student commented:

In one example, I have one classmate and we really talk a lot. She was advised by the same advising professor. We both asked for independent studies. She got the independent study and all she has to do is write a paper. I asked for an independent study and I was told, “Well, give me a compelling reason and I’ll consider it”. Now, it could be anything but it didn’t feel that way. . . (Evans, 2006).

Other students in this type of environment shared similar racialized experiences with white faculty. In some cases, the perception of the interaction itself is seen differently by white faculty and black students. In recent research conducted on white faculty and African-American students, Gonsalves states that often faculty and students interact and the African-American students feel as though they and their comments are not taken seriously or are not respected by white faculty members. In this silent engagement, misunderstandings as well as actual discrimination can occur (Gonsalves, 2002). African-American students are thus not as likely as white students to interact with white faculty outside the classroom. These students are not as likely to be well integrated into campus life as white students (Guiffrida, 2005).

In the predominantly white setting, exclusion is not the only experience experienced by African-American students. Another everyday reality for African-American students attending predominantly white universities is facing stereotyped notions about what it is to be black. These imposed racial identities not only can be offensive but also can serve as a means of differential racial treatment. For instance, research in this area indicates that African-American students, particularly men, are often stereotyped as athletes and not as serious students. Several black students interviewed in a recent study suggest that the assumption by white students that African-American males are only attending universities for athletics or because of athletic ability seriously misrepresents their academic achievements. Several male students mentioned that they are approached and asked blatantly, “what sport do you play;” contributing to the existing stereotype of blacks as athletes at white colleges and universities

(Evans, 2006). This experience is not limited to men; several young women have recounted similar incidents in reference to their campus identities as something other than a serious student.

To continue our understanding of African-American students in higher education, we now turn to a discussion of black students at predominantly black universities.

Students at Predominantly Black Universities

Student experiences at HBCUs are often vastly different from student experiences at predominantly white institutions. The historically black university student, undoubtedly, experiences issues with gender, sexuality, and social class; however, in regard to racism in the classroom or on campus, students there report experiences that are quantitatively and qualitatively different. Social science research suggests that attending an HBCU contributes significantly to African-American student outcomes in reference to educational experiences and attainment (Laird *et al.*, 2007). In this environment, black students are unlikely to experience an educational environment where they feel unwanted, but rather an environment in which they feel generally welcomed.

African-American students there mention the atmosphere of the university as one of racial equality, not inequality. For instance, students often state that “it feels good to be here and see that we are all working toward the same goal” (Evans, 2006). In addition, HBCUs provide an environment rich in history and positive associations with the African-American community and culture. Students often find themselves empowered by their campus surroundings with other African-American students and faculty of color providing significant positive role models.

Students at historically black universities not only mention having a different academic experience than those attending predominantly white universities, but they also mention having a different type of relationship with administrators, faculty, and staff. As mentioned earlier, faculty relationships can have a significant effect on college student outcomes; some students at historically white institutions mention feeling as though faculty members do not care about their progress in their respective programs or at the university. Conversely, students attending black universities frequently mention having a positive relationship with their faculty members; specifically the African-American faculty members. Students often speak of faculty in a way that includes mentorship, encouragement, and assistance. For instance, one female student that had been attending an HBCU for 3 years mentions:

Most of my professors or better yet all my professors know our first and last names; they know where we are

from; they know our career goals. A lot of my professors just push us and instill in us the desire to do better. They talk about things that are in the news with us. It is like a parent-child relationship. Because not only do they teach us, they mentor us and help us view things in a bigger picture. Black professors are more interested in what you want to do and helping you achieve those goals. They understand you because they may have had some of the same or similar struggles (Evans, 2006).

In conjunction with better faculty relationships, students often mention faring better academically. Although HBCUs make up slightly over 2% of institutions of higher learning, they are awarding one in five baccalaureate degrees to African-American students (St. John, 2000). By reducing or nearly removing the racial tension perceived by African-American students at predominantly white universities, the atmosphere is more conducive to student learning in these cases. In other words, the experiences of African-American students at HBCUs are usually more beneficial educationally than the experiences of African-American students at predominantly white universities (Laird *et al.*, 2007). In order to benefit from the environment that colleges and universities have to offer, young African-American high school graduates and their parents are often faced with the agonizing dilemma of choosing a historically white college location for their children.

An Agonizing Dilemma

In recent research conducted at HBCUs, students reported making the decision to attend this type of university specifically as a means of avoiding institutions perceived to be predominantly white and overtly racist. Although most of the students interviewed had no prior experience attending predominantly white colleges, they mentioned that such advice to attend an HBCU often came from parents and peers with experience at institutions that are predominantly white. For instance, one student had a parent who obtained a 4-year degree from an institution that was historically white, while her father attended an HBCU. When it was time to apply for college, she noted that

Well I always knew it was a HBCU and that was the prime concern with my parents. My mother went to a predominantly white college and my dad went to a HBCU and there just seemed to be more of a connection with my father and his school. My mother even wanted me to go to a black school.

In research conducted at a major university, sociologists Feagin, Vera, and Imani found that there were occasions where students and parents compared the environments

and situations at predominantly white colleges with those of HBCUs. Feelings of not belonging or being surrounded by whiteness versus feelings of a sense of belonging and of worthiness can make a huge difference to parents and students of color (Feagin *et al.*, 1996). Moreover, African-American students fare typically better socially and academically at historically black institutions because of the fact that these institutions often emphasize a general educational and achievement philosophy similar to that accented in African-American communities (Chavous *et al.*, 2004).

African-American students often share the concerns of their parents when making the decision to go to college. When having these conversations, parents report talking to their children about being careful and being aware of their surroundings if attending a university that is predominantly white. Other parents suggested that their children attend a university with a large African-American student population in order to have an experience with college that allows them “to focus on academics and not on how people treat you” (Evans, 2006). Collectively, African-American parents are aware of the environment on white campuses and feel an agonizing dilemma when sending their children into this type of white environment. Nevertheless, many African-American students do attend a university that is predominantly white, and there are often severe consequences for them. In this next section, we examine some of the psychological consequences.

Consequences of Racism in Education

Stark differences in racial experiences are observed in research on African-American students attending predominantly black colleges and those attending predominantly white colleges. These differences can have vastly different consequences. When it pertains to predominantly white universities, racial discrimination and the racist ideologies spouted by some white students, faculty, and administrators often lead to undue stress for African-American students. The vast differences in racial climates can lead to distinct psychological consequences, depending on the type of school attended.

For African-American students attending a predominantly white university, the general campus climate can be one that is seen as racist and unwelcoming to students of color. Not only is this environment seen as one of whiteness to African-American students, but also white students, faculty, and administrators often do not perceive any racial hostility or discrimination in the campus environment. Because of this, to succeed African-American students must develop various ways to cope with the racialized stress produced on these historically white campuses. Moreover, experiences with racism can lead to psychological problems, excessive alcohol consumption, and problems with classes and grades (Bynum *et al.*, 2007).

In addition to the aforementioned factors, racialized stress at predominantly white universities can lead to feelings of inadequacy and poor classroom performance, which in turn can lead to decisions to leave their current university. Generally, the academic environment at historically white institutions can be a source of stress because of factors such as racial hostility and discrimination, and African-American students there often attribute their higher levels of stress and poor psychological well-being to such unwelcoming environments (Heppner *et al.*, 2004).

For African-American students attending predominantly black universities, psychological problems are mostly caused by factors outside the realm of racial discrimination. They may simply come from the nature of higher education, as for white students at historically white institutions. However, even at the HBCUs, the surrounding white-dominated environments may cause them educational problems. Thus, students at HBCUs often report no experiences with discrimination or racist ideologies on campus, but often note these negative experiences when off campus and not in this safe environment of the predominantly black campus. For instance, students attending one HBCU in Texas cited numerous examples of overt racial discrimination off campus. Yet, such students can often rely on the counsel of black peers and faculty in dealing with these off-campus issues. In such cases, the university can become a safe space for getting help in dealing with outside racism or avoiding that racism.

Conclusion

When it comes to higher education and African-American students, several factors affect their educational lives. First, the fundamental role of systemic racism in the experiences of these African-American students must be considered. At most predominantly white universities throughout the United States, the general campus racial climate can be very uncomfortable to African-American students. Due to the historically exclusive and often racist nature of these schools, this racial foundation can still manifest itself in the everyday campus climate. The consequences for African-American students can be detrimental to their futures in academia and their educational success.

Conversely, students attending an HBCU often have a different experience when it comes to the realities of a racist environment and to educational achievement. There are better relationships with faculty, administrators, and peers that take place in this environment in comparison to schools in which African-American students make up a minority of the population. At these historically black schools, students generally report greater social and academic integration because they do not face significant feelings of exclusion and disrespect.

In regard to the broader meanings of racist experiences on predominantly white college campuses, we should consider the significance of breaking down the hostile racial climate and increasing the multiracial curriculum in all areas. Educating students on cultural diversity and the importance in maintaining and accenting this diversity is a necessity if the United States is to prosper in the future. Understanding the role of systemic racism in historically white educational settings, thus, can contribute to the changes needed in recruiting and retaining students of color. By teaching the significance of respecting various racial and cultural differences, white racial hostility and discrimination on historically white college campuses and the negative consequences for students of color can perhaps be diminished.

Bibliography

- Borden, V. M. H. (2005). The top 100: Interpreting the data. *Black Issues in Higher Education* 22(8), 36–39.
- Bynum, M. S., Burton, E. T., and Best, C. (2007). Racism experiences and psychological functioning in African American college freshmen: Is racial socialization a buffer? *Cultural Diversity and Ethnic Minority Psychology* 13(1), 64–71.
- Chavous, T. M., Harris, A., Rivas, D., Helaire, L., and Laurette Green, L. (2004). Racial stereotypes and gender in context: African Americans at predominantly black and predominantly white colleges. *Sex Roles* 51(1/2), 1–16.
- Evans, L. (2006). *When Black is White: The Social Construction of Race at Predominantly Black and Predominantly White Colleges*. Arlington, TX: The University of Texas at Arlington. Unpublished manuscript.
- Feagin, J. R. (2006). *Systemic Racism: A Theory of Oppression*. New York: Routledge.
- Feagin, J. R., Vera, H., and Imani, N. (1996). *The Agony of Education: Black Students at White Colleges and Universities*. New York: Routledge.

- Gonsalves, L. M. (2002). Making connections: Addressing the pitfalls of White faculty/Black male student communication. *College Composition and Communication* 53(3), 435–465.
- Guiffida, D. (2005). Othermothering as a framework for understanding African American students' definitions of student-centered faculty. *Journal of Higher Education* 76(6), 702–723.
- Heppner, P. P., Ji, P., Neville, H. A., and Thye, R. (2004). The relations among general and race-related stressors and psychoeducational adjustment in Black students attending predominantly White institutions. *Journal of Black Studies* 34(4), 599–618.
- Houts, P. L. and Feagin, J. R. (2007). *Two-Faced Racism: Whites in the Backstage and Frontstage*. New York: Routledge.
- Stewart, D. L. (2008). Being all of me: Black students negotiating multiple identities. *Journal of Higher Education* 79(2), 183–207.
- US Census (1890). *Report on Education in the United States: Table 5: Statistics of Schools, Census of 1880: Number of Pupils Attending Public Common Schools*. <http://www2.census.gov/prod2/decennial/documents/1890c1-02.pdf>. (accessed June 2009).

Further Reading

- Allen, W. R. (1985). Black student, White campus: Structural, interpersonal, and psychological correlates of success. *Journal of Negro Education* 54(2), 134–147.
- Allen, W. R. (1991). *College in Black and White: African American Students in Predominantly White and in Historically Black Public Universities*. Albany, NY: State University of New York Press.
- Johnson, C. A. (2005). *Guests at the Ivory Tower: The Challenges Black Students Experience While Attending a Predominantly White University*. Lanham, MD: University Press of America.
- Palmer, R. and Gasman, M. (2008). It takes a village to raise a child: The role of social capital in promoting academic success for African American men at a Black college. *Journal of College Student Development* 49(1), 52–70.
- Reid, L. D. and Radhakrishnan, P. (2003). Race matters: The relation between race and general campus climate. *Cultural Diversity and Ethnic Minority Psychology* 9(3), 263–275.

Bologna Process: On the way to a Common European Higher Education Area

A V Rezaev, St. Petersburg State University, St. Petersburg, Russian Federation

© 2010 Elsevier Ltd. All rights reserved.

Glossary

AUCC – Association of Universities and Colleges of Canada.

Bologna Declaration – The Joint Declaration of the European Ministers of Education signed in Bologna, Italy, 19 June 1999.

Bologna Follow-up Group (BFUG) – The administrative group organized in order to oversee the process of policy actions in European higher education between the ministerial conferences.

Bologna Process – Policy actions to create Common European Higher Education Area by 2010.

CEHEA – Common European Higher Education Area.

Communiqué – The official document resulting from biannual meetings of European Minister of Higher education with recommendations for acceptance of new members and a working plan for the subsequent 2 years.

EHEA – European Higher Education Area.

IHEP – Institute for Higher Education Policy.

Lisbon Strategy – The European Union (EU) Strategy that aims at making the EU the most competitive and dynamic knowledge-based economy in the world.

new dynamic to German universities by introducing the idea of a research university – ideas of freedom in teaching and learning (Rüegg, 1992/2004). This is the case, at present, in the Europe of the early years of the twenty-first century.

The last year of the twentieth century brought to public and professional attention a new concept describing the state of affairs in European higher education, which is the concept called the Bologna Process.

The purpose of this article is to look at the concept and meaning of the Bologna Process and to consider the strategies of European integration in the field of higher education. It includes an exploration of the definition and the meaning of the term and related notions, discussion of the rationales for the new plans, and policy actions called upon to transform higher education in Europe.

The primary objective is to construct a way of approaching the issues, to provide a framework for taking stock of what the appropriate paths are for establishing a common European higher education space, and to decipher what some of the challenges will be for European higher education following 2010, when the Bologna Process's ideas have to be implemented. Another important question involves how far the implementation of various policies in European higher education impacts the processes of tertiary learning in international perspectives.

Definition of the Bologna Process

The Bologna Process describes the active and comprehensive policy actions taken by European governments and universities leaders in order to modernize higher education structures in Europe and to create a Common European Higher Education Area (CEHEA). These actions commenced in 1999 and it was announced that activities of the process are to be completed by 2010.

The overall objective of the Bologna Process to create a CEHEA is detailed in the following action lines:

1. adopting a system of easily readable and comparable degrees;
2. adopting a system essentially based on two cycles (undergraduate and graduate);
3. establishing a system of transferable credits;
4. promoting academic mobility;
5. promoting European co-operation in quality assurance;

Introduction

The key concepts that have been dominant in professional literature on higher education in Europe have changed over time. Academic freedom and state, curriculum and methods of instruction, elite versus mass education, university governance, and funding have each, in turn, been examined in a series of debates and a plethora of reports, books, articles in professional journals, dictionaries, and encyclopedias (Rothblatt and Wittrock, 1993).

Philosophers, public thinkers, and political leaders have been involved in constant debates concerning the present state and the future of higher education in Europe (Bourdieu, 1996). This was, likewise, the case in the eleventh century when the idea of the university was introduced to Europe and to Western civilization. That was also the case in the nineteenth century when Wilhelm von Humboldt gave a

6. promoting the European dimension in higher education;
7. encouraging lifelong learning;
8. developing partnerships between higher education institutions and students;
9. promoting the attractiveness of the CEHEA; and
10. building programs in doctoral studies and the synergy between the CEHEA and the European Research Area.

The Bologna Process started as an initiative of European Ministers and European university leaders; however, it is not an official European Union (EU) project. Despite this, there is a great deal of overlap between policy actions in the framework of the Bologna Process and various official EU projects, including the EU's Lisbon Strategy that aims at making the EU "the most competitive and dynamic knowledge-based economy in the world" (Lisbon Strategy 2000). The reforms involved in this process attempt to address some of Europe's social and economic challenges by enhancing the quality of its education, research capacity, and graduate employability.

The Bologna initiatives stand for a comprehensive modernization of study programs and for internationally comparable degrees. They also entail a change in perspective toward learners and their development of competences. At the end of this path, European higher education leaders hope to find improvements in the quality of teaching for the sake of strengthening student's professional knowledge and skills.

The Set-Up and the Development of the Bologna's Initiatives

Universities have always reacted to environmental pressures for change, and – in the long run – they have done it quite successfully; that is why, at present, universities are among the oldest of structural organizations that are known in the world.

Much was heard in academic discourse with regard to reforming or restructuring, and concerning the reorganizing or modernizing that European countries have experienced in higher education since the mid-1980s. A great deal of academic research and writing was aimed at finding new ways to analyze the relationship between socioeconomic changes and higher education policies. All of this was sparked by perceptions of turbulent change

in contemporary Europe that has not been felt in centuries. The changes were most dramatically evident in the higher-learning institutions of Europe following the collapse of the old regimes in Eastern and Central Europe in the late 1980s, and the disintegration of the Soviet Union in 1991.

The field of education became most sensitive to simultaneously occurring revolutions: a political revolution that created new nation-states in Europe, an ideological revolution that ended the communist regimes, and an economic revolution that replaced state-administrative systems of production, distribution, and consumption with a market-driven economic order. Indeed, socioeconomic, political, and technological processes at the end of the twentieth century brought to the fore a completely new agenda for European education.

New national borders in Europe, new migration processes, and new structures in the labor market present new opportunities for higher education systems and international academic collaboration, while advances in communication technologies allow much faster exchanges and greater flexibility in the creation of joint-study programs. However, these exciting opportunities for international cooperation in higher education, at once, represent a host of challenges related to issues such as academic recognition, quality-assurance standards, and compatibility of qualification frameworks.

In June 1999, ministers of education from 29 European countries gathered in Bologna – the birthplace of Europe's oldest university – to sign the Joint Declaration of the European Ministers of Education (otherwise known as the Bologna Declaration). The policy actions proclaimed in the Declaration aimed at "increasing the international competitiveness of the European system of higher education" and at creating "the European area of higher education as a key way to promote citizen's mobility and employability and the Continent's overall development" (The Bologna Declaration, 1999). The Bologna Process had been started (**Figure 1**).

The Bologna Process has become the most important initiative in the field of higher education for contemporary Europe. Ministers of higher education from all country-signatories of the Bologna Declaration discuss policy actions, coordination efforts, and general strategies at biannual meetings (**Table 1**). The results of the Ministerial meetings are memorialized in documents called *communiqués*.

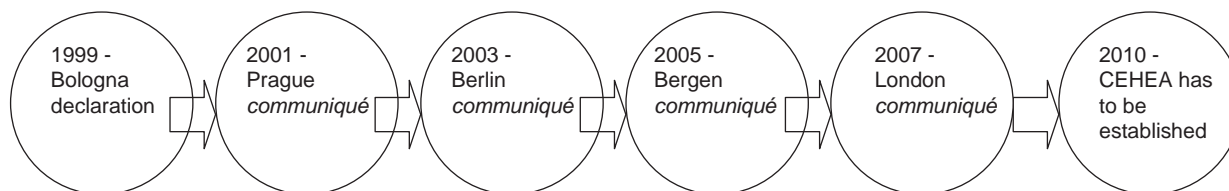


Figure 1 The Bologna Process.

Table 1 Country-signatories of the Bologna Declaration

Since 1999		Since 2001	Since 2003	Since 2005	Since 2006
Austria	Lithuania	Croatia	Albania	Armenia	Montenegro (as an independent State)
Belgium	Luxembourg	Cyprus	Andorra	Azerbaijan	
Bulgaria	Malta	Liechtenstein	Bosnia and Herzegovina	Georgia	
Czech Republic	Netherlands	Turkey	Holy See	Moldova	
Denmark	Norway		Russian Federation	Ukraine	
Estonia	Poland		Serbia and Montenegro		
Finland	Portugal		The former Yugoslav Republic of Macedonia		
France	Romania				
Germany	Slovak				
Greece	Republic				
Hungary	Slovenia				
Latvia	Spain				
Ireland	Sweden				
Italy	United Kingdom				
Iceland	Switzerland				

The *Communiqué* sets out recommendations for acceptance of new members, future priorities, and offers a working plan for the subsequent 2 years. The ideas of the *Communiqué* are to be implemented in the individual countries in accordance with their relevant legal contexts and national priorities.

The Bologna Follow-up Group (BFUG) was established in order to oversee the process between ministerial conferences. BFUG is composed of: (1) Representatives of all countries participating in the process of creating the CEHEA; (2) The European Commission as an additional full member; (3) Eight consultative members, namely – the Council of Europe, UNESCO's European Centre for Higher Education, the European University Association, the European Association of Institutions in Higher Education, the European Students' Union, the European Association for Quality Assurance in Higher Education, the Education International Pan-European Structure, and Business Europe. The BFUG meets at least once every 6 months and is chaired by the country holding the Presidency of the EU.

Economic, Social, and Academic Rationales for Unified European Higher Education

The Bologna Process should not be viewed in isolation from global processes – it is an example of a response to the internationalization of higher education and global integration; it is determined by economic, social, and academic rationales. The Bologna Process initiatives should also not be viewed in isolation from the sociopolitical processes in contemporary Europe.

European higher education structures are undergoing a substantial, comprehensive change as part of the international higher education system. The scope of this change is as great, or even universal, as the change that higher education systems experienced at the end of nineteenth century (Slaughter and Leslie, 1997), when vast advances in the industrial mode of production and an overwhelming transformation of social relationships fundamentally reshaped the mission, structure, and practices of higher education. At present, the expansion of markets and market relationships, globalization, and the further development of technology entail higher education-transformation processes in three major interrelated directions: the marketization of higher education, structural modification, and the establishment of new forms of learning at advanced levels. The market dominance and sharp decrease of government spending on higher education in contemporary European societies necessitates greater reactivity in higher education systems to market signals; especially those that come from the labor market. Higher education institutes (HEI) are forced to adjust to market conditions, and they do so by developing new organizational structures and altering the preceding structures, policies, and practices. The end of the twentieth century in Europe was marked by the emergence of a new type of HEI that has been recently referred to as an innovative, for-profit, or entrepreneurial university (Clark, 1998; Ruch, 2001). Finally, the introduction of new forms of higher education also provides evidence of the ongoing process of transforming traditional patterns of higher education: e-Learning, distance learning or distance education, life-long education, and nonuniversity forms of tertiary education. Europe is confronted with the reality of change when knowledge is vital to the welfare of its countries.

The scope, diversity, and contradictory nature of the aforementioned changes characterize the great interest of policymakers in terms of the pathways and mechanisms of transformation, its driving forces, determinants, and agents. The socioeconomic realities of contemporary Europe determine the necessity for concrete policy actions and projects that will help with the realization of an efficient and performing knowledge enterprise that generates increasing output with decreasing input.

It has also to be taken into account that the forming of the EU as a new sociopolitical unit now accounts for about 500 million citizens and about one-third of the world's gross domestic product (GDP). Although it arguably constitutes a technical process, the complexity of uniting the diverse cultural regions and ethnic and national traditions in common socialization processes, at the same time, requires active policy actions for creating a new European identity. Educational policy is the most direct means to foster an environment for developing, in the young generations, a feeling of unity in diversity.

Thus, a specific approach was needed to create a new European educational model that would value a variety of countries and nations and yield the maximum possible results for establishing the new EU. A shared long-term vision and the objectives of universities in European society as agents for constructing a united Europe in the global context came to the political agenda.

Globalization Effects on European Higher Education

How does globalization affect higher education policy and academic structures in Europe? Social analysts disagree on the essence and critical features of globalization and even whether it was unique historically, although most agree on certain basic ideas with respect to the relations between globalization and higher education and on the factors that shape the realities of the tertiary education systems in the twenty-first century.

Among the most important are the assumptions that the driving force for higher education in the time of globalization are the market and commercialization. The factors that influence the dynamics of higher education systems include new ways of financing higher education, new managerial techniques, and new rules of the game for people who make up the schools, unprecedented mobility for students and professors, the role of English as the *lingua franca* for higher education and science, advanced information, and communication technology (Altbach, 1998).

In Europe, education, in general, and higher education, in particular, were the great agents of nation-state building. Eugene Weber – in a classic account of nation building – has documented how people in France were molded into a

nation. The most important element that sped this process was a national educational system (Weber, 1976).

Ernest Gellner's study of nations and nationalism further stressed the relationship between nation building and education (Gellner, 1983). As Richard H. Robbins points out, "Gellner went so far as to suggest that education became the ultimate instrument of state power, that the professor and the classroom came to replace the executioner and the guillotine as the enforcer of national sovereignty, and that a monopoly on legitimate education became more important than the monopoly on legitimate violence to build a common national identity" (Robbins, 2005).

Thus, European academic systems and structures that – back in history – were closely related with nation-states and grew within national boundaries, in the time of globalization, find themselves competing internationally. The national languages that were major factors for comprising educational systems compete with English even within national borders.

Globalization poses exciting new opportunities for policymakers. Despite the fact that policymakers worry that they are losing yet more of their control over their own higher education systems, they are, nevertheless, turning increasingly to regional or international institutions in the hope that coordinated solutions will provide some new dynamics.

Globalization as a concept and analytical instrument – while used for understanding processes of change in higher education – focuses on the power relations between nation-states and global and regional actors, on the economic and social interactions between HEIs, governments, business enterprises, and nongovernmental organizations (NGOs).

The current debates on higher education and globalization in Europe have two distinct tendencies that might be called positive and negative. One group of scholars, political leaders, and educational practitioners praise globalization for opening new opportunities for higher education through trade liberalization, global competition, and commercial interactions. They put forward an argument that globalization serves as a means to eliminate the effects of national isolation. In the time of significant public cuts to systems of higher education, globalization is viewed by this group as a kind of new wind for the canvases of educational ships. Another group of academics argues that economic globalization is strictly alien and even conflicting with traditional goals and fundamental values of higher education. Globalization, in their view, makes the Humboldtian/European university model obsolete. The university may well survive as an educational corporation pushed forward by globalization, yet – in such a case – it would no longer be the university as it was known throughout European history (Higher Education in Europe, 2001).

The implementation of the Bologna Process's ideas demanded a strong governmental influence on higher

education. Thus – on the one hand – globalization brings to Europe the idea of a declining role of the state in university life. Yet, on the other hand, the government – irrespective of whether it is federal or unitary, centralized or decentralized – is in constant demand as a leading actor in facilitating the Bologna policy actions to modernize its higher education system by setting new parameters and providing financial support and legal regulations.

A strange symbiosis of globalization processes and the mounting role of nation-states for modernizing higher education systems throughout Europe tends to reinforce the fears of those who are concerned about the increasing homogenization and standardization of higher learning.

There are other challenges of globalization that have direct and profound effects on higher education in Europe: technological change and environmental threats, global risks and threats to the well-being of society, migration, and the security of the population (i.e., terrorism).

The Bologna Process is an attempt to meet these challenges, to answer the question of how higher education can become an active part in shaping the future of European societies.

The Bologna Process in the International Setting

The European higher education-reform model has created considerable interest in other parts of the world. In some countries, it generated public debate concerning the possibilities of implementing Bologna-like reforms in domestic higher education. A number of professional higher education associations around the world have acknowledged the significance of the emerging CEHEA, and have decided to undertake a course of action to address the implications of the Bologna Process for the development of their national universities (AUCC, 2008).

There are international initiatives aiming to facilitate the convergence of higher education systems at a regional (supranational) level. The inaugural Asia-Pacific Education Ministers' meeting was held in Brisbane, Australia, in April 2006. The Ministers signed a Brisbane *Communiqué* and agreed to promote regional student and academic mobility.

The Bologna Process has also stimulated a discussion between European and international partners with regard to the mutual recognition of qualifications. In the USA, for example, the Institute for Higher Education Policy (IHEP), which is based in Washington DC, has launched an initiative to create a new understanding of the rapidly changing global context for higher education learning and credentialing and the impact of these changes on US higher education.

While the Bologna initiatives are geared toward endorsing Europe's own competitiveness and attractiveness to non-European students and faculty, stakeholders within

Europe are becoming increasingly aware of the need to look beyond Europe's borders to broaden cooperation with other regions of the world. In the most recent years, country-signatories of the Bologna Declaration have also commenced to actively promote the ideas of the Bologna Process throughout the globe. In 2007, at the Ministerial conference in London, The European Higher Education Area in a Global Setting strategy was adopted. It encompasses the following five core policy areas:

1. improving information on the CEHEA;
2. promoting European higher education to enhance its worldwide attractiveness and competitiveness;
3. strengthening cooperation based on partnerships;
4. intensifying policy dialog; and
5. furthering recognition of qualifications.

Problems in the Implementation of the Bologna Process

The Bologna Process only provides the administrative framework for reforms, yet there is no homogeneous situation throughout Europe with regard to the implementation of strategic policies. The starting point for each country-signatory of the Bologna Declaration was different. Some countries already had a two-cycle (undergraduate–graduate) structure or a credit system in place while others did not have it. Countries also started at different times and with different economic potentials. The countries that joined the Bologna Process after 2003 are not yet advanced in implementation of the announced objectives. The former Soviet Union countries face multiple economic, social, and political problems in their attempts to incorporate Bologna's ideas into the reality of the legacies of the former centrally planned type of society and ideology-based education (Rezaev, 2005).

Following several years of Bologna Process activities aimed to transform European higher education, most European universities are still not competitive on a global scale with the HEIs of its major competitors; neither in access for students not from Europe, nor in attractiveness for students around the world, nor in the excellence of education and research. It is a general perception that the productive reforms cannot be accomplished within the current levels and patterns of investment in European higher education. The total (public and private) spending on higher education in the EU, in 2002, was 1.3% of GDP, compared to 2.8% in the USA. To close the spending gap with the US, the EU would have to commit an additional 140 billion Euro per year – securing, in particular, substantial investment from the private sector. One of the remedies to improve the situation in European higher education will be to diversify financial incomes and investments.

The major higher education initiatives that have been undertaken in the framework of the Bologna Process,

in Europe, in recent years have had different assessments. The optimists look at CEHEA as a real step toward constructing a new society based on common European ideas and values. The pessimists see the Bologna Process as an obstacle for real modernization of the national systems of higher education. Then a camp of pragmatists assesses CEHEA as a response to globalization processes and Europe's need for self-improvement and to make the higher education system more dynamic under market constraints. The latter group argues that CEHEA is the only way for European universities to become a real alternative to the American educational system – which takes in more and more students from around the world, including Europe.

Whatever view characterizes the current status – great success, success, partial success, or failure – nonpartial observers agree that implementation of Bologna's objectives is a complex process and that it needs more energy and constructive actions.

What Lies Ahead?

The actual problems for European higher education leaders to successfully establish CEHEA are quite numerous. To name a few that are most obvious: (1) Higher education systems in European countries continue to be a responsibility of national governments. The major portion of funding still comes to HEIs from national governments. (2) Most European HEIs are provincial, oriented toward training graduates from a specific region for a specific regional labor market. (3) The large majority of students – specifically from Central and Southern Europe – do not study abroad and could not effectively study in English but only in their native language. (4) The European labor market for graduates is not open to all professional fields but to relatively few. (5) European higher education continues to be highly Eurocentric, causing real problems for students from Muslim countries such as Turkey or Azerbaijan. (6) Foreign students are unevenly spread in European HEIs – with the UK, Germany, and France hosting the majority of international students. Other countries, particularly the member states of Eastern and Southern Europe, are highly underrepresented.

The realities of creating the new European system of higher education manifest the necessity to make this system open to real diversification. Five elements will constitute and determine the world of diversity in European higher education following 2010:

1. diversity of funding base;
2. diversity of curricula and programs;
3. diversity of methodology and teaching techniques;
4. diversity of departments and departmental periphery (interdisciplinary centers and professionalized organizational units to work with outside organizations and groups); and

5. diversity of faculty and students in terms of ethnicity, gender, cultural, and sub-cultural milieu.

Summary

The Bologna Process, aiming to create CEHEA by 2010, is an ambitious project that is complex and subtle. It is a project that answers objective economic, social, political, and academic challenges. The CEHEA is a significant challenge for the EU and for the unity of countries that share national borders, citizens, or values. It requires cooperation and collaboration among units that are often self-interested instead of focused on communal activity.

What is possible in the coming years will be achieved only through ambitiously tackling the obstacles to coherent growth and through self-sacrifice to the common goals. The Bologna Process has gained the allegiance of 46 countries, including those like Russia and the former Soviet Union countries that are not formally part of the EU and are struggling with the Soviet legacies. The opportunity of improving interactivity between HEIs through joint projects, research and development, scientific cooperation, and sharing of resources is already evident. How much unity can be found in the diversity and whether or not modern higher education will be fundamentally altered as a result of new technologies, systems, and programs are still debatable questions in Europe. In some senses, CEHEA redefines the meaning of university and seeks to establish a multiversity wherein diverse voices are represented on their own sovereign and independent terms.

The Bologna Process is a specifically European answer to global issues; it is obvious that the Bologna Process also contributes to international public goods. Despite this, the necessity for more institutional and conceptual cooperation with the rest of the world is of great importance. It is almost obvious that there is a necessity to open much wider the doors of the European higher education labor market to the rest of the world.

The year 2010 is the year when the Bologna Process is scheduled to be fully implemented and it appears that Europe will meet yet another process rather than a final event.

See also: Curriculum and Globalization: Higher Education; Internationalization of Higher Education.

Bibliography

- Altbach, P. G. (1998). *Comparative Higher Education*. Norwood, NJ: Ablex.
- AUCC (Association of Universities and Colleges of Canada) (2008). *Canadian Universities and the Bologna Process*. AUCC Statement. http://www.aucc.ca/_pdf/english/statements/2008/bologna_process_background_06_20_e.pdf.
- Bourdieu, P. (1996). *Homo Academicus*. Cambridge: Polity Press.

- Clark, B. (1998). *Creating Entrepreneurial Universities: Organizational Pathways of Transformation*. Oxford: Elsevier.
- Gellner, E. (1983). *Nations and Nationalism*. Ithaca, NY: Cornell University Press.
- Higher Education in Europe (2001). Globalization: A new paradigm for higher education policy. Vol. XXVI, No. 1, UNESCO-CEPES, European Centre for Higher Education. *Lisbon Strategy*. http://www.europarl.europa.eu/summits/lis1_en.htm.
- Rezaev, A. (2005). Diversification in Russian higher education: Profile, foundations and outlooks. In Allen, W. and Teranishi, R. (eds.) *Higher Education in Global Society: Achieving Diversity, Equity and Excellence*. Oxford, UK: Elsevier.
- Robbins, R. H. (2005). *Global Problems and the Culture of Capitalism*, 3rd edn. Boston, MA: Pearson Education.
- Rothblatt, S. and Wittrock, B. (eds.) (1993). *The European and American University since 1800: Historical and Sociological Essays*. New York: Cambridge University Press.
- Ruch, R. (2001). *The Higher Education, Inc: The Rise of for-Profit Education*. Baltimore, MD: The Johns Hopkins University Press.
- Rüegg, W. (ed.) (1992/2004). *A History of the University in Europe*, vol. I, II, III. New York: Cambridge University Press.
- Slaughter, S. and Leslie, L. (1997). *Academic Capitalism: Politics, Policies and the Entrepreneurial University*. Baltimore, MD: Hopkins University Press.
- The Bologna Declaration, (1999). *Bologna Declaration of 19 June 1999: Joint Declaration of the European Ministers of Education*. <http://ec.europa.eu/education/policies/educ/bologna/bologna.pdf>.
- Weber, E. (1976). *Peasants into Frenchmen: The Modernization of Rural France, 1870–1914*. Stanford, CA: Stanford University Press.
- Dale, R. and Robertson, S. (2008). *Globalization and Europeanization of Education*. Oxford: Symposium Books.
- Gellert, C. (ed.) (1995). *Diversification of European Systems of Higher Education*. Frankfurt/M: Lang.
- Kogan, M., Bauer, M., Bleiklie, I., and Henkel, M. (2000). *Transforming Higher Education: A Comparative Study*. London: Jessica Kingsley.
- Kohler, J. and Huber, J. (eds.) (2006). Higher education governance between democratic culture, academic aspirations and market forces. *Council of Europe Higher Education Series Nr. 5*. Council of Europe Publishing.
- Marga, A. (2005). *University Reform Today*, 4th edn. Cluj-Napoca: Cluj University Press.
- Marginson, S. and Considine, M. (2000). *The Enterprise University*. Cambridge: Cambridge University Press.
- Moore, R. (2004). *Education and Society: Issues and Explanations in the Sociology of Education*. Cambridge: Cambridge Polity Press.
- Neave, G. (2003). The Bologna declaration: Some of the historic dilemmas posed by the reconstruction of the community in Europe's systems of higher education. *Educational Policy* 17(1), 141–164.
- OECD (1998). *Redefining Tertiary Education*. Paris: OECD.
- Teichler, U. (2007). *Higher Education Systems: Conceptual Frameworks, Comparative Perspectives, Empirical Findings*. Rotterdam and Taipei: Sense Publishers.
- Weber, L. and Bergan, S. (eds.) (2005). The public responsibility for higher education and research. *Council of Europe Higher Education Series Nr. 2*. Council of Europe Publishing.
- Witte, J. (2006). *Change of Degrees and Degree of Change. Comparing Adaptations of European Higher Education Systems in the Context of the Bologna Process*. Enschede: CHEPS/Universiteit Twente.

Further Reading

- Beerens, H. J. J. G. (2004). *Global Opportunities and Institutional Embeddedness: Higher Education Consortia in Europe and Southeast Asia*. Enschede: CHEPS.
- Brennan, J., King, R., and Lebeau, Y. (2005). *Universities and the Transformation of Societies*. London: Open University and the Association of Commonwealth Universities.
- Calhoun, C. (2006). The university and the public good. *Thesis Eleven* 84(7), 7–43.

Relevant Websites

- <http://www.ond.vlaanderen.be> – Bologna Process.
- <http://www.bologna-bergen2005.no> – Bologna Process/Bergen.
- <http://ec.europa.eu> – Education and Training/ European Commission.
- <http://ec.europa.eu> – European Commission.
- <http://www.esib.org> – European Students' Union.
- <http://www.europeunit.ac.uk> – Europe Unit.
- <http://www.eua.be> – European University Association.

Critical Race Theory in Education Research

R Vue and C B Newman, University of California, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Colorblindness or colorblind ideology – Refers to a worldview that disregards race or racial characteristics. In practice it is a method of disregarding race as a criterion used for selection of participation or receipt of services.

Jurisprudence – Refers to the theory or philosophy of law. In this article it primarily refers to the dominant legal thought which excludes and marginalizes racial minorities in the United States.

Meritocracy – Refers to a system of rewards based on the ability and talent of individuals, leaving unaccounted for the structural advantages and disadvantages conferred to different groups in society.

Positivism – As it concerns social reality and knowledge is the notion that both exist objectively and independently of social actors.

Race – Refers to a socially constructed classification used to categorize people into groups based on phenotype and characteristics assumed to be inherently shared.

Critical race theory (CRT) is an evolving movement formalized in the legal field during the 1970s in order to grapple with the nexus between race, racism, and American jurisprudence, in a better manner. CRT's emergence represents a response to the retrenchment of civil rights reform. This accounts for critical race scholars' critique of civil rights doctrine (which espouses equal treatment) and dissatisfaction with traditional legal reform; Both legal ideals viewed as inadequate for the emancipation of people of color and susceptible to discourses supporting neoconservative agendas to reinstate racial subordination. Accordingly, CRT was formulated by scholars whose primary concern was advancing a theory and method for examining race and overcoming racial inequality through critical analysis of legal doctrines. While CRT is largely used in the legal field, the theory has been adopted by scholars in fields like sociology and education. Scholars in these fields continue to utilize and advance CRT's core tenets – most notably the centrality of race and experiential knowledge. This article discusses CRT as a theoretical framework for education research.

Although an exhaustive list of the theoretical elements and arguments of CRT scholars is beyond the

scope of this article, the key principals presented provide a basic understanding of theoretical underpinnings. First, a brief summary of CRT's emergence is provided; we then describe two central features of CRT and address how CRT has evolved. Finally, a description of its application to education presents examples of scholarship where CRT is employed to understand issues in the field of education. We conclude with implications for the future of educational research as it relates to a critical race framework.

Though critical race theorists vary in the scope of their study, at least five central tenets guide its use in education (theory, research, pedagogy, curriculum, and policy): "Centrality of race and racism and their intersectionality with other forms of subordination, the challenge to dominant ideology, the commitment to social justice, the centrality of experiential knowledge, and the transdisciplinary perspective" (Solórzano, Ceja and Yosso 2000, p. 63). The use of CRT in education is appropriate for several reasons; however, the main justification extends its emergence out of critical legal studies (CLS); both in its points of alignment and departure. Thus, before addressing CRT in education some discussion of its emergence merits attention.

Critical Legal Theory

CRT is considered to have its origins in a departure from the theorizing of CLS. While the critical race movement's inception is marked by critiques of legal scholars, most of whom were scholars of color (such as Derrick Bell, Richard Delgado, Allan Freeman, and later scholars such as Kimberly Crenshaw, Angela Harris, and Mari Matsuda) within the legal studies movement, the formulation of CRT is inherently informed by a myriad of previous movements – including (among others) feminism and the spirit of the civil rights movement encompassing a long legacy of resistance to racial subordination (Delgado and Stefancic, 2001) – and literature across multiple disciplines (Yosso, 2005). However, legal scholars were instrumental in informing the constructs that ultimately have come to define the critical race movement and theory.

The emergence of CRT is more than simply an extension of CLS: it serves as a critique. Although critical race theories and critical legal theories are separate and in contention with one another, they share some notions about how law operates in society. Broadly, the critical

legal movement emerged to counter presumptions that the law operates in a neutral and objective manner (Hunt, 1986). Specifically, it advocates notions of realism, a movement “developed out of the dissatisfaction with tenets of classical legal thought that cast judicial decision as the product of reasoning from a finite set of determined rules” (White, 1972 as cited in Tate 1997, p. 207). Thus, CLS provides an analytical lens to critically examine legal ideology and discourse as a mechanism to socially reproduce and legitimate inequality within society (Hunt, 1986; Tate, 1997). In alignment with this tradition, Tate’s (1997) review of CRT origins notes that critical legal theories provide two important critiques that would ultimately inform CRT; these include a critique of formalism and objectivism. The first is characterized by “impersonal purposes, policies, and principles that serve as the foundation for legal reasoning” while the latter assumes the legal system provides an intelligible moral order (Tate, 1997, p. 208). Therefore, critical legal theorists offer a critique of law which acknowledges legal ideology and legal structures of authority as producing and maintaining social inequality, as opposed to functionalist views viewing social inequality as natural. This critique has served critical race theorists in their interrogation of inconsistencies within legal doctrines, specifically explanations of the contradictions and therefore negative consequential effects of civil rights law on progressive racial reform (i.e., Bell, 1976; 1980; Freeman, 1995). For these reasons, Matsuda (1995) describes, “[t]he willingness to at least consider the utopian prospect and the passionate criticism of existing conditions of racism and poverty attract non-white readers to CLS” (p. 64). Given the legacy of racial inequality within the United States, largely maintained through the construction of laws, critical legal theories share an important perspective with the critical race theorist, which is necessary for understanding the nexus between law and oppression.

Despite its utility, critical race theorists point to limitations of critical legal theories, which render them inadequate for understanding experiences of populations of color. Dalton (1995) writes, as scholars of color within the CLS movement they “are always invited in for tea, but rarely invited to stay for supper, lest we use the wrong intellectual fork” (p. 80), specifically speaking to the limits and determinism that CLS’s hegemonic view of rights doctrine poses for people of color. The collective response to sentiments of disempowerment and intellectual differentiation resulted in CRT; a movement which foregrounds the lived experiences of people of color in order to both acknowledge their marginalization in American jurisprudence but also enable a vision of law that holds the potential for social justice in terms of equal rights, status, and material outcomes. While there are multiple theoretical discontents, including inability to offer structures for social change, a central critique that

critical race theorists have of critical legal theories is the assumption that racism is the same as other forms of oppression (Delgado and Stefancic, 1993). Utilizing legal theorists’ critiques of legal doctrine as a foundation for their own work, critical race theorists brought the racial dimension to the forefront of the analysis: race, they assert, is crucial for understanding and ultimately overcoming inequality given racial saliency in the United States.

Grounded in traditions of the critical legal theorist, Bell’s early writings (including Bell, 1976; 1980) have been identified as “the intellectual starting point of Critical Race Theory” (Crenshaw, Gotanda, Peller, and Thomas, 1995: 3). Crenshaw and colleagues believe that while Bell’s critiques of civil rights discourse represent critical legal critiques of the time, his use of a race-conscious perspective that (1) focused specifically on the effect on Blacks and (2) presented legal decisions as reflective of White racial interests, was a departure in perspective, marking the development of CRT. Specifically, Bell’s (1980) concept of interest convergence – the argument that basic rights are provided only when they meet the shared interests of and do not harm the benefits of Whites – was significant in propelling forward the critical race movement, especially as scholars of color worked to construct a theory in which race centered as an important, necessary basis of analysis.

Centrality of Race

The centrality of race is one of the distinguishing factors of CRT (Delgado and Stefancic, 1993; Solórzano *et al.*, 2000). Crenshaw (1988), points to the importance of race as a basis of analysis, given the legacy of racial oppression in the United States. She argues to ignore that race is illogical, especially when race has become so deeply embedded in how institutions operate to produce social inequality. For that reason, critical race theorists argue that race is significant and excluding race from the analysis merely veils rather than halts its influence in research, policy, and practice. By bringing race to the forefront of the analysis, critical race theorists explicate the presence use of race. Therefore, revealing race consciousness will illuminate what has traditionally been ignored or hidden in past analysis.

When applied specifically to education, the issues of race often fall into the colorblind discourse (Crenshaw, 1988). Colorblind rhetoric and rationale is misleading because although it follows civil rights doctrine, it is often used by conservatives to garner support for repealing race-based initiatives aimed at addressing race inequities. By evoking principals of equality while ignoring the cumulated effects of race reification in material conditions, the concept of colorblindness is deployed to eliminate progressive gains of racial minorities. A well-known and debated contemporary example of this is present in

debates over the issue of affirmative action in United States higher education. Opponents of affirmative action argue that race-based sanctions are unjust, because it betrays a fair approach by privileging people of color over Whites. To ground their arguments, opponents call forth the contradictions that affirmative action poses to treating all people as equal. Embedded within this discourse is the notion of a meritocracy, which is a system of rewards believed to be based on demonstrated talent and ability of an individual. Both simultaneously work to undermine social reform addressing race inequities. The use of colorblind ideology to dismantle affirmative action policies exemplifies critical race legal scholars' critiques of civil rights doctrine, and more broadly liberalism, which is ambiguous and potentially serves to disarm progressive efforts toward social justice. In principle alone, arguments grounded in colorblind discourses may appear to adhere to moral reasoning; however this rationale leaves accounted for the historical and sociocultural contexts that justify the need for affirmative action practices.

Crenshaw (1988) addresses the notion of equality by framing it in terms of equal process and equal outcomes. Using the terms restrictive and expansive equality Crenshaw argues that equal processes are not always indicative of equal outcomes. In accord with the earlier work of Freeman (1995), who uses the competing concepts of perpetrator and victim perspectives of racism to explicate race (and consequently non-neutral nature) in law, Crenshaw points to the absence of historical continuity in rationalizing colorblind approaches to equality. The notion of equality as used to support the elimination of race-based initiatives fails to acknowledge the presence of different opportunities, erroneously positioning Whites as the true victims. As such, race as used in the case of explaining affirmative action, and more generally law in the United States is important because it explicates the presence of race and the role of racism in ideology, structures, and practices.

Similarly, Tate (1997) raises the issue of race as critical for understanding law, education, and society. Arguments for including race are based on the premise that a colorblind approach requires a consciousness of race to be able to reject its consideration (Aleinikoff, 1991 as cited in Tate 1997: 203). In this regard, ignoring race merely disguises rather than eliminates its significance to debates. In citing Justice Blackmun's dissenting opinion in the *Bakke* decision, this argument is exemplary of critical race thought, "In order to get beyond racism [to color blindness], we must first take into account race. There is no other way" (*Regents of the University of California vs. Bakke*, 1978: 407 as cited in Tate, 1997: 203). More importantly, recognizing that race exists unveils the construction of whiteness as normative. Critical race theorists put forth race as a strong analytical tool for understanding and overcoming racism. Therefore, CRT's advancement in the legal field serves to further deconstruct the law,

which was formed (and evolved) to maintain racial privileges of one group and consequently legalize the oppression of another; all factors which betray the objectiveness it presupposes.

Although race is a central feature of CRT critical race scholars recognize the role of multiple forms of subordination including class, gender, and sexuality operating in society and law. Therefore, embedded within the critical race framework is a rejection and critique of essentialism (Delgado and Stefancic, 2001). Anti-essentialism is the notion there is no singular experience for people of color, Blacks, or any other socially constructed group. To further advance the anti-essentialism thesis, critical race feminist scholars proposed intersectionality to more meaningfully account for the unique ways in which multiple forms of oppression interact to shape the lived experiences of those with multiple identity. Intersectionality does not merely suggest that multiple identities are additive, but that their intersectional nature create a unique social location of marginality as well as the potential for ground breaking social and political empowerment. For example, Crenshaw (1995) describes how both feminists and antiracist discourses have failed to adequately consider women of color's intersectional identities. Therefore, the centrality of race is not to minimize the role of these other forms of oppression, but instead to recognize its social construction largely through legal processes (among other means), such as the formation of the constitution, which continue to influence society. Critical race scholarship works under the notion that racial subordination is intricately linked to other forms of oppression; thus dismantling racial inequity is a step towards achieving social justice.

The Centrality of Experiential Knowledge

Within the CRT movement is the commitment to voice scholarship, which centers experiential knowledge as important. Critical race scholars emphasize the importance of localized narratives, which helps to contextualize and provide understanding. This incorporates a methodological and theoretical approach of perspective taking that requires "looking to the bottom," – the oppressed of society – to effectively elucidate and appreciate the complexities of their struggle, both material and ideological (Matsuda, 1995). Further, Delgado (1989) argues that a method of including the experiential knowledge of people of color counters the way traditional legal scholarship conducts analysis. Given that many legal scholars embrace universalism over particularity, the perspective tends to minimize anything that is historical, contextual, or specific with the unscholarly descriptors of literacy or personal (Tate, 1997: 220). Critical race theorists argue for the inclusion of these factors in helping to explain the realities of people

of color, which are often marginalized in the presentation of law and society generally.

Bringing attention to whose voice is included and reflected in the dominant narrative points to the larger issue of what is considered legitimate knowledge in scholarship. Bringing radicalized lenses to the discussion not only gives voice to marginalized communities, but also offers a critique of the dominant ideology of what is understood as valid knowledge. The inclusion of narratives of people of color in understanding and reforming law counters the dominant discourse and challenges the racial construction of whiteness as the norm by which standards and merit are judged. Universal truths and methods of reasoning as they have been constructed in law are inherently biased toward privileging the dominant group. This is exemplified in Freeman's (1995) analysis of court decisions on civil rights cases of the past century, asserting that the court often adopted a perpetrator perspective of racial discrimination – viewing “racial discrimination not as a social phenomenon but merely as the misguided conduct of particular actors” (p. 30) – a process that consequently negates the victim perspective of discrimination which names the inequities in outcomes and conditions. Accordingly, CRT scholars emphasize subjectivity in experiences, which account for the differences in reality (Delgado, 1989). Thus, where they previously have been silenced, centering on the lived experiences of people of color, voice is a crucial concern of analysis and methodology, both for legal policy and for educational research.

Related to this concept of voice, as it involves inclusion of people of color, is the role of the scholar and researcher. This aspect has been addressed by critical race scholars, who find themselves as outside scholars, marginal within their fields of study (Delgado, 1984; 1992). Specifically, scholars like Delgado (1984) have called attention to how the ethos of academia perpetuates dominant narratives and inadvertently work to de-legitimize and silence scholars of color, particularly in discussions concerning minority rights (i.e., civil rights). Additionally, this critique has led to employing methodology which includes reflexive processes of self-interrogation (particularly for White scholars), interrogating how scholars' worldviews inform, understand, and contextualize the textual product. These aspects of theory and methodology are at the heart of CRT.

Narratives, Stories, and Counterstories

Within the construct of voice for CRT is the application of storytelling. The critical race movement was founded on the works of scholars like Derrick Bell and Richard Delgado, who employed and established a narrative tradition, which departed from mainstream legal scholarship and was more commonly found in literature. These include the use of “parables, chronicles, stories, counterstories, poetry,

fiction, and revisionist histories” (Ladson-Billings and Tate, 1995). These narrative forms enable scholars to challenge dominant discourses that propagate stereotypes and distorted images of people of color. Appropriately termed counterstories, they exist as a method for conveying the untold experiences of marginalized groups, and also function as a tool for analyzing dominant narratives presumed to be reality. Among the reasons for utilizing counterstories include (1) community building; (2) challenging the perceived wisdom of the those at the center of society; (3) self-preservation for marginalized groups; (4) constructing a richer description of issues and stories that reality alone may be unable to provide; and (5) provide a context for understanding and challenging established belief systems (Delgado, 1989; Lawson, 1995 as cited in Solórzano and Yosso, 2002). While stories are a critical part of the critical race tradition, their use is highly contentious and is not universally accepted as appropriate for the academy. Still, CRT scholars argue that this method is valid and critiques originate from a positivist paradigm which privileges the normative nature of whiteness as the status quo. Further, they argue that counterstories are necessary particularly since reality is socially constructed and therefore stories from people of color's lived experiences can illuminate their circumstances, since the experiences of Whites are privileged in the dominant or majoritarian narrative. (Delgado, 1989).

Branches of CRT and Related Literature

Although CRT originated in the legal field, the theory has made its way into multiple fields including sociology and education. It has also been advanced to offer new insights and innovative critiques of previous theories as well as issues within these fields. Likewise, along with its growing use, CRT as a theoretical construct has expanded to incorporate the experiences of women (FemCrit), Latinas/os (LatCrit), Native Americans (TribalCrit), Asian Americans (AsianCrit), sexualities, genders, and other social identities that fall outside normative identities (QueerCrit) and even White scholars (WhiteCrit), which consists of studies of Whiteness and critically examines White privilege. These branches of CRT move beyond the Black/White binary that CRT critiques tended to follow understandably as their critiques were originally focused on civil rights legislation (Yosso, 2005). Yet, just as each of these expansions represents the varied issues and priorities within and between communities of color and socially marginalized groups, CRT scholars acknowledge, “the institution of slavery is foundational in American racism” and instructive for the struggle against multiple oppressive isms beyond the Black–White paradigm (Matsuda, 2002: 397). Therefore, CRT has become more nuanced as it is applied to multiple experiences of people of color. Finally, due to the

interdisciplinary focus, CRT literature overlaps with literature from fields such as women studies, ethnic studies, history, law, and sociology.

CRT in Education

CRT is used in education to frame how institutions maintain racism, sexism, and classism through structures, processes, and discourses. Ladson-Billings and Tate (1995) initially introduced CRT to the field of education, arguing that race has gone virtually untheorized in understanding issues of education. Together, they put forth CRT as a theory for understanding and overcoming school and educational inequity. Since then, a growing number of scholars began employing CRT in education to elucidate how educational practices and widely accepted theories minimize the importance of race and/or perpetuate a system which oppresses, subordinates, and marginalizes African-American, Latino(a), Asian-American, and Native-American students. Solórzano and Yosso (2001) contend that stereotypes and deficit frames used in elementary and secondary educational practices reproduce social inequities. Educational practitioners, who subscribe to these stereotypes, suggest that students of color are intellectually inferior, lack motivation, and are destined to become engaged in criminal activity. Additionally, educational practitioners have used deficit-framed perspectives to explain the academic underachievement of students of color. These frameworks often put the onus of blame on student and/or his or her culture for poor academic performance. For example, deficit frame employers may state, "African Americans (or Latinos) do not perform academically because African American (or Latino) culture does not emphasize education." Clearly, the unequal resources, the quality of education, and societal inequities are not considered in these deficit frames. As a result, these educational inequities may be manifested in low expectations for African-American, Latino(a), and Native-American students, tracking, remediating the curriculum, and limiting access to high-level or advanced placement courses (Solórzano and Yosso, 2001; Darling-Hammond, 2001).

Reframing Dominant Educational Theory through a Critical Race Perspective

CRT is used to reconceptualize dominant educational theory, which has been framed around European American cultural values. A noteworthy example is Yosso (2005), who uses CRT to reframe the Bourdieuean theory of cultural capital, which is a narrowly defined theory based on "White, middle class values, and is more limited than wealth – one's accumulated assets and resources" (p. 77). Yosso expands this view to include cultural capital more

representative of communities of color, which include aspirational capital (maintain hopes and dreams), familial capital (cultural knowledge and community history), linguistic capital (skill gained from communicating in more than one language), social capital (networks and community resources), navigational capital (negotiate social institutions), and resistant capital (knowledge and skill gained from opposing inequality). Collectively, Yosso refers to this new cultural capital frame as a community of cultural wealth.

Similarly, Patton *et al.* (2007) use a critical race frame to problematize racelessness (p. 43) in prevailing college student development theories. An example of a raceless college student development theory is in Chickering and Reisser's (1993) *Education and Identity*, which gives limited treatment to the role of race in identity development. Additionally, Patton *et al.* point out that some scholars have used homogeneously White samples to inform the theories they developed. Subsequently, these theories are applied to college students in general and do not account for the experiences of college students of color. Patton *et al.* suggest "Color-blind racism and racial indifference must consistently be challenged through exposing the manner in which racial advances often come at the cost of promoting or feeding into White self-interests" (p. 43).

Exemplified through the works reviewed, CRT can aid scholars in reshaping the discourse around students of color in educational research. Previously, the norms and values of dominant cultural groups were juxtaposed with those of communities of color, resulting in an assimilationist approach to dealing with students of color. This approach suggests that students must discard cultural values not aligned with the dominant frame in order to succeed. As a result, students' (of color) culture is to blame for their academic underperformance and not the system maladaptive to multiple points of view. CRT has given scholars and practitioners the tools to reevaluate theoretical perspectives and educational practices from multiracial and multicultural perspectives.

CRT in Empirical Educational Research

Educational researchers have effectively utilized CRT to frame empirical research. For example, CRT-based research has been used to explain educational inequities in K-12 schooling, to evaluate how educational policies shape the educational pipeline from primary education through graduate school, the importance of family and culture in education, and the role of racial micro-aggressions in the perceived college-campus climate of students of color. In these contexts, CRT goes beyond simply describing a problem or challenges associated with educational outcomes; it explicates the sources of persisting problems, which is very often grounded in systemic racism and oppression.

In examining educational policy and educational pipeline Solórzano and Ornelas (2004) use CRT as a lens to investigate enrolment patterns of advanced placement (AP) courses generally in the top 50 California high schools and specifically in four Los Angeles Unified School District high schools. The findings suggest Latino and African-American students are disproportionately underrepresented in these AP classrooms. Solórzano and Ornelas apply CRT to suggest that AP courses reproduce societal inequities because Whites and some Asian Americans are afforded advantages in college admissions due to the boost that AP courses give to grade point averages. As a result, African-American and Latino students, who are underrepresented in AP courses, are less competitive in the college-application process. Solórzano and Ornelas are encouraged by the fact that African-American and Latino families are demanding better educational opportunities and have filed lawsuits against school districts and the University of California regents.

The disparities found in AP course enrolment is further investigated in the study by Solórzano *et al.* (2005) study, which employs CRT to examine how educational structures, policies, and practice impact Latino (a) students' educational attainment in the United States. The quantitative analysis shows that for every 100 Latino(a) students who begin primary education, only 48 graduate from high school and only 10 graduate from a 4-year college or university. Solórzano *et al.* (2005) focus the inquiry on California, which has had clear policy changes in anti-affirmative action legislation. In California, for every 100 Latino(a) high school graduates, only 40 continue on to public 4-year institutions with 30 attending 2-year community colleges. Using CRT as a lens, Solórzano, Villapando, and Oseguera suggest that policies designed to eliminate race-based-admissions decisions and emphasize merit are in fact still emphasizing race. For instance, persistent inequities in the educational and social system make Latino(a) students more likely to drop out of high school – and if they graduate, to be diverted to 2-year community colleges. In this regard, presumably race neutral policies and approaches perpetuate the achievement gap between Whites and underrepresented racial minorities.

Focusing on ethnographic studies of Latino families and education, Villenas and Deyhle (1999) use CRT to position schools' and larger society's negative perceptions of cultural differences in family socialization and education within a framework of power relations and the castification of Latinos in the United States. They find that Latino families play key roles resisting assaults on Latino culture. The history of Latinos as taught from the dominant White ideology, propagates views of Latinos as illegal immigrants and foreigners. The ethnographic studies demonstrate how Latino families can take collective action to transform the schooling of their children. Clearly, CRT can be used to identify counternarratives that "challenge

dominant culture's 'stock stories' which legitimize the world views of White, upper-class United Statesians" (p. 441).

Finally, A CRT perspective can also elucidate the atmospheric nature of racism in educational institutions as hostile climates for students of color. Solórzano *et al.* (2000) use CRT to describe the experiences of African American students attending predominantly White research universities. Participants in this study shared their experiences with everyday subtle forms of racism and stereotypes (micro-aggressions). Examples of these micro-aggressions include feeling invisible as one of only a few African Americans in a particular college classroom, faculty having low expectations, challenges in joining study groups with non-African American students, and feeling a sense of hyper-surveillance because of perceived criminality. Micro-aggressions created a negative campus racial climate, which led to feelings of frustration, isolation, self-doubt, and exhaustion. Many students in this study coped with the negative campus racial climate by creating counterspaces where they could be surrounded by supportive peers and collectively reject the notions of intellectual inferiority and stereotypes of deviance or criminality.

Implications of CRT for Future Educational Scholarship

While CRT has emerged as a powerful framework for understanding educational issues, it is yet to be fully realized. Scholars who employ a CRT framework most notably utilize the concept of permanence of racism and voice scholarship through counter-storytelling. For example, voice scholarship has included examining the experiences of students of color as well as the inclusion of scholars of color in theorizing space. However, the presence of an internal struggle over the materialist concerns of CRT has pressing implications for CRT scholars. Namely, it points to CRT inquiry, which has transcended from its material origins to its ideological concerns. Yet, these two forms of inquiry are inextricably connected, merely posing as two sides of the same coin. For example, while Crenshaw (1995) acknowledges identities as ideological in nature, she argues their consequences on lived experiences necessitate their consideration, in order to explicate and positively change ones marginalized condition by embracing and enacting on ones identity. In other words, although race is socially constructed and not a social fact in and of itself, the material consequence of race convey its social significance. Therefore, challenging both the social construction of race and the enactment of power based on race is significant. Still, the concern that scholarship in which CRT engages in may never reach those who are most affected, the very populations for which CRT is meant to account for (Ladson-Billings, 1999), exists as a valid concern which scholars must consciously engage

themselves in. One challenge then is in examining and asserting how ideological struggles are linked to material consequences. Harris' (1995) concept of Whiteness as property, further articulated by Ladson-Billings and Tate (1995), presents the opportunity to identify (ideological and structural) processes that consequently result in concrete differences in schooling opportunities. Yet, it has been underutilized, presenting a gap for future work to fill.

Additionally, Bell's (1980) concept of interest convergence also has important implications for understanding numerous issues, including the discourse centered on Asian Americans, especially within the context of affirmative action debates. A CRT approach which recognizes and examines how discursive practices construct, impose, and empower identity as well as shape definitions of minority, diversity, or social justice are important for understanding contemporary debates over the allocation of college-admission slots. First, it can challenge the social construction of race and further examine how particular definitions of race are privileged in certain instances and used to inform a limited definition and application of diversity that ultimately support conservative agendas (Omi and Winant, 1994). A CRT perspective can also elucidate the larger social context in which education issues are understood, explicating how foreign relations often shape experiences of Asian Americans and people of color in the US that ultimately support conservative agendas (Omi, 2008). For example, the peril of Asian Americans as invading and overrunning higher education institutions requires a critical race approach to elucidate that they mirror the larger societal sentiments seen in foreign relations. Finally, CRT and interest convergence can further aid, particularly in understanding how, despite these sentiments, Whiteness can be thrust upon Asian Americans (Omi, 2008) when it aids in the larger social and political interests of Whites. The case of Asian Americans serves as just one example of how CRT theoretical constructs can elucidate and explicate how ideologies of race and discursive practices have real consequences and are directly connected to material interests.

Methodologically, in part due to its role in advancing voice scholarship, CRT is less often applied to examining quantitative data. This is understandably so as quantitative research, until recently, was viewed as the only legitimate scientific method. The struggle over legitimate scientific inquiry and use of qualitative research has a tradition of giving voice to the experiences of marginalized populations. While CRT can be a useful framework for understanding educational inequity, the underutilization of CRT to examine quantitative work presents a glaring gap. The above-reviewed work of Solórzano and Ornelas (2004) and Solórzano *et al.* (2005) serve as examples of how CRT can be used to understand the hard facts of inequalities. We appreciate the epistemological struggle to legitimize qualitative research; therefore application of CRT to quantitative data must be approached cautiously. It is

imperative to include continuous interrogation of how numbers can be construed and recognize inherent biases in how numbers are collected, categorized, and understood. Such can be seen in the aggregation of Asian Americans as a monolithic group, which presents the misguided notion that they are outwhiting Whites in higher education (CARE Report, 2008). However, coupled with a narrative tradition, the continued use of CRT to examine quantitative data can further elucidate education issues.

To conclude, education scholars employ CRT to understand school inequity and educational issues while advancing the evolution of CRT in significant ways. Given its origins in and departure from critical legal theories, aimed at understanding the interplay of race, racism, and the law, a critical race analysis is appropriate for the field of education on both philosophical and historical grounds. When considering that law was used to, first, ban African Americans and other marginal groups from obtaining education; then, institute segregation and later, repeal initiatives like affirmative action aimed at addressing the legacy of racism, the use of CRT in education is logical if the ultimate goal is to overcome societal inequity. While these reasons alone represent a compelling case for CRT, the need for a theory which advances race as an analytical tool for understanding issues in education treads much deeper. It calls into question the ideologies and jurisprudence which informs and sustains these laws. Much like the legal scholars of color who were instrumental in creating a CRT framework, the allure of CRT for scholars in education is the possibility for liberation. Specifically, we are drawn to idea that although schools and schooling practices (curriculum, instruction, pedagogy, etc.) are sites and tools of social reproduction, we remain hopelessly optimistic in achieving a transformative education: one that is grounded in our experiences, inclusive of our cultural knowledge, and intentionally fosters social praxis for the emancipation of marginalized communities. For these reasons, scholars of color in education find solace in the discontent, struggle, and discourse of legal scholars of color within their field, (not merely in their collective identity as people of color, but more principally) in their criticisms, yet unyielding commitment in the ideals of and belief in the very institution in which they hope to transform (theoretically, methodologically, and practically). Finally, and perhaps most importantly, CRT challenges scholars to interrogate their own socialization into the academy and confront their possible roles in reproducing dominant discourses. Therefore, CRT has far-reaching implications for theory, research, and practice in education.

Bibliography

- Aleinikoff, A. T. (1991). A case for race-consciousness. *Columbia Law Review* 91, 1060–1123.
 Bakke, V. (1978). *Regents of the University of California v. Bakke*, 438 U.S. 265.

- Bell, D. A. (1976). Serving two masters: Integration ideals and client interests in school desegregation litigation. *The Yale Law Journal* **84**, 470–516.
- Bell, D. A. (1980). Brown v. Board of education and the interest convergence dilemma. *Harvard Law Review* **93**, 518–533.
- Chickering, A. W. and Reisser, L. (1993). *Education and Identity*, 2nd edn. San Francisco, CA: Jossey-Bass.
- Crenshaw, K. (1988). Race, reform, and retrenchment: Transformation and legitimization in antidiscrimination law. *Harvard Law Review* **101**, 1331–1387.
- Crenshaw, K. (1995). Mapping the margins: Intersectionality, identity politics and violence against women of color. In Crenshaw, K., Gotanda, N., Peller, G., and Thomas, K. (eds.) *Critical Race Theory: The Key Writings that Formed the Movement*, pp 357–383. New York: New Press.
- Crenshaw, K., Gotanda, N., Peller, G., and Thomas, K. (1995). *Critical Race Theory: The Key Writings that Formed the Movement*. New York: New Press.
- Dalton, H. L. (1995). The clouded prism: Minority critique of the critical legal studies movement. In Crenshaw, K., Gotanda, N., Peller, G., and Thomas, K. (eds.) *Critical Race Theory: The Key Writings that Formed the Movement*, pp 80–84. New York: New Press.
- Darling-Hammond, L. (2001). New standards and old inequalities: School reform and the education of African-American students. *The Journal of Negro Education* **69**, 263–285.
- Delgado, R. (1984). The imperial scholar: Reflections on a review of civil rights literature. *University of Pennsylvania Law Review* **132**, 561–578.
- Delgado, R. (1989). Storytelling for oppositionists and others: A plea for narrative. *Michigan Law Review* **46**, 665–676.
- Delgado, R. (1992). The imperial scholar revisited: How to marginalize outside writing, ten years later. *University of Pennsylvania Law Review* **140**, 1349–1372.
- Delgado, R. and Stefancic, J. (1993). Critical race theory: An annotated bibliography. *Virginia Law Review* **79**, 461–516.
- Delgado, R. and Stefancic, J. (2001). *Critical race theory: An introduction*. New York: New York University Press.
- Freeman, A. D. (1995). Legitimizing racial discrimination through antidiscrimination law: A critical review of supreme court doctrine. In Crenshaw, K., Gotanda, N., Peller, G., and Thomas, K. (eds.) *Critical Race Theory: The Key Writings that Formed the Movement*, pp 29–45. New York: New Press.
- Harris, C. I. (1995). Whiteness and Property. In Crenshaw, K., Gotanda, N., Peller, G., and Thomas, K. (eds.) *Critical Race Theory: The Key Writings that Formed the Movement*, pp 357–383. New York: The New Press.
- Hunt, A. (1986). The theory of critical legal studies. *Oxford Journal of Legal Studies* **6**, 1–45.
- Ladson-Billings, G. (1999). Just what is critical race theory, and what's it doing in a nice field like education? In Parker, L., Deyle, D., and Villenas, S. (eds.) *Race is... Race Isn't: Critical Race Theory and Qualitative Studies in Education*, pp 7–30. Boulder, CO: Westview Press.
- Ladson-Billings, G. and Tate IV, W. F. (1995). Toward a critical race theory of education. *Teachers College Record* **97**, 47–68.
- Lawson, R. J. (1995). Critical race theory as praxis: A view from outside to outside. *Harvard Law Journal* **38**, 353–370.
- Matsuda, M. (1995). Looking to the bottom: Critical legal studies and reparations. In Crenshaw, K., Gotanda, N., Peller, G., and Thomas, K. (eds.) *Critical Race Theory: The Key Writings that Formed the Movement*, pp 63–79. New York: New Press.
- Matsuda, M. (2002). Beyond, and not beyond, black and white: Deconstruction has a politics. In Valdes, F., Culp, J. M., and Harris, A. P. (eds.) *Crossroads, Directions, and a New Critical Race Theory*, pp 393–398. Philadelphia, PA: Temple University Press.
- Omi, M. (2008). Asian-Americans: The unbearable whiteness of being? *The Chronicle of Higher Education* **55**(5). <http://chronicle.com/article/Asian-Americans-the/14271> (accessed October 2009).
- Patton, L. D., McEwen, M., Rendon, L., and Howard-Hamilton, M. (2007). Critical race perspectives on theory in student affairs. In Harper, S. R. and Patton, L. D. (eds.) *Responding to the Realities of Race on Campus, New Directions for Student Services*, No. 120, pp 39–53. San Francisco, CA: Jossey-Bass.
- Solórzano, D. G., Ceja, M., and Yosso, T. J. (2000). Critical race theory, racial microaggressions, and campus racial climate: The experiences of African American college students. *Journal of Negro Education* **69**, 60–73.
- Solórzano, D. G. and Ornelas, A. (2004). A critical race analysis of Latina/o and African American advanced placement enrollment in public high schools. *High School Journal* **87**, 15–26.
- Solórzano, D. G., Villapando, O., and Oseguera, L. (2005). Educational inequities and Latina/o undergraduate students in the United States: A critical race analysis of their educational progress. *Journal of Hispanic Higher Education* **4**, 272–295.
- Solórzano, D. G. and Yosso, T. J. (2001). From racial stereotyping and deficit discourse toward a critical race theory in teacher education. *Multicultural Education* **9**, 2–8.
- Solórzano, D. G. and Yosso, T. J. (2002). A critical race counter story of race, racism and Affirmative Action. *Equity and Excellence in Education* **35**, 155–168.
- Tate, W. F. (1997). Critical race theory and education: History, theory, and implications. *Review of Research in Education* **22**, 195–247.
- Villenas, S. and Deyhle, D. (1999). Critical race theory and ethnographies challenging the stereotypes: Latino families, schooling, resilience, and resistance. *Curriculum Inquiry* **29**, 413–445.
- White, G. (1972). From sociological jurisprudence to realism: Jurisprudence and social change in early twentieth-century America. *Virginia Law Review* **58**, 999–1028.
- Yosso, T. J. (2005). Whose culture has capital? A critical theory discussion of community cultural wealth. *Race Ethnicity and Education* **8**, 69–91.

Further Reading

- Bernal, D. D. (2002). Critical race theory, Latino critical theory, and critical raced-gendered epistemologies: Recognizing students of color as holders and creators of knowledge. *Qualitative Inquiry* **8**, 105–126.
- Bonilla-Silva, E. (2006). *Racism without racists: Color-blind racism and the persistence of racial inequality in the United States*, 2nd edn. Lanham, MD: Rowan & Littlefield.
- Crenshaw, K., Gotanda, N., Peller, G., and Thomas, K. (eds.) (1995). *Critical Race Theory: The Key Writings that Formed the Movement*. New York: New Press.
- Delgado, R. (ed.) (1995). *Critical Race Theory: The Cutting Edge*. Philadelphia, PA: Temple University Press.
- Dixon, A. D. and Rousseau, C. K. (eds.) (2006). *Critical Race Theory in Education: All God's Children Got a Song*. New York: Routledge.
- Feagin, J. R. (2006). *Systemic Racism: A Theory of Oppression*. New York: Routledge.
- Harper, S. R. and Patton, L. D. (2007). *Responding to the Realities of Race on Campus New Directions for Student Services*, No. 120. San Francisco, CA: Jossey-Bass.
- National Commission on Asian American and Pacific Islander Research in Education (2008). *Asian Americans and Pacific Islanders. Facts, Not Fiction: Setting the Record Straight*. New York: The College Board.
- Omi, M. and Winant, H. (1994). *Racial Formation in the United States: From the 1960s to the 1990s*, 2nd edn. New York: Routledge.
- Parker, L., Deyle, D., and Villenas, S. (eds.) (1999). *Race is... Race Isn't: Critical Race Theory and Qualitative Studies in Education*. Boulder, CO: Westview Press.
- Solórzano, D. G. (1998). Critical race theory, race and gender microaggressions, and the experience of Chicana and Chicano scholars. *Qualitative Studies in Education* **11**, 121–136.
- Solórzano, D. G. and Yosso, T. J. (2002). Critical race methodology: Counter-storytelling as an analytical framework for education research. *Qualitative Inquiry* **8**, 23–44.

Cultural Diversification and Japanese Education: Social Constructions of the New Diversity

R Tsuneyoshi, The University of Tokyo, Tokyo, Japan

© 2010 Elsevier Ltd. All rights reserved.

Social Constructions of Ethnic Differences in the Japanese Context

Research on ethnicity suggests that what societies define as ethnic differences are socially constructed, not a given reality (Barth, 1969; Jenkins, 1997). The arrival of newcomers often pressurizes existing social boundaries to change, and provides opportunities to observe the social reconstruction of difference at work. Kaplan (2001) shows how the Jewish Ethiopians, newcomers in a predominantly white society, Israel, are depicted as black or non-black, or whiter than other groups but not quite white, depending on who the object of comparison is and who is doing the comparison. Studies on American immigration have also suggested that successive waves of immigration have forced the reconstruction of what it means to be white or black (Fass, 1989; Oakes, 1985).

Japan provides an interesting case in this sense, since it has often been seen as one of the most homogeneous-minded societies in the world, and yet from the 1980s, it has experienced its largest postwar influx of new foreigners which challenge this image of homogeneity. One does not have to go far back to find an assumption of homogeneity in the Japanese official discourse (e.g., the former Prime Minister Nakasone's comment that "Japan is a mono-ethnic nation" in 1986), a discourse that has its counterparts in other nation-states (Hickman and Walter, 2001).

Indeed, needless to say, Japan has never been totally homogeneous. In the northern districts, there are the indigenous populations, the Ainu, and though ethnically Japanese, the descendants of Japan's feudal outcaste population have constituted one of Japan's largest minorities. A title of a representative work on this population, *Japan's Invisible Race* (DeVos and Wagatsuma, 1966) captures the nature of invisibility and discrimination of this minority status in Japan. The colonization of Korea in 1910 also brought many Koreans into Japan as forced labor, or as voluntary entrants.

A characteristic of Japanese cultural diversity has been that the visible indicators of difference (e.g., skin color) has tended to be more subtle than many other visibly diverse societies, since the major ethnic/racial minorities in Japan, such as the Koreans and Chinese, are not visibly distinguishable from ethnic Japanese, and adhere to relatively similar cultural heritages (e.g., Confucianism, Buddhism). Indeed, for a large part of the postwar period,

foreigners were in large part identified with a single ethnic group, the Koreans in Japan (see Figure 1).

As recent as the late 1980s, those of Korean nationality comprised almost 70% of the registered foreigners residing in Japan. These foreigners were mainly the descendants of the Koreans who came, either voluntarily or by force, because of the colonization of Korea in 1910. With the end of World War II, Koreans lost their Japanese nationality, though a significant number were led to remain in Japan. Since Japanese nationality is gained either through naturalization or through one's Japanese parent (previously it was limited to the male parent; the nationality law was revised in 1985 to make it possible to acquire citizenship from either parent), this led to a situation in which there are a number of Korean nationals within Japan, who were born and raised in Japan, and who are the counterparts of later-generation immigrants in other countries, but who are foreign in nationality. This population has been called sociologically Japanese by Kajita (1998) since they look like Japanese, speak Japanese, and behave Japanese, but are foreign in nationality.

Since the Koreans in Japan cannot be distinguished by the color of their skin from the majority Japanese, and nationality is hidden in daily life, what signals that one is ethnically Korean is a Korean name. However, with the pressure to assimilate, since the vast majority of Koreans use their Japanese names instead of their Korean names, there is often no visible indicator differentiating the later-generation Koreans in Japan from the majority ethnic Japanese. Ethnicity in Japan is thus often concealed on the surface, visible on a daily basis only in those areas where there is a concentration of a certain population. It is thus not a surprise that the Korean movement has tended to emphasize the importance of using one's real Korean name. Where visible indicators of ethnicity are scarce, the use of an alien name is a powerful indicator of ethnic pride.

Due to the situation above, being ethnically different in Japan has tended to be identified with being foreign, not with being an ethnic minority. Nationality is identified with ethnicity, and although the majority of Japanese are aware of differences based on nationality, ethnicity is assumed to be a reflection of nationality. This paradigm, however, is highly problematic when applied to ethnic minorities living for generations in Japan.

Yet, since many Koreans in their third, fourth, and later generations, whose daily language is Japanese, are

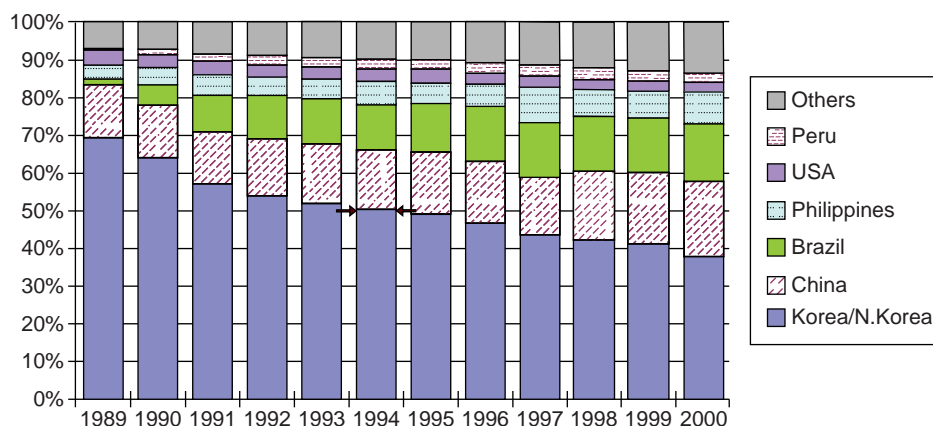


Figure 1 Number and distribution of registered foreigners by nationality and the region of origin. Source: Immigration Bureau, Ministry of Justice, Japan.

still foreign nationals (though an increasing number are Japanese nationals through naturalization and as children of intermarriages between Koreans and Japanese), they are, in the Japanese mind, foreigners. Such reasoning can be seen everywhere, not the least in governmental documents which refer to these populations as foreigners and count the ethnically different populations based on nationality (e.g., based on the assumption that Korean nationality means Korean ethnicity). School statistics are the same, recording by nationality, thus there is no recognition of an ethnic Japanese (e.g., Korean-Japanese); the categories available are Japanese and foreigners (foreign nationals). This dichotomous understanding of ethnicity as nationality, as well as the invisibility of Japanese ethnicity, was the background against which the Japanese returnees in the 1970s and the new foreigners in the 1980s emerged on the cultural diversity scene.

The Japanese Returnees

Where visible differences are not stark, subtle differences can become a major issue as seems to be the case of the Japanese returnees. These are children of Japanese living abroad temporarily (e.g., children of businessmen stationed abroad), who spent part of their school age in another country and returned to Japan. These children were Japanese in nationality, but if they were brought up in a local school abroad, they often had internalized parts of the culture of the host country and thus acted non-Japanese or were able to speak a foreign language, and were different from the majority Japanese.

Around the 1960s, Japanese companies moved out into various parts of the world, stationing their employees in foreign countries. More Japanese school-age children started to grow up abroad, and they returned and were called the *kikokushijo* (the Japanese returnees). The number of school-age children at the compulsory-education

level who were residing in another country for 3 months or more (and who were not the children of Japanese permanently residing abroad), was a mere 6662 in 1971, shooting up above 20,000 in less than a decade, and leveling out near 50,000 from around 1990. In 2007 April, there were 58,304 such children abroad. When these children come back to Japan, they become the *kikokushijo*, more than 10,000 in a school year.

The returnees quickly became a social problem. Despite the fact that they came from middle-class (often elite) business-person families, and were quite Japanese in the regular sense of the word (nationality), many *kikokushijo* came home to find that Japanese society was not ready for them (e.g., forced to attend a grade lower than their age). Although schools in districts with a concentration of the Koreans in Japan or other historically discriminated-against groups had experience with issues of cultural diversity and human rights issues, this was not the case for most other areas. While the *kikokushijo* were undergoing a personal cultural shock upon (re)entry to Japan, the emergence of a different type of Japanese child presented a cultural shock for many educators and researchers. The Intercultural Education Society (*Ibunkakan Kyoiku Gakkai*), for example, started in 1981 with many of its members researching *kikokushijo*. Today, the research themes have diversified as the new foreigners have arrived, and more interest has turned to other populations such as foreign students. The *kikokushijo* provided an inspiration for educators in international understanding education as well.

This also coincided with the increasing efforts of the government to emphasize internationalization. Various measures were institutionalized to meet the needs of Japanese children abroad. For example, schools following the national curriculum were erected in developing countries, and Japanese teachers were sent by the government, exporting Japanese compulsory education abroad. These teachers came back and practiced international understanding education; the Association of Japanese

Returnees (*Zenkoku Kaigaishijyo Kyoiku Kenkyu Kyogikai*) was established in 1974, and today it has added international understanding education to its title (*Zenkoku Kaigaishijyo Kyoiku Kokusairikai Kyoiku Kenkyu Kyogikai*) reflecting a widening of focus from its original narrow focus on returnees (the organization is a nonprofit organization (NPO) since 2005). Schools started to adopt special admission procedures for returnees (*kikokushijyo waku*) at various levels. Special schools were erected to focus on returnee education. In public schools, Japanese-language classrooms were erected and provided pull-out instruction for returnees who required assistance in Japanese.

The 1970s–1980s was the era of increased internationalization, and the term seemed to acquire instant media popularity. Government policies called for the internationalization of Japan, the expansion of foreign student exchange, the teaching of communicative English, and the strengthening of international understanding. The national rhetoric surrounding the *kikokushijyo* shifted from *kikokushijyo* as objects to be assimilated and saved, to *kikokushijyo* as pioneers of internationalization, and assets for international understanding education. As children of Japanese businessmen who had led Japanese society to its economic recovery, these *kikokushijyo* had powerful agents lobbying for their cause. Goodman (2003: 177) argues that “no other group’s image and status have changed as rapidly.” Once problems, some members of the *kikokushijyo* group were now looked up to as champions of internationalization, though behind these privileged few were many more who could not fully adjust to Japanese society.

The New Foreigners

The *kikokushijyo*, despite the fact that they are Japanese both in terms of ethnicity and nationality, have served as a reference point in conceptualizing education for the new foreigners. On the one hand, these new foreigners are foreigners like the Koreans in Japan before them. Yet, in other contexts, they are put together with the Japanese returnees as agents of internationalization. The manner in which the concept of the new foreigners shifts, is one indication of how the majority society is struggling to define the emerging cultural diversity.

The new foreigners were very diverse in terms of their purpose and country of origin. They came as foreign laborers, Chinese returnees, Indochinese refugees, and spouses of Japanese. From around the 1980s onward, certain districts became associated with a concentration of a certain group of new foreigners. Most famous perhaps are the districts with Latin American of Japanese descent, who are Japan’s largest new-foreigner group, and who typically come as foreign labor. For example, Oizumi and Ota in the northern Gunma prefecture, and Hamamatsu City in Shizuoka are well known (Matsuo, 1997;

Noyama, 1997). Certain public housing too, such as the prefectural public housing Ichō Danchi in Kanagawa prefecture, relatively near the former refugee-placement center, is now known to have high concentrations of the Indochinese as well as the Chinese returnees (Shimizu and Shimizu, 2001). There are, of course, the Ainu in northern Japan, and the children of international marriages of US army persons in Okinawa, or urban areas with concentrations of Asian workers (Okuda and Tajima, 1991, 1993), and rural communities with Asian brides (Shibata, 1997). Research on such districts is starting to accumulate (Ebashi, 1993; Watabe, 1995; Komai and Watodo, 1997; Shiramizu, 1996).

Compared to the existing cultural minorities, these new foreigners tended to be more visible at first sight. First of all, if they came straight from their country, there were language differences. In some cases, as in Southeast Asians, the skin color immediately signaled a difference. The Islamic residents constitute a small but visible population, whose religious customs and clothing tend to be very alien from the majority Japanese. At the same time, the new foreigners, as were the Koreans and Chinese in Japan before them, are predominantly from Asia. In the case of Latin Americans, because of the Japanese governmental policies that placed preference on Japanese descent, there were many who were difficult to distinguish by their skin color. Due to this invisibility, there are observations that reinstituting markers of difference was one reaction to dealing with Japanese expectations of homogeneity. For example, Brazilian youth have been observed to deliberately exhibit their Brazilian-ness by the way they dress and act, in order to avoid the expectations that would be imposed on them if people assumed they were Japanese (Tsunda, 2000).

The inflow of new Koreans also resulted in a diversification of the Korean category. There were not only the existing Koreans in Japan, but the new Koreans. The same held true for Chinese. The Koreans in Japan were in their later generations, born and raised in Japan. The new foreign Koreans were foreign laborers, spouses of Japanese, and foreign students, among others. Attempting to distinguish the differences among foreigners, researchers began calling the existing cultural minorities the old-comers (*orudo kama*) and contrasted this group to the new foreigners by calling them the newcomers (*nyukama*) based on the period of their entry, which was linked to the reason of their residence in Japan. However, this meant that Latin American children, a child of an intermarriage between a Japanese and foreign national, and a foreign worker from the Middle East are all together in the same category. It also meant breaking up members of the same ethnic group into different categories, such as the case with newcomer Koreans and oldcomer Koreans. Other attempts include distinguishing between permanently residing foreigners and the others, the permanently residing foreigners including the Koreans in Japan.

Another attempt at distinguishing between the foreigners was to link the concept to Japanese as a second language in education. The Japanese government started to compile statistics on the education of the children of the new foreigners since the revision to the Immigration and Refugee Registration Law in 1990, after which there was an increase in the number of Latin Americans of Japanese descent. Since this group was more likely than others to bring their children with them, there was an increase in the number of Latin American (mother tongue Spanish or Portuguese) children in Japanese public schools (see Figure 2). The government devised a term, foreign pupils and students who require instruction in the Japanese language and defined it as “those pupils or students who cannot sufficiently manage daily conversations in Japanese, or who can hold daily conversations, but who lack grade-level instructional vocabulary, and thus have difficulty in participating in learning activities, and who require instruction in the Japanese language.” In 2006, this population totaled to 22,413 students from elementary to high school. The hidden contrasts are those that do not require instruction in Japanese language, which presumably refers to the oldcomers as well as the Japanese.

Yet, reality seems to be outpacing the utility of this categorization as well. Though at one point, these figures were used to capture the number of newcomer children in Japanese public schools, there is an increasing number of newcomers who are raised in Japan and whose Japanese at a daily level is not a problem, and who do not appear in these statistics. Like the ethnic Korean who acquires a Japanese nationality and becomes Japanese, this concept creates an area where certain children fall through the gaps and disappear from the social categories of difference.

All of these experiments in conceptualizing the new diversity seem destined to change. For example, even if

the initial waves of immigration in the 1980s could be called the newcomers in contrast to existing ethnic minorities, albeit in a very gross way, what should the society call the migrants who are entering Japan today? The new newcomers? Many of the newcomer children are now not that new, since as in the example of Indochinese refugees, there is a generation that is now born and raised in Japan.

Internationalization as an Umbrella Term

If the categories of newcomers and oldcomers were terms that tried to distinguish between foreigners, there have also been attempts to categorize the newcomers using the label of internationalization. Needless to say, Japanese returnees and the newcomers are different in many ways, not the least because of differences in nationality and ethnicity. Japanese returnees are mainstream in both these factors. Yet, ironically, the newcomers are often placed under the same category as the Japanese returnees using the umbrella term of internationalization. For example, the 1974 May Central Council for Education (Japan’s main governmental committee for reform) report on International Exchange in Education, Academics, and Culture (*Kyoiku, gakujutsu, bunka ni okeru kokusai koryu nitsuite*) called for fostering Japanese who can become part of the international community. The report referred to the promotion of international understanding education, strengthening foreign-language (in effect English) education, and addressing issues seen as relating to internationalization, including the treatment of Japanese returnees and Japanese schoolchildren abroad, foreign-student education, assistance to developing countries, and Japanese-language lessons to foreign students. Since those foreigners who require instruction in the Japanese language are the

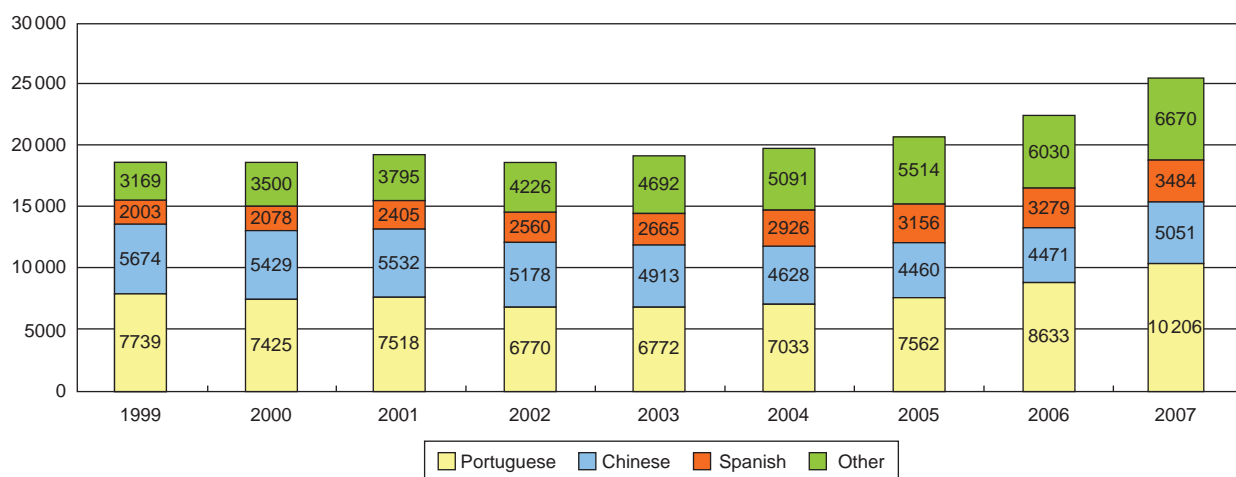


Figure 2 Numbers by mother tongue from 1999 to 2006. Yellow, Spanish; light green, Chinese; dark green Portuguese; and brown, others. Results are from the 2006, September Survey on the Acceptance, etc. of Foreign Pupils and Students Who Require Instruction in the Japanese Language (*Nihongo shido ga hitsuyona gaikokujin jidoseito no ukeire jokyō nado nikansuru chosa* (H18) *no kekka*). Japan: Ministry of Education, Culture, Sports, Science and Technology.

newcomers, it is this group of foreigners that is being associated with internationalization in this context.

This line of thinking has persisted. For example, the 1996 July report of the Central Council for Education on *The Education of Our Country Towards the 21st Century* (21 seiki o tenboshita waga kuni no kyoiku no arikata nitsuite) contained sections on international understanding education, English, Japanese returnees and Japanese schoolchildren abroad, and the Japanese language instruction of foreigners staying in Japan as related to internationalization. Now, if there is a similarity between the Japanese returnees and the new foreigners, it is the experience of growing up in another culture. Of course, this does not hold for the newcomer foreign children who grew up in Japan, but even in this case, if the parent(s) is from another country, he/she/they would be able to talk about their culture. The common image is someone coming from abroad. Thus, the models that have been used to understand Japanese returnees, in other words, acquisition of Japanese language (*nihongo shido*), learning Japanese ways (*tekio shido*), and respecting the traits that the child acquired in another culture as assets for international understanding, have all been applied to the newcomer education (Mabuchi, 2002: 92).

The link between the two categories can be seen in another aspect. The unit overseeing the returnees took charge of the foreigners requiring Japanese-language assistance. In 2001, when the organization of the ministry of education was restructured, the overseas and returnee unit (*Kaigai shijyo kyoiku ka*) was restructured and became part of the international-education unit (*Kokusai kyoiku ka*). Correspondingly, the designated schools formally labeled the cooperative schools for returnee education (*kikokushijyo kyoiku kenkyu kyoryoku ko*) was restructured as the districts that promote the internationalization of education with the returnees and foreign pupils and students (*kikoku/gaikokujin jido seito to tomoni susumeru kyoiku no kokusaika suishin chiki*) to include foreigners (Mabuchi, 2002). The national intracollege center for international education was formally established to research Japanese children abroad and returnees. Founded in 1978, its previous name was the Center for Japanese Returnee Education (*Kaigai shijyo kyoiku senta*). With the internationalization of Japanese society (from its homepage, it was renamed the Center for International Education (*Kokusai kyoiku senta*) in 2002, a unit was erected for the education of foreign children. As mentioned previously, the association for returnees (*Zenkoku Kaigaishijyo Kyoiku Kokusairi Kyoiku Kenkyu Kyogikai*) that gathers international-understanding-education-oriented educators, which formally referred to only the Japanese children abroad and returnees, added international understanding education to its title to enlarge its focus to the foreigners.

Although understanding foreigners using this internationalization framework works relatively well for the

newcomers who come from another country, what it has trouble incorporating is the oldcomers who have lived in Japan preceding the waves of the so-called internationalization. Japanese international understanding education continues to be criticized for its inability to incorporate the oldcomers (Nakajima, 1988). These latter groups are more strongly associated with human-rights education, emancipation education (*kaibo kyoiku*), or other practices that have a strong human-rights focus. In comparison, newcomers are more easily incorporated into the above-stated framework.

At the same time, though Japanese returnees and the newcomers are often placed under the same category of internationalization, there is a clear status difference between the two in terms of nationality and social class, which also has its counterpart in people's perceptions. Sekiguchi (2003: 92–93) compares the image Japanese have of Japanese returnees and the Latin American of Japanese descent, and notes that though both are linked to Japanese roots, the former is identified more with a first-world image, English, and an international elite (e.g., active, smart, progressive), while the latter is identified with the image of a developing country, Portuguese, and foreign worker.

Conclusion

In another paper, the author did a content analysis of how foreigners were portrayed in the latest version of the elementary school social studies textbook (2006 version) of the company which has more than half the national market share (Tsuneyoshi, 2007). There was an association between the type of foreigner being discussed in the unit, and the concept that was associated with it. The newcomer foreigners were linked to an image of internationalization and basically seen as people who came and left. On the other hand, when the foreigners discussed were the oldcomers, they tended to appear in sections which dealt with welfare, peace, human rights, and discrimination. These were the foreigners that stayed in Japanese society, and were associated with human-rights issues.

The entry of the new foreigners into the Japanese cultural diversity scene has brought about changes in how cultural boundaries are drawn. The ways in which the new foreigners are sometimes foreigners based on their nationality, and sometimes agents of internationalization together with a segment of the majority Japanese population, shows how such categories can cut across ethnic, racial, religious, and even nationality divisions. In a society where ethnicity or race tends to be the dominant dividing factor between culturally different groups, such as the United States, it may be hard to conceive of a category that cuts across the same ethnic/racial group, for example, as in the case of the newcomer Koreans and

the oldcomer Koreans. It may also seem absurd to put together two groups together, when the only thing they seem to have in common is their experience in another culture. However, this is the logic that places the newcomers in the same internationalization category as Japanese returnees, despite the fact that they differ in their nationality, ethnicity, and social positions in Japanese society.

Such social differentiation by the Japanese society can be seen as one example of how definitions of difference are socially constructed. At the same time, they also reflect existing views of how diversity has been understood in Japan.

See also: Ethnic Minority Identity and Educational Outcomes in a Rising China; Race, Critical Race Theory and Whiteness; The Role of the State in School Reform: Responding to Increasing Ethnic Diversity in Britain.

Bibliography

- Barth, F. (1969). *Ethnic Groups and Boundaries. The Social Organization of Cultural Difference*. London: Allen and Unwin.
- DeVos, G. and Wagatsuma, H. (1966). *Japan's Invisible Race: Caste in Culture and Personality*. Berkeley, CA: University of California Press.
- Ebashi, T. (eds.) *Gaikokujin wa jyunin desu (Foreigners are Residents)*. Tokyo: Gakuyo Shobo.
- Fass, P. S. (1989). *Outside In: Minorities and the Transformation of American Education*. New York: Oxford University Press.
- Goodman, R. (2003). The changing perception and status of Japan's Returnee Children (kikokushijo). In Goodman, R., Peach, C., Takenaka, A., and White, P. (eds.) *Global Japan: The Experience of Japan's New Immigrant and Overseas Communities*, pp 177–194. London: RoutledgeCurzon.
- Hickman, M. J. and Walter, B. (2001). Deconstructing whiteness: Irish women in Britain. In Goulbourne, H. (ed.) *Race and Ethnicity: Critical Concepts in Sociology*, vol. II, pp 139–152. London: Routledge.
- Jenkins, R. (1997). *Rethinking Ethnicity: Arguments and Explorations*. New York: Sage.
- Kajita, T. (1998). The challenge of incorporating foreigners in Japan: 'Ethnic Japanese' and 'Sociological Japanese'. In Weiner, M. and Hanami, T. (eds.) *Temporary Workers or Future Citizens? Japanese and US Migration Policies*, pp 120–147. London: Macmillan.
- Kaplan, S. (2001). Can the Ethiopian change his skin? The Beta Israel (Ethiopian Jews) and racial discourse. In Goulbourne, H. (ed.) *Race and Ethnicity: Critical Concepts in Sociology*, vol. II, pp 117–133. London: Routledge.
- Komai, H. and Watodo, I. (1997). *Jichitai no gaikokujin seisaku: Uchinaru kokuzaika e no torikumi* (The Policies of Foreign Residents by Localities). Tokyo: Akashi Shoten.
- Mabuchi, H. (2002). *"Ibunka rikai" no discosu: Bunka honshitsushugi no otoshiana*. Kyoto: Kyoto Daigaku Gakujyutsu Shuppankai.
- Matsuo, R. (1997). Hamamatsushi (Hamamatsu City). In Komai, H. and Watodo, I. (eds.) *Jichitai no gaikokujin seisaku*, pp 163–180. Tokyo: Akashi Shoten.
- Noyama, H. (1997). Otashi/Oizumi machi (Ota and Oizumi City). In Komai, H. and Watodo, I. (eds.) *Jichitai no gaikokujin seisaku*, pp 183–216. Tokyo: Akashi Shoten.
- Oakes, J. (1985). *Keeping Track: How Schools Structure Inequality*. New Haven, CT: Yale University Press.
- Okuda, M. and Tajima, J. (1991). *Ikebukuro no ajiakei gaikokujin: Shakaigakuteki jittai hokoku* (Asian Foreigners in Shinjuku). Tokyo: Mekon.
- Sekiguchi, T. (2003). *Zainichi Nikkei Burajirujin no kodomotachi—ibunkakan ni sodatsu kodomo no aidentiti keisei*. Tokyo: Akashi Shoten.
- Shibata, G. (1997). Kokusai kekkon no shinten ni yoru noson shakai no kokuzaika (The Internationalization of Rural Communities from the Viewpoint of International Marriages). In Komai, H. and Watodo, I. (eds.) *Jichitai no gaikokujin seisaku: Uchinaru kokuzaika eno torikumi*, pp 371–389. Tokyo, Akashi Shoten.
- Shimizu, K. and Shimizu, M. (eds.) (2001). *Nyukama to kyoiku: Gakko bunka to esunishi no katto o megutte*. Tokyo, Akashi Shoten.
- Shiramizu, S. (1996). *Ethnic media: Tabunka shakai Nihon o mezashite* (Ethnic Media). Tokyo: Akashi Shoten.
- Tsunda, T. (2000). Acting Brazilian in Japan: Ethnic resistance among return migrants. *Ethnology* **39**(1), 55–71.
- Tsuneoyoshi, R. (2007). The portrayal of 'foreigners' in Japanese social studies textbooks: Self-images of Mono-Ethnic Pluralism. *Educational Studies in Japan* **2**, 31–44.
- Watabe, M. (1995). *Dekasegi Nikkei Burajirujin*, vols. 1 and 2 (Brazilian Guest Workers of Japanese Descent). Tokyo: Akashi Shoten.

Further Reading

- De Carvalho, D. (2003). *Migrants and Identity in Japan and Brazil*. London: RoutledgeCurzon.
- Douglas, M. and Roberts, G. (eds.) (2003). *Japan and Global Migration: Foreign Workers and the Advent of a Multicultural Society*. Honolulu, OH: University of Hawaii Press.
- Goodman, R. (1993). *Japan's 'International Youth': The Emergence of a New Class of Schoolchildren*. Oxford: Clarendon Press.
- Linger, D. T. (2001). *No One Home: Brazilian Selves Remade in Japan*. Stanford, CA: Stanford University Press.
- Weiner, M. and Hanami, T. (eds.) (1998). *Temporary Workers or Future Citizens? Japanese and US Migration Policies*. London: Macmillan Press.
- White, M. (1988). *The Japanese Overseas: Can They Go Home Again?* New York: The Free Press.

Relevant Websites

- <http://cie.u-gakugei.ac.jp> – Center for Research in International Education, Tokyo Gakugei University (*Kokusai Kyoiku Senta*). An inter-university center which is open to the public situated in the national Tokyo Gakugei University. Added the foreign pupil/student education unit from 2001 and changed its name since 2002 to Center for the Education of Japanese Children Abroad (*Kaigaishijo Kyoiku Senta*). <http://www.eng.jp-clarinet.org> – CLARINET (Children Living Abroad and Returnees Internet). Started in 1997, and is managed by the Ministry of Education, Culture, Sports, Science and Technology (the International Education Unit). This is a database for Japanese children living abroad and returnees. It has expanded to include foreign children education.
- <http://www.joes.or.jp> – Japan Overseas Educational Services (*Kaigaishijo Kyoiku Shinko Zaidan*). This organization was founded by the support of Japanese corporations in 1971, with the approval of the Ministry of Foreign Affairs and the Ministry of Education. It assists the education of Japanese children and their families overseas by providing information, consultation, and correspondence education. Has links to other organizations.
- <http://www.zenkaiken.net> – The All Japan International Educational Research Conference – NPO from 2005. From its web page, the only international educational organization in Japan for foreign children, children returning from other countries, children living overseas, and Japanese language education. It is concerned with education from kindergarten through university. Has links to other similar organizations.
- <http://www.zenkaiken.net> – The All Japan International Educational Research Conference (*Zenkoku Kaigaishijo Kyoiku Kokusai Kyoiku Kenkyu Kyogikai*) (*Zenkaiken*) NPO from 2005. From its web page, the only international educational organization in Japan for foreign children, children returning from other countries, children living overseas, and Japanese language education. It is concerned with education from kindergarten through university. Has links to other similar organizations.

Education in War Countries: A Disempowerment of Society

J L Nkata, Makerere University, Kampala, Uganda

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Education has been a central and leading instrument that has assisted humans, from the very beginning, to prepare and guarantee his survival through innovativeness and development. The creation of knowledge, skills, attitudes, and values, as central psychological and social precepts of any tribe, society, or country, is supported by educational constructs and systems that are underpinned by social goals and purpose, proven and reliable. In this respect, education – whether informal or nonformal – possesses psychological and social liberating potential that underlies life development and sustainability. This is possible when society moulds its activities around a cherished social philosophy that is informed by a salient value premise and goals, destining to improve the intensity of knowledge, skills, values, and social attitudes. It is a social responsibility of a society to sustain education because education upholds social values, perceived morality, ethics, economics, culture, and politics leading to the preservation of the environment in society and outside society.

To achieve this, society needs to be organized and controlled through viable systems and structures that are both open and porous to entertain innovation, creativity, and strategic interventions. This is the ideal environment for systematic formal education systems and cultures; it allows the enrichment of values to enable such society to remain relevant in national and international aspirations. The nature of planning and management should be organic and should strive to achieve effectiveness and efficiency. Unfortunately, many countries are far from realizing such environment because they are constantly in war times, and this threatens their continued survival, global peace, and mutual coexistence. Such environment is not favorable to enable planned education to carry out its central role of empowering people to becoming more purposeful to development initiatives and to instinctively live on. Wars are pollutants to the utopia because they contaminate only meaningful viability for people to carry out activities that would redeem them to be self-supportive regarding food, health, education, security, democracy, and other areas. They become less empowered.

In this article, education is an organized entity or enterprise with definite goals, especially goals of redemption of human efforts, geared toward their survival and relevance. Education is further seen as an organization with structures and systematic programs whose effectiveness and efficiency are redeemed by corporate efforts of

the internal environment and outside factors. War, therefore, is seen as a constituent of both the intra- and extra-environments whose causes and prevalence affect the integrity of society and the environment.

This thought is supported by the systems literature that puts emphasis on systematic and effective management of organizations if goals and values that support them are to be achieved (Nkata, 2004). It underscores the rationale and importance of effective performance of substructures of the larger organization and their subsequent translation as effective constituents of the organization into its success, viability, and development (Kaufman, 1972; Etzioni, 1975; Mullins, 2002; Simon, 1964). In this respect, a country is understood in terms of its complexity and as an organized entity whose activities are assumed to be systematic and focused to social perceived goals.

This article examines the theoretical underpinnings of wars and the impact of wars on educational systems in those countries that have continued strife and armed confrontations, Uganda and Iraq in particular. It highlights the possible causes of war, its effects to education structure and systems, and the consequent disempowerment of societies and countries. Finally, recommendations are proposed via the path of possible solutions as mitigating measures to prevent occurrence or to subside war intensity across the globe.

Theoretical Considerations on the Causes of Wars or Conflicts

The central precept and agency in conflicts, it is agreed across disciplines, involves the role of human as an individual and humans in groups. In this respect, humans are seen as a leader or a participant in programs, issues, assignments, and agenda. Several schools of thought advance the genesis of conflicts as a possible missing link between what ought to be and what is or was. They see conflicts as being sparked by anxieties as a result of fear of extinction, denial, hatred, segregation, social ambiguity, terror, hidden truth, and confusion. Wars and conflicts are a result of a mismatch between people's aspirations and what is given to them or what they receive.

Psychologists

Psychologists such as Sigmund Freud, Erik Erikson, Harry Stack Sullivan, Alfred Adler, and Carl Jung

advanced theories that emphasize the role of individuals as leaders and participants in conflicts and aggressions (Vinacke and Reinhold, 1968). They see survival of the individual and the group as being causal and effective in flagging off social and psychological conflicts. This also implies, in part, that humans have two sides of the self: the bright side and the dark side (Jung, 1956). When everything is fine, the bright side shines but when there is deprivation of psychosocial needs, the dark side of human nature shadows the bright side. In such cases, the Freudian school of thought believes that gratification and survival are derived from aggression. Human being is essentially instinctive, and his/her survival instinct will lead him/her to anything when his/her harmony, freedom, and tranquility are threatened. It is now very clear that war is a reinforcement replacing the positive side to signal society that survival and esteem are being trampled upon. Finally, psychologists believe that pressures on self-esteem tend to reduce the individual's motivation and identity, yielding to aggression (Adler, 1998) for social dominance. Therefore, in order to uphold the individual's identity, emphasis should be placed on issues that reduce social ambiguity in society.

Brown (2001) identifies major classifications of variables that may lead to conflicts and wars in some regions or locations in the world: (1) structural variables (e.g., weak states, intrastate security systems, ethnicity, and geography); (2) political variables (e.g., discriminatory political institutions, exclusionary national ideologies, intergroup politics, and elite politics); (3) economic and social variables (e.g., economic problems, discriminatory economic environment, social dualism, and modernization); and (4) cultural/perceptual variables (e.g., patterns of cultural discrimination and problematic group histories). All these variables tend to underpin ethnic and identity domains of social behavior. Ratzburg (2000), however, observes that wars and conflicts can come by as a result of communication failure, value differences, goal differences, substandard performance, lack of cooperation, differences regarding responsibility, competition over resources, and noncompliance with rules.

All of these theoretical considerations confirm the presence of two parties in a conflict situation, that is, the happy and the unhappy, the satisfied and the unsatisfied, the oppressor and the oppressed, and so on.

Sociologists

Social psychologists emphasize the role of individuals (people) in groups (formal or informal), observing that groups can bring out the best as well as the worst in people. Group contagion can translate into uncertainty, panic, mindlessness, and collective maladjustment with misjudgment (Milgram, 1994). This theory is in line with the psychological assertion of Jung (1956) in relation to

the original and essential nature of people whose image is perceived as bright or dark, good or bad, righteous or evil. The implicit translation is that people can be good or bad, peaceful or war-like, depending on the nature of their environment and how they operate.

Biological Theories

Aggression is essentially in human beings for their survival as living things. It is like animals that have natural instincts and endowments for their protection from danger. It is unfortunate that humans do not apply homogeneous means of aggression. With developing technology, human survival can turn out to be very destructive and violent to others and to the environment. Humans invent weapons that can be used during communal aggression (Lorenz and Wilson, 1966). He/she does this to defend himself when confronted or when his/her psychological integrity and esteem are subjected to danger. However, according to Barbara Ehrenreich, anxieties and fears were naturally hardwired into prehuman species during the times when wild beasts feasted on the *Homo sapiens* to inherit such fears, and these fears were transferred to the humans.

Social Scientists

Some political scientists, such as Karl Marx, believe that the nation is the cause of war. That is, the nation and the people (masses) are in constant struggle for the ever-limited resources. Other social scientists believe that inflexibility and weak leadership can also lead to situations of conflicts and wars (Tuchman, 2004). However, Emmanuel Wallerstein observes that there is a strong diametrical opposition between the center and periphery, cities and villages, leaders and the led, employers and employees – the center exploiting the resources of the periphery, causing imbalances. In such cases, conflicts are rife and consequently, wars. On the other hand, violence results when human communities are under economic, social, and environmental stress or when they engage in ethnic and other social competitions such as religiosity.

Researchers on Stability and Peace

The New International Economic Order (NIEO) that led to globalization is also criticized on the premise that it created and is continuing to create imbalances and epitomizing exploitation, inequality, and aggravated syndromes of dependence. Groups and societies find themselves belonging to the wave of globalization. In attempts to emancipate themselves, conflicts and violence remain but a necessary imperative (Galtung, 1997). This also

implies that transformations on policies, regulations, and goals become incongruent to individual and social values as a result of the bandwagon effect.

Impact of War on Education

Educated people are those whose states of mind and behavior exhibit knowledge and understanding. However, they are not born this way and neither do they become this way through a process of maturation; they achieve it through learning. Often the teaching of others stimulates this learning (Peters, 1980). If the environment of teaching is turbulent, then the importance of learning is ruined or undermined. It will be very difficult to establish schools, train teachers, and enrol students. This is increasingly becoming a global concern due to issues of social and economic interdependence. This article cites Uganda and Iraq as representatives of countries that have been in persistent wars, whose effects on the youth, who are the majority and future generation, threaten local and global interdependency and socioeconomic sustainability.

Uganda

Two decades of armed conflict between the Lord's Resistance Army (LRA) and the Ugandan government in Northern Uganda has caused nearly 2 million civilians to be displaced from their homes. Ordered into so-called protected camps, internally displaced persons (IDPs) face heightened insecurity, appalling living conditions, and lack of means for subsistence. Abuses against these communities are rampant; the LRA has committed war crimes and gross violations of human rights, including the abduction of over 20 000 children; widespread maiming, rape, and murder against the civilian population. These attacks have led to secondary displacement in which up to 40 000 children commute nightly from the camps to sleep in the relative safety of town centers.

While the LRA has been the perpetrator of these crimes, the Ugandan People's Defense Forces (UPDF), the national army, has also committed human rights violations against civilians that include arbitrary detention, extrajudicial killing, torture, and rape. The UPDF, whose mandate is to protect civilians, has not only failed to prevent attacks and abductions by the LRA but has also perpetrated grave abuses against civilians in a climate of impunity (Between Two Fires, *Torture and Displacement in Northern Uganda*, Human Rights Focus-HURIFO). The impact of armed conflicts on children and youth development, in terms of their appropriate health, nutrition, security, psychological and sociological esteem, and integrity, has so far remained a dream beyond the periphery for the over 30 000 young children of this area. The International Rescue Committee (IRC Report, 2006)

notes that more than 30 000 children were kidnapped to serve as soldiers and slaves by the LRA. Boys in captivity were forced to loot, torture, and kill neighbors. Abducted girls were routinely raped and became sex slaves or wives of rebel commanders. Since 2006, efforts to reconcile and reach a peaceful resolution between the LRA and the Ugandan government have been slow to resettle over 1.8 million displaced people, who have been in protected villages. Unfortunately, much damage was done on the psychosocial dimension of their lives besides other facets of economics, culture, morality, and ethics.

There is insufficiency in terms of teachers who can handle the normal school curriculum, as well as assist with rehabilitation lessons and programs. This is why the Africa Education Initiative (AEI) has so far trained over 9000 teachers in conflict-affected areas in this region. This claim is supported by what the National Assistant Minister of Education and Sports found on ground when he recently visited some schools in the area (The Monitor Newspaper, 11th February 2008) to get firsthand information. At Orum Primary School, only two teachers were at the school during class hours. Over 80% of the staff (teachers) were absent or missing. Hon. Peter Lokeris who arrived at 11:00 a.m. also saw loitering students, yet this was already a month since schools opened for the first term in January. He was shocked to learn that out of the 28 pupils who had registered for Primary Leaving Examinations (public examination), no one had passed! He also learned that primary six students expected for term one, 2008, were 58 in number but only 21 had reported for the new term and year. This appears to be the reality in the war-torn northern region of Uganda where virtually every physical and social infrastructure and human resources including schools, roads, and teachers were either destroyed, displaced, or mismanaged, therefore calling for emergency assistance and resuscitation. As a result of the war, the region has failed to meet national set targets of education performance. For example, if one applied the Education For All (EFA) goals and targets of the Jomtien Declaration of 1990 (Education For All by the year 2000) to judge the performance of education, via the 18 indicators that underpin EFA targets, they would be appalled. The National Ministry of Education and Sports, on comparing the regional gross enrolment ratios (GERs) of 1991 and 1995, found out that whereas the GER of eastern, central and western regions of the country were progressively rising, that of the northern region was appallingly, reduced from 71.8% to 56.3% (see **Table 1** and **Figure 1**).

Again, the Ministry of Education admitted that standards fell in this region where turbulence has been going on for the last two decades. This was reached after a survey on performance of pupils in science and social studies, on a comparative basis by region, conducted in 1997 to establish reference points for Universal Primary Education (UPE), which was launched in this same year.

Northern region performed poorest among the four regions of the country. One of the outstanding reasons adduced to this phenomenon is armed insurgency (see Table 2).

Insecurity in this part of Uganda has lent to massive diversion of resources (financial and otherwise) to contain the situation and this is a great opportunity cost to the social, political, and economic performance of the country.

Table 1 Regional gross enrolment rates

Region	1991	1995
Northern	71.8	56.3
Eastern	78.5	86.4
Central	76.7	88.9
Western	78.8	86.0
Total	76.5	79.4

From Educational statistical abstract (ESA).

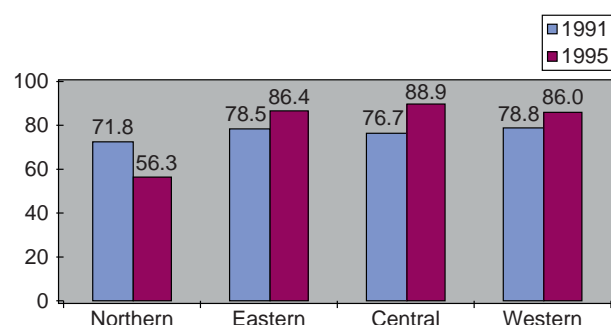


Figure 1 Regional gross enrolment rates. Looking at the regional gross enrolment ratios for 1991 and 1995, we note from Table 1 and Figure 1 that GER for the Northern region, which was as high as 71.8% in 1991, had significantly reduced to as low as 56.3% by 1995. One of the possible causes of this fall is the insurgency that has been in the region for many years now. This insurgency has forced many children out of school. For all the other regions, the gross enrolment ratio has increased over the same period (ESA).

This has been reflected in national budgets where the vote for defense is still leading other sectoral expenditure. In addition, international donor agencies, community-based initiatives, nongovernmental organizations and nonprofit-making bodies have continually come in to provide various interventions.

Through United States Agency for International Development (USAID), support is extended to the following areas to intervene into the wanting social-service sector:

- Improving policing in the north of the country through strategic functional training programs to civil police to reduce crime incidence and occurrence, such as illicit trafficking in conventional arms. The intention is to increase security and reduce intensity of criminal lawlessness, creating a safer environment for social and economic services/activities such as education.
- Supporting the reintegration of ex-combatants – majority of who are children and youth, through deliberate programs to increase their awareness and potential for greater coexistence, morality, and ethics.
- Distributing emergency assistance in collaboration with the World Food Programme and other private voluntary organizations. Over 64 000 metric tons of food aid has been provided to more than 270 000 returnees in the region; in addition, USD 19 000 000 for health and agro interventions has so far been provided.

Besides the USAID assistance to the northern districts of Uganda, many projects have been started with hope that the environment will be revamped, to enable the region stand again on its feet to support social structures such as those of formal education. These include (among others):

- Uganda Girls' Education Initiative Project, established by Acholi Education Initiative (AEI) and being funded by Pindof Project of Canada.
- World Council of Churches (WCC), through the local AEI supports child mothers with scholarships, uniforms,

Table 2 Overall performance of pupils in science and social studies by region

Class	Grade	North	West	Central	East	All
Science						
P6	Distinction	7 (2%)	25 (8%)	40 (15%)	22 (8%)	94 (8%)
	Pass	66 (22%)	109 (36%)	57 (21%)	82 (30%)	314 (27%)
	Fail	230 (76%)	171 (56%)	172 (64%)	168 (62%)	741 (65%)
Social studies						
P6	Distinction	46 (15%)	136 (45%)	99 (38%)	88 (32%)	369 (32%)
	Pass	82 (27%)	86 (29%)	62 (23%)	66 (24%)	296 (26%)
	Fail	173 (57%)	80 (27%)	109 (40%)	120 (44%)	482 (42%)

From: National Assessment of Progress in Education (NAPE), 1998.

books, and pens for their secondary period of 4 years. These girl mothers are given access to psychological counseling services and other support to meet healthcare and other basic needs in form of food, shelter, and clothing. Funds for their rehabilitation are also given to them through assortments of out-reach programs.

- Northern Uganda Child Legal Defense Project operates in the districts of Kitgum and Gulu with the sponsorship of the Association of Women Lawyers and Human Rights Focus, under the Legal Project of the Uganda Law Society.

The surest way of restoring order is to resuscitate the structure of education in the region, to enable its pivotal role of empowerment and awareness to take place among people – majority being youth and children. Education assists society to hand over the desired skills, knowledge, values, and attitudes to young generations, for its development and sustainability. It is therefore a tool that can be used to prevent and stop such insurgency.

Iraq

From the 1970s to 1980s, Iraq was a leading and fast-growing country in the Middle East, with per capita income of up to USD 3600. The successive wars, among other factors, have now stifled growth and development, debilitating basic social services such as education. By 2003, gross domestic product (GDP) per capita lowered to a mere USD 480–630. The most recent war that ousted Saddam Hussein just aggravated the situation and as a result the social sector, including education was affected. United Nations Educational, Scientific and Cultural Organization (UNESCO) reports indicate that before the First Gulf War in 1991 Iraq's education system and performance was one of the best in the Gulf region, with a GER of 100%, and literacy rates were significantly high. As a result of wars the education system and structure has suffered severely, sanctions and instability have complicated the situation, and yet the key asset of the country lies in its educated and experienced personnel. Unlike Uganda and many other countries emerging from war conflict or crisis, Iraq is well endowed with natural resources. The immediate challenge will be to restore its self-confidence and recover a sense of normalcy. The resumption of educational services is a prerequisite for all initiatives aimed at promoting national reconciliation, policy dialog, and the democratization process.

Education is a necessary social tool that can help society maintain its order. To do this its structure and that of the environment in which it operates must be mutually systematic and interactive. This supports their existence as vibrant entities. In the case where the environment is

hostile and nonsupportive, social development is not guaranteed. Countries should endeavor to prevent situations and circumstances that can lead them to war because war is destructive. In the face of its prevalence, in many countries across the globe something must be done.

Way Forward

Since war affects both the internal and external environments of warring countries, it should be observed that corporate global efforts should be concerted as much as possible to stabilize the environment of educational operations, performance, and development. This is possible by:

- Putting in place revised controls as much as possible to prevent wars. Such controls can be in form of regional cooperations with sufficient and genuine awareness programs – these stand out to empower the leaders and masses as privileged choosers of their destiny.
- Warring groups should be compelled by the global fraternity to become more accountable to their people to the extent of enabling simple but progressive core subjects, relevant to war situation such as empowerment lessons. The people, besides knowing why they are in war, should by right, be involved in resolutions that can lead to its end. Failure by war leaders to enable this right to be realized should be punishable by competent courts after the war. It should be taken as an abuse of fundamental right to humanity.
- With advanced technology, hi-tech researchers should evolve gadgets in form of microcomputers with programs that can facilitate learning during war times. A small computer like a matchbox is more convenient than a walkie-talkie to a soldier. People should not be kept in ignorance and information blackout.

Finally, all formal classes, during the good days when there is no war, should involve Development and Liberal Education (DLE) programs, across all levels of education to impart ideals of development, morality, ethics, and desired discipline to the learners. Currently in many countries in the world, emphasis is put on academic excellence. Development education is ignored and yet it would assist to produce reliable, upright, and responsible focused personalities who would contribute to the reduction of conflicts and wars. Formal education should also endeavor to provide quality basic education and training; reduce adult illiteracy, thereby increasing the acquisition by families and individuals of developmental knowledge, skills, values, and attitudes required for better living; and improve learning achievements. It is hoped that with the ever-increasing quality of education among countries, wars will be prevented due to its empowerment potential.

Bibliography

- Adler, A. (1998). *Understanding Human Nature*.
- Brown, M. (2001). *Ethnic and Internal Conflicts in Turbulent Peace: The Challenges of Managing International Conflict*, Crocker, C. A., Hampson, F. O., and Aall, P. R. (eds.). Washington, DC: USIP Press.
- Etzioni, A. (1975). *A Comparative Analysis of Complex Organizations: On Power, Involvement and Their Correlates*. New York: Free Press.
- Galtung, J. (1997). *Handsburg Peace and Conflict Studies*. Taylor & Fires group.
- IRC Report (2006).
- Jung, C. G. (1956). *The Undiscovered Self in Civilization in Transition*. Princeton University Press.
- Kaufman, R. A. (1972). *Educational Systems Planning*. Englewood Cliffs, NJ: Prentice Hall.
- Lorenz, C. and Wilson, E. O. (1966). On Aggression. *New York Times*.
- Milgram, S. (1994). Obedience to Authority. *New York Times*.
- Monitor Newspaper, Vol. 13, 11 February 2008.
- Mullins, L. J. (2002). *Management and Organizational Behavior*. Englewood Cliffs, NJ: Prentice Hall.
- Nkata, J. L. (2004). *Emerging Issues in Education Management*. Kampala, Uganda.
- Peters, R. S. (1980). *The Philosophy of Education*. Oxford: Oxford University Press.
- Ratzburg, W. H. (2000). Causes of Conflict. <http://www.geocities.com/Athens/Forum/1650/htmlcausesofconflict.html> (accessed July 2009).
- Simon, H. A. (1964). Concept of organizational goals. *Administrative Science Quarterly* 9, 1–22.

- Tuchman, B. (2004). *The guns of August*, Presidio Pr.
- Vinacke, E. and Reinhold, V. N. (1968). *Foundations of Psychology*. New York.

Further Reading

- Erik, E. H. (1963). *Childhood and Society*, 2nd edn. New York: W W Norton.
- Etzioni, A. (1964). *Modern Organizations*. Englewood Cliffs, NJ: Prentice Hall.
- Pourton, J. W. (1990). *Conflict Resolution and Prevention*. New York.
- UNESCO (2003). Situation Analysis of Education in Iraq. <http://unesdoc.unesco.org/images/0013/001308/130838e.pdf> (accessed August 2009).

Relevant Websites

- <http://www.un.org/events/tenstories> – Ten Stories the World Should Know More About, 2007.
- www.state.gov/prs/2007
- www.uls.org.ug – University Liggett School.
- www.worldbank.org
- <http://en.wikipedia.org> – Education in Iraq, Wikipedia.

Educational Attainments of U.S. Black Males and Females: 1971 to 2003

F B Bonner, M-C Jipguep-Akhtar, and R J Harrison, Howard University, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The participation and progress of females in higher education in the United States has been followed closely for more than three decades. Debates over closing the gender gap in postsecondary education continue. Several arguments have been put forward advancing the proposition that women are succeeding at the expense of males. In the US, concerns have been greatest for the growing gap in college enrolment and completion between black men and women. While black women comprised 55% of all blacks enrolled in 4-year institutions in 1971, they constituted 60% of the black students enrolled in 4-year institutions in 2004 (Allen *et al.*, 2005). The gap is even larger in all degree-granting undergraduate institutions, where 64.3% of all blacks enrolled in 2004 were women.

The female to male differential in enrolment and graduation rates has been reported to be as high as 6:1 in some historically black colleges and universities (HBCUs; Kaba, 2005), creating a gap that some find alarming. The black gender differential has often been couched in the wider context of several indicators on which the relative status of black males is poor or deteriorating, including employment, health, and incarceration measures (Sum *et al.*, 2007). A common inference is that special programs or policies are needed to target young black males, and to address specific barriers or dysfunctions to which they are presumed to be subject (Sum *et al.*, 2007).

Our research tracks the enrolment of males and females in higher education over a 30-year period in an attempt to shed light on the problem. This article notes that the gender gap in college enrolment and completion among African Americans is consistent with – though much larger than – similar gaps among all major race and ethnic groups in the United States, and is thus – in important senses – not solely just one more socioeconomic or cultural problem that is specific to African Americans or African-American males. We demonstrate that for African-Americans and other major race and ethnic groups, the gender gap – as measured by undergraduate fall enrolment – is larger than the gap measured by the percentages of all 18- to 24-year-olds and of 18- to 24-year-old high school graduates who entered or completed college. This would strongly suggest that differential retention is important to explain the gap, and that,

policymakers and educational institutions may need to focus more on identifying and addressing issues that might affect retention – including academic and financial resources, social support issues, and competing employment needs or opportunities.

A Glance into Black Male–Female Participation in Higher Education

Since the 1970s, both black females and males in the United States have made substantial gains in postsecondary education, but the gains of females have clearly been greater. The belief that females have benefited more from educational gains in recent years and are outperforming males at all levels is widespread (Association of Teachers and Lecturers, 2004). Historically, more black women have attended college than black males, and these women tend to achieve higher grade point averages (GPAs) and receive more academic honors than black men (Jackson, 1986). In 1976, nearly 54% of black undergraduate enrollees were female. Since 1976, the number of both males and females in undergraduate programs has increased. By 1980, the percentage of all females enrolled as undergraduates had surpassed that of all males. The largest difference between male and female enrolment was for black students. In 2002, the percentage of women with high school degrees surpassed the percentage for men, and in 2005 slightly more females (84.6%) than males (83.6%) reported they had earned a high school diploma (American Community Survey (ACS), 2005). Black females continued to enrol in degree-granting institutions in larger numbers than black males, and in 2004, females accounted for 64% of all black enrollees (Kewal Ramani *et al.*, 2007). This trend continues (Crissey *et al.*, 2007).

A Note on Gender, Race, and Educational Attainment in the United States

In global and national contexts, education has become increasingly important as a means of improving societal well-being through its role in increasing understanding and skills development as well as improving social cohesion and reducing social inequalities (Power, 2007). In the last 30 years, there has been increasing awareness of the

relationship between gender, race, and ethnicity and educational inequalities (Power, 2007). Existing studies have posited that involuntary minority groups with a history of oppression in the United States, specifically African Americans, are more likely to resist educational goals in opposition to the values of the dominant society. Such arguments have been used to explain the lack of congruence between black students' high educational aspirations and poor school performance (Ogbu, 1978; Ainsworth-Darnell and Downey, 1998; Hirschman and Lee, 2005). Early work on the educational attainment of US blacks also focused on females compared to white males; but, another debate emerged in the black feminist literature. This debate focused on the disaggregation of data to provide a clearer picture of the participation and progress of women of color compared to white males and white females.

Until recently, there has been little examination of the participation and progress of black males compared to females in these data. This changed with the publication of a few obscure articles, in 2003, on the disproportionate number of females to males enrolled at HBCUs. Research showed that, in many instances, the ratio of US female-to-male participation in higher education was as high as 6:1. However, most early studies ignored the fact that females outnumbered males in post-secondary education in nearly every country of the world. The alarm was sounded and the debate raged on – fueled, at that time, by the struggle over the reaffirmation of Title IX. Recent research on gender equity in education from the American Association of University Women (AAUW) posits that girls' successes have not come at the expense of boys' (Corbett *et al.*, 2008). The AAUW's (2008) comprehensive assessment of girls' and boys' educational progress during the past 35 years concludes that a boys crisis in US schools may be a myth and that both sexes have stayed the same or improved on standardized tests in the past decades (Strauss, 2008). This conclusion challenges the notions that the subgroup with the most persistent lag in education is black males, and stands in stark opposition to education statistics that consistently reveal that black males cluster at the bottom of the distribution of virtually every indicator of school failure such as enrolment and dropping out (Jordan and Cooper, 2002).

Educational Attainment and Black Males

The relative educational status of black males in the US has been characterized as problematic, with disparities evident across many educational metrics, including graduation rates and skills (Levin *et al.*, 2006). While the extant literature on gender differences in educational attainment among African-Americans suggests that males have deplorable high school graduation and dropout rates (Weatherspoon, 2006), statistics show that the percentage of black males

completing high school has increased at a rate comparable to the percentage for black women, and both have dramatically closed the high school completion gap with whites since 1965. At that time, only about one-half of all black women (50.3%) and black men (50.2%) in the age range of 25 and 29 years had completed high school, compared to nearly three-quarters of white women (72.8% and 72.7%). By 2007, high school completion was almost universal among Asians (97.2%) and whites (93.5%), and African Americans (87.4%) were not far behind. The female-to-male high school completion ratio among African Americans, in 2007 (1.01), was comparable to those for whites (1.02) and Asians (1.03), and none statistically differed from unity. The female-to-male high school completion ratio for African Americans has fluctuated over the years, reaching 1.06 as recently as 2006, as well as back in 1967 and 1970, and 1.12 in 1971. However, it has also dropped as low as 0.87 in 1967, 0.92 in 1968, and 0.93 in 1974 and 1977, and the average ratio in the past decade (1998–2007), the past 20 years (1988–2007), and the past 30 years (1978–2007) does not statistically differ from parity.

The data thus suggest that black males have made similar advances in their rates of completing high school as black females, and that both have dramatically closed the gaps with whites in high school completion. *The Journal of Blacks in Higher Education* (2006) notes that 758 400 black males were enrolled in institutions of higher education in 2004. They represented 10% of all male enrollees and 4.4% of all enrollees in higher education. In 1990, 484 700 black males were enrolled in higher education, and they comprised only 6.4% of all male enrollees and 3.5% of all enrollees in institutions higher education. In 1976, of the 1 033 000 black students enrolled in higher education institutions, 563 100 were women. The enrolment figure of black women in higher education institutions increased to 917 800 and, in 1999, of the 1 640 000 black students enrolled in institutions of higher education, 1 037 700 were women (National Center for Education Statistics (NCES), 2002; Kaba, 2005). Thus, not only black females, but also black males, have posted substantial gains in college enrolment in recent decades.

Enrolment in Degree-Granting Institutions

Enrolment statistics seem to tell a sharply different story, however. Black males have fallen from 45.7% of black enrolment in 1976 to only 35.7% in 2004 (see **Figure 1(a)**). In 2004, there were 1.8 black females in degree-granting institutions for every black male. This ratio was far higher than that for whites (1.27) and Asians (1.16), but did not statistically differ from the ratios for Hispanics (1.41) and Native Americans (1.56).

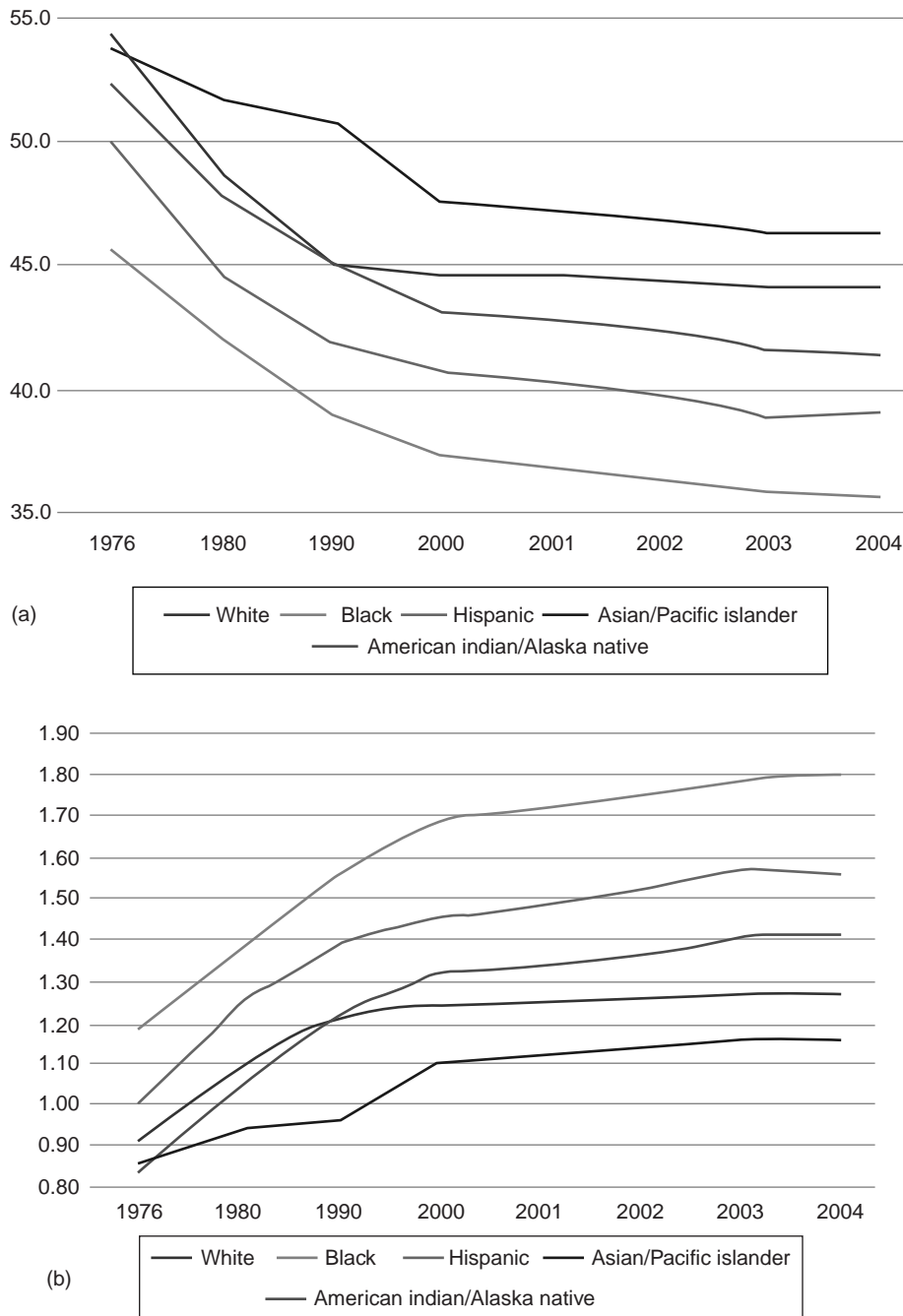


Figure 1 (a) Percent male of fall undergraduate enrolment in degree-granting institutions, by race and ethnicity, 1976–2004. (b) Ratio of female to male enrolment in degree-granting institutions, by race and ethnicity, selected years, 1976–2004. From US Department of Education, National Center for Education Statistics, Digest of Education Statistics, 2005 (NCES 2006-030), table 205 'Fall enrollment in colleges and universities'.

The female-to-male enrolment ratio grew by over 50% (51.4%) among blacks (see **Figure 1(b)**). However, it is worth noting that the ratio grew as rapidly among Hispanics (68.1%) and Native Americans (55.4%), and that the growth for blacks did not statistically differ from the growth among whites (39.2%) or Asians (35.5%). The trend among blacks thus seems to represent developments

common to all major race and ethnic groups, and is no stronger than those in the other minority populations – Hispanics and Native Americans – with lower educational attainments than whites and Asians.

One might, nevertheless, ask: does the much higher female-to-male college enrolment ratio among blacks (1.8 females per male) not confirm that the problem is at least

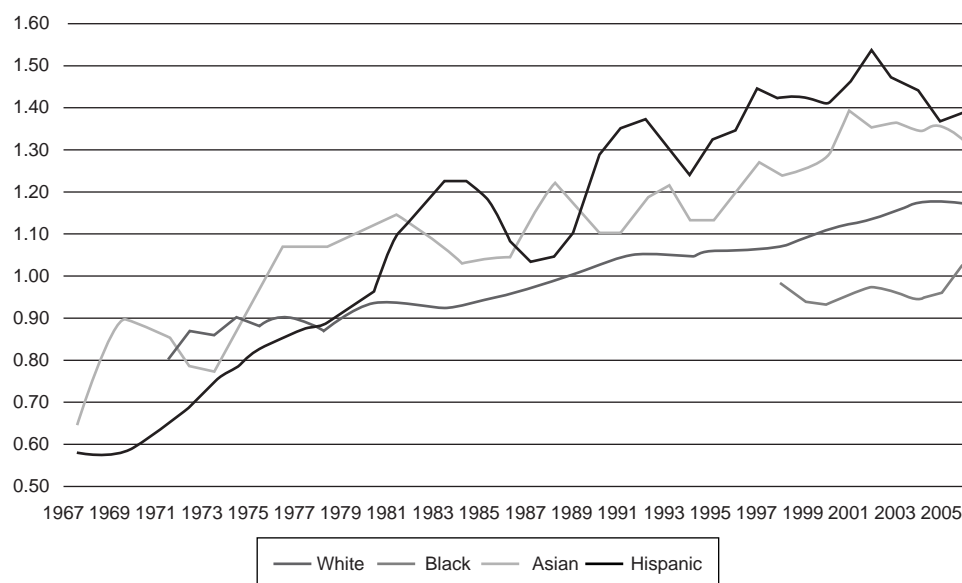


Figure 2 Ratio of female-to-male 18–24-year-olds enrolled in degree-granting institutions by race and ethnicity, 1967–2006 (3-year moving average). From US Census Bureau, current population survey, 1967–2006.

more severe in the black population, even though the gap is substantial in the major racial and ethnic populations, and also grew within them between 1976 and 2004? In the remainder of this article, we present evidence that suggests that the higher female-to-male ratios for blacks reflects at least two additional factors that have received too little attention and merit further analysis. First, a larger part of the gender differential among blacks seems attributable to enrolments of black women older than 25 years in institutions of higher learning: enrolment ratios are substantially lower for black 18–24-year olds. Second, the ratios seem even lower when one examines percentages of 14–24 and of 25–29-year-olds who attended or completed college. This suggests that some of the gender gaps in the stock of black females and males enrolled at any point in time may reflect differences in the flows of black women and men into and out of degree-granting institutions. Both more rapid completion of degrees and higher attrition could contribute to black men entering and leaving degree-granting institutions more rapidly than black women, contributing to a higher stock of black women at any given point in time.

18–24-Year-Olds Attending or Having Attended a Degree-Granting Institution

The statistics cited above indicate that in 2004, for every black male enrolled in a degree-granting institution there were 1.8 black females. However, among 18–24-year olds, only 1.34 black women were enrolled in degree-granting institutions for every black male, and 1.33 in 2006 (**Figure 2**). This suggests that more than half of the gap between black males and females reflects much higher enrolments of

black women than black men who are older than 24 years, either because they are in graduate or professional schools, or are returning to school – in all likelihood after some break in their education – or because they taking more years to complete their degrees – perhaps as part-time students. In contrast, the female-to-male enrolment ratio for whites of all ages (1.27) did not statistically differ from that for 18–24-year-olds (1.17) in 2004 (**Figure 2**).

The differences in enrolment for all ages compared to 18–24-year-olds in 2004 were also not statistically significant for Hispanics (1.41 and 1.46) or for Asians (1.16 vs. 0.93). Much of the unique dominance in the enrolment of black women over black men in degree-granting institutions thus seems to reflect a uniquely large enrolment of black women older than 24 years. The ratio of female-to-male enrolment among 18–24-year-olds for blacks (1.34), in fact, was not statistically higher than that for whites (1.17) in 2004 (**Figure 2**). The problem – in the light of these statistics – seems to lie less in the enrolment rates or behavior of young black men than in what one must infer is much higher enrolments among black women over 25 years still completing or returning to their education.

The statistics cited here on black female and male enrolment are drawn from fall enrolment in degree-granting institutions – a measure of the stock at each point in time. The percentage of 18–24-year-olds and of 14–24-year-old high school graduates who – in a given year – were enrolled in or had completed some college provides an indicator that reflects flows as well as stocks. The black female-to-male ratio for enrolment in or completion of college was only 1.20 in 2004, and remained below 1.25 until 2006 when it surged to 1.29 (**Figure 3**). This was statistically identical to the corresponding ratio for whites,

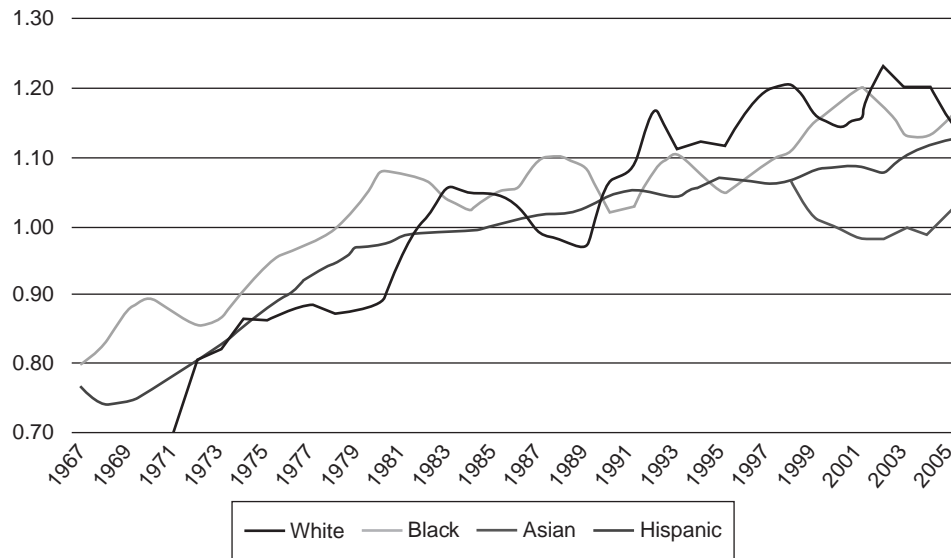


Figure 3 Female-to-male ratio of 14–24-year-old high school graduates who attended or completed college, by race and ethnicity, 1967–2006 (3-year moving average). From US Census Bureau, current population survey, 1967–2006.

which was 1.12 in 2004, and 1.09 in 2001 – when the ratio for blacks was 1.20 (**Figure 3**). The black female-to-male ratios for enrolment in or completion of some college are comparable for 14–24-year-olds and for 18–24-year-olds – remaining below 1.20, except in 2001 and in 2006, when the ratio climbed to 1.26. The female-to-male enrolment and completion ratios for 18–24- and for 14–24-year-olds both indicate that the behavior of young black male and female high school graduates do not differ greatly from those of their white counterparts, or, in most years, from Hispanics either.

College Completion

The fact that the percentages of 18–24-year-old black men who were enrolled in or had completed some college seems closer to the corresponding percentages for black women in most years than their enrolment rates, at any given point in time, suggests that black men may spend fewer years on campus than black women. This might arise from lower retention rates (higher college-dropout rates) for black males, or because black males take less time, on average, to complete their degrees. If college dropout rates were notably higher for black men than for black women, one would expect that the ratio of black females to black males with bachelors' degrees would be considerably higher than the corresponding ratio for those who enrolled in or completed at least some college.

There is no strong evidence of this in the percentages of 25–29-year-old black women and men with BAs. The 3-year moving average of the ratio was 1.29 in 2004 (and in 1997), and reached 1.46 in 2005 before falling to only 1.11 in 2007. With the exception of a few years adjacent to these

peaks, the ratio of black female to male BAs in this cohort has seldom exceeded 1.20 since the mid-1970s, and, in fact, briefly fell below 1.00 between 1988 and 1991. These ratios are well within the same range as those for the black female-to-male ratios for those enrolled in or having completed some college in the 14–24-year-old and the 18–24-year-old cohorts. The ratios, since the 1990s, for blacks may look high in comparison with the period from 1974 to 1984, during which the ratio remained under 1.10. However, the female-to-male BA ratios for black 25–29-year-olds, although higher, have also remained well within the range of those for whites, which have steadily risen from unity in 1990 (when the ratio for blacks was 0.94), to 1.23 in 2007 – a ratio seemingly higher than the 1.11 female-to-male BA ratio for blacks in 2007. Indeed, the black and white female-to-male college-completion ratios did not statistically differ from one another between 1989 and 1992, and between 2000 and 2003.

College-dropout rates may well have been higher for black men than for black women during some of these years, and this must certainly be an item of high priority for research on gender differentials in college enrolment and completion in recent decades. However, the comparability of the ratios of black 25–29-year-old women and men holding BAs to ratios of those who enrolled in or completed some college suggests instead that black women may be taking longer to complete their degrees. Any slower flow of black women through the institutions while completing their degrees would contribute to their higher enrolments, and might also help explain why the ratio of black female-to-male enrolment in college is higher than the corresponding ratio for young black adults (25–29-year olds) who have completed college. This too should be high on the research agenda.

Discussion and Conclusion

The fact that nearly two-thirds of all black students enrolled in degree-granting institutions in 2004 were women, and that there are approximately 1.8 black women for every black man on campuses in the United States has frequently been presented and discussed as yet another indicator where black men seem to lag far behind men and women in most major racial and ethnic groups in their educational achievements and attainments. We have presented evidence here that the largest factor in the differential enrolment of black men and women might lie in uniquely high enrolment by black women over 25 years old: the black female-to-male ratio among 18–24-year-olds is 1.40:1 – only half the ratio for blacks of all ages. The ratio falls even further, to – at most – 1.30:1 and usually 1.20:1, when one compares black males and females in the 18–24-year-old and 14–24-year-old cohorts who have enrolled in or completed some college. The ratios for enrolment and completion could be lower than those for enrolment alone if black men were dropping out of college at higher rates than black women, or if they were completing their degrees in less time. The ratios of black women to black men in the 25–29-year-old cohort who have completed a bachelor's degree also fall within the same range as the ratios for enrolment and completion. This suggests that higher dropout rates probably contribute less to the differential in most years than would faster completion of the degree by black men. Once these factors are considered, the gender differentials in the college enrolment and completion rates for blacks – and Hispanics – (usually no more than 1.30:1 and usually less than 1.20:1) do not differ dramatically from the differentials for whites (no more than 1.25:1 and usually above 1.10:1).

The gender differentials in enrolments might be telling us much more about women, and especially black women, than about black men and their lower educational attainments. In 2006, the mean earnings of a female black (\$22 643), white (\$23 805), Asian (\$25 696), or Hispanic (\$20 608) high school graduate falls short of or barely reaches the mean earnings range of a white, male high school drop-out (\$26 100). Completing some college or an Associate Arts degree barely raises the mean earnings of black (\$28 706), white (\$28 322), Asian (\$29 415), and Hispanic (\$26 260) women above the mean earnings of the white male dropout (\$26 100). Black (\$44 326), white (\$43 473), Asian (\$44 932), and Hispanic women (\$38 825) can roughly double what they would earn with a high school degree by completing college (even though this only gives them mean earnings comparable to those of white males with some college or an associate arts degree – \$43 589).

One might hypothesize that as social and cultural changes have greatly increased the likelihood that women – even those raising children – will work for most of their

adult years, they have also seen completing at least some college or earning an associate's or bachelor's degree as essential to making work worthwhile. In contrast, with only a high school degree, even black (\$30 122), Asian (\$32 528), and Hispanic (\$32 148) males can earn more than minority women who have completed some college or an associate arts degree (\$26 000–\$29 500). Many of these men might think it better – or might feel pressed by family responsibilities – to enter the labor market and begin earning wages that a woman could not expect to earn without completing some college or a degree. One might also hypothesize that the need to increase earnings potential may have led black women – who have had historically higher rates of single parenthood and contributed higher shares to family income when married – to increase their efforts (compared to males) to attend and complete some college before women in other racial and ethnic groups did so. The necessities of attending or completing college while working – and, for others, of returning to school to improve earnings – might be examined as factors that might help explain the much higher college enrolment rates of black women over 25 years of age.

Bibliography

- Ainsworth-Darnell, J. W. and Downey, D. B. (1998). Assessing the oppositional culture explanation for race/ethnic differences in school performance. *American Sociological Review* **63**(4), 536–553.
- Allen, W., Jayakumar, U. M., Griffin, K. A., Korn, W. S., and Hurtado, S. (2005). *Black Undergraduates from Bakke to Grutter: Freshman Status, Trends and Prospects 1971–2004*. Los Angeles: Higher Education Research Institute, UCLA.
- Association of Teachers and Lecturers (2004). (H. Claire, Editor). http://www.askatl.org.uk/atl_en/images/Genderpdf_tcm2-10600.pdf (accessed August 2009).
- Corbett, C., Catherine, H., and St. Rose, A. (2008). *AAUW Educational Foundation*. <http://www.aauw.org/research/upload/whereGirlsAre.pdf> (accessed August 2009).
- Crissey, S. R., Scanniello, N., and Shin, H. B. (2007). *U.S. Bureau of Census*. Housing and Household Economic Statistics Division: <http://www.census.gov/population/www/socdemo/educ-attn.html> (accessed August 2009).
- Hirschman, C. and Lee, J. C. (2005). Race and ethnic inequality in educational attainment in the United States. In Rutter, M. and Tienda, M. (eds.) *Ethnicity and Causal Mechanisms*, pp 107–138. Cambridge: Cambridge University Press.
- Jackson, E. (1986). *Black Education in Contemporary America-A Crisis in Ambiguity*. Bristol, IN: Wyndham Press.
- Jordan, W. J. and Cooper, R. (2002). *CRESPAR*. <http://www.csos.jhu.edu/crespar/techReports/Report60.pdf> (accessed August 2009).
- Kaba, A. (2005). *Education Policy Analysis Archives (EPAA)*. Education Policy Analysis Archives (EPAA): <http://epaa.asu.edu/epaa/v13n25/v13n25.pdf> (accessed August 2009).
- Kewal Ramani, A., Gilbertson, L., Fox, M. A., and Provasnik, S. (2007). *National Center for Education Statistics*. http://nces.ed.gov/pubs2007/minoritytrends/ind_6_23.asp (accessed August 2009).
- Laird, J., DeBell, M., and Chapman, C. (2006). http://www.hslda.org/docs/nche/Issues/D/dropout_rates_2004.pdf (accessed August 2009).
- Levin, H. M., Belfield, C., Muennig, P., and Rouse, C. (2006). http://www.cbcse.org/media/download_gallery/Public%20Returns%20Feb07.pdf (accessed August 2009).

- National Science Board (2002). <http://www.nsf.gov/statistics/seind02/c1/c1s9.htm#c1s9l1> (accessed August 2009).
- Ogbu, J. U. (1978). *Minority Education and Caste*. New York: Academic Press.
- Power, S. (2007). *European Commission, Directorate General for Research, Information and Communication Unit*. http://ec.europa.eu/research/social-sciences/pdf/policy_reviews/eur22090_en.pdf (accessed August 2009).
- Strauss, V. (2008). No Crisis For Boys In Schools, Study Says-Academic Success Linked to Income. *Washington Post*.
- Sum, A., Khatiwada, I., McLaughlin, J., and Tobar, P. (2007). The educational attainment of the nation's young black men and their recent labor market experiences: What can be done to improve their future labor market and educational prospects? Center for Labor Market Studies, Northeastern University. http://www.clms.neu.edu/publication/documents/Ed_Attainment_of_Black_Males.pdf (accessed August 2009).
- The Journal of Blacks in Higher Education (2006). http://www.jbhe.com/news_views/52_black-highereducation.html (accessed August 2009).
- U.S. Bureau of Census (2006). *Current Population Survey (CPS). A joint effort between the Bureau of Labor Statistics and the Census Bureau*. <http://www.census.gov/cps/> (accessed August 2009).
- U.S. Bureau of Census (2006). *School Enrollment – Social and Economic Characteristics of Students: October 2005-Detailed Tables*. <http://www.census.gov/population/www/socdemo/school/cps2005.html> (accessed August 2009).
- U.S. Census Bureau (2007). *Current Population Survey (CPS)*. <http://www.census.gov/cps/> (accessed August 2009).
- Weatherspoon, F. D. (2006). <http://users.law.capital.edu/fweatherspoon/Publications/racialjusticeandequity.pdf> (accessed August 2009).

Educational Preparation: Fostering the Self-Efficacy and Resilience of Urban Adolescent Youth

M B Spencer and B Tinsley, University of Pennsylvania, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Guaranteeing that a society's youth are prepared to take on adult tasks, through educational preparation and appropriate socialization opportunities, represents an important duty deserving substantial attention and support. The preparatory content may be framed and implemented with degrees of variation given the unique cultural traditions. However, it is universally understood that adolescence is a pivotal period of academic, social, and emotional adjustment. These anticipated changes arise in response to normative sets of challenges accompanying youths' preparations for transitioning into adulthood.

As detailed elsewhere, all humans are vulnerable (i.e., embody both risk and protective factors) (see Spencer, 2006, 2008). As a social minority, particular youths' normative experiences and, thus, vulnerability specific to the adolescent period of development, are intensified. Furthermore, belonging to a specific group that is perceived by the majority (or those in power) as inferior, in some manner, matters. For example, living in a family with few social and economic resources worsens the normative development-linked risks and their expression as myriad challenges in need of coping responses during this period. Additionally, although not often discussed, the potential impact may be moderated by adequate levels of protective factors manifested as supports. That is, resiliency is possible (i.e., the achievement of good outcomes in the face of significant challenge) as a function of the availability of and access to significant levels of support; the resources serve to offset the cumulative and significant levels of risks encountered. However, a question infrequently posed is whether one might have access to excessive and unacknowledged protective factors and, accordingly, too many supports?

Group membership and the availability of resources as supports and assets, as suggested, matter. Consequently, significant resource availability may either serve as protective factors which balance out normative risks or, on the other hand, exacerbate the risk level. With the exception of Luthar and colleagues (e.g., Luthar and Becker, 2002; Luthar and Latendresse, 2002), this theme is infrequently addressed; nonetheless, significant and unacknowledged social, emotional, or economic opportunity and asset access (e.g., as group status associated entitlements or privileges) may further complicate youths' healthy transition into

adulthood and compromise self-efficacy efforts. Simply stated, an overprotection against risk exposure and normative youth experience may contribute to inadequate learning opportunities, ordinarily afforded by confronting challenges.

For example, affluence and socially bestowed entitlements obscure learning opportunities obtained from instances requiring the confrontation of challenges and appropriately responsive and in-the-moment positive-reactive coping. Confronting and coping with normative developmental challenges contribute to (1) coping-skill development, (2) nurturing of competence (given successful results), and (3) establishing positive self-efficacy beliefs. On the other hand, the diminishment or absence of coping opportunities may result in maladaptive responses (e.g., drug use, suicide, and an inaccurate self-concept) when effective coping is most beneficial. In such circumstances, proactive learning opportunities are denied or sheltered against by disproportionate access to socially situated, significant, and unearned privilege (see Luthar and Becker, 2002; Luthar and Latendresse, 2002) (**Figure 1**). As described in **Figure 1**, habitually low-risk experiencing youth, although having a significantly high presence of protective factors or supports, are described as untested and having undetermined vulnerability. On the other hand, individuals with high risk but possessing high and significant levels of protective factors, are viewed as resilient (i.e., they obtain good outcomes in the face of significant challenge) (see Spencer, 2006, 2008; Spencer *et al.*, 2006).

As indicated, all humans experience some level of danger, represented as different levels of vulnerability with varying outcomes (see **Figure 1**) thus, the character and content of socialization and training opportunities is critical. In fact, the role of preparing youth for the future is an especially vital function that requires patience, commitment, and adequate knowledge. It is a responsibility that cannot be taken lightly given the many and varied societal challenges that exist. Accordingly, adolescents developing in exceptionally challenging neighborhoods may find it especially difficult to focus on issues of consequence (e.g., school achievement) that are critical for adult success. Not surprisingly, varying contextual obstacles may present difficulty for financially disadvantaged youth, and particularly, those of minority status. Considered collectively, adverse factors, such as a lack of personal

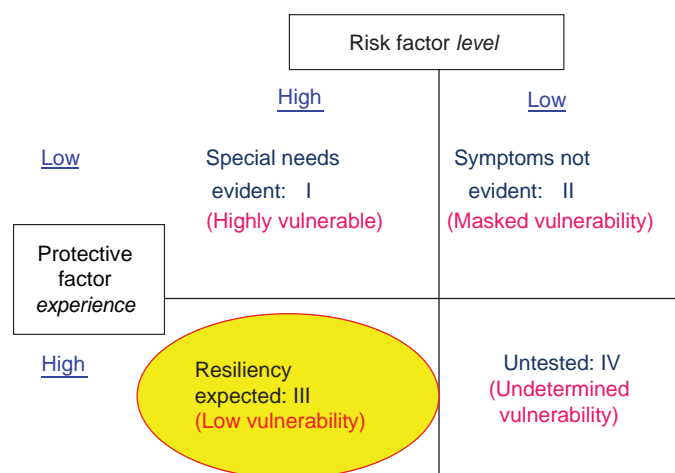


Figure 1 Phenomenological variant of ecological systems theory (PVEST)-linked vulnerability level and resiliency-prediction quadrant model. From Spencer, M. B. (2006). Phenomenology and ecological systems theory: Development of diverse groups. In Damon, W. and Lerner, R. (eds.) *Handbook of Child Psychology*, 6th edn, vol. 1, pp 829–893. New York: Wiley Publishers; Spencer, M. B. (2008). Phenomenology and ecological systems theory: Development of diverse groups. In Damon, W. and Lerner, R. (eds.) *Child and Adolescent Development: An Advanced Course*, pp 696–735. New York: Wiley Publishers; Spencer, M. B. 2004; and adapted from Anthony, James (1974). Introduction: The syndrome of the psychologically vulnerable child. In Anthony, E. J. and Koupernik (eds.) *The child in his family: Children at psychiatric risk*, pp 103–148. New York: Wiley.

financial resources, awareness of the inequality of opportunity, along with low-quality schooling, and harsh neighborhood contexts, reinforce persistent achievement gaps (Spencer, 1999, 2001).

The prevalence of an achievement gap between affluent and financially disadvantaged students is well documented (Hilliard, 2003). There is substantial evidence that students of color from low-socioeconomic backgrounds are disproportionately taught by less-qualified teachers and attend deteriorated schools that are racially and socioeconomically isolated (Darling-Hammond, 1997). In comparison with the college campus-like educational facilities provided by many affluent suburban communities, the attendant risks associated with economically strained urban schools and contributed to by teaching inadequacies and physical plant character, are obvious (see Spencer *et al.*, 2004; Spencer and Harpalani, 2004, 2008).

Accordingly, youth who develop in exceedingly challenging social contexts require attention that may look different from the mainstream. In order to address this issue, practitioners and researchers alike may benefit from understanding the distinctiveness of schooling experiences and therefore, adjust their actions. In this way, teachers, administrators, and parents can promote and make use of supportive strategies, which produce positive and productive youth development in the face of hardship by affording the necessary level and specific character of social supports. The specification is important given that all support, in fact, may not be experienced as supportive, given the history of undervalued American minorities (see Harpalani and Spencer, 2005; Spencer, 2005, 2008). Consistent with White's (1959, 1960) competence theorizing and explanation of

effectance motivations, adolescent students may positively and appropriately respond to the multitude of challenges and exhibit resiliency (i.e., obtain positive outcomes in the face of challenge) (refer to **Figure 1**). This most often occurs when young people believe that their efforts have a positive impact on the environment (i.e., manifested effectance motivation) (see White, 1959, 1960). Consequently, the level of academic and social resiliency of youth relies upon preparation and the quality and character of support offered to them by significant others in the varied settings transected by youth. Thus, the school environment should not be merely reproductive of traditional academic norms and standards but, instead, be transformative in the manner in which youth are taught and competence is achieved.

Accordingly, this article focuses on what factors contribute to a functional support system. The following ideas offer a theoretical perspective on how adolescents may best develop a sense of competence given any number of burdening factors. In particular, we consider how youth, particularly low-resource and minority-status youth of color, obtain and experience educational preparation as a source of support given the goal of academic self-efficacy and resilience. The solidification of these concepts endows youth with a sense of power and status, and facilitates a sense of competence. Fundamentally, educators benefit from understanding that learning is a developmental process and a risk-taking exercise. The context should provide a setting which communicates a sense of trust for engaging in learning as risk taking. Additionally, in order to serve youth more adequately, educators and school administrators would be wise to consider the perception-based cognitive dynamics around the formation of adolescent beliefs

about power, equity, and status. If not adequately informed by the insights described, seemingly proactive administrative and teaching efforts may waste resources, that is, and as suggested, intended supports may not be experienced as such and, instead, actually result in feelings of greater challenge due to unintentional systems injury (Spencer, 2006, 2008).

Background

Despite what has come to be the general public's perception of African-American children's desire to achieve (see review by Spencer and Harpalani, 2008), Black youth in America set high academic goals for themselves (see Spencer *et al.*, 2003). Yet, these socialized values and aspirations may be disputed as a function of matriculating through an educational system that, among other things, may neither financially nor socially support youth's academic intentions and, instead, communicate low expectations. The dilemma continues to be particularly acute for African-American males (Spencer, 2002; Youngblood and Spencer, 2003).

In 1990, Simmons and Grady reported that in a southern Maryland suburb located just outside of Washington, DC, Black males and females performed equally as well as their white counterparts in both mathematics and reading until the third grade. However, during the fourth grade year, a sharp decline in criterion-referenced mathematics and reading tests results were reported. Additionally, the percentage of Black males in the top reading groups dropped significantly (from 23% to 12%) from grade four to grade six (Lewis *et al.*, 2006; Simmons and Grady, 1990).

Unpacking and determining the causes of such findings is an imperative. Hypotheses about the most salient contributing factors include the examination of the school environment, inadequate self-reflection opportunities, and limited identity-exploration opportunities available to students. The implementation of strategies that take into account the factors noted, are thought to have implications for enhanced self-concepts and the development of increasingly positive attitudes and behaviors toward education. Support, experienced as positive change, can lead to broadened intellectual interests, improved self-regulated learning, greater commitment to education, and the framing of specific vocational plans and long-term aspirations (Fournet *et al.*, 1998). As proposed, greater specification and exposure to supports, overall, have implications for resilient outcomes in spite of the persistence of socially constructed risk factors.

In fact, contrary to common public perception, African-American boys wish to complete school and many actually desire classes that are academically challenging. In a survey of 2250 African-American males in New Orleans, 95% reported an expectation to graduate from high

school. Forty percent stated that they believed their teachers did not set high goals for them and 60% expressed a desire for their teachers to have higher expectations for them (Lewis *et al.*, 2006). Yet, despite these findings, unfortunately, low teacher expectations are inferred by youth and begin to take shape at startlingly early ages. Garibaldi (1992) reported that of 318 teacher respondents, 60% of whom taught in elementary schools, approximately six in ten stated that they did not believe their Black male students would go to college. Strikingly, 70% of these teachers had taught for 10 years or more and 65% of these teachers were African American. These detrimental biases are also evident as a function of the rate at which Black males are excluded from the classroom. In particular, African-American male-student statistics indicate that they are disproportionately suspended and/or expelled, show poor scholastic performance, avoid academic engagement, and represent decreasing rates of college attendance (Garibaldi, 1992).

Children are undoubtedly perceptive, conscious and aware of many of the biases that surface in the classroom. As a result, they are unavoidably sensitive to their central placement within the social dynamic. Thus, experiencing these hardships may be detrimental to positive youth development. Experiencing the noted biases while entering into and navigating through adolescence may be even more salient for such youth. That is, the outcomes lay the groundwork for additional critical-developmental tasks to be achieved for successful transitioning into late adolescence and early adulthood.

As previously noted, adolescence is a difficult stage to navigate due to the constant changes in one's physical and cognitive development. It is a time of heightened sensitivity to the many changes occurring both internally and externally. Consequently, adolescence may also bring about negative changes, including increased anxiety about school performance, social comparisons as a basis for assessing ability, confusion about the causes of one's academic outcomes, and declines in intrinsic motivation (Zimmerman *et al.*, 2000).

The negative psychological and behavioral outcomes noted may be exacerbated, while sources of protective factors, such as acknowledging youths' strengths, abilities, and cultural traditions, often go ignored. The lack of attention to cultural variations and opportunities to utilize them as strengths and sources of support may, in fact, cause their value or inferred relevance to a particular cultural context to wane. When Black youth are socialized to devalue the experiences of their own culture and ignore the historical and contemporary significance of issues facing their own ethnic groups, their psychological and behavioral well-being is affected (Baldwin *et al.*, 1987; Lewis, 2001).

For a better understanding of the phenomena of adolescent self-efficacy and resilience as it relates to perceived

competence, invoking a theoretical perspective is useful. In this instance, Spencer's phenomenological variant of ecological systems theory (PVEST) (Spencer, 1995, 2006, 2008) is useful when assessing how adolescents perceive, interpret, and act upon the stimuli that form their educational context. Given the task at hand, PVEST is beneficial due to its emphasis on meaning making as a potential contributor to youths' vulnerability as well as resiliency. Additionally, the systems theory views the self-system processes as dynamic, as well as recognizes the bridge between the individual's actions and societal context.

Taking into account that all humans are vulnerable, the PVEST framework consists of five components linked by bi-directional processes (Figure 2) (see Spencer, 1995, 2006, 2008), forming a cyclic, recursive model that explains identity development throughout the life course. The first component, net vulnerability level, represents the relationship or potential balance between risk factors versus protective-factor presence (e.g., protective factors as cultural capital). For urban minority youth, risks may be exacerbated by low socioeconomic conditions or sociocultural factors such as race and sex stereotypes. The second component, net stress level (i.e., challenges vs. supports), mediates between net vulnerability level and component three (i.e., reactive coping responses). The latter represents the state of balance between in-the-moment responses of maladaptive as opposed to adaptive strategic responses to confronted challenges. Particularly important at adolescence,

an individual's patterned reactive coping response (i.e., either maladaptive or adaptive), over time, becomes an internalized emergent identity (i.e., component four of PVEST). As illustrated in Figure 2, the emergent identifications can be either negative or positive. It is the consistent identity-linked set of behaviors that lead to component five (i.e., either unproductive or productive outcomes). Component five represent patterned sets of outcomes which have life-course implications. That is, as a productive or unproductive pattern, given the sets of developmental tasks associated with each developmental period of the life course, outcomes at one stage contribute to the subsequent period's status of vulnerability (i.e., risk vs. protective factors) (refer to component one of PVEST in Figure 2). As suggested by its cyclic character, the PVEST framework recycles and recurses through the lifespan as individuals balance new risks against protective factors, engage new stress levels, given challenges potentially offset by supporters, try different coping strategies, and redefine how they and others view themselves. Throughout the process, as individuals navigate across time and place, new developmental tasks are presented, which require responses. Thus, in many ways, the theory postulates an unchanging state of human coping. As noted, vulnerability (i.e., both protective and risk factors) is consistently a part of human development across the life course. Adolescence, in and of itself, holds rather fascinating challenges given pubertal changes and the

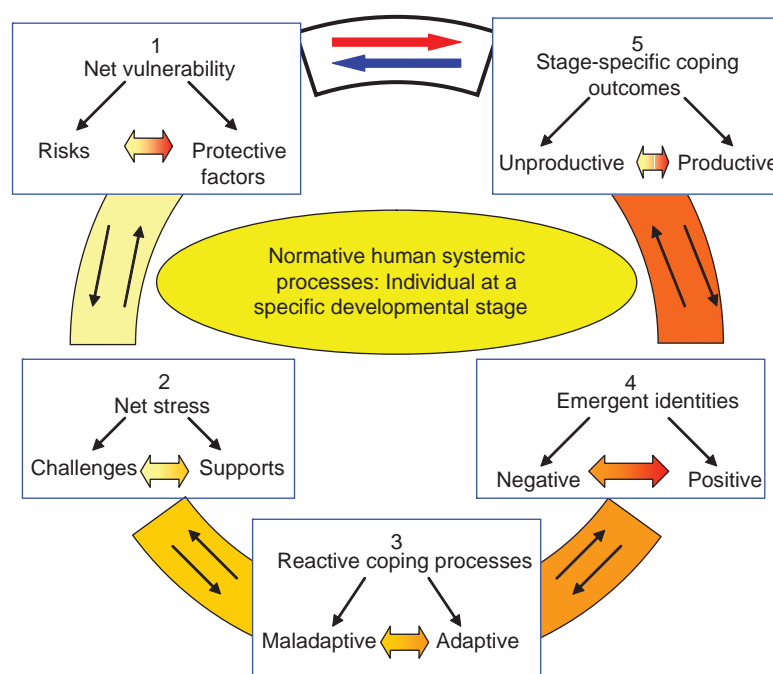


Figure 2 Phenomenological variant of ecological systems theory (PVEST). From Spencer (1995); (2004) Rev; Spencer, M. B. (2006). Phenomenology and ecological systems theory: Development of diverse groups. In Damon, W. and Lerner, R. (eds.) *Handbook of Child Psychology*, 6th edn., vol. 1, pp 829–893. New York: Wiley Publishers; and Spencer, M. B. (2008). Phenomenology and ecological systems theory: Development of diverse groups. In Damon, W. and Lerner, R. (eds.) *Child and Adolescent Development: An Advanced Course*, pp 696–735. New York: Wiley Publishers.

period's critical transitional role in preparing youth for the longest period of the life course – adulthood. Self-efficacy is an important protective factor that promotes the process.

Self-Efficacy

A number of factors, including stress and anxiety, may decrease one's level of self-efficacy. Particularly for socially marginalized youth (e.g., minority youth of color, immigrant, and special-needs young people), the additional social and financial barriers frequently encountered can heighten levels of anxiety and stress, which may interfere with academic performance and thus, the development of positive perceptions of self as a learner.

Cultural dissonance, which is understood to be the cognitive and emotional stresses resulting from conflicting cultural values or norms confronted by individuals in culturally salient situations, are hypothesized to be linked with unavoidable social cognitive processes (Spencer, 1982, 1984, 1985; Spencer and Markstrom-Adams, 1990). That is, one's cognition-based awareness of inequalities contributes to experienced cultural dissonance given personal values, desires, and interests.

A prominent hypothesis used to describe the discontinuity or discontinuity between inferred-context character and individuals' goals and desires has been described by Steele (1995) as a stereotype threat. This occurs when an individual of a socially stigmatized group is aware of certain task-related negative perception(s) held about their group. This consequently influences their ability to effectively conceive of themselves as able to successfully perform particular tasks. Otherwise stated, competing assumptions contribute to self-questioning about one's inferred level of efficacy.

In order to establish a sense of competency, adolescents' feelings of self-worth must remain intact and is associated with a belief that their efforts make a difference (see White, 1950, 1960). Hence, an undermining of confidence interferes with the ability to sustain a sense of power and/or control in a given situation; particularly those that are related to the development of academic self-esteem (see Hare and Castenell, 1985). This sense of authority over the direction of one's life trajectory and the various elements that may arise as daily hassles, often referred to as locus of control, endows youth with an important sense that they can enact change, or as classically described by White (1959, 1960), suggests effectance motivation. Accordingly, youth realize that they have the ability to function as active agents in their own lives, rather than existing as passive beings without a sense of agency. As indicated, past research has established academic self-efficacy as a fundamental adolescent coping mechanism used to navigate the educational system. Coping methods result in specific

products that may then be defined as demonstrations of competence (Anthony, 1987; White, 1959). Successful outcomes are often underacknowledged when associated with significant risk conditions and social challenges; that is, resiliency is underacknowledged. However, resiliency is a normal aspect of life for many successful young people.

Resiliency

As suggested, although resilience characterizes the daily experience of many lower-income minority adolescents (i.e., good outcomes are achieved in the face of significant challenge), for highly vulnerable youth, the stress that accompanies poverty, ethnicity-related social inequities, and chronic unchecked violence, may undermine self-efficacy beliefs and achievement efforts (Chestang, 1972; Hacker, 1992). Still, relatively little research has been conducted that focuses on the underlying processes of resiliency as observed among minority-status children of color. However, protective factors are also present and certainly a matter for offsetting the risks described (see Spencer *et al.*, 2006). Spencer *et al.* (1998) found that one's in-group cultural-identity processes support greater resilience of youth during periods of unusual stress. Their findings indicate that both academic and mental-health outcomes were positive and significantly higher for middle childhood and early adolescent youth who valued their own ethnic heritage as opposed to identifying with the dominant group.

Educational Preparation

Undoubtedly, education is one of the most influential tools for shaping and replicating a number of societal dynamics, namely, cultural norms and attitudes. Schools are one of the most prominent institutions for promoting mainstream ideology by contributing to students' development of a sense of organization and expectation. In fact, they are determined by the curriculum and structure of the school, as well as, the character of interactions experienced among students, teachers, and parents (Adler *et al.*, 1992).

The notion of educational preparation is concerned largely with the process of meaning making. Fundamentally, the manner in which individuals make meaning of the world is linked to their experiences with education, preparation, and competence. There has been relatively little research conducted in the area of educational preparation of urban-minority students. Furthermore, the research produced is not without flaws.

Essentially, educational preparation consists of the development of skills, knowledge, and the characters of youth. The formation of these elements enables adolescents

to create an environment in which they play a role in the planning, implementation, and modification of their own learning processes. Thus, while serving in diverse educational roles, they can reflect on their growth and, in turn, respond proactively to what they deem as societal needs.

Accordingly, socializing agents as individuals in position to influence outcomes should (1) seek to develop a knowledge base that is comprehensive and directed to the individual's needs; (2) identify and enact dispositions which enable them to be perceived and experienced as understanding, respectful, and inclusive in their creation of nurturing learning environments for diverse learners; and (3) hone skills which enable them to plan, implement, and assess appropriate instruction (Gagne *et al.*, 1988). Therefore, it is critical that educators prepare to be effective instructors who are responsive to the needs of ethnically diverse learners (e.g., including students of color). This requires sustaining students' sense of competency and enhancing resiliency outcomes. The process is facilitated by providing youth with supportive experiences, which produce a sense of personal causation (De Charms, 1953) or a belief that they can have a constructive impact on their environment.

Too often, there is a lack of understanding of cultural context. The patterned failure to appreciate the cultural components of the lives of young people contributes to misunderstanding about their outward expressions of self-hood (i.e., behaviors and attitudes). Furthermore, behaviors and struggles that result in success are often overlooked by authority figures since too many adults, specifically responsible for youth socialization, remain ill-equipped to recognize strengths as well as challenges. This significant shortcoming contributes to a failure to accurately assess these populations of students. The identification of influential and positive factors and their incorporation into the implementation of constructive interventions continue to be underrecognized albeit pivotal for the positive development of minority and highly vulnerable adolescents more generally.

Specifically, providing and teaching from a culturally relevant text is essential for the positive development of African-American youth. Education is aided when youth are provided with philosophical and academic material that is grounded in historically accurate content, which then facilitates the transmission of African and African American culture (e.g., see Spencer, 1990; Spencer and Dornbusch, 1990; Spencer and Markstrom-Adams, 1990; Spencer *et al.*, 2003). Teaching students to think critically and to understand their place in history can aid in the acquisition of skills and the accomplishment of this goal. In this way, children are able to recognize in a better manner, the importance of African heritage and the group's accomplishments under conditions of patterned adversity. Therefore, all of the above are important elements in developing socially responsible and engaged youth.

Overall, the goals of educational settings may well benefit from perspectives that recognize and enhance communal orientation, increase youths' sense of school connectedness, increase their motivation to achieve, increase their involvement in positive social change, and thus, improve their academic achievement. Although the noted values are frequently acknowledged, unfortunately, themes such as communalism frequently remain unacknowledged and are infrequently addressed.

Communalism, according to Lewis *et al.* (2006), refers to "attitudes and beliefs that one maintains concerning the significant and meaning of communal world experience and how one defines oneself in relation to one's community" (p. 5). Individuals with a communal orientation act in ways which communicates the notion that duty to one's social group is more important than individual rights and privileges (Boykin and Bailey, 2000). School connectedness refers to the extent to which one feels personally accepted, valued, supported, and encouraged by peers, teachers, and other adults in the school social environment (Goodenow and Grady, 1993). Additionally, an individual's degree of connectedness refers to the extent to which one feels a part of or aligned with the activity of one's context. For youth, the overwhelming majority of their context includes the classroom setting. As a function of their perceived and dynamic sense of synchrony, a young person's feelings of high connectedness correlates with open communication, high expectations from teachers, positive peer interactions, and motivation to achieve academic excellence (Arroyo and Zigler, 1995).

Accordingly, those students who feel alienated, are less likely to be motivated to achieve. Logically, they may demonstrate their sense of alienation by inserting their presence in other ways that are misunderstood or inferred to communicate disinterest in learning (e.g., acting out in class) (see Spencer, 1999, 2001, 2006).

Academic motivation refers to one's desire to learn and excel (Donohue and Wong, 1997).

Student motivation to achieve is among the top quartile of variables having strong effects on student achievement outcomes (Hudley, 1997; Weiner, 1990). Heightened motivation to achieve is of particular salience for Black adolescents in traditional mainstream educational environments (Hudley, 1997).

Knowledge as Power and Status

Understandably, the involvement of adolescents in some amount of change (e.g., social, educational, and emotional) in their environment may serve to fuel their desire to learn. The idea that one can alter the dynamics of a specific entity may serve as a fuel for one's desire for knowledge. Youth may have a stronger desire to achieve if they perceive their efforts and accomplishments to have

a significant impact on the world around them. Too frequently and as a function of marginalization, minority youth in particular, may feel that the efforts they put forth have no impact on the world around them, thus feeling that these efforts are in vain. Hence, endowing youth with the sense that their knowledge and efforts can generate positive change, can potentially increase their willingness to participate in the learning process.

Additionally, knowledge can be viewed as an indication of social position and, thus, a tool for altering one's status. As stated previously, this can be accomplished, in part, through the structure of school programming. One of the ways speculated for carrying out the task of providing cultural competence is via emancipatory education (Freire, 1998), generally described as African-centered education (see Madhubuti and Madhubuti, 1994). As referred to in the literature, both emancipatory and African-centered schooling involve a process of training that is intended to liberate underrepresented groups from racist ideologies and social institution in contemporary society (Lewis, 2004). Emancipatory schools attempt to connect African-American culture with Africa through the introduction and infusion of Afro-centric cultural approaches into curriculum and training (Akoto, 1992).

Emancipatory education, specifically, has been shown to increase communal orientation, school connectedness, motivation, and social-change involvement (Lewis *et al.*, 2006). However, participants in the program also felt more individualistic and competitive, which are prominent characteristics of mainstream American society and, when adopted, show utility for obtaining success in and beyond the classroom. Thus, adolescents who adopted either of these values more readily excelled in school when compared against those who are more collectivistic and communal in their work habits.

Conclusion

Teachers and parents should encourage youth to engage in self-appraisals regarding their academic competence and efficacy. The strategy has been shown to be positively related to achievement motivation and academic persistence (Bandura, 1977). As described, teachers and administrators obtain greater success when they adopt the use of strategies that are supportive of culturally responsive pedagogy. The strategy described facilitates the awareness of students' cultural needs while enhancing youths' interests and experiences. The suggested strategy builds trust and facilitates the achievement of core competence and critical competencies (i.e., particularly through the use of effective questioning strategies), thus, increasing opportunities for engagement in constructive dialog around student needs (Villegas and Lucas, 2002).

The adoption and integration of the several ideas espoused by emancipatory-education pedagogy may serve to bridge the achievement gap. The perspective offered counters the long-term void between the mainstream norms that schools continue to practice and also acknowledges the culturally specific and contextual needs of minority students (see Lee *et al.*, 2003; Rogoff, 2003).

As put forth by Spencer (2001), conceptual models are needed for a better understanding of the benefits of resiliency-promoting factors from a contextually and developmentally sensitive perspective. Spencer (2001) states that individual behavior is guided not only by opportunities but also by the perceptions of those opportunities. Thus, resilient outcomes can be achieved in spite of barriers if one perceives those barriers to be surmountable.

Bibliography

- Adler, P. A., Kless, S. J., and Adler, P. (1992). Socialization to gender roles: Popularity among elementary schools boys and girls. *Sociology of Education* **65**, 169–187.
- Akoto, K. A. (1992). *Nationbuilding: Theory and Practice in African Centered Education*. Washington, DC: Pan African World Institute.
- Anthony, E. J. (1987). Risk, vulnerability, and resilience: An overview. In Anthony, E. J. and Cohler, B. J. (eds.) *The Invulnerable Child*, pp 3–48. New York: Guilford Press.
- Arroyo, C. G. and Zigler, E. (1995). Racial identity, academic achievement, and the psychological well-being of economically disadvantaged adolescents. *Journal of Personality and Social Psychology* **69**(5), 903–914.
- Baldwin, J. A., Duncan, J. A., and Bell, Y. R. (1987). Assessment of African self-consciousness among black students from two college environments. *Journal of Black Psychology* **13**(2), 27–41.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review* **84**, 191–215.
- Boykin, A. W. and Bailey, C. T. (2000). The role of cultural factors in school relevant cognitive functioning: Description of home environmental factors, cultural orientation, and learning preferences. *Technical Report No. 43*. Baltimore, MD: Center for Research on the Education of Students Placed at Risk.
- Chestang, L. W. (1972). Character development in a hostile environment. (Occasional Paper No. 3) Chicago: University of Chicago, School of Social Service Administration.
- Darling-Hammond, L. (1997). Teachers and teaching: Testing policy hypothesis from a national commission report. *Education Researcher* **27**, 5–15.
- Fournet, D. L., Wilson, K. L., and Wallander, J. L. (1998). Growing or just getting along? Technical and adaptive competence in coping among adolescents. *Child Development* **69**(4), 1129–1144.
- Freire, P. (1998). *Education for Critical Consciousness*. New York: Continuum.
- Gagne, R. M., Briggs, L. J., and Wagner, W. W. (1988). *Principles of Instruction*, 3rd edn. New York: Holt, Rinehart and Winston.
- Garibaldi, A. M. (1992). Educating and motivating African American males to succeed. *Journal of Negro Education* **61**(1), 12–18.
- Goodenow, C. and Grady, K. E. (1993). The relationship of school belonging and friends values to academic motivation among urban adolescent students. *Journal of Experimental Education* **62**(1), 60–71.
- Hacker, A. (1992). *Two Nations: Black and White, Separate, Hostile, Unequal*. New York: Scibners's Sons.
- Harpalani, V. and Spencer, M. B. (2005). Racism. In Fisher, C. B. and Lerner, R. M. (eds.) *Applied Developmental Science: An Encyclopedia of Research, Policies, and Programs*, pp 905–909. Thousand Oaks: Sage.

- Hilliard, A. G. (2003). No mystery: Closing the achievement gap between Africans and excellence. In Perry, T., Steele, C., and Hilliard, A. G. III, (eds.) *Young, Gifted, and Black: Promoting High Achievement among African-American Students*, pp 131–165. Boston: Beacon Press.
- Hudley, C. A. (1997). Teacher practices and student motivation in a middle school program for African American males. *Urban Education* 32(2), 304–319.
- Lee, C., Spencer, M. B., and Harpalani, V. (2003). Every shut eye ain't sleep: Studying how people live culturally. *Educational Researcher* 32, 6–13.
- Lewis, K. M. (2004). Black students in public education: Outsiders in their won classroom. *Psych Discourse* 35(11), 8–9.
- Lewis, K. M., Sullivan, C. M., and Bybee, D. (2006). An experimental evaluation of a school-based emancipatory intervention to promote African American well-being and youth leadership. *Journal of Black Psychology* 32(1), 3–28.
- Luthar, S. S. and Becker, B. E. (2002). Privileged but pressured? A study of affluent youth. *Child Development* 73, 1593–1610.
- Luthar, S. S. and Latendresse, S. J. (2002). Adolescent risk: The cost of affluence. *New Directions in Youth Development* 95, 101–121.
- Madhubuti, H. and Madhubuti, S. (1994). *African-Centered Education: Its Value, Importance, and Necessity in the Development of Black Children*. Chicago: Third World Press.
- Rogoff, B. (2003). *The Cultural Nature of Human Development*. Oxford: Oxford University Press.
- Simmons, W. and Grady, M. (1990). Black male achievement: From peril to promise. *Report of the Superintendent's Advisory Committee on Black Male Achievement*. Prince Georges County, MD: Prince Georges County Schools.
- Spencer, M. B. (1990). Development of minority children: An introduction. *Child Development* 61(2), 267–269.
- Spencer, M. B. (1999a). Transitions and continuities in cultural values: Kenneth Clark revisited. In Jones, R. L. (ed.) *African American Children, Youth and Parenting*, pp 183–208. Hampton, VA: Cobb and Henry.
- Spencer, M. B. (1999b). Social and cultural influences on school adjustment: The application of an identity-focused cultural ecological perspective. *Educational Psychologist* 34, 43–57.
- Spencer, M. B. (2001a). Identity, achievement orientation and race: Lessons learned about the normative developmental experiences of African American males. In Watkins, W., et al. (eds.) *Race and Education*, pp 100–127. Needham Heights, MA: Allyn and Bacon.
- Spencer, M. B. (2001b). Resilience and fragility factors association with the contextual experiences of low resource urban African American male youth and families. In Booth, A. and Crouter, C. (eds.) *Does It Take a Village?* pp 51–77. Mahwah, NJ: Lawrence Erlbaum Associates.
- Spencer, M. B. (2005). Crafting identities and accessing opportunity post Brown. *American Psychologist* 60(8), 821–830.
- Spencer, M. B. (2006). Phenomenology and ecological systems theory: Development of diverse groups. In Damon, W. and Lerner, R. (eds.) *Handbook of Child Psychology*, 6th edn, vol. 1, pp 829–893. New York: Wiley Publishers.
- Spencer, M. B. (2008). Phenomenology and ecological systems theory: Development of diverse groups. In Damon, W. and Lerner, R. (eds.) *Child and Adolescent Development: An Advanced Course*, pp 696–735. New York: Wiley Publishers.
- Spencer, M. B., Cross, W. E., Harpalani, V., and Goss, T. N. (2003). Historical and developmental perspectives on Black academic achievement: Debunking the acting White myth and posing new directions for research. In Yeakey, C. C. and Henderson, R. D. (eds.) *Surmounting All Odds: Education, Opportunity and Society in the New Millennium*, pp 273–304. Greenwich, CT: Information Age Publishers.
- Spencer, M. B., Dobbs, B., and Swanson, D. P. (1998). Afro-American adolescents: Adaptational processes and socioeconomic diversity in behavioral outcomes. *Journal of Adolescence* 11, 117–137.
- Spencer, M. B. and Dornbusch, S. (1990). Challenges in studying minority youth. In Feldman, S. and Elliot, G. (eds.) *At the Threshold: The Developing Adolescent*, pp 123–146. Cambridge, MA: Harvard University Press.
- Spencer, M. B., Fegley, S. G., Harpalani, V., and Seaton, G. (2004). Understanding hypermasculinity in context: A theory-driven analysis of urban adolescent males' coping responses. *Research in Human Development* 1(4), 229–257.
- Spencer, M. B. and Harpalani, V. (2004). Nature, nurture, and the question of How?: A Phenomenological Variant of Ecological Systems Theory. In Garcia-Coll, C., Kearer, K., and Lerner, R. (eds.) *Nature and Nurture: The Complex Interplay of Genetic and Environmental Influences on Human Behavior and Development*, pp 53–77. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Spencer, M. B. and Harpalani, V. (2008). What does “acting White” actually mean? Racial identity, adolescent development, and academic achievement among African American youth. In Ogbu, J. U. (ed.) *Minority Status, Collective Identity and Schooling*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Spencer, M. B., Harpalani, V., Cassidy, E., et al. (2006). Understanding vulnerability and resilience from a normative development perspective: Implications for racially and ethnically diverse youth. In Cicchetti, D. and Cohen, D. J. (eds.) *Handbook of Developmental Psychopathology*, vol. 1, pp 627–672. Hoboken, NJ: Wiley Publishers.
- Spencer, M. B. and Markstrom-Adams, C. (1990). Identity processes among racial and ethnic minority children in America. *Child Development* 61, 290–310.
- Steele, C. M. and Aronso, J. (1995). Stereotyped threat and the intellectual test performance of African Americans. *Journal of personality and Social Psychology* 36, 21–30.
- Villegas, A. M. and Lucas, T. (2002). Preparing culturally responsive teachers: Rethinking the curriculum. *Journal of Teacher Education* 53(3), 20–32.
- Weiner, B. (1990). History of motivational research in education. *Journal of Educational Psychology* 82(4), 616–622.
- White, R. (1959). Motivation reconsidered: The concept of competence. *Psychological Review* 66, 297–333.
- White, R. (1960). Competence and psychosexual development. In Riley, M. R. (ed.) *Nebraska Symposium on Motivation*, pp 3–32. Lincoln, NE: Nebraska Press.
- Youngblood, J. and Spencer, M. B. (2002). Integrating normative identity processes and academic support requirements for special needs adolescents: The application of an identity-focused cultural ecological (ICE) perspective. *Journal of Applied Developmental Science* 6(2), 95–108.

HIV Education for Low-Literate People: Transforming Students and Communities through Paulo Freire's Praxis and the Pedagogy of Action

N Z Haniff, University of Michigan, Ann Arbor, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Activism – To become advocates of a cause related to social justice.

Consciousness – To be aware of the complex systems and ideas that drive social injustice in people's lives and the awareness that overcoming these injustices requires your own participation.

Empowerment – Ensuring that the least privileged is provided with the skills and knowledge that they can own and use to improve their lives.

Exoticism – To reduce a culture to essentialist caricatures, for example, one does not really see Africa unless one sees animals, or all Africans live in huts.

Low literacy – Population which is literate but only functionally literate.

Pedagogy – A method of teaching which facilitates the learners engagement in the teaching.

Praxis – Putting theory into action.

Introduction

This project asks some key question relating to transformation of students and communities: How young men and women can be conscientized about social injustice (race and gender) and translate that consciousness as a core principle of their own transformation as activists? How can a teacher who is a gender and human immunodeficiency virus (HIV) activist in South Africa, the Caribbean, and the United States use the nexus between privileged students and the community to develop consciousness in these communities to see themselves not as victims but as participants who can solve the problems faced by their communities if given the right tools? How can empowerment be achieved not as an academic exercise but as an act of human rights advocacy for the least literate and most marginalized?

The response to this question has been a three-course curriculum lasting a year. The first part of this curriculum is the course Pedagogy of Empowerment: Activism in Race, Gender, and Health, which educates the student about the HIV menace among African-Americans. There are four very important objectives for this class. First, the

student must become fully informed about HIV and African-Americans. Second, this information provides the context for a lively discussion on race and gender. Third, the students must be theoretically grounded in both race and gender in the HIV epidemic in the African community. Finally, from this process, the students become involved in praxis through teaching the low-literate module on HIV to communities outside the university.

The module was developed in the Caribbean when the author was engaged in creating low-literacy modules to teach women about their bodies. The women in Guyana asked to be taught about HIV. At the heart of this methodology is the philosophy that people can take responsibility for themselves if given the right tools, and that literacy was not a requirement for education. This module was then taught in Guyana, St. Lucia, Grenada, Dominica, St. Vincent, and Jamaica with great success. At that time, the epidemic was beginning to emerge as a great threat in Africa, and South Africa was then on its way to becoming the country with the largest number of infections (today there are 5 million HIV-infected persons in South Africa). The author is of the opinion that this module could have had some impact in South Africa, since there were large numbers of people who were not literate and needed HIV prevention education. Through her teaching at the University of Michigan, the author intends to create cadres of student activists from among the students who took her classes and who then would learn this pedagogy and then go into these communities.

To a certain extent, many of the students would embark on a kind of peer education where they would not only teach their peers the module but any community or organized group that would listen. This perception of community engagement or community involvement is often presumed to be simply a matter of acquiring skills. Campbell and MacPhail (2002) argue that "Peer education has been described as a method in search of a theory." They then further discuss a conceptual framework of community and peer education based on the work of Paulo Freire, whose conceptualization of empowerment adds a more cognitive intellectual dimension, focusing on people's intellectual analyses. He argues that a vital precondition for positive behavior change by marginalized social groups is the development of critical consciousness. Freire's book *The Pedagogy of the Oppressed* is the theoretical text used in all of the three courses of this project.

The students must first be conscientized by examining their own education and by challenging their perception of activism, including their own privileged status. Who and what is their community? How do they define community? Do they have a community? The class itself is not structured in the traditional didactic educational style but through dialogs about race, HIV, and gender. The issues Freire raises drive the theory of their ultimate project, teaching the HIV module to their own communities. These discussions are sometimes difficult but they happen nonetheless. This is the backbone of the pedagogy of action. Freire's conceptual framework is used in all three of the courses building layers of understanding that enhance the students own critical consciousness and their own emerging awareness of the intrinsic value of this consciousness in teaching their own communities and communities outside the United States.

The reflection on conscientization and its role in teaching the module provides an important insight into the student's own inability to depart from systematic education. It is here that the concept of praxis, putting theory into action, becomes critical. The student no longer sees teaching the module as an act of charity but as an act of building democracy, allowing the silenced to have a voice and to own knowledge rather than become a receptacle of that knowledge. Here, HIV work is transformed into human rights work, making both the beneficiaries and the educators, human rights activists. The act of teaching HIV prevention is inextricably linked to empowerment of the community and their ownership of the information.

The module is oral in nature and requires very little writing, and it is resource independent, there is nothing to read, no handouts, and no literature. At the most it lasts 20 min and takes about 3 days to learn and become teachers for those who are totally illiterate and about 2 h for those who are educated. The groups must be small with no more than 10 persons. A normal session of teaching would last 3 half-day sessions lasting 2–3 h. During that time, the student will learn the module and teach it back to their peers first in English and then develop the module in their own language when it is done in South Africa. Through this methodology, thousands of people have been taught by first- and second-generation teachers.

The HIV Module

On paper the module looks like this.

Science Lesson

As this module was developed to help low-literate people teach HIV prevention, it really has no meaning as a written text. To write about this module is to write about the pedagogy of these five simple lines. It is elementary in nature – it is like looking at the alphabet – structured

in such a way that the least-literate person can do it. A is for apple, H is for human, I is for ink or I is for a big long word immuno-deficiency, and V is for virus. First you begin with a story. Let us say I am in the kitchen and I am cutting cabbage and I accidentally cut my finger; what will I see? Blood. If you take a big glass and look at your blood, you will see that your blood is made up of little dots. These dots are called cells. There are two kinds of dots: red dots called red blood cells, and white dots called white blood cells. The red cells nourish your body and send oxygen throughout your body. To demonstrate this, we have to ask everyone to breathe in and out saying that we are breathing in oxygen and exhaling as it travels around our body. The white cells protect one's body. Here, we draw soldiers and explain how the white cells protect our bodies. We say, for example, if you fall and cut your knee then the white blood cells would rush to that area and form a scab, which means that it is now protecting that area from further harm, and begin the process of healing the cut. We now go to HIV (Figure 1).

H is for human, which means us, not dogs or cats or goats but us, humans. I stands for a big long word, immuno-deficiency, which means that all those white dots or soldiers who form a wall of protection against any disease or illness our bodies acquire, get together and fight it off so we can defeat the illness. So immuno is the white blood cells – the soldiers protecting one's body. Deficient simply means that the wall is now weakened – its not working well, something is wrong with the white blood cells.

At this point, we will show that a soldier has lost an arm or another one has lost a leg, demonstrating deficiency. The V in HIV stands for virus. The virus is explained as bacteria that have entered the body (Figure 2).

The Four Ways of Transmission

Here, we discuss the four ways in which one can contract HIV:

H	1	A	F	L
I	2	I	A	U
V	3	D	C	C
	4	S	T	K
			S	



Figure 1 The immune system: The soldiers (white blood cells).



Figure 2 Immune deficiency: The soldiers (white blood cells) are compromised.

1. The first way is through sex. Depending on the audience we are speaking to, we describe sex as friction. During this time, we ask participants to put their hands together and rub them together. This is known as friction, which represents the process that facilitates the cuts and tears that can occur during sex. Since we are not able to see them, fluids, including blood from one partner, can enter the small cuts and openings in the other partner and HIV can be transmitted this way.

2. The second way is through needles. We ask them to make a cylindrical tube while joining their two hands (as a fist) together. We explain that the needle is a tube, in which blood can remain, which if it is shared can transmit blood and other fluids.

3. The third way is mother to child. Although there have been preventative measures to reduce this, we still discuss it as a way of transmission. An HIV-infected mother can give her child the virus through breast-feeding. Breast-feeding is very important because it is the best nutrition a child can receive; therefore, it is important that mothers are tested so that they can breast-feed their child without fear of transmitting any infection. The child can contract HIV from the mother in the womb, if there is any kind of disruption in the womb, before delivery. The other way a mother can transmit HIV to a child is through labor. The infection occurs when the child is being delivered. The child can exchange fluids from the mother through the child's tender skin and openings such as the eyes, mouth, and nose. This is very important from a woman's point of view, so that she knows how to protect herself and her baby during pregnancy.

4. Finally, we discuss the fourth way, namely blood transfusion. Although this is very rare too, if blood is not screened for HIV, transmission can be facilitated. In a hospital, blood transfusion is done to patients, at times, in a life-or-death situation. If that blood is not tested then the person receiving the blood can become infected that way. Nowadays, this is not such a big problem because, in most places, the blood is tested for the HIV virus.

AIDS

Next, we link HIV with AIDS. The acronym can be explicated as:

- A in AIDS stands for acquired; the word, acquired, means that something is given from one person to another person.
- I in AIDS stands for immunodeficiency, just like in HIV, and we break it into two parts. The immune part stands for immune system and the deficiency in AIDS means that one's soldiers are not only losing an arm or leg but they are dying all together. Here, we wipe out all the soldiers we draw, except one or two, explaining the

soldiers cannot fight the virus anymore, which accounts for deficiency that is now more extreme.

- D is for deficient, which means the white blood cells are defeated. Some people can fight the disease for a very long time and some cannot. If one's soldiers can no longer fight the HIV virus, then one contracts AIDS.
- S in AIDS stands for syndrome, meaning a collection of all the illnesses that the person begins to experience – diarrhea, losing weight, rashes, and many other illnesses, which can eventually cause death since the body can no longer resist infections. No one really dies of AIDS but they die from one of these illnesses, such as a cold, which can turn into pneumonia and cause death, sores that cannot heal, weight loss, or constant diarrhea.

FACTS – Prevention of HIV

FACTS are some things one can use to protect themselves, and can be defined as follows.

- F stands for faithful, which means staying faithful to oneself and to others. This is a feminist interpretation of faithful. We say the first act of faithfulness is to oneself; if life's dream involves going to school, getting into university, or being employed in a good job, then it is important to be faithful to this dream. In this way, one is faithful to oneself by making his/her life better and by making good decisions about one's priorities. If your partner wants to have sex without protection or wants you to have a child before you are ready then if you make the wrong decisions you will put your future and your life in danger. If you have one partner then you must stick with one partner as well. This might mean–
- A abstaining, having no sex until you are ready, not when your partner is ready. This is especially important when talking to young people but does not exclude adults who want this as a choice.
- C stands for using a condom or, in South Africa, condomise. If you are ready then you should use a latex condom.
- T stands for testing, which means get tested for HIV. It also means to get tested for STDs like gonorrhea and syphilis, which cause open sores on one's private parts, which can help in the transferring of HIV. Finally, S stands for 'stay away from drugs and alcohol'. These substances can cause inhibitions to be lowered, causing one to possibly make the wrong judgment. If one plans to participate in these activities, we encourage the person to always have a trusted friend who will help you make right decisions.

Then if you know a person who is HIV positive you must be a source of LUCK to them. This is a critical part of the module since it deals with stigma which is still affecting

HIV-positive people and preventing others from seeking help because of fear of rejection.

LUCK – Stigma Reduction

- L is for love; show love to the infected person. Sit with them, talk to them, if you loved them before they had HIV, you should love them after they have HIV – they are still the same person. Do not reject them or turn your back on them.
- U is for understanding. Understand what the disease is, be educated about it, and educate others about it. Know that you cannot get HIV from touching or caring for them, that you cannot get it from drinking from the same cup or sitting on the bed or toilet. You can hug them and talk to them. The virus cannot live very long outside the body and so it is important you understand all of this, the ways you cannot get HIV. Understand also what they are going through, how they must be feeling having HIV or AIDS.
- C and K are for caring and kindness. Take the time to show them love, take them to church, take them for a drive, visit with them, and help in any way you can to make their lives more comfortable and show other people they should not be rejected but accepted. Show them love understanding caring and kindness.

Today there are many drugs, herbs, and foods to help your white blood cells fight this virus. Therefore, it is important that you get tested as soon as possible. To get tested does not mean that you are going to die, but that you are going to live because now there are many ways to manage this virus than before. So know the facts and be the bearer of LUCK. And be a role model who will stand up for an HIV-positive person.

Teach-Back Methodology

Students who take The Pedagogy of Empowerment course must then teach this module to a community outside the university and have at least two people teach it back. There have been many problems with this component of the class. Many students do not have a community they can go to and are afraid to do this. Many times they teach their peers within the university. Others try to teach youth groups in their churches, often being disappointed because they cannot teach sex in the church. Students also try to teach in their high schools and find it is complicated, because the schools also cannot teach about sex unless there is a long procedure that entails getting each parent to sign off or the school board has to decide. What seem apparently easy turns out to be very complicated.

The teach-back component of this methodology is critical to its success. It is here that didactic education is

most challenged. When the students in the class first learn the module, all of them must teach it back. They cannot use notes; they must look at their audience and teach the same way they were taught, that is without writing or turning their backs to the audience to write on the board. Students are so unfamiliar with this kind of presentation without notes that they intuitively want to write the acronyms out. Thus, for example, the F in FACTS they will write out facts on the board. This is strongly discouraged since for low-literate audiences this communicates that they too must write on the board. We discourage this because it slows down the story, the orality of the module, and diminishes their ability to feel empowered to teach this module. The transformation for the farmer, for example, when he/she gets up to teach is that they have taken on the role of the teacher and this act alone is empowering. That they could teach such information is, for them, the first act of egalitarianism the module facilitates. Most important, however, is the almost universal reticence to get up and teach even among the very literate. After two or three times of hearing the module, everyone thinks they can teach this, but when they are asked to do so very few actually get up. It is at this point that Freire's idea about the educational system becomes a living testimony. The hierarchy of the teacher as banker and the student as depositary is an idea so intractable in education that students become resentful and afraid when they are asked to become owners of knowledge rather than receptacles. They are often afraid and this is exactly the same when community members are asked to teach. This reticence is seen equally among the educated as well as the illiterate.

The peasant feels inferior to the boss because the boss seems to be the only one who knows things and is able to run things. They call themselves ignorant and say the "professor" is the one who has knowledge and to whom they should listen. The criteria of knowledge imposed upon them are the conventional ones. "Why don't you" said a peasant participating in a culture circle, "explain the pictures first? That way I'll take less time and won't give us a headache" Almost never do they realize that they too, "know things." (Freire, 2002)

This process provides a pivotal teaching moment for the Pedagogy of Empowerment. It is at this point that the students examine their own education. Are they being educated for themselves or to serve their communities? Why do they have few contacts with their communities other than church and high schools? The Freirean methodology and the theory of this pedagogy become apparent for now the limits of the didactic educational system and the illiteracy of their own social justice skills are exposed.

Only by working with the people could I achieve anything authentic on their behalf. Never had I believed that the

democratization of culture meant either its vulgarization or simply passing on to the people the prescription formulated in a teacher's office. I agreed with Mannheim that "as democratic processes become widespread it becomes more and more difficult to permit the masses to remain in a state of ignorance". Mannheim would not restrict his definition of ignorance to illiteracy, but would include the masses' lack of experience at participating and intervening in the historical process. (Freire and Macedo, 2000)

There are schools dedicated to interactions with the community, but the agenda is to teach the student to become a professional person in that community, to be a social worker, to be a clinician, a nurse, a teacher, etc. Many classes require community participation where the community provides the reality context for the student's education. The greatest beneficiary of all these 'projects' is the student ("remember you are part of a system, the field placement system which uses a variety of actors to accomplish the educational mission of the school") (Gorbman, 2002). Supervision and time for the student who is being placed to 'help' the community is done at the largesse of the community organization, which is not, as a general rule, paid to do this work, that they are getting a 'university student' to work with them is seen as payment enough (Barton *et al.*, 2005). Many workers in these organizations feel burdened by this; often the students themselves are left to muddle through simply because teaching them requires time, which organizations under stress do not have. This does not mean that the experience is not successful for many but, as a political commentary, it is important to note that the students leave, complete their education at the university, get their degree, and become professionals.

Discussions with students who volunteer in many humanitarian and nongovernmental organizations reveal that many are often loath to consider themselves beneficiaries at all. They feel they are good people and that the Freirean methodology of the class challenges their good work. It is not that their good work is not valued, but one must be able to distinguish the difference between charity and development. This is not an easy pill to swallow for young people who are caring and want to make a difference. The difference between criticism and challenge often becomes blurred. Even though it is emphasized that they should not stop their charity work because it has intrinsic value, it is still difficult, because the intent of the challenge is to make them see how their privilege still makes them beneficiaries even when they are giving. This in part is what Freire refers to as "problem posing education." It is about reflection on action and then again action and reflection. The process of conscientization begins with problem posing education. And it is this that is at the core of the class on the pedagogy of empowerment.

Before the class on the pedagogy of empowerment, they had no idea of the level of the HIV crisis in the

African-American communities. This information was shocking to them. They felt HIV was the disease of the 'other' in Africa. They raised money for Africa and felt that because they were Americans this had been addressed. It had not. At the 2008 Aids International Conference, the CDC MMWR (2008) reported the HIV epidemic was much larger in the African-American community than they had thought.

The new estimates show that MSM (men who have sex with men) of all races and ethnicities, and African-American men and women, are the groups most affected by HIV. Fifty-three percent of all new infections in 2006 occurred in MSM. African-Americans, while comprising 13% of the US population, accounted for 45% of the new HIV infections in 2006, with an annual infection rate seven times higher than whites and almost three times higher than Latinos (CDC MMWR, 2007).

This information confirms what those of us who have been working in this field for many years knew. The numbers were high but few people were paying attention on a national scale. The CDC notes that the new estimate does not represent an actual increase in the number of new HIV infections. Analyses conducted by the agency indicate the annual number of new infections has remained fairly stable over the past decade, albeit significantly higher than was originally thought (CDC MMWR, 2008).

Still the HIV epidemic has not reached a tipping point in the African-American community, and these issues were the center of many dialogs in the class. The greatest significance of this awakening was the realization "going to Africa to work on HIV" is not more important than working on HIV in the United States. The first principle of the Pedagogy of Action is that HIV in South Africa is not to be seen as a site for charity. The commitment to addressing HIV should be in the United States as well as in Africa. HIV has been exoticised as an "African disease" and the cultural politics of the racialization of disease has led to its marginalization in the United States as an African-American or American problem. This has clearly contributed to the ongoing perception that Africa is a place which must be rescued.

Preparation for the Field

Before students could embark on the culmination the Pedagogy of Action course, they had to first enroll in the class entitled 'Preparation for the Field'. This was a critical class because, once the students had been conscientized about HIV, race, gender, health, and empowerment in the United States, they had now to confront the hegemony of the United States and its image in the world. Readings and discussions about the issues, ideas, culture, and politics of the young American student engaged in communities abroad were rigorously addressed. What and where is South Africa?

What are the myths and perceptions of this African country? Is Africa a place to be exoticised or to be taken seriously as a site of cultural significance, a place of ideas and learning? What does it mean to be an American in the world today? How do South Africans see Americans? What kind of American do you want to be and present to the world?

Although the majority of the students would have taken the basic The Pedagogy of Empowerment class, a few will have applied to the 'study abroad' program and this class with those who had already taken the Pedagogy of Empowerment class will then form a cohort to bring the other students up to speed. They will also have to learn the module and teach it in the United States before they left for South Africa. The readings covered the latest information on the epidemic and its history in South Africa. The political life stories and struggles of Nelson Mandela and Bishop Desmond Tutu were required texts. The ideal number of students in this group was 10, most were undergraduates in their last 2 years and a graduate student who worked as an assistant. They ranged from engineering students to the social sciences and humanities.

A critical examination of the United States was essential in a pedagogy in which American students go into the developing world to see the world and to engage in projects. The questions they must examine were the following: Is anything or anyone better than America? If you have been told all your life that the United States is the greatest country in the world and that you are the greatest, is it possible then for you to see anyone or anywhere else as your equal? (Haniff, 2008). These questions formed the ideological basis of the preparation for the field.

The Pedagogy of Action

The Final semester is spent on the project in which the students put all these ideas into action in South Africa. The student who is now more thoughtful and reflecting is now taken into an environment that is even more radical. To teach HIV prevention, for example, in the townships in South Africa is both a frightening and exciting possibility. On the one hand, to presume that privileged American students can actually come to South Africa and teach people who speak in many different languages was a kind of arrogance or naiveté. On the other hand, it was such a challenge as to achieve it would itself be an act of supreme transformation for the student and maybe the community. Since the fall of apartheid, the universities have been developing ways to include service as a part of academic education. The tension between learning as knowledge in academia and learning as knowledge grounded in the community is at the nexus of education for all developing countries, especially in newly emerging democracies like South Africa. The Pedagogy of Action is an example of this praxis, that responsibility to one's community does not

exclude theoretical rigor or academic excellence. This is why the University of Witwatersrand and the University of Zululand were keenly interested in the Pedagogy of Action. The numbers have increased every year from 60 graduates in the HIV module the first year to a 110 the following year at the University of Zululand. At the University of Witwatersrand, students developed the module in 10 different languages, which include Sotho, Zulu, Xhosa, Yoruba, Lingala, and so on. In 2007 we met 10 of 60 students who graduated in our program in 2006 – out of their own commitment to act, they taught over 4000 people in their home villages in Kwa-Zulu Natal. This is how education can be a force for true citizenship and democracy.

Transformation

The impact of the pedagogy of action can be looked at through the evaluations conducted at various phases of the pedagogy. Evaluations of this pedagogy have been done in the Caribbean and in South Africa. The difference is that in the Caribbean evaluation represents the voices of community activists who had used the pedagogy over 4 years at the time of the evaluation (Haniff, 2004). The South Africa component was evaluated a year after the students had taught the module in their communities and reflects the voices of the teachers and not the communities who were taught. The transformation of students is expressed in their project papers where they recollect and analyze incidents pivotal to their new thinking and awareness.

The Community Activist

The early development of the module when it was taught by the author and two facilitators directly to communities did not include the capacity-building component. The sustainability of the module was left directly in the hands of the communities themselves through activists, nurses, teachers, and the outreach workers. However, this was found to have limited appeal since these activists could not go out of the communities and were not in a position to train others. They could, however, continue to teach. It was after these experiences with the module that we decided to train, for example, university students in South Africa – who were basically peers of my students who were the first-line teachers. The ideas and analyses of the impact of the pedagogy on the community as a part of the evolution of the pedagogy provide an important voice in the power of the pedagogy of action.

Although revolutionary leaders may have to think about the people in order to understand them better, this thinking differs from that of the elite; for in thinking about the people in order to liberate (rather than dominate) them the leaders give of themselves to the thinking of the

people. One is the thinking of the master; the other is the thinking of the comrade. (Freire, 2002: 132)

It is good because going to other workshops where they give you handouts and the speaker get up then I say I have everything already so you don't pay attention, but this workshop you had to listen so you had to take in what was said. You could always jot down whatever you wanted but whatever was being said, it had to remain in your head. It also develops your listening skills, so you can listen carefully. I think that method was very good. Too many times we are given handouts and after the workshop we chuck it there and then we don't have time to go back to it and whatever knowledge was said in this workshop it remain with us because we were given the facilities to give it back and it remains with you rather than a handout.

You see when I tell people what I know they say how I know that and I am not a doctor, not even a nurse because you know what happen. Because in learning to know about health I learn to know about my own self. I learn that nobody is higher, we at the same level. I can do whatever any other person can do.

Having to go up (to teach the module first time) you always feel a little nervous, in fact in truth you always feel a little shaky. Another thing that made us able to do it like we were on one level. Like you and the facilitator came, make yourself at the same level, like we were one, but sometimes you go to workshops where you have people of different height, (class) but it is for them (the education is for them) but you are not them because you are a different class. So you feel you are not comfortable among these people, so you not open up to receive, to grasp much. But you all come and make yourself part of us, just like us so we felt much more comfortable with you so therefore we were able to grasp and to learn faster that keep people back when you go to a workshop and they rank you (look down on you) and they don't take heed of you.

The New Teachers at the University of Zululand

These are the voices of the students who were taught by the University of Michigan students in South Africa. They were not community activists but were the peers of the University of Michigan students, at the University of Zululand. These new teachers taught in their villages and field placements in the rural areas of Kwa-Zulu Natal. The module was developed by these students into Zulu and then taught to their communities using their own language.

One cannot expect positive results from an educational or political action program which fails to respect the

particular view of the world held by the people. Such a program constitutes cultural invasion, good intentions notwithstanding. (Freire, 2002: 95)

Well, I already know about HIV, but what I learned was the simplest way to make HIV seem not as bad and not as complex as it is. It gives me, a teacher, like an imagination, even when I explain it to little kids. You know, kids imagine what HIV is. You don't know how to explain, but this just gives me a picture of HIV. And it's not as complex.

Also, again, I would really love, especially, foreigners or donors to make it a point that when they come to Africa, assisting people, they should empower people, not let people be in a position that they will fold their arms and just receive whatsoever. It could be information, but foreign people, they must empower people to be effective.

You are now even free to tell the whole community-you are positive. They won't, even, like, maybe hit you or do something or hate you. I mean, this program or this module has made people realize that the fact that you're HIV-positive, it doesn't mean that you are a bad person in the community. Maybe they thought you are unacceptable within the people or you are different from other people. Now people are learning to accept one another.

The Students

This is an excerpt from a student's writing on the topic 'faith', which was assigned to examine whether students really trusted that their 'low literate' audience could be empowered to teach.

The generosity of the oppressors is nourished by an unjust order, which must be maintained in order to justify that generosity. Our converts on the other hand, truly desire to transform the unjust order; but because of their background they believe that they must be the executors of the transformation. They talk about the people, but they do not trust them; and trusting the people is the indispensable precondition for revolutionary change. A real humanist must be identified more by his trust in the people, which engages them in his struggle than by a thousand actions in their favor without that trust. (Freire, 2002, 60)

A lot of things changed with a man named Timothy "Teaspoon" Zulu. On the first day of teaching, Teaspoon was the man that had said to us that he would "never wear a condom" because he was an "African man," and a real African man cannot wear a condom. As soon as I heard this, I lost any faith in him that I had in the first place, which probably was not much to begin with. How was I going to change his mind? Nevertheless, after a couple of days, he would become the most phenomenal teacher I have seen to emerge from our work using the module.

I use the term “emerge” to purposefully denote the fact that the teacher inside of him was always there, but it had now been finally given a chance to come out. Teaspoon taught the module in a way I have never seen anyone teach it before or since. He had charisma, confidence. And he knew what he was talking about. If someone would have told me on the first day that Teaspoon would be the one to get up in front of the executive board—a room full of white men, and teach them the module so perfectly that it brought tears to my eyes, I would never have believed them. When I saw Teaspoon do this, I had two feelings: pride and surprise. Pride because I was so proud of what he was accomplishing, and surprise because I had such little faith in him to begin with that I was surprised at what he accomplished. Nevertheless after the feeling of surprise had left, a little bit of faith came to take its place. Teaspoon had proven me wrong, and shown me that I should have had faith from the beginning.

Through these experiences, I have come to see that faith is not a matter of science or religion, it is a matter of choice. I can choose to have faith in the people I work with, and it will help them and me in the long run. When I met Mr. Ahmed Kathrada the other day, he said some of the most [profound words that I have ever heard: “there is no such thing as defeat if you are fighting for a just cause”. I believe this means that there is absolutely no harm, but only good, in believing in people, trusting them, and putting your faith in them. It is a just cause to educate, and especially to empower, and thus no defeat may come from any endeavor to do so. In addition, I have learned from these experiences that I should not need proof to have faith in something. It is quite arrogant to not believe in something that cannot be proven, for who am I to decide what “proof” is? And I know that as a white person, I have been socially trained to be doubtful of black people and their capabilities. However, I never knew, or possibly wanted to believe, how much this social training really manifested itself in my daily life. I now see it is my responsibility to overcome this, and to have faith in people, for as Freire has so poignantly said, the person who works for the liberation of an oppressed group but continues to regard them as totally ignorant, is doing nothing but false generosity.

HIV AIDS and Impact of the Pedagogy of Action

These voices articulate the impact of the pedagogy of action. It is the immediate response to problem posing education, which is a method not only for the community but also for those in higher education. To challenge the acquisition of knowledge at the university level is particularly important for those coming from marginalized communities and those who are unconscious of the

consequences of the didactic form of education. It is an ideological position that premises education as praxis rather than education as careerist. Whose responsibility is it to transform the world and to ensure those locked out of literacy and formal education not become victims of their circumstances but rather empowered to act? It is the educated who must do this and it is the institution of education which must develop pedagogies that provide tools to equip the actors for social justice.

Many nongovernmental organizations have been struggling on the front lines with illiterate and functionally literate populations, predominantly the female. The victims of HIV are growing increasingly female and people of color. We cannot wait until these populations become literate. There have been innovative methodologies of empowerment that have addressed the immediate issues of poverty, education, and health. The Barefoot College championed by ‘Bunker Roy’ is one example.

The philosophy of Mahatma Gandhi is reflected in the work style and lifestyle of the College. Traditional knowledge, village skills and practical wisdom are given greater respect than paper qualifications, and reading and writing are not seen as essential. “Just because someone cannot read or write does not mean he or she is uneducated,” Roy points out. The barefoot approach clearly works, and there are now 20 such colleges in 13 states in India. Plus the college is now involved in training villages from other countries, including Afghanistan, Ethiopia, Bhutan, Senegal and Sierra Leone. “But the real achievement,” says Roy, “is the process rather than the result.” (Milnes, 2006)

The microcredit movement, which started in Bangladesh, also wrestled with dehumanizing poverty by developing methodologies that empowered the very poor to sustain and support themselves ensuring that their integrity was not compromised by systems that undermined their autonomy. Yunus (2001), the catalyst of the Grameen Foundation, has said:

It is not microcredit alone which will end poverty. Credit is one door through which people can escape from poverty. Many more doors and windows can be created to facilitate an easy exit. It involves conceptualizing about people differently; it involves designing a new institutional framework consistent with this new conceptualization.

The Pedagogy of Action has attempted to change not only the uneducated but more importantly the educated by teaching them the basic principles “of conceptualizing about people differently.” They must first respect the uneducated and be given new eyes to see the value in traditional knowledge and practical wisdom. “Just because someone cannot read or write does not mean he or she is uneducated” (Roy). The Freirean theoretical principles used in the Pedagogy of Empowerment, and the Pedagogy

of Action were developed to create a body of students who would not only become actors for good in the world but actors on the world stage where these American students must manage the hegemony of their nationality. Problem posing education as a methodology in developing 'HIV education for low-literate people' transformed both students and the communities. It is a pedagogy that has attempted to address the problematics of the privileged who must construe their responses to inequality and social injustice in their own communities and in the world.

Bibliography

- Barton, H., Bell, K., and Bowles, W. (2005). Help or hindrance? Outcomes of social work student placements. *Australian Social Work* **58**(3), 301–312.
- Campbell, C. and MacPhail, C. (2002). Peer education gender and the development of critical consciousness: Participatory HIV prevention by South African youth. *Social Science and Medicine* **55**(2), 331–345.
- CDC MMWR (2008). Analysis provides new details on HIV incidence in U.S. populations. MMWR September, 2008 Fact Sheet.
- CDC MMWR (2007). Racial/ethnic disparities in diagnoses of HIV/AIDS—33 States, 2001–2005. *MMWR Weekly* March 9, 2007.
- Freire, A. M. A. and Macedo, D. (eds.) (2000). *The Paulo Freire Reader*. New York: Continuum Press.
- Freire, P. (2002). *Pedagogy of the Oppressed*. New York: Continuum Press.
- Haniff, N. Z. (2004). Privileging literacy: Methodologies developed and used in the Caribbean and South Africa to teach HIV prevention and catalyze community empowerment to low and non-literate populations (oral presentation). American Public Health Association, November.
- Haniff, N. Z. (2008). *"Ten Thoughts" Convocation Address GIEU*. Ann Arbor, MI: University of Michigan.
- Milnes, A. (2006). Community Hero: Bunker Roy. http://www.myhero.com/myhero/hero.asp?hero=bunker_roy_06 (accessed September 2009).
- Yunus, M. (2001). *Banker to the Poor*. London: The University Press.

Further Reading

- Farmer, P. (2008). Challenging orthodoxies: The road ahead for health and human rights. *Health and Human Rights, an International Journal* **10**(1), 5–19.
- Grobman, L. M. (ed.) (2005). *What You Need to Know to Get the Most from Your Social Work Practicum*. Harrisburg, PA: White Hat Communications.
- Larson, G. and Allen, H. (2006). Conscientization – the experience of Canadian social work students in Mexico. *International Social Work* **49**(4), 507–518.
- London, L. (2008). What is a human – rights based approach to health and does it matter? *Health and Human Rights, an International Journal* **10**(1), 65–80.
- MacPhail, C. and Campbell, C. (2001). "I think condoms are good but aai, I hate those things": Condom use among adolescents and young people in a Southern African township. *Social Science and Medicine* **52**(11), 1613–1627.
- Meguid, T. (2008). Notes on the rights of a poor woman in a poor country. *Health and Human Rights an International Journal* **10**(1), 105–108.
- Roy, B. (2005). Education international development. <http://itc.conversationnetwork.org/shows/detail783.html> (accessed September 2009).
- Shevchenko, O. and Fox, R. C. (2008). "Nationals" and "expatriates": Challenges of fulfilling "sans frontiers" ("without borders") ideals in international humanitarian action. *Health and Human Rights an International Journal* **10**(1), 109–121.
- UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance (2008). Epidemiological Fact Sheet on HIV and AIDS: Core data on epidemiology and response. http://www.who.int/globalatlas/predefinedReports/EFS2008/full/EFS2008_ZA.pdf (accessed September 2009).

Relevant Websites

- <http://web.mac.com> – Sanjit (Bunker) Roy, Design with India.
- http://hdrstats.undp.org/en/countries/country_fact_sheets/cty_fs_ZAF.html.

How Are Discussions of Interdisciplinary Studies Linked to Diversity Discourse?

C N Spence, Spelman College, Atlanta, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

The introduction to varied ways of knowing and seeing the world is lauded as the cornerstone of the liberal arts learning environment. Throughout the academy, interdisciplinary and multidisciplinary scholarship and research are recognized as progressive academic initiatives. These initiatives are tailored to encourage understanding and respect for the healthy tensions that exist among the various disciplines. It is clear that recognition of the value of embracing diverse ways of knowing and seeing the world through the examination of academic disciplinary lenses is a priority within the higher education community. What remains a challenge is an equal priority to expose students to voices of difference represented by faculty, students, and senior administrative staff.

While most institutions of higher education are not required to welcome all who seek admission or to acknowledge the experiences and opinions of all who would be considered different, it is expected that the academy will be a place where diversity is represented and elevated as an institutional goal. Many institutions of higher education profess to share a common goal to be agents of intellectual, personal, and social transformation for those who inhabit their walls. The challenge for many of these institutions is how to accurately define, construct, and implement initiatives that are inclusive and responsive to the particulars of their institutional history, setting, and culture. Second, institutions of higher education are charged with ensuring that inequities that exist outside the walls of the academy are not replicated within their teaching and learning environments. Implicit in the interdisciplinary and diversity language rests an assumption that the optimal teaching and learning environment is a space for free-flowing ideas that are representative of a collection of ideas and students and faculty representing varied personal, social, and cultural experiences.

It is assumed that the exchange of sometimes similar and many times dissimilar voices serves to enrich the teaching and learning environment. Such an opinion was expressed by Supreme Court Justice Sandra Day O'Connor in the opinion of the *Grutter v. Bollinger*, 539 US 306 (2003) case when she asserted that the contested plan implemented by the University of Michigan law school met the compelling state interest test applied in instances where intentional preferences based on race or some other protected category are operative. In her opinion, she cited the research and assertions of others who found that there

are instructional benefits associated with diversity in the classroom. (A number of *amicus* briefs were reviewed by the Supreme Court. Briefs prepared by the military and representative of Corporate America spoke the benefits of having a diverse workforce.) The cited benefits included breaking down of racial stereotypes and increasing cross-cultural understanding (539, US 306). It is assumed that the classroom becomes a safe space for diverse views and life experiences as represented in course content and student composition. The classroom supports the necessary tensions that develop as difference is accounted for and accepted. The classroom can become a social laboratory where difference becomes the common denominator as diverse ideas, theories, and individual perspectives and experiences frame the teaching and learning environment.

The Legal and Social Context

Various amendments to the Constitution have provided clear legal standards for the equitable treatment of persons regardless of race, class, or gender. While these standards continue to be legally acknowledged, the day-to-day realities of persons of color and women operate within personal, social, and political spaces that reflect what Hooks (2004) refers to as "imperialistic white supremacists capitalistic patriarchal culture." Such recognition requires us to remain ever vigilant in our quest to demand race, class, and gender equity. Those of us who hold positions of influence within the academy must not shy away from discussions that interrogate racist and sexist ideas, social policies, and behaviors. We must enlighten others to the realization that the absence of blatant discrimination does not in any way suggest that marginalization and disenfranchisement based on difference is not present. Bonilla-Silva (2006) speaks very specifically about the absence of honest discussions about the existence of racism within society. His assertions can easily be transferred to the absence of honest discussions about sexism. This absence of discussion is supported by those who would suggest that the most significant indicators of a racist and sexist society have disappeared or at least substantially dissipated. This presumption is based on a number of social indicators and number counting that denies the institutionalized systems of racism and sexism in our society. Uplifting all those who have made it

obscures the reality of those who remain on the margins of society and asserts a false sense of color and gender-blind ascension for those who hold coveted places within narrowly constructed windows of opportunity.

Institutions of higher education can in many ways become complicit in this false sense of progress. The creation of an elite class of educated individuals, whose opportunities for status mobility far exceed their parents,' suggests to some that things have really changed. The challenge for the academy is to create a space where these realities are celebrated but are not seen as clear and convincing evidence that such critiques are irrelevant. Many classrooms across this country become the forums for critical discussions of the existence of systemic racial, class, and gender inequities that permeate all sectors of society, inform dominant discourse, and serve as a lens through which the day-to-day life experiences of persons of color and women are evaluated. The scholarship of teaching and learning provides the canvas for the active engagement of students in this process of societal evaluation and self-discovery that is anchored in a very intentional reliance upon interdisciplinary analysis of issues of race, class, and gender.

The task then of the academy is to help students not only to understand race, class, and gender discrimination law and how it impacts their individual and social lives, but also to empower students to see education as a tool for social transformation. Many scholars and theorists point to what they consider the obligation of higher education to make individuals become active agents in the learning process (Freire, 1993; Hooks, 1994). A second task of the academy is to create a setting where patterns associated with patriarchy, racism, sexism, classism, and heterosexism are not modeled and reinforced. The classroom must become a safe space for the interrogation of these systems of oppression and subordination. Such a belief is not shared throughout the academy.

Interdisciplinary Studies and Diversity Discourse

While interdisciplinarity is lauded within the academy as a way to embrace the various ways of seeing and knowing afforded by diverse disciplinary lenses, the reality of the richness afforded by differing racial, ethnic, class, and sexual preference realities is not given the same accolades. Even within the context of the discussion of the benefits of interdisciplinary teaching and learning environments, preference is assigned to those disciplinary thinkers who do not necessarily represent the mosaic of ideas and theories espoused by scholars of color or women. Thus, interdisciplinary knowledge is often narrowly confined to the ideas of those who have informed dominant discourse within their particular disciplinary field of study. Many

see the structure and methods of the US higher educational system as a mirror of existing societal systems based on hierarchy wedded to capitalist ideological perspectives that are Darwinian at best and are marked by vacuous discussions of diversity that merely reflect a commitment to increasing numbers of brown and black faces only. Such a system denies the importance of not only discussing but also advancing race, class, and gender equity in the context of the classroom. This denial and/or delayed response to discussions of race, class, and gender equity are fueled by the historical commitment to the exclusion of blacks and women in higher education. Contexts, however, change either voluntarily or involuntarily. The presence of minority-serving institutions, including women's institutions, and the current reality of the imbalanced ratio of males to females attending institutions of higher education call for a reexamination of the ways institutions of higher education can demonstrate a commitment to racial, class, and gender equity in structure, content, and process and within the teaching and learning environment.

The Women's College Context

The question of how best to educate women presents particular challenges as considerations are made regarding how to prepare students for their role in a society that does not necessarily reflect a commitment to gender equality. How functional is it to create what Maher and Tetreault (2001) refer to as a feminist classroom? A feminist classroom is women centered and engages students in discussions that provoke introspection and external inspection of social systems that dominate and constrain the lives of women and girls. A feminist classroom creates a safe space for women to engage theory in ways that are specifically meaningful in the context of the designated discipline, and even more importantly, in the context of their lives. Issues of gender equity are at the center of all classroom explorations.

Classrooms that mirror society replicate and reinforce dominant societal norms, values, and beliefs. The subordination and disenfranchisement of women and persons of color inform and frame most of our social institutions, including the higher education community. Thus, paying particular attention to the voices of women and persons of color becomes a priority within classrooms committed to individual and social transformation. Empowering students to challenge the *status quo* becomes an intentional progressive revolutionary tactic inside of some classrooms. Bell Hooks refers to this form of pedagogy in *Teaching to Transgress* (1994) as "the practice of freedom." She suggests that a progressive classroom is one where students are taught to challenge established boundaries and to push beyond them in an effort to find truth. Is it important to create a classroom space that is women

centered in pedagogy and context? If so, how does one construct this environment in a way that emphasizes the significance and value of epistemologies that are framed by what Foucault refers to as power/knowledge? The sharing of knowledge that is empowering and that changes the student to examine the origins of dominant thought.

The creation of a feminist classroom is antithetical to the dominant model of higher education. One must remember that the education of women was not the primary goal of higher education. This reality continues to influence classroom dynamics across the country, particularly within some specialist professional areas. Guinere (1997), when examining the law school classroom, found it to be a sometimes hostile space for women students. In describing the law school setting, she found that it was a male-dominated space. The structure, pedagogy, culture, and general patterns of interaction ran counter to what she felt were reflections of the commitment to gender equity. Law school classrooms have historically been disproportionately male spaces until the most recent decade.

A recent American Association of University Professors report points to the fact that women represent 39% of university faculty (Gender Equity Report, 2006). One assumes that the face of the professoriate will inevitably change as the number of females attaining PhDs continues to outnumber the male turnout. However, the increase in the number of females receiving PhDs does not fully address concerns related to the achievement of racial equity in society in general and within the academy. The percentage of persons of color attaining a PhD remains below 7%.

The academy has not necessarily been seen as a space where students would critique systems of inequality and to dismantle the compound effects of patriarchy, racism, sexism, and heterosexism for the decade or more, discussions focusing on diversity in the higher education community have taken on increased significance. Most of these discussions have been framed by questions of access and accommodation primarily for students of color, with particular emphasis on African-American students. The development of and the challenging of special initiatives and programs to ensure the presence of students of color have dominated campus discussions and have resulted in the creation of some diversity initiatives and the dilution and dissolution of others. These discussions have not paid particular attention to gender equity. This reality may be attributable to the fact that when discussions of affirmative action have been raised, the impact upon gender equity in higher education has been a secondary discussion.

A Ford Foundation white paper, summarizing deliberations with diversity scholars and higher education diversity leaders, found several criteria that should be evident within institutions sincerely committed to a diversity agenda. (A white paper written by Cynthia Neal Spence, PhD, titled 'Reframing the diversity narrative: Achieving

social justice within the academy' was prepared for the Ford Foundation project, Diversity, Inclusion, and Institutional Change.) First and foremost, the commitment to diversity must be communicated from the top. The senior leadership of the institution must not only articulate a diversity agenda, but also provide tangible evidence of the institutional commitment. Second, the institutional mission statement must reflect and/or provide room for interpretation that includes an institutional commitment to actualize inclusion and diversity. Third, institutions of higher education must tangibly acknowledge the importance of diversity through the development of core curriculum initiatives that challenge the *status quo* Western civilization paradigm and force students and faculty to step outside conventional notions of how to situate themselves in a world community. A fourth consideration is one focusing on the equity of outcomes for underrepresented students, faculty, and staff. Not only must their representation be increased, but they should also flourish and succeed at the institution at the same rate as whites within the academy. (Access to honors, tenure and promotion, fellowships and scholarships, research opportunities, and significant committee leadership for underrepresented populations should mirror the access and accomplishments of white students, faculty, and staff.) Finally, and most significantly, institutions of higher education must move beyond providing mere access to the university to diverse groups, and must address climate issues that provide greater guarantees that diverse groups' experiences within and outside of the classroom will be conducive to learning and succeeding in the academic environment. This final criterion speaks to the institution's demonstrated commitment to recruitment and retention of diverse faculty, administrative staff, and students.

It is these two final criteria that colleges and universities find most challenging. Institutional promotional materials often exhibit the faces of diversity. Many can speak of the number of students of color and women admitted to their institutions. The Ford Foundation report suggests that mere numbers and faces are not enough. Persistence rates and the quality of the educational experience are equally important to those institutions that are sincerely committed to diversity. Thus, the recruitment and retention of faculty, students, and administrative staff of color must be matched by an equal commitment to ensure a welcoming and supportive environment. Strategic initiatives must be developed and implemented that include outreach, recruitment, and retention of faculty, administrative staff, and students of color.

The nation's top-tier institutions operate in a competitive environment as each seeks to attract the best and brightest faculty, students, and staff. This competition among the nation's top-tier institutions has a profound impact upon the ability of second tier and small liberal arts colleges as they seek faculty, staff, and students from

the same small pipeline of potential applicants. The shared goal of all institutions that enter this competition is to create an institutional environment that is reflective of the demographics of the larger society.

The language of diversity, while symbolically representing a shared meaning that is very much situated in race and ethnicity discourse, takes on a particular contextual significance through institutional history, programming, and implementation. Often, left out of the conversations or rendered secondary are discussions of other identity markers, including faith traditions, sexual preference, disabilities, and other life experiences and choices that may serve to marginalize particular segments of the population. The dominant diversity discourse is consistent with the US Constitution's affirmation of protected classes within society. This language of protected classes includes those who would be discriminated against on the basis of race, ethnicity, sex, and religion. Thus, diversity language and practice are often more attentive to ensuring that blatant discrimination cannot be found against faculty, staff, and students because of their race, ethnicity, sex, or religion. It is important to note that gender is not a protected class but is addressed by the language of protection based on sex discrimination. Of the specified protected classifications, racial discrimination and discrimination of groups based on ethnicity have been the most dominant consideration in diversity discussions, particularly when discussed within the historically white colleges and university context. The sociohistorical significance of race and racial discrimination within the American context continues to take precedence over other identity markers.

Discrimination based on sex, which is also very deeply rooted in American society, receives almost equal consideration in discussions about diversity. One must note, however, the increasing gender imbalance nationwide within undergraduate student populations, with more than 50%

women and, in some instances, particularly historically black institutions, more than 60% female population. This phenomenon even has some institutions and special higher education programs considering affirmative action plans for men. Creating a diverse environment is often translated into making sure that persons of color with specific intent to include African Americans and other persons of color are present in the institutional mix. Some recent research reports, however, indicate that at many of the more elite institutions, racial diversity is manifested by a student body who are blacks, who are foreign born, or are the children of foreign-born parents. The faculty diversity numbers are also influenced by foreign-born blacks and other persons of color. According to US Census data, more than 12% of all black undergraduate students enrolled at US colleges and universities were born in a foreign land. Among black graduate students, 18.7% are foreign born. Most recently, a commitment to diversity is materializing through special initiatives that focus on the socioeconomic status of potential students. Thus, social class becomes the dominant consideration for persons who would not ordinarily be represented in the student body. The question that is at the center of most diversity discussions is: What is the impact upon the institutional climate and teaching and learning environment?

Bibliography

- Bonilla-Silva, E. (2006). *Racism without Racists*. New York: Rowan and Littlefield.
- Freire, P. (1993). *Pedagogy of the Oppressed*. New York: Continuum.
- Grutter, v. Bollinger 539 U.S. 306. (2003).
- Guinere, L. (1997). *Becoming Gentleman*. Boston, MA: Beacon Press.
- Hooks, B. (1994). *Teaching to Transgress*. New York: Routledge Press.
- Maher, F. A. and Tetreault, M. K. (2001). *The Feminist Classroom: Dynamics of Gender, Race and Privilege*. New York: Rowman and Littlefield.

Identifying Patterns of Doctoral Attrition Across Academic Fields

A James, Loyola Marymount University, USA

K J Fosnacht, E B Stolzenberg and D Thirumalai, University of California, Los Angeles, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Background/Purpose

As individual students are the most affected by doctoral attrition (Lovitts, 2001), in the past, university leaders were not necessarily concerned with the role the institution played in the students' decisions to leave their programs. However, economic constraints are forcing institutions to (re)evaluate all facets of the higher education enterprise, including doctoral education. In this era of perpetual fiscal crisis, institutional leaders are realizing that doctoral education is one of the most expensive areas of higher education due to the costs of recruiting and enrolling top students, maintaining small classes, expensive equipment, and facilities, and the individualized attention provided by faculty (Baird, 1993). Consequently, attrition from doctoral programs becomes a question of the efficient use of a university's scarce resources (Kluever, 1997; Lovitts, 2001).

When talented doctoral students leave before graduating, institutions fail to benefit from the significant roles noncompleters would have played in furthering research and educating undergraduates in their respective fields (Haynes, 2004). In addition to the negative effects of doctoral attrition on institutions, society also suffers when qualified doctoral students leave before completing the degree. While society can benefit tremendously when students' research sheds light on a problem (Katz, 1997), society also loses out when the social and scientific perspectives, questions, and ideas of noncompleters are not pursued (Lovitts, 2001).

Literature

Researchers agree that the program, department, and discipline are all central to the persistence of doctoral students. Benkin (1984) found that after controlling for demographic and academic factors, the student's field of study was the most significant predictor of doctoral degree completion. Bowen and Rudenstine (1992) also demonstrated that differences in completion rates are most typically associated with the field of study than any other factor. Various studies have shown that completion rates are lower in the humanities and social sciences and that completion rates are higher in engineering and the natural sciences (Baird, 1997; Bowen and Rudenstine, 1992; Ferrer de Valero, 2001; Haynes, 2004; Hodgson and Simoni, 1995; Nerad and Cerny, 1993; Nerad and Miller, 1997; Smallwood, 2004).

In addition, certain characteristics of academic fields play a role in doctoral persistence, such as the nature of research and the dissertation, funding traditions/opportunities, relationship and amount/type of contact with the advisor and other faculty, and individual characteristics that lead students toward graduate study in a particular field (Bair and Haworth, 2004). For example, women and non-Asian minority students tend to be more represented in programs that have lower completion rates, such as those in the social sciences and humanities (Nerad and Cerny, 1993). Further, women, minority students, and American students drop out at higher rates than men, White students, and international students, respectively (Bowen and Rudenstine, 1992; Lovitts, 2001; Smallwood, 2004).

The structure of the discipline shapes both the academic and social interactions of its members and is also responsible for the long-term patterns of attrition across academic disciplines (Bowen and Rudenstine, 1992). For example, the sciences are highly structured disciplines, advisers are selected early, funding is more stable, and there is frequent contact between the student and the faculty and among students (Bowen and Rudenstine, 1992). On the other hand, the humanities and many of the social sciences are more loosely structured, research can be isolating, students select advisers later on in the process, and there is not much social and academic support from faculty and other students.

Past research has shown that failure to complete the doctoral degree is not solely a function of intelligence, academic preparation, or finances (Baird, 1991, 1993; Ferrer de Valero, 2001; Lovitts, 2001; Smallwood, 2004; Tinto, 1993). Students with fellowships do not complete the degree at a higher-than-average rate, as National Science Foundation fellowship recipients complete at a rate of only 75%, comparable to nonfellowship recipients in the sciences (Smallwood, 2004). Types of financial assistance can have differential effects on student outcomes depending on the student's stage of doctoral study. For example, when beginning doctoral study, an assistantship may be more beneficial for persistence than a fellowship, as a research position can facilitate both academic and social integration. However, during the dissertation stage, a research assistantship may not allow the student enough time dedicated to completing the dissertation (Tinto, 1993).

With respect to academic preparation, Lovitts (2001) found that completers and noncompleters enter their doctoral programs with comparable undergraduate grade-point

averages (GPAs), but women had higher GPAs, with female noncompleters reporting the highest grade-point averages. She also found that those who did not complete the doctorate appear to be as professionally qualified, if not more so, at time of entry (with respect to prior socialization experiences, such as being mentored by faculty and subscribing to journals).

The relationships between students and faculty are crucial in the doctoral-education process. The adviser is often the most influential person throughout the doctoral-education experience. Students learn the nature of the academic profession from their adviser, and the adviser plays a key role in the students' socialization into the roles of researchers and teachers. The dissertation topic, scope, and its quality are also shaped by the students' relationship with the adviser. Subsequent job placement can also be affected by the nature of this relationship: "Affiliation with the proper adviser can often spell the difference between completion and noncompletion" (Lovitts, 2001: 131).

Theoretical Frameworks

Three theoretical frameworks underlie this study – doctoral persistence, graduate-student socialization, and disciplinary/department culture. Tinto (1993) believes that due to disciplinary differences and differences in the stages of the doctoral process, persistence at the graduate level cannot be described by any one developmental model. Factors that are significant at one stage may not be significant at the next stage.

Graduate-student socialization is not a linear, predictable process, and no two students are socialized in the same way (Wulff and Austin, 2004). Doctoral-student socialization is a dynamic process in which students influence, and are influenced by, their environment (Austin, 2002). Socialization in graduate school is the process through which students gain the necessary knowledge, skills, and ideals to enter a career that requires advanced knowledge and specialized skills (Weidman *et al.*, 2001). The socialization process is affected by the chosen discipline, the institution in which it takes place, and the structure of the graduate program (Weidman *et al.*, 2001). The point of reference for integration for undergraduates is the institution or campus community; however, doctoral students associate more closely with the academic department or program (Golde, 2000).

Finally, academic disciplines have distinct cultures that affect faculty and student beliefs and behaviors, such as values, faculty work, training of students, the nature of research (types of research questions and research methods), and faculty–student interaction (Austin, 2002; Golde and Dore, 2004; Lovitts, 2001; Smart *et al.*, 2000;

Turner *et al.*, 2002). Knowledge of distinct disciplinary settings is fundamental to understanding and improving key aspects of doctoral education, such as student retention and satisfaction (Golde and Dore, 2004; Smart *et al.*, 2000).

Research Questions

The aforementioned literature and theoretical frameworks guided the development of these research questions:

1. During what years are doctoral students most likely to depart?
2. What individual and department factors help predict doctoral attrition?
3. Do departments vary in their patterns of doctoral student departure?

Data Sources/Sample

To address these research questions, this study utilizes institutional records from a large, public research university in the western United States. The institution is divided into 14 broad academic fields, housing more than 80 doctoral programs. The university frequently participates in national studies of doctoral-student outcomes. The data set includes all students pursuing a doctoral degree (PhD, Dr.PH, DMA, or DEnv.) who entered the institution between the 1992–93 and 2004–05 academic years. Excluded from the study are students enrolled in interdepartmental programs and programs not administered by the graduate school. The institutional records comprising the data set include data from admissions, enrolment, degree, financial support, and employment records, combined with student-survey data. Faculty demographic data will also be included.

On average, about 782 doctoral students enter each year, resulting in a total sample of 10 166 students. These students are spread among 59 academic departments. **Table 1** contains the list of academic departments by broad field. About 44% of the sample is female. Over a quarter (28%) are international students, and about 11% are domestic underrepresented minority students (as designated by the institution, which includes Black/African American, Native American, Chicano/Mexican American, Latino, and Filipino students).

Methods

As doctoral attrition is an event that occurs over time, we first examined life tables and graphs of the student-departure process for the campus and each academic field.

Table 1 Departments included in the analyses by broad academic field

Health science professional
Nursing
Public health
Humanities
Applied linguistics
Art history
Asian languages
Classics
Comparative literature
English
French
Germanic languages
Italian
Linguistics
Musicology
Near east languages
Philosophy
Slavic languages
Spanish
Life sciences
Biological chemistry
Biology
Biomathematics
Human genetics
Microbiology
Pharmacology
Molecular biology
Neurobiology
Psychology
Physical sciences
Chemistry and
biochemistry
Earth and space
sciences
Mathematics
Physics and astronomy
Statistics
Professional schools
Architecture
Computer science
Education
Engineering
Ethnomusicology
Film studies
Information studies
Management
Music
Social welfare
Theater
Urban planning
World cultures
Social science
Anthropology
Economics
Geography
History
Political science
Sociology

To preserve confidentiality, departments have been renamed to generic descriptions and the departments within the schools of Engineering and Public Health have been removed.

The tables and graphs were subsequently compared to investigate when students are most likely to depart and if these patterns vary by academic field.

However, conducting the multivariate analyses on student attrition presents two methodological challenges. First, student attrition is an event that occurs not at a static moment in time, but rather over time. The failure to control for the time will result in biased estimates (Allison, 1984). Additionally, the students attending the university are nested in academic departments and not a random sample. The nesting of students violates the assumptions of traditional multivariate methods such as logistic regression (Raudenbush and Bryk, 2002).

Taking these concerns into account, we decided to conduct a discrete-time hazard analysis in a multilevel framework (Barber *et al.*, 2000). (Hazard analysis is also commonly referred to as survival analysis, event history modeling, and duration analysis, see DesJardins, 2003, for an overview of its application in education research.) This study represents the first application of this analytic technique to student attrition at either the undergraduate or graduate level. Discrete-time hazard models avoid the assumption of cross-sectional models that assume statistical equilibrium exists over time and allow for the inclusion of time-varying regressors leading to less-biased results (Allison, 1984; Singer and Willett, 1993). The models also allow for the investigation of when doctoral students are most at risk for departure. Additionally, in the study of student persistence and attrition, a discrete-time model is preferable to a continuous-time model, such as the Cox model, because the exact time of departure within a semester or year is unknown (Yamaguchi, 1993).

This study uses the academic year as the unit of time. To conduct the multilevel hazard model, the initial data set was converted to person-period form, where each observation represented one academic year. The first time period for all students is their first year of enrolment. A central concept behind hazard models is the risk set – the set of individuals who may experience the event of interest. Consequently, three types of observations were then deleted from the sample. First, observations occurring after departure were removed as the student was no longer at risk to drop out of the institution again. Additionally, two classes of observations were excluded due to right censoring. After receiving their doctoral degree, the student is no longer at risk, and these future time periods are right censored. Similarly, for students currently enrolled with observation periods not yet observed, the unobserved periods are also right censored. No observations were left censored as students cannot depart from college before enrolling. After the deletion of the censored observations, the study yielded a final sample of 44 962 person-year observations from 10 130 students.

The multilevel component of the analysis satisfies the assumption of independence for hazard models and provides an opportunity to investigate if doctoral-student attrition varies between academic departments. We employed a two-level random-intercept model using logistic regression.

The general form of the model can be expressed as

$$\text{logit}(p_{nij(y=1)}) = \beta_{0j} + \beta_1 \lambda_r + \beta_2 (X_{ij} - \bar{X}_{\cdot j}) + \beta_3 (X_{tij} - \bar{X}_{\cdot j}) + \beta_4 X_{\cdot j} + \beta_5 X_{tj} + \varepsilon \quad \beta_{0j} = \beta_0 + \mu_j \quad [1]$$

where β_{0j} is the intercept term for department j , β_1 is the coefficient associated with a vector of dummy variables representing the hazard function, β_2 is the coefficient associated with a vector of time-invariant group mean-centered variables for individual i in department j , β_3 is the coefficient associated with a vector of time-varying group mean-centered variables at time t for individual i in department j , β_4 is the coefficient associated with a vector of time-invariant variables for department j , β_5 is the coefficient associated with a vector of time-varying variables at time t for department j , μ_j is the random effect for department j , and ε is the error term.

Analysis Strategy

To begin the study, we first examined the life table for the sample, which summarizes the patterns of attrition by observation period. Derived from the life table is the hazard rate, which forms the basic concept behind the hazard model. The hazard rate or more commonly, the hazard, is the probability that “an event will occur at a particular time to a particular individual, given that the individual is at risk at that time” (Allison, 1984: 16). For this study, the hazard refers to the probability that a student will depart from the institution without receiving his/her doctoral degree in a particular academic year, provided he/she is at risk for attrition. From the life table, the survivor function, which is the proportion of the sample that has not yet departed and is at risk to depart in an academic year, is also derived.

Our data-analysis strategy proceeded to build a final model by level. First, we fitted an initial model with just the hazard function and the random intercept. Next, we added a number of student-level variables to the first model. Subsequent models were fitted to include the context the student experienced while enrolling, the academic field of the department, and a number of interaction effects. Finally, a subset analysis of students in departments requiring the Graduate Record Examination (GRE) for admission was conducted.

Dependent Variable

As stated above, the dependent variable is a dichotomized variable indicating student attrition in a given academic

year. This study uses the institutional definition of attrition, which is when a student does not enrol or file for a leave of absence for two consecutive years without receiving the doctorate.

Independent Variables

A full list of the independent variables examined in this study and their coding schemes can be found in **Table 2**. Variable selection was guided by previous research and the availability of data. In addition to variables indicating the hazard function, student-level variables such as demographic traits including gender, age, ethnicity, and citizenship were included in the analyses. Additionally, data on the student's prior history were assessed by including variables indicating baccalaureate institution type and prior graduate-degree achievement. The total amount of net financial support (after deducting mandatory tuition and fees) received by the student per year was included in the analysis. These values were adjusted for inflation to the 2000 calendar year using the consumer price index (CPI) measure for the local metropolitan area. Moreover, dichotomous variables indicating whether the student was a teaching or research assistant during the observation period were included. All student-level variables were centered at their department's mean.

We also included variables reflecting the department/program environment, such as the demographic make up of the department's ladder rank, faculty, and student body. In addition, student satisfaction with the quality of instruction was examined. The analysis included department means for a number of the student-level variables to examine contextual effects. Finally, GRE percentile scores at the individual and department level were examined in the subset analyses described above.

Results

Descriptive Statistics

Table 3 contains the life table for the total sample, and **Figure 1** demonstrates the survival function over the sample's time period. The median life span for the sample is 11 years, indicating that half of all students fail to finish their doctoral degrees by year 11. This statistic is similar to the national average for all doctoral students (Bowen and Rudenstine, 1992; Lovitts, 2001).

The hazard function can be found in the last column of **Table 3** and graphically represented in **Figure 2**. The risk of drop out in the first year is 6% and rises to 9% in the second year. Then, the hazard rate gradually declines until year 8, when it begins to rise again. The hazard peaks in year 13; however, this rate should be interpreted with caution as only 35 students were observed during this

Table 2 Variable list and coding scheme

<i>Variable</i>	<i>Description</i>	<i>Time-varying</i>	<i>Centered</i>	<i>Coding</i>
Year 2	Time period 2			(1 = yes, 0 = no)
Year 3	Time period 3			(1 = yes, 0 = no)
Year 4	Time period 4			(1 = yes, 0 = no)
Year 5	Time period 5			(1 = yes, 0 = no)
Year 6	Time period 6			(1 = yes, 0 = no)
Year 7	Time period 7			(1 = yes, 0 = no)
Year 8	Time period 8			(1 = yes, 0 = no)
Year 9	Time period 9			(1 = yes, 0 = no)
Year 10	Time period 10			(1 = yes, 0 = no)
Year 11	Time period 11			(1 = yes, 0 = no)
Year 12	Time period 12			(1 = yes, 0 = no)
Year 13	Time period 13			(1 = yes, 0 = no)
Reference group = year 1				
URM	Ethnicities: Black, Hispanic, Native America, or Filipino		*	(1 = yes, 0 = no)
Foreign	Non resident visa holder		*	(1 = yes, 0 = no)
Reference group = Other Domestic				
Female	Gender		*	(1 = yes, 0 = no)
Age at entry	Age at matriculation		*	Continuous
Previous graduate degree	Held graduate degree at matriculation		*	(1 = yes, 0 = no)
BA Inst: Tier 1	BA institution: US N & WR Top 25 Doctoral or Top 10 liberal arts		*	(1 = yes, 0 = no)
Financial support	Inflation adjusted financial support in \$1000s	*	*	Continuous
Loans	Inflation adjusted loan amount in \$1000s	*	*	Continuous
Cohort size	Number of new registrants in dept. during the student's first year		*	Continuous
Time to degree	Department's median time to degree			Continuous
Instruction quality	Graduates' assessment of instruction quality			
% Female faculty	Percent of faculty: Female			Continuous
% Female	Percent of enrolment: Female	*		Continuous
% Foreign students	Percent of enrolment: Foreign	*		Continuous
% Underrepresented minority	Percent of enrolment: Underrepresented minority	*		Continuous
Student faculty ratio	# of students/# of faculty	*		Continuous
Selectivity	# of Admits/# of applicants			Continuous
Humanities	Field: Humanities			(1 = yes, 0 = no)
Health sciences professional	Field: Health sciences professional			(1 = yes, 0 = no)
Professional schools	Field: Professional schools			(1 = yes, 0 = no)
Social sciences	Academic field: social sciences			(1 = yes, 0 = no)
Reference group = Life sciences				
GRE verbal	Student's GRE verbal percentile score		*	Continuous
Dept. GRE verbal	Department's mean GRE verbal percentile score			Continuous
GRE quantitative	Student's GRE quantitative percentile score		*	Continuous
Dept. GRE quantitative	Department's mean GRE quantitative percentile score			Continuous

time period. Nonetheless, the hazard rate indicates that doctoral students at this institution are most likely to depart in their first 3 years and 9 years after entry.

Figure 3 contains the survival-function graph by broad academic field. The survival functions demonstrate differences in student departure across the academic fields. The survival curves for the humanities, health science professional, professional schools, and the social sciences appear to have similar, linear slopes. In contrast, the graphs for the life and physical sciences have a downward curve that

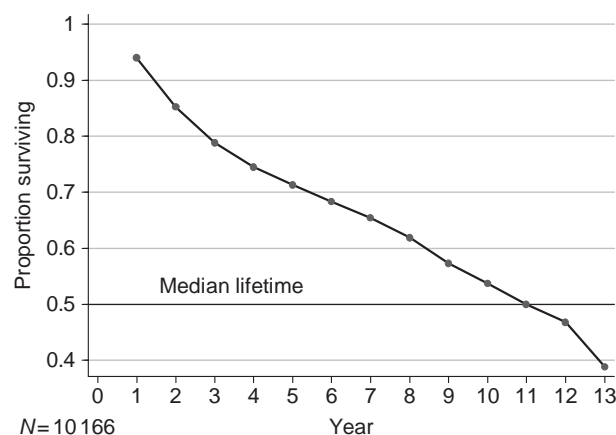
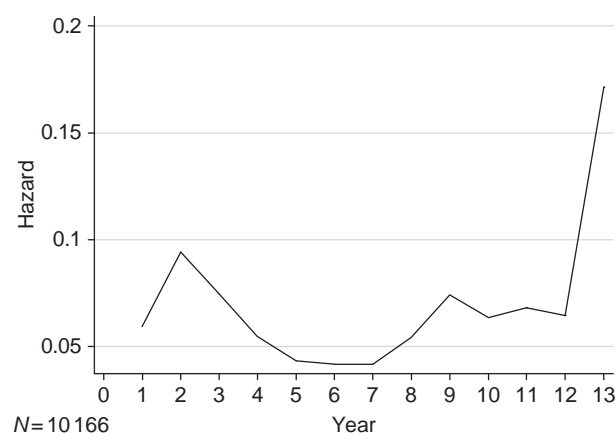
levels off after year 6. (The drop in attrition during year 9 for the physical sciences is based on a sample of just 46 and is within the standard error compared to year 8.)

Multivariate Results

A multilevel model was created to explore the impact of various factors on doctoral students' attrition at different time periods. The first model includes the baseline hazard of leaving the doctoral program at different time periods

Table 3 Life table of doctoral attrition, 1992–2004 entering cohorts

Year	Enrolled at beginning of year	Attrition	Censored	Survival function	Hazard function
0	10 130	–	–	–	–
1	10 130	573	679	0.94	0.06
2	8 878	835	719	0.85	0.09
3	7 324	548	797	0.79	0.07
4	5 979	326	923	0.74	0.05
5	4 730	205	1 175	0.71	0.04
6	3 350	140	1 113	0.68	0.04
7	2 097	88	833	0.65	0.04
8	1 176	64	464	0.62	0.05
9	648	48	254	0.57	0.07
10	346	22	148	0.54	0.06
11	176	12	71	0.50	0.07
12	93	6	52	0.47	0.06
13	35	6	29	0.39	0.17

**Figure 1** Survival curve of doctoral attrition, 1992–2004 entering cohorts.**Figure 2** Hazard rate of doctoral attrition, 1992–2004 entering cohorts.

compared to year 1 of doctoral study. Subsequently, additional models added several student-level characteristics presented at the time of admission to the doctoral program, departmental-level characteristics, broad field-level characteristics, and interaction effects to the baseline hazard model. The study results are contained in **Table 4**.

The baseline hazard of leaving doctoral study is highest at year 13 followed by years 2 and 3 compared to the reference group, which is year 1 of doctoral study. The odds of attrition drop down by year 4 and increase again after year 8. Students at year 13 are 3.4 times more likely to drop out than the reference group, but as mentioned earlier, only 35 students were observed during this time period. Hazards of attrition at years 2 and 3 remain significant after accounting for the effects of student, department, and field characteristics and interaction effects to the baseline model. In all levels of the model, students are less likely to depart at years 6 and 7 as compared with year 1.

The second model includes group-centered mean variables for individuals based on their departmental affiliation. Taking group-centered mean values by department accounts for the clustering effect of the variables by department. Foreign students and students with a previous graduate degree are predicted to have a lower probability of attrition across all models at $p < 0.001$. There is no significant difference in the probability of attrition between students who received baccalaureate degrees from top-tier schools compared to other less-prestigious schools. The probability of attrition for domestic under-represented minority students (URM) is higher than other students holding all else constant. In the second model that contains the students' basic demographic and academic characteristics, the older the student was at the beginning of doctoral study, the more likely he/she was to

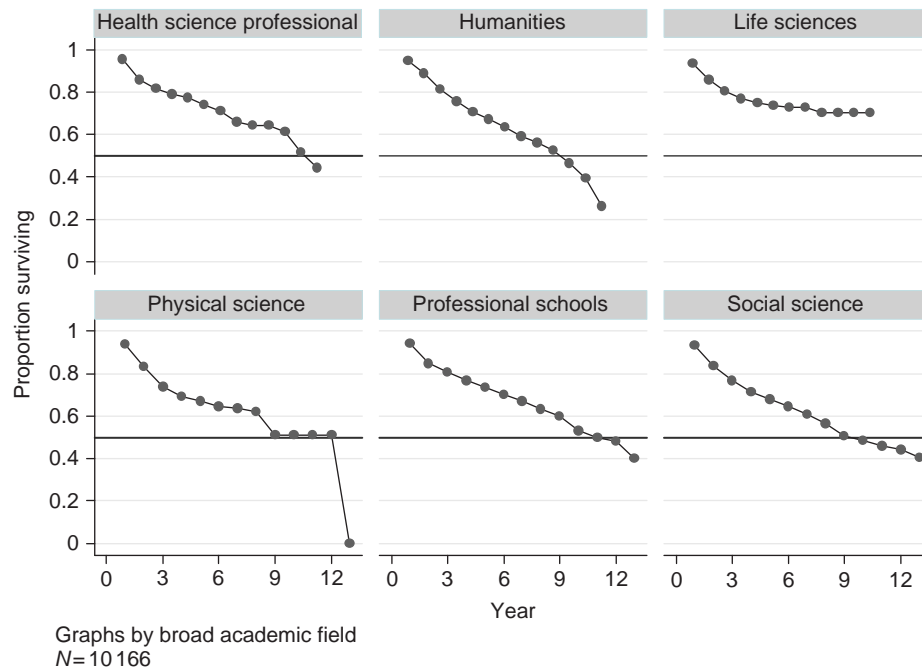


Figure 3 Survival curves for doctoral students by broad academic field, 1992–2004 entering cohorts.

depart before completing the degree. However, after accounting for the amount of student support received and the departmental context, the direction of the effect changed, and the older the student was at the start of doctoral study, the less likely he/she was to leave before completing the degree.

The third model includes the baseline, department- and student-level variables. Not surprisingly, the more total financial support the students received, the less likely they were to drop out. Interestingly, the total amount of loans received had the same effect, perhaps a reflection of the students' commitment to, and investment in, their education. Students who are enrolled in departments with higher student-rated levels of instructional quality are less likely to drop out of their doctoral studies. However, this effect becomes significant only after controlling for the broad academic fields, indicating that instructional quality varies between academic fields.

The next model includes dichotomous variables indicating the broad academic field in addition to the other independent variables discussed above. Using the life sciences as the reference group, students in all other fields, except for the professional health sciences, are predicted to be more likely to drop out, holding all else constant. Students in the humanities and social sciences are approximately 65% more likely to drop out of their doctoral program compared to life-science students. On the contrary, students from the professional health sciences field are 25% less likely to depart from their program, but the difference is not statistically significant.

The full model includes all the variables from the previous models and interaction effects. The results indicate that the impact of financial support varies across time periods. The interaction effects indicate that financial support plays a greater role in doctoral student departure during the third, fourth, and seventh year of enrolment. The other significant interaction effect observed was between females and the percentage of female students in the program. As the percentage of female students increases in a department, females become less likely to withdraw from their doctoral study.

Across all models, inclusion of the random effect was significant at $p < 0.001$. Consequently, we can conclude that doctoral-student attrition varies between academic departments.

In addition, we fitted a subset model that included the full model plus individual and mean department GRE percentile scores (for departments that required the GRE for admission). The results indicate that both individual and mean department GRE quantitative scores have no significant effect on doctoral-student attrition. However, individual GRE verbal percentile scores are positively related to doctoral-student attrition, indicating that students with higher GRE verbal percentile scores were more likely to drop out before completing the degree.

Discussion and Conclusion

This study applied multilevel hazard analysis to examine doctoral-student attrition at a university in the western

Table 4 Random effect predictors of doctoral student attrition

<i>Variable</i>	<i>Model 1</i>	<i>Model 2</i>	<i>Model 3</i>	<i>Model 4</i>	<i>Model 5</i>	<i>Model 6</i>
Year 2	0.556*** (0.056)	0.559*** (0.057)	0.889*** (0.059)	0.888*** (0.059)	0.842*** (0.059)	0.915*** (0.067)
Year 3	0.312*** (0.062)	0.320*** (0.062)	0.671*** (0.065)	0.670*** (0.065)	0.498*** (0.072)	0.621*** (0.081)
Year 4	–0.0180 (0.072)	–0.00849 (0.072)	0.315*** (0.074)	0.313*** (0.074)	0.0181 (0.090)	0.0455 (0.11)
Year 5	–0.254** (0.084)	–0.249** (0.084)	–0.129 (0.086)	–0.130 (0.086)	–0.134 (0.085)	–0.106 (0.098)
Year 6	–0.290** (0.097)	–0.294** (0.097)	–0.412*** (0.099)	–0.417*** (0.099)	–0.395*** (0.099)	–0.439*** (0.12)
Year 7	–0.294* (0.12)	–0.308** (0.12)	–0.631*** (0.12)	–0.634*** (0.12)	–0.250 (0.15)	–0.318 (0.18)
Year 8	–0.0325 (0.14)	–0.0597 (0.14)	–0.468*** (0.14)	–0.472*** (0.14)	–0.418** (0.14)	–0.446** (0.16)
Year 9	0.293 (0.16)	0.256 (0.16)	–0.268 (0.16)	–0.270 (0.16)	–0.214 (0.16)	–0.229 (0.18)
Year 10	0.126 (0.23)	0.0752 (0.23)	–0.491* (0.23)	–0.492* (0.23)	–0.430 (0.23)	–0.483 (0.26)
Year 11	0.205 (0.30)	0.152 (0.30)	–0.420 (0.31)	–0.417 (0.31)	–0.349 (0.31)	–0.378 (0.36)
Year 12	0.141 (0.43)	0.0772 (0.43)	–0.550 (0.43)	–0.547 (0.43)	–0.441 (0.43)	–0.403 (0.48)
Year 13	1.231** (0.45)	1.135* (0.45)	0.507 (0.46)	0.520 (0.46)	0.616 (0.46)	1.085* (0.49)
URM		0.106 (0.064)	0.229*** (0.066)	0.228*** (0.066)	0.226*** (0.066)	0.269*** (0.080)
Foreign		–0.333*** (0.055)	–0.427*** (0.057)	–0.467*** (0.057)	–0.454*** (0.057)	–0.405*** (0.072)
Female		0.0534 (0.043)	0.0591 (0.045)	0.0593 (0.045)	0.520*** (0.11)	0.483*** (0.12)
Age at entry		0.00841* (0.0039)	–0.0100* (0.0040)	–0.0100* (0.0040)	–0.0101* (0.0040)	–0.0158** (0.0050)
Previous graduate degree		–0.185*** (0.048)	–0.171*** (0.049)	–0.171*** (0.049)	–0.171*** (0.049)	–0.184** (0.058)
BA Inst: Tier 1		0.0143 (0.049)	0.0106 (0.051)	0.0104 (0.051)	0.0128 (0.051)	–0.0395 (0.058)
Loans			–0.0251*** (0.0049)	–0.0250*** (0.0049)	–0.0233*** (0.0049)	–0.0320*** (0.0057)
Cohort size			0.00405 (0.0034)	0.00528 (0.0033)	0.00505 (0.0033)	0.00589 (0.0036)
Financial support			–0.0777*** (0.0020)	–0.0778*** (0.0020)	–0.0677*** (0.0025)	–0.0764*** (0.0029)
Time to degree			0.109* (0.048)	0.00403 (0.053)	–0.00634 (0.053)	0.00491 (0.072)
Instruction quality			–0.501 (0.27)	–0.659** (0.24)	–0.681** (0.24)	–0.932*** (0.28)
% Female faculty			–0.0764 (0.39)	–0.0530 (0.34)	–0.0670 (0.34)	0.166 (0.46)
% Female students			–0.279 (0.25)	–0.0502 (0.24)	–0.0517 (0.24)	–0.368 (0.34)
% Foreign			0.0348 (0.25)	–0.164 (0.24)	–0.180 (0.24)	–0.112 (0.32)
% URM			0.356 (0.45)	0.242 (0.42)	0.242 (0.43)	–0.883 (0.65)
Student faculty ratio			0.0163 (0.24)	–0.0180 (0.23)	–0.0163 (0.23)	–0.0183 (0.25)
Selectivity			0.441 (0.35)	0.803* (0.33)	0.781* (0.33)	1.382** (0.42)
Humanities				0.512** (0.18)	0.511** (0.18)	0.483* (0.22)

Continued

Table 4 Continued

<i>Variable</i>	<i>Model 1</i>	<i>Model 2</i>	<i>Model 3</i>	<i>Model 4</i>	<i>Model 5</i>	<i>Model 6</i>
Health science professional				−0.310 (0.21)	−0.314 (0.21)	−0.473* (0.24)
Physical science				0.354 (0.18)	0.335 (0.18)	0.236 (0.19)
Professional schools				0.217 (0.15)	0.203 (0.18)	0.00395 (0.19)
Social sciences				0.497** (0.17)	0.490** (0.17)	0.429* (0.20)
Year 3* financial support					−0.0285*** (0.0050)	−0.0255*** (0.0057)
Tear 4* financial support					−0.0433*** (0.0063)	−0.0446*** (0.0075)
Year 7* financial support					0.0323** (0.010)	0.0344** (0.013)
Female*% female students					−1.019*** (0.23)	−0.847** (0.26)
GRE verbal						0.00311** (0.0010)
GRE quantitative						−0.000855 (0.0015)
Dept. GRE verbal						0.00347 (0.0073)
Dept. GRE quantitative						−0.00521 (0.0084)
Constant	−2.847*** (0.061)	−2.862*** (0.061)	−2.481** (0.92)	−1.635 (0.88)	−1.395 (0.88)	−0.602 (1.60)
Variance of random effect	−2.335*** (0.26)	−2.337*** (0.26)	−2.289*** (0.27)	−2.915*** (0.34)	−2.921*** (0.34)	−3.032*** (0.41)
SD of random effect	0.311	0.311	0.318	0.233	0.232	0.220
Residual ICC	0.029	0.029	0.030	0.016	0.016	0.014
N	44962	44962	44962	44962	44962	35518
N Departments	59	59	59	59	59	47
LL	−10490	−10449	−9618	−9608	−9555	−7329

Standard errors in parentheses

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

United States. The results indicate that the point at which doctoral students depart varies. While it is commonly understood that the doctoral process in all fields involves stages of degree progress (coursework, exams, candidacy, and dissertation), the timing of these stages is highly variable. Utilizing academic year as the unit of time instead of stage-of-degree progress adds another level of clarity to the analysis.

Not surprisingly, the study also found that student support is critical to doctoral-student success. One difference from past research is the finding that the sum of student loans seemed to be a predictor of student commitment to degree obtainment, rather than a reflection of the burden of student debt. The finding that the impact of financial support changes over time requires further analyses, though Tinto (1993) believed this to be true as well.

The fact that there is more variability between academic departments than academic fields highlights the need to understand department and doctoral program culture better, including the nature of student–faculty relationships, mentoring, and relationships among students. Additionally,

the results suggest that female students are less likely to drop out before completing the doctorate in departments with higher percentages of female students. This is an important contextual effect that needs to be examined more closely, especially as it relates to fields in which women are underrepresented.

Implications for Policy and Practice

The results contain a number of findings that can be used to inform policy and practice on graduate-student issues. First, most doctoral students at the institution studied are at the highest risk for departure during their second and third years. After declining between years 3 and 7, the risk of drop out subsequently increases again during year 8. Increased focus on retaining students should occur during the second and third years of study when students are making the transition from the coursework to dissertation phase of their doctoral studies and after 8 years of enrolment, when students have completed all but the dissertation.

The results support previous research indicating that adequate student support is critical to student success (Bair and Haworth, 2004). Apart from the effect of time, financial assistance was the most significant factor in the model. Financial support appears to play an even more critical role during the third and fourth year of enrolment. These years represent the time when many students are beginning the dissertation process; consequently, when students fail to receive adequate financial support during the transition from class work to candidacy, they are more likely to drop out.

The present study both supports and contradicts what has been found in previous studies: that women, minority students, and American students drop out at higher rates than men, White students, and international students, respectively (Bowen and Rudenstine, 1992; Lovitts, 2001; Smallwood, 2004). Across all models in this study, foreign students are less likely to depart compared with majority domestic students. Initially, domestic URM students are not predicted to depart more than their majority counterparts; however, after taking into account the departmental context, the model predicts that URM students will be more likely to leave before completing their doctoral studies. Finally, gender was not a significant predictor of attrition; however, as mentioned below, gender does come into play with respect to the structural characteristics of the department.

The results support the previous research indicating that the structural characteristics of the department have little relationship to doctoral-student attrition. From the faculty-to-student ratio to the percentage of underrepresented students in the program, the basic observable characteristics appear to be unrelated to drop out. The exception is the finding that female students are less likely to drop out in departments with higher percentages of female students.

The study also suggests that academic credentials of applicants are not the best predictors of student success, confirming previous research (Bair and Haworth, 2004). While having a previous graduate degree is consistently associated with a declining probability of student departure, GRE quantitative scores at the individual and department levels appear to be unrelated to doctoral-student attrition. However, individual GRE verbal scores are modeled to be positively related to doctoral-student attrition, indicating that students with higher verbal scores are more likely to not complete their studies. This finding raises questions about the validity of GRE scores as a predictor of graduate-student success.

Bibliography

Allison, P. D. (1984). *Event History Analysis: Regression for Longitudinal Event Data* (Sage University Papers: Quantitative Applications in the Social Sciences, 07-046). Newbury Park, CA: Sage.

- Austin, A. E. (2002). Preparing the next generation of faculty: Graduate school as socialization to the academic career. *Journal of Higher Education* **73**(1), 94–122.
- Bair and Haworth (2004). Doctoral student attrition and persistence: A meta-synthesis of research. In Smart, J. C. (ed.) *Higher Education: Handbook of Theory and Research*, vol. 19, pp 481–534. Netherlands: Kluwer Academic Publishers.
- Baird, L. L. (1991). The melancholy of anatomy: The personal and professional development of graduate and professional school students. In Smart, J. C. (ed.) *Higher Education: Handbook of Theory and Research*, vol. 6, pp 361–392. New York: Agathon Press.
- Baird, L. L. (1993). Using research and theoretical models of graduate student progress. In Baird, L. L. (ed.) *New Directions for Institutional Research: No. 80. Increasing Graduate Student Retention and Degree Attainment*, pp 3–12. San Francisco, CA: Jossey-Bass Publishers.
- Baird, L. L. (1997). Completing the dissertation: Theory, research, and practice. In Goodchild, L. F., Green, K. E., Katz, E. L., and Kluever, R. C. (eds.) *New Directions for Higher Education: No. 99. Rethinking the Dissertation Process: Tackling Personal and Institutional Obstacles*, pp 99–105. San Francisco, CA: Jossey-Bass.
- Barber, J., Murphy, S., Axinn, W., and Maples, J. (2000). Discrete-time multilevel hazard analysis. *Sociological Methodology* **30**, 201–235.
- Benkin (1984). *Where Have All the Doctoral Students Gone?: A Study of Doctoral Student Attrition at UCLA*. Doctoral dissertation, University of California: Los Angeles.
- Bowen, W. G. and Rudenstine, N. L. (1992). *In Pursuit of the PhD*. Princeton, NJ: Princeton University Press.
- DesJardins, S. L. (2003). Event history methods. In Smart, J. C. (ed.) *Higher Education: Handbook of Theory and Research*, vol. 18, pp 421–472. London: Kluwer.
- Ferrer de Valero, Y. (2001). Departmental factors affecting time-to-degree and completion rates of doctoral students at one land-grant research institution. *Journal of Higher Education* **17**(3), 341–367.
- Golde, C. M. (2000). Should I stay or should I go? Student descriptions of the doctoral attrition process. *Review of Higher Education* **23**(2), 199–227.
- Golde, C. M. and Dore, T. M. (2004). The survey of doctoral education and career preparation: The importance of disciplinary contexts. In Wulff, D. H., Austin, A. E., et al. (eds.) *Paths to the Professoriate: Strategies for Enriching the Preparation of Future Faculty*, pp 18–44. San Francisco, CA: Jossey-Bass.
- Haynes, V. D. (2004). The doctorate is out for more grad students. *The Chicago Tribune* p 8.
- Hodgson, C. S. and Simoni, J. M. (1995). Graduate student academic and psychological functioning. *Journal of College Student Development* **36**(3), 244–253.
- Katz, E. J. (1997). Key players in the dissertation process. In Goodchild, L. F., Green, K. E., Katz, E. L., and Kluever, R. C. (eds.) *New Directions for Higher Education: No. 99. Rethinking the Dissertation Process: Tackling Personal and Institutional Obstacles*, pp 5–16. San Francisco, CA: Jossey-Bass.
- Kluever, R. C. (1997). Students' attitudes toward the responsibilities and barriers in doctoral study. In Goodchild, L. F., Green, K. E., Katz, E. L., and Kluever, R. C. (eds.) *New Directions for Higher Education: No. 99. Rethinking the Dissertation Process: Tackling Personal and Institutional Obstacles*, pp 47–56. San Francisco, CA: Jossey-Bass.
- Lovitts, B. E. (2001). *Leaving the Ivory Tower: The Causes and Consequences of Departure from Doctoral Study*. Lanham, MD: Rowman & Littlefield.
- Nerad, M. and Cerny, J. (1993). From facts to action: Expanding the graduate division's educational role. In Baird, L. L. (ed.) *New Directions for Institutional Research: No. 80. Increasing Graduate Student Retention and Degree Attainment*, pp 27–40. San Francisco, CA: Jossey-Bass.
- Nerad, M. and Miller, D. S. (1997). The institution cares: Berkeley's efforts to support dissertation writing in the humanities and social sciences. In Goodchild, L. F., Green, K. E., Katz, E. L., and Kluever, R. C. (eds.) *New Directions for Higher Education: No. 99. Rethinking the Dissertation Process: Tackling Personal and Institutional Obstacles*, pp 75–89. San Francisco, CA: Jossey-Bass.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edn. Thousand Oaks, CA: Sage.

- Singer, J. D. and Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study the duration and the timing of events. *Journal of Educational Statistics* **18**(2), 155–195.
- Smallwood, S. (2004). Doctor dropout: High attrition from PhD. programs is sucking away time, talent, and money and breaking some hearts, too. *Chronicle of Higher Education*, p A10.
- Smart, J. C., Feldman, K. A., and Ethington, C. A. (2000). *Academic Disciplines: Holland's Theory and the Study of College Students and Faculty*. Nashville: Vanderbilt University Press.
- Tinto, V. (1993). *Leaving College: Rethinking the Causes and Cures of Student Attrition*, 2nd edn. Chicago, IL: University of Chicago Press.
- Turner, J. L., Miller, M., and Mitchell-Kernan, C. (2002). Disciplinary cultures and graduate education. *Emergences* **12**(1), 47–70.
- Weidman, J. C., Twale, D. J., and Stein, E. L. (2001). Socialization of graduate and professional students in higher education: A perilous passage? *ASHE-ERIC Higher Education Report*, Vol. 28, no 3. San Francisco, CA: Jossey Bass.
- Wulff, D. H. and Austin, A. E. (2004). Future directions: Strategies to enhance paths to the professoriate. In Wulff, D. H. and Austin, A. E. and Associates, *Paths to the Professoriate: Strategies for Enriching the Preparation of Future Faculty*, pp 267–292. San Francisco, CA: Jossey-Bass.
- Yamaguchi, K. (1991). *Event History Analysis*. Newbury Park, CA: Sage Publications.

Further Reading

- Gravois, J. (2007). New data offer a rosier picture of Ph.D.-completion rates. *Chronicle of Higher Education*. <http://www.chronicle.com/daily/2007/07/2007071703n.htm>. (accessed July 2007).
- Ishitani, T. T. (2008). How do transfers survive after "transfer shock"? A longitudinal study of transfer student departure at a four-year institution. *Research in Higher Education, Online First* February 14.

Male Underachievement in Education Across the Globe: A Shift in Paradigm for Gender Disparities Regarding Academic Achievement

J F L Jackson, University of Wisconsin–Madison, Madison, WI, USA

J L Moore III, The Ohio State University, Columbus, OH, USA

R A Leon, University of Wisconsin–Madison, Madison, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Hegemonic masculinity – Refers to the belief in the existence of a culturally normative ideal of male behavior.

Summarizing academic research relevant to the decline of male academic performance in education is an important step toward understanding why males appear disengaged, uninterested, and failing at high rates (Sutherland, 1999). To this end, it is critical that education scholars and practitioners gain better understanding of current research and use it to closely analyze and identify explanations for the male educational crisis. Without this understanding, how could we realistically address this problem? Arguably, more so than any other time in history, human capital is one of society's most precious commodities. Countries, such as the United States, Canada, England, and France, can no longer stand on the sidelines watching certain demographic groups underachieve or low achieve in their educational systems.

Today's knowledge-based economy mandates continuous investment in education to ensure that countries are able to compete in the new global market. Increasingly, the government and business leadership of countries around the world realize that they must improve their educational systems and address the issue of underachievement or low achievement among certain demographic groups, including males. These countries also recognize that they must invest more resources in science, technology, engineering, and mathematics (STEM). Nations, such as China and India, once seen as poor, antiquated, and underdeveloped, are now viewed as major players in the twenty-first century.

Over the last decade or so, these two countries have invested major financial resources in improving their educational system, particularly in the STEM areas. In this short span of time, the two countries have reaped major benefits from these investments. More specifically, they have allowed China and India to transform themselves from struggling, poor nations to global economic engines, where other countries want to partner with them.

In these countries, men and boys are front and center and are the primary reasons for the transformation. As in most parts of the world, there are gender disparities; however, the biggest difference is that boys are leading the way educationally, which is the complete opposite for other countries, such as the United States and England. Instead, in the United States and England, girls are outperforming boys (Jha and Kelleher, 2006). The former pattern, not the latter, is evident for other countries that have "achieved universal access and have high participation rates for both girls and boys, at least at the primary stage of schools (including a number of Commonwealth countries in the Caribbean, Europe, East Asia, and the Pacific, and some sub-Saharan Africa and South Asia)" (Jha and Kelleher, 2006: 4).

To ensure that countries have a place at the table in the global market, it is important that their social scientists, educators, and policymakers take note of the educational trends for males and equally work to understand them better. Similar to Jha and Kelleher (2006) in *Boys Underachievement in Education: An Exploration in Select Commonwealth Country*, we framed this article on participation and performance. Based on our review, these two points were the most salient issues for males as it relates to their underachievement or low achievement.

Establishing the Problem

Scholarly journal entries (e.g., Jackson and Moore, 2006, 2008) and popular media excerpts (e.g., Patterson, 2006) present a compelling argument regarding the academic decline of males in education. These arguments are largely based on national standardized test scores, high dropout rates, disciplinary rates, disproportionate numbers in special education, and falling college enrolment (Weaver-Hightower, 2003). In order to interpret the male underachievement concept, a synopsis of the research literature depicting the shift in concern toward the educational experiences of males across the world is undertaken. Of particular interest are the perceptions of this phenomenon as a crisis directly related to the performance of the girls and women. By exploring this specific contrast, we are able to find inconsistencies across the

literature when attempting to characterize and interpret the nature of what is defined as male underachievement.

The Boy Concern

The shift toward the study of the educational performance of males emerged as a global concern only in the mid-1990s, when a focus on the schooling experiences of boys took the central stage of educational debate (Weaver-Hightower, 2003). Preoccupation, within the United States, was echoed by concerns in countries, such as England and Australia, where this issue acquired the label of a national crisis. In particular, national anxiety in England occurred because girls surpassed boys at A level for the first time on the General Certificate of Secondary Education (GCSE) examinations in 2000 (Myhill, 2002).

The practice-based focus on boys was accompanied by an empirical research component. Weaver-Hightower (2003) defines this transition toward the study of males as the boy turn. He acknowledged that shifting attention from girls to boys could be seen as unwelcomed, but it also provides a space for policy and research to face the current academic struggles of this population. Concurrently, despite the relatively nascent nature of research on males, most of the current research in this area is dominated by the study of the experiences of black/African-American males. In certain London districts, the black expulsion rate is 15 times that of whites. Unfortunately, less than 25% of British secondary schools had a specific policy to address low black student educational achievement (*Journal of Blacks in Higher Education*, 2002).

Studying the experiences of the black/African-American male student population is also occurring in the United States. Cross and Slater (2000) assert:

It is true that men of *all* races are increasingly earning smaller share of all degrees awarded in the United States. In large part this is the product of a society that has rather recently opened up employment and professional opportunities to women generally. But the declining performance of men in higher education in relative terms took hold much earlier for [B]lack men than it did for [W]hite men. (p. 82)

Further supporting and perhaps explaining the reasons for a particular interest in the underachievement of black/African-American males, it is necessary to acknowledge that the black/African-American male college completion rate has dropped from 35% to 31% since 1995 (Cross and Slater, 2001).

This decline coincides with the time period (i.e., mid-1990s) when educational research started focusing on the experiences of males, and it is not difficult to suspect that some relationship could be drawn between the low academic performance of black/African-American males and

their current struggle in other social domains. For example, in 1999 it was reported that 25% more black/African-American men were in prison than enrolled in higher education (*The Journal of Blacks in Higher Education*, 2003). This occurrence could be an indication of the implications of low academic performance and its broader repercussion in a world that continuously places a higher value on postsecondary academic preparation. Data on the experience of black/African-American males in education worldwide are troublesome. It seems as if black/African-American males are edified by male underachievement.

Problematising the Label of Male Underachievement

It must be noted that there are challenges to the current perception and applicability of the label male underachievement. It is highly suggested that current definitions be reviewed, and in turn, alternative interpretations of the phenomenon considered. It is believed that focusing on the underachievement of boys' obscures differences within genders by social class and ethnicity (Tinklin, 2003). Likewise, Smith (2003) found that the term is largely used by politicians, journalists, academics, and practitioners to describe relatively poor academic performance. Nonetheless, a review of scientific and popular literature does not yield a consensus on its definition and measurement. He further states that "gender accounted for less than 1% of the variance in examination outcome," which has deep implications for "commentators who place gender, and the underachievement of boys in particular, at the heart of the standards debate and should provide a cautionary note for all those who perpetuate the binary notion of boys versus girls" (Smith, 2003: 585).

This statement, in particular, voices a powerful criticism of the current notion of underachievement. A specific focus on understanding what the underachievement is in relation to and the use of more appropriate terms such as low achievement, or differential achievement, to describe the male performance, ought to be considered. It appears evident that redefining how the gender gap is currently conceived and studied is imperative to move into a deeper exploration of this phenomenon, particularly avoiding a shortcoming where previous research may have been attempting to explain differences whose nature was incompletely understood. This compels researchers to further explore the potential socioeconomic, classroom, and individual determinants of these gender gaps (Gorard *et al.*, 2001), while also moving away from a surge of media interest positing a number of interpretations as a cause of the gender gap (Renold, 2001).

For that reason, in the following section, we offer a review of important evidence across five themes within

the literature that could offer valuable clues to understand the various aspects related to the educational performance of males and reasons for their current academic predicament.

Identified Reasons of Male Underachievement

Scholarly research examining topics, such as social identity, parent involvement, peer influence, classroom behavior, lack of engagement, and school curriculum, among other themes are scrutinized as potential underlying reasons contributing to male underachievement. Our exploration of the research literature – focused on academic performance of the male population – opens new avenues for understanding this challenge.

Family Composition

A body of academic research explores family composition and its impact on behavior problems of inner-city boys. Florsheim *et al.* (1998) found that boys in single-mother families were at greater risk of developing problems than boys in two-parent families. However, the risks associated with single motherhood were offset by a structured family environment, an effective disciplinary strategy that allowed for some degree of adolescent autonomy, and the positive involvement of a male family member. In addition, it is necessary to note that they also found that higher economic status among these single-mother families did not deter negative behaviors. This finding contrasts with previous findings that suggest underachievers are mainly working-class boys.

These findings are relevant in relation to male underachievement, because it is imperative to understand that there are multiple family risk factors that play a role in the development of such problems. Stated differently, the accumulation of risks elevates the likelihood that behavior disorders will occur. This understanding offers an approach to better interpret and draw connections among the current reasons for male underachievement.

Identity Development

A student's identity development is influenced by the environment in which they interact. Further, current conceptions of masculinity influence attitudes across various individual and social contexts. In these contexts, students are continuously attempting to fit in with common stereotypes that guide their behaviors within the school setting. Dumais (2002) proposes that researchers should consider both one's resources (capital) and the orientation one has toward using those resources (habitus) to truly understand the academic potential of students. This differentiation is

important considering that male students continuously attempt to downplay, or disguise these attributes, fearing that their peers will label them as sissies, in turn, preventing these boys from transforming their cultural knowledge skills into success in the classroom. Stereotypes are structured upon concepts of masculinity that dominate the educational experiences of boys. Therefore, understanding the context of these interactions could explain why male identity is guided through a path that is in opposition to academic success, dedication, and hard work.

Hegemonic masculinity

The concept of social identity is introduced as a predominant factor with potential to explain the roots of male underachievement. More recent concepts of masculinities and femininities are often constructed in opposition to each other (Renold, 2001). Renold's (2001) ethnographic study explores the interactions of students in a rural, white middle-class town in England, where for a boy being academically oriented is often devalued and denigrated because it is equated with femininity which conflicts with conventional forms of hegemonic masculinity. From these notions, and framing male identity development within the context of hegemonic masculinity, it is easy to observe aggressive behavior as a key trait for male students to protect and reassure their male identity. An increasing number of high-achieving boys are "strategically dissociating from the activities of their nonhegemonic peers and investing heavily in dominant masculine practice such as fighting, rule-transcending and football" (Renold, 2001: 377). Foney and Cunningham (2002) argue that males' status may be increased when they express violence that encourages boys to recreate these types of behavior, while females are the recipients of rejection at the expression of their aggression.

Racialized perspectives of masculinity

Foney and Cunningham (2002) also identified race as a significant factor affecting the male academic performance. In particular, they explored the connection between concepts of hegemonic masculinity and the experience of black/African-American students. Relevant to student behavior, they note as follows: "Although the 'daring' behaviors may not ordinarily be labeled 'aggressive behavior' in majority cultures, in Black/African American males, these daring behaviors may be viewed as antisocial and aggressive" (Foney and Cunningham, 2002: 145). Therefore, their interpretation, among other cultures, produces a negative impact that triggers attitudes within this population such as the development of maladaptive coping mechanisms which are linked to behavior that affect academic performance.

Wright *et al.* (1998) focused on black Caribbean males and recognized that current perceptions of their masculinity are constructed and influenced by other male peers and white teachers. However, they emphasize that these

expressions of black male masculinity triggering exclusion from the educational system should not be interpreted as an inability to attain white concepts of masculinities. These particular findings are in direct opposition to what Horvat and Lewis (2003) assert in their analysis of factors that inhibit academic excellence. Horvat and Lewis rely upon Fordham and Ogbu's (1986) work, *Black Students' School Success: Coping with the Burden of Acting White*, and conclude that black/African-American students continue to underperform in school because of their cultural opposition to acting white, particularly believing that they do not receive comparable returns for their hard work in school relative to whites.

Classroom Behavior

From acting cool to acting dumb

Linked to the process of identity development, student's behaviors and practices in classroom settings are important factors to consider when exploring academic achievement. Tinklin (2003) notes that key factors that help to explain gender differences in attainment is that girls take school more seriously than boys. This finding is related to previous research that suggests: "[w]hile academic success was valued by girls, doing too well at school could disadvantage some boys among their peers" (p. 321). As a result, poor classroom behavior could be seen as a way of compensating for losing popularity or being at risk of negative labels associated with strong academic performance. This opens doors for debate regarding the correlation of classroom behavior and the decline of male academic performance.

Moreover, Martino's (2000) research on adolescent boys in a Catholic coeducational school in Australia affirms that boys develop versions of masculinity for themselves through specific social practices such as mucking around in class, giving crap, and acting cool. In turn, these practices themselves establish norms for fashioning a desirable heterosexual cool masculinity reinforced by a system of verbal abuse and put-downs which establishes a hierarchy of masculinities (Frank, 1987). Martino (1999) also found that the cool boys act dumb to establish a hegemonic form of masculinity through which they can demonstrate their opposition to the values embodied in the aims of formal education. Osborne (1999) further elaborates and reflects upon these notions of being cool as opposition to academic achievement from a psychological aspect. He refers to the adoption of a cool pose in opposition to the stigmas attached to good academic performance as central matter that explains black/African-American male disengagement with school.

Student engagement

Educators play an important role in a classroom, impacting on students' learning and influencing their attitudes

toward achievement (Moore, 2006; Moore *et al.*, 2008). One would expect that teachers treat their students equitably and their teaching methods are structured in such a way that will provide equal treatment to both male and female students. However, referring to the implicit gender difference illustrated within the literature, Younger *et al.* (1999) state:

Research in this broader context suggests that most teachers believe that they give equal treatment to girls and to boys, particularly in support of their learning, but focus group interviews with students and classroom observation suggest that this is rarely achieved; in most schools, boys appear to dominate certain classroom interactions, while girls participate more in teacher-student interactions which support learning. (p. 325)

Therefore, student engagement must be analyzed as a central theme when referring to academic achievement. However, making adequate differentiations among types of positive and negative engagement is necessary to comprehend current evidence that highlights difference patterns and levels of engagement between genders. Marks' (2000) work states, "Female gender (at all grade levels) contributes positively to engagement, although the effect is attenuated in the orientation toward school and forms of social support models" (p. 171). In contrast, males are associated with patterns of engagement dominated by disruption. This results in animosity between students and instructors, and as a consequence, teachers' attention toward boys is dominated by a negative connotation, and they start to perceive girls as their ideal students. That is, girls embody characteristics such as confidence, self-learning, and articulation and are more inclined to take initiative in promoting their own learning (Younger *et al.*, 1999).

Educator's Perceptions

Younger *et al.* (1999) note that the idea of teacher's negative attention toward male students is rooted in teachers' interpretation of boys' behavior. That is, boys are more vocal, boisterous, less advanced for their years, more easily distracted than girls. This in turn, is seen as a direct challenge to their authority and efforts to create a learning environment. Consequently, this translates in to direct frustration and anger toward male students reinforcing teachers' perceptions that male students have other interests above and more important than school work (Younger *et al.*, 1999).

This particular negative perception of male students is also evident when referring to black/African-American males. Teachers interpreted the behavior of black/African-American males within a sinister and intentional context, punishing them as adults, while judging the rule-breaking attitudes of white boys as childish and innocent

acts (Ferguson, 2000; O'Connor, 2001). Along these lines, Hubbard (1999) argues that teacher's low expectations are particularly prevalent for the black/African-American male student population. These beliefs are part of a cycle of disengagement where teacher's low expectations will diminish student involvement that further caused teachers to have even lower expectations regarding the student's ability creating a self-fulfilling prophecy of failure. The reverse must also be considered regarding the negative connotations of student engagement and teacher's expectations. It is worth recognizing that classroom interactions and boys' (negative) participation in classrooms could be framed in a context where positive contribution to student learning is possible.

Myhill (2002) concluded that there are relationships between underachievement and levels of interaction and response in whole-class teaching episodes. Myhill recognized that there are exceptions within each gender, where "Whilst it appears to be true that boys do dominate calling out, high achievers dominate positive learning interactions whilst underachievers dominate the more negative classroom interactions" (p. 347). Therefore, dominance of high achievers, particularly males, offers hope for school reform that will acknowledge the potential of canalizing boys' energy and learning styles in classroom settings. In turn, this presents the opportunity to develop teaching styles that acknowledge learning and behavior differences among each student and contribute to the overall student's social and intellectual achievement.

Curriculum and Class Structure

A concern with the apparent underachievement of boys had led schools to review the potential benefits of single-sex classes and gender-specific curriculum as interventions to improve the performance of students. However, it is essential to mention that in countries such as England, the debate about the merits of single-sex education has focused on girls and less so on boys (Warrington and Younger, 2001). For that reason, considering the current concern for the performance of males, it is imperative to understand the benefits and limitations of class composition and the type of curriculum that is employed. "Curriculum-as-usual boys' classes do nothing to challenge the problematic macho male cultures inherent in schools; indeed, it may be the case that they actually exacerbate them" (Jackson, 2002: 37).

From a curriculum perspective focusing on school literacy and masculinity, Young and Brozo (2001) levied two arguments. One is arguing for the benefit of finding "boy books for boys and girls books for girls," and the other one stating the benefits of exposing all students to multiple kinds of texts, helping boys and girls to analyze the implicit notions of masculinity and femininity embedded in the literature. Despite this discrepancy, these authors

acknowledge that it might be beneficial to develop and start a curriculum that does not pit boys against girls. This notion was based on the recognition of the social construction of gender and its implications on current literacy practices in schools.

This notion also advanced the discussion regarding the benefits of single-sex versus coeducational schools for boys and girls. Findings across many countries appear to echo the criticism of the benefits of single-sex school for boys. Wong *et al.* (2002) examined the gender differences in educational achievements and concluded that, "after controlling for initial ability, that girls benefited from studying in single-sex schools whereas boys benefited from co-educational schools" (p. 827). These assertions raise a question of the potential effectiveness of school reforms that focused on single-sex class environments.

However, it must not be ignored that most studies have shown no advantage for boys of all-male classrooms. In turn, no studies have documented that coeducational schooling is any better for males than single-sex classes (Singh *et al.*, 1998). As a result, understanding the implications beyond class structures will help to better comprehend the learning process of students, stepping away from research paradigms that ignore implicit connections of class composition with other student and teacher characteristics that must be factored to understand student performance.

Conclusion and Proposed Solutions

Currently, we are venturing into a field of research where discrepancies, different criteria, and uncertainty create a conflictive scenario. Consequently, approaching male academic performance as a multilayer issue is problematic, especially when it is not studied from different academic perspectives. Likewise, attempting to understand each of its components as a smaller part of a bigger and interrelated matter raises many eminent concerns. Therefore, understanding this phenomenon of male underachievement is a difficult task, because it evolves around ideas, such as "one of relative rates of improvement for both sexes" (Younger *et al.*, 1999: 326). It is quite likely that this could help to better comprehend why some males have failed in their quest to conquer academic excellence, while others have succeeded.

It is also worth asking if we are currently improperly labeling the experience of all males in education as a crisis, ignoring that for policy and educational research to effectively target the specific needs of identified student populations, we must consider unique aspects of the experience of each group, and variances within its own population. From this literature review, it is undeniable that male students face various challenges that currently

affect their academic performance. However, placing all these male underachievement indicators under one category constitutes a fatal error when attempting to provide solutions to low academic performance.

O'Connor (2001) affirms that current research and structural theories dealing with achievement: "have failed to account for within-group variation in achievement because they are rendered as if class or race unilaterally positions people in the social world. But people are not simply raced or classed. They are raced and classed and gendered" (p. 164). This reasoning provides us with a valuable alternative to the current approach to this problem, where future research must start examining internal group variation within each academic cohort of students, acknowledging that "in explaining achievement outcomes, we have to attend to the fact, for example, that being black in a public school is different from being black in a Catholic school, being black in a low-ability group is different from being black in a high-ability group, and being black in one family is different from being black in another" (O'Connor, 2001: 164).

Considering the current concern about boy's performance and its many implications of hegemonic masculinity in the identity development process of this population, it is imperative for schools to create environments that positively engage, offer learning opportunities, and challenge "the impact and effects of hegemonic masculinities in school" (Martino, 1999: 259). Drawing upon the themes that emerged as potential reasons that trigger current academic decline among the male population within this literature review could be seen as a solid point to redirect current research, where each theme is explored and placed into the context of a specific population (e.g., black/African American) under analysis, attempting to find evidence of each category represented in the experience of male students that live, study, and go through a process of social, cultural, and educational development that complicates current notions of male underachievement as a single-tier problem. Tinklin (2003) suggests that "a more complex definition of the problem of underachievement is needed" (p. 323).

We do not posit that this review of the literature is exhaustive; however, we do believe that it is one of the most comprehensive examinations of male underachievement or low achievement found in the research literature that goes beyond the borders of the United States. In closing, we hope that this article provides the reader with a starting point to further explore the topic of male underachievement or low achievement. That is, we desire to instigate research that will deepen our knowledge on this topic. Likewise, we hope educational practitioners and policymakers will look at this article as a challenge to develop new strategies and improve the current ones to address education outcomes for males across the globe.

Bibliography

- African-American scholar tackles the educational problems of young Black males in British schools (2002). *Journal of Blacks in Higher Education* **35**, 49–50.
- Cross, T. and Slater, R. B. (2000). The alarming decline in the academic performance of African-American men. *Journal of Blacks in Higher Education* **27**, 82–87.
- Cross, T. and Slater, R. B. (2001). The troublesome decline in African-American college student graduation rates. *Journal of Blacks in Higher Education* **33**, 102–109.
- Dumais, S. A. (2002). Cultural capital, gender, and school success: The role of habitus. *Sociology of Education* **75**, 44–68.
- Ferguson, A. A. (2000). *Bad Boys: Public Schools in the Baking of Black Masculinity*. Ann Arbor, MI: University of Michigan Press.
- Florsheim, P., Tolan, P., and Gorman-Smith, D. (1998). Family relationships, parenting practices, the availability of male family members, and the behavior of inner-city boys in single-mother and two-parent families. *Child Development* **69**, 1437–1447.
- Foney, D. M. and Cunningham, M. (2002). Why do good kids do bad things? Considering multiple contexts in the study of antisocial fighting behaviors in African American urban youth. *Journal of Negro Education* **71**, 143–157.
- Fordham, S. and Ogbu, J. (1986). Black students' school success: Coping with the burden of "acting White." *Urban Review* **18**, 176–206.
- Frank, B. (1987). Hegemonic heterosexual masculinity. *Studies in Political Economy* **24**, 159–170.
- Gorard, S., Rees, G., and Salisbury, J. (2001). Investigating the patterns of differential attainment of boys and girls at school. *British Educational Research Journal* **27**, 125–139.
- Horvat, E. M. and Lewis, K. S. (2003). Reassessing the "burden of 'acting White'": The importance of peer groups in managing academic success. *Sociology of Education* **76**, 265–280.
- Hubbard, L. (1999). College aspirations among low-income African American high school students: Gendered strategies for success. *Anthropology and Education Quarterly* **30**, 363–383.
- Jackson, C. (2002). Can single-sex classes in co-educational schools enhance the learning experiences of girls and/or boys? An exploration of pupils' perceptions. *British Educational Research Journal* **28**, 37–48.
- Jackson, J. F. L. and Moore, J. L. III. (2006). African American males in education: Endangered or ignored. *Teachers College Record* **108**, 201–205.
- Jackson, J. F. L. and Moore, J. L. III. (2008). The African American male crisis in education: A popular media infatuation or needed public policy response. *American Behavioral Science* **51**, 847–853.
- Jha, J. and Kelleher, F. (2006). *Boys' Underachievement in Education: An Exploration in Selected Commonwealth Countries*. London: Commonwealth of Learning and the Commonwealth Secretariat.
- Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal* **37**, 153–184.
- Martino, W. (1999). 'Cool boys', 'party animals', 'squids' and 'poofers': Interrogating the dynamics and politics of adolescent masculinities in school. *British Journal of Sociology of Education* **20**, 239–263.
- Martino, W. (2000). Mucking around in class, giving crap, and acting cool: Adolescent boys enacting masculinities at school. *Canadian Journal of Education* **25**, 102–112.
- Moore, J. L. III. (2006). A qualitative investigation of African American males' career trajectory in engineering: Implications for teachers, school counselors, and parents. *Teachers College Record* **108**, 246–266.
- Moore, J. L. III, Henfield, M. S., and Owens, D. (2008). African American males in special education: Their attitudes and perceptions toward high school counselors and school counseling services. *American Behavioral Scientist* **51**, 907–927.
- More Black men are in jail than are enrolled in higher education. (2003). *Journal of Blacks in Higher Education* **41**, 62.
- Myhill, D. (2002). Bad boys and good girls? Patterns of interaction and response in whole class teaching. *British Educational Research Journal* **28**, 339–352.

- O'Connor, C. (2001). Comment: Making sense of the complexity of social identity in relation to achievement: A sociological challenge in the new millennium. *Sociology of Education* **74**, 159–168.
- Osborne, J. W. (1999). Unraveling underachievement among African American boys from an identification with academics perspective. *Journal of Negro Education* **68**, 555–565.
- Patterson, O. (2006). A poverty of the mind. *The New York Times* **March 26**, 4–13.
- Renold, E. (2001). Learning the 'hard' way: Boys, hegemonic masculinity and the negotiation of learner identities in the primary school. *British Journal of Sociology of Education* **22**, 369–385.
- Singh, K., Vaught, C., and Mitchell, E. W. (1998). Single-sex classes and academic achievement in two inner-city schools. *Journal of Negro Education* **67**, 157–167.
- Smith, E. (2003). Understanding underachievement: An investigation into the differential attainment of secondary school pupils. *British Journal of Sociology of Education* **24**, 575–586.
- Sutherland, M. B. (1999). Gender equity in success at school. *International Review of Education* **45**, 431–443.
- Tinklin, T. (2003). Gender differences and high attainment. *British Educational Research Journal* **29**, 307–325.
- Warrington, M. and Younger, M. (2001). Single-sex classes and equal opportunities for girls and boys: Perspectives through time from a mixed comprehensive school in England. *Oxford Review of Education* **27**, 339–356.
- Weaver-Hightower, M. (2003). The "boy turn" in research on gender and education. *Review of Educational Research* **73**, 471–498.
- Wong, K., Lam, Y. R., and Ho, L. (2002). The effects of schooling on gender differences. *British Educational Research Journal* **28**, 827–843.
- Wright, C., Weekes, D., McGlaughlin, A., and Webb, D. (1998). Masculinised discourses within education and the construction of black male identities amongst African Caribbean youth. *British Journal of Sociology of Education* **19**, 75–87.
- Young, J. P. and Brozo, W. G. (2001). Conversations: Boys will be boys, or will they? Literacy and masculinities. *Reading Research Quarterly* **36**, 316–325.
- Younger, M., Warrington, M., and Williams, J. (1999). The gender gap and classroom interactions: Reality and rhetoric? *British Journal of Sociology of Education* **20**, 325–341.
- Collins, C., Kenway, J., and McLeod, J. (2000). *Factors Influencing the Educational Performance of Males and Females in School and Their Initial Destinations after Leaving School*. Canberra: Commonwealth of Australia Department of Education, Training and Youth Affairs.
- Cousins, L. H. (1999). "Playing between classes": America's troubles with class, race, and gender in a black high school and community. *Anthropology and Education Quarterly* **30**, 294–316.
- Davis, J. (2002). Boys to men: Masculine diversity and schooling. *Paper Presented at the School Leadership Centre of Trinidad and Tobago*.
- Figueroa, M. (2000). Making sense of the male experience: The case of academic underachievement in the English-speaking Caribbean. *IDS Bulletin* **31**, 68–74.
- Gillborn, D. and Gipps, G. (1996). *Recent Research on the Academic Achievements of Ethnic Minority Pupils*. London: HMSO.
- Jackson, D. and Salisbury, J. (1996). Why should secondary schools take working with boys seriously? *Gender and Education* **8**, 103–115.
- Jones, S. and Myhill, D. A. (2004). *Seeing Things Differently: Boys as Underachievers*. *Gender and Education* **16**, 531–546.
- Martino, W. and Berrill, D. (2003). Boys, schooling and masculinities: Interrogating the 'right' way to educate boys. *Educational Review* **55**, 99–117.
- McNamara, F. (2002). Boys: Getting some of it right. *Professional Magazine* **19**, 8–9.
- Salisbury, J. and Jackson, D. (1995). *Challenging Macho Values: Ways of Working with Boys in Secondary Schools*. Philadelphia, PA: Taylor and Francis.
- Salisbury, J., Rees, G., and Gorard, S. (1999). Accounting for the differential attainment of boys and girls: A state of the art review. *School Leadership and Management* **19**, 403–427.
- Tyre, P. (2006). The trouble with boys. *Newsweek* **147**, 44–52.
- Warrington, M. and Younger, M. (2004). We decided to give it a twirl: Single-sex teaching in English comprehensive schools. *Gender and Education* **15**, 339–350.
- Whitmire, R. (2006). Boys and books: Boy trouble. *The New Republic*. <http://www.ssl.tnr.com/p/docsub.mhtml?i=20060123&s=whitmire012306> (accessed October 2009).

Relevant Websites

- <http://www.madeinbritaincampaign.org.uk> – Black Male Forum: Made in Britain Campaign.
- <http://www.boysandschools.com> – Boys and Schools.
- <http://www.boyslearning.com.au> – Boys' Learning.
- <http://www.boysproject.net> – Boys Project.
- <http://www.boystomen.info> – Boys to Men.
- <http://www2.edc.org/GDI/Newsletter/Issue2/fpage.htm> – Gender and Diversities Institute.
- <http://www.singlesexschools.org/home.php> – National Association for Single Sex Public Association.
- <http://www.literacytrust.org.uk/> – National Literacy Trust.
- <http://www.rba.educ.cam.ac.uk> – Raising Boy's Achievement.
- <http://mensbiblio.xyonline.net> – The Men's Bibliography.
- <http://oma.osu.edu/current-students/bell-resource-center> – Todd Anthony Bell National Resource Center on the African American Male.

Further Reading

- Bailey, D. F. and Bradbury-Bailey, M. E. (2007). Promoting achievement for African American males through group work. *Journal for Specialist in Group Work* **32**, 83–96.
- Bailey, D. F. and Moore, J. L. III. (2004). Emotional isolation, depression, and suicide among African American men: Reasons for concern. In Rabin, C. (ed.) *Linking Lives Across Borders: Gender-Sensitive Practice in International Perspective*, pp 186–207. Pacific Grove, CA: Brooks/Cole.
- Bleach, K. (ed.) (1998). *Raising Boys' Achievement in Schools*. Stoke-on-Trent: Trentham Books.
- Callan, P. (2006). *Measuring Up 2006: The National Report Card on Higher Education*. San Jose, CA: The National Center for Public Policy and Higher Education.

Noncognitive Measures for Higher Education Admissions

W E Sedlacek, University of Maryland, College Park, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Noncognitive variables – Noncognitive is used here to refer to variables relating to adjustment, motivation, and student perceptions, rather than the traditional verbal and quantitative (often called cognitive) areas typically measured by standardized tests.

Nontraditional students – Those with cultural, racial, and gender-related experiences different from 18 to 22 year olds with white middle and upper-middle class Euro-centric experiences.

Variance – A statistical term referring to variability on a measurable dimension.

Introduction

We appear to have forgotten why tests were created in the first place. While they were always considered to be useful in evaluating candidates, they were also considered to be more equitable than using prior grades because of the variation in quality among preparatory schools. The College Board has long felt that the Scholastic Aptitude Test (SAT) was limited in what it measured and should not be relied upon as the only tool to judge applicants (Angoff, 1971).

In 1993, the verbal and mathematical reasoning sections of the SAT were lengthened and the multiple-choice Test of Standard Written English was dropped. The name was changed from Scholastic Aptitude Test to Scholastic Assessment Tests, while retaining the SAT initials. Currently, it is just called the SAT-I. In 2003, the College Board announced that an essay would be added and the analogies item type removed as of 2005. Despite various changes and versions over the years, the SAT in essence measures what it did in 1926, verbal and math ability; it is basically still a general intelligence test (Sedlacek, 2003, 2004).

However, we have come to the point where the Big Test has become the focal point in our schools (Lemann, 2000). It has become the standard by which we judge ourselves and others. Many assume that if an individual has high SAT scores, or if a school has high mean SAT scores, the students must be learning something, and the school must be good. To cite that common metaphor, the tail is wagging the dog.

Test results should be useful to educators, student service workers, and administrators, by providing the basis to help students learn better and to analyze their needs. As currently designed, tests do not accomplish these objectives. Many teachers tend to teach to get the highest test scores for their students, student service workers may ignore the tests, and too many administrators are satisfied if the average test scores rise in their school. We need some things from our tests that currently we are not getting. We need tests that are fair to all and provide a good assessment of the developmental and learning needs of students, while being useful in selecting outstanding applicants. Our current tests do not do that.

Keeping Up with Change

The world is much different than it was when the SAT and other tests were developed in the last century. International students, women, people of color, gays, lesbians, and bisexuals, among others, are participating in higher education in more extensive and varied ways (Knapp *et al.*, 2002). Commonly employed tests have not kept up with these changes (Sedlacek, 2004).

We need a fresh approach. It is not good enough to feel constrained by the limitations of our current ways of conceiving of tests. Instead of asking, "How can we make the SAT and other such tests better?" we need to ask, "What kinds of measures will meet our needs now and in the future?" The purpose of this article is to present the underlying logic and research supporting a method that yields such measures. We do not need to ignore our current tests, we need to add some new measures that expand the potential we can derive from assessment.

Noncognitive Variables

Noncognitive is used here to refer to variables relating to adjustment, motivation, and student perceptions, rather than the traditional verbal and quantitative (often called cognitive) areas typically measured by standardized tests (Sedlacek, 1998a, 1998b, 2004). While noncognitive variables are useful for all students, they are particularly critical for nontraditional students, since standardized tests and prior grades may provide only a limited view of their potential. Below is a discussion of the eight variables recommended to be included in admissions assessment systems (see **Table 1**).

Table 1 Description of noncognitive variables

Variable #	Variable name
1	<i>Positive self-concept</i> • Demonstrates confidence, strength of character, determination, and independence.
2	<i>Realistic self-appraisal</i> • Recognizes and accepts any strengths and deficiencies, especially academic, and works hard at self-development. Recognizes need to broaden his/her individuality.
3	<i>Understands and knows how to handle racism (the system)</i> • Exhibits a realistic view of the system based upon personal experience of racism. Committed to improving the existing system. Takes an assertive approach to dealing with existing wrongs, but is not hostile to society, nor is a “cop-out”. Able to handle racist system.
4	<i>Prefers long-range to short-term or immediate needs</i> • Able to respond to deferred gratification, plans ahead and sets goals.
5	<i>Availability of strong support person</i> • Seeks and takes advantage of a strong support network or has someone to turn to in a crisis or for encouragement.
6	<i>Successful leadership experience</i> • Demonstrates strong leadership in any area of his/her background (e.g., church, sports, noneducational groups, and gang leader).
7	<i>Demonstrated community service</i> • Participates and is involved in his/her community.
8	<i>Knowledge acquired in or about a field</i> • Acquires knowledge in a sustained and/or culturally related ways in any field.

Positive Self-Concept

There is evidence that the way students feel about themselves is related to their adjustment and success in college (Sedlacek, 2003, 2004). A strong self-concept is particularly important for students of color, students with disabilities, and women returning to school.

A number of studies have shown that a positive self-concept correlates with college grades, retention, and graduation, particularly the later two, for regularly admitted African-American students. O’Callaghan and Bryant (1990) found self-concept important for the success of black American students at the US Air Force Academy. A positive self-concept has also shown to be important for Asian and Pacific Islanders, Latinos, international students, and women, among others (Ancis *et al.*, 2000).

Realistic Self-Appraisal

Realistic self-appraisal is the ability to assess one’s strengths and weaknesses and allows for self-development. Realism in self-appraisal by nontraditional persons does not connote cultural, racial, or gender deficiency or inferiority.

White students may do well pursuing their own interests (internal control) in a society designed to meet their needs, while students of color need to also be aware of the external control on their lives which requires them to negotiate the racism in the system (Sedlacek, 2003, 2004, 2005). Perceived intrinsic interest in a field has been shown to be a major barrier to achieving career goals for whites and Asian Americans, while African-Americans, Latinos, and Native Americans cited personal finances as their major barrier.

Realistic self-appraisal has been found to correlate with college grades, retention, and graduation for students of all races, but the relationships were particularly strong for African-Americans. Women who are able to make realistic self-appraisals have been shown to get higher grades in a university than those who have difficulty with such assessments (Sedlacek, 2004).

Understands and Deals with Racism

The successful nontraditional student is a realist based on a personal experience with discrimination, is committed to fighting to improve the existing system, and is able to handle a racist system. Institutional racism is defined as the negative consequences that accrue to a member of a given group because of the way a system or subsystem operates in the society (e.g., college admissions), regardless of any other attributes of the individual (Sedlacek, 2003, 2004). Racism can take many forms and is used here to cover all types of ‘isms’ (e.g., sexism, ageism, and athleticism). While racism can be individual rather than institutional, the primary concern here is for dealing with the policies, procedures, and barriers, intentional or not, that interfere with the development of people.

For traditional students, the variable takes the form of handling the system without the addition of racism (Sedlacek, 2003, 2004, 2005). How one learns to handle the circumstances with which they are confronted tells us much about their ability and potential.

Steele’s (1997) work on stereotype threat supports the importance of the psychological set with which examinees approach a test. If African-Americans are told that they do

not usually do well on a test, they do less well than if a more positive set is given. Therefore, for African-Americans, the act of taking a test probably involves dealing with the racism that may have been involved in helping to create a stereotype threat in the first place. Hence, part of the variance that is being measured when an African-American takes the SAT is likely to relate to how that person handles racism.

Prefers Long-Range Goals

Having long-range goals will predict success in college for students. Since role models often are more difficult to find, and the reinforcement system has been relatively random for them, many nontraditional students have difficulty in understanding the relationship between current efforts and the ultimate practice of their professions (Sedlacek, 2003, 2004, 2005).

A significant relationship has been found between setting long-range goals and grades and retention for international students, including international community college students (Boyer and Sedlacek, 1988).

Availability of a Strong Support Person

Students who have done well in school tend to have a person of strong influence who provides advice to them, particularly in times of crisis (e.g., Sedlacek, 2003, 2004, 2005). This individual may be in the education system, but may be a relative or a community worker. Evidence is available for women, athletes of all races, and international students (Bennett and Okinaka, 1990). This variable involves having someone beyond a role model, although that may be part of the relationship. A strong support person takes an active role in advising and directing someone.

Successful Leadership Experience

Students who are most successful in higher education have shown an ability to organize and influence others. The key here is that nontraditional students may show evidence of leadership in different ways than their white counterparts (Liu and Sedlacek, 1999).

Asian-American students have unique and culturally related ways of expressing their leadership. If an applicant succeeds in his or her culture and is now ready to take on college, there is evidence that the student has the potential to succeed. There is evidence of the value of leadership in the retention of Latinos and Native Americans, and African-Americans in undergraduate and medical school programs (Webb *et al.*, 1997). Leadership has also been identified as a correlate of success for women and for female and male international students (Sedlacek, 2004).

Community

Having a community with which students can identify, and from which they can receive support, is critical to their academic success. For white students, there tend to be a number of opportunities to find a community, in or out of school. The community for nontraditional students often is based on racial, cultural, or gender-related variables. Students of color, women, and other persons with nontraditional experiences who are active in a community learn how to handle the system, exhibit leadership, and develop their self-concepts in such groups. For example, it has been found that African-American students who used campus athletic facilities and certain student union programs were more likely to stay in school than those who did not (Mallinckrodt and Sedlacek, 1987).

Identification with a community has been shown to be important for Asian-American, Latino, and Native-American success in school, and for male and female athletes of all races. Community also has been shown to be a correlate of success for undergraduate women, teachers in training, and international students (Sedlacek, 2004).

Nontraditional Knowledge Acquired

The ability of someone to learn from experiences outside the classroom correlates with their success in school. Persons of color are more apt to learn and develop using methods that are less traditional and are outside the education system. The methods may be culture or gender-related, and the field itself may be nontraditional students and women (Sedlacek, 2004). This learning should be demonstrated in an admission or scholarship application, rather than just listed as a job or internship. Several schools and programs use this method of assessment as discussed below.

Measuring Noncognitive Variables

The Noncognitive Questionnaire (NCQ) was designed to assess the eight noncognitive variables discussed above and shown in **Table 1** (Sedlacek, 1996). Several forms of the NCQ have been developed and employed in different selection contexts 2-week test-retest reliability estimates on NCQ scores ranging from .74 to .94, with a median of .85 (Sedlacek, 2004). Interrater reliability on scores from the three open-ended NCQ items ranged from .73 to 1.00.

The variables shown in **Table 1** have been successfully assessed in ways other than a version of the NCQ. In the Gates Millennium Scholars program, a review of an entire application is scored on the noncognitive variables, and makes up about 80% of the weight used in selection. The application includes short-answer questions based on each of

the noncognitive variables shown in **Table 1**, a personal statement by the applicant, letters of recommendation by the nominator and another person, and demographic, background, and activity questions. Raters were trained to identify and consider all this information in scoring each of the eight noncognitive variables. The raters were educators of color, familiar with multicultural issues in education, and in working with the kinds of students that were applying. Inter-judge reliability was estimated at .83 for a sample of raters in the first year (Sedlacek and Sheu, 2004). More than 11 000 Gates scholars have attended more than 1450 different colleges and universities with a 97% first-year retention rate, an 87% 5-year retention rate, and a 78% 5-year graduation rate. More than 60% are majoring in STEM (STEM: science, technology, engineering, mathematics) fields. Their realistic self-appraisal score has a significant relationship with their first-year college grade point average (GPA) and their leadership score has a significant relationship with engaging in academic activities while in college (Sedlacek and Sheu, 2008).

Oregon State University (OSU) has developed a selection system based on the noncognitive variables shown in **Table 1**. The OSU application contains six short-answer questions which cover the eight noncognitive variables. Responses are limited to 100 words and are scored independently from other application materials. Raters from many parts of the campus are trained to score the six questions. Inter-judge agreement was estimated at .85. OSU uses its system in selection, academic advising, student services, on and off campus referrals, financial aid, and teaching. OSU noncognitive scores correlate with retention, and since employing noncognitive variables the OSU retention rate is higher, there is more diversity in the applicant pool and first-year class, campus offices are working better together, applicant GPA is up, referrals are better, and new courses and student services have begun based on the noncognitive information.

The noncognitive variables can be used along with any other variables, models, or techniques that are employed in whatever role or type of mentoring or advising is involved. Teachers, advisors, or counselors who use the system can expect to obtain better student outcomes in terms of grades, retention, and satisfaction, as well as greater satisfaction themselves in employing something systematic with demonstrated utility in an area that often produces confusion and anxiety. Sedlacek *et al.* (2007) provide examples and case studies of how noncognitive variables can be used in postmatriculation programs.

- Angoff, W. H. (1971). *The College Board Admissions Testing Program*. New York: College Entrance Examination Board.
- Bennett, C. and Okinaka, A. M. (1990). Factors related to persistence among Asian, Black, Hispanic, and White undergraduates at a predominantly White University: Comparison between first and fourth year cohorts. *Urban Review* **22**, 33–60.
- Boyer, S. P. and Sedlacek, W. E. (1988). Noncognitive predictors of academic success for international students: A longitudinal study. *Journal of College Student Development* **29**, 218–222.
- Knapp, L. G., Kelly, J. E., Whitmore, R. W., Wu, S., and Gallego, L. M. (2002). *Enrollment in Postsecondary Institutions, Fall 2000 and Financial Statistics, Fiscal Year 2000* (NCES 2002-212). Washington, DC: National Center for Education Statistics.
- Lemann, N. (2000). *The Big Test: The Secret History of the American Meritocracy*. New York: Farrar, Straus & Giroux.
- Liu, W. M. and Sedlacek, W. E. (1999). Differences in leadership and co-curricular perception among male and female Asian Pacific American college students. *Journal of the Freshmen Year Experience* **11**, 93–114.
- Mallinckrodt, B. and Sedlacek, W. E. (1987). Student retention and the use of campus facilities by race. *National Association of Student Personnel Administrators Journal* **24**(3), 28–32.
- O'Callaghan, K. W. and Bryant, C. (1990). Noncognitive variables: A key to Black-American academic success at a military academy. *Journal of College Student Development* **31**, 121–126.
- Sedlacek, W. E. (1996). An empirical method of determining nontraditional group status. *Measurement and Evaluation in Counseling and Development* **28**, 200–210.
- Sedlacek, W. E. (1998a). Multiple choices for standardized tests. *Priorities* **10**, 1–16.
- Sedlacek, W. E. (1998b). Admissions in higher education: Measuring cognitive and noncognitive variables. In Wilds, D. J. and Wilson, R. (eds.) *Minorities in Higher Education 1997–98: Sixteenth Annual Status Report*, pp 47–71. Washington, DC: American Council on Education.
- Sedlacek, W. E. (2003). Alternative measures in admissions and scholarship selection. *Measurement and Evaluation in Counseling and Development* **35**, 263–272.
- Sedlacek, W. E. (2004). *Beyond the Big Test: Noncognitive Assessment in Higher Education*. San Francisco, CA: Jossey-Bass.
- Sedlacek, W. E. (2005). The case for noncognitive measures. In Camara, W. and Kimmel, E. (eds.) *Choosing Students: Higher Education Admission Tools for the 21st Century*, pp 177–193. Mahwah, NJ: Erlbaum.
- Sedlacek, W. E., Benjamin, E., Schlosser, L. Z., and Sheu, H. B. (2007). Mentoring in academia: Considerations for diverse populations. In Allen, T. D. and Eby, L. T. (eds.) *The Blackwell Handbook of Mentoring: A Multiple Perspectives Approach*, pp 259–280. Malden, MA: Blackwell.
- Sedlacek, W. E. and Sheu, H. B. (2004). Correlates of leadership activities of Gates Millennium Scholars. *Readings on Equal Education* **20**, 249–264.
- Sedlacek, W. E. and Sheu, H. B. (2008). The academic progress of undergraduate and graduate Gates Millennium Scholars and non-scholars by race and gender. *Readings on Equal Education* **23**, 143–177.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist* **52**, 613–629.
- Webb, C. T., Sedlacek, W. E., Cohen, D., *et al.* (1997). The impact of nonacademic variables on performance at two medical schools. *Journal of the National Medical Association* **89**(3), 173–180.

Bibliography

- Ancis, J. R., Sedlacek, W. E., and Mohr, J. J. (2000). Student perceptions of the campus cultural climate by race. *Journal of Counseling and Development* **78**(2), 180–185.

Further Reading

- Ancis, J. R. and Sedlacek, W. E. (1997). Predicting the academic achievement of female students using the SAT and noncognitive variables. *College and University* **72**(3), 1–8.

- Bennett, C. I. (2002). Enhancing ethnic diversity at a Big Ten University through project TEAM: A case study in teacher education. *Educational Researcher* **31**, 21–29.
- Fuertes, J. N. and Sedlacek, W. E. (1995). Using noncognitive variables to predict the grades and retention of Hispanic students. *College Student Affairs Journal* **14**, 30–36.
- Fuertes, J. N., Sedlacek, W. E., and Liu, W. (1994). Using the SAT and noncognitive variables to predict the grades and retention of Asian American university students. *Measurement and Evaluation in Counseling and Development* **27**, 74–84.
- Harvey, W. B. (2002). *Minorities in Higher Education 2001–2002: Nineteenth Annual Status Report*. Washington, DC: American Council on Education.
- McTighe Musil, C., García, M., Hudgins, C. A., et al. (1999). *To Form a More Perfect Union: Campus Diversity Initiatives*. Washington, DC: Association of American Colleges & Universities.
- Mohr, J. J. and Sedlacek, W. E. (2000). Perceived barriers to friendship with lesbians and gay men among university students. *Journal of College Student Development* **41**, 70–79.
- Moore, S. K. (1995). *Indicators of Academic Success and the Student Characteristics of International Students at Santa Monica College*. Unpublished doctoral dissertation, Pepperdine University.
- Neville, H. A., Heppner, P., and Wang, L. (1997). Relations among racial identity attitudes, perceived stressors, and coping styles in African-American college students. *Journal of Counseling and Development* **75**, 303–311.
- Patterson, A., Sedlacek, W. E., and Scales, W. R. (1988). The other minority: Disabled student backgrounds and attitudes toward their university and its services. *Journal of Postsecondary Education and Disability* **6**, 86–94.
- Sandlin, M. (2008). *Insight Resume Training*. Corvallis, OR: Oregon State University. <http://oregonstate.edu/admissions/2004req/2004profile.html> (accessed October 2009).
- Sedlacek, W. E. (1994). Issues in advancing diversity through assessment. *Journal of Counseling and Development* **72**, 549–553.
- Sedlacek, W. E. (1995). Using research to reduce racism at a university. *Journal of Humanistic Education and Development* **33**, 131–140.
- Sedlacek, W. E. and Adams-Gaston, J. (1992). Predicting the academic success of student-athletes using SAT and noncognitive variables. *Journal of Counseling and Development* **70**(6), 724–727.
- Sedlacek, W. E. and Brooks, G. C., Jr. (1976). *Racism in American Education: A Model for Change*. Nelson-Hall: Chicago, IL.
- Ting, S. R. (1997). Estimating academic success in the first year of college for specially admitted White students: A model combining cognitive and psychosocial predictors. *Journal of College Student Development* **38**, 401–409.
- Webb, C. T., Sedlacek, W. E., Cohen, D., et al. (1997). The impact of nonacademic variables on performance at two medical schools. *Journal of the National Medical Association* **89**(3), 173–180.
- White, C. J. and Shelley, C. (1996). Telling stories: Students and administrators talk about retention. In Johnson, I. H. and Ottens, A. J. (eds.) *Leveling the Playing Field: Promoting Academic Success for Students of Color*, pp 15–34. San Francisco, CA: Jossey Bass.

Peace Education

J Page, Southern Cross University, Lismore, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

The Problem of Defining Peace Education

Peace education is notoriously difficult to define. There are arguably three levels of understanding peace education. The primary concern of peace education is to prevent the suffering and wastage associated with warfare. A secondary concern is the linkage with cognate social concerns, such as those reflected in development education, education for international understanding, human-rights education, futures education, inclusive education, education for social justice, and environmental education. A third level of peace education is what might be called the intrapersonal and interpersonal aspects of peace education, dealing with self-understanding, self-fulfillment, and how we interact with each other and our environment at a personal level.

One of the noteworthy after developments in peace education has been an increasing interest in the above interpersonal and intrapersonal aspects of peace education. In other words, we need not be merely concerned with the prevention of violence on a governmental and social level, but we need also to be concerned with local and domestic violence, and with the quality of our everyday relationships. Ultimately, the personal and social dimensions of peace education are only classifications, and each is uniquely linked. Indeed, one of the overall lessons of peace education is that the personal is uniquely important: we teach peace education substantially through who we are.

Maximalist Versus Minimalist Peace Education

Another way of understanding peace education is through the notion of maximalist versus minimalist peace education. The maximalist understanding of peace education involves educating individuals toward sensitivity and responsibility in a range of areas, including personal fulfillment, the creation of a just and cooperative society, inclusiveness, and care for the environment. Minimalist understandings of peace education involve educating individuals toward learning to reject war and aggression, and learning to reject militarism and arms races. Both understandings pose challenges. Clearly, it is insufficient to educate people to reject war without encouraging an awareness of social injustice and inequality. However, at the same time, one problem is that peace education can be

seen as encompassing all educational endeavor so that peace education merely becomes a synonym for education.

The definitional problem of peace education can be seen as an extension of the definitional problem of peace itself. The Norwegian peace researcher Johan Galtung is widely recognized for emphasizing the difference between negative peace, that is, the absence of armed conflict, and positive peace, that is, the presence of cooperative and harmonious relationships. Johan Galtung has also developed a threefold classification of direct, structural, and cultural peace: direct peace denoting the absence of armed conflict, structural peace denoting the presence of justice and just structures, and cultural peace denoting the presence of a vital culture of peace.

The above classifications can be useful for peace education. Peace education can be seen as promoting both negative and positive peace. Peace education can be seen as encouraging direct peace, that is, educating individuals to see the dangers of war; encouraging structural peace, that is, educating individuals to see the importance of social justice; and encouraging cultural peace, that is, educating individuals to see the importance of a culture of peace.

A Brief History of Peace Education

Peace education arguably has a long history. If we view the major world religions as striving to encourage cooperation and harmony, if only internally, then the propagation of these world religions serves as a form of peace education. Similarly, one of the interesting historical phenomena of European history has been recurring peace plans, aiming to prevent wars between European monarchs. These too can be seen as a form of peace education. The beginning of a self-conscious idea of education as a means of preventing war can be traced to the philosophy of education of John Dewey and Maria Montessori. Both were motivated very much by reflection upon modern warfare, either a world war just ended (for John Dewey) or a new world war on the horizon (for Maria Montessori).

In more recent times, the importance of peace education has become directly recognized in international law, namely, in United Nations documents and declarations. One of the interesting aspects of international declarations on peace education is that these have become more explicit in recent years. Peace education is now increasingly recognized as a professional field, with the Peace Education Commission recognized as one of the Commissions

of the International Peace Research Association. Peace education is widely represented in scholarly publication, with the publication of numerous monographs, and with articles on peace education common in many educational journals. There is also now a specialist international scholarly periodical, the *Journal of Peace Education*.

Peace Education and International Law

It is easy to decry peace education as being at best an idealistic nonessential aspect of educational endeavor and at worst political indoctrination in the guise of education. This is why the recognition of the importance of peace education within international law is so important. International law has no enforceable authority of its own. However, there is a powerful moral and symbolic force within international law, and the recognition of the importance and role of peace education is an indicator that peace education is not merely an invention of particular interest groups and is indeed an essential part of education.

The Charter of the United Nations remains one of the core documents of international law, and the Charter outlines the special purpose of the United Nations in preventing future war. Peace education is one crucial means by which this aim can be fulfilled. The Preamble to the Charter refers to reaffirming faith in the dignity of the human person and establishing conditions under which justice and respect for international obligations may be maintained. It is difficult to see the above task of reaffirming faith in the dignity of the individual or establishing respect for international obligations as being anything other than an educational task.

Another fundamental document of international law is the Universal Declaration of Human Rights. Article 26 indicates that education shall be directed to strengthening respect for human rights and fundamental freedoms, promoting understanding, tolerance, and friendship, and education shall further the activities of the United Nations for the maintenance of peace. The idea of maintaining peace suggests the more limited notion of negative peace, although the other elements of Article 26 clearly refer to a wider notion of encouraging what we might describe as education for positive peace. Within both of the above documents, there is an implicit endorsement of the importance of education for peace.

Peace Education and UNESCO

The agency within the United Nations with primary responsibility for educational policy is the United Nations Educational, Scientific and Cultural Organization (UNESCO). The Preamble of the UNESCO Constitution indicates that the organization was formed for the promotion of a culture of peace; that is, as war commences in the minds of

individuals, so too defenses against war must be similarly constructed in the minds of individuals. UNESCO has regularly issued documents confirming the importance of peace education, most recently as the lead agency in the United Nations commitment to encouraging a culture of peace, such as in the International Year for the Culture of Peace, and the International Decade for a Culture of Peace and Nonviolence for the Children of the World.

Within United Nations programs on education for tolerance, there are numerous commitments to peace education. The 1995 UNESCO Declaration of Principles on Tolerance and the 1996 United Nations General Assembly Follow-up to the United Nations Year for Tolerance make it clear that tolerance is an essential component of peace. Article 1 of the UNESCO document describes tolerance as the virtue which makes peace possible and which contributes to the replacement of a culture of violence by a culture of peace. Article 4 of the UNESCO document involves a commitment to improve teacher training, curricula, textbooks, lessons, and educational materials in order to create caring and responsible citizens who are open to other cultures, able to appreciate the value of freedom, respectful of human dignity and differences, and able to prevent conflicts or resolve such conflicts through nonviolent means.

Peace Education and Disarmament Education

Disarmament education can be considered a subfield within peace education. The 1978 United Nations Special Session on Disarmament I, the 1980 World Congress on Disarmament Education, and the 1982 United Nations Special Session on Disarmament II all urged governments and international organizations to develop programs in disarmament and peace education. The Special Session II initiated the World Disarmament Campaign, which in turn became a catalyst for many peace movement initiatives of the following decade. The World Disarmament Campaign was an example of how intergovernmental action can impact upon popular culture and of how it is possible to lead in promoting peace.

Since that time, the World Disarmament Campaign has been succeeded by a permanent agency, the United Nations Office of Disarmament Affairs (UNODA), the focus of which has been on diplomacy rather than education. However, in 2002, the United Nations General Assembly commissioned an expert report, The United Nations Study on Disarmament and Non-Proliferation Education. The report renews the United Nations commitment to peace education, specifically in Articles 6–10. Article 20 usefully describes peace and disarmament education as a base of theoretical and practical knowledge, allowing individuals to choose values which reject violence, resolve conflicts peacefully, and sustain a culture of peace.

Peace Education as a Right

If it is tempting to see peace education as merely an ideal, one counter to this is to view peace education as a right. The 1959 Declaration of the Rights of the Child emphasizes the rights children have to protection and education, and Principle 7 expressly states that a child has a right to an education that will develop a sense of moral and social responsibility. One could argue that a corollary of this is that a child has a right to peace education. The same thrust can be discerned in the 1989 Convention on the Rights of the Child, with Article 29.1(d) indicating that education should be directed toward the preparation of the child for responsible life in a spirit of understanding, peace, equality, and friendship among all peoples.

If we see peace education as a right, then it is not too difficult to see the connection between a right to peace and a right to peace education, and the connection between peace education and human rights education. The 1984 United Nations General Assembly Declaration on the Right of Peoples to Peace recognized peace as a right. Rights are meaningless if individuals do not know of these rights, and in particular if peace is a right, then it follows that individuals have a right to know about this particular right. If peace is a right, it follows that peace education ought also to be regarded as a right. It is thus not surprising that the 1993 World Convention on Human Rights in Vienna recognized peace education as being part of human rights education. One flows from the other.

Peace Education and the Culture of Peace

The expansion of the concept of peace education has been reflected more recently in the culture of peace programs of the United Nations. The rationale behind the culture of peace movement is that peace involves more than governmental action; it is a civil and cultural process, encompassing all sectors of society. Peace is not merely institutional but personal. Thus peace education must ultimately involve teaching to encourage a culture of peace, however complex and difficult this might be.

This integrated understanding of peace and peace education is reflected in the 1999 United Nations General Assembly Declaration and Programme of Action on a Culture of Peace. Article 4 of the Declaration indicates that education is one of the principal means of building a culture of peace. Article 9 of the program of action contains specific actions for fostering a culture of peace through education, including involving children in activities for instilling the values and goals of a culture of peace, revision of curricula and textbooks with regards to peace, encouraging and strengthening efforts in developing skills and values supporting a culture of peace, and expanding the culture of peace initiatives in institutions of higher education.

Delimiting Peace Education

Any discussion of peace education ought to include the relationship of peace education to peace advocacy, and the relationship of peace education to peace research. Peace advocacy is practiced by peace groups seeking to emphasize the importance of peace and the destructiveness of war. Peace education similarly seeks to emphasize the importance of peace and the destructiveness of war. However, the difference is methodology. An important aspect of the educational task is that the educator must always respect the autonomy of the learner; thus, there is a fundamental difference between preaching and teaching peace. What makes this distinction more complex is that the task of emphasizing peace is no doubt better accomplished through recognizing the importance of the autonomy of learning.

The relationship of peace education to peace research is similarly a complex one. If we define peace research as normative applied research with the aim of emphasizing the importance of a commitment to peace, then it becomes clear that this is not far removed from the aims of peace education itself. It is for this reason that peace research and education are often linked together. Peace research is ultimately an educative endeavor. Similarly, peace education aims to clarify the causes of war and social injustice, and how we might work to prevent these in the future. Thus peace education may also be seen as a research process.

The Philosophy of Peace Education

On a fundamental level, one of the problems facing peace education has been the dearth of developed educational rationale or philosophy of peace education. One of the reasons for this is that those who are involved in peace education are likely to see the need for peace education as obvious, and thus the temptation is to see the need to articulate a philosophy of peace education as unnecessary. However, one of the marks of any valid educational endeavor is the existence of a developed educational rationale. Without such a rationale, peace education can be dismissed as indoctrination or political correctness.

There have been recent attempts to develop a coherent philosophy of peace education. Maltese educationist Joachim James Calleja sees a possible basis for peace education in the notion of duty and specifically in the Kantian categorical imperative. The fact that Immanuel Kant was a peace proponent adds weight to this specific line of argument. Australian educationist James Page identifies five possible ethical foundations for peace education: virtue ethics, consequentialist ethics, conservative political ethics, esthetics ethics, and the ethics of care. It is argued that none of these is conclusive in themselves, but each is part of a whole and credible rationale for peace education.

The Methodology of Peace Education

Just as there are layers of levels to peace education, so too one can suggest there are layers to the methodology of peace education, involving curriculum, structures and process, and personal leadership. The curriculum level is perhaps most appropriate to secondary and tertiary education. For instance, it is important to have a curriculum that does not necessarily follow a narrow nationalistic agenda and that allows for a genuinely international perspective. It is important to have a curriculum which acknowledges dispossession and marginalization of peoples. The teaching of history and civics is where this comes into focus most clearly, although it is also important in a range of other subjects and fields.

Structures and processes are important in that without just structures and processes within educational institutions, then all the discourse about peace within any curriculum can remain mere rhetoric. Indeed, without just structures and processes, discourse about peace within a curriculum can be destructive in that it can engender cynicism and disengagement. Peace and justice go together. If one has an educational program with much discourse about peace, and the educational context of that program is an educational institution with a dearth of justice in structure and processes, then clearly students will tend to be cynical about the peace education program.

The final level of peace education – personal leadership – is the most challenging. We learn through example and role modeling. We learn peace through seeing peaceful people in action. If someone aspiring to teach peace has a domineering teaching style, then this clearly is counterproductive. Peace education is concerned with the empowerment of the other. The role of a peaceful and encouraging teacher is not a simple one, especially in demanding and often conflict-ridden educational settings. Yet, there are steps which can be taken to empower and support the teacher. The methodology of peace education, and indeed peace education itself, remains an emerging field. Overall, more research, both theoretical and practical, is needed within this field for the future.

See also: Curriculum and Human Rights; Education in War Countries: A Disempowerment of Society; Justice and Care; Peace Education; School Choice and the Common School.

Further Reading

- Bar-Tal, D. (2002). The elusive nature of peace education. In Salomon, G. and Nevo, B. (eds.) *Peace Education: The Concept, Principles and Practices around the World*, pp 27–36. Mahwah, NJ: Lawrence Erlbaum Associates.
- Boulding, E. (1988). *Building a Global Civic Culture: Education for an Interdependent World*. New York and London: Teachers College Press.
- Boulding, E. (1989). Can peace be imagined? In Forcey, L. R. (ed.) *Peace: Meanings, Politics, Strategies*, pp 73–86. New York: Praeger.
- Calleja, J. J. (1991). *A Kantian Epistemology of Education and Peace: An Examination of Concepts and Values*. Unpublished PhD Thesis: University of Bradford. Department of Peace Studies.
- Dewey, J. (1967–91). *The Collected Works of John Dewey, 1882–1953*. In Boydston, J. O. (ed.). Carbondale: Southern Illinois University Press.
- Galtung, J. (1975–88). *Essays in Peace Research* (6 volumes). Copenhagen: Ejlers.
- Galtung, J. (1983). Peace education: Learning to hate war, love peace and do something about it. *International Review of Education* 29(3), 281–287.
- Galtung, J. (1996). *Peace by Peaceful Means: Peace and Conflict, Development and Civilization*. London: Sage Publications.
- Gordon, H. and Grob, L. (eds.) (1987). *Education for Peace: Testimonies from World Religions*. Maryknoll: Orbis Books.
- Gur-Ze'ev, I. (2001). Philosophy of peace education in a postmodern era. *Educational Theory* 51(3), 315–336.
- Haavelsrud, M. (1975). *Principles of Peace Education*. Oslo: University of Oslo.
- Haavelsrud, M. (1981). *Approaching Disarmament Education*. Guildford: Westbury House.
- Haavelsrud, M. (1983). An introduction to the debate on peace education. *International Review of Education* 29(3), 275–280.
- Harris, I. M. and Morrison, M. L. (2003). *Peace Education*, 2nd edn. Jefferson: McFarland.
- Hicks, D. (ed.) (1988). *Education for Peace – Issues, Principles and Practice in the Classroom*. New York: Routledge.
- Hutchinson, F. P. (1996). *Educating beyond Violent Futures*. London and New York: Routledge.
- Höffe, O. (ed.) (2004). *Immanuel Kant, Zum ewigen Frieden*. Berlin: Akademie Verlag.
- IBE (1994). *Final Report of the 44th Session of the International Conference on Education*. Geneva: International Bureau of Education.
- IBE (1995). Declaration and integrated framework of action on education for peace, human rights and democracy. *From the 44th Session of the International Conference on Education*. Geneva: International Bureau of Education.
- Lin, S. (1999). *Vorurteile überwinden – eine friedenspädagogische Aufgabe*. Weinheim and Basel: Beltz Pädagogik.
- McCarthy, C. (2002). *I'd Rather Teach Peace: Lessons from the School of Nonviolence*. Maryknoll: Orbis Books.
- Montessori, M. (1949). *Educazione e pace*. Milano: Garzanti.
- Page, J. S. (2000). Can history teach us peace? *Peace Review* 12(3), 441–448.
- Page, J. S. (2001). The international year for the culture of peace: Was it worthwhile? *International Journal of Cultural Studies* 4(3), 348–351.
- Page, J. S. (2004). Peace education: Exploring some philosophical foundations. *International Review of Education* 50(1), 3–15.
- Page, J. S. (2007). Teaching peace to the military. *Peace Review* 19(4), 189–196.
- Page, J. S. (2008). *Peace Education: Exploring Ethical and Philosophical Foundations*. Charlotte: Information Age Publishing.
- Page, J. S. (2008). The United Nations and peace education. In Bajaj, M. (ed.) *Encyclopedia of Peace Education*, pp 75–83. Charlotte: Information Age Publishing.
- Reardon, B. A. (1989). *Comprehensive Peace Education*. New York: Teachers College Press.
- Reardon, B. A. (1997). Human rights as education for peace. In Andreopoulos, G. J. and Pierre Claude, R. (eds.) *Human Rights Education for the Twenty-First Century*, pp 255–261. Philadelphia: University of Pennsylvania Press.
- Reardon, B. A. and Cabezudo, A. (2002). *Learning to Abolish War: Teaching Towards a Culture of Peace*. New York: Hague Appeal for Peace, Global Campaign for Peace Education.
- Roche, D. (2003). *The Human Right to Peace*. Toronto: Novalis.
- Smith, D. C. and Carson, T. R. (1998). *Educating for a Peaceful Future*. Toronto: Kagan and Woo.
- Synott, J. P. (2005). Peace education as an educational paradigm: Review of a changing field using an old measure. *Journal of Peace Education* 2(1), 3–16.
- UNICEF (1996). *State of the World's Children*. New York: Oxford University Press.
- UNICEF (1999). *Peace Education in UNICEF*. Prepared by Susan Fountain. New York: UNICEF. Staff Working Paper Series.

- UNESCO (1945). *Constitution of the United Nations Educational, Scientific and Cultural Organization*. London: UNESCO. Adopted on 16/11/45.
- UNESCO (1974). *Recommendation Concerning Education for International Understanding, Co-operation and Peace and Education Relating to Human Rights and Fundamental Freedoms*. Paris: UNESCO. Recommendation adopted by the General Conference of UNESCO at the 18th Session, on 19/11/1974.
- UNESCO (1980). *World Congress on Disarmament Education: Final Document and Report*. Paris: UNESCO.
- UNESCO (1994/5). *Declaration and Integrated Framework of Action on Education for Peace, Human Rights and Democracy*. Paris: UNESCO. Declaration of the 44th Session of the International Conference on Education, Geneva, October 1994, endorsed by the General Conference of UNESCO at its 28th Session, October-November, 1995.
- UNESCO (1995). *Declaration of Principles on Tolerance*. Paris: UNESCO. Adopted by the General Conference of UNESCO on 16/11/95.
- UNGA (1959). *Declaration on the Rights of the Child*. New York: United Nations General Assembly. Resolution A/14/1386.
- UNGA (1978). *Final Document of the Tenth Special Session of the United Nations General Assembly (SSODI-Special Session on Disarmament I)*. New York: United Nations General Assembly.
- UNGA (1984). *Right of Peoples to Peace*. New York: United Nations General Assembly. Resolution A/51/201.
- UNGA (1993). *Vienna Declaration and Programme of Action (World Conference on Human Rights)*. New York: United Nations General Assembly. A/CONF. 157/23. 12/7/93.
- UNGA (1995). *International Decade for a Culture of Peace and Non-violence for the Children of the World (2001-2010)*. New York: United Nations General Assembly. Resolution A/53/25.
- UNGA (1996). *Final Report on the United Nations Year for Tolerance*. New York: United Nations General Assembly. Resolution A/51/201.
- UNGA (1998). *International Decade for a Culture of Peace and Non-Violence for the Children of the World*. New York: United Nations General Assembly. Resolution A/53/25.
- UNGA (1999). *Declaration and Programme of Action on a Culture of Peace*. New York: United Nations General Assembly. Resolution A/53/243.
- UNGA (2000). *International Decade for a Culture of Peace and Non-violence for the Children of the World: Report of the Secretary-General*. New York: United Nations General Assembly. Reference A/55/377.
- UNGA (2002). *United Nations Study on Disarmament and Non-proliferation Education: Report of the Secretary-General*. New York: United Nations General Assembly. Resolution A/57/124.
- UNICEF (1999). *Peace Education in UNICEF*. Prepared by Susan Fountain. New York: UNICEF. Staff Working Paper Series.

Remote Control: Africa's Development and Outflow to Developed Nations of Technical Talent

F Haniff, Wits International Office, University of the Witwatersrand, Johannesburg, South Africa

© 2010 Elsevier Ltd. All rights reserved.

Moving to South Africa in 1994 was both a political and personal victory, to live in a free South Africa, when the past 10 years were spent, like millions of activists, in the hope of a free South Africa, but on the education front, where do we find ourselves, ahead, with the pack, or otherwise?

The author's field of interest has always been in the area of people development; in the traditional corporate human resource (HR) profession; or in change management via people development interventions.

In 1995 the author found her first job in South Africa (SA) in the higher education sector, but in her HR field. This was the eye-opening experience that left the author with the why question. This experience was further enriched by exposure to her partner's job at the University of the Witwatersrand, Johannesburg, fondly known as Wits or Wits University.

The more than 5 years the author spent in the corporate world reaffirmed the thinking that education is the cornerstone for intellectual and economic stimulation and development. This reaffirmation was frighteningly confirmed during her consultancy stint via the audit of the HR interventions by the Tertiary Education Linkage Program (TELP) between 2000 and 2005 in SA higher education (HE) institutions.

In her current role as director of the Wits International Office (WIO) at the University of the Witwatersrand, Johannesburg and her recent immediate past president portfolio at the International Education Association of South Africa (IEASA), she came full circle to the place of departure of her thinking about the role of aid, if the aid is indeed done in the interest of the 'developing' country that is supposed to benefit from it; in the race to secure economic dominance via access and control of the world's pool of talent.

Recent statistics has indicated that Africa's contribution to the world's pool on scientific and technological research is less than 1%, in fact 0.97% of this contribution, South Africa contributes 90%. This then translates into the reality that, outside of South Africa, the rest of Africa does not really feature as developing the talent that they need, or more accurately, retaining talent.

Research in the return rate of African nationals taking up studies in Europe, North America, and Australia revealed that over 80% of this talent pool do not return to their home country, even when their education is

financially supported and often part of a national scheme to revitalize their home country's talent pool.

The new study says some 65 000 African-born physicians and 70 000 African-born professional nurses were working overseas in developed countries by the year 2000. It says this represents about one in five African-born physicians and about one-tenth of African-born professional nurses. According to the British government, more than 17 000 doctors and nurses from Africa were recruited in 2007 to work in Britain. (Tendai Maphosa London, VOA, 11 January 2008)

What this illustrates is that African universities are quite capable in producing quality talented individuals; however, the retention of such individuals is not factored in by any schemes either by government or donors. This leads to support, especially in areas of infrastructure, salaries, benefits, and good working conditions. While many western donors are putting together elaborate schemes to develop talent under many schemes, the majority of the schemes are for the participants to study abroad, where such talent is recruited by those institutions and such talent rarely returns to the home country. This is clearly a flawed solution and lends itself more to controlling what they want to develop for themselves.

While institutions like UCLA working in Africa focused schemes that support postgraduate programs that are on condition that the recipients of the scholarship must return to their home country on completion of the program, fellows are expected to return to their home countries (African Development Dissertation Workshop Program: Institute of International Studies; UC Berkeley Funding Opportunities for Africans and Foreign Nationals). This example is rare. However, it only goes so far; an improvement would be for such programs to financially assist the African university to pay for a lab or the professor's salary for the first year, to create a welcoming environment to assist the professor's re-entry.

The Southern African Norwegian Training, Education and Development (SANTED) initiative, fully funded by the Norwegian government, provided an open forum for Southern African countries in the HE sector to look for creative solutions for indigenous projects that addressed regional needs and solutions. This funding created capacity development for all participating institutions.

The collaboration that the University of the Witwatersrand, Johannesburg, fondly known as Wits, collaborated with three other institutions, primarily the University of Namibia, University of Eduardo Mondlane, and in one subject area, the University of Botswana. The collaboration was in economics and labor studies, biological sciences, and engineering. With additions of new projects to further build capacity in the region, fulfilling the goals of SANTED and working toward the realization of the SADC Protocol on Education.

The institutional cooperation is intended to develop high-level capacity through the production of quality graduates and postgraduates who engage in advanced research projects. Apart from meeting elements of the SADC Protocol for HE, the project has obvious benefits for the institutions themselves and for the region as a whole. The specific projects chosen have either become, or will be, at the end of the project, embedded in the institutions and will be sustainable beyond the project funding.

All three universities bring to the project skills, expertise, and capacities; human, capital, and infrastructural resources; and challenges and limitations that will be addressed in this report.

The participating institutions, universities of Namibia, Eduardo Mondlane, and the Witwatersrand have similar backgrounds in their national demands; education and training, as well as research and development programs have been designed to meet national and international, human resource requirements. This is done through quality teaching, research, consultancy, and community service. The university continues to concentrate on key objectives of enabling students to reach their optimum levels of academic performance and improvement with respect to independent thinking and practical learning, communications skills, knowledge, networking for problem solving and trans-disciplinary approaches, research capabilities, publications, and writing skills. The university's graduate employment record continues to be encouraging. However, without a deliberate plan to attract, develop, and retain talent, the country's ability to grow its intellectual pool to respond to its domestic and regional needs will continue to lose talent to the west. This is the case of every African HE institution, the country, and the continent.

Wits University was identified as the coordinating host institution for the SANTED 'NEW' II project, which was championed by Professor Thandwa Mthembu, DVC (Prof. Mthembu is now the vice chancellor of Central University of Technology); currently, the project falls under the portfolio of Prof. Yunus Ballim, who is the deputy vice-chancellor (academic) and vice principal. The coordinating project team from the Wits International Office included Ms. Fazela Haniff, Mr. Lebethe Malefo, and Ms. Cindy Fredericks and the Late Ms Deborah McCONNell, who was part of the first SANTED team.

A unique element of the program lies in the biological science collaboration, where team teaching at a remote research location was developed. This allowed for faculty development via team teaching and the students experiencing the teaching styles and technique from different institutions. One of the highlights of this section of the project was seen by the group presentations of the students, to name a few, in the areas of:

- The distribution of phytoplankton in the Benguela.
- Variation in densities of *Welwitschia* and occurrence of small mammals in the gravel plains of the Namib.
- Comparisons of the morphological adaptations of *Meriones* lizards in the Namib desert.
- Changes in species richness, composition, and cover of terrestrial vegetation from coastline inland in the Namib desert.
- Namibian rocky shores: the nature of zonation on the northern Namibian coast.

The university senate approved the program as recommended by the postgraduate committee. Plans are underway for the program to be taught in 2009. Lecturers will be sought from all participating universities to deliver the lectures.

The challenges of communication and proper reporting, particularly from UEM, remain the greatest obstacles and limitations. Capacity has been increased in the Wits International Office to galvanize more support to all the projects managed from that office.

Like any collaboration project, it is imperative that commitment and buy-in from all stakeholders is solicited concurrently in order to achieve the objectives of this project. This is an ongoing process. SANTED 'NEW' II offers the SADC region a tremendous opportunity in the development of skills and building of capacity for all stakeholders. Once this is internalized, the project model will offer the region and other African institutions another option to maximize the value of the talent already on the continent.

The program funding provided resources in the following subject areas and categories. In the biological sciences, engineering, and economics, the program included curriculum development, support for materials, computers, books, research field trips, relief teaching, translation of course packages, support for university structures to approve masters curriculum, faculty development at masters and PhD levels for all participating institutions, and postgraduate and administrative coordination.

The example at UCLA, the joint program of SANTED 'NEW' II with Wits, UNEM, and UEM funded by the Norwegian government are few of the good things that are happening for African HE. However, the latter example of SANTED II will be soon closing, and there is practically no funding program that allows only African universities to develop and deliver their model of development.

The new order is that funding is no longer directly given to institutions unless there is a 'home-based' benefit. For the funding home country, the African university must partner with the funder's home-country institution. However good the intention to share skills, knowledge, etc., this scenario still benefits the funding home country's knowledge base and ultimately remotely controls how Africa develops, at what pace, and even what is developed – usually mostly to the benefit of the home country.

Within efforts to participate in and initiate initiatives to resuscitate either country development plans, to strengthen HE institution's capacity in Africa the west's intervention has been calculating and quite deliberate to favor schemes that primarily focus on strengthening their diversity goals, their quest to access data, and research intellectual capital to maintain their dominance in this sector. The most common misinformation is that foreign-aid money is used to assist others. This is primarily the most successful international marketing scheme of selling a product that does not exist. This is primarily a product to employ nationals of rich countries to spread some crumbs under the guise of international development programs.

Whilst the above example illustrates the uniqueness of regional capacity building, to assist with indigenous relevance and capacity development; the shift now by the Nordic countries, including Norway, and the European and North American philanthropical organizations to pull away from these models shows a reluctance or deliberate attempt to remotely control the pace and the path of HE development in Africa. The USAID's recent focus on Africa to dish out a mere one million dollars for all of African intuitions to access these funds – only if they participate in partnerships with other US institutions and only if that institution's area of interest happens to fall within the needs of the funding institution or country – is based on such a joint program for funding. The same trends are also now being followed by the Scandinavian and European countries. If one is not working with a funding-country institution, funding is primarily not available.

In the philanthropy world, the trend is youth centered; however, the skills to help the youth, counseling, research, intellectual capital at masters or PhD level, and support to strengthen universities to attract and retain professors and researchers have virtually been depleted.

Martin Davidson, Chief Executive of the British Council, said: "Fundamentally, this report shows the shift of axis of our education system from one that operates predominantly domestically to one that operates on a truly international basis. However, our position is vulnerable. Unless we start taking education much more seriously as a global business, we will lose out to other countries who understand the value of education to their economy much better than we do."

From Davidson's remarks, international education is moving toward running universities as a business, one that values talent and develops talent for national interest.

In the UK scenario, the British government and business community responds to this need by financially supporting their HE institutions. The growing need to quench the thirst for knowledge by workers, many governments are shifting gears to streamline their outreach to developing countries to 'help' Africa, or to help themselves. Most schemes are developed to hand out scholarships for African nationals to attend their universities; however, going back to our initial position, all these initiatives are assisting these governments and their institutions to have access to talent and ultimately retain this talent, to foster their diversity, and take control of knowledge and the knowledge economy.

Bernard makes the point that the categories of prior residence and country of prior education limit the data to be conclusive about its correlation to mobility and the residency, after the degree. Is this because the data-collection bodies, normally funded by those funding sources do not want to show that these schemes take talent away from developing nations, in this case Africa, the same group that such schemes are supposed to develop their talent and enable them to return to the home country to rejuvenate their intuitions and ultimately their economies? (Higher Education on the Move: New Developments in Global Mobility: Bernd Wachter, p. 59)

Further, in the same publication, Jane Knight also makes the case that graduates from top universities and their employability are closely linked. With investment in talent, and such talent not returning home, this situation then takes away the 'development' from African economies – hence remotely controlling the HE sector and economic development in developing countries.

The South African Network of Skills Abroad (SANSA) is an example. Through its website, it invites professional South Africans to sign up. It reports that at least 22 000 graduates from five major South African universities resident abroad remain in touch with the universities. SANSA estimates that about 60 per cent of the country's expatriate graduates are located in six countries, with Australia, the UK and the US accounting for more than half of them. Looking at the nature of their skills, the group estimates that about 30 per cent of the University of Cape Town's contactable doctoral graduates are living overseas. They comprise significant proportions of the university's graduates in medicine, commerce, education and engineering, all areas in which South Africa has an acute shortage of skills. (Gumisai Mutume, 2003: 1)

This further illustrates that the goodwill of education either within or outside South Africa is not enough in the equation to attract and retain talent. While it has taken the aid community and the government has taken some time to find a solution; there are some pockets of Industrialised countries are in growing need of two types of immigrant labour—those willing to do poorly paid, dirty and dangerous

jobs that their own nationals scorn, and highly specialized professionals, such as software specialists, engineers, doctors and nurses. The US has 126 000 fewer nurses than it needs and government figures show that the country could face a shortage of 800 000 registered nurses by 2020. Because of such shortages, industrialized nations have embarked on massive international recruitment drives. South Africa recently had to appeal to the government of Canada to desist from recruiting its medical professionals. In the rural province of Saskatchewan, Canada, more than 50 per cent of doctors are foreign trained and at least 1 in 5 of the 1 530 doctors there earned in Northern countries is forecast to grow, as younger people are needed to maintain productivity. In poorer countries, millions will continue to seek opportunities in richer countries to find better paying jobs and raise their standards of living. And in a globalizing world, where the dominant economic paradigm promotes the free movement of capital, it will become increasingly difficult to restrict the free movement of skilled labour their first medical degree in South Africa. However, it may become even tougher to stem the outward flow of skilled professionals from developing countries in future. With falling birth rates and aging populations, demand for labour. (Gumisai Mutume, 2003: 4)

The trend now is for western countries to offer wonderfully packaged schemes for teams of African universities to pool their talent to willingly participate in different consortiums, research groups to offer their findings for the 'global' funding agency. Long gone are the opportunities for such agencies to support Africa to develop and retain its own talent. The October 2008 USAID meeting in Kigali to address their African development agenda only promised \$1M for projects with African universities that must include a US university partner institution. This scenario clearly gives the US partner university the upper hand in the relationship.

Even the great idea of several foundations joining hands to deliver bigger and better funding initiatives now runs the risk of being managed as a corporate, with participating foundations taking the role of corporate board members to 'be kept happy', risking the initial intent. The author truly hopes that it can stay true to its founding principles.

Relative Merits of Investing in Institutions Rather than Individuals

At a relatively recent Grant-makers Affinity Group retreat (Johannesburg, South Africa, February 2008), there was a brief discussion on grant-making of institutions or individuals and the relative merits of each. Rather than dichotomizing the issue into institution versus individual,

it was agreed that consideration should be given to what is the most effective means of accomplishing the goal. If one desires institutional capacity building, institutions are more important because funding to individuals may not necessarily lead to institutional capacity building. Moreover, attention to individuals alone will not address the structural issues affecting universities. However, there is need for an interlocking system. Funding individuals is significant because attitudes, behavior, cultures, and traditions shape institutions. When individuals receive scholarships or funding, it changes their perspectives and subsequently affects institutional culture. However, when individuals study abroad and return to their home institutions, too often they find these institutions to be dysfunctional or limited in their capacity to support them and eventually decide to leave. Finally, the decision whether to invest in institutions or individuals is dependent on the availability of funds as investments in institutions are more sizable than in individuals.

Even as the grant makers ponder the institutions versus individuals aspects, the question does arise as to why choose? After all this group cannot be concerned about money when they have put together billions to create this body, for the common good, to pool their resources to make an effective contribution. A decision was not forthcoming; was this because of lack of foresight among the people who really care for developing Africa or was it politics?

In every situation, African institutions do, in some way, be it small, benefit from the different international partnerships and African scholars get the opportunity to study overseas. What the author cannot understand is why the failure to support the grand program or, say, 10 000 post-graduate scholars annually? Why only support a few hundred? Surely this cannot make a difference in one country, much less the whole continent of Africa. Is this the real plan to move forward Africa's economy and for reduction in poverty? Is funding the universities to rejuvenate their faculty, infrastructure, and their capability to increase their capacity to deliver more graduates to support economic growth inconsistent with real capacity development? Do the corporate companies who need the talent to develop their businesses not see the value in making a more meaningful contribution? Surely they do not want their business ventures to fail. Is the author too naive? Does the donor community not see that giving support for more African intellectuals to be home grown will enhance all the other aid programs? Is it a treat, to the funders/developed world, to help Africa to become self-sufficient?

The question is whether the schemes to develop talent are enough? In the author's view, the answer is no; there is a matrix of concurrent activities that must take place. Developing talent without an environment to use such talent and such individuals playing their part in broader society are the missing links. If we develop scholars

and the home institution has no way of attracting and competing with its international counterpart, we do not need to do any fancy research to know the result. The funding agencies need to consider the role they play in supporting talent to leave Africa and should engage in strategies to assist in the retention of talent. Without this shift in the grant-making world, be it philanthropical organizations, the business communities, or the governments, Africa will continue to be controlled remotely by those who purport to be its friends.

While the author presents a window that may reveal grant makers as indifferent to the parallel constraint, she does so with the intention that they will prove her wrong in her observations.

Further Reading

- African Development Dissertation Workshop Program: Institute of International Studies; UC Berkeley Funding Opportunities for Africans and Foreign Nationals.
- Department of Education Title SA: Pandor: Global Perspectives in Education Workshop Developing and Retaining the Next Generation of Academics in African Universities, Funders' Perspectives, Africa Grant-makers Affinity Group Retreat, Johannesburg, South Africa 7 February 2008.
- Mutume, G. (2003). Reversing Africa's 'brain drain, new initiatives tap skills of African expatriates. *Africa Recovery* 17(2), 1.
- Mutume, G. (2003). Reversing Africa's 'brain drain, new initiatives tap skills of African expatriates. *Africa Recovery* 17(2), 4.
- Pandor, N. (2007). *Global Perspectives in Education Workshop*, October 2007.
- Maphosa, T. (2008). London: VOA.

Scholar-Baller®: Student Athlete Socialization, Motivation and Academic Performance in American Society

C K Harrison, S Bukstein and J Mottley, University of Central Florida, Orlando, FL, USA

E Comeaux, University of Kentucky, Lexington, KY, USA

J Boyd, Arizona State University, Tempe, AZ, USA

C Parks, Long Beach Unified School District, Long Beach, CA, USA

D Heikkinen, Eastern Michigan University, Ypsilanti, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Movies such as *Varsity Blues*, *Friday Night Lights*, and *Coach Carter* portray the American socialization of high-school student-athletes who have aspirations of attending college with the common ultimate goal of playing professional sport. Many of the characters in these three feature films prioritize athletics over academics and have an unrealistic understanding of their chances of making it to the pros. These representations of sport and student-athlete focus in American society enable youth, teachers, coaches, and fans to examine the values that sport and society teach through athletics and other cultural organizations (Coakley, 2004).

In American higher education, a greater emphasis is frequently placed on the athletic abilities of student-athletes and the success of athletic teams, as compared to the classroom achievements of student-athletes. The primary purpose of higher education is too often overlooked, and the academic success of student-athletes is too often overshadowed. As high-school and college athletics become increasingly commercialized, secondary and post-secondary institutions are now facing the challenge of addressing an increasing lack of academic productivity among some student populations. Specifically, the student-athlete culture in higher education possesses certain sub-cultures that tend to underachieve educationally (Bowen and Levin, 2003). This issue, compounded by the National Collegiate Athletic Association (NCAA) Academic Reform Movement, requires the creation of new cultural paradigms and educational programs that have the ability to challenge student-athletes to apply a competitive spirit not only on the playing field but also in the classroom and in the development of life skills (Brand, 2003).

A lack of commercial and institutional emphasis on education still remains. This lack of focus on branding and marketing student-athlete academic success permeates into all sectors of high-school and collegiate sport and influences how student-athletes are presented in the media and perceive themselves. To solve some of the problematic issues relating to student-athlete academic performance and motivation, a group of colleagues, all of whom are former student-athletes, created a nonprofit

organization called Scholar-Baller®, which was established in 1995 as a working theory and concept. Scholar-Baller has developed and implemented culturally relevant incentive-based educational programs at the high-school and college levels to help bridge the gap between education, sport, and popular culture, to help reposition the current model of sport in American society and to help create a new mindset among student-athletes in terms of their perceptions about education, sport, and career aspirations.

The curriculum intervention programs and related initiatives developed by Scholar-Baller serve as an excellent example of a marketing platform to reposition student-athletes and to tell true and meaningful stories about their successes on and off the playing field. Scholar-Baller, which is endorsed by the NCAA, recognizes the academic achievements of student-athletes who attain a high *grade point average* (GPA) or who demonstrate significant academic improvement. For some academic institutions who have partnered with Scholar-Baller, marketing and positioning of the student-athletes takes place in part via the display of a uniform patch or a helmet sticker called ThinkMan™ or ThinkWoman™ (see **Figures 1** and **2**), which helps to draw attention to the academic performances of the student-athletes while still allowing the players to demonstrate their athletic skills – truly epitomizing a student-athlete. In addition, Scholar-Baller, in conjunction with the National Consortium for Academics and Sports, sponsors the Academic Momentum Award, which recognizes 16 student-athletes each year for their academic accomplishments. Many institutional partners of Scholar-Baller have begun to market the academic successes of their student-athletes. Images and stories relating to the academic performance of student-athletes can be found in school media guides and on websites of college athletic programs. What follows is an overview of what constitutes a Scholar-Baller as well as further discussion of the Scholar-Baller program.

Scholar-Baller

Dr. C. Keith Harrison, director of the Paul Robeson Research Center for Academic and Athletic Prowess,



Figure 1 ThinkMan™.



Figure 2 ThinkWoman™.

first introduced the term Scholar-Baller in his continued discussion and academic research on Paul Robeson's incredible commitment to education and sport (Harrison, 1995, 1996).

Scholar is a term used to describe an individual who possesses academic prowess or a commitment to education and learning. For example, Paul Robeson, a 1919 graduate

of Rutgers University, possessed the intellectual capacity and political awareness to earn the status of Phi Beta Kappa scholar, valedictorian of his class, and legendary political icon. Robeson represented a scholar on many levels, as he earned his law degree from Columbia University and gave numerous speeches throughout the world.

Baller is an urban term that resonates with individuals of all origins and has taken on global meaning in recent years. In popular culture, baller has been mainstreamed on ESPN, Music Television (MTV), in major newspapers such as USA Today, and in speeches by President Obama. The term baller can be used as a noun, adjective, or verb to signify aspects of achievement or success. For example, an individual who is considered a high achiever in any task or vocation can be labeled as a baller. Many of today's student-athletes view a baller as someone who has gone to the top, won the biggest prize, and who has simply just made it. Ask a room full of high-school or college football players, "Which one of you is a baller?" and nearly all the hands in the room reach sky high. It is a concept student-athletes understand and embrace. Paul Robeson was a baller, as he played professional football and was a renowned actor.

Scholar-Baller essentially means an educated individual who also participates in sport, art, music, or any other extracurricular activity. The Scholar-Baller concept promotes the willingness of students to accept the challenge of harmonizing academics and athletics. At a pragmatic level, Scholar-Baller is about cultivating education, sport, and entertainment consumption into one lifestyle. Any misunderstanding about the term or any uneasiness has come from coaches and administrators who do not identify with young people and what they are into. Dr. Harrison has published numerous articles and book chapters relating to how popular culture (e.g., sports and entertainment imagery) impacts the educational aspirations and perceptions of youth in society.

Scholar-Baller Paradigm: Promoting Critical Literacy Through Engaging Popular Culture with Urban Youth

The Center for the Study of Sport in Society (1997) reported that 66% of African-American male youth between the ages of 13 and 18 stated they perceived that their first career would be as a professional athlete. Another related study found that 48% of African-American inner-city youth responded that hip-hop/rap music is their favorite musical genre. Just over one out of three (34%) of the respondents in that study reported watching 4 h or more of television per day, and 62% of the respondents reported going to the movies two or more times a month. According to the study's findings, urban youth disproportionately

spent their disposable income on the following items: shoes, music, and jewelry (MEE Report, 1993).

Teachers, coaches, and administrators need to be able to interact with today's student-athletes in a culturally diverse and impactful manner that allows the students to be prepared for the next level and for life after sports. Teachers, coaches, and administrators must engage student-athletes and must re-emphasize the importance of education. The message that school is cool must get out to all student-athletes. As those at Scholar-Baller often say, it is imperative that we help "make the cool students smart and the smart students cool."

People in academic circles must know what they teach and why the material is relevant to today's student-athletes. For this to happen, it is essential for educators to identify desired results (i.e., what do we want students to know?), determine acceptable evidence (i.e., what counts as proficiency?), and plan learning and instruction (i.e., what do we do and how do we teach it?). There are a few things that allow learning to be maximized when trying to create a lifelong learner. Course content must be relevant and meaningful. Teachers must also find ways to be engaging and should utilize a variety of teaching styles that incorporate concepts from the sport and entertainment industries.

As utilized by scholars such as Taylor (1999), this concept suggests that sport and athletic principles (i.e., character, determination, perseverance, and commitment) should be applied to academics, which should improve the motivation of student-athletes in the classroom. This motivation for academic excellence is enhanced through utilizing aspects of popular culture entertainment (i.e., music, film, athletics, and the arts) to challenge and motivate student-athletes. Scholar-Baller uses popular culture as a framework to engage and motivate student-athletes. Numerous colleges and universities have implemented the Scholar-Baller curriculum intervention program, which uses DVDs, CDs, social media platforms and other relevant mass media materials that student-athletes already consume and understand.

Implications for Practice: A Collegiate-Level Case Study

Pilot implementation of the Scholar-Baller program began in August 2001 at a major NCAA Division I institution with their football program. Student-athletes on the football team were first introduced to the concept of Scholar-Baller during fall training before the 2001 season. The team was asked how many of the student-athletes considered themselves to be ballers; approximately 85 out of 105 raised their hands. When asked how many of them saw themselves as Scholar-Ballers, only a few raised their hands. The students' responses illustrate the fact that

many college student-athletes who participate in a revenue sport such as football or basketball tend to focus narrowly on their athletic prowess while trivializing their academic and social development.

However, with the introduction of the Scholar-Baller program, it was immediately evident that the concepts, messages, and approach resonated with the student-athletes, in this case those who participated in the revenue sport of football. During the weeks following the initial introduction of the Scholar-Baller program, multiple student-athlete football players approached both the head football coach and the football academic counselor and communicated sincere interest in attaining Scholar-Baller status and recognition. The football student-athletes at this institution were challenged to employ their competitive spirit in the arenas of academics and social development.

A series of strategies were utilized to establish a consistent message that subpar (or even an average performance) in academics was simply unacceptable. Some of these strategies included helping student-athletes examine their social- and self-identities to reinforce that they were complete human beings with a multiplicity of abilities beyond athletics. Student-athletes were exposed to Scholar-Baller icons (i.e., Paul Robeson, Ralph Bunche, Vince Carter, and others) via innovative lesson plans. Academic goals for the football program were displayed in the locker room alongside football goals. In addition, team members engaged in an academic team competition in which groups of football student-athletes competed against each other throughout the school year.

Additionally, the head football coach and assistant coaches embraced the Scholar-Baller concept and soon incorporated it into some of their pedagogical vernacular, including utilizing the term in memorandums and letters to the team during the off-season, encouraging academic excellence during team meetings, and including the term in the school's players manual. More importantly, an incentive and disincentive system was established that rewarded the student-athletes for high performance in the classroom. This same system parallels internally what the NCAA Academic Reform Movement is attempting to do externally with its incentive/disincentive system to increase academic achievement and graduation rates.

At this particular institution, a Scholar-Baller was defined as a person in a given academic year who earned a fall, spring, or cumulative GPA of 3.0 or above. First-year Scholar-Ballers earned a T-shirt. After completion of the academic team competition during the second year of the program, the top three academic teams earned sweat suits that lauded them as "Scholar-Ballers competing in the classroom." Needless to say, the entire Scholar-Baller program was very well received and reinforced that it really was a benefit to compete and perform well in school. Scholar-Ballers were acknowledged by the head football coach at the beginning of each semester, and at the end of

a particular academic term the Scholar-Ballers were rewarded with a steak dinner, hosted by the athletic director. In fall 2004, football student-athletes who achieved Scholar-Baller status were recognized with the ThinkMan logo jersey patch, the first known time in the history of NCAA Division I football that academic prowess had been acknowledged on the jersey of student-athletes.

In the first 3 years that the Scholar-Baller program was implemented (2001–04), there were significant increases in team-cumulative GPA, fall-team GPA, freshman-cohort GPA, and retention rates of student-athletes. For example, retention rates skyrocketed to 80% for the three classes that entered during the first 3 years of the program. The number of football student-athletes earning a GPA of 3.0 rose from figures in the teens in 2000 to 38 in 2004. Another shift occurred with decreases in the number of football student-athletes on probation, which went from five after spring semester in 2000 to zero in 2004. Finally, 33 student-athletes earned B or better averages during the academic semester that coincided with the 2004 football season (see Steinbach, 2004).

Student-athletes in this football program understood from day 1 that they had entered a culture in which it was not acceptable to do the minimum work required to pass classes and remain eligible. Academic excellence became the verbal and written goal of most of the student-athletes. The following are some quotes and narratives from 2004 Scholar-Baller participants (28 football student-athletes) who were interviewed about becoming the first NCAA Division I school to acknowledge academic success on the front of a football jersey:

- It feels good because of the stereotypes that have been set that football players can't be smart but society always changes. Happy to represent [support the program's image and concept] by wearing the patch.
- It feels good to be recognized for something positive. Many people think of football players as dumb jocks so this was a way to prove that they were wrong.
- It gave me something to push for. I feel it is a great honor to wear this patch. The patch is not about separating yourself from others, yet its purpose is to glorify those who work hard in all phases of life. This is a motivation for those who do not have one.

Implications for Practice at the High-School Level

As implied by Scholar-Baller, it is assumed that this paradigm will serve as a resource for academic and athletic success for all student-athletes at any educational institution across the nation. Because this model is currently most widely utilized at the intercollegiate level and has

been proven to be successful, it can be assumed that it would be effective if implemented on a wide-scale level at some of the nation's high schools that are most in need of Scholar-Baller's program and message. However, the model would have to be flexible enough to fit the needs of any interscholastic program in which implementation would take place.

There are many things to consider related to the implementation of the Scholar-Baller program at the high-school level, many of which have already been discussed above. Other aspects to be considered in the implementation process of Scholar-Baller would include the socioeconomic status, ethnic background, demographic region, and overall morale of the students at the particular high school in which implementation of the Scholar-Baller would take place.

The Scholar-Baller implementation process begins with high-school coaches, administrators, counselors, and teachers involved in a life-skills curriculum designed to challenge student-athletes to create powerful visions for their future. The typical curriculum covers content standards such as: (1) self-identity, (2) the competitive spirit, (3) the Scholar-Baller paradigm, (4) purpose/vision/mission and goals, (5) decision-making system, and (6) living the Scholar-Baller way. The goal of Scholar-Baller is to create an environment on high-school campuses that will allow student-athletes as well as other students to view education and athletics as means to success. The Scholar-Baller program empowers student-athletes to be better prepared for the rigors of college and the many challenges that college provides; GPAs will likely continue to increase, and parents will be empowered to support their children not only athletically but also academically. Finally, Scholar-Baller should also help students develop critical life skills and should equip student-athletes with the ability to make positive future decisions relating to career and other aspects of life.

The Scholar-Baller Patch and Related Incentives: The New Badge of Coolness

As mentioned above, the Motivational Entertainment Educational (MEE) Report (1992/93) found that 387 urban youth, ages 13 to 18, spent most of their disposable income on clothes, food, music, shoes, and jewelry. This finding has significant implications. First, urban youth set the trends for popular culture, and hip-hop culture is currently one of the most consumed products and commodities in American society (Boyd, 2004). Specifically, marketing, fashion, language, and urban styles permeate the most mainstream cultural spaces (Simmons, 2002). In addition, athletic identity is a valued status symbol in secondary and post-secondary education. Thus, the desire to be cool often involves the attainment of some of the

material items found in the MEE Report. Through the previously mentioned cultural facts about urban styles and athletics in secondary and post-secondary education, we hope that the Scholar-Baller, ThinkMan, and ThinkWoman logos (see **Figures 1 and 2**) will become a tipping point and trend (Gladwell, 2000) in American culture. This is possible if Scholar-Baller continues to become aligned, associated with, and embedded in social activities and academic programs that today's student-athletes value, understand, and embrace.

The essence of the primary Scholar-Baller logos, such as the ThinkMan and ThinkWoman images, are of a student-athlete sitting at a desk studying – representations that are powerful in their simplicity, with or without additional wording.

These logos, when applied to a piece of apparel, are meant to be a badge of honor that can be proudly worn, a statement of the acceptance of the Scholar-Baller ideals and the related accomplishments made by each individual. The Scholar-Baller logos and their related marks and graphics will continue to evolve with recognition, acceptance, specific individual or team needs, and style trends.

To date, ThinkMan and ThinkWoman have adorned team jerseys, football helmets, polo shirts, sweatshirts, sweat suits, shorts, hats, shoes, coffee cups, duffel bags, dress shirts, visors, and more. For instance, Arizona State University became the first NCAA Division I football team to recognize academic achievement during athletic competition on 2 September 2004. Hampton University and Morgan State University also sported the patch during the New York Urban League classic, becoming the first of the historically black colleges and universities to participate in the Scholar-Baller uniform-patch recognition program. Coincidentally, during the 2004 Vitalis Sun Bowl, Arizona State University and Purdue University became the first bowl teams to recognize academic achievements on student-athlete jerseys based on the Scholar-Baller concept. Finally, one of the historical moments of the NCAA and higher education occurred on 10 November 2006 when Southeast Missouri State University and Tennessee State University became the first two institutions with the same symbol/patch to honor student-athletes for their academic success immediately before an athletic contest; what made this recognition ceremony unique is that both teams interacted with the administration, staff, and the Ohio Valley Conference Commissioner at mid-field.

Much more is planned for the Scholar-Baller logo set, including continued evolution of the base trademarked logos, creation of complementary logos such as the emerging It's Cool to be Smart, AthletesThink, and ThinkKids logos, and development of additional recognition and motivational items. Not only do we want the Scholar-Baller ideals and related curriculum to be

embraced by many, but we also want those that are deemed to be Scholar-Ballers to be able to proudly express this recognition in a variety of ways.

Conclusion and Implications

For over a century, improving the academic success of student-athletes has been a complex challenge for high-school and college educators and administrators. There are major cultural disconnects between the educational, athletic, and entertainment communities in America. This gap and lack of synergy between the three areas impacts the lifelong learning and consumption patterns of student-athletes. Two obvious questions are (1) when did this disconnect begin? and (2) how do we connect education, sport, and entertainment in a constructive way so that student-athletes become intellectually engaged and dedicated to lifelong learning? The disconnect and gap between academics, athletics, and popular culture are poignantly articulated by Coleman (1960) in three powerful narratives. First, speaking of adolescence, Coleman said,

What our society has done is set apart, in an institution of their own, adolescents for whom home is little more than a dormitory and whose world is made up of activities peculiar to their fellows. They have been given as well many of the instruments which can make them a functioning community: cars, freedom in dating, continual contact with the opposite sex, money, and entertainment, like popular music and movies, designed especially for them. The international spread of rock-and-roll and of so-called American patterns of adolescent behavior is a consequence, I would suggest, of these economic challenges which have set adolescents off in a world of their own. (p. 338)

This world of their own often consumes American youth in nearly all cultural forms, except in educational engagement and intellectual development. Student-athletes are especially influenced by and susceptible to the bombardment of messages about material gain based on the athletic and entertainment identity of professional athletes, actors, and musical artists (Gerdy, 1997). What are the effects of a cultural system that reinforces athleticism and not intellectualism? The second narrative by Coleman is key to understanding and responding to this question.

Coleman (1960) continues to frame the incentive and reward system of American education and society by having a vision that

the fundamental change which must occur is to shift the focus: to mold social communities as communities, so that the norms of the communities themselves reinforce educational goals rather than inhibit them, as is at the present case. (p. 338)

Presently, the social communities have learned to consume athletics, material objects, and immediate gratification at such an influential rate that education is overlooked and neglected as a viable option for success (Harrison, 2002). In a Coleman culture the movies, music, video games, and athletic contests would compliment the pedagogy of school systems with competency in reading, writing, and arithmetic (commonly known as the three Rs). This method could easily enhance educational goals instead of inhibiting them – hence, a method that inspires youth to desire learning and intellectual development throughout their lifetime. This leads to the third and final narrative by Coleman.

In the final analysis, Coleman (1960) exposes the bias in American sport and entertainment by indicating the cultural fact that “the outstanding student has little or no way to bring glory to his school” (p. 347) (in comparison to athletics). This is where the Scholar-Baller paradigm may theoretically and practically influence the culture of American sport by ending the silence about the importance of student-athlete academic success. This is an approach we hope will become the new roadmap for success in American higher education. We must continue to inspire youth to excel in education and life by making the most of their cultural interests in sport and entertainment. Having visual evidence such as the ThinkMan logo visible on the uniform and discussions of Scholar-Ballers in print stories, on the Internet, and on television are essential in repositioning high-school and college athletes as student-athletes and ultimately establishing that image as the norm rather than the exception. More middle schools, high schools, and colleges need to implement culturally relevant programs that incentivize and motivate student-athletes to balance playbooks with textbooks.

Postscript: Coach Carter Film Screening

On 13 January 2005, the triangle of success (education, sport, and entertainment) celebrated a collaborative moment of combining influential cultural worlds. Scholar-Baller® and its team of educators, Rush Philanthropic led by Russell Simmons, the *Women's National Basketball Association*

(WNBA) represented by Diana Taurasi, and the NCAA represented by Senior Vice President, Bernard Franklin, all partnered together for a film screening of *Coach Carter* for over 300 urban and inner-city youth in New York City. Educational and critical dialog followed, with a panel and group discussion about the influence of sport and entertainment on the educational aspirations of America's youth.

Bibliography

- Bowen, W. and Levin, S. (2003). *Reclaiming the Game: Sports and Our Educational Values*. Princeton, NJ: Princeton University Press.
- Boyd, T. (2004). *Young, Black, Rich and Famous*. New York: New York University Press.
- Brand, M. (2003). Sustaining the collegiate model of athletics. *Paper Presented to the NCAA Membership*. <http://symposium.tulane.edu/keynote.txt>.
- Coakley, J. (2004). *Sport in Society*. Denver, CO: McGraw Publishers.
- Coleman, (1960). The adolescent subculture and academic achievement. *American Journal of Sociology* 4(35), 337–347.
- Gerdy, J. (1997). *The New Standard*. Phoenix, AZ: Onyx Press.
- Gladwell, M. (2000). *The "Tipping Point": How Little Things Make a Big Difference*. New York: Back-Bay Books/Little, Brown.
- Harrison, C. K. (1995). African American male student-athletes. *Paper Presented at the North American Society for the Sociology of Sport Annual Meeting*, Sacramento, CA.
- Harrison, C. K. (1996). Time capsule: Perceptions of Paul Robeson, Arthur Ashe, and Grant Hill. *Paper Presented at North American Society for Sport History*, Auburn, AL.
- Harrison, C. K. (2002). Scholar or baller: A visual elicitation and qualitative assessment of the student-athlete mindset. *National Association of Student Affairs Professionals Journal* 8(1), 66–81.
- Motivational Educational Entertainment Report (1993). Philadelphia, Pennsylvania.
- Steinbach, P. (2004). Patch work: “Scholar Baller” logos on ASU football jerseys, which honor B students, may start a trend. *Athletic Business* 28(11), 28–32.
- Taylor, E. (1999). Bring in da noise: Race, sport and the role of schools. *Educational Leadership* 56, 75–78.

Further Reading

- Harrison, C. K. (2007). “Drop it like it's hot”: Releasing Scholar-Baller voices and changing the “character of the discourse” in North America's post-secondary institutions. *Journal for the Study of Sports and Athletes in Education*.
- Simmons, R. and George, N. (2002). *Life and Def*. New York: Three Rivers Press.

Service-Learning and Diversity as Mediums to Foster Civic and Community Engagement in Undergraduates

L Oseguera, University of California, Irvine, Irvine, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Campus Compact – A national coalition of more than 1100 college and university presidents, who are committed to fulfilling the civic purposes of higher education. Campus Compact provides leadership, resources, and advocacy to support civic education, community building, and campus engagement.

Modern racism scale – The concept of modern racism has also been termed symbolic racism. Symbolic racism is the expression in terms of abstract ideological symbols and symbolic behaviors of feeling that blacks (and other racial/ethnic minority groups) are violating cherished values and making illegitimate demands for changes in the racial *status quo*. The label was changed from symbolic racism to modern racism in order to emphasize the contemporary, post civil rights movement nature of the beliefs and issues.

Pluralistic orientation – The ability to see multiple perspectives; ability to work cooperatively with diverse people; ability to discuss and negotiate controversial issues; openness to having one's views challenged; and tolerance of others with different beliefs.

Project pericles – A not-for-profit organization that encourages and facilitates commitments by colleges and universities to include education for social responsibility and participatory citizenship as an essential part of their educational programs, in the classroom, on the campus, and in the community.

Background and Introduction

There is a common concern among scholars, public officials, and the general public that young adults today are apathetic, uninvolved, and dispassionate. Public perception that college students are increasingly becoming more 'me oriented', has coincided with a general discontent with how colleges and universities are addressing (or failing to address) this concern. Educating, training, and preparing future leaders for many segments of society continue to be the responsibilities of colleges and universities. Hurtado

et al. (2002) report that at college entry students are not exhibiting certain characteristics and understandings, fundamental to life in a diverse democracy (e.g., views about the importance of engaging in social action activities). Thus, the college experience may be a critical juncture in the development of young people into actively engaged citizens, leaving higher education institutions with a gravely important responsibility. In recent years, higher education has been faced with great scrutiny, as institutions have been challenged to be more socially responsible and to produce more socially responsible graduates. Indeed, as the economic and social importance of higher education becomes greater, the public is more critical of what universities are doing and how they are doing it.

Since the 1990s, there has been an increased interest in college students' community and civic involvement and their attitudes toward members of other races and ethnicities. In educating a more diverse student body, colleges have undertaken numerous initiatives to increase student engagement with social diversity in order to achieve desired outcomes. What is known is that meaningful opportunities for diversity exposure and a variety of civic engagement activities/experiences produce individuals who are better able to function in a diverse and global society. Salient approaches to equipping students with the faculty to transform society include civic and community engagement, civic goals and value development, and social issues awareness or democratic sensibilities. The two main areas that researchers focus on to better develop these democratic competencies are diversity related, ranging from diversity coursework to increasing structural representation of diverse groups on college campuses to structured programs on campus created to develop more socially responsible citizens. Examples of these structured programs include the provision of service learning and other community-based experiential learning opportunities. Specifically, the aim is to review the relationship between service learning, multicultural competencies, and subsequent civic engagement and is framed by a larger interest in the role of educational institutions in preparing students to assume the responsibilities of citizenship in a democratic society. Civic engagement is essential to foster in students because it allows them to use their knowledge and values to solve societal problems. Thus, utilizing current research in the sociology of education, this article provides a synthesis of the existing literature on US colleges and universities' roles in producing a civically engaged citizenry who can function

in a diverse, pluralistic society highlighting research that focuses on systematic attention to educating students for such a society.

Diversity Exposure and Preparation for a Diverse Democracy

Racial and ethnic diversity research has received increased attention in the last 20 years. Research on racial and ethnic diversity has been concerned with how diversity experiences affect the quality of the undergraduate experience and, in particular, the extent to which exposure to diversity alters students' appreciation and understanding of different races, ethnicities, and cultures. There is ample evidence of the educational benefits of campus diversity for students and society. In a report filed on behalf of the University of Michigan, Gurin (1999) provided a comprehensive explanation of how diversity enables educational institutions to fulfill their mission by improving student learning, instilling civic virtues, and promoting democratic ideals (Jayakumar, 2007).

Although the 2003 US Supreme Court's ruling in *Gratz v. Bollinger* struck down the method the University of Michigan had used to achieve a diverse undergraduate student body, it did support the educative value of diversity. The court not only affirmed the importance of diversity in higher education, but also reminded us that higher education institutions have a responsibility to train their students to become leaders across a diverse, pluralistic society and that these future leaders should also be diverse.

Thus far, diversity research has been divided into three major strands. These strands are often presented as distinct categories, yet they are best understood in relation to one another. Jayakumar (2007) succinctly organizes and reports these as:

Structural diversity or numerical representation of students of Color as the starting point from which benefits emerge. However, increasing the enrollment of students of Color will not, in and of itself, lead to positive outcomes. A second dimension of diversity is the extent and quality of one's engagement with people of different racial/ethnic backgrounds, called both *interactional diversity* (Gurin *et al.*, 2002) and *cross-racial interaction* (Chang, 1999) in the extant literature. The third component encompasses formal exposure to diverse peoples through curricular and cocurricular offerings, or *classroom diversity*, also referred to as *curricular diversity* (Gurin *et al.*, 2002; Jayakumar, 2007: 22)

This section thus summarizes the literature on the benefits of diversity experiences across the three aforementioned dimensions on students' democratic skills including civic responsibility.

Researchers have found that the structural diversity of a campus is positively related to whether students will socialize with someone of a different race and develop greater tolerance and awareness of different racial/ethnic groups. Although institutional-type variables (e.g., mission, size, selectivity, etc.) show little relation to changes in racial/ethnic attitudes, peer, faculty, and administrator support of diversity issues on campus do lead to students' reports of lower racial tension on campus. Another manner in which the higher education institution can enable change in student attitudes is through attention to the peer and faculty context of the institution. The more the student and the faculty body deem service important and exhibit service commitments and action, the positive net benefits that accrue to the individual student as well. The evidence is rather conclusive, with a few exceptions, that typical characteristics that differentiate higher education institutions (e.g., mission, size, type, selectivity, etc.) do not predict stronger civic competencies reminding the reader that all higher education institutions can and do aid in the development of socially responsible citizens.

Providing opportunities for cross-racial interaction (i.e., dining, classroom, friendships, and studying) also results in enhanced leadership skills, cultural awareness, and civic interest. Several forms of classroom diversity have been evaluated including classroom racial composition, classroom interactions and dialog, and content of the curriculum. The effects of classroom diversity have also been investigated across multiple institutions, particular classrooms within an institution, and single institutional-wide programmatic efforts. Regardless of the approach or the context, all of these studies unanimously support the value of classroom diversity on students' multicultural competencies. In most cases, classroom diversity alone does not appear to maximize the benefits of campus diversity with regard to civic and democracy-related skills unless attention is also paid to the quality of the intergroup relations as that has a deeper impact on democratic outcomes than classroom diversity. Further, in her meta-review of the literature on classroom diversity, Jayakumar (2007) concluded that structural diversity in the classroom, much like at the institutional level, is most beneficial when accompanied by efforts to make the curriculum relevant for all students.

Other campus-facilitated activities to bring awareness and action to social issues and diversity include diversity-oriented workshops. Benefits accrue not only for students of color but also for white students. Additional programmatic factors on campus that impact multicultural competence include curriculum offerings. Analyzing the relationship between freshman to senior-year changes in a student's social activism orientation, Astin (1993) found that enrollment in ethnic and women's studies courses, discussions of racial/ethnic issues with other students, and participation in campus-based racial or cultural awareness

workshops increased students' orientation to social activism and concluded that this was not due to social change or maturation. Who teaches the course may also impact the intended outcomes. In other words, how the instruction is delivered and by whom may be just as important as what the curriculum is. Researchers have concluded that instruction of diversity-related issues by female instructors and instructors of color produce more meaningful changes in students' attitudes.

Civic Engagement and the Development of Democratic Competencies

In addition to the positive benefits of diversity opportunities on civic competencies, the next broad area attended to includes specific engaged learning opportunities that facilitate the development of civic competencies. The emphasis of this section is on how various forms of service experiences lead to the preparation and development of global citizens. The section begins by operationalizing what falls under the term civic engagement and then follows with existing research in the area. Typically, the research in the area of civic engagement practices examines everything from moral development, to cognitive development to understanding of difference. Because this article's aim is to examine higher education's role in developing socially responsible citizens by way of cross-cultural understanding, the research on civic engagement will necessarily be limited to work that examines students' experiences, attitudes, and goals toward pluralistic orientation and democratic ideals and work related to diverse groups and perspectives.

Specifically, experiences that campuses utilize to engender social responsibility and social awareness in students are reviewed here. The previous section presented diversity-related experiences, broadly speaking, while this section presents civic engagement practices such as service learning and volunteerism and campuses' roles in facilitating these opportunities for undergraduates to prepare them to live and work in a diverse society. The primary interest is on civic engagement as an outcome with service or volunteering as the actual process. These are different practices but often used interchangeably in the literature, thus leading to confusion. In the higher education literature, the term civic engagement encompasses a wide range of activities, motivations, and commitments. Traditionally, civic engagement referenced political engagement (such as voting in an election or running for political office); however, today this term has been expanded to include other types of social engagement such as protests, community advocacy, and service learning. Below is a rubric adapted from Astin *et al.* (2006) who identify multiple categories of engagement. This work is not meant to defend what constitutes civic engagement but rather it includes a synthesis

of work that also employs these varied definitions. Again, the ultimate aim is concerned with how college service and volunteer experiences affect students' civic engagement and simultaneously undergraduates' readiness for participation in a diverse society. This could range from out-of-class activities to volunteering to internships.

Astin *et al.* (2006) conceptualize engagement into three broad categories:

1. *Political Engagement.* This includes voting behavior, commitment to political action, and political activism and expression. This article does not focus on the traditional political science view of civic engagement, but it is rather expanded to include other types of service activities that lead to developing competencies for successful participation in a democracy. What we do know about young adults is that they are less likely to vote in elections and possess less political knowledge than older citizens. In fact, when compared to older adults, young adults are less attentive to news, less likely to volunteer for political organizations, and are less likely to contribute money to various causes.
2. *Civic and community engagement.* This includes forms of service learning and community involvement including volunteerism as well as giving of time and money to political and social causes. Service learning is viewed as one form of engaged learning that can enhance the skills and motivation necessary for civic engagement.
3. *Civic goals and values.* This includes pluralistic orientation, self-efficacy, and developing racial understanding.

This section narrows its focus to include how service practices contribute to undergraduates' preparation for and readiness for a diverse and global society.

Young Adults and Civic Participation

Generally, what do we know about civic engagement and what do today's US college students look like? Citizens in general are becoming less involved in community efforts, less likely to take on leadership positions and less likely to participate in traditional political activities. Previous research finds that the period of young adulthood is one of less engagement in civic and voluntary endeavors, compared to other stages in one's life. However, evidence suggests that young adults tend to participate in nonpolitical volunteer work at rates close to those of older adults. In fact, today's college students are increasingly participating in alternative and new forms of engagement, such as buy-cotting (intentionally purchasing a product because of the values of the company), or utilizing media, such as Facebook, to engender social or political consciousness and awareness. Historically, educational institutions have played a fundamental role in cultivating an educated citizenry in a democratic society. In general, the research is unanimous that with increased educational levels so do the

rates of civic participation and many educators and policy-makers believe that civic participation, in particular, is best fostered and increased through student engagement in service learning. Service learning includes a structured service component tied to classroom content. Included in this section is also attention to generic service and volunteering activities as research also demonstrates the positive benefits that accrue even if a formalized classroom component is not attached.

Relationship to Student Attitudes, Values, and Behaviors (During College Years)

This section offers a brief overview of what we have learned about service participation generally, and then it follows with specific attention to service participation's contribution to diverse perspectives and sensibilities. The research evidence is overwhelming regardless if the study was conducted using large, national, databases, single-institution-wide evidence, or small-scale surveys. Further, these results are consistent when researchers use established scales or rely on student self-reports. Finally, the effects of service participation are evidenced over a variety of time measures. That is, whether the outcome was measured during the course, after the semester, at the end of the year, or postcollege. Students who perform service learning show increases in their plans for future civic action, assessments of their own interpersonal, problem solving and leadership skills, and agreement with items emphasizing society factors that affect individual outcomes (i.e., social justice).

Based on a variety of civic participation scales, students score significantly higher on civic participation scales after engaging in a service-learning course and state that service provides them an opportunity for understanding the communities they serve more thoroughly. Over the course of the service-learning experience, students become advocates of community service, become personally and emotionally engaged, learn about their positions of privilege in society, recognize the assumptions that they bring to the class, and strengthen feelings of social responsibility. These findings are confirmed even after controlling for freshmen-year pretests, service propensity, academic major, race, ethnicity, gender, and structural characteristics of the institution. Overwhelmingly, outcomes are favorably influenced by service participation.

In terms of enhancing student learning, the effects of service learning and community service on college undergraduates' cognitive and affective development have also been explored, and it has been found that service participation shows significant positive effects on multiple outcome measures with service participation having its strongest effect on students' decisions to pursue service careers. Most students believe their service made a difference. However, students' degree of interest in the

subject matter was the most important factor associated with positive experiences. Further, both faculty and students develop a heightened sense of civic responsibility as a result of their service experience.

Civic Engagement and Multicultural Competence

We move forward with civic participation and attitudes toward diversity and creating democratic citizens. Students engaged in service work also show increases in their civic responsibility defined as commitment to the life goals of helping others and promoting racial understanding and increases in their life skills which include critical thinking, interpersonal skills, leadership skills, social self-confidence, knowledge of different races or cultures, and conflict resolution skills. Not surprisingly, the more time devoted to service the more positive the effect on students.

Service learning affected students in their awareness and involvement in the community and sensitivity to diversity. Other outcomes related to civic participation that contribute to diversity awareness include an awareness of personal philosophy regarding racial issues and concerns regarding specific multicultural or race-related incidents. Engaging in service also showed reductions in contemporary racism attitudes measured using the Modern Racism scale. Neither political orientation and gender nor race were predictors of change on the Modern Racism scale, suggesting that the learning component of service is more predictive than background attributes and characteristics. Thus, US colleges and universities can structure opportunities for the development of civic engagement regardless of gender or social-class backgrounds.

In addition to the positive benefits that accrue to the students who engage in service participation, positive benefits also accrue for the community agencies and the faculty involved in instructing these courses. Researchers stress the importance of understanding not only the influence of students but also of the agencies in which students serve out their roles. The impact on community agencies was evident in that they perceived an effect on their capacity to serve clients, received economic and social benefits, and were satisfied with student interactions. Finally, faculty members felt that community service experiences could be fertile ground for research and other scholarly work.

This section and much of the literature on civic engagement evaluate within and during college changes. Less is known about the long-term effects. Clearly, sustained skills are important as we expect undergraduates to continue their attention to serving communities. The next section highlights our knowledge of the long-term effects.

Impact on Student Attitudes, Academic Skills, and Behaviors (Long-Term Effects)

Astin *et al.* (1999) show that students who spent 6 or more hours a week in volunteering during their last year of college were almost twice as likely to perform volunteer work after college. Nine years after graduating, the frequency of volunteering still correlated with the amount of volunteering during high school. Volunteering was associated with the following 13 behavioral and value outcomes: attending graduate school, donating money to the undergraduate college, frequency of socializing with diverse people, helping others in difficulty, developing a meaningful life philosophy, promoting racial understanding, participating in community action programs, participating in environmental cleanup, a sense of efficacy, highest degree earned, hours spent volunteering, career preparation, and degree aspirations.

In a recent longitudinal study examining the effects of college service participation on young adults' service participation 4 years out of college, Astin *et al.* (2006) show that participating in service-learning courses and generic service during college appears to have positive effects on postcollege outcomes including working with communities, volunteering, and commitment to political/social change. Unique positive effects of service learning (over and above the effect of generic service) are associated with three postcollege outcomes: civic leadership, charitable giving, and overall political engagement.

Campus Initiatives to Increase Engagement: Why/How Can We Make Civic Engagement More Meaningful

We know service matters, but how do students come to understand diversity in the context of service learning. Findings suggest that it is the importance of relationships that are developed initially through finding common ground and then strengthened as efficacy is enhanced and empathy and compassion are nurtured. A key factor in whether or not service learning has any impact on students appears to be the quality of the experience. The design and quality of service-learning programs play a critical role in determining the effectiveness of these programs. There is overwhelming evidence that service learning and community service influence students' democratic and multiracial competencies. The next section provides research rationales for how and why service and other forms of diversity-related experiences matter. Several factors shape the extent of learning outcomes. The consensus reached is summarized below. First, is the quality of placement or the sense of what the student obtained from their experience as well as the challenge and variety of work performed. Second, the effects of both service learning and generic community service appear

to be mediated in part by the use of reflection. Reflection can take the form of journaling, writing essays, and engaging in formal and informal discussions with peers and faculty. Third, the number of hours spent in service. The more time devoted to service, the more positive the effect on students. Fourth, the degree to which course objectives and content aligned with actual service behavior. In problem-solving interviews, students had the chance to demonstrate their analysis of a social problem linked to their service. Over the course of a semester, students in service-learning classes in which service and academic study were continuously and closely linked showed significantly more change in the complexity of their problem analysis, their assessment of the locus of the problem and solution, and in their critical thinking ability than did students in programs with little linkage between the service option and the course of study or students with no service options. Students in the well-integrated service-learning courses were also more likely to apply subject-matter knowledge to their problem analysis and to have well-developed practical strategies for community action. In both the survey and the single interviews, students reported greater learning when they had higher quality experiences (Eyler and Giles, 1999). Fifth, having diverse perspectives resulted in greater learning for students. Finally, attention to the preparation students receive is important in enabling them to fulfill their service opportunities as the most beneficial learning outcomes occur when students are adequately trained for service.

A small body of evidence finds that mandatory volunteerism can reduce the impact of the positive effects. For this reason, campuses need to stress upfront the learning outcomes and what the expectations are for students to become engaged.

Recommendations for Campuses to Engage More Students

Since we know that students who are more engaged also report greater civic and pluralistic competencies, it is important to offer recommendations for how campuses can create or bolster environments to engage more students.

The content, skills, pedagogy, and structure that should guide the design of service-learning curricula for multicultural citizenship education are highlighted below. For service learning to teach youth about their responsibilities as citizens in a democratic society, content and strategies must model and support democratic principles. Battistoni (1997) and others offer some recommendations of ways to institutionalize and prioritize service and diversity on campus.

1. *Involve faculty.* Faculty members are important because they teach the courses that can prepare students for

- civic roles. Thus, it is important for faculty members to connect their work to community issues through the application of their research. Checkoway (2001) stresses the importance of faculty involvement and of diversity in renewing the institution's civic mission. Faculty should establish collaboration with students to allow the sharing of ideas that may lead to establishing new methods for solving problems. By working in groups, students develop the ability to work with a group to "analyze a problem, plan and implement a strategy, and work with others to evaluate results" (Schneider, 2001).
2. *Increase institutional capacity.* Capacity can be increased by creating collaborations with communities. Service-learning programs, if planned and executed correctly, benefit both the community and the institution. By creating learning partnerships both communities and universities benefit: communities are provided with consultation and students are given the ability to engage in the real world and learn from practice. Service learning serves to meet both the objectives of class courses and community needs. Campus-community partnerships are conceptualized as relationships that must maintain effective communication and a clear purpose. In order for campuses and communities to develop partnerships that benefit both students and communities, both must develop and maintain healthy relationships through a clear mission, compatibility, effective communication, and skilled staff.
 3. *Connect democracy and diversity.* Diversity must be recognized and implemented in education so that students can learn to respect and work with a diverse America. Students should interact with diverse groups, as diversity provides them with the ability to understand, communicate, and collaborate with multicultural groups. This can be accomplished by engaging students in research projects that address important issues.
 4. *Offer service-learning curricula.* Students should be involved in service-learning courses in which they serve the community and learn from experience. Institutions should also create academic courses/programs that discuss such issues. By adding new courses (i.e., cultural studies, gay and lesbian studies, and peace studies), institutions can engage students to think about issues that are prevalent in society. Hands-on programs that are combined with academic courses can establish connections between scholarly and applied knowledge. Who teaches the course may also impact the intended outcomes. In other words, how the instruction is delivered and by whom may be just as important as what the curriculum is.
 5. *Provide cocurricular service opportunities.* Students should be involved in cocurricular activities with a strong civic purpose. Integrating service into student clubs and organizations reinforces the importance of service to others.
 6. *Encourage residential living.* The available evidence also recommends residential living. The size of the impact varies but the evidence does suggest that exposure to 'others' influences understanding of others and openness to diversity. The most dramatic effects not surprisingly occur in students who live in residence halls designed specifically to address social issues and themes related to multicultural understanding.
 7. *Establish a permanent institute for education for democracy.* This is a forum for research dissemination and learning from other campuses and organizations.
- Another important component to civic engagement to consider is the relationship to policy making and understanding. The organization Campus Compact has a goal of involving students in various agencies so that they can understand "what it takes to be successful in influencing policy." Campus Compact integrates academics and hands-on experiences with the hope of creating civically engaged students. Campuses that do not collaborate with a national organization have helped students create strong relationships with one or two organizations in their local community. By emphasizing student volunteerism, campuses hope that students will help the social conditions of the communities while simultaneously gaining a better understanding of society. Caputo (2005) also argues that faculty and administrators must support the civic engagement efforts of their institutions. Project Pericles stresses civic engagement on the part of an institution's board of trustees and faculty. By engaging the institution as a whole, students can have easier access to activities and projects on campus and in their communities that lead them to develop their civic engagement. Regardless of the approach used, institutions should expect certain goals. These goals include: (1) increasing the number of individuals who are engaged in policymaking and debates; (2) ensuring that students have a clear understanding of the effects of their participation in policy making; (3) as students learn about and participate in their communities they should want to become civically engaged later in life; and (4) increasing civic engagement should lead to the development of stronger public policies which in turn should reinforce engagement and democracy. In order to achieve these goals Caputo (2005) lists the following recommendations: institutions should emphasize civic engagement from the first year through the last year, students should receive more applied experience (i.e., through an internship), and students should be allowed to participate in group projects so that he or she can acquire leadership skills.
- The development of successful partnerships is an effort of community leaders, students, and faculty. If the goal of the university is to institutionalize service learning, the following are additional areas to consider: creation of a service-learning committee; development of clear mission

and goals for on-campus service learning; educating faculty members regarding the benefits of service learning; generation of enthusiasm among student body; educating the community and community agencies about the potential benefits of service learning; and continued outreach to all areas of campus and community by distributing information regarding current service-learning projects (Merger and Brungardt, 2007).

Conclusion

This article can help to revitalize higher education's mission in preparing a diverse student body, merge research and practice in the effort of meeting the campus needs, and assist campuses in overcoming a student culture of disengagement to cultivate citizenship for the future. Institutional research can play a pivotal role in coordinating the curricular and cocurricular activities that enhance the campus-wide efforts to improve students' preparation for a diverse democracy. For example, institutional researchers can examine student enrollment, funding, and diversity of academic programs to determine the institution's readiness to prepare students for a pluralistic society. As our national community diversifies, institutional researchers must create partnerships with administrators and faculty to assure that students can emerge as active participants in a diverse democracy.

The research evidence reviewed reaffirms the role of service learning and diversity in fostering civic engagement because it encompasses learning and teaching, leads to additional forms of civic engagement, and involves students in meaningful activities that address common issues.

As Boyer (1997) remarked, scholarship of engagement has meaning at two levels: (1) connecting the university's rich resources to the most pressing social, civic, and ethical problems, making it the staging ground for action; and (2) creating a climate in which academic and civic cultures communicate more continuously and creatively, enlarging the universe of human discourse and enriching the quality of life for all.

Bibliography

- Astin, A. W. (1993). *What Matters in College: Four Years Revisited*. San Francisco, CA: Jossey-Bass.
- Astin, A. W., Sax, L. J., and Avalos, J. (1999). Long term effects of volunteerism during the undergraduate years. *Review of Higher Education* 22(2), 187–202.
- Astin, A. W., Vogelgesang, L., Misa, K., et al. (2006). *Understanding the Effects of Service-Learning: A Study of Students and Faculty*. Los Angeles, CA: Higher Education Research Institute, UCLA.
- Battistoni, R. M. (1997). Service learning and democratic citizenship. *Theory into Practice* 36, 150–156.
- Boyer, E. L. (1997). The scholarship of engagement. *Journal of Public Service and Outreach* 1(1), 11–20.

- Caputo, D. A. (2005). A campus view: Civic engagement and the higher education community. *National Civic Review* 3–9.
- Chang, M. J. (1999). Does racial diversity matter? The educational impact of a racially diverse undergraduate population. *Journal of College Student Development* 40, 377–394.
- Checkoway, B. (2001). Renewing the civic mission of the American research university. *Journal of Higher Education* 72, 125–147.
- Eyler, J. S. and Giles, D. E., Jr. (1999). *Where's the Learning in Service-Learning?* San Francisco, CA: Jossey-Bass.
- Gurin, P. (1999). Expert Report of Patricia Gurin, in the compelling need for diversity in higher education. Gratz et al. v. Bollinger et al., No. 97-75321 (E.D. Mich.) Grutter et al. v. Bollinger et al., No. 97-75928 (E.D. Mich.). Ann Arbor, MI: University of Michigan.
- Gurin, P., Dey, E. L., Hurtado, S., and Gurin, G. (2002). Diversity in higher education: Theory and impact on educational outcomes. *Harvard Educational Review* 72(3), 330–366.
- Hurtado, S., Engberg, M., Ponjuan, L., and Landreman, L. (2002). Students' precollege preparation for participation in a diverse democracy. *Research in Higher Education* 43, 163–186.
- Jayakumar, U. M. (2007). *Can higher Education meet the Needs of an Increasingly Diverse Society and Global Marketplace? Campus Diversity and Cross-Cultural Workforce Competencies*. Unpublished dissertation. Los Angeles: University of California.
- Merger, D. K. and Brungardt, C. (2007). Case study: Institutionalizing service-learning at Fort Hays University. *National Civic Review Summer*, 52–54.
- Schneider, C. G. (2001). Toward the engaged academy. *Liberal Education* 87, 18–27.

Further Reading

- Allen, W., Bonous-Hammarth, M., and Teranishi, R. (2006). *Higher Education in a Global Society: Achieving Diversity, Equity, and Excellence*. Oxford: Elsevier.
- Astin, A., Vogelgesang, L. J., Ikeda, E. K., and Yee, J. A. (2000). *How Service Learning Affects Students*. Los Angeles, CA: Higher Education Research Institute, UCLA.
- Bringle, R. G. and Hatcher, J. A. (2000). Institutionalization of service learning in higher education. *Journal of Higher Education* 7(3), 273–290.
- Bringle, R. B. and Hatcher, J. A. (2002). Campus–community partnerships: The terms of engagement. *Journal of Social Issues* 58, 503–516.
- Chang, M. J., Astin, A. W., and Kim, D. (2004). Cross-racial interaction among undergraduates: Some consequences, causes, and patterns. *Research in Higher Education* 45, 529–553.
- Chang, M. J., Witt, D., Jones, J., and Hakuta, K. (2003). *Compelling Interest: Examining the Evidence on Racial Dynamics in Colleges and Universities*. Stanford, CA: Stanford University Press.
- Driscoll, A., Holland, B., Gelmon, S., and Kerrigan, S. (1996). An assessment model for service-learning: Comprehensive case studies of impact on faculty, students, community, and institutions. *Michigan Journal of Community Service Learning* 3, 66–71.
- Eyler, J. S. (2000). *Special Issue: What Do we Most Need to know about the Impact of Service-Learning on Student Learning?* *Michigan Journal of Community Service Learning* Fall, 11–17.
- Galston, W. (2001). Political knowledge, political engagement, and civic education. *Annual Review of Political Science* 4, 217–234.
- Gray, M. J., Ondaatje, E. H., Fricker, R., et al. (1998). *Coupling Service and Learning in Higher Education: The Final Report of the Evaluation of the Learn and Serve America, Higher Education Program*. Santa Monica, CA: The RAND Corporation.
- Hodge, G., Lewis, T., Kramer, K., and Hughes, R. (2001). Collaboration for excellence: Engaged scholarship at Collin County Community College. *Community College Journal of Research and Practice* 25, 675–690.
- Hurtado, S. (2003). Preparing college students for a diverse democracy. *Final Report to the U.S. Department of Education, Office of Educational Research and Improvement, Field Initiated Studies Program*. University of Michigan.

- Moely, B. E., McFarland, M., Miron, D., Mercer, S., and Ilustre, V. (2002). Changes in college students' attitudes and intentions for civic involvement as a function of service-learning experiences. *Michigan Journal of Community Service Learning* 9(1), 18–26.
- Myers-Lipton, S. J. (1996). Effect of a comprehensive service-learning program on college students' level of modern racism. *Michigan Journal of Community Service Learning* 3, 44–54.
- Putnam, R. (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York, NY: Simon and Schuster.
- Springer, L., Palmer, B., Terenzini, P. T., Pascarella, E. T., and Nora, A. (1996). Attitudes toward campus diversity: Participation in a racial or cultural awareness workshop. *Review of Higher Education* 20(1), 53–68.
- Yates, E. L. (2000). Helping students build bridges across racial divisions. *Black Issues in Higher Education* 17(15), 36–38.
- Zlotkowski, E. (2002). *Service-Learning and the First-Year Experience: Preparing Students for Personal Success and Civic Responsibility. The First-Year Experience Monograph Series*. Columbia, SC: South Carolina University, National Resource Center for the First-Year Experience and Students in Transition.

Related Websites

- <http://www.campuscompact.org> – Campus Compact.
- <http://www.servicelearning.org> – Learn and Serve America's National Clearinghouse.
- <http://www.civicyouth.org> – The Center for Information and Research on Civic Learning and Engagement (CIRCLE).

Success and Failure in Education and Criminal Justice: Identifying Common Mechanisms

S Ramirez and A Harris, University of Washington, Seattle, WA, USA

© 2010 Elsevier Ltd. All rights reserved.

The aim of this article is to investigate links between academic failure in the United States educational system and incarceration in the United States criminal and juvenile justice systems. From a sociological perspective, we focus in particular on the disproportionate dropout and incarceration rates among young people of color in high school and prison, respectively. The purpose is to highlight current research that speaks to the obvious – yet little understood – link between the two institutions, and to shed light on related mechanisms that produce racial disparities and negative outcomes. One glaring piece of the pipeline linking these two institutions that has had little attention is the phenomena of alternative schools. These schools are public programs which provide individualized educational options for students, many of whom have been expelled from mainstream schools. Much of this article reviews research investigating these unconventional programs and discusses how alternative schools can both lead to educational failure and criminal justice embroilment.

Recently, the United States Department of Education highlighted the fact that the United States has had the highest high school graduation rate ever at 85%. In 2003, the total graduation rate for adults aged 25 and above was 85%, and among this same-age population, 27% had college degrees. However, we see varying completion rates by racial/ethnic groups. Among whites, the graduation rate was 89%, African-Americans it was 80%, and among Latinos the rate was 57% (US Department of Education, 2007).

It is important to note that if we do not include other types of credentials for high school graduates in these statistics (such as General Educational Development (GED) degrees), the graduation rates by racial/ethnic group are much more bleak. A recent analysis was conducted that estimates graduation rates for 1998 (Green, 2001). This investigation includes only those individuals who received high school diplomas and is based on the number of students enrolled in public schools in eighth grade, thus accounting for high school dropouts. The study estimates a 71% completion rate among United States high school graduates, compared to the National Center for Education Statistics that show an 86% graduation rate for the class of 1998 (Greene, 2001). Furthermore, the study finds that among white students the high school completion rate was 78%, while it was 56% for

African-American students, and 54% for Latino students. State-level differences among the racial/ethnic categories varied dramatically. For example, African-American students had the lowest graduation rate at 40% in Wisconsin, yet had a 71% graduation rate in West Virginia. The lowest graduation rate among Latinos was 32% in Georgia and peaked to 82% in Montana (Greene, 2001).

In contrast, the point is that traditional measures for generating high school education statistics ignore both the dropout rate, as well as obfuscates the number of students who are not receiving high school credentials through traditional schooling channels. Furthermore, the traditional measures do not indicate (nor include in their calculations) the relatively high dropout rate among students of color. For example, the 2003 dropout rate among whites was 6.3%, and was 10.9% for African-Americans, and 23.5% for Latinos (US Department of Education, 2007).

At the same time we see these racial/ethnic disparities in graduation and dropout rates, we see even more dramatic disparities in juvenile and criminal justice conviction and incarceration rates. As of 2001, it is estimated that there were 5.6 million US adults (2.7% of the adult population) who served time in state or federal prison (BJS, 2003). The prevalence of imprisonment was higher for African-American males (16.6%) and Latino males (7.7%) than for white males (2.6%). Of those in prison in 2001, 64% were people of color (nonwhite). The lifetime chances of a person going to prison are highest among African-Americans (18.6%) and Latinos (10%), than for whites (3.4%). An even more dramatic figure is that estimates show that one out of every three African-American men will be sentenced to state or federal prison in their lifetime (Travis, 2005).

Several questions are raised by these statistics. What is leading to the relatively high dropout rate among youth of color? What does the pathway look like for students who take a nontraditional route to receiving high school credentials (via alternative school programs and GED programs)? What are the mechanisms that lead to the bleak incarceration prospects for African-Americans? Are the mechanisms that lead to failure in the educational system similar to those that lead to selection into the criminal justice system?

We know that there is a direct relationship between education and incarceration. For example, among men born between 1965 and 1969, 3% of whites and 20% of

African-Americans served time in prison by their 1930s. Furthermore, among this population 30% of those without a college education and nearly 60% of high school dropouts went to prison (Pettit and Western, 2004). This trend appears to be consistent among more recently incarcerated populations as well: of those in state prison facilities in 2001, only 57% had a high school diploma or its equivalent (BJS, 2003). Furthermore, schooling has been found to significantly reduce the probability of incarceration and arrest (Lochner and Moretti, 2004).

Thus, individuals who fail at education – either by dropping out of school, or by not continuing on to higher education – have much higher likelihoods of having contact with the criminal justice system and of being incarcerated. However, what is less clear are the mechanisms that create this linkage. Research has shown dramatic negative consequences for individuals from decreased wages (Western and Pettit, 2005), unstable families (Uggen *et al.*, 2005), loss of citizenship rights (Manza and Uggen, 2006), and health-related problems (Travis, 2005). Thus, it is increasingly important that research turn toward identifying factors that connect the two institutions and produce the disparate outcomes.

The remainder of this article investigates research focusing on the mechanisms that may link the United States educational and criminal justice institutions. We begin with a discussion of the reciprocal relationship between the two systems – how failure in one leads to failure in the other, and vice versa. More specifically, we move to explore alternative or probationary schools – common options for youth who have had disciplinary or learning problems in mainstream schools. We conclude with a discussion about future research that should be explored to further our understandings of the linkages between education and incarceration.

The Link between Education and Incarceration

The juvenile justice system was originally intended to provide rehabilitation in the way of schooling, work experience, and vocational skills to wayward children who were engaging in delinquent behavior or were at risk for child abuse and neglect (Platt, 1967). Despite this noble intent to provide guidance and education to such children, evidence suggests that youth under the supervision of the juvenile justice system, upon reentering society actually have lower performances in school, experience high dropout rates, and thus have limited opportunities for employment in their transition to adulthood.

For example, as contact with the justice system increases, the performance in math and reading decreases (Hirschfield, 2004). Furthermore, first arrests have unclear effects on grade 7- and 8-level math and reading

scores, but the chances of having to repeat grade 8 sharply increases. First arrests which occur in grade 9 not only decrease school performance in reading and math, but also increase the likelihood of dropping out of school or of greater instances of truancy. For those children who experience detention or incarceration or who are arrested multiple times, drop in school performance, likelihood of truancy, and dropping out are experienced at even higher rates. Most likely spending time away from school during a stint of detention, or having to attend court hearings likely explains why it is easy for these children to fall behind. However, such evidence suggests that a justice system originally meant to deter future behavior actually decreases the chances of being able to succeed in school after a youth's first arrest.

Just as involvement in the justice system is an indicator for poor school performance, low school performance is also a significant predictor of future delinquency or more serious criminal behavior. It is well documented in criminological literature that school is an extremely important socializing institution for children in adolescence (Warren, 2000). Research indicates that irregular school attendance and school behavioral problems are a good predictor of delinquency as well as recidivism for youth (Sieverdes and Cunningham, 1977). Indicators such as school performance and involvement in school are even better predictors of delinquency than measures of self-control and deviant peer pressure (McGloin *et al.*, 2004); these measures of school performance are also important in predicting future criminality (LeBlanc, 1994).

In sum, prior research suggests that not only school performance is a significant predictor of delinquency and future criminal behavior, but also involvement in the justice system decreases the ability for adolescents to succeed in school.

School Dropouts

Similarly to research on school performance, research on youth who drop out of school reveals interesting implications for the relationship between the educational system and the criminal justice system. Early signs of delinquency in the form of drug use are associated with high school dropout rates as well as other adult life-course markers experienced at an early age, such as teenage pregnancy. In turn, dropping out of school increases the risk of drug and alcohol use, especially for males.

It is important to note that dropout rates tend to affect specific social groups, such as those living in poverty. For example, dropout rates for children living on public assistance have dramatically increased over the past 10 years (Orthner and Randolph, 1999). Transitions off welfare and consistent parental employment decrease the risk of dropping out of school. Such evidence suggests that there are implications for the relationship between poor

or incomplete schooling and the most impoverished segments of society.

Together, evidence from research on the role of school, dropping out of school, and low school performance suggests that not only is involvement in school an important factor for delinquency and future criminality, but also that initial delinquency and involvement in the juvenile justice system does not promote success in future educational achievement. Furthermore, consequences of this nexus between the justice system and education system seem to be exacerbated for people of color (namely African-Americans and Latinos).

Race-Specific Consequences

Sociologists have well documented the disproportionate trends in sentencing and other justice-outcome processes for people of color (Krisber, 2005). African-Americans are overrepresented in almost every area of the criminal justice system, including arrest, likelihood of being adjudicated delinquent, severity in sentence types, probation restrictions, and violations (Leiber, 2002). In addition, Bridges and Steen (1998) find that probation officers' construct attributions differently for black and white youth. In an analysis of probation narratives, they found that probation officers were more likely to rely upon external reasons to explain white youth's delinquency, such as parental involvement, influence of drugs, and peer pressure. In contrast, probation officers relied on internal characterizations to make judgments about African-American youth's offending, such as having low self-control or poor character. In turn, these attributions had substantially different justice consequences for the two groups of youth.

Along similar lines, recent work suggests that youth of color experience hyper-criminalization in the juvenile justice system. For example, Rios (2006) found that Latino and black youth experience sentences common for violent offenders even for nonviolent offences. Rios suggests that this stigma translates to other institutions such as the educational system. Thus, even in institutions built to nurture youth, youth of color experience a racialized socialization process.

Similarly to the justice system, within the educational system, indicators illustrate that youth of color experience poor performance disproportionately to white youth. Katsinas (1989) finds that in Illinois, higher education has an under-enrolment of Latino youth that results from extremely high dropout rates of 60%. Katsinas also suggests that the dropout rate is influenced by socioeconomic status and lack of Latino teachers. These results are even more exaggerated for Mexican immigrants who achieve only 50% of the educational levels of their native-born counterparts and non-Latino whites. In addition, Mexican immigrants are 300–400% more likely to drop out of school (Charvez, 1989).

Harris and Allen (2003) show that similar patterns exist for African-Americans. In an analysis of the link between the overrepresentation of black youth in the justice system and underrepresentation of black youth in education, Harris and Allen show that African-Americans represent only 8% of high school graduates in California; furthermore, only 27% of these graduates completed the required coursework for college eligibility. In addition, Maralani (2006) finds that those with a GED certification rather than a traditional completion of a high school diploma are less likely to go to college. While Maralani (2006) does not specifically talk about racial disparities, given what we know that youth of color are more likely to drop out of school or have less success in high school, and thus may have to rely on nontraditional ways of completing a high school equivalent education such as the GED certificate, this link between the justice system exposure and educational exposure may help explain the under-enrolment of youth of color in higher education. This is likely to have implications for future employment as youth make the transition out of school and into the workforce.

Overall, literature on educational failure points to some common elements that may explain why youth of color have a disadvantage for school success. Specifically, research suggests that youth of color do not receive as much support in school as their white counterparts due to the lack of opportunity to attend quality schools. Furthermore, schools with high percentages of youth of color tend not to have the necessary resources for a well-rounded education, nor do these schools have culturally appropriate curricula or ethnically diverse teachers (Hirschfield, 2004; Katsinas, 1989; Habteyes and Steinkamp, 1985).

Not only do school experiences in general seem to negatively impact youth of color, but also zero-tolerance policies for inappropriate conduct in school generally target youth of color (Johnson, 2003). Such policies, which automatically punish students for a host of school infractions, have also been found to disproportionately punish youth of color. Some argue that this is because youth of color are more likely to be disproportionately involved in bad school behavior. However, Johnson (2003) argues that socioeconomic status underpins this disproportionality. Regardless, zero-tolerance policies increase the rate of high school dropouts and thus the proportion of youth of color involved in the juvenile justice system.

Alternative/Probationary Schools

How schools deal with students identified as discipline problems has had limited attention in sociological literature. However, the presence of alternative schools is important in any discussion regarding the link between educational failure, involvement in the justice system, and how these institutions disproportionately disadvantage

youth of color. Alternative schools were first seen as an option for parents to have their children educated. These schools were looked to as alternatives for students to specialize early on in certain subject areas, work at their own learning levels, and explore learning in a more hands-on manner. However, today, alternative schools are often used as a holding area for children deemed difficult by traditional school staff. The following section outlines some of the literature about alternative school education and how they may be an important link between educational failure and involvement in the justice system, incarceration, and future criminal behavior.

Alternative or probationary schools have typically been framed as institutions which provide an avenue for parents and school systems to manage difficult students. These students are typically disruptive in a traditional school setting and/or are in transition from juvenile detention. These schools were thought to be useful in curbing juvenile crime (Gottfredson, 1987) given that these schools would provide disruptive children with a more supportive environment.

While there is no standard format for alternative school programs, common elements include low student-teacher ratios, individualized self-paced instruction, non-competitive performance assessment, and less-structured classrooms (Cox, 1999). These elements allow for a more individualized approach to student learning that was thought to improve school attendance, motivation for school instruction, and school performance. While these alternative programs are available to all students in public school systems in the United States, scholars have argued that these programs become a dumping ground for disruptive children that traditional schools find difficult to handle (Cox, 1999; Lehr Lanners and Lange, 2003).

Mixed Results about Effectiveness

Evidence for the success of alternative school programs in increasing school performance, motivation for school involvement, and decreasing involvement in delinquent activities has been mixed. Early evaluations of alternative school programs provide little evidence that these programs have any positive effect on delinquency or school performance for students. Raywind (1981) found that the lack of improper effects were a result of school officials using these alternative programs as punishment for students who were uncontrollable in traditional classrooms. More recent analysis suggests that improvement in school performance was short lived for students at an alternative school program. Cox (1999) finds that students in an alternative school showed marked improvement in school grades, school attendance, and self-esteem; however, at a 1-year follow-up, these effects did not remain. Cox also finds no evidence that the alternative program had any

observed effect on attitudes toward school or delinquent behavior.

Some research has suggested that alternative schools do well in their ability to discipline and train students for lower aspirations. Duke and Perry (1978) find that alternative school staff are less likely to rate behavior problems as a major school issue compared to traditional school counterparts. Further, they suggest that these schools markedly improve problem behavior and school performance in youth enrolled in alternative schools. Furthermore, a recent ethnographic study (Inderbitzen, 2007) found that school staff at one alternative program attempted to refocus students' conceptions of success and aspirations to something more achievable. Specifically, these students were taught to aim low for opportunities that were more consistent with what educators felt was more appropriate and realistic. While this case study may not be generalizable to other alternative programs, these findings indicate that important questions should be asked about the long-term trajectories for delinquency, criminal behavior, and future employment for youth involved in alternative school programs.

Research on alternative schools and their students is limited in sociological and criminological literature. As such, we know very little about the affect these alternative school programs have on delinquent youth or at-risk youth in educational outcomes, propensity for further delinquent behavior, or transition to adult roles in employment. However, consistent in both academic literature and high school evaluations is that alternative schools are perceived as a dumping ground for students who are labeled as uncontrollable in the traditional school system and those who are involved in delinquent activity. As such, alternative schooling might be perceived as punishment rather than an alternative opportunity to become an educated citizen.

For example, annual reports from one alternative school in an urban school district in Washington State suggest that in all academic performance indicators, students in the alternative schools performed well under the state standard for academic performance. For example, in 2004, 48.6% of students in traditional schools performed at the state standard for grade 10 reading ability, while only 28.6% of students in the alternative high schools met the state standard. Similar patterns can be observed for mathematics. These results are based on the Washington State Assessment for Student Learning (WASL) administered to eighth and tenth graders. The alternative schools also reported lower number of Scholastic Aptitude Test (SAT) test-takers and graduation rates than traditional schools in the same district (Seattle Public Schools Annual Reports, 2007). Demographic information on the alternative high schools suggests that a typical student is African-American, not living with both parents, enrolled in a reduced or free lunch program, and almost

Table 1 Performance indicators for WA alternative and traditional schools percent meeting state standards for WA State WASL

	Year	District (%)	Traditional (%)	Alternative (%)
Reading	2007	81	79	52
Writing	2007	84	83	54
Mathematics	2007	50	52	15
Science	2007	36	33	8
Dropout rate ^a				
	2005	29	18	62
	2006	28	15	54
	2007	28	17	57
Graduation rate ^b				
	2005	70	83	35
	2006	69	86	39
	2007	63	76	27
Students of color ^c	2007	58	59	71
Number of schools	2007	17	12	5

^aBased on all students who enrolled since 2002 who expected to graduate in 2005, 2006, and 2007.

^bDoes not include students transferred to other schools.

^cIf students do not meet standards in 12th grade they will not graduate (first class 2010).

50% of the enrolled students have been placed in special-education curriculum (Seattle Public Schools Annual Reports, 2007). **Table 1** compares the percent of students meeting state performance indicators at a tenth-grade level between traditional and alternative schools, and dropout and graduation rates. Among the alternative schools, rates are considerably lower for performance indicators, graduation rates, and dropout rates than for traditional schools.

To conclude, Rios (2006) suggests that the juvenile justice system is an intermediary between school and future criminal behavior and as such, schools play a vital role in the trajectory of youth toward criminal careers. Disproportionately, people of color have been ensnared in the juvenile and criminal justice systems, suggesting that when certain schools, such as alternative ones, become a holding area for youth of color with actual or perceived behavior problems and, possibly, current justice system monitoring, further justice system involvement and educational failure are likely to persist.

Common Mechanisms

Clearly, a link exists between youth's experiences in the educational system and experiences in the justice system. Juveniles who fail or have difficulties in school are more likely to experience some kind of supervision from the justice system such as probation or detention

(Hirschfield, 2004). In addition, failure in school is a well-known predictor of future delinquency and/or criminal behavior (Hirschfield, 2004; Krohn *et al.*, 1997). Conversely, those who have experienced justice system supervision are less likely to be able to succeed in school (Hirschfield, 2004; Rios, 2006). These youth may have to turn to alternative methods of receiving high school equivalent education such as GED certification, or may be forced to attend alternative schools rather than traditional high school programs.

At this time, there is not enough evidence to demonstrate a causal link between failure in one institution and failure in the other; however, with an examination of the common mechanisms within the two systems, we gain understanding on how these two institutions are related and also why youth of color are overrepresented in the justice institutions and underrepresented or are failing in educational institutions. This has important implications for future behavior, success, and other life-course stages of children and young adults.

At the very least, the above review shows that common to both failure in school and criminal/juvenile justice outcomes is the quality of schooling that programs such as alternative schools may be providing students. This is particularly important for students of color, and students who have been placed in alternative education due to delinquency or a lack of success in traditional school.

Furthermore, young adults' attachment to conventional institutions and norms may be mechanisms linking the two institutions. For example, dropout rates are significant for students from low socioeconomic families. Perhaps these students are encouraged to drop out from school in favor of taking up paid employment in order to support family income. In sociology, two differing perspectives on early work experienced dominate the literature. First, some argue that early work experiences (as a part-time job while still in school) are beneficial by building confidence and responsibility, thus aiding the transition from adolescence to adulthood (Staff, 2005). Conversely Paternoster *et al.* (2003) argue that teenage work experiences remove children from school and thus make them more vulnerable to delinquency by placing them in low-skilled jobs with little stability and security. Not only do these students become more vulnerable to delinquency but also vulnerable to other problem behavior and future criminal behavior (Paterneoster *et al.*, 2003). Assuming that high school dropouts leave school and find some kind of employment, legitimate or illegitimate, Paternoster's (2003) argument may illuminate an important link between delinquency, school participation, and the transition into adulthood, especially where perpetuating deviant behavior might be concerned.

Involvement in the juvenile justice system may be a risk factor in and of itself for school dropouts since justice system requirements take children out of school. Court

appearances and detention remove youth from school from both short and extended periods of time and therefore disrupt educational involvement. Control theorists might argue that high school dropouts have a weak commitment and attachment to school as a socializing institution and thus it is not surprising that these youth are at a higher risk for delinquent involvement (Hirschi, 1969).

No discussion about the link between education and the justice system can be complete without noting the trends in government funding to these institutions. Western *et al.* (2003) argue that while states spend more on education than corrections in general, the growth in spending rates have favored spending on corrections and legal systems rather than education. For example, from 1977 to 1999, funding for corrections has increased by 946% and funding for the legal system has increased by 1518%. In contrast, funding for education has increased only by 370%.

Discussion and Conclusion

Clearly, rates of low school performance, quality of available education, and high dropout rates affect students of color disproportionately. At the same time, the association of high rates of youth of color represented within the justice system cannot be ignored. Yet, sociologists and criminologists have not given the apparent reciprocal link between education and incarceration and justice system involvement much attention. Not only is it clear that educational failure is associated with involvement in the justice system, but detention and incarceration as well as arrest as criminal justice supervision are also associated with future educational failure. Ironically, the juvenile justice system was designed to rescue wayward children and reform them into valuable citizens for society; yet, mounting evidence suggests involvement in the justice system impedes adolescents' ability for future success.

Very little sociological or criminological research has been done on alternative school programs. If students in alternative schools perform academically lower than traditional schools, it is unlikely that these programs promote further educational achievement such as completing a high school diploma, or admission into higher education. Therefore, how much do alternative schools perpetuate the disproportionate rate of people of color in the criminal justice system? While education seems to play an important role for avoiding incarceration and gaining employment, might not the quality of education also play a role?

International Perspective

As sociologists, we must be careful in assuming that other nations have experienced the extent of discrimination

toward people of color that has become part of the socialization process in the United States. However, it is clear that there are lessons to be learned about how education systems and justice systems influence each other and may be able to work together in order to combat social inequality, discrimination, and widespread social problems.

Similar trends in low school performance have been found in the UK and Australia, suggesting that these indicators are also closely linked to delinquency, teenage pregnancy, drug use, and future criminal behavior (Dwyer and Wyn, 2001). In particular, Australian Aboriginals have experienced high dropout rates and high incarceration rates, especially in rural areas where aboriginal communities are concentrated (Palmer, 1999; Michalos and Orlando, 2006).

In order to amend discriminatory practices within the justice system for people of color, the UK, Australia, and New Zealand are using or currently experimenting with new ways to do justice (Consedine, 1999; Braithwaite and Mugford, 1994). In particular, restorative justice practices that involve more participation on behalf of the offender's family, victim, community members, and teachers are designed to provide a different experience of social control for groups in society that have experienced adverse outcomes and extreme consequences, such as suicide and other forms of mental illness, and failure in education and employment (Grabosky, 1998).

Restorative practices such as these may provide one way for closer collaboration between justice institutions and educational institutions, especially for youth of color. If teachers do provide motivation for success in school and quality of education as research suggests (Cox, 1999; Backstrom, 2004), having offenders and teachers involved in justice system outcomes may provide necessary motivation to continue to engage or reengage them in education and thus help contribute to less-frequent contacts with police and detention (Weidjers and Duff, 2004).

In another example, post-apartheid South Africa has made several changes to the availability of basic social institutions such as education to ethnic minorities. However, because of such widespread and deep-rooted discrimination practices of the past, cities and communities still remain quite segregated and thus quality of education and involvement in the justice system are still experienced disproportionately for South African blacks (Marx, 1998). As such, given the trends and research discussed here, sociologists might expect a similar relationship between educational failure, high dropout rates, and justice system involvement in South Africa that is disproportionately experienced by people of color. Consequently, future policy might explore the possibility of a closer working relationship between educational systems and branches of criminal justice.

Bibliography

- Backstrom, R. J. (2004). Examining teachers' experiences in a small middle college. *Dissertation Abstracts International A* **65**(4), 1312.
- Braithwaite, J. and Mugford, S. (1994). Conditions of successful reintegrative ceremonies: Dealing with juvenile offenders. *British Journal of Criminology* **34**(2), 139–171.
- Bridges, G. and Steen, S. (1998). Racial disparities in official assessments of juvenile offenders. *American Sociological Review* **63**(4), 554–570.
- Charvez, L. (1989). Tequila sunrise: The slow but steady progress of Hispanic immigrants. *Policy Review* **48**, 64–67.
- Consedine, J. (1999). *Restorative Justice: Healing the Effects of Crime*. Lyttelton: Ploughshares Publications.
- Cox, S. M. (1999). An assessment of an alternative education program for at-risk delinquent youth. *Journal of Research in Crime and Delinquency* **36**(3), 323–336.
- Duke, D. L. and Perry, C. (1978). Can alternative schools succeed where Benjamin Spock, Spirrow Agnew and B. F. Skinner have failed? *Adolescence* **13**(51), 375–392.
- Dwyer, P. and Wyn, J. (2001). *Youth Education and Risk*. London, UK: Routledge Farmer.
- Greene, J. (2002). High school graduation rates in the United States. *Revised Report, The Manhattan Institute for Policy Research*. Original Report: November 2001.
- Habteyes, Y. and Steinkamp, M. (1985). Sociocultural factors and achievement motivation in the United States Virgin Islands. *Ethos* **13**(1), 75–93.
- Harris, A. and Allen, W. (2003). Lest we forget thee . . . the under- and over-representation of Black and Latino youth in California higher education and juvenile justice institutions. *Race and Society* **6**, 99–123.
- Hirschfield, P. J. (2004). Preparing for prison? The impact of legal sanctions on educational performance. *Dissertation Abstracts International A* **65**(1), 294.
- Inderbitzen, M. (2007). Inside a maximum security juvenile training school. *Punishment and Society* **9**(3), 235–251.
- Johnson, V. (2003). *Fundamental Differences*. Lanham, MD: Rowman and Littlefield.
- Katsinas, S. G. (1989). Educational arrears: Addressing the underenrollment of Hispanics in Illinois higher education. *Urban Review* **21**(1), 35–50.
- Krisber, B. (2005). *Juvenile Justice: Redeeming Our Children*. Thousand Oaks, CA: Sage.
- Krohn, M. D., Lizotte, A. J., and Perez, C. (1997). The interrelationship between substance use and precocious transitions to adult statuses. *Journal of Health and Social Behavior* **38**(1), 87–103.
- LeBlanc, M. (1994). Family, school, delinquency and criminality, the predicative power of an elaborated social control theory for males. *Criminal Behavior and Mental Health* **4**(2), 101–117.
- Leiber, M. (2002). Disproportionate Minority Confinement (DMC) of youth: An analysis of state and federal efforts to address the issue. *Crime and Delinquency* **48**(1), 3–45.
- Maralani, V. (2006). From GED to college: The role of age and timing in educational stratification. *California Center for Population Research*.
- Marx, A. W. (1998). *Making Race and Nation: A Comparison of the United States, South Africa and Brazil*. New York, NY: Cambridge University Press.
- McGloin, J. M., Pratt, T. C., and Maahs, J. (2004). Rethinking the IQ–delinquency relationship: A longitudinal analysis of multiple theoretical models. *Justice Quarterly* **21**(3), 603–635.
- Michalos, A. C. and Orlando, J. A. (2006). Quality of life of some under-represented survey respondents: Youth, aboriginals and unemployed. *Social Research Indicators* **79**(2), 191–213.
- Orthner, D. K. and Randolph, K. A. (1999). Welfare reform and high school dropout patterns for children. *Children and Youth Services Review* **21**(9–10), 881–900.
- Palmer, D. (1999). Indigenous young people and victimhood. *Social Alternatives* **18**(2), 52–57.
- Paternaoster, R., Bushway, S., Brame, R., and Apel, R. (2003). The effect of teenage employment on delinquency and problem behaviors. *Social Forces* **82**(1), 297–335.
- Pettit, B. and Western, B. (2004). Mass imprisonment and the life course: Race and class inequality in U.S. incarceration. *American Sociological Review* **69**(2), 151–169.
- Platt, A. (1967). *The Child Savers: The Invention of Delinquency*. Chicago, IL: University of Chicago Press.
- Rios, V. M. (2006). The hyper-criminalization of Black and Latino male youth in the era of mass incarceration. *Souls* **8**(2), 40–54.
- Sieverdes, C. and Cunningham, O. (1977). Family adjustment, school performance, and recidivism among official juvenile offenders. *Mid-South Sociological Association Annual Meeting*.
- Staff, J. (2005). Precocious maturity and the process of occupational attainment. *Dissertation Abstracts International A* **65**(8), 3176.
- Travis, J. (2005). *But They All Come Back: Facing the Challenges of Prisoner Reentry*. Washington, DC: The Urban Institute Press.
- Uggen, C., Wakefield, S., and Western, B. (2005). Work and family perspectives on reentry. In Travis, J. and Visser, C. (eds.) *Prisoner Reentry and Crime in America*, pp 209–243. New York: Cambridge University Press.
- Warren, P. Y. (2000). Academic achievement, strain and delinquency. *Southern Sociological Society Annual Meeting*.
- Western, B. and Pettit, B. (2005). Black-White wage inequality, employment rates, and incarceration. *American Journal of Sociology* **111**(2), 553–578.
- Western, B., Schiraldi, V., and Zidenberg, J. (2003). Education and Incarceration. *Justice Policy Institute, Policy Report, August 28, 2003*. Washington, DC: Justice Policy Institute.
- US Department of Education, National Center for Education Statistics (2007). *Digest of Education Statistics 2007*. Washington, DC: National Center for Education Statistics.

Further Reading

- Travis, J. and Visser, C. (2005). *Prisoner Reentry and Crime in America*. New York: Cambridge University Press.
- Weijers, I. and Duff, A. (2002). *Punishing Juveniles: Principle and Critique*. Portland, OR: Hart Publishing.
- Winters, C. A. (1997). Learning disabilities, crime, delinquency, and special education placement. *Adolescence* **32**(126), 451–462.

The History of Education: Race and Education

J O Jewell, Loyola Marymount University, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Historically, schools have been viewed as both the object and the means of struggle between groups, facilitating the social control of various populations while simultaneously embodying the potential to ameliorate various types of social inequality, such as class, gender, and race. In the case of those social systems structured by race, the distribution of educational opportunities has tended to be shaped by a durable but malleable racial structure. The current body of social science literature on education focuses on the role that schools played in reinforcing and legitimizing hierarchies of race through a variety of means including, but not limited to, curricular choices, patterns of organization, and policies governing school funding.

What Is Race?

Race is best understood as the complex set of social meanings ascribed to real or perceived phenotypical and/or cultural differences among human populations that generate the unequal distribution of valued resources. Among scholars who study race and its impact upon human societies, three general principles guide current research. First, within social systems structured by race, all social actors are familiar with and subjected to racialized schemes of classification, systems of belief, and cultural strategies that tend to construct and reproduce racial meanings over time. Second, the socially constructed nature of racial categories and its accompanying meanings, evinced by sociohistorical variations and shifts in the number and composition of racial categories, mean that actors and groups of actors have a vested interest in either maintaining or substantially rearticulating racial identities. Third, the pervasive and entrenched nature of racial hierarchies in social systems means that they interact systemically and frequently as structures with other hierarchical systems such as class, ethnicity, gender, and sexuality, with which they are deeply articulated and in ways that mutually construct and reproduce their existence over time.

Schools as Raced Spaces

Scholarship on race and education contends that racial hierarchies are frequently reflected in the distribution of

educational opportunities. However, a more recent body of scholarship posits that within racialized social systems, schools operate as raced spaces, in which students are actively assigned to racial categories which have significance for the content of school experiences and the broader social context in which those experiences take place. As such, schools are implicated in the construction and ongoing maintenance of racial hierarchies. In the late nineteenth and early twentieth centuries, the growth of the educational system in the United States occurred within a context of large-scale change to the nation's racial structure due to rapid industrialization, urbanization, foreign immigration, westward expansion, and emancipation. These changes generated extensive debate about how, to what extent, and in what social context children from various racial groups should be educated (Fass, 1989; Spring, 2004; Tyack, 1974).

African Americans, who were considered chattel property until the late nineteenth century, were almost totally excluded from educational access. In the postbellum South, educational access for blacks became a primary issue as its outcomes would affect not only the chances of African Americans for mobility within a new economic and political regime, but also the future of the South as a region. Conflicts among those who wished to maintain blacks' subordinate social position and those who desired educational opportunities commensurate with full and equal citizenship were common in this period. Ultimately, despite the efforts of blacks and their progressive white allies, those who advocated for limited access in racially separate institutions held sway for nearly more than a century. In the north, state-sanctioned practices of *de facto* segregation, in which blacks were concentrated in schools that received fewer economic and physical resources, had much the same result, effectively limiting black economic and occupational mobility.

Native peoples, long seen as both cultural and physical obstacles to nation building and modernization, were targeted by the mission agencies of religious bodies with an educational program designed to supplant native cultures with that of the Anglo-Protestant majority. Native-American education became a primary focus with westward expansion in the mid- to late nineteenth century. Americanization or civilization of native children continued through a system of boarding schools that functioned as total institutions – teaching vocational skills while eradicating native culture. These institutions grew in number in the late nineteenth century, numbering 25 federally

funded off-reservation schools in the continental United States and more on-reservation community schools which employed the Americanization curriculum until 1928. However, educational opportunities remained limited as late as the 1970s.

Asian Americans, who had come to the Western territories in large numbers since the 1840s, experienced similar treatment. Initially, no educational provision had been made for the children of Chinese, Japanese, Korean, and East Indian laborers, leaving them dependent upon voluntary or community efforts or the missionary agencies of various religious bodies. In 1885, the California Supreme Court's decision in *Tape vs. Hurley*, which gave Chinese Americans access to the public schools of California, was modified that same year by the State legislature with an emergency provision that established separate schools for children of Mongolian or Chinese descent. Twenty years later, the San Francisco School Board created separate schools for Chinese, Japanese, and Korean students that persisted until the 1960s.

Mexican Americans, who found themselves incorporated into the United States as a result of international warfare, also experienced various forms of exclusion owing to both ideologies of race and culture. Seen by the new Anglo-American English-speaking power structure as cultural outsiders of indeterminate racial status, patterns of segregation similar to that experienced by blacks and other minorities emerged. In 1855 and 1870, California and Texas, the two largest states with Spanish-speaking populations, passed laws requiring that English be the sole language of instruction in all public schools. This later enabled the state-sanctioned establishment of separate schools for all students of Mexican descent on the grounds that their cultural peculiarities might hinder the educational progress of native-born Anglo children. As with blacks and other groups, these Mexican schools were often deprived of material resource and employed limited curricula geared toward vocational training. This Americanization plan to erode cultural differences and place Spanish-speaking populations within the US racial hierarchy would subsequently include migrant laborers and refugees from Mexico and later Puerto Ricans with the annexation of the island after the Spanish–American War.

The successive waves of European immigrants, who came to the United States in the late nineteenth century, prompted various types of reactions from the Anglo-Protestant majority. Those coming from Western Europe, who shared some degree of familiarity with Anglo-Protestant culture were seen as culturally (though not always intellectually) compatible with the nation's educational institutions and encountered little separation that was not a voluntary attempt on their part to preserve language and culture in the face of structured Americanization (read assimilation). By contrast, those who came from countries in Southern and Eastern Europe, Catholics and Jews in

particular, were seen as threats to the cultural homogeneity of the United States and were often subjected to exclusionary practices in schooling with limited educational opportunities.

The state-sanctioned unequal distribution of political resources, such as the vote and citizenship rights along racial lines, undermined the ability of racial minorities to secure public education and influence educational policies, ultimately contributing to educational disparities between groups. However, these populations continued to find ways to influence educational policies and to improve access to educational resources. As the state gradually conceded to the demands of these populations for access to schooling that would at the very least Americanize the other, those who perceived the education of racial minorities as a threat to white social and political dominance relied upon the use of ideologies that constructed these populations as innately inferior and best suited for subordinate places in the economy, often performing low-wage or manual labor. As a result, blacks, Mexicans, Native Americans, Asians, and some European immigrants were cast as moral and intellectual inferiors, whose educational content and conditions required high levels of external control and spatial separation.

Race, the State, and the Distribution of Educational Opportunities

The distribution of educational opportunities and resources by the state, in the form of assigning students to particular schools and making material resources available to schools serving various populations, has historically been structured by class and race. Although the processes by which educational opportunities and resources are distributed may shift over time, and they may operate differently in varying social contexts, their end result has been maintenance of durable systems of inequality. What follows is an examination of these mechanisms as they relate to the construction and maintenance of racial inequality.

Racial Segregation

In those US states and territories with large concentrated populations of immigrant, domestic, and colonized minorities, it was common practice to assign children to schools on the basis of race. Despite residing in communities where they often represented significant parts of the population, racialized minorities could never assume that they had the ability to control their educational destinies – neither in choice of curricula and teachers nor the availability of school facilities, or the length of the school term. In the case of the United States, although segregationist practices existed in some form or fashion in various regions of the country directed against Asians, Native Americans,

Mexican residents of the conquered Western territories, and African Americans (Adams, 1997; Anderson, 1988; Iber and De León, 2006; Nakanishi and Nishida, 1995), they were given legal backing in the US Supreme Court's 1896 ruling in *Plessy v. Ferguson*, which under the separate-but-equal doctrine allowed states to effectively separate access to educational resources on the basis of race. Even in those states where it was not codified in law, *de facto* practices or social conventions sanctioned the practice, facilitating the unequal distribution of symbolic and material resources. Although the ruling stated that the resources provided to each racial group must be equal, few, if any, legal enforcements were typically made to ensure and maintain equality between schools.

Nevertheless, minority populations used these segregated institutions with varying degrees of success to challenge racial inequality, producing activist leadership who petitioned the state for change. This was especially true in the Southern United States, where for nearly a century the racially segregated private and publicly funded colleges and universities for blacks also fostered the growth of a black leadership class that figured largely in the struggle for racial equality (Jewell, 2007). For this reason, local and state authorities often maintained a close watch on the curricular and extracurricular activities of students within such institutions, withholding financial and other resources as a method of social control.

Intelligence Testing

In the late nineteenth and early twentieth centuries, multi-ethnic and multiracial states continued to justify the unequal distribution of educational resources along racial lines using new, scientifically imbued mechanisms of inequality. The growth of scientific racism and the popularity of the eugenics movement fueled the development of standardized examinations, which claimed to objectively measure cognitive ability and learning potential. As the twentieth century progressed and educational systems expanded due to industrialization, urbanization, immigration, and imperialism, these instruments were considered to be effective tools for organizing public schooling (Fass, 1989; Kliebard, 1995; Spring, 2007; Tyack, 1974; Tyack and Cuban, 1995). Despite their claims to universality, however, such tests typically reflected contemporary racial and cultural biases, having been developed by native-born white elites (see Young, 1922; Binet and Simon, 1916). Not surprisingly, these tests tended to label the children of racialized minority populations as intellectually and/or morally deficient, and in need of distinct types of instruction best carried out in separate educational facilities. As an alternative to separation between schools, minority children could be and often were placed in separate classes or curricular tracks in the same schools on the basis of their

perceived ability. Within these classes, they were prepared for positions in the economy as low-level or menial workers.

Race and Educational Reform

Desegregation

State-supported racial segregation in schools lasted for nearly 60 years, during which time minority populations lagged behind whites on most indicators of educational progress. In the mid-twentieth century, the United States Supreme Court, as a result of continued legal challenges to state and local practices of segregation and the findings of social scientists, decided unanimously in *Brown v. Board of Education of Topeka, Kansas* (1954) that the practice of segregated schooling was not only unconstitutional, but also inherently damaging to children with dire effects for society as a whole. The ruling in the *Brown* case forced widespread reassessments of the ways in which educational opportunities had been made available for racial minorities. Similar cases concerned with the detrimental nature of separate or hierarchical schooling on other minority groups both preceded and followed *Brown*, including *Mendez et al. v. Westminster School District of Orange County* (1946), *Delgado v. Bastrop Independent School District* (1948), and *Cisneros v. Corpus Christi ISD* (1971), which addressed Mexican Americans and other Latinos; *Lau v. Nichols* (1974), which dealt with the issue of bilingual instruction for Asian Americans; and the 1969 publication of *Indian Education: A National Tragedy – A National Challenge* by the US Senate Committee on Labor and Public Welfare, which examined the failings of education offered to Native Americans (US Senate Committee on Labor and Public Welfare, 1969). The ruling in the *Brown* case and the subsequent reexaminations of the American educational system it inspired called attention to the complex effects of race on academic access and achievement. The Elementary and Secondary Education Act of 1965, in which schools with a significant proportion of poor or disadvantaged students received additional federal monies, grew out of the concerns expressed in the *Brown* case.

Perhaps one of the most influential works of social science research in the immediate post-*Brown* era was *Equality of Educational Opportunity* by Coleman *et al.* (1966), sponsored by the US Department of Health, Education, and Welfare, to evaluate indicators of educational opportunity among white and minority students (blacks, Puerto Ricans, Mexican Americans, Asian Americans, and Native Americans) by assessing educational quality in terms of the availability of curricula, school facilities, academic practices, and the academic characteristics of teachers and student bodies in schools. Among Coleman's chief findings was that while persistent and unequal separation by race was found to be detrimental as per the Supreme Court's ruling in *Brown*, it was the combination of the socioeconomic

composition of the school, the familial and socioeconomic background of the students, and the nature of their peer groups together that accounted for the majority of the differences observed in academic achievement between white and minority schoolchildren. Racial integration alone, Coleman and his associates concluded, could not improve the academic performance of poor minority children. The effectiveness of these findings in dismantling racial segregation was, however, limited due to continued institutionalized resistance on the part of many state governments and local districts and the common focus on the *de jure* segregation common in certain parts of the country. As a result, the *de facto* segregation that had been practiced in the Northeastern and Midwestern states remained largely unaddressed.

The Post Civil Rights Era and Market-Based Reforms

Race and legacies of racial inequality continued to inform educational reforms long after the end of the Civil Rights Era. Influenced by the findings of the Coleman Report and other studies which took a dim view of compensatory education, neoconservative reformers sought to remedy race and class achievement gaps through what they described as a focus on standards and accountability, and an application of market forces to education. The publication of *A Nation at Risk*, which marked what some consider the end of liberal hegemony in American education, cited generally lower rates of performance among American schoolchildren and a persistent achievement gap by race and class as proof of the failure of liberal reforms which had tried to address inequalities stemming from race and class.

Many scholars contend that racial inequalities of a previous era continue to be perpetuated within and between schools. Patterns of residential segregation, in which whites leave racially mixed urban districts for homogeneous suburban ones, continue to influence the distribution of educational resources and opportunities. Within racially mixed schools, minority students often find themselves tracked into low-performance courses on the basis of perceived ability using newer versions of the standardized tests that have become *de rigeur* in assessing educational ability and outcomes. Rather than continuing to search for governmental solutions for failing urban schools, where the majority of African-American and Latino students are concentrated, the parents of poor and minority schoolchildren are encouraged to exercise their rights as consumers of public education and to improve their educational prospects by leaving failing schools and choosing those schools (whether public, private, or charter) with demonstrated levels of achievement. The 1994 Improving America's Schools Act (Public Law 89–10) also endorsed the establishment of charter schools

as means of improving teaching and student performance. This pattern of reform has greatly expanded under the 2002 No Child Left Behind Act (Public Law 107–110). NCLB has as its expressed goal to close the persistent race and class gap in educational achievement using legislative guidelines for the use of Title I funds rather than increasing the amount of federal support for public education.

Conclusion

The American educational system's foundations, as an unequal system rooted in hierarchies of race, have both inspired and limited attempts at reforms that seek to remedy the long-term effects of these inequalities. Over the course of the twentieth century, the expansion of the nation's educational system has been implicated in the ongoing construction and maintenance of a racialized social system through a variety of mechanisms, including physical separation, curricular differentiation, and unequal funding. The efforts of minority populations, both immigrant and domestic, to address these inequalities through self-help, petitioning the state, and activist traditions were central in effecting reform, but were often limited by the efforts of state and local bodies to preserve privilege within these hierarchical systems, demonstrating that schools are both the source and the means of effecting inequality in racially stratified systems.

Bibliography

- Adams, D. W. (1997). *Education for Extinction: American Indians and the Boarding School Experience, 1875–1928*. Lawrence, KS: University Press of Kansas.
- Anderson, J. D. (1988). *The Education of Blacks in the South, 1865–1930*. Chapel Hill, NC: University of North Carolina Press.
- Binet, A. and Simon, T. (1916). *The Development of Intelligence in Children*. Baltimore, MD: Williams & Wilkins Company.
- Coleman, J. (1966). United States Office of Education, and National Center for Education Statistics. *Equality of Educational Opportunity*. Washington, DC: US Department of Health, Education, and Welfare.
- Fass, P. (1989). *Outside In: Minorities and the Transformation of American Education*. New York, NY: Oxford University Press.
- Iber, J. and De León, A. (2006). *Hispanics in the American West*. Santa Barbara, CA: ABC-CLIO.
- Jewell, J. O. (2007). *Race, Social Reform, and the Making of a Middle Class: The American Missionary Association and Black Atlanta, 1870–1900*. Lanham, MD: Rowman & Littlefield.
- Kliebard, H. M. (1995). *The Struggle for the American Curriculum, 1893–1958*. New York, NY: Routledge.
- Nakanishi, D. T. and Nishida, T. Y. (1995). *The Asian American Educational Experience: A Source Book for Teachers and Students*. New York, NY: Routledge.
- Spring, J. (2007). *Deculturalization and the Struggle for Equality: A Brief History of the Education of Dominated Cultures in the United States*. New York: McGraw-Hill.
- Tyack, D. (1974). *The One Best System: A History of American Urban Education*. Cambridge, MA: Harvard University Press.
- Tyack, D. and Cuban, L. (1995). *Tinkering Toward Utopia: A Century of Public School Reform*. Cambridge, MA: Harvard University Press.

US Senate Committee on Labor and Public Welfare (1969). *Indian Education: A National Tragedy – A National Challenge*. Washington, DC: United States Government Printing Office.

Young, K. (1922). *Mental Differences in Certain Immigrant Groups: Psychological Tests of South Europeans in Typical California Schools with Bearing on the Educational Policy and on the Problems of Racial Contacts in this Country*. Eugene, OR: University of Oregon.

Further Reading

Lemann, N. (1999). *The Big Test: The Secret History of the American Meritocracy*. New York, NY: Farrar, Straus & Giroux.

Oakes, J. (1984). *Keeping Track: How Schools Structure Inequality*. New Haven, CT: Yale University Press.

Ogbu, J. (1978). *Minority Education and Caste: The American System in Cross-Cultural Perspective*. New York, NY: Academic Press.

Orfield, G. and Eaton, S. (1996). *Dismantling Desegregation: The Quiet Reversal of Brown v. Board of Education*. New York, NY: The New Press.

Patterson, J. T. (2001). *Brown v. Board of Education: A Civil Rights Milestone and Its Troubled Legacy*. New York: Oxford University Press.

Peterson, P. E. and West, M. R. (eds.) (2003). *No Child Left Behind? The Politics and Practice of School Accountability*. Washington, DC: Brookings Institution Press.

United States National Commission on Excellence in Education (1983). *A Nation at Risk: The Imperative for Educational Reform*. Washington, DC: The Commission on Excellence in Education.

Walters, P. B. (2001). Educational access and the state: Historical continuities and discontinuities in racial inequality in American education. *Sociology of Education* (Extra Issue): 35–49.

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

This page intentionally left blank

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

EDITORS-IN-CHIEF

PENELOPE PETERSON
School of Education and Social Policy
Northwestern University
Evanston, IL
USA

EVA BAKER
UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

BARRY MCGAW
McGaw Group Pty Ltd
Mont Albert Nth, VIC
Australia



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
The Boulevard Langford Lane, Kidlington, Oxford OX5 1GB, UK

Copyright © 2010 Elsevier Ltd. All rights reserved

LATVIA

Copyright © 2010 A Joma

STUDENT DATA SYSTEMS AND THEIR USE FOR EDUCATIONAL IMPROVEMENT

Copyright © 2010 J C Wayman, V Cho, and M P Richards

QUALIFICATIONS FRAMEWORKS AND THEIR ROLE IN THE REFORM OF EDUCATION AND TRAINING

Copyright © 2010 M Young

LANGUAGE OF VOCATIONAL EDUCATION AND TRAINING

Copyright © 2010 J Misko, J MacKenzie, and R-A Polvere

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44)(0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at (<http://elsevier.com/locate/permissions>), and selecting *Obtaining permissions to use Elsevier material*

Notice

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Catalog Number

A catalogue record for this book is available from the Library of Congress

ISBN: 978-0-08-044893-0

For information on all Elsevier publications
visit our website at books.elsevier.com

PRINTED AND BOUND IN THE UNITED KINGDOM

09 10 11 12 13 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

PREFACE

A preface usually provides a brief introduction to a work, intended to set the stage, provide some background insight, and whet the appetite of the reader. In our case, however, this preface has to address a fundamental question, one that was in our minds at the time we were recruited as Editors-in-Chief for the International Encyclopedia of Education. The question was “Why do we need an Encyclopedia? Its subtext was inspired by the ever-growing, ever-popular Internet. We believe that *this* Encyclopedia is desperately needed and will become a valued resource in education and associated social sciences and arts. The reasons are intellectual and procedural. Anyone with a modicum of knowledge knows that finding and trusting information gleaned from the Internet are two separate actions. The reliance on browsers to help discover references and comments result in resources based on popularity not quality. Pithy titles catch the eye and references rise in the ranks of browser searchers. Related to this is the “editing” in the Internet realm of populist efforts at encyclopedia, references, and other compilations. Once again, after removing offensive material, the accuracy, completeness, lack of bias, and other provenance for entries simply do not exist. Experienced researchers in education can sort through and make intelligent choices. Novices and many journeyman, or practitioners, parents, and policy makers cannot. Contrast how this Encyclopedia was built. Key domains of educational research were identified, and a tentative list of sub-domains or useful applied areas was posited. Then the Editors-in-Chief (apologies for the awkwardness of the term) identified the leading researcher in a particular domain, and with surprisingly little effort, recruited them to participate. They in turn identified the two best researchers in a sub-domain, such as formative assessment or the training of pre-school teachers. The authors of the sections of the Encyclopedia do not represent a collective group of friends and acquaintances, although friendships have been made. Rather they embody a deep and broad scholarly community. The difference from compiled Internet resources is the built-expertise and intellectual engagement of the authors. The summary of the developments and futures in their personal areas of scholarship have been filtered through their years of experience, both as scholars and communicators. Quality, then, is endemic to each piece, developed through this top-down identification of expertise, and made indelible by the bottom-up application of high standards from people leading the sub-domains – the authors, and the domains themselves, the section editors.

On a procedural level, the publishers early committed to the notion that this Encyclopedia would also be an online resource, and access would be available through print, for those with strong bookcases and the persisting love of turning real pages. The Internet version will allow multiple prisms through which the reader may access articles and provide, as it were, an emulation of the Internet in our field, albeit bounded by expertise and high quality.

What must be underscored in the assessment of this effort are the Editors-in-Chief and the publishers’ commitment to find excellence worldwide. We tried very hard to persuade notable scholars from all parts of the world to make contributions. Less than to fulfill the title of “International,” we were on the hunt for perspectives that would enrich the scope and depth of the sections. Our section editors put in enormous time attempting to find the best in the field, wherever they resided. Yet, not everyone is in the volume. Some were overcommitted. Many were not fully confident of their English, and the automated translation software has not yet met standards for technical writing. We believe that such writing and editing tools will make the outreach to an even broader International group of scholars possible in future revisions, or online updates. Furthermore, the birth of the World Educational Research Association (in 2009) will provide a better set of interlocking networks to find and evaluate scholarship from any place on the globe.

Finally, the scope of the effort must be acknowledged: 28 section editors, 926 articles were commissioned, drafted, reviewed, redrafted, edited, and put together in the space of four years. The publishers underwent some internal changes, and alterations in management. We as Editors-in-Chief, changed roles, moved, and also had to keep our own research and development enterprises afloat. Deadlines wobbled; authors dropped from view and had to be replaced.

Yet, at times frustrating as all development is, we find the final product exhilarating. We are enthusiastic not simply because it came into being at all, but because the collective light of the minds that wrote have left a bright resource for the future, one that will impact the way our colleagues understand and experience the educational knowledge, improvement, and impact in the future.

Eva Baker
Penelope Peterson

EDITORS

EDITORS-IN-CHIEF

Penelope Peterson

School of Education and Social Policy
Northwestern University
Evanston, IL
USA

Eva Baker

UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

Barry McGaw

McGaw Group Pty Ltd
Mont Albert Nth
Melbourne, VIC
Australia

SECTION EDITORS

Walter Allen

HEOC Division
University of California, Los Angeles
Los Angeles, CA
USA

Vibeke Grøver Aukrust

Institute for Educational Research
University of Oslo
Oslo
Norway

Dominic J Brewer

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

Kai-Ming Cheng

Faculty of Education
The University of Hong Kong
Hong Kong

Gregory Chung

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Penny Enslin

Department of Educational Studies
University of Glasgow
Glasgow
UK

Noelle Griffin

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Dennis Hocevar

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

David Hopkins

Institute of Education
University of London
London
UK

Mary James

Faculty of Education
University of Cambridge
Cambridge
UK

Sanna Järvelä

Department of Educational Sciences and
Teacher Education
University of Oulu
Oulu
Finland

Thomas Karmel

National Centre for Vocational Education Research
Adelaide, SA
Australia

Harold G Levine

School of Education
University of California, Davis
Davis, CA
USA

Robert L Linn

University of Colorado at Boulder
Ouray, CO
USA

Allan Luke

Faculty of Education
Queensland University of Technology
Brisbane, QLD
Australia

Ann M Mastergeorge

Human Development and Family Studies
University of California, Davis
Davis, CA
USA

Patrick J McEwan

Department of Economics
Wellesley College
Wellesley, MA
USA

Barbara Means

Center for Technology in Learning
SRI International
Menlo Park, CA
USA

Ingrid Moses

University of Canberra
Canberra, ACT
Australia

Bill Mulford

Faculty of Education
University of Tasmania
Hobart, TAS
Australia

Yu Jing Ni

Faculty of Education
The Chinese University of Hong Kong
Hong Kong

Festus Obiakor

Department of Exceptional Education
University of Wisconsin-Milwaukee
Milwaukee, WI
USA

P David Pearson

Graduate School of Education
University of California, Berkeley
Berkeley, CA
USA

Colin Power

Graduate School of Education
University of Queensland
Brisbane, QLD
Australia

Kjell Rubenson

Faculty of Education
The University of British Columbia
Vancouver, BC
Canada

Sandip Sinharay

Education Testing Services
Princeton, NJ
USA

Collette Tayler

Queensland University of Technology
Brisbane, QLD
Australia

Nico Verloop

ICLON, Graduate School of Education
Leiden University
Leiden
Netherlands

HOW TO USE THE ENCYCLOPEDIA

The International Encyclopedia of Education is intended for use by students, research professionals, and interested others. Articles have been chosen to reflect major disciplines in the study of education and common topics of research by academics in this domain. Each article serves as a comprehensive overview of a given area, providing both breadth of coverage for students, and depth of coverage for research professionals. We have designed the encyclopedia with the following features for maximum accessibility for all readers.

The contents of the encyclopedia are arranged alphabetically by section, and within sections, alphabetically by article. The Subject Index is located in Volume 8. Some topics are covered in a multitude of articles from differing perspectives, while other topics may have only one entry. We encourage use of the index for access to a subject area, rather than use of the Contents list alone, so that a reader has a full notion of the coverage of that topic.

The articles include cross-references to other related encyclopedia articles, suggested further readings where applicable, and many contain relevant websites for additional information. We encourage readers to use the cross-references to locate other encyclopedia articles that will provide more detailed information about a subject.

The Further Reading sections include recent secondary sources to aid the reader in locating more detailed or technical information. Review articles and research articles that are considered of primary importance to the understanding of a given subject area are also listed. These suggested further readings are not intended to provide a full reference listing of all material covered in the context of a given article, but are provided as next steps for a reader looking for additional information.

This page intentionally left blank

CONTENTS

Preface	v–vi
Editors	vii–viii
How to use the Encyclopedia	ix

VOLUME 1

A

Adult Education

Adult Education Overview	<i>K Rubenson</i>	1
--------------------------	-------------------	---

Adult Education – Adult Learning, Instruction and Program Planning

Adult Learning	<i>S B Merriam</i>	12
Adult Learning and Instruction: Transformative Learning Perspectives	<i>P Cranton</i>	18
Adult Learning in a Biographic Perspective	<i>L West</i>	25
Adult Learning, Instruction and Programme Planning: Insights from Freire	<i>P Mayo</i>	31
Characteristics of Adult Learning	<i>K Illeris</i>	36
Informal Learning: A Contested Concept	<i>P Hodgkinson</i>	42
Organizational Learning	<i>P-E Ellström</i>	47
Program Planning	<i>A L Wilson and R M Cervero</i>	53
Workplace Learning Frameworks	<i>S Billett</i>	58

Adult Education – as a Field of Study

Class Analysis in Adult Education	<i>T Nesbit</i>	64
Gender Analysis	<i>J Stalker</i>	70
Race and Ethnicity in the Field of Adult Education	<i>J Johnson-Bailey and D Drake-Clark</i>	77
Trends in Workplace Learning Research	<i>T Fenwick</i>	83

Adult Education – Concepts

Lifelong Learning	<i>J Field</i>	89
Rewriting the History of Adult Education: The Search for Narrative Structures	<i>B J Hake</i>	96

Adult Education – Domains and Provision

Adult Basic Education: A Challenge for Vocational Based Learning	<i>R Höghelm</i>	102
Adult Literacy Education	<i>L Tett and R St.Clair</i>	107
Citizenship and Immigrant Education	<i>D Schugurensky</i>	113
Community Based Adult Education	<i>B Connolly</i>	120
Continuing Professional Education: Multiple Stakeholders and Agendas	<i>R M Cervero and B J Daley</i>	127
Financing of Adult and Lifelong Learning	<i>Å Sohlman</i>	133
Labor Education	<i>K Forrester</i>	139
Learning Cities and Regions	<i>C Duke</i>	144
Museums as Sites of Adult Learning	<i>R S Grenier</i>	150
Overview of Lifelong Learning Policies and Systems	<i>D-B Kwon and C H O Daeyeon</i>	156
Popular Adult Education	<i>S Tøsse</i>	162
Provision of Prior Learning Assessment	<i>P Andersson</i>	169
The Age of Learning: Seniors Learning	<i>P Jarvis</i>	175
University Adult Continuing Education: The Extra-Mural Tradition Revisited	<i>M Slowey</i>	183

Adult Education – Economy and Society

Adult Education and Civil Society	<i>J Crowther and I Martin</i>	191
Adult Education and Nation-Building	<i>S Walters</i>	198
Economic Outcomes of Adult Education and Training	<i>A Ferrer and W C Riddell</i>	204
Health and Adult Learning	<i>P Bélanger and M Robitaille</i>	211
Modernization Processes and the Changing Function of Adult Learning	<i>H S Olesen</i>	217
The Political Economy of Adult Education	<i>G Rees</i>	223
Wider Benefits of Adult Education	<i>R Desjardins and T Schuller</i>	229

Adult Education – Participation

Barriers to Participation in Adult Education	<i>K Rubenson</i>	234
Participation in Adult Learning	<i>R Desjardins</i>	240

C**Comparative Education**

Comparative and International Education: English-Language Scholarly Publications	<i>D Post and M Fedri</i>	251
Comparative Education: Philosophical Issues and Concepts	<i>M Mason and C W Evers</i>	257
Comparative Education: Societies and Associations	<i>M Bray</i>	266

Curriculum Development

Curriculum and Religion	<i>C Kapitzke and P Graham</i>	272
Gender and Curriculum	<i>J Marsh</i>	277
Local Teacher-Based Curriculum Development	<i>M D Osborne and H B Barker</i>	283

Curriculum Development – Contexts

Curriculum and Human Rights	<i>C Lohrenscheit</i>	287
Curriculum and the Education of Cultural and Linguistic Minorities	<i>S May</i>	293
Curriculum in Postcolonial Contexts	<i>A Hickling-Hudson</i>	299
Curriculum, Economic and Cultural Development	<i>A Rampal</i>	307
Globalization and Curriculum	<i>R Dale</i>	312
Popular Culture	<i>K Schultz and R Throop</i>	318

Curriculum Development – Evaluation and Research

An Overview of Research in Curriculum Inquiry	<i>F M Connelly and S J Xu</i>	324
Curriculum Evaluation: Approaches and Methodologies	<i>V Klenowski</i>	335
Curriculum Studies, Discourse Analysis, and the Construction of Historical Time	<i>J Laspina</i>	342
Curriculum Use in the Classroom	<i>M Ben-Peretz and B Eilam</i>	348
The Roots and Routes of Teacher-Based Action Research and Curriculum Inquiry: An Historical Perspective	<i>L G Putney and J L Green</i>	355

Curriculum Development – Planning and Development

Curriculum and Syllabus Design	<i>A Woods, A Luke, and K Weir</i>	362
Curriculum and the Publishing Industry	<i>J Larson, A-R Allen, and D Osborn</i>	368
Curriculum and Teacher Change	<i>J G Ladwig</i>	374
Curriculum Governance and Planning	<i>B Levin</i>	379
Curriculum Planning and Systems Change	<i>Z Deng</i>	384
Curriculum, Digital Resources and Delivery	<i>Y Zhao, G Zhang, and C Lai</i>	390
Textbook Development and Selection	<i>P Shannon</i>	397

Curriculum Development – Subject Matter

Curriculum and Globalization: Higher Education	<i>K H Mok</i>	403
Curriculum Development as Subject Matter: Social Studies	<i>J Brophy and J Alleman</i>	411
Curriculum Development in the Area of Reading	<i>C L Madda and T E Raphael</i>	415
Curriculum in the Arts	<i>C M Thompson, L Bresler, and T E Costantino</i>	423
Curriculum: Foreign Language Learning	<i>Angel M Y Lin</i>	428
First Language Learning	<i>T Rogers and L Moffatt</i>	435
Learning a Second Language in First Language Environments	<i>E Bernhardt and D J Tedick</i>	441
Literature	<i>S Hynds</i>	447
Mathematics	<i>B Kaur and C P Vistro-Yu</i>	453
Physical Education and Sports	<i>D Kirk</i>	459
Primary and Elementary/Middle Grades Reading	<i>D E Alvermann and C A Mallozzi</i>	464
Science	<i>Y Kali and M Linn</i>	468

Curriculum Development – Theory

Bilingual Learning (Learning L1 and L2 in an L1 and L2 Environment)	<i>D Schwinge</i>	475
Curriculum and Complex Systems Theory	<i>W-M Roth and J Thom</i>	481

Curriculum and Constructivism	<i>B Davis and D Sumara</i>	488
Curriculum and Critical Theory	<i>L Yates</i>	494
Curriculum and Poststructuralist Theory	<i>J L Miller</i>	499
Curriculum and Structuralist Sociology: The Theory of Codes and Knowledge Structures	<i>J Muller and J Gamble</i>	505
Curriculum Reform	<i>R Gilbert</i>	510
Early Childhood Curriculum and Developmental Theory	<i>S M-Y Lim and C Genishi</i>	514

D

Demography and Social Change

An Evaluation Framework for E-Learning Effectiveness in the Arab World	<i>W AbuSneineh and M Zairi</i>	521
Globalization and Social Justice in Higher Education	<i>E Kimura-Walsh</i>	536
Racial Inequality and the Economic Payoffs of Higher Education	<i>S A Lachtman</i>	542
School Health Education	<i>J S Yang</i>	547

Demography and Social Change – Demography

A Paradigm of Contradictions: Racism and Science Education	<i>B A Brown and J W Mutegi</i>	554
Africana Studies: Past, Present And Future	<i>B R Hare, A Young, J B Stewart, and D P Aldridge</i>	565
Beyond Black Nationalism to Black Internationalism: The Compelling Case for Examining Black Educational Challenges Globally	<i>K Freeman</i>	573
Beyond Racial, Ethnic, and Gender Bias in Education Statistics	<i>J T Lott and C Bennett</i>	581
Caste and Education in India	<i>G G Wankhede</i>	589
Demography, Community, and the Education of Asian American and Pacific Islander Students	<i>R T Teranishi</i>	595
Educational Attainment and Mortality Differentials	<i>H D Horton, L Hixson, and S M Monnat</i>	602
Educational Opportunity and Latino/Chicano College Choice: New Findings, and Theoretical Perspectives	<i>M A Ceja and P A Pérez</i>	610
Ethnic Minority Identity and Educational Outcomes in a Rising China	<i>G A Postiglione</i>	616
Ethnic Studies and Women's Studies: From the Past to the Present, Providing a Space in the Academy for Scholarship Focused on the Histories and Experiences of Marginalized People	<i>C A Brown</i>	623
Gender Equity in Higher Education: Challenges and Celebrations	<i>L Morley</i>	629
Locating Space and Place in the College Access Debate: New Tools for Mapping and Understanding Educational Inequity and Stratification	<i>B Cantwell and J F Milem</i>	636
Public Policy and Inequality in Postsecondary Opportunity: Educational Statistics and the Failure of Education Reform	<i>E P St. John, K Williams, and K Moronski</i>	649
The Education of Indigenous Students	<i>J W Tippeconnic and S C Faircloth</i>	661
The Health Advantages of Educational Attainment	<i>G Y Dinwiddie</i>	667
To Go or Not to Go: Access and Barriers to Ethnic Minority Participation in Higher Education within the United Kingdom	<i>P Brug and T L Oliver</i>	673
Universal Secondary Education and the Two-Tiered School System in the English-Speaking Caribbean	<i>P Anderson and J Devonish</i>	679
Women of Color in Science, Technology, Engineering and Mathematics (STEM): Refining the Concepts, Reframing the Issues	<i>C B Leggon</i>	686

Demography and Social Change – Social Change

Educational Diversity in U.S. Law Schools	<i>C E Daye, A T Panter, W R Allen, M E Deo, and L F Wightman</i>	691
Educational Reform	<i>F Lopez-Segrera</i>	698
Emergence of For-profit Higher Education	<i>X Li</i>	705
Engaged Education: Experiential Learning, Intensive Field Experiences, and Social Change	<i>T R Rone</i>	711
Higher Education and Social Change in Mexico	<i>A Mungaray, G Estrella, P Moctezuma, and M Ramírez-Urquidy</i>	719
The Changing Role of the State in Higher Education	<i>K Gürüz</i>	728
The Role of the State in School Reform: Responding to Increasing Ethnic Diversity in Britain	<i>P Ratcliffe</i>	737
Transforming Mindsets Through Education for Sustainable Development	<i>V M Segovia</i>	746

Demography and Social Change – Sociology

Achievement Motivation in Ethnic Minority Youth	<i>S Graham and E Bryant</i>	753
Affirmative Action and Higher Education in Brazil	<i>J T dos Santos and D M Queiroz</i>	760
Black in White: Black Students at White and Black Colleges	<i>L Evans and J R Feagin</i>	767
Bologna Process: On the way to a Common European Higher Education Area	<i>A V Rezaev</i>	772
Critical Race Theory in Education Research	<i>R Vue and C B Newman</i>	779
Cultural Diversification and Japanese Education: Social Constructions of the New Diversity	<i>R Tsuneyoshi</i>	787
Education in War Countries: A Disempowerment of Society	<i>J L Nkata</i>	793
Educational Attainments of U.S. Black Males and Females: 1971 to 2003	<i>F B Bonner, M-C Jipguep-Akhtar, and R J Harrison</i>	799
Educational Preparation: Fostering the Self-Efficacy and Resilience of Urban Adolescent Youth	<i>M B Spencer and B Tinsley</i>	806
HIV Education for Low-Literate People: Transforming Students and Communities through Paulo Freire's Praxis and the Pedagogy of Action	<i>N Z Haniff</i>	814
How Are Discussions of Interdisciplinary Studies Linked to Diversity Discourse?	<i>C N Spence</i>	823
Identifying Patterns of Doctoral Attrition Across Academic Fields	<i>A James, K J Fosnacht, E B Stolzenberg, and D Thirumalai</i>	827
Male Underachievement in Education Across the Globe: A Shift in Paradigm for Gender Disparities Regarding Academic Achievement	<i>J F L Jackson, J L Moore III, and R A Leon</i>	838
Noncognitive Measures for Higher Education Admissions	<i>W E Sedlacek</i>	845
Peace Education	<i>J Page</i>	850
Remote Control: Africa's Development and Outflow to Developed Nations of Technical Talent	<i>F Haniff</i>	855
Scholar-Baller®: Student Athlete Socialization, Motivation and Academic Performance in American Society	<i>C K Harrison, S Bukstein, J Mottley, E Comeaux, J Boyd, C Parks, and D Heikkinen</i>	860
Service-Learning and Diversity as Mediums to Foster Civic and Community Engagement in Undergraduates	<i>L Oseguera</i>	866
Success and Failure in Education and Criminal Justice: Identifying Common Mechanisms	<i>S Ramirez and A Harris</i>	874
The History of Education: Race and Education	<i>J O Jewell</i>	881

VOLUME 2**E****Early Childhood Care and Education Theory**

- Early Childhood Care and Education: The Family, The Market, and The State *S B Kamerman and M J Neuman* 1
- Gender Issues in Early Childhood Education and Care *B Pocock* 11

Early Childhood Education and Care Governance

- Early Childhood Risk, Protection and Abuse Prevention *E Khoo* 17
- Evaluating Early Childhood Education and Care Programs *M Janus and S Brinkman* 25

Early Childhood Education and Care Investment

- Investing in Early Childhood Education and Care: The Health and Wellbeing Case *T Moore and F Oberklaid* 32
- Cost-effective Early Childhood Programs from Preschool to Third Grade *A J Reynolds, J A Temple, and B White* 38
- Early Childhood in Post-Modern Cultures: Thoughts and Some Concerns *P Gammage* 49
- Investing in Early Childhood Education and Care: Some Policy Implications *J Bennett* 55
- Investing in Early Childhood Education and Care: The Economic Case *G Cleveland and M Krashinsky* 63

Early Childhood Education and Care Programs

- Early Care and Education Programs for Infants and Toddlers *B Nyland* 69
- Literacies in Early Childhood: The Preschool Period *M Flear and B Raban* 75
- Participation in Early Childhood Education and Care Programs: Equity, Diversity and Educational Disadvantage *M Vandenbroeck* 81
- Teaching in Early Childhood Centers Instructional Methods and Child Outcomes *I Siraj-Blatchford* 86

Early Childhood Education and Care Provision

- Early Childhood Education and Care in the People's Republic of China *Zeng Xiaodong* 93
- Home-Based and Institutional Early-Childhood Education and Care Services *R Fukkink and H Blok* 98
- Integrated Early Childhood Education and Care Services – Care, Upbringing, Education and Health *S Colley* 104
- Issues of Access and Program Quality *J T Hustedt and W S Barnett* 110
- Kindergarten Transitions and Linkages to Primary School – Readiness Re-conceptualized *A Petriwskyj* 120
- Parent Support in Early Childhood – Approaches and Outcomes *K S Howard, K A Beckmann, and J Brooks-Gunn* 126

Early Childhood Education and Care Research

- Development and Implementation of Early Learning Standards in the United States *C Scott-Little* 132
- Global Status of Early Learning and Development Standards *P R Britto and S L Kagan* 138
- Rethinking Early Childhood Education and Care: Implications for Research and Evaluation *K Thorpe, D Cloney, and C Tayler* 144

Early Childhood Education and Care Theory

- Child Rearing and Early Education: Parents and Professionals: Theoretical and Historical Influences *H Penn* 151

Early Childhood Education and Care Workforce

- Early Childhood Education and Care Professional Profiles in the United Kingdom: Multiagency Teamwork and Transdisciplinary Challenges *A Anning and P Calder* 157
- Early Childhood Education and Care Staff Support: Mentoring and Professional Development in Finland *P Lindberg* 162

Economics of Education

- Education and Health *P Muennig* 169
- Data *M Perez and M Socias* 178
- Empirical Research Methods in the Economics of Education *P J McEwan* 187
- Theoretical Concepts in the Economics of Education *D J Brewer, G C Hentschke, and E R Eide* 193

Economics of Education – Costs and Financing of Schooling

- Cost–Benefit Analysis and Cost–Effectiveness Analysis *C Belfield and H M Levin* 199
- Economic Approaches to Adequacy *J Imazeki* 204
- Economic Approaches to School Efficiency *L L Taylor* 210
- School Finance Reform *T Downes* 216
- School Finance: An Overview *J K Rice, D H Monk, and J Zhang* 222
- The Economics of Tuition and Fees in Higher Education *R G Ehrenberg* 229

Economics of Education – External Benefits of Schooling

- Education and Civic Engagement *T S Dee* 235
- Education and Crime *L Lochner* 239
- Education and Economic Growth *E A Hanushek and L Wößmann* 245
- Parental Socioeconomic Status, Child Health, and Human Capital *J Currie and J Goodman* 253
- The External Benefits of Education *W W McMahon* 260

Economics of Education – Labor Markets, Education and Earnings

- Education and Inequality *J Blanden and S Machin* 272
- Human Capital *E R Eide and M H Showalter* 282
- Race Earnings Differentials *M Carnoy* 288
- Returns to Education in Developed Countries *M Gunderson and P Oreopoulos* 298
- Returns to Education in Developing Countries *H A Patrinos and G Psacharopoulos* 305
- School Quality and Earnings *J R Betts* 313
- Signaling in the Labor Market *M E Page* 321
- The Economics of High School Dropouts *R W Rumberger* 325

Economics of Education – Markets, Incentives and School Choice

- Competition and Student Performance *A A Payne* 331
- Educational Privatization *C Belfield and H M Levin* 337

Student Incentives	<i>E R Eide and M H Showalter</i>	342
The Economic Role of the State in Education	<i>D N Plank and T E Davis</i>	347
The Economics of Catholic Schools	<i>W Sander and D Cohen-Zada</i>	354
The Economics of Charter Schools	<i>R Zimmer and R Buddin</i>	360
The Economics of Parental Choice	<i>R Chakrabarti and J Roy</i>	367
The Economics of School Accountability	<i>D N Figlio and H F Ladd</i>	374
The Efficacy of Educational Vouchers	<i>R Zimmer and E P Bettinger</i>	380
Tiebout Sorting and Competition	<i>T J Nechyba</i>	388

Economics of Education – Production of Schooling

Desegregation, Academic Achievement and Earnings	<i>S Rivkin</i>	394
Education Production Functions: Concepts	<i>D N Harris</i>	402
Education Production Functions: Developed Country Evidence	<i>E A Hanushek</i>	407
Education Production Functions: Evidence from Developing Countries	<i>P Glewwe and S Lambert</i>	412
Family Environment in the Production of Schooling	<i>R Rothstein</i>	423
Neighborhoods and Peers in the Production of Schooling	<i>J Vigdor and J Ludwig</i>	431
Teacher Quality in Education Production	<i>I Pelayo and D J Brewer</i>	438
The Economics of Class Size	<i>D W Schanzenbach</i>	443
The Economics of Early Childhood Interventions	<i>M Nores</i>	450

Economics of Education – Teaching Labor Markets

An Overview of Teacher Labor Markets	<i>H Lankford and J Wyckoff</i>	456
Compensating Differentials in Teacher Labor Markets	<i>J Chambers</i>	465
Economic Approaches to Teacher Recruitment and Retention	<i>S Loeb and J Myung</i>	473
Teacher Incentives	<i>L Santibañez</i>	481
Teacher Supply	<i>P Dolton</i>	489
Teacher Training and Preparation in the United States	<i>M Podgursky, M G Springer, and R Hutton</i>	498
Teachers in Developing Countries	<i>F H Rogers and E Vegas</i>	504
The Economics of Teachers' Unions in the United States	<i>K Strunk</i>	512

Education of Children with Special Needs

Adapted Physical Activity for Students with Special Needs	<i>S R Hodge</i>	519
Assessment Accommodations for Children with Special Needs	<i>R J Kettler and S N Elliott</i>	530
Assistive Technology and Educational Practice	<i>H P Parette and G R Peterson-Karlan</i>	537
Augmentative–Alternative Communication	<i>C Mason and K M Chinn</i>	544
Community Integration and Employment of Youth with Special Needs	<i>B A Ford and A W Daviso III</i>	550
Cooperative Learning for Children with Special Needs	<i>E A Dooley and K R Kossar</i>	555
Co-Teaching	<i>E D Cramer</i>	560
Cultural and Linguistic Diversity and Children with Special Needs	<i>J M Patton, N Day-Vines, and A McSwain</i>	565
Curriculum-Based Assessment and Students with Special Needs	<i>D J O'Shea and J D Katsafanas</i>	571

Diagnosis and Classification	<i>M Sapp</i>	577
Direct Instruction and the Education of Children with Special Needs	<i>T Mehring</i>	584
Disabilities and Gifted Education	<i>V I Daniels and M J McCollin</i>	588
Early Childhood Special Education: Birth to Eight Years Old	<i>T S Lovelace, R Gardner III, J B Axe, and D M Sainato</i>	594
Early Intervention for Students with Special Needs	<i>P A Grant</i>	608
Early Transition of Children with Special Needs	<i>S A Fowler</i>	613
Educating Children Who are Deaf-Blind	<i>J Wood and K M Chinn</i>	621
Educating Children with Deafness and Hearing Impairments	<i>G M Mukuria and C J Eleweke</i>	628
Educating Children with Other Health Impairments (ADHD)	<i>G M Mukuria and F E Obiakor</i>	634
Educating Children with Physical Disabilities	<i>Z P Nkabinde, F E Obiakor, M T Ofor, and D J Smith</i>	640
Educating Children with Specific Learning Disabilities	<i>P A Grant and P B Grant</i>	646
Educating Children with Speech and Language Disorders	<i>S L Green and C D Qualls</i>	654
Educating Students with Autism and Related Disorders	<i>T T Dyches</i>	661
Educating Students with Cognitive Disabilities	<i>J P Bakken</i>	669
Educating Students with Emotional and Behavioral Disorders	<i>S W Smith and G G Taylor</i>	678
Educating Students with Special Needs: An Overview	<i>F E Obiakor, M K Harris, P Harris-Obiakor, and G M Mukuria</i>	688
Educating Students with Traumatic Brain Injury	<i>S Burkhardt and A Rotatori</i>	695
English Language Learners with Special Needs	<i>R Rueda and G Ragusa</i>	701
Functional Behavioral Assessment	<i>J P Bakken</i>	708
History of Special Education	<i>L M Garrick Duhaney and S J Salend</i>	714
Inclusion of Students with Special Needs in General Education Classrooms	<i>M A Prater</i>	721
Instructional Accommodations for Children with Special Needs in Inclusive Settings	<i>A L Yurick, S Musti-Rao, and G Cartledge</i>	727
Interagency Agreements	<i>L A Owens</i>	733
Interdisciplinary Collaboration in Early Childhood	<i>N File</i>	739
Language Development in Children with Special Needs	<i>M S L Stewart</i>	745
Large Scale Assessment and Accountability and Students with Special Needs	<i>M L Thurlow</i>	752
Literacy Instruction for Students with Special Needs	<i>L H Mason</i>	759
Math Instruction for Children with Special Needs	<i>B A Bottge</i>	767
Overrepresentation of Culturally and Linguistically Diverse Students in Special Education	<i>S C Trent</i>	774
Paraprofessionals in Special Education	<i>B Y Ashbaker, S C Dunn, and J Morgan</i>	780
Parent and Family Involvement in the Education of Children with Special Needs	<i>L P Prater</i>	788
Peer Relations and Socialization of Children and Adolescents with Special Needs	<i>L M Bullock, M Wong-Lo, and R A Gable</i>	794
Peer-Tutoring	<i>L J O'Shea and D J O'Shea</i>	802
Postsecondary Participation of Students with Special Needs	<i>R A Stodden and K D Roberts</i>	808
Prereferral Strategies	<i>C G Simpson and J P Bakken</i>	815

Preventing Antisocial Behavior and Delinquency: A Comprehensive School-wide Approach <i>P R Hallam, K R Young, P Caldarella, D G Wall, and L Christensen</i>	820
Programs and Instructional Strategies for Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	829
Related Services for Children with Special Needs <i>S O Obi</i>	834
Risk and Resilience Frameworks in Understanding Special Education <i>W-Y Hsieh and M L Donahue</i>	840
School Dropout Prevention <i>F E Obiakor</i>	845
School-Based Services for Children with Special Needs <i>S O Obi</i>	850
Self-Determination <i>M L Wehmeyer</i>	857
Social Competence <i>L A Owens and S Johnston-Rodriguez</i>	865
The Identification of Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	870
Transition from School to Adult Life <i>M J McCollin and F E Obiakor</i>	876
Visual Impairment <i>W Sapp</i>	880
Workforce Issues in Special Education <i>E Carlson and B Billingsley</i>	886

VOLUME 3

Education of Professionals

An Overview of Accreditation, Certification, and Licensure Processes <i>J J Norcini and M van Zanten</i>	1
Educating America's Military Officers <i>M H Adamshick</i>	7
Education for Architecture in the United States and Canada <i>S Schreiber</i>	13
Education for Medicine <i>M B Anderson</i>	19
Education for the Legal Profession <i>J W Wegner</i>	25
Education of Clergy <i>D O Aleshire</i>	31
Framing Professions Education <i>P S O'Sullivan</i>	37
Issues in Accounting/Business Education <i>W H Gijsselaers and R G Milter</i>	39
Professions – Engineering <i>J Heywood</i>	45
Professions – Health-Related <i>G M Jensen and C B Royeen</i>	51
Professions – Pharmacy <i>D E Beck</i>	58
Teaching for the Complex Practice of Nursing <i>P Benner and M Sutphen</i>	65
Veterinary Education <i>K M Edmondson</i>	71

Education Research Methodology: Quantitative Methods and Research

Abductive Research Methods <i>B D Haig</i>	77
Field Experimentation in Education <i>M A Constan</i>	83
Internal Validity <i>M H Clark and S C Middleton</i>	90
Methods for Approximating Random Assignment <i>B Schneider and S-K McDonald</i>	97
Methods for Approximating Random Assignment: Regression Discontinuity and Propensity Scores <i>L M Scheier</i>	104

Multitrait–Multimethod Designs	<i>M Bong and S Hong</i>	111
Multivariate Longitudinal Data Analysis	<i>H Liu and L Chang</i>	120
Quasi-Experimentation: Two Group Design	<i>E Hong</i>	128
Regression Discontinuity Designs	<i>A Kelava, S Rohrmann, and V Hodapp</i>	134
Sampling	<i>R S Brown</i>	142
Statistical Conclusion Validity	<i>P L Busk</i>	147
Survey Research Methods	<i>A Fink</i>	152

Educational Assessment

An Overview of Educational Assessment	<i>M James</i>	161
Assessment and the Evaluation of Institutional Effectiveness	<i>S M Thomas</i>	172
Assessment in Schools – Dispositions	<i>R D Crick</i>	181
Assessment in Schools Related To Literacy: Reading	<i>K J Topping</i>	189
Impact of Assessment on Students' Learning Strategies and Implications for Judging Assessment Quality	<i>M Segers, F Dochy, and D Gijbels</i>	196

Educational Assessment – Assessment Activities

Alternative Assessment	<i>G Stobart and C Gipps</i>	202
Classroom Assessment Tasks and Tests	<i>J Phelan and J Phelan</i>	209
Dynamic Assessment	<i>J G Elliott, E L Grigorenko, and W C M Resing</i>	220
Instructional System Provided Feedback	<i>H F O'Neil, S Chuang, and T Huang</i>	226
Peer and Self-assessment	<i>C Harrison</i>	231
Portfolio Assessment	<i>V Klenowski</i>	236
Records of Achievement: Beyond Traditional Tests	<i>P Broadfoot</i>	243

Educational Assessment – Assessment in Domains

Assessment in Higher Education	<i>D R Sadler</i>	249
Assessment in Schools – Affective Domain	<i>M Ventura, R D Roberts, and M Zeidner</i>	256
Assessment in Schools – Creative Subjects	<i>T E Costantino and L Bresler</i>	262
Assessment in Schools – Learning to Learn	<i>A Hautamäki, J Hautamäki, and S Kupiainen</i>	268
Assessment in Schools – Literacy Writing (Extended)	<i>J M Parr and L Earl</i>	273
Assessment in Schools – Mathematics	<i>M Brown and J Hodgen</i>	279
Assessment in Schools – Oracy	<i>A L Bailey</i>	285
Assessment in Schools – Primary Science	<i>W Harlen</i>	293
Assessment in Schools – Secondary Science	<i>J Shemwell, A Fu, M Figueroa, R Davis, and R Shavelson</i>	300
Assessment in Schools – Technology Education and ICT	<i>A Jones, B Cowie, and J Moreland</i>	311
Assessment in the Early Years	<i>M J Drummond</i>	316
Assessment in the History, Civics and Social Studies Domains	<i>P G Avery</i>	323
Assessment in the Workplace of Performance, Developing Expertise and Competence	<i>A J Brown</i>	330
Assessment in Vocational Education	<i>K Ecclestone</i>	337

Educational Assessment – Concepts and Issues

Assessing Group Work	<i>M Galton</i>	342
Assessment and the Regulation of Learning	<i>L Allal</i>	348
Challenges of Developing and Implementing Formative Assessment Practices in Schools	<i>R Sutton</i>	353
Formative Assessment	<i>P Black</i>	359
Formative Assessment and Instructional Planning	<i>B Marshall</i>	365
Progression and Assessment: Developmental Assessment	<i>M Forster and G Masters</i>	369
Students' Rights in Assessment and Evaluation	<i>J Gardner and B Holmes</i>	378
Summative Assessment by Teachers	<i>R Daugherty</i>	384
The Multiple Purposes of Assessment	<i>P E Newton</i>	392
The Relationship between Assessment and the Organization and Practice of Teaching	<i>L B Resnick, D William, R Apodaca, and E S Rangel</i>	397

Educational Assessment – Enhancing Quality and Use

Assessment Conversations: Reading and Writing Conferences with Students and with Parents	<i>M Barrs</i>	403
Assessment: Pre-service and In-service Teacher Education	<i>A Davies</i>	410
Classroom Assessment in Policy Context (Australia)	<i>J J Cumming</i>	417
Classroom Assessment in Policy Context (England and France)	<i>M Osborn and M Raveaud</i>	425
Classroom Assessment in Policy Context (French Sources)	<i>D Laveault</i>	432
Classroom Assessment in Policy Context (Hong Kong)	<i>D Carless</i>	438
Classroom Assessment in Policy Context (New Zealand)	<i>T J Crooks</i>	443
Classroom Assessment in Policy Context (Sub-Saharan Africa)	<i>S J Howie, L Zimmerman, and K Draper</i>	449
Moderation of Student Work by Teachers	<i>G S Maxwell</i>	457
School Policies and Practices to Support Effective Classroom Assessment for Learning	<i>D Pedder</i>	464
Student Assessment: Policy and Practice in Eastern Europe	<i>G Bethell</i>	472

Educational Assessment – Key Relationships

Assessment Practice in Policy Context: Latin American Countries	<i>F M Rizo</i>	479
Impact of Assessment on Learner Groups (Boys/Girls)	<i>J Elwood</i>	486
Impact of Assessment on Learner Groups (Disabilities)	<i>M Rouse</i>	492
Impact of Assessment on Students' Test Anxiety	<i>P Buchwald and C Schwarzer</i>	498
Impact of Assessments on Classroom Practice	<i>J L Herman</i>	506
Social Practices in School Assessment and their Impact on Learner Identities	<i>A Pollard and A Filer</i>	512

Educational Evaluation – Concepts, Practice, and Future Directions

Educational Evaluation: Concepts, Practice, and Future Directions	<i>Y J Ni</i>	518
---	---------------	-----

Educational Evaluation – Contemporary Themes in Evaluation

Evaluating E-Learning	<i>D D Williams and C R Graham</i>	530
Evaluating Schools as Learning Communities	<i>S M Thomas</i>	539
Evaluation of Adult Education and Training Programs	<i>G McNamara, P Joyce, and J O'Hara</i>	548
Evaluation of Integrated Health Programs in School	<i>T M Franke</i>	555

Formative Assessment in Teacher Education and Teacher Professional Development	<i>H Tillema</i>	563
Informal Education and Evaluation	<i>T Jeffs</i>	572
Methods to Evaluate Technology	<i>W L Bewley</i>	578

Educational Evaluation – Contexts of Educational Evaluation

Cost Analysis in Evaluation Studies	<i>J D Fletcher</i>	585
Cultural Issues that Can Affect the Validity of Educational Evaluations	<i>R P Durán</i>	592
Evaluation and Accountability	<i>E Sato and S N Rabinowitz</i>	600
Evaluation Governance and Planning	<i>R Picciotto</i>	607
Evaluation Use	<i>V G Thomas</i>	614
Moral and Ethical Issues in Evaluation	<i>M Morris</i>	622
The Role of Stakeholders in Educational Evaluation	<i>S Taut and M C Alkin</i>	629

Educational Evaluation – Evaluation Domains

Curriculum Evaluation	<i>T Levin</i>	636
Evaluating Education in Three Policy Eras	<i>M A Peters, R Britez, and S M Weber</i>	645
Evaluation of Mathematics Education Programs	<i>J Cai</i>	653
Evaluation of Science Education Programs	<i>G Berberoğlu</i>	660
Evaluation of Social Studies and Civic Education Programs	<i>S G Grant</i>	667
Evaluation of Teacher Quality and Practice	<i>W Robinson and J Campbell</i>	674
International Evaluations	<i>A Schleicher</i>	681
National Assessment Programs: The Example of the U.S. National Assessment of Educational Progress	<i>M Schneider</i>	690
Principal Evaluation: Concepts, Needs and Realities	<i>N S Glasman and L D Glasman</i>	698
Program Evaluation	<i>C A Christie and L A Fierro</i>	706
School Based Evaluation: Purposes, Protocols and Processes	<i>J MacBeath</i>	713
School Inspection/External School Evaluation	<i>C Chapman and P Earley</i>	719
The International System for Teacher Observation and Feedback: A Theoretical Framework for Developing International Instruments	<i>L Kyriakides, B P M Creemers, C Teddlie, and D Muijs</i>	726

Educational Evaluation – Evaluation Methods

Evaluation Methodology	<i>M Chatterji</i>	735
Evaluation Reporting and Communicating	<i>R T Torres</i>	746
Longitudinal Evaluation Designs	<i>X Ma</i>	757
Meta-evaluation: Purpose, Prescription, and Practice	<i>L A Wingate</i>	765

Educational Evaluation – Purposes and Models of Educational Evaluation

Defining Quality in Evaluation	<i>P Dahler-Larsen</i>	775
Internal and External Evaluation	<i>D Nevo</i>	781
Needs Assessment in Education	<i>J W Altschuld and T L Lepicki</i>	786
The Purpose of Educational Evaluation	<i>S Mathison</i>	792
Understanding Approaches to Evaluation	<i>A J Love</i>	798

VOLUME 4**Educational Measurement**

Ability Testing	<i>P B Tymms</i>	1
Admissions Testing	<i>R Zwick</i>	7
Appropriateness Measurement: Person Fit	<i>R R Meijer and A M Weekers</i>	15
Automated Essay Scoring: Writing Assessment and Instruction	<i>M D Shermis, J Burstein, D Higgins, and K Zechner</i>	20
Classical Test Theory Reliability	<i>M D Miller</i>	27
Criterion-Referenced Measurement	<i>S A Shrock and W C Coscarelli</i>	31
Differential Item Functioning	<i>M D Hidalgo and J Gómez-Benito</i>	36
Educational Measurement: Overview	<i>R L Linn</i>	45
Equating and Scaling	<i>A A von Davier</i>	50
Fairness	<i>C Gipps and G Stobart</i>	56
Generalizability Theory	<i>R L Brennan</i>	61
Impact of Assessment on Classroom Practice	<i>J L Herman</i>	69
Intelligent Systems	<i>V J Shute and D Zapata-Rivera</i>	75
Item Response Theory	<i>W J van der Linden</i>	81
National Assessments	<i>T J Crooks</i>	89
Norm-Referenced Measurement	<i>H D Hoover</i>	95
Progress Monitoring	<i>L S Fuchs and D Fuchs</i>	102
Rasch Models	<i>D Andrich</i>	111
Scoring Rubrics	<i>J A Arter</i>	123
Second Language Assessment	<i>L F Bachman</i>	140
Self Adaptive Testing	<i>J Revuelta</i>	148
Student Test Results in School Accountability	<i>L S Hamilton</i>	153
Test Development	<i>S M Downing</i>	159
Test Translation and Adaptation	<i>A Allalouf and P Hanani</i>	166
Testing Creativity	<i>M A Runco</i>	170
Testing in History	<i>S Wineburg and B Fogo</i>	175
Validity	<i>R L Linn</i>	181
Validity of Achievement Gains on High-Stakes Tests	<i>D Koretz</i>	186
Validity of Educational Indicators	<i>B P M Creemers and L Kyriakides</i>	193

G**Globalization of Universities and Institutions**

Globalization, Universities and Medium of Instruction	<i>J Lo Bianco</i>	201
The Washington Accord on Engineering Curriculum	<i>E Ko</i>	209

H

Higher Education

Higher Education: An Overview	<i>P Scott</i>	217
-------------------------------	----------------	-----

Higher Education – Management, Leadership and Governance

Academic Managers – Heads of Departments/Schools	<i>S Marshall</i>	229
Leading and Managing the University – Presidents and their Senior Management Team	<i>R Middlehurst, T Kennie, and S Woodfield</i>	238
Management of and in Higher Education Institutions	<i>B Sporn</i>	245
Managerialism and Collegialism in Higher Education Institutions	<i>M Shattock</i>	251
Strategic Planning in Higher Education	<i>L Zechlin</i>	256
The Changing Role of University Presidents, Vice-Chancellors and Rectors	<i>M F Green and P D Eckel</i>	264
The Changing Roles of University Governing Boards and Councils	<i>J Fielden</i>	273

Higher Education – Research and Scholarship

Funding of University Research	<i>G Harman</i>	279
Mobility of PhD Students and Scientists	<i>S Avveduto</i>	286
National Science Policy and Universities	<i>M Henkel and U Kogan</i>	294
Research Commercialization	<i>S Slaughter</i>	300
Research Quality Assessment	<i>I McNay</i>	307
The Management of University Research	<i>J Kirkland</i>	316
The Training of New Scientists	<i>R Neumann</i>	322

Higher Education – Society

Academic Freedom	<i>S F Macintyre</i>	328
Access and Equity in Higher Education	<i>G Baldwin and R James</i>	334
Boundaries of Institutional Autonomy and their Impact on Higher Education	<i>V L Meek</i>	341
Diversity of Higher Education	<i>U Teichler</i>	347
Employability of University Graduates and Graduate Outcomes	<i>J Brennan</i>	354
Higher Education and the Knowledge Society	<i>J Välimaa</i>	360
Higher Education and the Labor Market	<i>J Enders</i>	365
Higher Education and the Transformation of Society	<i>P Scott</i>	370
Quality Assurance in Higher Education – Practices and Issues	<i>D Dill</i>	377
The Role of Accreditation of Higher Education Institutions	<i>J S Eaton</i>	384
Universities and Regional Development	<i>J Goddard and J Puukka</i>	390
Universities' Engagement with Society	<i>D Watson</i>	398

Higher Education – Teaching and Learning in Higher Education

Distance Education and Open Universities	<i>A Kanwar and J Daniel</i>	404
Education Development in Higher Education	<i>Y Ryan and K Fraser</i>	411
Effective Use of Technology in Teaching and Learning in HE	<i>D Laurillard</i>	419

Facilitating Students' Success – Student Support and Learning Resources in Universities	<i>E Woodberry and C Barnes</i>	427
Faculty Values and Expectations and Their Implications for Teaching, Learning and the Organization of Higher Education Institutions	<i>K Harman</i>	433
Flexible Learning in Higher Education	<i>S Alexander</i>	441
Innovation in Teaching and Curriculum Design	<i>J M Dearn</i>	448
Productivity of University Faculty Staff	<i>S Kyvik</i>	455
Promoting Effective Student Learning in Higher Education	<i>K Trigwell</i>	461
Student Attrition from Higher Education Institutions	<i>M Hayden</i>	467
University Students' Experiences of Higher Education	<i>C McInnis</i>	473
What Do University Students Learn?	<i>G D Kuh and B K Bridges</i>	479

Higher Education – Tertiary Education in a Globalized World

Associations of Universities	<i>G F Moodie</i>	485
Collective Bargaining in Higher Education	<i>C Allport, K McAlpine, and S Roberts</i>	491
Financing of Higher Education	<i>B Chapman and P Tulip</i>	499
Higher Education Crossing Borders	<i>J Knight</i>	507
Higher Education in Post-Conflict Societies	<i>M Daxner</i>	514
Higher Education in the New Economies	<i>J Fielden</i>	520
Human Resource Issues in Higher Education	<i>E El-Khawas</i>	527
International Accreditation in Higher Education	<i>A Stella and D Woodhouse</i>	533
Internationalization of Higher Education	<i>M van der Wende</i>	540
National and International Rankings of Higher Education	<i>S Marginson</i>	546
Steering of Higher Education Systems – The Role of the State	<i>K Yokoyama and V L Meek</i>	554
Student and Faculty Transnational Mobility in Higher Education	<i>U Lanzendorf and B M Kehm</i>	559
The Bologna Process in European Higher Education	<i>D Rich</i>	566
The Economics and Finance of Higher Education	<i>A Wolf</i>	573
The Role of the OECD in the Development of Higher Education in a Globalized World	<i>R Yelland</i>	584
Transforming Higher Education in Developing Countries: The Role of the World Bank	<i>J Salmi and R M Bassett</i>	590
UNESCO's Role in the Development of Higher Education in a Globalized World	<i>S Uvalić-Trumbić</i>	597

Higher Education – The Organization and Financing of Higher Education

Higher Education in Federal Systems	<i>M A M Cortés and U Teichler</i>	603
Pathways and Articulation into Higher Education	<i>G F Moodie</i>	609
Unitary, Binary and other National Systems of Higher Education	<i>L Goedegebuure</i>	616
Viewing Private Higher Education: How Much, Where, Why, and What?	<i>D C Levy</i>	622

I

Immigration

Children of Migrant Populations	<i>M M Suárez-Orozco and C Suárez-Orozco</i>	629
Third Culture Kids	<i>R E Van Reken</i>	636

International Comparisons

IEA Studies in ICT	<i>N Law</i>	644
IEA Studies in Mathematics and Science	<i>F K S Leung</i>	650
IEA Study in Civic Education	<i>J Torney-Purta, J-A Amadeo, and J Schwillie</i>	656
IEA: Globalization and Assessment	<i>H Wagemaker</i>	663

International Organizations in Education

Education and the European Union	<i>S Hegarty</i>	669
National University on Educational Planning and Administration, New Delhi	<i>J B G Tilak</i>	676
On School Management – International Council on School Effectiveness and Improvement, Commonwealth Council on Educational Administration and Management	<i>A Walker and T Townsend</i>	681
UNESCO's International Institute for Educational Planning	<i>S J Hite and M Bray</i>	688

L**Leadership and Management**

Leadership and Management Overview	<i>B Mulford</i>	695
------------------------------------	------------------	-----

Leadership and Management – as a Field of Study

Contexts for Leadership at the Beginning of the 21st Century	<i>M Bottery</i>	704
Educational Leadership: Philosophical Issues	<i>C W Evers</i>	710
Ethical/Moral Issues in Educational Leadership	<i>R J Starratt, L Langlois, and P Duignan</i>	717
History of Educational Leadership/Management	<i>R Bates</i>	724
Educational Leadership as a Field of Study: Professional Journals 1994–2006	<i>A R Thomas</i>	731
Research on Educational Leadership – Approaches/Promising Directions	<i>W A Firestone and V M J Robinson</i>	740

Leadership and Management – Leader Preparation

Leaders for Productive Schools	<i>J Murphy, S N Elliott, E B Goldring, and A C Porter</i>	746
New Approaches in Preparing School Leaders	<i>S G Huber</i>	752
Research for Leaders: The National College for School Leadership	<i>G Southworth</i>	762
School Leadership Standards	<i>L Ingvarson</i>	768

Leadership and Management – Leadership Issues

Effective Leadership in Challenging Schools	<i>A Harris</i>	777
Esthetics, Educational Leadership and Management	<i>R Bates</i>	783
Leader Mistreatment of Teachers	<i>J Blase and J Blase</i>	790
Leadership and Gender	<i>J Blackmore</i>	797
Leadership and Student Voice	<i>C Lanskey and J Rudduck</i>	803
Leadership and Urban Education	<i>K A Riley</i>	809
Leadership for Learning	<i>J MacBeath</i>	817
Leadership in Diverse Cultures	<i>A Walker and P K Yu-kwong</i>	824

Leading Diverse Schools	<i>J Ryan</i>	832
Supervision and Personnel Administration	<i>P D Tucker and D G Pounder</i>	839
The Big Challenge: Leadership for Inclusion	<i>M Ainscow and A Sandill</i>	846
The Educational Leader's New Work	<i>P Gronn</i>	852

VOLUME 5

Leadership and Management – Leadership Types

A Distributed Perspective on School Leadership and Management	<i>J P Spillane</i>	1
Leadership: Authentic	<i>P T Begley</i>	7
Leadership: Democratic	<i>J Møller</i>	12
Leadership: Instructional	<i>P Hallinger, K Leithwood, and R H Heck</i>	18
Leadership: Transformative	<i>C M Shields</i>	26
Strategic Leadership	<i>B Davies and B J Davies</i>	34
Teacher Leadership and Organizational Development	<i>A Harris</i>	40
Understanding How Leadership Influences Student Learning	<i>K Leithwood and B Levin</i>	45

Leadership and Management – Politics and Governance

Equity and Educational Effectiveness	<i>P Sammons</i>	51
Leading/Managing Schools in Communities Made Poor	<i>P Thomson</i>	58
Principals Role in Restructuring Schools	<i>G M Crow and K D Peterson</i>	66
School Reform and Restructuring: Self Managing School	<i>B J Caldwell</i>	72
System Leadership	<i>D Hopkins and R Higham</i>	78

Leadership and Management – School and Community

Educational Leadership and Social Capital	<i>J West-Burnham</i>	85
Networked Learning Communities School-to-School Collaboration as an Essential Component of a System Reform Strategy	<i>D Jackson</i>	92
Networks and Communities of Knowledge	<i>T Schuller and H Theisens</i>	100
School Community Relations	<i>R L Crowson and E B Goldring</i>	106
Social Capital, Educational Institutions and Leadership	<i>S Kilpatrick, S Johns, and B Mulford</i>	113

Leadership and Management – School Effectiveness and Improvement

Leadership and Technology	<i>C F Webber</i>	120
Leadership in the Implementation of Innovations	<i>I E Bogotch, T Townsend, and M Acker-Hocevar</i>	128
Leadership: School Improvement	<i>R H Heck and P Hallinger</i>	135
Organizational Learning in Schools	<i>B Mulford and H Silins</i>	143
Professional Learning Community	<i>L Stoll</i>	151
Transformational School Leadership	<i>K Leithwood</i>	158

Learning and Cognition – Issues and Concepts

Attention in Cognition and Early Learning	<i>M Heimann, T Tjus, and K Strid</i>	165
Concept Learning	<i>S J Unsworth and D L Medin</i>	172
Cooperative Learning	<i>R E Slavin</i>	177
Intelligence	<i>R J Sternberg</i>	184
Learning from Multiple Information Sources	<i>H I Strømsø and I Bråten</i>	191
Memory	<i>S Magnussen and T Brennen</i>	197
Metacognition	<i>L Baker</i>	204
Personal Epistemology in Education: Concepts, Issues, and Implications	<i>I Bråten</i>	211
Piaget: Recent Work	<i>L Smith</i>	218
The Adult Development of Cognition and Learning	<i>E A L Stine-Morrow and J M Parisi</i>	225
Vygotsky and Recent Developments	<i>J V Wertsch</i>	231

Learning and Cognition – Issues and Concepts – Focus on Cognition

Cognition and Emotion	<i>F Pons, M de Rosnay, and F Cuisinier</i>	237
Cognition: Overview and Recent Trends	<i>J Wiley and B D Jee</i>	245
Development of Creativity	<i>M A Runco</i>	251
Knowledge Domains and Domain Learning	<i>L Maggioni and P A Alexander</i>	255
Problem Solving and Human Expertise	<i>T J Nokes, C D Schunn, and M T H Chi</i>	265
Problem Solving and Reasoning	<i>R E Mayer</i>	273
Reasoning, Moral and Social	<i>N S Jampol, C B Richardson, and M Killen</i>	279
Remembering as Social Activity	<i>D Middleton</i>	285

Learning and Cognition – Issues, Concepts, Types

Computer-Supported Collaborative Learning: Basic Concepts, Multiple Perspectives, and Emerging Trends	<i>S R Ludvigsen and A I Mørch</i>	290
Learning as Inquiry	<i>S R Goldman, J Radinsky, S Tozer, and D Wink</i>	297
Learning Through Narrative Socialization	<i>A Fasulo</i>	303
Neuroscience Bases of Learning	<i>M H Immordino-Yang and K W Fischer</i>	310
The Neuroscience of Reading	<i>U Goswami</i>	317

Learning and Cognition – Issues, Concepts, Types – Focus on Learning

Learning Strategies	<i>C E Weinstein, J Jung, and T W Acee</i>	323
Learning Through Play	<i>K Aronsson</i>	330
Situated View of Learning	<i>A Collins and J G Greeno</i>	335
The Neuroscience of Aging and Cognition	<i>W J Hoyer and B Bessette-Symons</i>	340

Learning and Cognition – Language, Literacy and Subject-Related

An Overview of Language and Literacy in Educational Settings	<i>V G Aukrust</i>	345
Bilingualism and Learning	<i>L S Verplaetse and E Schmitt</i>	355
Classroom Discourse and Student Learning	<i>L Hemphill</i>	361
Early Writing	<i>B E Hagtvet</i>	367

First Language Acquisition	<i>N B Ratner</i>	375
Foreign Language Learning	<i>C Kinginger</i>	382
Learning Science	<i>J Leach and P Scott</i>	387
Learning to Read	<i>M S Burns and J K Kidd</i>	394
Mathematics Learning	<i>L Verschaffel, B Greer, and E De Corte</i>	401
Music Learning	<i>S Hallam</i>	407
Reading Comprehension: Reading for Learning	<i>C E Snow</i>	413
Second Language Learning	<i>J S Arnfast, J N Jørgensen, and A Holmen</i>	419
Writing: Advanced	<i>R Andrews</i>	426

Learning and Cognition – Other Issues

Gender and Learning	<i>L Yates and J McLeod</i>	432
Parent–Child Relationships in Early Learning	<i>C P Edwards, S M Sheridan, and L L Knoche</i>	438
Peer Interaction and Learning	<i>S Blum-Kulka and S Dvir-Gvirsman</i>	444

Learning and Cognition – Places of Learning

Learning in a Cross-Cultural Perspective	<i>L-F Zhang and R J Sternberg</i>	450
Learning Outside of School	<i>T Nunes</i>	457
Reasoning and Explanatory Talk: Learning in Everyday Settings	<i>D E Beals</i>	464

Learning and Cognition – Theoretical Perspectives – Learning

Apprenticeship Approach to Learning	<i>K Nielsen</i>	469
Connectionism and Learning	<i>T R Shultz</i>	476
Constructivism and Learning	<i>S Sjøberg</i>	485
Cultural–Historical Activity Theory	<i>A Havnes</i>	491
Learning in a Sociocultural Perspective	<i>R Säljö</i>	498
Self-Regulated Learning and Socio-Cognitive Theory	<i>P H Winne and A F Hadwin</i>	503

N

National Systems of Education

Australia	<i>C N Power</i>	510
Austria	<i>M J Gaertner</i>	517
Bangladesh	<i>R Z Dewan</i>	524
Barbados	<i>R Rudder</i>	534
Bulgaria	<i>A Chavdarova and D Vesselinov</i>	540
England	<i>C Whatford</i>	550
Estonia	<i>M Keränen and J Kõrgesaar</i>	555
Fiji	<i>I Senibulu</i>	563
Finland	<i>J E Tuovinen</i>	569
France	<i>M Sellier</i>	576

Germany	<i>A Hüfner and K Hüfner</i>	582
Ghana	<i>R Palmer</i>	590
Greece	<i>K Toura and A Karagiannis</i>	599
Iceland	<i>M Hardardottir and M Gunnlaugsdóttir</i>	605
Indonesia	<i>A C Alwasilah and Furqon</i>	611
Israel	<i>A Volansky</i>	618
Jamaica	<i>E Hannam</i>	627
Latvia	<i>A Joma</i>	630
Lebanon	<i>N A Mneimneh</i>	636
Lithuania	<i>J Navickaitė, L Paurienė, R Zablackė, A Tijūnelytė, and V Kožemiakina</i>	656
Malawi	<i>F R Mkandawire and D Mulera</i>	666
Malta	<i>C Farrugia</i>	673
Mongolia	<i>B Yembuu</i>	681
Namibia	<i>J Ellis</i>	687
New Zealand	<i>R Baker and C Wylie</i>	692
Niger	<i>A A I Daouda</i>	699
Norway	<i>H Farstad</i>	709
Oman	<i>S A Issan</i>	718
Pakistan	<i>N Mahmood and R Malik</i>	727
Poland	<i>A Janowski</i>	737
Qatar	<i>S R Salman</i>	744
Russia	<i>V M Filippov and G F Tkach</i>	750
Rwanda	<i>E Bahizi</i>	758
Scotland	<i>D Raffe</i>	770
Singapore	<i>P T Ng</i>	776
Slovenia	<i>T Plevnik and A B Lakota</i>	782
Solomon Islands	<i>D P Malasa</i>	790
South Africa	<i>P Kgobe and J Pampallis</i>	799
Spain	<i>A Tiana</i>	806
Sri Lanka	<i>B M J Balasooriya, M M Wehella, and S U Wijeratne</i>	813
Suriname	<i>H R Ori</i>	826
Taiwan	<i>C-M Lee</i>	834
Tanzania	<i>R Wedgwood</i>	841
Thailand	<i>C Rukspollmuang</i>	847
The Gambia	<i>Y A-M Jobe</i>	858
The Philippines	<i>E A P Valenzuela</i>	865
Uruguay	<i>B Guinovart and C Troncoso</i>	873
USA	<i>B Pusser and M E R Lloyd</i>	888
Zanzibar	<i>I M Hija</i>	896

VOLUME 6**P****Philosophy of Education**

Philosophy of Education: Overview	<i>P Enslin</i>	1
-----------------------------------	-----------------	---

Philosophy of Education – Contemporary Issues

Neoliberalism, the Market and Performativity	<i>M A Peters</i>	11
Parents and Children's Rights	<i>M Clayton</i>	17
Peace Education	<i>I Gur-Ze'ev</i>	22
Philosophy and Educational Research	<i>D Bridges</i>	29
Race, Critical Race Theory and Whiteness	<i>B Applebaum</i>	36
School Choice and the Common School	<i>G Haydon</i>	44

Philosophy of Education – Philosophical Perspective

African Philosophy of Education	<i>K Horsthemke</i>	50
Feminism	<i>L Stone</i>	56
Hermeneutics	<i>C Bingham</i>	63
Islamic Education	<i>Y Waghid</i>	69
Liberalism and Education	<i>W Feinberg</i>	74
Phenomenology and Existentialism	<i>B Levering</i>	80
Pragmatism	<i>B Thayer-Bacon</i>	86
Psychoanalysis and Education	<i>P Smeyers</i>	92
Social and Cultural Capital in Education	<i>M Print</i>	97
The Aims of Education	<i>R Marples</i>	103
The Analytical Tradition	<i>D Carr</i>	108

Philosophy of Education – Philosophical Themes

Adult Learning: Philosophical Issues	<i>D Beckett</i>	114
Autonomy	<i>C A Wringe</i>	121
Bildung	<i>J Masschelein and N Ricken</i>	127
Critical Theory and Pedagogy	<i>M Mason</i>	133
Critical Thinking	<i>H Siegel</i>	141
Environmental Education	<i>M Bonnett</i>	146
Identity	<i>M S Merry</i>	152
Inclusion	<i>R Cigman</i>	158
Justice and Care	<i>M Tjiattas</i>	164
Lifelong Learning	<i>J Halliday</i>	170
Post-Structuralism and Education	<i>M Mason and M Clarke</i>	175
Reason and Rationality	<i>P Hogan</i>	183

Religious and Spiritual Education	<i>J C Conroy and S J McKinney</i>	189
The Capabilities Approach	<i>L Terzi</i>	194

Primary and Secondary Education – Learning and Teaching in School Age Education

Class Size – Arguments and Evidence	<i>P Blatchford and K C Lai</i>	200
Continuing Professional Development of Teachers	<i>P Earley</i>	207
Homeschooling	<i>P Farenga</i>	214
Innovations and Early Intervention in the Teaching of Literacy	<i>R E Slavin</i>	221
Personalized Learning in School Age Education	<i>D Hopkins</i>	227
Perspectives on Schooling in the Middle Years	<i>V J Russell</i>	233
Reform of the School Workforce	<i>H M Gunter and G Forrester</i>	242
Secondary Education in Developing Countries	<i>P Hughes and A Walker</i>	249

Primary and Secondary Education – Social Justice Issues in School Age Education

Community Focused Schools	<i>A Dyson</i>	255
---------------------------	----------------	-----

Primary and Secondary Education – The Effectiveness of School Systems

Reforming the American School System	<i>M Barber</i>	261
School Effectiveness in Developed Societies	<i>D Reynolds</i>	269
School to School Collaboration: Innovation and Improvement	<i>R Higham</i>	274
Studies of School Improvement in Developing Countries	<i>S E Anderson</i>	281
Sustainable Educational Reform	<i>A Hargreaves</i>	289
The Role of the District in Tri Level Reform	<i>M Fullan</i>	295
Whole School Designs for Enhancing Student Achievement	<i>S Stringfield and J Nunnery</i>	303

Q

Qualitative Research

Action Research in Education	<i>E Stringer</i>	311
Classroom Ethnography	<i>F Erickson</i>	320
Computer Assisted Qualitative Data Analysis	<i>C Silver and A Lewins</i>	326
Conversational Analysis	<i>H G Levine</i>	335
Critical Ethnography	<i>D Foley</i>	337
Critical Race Theory	<i>D Gillborn and G Ladson-Billings</i>	341
Critical Theory	<i>J J Ryoo and P McLaren</i>	348
Cultural Studies	<i>R Gaztambide-Fernández and K Cairns</i>	354
Cultural–Historical Activity Theory	<i>Laboratory for Comparative Human Cognition</i>	360
Design Experiments	<i>A Collins</i>	367
Discourse Analysis	<i>P Nikander and S Vehviläinen</i>	373
Document Analysis	<i>T Rapley and K N Jenkins</i>	380
Ethnography	<i>M Hammersley</i>	386

Ethnomethodology in Education Research	<i>D Macbeth</i>	392
Focus Groups	<i>D L Morgan and K Hoffman</i>	401
Grounded Theory	<i>K Charmaz and A Bryant</i>	406
Hermeneutics	<i>J Dyer</i>	413
Interpretive Research	<i>E Rauscher and M E Graue</i>	419
Interviews and Interviewing	<i>A Marvasti</i>	424
Life History	<i>J A Hatch and S K Newsom</i>	430
Narrative Inquiry	<i>D J Clandinin and J Huber</i>	436
Participant Observation	<i>J Schostak</i>	442
Phenomenology	<i>M van Manen and C A Adams</i>	449
Qualitative Case Studies	<i>S B Merriam</i>	456
Qualitative Data Management	<i>C Grbich</i>	463
Semiotics	<i>M Danesi</i>	475
Symbolic Interactionism, Naturalistic Inquiry, and Education	<i>J A Forte</i>	481
The Quality of Evidence in Qualitative Research	<i>S K Biklen</i>	488
Visual Data in Education Research	<i>J Wagner</i>	498

Qualitative Research – Standards of Proof

Validity: Mapping Diverse Perspectives	<i>G B Rossman, S F Rallis, and A M Kuntz</i>	505
--	---	-----

Quantitative and Qualitative Approaches to Research – Integration

Mixed Methods	<i>A Tashakkori and I Newman</i>	514
---------------	----------------------------------	-----

Quantitative Methods and Research Design

External Validity	<i>G E Matt, A Brewer, and M Sklar</i>	521
-------------------	--	-----

R

Role of NGOs in Globalization

NGOs and Globalization of Education	<i>M Ahmed</i>	529
-------------------------------------	----------------	-----

S

Social and Emotional Aspects of Learning

Achievement Goal Theory: Definitions, Correlates, and Unresolved Questions	<i>T Urdan</i>	537
Affect, Mood and Emotions	<i>S Schnall</i>	544
Anxiety	<i>M Zeidner</i>	549
Attribution Theory	<i>B Weiner</i>	558
Children's Friendship	<i>S R Asher and K L McDonald</i>	564
Coping with Stressful Situations: An Important Aspect of Self-Regulation	<i>M Boekaerts</i>	570
Culture in Motivation Research: A Challenging and Enriching Contribution	<i>K Kimmel and S Volet</i>	576
Early Social Development and Schooling	<i>G M van der Aalsvoort</i>	585

Emotion in Educational Contexts	<i>P A Schutz, P D Quijada, S de Vries, and M Lynde</i>	591
Empathy	<i>R Arnold</i>	597
Flow in Education	<i>J A Schmidt</i>	605
Interest	<i>M Ainley</i>	612
Intrinsic and Extrinsic Motivation	<i>C Levesque, K J Copeland, M D Pattie, and E L Deci</i>	618
Motivating Students in Classrooms	<i>J Brophy</i>	624
Motivation Regulation	<i>C A Wolters and S A Mueller</i>	631
Peer Learning in the Classroom	<i>N M Webb</i>	636
Poverty, Effects of on Social and Emotional Development	<i>N Philipsen Hetzner, A D Johnson, and J Brooks-Gunn</i>	643
Seeking Help as an Adaptive Response to Learning Difficulties: Person, Situation, and Developmental Influences	<i>S A Karabenick and R S Newman</i>	653
Self-Concept in Learning: Reciprocal effects model between academic self-concept and academic achievement	<i>H W Marsh and L F Scalas</i>	660
Self-Efficacy Beliefs	<i>D H Schunk and F Pajares</i>	668
Social and Emotional Outcomes of Learning	<i>J Vorhaus</i>	673
Social Aspects of Collaborative Learning	<i>M Arvaja and P Häkkinen</i>	685
Social Development and Schooling	<i>G W Ladd</i>	691
Social Interaction and Learning	<i>K Littleton</i>	698
Social Networks and the Education of Children and Youth	<i>B Schneider, T G Ford, and L Perez-Felkner</i>	705
Sociocultural Issues in Motivation	<i>R A Walker</i>	712
Socioemotional Aspects of Technology-Supported Learning	<i>S Järvelä</i>	718
Volitional Control of Learning	<i>J Husman and L Corno</i>	724
Wellbeing	<i>T Hascher</i>	732

VOLUME 7

Statistics

An Overview of Statistics in Education	<i>S Sinharay</i>	2
Analysis and Interpretation of Multivariate Data	<i>D J Bartholomew</i>	12
Analysis of Covariance	<i>N T Longford</i>	18
Analysis of Extreme Values in Education	<i>J Beirlant, G Dierckx, K Luwel, and P Onghena</i>	25
Analysis of Variance	<i>B M King</i>	32
Bayesian Statistical Analysis	<i>G Kokolakis</i>	37
Bootstrap Method	<i>K Singh and M Xie</i>	46
Canonical Correlation	<i>A Basu and A Mandal</i>	52
Categorical Data Analysis	<i>S E Fienberg and B W Junker</i>	58
Causal Inference	<i>D B Rubin</i>	66
Cluster Analysis: Overview	<i>S Landau and I Chis Ster</i>	72

Cognitive Psychology and Educational Statistics	<i>H S Kim and S Embretson</i>	84
Computational Statistics	<i>J E Gentle</i>	93
Continuous Probability Distributions	<i>S Sinharay</i>	98
Correspondence Analysis	<i>M Greenacre</i>	103
Data Mining	<i>R S J d Baker</i>	112
Decision Theory	<i>M Mendoza and E Gutiérrez-Peña</i>	119
Design of Experiments	<i>C A Albers and T R Kratochwill</i>	125
Discrete Probability Distributions	<i>S Sinharay</i>	132
Discrimination and Classification	<i>R C Neath and M S Johnson</i>	135
Empirical Bayes Methods	<i>E Ip and G Molenberghs</i>	142
Evaluation Research	<i>T Kellaghan</i>	150
Exploratory Data Analysis	<i>A Unwin</i>	156
Factor Analysis: An Overview and Some Contemporary Advances	<i>F A N Alhija</i>	162
Generalized Linear Mixed Models	<i>S Rabe-Hesketh and A Skrondal</i>	171
Generalized Linear Models	<i>J J Faraway</i>	178
Generating Random Numbers	<i>P J Pashley and A Amodeo</i>	184
Goodness-of-Fit Testing	<i>A Maydeu-Olivares and C García-Forero</i>	190
Graphical Models	<i>R G Almond</i>	197
Growth Modeling	<i>I Plewis</i>	203
Hierarchical Linear Models	<i>D Rindskopf</i>	210
Hypothesis Testing and Confidence Intervals	<i>T R Konold and X Fan</i>	216
Instrumental Variables	<i>E P Bettinger</i>	223
Jackknife Methods	<i>S Sinharay</i>	229
Large-sample Statistical Methods	<i>J M Singer and D F Andrade</i>	232
Latent Class Models	<i>J K Vermunt</i>	238
Markov Chain Monte Carlo	<i>A M Johansen</i>	245
Matrix Algebra	<i>H Yanai and H Ishii</i>	253
Measure of Association	<i>G-H Huang</i>	260
Measures of Central Tendency	<i>D Clark-Carter</i>	264
Measures of Dispersion, Skewness and Kurtosis	<i>D Rindskopf and M Shiyko</i>	267
Meta Analysis	<i>J Sánchez-Meca and F Marín-Martínez</i>	274
Missing Data	<i>C A W Glas</i>	283
Model Selection	<i>S J Haberman</i>	289
Monte Carlo Methods	<i>A M Johansen</i>	296
Multidimensional Scaling	<i>Z Zhang and Y Takane</i>	304
Multiple Comparisons	<i>C Lewis</i>	312
Multivariate Analysis of Variance	<i>N T Longford</i>	319
Multivariate Linear Regression	<i>S Ganesh</i>	324
Multivariate Normal Distribution	<i>J Tacq</i>	332

Nonlinear Regression Analysis	<i>H-H Huang, C K Hsiao, and S-Y Huang</i>	339
Nonparametric Statistical Methods	<i>K Sijtsma and W H M Emons</i>	347
Observational Studies	<i>P R Rosenbaum</i>	354
Order Statistics	<i>I Bairamov</i>	360
Point Estimation Methods with Applications to Item Response Theory Models	<i>F Bartolucci and L Scrucca</i>	366
Principal Components Analysis	<i>D J Bartholomew</i>	374
Probability Theory	<i>T Rudas</i>	378
Recursive Partitioning	<i>S Landau and S Barthel</i>	383
Sampling	<i>J Qian</i>	390
Sequential Probability Ratio Test	<i>T J H M Eggen</i>	396
Sequential Testing	<i>H J Vos</i>	401
Signal Detection Theory	<i>H Abdi</i>	407
Small Area Estimation	<i>A Noble</i>	411
Statistical Analysis of Functional Data	<i>F Ferraty, P Sarda, and P Vieu</i>	417
Statistical Inequalities	<i>H M Mahmoud and J L Gastwirth</i>	426
Statistical Paradoxes	<i>Y-K Tu</i>	429
Statistical Power Analysis	<i>L V Hedges and C Rhoads</i>	436
Statistical Significance Versus Effect Size	<i>X Fan and T R Konold</i>	444
Stochastic Processes	<i>F Baudoin</i>	451
Structural Equation Models	<i>S-Y Lee and X-Y Song</i>	453
Survival Data Analysis	<i>A F De Champlain</i>	459
The Normal Distribution and its Applications	<i>J Tacq</i>	467
Time Series Analysis	<i>P J Brockwell</i>	474
Univariate Linear Regression	<i>M L Hazelton</i>	482
Value-Added Models	<i>D F McCaffrey</i>	489

T

Teacher Education – Evaluation in the Teacher Education Program

Accreditation and Standards in Teacher Education	<i>R J Yinger and K L Daniel</i>	495
Determining Long Term Effects of Teacher Education	<i>C N Brouwer</i>	503

Teacher Education – Evaluation of the Prospective Teacher

The Evaluation of Prospective Teachers in Teacher Education	<i>D E H Tigelaar and J van Tartwijk</i>	511
---	--	-----

Teacher Education – Macro Factors

Diversity in Teacher Education	<i>T Wubbels</i>	518
Social and Cultural Influences on Teacher Education	<i>M Ben-Peretz and R Lotan</i>	525
The Role of Governance in Teacher Education	<i>G E Karlsen</i>	532

Teacher Education – Post-Initial Professional Development

Action Research as a Tool for Teachers' Professional Development	<i>P Ponte</i>	540
Contemporary Approaches to Teacher Professional Development	<i>H Borko, J Jacobs, and K Koellner</i>	548
School Development for Teacher Learning and Change	<i>P Sleegers and K Leithwood</i>	557
Teacher Induction	<i>D Beijaard, J Buitink, and C Kessels</i>	563
Teacher Learning as Workplace Learning	<i>J Imants and K van Veen</i>	569
Teacher Learning with Lesson Study	<i>A Murata</i>	575
Teachers Career Stages and Professional Development	<i>R Fessler and E Rice</i>	582

Teacher Education – Preservice Teacher Education

A Pedagogy of Teacher Education	<i>J Loughran</i>	587
Case Methods in Teacher Education	<i>M Sato and C Rogers</i>	592
Conceptions of Teacher Education	<i>J A Whitcomb</i>	598
Inquiry-Oriented Teacher Education	<i>R Schulz</i>	604
Narratives and Biography in Teacher Education	<i>G Kelchtermans</i>	610
Supervision in Teacher Education	<i>J McCarthy and L F Quinn</i>	615
Taking Prospective Teachers' Beliefs into Account in Teacher Education	<i>M Valcke, G Sang, I Rots, and R Hermans</i>	622
Teacher Education and Models of Teacher Reflection	<i>V K LaBoskey</i>	629
Teacher Education as Teaching for Understanding with New Technologies	<i>M S Wiske and D E Spicer</i>	635

Teacher Education – Preservice: The Knowledge Base

Experienced Teachers' Craft Knowledge	<i>P C Meijer</i>	642
Moral Values in Teacher Education	<i>W Veugelers</i>	650
Pedagogical Content Knowledge	<i>J H Van Driel and A Berry</i>	656
Teacher Education and the Educational Foundations Knowledge Base	<i>F B Murray</i>	662
The Relationship Between Theory and Practice in Teacher Education	<i>F A J Korthagen</i>	669

Teacher Education – Teacher Educators

Characteristics, Scholarship and Research of Teacher Educators	<i>M Lunenberg</i>	676
Professional Development of Teacher Educators	<i>P K Smith</i>	681
Self-Study by Teacher Educators	<i>T Russell</i>	689

Teacher Education – Teaching Specific Domains

Arts Teacher Education	<i>L Beudert</i>	695
Mathematics Teacher Education	<i>K Krainer and S Llinares</i>	702
Music Teacher Education	<i>M W-y Wong</i>	706
Science Teacher Education	<i>J H Van Driel and S K Abell</i>	712
Social Studies Teacher Education	<i>A-L Halvorsen and S M Wilson</i>	719
Teacher Education for Elementary Education	<i>K D Brown and G N Carnes</i>	725

Teacher Education for Teaching the Gifted	<i>A Y Baldwin</i>	732
Teacher Preparation in Mother Tongue Education	<i>M Kooy and H Hulshof</i>	738
Vocational Teacher Education	<i>S Johansson</i>	744

Teacher Education – The Cooperating School

Mentoring in Teacher Education	<i>P D Tomlinson, A J Hobson, and A Malderez</i>	749
Partnerships Between Schools and Higher Education	<i>C Day and L Smethem</i>	757

VOLUME 8

Technology and Learning

An Overview of Technology and Learning	<i>B Means and J Roschelle</i>	1
Educational Data Modeling	<i>S Iksal</i>	11
Student Data Systems and Their Use for Educational Improvement	<i>J C Wayman, V Cho, and M P Richards</i>	14
Technology and Learning: Access in Schools Around the World	<i>R Anderson and J Ainley</i>	21

Technology and Learning – as a Subject of Instruction

Conceptions of Technology Literacy and Fluency	<i>O Erstad</i>	34
--	-----------------	----

Technology and Learning – Assessment

Technology and Formative Assessment	<i>J W Pellegrino</i>	42
Technology for Large-Scale Assessment	<i>R E Bennett</i>	48
Technology Supports for Assessment Design	<i>R J Mislevy, F I Winters, I I Bejar, R E Bennett, and G D Haertel</i>	56

Technology and Learning – Education Reform and Economic Development

National Strategies to Build a Technology Workforce	<i>H-M Cheah</i>	66
Public–Private Partnerships for Educational Reform	<i>V Quah, C P Lim, and C Brook</i>	73
Relating Technology, Education Reform and Economic Development	<i>R B Kozma</i>	81

Technology and Learning – Evaluating Technology

Examining the Effects of Educational Technology Programs: Challenges and Strategies	<i>M Russell</i>	88
---	------------------	----

Technology and Learning – Global Trends

Equity in Technology Access and Opportunities	<i>T Matuchniak and M Warschauer</i>	95
Internet-Based Education	<i>M Bakia</i>	102

Technology and Learning – Issues

Media Use and School Achievement	<i>K Roe</i>	109
Technology and Physical and Social Health	<i>K Subrahmanyam</i>	112

Technology and Learning – Meeting Special Students’ Needs

Universal Design for Learning *D H Rose and J W Gravel* 119

Technology and Learning – School Leadership for Technology Integration

Technology as a Support for School–Community Connections *K M Culp* 125

Technology and Learning – Supports for Managing Instruction

Classroom Uses of Technology to Manage Instruction *W R Penuel* 133

Technology and Learning – Supports for Skill Learning

New Media, Learning from *W Schnotz and H Horz* 140

Reading and Technology *N DeFrance, D Khasnabis, and A S Palincsar* 150

Technological Supports for Acquiring Twenty-First-Century Skills *C Dede* 158

Technology Supports for Acquiring Inquiry Skills *T de Jong* 167

Technology Supports for Acquiring Mathematics *M J Nathan* 172

Technology Supports for Lifelong Learning *A Collins and R Halverson* 184

Technology and Learning – Supports for Student Collaboration

Knowledge Communities in the Classroom *J D Slotta and H Najafi* 189

Technology and Learning – Supports for Subject Matter Learning

Technology Supports for Science Learning *C Delgado and J Krajcik* 197

Technology Supports for Second Language Learning *D Nunan* 204

Technology and Learning – Teacher Learning

Teacher Skills and Knowledge for Technology Integration *N Law* 211

Technology in Preservice Teacher Education *N Davis* 217

Technology Resources for Teacher Learning *J E Hinojosa, C Labbé, and L López* 222

V

Vocational Education and Training

An Overview of Vocational Education and Training *T Karmel* 229

Informal and Non-Formal Learning in Vocational Education and Training *F Beddie and S Halliday-Wynes* 240

School-Based Vocational Education and Training *B Knight and R Sweet* 247

The Economics of Vocational Education and Training *P Carneiro, L Dearden, and A Vignoles* 255

Vocational Education and Training – Individuals and Students

Age Equality in Education and Training *P Taylor* 262

Career Guidance in Vocational Education and Training and in the Workplace *P Plant* 269

Displaced Workers, Unemployed and Vocational Education and Training *B Kriechel* 275

Immigrant Investment in Host Country Vocational Education and Training *I Kogan* 280

Indigenous People and Vocational Education and Training	<i>R G Schwab</i>	289
Persons with a Disability and Vocational Education and Training	<i>F Ferrier and C S Smith</i>	294
Prisoner Education and Training	<i>S Dawe, S H Wynes, and J Misko</i>	300
Vocational Education and Training and the School-to-Work Transition	<i>G Hayward</i>	306

Vocational Education and Training – Industry and Employers

Apprenticeships	<i>E Smith</i>	312
Skills Shortages: Concepts, Measurement and Policy Responses	<i>C Shah and G Burke</i>	320
Training and Learning in the Workplace	<i>R L Jacobs</i>	328
Training in Small Businesses	<i>H Matlay</i>	337

Vocational Education and Training – Research on VET

Language of Vocational Education and Training	<i>J Misko, J MacKenzie, and R-A Polvere</i>	343
Research on Vocational Education and Training	<i>J Robson and B Bailey</i>	352
Vocational Education and Training Information Resources	<i>L-A Harris, R Naidu, and M Saunders</i>	358

Vocational Education and Training – Teaching and Learning

Employability Skills	<i>E Smith</i>	368
Learning Styles in Vocational Education and Training	<i>J Vorhaus</i>	376
Modularization in Vocational Education and Training	<i>H Ertl and G Hayward</i>	383
Personalized Learning and Vocational Education and Training	<i>I Duckett</i>	391
Recognition of Prior Learning and Vocational Education and Training	<i>J Keating</i>	397
The Competency Model	<i>A Gonczi and P Hager</i>	403

Vocational Education and Training – VET in Context

Globalization and Vocational Education and Training	<i>P Loveder</i>	411
Learning to Work for the Future	<i>J Fien, R Maclean, and D N Wilson</i>	416
The Relationship between Technical and Vocational Education and Training and Innovation	<i>K Ruth and L Deitmer</i>	423
The Status of Vocational Education and Training	<i>J Lasonen</i>	429
Vocational Education and Training in Post-Conflict Countries	<i>E Lyby</i>	441

Vocational Education and Training – VET System

Dual System	<i>T Deissinger</i>	448
Industry Involvement in the Vocational Education and Training System	<i>M Hughes and C Hughes</i>	455
Non-Completion in Vocational Education and Training: Issues, Patterns and Possibilities	<i>S Lamb and A Walstab</i>	463
Planning and Policy Development for Technical Vocational Education and Training Systems	<i>R Maclean and M Pavlova</i>	469
Qualifications Frameworks and their Role in the Reform of Education and Training	<i>M Young</i>	476
Quality Assurance in Vocational Education and Training	<i>D Gunning</i>	482

Training Finance	<i>A Ziderman</i>	489
Training Markets	<i>P Ryan</i>	496
Vocational Education and Training Teacher Training	<i>S Nielsen</i>	503
Vocational Education and Training Workforce	<i>D L Wallin</i>	513
Subject Index		519
Contributors		747

EARLY CHILDHOOD CARE AND EDUCATION THEORY

Contents

Early Childhood Care and Education: The Family, The Market, and The State

Gender Issues in Early Childhood Education and Care

Early Childhood Care and Education: The Family, The Market, and The State

S B Kamerman, Columbia University, New York, NY, USA

M J Neuman

© 2010 Elsevier Ltd. All rights reserved.

Introduction and Overview

Historically, the development of education systems has spurred ongoing tensions with regard to the appropriate roles of the family, the market, and the state in the provision, funding, or regulation of services. These debates have been even more heated when it comes to the early years of children's lives. Unlike schooling – which has been recognized as a public good for children from about 6 years of age – most countries have a long tradition of informal and private services for children of preschool age. The role of the family continues to be strong, and government is, often, ambivalent about its responsibility. In recent years, many countries have developed strong publicly funded early-childhood systems as part of their welfare states or social protection systems, whereas others have left this responsibility to family and to the market.

Early-childhood education and care (ECEC) straddles many policy domains. As revealed in the Organization for Economic Cooperation and Development (OECD) thematic review of 20 advanced industrialized countries carried out between 1998 and 2006 (*Starting Strong I & II*, 2001 & 2006) and the United Nations Educational, Scientific and Cultural Organization (UNESCO) reviews of developing countries, the funding and delivery of ECEC involves parents and families, civil society and the market, and government (employment policies, income transfers and the prevention of social exclusion, health, education, care and leave policies, etc.). In some developing

countries, conditional cash transfers (CCTs) can also be viewed as a component of education policy, as income supplements for poor families are provided as long as they follow certain behaviors such as enrolling their children in school (or bringing them to regular healthcare check-ups).

This article focuses on the roles of the family, the market, and the state in the provision of ECEC for children younger than the compulsory school age – involving elements both of physical care and education (socialization as well as cognitive stimulation). Throughout the world, ECEC services may be publicly funded and delivered, publicly funded and privately delivered (by for profit or nonprofit organizations), or privately funded and delivered; and they tend to be heavily subsidized by government in most OECD countries. ECEC programs include a wide range of part-day and full-day programs under education, health, and social welfare auspices. The major cross-national differences have to do with such variables as the locus of policymaking authority (national or local), administrative auspices (education, health, social welfare, or a combination), scale of coverage, age-group served (infants and toddlers, preschoolers, primary school aged, etc.), other eligibility criteria (poor, with working parents, etc.), funding strategies (government, employer, parents fees, combination, etc.), delivery strategies (supply or demand), locus of care, primary caregiver (professional, paraprofessional, parent, etc.), and program philosophy, as relevant (**Figure 1**).

The US has adopted a different model from most OECD countries, which assumes that parents will carry personal responsibility for the provision of an adequate standard of living, and that they will achieve this through the workings of the market, primarily. US policies toward ECEC follow a similar pattern. The US provides free primary and secondary public school education for children aged 6–16, and kindergarten, for almost all 5-years-olds; but it does not provide the equivalent for younger children. Several states (Georgia, Oklahoma, and New York) provide universal prekindergarten, part-day programs for 4-year-olds, often in a range of settings, including schools and community-based programs.

Nor does the US ensure universal, equitable access to ECEC. In 2005, 72% of children under age 5 had a primary nonparental care arrangement in the US primarily care by a relative. About half of the 19 million children in the US under age 5 were cared by a relative, usually a grandparent or a parent. Among the 11.3 million children younger than 5 whose mothers were employed, 30% were cared for on a regular basis by a grandparent during their mother's working hours. A slightly higher percentage was cared for in an organized facility such as a center or preschool (US Census Bureau News, February 28 2008). One-third of infants under age 1 are in relative care and those in nonrelative care are often in family day care (child mindres). Furthermore, working parents often choose to work different shifts in order to ensure parental care for their young children. When they do purchase child care, US poor and near-poor parents pay a disproportionate share of their income for ECEC services (in 2008, 29% and 15% respectively) in contrast to 6% for upper income families. (US Census, 2008). In the US, families continue to bear a far larger part of the operating costs of ECEC services than parents in other industrialized countries.

Figure 1 ECEC in the United States: A market approach.

Around the world, the family continues to carry the major responsibility for childcare, early-childhood education, child rearing, and socialization. However, increasingly – both in developed and developing countries – with the rise of women's participation in the labor force, the family's role is supplemented by other institutions. The trend toward the greater involvement of nonfamily resources is characteristic of the industrialized countries, in particular, but there is great variation with regard to which institutions, employing which roles, how they are paid for, and by whom.

This article, next, reviews the distribution of responsibility for the education, care, support, and socialization of young children in the OECD countries among the family, the market, and the government. Fundamental questions remain with regard to the appropriate role of each in the early years of young children's lives. In particular, two cross-national issues are: how should responsibility be allocated among these institutions with regard to policy strategies such as funding, production, and regulation of service quality? Also, what are the implications for quality, access, and equity?

The Role of the Family

The family continues to have the primary responsibility for children and their support internationally – as it has historically, but not exclusively. Several different strategies have been employed to support the family in its role. Most importantly, women (married as well as single) have entered the labor market to compensate for the inadequacy of the traditional family model of two parents

and one wage-earner – because one average wage is not adequate to support a family with children. In addition, there has also been a rise in female-headed, single-parent households.

In addition to the financial support provided by mothers entering the labor force, major policy strategies providing financial support for family care include those that the government gives to the family to offset the costs of childcare, early education, and child rearing. What the European countries, for example, have done to support family care – where they believe it is especially important – is to enact policies that subsidize the family in carrying out caring tasks.

Child Allowances and Tax Benefits

One form of financial support for the provision of family care is a child or family allowance – a cash benefit provided to families with children, based on the presence, number, age of children, and, sometimes, the ordinal position of the child and the employment status and income of the parents. These cash benefits are frequently employed for supplementing earned income for families with children (SSWTP, 2007; Kamerman and Gatenio, 2002; Kamerman *et al.*, 2003; Kamerman, 2005).

A second form is a tax benefit, credit, or allowance – another device for supplementing earnings or off-setting some of the costs of child rearing or childcare. In some countries, family allowances (and tax exemptions for dependent family members) have been replaced or supplemented by tax credits. For example, the United States provides a tax benefit for modest-income working

families with children, and another tax benefit for families with incomes above the tax threshold where children are present. However, the United States has not provided a universal cash benefit as in other countries, nor does it have any other fully universal child benefit. The provision of the Earned Income Tax Credit (EITC) and the Child Tax Credit offer some evidence that the United States acknowledges some level of responsibility for assuring families with children adequate income, but it is only partial, at a limited level, and only for some families with children. The US family-benefit package provided through the tax system is nowhere near as generous as the comparable tax and cash-benefit packages provided by many other countries (Kamerman and Gatenio, 2002).

Third, cash benefits are provided as a guaranteed, minimum child-support benefit advanced by government if the noncustodial parent does not pay maintenance or pays it irregularly or at an inadequate level – and then collected to the extent possible from the noncustodial parent.

Leave Policies

A fourth form of financial support is the enactment of paid and job-protected leaves from employment for working parents – supporting them in their parenting and infant-care roles, providing financial support, and parental care, etc. In OECD countries, the under 3s are especially likely to be cared for by family members, or if not, by family day-care providers. In recent decades, many European countries have enacted a different approach to supporting the family's role in providing infant and toddler care. This comprises a policy of paid and job-protected maternity parental, child rearing, or childcare leaves following childbirth or adoption – lasting 1–3 years, sometimes limited to working parents and, other times, for all parents – fully or partially replacing wages foregone while on leave (Kamerman, 2000a, 2000b; Waldfogel, 2001). The European Commission (EC) enacted a maternity leave-policy directive in 1986, providing for a 14-week paid and job-protected leave for working women at the time of childbirth, followed, in 1992, with a directive providing for a 3-month unpaid albeit job-protected parental leave for each parent.

Sweden is known for its generous leave policies. It provides an 18-month job-protected parental leave following childbirth or adoption. For 1 year on leave, a working parent receives a cash benefit replacing 80% of prior wages. For the next 3 months, a parent on leave receives a flat-rate cash benefit and the final 3 months are unpaid. In addition, fathers can obtain an additional 2 months of paid leave, but this is only available to fathers; if not claimed by the father, it is lost. The objective is to make it possible for parents to take care of their infants at home instead of placing them in out-of-home care, without being subject to a substantial economic penalty.

In October, 2008, the at present 27-country European Union (EU) unveiled a package of proposals in the social policy area. The proposed family-policy package amends the existing maternity and parental leave policies by extending the existing maternity leave from 14 to 18 weeks, with a benefit replacing full wage, and with more flexibility with regard to when the leave can be taken (before or after childbirth). The Commission announced that it would also launch negotiations on other forms of family-related leaves with the social partners (employers and employees organizations) such as paternity leave (a brief period of leave for fathers at the time of birth or adoption of a child), adoption leave (similar to maternity leave following childbirth), filial leave, (a leave to care for dependent family members), and parental leaves. The aim is to conclude negotiations and reach consensus within 9 months. With regard to such policies as well, the United States follows a different path, stressing family responsibility without societal support.

Across Europe, care leaves allow parents to take job-protected leave (often paid at a low, flat rate) and to extend maternity leaves for up to 3 years. These policies provide an alternative to subsidizing expensive infant care. Hungary initiated such a policy – first, in 1967 – while Finland also enacted a 3-year job-protected, modestly paid leave, in 1990, contingent on nonuse of publicly funded childcare or, as an alternative, with the right to a place in public childcare from the age of 1 year. Yet another policy package is that provided by Iceland – where a job-protected, fully paid 9-month leave is provided. Three months are assigned to the mother, another three months to the father, and the final three months are a family leave, with the parents able to decide who should take primary responsibility. Increasingly, attention is being paid to fathers – with the goal of achieving gender equity. The objective is for fathers to participate equally in providing childcare as well as economic support. Parental leaves of long duration have also been used to help fight unemployment by providing incentives for mothers to stay out of the workforce (Morgan, 2006a).

Across countries, most of the care provided by family members is unpaid, but this does not mean it comes without costs. Parents pay for the care they provide themselves by forgoing current earnings from the labor market and by reducing their likely future earnings; they, often, pay for the care provided by other family members by trading goods or services. These costs can be substantial and need to be taken into account when thinking about the resources expended in this sector (Young and Nelson, 1973). Most of the financial, time, and emotional burdens of rearing children (and the pleasures) are borne by parents, sometimes aided by other relatives as well. However, an increasing share is provided outside the family, either purchased in the market or provided by or funded by government, or provided by the family but with the help

Table 1 Enrolment rates in childcare and early education for children under six (2004)

	<i>Enrolment in daycare for the under 3s and preschool from 3 to 6 years (%)</i>				<i>Expected years in education for 3–5-year-olds</i>
	<i>Under 3 years</i>	<i>3 years</i>	<i>4 years</i>	<i>5 years</i>	<i>3–5 years</i>
Australia ^a	29.0	55.0	64.6	90.9	1.8
Austria	4.1	45.9	82.1	93.1	2.2
Belgium	38.5	99.3	99.9	99.7	3.1
Canada ^b	19.0
Czech Republic	3.0	68.0	91.2	96.7	2.6
Denmark ^a	61.7	81.8	93.4	93.9	2.7
Finland ^c	22.4	37.7	46.1	54.6	1.4
France ^d	26.0	100.0	100.0	100.0	3.2
Germany ^b	9.0	69.5	84.3	86.7	2.4
Greece ^c	7.0	..	57.2	84.1	1.4
Hungary	6.9	71.0	92.3	97.8	2.6
Iceland ^c	58.7	93.3	95.1	95.9	2.8
Ireland ^e	15.0	48.0	46.6	100.0	1.5
Italy ^e	6.3	98.7	100.0	100.0	3.0
Japan	15.2	67.3	95.2	96.6	2.6
Korea ^a	19.9	59.5	66.4	88.7	0.9
Luxembourg ^c	14.0	37.9	83.5	96.9	2.2
Mexico ^c	3.0	22.1	66.4	95.9	1.8
Netherlands	29.5	32.3	74.0	98.4	1.7
New Zealand	32.1	82.1	95.1	100.0	2.8
Norway ^c	43.7	79.4	86.9	89.0	2.6
Poland ^b	2.0	26.1	35.7	46.2	1.1
Portugal	23.5	63.9	79.9	90.2	2.3
Slovak Republic ^c	17.7	60.3	71.7	84.7	2.2
Spain	20.7	95.9	100.0	100.0	3.1
Sweden	39.5	82.5	87.7	89.7	2.6
Switzerland	..	7.2	34.4	89.7	1.3
Turkey	..	1.7	3.4	26.2	0.3
United Kingdom	25.8	50.2	92.0	98.2	2.4
United States ^a	29.5	41.8	64.1	77.0	1.8

^aYear of reference 2005.^bYear of reference 2001.^cYear of reference 2003.^dYear of reference 2002.^eYear of reference 2000.

From OECD Family and Education databases.

of government – as has already been discussed in the previous sections. Next, the role of the market is discussed (Table 1).

The Role of the Market – The Private Sector

(This section draws on Kamerman and Waldfogel in Nelson (2005) and Kamerman and Kahn (1989).)

The terms public and private childcare are commonly used to distinguish among the main auspices of ownership of regulated childcare. These are, generally, viewed as including three main types:

- Public – owned and operated by a level of government (national, state, or local).

- Private nonprofit sector (or voluntary sector) – owned and operated by a nongovernmental educational or social service organization (nongovernment organization (NGO)).
- Private for-profit sector – owned and operated by an individual, a partnership, or a small or large company. A for-profit childcare company may be a publicly traded company (corporate childcare) traded on a stock exchange or a private company.

According to the *UNESCO Statistical Yearbook*, a formal – albeit slightly different – distinction was first made between public and private ECEC programs in 1961. Public preschools were those operated by a government agency, irrespective of whether or not funded by the government; and private preschools were those operating

under private auspices, whether or not receiving private funds. Private providers continue to be of two types: (1) those established by NGOs, women's groups, and religious institutions; and (2) those established by private individuals for profit. Private establishments tend to be regulated when subsidized by government or included in the public system.

Private Provision of ECEC Services

How large is the private sector for ECEC today? In most of the OECD countries, ECEC programs are publicly funded and delivered. The portion of children enrolled in private programs – largely under religious auspices or other types of nonprofit sponsors – range widely across countries from being a major component of the delivery system in Australia, the United States, the United Kingdom, New Zealand, Germany, and the Netherlands to playing an insignificant role in the Nordic countries. In Australia, for example, the private, for-profit sector provides about 70% of full-day child-care centers (~70%), and at least 25–30% of these are held by the ABC Learning Corporation that recently went bankrupt (see below).

In Europe, the privatization of early care and education, since the 1980s, has occurred in partnership – rather than in competition – with the state. A range of alternative forms of provision has emerged, albeit unevenly, to accommodate diverse families and diverse communities. In France, nonprofit and voluntary organizations involved in child care proliferated during the 1980s and 1990s (Ullman, 1998). Since the late 1970s, nonprofits have grown from a little over 10% to more than 40% of all childcare providers. These organizations tend to be small – and in contrast to the past – most are not religiously affiliated. Much of the growth in the nonprofit sector is linked to the development of parent cooperatives (Eme and Fraisse, 2005). French local policymakers increasingly recognize the potential of private programs to provide flexible services for parents who work nonstandard hours or live in rural areas where there are not many family supports. In 2003, for the first time, private, for-profit *crèches* could receive public subsidies to address the supply shortages, but the issue remains ideologically charged (Neuman, 2007).

In Sweden, the proportion of children in preschool who attend nonmunicipal provision grew from 5%, in 1990, to 17%, in 2002. Parent cooperatives still account for nearly half the places in nonmunicipal preschools. Pysslingen – the largest for-profit early-care and education provider in Sweden (established in 1984) – expanded from six preschools in 1991 to more than 60 in 2007. Two-thirds of private preschools are cooperatives and one-third are for profit. Even if the private sector constitutes less than 20% of the total provision, it remains an influential part of the welfare state contributing to the

expansion and diversification of early-care and education services (Vamstad, 2005). It is important to emphasize that the growth of the private sector, in most of Europe, has occurred within the infrastructure of a carefully publicly regulated and funded system, so that even the differences between France and Sweden, for example, will be small relative to those nations that have adopted a more free-market approach to ECEC.

The standard arguments in favor of a private sector role in ECEC include: (1) The likelihood of the private market producing a larger supply; (2) freedom of choice for consumers; (3) the greater efficiency among for-profit providers; (4) support for individual choices versus societal choices; and (5) the greater responsiveness of the private market to consumer preferences. The major arguments against a role for the private sector, or limiting its role, include: (1) the inadequate supply of ECEC services produced by the private market; (2) inequitable access to ECEC services; (3) the reluctance of market providers to invest in poor neighborhoods exacerbating the problem of inequity in access, especially with regard to low-income families; and (4) the inadequate quality of ECEC services produced by the private market. This section focuses on the impact of privatization on quality and access issues.

Impact of the market on quality

There has been a long-standing debate in the childcare literature as to whether for-profit providers operate differently and – in particular, whether they offer lower-quality care, on average – than nonprofit providers. Most of the research has focused on countries with large for-profit sectors (Australia, Canada, New Zealand, United States, etc.). Although generalizations are hard to draw – given that there is a range of quality in both settings – the weight of the evidence suggests that these providers do operate differently and that on average for-profit providers do offer lower quality care than nonprofit providers (see, e.g., Gormley, 1995; Smolensky and Gootman, 2003; Cleveland and Krashinsky, 2003). In the landmark Cost, Quality, and Outcomes Study in the United States, for example, states with low licensing standards had lower quality for-profit centers than nonprofit centers (Morris and Helburn, 2000).

To make a profit, centers may cut costs by hiring poorly trained staff at low wages – leading to higher staff turnover, which jeopardizes quality. A recent research review of privately owned and community-owned early-childhood education provision in New Zealand found statistically significant differences in the employment of staff holding a teaching qualification, and of staff holding no early-childhood qualification between community- and privately owned education and care centers (Mitchell, 2002). Private centers tend to employ fewer staff with ECEC qualifications and they also tend to pay their staff less than community centers. These findings are consistent

with research in Canada and the United States. An analysis of four Canadian datasets with information about quality of childcare services and for-profit/nonprofit status of centers found significant quality differences between nonprofits and for-profits across Canada. In larger or thick markets, they are able to attract better staff, directors, and leadership because of their mission and ability to produce better-quality care for children at an affordable price. Even with financial resources held constant, nonprofits are better at producing quality services, with an advantage ranging from 7.5% to a 22% across the datasets. Only in thin markets, with less competition between public and private centers – and with inadequate parental ability to pay for higher quality care – did the researchers find no average quality difference between for-profits and nonprofits (Cleveland and Krashinsky, 2005).

It should be noted that most countries require private and public providers to follow the same quality standards if they are to receive public subsidies and, in many cases, if they are to operate at all (e.g., Belgium and Denmark). Regulating private early-childhood services, in France and Sweden, appears to help reduce quality disparities across different forms of provision. In both France and Sweden, the nonprofit sector was viewed as a model for parental involvement and innovative pedagogy – although these programs also have difficulty meeting the same quality standards as public services. By increasing innovation and responsiveness to local needs of families (e.g., parental involvement and different pedagogies), privatization improves some aspects of quality. The regulation of staff training and ratios are critical as these two variables are among the strongest predictors of quality (Neuman, 2007).

Impact of the market on access

An earlier analysis (Kamerman and Kahn, 1989) found that childcare services constitute the prototypical illustration of privatization and a stress on the private market as an explicit social service policy. The childcare industry has always been a mixed economy in the United States as in Australia, Canada, New Zealand, and the United Kingdom. Privately funded and operated programs have always coexisted with totally public programs. Publicly funded, but privately delivered, child-care services came to dominate the childcare market in the United States as the supply of services increased during the 1970s and 1980s, in response to federal regulations which permitted public funds to be used to purchase services from private nonprofit providers, and the Child and Dependent Care Tax Credit permitted working parents to purchase services from for-profit providers and still benefit from the tax credit. Subsidizing demand created an incentive for parents/consumers to purchase care in the open market and encouraged private sector providers to produce and deliver services. It is also worth emphasizing that the private sector in the United States consists of more for-profit

providers, and fewer nonprofit providers, than is the case in many other countries.

Another difficulty is when one corporation becomes a relative monopoly by purchasing other smaller providers that fail to make a profit. In Australia, the world's largest childcare company went bankrupt in November 2008 – when it could not repay the \$1.54 billion debt that it acquired during an expansion into the United States and the United Kingdom. ABC had expanded rapidly, since the late 1990s, when the conservative government ended subsidies to nonprofit providers in favor of a market-based approach that gave parents tax rebates for purchasing care. The downfall of a company which covers 25% of the childcare population will have potential consequences for the tens of thousands of working parents stranded without childcare services – and, of course, to the 120 000 children enrolled in ABC Centers. (In contrast, the largest child care providers in the United States and Britain control only about 2% or 3% of child-care places.) Children's advocacy groups have called on the government to rethink its market-based approach to childcare which they argue has led to higher costs, fewer options, and lower quality (Foley, 2008).

If ECEC is viewed as an important and increasingly essential service, the challenge is how to ensure equitable access. Leaving ECEC to the market will not redress that problem, since – without governmental intervention – only the more affluent families will be able to afford high-quality care. A further concern is that the private sector or the voluntary agencies may siphon off the more affluent consumers or the more able children while compromising access and quality for the poor by leaving the more difficult cases in the hands of an underfunded voluntary sector or an understaffed public sector. In countries such as France, Germany, Sweden, and the Netherlands, the private sector (nonprofit and, to a lesser extent, for-profit) expanded supply, as well as diversifying the types of programs available (e.g., part-time and parent cooperative). There are equity concerns – as low-income families tend to be underrepresented in private early-care and education programs, including parent cooperatives. In sum, greater market involvement can expand options and choices for parents, it may also foster inequities of access, quality, and cohesion between public and private services (OECD, 2001). Although there has been some increase in private provision recently, the public role remains overwhelmingly dominant. Governments expand their supply by funding and operating more programs or by increasing subsidies offered to providers.

The Role of Government

The role of the government in ECEC has increased as a means to support working parents – especially mothers – with young children, as well as to promote young

children's early development and learning and their success in formal schooling. As a result, several OECD countries now guarantee children a place in ECEC when they reach a certain age (age 1 in Denmark and Sweden; age 2.5 in France and Belgium; age 3 in Germany and Italy). More than 60% of 3-year-olds in the EU were enrolled in day-care or preschool programs, in 2004, and 21% of the under 3s (OECD Family Data Base).

Governance or administrative auspice is another key dimension affecting ECEC policies and programs. The major difference is whether the auspice is education, health, social welfare, or some combination, and, if a combination, whether responsibility is carried out sequentially (as children get older the auspice shifts) or simultaneously. The dominant European pattern (and increasingly in developing countries) is one in which the programs serving the 3–6-year-olds, until compulsory school begins, are cared for under education auspices while the younger children are cared for under health or social welfare auspices. There is an emerging trend in Europe to include care of the under 3s also under education – as can be seen in England, Norway, Scotland, Slovenia, Spain, and Sweden – in addition to such non-European countries as Brazil, Kazakhstan, and New Zealand.

As provision of care and education of young children increasingly shifts from families to nonfamily institutions, one question is what role belongs to the government? In reviewing the role of government in ECEC internationally, government clearly plays a major role – in the funding, production/delivery, and regulation of ECEC. In terms of funding, in most of Europe, the costs of enrolling a child in ECEC is heavily subsidized, and fees are often determined on a sliding-scale basis according to parental income. Sweden enacted a policy whereby a maximum fee for ECEC programs for all children is set at no more than 2–3% of family income to shift some of the cost burden from families to the state. (In the 1990s, during difficult economic times, the parent share of the total cost of preschool had increased from 10% to about 18% (Blomqvist, 2004).) Preschool in Belgium, France, and most of Italy, covering all children in the age range of 3–6, is universal and free as a part of the educational system. Universal, free preschool for 4–5-year-olds was introduced in England and Sweden in more recent years.

What is the government's role in directly providing these services? All the advanced industrialized countries have moved toward government having the dominant role with regard to primary school and some countries have gone further by beginning primary school at the age of 5 or 4 (e.g., United Kingdom and Netherlands). Even in the United States, there is a good deal of movement toward universal preschool at age 4 – a position adopted by governors of 43 states. Government's role is especially strong in those countries that have adopted an educational framework or model for their ECEC programs, especially

for those serving children from 3 to compulsory school age, whether by delivering ECEC services directly or publicly funding NGOs to deliver these services. ECEC preschool programs serving children of this age are operated, largely, under education ministries or departments and viewed as part of the education system.

In contrast, ECEC services for the under 3s are more diversified, involving governmental support but also voluntary agencies and the market as well. Moreover, in some countries, nonfamily care of the under 3s is largely family day care or child minding, with funding and delivery, largely, private but, sometimes, with governmental regulation and monitoring – another role assigned to government. Although there is little debate with regard to a significant role for government in preschool programs, the measures used for distributing the funds remains a matter for debate, especially among programs serving the under 3s.

A second question is whether government should provide financial support for care delivered by the private sector (nonprofit or for-profit). If government does not provide the services directly, what is its responsibility for funding these services? In addition, if it does provide financial support for these services, a third question is – what measures should government use for distributing the funds, especially among programs serving the under 3s? Should it use demand-side or supply-side measures to allocate its investment in ECEC – should it subsidize the parent/consumer of ECEC services or the provider? The main issues are choice, quality, and access and whether these dimensions are enhanced or diminished in the private sector or through demand-side subsidies?

The Netherlands has opted for a unique demand-side approach with a tripartite financing of government, employers, and parents. There has been spectacular growth in the number of childcare places – from 22 000 places in 1989 to 206 714 in 2004. In 2004, each place was occupied by about 1.6 children. The 2005 Childcare Act seeks to further stimulate supply by creating more choice for parents and a greater scope for market forces. Employers are required to cover one-sixth of childcare costs per parent; the government steps in only if the employer fails to contribute. The Act includes only limited quality requirements. Services compete in terms of price and quality, and the impact for children needs to be monitored (SZW, 2005). Hasan (2007) cautions that demand-side subsidies (tax benefits or vouchers) assume that parents have adequate information to choose from among the alternative programs. However, parents may not have adequate knowledge to select the best-quality program and poor-quality providers may drive quality providers out of the market. In contrast, supply-side measures may end up involving higher costs.

The OECD reviews (OECD, 2001, 2006) pointed to the need for substantial public investment in ECEC

services and the infrastructure to support a sustainable system of quality, accessible services. Without governmental investment, regulation, and supervision, children's services will be disorganized, and of low quality. Government alone may not guarantee quality, but without government funding or regulation, quality is likely to be poor, programs weakly monitored, and access constrained (Bennett, 2007).

A final issue is with regard to government's role in relation to parental leave policies, involving labor-market policies with regard to job-protected and paid parental leaves from employment. Clearly, if there is no statutory provision, leaves are likely to be inadequate, access limited, and coverage sparse – as can be seen in the United States and Australia, for example.

Conclusions

Although the family continues to have primary responsibility for the care and rearing of its young children, it is no longer the institution with exclusive responsibility and the role of nonfamily institutions continues to increase. Given the increased time that children spend in care, education, and socialization carried out by nonfamily institutions, adequate and equitable access to decent-quality care is essential. The role of the family in ECEC is obviously important; however, the family can no longer do the job by itself – meaning that government or the market must step in – because if decent-quality ECEC services are not available for preschool-age children, the children, the schools, and the society will suffer.

History shows us that when families cannot meet the education and socialization needs of their children, the government has stepped in – as it did when establishing public primary schools. Experience has shown that when families cannot meet the care and education needs of their preschool-age children on their own, the private sector alone will not do an adequate job. Government must play a role – whether as a provider of services directly or as a funder and regulator of services provided by the private sectors.

The trend toward the greater involvement of nonfamily resources in ECEC has been observed in all industrialized countries. Where countries have diverged is in who has picked up this responsibility. In the United States, policy in this area has tended to look first to the family and private sector and to look to government only in the event that the market fails or if there are compelling reasons of public interest for the government to get involved. In European countries, in contrast, policy has tended to look first to the government, and the role of the private sector in the ECEC sector has been minimal. Where the private sector exists, it has developed within the framework of a publicly regulated and funded system which minimizes disparities in access and quality.

If society values quality ECEC services, either government will provide the services, or it will provide the resources and ensure adequate quality, or the private sector will provide the services, and the government will subsidize access and regulate them. Thus, key policy issues are to ensure the adequacy of the resources provided (largely requiring public funds); expanding equitable access irrespective of income; recognizing differential quality and good management and accountability to ensure quality; and acknowledging the importance of government supporting families in their parenting roles.

Access (the ease of obtaining a place in an ECEC program) and coverage (the percentage of an age-cohort enrolled in these programs) are clearly important indicators of a country's commitment. The most extensive access and most complete coverage are in those countries with a strong governmental role both in funding and delivery of ECEC. The highest rates are for children enrolled in programs under public education auspices. Between 95% and 99% of this age group are enrolled in the public, voluntary, free, and universal programs in Belgium, France, Italy, and Spain – covering the normal school day, often with supplementary after-school services as well. The EU Barcelona targets of at least 90% coverage of 3–6-year-olds in ECEC by 2010 have largely been reached, with strong governmental support, although not all coverage is for a full workday. The target for the under 3s, at 33%, is close to being reached if coverage by parental leave policies are included. Looking ahead, the missing link in the Barcelona targets are guidelines for adequate resources and appropriately trained and qualified staff – critical components of quality.

See also: Child Rearing and Early Education: Parents and Professionals: Theoretical and Historical Influences; Gender Issues in Early Childhood Education and Care; Investing in Early Childhood Education and Care: Some Policy Implications; Investing in Early Childhood Education and Care: The Economic Case.

Bibliography

- Bennett, J. (2007). *Results Form the OECD Thematic Review of Early Childhood Education and Care Policy (1998–2006)*. Paris: UNESCO Policy Brief.
- Blomqvist, P. (2004). The choice revolution: Privatization of Swedish welfare services in the 1990s. *Social Policy and Administration* 38(2), 139–155.
- Cleveland, G. and Krashinsky, M. (2003). *Financing ECEC Services in OECD Countries*. Paris: OECD Discussion Paper.
- Cleveland, G. and Krashinsky, M. (2005). *The Nonprofit Advantage: Producing Quality in Thick and Thin Child Care Markets*. Toronto: University of Toronto at Scarborough, Department of Management.
- Foley, M. (2008). Should child care be at the mercy of the market? *International Herald Tribune*. <http://www.ihrt.com/bin/printfriendly.php?id=18210827> (accessed September 2009).
- Gormley, W. T. Jr (1995). *Everybody's Children: Child Care as a Public Problem*. Washington, DC: The Brookings Institution.

- Kamerman, S. B. (2000a). Parental leave policies: An essential ingredient in early childhood education and care policies. *Social Policy Report* 14(2), 3–15.
- Kamerman, S. B. (2000b). From maternity to parenting policies: Women's health, employment, and child and family well-being. *Journal of the American Women's Medical Association* 55, 96–99.
- Kamerman, S. B. and Gatenio, S. (2002). Tax day: How do America's child benefits compare? *Issue Brief* 2002 Springer 2000. Columbia University.
- Kamerman, S. B. and Gatenio, S. (2003). Overview of the current policy context. In Cryer, D. and Clifford, R. (eds.) *Early Childhood Education and Care in the United States*, pp 1–30. ECEC Background Report. Baltimore, MD/Paris: Brooks Publishing/OECD.
- Kamerman, S. B. and Kahn, A. J. (eds.) (1989). *Privatization and the Welfare State*. Princeton, NJ: Princeton University Press.
- OECD (2001). *Starting Strong: Early Childhood Education and Care*. Paris: OECD.
- Waldfoegel, J. (2001). What other nations do: International policies toward parental leave and child care. *Future of Children* 11(1), 99–111.
- Young, D. R. and Nelson, R. N. (1973). *Public Policy for Day Care of Young Children: Organization, Finance, and Planning*. Lexington, MA: Lexington Books.
- Kaplan, J. (2002). Child care funding and policy issues. *WIN, Reauthorization Notes* 2(1).
- Karoly, L. A., Greenwood, P. W., Everingham, S., et al. (1998). *Investing in Our Children: What We Do and Don't Know About the Costs and Benefits of Early Childhood Interventions*. Santa Monica, CA: RAND.
- Layzer, J. and Collins, A. (2000). National study of child care for low-income families: State and community substudy. *Interim Report*. Cambridge, MA: Abt Associates.
- Magnuson, K. A., Meyers, M., Ruhm, C., and Waldfogel, J. (in press). Inequality in Preschool Education and School Readiness. *American Educational Research Journal*.
- Magnuson, K. A., Ruhm, C., and Waldfogel, J. (2004). Does prekindergarten improve school preparation and performance? *NBER Working Paper* 10452, April 2004.
- Meyers, M. K., Rosenbaum, D., Ruhm, C., and Waldfogel J. (in press). Inequality in early childhood education and care: What do we know? Forthcoming. In Neckerman, K. (ed.) *Social Inequality*. New York: Russell Sage Foundation Press.
- Mitchell, L. (2002). Differences between community owned and privately owned early childhood education and care centres: A review of evidence. *NZCER Occasional Paper* 2002/2. Wellington: New Zealand Council for Educational Research.
- NICHD Early Child Care Research Network and Duncan, G. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development* 74, 1454–1475.
- OECD (2000). Twelve country thematic review of early childhood education and care policy. *Country Background Reports and Country Notes*. Paris: OECD.
- Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., et al. (1999). *The Children of the Cost, Quality, and Outcomes Study Go to School*. Chapel Hill, NC: University of North Carolina.
- Peth-Pierce, R. (1998). *NICHD Study of Early Child Care*. Washington, DC: NICHD.
- Penn, H. (2007). Childcare market management: How the United Kingdom has reshaped its role in developing early childhood education and care. *Contemporary Issues in Early Childhood* 8(3), 192–207.
- Phillips, D. and Adams, G. (2001). Child care and our youngest children. *Future of Children* 11(1), 35–52.
- Phillips, D. and Mekos, D. (1993). The myth of child care regulations: Rates of compliance in center-based child care settings. *Paper Presented at the Annual Research Conference of the Association for Public Policy Analysis and Management*. Washington, DC, October 28–30.
- Pistillo, F. (1989). Pre-primary education and care in Italy. In Olmsted, P. and Weikart, D. (eds.) *How Nations Serve Young Children*, pp 151–202. Ypsilanti, MI: High Scope.
- Pringle, M. and Naidoo, S. (1975). *Early Child Care in Britain*. New York: Gordon and Breach.
- Rose-Ackerman, S. (1983). Unintended consequences: Regulating the quality of subsidized care. *Journal of Policy and Analysis and Management*. 3(1), 14–30.
- Rose-Ackerman, S. (1986). Altruistic nonprofit firms in competitive markets. *Journal of Consumer Policy* (9), 291–310.
- Shonkoff, J. P. and Phillips, D. A. (2000). *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, DC: National Academy Press.
- Smelser, N. and Halpern, S. (1978). The historical triangulation of family, economy, and education. In Demos, J. and Bookcock, S. S. (eds.) *Turning Points*. Chicago, IL: University of Chicago Press.
- Smith, K. (2000). U.S. who's minding the kids? Child care arrangements, fall, 1995. *Current Population Reports*, pp 70. Washington, DC: Bureau of the Census.
- Smith, K. (2002). U.S. who's minding the kids? child care arrangements, fall, 1997. *Current Population Reports*, pp 70–86. Washington, DC: Bureau of the Census.
- Smolensky, E. and Gootman, J. A. (eds.) (2003). *Working Families and Growing Kids: Caring for Children and Adolescents*. Washington, DC: National Academies Press.
- Steinfeld, M. O'B. (1973). *Who's Minding the Children? The History and Politics of Day Care in America*. New York: Simon and Shuster.

Further Reading

- Adams, G. and Rohacek, M. (2002). Child care and welfare reform. In Weil, A. and Feingold, K. (eds.) *Welfare Reform, the Next Act*, pp 121–142. Washington, DC: Urban Institute Press.
- Barnett, S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *Future of Children* 5, 25–50.
- Basile, K. C., Henderson, L. W., and Henry, G. T. (1998). *Pre-Kindergarten Longitudinal Study 1996–97 School Year*. Georgia State University. Applied Research Center, School of Policy Studies.
- Beatty, B. (1995). *Preschool Education in America*. New Haven, CN: Yale University Press.
- Berfenstam, R. and William-Olssen, I. (1973). *Early Child Care in Sweden*. New York: Gordon and Breach.
- Blau, D. M. (2001). *The Child Care Problem: An Economic Analysis*. New York: Russell Sage Foundation.
- Committee on Economic Development (2002). *Preschool for All: Investing in a Productive and Just Society*. New York: CED.
- Cooper, C. (1999). *Ready to Learn: The French System of Early Education and Care*. New York: French American Foundation.
- Currie, J. (2001). Early childhood intervention programs: What do we know? *Journal of Economic Perspectives* 15, 213–238.
- David, M. and Lezine, I. (1975). *Early Child Care in France*. New York: Gordon and Breach.
- Helburn, S. and Bergmann, B. (2002). *America's Child Care Problem: The Way Out*. New York: Palgrave.
- Helburn, S., Culkin, M. L., Morris, J. R., et al. (1995). *Cost, Quality, and Child Outcomes in Child Care Centers*. Denver, CO: University of Colorado.
- Hwang, D. P. and Broberg, A. G. (1992). The history and social context of child care in Sweden. In Lamb, M., Sternberg, K. J., and Broberg, A. G. et al. (eds.) *Child Care in Context*, pp 27–54. Hillsdale, NJ: Lawrence Erlbaum.
- Jamieson, A., Curry, A., and Martinez, G. (2001). School enrollment in the United States – Social and economic characteristics of students. *Current Population Reports*. pp 20–533 Washington: DC.
- Kamerman, S. B. (Forthcoming). Europe advanced while the U.S. lagged. In Heyman, J. (ed.) *Work, Family, and Democracy*. New York: The New Press.
- Kamerman, S. B. and Kahn, A. J. (1976). *Social Services in the United States*. New York: Temple University Press.
- Kamerman, S. B. and Kahn, A. J. (eds.) (1997). *Family Change and Family Policies in Great Britain, Canada, New Zealand, and the U.S.* Oxford: Oxford University Press.
- Kamerman, S. B., Neuman, M., Waldfogel, J., and Brooks-Gunn, J. Social policies, family types, and child outcomes: A review of the literature. *OECD Occasional Paper*.

- Sumsion, J. (2006). The corporatization of Australian childcare. *Journal of Early Childhood Research* 4(2), 99–120.
- Thompson, T. (2002). Testimony before the U.S. House of Representatives Committee on ways and means. *Hearing on the President's Plan to Build on the Success of Welfare Reform*. March 12, 2002.
- Tizard, J., Moss, P., and Perry, J. (1976). *All Our Children; Preschool Services in a Changing Society*. London: Maurice Temple Smith.
- US Bureau of the Census (1987). *Current Population Reports, Series P-70 No. 9. Who's Minding the Kids? Child Care Arrangements Winter 1984–85*. Washington, DC: Government Printing Office.
- US Department of Health and Human Services (2002). *Head Start: Promoting Early Childhood Development*. <http://www.hhs.gov/news/press/2002/pres/headstart.html> (accessed September 2009).
- US House of Representatives Committee on Ways and Means (2004). *The Green Book 2004*. Washington, DC: GPO. <http://waysandmeans.house.gov/Documents.asp?section=813> (accessed September 2009).
- Vandell, D. L. and Wolfe, B. (2000). *Child Care Quality: Does it Matter and Does it Need to be Improved?* IRP, University of Wisconsin – Madison.
- Waldfogel, J. (2002). Child care, women's employment and child outcomes. *Journal of Population Economics* 15, 527–548.
- Woodrow, C. and Press, F. (2008). 'Investigating the reach and implications of the corporatisation of childcare in Australia: The giant in the playground'. *Seminar Presentation, International Centre for the Study of the Mixed Economy*. University of East London.

Gender Issues in Early Childhood Education and Care

B Pocock, University of South Australia, Adelaide, SA, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Carer-citizens – The notion that citizenship can be built around care activities, as well as around other activities like paid work. Contests the notion that citizenship should require participation in paid work.

Dependency ratios – The ratio of those of nonactive age (i.e., not in paid work) to those of active age in a given population.

Gender contracts – The unwritten cultural, social, and economic ‘contracts’ between the sexes which distribute paid and unpaid work and care, and other social and economic resources unevenly between women and men, disadvantaging women.

Gender regime – The prevailing set of social, cultural, and economic policies within which male and female citizens live and work, and which differentially shape outcomes for men and women, often disadvantaging women.

Male-breadwinner households – Households where men are the breadwinners and women the unpaid carers and domestic workers.

Work-care chains – These are the chains of care that are created internationally by changing work patterns as immigrant workers are drawn into caring for the children of those in higher income households and countries, leaving behind care deficits in their countries and households of origin.

‘Work first’ policies – Social and labor market policies that require participation in paid work in exchange for state support, or which give priority to entering paid work.

Working-carer citizens – Workers with caring responsibilities who need access to full citizenship that recognizes these joint responsibilities.

The care and education of young children has, for long, fallen unevenly between the sexes, making the gendered nature of care – and its consequences for women and men – an important aspect of the discussion and practice of the care of children. As mothers, women have been the major carers of children from time immemorial. As the majority of paid workers in childcare and early-education systems around the world since the industrial revolution, women are also differentially affected from men as care-providers, educators, and workers. The value placed on care and care work and its terms and conditions – in both the world of the

private household and the world of the public workplace and nation – significantly shape gender inequality. Care regimes – that is the complex cultural, social, and institutional practices of care – help constitute gender regimes (and vice versa). Beyond the nation-state, cross-country work-care chains also construct international chains that link-up workers who emigrate to provide care in richer, industrialized countries to those who provide and receive care in poorer, less industrialized countries. These work-care chains are also gendered. Thus, early-childhood education and care (ECEC) of children is highly gendered – and the nature of care and work arrangements around children have shaped the spoken and unspoken gender contracts between men and women in many countries – both individually and at the wider national level.

In this article, perspectives on care and women’s equality, the nature of the gender contract, the search for new ECEC policy principles to underpin greater equality for women, some comparative ECEC international experiences and their impacts upon women, and the notion of international work/care trades and gender equity are considered.

Care and Women’s Liberation

Education and care arrangements for children have been central concerns of women’s movements in both their first-wave (in the eighteenth and early nineteenth century) and second-wave configurations (since the 1960s).

A prominent first-wave feminist Mary Wollstonecraft – writing in 1792 – argued for the education of women, including girls, partly on the grounds that better-educated women would make better-educated mothers and carers. More equality for women was justified on the grounds that it would improve women’s mothering and care. Feminist writers, such as Carole Pateman, have coined the term Wollstonecraft’s dilemma to describe the difference between arguments – such as those advanced by Wollstonecraft – for women’s equality based on their biological and social responsibility for social reproduction and care, versus arguments rooted in women’s right to be treated as equal citizens, irrespective of their biology or maternity. Should action fostering women’s equality rest on responses that deal with their differences from men, or should such action aim at equal treatment with men, treating them the same as men without discrimination? The answer to Wollstonecraft’s dilemma is obvious when it comes to consideration of care issues: to ignore their differential effects

on women and men, as if men and women are the same, cloaks one of the main drivers of gender inequality – the differential impact of care – under a mistaken discourse of equal treatment.

From the early 1970s and the first days of second-wave women's liberation, the gendered nature of the care of children has been a concern of women and their political organizations – attracting attention as a practical policy issue as well as a theoretical concern at the heart of understanding gender relations.

For example, among the earliest writings of the women's liberation movement in 1969–72 in Britain, concern with regard to the care of children figured prominently. Alongside equal education and opportunity, equal pay, and free contraception and abortion, 24-h nurseries were among The Four Demands made by the first national women's conference at Ruskin College Oxford in February–March 1970.

For gender theorists, the care of children is a critical issue that sits at the interface where the public and private worlds meet. The male-breadwinner family form – dominant in many industrialized countries in the postwar years – included a female carer, naturalized as the normal child-carer. As Germaine Greer put it in *The Female Eunuch*, in 1970, “Mother duck, father duck and all the little baby ducks. The family, ruled over and provided for by father, suckled and nurtured by mother seems to us inherent in the natural order” (p. 219). The unwinding of the male-breadwinner household in many countries, since the 1970s, has made the issue of caring for children of vital importance to discussion of gender equality in both theory and practice.

The Gender Contract: The Slow Decline of the Male Breadwinner

The contract at the core of the male-breadwinner/female-carer model states that men will work for pay while women will provide care at home and undertake all the tasks of social reproduction – of children, the male worker, the female carer herself, and much of the community in which they are located. In this model, the state may also play some role, for example, helping to provide some public childcare or assistance with paid maternity leave. However, the nature of this private/state sharing varies from nation to nation.

To work, the male-breadwinner model relies on several factors: household/relationship stability; women's willingness to accept economic dependence on a male breadwinner; citizenship for women conferred primarily through household life, care, and maternity; and male earnings high enough to sustain a household with two adults and children (what is, often, referred to in the United States as a family wage). Many threads of this model have unraveled in Western countries in the past 30 years with changing female aspirations, women's active pursuit of independence, rising divorce

rates, declining real individual wages that cannot sustain a family, and a labor market hungry for women's contribution to paid work.

While countries such as Australia embedded the male breadwinner at the core of their wage-fixing system for most of the twentieth century, the proportion of male-breadwinner households has fallen sharply so that they are now in a minority. In most Organization of Economic Cooperation and Development (OECD) countries, the proportion of dual-earner households has increased rapidly, since the 1970s, along with the proportion of single-parent households.

In the context of rapid change in household shape and labor-force participation rates, arrangements for ECEC have been under increasing pressure in many places.

Complex, Competing Drivers for Increased ECEC

Childcare has become a major policy concern in OECD countries, reflecting several drivers. These include increasing female participation in paid work and a push to increase female labor supply; an interest in increasing (or at least moderating the rate of decline in) fertility rates; public policy interest in increasing children's welfare and well-being and protecting the rights of the child; recognition of the importance of quality care early in life to children's optimal development; and a desire to narrow inequality between young children through the provision of quality early-childhood services.

Improvements in ECEC systems have also been motivated by a desire to reduce inequality between women and men, and reshape the gendered nature of economic and social citizenship. However, the relative strength of this motivation varies greatly between countries and economic contexts.

These multiple policy goals do not always sit easily alongside each other: for example, it is possible to rapidly increase childcare provision in ways that compromise quality and, thus, the well-being of children and it is possible to provide care of infants by paying mothers to stay at home and out of paid work, but the implications for gender equity are not always positive.

Who Should Care? Mothers, the State, and the Market?

The balance of ECEC provision by the private household (mostly through the work of mothers), the state, and the market has been shifting in many countries in the face of women's rising labor-market participation. This balance is affected by a complex set of factors including the nature of the gender regime that prevails (e.g., whether supportive

of maternal care or public ECEC provision), the nature of work arrangements and supports (including the availability of paid maternity and parental leave, and flexible work arrangements), and the extent and availability of public childcare, as well as prevailing tax and welfare systems. These arrangements vary widely from country to country. Consider the examples of Denmark, Iceland, Germany, and Australia.

Denmark

In Denmark, in dual-earner households both partners are likely to work full-time prior to and following their children. Parents, generally, have a year off with the birth of each child, earning around 80–90% of their usual earnings in that year. However, as in most countries (even where paternity leaves are relatively generous), men take only a small portion of this leave. There is, in effect, little earnings or career penalty for either parent when they take time out to look after children, and the tax system encourages dual earners. Core working hours, in Denmark, are 37 and long hours (such as those affecting significant numbers of men in the United Kingdom, Japan, South Korea, and Australia) are less common than in many countries. What is more, quality, well-regulated childcare options are available in Denmark. Following the first year of an infant's life, publicly available childcare is, generally, available and its use is socially accepted. Indeed, public childcare is seen by some as preferable to parental care because of the high-quality childcare standards, removing social sanctions against the use of nonmaternal care. One consequence of all of this is that women who take breaks from paid work around their children suffer relatively slight negative effect upon their life-time earnings in Denmark.

Alongside this, couples in Denmark tend to share housework and unpaid care more than in most other developed countries, although still far from equally. While women in Denmark, often, shoulder a full-time job as well as the main load of domestic work – making them tired and pressed for time – they have a high level of social support from the state, a narrow gender pay-gap (women earn around 89% of men's earnings), and the prospect of a retirement without poverty. Interestingly, Denmark enjoys both high female-participation rates (71.7%) and high fertility (total fertility rate (TFR) of 1.8).

Iceland

Even better than Denmark is the example of Iceland, which has the highest birth rate in Europe, high labor-market participation among women, and tops the world on measures of overall human development (such as expectancy, education, and standard of living) and happiness – at least prior to the Global Financial Crisis of 2008/09. With the highest divorce rate in Europe, the country also has a

culture that is relatively accepting of change in family shape, and supports transitions, rather than penalizing them. Mothers and fathers enjoy 9 months shared paid parental leave at close to replacement earnings and some attribute the high levels of well-being in the country to the relative equality that women enjoy.

Germany

Compare this with Germany – ranked much lower on human development measures – where the lack of childcare, marginal part-time work for mothers, a family-based rather than individually based tax system, and – until very recently – limited maternity leave are associated with both low fertility (TFR of 1.3) and low female participation in paid work. Gender pay-inequity is wide, with women earning around 74% of men's earnings.

Australia

Australia scores well on general human development and happiness scales but mothers have been increasing their attachment to paid work and retaining responsibility for domestic work and care at the cost of their personal time – with the majority of mothers, often, feeling rushed and pressed for time (Pocock *et al.*, 2009). Public childcare has increased in availability in Australia in recent years. However, much of it is provided not by the state but by large monopolistic corporate for-profit providers. The proportion of Australian women – who are working before their child is 1 year old – doubled between the mid-1970s and 2001, to reach 36% and many parents look to grandparents for help with childcare because of problems of cost, quality, and accessibility associated with formal care. The gender pay-gap is narrow by international standards but Australia's occupationally based superannuation system implies that only a minority of Australian women can look forward with confidence to a comfortable retirement – given that women work, on average, around half as many effective years as men (20–38). Lively debates with regard to the morality of nonmaternal care for children in Australia have been overtaken by women's behavior with increasing use of ECEC services and most Australians, at present, experiencing some type of formal childcare prior to their entering school.

Women's 'Choice' in Different Contexts

As in most countries, women in Denmark are more time pressured than men, but they do not have to choose between maternity and labor-market attachment, unlike many women in Germany, with long-term consequences for their earnings, pensions, and the welfare and income available to them and their children – especially if they

divorce. These brief examples illustrate how the terms for combining work and family are much more positive for women in some countries than others – with important consequences for gender equity. In this context, cultural norms and institutional arrangements (including parental leave from paid work, the availability of ECEC services, and tax-and-benefit systems) matter a great deal in terms of their implications for gender equity.

The Role of Men in Domestic Work and Care

In most countries – irrespective of patterns of participation in paid work – women do much more domestic work and care than men. According to the OECD, women do 80% of domestic work and childrearing in the European Union (EU), and men's behavior has barely evolved since good surveys of behavior and time-based activities began – except in the United States and the Nordic countries.

In 2006, Australian women, on average, did twice as much unpaid work and care as men and this hardly changed between 1992 and 2006. It appears that women's increasing entry to paid work has hardly shifted the distribution of unpaid work, leaving most of it with women. The hardy persistence of these gendered norms – and men's resistance to its redistribution – remains one of the great puzzles of masculinity and a significant barrier to gender equality, especially in relation to housework and the care of children.

One of its consequences is the low value – rhetoric aside – placed on mothering and all care work. This valuation casts a long shadow in the labor market, with many forms of feminized work – such as care and early-childhood education work – paid much less than other forms of work that require less skill, training, and effort and involve less responsibility. In Australia, for example, childcare workers are among the country's lowest paid workers, despite the demands of their work.

Conceptualizing Gender Equity in a More Individualized World of Work

The decline of the male breadwinner/female carer model throughout OECD countries has given rise to discussion of an adult worker model – where it is assumed that individualized men and women both work and both care. The key issue in considering gender outcomes in this adult-worker state is the actual distribution of work and care and the terms on which they occur. Variable conditions of work and care lie beneath summary measures of participation in either work or care. For example, many women who manage to enter paid work do so through insecure part-time – or by undertaking complex juggles of work/care – arrangements which leave them pressed for

time and, sometimes, making use of poor quality or unsafe childcare. In this context, the quality of paid work, its rewards, and the ease of its combination with care become very important, as does the capacity of the state (or the market) to help meet the care gap when more adults work.

While some suggest that decisions around care and paid work are simply a matter of choice, the country differences described above show that such choices are socially embedded and constructed by cultural contexts (especially dominant beliefs with regard to the role of women, mothers, and fathers), institutional arrangements, and social policies. Real choice – especially for women – depends upon considerably more than personal decisions. In addition, gender equality depends on much more than equal participation in paid work. For many women, this means the right to add paid working hours to an unchanging and disproportionately large care and housework load.

Work, Care, Gender, and Citizenship

Many OECD countries are introducing work and social policies that – embedded in a discourse of rights and responsibilities – require work of those who can do it including mothers in exchange for social supports (for example, work-first policies). However, the attachment of the rights of full citizenship to the working citizen should not exclude from full citizenship those who withdraw from work to care, or those who combine work with various forms of care over the life cycle – as carer-citizens and working-carer citizens, respectively. In addition, it cannot be assumed that all care can or should be commodified – given that parental care of children is much more than mere work: it is relational and emotional.

In this light, some writers have suggested that all workers should have both the capability (i.e., the real choice underpinned by necessary supports) to work and to care. Achieving this relies on a menu of commodified care options, some of which are within the family (such as paid leave and/or payments to parents allowing them to withdraw from paid work to provide care) and some of which are defamilialized through the establishment of ECEC systems and/or subsidies to parents to organize nonparental care. The outcomes of these various policy options have different gender effects: for example, the provision of paid maternity leave will assist women but may embed mothers as carers unless accompanied by paid paternity leave, with some incentive to fathers to take it. Cash payments to parents to stay at home to look after children, in place of using a state-provided childcare place, increase pressure on mothers to withdraw from paid work – to provide care, with long-term effects on life-time earnings and employment possibilities. In other words, different policy settings and options have contrasting effects on gender equity.

The EU has set a female participation-rate goal of 60%, by 2010, to meet economic objectives. However, achieving this without widening gender inequality or facilitating the movement of children into poor-quality care relies on an expansion of the supply of quality non-familial care as well as quality jobs. At present, European approaches to these challenges are diverse, with countries in Eastern Europe moving to reduce state-based supports for services such as long, paid parental leaves and child-care, while western European countries have approaches that vary from the traditional family-care model of Southern Europe to the caring-state models of the Nordic countries. It appears, however, that there are some signs of convergence in Western Europe and the United Kingdom, New Zealand, and Australia, with greater public responsibility for childcare and increasing care-related social rights (including improved rights to paid leave).

Policy Directions for Gender Equity When Mothers Work

Societies that are to enjoy high levels of gender equity – while increasingly relying on the paid work of mothers – have to provide a number of practical policy supports. These include quality, affordable, accessible ECEC; long parental leaves; quality part-time jobs and flexibility at work; and work practices that permit workers – especially women – to reconcile paid work with care. Real choices for mothers – with regard to their work and care options – rest on an active state that meets the differences in the situations of women and men and mothers and fathers, with support. While not all households will evenly share unpaid housework and care – and this gendered inequality has important effects on women’s labor-market standing – without significant provisions such as publicly provided ECEC systems, there are few prospects of genuine equality for women. What is more, the effects of poor ECEC systems do not fall evenly in terms of social class – with poor women most likely to have to make use of inferior-quality care, to take poor-quality work, or return to work more quickly than they would prefer following the birth of a baby, in pursuit of an adequate income.

International Care Trades to Underpin Work

Beyond the national level, international gendered circuits of care and work are very visible. A gendered global care economy exists, underpinning labor markets. Rising dependency ratios in every OECD country (defined as the ratio of those of nonactive age to those of active age in a given population) are predicted to increase significantly over

coming decades, especially following 2010 and particularly in the EU and Japan. This reflects the declining birth rate and increasing life expectancy. Thinned care capacities do not speak well for the future of care in some OECD countries. As a result, many are entering new international care trades with poorer countries – with particular implications for women.

Rapid population growth in poorer countries – accompanied by high levels of national poverty – is accelerating immigration to countries facing a shortage of carers. This has important implications for global care and work-flows. Rachel Salazar Parrenas, in her 2001 book, *Servants of Globalisation*, describes how 6.5 million Filipino immigrants, mostly women, are at present residing in over 130 countries as domestic workers, cleaning and caring for children in wealthy countries and households.

These international circuits of care are not only gendered: they are also racialized. Rich white nations (such as the United States, Canada, the EU, and Australia) draw on, mostly, women of color from Asia and Africa to do their care and dirty work – usually at very low pay. This creates an international, gendered distributional care injustice, as care is drawn upward and away from the country of origin, where a local care-deficit results.

This has important personal implications in the family networks and dependents in families, where poor workers of color provide care to the white rich at the cost of caring for their own families who may miss out on adequate care as a result – resulting in a tiered transfer of reproductive labor and care. The quality of care – and certainly its rewards – often, deteriorates as care is passed down an international chain of care and the poorer Third World takes on a share of the care and reproductive work essential to labor participation in richer First World countries.

In Arlie Hochschild’s words, a ‘care drain’ thus enjoins a ‘brain drain,’ as private acts of immigration are undertaken to meet the public problem of a First World care-deficit – at great emotional cost arising from what she calls ‘global heart transplants.’ At the bottom of the international hierarchy of reproductive care are the children of the end-carers in the Third World. Rich nations can mitigate the care deficit, created by women’s entry into paid work, by drawing on imported female carers, but only by creating a care deficit down the global care chain – one that poorer communities are least equipped to meet. A labor of grief is created both for the mothers who leave their children behind and the children who plead for their mothers to come home. Some argue for a global ethic of care to meet this problem which creates obligations for those who draw on the care of poor female immigrants and drain the quality of care in the country and family network of origin. The international character of care chains creates a strong argument for international cooperation around the creation of good national systems of ECEC and improving gender equality by providing decent employment standards for carers, the

fair remuneration of care work, and good national public systems of care to underpin the growing labor-market participation of women – society's traditional carers.

See also: Early Childhood Care and Education: The Family, The Market, and The State; Investing in Early Childhood Education and Care: Some Policy Implications.

Bibliography

Pocock, B., Skinner, N., and Ichii, R. (2009). *Work, Life and Workplace Flexibility: The 2009 Australian Work and Life Index*, Centre for Work + Life. Adelaide: University of South Australia.

Further Reading

Crompton, R., Lewis, S., and Lynette, C. (eds.) (2007). *Women, Men, Work and Family in Europe*. New York: Palgrave Macmillan.

Ehrenreich, B. and Hochschild, A. R. (2002). *Global Woman: Nannies, Maids, and Sex Workers in the New Economy*. New York: Henry Holt and Company.

Hill, E., Pocock, B., and Elliot, A. (2007). *Kids Count: Better Early Childhood Education and Care in Australia*. Sydney: University of Sydney Press.

Hochschild, A. R. (2003). *The Commercialization of Intimate Life. Notes from Home and Work*. Berkeley, CA: University of California Press.

Hochschild, A. R. (2005). Love and gold. In Ricciuti, L. (ed.) *Women, Power and Justice: A Global Perspective*. Toronto: Zed/Innana Books.

Kittay, E. F., Jennings, B., and Wasunna, A. A. (2005). Dependency, difference and the global ethic of longterm care. *Journal of Political Philosophy* **13**(4), 443–469.

Lewis, J. and Giullair, S. (2005). The adult worker model family, gender equality and care: The search for new policy principles and the possibilities and problems of a capabilities approach. *Economy and Society* **34**(1), 76–104.

Lister, R., Williams, F., Anttonen, A., et al. (2007). *Gendering Citizenship in Western Europe: New Challenges for Citizenship Research in a Cross-National Context*. Bristol: The Polity Press.

McDowell, L. (2004). Work, workfare, work/life balance and an ethic of care. *Progress in Human Geography* **28**(2), 145–163.

OECD (2006). *Starting Strong II: Early Childhood Education and Care*. Paris: OECD.

OECD (2007). *Babies and Bosses – Reconciling Work and Family Life: A Synthesis of Findings for OECD Countries*. Paris: OECD.

Parrenas, R. S. (2001). *Servants of Globalisation. Women, Migration and Domestic Work*. Stanford, CA: Stanford University.

Pocock, B. (2003). *The Work/Care Collision*. Sydney: Federation Press.

Pocock, B. (2005). Work/care regimes: Institutions, culture and behaviour and the Australian case. *Gender, Work and Organisation* **12**(1), 32–49.

Pocock, B. (2006). *Jobs, Care and Justice: A Fair Work Regime for Australia*. Clare Burton Memorial Lecture 2006. <http://www.unisa.edu.au/hawkeinstitute/cwl/documents/Burton2006.pdf> (accessed October 2009).

Salazar Parrenas, R. (2001). *Servants of Globalization. Women, Migration and Domestic Work*. Stanford, CA: Stanford University Press.

Tronto, J. (1993). *Moral Boundaries: A Political Argument for an Ethic of Care*. London: Routledge.

EARLY CHILDHOOD EDUCATION AND CARE GOVERNANCE

Contents

Early Childhood Risk, Protection and Abuse Prevention
Evaluating Early Childhood Education and Care Programs

Early Childhood Risk, Protection and Abuse Prevention

E Khoo, Umeå University, Umeå, Sweden

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Early childhood is an important social, developmental, and educational stage in a child's life. Young children from preschool to early elementary school age (generally categorized from age 3 to 8 years) are most vulnerable to conditions of risk, abuse, and neglect, and are the age group most often referred to social services agencies because of concerns for their well-being. Conditions that put a child at risk of abuse or neglect also influence a child's ability to reach important life milestones.

For these reasons, organizations involved in providing services to children – from social service agencies to schools – share an interest in responding to concerns about need, risk, and abuse. However, the ways in which different organizations respond to concerns depend on the context in which services are provided – from the organization itself, to the community, society, and national contexts in which these are situated. Furthermore, the purposes of service provision depend on underlying assumptions behind definitions of risk, prevention, and abuse in these contexts.

Although there are thousands of peer-reviewed articles on the broad subject of child abuse, there has been relatively less conceptualizing about the challenges of reaching a shared understanding of the concepts of risk, protection, and abuse in childhood. Classifying and distinguishing between concepts in child welfare and child protection are hindered by the lack of agreement on definitions. It is not, however, simply a difference in opinions on definitions between disciplines such as education, psychology, and social work. Arrival at shared definitions is also hindered by the complexity of the political, social, and historical contexts in which definitions are created.

Given the amount of contact between children and educators, this is as much an educational concern and it is important, therefore, to understand approaches to

understanding key concepts connected with children at risk. This article offers a means to approach the complexity of early-childhood risk, protection, and abuse prevention by highlighting some of the key areas wherein concepts of need, risk, protection, and abuse can be elucidated.

Forms of Child Maltreatment

Child maltreatment is a social phenomenon embedded within the particular historical, political, and legal contexts in which it is identified. It appears in the milieu of class, gender, age, and race and is influenced by the key processes of modernization and individualization. The ways in which the problems of child abuse and neglect are defined by service providers and the wider society lead to particular ways of constructing responses to the problem and determining the grounds for state intervention in the protection of a child's well-being.

Although child maltreatment is a social construction, there is an increasing acceptance of the definition of child abuse or maltreatment as any act or failure to act – by a parent or other person in the context of a relationship of responsibility, trust, or power – which results in serious physical and/or emotional harm, sexual abuse, neglect or negligent treatment, or commercial exploitation resulting in actual or potential harm to the child's health, survival, development, or dignity.

Physical Abuse

Physical abuse is normally defined as a nonaccidental and substantial physical injury to a child, for example, striking, kicking, burning, or biting the child, or any action that results in a physical impairment of the child. It may also include acts or circumstances that threaten the child with harm or create a substantial risk of harm to the child's physical health or welfare.

Sexual Abuse/Exploitation

Some jurisdictions refer – in general terms – to sexual abuse, while others specify various acts, from invitation to sexual touching to incest, as sexual abuse. Sexual exploitation is an element of the definition of sexual abuse in most jurisdictions. Sexual exploitation includes allowing the child to engage in prostitution or in the production of child pornography.

Emotional Abuse

Emotional abuse has, more recently, entered into the scope of definitions of child abuse and neglect. Typically, it is defined as an injury to the psychological capacity or emotional stability of the child as evidenced by an observable or substantial change in behavior, emotional response, or cognition, or as evidenced by anxiety, depression, withdrawal, or aggressive behavior.

Neglect

Neglect is frequently defined in terms of deprivation of adequate food, clothing, shelter, medical care, or supervision. Some jurisdictions also include failure to educate the child as required by law in their definition of neglect.

Parental Substance Abuse

Parental substance abuse is an element of the definition of child abuse or neglect in some states. Parental behavior that exposes a child to harm is considered neglect, whereas it is considered abusive to intentionally or carelessly distribute alcohol or other illicit substances to a child.

As knowledge of child development and children's needs has increased, so too has our recognition of other kinds of abuses they may experience. These include: medical neglect, child prostitution and pornography, honor and race-related violence, bullying, Internet sexual exploitation, and institutional abuse.

Risk Factors in Early Childhood

Certain factors in a child's life are associated with the likelihood or risk of child maltreatment. These factors are often termed risk indicators. Researchers have noted that it is easier to identify indicators for physical abuse than sexual abuse and emotional abuse or neglect because several of the risk indicators for physical abuse are demographic variables. Risk factors apply to any variables (whether child, parental, or social) associated with an increased risk for child maltreatment. Because much of our information concerning these risk factors is based on retrospective accounts, these should be referred to as correlates rather than causal factors. Furthermore, these factors which place

Table 1 Risk factors

<i>Early-childhood risk factors</i>	
<i>Child</i>	<ul style="list-style-type: none"> • Excessive health or medical problems • Developmental delays • Diagnosed handicapping condition • Temperament
<i>Parent</i>	<ul style="list-style-type: none"> • Parents experienced poor parenting during their early childhood • Parent has attachment issues stemming from childhood • Perceptions of child • Inappropriate child expectations around development • Lower educational level, low school achievement • Authoritarian • Substance abuse • Chronic illness (physical, mental, or emotional), particularly depression • Single parenthood, Teen parenthood • Unemployment/low family income • Aggressive behavior • Incarceration
<i>Family</i>	<ul style="list-style-type: none"> • Coercive child-rearing practices • Family density (number of members) • Few positive interactions • Family lacks leadership, closeness, and negotiating skills • Language deficiency: family from immigrant background
<i>Environment</i>	<ul style="list-style-type: none"> • Lack of caring individual in child's life • Neighborhood lacks both formal and informal social supports • Living in a violent community • Lack of connection with community
<i>Macro system</i>	<ul style="list-style-type: none"> • Attitudes toward how mothers should behave as parents • Cultural values that support violence • Culturally promoted attitudes and behaviors about parental rights to physically punish

children at risk of maltreatment are also connected with children becoming educationally disadvantaged and, therefore, in need of special assistance. **Table 1** presents an ecological perspective of risk factors related to the child, parents, family, community, and wider social contexts.

Approaches to Assessing Risk and Need

When an allegation or suspicion of child abuse or neglect is reported to statutory authorities, social workers or other professionals gather information during this initial contact to determine if an investigation is warranted and, if so, whether the child needs further protection during the process. Authorities use a number of screening procedures that establish thresholds for the level and immediacy of the response – ranging from an immediate investigation to follow-up or a referral to community supports. Some of

these screening approaches are integrated with standardized and structured risk-assessment instruments.

Structured risk assessment

In the face of growing demands for services, jurisdictions around the world are increasingly turning to structured risk assessment to assist child-protection authorities in rationing their services. This movement toward structured risk assessment is representative of trends toward evidence-based practice (emphasizing methods built on a solid foundation of scientific research) and demands improved practice procedures.

Risk-assessment models can be divided into two kinds: actuarial and consensus-based, with some models combining elements of both. Actuarial models are based on the empirical study of child-protection cases and their future maltreatment outcomes. The object of actuarial research is to identify factors known to be statistically predictive of future maltreatment and to use this information in the construction of an instrument that can be scored in a mechanical fashion. In consensus-based risk assessment, workers rate selected characteristics originally identified by consensus among experts and, subsequently, these factors are processed using professional judgment instead of a standard algorithm. Whether actuarial or consensus-based, an often-cited benefit of structured risk assessment is that it leads to greater consistency of evaluation and reliability of response among child-protection workers.

Consensus approaches dominate in child-protection systems. The instruments in these approaches require assessors to determine if there are grounds for service and whether the child is in immediate danger, as well as assess a wide range of risk indicators – first to determine which indicators are present and then to determine the level or severity of potential risk.

Critics have argued that – despite their increasing use – risk-assessment approaches are based on instruments that have questionable reliability, have never been validated, and for which no empirical evidence exists. For example, a risk-assessment instrument cannot determine whether to remove a child from a parent's care or how much intervention to provide. Yet, sometimes, professionals use risk assessments to guide their judgments about such decisions (Figure 1).

Ecological approaches to needs assessment

Standardized risk assessment is not always the best way to safeguard the well-being of children. Recognizing this, many countries and jurisdictions – from England, Canada, and Australia to Denmark and Sweden – have begun to use holistic approaches rooted in the ecology of child development, resilience, and attachment theories. These approaches and child-welfare research inform us that significant life events, as well as quality of parenting, affect individual development most critically in childhood. Moreover, these approaches are child-centered, concerned with equality, and equip social workers with tools to more systematically document and evaluate children's circumstances. They are further promoted as quality-assurance approaches wherein child-welfare investigation, planning, and follow-up can be improved while concomitantly strengthening children's rights.

Ecological approaches contain records for use by social workers wherein gathered information is from the first contact to closure and recorded in a structured manner. The overall ecological approach is illustrated in an assessment-framework triangle (Figure 2) where the three different domains represent: (1) the child's developmental needs, (2) the parent's capacity to meet the child's needs, and (3) factors in the family and the environment which may

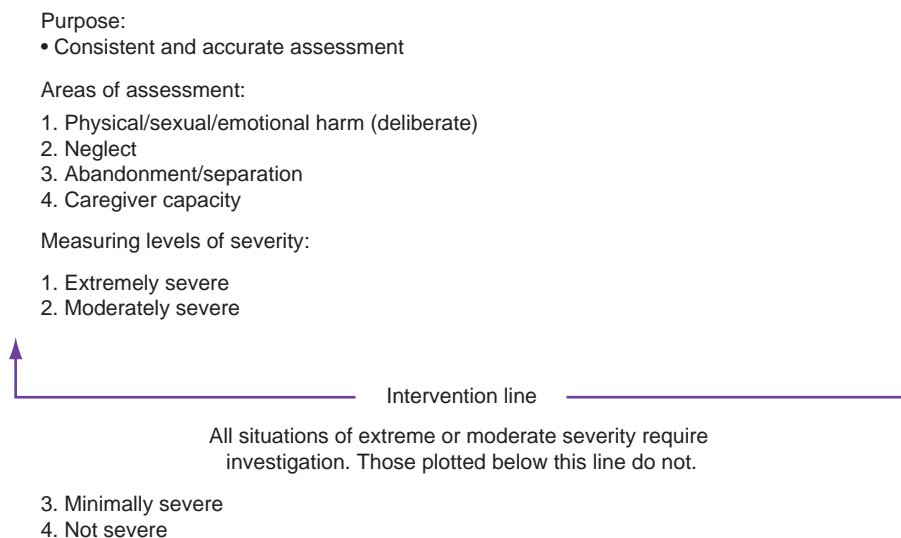


Figure 1 Risk-assessment approaches.

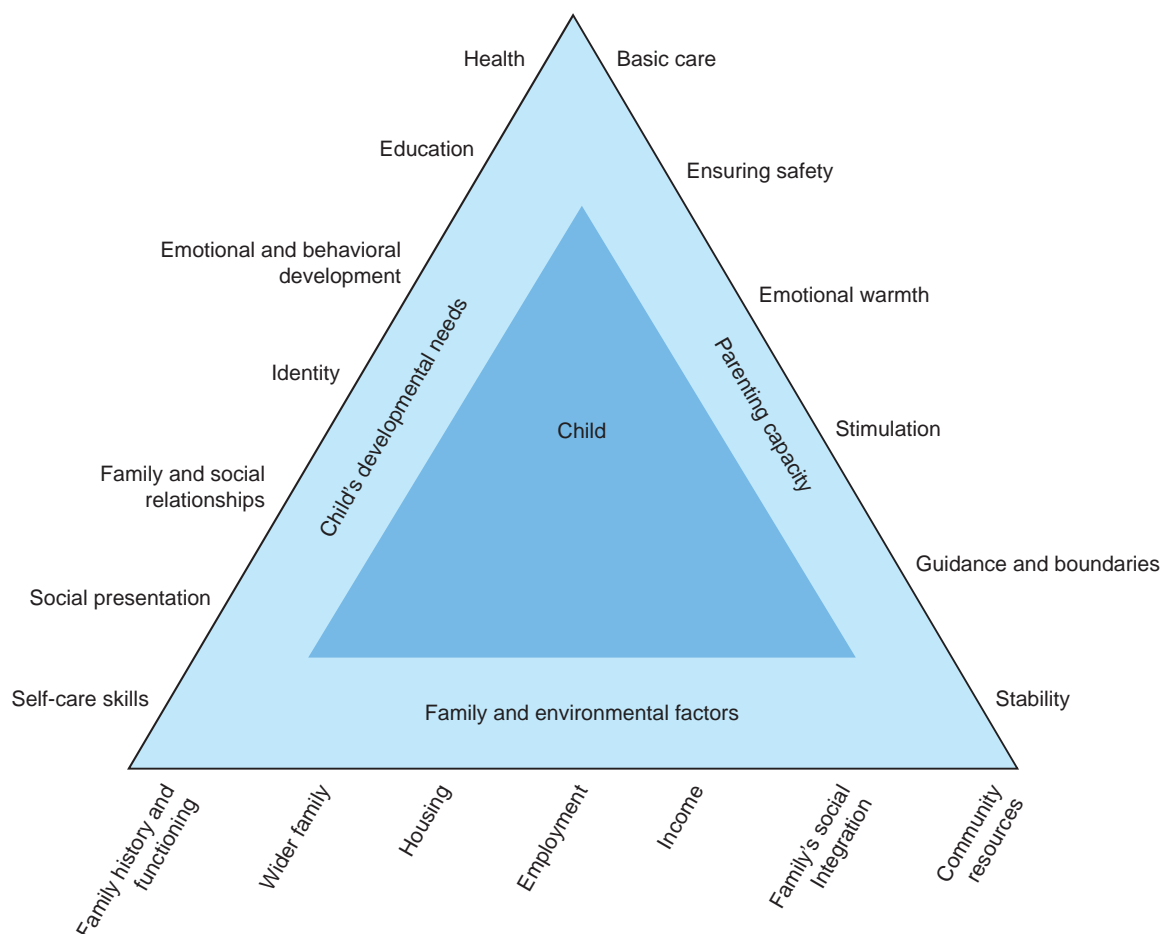


Figure 2 Needs-assessment triangle.

influence parenting capacity and hence also the overall situation of the child. These approaches also contain special procedures and protocols to facilitate work with other professionals – in particular, with teachers.

Effects of Child Abuse and Neglect

Accumulated research evidence indicates that children who are maltreated, often experience disrupted growth and development. Adverse effects have been identified in maltreated children's physical, cognitive, emotional, and social development, and these adverse effects accumulate over time.

The psychological, emotional, or physical damage that a child suffers as a result of maltreatment depends on aspects of the abuse itself and on the child's stage of development. In addition, individual children respond differently to the same kinds of abuse. Some children do not appear to show significant effects from maltreatment. These resilient children may have personality characteristics (high cognitive ability, optimism, high self-esteem, hopefulness, etc.) that buffer against harm. Damaging

effects may be mitigated if the abuse occurs only once, or if a supportive adult is available who lets the child feel he or she is believed and will be protected.

In other cases, the consequences can be longlasting. For example, physically abused children tend to have greater difficulty in relationships with peers, show less ability to empathize, and are more aggressive toward peers and adults. Studies of neglected toddlers show that they, often, have an impaired ability to trust others which may, in turn, lead to feelings of being unloved and unwanted. They may be less able to develop the social skills required to form healthy relationships with other children and adults. When a child cannot master developmental tasks (such as learning to trust) at the appropriate age, it becomes more difficult to accomplish later developmental tasks throughout the life span.

As they get older, abused and neglected children are more likely to perform poorly in school. The younger the child is at the onset of maltreatment, the more important it is to accurately assess and ameliorate the effects of the experience so that the child can recover and go on to master other life tasks successfully.

There are two main limits to these general research findings. First, most research on maltreated children comes through clinical studies of young children who have been referred for treatment and who exhibit the most serious behavioral problems. Additionally, most research is of children involved with public child welfare agencies and who belong to families of lower socioeconomic status and minority populations. Therefore, these findings do not represent the entire – and often hidden – population of abused and neglected children.

Responses to Child Maltreatment

Child Welfare and Child Protection

Child protection is a term used to describe philosophies, policies, standards, guidelines, and procedures to protect children both from intentional and unintentional harm. The definition reflects the duty of organizations – and the people in them – toward children in their care.

Child protection emphasizes the legal grounds upon which social workers and police can intervene to identify children who have been harmed or are likely to be harmed, and to then intervene to protect children from future harm. Child protection is often used to delimit specific services for children in need of protection from abuse and neglect. It is a term more often used in Anglo-Saxon welfare states. Child protection can be characterized as social work practice – guided by large amounts of official written material and lengthy procedural guidance used locally – but tightly coordinated and controlled by the central authority. Child-protection systems are criticized for their lack of social solidarity in comparison to child-welfare systems.

Child welfare is a term most often used in a general sense to encompass the broad scope of involvement by the state and its authorized professionals in assisting children and their families – both in cases of abuse and neglect and when families and children are found to be in need. In this sense, child welfare is concerned with protecting children through general prevention services including health, education, recreation, family support, and treatment services.

Depending on the context, safeguarding children and child and family services have replaced the terms child welfare and child protection or are used interchangeably with them. A child in need of protection may also be a child with welfare needs that could be successfully addressed through the provision of preventive services.

Early-childhood centers and schools

Child protection in early-childhood educational settings involves both responding to concerns about abuse and neglect and also proactively helping children to develop

personal skills and values with which to avoid or resist dangerous situations and abuse.

Responsibility for Child Welfare and Child Protection

The state

Legislative responsibility for child and family services rests with the different levels of government from provinces/states in federal systems to national government. Each may have its own specific legislation that provides protection to abused and neglected children. Policy describes the roles and responsibilities of statutory authorities, and establishes the framework for the organizational structures, tools, and resources used by authorities in carrying out their responsibilities.

When considering prevention and screening measures, effective governance may reduce the number of children at risk. Governance also provides policies and guidelines that facilitate the deployment of people and resources to effectively respond to the needs of children at risk. Good governance can also help to ensure the safety of children by enforcing best practices in the organization and enactment of child-welfare and child-protection policy, legislation, and service delivery.

Although the goals and philosophies of child welfare and child-protection services vary, they reflect the principle that families have the primary responsibility for the care of children. These approaches also share the general position that children should have certain rights to be protected from abuse and neglect and that governments have a role in protecting children from harm.

Different jurisdictions, generally, share some basic belief that the best interests of the child should be a primary consideration and that the least intrusive form of intervention is best. When intervention does take place, the range of services available can vary enormously but can be broadly categorized as voluntary or compulsory services. Services begin with the investigation of concerns about need or allegations/suspicions of abuse or neglect using a variety of approaches established according to the legislation, policy, and protocols of the jurisdictions in which they are found. Where appropriate and necessary, the state or its delegates provide services ranging from counseling and support to removal of a child from parental care on temporary or permanent bases. Further, there is, generally, oversight of child welfare and child-protective services, ensuring that standards and requirements of treatment and care services are met.

Interagency cooperation

Many jurisdictions have begun to formalize committees responsible for collaborating with the care and protective needs of vulnerable children. These committees are often

comprised of social workers, healthcare providers, and educators.

Increased interagency cooperation has led to the development of protocols or agreements between collaborators which define and give direction on concerns about children. Protocols may address the reporting of concerns about a child (ranging to concerns about children in need to more serious concerns about harm caused to a child). Protocols may operate internally within multidisciplinary or multifunctional agencies, locally to coordinate the activities of different community actors, or on a wider scale, involving different authorities or government departments. The purpose of protocols is to clearly define the responsibilities of each collaborator, to minimize harm to children, coordinate responses, and also to inform the public of different mandates. Protocols help organize the activities of professionals in various capacities including day care centers, schools, hospitals, policy, and child-welfare agencies.

Mandatory reporting

Mandatory reporting refers to a legal obligation to report certain kinds of concern to child-welfare authorities. The intent of mandatory reporting is to encourage the good faith (without malice and with reasonable cause) reporting of concerns for the safety and needs of children. Sometimes, reporting standards apply to professionals and, at other times, to both professionals and the public at large. In some places, only social workers, healthcare providers, and the police are legally required to report that a child is in need of care and protection. In other places, local guidelines and professional codes of conduct (e.g., teachers) indicate the duty to report. However, individuals are not bound to report as a matter of law.

Recognizing and reporting suspicions of abuse then becomes an ethical obligation that requires early-childhood educators and teachers to have the appropriate training, to be familiar with symptoms of abuse/neglect, and to be able to follow legal requirements and community procedures to protect children.

Educators are the group of professionals that report most cases of concern to child-protection authorities, although research shows that they only report about 25% of suspected concerns. Reasons for not reporting may include a lack of knowledge about how to do so, fear of making an inaccurate report, or a belief that child-protective services do not help families and children. Mandatory reporting systems have faced criticism for being overburdened and ineffective in responding to the concerns of teachers and other professionals.

Abuse Prevention

Abuse prevention can be considered on a continuum from primary prevention to tertiary intervention. Primary

prevention is targeted at the stress factors shown to have the capacity to lead to abuse. Poverty, unemployment, single-parenthood, lack of social networks, and lack of access to healthcare can all contribute to the conditions in which abuse can develop. Therefore, social policy, healthcare policy, and family policies that address these strategies can be viewed as forms of primary prevention. In education, primary prevention may include general public education campaigns, educational information to new parents, and education programs for children (e.g., Safe Touch programs to prevent sexual abuse). Early education, open/free preschools, and community-based services can enhance family stability and reduce the stresses and strains in the family circle.

Secondary prevention is targeted at high-risk families or the risk factors that have a cumulative risk characteristic. In this instance, we can think of programs for teenage mothers, access to birth control, home visitors, parent relief, day care, or preschool as secondary prevention measures. These measures enhance the chances for positive development in children and are not associated with the stigmatizing and labeling features often associated with child-welfare services.

Tertiary prevention is action taken to help families after abuse happens. Child-welfare agencies provide case management, family support, and child-in-care services while court services handle legal issues. Tertiary prevention may also include therapeutic intervention – sexual abuse-treatment groups, nonoffender treatment groups, groups for violent men, family therapy, and intensive family-preservation interventions.

Primary and secondary prevention have shown the most promise, according to research, of having an impact on child maltreatment. Tertiary prevention has not had a positive impact on re-abuse rates and there is little evidence to suggest that tertiary prevention is effective in child neglect. An ongoing problem is that few studies have systematically evaluated these interventions in a methodologically rigorous way. There is a lack of experimental designs, small sample sizes, high attrition rates, and inadequate outcome indicators.

Educational Responses

Early-childhood professionals are particularly important in risk, protection, and abuse prevention because many abuse victims have, at some point, attended early-childhood-education centers and almost all, eventually, are in schools. Most often, children are victimized by family members, although child maltreatment occurs much less frequently in the contexts of schools, childcare, and organized after-school activities.

Outside of the family, educational professionals have the most contact with young children. Young children are the most vulnerable to all forms of abuse – given that

they are easily manipulated and powerless, being wholly dependent on adults for their care and safety. In their positions and with their knowledge of child development, early-childhood professionals are in a position to be able to observe children, detect behavioral changes, and respond to their needs. Thus, early-childhood educators and teachers play important roles in the prevention, identification, and reporting of abuse, as well as in providing support to child victims of abuse and neglect.

Primary preventive programs include ensuring safe and high environmental standards within schools and childcare settings, developing appropriate standards to reduce the potential of abuse within these settings, and establishing policies of inclusion of parents in developing programs. Educators can also help protect children from abuse and neglect by advocating for children and their families for access to well-designed, sufficiently funded, and effectively implemented programs, supports, and regulations.

Secondary preventive approaches can support and build family strengths and understandings of child development. When concerned about the potential of risk to a child, communication with parents and families becomes critical. Therefore, the aim of educators is to be supportive to families while providing information, referrals, and follow-up. Collaborating with other professionals, including child-welfare professionals, promotes a shared understanding of child development and supports and empowers families.

Educators can also play an important role in abuse prevention by introducing an explicit and developmentally appropriate child-protection program. Young children can be taught personal safety skills and will practice them if activity methods are used with reinforcement. They will also be more confident, know their rights, and understand what constitutes abusive behavior. Children with the best safety knowledge and skills are those who have undertaken a comprehensive child-protection program with parental reinforcement. Parents who are told what abuse prevention is and why it is being taught are less likely to undermine the program by promoting secrecy, not listening to children, or defending those who touch children inappropriately.

See also: Community Focused Schools; Participation in Early Childhood Education and Care Programs: Equity, Diversity and Educational Disadvantage; Risk and Resilience Frameworks in Understanding Special Education.

Further Reading

- Ainsworth, F. and Hansen, P. (2006). Five tumultuous years in Australian child protection: Little progress. *Child and Family Social Work* 11(1), 33–41.
- Barber, J., Trocmé, N., Goodman, D., et al. (2007). *The Reliability and Predictive Validity of Consensus-Based Risk Assessment*, 34p. Toronto, ON: Centre of Excellence for Child Welfare.

- Bronfenbrenner, U. (2005). *Making Human Beings Human: Bioecological Perspectives on Human Development*. London: Sage.
- Clark, M. M and Waller, T. (eds.) (2007). *Early Childhood Education and Care. Policy and Practice*. London: Sage.
- Cleaver, H. and Walker, S. (2004). *Assessing Children's Needs and Circumstances. The Impact of the Assessment Framework*. London: Jessica Kingsley.
- Corby, B. (2000). *Child Abuse: Towards a knowledge Base*. London: Open University Press.
- DePanfilis, D. and Scannapieco, M. (1994). Assessing the safety of children at risk of maltreatment: Decision-making models. *Child Welfare* 73(3), 229–245.
- Department of Health (2000). *Framework for the Assessment of Children in Need and Their Families. Guidance Notes and Glossary for: Referral and Initial Information Record, Initial Assessment Record and Core Assessment Record*. London: The Stationary Office.
- Drury-Hudson, J. (1999). Decision making in child protection: The use of theoretical, empirical, and procedural knowledge by novices and experts and implications for fieldwork placement. *British Journal of Social Work* 29(1), 147–169.
- Egelund, T. (1996). Bureaucracy or professionalism? The work tools of child protection services. *Scandinavian Journal of Social Welfare* 5(3), 165–174.
- Ferguson, H. (2001). Promoting child protection, welfare and healing: The case for developing best practice. *Child and Family Social Work* 6(1), 1–12.
- Flynn, R. J., Dudding, P. M., and Barber, J. G. (eds.) (2006). *Promoting Resilience in Child Welfare*. Ottawa: University of Ottawa Press.
- F/P/T Working Group (2002). Child welfare in Canada – 2000. The role of provincial and territorial authorities in the provision of child protection services. <http://www.hrsdc.gc.ca/eng/cs/sp/sdc/socpol/publications/reports/2000-000033/page01.shtml>.
- Government of Ontario (1997). *Risk Assessment Model for Child Protection in Ontario*. Ontario: Ministry of Community and Social Services.
- Hill, M., Stafford, A., and Green Lister, P. (eds.) (2002). *International Perspectives on Child Protection*. Edinburgh, Scotland: Scottish Executive.
- Houston, S. (2001). Transcending the fissure in risk theory: Critical realism and child welfare. *Child and Family Social Work* 6(2), 219–228.
- Jack, G. (1997). Discourses of child protection and child welfare. *British Journal of Social Work* 27(5), 659–678.
- Leventhal, J. M. (2001). The prevention of child abuse and neglect: Successfully out of the blocks. *Child Abuse and Neglect* 25, 431–439.
- OACAS (2006). *Ontario Child Welfare Eligibility Spectrum* (revised). Toronto: Ontario Association of Children's Aid Societies.
- OECD (Organization for Economic Co-operation and Development) (2001). *Starting Strong. Early Childhood Education and Care*. Paris: OECD.
- Otway, O. (1996). Social work with children and families. From child welfare to child protection. In Parton, N. (ed.) *Social Theory, Social Change and Social Work*, pp 154–171. London: Routledge.
- Parton, N. (1998). Risk, advanced liberalism and child welfare: The need to rediscover uncertainty and ambiguity. *British Journal of Social Work* 28(1), 5–27.
- Pringle, K. and Harder, M. (eds.) (1997). *Protecting Children in Europe*. Aalborg: Aalborg University Press.
- Scottish Executive (2002). It's everyone's job to make sure I'm alright. *Report of the Child Protection Audit and Review*.
- Swift, K. (1995). Missing persons: Women in child welfare. *Child Welfare* 74(3), 486–505.
- Thomlison, B. (2003). Characteristics of evidence-based child maltreatment interventions. *Child Welfare* LXXXII(5), 541–569.

Relevant Websites

- <http://actagainstviolence.apa.org> – Adults and Children Together Against Violence.
- <http://www.childcarecanada.org> – Childcare Resource and Research Unit.

<http://www.childwelfare.gov> – Child Welfare Information Gateway (USA).
<http://www.earlychildhoodaustralia.org.au> – Early Childhood Australia.
<http://www.everychildmatters.gov.uk> – Every Child Matters (UK).
<http://www.teachers.tv> – Early Years-How Do They Do It In Sweden?
<http://www.naeyc.org> – National Association for the Education of Young Children.

<http://www.aifs.gov.au> – National Child Protection Clearinghouse (Australia).
<http://www.dh.gov.uk> – Publications policy and guidance: Department of Health.
<http://www.nspcc.org.uk/> –The National Society for the Prevention of Cruelty to Children (NSPCC).

Evaluating Early Childhood Education and Care Programs

M Janus, McMaster University, Hamilton, ON, Canada

S Brinkman, University of Adelaide, Adelaide, SA, Australia

© 2010 Elsevier Ltd. All rights reserved.

Early childhood education and care has been gaining recognition as a very important factor in influencing lifelong human development and outcomes. Sheer numbers of preschools, child-care centers, playschools, and early childhood programs have been steadily growing, both in the public and for-profit sectors. Many claim that their curriculum or principles are based on the available evidence of what works best for young children. Nevertheless, formal evaluations of new or existing programs are completed less frequently than would be expected in an area that deals with such a crucial and vulnerable period in child development.

There are many factors that have led to the current proliferation of early childhood services. Since the inception of Head Start in the United Kingdom and Early Head Start in the United States similar programs have become popular, and the model and ideas have been adopted in socially conscious programs and policies. Increased workforce participation by women and the degeneration of traditional family networks, contributed to the demand for such early childhood services. Parents are seeking preschool environments for their children more frequently now than they did 30 years ago, whether for child care, or as a supplement to their children's activities. Much of this growth has been closely followed by the increase in evidence on the importance of high-quality early childhood education and care (ECEC).

In the first section of this article we briefly describe the types of evaluation and their relevance to the ECEC; in the second, we present current challenges and concerns in evaluating ECEC; and in the third, we describe the international perspectives on ECEC.

Types of Evaluation and Their Relevance to ECEC

Any educational structure or service can be evaluated based on general principles. The ECEC systems, however, are unique for at least three reasons. First, unlike school-based education, they are rarely governed by uniform legislative bodies, even though they are often centrally licensed or accredited. Second, there is no discernible consensus among professionals as to the most optimal assessment. Third, the types of early education and care vary dramatically among providers. These reasons make it

especially challenging to summarize the current state of knowledge on this topic.

The categories of evaluation are broadly based on: What are the goals of assessment (e.g., monitoring of the outcomes or implementation)? Who is doing the evaluation (e.g., government, external consultants, etc.)? Who is providing the data (parents, providers, managers)? Also, what are its intended uses (summative, formative, for accountability reasons)? These categories are not mutually exclusive. Alternatively, evaluation studies could be categorized based on methodologies used.

Target of the Evaluation

Impact evaluation

The majority of evidence regarding the evaluation of ECEC comes from outcome-oriented approaches. In this category, the evaluation is being undertaken to assess whether the particular program or system improves outcomes for the participants. Outcomes need to be defined as a type of skill or developmental area so that the appropriate methodology can be applied. This type of evaluation is known as an impact evaluation, and thus whether the program is fulfilling the intended goals is the target of evaluation. The methodology used is most often quantitative.

Quality evaluation

The goal of quality-oriented approaches is to monitor how the programs are delivered, thus, they are focused on how the outcomes may be achieved rather than on what they actually are. This type of evaluation tends to consider the quality of classroom environments, staff training, parent satisfaction, and uses either only qualitative, or a combination of both qualitative and quantitative methodology. Recently, there have been an increasing number of voices advocating for the inclusion of children's perspectives on aspects of the programs they attend, not in the more established format of test results, but in terms of their narrative impressions.

Process evaluation

A process evaluation aims to evaluate the implementation aspects of a program, and thus essentially it is also quality oriented. It is claimed, however, that this particular type is best suited to a program in its initial stages, and in fact,

that its goal is to establish whether the program is ready for the impact evaluation.

Confirmatory program evaluation

A variation on impact evaluation is a confirmatory program evaluation. According to its proponents, this is a systematic, theory-driven investigation on how program participation leads to the outcomes. Since its clear goal is to confirm that the program works in the way its philosophy predicts, the hypotheses are narrower than in the pure impact evaluation, and they do not include the evaluation of implementation. Nevertheless, the more complex methodology suggested for this more strictly causal evaluation approach (regression, path modeling) allows to create clear criteria for interpreting findings.

Evaluators

Evaluations are often legislated into state-funded ECEC programs. Such evaluations may be contracted to independent consultants, or sometimes, government systems establish their own evaluation units or departments whose entire mandate is to carry out and report on the evaluations of the funded services. In a number of jurisdictions in the United States, Canada, Australia, and the United Kingdom, the consultants tend to be affiliated with education departments at local universities, thus providing a solid research base and some continuity of effort.

The identity of the evaluating body can also be dependent on the goals. Short-term evaluations may be run by government agencies for accountability reasons, while a programmatic, long-term evaluation may be carried out by an independent agency.

Data Sources

Clearly, the data sources for any evaluation are determined by its type and goal. Nevertheless, either one or many types of informants can be engaged in the process of providing the data: children themselves (through tests, direct assessments, observations, or interviews); teachers/providers (through their standardized assessments of children, observational assessment such as the Early Development Instrument (EDI), program delivery, environments, and own experiences); parents (through their assessment of children, teachers, environments, and satisfaction); or managers (through their assessment of the program). A good system evaluation would include at least two or more informants, in order to ensure reliability of the collected data. Of particular importance is the fact that the ultimate target population – the young children who participate in the ECEC programs – are on a steep developmental curve, and outcomes of any assessment can rapidly change with age.

Intended Uses of the Evaluation

Three intended uses are most frequent in evaluating ECEC programs: formative, summative, and programmatic evaluation.

Formative evaluation

Formative evaluation is usually undertaken early in the development of the program to inform the providers and stakeholders about the trends in results, whether the goals of the program are likely to be fulfilled, and to identify the barriers and facilitators of implementation. Results of the formative evaluation are then incorporated into the program with the necessary adjustments made to improve program implementation. These evaluations are usually less formal, and more likely to be internal, than the summative evaluations, and while they are often mentioned in descriptions of new programs, the evidence is rarely published.

Summative evaluation

Summative evaluations are intended to provide a package of results used to assess whether a program works or not. These types of evaluations are dominating the field when it comes to evaluating ECEC programs. While the timing of a summative evaluation has to allow the program to have a reasonable chance to achieve its goals, it is often carried out for the evaluation of short-term goals. In general, summative evaluations provide quantitative data and are focused on outcomes. However, alongside the developmental, behavioral, or cognitive outcomes for children these evaluations often also include program statistics, for example, attendance, staff characteristics, funding, and cost-effectiveness data. Summative evaluation can form part of an impact evaluation, or be carried out in conjunction with a qualitative or process evaluation to provide complementary evidence.

Programmatic evaluation

Programmatic evaluation is meant to provide a broad, inclusive, yet detailed and definitive assessment of a program in relation to both whether and how it fulfills its goals. These evaluations tend to involve long-term aspects of programs, and by the very nature of their purpose are fairly costly. By definition, they should also consider how well the program's implementation is following the program's philosophy as set out in the design of the program. Programmatic evaluation is most useful in evaluating ECEC programs that have considerable longevity (and thus sustainable funding sources) as well as long-term goals. Perhaps due to these reasons, government-funded initiatives are most likely to be the subject of such evaluations.

Methodologies

The goals of an evaluation mandate the basic methodology. Generally, impact evaluation requires quantitative data collection, while a quality or process evaluation cannot be undertaken without qualitative strategies. The interpretability of the evaluation outcomes is inherently linked to the qualitative data and research designs utilized.

Quantitative Methodologies

The basic evaluative tool for assessment of outcomes is a test of children's abilities. Such a tool could be administered directly, through observations, or based on teacher report. What makes a test score a component of evaluation is its comparison with another test score. The change in outcome for a child who attended the program could be compared with the change in outcome of a child who did not attend this program (comparison group). The children's scores are taken prior to attendance and then compared over time. They can also be compared to a standardized score for a child meeting specific characteristic (usually age and gender).

Comparison group

By far the most scientifically grounded methodology is to include a well-selected comparison group. Randomized control trials (RCTs) prescribe a longitudinal involvement with the sample, and a random assignment to the experimental (program) or control (no program) group. While excellent for statistical reasons, this method rarely works with ECEC due to logistics, but most importantly, ethics – some parents would have to be denied the opportunity to participate in the program. Evaluators of ECEC often seek to modify an RCT approach to make it acceptable, feasible, and yet scientifically sound. One approach taken is to include families on the waiting list. The positive aspect of this selection is that these are families who are motivated to participate in the program and they are easily accessible, so their potential participation in other programs while waiting could be monitored (as this might influence the results). An approach that is gaining an increasing popularity is the regression discontinuity (RD) design for evaluating causal effects of interventions. The RD design is an experimental pre-test/post-test comparison group strategy. In RD designs, participants are assigned to program or comparison groups based on a score taken before the program participation. Thus, program can be targeted toward those who most need it and avoid the ethical RCT issues. Another approach is to select families with similar characteristics (e.g., same neighborhood and similar risk factors), whose children do not attend the program. While also acceptable, this choice is less optimal due to the possible levels of parental

motivation, lower interest in the program, as well as decreased access. An added difficulty is the need to specify whether the group of nonattendees is composed of children who do not attend any program, or just do not attend the program being evaluated. Selecting the best control or reference group is challenging, and there are documented cases where over the course of evaluation the composition of such groups changed so much as to render the evaluation outcomes uninterpretable. The programs with the most longitudinal high-quality evaluation follow-up evidence – Perry Preschool, Carolina Abecedarian, Infant Health and Development Program, and Chicago Child-Parent Centers – all have employed the control group design with well-matched ethnic and socioeconomic characteristics.

Pre-test/post-test methodology

For best results, the pre-test/post-test methodology also requires an appropriate comparison group. Considering the age of children, the maturational changes are very likely to contribute to the increase in children's test scores from pre- to post-test. Such change can be statistically controlled for, if the test is standardized and well documented, however, the evaluators would need to have sound evidence that the groups on which the test was standardized were similar to those tested (and thus the maturational growth over time comparable between them). Inclusion of a comparison group has to be guided by all the same considerations as in the previous paragraph. A well-chosen comparison group can make the outcomes of a pre-/post-test methodology a solid base for program improvement and policy recommendations.

Comparison to standards

Some early childhood evaluations follow the practices used in higher-level educational systems by using established norms (standards) as a comparison. While it appears to be an acceptable compromise, the cautions are numerous. Apart from those already mentioned above, such as the suitability of the test or maturational influences, the possibility of bias has to be taken into account: first, in terms of the group tested, and second, in terms of opening the possibility of teachers developing the teaching to the test habits.

Group randomization

Recently, the “randomize groups not individuals” was proposed as an overall program strategy in assessing ECEC (St. Pierre and Rossi, 2006). Its authors claim that it provides the most effective methodology to answer the question “what works better?” rather than the more common “whether” a program works. They reject the feasibility of a no treatment group with respect to ECEC as unrealistic. In this context, the group randomization offers

a much more effective alternative, even if more costly. Since groups (schools, neighborhoods) would be randomized to participate in different programs, there are no ethical issues; the necessary larger sample, though more costly, offers less attrition and larger variation than randomization of individual children and families. The current availability of sophisticated statistical methodologies like multilevel modeling, makes it possible to account for the impact of within-group, and between-groups factors. The major attraction of this methodology, its breadth, is also its drawback, as it requires extensive and long-term funding investment. The proponents of this strategy suggest that the US Head Start has infrastructure, resources, and the will to engage in such an evaluative effort.

Qualitative Components

Many evaluations use qualitative methodology as a component complementary to the quantitative data. The results of parent satisfaction surveys, open-ended interviews with parents and service providers, are often found in evaluation reports, however, not reported as frequently and as fully as the results of quantitative impact evaluations. This could ultimately lead to the low level of sharing knowledge on the nonquantifiable aspects of the ECEC, and potentially impede progress. In fact, there is some evidence that early childhood practitioners are concerned that quantitatively oriented evaluations are incomplete and inadequate to inform practice. The current recommendations in the literature require the qualitative input from child, family, service providers, and the community to an integrated component of any evaluation.

Challenges

Evaluating ECEC programs is not without its challenges. A paper by a group of early childhood educators and professionals from diverse international backgrounds (Jalongo *et al.*, 2004) suggests seven dimensions for a global vision of ECEC. The seventh dimension is rigorous evaluation, which should be comprehensive, ongoing, longitudinal, and broad-minded enough to consider the value-added characteristics. The value-added component contributes program-wide and individual family perspectives for both expected and unexpected outcomes. A program-wide perspective would include impact on children's health even if that was not the focus of the program, or the fact that by providing high-quality, affordable child care, the mothers were encouraged to seek more education. Individual family perspectives need to be added through interviews and case stories, and can be used for advocacy and recommendations.

Contextual Information

The necessity to include contextual information is perhaps most strongly emphasized within the international context (see section titled 'International perspective' below), however, it could be also crucial for comparisons of programs within a country or state, or over a number of years. For example, changes in the parental leave policies can have major impact on the demand for early childhood services; fluctuation in economy can affect the numbers of families qualifying for state assistance in child-care funding; social awareness about the impact of early childhood could influence the quality of child care; levels of immigration could be crucial in determining the language needs of families. Even within scientifically sound evaluation designs, context is often considered solely at the individual level (such as child characteristics) and in some circumstances the family level. The ability to control individual- and family-level contexts are basic requirements in rendering the outcomes of any evaluation usable. Nevertheless, the population and community context is often just as crucial.

Policy Ramifications

A theme reoccurring in many recent publications in the field is that very few countries have an established policy framework with stated goals and philosophy, which would provide the basis for evaluation. Without such a framework, it is more difficult to ensure that results of any evaluation are being translated into improvements in early childhood education.

While the majority of state-funded programs have formal evaluations mandated in legislation, in reality, few of those are well-conducted or well-documented impact or quality evaluations. Evaluation as a process is often misunderstood by legislators and providers, and interpreted as the need to carry out financial reviews, informal surveys of parent satisfaction, or assessment of participation rates. In the United States, of the 33 states identified to have had a state-funded preschool program, 22 reported having completed a formal evaluation, of which only 13 had data usable for a meta-analysis (39%). According to another statistic from the United States, less than 1% of total expenditures for social programs are spent on research and evaluation.

Applied Research Perspective

The goals of an evaluation can vary, and therefore the process of evaluating can vary according to those goals. Nevertheless, evaluation can be considered an exercise in applied research, and has to lead to answers to fairly routine questions. These could be "what works?", "what works better?" (impact), or "how well do we make it work?" (quality), but the evaluation process and underlying research

should result in answers that can be used for program improvement. These can range from “it does work in this area” to “we do not know whether it works, because . . .” but the answers have to be pragmatic enough to allow a follow-up: with recommendations for program improvement, or with recommendations for another, better evaluation. There is a concern that evaluation practices in the early childhood education field are underdeveloped in comparison to evaluation in other educational areas. Inaccuracies in understanding the process of evaluation while designing the ECEC programs can have long-reaching detrimental consequences not only in terms of impact on child development but also in misuse or misallocation of funds available for early childhood. The lack of solid evidence that evaluation results are routinely incorporated into action plans seems to highlight this issue.

Accountability Controversy

Increased emphasis on accountability in education as a whole, which also appears to include the education of younger children, raises the concern in some professionals that it not only impacts the quality of the education, but also shortchanges the evaluation process. Too strict and prescribed achievement standards, especially for the younger children, can easily lead to teaching to those standards, with the sole purpose being, to enable children to pass the test. This concern has a lot of validity in early childhood education for a number of reasons: the development of young children can be variable, and the domains of development are interconnected; thus, assigning a value-laden consequence to results of any one test (or achieving any one or two cognitive skills) is misguided. Moreover, the role of the teacher is especially important in the early years, and there are fears that the focus on accountability leads to a pressure to teach more and sooner, within a narrowing, standardized range, thus de-professionalizing the teachers making their profession akin to those of technicians monitoring discrete bits of knowledge or behavior rather than enabling them to use their thinking, judgment, and insight to tailor programs to each child's learning needs. For the children themselves, this approach seems to value performance over learning. It appears that in the early childhood education field more than in the education of older children, there is a strong need to develop a consensus in understanding the differences between accountability and evaluation. This should be communicated clearly through the governance of early education and care.

Perspectives on Program Quality

While there is an improvement in the level of understanding that test scores do not equal program quality, definitions

of quality still raise concerns as to the appropriateness of evaluative efforts. Literature suggests that there are three basic concepts of program quality evaluation: (1) outcome-based, where the quality of the program is judged on the basis of the outcomes usually obtained through children's results on tests scores; (2) standards-based, where the quality of the program is judged based on how it complies with externally developed standards (which routinely include tools like the Early Childhood Environment Rating Scale (ECERS) to rate the quality of the classroom, sometimes tools to assess caregiver behavior, or program implementation or state/national standards); and finally, (3) quality based on the program's utilization of developmental appropriateness practices (DAPs), usually in relation to distinctions outlined in the US National Association for the Education of Young Children guidelines. Ironically, while the DAP guidelines are commendable in principle (and in many ways represent the full pendulum swing from skill-oriented practices), if they are used as a sole evaluation criterion, they share the major problem with evaluating the ECEC programs solely on test-based outcomes: children's development in the early years is fluid, uneven, and idiosyncratic, and at any given point in time their needs and performance may dramatically differ from a day before or after. They have been criticized for relying too heavily on the dominant, Western, ideals. Nevertheless, the inclusion of broadly based, inclusive DAP among the criteria for evaluating program quality, is very likely the avenue for future evaluations.

International Perspective

The dimensions of high quality in early childhood education are universal. The knowledge as to the extent to which they are implemented and evaluated at the national level is very poor. Most evidence is available for English-speaking countries, simply because most accessible results are published in English. This constitutes a very significant bias in assessing international perspectives, as lack of accessible evidence does not mean that there is no evidence at all.

What can be learned from the limited international evidence is that there are both cultural and societal differences in the quality and effectiveness of ECEC and also differences in the infrastructure on which the ECEC programs and their evaluations are built. Research for the Organization for Economic Cooperation and Development (OECD, 2006) Report, *Starting Strong II: Early Childhood Education and Care*, indicates that a critical mass of academics in the field of early childhood education is imperative for high-quality research and evaluation of ECEC. The report suggests that a number of developed countries have established comprehensive, early-childhood-oriented research agendas through which universities actively participate in

partnership with government agencies. United States is clearly in the lead with the longitudinal agenda of following children, thus creating the opportunity to provide evaluative evidence on the impact of ECEC on later development, but countries like Canada, Australia, the UK, and Korea have longitudinal study cohorts that have been (or currently are) followed through childhood. In the Canadian National Longitudinal Study of Children and Youth, the oldest children are now young adults; in the Ontario Child Health Study (a provincial study in Canada), recently published research examined the impact of early childhood living conditions (neighborhood and family) on the study participants' life as adults. The Longitudinal Study of Australian Children is in its fourth year. While these studies offer opportunities to examine the impact of children's participation in a broad range of ECEC services, they cannot evaluate any one type of program, as they were never originally designed for such purpose. Longitudinal studies more closely linked with specific early childhood education goals, like the UK's Effective Provision of Pre-School Education Project, conducted on children's development between 3 and 7 years of age is an example of a much more focused evaluative research activity. While not evaluation in the strict meaning of the term, practitioner-based research allowing for self-evaluation and reflection on their own practice is utilized in countries like Norway, Belgium, and Italy, and has been incorporated into professional development in Belgium, Norway, Sweden, Finland, and the United Kingdom. If supported by established and rigorous methodology, this research can often contribute to a fuller picture of ECEC.

In their vision for a globally blended high-quality ECEC systems, Jalongo *et al.* recommended that each country should establish its own framework guiding the provision of culturally responsive and sensitive programs, which at the same time would have demonstrated effectiveness and responsiveness to the needs of all types of families. A greater level of international communication and sharing of experiences could only be beneficial in approaching the evaluation of ECEC programs from the sociocultural perspective. The dominant models of early childhood education are often criticized for their rigidity in some aspects. The process of evaluation has to evolve with the changing composition and needs of a society and its children. Some basic truths remain the same – children have an unshakeable right to develop optimally and be provided with the best environment in which to do so. However, some of the context has changed: the understanding of the importance of the early years has grown among the general public, children with special needs are no longer isolated from the mainstream education, and the diversity of societies in which children grow has increased, to name just a few of the important factors. Evaluating the ECEC programs that serve the youngest members of society has to be flexible enough to incorporate the contextual

information, rigorous enough to produce interpretable results, and articulate enough to be incorporated into action.

See also: Early Care and Education Programs for Infants and Toddlers; Early Social Development and Schooling; Educational Evaluation: Concepts, Practice, and Future Directions; Integrated Early Childhood Education and Care Services – Care, Upbringing, Education and Health; Investing in Early Childhood Education and Care: Some Policy Implications; Investing in Early Childhood Education and Care: The Economic Case; Investing in Early Childhood Education and Care: The Health and Well-being Case; Literacies in Early Childhood: The Preschool Period; Parent–Child Relationships in Early Learning; Participation in Early Childhood Education and Care Programs: Equity, Diversity and Educational Disadvantage; Program Evaluation; Qualitative Case Studies; Teaching in Early Childhood Centers Instructional Methods and Child Outcomes.

Bibliography

- Jalongo, M. R., Fennimore, B. S., Pattnaik, J., *et al.* (2004). Blended perspectives: A global vision for high-quality early childhood education. *Early Childhood Education Journal* **32**, 143–155.
- OECD (2006). *Starting Strong II: Early Childhood Education and Care*. Paris: OECD.
- St. Pierre, R. G and Rossi, P. H (2006). Randomize groups, not individuals. A strategy for improving early childhood programs. *Evaluation Review* **30**, 656–685.

Further Reading

- Aber, J. L., Bishop-Josef, S. J., Jones, S. M., McLearn, K. T., and Phillips, D. A. (eds.) (2007). *Child Development and Social Policy: Knowledge for Action*. Washington, DC: American Psychological Association.
- Bocock, S. S., Barnett, W. S., and Frede, E. (2001). Long-term outcomes of early childhood programs in other nations: Lessons for Americans. *Young Children* **56**, 43–50.
- Bredenkamp, S. and Copple, C. (eds.) (1997). *Developmentally Appropriate Practice in Early Childhood Programs*. Washington, DC: National Association for the Education of Young Children.
- Epps, S. and Jackson, B. J. (2000). Strategies for effective program evaluation: A critical factor in quality services. In Epps, S. and Jackson, B. J. (eds.) *Empowered Families, Successful Children: Early Intervention Programs that Work*, pp 229–248. Washington, DC: American Psychological Association.
- Gilliam, W. S. and Zigler, E. F. (2000). A critical meta-analysis of all evaluations of state-funded preschool from 1977 to 1998: Implications for policy, service delivery and program evaluation. *Early Childhood Research Quarterly* **15**, 441–473.
- Hatch, J. A. and Grieshaber, S. (2002). Child observation and accountability in early childhood education: Perspectives from Australia and United States. *Early Childhood Education Journal* **29**, 227–231.
- Gorey, K. M. (2001). Early childhood education: A meta-analytic affirmation of the short- and long-term benefits of educational opportunity. *School Psychology Quarterly* **16**, 9–30.
- Groark, C. J., Mehaffie, K. E., McCall, R. B., and Greenberg, M. T. (eds.) (2006). *Evidence-Based Practices and Programs for Early Childhood Care and Education*. Thousand Oaks, CA: Corwin.

- Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* **142**, 615–635.
- Kagan, S. L. and Cohen, N. E. (eds.) (2004). *Reinventing Early Care and Education: A Vision for a Quality System*. San Francisco, CA: Jossey Bass.
- Kagitcibasi, C. (1996). *Family and Human Development Across Cultures: A View from the Other Side*. Mahwah, NJ: Erlbaum.
- Kammerman, S. (ed.) (2001). *Early Childhood Education and Care: International Perspectives*. New York, NY: Columbia University, The Institute for Child and Family Policy.
- Lee, J. -H. and Walsh, D. J. (2004). Quality in early childhood programs: Reflections from program evaluation practices. *American Journal of Evaluation* **25**, 351–373.
- Leseman, P. (2002). *Early Childhood Education and Care for Children from Low-Income or Minority Backgrounds*. Paris: OECD.
- Love, J. M., Kisker, E. E., Ross, C., et al. (2005). The effectiveness of early head start for 3-year-old children and their parents: Lessons for policy and programs. *Developmental Psychology* **41**, 885–901.
- OECD (2001). *Starting Strong: Early Childhood Education and Care*. Paris: OECD.
- Reynolds, A. J. (2005). Confirmatory program evaluation: Applications to early childhood interventions. *Teachers College Record* **107**, 2401–2425.
- Shonkoff, J. P. and Phillips, D. A. (2000). *From Neurons to Neighborhoods: The Science of Early Childhood*. Washington, DC: National Academy Press.
- Spodek, B. and Saracho, O. N. (eds.) (2006). *Handbook of Research on the Education of Young Children*, 2nd edn. Mahwah, NJ: Erlbaum.
- Teti, D. M. (ed.) (2005). *Handbook of Research Methods in Developmental Science*. Malden, MA: Blackwell.
- Woodhead, M. (1998). “Quality” in early childhood programs: A contextually appropriate approach. *International Journal of Early Years Education* **6**, 5–17.

EARLY CHILDHOOD EDUCATION AND CARE INVESTMENT

Contents

Investing in Early Childhood Education and Care: The Health and Wellbeing Case
Cost-effective Early Childhood Programs from Preschool to Third Grade
Early Childhood in Post-Modern Cultures: Thoughts and Some Concerns
Investing in Early Childhood Education and Care: Some Policy Implications
Investing in Early Childhood Education and Care: The Economic Case

Investing in Early Childhood Education and Care: The Health and Wellbeing Case

T Moore, The Royal Children's Hospital, Melbourne, Australia

F Oberklaid, The Royal Children's Hospital, Parkville, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

The dramatic social and demographic changes that have occurred in developed nations over the past several decades have led to significant changes in the circumstances in which families are raising young children, and in the conditions that young children experience as they grow up. In parallel with these social and demographic changes, there is evidence of worsening health and developmental outcomes for many children and young people (Perrin *et al.*, 2007). These have associated social and economic costs that undermine the general productivity and well-being of countries, and the fact that they have occurred in those very countries that have benefited most from economic and technological advances has been termed modernity's paradox (Keating and Hertzman, 1999).

There has been extensive research into those biological and environmental factors that increase the risk of poor outcomes, in children and throughout the life course; similarly, much is known about the protective factors that foster resilience in children who are otherwise at risk. While the pathways that lead to health, developmental, and social problems have been well described, the research about effective interventions is less than robust, both at an individual and, especially, at a population level.

Factors That Influence Child Health and Development

There are many factors, both biological and environmental, that impact on child health and development – genetic, familial, environmental, community, and societal. It is rare that any one of these factors determines child

health outcomes, or that they act in a simple causal fashion. While there are some genetic causes that act in a direct, linear causal fashion to threaten health and well-being, multiple factors usually interact with each other in a more complex fashion to influence outcomes. This representation of interactions among multiple influences that affect children's health has been called a kaleidoscope model (National Research Council and Institute of Medicine, 2004). These interactions between factors may occur within single domains, such as behavior or social environment, or they may extend beyond domains, for example, the child's social environment, behavior, and local or national policies. Changes in any one factor may influence others, thereby giving rise to a complex interaction of factors that may all play a role in determining the child's current and future health. Each influence, in turn, interacts with other influences to form a pattern, which in turn sets up the substrate for future patterns. In this way, children's health and well-being are determined by the prior state of the child's health, and by the presence or absence of risk and protective factors, and their interaction.

Genetic Factors

Genes have their effect either in correlation with or in interaction with the environment. Individual differences in human development may to a large extent be explained by gene–environment interactions that result in differential susceptibility. Common childhood conditions, such as obesity, asthma, and attention-deficit hyperactivity disorder (ADHD), have all been described as having a strong genetic foundation. Nonetheless, changes in the gene pool

cannot explain the recent dramatic growth of these conditions, suggesting that the strongest etiological pathways are environmental (Perrin *et al.*, 2007).

Familial Factors

Poverty

One of the most consistent associations in developmental science is the relationship between economic hardship and compromised outcomes in children's health, well-being, and development. Furthermore, it is suggested that "the malleability of young children's development and the overwhelming importance of the family . . . context suggest that economic conditions in early childhood may be far more important for shaping children's ability, behavior and achievement than conditions later in childhood" (Shonkoff and Phillips, 2000).

Maternal employment

There is some accumulating evidence that maternal employment in the child's first year of life, especially if the mother works long hours, can be a negative factor for infant development. However, beyond the first year, there is evidence that children may actually benefit from maternal employment, especially those from disadvantaged families, and particularly with regard to cognitive outcomes. In this context, the out-of-home care that these children experience can be regarded as a form of early intervention. Overall, parental employment can be seen as being either a positive or a negative factor for young children's development; this depends on the nature and structure of the job, the income it generates, and especially on the environments and relationships that children experience when they are not in the care of their parents.

Parental education

There is a strong consistent correlation between parental education level and their children's achievement and behavior. Parental education levels are strongly associated with the home literacy environment, parental teaching styles, and investment in a variety of resources that promote learning. Children of parents with limited education, especially maternal, are at an increased risk of behavior problems, poor literacy and school performance, and developmental delay.

Family structure

There is little evidence that family structure, in and of itself, is a significant factor in determining child outcomes. While growing up in a single-parent family may increase the risk of school difficulties and behavior problems, these are related to the socioeconomic realities of single parenthood – lower income and less parental time from both mothers and fathers – rather than from any direct effect of living with only one parent. There is no

evidence that children growing up in nontraditional family forms are at any increased risk of poor outcomes.

Parenting

An increasing body of research has documented the relationship between parenting and the development of a wide range of health, developmental, and behavior problems both in childhood and through to adulthood (Richter, 2004). For example, the quality of relationships that parents have with their children predict healthy eating, and the only programs that are successful (albeit modestly so) in treating childhood obesity are those that focus on parenting skills as well as lifestyle advice. Adverse parenting is also a risk factor for the adoption of smoking and alcohol use, teenage pregnancy, and poor mental health in children and adolescents. These relationships appear to be independent of socioeconomic status. It has been suggested that poor parent-child relationships have an adverse impact on the areas of the brain that deal with emotional and social functioning and with the physiological response to stress. Healthy and unhealthy patterns of relationships in the early years seem to be hard-wired into the brain early in life, dictating subsequent resilience of vulnerability to stress, and proving difficult, if not impossible, to moderate later in the life cycle.

Community Factors

Evidence on the impacts of community and neighborhood environments on child development and health is complex. Evidence suggests that dramatic changes – for example, moving from high-poverty to low-poverty neighborhoods – can enhance the physical and psychological health of children. Whether smaller, more easily achieved changes in neighborhood conditions produce improvements in children's health, well-being, and development is less clear.

Social support

Numerous studies of children and families have shown that social support has a direct influence on the well-being of children and families. Social support has been found to be linked to a number of child and family outcomes, including low birth weight, child abuse, child neglect, maternal adjustment, mental health, and physical health (Cooper *et al.*, 1999). Those families who are most in need of social support and who would benefit from it are often those that are the most isolated.

Societal Effects

Health is affected by environmental and social processes as well as by sociological factors, and the society and community in which a child lives are major determinants of health. Therefore, it has been argued that, because the primary determinants of disease in the twenty-first

century are mainly economic and social, the remedies must also include economic and social interventions (Rose, 1992). At a population level, the contribution of traditional medical care is modest, and health-care needs to focus more on prevention. “This involves community approaches as well as individual health care, and must take into account the physical and mental health of the adults who interact with children and young people” (Hall and Elliman, 2003).

Adverse health outcomes have been shown to be associated with environments that threaten personal safety, that limit the ability to develop strong social ties, or that are characterized by conflictual, violent or abusive interpersonal relationships. These effects occur across the lifespan. Positive health outcomes are associated with environments that provide safety, opportunities for social integration, and the ability to predict and/or control aspects of that environment.

Social capital

Social capital, defined as the networks of social relations that are characterized by norms of trust and reciprocity, has been shown to be related to health outcomes. In communities that are high in social capital, there are strong connections between members of the community based on mutual trust and reciprocal exchanges. Like social support, social capital has been linked to a number of factors, including improved health, greater well-being, better care for children, and lower crime rate (Kroll, 2008).

Social gradients

There is clear evidence of the relationship between social gradients and health outcomes: where we stand in the social hierarchy is intimately related to our chances of getting ill and to how long we live. The Commission on Social Determinants of Health (2008) took a holistic view of the social determinants of health:

The poor health of the poor, the social gradient in health within countries, and the marked health inequities between countries are caused by the unequal distribution of power, income, goods, and services, globally and nationally, the consequent unfairness in the immediate, visible circumstances of people’s lives – their access to health care, schools, and education, their conditions of work and leisure, their homes, communities, towns, or cities – and their chances of leading a flourishing life. This unequal distribution of health-damaging experiences is not in any sense a ‘natural’ phenomenon but is the result of a toxic combination of poor social policies and programs, unfair economic arrangements, and bad politics. Together, the structural determinants and conditions of daily life constitute the social determinants of health and are responsible for a major part of health inequities between and within countries.

How do social gradients affect health? Wilkinson (2005) argues that inequality is socially corrosive and affects health because the quality of social relations is crucial to well-being. In wealthy countries, health is not simply a matter of how material circumstances determine quality of life and access to healthcare; it is how social standing makes a person feel. Low social status – being devalued and looked down on – is stressful and can have devastating effects on people’s lives and communities. More unequal societies have poorer communal environments, which in turn is related to a range of social issues from higher levels of violence to more widespread depression.

Environmental Factors

Child health and development may be affected directly or indirectly by a number of environmental factors.

Climate change

Climate change poses direct and indirect risks to health, impacting on exposure to sun and extremes of climate, food safety, water supplies, and a higher risk of natural disasters such as fires and floods. There is almost no component of health and well-being that will be untouched by climate change. However, it is suggested that the risk is distributed unequally across society, as vulnerability to the effects of climate change depends on the degree of exposure, sensitivity, and adaptive capacity.

Environmental toxins

It is suggested that, over the past few decades, environmental exposures are contributing to children’s declining health status. Environmental agents that we know to cause health and developmental problems in humans include alcohol, nicotine, lead, mercury, arsenic, solvents, pesticides, and many others. Many of these are organic and inescapable; they enter the food chain from sources such as pesticides, chemical manufacturing, and incinerated waste, and accumulate in animals higher up the chain. However, there are many more chemicals in common use whose effects on children, either singly or in combination, are unknown because they have never been tested (Collaborative on Health and the Environment Learning and Developmental Disabilities Initiative, 2007).

Children are often more susceptible than adults to the effects of exposure to environmental agents. This is a particular issue during pregnancy, where the fetus is exposed to larger doses relative to bodyweight (International Scientific Committee of the International Conference on Foetal Programming & Development Toxicity, 2007). Exposure during fetal development can adversely affect health and well-being and can lead to lifelong functional deficits and increased disease risks.

Exposure to environmental pollution, although a major source of health risk throughout the world, is particularly problematic in developing countries where unsafe water, poor sanitation, and poor hygiene, along with indoor air pollution, are major sources of exposure.

Associations between environmental pollution and health outcomes are complex; individual pollutants may be implicated in a wide range of health effects, whereas few diseases are directly attributable to single pollutants. However, most of our exposures to these chemicals are not from sources traditionally regulated, such as remote waste sites and factories. Rather, the primary sources are close to us: within our indoor environments, and the personal activities, products, and materials inside those environments. The sources of these pollutants are largely unregulated – meaning that our environmental regulations, designed to protect and promote human health, are missing major sources of health risks.

Changes in urban environments

There is growing recognition that the built environment – the man-made physical structures and infrastructure of communities – has an impact on health. A good example of this is the opportunities children have for physical activity. Increased urban density, and reduced access to parks and safe places for children to play, accompanied by parental concern about children's safety, is said to have contributed to the increase in obesity (Perrin *et al.*, 2007). Physical environments and community recreation facilities have an impact on how children use their time, and particularly their likelihood of physical activity.

Changes in home living environments

Many allergies and immune system diseases have increased significantly in prevalence in the past few decades; asthma, hay fever, and food allergies the most notable among them significantly (Gibson, 2009). While the exact cause is debatable, it has been suggested that all of them may have a common explanation rooted in aspects of modern living. One theory, termed the hygiene hypothesis, suggests that these increases are due to children growing up in an increasingly sterile environment; other causes postulated include changes in diet, air pollution, and increasingly sedentary lifestyles. Evidence for the hygiene hypothesis comes from studies demonstrating lower rates of allergies in children who live on farms and whose mothers lived on a farm during pregnancy.

Changes in food consumption

There have been considerable changes in children's eating habits, including increases in high-energy foods, meals and snacks eaten outside the home, and increased

portion size. The so-called fast foods tend to have low-quality carbohydrates and fats, little fiber, few essential nutrients, and high-energy density. Fast-food outlets are more highly concentrated in lower socioeconomic areas, contributing to the higher rates of obesity in disadvantaged communities. Changes in food production and food consumption have seen increases in the use of food additives, reduction in fruit and vegetable intake, and an increase in the consumption of sugar-sweetened beverages.

Evidence of Long-Term Impacts of Early Experiences

Developmental pathways originate in the complex interplay between biology and experience; there is accumulating evidence of the child's immediate environment in the early years as having a major impact (Hertzman, 2004). The evidence suggests that early behavior and functioning are predictive of later behavior and functioning to the extent that children's social and physical environments remain unchanged. In other words, it is difficult for children and families to extricate themselves from adverse circumstances, especially when there are multiple risks or adverse factors in their lives.

However, there is also evidence of developmental plasticity over the life span. Plasticity is the potential for change in intrinsic characteristics in response to environmental stimuli. Children's development continues to be shaped by experiences throughout the course of childhood. Moreover, there is emerging evidence that suggests that the brain can change itself or can be changed by experience to a much greater degree than was previously recognized. For instance, stress neurobiology is highly responsive to changes in the environment: although very sensitive to early social contexts, it is not a fixed or inflexible system, but reflects both the organism's epigenetic history and its new circumstances. Improved living conditions, enriched environments, and corrective emotional experiences can reverse the adverse consequences of early adversity.

Hertzman and Power (2003) describe three mechanisms through which exposure to both beneficial and adverse circumstances over the life course impact on health and development.

Latency or Sleeper Effects

This hypothesizes that there is a relationship between exposure at one point in the life course and its impact on health many years later. In recent years, there has been an emerging body of research suggesting that the roots of adult disease lie in fetal and neonatal development (Barker, 1992).

It is suggested that the mechanism for this to occur is that nutrition or exposure to environmental toxins *in utero* in the neonatal period may affect the programming of tissue function that occurs during development. This concept is called the developmental basis of health and disease, and the process by which this occurs has been termed biological embedding (Keating and Hertzman, 1999).

It has been shown that low birth weight (small for gestational age) strongly predicts the subsequence of hypertension, hyperlipidemia, insulin resistance, type 2 diabetes, and ischemic heart disease (Heindel, 2007). Low birth weight is taken to be an indicator of poor nutrition during pregnancy, and it is hypothesized that the fetus permanently changes its structure and metabolism as an adaptation to a limited supply of nutrients. The metabolic demands of the growing brain and heart are favored at the expense of other tissues.

Cumulative Effects

In this model, there are either multiple exposures to a single recurrent factor (such as poverty) or a series of exposures to different factors; these exposures to risk factors (and protective factors) may accumulate over the life course. While in some cases exposure occurs in a dose-response manner – for example, the health effects of exposure to a toxic substance usually increase with the duration and intensity of exposure – the biological mechanisms by which disadvantages and inequities carried over a life course of differential exposures leading to health disparities are not well understood. One mechanism might be the body's biological response to stress, involving the autonomic nervous system, the hypothalamic-pituitary-adrenal (HPA) axis, and the cardiovascular, metabolic, and immune systems.

Many studies have shown the cumulative impact of risk and protective factors on young children's development. These studies show that the risk of core developmental outcomes in children and adolescents increases in a linear fashion as the number of environmental risks increases. The corollary is that the incidence of positive developmental outcomes increases as the number of protective factors in children's lives increases. There are well-established correlations between a range of adverse childhood experiences, including abuse, neglect, and household dysfunction, and later health problems such as ischemic heart disease. This effect is cumulative – the more the number of adverse childhood experiences, the more the likelihood of the child developing ischemic heart disease later in life. These adverse early experiences are more strongly predictive than traditional risk factors such as smoking, obesity, diabetes, and hypertension. Adverse events in childhood have also been shown to be

directly correlated with other problems in adulthood, including alcoholism, illicit drug use, and mental health problems such as depression.

Pathway Effects

In this model, exposure to risk factors at one stage of the life course influences the probability of other exposures later in the life course. Early events may influence the life-course trajectory, and once a child is established on that trajectory he or she is more likely to experience other exposures that strengthen the likelihood of poor outcomes. For example, children from disadvantaged backgrounds have a greater risk of poor school readiness, as measured by cognitive and social-emotional competencies at school entry. Due to this they are more likely to experience problems at school, with a subsequent increased risk of unemployment, poor self-esteem, and mental health problems. It is hypothesized therefore that early adverse experiences disrupt neurological development, resulting in social, emotional, and cognitive impairments, which in turn lead to increased risk to health behaviors that result in disease, disability, and social problems. Mechanisms for this are uncertain, although one suggestion is that a cascade of risks is created early in life that exacerbates certain genetically based vulnerabilities. This may lead to deficits in children's control of and expression of emotions and social competence, and also to the disturbance of physiological and neuroendocrine system regulation that can have cumulative, long-term adverse effects.

The three models described are not mutually exclusive, but are present simultaneously in any individual's life course. Environmental lead exposure can be used to illustrate this. This can have a cumulative impact because lead is retained in the body, and there is a dose-response curve so that the more it is retained, the greater the health risk. At the same time, lead exposure *in utero* may also have a latent effect on subsequent health and well-being through inhibiting the production of brain cells. Finally, there may also be a pathway effect since the socioeconomic position will affect the probability of exposure to lead *in utero* and in childhood, which may impact negatively on a range of outcomes throughout the life course.

Each of the three sets of influences on health described by Hertzman and Power (2003) – latency, cumulative, and pathway – carries with it a strategic message for policies and interventions to improve population health. The message of latency is, the earlier the better; the message of the cumulative model, is intervene wherever there is an effective intervention; and the message of the pathways model is, intervene at strategic points in time. All three of these forms of intervention will be needed if broad-ranging improvements in health and developmental outcomes for children are to be achieved.

Summary and Conclusions

Humans are unique in the degree to which they can adapt to their environments. Young children's neurology and biology are particularly adaptable, designed as they are to learn from their prenatal and early postnatal environments. They are highly sensitive to both positive and negative experiences, which lay down biological, neurological, and behavioral patterns that become increasingly difficult to change, and which can have lifelong consequences for health and well-being.

Given the evidence of the links between early experience and later health outcomes, there is a strong argument for greater investment in the early years to reduce the long-term burden of poor health in adulthood. Given the evidence that the primary determinants of poor health outcomes in the twenty-first century are mainly social and economic, these investments should focus not so much on traditional forms of medical care, but on addressing the social and economic conditions under which families are raising young children.

Bibliography

- Barker, D. J. P. (1992). *Fetal and Infant Origins of Adult Disease*. London: British Medical Publishing Group.
- Collaborative on Health and the Environment's Learning and Developmental Disabilities Initiative (2007). *Scientific Consensus Statement on Environmental Agents Associated with Neurodevelopmental Disorders*. Freeland: Institute for Children's Environmental Health.
- Commission on Social Determinants of Health (2008). Closing the gap in a generation: Health equity through action on the social determinants of health. *Final Report of the WHO Commission on Social Determinants of Health*. Geneva, Switzerland: World Health Organisation.
- Cooper, H., Arber, S., Fee, L., and Ginn, J. (1999). *The Influence of Social Support and Social Capital on Health*. London, UK: Health Education Authority.
- Gibson, G. (2009). *It Takes a Genome: How a Clash Between Our Genes and Modern Life Is Making Us Sick*. Upper Saddle River, NJ: FT Press Science.
- Hall, D. M. B. and Elliman, D. (eds.) (2003). *Health for All Children*, 4th edn. Oxford, UK: Oxford University Press.
- Heindel, J. J. (2007). The developmental basis of health and disease. *Reproductive Toxicology* **23**(3), 257–259.
- Hertzman, C. (2004). *Making Early Childhood Development a Priority: Lessons from Vancouver*. Ottawa, Ontario: Canadian Centre for Policy Alternatives. <http://www.policyalternatives.ca/bc/early-childhood.pdf>.
- Hertzman, C. and Power, C. (2003). Health and human development: Understandings from life-course research. *Developmental Neuropsychology* **24**(2–3), 719–744.
- International Scientific Committee of the International Conference on Fetal Programming and Developmental Toxicity (2007). The Faroes statement: Human health effects of developmental exposure to environmental toxicants. *Proceedings of the International Conference on Fetal Programming and Developmental Toxicity*, Tórshavn, Faroe Islands, May 20–24.
- Keating, D. P. and Hertzman, C. (1999). Modernity's paradox. In Keating, D. P. and Hertzman, C. (eds.) *Developmental Health and the Wealth of Nations: Social, Biological, and Educational Dynamics*. New York: The Guilford Press.
- Kroll, C. (2008). *Social Capital and the Happiness of Nations: The Importance of Trust and Networks for Life Satisfaction in a Cross-National Perspective*. Oxford, UK: Peter Lang Publishing.
- National Research Council and Institute of Medicine (2004). *Children's Health, The Nation's Wealth: Assessing and Improving Child Health*. Committee on Evaluation of Children's Health, National Research Council. Washington, DC: National Academies Press.
- Perrin, J. M., Bloom, S. R., and Gortmaker, S. L. (2007). The increase of childhood chronic conditions in the United States. *Journal of the American Medical Association* **297**(24), 2755–2759.
- Richter, L. (2004). *The Importance of Caregiver–Child Interactions for the Survival and Healthy Development of Young Children: A Review*. Geneva: Department of Child and Adolescent Health and Development, World Health Organisation.
- Rose, G. (1992). *The Strategy of Preventative Medicine: The Distribution of Social Problems in Society*. Oxford, UK: Medical Publications.
- Shonkoff, J. P. and Phillips, D. A. (eds.) (2000). From neurons to neighborhoods: The science of early childhood development. Committee on Integrating the Science of Early Childhood Development, National Research Council and Institute of Medicine. Washington, DC: National Academy Press.
- Stein, R. E. K. (2005). Children's health, the nation's wealth: Assessing and improving child health. *Ambulatory Pediatrics* **5**(3), 131–133.
- Wilkinson, R. G. (2005). *The Impact of Inequality: How to Make Sick Societies Healthier*. New York: The New Press.

Further Reading

- Baker, N. (2008). *The Body Toxic: How the Hazardous Chemistry of Everyday Things Threatens Our Health and Well-being*. New York: Farrar, Straus and Giroux.
- Barker, D. J. P. (1998). *Mothers, Babies and Health in Later life*. Edinburgh: Churchill Livingstone.
- Doide, N. (2007). *The Brain That Changes Itself*. New York: Viking.
- Gluckman, P. and Hanson, M. A. (2005). *The Fetal Matrix: Evolution, Development and Disease*. Cambridge, UK: Cambridge University Press.
- Gluckman, P. and Hanson, M. (2006). *Mismatch: Why Our World No Longer Fits Our Bodies*. Oxford, UK: Oxford University Press.
- Golombok, S. (2000). *Parenting: What Really Counts?* London, UK: Routledge.
- Graham, H. and Power, C. (2004). *Childhood Disadvantage and Adult Health: A Lifecourse Framework*. London, UK: Health Development Agency.
- Gunnar, M. and Quevedo, K. (2007). The neurobiology of stress and development. *Annual Review of Psychology* **58**, 145–173.
- Halfon, N. and Hochstein, M. (2002). Life course health development: An integrated framework for developing health, policy, and research. *The Milbank Quarterly* **80**(3), 433–479.
- Heinzerling, L. (2008). Health regulation and governance: Climate change, human health, and the post-cautionary principle. *Georgetown Law Journal* **96**, 445–460.
- Kuh, D. and Ben-Shlomo, Y. (eds.) (2004). *A Life Course Approach to Chronic Disease Epidemiology: Tracing the Origins of Ill-Health from Early to Adult Life*, 2nd edn. Oxford: Oxford University Press.
- Marmot, M. (2004). *Status Syndrome: How Your Social Standing Directly Affects Your Health and Life*. London, UK: Bloomsbury.
- Prüss-Ustün, A. and Corvalán, C. (2006). *Preventing Disease Through Healthy Environments: Towards an Estimate of the Environmental Burden of Disease*. Geneva: World Health Organisation.
- Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.
- Richardson, S. and Prior, M. (2005). Childhood today. In Richardson, S. and Prior, M. (eds.) *No Time to Lose: The Wellbeing of Australia's Children*. Melbourne, VIC: Melbourne University Press.
- Rutter, M. (2006). *Genes and Behaviour: Nature-Nurture Interplay Explained*. Oxford, UK: Blackwells.
- Stewart-Brown, S. (2008). Improving parenting: The why and the how. *Archives of Disease in Childhood* **93**(2), 102–104.
- Wilkinson, R. G. and Pickett, K. E. (2009). *The Spirit Level: Why More Equal Societies Almost Always Do Better*. London: Allen Lane.

Cost-effective Early Childhood Programs from Preschool to Third Grade

A J Reynolds, J A Temple, and B White, University of Minnesota, Minneapolis, MN, USA

© 2010 Elsevier Ltd. All rights reserved.

The positive effects of early-childhood development (ECD) programs on child well-being have been documented in hundreds of studies and reviews since the 1960s (Karoly *et al.*, 2005; Reynolds *et al.*, 2003; Zigler *et al.*, 2006). Advances in scientific knowledge have contributed not only to the establishment of early-childhood programs but have helped spur recent expansion of programs and services in states and localities (Barnett *et al.*, 2007; Reynolds and Temple, 2008; Zigler *et al.*, 2006). Increased attention to the early years of life also has sparked greater interest in the transition to school and the experiences in the early grades that can reinforce preschool gains and strengthen school achievement (Bogard and Takanishi, 2005; Reynolds *et al.*, 2003). The creation of early-childhood systems and practices that enhance the continuity of development over the first decade of life is increasingly viewed as fundamental to shaping children's well-being.

Scope of Review

In this article, evidence is presented on the effectiveness and cost effectiveness of ECD programs for the outcomes of school readiness, school achievement, and performance, and long-term life-course development. The primary focus is on preschool or pre-kindergarten programs for 3- and 4-year-olds and early school-age programs including preschool to third-grade programs (PK-3) and practices. Two major questions are addressed:

1. What are the effects and economic benefits of ECD programs from preschool to the early elementary grades?
2. Which elements and principles of effectiveness are key to long-term effects?

In this article, ECD is defined broadly to include most of the first decade of children's life. Findings from prenatal and infant programs are beyond the scope of the article, although we summarize evidence comparing the effects of birth-to-3 program to preschool and early school-age programs. A much longer discussion of some of these programs can be found in Reynolds and Temple (2008).

Although findings on achievement and other short-term outcomes are reported, we emphasize the results of cost-benefit analyses (CBAs) of the effects of programs. There are three reasons for the focus on CBAs.

First, economic benefits relative to costs are the most relevant indicator for policy development. The value of

public investments can be judged, at least in part, on efficiency (Heckman, 2000). This is especially true in a time of scarce resources for educational and social programs. Second, in the economic-benefit approach, program effects on multiple outcomes can be converted into the metric of dollars and cents (Levin and McEwan, 2001). Other metrics such as standard deviation units or percent change cannot be used to combine or compare outcomes.

Finally, cost-benefit analyses emphasize longer-term effects of programs and practices. A focus on immediate and shorter-term effects, while an important first step, is not the ultimate program goal. Longer-term effects are a major focus of early-childhood programs. A major question for social policy is whether short-term effects translate into long-term effects of adaptive life skills and behavior.

How ECD Programs Influence Outcomes

Considerable research has documented that ECD programs impact later well-being through at least one of five processes or pathways (Reynolds, 2000; Reynolds *et al.*, 2004). In short, these can be viewed as the active ingredients contributing to impacts on child development. They have been conceptualized from the beginning of the early childhood as the primary mechanisms of intervention effects. As shown in **Figure 1**, the cognitive-advantage pathway indicates that the longer-term effects of ECD programs are primarily due to the enhancement of cognitive skills, including literacy, school readiness, language, and numeracy.

The family-support pathway indicates that impacts on child outcomes derive from greater parental investments in children's development, such as greater parent involvement in education, increased parenting skills, and greater resource supports for parents.

The school-support pathway suggests that longer-term effects would occur to the degree that post-program school experiences reinforce learning gains. Enrolment in higher-quality schools and schools with positive learning environments would strengthen or maintain learning gains while enrolment in schools lower in quality would neutralize earlier learning gains.

The social adjustment and motivational-advantage hypotheses indicate that noncognitive skills can be the

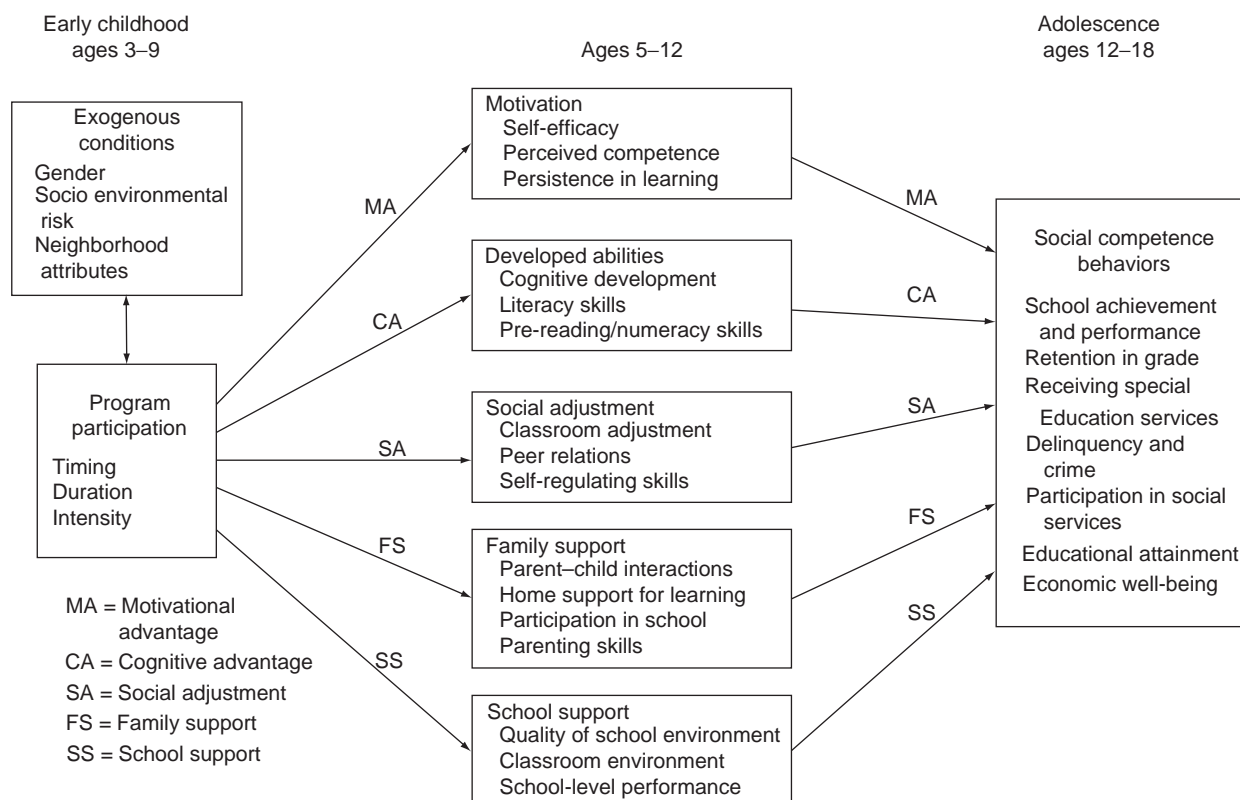


Figure 1 Pathways from early-childhood programs to long-term outcomes.

mechanism of effects of ECD programs, such as increased classroom and peer social skills, positive teacher-child relationships, achievement motivation, and school commitment. The greater the magnitude of effect of program experiences on a particular pathway or multiple pathways, the more likely that enduring effects would occur.

Notably, programs that provide comprehensive services would be expected to impact several of the pathways simultaneously. This is one explanation for why (or simply why) comprehensive programs are found to be more likely to have longer-term effects.

Cumulative Evidence of Effects

Given the size of the knowledge base, the effects of preschool ECD programs are summarized through findings from 19 reviews of preschool impacts published in the past decade (1995–2006; Reynolds and Temple, 2008). These reviews were selected as among the most thorough in assessing short- and longer-term effects of both model and large-scale programs (contact the author for additional information). **Table 1** shows the most frequently cited programs demonstrating beneficial effects along with the last age of follow-up.

Two major conclusions are evident from the reviews. First, many programs have assessed long-term effects into adulthood. Three-quarters of the reviews reported effects at

5 or more years after the end of participation. This is rare for social programs and indicates that impacts on life-course development and economic benefits can be accurately assessed. Second, the accumulated evidence includes both model programs, developed for research demonstration, and large-scale programs, developed for routine implementation by schools and other institutions. Consequently, the generalizability of the evidence for policy recommendations is much stronger today than a decade ago.

What are the main findings of the reviews? Of the hundreds of studies synthesized, there is substantial evidence that preschool programs for mostly children at risk, positively impact cognitive skills, school achievement, social and emotional development, as well as educational attainment, employment, and later social behavior. The average effect size on cognitive skills at or near school entry was 0.42 standard deviations (sd), which is roughly equivalent to one-half of a year of growth associated with preschool participation. Average effects were also statistically and practically significant for social and emotional development (0.24 sd), school achievement (0.35 sd), delinquency and crime (–0.22 sd), grade retention (–0.21 sd), special education (–0.46 sd), school completion (0.27 sd), and employment and earnings (0.37 sd). As a number of these significant outcomes can be translated to dollar terms, the method of benefit-cost analysis is well suited for analyzing the net benefits to society of these preschool programs.

Table 1 Most frequently cited early-childhood-education programs from research reviews (1995–2006)

<i>Program</i>	<i>Type</i>	<i>Age</i>	<i>Citations</i>
High/Scope Perry Preschool Program	Model	40	19
Chicago Child–Parent Centers (CPC)	Large scale	24	14
Carolina Abecedarian Project	Model	21	13
Houston Parent–Child Development Center (PCDC)	Model	11	12
Infant and Health Development Program	Model	18	11
Comprehensive Child Development Program (CCDP)	Large scale	5	8
Early Training Project	Model	20	8
Prenatal/Early Infancy Project (PEIP)/Nurse–Family Partnership Program(NFP)	Model	15	8
Milwaukee Project	Model	14	8
Philadelphia Project	Model	18	7
Consortium for Longitudinal Studies	Model	27	6
Educational Testing Service Head Start Study	Large scale	8	6
New Haven Follow-Through Study	Large scale	17	6
Institute for Developmental Studies	Model	13	5
Louisville Experiment (Head Start)	Model	16	5
Even Start	Large scale	7	4
Harlem Training Project	Model	12	4
Maryland Head Start	Large scale	17	4
Yale Child Welfare Research Project	Model	10	4
Avance Family Support and Education	Large scale	5	3
New York State Experimental Prekindergarten	Large scale	9	3
PSID Head Start Longitudinal Study	Large scale	25	3

Effects and Economic Benefits of Three Preschool Programs

Based on their prevention focus and strong evidence of enduring effects, three early-childhood programs listed in **Table 1** have been the subjects of extensive cost-benefit analyses: Carolina Abecedarian Project (ABC), High/Scope Perry Preschool (PPP), and the Child–Parent Center Program (CPC). Although there were important differences, all three programs provided high-quality educational enrichment to children at risk in group settings characterized by small class sizes, a focus on language and cognitive skills, and well-qualified and compensated teachers. ABC was the most intensive and lengthy, providing full-day, year-round care for 5 years (Campbell and Ramey, 1995). PPP provided the most established and organized curriculum, which followed the Piagetian cognitive principle of child-initiated learning (Schweinhart *et al.*, 1993). CPC provides the most comprehensive services by implementing an intensive parent-involvement component, outreach services, and attention to health and nutrition (Reynolds, 2000; Reynolds *et al.*, 2002). It also is the only program that became established in public schools and is still in existence. The major long-term findings of the studies generating economic benefits are shown in **Table 2** (see also Reynolds and Temple, 2008). Not shown are the substantial effect sizes for program participation on cognitive skills at kindergarten entry, and on school achievement through the elementary grades. CPC participation also was associated with higher levels of parent involvement in school.

CBA in ECD Programs

CBA is an economic approach for estimating the value of alternative programs and policies relative to costs. Levin and McEwan (2001) define CBA as the “evaluation of alternatives according to their costs and benefits when each is measured in monetary terms” (p. 11). Using CBA, program and intervention investments can be ranked according to their effectiveness per dollar of expenditure. The most frequently used outcome measures in CBAs of ECD programs and their data sources are shown in **Table 3**.

CBA is a major departure from traditional measures of effect size, such as the *d* statistic and percentage-change metrics, which take into account only program effects while ignoring their costs. The use of CBA in the field of ECD has increased dramatically. To illustrate estimation, relative to the control group, participants in the Child–Parent Center preschool program spent an average of 0.7 fewer years in special education from kindergarten to high school. The effect size is 0.29 standard deviations. This translates to an average savings in special education of US\$5317 per program participant, which was calculated by multiplying the program effect of 0.7 years by the average annual cost per child for special education services above and beyond regular instruction estimated in 2007 dollars for the Chicago school district in which they were enrolled and discounting the cost, occurring at an average age of 10, to age 3 by an annual rate of 3% (Reynolds *et al.*, 2002).

The major advantage of CBA is that benefits for multiple outcomes can be summarized in dollar terms, either

Table 2 Adjusted means or percentages for program and comparison groups on key outcomes for cost-benefit analysis (CBA)

<i>Outcome</i>	<i>Perry Preschool^a</i>	<i>Abecedarian</i>	<i>Child-Parent Centers</i>
Original sample sizes (program, comparison)	58, 65	57, 54	989, 550
Sample recovery for high school completion (%)	94	95	87
Special-education services by age 15/18 (%)	15 vs. 34	25 vs. 48	14 vs. 25
Grade retention by age 15 (%)	ns	31 vs. 55	23 vs. 38
Child maltreatment by age 17 (%)	n/a	n/a	7 vs. 14
Arrested by age 19 (%)	31 vs. 51	ns	17 vs. 25
Highest grade completed by age 21/27 (mean)	11.9 vs. 11.0	12.2 vs. 11.6	11.3 vs. 10.9
High school completion by age 21/27 (%)	71 vs. 54	70 vs. 67 (graduation)	66 vs. 54
Attended college by age 21/27 (%)	33 vs. 28	36 vs. 14 (4-year college)	24 vs. 18
Employed at age 21–27 (%)	71 vs. 59	70 vs. 58	ns
Monthly earnings at age 27 (mean)	US\$1219 vs. US\$766	n/a	n/a

^aFor the Perry Preschool Study, special education is for equated monthly instalment (EMI) placements by age 15. Ages for educational attainment and employment are 27 for Perry, 21 for Abecedarian, and 22 for Chicago. Employment for the Chicago study is full-time employment.

ns – not significant; n/a – not available.

Table 3 Examples of monetized benefits in cost-effectiveness analyses of early-childhood programs

<i>Benefit category</i>	<i>Treatment effect</i>	<i>Estimate</i>
Grade retention	Reduced rates of grade retention	Expenditure for 1 year of school
Special education	Fewer years of special education	Weighted average annual cost for special-education services
Childcare expenditures	Average duration of program participation	Annual opportunity cost of parents time
Child maltreatment	Fewer indicated reports of abuse/neglect	Weighted average annual cost for in-home services and out-of-home services, including administrative and investigation costs
Public assistance	Fewer months of public-assistance participation	Average monthly payment for AFDC, food stamps, and Medicaid, including administrative costs
Adult health	Lower rates of smoking	Monetary value of an additional year of life
Lifetime earnings	Higher rates of high school completion	Projected increase in lifetime earnings associated with high school graduation
Juvenile crime	Fewer petitions to juvenile court	Estimated juvenile justice system expenditures per court petition

AFDC – Aid to families with dependent children.

the net return (benefits minus costs) or return per dollar invested (benefits divided by costs). However, the ability to conduct a CBA depends on whether or not it is possible to reflect program benefits in dollar terms. Researchers have a long tradition of estimating the benefits of increased graduation rates and reductions in crime. Often program budgets contribute information used to create estimates of the benefits of reductions in services such as special education or child welfare. It is more difficult to monetize the benefits of higher test scores or problem behaviors because relatively few studies link test scores or behavior to the more monetizable outcomes of higher future incomes or fewer crimes committed. When program outcomes cannot be easily converted to monetary terms, cost-effectiveness analysis is the recommended approach (see Levin and McEwan, 2001).

Cost-effectiveness analysis typically assumes that there is one main outcome of interest such as test scores.

Summary of Results of Three Cost-Benefit Analyses

At a minimum, the economic return should equal the amount invested in the program – a return of at least one dollar per dollar invested. CBAs can illustrate the distribution of the benefits across different segments of society. Benefits to participants are returned to the child and parent attending the program but do not directly benefit others in society. These benefits include increased earnings capacity in adulthood projected from educational attainment as well as the benefit to parents from

the provision of part-day care for children. Benefits to the general public include averted expenditures for remedial education and social welfare spending by governments, reduced tangible expenditures to crime victims as a result of lower rates of crime, and increased tax revenues to state and federal governments as a result of higher earnings capacity. Benefits to society at large include the sum of benefits to program participants and to the general public. While societal benefits frequently are emphasized because they represent the total economic benefit to society, the existence of benefits to the general public (not including the program participants) often is used to justify government involvement.

As shown in **Table 4**, all three programs showed substantial economic returns of preschool into adulthood through government savings in education, justice system, and health expenditures and in increased economic well-being.

Although the costs of the programs are significantly different from each other, the economic returns of each program far exceeded the initial investment. As a ratio of benefits to costs, all three programs showed a large return on investment based on data collected into adulthood, ranging from a total societal benefit of US\$2.49 per dollar invested to US\$16.14 per dollar invested. These are equivalent to a 249% to 1614% return on the dollar. The CPC program showed a relatively high benefit-cost ratio at US\$10.15, in part reflecting its relatively lower costs. The lower costs are primarily a result of a higher child to staff ratio in the classroom (8.5 to 1 vs. less than 6 to 1 for PPP and ABC). That a routinely implemented school-based program demonstrates positive returns indicates that wide-scale programs can be cost effective. The other school-based program, PPP, demonstrated an economic return of US\$16.14 per dollar invested as assessed at age 40. At US\$2.49 per dollar invested, ABC had the lowest benefit-cost ratio. This is not surprising given its high cost. In terms of public benefits alone (i.e., government and

crime victim savings), benefit-cost ratios ranged from US\$0.50 to US\$12.90 per dollar invested.

The consistent findings of the economic analyses of the PPP, CPC, and ABC programs, despite their major differences in social context and instructional approach are encouraging evidence in favor of expanding preschool access. Nevertheless, the participants of the three programs were almost exclusively low-income, African American children. While there is no comparable evidence from studies of middle-income families or from more diverse samples, research on the short-term effects of state-funded preschool programs, which include more diverse samples by socioeconomic status and race/ethnicity, find positive impacts. For example, effects sizes on school readiness (average of reading and math scores) of state pre-kindergarten programs for 4-year-olds implemented from 2002 to 2006 in New Mexico, Arkansas, New Jersey, and Oklahoma ranged from 0.26 to 0.58. These are statistically and educationally meaningful (for details, see Reynolds and Temple, 2008).

CBA from Policy Simulations

To estimate the economic benefits of high-quality but routinely implemented preschool programs, several researchers have conducted cost-benefit simulations that either (1) modify assumptions of existing CBAs from longitudinal analyses of model programs such as the Perry Preschool Project or the Chicago Child-Parent Centers or (2) make projections of benefits from predicted changes in educational attainment, income, or criminal behavior using information from other studies that have shown correlations between these adult outcomes and the observed short-term outcomes such as achievement scores. Three such analyses are summarized below. All indicate that more widely implemented preschool programs for 3- and 4-year-olds would be likely to yield benefits than significantly exceed costs (also see Reynolds and Temple, 2008).

Table 4 Summary of costs and benefits per participant in 2007 dollars for three early interventions

<i>Program cost (US\$)^a</i>	<i>High/Scope Perry Preschool</i>	<i>Chicago Child-Parent Centers</i>	<i>Abecedarian Project</i>
Average per participant	18 261	8 512	73 159
For 1 year of participation	10 283	5 434	16 020
Program benefits (\$)			
Total benefits	294 772	86 400	182 422
Net benefits (benefits – cost)	276 511	77 899	109 263
Public benefits	235 544	58 476	36 429
Net public benefits	217 283	49 964	(36 730)
Total benefit per dollar invested	16.14	10.15	2.49
Public benefit per dollar invested	12.90	6.87	0.50

^aCosts are program expenditures and do not include estimated costs for comparison-group experiences. Ages of study participants for economic analysis of the Perry Preschool Program, Chicago CPC Program, and the Abecedarian Project were 40, 21, and 22 respectively. The public benefit of the Abecedarian Project is estimated by assuming that 25% of the earnings benefit to participants, parents, and future generations are tax revenues. In addition, the public is assumed to be responsible for two-thirds of the cost associated with increased college attendance /adult education.

Using short- and long-term data from 58 evaluation studies published from 1967 to 2003, Aos *et al.* (2004) estimated an economic return of US\$2.36 return per dollar invested for preschool programs for low-income 3- and 4-year-olds.

Karoly and Bigelow (2005) estimated the economic benefits of universal access to 1 year of preschool education at age 4 in California. Based in part on cost-benefit findings from the CPC program and assuming a 70%-participation rate, the estimated return to California society at large was US\$2.62 per dollar invested. The most conservative estimates were about US\$2 return per dollar invested and the most liberal were about a US\$4 return per dollar invested.

A broader national analysis by Lynch (2007) used modified estimates from the CBA of the CPC program (Reynolds *et al.*, 2002) to generalize across states and in the country at large. It was estimated that by the year 2050, a high-quality targeted preschool program for 3- and 4-year-olds would cost US\$6479 (2007 dollars) per child and provide a return per tax dollar invested of US\$3.10 in government budget savings alone. For a universal access program, the return per tax dollar invested was estimated at US\$2.00 for government budget savings. Considering all societal benefits, the long-range annual benefit per tax dollar invested was estimated at US\$12.10 for a targeted program and US\$8.20 for a universal-access program.

The Effects of Full-Day Kindergarten

The effects of full-day kindergarten (FDK) compared to half-day kindergarten, are well documented. Many studies have examined achievement gains at the end of kindergarten and in the early-school grades. Aos *et al.* (2007) synthesized the results of 23 well-designed comparison-group studies on the effects of FDK on academic achievement. The studies were published since 1970 and excluded those in which FDK was part of other interventions or the selection into groups was undocumented or uncontrolled. The average effect size of FDK on achievement at the end of kindergarten was 0.18 standard deviations for all children and 0.17 for economically disadvantaged children. This is equivalent to a 2-month achievement gain.

This relatively small advantage largely disappeared by first grade and did not reemerge later. The average effect size was 0.01 at the end of first grade, 0.048 at second to third grade, and 0.00 at fourth and fifth grade. Although no formal cost-effectiveness studies have been done, the lack of long-term achievement gains would be expected to yield a benefit-cost ratio close to zero.

Class-Size Reductions

In the most extensive study of class-size reduction, Project Student Teacher Achievement Ratio (STAR) in Tennessee

experimentally investigated the impact of enrolment in class sizes limited to 13–17 students from kindergarten to third grade relative to enrolment in class sizes of 22–26 students with and without teacher aides. A total of over 6300 kindergarten students in 79 schools were included. Although 1 or more years in small classes was associated with higher achievement in the short-term, longer-term effects by eighth grade were found only for students with 3 or 4 years of reduced class sizes (Finn *et al.*, 2001). The 3-year group had median effect sizes of 0.17 standard deviations in grades 4 to 8. The four-year group had median effects sizes of 0.25. Relative to control groups, only low-income students with 3 or 4 years of small classes had higher rates of high school graduation (Finn *et al.*, 2005).

Based on Krueger (2003), Project STAR was found to have an economic return of US\$2.83 per dollar invested in the program. The source of this benefit is an increase of 0.2 standard deviations in test scores, which is associated with a 1.6% increase in adult earnings.

In a recent synthesis of 38 studies of class-size reductions, Aos *et al.* (2007) reported that reducing class sizes in kindergarten through second grade was more cost effective than reducing class sizes in third through sixth grades, middle school, or in high school. A reduction from 25 to 15 students in kindergarten, for example, was found to increase achievement by 0.19 standard deviations. Based on higher test scores, Aos *et al.* estimated that the economic return of small classes in kindergarten through second grade was US\$2.79 per dollar invested and US\$1.38 for small classes in third through sixth grades.

Additional evidence on small class sizes comes from the school-age program of the Child-Parent Centers, of which the main program element was a reduction in class sizes from 35–1 to 25–2 (teacher and aide) during grades 1–3. The school-age program also included instructional resources to promote reading and math achievement and family-support activities under the direction of a program coordinator. The economic return for 2 years of school-age intervention was US\$2.13 per dollar invested (Reynolds *et al.*, 2002). While this return per dollar invested is much lower than that of the CPC preschool program, it is within the range of that found for Tennessee STAR.

Other School-Age Programs and Practices

Given their implications for cost effectiveness, two additional ECD programs are covered. The Skills, Opportunities and Recognition (SOAR) Program, formerly the Seattle Social Development Project, is designed to promote social and emotional skills (Hawkins *et al.*, 1999). Starting in grade 1 and continuing to grade 6, the supplemental classroom-based program includes cooperative, developmentally appropriate teaching practices and optional

parent-education classes. Six years after the end of the program, participants had greater attachment to school, higher-achievement test scores, and lower rates of delinquency, and lower rates of alcohol misuse. The economic return was US\$3.14 per dollar invested.

Reading Recovery, an instructional tutoring program for first-grade students who are having difficulty learning to read, provides 30 min of one-on-one daily instruction with a teacher outside of the regular school class. Students in the bottom 20% in reading performance are enrolled in the program. In more than 30 studies that have been conducted of Reading Recovery (D'Agostino and Murphy, 2004), findings are generally consistent showing that the program increases participants' reading performance and helps close the gap with more typically performing students but the benefit is reduced substantially by fourth grade. Although formal CBAs have not been reported, Shanahan and Barr (1995) estimated that the program would, at best, be expected to return about 30% (or US\$0.30 per dollar invested) of its costs through reductions in special-education placement as a consequence of the short-term achievement effects.

PK-3 Programs

A key rationale for transition programs and practices in the early school-age years is that elementary schools play an important role in sustaining the benefits of early-childhood programs, and a continuation of programs into the primary grades will promote successful transitions. PK-3 programs are the most comprehensive approaches for enhancing transitions and promoting positive child development. Four extended early-childhood programs have shown evidence of positive effects on school achievement and child well-being above and beyond that of preschool participation. These are the Head Start/Follow through Program, National Head Start/Public School Transition Demonstration Project, Abecedarian Project, and the Child-Parent Center extended intervention program.

Only the CPC extended intervention program has been the subject of CBA. Compared to participation in less-extensive CPC services (0–3 years of intervention), the CPC extended program returned US\$9.13 per dollar invested through reduced remedial education and child maltreatment, lower juvenile arrest for violence, and higher levels of educational attainment. The US\$6.11 return associated with the extended program is the public benefit cost ratio – not the return exclusive of intangible crime victim savings. This is the public benefit cost ratio reported in the age 21 CBA. This is the public return per dollar invested exclusive of intangible crime victim savings.

The school-age program alone demonstrated a return of US\$2.13 per dollar invested primarily through participants requiring fewer remedial-education services (Reynolds *et al.*, 2002). The US\$1.66 return associated with the school-age program is the public benefit cost ratio – not

the return exclusive of intangible crime victim savings. This is the public benefit cost ratio reported in the age 21 CBA. This is the public return per dollar invested exclusive of intangible crime victim savings. The main components of the school-age services were reduced class sizes, family services, and instructional support to classrooms.

Summary of Cost Effectiveness

Table 5 summarizes the economic returns of ECD programs from preschool to third grade. Preschool for 4-year-olds has demonstrated the highest returns across many programs in different contexts, decades, service systems, and curricular philosophy. New evidence from state prekindergarten programs shows positive effects on school readiness.

Small classes in the early grades show positive returns from test-score gains (Krueger, 2003) but long-term effects were limited to low-income children with 3 or 4 years of enrolment (Finn *et al.*, 2005). Substantial reductions in class sizes (i.e., 25–15) are needed to meaningfully improve achievement (Aos *et al.*, 2007).

The CPC PK-3 intervention shows substantial benefits above and beyond earlier intervention. The SOAR social-skills program also has demonstrated positive long-term effects leading to cost-effectiveness. Neither Reading Recovery or full-day kindergarten have findings suggestive of cost effectiveness. Across many studies, the average effect of full-day kindergarten is not detectable at the end of kindergarten.

For completeness, CBA estimates are available for select prenatal and home-visitation programs in the first 3 years of life. The special supplemental nutrition program for Women, Infants, and Children (WIC) provides nutrition education, referrals to social services, and a variety of food supplements to low-income families. A meta-analysis of 15 studies in different states by Avruch and Cackley (1995) found that WIC participation was associated with a 25% reduction in the rate of low birth-weight births, which significantly reduced hospital costs paid by insurers in the first year of life. The economic return in savings was estimated to be US\$3.07 per dollar invested.

The Nurse-Family Partnership (Olds *et al.*, 1993) is an intensive nurse home-visitation program for young mothers having their first child. For the high-risk sample. (unmarried and low-income mothers having their first child), Karoly *et al.* (1998) it was found that participation from prenatal development to age 2 was associated with lower rates of criminal behavior for both mothers and target children, lower rates of substantiated child maltreatment, higher-earnings capacity for the mothers, and increased tax revenues projected into adulthood. The estimated economic return was US\$5.06 for every dollar invested. For the lower-risk sample, the economic return

Table 5 Cost-effectiveness estimates for early-childhood programs, birth to third grade

Development stage	Source	Focus	Location	2007 Dollars ^a			
				Benefits	Costs	B–C	Ratio
<i>Birth to age 3</i>							
WIC ^b	Avruch and Cackley (1995)	Targeted	National	1 206	393	813	3.07:1
NFP, Low SES	Glazner <i>et al.</i> (2004)	Targeted	Elmira, NY	83 850	16 727	67 123	5.01:1
NFP, Higher SES	Glazner <i>et al.</i> (2004)	Targeted	Elmira, NY	25 317	16 727	8 590	1.51:1
<i>Preschool</i>							
Child–Parent Centers	Reynolds <i>et al.</i> (2002)	Targeted	20 Chicago sites	86 401	8 512	77 889	10.15:1
Perry Preschool	Belfield <i>et al.</i> (2006)	Targeted	1 Ypsilanti site	294 772	18 261	276 511	16.14:1
Abecedarian ^c	Barnett and Masse (2007)	Targeted	1 NC site	182 422	73 159	109 263	2.49:1
RAND Study of Preschool in CA	Karoly <i>et al.</i> (2005)	Universal	State of CA	12 818	4 889	7 929	2.62:1
National Pre-K synthesis for 2050 ^d	Lynch (2007)	Targeted	National	20 085	6 479	13 606	3.10:1
Synthesis study	Lynch (2007)	Universal	National	12 958	6 479	64 479	2.00:1
Kindergarten	Aos <i>et al.</i> (2004)	Targeted	58 programs	19 826	8 415	11 411	2.36:1
Full-Day K synthesis ^{e,f}	Aos <i>et al.</i> (2007)	Universal	23 programs	0	2 685	–2 685	0:1
<i>School-Age</i>							
Tennessee STAR (class size reduction, K-3)	Krueger (2003)	Universal	79 schools	27 561	9 744	17 817	2.83:1
Synthesis of reduced class sizes, K-2 ^{e,g}	Aos <i>et al.</i> (2007)	Universal	38 studies	6 847	2 454	4 393	2.79:1
Synthesis of reduced class sizes, grade 3-6 ^{e,g}	Aos <i>et al.</i> (2007)	Universal	38 studies	3 387	2 454	933	1.38:1
Child–Parent Centers School-Age program	Reynolds <i>et al.</i> (2002)	Targeted	20 Chicago sites	8 089	3 792	4 297	2.13:1
Reading Recovery ^e	Shanahan and Barr (1995)	Targeted	General	1 679	5 596	–3 151	0.30:1
Skills, Opportunities and Recognition	Aos <i>et al.</i> (2004)	Universal	Seattle schools	16 256	5 172	11 084	3.14:1
<i>PK-3 Intervention</i>							
Child-Parent Centers Extended program	Reynolds <i>et al.</i> (2002)	Targeted	20 Chicago sites	47 161	5 163	41 998	9.13:1

^aAll original estimates are converted to 2007 dollars using the Consumer Price Index for All Urban Consumers (CPI-U).

^bEstimates are based on a meta-analysis of studies investigating the effects of WIC.

^cThe cost for the Abecedarian Program represents the total costs of the intervention. Apply the marginal cost for the Abecedarian Program results in a return of \$3.64 for every dollar invested in the program.

^dEstimates for Lynch's (2007) synthesis of target and universal preschool represent per pupil program costs and associated annual government budget benefits. Total accrued benefits to government, the general public, and program participants and their parents relative to costs are \$12.10:1 and \$8.20:1 for the targeted and universal programs, respectively.

^eEstimates are not based on formal cost-benefit analyses.

^fThe cost of full-day kindergarten is relative to the cost of half-day kindergarten in Washington State.

^gEstimates from syntheses of reduced class sizes assume a reduction from 25 to 15 pupils per class.

was US\$1.10 per dollar invested. A more recent analysis of the Nurse-Family Partnership by Glazner *et al.* (2004) showed an overall return of US\$3.93 for every dollar invested in the Elmira, NY program.

In summary, findings show that a variety of ECD investments are associated with positive economic returns. Preschool programs for 3- and 4-year-olds have had the most research, and generally show the highest returns. This is illustrated in Figure 2 as the economic return per dollar invested for the accumulated research as a function of the age of entry into intervention. Age 0 corresponds with prenatal development (WIC and NFP). Early school-age programs show positive returns of close to US

\$3 dollars per dollar invested (Tennessee STAR). Combinations of programs such as PK-3 are not represented in Figure 2. These are other synergies that warrant further investigation. Nevertheless, cost effectiveness is one of many criteria for prioritizing investments. Social importance, feasibility, and capacity for sustainability also are important to consider in policymaking.

Key Principles of Effectiveness of ECD Programs

Findings summarized in this review indicate that greater investments in high-quality preschool and school-transition

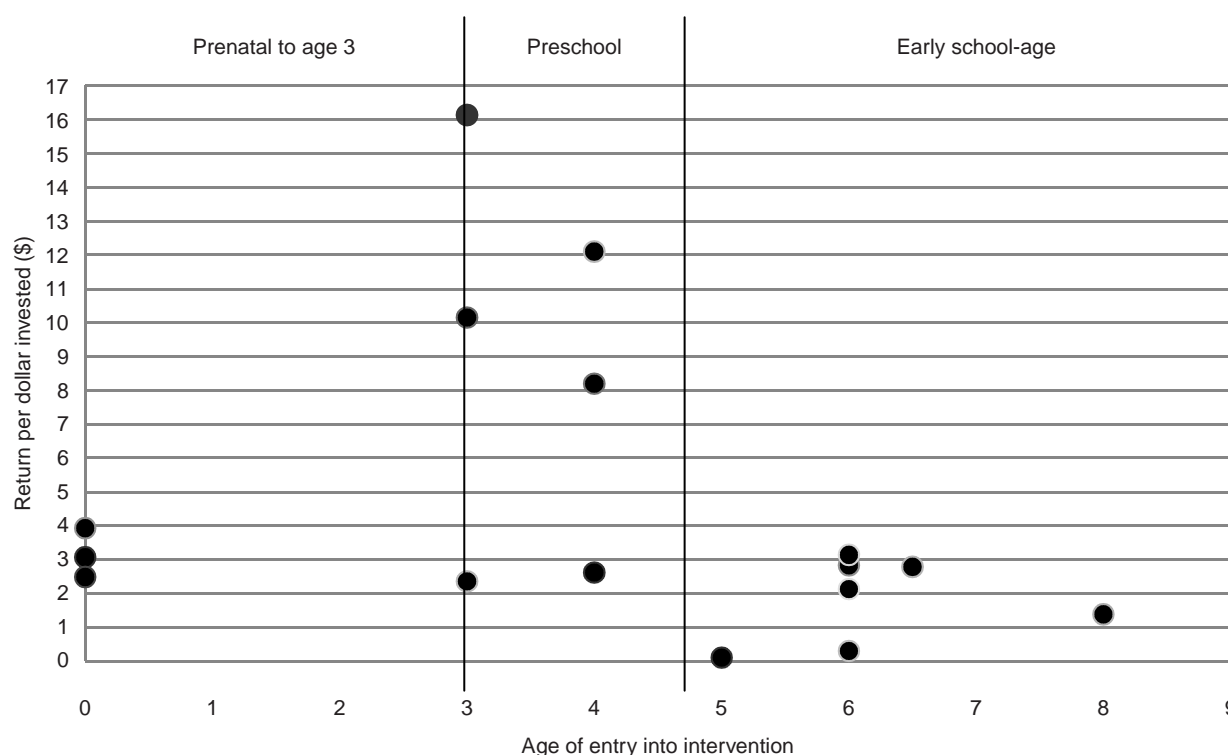


Figure 2 Return per dollar invested by age of entry into intervention. The median return per dollar invested in presented in the figure is US\$2.81. The average return is US\$4.61 per dollar invested.

programs are warranted. Since relatively large percentages of children do not enrol in center-based preschool programs (Barnett *et al.*, 2007), and the quality of services that many receive is not high, the ECD programs summarized in the article provide effective models to be used in the design of coordinated early childhood systems. As shown in **Figure 1**, attention to the causal mechanisms of change can also strengthen program effectiveness. Research on three programs reviewed (ABC, PPP, and CPC) and many others suggest five major principles that can enhance the effectiveness of ECD programs and to increase long-term economic benefits.

The first principle is that a coordinated system of early education is in place beginning at age 3 and continuing to the early-school grades. Program implementation within a single administrative system in partnerships with communities can promote stability in children's learning environment which can provide smooth transitions from preschool to kindergarten and from kindergarten to the early grades. The three major programs we reviewed were either housed in elementary schools or provided continuity of services between preschool and formal schooling. This is a first-decade strategy of promoting child development. Today, most preschool programs are not integrated within public schools and children usually change schools more than once by the early grades. The CPC program, for example, is an established program in the third-largest school system in the nation. Findings from the CBA of a complete cohort

of CPC participants give a good indication of the size of effects that could be possible in public schools.

A second major principle of effective ECD programs is that the teaching staff should be trained and compensated well, preferably with earned bachelor's degrees, certification in early childhood, and competitive salaries. It is no coincidence that the three major programs reviewed in the article followed this principle. Being located in public schools, the Perry and CPC programs were implemented by teachers with at least bachelor's degrees and appropriate certification in early childhood. They were paid on the public-school salary scale, and Perry teachers received a 10% bonus for working in the program. In the Abecedarian program, teachers were compensated at level that was highly competitive with public schools. Turnover was low in all three programs. In most other early education programs, from child care to Head Start, staff did not have this level of education, training, and compensation, and turnover was significantly higher.

Third, educational content should be responsive to all of children's learning needs but special emphasis should be given to cognitive and school-readiness skills through a structured but diverse set of learning activities. All of the cost-effective programs reviewed had a strong emphasis on the development of cognitive and language skills necessary to do well in school within a responsive learning environment. Child to staff ratios of less than 9 to 1 in preschool helped as well. The curriculum appeared to be

less important since the programs spanned from Perry's child-initiated approach to Chicago's blended, teacher-directed approach. Extrapolating these findings, preschool and other social programs are more likely to have enduring effects if they provide services that are intensive and are dedicated to the enhancement of educational and social skills.

A fourth principle of effectiveness is that comprehensive family services should be provided to meet the different needs of children. As child development programs, preschool, kindergarten, and school-age programs must be tailored to family circumstances and thus provide opportunities for positive learning experiences in school and at home. Those with special needs or who are most at risk benefit from intensive and comprehensive services. Each of the cost-effective preschool programs discussed in the article provided family services. Abecedarian provided medical and nutritional services. The Perry preschool had weekly home visits by teachers. In the CPC program, parent involvement is more intensive as each center has a parent-resource room run by a certified teacher and provides school-community outreach. Parents' own educational and personal development are important program goals.

Finally, greater commitment to ongoing evaluations of effectiveness and cost effectiveness is needed. Even today, CBAs are rarely conducted. This state of affairs limits full consideration of the effects of alternative programs. Paramount in conducting CBAs is the availability of longitudinal studies of programs for children and youth. These studies are more likely to accurately assess the total impact of program participation. In addition, more studies are needed that address the differential effects of participation across a range of child, family, and program attributes.

See also: Home-Based and Institutional Early-Childhood Education and Care Services; Investing in Early Childhood Education and Care: Some Policy Implications; Investing in Early Childhood Education and Care: The Economic Case.

Bibliography

- Aos, S., Lieb, R., Mayfield, J., Miller, M., and Pennucci, A. (2004). *Benefits and Costs of Prevention and Early Intervention Programs for Youth: Technical Appendix*. Olympia, WA: Washington State Institute for Public Policy.
- Aos, S., Miller, M., and Mayfield, J. (2007). *Benefits and Costs of K-12 Education Policies: Evidence-Based Effects of Class Size Reductions and Full-Day Kindergarten*. Olympia, WA: Washington State Institute for Public Policy.
- Avruch, S. and Cackley, A. P. (1995). Savings achieved by giving WIC benefits to women prenatally. *Public Health Reports* **110**, 27-34.
- Barnett, W. S., Hustedt, J. T., Friedman, A. H., Boyd, J., and Ainsworth, P. (2007). *The 2006 State Preschool Yearbook*. New Brunswick, NJ: National Institute of Early Education Research.
- Bogard, K. and Takanishi, R. (2005). PK-3: An aligned and coordinated approach to education for children 3 to 8 years old. *Social Policy Report*, XIX, No. III. Washington: Society for Research in Child Development.
- Campbell, F. A. and Ramey, C. T. (1995). Cognitive and school outcomes for high risk African-American students at middle adolescence: Positive effects of early intervention. *American Educational Research Journal* **32**, 743-772.
- D'Agostino, J. V. and Murphy, J. A. (2004). A meta-analysis of reading recovery in United States schools. *Educational Evaluation and Policy Analysis* **26**, 23-38.
- Finn, J. D., Gerber, S. B., Achilles, C. M., and Boyd-Zaharias (2001). The enduring effects of small classes. *Teacher College Record* **103**, 145-183.
- Finn, J. D., Gerber, S. B., and Boyd-Zaharias, J. (2005). Small classes in the early grades: Academic achievement and graduation from high school. *Journal of Educational Psychology* **97**, 214-223.
- Glazner, J., Bondy, J., Luckey, D., and Olds, D. (2004). *Effects of the Nurse Family Partnership on Government Expenditures for Vulnerable First-Time Mothers and Their Children in Elmira, New York, Memphis, Tennessee, and Denver, Colorado*. Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services.
- Hawkins, J. D., Catalano, R. F., Kosterman, R., Abbott, R., and Hill, K. G. (1999). Preventing adolescent health-risk behaviors by strengthening protection during childhood. *Archives of Pediatrics and Adolescent Medicine* **153**, 226-234.
- Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics* **54**, 3-56.
- Karoly, L. A. and Bigelow, J. H. (2005). *The Economics of Investing in Universal Preschool Education in California*. Santa Monica, CA: RAND.
- Karoly, L. A., Greenwood, P. W., Everingham, S. M. S., et al. (1998). *Investing in Our Children: What We Know and Don't Know about the Costs and Benefits of Early Childhood Interventions*. Santa Monica, CA: RAND.
- Karoly, L. A., Kilburn, M. R., and Cannon, J. S. (2005). *Early Childhood Interventions: Proven Results, Future Promise*. Santa Monica, CA: RAND.
- Krueger, A. B. (2003). Economic consideration and class size. *Economic Journal* **113**, F34-F63.
- Levin, H. M. and McEwan, P. J. (2001). *Cost-Effectiveness Analysis: Methods and Applications*, 2nd edn. Thousand Oaks, CA: Sage.
- Lynch, R. G. (2004). *Exceptional Returns: Economic, Fiscal, and Social Benefits of Investing in Early Childhood Development*. Washington, DC: Economic Policy Institute.
- Lynch, R. G. (2007). Enriching children, enriching the nation: Public investment in high quality prekindergarten. Washington, DC: economic Policy Institute.
- Olds, D. L., Henderson, C. R., Phelps, C., Kitzman, H., and Hanks, C. (1993). Effects of prenatal and infancy nurse home visitation on government spending. *Medical Care* **31**, 155-174.
- Reynolds, A. J. (2000). *Success in Early Intervention: The Chicago Child-Parent Centers*. Lincoln, NE: University of Nebraska Press.
- Reynolds, A. J., Ou, S., and Topitzes, J. (2004). Paths of effects of early childhood intervention on educational attainment and juvenile arrest: A confirmatory analysis of the Chicago child-parent centers. *Child Development* **75**, 1299-1328.
- Reynolds, A. J. and Temple, J. A. (2008). Cost-effective early childhood development programs from preschool to third grade. *Annual Review of Clinical Psychology* **4**, 109-139.
- Reynolds, A. J., Temple, J. A., Robertson, D. L., and Mann, E. A. (2002). Age 21 cost-benefit analysis of the title I Chicago child-parent centers. *Educational Evaluation and Policy Analysis* **24**, 267-303.
- Reynolds, A. J., Wang, M. C., and Walberg, H. J. (eds.) (2003). *Early Childhood Programs for a New Century*. Washington, DC: CWLA Press.
- Schweinhart, L. J., Barnes, H. V., and Weikart, D. P. (1993). *Significant Benefits: The High-Scope Perry Preschool Study through Age 27*. Ypsilanti, MI: High/Scope Press.
- Shanahan, T. and Barr, R. (1995). Reading recovery: An independent evaluation of the effects of an early instructional intervention for at-risk learners. *Reading Research Quarterly* **30**, 958-996.
- Zigler, E. F., Gilliam, W. S., and Jones, S. (eds.) (2006). *A Vision for Universal Preschool Education*. Cambridge, MA: Cambridge University Press.

Further Reading

- Barnett, W. S. (1985). Benefit–cost analysis of the Perry preschool program and its policy implications. *Educational Evaluation and Policy Analysis* **7**, 333–342.
- Barnett, S. W. (1996). *Lives in the Balance: Age-27 Benefit–Cost Analysis of the High/Scope Perry Preschool Program*. Ypsilanti, MI: The High/Scope Press.
- Belfield, C. R., Nores, M., Barnett, S., and Schweinhart, L. (2006). The high/scope Perry preschool program: Cost–benefit analysis using data from the age-40 follow-up. *Journal of Human Resources* **41**, 162–190.
- Boardman, A. E., Greenberg, D. H., Vining, A. R., and Weimer, D. L. (2006). *Cost–Benefit Analysis: Concepts and Practice*, 3rd edn. Upper Saddle River, NJ: Pearson Education.
- Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives* **15**, 213–238.
- Hummel-Rossi, B. and Ashdown, J. (2002). The state of cost-benefit and cost-effectiveness analyses in education. *Review of Educational Research* **72**, 1–30.
- Masse, L. N. and Barnett, W. S. (2002). *A Benefit Cost Analysis of the Abecedarian Early Childhood Intervention*. New Brunswick, NJ: National Institute for Early Education Research.
- Schweinhart, L. J., Montie, J., Xiang, Z., et al. (2005). *Lifetime Effects: The High/Scope Perry Preschool Study through Age 40*. Ypsilanti, MI: High/Scope Press.
- Temple, J. A. and Reynolds, A. J. (2007). Benefits and costs of investment in preschool education: Evidence from the child–parent centers and related programs. *Economics of Education Review* **26**, 126–144.
- Temple, J. A., Reynolds, A. J., and White, B. A. (2007). *Age 21 Cost–Benefit Analysis of the Title I Chicago Child–Parent Centers: Technical Report*. Minneapolis, MN: University of Minnesota. <http://cehd.umn.edu/icd/clis/Publication.htm> (accessed August 2009).

Early Childhood in Post-Modern Cultures: Thoughts and Some Concerns

P Gammage

© 2010 Elsevier Ltd. All rights reserved.

Into my heart an air that kills
From you fair country blows:
What are those blue remembered hills,
What spires, what farms are those?

That is the land of lost content,
I see it shining plain,
The happy highways where I went
And cannot come again.

A E Housman

The Overriding Influences

Stark Contrasts

If one takes an ecological view of childhood and, having placed the child at the center of a diagrammatic map or target, looks first at the outer pressures on our central concern, children, there are clearly two overriding contextual features that dominate everything throughout the world; and though they do not sharply divide, but merge into one another, these broad health, and socioeconomic extremes may be represented fairly starkly. All the intermediate and close immediate pressures on children are shaped, result from, or are affected by them.

About 9–10% of children are born each year into somewhere in the relatively rich 30 Organization for Economic Cooperation and Development (OECD) countries, and despite the diseases of affluence (diabetes and heart disease, caused largely by underexercise and overeating fatty foods), they will live for about 80 or more years and will normally have opportunities for education up to and beyond university level. (In the OECD, there are 30 rich countries as members.) The contrast is with that of the remaining 90%, who are born into extreme poverty, sometimes poor sanitation and certainly inadequate diets and into cultures replete with diseases such as acquired immune deficiency syndrome (AIDS), malaria, polio, and tuberculosis. War, genocide, and famine will often also be their lot. Such children will have, in extreme cases (such as Rwanda in Africa), a life expectancy of half that of their rich cousins, that is, not much more than 40 years, and their education will be sporadic, tenuous, serendipitous, or nonexistent.

Of course, there are many variations and many within-group differences. Cuba, a relatively impoverished and embargoed country, provides tolerably well for the care

and education of its children, possibly more consistently than does a generally extremely affluent Australia. The latter has tended to rely on both state and central government provision mixed with a substantial commitment to private enterprise. Approximately 70% of all Australian child care was (in 2008) private, corporate, or requiring some form of substantial financial input from parents. Consequently, some of the poor children (and there are approximately 15% such in Australia, as there are in Canada and the UK; OECD, 2001; UNICEF, 2005) fall through the net and do not receive the support they deserve. This is especially true of Australian aboriginal children, whose conditions are more often similar to those of a Third-World country (Australian Institute of Health and Welfare, 2008). A major initiative of the Australian federal government (in cooperation with the states and territories, 2008/2009) is to improve early childhood provision, to assay its quality, and to regulate and guide it better.

Cuba, despite many financial difficulties, has within its overarching, official approach, a system of child-care/preschool institutions called Children's Circles and these are available from the age of about 12 months. For those who cannot benefit, or who are in rural areas where it is difficult to service completely, the education authorities have a system of trained coaches working within the community, who not only support and stimulate the family's approaches to the child's development and learning but also provide the equivalent of toy libraries and materials as well as advice and health support. There are no figures available for proportions of Cuba's gross domestic product (GDP) spent on such a system and it clearly has its gaps, but it is a basic right of the people and does not depend upon income (UNESCO, 2007). Moreover, as in many countries nowadays, Cuba has a ministry of education that embraces child care and sees it as an integral part of education, much as do the Tasmanian, South Australian, and Victorian states in 2009.

Thus, the broad outer circle of any contextual/ecological diagram, which places children at the center, would have to be one which takes clear account of affluence, or poverty, and how these are dealt with politically. A rich country can still display a relative disregard for its youngest and most vulnerable, and may especially, disregard the impact of seeming choice on the minority poor in a thriving market economy, whereas a poor country may have ingenious ways of supporting its young and their families and see systems of support as a major political imperative, such that no

child is left uncared for within the limits of its finance and power. Choice is a favored term in conservative political systems, which tend to regard education and child care as simple commodities, which can be bought and sold in the market place. According to such a view, one has the choice, to buy or not to buy. The politics of it all are themselves embedded in ideology and often of themselves provide the onlooker with a key to style and type of commitment, with (mainly) rich socialist countries viewing early childhood education and care (ECEC: the OECD term) as a definite part of tax-based public provision and rich conservative countries tending to emphasize the importance of family responsibility and free-market approaches. Even in Norway and Sweden, which have a high reputation for social and educational care there are mixed-market (though well-regulated) approaches to ECEC.

Intermediate Influences: The Meso-System

Demography, Families, and Change

The context of childhood is changing. In the affluent world, children are clearly a decreasing commodity. For instance, in the 30 rich countries of OECD, the total fertility rate (TFR) is, for the most part, well below replacement level (replacement being approximately 2.2 children per female child-bearing life). Indeed, in Australia and England the TFR hovers around 1.7 and in Europe as a whole is currently (2006 figures) about 1.46. Moreover, the TFR is hard to raise and, partly as a consequence of much later first parturition (now at about age 29.9 years per Western female), over one-third of the child-rearing age is statistically largely barren. This means that, in effect, many families have only one child especially in a family where both parents work. It also means that many children are in effect only, lonely children and it makes imperative the provision of good systems of community support, so that the child can have what might be appropriately termed pseudo-siblings; put simply, the chance to develop, share, and play with companions (Gammage, 2008).

Such a change is partly the impact of contraception and the efficiency and availability of the pill since the early 1960s, which in some ways has given women more power over their own bodies and changed the nature of planned parenthood. It is also partly the impact of economic change and the globalization of the economies and of the drive toward house purchase and technological advance generally. There are many interrelated causes. However, whatever these are (and the birth-rate is also low in Catholic countries: e.g., Italy 1.55 and Portugal 1.46), they have helped change the nature of childhood itself, since childhood as described within each culture is largely a cultural artifact. With both parents at work, it means that many children have to be involved in systems of care, be it Family

Day Care, or being looked after by a friend or relation, or placed in a formal child-care institution with qualified carers and teachers. Moreover, in a world of globalized expectation (via commonly accessed media), mobility of families and high mortgages, this means that many women feel the financial imperative to return to work as soon as possible after the birth of their child to help secure the family's finances. Under such circumstances, child care becomes a necessity, which is no longer a luxury afforded only by the upper classes; it becomes an integral part of state provision. Its vitality and coverage is as important as the provision of elementary or primary education, with which it now merges practically and philosophically.

Within this contextual, middle category of influence come divorce and the changing nature of families themselves. Divorce rates currently run at about 40% plus in much of Europe and appear in Australia, as well. (Eurostat, 2008) Thus, nowadays, many families are remade or restructured at certain times and the art (diplomacy and psychology) of step-parenting becomes crucial to the happiness and welfare of the child.

Day care, family centers, child-care institutions, kindergartens, *crèches* and nurseries, and many other terms are used for the various provisions subsumed under the term ECEC (the OECD and the United National Educational, Scientific and Cultural Organization (UNESCO) term for care and education from birth to about 8 years) and such institutions help in supporting revised family structures, working parents, patterns of shift work, economic needs, and remade families. They have, in the last decades of the twentieth century and the first of the twenty first, quickly become an essential part of a normal modern society.

There are other demographic effects, too. These are that in current affluent societies there are greater cadres of the elderly. These older people have to be looked after by the declining financial (tax) and resource (time) base of those (apparently now getting fewer) in work. Partly this may be compensated for by raising the age at which people retire – a relatively recent, but common, ploy in political systems in Europe and Australia. Such an approach may not be bad, since it recognizes greater longevity (and fitness for purpose) among the elderly and restores some status and dignity to the values and purposes of that group.

Therefore, families have altered and what matters is less the structure of the family and more its function; and how children may be best loved and socialized in a secure environment becomes a fundamental political issue. There will be families of all sorts and the child-care institutions need to be attuned to support and community involvement in a way that takes this into account, without stigma or prejudice perhaps, without excessive expense on the part of the individual. The whole society is responsible for the shift in mores, needs, and institutions. It seems that divorce and separation are inevitable, a part of adult choice

and of our mobile, longer-living, modern, postnatural societies (Giddens, 1999); and family change has to be viewed constructively, for the sake of both the children and the mental health of the adults involved. Surely, therefore (or so runs the argument), we need to find collective societal solutions.

The Close Inner Pressures

Technology and the Media

Children are increasingly socialized by the media. While this clearly provides for greater democratization of knowledge and the accessibility of sophisticated comment (at best), it can also result in the domination of information and attitudes by forces that are sometimes inimical to sound psychological and social health. Violence is a particular issue on our television (TV) screens and with many video games, as is the assumed consensus of beliefs and the dumbing down of complexity and difference. Coarseness of language and populism can provide a poor basis for a young child's early development. However, popular culture is big business and music and iconic performances by chosen stars are manipulated regularly for the benefit of marketization, almost regardless of merit. The technology of the home and the classroom is increasing, with interactive white boards in kindergartens and the attempts (in USA) to encourage lap-ware (the use of laptop software for soothing and educating babies). TV represents an especially potent force, in that the hours spent in front of the screen are probably now well in excess of the hours spent in child care, or with mother or father. (The *New Media* publication suggested, almost a decade ago, that about 20-plus hours viewing per week appeared to be the average for a South Australian 3-year-old.)

Mobile phones, game machines, computers, and personal players are common to many 5-year-olds and the interface with technology has a commanding hold on the developing brain, even if one is legitimately worried about the quality and amount of genuine interaction thus encouraged.

The Center

Early Brain Development

In an earlier publication, the author has discussed something akin to the following, and subsequent, more recent research has tended to support such a position, so the author reiterates it. It is especially important that we do not take too deterministic a view of early child development, although it is vital. Remediation and change are possible at almost any stage of human life. The disciplines of psychology and psychiatry are to some extent built on such assumptions.

There are somewhat contradictory views taken about the importance of what McCain and Mustard (1999) referred to as critical periods in the development of the brain. Research with small children is extremely difficult in this domain; and although there is some evidence of critical periods of development for some animals other than humankind, there is little that is specific to our children: for example, those for critical periods are McCain and Mustard (1999) and Wynder, among others. Wynder said, "It is the consensus of the participants that a 'critical period' exists during which the synapses of the dendrites are most ready for appropriate stimulation, be it through words, music, love, touch, or caring. If these synapses are not so stimulated early, they may never fully develop" (Wynder, 1998: 166). There are those less enthusiastic, however. "I suggest that the importance of timing lies not within a set of age parameters but rather in the match between experiences provided, the child's developmental status, and the child's need or readiness to learn a particular skill or concept" (Bailey, 2002: 2–3) (vide Gammage, 2003: 340).

What we clearly do know and where there is much consensus is that the child's brain is immensely plastic. During the first 3 years or so of life, some trillions of connections are gradually established, forming the maps of response, coordination, and information processing. Constant change is occurring as the networks grow in parallel and in sophistication through the infant's daily observation and curiosity. Processes of selective amplification occur in direct relationship to the frequency, intensity, and salience of the stimuli. Some networks are pruned and adapted. All these are embedded in the quality of attachment and consistency. Indeed the trinity of care might well be described as one of attachment (love), consistency (prediction and reliability), and boundaries (knowing what counts). These are the difficult areas that all parents have to chart and recognize in the upbringing of any small child. They are affected by culture and by personal disposition; and sensitivity to and awareness of them are not necessarily simple correlates of pregnancy. Parents often need support and knowledge to make the most of their own inclinations and attributes. Some parents might be considered almost lethal in their styles of child rearing, as many psychologists could point out. Such comment, if made publicly, is commonly frowned upon in a democracy and regarded as unsupportively critical and politically incorrect. However, human sciences casework would testify to its accuracy. There is such a thing as bad, ignorant, or inadequate parenting and we do a disservice to future generations if we fail to recognize that and plan accordingly. Some of recent policy changes in England and Australia can be laid directly at the door of political anxiety about what to do if a parent really wants to harm his/her child. It is doubtful whether any system is so good that it could assure that no child would ever be harmed. Scapegoats in the public services are common, but the problem lies very often in

the cyclic nature of psychological dysfunction within families themselves.

Boys and girls tend to excel at somewhat different tasks and though the density of their brains differs in some respects, no one has adequately explained the difference between the sexes. Recently, however, work on epilepsy has shown markedly different synapse density in actual synapse connection and clusters in different areas of the brains of men and women. It may be that such differences give some explanation as to why men and women show different levels of skill at different tasks (New Scientist, 2008). A critical feature seems, however, to be the overlay of cultural expectation as well.

Research on the development of the brain, while not offering what some people may claim in terms of explicit programs and curricula for the very young, or for different genders, does suggest that money is spent at the wrong end of the developmental curve. Put crudely, we tend to spend downstream, rather than upstream. Upstream spending would be a better investment and of great benefit to society. There is no doubt that what goes on in early child development is highly predictive of later adult behavior and there is considerable evidence from cohort studies to support this view. Such a perspective also tends to confirm what many teachers and carers have averred for many years, that high-quality social interaction and play are the absolute best basis for secure learning. Elkind (2008) sees play as the way in which dysfunctionality is countered and useful schema interpreted and established. He suggests that play leads to healthier, happier adults. Dysfunctionality would seem to be internalized young and has major implications for society both then and later, as many a psychiatrist and criminologist would testify and as many cohort studies clearly demonstrate. Perhaps how the 3-year-old behaves in the long-day care center tells one a great deal about how (later) the 23-year-old will behave in an adult environment (Silva and Stanton, 1996)! Yet, we have to be careful. It is all too easy to fit the Western developmental expectation coat upon every child, regardless of his/her experiences, perceptions, and his/her needs (Yelland and Kilderry, 2005). Prediction has its dangers.

The Child, Happiness, and Early Pursuits

The message that the early years are vital has now caught the attention of most politicians and policymakers. After all, few could ignore the two-stage study by the OECD, *Starting Strong* (OECD, 2001, 2006), or the constant ECEC papers provided by UNESCO on different country values and approaches. However, one suspects that some policymakers have entertained only certain parts of the message. They now know that learning is crucial from birth and they therefore see the child as an appropriate

investment for the future. They tend, however, to see the provision of early formal curricula as the best approach to achieving this. Regardless of the demonstrably poor match between such studies as the Programme for International Student Assessment (PISA) and the formal ECEC curriculum (the Finns do very well in PISA without formal ECEC curricula), they avidly pursue a diet of watered-down school approaches; and fail to recognize the here and now of a small child's life, the vital place of joy in motivation, or the need for well-trained staff (quality comes from quality staff – an almost universal research finding). Such policymakers adore measurement, even though the measures may not tell us much, or even when the measures drive and delimit the curriculum such that children are bored and joyless (Willis, 2007). Workers with children know that often what is difficult to measure is that which really counts in the child's progress to a healthy adulthood. Well-being, persistence, humor, empathy, engagement, and locus of control are all difficult to measure, yet form the vital core of children's drama, play, listening, sharing, and social activities and may be equally as powerful as success in literacy or numeracy. Even in early literacy, we now know from detailed studies that the emotional bonds are often key to the success of the child.

It cannot be emphasized enough that children need variation, individual attention, systematic, yet noncompetitive atmospheres, and ample opportunity to choose and to persist. Success lies in the fostering of dispositions and attitudes which lead to autonomy, yet sensitivity, creativity, yet responsibility. The true curriculum for small children is, in its core, that of relationships and democracy. These, as the Swedish policy documents aver, are the basis of a just and humane society (vide Bennett, 2007).

Politics and Policies: Freedom, Risk, Accountability, and Control

Childhood and being a child are in some respects under siege in many modern societies, since children are viewed (often at best) as little investments waiting to happen; at least, that is the rationale commonly offered by politicians. Much lip service is paid to children's rights, but few children are asked what they would really like to experience. They are done unto with all the zeal of religious reformers. The child must surely have the opportunity to be as well as become (Moss and Petrie, 2002; Morrill, 2003; Willis, 2007). Two compulsory institutions commonly exist in modern societies: school (which is starting ever-earlier in the child's life) and prison. Both are implicated in the withdrawal (necessarily, or otherwise) of human rights.

Although we know much more about early childhood and the developing brain, one sometimes wonders whether we should be more concerned over whether the brains of adults are actually equal to the task. As Erasmus said, "a

person is not born human, but becomes one.” What are we doing to make sure that the impact of the environment and opportunities we provide are equal to the task (OECD/CERI, 2007)?

There are quite contrasting perspectives among policymakers. While we should be clear that politics shape policies, this is not an easily discerned, nor necessarily rational, tidy, contiguous process. Moreover, ideology is frequently employed as theory, even at the very highest level. People know how they think children best learn, or are best brought up. They know it vehemently, with fervor and conviction. Research is usually not allowed to get in the way of such expressed convictions. In most situations, governments try to establish the facts, so that they have knowledge and the best up-to-date information before they act and before legislation. In the ensuing melee, however, they usually act according to pressures, lobby groups, personal beliefs, and values. In reality, while these seem to be very different perspectives . . . (cool information gathering, as opposed to strong conviction), most policy decisions have to try to accommodate all of them.

In all this, the media have increasing power, such that their moguls are increasingly courted and heeded by those elected to power. There is, in the public service, in the media and among politicians, a prevalent use of weasel words. These are words that summon up much spirit, but often mean relatively little, or are especially difficult to define. They can be insinuated into the discussion and work their magic helpfully, it seems. Thus, some politicians talk of curriculum delivery, spreading perhaps the image of simplicity, of Postman Pat going about his cheerful rounds with a great, easily described, and attractively shaped parcel of knowledge. Reality is not like that and each child requires curricula that match and mesh with the child’s needs and his/her stage of thinking. The parcel, however, fits the folklore view of teaching as filling up children with vital knowledge and, as an image, has much power and salience with the press. Many such words abound, partly, perhaps, because of the current preoccupation with management and leadership theory, or of the sound-bite approach to successful marketing. The current political vogue for target setting, spawns many such, from performance managing to children’s well-being, from curriculum framework (so as not to frighten those who are child-centered in approach?) to essential learnings. A lifetime in the field would yield some interesting words and fashions that last relatively short periods of time, yet have power and collective strength in the way they are used to persuade children and teachers that this or that is best for them. They are usually constructed to mesh well with populist views of education and care.

Closely following the media expositions of childhood (especially those of damaged or aberrant childhood) and its context comes the legal profession. They are above politics in any sane democracy, but there are

many interstices in which they can act for, defend, and indeed sometimes profit from, the concerns of the public. This has given rise, in some countries (there are marked differences), to what might be described as the culture of blame; accountability, risk management, and the overreliance on child outcomes (school achievement) being directly contiguous with teacher skill and effort. Accountability is commonly defined as being responsible for and able to explain one’s actions. In major parts of the Western world (broadly the English-speaking ones), principals, headteachers, and others in responsible positions have to be extremely aware of the ways that litigation can be employed destructively against a college, school, or center. This is not helped by the attitude of many politicians and policymakers who seem to think that there is clear connective tissue between the actions of a teacher and the intellectual outcomes for the child. While it is important that any professional should be accountable for his/her actions and certainly able to explain them rationally, the subtlety, the timing, and the myriad intervening variables are so often not clearly perceived in such simple equations. Moreover, the responsibility of the family and of others can so easily be lost in any ensuing discussion, as can the responsibility of those who talk of local management in education and yet support rigid central control, or inappropriate systems of assessment. Professionals make judgments, using the best knowledge available at the time – sometimes those judgments are wrong. The role of a surgeon, general practitioner, teacher, psychiatrist, and many others is not necessarily helped by the specter of litigation. In the beginning, somewhat punitive, approaches taken by the Office for Standards in Education, Children’s Services and Skills (Ofsted; the inspectorate system) in England in the early 1990s, contrived to make being accountable take on a meaning that may have directly led to the illness of leaders and even in some cases, suicide.

Risk is in fact a biological necessity for the human organism to progress and develop. Challenge is important. Small triumphs and occasional failures are the weft and warp of the growing child’s life. The question is, how much risk; and is it calculated sensibly and sensitively? Schools and children’s centers provide, in effect, substitute homes and pseudo-siblings for their charges. They also provide role models and control, education and care, and many other opportunities to develop and learn. In all that, there must be at times measures of calculated risk. The Outward Bound movement of the late Kurt Hahn would have some difficulty with the curtailing of enterprise and calculated risk, as would the international colleges like those in Singapore, Canada, and UK. Different countries clearly have different approaches to risk management. Some systems, typically those of the Nordic countries and Italy and Portugal, use it in what might be termed light-touch approaches, whereby children are still allowed to climb trees, light candles, play in the snow, and

do things which require care, self-control, and some deployment of good sense. Countries such as the UK, USA, and Australia are nowadays much more cautious. Even traditional playground apparatus, such as seesaws, being frowned upon, or well insulated from occasional bruising or harm by the judicious application of cork or wood chips around them. In all these, one is left with the question: How does one provide challenge, independence of action, responsibility, and safety, all in one institution? How does one make the context of childhood optimal? How does one cater to adventure and risk, for stimulation and excitement? Accountability and risk management, if too zealously applied, may make for exactly the opposite of what we want for our children – and what they need; good care and imaginative, motivating, and creative education.

Education is, rightly, always at the cross-roads (Maritain, 1943), but it is a crossroads as much, as Maritain pointed out, about ethics, as pragmatics and deserves constant vigilance and cautious attention by politicians and policymakers, as by teachers, pupils, and families themselves. ECEC is becoming a major part of Western culture, can be seen as supporting, or usurping, changed family functions, and structures. The joy of childhood must not be allowed to disappear in the midst of many well-meaning reforms which (sometimes) further restrict our children and their childhood. The joy of childhood, its partial freedom from responsibility, its soaring imagination, and creativity are part of our very humanity. Whatever is done to and for children must surely take their essential welfare and freedoms into account. According to Morris (1972)

The development of confidence in one's own genuine powers – however limited these may be – is the first essential of personal growth, and such confidence is rooted in an attitude to the world which finds it a good place and the people in it worthy of trust and love. In the beginning, belief in ourselves depends on someone else having believed in us, having cared for us, having loved us. This is the first responsibility of parents and teachers. . . (p. 261)

ECEC is now at the crossroads; its careful, but inevitable provision is as big a social revolution as the establishment of universal elementary education in the nineteenth century. "What is important is not the technology of it all, but the seeing. . ." (Maritain, 1943: 39).

It is up to us to make the spread of ECEC loving, joyful, and essentially worthwhile, so that our children

recognize their future gladly and see their essential humanity both now and into the future.

Bibliography

- Australian Institute of Health and Welfare (2008). Making progress: The health, development and wellbeing of Australia's children and young people. *Cat No PHE 104*, AHWI, Canberra.
- Bailey, D. B. (2002). Are critical periods critical for early childhood education? *Early Childhood Research Quarterly* **17**, 281–294.
- Bennett, J. (2007). Early childhood education for democracy and citizenship. *Asia-Pacific Journal of Research in Early Childhood Education* **1**(1), 65–82.
- Elkind, D. (2008). *The Power of Play: How Spontaneous, Imaginative Activities Lead to Happier, Healthier Children*. Cambridge, MA: DeCapo Lifelong.
- Eurostat (2008). *The Life of Women and Men in Europe: A Statistical Portrait*, 2008 edn. Brussels: Author.
- Gammage, P. (2003). The sacred and profane in early childhood: An Englishman's guide to context and policy. *Contemporary Issues in Early Childhood* **4**(3), 337–356.
- Gammage, P. (2008). *The Social Agenda and Early Childhood Education and Care: Can We Really Help Create a Better World? Online Outreach Paper No 4*. The Hague: Bernard van Leer. <http://www.bernardvanleer.org> (accessed September 2009).
- Giddens, A. (1999). *The BBC Annual Reith Lectures: 2 The Family*. Washington, DC, May.
- Maritain, J. (1943). *Education at the Crossroads*. New Haven, CT: Yale University Press.
- McCain, M. N. and Mustard, J. F. (1999). *Reversing the Real Brain Drain: Early Years Study: Final Report*. Toronto, ON: The Canadian Institute for Advanced Research.
- Morrill, R. (2003). Lessons from the Danes. *Phi Delta Kappan* **84**(6), 460–463.
- Morris, B. (1972). *Objectives and Perspectives in Education*. London: Routledge and Kegan Paul.
- Moss, P. and Petrie, P. (2002). *From Children's Services to Children's Places*. London: Routledge/Falmer.
- New Scientist (2008). Do synapses vary with gender? **199**(2673), 17.
- OECD (2001). *Starting Strong 1*. Paris: Author.
- OECD (2006). *Starting Strong 2*. Paris: Author.
- OECD/CERI (2007). *Understanding the Brain: The Birth of a Learning Science*. Paris: Author.
- Silva, P. A. and Stanton, W. R. (eds.) (1996). *Child to Adult: The Dunedin Multi-disciplinary Health and Development Study*. Oxford: Oxford University Press.
- UNESCO (2007). *Country Profile Commissioned for EFA Global Monitoring Report*. <http://www.efareport@unesco.org> (accessed September 2009).
- UNICEF (2005). *Child Poverty in Rich Countries*. Florence: Author.
- Willis, J. (2007). The Neuroscience of Joyful Education. *Educational Leadership* **64**. http://www.ascd.org/publications/educational_leadership/summer07/vol64/num09/The_Neuroscience_of_Joyful_Education.aspx (accessed September 2009).
- Wynder, E. L. (1998). Introduction to the Report on the Conference on the 'critical' period of brain development. *Preventive Medicine* **7**(27), 166–167.
- Yelland, N. and Kilderry, A. (2005). Postmodernism, passion and potential for future childhoods. In Yelland, N. (ed.) *Critical Issues in Early Childhood Education*, pp 243–248. Maidenhead: Open University Press and McGraw Hill Education.

Investing in Early Childhood Education and Care: Some Policy Implications

J Bennett, OECD, Paris, France

© 2010 Elsevier Ltd. All rights reserved.

The Childcare Transition: A Profound Social Change

In the past three decades, a profound change has taken place in how parents rear their children in the economically advanced countries. Following centuries of being a predominantly family affair, more than 80% of the rich world's 3–6-year-olds are now in early education and care by the time they enter school. Of children under the age of 3, the proportion using childcare is approximately 25%, the proportion using childcare is approximately 25%, for OECD countries as a whole, and in individual countries, such as Denmark and Sweden, more than 50% of 1–3-year olds can be enrolled. Across the industrialized nations, parents are increasingly placing their young children in out-of-home childcare at ever earlier ages and, often, in full-day care (UNICEF, 2008).

Powerful forces are driving this change – among which is the influx of women into the labor market. At present, women in the rich economies are more highly educated than men and desire a more active role in society. Their personal independence, social identity, and participation are based on finding salaried work that is viewed positively and remunerated. The new service economies – in which traditional divisions of labor between men and women have become redundant – provide this opportunity. In addition, women's work is, often, a necessity. The high cost of living in the industrialized world increases the need for dual-earner families. Women are conscious that their work contributes significantly to family budgets and, in many instances – in particular, when women become the sole providers for children – allows families to stay above the poverty line (OECD, 2006).

Childcare Outside the Home Brings Benefits to Economies, Women, and Young Children

In general, governments, business, and finance ministries strongly support the entry of women into the labor market as the larger workforce delivers important economic objectives. Greater participation by women boosts production, increases tax receipts, and reduces the number of families on welfare. Governments, therefore, encourage women to join the labor market, even if they have very young

children. Business, too, welcomes the entry of women into the labor market as the higher educational achievement of women and their relatively lower pay levels make them attractive employees. In parallel, as recent studies from California show, the growth of children's services can significantly expand economies and create many jobs and business opportunities at the local level (Centre for Community Economic Development, 2008).

Figure 1 the proportion of women, in employment, in selected OECD countries. The ratio ranges from a high of 81.2% of women working in Iceland to a low of 23.7% in Turkey.

In addition to serving the economy, investment in early-childhood services allows governments to strengthen other major policy fields – notably, social policy and education, for example, gender-equality goals (**Table 1**). (For a summary of research showing benefits from ECEC services, see OECD (2006), Annex D. This Annex summarizes research showing social, economic, labor market and educational returns from investment.) When governments ensure a sufficient supply of affordable childcare and early-education services, women are enabled to reconcile work outside the home with child-rearing responsibilities. This is good for women's autonomy, family budgets, and the national population–employment ratio. According to the research, the women who have most difficulty in accessing services and finding work are lone mothers and immigrant women – which leads, in general, to family and child poverty among these groups (European Commission, 2008). In addition, although access to work has emancipated many women, studies and data show that a level playing field does not exist for women where recruitment, promotions, and salaries are concerned (EUROSTAT, 2008; World Economic Forum, 2008). The difficulties of reconciling motherhood with a career also impacts on the economy, as many women engage in part-time work well below their potential or in the care sector where wages tend to remain low (Glass and Estes, 1997). **Figure 1** published by the World Economic Forum, in 2008, shows which countries have been successful in ensuring a certain equality for women across four important indicators: women's health and survival; economic participation and opportunity; educational attainment; and political empowerment (World Economic Forum, 2008).

Societies and governments also acknowledge the contribution that early education can make to education. The intense competition among trading nations has convinced

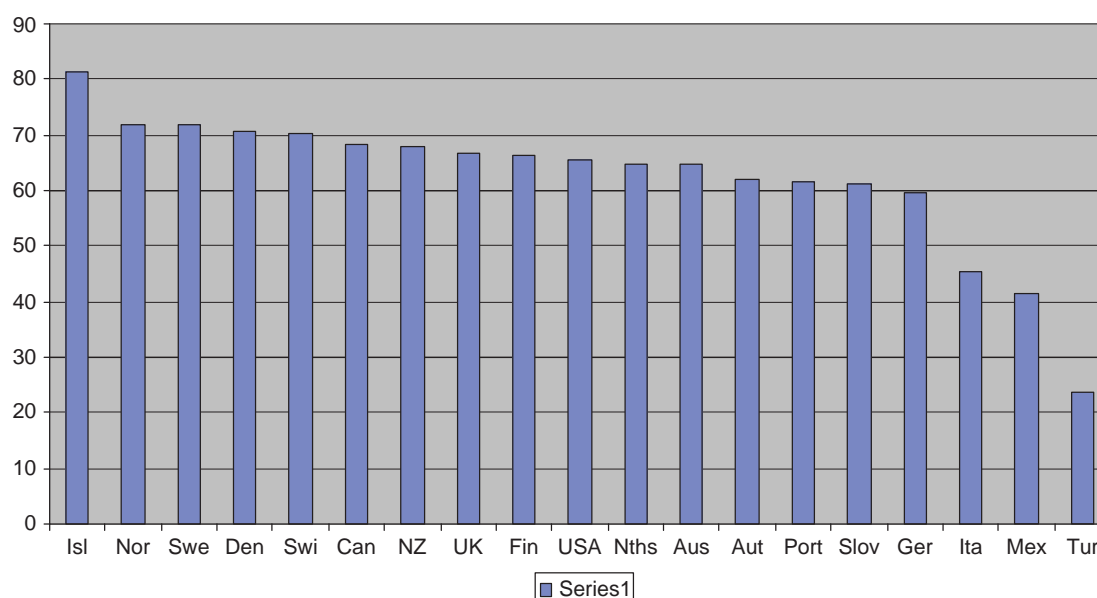


Figure 1 Women (15–64 years) in employment in selected OECD countries. From OECD (2008).

Table 1 The gender-equality gap, 2008

Country rankings	
1. Sweden	In 2007, Latvia (13) and Lithuania (14) made the biggest advances among the top 20 countries, gaining six and seven places respectively, driven by smaller gender gaps in labor-force participation and wages.
2. Norway	
3. Finland	
4. Iceland	
5. N. Zealand	The UK ranks No. 11; Australia ranks No. 17; Canada ranks No. 18; and the USA ranks No. 31
6. Philippines	
7. Germany	
8. Denmark	
9. Ireland	
10. Spain	

governments of the importance of human capital development. Increasingly, they see education as vital to future economic success, and look to the early years as the foundation stage of lifelong learning. Some countries, notably the continental European countries (Belgium, France, Italy, Spain, etc.) enrol 100% of children in early-childhood education services from the age of 3 years (Table 2).

Government Investment in Childcare Remains Weak

A clear disparity exists between investment in childcare services and in early education, particularly in the liberal economies. (The Nordic countries invest most in the care and education of young children. They have integrated childcare and early education both at administrative and centre level, and in general, provide financing directly to services.) In the liberal economics, childcare is often seen as a labour market tool to allow women to work, rather than as

Table 2 Enrolment rates of young children in early-childhood services

Country	0–36 months	36–60 months
Australia	29	71.5
Belgium	33.6	99.6
Denmark	61.7	89.7
Finland	35.0	46.1
France	28.0	101.9
Germany	9.0	80.3
Hungary	6.91	86.9
Iceland	58.7	94.7
Italy	6.3	100.3
Japan	15.2	86.4
Korea	19.5	60.5
Netherlands	29.5	70.2
Norway	43.7	85.1
Portugal	23.5	77.9
Spain	20.7	98.6
Sweden	39.5	86.6
United Kingdom	25.8	80.5
United States	35.5	62

Data taken from the OECD Family database, www.oecd.org/els/social/family/database.

a developmental and educational opportunity for young children. State investment in these services is relatively weak, with parents expected to purchase care in the childcare market – which is often weakly regulated. As a result, childcare services tend to be less developed in terms of coverage, and in some instances, have become a patchwork of small-scale providers and individual family day carers. Various reasons have been put forward to explain this reluctance to invest, for example, the notion that childcare is in the private sphere and not a question for government; the

reluctance of some governments to invest in social programs and equity; the desire to engage the private sector and to keep government small, etc. Whatever the reason, childcare services in many countries tend to be of poor quality: underfinanced and overcrowded, with poorly educated staff who are not able to provide high-quality pedagogical programming. Yet, countless pieces of research show the importance of the years from birth to age 3 in the human life cycle. Given a positive environment, infants and toddlers can make tremendous strides in this foundational period when the senses, language, and the higher cognitive functions are developing at an unprecedented rate (**Figure 2**).

Countries Prefer to Invest in Early-Education Services

The attitude of liberal governments toward early-education services is marked by greater engagement and investment. As children approach school age, targeted educational services – for example, Head Start in the United States of America or Sure Start in the United Kingdom – may be funded for children from low-income and/or dysfunctional families. In recent decades, all countries have become concerned by educational outcomes for low-income children. They are aware that low educational achievement among children does not serve the economy and may lead, later, to delinquency and dependence on social welfare. In response, they adopt a targeted approach to early-childhood services, creating, for example, special services for ‘at-risk’ children. Over the years, these services have become increasingly comprehensive – that is, they go beyond learning activities for young children to also focus on wider aspects of development, such as the general

health and the well-being of children, family support, and community environments. Typically, comprehensive services work in co-operation with other community services and pay particular attention to parents. Both Head Start and Sure Start are examples of comprehensive early-childhood services focused on the development and education of young children, but which can also provide parents with courses and advice on parenting, employment, job training, and leisure-time activities.

Despite this, critics note some fundamental weaknesses in targeted programs. First, services for the poor tend to become poor services, that is, they may not receive, over time, the funding, staff, and environments that disadvantaged children and families need (OECD, 2001). Educational expectations may remain low because of the exclusion of middle-class parents and children from these services. Second, there is the danger that governments, in their enthusiasm for interventions, may be treating only symptoms and not the underlying causes of child poverty – namely, social inequality and its attendant ills. Obviously, children at-risk of educational failure need high-quality programs (**Table 3**), but societies that tolerate extreme inequality and large-scale poverty are, at the same time, placing unacceptable numbers of children in a situation of disadvantage where educational achievement and life chances are greatly endangered.

To What Extent Do Governments Fund Early Education and Childcare?

To answer the question, it is useful to draw a distinction between early-education services, for children 3–6 years,

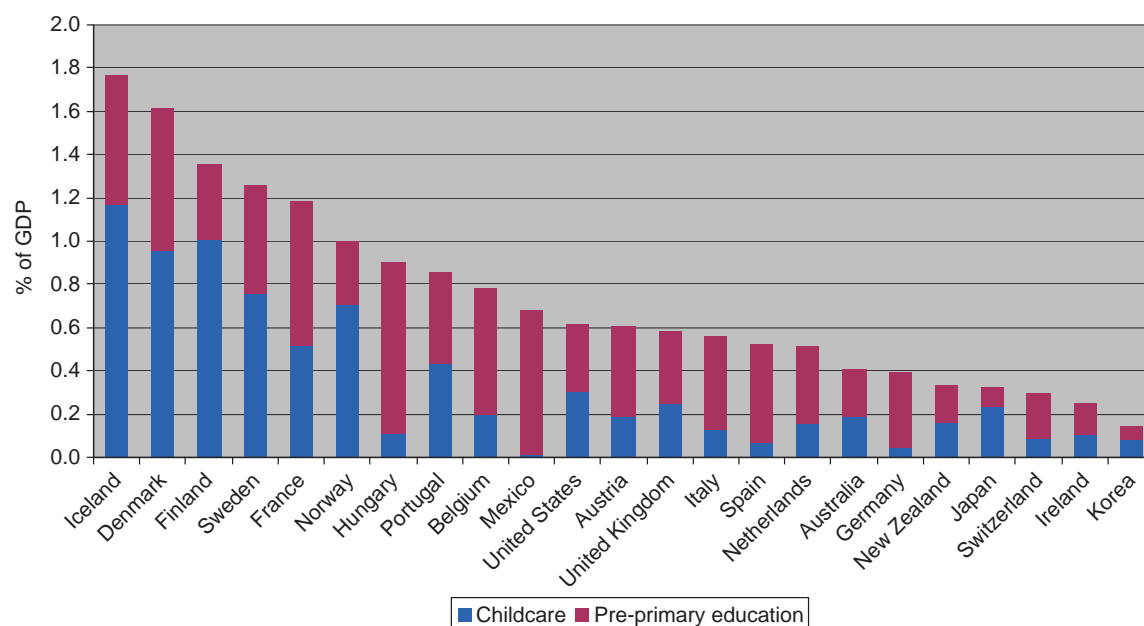


Figure 2 Human brain development in childhood.

Table 3 Investment estimates per child in high-quality early-childhood programs in 2005

<i>Source of estimate and year</i>	<i>Half-day, school year program</i>	<i>Full-day, school year program</i>	<i>Full-day, year round, with integrated child care</i>	<i>Reference</i>
Denmark, 2004			\$19 500 (this figure includes a parental contribution of c. 30%). The net public investment is \$13 650	BUPL, 2005
Finland, 2004			Over EUR 10 248 (not including parental contribution)	STAKES, 2005
Norway, 2005			EUR 12 520 (not including parental contribution)	BFD, 2005
France, 2008		EUR 4000.00 +	EUR 15 000 (including parental contribution)	Le Monde, 09/2008
Sweden, 2004			\$12 097 (not including parental contribution)	Ministry of Education and Culture, 2005.
Abecedarian Project, North Carolina		c. \$13 000 in 2002 costs		Masse and Barnett, 2003
Committee Economic Development, 2006	\$5100	\$8800	\$12 970	CED, 2006 www.ced.org
Head Start, 2005	\$8626 (Federal and local contributions combined).			Barnett and Hustedt, 2005
Kagan and Rigby, 2003	\$4000–\$6000	\$8000–\$12 000		Kagan and Rigby, 2003

From OECD (2006) *Starting Strong II*.

and childcare. Broadly speaking, governments in all countries take charge of the major costs of public early education from the age of 3, 4, or 5 years. In European countries, governments generally provide free early education for 4, 6, or 8 h daily for children from the age of 3 or 4 (Ireland, Netherlands, etc.) years. In Australia, Canada, Korea, and the United States, free education comes about 2 years later: governments provide free preschool educational services for children, from about the age of 5 years – although, in some instances, efforts are made to provide free half-day services for 4-year-olds either under state/provincial auspices or through targeted central (federal) government programming. The Nordic countries provide paying services from the end of parental leave (generally, 1 year or more) up to the age of 6 years, but charges are modest (parents fund between 9% and 15% of service costs). Certain parts of the day (early-education periods) may be free of charge, and costs are waived for poorer parents.

With the exception of the Nordic countries, the picture in childcare is different. The main costs of childcare are taken charge of by parents, with subsidization – to a greater or lesser extent – by government, depending on the country and on the income level of parents. For example, middle-class parents in Ireland, the United Kingdom, and the United States pay most of the costs of childcare; whereas, in the continental European countries, public subsidies take charge of well over half the costs. In the Nordic countries,

state and local government subsidies take charge of over 85% of costs (excepting Denmark). In addition, in Finland and Sweden, a *de facto* right to highly subsidized childcare services from the end of parental leave exists, with parents paying much less than 15% of costs (9% in Sweden).

What Is the Actual Investment as a Share of Gross Domestic Product?

According to the OECD Family Data Base (2007) chart below, the following ranges are found within the industrialized world:

- *% of GDP devoted to childcare 0–3 from public funds:* The percentage ranges from just less than 1.2% in Iceland to 0.2% in Ireland and Korea;
- *% of GDP devoted to early education (3–6 years) from public funds:* The percentage ranges from 0.8% of GDP in Hungary to 0.1% in Korea.

What Share Should Countries be Investing if Quality is to be Achieved?

In 1996, the European Commission Network on Childcare (EC Network on Childcare, 1996) recommended, to European countries, an investment level of at least

1% of gross domestic product (GDP) in early-childhood services. As can be seen from the chart above, only five countries out of the 20 reviewed have clearly reached this investment level: Denmark, Finland, France (the 1% investment attributed includes local authority investments; *école maternelle* for children from age 2; and child-care services), Norway, and Sweden. The case can be easily made that 1% of GDP is a minimum figure if adequate quality is to be achieved. France's investment of about 1% provides a well-organized and free access system but does not ensure adequate child:staff ratios. The French system employs only graduate-level teachers (high salary costs), depends almost entirely on public financing (with no clawback from parental fees) and enrolls almost 100% of children from 3 to 6 years, but in classes of 20–30 children per teacher. In contrast, Finland spends over 1.3% of GDP on far fewer children attending early-childhood services (see **Table 2**), charges parental fees, but is able to maintain child:staff ratios at a much lower rate.

Another way of looking at early-childhood education and care (ECEC) investment is to ask: What is the average investment per child in a good-quality program or in a public ECEC system known to have good services. This may be a more concrete benchmark for early-childhood managers. From the evidence at our disposal – for example, average child costs in Denmark, Finland, Norway, and Sweden; American Abecedarian and Head Start average costs; estimates made by the Committee for Economic Development (CED, 2002, 2006), or as proposed by Kagan and Rigby (2003), the annual cost works out at over US \$8000 per child per year in early education (3–6 years) for a full-day, school-year program (36–39 weeks) in which reasonable child:staff ratios are practiced, and a majority of certified educators are employed. The CED (CED, 2006) proposes \$5000 as a rough starting point for a child attending a part-day, part-year program. (Abecedarian costs run to \$63 476 per child over 5 years. See also Gormley and Phillips 2004 on Oklahoma Pre-K costs, and the analysis of

Barnett *et al.* 2005 of pre-K costs across five states (Gormley and Phillips 2004; Barnett *et al.*, 2005)).

More detailed costs have been estimated by a number of authoritative sources (Kagan and Rigby, 2003; Masse and Barnett, 2003 (Abecedarian); Barnett and Hustedt, 2005; Barnett and Robin, 2006; CED, 2006). They indicate that costs per child in a high-quality early-education service – with child:staff ratios equal to or less than 10 children per trained adult – range from US \$8000 to \$14000 annually per child 1–3 years, and between \$6000 and \$10 000 per child 3–6 years. As can be seen from **Figure 3**, recent figures from France place annual early-childhood costs at a far higher level: €4000 for a place at the *école maternelle* (child:staff ratio of between 20 and 30 children per teacher); €10 000 for a place at an accredited family day care; and €15 000 per year (11 months) in a licensed crèche (Le Monde, 2008).

These figures suggest that current financing of early childhood services is insufficient if high quality is to be maintained. Our observation is corroborated by the financing costs provided by Head Start, the Nordic countries and several municipalities interviewed during the *Starting Strong* OECD reviews. In sum, costs per child in ECEC services will become a real challenge in the years to come. To move from an average OECD investment per child of \$4741 in early education (EAG, 2007) to double that amount (the current Head Start investment if one includes the 20% local contribution) will demand a significant financial effort on the part of governments.

What Funding Methods Do Governments Use?

Funding in Early-Education Services for Children over 3 Years

The funding of public education services for children over 3 years (or older in the liberal economies) is similar across most countries. Supply-side funding directly to services is

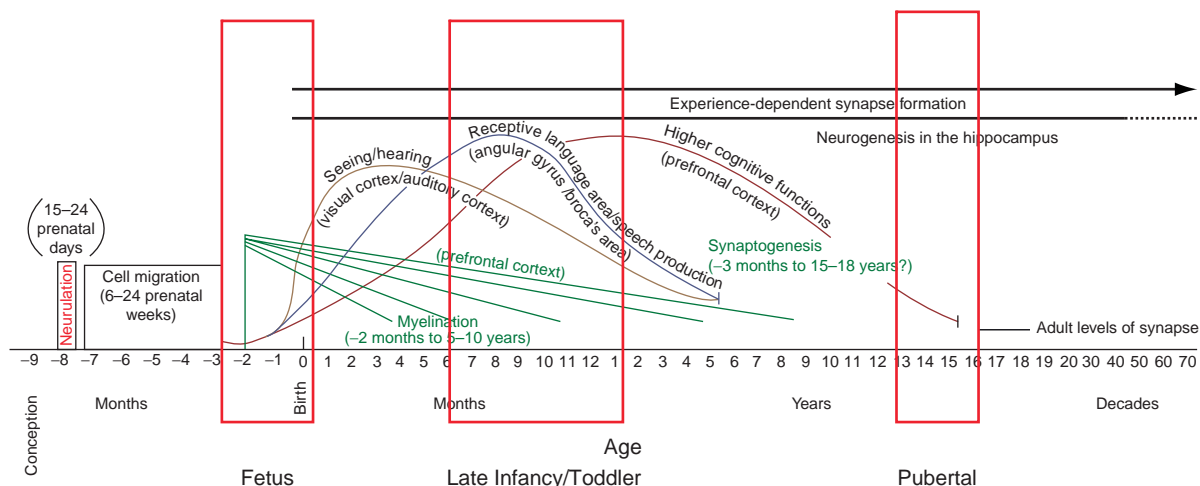


Figure 3 Public expenditure on childcare and early education as a percentage of GDP, 2003.

the major mechanism used. Services are often provided directly by government or local authorities, or by government dependents – that is, voluntary or private nonprofit organizations that receive more than 50% of their funding from government and who have agreed to run services in accordance with government regulations or to specific contractual obligations. The variety of accredited voluntary or private organizations used by governments to deliver services, in this way is varied. Possibly, the largest accredited government dependents in the early-education field for children 3–6 years – as in compulsory education – are faith-based organizations. In the Benelux countries, Germany, and Italy where church groups take charge of a third or more of all early-education services.

In such situations, it is customary for governments (central, state, or local) to provide funding directly to providers, linked to the number of children being served. Funding, generally, takes the form of operational subsidies, staff wages (or wage-enhancement grants), grants for capital equipment and supplies, supplementary grants made to services serving children with additional or special educational needs, and grants for the enhancement of quality or other public objectives. In general, the organizations that receive such subsidies come under strict regulation – although, in many of these services in Europe (excepting the Nordic countries), child-staff ratios can be high, and shortages of provision occur, particularly in growth urban areas. In most countries, governmental departments prepare legislation or regulatory frameworks, license programs, and set group sizes and staff qualifications. Depending on the country, government services may also be helped by provider management groups in monitoring quality and program standards, providing for educator support and professional development, and encouraging parental participation.

Funding for Children under 3 Years

For children younger than 3 years, both supply-side (funding to services) and demand-side funding (subsidies to parents) or a mixture of both models are used. In general, a division can be seen between the liberal economies and other countries. Countries in the former group (Australia, Canada, Ireland, Korea, the Netherlands, the UK, the US, etc.) tend to use, as their main financing mode, childcare subsidies to parents – such as cash benefits, vouchers, tax reductions, and so on. In contrast, the European continental countries use supply-side subsidies paid directly to services as their main mode of financing, although some countries – for example, Belgium and France – may also use tax credits to help parents meet childcare costs. In the social democratic countries, supply-side funding – and tax credits – are predominantly used. Direct municipal provision of day care services (including family day care) is also the rule in these

countries – except in Norway where, in 2004, private kindergartens already provided for the majority (57%) of children (Moser, 2005). However, family day carers in Denmark and the contracted private providers in Finland and Norway are not the independent (and often unregulated) operators found in the liberal economies, but are licensed and regulated by the local municipality.

Controlling Costs

As in other human services, the ratio of qualified staff employed generally sets the level of ECEC costs. This has led some governments either to cut back on qualified staff or to increase the number of children per qualified teacher. Thus, in the childcare sector, costs are often contained through the employment of poorly qualified and poorly paid staff – a feature found often in privatized childcare in the liberal economies. In early education, costs are, often, curtailed in public early-education services through allowing child:staff ratios to rise – for example, in France, Ireland, Korea, and Mexico where child:staff ratios in excess of 25:1 are to be found. Neither approach is adequate if the aim is to have services that provide high-quality education and care for young children.

A more positive approach to keeping costs at a reasonable level is to collect a financial contribution from parents and to build up team teaching. In some of the Nordic countries, university-trained, kindergarten educators form approximately a third (Finland) or half (Sweden) of staff in ECEC centers. They work in teams with trained children's nurses or child assistants. In this way, these countries can provide both appropriate child:staff ratios and graduate-led services. Staff knowledge and morale are maintained – especially for the lesser qualified staff – by acceptable work conditions and ongoing professional development tied to professional advancement. In the United States of America, Hartman (2008) proposes that savings could also be made through raising parental fees and through control of teacher benefits and pension rights, which are often far more favorable than for the general population.

Which Is More Effective: A Direct Funding-of-Services Model or a Parent Subsidy Model?

The experience of the OECD country reviews (OECD, 2001, 2006) suggests that, for the moment at least, a public supply-side investment model brings more uniform quality and superior coverage of childhood populations (1–6 years) than parent-subsidy models. The stratagem of directly funding parents, while politically attractive,

weakens both equity and governmental steering of the early-childhood field. Direct public funding of services (not to be confused with direct public provision of services) brings – in the majority of countries reviewed – more effective control, advantages of scale, better national quality, more effective training for educators, and a higher degree of equity in access and participation than consumer-subsidy models. In addition, advantages of scale can be considerable: public systems make it easier to enforce regulations, support educators, monitor quality and communicate good practice within the system. The comparison is striking when one compares the financing and organization of public early education with that of childcare in countries where these sectors are split. A similar difference in coverage and quality is also apparent, when marketized childcare models are compared with the predominantly public service model of the Nordic countries. The experience of Norway and Sweden also suggests that a public service model can accommodate private providers when they are properly contracted, regulated, and supported by public funding. However, a public service aiming at high quality is costly, as it implies employing well-qualified teachers, reasonable child:staff ratios, and adequate buildings and resources. Outside the nordic countries, these criteria are not always met in public services.

A possible solution to keeping down government costs is the contractual outsourcing approach, whereby private providers are brought into the provision network through public–private partnerships. This is the predominant approach, for example, in New Zealand and Norway. The approach can bring down the costs of services (particularly capital costs) and enlarge the choice of provision offered to parents. It can be acceptable also to ECEC workers, when the state supports a policy of higher qualifications and maintains a guaranteed wage structure for all qualified personnel, whatever their place of work. A similar situation pertains in the formal education system, where government dependents are contracted to deliver primary and secondary education. In many countries, these providers receive full government funding, but are not allowed to charge fees or (in the case of Sweden) fees greater than those charged by the public services. This is to avoid a growing disparity between services for low- and modest-income families and services for richer parents who can afford supplementary fees.

A more radical means of lowering costs is for governments to encourage an open, deregulated market in childcare services – with assistance given to poorer parents through parent subsidies. Up to the moment, results from such policies have not been encouraging (Prentice, 2005; Mitchell, 2002; Cleveland and Krashinsky, 2003, 2005). One reason is that the entry of private providers often leads to state disengagement from the field and a loosening of regulations. The situation, generally, leads to a

fragmentation of provision, a decline in quality, and clear inequalities in access and outcomes. The crux of the matter is that when public funding of the childcare system takes the form of subsidies paid directly to parents, the subsidies are, generally, too low to employ high-quality staff or to finance system infrastructure – for example, data collection, monitoring, inspection, in-service training, etc. In addition, the steering capacity of governments *vis-à-vis* services becomes considerably weaker than in funding-to-services systems. The conclusion reached in the PricewaterhouseCoopers (2004) report, on financing a universal ECEC system for England by the year 2020 is instructive:

Supply-side funding tends to be the dominant form of finance in countries with the best developed systems of early years education and care, such as Sweden, Denmark, France and New Zealand, whereas means-tested, demand-side funding is more typical of countries with less well-developed systems, such as the UK and the US. (Daycare Trust, 2004)

See also: Gender Issues in Early Childhood Education and Care; Investing in Early Childhood Education and Care: The Economic Case.

Bibliography

- Barnett, W. S. and Hustedt, J. T. (2005). Head start's lasting benefits. *Infants and Young Children* **18**(1), 16–24.
- Barnett, W. S., Jung, K., Wong, V., Cook, T., and Lamy, C. (2005). *The Effects of State Prekindergarten Programs on Young Children's School Readiness in Five States*. New Brunswick, NJ: Rutgers University, NIEER.
- Barnett, W. S. and Robin, K. B. (2006). *How Much Does Quality Pre-School Cost?* New Brunswick, NJ: NIEER/Rutgers.
- Bjorklund, A., Edin, P. A., Frederiksson, P., and Krueger, A. (2004). Education, equality and efficiency: An analysis of Swedish school reforms during the 1990s. *IFAU Report*, No. 1.
- CED (2002, 2006). *Preschool for All: Investing in a Productive and Just Society*. New York: New York Committee for Economic Development.
- Centre for Community Economic Development (2008). *The Economic Impact of the Early Care and Education Industry in Los Angeles County*, Los Angeles.
- Cleveland, G. and Krashinsky, M. (2003). *Financing ECEC Services in OECD Countries*. Paris: OECD.
- Cleveland, G. and Krashinsky, M. (2005). *The Non-Profit Advantage: Producing Quality in Thick and Thin Child Care Markets*. University of Toronto at Scarborough, Division of Management.
- Daycare Trust (2003). Costs and benefits of universal childcare. A preliminary economic analysis for the UK. http://www.daycaretrust.org.uk/mod/fileman/files/Cost_benefits__universal_childcare.pdf (accessed September 2009).
- EAG (2005). *Education at a Glance*. Paris: OECD/Author.
- EAG (2007). *Education at a Glance*. Paris: OECD/Author.
- EC Network on Childcare (1996). *A Review of Services for Young Children in the European Union 1990–1995*. Brussels: European Commission, DGV.
- European Commission (2008). *Child Poverty and Well-Being in the EU: Current Status and Way Forward*. Brussels: Author.
- EUROSTAT (2008). *The Life of Women and Men in Europe: A Statistical Portrait*, 2008 edn. Brussels: Author.

- Glass, J. L. and Estes, S. B. (1997). The family responsive workplace. *Annual Review of Sociology* **23**, 289–313.
- Gormley, W. T., Jr. and Phillips, D. A. (2005). The effects of universal pre-k in Oklahoma: Research highlights and policy implications. *Policy Studies Journal* **33**(1), 65–82.
- Hartman, K. E. (2008). Common-sense ways to improve education without a tax increase. *Education Week* September 17, 2008-09-18.
- Kagan, S. L. and Rigby, E. (2003). *Policy Matters: Improving the Readiness of Children for School: Recommendations for State Policy*. Washington, DC: The Children's Project.
- Le Monde (2008). *Garde d'enfants: bataille autour d'une promesse*. Paris: 19/09/2008.
- Masse, L. N. and Barnett, W. S. (2003). *A Benefit Cost Analysis of the Abecedarian Early Childhood Intervention* (Mitchell, 2002). New Brunswick, NJ: NIEER/Rutgers.
- OECD (2001). *Starting Strong: Early Childhood Education and Care*. Paris: OECD.
- OECD (2006). *Starting Strong II: Early Childhood Education and Care*. Paris: OECD.
- OECD Family Data Base (2007). OECD family data base. <http://www.oecd.org/els/social/family/database> (accessible September 2009).
- Prentice, S. (2005). For profit child care: Past present and future. *Childcare Resource and Research Unit, Occasional Paper 21*, Toronto, Canada.
- World Economic Forum (2008). *The Global Gender Gap Index, 2007*. Geneva: Author. <http://www.weforum.org/pdf/gendergap/report2007.pdf> (accessed September 2009).
- BUPL (2005). *Danish Federation of Early Childhood Teachers and Youth Educators: Communication: Annual Cost Per Child for Danish Children in Early Childhood Services*. Copenhagen, Denmark.
- EUROSTAT (2007). *Europe in Figures, Eurostat Yearbook 2006–07*. Brussels: Author.
- Kaga, Bennett and Moss (in press) *Caring and Learning Together*. Paris: UNESCO.
- Nelson, C. (2000). In Shonkoff, J. and Phillips, D. (eds.) *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, DC: National Academy Press.
- OECD Family Data Base (2008). OECD family data base. <http://www.oecd.org/els/social/family/database> (accessible September 2009).
- PricewaterhouseCoopers (2004). Universal early education and care in 2020: Costs, benefits, and funding options. *A Report for Daycare Trust and the Social Market Foundation*. London: Daycare Trust, Pricewaterhouse-Coopers, and Social Market Foundation.
- STAKES (2005). *Finnish National Research and Development Centre for Welfare and Health: Communication: Annual Cost Per Child for Finnish Children in Early Childhood Services*. Helsinki: Author.
- Swedish Ministry of Education and Culture (2005). *Communication: Annual Cost Per Child for Swedish Children in Early Childhood Services*. Stockholm: Author.

Relevant Websites

Further Reading

- Bennett, J. (2008). *Early Childhood Services in the OECD Countries: A Technical Background Paper for the Innocenti Report Card No. 8*, UNICEF, Florence.
- BFD (2005). Ministry of children and family affairs, *Communication: Annual Cost Per Child for Norwegian Children in Early Childhood Services*.
- <http://www.bernardvanleer.org> – Bernard van Leer Foundation.
- <http://www.childcarecanada.org> – Childcare Resource and Research Unit.
- <http://ec.europa.eu> – EUROSTAT, European Commission.
- <http://www.oecd.org> – OECD Family Database, Organisation for European Economic Co-operation (OECD).
- <http://titania.sourceoecd.org> – OECD Factbook 2009, Source OECD.
- <http://www.oecd.org> – Organization for European Economic Cooperation (OECD).

Investing in Early Childhood Education and Care: The Economic Case

G Cleveland and M Krashinsky, University of Toronto Scarborough, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

This article provides a guide to the economic case for public investment in early childhood education and care (ECEC) services. Public investment in ECEC can take many forms:

- Child-related leaves and associated benefits – including maternity, parental, paternity, and child-rearing leaves, with or without paid benefits, with or without full job protection.
- Publicly provided ECEC services – including preprimary education and ECEC services provided by public-sector bodies or nonprofit agencies (even with user fees so long as these fees are small for all users).
- Supply subsidies to ECEC services – operating grants, quality-enhancement grants, wage-enhancement grants, capital-equipment grants, tax benefits, and tax reductions given to ECEC services otherwise normally subject to taxation.
- Demand-side subsidies for the use of ECEC – subsidies to low-income families for the use of ECEC services, tax deductions of ECEC expenses or tax credits based on ECEC expenditure, vouchers for the purchase of approved types of ECEC services.

There are two broad sets of reasons why governments may invest in ECEC services. The first is to transfer resources to families with young children, as a matter of equity. Raising children is expensive and families with young children are typically themselves young; the incomes and assets of these families are low relative to what they may be later in life, and relative to the high cost of raising young children. The second is to protect and promote the public interest in the raising of children and the functioning of families; this can enhance economic efficiency by correcting market failure. While governments seek to preserve a wide scope of freedom for parents to decide what is in the best interests of children and family members, ECEC policies generally provide incentives, financial support, and regulation of services that are intended to help parents make decisions that are most positive for the interests of children, families, and society in the long run. This may go so far as to include the provision by governments of accessible, good-quality ECEC services with or without parent fees.

Why Private Markets May Fail in the Provision of ECEC Services

Thinking like an economist, the default position is to assume that all goods and services will be privately produced and sold in private markets, with private interests determining whether and under what conditions these goods and services will be bought and sold. In this manner of thinking, the arguments for public investment, regulation, and policy of ECEC services must proceed by explaining why and in what ways ECEC services should be an exception to this free-market pattern. This form of argument is particularly well suited to Anglo-Saxon countries with strong market orientations, which typically favor moderate amounts of demand-side (i.e., financial assistance directly to consumers) government intervention in ECEC markets. This form of argument may seem odd to those in countries who have always assumed that (publicly provided) ECEC is a public responsibility.

In a market economy, some goods and services can be bought and sold in competitive markets and, without substantial government involvement, buyers and sellers interact in ways that produce desirable results. In effect, buyers of these commodities weigh up the benefits and compare them to the costs when they make a decision to purchase. The sellers behave in ways that ensure that the prices faced by buyers reflect the true cost to society of producing this good or service. The private competitive market, without government action, allocates an appropriate amount of society's economic resources to the production of these goods. In these cases, competition keeps prices low and encourages producers to invest in new technologies and to seek new ways of satisfying customer demands. Furthermore, the multiplicity of potential sellers gives consumers maximum freedom of choice to find the combination of characteristics that suits them best.

However, certain goods or services do not have some or all of the economic characteristics required for competitive markets to deliver desirable results. For instance, a good or service may generate substantial positive external benefits which are not taken into account in private market transactions. Or a good or service may have characteristics that are difficult for potential consumers to judge

accurately and purchasing mistakes may be made with negative consequences. In these and similar cases, there is market failure, and these market failures can potentially be corrected by government action. Government action will be appropriate if the benefits gained by government policy are greater than its costs. Education, for example, is a service that, because of its characteristics, is typically provided through the public sector because private markets fail to deliver the appropriate amount of it to those who need it most; the benefits of publicly financed (and, typically, publicly provided) education are judged to exceed its costs.

Two broad areas of potential market failure exist in relation to ECEC: market failure in relation to the care and education of young children, and market failure in relation to the employment of parents (especially mothers) when children are young.

There is a public interest in children and in the care and education they receive when they are young. There would be no market failure related to children if this public interest completely mirrored the interest of parents and their ability to purchase good-quality ECEC services. However, an educated workforce is essential both for economic growth and for the maintenance of a healthy democracy, and these benefits spill over beyond the individual family to society as a whole. Further, society has an interest in assuring that children get an equal start in life; if ECEC decisions are financed entirely by the stretched incomes of young parents, this equal start is imperiled.

There is, in equal measure, a public interest in the employment attachment of young parents. Since mothers are likely to be primary caregivers for their young children, they are likely to have their employment relationships interrupted in children's early years. Mothers' employment, earnings, and related child-care decisions may be distorted by various forms of market failure. One distortion is the absence of perfect capital markets; in theory, since purchasing good-quality ECEC services for children should mean that those children prosper in the future, and because continued labor-force attachment for mothers should mean increased prosperity for these parents in the future, it should be possible to borrow the cost of ECEC services from a bank today, using the future prosperity as security for the loan. Of course, this is not possible. Public provision of ECEC services today, financed by future taxation on more prosperous parents and children, is an alternative, given the absence of perfect capital markets to bridge the present and the future.

Another important market distortion in countries in which ECEC is sold on the private market is the taxation of earnings which are used to purchase ECEC services; since the costs of ECEC are necessary costs of employment (rather than an expenditure of discretionary income), the failure to make these expenditures fully deductible from income before taxes are assessed produces an economically inefficient barrier to mothers' employment.

Another distortion, particularly important for lone-parent families, is provided by the tax-back rate, or benefit-reduction rate associated with social-assistance (or welfare) payments. These punitive benefit-reduction rates make employment unattractive for the majority of families on social assistance, leading to increased dependence over time. If, in addition to this, ECEC services are costly, this inefficient barrier to employment is still higher.

Finally, subsidization of ECEC may be desirable as a second-best response to gender discrimination in labor markets or within the family. If gender discrimination compels women to be the sole or main caregiver for young children, this discrimination may cause women and girls to make welfare-reducing decisions about labor-market participation, job promotion, occupational choice, and human-capital investment. Subsidization of ECEC services may provide a basis for greater gender equity in both work and home settings.

ECEC policies in many countries are directed toward reducing or eliminating one or another of these sources of market failure in relation to children and families. In general, ECEC policies that are directed toward both providing developmental/educational benefits for children and reducing employment barriers for parents will have a greater ratio of benefits to costs. Beyond these economic-efficiency concerns, ECEC policies are typically directed toward equalizing the economic and social opportunities of children from different backgrounds.

The Research on Benefits to Children

There are a very large number of factors that can affect a child's development and behavior. Isolating the separate effects of good-quality child care on children while holding all the rest of these factors constant is a difficult research task; many of the disagreements in this research area are methodological. The key research problem is that ECEC services are not normally randomly assigned to children. Most data come to researchers from a situation in which individual parents have chosen the type and quality level of ECEC that their child will use. As a result, it is difficult to separate the effects of family from the effects of ECEC. The research that is most trustworthy in evaluating the effects of ECEC on children is research that deals well with this issue (known as selection bias).

There have been several waves of research about the effects of child care on children. Many of the early studies were random assignment studies of children from lower-income families to good-quality child care centers. Overwhelmingly, these studies (e.g., the Perry Preschool Study, the Abecedarian Study, and the Chicago Parent-Child Center Program) found that good child care can have very positive effects on children and that these

advantages can be long-lasting. In particular, good child care can compensate, at least partially, for a disadvantaged home life.

A benefit–cost analysis of the Abecedarian Early Childhood Intervention has revealed how broad and long-lasting the effects of a well-designed ECEC program can be on disadvantaged children. The experiment involved 112 children, mostly African-American, whose family situations were believed to put them at risk of slowed development. On average, maternal education in experimental families was 10 years, maternal IQ was 85, and 55% of households were collecting social assistance. The benefits to participants (and the community) included:

- improved measures of intelligence and achievement over the long term, leading to higher earnings and fringe benefits now and in the future;
- lower levels of grade retention and placement in special-education classes, leading to cost savings in elementary and secondary education;
- improved employment and earnings of mothers of the children receiving early childhood education services;
- reduced probability of smoking and improved health; and
- reduced use of social assistance.

The costs of providing intensive high-quality child care in the Abecedarian program were high. For infants, there was one staff member to every three children; for 2- and 3-year-olds, there were two staff members for every seven children; for 4- and 5-year-olds, the ratio was one to six. All staff were paid competitive public school salaries. Nevertheless, the value of the benefits, discounted back to the present, was found to be considerably higher than the costs. At a discount rate of 3%, the project provided a 4 to 1 return on the investment of public resources targeted at a disadvantaged group (Masse and Barnett, 2003).

The advent of a universal 4-year-old prekindergarten program in Tulsa Oklahoma has demonstrated the strong positive cognitive and language effects of a well-designed prekindergarten. Nationally normed test instruments (Woodcock–Johnson subtests for letter-identification (prereading), spelling (prewriting), and applied problems (premath)) found effect sizes, from 1 year of a very good-quality prekindergarten program, of 0.79 standard deviation for prereading, 0.64 standard deviation for prewriting, and 0.38 standard deviation for premath. Strong positive effects were found for both disadvantaged and middle-class children, and children from different racial and ethnic groups (with, in general, somewhat larger effects for disadvantaged and both Black and Hispanic children, but substantial positive effects for all children).

The Oklahoma results were partly due to the very high quality of prekindergarten services provided. Prekindergarten services were provided in the schools by teachers who had both a teaching certificate and a certificate in

early childhood education, and who were paid at public-school rates. Classroom sizes were capped at 20 children, and, with one lesser-trained assistant, this meant that staff-child ratios were 1:10 (Gormley *et al.*, 2005).

A similar pattern has been found in prekindergarten programs in five states: Michigan, South Carolina, New Jersey, West Virginia, and Oklahoma. Some of these programs are universal and some targeted, but not as well resourced as the Tulsa program. In all five states, there were substantial gains in both prereading and premath skills (Barnett *et al.*, 2005).

Of course, these latter results are for prekindergarten (generally provided in schools), rather than child care or other preschool. However, there is good evidence that child care can have positive cognitive effects too, as long as it is of adequate quality. Recent studies in the USA conclude that the effects of quality are relatively small (0.04–0.08 of a standard deviation on cognitive outcomes for children), but that, in addition, use of center-based care in the third and fourth year of a child's life has an independent effect of about 0.25 on cognitive and academic-achievement outcomes. Taken together, the effect of a good quality center-based program would be substantial (less than half the size of the effects of parenting or home environment or being in poverty, but still sizeable) (NICHD-ECCRN and Duncan, 2003).

Other studies of center-based programs include various studies of Head Start in the US. Head Start children are a targeted low-income group with heterogeneous programs delivered to approximately 1 million children per year.

These studies are well designed to account for selection bias; the main technique used is sibling fixed effects. In other words, the control group for the effect of Head Start is composed of the sibling of each Head Start attendee who did not attend Head Start. This controls for unmeasured family characteristics that might affect both Head Start enrolment and child outcomes. Currie and Thomas (1995) found long-term (i.e., into the early school years) positive effects on school achievement for white Head Start attendees, but not for African-American children, apparently because of the later negative effects of their schooling. In a second paper (1999), Currie and Thomas found long-term positive effects on school achievement for Hispanics (Currie and Thomas, 1999). Examining the long-term effects of Head Start programs when children were in their early 1920s, Garces *et al.* (2002) found that white children who attended Head Start were more likely to complete high school, to attend college, and to earn more than other white children who did not attend Head Start. African-American children were more likely to complete high school and less likely to be arrested for criminal activity than similar children who did not attend Head Start.

Finally, a recent study examines the long-term effects of the introduction of universal access to ECEC for 3–6-year-olds in Norway in the 1970s. A difference-in-differences

design, based on differential implementation of the program by municipalities, finds strong positive effects on long-term child outcomes. An increase of 17 500 child-care spaces results in about 6200 additional years of education measured 30 years later. There are also significant increases in labor-market participation for these children when they grow up, and reduced welfare dependency (Havnes and Mogstad, 2009).

In summary, there are good reasons to believe that some forms of center-based child care/preschool/prekindergarten can have important positive effects on children, whether these children are disadvantaged and low income and from lone-parent families or whether these children are from middle-income and two-parent families. The effect sizes appear to be dependent on two main factors: the quality and type of child care/early education they receive, and the quality (support and stimulation) of the care the children would have alternatively received (often related to the family situation of the child). The age of the child moderates both of these factors, and the persistence of improved child outcomes will depend on later classroom experiences (Magnuson *et al.*, 2007a).

It is possible that child care can also have some negative effects on children that have to be weighed against the positive cognitive, academic, and language effects (Magnuson *et al.*, 2007b; Baker *et al.*, 2008). Recent US studies have found evidence that more hours in a child-care arrangement is associated, at 24 and 54 months and in kindergarten, with various behavior problems. Initial research found that more hours in child care up to age 2 were associated with more negative interactions with peers and more behavior problems as reported by caregivers, and less social competence as reported by mothers. Measured cumulatively up to 4.5 years, more hours in child care was associated with more negative play, less social competence, and more externalizing behaviors. Higher amount of hours in child care was associated with more teacher-child conflict in kindergarten. These effects were moderated by the quality of child care and the quality of parenting received by the child, but were still statistically significant and quantitatively important after controlling for these influences (NICHD-ECCRN, 2003).

The effects on behavior do not appear to be related to a threshold level of hours. However, it is only children in child care for more than 30 hours per week who, on average, had more than normal behavioral problems, and only children spending over 45 hours per week over the whole period from 3–54 months who displayed high levels of negative externalizing behaviors. The weight of these negative effects is uncertain. One recent comprehensive study (Duncan *et al.*, 2007) found that while math and reading skills and ability to pay attention at the time of school entry can matter dramatically for later academic success, internalizing and externalizing behaviors had no independent effect.

Effects of ECEC on Parents' Employment Situation

The way most societies have evolved, it is generally the mother's role to take the primary responsibility for both the provision of care to young children and the making of day-to-day decisions about their lives. Accordingly, it is nearly always the mother's career that is foregone if someone stays home with the children; it is the mother who works part time when children are young, who declines opportunities for advancement, and who neglects the acquisition of skills that might permit moving to a higher income. Of course, young children make life forever different for fathers as well; often fathers may work harder or longer hours, and there is a considerable amount of off-shifting, where fathers and mothers adjust work schedules to avoid having to hire paid caregivers while both work. The evidence, however, seems overwhelming that changes in ECEC policy will have more dramatic direct effects on the daily lives of mothers, and on fathers, more indirectly.

Several different dimensions of female labor-market behavior are potentially affected by changes in ECEC policies (1) labor-force participation, (2) hours of work, (3) acquisition of general labor-force experience and company-specific seniority, (4) human-capital acquisition, (5) earnings, and (6) occupational status. There is a considerable research literature on the effects of child care costs on mothers' labor-force participation, much less on hours of work, and very little on other dimensions of labor-market experience.

There is an approximate consensus among researchers that the price that parents have to pay for child-care services will negatively affect the mother's employment decisions. However, there is considerable difference on the size of this effect. Blau (2003) analyses a range of studies and concludes that the possible positive effects of public investment in ECEC on participation in the labor force and hours of work are both small (an elasticity of labor supply to child-care price of probably less than -0.20).

Several natural experiments have been studied to provide information about effects of the price of child care on mothers' employment. For the USA, Gelbach (2002) analyses the impact of public kindergarten. His estimates indicate that access to free-of-charge part-day kindergarten increases the probability of the mother being employed by 4–5 percentage points. Cascio (2009) finds that the introduction of (typically half-day) state kindergarten programs in the 1960s had a strong positive impact on the labor supply of single mothers with no younger children, but not on other mothers.

However, a natural experiment which began in 1997 in Quebec (Canada) has produced substantial labor-supply effects on the mothers of preschool children. The policy change involved several family policy measures, including

provision of free full-school-day kindergarten for 5-year-olds and center-based or home-based child care for \$5 per day for children from 0 to 4. Lefebvre and Merrigan (2008) find that by 2002, the policy changes had increased the labor-force participation rate of mothers with young children by about 13%. Annual hours worked and weeks worked had risen by even greater percentage amounts, driven by an increase in full-time work. Mothers' earnings were up correspondingly. Baker *et al.* (2008) confirm this general pattern using a different data set.

There is not much evidence about how ECEC costs, convenience, and quality affect whether a mother works full time versus part time. Powell (1997) reports on Canadian evidence suggesting that full-time work is quite sensitive to child-care costs, while part-time work is less so.

Government ECEC Policy Issues

Even when governments agree on the importance of public investment in ECEC, there remain important differences over the forms that public policy should take. One issue is whether public intervention will occur on the demand side or on the supply side. Usually, this decision comes down to a debate over the importance of maximizing parental choice versus maximizing control over the quality of ECEC provided, and about the effectiveness of alternative ways of improving quality.

It is important to keep in mind that the most important component of the early care of the young child will occur within the child's home and will be provided by the child's parents. Even the most extensive ECEC programs currently in operation generally provide no more than 40–50 hours a week of ECEC to children, implying that the rest of the day and nighttime hours (120 or more) are provided by parents. Most studies of child outcomes identify parents as the most important factor in determining how children turn out. A central concern for ECEC policy will be to determine the appropriate boundary between parental care and ECEC services outside the family. In particular, a decision needs to be made on what age is most appropriate to begin ECEC services, and what supports are necessary to families prior to that age to allow them to provide family care. Currently, countries vary widely in the nature of the parental and child-rearing leave programs they provide, and in the extent to which those programs encourage mothers and fathers to remain in the home to care for their children when they are very young.

There is no one best funding mechanism for all circumstances. However, it is often the case that disagreements on funding that appear to be only technical also reflect differences in philosophy. For example, individuals (and governments) may differ on the appropriate role of the state in setting standards for the care of children, and

on the appropriate roles for men and women within and outside the family. Different funding mechanisms will be consistent with a larger role for the market or for the state in ECEC and these different approaches may be more or less consistent with practices and beliefs of the party in power in different countries.

Concluding Observations

The evidence on the benefits and costs of ECEC does not suggest that all and any expenditure of public money on ECEC will generate benefits greater than costs. The precise design of ECEC financing programs matters. In particular, the ratio of benefits to costs is clearly affected by the quality of ECEC services available; benefits to children rise with quality level, not just up to some point, but apparently without obvious limit (Lamb *et al.*, 1998). However, costs also rise, in general, as the quality level of ECEC services rises. The ratio of benefits to costs is similarly affected by the degree of employment support for mothers that ECEC programs provide. Programs should be designed to maximize the excess of benefits over costs for any child, and should include all children and families for whom benefits exceed costs.

Bibliography

- Baker, M., Gruber, J., and Milligan, K. (2008). Universal childcare, maternal labor supply and family well-being. *Journal of Political Economy* **79**(4), 709–745.
- Barnett, W. S., Lamy, C., and Jung, K. (2005). *The Effects of State Pre-Kindergarten Programs on Young Children's School Readiness in Five States*. New Brunswick, NJ: National Institutes for Early Education Research.
- Blau, D. M. (2003). Child care subsidy programs. In Moffitt, R. (ed.) *Means Tested Social Programs in the U.S.* Chicago: University of Chicago Press.
- Cascio, E. (2009). Maternal labor supply and the introduction of kindergartens into American public schools. *Journal of Human Resources* **44**, 140–170.
- Currie, J. and Thomas, D. (1995). Does head start make a difference? *American Economic Review* **85**, 341–364.
- Currie, J. and Thomas, D. (1999). Does head start help Hispanic children? *Journal of Public Economics* **74**, 235–262.
- Duncan, G. J., Dowsett, C. J., Claessens, A., *et al.* (2007). School readiness and later achievement. *Developmental Psychology* **43**(6), 1428–1446.
- Garces, E., Thomas, D., and Currie, J. (2002). Longer term effects of head start. *American Economic Review* **92**, 999–1012.
- Gelbach, J. (2002). Public schooling for young children and maternal labor supply. *American Economic Review* **92**, 307–322.
- Gormley, W. Jr., Gayer, T., Phillips, D., and Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology* **41**, 872–884.
- Havnes, T. and Mogstad, M. (2009). No child left behind: Universal childcare and children's long-run outcomes. *Discussion Papers No. 582*, May 2009, Statistics Norway: Research Department.
- Lamb, M. E., Damon, W., Sigel, I. E., and Renninger, K. A. (1998). Nonparental child care: Context, quality, correlates and consequences. In *Handbook of Child Psychology*, 5th edn., vol. 4, pp 73–133. New York: Wiley.

- Lefebvre, P. and Merrigan, P. (2008). Child-care policy and the labor supply of mothers with young children: A natural experiment from Canada. *Journal of Labor Economics* **26**, 519–548.
- Magnuson, K. A., Ruhm, C., and Waldfogel, J. (2007a). The persistence of preschool effects: Do subsequent classroom experiences matter? *Early Childhood Research Quarterly* **22**, 18–38.
- Magnuson, K. A., Ruhm, C., and Waldfogel, J. (2007b). Does pre-kindergarten improve school preparation and performance? *Economics of Education Review* **26**, 33–51.
- Masse, L. and Barnett, W. S. (2003). *A Benefit-Cost Analysis of the Abecedarian Program*. New Brunswick, NJ: National Institute for Early Education Research.
- NICHD-ECCRN (2003). Does amount of time spent in childcare predict socioemotional adjustment during the transition to kindergarten? *Child Development* **74**, 976–1005.
- NICHD-ECCRN and Duncan, G. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development* **74**, 1454–1475.
- Powell, L. M. (1997). The impact of childcare costs on the labor supply of married mothers: Evidence from Canada. *Canadian Journal of Economics* **30**, 577–594.
- Cleveland, G. and Krashinsky, M. (2003). *Fact and Fantasy: Eight Myths about Early Childhood Education and Care*. Toronto: Childcare Resource and Research Unit: ISBN-1-896051-25-1.
- Cleveland, G. and Krashinsky, M. (2004). *Financing ECEC Services in OECD Countries*. OECD Directorate for Education. Paris: OECD.
- Cunha, F. and Heckman, J. (2009). The economics and psychology of inequality and human development. *National Bureau of Economic Research Working Papers No. 14695*.
- Gormley, W. T. Jr. (2007). Early childhood care and education: Lessons and puzzles. *Journal of Policy Analysis and Management* **26**(3), 633–671.
- Karoly, L. A., Greenwood, P. W., Everingham, S. S., et al. (1998). Investing in our children: What we know and don't know about the costs and benefits of early childhood interventions. *Report No. MR-898-TCWF*. Santa Monica, CA: RAND Corporation.
- OECD (Organization for Economic Cooperation and Development) (2006). *Starting Strong II: Early Childhood Education and Care*. Paris: OECD.
- Shonkoff, J. P. and Phillips, D. (2000). *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, DC: National Academy Press.
- Vandell, D. (2004). Early child care: The known and the unknown. *Merill-Palmer Quarterly* **50**(3), 387–411.

Further Reading

- Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs? *Social Policy Research* **XVII**(1), 3–15.

EARLY CHILDHOOD EDUCATION AND CARE PROGRAMS

Contents

Early Care and Education Programs for Infants and Toddlers

Literacies in Early Childhood: The Preschool Period

Participation in Early Childhood Education and Care Programs: Equity, Diversity and Educational Disadvantage

Teaching in Early Childhood Centers Instructional Methods and Child Outcomes

Early Care and Education Programs for Infants and Toddlers

B Nyland, RMIT University, Bundoora, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Early care and education programs for infants and toddlers can be defined by the wide variations that exist across countries and regions. Social attitudes to very young children will influence the type of education and care provision available and the priority that is given to programs for infants and toddlers. Increasingly research has highlighted the importance of the first 3 years and as knowledge of the social infant grows, there is increasing pressure to design programs that will have the greatest potential to support beneficial outcomes for children's development, growth and well-being. There is an increasing respect for the communicative competence displayed by infants and toddlers, and also a renewed emphasis on the family as a cultural unit, on work–family integration and demands on family members as current or future workers.

There are multiple agendas that early care and education programs are expected to meet. These include physical care, appropriate stimulation to support growth and well-being, family and workforce support, educational advantage, and early intervention for children, families and sometimes whole communities identified as disadvantaged. Infants and toddlers are no longer viewed as individuals who can easily be profiled by describing their developmental performance. Views have expanded as children are also seen as part of the culture, albeit novices, as citizens with rights and as communicators who can establish reciprocal relationships with those in their immediate environments. This powerful image of the child has come from theorists such as Malaguzzi, Magda Gerber, Piaget, and Vygotsky and examples of practice influenced by this image can be seen in the design of early childhood programs around the world, including Sweden, Italy, the United Kingdom, America, China, Australia, and New Zealand.

Globalization has altered some of the debates about group programs for infants and toddlers, which have abounded over the years. These include questions of education and care, demands for investment and regulation, development of a national curriculum appropriate for the age, increased research and a spreading of ideas about child development. Since the 1970s, there has been a demarcation in many countries about what constitutes care and education. This situation is changing. Language is changing as well to reflect recent thinking. In Spain, centers for the very young are now called *escuela infantil* emphasizing the educational role of childcare institutions. Denmark has developed training for a pedagog that combines care and education as inseparable units and the pedagog works with people from birth to old age. Many countries have integrated education programs for children from birth to school including Brazil, England, Scotland, Norway, New Zealand, and Sweden. Increasingly, early education and care is promoted as an economic consideration that is seen as an investment. Child-development theories, seen by many as a Western concept, have been more widely adopted as international agencies, such as the World Bank, Asia Development Bank, United Nations Educational, Scientific and Cultural Organization (UNESCO) and United Nations Children's Education Fund (UNICEF), support early education, health services, and the gathering of research data across the globe. Given the context of changing times, this article has been divided into the following sections:

- changing patterns of early care and education;
- research;
- changing family contexts for childrearing;
- prevailing theories, policies, and practices; and
- conclusion.

Changing Patterns of Early Care and Education

Depending on family circumstances, many infants and toddlers today are under nonparental care and education arrangements for a significant number of hours each week. Relative care is the most common, followed by home care by a nonrelative, and center-based care is the third in countries like the United States, United Kingdom, Canada, and Australia. This move outside the home for many babies has been accompanied by a renewed interest in the impact of the first years on a child's growth and well-being. There are two prevailing approaches to this acknowledgment of the value of positive experiences in the early years. One is a neoliberal view of the child as an asset to society, promoting the need to invest in early childhood services. The other is a participatory approach that would place children as protagonists in their own development, reciprocal members of the culture, and social partners establishing learning relationships.

The first approach to early care and education views the child as an asset to society and gives rise to investment models. One of these models is based on human capital theory, the proponents of which claim that high-quality education and care experience in the early years will aid the acquisition of skills for the individual and that skill begets skill. The cost of providing for children in the first 3 years will be beneficial to the individual and the society. This is an economic and political methodology and is characterized by language like scientific, evidence based, targeted interventions, learning outcomes, and mixed-mode funding. The latter may involve state money, sometimes as vouchers, being paid to the private sector along with users contributing to the service through the payment of fees. There has been criticism from the Organization for Economic Cooperation and Development (OECD), through the Starting Strong 1 and 11 reports, that these funding arrangements threaten quality, cost, and accessibility of children's services. Such an approach to the delivery of early childhood services may be favored by politicians. Targeted intervention is supported if the market cannot provide programs for all and particular populations, identified as needy, can receive services. Policymakers may find it ideologically easier to support the idea that infancy and toddlerhood are in the family domain and therefore are a private not a public good. Targeted interventions can support such a stance. The role of the state and whether early childhood services for infants and toddlers are a public good is a debate for the early childhood field, policymakers, politicians, and electorates.

A participatory approach is best known in early childhood literature through exemplary programs such as Reggio Emilia and the New Zealand early childhood curriculum *Te Whāriki*. *Te Whāriki* encompasses four principles: empowerment, holistic development, family

and community relationships, and five interwoven strands of well-being—belonging, contribution, communication, and exploration. The metaphor used is that of a woven mat. *Te Whāriki* was written as a bicultural, bilingual curriculum with space to promote a diversity of voices. The other example, the infant/toddler centers of Reggio Emilia, *nidos*, have as their cornerstone an image of the strong, competent, social infant. These infant/toddler centers have received much international attention as they have been designed for children specifically within the birth–3 age range and are an impressive example of articulating a social image of the child through physical surroundings, materials offered, and the documentation of relationships among babies and between babies and adults. This is one of the few examples of infant/toddler care that can claim to be strongly theorized for the particular age group.

There are many other approaches to group care for children from birth to age 3. A prevalent model that has had an international following is a health-focused model that emphasizes routines, health and nutrition, cognitive and language stimulation, and early intervention. Many early childhood programs are eclectic in the use of the theory on which to base practice. In England, part of the Sure Start initiative saw the establishment of local programs such as the Sure Start Local Programs (SSLPs), which integrated services, had multidisciplinary teams, and catered for local needs. Services, like those of Pen Green, combined concepts of children as protagonists in their own learning with a family-centered model and an interventionist approach to try to alleviate disadvantages within the targeted population.

There have been criticisms of psychologists having supremacy in relation to the design and establishment of early care and education programs. This may be justified as most of the theorists who have been influential in the early education field have been psychologists or from a related discipline. Among these, Piaget was particularly important as he emphasized the enormous development that occurs before formal language has become established. This led to an acknowledgment of infants as individuals who have preferences and desires. From the contributions of scholars like Piaget and Vygotsky observing and planning for very young children now has a broader base.

Infant and toddler education and care group programs are historically recent. Many of the theoretical influences are shared across borders as curriculum discourse becomes increasingly international. In the last 40 years, there has been a massive movement of people around the world and changing work demands as well as cultural interaction that have impacted upon traditional nuclear and extended family patterns. Such changes have brought about a need for out-of-home care for children as the families can no longer sustain care of the young and meet new demands. The growth of formalized childcare has occurred as a response to social change. As social

change in a globalized world is a reasonably recent phenomenon, the design of care and education programs for infants and toddlers is occurring simultaneously across the globe. Many similar influences impact on policymakers and scholars who respond to international research findings and economic demands to support the workforce and help provide the best start in life for the next generation. Current shared debates address the role of the state in providing services, issues of quality care and education, economics and social welfare, cultural and linguistic inclusion, and acknowledgment of a growing diversity of populations sharing the same services.

Research

Areas where research can support the growth and development of the early child education and care ECEC field include: (1) exploring the implications of the relationship between the environment, genetics, nature, nurture, and culture, and (2) studying the role of relationships in the development of social competence and resilience. Quantitative and qualitative research is necessary along with the emerging concept of mixed-mode research. Mixed methods research has the advantage of integrating quantitative and qualitative data to achieve sophisticated research study designs as well as combining data based on social science research and experimental psychology. For infants in group care, an advantage of combining methods has been the adding of the concept of well-being as a lens to view everyday happenings. Infants can therefore be studied in cultural contexts with an emphasis on what they can do and how they feel, and the lens includes both physical and psychological experiences. There is a greater awareness of infants' learning capabilities and how stress will affect physical and emotional growth, as well as perinatal risk factors that may impact on brain function and cognitive functioning.

In recent years, there have been a number of trends driving research impacting on infant and toddler programs. One is an emphasis on research to support practice and policy so decision making can be more evidence based. Such a research approach was enshrined in legislation in the United States in the No Child Left Behind Act of 2001. Since this legislation, there has been a revisiting of the concept of human capital theory encapsulated in the High/Scope Perry Preschool research of the 1960s and 1970s, as well as lessons from more recent scientific advances. Neuroscience, known in the field as brain research, has highlighted the threats that poor environments bring in the first 3 years and the impact upon the developing brain. Results can be measured by costs to the society in terms of skills, income, education levels, and even incarceration. This emphasis on empirical research and the privileging of the language of empiricism by

politicians and policymakers have had a profound effect. Canada and Australia have applied an early development index to aid the identification of populations at risk, warranting additional targeted interventions. Education systems have favored outcomes-based approaches, with concentration on goals, measurable standards, and the need to develop curriculum statements for early childhood settings, including those for babies and toddlers.

At the same time, qualitative research has become increasingly sophisticated and popular among educational researchers as naturalistic, interpretive inquiry designed to build understandings about meaning making in daily contexts. Qualitative research provides tools for exploring the meanings people attach to their actions and interpret their worlds. There is also a growing body of practitioner research and an interest in giving the child, including the preverbal child, a voice in the research process. This latter research has the potential to support the concept of rights as a living, active process in everyday experience.

Changing Family Contexts for Childrearing

Given that most nonparental care and education for very young children are carried out by relatives, with nonrelatives and formal institutional care being secondary forms of care, parenting programs for grandparents and others caring for infants and toddlers (and facilities like playgroups) are promoted by many governments and agencies. Research into grandparent care, for example, has highlighted gaps between services offered to parents caring for their children and other family members or friends who may be fulfilling a child-caring role. Grandparents are often the carers of choice when parents are working; however, there is an emerging pattern from the research suggesting that grandparents have their own special needs.

Changing demographics also impact upon parental care and education. Fertility rates are a concern of governments with reproductive levels driving some policies related to families. For example, China's restricted child policy has implications for early childhood experiences and family arrangements. Further, in response to low fertility rates, governments may endeavor to support and provide early childhood services as an impetus to population growth. In countries like Australia the advent of older mothers, higher education attainment, and smaller families have also led to a demand for better-quality services. However, what is high quality for infants and toddlers is a debated issue. Developmental psychologist Penelope Leach has long maintained that the best environment for the infant is home with a parent, although she acknowledges that this may be a luxury not open to many families. She argues that governments should take more responsibility

to ensure parents and babies can form strong relationships in the early months by making paid parenting leave available. Toddlers benefit from group programs, Leach claims, if they have a strong relationship with their parents.

The question of how best to support the development of babies and toddlers involves debate about paid parental leave, the quality of services, grandparents as carers, nannies and in-home services, the number of hours children are in centers, and possible developmental effects on children. There have also been studies of early childhood teachers' attitudes to infants and toddlers being placed in programs at such a young age. The prevailing wisdom is that it is the quality of early childhood programs that matters. The most considered quality factors are structural and organizational: staff training and child:staff ratios, the use of space, and the size of groups and industrial considerations about the work of childcare. All these become crucial for families seeking to ensure the right of their infants and toddlers to high-quality care and education in the early years.

Prevailing Theories, Policies, and Practices

A range of theories of childhood impinge upon local and global ECEC policies. The most dominant theories for infant and toddler programs follow the developmental approaches. Major theorists who belong to this school are Piaget, Vygotsky, and Bronfenbrenner. Piaget studied the growth of knowledge in the developing child. He viewed children as active learners who, through interactions with their surroundings, developed key concepts and understandings. Piaget's stage theory, based on children's intrinsic interests in their environment and the motivation to explore, has been criticized for taking too little notice of the role of culture in development. Whereas Piaget acknowledged the child in the culture and observed connections between language learning and culture, his key stages were predominantly based on cognitive understandings. As Piaget promulgated a theory of child competence and an awareness of the communicative nature of the infant, his theory has been associated with the cognitive revolution in education and developmentally appropriate practice (DAP), an early childhood framework promulgated by the National Association for the Education of Young Children (NAEYC).

In contrast, Vygotsky became known for promoting the concept of the social and cultural nature of learning wherein development consists of the use of signs and mediated understandings that occur through joint activity with other members of the culture. The preverbal child is communicative and can use signs. Vygotsky's work identified the developmental tasks of joint attention in infancy and toddlerhood that involved shared attention and engagement with another, for example, through eye gaze

of infants from a few weeks, and exploring objects from a few months, or pointing and social referencing by 9 or 10 months and the emergence of the first words, about the end of the first year. All of these signs are important steps in symbol use and are supported by more able members of the culture interacting with the infant. In the second year, the child consolidates the use of language as decontextualized symbols that can extend experience beyond the concrete present.

Bronfenbrenner suggested a theory of the ecological nature of human development. This model is often presented as a series of concentric circles with the infant in the middle circle. This infant is nested in an ecology which embeds relationships and experiences that gradually get further and further from the child's immediate family. The child's daily environment, and those members of the culture with whom the child shares frequent interactions, are described as the microsystem. This theory includes a number of systems. Some, such as a parent's workplace, are physically removed from the baby, yet will have an impact on its development. The macrosystem is the culture and society at large. For the infant and toddler, group education and care programs are potentially an important part of the microsystem and relationships within the setting should support the child's need for reciprocal interactions with others. Bronfenbrenner strongly supports the idea of a responsible state that will support families in their childrearing tasks.

These theories have a complex relationship to policy and practice as there is eclectic use of theory and stakeholders have divided opinions and different models. Educators often use a mix of Piagetian and Vygotskian theory with references to Bronfenbrenner, Dewey, Malaguzzi, Montessori, or Steiner. A growing discourse is on curriculum and academic outcomes, which has the potential to threaten the play-based practices that have been popular since the early twentieth century. Curriculum documents that are inclusive of infants and toddlers are becoming commonplace.

Internationally, the concept of curriculum for the early childhood years, birth–8, has become increasingly popular. This move toward formalizing learning experiences in the early years has implications for the birth–3 years where practices have not been strongly theorized. Other impacts on the efficacy of curriculum guidelines are the diversity that is found among care settings and the differing levels of experience and training of staff. Studies have found that curriculum statements on their own will not necessarily improve quality or enhance pedagogy. As learning and development occur in the social milieu, relationships are the most important attribute of the infant and toddler's early childhood experience.

The idea of lifelong learning has also had an effect on infant and toddler programs. Many countries include infants and toddlers in the learning continuum. However,

for children under 3, services can be a patchwork of care arrangements. Often in the private sector, services may attract poor regulation and have low standards and staff have little training. These factors are likely to produce poor-quality provision. As pedagogy for infants and toddlers is argued to depend for quality on positive, reciprocal relationships, the training, work conditions and child to staff ratios have to be considered core to positive pedagogies. Therefore, pedagogy has come to constitute relationships that exist between young children and other members of the culture, including ideas of agency and of the infant, or toddler, as a protagonist in their own development.

A further concept in the education discourse is the concept of rights. This is especially relevant to preverbal children in out-of-home care. These children are especially vulnerable to the social constructs of infancy. The most useful international instrument for the promotion of children's rights is the United Nations Convention on the Rights of the Child (UNCROc). This is the most ratified of the UN conventions, fostering a global discourse about rights and children and a level of political consensus that gives the convention universal relevance. The UNCROc can be divided into four sections, promotion, provision, protection, and participation, known as the four Ps. For ECEC programs, all sections are important but it is provision and participation that can be most closely linked to daily experience. In relation to provision, the UNCROc acknowledges that many parents work and a child is entitled to childcare services and facilities (article 18), all children have a right to benefit from early childhood services, (article 3, best interests of the child), and regulated high-quality care should be a right.

The participation rights espoused in articles 12, 13, 14, 15, 16, and 17 are the most innovative of the convention and it is these articles that promote the child's right to a point of view and to be heard. For infant/toddler programs, it is the responsibility of the government, the community, parents, and staff to ensure that the children's rights are upheld. Staff need to be well trained in child development, sensitive to individual children, and aware of the significance of a positive social context. Professional staff are expected to reflect on children having meaningful choices, respect a child's communicative competence, and work in partnership with families and other members of the child's culture. A rights perspective has the potential to reframe social images of the child and transform everyday relationships and practice.

Conclusion

This article has discussed ECEC programs for infants and toddlers and some of the major influences on changing patterns of care. The ECEC field, focusing on programs for children under the age of 3 until recently, has been

under theorized. Most of the attention has been focused on children in the preschool age group, 3–6 years. The terms pedagogy and curriculum are applied to everyday practice with infants and toddlers. These words need definitions specifically in relation to children of this age if they are to have meaning in the field.

The dominant theories for early childhood services are developmental theories, including Bronfenbrenner's ecologic approach. Seventeenth-century neoliberalism and economics theories favoring human capital development are also influential with governments and policymakers. An economics-oriented approach has also been presented in conjunction with the findings from the neurosciences so that both scientists and economists agree that the first 3 years are the most significant in relation to return on investment or developmental interventions.

Research to supply an evidence base for policy and practice has become more important, prompting new understandings of young children. Qualitative, quantitative, and cross-disciplinary research has brought new insights, not the least being that brain research, which confirms the first 3 years as of enormous developmental significance. Contemporary research efforts profile different voices, including those of infants and toddlers, raising awareness of the communicative competence of preverbal children. Finally, a concept of rights-based practice has been foreshadowed as a new direction for education and care programs for children from birth to age 3.

Further Reading

- Anning, A. and Bamm, M. (2008). *Improving Services for Young Children: From Sure Start to Children's Centres*. London: Sage.
- Bennett, J. (2003). Starting strong: The persistent division between care and education. *Journal of Early Childhood Research*. <http://ecr.sagepub.com/cgi/content/abstract/1/1/21> (accessed September 2009).
- Berk, L. (1999). *Infants, Children and Adolescents*. Needham Heights, MA: Allyn and Bacon.
- Brennan, M. (2007). Beyond child care – how else could we do this? Sociocultural reflections on the structure and cultural arrangements of contemporary Western child care. *Australian Journal of Early Childhood* **32**(1), pp 1–10.
- Bruer, J. (2002). *The Myth of the First Three Years: A New Understanding of Early Brain Development and Lifelong Learning*. New York: Simon and Schuster.
- Department of the Prime Minister and Cabinet (2008). *Families in Australia: 2008*.
- Fragher, J. (2006). *Childhood: Growth and Development in the First Three Years*. Melbourne: Scholastic Press.
- Gammage, P. (2008). *The Social Agenda and Early Childhood Care and Education: Can We Really Create a Better World?* The Hague. The Netherlands: Bernard van Leer Foundation.
- Gandini, L. and Edwards, C. (2001). *Bambini: The Italian Approach to Infant / Toddler Care*. New York: Teachers College Press.
- Gerber, M. (1998). *Dear Parent: Caring for Infants with Respect*. Los Angeles, CA: RIE.
- Gonzalez-Mena, J. and Eyer, D. (2004). *Infants, Toddlers and Caregivers: A Curriculum of Respectful, Responsive Care and Education*. Sydney: McGraw Hill.
- Heckman, J. and Masterov, D. (2007). The productivity argument for investing in young children. *Review of Agricultural Economics* **29**(3), 446–493.

- Kagan, J. (1998). *Three Seductive Ideas*. Cambridge: Harvard University Press.
- Lally, R., Mangione, P., and Greenwald, D. (eds.) (2006). *Concepts for Care: 20 Essays on Infant/Toddler Development and Learning*. San Francisco, CA: WestEd.
- Leach, P. (2009). *Child Care Today: Getting it Right for Everyone*. New York: Random House.
- Leavitt, R. (1995). The emotional culture of infant-toddler day care. In Hatch, J. (ed.) *Qualitative Research in Early Childhood Settings* pp 3–23. Westport: Praeger.
- Markova, G. and Legerstee, M. (2008). How infants come to learn about the minds of others. *Zero to Three* **28**(5), 26–31.
- McCain, M., Mustard, F., and Shakar, S. (2007). *Early Years Study 2: Putting Science into Action*. Toronto: Council for Early Childhood Development.
- McCaleb, M. (2004). A new learning story. Capability and vulnerability. The First Years: Nga Tau Tuatahi. *New Zealand Journal of Infant and Toddler Education* **6**(1), 5–8.
- Murray, L. and Andrews, L. (2001). *Your Social Baby: Understanding Babies Communications from Birth*. Melbourne: ACER.
- Nixon, D. and Gould, K. (1996). *Emerging: Child Development in the First Three Years*. Katoomba, NSW: Social Science Press.
- OECD (Organization for Economic Co-operation and Development) (2001). *Starting Strong: Early Education and Care Report*. Geneva: OECD.
- OECD (Organization for Economic Co-operation and Development) (2006). *Starting Strong 11: Early Education and Care Report*. Geneva: OECD.
- Penn, H. (2002). The world bank's view of early childhood. *Childhood* **9**(1), 118–132.
- Piaget, J. (1976). *The Child's Construction of Reality*. London: Routledge and Kegan Paul.
- Rochat, P. (2001). *The Infant's World*. Cambridge: Harvard University Press.
- Rogoff, B. (2003). *The Cultural Nature of Human Development*. New York: Oxford University Press.
- Shonkoff, P. and Phillips, D. (eds.) (2001). *From Neurons to Neighbourhoods: The Science of Early Childhood Development*. Washington, DC: The national Academy of Sciences.
- Valsiner, J. (1987). *Culture and Development of Children's Actions*. New York: Wiley.
- Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge: Harvard University Press.
- Woodhead, M., Faulkner, D., and Littleton, K. (1998). *Cultural Worlds of Early Childhood*. London: Routledge.
- Wittmer, S. and Petersen, S. (2009). *Infant and Toddler Development and Responsive Program Planning: A Relationship Based Approach*. Upper Saddle River, NJ: Prentice Hall.

Relevant Websites

- <http://www.education.auckland.ac.nz> – Faculty of Education-The University of Auckland.
- <http://www.zertothree.org> – Zero to Three.

Literacies in Early Childhood: The Preschool Period

M Fleer, Monash University, Melbourne, Australia

B Raban, The University of Melbourne, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In recent years, educators have looked for new ways of thinking about organizing learning for young children (e.g., Anning *et al.*, 2009). Hedegaard (2007) argues that children make sense of their environment within the context of their daily lives. Learning is situated. Children may act intuitively or enact a particular known routine during their everyday life. Through this lived experience, children build culturally relevant practices and develop working theories and concepts. However, at the psychological level, how these everyday concepts develop into literacy and numeracy concepts for preschool aged children is less well understood.

According to Vygotsky (1987), concept formation occurs when children can work with both situated everyday concepts and the academic concepts which make up the conceptual systems of literacy and numeracy. For example, children visiting a food market may not think conceptually about shopping places as being organized around a particular classification system (conceptual knowledge of comparison, systems, and sorting) or realize that they are using important literacy concepts when following arrows and pictorial cues on signage (conceptual awareness of print and symbol systems). At an everyday conceptual level, the children will draw upon their lived experience of moving around the food market simply looking for a product until they find it. However, if children have a mathematical concept of a classification system, they can be more systematic about how they move about the food market. If they have an understanding or concept that print holds meaning, then they may pay attention to the signs, and use these to guide them to the product they wish. Vygotsky (1987) argued that for children to develop conceptual understandings, they need experiences of both everyday concepts and academic or preschool/school-learned concepts of literacy and numeracy (e.g., those labeled in curriculum documents). Importantly, preschool children need knowledgeable adults who simultaneously help them pay attention to everyday contexts and who introduce them to the concepts which give meaning to their lived experiences in their moment-by-moment lives.

Hedegaard and Chaiklin (2005) state that these concepts inherent in our literacy and numeracy system of knowledge are empirical. Empirical knowledge is concerned with a single phenomenon abstracted from its context. Being able to consciously apply empirical knowledge already learned,

to an unfamiliar or new context is very useful for children. For example, knowing that a measuring instrument such as a ruler can give comparative and accurate information means that children can measure a table to see if it will fit through a doorway, rather than physically trying to move it through. The concept of measurement is developed as children experience informal and formal measurement through lived experiences, such as hearing about being a size six, or using measuring cups when cooking.

For young children, empirical knowledge that has not been connected to children's lives, or embedded in a child's world is of limited immediate worth. Children may be unable to consciously draw upon empirical knowledge in their everyday world, as they do not yet know what this knowledge can do for them. For example, learning to recognize and name single words at school or learning about classification in preschool through sorting beads assumes abstract concepts that are separated from everyday practical experiences in the real world. Learning isolated words is not necessarily going to be helpful when seeing print on a notice board in their community.

In this article, we draw upon everyday concept formation and academic concept formation in literacy and numeracy for a better understanding of how the learning of these knowledge systems can be best supported in the preschool years. We begin this discussion by drawing upon Vygotsky's (1987) theoretical works and those that have developed these ideas (Hedegaard, 2007) and discuss the importance of the adults in children's lives for mediating concept formation. We also review research literature which suggests that this is an area that has not been effectively undertaken in the past in preschools.

Concept Formation in Early Literacy and Numeracy

Hedegaard and Chaiklin (2005) suggest that in acquiring understandings of a literacy or numeracy knowledge systems, children need to gain an understanding of the central concepts that characterize particular forms of knowledge. For example, when a child understands classification, it has gained relational understandings of comparison, systems, and sorting.

To understand these important or central concepts, children must develop a conceptual model which shows the relations between the concepts that make up the

system of knowledge. In working through different, but related examples, children develop theoretical knowledge of the central concept, while keeping the related concepts visible in everyday practice. For example, in examining how to locate picture books in the local library, children learn about the classification system of their library, and develop a conceptual model for a classification system. The model can then be used in other contexts to navigate around other complex collections. That is, the children have acquired not just an everyday understanding of a library, but theoretical knowledge that libraries have classification systems to help people find specific items. As such, a child can go to any library and take into account the classification system in operation, thus transforming how it engages with its environment.

In drawing upon the work of Davydov (1988), Hedegaard and Chaiklin (2005) have advanced the theoretical idea of a double-move approach, which specifically emphasizes the “relations among children’s already acquired everyday concepts, subject-matter concepts (e.g., numeracy), and local knowledge” (p. 69). The double-move approach illustrates how educators can bring together literacy or numeracy concepts with children’s personal knowledge and their lived experiences. For example, a teacher notes a child’s everyday concept such as seeing a sign on a door (one move), at the same time as considering the academic knowledge she wishes the child to acquire such as print holds meaning (second move). By explicitly linking everyday concepts and academic concepts (double move), it is possible to develop core concepts, and through this, develop theoretical knowledge (Davydov, 2008). This double-move approach recognizes that the motives of the society and the child’s personal motives can be connected in everyday preschool and school practices:

The school child becomes oriented to topics that are valued by his [sic] parents, by the community, or that the child finds new and exciting to explore. (Hedegaard and Chaiklin, 2005: 80)

Others promote similar ideas for making literacy and numeracy learning meaningful to children. When a society specifically values literacy and numeracy, as is evident now in many industrialized communities, early childhood professionals are increasingly called upon to address specific academic concepts.

The Role of Adults in Everyday Practice

Along with children’s natural curiosity, persistence, and motives which are oriented to learning, what children learn during their first 4 or 5 years is not learned in isolation. A concept cannot be disembedded from the other known concepts which together give meaning for action.

Infants’ activities are complemented by relationships that encourage their gradual involvement in the activities of the family and community (see Rogoff, 2003). Caregivers may arrange the environment to promote learning (Schaffer, 1977: 73). A great deal of research has been conducted on how to guide children’s understanding and actions in new situations through emotional cues about the nature of the situation, nonverbal models of how to behave, verbal and nonverbal interpretations of events, and verbal labels to classify objects and events in everyday experiences (Rogoff, 1990, 2003; Walden and Ogan, 1988). Successful caregivers make efforts to build on what children know and extend their competencies by providing supporting structures (or scaffolds) for the child’s performance (Bodrova and Leong, 2001). Through bringing together everyday concepts and scientific concepts, adults are seen to be:

- supporting children to make sense of an activity by suggesting and talking about strategies that will help govern the number of steps required to solve a problem;
- engaging children in learning by making it meaningful, and keeping their attention through motivation and direction of the activity;
- encouraging and providing support for problem solving using scaffolding techniques that minimize frustration and risk, including telling and revealing;
- reinforcing the processes of problem solving and demonstrating the solution by providing examples, sharing ideas, and revisiting the process as a whole; and
- making conscious for young children the concepts that are embedded in their lived experiences, and setting the stage for abstract thinking in other contexts.

Along with these aspects of cognitive growth, social opportunities influence learning. Feeling that one is contributing something to others appears to be especially motivating (Schwartz and Brandsford, 1999). For example, young learners are highly motivated to tell and write stories of their own, and draw pictures that they can share with others. Some argue that effective mediation of young children’s learning appears to require adults to match their strategies to children’s intentions, knowledge, and understandings, and to phase in and out of more or less directive roles. Others have suggested that effective mediation occurs within the context of complete performances with more experienced others, as a form of guided participation (Rogoff, 1995).

Literacy and Numeracy: Opportunity and Development

Children’s literacy and numeracy behaviors develop as a result of social processes that are embedded in their relationships, activities, and everyday events. The vast majority of children today encounter a world of text and

numbers, and people using them. As Hall and Robinson (2000) point out, they are not confused by this symbolic world and in their own way are developing means of making sense of it. This is made easier for children because all around them are people who demonstrate the uses of literacy and numeracy – when they use it, what they use it for, why they use it, and how it is used. Adults and older siblings do this not by being explicit teachers, but simply by getting on with their daily lives.

Everyday activities embed opportunities for children to learn and develop through observation and apprenticeship. Young children learn through observation of and participation in the purposes, styles of interaction, and activities around literacy and numeracy. Yet, opportunities vary according to the community into which children are born. For instance, Neuman and Celano (2001) in their study of four neighborhoods asked the question: “What is the magnitude of the differences in access to literacy materials and resources across different income groups?” They found three times the number of retail outlets selling books, magazines, comics, etc. in the middle-income neighborhoods; and they found 44 times the number of different titles available to these children. In the schools there was twice the number of books available for children living in the middle-income neighborhood, four times the number of computers and trained librarians, while there were no trained librarians working in the low-income neighborhood. However, these researchers noted that resources alone are unlikely to improve achievement. Nevertheless, we would argue that differences in settings are likely to contribute to the considerable variations in patterns of early literacy and numeracy development, through the differential opportunities for mediated interaction with resources.

Descriptive studies of children’s early experiences of numeracy/literacy learning offer critical insights into the process of making sense of experiences with print and numbers, and of others’ attitudes toward numeracy/literacy (Heath, 1983; Purcell-Gates, 1996). Even rich environments provide limited opportunities for growth if children do not know how to interact with the artifacts around them. The modeling and demonstrations provided by surrounding adults are significant and important for children’s learning.

From a different perspective, with respect to print literacy, Heibert *et al.* (1998) from the Centre for the Improvement of Early Reading Achievement (CIERA) and other reviewers (e.g., Adams, 1990) indicate that powerful predictors of reading achievement are letter-name knowledge and phonemic awareness. However, in order to apply this knowledge in learning to read, children need to understand first the purposes and conventions of reading and writing. Merely teaching letter-name and sound knowledge without the child having an understanding of the alphabetic principle (Pearson, 1999) will not by itself make the difference.

As Clay (1975) observed:

As children learn to read and write there is a rich intermingling of language learning across levels, which probably accounts in some ways for the fast progress which the best children can make. A simplification achieved by dealing firstly with letters, then words, and finally with groups of words, may be easy for teachers to understand, but children learn on all levels at once. (p. 19)

The complexity of this intermingling of language levels that advances development in spoken and written language indicate that there will be more than a single route to literacy achievement (Konold *et al.*, 1999; Clay, 1998). This will be true for the early years of childhood and in formal schooling. For instance, with struggling readers, teachers may be tempted to postpone advanced and independent aspects of comprehension skills until lower-order skills and strategies such as fluency and accuracy are fully in place. Nothing can be more damaging to the comprehension growth of struggling readers. Indeed, Gray’s work (Gray, 1999) and that of Scull (2008) demonstrate clearly that both young children and older struggling readers benefit from explicit strategies which reveal the links across levels of language. With meanings embedded in text, this style of pedagogy addresses simultaneously, fluency, accuracy, and comprehension of what is being read.

Early Literacies and Preschool Practices

The role of the adult in children’s literacy and numeracy learning is highly significant, both at home and at preschool. However, teacher beliefs about the role of teaching concepts, determine whether or not literacy and numeracy concepts are used to give meaning to children’s everyday practices.

Demonstrating Interest/Awareness

In a study of UK preschool teachers, Hannon and James (1990) reported that the teachers surveyed did not see literacy as a central concern of the preschool curriculum. They were aware of parents’ interest in their children’s literacy development but worried about parents using inappropriate methods of teaching their children to read, and about parents putting too much pressure on children. Similar findings have been reported by Raban and Ure (1999) in Australia and Lynch (2009) in Canada. Preschool educators hold mixed views about preparing children for school. This has implications for their programming either in general or with respect to developing literacies. Clay (1998: 42) points out that some preschool teachers are committed to having children learn their letters, while others leave children unaware of such things prior to

entry into school. In research by Ure and Raban (2001), of a 156 preschool centers which were invited to join a literacy project in Australia, only 40 agreed to join the project because many centers saw literacy development as the domain of schooling and of no concern to them.

Reviewing Materials and Supports for Learning

In a US study of 3–4-year-olds and their access to literacy learning within prior-to-school contexts in Philadelphia, Neuman (1999) reported centers who refused to participate in a book-flood program. Of the 100 rooms included in the study, 21 rooms had book corners, 25 had bookshelves, 84 rooms showed print in the form of signs, alphabet letters, numbers, color names, and the like, with some of these signs at the children's eye-level, but most were not and only two rooms had writing centers.

Neuman concluded that young children need not only rich and diverse reading materials, but these need to be shared with more knowledgeable others, in order to acquire the complex set of attitudes, understandings, and behaviors associated with early literacy development.

In such print-enriched environments, however, young children were not merely passive recipients bombarded with stimuli. Rather, they appeared to be active agents in their own development; exploring, discovering, and using the physical environment as an important medium for transactions. Numerous observations often indicated that it was the children, even more than the adults, who generated the reading activity. Earlier studies also claim similar findings (Neuman and Roskos, 1990; Vukelich, 1990; Noble and Foster, 1993). The inclusion of literacy-enriched play centers increased, often dramatically, the amount of literacy-related activities in which children engaged during play. When materials only, and materials coupled with adult scaffolding were compared (Pickett, 1998), children engaged in significantly more literacy-related play when adults were present and involved.

Using Appropriate Pedagogy

The Preschool Literacy Project (Raban and Ure, 2000), conducted in 40 preschools in Australia, aimed to increase preschool teacher understandings of young children's early literacy development. Resources and activities designed to stimulate interest were introduced to give children literacy experiences that would promote their preparedness for school. Typically, primary school children in these areas of the state failed to achieve acceptable levels of literacy to support their further learning at school. However, by working with these preschool teachers over a 2-year period, supporting them in their growing understanding concerning young children's early literacy development, providing them with resources, discussion, and feedback high levels of student literacy were noted. Follow-up data reported by Raban and Coates (2004) revealed that

children from these project centers achieved higher scores on measures of both spoken and written language after 2 years in their primary schools.

In a study reported by Fleer and Raban (2006), undertaken in a childcare center to examine staff awareness of the literacy and numeracy concepts embedded in everyday center practices, it was found that the teachers could identify broadly based concepts, but many admitted that they did not actively stimulate or support literacy or numeracy learning through mediation. Rather, they assumed the resources alone would provide the learning. Evidence supporting the active role of adults has been shown in the UK Effective Provision of Preschool Education (EPPE) longitudinal study (Siraj-Blatchford *et al.*, 2002). The learning outcomes were linked to interaction strategies including sustained shared thinking time where teachers engage and extend the child's thinking, giving formative feedback during the activities in real time.

In the Harvard Home-School study, 74 young children from low-income families were followed through their preschool years from age 3 to high school. In synthesizing this study, Dickenson (2001) argues that preschool teachers who spent more time listening than talking supported children's superior language development. Successful preschool teachers were those who inclined to speak to children in ways that extended children's comments. Conversations around shared books that addressed issues such as the setting of the story, the attributes, and motivations of the characters, the order of events, and the like, also were found to be supportive of children's literacy development. The extent to which preschool teachers engaged children in intellectually challenging conversations around ideas and the meanings of new words were strongly related to measures of early literacy.

The study also found that preschool teachers were identified as of paramount importance – how they viewed their role, how they conversed with children, and the supports they provided for the children using language and literacy in varied ways. In line with the findings of Siraj-Blatchford (2004), successful preschool teachers exhibited a deep understanding of what children needed, they were skillful in providing appropriate experiences for children throughout the day, and they illustrated a willingness to expend the energy needed to support children's development. This UK research group identified these characteristics as the qualities of intentional and deliberate conceptual interactions for the early years. However, the preschool teachers in this study revealed little understanding of the developmental nature of early literacy development, of the place of oral language in supporting literacy, or of the critical role they as teachers played in fostering children's long-term language and literacy growth.

Conclusion

This article examined research literature to determine the importance of linking preschool children's everyday

concepts with early literacy and numeracy concepts. Critical for concept formation within the systems of literacy and numeracy is the close relations between children's everyday concepts they enact in practice, and the concepts that are central to literacy and numeracy knowledge. Children come to acquire conscious awareness, through deliberate and intentional interactions with adults that lead to learning and development. The theoretical writings of Vygotsky (1987) provide an important direction for understanding how concept formation can occur in the preschool years, and what kind of role the adult should take.

The literature demonstrated that successful learning in literacy and numeracy occurs when preschool teachers are active agents in:

- analyzing the everyday concepts children are experiencing through their play-based programs (as related to their everyday home and community experiences); and
- making academic literacy and numeracy concepts conscious to the child within these everyday situations.

However, the literature also showed the difficulties with promoting conscious realization of concepts in practice because the philosophical position of teachers tended to focus on taking a less-active role in concept formation. Rather than assuming that the resources or the experiences alone will provide concept formation for children, the research shows that the more knowledgeable others (early childhood professionals and families) have an important mediation role. Significantly, the research supports interactions which promote shared sustained thinking (Siraj-Blatchford, 2004) about the literacy and numeracy concepts that the children are experiencing in their everyday lives.

See also: Cultural-Historical Activity Theory; Early Childhood Curriculum and Developmental Theory; Learning in a Sociocultural Perspective; Learning Through Play; Learning to Read; Mathematics; Pedagogical Content Knowledge; Primary and Elementary/Middle Grades Reading.

Bibliography

- Adams, M. J. (1990). *Beginning to Read: Thinking and Learning about Print*. Cambridge, MA: MIT Press.
- Anning, A., Cullen, J., and Fleer, M. (eds.) (2009). *Early Childhood Education. Society and Culture*. London, UK: SAGE Publications.
- Bodrova, E. and Leong, D. J. (2001). Tools of the mind: A case study of implementing the Vygotskian approach in American early childhood and primary classrooms. *Innodata Monographs 7*. Geneva, Switzerland: UNESCO International Bureau of Education.
- Clay, M. M. (1975). *What Did I Write?* Auckland, NZ: Heinemann.
- Clay, M. M. (1998). *By Different Routes to Common Outcomes*. York, Maine: Stenhouse.
- Davydov, V. V. (1988). Problems of developmental teaching. *Soviet Education* 30(8), 6–97.
- Davydov, V. V. (1990/1972). Types of generalization in instruction: Logical and psychological problems in the structuring of school curricula. *Soviet Studies in Mathematics Education, Vol. 2: National*

- Council of Teachers of Mathematics*, Reston, Virginia, USA, [Translator, Joan Teller].
- Davydov, V. V. (2008). Problems of developmental instruction. A theoretical and experimental psychological study. In Robbins, D. and Lektorsky, V. (ser. eds.). *International Perspectives in Non-classical Psychology*. New York: Nova Science Publishers.
- Dickenson, D. K. (2001). Putting the pieces together: Impact of preschool on children's language and literacy development. In Dickenson, D. K. and Tabors, P. O. (eds.) *Beginning Literacy with Language*, pp 257–287. Baltimore, MD: Paul Brookes.
- Fleer, M. and Raban, B. (2006). A cultural-historical analysis of concept formation in early education settings: Conceptual consciousness for the child or only the adult? *European Early Childhood Education Research Journal* 14(2), 69–80.
- Gray, B. (1999). Scaffolding literacy with indigenous children in school. *SRP News* June 3. Australian Curriculum Studies Association, DETYA Canberra: Commonwealth Government of Australia.
- Hall, N. and Robinson, A. (2000). Play and literacy learning. In Barratt-Pugh, C. and Rohl, M. (eds.) *Literacy Learning in the Early Years*, pp 81–104, Sydney, NSW: Allen & Unwin.
- Hannon, P. and James, S. (1990). Parents' and teachers' perceptions of preschool literacy development. *British Educational Research Journal* 16(3), 259–272.
- Heath, S. B. (1983). *Ways With Words: Language, Life and Work in Communities and Classrooms*. Cambridge: Cambridge University Press.
- Hedegaard, M. (2007). The development of children's conceptual relations to the world, with a focus on concept formation in preschool children's activity. In Harry, D., Michael, C., and James, V. W. (eds.) *The Cambridge Companion to Vygotsky*, pp 246–275. New York, US: Cambridge University Press.
- Hedegaard, M. and Chaiklin, S. (2005). *Radical-Local Teaching and Learning: A Cultural-Historical Approach*. Denmark: Aarhus University Press.
- Heibert, E. H., Pearson, P. D., Taylor, B., Richardson, V., and Paris, S. G. (1998). *Every Child a Reader*. CIERA Ann Arbor, MI: University of Michigan.
- Konold, T. R., Juel, C., and McKinnon, M. (1999). *Building an Integrated Model of early Reading Acquisition*. CIERA. Ann Arbor, MI: University of Michigan.
- Lynch, J. (2009). Preschool teachers' beliefs about children's print literacy development. *Early Years: International Journal of Research and Development* 29(2), 199–203.
- Neuman, S. (1999). Books make a difference: A study of access to literacy. *Reading Research Quarterly* 34(3), 286–311.
- Neuman, S. and Celano, D. (2001). Access to print in low-income and middle-income communities: An ecological study of four neighbourhoods. *Reading Research Quarterly* 36(1), 8–26.
- Neuman, S. and Roskos, K. (1990). Print, play and purpose: Enriching play environments for literacy development. *Reading Teacher* 44(3), 214–221.
- Noble, E. and Foster, J. E. (1993). Play centres that encourage literacy development. *Day Care and Early Education* 21(2), 22–26.
- Pearson, P. D. (1999). A historically based review of preventing reading difficulties in young children. *Reading Research Quarterly* 34(2), 231–246.
- Pickett, L. (1998). Literacy learning during block play. *Journal of Research in Childhood Education* 12(2), 225–230.
- Purcell-Gates, V. (1996). Stories, coupons and the TV guide: Relationships between home literacy experiences and emergent literacy knowledge. *Reading Research Quarterly* 31, 406–429.
- Raban, B. and Coates, H. (2004). Literacy in the early years: A follow-up study. *Journal of Reading Research* 27(1), 15–27.
- Raban, B. and Ure, C. (1999). Continuity for socially disadvantaged school entrants: Perspectives of parents and teachers. *Journal of Australian Research in Early Childhood Education* 7(1), 27–38.
- Raban, B. and Ure, C. (2000). Literacy in the preschool: An Australian case study. In Hayden, J. (ed.) *Landscapes in Early Childhood Education: Cross-national Perspectives on Empowerment – A Guide for the New Millennium*, pp 375–392. New York: Peter Lang.
- Rogoff, B. (1990). *Apprenticeship in Thinking: Cognitive Development in Social Context*. New York: Oxford University Press.
- Rogoff, B. (1995). Observing sociocultural activity on three planes: Participatory appropriation, guided participation, and apprenticeship. In James, V. W., Pablo, D. R., and Amelia, A. (eds.) *Sociocultural Studies of Mind*, pp 139–164. Cambridge: Cambridge University Press.

- Rogoff, B. (2003). *The Cultural Nature of Human Development*. New York: Oxford University Press.
- Schaffer, H. R. (1977). *Studies in Infant–Mother Interaction*. London: Academic Press.
- Schwartz, D. and Brandsford, J. D. (1999). A time for telling. *Cognition and Instruction*, **17**(3), 249–282.
- Scull, J. (2008). *Learning to Read, Reading to Understand: Embedding Comprehension Within Reading Acquisition Processes*. Unpublished PhD Thesis. The University of Melbourne.
- Siraj-Blatchford, I. (2004). Quality teaching in the early years. In Anning, A., Cullen, J., and Fleer, M. (eds.) *Early Childhood Education. Society and Culture*, pp 137–148. London: SAGE Publications.
- Siraj-Blatchford, I., Sylva, K., Taggart, B., Sammond, P., and Melhuish, E. (2002). *Technical Paper 10 –Intensive Study of Selected Centres*. London: London Institute of Education.
- Ure, C. and Raban, B. (2001). Teachers' beliefs and understandings of literacy in the preschool: Preschool literacy project Stage 1. *Contemporary Issues in Early Childhood Education* **2**(2), 157–168.
- van Moorst, H. and Graham, S. (1995). *Kindergarten at the Crossroads: The Werribee Kindergarten Study*. Werribee, VIC: Victoria University of Technology Press.
- Vukelich, C. (1990). Where's the paper? Literacy during dramatic play. *Childhood Education* **66**(4), 205–209.
- Vukelich, C. (1994). Effects of play interventions on young children's reading of environmental print. *Early Childhood Research Quarterly* **9**, 153–170.
- Vygotsky, L. S. (1987). Thinking and speech. In Rieber, R. W. and Carton, A. S. (eds.); N. Minick (Trans.) *L. S. VYGOTSKY, The Collected Works of L.S. Vygotsky*, vol. 1, *Problems of General Psychology*, pp 39–285. New York: Plenum Press.
- Walden, T. A. and Ogan, T. A. (1988). The development of social referencing. *Child Development* **59**, 1230–1240.

Further Reading

- Bodrova, E. and Leong, D. J. (1998). Development of dramatic play in young children and its effects on self-regulation: The Vygotskian approach. *Journal of Early Childhood Teacher Education* **19**(2), 115–124.
- Chaiklin, S. and Hedegaard, M. (2009). Radical-local teaching and learning: A cultural-historical perspective on education and children's development. In Marilyn, F., Mariane, H., and Jonathan, T. (eds.) *Constructing Childhood: Global-local Policies and Practices*, pp 254–278. *World Year Book 2009*. New York: Routledge.
- Siraj-Blatchford, I. (2007). Creativity, communication and collaboration: the identification of pedagogic progression in sustained shared thinking. *Asia Pacific Journal of Research in Early Childhood Education* **1**(2), 3–23.
- Veresov, N. (2006). Leading activity in developmental psychology. Concept and principle. *Journal of Russian and East European Psychology* **44**(5), 7–25.

Relevant Websites

- <http://www.oecd.org> – Starting Strong II: Early Childhood Education and Care.
- <http://www.deewr.gov.au> – Early Childhood Learning: Supporting Children's Development.

Participation in Early Childhood Education and Care Programs: Equity, Diversity and Educational Disadvantage

M Vandenbroeck, Gent University, Gent, Belgium

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Child-centered curriculum – A curriculum or educational project that starts from a holistic view on the child and takes the child's needs, experiences and its context as its focus. It is often used in contrast to adult-centered curriculum.

Early childhood education and care – A generic term to define services for children from birth to compulsory school age, including both family based (family day care or childminders) and centre based services (e.g. child care centers and kindergarten)

Socio-economic status (SES) – A generic term, varying in how it is defined, but mostly including demographic variables such as income, educational level and profession.

Vouchers – Benefits for (most often low-income) families, that aim at enabling them to purchase ECEC places in non-funded provisions.

Introduction

In different countries, the Programme for International Student Assessment (PISA) studies have profoundly shocked the educational communities as well as policymakers – not only in cases where results were disappointing, but especially where they were showing a societal gap in educational achievements. This gap seems to run along socioeconomic and ethnic-cultural lines: children from ethnic minorities and children from poor families (and these are often – but not always – overlapping categories), more often, perform less well at school. Such findings threaten social cohesion. In postindustrial societies, social success tends to be associated with (or may be reduced to) one's possibilities on the labor market and these possibilities are increasingly dependent on educational achievement. In short: widely differential education outcomes (re)produce social inequality. A positive finding is that quality early-childhood education and care can make a difference, yet, it often does not. In this article, a first look is taken at why this is the case, or in other words: What is the problem? Then, an examination is undertaken of what quality care – in the context of equity and diversity – may mean for policy and practice.

The Problem: Disadvantage and Inequalities

According to Leseman (2002), in all Organization for Economic Cooperation and Development (OECD) countries, nearly 10–20% of the children are children with presumably normal potential, but who show developmental delays or are at risk for educational failure due to socioeconomic, cultural, and/or socio-linguistic factors. It is well documented that poverty impacts on children's development and readiness for school (for an overview, see, for instance, the UNICEF Report card 8 or Vandenbroeck, *et al.*, 2008). Poverty is harmful to the developing child – across all areas of development, and early childhood is a particularly important stage. One of the more salient examples is that children from families with low socioeconomic status (SES) at age 3 have a vocabulary of less than 500 words, while children of families with high SES have a vocabulary of almost 1200 words at the same age. It is troublesome that this gap tends to grow, rather than to reduce in school age. Recent research in neuroscience tends to confirm that the early years are crucial for the later development.

Many studies have shown the potentials of early-childhood education and care (ECEC) on children's development – especially for those who are at risk for educational disadvantage. In the US, the most renowned example is the National Institute of Child Health and Human Development-Early Child Care Research Network (NICHD-ECCRN) study showing beneficial effects on different domains of cognitive and language development (e.g., Vandell, 2002). In the United Kingdom (UK), the extensive, longitudinal Effective Provision of Pre-School Education (EPPE) study also showed that children accumulating several risk factors thrive well at school, when – prior to school – they have attended high-quality ECEC (Sylva *et al.*, 2004). These converging studies make it even more salient that the quality of care that young children receive outside their homes is one of the important pathways through which income affects children (Duncan and Brooks-Gunn, 2000). Indeed, despite the potential beneficial effects of ECEC, children from ethnic minorities and children from lower-income families are to be found, more often, in lower-quality care than those from middle-income and higher-income families (Phillips and Adams, 2001; Pungello and Kurtz-Costes, 1999). Over the last decades, it has been extensively discussed why this is the case. Whereas, initially, some scholars thought that

parental choices may explain these differences, it is now clear that reality is much more complex. Parents from all classes and ethnicities attach importance to good-quality care, but parental choices for a specific type of ECEC or for a specific provision are, to a large extent, molded by environmental constraints. Moreover, parents tend to stick to the choice they have made and tend to appreciate what they receive. Differences in preferences, in fact, reflect restricted childcare options and there is a need to criticize the concept of free choice in this matter. Put more bluntly: parents can only choose what is available to them and resign themselves to that (restricted) choice (for a more elaborated account of this discussion, see Vandenbroeck *et al.*, 2008). Wall and Jose (2004) have shown, for instance, that quality care is hardly accessible for immigrant families in Finland, France, Italy, and Portugal. Vandenbroeck *et al.* (2008) have shown that – in the case of Belgium – quality childcare is more readily available in affluent neighborhoods and that the priorities set by the management of the ECEC provisions favor double-income, white, middle-class families. The unequal distribution of quality care is also to be found in The Netherlands (Noailly *et al.*, 2007). These findings can be summarized by saying that early-childhood education and care can make a difference in the lives of young children, but that it very often does not.

Policy Matters

In conclusion, it might be said that accessibility of services is one of the first-quality criteria to be met – from the point of view of equity. Accessibility as a quality criterion includes availability, affordability, and desirability (Shlay *et al.*, 2005). This, obviously, is a matter of ECEC policy (Weinraub *et al.*, 2005). Some countries have, traditionally, considered ECEC as a private matter in which states are not supposed to intervene. Provisions mainly operate on the private market and local or central authorities complement or compensate the market mechanisms with voucher systems. This is the case for diverse countries, such as Chile (Camoy, 2002), the United States, Hong Kong and Taiwan (Lee, 2006; Yuen, 2007), and The Netherlands (Noailly *et al.*, 2007). Other countries consider ECEC as part of the formal education system (e.g., Sweden) or of the welfare system (e.g., Denmark) – in which every citizen is entitled to quality ECEC, entirely funded by local or central authorities. Many continental European countries have a mixture of both systems (OECD, 2006). Despite the claims of the former (private market-oriented) system, it is clear that – from the point of view of equity – it does not keep its promises (Moss, 2008). A first reason has to do with availability. Privatization of ECEC tends to increase the number of provisions in more affluent and more urban regions – leaving rural and underprivileged areas behind (Noailly *et al.*, 2007). A second reason has to do with quality.

One of the most important conditions for quality is the level of staff qualifications (Sylva *et al.*, 2004), while independent studies in different European countries have shown that private provisions tend to hire lower-qualified staff in order to reduce the costs (e.g., Misplon *et al.*, 2004; Osgood, 2004). The Netherlands is a case of a country where policy has shifted recently. Since 2005, ECEC provisions operate entirely on the private market. Longitudinal studies show that the quality – as measured with Infant/Toddler Environment Rating Scale Revised edition (ITERS-R) and Early Childhood Environment Rating Scale-Revised edition (ECERS-R) decreased since the privatization – especially on those items that have to do with the educational interaction between staff and children. The number of groups with an unsatisfactory quality – according to international standards – has risen from 6% to 30% (Vermeer *et al.*, 2005). In the case of the US, subsidized childcare also seems to be more efficient and effective in reaching disadvantaged children than private arrangements (Weinraub *et al.*, 2005). This is not to say that inequality is inexistent in countries that traditionally adopt a policy of state-funded ECEC (e.g., Sweden, Denmark, and France). The problem in these countries, however, is a shortage of places and a possible unequal distribution of ECEC provisions – especially for the youngest children (babies and toddlers). From an equal opportunities point of view, the solution is, rather, more subsidies than privatization.

Quality Matters

Early-childhood education matters, but not every early-childhood education matters in the same way. As the EPPE study (Sylva *et al.*, 2004) made perfectly clear, only high-quality ECEC makes a difference. The average is just not good enough. However, what constitutes high quality in a context of diversity? The EPPE study included qualitative observations in those centers that made a difference for children's outcomes. According to Siraj-Blatchford (2006) – a leading researcher in this study – respect for diversity is one of the quality criteria that correlates most with positive outcomes for disadvantaged children. This is not surprising. Although what constitutes good practice with respect for diversity may vary substantially from one setting to another (Vandenbroeck, 2007), some guidelines may be set on what exactly quality – in this matter – may be. Since the famous works of scholars including Bronfenbrenner and Vygotsky, and also thanks to the work of Barbara Rogoff and colleagues (e.g., Rogoff *et al.*, 2005), we know that cultural context plays a major role in learning processes. Laevers (1997) puts it this way: children only learn when they feel well and when they are involved in activities. Well-being and involvement are two major criteria for educational quality. This implies that children should have a feeling of belonging to the ECEC center. For many children, their introduction

with ECEC represents their first step into society. It presents them with a mirror on how society looks at them and thus how they may be looking at themselves – since it is only in a context of sameness and difference that identity can be constructed. It is in this public mirror that they are confronted with these essential and existential questions: Who am I? Is it OK to be who I am? A positive self-image is closely linked to well-being. Different scholars have documented that these essential and existential questions, with regard to identity, may be problematic for children from ethnic minorities and also how important it is that ECEC providers take into account family cultures in the curriculum (Vandenbroeck, 2001). Children are confronted with many messages about how it is to be who they are. These messages are, often, unconscious and unintentional. It may be that a child does not feel represented in all children's cultural items, such as books or dolls or other play materials; that the child's own language is not recognized as valuable; that the child's eating or sleeping habits are considered as strange; or that his or her family composition is never represented. Therefore, a child-centered curriculum is always also a family-centered curriculum. In practice, this implies that the curriculum needs to balance between two pitfalls: denial and essentialism (Preissing and Wagner, 2003; Vandenbroeck, 2001). Denial of diversity would imply that one treats all children the same. It may be common practice among educators valuing equality – especially in contexts in which a strong separation between the public and the private domains is welcomed, precisely in order not to discriminate any children (Brougère *et al.*, 2007). Yet, it is obvious that every child is different and an equal treatment, therefore, would imply that every child is treated differently. Moreover, in practice, treating every child the same may imply that the educator addresses what she (or once in a while he) considers to be an average child. Most often, this average child is what is constructed as an average through the dominance of developmental psychology: a middle-class, white child, living in a traditional nuclear family (see, for instance, Burman (1994) for a critique of this developmental average). This may easily lead to what is, sometimes labeled as racism by omission. The other (and opposite) common pitfall is essentialism. This implies that a child is reduced to its (ethnic or cultural) background. It may be common practice in traditional multicultural programs that wish to celebrate cultural diversity. It assumes that there is such a thing as Muslim practices or African culture – denying not only the huge diversity within cultures, but also the agency with which parents, as well as children, shape their own multiple belongings or multiple identities (Beck, 1997; Vandenbroeck, 2001). One cannot simply assume that a child of North African descent loves to eat tajine, refuses to have pork meat, or that her parents would appreciate that the staff addresses her in Arabic. Consequently, the curriculum of ECEC cannot be constructed upon the belief that the provision should be a

home away from home, as many parents precisely choose ECEC for those things in which it differs from the home (Vandenbroeck, Roets, and Snoeck, forthcoming). On the other side, an ECEC curriculum cannot be built on constructions of the average child either. This has far-reaching implication for daily practice. There is, indeed, a major difference between asking a child's parents to bring some music from their culture to the center, or asking them if there is any music they often play at home and they would like to share with the group. As this simple example shows, child-centered curricula can only be constructed with the participation of the child's family. A good summary of guiding principles for a respectful curriculum is given by the European Diversity in Early Childhood Education and Training (DECET) network. ECEC provisions need, according to DECET, to be places where:

Every child, parent and staff member should feel that he/she belongs. This implies an active policy to take into account family cultures when constructing the curriculum.

Every child, parent and staff member is empowered to develop the diverse aspects of his/her identities. This implies that the curriculum fosters multiple identity building and multilingualism by building bridges between the home and the institutional environment as well as with the local community.

Everyone can learn from each other across cultural and other boundaries

Everyone can participate as active citizens. This implies that staff develops an explicit anti-bias approach and takes appropriate action to involve all parents.

Staff, parents and children work together to challenge institutional forms of prejudice and discrimination. This includes a critical study of availability and access policies and structural inequities.

A Final Note on Participation

A shallow reading of the literature on anti-bias education or respect for diversity may lead to a plea for parent participation, followed by the complaint that some parents do not participate as they should. It is important to distinguish between involvement and participation. Participation may be considered as one specific way to involve parents. Very often, it entails group sessions – meetings in which one is expected either to listen to experts or to express oneself aloud. These meetings may be extremely valuable to discuss the curriculum as well as daily practices in ECEC. However, self-expression may be more common in some families than in others and, therefore, it may privilege the already privileged and reproduce existing power relations in subtle ways (Tobin, 1995; Vandenbroeck and Bouverne-De Bie, 2006). Involvement includes a wider

variety of possibilities to exchange with one another. A close, qualitative, semilongitudinal study of experiences of immigrant mothers shows that the first weeks prior to, and following upon, the start of ECEC play a significant role in constructing reciprocal relationships of trust (Vandenbroeck *et al.*, forthcoming). This first – mutual – habituation period also appears to be crucial for the construction of a sense of belonging, both for children and their parents. Involvement is, then, understood as the continuous daily negotiation of educational practices between staff and parents. Another example of how involvement goes far beyond participation is described eloquently by Dahlberg and Moss (2005), referring to pedagogical documentation and the practice in Reggio Emilia (see also Rinaldi, 2005). Pedagogical documentation means the concrete documentation of daily practices, making practices visible, and, therefore, subject to reflection within staff meetings as well as with parents. Dahlberg and Moss describe it as democratic practices, in which parents and staff try to make meaning of these practices, acknowledging that there are always multiple perspectives to look at educational practices. This asks for a reflexive, researching attitude from the staff, since pedagogical practices do not have a meaning by themselves – an essence – but meaning is made or created through the perspectives of the (multiple) people involved. This may very well be at the core of what building a curriculum – based on equity and diversity – may mean, when we acknowledge that respect for diversity is not about tolerating what is deviant from the norms, but about problematizing the norms that create deviance.

Bibliography

- Beck, U. (1997). Democratisation of the family. *Childhood* 4(2), 151–168.
- Brougère, G., Guénif-Souilamas, N., and Rayna, S. (2007). De l'usage de l'altérité pour comprendre le préscolaire. In Brougère, and Vandenbroeck, M. (eds.) *Nouveaux paradigmes pour repenser l'éducation des jeunes enfants*, pp 263–284. Bruxelles: Peter Lang.
- Burman, E. (1994). *Deconstructing Developmental Psychology*. London: Routledge.
- Camoy, M. (2002). Lessons of Chile's voucher reform movement. *Rethinking Schools*.
- Dahlberg, G. and Moss, P. (2005). *Ethics and Politics in Early Childhood Education*. London: Routledge.
- Duncan, G. and Brooks-Gunn, J. (2000). Family poverty, welfare reform, and child development. *Child Development* 71(1), 188–196.
- Laevers, F. (1997). Assessing the quality of childcare provision: "Involvement" as criterion. *Researching Early Childhood* 3, 151–165.
- Lee, I. F. (2006). Illusions of social democracy: Early childhood educational voucher policies in Taiwan. In Bloch, M., Kennedy, D., Lightfoot, T., and Weyenberg, D. (eds.) *The Child in the World/The World in the Child. Education and the Configuration of a Universal, Modern, and Globalized Childhood*. New York: Palgrave Macmillan.
- Leseman, P. (2002). *Early Childhood Education and Care for Children from Low-Income or Minority Backgrounds*. Paris: OECD.
- Misplon, S., Hedeboew, G., and Pacolet, J. (2004). *Financiële leefbaarheid van de minicrèches*. Leuven: HIVA.
- Moss, P. (2008). *Markets and Democratic Experimentalism. Two Models for Early Childhood Education and Care*. Gütersloh: Bertelsmann Stiftung.
- Noailly, J., Visser, S., and Grout, P. (2007). *The Impact of Market Forces on the Provision of Childcare: Insights from the 2005 Childcare Act in the Netherlands*. Den Haag: Centraal Planbureau.
- Organisation for Economic Co-operation and Development (OECD) (2006). *Starting Strong II. Early Childhood Education and Care*. Paris: O.E.C.D.
- Osgood, J. (2004). Time to get down to business? The responses of early years practitioners to entrepreneurial approaches of professionalism. *Journal of Early Childhood Research* 2(1), 5–24.
- Phillips, D. and Adams, G. (2001). Child care and our youngest children. *Future of Children* 11(1), 35–51.
- Preissing, C. and Wagner, P. (2003). *Kleine Kinder, keine Vorurteile? Interkulturelle und vorurteilsbewusste Erziehung in Kindertageseinrichtungen*. Freiburg: Verlag Herder.
- Pungello, E. and Kurtz-Costes, B. (1999). Why and how working women choose child care: A review with a focus on infancy. *Developmental Review* 19, 31–96.
- Rinaldi, C. (2005). In *Dialogue with Reggio Emilia: Listening, Researching and Learning*. London: Routledge.
- Rogoff, B., Moore, L., Najafi, B., et al. (2005). Children's development of cultural repertoires through participation in everyday routines and practices. In Grusec, J. and Hastings, P. (eds.) *Handbook of Socialization*, pp 490–515. New York: Guilford.
- Shlay, A., Tran, H., Weinraub, M., and Harmon, M. (2005). Teasing apart the child care conundrum: A factorial analysis of perceptions of child care quality, fair market price and willingness to pay by low-income, African-American parents. *Early Childhood Research Quarterly* 20, 393–416.
- Siraj-Blatchford, I. (2006). The impact of early childhood settings and families on children's development. *Paper Presented at the Promoting Diversity through Education*, Bratislava.
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., and Taggart, B. (2004). The effective provision of preschool education (EPPE) project: *Final Report*. Nottingham: DfES Publications - The Institute of Education.
- Tobin, J. (1995). The irony of self-expression. *American Journal of Education* 103, 233–258.
- Vandell, D. (2002). Early child care and children's development prior to school entry: Results from the NICHD study of early child care. *American Educational Research Journal* 39, 133–164.
- Vandenbroeck, M. (2001). *The View of the Yeti: Bringing up Children in the Spirit of Self-Awareness and Kindredship*. The Hague: Bernard Van Leer Foundation.
- Vandenbroeck, M. (2007). Beyond anti-bias education: Changing conceptions of diversity and equity in European early childhood education. *European Early Childhood Education Research Journal* 15(1), 21–35.
- Vandenbroeck, M. and Bouverne-De Bie, M. (2006). Children's agency and educational norms. A tensed negotiation. *Childhood* 13(1), 127–143.
- Vandenbroeck, M., De Visscher, S., Van Nuffel, K., and Ferla, J. (2008). Mothers' search for infant child care: The dynamic relationship between availability and desirability in a continental European welfare state. *Early Childhood Research Quarterly* 23(2), 245–258.
- Vandenbroeck, M., Roets, G., and Snoeck, A. (2009). Mothers crossing borders: immigrant mothers on reciprocity, hybridisation and love. *European Early Childhood Education Research Journal* 17(2), 203–216.
- Vermeer, H. J., van Ijzendoorn, M. H., de Kruij, R., et al. (2005). *Kwaliteit van de Nederlandse kinderdagverblijven: trends in de kwaliteit in de jaren 1995-2005*. Leiden - Amsterdam - Nijmegen: Nederlands Consortium Kinderopvang Onderzoek.
- Wall, K. and Jose, J. S. (2004). Managing work and care: A difficult challenge for immigrant families. *Social Policy and Administration* 38(6), 591–621.
- Weinraub, M., Shlay, A., Harmon, M., and Tran, H. (2005). Subsidizing child care: How child care subsidies affect the child care used by low-income African-American families. *Early Childhood Research Quarterly* 20, 373–392.
- Yuen, G. (2007). Vouchers in Hong Kong: A new milestone in early childhood education? *Contemporary Issues in Early Childhood* 8(4), 355–357.

Further Reading

Himmelweit, S. and Sigala, M. (2004). Choice and the relation between identities and behaviour for mothers with pre-school children: Some implications for policy from a UK study. *Journal of Social Policy* **33**(3), 455–478.

Huston, A. C., Chang, Y. E., and Gennetian, L. (2002). Family and individual predictors of child care use by low-income families in different policy contexts. *Early Childhood Research Quarterly* **17**, 441–469.

Mistry, R., Biesanz, J., Taylor, L., Burchinal, M., and Cox, M. (2004). Family income and its relation to preschool children's adjustment for families in the NICHD study of early child care. *Developmental Psychology* **40**(5), 727–745.

Relevant Website

www.decet.org – Diversity in Early Childhood Education and Training.
www.unaglobal.org

Teaching in Early Childhood Centers Instructional Methods and Child Outcomes

I Siraj-Blatchford, Institute of Education, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Curriculum differentiation – The provision of differentiated curriculum experiences as a response to the identification to the specific needs of individual or small groups of children.

ECERS-E – (Early Childhood Environment Rating Scale: Extension) A rating scale of Four subscales which assesses pedagogy and the curriculum within the setting, including the areas of mathematics, science, literacy, and diversity (whether staff plan to meet particular/individual needs).

ECERS-R – (Early Childhood Environment Rating Scale: Revised) A rating scale consisting of seven subscales which provide an overview of the preschool environment, covering aspects of the setting from furnishing to individuality of care and the quality of social interactions.

Formative assessment – The process whereby information is collected to identify the extent of children's learning, and subsequently applied in the provision of feedback and in adapting the curriculum and pedagogy to provide for their particular needs.

Multilevel modelling – A methodology that allows data to be examined simultaneously at different levels within a system (i.e., young children within preschool centres/primary schools, within LAs), essentially an extension of multiple regression.

Pedagogical content knowledge – Different pedagogic techniques are often required to make different forms of knowledge, skill, and understanding accessible to young children. In teacher education, the identification of appropriate strategies is often referred to as 'pedagogical content knowledge'. Requires knowledge of the 'subject' being taught and the child's level of learning.

Pedagogical interactions – Face-face interactions practitioners engage in with children; they may take the form of cognitive or social interactions.

Sustained shared thinking – An episode in which two or more individuals 'work together' in an intellectual way to solve a problem, clarify a concept, evaluate activities, extend a narrative, etc. Both parties must contribute to the thinking and it must develop and extend.

Systematic observations – Analysis from the timed *Target Child Observations*.

Introduction

This article summarizes the key findings of two major English studies related to teaching and learning in early-years centers. The Effective Provision of Preschool Education (EPPE) project identified centers that demonstrated good-to-excellent child-developmental outcomes, and the Researching Effective Pedagogy in the Early Years (REPEY) project (Siraj-Blatchford *et al.*, 2002, 2003) extended the EPPE analysis in an effort to identify the working practices that contributed most significantly to those outcomes.

The EPPE research project was initially funded as a 5-year longitudinal study following the progress of approximately 3000 children – from the age of 3 – in 141 preschools across England. (The studies were both funded by the Department for Education and Skills (DfES, 1997-2003). Subsequent funding has extended the study of the EPPE cohort and the current investigation is the Effective Pre-school, Primary and Secondary Education (EPPSE 3-14) project.) In addition to the range of quantitative data collected about children, their families, and their preschool centers, 12 effective preschool settings (selection based on child-developmental outcomes) were identified in the EPPE multilevel analysis for intensive, in-depth qualitative case study. The case studies were to reveal those teaching practices that are associated with better child outcomes. The REPEY study provided an opportunity to extend the case-study analysis further with both naturalistic and systematic observations, interviews, and focus-group discussions. Together, the studies provide a large-scale, sequential and explanatory mixed-method research-design investigating pedagogy in each of the Foundation Stage settings (at that time, providing for children in the age range of 3–5 in England).

This research has shown that the most effective pre-school settings (in terms of intellectual, social, and dispositional outcomes) enable children to spend two-thirds of their time in child-initiated activities – further achieving a balance between the pedagogical efforts provided for children in teacher-initiated group work and in the children's freely chosen (and potentially instructive) play

activities. The analysis also shows an association between children's outcomes and the frequency of sustained shared thinking observed between an adult and a child (or small group). Each of these features was found to contribute to the cognitive challenge. The evidence suggests that the better a setting does on each of these dimensions of pedagogical practice, the more effective it will be.

Researching the Effectiveness and Quality of Center-Based Practice

Children (and their families) were recruited into the EPPE study when they were 3 years old – the age at which many enter their first Foundation Stage group setting. Developmental progress was assessed regularly – beginning when the children entered the study and continuing through school entry – and at the end of their first and second years. Quantitative analysis (multi-level modeling) was applied to assess the contribution of the preschool setting to a child's cognitive and social/behavioral development – having controlled for child factors (e.g., gender, health) and family background characteristics (e.g., socioeconomic status, mother education). Full details of the research design may be found in EPPE Technical Paper 1 and in the Researching Effective Pedagogy in the Early Years (REPEY) Project Report (Siraj-Blatchford *et al.*, 2002).

The representative sample of 141 preschool settings were drawn from a range of providers (local authority day-nursery, integrated centers, playgroups, private day-nurseries, maintained nursery schools, and maintained nursery classes). A sample of home children (who had no, or minimal, group preschool experience) was also recruited to the study at entry to school for comparison with the preschool group. Analyses of the quantitative data collected on every child in the study revealed that, in some preschool centers, children made progress as expected, or better (or worse) than expected, given their individual and home characteristics. Settings for the case-study research were, therefore, selected based on their child-outcome data. This enabled an examination of the variation in child outcomes between centers – and the range of outcomes within centers – on eight identified cognitive and social development outcomes.

In addition to investigating the effects of preschool provision on young children's development, EPPE explored the quality of provision using the Early Childhood Environment Rating Scale: Revised (ECERS-R: Harms *et al.*, 1998) and the Early Childhood Environment Rating Scale: Extended (ECERS-E: Sylva *et al.*, 2003). EPPE has, therefore, been able to demonstrate the positive effects of high-quality provision on children's intellectual and social/behavioral development by describing gains during the preschool period (see Sammons *et al.*, 2002, 2003). Medium- and high-quality provision, as measured by the ECERS-E, has

continued to show a significant influence on attainment through to age 11, when disadvantaged children, in particular, have been found to benefit in terms of literacy and mathematics (Sylva *et al.*, 2008).

Quality and Diversity

In assessing the quality of provision, EPPE employed two rating scales – the ECERS-R and the ECERS-E. The ECERS-E was devised after wide consultation with UK experts and extensive piloting of the instrument. The word environment in both rating scales is taken in its broadest sense to include social interactions, pedagogical strategies, and relationships between children as well as adults and children. The construction of this new scale provided an opportunity to address the issue of the quality of provision for inclusion and diversity. The instrument was, therefore, developed to include a Diversity sub-scale along with those rating the provisions made for literacy, mathematics, and science.

The Diversity sub-scale recorded provision and planning for individual and cultural needs, and for the provision of positive role models. High scores on this sub-scale reflect a recognition of the importance of having high expectations for all children irrespective of socioeconomic class, ability, gender, or ethnicity. In reporting on the associations found between the qualities of provision for diversity made by the 141 preschools, and the children's cognitive and social outcomes, EPPE has provided the first evidence of the overall effectiveness of these curriculum approaches.

The study found that most early-childhood settings provided a relatively low-quality learning environment for children in terms of diversity (Sylva *et al.*, 1999). However, the quality of diversity provision was highest in combined centers (those that fully integrated care and education) and in some nursery schools where the overall ECERS-E scores were also relatively high. In terms of outcomes, strong patterns of association were found between the ECERS-E sub-scale for Diversity and children's attainment in Early Number and Non-Verbal reasoning. A positive relationship with prereading was also verging on statistical significance. In terms of social/behavioral outcomes, high scores on the Diversity sub-scale were also found to be correlated with Independence and (not quite significantly) with Co-operation and conformity. Diversity was, therefore, found to be associated with as many as five of the nine attainment outcomes. This was higher than for any of the other ECERS-R or ECERS-E sub-scales.

The Qualitative Case Studies

All of the settings selected for case study demonstrated a range of practices and all of them demonstrated some

above average outcomes. The settings were chosen from a range identified as good (if their children made slightly more developmental progress than expected based on their individual child and home characteristics) to excellent (where children made significant developmental progress above their projected developmental progress). The report, therefore, consistently refers to settings throughout as good (above average) or excellent (well above average), based on the child-outcome data.

In conducting the case studies, trained researchers – who were already familiar with the centers – spent 2 weeks in each center. Case-study data came from multiple sources to allow for assessment by source and the method of data collection. Data from policy documents was triangulated with manager and parent interviews, extensive naturalistic observations of staff (over 400 h), and systematic, focal, child observations of children (254 target child observations; Siraj-Blatchford *et al.*, 2002, 2003). The initial analysis was grounded and carried out blind, and this was, subsequently, built upon in an iterative process of triangulation with the ECERS quality, and other quantitative findings. Further details of the analytic procedures may be found in Siraj-Blatchford *et al.* (2003) and Siraj-Blatchford *et al.* (2006). This systematic process of iterative triangulation characterized the general approach that was taken in adopting a sequential explanatory mixed method approach to the study (Creswell, 2003).

A number of significant correlations were, in this way, identified between particular sub-scales and items in the ECERS, and the setting residuals (the complex value-added, multilevel model) analysis for each outcome (see Sammons *et al.*, 2002, 2003). This analysis has gone a long way in providing explanations for the patterns and associations between particular practices (as measured by the ECERS) and developmental outcomes (see Sammons *et al.*, 2002, 2003). Four patterns of association were identified for special attention, and a closer analysis of the data from systematic observations suggested that each of the following practices should be investigated further:

- Adult–child verbal interactions.
- Differentiation and formative assessment.
- Discipline and adult support in talking through conflicts.
- Parental partnership with settings and the home-education environment.

Pedagogical Practices Producing Improved Child Outcomes

Adult–Child Interactions

The excellent settings encouraged sustained shared thinking. This refers to an episode in which two or more individuals work together, in an intellectual way, to solve a

problem, clarify a concept, evaluate activities, extend a narrative, etc. Both parties must contribute to the thinking and it must develop and extend thinking. The research found that this does not happen very frequently. In excellent settings, there were significantly more sustained shared-thinking interactions occurring between staff and children than in the good settings. When it did occur, it extended children's thinking. Investigations of adult–child interaction have suggested that periods of sustained shared thinking are a necessary prerequisite for the excellent early-years practice – especially where this is also encouraged in the home through parental support.

In the excellent case-study settings, the importance of staff members extending child-initiated interactions was also clearly identified. In fact, almost half of all of the child-initiated episodes which contained intellectual challenge included interventions from a staff member to extend the child's thinking. The evidence also suggests that adult modeling is, often, combined with sustained periods of shared thinking, and that open-ended questioning is also associated with better cognitive achievement. (The process where early years educators provide a model in terms of their language, behaviors, skills, and/or attitudes for young children to imitate.) However, open-ended questions made up only 5.1% of the questioning used in the 14 case-study settings.

In the excellent settings, the balance of who initiated the activities (staff member or child) was nearly 2 times child initiated, revealing that the pedagogy of the excellent settings encourages children to initiate activities more often than the staff. In addition, staff regularly extended around two-thirds of all observed episodes and around one-half of all child-initiated activities, but did not dominate them. However, the children in reception classes experienced a different balance of initiation – with a much greater proportion of staff-initiated episodes. In all of the case-study settings, children spent most of their time in small groups. But observations show that sustained shared thinking was most likely to occur when children were interacting 1:1 with an adult or with a single peer partner. Freely chosen play activities, often, provided the best opportunities for adults to extend children's thinking. Adults need, therefore, to create opportunities to extend child-initiated playful learning as well as teacher-initiated small-group work – as both have been found to be important vehicles for promoting learning.

The findings reveal that qualified staff (i.e., almost all trained teachers) provided children with more experience of academic activities (especially language and mathematics) and they encouraged children to engage in activities with higher cognitive challenge. While the most highly qualified staff also provided the most direct teaching (instruction through demonstration, explanation, questioning, modeling, etc.), they were the most effective in their interactions with the children, using the most

sustained shared thinking. Furthermore, less well-qualified staff were found to be better pedagogs when they worked alongside qualified teachers.

Differentiation and Formative Assessment

The analysis of teacher observations suggests a positive association between curriculum differentiation, formative assessment, and the process of selecting activities to provide optimal cognitive challenge, and sustained shared thinking. The practice of adults modeling (or demonstrating) positive attitudes, behaviors, and the appropriate use of language, has also been identified as a valuable pedagogical strategy to be employed in early childhood. The best case-study settings maintained good records and engaged with parents with regard to their child's progress on a weekly or monthly basis. However, there was little evidence of detailed formative feedback to children during tasks. Feedback is a pedagogical strategy known to be effective in advancing cognitive outcomes in primary education.

Discipline and Adult Support in Talking Through Conflicts

The excellent settings adopted discipline/behavior policies that engaged staff in supporting children to rationalize and talk through their conflicts. In other words, a more problem-solving approach was taken. Three settings with positive social and behavioral outcomes had strong behavior-management policy which supported this practical approach. In settings that were less effective in this respect, observations showed that there was, often, no follow-up on children's misbehavior or conflicts and, on many occasions, children were distracted or simply told to stop.

Parental Partnership

The case studies indicate that, where a special relationship in terms of shared educational aims had been developed with parents – and pedagogic efforts were made by parents at home to support children – sound learning can take place even in the absence of consistently good pedagogical practice in the preschool setting. The excellent settings shared child-related information between parents and staff, and parents were often involved in decision-making concerning their child's learning program. This level of communication was particularly the case in private day nurseries. While settings providing for the needs of children from the higher socioeconomic groups benefited especially from this, the potential benefit of adopting a combined approach (good pedagogical practice within the setting and support for the home-learning environment) in settings serving more disadvantaged areas is also clear.

In more disadvantaged areas, staff in settings had to be proactive in influencing and supporting the home-education environment in order to support children's learning.

The evidence suggests that the excellent settings in disadvantaged areas recognize the importance of – and were proactive in – encouraging strong parental involvement in the educational process, by taking the time to share their curriculum, pedagogical strategies, and educational aims with parents. They offered advice on how parents could complement this within the home-learning environment and how this impacted on young children's development.

Some Overriding Conditions for Success

Knowledge of the Curriculum and Child Development

The case studies show that practitioners' knowledge and understanding of the particular curriculum area that is being addressed, are vital. Different pedagogical techniques are, often, required to make some forms of knowledge, skill, or understanding accessible to young children. A good grasp of the appropriate pedagogical content knowledge is a vital component of pedagogy and is just as important in the early years as at any later stage of education. Even in these good and effective settings, there were examples of inadequate knowledge and understanding of curriculum areas, especially in the teaching of phonological skills and science.

The study reveals that early-years staff may need support in developing their pedagogical content knowledge in the domains of the Early Learning Goals. Educators who demonstrate good pedagogical content knowledge display a firm knowledge and understanding of their curriculum content, but crucially, the most effective educators also demonstrated a knowledge and understanding of what part of that content was most significant and relevant to the needs of the children that they were teaching. They were also able to draw upon knowledge of the pedagogical strategies found to be most effective in teaching any particular content.

Management and Staff

The data showed that all the case-study preschool settings had strong leadership and long-serving staff. Most of the managers and staff had been in the settings for over 3 years. The private nurseries in the case-study sample had stability of staffing with retention between 3 and 9 years. In the other settings, staff – especially senior management – had been in their post even longer, and 10–20 years of service was not uncommon.

All the managers took a strong lead – especially in curriculum and planning. In most of the settings, the leadership was characterized by a clear philosophy for the setting that was shared by everyone working in the setting. The managers of the excellent centers also had a strong educational focus, valued the importance of

adult–child interaction, and supported their staff to develop better ways of engaging children. In excellent centers, the staff were encouraged to attend staff-development sessions. Although there was a great deal of variation in the training offered and what staff were able to access.

Summary of Findings

In summary, effective pedagogy – in the early years – was found to involve both the kind of interaction traditionally associated with the term teaching, and also the provision of instructive learning environments and routines.

The excellent settings provided both teacher-initiated group work and freely chosen, yet potentially instructive, play activities. Children's cognitive outcomes appear to be directly related to the quantity and quality of the teacher/adult-planned-and-initiated focused group work for supporting children's learning. The settings that viewed cognitive and social development as complementary, managed to achieve the best outcomes for children. Trained teachers were most effective in their interactions with children, using the most sustained shared-thinking interactions. Less well-qualified staff were better pedagogs when qualified teachers supported them.

The research showed that the best outcomes for children are linked to early years settings that:

- View cognitive and social development of children as complementary and do not prioritize one over the other.
- Have strong leadership and long-serving staff (3 years plus, this was even the case in the private day-care settings where the turnover of staff is normally the highest).
- Provide a strong educational focus with trained teachers working alongside and supporting less qualified staff.
- Provide children with a mixture of practitioner-initiated, small-group work and learning through freely chosen playful activities.
- Provide adult–child interactions that involve sustained shared thinking and open-ended questioning to extend children's thinking.
- Have practitioners with good curriculum knowledge combined with a knowledge and understanding of how young children learn.
- Have strong parental involvement, especially in terms of shared educational aims with parents.
- Provide formative feedback to children during activities and provide regular reporting and discussion with parents with regard to their child's progress.
- Ensure behavioral policies in which staff support children in rationalizing and talking through their conflicts.
- Provide differentiated learning opportunities that meet the needs of particular individuals and groups of children, for example, bilingual, special needs, and girls/boys.

Conclusions

The case studies have shown how diverse the early-years settings are in England. They have shown that there is no level-playing field, in terms of the training of staff, staff salaries and conditions of service, adult–child ratios, resources, or accommodation. EPPE identified moderate-to-excellent settings from among all the types of providers. However, few playgroups and local authority day-care settings demonstrated better child outcomes. Given the variation in staff pay, training, and development, this is unsurprising. In spite of this, the case-study centers all projected a positive ethos:

1. All case-study centers had a warm, caring, safe, secure, and supportive approach to their children. All the settings engaged children in a range of different groupings, individual and group play, group-focused table-top activities, interest areas, and class snack and story times.
2. All case-study settings had a welcoming appearance. The displays on the whole reflected the children's work. Children were, generally, treated with respect. The centers were warm and inviting places. The Staff appeared calm and engaged well with the children. All these centers were fairly well resourced and, although not always ideal, had enough space. However, the outdoor play environments varied greatly.

Many of the EPPE and REPEY findings echo those reported by Hohmann and Weikart (1995) and may be considered broadly consistent with the High/Scope curriculum program that emphasizes interaction, reasoning, reflection, and responsibility for self-learning. In High/Scope, efforts are made to ensure that learning experiences are developmentally appropriate for each child, and appropriateness is defined as the extent to which a learning activity:

- Exercises and challenges the capacities of the learner.
- Encourages and helps the learner to develop unique patterns of interest, talents, and reach goals.
- Presents experiences in which the learner is able to master, generalise, and retain concepts, skills, and knowledge and which relate to previous experiences, whilst linking to future learning expectations. (Adapted from Hohmann and Weikart, 1995:15)

Indeed, there are common characteristics in several, particularly popular and successful international early-childhood education (ECE) models of Early Childhood Education that may be considered consistent with the EPPE findings. If one considers the accounts of the three ECE models most clearly identifying their pedagogy in the Starting Strong Report (see **Table 1**; OECD, 2004), it can be seen that the particular strategies applied according to these accounts of the models by Ferre Leavers (Effective Early Learning), David Weikart (High-Scope), and

Table 1 OECD curriculum outlines

	<i>Teacher's initiating activities</i>	<i>Teacher's extending activities</i>	<i>Differentiation and formative assessment</i>	<i>Relationships and conflict between children</i>	<i>Sustained shared thinking</i>
<i>EEL</i> ¹	"Introducing new activities"	"Enriching interventions"	"Observe children"	"Work out sustaining relations"	"Engagement"
<i>High Scope</i>	"Sharing Control"	"Participation as partners"	"Plan- Do-Review"	"Adopt a problem solving approach"	"Authentic dialogue"
<i>Reggio Emilia</i>	"Development of short and long-term projects"	"Sustaining the cognitive and social dynamics"	"Teachers first listen don't talk"	"Warm reciprocal relationships"	"Reciprocity of interactions"
<i>EPPE/ EPEY</i>	<i>Correlations found with effective practice</i>	<i>Correlations found with effective practice</i>	<i>Correlations found with effective practice</i>	<i>Correlations found with effective practice</i>	<i>Correlations found with effective practice</i>

¹Effective Early Learning (EEL) (Pascal and Bertram, 1995), referred to as 'Experiential Education' (EXE) in Pramling *et al.* (2004).

Carla Rinaldi (Reggio Emilia) correlate very closely with the evidence from EPPE and REPEY:

Of course, each of the approaches should be considered ideal types and the practices in many settings will involve a combination of all of these approaches. The challenge for early-childhood educators is to provide a gradual and supportive transition – as the children become more capable – that stimulates learning and development while avoiding any risk of regression or failure.

The Early Years Foundation Stage (EYFS) guidance in England promotes a pedagogy which involves negotiating and co-constructing the curriculum through playful processes of sustained shared thinking that may be initiated by either the adult or the child. The question of who initiates this sustained shared thinking is actually irrelevant as long as both parties are committed to playing an equal part in determining its focus and direction (i.e., its co-construction) in collaborative free flow. In a sense, initiation is taken in turns as different material and symbolic resources are drawn upon, and each play is extended as a more or less unique improvisation. As children develop the capability and are motivated to play with peers, the EYFS guidance encourages one to continue to provide children with a rich range of experiences and resources to draw upon in their play and to support them in developing a greater awareness of their development and learning. Ultimately, children take pleasure in learning for its own sake and restrict their play to scheduled playtimes, more disciplined creative activities, and their involvement in a variety of games with more formal rules (Siraj-Blatchford, 2008; 2009).

Bibliography

Creswell, J. W. (2003). *Research Design Qualitative and Quantitative Approaches*, 2nd edn. London: Sage.

- Harms, T., Clifford, R. M., and Cryer, D. (1998). *Early Childhood Environmental Rating Scale*, revised edn (ECERS-R). New York: Teachers College Press.
- Hohmann, M. and Weikart, D. P. (1995). *Educating Young Children*. Ypsilanti, MI: High/Scope Press.
- Organisation for Economic Co-operation and Development (OECD) (2004). *Starting Strong: Curricula and Pedagogies in Early Childhood Education and Care: Five Curriculum Outlines*, Directorate for Education, OECD. <http://www.oecd.org/dataoecd/23/36/31672150.pdf> (accessed September 2009).
- Siraj-Blatchford, I. (2008). Understanding the relationship between curriculum, pedagogy and progression in learning in early childhood in Hong Kong. *Journal of Early Childhood Education* 7(2), 3–13.
- Siraj-Blatchford, I. (2009). Conceptualising progression in the pedagogy of play and sustained shared thinking in early childhood education: A Vygotskian perspective. *Educational and Child Psychology* 26(2) June (in press).
- Siraj-Blatchford, I., Sammons, P., Sylva, K., Melhuish, E., and Taggart, B. (2006). Educational research and evidence-based policy: The mixed method approach of the EPPE project. *Evaluation and Research in Education* 19(2), 63–82.
- Sammons, P., Sylva, K., Melhuish, E., *et al.* (2002). *The Effective Provision of Pre-School Education (EPPE) Project: Technical Paper 8a – Measuring the Impact of Pre-School on Children's Cognitive Progress*. London: Institute of Education, University of London.
- Sammons, P., Sylva, K., Melhuish, E., *et al.* (2003). *The Effective Provision of Pre-School Education (EPPE) Project: Technical Paper 8b – Measuring the Impact of Pre-School on Children's Social/ Behavioural Development*. London: Institute of Education, University of London.
- Siraj-Blatchford, I., Sylva, K., Muttock, S., Gilden, R., and Bell, D. (2002). *Researching Effective Pedagogy in the Early Years DfES: Research Report 365* (Queen's Printer). London: HMSO.
- Siraj-Blatchford, I., Sylva, K., Taggart, B., Sammons, P., and Melhuish, E. (2003). *The EPPE Case Studies – Technical Paper 10*. London: Institute of Education, University of London/Department for Education and Employment.
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., and Taggart, B. (2008). *The Effective Pre-School and Primary Education (EPPE 3–11) Project: Final Report. A Longitudinal Study Funded by the DCSF*. London: DCSF/Institute of Education, University of London.
- Sylva, K., Siraj-Blatchford, I., Melhuish, E., Sammons, P., and Taggart, B. (1999). *Technical Paper 6 – The Effective Provision of Pre-School Education (EPPE) Project: Characteristics of the Centres in the EPPE Sample: Observational Profiles*. London: DfEE/Institute of Education, University of London.
- Sylva, K., Siraj-Blatchford, I., and Taggart, B. (2003). *ECERS-E The Early Childhood Environmental Rating Scale Extension: 4 Curricular Subscales*. Stoke-on-Trent: Trentham Books.

Futher Reading

Pascal, C. and Bertram, A. (1995). *Evaluating and Developing Quality in Early Childhood Settings: A Professional Development Programme*. Worcester: Amber Publishing.

Pramling, I., Sheridan, S., and Williams, P. (2004). *Key Issues in Curriculum Development for Young Children in OECD – Starting Strong Curricula and Pedagogies in Early Childhood Education and*

Care: Five Curriculum Outlines, ch. 2. Directorate for Education, OECD. <http://www.oecd.org/dataoecd/23/36/31672150.pdf> (accessed September 2009).

Sylva, K., Siraj-Blatchford, I., and Taggart, B. (2006). *Assessing Quality in the Early y, Early Childhood Environment Rating Scale Extension (ECERS-E): Four Curricular subscales*, 2nd edn. London: Trentham Books.

EARLY CHILDHOOD EDUCATION AND CARE PROVISION

Contents

Early Childhood Education and Care in the People's Republic of China
Home-Based and Institutional Early-Childhood Education and Care Services
Integrated Early Childhood Education and Care Services – Care, Upbringing, Education and Health
Issues of Access and Program Quality
Kindergarten Transitions and Linkages to Primary School – Readiness Re-conceptualized
Parent Support in Early Childhood – Approaches and Outcomes

Early Childhood Education and Care in the People's Republic of China

Zeng Xiaodong, Beijing Normal University, Beijing, Peoples Republic of China

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Collective economy – A form of enterprise under the planned economy in China where the investments were collectively made and owned by a group of people, rather than by the government. Such entities are generally organized among households in the vast rural areas for farming, and among those in urban community with low skills to support themselves.

Large state syndicate – A group of companies combined to undertake a joint project. Here it refers to the dependence of state-run enterprises on state in the planned economy in China.

Self-kept profit – The portion of the business profit allowed to be kept by the state-owned enterprises to cover operating and nonoperating costs. It is based on the notion that all state-owned enterprises (to which no taxation applied) are supposed to hand over their business profits to the government under the planned economy in China.

Working institution – An institution under the planned economy in China where you worked and were entitled to a wide range of welfares, including housing, ECEC for children, pension, etc., in addition to the salary, which generally is quite limited. It might refer to a government agency, a public institution (academy, school, hospital, etc.), or an enterprise.

transition, as its concepts and institutional framework are being affected by stakeholders, market forces, governments, and the reflection on the overseas models in ECEC. China's ECEC has also to conciliate among many objectives, such as social compensation, equity, facilitating women's employment, and rural areas development. Given this, it is easily subject to changes in national social development policies or innovations of reform by the local governments. This article offers a description of the patterns of ECEC in China from three perspectives, that is, the origin of ECEC in China, its evolution and the propelling forces behind, and the driving and restrictive factors in formulating China's ECEC provision in future.

The Recent Origin of ECEC in China

The idea of ECEC had been introduced to China as early as 1920s, when the progressivist education campaign first came to settle in China, and it had even entailed some practices in few metropolitan cities.

However, China's ECEC service was officially launched only after the founding of the People's Republic of China in 1949, partly out of government's effort to help to liberalize women, that is, the focus of the ECEC then was more on enabling women to work, rather than on early childhood development alone. It was clearly stipulated in Article Two of The Provisional Regulations for Kindergartens, which was issued by the Ministry of Education in 1952, that the mandates of the kindergartens are to ensure the preschool children to have a sound physical and mindful development by raising them according to

Early Childhood Education and Care (ECEC) in the People's Republic of China is presently experiencing a

the New Democratic Guidelines on Education; and in the mean time, to enable their mothers to participate in the political life, production, and cultural and educational activities by easing their burden in childbearing. The Circular on Several Issues on Nurseries and Kindergartens, which was jointly issued in 1956 by the Ministry of Education, Ministry of Health, and the Ministry of Interior Affairs, emphasized the guidelines for the development of nurseries and kindergartens and pointed out that in order to help more mothers solve the problem of looking after and educating their children, the number of nursery and kindergarten needs to be increased.

The strategy to achieve the above objectives through ECEC provision under the planned economy was incorporated into the fundamental social organizational and management system, which was basically categorized into working institutions (the institutions where you work), the urban communities, and rural villages; that is, almost all aspects of life, for example, political and economic activities, as well as social welfares (including ECEC, housing, medical insurance, and pension) were organized or provided by one of these three categories of organizations. This social welfare system was the basis where the ECEC reform could set its orientation and tasks.

Financial and staffing support were arranged for such a system with working institutions and urban communities and rural villages as the ECEC service providers. As ECEC is basically regarded as a kind of work assurance and social welfare for the staff, its cost is part of the operating expense of that institution, with the ECEC staff also being the formal employees in that institution. At the time when the Directive No. 41 by the Ministry of Finance was issued on 15 May 1973, the ECEC cost was listed as a nonbusiness expense, affirming the notion that kindergartens come under the logistic expense and the public welfare nature of ECEC therefore enhanced. Until the 1970s China had improved and perfected a preschool education system which conforms to the planned economy mechanism. The ECEC run by urban communities and rural villages is financed and staffed at their own expenses.

In brief, until economic reform was later launched, ECEC had been clearly defined as part of the welfare offered by an institution for enhancing women's employment. In fiscal terms, ECEC expenses, being incorporated into the welfare and cost of fringe benefits, were deducted from the profit which is supposed to be handed over by an enterprise to the government. Such dispersed funding for ECEC was based on the notion of accounting the whole society as a large state syndicate, where enterprises and businesses also take on some government functions. Therefore, it made no difference whether ECEC was listed as an expense by enterprises or by the government. The implementation of the concept of the socialist market economy releases the individual institutions in

the state syndicate as independent accounting entities, by requesting enterprises to test their products or services in the market. This change shook the foundation of the accounting system of listing the expense of the welfare (including ECEC) offered by enterprises as nonbusiness expenses, and China's ECEC has thus been engaged into the process of economic reform.

Evolution of ECEC in China and the Propelling Drives Behind

The Reform and Opening-up policy that has been adopted since 1978 in China has profoundly affected the whole society and China's ECEC provision has also experienced significant changes, which can be categorized into three stages.

From 1978 to 1993: Consolidation of ECEC Provision

Since 1978, the society was going through an era of overall restoration and restructuring of orders in the aftermath of the Cultural Revolution, so was the ECEC as part of education undertakings. The Summary of National Working Meeting on Nurseries and Kindergartens (1979) reaffirmed the allocation of functions between the Ministry of Education and Ministry of Health. The Ministry of Education has been designated to be accountable for ECEC for children over 3 years, reflecting a focus on the education aspect in early childhood development. The Summary of National Working Meeting on Nurseries and Kindergartens stressed the principle of cooperation under the coordinated leadership, and proposed for the first time to coordinate with the civil construction and housing authorities so that nurseries and kindergartens could accommodate the demographics of the localities. It also requested the ECEC provided by enterprises to be open to the public. All this implies that the concept of open services of ECEC was starting to take root, alongside the strengthening of service provision. The Regulation for the Work of Urban Kindergartens (1979), and a number of comments on the Development of Rural Pre-school Education by the Ministry of Education (1983) clearly defined the standards in curricula, facilities, staffing, etc. in the urban and rural areas. Through all these professional efforts, the education aspect of ECEC was enhanced while its role in facilitating women's employment lessened. Till then, China's ECEC was still based on the dispersed funding and staffing by businesses and public institutions but with a unified professional guidance. Nevertheless, its adjustment in orientation provides the base for further reform of ECEC in China.

From 1993 to 2002: Market Economy Weakened the Funding Basis of China's ECEC

Under the planned economy, till 1993 when the government initiated the reform toward modern enterprise system, the public fund was channeled into ECEC in three ways. First, for ECEC services run by the government and army agencies, and by the public institutions, their operating cost was incorporated into the overall budget scheme in line with the total staffing quotas and service standards pre-approved by the government. Second, for those ECEC services run by the businesses, the funding of their operation was supported by part of its business profit that is allowed to be kept by that business for its own expense, while the majority of the business profit was taken by the government as public revenue. Such self-kept profit practice was featured in the planned economy as an incentive for a business to sustain its welfare services, including ECEC. Third, for ECEC services run by the urban communities and rural villages, their operation relied on the funding by the local communities drawn from the community-owned businesses. Their ECEC staffs are cost and paid at the level equivalent to the similar jobs.

In 1993 when the government started to build up the market economy system in China, all businesses were compelled to launch their transition toward the modern enterprise system by realigning their relations with the governments, easing off some social welfare functions so as to reveal real operation expense. Given this, the route to channel public funds to enterprise staff for welfare benefit was blocked and the funding base for business-run ECEC removed. Many enterprise-run kindergartens were compelled to close their businesses or operate at their own cost. It is estimated that a total of 10 800, or around 80%, of enterprise-run kindergartens were closed during 1995–2005, and most remaining enterprises-run kindergartens were transformed into kindergartens with other, sometimes uncertain, funding nature. That's why currently some ECEC institutions in China are referred to as kindergartens in transition, by which it implies that they are yet to come into a final classification. Probably these

kindergartens ought to be transformed into not-for-profit organizations; but respective laws are necessary to regulate them in governance, funding, operation, and quality assurance. Under the planned economy, the urban communities and rural villages ran their collective-owned enterprises and pay for their own welfare. These collective enterprises, most of which were small, welfare-based businesses, had also to face the fierce competition of the market economy, and many drowned in the ocean of the market economy, owing to their lack of competitiveness and/or ineffective governance. The land reform in the rural areas is another reason for the fading of the collective economy. Hence, the funding basis for ECEC services run by the urban communities and rural villages collapsed along with the collective economy. Only the ECEC services run by the governments, army agencies, and public institutions are still receiving steady public funding through government budgets. Not only did they survive the funding crisis, but also gradually evolved as the high-quality ones, by availing of the resources to which they are entitled.

In summary, the transition to the market economy has in some ways removed two out of the three funding sources for China's ECEC services. This is how we have come to the present pattern where the public funding is only made available to the ECEC operated by the governments and army units, public institutions, which is under frequent criticism.

From 2002 to the Present: Private Providers are Emerging as a Major Player in China's ECEC Services

While the market economy weakened the public ECEC provision, it has created enormous space for the market players, and, on the whole, the supply side of service capacity has increased dramatically. Since 2000, the private ECEC service has increased 52.6%. Currently, private ECEC providers make up around half of the total ECEC institutions, and children enrolled in private ECEC account for one-third

Table 1 China's ECEC provision under the planned economy and at present

<i>ECEC providers</i>	<i>Funding source in the planned economy</i>	<i>Funding sources at present</i>
<i>Run by working institutions</i>		
by Government and army agencies	Government budget	Government budget
by public institutions (academies, hospitals, universities, etc.)	Government budget	Government budget
by businesses/enterprises	Self-kept profit	Removed
<i>Run by collective economy</i>		
by urban communities	Profit by its business	Removed
by rural villages	Profit by its business	Removed
<i>Private</i>	Nonexistent	Private

of the total, emerging as an important supplier of ECEC in China.

Owing to the lack of basic rules to regulate the market supply, there are many problems associated with private ECEC services. First, the accounting norms currently in use in China do not make a clear distinction between for-profit and not-for-profit ECEC providers, and, as a result, much capital enters ECEC services trying to obtain the dual benefits by making profits while enjoying tax exemption. Second, private ECEC is spread on either the high or the low end of the market. The former is facing blames for its unwarranted high prices, while the latter frequently arouses safety concerns and poor quality, resulting from the low cost.

Uncertain Reform of ECEC Provision Ahead in China

Along with the launch of the second-stage piloting work by the Central-Government-owned enterprises to disengage from social functions in 2002, the reform of ECEC provision has also set out on an uncertain voyage. Here, it is referred to as uncertain in four senses. First, the goal of ECEC reform is still unclear. There has been no consensus as to whether it shall aim to strengthen the government's responsibility in ECEC provision, or to better align the current funding structure to improve the ECEC provision. Second, the balance between the role of government and market has not been resolved, owing to obscure expression of this issue in the Law for Non-Government Education. Third, given the multidiscipline nature of ECEC, no certain government agency has so far been designated the role as the focal point or as the leading authority for management and regulation of ECEC in China, thus causing difficulties in policymaking and coordination. Fourth, the early education part is having an upper hand over the child-development part in both urban and rural areas as people are anticipating an ever-fierce competing and globalized society, and the need in facilitating women's employment is often being overshadowed by various professionalized early education efforts, which are mainly concerned with how to impart knowledge and skills to early children.

One of the significant features in the transition period is the transformation and alternation of new and old systems and mechanisms. In the transition of ECEC provision from a working-institution-based model to a socialized and diversified model, the benefits of some social groups are harmed; in particular, the collapse of enterprise-run ECEC has negatively affected the welfare of the enterprises' staff, whereas the benefits of those who can access kindergartens operated by government bodies and public institutions have remained unchanged. The ECEC for urban communities and rural villages have succumbed to the market, given their collapsed funding base.

Economic reform has propelled ECEC into a reform pathway which has raised many urgent issues for attention. It is a justifiable objective to separate the social services of enterprises to allow them to perform their economic functions. However, the removal of social services from enterprises raises the question of how to redefine the nature of existing ECEC providers, and how to secure the working conditions of ECEC teachers. At once, private investment in ECEC raises the issue of how to regulate the nongovernment kindergartens. Another case: in some newly developed communities, ECEC services, unlike other public services such as primary schools, post offices, hospitals, and groceries, are yet to be provided. This is largely caused by the lack of a leading authority for ECEC development in the local government, as mentioned previously in this article, and thus no public funding available being allocated. All these problems in ECEC provision in China reflect the challenges for the government in adapting to the changed context in their delivery of public services to the society.

Some of these problems require urgent response by the education administration. For example, some ECEC teachers appealing for assistance, social resentment toward the unwarrantedly high price for some private ECEC services, escalating safety concerns over some other private ECEC services owing to the overcompetition among themselves through lowering the cost and quality, and the difficulties of enrolment to ECEC in some newly developed communities owing to the mismatched facilities. All these conflicting issues seem to leave no enough time for reform to take place in an easy manner.

Conclusion

At present and in the years to come, China's ECEC has to evolve in the mixed context of the planned and market economic circumstances. On the one hand, the welfare offered by the working institution under the planned economy will be further eroded, and on the other, the government has to become accustomed to managing the emerging players in ECEC under the market economy. It is likely that China can neither deter the transition of the publicly funded ECEC service, nor effectively manage the emerging private or transformed ECEC providers. In view of such a mixed and diversified provision in the context of lacking basic regulation, ECEC in China faces the hardship of long-term evolution: to clarify the objectives and make realistic choices. The new system should be built in a carefully discussed conceptual framework and on a clear and rational base of public service provision. Regulation and funding issues are more important than technical issues for ECEC in China at this time.

Bibliography

- Chinese Early Childhood Education Association (1999). *Collection of ECEC Policy in Republic China*. Beijing: Beijing Normal University Press.
- Dugger, W. M. and Sherman, H. J. (2000). *Reclaiming Evolution*. London: Routledge.
- Jinglian, W. (2004). *Economic Reform in Current China*. Shanghai: Shanghai Far East Publishing.
- Kasper, W. and Streit, M. E. (1999). *Institutional Economics: Social Order and Public Policy*. Northampton, MS: Edward Elgar.
- Mantzavinos, C. (2001). *Individuals, Institutions, and Markets*. Cambridge, UK: Free Press.
- Ministry of Finance (1973). *Directive No. 41 on Some Costs Accounting in State Enterprises*. <http://www.fayixing.com/lawcontent.jsp?id=119262> (accessed September 2009).
- OECD (2002). *Models of Public Budgeting and Accounting Reform*. *OECD Journal on Budgeting* 2(Supplement 1), 333–354.
- Tony Bertram Christine Pascal (1999). *The OECD Thematic Review of Early Education and Care: The National Report for the UK*. <http://www.oecd.org/dataoecd/48/16/2479205.pdf> (accessed September 2009).
- Xiaodong, Z. (2006). The evolution of ECEC provision in China: From Working Institution Welfare to social welfare. *Journal of BNU* 2006(2), 11–16.

Home-Based and Institutional Early-Childhood Education and Care Services

R Fukkink and H Blok, University of Amsterdam, Amsterdam, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Glossary

ECEC – Early-childhood education and care.

ES – The abbreviation for effect size, calculated as the difference between the experimental group and the control group, expressed in standard deviations. In the social sciences, values of approximately 0.20 are considered to indicate small effects. Effects of 0.50 are considered medium, and effects of 0.80 or higher are considered large.

Meta-analysis – The statistical method of summarizing the results from different studies by aggregating outcomes over individual studies.

Introduction

Children from disadvantaged families are generally less successful in school. Early-childhood education and care (ECEC) intends to prepare these children for a more favorable school career. The central thesis of ECEC is that the environment of disadvantaged children can be enriched in order to optimize child development. Specific goals include preparing children adequately for primary education. General goals include offering children opportunities to escape from their detrimental circumstances.

ECEC includes a diverse array of facilities and projects in different parts of the world. In addition to the famous and extensively researched Head Start program in the USA, other historic forerunners in early care, early education, and early intervention have focused on the development of young children in disadvantaged situations. Preschool and day care are but two of the many forms that are covered by ECEC, which also includes the half-day play groups of the Anglo-Saxon world, the French *école maternelle* (nursery school), the German *Krippe* and *Kindergarten*, home instruction for parents of preschool youngsters (HIPPY) projects from various countries, the internationally renowned Italian *Reggio Emilia* centers, the Japanese *yochien* (private preschool) and *boiku-en* (full-day programs), Kenyan community-based rehabilitation centers, and various early-childhood projects that have been initiated by the United Nations, United Nations Educational Scientific and Cultural Organization (UNESCO) and the World Bank in various parts of the world. The heading of ECEC

covers extensively documented and scientifically evaluated, small-scale demonstration projects from the USA, as well as the less-frequently described, large-scale *anganwadis* centers in India, in which 4.6 million children receive health check-ups, in addition to education (see Boocock, 1995; Guralnick, 2005; Odom *et al.*, 2003 for a few international examples).

According to the ecological model (Bronfenbrenner, 1979), biosocial developmental contextualism (Ramey and Ramey, 1998), the developmental-systems approach (Guralnick, 2005), and other theories, early intervention offers multiple pathways to improving the development of children. Some programs influence the parents; others provide direct educational experiences to children, and still others do both. Three corresponding program types have been distinguished in the literature: home-based, center-based, and a combination of home-based and center-based delivery (each of these types is discussed in more detail below). In addition to differences in the location of operations, these types of programs also differ according to the chosen manner of entry into the ecological system to combat disadvantages. Program characteristics vary widely within each of these three broad categories (home-based, center-based, and combined). Without any attempt to provide an exhaustive list, examples of these characteristics include the age of the children at the beginning and end of the program (and therefore the length of the program); the involvement of professionals, paraprofessionals, or volunteers; the application of an individual or a group approach; and program intensity, as measured by the average number of hours per week and weeks per year.

Effectiveness of ECEC

The effects of early-childhood interventions have been studied in a number of areas. At the level of the child, outcomes have been investigated in the cognitive, affective, behavioral, and physical domains. Longitudinal designs, in which children are followed for longer periods, often reveal a shift in the domains that are measured. The initial measurement usually involves scores on academic achievement tests or school-readiness tests, while later measurements focus on school marks and information involving the school career (e.g., grade retention, placement in special education, attendance, and dropout rates), in addition to social career and citizenship across the life span (e.g., income, criminality). Some studies also

include other types of variables which illustrate the diversity of programs and reflect the values that are promoted in different societies (Boocock, 1995; Odom *et al.* 2003). For example, a Swedish evaluation study concludes that early entrance into day care predicts a creative, socially confident, popular, open, and independent adolescent (Boocock, 1995: 102). Boocock also mentions in her international overview variables such as sharing and cooperating (Singapore) and a group-oriented sense of self (Japan). In places where ECEC programs also include health education and nutrition supplements, the studies have measured the effects on physical growth and infant mortality. The majority of research focuses on the effects of such programs on children, although outcomes for parents have also been charted for home-based programs with parent-education components. Interestingly, a growing body of research is emerging that charts the long-term retention of program effects.

Some older publications sketch an optimistic image of ECEC, based on a few results that were achieved in successful projects and on sometimes high expectations (see Evans, 1994). In the recent scientific literature, which has been enhanced by many empirical results, ECEC is often discussed from a more balanced perspective. According to Boocock (1995), recent reviews of studies (predominantly from the USA, but including research from other countries as well) clearly show that intensive center-based and combined programs reduce educational disadvantages, although they do not resolve them. While ECEC does contribute positively to the development of children in disadvantaged circumstances, it does not resolve disadvantages either completely or definitively. Meta-analytic studies report statistically significant but modest cognitive effects, with effect sizes ranging from small to medium ($ES = 0.18$ in Sweet and Applebaum (2004) for home-based programs; Blok *et al.*, 2005: 0.32; and Anderson *et al.*, 2003: 0.35–0.43). Effects are smaller in the socioemotional domain, which tends to be less prominent in program goals and conducted evaluations ($ES = 0.10$ for home-based programs in Sweet and Applebaum (2004); 0.05 in Blok *et al.* (2005); and 0.38 in Anderson *et al.* (2003)). Other important effects (e.g., in the social and medical domains), are yet to be investigated sufficiently to allow any strong conclusions (Anderson *et al.*, 2003).

According to some hypotheses, ECEC may have ripple effects (i.e., a positive effect in a specific domain has a positive influence on the development for a different domain) or even sleeper effects (i.e., effects that become apparent only over time). A rival hypothesis, however, predicts that immediate effects will fade over time, as the continually changing school environment continually calls for new competencies. For example, Farran (2000) emphasizes that most programs focus on concrete skills related to readiness tests, even though profound curricular changes

take place in the third and fourth grades. Similar argumentation is conceivable with regard to home-based programs. Various developmental psychologists have asserted that growing up represents a series of continually differing developmental tasks for children. This implies that parents are confronted with changing parenting tasks throughout the process of raising a child, and these tasks call upon different parenting skills.

Several reviews have indeed provided evidence that the positive results of early intervention tend to fade over time (Blok *et al.*, 2005; Farran, 2000; Gilliam and Zigler, 2000). It is nonetheless important to note that, although the positive effects are likely to fade, they do not disappear entirely. The classic and often-used term fade out is misleading in this respect. Moreover, longitudinal research shows that programs whose positive effects on cognitive test scores fade can have ripple effects in other domains. For example, children whose scores on school-readiness tests improve immediately after ECEC programs are less likely to be held back or referred for special education (Barnett, 1995).

Home-Based Programs

For home-based interventions (also known as home-visitation, home-enrichment programs, parent education, parenting programs, or maternal teaching programs), professionals, paraprofessionals, and in some cases, volunteers without special training visit parents in their homes over a period that lasts between 1 and 3 years, beginning when the children are young (between birth and 3 years of age). In some cases, the program starts while the mother is still expecting. Internationally recognized examples include the HIPPI projects and the parents-as-teachers program. Home-based programs are highly diverse and include an array of interventions that vary widely in goals, methods, and target groups.

Proceeding from the assumption that parents mediate the development of their children, home-based programs rely on indirect methods of stimulating children's development, and they often assume a cascading set of influences. It has been hypothesized that home-based programs lead to a change in parental attitudes, behavior, and school involvement. These changes result in more and higher-quality parent-child transactions and family-orchestrated child experiences, all of which have positive effects on the child. According to the philosophy of home-based programs, therefore, parent education goes hand in hand with child development.

Recent reviews of available evidence of the effects of home-based programs are not overly optimistic (Brooks-Gunn and Markman, 2005; Farran, 2000; Gomby *et al.*, 1999; Halpern, 2000; Sweet and Applebaum, 2004). The conclusion is that, while home-based programs do bring

about changes in both parents and children, these changes are quite modest. A fundamental question therefore involves whether home-based programs are actually capable of generating significant improvements for families in disadvantaged circumstances. For example, in a meta-analysis of 60 programs, Sweet and Applebaum (2004) found relatively modest benefits for parents ($ES = 0.14$ for parenting behavior and 0.11 for parental attitudes) and for children ($ES = 0.18$ for cognitive outcomes, and 0.10 for socioemotional outcomes). The results are more favorable for programs that are more intensive and for those that involve working with professionals. A meta-analysis by Bakermans-Kranenburg *et al.* (2005) also reports modest effects for home-based programs, as measured with the frequently used homeschooling opportunities to maximize effectiveness (HOME) instrument ($ES = 0.20$). This meta-analysis further shows that randomized studies reveal even smaller effects ($ES = 0.13$). On average, the greatest effects are observed in the area of increased parental involvement ($ES = 0.19$), cognitive stimulation through the availability of educational materials (0.20), and the availability of various experiences (0.16). The interventions proved less effective for families from the most disadvantaged circumstances (e.g., families with low socioeconomic status and adolescent mothers).

According to Gomby *et al.* (1999), one explanation for the modest results is the vulnerability of home-based interventions when they are implemented in practice. For example, in some cases, only half of the planned home visits actually take place, and the visits that do occur are often shorter than planned (lasting between 28 and 50 min instead of between 45 and 60 min). In addition, some parents are not capable of completing the planned home assignments with their children, and too many families (between 20% and 67%) drop out before completing the program. The effective dose is relatively modest, involving approximately 20–40 contact hours.

Center-Based Programs

Center-based programs (also known as institutional ECEC services, preschool, early childhood education, childcare, play group, kindergarten, educare, or the child-focused approach) provide children with direct educational experiences in day care, preschool, or similar settings. In many countries, institutional ECEC services are divided into two separate systems: child-development centers (preschool) and childcare centers (care). Stimulation and preparation for primary school are important objectives for child-development centers, which have, historically, connections to school systems. Childcare facilities were originally established with the goal of facilitating the labor-force participation of parents with young children. A frequent point of criticism is that childcare centers pay insufficient attention

to stimulation (whether cognitive or in other domains). For example, according to an Organization for Economic Cooperation and Development (OECD) report titled *Starting Strong-II* (OECD, 2006: 16), “Children’s well-being and learning are core goals of early childhood services, but services for children under 3 have often been seen as an adjunct to labour market policies, with infants and toddlers assigned to services with weak developmental agendas.” On the other hand, some critics argue that child-development centers place excessive emphasis on this type of stimulation. The same OECD report identifies “the use of programmes and approaches that are poorly suited to the psychology and natural learning strategies of young children” (p. 13) as one disadvantage of the readiness-for-school approach that is common in English and French-speaking countries.

Center-based programs are based on the idea that providing direct stimulation to children is the most effective strategy. It has been proposed that center-based programs should begin relatively early, as children from disadvantaged families already show educational disadvantages upon entering primary school, and these disadvantages do not tend to decrease throughout the educational process. Many center-based programs begin in the toddler years and continue through the early grades. Each program has its own curriculum for addressing cognitive aspects such as early literacy and numeracy, and many address socioemotional development as well. Special center-based programs are usually carried out with relatively high structural quality. Caregivers in these programs have relatively high levels of education and special training; caregiver-to-child ratios are favorable, and the groups are usually small. As a result, these programs compare positively to other national center-based services, which are faced with less-advantageous conditions. The role of parents is less crucial in center-based programs than in home-based programs, although parents are often involved in the operation of center-based programs. For example, in child–parent center (CPC) projects and similar projects, parents can choose from a wide range of activities, including serving as classroom aides, tutoring children, performing clerical tasks, or accompanying classes on field trips. Other well-known examples of center-based programs include Success for All and Head Start.

Findings in this field have shown that significant effects can be achieved in the area of developmental competence and achievement-test scores at the end of preschool and into kindergarten. Increasing evidence also suggests that direct provision through center-based approaches is more effective than is indirect provision through home-based approaches. For example, in a recent meta-analysis using a statistical model with multiple predictors, Blok *et al.* (2005) found that cognitive effects were one half of a standard

deviation greater in center-based programs than in home-based programs ($ES = 0.48$).

Combinations of Center-Based and Home-Based Programs

There is a growing awareness in the field that effective programs should target both children and their parents in order to promote stable effects on child development. In line with this more inclusive view, a combination of home-based and center-based components is offered in combined programs (also known as mixed-models, two-generation programs, multistrategy programs, or a middle approach between home-based and center-based programs). The central assumption of combined programs is that the combination of the two delivery formats may create more pathways for positive change in families' lives and that they may have synergetic effects (Ramey and Ramey, 1998). Well-known examples include the High/Scope program and the Perry preschool project. Both programs comprise half-day classroom sessions and educational home visits, in which teachers attempt to engage mothers in learning activities with their children, paralleling the approach used in the classroom.

The parental component focuses on the parent as the primary caregiver, the parent as individual, or a combination of the two. In the first variant, the home-based component centers on the child, as the parental component is explicitly oriented toward child rearing, and parenting skills are central to the approach. In the second variant, the parental component focuses on the parents as individuals and therefore includes social, economic, and other forms of support (e.g., parent self-esteem, employment training, English-as-a-second-language courses or other forms of adult education). Combined programs with a dual focus on parents as primary caregivers and as individuals constitute the most extensive type of program, including both a center-based component and parental component that focuses on both parenting skills and the provision of social and/or economic support. The intensity of the parental component is a particularly prominent source of variation between programs. Some projects involve mandatory adult education, while the parental component of other programs consists only of a referral service. In other cases, a program involves multiple ways of carrying out the parental component in practice, due to local variations or customization by case managers who coordinate the program supply with the needs of individual families (see St. Pierre *et al.*, 1995 for an overview).

Both narrative reviews (Barnett, 1995; Farran, 2000) and quantitative reviews (see, e.g., Blok *et al.*, 2005; Karoly *et al.*, 2005) have found convincing evidence that combined programs produce favorable educational results at the child level. In a meta-analysis, Blok *et al.* (2005) report a statistically significant advantage of combined programs

compared to center-based and home-based programs ($ES = 0.42$ with regard to center-based programs and 0.90 with regard to home-based programs). In a meta-analysis by Karoly *et al.* (2005), however, the difference is not significant, although the trend is in the same direction ($ES = 0.33$ for combined programs, as compared to 0.21 for parent-training-only programs). There is evidence to suggest that the cognitive benefits for children are also greater when the parental component focuses on the parent as primary caregiver ($ES = 0.69$), which lends itself well to theoretical explanation. Parental components that focus exclusively on social or economic support yield no advantages in the area of cognitive child outcomes. According to Yoshikawa (1995), combined programs are also the most effective with regard to realizing advantages in the social realm and preventing delinquency at later ages.

At the parent level, positive effects have been reported in the area of parenting skills (i.e., for the child-focused variant), corresponding to the results of the previously mentioned home-based programs. Effects of adult education in combined programs are apparently more difficult to achieve. In a discussion of six US programs, St. Pierre *et al.* (1995) report that results are demonstrated only for outcomes that were closely related to the program (e.g., completion of certificates or more frequent use of food stamps or similar welfare benefits after having taken classes that refer to them). No transfer effects emerged in the form of improved performance on standardized tests, employment, household income, depression, or parental self-esteem. The conclusion that emerges is that impressive results have been achieved with combined programs by coordinated efforts in the home and the ECEC center. Nonetheless, the situation of the parents apparently did not change drastically, at least not in the short term.

Conclusion

The field of ECEC has developed considerably since the advent of the pioneering US Head Start program, center-based programs, and various home-based programs in the 1960s. Although program developers can build upon the experiences of the past, current times require new approaches in order to make a positive difference for children who grow up in disadvantaged circumstances.

Remarkably, the programs are ambitious and thus broader in scope. One important matter that is directly related to this involves coordination and continuity. Combined programs, which have proved to be effective, involve goals at both the child and parent levels, and they have various program components that must be coordinated with each other. There is also a demand for integrated and continuing program lines in which services are seamlessly coordinated with each other (see, e.g., Guralnick, 2005). From a theoretical perspective, enhancing continuity

regarding transitions between the most important social contexts of the young child (i.e., home, preschool, and elementary school) seems very important for the optimization of developmental and learning gains. The issue of coordination also relates to a completely different area – the coordination of vocational education with the occupational sector. Lessons that have been learned through practice should be translated into vocational education and career training, with special attention to ECEC in order to impart the appropriate vocational skills to future personnel (Klein and Gilkerson, 2000). Additional attention to vocational education is also important as research shows that targeted training is effective in improving essential interaction skills among professional caregivers who work with children in the day-care sector (Fukkink and Lont, 2007).

Limitations of Studies

ECEC has, generally speaking, positive effects. Another positive conclusion is that longitudinal studies continue to find effects years after completion of a program; ECEC therefore continues to work long after the end of the program. How large an effect must be before it constitutes a meaningful result remains an open question. Guidelines from the research literature are not the only important considerations in this regard; other considerations include the importance of the variables themselves, the necessary investments in terms of time and money, and the availability of alternatives. As emphasized by various authors, the results that have been found in ECEC evaluations are encouraging, while also calling for realistic expectations on the part of policymakers and directly involved parties.

Despite a growing body of literature, the modest knowledge base remains an important limitation in discussions of the effectiveness in ECEC. First, the internal validity of studies has traditionally been a topic of discussion in this regard. In addition to randomized studies, the literature includes results from quasi-experimental designs, which suffer from methodological weaknesses. Other questions can be raised regarding external validity. One issue involves the extent to which findings from predominantly US studies can be generalized to other countries that differ from the USA in many respects. Projects conducted in other parts of the world should be systematically evaluated and should receive more attention in the research literature, which is currently dominated by Western (and largely USA) studies. Another issue involves the generalizability of results from the past to the present and near future. The demographic composition of populations is changing continually, ECEC programs are in transition, and there are shifts in the participation of various groups in home-based, center-based, and combined programs.

A related question involves the determination of which approach is more effective – which approach works the

best? The steadily increasing database of ECEC evaluations makes it possible to use meta-analysis to identify characteristics that are associated with greater or lesser effects. This is also interesting because various program designs have been evaluated with differing results. These lines of research show that combined programs are more effective than center-based programs from the perspective of education, and that center-based programs in turn are more effective than home-based programs (Blok *et al.*, 2005). Correlations have also been found with improved training for personnel, more favorable child-to-staff ratios, and greater program intensity (Karoly *et al.*, 2005). For home-based programs, possible success factors have been shown to vary with regard to parent or child outcomes. When child outcomes are the focus, more intensive programs, working with professionals and single-site programs are more effective, although the associations that have been found are not strong (Sweet and Applebaum, 2004).

The meta-analytic line of research also involves a number of limitations. First, although the primary data from intervention studies are experimental or quasi-experimental (and thus aimed at identifying causal relationships), the study of these programs through meta-analysis is correlational. The classic adage “correlation is not causation” is applicable in this regard. In addition, characteristics of programs are usually confounded, thereby making it impossible to determine the unique contribution of a given program characteristic. For example, as could be expected, combined programs offer more services than do other types of programs, and they tend to be characterized by higher intensity and thus by a higher dose of the intervention than home-based programs are. The unique contribution of delivery model or intensity to effectiveness is therefore difficult to interpret unambiguously. An additional complicating factor is that there is sometimes too little variation in background characteristics. For example, many home-based programs pursue similar goals for similar populations, making it impossible to make a true comparison on these points (Karoly *et al.*, 2005). An additional limitation is that associations between program characteristics and effects are usually investigated with regression models under the assumption of linearity. Whether this assumption is always valid remains an open question. For example, it is questionable whether the effects decrease constantly over time or whether each home visit in a home-based program always adds a constant effect, as assumed by a linear model. It could be that the relationships are not linear or that it is necessary to consider threshold values below which effects cannot be seen.

Briefly summarized, the empirical research that has been conducted provides answers (at least tentatively) to three important questions in the field: Does ECEC work? How long does it work? What works the best? Nonetheless, authors unanimously emphasize that the empirical basis for home-based, center-based, and combined programs

continue to have limitations, despite the impressive efforts of various involved parties.

See also: Cost-effective Early Childhood Programs from Preschool to Third Grade; Evaluating Early Childhood Education and Care Programs; Investing in Early Childhood Education and Care: The Economic Case; Parent Support in Early Childhood – Approaches and Outcomes.

Bibliography

- Anderson, L. M., Shinn, C., Fullilove, M. T., et al. (2003). The effects of early childhood development programs: A systematic review. *American Journal of Preventive Medicine* **24**, 32–46.
- Bakermans-Kranenburg, M. J., Van IJzendoorn, M. H., and Bradley, R. H. (2005). Those who have, receive: The Matthew effect in early childhood intervention in the home environment. *Review of Educational Research* **75**, 1–26.
- Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *Future of Children* **5**, 25–50.
- Blok, H., Fukkink, R. G., Gebhardt, E. C., and Leseman, P. P. M. (2005). The relevance of delivery mode and other program characteristics for the effectiveness of early childhood intervention with disadvantaged children. *International Journal of Behavioral Development* **29**, 35–47.
- Boocock, S. S. (1995). Early childhood programs in other nations: Goals and outcomes. *Future of Children* **5**(3), 94–114.
- Bronfenbrenner, U. (1979). *The Ecology of Human Development*. Cambridge, MA: Harvard University Press.
- Brooks-Gunn, J. and Markman, L. B. (2005). The contribution of parenting to ethnic and racial gaps in school readiness. *Future of Children* **15**, 139–168.
- Evans, J. L. (1994). Early childhood interventions. In Husén, T. and Postlethwaite, T. N. (eds.) *International Encyclopedia of Education*, 2nd edn. Oxford, UK: Pergamon.
- Farran, D. C. (2000). Another decade of intervention for children who are low income or disabled: What do we know now? In Shonkoff, J. P. and Meisels, S. J. (eds.) *Handbook of Early Childhood Intervention*, 2nd edn., pp 510–548. Cambridge, UK: Cambridge University Press.
- Fukkink, R. G. and Lont, A. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly* **22**, 294–311.
- Gilliam, W. S. and Zigler, E. F. (2000). A critical meta-analysis of all evaluations of state-funded preschool from 1977 to 1998: Implications for policy, service delivery and program evaluation. *Early Childhood Research Quarterly* **15**, 441–473.
- Gomby, D. S., Culross, P. L., and Behrman, R. E. (1999). Home visiting: Recent program evaluations – analysis and recommendations. *Future of Children* **9**, 4–26.
- Guralnick, M. J. (2005). *The Developmental Systems Approach to Early Intervention*. Baltimore, MA: Paul H. Brookes.
- Halpern, R. (2000). Early intervention for low-income children and families. In Shonkoff, J. P. and Meisels, S. J. (eds.) *Handbook of Early Childhood Intervention*, 2nd edn., pp 361–386. Cambridge, UK: Cambridge University Press.
- Karoly, L. A., Kilburn, M. R., and Cannon, J. S. (2005). *Early Childhood Interventions: Proven Results, Future Promises*. Santa Monica, CA: RAND.
- Klein, N. K. and Gilkerson, L. (2000). Personnel preparation for early childhood intervention programs. In Shonkoff, J. P. and Meisels, S. J. (eds.) *Handbook of Early Childhood Intervention*, 2nd edn., pp 454–483. Cambridge, UK: Cambridge University Press.
- Odom, S. L., Hanson, M. J., Blackman, J. A., and Kaul, S. (2003). *Early Intervention Practices around the World*. Baltimore, MA: Paul H. Brookes.
- OECD (2006). *Starting Strong II: Early Childhood Education and Care*. Paris: OECD.
- Ramey, C. T. and Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist* **53**, 109–120.
- St. Pierre, R. G., Layzer, J. I., and Barnes, H. V. (1995). Two-generation programs: Design, cost, and short-term effectiveness. *Future of Children* **5**, 76–93.
- Sweet, M. A. and Applebaum, M. I. (2004). Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. *Child Development* **75**, 1435–1456.
- Yoshikawa, H. (1995). Long-term effects of early childhood programs on social outcomes and delinquency. *Future of Children* **5**, 51–75.

Further Reading

- New, R. S. and Cochran, M. (2007). *Early Childhood Education: An International Encyclopedia*. Westport, CT: Praeger.
- OECD (2001). *Starting Strong. Early Childhood Education and Care*. Paris: OECD.
- Shonkoff, J. P. and Meisels, S. J. (2000). *Early Childhood Intervention*, 2nd edn. Cambridge, UK: Cambridge University Press.
- Shonkoff, J. P. and Phillips, D. A. (2000). *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, DC: National Academy Press.
- Zaslow, M. and Martinez-Beck, I. (2006). *Critical Issues in Early Childhood Professional Development*. Baltimore, MA: Paul H. Brookes.

Relevant Websites

- <http://www.acf.hhs.gov> – Administration for Children and Families: U.S. Department of Health and Human Service.
- <http://portal.unesco.org> – UNESCO, homepage.
- <http://www.unicef.org> – Unicef, homepage.
- <http://web.worldbank.org> – Worldbank.

Integrated Early Childhood Education and Care Services – Care, Upbringing, Education and Health

S Colley

© 2010 Elsevier Ltd. All rights reserved.

In recent years, international researchers have recognized that integrating early childhood education programs will support the development of young children in a better manner. Integrated early learning and child-care programs include a range of services and systems that are attempting to blend governance and practice in a number of fields. Since the publication of the Organization for Economic Cooperation and Development (OECD) Report *Starting Strong* (OECD, 2001), many countries have focused on blending their education and child-care systems and services. Some countries have also attempted to integrate other services – health and social services – but this article focuses on integration between education and care.

The *Starting Strong* OECD Reports (OECD, 2001, 2006) identified eight policy directions to encourage equitable access to quality early childhood education and care (ECEC). The recommendations focused on supporting the social needs of families and strengthening the foundations of lifelong learning for all children. The second policy direction was related to the need for greater integration – a strong and equal partnership with the education system. Specifically, the OECD recommended that governments review their professional profiles, improve recruitment levels, and strengthen the initial and in-service training of staff in both kindergarten and child care.

Other studies have found positive results for child development as a result of their integration initiatives. One international study of early-years-integration projects concluded that integrated care and education programs were beneficial for children from birth to 6 years, particularly for children facing multiple risks (Penn *et al.*, 2004). The Ontario, Early Years Study cited integration as an important stepping-stone in the quest for improved child-development outcomes. One of the influential projects that impacted on it was the South-East Grey Community Organization (SEGCO) established in rural Canada as a delivery model that could offer accessible, sustainable child care, junior kindergarten, and family-support services in a rural community. The agency created a hub model to deliver programs and services tailored to the community's requirements. The UK Sylva *et al.* study of preschool education reported that children who attended integrated early learning and care programs and those who attended nursery schools made better intellectual progress by the time they entered primary school than children

who attended regular child-care centers, family child-care programs, or other combinations of nonparental care and early-education programs.

The Integration Continuum

There is a recurring question of what is meant by integration. Some programs consider themselves integrated if they offer separate services in one location (or close by) and have regular communication. Some services consider themselves to be integrated if they co-locate services to provide a one-stop shopping opportunity for local residents with some loose coordination and regular communication. Others consider integration to be a coherent service equally accessible to all potential users with a common budget, staffing, policy environment, and a pedagogic framework for all programs. Or, it may mean combining care and health (or welfare), but not care and education programs. To achieve positive outcomes, it is generally acknowledged that something more than coordination is required, but that attention needs to be given to both structural and conceptual integration. Researchers have generally concluded that service integration involves an ongoing process that involves a continuum of awareness, communication, coordination, collaboration, and consolidation even though it is not necessary to go through each stage. The focus is on the extent to which joined-up working and collaboration is occurring. (Cohen *et al.*, 2003)

A general distinction can be drawn between those countries that have administratively amalgamated their education, child care, and family-support systems under one ministry and those countries that have maintained the historical separation.

Countries with Administrative Amalgamation of ECEC Services

As early as the 1960s, several European countries began to integrate their child care and preschool education programs under one government department. Most of the Nordic countries integrated them into their welfare systems, while several Mediterranean countries integrated them into education. A new wave of integration began to occur in the 1990s when several OECD countries then

moved their entire ECEC systems for children from birth to age 6 (or older) under the umbrella of education. This process took place in some countries that had already experienced a measure of integration and in others that had previously maintained a total separation between child care and public education.

Of the countries that have administratively amalgamated their ECEC programs under education, perhaps the most well-known example is that of Sweden. In 1996, responsibility for child care in Sweden was transferred from the ministry of health and social affairs to the ministry of education and science under the School Act. Among a number of reforms, the government also introduced a uniform framework for training for child-care teachers, school teachers, and school-age child-care providers (known as leisure-time pedagogues). In this process, there were concerns expressed about preschool becoming rigid and formal, losing its emphasis on play and children's holistic development. Child-care teachers were anxious about the threat to their profession as a result of the shift toward education. Yet, at the time of the change, administrators and politicians were cognizant of the need for early child development philosophy to influence the integration process. Prior to moving preschool under the ministry of education, a national study urged Swedish schools to become more responsive to children's individual learning needs and styles. It was argued that integrating preschools with primary schools would allow the former to transform the latter. This point was taken into consideration in the revision of the school curriculum, which took on many of the pedagogical practices of preschools. Learning came to replace teaching, shifting the overall focus from teaching to learners. The Swedish Prime Minister, Göran Persson, set this approach in motion when he announced his vision of lifelong learning for Sweden in 1996, stating that preschool education should be a part of the country's lifelong learning vision. Nevertheless, concern persists about the degree to which early childhood programs have become too schoolified, with too much emphasis on cognitive skills and not enough emphasis on the other domains of development: physical, cultural, social, and emotional development (Choi, 2002).

Sweden was not alone in integrating the administration of its ECEC programs. New Zealand, Spain, Brazil, Slovenia, Iceland, Norway, England, and Scotland, as well as some of the Canadian provinces and the Australia states, also moved to unify services under education departments. In England, the central government not only administratively integrated its nursery school and child-care programs, but also endeavored to integrate health and family-support programs to better support children, families, and communities. The government established Sure Start to work with parents and preschool children to promote the physical, intellectual, social, and emotional development of children, particularly those who were disadvantaged. The aim was to

ensure that the children were ready to thrive when they arrived at school. In each area where there is a Sure Start project, locally based programs are encouraged to build on what already exists to ensure a range of core services including:

- outreach services and home visiting;
- support for families and parents;
- good quality play, learning, and child care;
- primary and community health care, and advice about family health and child development; and
- support for those with special needs.

Local communities may provide extra services according to local needs, such as skills training for parents, personal development courses, and practical advice and support such as debt counseling, and language or literacy training.

In New Zealand, some community organizations have developed their own initiatives toward multiservice provision. This is a result of a new policy encouraging ECE centers to offer parental support and development. In 2006, the Ministry of Social Development funded eight pilot projects of ECE Centre Based Parent Support and Development. Funding was extended to a further ten pilot centers in 2007 (Mitchell, 2008).

Countries that Have Maintained the Historical Separation Between Public Education and Child Care

Countries, such as the United States, Canada, and Australia, are beginning to explore ways to integrate ECEC services, often through the establishment of demonstration projects that may further influence government policy.

In the United States, several states are undertaking integration initiatives as part of the No Child Left Behind Policy. For example, Smart Start in North Carolina is a statewide initiative to enhance child-care center quality and ensure partnerships with family-support programs. Smart Start, a public-private initiative, provides early education funding to all of the state's 100 counties. Smart Start funds are administered at the local level through nonprofit partnership organizations. The North Carolina Partnership for Children, Inc. (NCPC) is a statewide nonprofit organization set up to provide oversight and technical assistance for local partnerships. Smart Start funds are used to improve the quality of child care, make child care more affordable and accessible, provide access to health services, and offer family support. Services at the local level vary depending on local needs. Approximately, 80 local partnerships have been established throughout the state to administer funding and programs. Smart Start has garnered much national recognition and is considered a model for comprehensive early

childhood education initiatives. In 2003, the Smart Start evaluation team examined the impact on program quality, development of cross-sectoral partnerships, and child outcomes and found that children attending Smart Start centers do better on measures of cognitive, language, and social skills regardless of family or ethnic background.

In Canada, the Toronto First Duty Project (2001–06) was a demonstration research project that aimed to integrate and expand early education, family support, and child care in five neighborhoods. The project was designed to demonstrate and study how existing early childhood and family support programs could be transformed into an integrated system for children aged 0–6 years. The five sites engaged in activities to bring together and integrate early-years services to children and families in their respective communities. The findings from this demonstration project indicated that parenting capacity (including confidence in helping learning at home and involvement with the school) was greater at the first-duty sites than in schools with similar socioeconomic characteristics and traditional early childhood programs (Corter *et al.*, 2007).

Within the context of the Australia-government approach to improve the proportion of children acquiring the basic skills for life and learning, South Australia, among other states, has developed a strategic plan that incorporates an integration perspective. Twenty-four children's centers for early childhood development and parenting are under development. The strategic plan further emphasizes the need to establish and network Aboriginal children's centers in remote communities. South Australia Action Plan for Early Childhood and Child Care includes the development of children's centers organized as hub-and-spoke models and linking children's services to a broad spectrum of support and referral networks. The plan also proposes to upgrade the qualifications of the early childhood workforce, building new competencies for an integrated multidisciplinary early childhood system.

The need to establish hubs and networks to support a variety of children's service needs has also been recognized in Canada. The Healthy Babies, Healthy Children (HBHC) Program was designed as a universal prevention/early-intervention program to improve the well-being and long-term prospects of young children. Using a community-wide planning and implementation process that involved most or all organizations and agencies that serve families and children (prenatal to age 6), HBHC's goal was to ensure an effective system of prevention, screening, and early-intervention services. It emphasized the early identification and prevention of problems and built on the strengths of families and community members. Central to this project was an integrated system of services and supports for healthy child development and family well-being (Ryan and Robinson, 2005).

Ingredients of Service Integration

Whether countries have already integrated their ECEC administration systems or not, there is widespread recognition that integration is a process that goes beyond jurisdictional integration. Moving toward integration requires a willingness to leave old identities behind – both professional and programmatic.

In countries, such as the USA, Ireland, Australia, and Canada, where spending on ECEC continues to be below the OECD average, insufficient government funding remains a huge obstacle to the optimum development of ECEC services. In those countries, there is a major challenge to ensure that government resources are sufficient to achieve the outcomes promised through integrated ECEC programs. An expectation of higher-quality programs and greater effectiveness, and the associated costs that accompany this, is enshrined in movements toward integrated ECEC provision. Consequently, government funding remains a key ingredient for future progress.

The structural elements – funding, legislation, and regulation, are essential, but not enough. Two key ingredients for reform are workforce reorganization and the curriculum (or pedagogic approach). These policy areas require a fundamental change in order to achieve full integration.

Reorganizing the Workforce

One indicator of the success of integration is the unification of a divided education and care workforce (Cohen *et al.*, 2004). In those countries that have made the most progress in integrating early childhood services, a core profession has emerged. These professionals work across the early childhood age spectrum from birth to age 5 or 6 and undertake all aspects of education and care.

The OECD recommended a review of professional profiles, improved recruitment levels, and strengthening of the initial and in-service training of staff. As early childhood centers become the foundation for lifelong learning, staff in these centers will be required to deal sensitively with cultural issues, respond appropriately to children with special needs, and provide individualized support to every child in moments of vulnerability or stress. The success of integrated early learning and care programs depends primarily on the excellence of the human resources employed. Most proposals for integrated or seamless ECEC identify human resources as the critical place for transformation if integration activities are to succeed.

A core profession has emerged in countries that have gone furthest in integrating early childhood services. Dedicated and qualified professionals work across the early childhood age spectrum from birth to age 5 or 6 or 7 and engage in all aspects of education and care. Examples include Denmark, Finland, New Zealand, Norway, Spain,

and Sweden. Some countries are still engaged in this transformation. For example, in New Zealand, all early childhood teachers are required to have a teaching diploma by 2012. Incentive grants and prior learning assessment, together with time for phase-in, make upgrading possible. In Norway, considerable resources have been dedicated to improving quality and hiring more qualified teachers.

In reviewing the training issues, the OECD has raised key questions:

- Is integration of the workforce possible? If so, who will be the core professional and what will be the balance between this professional and others working in the system?
- What educational requirements will this professional have?
- Who will pay for training and who will pay for a properly qualified workforce?
- What kind of phase-in period will be established for the new core professional?
- Will all staff be required to have the same training – as in New Zealand – or will there be a role for an assistant trained at a lower level but with access to opportunities to gain additional qualifications?
- How will the workforce be reorganized to accommodate longer hours, and a longer year?

However these questions are answered, without reform to ensure adequate core funding, governments' ability to provide proper remuneration, training, and development for a substantially underpaid and undervalued workforce will be severely compromised.

Integrating Curricular and Pedagogical Frameworks

A second key challenge for the successful integration of ECEC services centers on the successful blending of curriculum and pedagogy. As Bennett points out in **Table 1**, there are two distinct traditions of the pedagogical approach in the field of early childhood education. These often present themselves in the two separate spheres: a favoring of the preprimary tradition by programs originating in education jurisdictions and a favoring of the social-pedagogic tradition by programs oriented to the care of very young children. The social-pedagogy approach tends to emphasize broad development goals, which contribute to children's overall development. The preprimary approach tends to place more emphasis on focused cognitive goals such as mathematical development, language, and literacy.

Educators rooted in the social-pedagogic tradition argue that effective pedagogy involves the provision of enriched learning and play environments, freely chosen activities by children, and responsive accompaniment of children by educators who guide, inform, model, and instruct, but do not dominate the child's thinking.

Aware that children may be able to take on the challenge of a focused cognitive curriculum, these educators oppose the prescribed detailed literacy goals for all young children as it threatens to put undue pressure on educators and children. When begun too early, literacy instruction may actually harm the self-concept of young children, leading to anxiety, low self-esteem, and mediocre literacy results.

Table 1 Pedagogical goals

<i>The direct instruction tradition</i>	<i>The Swedish constructivist tradition</i>
<p>1. The early childhood centre: Often seen as a junior school, a place of instruction, socialisation and preparation for school</p> <p>2. Approach to curriculum: Centralised development of curriculum – stressing autonomy and competition</p> <p>3. Organisation of curriculum: Often prescriptive: clear targets and outcomes, detailed competencies. . . Learning is stepped (sequential) – see first draft EYFS curriculum</p> <p>4. Focus of work: A focus on learning standards, teachers, classroom environment, preparation for school, on literacy. . . Children can or cannot read. . . Teaching subjects privileged</p> <p>5. Approach to outputs: It is necessary to reach government standards. High stakes assessments and sometimes testing required</p> <p>6. A focus on competence in the national language, is sought. Technical approach and assessment: oral, phonemic. . . with some formal skills in writing and reading from an early age</p>	<p>1. The early childhood centre: viewed as a life space, a place in which children and pedagogues “learn to be, learn to do, learn to learn, learn to live together”</p> <p>2. Approach to curriculum: Conceived as a broad central guideline stressing fundamental aims and values - with local development of curriculum</p> <p>3. Organisation of curriculum: Broad orientations rather than prescribed outcomes: play, music, arts, project work, interaction with peers and nature – the natural learning strategies of the child. Learning occurs informally (and in many places) and especially within the planned holistic experience (rich project work)</p> <p>4. Focus of work: on a range of developmental goals and living together in a learning community of educators and peers. . . the competent child. Focus on communication, literacies. . . Culturally valued ‘topics of learning’ privileged. . . child interest</p> <p>5. Approach to outputs: Parental and child satisfaction predominate. Holistic goals for each child to aspire to are set with unobtrusive, developmental assessment</p> <p>6. A growing focus on individual language and oral competence. . . with much ‘play’ reading and writing</p>

Source: Bennett, Montreal, 2009.

A curriculum helps to ensure that staff cover important learning areas, adopt a common pedagogical approach, and reach for a certain level of quality across age groups and regions. As a result, support for a blended curriculum is critical to successful integration and is also necessary to guide the staff in centers, especially when they have low certification and little training. Curricular frameworks also help to involve families in curriculum development and implementation. Although knowledgeable about their children, many parents need professional assistance to support their children's learning and development.

It is not yet clear that any jurisdiction has determined the appropriate amount of literacy, numeracy, and scientific/technological knowledge in an integrated early childhood curriculum or how it should be ensured. But experts are clear that programs must be oriented to the broad developmental goals and not just focused on cognitive goals (Bennett, 2004).

New Zealand is a case in point. After Iceland, New Zealand was the next country to administratively integrate its child care and kindergarten programs under one department. An important element in the integration process was the development of an integrated, national early childhood curriculum, Te Whaariki, which was developed in partnership with Maori and after months of consultation across the country. The aims of Te Whaariki include: well-being, belonging, contribution, communication, and exploration. Implementation requires staff understanding of theory and reflective practice. It has been incorporated into teacher education, professional development, and ECEC resources. The New Zealand Government also developed a strategic plan emphasizing improved quality and collaborative relationships. Although divisions still exist, integration is based on theory, philosophy, research, and political action. Status, recognition, and support for ECEC have increased greatly and the system has moved from an *ad hoc* to a planned approach. Researchers have found that general support for the curriculum is high but, as observed elsewhere, current regulations and funding make it difficult for centers to meet the quality expectations (Carr *et al.*, 2002). Several training institutions have begun using Te Whaariki as a framework, and evaluation and assessments are now ongoing (Mitchell, 2008).

The challenge now is to bring practice in line with the theory. The essential ingredients required to meet this challenge include professional development of teachers and a strong policy framework to support enhanced materials and approaches to quality provision.

Conclusion

There are many structural, political, professional, and historic barriers to overcome before there is an integrated

approach to ECEC services for young children in many OECD countries. The goal of integrated services is to strive for collaboration, mutual respect between sectors and communities, and develop options based on collective imagination. Many countries have already made considerable progress toward integration of their ECEC programs. The pioneering efforts of these countries may provide some direction to other countries seeking to move toward integrated approaches to ECEC provision.

See also: Early Childhood Education and Care Professional Profiles in the United Kingdom: Multiagency Teamwork and Transdisciplinary Challenges; Parent Support in Early Childhood – Approaches and Outcomes; Teaching in Early Childhood Centers Instructional Methods and Child Outcomes.

Bibliography

- Bennett, J. (2004). Curriculum in early childhood education and care. *UNESCO Policy Briefs on Early Childhood*. United Nations Educational Scientific and Cultural Organisation.
- Carr, M., May, H., and Podmore, H. (2002). Learning and teaching stories: Action research on evaluation in early childhood in Aotearoa-New Zealand. *European Early Childhood Education Research Journal* 10(2), 115–125.
- Choi, S. (2002). Integrating early childhood into education: The case of Sweden. *UNESCO Policy Briefs on Early Childhood*. United Nations Educational Scientific and Cultural Organisation, 2002.
- Cohen, B., Moss, P., Petrie, P., and Wallace, J. (2004). *A New Deal for Children? Re-Forming Education and Care in England, Scotland and Sweden*. Bristol: Policy Press.
- Cortier, C., Bertrand, J., Pelletier, J., *et al.* (2007). *Toronto First Duty Phase 2 Research Progress Report: December 2007*. Toronto: The Toronto First Duty Research Team.
- Mitchell, L. (2008). *Early Childhood Education Services in 2007: Key Findings from the NZCER National Survey*. New Zealand Education Research Council.
- OECD (2001). *Starting Strong: Early Childhood Education and Care*. Paris: OECD.
- Penn, H., Barreau, S., Butterworth, L., *et al.* (2004). *What is the Impact of Out-of-Home Integrated Care and Education Settings on Children Aged 0-6 and Their Parents?* Unpublished manuscript.
- Ryan, B. and Robinson, R. (2002). *Evaluation of the Healthy Babies Healthy Children Program Consolidated Report, end of Fiscal Year 2000–2001*. Unpublished document. Prepared for Ministry of Health and Long-Term Care, Government of Ontario.

Further Reading

- Ferguson, M. and Murray, J. (1998). Kids'n us: A review of challenges, lessons learned and strategies for community development. *Independent Review*. South East Grey Community Outreach, Markdale, Canada.
- Government of South Australia (2007). *South Australia's Action Plan for Early Childhood and Child Care*, Adelaide, SA.
- McCain, M. and Mustard, F. (1999). *The Early Years Study: Reversing the Real Brain Drain*. Toronto: Ontario Children's Secretariat.
- National Research Council (2001). *Eager to Learn: Educating Our Preschoolers*. Committee on Early Childhood Pedagogy. In Barbara, T., Bowman, M., Donovan, S., and Burns, M. S. (ed.) *Commission on Behavioral and Social Science and Education*. Washington, DC: National Academy Press.

New Zealand Ministry of Education (1996). *Te Whaariki Early Childhood Curriculum*, Wellington, NZ.

NICHD Early Child Care Research Network (1999). Child outcomes when child care center classes meet recommended standards for quality. *American Journal of Public Health* **89**(7), 1072–1077.

OECD (2002). *Babies and Bosses: Reconciling Work and Family Life: Australia, Denmark and the Netherlands*, vol. 1. Paris: OECD.

Schweinhart, L. J. and Weikart, D. P. (1997). The high/scope preschool curriculum comparison study through age 23. *Early Childhood Research Quarterly* **12**(2), 117–143.

Smart Start Evaluation (2003). *Smart Start and Preschool Child Care Quality in North Carolina: Change Over Time and Relationships to Children's Readiness*. Chapel Hill, NC: University of North Carolina at Chapel Hill, FPG Child Development Institute.

Sotolongo, J. (2008). Smart start pilot grants: Assuring better child health and development, May 2008, *Evaluation Report*. The North Carolina Partnership for Children.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., Taggart, B., and Elliot, K. (2003). *The Effective Provision of Pre-School Education (EPPE) Project: Findings from the Pre-School Period*. London: Institute of Education, University of London and SureStart.

Relevant Website

<http://www.surestart.gov.uk> – Department for Children, Schools and Families, Sure Start.

http://www.ontario.ca/en/initiatives/early_learning/ONT06_018865.

<http://www.oise.utoronto.ca/atkinson/UserFiles/File/IntegrationColley06.pdf>.

<http://www.skolverket.se/content/1/c4/09/44/00-531.pdf>.

<http://www.educate.ece.govt.nz/learning/curriculumAndLearning/TeWhariki.aspx>.

Issues of Access and Program Quality

J T Hustedt and W S Barnett, Rutgers University, New Brunswick, NJ, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Child care – Early childhood program designed to offer care to children but without a primary emphasis on providing educational services.

Early childhood education and care – Organized program offering educational services and/or care to children during early childhood. Here, we have focused on programs serving children from birth until entry into primary school.

Enrollment rate – Percentage of children from a pre-defined group who are on record as current participants in a school or early childhood education and care program.

Externalities – Spillover benefits of an early childhood program to people beyond those enrolled as children or their families.

Gross enrollment ratio (GER) – Percentage of children enrolled in the age group that corresponds to a particular grade in school.

Head start – A US federal government program providing education and other services to children in poverty, primarily from age 3 to 5, and their families.

Legal entitlement to education – Children with legal entitlements to education are guaranteed access to a program, typically based on their age.

Pre-primary education program – An educational program designed to serve children between ages 3 and the primary school entry age. These types of programs may be known by terms such as preschool, nursery education, infant education, prekindergarten, and kindergarten.

Targeted preschool education – Education program that is targeted to a particular subset of children within a given age range, often based on family income.

Universal preschool education – An education program available to all children within a specified age range (such as the year before kindergarten eligibility) whose parents wish to enroll them.

countries around the globe. There is consensus across a variety of fields that early childhood is a sensitive period of life, marked by development of critical cognitive, social–emotional, and physical skills. Over the past few decades, a large body of research has shown that participation in early childhood education programs can produce positive impacts on children’s development. Improvements in early learning and development in turn produce long-term improvements in achievement, educational attainment, and social behavior. Specific impacts shown in longer-term studies of high-quality preschool programs include decreased rates of grade retention and special education enrolment, increased school completion rates, and reductions in crime and delinquency rates.

Although the aims of ECEC initiatives vary considerably from country to country, a common priority of these programs is to offer educational services and/or care to children during the years prior to their entry into primary school, which occurs by age 6 in a majority of countries. Implementation of ECEC programs internationally has been marked by divisions between the types of services offered to children from birth to age 3 and those offered for children from ages 3 to 6. When government agencies are involved in the provision of ECEC, this age boundary is often replicated at the government level as well, with different agencies or ministries responsible for overseeing ECEC programs depending on a child’s age. Regardless of the level of government involvement, ECEC programs for children younger than age 3 tend to focus mainly on child care, while ECEC programs for children ages 3 and older tend to be conceptualized as pre-primary education.

Several Decades of Steady Growth in ECEC

Worldwide enrolment rates in preschool education programs have grown substantially since the 1960s. This rise in enrolment has occurred in the context of broad societal changes, including increased workforce participation by mothers of young children, higher numbers of single-parent households, decreasing birth rates, and urbanization. There has been burgeoning interest by both parents and governments in increasing the availability of educational opportunities for young children. One widely known example of an early government-sponsored ECEC initiative is the United States’ federal Head Start program, for young children in poverty and their families, which was started in 1965.

The Importance of Early Childhood Education and Care

Early childhood education and care (ECEC) programs for young children have become increasingly available in

Although many nations continue to target public resources on programs for disadvantaged children, there is a growing movement to ensure that all children can access preschool programs. Countries that were among the initial leaders in making ECEC widely accessible to all young children, regardless of socioeconomic background, include Belgium, France, and the Netherlands. In Belgium, nearly all children aged 3–6 were enrolled in ECEC programs by 1975. By 1980, almost all children in France aged 2–6 and almost all children aged 4–6 in the Netherlands were enrolled in ECEC. In general, differences in enrolment rates between countries have been diminishing, as ECEC becomes a higher priority internationally. Pre-primary enrolment rates are now about 3 times higher than in 1970, although rates are still low in less economically developed countries. Most countries in the Organization for Economic Cooperation and Development (OECD), which primarily counts industrialized countries as its members, are currently offering free pre-primary education programs of at least 2 years' duration.

Current Issues of Access to ECEC

International data on ECEC enrolment rates are gathered on an ongoing basis by the United Nations Educational, Scientific and Cultural Organization (UNESCO), which monitors progress toward a set of six Education for All goals adopted at the 2000 World Education Forum by 164 governments and 35 international institutions. These goals focus on increasing access to educational opportunities at all stages of life, as a human rights issue. The goal related to early childhood involves increasing access to comprehensive ECEC programs, and making improvements in these programs, with a particular emphasis on serving disadvantaged children. Data from UNESCO are reported in terms of the gross enrolment ratio (GER), a percentage of children enrolled from the age group matching a particular grade level in school. (A GER higher than 100% reflects enrolment by children whose age does not correspond to that expected at their current grade level, because they entered school at an atypical age or were retained.) The UNESCO data focus on enrolment in pre-primary education programs beginning at age 3 – programs offering a school-type, center-based environment, including programs variously known by such terms as preschool, nursery education, infant education, and kindergarten. Enrolment data for children from birth until age 3 are less widely available on an international level.

Pre-primary enrolment rates tend to be much higher in developed countries than in developing countries. The regions of the world with the highest pre-primary GERs in 2005 included North America and Western Europe (79%), Latin America and the Caribbean (62%), and Central and Eastern Europe (59%). In East Asia and the

Table 1 International pre-primary enrolment rates

Country or territory	Age group, 2005	Percent enrolled in pre-primary, 2005	Private as percent of total, 2005
<i>Arab States</i>			
Algeria	4–5	6	–
Bahrain	3–5	47	99
Djibouti	4–5	1	84
Egypt	4–5	16	31
Iraq	4–5	6 ^c	–
Jordan	4–5	31	95 ^b
Kuwait	4–5	73	37
Lebanon	3–5	74	77
Libyan Arab Jamahiriya	4–5	8 ^b	15 ^{b, c}
Mauritania	3–5	2 ^c	78 ^c
Morocco	4–5	54	100
Oman	4–5	8 ^b	100 ^b
Palestinian Autonomous Territories	4–5	30	100
Qatar	3–5	36	94
Saudi Arabia	3–5	10	45
Sudan	4–5	25	71
Syrian Arab Republic	3–5	10	74
Tunisia	3–5	22 ^{b, c}	–
United Arab Emirates	4–5	64	75
Yemen	3–5	1	49
<i>Central and Eastern Europe</i>			
Albania	3–5	49 ^b	5 ^b
Belarus	3–5	100	5
Bulgaria	3–6	79	0.3
Croatia	3–6	47 ^b	8 ^b
Czech Republic	3–5	100	1
Estonia	3–6	100	2
Hungary	3–6	83	5
Latvia	3–6	84	3
Lithuania	3–6	68	0.1
Poland	3–6	54	8
Republic of Moldova ^a	3–6	62	1
Romania	3–6	75	1
Russian Federation	3–6	84	1
Slovakia	3–5	95	1
Slovenia	3–5	79	1
TFYR Macedonia	3–6	33	–
Turkey	3–5	10	4
Ukraine	3–5	86	3
<i>Central Asia</i>			
Armenia	3–6	33	1
Azerbaijan	3–5	29	0.1
Georgia	3–5	51	–
Kazakhstan	3–6	34	5
Kyrgyzstan	3–6	13	1
Mongolia	3–6	40	1
Tajikistan	3–6	9	–
Uzbekistan	3–6	28 ^{b, c}	–
<i>East Asia and the Pacific</i>			
Australia	4–4	100	66
Brunei Darussalam	3–5	52	65
Cambodia	3–5	9	24
China	4–6	40 ^b	31 ^b

Continued

Table 1 Continued

Country or territory	Age group, 2005	Percent enrolled in pre-primary, 2005	Private as percent of total, 2005
Cook Islands ^a	4–4	91 ^{b, c}	22 ^{b, c}
Fiji	3–5	16 ^c	100 ^c
Indonesia	5–6	34	99
Japan	3–5	85	66
Kiribati	3–5	75 ^{b, c}	–
Lao People's Democratic Republic	3–5	9	26
Macao, China	3–5	92	95
Malaysia	5–5	100 ^b	45 ^b
Marshall Islands	4–5	50 ^{b, c}	18 ^{b, c}
Nauru ^a	3–5	71 ^{b, c}	17 ^{b, c}
New Zealand	3–4	93	98
Niue ^a	4–4	100 ^c	–
Palau ^a	3–5	64 ^c	20 ^c
Papua New Guinea	6–6	59 ^{b, c}	–
Philippines	5–5	41	45
Republic of Korea	5–5	96 ^b	77 ^b
Samoa	3–4	49 ^{b, c}	–
Solomon Islands	3–5	41 ^{b, c}	–
Thailand	3–5	82 ^b	21 ^b
Timor-Leste	4–5	16	–
Tokelau ^a	3–4	100 ^{b, c}	–
Tonga	3–4	23 ^c	12 ^c
Tuvalu ^a	3–5	99 ^b	–
Vietnam	3–5	60	58
<i>Latin America and the Caribbean</i>			
Anguilla	3–4	97 ^c	100
Argentina	3–5	64 ^b	27 ^b
Aruba ^a	4–5	99	77
Bahamas	3–4	31 ^{b, c}	79 ^{b, c}
Barbados	3–4	93	17
Belize	3–4	33	96
Bolivia	4–5	50 ^c	23 ^{b, c}
Brazil	4–6	63 ^b	29 ^b
British Virgin Islands ^a	3–4	90	100
Cayman Islands	4–4	93 ^c	91
Chile	3–5	54	48
Colombia	3–5	39	38
Costa Rica	4–5	69	10
Cuba	3–5	100	–
Dominica ^a	3–4	78	100
Dominican Republic	3–5	34	43
Ecuador	5–5	77 ^c	47 ^c
El Salvador	4–6	51	18
Grenada ^a	3–4	81 ^c	58 ^b
Guatemala	3–6	28	19
Guyana	4–5	100	3
Honduras	3–5	33 ^c	23 ^c
Jamaica	3–5	95	91
Mexico	4–5	93	13
Montserrat ^a	3–4	100	–
Netherlands Antilles	4–5	100 ^{b, c}	75 ^b
Nicaragua	3–6	37	16
Panama	4–5	62	18
Paraguay	3–5	31 ^b	27 ^b
Peru	3–5	62	21
Saint Kitts and Nevis ^a	3–4	100	59

Continued

Table 1 Continued

Country or territory	Age group, 2005	Percent enrolled in pre-primary, 2005	Private as percent of total, 2005
Saint Lucia	3–4	74	100
Saint Vincent and the Grenadines	3–4	86 ^c	100 ^c
Suriname	4–5	89	45
Trinidad and Tobago	3–4	87	100
Turks and Caicos Islands	4–5	100 ^c	65
Uruguay	3–5	62 ^b	20 ^b
Venezuela	3–5	58	18
<i>North America and Western Europe</i>			
Andorra ^a	3–5	100	2
Austria	3–5	91	27
Belgium	3–5	100	53
Cyprus ^a	3–5	65	40
Denmark	3–6	93	–
Finland	3–6	59	8
France	3–5	100	13
Germany	3–5	98	59
Greece	4–5	67	3
Iceland	3–5	94 ^b	8 ^b
Israel	3–5	92	4
Italy	3–5	100	30
Luxembourg	3–5	86	6
Malta	3–4	100	39
Monaco	3–5	–	19 ^b
Netherlands	4–5	90	70 ^b
Norway	3–5	88	42
Portugal	3–5	77	47
Spain	3–5	100	35
Sweden	3–6	88	14
Switzerland	5–6	99	8
United Kingdom	3–4	59	8
United States	3–5	61	38
<i>South and West Asia</i>			
Afghanistan	3–6	1 ^{b, c}	–
Bangladesh	3–5	11 ^b	53
Bhutan	4–5	–	100
India	3–5	41	4 ^b
Iran, Islamic Republic of	5–5	46	8
Maldives	3–5	49	38
Nepal	3–4	27 ^b	80 ^b
Pakistan	3–4	50	–
<i>Sub-Saharan Africa</i>			
Benin	4–5	5	37
Burkina Faso	4–6	2	–
Burundi	4–6	2	47
Cameroon	4–5	24	66
Cape Verde	3–5	54	–
Central African Republic	3–5	2 ^{b, c}	–
Chad	3–5	1 ^c	47 ^b
Comoros	3–5	3 ^c	62 ^c
Congo	3–5	6	77
Côte d'Ivoire	3–5	3 ^b	46 ^{b, c}
Democratic Rep. of the Congo	3–5	1 ^{b, c}	84 ^{b, c}

Continued

Table 1 Continued

Country or territory	Age group, 2005	Percent enrolled in pre-primary, 2005	Private as percent of total, 2005
Equatorial Guinea	3–6	41	49
Eritrea	5–6	12	48
Ethiopia	4–6	2 ^b	100
Gambia	3–6	18 ^{b, c}	100 ^{b, c}
Ghana	3–5	56 ^b	34
Guinea	3–6	7	91 ^b
Kenya	3–5	52	31
Lesotho	3–5	34	100
Madagascar	3–5	10 ^{b, c}	90 ^b
Mali	3–6	3	–
Mauritius	3–4	95	83
Namibia	3–5	29 ^{b, c}	100 ^{b, c}
Niger	4–6	1	32
Nigeria	3–5	15	–
Sao Tome and Principe	3–6	32	–
Senegal	4–6	8	68
Seychelles ^a	4–5	100	5 ^b
South Africa	6–6	37 ^b	7 ^b
Swaziland	3–5	18 ^b	–
Togo	3–5	2 ^{b, c}	59 ^{b, c}
Uganda	4–5	1	100
United Republic of Tanzania	5–6	30 ^b	2 ^b
Zimbabwe	3–5	43 ^b	–

^aEnrolment rates were calculated using national population data.

^bData are from a year other than 2005. In all cases, data from 2003 or later were used.

^cEstimated data were used.

For purposes of this table, gross enrolment rates exceeding 100% are capped at 100%. From United Nations Educational, Scientific and Cultural Organization (2007). *Education for All Global Monitoring Report 2008: Education for All by 2015. Will We Make It?* Paris: UNESCO.

Pacific, 43% of children were enrolled in pre-primary education; and in South and West Asia 37% of children were enrolled. Less than one-third of children were enrolled in pre-primary education programs in Central Asia (28%), the Arab States (17%), and sub-Saharan Africa (14%). From 1999 to 2005 (the years just after the adoption of the Education for All goals), pre-primary enrolment rates grew by 19% worldwide, with a 30% increase in transition countries, a 24% increase in developing countries, and a 6% increase in developed countries. Regions with the lowest enrolment rates at the beginning of that time period tended to show the greatest percentage increases, although these areas of the world continue to offer fairly limited access to ECEC programs for children between age 3 and primary school entry. See **Table 1** for country-by-country enrolment rates in 2005 from UNESCO. **Table 2** provides further details about enrolment in 20 OECD countries, focusing on legal entitlements and

enrolment rates for children from birth to age 3 and from age 3 until primary school entry.

Issues of ECEC Program Administration and Governance

Although, by definition, ECEC specifically encompasses services for young children starting at birth, the reality is that most policies and organized programs focus on children starting at age 3. Only about 53% of countries internationally have at least one type of program for children from birth to age 3; such programs are most prevalent in North America and Western Europe and are least prevalent in the Arab States and in Central and Eastern Europe. In countries that offer ECEC services prior to age 3, programs for older and younger preschoolers are often separately organized and administered, with those for younger children focusing on child care and those for older children focusing on education. Thus, programs for infants and toddlers may be separate from a country's pre-primary education programs rather than a downward extension of them.

The governmental divisions with responsibility for ECEC programs often differ depending on children's ages. From birth until age 3, ministries of health frequently administer child-care programs, whereas other ministries – particularly ministries of education – are more likely to be involved in ECEC for children age 3 and older. Research on early childhood education systems in France and Sweden suggests that as government education agencies took more responsibility for ECEC programs, this provided the resources for increased access for children and also helped to professionalize early childhood staffing.

The ways in which ECEC programs are paid for also tend to vary, depending on children's ages and whether the program is conceptualized as child care or education. For example, in OECD countries, governments bear the majority of the costs for pre-primary education programs. However, parents generally must contribute larger shares toward the total costs when their children attend ECEC programs for infants and toddlers or programs for older children that are considered to be child care. The parental share of costs for these programs can be very high in countries such as Canada and the United States, where parents may bear the entire cost unless they qualify for programs targeting families in or near poverty.

Pre-primary education programs are much more likely to be provided in public school settings than in private settings internationally. The use of private settings is most common in developing countries, where a median 47% of pre-primary education programs are in private settings. Less than 10% of programs in developed countries and almost none of the programs in transition countries occur in private settings. UNESCO reports that use of private

Table 2 ECEC programs from birth until primary school in OECD countries

<i>ECEC entitlements and enrolment rates</i>		
<i>Country</i>	<i>Birth to age 3</i>	<i>Age 3 until primary school entry</i>
Australia	There is no legal entitlement to ECEC. Child-care benefits are available for children enrolled in approved services. Accredited and family day care centers reach approximately 25% of children.	There is no legal entitlement to ECEC. Free or nearly free preschool is provided by most states at ages 4 and 5. Accredited and family day care centers reach up to 62% of 4-year-olds. Kindergarten or reception classes reach 17% of 4-year-olds and 84% of 5-year-olds. Primary school begins at age 6.
Austria	There is no legal entitlement to ECEC. About 9% of children attend family day care or crèche programs.	Children aged 3–6 have a legal right to fee-based kindergarten, and 80% are enrolled. Primary school begins at age 6.
Belgium	The legal right to schooling begins at age 2.5. Earlier, subsidized services are widely available, although supply is less than demand. In the Flemish community, family day care and crèches are most common, reaching 34% of children. In the French community, crèches and centers are most common, reaching 18% of children.	Children aged 2.5–6 are entitled to free school programs. At age 2.5, 90% of children are enrolled, and almost 100% are enrolled by age 3. Primary school begins at age 6.
Canada	Entitlements to schooling vary by province. In Quebec, the right to educational child care begins at birth, and 38% of children from birth to age 4 are enrolled. Programs are fee-based, though parent fees are very low. Outside Quebec, provincial support is weak.	Entitlements vary by province. Fee-based junior kindergartens enroll 40% of 4-year-olds in Ontario and 50% in Quebec. Free kindergartens enrol 95% of 5-year-olds in most provinces. Primary school begins at age 6.
Czech Republic	There is no legal entitlement to ECEC, and few crèche services are available.	There is no legal entitlement, but kindergarten programs are widely accessible. Beginning at age 3, 76% of children are enrolled. Free programs are available by age 4, with priority given to 5-year-olds. By age 5, 98% of children are enrolled. Primary school begins at age 6.
Denmark	Day care services for all children ages 1 to 5 are guaranteed by 87% of municipalities. Fee-based services may begin at 6 months and continue until age 6. Family day care centers enroll 45% of children, with an additional 15% in crèches and age-integrated facilities.	Until age 6, children may continue to attend day care programs. At age 6, children have a legal right to attend free preschool in a center-based setting or primary school. About 58% of children are enrolled in kindergartens from ages 3 to 6, and 98% of children are enrolled from age 6 until age 7. Primary school begins at age 7.
Finland	Beginning at birth, children have a legal right to enroll in a center- or home-based ECEC program. Family day care and municipal programs reach 28% of children from ages 1 to 2 and 44% of children aged 2–3. Programs are fee-based, with parents paying an average of 15% of costs in public day care programs.	Children have a legal right to enroll in a center- or home-based ECEC program. By age 5, 73% are enrolled. At age 6, preschool classes in centers and primary schools are free, and 96% of children are enrolled. Primary school begins at age 7.
France	There is no legal entitlement to ECEC, although subsidized services are widely available. About 30% of children are enrolled in various types of settings. Starting at age 2, about 35% of children are enrolled in free educational programs.	Children have a legal right to free ECEC services, and almost all children attend until primary school begins at age 6.
Germany	There is no legal entitlement to ECEC, but 37% of children in the former East Germany and about 3% of children in the former West Germany attend center-based crèche programs, for a total of 9% nationwide.	Children have a legal right to fee-based kindergarten programs from ages 3 to 6, and 90% are enrolled. Primary school begins at age 6.
Great Britain	There is no legal entitlement to ECEC. Under the responsibility of the Ministry of Education, 26% of children attend private nurseries or play groups, or are cared for by child minders.	At ages 3 and 4, children are legally entitled to enroll in a free, part-time ECEC program. At age 3, 95% of children attend play groups or nurseries, and at age 4 nearly 100% attend reception classes or nursery schools. Primary school begins at age 5.
Hungary	Children of working parents have a legal right to free ECEC at 6 months of age, but only 9% of children attend.	Children have a legal right to free kindergarten programs starting at age 3. From ages 3 to 5, 85% of children are enrolled. Compulsory kindergarten starts at age 5, and 97% of children are enrolled.

Continued

Table 2 Continued

Country	<i>ECEC entitlements and enrolment rates</i>	
	<i>Birth to age 3</i>	<i>Age 3 until primary school entry</i>
Ireland	There is no legal entitlement to ECEC. The majority of children are in family or informal child care, although 10–15% of children up to age 4 attend licensed family day care centers or nurseries.	Up to age 4, 10–15% of children attend licensed family day care centers or nurseries, and 4% of 3-year-olds attend pre-primary education. The legal right to free preschool begins at age 4, and 56% of children ages 3 to 6 are enrolled in public pre-primary programs. By age 5, enrolment rates are nearly 100%. Primary school begins at age 6.
Italy	There is no legal entitlement to ECEC. The majority of children are in family or informal child care, although 19% are enrolled in crèches.	Children have a legal right to enroll in school-based ECEC; programs in public settings are free. From ages 3 to 4, 98% of children are enrolled in regulated services (although rates vary by region); this increases to 100% from ages 5 to 6. Primary school begins at age 6.
Korea	There is no legal entitlement to ECEC. Family and informal child care are common, although 10% of children are enrolled in child care centers.	Four-year-olds from low income families gained a legal right to ECEC in 2006. Otherwise, the right to free ECEC begins at age 5. Child care centers enrol 31% of children ages 3 to 5. Ministry of Education kindergartens enrol 12% of children at age 3, 27% at age 4, and 45% at age 5. About 70% of 5-year-olds receive licensed services. Primary school begins at age 6.
Mexico	There is no legal entitlement to ECEC, and only 3% of children are enrolled in crèches.	Currently, more than 80% of children participate in ECEC at age 5, with enrolment rates for younger children much lower. Starting in 2009, free, school-based compulsory education will begin at age 3.
Netherlands	There is no legal entitlement to ECEC, although subsidies are available for children at risk. Family day care and child care centers cover 23% of children from birth to age 4, and another 5–10% attend municipal education services for children who are disadvantaged.	Although there is no legal right to ECEC until age 4, 89% of children aged 2–4 participate in play groups or receive other types of services. Nearly all children are enrolled in free pre-primary school starting at age 4, and pre-primary school is compulsory at age 5.
Norway	There is no legal entitlement to ECEC. Private and public kindergartens enrol about 48% of children.	There is no legal entitlement to ECEC. Private and public kindergartens are available in most areas, and about 88% of children ages 3 to 6 are enrolled in fee-based ECEC. Primary school begins at age 6.
Portugal	There is no legal entitlement to ECEC. About 13% of children are enrolled in crèche programs.	A legal entitlement to free ECEC programs begins at age 4. Average enrolment for children from age 3 to age 6 is 76%, with 60% of 3-year-olds enrolled and 90% of children enrolled from age 5 to age 6. Primary school begins at age 6.
Sweden	Beginning at 1 year of age, children of working or studying parents have a right to fee-based ECEC. Full-day programs enroll 45% of children from age 1 to age 2 and 86% of children from age 2 to age 3. Family day care programs enroll 8% of children from birth to age 6.	At age 3, bilingual children are entitled to free preschool. This right is being extended to all 5-year-olds and all 4-year-olds, progressively. All 6-year-olds are entitled to free preschool; enrolment in these programs is 91%, with the other 9% already in school. Primary school begins at age 7.
United States	There is no legal entitlement to ECEC. About 50% of children attend private child care centers or family day care settings. Of these children, 38% are in licensed settings.	There is no legal entitlement to ECEC. Most states offer free public prekindergarten at age 4, and sometimes at age 3 as well, but usually to targeted populations. At age 3, 40% of children are enrolled in some type of education program, and 70% are enrolled at age 4. Most school districts offer a year of free kindergarten at age 5, and more than 90% of children are enrolled. Primary school begins at age 6.

From Organisation for Economic Co-operation and Development (2006). *Starting Strong II: Early Childhood Education and Care*. Paris: OECD.

settings has changed little during the recent expansion of pre-primary education programs, with roughly equal proportions of countries increasing their use of private providers, decreasing their use of private providers, or

experiencing few changes in their use of private providers. Nevertheless, some countries rely heavily on private provision of ECEC programs during the pre-primary years. This is particularly the case in Arab States such as Bahrain,

Morocco, and Oman; Latin American and Caribbean countries such as Anguilla, the British Virgin Islands, and Saint Lucia; and some regions of sub-Saharan Africa including Ethiopia, Lesotho, and Uganda. In these countries, pre-primary education occurs almost exclusively in private settings. **Table 1** shows percentages of children in pre-primary programs who are enrolled in private settings by country.

Moving Beyond Enrolment Counts to Assure Access to High-Quality ECEC Programs

As mentioned previously, a substantial body of research indicates that children's early learning and development can benefit significantly from preschool education, leading to long-term improvements in achievement and educational attainment as well as in social behavior. Studies have found these results across a broad range of countries with more and less developed economies. There is some evidence that benefits are greater for more economically disadvantaged children and for those who are not native speakers of the language of the country in which they reside. Moreover, the precise benefits vary with the populations and with the characteristics of programs. These findings have two important implications.

First, it is not just the percentage of the population enrolled in ECEC that is important, but who is enrolled. Thus, international comparisons find that national average achievement test scores rise with preschool enrolment, but that the degree of inequality in scores begins to narrow only as enrolment rises above 60% of the population. One reason may be that high levels of enrolment are required to effectively reach most disadvantaged children. Another reason may be that there are peer effects from more socioeconomically mixed preschool classrooms and from a higher proportion of more able students in schools after preschool education.

Second, both the hours of preschool education and the quality of preschool education have implications for benefits. Longer hours are necessary if the programs are to facilitate parental employment and the increased income and productivity that result from this employment. Children's gains depend on what opportunities to learn are provided and how well these are provided. Programs that address children's needs more broadly have a broader range of positive effects. Thus, programs are more beneficial if they address both cognitive (academic achievement) and socioemotional (social behavior) development. Where children suffer from problems of inadequate health and nutrition, programs that address these needs as well as the need for education will be more beneficial. What teachers do in the classroom matters a great deal for the magnitude of

benefits to children's learning. As a result, the preparation, ongoing professional development, and supervision of teachers are important determinants of program benefits.

Very little comparative information has been reported about the quality of ECEC programs on a worldwide scale. However, due to differences in funding, policies, and standards, ECEC appears to be far from uniform with respect to both quality and quantity (age at start and length of day) both across and within nations. Program standards in ECEC are more variable than for K-12 education and often are quite low. Although some variation is to be expected because of variations in children's needs and the capacity of countries to provide for those needs, variations in which children are enrolled and in the characteristics of programs are likely to influence the effectiveness of early education in improving children's learning and development and supporting families.

Thus, it is important that monitoring of the national provision of ECEC programs moves beyond simply counting enrolment. Information should be available on enrolment for population subgroups – for example, children from low-income families, language minority children, and children of parents with low levels of education. Likewise, information is lacking on ECEC program characteristics including: program schedules, teacher education and training, class size, teacher–child ratio, the type and quality of children's classroom experiences, and other program services (e.g., nutritional supplementation). Direct measurement of program quality through observation of representative samples of classrooms is well advised, as classroom structural features are not strong predictors of actual practice. Conducting assessments of children's learning and development may also be useful, although practical limitations include the cost, difficulties obtaining valid assessment results for young children, and problems of attributing variations in children's assessment scores to particular program experiences.

Unfortunately, relatively few countries provide data on enrolment of subgroups, and quality data are rarely systematically collected and disseminated at national levels even in the wealthiest nations. The lack of detailed enrolment data is less of an issue, of course, in countries where enrolment is near universal. However, as mentioned earlier, even where enrolment is universal for older preschoolers it typically is far from universal for infants and toddlers. There are examples of private efforts to make such information available, including annual ratings of state preschool education programs in the United States by the National Institute for Early Education Research, and the Bertelsmann Foundation's ratings of preschool programs by state in Germany. Such efforts seek to inform both government policymakers and the general public. Examples of public efforts include government-sponsored rating systems for child-care programs that are provided

as guides to parents and may be linked to a schedule of government payments that vary with the rating attained.

Universal and Targeted ECEC Programs

Perhaps the most important policy decision governments make relating to the accessibility of ECEC programs is whether provision or subsidies for public programs should be limited to a particular population (typically, by means testing) or available to all. The arguments for targeting begin with the premise that the largest benefits accrue to children who are economically disadvantaged. If so, it is sensible to prioritize scarce public resources for those children. In addition, it is argued that governments are more able to provide higher-quality early education, with its higher cost per child served, if the total cost is kept down by serving a limited portion of the population. The arguments for public support of universal provision recognize that even if benefits are largest for disadvantaged children, the benefits to nondisadvantaged children may still far exceed costs and a universal program that serves all children together may better serve the disadvantaged children. Further, it is not always clear where to draw the line between children who are disadvantaged and those who are not. As the costs, takeup rates, and benefits of programs may vary by age and type of program (e.g., parenting education, paid parental leave, infant and toddler child care, and pre-primary education from age 3), countries may choose to target some programs and provide others to all children and families.

There are four major reasons to suppose that public support for universal preschool education may more effectively serve disadvantaged children. First, targeting has proven highly imperfect, and improving its accuracy is costly. In practice, universal programs can achieve better coverage of disadvantaged populations. When programs are targeted, there are costs of identifying and recruiting the intended population and excluding those who do not qualify. When means-testing is employed, the target population is constantly changing, as family income is not constant over time. A child who qualifies at the start of the year may not qualify 4 months later and vice versa. Frequent requalification of children for participation is costly and results in ineffective education if children are bounced in and out of an ECEC program during a single school year. This is the current policy in some programs that are operated primarily as child care. Moreover, any easily administered criterion such as means-testing fails to identify all of the children who are at elevated risk of school failure and other problems that preschool education might prevent or ameliorate. In addition, there may be social stigma associated with programs for the poor that lead some qualifying families to opt out. Second, because of the effects of peers on learning, disadvantaged

children appear to learn more when they attend programs that also include more advantaged children. Third, economic theory and experience suggest that the voters are more likely to support high-quality education if the program is available to all children and families.

Perhaps the strongest argument in favor of public support for universal preschool education is that all children benefit from high-quality preschool education, and there are substantial effects of subsidized high-quality child care on maternal employment in even the wealthiest countries. The externalities (spillover benefits to others that do not accrue to the child or parents) from high-quality preschool education are likely to remain significant for most of the nonpoor population. Even if benefits to more advantaged children are only a small fraction of those estimated for disadvantaged children, the benefits are likely to exceed the costs. If the effects on maternal labor-force participation and hours are somewhat smaller, this may be offset to some extent by the higher hourly earnings and higher taxes paid on those earnings of more advantaged mothers. In addition, countries seeking to support parents as a way of increasing the number of children born are likely to want this incentive to extend across the entire population. It should be noted that within a universal public program, the level of subsidy and intensity of service can be varied according to the individual needs of children and families. As with targeting, there are costs and practical difficulties in effectively delivering highly differentiated services.

One way of making ECEC programs universally available is by adopting policies that lower the mandatory school entry age to encompass the pre-primary years. UNESCO reports that by 2006, 30 countries had adopted policies for compulsory pre-primary education. This is a recent development, with fully half of the policies for required pre-primary attendance enacted since 1995. Countries that have lowered their school entry ages to include the pre-primary years are heavily concentrated in Latin America and the Caribbean, and also in Central and Eastern Europe. Attendance rates from countries with compulsory pre-primary attendance policies indicate that these countries, while signaling a commitment to ECEC, have not necessarily achieved universal attendance. In addition, a number of countries (particularly in Europe) offer universal programs with 4-year-old enrolment rates that exceed 90% even in the absence of mandatory attendance laws. Although national policies that address whether programs are universal or targeted offer an important starting point, other factors weigh in to accessibility of ECEC as well – including cultural attitudes toward early childhood education, government funding levels, whether ECEC programs are integrated within systems for primary education, and availability of appropriate ECEC facilities.

Conclusions

While international enrolment rates in ECEC programs have increased substantially over the past few decades, there is still wide variability across countries with respect to ECEC policies. Within this sector, countries have generally prioritized education programs for children beginning at age 3. There has been less emphasis on providing programs for infants and toddlers, and the initiatives that do exist have often focused on child care. As a result, much of the information about the availability of ECEC on an international level has focused on pre-primary education programs for children from ages 3 to 6. Most pre-primary education programs tend to occur in public settings around the world, although there is great variability from country to country, and some countries rely almost exclusively on private settings.

One complication in international comparisons of pre-primary education is that the primary school starting age varies considerably among countries. This is true even among countries where there is universal participation in the primary grades, with the compulsory primary school entry age ranging from age 4 (Northern Ireland) to age 7 (Finland, Sweden, and many other countries). Thus, 2 years of pre-primary education in some countries might still leave children with the same level of access to education as in other countries that start school at age 5 and have no pre-primary education. Interestingly, compulsory school starting age is positively related to international differences in test scores, but less strongly so than is the number of years of preschool education. This may be because universal preschool education policies to some extent compensate or negate the need for earlier compulsory schooling. At some point, the line is blurred between pre-primary and primary education, and what is important is the extent to which children can access programs and the content and methods of education rather than whether programs are administered within a primary or pre-primary education system, although the administrative auspices of a program may well influence who it serves and how it operates.

Based on the available international data, it is currently difficult to move beyond tallying enrolment rates from country to country. There is little comparative information available on a worldwide scale about which children are enrolled in ECEC, what specific services they receive, and the quality of programs they attend. These additional types of data are key to further examination of the educational opportunities available to children internationally, during the years before they begin primary school.

See also: Early Childhood Care and Education: The Family, The Market, and The State; Family Environment in the Production of Schooling.

Further Reading

- Barnett, W. S. (1998). Long-term effects on cognitive development and school success. In Barnett, W. S. and Boocock, S. S. (eds.) *Early Care and Education for Children in Poverty: Promises, Programs, and Long-Term Results*, pp 11–44. Albany, NY: SUNY Press.
- Barnett, W. S. (2007). Benefits and costs of quality early childhood education. *Children's Legal Rights Journal* 27(1), 7–23.
- Barnett, W. S., Brown, K., and Shore, R. (2004). The universal vs. targeted debate: Should the United States have preschool for all? *Preschool Policy Matters*, 6. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.
- Barnett, W. S., Hustedt, J. T., Friedman, A. H., Stevenson Boyd, J., and Ainsworth, P. (2007). *The State of Preschool 2007: State Preschool Yearbook*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.
- Barnett, W. S. and Masse, L. N. (2007). Comparative benefit–cost analysis of the Abecedarian program and its policy implications. *Economics of Education Review* 26(1), 113–125.
- Berlinski, S. and Galiani, S. (2007). The effect of a large expansion of pre-primary school facilities on preschool attendance and maternal employment. *Labour Economics* 14, 665–680.
- Berlinski, S., Galiani, S., and Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics* 92, 1416–1440.
- Boocock, S. S. and Lerner, M. (1998). Long-term outcomes in other nations. In Barnett, W. S. and Boocock, S. S. (eds.) *Early Care and Education for Children in Poverty: Promises, Programs, and Long-Term Results*, pp 45–76. Albany, NY: SUNY Press.
- Engle, P. L., Black, M. M., Behrman, J. R., et al. (2007). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *Lancet* 369, 229–242.
- European Child Care and Education Study Group (ECCE) (1997). European child care and education study: Cross national analyses of the quality and effects of early childhood programmes on children's development. *Final Report*. Berlin: Freie Universität.
- Gormley, W. T., Gayer, T., Phillips, D., and Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology* 41, 872–884.
- Melhuish, E. and Petrogiannis, K. (eds.) (2006). *Early Childhood Care and Education: International Perspectives*. New York: Routledge.
- Montie, J. E., Xiang, Z., and Schweinhart, L. J. (2006). Preschool experience in 10 countries: Cognitive and language performance at age 7. *Early Childhood Research Quarterly* 21, 313–331.
- Myers, R. G. (2003). Early childhood education: International context. In *Encyclopedia of Education*, 2nd edn. vol. 2, pp 626–634. New York: Macmillan.
- Myers, R. G. (2004). In search of quality in programmes of early childhood care and education. *Background Paper for Education for All Global Monitoring Report 2005*. Paris: UNESCO. www.unesco.org/education/gmr_download/references_2005.pdf (accessed August 2009).
- Neuman, M. J. (2008). The politics of integrating the care and education of young children in France and Sweden. In Zimanyi, L. (moderator) *Politics and Policy Planning in ECD*. Panel conducted at the meeting of the Comparative and International Education Society, New York.
- Neuman, M. J. and Peer, S. (2002). *Equal from the Start: Promoting Educational Opportunity for All Preschool Children – Learning from the French Experience*. New York: French-American Foundation.
- Olmsted, P. P., Montie, J., Claxton, J., and Oden, S. (eds.) (2001). *Early Childhood Settings in 15 Countries: What Are Their Structural Characteristics?* Ypsilanti, MI: High/Scope Press.
- Organisation for Economic Co-operation and Development (2006). *Starting Strong II: Early Childhood Education and Care*. Paris: OECD.
- Raine, A., Mellingen, K., Liu, J., Venables, P., and Mednick, S. A. (2003). Effects of environmental enrichment at ages 3–5 years on schizotypal personality and antisocial behavior at ages 17 and 23 years. *American Journal of Psychiatry* 160(9), 1627–1635.
- Rindermann, H. and Ceci, S. J. (2008). *Education Policy and Country Outcomes in International Cognitive Competence Studies*. Graz, Austria: Institute of Psychology, Karl-Franzens-University Graz.

- Schutz, G., Ursprung, H. W., and Wossmann, L. (2008). Education policy and equality of opportunity. *Kyklos* **61**, 279–308.
- Schweinhardt, L. J., Montie, J., Xiang, Z., et al. (2005). *Lifetime Effects: The High/Scope Perry Preschool Study through Age 40*. Ypsilanti, MI: High/Scope Press.
- Tietze, W. and Ufermann, K. (1989). An international perspective on schooling for 4-year-olds. *Theory into Practice* **28**(1), 69–77.
- United Nations Educational, Scientific and Cultural Organization (2006). *Education for All Global Monitoring Report 2007: Strong Foundations. Early Childhood Care and Education*. Paris: UNESCO.
- United Nations Educational, Scientific and Cultural Organization (2007). *Education for All Global Monitoring Report 2008: Education for All by 2015. Will We Make It?* Paris: UNESCO.
- Weikart, D. P., Olmsted, P. P., Montie, J., Hayes, N., and Ojala, M. (eds.) (2003). *A World of Preschool Experience: Observations in 15 Countries*. Ypsilanti, MI: High/Scope Press.
- Zigler, E., Gilliam, W. S., and Jones, S. M. (eds.) (2006). *A Vision for Universal Preschool Education*. New York: Cambridge University Press.

Relevant Websites

- <http://www.bernardvanleer.org> – Bernard van Leer Foundation.
- <http://www.ecdgroup.com> – Consultative Group on Early Childhood Care and Development (CGECCD).
- <http://www.highscope.org> – High/Scope Educational Research Foundation.
- <http://nieer.org> – National Institute for Early Education Research (NIEER).
- <http://www.oecd.org> – Organisation for Economic Cooperation and Development (OECD).
- <http://portal.unesco.org> – United Nations Educational, Scientific and Cultural Organization (UNESCO).

Kindergarten Transitions and Linkages to Primary School – Readiness Re-conceptualized

A Petriwskyj, Queensland University of Technology, Brisbane, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Horizontal transition – Transitions across one point in time such as within one day or week (e.g., kindergarten, therapy and outside-school-hours care).

Linkages – Connections between stakeholders (e.g. home and ECE, preschool and school) to share information.

Multi-faceted or multi-layered transition – Transition processes that take into account multiple changes in children's lives, and multiple strategies to improve continuity of experience.

Preparedness – Shared preparation of children, families, schools and communities for children's transition to school.

Readiness for school – Individual child characteristics such as maturity and pre-academic skills that are anticipated to enhance school adjustment and achievement.

Ready schools – Schools that have undertaken preparation to cater for the diversity of school entrants.

Structural provision – Organizational or system level provision such as set age of school entry, transition class, or transition curriculum.

System coherence – Consistency and communication between systems and agencies working with families and children.

Transition to school – On-going and non-stigmatising process of mutual adaptation by children, families, schools and communities to facilitate all children moving successfully from home and ECEC centres into and through the early years of school.

Vertical transitions – Transitions across time and between education levels (e.g., preschool, kindergarten and the first grade of school).

the growth of knowledge about children, learning, and achievement and adjustment. At the beginning of a new century, there is evidence of school entry and adjustment being conceptualized as longer-term and more complex processes. Traditional school-readiness concepts primarily focused on the individual child and normative expectations that were linked to child age. Contemporary constructions of transition have, to some extent, emerged as a response to the challenges teachers encounter as they work with the cohorts of diverse children entering a program, especially programs that use traditional age-based content. This article, therefore, primarily addresses the transitions of children from minority groups.

Since the 1990s, there has been increasing evidence of a re-framing of readiness for school toward the construct transition to school. Transition to school is thought to address more broadly the changes that occur as children move across early years' services and into school. Transitions can be broadly defined as border crossings that involve ongoing processes of mutual adaptation by children, families, communities, and schools (Ramey and Ramey, 1998). This article outlines some conceptualizations of transition to school that are apparent in early childhood education and care (ECEC) systems at this time including single time-limited change events, teaching practices, continuity of experience, multilayered and multifactorial processes, dynamic relationship-based processes, the engagement of children's resources, and the enhancement of resilience (Petriwskyj *et al.*, 2005; Dunlop and Fabian, 2007). The role of school readiness and preparedness is also discussed because readiness and preparedness constructs continue to have relevance in some systems, particularly in the US, Asia, and Australia.

Readiness and Preparedness

Attempts by professionals to ameliorate the perceived disadvantage of children from minority groups have focused on program content that is designed to optimize children's academic achievement and adjustment (Sammons *et al.*, 2004). For example, pre-academic programs, English-language instruction, and health or social-development programs have been used to enhance children's well-being and assist children's preparedness for formal schooling. Children with disabilities who have attended segregated early-intervention programs often receive specific instruction

Concern for the difficulties faced in the early years of school by diverse children, especially children from economic, social, and cultural minorities, and children with disabilities or learning difficulties is not new. Over several decades, a focus on individual children's readiness for school has altered in light of changing contexts and

in social and academic skills for school entry and social inclusion (Hanson *et al.*, 2000).

Child-readiness approaches emphasize a child's social maturation and content knowledge. An orientation toward child readiness is a feature of education systems within the United States and Australia. When political concern about the educational achievement of minority groups is addressed by focusing on child readiness, then children themselves become the primary targets for change (Burchinal *et al.*, 2002; Petriwskyj *et al.*, 2005). Support for readiness approaches is also apparent among Asian countries, and communities with Asian roots, where cultural expectations of a child's academic success may be high (Sy, 2003). The persistence of child-focused-readiness understandings among teachers, parents, communities, and policymakers can occur where there are also perceptions about school systems being inflexible, beliefs that each chronological age level requires the presence of uniform child-development outcomes and where there are understandings of preschool and kindergarten programs as getting children ready for school (Gill *et al.*, 2006). However, incompatibilities arise as readiness constructs, inclusive educational policies, and the use of selected, normative readiness assessments of young children coexist in an education system. In response, and sometimes to overcome problems related to individual assessments of risk or deficit, population-level assessments of child development may also be used to determine how best to allocate additional support resources for children in the early years. For example, in Canada, the early development instrument (EDI) and in Australia the Australian early development index (AEDI) are used to characterize communities that may benefit from enhanced school-readiness efforts.

There is evidence that some readiness-linked practices have generated equivocal results, including evidence that the practice of grade-retention harms child progress and self-esteem (Carlton and Winsler, 1999; Graue *et al.*, 2002). Increasing the age of school entry has not changed perceptions of advantage to older over younger members of a cohort, and delayed school entry has been challenged on the grounds that late school commencement may further disadvantage children in unfavorable home circumstances.

In the context of inclusive policies, there is pressure for schools to adapt to the diverse learning requirements of children rather than assume a set of prespecified academic and social skills that is required for young children at school entry. A shift in ECEC away from age-related approaches based on normative expectations of young children has prompted a reframing of readiness as preparedness (Dockett and Perry, 2007). Contemporary preparedness is interactionist in nature, context specific, and focused on the preparation of schools and families as well as children (Graue *et al.*, 2002). Both prior-to-school and school programs require attention to ensure preparedness. Shifting from a singular focus on the child to a comprehensive

analysis of the educational environment highlights the range of issues that affect transitions, children's subsequent learning, and their adjustment in new settings.

Transition as a Single Time-Limited Change Event

Transition to school can be construed as vertical steps: steps over time between different programs or program levels, for example, moving from child-care, through pre-school, and kindergarten class to the first primary-school grade. The step into a school-based kindergarten class from home or prior ECEC service is a key point of interest at system level, particularly with regard to children with perceived environmental or developmental disadvantage. Education systems vary in setting the age at which children must enter the statutory school education. A child's transition from informal preschool contexts to a formal kindergarten or primary school class can vary from age 3 to age 6 or 7 (Fabian, 2002; OECD, 2006). Transitions from a school-based kindergarten class into the first primary school grade may receive less attention, despite changes in expectations (Entwisle and Alexander, 1998), and further transitions throughout the early school years have only recently been addressed.

The construction of transition to school as a time-limited change event for children is consistent with an emphasis on initial adjustment to the school context, and to specific teaching practices that could improve adjustment. Some approaches to reducing the shock of this change event have reframed the prior preschool and kindergarten programs to be more school like. Other approaches have focused on supporting children's self-confidence, or have changed the first primary school year to achieve a more graduated change (Brostrom, 2002).

Short transition timeframes of a few days or weeks are evident in transition approaches that invest heavily in preparedness, single change events such as kindergarten entry and initial child-adjustment indicators. Longer transition timeframes are associated with practices that involve more sustained and complex processes surrounding learning continuity. These broader constructions of transition to school involve children, families, and schools and the processes shift from a primary attention to children's maturation and skills toward more complex interweaving of different facets of transition, including the structure and culture of family and school (Brooker, 2002).

Transition as Set School Practices

Concern for children's adjustment to school has resulted in a number of systems making structural changes that

include raising the age of school commencement and introducing transition classes in order to enhance school-entrant homogeneity. Accelerated school entry may also be justified in cases where commencement issues are substantial, such as for gifted children who encounter low-level curriculum relative to their skills. A focus on adjustment has given rise to various induction practices including prior-to-entry school visits by children or visits by school teachers to nearby preschools in order to familiarize the children and families with the local school context (Hanson *et al.*, 2000; Pianta and Kraft-Sayre, 2003).

Greater responsibility for induction is normally taken by the sending setting (e.g., kindergarten) than by the receiving setting (e.g., year 1 class). School-induction emphases may include introducing children and families to their local school, transferring information about children between teachers, and orienting children to the physical facilities and expectations of the new context (Dockett and Perry, 2007). These induction practices have been valuable in facilitating children's adjustment by enhancing confidence in the new environment, especially in situations where individual child support decreases at the same time as the social and academic demands increase.

Contemporary school-transition practices have incorporated wider communication linkages between families, schools, and ECEC services together with broader processes for supporting children as they move into kindergarten and primary school. Family-school interviews, family-induction meetings, and parent-orientation visits to school are also components of introductory practices (Brostrom, 2002). The unequal family-school power relationships implicit in some practices deserve careful consideration if diverse family circumstances are to be addressed. Sharing of information between teachers through meetings, transfer of children's assessment records, and other communication about individuals and curriculum are less frequent than parent-child orientation programs; yet they are valuable in the achievement of seamless transitions (La Paro *et al.*, 2000).

Transition as Continuity of Experience

Continuities in children's experiences as they move across ECEC and into school programs help children feel confident and secure. Although some level of dissonance can stimulate or positively challenge children, the negative impact of extreme discontinuity can threaten their progress in educational programs, especially children from social, economic, and cultural minorities (Dunlop, 2007; Podmore *et al.*, 2003). For example, the gap between ECEC and traditional school curricula and pedagogies can be significant. Efforts toward better curricula alignment continue to be made in many systems. These may involve changes to both ECEC and school curricula

(Brostrom, 2002; Neuman, 2001). There is awareness of the importance of pedagogy, especially achieving coherence of interaction style and language usage within and between ECEC and school programs. A focus on pedagogical continuity enables smooth transitions between the play pedagogies of ECEC and the more structured pedagogies of schools.

Another aim of systems that are sensitive to continuity is ensuring cultural coherence and inclusion. It is important to have in place program interactions that acknowledge diverse family-interaction patterns while also facilitating the development of common usage for learning in school settings. Ensuring cultural coherence across institutions and a continuity of experience for families and children are particular priorities in education systems working to facilitate the success of minority groups. Cultural variations and different expectations set within specialist and general-education services (e.g., early intervention program and kindergarten), can increase the challenge confronting children making transitions across the early years (Brooker, 2002). In many systems, there is a movement toward partnership with families, respecting continuity in learning home languages, and incorporating valued experiences of specific communities in the institutional settings (e.g., indigenous communities).

The need for increased overall system coherence has been raised in response to the lack of continuity of processes, policies, expectations, and quality between various ECEC and school systems. Where early childhood programs emerge from a range of agencies under different jurisdictions and with different mandates, discontinuities also emerge that may be confusing for families and children (Neuman, 2001). In the US, the full-service school or a school with a central community role has been suggested as a potential solution to this fragmentation. Approaches can also be made to integrate systems within a limited geographic area as a means of smoothing transitions although such approaches may assume family stability that is not warranted in times of increasing family geographic mobility (Fabian, 2002). In some European countries, more coordinated and systematically planned services and programs have been developed to ensure higher levels of system coherence for families and children (Neuman, 2001; OECD, 2006).

Transition as Multilayer Multiyear Process

Multilayered approaches to transition are grounded in a longer-term view of the transition period as a multiyear process. Recent literature frames transition into school as an extended process, ranging from 6 months to 3 years or an ongoing series of border crossings (Dunlop and Fabian, 2007; Raban and Ure, 2000). This longer frame of reference may have its foundation in the US experience of the

difference between short-term impacts on initial school adjustment and achievement and longer-term effects on developmental trajectories and life outcomes in programs such as Headstart (La Paro *et al.*, 2000; Peisner-Feinberg *et al.*, 1999).

The framing of school transition as a multiyear experience has occurred concurrently with the conceptualization of transition as a multifaceted process engaging a range of stakeholders (Graue *et al.*, 2002; Yeboah, 2003). The theoretical models of transition developed by Ramey and Ramey (1998), Fabian (2002), and Pianta and Kraft-Sayre (2003) share an ecological base that considers the interactions of factors in the child and family, the community, the school, and ECEC services. Conceptualizations of transition as both multiyear and multifaceted are also evident in recent investigations of the developmental trajectories of high-risk groups (Burchinal *et al.*, 2002). Such conceptualizations go well beyond earlier readiness issues toward a variety of ameliorating effects on educational disadvantage. Strategies include the provision of high-quality ECEC programs prior to school entry, high-quality educational programs in schools (Sammons *et al.*, 2008), and effective transition processes at a range of transition points.

Young children entering school undertake not only vertical transitions (from class to class) but also horizontal transitions, that is, across time within a class on any day (e.g., moving from general class lessons to specialist lessons such as music) or week (e.g., attending a regular kindergarten class part-time, a specialized class part-time, and an outside-school-hours care program). Horizontal transitions can pose challenges for children as they are added to other transitions in children's family lives and on vertical transitions (e.g., kindergarten into first grade) (Neuman, 2001). Children who are frequently withdrawn from their home classroom for learning support, additional language classes, extension programs, or behavior guidance may also spend undue time moving from place to place, thereby lowering their engagement in general classroom learning. Further, horizontal transitions between school and outside-school-hours care programs may involve substantial changes in expectations, location, relationships, and resourcing (Dockett and Perry, 2007).

Multiple transitions can cause stress by adding to the complexity of lives and increasing the chances of confusion (Neuman, 2001). Uncertainty about expectations in different settings has the potential to reduce the effectiveness of support programs and increase the likelihood of behavior challenges. Research that addresses the impact of multiple transitions is needed in order to guide policy development and educational practices. Teachers in the early years need to consider children's responses to multiple, overlapping transitions when planning curriculum and daily-program schedules.

Transition as a Dynamic Relationship-Based Process

Within both ECEC and inclusive education, dynamic, relationships-based transition processes have emerged, particularly in Europe (Niesel and Griebel, 2007). A key contribution of relationships-based approaches is their capacity to support young children and families who face multiple changes in their lives. The relationships considered include peer interactions, teacher–child relationships, family–school communication, linkages between teachers, and community involvement in the school. Within-school relationships have been the focus although wider support is now considered important (Dockett and Perry, 2007).

Studies of children's adjustment to school in Australia, New Zealand, and Europe have emphasized social and emotional issues, either because of the research emphasis on children from minority backgrounds with limited awareness of the rules and culture of schools or because of a perceived need to address academic readiness pressures from schools. There is some evidence indicating that teachers construct any atypical behavior of children who have delayed development or are from socially disadvantaged backgrounds as problematic (Skinner *et al.*, 1998). The contribution of social relations in the classroom and the playground, including both child–child and teacher–child relationships, appears to be vital not only to a child's social–emotional adjustment but also to academic achievement and self-esteem (Smith, 2002; Van den Oord and Rossem, 2002).

The continuity of children's friendships has been highlighted as important to a sense of security at school entry, meriting attention in class formation (Corsaro *et al.*, 2003). However, there is evidence that friendships alone do not necessarily make transition easier. Broader relationships issues require attention (Ledger *et al.*, 2000).

Greater involvement of the community in the life of schools is a feature of more extensive relationships-based transition processes. Community supports and connections that enhance bridging social capital improve the likelihood of successful transition to school (Dockett and Perry, 2007; Niesel and Griebel, 2007). Communities high in social capital may buffer and support children during transitions, through information networks, supportive webs of relationships, and reduction of social isolation for families facing challenges. Transition approaches that emphasize system-level coherence and integration of services reflect a democratic political view of transition as a community responsibility. This conception is powerful in government systems that emphasize more integrated national services, such as those in Scandinavian countries (OECD, 2006). In federated countries such as Australia and the US where ECEC and school policy and provision is segmented into separate federal, state, and local departmental jurisdictions (e.g., social welfare, education, and health), the community engagement challenge may be greater.

Child and Family Resource and Resilience

Reconceptualization of school transition has given rise to changes in the way successful transitions are determined and stakeholders are included. Successful school entry that was dependent on readiness as a maturational characteristic of the individual child, or a series of pre-academic skills, differs from more recent constructions of transition. Success is no longer defined solely as individual social and emotional adjustment and academic achievement against age-based curricula. Making successful transitions encompasses broader educational and personal purpose, taking into account the different participant groups (parents, teachers, and assistants) and learner diversity.

With greater recognition of differences in children's initial adjustment success at school entry, problems such as medium-term fade-out in achievement by some children over time, and the evidence of long-term improvement in broad life outcomes from effective ECEC programs, there is greater attention to the complexity of transitions.

It seems clear that successful transitions are important to ongoing progress, lifelong learning, and meaningful participation. There is an interest at this time in a new century to set more positive developmental trajectories for all children. Some shift in images of children from vulnerable toward competent has been accompanied by shifts in transition approaches from those emphasizing deficit and needs to those focused on personal and cultural resources and resilience (Dunlop, 2007). It is seen as important to capitalize on the resources children bring to the school in the transition process, ensure effective communication and collaboration with families and communities, and enact an outward-looking approach in both schools and communities.

Conclusion

Expectations of homogeneity in school entrants are giving way to recognition of the diversity in young children, families, and communities. Contemporary constructions of transition recognize the potential for child and family diversity to enrich teaching and learning. There is a focus on inclusive policies and a recognition that homogeneous classes of children ready for school entry denies the reality of young children's developmental, cultural, and linguistic diversity. A focus on single-transition issues such as teacher practices or time-limited change events has given way to more complex, multifaceted views incorporating a variety of transition points and a range of stakeholders. Studies of successful transitions now take into account long-term trajectories as well as initial adjustment to the school context. Alongside expanded conceptions of transition have come questions related to the coherence of curriculum, pedagogy and service

systems, and of authentic partnerships between families, communities, schools, and ECEC services.

There is also a persistence of attention to child-level readiness and preparedness at a policy, school, and community level. Teacher beliefs, public perceptions, and policy formation are resistant to change, and provision of universally accessible, high-quality ECEC programs confronts the continuing challenge of mechanisms that allow all children, including diverse minority groups, to succeed.

See also: Literacies in Early Childhood: The Preschool Period; Participation in Early Childhood Education and Care Programs: Equity, Diversity and Educational Disadvantage.

Bibliography

- Brooker, L. (2002). *Starting School: Young Children Learning Cultures*. Buckingham: Open University Press.
- Brostrom, S. (2002). Communication and continuity in the transition from kindergarten to school. In Fabian, H. and Dunlop, A. -W. (eds.) *Transitions in the Early Years: Debating Progression and Continuity for Children in Early Education*, pp 52–63. London: Routledge-Falmer.
- Burchinal, M., Peisner-Feinburg, E., Pianta, R., and Howes, C. (2002). Development of academic skills from preschool through second grade: Family and classroom predictors of developmental trajectories. *Journal of School Psychology* 40(5), 415–436.
- Carlton, M. and Winsler, A. (1999). School readiness: The need for a paradigm shift. *School Psychology Review* 28(3), 338–352.
- Corsaro, W., Molinari, L., Hadley, K., and Sugioka, H. (2003). Keeping and making friends: Italian children's transition from preschool to elementary school. *Social Psychology Quarterly* 66(3), 272–292.
- Dockett, S. and Perry, B. (2007). *Transition to School: Perceptions, Expectations, Experiences*. Sydney: University of New South Wales Press.
- Dunlop, A-W. (2007). Bridging research, policy and practice. In Dunlop, A. -W. and Fabian, H. (eds.) *Informing Transitions in the Early Years: Research, Policy and Practice*, pp 151–168. Maidenhead: Open University Press.
- Dunlop, A. -W. and Fabian, H. (2007). *Informing Transitions in the Early Years: Research, Policy and Practice*. Maidenhead: Open University press/McGraw Hill.
- Entwisle, D. and Alexander, K. (1998). Facilitating the transition to first grade: The nature of transition and research on factors affecting it. *Elementary School Journal* 98(4), 351–364.
- Fabian, H. (2002). *Children Starting School*. London: David Fulton.
- Gill, S., Winters, T., and Friedman, D. (2006). Educators' views on pre-kindergarten and kindergarten readiness and transition processes. *Contemporary Issues in Early Childhood* 7(3), 213–226.
- Graue, E., Kroeger, J., and Brown, C. (2002). Living the 'gift of time'. *Contemporary Issues in Early Childhood* 3(3), 338–353.
- Hanson, M., Beckman, P., Horn, E., et al. (2000). Entering preschool: Family and professional experiences in this transition process. *Journal of Early Intervention* 23(4), 279–293.
- La Paro, K., Pianta, R., and Cox, M. (2000). Kindergarten teachers' reported use of kindergarten to first grade transition practices. *Elementary School Journal* 101(1), 63–78.
- Ledger, E., Smith, A., and Rich, P. (2000). Friendships over the transition from early childhood centre into school. *International Journal of Early Years Education* 8(1), 57.
- Neuman, M. (2001). Hand in hand: Improving the links between ECEC and schools in OECD countries. In *Early Childhood Education and Care: International Perspectives. The Report of a Consultative Meeting*, New York, May 11–12, 2000.
- Niesel, R. and Griebel, W. (2007). Enhancing the competence of the transition system through co-construction. In Dunlop, A. -W. and

- Fabian, H. (eds.) *Informing Transitions in the Early Years: Research, Policy and Practice*, pp 21–32. Maidenhead: Open University Press/McGraw Hill.
- OECD (Organization for Economic Cooperation and Development) (2006). *Starting Strong 11. Early Childhood Education and Care*. Paris: OECD.
- Peisner-Feinberg, E., Burchinal, M., Clifford, R., et al. (1999). *The Children of the Cost, Quality and Outcomes Study go to School. Technical Report*. Chapel Hill, NC: University of North Carolina.
- Petriwskyj, A., Thorpe, K., and Tayler, C. (2005). Trends in construction of transition to school in three western regions 1990–2004. *International Journal of Early Years Education* **13**(1), 55–69.
- Pianta, R. and Kraft-Sayre, M. (2003). *Successful Kindergarten Transition*. Baltimore, MD: Paul H. Brookes.
- Podmore, V., Sauvaio, L., and Mapa, L. (2003). Sociocultural perspectives on transition to school from Pacific Islands early childhood centres. *International Journal of Early Years Education* **11**(1), 33–42.
- Raban, B. and Ure, C. (2000). Continuity for socially disadvantaged school entrants: Perspectives of parents and teachers. *Australian Research in Early Childhood Education* **7**(1), 54–65.
- Ramey, S. and Ramey, C. (1998). The transition to school: Opportunities and challenges for children, families, educators and communities. *Elementary School Journal* **98**(4), 293–295.
- Sammons, P., Sylva, K., Melhuish, E., et al. (2004). *The Effective Provision of Pre-School Education (EPPE) Project Technical Paper 11-Report on the Continuing Effects of Preschool Education at Age 7*. London: Institute of Education, University of London.
- Sammons, P., Sylva, K., Melhuish, E., et al. (2008). *The Influence of School and Teaching Quality on Children's Progress in Primary School*. London: Institute of Education, University of London.
- Skinner, D., Bryant, D., Coffman, J., and Campbell, F. (1998). Creating risk and promise: Children's and teachers' co-constructions in the cultural world of kindergarten. *Elementary School Journal* **98**(4), 297–310.
- Smith, N. (2002). Transition to the school playground: an intervention programme for nursery children. *Early Years* **22**(2), 129–145.
- Sy, S. (2003). Parent involvement and children's transition to school in Asian American and European American families. *Dissertation Abstracts International* **63**(7-B), 3499.
- Van den Oord, E. and Rossem, R. (2002). Differences in first graders' school adjustment: The role of classroom characteristics and social structure of the group. *Journal of School Psychology* **40**(5), 371–394.
- Yeboah, D. (2003). Enhancing transition from early childhood phase to primary education: Evidence from the research literature. *Early Years* **22**(1), 51–67.

Further Reading

- Thorpe, K., Tayler, C., Bridgstock, R., et al. (2004). *Preparing for School: Report of the Queensland Preparing for School Trials 2003/2004*. Brisbane: Queensland University of Technology.
- Vogler, P., Crivello, G., and Woodhead, M. (2008). *Early Childhood Transitions Research: A Review of Concepts, Theory and Practice*. The Hague: Bernard van Leer Foundation.

Parent Support in Early Childhood – Approaches and Outcomes

K S Howard, Marymount Manhattan College, New York, NY, USA

K A Beckmann and J Brooks-Gunn, Columbia University, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Effect size – A number representing the magnitude of an effect in standard deviation units. One of the most common statistics for reporting effect size is known as Cohen's *d*; scores from 0 to 0.3 are considered small effects, 0.4 to 0.6 moderate, and 0.7 to 1.0 large.

Early childhood is a time of special vulnerability and opportunity. Because young children are particularly sensitive to environmental inputs, characteristics of early experiences, including the quality of parent–child interactions and home environments, potentially may have effects that last well into adulthood. Due to the importance of the early years and the fundamental role that parents play during this time, an abundance of research has focused on highlighting the specific ways in which parents influence their children. As a result, the notion that children's outcomes can be improved by altering parenting practices has become the fundamental belief at the core of parent-education programs.

Although parenting programs share a common goal of improving parental capacity in order to positively affect children, programs greatly differ with regard to their methods, staff, and the domains in which outcomes have been observed. Some are home based, while others consist of group meetings, child care, or combinations thereof. Qualifications of program staff range from paraprofessionals to public health nurses or social workers. Over the last several decades, as parenting programs have been implemented around the world, different patterns of results have been observed depending on the characteristics of the program and families served.

We now turn to a review that represents the wide variation in parenting programs, with international perspectives on parent training and supports for low-income families. The review has been organized to offer comparisons of programs around the world with similar programs in the United States. In order to focus the discussion on program effectiveness, we only consider outcomes within three domains: maternal parenting (warmth, sensitivity, and reduced harshness), child health, and children's

cognitive and social development. Whenever possible, we report Cohen's *d* to indicate effect sizes for significant findings.

Early Start and Healthy Families

Early Start, New Zealand

Early Start is a parenting support program in Christchurch, New Zealand that is part of the larger network of family support programs throughout the country. The program provides home-based services to families of infants and young children under age 5, with enrolment just after birth. Based on the Healthy Families America (HFA) model, those who are at the highest risk for poor parenting receive the most home visits, averaging an hour per month for up to 2 years. Services provided by trained family support workers (FSWs) balance the limitations and strengths of each family. FSWs who conduct the home visits have degrees in nursing or social work, and receive additional program-specific training. The program aims to build positive relationships, problem-solving skills, and provide support and advice to families throughout the preschool years (Fergusson *et al.*, 2005).

The randomized trial of Early Start included 443 families who were evaluated at 6, 12, 24, and 36 months after program enrolment. A number of significant effects were observed with respect to maternal parenting and children's health and development, with small to moderate effect sizes ranging from 0.03 to 0.31 (Fergusson *et al.*, 2005). Early Start families reported more positive and nonpunitive maternal parenting attitudes relative to control families. Children in treatment families were more likely to have had regular doctor and dental visits and had fewer accidental injuries than children in the control group. Early Start children also exhibited fewer behavior problems, particularly in the area of internalizing (Fergusson *et al.*, 2005).

Healthy Families America

After being initially developed in Hawaii, the HFA program has expanded throughout the US and Canada with help from Prevent Child Abuse America. The primary goal of HFA is to prevent child abuse and neglect by providing home-based supports to mothers of young

children. Trained paraprofessionals provide home visits to families who are considered at risk for abuse or neglect until the child is 5 years old or begins preschool.

Despite its popularity and fairly widespread implementation, randomized trials of HFA have identified few benefits associated with home-visiting services. In the initial evaluations of Hawaii Healthy Start, no significant program effects were found for either parenting behaviors or children's development. However, the evaluation of HFA in New York State has shown relatively small effects on mothers' use of harsh parenting. The observed effects were stronger for the most at-risk subgroups in the sample (DuMont *et al.*, 2008).

Early Start has roots in the HFA program. A key similarity is that families are screened for eligibility and offered services on the basis of family risk. Although both utilize home visiting as the method of service delivery, they differ in home-visitor qualifications. In HFA, paraprofessionals provide home visits, whereas in Early Start, visits are conducted by nurses or social workers. This probably accounts for the different pattern of results observed across programs.

Jamaican Home-Visiting Programs and Nurse–Family Partnership

Home Visiting, Jamaica

While many early education programs in developing countries contain parenting education components, few have been rigorously evaluated. In the 1970s, Powell and Grantham-McGregor began empirical tests of home-visiting programs for low-income families in urban areas of Jamaica. The programs aimed to improve early child development while teaching mothers more effective parenting skills; variations were based on program design (i.e., frequency of visits). Performed by paraprofessionals who were supervised by nurses at local health care centers, home visits typically focused on enhancing mother–child interactions and the self-esteem of both mothers and children. Home visitors modeled different ways for parents to engage their children through culturally relevant language, games, songs, and art activities (Powell and Grantham-McGregor, 1989).

Each separate home-visiting program was associated with beneficial effects for children's development. However, greater frequency of home visits was associated with the most positive outcomes. In one randomized study of 56 families who either received weekly home visits or were assigned to a control group, home-visited children displayed significantly higher scores on developmental tests 1 year after the intervention concluded. Despite the small sample, effect sizes were large in magnitude (0.52–1.09; Powell and Grantham-McGregor, 1989). Measures of maternal parenting were not examined, although parenting

was the hypothesized pathway by which children's development was impacted.

Nurse–Family Partnership, United States

The Jamaican home-visiting programs were not explicitly based on an American program, but they were similar to the Nurse–Family Partnership (NFP) – one of the most extensively tested home-visiting programs in the United States. The NFP employs public health nurses as home visitors in order to improve pregnancy outcomes, maternal economic self-sufficiency, and children's health and development by helping low-income parents provide sensitive and competent care. The first trial of NFP began in 1977 in Elmira, New York with a sample of primarily rural, white women. The Elmira study demonstrated long-term benefits of the intervention, but most effects were concentrated in the highest-risk subgroups (i.e., poor, single mothers). In particular, at the 15-year follow-up, nurse-visited families evidenced a 48% reduction in state-verified reports of child abuse and neglect. Although there were no other significant effects on maternal parenting or children's development, the 15-year follow-up also showed that adolescents of poor, single mothers who had received home visits had fewer arrests (0.20 vs. 0.45; $p = 0.03$) and fewer convictions (0.09 vs. 0.47; $p < 0.001$) than their counterparts in the control group (Olds *et al.*, 1999).

Two replications of NFP have been evaluated in Memphis and Denver. The Memphis sample included higher proportions of urban, African-American parents, whereas the Denver sample included more adult mothers. The pattern of results has largely been inconsistent across the three trials. Although both the Elmira and Memphis trials found evidence for fewer hospital visits for child injuries or ingestions (Olds *et al.*, 1999), this effect was not replicated in Denver. However, Denver was the only site that identified a benefit in terms of children's language skills ($d = 0.31$ at age 4; Olds *et al.*, 2004). A vast majority of program effects for NFP were isolated to the most vulnerable families within each site.

Since the use of public health nurses as home visitors is a central component of the NFP program, the relative effectiveness of nurse home visitors and paraprofessionals was examined in the Denver study. Although families who received home visits from paraprofessionals, in general, fared better than controls, nurse home visitors were associated with the most benefits (Olds *et al.*, 2002). Similarly, the Jamaican studies showed that nurse home visitors were associated with large benefits for children in the treatment group, relative to controls (average gain of 13 points on a developmental index). Since the effects with nurses were so large, subsequent programs employed paraprofessionals – resulting in smaller, but still significant, effects and a more cost-effective program that could be scaled to reach more families (Grantham-McGregor and Desai, 1975).

Sure Start and Early Head Start

Sure Start, United Kingdom

In 1999, the UK government initiated the Sure Start program as an early intervention strategy aimed at reducing child poverty and social exclusion (to address both financial hardship and access to civil, political, and social rights; Aber *et al.*, 2002). In order to limit social stigma associated with targeting less-advantaged individuals within communities, Sure Start was provided to all families with children under age 4, who lived in selected disadvantaged districts (Rutter, 2006).

Within the broad family of programs that comprised Sure Start, the government mandated provision of community-based supports for participating families and children, including outreach, home visiting, and quality child care and healthcare, with an emphasis on providing supports for families of children with special needs. Each Sure Start Learning Program (SSLP) had the autonomy to work with surrounding communities to improve existing services and create new initiatives where gaps existed. As such, SSLPs initially did not have prescribed curricula or standard services; individual programs greatly varied in implementation, content, and quality (Rutter, 2006).

The first phase of the National Evaluation of Sure Start was based on a nonrandomized, cross-sectional design in which 150 of 260 SSLP areas were randomly selected and compared to 50 areas waiting to be designated as SSLP. After less than 3 years of program participation, study families were randomly selected based on whether children were 9 or 36 months of age resulting in approximately 14 000 9-month-old and 5000 36-month-old children and their families. The evaluation focused on the effects of SSLPs on community service utilization, family functioning, child development outcomes, and program effectiveness (Belsky *et al.*, 2006).

The results showed small or nonsignificant effects of SSLPs on supportive parenting and child behavior. Among families of 9-month-olds, those in SSLPs displayed less household chaos than controls ($d = 0.14$). However, in some cases, Sure Start had adverse effects on the most-disadvantaged 3-year-old children and their families (Rutter, 2006). For example, children from unemployed and single-parent SSLP households scored lower on verbal ability than their non-SSLP peers (Belsky *et al.*, 2006).

From 2004 to 2006, SSLPs evolved into Sure Start Children's Centers, signaling a clearer set of program goals and more systematic services for families. With the maturation of Sure Start programs, a second phase of the National Evaluation was undertaken with a quasi-experimental comparison of 5883 families residing in 138 Sure Start areas to 1879 participants from the Millennium Cohort Study who resided in demographically similar, non-Sure-Start areas (Melhuish *et al.*, 2008). The results of the second phase of evaluation were more favorable than the first.

After 3 years of program exposure, children demonstrated more positive social behavior ($d = 0.19$) and greater self-regulation ($d = 0.17$) than nonprogram children. Parents in Sure Start areas also exhibited fewer negative parenting behaviors ($d = 0.44$) and provided more stimulating home-learning environments than comparison parents ($d = 0.27$). Further, the effects on parenting partially mediated the effects of Sure Start on children's social behavior (Melhuish *et al.*, 2008).

Early Head Start

Implemented 4 years prior to Sure Start in 1995, Early Head Start (EHS) was created by the US federal government to serve pregnant women and children under age 3 in disadvantaged populations (initiated in 1965, Head Start serves 3- and 4-year-olds and their families). In order to best tailor services, Early Head Start performance standards provide four program options—home based, center based, a combination of the two, or locally designed alternatives. The home-based programming includes weekly home visits and biweekly group meetings, while the center-based model incorporates center-based care and education with at least two annual home visits (Love *et al.*, 2005).

More specific than Sure Start's core components, the EHS performance standards mandate that programs explore language and culture; encourage parent involvement and positive consistent relationships with children; and promote healthy development along with early detection and prevention of developmental concerns. EHS is required to serve children with disabilities; facilitate smooth transitions into Head Start; and collaborate with other providers in order to enhance services (Love *et al.*, 2005).

With evaluation built into the federal program since its conception, a randomized control trial began in 1996 with a diverse sample of 3001 low-income families in 17 programs from urban, suburban, or rural communities. Overall, the study demonstrated positive, but small, effects of EHS participation for parenting and children's cognitive, language, social, and emotional development at 3 years. Parents who received EHS services were more emotionally supportive than control parents, demonstrated greater warmth, and provided more language and learning stimulation ($d = 0.15$, 0.09 , and 0.10). While EHS had smaller impacts on hostility and insensitivity, parents who participated in the program were less likely to use physical punishment as a means of discipline ($d = -0.14$). EHS children performed significantly better than controls in both cognitive and language development ($d = 0.10$ – 0.13). The children also scored higher on emotional engagement and sustained attention and lower in aggressive behavior ($d = 0.20$, 0.16 , and -0.11). The greatest impacts were in programs that offered a mix of home-visiting and center-based services that fully implemented EHS performance standards (Love *et al.*, 2005).

Despite being based on EHS and Head Start, the results of Sure Start were not as positive as those seen in the US. The main reason for the differences is likely due to the greater specificity of EHS services. Since Sure Start was so flexible across areas, different patterns of success likely emerged in different regions. Differences in treatment may have only been exacerbated by not having a randomized comparison group for examining program effectiveness, as was the case with EHS. However, results from smaller studies of Sure Start within particular areas, which used a set curriculum, yielded more positive results. We now turn to a discussion of the Incredible Years, and the results associated with it in Sure Start areas in Wales.

The Incredible Years

The Incredible Years is a prevention-oriented intervention that has received extensive empirical support for changing parenting practices that lead to reductions in child behavior problems. The program trains parents in positive communication and child-directed play skills, consistent and clear limit setting, and nonviolent discipline strategies. The Incredible Years promotes parental self-care, social and emotional child development, and secure attachment, while educating parents on how to teach children problem-solving skills and anger management (Barth *et al.*, 2005). The curriculum for Incredible Years parent programs is based on child age: babies and toddlers (0–3), Basic early childhood (3–6), Basic school age (7–12), and Advance (6–12). The Advance curriculum is an adaptation of the BASIC school-age program that has a special emphasis on problem solving, anger management, and support systems. Employed in different populations within New Zealand, Canada, Norway, the United States, and the United Kingdom, modified Incredible Years videos are available that are diverse in terms of ethnicity and language (Webster-Stratton, 1994).

The Incredible Years, Wales

In Wales, a randomized trial was conducted based on 11 programs that implemented the Incredible Years Basic within the Sure Start framework. In contrast to the results of the first phase of the National Sure Start Evaluation, this study showed benefits for Sure Start families relative to controls in terms of both positive parenting practices and fewer behavior problems for children (Hutchings *et al.*, 2007), suggesting the benefits of utilizing a systematic, evidence-based curriculum. Specifically, the evaluation demonstrated significant improvements in positive parenting ($d = 0.57$), decreases in child antisocial behavior and hyperactivity ($d = 0.30$ and 0.41), and increases in child self-control ($d = 0.38$) for those who had participated in Incredible Years (Hutchings *et al.*, 2007).

The Incredible Years, United States

Providing additional support for The Incredible Years model as an approach to altering parenting practices, randomized controlled trials with Head Start populations in the United States have demonstrated that harsh discipline and criticism were reduced ($d = -0.27$ and -0.18) and positive parenting behaviors were enhanced ($d = 0.17$) after program completion (Reid *et al.*, 2001). Parents who participated in The Incredible Years Advance Program showed increases in marital problem solving and their children improved in social problem solving as compared to controls and to those who received the Incredible Years Basic, with most effects retained for at least 1 year (Gross *et al.*, 2003).

Home Instruction Program for Preschool Youngsters

Founded on the belief that parents are children's first teachers, Home Instruction Program for Preschool Youngsters (HIPPY) was created at Hebrew University in the late 1960s. Over the past decades, HIPPY has spread to 12 countries, serving more than 22 000 families worldwide, and has built an international support network providing program-related services, including curriculum development and coordination of research efforts. HIPPY parenting education follows a coordinated approach employing local coordinators to train paraprofessionals who, in turn, use role-playing methods to educate parents (Westheimer, 2003). HIPPY strives to enhance home literacy environments, parent-child communication, and parental ability to support early learning, using a 2-year structured curriculum targeted to 4- and 5-year-olds. There are two core elements of the program. First, home visits by paraprofessionals and group meetings with parents, paraprofessionals, and local HIPPY coordinators occur on alternate weeks. These sessions include role-playing, sharing questions and concerns, meeting other parents, and enrichment activities such as presentations by school officials. Second, HIPPY includes 30 weeks of educational activities for parents to complete with their children. Age-appropriate materials are designed to support language, sensory, and perceptual development, as well as the visual-motor coordination and problem-solving skills that children need to succeed in school (Baker *et al.*, 1999).

Introduced in the United States in 1984, HIPPY-USA has program sites in 25 states and the District of Columbia. The primary US evaluation involved two cohorts of HIPPY and comparison families at sites in New York City and Arkansas. In New York, a diverse sample of 182 low-income families was recruited across both cohorts of the randomized control trial. As the Arkansas trial was quasi-experimental and replicated the New York results, we report only outcomes from the New York evaluation (Baker *et al.*, 1999). With effect sizes ranging from 0.56

to 0.76, children who participated in HIPPY performed significantly better on measures of cognitive skills in kindergarten, reading comprehension at the end of first grade, and classroom adjustment in first and second grades. However, these findings were not replicated in the second cohort and subsequent analyses could not account for the differences in results. Although data regarding parent participation are not available, it has been suggested that cohorts differed in the level and quality of parental engagement (Baker *et al.*, 1999).

Turkish Early Enrichment Program

In Turkey, the HIPPY curriculum has been incorporated into a larger program known as the Turkish Early Enrichment Program (TEEP). TEEP focuses on families with children ages 3–5 to promote cognitive and socioemotional development through maternal education. Paraprofessionals conduct home visits and group meetings in order to model strategies for active engagement with children. The program also seeks to cultivate human capital by building agency and communication skills in mothers (Kagitcibasi *et al.*, 2001).

The randomized longitudinal evaluation included 255 families who were studied over a period of 4 years and then followed up 6 years later. In the first year, baseline data were collected, then mothers were randomly assigned to treatment ($n = 90$) and control groups. In the second and third years, mothers in the treatment group participated in a 60-week, home-based training that used HIPPY materials and focused on maternal support and child cognition (Kagitcibasi *et al.*, 2001). In the fourth year, parent and child outcomes were evaluated.

Children whose mothers had received TEEP training had higher intelligence quotient (IQ) scores than children in the control group ($d = 0.55$). However, the picture regarding academic achievement and grades was more mixed. Treatment group children scored higher on tests of general ability ($d = 0.27$), but not on math or Turkish. The effects on grades were moderate among 5-year-olds ($d = 0.36$ – 0.47). Program mothers reported higher levels of self-efficacy because they were able to act as their children's teachers and interact more effectively with their children by being less punitive, more responsive, and more cognitively stimulating. Program effects were also observed at a 6-year follow up with superior academic performance among children in the treatment group ($d = 0.43$ – 0.47 ; Kagitcibasi *et al.*, 2001).

Parenting Support and Education: Lessons Learned

The picture that emerges when taking a broad view of parenting education and support programs is that the

effects are neither strong nor consistent across programs, domains, or even within programs. Some programs, based on similar models of intervention, find very different patterns of results. For example, while the NFP showed minimal effects on children's development, the Jamaican Home-Visiting programs showed large effects. However, the NFP results were strongest among the most high-risk families in the sample. Similarly, Jamaican families were at such high risk that even scaled-back versions of the intervention produced significant effects. In another case, the reason that Early Start in New Zealand showed more positive effects than HFA is likely due to the fact that despite utilizing the same program model, HFA employed paraprofessionals to deliver the intervention, whereas Early Start used nurses and social workers.

The programs with the fewest overall effects were those that lacked a strong theoretical model tied to the curriculum. Sure Start, with its flexibility to meet the needs of individual communities, produced very few results. Similarly, one of the reasons that Early Head Start has shown few effects is likely also due to its lack of a clear focus as the program was implemented around the country. The Incredible Years demonstration in Wales clearly showed that the Sure Start model had stronger effects when a specific curriculum was put in place than when no structured curriculum was utilized. Similarly, the theoretically based HIPPY curriculum has been associated with benefits for children and families in Israel, the United States, Turkey, and in several other countries. The differences in programs and their effectiveness can be translated to unique points of consideration for future research on parenting education and support programs.

The majority of the programs that were discussed in this review have been conducted on a large scale. One potential problem in offering services broadly is that relatively swift program expansion without adequate resources can hinder quality as well as impact. Bringing programs to large-scale implementation can pose challenges to maintaining fidelity and consistency in program quality, especially when a component of program effectiveness is flexibility. This was one of the primary problems of Sure Start – programs were rolled out across the country without a clear structure or plan. In contrast, the NFP programs have been slowly adding program sites in order to expand services, at the same time maintaining quality and strict adherence to the program model.

Another important consideration in applying research on parenting programs is understanding the cultural context in which the program was designed and will be applied. Research has suggested that although culturally adapted parenting programs may have positive effects, the effects may be reduced or observed in different domains than what would have been identified in the original program (Kumpfer *et al.*, 2002); caution is warranted when attempting to transfer intervention programs across cultures. It may be

that programs are not carried out in exactly the same manner in different contexts, or that they are experienced differently by families from different cultures. One of the reasons that programs such as HIPPY and the Incredible Years have been successful in different contexts is that adherence to their theoretical models was not compromised while cultural adaptations were made.

It is important to recognize that the research on parenting programs does not suggest a one-size-fits-all approach. Even so, tailoring programs to the needs of families within a specific community should not outweigh the importance of utilizing specific curriculum and targeting specific behaviors. The research on parenting programs to date shows that the most effective programs are those that target families with specific needs and use theoretically driven curriculum to enhance family and child functioning. Although most parenting programs have stronger effects on parenting practices than on children's development, it remains to be seen how the effects of parenting may continue to positively alter the child's experience into adolescence.

See also: Child Rearing and Early Education: Parents and Professionals: Theoretical and Historical Influences; Cost-effective Early Childhood Programs from Preschool to Third Grade; Home-Based and Institutional Early-Childhood Education and Care Services; Investing in Early Childhood Education and Care: The Economic Case.

Bibliography

- Aber, J. L., Gershoff, E. T., and Brooks-Gunn, J. (2002). Social exclusion of children in the United States: Identifying potential indicators. In Kahn, A. J. and Kamerman, S. B. (eds.) *Beyond Child Poverty: The Social Exclusion of Children*, pp 245–286. New York, NY: Columbia University.
- Baker, A., Piotrkowski, C., and Brooks-Gunn, J. (1999). The home instruction program for preschool youngsters (HIPPY). *Future of Children* **9**, 116–133.
- Barth, R., Landsverk, J., Chamberlain, P., et al. (2005). Parent-training programs in child welfare services: Planning for a more evidence-based approach to serving biological parents. *Research on Social Work Practice* **15**, 353–371.
- Belsky, J., Melhuish, E., Barnes, J., et al. (2006). Effects of Sure Start local programmes on children and families: Early findings from a quasi-experimental, cross-sectional study. *British Medical Journal* **332**, 1476–1480.
- DuMont, K., Mitchell-Herzfeld, S., Greene, R., et al. (2008). Healthy families New York (HFNY) randomized trial: Effects on early child abuse and neglect. *Child Abuse and Neglect* **32**, 295–315.
- Fergusson, D. M., Grant, H., Horwood, L. J., and Ridder, E. M. (2005). Randomized trial of the Early Start program of home visitation. *Pediatrics* **116**, 803–809.
- Grantham-McGregor, S. and Desai, P. (1975). A home-visiting intervention programme with Jamaican mothers and children. *Developmental Medicine and Child Neurology* **17**, 605–613.
- Gross, D., Fogg, L., Webster-Stratton, C., et al. (2003). Parent training with families of toddlers in day care in low-income urban communities. *Journal of Consulting and Clinical Psychology* **71**, 261–278.
- Hutchings, J., Bywater, T., Daley, D., et al. (2007). Parenting intervention in Sure Start services for children at risk of developing conduct disorder: Pragmatic randomised controlled trial. *British Medical Journal* **334**, 678–682.
- Kagitcibasi, C., Sunar, D., and Bekman, S. (2001). Long term effects of early intervention: Turkish low-income mothers and children. *Applied Developmental Psychology* **22**, 333–361.
- Kumpfer, K. L., Alvaro, R., Smith, P., and Bellamy, N. (2002). Cultural sensitivity and adaption in family-based prevention interventions. *Prevention Science* **3**, 241–246.
- Love, J. M., Kisker, E. E., Ross, C., et al. (2005). The effectiveness of Early Head Start for 3-year-old children and their parents: Lessons for policy and programs. *Developmental Psychology* **41**, 885–901.
- Melhuish, E., Belsky, J., Leyland, A., and Barnes, J. The National Evaluation of Sure Start Research Team (2008). The impact of Sure Start local programmes on 3-year-old children and their families living in England: A quasi-experimental observational study. *Lancet* **372**, 1641–1647.
- Olds, D. L., Henderson, C. R., Kitzman, H. J., et al. (1999). Prenatal and infancy home visitation by nurses: Recent findings. *Future of Children* **9**, 44–65.
- Olds, D. L., Robinson, J., O'Brien, R., et al. (2002). Home visiting by paraprofessionals and by nurses: A randomized, controlled trial. *Pediatrics* **110**(3), 486–496.
- Olds, D. L., Robinson, J., Pettitt, L., et al. (2004). Effects of home visits by paraprofessionals and by nurses: Age 4 follow-up results of a randomized trial. *Pediatrics* **114**, 1560–1568.
- Powell, C. and Grantham-McGregor, S. (1989). Home-visiting of varying frequency and child development. *Pediatrics* **84**, 157–164.
- Reid, M., Webster-Stratton, C., and Beauchaine, T. (2001). Parent training in Head Start: A comparison of program response among African American, Asian American, Caucasian, and Hispanic mothers. *Prevention Science* **2**, 209–227.
- Rutter, M. (2006). Is Sure Start an effective preventive intervention? *Child and Adolescent Mental Health* **11**, 135–141.
- Webster-Stratton, C. (1994). Advancing videotape parent training: A comparison study. *Journal of Consulting and Clinical Psychology* **62**, 583–593.
- Westheimer, M. (2003). Introduction: A decade of HIPPY research. In Westheimer, M. (ed.) *Parents Making a Difference: International Research on the Home Instruction for Parents of Preschool Youngsters (HIPPY) Program*, pp 19–45. Jerusalem: The Hebrew University Magnes Press.

Further Reading

- Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report, Society for Research in Child Development* **17**(1), 3–14.
- Brooks-Gunn, J. and Markman, L. (2005). The contribution of parenting to ethnic and racial gaps in school readiness. *Future of Children* **15** (1), 138–167.
- Gray, R. and Francis, E. (2007). The implications of US experiences with early childhood interventions for the UK Sure Start programme. *Child: Care, Health and Development* **33**, 655–663.
- Howard, K.S. and Brooks-Gunn, J. (in press). Home-visiting programs: The prevention of child abuse and neglect. *The Future of Children*.
- Sweet, M. A. and Appelbaum, M. I. (2004). Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. *Child Development* **75**, 1435–1456.

Relevant Websites

- <http://www.earlystart.co.nz> – Early Start.
- <http://www.healthyfamiliesamerica.org> – Healthy Families America.
- <http://www.acf.hhs.gov> – Head Start/Early Head Start.
- <http://www.hippy.org.il> – Home Intervention Program for Preschool Youngsters (HIPPY).
- www.hippyusa.org – HIPPY USA.
- <http://www.incredibleyears.com> – Incredible Years.
- <http://www.nursefamilypartnership.org> – Nurse–Family Partnership.
- <http://www.surestart.gov.uk> – Sure Start.

EARLY CHILDHOOD EDUCATION AND CARE RESEARCH

Contents

Development and Implementation of Early Learning Standards in the United States

Global Status of Early Learning and Development Standards

Rethinking Early Childhood Education and Care: Implications for Research and Evaluation

Development and Implementation of Early Learning Standards in the United States

C Scott-Little, University of North Carolina at Greensboro, Greensboro, NC, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Early-learning guidelines – The documents that describe standards for what children should know and be able to do before they enter grade school/ kindergarten. Most commonly refer to standards for very young children (infants and toddlers), see also Early learning standards.

Early-learning standards – The documents that describe standards for what children should know and be able to do before they enter grade school/ kindergarten. These are sometimes called early-learning guidelines.

Early Learning Standards (ELS) – documents that define expectations for what children from birth through to age 5 should be know and be able to do – are increasingly common in the United States. Since the beginning of this century, every state in the country as well as the District of Columbia has developed ELS for preschool-age children. A majority of states have – or are developing – similar documents for infants and toddlers. This article explains why ELS have become so common within the field of early care and education in the United States, how the documents have been developed, the content that has been included in the ELS, and how they are being implemented.

The Momentum of the ELS Movement in Early Care and Education

Standards-based education has been the norm within the United States' K-12 education arena – the educational system

serving older children – for almost two decades. Starting in the late 1980s, states developed standards to define what teachers should teach, what students should learn, and what should be assessed to determine the degree to which schools have been effective in helping students learn the content articulated in the standards. For several years, the early-care and education field was excluded from the effort to define standards for what children should learn, but this changed shortly after the turn of the century. Educators and policymakers began to focus their attention on developing ELS for use with children before they enter the K-12 education system.

The shift to developing ELS for younger children emanated from four developments:

1. extension of the standards-based education movement downward to younger ages;
2. increasing investments of public funds in services for young children, which led policymakers to ask what child outcomes should be expected from the public investments in early care and education;
3. increasing recognition of the potential benefits of early care and education in closing the achievement gap noted among children who enter school behind their more advantaged peers; and
4. increasing knowledge – with regard to the capabilities of young children – gained through research and articulated in seminal publications such as *Eager to Learn*, a report prepared by the United States Department of Education and published in 2001.

Taken together, these factors created a climate that was ripe for work on ELS, and the federal government added to the momentum by introducing the Good Start, Grow Smart early-childhood initiative in 2002. The Bush Administration's Good Start, Grow Smart initiative directed states

to include plans for developing voluntary early-learning guidelines in the areas of language and early-literacy skills as part of their plans for their Child Care Development Funds (CCDF), the plan states submit to describe how they are using childcare funds received from the federal government. Good Start, Grow Smart specified that these voluntary early learning guidelines should be aligned with K-12 standards and should be written in a format that would be applicable in various childcare settings. A combination of policy-related developments – combined with developments within the field of education – led states to develop ELS.

The momentum to develop ELS was not, however, without critics. Standards to define what very young children should know and be able to do seem inherently at odds with several long-standing tenants of early-childhood education. First, it has long been recognized that development at this age proceeds in an uneven and episodic manner. Individual children, typically, develop faster in some areas of development than others, and their development may proceed unevenly – with periods of rapid progress, followed by periods where their development seems to lag or not progress rapidly. Furthermore, children's development is highly influenced by their environment and their experiences – which vary tremendously from one child to another. Children who experience an enriched environment where they are exposed to stimulating materials and interactions at home and/or in an early-care-and-education setting will exhibit vastly different progress than children who experience impoverished or unstimulating environments. Given the high degree of variability expected in children, critics wondered how ELS that define uniform expectations for children's development could be written, and pointed out great potential for misuse of the documents. Some educators feared that standards to define expectations for children's learning and development before they enter kindergarten would be used inappropriately to make judgments concerning individual children or to keep children who do not demonstrate that they have met the standards from entering kindergarten.

ELS also seemed to be at odds with traditional early-childhood pedagogy – which emphasizes individualized learning experiences planned for children based on observations of the children's interests and their current knowledge and skills. Critics wondered if the rich tradition of developmentally appropriate practices – based on the tenets of learning experiences that are individually, culturally, and developmentally appropriate – could continue in a standards-based education approach. While the field has accepted the importance of program standards to define basic requirements for the services provided through early-childhood programs, the idea of using ELS to define what children should learn seemed counter to traditional early-childhood pedagogy. Some educators

worried that early-childhood curricula would become increasingly focused on academic skills, giving less attention to the developmental orientation and leaving little room for individualization.

Despite these concerns, many educators within the field recognized the potential benefits of high-quality ELS that are implemented appropriately. ELS that define appropriate expectations for young children can be used to educate teachers/caregivers with regard to child development, to engender more intentional teaching practices, to establish credibility and the importance of learning that takes place before children enter school, and to offer common goals that could be shared across the many different types of programs serving children before they enter school. Recognizing the tremendous potential benefits as well as the potential risks of ELS, two national organizations joined forces to call attention to the importance of ELS and provide guidance on issues related to ELS. In 2002, the National Association for the Education of Young Children (NAEYC) and National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE) published a position statement to guide states as they develop ELS. The position statement addresses the process that should be used to develop ELS, the content of such standards, and how the standards should be used.

Development of ELS in the United States

There has been a tremendous momentum toward developing ELS within the United States. Whereas in 2002, only 27 of the 50 states had published ELS documents for preschool-age children, in 2008, all 50 states, plus the District of Columbia, had published ELS for preschool-age children. What is more, the momentum has spread to programs serving younger children. While, in 2002, only four states had ELS to define expectations for infants and toddlers (most commonly called Early Learning Guidelines for children at this age), in 2008, 22 states had developed infant-toddler ELS and ten were in the process of developing such documents. By the close of 2009, all 50 states and the District of Columbia will have ELS for preschool-age children and well over half of the states will have infant-toddler ELS. This section provides information on how these documents have been developed and the resources that are available to states as they develop ELS.

ELS Development Process

In 2002, a national study was conducted to collect data on how states were developing their ELS. The authors

interviewed multiple informants from each state to collect data on the standards-development process. Results indicated that the process varied from state to state, but that most could be described as highly collaborative and lengthy.

States varied in their reasons for initiating the standards-development process. In nine states, there was a legislative mandate to develop ELS, and in four states, there was a directive from the state's department of education (the agency responsible for the K-12 education system). In 14 states, the process did not begin with a mandate from the legislature or a governmental agency, but emerged because early educators felt the need for ELS. In some cases, respondents indicated that starting the ELS-development process was a preemptive move. Early educators in their state believed that a mandate for ELS was forthcoming from policymakers and started the process in order to be proactive and have more control over the process than they would if they were responding to a mandate.

Irrespective of how the process started, most state respondents noted that, as the standards development work unfolded, it was inclusive of stakeholders. Committees were formed that included representatives from the state's department of education, local school districts, department of social services, higher education faculty, parents, content experts from a variety of disciplines, and resource persons – such as children's librarians. Respondents indicated they felt it was important to include a wide array of stakeholders in order to ensure that the ELS reflect the values and priorities of their own state. Furthermore, typically states engaged the public in the development process – seeking input via open forums and public hearings. Input from parents took different forms – with parents engaged in committee work in some states and serving on advisory committees or participating in focus groups in others.

Respondents to the survey reported that the typical timeframe for the ELS development process was 1–2 years, with one state reporting that it took 5 years for them to develop their ELS. Given the number of stakeholders involved and the complicated nature of the task, it is not surprising that the development of ELS required a lengthy process. The ELS-development process also required significant resources – both in terms of personnel time and financial support. Although respondents from three states indicated that there were no funds earmarked for standards development, the majority of states did procure financial resources to support the process. Typically, states combined funds from several sources – often using the federal Child Care and Development Block Grant (CCDBG) and Individuals with Disabilities Education Act (IDEA) funds. The state's department of education was the most common source of funding cited for the development process, and some states used private sources of funding – such as grants from foundations or corporations.

Resources Used to Develop the Content of ELS

State work-groups, typically, consult a variety of resources when developing their ELS. The mostly commonly cited resource is the state's own K-12-standards document. ELS-development committees consult their state's standards for kindergarten as they develop the ELS for prekindergarten-age children to promote consistency between expectations for prekindergarten-age children and kindergarten-age children. State committees developing ELS also review the work of various developmental theorists, national frameworks or standards, such as the Head Start Outcomes Framework, the work of the National Education Goals Panel, and publications from the National Association for the Education of Young Children as they define ELS. In addition, committee members developing ELS, often, seek to learn what other states have addressed in their ELS. Committees writing ELS, often, look at ELS documents from another state, and, sometimes, use standards from other countries or provinces – particularly British Columbia – as examples.

A number of other resources are available to states that are developing or revising the content of their ELS. Starting with the NAEYC/NAECSSDE position statement, in 2002 (NAEYC & NAECSSDE), various national organizations have developed resources to assist states as they write or revise their ELS documents. The Council of Chief State School Officers (CCSSO) maintains a website that catalogs standards developed by each state and provides online linkages to each state's document. The National Center Child Care Information and Technical Assistance Center (NCCIC) also posts lists of ELS from states and has developed several documents to describe issues ELS writers might need to address. Zero to Three convened a national panel of experts to offer guidance to state agencies developing infant–toddler ELS. The resulting document – entitled *Early Learning Guidelines for Infants and Toddlers: Recommendations for States* – is available on the Zero to Three website. As the number of states that have published ELS documents has increased, so too has the number of resources available to support the development process. Websites for these resources are listed at the end of this article.

The Content of ELS Documents

The advent of ELS is a relatively new development within the field of early care and education in the United States. Given the potential use of the documents in early-care and education programs, it is important to understand the content of these documents. Two studies have been conducted to examine the content of ELS. One looked at the content of ELS written for

preschool- or prekindergarten-age children, and the second study examined the content of ELS written for infants and toddlers.

The Content of Preschool ELS

Content analyses were conducted on 46 ELS documents published as of January 2005. An analytic framework that explicated various elements of development and learning within five developmental domains – physical and motor development, social and emotional development, approaches toward learning, language and communication development, and cognition and general knowledge – was used to analyze the content of the ELS documents. Each individual standards item within each of the ELS documents was coded according to where the content of the item fit on the analytic framework. The percentage of ELS items coded for each of the five developmental domains, and for specific indicators within each domain, was then calculated and analyzed.

States have, generally, focused more attention on two domains of learning – cognition and general knowledge, and language and communication. On average, approximately 40% of states' ELS items addressed children's cognitive development and knowledge they need to acquire about the world. ELS coded within this category included children's knowledge of factual information such as naming colors, mathematical skills, and their knowledge of social conventions. By far, most of the ELS items coded within this domain addressed knowledge of the physical world and knowledge related to mathematics. On average, approximately 30% of the ELS items addressed the language and communication domain – which includes skills and knowledge related to receptive and expressive language development and early-literacy development. The ELS standards within this domain most frequently addressed children's social uses of language, their writing skills, and their knowledge of print.

ELS written for preschool-age children have also addressed the remaining developmental domains (physical development, social-emotional development, and approaches toward learning), but to a lesser extent. The physical development domain was addressed least often, with only about 8% of the ELS items addressing this domain. Social-emotional development and approaches toward learning have been addressed to a greater extent, but each of these two domains account for only ~11% of the ELS items. Within the ELS items written for children's physical development, the vast majority addressed children's motor development. Social skills with peers, emotional expression, and self-concept were most commonly addressed within the social-emotional domain. Within the approaches toward learning domain, states have most often addressed children's ability to reflect and interpret information, their curiosity, and their creative abilities.

The Content of Infant-Toddler ELS

Similar analyses were conducted on 21 infant-toddler ELS documents to examine the content of ELS written for infants and toddlers. The analytic framework used for the preschool-age ELS-content analyses was modified significantly to reflect the learning and development of infants and toddlers. Each of the ELS items written for infants and toddlers was coded according to where it fit on the framework, and then percentages were calculated in the same manner as they were in the preschool ELS study.

The content of the infant-toddler ELS items included in the analyses was relatively balanced across four of the five developmental domains. The mean percentage of ELS items coded as physical development, social-emotional development, language and communication development, or cognitive development, ranged from 20% to 25%. The balance of ELS items across the four domains differed, however, by the age for which the ELS were written. Physical and social-emotional development tended to be emphasized more heavily in ELS written for children ages birth through 18 months, while language and communication and cognitive development were the domains emphasized most within ELS written for children ages 19–36 months.

Far fewer ELS items addressed children's approaches toward learning within the infant-toddler ELS documents. On average, only about 5% of the ELS items addressed children's approaches toward learning. It is unclear why the state ELS for infants and toddlers have not addressed the approaches toward learning domain to the same extent as other domain areas. It is possible that the committees developing the infant-toddler ELS did not see children's approaches toward learning – characteristics such as interest in the environment, persistence, and problem solving – as important for children of this age, or perhaps they considered these types of skills as so much a part of the development of children of this age that they did not need to address this area in the ELS.

Implementation of ELS

State agencies develop ELS with the intention of improving the quality of care and education children receive in various early-education settings. By defining the skills and knowledge that are important for young children, ELS developers hope to improve the knowledge and skills of teachers, and engender more intentional teaching in early-education classrooms. To this end, state agencies that develop and implement ELS intend for teachers to use the documents in their curriculum planning. Prekindergarten programs operated with state funding are

most commonly required to use the ELS. States have also promoted the use of early-learning standards in a number of other programs, including private childcare, Head Start, IDEA programs serving children with disabilities, and Even Start family literacy programs, but the use of the ELS in these programs is typically on a voluntary basis. States make the documents available to other programs and encourage them to use the ELS, but, most often, it is the state's pre-kindergarten programs that are required to use them.

In addition to requirements that programs use the ELS, state agencies have developed various types of resources to support teachers in their efforts to use the documents. New in-service training initiatives have been developed to introduce teachers to the content of the ELS, and more in-depth training has been developed in many states to teach teachers how to use the ELS to plan intentional, goal-directed learning activities for children. Several states have also developed online and print resources to help teachers and administrators learn about the ELS. Often known as a Tool Box, these materials provide teachers with access to information with regard to the ELS and with suggestions on how they can use the ELS. Several states have developed resources for parents. These documents are intended to help parents understand the content of the ELS, and to learn how they can help children make progress in the areas of child development described in the ELS.

The ELS have also been incorporated into the preservice professional development provided in many institutions of higher education. Early-childhood teacher-education programs have incorporated the documents into courses, with some faculty using the ELS as a required or recommended textbook in their courses and others requiring that students use the ELS in the lesson plans they write as part of their course assignments or practical experiences. One relatively new development within the field of early care and education is the advent of standards that define what teachers should know and be able to do in order to teach effectively. These documents – generally known as professional competencies – articulate the content knowledge, pedagogical knowledge, and skills that the state requires of teachers in order to be credentialed. To encourage the use of the ELS in preservice professional development, more than half of the states have developed or revised their professional competencies based on the content of the ELS. Teachers are required to learn about the ELS and to demonstrate teaching skills that will help children make progress in the areas described within the ELS. ELS have become a central component of the in-service and preservice professional development offered in many states, and are being used to define requirements for the competencies teachers are expected to demonstrate.

Summary

ELS are a recent and important development within the early-care and education system within the United States. Within the past decade, all 50 states have developed ELS for preschool-age children and over half of the states have similar documents to define expectations for infants and toddlers. States have invested significant efforts in developing these documents, convening stakeholder groups to develop standards and collecting public input on draft documents. The content of preschool-age ELS tends to focus more attention on content related to children's knowledge about the world and their language and literacy development. The content of infant-toddler ELS is more balanced across the developmental domains, with roughly equivalent numbers of standards to address children's physical, social-emotional, language, and cognitive development.

ELS are an important component of efforts to improve the quality of programs serving young children. States have developed in-service training and are incorporating ELS into courses offered in institutions of higher education to teach teachers and caregivers how to use the ELS in their curriculum planning. The hope is that using ELS can improve the quality of teaching and learning in early-childhood-education settings by engendering more intentional teaching practices to address the content of the standards.

Along with these developments within the United States, ELS are increasingly common in countries around the globe. It stands to reason that the content of ELS developed by different countries may vary, but the purposes of the documents may be common across the countries. ELS are viewed as a strategy to help adults foster the skills, knowledge, and characteristics within children that are valued and expected. As such, they have great potential to shape the course of early care and education in the United States and around the world.

See also: Early Care and Education Programs for Infants and Toddlers; Global Status of Early Learning and Development Standards; Literacies in Early Childhood: The Preschool Period.

Further Reading

- Kagan, S. L. and Scott-Little, C. (2004). Early learning standards: Changing the parlance and practice of early childhood education? *Phi Delta Kappan* 85(5), 388–396.
- National Association for the Education of Young Children (NAEYC), National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE) (2002). *Early Learning Standards: Creating the Conditions for Success*. http://ericps.crc.uiuc.edu/naecs/position/creating_conditions.pdf (accessed September 2009).
- National Education Goals Panel (1995). Reconsidering children's early development and learning: Toward common views and vocabulary.

- In Kagan, S. L., Moore, E., and Bredekamp, S. (eds.) *Goal 1 Technical Planning Group*. Washington, DC: US Government Printing Office.
- National Research Council (2001). *Eager to learn: Educating our preschoolers*. In Bowman, B. T., Donovan, M. S., and Burns, M. S. (eds.) *Committee on Early Childhood Pedagogy. Commission on Behavioral and Social Sciences and Education*. Washington, DC: National Academy Press.
- Petersen, S., Jones, L., and McGinley, K. A. (2008). *Early Learning Guidelines for Infants and Toddlers: Recommendations to States*. Washington, DC: ZERO TO THREE. <http://www.zerotothree.org/site/PageServer> (accessed September 2009).
- Scott-Little, C., Kagan, S. L., and Frelow, V. S. (2003). Creating the conditions for success with early learning standards: Results from a national study of state-level standards for children's learning prior to kindergarten. *Early Childhood Research and Practice* 5(2). <http://ecrp.uiuc.edu/v5n2/little.html> (accessed September 2009).
- Scott-Little, C., Kagan, S. L., and Frelow, V. S. (2003). Standards for preschool children's learning and development: Who has standards, how were they developed, and how are they used? Tallahassee, FL: SERVE <http://www.serve.org/Products/ProdPub.php> (accessed September 2009).
- Scott-Little, C., Kagan, S. L., and Frelow, V. S. (2005). *Inside the Content: The Depth and Breadth of Early Learning Standards*. University of North Carolina at Greensboro: SERVE Center for Continuous Improvement. <http://www.serve.org/Products/ProdPub.php> (accessed September 2009).
- Scott-Little, C., Kagan, S. L., and Frelow, V. S. (2006). Conceptualization of readiness and the content of early learning standards: The intersection of policy and research? *Early Childhood Research Quarterly* 21, 153–173.
- Scott-Little, C., Kagan, S. L., Frelow, V. S., and Reid, J. (2008). *Inside the Content of Infant-Toddler Early Learning Guidelines: Results from Analyses, Issues to Consider, and Recommendations*. Greensboro, NC: University of North Carolina at Greensboro. <http://www.policyforchildren.org/policy.html> (accessed September 2009).
- Scott-Little, C., Kagan, S. L., Frelow, V., and Reid, J. (2009). Infant-toddler early learning guidelines: The content states have addressed and implications for programs serving children with disabilities. *Infants and Young Children* 22, 87–99.
- Scott-Little, C., Lesko, J., Martella, J., and Milburn, P. (2007). Early learning standards: Results from a national survey to document trends in state-level policies and practices. *Early Childhood Research and Practice* 9(1). <http://ecrp.uiuc.edu/v9n1/little.html> (accessed September 2009).

Relevant Websites

- <http://www.ccsso.org> – Council of Chief State School Officers: Early Childhood Assessment Consortium.
- <http://www.naeyc.org> – National Association for the Education of Young Children: Position Statements.
- <http://nccic.acf.hhs.gov> – National Child Care Information and Technical Assistance Center.
- <http://www.zerotothree.org> – Zero to Three.

Global Status of Early Learning and Development Standards

P R Britto, Yale University, New Haven, CT, USA

S L Kagan, Columbia University, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Early childhood development – The age span below 8 years.

Early learning and development standards – Statements of expectation for what children should know and be able to do.

Validation – Methodology to determine accuracy of information.

Internationally, early childhood development (ECD) is coming of age, accompanied by a growing recognition of the importance of the early years. Fueled by the latest developments in science and a host of recent reports on ECD, attention to the field is rapidly burgeoning. This work has indicated that a majority of children, 8 years of age and under, are not developing to their fullest potential. These children primarily reside in the developing world and bear the greatest burden of poverty, disease, and violence. The staggering statistics, coupled with the empirically grounded importance of the early years, have led to a call for action in discovering the most effective and optimal approaches to improve outcomes for young children across the globe.

Going Global with Early Learning and Development Standards (henceforth referred to as Going Global), an initiative rooted in Early Learning and Development Standards (ELDS), is one such approach that resonates with this call for action. Standards, in general, are statements that specify an expectation for achievement. ELDS are statements that describe expectations for children's behavior and performance across multiple dimensions of learning and development, reflecting 'what children should know and be able to do.' For supporting countries engaged in the development and utilization of ELDS, Going Global is a culturally grounded, inclusive, and child-centered approach to standards. It uses a partnership model, between government, the development agencies, primarily the United Nations Children's Fund (UNICEF), the private sector, academia, and other relevant stakeholders to develop and implement ELDS. It posits that, once developed and validated, the ELDS can be used for multiple purposes, from improving the quality of learning environments and supporting parenting, to providing data for policy decision. In so doing, the ELDS not only address the needs and rights of children, but also foster a truly integrated approach to ECD.

In this article, we provide a conceptual background to the Going Global initiative; briefly describe the process of developing and implementing ELDS; present a brief status report on the countries participating in this initiative and their achievements; and finally present a set of lessons learned from this initiative.

Conceptual Background of ELDS

Standards, in and of themselves, are not a new concept. They have been the *norm-de-rigor* of efficiency and accountability in several industries, including many human services. In education, they made their appearance in the early 1980s and by the late 1990s were being considered by early childhood educators. Given the newness of the concept of standards to the field of early childhood, there has been confusion around this concept, both in terms of the different types of standards applicable for ECD and the uses to which these standards can be put to improve child outcomes. To that end, in work for the National Governors Association in the United States, Kagan sought to distinguish among different kinds of standards and developed a typology of standards, germane to ECD, as follows:

Type I: ELDS. These are typically a set of statements that inform various audiences about children's learning and development accomplishments. They are statements that indicate what children of specific ages should know and be able to do.

Type II: parenting standards. These define what parents should know and do to advance their young children's health, development, and education. Parenting standards provide information on parenting behaviors, caregiving practices, stimulating and supportive experiences, and opportunities for learning, for example, standards, such as the frequency of shared book reading, or the presence of printed materials in the home.

Type III: teacher standards. These define what teachers should know and do to advance their students' learning. Teacher standards provide information on teaching practices, educational interactions, pedagogy, and creating opportunities for learning. An example of teacher standards is: all teachers should know how to assess their students' competence and report such findings to parents.

Type IV: social indicator standards. These are those that yield information about the overall social and economic

conditions of children and their families. For example, social indicator standards can include the number of children living in poverty, the number of children born to single parents, or the number of live births.

Type V: access to services standards. This reflects the nature of services that exist and the access that children have to these services. Examples include the percent of children who have access to preschool services, the percent of children living in families that receive income subsidies, and the percent of pregnant women with access to prenatal care.

Type VI: system effectiveness standards. These define the degree to which services and programs work together to improve their effectiveness. For example, the degree to which early childhood programs, launched by a ministry of health and a ministry of social welfare, might work together to improve the learning and development of young children and family functioning.

Using this typology as a guide, the Going Global effort focuses on type 1 standards or ELDS (what young children should know and be able to do). This focus is predicated on the understanding that the heart of all standards should be child based. Once type 1 standards are developed and used, other types can be created, thereby providing an integrated approach to ECD. A second key premise underlying ELDS is that it is rooted in cultural and national expectations of what children, residing in a given country, should know and be able to do at predetermined ages. Standards emanate from the socioeconomic, cultural, and political context within which the children live. No single set of standards is, therefore, suitable for all of the world's countries. The standards-based approach is founded in research and scientific knowledge of the processes and outcomes of early learning, taking into consideration cultural, linguistic, and socioeconomic differences, as well as children with special needs. A third key premise of ELDS is that it must respect holistic child development and therefore must embrace multiple domains, including, but not limited to, physical and motor development, social and emotional development, cognitive development, language development, and approaches toward learning.

Process for Developing and Implementing ELDS

As the ELDS can be used for so many purposes and because they need to include a variety of domains, an inclusive, systematic process of development is recommended. The four-phase process includes:

1. making initial decisions;
2. drafting the standards;
3. validating the standards; and
4. implementation.

With the understanding that the skill set required for each phase might vary, all phases must be highly inclusive with representation of multiple sectors and disciplines.

It should be noted that the process of developing ELDS commences with a reflective and critical dialog about country cultural, social, and political values. The values conversation focuses on articulating country and cultural traditions, heritage, and ideals that should be carried forth through future generations, the place of early childhood and view of young children in the country. Other concomitant values questions are discussed during this phase. The goal of this prerequisite stage is to arrive at a set of value statements that will help guide the content, development, and implementation of ELDS.

Phase 1: Making Initial Decisions

The foundation of ELDS is laid at the initial decision-making stage with four decisions that are undergirded by country values: guiding principles; domains of development; ages covered by ELDS; and how ELDS will be used.

Decision 1: Guiding principles

A set of principles needs to be developed that provides the answer to the question, 'What principles should guide our work?' or, 'What do we believe in?' Principles can be developed for several aspects of the ELDS – content, development process, implementation, and utilization.

Decision 2: Domains of development and learning

Domains are the broadest conceptualization of an area of learning and development and are then decided on the basis of culling together the values and the scientific literature on ECD. The literature is used to provide a framework for amalgamating the values into a scientifically recognized typology of domains, while also providing the rationale for inclusion and substantiation of values. Therefore, domains of development, included in a set of standards, could be considered a reflection of a country's priorities and expectations for their young children.

Decision 3: Ages covered by ELDS

The age-range and age-grouping decisions are linked to several factors, such as country values, the developmental literature, the national policies and service provision structure in a country, and the purpose for which the standards are being developed. Given that standards are statements of what children should know and be able to do, specifying the ages of the children is of pivotal importance, as expectation is intrinsically linked to age. An important set of decisions, regarding the age ranges to be covered by ELDS, relates not just to the overall age coverage, but also to the age groupings that will be addressed by the indicators (an element of ELDS).

Decision 4: Uses of ELDS

As stated earlier, ELDS can be used for several purposes:

- *Improve instruction.* Standards can form the basis for revising the educational program for young children according to their needs and progress.
- *Support parenting skills and behaviors.* Standards can be used to help parents see what is and can be realistically expected from young children.
- *Improve curriculum.* Standards can be used to guide the development and revision of curriculum of early childhood and preschool programs.
- *Improve teacher preparation.* Standards can guide the development of curricula for teacher-training programs by providing information about what teachers of young children should know and be able to do in their roles as teachers.
- *Evaluate program effectiveness.* Standards can form the basis for selecting or developing instruments for assessing children's progress as part of a comprehensive evaluation of a program's effectiveness.
- *Monitor national progress of children and families.* Standards can be used to collect national data on performance of children to determine how the nation's children and families, as a whole, are doing.
- *Improve public knowledge of children's development.* The standards can be used for public information materials, for example, television, newspaper, and public service announcements.

The decision regarding uses is an important one because many of the subsequent stages of development are guided by the primary uses. However, it should be noted, given the versatility of standards, that they could be developed for several purposes.

Phase 2: Drafting the Standards

After the foundational decisions are made, the standards document is drafted. Country standards are articulated in a document that provides a framework for understanding the expectations for what children should know and be able to do. Typically, the ELDS document provides background information on the process of development, a brief statement of values, and lists guiding principles. The framework for presenting the standards contains several components, such as the domains of development and learning; the standards statements themselves; indicators, which are measurable and observable, and are linked to the standards; and some preparatory learning activities for the adults that would help the children achieve the standards (see **Table 1** for an example). Several standards are required per domain and several indicators per standard.

In general, the process of developing standards is iterative, going through stages of review, reflection, and revision prior to finalization.

Table 1 Example of an ELDS framework

<i>Domain 1:</i> Language and literacy development.
<i>Standard 1:</i> Children develop skills in listening and in understanding language.
<i>Indicator:</i> Child is able to follow directions that involve a two- or three-step sequence of actions.
<i>Preparatory learning activities:</i>
1. Give oral directions and play a game like caregiver says.
2. Make the children give simple directions to each other.

Phase 3: Validation

The final stage in the development of standards is validation. With reference to ELDS, validity refers to the extent to which the standards are what they say they are. Validation, a cornerstone of good practice in the development of standards, measures, and instruments, is essential for ensuring the accuracy of the standards. Therefore, before standards can be implemented, they need to undergo a process of validation to ensure that the content and age expectations of the standards are accurate and valid for the children of the country.

The ELDS validation model has three tiers which build sequentially on the previous tier, and also increase in level of methodological sophistication. There are three types of validation, recommended for ELDS, linked with the uses of the standards. The model has been built on the empirical literature with applicability for the many uses of the standards.

Tier 1 of the model ascribes that content validation be conducted for all seven uses of ELDS. Content validity is the extent to which the ELDS represent national values, expectations, and the scientific conceptualizations of child development. The rationale for conducting content validity is to ensure, that for all uses to be nationally relevant, the content has to be accurate and valid. Therefore, content validity is required for all uses.

Tier 2 of the model is required for all uses, except for building public knowledge and will power. Two types of validation are required: content (described previously) and age validation. Age validity is the extent to which the indicators are age appropriate for most of the population. Age validation is concerned with the age appropriateness of the indicators (e.g., Are they too easy, too hard, or just right for the children of a given age?), based on a consensually predetermined percentage of children that can attain them. The rationale for conducting age validity is to ensure that expectations for children's learning and development are correct for their age, given the uses for education, parenting, program evaluation, and national monitoring.

Tier 3 of the model is the most advanced and is required for national monitoring and program evaluation. Three types of validation are required for these two uses: content, age (describe above), and predictive validity.

Predictive validity, with respect to ELDS, is defined as the extent to which the indicators demonstrate a significant degree of association with future desired outcomes. Data from national monitoring, which reports on the status of children, are typically used for policy mobilization. Most ECD programs are based on the premise that they are preparing children for future success in life, and, most immediately, to succeed in school. Consequently, indicators, selected for program evaluation and national monitoring, should have demonstrated predictive validity.

Scientifically accepted methodologies are recommended for conducting content, age, and predictive validation studies, in keeping with the resources and capacity in the country to conduct such work.

Phase 4: Implementation

Once ELDS are validated, they are ready to be implemented for one or more of the several uses, outlined above. Typically, the process of implementation requires the selection of a set of standards, most appropriate for use, and transforming them in implementation tools. For example, if standards are to be used in parenting programs, standards most applicable for the age which the program is targeting and the aspects of parenting which the program focuses on, are selected from the ELDS document, through a systematic item-selection process, and converted into parenting materials, such as brochures, posters, and pamphlets. The process of conversion from the ELDS document to the implementation document is dependent on the use and, therefore, varies considerably from country to country. In addition, ELDS are also implemented in alignment with other sets of existing standards, such as program standards and curriculum standards.

Status Report on Countries Participating in ELDS

Going Global, a multi-country initiative, was launched to create integrated systems for ECD. In each country, this initiative has been adapted in keeping with the economic-political context and sociocultural ethos. However, several basic tenets of the initiative have been maintained. The initiative took roots in 2003 with six pilot countries: Brazil, Ghana, Jordan, Paraguay, Philippines, and South Africa. These countries represented geographic and economic diversity. Currently, in 2008, this initiative is alive and contributing to the ECD landscape in almost every region of the world: East Asia and Pacific (e.g., Cambodia, China, Fiji, Thailand, and Vietnam); South Asia (e.g., Bangladesh, Maldives, Nepal, Pakistan, and Sri Lanka); Central Europe and Commonwealth of the Independent States (e.g., Bulgaria, Georgia, Kosovo, Moldova, and

Tajikistan); and Latin America (e.g., Chile and Uruguay). (This is not an exhaustive list of countries. Examples have been listed to highlight regional diversity and expansion of the initiative.) The number of countries participating in this initiative continues to grow and expand, in real time, thereby precluding a final tally count.

The countries are at different stages of development and implementation with their ELDS. Most of the six original pilot countries have completed the development of their standards, including validation and are currently implementing them for multiple uses. For example, ELDS are being used: to revise preschool and ECD curricula in Ghana, Philippines, and South Africa; to inform parenting programs in Jordan, for National Monitoring in Jordan; and to revise teacher-preparation curriculum in Ghana.

The second cohort of countries to join this initiative was primarily from East Asia. Most of the countries in that region are nearing the end of validation, and many have begun implementation in uses beyond those envisioned by the initiative. For example, in Cambodia, the ELDS standards have been used to ensure that the educational programming for children on television is rooted in the standards, and thereby, ensuring cultural relevance and allegiance to the country's values.

The next set of countries in Central Europe and the Commonwealth of Independent states are near the end of the development process, with interesting applications emerging in some of the countries. For example, Moldova has created a comprehensive set of standards and measurements tool to align child and instructional standards used in preschool classrooms.

The newest set of countries from South Asia are either just beginning the process (e.g., Nepal) or are at the writing (e.g., Bangladesh) or validation phase (e.g., Pakistan). It is anticipated that, in 2009, several countries from East and South Africa will join this initiative.

Lessons Learned

The lessons learned from the initiative hold valuable recommendations for future direction not only for ELDS, specifically, but also for ECD programs and policies in general. The following aspects are discussed: conceptual; procedural; and implementation.

Conceptual Gleanings

Each country has been developing their ELDS, rooted in their values, cultural ethos, and child development research, through an inclusive broad-based process. As a result, we now have a large corpus of standards that provide a window into interesting and unique conceptualizations of ECD. Initial, and informal content analyses have revealed, while many similarities are noted across

country conceptualizations of ECD (e.g., motor development and language development), unique domains of development have also emerged (e.g., national values and moral development). This corpus has tremendous potential to widen our vistas on ECD and create a South-to-North transfer of knowledge regarding our conceptualization of ECD.

Another valuable insight gained during the course of this initiative is the varied reaction to the term standards. In some parts of the world, where standards were part of the system, the term met with no resistance, whereas in other parts of the world, it has been off-putting. As a result, countries have elected to use their own lexicon and terminology to label their ELDS, including benchmarks, competencies, and guidelines.

Procedural Gleanings

A key ingredient of success in the development and implementation of ELDS has been government involvement and support. If the initiative occurred in the absence of government participation, the process either was disbanded or implementation was limited. Other factors that made for success were consistent and solid leadership, inclusivity extended to all traditional and nontraditional stakeholders, and finally, and most importantly, a need for ELDS in the country.

Through the processes of developing ELDS, participating countries derived many benefits. First, by making standards development an inclusive process, capacity for ECD is built in the country. There is now a large cadre of professionals, not just those with an ECD background, but also those who are knowledgeable about latest developments and the importance of ECD (e.g., Philippines). Second, ELDS has generated a base for implementation support. By information from professionals across sectors, there is far greater synergy and integration in addressing ECD issues (e.g., Ghana). Third, data collected during the age validation phase of the work could be used as comprehensive baseline data for children's development (e.g., China). In addition, the process used for age validation provided the platform for future national data collection. These byproducts have worked toward the success of this initiative in individual countries.

Implementation Gleanings

ECD implementation challenges, in most parts of the world, are instructive. First, by linking ELDS to other ongoing initiatives, the linkage not only smoothens the implementation process, but also creates for integration across programs. For example, in Cambodia, ELDS were used to create expectations for children school readiness, thereby linking a successful preschool program in the country to primary school entry. In Thailand, as part of the national

policy plan, ELDS have been linked to family development and gender development programs. By creating this alignment, it has become possible to train 30 000 caregivers to improve the quality of those programs, based on ELDS.

ELDS, also has valuable lessons for ECD in general. Internationally, school readiness is gaining rapid currency and is built upon three pillars of readiness – children ready for school, schools ready for children, and families and communities ready for school. The application of ELDS simultaneously (for multiple uses – curriculum, instruction, teacher preparation, and parenting programs) addresses these three pillars of school readiness, holistically.

Finally, ELDS has been a positive experience in countries, as noted by the evidence. Many countries report that ELDS has set a historic precedent by bringing together multiple sectors in a shared vision for ECD. The highly inclusive process has pushed ECD policy and advocacy among broad-based communities, including health, education, labor ministries, and community leaders. This approach is a lasting legacy in many countries.

Acknowledgments

We would like to thank UNICEF for the funding they have provided for the Going Global initiative. We owe a debt of gratitude to the UNICEF staff and members of the ELDS teams in countries that are part of this initiative.

See also: Development and Implementation of Early Learning Standards in the United States; Evaluating Early Childhood Education and Care Programs.

Further Reading

- Britto, P. R. (2008). School readiness: A means to achieving child, national and international development. *UNICEF Report*.
- Engle, P. L., Black, M. M., Behrman, J. R., et al. (2007). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *Lancet* **369**, 229–242.
- Gall, M. D., Gall, J. P., and Borg, W. R. (2003). *Educational Research: An Introduction*, 7th edn. Boston, MA: Allyn and Bacon.
- Grantham-McGregor, S., Cheung, Y. B., and Cueto, S., and the International Child Development Steering Group Developmental (2007). Potential in the first 5 years for children in developing countries. *Lancet* **369**, 60–70.
- Heckman, J. J. and Krueger, A. B. (eds.) (2003). *Inequality in America: What Role for Human Capital Policies?* Cambridge, MA: MIT Press.
- Irwin, L. G., Siddiqi, A., and Hertzman, C. (2007). Early child development: A powerful equalizer. *Final Report for the WHO's Commission on Social Determinants of Health*. Geneva, Switzerland: World Health Organization, Commission on Social Determinants of Health.
- Kagan, S. (1996). *By the Bucket: Achieving Results for Young Children*. (Issue Brief). Washington, DC: National Governors' Association.
- Kagan, S. L. and Britto, P. R. (2007). Frequently occurring constructs: An analysis of ELDS. *UNICEF. Technical Report*.
- Kagan, S. L., Moore, E., and Bredekamp, S. E. (1995). *Reconsidering Children's Early Development and Learning: Toward Common Views*

- and vocabulary*. Washington, DC: National Education Goals Panel, 95-03, Goal 1 Technical Planning Group.
- Proctor, R. W. and Capaldi, E. J. (2006). *Why Science Matters: Understanding the Methods of Psychological Research*. Oxford, UK: Blackwell.
- Shonkoff, J. P. and Phillips, D. A. (2000). *From Neurons to Neighborhoods: The Science of Early Childhood Development*. Washington, DC: National Academy Press.
- Somekh, B. and Lewin, C. (eds.) (2005). *Research Methods in the Social Sciences*. Thousand Oaks, CA: Sage.
- UNESCO (2007). Strong foundations: Early childhood care and education. *EFA Global Monitoring Report*. Paris, France: UNESCO Publishing.
- UNICEF (2007). *State of the World's Children*. New York: UNICEF.

Rethinking Early Childhood Education and Care: Implications for Research and Evaluation

K Thorpe, Queensland University of Technology, Brisbane, QLD, Australia

D Cloney and C Tayler, University of Melbourne, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Evidence from a diversity of research studies in developmental neuroscience and developmental psychology provides a consistent, strong, and compelling argument for the role of early experience in directing children's trajectories of achievement and well-being (Shonkoff and Phillips, 2000; Schweinhart *et al.*, 2005; Center on the Developing Child, 2007). Early experiences affect the child's developing brain architecture. The range and quality of early experiences determine the strength of neural foundations set for future learning, social behavior, and relationships. Early learning experiences establish the pathways for children's motivation for school learning and long-term scholastic attainment, while early interactions and relationships with adults and peers establish pathways for their emotional security, sense of agency, self-regulation, and social behavior.

Early Childhood Education and Care (ECEC) programs present an important opportunity for improving the quality of early experience and affecting long-term outcomes. Randomized control trials of early education programs among disadvantaged children have demonstrated a causal relationship between early education and life course (Schweinhart *et al.*, 2005; Reynolds, 1994; Campbell and Ramey, 1994). Longitudinal tracking of participants in these trials indicates that, in adulthood, children who received well-resourced preschool education had more life successes (e.g., high school completion and higher lifetime earnings) and experienced less adversity (e.g., need for special education, participation in crime, early parenthood, and unemployment) than comparable children who had not experienced an early educational program (see Schweinhart *et al.*, 2005). A key question for research is how these findings might apply for all children. There is a need to understand how the potential of ECEC programs demonstrated in these studies of targeted populations might generalize to broader applications of educational policy and practice.

In this article, we examine the question: What ECEC programs, and which of their components, most effectively deliver optimal outcomes for children and society? We examine the conceptualizations and policy directions in English-speaking Western economies, propose an evidence-based approach to define optimal program outcomes, and

outline approaches to conducting research exploring the quality and effect of programs.

Changing Interests in ECEC Programs

The evidence that ECEC programs can have a long reach into adulthood has provoked a rethinking of their place in early childhood development and learning. In English-speaking developed economies (e.g., USA, Australia, and UK), until recent times, early education and child care provisions have been conceptualized separately, serving two quite separate functions often with different political and governance structures (see OECD, 2006). In this article, we refer to this as a differentiated service orientation. The provision of care (child care) has been constructed primarily as delivering the immediate goal of enabling parent participation in the paid workforce. The benefits for children have been viewed as indirect; preventing the adverse effects of poverty and contributing to the material well-being of the family. Commensurate with this focus, the regulation of child care has favored attention to aspects such as safety and hygiene, giving relatively little attention to a child's education. In contrast, early education or preschool programs, in English-speaking countries, focused on short-term goals: particularly preparing children for formal schooling. The role of preschool has been seen by many as providing children with social experiences and cognitive preparation for literacy and numeracy learning on entry to school. Further, many of these programs operate in sessional or part-time modes, making it difficult for working parents to ensure a child's attendance and assure that a care program is available for the remainder of a working day. This dual system of child-care (care) and preschool (education) provision has contributed to inequities and served to create a tension for parents between providing for their children's material needs and their children's educational needs. Research questions, mirroring the dual nature of provision, have tended to focus on either child care or preschool provision. Those focusing on child care have been concerned with the effects on children and asked whether nonparental care causes harm. Research on preschool education has predominantly focused on the narrow remit of school readiness.

The alternative conceptualization to a differentiated service orientation is a comprehensive program orientation. This orientation focuses on the evidence from developmental science, particularly longitudinal studies, that has highlighted the capacity of early education to benefit the child, families, and the community in both the short term and across the life span by promoting economic productivity (through human capital formation) (e.g., Keeley, 2007) and social inclusion and participation (through social capital formation) (Schweinhart *et al.*, 2005; Reynolds, 1994; Campbell and Ramey, 1994). To this end, all programs, regardless of their stated goal, are assessed in terms of this broader framework. Evidence clearly shows the potential for substantial benefits to be accrued through the provision of programs at this early developmental stage. Further, if the opportunity to maximize these benefits is forgone in the cost of achieving the same benefits later in life are greater, there is a high opportunity cost associated with achieving the same benefits later in life (Heckman, 2007).

Figure 1 illustrates the theorized difference in lifelong effect when ECEC programs are offered within the above-mentioned paradigms of either differentiated service orientation (a dual system of care and education where programs are offered primarily to meet the single goals such as parental labor participation or school readiness) or an comprehensive service orientation (where both short- and long-term goals are being addressed and

the rights of children to optimal educational experiences are enhanced).

Competing policy interpretations have grown as a result of viewing ECEC in the long term and adopting a comprehensive conceptualization of ECEC: targeted and universal provision. Targeted provision focuses available and limited resources to ensure the provision of programs for disadvantaged populations only (Penn, 2005). This interpretation is a direct application of randomized control trial studies to new settings and is underpinned by a compensatory approach in which poorer home environment is supplemented by rich educational experience outside the home. In contrast, universal provision takes a broader view of existing research findings and directs focus to social equity by highlighting the potential benefit of programs for all children alongside the relatively greater effects for those who are disadvantaged. In this construction, access to educational programs is presented as the right of all children (OECD, 2006).

Targeted provision is proposed on the premise that it is cost effective, best serves those with greatest need, and more effectively reduces social inequity (e.g., see Rutter, 2006). However, success of programs may be undermined by poor access or poor uptake by those groups who might benefit. Evidence from Early Head Start (Love *et al.*, 2002) and Sure Start UK (NESS, 2005) indicates that the most disadvantaged were less able to benefit. For example, teen parents, lone parents, and

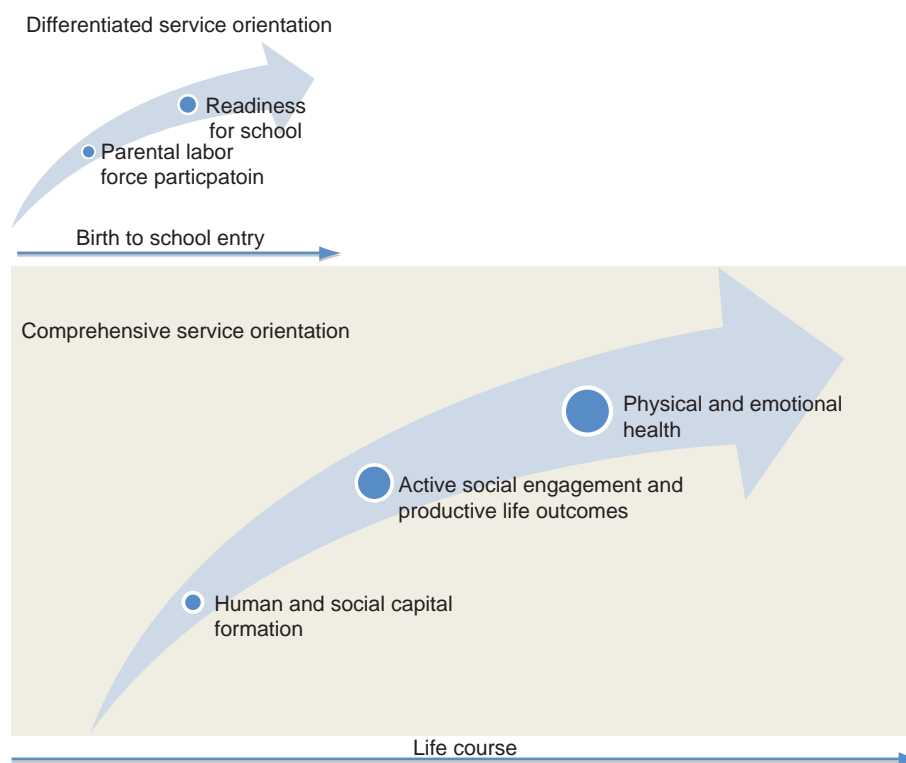


Figure 1 Two conceptions of the ECEC policy framework associated with the effect of programs over the life course.

workless households living in Sure Start UK intervention areas were found to make poor gains or even regress relative to others (NESS, 2005). These findings are likely to reflect poorer uptake and access by the most disadvantaged. Targeting is criticized for being potentially stigmatizing, and it may also be inaccurate. Rutter (2006), for example, argues that one difficulty with Sure Start is that it targeted geographical areas statistically defined as disadvantaged when in the UK many disadvantage families reside within statistically advantaged areas. Moreover, clustering targeted groups in educational programs may exclude social and learning opportunity. When the social composition of an early childhood class is more socially diverse, greater positive effects have been found for children's scholastic and social learning (EPPE, 2002).

Universal provision is based on the tenet that ECEC benefits all children and that when programs are presented as normative, access is increased and uptake by marginalized groups is more likely to be achieved. Hustedt and Barnett argue that it is not the percentage of enrolment in ECEC programs but who is enrolled that makes a difference to social equity. They state that a minimum of 60% of the population should have access before a program is effective in reducing inequality in achievement scores because greater coverage of the population as a whole is required to reach the most disadvantaged. Thorpe *et al.* (2004) report that free and universal provision of a full-time program for 4-year-olds improved performance of all children. In addition, such provision closed the gap in social and cognitive achievements across a school year between disadvantaged children and others. Thorpe *et al.* (2004) proposed the effects reflect improved access for disadvantaged groups into a universal, free program that engages the whole age cohort.

Currently, there are no studies examining the impact of population-based ECEC programs with outcomes into adulthood. The Effective Provision of Preschool Education (EPPE) study provides longitudinal data on the effects of a range of ECEC programs that were delivered to children in the UK. Results of the study have been tracked through to the middle years of school. These findings indicate that not all early education programs deliver positive outcomes in the long term; some have little or only short-term effects (EPPE, 2007). A key task for research is to clearly define outcomes and learning standards for ECEC programs and then identify the programs, as well as features of these programs that best meet these goals.

Defining Optimal Outcomes

We propose, on the basis of a comprehensive conceptualization of ECEC, that there are four outcome criteria against which the success of ECEC programs should be assessed:

1. Access.
2. Benefit to all children.
3. Long benefit.
4. Wide benefit.

Access

Evaluations of ECEC programs very often do not take account of access and yet a program, no matter how well resourced, cannot be of benefit if it is not attended by those for whom it was designed and implemented. The randomized control trials of preschool education conducted in the USA demonstrate the stark effects of not accessing preschool education among disadvantaged populations. In these studies, children who did not access programs had significantly poorer achievement and social outcomes across their lifetime. In a nonexperimental setting, where there is likely to be lower impetus to attend an ECEC program, one key criterion of success is the level of uptake and sustained attendance in the program across the diversity of the population. Access is particularly pertinent when considering marginalized groups and developing ECEC programs as potential meeting places for diverse social groups (see OECD, 2006). For example, in the EPPE study, the children who did not access ECEC programs were a select subgroup with very different characteristics when compared to children who experienced ECEC programs. Children in the EPPE study, who did not experience ECEC programs, were more likely to be from an ethnic minority, have English as an additional language in the home, come from larger families, and have mothers with no formal qualification (EPPE, 2002). Similarly, in Australia, there is disproportionate nonattendance at preschool for Indigenous children (Biddle, 2007). Location and physical access to program may also affect access and need to be considered as part of the features that define a program's quality – that is, its reach. In the Australian context, for example, where 38.5% of the population of 3–4-year-olds do not access preschool education (OECD, 2006), geographical barriers are almost certainly a factor. Cost, timing (number of days per week), and availability (hours per day) of program may also serve as barriers particularly if there is a disjuncture between hours of program provision and the needs of working parents (Thorpe *et al.*, 2004).

Benefit to All Children

Not all ECEC programs deliver benefit to children. Further, the magnitude of benefit varies considerably across programs (Thorpe *et al.*, 2004; EPPE, 2002). The success of a program should be assessed by measurement of the degree of impact on all children from a given sampling frame or community of interest from baseline measurement through repeated measures after the course of the program and to a nominal point in the future. Measurement must take account of other factors that may influence

children's development, including family background and home learning environment (Duncan and Gibson-Davis, 2006; Hungerford and Cox, 2006). The outcome measures would include a range of social and attainment outcome domains. Snow and Van Hemel (2008) suggest a categorization of child outcomes that includes:

1. physical well-being and motor development;
2. socioemotional development;
3. approaches to learning;
4. language (and emergent literacy); and
5. cognitive skills, including mathematics (p. 58).

Long Effects

A key premise of the comprehensive approach to early education is that the early years are vital foundation years that direct life chances. The test of a successful program, therefore, is that its effects endure beyond the course of the program or the period of school entry which follows immediately after: the effect must be sustained. The effects may have direct pathways in which early numeracy and literacy preparation support higher levels of scholastic achievement into the school years. There may also be moderating or mediating effects. For example, the EPPE study indicates that higher-quality education programs served to make children more resilient to poor educational experiences in the primary years of school than those who had poorer forms of provisions (EPPE, 2007).

Wide Effects

Early successes increase the likelihood of future successes in a range of related areas, while limitations on learning can increase the likelihood of compounding difficulties. Language development provides a good case in point. Analyses of the 1970 British birth cohort parsons *et al.*, 2009 indicate that children who had receptive language difficulties at age 5 yet achieved, through educational experience, competence in reading at age 10 were more likely to have an easier pathway into adulthood and to achieve higher levels of socioeconomic status than those who did not achieve reading competence at age 10. Language difficulties have also been shown to link to poorer mental health: the pathway being through decreasing levels of achievement and increasing social exclusion (Rutter and Mawhood, 1991). Assessment of program success should therefore capture not only direct child outcomes (see Snow and Van Hemel, 2008) but also the indirect pathways to other life attainments. Little attention has been paid to long-term health outcomes that might derive from ECEC programs but, given the high association between health and social economic status, this would be an important new direction for assessment of the wide effect of ECEC programs.

The effects of ECEC programs also extend beyond the achievements and well-being of the individual child to include family, community, and societal effects. Engagement between home and school may serve to not only benefit the child (Thorpe *et al.*, 2004) but also engage families in social participation. This is a less studied effect of ECEC programs but one identified by the Organization for Economic Cooperation and Development (OECD) as a potential benefit, and worthy of inclusion as a measure of outcome. The potential societal benefits of ECEC programs are perhaps best demonstrated in the experimental studies. The long-term effects measured were wide and included the prevention of a range of negative outcomes. The reduction of such negative behaviors spills over to produce communities with both tangible and intangible improvements. An example of tangible improvements includes the reduced costs of repairing damage caused by crime and associated lower insurance costs. Intangible benefits can include a better sense of community or safer environment. These are examples of ECEC's ability to produce positive externalities realized as societal benefits.

Identifying Optimal Programs

The claim that high-quality early education delivers positive and enduring outcomes, at least for disadvantaged groups, is ubiquitous in the early education literature. Yet there is, at best, only partial empirical evidence to enable the definition of relevant models of quality ECEC that can be applied to a broad population rather than specific at-risk groups. It is not yet known what model or models of ECEC delivered in a general population yield optimal outcomes relative to each other. This is a key agenda for current research focus.

We propose that a three-staged analytic strategy is required to provide a strong evidence-base for any universal delivery of ECEC programs:

1. Deriving a model of quality in ECEC.
2. Testing alternative delivery and access modes for the model.
3. Evaluating delivery of the derived model.

Deriving a Model of Quality in ECEC

The experimental studies of ECEC have tested the effects of single models of highly resourced ECEC programs against counterfactual conditions (the children in the control group do not receive the experimental program, but may have other educational and care experiences outside the home). The outcomes of these studies indicate that the programs have been extremely successful against three of the four criteria set out in this article for success

in an ECEC program: all report immediate, long, and wide benefits but data on access are limited. High/Scope Perry Preschool Study reported a 69% full-attendance rate but none of the early intervention experimental studies reports on attrition rate (Wise *et al.*, 2005). Despite the gains made and attributed to the treatment condition (model program), these studies cannot provide evidence that the model tested is the most effective model of ECEC even within the population they target. The experimental programs implemented are informed by anecdotal evidence, subject-matter experts, and literature; there is no empirical analysis of the components of program that are most efficacious or produce the greatest effect when combined together. These experimental designs test the efficacy of a model program rather than the best model program. To empirically define quality in ECEC requires comparison of alternative models of ECEC provision and examination of their relative effects. Variability between models of program (contrasting different approaches to ECEC delivery) and within models of program (contrasting different ECEC delivery of the same type) need to be exploited to identify program components and combinations of components that most effectively deliver successful outcomes to a given population of interest. To this end, the use of effectiveness designs, in which the performance of multiple existing ECEC programs are compared against each other, may be a more potent first step before the rollout of experimental designs testing specific models of ECEC delivery.

Unlike experimental designs, where randomization to groups is used to ameliorate potential systematic bias on observed and unobserved variables, effectiveness designs use the natural assignment to ECEC programs that occurs in the community. While this has the advantage of allowing the examination of access and uptake of programs, it also means that statistical approaches are needed to control for selection bias. The National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development (NICHD henceforth) is the most significant example of a nonexperimental design in ECEC. NICHD looks at variability within a program model to estimate the effect on quality in predicting child outcomes. NICHD has tested several approaches to accounting for the effect of background characteristics that lead to selection bias and have categorized them in the groups of: characteristics of the child, of the home, and characteristics of the family. Some of these approaches, such as the difference model, also account for unobserved differences within these categories. By controlling for these background characteristics, it is possible to begin to disentangle the effects of nonrandom selection into groups from the effects of the program on children's outcomes (NICHD and Duncan, 2003; Duncan and Gibson-Davis, 2006). Similar approaches have been used in the Australian Preparing for School study (Thorpe *et al.*, 2004) and also in research

conducted across multiple studies of preschool education in the United States (Early *et al.*, 2007; Mashburn *et al.*, 2008).

Although effectiveness designs have been criticized for their inability to make inference of causality (National Forum on Early Childhood Program evaluation, 2007), they provide the much needed data on the relative effect of program and allow the empirical definition of program quality. Such an approach allows the specification of an optimal model of quality that can be translated into a program. This minimizes the likelihood of falsely concluding that a program does not have an effect (type II error) or of understating the effect of quality ECEC on children's development in a universal population.

Evaluating Alternative Delivery and Access Modes for the Model

As access to programs and maintenance of attendance is a key success criterion to ECEC programs, understanding how any apparent barriers to access might be overcome is an important issue to be addressed by research. Effectiveness designs, because they compare existing programs, are able to identify barriers and biases in uptake across the various program provisions. The identification of key success elements presents the possibility of experimentation with provision to overcome barriers that might relate to geographical, time, and cultural barriers. For example, in the Australian context, the need to examine ways to deliver preschool education to those in remote locations while retaining core quality components could be investigated. Thus, increased attendance at preschool programs by indigenous children in Australia has been found to be associated with the presence of an indigenous preschool worker (Biddle, 2007). Core elements of quality would be trialed alongside component additions and different access modes, depending on the context. This stage does provide an understanding of the functioning of identified high-quality or successful programs in different contexts and populations.

Evaluating Delivery of a Derived Model

The National Forum on Early Childhood Program Evaluation (2007) indicates that experimental and regression discontinuity designs (RDDs) are the gold standard for testing the success of ECEC programs. (RDD uses exogenous variation such as an age cutoff or policy change to allocate participants to groups in a way that approaches random assignment. Such designs can be used to make causal inferences.) Although studies, such as the EPPE study in the UK, have demonstrated that effectiveness designs can be powerful in their own right and make important contributions to the field, the use of experimental and RDD methods are able to provide estimates of causality and are powerful in evaluating the independent

effect of a model program or intervention. It should be acknowledged that implementing experimental research designs of ECEC programs is not free of difficulties. There are ethical dilemmas about the rights of those randomized to the control conditions. There is also difficulty in specifying the control condition as it is seldom that parents of young children would withhold all out-of-home ECEC experiences: unless the experimental intervention is specified as a particularly efficacious program there is potential that parents may source equally good or better ECEC elsewhere. The alternative, RDD designs require conditions that exploit naturally occurring cut-points in provision (e.g., children who sit either side of an age criterion for inclusion/exclusion or introduction of a new program). Although appealing, these require the model program to be implemented such that a hard cutoff point is established. This is often not the case in noncompulsory ECEC, particularly at the pre-K level. In the Preparing for School study (Thorpe *et al.*, 2004), many transgressions of set age criteria were noted in the two preschool programs that were compared. It is therefore vital that robust and powerful model ECEC programs be specified before progressing with experimental designs – failing to do so weakens the cogency of the research conducted.

Conclusion

Despite significant evidence of the potential positive benefits of ECEC programs, there remains no empirical framework from which to design model programs that can be distinguished from other forms of provision as being more powerful or of having a greater impact on children's development. This is particularly true in the general population: quality programs are currently defined by what has been historically seen as important, the subjective inputs of experts (academics, teachers, governments, and businesses), and by imitating programs found to be successful in at-risk populations. The task for research is to determine those programs and program conditions that allow access for all children and deliver benefits that yield enduring and far-reaching outcomes for children, families, and society. We propose that a first stage is to empirically define key components of quality in ECEC through comparison of the effectiveness of existing programs. In addition, research studies need to explore barriers to access and sustained attendance in diverse contexts. From such work, model programs can be specified and tested experimentally. Under such conditions, it is possible to comment on the potential efficacy of ECEC programs in general rather than simply that of programs touted to be high quality but with little evidence of their relative superiority in a wider context.

See also: Evaluating Early Childhood Education and Care Programs; Integrated Early Childhood Education and Care Services – Care, Upbringing, Education and Health; Investing in Early Childhood Education and Care: Some Policy Implications.

Bibliography

- Biddle (2007). Indigenous Australians and pre-school education. *Australian Journal of Early Childhood* **32**, 9–16.
- Campbell, F. and Ramey, C. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development* **65**, 684–698.
- Center on the Developing Child (2007). *A Science Based Framework for Early Childhood Policy: Using Evidence to Improve Outcomes in Learning, Behaviour and Health for Vulnerable Children*. <http://www.developingchild.harvard.edu> (accessed October 2009).
- Duncan, G. J. and Gibson-Davis, C. M. (2006). Connecting childcare quality to child outcomes. Drawing Policy lessons from non-experimental data. *Evaluation Review* **30**(5), 611–630.
- Early, D., Maxwell, K. L., Burchinal, M., *et al.* (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development* **78**(2), 558–580.
- Effective Provision of Preschool Education (EPPE) (2002). Technical Paper 8a - Measuring the Impact of Pre-School on children's cognitive progress over the pre-school period.
- Effective Provision of Preschool and Primary Education (EPPE) (2007). *Influences on Children's Attainment and Progress in Key Stage 2*. Research briefs DCSFRB828 (Cognitive) and DCSFRB007 (Social/Behavioural). London: Department for Children, Schools and Families.
- Heckman, J. (2007). The economics technology and neuroscience of human capital formation. *IZA Discussion Paper Series No. 2875*. Bonn: The Institute for Study of labour (IZA).
- Hungerford, A. and Cox, M. (2006). Family factors in child care research. *Evaluation Review* **30**(5), 631–655.
- Love, J., Kisker, E. E., Ross, C. M., *et al.* (2002). *Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start. Volume 1: Final Technical Report*. Princeton, NJ: Mathematica Policy Research. <http://www.mathematic-mpr.com/PDFs/ehsfinalvol1.pdf> (accessed October 2009).
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., *et al.* (2008). Measures of classroom quality in prekindergarten and children's development of academic, language and social skills. *Child Development* **79**(3), 732–749.
- NESS (National Evaluation of Sure Start Team) (2005). *Report 13 Early Impacts of Sure Start Local Programmes on Children and Families*. London: HMSO.
- NICHD and Duncan (2003). *Modelling the Impacts of Childcare Quality on Children's Preschool Cognitive Development*. Rockville, MD: NICHD.
- OECD (2006). *Starting Strong II*. Paris: OECD.
- Parsons, S., Schoon, I., Rush, R., and Law, J. (2009). Long-term outcomes for children with early language problems: beating the odds. *Children and Society, early view*.
- Penn, H. (2005). *Childcare and Early Childhood Development Programs and Policies: Their Relationship to Eradicating Child Poverty*. London: Childhood Poverty Research and Policy Centre.
- Reynolds, A. (1994). Effects of a preschool plus follow-on intervention for children at risk. *Developmental Psychology* **30**(6), 787–804.
- Rutter, M. (2006). Is sure start an effective preventive intervention? *Child and Adolescent Mental Health* **11**(3), 135–141.
- Rutter, M. and Mawhood, L. (1991). The long-term psychosocial sequelae of specific developmental disorders of speech and language. In Rutter, M. and Casaer, P. (eds.) *Biological Risk Factors for Psychosocial Disorders*, pp 233–259. New York: Cambridge University Press.

- Schweinhardt, L. J., Montie, J., Xiang, Z., et al. (2005). *Lifetime Effects: The High/Scope Perry Preschool Study through Age 40*. Ypsilanti, MI: High/Scope Press.
- Shonkoff, J. P. and Phillips, D. A. (2000). *From Neurons to Neighbourhoods: The Science of Early Childhood Development*. Washington, DC: Office of Educational Research and Improvement.
- Snow, C. and Van Hemel, S. B. (eds.) (2008). *Early Childhood Assessment: What, Why and How? Assessing National Research Council of the National Academies*. Washington, DC: The National Academies.
- Thorpe, K., Tayler, C., Greishaber, S., et al. (2004). *External Evaluation of the Preparing for School Trial*. Brisbane: Go Press, Queensland Government.
- Wise, S., da Silva, L., Webster, E., and Sanson, A. (2005). The efficacy of early childhood interventions. *Report Prepared for the Australian Government Department of Family and Community Services*. Canberra: Australian Institute of Family Studies.

Further Reading

- Sylva, K., Melhuish, E. C., Sammons, P., Siraj-Blatchford, I., and Taggart, B. (2004). The Effective Provision of Pre-School Education (EPPE) project: *Technical Paper 12 – The Final Report: Effective Pre-School Education*. London: DfES/Institute of Education, University of London.

EARLY CHILDHOOD EDUCATION AND CARE THEORY

Child Rearing and Early Education: Parents and Professionals: Theoretical and Historical Influences

H Penn, University of East London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Diverse Families

The traditional nuclear family of a married mother, father, and two or three children has been a model for much policymaking in the global North (Jenson, 2008), but it is a global experience that families are becoming more diverse. There are more parents choosing not to marry; more divorce, more second marriages, and more single parents; more role reversals between men and women, with men choosing to stay at home, and women choosing to work outside of the home; and there are more older mothers, with more educated women choosing to have families later or not at all. There is more mobility within and across countries (Bianchi *et al.*, 2003, Heymann, 2006). Even in the global North, caring arrangements for children may fluctuate considerably in accordance with family needs and societal circumstances. In the global South, family life has generally been much more communal and shared (DeLoache and Gottlieb, 2000), but the most significant change in the South is migration–urbanization – where men and women from the countryside migrate to towns and cities. Stripped of their rural support networks, they struggle to make a living and bring up children in the shanty towns that surround every city of the South (UNFPA, 2008). Discontinuity of caring arrangements in migrant or disrupted societies may be the norm and young children may move between people and places – even continents. Newman – writing about South Africa, for instance – comments that “commonly households are in a state of flux with the children moving between the relatives that are best able to care for them at particular times of their lives” (Newman, 2006: 1121).

Working Mothers

One of the biggest changes that has taken place in developed countries has been the rise of women in the workplace. In such countries, gender equity has become a major issue (OECD, 2006; EU, 2007, 2009). Maternity, paternity, and parental leave policies – along with the availability of early education and childcare provision – are crucial in

enabling women to try to reconcile their position in the workforce with familial obligations. The number of mothers in the workforce differs considerably across countries. This is, partly, for historical and cultural reasons, but, partly, also that if childcare and other benefits are not available, mothers are less likely to work.

One reason for encouraging mothers into the labor market in developed countries is that social security payments to single mothers and mothers in low-income households are regarded as a drain on the national economy, but once in work, such mothers contribute, instead, to tax revenues. Another is that the poorest households tend to be workless households, and encouraging mothers to work critically augments family income. Evidence from diverse countries suggests that governments have an interest in encouraging mothers to work outside of the home, and in providing childcare to facilitate their entry into the workforce (Kucera and Bauer 2001, Cleveland and Krashinsky, 2003).

However, a recent study in Australia, by Craig (2007) on mother's time-use, illustrates how a woman's work impacts on her time, and how – without compensatory social policies such as adequate childcare – most mothers who attempt to combine work and childrearing do so at considerable personal cost – lack of sleep, lack of leisure, and high levels of stress – whereas a father's employment status is likely to make little difference to the organization of his daily life. Meda (2001) using European data reaches similar conclusions. Children have a highly differential impact on mothers, fathers, and, indeed, other family carers. Role reversal is, of course, possible, with fathers shouldering the burden of childcare, but, typically, it is women whose economic and social position changes with childrearing (Paull, 2007). Although the word parent is commonly used in order to avoid gender stereotyping, it also obscures very real gender differences. Where possible, mother is used, in this article, in preference to parent for this reason.

As Heymann (2003) points out, these circumstances are exacerbated in the global South – where there are few state benefits, and employment is largely informal. In her four-country survey, 29% of working mothers in shanty towns left their children unattended or in the care of siblings.

It is both – for equity reasons, that is, to reconcile family life with employment, and educational rationales, that is, to improve children's school performance and outcomes – that the expansion has come about in early education and care services, and to the growth of a professional cadre of people who are concerned with children's cognitive, social, and emotional well-being.

The Intimacies of Caring

Caring, by itself, is an activity arising out of consanguineous relationships. Studies of caring – and of mothering, in particular – suggest that mothers focus on the material and emotional welfare of their children, that they pursue a different ethic of care from that of teachers. Mothers' knowledge of their children and relationship with them are not scientific and generalized, but anecdotal, subjective, *ad hoc*, and continuous – developing and changing over time within a specific context. It is also intensely physical. Care requires responsive ministration to physical as well as emotional and cognitive demands. It is this dirty physical work – feeding young children, keeping them clean and maintaining their bodily functions, and guarding them against danger – that has paradoxically led to care activities being devalued. Mothers have intense and intimate relationships with – and knowledge of – their children, especially when they are very young (Ebtehaj *et al.*, 2006, Noddings, 1984).

Caring is also a reciprocal relationship, in which the person caring and the person cared for offer each other feedback and support – albeit of differing kinds (Noddings, 2003). Mothers care for their children and expect love in return. The sociobiologist Sarah Hrdy argues that mothering has a strongly protective function, and that infants – in turn – reward their caretakers in a virtuous cycle of physical intimacy.

Maternity is inextricably inter-twined with physical sensations, and it is an infant's business, through grunts and coos, touches and smells, to make the most of Mother Nature's reward system, which conditions a woman to make this infant a top priority. Evolutionary logic is firmly on the side of mothers who enjoy the sensual side of mothering for its own sake. (Hrdy, 1999: 538)

Hrdy – drawing on her work with primates – also argues that the work of infant rearing is burdensome, too much for a single individual, and infants who are reared in multiple-female households – aunts, grannies, and cousins – in evolutionary terms, stand a better chance of survival.

Teachers and professionals, on the other hand, tend to hold more abstract, norm-related knowledge and expectations of children, unrelated to context, and without expectations of reciprocity or continuity. They are not

swayed by the storms of the moment, and they limit physical care (nursery education, formally, begins in almost all countries when children are no longer incontinent). They are more professional and detached. Teachers and other professionals operate in public institutionalized settings, mothers and carers in the private enclosed setting of the home. This boundary between public and private is a sensitive frontier which both professionals and parents cross with caution (see below). For example, in the 1970s, in the UK, the eminent psychologist Jerome Bruner argued in favor of the playgroup movement, and for mothers to be involved with their children:

when we are planning for the care of our young children we must also enable care in its widest sense to be given to those who have day to day responsibility, their parents and in particular their mothers. (Bruner, 1980: 231)

Countries with maternalist welfare traditions (i.e., holding beliefs about the natural role of women as mothers and the importance of mother-child attachments in early childhood) have encouraged mothers to stay at home with their young children and, until relatively recently, discouraged the use of alternative services except on a very part-time basis (Schweie and Willekens, 2009). Some commentators – who have attempted to extrapolate from neuroscientific studies of the brain – argue that the mother's role as a responsive carer is a key one in stimulating cognitive growth, although others are considerably more skeptical about the use of such studies in justifying particular approaches to parenting (Bruer, 1999, Thompson and Nelson, 2001).

In a classic study in the UK, Tizard and Hughes (2003) tape-recorded the conversations of 3-year-old girls at home with their mothers and compared their conversations with those they had at nursery school. They found that whatever the background of the child – working class or middle class – and however limited the conversation, the child's language at home tended to be richer, more extended, and drew on a much wider range of events, and showed more evidence of logical thinking than any talk that took place in the nursery school. This conversation happened in a leisurely, almost casual way in the course of everyday activities and in getting on with daily life – they were not scheduled play sessions, in which mothers focused exclusively on their children's talk. The researchers discuss the complementary roles of home and school and conclude that “the school's problem, as we see it, is to foster, harness and satisfy the curiosity which children show at home . . . instead of the emphasis on fostering play, on devising ingenious ways of using play materials, and on questioning children about their play, a higher emphasis would have to be given to widening the children's horizons, extending their general knowledge, and listening to them talk” (2003: 219). This finding – although powerfully evidenced – has been,

generally, ignored by professionals, possibly, because it is perceived to undermine their expertise in fostering children's cognitive and linguistic development. The study was originally carried out in the early 1980s, and it may also be the case that more recent changes in family life and conduct – the reliance on television (TV) and, in some areas, increasing drug use – have meant that the kind of intimate talk recorded by the researchers is less frequent.

Mothering Styles and Children's School Outcomes

The evidence suggests that mothers from poor homes do worse in preparing their children for the specific requirements of school – irrespective of ethnicity or any other variables. Ermisch (2008), an economist, has analyzed the Millennium cohort data in the UK, and concluded that differences in cognitive ability and behavioral development at age 3 are correlated with parental income. The lower the parental income, the poorer the scores are on standard cognitive and behavioral tests. Using a production-function framework, he argues that these differences can be partly explained by parenting styles – with low-income parents demonstrating less interest in cognitive-promoting activities such as reading. Sylva *et al.* (2007) reach a similar conclusion in their EPPE study (EPPE, the Effective Provision of Pre-School Education), arguing that parental style is a more powerful determinant of subsequent child outcomes than any educational intervention – although early-education interventions also do make a difference to outcomes. Feinstein *et al.* (2008) argue that a mother's educational level and social class are correlated with child outcomes. The differences in the willingness or capabilities of families to take advantage of educational opportunities exacerbate social class differences and limit actual equality of opportunity. Feinstein argues, for instance, that mothers from poor families have the least resources to deal with children who have learning problems, and as a result, their children are disproportionately disadvantaged (an estimated one in six children from across the social spectrum has some kind of disability or problem that may temporarily or permanently disrupt their learning). These three longitudinal studies come from the UK, but on a broader level, the Organization for Economic Cooperation and Development (OECD) argues that inequality and educational outcomes are linked; the more unequal a society, the more marginalized groups, and the higher the rates of poverty, the poorer the educational outcomes (OECD, 2008).

If the mother's parenting style – or, rather, the interface between expectations of school and home – plays a crucial role in determining a child's initial progress and subsequent readiness, it makes sense to focus on the home environment

and home-school relationships in the early years. The importance of the home environment – and, in particular, the vulnerability of children from poor and/or dysfunctional homes – have led some countries to invest heavily in home visiting and parental education programs – most notably, the Sure Start program in the UK. However, recent evidence summarized by Waldfogel (2004) suggests that home visiting and parent education do not significantly affect children's outcomes, although they may, in some cases, alter parental behavior. In addition, the investigation of such outcomes is problematic, since it is unclear what psychological mechanisms are at play in producing change when mothers receive support from home visitors or other support programs (Rutter, 2002, 2006). Poverty and inequity severely depress expectations both for mothers and their children (Bradshaw and Bennett, 2007), and poor parenting is likely to be multicausal, not merely a matter of improving mothers' emotional relationships with their children or assisting mothers to engage in cognitively stimulating games with them. In addition, the mothers most likely to be visited at home are those about whom there is concern – mostly from low-income households – whereas the professionals carrying out the visits are, by definition, middle class or at least trained and skilled in knowing about and interpreting the norms and expectations that their society has of children.

Training for Work with Young Children

There has been some discussion concerning the extent to which these different kinds of responsiveness to – and knowledge about – children are distinct or overlap (Noddings, 2003; Lynch *et al.*, 2007). Looking after one's own children requires a particular kind of contextual knowledge, responsiveness, and protectiveness, on a continuous 24-h basis, but these mothering skills cannot be easily generalized to other nonrelated children who may only be with the carer or educator for a relatively short period of time. Most countries choose to emphasize the importance of training for work with children, even for those women who are childminders or family day-carers working from their own home (OECD, 2006).

Training for workers in some – mainly English-speaking – countries reflects the different traditions of care (substitute mothering including physical care for those unable to care for their own children) and education (developing children's cognitive, linguistic, and other crucial skills in preparation for mainstream schooling). Generally, care training is at a low, mainly vocational level, and poorly paid. Education training tends to emphasize degree or postdegree standards. However, models of training also vary on a continuum between social pedagogic models which explore and emphasize the emotional relationships which underpin learning, and teacher models which

emphasize the knowledge content of what is to be learned. Overall, there is considerable variation between countries on the knowledge content, pedagogical methods and length of time, and the academic standard required for training to work with young children; in other words, there is no consensus with regard to what constitutes professionalism in the field. OECD (2006) has produced a comparative table of training which indicates the range of variation. Some countries have chosen to emphasize age-related training, with particular skills being required for particular age groups, and an emphasis on caring skills for younger children. In yet others – most notably in the famous groups of nurseries in Reggio Emilia, Northern Italy – the emphasis is on in-service training, on learning from, and reflecting continuously on daily practice with children.

New and Cochran (2007) review training initiatives across ten countries, including China, Japan, Brazil, and Sweden. Training, in these countries, is also firmly located within national policies and traditions, and there is considerable variation in the expectations of what professionals will provide, and the level to which they should be trained.

In all countries, working with young children is a profession which attracts women rather than men. Although there is considerable rhetoric about the importance of male role models – especially for children from female-headed households – hard evidence of the impact of gender inequity among professionals on young children or on their mothers or carers is lacking thus far.

Multidisciplinary Working

Welfare, health, and educational professionals, all have a role with respect to young children's upbringing and well-being, and – in some countries, most often those with relatively fragmented and low levels of services – multidisciplinary or multiprofessional working is considered to be the key to successful practice. The health and nutritional status of young children has always been regarded as an important aspect of child development, particularly in countries where child poverty is rife. The recent UNICEF-IRC (the United Nations Children's Fund-Innocenti Research Centre; 2008) report *The Child Care Transition: a League Table of ECEC in Economically Advanced Countries* gives ten benchmarks for a decent level of service, which include a spectrum of support for parents – parental leave arrangements, the need for a national integrated plan for young children, specified levels of service for children aged 3 and 4, antipoverty strategies, and health-outreach work. A country that successfully serves young children needs to have all these aspects of provision in place. Arguably, the most well-integrated schemes were those to be found in the ex-communist countries – where it was routine practice to

provide weekly health checks in kindergartens, exercise and rest regimes, and where nutritional intake was closely controlled.

Some of the current discourses about multiprofessional work take their cue from the activity theory of Engstrom (Anning and Edwards, 2006). Engstrom's theory – derived, in turn, from Vygotskian cultural theory – emphasizes client-focused cross-disciplinary collective problem solving, rather than co-ordination of specific areas of professional expertise, which is what is usually implied by multiprofessional work (Sanning *et al.*, 2009). In the case of early years, the argument has been whether to create a core professional whose training encompasses a wide range of disciplinary approaches – albeit at a relatively basic level – or whether to try to create projects where a range of professionals come together to share their expertise in dealing with the complex circumstances of childcare, early education, and family life (OECD, 2006).

Mothers and Professionals Working Together

Historically, parental involvement had a narrow meaning for professionals; a mother's (and occasionally father's) role is to be a loyal supporter of the activities of the nursery or school, or to carry out those tasks which the school deems especially suitable to foster learning, such as reading with the child or going to the library. Professionals, frequently, see themselves as in possession of knowledge and expertise in the field of child development and schooling, which mothers – especially poor mothers – lack. From this point of view, the challenge for professionals is to impart such knowledge and skill to mothers – hence parenting classes, early stimulation, and many other special lessons for mothers. Communication is one-way – from the school or setting to the mother – rather than mutual.

Bouve (2007) discusses moves in France to increase family involvement in services, and notes that the process has been rather different for working-class and upper-class families. For the former, the dominant ethos has, historically, been social control and normalization, whereas better-off socioeconomic groups have been able to exert a stronger influence on the pedagogy and other aspects of the *ecoles maternelle* system. Penn (2004) also notes a similar historical class division in the UK. The OECD report *Starting Strong* (2006) stresses that social equity and a higher standard of service education and care services are more likely to be achieved where the priority is universal services for all children, rather than targeted services for certain classes of children and their families. The take-up of universal nursery education in most countries is very high, but mothers are much more ambivalent about using targeted services which they, often, see as stigmatizing or patronizing.

However, others have cast the challenge of parent–professional relationships rather differently. Given the pressures mothers – especially working mothers – encounter, the challenge for early-childhood services is how to support mothers (and fathers and other carers), especially those living in vulnerable circumstances, by acknowledging their rights as mothers/parents and their intimate contribution as carers. Dahlberg and Moss (2005) have argued that services should ideally recognize the mother's and father's rights within services; their right to be informed, to comment, and to participate in key decisions concerning their child. This echoes Dewey's idea of the school as a democratic institution where citizenship is actively learned and practiced (Pring, 2007). From this point of view, early-childhood services are democratic spaces trying to forge mutual discussion and decision-making between mothers, fathers, and professionals – and, indeed, with children themselves – with regard to the range and quality of the service they use or practice in, or attend, or rely on. This ideal is a demanding one – although, for example, the services in Reggio Emilia and other towns in Northern Italy do manage to operate on this basis (Bloomer and Cohen, 2008). Many early-childhood services in the Nordic countries have also devolved and have democratic decision-making at the level of the nursery and view mothers and fathers as rightful partners in the joint enterprise of care and education of young children (OECD, 2006).

The Tensions of the Marketplace

In market-driven early education and care services, the notion of parental involvement has become conflated with parental choice. In a situation where early education and care is purchased within the market – as is increasingly the case in neoliberal economies (Penn, 2009) – parents can choose between one kind of care rather than another, but they are essentially customers making a purchase between satisfactory or less satisfactory options, rather than partners who might help shape the service they use. Vincent and Ball (2006) have demonstrated how such choices are also class driven. Middle-class parents have the resources – both financial and social – to make strategic choices of education and care, whereas working-class parents are not only financially constrained, but make choices on the basis of the local and familiar. The result is that, in a market-driven situation – where parents are funded directly through tax credits or vouchers in order to choose early education and care (as opposed to supply-side funding where services are directly funded) – social stratification becomes amplified. The rich go in search of what they consider to be the best, whereas the poor stay put and take whatever is on offer locally.

In the global South – where most early education and care provision is provided by local entrepreneurs, and relies entirely on parental fees – cost is directly related to quality; the poorer the mother, the less that can be afforded. Those who can pay get relatively good services, but those who cannot do so get the poorest quality services – thus, increasing inequity (Penn, 2008).

The financial transactions over children between the parents and the teachers or child-carers who look after them, are frequently characterized by fracture and dissension. Staff are ambivalent about working mothers and feel exploited, while mothers try to prioritize an untroubled relationship with staff – even when they are unhappy with what seems to be on offer. On the other hand, mothers may also seem to devalue the work of those in charge of their children (Vincent and Ball, 2006). For example, the UK practitioner magazine *Nursery World* increasingly carries articles concerning disputes over fees between nursery owners and parents (Penn, forthcoming).

In a marketplace, early education and childcare are products which are bought; and the child will lose his or her place if the mother and father are unable or unwilling to continue with the purchase.

Conclusion

This article has very briefly considered some of the theories and debates on the topic of child rearing and early education. Conceptual distinctions have been made between the care offered by mothers and other carers, and that offered by professionals; however, in rapidly changing family environments, new accommodations need to be made. This is especially true with respect to working mothers. The growth of a substantial private market in many countries has been a further complication. Although the debate on mothers, fathers, and professionals, has mainly derived from the global North or rich countries, many of the considerations are also pertinent for the global South.

See also: Early Childhood Care and Education: The Family, The Market, and The State; Gender Issues in Early Childhood Education and Care; Investing in Early Childhood Education and Care: The Economic Case; Literacies in Early Childhood: The Preschool Period; Teaching in Early Childhood Centers Instructional Methods and Child Outcomes.

Bibliography

- Anning, A. and Edwards, A. (2006). *Promoting Learning from Birth to Five: Developing Professional Practice in the Pre-School*, 2nd edn. Buckingham: Open University Press.

- Bianchi, S., Casper, L., and King, R. B. (eds.) (2003). *Work, Family, Health, and Well-being*. National Institute of Child Health and Human Development (U.S.). London: Routledge.
- Bloemer, K. and Cohen, B. (2008). *Young Children in Charge*. Edinburgh: Children in Scotland.
- Bouve, C. (2007). Family involvement in France. In New, R. and Cochran, M. (eds.) *Early Childhood Education; An International Encyclopaedia*, pp 1072–1076. New York: Greenwood Press.
- Bradshaw, J. and Bennett, F. (2007). *UK: Tackling Child Poverty and Promoting the Social Inclusion of Children: A Study of National Policies*, European Commission, DG Employment, Social Affairs and Equal Opportunities.
- Bruer, J. (1999). *The Myth of the First Three Years*. New York: Free Press.
- Bruner, J. (1980). *Under Five in Britain*. London: Grant McIntyre.
- Commission, E. U. (2007). *Opportunities, Access and Solidarity for 21st Century Europe*, 726 final. Brussels: COM.
- Craig, L. (2007). *Contemporary Motherhood: The Impact of Children on Adult Time*. Aldershot: Ashgate Publishing.
- Dahlberg, G. and Moss, P. (2005). *Ethics and Politics in Early Education*. London: Routledge.
- DeLoache, J. and Gottlieb, A. (eds.) (2000). *A World of Babies: Imagined Childcare Guides for Seven Societies*. Cambridge: Cambridge University Press.
- Ebtehaj, F., Lindley, B., and Richards, M. (ed.) (2006). *Kinship Matters*. Oxford: Hart Publishing.
- Ermisch, J. (2008). Origins of social immobility and inequality: Parenting and early child development. *National Institute Economic Review* **205**, 62–71.
- EU (2009). *Early Childhood Education and Care: Key Lessons from Research for Policy Makers*. Brussels: European Commission/ NESSE June 2009. http://www.nesse.fr/nesse/nesse_top/tasks (accessed September 2009).
- Feinstein, L., Duckworth, K., and Sabato, R. (2008). *Education and the Family: Passing Success across the Generations*. London: Routledge.
- Heymann, J. (2003). *The Role of ECCE in Ensuring Equal Opportunity, UNESCO Policy Brief No 18*. Paris: UNESCO.
- Heymann, J. (2006). *Forgotten Families: Ending the Growing Crisis Confronting Children and Working Parents in the Global Economy*. Oxford: Oxford University Press.
- Hrdy, S. (1999). *Mother Nature: Maternal Instincts and How They Shape the Human Species*. New York: Ballantine Books.
- Jenson, J. (2008). Writing women out, folding gender in: The European union “Modernises” social policy. *Social Politics* **15**(2), 131–153.
- Kucera, K. M. and Bauer, T. (2001). *Costs and benefits of Child Care Services in Switzerland-Empirical Findings from Zurich*, 257pp (Cited in OECD (2006) *Starting Strong II*). Paris: OECD.
- Lynch, K., Lyons, M., and Cantillon, S. (2007). Breaking silence: Educating citizens for love, care and solidarity. *International Studies in Sociology of Education* **55**(3), 550–570.
- Meda, D. (2001). *Le Temps des femmes; pour un nouveau partage des roles*. Paris: Champs/Flammarion.
- New, R. and Cochran, R. (2007). *Early Childhood Education: An International Encyclopaedia*, vol. IV. New York: Greenwood Press.
- Newman, M. (2006). Family involvement in early childhood development in south africa. In Cochran, M. and New, R. (eds.) *Encyclopaedia of Education*, vol. 4, pp 1211–1215. Westport, CT: Praeger.
- Noddings, N. (1984). *Caring: A Feminine Approach to Ethics and Moral Education*. Berkeley, CA: University of California Press.
- Noddings, N. (2003). *Happiness and Education*. Cambridge: Cambridge University Press.
- OECD (2006). *Starting Strong II: Early Childhood Education and Care*. Paris: OECD.
- OECD (2008). *Growing Unequal: Income Distribution and Poverty in OECD Countries*. Paris: OECD.
- Paull, G. (2008). Children and women’s hours of work. *Economic Journal* **118**(526).
- Penn, H. (forthcoming). Gambling on the Market: The role of for-profit provision in early childhood education and care. *Journal of Early Childhood*.
- Penn, H. (2004). Round and round the mulberry bush: The balance of public and private in early education and childcare in the twentieth century. In Aldrich, R. (ed.) *Public or Private Education: Lessons from History*, pp 67–92. London: Woburn Press.
- Penn, H. (2008). Working on the impossible: Early childhood policies in Namibia. *Childhood* **15**(3), 378–398.
- Penn, H. (2009). International perspectives on quality in mixed economies of childcare. *National Institute Economic Review* **207**, 83–89.
- Pring, R. (2007). *John Dewey: Continuum Library of Educational Thought*. London: Continuum.
- Rutter, M. (2002). Nature, nurture and development: From evangelism through science towards policy and practice. *Child Development* **73**(3), 1–22.
- Rutter, M. (2006). Is sure start an effective preventative intervention? *Child and Adolescent Mental Health* **11**(3), 135–141.
- Sanning, A., Daniels, H., and Gutierrez, K. (eds.) (2009). *Learning and Expanding with Activity Theory*. Cambridge: Cambridge University Press.
- Schweie, K. and Willeken, H. (eds.) (2009). *Childcare and Preschool Development in Europe: Institutional Perspectives*. London: Palgrave MacMillan.
- Sylva, K., Stein, A., Leach, P., Barnes, J., Malmberg, L. -E., and FCCC, team (2007). Family and child factors related to the use of infant care: An English study. *Early Childhood Research Quarterly* **22**, 118–136.
- Tizard, B. and Hughes, M. (2003). *Young Children Learning: Inside Children’s Worlds*, 2nd edn. Oxford: Blackwell.
- Thompson, R. and Nelson, C. (2001). Developmental science and the media. *American Psychologist* **56**, 5–15.
- UNICEF-IRC (2008). *Report: The Child Care Transition: A League Table of ECEC in Economically Advanced Countries. Report Card No 8*. Florence: UNICEF-IRC.
- United Nations Population Fund (UNFPA) (2008). *State of the World’s Population 2008: Reaching Common Ground: Reconciling Culture, Gender and Human Rights*. New York: UNFPA.
- Vincent, C. and Ball, S. (2006). *Childcare, Choice and Class Practices*. London: Routledge.
- Waldfoegel, J. (2004). *Social Mobility, Life Chances and the Early Years. CASE Paper 88*. London: Centre for the Analysis of Social Inclusion.

Relevant Websites

<http://www.oecd.org> – OECD (Organization for European Economic Co-operation).

EARLY CHILDHOOD EDUCATION AND CARE WORKFORCE

Contents

Early Childhood Education and Care Professional Profiles in the United Kingdom: Multiagency Teamwork and Transdisciplinary Challenges

Early Childhood Education and Care Staff Support: Mentoring and Professional Development in Finland

Early Childhood Education and Care Professional Profiles in the United Kingdom: Multiagency Teamwork and Transdisciplinary Challenges

A Anning, University of Leeds, Leeds, UK

P Calder, London South Bank University, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Policy Context

There are different legislative frameworks and qualifications for Early Childhood Education and Care (ECEC) practitioners within the four countries of the United Kingdom. In 1999, devolution of powers to the national governments of Scotland, Wales, and Northern Ireland resulted in complex layers of legislation. For example, in Scotland, in the area of early years and child-care policy, education, and health are devolved. Yet employment and the Department for Work and Pensions, including the child benefits system, as well as maternity and paternity leave, remain the remit of the UK government. Childcare Tax Credits, designed to promote a consumer demand-led approach to developing child care, are UK-wide. Aspects of workforce qualifications are both UK-wide and specific to the four countries, depending on the workforce and level of qualification. The complex legislative arrangements result in complex administrative processes in the ECEC field in the United Kingdom.

However, the context of redefining ECEC professional profiles in the four nations is similar. There is common inheritance of occupational segregation, low pay and status, and a predominantly female workforce. Social services have had responsibility for day care, child-minding, and child protection. Education has had responsibility for nursery (for 3-year-olds) and reception classes (for 4-year-olds) in primary schools (catering to 3–11-year-olds) and a few nursery schools (catering to 3–4-year-olds). The voluntary sector has provided preschool playgroups and many family support services. Health authorities have had responsibility for universal, statutory home visits, health checks to newborns and toddlers (through midwives working from hospitals), and health visitors (working from general

practitioner doctors' surgeries). Responsibility for play services has been with the recreation/leisure services. Each agency has had distinct training routes, qualifications, pay and conditions, knowledge and skills bases, beliefs, values, and languages. As a result, professionals working for these agencies hold different beliefs about the causes of problems presented by families and they have prescribed different treatments to put them right.

Since 1997 'Education, Education, Education' and 'Joined-up Thinking' have been mantras driving the Labour government reforms of children's services. However, within as well as outside government and its agencies, interpretations of these slogans have been multiple and contradictory. There have been disputes over the purposes and definition of early childhood education: Is it preparation for life, for school, for employment, or a phase in its own right? (Dahlberg *et al.*, 1999). There have been disputes as to what should be joined up and how: How infrastructures and funding streams should be redefined? How different pay and conditions of service reconciled? Also, How settings should be reshaped?

Overarching the rhetorical aims has been an ideology of public-private initiatives. These are to fund new developments to replace public sector services, and to reorientate the work of employees, including the use of short-term contracts, commissioning out, and employment of corporate rather than specialist agencies to manage the strategic planning and delivery of services (Ball, 2007).

Policy

In the United Kingdom, national policy aspirations have been to address poverty and inequalities by providing

high-quality, universal, integrated services for families. The justifications for investing in early years services were on social justice grounds (redressing the widening gap in life chances between the rich and poor) and for economic gain (money spent at the early stages of a child's development is likely to enhance life trajectories in health and education, enhance employment outcomes, and, ultimately, reduce government spending).

Many initiatives were launched to address these aspirations. In the late 1990s, Early Excellence Centres (EECs) were set up as one-stop shops to model services for the education and care of children under four, the training and employment of their parents, the health and welfare of the whole family, and support for parenting. From 1998 (eventually more than 500) 'sure start local programs' (SSLPs) were established as a flagship antipov-erty program in the most disadvantaged areas of England. They provided outreach and home visiting, support for families with preschool-age children, good quality play/early learning/childcare, primary and community health-care, and support for families with young children with additional needs and disabilities. The mode of delivery of SSLP services was through multiagency teamwork. In 2004, the 'ten-year strategy' made a commitment to providing 2500 children's centers by 2008 and 3500 by 2010. In 2006, partly in response to disappointing early findings on impact from the National Evaluation of Sure Start (NESS, 2005), SSLPs were rebranded as Sure Start Children's Centres. Further, primary schools are expected to take on a broader role in working with families with children aged 4–11 years as they expand into extended schools. Extended schools will provide before- and after-school care, leisure activities, family support services (including parenting advice and training and signposting to employment opportunities), and provision for children with additional needs and disabilities. Social work and health professionals are required to work closely with those delivering education and care services.

Children's trusts within all English local authorities (the bodies responsible for commissioning, coordinating, and monitoring the delivery of all services for local communities) are expected to ensure a suitably diverse range of providers of children's services. They must commission services of uniformly high quality equitably from the private, voluntary, and maintained sectors, in line with parental choice (Her Majesty's Government, 2006). In sum, the policy context has generated major structural reform of the ECEC field, impacting the professional profiles of all those engaged.

Legislation to Support Policy Change

Reforms in education and care services were bound together by a seminal Children Bill and Children Act in 2004 which enacted principles outlined in the Green Papers

Every Child Matters (DfES, 2003), the Ten-Year Strategy (Her Majesty's Treasury, 2004), and *Every Child Matters: The Next Steps* (DfES, 2004). There are five key principles embedded in the outcomes specified in the legislation: being healthy; protection from harm and neglect; enjoying and achieving; making a positive contribution; and economic well-being. The principles are embedded in the Office for Standards in Education (Ofsted) affirming the inspection of all settings delivering education and care for children from birth to school-leaving age. In 2007, an ambitious children's plan was launched by the Department for Children, Schools and Families, with goals to improve schools and the way families and children are supported.

Education

For children aged 3 and 4 years, there is an entitlement to two-and-a-half (soon to be three) hours of free pre-school education a day for 38 weeks a year. There is currently a pilot scheme offering part-time, free daycare for 2-year-olds from families in poverty. The entitlement may be delivered by a range of providers. The Early Years Foundation Stage, which was introduced in 2008, replaces earlier legislation, and provides a statutory curriculum framework for children from birth to 5 in early years provision.

Under the duties of the Childcare Act 2006, practitioners in England are required to complete a Foundation Stage profile for each child, recording the child's progression in physical, intellectual, emotional, and social development against 13 assessment scales. By the start of compulsory schooling (at age 5), children are expected to attain a set minimum score in the personal/social/emotional and the communication/language/literacy scales. Local authorities are required to return data to the Department for Children, Schools and Families at institutional level. There is heightened concern by staff about accountability for the results and increased scrutiny of ECEC professional profiles.

Childcare

Reforms have included a National Childcare Strategy (in 1998), a Ten-Year Strategy for Childcare in 2005 (DfES, 2005), and the Childcare Act in 2006. (These reforms have enabled the expansion of child-care services, mainly through the private sector, to ensure that parents (in particular those on benefits) can return to work.) The Neighbourhood Nursery Initiative was launched in 2000 to increase the supply of day care in poor neighborhoods with a requirement that their funding should be commissioned from the private sector. In 2007, in England, 80% of child care was provided by the private sector. The globalization of corporate child care has resulted in new risk: How will the imperatives of social justice be reconciled with those of making profit? (See Williams (2004), for a discussion of the ethics of social care.) There is concern that

recruitment and retention of more highly educated and better-trained professional staff will be incompatible with a private system which aims to make profits and where parents are usually responsible for paying a high proportion of the costs.

Either the government will have to subsidize equitable salaries between comparable professionals, such as Early Years Professionals (EYPs), and qualified teachers, or private providers go out of business as parents will not be able to afford the fees or staff careers and salaries will be dictated by the settings in which they work. Those with guaranteed salaries and career structures, such as teachers in maintained schools, will be reluctant to retrain to work in new (private?) multiprofessional settings where there are no such protections.

Integrating Early Years Care and Education

For a brief period (1997–2003), there appeared to be a possibility of developing a new concept of integrated care and education with a new working-framework, structures, and curriculum. Initiatives such as EECs and SSLPs had been seen as moving toward such integration. The extension of children's centers from targeted disadvantaged areas to all areas, through the ten-year strategy, appeared to recognize that providing integrated early care and education was good for all families. Aspirations were for "high quality provision with a highly skilled childcare and early years workforce, among the best in the world" (Her Majesty's Treasury, 2004: 1). However, the implementation of these ideas indicates that since 2004 forces driving care and education apart have strengthened. Even though new structures were set up with the purpose of integrating children's services, schools and teachers (including nursery teachers) were left outside them.

When there is emphasis on children of all ages and the integration of all children's services, the strategic scale of the work may be so great that it risks destabilizing and overlooking or alienating core subgroups to the extent that more positive outcomes are compromised. At the core, professional work in integrated services must translate into practical actions that advance the outcomes of children and families.

Multiagency and Transdisciplinary Teams Provision and Settings

Partly by chance and partly because of government commitment to offering choice to consumers, policy reforms in the United Kingdom have ensured service users have a mixed menu of provision. Three types of settings exist: private-for-profit provision (private schools with day care and nursery education attached, day care, and childminders), voluntary sector provision (preschools and family

centers), and publicly funded (maintained) nursery schools, Sure Start Children's Centres, and early years units in primary schools. The settings differ in whether they offer full-day care or sessional care, and whether they provide free educational entitlement.

Many young children attend several types of provision before they enter formal school. This has consequences, given the variable quality of experiences provided in preschool services. Very young children may encounter inconsistencies in carers, frequent changes in routines and familiar places, and exposure to changing cultures and values in settings. Parents and carers have to make complex arrangements for children, taking part for varying durations, in different types of settings.

The nature of the setting has consequences for the type and qualifications of staffing. A significant driver of quality is the level of qualifications of the workforce and yet the level of qualifications in the ECEC workforce is still low. For example, in maintained nurseries, 80% of the staff were found to be qualified with A levels, or National Vocational Qualification 3 (NVQ3) or higher, while for voluntary and private sector nurseries levels fell to 51% and 46%, respectively (Coxon *et al.*, 2007).

Workforce Reform

An aim coordinated by the Children's Workforce Development Council (CWDC) is to upgrade the workforce qualifications so that all supervisors of full day-care provision hold a relevant level 3 qualification and at least 50% of the remaining child-care staff are qualified to level 2. In recognition of the complexity of managing integrated services settings, a master's level qualification (the National Professional Qualification for Integrated Centre Leadership) for present and potential managers was made available. Other initiatives include a 2-year foundation degree offered to the existing early years practitioners who have limited formal training with recognition of excellence as senior practitioner serving as an incentive for such staff to study. As a further encouragement to career progression, an additional 2-year route was offered to gain qualified teacher status. Further a graduate professional status aims to create an EYP for work in every children's center and in every full-day-care setting to provide an integrating role as well as taking responsibility for planning, assessing, and recording children's progress through the Early Years Foundation Stage. The training of teachers remains separate. In practice, teachers working in integrated settings, but trained to work in schools, have found the transition into a different working culture difficult. For many, the birth of EYPs raised the specter of a different kind of teacher – one for example where the training, unlike teacher training, is assessed through attaining competency standards which do not include being supervised while interacting directly with children (CWDC, 2006) and can result in lower paid staff

for children under the age of 5 in English preschool settings. There are currently fierce debates in England at central and local government levels about applying a common funding formula to all settings providing education and child care.

In England, development of an Integrated Qualifications Framework is emerging, described as a climbing frame of qualifications, which will allow staff working with families and young children to move within and across occupations. The aim is for clear career pathways and qualifications not only for a chosen job, but also for routes to move, for example, from being a nursery nurse in a preschool setting to becoming a care assistant in a children's home – without having to start all over again.

Working in Multiagency and Transdisciplinary Teams

'Joined-up Thinking' that has underpinned workforce reform in the last decade has left questions. There is confusion about the nature of teams working together in integrated service settings. Teams may be labeled as multiagency with the team made up of some specialists and some generalists. Within each agency team there may be workers from several disciplines (e.g., in a hospital-based team representation from psychology, medicine, and speech and occupational therapy). Each layer within agencies and disciplines provides differing levels of challenge to those charged with working together.

There is evidence of four common dilemmas in making multiagency teams work (e.g., Anning *et al.*, 2006; Anning *et al.*, 2007; Siraj-Blatchford *et al.*, 2007). The first relates to the structural features of the team. For example, some professionals are seconded to multiagency teams for short periods or on part-time contracts. Those appointed as core members of a team may be on time-limited contracts. Where a professional feels peripheral due to their working hours or contracts, full commitment to a team may be compromised. Arrangements for professional appraisal and supervision may have remained with the agency from which staff were released. There can be tensions between the priorities of their lead agency and those of the multiagency team. For example, a health visitor may be expected to take on a more active role in offering parenting support in her multiagency team, to the point where she feels her attention is diverted from her mainstream agency role of monitoring children's health and well-being. The tensions between core and peripheral team members are often exemplified in the allocation of space and resources for staff to carry out activities.

The second dilemma is ideological. A social worker may have different constructs of children and families, ascribing blame for problems with children's behavior to society, while a teacher in the same setting ascribes blame to parents. The cultures of medicine and preventative health workers may collide; the medical model focuses

on diagnosis of the medical condition and applying treatments in line with scientific evidence, while the health worker may see the patient within the social network of a particular culture, community, and family. For professionals expected to work in a joined-up way, considerable time may be spent clarifying the distinct beliefs and values of team members. It can be challenging for many (particularly those in lower-status occupations) to have to articulate why they do things in certain ways and how this relates to what they know. They may never have been asked to do so before. Opening up to possibilities, especially when they are cloaked in professional jargon, involves a real effort and can destabilize a professional sense of self. Yet, exploring the knowledge and beliefs of those from other agencies and disciplines can be rewarding, leading to the generation of new and more expansive forms of knowledge both for individuals and the whole team. This is a benefit to service users.

A third set of dilemmas are procedural, related to the nitty-gritty of daily activities of delivering services. The pace of change in children's services may cause stress. The distinct approaches of staff from different agencies to documentation and record keeping, and related attitudes to confidentiality of information can exacerbate tension. Different work cultures can make workers particularly anxious about sharing information. Yet, one of the pivotal aspects of making interagency collaboration work is sharing information. At one level, this is about professional skills and expertise and the knowledge base underpinning actions, and, at another, this concerns the specifics of individual cases for whom the team is responsible. A further concern is how specialist and generalist staff are deployed. For those with specialist (and by implication higher-status) training and qualifications, taking on generic team functions may feel like being de-skilled. An example is when a teacher in a children's center is scheduled to take on routine, caring aspects of work with children, whereas their perceived primary role is that of an educator. Carers may feel stressed about expectations that they should teach. Such perceptions underline the difficulties of trying to introduce a new role.

Specialists may question the competence of generalist workers to take on aspects of work they perceive to be the preserve of specialists. An example is when a speech therapist is charged with training mainstream nursery nurses or teaching assistants to support children with language delay or impairment. In this female-dominated workforce, there is also discomfort with confronting disagreements and conflicts. Where individuals and teams are not guided by skilled managers in how to deal with conflict, teams continue to work in parallel with each other, each defending their own disciplinary and agency boundaries inherited from their past working lives, thereby not moving the promised gains of genuine multiagency teamwork.

The fourth set of dilemmas relates to building inter-professional relationships over time. The speed at which

multiagency teams is formed and expected to operate, the rapid shifts in professionals' roles and responsibilities (often without training) can leave individuals feeling that their professional identity is at risk and their job satisfaction eroded. A professionals' sense of self, of 'Who I am', is deeply embedded in their work – 'What I do'. Leaders and managers need to support professionals when they feel destabilized by workforce reforms and be aware of tensions arising from the differing pay and working conditions of their team members. It is not unusual, for example, for a teacher in an integrated service setting to be earning more than her (social-work-trained) center manager and yet still be entitled to the long holidays and short working days afforded to teachers.

Professionals may need to retain links with their specialist agencies, in order to ensure a career trajectory in the home agency. This means that managers have to make judgments about reconciling the training needs of a whole multiagency team with those of individual members of the team.

Implications

Proposals for reforms in service provision and workforce qualifications for ECEC are visionary and show commitment to improving the life chances of young children and providing day care for parents wanting to return to work.

As we have argued, the case study of ECEC reform in the United Kingdom indicates that there is a risk that the Every Child Matters agenda and, more recently, the Children's Plan have been over ambitious. For example, the commissioning of services by Children's Trusts to ensure a diverse range of providers is a challenge when day care is largely in the hands of the private sector and education in the hands of the maintained (public) sector; and to afford to staff and resource its services to deliver the Early Years Foundation Stage. There is growing evidence of tensions between running a for-profit service and maintaining high-quality provision of day care, and of the challenges of regulating services offered by independent companies. In England, the maintained sector, offering most of preschool education services, is better resourced and, in particular, has more highly qualified and expensive staff than the day-care sector. Voluntary, private, and maintained sectors are subject to rigorous monitoring, regulation, and inspection. But how equitable are services for families using different types of provision?

The creation in England of a new EYP raises unanswered questions about how they will be integrated into the current workforce. For example, how will an EYP equate to an Early Years Teacher? At a more visionary

level, how will a new breed of ECEC professional drawing from the sectors of health, education, care, and family support learn to work together for the benefit of families with young children if there is little clarity about training at initial service level or retraining at in-service level? Who will fund the new qualifications and enhanced salaries?

See also: Integrated Early Childhood Education and Care Services – Care, Upbringing, Education and Health.

Bibliography

- Anning, A., Cottrell, D., Frost, N., Green, J., and Robinson, M. (2006). *Developing Multi-Professional Teamwork for Integrated Children's Services*. Maidenhead: Open University Press.
- Anning, A., Stuart, J., Nicholls, M., Goldthorpe, J., and Morley, A. (2007). Understanding variations in sure start local programmes. *National Evaluation of Sure Start Research Report NESS/2007/FR/024*. London: HMSO/Nottingham: DfES Publications. <http://www.ness.bbk.ac.uk> (accessed September 2009).
- Ball, S. (2007). *Education plc: Private Sector Participation in Public Sector Education*. London: Routledge.
- Children's Workforce Development Council (CWDC) (2006). *Early Years Professional National Standards*. http://www.cwdcouncil.org.uk/pdf/Early%20Years/EYP_National_Standards_July_2006.pdf (accessed September 2009).
- Coxon, K., Sigala, M., and Smith, T. (2007). 102 Neighbourhood nurseries on the ground: The NNI implementation study. In *NNI Research Team (2007) National Evaluation of the Neighbourhood Nurseries Initiative: Integrated Report Research Report SSU/2007/FR/024 HMSO*. Nottingham: DFES Publications.
- Dahlberg, G., Moss, P., and Pence, A. (1999). *Beyond Quality in Early Childhood Education and Care: Postmodern Perspectives*. London: Falmer.
- DfES (Department for Education and Skills) (2003). *Every Child Matters*. London: HMSO.
- DfES (Department for Education and Skills) (2004). *Every Child Matters: The Next Steps*. London: HMSO.
- DfES Department for Education and Skills (2005). *Ten Year Strategy for Childcare*. London: HMSO.
- Her Majesty's Government (2006). *Childcare Act 2006*. London: HMSO.
- Her Majesty's Treasury (2004). *Choice for Parents, the Best Start for Children: A Ten Year Strategy for Childcare*. London: HMSO.
- National Evaluation of Sure Start (NESS) (2005). *Early Impacts of Sure Start Local Programmes on Children and Families*. Nottingham: DfES.
- Siraj-Blatchford, I., Clarke, K., and Needham, M. (2007). *The Team Around The Child: Multi-Agency Working in the Early Years*. Stoke on Trent: Trentham Books.
- Treasury (2004). *Choice for parents, the best start for children: A ten year strategy for childcare*. http://www.hm-treasury.gov.uk/d/pbr04childcare_480upd050105.pdf.
- Williams, F. (2004). *Rethinking Families*. London: Calouste Gulbenkian Foundation.

Relevant Websites

- www.cwdcouncil.org.uk – Children's Workforce Development Council (CWDC).
- www.everychildmatters.gov.uk – Every Child Matters: Department for Children, Schools, and Families.

Early Childhood Education and Care Staff Support: Mentoring and Professional Development in Finland

P Lindberg, National Institute for Health and Welfare, Helsinki, Finland

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Curriculum in ECEC – A learning and teaching plan including developmental aims, pedagogical processes, and contents. In the Finnish context, a framework or guidelines for implementing ECEC based on the society's values; enhances child's rights and well-being.

ECEC (Early childhood education and care) – A holistic system that covers both the day-care arrangements offered to families and the goal-oriented early childhood education open to children. Care and education are brought together in an integrated way also in the daily pedagogical activities.

E-learning / working environments – Learning environments that use specific software to provide web-based tools to aid learning processes. Also referred to as virtual learning environments (VLE).

E-mentoring – Electronic mentoring that is carried out by using electronic forms of communication, such as emails or media using Internet connection.

Group mentoring – Mentoring where the mentor is a learning leader of a team or a group. In a group mentoring situation everyone can benefit from the diverse knowledge and experience of the learning leader.

ICT – Information and communication technologies; may also be referred to as information technologies (IT).

Open source – Open source is a development method for software that enables access to the source code and free distribution.

Peer mentoring – A form of group mentoring where each participant is equal and benefits from the mentoring situation. Participants in peer mentoring complement each others' knowledge and skills in a reciprocal dialogue.

Reflection – Critical way of observing and evaluating actions in the working community. It helps to develop practices by questioning current ways of working. Reflection makes values, tacit knowledge, and experiences more visible.

The Need for Staff Support

Well-educated staff with ongoing access to professional development opportunities have been raised as a core element in the provision of high-quality early-childhood education and care (ECEC) services (OECD, 2006). The foundation for professional knowledge is laid during the initial education of the staff. However, policies focused only on increasing teachers' education level cannot suffice for improving quality. To keep the professional skills of the staff in balance with the changing context and situation of the society, professional development should be a continuous and intentional process. This process should be formed by a number of relevant support measures and clear structures for staff development.

The required staff competencies are numerous. Staff should have knowledge of what is needed to foster the well-being, development, and learning of children and be aware of the rights and potentialities of young children. Staff should also be skilled for operating in partnerships with a variety of parents from a number of social, cultural, and educational backgrounds, and with other professionals in schools, healthcare, and social services. Many children have special needs. Furthermore, families may have problems that require either support from the staff, or at least advice on where to get support. In the information-based global world, ECEC staff also need to be capable of using new technologies – such as computers and other digital equipment. Furthermore, staff in ECEC should have reflective skills and they should be competent in team work.

The Organization for Economic Cooperation and Development (OECD) (2006) demonstrates that ECEC staff, in many countries, face several practical challenges in accessing in-service training. There are too few professional development opportunities available – both in the public sector and in parts of the private sector. Considering this, it is important to develop support models that can be brought closer to the ECEC practice, for example, by utilizing modern technologies, or by applying methods where the expertise of the staff and other regionally available sources can be employed. Moreover, the everyday working environment of the ECEC staff should be used in the development of support measures because it forms an important context for building expertise.

Mentoring provides a functional means for holistic professional learning, taking into consideration the different developmental needs of staff. Mentoring can also have an impact on the development of ECEC quality by providing an instrument for knowledge transfer, translation, and reflection. The initial education of the ECEC staff provides a theoretical basis for the work, while mentoring may augment the reflective element that affords a means to deepen the understanding of the theories-in-use and, thus, improve the intentionality of the actions taken.

This article adopts a perspective of mentoring as staff support in ECEC. A brief introduction into mentoring, in general, is given, but the focus is on the development of working communities, and e-mentoring. To provide practical illustration, two examples of how different mentoring models have been used in professional development in Finland are presented.

Mentoring in Professional Development

Mentoring is a versatile method used for professional development in a variety of areas. In general, mentors are seen as more experienced and competent people who – in different contexts – can act as facilitators for the less experienced. In this article, only work-related mentoring – in the context of education – is discussed, although mentoring has been used also for other purposes, such as supporting young people to realise their potential in life.

Mentoring has been used extensively in the rapidly changing business world – the environment in which many of the mentoring theories were originally developed. In the education sector, mentoring is widely used – especially as a method for teacher students and for beginning teachers in their early career development (Scherer, 1999). Observation of the Internet pages of a number of universities and other teacher education providers demonstrates well the contemporary relevance of this support method. Staff-support methods using different forms of mentoring have also been applied in ECEC in many countries. Mentoring has expanded from the more traditional one-to-one mentoring processes into a variety of models for deepening the expertise of entire working communities. Mentoring provides a tool for employing local experts as mentors in the professional development processes (Lindberg, 2007).

What is Mentoring?

Over time, there has been an array of definitions on mentoring. Due to the multilateral nature of mentoring, a simplified categorization of different mentoring models is difficult. According to Clutterbuck (2004), the differences in definitions derive from the purpose of the relationship, the

expectations of the mentoring pair, and the context in which they operate. The two main mentoring traditions – the US and the European tradition – originate from different assumptions concerning the role and nature of mentoring.

Kram (1985) – who represents the US tradition – has divided mentoring into two main categories: career and psychosocial mentoring. Career mentoring – which is also called sponsorship – is a top-down model where the mentor is typically in a higher position in the same organization with the mentee and has power to influence the mentee's career. The mentor's tasks include coaching, protecting, and supporting the mentee, arranging challenging tasks, and helping the mentee to meet important people. Psychosocial mentoring is about strengthening the mentee's identity, self-respect, and feeling of competency, with the mentor being a role model, guide, and friend. Even if the mentoring task would focus only on one of these categories, in most mentoring cases, the other category is included at some stage of the process.

Meggison *et al.* (2006) – who represent the European tradition – view mentoring as a relationship where one person helps another to develop in the process of knowledge, work, and thinking. The mentor is often more experienced than the mentee, but not necessarily in a direct supervisory position. The focus of the mentoring process is on providing the time and space to take a wider view, and on helping the mentee to find his or her own solutions. The mutuality of learning between the mentor and mentee is emphasized (Clutterbuck, 2004).

Mentoring redefined

Many of the contemporary definitions seem to point out that the distinction between the mentoring traditions has diminished as the North American tradition has become closer to the more holistic European one. The mentoring process is more reciprocal than in a traditional apprenticeship model, and the mentor's role has changed from being an expert and adviser to a listener and promoter of reflective thinking (Boreen *et al.*, 2000). The mentor is also seen as a learner and, subsequently, the mentoring relationship brings advantages to all parties: the mentee, the mentor, and the organization. (Kram and Hall, 1996).

Group and peer mentoring

In addition to the traditional one-to-one mentoring models, alternatives have been developed. This has been important especially when organizations have become flatter and the individuals more self-reliant. In group-mentoring, the mentor is a learning leader of a team or a group. In a group-mentoring situation, everyone can benefit from the diverse knowledge and experience of the learning leader. Group-mentoring, often, also includes peer mentoring – where each participant is equal and benefits from the mentoring situation. Participants in the

mentoring complement each others' knowledge and skills in a reciprocal dialog (Kaye and Jacobson, 1996).

E-mentoring

E-mentoring is – at least partly – a virtual process. It is carried out by using electronic forms of communication, such as emails or other media using Internet connection (Kasprisin *et al.*, 2003). In synchronous e-mentoring, the participants in the mentoring process are online simultaneously. The interaction takes place using tools that enable immediate response and feedback – such as telephone, chat-channels, the Skype, or video conferencing. In asynchronous e-mentoring, the mentoring parties are engaged in the mentoring process at differing times, for example, via email, blogs, or discussion forums.

E-mentoring offers an option for a mentoring relationship, especially when there are no opportunities for face-to-face mentoring. E-mentoring is easier to access – being independent of time and place – and it is more egalitarian than traditional mentoring. It can also be less costly, at least from the point of view of the mentoring process (Bierema and Hill, 2005). However, its cost-effectiveness has been a point of debate, since setting up websites and creating software is time consuming and expensive.

E-mentoring is, increasingly, used to strengthen the relationship between education and working life. In addition to one-to-one mentoring using web-tools and email, the development of e-learning and e-working environments enables virtual group and peer mentoring. Especially in the ECEC field – where professional isolation may occur and where there may be a lack of local experts to support professional development – e-mentoring can offer better opportunities for finding mentors and bringing mentoring processes within the reach of the staff. Internet environments enable the involvement of a greater number of people in the mentoring processes.

Mentoring and reflection

For an individual, the development of expertise means responding to challenges and deepening understanding by being open to new ideas, by continuously seeking more information, and by learning from others in dialog and reflection. This is, however, not an easy task. First, a long working history does not necessarily guarantee improved expertise. Instead, it may lead to routines that create resistance to change and restrains development (Hakkarainen *et al.*, 2004). Second, acquiring new information and theoretical knowledge is made easy at present, with a variety of sources – such as educational courses, reading, conferences, or workshops – being available. New information may seem relevant and useful. However, linking theory and practice is not easy, and thus it is important to ask how new ideas and information might create new thinking and change practice.

According to Argyris and Schön (1974), theories of action are used to link our thoughts with our action. Theories of action can be espoused (those we know) or they can be theories-in-use – which are tacit in nature. One problem with transforming new ideas into practice is that the behavior of staff is governed by theories-in-use. To develop the working communities and to proceed in professional development, the tacit theories-in-use have to be made explicit by interaction and reflection in the working community.

When adopting reflective practices, the ECEC personnel start to think more of the implications their choices have for different children. Consciousness in the educational process means that the staff members are capable of examining their own pedagogical thinking and view of learning, that is, the capacity to ponder over what is best for the child, how the child learns, why they are doing what they are doing, what are the aims of education and the methods to reach the goals.

To achieve progress in the reflective practices, a promoter appears to be needed. This is where a mentor can play a fundamental role. The mentor can be internal or external to the work setting. Especially when the mentor's task is to promote change, the distance of an external mentor from the organization, sometimes, allows better handling of challenging and difficult issues. An external mentor works as an outsider who questions trivial issues, and who initiates dialog and common reflection in the working community. Sometimes, the use of internal mentors – who may be too committed to the working community's ways of action – can halt the development of working practices and values (Arhén, 1991), especially in the early-childhood field where the development of new thinking often collides with a this-is-how-we-have-always-done attitude. Even where members of the staff want to be reflective and begin to develop their practices, it may prove difficult. In the hectic daily working situation, old habits and routines easily take over if there is no one to support and encourage the process of change. A mentor can help to sustain the reflective atmosphere and support the systematically progressing change.

Mentoring the Working Communities

Mentoring brings certain advantages to professional development compared to traditional staff-support methods. Different mentoring models can help to develop staff competencies when there are no options for organizing other long-term staff educational programs. Using local expertise in mentoring can promote reflection processes among staff, especially when there is variation in the educational background of the staff members – both in the level of education and in the knowledge and skills they hold. In geographical areas with long distances and

low population density that often also have shortage of qualified staff, using external mentors from nearby locations or exploiting e-mentoring may be helpful in the provision of staff support. The National Research and Development Center for Welfare and Health in Finland developed two mentoring programs for ECEC, supplying national and regional processes for professional development. To better understand the function of these models, it is worth knowing that, in Finland, over 90% of the ECEC services are publicly provided. The services are regulated nationally and provided by municipalities. Services include center- and family-based ECEC.

A Coordinating Mentor in the ECEC Curriculum Process

Finland's national curriculum guidelines provide a frame for the content of the ECEC that has to be discussed and reflected upon at the municipality and individual service levels. The idea of the guidelines document is to promote an ongoing development process at each level, to increase the professional awareness of ECEC staff, and to, thus, raise the quality of the services. To implement such a framework curriculum, requires a well-structured professional development program – including both the provision of theoretical and practical knowledge, and multiple staff-support methods. The ambitious goal of the curriculum framework is to go beyond the daily discussion of functions, and to argue what is important in ECEC and how a vibrant community of children and adults can be built. (Lindberg, 2007).

The structure of the mentoring models

As a structure to support the professional development of ECEC staff throughout the country, a mentoring-network approach was developed. The mentors in this network were municipal ECEC experts with high qualifications and long-term working experience. The core of the approach was formed by eight 2-day mentor training sessions organized at the national level at the beginning of the process, in 2003. During the training sessions, the mentors were educated about both the content of the

curriculum and providing support for the ECEC staff in the implementation process.

Following the training, the mentors had two main tasks in the curriculum process to fulfill: (1) to start a networking project in their own municipalities or in larger geographical areas with a number of municipalities; and (2) to act as a group and peer mentor for the mentees in their areas. Thus, in addition to being experts to guide the staff for the content of the curriculum and to start reflection processes to change their practices accordingly, the mentors were coordinators of the municipal curriculum processes. The municipalities adjusted the group mentoring in accordance to the size and local needs into three different models. All models had several layers, the number of them depending on the size and other features of the municipality (Figure 1).

Model 1

Model 1 utilized group mentoring where one municipal mentor mentored several mentees in the municipality. Depending on the size and resources of the municipality, each mentor could have one or several groups of mentees. The mentoring inside the municipalities was organized using one of two alternatives: (1) the groups of mentees were usually staff from center-based ECEC settings (one group = one center) or a group of family child-minders; (2) each of the ECEC centers sent one or more members of staff into a mentoring group to be inducted into the mentoring process. They then became mentors for the second mentoring layer. Usually, a family child-minder (family-day care) coordinator also belonged into this group. The second layer of mentoring was then formed inside the centers between the mentor who participated in the induction meeting/s and other members of staff. This model was, especially, utilized in large cities.

Model 2

In model 2, a single mentor took care of several municipalities. This model was mainly used in small municipalities that did not have enough critical mass of adequately trained experts to run individual municipal mentoring programs. The municipalities combined their resources

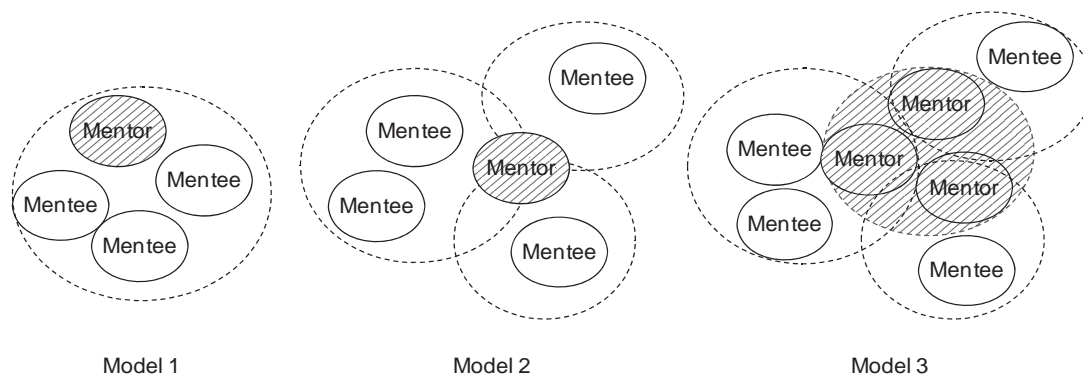


Figure 1 Local-level group-mentoring models.

and educated one mentor to co-ordinate a larger area of several municipalities. Often, each of the municipalities on the area had only one center, or their system consisted only of family child-minders.

Model 3

Model 3 effectively combined group and peer mentoring. Each mentor in the network had their own municipalities to mentor. In addition, mentors from several municipalities formed a network that utilized peer mentoring among the mentors by organizing their own meetings and training.

For example, in Northern Finland the regional Social Competence Center coordinated a model where 22 municipalities formed a network, together with the regional university – which, in this case, provided the mentors. Some of these municipalities were very small and did not have their own experts.

The mentors

The municipal mentors were skilled professionals with long-term working experience. Many of them had higher education than ECEC staff, in general – often possessing a master's degree in early education. Many mentors held day care-center leadership positions, and – especially in the smaller municipalities – had municipal-level leadership positions. In addition to being the process organizers, the mentors were the supporters of groups. These tasks demanded strong organization skills and extensive knowledge and experience in ECEC work. The leader or senior-expert positions held by these mentors gave them more credibility to mentor the other staff members.

Many of the staff members participating in the process as mentees were also highly educated and skilled with long-term working experience. This made the mentoring processes challenging and created a pressure to even more reciprocal reflection processes – based on dialog and peer mentoring.

Web-based support

The mentoring network was provided with its own extranet environment within the national public ECEC website. The extranet contained all training material, other important information, and a discussion forum. Via the extranet, the mentors had the opportunity to interact with mentors from different municipalities. They could, thus, share their visions and understanding, as well as problems, with the network. The open website also had an essential role in providing working tools and content material to assist the municipalities in their own curriculum processes.

Promoting Professional Development by Group-Based E-Mentoring

An e-learning environment for the professional development of ECEC staff was piloted in 2002–04. The environment consisted of material and process tools to promote

professional development in center-based ECEC settings. The centers were from four municipalities in the South of Finland – 12 centers in all. The primary goal was to offer the centers an opportunity for self-reflection with the help of the tools and e-mentoring. The process involved all the staff in the centers, and thus also aimed at increasing the sense of belonging in the working community. The e-learning environment, as a whole, supported professional development processes and encouraged reflection among members of staff – evidenced by the content of each center's documentation using the preset documentary tools.

E-mentoring was provided within the learning environment by a network of seven mentors. All the mentors in the network were experts working in the municipalities that took part in the process, and were highly educated with long-term career background. Each of the mentors also had a specific topic of expertise – such as children's special needs. The mentors' tasks were twofold:

1. To follow the documentation of the centers and to provide the centers help by offering reflective questions, comments, information, or suggestions for material to be utilized. All these contacts were made on an asynchronous discussion channel inside the learning environment. All activities were open to the centers involved, which allowed the centers also an opportunity to support each other in the form of peer mentoring.
2. To respond when a center actively asked for support in the form of questions for the network of mentors.

The mentors acted as a team. One of the mentors at a time took the main responsibility for reacting to questions posed on the network, for coordinating the discussion in the network, and for ensuring that the answers were completed in time. A rota of biweekly periods was arranged for this. The mentors also had their own forum inside the learning environment where only they had access. This allowed mentors themselves an opportunity for peer mentoring: to discuss and analyze the issues, expand their knowledge, and thus provide better support to the centers.

Developing working communities

A common feature in each mentoring-model example is that they emphasize the importance of reflection, dialog, and reciprocity in all actions taken. The mentors' tasks are to support the self-reflection of the mentees in ways that encourage them to reach their own conclusions and to find their own solutions within the working communities. The mutual dialog within the working communities is of crucial importance in the process. This dialog covered issues such as the basic values of the work, the importance of the being clear about the goals of education and the effect of pedagogical choices for the child's well-being,

learning, and development. These were seen as the key issues in promoting the development of the ECEC work.

Conclusion

The ECEC systems in different countries are strongly socially and culturally bound. It would be overly simple to claim that there can be universally suitable staff-support methods to promote professional development. However, certain aspects of mentoring can offer well-functioning tools for professional development programs in different countries, especially when combined with other staff-support methods.

In the ECEC sector, e-mentoring has much to offer in the future. However, it faces a number of challenges that have to be resolved if e-mentoring is to reach full potential. Contrary to face-to-face mentoring programs that can be set up – for instance, by a single center – e-mentoring needs a coordinating organization to fund and build the mentoring environment, and to manage and develop it. Although the problem of building the environment can be overcome by using environments based on open source, resources to educate the staff about effective use of the environment are required. In many countries, ECEC staff is not yet proficient in the use of information and communication technology (ICT), which creates further pressure for these kinds of proceedings. Asynchronous e-mentoring may be an easier and less expensive option to commence. Most ECEC professionals are familiar with email and discussion channels, and, therefore, the threshold to use them is lower.

An important aspect in building successful mentoring models is the training of mentors. Many of the websites providing information about mentoring in the education sector pay substantial amount of attention to mentor training. As mentors in ECEC communities are, mostly, professionals who have no formal education for adult teaching, educating the mentors is a crucial task for successful mentoring processes. One of the important aspects in the presented Finnish professional development processes was to help the mentors to understand, clearly, the idea of mentoring: what does it mean to become a mentor, what are the mentor's obligations, and how time consuming mentoring is. Mentors need support in their work and various peer-mentoring options can help give this support. It is also important for mentors to feel they belong to their respective working communities, even if they are external mentors.

Personality and individual skills and knowledge are important elements concerning professional expertise. In most of the current mentoring theory and discussion, mentoring supports an individual's professional development – either by providing help in one-to-one relationships or in groups. Less attention has been paid to the development of the entire working community, its values, and core ideas.

These aspects, however, should not be exclusive, because, on one hand, the working environment has an important role in the development of professional expertise (Karila, 2008) and, on the other hand, each individual has an effect on the working environment, its culture, and its course of development. Instead of viewing mentoring only as a process – where an apprentice is mentored into the existing, unchangeable working culture – mentoring can provide tools for change and learning. To realize the potential of mentoring, members of staff are required to have the courage to think, to take intellectual risks, and to exchange ideas with others openly in the process of developing expertise (Hakkarainen *et al.*, 2004). As such, mentoring can be an empowering process for the individual members of the working community.

Mentoring can decrease resistance to change by providing support and encouraging reflection and open dialog. In this way, mentoring helps to create an intelligent working community based on mutual, continuous learning.

See also: Integrated Early Childhood Education and Care Services – Care, Upbringing, Education and Health; Teaching in Early Childhood Centers Instructional Methods and Child Outcomes.

Bibliography

- Argyris, C. and Schön, D. (1974). *Theory in Practice: Increasing Professional Effectiveness*. San Francisco, CA: Jossey-Bass.
- Arén, G. (1991). *Mentorskap. Viktig del i all chefsutveckling*. Malmö: Liber Ekonomi. (Mentoring. An Important Part of all Leader Development.) Malmö: Liber Ekonomi.
- Bierema, L. and Hill, J. (2005). Virtual mentoring and HRD. *Advances in Developing Human Resources* 7(4), 556–568.
- Boreen, J., Johnson, M. K., Niday, D., and Potts, J. (2000). *Mentoring Beginning Teachers: Guiding, Reflecting, Coaching*. Portland: Stenhouse.
- Clutterbuck, D. (2004). *Everyone Needs a Mentor. Fostering Talent in Your Organisation*, 4th edn. London: Chartered Institute of Personnel and Development.
- Hakkarainen, K., Lonka, K., and Lipponen, L. (2004). *Tutkiva oppiminen. Järki, tunteet ja kulttuuri oppimisen sytyttäjinä. (Explorative Learning. Sense, Emotions and Culture Inspiring Learning.)* Porvoo: WSOY.
- Karila, K. (2008). A Finnish viewpoint on professionalism in early childhood education. *European Early Childhood Education Research Journal* 16(2), 210–223.
- Kasprisin, C. A., Boyle Single, P., Single, R., and Muller, C. (2003). Building a better bridge: Testing e-training to improve e-mentoring programmes in higher education. *Mentoring and Tutoring* 11(1), 67–78.
- Kaye, B. and Jacobson, B. (1996). Reframing mentoring. *Training and Development* 50(8), 44–47.
- Kram, K. (1985). *Mentoring at Work. Developmental Relationships in Organizational Life*. Glenview, IL: Scott, Foresman and Company.
- Kram, K. and Hall, D. (1996). Mentoring in the context of diversity and turbulence. In Kossek, E. and Lobel, A. (eds.) *Managing Diversity: Human Resource Strategies for Transforming the Workplace* pp 108–136. Cambridge, MA: Blackwell.
- Lindberg, P. (2007). The ECEC curriculum guidelines and its implementation in Finland. In O'Brien, S., Cassidy, P., and Schonfeld, H.

- (eds.) *Proceedings of CECDE International Conference 2007: Vision into Practice*, pp 26–35. Dublin: CECDE.
- Meggison, D., Clutterbuck, D., Garvey, B., Stokes, B., and Garret-Harris, R. (2006). *Mentoring in Action – A Practical Guide for Managers*, 2nd edn. London: Kogan Page.
- OECD (Organisation for Economic Cooperation and Development) (2006). *Starting Strong II. Early Childhood Education and Care*. Paris: OECD.
- Scherer, M. (ed.) (1999). *A Better Beginning: Supporting and Mentoring New Teachers*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Hezlet, S. A. and Gibson, S. (2005). Mentoring in the millennium. *Theory into Practice* **39**(1), 50–56.
- Higgins, M. and Kram, K. (2001). Reconceptualizing mentoring at work: A developmental network perspective. *Academy of Management Review* **26**(2), 264–288.
- Hunt, K. (2005). E-mentoring: Solving the issue of mentoring across distances. *Development and Learning in Organizations* **19**(5), 7–10.
- Penn, H. (2005). *Understanding Early Childhood. Issues and Controversies*. Berkshire: Open University Press.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. Cambridge: Cambridge University Press.

Further Reading

- Bierema, L. and Merriam, S. B. (2002). E-mentoring: Using computer mediated communication to enhance the mentoring process. *Innovative Higher Education* **26**(3), 211–227.
- Dahlberg, G., Moss, P., and Pence, A. (1999). *Beyond Quality in Early Childhood Education and Care: Postmodern Perspectives*. London: Falmer.
- Darwin, A. (2000). Critical reflections on mentoring in work settings. *Adult Education Quarterly* **50**(3), 197–211.
- Daugall, M. and Beattie, R. S. (1997). Peer mentoring at work. The nature and outcomes of non-hierarchical developmental relationships. *Management Learning* **28**(4), 423–437.
- Hakkarainen, K., Palonen, T., Paavola, S., and Lehtinen, E. (2004). *Communities of Networked Expertise. Professional and Educational Perspectives*. Oxford: Elsevier.

Relevant Websites

- <http://www.cecde.ie> – Centre for Early Childhood Development and Education.
- <http://www.empire-leonardo.org> – EMPIRE empowerment through mentoring.
- <http://www.eecera.org> – European Early Childhood Education Research Association.
- <http://www.amitie.it> – MAITRE, a Leonardo da Vinci Project.
- <http://www.oecd.org> – Organization for Economic Cooperation and Development: Early Childhood Education and Care.
- <http://www.opensource.org> – Open Source Initiative.
- <http://www.edu.salford.ac.uk> – University of Salford, a Greater Manchester University: Mentoring.
- <http://varttua.stakes.fi> – Varttua: a web service for ECEC.

ECONOMICS OF EDUCATION

Contents

Education and Health

Data

Empirical Research Methods in the Economics of Education

Theoretical Concepts in the Economics of Education

Education and Health

P Muennig, Columbia University, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Health-related quality of life (HRQL) score –

A numerical valuation of life in a given health state anchored between 0 (death) and 1 (perfect health). These scores are used to adjust life years to quality-adjusted life years (QALYs). Thus, a year of life lived at an HRQL of 0.7 is equivalent to 0.7 QALYs (or 0.7 years of perfect health).

Medical expenditure panel survey (MEPS) – An annual survey of around 40 000 households in the United States. It includes a household component with demographic and health information of the participants. It also contains a provider component that obtains the cost of medical visits from providers.

Prospective studies – The studies in which subjects are followed over time. Typically, this involves following subjects with and without putative risk factors for disease to ascertain who does and does not develop the disease.

Quality-adjusted life expectancy (QALE) – The life expectancy in perfect health. It is equal to the product of life expectancy and the age-adjusted health-related quality-of-life score.

Quality-adjusted life year (QALY) – A year of life lived in perfect health.

Randomized controlled trial – A form of experimental study in which subjects are randomly assigned to either a treatment condition or a control condition.

Risk ratio – The ratio of the probability of an event occurring in an exposed group relative to the probability of the event's occurring in the unexposed group.

Education and health together account for a quarter of the gross domestic product in the United States, and are perhaps the two most complex areas of domestic policy. They are also intertwined. Not only is health critical for learning, but also education appears to be a major determinant of health.

This article begins with a brief overview of the research exploring whether the relationship between education and premature mortality is causal. The primary causes of death from which the less educated are more likely to die are then explored. These causes of death point to six health risk factors associated with lower educational attainment: higher levels of stress, lower social standing, social deprivation, behavioral risk factors, lower health insurance coverage rates, and suboptimal cognitive skills.

Overview of Causal Relationships

It has been recognized since the time of Hippocrates that social conditions shape the ecological niche in which people live or die. Some 2300 years after Hippocrates, Horace Mann proposed that education might serve as a tool for repairing social ills. He observed that education is “the great equalizer,” providing people with the knowledge and technical skills needed to survive. While Mann was referring to income and social prestige, education also appears to be an equalizing force in health.

Evidence of a causal relationship draws upon a single, randomized, controlled trial, instrumental variable analyses, and a large number of prospective studies. The lone, randomized, controlled trial with long-term follow-up is the Perry Preschool Program, a study consisting of 123 students randomized to 2.5 h of preschool at ages 3 and 4 years or no intervention. Although the study was small,

children randomized to the preschool intervention were much more likely to report better health status and fewer behavioral risk factors for health (such as seatbelt use or smoking). By age 40, two of the children in the experimental group had died and five in the control group had died.

Most evidence arises from natural experiments, which primarily rely on instrumental variable analyses. Lleras-Muney (2005) conducted what is perhaps the best known of these studies, using compulsory schooling laws to examine the causal relationship between educational attainment and mortality. Her instrumental variable analysis exploits the fact that compulsory schooling laws should have a direct influence on educational attainment, but not on mortality rates. The analysis first examines the effect of compulsory schooling laws on educational attainment, and then estimates the effect of these laws on mortality. The ratio of the effect of the laws on mortality to the effect on educational attainment produces a direct estimate of the causal effect of educational attainment on health. Using a synthetic cohort derived from census data, Lleras-Muney (2005) found that compulsory schooling and child labor laws reduced mortality rates by as much as 60%. Given that adjustment for income or occupation produced little effect on the analyses, it was hypothesized that improved cognition, manifested as improved medical decision making, was the causal factor through which educational attainment produces health.

This same methodology has since been replicated in a number of other countries with similar results. It was also later explored using a larger census sample as well as using data from the Survey of Income and Program Participation (SIPP). This subsequent analysis of the census data raises concerns that the instrument might be picking up smooth cohort trends in educational attainment rather than discrete increases induced by more stringent compulsory schooling laws. Nonetheless, using the SIPP, a strong causal association between educational attainment and self-reported health was also found.

Beyond the handful of domestic and international natural experiments and randomized studies with proximal outcome measures, there are few studies specifically examining the causal association between education and health. However, there are a number of reasons to believe that the relationship between educational attainment and health is causal in nature. First, these studies do not reveal patterns in the relationship between health and third variables, those other than education, so little of the association is explained by unobserved heterogeneity in correlational analyses. Second, because children and adolescents have very low rates of chronic disease, sickness is not a major cause of low educational attainment. Thus, we can be more confident that most of the difference in life expectancy by educational attainment is explained by education rather than poor health. Third, the association is strong and consistent across cultures and time, and

fourth, the primary causes of death between those with more and less education can be readily explained by the very risk factors for diseases for which less-educated persons are most at risk.

The education–health association has been explored by epidemiologists, sociologists, economists, urban planners, social psychologists, and even neuroimmunologists. Once this work is connected, a rich picture emerges. Specifically, the risk factors most obviously linked to educational attainment are also risk factors for the diseases responsible for the 6–9-year gap in life expectancy between those with and without a high school diploma (Wong *et al.*, 2002).

These diseases include cardiovascular disease (35% of all deaths), cancer (27% of all deaths), infection (9% of all deaths), injury (5% of all deaths), lung disease (5% of all deaths), and diabetes (4% of all deaths). With the exception of injury, these risk factors overlap with respect to the diseases they cause (**Figure 1**). Many of these underlying risk factors are remarkably consistent with new evidence from the fields of psychology, neuroanatomy, neurophysiology, molecular biology, sociology, and epidemiology. Others, such as smoking or lower rates of health insurance, have been widely discussed. However, the interdisciplinary research helps us better understand and contextualize the linkages between these factors and educational attainment as well.

The following sections set the stage by discussing the relationship between poor health and poor education outcomes, then explore the evidence related to the six main pathways potentially linking education to health.

Health, Education, and Poverty

Before exploring each mechanism, it is important to consider the environmental factors giving rise to poor outcomes in both health and education. Poor health and poor education are tightly intertwined concepts. Native-born parents lacking a high school diploma tend to have lower than average income, less-healthy children, and children who are, themselves, at risk of dropping out of high school. Factors contributing to higher dropout rates among such children include attendance at poorly funded schools, exposure to lead paint, abuse, overcrowded living conditions, and a host of other health and environmental factors. The challenge, therefore, is to tease apart the primary ways in which education produces health independent of the influence of these factors on health.

For instance, a lead-abatement intervention may both improve health and educational attainment in children, but we would point to the lead-abatement program as the causative agent for any observed improvements in both health and education, not to the additional education received by the children in the abatement program. On the other hand, if a child received an educational

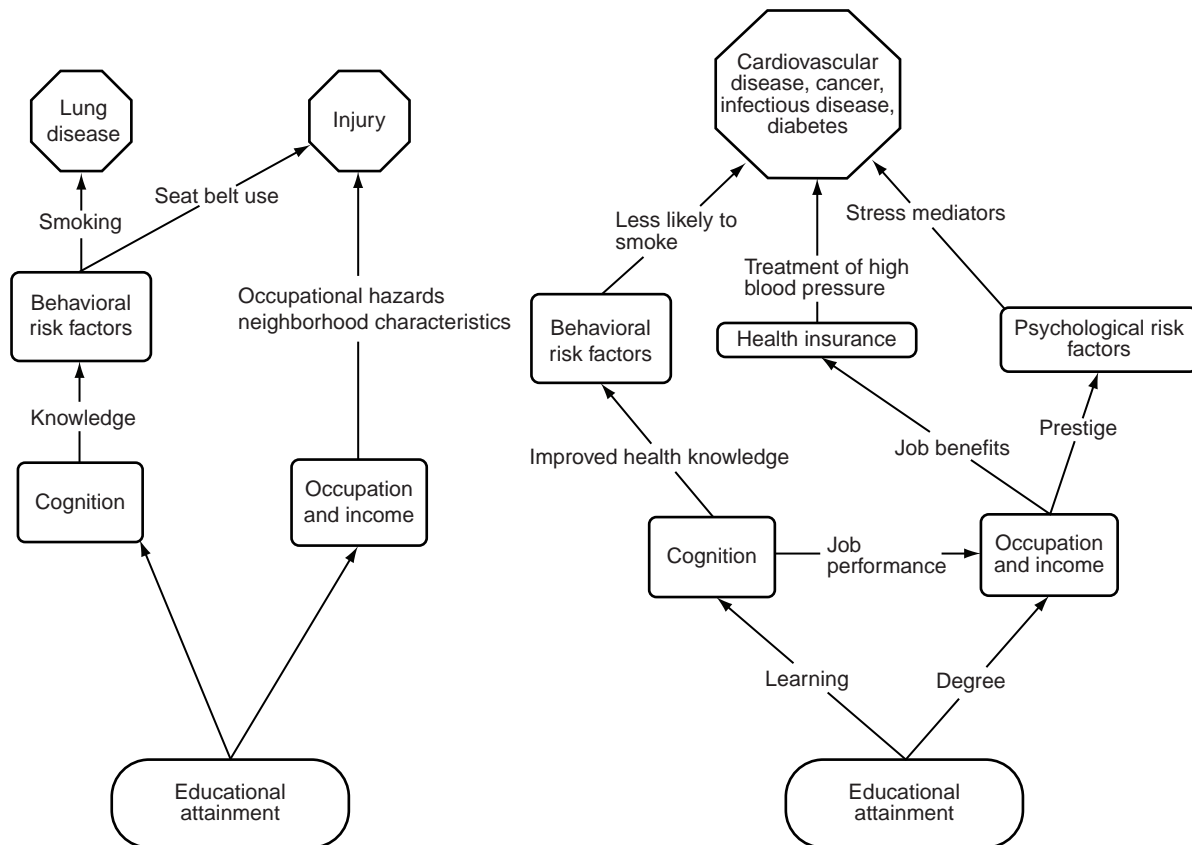


Figure 1 Selected plausible pathways through which education works to improve health.

intervention that induced him or her to graduate from high school, that individual will be more likely to obtain a quality job and enjoy the benefits of health insurance, a quality job, a lower-stress lifestyle, good housing stock, and other health-inducing social benefits. In the latter case, it was the education, not the intervention, which was central to the production of health.

Stress

Early associations between type-A lifestyles and heart disease led to the popular misperception that affluence and education naturally led to stressful lifestyles. Wealthier, more educated persons indeed have stress in their lives. However, while those with a postgraduate degree report higher levels of having too many things to do (relative to high school dropouts), nationally representative self-report surveys suggest that the better educated also have fewer concerns about money, health, leisure time, environmental noise, or problems with children. Self-report surveys also suggest that those with more education enjoy lower levels of anger, distress, aches, pains, and other factors that interfere with subjective quality of life than do those with less education. Finally,

those with more education are more likely to have fulfilling, rewarding jobs, a high sense of control in life, and higher levels of social support, all of which are associated with higher self-reported health and physical functioning.

If real, how might these differences in self-reported stressors and annoyances translate into poor health? Functional magnetic resonance imaging (fMRI) studies offer one view. These studies expose small numbers of subjects to stimuli that are intended to evoke an emotional response or a cognitive appraisal of an event. The scanner then detects which areas of the brain receive an increase in blood flow, and statistical testing is performed to ascertain which areas are significantly more active than others. fMRI studies suggest that, when one is exposed to a stressor, such as persistent horn honking or a difficult boss, parts of the brain that give these perceptions emotional valence become activated. These emotional centers, collectively known as the limbic system, in turn activate other circuits in the brain responsible for regulating heart rate, blood pressure, and the production of stress mediators.

We can observe the effects of these neural processes on the body by examining the relationship between levels of stress mediators in the blood and educational attainment (McEwen, 1998). When subjects of differing social class

(and, by extension, educational attainment) are exposed to mild stressors such as a line-tracing task, persons with a low income and low education tend to exhibit an abnormal stress response. Many of these stress mediators (e.g., cortisol, interleukin-6, catecholamines, C-reactive protein, and fibrinogen) are putative risk factors for cardiovascular disease, hypertension, diabetes, and infectious disease – four of the education-related diseases in **Figure 1** (Muennig, 2007). Nevertheless, it is possible that these stress mediators are merely markers of other processes or that their association with education-related disease is simply the result of spurious association.

In addition to leading to high blood pressure, diabetes, and disruption of the immune system, stress may cause oxidative damage, which increases the rate of human-cell aging and possibly increases one's risk of cancer. The first study in this area recruited unstressed subjects and compared them to subjects with a chronically ill child to ensure large differentials in self-reported stress (Epel *et al.*, 2004). They then examined a marker of cell aging, called the telomere, to ascertain whether there was an association between self-reported stress and cell age in the immune system. They found that subjects with high levels of self-reported stress show chromosomal changes that are consistent with an entire decade of additional biological life relative to those with the lower levels. This study was potentially confounded by their selection criteria; parents with a chronically ill child may be more likely to suffer from genetic disease and therefore be more prone to premature cell aging.

A subsequent study overcame this limitation by examining twins who had different levels of educational attainment (along with other markers of socioeconomic status (SES)). While the outcome measures they included in the second study were different, this group of researchers effectively confirmed the finding that stress increases cell aging. These authors found that the increased aging among less-educated, poorer twins could be accounted for by increased rates of smoking and obesity, and lower rates of exercise. While these authors did not fully disentangle these behavioral risk factors from psychosocial stress, the study raises the possibility that behavioral risk factors and stress are intertwined.

In a related series of studies, subjects with various forms of stress, including job-induced stress, were found to have higher levels of DNA damage and higher blood levels of a marker for cancer. This damage is thought to occur when the acute stress response causes the physiological release of oxidative chemicals into the bloodstream. A subanalysis of these data found that effective cognitive coping skills for stress (measured using a validated instrument) were associated with lower levels of DNA damage. Given that increased educational attainment is associated with improved cognitive coping skills and social network size, it is plausible

that education may reduce the incidence of disease though this association.

Social Standing

Social identity theory suggests that individuals tend to categorize other people on the basis of characteristics such as educational attainment as a means of self-comparison. Differences in relative standing – whether due to less education or less wealth – can be a source of anger, envy, or stress. The long human history of bloody wars fought over social class, such as the Bolshevik Revolution, suggests both that these feelings can be powerful and that they can have direct consequences for health and mortality. It has been suggested that the modern manifestations of lower social standing include both internalized physiological disruptions and crime. This hypothesis was initially drawn from heavily confounded ecological studies assessing the effects of income inequality on mortality and crime, but it is now supported by a stronger base of outcome measures and research designs.

The idea that social status affects health came to light in studies of government workers in England. These studies found that, among persons with good jobs in the same government department and equal access to healthcare, occupational class was still inversely linked to premature mortality. More surprisingly, this health gradient extended all the way into white-collar jobs; even those with high-level jobs were at greater risk of premature mortality than those in the most prestigious jobs. Similar gradients were subsequently found for income and education, and these gradients have been observed in a wide variety of cultural and economic contexts. In fact, it is possible that the social prestige conferred by an educational degree is a more important determinant of life expectancy than the skills acquired with each year of education (see **Figure 2**).

These findings are corroborated by animal studies. Primates, low in social standing, measured in part by the size of the animals and dominance behavior, have higher levels of cholesterol and higher levels of stress mediators. When a dominant male is removed, the subordinate's laboratory tests improve. Likewise, when a dominant male is put among still larger males, his laboratory tests deteriorate. Such findings are, of course, only consistent among stable social hierarchies. Dominant primates whose authority is constantly challenged show opposite results, but most human primate hierarchies (e.g., corporate or government offices) more closely resemble stable nonhuman primate hierarchies than unstable ones.

The implications of social-standing research for education policy are unclear. All of human society is hierarchical, and the top and bottom of the hierarchy may be better delimited by those in one's immediate environment than by the relative social distance between the wealthiest and poorest members of society.

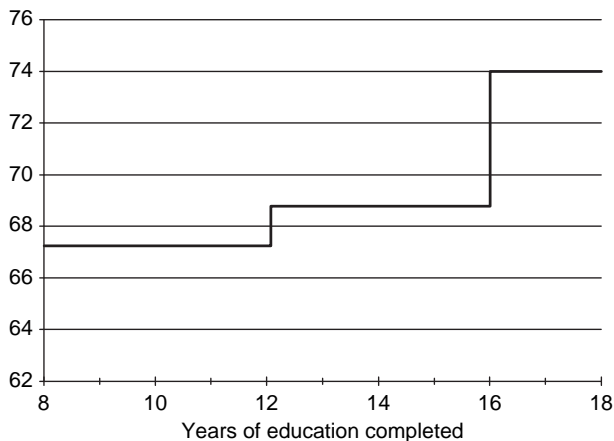


Figure 2 Change in life expectancy with educational attainment among males. Life expectancy remains relatively flat between educational milestones, but increases greatly when degrees are conferred. Mortality risks were converted to life expectancy values. All data points are adjusted for age, race, household size, marital status, employment status, and occupation. A similar relationship is seen for females, but the data are not presented here for simplicity. Adapted from Backlund, E., Sorlie, P. D. and Johnson, N. J. (1999). A comparison of the relationships of education and income with mortality: The National Longitudinal Mortality Study. *Social Science and Medicine* 49, 1373–1384.

Social Deprivation

For children with poorly educated parents, the effects of social deprivation manifest early in life. Harlow's (1964) classical work on animal bonding suggests that an individual's psychosocial troubles can begin with poor parenting. For instance, socially deprived monkeys tend to be more hostile and have predilections that lead them toward risk taking and alcohol consumption in experimental situations. These problems show a dose–response effect, with monkeys raised by a furry stuffed doll that administers milk fairs better than those raised by a similarly equipped surrogate mother made of chicken wire.

These problems can begin before birth; behavioral problems are also observed among the offspring of monkeys stressed by prolonged physical restraint during pregnancy. Likewise, humans raised in busy orphanages with less social bonding tend to make few friends and exhibit more social pathologies. Nicolae Ceausescu famously relied on orphans to supply the Romanian Secret Police because they had few social contacts and he correctly judged that they would be able to kill without remorse.

From childhood to death, persons born into poverty face a series of psychological and social obstacles that often include neglect, abuse, failing schools, bullying, low graduation rates, low wages, and often multiple jobs. This does not portend a healthy psychological milieu in which children or adults can thrive. Educational and

economic deprivations such as these are associated with a cascade of biochemical mediators in childhood, which may lead to cardiovascular disease later in life. Many childhood events, such as exposure to lead paint and abuse, are largely irreversible and simultaneously affect cognitive development and health. Lead paint has long been known to be a neurotoxin, and even small quantities of lead cause intelligence quotient (IQ) deficits. Likewise, children who suffer from physical and emotional abuse have been shown to have smaller brain structures than children who do not.

It is possible that weaker social networks are merely markers for childhood deprivations, and that such traumas are the primary reason why weak social networks are associated with poor health. However, it is likely that poor social networks are also independent risk factors for poor health.

Social deprivations and their associated feelings of anger or hostility, like other psychosocial stressors, are associated with higher rates of cardiovascular disease, infectious disease, lung disease, and diabetes. Although speculative, risk taking and anger may also partially explain higher mortality; persons of lower educational attainment sustain significantly more injuries, and display higher rates of smoking and drinking. Data from the aforementioned Perry Preschool Program randomized controlled trial (along with many other quasi-experimental studies) confirms much higher rates of crime, injury, and death among those receiving less education.

The Perry Preschool Program randomized experiment suggests that these early deprivations are somewhat reversible with early intervention. By age 5, 67% of the children in the experimental group, relative to just 28% of the children in the control group, had a measured IQ greater than 90. By age 40, those in the intervention group were a third less likely to have multiple arrests and were a third more likely to have earnings over US\$20 000 per year. Relative to those who did not receive the intervention, males who received the intervention were nearly twice as likely to raise their own children, less likely to use drugs, and more likely to report satisfactory relationships with their children. Thus, the intervention appeared to modulate the intergenerational transfer of social pathology.

Likewise, one nonrandomized age-matched trial of 1539 children showed decreased arrest rates at age 20 years among those children assigned to a prekindergarten (plus family interventions and health services) intervention group relative to children receiving a less-intensive preschool intervention. While the rates of social pathology were reduced in both instances, it is unknown whether the benefits of these preschool interventions extended to other health measures.

Finally, there is evidence that other social interventions reduce social pathology as well. In one natural experiment, Costello *et al.* (2003) examined the effects of

a casino opening on social pathology rates among Native American on a reservation. The investigators examined the rate of mental health diagnoses among 1420 children in the community according to income level before and after the casino redistributed large amounts of cash to Native American families registered by the tribe. After 8 years of follow-up, results showed the rate of conduct and oppositional/defiant disorders, such as physical abuse, among the newly rich families fell to levels seen in families that were well-off before the intervention. On the other hand, rates of conduct and oppositional/defiant disorders remained high among those families not receiving the income distributions.

To the extent that educational attainment improves social networks and reduces social pathology, we would expect education interventions to improve health through reduced negative emotions and improved behavioral risk factors.

Behavioral Risk Factors

The combined effects of smoking, eating poorly, and lack of exercise are thought to explain somewhere between 12% and 50% of the association between educational attainment (or other socioeconomic risk factors) and mortality (Lantz *et al.*, 1998). About 44% of white males without a high school diploma and 32% of white males with a high school diploma (but no higher degree) are self-reported smokers. Although the overall prevalence of smoking is lower among Hispanics, African-Americans, and females, the percent reduction in smoking is similar across these populations among those who have a high school diploma relative to those who do not.

It has been pointed out that most people, regardless of their level of educational attainment, know that smoking, drinking, and eating greasy food is not healthy (Mechanic, 2002). One might speculate, however, that, in addition to reducing fatalism, education could increase the total exposure to preventative health information, thus normalizing it. Education might similarly improve cognitive appraisal of this information. For instance, people may be more likely to eat healthy food if they understand that cholesterol and saturated fat clog arteries, and can visualize these arteries, instead of just having a vague, abstract notion that these substances are harmful. Better-educated persons might also be better equipped to balance health information against messages from the fast food, tobacco, and alcohol industries.

As increased educational attainment improves income, it may also exert positive effects on health behavior through upward mobility to neighborhoods where healthier foods are available in stores and their consumption is normative. Given lower rates of crime, wealthier neighborhoods also afford more opportunity for exercise. One large, randomized, controlled multisite trial evaluated

the health, crime, and other social effects of vouchers for housing that allowed recipients to move out of low-income neighborhoods. Five years after randomization, those who received the vouchers had significant reductions in obesity and improvements in measures of mental health relative to those randomized to receive no vouchers.

Finally, prospective examination of the contacts that people make in their lives suggest that improved diet and exercise as well as reduced smoking and obesity rates are contagious. That is, when an individual develops a contact with a peer that does not smoke or drink, he or she is much less likely to smoke or drink (Christakis and Fowler, 2007). Thus, the effects of educational attainment on behavioral risk factors appear to be transmissible to less-educated peers.

Genetic Risk Factors

Individual characteristics are determined by a combination of genetic predispositions and environmental influences. Genetic predispositions that influence health behaviors may also influence success in school. For instance, twin studies suggest that a person's ability to cope with stress is partly determined by genetic factors; further, the probability of having a life event perceived as stressful in the first place has also been partly linked to genetic factors.

Environment also plays a role; while twin siblings reared apart show some concordance on measures of IQ and income, twins who fall on hard times socioeconomically have been found to have more cardiovascular risk factors (e.g., higher blood pressure and cholesterol levels) than identical siblings not raised in poverty. Twin studies strongly suggest that measures of health status and cognitive ability are partially environmentally determined. Unfortunately, such studies cannot produce a reliable estimate of the contributions of genes, environmental factors, or of the interaction of the two (Boomsma *et al.*, 2002).

First, twins reared apart are nonetheless exposed to similar environments *in utero*, and sometimes the same environment in childhood, both of which are strong predictors of adult mortality by SES and social pathology. Therefore, upon comparing adult twins reared apart, it is difficult to disentangle genetic effects from early environmental effects. For instance, fetal alcohol and lead paint exposure are both risk factors for delayed development, and both are strongly associated with parental educational attainment.

Second, virtually all of the conceivable genetic contribution to the education–health gradient described here can be attributed to a large number of genetic foci. The expression of these genes, in turn, is highly influenced by environmental factors in a dose–response manner. Consider a hypothetical experiment in which a group of

fetuses is randomized to ideal childhood conditions and another group is randomized to a harsh childhood. Among children with loving parents, excellent schools, and ideal nutritional intake, genetics will play a major role in determining which children do well in school and live a long and prosperous life. Among children with abusive parents and peers who are exposed to drugs, unhealthy food, lead paint, and bad schools, the degree of exposure to positive environmental conditions (such as preschool programs or mentors) will be a major determinant of their longevity and prosperity. In these latter situations, genetics play a small role in explaining educational attainment.

In a groundbreaking twin study, Turkheimer *et al.* (2003) set out to test this hypothetical scenario using biometric analyses. They examined the contribution of (1) genetics, (2) shared environment, and (3) nonshared environment on IQ as measured by the Wechsler Intelligence Scale for Children administered at age 7 years. Their models examined the interaction of these three characteristics and SES using the National Collaborative Perinatal Project data set. This prospective study included 48 197 pregnant women and their 59 397 children. Their measure of SES was based on a linearly combined measure of parental educational attainment, occupation, and income. Their intent was to measure the interaction between genotype and environmental conditions (as measured by SES). They accounted for the possibility that SES is genetically determined by including the main effect of the moderating variable in their model. They found that, among poor families, IQ was determined almost entirely by the childhood environment. Among wealthy families, on the other hand, genes were almost entirely predictive of IQ. Of course, this study's conclusions hinge on the assumption that SES is an adequate proxy measure of a harsh versus an ideal childhood environment.

In another compelling study, Korean adoptees – who were essentially randomly assigned to families of varying SES – were studied later in life (Sacerdote, 2004). Education and health outcomes were then compared with those of their new parents and siblings. Clearly, Korean children adopted into non-Asian families tend to stand out in social situations, and this likely exerts an influence on their development. Nonetheless, adoptees assigned to better-educated parents do better in school, go further, and are healthier by some measures than adoptees assigned to parents with less education. Most strikingly, adoptees were just as likely as their nonadopted siblings to take up smoking and drinking, providing strong evidence that these behavioral risk factors might not be inherited.

In sum, both genetic factors and environmental factors influence characteristics of individuals that are critical in determining both SES and health status later in life. Study design problems make it difficult to quantify the effect of genetics on health. However, both current scientific

evidence and logic suggest that genetic factors will be stronger determinants of health among more affluent children and much weaker determinants among poor children. Thus, intellectual capacity will be likely be optimized when educational interventions are targeted toward low-income families and schools in low-income communities. Clearly, to maximize environmental variables, it is also important to optimize other aspects of childhood and adult environmental conditions, such as access to healthy foods, good housing, safe transportation, and medical care.

Health Insurance

Of these childhood environmental characteristics, medical care for children has received almost as much attention as education interventions. Massachusetts, Pennsylvania, and California all have planned health insurance reforms that are primarily driven by concerns surrounding child health. Nonetheless, while it is clear that families with poorly educated parents simultaneously lack health insurance and suffer from poor health outcomes, there is as of yet limited evidence showing that the possession of health insurance is causally linked to improved health status.

Among 18–64-year olds, 7.3% of persons with at least a bachelors degree lack health insurance, compared with 27.6% of those without a high school diploma. This lower rate of insurance among those with less education is probably attributable to fact that less-educated persons tend to have lower-quality jobs, which usually do not offer health insurance.

Access to medical care increases access to medications and treatments that are known to reduce morbidity and mortality. Of those diseases prominent in the education gradient, access to medications that reduce cholesterol, blood pressure, and diabetes may be most important (see **Figure 1**). The best evidence of the efficacy of health insurance to date suggests that insurance may improve health through these modalities.

In the 1982 Rand health insurance experiment, 3958 healthy but uninsured subjects were randomly assigned to either receive a premium health insurance policy or a policy that required financial contributions on the part of patients before they could receive care (Brook *et al.*, 1983). Subjects were assigned to their insurance plans for 3 or 5 years, and then evaluated for health outcomes, including mortality. These authors found that mandatory patient contributions reduced healthcare utilization relative to those who had no such requirement, but this barrier to care only produced a calculated 10% increased risk of death among high-risk subjects with hypertension. No changes were found in other measures of health outside of an improvement in vision through corrective lenses.

This study is somewhat dated, however, and over the past 24 years, a wide range of medications that prevent

heart disease, infections, and cancer have become available (e.g., statins, vaccines, and smoking cessation technologies, respectively). These medications, being expensive, are almost exclusively used by persons with health insurance. Recent correlational analyses show a 25–67% reduction in mortality among the uninsured (Muennig *et al.*, 2005). However, this reduction in mortality might be explained by model endogeneity, or other factors associated with insurance, such as the economic protection and peace of mind that health insurance affords, rather than the benefits associated with receiving medical attention itself.

Given that preventive modalities, such as antihypertensive medications, seem to matter most, and that these modalities reduce the incidence of disease in the education–health gradient (**Figure 1**), health insurance seems a logical contender for an explanatory variable in the education–health gradient. If accurate, this 25–66% reduction in mortality would account at most 3–9 months of the roughly 6–9-year difference in life expectancy between those with and without a high school diploma. It should be noted though, that the effects of educational attainment on cognition should synergize with health insurance effects; in addition to improving access to medications, more-educated people are also more likely to comply with prescribed dosages and intervals.

In sum, in industrialized nations, healthcare likely plays a small but significant and growing role in reducing health disparities by SES. While health disparities by educational attainment are smaller in nations that offer health insurance, these countries also tend to offer more social services like childcare and parental leave, making it difficult to disentangle the effect of health insurance from the effects of other social programs on reducing educational disparities in health. Moreover, increasing health insurance would be significantly less cost effective than implementing effective education interventions, such as small class sizes (Muennig and Woolf, 2007).

Nonetheless, universal health insurance coverage may be one potential policy approach to addressing the health–education gradient. The findings of the Rand health insurance experiment notwithstanding, universal healthcare may in of itself prove to be an effective education intervention. After all, it is difficult for children to learn if they are sick. Moreover, while health may improve educational attainment, increasing educational attainment may improve the utilization of healthcare among those who are already insured.

Enhanced Cognitive Ability

Even among those with access to medical care, knowledge of screening test availability is predictive of the use of such tests. Those with more education are also more likely to be compliant with their medications and otherwise manage their disease better than those with less

education. In addition to increasing the uptake and proper use of health technologies, improved cognition also likely reduces day-to-day errors that can lead to automotive or household accidents. In addition to increasing one's access to health information and improving one's comprehension of the information, cognitive ability may also influence the so-called future discount rate, that is, a value assigned to consequences and events far in a person's future, such as retirement funds or smoking-induced lung cancer.

Conclusions

Lower educational attainment likely affects health both through a higher-stress lifestyle and through material deprivation. Social stressors both partially originate from, and are compounded by, a higher likelihood of exposure to social pathology, weaker social support networks, and lower social standing. Likewise, lower rates of health insurance, weaker cognitive coping mechanisms, lower job quality, and an inability to fully understand medical diagnoses or treatments all seem to play a role.

Less explored are the effects of fatalism, anger, and hostility among those with less educational attainment. These negative emotional styles may contribute to higher rates of smoking, drinking, and eating poorly among those with less education. Recent evidence suggests that influences from educated friends and family members also play a large role in behavioral risk factors.

In sum, the psychosocial factors linked to educational attainment are also major medical risk factors for cardiovascular disease, cancer, infection, injury, lung disease, and diabetes – the diseases contributing most to the 6–9 years of life expectancy separating those with and without a high school diploma.

See also: Education Production Functions: Concepts; Family Environment in the Production of Schooling; Human Capital; The External Benefits of Education.

Bibliography

- Backlund, E., Sorlie, P. D., and Johnson, N. J. (1999). A comparison of the relationships of education and income with mortality: The National Longitudinal Mortality Study. *Social Science and Medicine* **49**, 1373–1384.
- Boomsma, D., Busjahn, A., and Peltonen, L. (2002). Classical twin studies and beyond. *Nature Genetics* **3**(11), 872–882.
- Brook, R. H., Ware, J. H. Jr., Rogers, W. H., *et al.* (1983). Does free care improve adults' health? Results from a randomized controlled trial. *New England Journal of Medicine* **309**(23), 1426–1434.
- Christakis, N. A. and Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine* **357**(4), 370–379.
- Costello, E. J., Compton, S. N., *et al.* (2003). Relationships between poverty and psychopathology: A natural experiment. *JAMA* **290**(15), 2023–2029.

- Epel, E. S., Blackburn, E. H., Lin, J., *et al.* (2004). Accelerated telomere shortening in response to life stress. *Proceedings of the National Academy of Sciences of the United States of America* **101**(49), 17312–17315.
- Harlow, H. F. (1964). Early social deprivation and later behavior in the monkey. In Abrams, A., Gurner, H. H., and Tomal, J. E. P. (eds.) *Unfinished Tasks in the Behavioral Sciences*, pp 154–173. Baltimore, MD: Williams and Wilkins.
- Lantz, P. M., House, J. S., Lepkowski, J. M., *et al.* (1998). Socioeconomic factors, health behaviors, and mortality: Results from a nationally representative prospective study of US adults. *JAMA* **279** (21), 1703–1708.
- Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *Review of Economic Studies* **72**(1), 189–221.
- McEwen, B. S. (1998). Protective and damaging effects of stress mediators. *New England Journal of Medicine* **338**(3), 171–179.
- Mechanic, D. (2002). Disadvantage, inequality, and social policy. *Health Affairs (Millwood)* **21**(2), 48–55.
- Muennig, P. (2007). How education produces health: A hypothetical framework. *Teachers College Record* **14606**, 1–17.
- Muennig, P., Franks, P., and Gold, M. (2005). The cost effectiveness of health insurance. *American Journal of Preventive Medicine* **28**(1), 59–64.
- Muennig, P. and Woolf, S. (2007). The cost-effectiveness of education as a health intervention: An analysis of the health and economic benefits of reducing the size of classes. *American Journal of Public Health* **97**(11), 2020–2027.
- Sacerdote, B. (2004). What happens when we randomly assign children to families? *NBER Working Paper No. 10894*.
- Turkheimer, E., Waldron, A. H., Onofrio, B., and Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science* **14**(6), 623–628.
- Wong, M. D., Shapiro, M. F., Boscardin, W. J., and Ettner, S. J. (2002). Contribution of major diseases to disparities in mortality. *New England Journal of Medicine* **347**(20), 1585–1592.
- Haynes, S. G., Levine, S., Scotch, N., Feinleib, M., and Kannel, W. B. (1978). The relationship of psychosocial factors to coronary heart disease in the Framingham study. I. Methods and risk factors. *American Journal of Epidemiology* **107**(5), 362–383.
- Irie, M., Asami, S., Nagata, S., *et al.* (2001). Psychosocial factors as a potential trigger of oxidative DNA damage in human leukocytes. *Japanese Journal of Cancer Research* **92**(3), 367–376.
- LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cellular and Molecular Neurobiology* **23**(4–5), 727–738.
- Levin, H., Belfield, C., Muennig, P., and Rouse, C. (2007). The public returns to public educational investments in African-American males. *Economics of Education Review* **26**, 700–709.
- Link, B. G. and Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior* **80**–94.
- Muennig, P., Sohler, N., and Mahato, B. (2007). Socioeconomic status as an independent predictor of physiological biomarkers of cardiovascular disease: Evidence from NHANES. *Preventive Medicine* **45**, 35–40.
- Reynolds, A. J., Temple, J. A., Robertson, D. L., and Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *JAMA* **285**(18), 2339–2346.
- Roberts, H. (2000). What is sure start? *Archives of Disease in Childhood* **82**(6), 435–437.
- Ross, C. E. and Wu, C. L. (1995). The links between education and health. *American Sociological Review* **60**, 719–745.
- Sapolsky, R. M. (2005). The influence of social hierarchy on primate health. *Science* **308**(5722), 648–652.
- Schweinhart, L. J. (2004). *The High/Scope Perry Preschool Study through Age 40*. Ypsilanti, High/Scope.
- Smith, G. D., Shipley, M. J., and Rose, G. (1990). Magnitude and causes of socioeconomic differentials in mortality: Further evidence from the Whitehall study. *Journal of Epidemiology and Community Health* **44**(4), 265–270.
- Wilkinson, R. G. (1999). Health, hierarchy, and social anxiety. *Annals of the New York Academy of Science* **896**, 48–63.

Further Reading

- Cassel, J. (1976). The contribution of the social environment to host resistance: The Fourth Wade Hampton Frost lecture. *American Journal of Epidemiology* **104**(2), 107–123.
- De Bellis, M. D., Keshavan, M. S., Clark, D. S., *et al.* (1999). A. E. Bennett Research Award. Developmental traumatology. Part II: Brain development. *Biological Psychiatry* **45**(10), 1271–1284.
- Deaton, A. (2002). Policy implications of the gradient of health and wealth. An economist asks, would redistributing income improve population health? *Health Affairs (Millwood)* **21**(2), 13–30.

Relevant Websites

- <http://www.macses.ucsf.edu> – MacArthur Network on SES and Health summarizes the seminal works in pertaining to the education-health relationship, and provides information on many of the key researchers in this area.
- <http://www.iserp.columbia.edu> – The Institute for Social and Economic Research and Policy (ISERP) is a think tank based at Columbia University that focuses on the policy implications of social research.

Data

M Perez, Stanford University, Stanford, CA, USA

M Socias, American Institutes for Research, Palo Alto, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

As research in economics of education moves toward causal research methods, the importance of high-quality data on educational institutions, students, teachers, and policies has increased significantly over time. This article offers an overview of the available national, state, and school district databases that might be of interest for researchers and policymakers alike. In doing so, we center our attention on US primary and secondary education data sources. (For a discussion of the different empirical research methods in education, see the article on empirical research methods in this encyclopedia. McEwan (2007), Barrow and Rouse (2005), Angrist (2004), Loeb and Strunk (2003), and Meyer (1995), also provide a thorough discussion.)

This article builds on Dee *et al.* (1999), paying special attention to national longitudinal surveys and state longitudinal administrative datasets. (The national longitudinal survey of youth (NLSY) and the panel study of income dynamics (PSID) are excluded from this review. Both have been extensively used and described in prior research studies. See, e.g., Brown *et al.*, (1996) for a description of PSID, and visit the NLSY website to access the NLS handbook that contains an overall picture of the seven cohorts of this survey.) Without intending to be exhaustive, it offers a comprehensive description of the main data sources that may be of interest to researchers in the economics of education field. It is organized as follows. We first present an overview of the different research questions that can be answered with the type of data systems described in this article. We then describe the most relevant nationally representative data systems. We follow it up with a description of state and district data systems that contain longitudinal information and briefly describe some international data systems. The last section offers concluding remarks.

Overview

Before describing the available datasets, it is important to provide a general overview of the type of research questions economists of education focus on, and how these data sources may be used to answer them. Some basic questions posed by economists of education are: (1) what are the effects of education inputs (e.g., class size, teachers'

qualifications, and professional development), competition among schools (i.e., vouchers), and accountability measures on academic achievement; (2) of cognitive abilities and education on earnings; (3) of pay incentives, school characteristics, and testing on teacher characteristics; and (4) of family background characteristics (i.e., income and parental education) on student academic attainment? In order to address these questions, researchers need state- or nationally representative data on student test scores, class and school size, characteristics of peers, teacher education and experience, incentives associated to the federal and state accountability system, and/or family background. In addition, researchers need to be able to link these different variables with each other, which means linking students to teachers, students to peers, and/or follow students over time. The different nationally representative datasets that are featured in this article, which are actively maintained by the National Center for Education Statistics (NCES), tend to specialize in some of the areas mentioned above. For instance, the National Assessment of Education Progress (NAEP) provides extensive data on student test scores, while the School and Staffing Survey (SASS) collects data on teacher-related topics. Therefore, being able to merge information across datasets is a basic prerequisite for addressing these research questions at a national level. In the case of state or district databases, generally a single data collection effort covers student, teacher, and school information.

National Data Systems

The observed emphasis on scientifically based research in education takes place within a tradition of extensive use and collection of data in the field. Since the 1980s, the US has used complementary approaches to analyze the education process from different perspectives. In this section, we present a brief description of the main nationally representative databases that are maintained by the federal government. We present some examples of empirical research that have used these data sources, and also provide relevant information on how to access these data.

CCD and SASS Databases

The Common Code of Data (CCD) is the information backbone of the US education system. It contains the list of all public primary and secondary schools in the country, as well as all public school districts. This database

identifies all charter, magnet, Title I, special education, vocational education, and alternative schools in the country. It also keeps track of enrolment figures by primary ethnicity, gender, and eligibility for free and reduced-price lunch. The number of classroom teachers is also provided. At the district level, staffing information is more extensive, including instructional, administrative, library, and support staff, as well as guidance counselors. At the district level, financial information is also available. (The fact that financial data are not available at the school level is probably one of the areas of this information system that should be addressed in the future. Without it, it is not possible to know in a systematic way how much additional resources are provided to schools under certain policy interventions.) NCES has also generated a district-level longitudinal data set – the Longitudinal School District Fiscal-Nonfiscal File – from 1989–90 through 2001–02.

Unlike the CCD, the SASS does not collect information on all schools in the country and across all years. It relies instead on a state-level representative sample of public schools. (SASS data are also representative at the private school affiliation level for private school respondents.) The trade-off for this smaller sample size is a rich dataset of teacher-related topics such as teacher shortage and demand, general characteristics of elementary and secondary teachers, teacher workplace conditions, and teachers' career paths (since 2003–04). In addition, SASS has a small longitudinal component called the teacher follow-up survey (TFS). A subset of the teachers surveyed in SASS is surveyed again the following year, even if they leave the teaching profession.

SASS goes well beyond teacher-related topics; it includes principal, school, and district's questionnaires. It also collects data on school programs and policies, admission requirements, tuition, student and class organization, special programs and services offered, and student instructional time, among others. The years covered by SASS have thus far been 1987–88, 1990–91, 1993–94, 1999–2000, and 2003–04.

Given the nature of SASS, it has been used extensively to analyze teacher-related questions. Ballou and Podgursky (1998) used this source of information to analyze teacher recruitment and retention in public and private schools. Angrist and Guryan (2004) studied the effect of teacher testing on teacher characteristics. Figlio (2002) addressed the question whether public schools could attract better-qualified teachers increasing teacher salaries, and Figlio and Rueben (2001) used these data to analyze the relationship between tax limits and qualifications of new teachers.

NAEP Assessment Data

Even though SASS, CCD and its private school counterpart, the private school universe survey (PSS), generate

rich information on student, teacher, and school characteristics, neither of them collects data on academic achievement. To fill this gap, NAEP provides assessment measures in mathematics, reading, science, writing, the arts, civics, economics, geography, and US history every 2 years. (Note that, since 1969, NAEP was carried out annually until 1979.) NAEP relies on a complex sampling design to ensure state- and national-level representativeness of academic performance estimates of public and nonpublic schools, as well as student subgroups by primary ethnicity, gender, and parental education for fourth, eighth, and twelfth graders. In fact, every 2 years, NAEP generates two different samples: one for state-level estimates used to compare state performances at a given moment in time, and another for national-level measures. (Not all subjects are covered in exactly the same years.) In addition, every 4 years a third sample is drawn with the purpose of identifying changes in academic performance of students, aged 9, 13, and 17 years, in mathematics and reading. This component of NAEP is known as the NAEP long-term trend (LTT) assessment. Of course, this long-term approach implies that assessment instruments need to be consistent over time, and therefore cannot change as the state and national cross-section NAEP assessments do, as they try to adapt to ongoing changes in curricula.

The increased emphasis on accountability since the 2001 No Child Left Behind (NCLB) Act has certainly influenced NAEP. Now states receiving Title I funding are required by law to participate in the state NAEP in mathematics and reading in grades 4 and 8 every 2 years. The accountability movement has also influenced the type of research questions that have been addressed using NAEP data during the last years. For instance, for cross-state comparisons of the percent of students meeting or exceeding proficiency (this is the key performance measure of the recent standards-based reforms) in a given subject to be consistent, it is crucial to know which states set higher standards than others. McLaughlin and Bandeira de Mello (2002) map state standards to the NAEP scale in order to compare them. Carnoy and Loeb (2002) use NAEP data to study the effect of accountability measures on student achievement.

National Longitudinal Studies: NLS-72, HS and B, NELS: 88, and ELS: 2002

Four longitudinal studies have been carried out in the US so far, including the national longitudinal study of the high school class of 1972 (NLS-72), high school and beyond (HS&B), the national education longitudinal study of 1988 (NELS:88), and the educational longitudinal study of 2002 (ELS:2002).

The first of them, the NLS-72, surveyed 21 000 high school seniors (i.e., twelfth graders) in 1972. This data collection effort not only gathered background information

on students and their schools, but also carried out cognitive tests during the base year. This cohort was resurveyed again in 1973, 1974, 1979, and 1986, as it entered higher education and the workforce. The first follow-up survey of 1973 also added 4500 new individuals to the sample. The second longitudinal study, HS&B, besides surveying and testing high school seniors, also collected information on sophomores (i.e., tenth graders). That is, it started collecting information of individuals earlier on in their life cycle. The younger cohort of HS&B was followed through 1992, and surveyed in 1982, 1984, 1986, and 1992. The high school senior cohort of 1980 was only followed through 1986 and surveyed every 2 years.

The third longitudinal study was initiated in 1988. The NELS of 1988 continued the trend of following individuals earlier on in their life cycle. The base year included a sample of about 25 000 eighth graders attending 1000 schools throughout the country. As the previous longitudinal studies, this sample was nationally representative of that particular cohort of students. Two years later the sample was expanded to allow national representativeness of estimates of tenth graders in 1990. The students were resurveyed in 1990, 1992, 1994, and 2000. The fourth and last longitudinal data collection effort, ELS: 2002, followed the cohort of tenth graders in 2002 over time. Again, student academic achievement was measured during the base year, as well as during the first follow-up in 2004. The last data point available (2006) followed this cohort into postsecondary education and the labor market.

The research agenda involving NLS-72 has focused extensively on the effect of college attendance on earnings. Frazis (1993) estimated the college premium trying to correct for selection bias, while Kane *et al.* (1999) proposed a methodology to identify measurement errors when schooling is self-reported (as it is in NLS-72 and HS&B), and analyzed its impact on the estimates of returns to schooling. Kane and Rouse (1995) studied the economic return of 2- and 4-year colleges. Altonji (1995) used NLS-72 to estimate the effect of high school curriculum on education and labor market outcomes. Finally, Acemoglu and Pischke (2001) used NLS-72, as well as HS&B, to analyze the link between family income and college education (Note that other research has been conducted using these data. This list is by no means exhaustive.).

The other longitudinal studies have provided invaluable information for analyzing other topics in the field. HS&B has been used by researchers to estimate the effect of Catholic schools on student test scores (see Coleman *et al.*, 1982; Coleman and Hoffer, 1987) and educational attainment (see Evans and Schwab, 1995). Other examples of studies involving HS&B are Murnane *et al.* (2000), which studied the impact of cognitive abilities on earnings, and Dee (2004), which addressed the question of civic returns to education.

The research using NELS: 88 has been somewhat more focused on teachers. Ehrenberg *et al.* (1995) used NELS: 88 data to evaluate the relationship between teacher and student ethnic and gender match and student learning. Goldhaber and Brewer (1995) studied the impact of teacher certification on student achievement using NELS: 88. Other researched topics with NELS: 88 have been high school size and academic achievement (see Lee and Smith, 1997), and the effect of unobserved school, teacher, and class characteristics on student achievement (see Goldhaber and Brewer, 1997). Given the more recent availability of ELS: 2002, its literature is very limited. Dee *et al.* (2006) used this data source to study the effect of school size on parental involvement in education.

Crime and Safety Surveys

An additional data source that may provide important contextual information on schools is constituted by the crime and safety surveys (CSSs) maintained by NCES. Two types of national surveys regularly address the crime and safety situation at American schools: the school survey on crime and safety (SSOCS), and the school crime supplement (SCS) to the national victimization survey (NCVS). These surveys collect data on frequency and type of crimes, disciplinary actions, prevention mechanisms, and the general safety climate at schools.

Accessibility

Extensive portions of all NCES databases are publicly available. Individuals and institutions can either download these publicly available data from the NCES website or request a CD-ROM at no cost. Researchers should note that some portions of these NCES databases are classified as restricted use, which means that researchers need to request a license in order to get access. Without this license, it is not possible to link school information in SASS to CCD or PSS, or analyze student-level records in any of the longitudinal surveys.

It is also important to mention that NCES is not the only institution involved in collecting and generating national education statistics. The US Census Bureau carries out the current population survey (CPS) and makes the public-use micro-data samples (PUMS) of the decennial census available. The CPS is a monthly survey of approximately 50 000 households, and NCES funds the October supplement survey, which collects data on educational attainment. Additional items in the CPS funded by NCES involve language proficiency, disabilities, computer use, student mobility, and private school tuition. The PUMS also contains data on a sample of households in the country. Education attainment estimates by age and race can be generated using this data source. For an overview of all these national education data systems please see **Table 1**.

Table 1 Summary of national data systems

<i>Dataset</i>	<i>Acronym</i>	<i>Link</i>	<i>Census or sample</i>	<i>Data frequency</i>	<i>Years available</i>
Common core of data	CCD	http://nces.ed.gov/ccd	Census	Annually	1986–87 through 2005–06 (for previous years contact NCES)
1. Public elementary/secondary school universe survey	CCD	http://nces.ed.gov/ccd/pubschuniv.asp	Census	Annually	1986–87 through 2005–06
2. Local education agency (school district) universe survey data	CCD	http://nces.ed.gov/ccd/pubagency.asp	Census	Annually	1986–87 through 2005–06
3. Local education agency (school district) finance survey (F-33) data	CCD/ F-33	http://nces.ed.gov/ccd/f33agency.asp	Census	Annually	1992 through 2005
4. Local education agency (school district) universe survey dropout and completion data	CCD	http://nces.ed.gov/ccd/drpagency.asp	Census	Annually	1997–98 through 2003–04
5. State nonfiscal public elementary/secondary education survey data	CCD	http://nces.ed.gov/ccd/stnfis.asp	Census	Annually	1986–87 through 2005–06
6. National public education financial survey data	CCD	http://nces.ed.gov/ccd/stfis.asp	Census	Annually	1987 through 2005
Schools and staffing survey	SASS	http://nces.ed.gov/surveys/sass	Sample	Every 3–5 years	1987–88, 1990–91, 1993–94, 1999–2000, 2003–04, 2007–08
Private school survey	PSS	http://nces.ed.gov/surveys/pss	Census	Every 2 years	1991–92, 1993–94, 1995–96, 1997–98, 1999–2000, 2001–03, 2003–04
Longitudinal school district fiscal–nonfiscal file	FNF	http://nces.ed.gov/edfin/finance_data.asp	Census	Annually	1990 through 2002
National assessment of educational progress	NAEP	http://nces.ed.gov/nationsreportcard	Sample	Annually until 1979, then every 2 years	1969 through 2006
National longitudinal study of 1972	NLS-72	http://nces.ed.gov/surveys/nls72	Sample	Variably	1972, 1973, 1974, 1976, 1979, 1986
High school and beyond	HS&B	http://nces.ed.gov/surveys/hsb	Sample	Variably	1980, 1982, 1984, 1986, 1992
National education longitudinal study of 1988	NELS: 88	http://nces.ed.gov/surveys/nels88	Sample	Variably	1988, 1990, 1992, 1994, 2000
Education longitudinal study of 2002	ELS: 2002	http://nces.ed.gov/surveys/els2002	Sample	Variably	2002, 2004, 2006
Crime and safety surveys	CSS	http://nces.ed.gov/programs/crime	Sample	Variably	1989 through 2005
1. School survey on crime and safety	SSOCS	http://nces.ed.gov/surveys/ssocs	Sample	Variably	1999–2000, 2003–04, 2005–06
2. School crime supplement	SCS	http://nces.ed.gov/programs/crime/student_data.asp	Sample	Variably	1989, 1995, 1999, 2001, 2003, 2005
Current population survey, October supplement	October CPS	http://nces.ed.gov/surveys/cps	Sample	Annually	1968 through 2005
Public-use microdata samples	PUMS	http://www.census.gov/main/www/pums.html	Sample	Every 10 years	1940 through 2000

State and School District Databases

With NCLB, the pressure for state and district officials to develop and maintain student-level longitudinal databases that track student progress over time has increased. In fact, according to a report in the *Education Week* (2006)

about the study of states' data, 39 states have longitudinal systems that can track individual academic progress and characteristics over time. As the quality of the data and the number of available years rise, we would expect to see an increase in the access and usage of these data systems. In this section we describe state and school district

administrative data systems that allow following students over time and that have been used to conduct empirical studies.

The National Center for Analysis of Longitudinal Data in Education Research

The Center for Analysis of Longitudinal Data in Education Research (CALDER) is a federally funded joint project of the Urban Institute and scholars at Stanford University, the University of Florida, the University of Texas at Dallas, Duke University, the University of Missouri, and the University of Washington. The mission of this center is to inform education policy leaders through analysis of longitudinal administrative databases on individual students and teachers. Their research focuses on how teacher policies, governance policies, and social and economic conditions affect teacher and student outcomes. One of the unique strengths of this center is the potential to explore similar questions across different states. This will not only increase the external validity of their findings, but also the reliability of their methodological designs and results. Having access to multiple data will allow them to test findings under different policy and context conditions. CALDER makes use of administrative longitudinal state databases that have been created and maintained by certain state officials, as well as state longitudinal databases that have been compiled by scholars. The states from which CALDER has longitudinal information include Florida, Texas, New York, North Carolina, Missouri, and Washington. The main characteristics of each of these databases are briefly described below.

Florida

Florida has been at the forefront of tracking individual students over time. The Florida Department of Education maintains longitudinal records on all Florida public school students and their teachers, from preschool through college, beginning with the 1995–96 school year. This education data warehouse (EDW) contains student information, such as test scores and student demographic data, course grades, courses taken, home language, disciplinary actions, and participation in programs designated for English learners. It also contains teacher information, such as education level, certification status, years of experience, and participation in professional development activities. In addition, it includes school and district financial records. These data have a unique characteristic. They can precisely link both students and teachers to specific classrooms at all grades, and at the same time, provide information on the percentage of time spent in each classroom. (Hansen (2006) provides a detailed account about the origins of Florida's comprehensive database as well as

how it has been enhanced and maintained by the Florida Department of Education.)

Florida longitudinal data have been used to address a wide array of policy-relevant questions. For example, Sass (2006) used Florida student information to study the performance of charter schools using student-fixed effects. Figlio (2006) investigated if schools employ discipline for misbehavior as a tool to increase student test performance. It is possible to obtain access to this confidential dataset by soliciting permission to the Florida Department of Education.

Texas

The Texas School Project (TSP) funded in 1992 by John Kain, is housed today at the University of Texas at Dallas. In 2007, the TSP was officially designated as a State of Texas Education Research Center. The TSP dataset combines individual-level information from multiple administrative datasets from state agencies and school districts. These data include student-level test scores starting with the 1992 third-grade cohort, enrolment information, demographic characteristics, and other K-12 public school data. They also contain community college and public university attendance and completion, teacher characteristics and teacher test scores, as well as higher education tuition and financial aid data.

Hanushek, Kain, and Rivkin have used these data in several articles. They have looked at the effects of charter schools, teachers' contributions to student outcomes, peer groups effects, the costs and benefits of switching schools, among others (Hanushek *et al.*, 1998, 2002; Hanushek and Rivkin, 2003). Interested researchers can gain access to these data by submitting a research proposal to the Texas School Project center, which will be reviewed for feasibility. The final decision is made by state authorities.

New York

Boyd, Lankford, Loeb, and Wyckoff have collected a dataset that follows teachers and administrators in New York State since 1969. This dataset contains state-level longitudinal teacher administrative records, City University of New York (CUNY) and longitudinal New York City student records. Teacher records include undergraduate institutions, teacher preparation programs, years of experience, education level, demographic characteristics, and certification area and scores, among other records. These data can be linked to New York City student-level test scores and demographics. This group of researchers has used these data to investigate the matching process of teachers and schools, the distance from where teachers attended high school to their first teaching position, and teachers' career paths, among others (see Lankford *et al.*, 2002; Boyd *et al.*, 2005). Interested researchers who would like to use New York data need to contact directly the New York Department of Education for approval.

North Carolina

The North Carolina Education Research Data Center (NCERDC), located at the Center for Child and Family Policy at Duke University, was established in 2000–01 to house several data sources in a unique location. The NCERDC stores data from the North Carolina Department of Public Instruction (DPI), NCES, and other agencies. The main goal of this center is to provide researchers with access to rich longitudinal data to inform policy-relevant research. These data have the capability to link students' test scores over time, analyze students' transitions to middle and/or high school, and identify students who move into different school systems (either charter schools or different LEAs), for example. Researchers can also link students to teachers, evaluate teacher's qualifications, payment and education levels, as well as follow teachers and identify their career paths.

Several researchers have used these data. For example, Bifulco and Ladd (2006) used North Carolina data to study the impact of school choice on racial segregation. Clotfelter *et al.*, (2004) studied the impact of the state's school-based accountability on the retention of high-quality teachers in low-performing schools. (NCERDC provides clear guidelines to access these data. Institutions of higher education, nonprofit research institutions, and government agencies can gain access to use their confidential data.)

Missouri

Missouri state education agencies maintain several administrative databases. For instance, the Missouri Department of Higher Education maintains student-level information on graduates from public higher education institutions. The Missouri Department of Elementary and Secondary Education maintains school-level K-12 information, as well as teacher-level data. Michael Podgursky, who is part of CALDER, has developed a panel data that matches higher education student-level data to teaching-level information. This dataset contains information on students enrolled in public 2- and 4-year colleges; it includes high school enrolment information, high school GPA scores, class rank, college entrance exams, college coursework, college major, and degree earned. Teacher information includes certification, salary, title, and demographic information. These data have been used to study labor-market decisions of recent college graduates, specifically their decision to enter the teaching profession (Podgursky *et al.*, 2004).

Washington

CALDER's website provides a description of a longitudinal dataset of Washington State that includes longitudinal teacher- and school-level information. However, official information about this dataset could not be found at the time of this article.

Other State and District Administrative Databases

In this section, we feature some additional longitudinal state and district databases that have been used to conduct empirical research. This list is by no means exhaustive. It intends to give an idea of the type of data that exists, and the steps some states have taken to build a longitudinal system.

Chicago

The Chicago public schools (CPS) database allows following students over time as long as they remain in the Chicago public school system. It includes test results (Iowa test of basic skills), students' demographic and family characteristics, language status, grade retention, summer school attendance, and special education eligibility. The district also maintains fiscal information, as well as teacher data that can be linked over time at the school level. Jacob and Lefgren (2004) studied the effect of remedial education on student academic achievement with CPS data.

Arizona

The Arizona Department of Education maintains longitudinally linked student-level data. These data contain the Stanford Achievement Test result, ninth edition (SAT9) starting in 1997, and demographic characteristics such as race, gender, grade level, number of absent days, years in the same district, and program participation. Solmon *et al.* (2001) used these data to evaluate the effect of charter schools on student academic achievement.

California

California does not have a statewide system to track students' performance over time. However, California has a myriad of databases that allow for school- or district-level longitudinal analyses, such as the standardized testing and reporting (STAR) program, the academic performance index (API) databases, and the California basic educational data system (CBEDS). California has already begun to take important steps toward building a student- and teacher-level longitudinally linked data systems. The state enacted Senate Bill 1453 to create the California longitudinal pupil achievement data system (CALPADS), and Senate Bill 1614 to create the California longitudinal teacher integrated data education system (CALTIDES).

Despite that a longitudinal data system is still not available statewide, the capability has existed in several school districts within the state. For example, Los Angeles Unified School District, Fresno Unified, Long Beach Unified, San Diego Unified, Oakland Unified, San Francisco Unified, Elk Grove Unified, and Santa Clara Unified, to mention some, have longitudinal student-level information going back, in some cases, to 1998. Some of these databases have already

Table 2 Links to international data systems

Dataset	Acronym	Link
Trends in international mathematics and science study	TIMSS	http://www.timss.org
Progress in international reading literacy study	PIRLS	http://pirls.org
Program for international student assessment	PISA	http://www.pisa.oecd.org

been used by researchers. For example, Parrish *et al.* (2006) used the Los Angeles Unified School District longitudinal data system to study the effects of Proposition 227 on the education of English learners.

Other states

Nevada and Alaska are following Florida's data system model and are in the process of implementing a state-wide data system. The state of Nevada has created the statewide management of automated record transfer (SMART) system, which tracks students over time. Alaska has also developed a student-level linkable database called the on-line Alaska school information system (OASIS).

International Data Systems

Even though this article has focused on the US, researchers might also want to carry out international comparisons. The Trends in International Mathematics and Science Study (TIMSS) compares math and science academic achievement of 4th and 8th graders around the world (in 2003, TIMSS involved 46 countries). It has been collected in 1995, 1999, 2003, and 2007. This dataset is complemented by the Progress in International Reading Literacy Study (PIRLS), which focuses, as its name suggests, on reading skills. PIRLS data for 2001 and 2006 are available. On the other hand, the Program for International Student Assessment (PISA) focuses on older students (15-year-olds), and includes assessments on math, science, reading, and the cross-disciplinary problem-solving subjects. More than 40 countries have participated in each of the three PISA studies carried out so far (i.e., 2000, 2003, and 2006). These three international data systems are somewhat equivalent to the topics that NAEP covers within the United States. Examples of research studies that have used these data are Hanushek and Woessmann (2006), and Afonso and Aubyn (2006). **Table 2** shows the link to each of these international datasets.

Conclusions

In this article we have centered our attention on US primary and secondary education data sources, and some major international data-collection efforts. However, additional longitudinal data sources can be often found in other countries. For example, Chile has been used to analyze the impact of a national voucher plan, Israel to evaluate the effect of teacher training and use of technology in the classroom, and Indonesia to study the impact of significant expansions of the education system. Of course, these are only some examples, but they highlight the importance of finding new sources of variation of education inputs and outputs beyond the frontiers of the United States.

See also: Empirical Research Methods in the Economics of Education; Theoretical Concepts in the Economics of Education.

Bibliography

- Acemoglu, D. and Pischke, J. S. (2001). Changes in the wage structure, family income, and children's education. *European Economic Review* **45**(4–6), 890–904.
- Afonso, A. and Aubyn, M. S. (2006). Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs. *Economic Modeling* **23**(3), 476–491.
- Altonji, J. G. (1995). The effects of high school curriculum on education and labor market outcomes. *Journal of Human Resources* **30**(2), 409–438.
- Angrist, J. D. (2004). American education research changes track. *Oxford Review of Economic Policy* **20**(2), 198–212.
- Angrist, J. D. and Guryan, J. (2004). Teacher testing, teacher education, and teacher characteristics. *The American Economic Review* **92**(2), 241–246.
- Ballou, D. and Podgursky, M. (1998). Teacher recruitment and retention in public and private schools. *Journal of Policy Analysis and Management* **17**(3), 393–417.
- Barrow, L. and Rouse, C. E. (2005). Causality, causality, causality: The view of education inputs and outputs from economics. *Working Paper Series*, Federal Reserve Bank of Chicago, WP-05-15.
- Bifulco, R. and Ladd, H. (2006). School choice, racial segregation and test-score gaps: Evidence from North Carolina's Charter school program. *Journal of Policy Analysis and Management* **26**(1), 31–56.
- Boyd, D., Lankford Loeb, S., and Wyckoff, J. (2005). The draw of home: how teachers' preferences for proximity disadvantage urban schools. *Journal of Policy Analysis and Management* **24**(1), 113–132.
- Brown, C., Duncan, G. J., and Stafford, F. P. (1996). Data watch: The panel study of income dynamics. *Journal of Economic Perspectives* **10**(2), 155–168.
- Carnoy, M. and Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis* **24**(4), 305–331.
- Clotfelter, C. T., Ladd, H., and Vigdor, J. (2004). Do school accountability systems make it more difficult for low performing schools to attract and retain high quality teachers? *Journal of Policy Analysis and Management* **23**(2), 251–271.
- Coleman, J. and Hoffer, T. (1987). *Public and Private Schools: The Impact of Communities*. New York: Basic Books.
- Coleman, J., Hoffer, T., and Kilgore, S. (1982). *High School Achievement: Public, Catholic, and Private Schools Compared*. New York: Basic Books.
- Dee, T. S. (2004). Are there civic returns to education? *Journal of Public Economics* **88**(9–10), 1697–1720.

- Dee, T. S., Evans, W. N., and Murray, S. E. (1999). Data watch: Research data in the economics of education. *Journal of Economic Perspectives* **13**(3), 205–216.
- Dee, T. S., Wei, H., and Jacob, B. A. (2006). The effect of school size on parental involvement and social capital: Evidence from the ELS:2002. *Brookings Paper on Education Policy 2006/2007* 77–97.
- Editorial Projects in Education Research Center (2006). The information edge: Using data to accelerate achievement. State Technology Report: California. *Education Week*.
- Ehrenberg, R. G., Goldhaber, D. D., and Brewer, D. J. (1995). Do teachers' race, gender, and ethnicity matter? Evidence from the national educational longitudinal study of 1988. *Industrial and Labor Relations Review* **48**(3), 547–561.
- Evans, W. N. and Schwab, R. M. (1995). Finishing high school and starting college: Do Catholic schools make a difference? *Quarterly Journal of Economics* **110**(4), 941–974.
- Figlio, D. (2002). Can public schools buy better-qualified teachers? *Industrial and Labor Relations Review* **55**(4), 686–699.
- Figlio, D. (2006). Testing, crime and punishment. *Journal of Public Economics* **90**, 837–851.
- Figlio, D. and Rueben, K. S. (2001). Tax limits and the qualifications of new teachers. *Journal of Public Economics* **80**(1), 49–71.
- Frazis, H. (1993). Selection bias and the degree effect. *Journal of Human Resources* **28**(3), 538–554.
- Goldhaber, D. D. and Brewer, D. J. (1995). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis* **22**(2), 129–145.
- Goldhaber, D. D. and Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservable on educational productivity. *Journal of Human Resources* **32**(3), 505–523.
- Hansen, J. (2008). The availability and transparency of education data in California. *Education Finance and Policy Winter 2008* **3**(1), 41–57.
- Hanushek, E. A., Kain, J., and Rivkin, S. (1998). *Teachers, Schools and Academic Achievement*. NBER Working Paper No. w3691. Cambridge, MA: National Bureau of Economic Research.
- Hanushek, E. A. and Rivkin, S. (2003). Does public school competition affect teacher quality? In Hoxby, C. M. (ed.) *The Economics of School Choice*. pp 23–47. Chicago, IL: University of Chicago Press.
- Hanushek, E. A. and Woessmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *Economic Journal* **116**, C63–C76.
- Hanushek, E. A., Kain, J., and Rivkin, S. (2002). *The impact of charter schools on academic achievement*. Working paper, 2002.
- Jacob, B. A. and Lefgren, L. (2004). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources in Chicago* **39**(1), 50–79.
- Kane, T. J. and Rouse, C. E. (1995). Labor-market returns to two- and four-year college. *The American Economic Review* **85**(3), 600–614.
- Kane, T. J., Rouse, C. E., and Staiger, D. (1999). *Estimating returns to schooling when schooling is misreported*. NBER Working Paper, No. 7235.
- Lankford, H., Loeb, S., and Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis* **24**(1), 37–62.
- Lee, V. E. and Smith, J. B. (1997). High school size: Which works best and for whom? *Educational Evaluation and Policy Analysis* **19**(3), 205–227.
- Loeb, S. and Strunk, K. (2003). The contribution of administrative and experimental data to education policy research. *National Tax Journal* **56**(2), 415–438.
- McEwan, P. (2008). Quantitative research methods in education finance and policy. In Ladd, H. F. and Fiske, E. B. *Handbook of Research in Education Finance and Policy*. New York: Routledge.
- McLaughlin, D. and Bandeira de Mello, V. (2002). *Comparison of state elementary school mathematics achievement standards, using NAEP 2000 Paper presented at the annual meeting of the American Education Research Association, New Orleans, LA*.
- Meyer, B. D. (1995). Natural and quasi-experiments in economics. *Journal of Business and Economic Statistics* **13**(2), 151–161.
- Murnane, R. J., Willett, J. B., Duhaldeborde, Y., and Tyler, J. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings. *Journal of Policy Analysis and Management* **19**(4), 547–568.
- Parrish, T. B., Merickel, A., Perez, M., et al. (2006). *Effects of the Implementation of Proposition 227 on the Education of English Learners, K-12: Findings from a Five-year Evaluation. Final Report for AB 56 and AB 1116, Submitted to the California Department of Education*. Palo Alto, CA: American Institutes for Research.
- Podgursky, M., Monroe, R., and Watson, D. (2004). The academic quality of public school teachers: An analysis of entry and exit behavior. *Economics of Education Review* **23**, 507–518.
- Sass, T. R. (2006). Charter schools and student achievement in Florida. *Education Finance and Policy* **1**(1), 91–122.
- Solmon, L., Paark, K., and Garcia, D. (2001). *Does Charter School Attendance Improve Test Scores? The Arizona Results*. Phoenix, AZ: Goldwater Institute.

Further Reading

- Chaplin, D. D., Turner, M. D., and Pape, A. D. (2003). Minimum wages and school enrollment of teenagers: A look at the 1990's. *Economics of Education Review* **22**(1), 11–21.
- Cho, D. (2007). The role of high school performance in explaining women's rising college enrollment. *Economics of Education Review* **26**(4), 450–462.
- Clofelter, C. T., Glennie, E., Ladd, H., and Vigdor, J. (2008). Teacher bonuses and teacher retention in low performing schools: Evidence from the North Carolina \$1,800 Teacher Bonus Program. *Public Finance Quarterly* **36**, 63–87.
- Dee, T. S. (1998). Competition and the quality of public schools. *Economics of Education Review* **17**(4), 419–427.
- Dee, T. S. (2000). The capitalization of education finance reforms. *Journal of Law and Economics* **43**, 185–214.
- Eide, E. R. and Showalter, M. H. (2001). The effect of grade retention on educational and labor market outcomes. *Economics of Education Review* **20**(6), 563–576.
- Eide, E. R., Brewer, D. J., and Ehrenberg, R. G. (1998). Does it pay to attend an elite private college? Evidence on the effects of undergraduate college quality on graduate school attendance. *Economics of Education Review* **17**(4), 371–376.
- Grubb, W. N. (1993). The varied economic returns to postsecondary education: New evidence from the class of 1972. *Journal of Human Resources* **28**(2), 365–382.
- Haller, E. J., Monk, D. H., Spotted Bear, A., Griffith, J., and Moss, P. (1990). School size and program comprehensiveness: Evidence from high school and beyond. *Educational Evaluation and Policy Analysis* **12**(2), 109–120.
- Harris, D. and Sass, T. (2007). *Teacher training, teacher quality and student achievement. Working Paper 3. National Center for Analysis of Longitudinal Data in Education Research*.
- Holmes, G. M., DeSimone, J., and Rupp, N. G. (2003). Does school choice increase school quality. *NBER Working Paper*, No. 9683.
- Leigh, D. E. and Gill, A. M. (1997). Labor-Market returns to community colleges: Evidence for returning adults. *The Journal of Human Resources* **32**(2), 334–353.
- Martorell, P. and McFarlin, I. (2007). *Help or hindrance? The effects of college remediation on academic and labor market outcomes*. Unpublished manuscript.
- Sandy, J., Gonzalez, A., and Hilmer, M. (2006). Alternative paths to college completion: Effect of attending a 2-year school on the probability of completing a 4-year degree. *Economics of Education Review* **25**(5), 463–471.

Zimmer, R., Buddin, R., Chau, D., *et al.* (2003). *Charter School Operations and Performance: Evidence from California*. Santa Monica, CA: RAND Education.

<http://www.nces.ed.gov> – National Center for Education Statistics (NCES).

<http://www.teacherpolicyresearch.org> – Teacher Policy Research.

<http://www.pubpol.duke.edu> – Terry Sanford Institute of Public Policy, Duke University.

<http://www.bls.gov/nls> – US Bureau of Labor Statistics.

Relevant Websites

<http://www.caldercenter.org> – CALDER website.

Empirical Research Methods in the Economics of Education

P J McEwan, Wellesley College, Wellesley, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Empirical research in the economics of education often seeks to identify causal relationships between policies or investments, and a variety of school or individual outcomes. Do additional years of schooling cause individuals' earnings to increase? Does attending private instead of public schools cause students' test scores to improve? This article summarizes the methods used by economists to address these and other causal research questions. It is based on the more extensive description of empirical methods in McEwan (2008). The methods include regression analysis, but they are increasingly supplemented by experimental and quasi-experimental research designs.

Common Research Terms

The results of cause-testing research, regardless of the method, are commonly judged by two criteria: internal validity and external validity (Shadish *et al.*, 2002). A result is internally valid when it identifies a believable causal link between a policy or program and an education outcome. The policy or program is often generically called a treatment. A causal result is externally valid when it can be generalized to modified versions of the policy treatment, to alternate measures of student outcomes, to diverse populations of students or schools, or to different policy contexts. This article mainly emphasizes the importance of improving internal validity.

The causal effect of a policy treatment is the difference between students' outcomes when treated, and the same students' outcomes when not treated (commonly referred to as the counterfactual). Research methods are employed to "create reasonable approximations to the physically impossible counterfactual" (Shadish *et al.*, 2002: 5). Researchers estimate counterfactual outcomes by identifying a separate group of untreated students, called a control group or comparison group (the former term is reserved for randomized experiments, but is used loosely). Members of the control or comparison group should be similar to their treated counterparts, in every respect but for exposure to the treatment.

In practice, the groups are often dissimilar, in ways that affect the outcomes of interest but have nothing to do with the treatment (e.g., students' families have different incomes, because higher-income families were more or

less likely to choose the treatment). If this occurs, then the outcomes of comparison of students incorrectly estimate the counterfactual. Thus, a simple difference in the average outcomes of treatment and comparison students can yield a misleading estimate of the treatment's causal effect, and is said to suffer from selection bias. Selection bias is a pervasive challenge to internal validity, and the research methods discussed in this article are aimed at lessening it.

Economists use two broad approaches, usually in combination, to ensure that members of treatment and control/comparison groups are similar, on average. First, they make statistical controls for observed differences between students, often using regression analysis. Second, they influence how students are assigned to treatment and control groups. Cause-testing research is often lumped into three broad categories: experimental, quasi-experimental, and nonexperimental. The essential difference among categories is the degree of control exerted by the researcher over who is assigned to the policy treatment (whether students, teachers, schools, districts, or states), and who is assigned to the control/comparison group.

In experimental research – often referred to as a randomized, controlled trial (RCT) – the assignment is entirely determined by luck of the draw, as in a researcher's flip of a coin. In quasi-experimental research, the broadest category of research, assignment may contain elements of randomness or purposeful assignment by the researcher, but some might be due to the individual choices of students, parents, or administrators (called selection). In nonexperimental research, the researcher exerts absolutely no influence, and assignment is entirely due to selection. When greater control is exerted, the causal results often possess greater internal validity.

Methods for Answering Causal Questions

Statistical Controls for Observed Variables

Suppose that researchers collect nonexperimental data from students who attend either private or public schools, but were not encouraged or coerced to do so by researchers. The causal question is whether attending private school improves test scores. A naïve researcher would simply estimate the difference between the average test scores of private students and public students, and ascribe

it to the causal effect of school type. But the difference could be explained by preexisting differences in students that are the result of selection. For one, private students in tuition-paying schools have higher incomes, on average, which might be associated with higher test scores.

Regression analysis

Regression analysis is the first line of defense against this kind of selection bias. Here, regression analysis implies ordinary least-squares (OLS) regression. In education policy, it is common to apply multilevel or hierarchical models (Raudenbush and Bryk, 2002) that model error components and account for the potential correlation of errors within classrooms, schools, communities, or states. In so doing, they avoid understating standard errors of coefficients and overstating their statistical significance – the models do not necessarily, as is sometimes assumed, remove selection bias. Economists are more likely to report OLS coefficient estimates accompanied by adjusted Huber–White standard errors that allow for arbitrary correlations among units within clusters (Wooldridge, 2002). In comparisons, OLS with adjusted standard errors and other multilevel models yield similar results, though OLS with standard errors not adjusted for clustering can dramatically underestimate standard errors (Angeles and Mroz, 2001). This issue is not discussed further, but the research cited in this article generally reports cluster-adjusted standard errors. A basic regression model can be written as:

$$A_i = \beta_0 + \beta_1 P_i + \beta_2 X_i + \varepsilon_i$$

where A represents the test score of each student in the entire sample (the subscript i might range from 1 to 1000); P indicates whether a student attends a private school ($P = 1$) or public school ($P = 0$); (This assumes a binary policy intervention (treated or not), though the discussion can be generalized to continuously measured policy interventions (e.g., class sizes of 1–50).) X indicates the value of a control variable, like family income, that one wishes to hold constant; and ε is an error term unique to each student. The error term captures the notion that test scores vary, for unobserved reasons, even among students attending the same school type and with the same incomes.

Using the method of OLS, researchers estimate the regression coefficients β_0 , β_1 , and β_2 . In the absence of controls for X , the estimate of β_1 would be interpreted as the average difference between private and public students' achievement (the naïve estimate from before). Upon controlling for X , it is the average difference holding constant family income. The immediate question is whether the difference can now be interpreted as the causal effect of private school attendance.

The causal interpretation rests on an assumption of regression analysis: that private school attendance (P),

controlling for X , is uncorrelated with positive or negative shocks in test scores captured in the error term, ε . (Formally, the assumption is that $\text{COV}(P_i, \varepsilon_i) = 0$.) What could produce such correlations? Suppose that an unmeasured variable, M , gauges parent motivation. Further suppose that children of motivated parents disproportionately attend private schools (M and P are positively correlated, even controlling for X) and that the children of motivated parents obtain higher test scores (M and A are positively correlated, even controlling for X).

In regressions that do not control for M , the net result is that attending private schools tends to be accompanied by positive shocks in students' test scores, the (noncausal) influence of greater unobserved motivation among their parents. In this example, estimates of the coefficient β_1 would be too big because of selection bias, leading to overly optimistic causal conclusions about private school effects. Yet, selection bias could also work in the opposite direction, depending on the sign of partial correlations between the excluded variable(s), the dependent variables (A), and the key treatment variable (P). Omitting variables creates no bias in β_1 if (1) the omitted variables are uncorrelated with A , or (2) the omitted variables are uncorrelated with P (The goal of randomized experiments is to ensure that condition (2) holds by design.)

In nonexperimental research settings, researchers have few remaining options. One is to collect and control for additional variables in the regression, but that is cold comfort to users of existing data sets. Even controlling for hundreds of variables proves unconvincing in many contexts. For example, the archetypal nonexperimental study in education policy regresses student test scores on student, family, and school variables. The third category often includes a key policy variable such as class size or student/teacher ratio. Yet, test scores (say, at the end of grade 6) are the cumulative product of thousands of family and school inputs received by students from birth onward (Todd and Wolpin, 2003). Data sets can never fully and accurately measure all inputs.

In test score regressions, researchers often resort to controlling for test score measurements taken at earlier moments in students' careers (say, at the beginning of grade 6). By controlling for pretests in so-called value-added regressions, researchers hope to implicitly control for the all inputs that affected test scores until that moment, thereby reducing the scope of bias. Nonetheless, there is no guarantee that omitted variables during grade 6, or earlier ones not captured by error-ridden pretests, do not continue to bias estimates.

Propensity score matching

An alternative method of controlling for observed variables is propensity score matching. Researchers first estimate a propensity score for each student (or other unit) in the sample (Rosenbaum and Rubin, 1983). The score is a

predicted probability that students receive a treatment, given their observed characteristics. So, in the prior example, researchers would estimate probabilities, using a probit or logit regression, that students attend a private school, given their family income (X) and other observed variables thought to influence propensities.

In a common approach, private students would then be matched to similar public students, based exclusively on values of their propensity scores. If students cannot be matched to a counterpart, then they are discarded from the sample (e.g., this might happen if children of millionaires, or with other observed characteristics, always attend private schools). Estimates of private school effects are based on comparisons of average outcomes across students in propensity-score matched treatment and control groups.

The method's virtues are at least twofold (Ravallion, 2005). First, it imposes no arbitrary assumption of linearity on the relationships between outcomes, policy variables, and other controls, as in most regression models. Second, it removes treated (or untreated) students from the samples that have no obvious match in the other group. Intuitively, the observed uniqueness of such students implies that they are also unique in unobserved ways that could introduce selection bias. Yet, like regression analysis, the causal interpretation of propensity score matching results rests on the unverifiable assumption that no unobserved variables are correlated with outcomes and with the probability of receiving a treatment. In this regard, it is no panacea for causal research.

Randomized Assignment

In the 1990s, economists became disenchanted with the ability of statistical controls for observed variables to eliminate selection bias in nonexperimental data. (A growing literature finds that nonexperimental statistical approaches, including regression and propensity score matching, do a poor job of replicating experimental results (Agodini and Dynarski, 2004; Glazerman *et al.*, 2003).) A turning point was the widespread analysis and debate of results from a randomized experiment in Tennessee that identified the causal effect of smaller class sizes on student test scores (e.g., Krueger, 1999).

In the classic instance of randomized assignment, researchers flip a coin to determine which students are treated, and which are not (note that teachers, schools, districts, or even entire towns could also be randomly assigned). Thus, each student's probability of receiving the treatment is an identical 0.5. (A coin flip or similar mechanism is only the simplest approach to designing randomized assignment. The essential point is that students or other units have well-defined probabilities of being assigned to the treatment.) The virtue of this approach is that it balances the distribution of students' observed and unobserved characteristics across treatment

and control groups. The two groups are not identical, of course, but they should be similar, on average. Control group members are similar and therefore, except for their exposure to a treatment, they provide an ideal counterfactual estimate of outcomes.

To obtain an internally valid estimate of causal effects, one estimates the mean difference between the outcomes of treated and untreated units. One could further apply regression analysis to control for observed differences between groups. If randomization proceeded without a hitch, then doing so is not strictly necessary to eliminate selection bias, although it reduces the standard errors of estimates of causal effects.

Besides class size, randomized experiments have been used to explore the effects of many policy treatments on student test scores. For overviews of methods and findings, see Borman (2002) and Glewwe and Kremer (2006). Randomized experiments are not without pitfalls (for varied opinions, see Burtless, 1995; Heckman, 1995). One common critique is that attrition from treatment or control groups could reintroduce selection bias into experimental estimates. To be fair, attrition is a widespread problem in social science research, and is hardly confined to randomized experiments.

A second critique is that experiments, especially small-scale ones, yield causal conclusions of limited external validity (Shadish *et al.*, 2002). As one example, it is unlikely that treatments affect all students similarly. A typical randomized experiment identifies the average causal effect among heterogeneous students (some of who are strongly affected, and others not at all). Researchers with large enough samples can estimate causal effects within subsamples of students, perhaps dividing them by location, income, or race. Yet, even average effects in experimental samples may not be generalizable to the average student in the entire population, since initial samples are not always a random draw. In fact, many research studies pragmatically begin with volunteers, either students or schools. Results from volunteer students or schools may be harder to generalize to the broader population.

Discontinuity Assignment

One of the most credible quasi-experimental methods is the regression-discontinuity design (RDD) (Hahn *et al.*, 2001; Shadish *et al.*, 2002). In the RDD, researchers assign students to treatment or control groups on the basis of a single assignment variable – often a test score, but potentially any continuous variable – and a specified cutoff value. To provide an illustration, suppose that a thousand students vie for college financial aid by taking a pretest (the assignment variable). Students with scores of 50 or above (the cutoff) receive aid, and those with scores below 50 do not. Note that the assignment is not randomized, as in the flip of coin, but neither is it due to selection.

This provides sufficient leverage to identify the causal effect of financial aid on some students' subsequent outcomes.

The causal effect is estimated by comparing the outcomes of treatment and control students whose values of the assignment variables are close to 50 (van der Klaauw, 2002). The intuition is that such students should be very similar, not just in their values of the pretest, but in other observed and unobserved ways. At the very least, observed and unobserved characteristics of the students should not vary sharply in the vicinity of the cutoff. In short, control students (just to the left of the cutoff) provide a good counterfactual estimate of outcomes for treated students (just to the right). Thus, any sharp – or discontinuous – changes in outcomes near the cutoff can be attributed to the financial aid treatment.

A hallmark of recent papers is that researchers do not specify cutoffs or implement the assignment process. Instead, researchers take advantage of cutoff-based assignment that administrators used to allocate resources in a transparent, fair, or efficient way (i.e., needy or meritorious students receive financial aid, low-scoring schools receive assistance or sanctions, and high-scoring ones receive rewards, less-effective teachers receive training, and so on). The unintended usefulness of such rules to researchers has only recently been noted in many cases, even when discontinuity assignment has a long history.

What are the potential pitfalls of using discontinuity-based assignment? The most serious, related to internal validity, is that students, or others subject to discontinuity assignment, are familiar with the potential intervention, the assignment variable, and the value of the cutoff. If they have incentives to receive the treatment, or not, then they may well attempt to manipulate their values of the assignment variable (Lee, 2008). As in the nonexperimental context, the concern is that manipulation may introduce selection bias into estimates of causal effects. For example, suppose that students with rich parents are aware of financial aid assignment rules and obtain extra pretest tutoring for their children. The result is that treated children, just to the right of the cutoff, also happen to be somewhat wealthier, and perhaps more likely to attend college even without the treatment.

Precise manipulation of many continuous assignment variables is actually harder than it might seem (Lee, 2008). While most families can probably influence their child's pretest score, random errors in testing make it unlikely that they can affect it within a very narrow band of scores around the assignment cutoff. Students within this narrow band contribute the most to estimates of causal effects, implying that assignment variable manipulation must be very precise to bias regression-discontinuity effects. To test for manipulation, researchers typically search for suspicious clustering of students on either side the cutoff. They also compare students' observed characteristics near the cutoff, which should vary smoothly across the break, in the absence of manipulation.

Instrumental Variables

In the majority of cases, the assignment of students, schools, or other units to policy treatments is neither random nor based on values of an observed assignment variable. Besides controlling for observed characteristics, what remaining methods are available to identify the causal effect of policies on outcomes? One of most popular among economists has been instrumental variables (IV).

In nonexperimental data, the receipt of policy treatments is usually correlated with unobserved characteristics of individuals that affect outcomes. Therein lies the empirical dilemma. Yet, some individuals in the sample might receive a treatment because of luck or because they were encouraged to do so for reasons unrelated to outcomes. The challenge is to base estimates of causal effects entirely on clean variation in treatment status – that is, variation uncorrelated with unobserved characteristics that affect outcomes.

One must identify an instrumental variable, or instrument, that fulfills two conditions (Wooldridge, 2002). First, it must be strongly correlated with the probability of receiving an intervention. This condition, straightforward to test in the data, is needed to ensure that the instrument actually induces students or schools to alter their treatment status. Second, the instrument cannot be correlated with unexplained variation in the outcome variable (i.e., the variation in outcomes that remains after controlling for other independent variables). The validity of the second assumption, more complicated to empirically test, usually rests upon the compelling reasoning of the researcher.

In applications to education policy, instruments are often related to features of geography or students' location, which are assumed to be random in some regard, and thus viable candidates to fulfill the second condition. Toward estimating private school effects, for example, Figlio and Ludwig (2000) show that the availability of subway transportation in metropolitan areas affects the probability that families, especially poorer ones, choose private schools. Using this as an instrumental variable, their analysis suggests that private school attendance has strong effects on reducing some risky teenage behaviors. Given the second IV condition, they must assume that transportation availability is uncorrelated with student outcomes, after controlling for other variables like family income. The validity of this condition is hard to prove. For example, do metropolitan areas with extensive subways have progressive mayors that invest in public schools? In the most convincing IV analyses, there are *a priori* reasons to believe that instruments are uncorrelated with unexplained outcomes.

Difference-in-Differences

Difference-in-differences (DD) methods attempt to control for unobserved variables that bias estimates of causal

effects, aided by longitudinal data collected from students, school, districts, or states. Researchers employ two varieties of longitudinal data. Panel data track the progress of the same students or teachers in successive months or years. Repeated cross-section data follow different groups of individuals (e.g., second-graders in successive years) that are clustered within the same schools, districts, or states.

The logic of DD causal inference is best communicated with an example based on repeated cross-section data. Of two states, Massachusetts and Maine, suppose that the former implements a finance reform – increasing state financing of local public school districts – and the latter does not (Dee and Levine, 2004). To estimate the reform's impact on district outcomes in Massachusetts, a naïve approach would compare outcomes across states, within a single year of post-reform data. The comparison is likely biased by selection, since unobserved differences across states could also affect outcomes.

Now consider the same comparison of outcomes, but within an earlier, pre-reform year of data. Evidently any differences in outcomes cannot be attributed to a Massachusetts reform that is yet to occur. Pre-reform differences in outcomes are perhaps due to unobserved differences across states that contaminated the previous, naïve estimate. To control for these unobserved variables, the DD estimate of the reform's effect subtracts the second difference from the first. The remaining DD could be plausibly attributed to the reform. For this to be credible, the change in Maine's outcomes must be a good counterfactual for Massachusetts' outcomes. Yet, suppose that Massachusetts' outcomes rose more quickly than other states, even before the reform, because of economic growth due to a strong biotech industry. The DD will nonetheless attribute faster outcome growth in the treated state to the causal effect of reform.

In light of this pitfall, one of the best ways to assess the internal validity of DD results is to compare the trends of outcome variables across treatment and control groups before application of the treatment. Evidence of similar trends bolsters confidence in the DD assumption. Dee and Levine (2004) estimated the effect of Massachusetts' state finance reform on districts' per-pupil state revenues. As controls, they used districts in Maine and Connecticut, which did not apply reforms, over the same pre- and post-reform period. DD estimates showed significant effects of the reform on local revenues. To support the use of these comparison groups, they showed that the outcome variables had similar trends in the three states in years prior to the reform.

Researchers increasingly apply DD methods to student-level panel data on test scores, applying a similar logic of causal inference. Some students' outcomes are observed before and after exposure to a treatment. Their outcomes are compared to students never exposed to the treatment. Rouse (1998) provides an example of this approach in her

re-analysis of data from the Milwaukee voucher program. The treatment group consists of students, observed before and after their selection to receive a private school voucher. The comparison groups consist of (1) students that were denied a private school voucher or (2) students in public schools that never applied for one. The DD estimates suggest that treated students have faster gains in math scores, but not in reading, than students in both comparison groups. The necessary assumption, as in previous analyses, is that treated students would have had trends in achievement similar to untreated students in the absence of the treatment.

Combining Methods to Improve Causal Inference

Researchers often apply multiple methods in the same study to improve internal validity. Almost every study employs statistical controls for family and student characteristics that affect outcomes. DD methods are frequently combined with experiments (Skoufias, 2005), the RDD (Chay *et al.*, 2005), and IV methods (Kuziemko, 2006). Finally, researchers often combine IV methods with randomized experiments and the RDD, especially to address imperfect compliance of students with random or cutoff-based assignment to policy treatments. In New York City's voucher experiment, for example, students were randomly assigned to receive a voucher offer, but not all students accepted the offer and actually attended a private school (Krueger and Zhu, 2004). To recover an estimate of the treatment-on-the-treated (i.e., the effect of actually attending a private school), researchers used the voucher offer as an instrument for private school attendance. The resulting IV estimate provides a credible estimate of the effect of private school attendance on those induced to accept it by the voucher offer.

Conclusions

This article has described empirical research methods used to estimate the causal effect of education investments on outcomes. Some, like regression analysis with nonexperimental data, are common but not always capable of delivering strong causal conclusions. Others, like experimental and discontinuity research designs, are increasingly common in the economics of education and better suited to addressing causal questions. Finally, it bears emphasis that good causal research is a necessary but not sufficient condition for designing and implementing good policy, a point not always clear in recent debates over scientific research (US Department of Education, 2003).

See also: Data; Theoretical Concepts in the Economics of Education.

Bibliography

- Agodini, R. and Dynarski, M. (2004). Are experiments the only option? A look at dropout prevention program. *Review of Economics and Statistics* **86**(1), 180–194.
- Angeles, G. and Mroz, T. A. (2001). *A Guide to Using Multilevel Models for the Evaluation of Program Impacts. Measure Evaluation Working Paper 01-33*. Chapel Hill, NC: Carolina Population Center/University of North Carolina.
- Borman, G. D. (2002). Experiments for educational evaluation and improvement. *Peabody Journal of Education* **77**(4), 7–27.
- Burtless, G. (1995). The case for randomized field trials in economic and policy research. *Journal of Economic Perspectives* **9**(2), 63–84.
- Chay, K. Y., McEwan, P. J., and Urquiola, M. (2005). The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review* **95**(4), 1237–1258.
- Dee, T. S. and Levine, J. (2004). The fate of new funding: Evidence from Massachusetts' education finance reforms. *Educational Evaluation and Policy Analysis* **26**(3), 199–215.
- Figlio, D. N. and Ludwig, J. (2000). *Sex, Drugs, and Catholic Schools: Private Schooling and Non-Market Adolescent Behaviors. Working Paper No. 7990*. Cambridge, MA: National Bureau of Economic Research.
- Glazerman, S., Levy, D. M., and Myers, D. (2003). Nonexperimental versus experimental estimates of earnings impacts. *Annals of the American Academy of Political and Social Science* **589**(1), 63–93.
- Glewwe, P. and Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. In Hanushek, E. and Welch, F. (eds.) *Handbook of the Economics of Education*, vol. 2, pp 945–1017. Amsterdam: Elsevier.
- Hahn, J., Todd, P., and van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* **69**(1), 201–209.
- Heckman, J. J. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives* **9**(2), 85–110.
- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics* **114**(2), 497–532.
- Krueger, A. B. and Zhu, P. (2004). Another look at the New York City voucher experiment. *American Behavioral Scientist* **47**(5), 658–698.
- Kuziemko, I. (2006). Using shocks to school enrollment to estimate the effect of school size on student achievement. *Economics of Education Review* **25**(1), 63–75.
- Lee, D. S. (2008). Randomized experiments from non-random selection in US House elections. *Journal of Econometrics* **142**(2), 675–697.
- McEwan, P. J. (2008). Quantitative research methods in education finance and policy. In Ladd, H. F. and Fiske, E. B. (eds.) *Handbook of Research in Education Finance and Policy*, pp 87–104. New York: Routledge.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edn. Thousand Oaks, CA: Sage.
- Ravallion, M. (2005). *Evaluating Anti-poverty Programs*. Policy Research Working Paper No. 3625. Washington, DC: World Bank.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**(1), 41–55.
- Rouse, C. E. (1998). Private school vouchers and student achievement: An evaluation of the Milwaukee Parental Choice Program. *Quarterly Journal of Economics* **113**(2), 553–602.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.
- Skoufias, E. (2005). *PROGRESA and Its Impacts on the Welfare of Rural Households in Mexico. Research Report No. 139*. Washington, DC: International Food Policy Research Institute.
- Todd, P. E. and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal* **113**(485), F3–F33.
- US Department of Education (2003). *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User Friendly Guide*. Washington, DC: US Department of Education/Institute of Education Sciences.
- van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression-discontinuity approach. *International Economic Review* **43**(4), 1249–1286.
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Further Reading

- Angrist, J. D. (2004). American education research changes tack. *Oxford Review of Economic Policy* **20**(2), 198–212.
- Angrist, J. D. and Krueger, A. B. (1999). Empirical strategies in labor economics. In Ashenfelter, O. and Card, D. (eds.) *Handbook of Labor Economics*, vol. 3A, pp 1277–1366. Amsterdam: Elsevier.
- Barrow, L. and Rouse, C. E. (2005). *Causality, Causality, Causality: The View of Education Inputs and Outputs from Economics. Working Paper 2005-15*. Chicago, IL: Federal Reserve Bank of Chicago.
- Campbell, D. T. and Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally.
- Duflo, E., Glennerster, R., and Kremer, M. (2006). *Using Randomization in Development Economics Research: A Toolkit*. Unpublished manuscript, MIT.
- Hanushek, E. A. (2002). Publicly provided education. In Auerbach, A. J. and Feldstein, M. (eds.) *Handbook of Public Economics*, vol. 4, pp 1197–2042. Amsterdam: Elsevier.
- Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* **142**(2), 615–635.
- Ludwig, J. (2001). Problems in the estimation of school effects: Insights from improved models. In Monk, D. H., Walberg, H. J., and Wang, M. C. (eds.) *Improving Educational Productivity (Research in Educational Productivity)*, vol. 1, pp 209–236. Greenwich, CT: Information Age Publishing.
- Meyer, B. D. (1995). Natural and quasi-experiments in economics. *Journal of Business and Economic Statistics* **13**(2), 151–161.
- Orr, L. L. (1999). *Social Experiments*. Thousand Oaks, CA: Sage.
- Ravallion, M. (2001). The mystery of the vanishing benefits: An introduction to impact evaluation. *World Bank Economic Review* **15**(1), 115–140.
- Shavelson, R. J. and Towne, L. (eds.) (2002). *Scientific Research in Education*. Washington, DC: National Academy Press.

Theoretical Concepts in the Economics of Education

D J Brewer and G C Hentschke, University of Southern California, Los Angeles, CA, USA
E R Eide, Brigham Young University, Provo, UT, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Over the past two decades, the economics of education has grown rapidly as a field. Previously, scholars and policymakers tended to view education and economics as separate realms, with economics applied to the study of private goods and education as a public good. Economics has been characterized as cold and impersonal due to its focus on firms, rational self-interested individuals, and cost-benefit decision making, all of which on the surface appear to be unrelated to the social and moral values associated with educating children. As school systems in developed countries have come under pressure to improve quality and scale, the distance between the two realms has narrowed. It is well documented that better-educated workers have more favorable labor-market outcomes than those with less schooling. Moreover, a well-educated labor force is critical for a nation to compete in an increasingly global economy that rewards knowledge and skills. Given concerns with the productivity of educational institutions and the fact that the study of incentives, choice, and competition lie at the heart of economics, economists have become more relevant to education-reform debates. They bring increased attention to resource allocation and decision making at the school level, take the view of educational organizations as potentially competitive enterprises, and those running them as entrepreneurs. The study of incentives lies at the heart of economics, and an understanding of how actors in large complex systems respond to incentives, and changes in incentives, helps shed light on how teachers might react to merit pay incentives or schools might react to increased competition from choice programs and charter schools.

In this article, we briefly review several of the most important theoretical concepts in the economics of education. First, we define economics and then review three of the most commonly used ideas – human capital, markets, and education production. We focus here on explicating the major underlying theories (in a nontechnical manner), rather than their application. All three have been utilized in numerous empirical studies. These are reviewed in greater depth elsewhere in the encyclopedia.

Economics Defined

Economics is often defined as “the study of the allocation of scarce means to satisfy competing ends” (Gary Becker

as quoted by Walberg and Bast, 2003:182). Economists study how individuals, organizations, and societies employ time, money, and effort. In the case of education, economists are interested in how society organizes and uses scarce resources to produce various types of knowledge and skills through formal schooling, and how these types of knowledge and skills are distributed to various groups in society. This broad definition means that many social and political issues and topics can fall under the purview of economics. John Maynard Keynes once wrote that economics was a “way of thinking” and it is this lens that has been brought to bear on a wide array of traditionally non-economics topics, including education policy.

Economists typically begin an explanation of observed phenomena by building a theory or a model in order to simplify reality and highlight key characteristics. A model contains a set of assumptions, and yields predictions, *ceteris paribus* (all other things being equal). Often this abstraction causes concern among noneconomists, but such simplifications are essential to understanding real-world settings. Economists would argue that what matters is whether the predictions of a model are correct on average rather than whether the assumptions underlying it are realistic. Economics, then, sits firmly within the tradition of theory-testing scientific method-based disciplines: a question is framed; a model/theory developed to explain behavior; and the hypotheses or predictions of that model/theory are then tested empirically using real-world data. It is often described as concerned with positive rather than normative issues, where the former are empirically testable and the latter are dependent on value judgments. The emphasis on hypothesis testing makes economists almost always use research designs that are quantitative in nature, attempting to discern whether predictions of cause and effect are valid, and the degree to which they are generalizable.

Economic theories are typically built on three basic foundations: scarcity, rationality, and optimization. Scarcity refers to the assumption that individuals and society will never have enough resources to completely satisfy their unlimited wants. Rationality refers to people's ability to make decisions in a systematic and purposeful way. It implies a “consistency of response to general economic incentives and an adaptability of behavior when those incentives change” (Ehrenberg and Smith, 2006: 4). The last assumption is the idea of optimization – either profit or goal maximization with reference to organizations or

utility maximization with reference to individuals. Individuals and groups have particular goals – be they happiness, profit, market share, or some combination of these or others – and will make choices that will maximize these benefits, subject to the constraints that they face (e.g., their income). This does not mean, however, that economists only care about selfish individuals; personal values are viewed more broadly, including all that individuals care about. Individuals behave subject to the constraints they face, the context in which they find themselves, and their perceptions of the consequences of alternative choices they make.

Economics provides a framework for understanding the behavior of individuals and organizations as they generate and allocate human, material, and financial resources. Using this perspective, economists have examined a wide range of education-related topics, and in the remainder of this article we discuss three of the major education questions of interest and the concepts that have been used to shed light on them. First, how much education (does and) should an individual acquire? This entails the notion of human capital. Second, how should education be produced and allocated by a society? This broad question examines conditions, characteristics, and behavior under alternative organizational forms, including both markets and hierarchies. Third, can we be more efficient and effective in organizing the production of education? The idea of ‘education production’ is helpful in answering this question. We discuss each of these in turn.

Human Capital

A primary research area within the economics of education is the association between schooling and individual outcomes, especially those associated with the labor market. Education (and training) is modeled as an individual investment decision that will receive a monetary return in the labor market, typically in the form of higher lifetime earnings. This notion of human capital has a rich history, with early economists such as Adam Smith, John Stuart Mill, and Alfred Marshall suggesting that individual’s skills could contribute to their economic status. In 1776, Smith laid the foundation for human capital theory when he wrote that human effort lies at the root of all wealth. In 1848, Mill built upon Smith’s notion; he considered human abilities as means to wealth (Sweetland, 1996). Modern-day human capital theory has further extended the central insight through the pioneering work of Schultz (1963), Becker (1964), and Mincer (1958, 1962).

Knowledge and skills acquired through educational investments increase human productivity. With each investment, one may incur costs in the form of out-of-pocket expenses, foregone earnings, and psychic costs associated with the pressure of studying and examinations.

Benefits accrue later in life through enhanced earnings in the labor market, access to better jobs, a higher likelihood of being employed, and better health. There are also psychic benefits from enhanced social status and the prestige associated with higher levels of education. Although individuals’ motivation for pursuing schooling may differ, and the psychic costs and benefits may be quite varied depending on personality, expectations of returns, and other traits, economists hypothesize that, other things equal, the more the education acquired, the higher the earnings achieved after the schooling is completed.

Prima facie evidence for human capital theory is to be found in the strong positive relationship between education levels and earnings that exist in almost every developed country. Generally, earnings rise with education level and they increase at an increasing rate in the immediate post education years, continue to increase at a slower pace, and then flatten as individuals approach retirement (Ehrenberg and Smith, 2006). This general pattern of earnings by education level holds for almost all subgroups, including men and women, and different racial and ethnic groups, but it is the differences among these groups that often fuels education policy debates about the distribution of education subsidies and services. Economists have devoted considerable attention to the challenge of estimating the returns to schooling taking account of these other factors. Analysis of such returns generally reveals a consistent positive relationship between investment in education and increased earnings for individuals, with an estimate of the average rate of return to an additional year of schooling of about 10% (Psacharopoulos and Patrinos, 2002). An overview of the empirical literature is provided elsewhere in the encyclopedia.

Economic research has also found nonmonetary benefits, both private and public, associated with educational attainment. Individuals who have invested in education and job training often have more job stability, improved health (e.g., exercise regularly, smoke less, and eat better), are more likely to receive employer-provided health insurance and pension benefits, are more inclined to vote, and have generally increased social and cultural capital that often enables upward mobility. These benefits are reviewed elsewhere in the encyclopedia.

Markets and Market Failure

There is general consensus that national investments in education lead to economic growth (for a review of the literature, see Sturm, 1993; Hanushek and Kimko, 2000). Countries spend a sizable percentage of their gross domestic product (GDP) on education each year. Educational spending can be undertaken by private individuals and by governments through public expenditures. The decision as to how education at different age levels should

be allocated is at its core an economic decision about how best to allocate scarce resources in order to maximize output (i.e., education).

Goods and services may get allocated in many different ways – for example, by tradition, force, or lottery. In modern societies, resources are allocated either by markets, by governments, or frequently through the interplay of both. K-12 schooling has traditionally been allocated by government at the federal, state, and local levels, with postsecondary education allocated by some combination of markets and government. As concerns about the effectiveness of existing schools have risen, policymakers have questioned the central role and functions of government in the allocation of educational resources, and turned toward market or market-related mechanisms.

A market is defined in an economic context as a collection of buyers who purchase and sellers who produce and sell goods and services; the interaction of buyers and sellers results in the possibility of exchange and, hence, in the allocation of goods and services. The transaction is facilitated through agreement on price. A graphical illustration of a market shows a downward-sloping demand curve and an upward-sloping supply curve. Sellers want to maximize profits, while buyers want to maximize satisfaction based on their preferences and budget constraints. The higher the market price, the more of a good or service a seller is willing to supply, but the lower quantity of that same good or service a buyer will demand, other things being equal. The function of a market is to adjust price to accommodate changes in supply and demand as efficiently as possible. When the price in a market reaches a level where the quantity that buyers want to purchase equals the quantity that sellers want to supply, then the market is said to be in equilibrium. Markets also act to keep prices low. Producers that fail to offer consumers what they want, or who charge too high a price, will lose business and eventually close. The dynamics of markets means a continuous process of adjustments that includes shortages and surpluses, and consumers and producers entering and exiting the market.

In many circumstances, markets are the preferred method for allocating resources because they are able to coordinate many buyers and sellers, give consumers considerable influence over price, characteristics, and quantity, and avoid relying on a handful of arbitrary decision makers. Under these circumstances, markets are an efficient mechanism for allocating resources, meaning that no more could be produced with the same resources, and the same output could not be produced with fewer resources. Efficiency is a specific criterion for judging an allocation mechanism. It does not say whether the resulting distribution of resources meets goals other than satisfying buyers and sellers (e.g., whether it is fair). Clearly, consumers of education have multiple goals (Gill *et al.*, 2001) and these need to be considered in deciding what is the

best mechanism for allocating education from society's standpoint.

When markets do not efficiently organize production or allocate goods/services to consumers, then market failure is said to occur. There are several reasons why markets fail. First, market power may arise when a supplier of a good/service has the ability to control price. A monopoly is an example of such market power. Perfectly competitive markets have many buyers and sellers, so no single buyer or seller has a big impact on price. While certain inputs to schooling may be more characteristic of perfect markets than others (e.g., school supplies), markets in K-12 schooling are quite imperfect.

A second type of market failure is when consumers have incomplete information about price and product quality, in which case the market cannot respond efficiently and correctly. Under incomplete information, parents may or may not choose schools based on outputs important for broader society. (There is some evidence, e.g., that many parents care not only about student achievement but also the social and racial profile of a school's students, preferring settings where there are most students like their own child.) Hence, although their preferences may be satiated in a market setting, some may judge that these preferences are not desirable from society's standpoint.

Third, externalities exist when consumption or production have an indirect effect on others that is not reflected in market prices. In the case of education, the decision maker (e.g., an individual student) does not bear all the costs or reap all the rewards from his or her decision about how much education to obtain. Even though society may benefit more from an educated person, the person making the educational decisions may not see those benefits as his or her own. Thus, the good (education) will be underconsumed from the perspective of the market. This presence of social benefits arising from basic education is perhaps the chief reason why governments have typically made schooling compulsory at elementary and secondary levels.

Fourth, markets may fail for public goods – those that can be made available to additional people without additional cost (nonrival), and once provided are difficult to prevent others from consuming (nonexcludable). Schooling is to some degree a public good. As with externalities, markets will tend to undersupply public goods.

The possibility of market failure, especially the underconsumption of education by private individuals from society's standpoint, as well as the importance of educational goals other than efficiency, has historically led to significant government intervention in the education sector through regulation, financing, and operation. Regulation can take different forms including setting safety standards, mandating curriculum or student assessments, and requiring teacher credentials. Financing can be in the

form of direct funding to schools or various forms of financial aid to individuals. Revenues for schooling may be generated from general taxation, rather than user fees, such that there is no clear relationship between receipt of the service and the payment for it. The government may also directly operate educational enterprises, which means that the delivery units are embedded within a larger government hierarchical infrastructure controlled by political mechanisms, owned by the state, and in which the employees are civil servants.

Typically, regulation, finance, and operation have been combined in a vertically integrated public sector system. Further, because government-operated schools have been designed to serve all students in a geographic area, in that locality they constitute a virtual monopoly. From a market perspective, this means that schools do not face competitive pressure to keep quality high and costs down. In addition, many families (particularly low-income and minority families) do not have much choice over the schooling options for their children. Recent educational reforms are to some extent characterized by an attempt to unbundle regulation, finance, and operation – ranging from tax credit schemes, to magnet schools, to controlled-choice programs, to charter schools, and voucher programs. Several of these are discussed in much greater depth elsewhere in the encyclopedia.

Education Production

Economists have sought to understand how education is produced. This has taken two different forms. One is to treat education as a production function wherein schooling inputs are processes from which outputs are produced. In this formulation of schooling, processes occur within a black box of the school system. The second approach explicitly looks inside the black box and examines the organization as a web of interpersonal contracts wherein individuals seek to coordinate others (and are in turn coordinated by others) in the performance of work. This latter arena is most often referred to either as transaction cost economics (accounting for the newly recognized costs of coordination or transactions) or as applications of principal–agent theory (seeking to capture the complex issues of delegation of decision rights between bosses or principals, and their subordinates or agents).

The production function approach uses an input–output framework to help think about schooling. The main inputs may include teachers, administrators, supplies, and facilities while the main outputs are student achievement (knowledge, skills). The relation between the educational inputs and outputs is usually statistically estimated using multiple regression techniques. While the education production function is simple in theory, it is very complex in practice (Goldhaber and Brewer, 1997). For example, it is hard to

identify and measure all inputs and outputs of schooling. Multiple outputs (e.g., basic skills, vocational skills, creativity, and attitudes) are valued, may accrue in a cumulative manner, and may only be discernable many years into the future. Inputs can be hard to measure, and the dimensions most easily measurable may not capture the important features of that input adequately. For example, the way in which a teacher interacts with students is important in the students' learning process; however, the characteristics of effective teachers may not be well captured by readily available proxies such as years of experience or qualifications. Nonschool inputs, such as peer influence and family background clearly affect how much students learn. Moreover, outputs are themselves joint products (i.e., students experience multiple teachers and carry with them knowledge from other classes and from home). The value of the production function approach, however, is as a framework for thinking about what resources, in which combinations, make a difference for student outcomes. Many studies have attempted to determine the relationship between inputs and outputs as currently exists in the United States and elsewhere.

The second economics-oriented perspective on organizations, achieved largely through applied principal–agent theory, was originally conceived by economist Nobel Laureate Ronald Coase in the early 1930s, who argued that markets and hierarchies, heretofore examined as separate topics, were in effect, substitutes for each other. The factors in a specific firm or division of a firm that made one alternative superior to another were often associated with the differing costs of coordination (the costs associated with transactions among individuals). This perspective has been extended to include a third substitutable form of organization, alliances or networks. In agency theory (Moe, 1984; Ferris and Winkler, 1991) principals (superiors in organizations, e.g., school superintendents) seek to ensure that agents (subordinates in organizations, e.g., school principals) carry out the principal's goals, in recognition of four primary factors that make this difficult. An adverse selection problem occurs when principals (e.g., school superintendents) are not fully informed about the abilities and values of the agents (e.g., school principals) and select agents that are not the best choice. A diverse objectives problem occurs when agents pursue their own objectives at the expense of pursuing the principals' objectives. This problem is compounded when compliance is achieved only by costly monitoring and controlling of the agents. An information asymmetry problem occurs when information within the accountability relationship is not evenly distributed. The agent typically has the information advantage. Finally, a weak incentives problem occurs when principals lack sufficient decision rights to cause the agents to either share principals' values or to behave as if they did. Although the full implications of agency theory and the

concepts embedded with it are still being explored, they have become common in analyses of current educational policies.

Concluding Thoughts

Economics, as a framework for understanding the allocation of resources, and human and organizational behavior more generally, has made important contributions to the study of education. The core concepts of human capital, markets, and education production have over several decades become increasingly common in education policy debates as well as familiar to practitioners. The contributions and the impact of economics on education (past and projected) are uneven. The concept of human capital, for example, is well established, and has expanded beyond simple rate of return calculations to considerations useful for policymakers – for example, in determining the relative contributions of different types of institutions or programs of study. The interplay between markets, regulation, and individual and organizational performance in education has attracted considerable academic interest, and that is expected to continue, albeit frustrated by the absence of large-scale demonstrations of different structures of the sort that would permit more definitive statements about what works. Applying economics to policies affecting the design of education organizations has been a mixed bag. With some exceptions, production function and cost effectiveness studies have had limited policy impact. Agency theory and transaction cost economics, intuitively appealing concepts, have not yet been rigorously applied to education organizations. However, as the education world begins to flatten this field should be expected to grow.

See also: Education and Economic Growth; Education Production Functions: Concepts; Human Capital; Returns to Education in Developed Countries; Returns to Education in Developing Countries; The Economic Role of the State in Education; The Economics of Charter Schools; The Efficacy of Educational Vouchers.

Bibliography

- Becker, G. S. (1964). *Human Capital – A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: National Bureau of Economic Research.
- Ehrenberg, R. G. and Smith, R. (2006). *Modern Labor Economics: Theory and Public Policy*, 9th edn. Reading, MA: Addison-Wesley.
- Ferris, J. and Winkler, D. (1991). Agency theory and decentralization. In Prud'homme, R. (ed.) *Public Finance with Several Layers of Government*, pp 155–166. The Hague: Foundation Journal.
- Gill, B. P., Timpane, P. M., Ross, K. E., and Brewer, D. J. (2001). *Rhetoric Versus Reality: What We Know and What We Need to Know About Vouchers and Charter Schools*. Santa Monica, CA: RAND.

- Hanushek, E. and Kimko, D. (2000). Schooling, labor force quality, and the growth of nations. *American Economic Review* **90**(5), 1184–1208.
- Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy* **66**, 281–302.
- Mincer, J. (1962). On-the-job training: Costs, returns and some implications. *Journal of Political Economy* **70**(supplement), 50–79.
- Moe, T. (1984). The new economics of organization. *American Journal of Political Science* **28**, 739–777.
- Psacharopoulos, G. and Patrinos, H. A. (2002). *Returns to Investment in Education: A Further Update*. Washington, DC: The World Bank.
- Schultz, T. W. (1963). *The Economic Value of Education*. New York: Columbia University Press.
- Sturm, R. (1993). *How Do Education and Training Effect a Country's Economic Performance? A Literature Survey*. Santa Monica, CA: RAND.
- Sweetland, S. R. (1996). Human capital theory: Foundations of a field of inquiry. *Review of Educational Research* **66**(3), 341–359.
- Walberg, H. J. and Bast, J. L. (2003). *Education and Capitalism: How Overcoming Our Fear of Markets and Economics Can Improve America's Schools*. Stanford, CA: Hoover Institution Press.

Further Reading

- Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for private schooling in Columbia: Evidence from randomized natural experiment. *American Economic Review* **92**(5), 1535–1558.
- Angrist, J., Bettinger, E., and Kremer, M. (2006). Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia. *American Economic Review* **96**(3), 847–862.
- Ashenfelter, O. and Krueger, A. (1994). Estimates of the economic return to schooling from a new sample of twins. *American Economic Review* **84**, 1157–1174.
- Ashenfelter, O. and Rouse, C. (1998). Income, schooling and ability: Evidence from a new sample of identical twins. *Quarterly Journal of Economics* **113**, 253–284.
- Barrett, N. S. (1991). Women. In Hornbeck, D. W. and Salamon, L. M. (eds.) *Human Capital and America's Future*, pp 61–94. Baltimore, MD: The Johns Hopkins University Press.
- Behrman, J., Rosenzweig, M., and Taubman, P. (1994). Endowments and the allocation of schooling in the family and in the marriage market: The twins experiment. *Journal of Political Economy* **102**, 1131–1174.
- Bettinger, E. P. (2005). The effect of charter schools on charter students and public schools. *Economics of Education Review* **24**(2), 133–147.
- Betts, J. and Hill, P. (2006). *Key Issues in Studying Charter Schools and Achievement: A Review and Suggestions for National Guidelines*. Seattle, WA: National Charter School Research Project Charter School Achievement Consensus Panel.
- Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J. (2005). The draw of home: How teachers' preferences for proximity disadvantage urban schools. *Journal of Policy Analysis and Management* **24**(1), 113–132.
- Brewer, D. J. (1996). Career paths and quit decisions: Evidence from teaching. *Journal of Labor Economics* **14**(2), 313–339.
- Brewer, D. J., Eide, E. R., and Ehrenberg, R. G. (1999). Does it pay to attend an elite private college? Cross-cohort evidence on the effects of college type on earnings. *Journal of Human Resources* **34**(1), 104–123.
- Brewer, D. J. and Wohlstetter, P. (2005). Charter schools come of age. *Urban Ed Fall/Winter* 15–19.
- Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling. In Christofides, L. N., Grant, E. K., and Swidinsky, R. (eds.) *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*, pp 201–222. Toronto, ON: University of Toronto Press.
- Card, D. (1999). The causal effect of education on wages. In Ashenfelter, O. and Card, D. (eds.) *Handbook of Labor Economics*, vol. 3, pp 1801–1863. Amsterdam: Elsevier.

- Carnoy, M. and Loeb, S. (2003). Does external accountability affect student outcomes? A cross-state analysis. *Education Evaluation and Policy Analysis* **24**(4), 305–331.
- Chay, K., McEwan, P. J., and Urquiola, M. (2005). The central role of noise in evaluating intervention that use test scores to rank schools. *American Education Review* **95**(4), 1237–1258.
- Cohn, E. and Geske, T. (1990). *The Economics of Education*. New York: Pergamon.
- Coons, J. and Sugarman, S. (1971). Family choice in education: A model state system for vouchers. *California Law Review* **59**, 321–438.
- Dale, S. B. and Krueger, A. B. (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *Quarterly Journal of Economics* **117**(4), 1491–1528.
- Dee, T. S. (2004). Teachers, race and student achievement in a randomized experiment. *The Review of Economics and Statistics* **86**(1), 195–210.
- Ehrenberg, R. G. and Brewer, D. J. (1994). Do school and teacher characteristics matter? Evidence from high school and beyond. *Economics of Education Review* **13**(1), 1–17.
- Ehrenberg, R. G., Brewer, D. J., Gamoran, A., and Willms, D. (2001). Does class size matter? *Scientific American* **285**(5), 66–73.
- Eide, E. R. (1994). College major choice and changes in the gender wage gap. *Contemporary Economic Policy* **12**(2), 55–63.
- Eide, E. R. (1997). Accounting for race and gender differences in college wage premium changes. *Southern Economic Journal* **63**(4), 1039–1050.
- Eide, E. R. and Showalter, M. S. (1998). The effect of school quality on student performance: A quantile regression approach. *Economics Letters* **58**(3), 345–350.
- Figlio, D. N. and Rouse, C. E. (2005). Do accountability and voucher threats improve low-performing schools? NBER Working Papers 11597, National Bureau of Economic Research, Inc.
- Friedman, M. (1962). *Capitalism and Freedom*. Chicago, IL: University of Chicago Press.
- Goldhaber, D. and Brewer, D. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources* **32**(3), 505–523.
- Goldhaber, D. D. and Brewer, D. J. (1997). Evaluating the effect of teacher degree level on educational performance. In Fowler, W. (ed.) *Developments in School Finance*, 1996, pp 197–210. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Goldhaber, D. D. and Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis* **22**(2), 129–145.
- Goldhaber, D. D. and Eide, E. R. (2002). What do we know (and need to know) about the impact of school choice reforms on disadvantaged students? *Harvard Educational Review* **72**(2), 157–176.
- Goldhaber, D. D. and Eide, E. R. (2003). Methodological thoughts on measuring the impact of private sector competition on the educational marketplace. *Educational Evaluation and Policy Analysis* **25**(2), 217–232.
- Greenwald, R., Hedges, L. V., and Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research* **66**(3), 361–396.
- Grogger, J. and Eide, E. R. (1995). Changes in college skills and the rise in the college wage premium. *Journal of Human Resources* **30**(2), 280–310.
- Hanushek, E. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis* **19**(2), 141–164.
- Hentschke, G. (2007). The role of government in education – Enduring principles, new circumstances and the question of self-life. In Ealy, L. and Endlow, R. (eds.) *Liberty and Learning: Friedman's Voucher Idea at Fifty*, pp 11–23. Washington, DC: Cato Institute.
- Hentschke, G. and Wohlstetter, P. (2004). Cracking the code of accountability. *Urban Ed* Spring/Summer, 17–19.
- Hilmer, M. J. (1998). Post-secondary fees and the decision to attend a university or a community college. *Journal of Public Economics* **67**(3), 348–372.
- Hornbeck, D. W. and Salamon, L. M. (eds.) (1991). *Human Capital and America's Future*. Baltimore, MD: The Johns Hopkins University Press.
- Hoxby, C. (1999). The productivity of schools and other local public goods producers. *Journal of Public Economics* **74**(1), 1–30.
- Hoxby, C. (2003). Does competition among public schools benefit students and taxpayers? *The American Economic Review* **90**(5), 1209–1238.
- Hoxby, C. (2004). School choice and school competition: Evidence from the United States. *Swedish Economic Policy Review* **10**, 11–66.
- Jacob, B. and Lefgren, L. (2004a). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics* **86**(1), 226–244.
- Jacob, B. and Lefgren, L. (2004b). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources* **39**(1), 50–79.
- Kane, T. J. and Rouse, C. (1993). Labour market returns to two- and four-year college: Is a credit a credit and do degrees matter? *American Economic Review* **83**(3), 600–613.
- Karoly, L. (2001). Investing in the future: Reducing poverty through human capital investments. In Danziger, S. H. and Haveman, R. H. (eds.) *Understanding Poverty*, pp 314–359. Cambridge, MA: Harvard University Press.
- Levitt, S. D. and Dubner, S. J. (2005). *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*. New York: HarperCollins.
- Loeb, S. (2001). Estimating the effects of school finance reform: A framework for a federalist system. *Journal of Public Economics* **80**(2), 225–247.
- Loeb, S. and Page, M. (2000). Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation. *Review of Economics and Statistics* **82**(3), 393–408.
- McEwan, E. K. and McEwan, P. J. (2003). *Making Sense of Research: What's Good, What's Not, and How to Tell the Difference*. Thousand Oaks, CA: Corwin Press.
- Murnane, R. J., Willett, J. B., and Levy, F. (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics* **77**(2), 251–266.
- Murray, S. E., Evans, W. N., and Schwab, R. M. (1998). Education finance reform and the distribution of education resources. *American Economic Review* **88**(4), 789–812.
- Ouchi, W. and Segal, L. (2003). *Making Schools Work: A Revolutionary Plan to Get Your Children the Education They Need*. New York: Simon and Schuster.
- Phelps, E. S. (2006). Dynamic capitalism. *Wall Street Journal* A14. <http://online.wsj.com/article/SB116043974857287568.htm> (accessed Jun. 2009).
- Rees, D., Argys, L., and Brewer, D. J. (1996). Detracking America's schools: Equity at zero cost? *Journal of Policy Analysis and Management* **15**(4), 623–645.
- Rouse, C. E. (1998). Schools and student achievement: More on the Milwaukee parental choice program. *Economics Policy Review* **4**(1), 61–78.
- Rouse, C. E. (2006). Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics* **90**(1–2), 239–255.
- Wolfe, B. and Zuvekas, S. (1997). Nonmarket outcomes of schooling. *International Journal of Educational Research* **27**(6), 491–501.
- Zimmer, R., Buddin, R., Chau, D., et al. (2003). *Charter School Operations and Performance: Evidence from California*. Santa Monica, CA: RAND.

ECONOMICS OF EDUCATION – COSTS AND FINANCING OF SCHOOLING

Contents

Cost–Benefit Analysis and Cost–Effectiveness Analysis

Economic Approaches to Adequacy

Economic Approaches to School Efficiency

School Finance Reform

School Finance: An Overview

The Economics of Tuition and Fees in Higher Education

Cost–Benefit Analysis and Cost–Effectiveness Analysis

C Belfield, Queens College, City University of New York, Flushing, NY, USA

H M Levin, Teachers College, Columbia University, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Economic Evaluation of Education

Cost–benefit analysis (CBA) and cost–effectiveness analysis (CEA) are the two main methods used by economists to evaluate educational investments.

CBA is an important method for evaluating educational investments and for deciding on the provision of public goods. CBA compares the costs and outcomes of alternatives. Critically, these outcomes must be expressed in monetary terms (Mishan, 1976). Both costs and benefits must be expressed in present values, that is, monetary values that take account of the difference between income received now and received in the future. At its most general level, CBA has been applied widely to evaluating educational investments at the national and subnational levels, most commonly in evaluating the ability of such investments to raise productivity and earnings.

CBA enables a direct comparison of the costs and benefits of an alternative or a comparison of their magnitudes with those of other types of social investments in education or in other sectors. Not only is it possible to ascertain which educational investment has the largest benefits relative to costs, but it is also possible to compare these results with investments in health, transportation, physical capital, such as plant and equipment, and other sectors. In this way, governmental units can use CBA to compare the desirability of alternative educational investments, as well as to determine the balance between investing in education and other sectors.

CEA is a method of comparing decision alternatives in which both the costs and the effects are taken into account in a systematic way (Levin and McEwan, 2001). The

method is used to compare the efficiency of alternative ways to achieve an educational objective. For example, if the objective is to raise the rate of high school graduation, this method of analysis can be used to determine whether investment in (as examples) preschool reading programs or smaller elementary school classes is more cost effective. In principle, the method is simple for each investment, the full costs are calculated (C) and the effect on graduation rates is identified (E); and so the most cost-effective investment is the one with the lowest C/E ratio.

Clearly, CBA and CEA are closely related to each other. The essential difference is that for CBA, the benefit must be measured in money terms and for CEA, the effect is measured directly in educational results. For example, vocational education programs are often solely intended to raise future earnings and are amenable to CBA. Alternatively, educational programs may be designed to improve, for example, reading skills as a desirable end per se, whether or not there is an associated monetary value. Sometimes, the effects of an educational program can be translated into monetary benefits: for example, a program that has the effect of raising the high school graduation rate will most likely also yield a benefit of thousands of dollars in extra income for the new graduates. The advantage of CBA is that – because the metric used is money – many more educational programs can be compared with each other. A program to boost reading skills may be compared to one that reduces grade retention rates; and more generally, investments in education can be compared to other social investments. However, many educational alternatives are dedicated to improving achievement or some other educational outcomes that cannot be easily converted into

monetary terms. When this is the case, CEA is the more appropriate method.

The purpose of both these forms of economic evaluation in education is to ascertain which program can achieve a set of particular objectives at the lowest cost. The implicit assumption is that different alternatives are associated with different costs and different educational effects or monetary benefits. By choosing those with the least cost for a given effect or benefit, society is using its resources more efficiently. That said, it is important to know if economic differences are large or small. If the differences are small, it may be wise to weigh more fully other criteria in deciding between educational investments (such as ease of implementation).

Methodology

The basic technique is to measure effects or benefits for each educational alternative using high-quality evaluation procedures, that is, ones that can causally identify consequences or ones that apply rigorous econometric techniques (Levin and McEwan, 2001). Separately, the costs for each alternative should be measured using the ingredients method (Levin and McEwan, 2001); these costs should include all the resources that are given up by society to implement the intervention. In practice, measurement of both cost and benefit/effects is challenging.

In order to adjust for the time pattern of costs, a discount rate is used to make past, present, and future costs comparable. A discount rate does not adjust for inflation in prices. What it does do is enable one to compare costs that have different time patterns by establishing their present values (Levin and McEwan, 2001: 90–94). The specific discount rate for public investments is a matter of some contention, although Moore *et al.* (2004) provide a strong argument for using a rate of 3.5%.

Measuring Costs

CBA and CEA both require that the costs of the investment are accurately measured using the ingredients or resource-recovery method (Levin and McEwan, 2001). The costs of an intervention are defined as the value of the resources that are given up by society to implement the intervention. These resources are referred to as the ingredients of the intervention, and it is the social value of those ingredients that constitute its overall cost. At a later stage, one can assess the distribution of these costs between the decision-making agency and other agencies. Accordingly, the method sets out systematically to identify and ascertain the value of the ingredients that are required for each alternative that is under consideration. From an economic perspective, the costs of an action are determined by the value of the resources that are entailed in their best alternative use.

This is known as opportunity cost, in the sense that it is the value of the best forgone opportunity that must be considered when one refers to the cost of the intervention. The ingredients method represents a straightforward way of estimating cost.

The ingredients approach to cost estimation entails three distinct phases: (1) identification of ingredients; (2) determination of the value or cost of the ingredients; and (3) an analysis of the costs in an appropriate decision-oriented framework.

The identification of the ingredients refers to the delineation of all of the resources that are required for the particular intervention that forms the basis for the economic analysis. For example, in the case of a study of expansion of technical education, the personnel, facilities, materials, and other ingredients required for the activity need to be identified. For educational programs, the main ingredient is often labor services of which the largest component is that of the teacher. Yet, teachers and administrators must prepare classes as well as manage the educational enterprise; so the ingredients are more than just their instructional time.

Educational interventions use not only these resources but also the time of the students who might otherwise be able to use their time productively in the workplace or through self-employment. Even the time of elementary school children is likely to have some value in developing countries in terms of household or farm tasks. If a student is required to spend four additional years in the educational system, those years could have been used for other productive activities that, when forgone, have a cost to both the individual and to society. Hence, student time is an important ingredient in economic analysis.

Once the ingredients have been identified and stipulated, it is necessary to ascertain their costs (Levin and McEwan, 2001). In doing this, all ingredients are assumed to have a cost, including donated or volunteer resources. That is, they have a cost to someone, even if the sponsoring agency did not pay for them in a particular situation. At a later stage, the costs will be distributed among the constituencies who paid them, but at this stage the aim is to ascertain the total costs of the intervention.

Ingredients can be divided into those that are purchased in reasonably competitive markets, and those that are obtained through other types of transactions. In general, the value of an ingredient for costing purposes is its market value. In the case of personnel, the market value may be ascertained by determining what the costs would be for hiring a particular type of person. Such costs must include not only salary, but also fringe benefits and other employment costs that are paid by the employer. Many of the other inputs can also be estimated using their market prices.

There are a number of techniques for determining the values of resources that cannot be readily priced in the market place. To cost out student time, the usual measure

is the earnings forgone by the student by enrolling in school or a training program rather than being employed in the workplace. This should approximate to the value of productive activity that was lost because the student was in school. Typically, the cost of student time is estimated from the earnings of persons of similar gender, race, and socioeconomic status who are in the workforce rather than in schools or training programs. Such estimates must be adjusted for the probability of unemployment.

However, the quality of costs data is often poor. Although the market prices of some ingredients can be obtained from accounting data for educational enterprises, such data are not reliable sources for ascertaining overall social costs. The accounting systems that are used by schools are designed for ensuring consistent reporting to state agencies rather than for providing accurate and consistent costs data on specific educational interventions. Important discrepancies are the costs of volunteers' time and other donated resources, as well as the costs of capital facilities that are typically charged to budgets at the time of expenditure rather than amortized over the life of the school. Besides, the issue of who pays the costs is important: decision makers may pay only part of the cost for one intervention, but the whole cost for another intervention.

Lastly, the issue of scale is important. In general, those alternatives with high fixed costs such as those with large investments in facilities and equipment will require a high enrolment or utilization to be most efficient. In contrast, alternatives that are constituted largely of variable costs such as personnel will have costs that are less sensitive to the scale of output. Thus, an economic evaluation of alternatives that differ in terms of their intensities of fixed versus variable costs may produce very different results depending on the scale of enrolment or output.

Benefits and Effects

Since CBA requires that the benefits be expressed in monetary units, it is only possible to apply this to interventions where this is feasible. CBA lends itself especially well to those alternatives or interventions in which the outcomes are market oriented. For example, educational and training programs that are designed to improve earnings or reduce poverty can be evaluated with a cost-benefit approach when the benefits are the additional earnings associated with the interventions. But all benefits must be considered. Yet, Herrnstein (1997) points out a saliency mismatch, whereby it is hard to envisage some of the benefits (or even some of the costs) and this may falsely diminish their significance.

Some studies have looked specifically at the monetary benefits of education, without directly relating these to the costs of a particular program that might influence education. An important early study was by Haveman and Wolfe (1984) which itemized the potential economic

benefits of education and derived a simple method by which these could be monetized. Research in an edited volume by Belfield and Levin (2007) shows in more detail the link between high school graduation and earnings, health, welfare, and crime; this volume also calculates the present value benefits for each additional high school graduate that is produced in the United States—these benefits are substantial.

For cost effectiveness, it is necessary to measure effects. For some programs, the average effect is of interest; for others it is the marginal effect of a program. In addition, many programs have multiple effects and these effects may not be found for all students. For example, some students may have higher test scores and others may exhibit improved classroom behavior or social skills. Deciding which effect is more important—or placing weights on each—is a challenge that might be undertaken using utility analysis (Levin and McEwan, 2001).

Examples of Applications

Cost-benefit analysis

The most popular area for CBA in education is that of preschool programs. Belfield *et al.* (2006) conducted an updated CBA of the High/Scope Perry Preschool Program, using data on individuals aged 40. Delivered in the 1960s, this program involved random assignment to a treatment or control group; the treatment was a preschool program. The treatment group received program provision for one or two short academic years composed of three parts: (1) a center-based program for 2.5 h per day for each weekday, with a child:teacher ratio of 5:1; (2) home visiting for 1.5 h⁻¹ weekday; and (3) group meetings of parents. Overall, the program represented a relatively intensive and structured investment in supporting the preschool development of the participants. The participants entered the study as 3- and 4-year-old African-American children, with no physical handicaps, but were selected on the basis of low parental education and disadvantaged background. Program costs were compared against treatment impacts on educational resources, earnings, criminal activity, and welfare receipt where these impacts were expressed in money terms. Net present values were calculated for participants, the general public, and society. Treatment group individuals obtained significantly higher lifetime earnings. For the general public, higher tax revenues, lower criminal justice system expenditures, and lower welfare payments easily outweighed program costs; they repaid US\$12.87 for every US\$1 invested. However, program gains were mainly in the form of reduced crime by males.

CBAs of preschool interventions have also been performed for other preschool programs, such as the Chicago Child-Parent Center program. This too showed a very high cost-benefit ratio: the total benefits were 10.15 times greater than the costs (Reynolds and Temple, 2007).

Another program, which also involved more comprehensive child care for younger children, is the Abecedarian Early Childhood Intervention. It also had a positive net present value where benefits greatly exceed costs (Barnett and Masse, 2007).

Potentially, CBA could be performed on any educational investment. For example, for the US, another example is by Bishop (1996), who estimates the costs and benefits of lengthening the school day and finds that the net present value of benefits is approximately five times the costs. Increasingly, CBA is being performed on educational investments in developing countries to guide state governments. However, ensuring that these economic evaluations are of high quality is far from straightforward (see Vawda *et al.*, 2003). An early attempt for rural Brazil – comparing different resource reallocation strategies – was performed by Harbison and Hanushek (1992). More recently, Machin and McNally (2007) calculated a benefit–cost ratio of the National Literacy Strategy in England; this reform changed the way literature was taught (with a focus on classroom management and effective teaching). As it was beneficial but relatively low cost, the reform yielded a very high benefit–cost ratio.

Further application of CBA

CBA is slowly being implemented, but it is still not a straightforward method to apply (e.g., for comprehensive school reform, see Levin, 2002). The main advantage from additional years of education appears to be in terms of the extra income a person can earn. However, this type of analysis is often done in the form of a rate-of-return analysis rather than a cost–benefit ratio or net present value (see Krueger's (1999) rate of return calculation for reducing class sizes in the elementary grades). That is, the income benefits are expressed in terms of the equivalent of an annual return on investment. Nevertheless, the basic data used to estimate rates of return can be used to calculate net present values (the difference between the present values of benefits and costs) or cost–benefit ratios.

Cost–effectiveness analysis

CEA can be applied to any educational interventions that have broadly the same objective. However, in practice its application has been very limited, not least because of the methodological challenges described above. Yet, this is in contrast to health economics, where CEA is much more prevalent (Muennig, 2002).

Curriculum is an area that is very appropriate for CEA. In the quest for educational improvement, schools face numerous ways in which they can organize the mode of instruction. Some are likely to be more effective than others, and there may also be substantial differences in resource requirements with respect to ingredients such as teacher time, materials, equipment, and so on. An excellent cost–effectiveness study of fifth-grade mathematics

curriculum was performed by Quinn *et al.* (1984). Their evaluation compared two approaches to teaching mathematics, a traditional curriculum and an alternative one. Using the ingredients method, the resources found that the alternative math program cost about 50% more than the traditional program, but the effectiveness of the alternative program was also higher. Depending on how student achievement was measured, the alternative program was found to be from 60% to 300% more cost effective (cost per point of achievement score) than the traditional program.

Levin *et al.* (1987) investigated four reforms for which there were adequate effectiveness data as well as a basis for estimating costs. These were: computer-assisted instruction (10 min of drill and practice per day); a longer school day (of 1 h divided equally between instruction in math and reading); smaller class sizes (changes in class size of various ranges from 35 to 20); and peer tutoring (of second-grade students for 15 min^{−1} day by fifth- and sixth-grade students).

Costs were estimated using the ingredients approach. The least costly interventions were reductions in class size of five students and increasing the length of the school day. The most expensive was peer tutoring because it requires adult coordinators. The costs of peer tutoring were about four times as great as reducing class size by five students or increasing instructional time by one-half hour per day per subject and twice as great as the daily sessions of computer-assisted instruction. In terms of effectiveness, peer tutoring showed the largest effects, nearly a year of achievement gain in math and about one-half year in reading. The longer school day and reduction of class size by five students showed the smallest effects. Computer-assisted instruction was associated with gains in the middle of the range of results. Cost–effectiveness was determined by combining costs and effectiveness in terms of the estimated annual costs to obtain an additional month of student achievement per year of instruction. Peer tutoring was the most cost-effective intervention and a longer school day the least cost effective; the former was almost ten times more cost effective in months of math achievement than the latter (and almost double for reading achievement). Computer-assisted instruction was almost as cost-effective as peer tutoring in terms of reading achievement, but reducing class size was very inefficient with respect to reading gains. Although class-size-reduction policies are very effective, they are also very expensive.

Further application of CEA

McEwan (2002) describes the many potential pitfalls and itemizes a set of technical criteria with which to evaluate cost–effectiveness studies. Clune (2002) evaluates the method as applied in 541 studies that refer to cost effectiveness; in over three-quarters of the studies the CEA is

rhetoric or minimal. A more recent exercise by Ross *et al.* (2007) yields a similarly low number of high-quality economic evaluations. The reasons for this are described in detail by Levin (2001).

See also: Economic Approaches to School Efficiency; Empirical Research Methods in the Economics of Education; Returns to Education in Developed Countries; Returns to Education in Developing Countries; The Economics of Class Size; The Economics of Early Childhood Interventions.

Bibliography

- Barnett, W. S. and Masse, L. N. (2007). Comparative benefit-cost analysis of the Abecedarian program and its policy implications. *Economics of Education Review* **26**, 113–125.
- Belfield, C. R. and Levin, H. M. (eds.) (2007). *The Price We Pay: The Costs to the Nation of Inadequate Education*. Washington, DC: Brookings Institution Press.
- Belfield, C. R., Nores, M., Barnett, W. S., and Schweinhart, L. (2006). The Perry pre-school 40-year follow-up cost-benefit analysis. *Journal of Human Resources* **42**, 215–246.
- Bishop, J. H. (1996). Incentives to study and the organisation of secondary instruction. In Becker, W. E. and Baumol, W. J. (eds.) *Assessing Educational Practices: The Contribution of Economics*, pp 99–160. London: Russell Sage.
- Clune, W. (2002). Methodological strength and policy usefulness of cost-effectiveness research. In Levin, H. M. and McEwan, P. J. (eds.) *Cost-Effectiveness Analysis and Educational Policy*, pp 55–70. Larchmont, NY: AEFA Yearbook: Eye on Education.
- Harbison, R. and Hanushek, E. A. (1992). *Educational Performance of the Poor: Lessons from Rural Northeast Brazil*. New York: Oxford University Press.
- Haveman, R. and Wolfe, B. (1984). Schooling and economic well-being: The role of non-market effects. *Journal of Human Resources* **19**, 377–407.
- Hernstein, R. J. (1997). *The Matching Law*. Beverly Hills, CA: Russell Sage Foundation.
- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics* **CXIV**, 497–532.
- Levin, H. M. (2001). Waiting for Godot: Cost-effectiveness analysis in education. In Light, R. (ed.) *Evaluation Findings That Surprise*. New Directions for program Evaluation, 90, 55–68.
- Levin, H. M. (2002). Issues in designing cost-effectiveness comparisons of whole school reforms. In Levin, H. M. and McEwan, P. (eds.) *Cost-Effectiveness Studies in Education, Yearbook of the American Educational Finance Association*, pp 71–96. Larchmont, NY: Eye on Education.
- Levin, H. M., Glass, G. V., and Meister, G. (1987). A cost-effectiveness analysis of computer-assisted instruction. *Evaluation Review* **11**, 50–72.
- Levin, H. M. and McEwan, P. J. (2001). *Cost-Effectiveness Analysis: Methods and Applications*. Thousand Oaks, CA: Sage.
- Machin, S. and McNally, S. (2007). The literacy hour. *Journal of Public Economics* **92**, 1441–1462.
- McEwan, P. J. (2002). Are cost-effectiveness methods used correctly? In Levin, H. M. and McEwan, P. J. (eds.) *Cost-Effectiveness Analysis and Educational Policy*, pp 37–54. Larchmont, NY: AEFA Yearbook: Eye on Education.
- Mishan, E. J. (1976). *Cost-Benefit Analysis: An Informal Introduction*. New York: Praeger.
- Moore, M. A., Boardman, A. E., Vining, A. R., Weimer, D. L., and Greenberg, D. H. (2004). Just give me a number! Practical values for the social discount rate. *Journal of Policy Analysis and Management* **23**, 789–812.
- Muennig, P. (2002). *Designing and Conducting Cost-Effectiveness Analysis in Health and Medicine*. San Francisco, CA: Jossey-Bass.
- Quinn, B., VanMondfrans, A., and Worthen, B. R. (1984). Cost-effectiveness of two math programs as moderated by pupil SES. *Educational Evaluation and Policy Analysis* **6**, 39–52.
- Reynolds, A. J. and Temple, J. A. (2007). Benefits and costs of investments in preschool education: Evidence from the Child-Parent Centers and related Programs. *Economics of Education Review* **26**, 126–144.
- Ross, J. A., Barkoui, K., and Scott, G. (2007). Evaluations that consider the cost of educational programs: The contribution of high quality studies. *American Journal of Evaluation* **28**, 477–492.
- Vawda, A. Y., Moock, P., Gittinger, J. P., and Patrinos, H. A. (2003). Economic analysis of World Bank education projects and project outcomes. *International Journal of Educational Development* **23**, 645–660.

Further Reading

- Behrman, J. and Stacey, N. (1997). *The Social Benefits of Education*. Ann Arbor, MI: University of Michigan Press.
- Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review* **91**, 795–813.
- Levin, H. M. and McEwan, P. J. (eds.) (2002a). *Cost-Effectiveness Analysis and Educational Policy*. Larchmont, NY: AEFA Yearbook: Eye on Education.
- Levin, H. M. and McEwan, P. J. (2002b). Cost-effectiveness and educational policy. In Levin, H. M. and McEwan, P. J. (eds.) *Cost-Effectiveness Analysis and Educational Policy*, pp 1–20. Larchmont, NY: AEFA Yearbook: Eye on Education.
- McMahon, W. W. (1998). Conceptual framework for the analysis of the social benefits of lifelong learning. *Education Economics* **6**, 309–347.
- World Bank (2006). Early childhood education and development in Indonesia: An investment for a better life. *World Bank Policy Research Working Paper* (37531).

Economic Approaches to Adequacy

J Imazeki, San Diego State University, San Diego, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cost of adequacy – It is the minimum amount of money that school districts must spend in order to achieve some minimum-outcome goal.

Educational cost function – It is the estimated relationship between dollars per pupil, educational performance, and the characteristics of the school district and its student body.

Marginal costs of adequacy – They are the additional costs associated with specific student or district characteristics (such as poverty, English learners, and special education), above and beyond the base cost of adequacy in a district with none of these special needs.

Throughout the 1970s and early 1980s, school finance court cases and state-level school finance reforms across the United States focused primarily on the equalization of resources across districts. The underlying assumption of these efforts seemed to be that a more equal distribution of dollars would lead to a more equal distribution of educational outcomes. However, during the 1980s and 1990s, a growing awareness that equal dollars do not lead to equal outcomes led to new court challenges in several states, based on adequacy grounds. In these cases, plaintiffs maintained that states were required to ensure students had equal access to an adequate education, with adequacy defined in terms of some minimum standard of student performance or skills. In particular, these cases recognized the differential costs of education for schools serving different student populations, highlighting the fact that some schools require more resources than others to meet the same performance standard.

The cost of education can be defined as the minimum amount of money that a school district must spend in order to achieve a given educational outcome, such as reading at a grade-appropriate level. Costs generally differ across school districts for reasons that are outside the control of local school boards or state governments, such as the number of children with special needs. All else equal, districts with higher costs will need to spend more than districts with lower costs in order to achieve any given outcome. Thus, costs are specifically linked to outcomes. Costs are also distinct from expenditures. Actual expenditures in a district are influenced by costs

but they also reflect choices made by local school boards or district officials regarding how to allocate and organize resources to achieve selected objectives. Thus, a school district with below-average costs could have above-average expenditures because it chooses to offer a particularly broad range of advanced classes, or because it is relatively inefficient in its use of resources.

Measuring the Costs of Adequacy

Designing a school finance system that guarantees all students an adequate education requires information on the costs of providing that education in each district. To that end, an increasing number of costing-out adequacy studies have now been conducted in a large number of US states. (The literature discussed here focuses on cost studies conducted with data from the United States. For a review of the literature on costs in other countries, see Tsang (1988, 1997).) Researchers have used a number of different methodologies to estimate educational costs, but most studies involve one of four methodological approaches: professional-judgment, evidence-based, successful-schools, or the cost-function (or econometric) approach. As discussed extensively in Baker *et al.* (2004) and Duncombe *et al.* (2004), each of these approaches has distinct advantages and disadvantages.

Researchers conducting a professional-judgment study organize from one to several teams of educators within a state and ask them to design an educational program that will achieve the state's educational goals. Once team members have identified the required set of inputs, researchers determine how much money will be needed to fund those inputs; typically, the panels themselves are not required to consider how much their suggested program will cost. This has led to concerns that these panels tend to overestimate resource needs. The evidence-based approach is similar except that the sources of expert opinion are not panels of professionals but research evidence on strategies that have been proven effective. Both professional judgment and evidence-based studies have the appeal of transparency; that is, not only do they provide a total cost of education in dollars, but they specify how those dollars are to be spent. However, these methods focus on the selection of inputs, the estimated costs may be only weakly connected to specific outcomes. This could be considered a strength, as it allows for the consideration of performance outcomes that may be difficult (or impossible) to measure, but it could also be

considered a weakness in a policy environment focused increasingly on specific accountability standards. Finally, although some professional judgment studies address the issue of extra resources that may be required for certain types of students, such as those from low-income families or with limited English proficiency (i.e., marginal costs), it may be problematic to estimate costs for actual districts that differ significantly from the typical school or district for which the original program was designed.

Sonstelie (2007) modifies the professional judgment methodology to address some of the concerns with the more traditional approach. Using a large sample of California teachers, principals, and superintendents (rather than a few selected panels), these professionals are presented with a school budget and characteristics for a hypothetical school, and then asked how they would allocate those resources. Once that allocation is completed, these same professionals are asked to predict the performance level of students in the re-designed school. The researcher can estimate relationships between the resource allocations, school characteristics, and outcomes because of the large sample. Thus, this approach can produce cost estimates that are tied to a particular outcome level, generate estimates of the marginal costs associated with specific student characteristics, and also provide insights into how dollars should be spent. One drawback is that the estimates are still built on predictions of individuals who may have unknown biases, rather than on observed correlations between dollars and outcomes.

Cost studies using the successful-schools approach start by identifying a set of high-performing schools (with performance generally based on the state's educational goals). Estimates of the cost of providing a high-quality education are then based on the lowest level of per-pupil spending among this set of successful schools. This approach has the advantage of directly linking costs and outcomes, but generally says little about how resources ought to be allocated within the school or district. A few successful schools studies have addressed the issue of additional costs for low-income students by identifying high-performing schools among schools with high proportions of students in poverty but, as with the professional-judgment method, it may be challenging to extrapolate to districts that are significantly different from the original set of successful schools.

The cost-function, or econometric, approach utilizes data on per-pupil school expenditures, student performance, and various characteristics of students and school districts, from all school districts within a state. Following Bradford *et al.* (1969), the output of public schools is assumed to be a function of school inputs (e.g., teachers and textbooks), characteristics of the student body, and the family and neighborhood environment in which the students live. This relationship can be represented by

$$S_{it} = g(X_{it}, Z_{it}, F_{it}) \quad [1]$$

where S_{it} represents an index of school output, X_{it} is a vector of direct school inputs, Z_{it} is a vector of student characteristics, and F_{it} is a vector of family and neighborhood characteristics. The subscript i refers to the school district and subscript t refers to the year.

The amount that a district will need to spend in order to produce a particular level of output is a function of school inputs and the prices of those inputs. This is shown in eqn [2], where per-pupil expenditures, E_{it} , are considered as a function of school inputs, a vector of input prices, P_{it} , and ε_{it} , a vector of unobserved characteristics of the school district:

$$E_{it} = f(X_{it}, P_{it}, \varepsilon_{it}) \quad [2]$$

Finally, by solving the production function in eqn [1] for X_{it} , and plugging X_{it} into eqn [2], we arrive at the cost function represented by eqn [3], where u_{it} is a random error term:

$$E_{it} = b(S_{it}, P_{it}, Z_{it}, F_{it}, \varepsilon_{it}, u_{it}) \quad [3]$$

Once a functional form is chosen for eqn [3], it can be estimated with district-level data for a given state, using regression techniques. The resulting coefficients indicate the contribution of various district characteristics to the cost of education, holding constant the level of performance. The cost function can then be used to predict the cost of any given level of performance; this is done by multiplying the cost function coefficients by the actual values of the student and district characteristics while setting the performance variables equal to the desired level. Thus, for each district, the cost function can predict the minimum amount of money necessary to achieve various educational performance goals, given the characteristics of the school district and its student body. As discussed in Reschovsky and Imazeki (1997), these predictions can be used to generate an index that summarizes, in one number, the cost differentials for each district in the state. Such an index is then easily incorporated into a state foundation aid formula (Ladd and Yinger, 1994). The marginal costs of specific student or district characteristics are also easily determined since the coefficients quantify the additional spending required for higher values of a specific cost factor, holding constant everything else.

Although generally considered more complex than other methods, the econometric approach has the advantage of directly quantifying the relationship between outcomes and costs for districts with a variety of characteristics. This approach is particularly appealing for states such as California, Texas, and New York, which are quite large and have tremendous diversity in both student and school district characteristics. However, as with the successful-schools method, cost functions are black boxes that do not illuminate how districts should organize their resources. Cost

functions are also limited by the same problems that can plague any statistical analysis, including errors in estimation and availability of high-quality data.

Cost-Function Estimation

As discussed in Imazeki and Reschovsky (2005), any empirical analysis requires that the researcher make a number of choices. Even when the basic methodology is the same, different researchers may make legitimately different choices that lead to different results. For example, before estimating a cost function, one must first specify a functional form (e.g., linear, log-linear, etc.). Such choices have implications for the final cost function results and any subsequent cost predictions that are derived from those results. Imazeki and Reschovsky (2005) present a detailed comparison of two cost functions estimated using data for Texas (Imazeki and Reschovsky, 2004; Gronberg *et al.*, 2004) to highlight some of the more common decisions that must be made in cost-function analyses. These include the choice of which variables to include, how to measure them, and the functional form of the cost function; how to address school district inefficiencies; and whether and how to account for the endogenous nature of district spending and student performance.

Variables and Functional Form

Although guided by an underlying model of public decision making and assumptions about cost-minimization, researchers still have considerable latitude in choosing which specific variables to include, how to measure them, and what functional form to assume. First, as noted by Baker *et al.* (2004), “A central difficulty of performance-oriented analysis involves the politics of achieving consensus regarding *important outcomes* (emphasis in original) and the empirics of precisely measuring those outcomes” (p. 20). The vast literature on educational production functions tends to focus on student cognitive achievement as measured by standardized test scores. Thus, a commonly used measure of school output or performance is test scores from achievement tests administered to all students. However, cost-function studies differ in whether they use a value-added measure of performance (i.e., including a lagged value of the performance measure), as in Reschovsky and Imazeki (1998, 2003) and Gronberg *et al.* (2004), or a current level measure, as in Duncombe and Yinger (2005). The primary drawback to the level measure is that it may bias the coefficients on student characteristics (see Meyer, 1996, for a discussion of value-added versus level student output measures); on the other hand, it may be more straightforward for policymakers who are interested in

the total cost of achieving some target level of student performance. Studies also vary in whether they include other measures of student outcomes such as dropout or graduation rates (Duncombe *et al.*, 1996) or the number of advanced courses offered (Reschovsky and Imazeki, 1998).

All cost-function studies include measures of district and student characteristics, typically district size, the percentage of students in poverty, limited English proficiency, or disabled, and teacher salaries. However, there are differences in how these variables are measured. For example, Reschovsky and Imazeki (1998, 2003) typically measure district size with a quadratic term (thus allowing for economies and diseconomies of scale) while Duncombe and Yinger (2005) have included indicator variables for various enrolment categories (e.g., below 1000, 1000–2000, 2000–5000, etc.). Duncombe and Yinger (2005) also compare different measures of poverty, using both the proportion of students in the free and reduced-price lunch program (the more common measure in other studies) and the child poverty rate from the census.

Most cost-function studies have assumed a log-linear functional form (e.g., Imazeki and Reschovsky, 2006; Duncombe and Yinger, 2005). With this function, the coefficients can be interpreted directly as the marginal effect of a one-unit change in the variable. In contrast, Gronberg *et al.* (2004) estimate a translog cost function for Texas. The translog is highly flexible, including interactions among all the variables and quadratic terms for all the variables; the trade-off is that the resulting coefficients are more difficult to interpret as the marginal effects of individual variables will depend on the values of all the variables in the model.

Efficiency

Although cost functions are built from a theoretical model that assumes schools allocate resources in order to minimize cost, in reality, some schools may operate inefficiently. These potential inefficiencies are not directly observed and, if not addressed, may bias the cost-function estimates. However, the measurement of efficiency in the context of elementary and secondary education is exceedingly complex; in order to accurately measure school district efficiency, it is necessary to accurately identify and quantify both the educational goals of each school district and the factors that contribute to the achievement of these goals and to school district expenditures.

One approach is to use statistical techniques such as stochastic frontiers or data envelopment analysis to attempt to identify spending that is high, relative to spending in districts with similar performance and costs; see, for example, Duncombe *et al.* (1996), McCarty and Yaisawarng (1993), and Deller and Rudnicki (1993). However, the measurement of school district efficiency using these statistical methods is highly sensitive to the way that

school district goals are measured. For example, in school districts that emphasize vocational education, or arts and music – subjects not directly measured by standardized tests – money spent on these alternative educational objectives will be counted as inefficient spending.

An alternative approach to addressing district inefficiencies is to include control variables that are likely to be correlated with efficiency; for example, Imazeki and Reschovsky (2004, 2006) use a measure of district competition while Duncombe *et al.* (2003) use district characteristics associated with tighter or looser budget constraints (such as property values and state aid). Although this does not require as strict a set of assumptions as the statistical approach, there is also a greater potential that inefficiency will be underestimated and resulting estimates will still contain bias.

Estimation Methods

When school districts strive to achieve higher levels of student performance, such goals presumably will require additional spending. At the same time, decisions by local school boards about per-pupil spending are likely to have an impact on student performance goals. That is, school districts often make decisions about spending levels and student outcome levels simultaneously. If not addressed, this simultaneity can lead to bias in the estimated cost-function coefficients. However, statistical tests can be used to assess whether this bias is, indeed, present. For example, Gronberg *et al.* (2004) and Imazeki (2007) rule out the need to adjust for simultaneity in their data.

When needed, the most common approach to the issue of simultaneity is estimation with two-stage least squares (i.e., instrumental variables). The challenge is to find instruments that are correlated with student outcomes but not correlated with school district spending. Imazeki and Reschovsky (2004) follow the literature on demand for education and use variables such as median income and tax price. Duncombe and Yinger (2005) rely on geographic proximity, using measures of fiscal capacity, student need, physical conditions, and student performance in adjacent districts.

Data Requirements

Although not under the control of the analyst, another issue for estimation of educational cost functions is that such estimation requires data on quantifiable outcome measures and student and district characteristics from a large number of districts, ideally over multiple years. The cost-function model also assumes districts allocate resources in order to maximize outcomes/minimize cost. When good data are not available, or districts are particularly constrained in their ability to allocate resources

efficiently, other methods for assessing the costs of adequacy are likely to be more appropriate. (See Imazeki, 2007, for a discussion of the biases that can arise in cost functions when districts face such constraints.)

Cost-Function Results

Although cost-function studies vary in how they handle these estimation issues, all of the studies are consistent in at least a few ways. Costs are consistently found to be higher for students in poverty, for English learners, for special education, and for very small districts, although there is variation in the magnitude of the cost differentials. For example, **Table 1** shows the base and marginal costs from four recent cost-function studies. Base costs refer to the costs for a low-need district to achieve the state standard; that is, a district with relatively low levels of poverty, few English Learners, etc. It is important to keep in mind that base costs may vary across time or across states because of differences in standards (e.g., if states raise their performance standards, the base cost will increase) or differences in regional price levels (e.g., southern states may have lower base costs than northeastern states); thus, comparisons across states must be viewed with much caution. But in a given year and state, the base cost represents the minimum bar for per-pupil spending within that state.

Marginal costs refer to the additional costs associated with specific student or district characteristics (such as poverty, English learners, and special education), above and beyond the base cost in a district with none of these special needs. For example, suppose the marginal cost of a student in poverty is determined to be 10%. If the base cost for a student with no special needs is US\$8000 per pupil, then the cost for a poor student is US\$8800, or 10% more. Many state-aid formulas try to account for marginal costs by assigning extra weight, and thus extra revenue, to students in certain categories. Following the same example, a poor student may be given an extra weight of 0.1 and this generates 10% revenue for that student, relative to the revenue allocated for a non-poor student. One advantage of cost functions is that these marginal costs are straightforward to calculate from the cost function coefficients. (See Imazeki, 2007, for a complete description of how the weights in **Table 1** are calculated, and for an expanded comparison of the estimates.)

The different choices made by the researchers are apparent in **Table 1**, since differences in the cost functions estimated in each state have large implications for the resulting pupil weights. For example, the 2004–05 Texas study and the New York study do not include separate measures of students with severe disabilities; it is thus not possible to calculate a separate weight for these students. (In the New York study, the special education

Table 1 Comparison of base and marginal costs^a

	<i>California Cost function, 2004–05</i>	<i>Texas^b Cost function, 2004–05</i>	<i>Texas^c Cost function, 2002–03</i>	<i>New York^{d,e} Cost function, 2000</i>
Base costs				
Number of studies	1	1	1	1
Mean	\$5163	\$3528	\$6174	\$4779
Minimum				
Maximum				
Marginal Costs: Pupil weights				
poverty	0.30	0.51	0.38	1.22
English Learners			0.24	1.01
Spanish speakers	0.08			
non-Spanish speakers	0.24			
Special education		0.71		2.05
All disabilities	1.13		0.72	
High-cost disabilities	6.68		1.03	

^aAll costs are in 1999 dollars and adjusted for regional variation in wage costs using the National Center for Education Statistics' Comparable Wage Index; base year = 1999.

^b2004–05 Texas cost index and pupil weights from author's calculations, based on cost function in Imazeki and Reschovsky (2006).

^c2002–03 Texas cost index from Gronberg, *et al.* (2004); pupil weights based on Gronberg *et al.* table 5.

^dNew York cost index from Duncombe, based on cost function in Duncombe, Lukemeyer and Yinger (2003).

^eNew York pupil weights from Duncombe and Yinger (2005, table 6).

variable does not include minor disabilities and thus captures something in between the two variables used in the California and 2002–03 Texas cost studies.) Overall, the marginal cost estimates from New York are larger than in the other states but all of the studies find that students in poverty and in special education require significantly more resources, and three of the four studies find that English learners also require additional funds. (Imazeki and Reschovsky, 2006 find a negative relationship between English learners and costs. They suggest this may be due to economies of scale, since Texas regulations require districts to provide a certain level of services if there is even one English learner in the district.) It is also worth noting that the cost differentials shown in **Table 1** are all larger than the existing state aid for student needs in each state; for example, in California, the average district receives only about 10% more aid for students in poverty while the California cost function suggests at least 30% more is required.

Although there are multiple ways to identify the costs of an adequate education, economists have tended to focus on statistical methods and the estimation of educational cost functions. Of course, no method is perfect, and even researchers using similar methods may make different choices that influence their results. However, the growing body of adequacy research consistently supports the need for additional resources for students from particular backgrounds and in particular educational settings. School finance systems designed to ensure that all students have equal access to an adequate education must recognize these differential costs.

See also: Compensating Differentials in Teacher Labor Markets; Economic Approaches to School Efficiency; Education Production Functions: Concepts; School Finance Reform; School Finance: An Overview.

Bibliography

- Baker, B. D., Taylor, L. L., and Vedlitz, A. (2004). Measuring educational adequacy in public schools. *A Report Prepared for the Texas Legislature Joint Committee on Public School Finance, The Texas School Finance Project.*
- Bradford, D. F., Malt, R. A., and Oates, W. E. (1969). The Rising Cost of Local Public Services: Some Evidence and Reflections. *National Tax Journal* **22**, 185–202.
- Deller, S. C. and Rudnicki, E. (1993). Production efficiency in elementary education: The case of Maine public schools. *Economics of Education Review* **12**, 45–57.
- Duncombe, W., Lukemeyer, A., and Yinger, J. (2003). Financing an adequate education: A case study of New York. In *Developments in School Finance*, pp 129–153. Washington, DC: National Center on Education Statistics, US Department of Education.
- Duncombe, W., Lukemeyer, A., and Yinger, J. (2004). *Education Finance Reform in New York: Calculating the Cost of a "Sound Basic Education" in New York City*. New York: Syracuse University, Center for Policy Research, Maxwell School of Citizenship and Public Affairs, Policy Brief No. 28.
- Duncombe, W., Ruggiero, J., and Yinger, J. (1996). Alternative approaches to measuring the cost of education. In Ladd, H. F. (ed.) *Holding School Accountable; Performance-Based Reform in Education*, pp 327–356. Washington, DC: Brookings Institution.
- Duncombe, W. and Yinger, J. (2005). How much more does a disadvantaged student cost? *Economics of Education Review* **24**, 513–532.
- Gronberg, T. J., Jansen, D. W., Taylor, L. L., and Booker, K. (2004). School outcomes and school costs: The cost function approach. *A Report Prepared for the Texas Legislature Joint Committee on Public School Finance, The Texas School Finance Project.*

- Imazeki, J. (2007). Assessing the costs of K-12 education in California public schools. *Getting Down to Facts Project*. Palo Alto, CA: Stanford University.
- Imazeki, J. and Reschovsky, A. (2004). Estimating the costs of meeting Texas education accountability standards. *Report Submitted to the Plaintiffs as Evidence in West Orange-Cove et al. v. Neeley et al.* District Court of Travis County, Texas, revised version, July 8.
- Imazeki, J. and Reschovsky, A. (2005). Assessing the use of econometric analysis in estimating the costs of meeting state education accountability standards: Lessons from Texas. *Peabody Journal of Education* **80**, 96–125.
- Imazeki, J. and Reschovsky, A. (2006). Does no child left behind place a fiscal burden on states? Evidence from Texas. *Education Finance and Policy* **1**, 217–246.
- Ladd, H. F. and Yinger, J. (1994). The case for equalizing aid. *National Tax Journal* **47**, 211–224.
- McCarty, T. A. and Yaisawarng, S. (1993). Technical efficiency in New Jersey school districts. In Fried, H. O., Lovell, C. A., and Schmidt, S. S. (eds.) *The Measurement of Productive Efficiency: Techniques and Applications*, pp 271–287. New York: Oxford University Press.
- Meyer, R. H. (1996). Value-added indicators of school performance. In Hanushek, E. A. and Jorgeson, D. W. (eds.) *Improving the Performance of America's Schools*, pp 197–223. Washington, DC: National Academy Press.
- Reschovsky, A. and Imazeki, J. (1997). The development of school finance formulas to guarantee the provision of adequate education to low-income students. In *Developments in School Finance*, 1997, pp 121–148. Washington, DC: National Center on Education Statistics, US Department of Education.
- Reschovsky, A. and Imazeki, J. (2003). Let no child be left behind: Determining the cost of improving student performance. *Public Finance Review* **31**, 263–290.
- Sonstelle, J. (2007). Aligning school finance with academic standards: A weighted-student formula based on a survey of practitioners. *Getting Down to Facts Project*. Palo Alto, CA: Stanford University.
- Tsang, M. C. (1988). Cost analysis for educational policymaking: A review of cost studies in education in developing countries. *Review of Educational Research* **58**, 181–230.
- Tsang, M. C. (1997). Cost analysis for improved educational policymaking and evaluation. *Educational Evaluation and Policy Analysis* **19**, 318–324.

Further Reading

- Downes, T. A. and Pogue, T. F. (1994). Adjusting school aid formulas for the higher cost of educating disadvantaged students. *National Tax Journal* **47**, 89–110.

Economic Approaches to School Efficiency

L L Taylor, Texas A&M University, College Station, TX, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Allocative efficiency – Allocative efficiency is an aspect of school efficiency. Schools are allocatively efficient when the relative productivity of educational inputs equals the relative price of those inputs. Thus, if administrators are paid twice as much as teachers, then allocative efficiency requires the school to choose a mix of teachers and administrators so that administrators are also twice as productive as teachers.

Data envelopment analysis (DEA) – It is a mathematical (as opposed to statistical) strategy for identifying a best-practice frontier. The best-practice frontier is a series of linear segments connecting the decision-making units – in the education context, the schools, school districts, or universities – that maximize the stated objectives given the stated constraints. For example, the objective could be to maximize a set of school outputs given the observed quantities of school inputs.

Scale efficiency – Scale efficiency is an aspect of school efficiency. Schools are scale efficient when there are no productivity gains from changing the size of the school.

Technical efficiency – Technical efficiency is an aspect of school efficiency. Schools are technically efficient when it is not possible to increase educational outputs without increasing inputs. Equivalently, schools are technically efficient when it is not possible to reduce all inputs proportionately without reducing school output.

efficient when they are using the right mix of inputs, given the relative prices and productivity of those inputs. Thus, allocative efficiency requires that when teachers are twice as expensive as teacher aides, they must also be twice as productive. The third aspect of school efficiency is scale efficiency. Schools are scale efficient when there are no productivity gains from changing the size of the school. Although most of the literature focuses on technical efficiency, researchers have found evidence that schools are inefficient in all three dimensions.

Econometric Approaches

Researchers have used a number of analytic techniques to examine school efficiency. The three most common strategies are data envelopment analysis (DEA), stochastic frontier analysis (SFA), and corrected ordinary least-squares analysis (C-OLS). All three approaches have been used to identify best practices among schools or school districts. Deviations from best practices are then attributed to inefficiency (DEA and cross-sectional C-OLS) or to a combination of technical inefficiency and random error (SFA and panel C-OLS). In all cases, inefficiency is measured relative to current best practice, which may fall short of the theoretical ideal.

Data Envelopment Analysis

DEA uses a mathematical optimization technique called linear programming to identify a best-practice frontier. The best-practice frontier is a series of linear segments connecting the schools, school districts, or universities that maximize output (or minimize costs) given a set of constraints. Typically, the best-practice frontier represents the highest level of school output achievable with a given set of school inputs, although other specifications have been used to measure school efficiency (e.g., Ruggiero and Vitaliano, 1999; Grosskopf *et al.*, 1999). Schools that are on the frontier are deemed technically efficient, while schools that are not on the frontier are inefficient. The closer the school is to the best-practice frontier, the more efficient it is. Although DEA can be used to measure allocative and scale efficiencies (e.g., Grosskopf *et al.*, 1999), it is primarily used to evaluate technical efficiency.

DEA has a number of attractive properties. The methodology can accommodate multiple inputs and outputs, allowing it to reasonably reflect the educational process.

Efficiency in an Education Context

Many researchers have found little or no empirical relationship between school resources and school outcomes. This lack of correlation implies that schools are economically inefficient, and has given rise to a substantial literature examining the nature and causes of such inefficiency.

This literature focuses on three important aspects of school efficiency. The first is technical efficiency. Schools are technically efficient when it is not possible to reduce all inputs proportionally without reducing output. The second is allocative efficiency. Schools are allocatively

In addition, DEA does not require the researcher to specify a functional form for the relationship between inputs and outputs or for the distribution of school efficiency. As such, the method is largely insulated from specification errors (although like all other approaches it can yield biased estimates of efficiency when relevant variables are omitted). Research also suggests that DEA yields more accurate efficiency estimates than C-OLS when an input is endogenous (Bifulco and Bretschneider, 2003; Ruggiero, 2003).

One downside to DEA arises from the method's sensitivity to extreme observations. DEA identifies all schools with a unique mix of inputs or outputs as frontier schools because there are no other schools to use for reference. Any increase in the number of inputs or outputs tends to increase the number of schools that are unique in some dimension, and therefore the number of schools on the DEA frontier. To limit the number of schools that are efficient by default, researchers must use a relatively parsimonious model of school behavior.

The need for parsimony has led many researchers to adopt a two-stage estimation strategy. The first stage uses DEA with a minimal number of inputs and outputs to generate efficiency scores. The second stage uses regression analysis to adjust those efficiency scores for student demographics and other factors that were omitted from the first stage. As discussed in Johnes (2006), this two-stage approach implies that the variables omitted from the first stage of the analysis influence school efficiency, but not the educational production process itself. As such, it is not always obvious which variables belong in which stage of the analysis. Excessive parsimony in the first stage of the analysis can lead to biased estimates of school efficiency, even after the second-stage adjustments. Furthermore, Simar and Wilson (2007) argue that correlations among the DEA efficiency scores invalidate many of the regression techniques commonly used in second-stage analyses.

More problematically, DEA is an approach that presumes all inputs and outputs are measured without error. Measurement error with respect to the schools that make up the best-practice frontier leads to biased estimates of school efficiency, even for schools that are measured without error. If a school is on the frontier because its output is overestimated (or its input is underestimated), then DEA will overstate the inefficiency of all schools that use the mismeasured school as a reference. Research suggests that the inaccuracy can be substantial, even when the measurement error is relatively modest (e.g., Bifulco and Bretschneider, 2003). Exploiting multiple observations for each decision-making unit – either by relying on multiple years of data or by aggregating school data to the school district level – can increase the accuracy of the efficiency estimates if the measurement error is random (Ruggiero, 2006), but not if the measurement error is systematic.

Stochastic Frontier Analysis

SFA is a regression technique that formally incorporates both technical efficiency and random error into the estimation of a best-practice frontier. Estimation error from the regression is presumed to represent two distinct parts. The first is the normal random error that is associated with any regression-based estimation. The second represents inefficiency and is therefore only found on one side of the estimated frontier. The estimated relationship between inputs and outcomes is based on best practice within the data because the composite error is asymmetric.

As with any regression-based analysis, researchers using SFA must designate the functional relationship between inputs and outputs. Many researchers choose highly flexible functional forms to avoid imposing undue structure on the relationship. However, in small samples, such flexibility can bias the estimates of technical efficiency when the underlying relationship between inputs and outputs is more parsimonious (Jensen, 2005).

Researchers also must specify a distribution for the inefficiency term. Most researchers presume that school inefficiency follows a half-normal distribution, but other distributions have also been used. Jensen (2005) finds that specifying a half-normal distribution for the inefficiency term generates more reliable estimates of technical efficiency than other assumptions about the distribution of inefficiency.

Corrected OLS

Corrected OLS is exactly what it sounds like it is. With this approach, researchers use OLS regression to estimate a production, cost, or distance function, and then make a correction to the intercept term to reflect school inefficiency. The predicted values from this adjusted regression represent the C-OLS frontier. As with DEA, all deviations from the C-OLS frontier are interpreted as inefficiency. However, unlike DEA or SFA, the relationship between inputs and outputs along the frontier is determined by the average practice of schools in the data rather than the best practice.

As with SFA, the researcher must designate the functional form of the educational production process. However, the researcher has no discretion over the distribution of school efficiency. The C-OLS methodology imposes a normal distribution on the technical efficiency estimates. While there is seldom any *a priori* reason to believe that efficiency is normally distributed, research suggests that in a great many cases, C-OLS produces more accurate efficiency rankings than does SFA (Jensen, 2005).

Regression analysis using multiple years of data and indicator variables (i.e., fixed effects) for each school or district (as in Schwartz and Zabel, 2005) can be thought of

as a special case of C-OLS. In this case, researchers must assume that school inefficiency is constant over time – a particularly strong assumption in a changing policy environment – but face no other restrictions on the distribution of school efficiency.

Specifications and Measurement

A variety of specifications have been used to estimate school or school district efficiency using DEA, SFA, or C-OLS. Some researchers estimate a production function, wherein a composite (or representative) school output is modeled as a function of school input quantities, environmental factors like student poverty, and technological factors like school size. Other researchers estimate cost functions, wherein school expenditures are a function of one or more school outputs, input prices like teacher wages, environmental factors, and technological factors. Still other researchers use distance functions like the input distance function, which is dual to (i.e., a mathematical translation of) the cost function but requires data on input quantities rather than input prices.

Regardless of the specification or estimation strategy, researchers must struggle with three important measurement issues. First, researchers need reliable measures of school output. Second, researchers need reliable measures of school inputs. Finally, given the significant role that home and family play in student learning, researchers need a credible strategy for incorporating differences in the educational environment.

School Outputs

Most researchers rely on standardized tests to measure school outputs. Test scores are readily available, tractable, and consistently defined measures of student performance in subject areas that have been associated with higher lifetime earnings (e.g., Blau and Kahn, 2005). Many states incorporate test scores into their accountability systems, making test performance a direct objective of school administrators. Research also suggests that differences in test scores are capitalized into housing values, implying that the home-buying public also interprets test scores as an indicator of school quality (Taylor, 2005).

No one believes that test scores measure all of the important differences in school outcomes, however. Many of the things schools do, such as arts, music, and other enrichment activities, are not directly measured by the minimum competency tests available to researchers. Taylor (2005) finds that test scores and student demographics can explain less than one-third of the premiums buyers pay for homes in specific school-attendance zones, suggesting that the school outcomes not captured by test scores are far from negligible.

Test scores are also a questionable indicator of school outcomes because they do not reflect exclusively the performance of schools. Education researchers have long recognized that students from advantaged backgrounds can score well on standardized tests despite ineffective schools, while students from disadvantaged backgrounds can score poorly even when the schools are highly effective. As discussed in Hanushek and Taylor (1990), unadjusted test scores are biased estimates of the marginal effects of schools.

Unfortunately, reliable estimates of school effects can be hard to come by. A student's performance in the current year is a function not only of current school and family inputs, but also of the student's complete history of school and family inputs, and therefore the data required to estimate school effects can be substantial. Most researchers did not have access to the necessary data on individual student performance, demographics, and history.

Measurement error is clearly a concern, because school effects must be estimated. Kane and Staiger (2001) find that much of the variation in estimated school effects – either performance-level effects or performance-gain effects – may be attributable to nonpersistent factors and noise. Furthermore, the measurement error increases as the size of the group under analysis falls; so, the measurement error is most pronounced for small schools and student subgroups within schools. However, Kane and Staiger (2001) also demonstrate that averaging student performance over time can greatly reduce the measurement error in estimated school effects.

At the university level, there are few measures of student performance from which to estimate marginal effects. Therefore, researchers tend to measure teaching output using quantity measures such as the number of degrees awarded or the number of full-time-equivalent students. Most researchers also include measures of research activity as a second type of school output. The ready availability of research ratings for British universities may explain why analyses of the UK dominate the literature on higher education efficiency.

School Inputs

While researchers generally use some measure of test scores as an output, there is much less consensus in the literature regarding the appropriate measures of school inputs (Worthington, 2001). Some researchers use real input measures such as the number of full-time-equivalent teachers, the number of administrative staff per pupil, the number of books in the school library, or the pupil-teacher ratio. Others use expenditure measures, such as instructional expenditure, teaching expenditure per pupil, or current operating expenditure.

Researchers who use real measures of personnel inputs seldom deal explicitly with the heterogeneity of the teacher corps. A teacher is a teacher whether she is a rookie with a newly minted bachelor's degree or a classroom veteran with 20 years experience and two advanced degrees under her belt. Arguably, observable characteristics like experience and educational attainment have little power to differentiate effective teachers from ineffective ones, so greater precision may not be possible. However, even unavoidable imprecision introduces measurement error and has a detrimental effect on estimates of school efficiency.

Researchers who use expenditure as a proxy for real inputs implicitly assume that the prices of such inputs are constant across the schools under analysis. While such an assumption is plausible for school supplies, it is much less plausible for labor inputs. Wages vary dramatically from one labor market to another, even within a single state (Taylor and Fowler, 2006). Any variation in prices introduces noise into expenditure-based measures of school inputs.

Variation in input prices is a key element of cost function analysis, and reliable measures of input prices are essential to any analysis of allocative inefficiency. It is particularly important to measure educator wages accurately because education is such a labor-intensive industry. Average wages are a poor proxy for the price of educator labor because differences in average wages reflect differences in staff composition from one school to the next. Instead, theory favors a wage index that indicates the price schools must pay to hire an educator with specified characteristics (such as a teacher with 5 years experience and a bachelor's degree). Average wages for beginning teachers are a common proxy for such an index, although this approach can be problematic if some schools do not have beginning teachers, or if variations in beginning teacher wages reflect endogenous variations in teacher quality.

Environmental Factors

The influence of student and family characteristics on educational outcomes is well known. Therefore, virtually all researchers incorporate some measure of the educational environment into their analyses. The most common measures of environmental conditions are indicators of parental education and socioeconomic status (Worthington, 2001).

These environmental factors can be thought of as indirect measures of the student and family inputs to the educational production process, or as shift factors that capture differences in the education technology. When outputs are measured with test scores rather than school effects, these factors (at least in a stochastic setting) also serve to partially control for measurement error in the outputs.

Whatever role the common environmental indicators play in the analysis, it is unlikely that they are sufficient to perform the task. A single summary statistic like family income cannot possibly reflect the array of abilities and motivations that students bring to the classroom. The richness of the literature on peer effects makes the treatment of this issue in the school efficiency literature seem rather paltry. Student-selection issues are largely ignored, which is particularly problematic for analyses that compare efficiencies across magnet, charter, private, and post-secondary schools. A careful analysis of the appropriate role for student characteristics in school efficiency measurement is long overdue.

Findings Regarding School Efficiency

Much of the empirical literature on school efficiency has focused on estimating and comparing technical efficiency scores. These studies find that there is significant inefficiency in the education process in the United States, the United Kingdom, and most other countries for which such studies have been conducted. Researchers typically find average technical inefficiencies in the education sector that are between 5% and 20%. As such, the evidence suggests that the average technical inefficiency of the education sector is similar in magnitude to the average technical inefficiency outside of education (Ruggiero and Vitaliano, 1999; Johnes, 2006).

Occasionally, researchers have estimated technical inefficiency scores using multiple econometric techniques. Generally, these researchers find that efficiency scores are sensitive to differences in specification and methodology, and that there is more correlation among efficiency rankings than efficiency scores.

Researchers have also found evidence of significant allocative inefficiency. The DEA has been the most common tool for analyses of allocative inefficiency because it can yield direct estimates of the optimal resource allocation (e.g., Grosskopf *et al.*, 1999). However, stochastic approaches can also support estimates of allocative inefficiency (e.g., Grosskopf *et al.*, 2001). Intriguingly, despite popular rhetoric about excessive administration in education, schools appear only slightly more likely to overuse administrators and administrative staff (given their relative wages) than to overuse teachers and other instructional staff (e.g., Grosskopf *et al.*, 1999, 2001).

The literature on scale efficiency in education has developed separately from the literature on technical and allocative efficiency. Most researchers interested in economies of scale have estimated cost and production functions using variants of OLS estimation rather than frontier methodologies. However, their decision not to recover the C-OLS efficiency scores from such analyses

does not affect the applicability of their findings on scale efficiency. As discussed by Andrews *et al.* (2002) researchers have generally found that there are economies of scale in elementary and secondary education, but that schools with more than 600–1000 students (depending on grade level) and school districts with more than 6000 students are inefficiently large. Analyses of higher education also tend to indicate economies of scale.

The Determinants of School Inefficiency

There are many possible explanations for the demonstrated inefficiency of the education sector. One of the most frequently explored is the perceived lack of competition at the elementary and secondary school levels. A number of researchers have found evidence that inefficiency is greater in less-competitive markets. For example, Grosskopf *et al.* (2001) find that allocative inefficiency rises with market concentration, and that schools in concentrated education markets are at least 16% more allocatively inefficient than schools in competitive markets.

Another possible reason for school inefficiency arises from the principal-agent problems common to public institutions. School administrators may have some objectives – such as increasing their bureaucratic power or maximizing their budgets – that are inconsistent with efficient operation of schools. Administrators are more likely to pursue their own objectives when it is difficult for the public to monitor them effectively. Researchers have found that common proxies for monitoring activity (such as home-ownership rates, tax rates, and the educational attainment of the adult population) are significantly correlated with school efficiency, and that schools are less efficient where monitoring is less likely (e.g., Grosskopf *et al.*, 2001).

Regulation may also partially explain school inefficiency. Many of the parameters of the educational production process are set by state agencies or statutes. State policies determine maximum class sizes and minimum teacher salaries; require schools to hire full-time nurses and counselors; specify the length of the school year; and dictate the minimum square footage per pupil in each classroom. The restrictions imposed on schools, districts, and universities may make it impossible for them to behave efficiently. Comparative analyses of charter schools (which are a largely unregulated form of traditional public schools), private schools, and private universities may be the key to more fully understanding the impact of regulation on public school efficiency.

Union activity is another possible explanation. Union contracts typically limit school flexibility with respect to the allocation and compensation of labor. Work rules and seniority protections can make it very difficult to fire ineffective teachers, or close dysfunctional schools.

To the extent that unions lead schools to diverge from best practice, they contribute to measured inefficiency.

A lack of population density can explain some types of inefficiency. It is not possible to operate a scale-efficient school when there are only a handful of sixth graders within a 100-mile radius. In turn, inefficiently small schools find it very difficult to avoid allocative inefficiencies, particularly with respect to labor.

Finally, measured inefficiency may reflect nothing more than the general inadequacies of available data. All of the efficiency estimation strategies are vulnerable to omitted variables bias and measurement error. In the education sector, omitted variables are common, and measurement errors are pervasive. As a result, apparently efficient schools may simply be schools that rely heavily on an unmeasured input or that operate in a favorable environment. Similarly, some apparently inefficient schools may actually be best practitioners that are either focusing their attention and resources on an unmeasured output, or operating in a very unfavorable educational environment.

Conclusions

An extensive literature has developed to explore the efficiency of the education sector. Much of that literature has focused on applying existing tools for efficiency measurement to schools, school districts, and universities. An array of researchers report finding significant technical, allocative, or scale inefficiencies in the education sector.

The literature on the determinants of inefficiency is less well developed, but potentially more influential. Researchers have identified a number of institutional structures that appear to foster inefficiency. Changes to those structures could lead to significant improvements in the efficiency of the education sector. Given the tight fiscal situation facing most governments, improvements in efficiency may be a necessary precondition for any improvements in public support for education. Therefore, research into the determinants of school inefficiency may be particularly useful going forward.

The primary challenge facing researchers, however, rests with the measurement of school characteristics, not with the analysis of school efficiency *per se*. No measure of school efficiency can be more reliable than the underlying data, and the data underlying many efficiency analyses are questionable. Fortunately, the school accountability movement is generating a flood of data that will allow researchers to improve their measures of school characteristics. In addition, wider adoption of analytic strategies designed to reduce the influence of measurement error, such as panel-data analyses or the estimation of confidence intervals for efficiency scores, will greatly enhance the credibility of any future analyses of school efficiency.

By refining their measures of school inputs, outputs, and environmental characteristics, researchers should be able to greatly improve our understanding of the nature and determinants of school efficiency.

See also: Cost–Benefit Analysis and Cost–Effectiveness Analysis; Economic Approaches to Adequacy; Education Production Functions: Concepts; Education Production Functions: Developed Country Evidence; School Finance: An Overview.

Bibliography

- Andrews, M., Duncombe, W., and Yinger, J. (2002). Revisiting economies of size in American education: Are we any closer to a consensus? *Economics of Education Review* **21**(3), 245–262.
- Bifulco, R. and Bretschneider, S. (2003). Response to comment on estimating school efficiency. *Economics of Education Review* **22**(6), 635–638.
- Blau, F. D. and Kahn, L. M. (2005). Do cognitive test scores explain higher U.S. wage inequality? *Review of Economics and Statistics* **87**(1), 184–193.
- Grosskopf, S., Hayes, K. J., Taylor, L. L., and Weber, W. L. (1999). Anticipating the consequences of school reform: A new use of DEA. *Management Science* **45**(4), 608–620.
- Grosskopf, S., Hayes, K. J., Taylor, L. L., and Weber, W. L. (2001). On the determinants of school district efficiency: Competition and monitoring. *Journal of Urban Economics* **49**(3), 453–478.
- Hanushek, E. A. and Taylor, L. L. (1990). Alternative assessments of the performance of schools: Measurement of state variations in achievement. *Journal of Human Resources* **25**(2), 179–201.
- Jensen, U. (2005). Misspecification preferred: The sensitivity of inefficiency rankings. *Journal of Productivity Analysis* **23**(2), 223–244.
- Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review* **25**(3), 273–288.
- Kane, T. J. and Staiger, D. O. (2001). Improving school accountability measures. *NBER Working Papers No. 8156*. National Bureau of Economic Research.
- Ruggiero, J. (2003). Comment on estimating school efficiency. *Economics of Education Review* **22**(6), 631–634.
- Ruggiero, J. (2006). Measurement error, education production and data envelopment analysis. *Economics of Education Review* **25**(3), 327–333.
- Ruggiero, J. and Vitaliano, D. F. (1999). Assessing the efficiency of public schools using data envelopment analysis and frontier regression. *Contemporary Economic Policy* **17**(3), 321–331.
- Schwartz, A. E. and Zabel, J. E. (2005). The good, the bad and the ugly: Measuring school efficiency using school production functions. In Stiefel, L., Schwartz, A. E., Rubenstein, R., and Zabel, J. E. (eds.) *Measuring School Performance and Efficiency: Implications for Practice and Research*, pp 37–66. Larchmont, NY: Eye on Education.
- Simar, L. and Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics* **136**(1), 31–64.
- Taylor, L. L. (2005). Revealed-preference measures of school quality. In Stiefel, L., Schwartz, A. E., Rubenstein, R., and Zabel, J. E. (eds.) *Measuring School Performance and Efficiency: Implications for Practice and Research*, pp 163–183. Larchmont, NY: Eye on Education.
- Taylor, L. L. and Fowler, W. J., Jr. (2006). A comparable wage approach to geographic cost adjustment. *Report No. 2006-321*. National Center for Education Statistics Research and Development.
- Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. *Education Economics* **9**(3), 245–268.

Further Reading

- Hanushek, E. A., Rivkin, S. G., and Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *Review of Economics and Statistics* **78**(4), 611–627.
- Johnes, J. (2004). Efficiency measurement. In Johnes, G. and Johnes, J. (eds.) *International Handbook on the Economics of Education*, pp 613–742. Northampton, MA: Elgar.
- Ruggiero, J. (2007). Measuring the cost of meeting minimum educational standards: An application of data envelopment analysis. *Education Economics* **15**(1), 1–13.
- Stiefel, L., Schwartz, A. E., Rubenstein, R., and Zabel, J. E. (eds.) (2005). *Measuring School Performance and Efficiency: Implications for Practice and Research*. Larchmont, NY: Eye on Education.
- Todd, P. E. and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal* **113**(485), F3–F33.

School Finance Reform

T Downes, Tufts University, Medford, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Capitalization – Change in the price of a house (or any other asset) as a result of differences between communities with respect to taxes or the provision of public services.

Tax price – The cost to local taxpayers in additional property taxes, of an additional dollar of spending.

Introduction

During the past several decades, federal and state governments have pursued a variety of redistributive policies aimed at fostering equality of economic opportunity – the idea being that although people's incomes may vary, this variance should be primarily due to factors such as individual ability and effort, not due to differences in circumstance. Many in the policy arena have suggested that opportunities could be further equalized via the implementation of changes in the way elementary and secondary education is financed. The article provides a brief review of the multifarious effects of large-scale modifications, typically stimulated by court challenges to financing systems, to the formulas states use to determine aid to local school districts.

For many years, those concerned with the persistence of income inequality in the US have argued for reforms in the method of financing public elementary and secondary schools that would make education spending more equal. These arguments, which have been buttressed by substantial evidence that pre-market factors play a significant role in determining subsequent labor-market outcomes, have been cited by those who have argued in the courts for fundamental reforms of the way in which public schools are financed. These court challenges have experienced a resurgence in the last decade, with state supreme court decisions mandating equalization in states such as Arkansas, Kansas, New York, Texas, Montana, and New Hampshire further altering a school finance landscape that has changed dramatically since 1970 (Yinger, 2004). The end result of these court challenges is that, in almost every state in the nation, the system of financing the public schools has been fundamentally altered, with state governments becoming an ever-more important part of the educational financing landscape.

Because court decisions have been the antecedent of many of the school finance reforms that have occurred over the last four decades, one common strategy in empirical studies has been to classify reforms on the basis of whether they were a direct response to a court decision. This method of classifying reforms represents but one of many attempts to develop a simple rubric to reduce to one or two dimensions what are very heterogeneous policy changes. And while this classification method is as imperfect as any other, the focus on court-ordered reforms does highlight one aspect of the policy environment that may explain why some reforms seem to have large effects, while others do not. The combination of the possibility of free-riding on the policy innovations of others and the down-side risk associated with failed policy innovations will tend to result in less policy innovation than is socially optimal. In the case of school finance, this may mean that, unless they are pushed by outside forces, policymakers are likely to make modifications to aid formulas that are less ambitious than may be necessary. Court decisions, threats of a court decision, or a politician's dare (as in the case of Michigan), can provide politicians with the cover to make reforms that may be risky but necessary. Thus, finance reforms that result from intervention outside the traditional political arena would be expected to have larger effects than reforms that do not result from such intervention. And such reforms can reasonably be thought of as exogenous, since the placement of outside interventions in place and time is effectively random. Much of the empirical literature on the impact of finance reforms has taken advantage of this fact to quantify the effects of reforms.

The remainder of this article documents the results of this empirical literature, starting with a review of the evidence on the impact of finance reforms on the level and distribution of spending. Next the effects of finance reforms on student performance are discussed. The article concludes by providing a brief overview of the growing literature documenting some of the other effects of finance reforms.

Effects of School Finance Reforms on Spending

Impact of Reforms on the Distribution and Level of Spending

Starting with the decisions of the California Supreme Court in the Serrano versus Priest case, the earliest

court rulings on the constitutionality of state school finance systems focused on the cross-district equity of education spending. Even more recent decisions that have focused on whether the finance system makes it possible for all districts to provide an adequate education have been described by researchers as effectively being equity decisions. Thus, quantifying the equity implications is a natural starting point for evaluating any finance reform.

Uniformly, analyses of court-ordered finance reforms find that these reforms result in more equal spending. The literature has not, however, reached consensus on whether this greater equality is attributable to raising up the spending of districts at the bottom of the spending distribution or of leveling down the spending of districts at the top of the spending distribution. Research that has quantified the effect of finance reforms on mean per-pupil spending in a state has established that there is significant cross-state variation in the impact of court-ordered finance reforms on mean spending. That cross-state variation depends on characteristics of the state's population that affect the political landscape of the state.

National-level studies that have used district-level (as opposed to state-level) data have tended to confirm the conclusions drawn from state-specific analyses. Reforms that followed court mandates have equalized per-pupil spending, though reforms that followed court decisions based on adequacy grounds have resulted in less equalization (Corcoran and Evans, 2007). The district-level data used in these studies have made possible relatively nuanced examinations of the leveling-up/leveling-down question. Analyses that treat court-mandated reforms as discrete, exogenous events have generally concluded that equalization happened through leveling-up, with spending increasing everywhere in the spending distribution (Murray *et al.*, 1998). On the other hand, analyses that attempt to account for the political nature of spending decisions tend to echo the conclusion of cross-state variability reached in some of the studies that used state-level data. Research has begun to establish that one critical determinant of the impact of a reform on spending in a district is the impact of that reform on the district's tax price (Hoxby, 2001).

How Are Additional Dollars Spent?

Recent research also has begun to move beyond the more traditional focus on the level and distribution of spending to analyze how districts use any additional dollars that they get. Districts that receive relatively more dollars after a finance reform use a substantial fraction of those additional dollars to increase current expenditure. The increased spending in these districts has been allocated primarily to instructional expenditure and expenditure on student support services.

Impact of School Finance Reform on Student Performance

The earliest work on the impact of school finance reforms on student performance focused on the extensive school finance reforms in California in the late 1970s (Downes, 1992). Those reforms generated greater equality across school districts in per-pupil spending but not greater equality in measured student performance. The problem with using the California case as a benchmark is that it has proven to be the exception, not the rule. First, the limits imposed on local control over spending have not been duplicated in any other state. Even in Michigan and Vermont, the states in which the most extensive post-Serrano reforms have been implemented, some degree of local control over taxes and spending is permitted. Further, the population of students served by California schools changed more dramatically than the population of students in any other state in the nation. These demographic changes make it difficult to quantify the impact of the finance reforms in California on the cross-district inequality in student achievement.

In response to the realization that lessons from California may not be generalizable and in an effort to examine the impact of finance reforms on mean student performance in reform states, a number of researchers attempted to use national-level data to determine how the level and distribution of student performance in a state was affected by a finance reform. These studies have varied in how they measured student performance, how they classified the nature of finance reforms, and how they characterized the distribution of student performance. Further, because the national-level nature of these analyses made cross-state comparisons possible, these studies tended to focus much of their attention on the impact of reforms on mean performance. These studies have not, however, produced a consistent vision of the impact of finance reforms on student performance. Several national-level studies indicate that relative performance may have declined and the extent of variation in performance may have increased in states in which finance reforms have been implemented. Other research using national-level data offer a more sanguine picture of the effects of such finance reforms. Test scores of those students who were most likely to benefit from finance reforms show evidence of relative improvement, and dropout rates appear to have declined in those school districts that benefited most from the finance reforms.

The fundamental reason for the absence of a consistent picture of the impact of school finance reforms is the tremendous diversity of these reforms. Even though there is general consensus that the key elements of a finance reform are the effects of the reform on local discretion, the effects of the reform on local incentives, and the

change in state-level responsibilities in the aftermath of reform, in national-level studies, different authors have taken different approaches to account for the heterogeneity of the reforms. The result is variation in conclusions reached in studies that are asking the same fundamental question.

For that reason, a number of authors have chosen to analyze carefully the impact of reforms in individual states. Analyses of individual states also allow for examination of the effects of reforms in those districts likely to benefit most from the changes in the financing system. While the contexts and the methods of analysis vary, most of these studies document relatively small reductions in variation in student performance in the aftermath of finance reforms.

Strikingly, the results of the state-specific analyses of finance reforms do not appear to depend on the method used for quantifying the impact of the reforms. A number of authors use an education production function-style analysis that relates spending to performance, other authors use an event analysis that compares performance before and after the reform for different types of districts, and still other authors use both methods. The estimated effects of the reforms appear to depend far more on the context than the method, a conclusion that is particularly strongly supported by the results of those studies that use both methods to evaluate the same reform.

Other Effects of School Finance Reforms

While much of the focus of the empirical work on school finance reforms has been on the impact of these reforms on education spending and on student performance, a significant body of research has developed that documents the variety of responses of states, localities, and private citizens to school finance reforms and that show that these responses vary in type and magnitude. Among the effects that have been examined are those on spending on other public services, community composition, housing prices, private school enrolment, and private contributions to public schools. The remainder of this article briefly summarizes the current state of knowledge on the nature of each of these effects.

Spending on Other Public Services after Finance Reforms

Analysis of the impact of school finance reforms on local spending shows that, while state aid for K-12 education increases after a court-mandated school finance reform, state aid for other public services declines. Spending on K-12 education appears to be different from every other spending category, with state resources being shifted away from other spending toward education in the aftermath of

a finance reform. On average, however, the composition of spending at the local level appears to be unaffected. In particular, in counties located in states in which there had been a court-mandated school finance reform, per capita expenditures on education, on police and criminal justice, on public welfare, health, and hospitals, and on other services (such as parks and recreation and libraries) appear to be unaffected. Highway expenditures might decline, but that decline appears to be small quantitatively.

The effects of reforms are not, however, uniform. There is evidence of heterogeneity in the extent to which spending on other services may change in the aftermath of finance reforms. This cross-locality variation may result because counties in which median family income is higher tend to benefit less from the increases in state education aid and tend to experience larger reductions in state aid for other services.

These results confirm a conclusion hinted at by other research: increases in spending on services that might substitute for education spending would not be expected in the aftermath of court-mandated school finance reforms. No substitution would be expected both because finance reforms impose, at most, minimal constraints and because state governments limit the impact of reforms on high-income communities by reducing state-level expenditure in other areas of spending.

Finance Reforms and Community Composition

Empirical research on the impact of finance reforms on community composition built off of a body of theoretical work that extended the classic Tiebout (1956) model to develop predictions concerning the impact of school finance reforms on community composition, housing prices, and private school enrolment. These theoretical papers indicate that restrictions on the ability of individuals to consume their desired level of education services in the public sector will tend to break down the tendency of individuals to sort on the basis of income or parental education. Further, many of the high-income individuals who move from previously high-spending to previously low-spending school districts will, at the same time, be choosing to opt out of the public schools.

This tendency of finance reforms to induce significant changes in community composition depends, however, on the extent to which the ability of schools to produce outcomes depends on the quality of peers in the schools. While the existence of peer effects need not accentuate the tendency to sort, the degree of sorting will depend critically on the benefits that high-income or high-ability individuals get from mixing with lower-income or lower-ability individuals. If the benefits of mixing are perceived to be low or if the benefits of having high-ability peers are sufficiently strong, individuals will either opt out of the

public sector or, by extension, will only choose the public sector if they are able to reside in homogeneous communities. Pursuing the logical implication of this reasoning, if parents feel high-ability peers are sufficiently important, school finance reforms could accentuate the tendency of individuals to sort both across communities and across schools.

Empirical research on the impacts of finance reforms on the degree of homogeneity of affected districts has shown that these reforms have relatively modest effects on community composition. The earliest work indicated that only in states where the constitutionality of a finance system was upheld, was there any consistent evidence of changes in community composition (Aaronson, 1999). More recent work, that has focused more directly on the composition effects of finance reforms in those districts, where reforms have had the largest impact and have considered measures of community composition other than income, has documented evidence of substantive changes in the extent of homogeneity of school districts. The evidence on the nature of these composition effects is equivocal, with national-level analysis indicating that court-mandated finance reforms may have led to a higher number of minority, lower-income, less-educated, and non-English-speaking students in previously low-spending districts, while state-level analysis indicates that the extent of sorting by education and income may have been reduced. These studies do, however, consistently show that these composition effects appear to be concentrated in those public school districts that were low-spending prior to the reforms. They produce no consistent evidence of composition changes in school districts that had been high-spending.

Housing Prices and Finance Reforms

The capitalization of the spending changes associated with court-mandated finance reforms into housing prices and rents may offer an explanation for the absence of consistent evidence of composition effects in the aftermath of finance reforms. Changes in the gap in housing costs between high-spending and low-spending school districts will influence the relative attractiveness of districts that were, before the finance reform, low-spending or high-spending. Theoretical models have established that changes in equalizing intergovernmental aid will be capitalized into property values. These capitalization effects can have dramatic effects on the impact of the changes in the aid program on consumer well-being. The potential impact of capitalization on community composition would be one such effect.

Research using national-level data on property values drawn from the decennial census has shown that the effects of court-mandated finance reforms are capitalized into property values (Dee, 2000). State-level analyses

have tended to confirm these findings. Further, the state-level analyses indicate that the capitalization effect is isolated in high-wealth communities. This loss in value in high-wealth communities could ultimately affect the sustainability of the finance reform, both because it affects revenue available to finance the schools and popular support for the system of finance.

Do Finance Reforms Lead to Shifts of Enrollment to Private Schools?

Finance reforms may not only result in movement from one school district to another; they may also result in movement to and from private education. While early work on the impact of finance reforms on private school enrolment suggested that families making schooling choices did not appear to have been sensitive to reforms in public school finance, more recent research has indicated that finance reforms can result in students shifting from public to private schools in significant numbers. For example, about 45% of the increase from 1970 to 1980 in the private school share in unified (K-12) districts in California could be attributable to the combined effect of Proposition 13, a tax limitation measure, and Serrano (Downes and Schoeman, 1998). While the long run shift to private schools in California is less than the shift that occurred in the immediate aftermath of the constraints, there was a permanent increase in the private school share.

A growing body of work has suggested that the growth in private school enrolment that occurred in California was not unique to that state. Constraints on spending imposed on some districts after a finance reform can lead some students to leave the public sector for the private sector. And those who depart may potentially be the highest-ability students.

Finance Reforms and Private Contributions to Public Schools

The other private response that has been studied extensively is the growth in private contributions to public schools that follows the imposition of constraints. This research has established that, while private contributions are a substantial portion of district revenues in a small number of school districts, these contributions typically have little impact on the post-finance reform distribution of education expenditures. For example, in California, the state in which private contributions have been most studied, 81.75% of students attended schools for which the combined contributions to the school and the district were less than US\$50 per pupil. Further, only 1.2% of students attended schools for which combined contributions were at or above US\$500 per pupil. Overall, contributions per pupil in California were only US\$39 in 2001 (Brunner and Imazeki, 2005).

Only in Vermont is there any evidence that private contributions were of sufficient magnitude to have muted the effect of finance reforms. In the state of Vermont, the court-mandated finance reform enacted in 1997, provided particularly strong incentives in some districts for private contributions to be a major source of local revenue. As a result, for a small subset of districts, contributions went from being inconsequential to comprising almost a fifth of revenues. Nevertheless, while contributions may have muted the impact of constraints slightly, the overall effect of contributions on the distribution of spending in Vermont was limited. Further, once sharp incentives to contribute were removed by changes in the finance system that took effect in 2004, private contributions went back to being inconsequential in all districts in Vermont.

Summary

Over the past four decades, school finance reforms, stimulated primarily by state supreme court decisions invalidating existing financing systems, have dramatically changed the educational landscape. These finance reforms have succeeded in reducing within-state variation in spending. And, while a fraction of the additional dollars that have flowed to previously low-spending districts have been used to provide property tax relief, the majority of these new dollars have been used to increase instructional expenditures. The effect these reforms have had on mean per-pupil spending has been less uniform, with the impact of the reforms in any state depending both on the context in which they have been implemented and the changes in incentives that have resulted from the reforms.

Analyses of the impact of finance reforms on student achievement indicate that the reforms tend to result in small, but significant, reductions in within-state variation in student achievement. Less clear is the effect of a finance reform on the mean level of achievement in a state. Undoubtedly, this lack of clarity is attributable to the heterogeneity of reforms and the dependence of the impact of a reform on the individual state context.

The changes in incentives that are a by-product of finance reforms have resulted in a number of effects of reforms both inside and outside the education sector. In general, the reforms have not resulted in significant substitution of other public expenditure for education expenditure or growth in private giving to public schools that is sufficient to undo the equalization that is the intent of the new financing system. The share of students attending private schools has, however, grown in the aftermath of court-ordered finance reforms, and property values have declined in those districts which had high levels of spending and property wealth prior to these reforms.

Finally, while the number of papers examining these other effects of finance reforms has grown dramatically in

recent years, much still remains to be learned about these other effects. Improved data and the abundance of recent finance reforms will make it possible for researchers to continue to deepen our knowledge of the full range of effects of finance reforms.

See also: Economic Approaches to Adequacy; Neighborhoods and Peers in the Production of Schooling; School Finance: An Overview; School Quality and Earnings; The Economics of Catholic Schools; Tiebout Sorting and Competition.

Bibliography

- Aaronson, D. (1999). The effect of school finance reform on population heterogeneity. *National Tax Journal* **52**, 5–30.
- Brunner, E. J. and Imazeki, J. (2005). Fiscal stress and voluntary contributions to public schools. In William, C. F. (ed.) *Developments in School Finance: 2004*, pp 39–54. Washington, DC: National Center for Education Statistics.
- Corcoran, S. and Evans, W. (2007). Equity, adequacy, and the evolving state role in education finance. In Helen, F. L. and Edward, B. F. (eds.) *Handbook of Research in Education Finance and Policy*, pp 332–356. New York: Routledge.
- Dee, T. S. (2000). The capitalization of education finance reforms. *Journal of Law and Economics* **43**, 185–214.
- Downes, T. A. (1992). Evaluating the impact of school finance reform on the provision of public education: The California case. *National Tax Journal* **45**, 405–419.
- Downes, T. A. and Schoeman, D. (1998). School financing reform and private school enrollment: Evidence from California. *Journal of Urban Economics* **43**, 418–443.
- Hoxby, C. M. (2001). All school finance equalizations are not created equal: Marginal tax rates matter. *Quarterly Journal of Economics* **116**, 1189–1231.
- Murray, S. E., Evans, W. N., and Schwab, R. M. (1998). Education finance reform and the distribution of education resources. *American Economic Review* **88**(4), 789–812.
- Tiebout, C. M. (1956). A pure theory of local public expenditures. *Journal of Political Economy* **64**, 416–424.
- Yinger, J. (ed.) (2004). *Helping Children Left Behind*. Cambridge, MA: MIT Press.

Further Reading

- Bénabou, R. (1993). Workings of a city: Location, education, and production. *Quarterly Journal of Economics* **108**, 619–652.
- Briffault, R. (2006). Adding adequacy to equity: The evolving legal theory of school finance reform. *Columbia University Law School Public Law and Legal Theory Research Paper Series Research Paper No. 06-111 and Princeton University Program in Law and Public Affairs Research Paper Series Research Paper No. 06-013*, Fall Semester.
- Brunner, E. J. and Sonstelie, J. (2003). School finance reform and voluntary fiscal federalism. *Journal of Public Economics* **87**(9–10), 2157–2185.
- Card, D. and Payne, A. (2002). School finance reform, the distribution of school spending, and the distribution of student test scores. *Journal of Public Economics* **83**, 49–82.
- Courant, P. N. and Loeb, S. (1997). Centralization of school finance in Michigan. *Journal of Policy Analysis and Management* **16**, 114–136.
- Dee, T. S. and Levine, J. (2004). The fate of new funding: Evidence from Massachusetts' education finance reforms. *Educational Evaluation and Policy Analysis* **26**(3), 199–215.

- Downes, T. A. and Shah, M. (2006). The effect of school finance reform on the level and growth of per pupil expenditures. *Peabody Journal of Education* **81**(3), 1–38.
- Fernández, R. and Rogerson, R. (1996). Income distribution, communities, and the quality of public education. *Quarterly Journal of Economics* **111**, 135–164.
- Hoxby, C. M. and Kuziemko, I. (2004). Robin Hood and his not-so-merry plan: Capitalization and the self-destruction of Texas's school finance equalization plan. *NBER Working Paper No. 10722*. Cambridge, MA: National Bureau of Economic Research.
- Husted, T. A. and Kenny, L. W. (2000). Evidence on the impact of state government on primary and secondary education and the equity-efficiency trade-off. *Journal of Law and Economics* **43**, 285–308.
- Nechyba, T. J. (2000). Mobility, targeting, and private school vouchers. *American Economic Review* **90**, 130–146.
- Sonstelie, J., Brunner, E., and Ardon, K. (2000). *For Better or Worse? School Finance Reform in California*. San Francisco, CA: Public Policy Institute of California.

Relevant Websites

- <http://www.cpre.org> – Consortium for Policy Research in Education (CPRE).
- <http://nces.ed.gov> – Education Finance Statistics Center (EDFIN) of the National Center for Education Statistics.
- <http://www.schoolfunding.info> – National Access Network.
- <http://irepp.stanford.edu> – Research on Education Policy and Practice, Getting Down to Facts: A Research Project Examining California's School Governance and Finance Systems.
- <http://www.ruraledu.org> – The Rural School and Community Trust.
- <http://www.albany.edu> – University at Albany, Education Finance Research Consortium (EFRC).

School Finance: An Overview

J K Rice, University of Maryland, College Park, MD, USA

D H Monk, The Pennsylvania State University, University Park, PA, USA

J Zhang, University of Maryland, College Park, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

School finance – The process of raising, distributing, and utilizing money for the purpose of providing educational services.

Introduction

School finance is a broad and evolving field encompassing three resource-related functions – raising revenue, allocating resources, and using resources – all aimed at providing educational opportunities and producing educational outcomes. All of these activities occur in a broader context of educational goals and societal values that shape how finance systems are structured and executed. In this article, we provide an overview of school finance, emphasizing the enduring challenges and highlighting new ways of thinking about them. We begin by exploring contextual factors that influence school-finance decisions; this includes a discussion of the goals and purposes of public education and the broader societal values that frame public finance. The next section describes traditional and contemporary mechanisms for raising revenue to support education systems, and examines the role of different levels of government in supporting public education. The section that follows discusses how resources are allocated across education systems, and emphasizes the evolution of equity and efficiency considerations in our allocation decisions. We conclude with a set of critical resource-utilization issues that are at the center of current school-finance-policy deliberations.

Throughout the article, our goal is to provide a sense of how this field has evolved, and we emphasize longstanding challenges and new ways of thinking about enduring issues. With the goal of offering an overview of the field, we prioritize breadth over depth and refer readers to other sources for additional information on the complexities and nuances of the topics covered. In addition, many of the ideas touched on in this article are given more in-depth attention elsewhere in this encyclopedia. Our focus here is on K-12 public-education systems in the United States. Issues related to financing higher education, nonpublic education, and education in international contexts are addressed in other articles elsewhere in this encyclopedia.

School Finance in Context

The goals and purposes of public education are to meet both individual and societal demands for schooling. In general, purposes of education are to produce individuals who can contribute to the economic, political, civic, social, and cultural institutions in our society. As such, we expect high school graduates to have acquired a wide range of competencies, skills, and personal qualities. These ideas are consistent with those of early proponents of public education including Horace Mann and Thomas Jefferson. Further, they reflect the work of international efforts to identify the array of “key competencies that contribute to a successful life and a well-functioning society” (Rychen and Salganik, 2003).

While many of the benefits of formal education are enjoyed by individuals in the form of better employment opportunities, higher wages, better health, expanded options for leisure time, and a better life for themselves and their children, it is the social benefits of education that justify a publicly financed system of schools. These social or collective benefits include all of the benefits enjoyed by individuals in the society, plus additional benefits that are uniquely collective in nature (Cohn and Geske, 2004). For instance, from a societal perspective, investments in public education pay returns in the form of national economic productivity and growth, good citizenship and a working democracy, a more peaceful society, and lower costs associated with prisons and social services. Analysts typically classify educational benefits into a four-cell matrix, as shown in **Table 1**.

Calculating the returns on investments in education has been the focus on research, and estimates depend on a number of contextual factors including the developmental status of the nation, and the level and type of education (Psacharopoulos, 2006). Taken together, studies show that education is a good investment for individuals and for society, with a rate of return typically exceeding 10% (Becker, 1993). Estimating the full returns to education has long been a challenge due to the difficulties of quantifying many of the nonmarket benefits, though some progress has been made capturing the value of benefits like lower crime rates, better child education, and household health resulting from investments in education (e.g., Baum and Payea, 2005; Barnett and Masse, 2007).

Table 1 Classification of the benefits of education

<i>Benefit type</i>	<i>Private</i>	<i>Social</i>
Market	Employability Higher earnings Labor-market flexibility	Higher productivity Higher net tax revenue Less reliance on government financial support
Nonmarket	Greater mobility Greater consumer efficiency Better household health Nonwage remuneration (fringe benefits and working conditions) Future opportunities for children	Lower crime rates Less spread of infectious diseases Better social cohesion Civic participation (Voter participation, volunteerism, charity; service in public agencies) Technological change

From Wolfe, B. and Zuvekas, S. (1997). Non-market effects of education. *International Journal of Education Research* 27(6), 491–502; Psacharopoulos, G. (2006). The value of investment in education: Theory, evidence, and policy. *Journal of Education Finance* 32(2), 113–136.

The estimated benefits of education, particularly education at the elementary level that gives rise to basic literacy skills, make a compelling case for a public investment in education. Each year, hundreds of billions dollars in federal, state, and local revenues are dedicated to public K-12 education (NCES, 2006). From 1982 to 2003, total revenues allocated to public K-12 education in the United States steadily rose from around US\$117 billion to US\$462 billion. Decisions about how best to raise and allocate these resources are influenced by three broad and sometimes competing goals: efficiency, equity, and liberty (Garms *et al.*, 1978). While the ideas presented here are relatively straightforward, each of these goals is multifaceted and complex. For instance, our definition of efficiency deals with production efficiency, but efficiency in exchange is also an important consideration (see Monk, 1990). The goal of efficiency holds that resources should be used to pursue the best set of outcomes in ways that minimize the use of resources. The goal of equity emphasizes the fairness in the distribution of a good, service, or burden. The goal of liberty holds that revenue generation and resource allocation should be conducted in a way that properly balances individual in contrast to collective interests. While each of these goals is important in its own right, they are often in tension with one another requiring policymakers to strike a reasonable balance among them as they consider options for raising, distributing, and utilizing resources to realize the goals of public education.

Raising Revenue: Multiple and Evolving Roles

Historically, school finance in the United States has largely been a local function. For a history of education finance, see Guthrie *et al.* (2007). Families and communities raised resources to provide local schools so that children would learn the knowledge, skills, and values needed to be competent and productive adult members of the community. As the broader civic, social, and economic benefits of public education became more apparent, state laws requiring children to attend school were adopted. Compulsory-education laws, first enacted in the mid-1600s, obligated states to establish school systems to finance and administer public education. By the mid-1800s, systems of universal, tax-support education involving multiple levels of government spread throughout the country. Most often, these systems were organized around school districts that were responsible for raising revenue, typically through the use of the local property tax, and providing educational services for students. This local system of education finance coupled with the uneven distribution of wealth across districts, however, resulted in large disparities across communities in education spending, services, and outcomes. Despite the efficiencies expected to result from this decentralized system of school finance (i.e., the potential for local systems to better meet the preferences of their constituencies), the inherent inequities associated with such a heavy reliance on the local property tax for revenue generation led to court cases challenging the legality of state systems of school finance. For a detailed review of school-finance litigation, see Minorini and Sugarman, 1999. One outcome of these court cases was a gradual shift to more state involvement in revenue generation for public education. Over the period from 1919 to 2003, the state share of revenue for public education increased from 17% to 49%, while the local share decreased from 83% to 43% (NCES, 2006). This shift in the share of revenue provided by states and districts reflects a normative shift in school finance from the traditional emphasis on liberty to a greater emphasis on equity that could only be ensured through an increase in state involvement.

Since education is fundamentally a state responsibility, the federal role has always been relatively modest, typically accounting for about 7% of total government revenues for K-12 public education (NCES, 2006). In general, federal funding for education has been motivated by three key concerns: ensuring opportunity for all students, countering underinvestment that might result in national labor shortages, and realizing scale economies through national research and development efforts (Guthrie *et al.*, 2007). Federal expenditures to promote equity include programs like Head Start that provides early educational opportunities for disadvantaged students, the Elementary and Secondary Education Act that supports education for

low-income students, and Public Law 94–142 that funds education for handicapped students. Federal policies to promote efficiency include vocational-education initiatives, the National Assessment of Educational Progress (NAEP) and other efforts to monitor the achievements of schools, and data collection and research efforts to guide decision making. Finally, the federal government has promoted the goal of liberty through policies that enhance freedom and choice. Arguably, the No Child Left Behind Act of 2001 (NCLB) advances all three goals by making federal funds for low-income students contingent on demonstrated performance (adequate yearly progress) for all subgroups of students served, and providing alternatives for students in chronically failing schools. However, more than 5 years since its adoption, NCLB has fallen short on a number of counts and the level of federal investment in public education has become a matter of contentious debate (Sunderman and Orfield, in press).

In addition to local, state, and federal revenue sources, many schools benefit from funding and resources from nongovernment sources. Since states vary in their reporting requirements, available information about the types, amounts, and distributions of nongovernmental resources is limited and uneven. Existing evidence shows that private contributions to public schools can be substantial, giving rise to related questions about the implications these resources have for efficiency, equity, and liberty (Schwartz, *et al.*, 2002).

The different levels of government tend to rely on different types of revenue-raising instruments, and these instruments have been studied in terms of their effects on equity and efficiency. As noted above, local districts tend to rely most heavily on the property tax, states have typically used a combination of sales and income taxes, and the federal government relies primarily on income taxes. Evaluations of these taxes tend to focus on five criteria: tax base, yield, equity, economic effects, and administrative and compliance costs (Odden and Picus, 2004). The most desirable taxes are progressive approaches that have a broad base and a low rate, a stable yield, minimal economic effects, and low administrative costs. Education-finance systems typically incorporate a combination of taxes across levels of the education system, balancing the strengths and weakness of each instrument. Further, as is discussed in the next section, intergovernmental grants are often used by federal and state agencies to encourage local districts to assume their share of the tax burden for public education.

In addition to the more traditional taxes used to raise funds for education, lotteries have been considered by many state legislatures as a potential revenue source for public education. Lotteries provide fungible revenues that are often earmarked for public education (Novarro, 2005). Lotteries are popular alternatives in the face of constrained resources, and earmarking lottery profits for K-12

education tends to increase spending (Evans and Zhang, 2007). However, revenue from lotteries tends to be unstable and regressive, qualities that make this strategy less attractive – on both practical and normative grounds – as a sustainable means to generate revenue for public education.

Distributing Resources: Multiple and Competing Goals

As described above, states assumed a more active role in school finance following court cases that challenged the equity of heavy reliance on the local property tax to fund education. In this section, we describe the dominant approaches used by states over the years to allocate funds for public education. Next, we review school finance litigation to highlight the key equity considerations in the distribution of resources for public education. Finally, we discuss issues related to the efficient use of educational resources.

Mechanisms for Distributing Revenue Across School Districts

Several state equalization formulas have become common tools to redistribute funds in more equitable ways. Four approaches have dominated the landscape. A more detailed treatment of these formulas is provided by Odden and Picus (2004) and Monk (1990). Flat grants have been used to ensure that all schools are funded to provide a basic education for students. These programs provide all schools a grant of equal size (either by school or by enrolment) to help support elementary education programs. However, by treating all schools the same, flat grants fail to recognize the varying fiscal capacity of different districts that gives rise to inequities in the first place.

Foundation programs were introduced in the early 1900s in direct response to the shortcomings of flat grants. The foundation program is rooted in the philosophy that states have an obligation to provide a minimum level of education. The state sets a foundation, the per-pupil expenditure needed to provide a minimum quality education, and requires a minimum tax rate to ensure local effort as a condition for state aid. States fund the difference between the per-pupil expenditure that districts are able to generate at the minimum tax rate and the state-established foundation per-pupil expenditure. In this way, states distribute funds inverse to local wealth with the goal of helping all school districts to provide a minimum foundation of education. Districts are free to tax themselves above the minimum required rate to supplement this foundation level of spending, but the revenues generated beyond the required tax rate are a function of local wealth alone. So long as the foundation level and

the minimum tax rates are set at relatively low levels, inequities stemming from differences in the fiscal capacity of school districts remain a problem.

Guaranteed tax base (GTB) programs were introduced in the 1970s in response to school-finance litigation based on unequal local fiscal-capacity arguments. A GTB program guarantees, through the allocation of state aid, that each school district in a state can function as if it had an equal tax base. Essentially, the state establishes the tax base that will be guaranteed for all school districts and provides aid such that any district with an actual tax base less than the GTB will generate revenue as if they had the GTB. This program grants school districts the liberty to determine their own tax rate, and equalizes fiscal capacity up to the level of the GTB. State aid is awarded inversely to local wealth, thereby improving equity without limiting liberty. However, the cost to the state can be excessive.

The fourth formula that states have used to allocate funds to support education is a combination of the foundation program and GTB program. This combination approach overcomes the problems of each of the formulas; the foundation program requires that districts provide at least a basic level of education, and the GTB ensures equity if districts choose to spend beyond the foundation level.

Beyond these four basic formulas, states have employed a variety of strategies to promote greater equity. For instance, state-determined-spending programs prescribe the per-pupil expenditure across districts in the state. This per-pupil expenditure may be funded by the state (e.g., Hawaii) or some combination of state and local revenues (e.g., Vermont). States have also placed revenue limits on local school districts to restrict the difference in spending across jurisdictions (e.g., California). While these sorts of programs are intended to promote greater equity, they can have serious implications for liberty.

Regardless of the state formula used to distribute funds to districts, additional adjustments are commonly made to account for student needs and geographic cost differences across school districts. Adjustments for special-needs students tend to take the form of weights that recognize the higher-than-average costs associated with educating low-income, limited-English-proficient, and special-education students. If these students were evenly distributed across schools and districts, such weights would not be necessary. However, some districts have higher concentrations of such students and require additional funding to provide education programs. States provide this additional funding through weighted student adjustments, but some research indicates that states tend to underestimate the influence of poverty and limited-English-proficiency status on cost (Baker and Duncombe, 2004). Determining student eligibility and the magnitude of the weights to be assigned to various types of students is a process that varies across

states and involves a number of assumptions about the costs of providing education services to students with special needs. For additional discussion, see Chambers *et al.* (2002) and Duncombe and Yinger (2005).

A second type of adjustment recognizes differences in what a dollar can purchase across different jurisdictions. There is broad consensus that the cost of educational inputs varies geographically due to differences in local labor markets, house prices, transportation costs, energy prices, and the like. Despite the intuitive appeal, measuring cost differentials is difficult due to the scarcity of geographically detailed data on the prices and quality of inputs purchased. Nonetheless, some headway has been made and various estimates of geographic cost differentials in education, mostly focused on teachers, now exist. A prime example is the geographic cost of education index (GCEI) developed by Chambers (1997) which adjusts teacher salary to account for differences in workplace characteristics. More recently, Taylor and Fowler (2006) developed a comparable wage index (CWI) to adjust for differences in the cost of education that are beyond the control of school districts. The CWI is constructed on salary differentials of noneducators and reflects the systematic, regional labor-cost variation due to differences in both the cost of living and local-community characteristics (such as crime rate).

Equity and Adequacy

The distribution formulas and weighting mechanisms that states use to allocate resources across school districts are a direct response to demands for greater equity in education finance. But defining what specifically, constitutes equity has been an evolving process, one driven largely by the courts. School-finance litigation has been categorized into three waves (Guthrie *et al.*, 2007). During the first wave, which spanned the 1960s through 1973, plaintiffs challenged school-finance systems through the federal Constitution's equal protection clause. The second wave, from 1973 through 1989, was characterized by challenges based on state-education clauses, equal-protection clauses, or a combination of the two. The third wave, from 1989 to the present, includes cases based on claims that school-finance formulas prevent poor school districts from providing an adequate education as defined by state-education clauses. Research suggests that litigation, and the public and judicial pressure resulting from litigation, has improved equity. In most cases, this is a result of leveling up, or increasing overall state funding such that the additional funds are allocated to lower-wealth districts (Evans *et al.*, 1999).

The school-finance adequacy litigation expands the traditional equity paradigm to consider what is needed to achieve a "high-minimum quality education for all"

(Minorini and Sugarman, 1999: 188). Several methodologies have been developed to determine an adequate expenditure level, including the professional-judgment approach (Guthrie and Rothstein, 1999), the successful-district approach (Augenblick, 1997), and the cost-function approach (Reschovsky and Imazeki, 2003). Despite the efforts to estimate the cost of an adequate education, no consensus exists as to how much money is needed to reach the adequate level. Some researchers suggest that the evidence is converging (Odden *et al.*, 2004).

Efficiency

Earlier we noted that the goal of efficiency is served to the degree that the best outcomes are being produced in ways that minimize the use of resources. Efficiency is always a matter of degree and policymakers at all levels of educational systems have an ongoing responsibility to improve efficiency within their areas of responsibility.

While the efficiency concept is seemingly straightforward, it is actually riddled with complexities. Notice, for example, the stipulation about the selection of the best outcomes. A system could be minimizing the use of resources in pursuit of useless outcomes and fall far short of the efficiency goal. Similarly, a system could be pursuing the right outcomes, but by doing so in ways that are wasteful, thereby again fail to achieve efficiency. Moreover, in the real world, it can be difficult to reach consensus about what counts as best. Measurement problems also abound and it can be devilishly difficult to disentangle available fiscal data so that a good map can be made showing what resources are actually being expended in pursuit of what goals.

There has been great interest recently in measuring the value added by educational resources, particularly teacher resources (Sanders and Rivers, 1996). The logic is that teachers vary in how much value they add to their students' performance. If these value-added contributions can be measured accurately, then policymakers could take steps to increase the efficiency of the system. Shifts could be made in personnel policies to increase the incidence of teachers with high value-added scores. But these high value-added teachers may not be available at prevailing wages. If it becomes necessary to pay higher salaries in order to attract and retain the high value-added talent, the efficiency gains associated with having the more productive teachers in place could be undermined.

Attempts are also being made to use accountability systems to stimulate efficiency, although typically, the focus is more on achieving performance targets rather than on minimizing costs. Efficiency will be enhanced to the degree that the accountability system stimulates performance, regardless of the means, without adding to costs.

Utilization of Resources: Current Policy Issues for School Finance

Education expenditures are of interest to the extent that they are used to purchase resources that translate into meaningful learning experiences for students, and, ultimately, desired outcomes like economic productivity, social responsibility, and civic participation. In this section, we address three current policy issues that have implications for education-finance systems and, specifically, how educational resources are utilized.

Teacher Compensation

A first set of policy issues relates to teachers and teacher compensation. Teacher compensation consumes more than half of K-12 public-education operating expenditures. This substantial investment in teachers is justified on the grounds that teachers are the most important school resource provided to students (e.g., Sanders and Rivers, 1996). Concern about the supply and quality of teachers, particularly in geographic and subject-shortage areas, has induced policymakers and researchers to consider how compensation can be used to attract and retain quality teachers into the profession and to the schools that need them the most.

While most school districts rely on traditional single-salary schedules, some districts have begun to experiment with alternative compensation structures that may be more efficient and equitable for teachers, students, and taxpayers (see Podgursky and Springer, 2007). In addition, districts have used a variety of economic incentives – for example, signing bonuses, tuition remission for university credits, and housing assistance – to attract teachers to geographic and subject-shortage areas, but we have little evidence on the impact of these investments (Guarino *et al.*, 2004). Clearly, more work is needed to understand how to structure teacher compensation in ways that advance the overarching goals of public education.

Special Education

A second highly contested area in the utilization of resources is special-education finance. Almost six-million children receive federally funded special-education services with total expenditure estimated around US\$50 billion (Chambers *et al.*, 2004). States and districts support special education with a variety of funding sources and distribution formula, and studies show that the proportion of total education expenditure spent on special education has increased over time (Kolbe and O'Reilly, 2005).

The level of funding needed for special education is a function of the number of students in need of services, and the cost of those services. Both identifying eligible students and determining the appropriate funding levels

for those students have presented challenges to policy-makers. The rising number of students in less-severe, high-incidence special-education categories accounts for the majority of growth in special-education enrolments and spending (Parrish and Esra, 2006). Researchers have considered the extent to which state policies provide fiscal incentives and disincentives that might affect the identification of special-education students (Mahitivanichcha and Parrish, 2005), giving rise to debates about how state-funding structures might lead to overidentification.

In addition, policymakers struggle with challenges related to the most effective and efficient delivery of special-education services. This includes designing and implementing more effective alternatives to provide special education in the general-education setting, particularly for students with mild-to-moderate learning disabilities. More research is needed on the relative cost-effectiveness of different program and service options (Kolbe and O'Reilly, 2005). Establishing reasonable special-education service standards could help control costs, narrow the variation in per-pupil expenditures across systems, and guide determinations about special-education funding adequacy. For a detailed discussion of special education adequacy, see Harr *et al.* (2006). This requires a better system of periodic, reliable data on what is spent on special education and the student outcomes that result.

Expanding the Scope of Education

A final example of resource utilization in school finance relates to expanding the scope of educational services. Initiatives like investing in early-childhood education and linking education with other social services are typically advanced as efforts to improve the education opportunities for students from economically disadvantaged families. Research suggests that these kinds of initiatives may also have important implications for efficiency.

While policies aimed at narrowing the achievement gap typically target school-age children, research shows that a substantial gap exists at the outset of formal schooling (Lee and Burkham, 2002). Proponents of early-education programs recognize that investments in the young have relatively high returns (Carneiro and Heckman, 2003). Cost-benefit analyses of preschool programs for disadvantaged children have shown that the long-term effects of early education are associated with monetary benefits that exceed the costs of the programs (see Belfield *et al.*, 2006; Temple and Reynolds, 2007; Ludwig and Phillips, 2007). While the price tag associated with these programs may be high, the investment may be wise on both equity and efficiency grounds.

The conventional bounds of public education finance are also challenged by proposals to expand the role of schools. The idea is that efforts to narrow the achievement

gap must recognize the wide range of social and economic factors – including healthcare quality, nutrition, housing quality and stability, parental occupation and aspirations – that affect student achievement (Rothstein, 2004). Proposals to address these multiple influences may involve greater collaboration between schools and other social service fields, or they might require providing a broader set of services (e.g., health and dental care) to low-income students within the school setting. Either approach has implications for fiscal policy and the potential to advance goals of efficiency, equity, and liberty.

See also: An Overview of Teacher Labor Markets; Economic Approaches to Adequacy; Economic Approaches to School Efficiency; School Finance Reform; The Economic Role of the State in Education; The Economics of School Accountability.

Bibliography

- Augenblick, J. (1997). *Recommendations for a Base Figure and Pupil-Weighted Adjustments to the Base Figure for Use in a New School Finance System in Ohio*. Columbus, OH: Ohio Department of Education.
- Baker, B. and Duncombe, W. D. (2004). Balancing districts needs and student needs: The role of economies of scale adjustments and pupil need weights in school finance formulas. *Journal of Education Finance* 29(3), 195–222.
- Barnett, W. S. and Masse, L. N. (2007). Comparative benefit-cost analysis of the Abecedarian program and its policy implications. *Economics of Education Review* 26, 113–125.
- Baum, S. and Payea, K. (2005). Education pays 2004: The benefits of higher education for individuals and society. *Trends in Higher Education Series. The College Board*. http://www.collegeboard.com/prod_downloads/press/cost04/EducationPays2004.pdf (accessed June 2009).
- Becker, G. S. (1993). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, 3rd edn, ch. 2, pp 15–25. Chicago, IL: University of Chicago Press.
- Belfield, C. R., Nores, M., Barnett, S., and Schweinhart, L. (2006). The high/scope Perry preschool program: Cost-benefit analysis using data from the age-40 follow-up. *Journal of Human Resources* 41(1), 162–190.
- Chambers, J. (1997). *Geographic Variations in Public School Costs*. Washington, DC: US Department of Education, National Center for Education Statistics.
- Chambers, J., Parrish, T., and Harr, J. J. (2000). What are we spending on special education services in the United States, 1999–2000. *Report*. Palo Alto, CA: Center for Special Education Finance, Special Education Expenditure Project.
- Cohn, E. and Geske, T. G. (2004). *The Economics of Education*, 3rd edn. Mason, OH: South-Western Publishing, Thomson Learning.
- Duncombe, W. D. and Yinger, J. M. (2005). How much more does a disadvantaged student cost? *Economics of Education Review* 24(5), 513–532.
- Evans, W. N., Murray, S. E., and Schwab, R. M. (1999). The impact of court-mandated school reform. In Ladd, J., Chalk, R., and Hansen, J. (eds.) *Equity and Adequacy in Education Finance: Issues and Perspectives*, pp 72–98. Washington, DC: National Research Council.
- Evans, W. N. and Zhang, P. (2007). The impact of earmarked lottery revenue on K–12 educational expenditures. *Education Finance and Policy* 2(1), 40–73.
- Garms, W. I., Guthrie, J. W., and Pierce, L. C. (1978). *School Finance: The Economics and Politics of Public Education*. Englewood Cliffs, NJ: Prentice-Hall.

- Guarino, C., Santibanez, L., Daley, G., and Brewer, D. (2004). *A Review of the Research Literature on Teacher Recruitment and Retention (No. TR-164-EDU)*. Santa Monica, CA: RAND Corporation.
- Guthrie, J. and Rothstein, R. (1999). Enabling adequacy to achieve reality: Translating adequacy into school finance arrangements. In Ladd, H. F., Chalk, R., and Hansen, J. (eds.) *Equity and Adequacy in Education Finance: Issues and Perspectives*, pp 209–259. Washington, DC: National Academy Press.
- Guthrie, J. W., Springer, M. G., Rolle, R. A., and Houck, R. A. (2007). *Modern Education Finance and Policy*. Boston, MA: Pearson Education.
- Harr, J. J., Parrish, T., Chambers, J., Levin, J., and Segarra, M. (2006). *Considering Special Education Adequacy in California*. Washington, DC: American Institutes for Research.
- Kolbe, T. and O'Reilly, F. (2005). *OSEP Technical Assistance in Planning to Obtain Special Education Finance Data and in Dissemination of Currently Available Data*. Cambridge, MA: Abt Associates.
- Lee, V. and Burkham, D. (2002). *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School*. Washington, DC: Education Policy Institute.
- Ludwig, J. and Phillips, D. A. (2007). *The Benefits and Costs of Head Start*. Ann Arbor, MI: The National Poverty Center.
- Mahitivanichcha, K. and Parrish, T. (2005). The implications of fiscal incentives on identification rates and placements in special education: Formulas for influencing best practice. *Journal of Education Finance* **31**(1), 1–22.
- Minorini, P. A. and Sugarman, S. (1999). Educational adequacy and the courts: The promise and problems of moving to a new paradigm. In Ladd, H. K., Chalk, R., and Hansen, J. (eds.) *Equity and Adequacy in Education Finance: Issues and Perspectives*. Washington, DC: National Academy Press.
- Monk, D. H. (1990). *Educational Finance: An Economic Approach*. New York: McGraw-Hill.
- NCES (National Center for Education Statistics) (2006). *Digest of Education Statistics*. Washington, DC: National Center for Education Statistics.
- Novarro, N. K. (2005). Earmarked lottery profits: A good bet for education finance? *Journal of Education Finance* **31**(1), 23–44.
- Odden, A. R. and Picus, L. O. (2004). *School Finance: A Policy Perspective*, 3rd edn. New York: McGraw-Hill.
- Odden, A., Picus, L. O., Fermanich, M., and Goetz, M. (2004). An evidence-based approach to school finance adequacy in Arizona. *Report prepared for the Rodel Charitable Foundations*. Phoenix, AZ: Rodel Charitable Foundation of Arizona.
- Parrish, T. and Esra, P. (2006). The special education expenditure project (SEEP): Synthesis of findings and policy implications. In Forum.
- Podgursky, M. and Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management* **26**(4), 909–949.
- Psacharopoulos, G. (2006). The value of investment in education: Theory, evidence, and policy. *Journal of Education Finance* **32**(2), 113–136.
- Reschovsky, A. and Imazeki, J. (2003). Let no child be left behind: Determining the cost of improved student performance. *Public Finance Review* **31**(3), 263–290.
- Rothstein, R. (2004). *Class and Schools: Using Social, Economical and Educational Reform to Close the Black–White Achievement Gap*. Washington, DC/New York: Economic Policy Institute/Teachers College, Columbia University.
- Rychen, D. S. and Salganik, L. (2003). *Key Competencies for a Successful Life and a Well-Functioning Society*. Toronto, ON: Hogrefe and Huber.
- Sanders, W. L. and Rivers, J. C. (1996). *Research Progress Report: Cumulative and Residual Effects of Teachers on Future Student Academic Achievement*. University of Tennessee Value-Added Research and Assessment Center.
- Schwartz, A. E., Amor, H. B. H., and Fruchter, N. (2002). Private money/public schools: Early evidence on private and non-traditional support for New York City public schools. In Rice, J. K. and Roellke, C. F. (eds.) *Fiscal Policy Issues in Urban Education*, pp 231–253. Greenwich, CT: Information Age Publishing.
- Sunderman, G. and Orfield, G. (in press). No child left behind reforms and the state administrative response. In Rice, J. K. and Roellke, C. F. (eds.) *The Costs of Accountability*. Information Age Publishing.
- Taylor, L. L. and Fowler, W. J. Jr. (2006). *A Comparable Wage Approach to Geographic Cost Adjustment (NCES 2006-321)*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Temple, J. A. and Reynolds, A. J. (2007). Benefits and costs of investments in preschool education: Evidence from the child–parent centers and related programs. *Economics of Education Review* **26**, 126–144.
- Wolfe, B. and Zuvekas, S. (1997). Non-market effects of education. *International Journal of Education Research* **27**(6), 491–502.

Further Reading

- Alexander, K. and Salmon, R. G. (1995). *Public School Finance*. Boston, MA: Allyn and Bacon.
- Baker, B. D., Green, P., and Richards, C. E. (2008). *Financing Education Systems*. Upper Saddle River, NJ: Pearson Education.
- Burtless, G. (1996). *Does Money Matter: The Effect of School Resources on Student Achievement and Adult Success*, pp 97–140. Washington, DC: Brookings.
- King, R. A., Swanson, A. D., and Sweetland, S. R. (2003). *School Finance: Achieving High Standards with Equity and Efficiency*, 3rd edn. Boston, MA: Pearson Education.
- Ladd, H. F. and Hansen, J. S. (1999). *Making Money Matter: Financing America's Schools*. Washington, DC: National Academy Press.
- Levin, H. M. and McEwan, P. J. (2002). *2002 Yearbook of the American Education Finance Association: Cost-Effectiveness and Educational Policy*. New York: Eye on Education.
- Odden, A. and Busch, C. (1998). *Financing Schools for High Performance: Strategies for Improving the Use of Educational Resources*. San Francisco, CA: Jossey-Bass.
- Odden, A. and Kelley, C. (2002). *Paying Teachers for What They Know and Do: New and Smarter Compensation Strategies to Improve Schools*. Thousand Oaks, CA: Corwin Press.
- Swanson, A. D. and King, R. (1997). *School Finance: Its Economics and Politics*, 2nd edn. New York: Longman.
- Thompson, D. C. and Wood, R. C. (1998). *Money and Schools: A Handbook for Practitioners*. New York: Eye on Education.

The Economics of Tuition and Fees in Higher Education

R G Ehrenberg, Cornell University, Ithaca, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Net tuition – Tuition paid by students minus the average amount of institutional and other grant aid they receive.

Tuition discount rate – Percentage of undergraduate tuition revenue that goes back to undergraduate students in the form of institutional grant aid.

Introduction

Tuition and fee policies in higher education vary widely across countries and, in some countries, widely across academic institutions within the country. This article focuses on one country, the United States. Readers interested in policies in other nations can consult the website of the Center for International Higher Education at Boston College, which provides a comprehensive listing of international higher education journals and has numerous working papers dealing with the subject.

The American higher education system is a highly decentralized one. There are over 4200 degree-granting institutions in the United States. While the majority of institutions are private, not for profits, about two-thirds of all undergraduate students enrolled in four-year colleges attend public institutions; this share rises to about four-fifths when one takes into account the students enrolled in 2-year colleges, which are predominately public institutions. A small but growing share of students are enrolled in private for-profit degree-granting institutions.

For a number of years, the College Board has been collecting data on tuition and fees and publishing it in annual volume *Trends in College Pricing*. In 2007–08, the average tuition and fees at 2-year public institutions, 4-year publics, and 4-year privates were US\$2361, US\$6185, and US\$23 712, respectively (College Board, 2007). The figures for public institutions are for students who reside in the same state as the institution is located; students from out of the state and foreign countries are charged higher levels of tuition at the publics. The average public 4-year out-of-state tuition and fees was US\$16 640 during that period.

These averages mask considerable variation in the tuition and fees that students pay to attend both public

and private institutions. For example, in 2007–08 about 19% of the students attending private 4-year institutions faced tuition and fees that were over US\$33 000, while almost 18% faced tuition and fees that were less than US\$15 000. Similarly, while 45% of the in-state and out-of-state students attending public 4-year institutions faced tuition and fee levels that were less than US\$6000, 13% faced tuition and fee levels that were more than US\$12 000.

It is important to stress that an institution's tuition and fee level does not reflect the full cost that the institution incurs in educating undergraduate students. Gordon Winston from Williams College has very carefully documented that no undergraduate student at a public or private non-profit college or university (including those who receive no financial aid) pays the full cost of his or her education because of subsidies that are provided to them (Winston, 1999). In private higher education, these subsidies come from the income that the endowments of the university provide, from annual-giving streams to the institutions provided by alumni, foundations, corporations, and other donors, and from the value of the services of the buildings on campus that were constructed using funds from external donors. In public higher education, these subsidies come from all of the previous sources, as well as from current state appropriations and previous state appropriations for buildings and capital equipment. For future reference, it is important to note that Winston has also documented that the subsidies students receive are largest in absolute value at the nation's most selective (in terms of admission standards) and wealthiest private colleges and universities.

An institution's tuition and fee level also overstates the costs that many students incur to attend the institution because of financial aid provided to students by the federal government, state governments, the institutions themselves, and private donors. At the federal level, aid currently takes the form of grant aid for students from lower and lower-middle income families, subsidized loans for students from lower and middle income families, and tax credits. A number of states have grant-aid programs for students who reside in their state and attend colleges within the state. Sometimes, these grant programs are restricted to students attending public colleges in the state. Depending upon the state, these grant programs may be either need-based or merit-based; an example of the latter is the Georgia HOPE scholarship program. HOPE provides grant aid to Georgia students who attain a B average in high school and attend either a public or private academic

institution in the state. Continuation of the aid after a student's first year of college is contingent on the student's maintaining a B average in college. A number of states also have special scholarship programs or loan-forgiveness programs for students who are employed in the state after graduation in relatively low-paying occupations that are deemed to be of critical importance to the state.

Researchers and policymakers have long worried whether academic institutions will try to capture some of the grant aid that governments provide to help students afford college by raising their tuition levels. To date, the empirical evidence on whether this has occurred is ambiguous (Long, 2004).

Institutions also provide grant aid to students. Grant aid may be based upon financial need (as at the nation's most selective private colleges and universities) or it may be based upon merit. Merit is broadly defined here to include efforts by institutions to attract students with strong academic or strong athletic backgrounds. For a number of years, the National Association of College and University Business Officers (NACUBO) conducted an annual tuition-discounting survey to estimate how undergraduate grant aid provided by an institution compared to the tuition revenue that the institution was receiving; recently, this survey has been taken over by the College Board. These surveys suggest that the typical private American college or university has a discount rate in the range of 33%, in the sense that it returns roughly this share of the tuition revenue that it receives back to undergraduate students in the form of grant aid (Baum and Lapovsky, 2006). Of course, at the richest private institutions, much of these grant funds come from endowment income. Increasingly, public colleges and universities also provide grant aid to students; in recent years, the average tuition discount rate at public 4-year institutions was in the range of 15% of tuition revenues.

Tuition Keeps Rising in Private Higher Education

In 1967, William Bowen published an important book that examined what had happened to tuition and fee levels at a set of selective private colleges and universities in the United States during the first two-thirds of the twentieth century (Bowen, 1967). He found that, on average, the institutions increased their tuition and fee levels by 2–3% more than the rate of inflation (as measured by the increase in the consumer price index (CPI)) each year. He attributed this partially to the growth of new knowledge and graduate programs, which added costs to the academic institutions. But first and foremost, he attributed this to the fact that higher education institutions were not sharing in the productivity gains that were occurring in the rest of the economy.

To understand Bowen's argument, consider a very simple model in which tuition is the only source of revenue for an academic institution and faculty salaries are the only cost for the university. Suppose also that the institution believes that a high-quality undergraduate education is dependent upon maintaining a fixed student/faculty ratio. In such a world, faculty members' productivity does not increase over time; each year they educate the same number of students. However, in the rest of the economy, productivity is increasing because of investments in new capital equipment and changes in technology and since real wages (wages adjusted for inflation) depend upon productivity growth, they too are increasing.

In such a world, the administrators at the academic institutions face a dilemma. If they restrict tuition increases to the rate of inflation, salaries of faculty would remain constant (in inflation-adjusted terms) and would fall behind the earnings of people in other professional occupations. This would make it difficult to retain existing faculty members and to attract new people into graduate study and then faculty positions in academia. Inevitably then, the quality of higher education would decline. Alternatively, the administrators could bite the bullet, raise tuition by more than inflation, and try to keep the salaries of faculty members competitive with salaries in the rest of the economy. The pattern of tuition increases during the first two-thirds of the twentieth century suggests that they chose the second option.

As documented in *Tuition Rising* (Ehrenberg, 2000), throughout this period of time, tuition and fees at selective private higher education in the United States did not increase as a share of median family income, because median family income also increased more rapidly than inflation because of the growth in labor productivity and the growth in the number of two-income-earner families in the United States. However, the decade of the 1980s was one of virtually no real income growth in the United States, yet tuition kept increasing throughout this decade at rates greater than the rate of inflation. As a result, while tuition and fees at the typical selective private college and university in the United States was in the range of 30% of median family income in 1975, today it is over 50% of median family income. While the growth in financial aid has caused net tuition to rise by a smaller amount, it too is still much higher today as a share of median family income than it was 30 years ago.

In *Tuition Rising* there has been a detailed description of the forces that have allowed the selective private colleges and universities to keep increasing their tuition and fees at rates that substantially exceed the rate of increase in the CPI. A few of the more important ones have been highlighted in this article.

First, the selective private colleges and universities have but one objective; they want to be the very best that they can in every dimension of their activities. They

want to have the best instructional and research facilities, attract the best students and faculty, provide the highest quality education and support services, and the like. To do all these things takes money. While they try to diversify their revenue streams (by attracting more annual giving to support current operations, to build the endowment and to finance new construction, and by trying to commercialize their faculty members' research findings), in the absence of any market forces that limit tuition increases, increasing tuition is an easy way to generate increased revenues. After all, the institutions can always price discriminate (and they do); they can use a share of the increased revenue that they generate from a tuition increase to provide more grant aid to students who otherwise would not be able to afford to attend the institution to ensure that the tuition increase does not restrict access.

Each year, the number of students who apply to attend these institutions keeps getting larger, while the number of positions in their first-year students has increased only slightly. These institutions have focused primarily on maintaining or increasing their quality rather than increasing their size (although several are now marginally doing so) and despite their increased tuition and fees, students increasingly flock to them. This has occurred in large part because, as the distribution of earnings becomes more disperse in the United States, students and their parents instinctively understand that where one goes to college may matter as much as whether one goes to college. In what has become an increasingly winner-take-all society, there is increased pressure on students and their families, in the words of my colleague Robert Frank "to buy the best" (Frank and Cook, 1995). And empirically, with one exception, all empirical research on the topic suggests that students who attend selective private colleges and universities in the United States, which Winston's research indicates, provide the greatest subsidy to their students, are making a rational economic choice; they receive benefits in the form of higher post-college earnings and increased probability of being admitted to high-quality professional and graduate programs than would otherwise be the case. Put simply, as long as long lines of students keep knocking on the doors of these selective private institutions clamoring to gain admission to them, there are no market forces to hold the rate of increase in tuition down at these institutions.

Of course, these institutions could try to be socially responsible, improve the efficiency of their operations, and get better by substitution rather than by increasing expenditures. Often they do try to do this, especially in the nonacademic sphere of their operations. However, there are a number of forces that prevent them from doing this in the academic sphere of their operations. One is the system of shared governance that prevails within them that gives faculty members an extraordinary amount of influence over academic decisions. Faculty

members are crucial to the reputation of these institutions because they are the creators of new knowledge and the ones who educate undergraduate and graduate students. Administrators strive to keep the faculty happy because the labor market for faculty at these top institutions is very competitive and top faculty are mobile; often, this requires the administrators to make greater expenditures than they otherwise would prefer to do.

Another factor is the important role that alumni and other external constituents play. Alumni are vital to these institutions because they help recruit new students, provide internship opportunities and summer employment opportunities for existing students, provide job opportunities for graduates, and provide financial support in the form of donations for current operations, endowment, and new buildings. However, alumni often have very strong preferences for specific programs (often ones that they were involved with when they were students) and any attempt to cut back the scale of, or eliminate a program, may lead to a threat to withhold contributions or other sources of support for the institution if the action is undertaken. This makes it difficult for administrators to cut any program that an external constituent really cares about.

A third factor is the role that the annual *U.S. News & World Report* (*USNWR*) rankings, which in many respects is now the gold standard in rankings, now plays in American higher education. Empirical research shows that when an institution improves in the rankings, other factors held constant, it attracts more applicants, can be more selective and admit a smaller fraction of these applicants, a greater fraction of the students it admits will enroll at the institution, these enrolled students will have higher test scores, and the institution will be able to spend less on grant aid to attract the students (Monks and Ehrenberg, 1999). Conversely, when an institution falls in the rankings, just the reverse occurs. Administrators passionately care about these outcomes and while administrators often claim that they pay no attention to the *USNWR* rankings and that the rankings do not influence anything that goes on at the university, such statements are simply false.

The formula that *USNWR* uses to compute its rankings gives positive weight to the educational expenditures per student that an academic institution makes. Hence, any academic institution that unilaterally reduced its expenditures per student, or even the rate of growth of its expenditures per student relative to its competitors' rate of growth, would find that it would fall in the rankings. Hence, the *USNWR* ranking methodology does not reward academic institutions for holding their costs down; it puts pressure on the institutions to spend more.

Of course, to say that the selective private colleges and universities continue to have long lines of applicants clamoring to gain admission is not to say that most private colleges in the United States face this situation. In fact,

many private colleges and universities admit large fractions, if not virtually all, of the individuals that apply to them. Yet, these other private colleges and universities, for the most part, have increased their tuitions at roughly the same rates over time as the most selective privates.

They have been able to do this in part because of the belief that in the higher education market, posted price (tuition and fees) is taken by many to be an indicator of quality; if an institution lets its tuition fall relative to its competitors, it runs the risk of being perceived as an inferior institution. In part, they have been able to do this because students and their families do not seem to respond in a symmetric fashion to tuition increases and grant aid increases; Christopher Avery and Caroline Hoxby have found that in awarding accepted students a merit-based grant has a larger effect on their enrolment probabilities than reducing tuition by an equal amount (Avery and Hoxby, 2004).

Not surprisingly, tuition discount rates tend to be higher at smaller less-selective private liberal art colleges than they do at the more-selective private liberal art colleges and the private research universities. The more selective an institution is, the less it needs to use grant aid to try to attract students to it.

Tuition Keeps Rising at Public Institutions

During the last 30 years, rates of tuition and fee increase at 4-year public higher education institutions in the United States have been equal to or greater than the rates of tuition increases at the 4-year privates. However, while tuition increases at the privates have always been associated with increases in expenditures per student, tuition increases at the publics have often been associated with decreases in expenditures per student. This occurs because in addition to the factors described above that influence tuition growth in private higher education, tuition growth in public higher education is influenced by changes in the state appropriations that public higher education institutions receive. As described in *The Perfect Storm* (Ehrenberg, 2006), tight budgets faced by the states during the period due to several recessions, limitations on the willingness of the public to increase state tax rates, competing demands on state budgets from elementary and secondary education, Medicaid, and the criminal justice system, and large increases in college-age populations, have limited the ability of states to increase their support for their higher education institutions. On a real per-student basis, state appropriations per student today are only slightly greater than they were at the start of the period. As a result, tuition has risen and state appropriations have fallen as a share of public college and university budgets in the United States.

Although tuition has risen at roughly the same rate (in recent years roughly 3% more than the rate of increase in the CPI) at public and private higher education institutions in the United States, because state appropriations have failed to grow at a similar rate, expenditures per student have fallen in public higher education relative to that in private higher education. As a result, faculty salaries in public higher education have fallen relative to faculty salaries in private higher education, which makes it more difficult for the public institutions to attract and retain high-quality faculty. Student–faculty ratios have also risen in public higher education relative to those in private higher education. Great concern has been expressed that these changes are causing a decline in the relative quality of public higher education in the United States.

To try to reduce these financial pressures, public higher education institutions are pursuing a number of strategies. Public colleges and universities have always charged higher tuition and fee levels to students enrolling in their institutions who are not residents of the state in which the institution is located. Many public institutions have substantially increased the dollar gap between what in-state and out-of-state students are charged and, when possible, have sought to increase the share of their students they enroll from out-of-state as a way of generating additional revenues.

A number of them have moved toward charging differential tuition and fee rates for undergraduate students depending upon the program in which the student is enrolled. Historically, colleges and universities charged the same rate for all students; the notion being that students should be able to choose what to study based upon their intellectual interest, not based on costs. However, financial pressures have led public institutions to begin to charge higher tuition and fees for high-cost majors (due to small class sizes or the need for expensive equipment) and/or for majors that promise to lead to jobs with high economic returns. The intellectual justification for such strategies was articulated long ago by Stephen Hoenack and William Weiler (Hoenack and Weiler, 1975).

Graduate and Professional Program Tuition and Fees

The discussion above all deals with undergraduate education. American colleges and universities also enrol graduate students in professional degree programs in law (JD), medicine (MD), business and management (MBA), and a host of other fields. They also enrol students in research masters (MA or MS) and PhD programs. Tuition and fees for professional degree programs are often market determined and are often higher than the tuition charged to undergraduate students at the same institution. The

higher tuition levels are justified by the high earnings that entry into these professional fields promises to potential students. Students bear the cost of tuition and fees in most professional degree programs, although increasingly professional programs award both need- and merit-based aid to attract a diverse student body and to improve its academic profile.

In 2007–08, annual tuition and fees at many top private law, graduate business, and medical schools were in the US\$40 000–US\$45 000 range. With tuition levels this high, many students have extremely large loan burdens upon graduation. To encourage law students to enter low-paying public service law careers, many of the selective private law schools have funded their own programs that provide for loan forgiveness if graduates are employed in public service positions for a specified number of years. A recent study by Erica Field looked at an interesting quasi-experiment that the NYU law school conducted (Field, 2006). Applicants who expressed interest in entering public law careers were offered either a loan-forgiveness program or free tuition, with the understanding that the tuition would have to be paid back if they did not enter public interest law careers upon graduation. The two options were developed to have identical present value to participants. Field's striking finding was that providing tuition remission upfront was much more effective in inducing students to enter public interest law careers.

Tuition levels for students enrolled in most PhD programs are much more similar to undergraduate tuition levels. However, at the best private and public universities, very few students actually pay tuition on their own. Doctoral programs must provide financial support to doctoral students to induce them to enrol in their programs because the economic return upon graduating from a PhD program in many fields is relatively low. This support often comes in the form of a multi-year guarantee, with the support being either a fellowship (no work expectation), a teaching assistantship, or a research assistantship; students often have a variety of these forms of support during the time as doctoral students. Tuition remission and increasingly the provision of health insurance are part of these packages.

Concluding Remarks

Looking to the future, the most pressing issue relating to tuition and fees at American colleges and university is whether undergraduate tuition and fees can continue to increase in the years ahead at rates that exceed the rate of inflation. America no longer leads the world in the fraction of its recent cohorts of young people who attain college degrees. The major growth in its populations is coming from groups that traditionally have been under-represented in higher education, who are least able to

afford higher education. Congressional debate over the rate of increase in college tuitions has intensified and recent legislation winding its way through Congress, if enacted, would begin to publicize those institutions whose tuition increases are above average. A decade from now, an article in the next edition of this encyclopedia on tuition might look very different.

Bibliography

- Avery, C. and Hoxby, C. (2004). Do and should financial aid affect students' college choices? In Hoxby, C. (ed.) *College Choice*, pp 239–299. Chicago, IL: University of Chicago Press.
- Baum, S. and Lapovsky, L. (2006). *Tuition Discounting: Not Just a Private College Practice*. New York: The College Board.
- Bowen, W. G. (1967). *The Economics of the Major Private Research Universities*. Berkeley, CA: Carnegie Commission on Higher Education.
- College Board (2007). *Trends in College Pricing 2007*. New York: The College Board.
- Ehrenberg, R. G. (2000). *Tuition Rising: Why College Costs so Much*. Cambridge, MA: Harvard University Press.
- Ehrenberg, R. G. (2006). The perfect storm and the privatization of public higher education. *Change* **38**, 47–53.
- Field, E. (2006). Educational debt burdens and career choice: Evidence from a financial aid experiment at NYU Law School. National Bureau of Economic Research Working Paper No. 12282.
- Frank, R. H. and Cook, P. J. (1995). *The Winner-Take-All Society*. New York: Free Press.
- Hoenack, S. A. and Weiler, W. C. (1975). Cost-related tuition policies and university enrollments. *Journal of Human Resources* **10**, 332–360.
- Long, B. T. (2004). How do financial aid policies affect colleges? The Georgia HOPE scholarship. *Journal of Human Resources* **39**, 1045–1066.
- Monks, J. and Ehrenberg, R. G. (1999). U.S news and world reports college rankings: Why do they matter? *Change* **31**, 42–57.
- Winston, G. C. (1999). Subsidies, hierarchies and peers: The awkward economics of higher education. *Journal of Economic Perspectives* **13**, 13–36.

Further Reading

- Brewer, D. J., Eide, E. R., and Ehrenberg, R. G. (1999). Does it pay to attend an elite private college? Cross-cohort evidence on the effects of college type and earnings. *Journal of Human Resources* **34**, 104–123.
- Dale, S. B. and Krueger, A. B. (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *Quarterly Journal of Economics* **117**, 1491–1527.
- Ehrenberg, R. G. (ed.) (2004). *Governing Academia*. Ithaca, NY: Cornell University Press.
- Ehrenberg, R. G. (ed.) (2007). *What's Happening to Public Higher Education?* Baltimore, MD: Johns Hopkins University Press.
- Eide, E. R., Brewer, D. J., and Ehrenberg, R. G. (1998). Does it pay to attend an elite private college? Evidence on the effects of undergraduate college quality on graduate school attendance. *Economics of Education Review* **17**, 371–376.
- Groen, J. A. and White, M. J. (2004). In-state versus out-of-state students: The divergence of interest between public universities and state governments. *Journal of Public Economics* **88**, 1793–1814.

- Long, B. T. (2004). How do financial aid policies affect colleges? The institutional impact of the Georgia HOPE scholarship. *Journal of Human Resources* **39**, 1045–1066.
- Rizzo, M. J. and Ehrenberg, R. G. (2004). Resident and nonresident tuition and enrollment at flagship state universities. In Hoxby, C. (ed.) *College Choice*, pp 303–349. Chicago, IL: University of Chicago Press.
- Singell, L. D. and Stone, J. A. (2007). For whom the pell tolls: The response of university tuition to federal grants-in-aid. *Economics of Education Review* **26**, 285–295.

Relevant Websites

<http://www.bc.edu> – Center for International Higher Education at Boston College. A leading source for information on international higher education.

<http://www.ilr.cornell.edu> – Cornell Higher Education Research Institute. A large number of conference and working papers dealing with the economics of higher education are available on this site.

<http://www.grapevine.ilsu.edu> – Grapevine/Center for the Study of Education Policy at Illinois State University. An annual compilation of state tax appropriations for higher education institutions in the United States.

<http://www.sheeo.org> – State Higher Education Executive Officers (SHEEO). The reader can download its annual “Survey of State Tuition, Fees and Financial Assistance Policies for Public Colleges and Universities.

<http://www.collegeboard.com> – The College Board web site. The reader can download the most recent edition of the College Board’s publications *Trends in College Pricing* and *Trends in Student Aid* from this site.

<http://webcaspar.nsf.gov> – WebCaspar. A searchable institutional data base that provides annual data on tuition and fees (and many other variables) for all higher education institutions in the United States.

ECONOMICS OF EDUCATION – EXTERNAL BENEFITS OF SCHOOLING

Contents

Education and Civic Engagement

Education and Crime

Education and Economic Growth

Parental Socioeconomic Status, Child Health, and Human Capital

The External Benefits of Education

Education and Civic Engagement

T S Dee, Swarthmore College, Swarthmore, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

One of the most fundamental and recurring questions in economics involves how we should understand the proper division of human activity across the public and the private sectors. The field of education is certainly no exception. Public-sector involvement in virtually all forms of formal educational activity is both diverse and extensive. Correspondingly, the intellectual justifications for education-related public policies and institutions reflect a broad variety of distinct normative concerns. However, a conjecture that is central to claims about the importance of government involvement in education is the substantive role that educational investments are thought to play in the formation of an engaged and enlightened citizenry. For example, the claim that widespread education is critical for a stable, functioning democracy was a key motivation for the dramatic growth in access to formal schooling in the United States during the nineteenth century. More recently, the putative civic benefits of investments in education have motivated arguments for private-school vouchers as well as for the Great Society investments in higher education.

From an economic perspective, the basic argument is that individual investments in education may generate a positive externality through improvements in civic engagement that effectively benefit the entire society. This article provides a critical overview of the available evidence that has examined the effects of schooling on subsequent measures of civic engagement. The available evidence is largely, although not exclusively, based on data from the United States. However, it is also important to note that, while the existence of a civic externality from schooling can motivate government intervention in education, it does not have unambiguously clear implications for the exact form of that intervention. The existence of

civic externalities from schooling also does not speak directly to the important question of how public and private schools compare with regard to promoting subsequent civic engagement. This article also discusses the more limited and recent evidence that touches on these issues and concludes with some promising directions for future research.

The Civic Returns to Educational Attainment

Civic engagement refers to a diverse set of behaviors, attitudes, and knowledge that constitute effective citizenship. For example, one particularly prominent dimension of civic engagement involves the allocation of time: participation in voting, volunteering, membership in civic organizations, and engagement with elected representatives. A second key component of civic engagement is a belief in the validity and desirability of democratic, pluralistic institutions, and related values such as tolerance and respect. A third dimension involves having both the requisite cognitive skills and an awareness of current events that make an informed deliberation on complex social and technological issues possible.

Investments in education are widely thought to be a critical determinant of all dimensions of civic engagement, and, by implication, of maintaining a stable, well-functioning democracy. In particular, increases in educational attainment are thought to promote civic engagement both by inculcating students with a sense of civic responsibility and by providing them with the skills and knowledge that allow them to make informed decisions. However, basic economic theory suggests a number of other, potentially contrary

ways in which additional schooling may shape civic engagement. For example, the well-documented effect of schooling on wages implies that educational attainment increases the opportunity cost of time. This increase should reduce the willingness to invest time in civic activities such as voting and volunteering. This effect may be particularly relevant for volunteering, which, relative to voting, implies a more intensive and sustained commitment of time. Additional schooling could also reduce voter participation if it makes voters more aware of the paradox of voting. More specifically, additional school could encourage voters to view voting as a largely expressive and meaningless act rather than an instrumental one with a nontrivial likelihood of directly influencing an election's outcome. Finally, it should be noted that the effects of additional schooling on civic responsibility may work through the changes in peer-groups and social norms rather than through the effects of schooling *per se*.

A large number of older, empirical studies, mostly from the field of political science, have examined the effects of educational attainment on dimensions of civic engagement in the United States (Converse, 1972; Wolfinger and Rosenstone, 1980; Nie *et al.*, 1996; Putnam, 2001). These studies generally find that, conditional on several observable traits, higher levels of educational attainment are associated with substantial increases in multiple dimensions of adult civic engagement. In fact, several authors have stressed the central role that education seems to have among all determinants of civic engagement (Converse, 1972; Putnam, 2001). The apparent gains from increases in educational attainment appear to exist at both the secondary and the postsecondary levels. Furthermore, the strong partial correlations between educational attainment and civic engagement also appear to exist internationally (e.g., Franklin, 1996).

However, several recent studies on this topic have been motivated by the concern that the partial correlations between levels of schooling and civic engagement may be misleading. The basic concern is these correlations could, quite plausibly and to an unknown degree, reflect inherently unobservable individual, family, and community traits that influence both educational attainment and subsequent civic engagement. For example, families and communities that foster and encourage educational success may also be more likely to promote the development of civic virtue. One recent study addressed this potential source of bias by conducting comparisons within 85 pairs of monozygotic twins from New Zealand and found that an additional year of schooling appeared to lower the likelihood of volunteering by roughly 12.5% (Gibson, 2001). This contrarian finding is consistent with the effects of schooling on the opportunity cost of time.

Two other recent studies have examined the causal effects of additional schooling on more diverse measures

of civic engagement by exploiting the credibly independent variation in educational attainment generated by teen exposure to child labor and compulsory schooling laws (Dee, 2004; Milligan *et al.*, 2004). Using different data sets of respondents from the United States, these studies find that increases in educational attainment do appear to generate substantial increases in the likelihood of voting. Furthermore, these studies present evidence that additional schooling increases the quality of civic participation both in the United States and in the United Kingdom as measured by awareness of public affairs and support for free speech (Dee, 2004; Milligan *et al.*, 2004).

Additional evidence has focused on the civic returns from additional schooling at the postsecondary margin in the United States, using geographic access to 2-year and community colleges as a natural experiment (Dee, 2004). The evidence from this approach indicates that college attendance leads to substantial increases in rates of subsequent voter participation. However, the estimated effects of college attendance on volunteering in adulthood were negative and statistically imprecise. Furthermore, this study did not address the effects of college attendance on other dimensions of civic engagement (e.g., civic knowledge and values) because of the apparent lack of a data set that simultaneously accommodates the use of a credible natural experiment and includes the requisite outcome variables.

Comparisons of Public and Private Schools

The available empirical evidence indicates that increases in educational attainment lead to substantial increases in several measures of civic engagement in adulthood, with the possible and notable exception of volunteering. In other words, the widely held view that investments in schooling generate some meaningful and positive civic externalities appears to be valid. These results imply that schools and classrooms are an important and highly effective setting for the development of civic engagement. However, the effects of additional schooling on civic engagement are often used to motivate continued or additional financial support for publicly managed elementary and secondary schools. From an economic standpoint, this policy interpretation is not clearly justified. Narrowly interpreted, the existence of these civic externalities argues only for the existence of corrective financial subsidies that encourage individuals to acquire additional schooling. More specifically, a subsidy equal to the value society places on the civic engagement created by additional schooling could encourage individuals to choose the socially optimal level of schooling (i.e., a level that reflects both the internal and external benefits of schooling). In fact, the seminal argument for private-school

vouchers appealed to the existence of civic externalities as an important justification (Friedman, 1955).

A more compelling economic argument for an extensive and publicly managed system of elementary and secondary schools, similar to that observed in the United States, would turn on evidence that the public sector is more effective than the private sector at promoting the development of civic engagement. There are a number of reasons to suspect that, even with regulation, private schools will not promote a socially desirable level of civic engagement. For example, because private schools are largely accountable only to the parents of their students, they may place more emphasis on skills and knowledge with clear individual benefits and less emphasis on outcomes with larger external benefits. Furthermore, the sectarian and, sometimes, segregated nature of many private schools may also harm key dimensions of civic engagement. However, there are at least two ways in which private schools could be more effective at promoting the development of civic engagement. First, private schools may be more successful at developing civically relevant cognitive skills. Second, private schools may be more effective at increasing student exposure to social capital (i.e., shared norms and trust), which, in turn, promotes the development of civic engagement.

Despite the importance of these issues, the comparative effects of public and private schools in promoting civic engagement have been the subject of surprisingly little empirical study. However, the limited evidence that is available uniformly suggests that private schools are actually more effective than public schools. For example, two recent studies based on nationally representative survey data found that attending a Catholic high school instead of a public high school was associated with improvements in civic participation, knowledge, and attitudes (Greene, 1998; Campbell, 2001). Another study based on data from college students in Texas also found a correlation between measures of political tolerance and private-school attendance (Wolf *et al.*, 2001). However, as with the strong correlations between increased schooling and civic engagement, this evidence could merely reflect selection biases. In other words, the students who attend private, instead of public, schools may experience unobserved family and community environmental traits that promote higher civic engagement in adulthood.

A recent study based on detailed data from the High School and Beyond longitudinal study examined the effects of attending a Catholic high school on several forms of civic participation, using specifications that addressed the potential influence of selection bias (Dee, 2005). In particular, this study used residence in a county with a Catholic high school as a kind of natural experiment (i.e., an instrumental variable) for attending a Catholic high school. The results of this approach suggested that attending a Catholic high school significantly increased the likelihood of voting in

adulthood. However, the estimated effects of Catholic school attendance on volunteering were smaller and statistically insignificant.

A critical assumption implicit in this analysis was that the geographic distribution of Catholic schools is unrelated to the unobserved propensities for civic engagement in adulthood. The fact that the presence of Catholic high schools was driven in large part by patterns of immigration in the late nineteenth and early twentieth centuries is consistent with this approach. Furthermore, a confounding bias would occur if communities with Catholic high schools also tended to have higher levels of civic engagement. However, Dee (2005) found that counties with Catholic high schools actually have lower voter-turnout rates. Nonetheless, Dee (2005) also estimated the degree of potential bias directly by assessing how the instrumental variable influenced the adult civic outcomes among students for whom the presence of Catholic high schools was largely irrelevant (e.g., those attending public school for eighth grade). The results provided no evidence of statistically significant bias from this approach. However, at least two other caveats with regard to this study are relevant. First, due to data constraints, this study did not provide any evidence on how Catholic schools influenced other key civic outcomes, such as tolerance and support for democratic values. Second, it is not clear whether the apparent civic advantage of Catholic high schools generalizes to other types of private schools.

Nonetheless, this evidence indicates that public schools are not as effective as they could be with regard to fostering voter participation in adulthood. Interestingly, the relative ineffectiveness of public schools does not appear to be due solely to being less effective at promoting educational attainment. More specifically, the estimated effectiveness of Catholic schools in promoting high school completion combined with the estimated effects of completing high school on civic outcomes explains less than half of the apparent Catholic school advantage (Dee, 2005). These effect-size comparisons imply that other factors, such as social capital within school communities and instructional practices, play an important role in the relative success of Catholic schools in increasing subsequent voter turnout.

This discussion has focused on how public and private schools promote the civic development of the students they serve. However, it should also be noted that a provocative recent study has argued that public schools play a critical role in sustaining and promoting civic engagement among adults in their communities (Fischel, 2006). More specifically, this study argues that local public schools provide an important complement to local social capital by facilitating connections among parents that, in turn, lower the transaction costs involved in a variety of civic activities. Some descriptive empirical evidence, such as the correlation between the levels of social capital and the presence of smaller school districts, is consistent with this

claim (Fischel, 2006). This study also suggests that the extensive use of private-school vouchers would harm the production of social capital and civic engagement by dispersing students from their communities and increasing the social isolation of neighboring parents.

Summary and Future Directions

Investments in schooling appear to lead to substantial increases in voter participation and civic awareness, although not necessarily in volunteering. As it is not clear how to monetize improvements in civic engagement, the exact implications of these results for the social rate of return to education are not clear. Nonetheless, these results confirm the conventional wisdom that what occurs within schools and classrooms plays a vitally important role in the civic development of students. Furthermore, the existence of some civic externalities provides an important motivation for policies designed to promote high school completion and to expand access to higher education.

However, the limited evidence that is available also suggests that public schools are inferior to Catholic schools with respect to promoting adult voter participation and, possibly, other dimensions of civic engagement. These comparative results imply that public schools are, on average, underperforming with regard to one of their core goals. Therefore, the identification of educational policies and practices that are effective at promoting civic engagement will undoubtedly address an important policy need. One prominent way in which schools have been increasingly embracing their civic missions is through the encouragement of volunteer activity among students. A key argument made in support of promoting or requiring community service among students is that early engagement with volunteering may create lifelong volunteers. However, there is little in the way of credible evidence that these efforts have actually been effective in promoting sustained civic engagement. In particular, the evidence on college-level initiatives to promote volunteering and service learning is largely descriptive (Dee, 2008). Furthermore, a recent study examined the effect of Maryland's mandatory community-service requirement for high school graduates and found that it influenced only the timing, and not the overall level, of volunteering (Helms, 2007).

See also: Human Capital; The Economics of Catholic Schools; The Efficacy of Educational Vouchers; The External Benefits of Education.

Bibliography

Campbell, D. E. (2001). Making democratic education work. In Peterson, P. E. and Campbell, D. E. (eds.) *Charters, Vouchers, and Public Education*, pp 241–267. Washington, DC: Brookings Institution Press.

Converse, P. E. (1972). Change in the American Electorate. In Campbell, A. and Converse, P. E. (eds.) *The Human Meaning of Social Change*, pp 263–331. New York: Russell Sage.

Dee, T. S. (2004). Are there civic returns to education? *Journal of Public Economics* **88**(9), 1697–1720.

Dee, T. S. (2005). The effects of Catholic schooling on civic participation. *International Tax and Public Finance* **12**(5), 605–625.

Dee, T. S. (2008). Assessing the college contribution to civic engagement. In McPherson, M. S. and Schapiro, M. O. (eds.) *Succeeding in College: What It Means and How to Make It Happen*. New York: College Board.

Fischel, W. A. (2006). Why voters veto vouchers: Public schools and community-specific social capital. *Economics of Governance* **7**, 109–132.

Franklin, M. N. (1996). Electoral participation. In LeDuc, L., Niemi, R. G., and Norris, P. (eds.) *Comparing Democracies: Elections and Voting in Global Perspectives*, pp 216–235. London: Sage.

Friedman, M. (1955). The role of government in education. In Solo, R. A. (ed.) *Economics and the Public Interest*, pp 127–134. New Brunswick, NJ: Rutgers University Press.

Gibson, J. (2001). Unobservable family effects and the apparent external benefits of education. *Economics of Education Review* **20**(3), 225–233.

Greene, J. P. (1998). Civic values in public and private schools. In Petersen, P. E. and Hassel, B. C. (eds.) *Learning from School Choice*, pp 83–106. Washington, DC: Brookings Institution Press.

Helms, S. E. (2007). Involuntary Volunteering: The Impact of Mandated Service in Public Schools, mimeo. University of Alabama at Birmingham.

Milligan, K., Moretti, E., and Oreopolous, P. (2004). Does education improve citizenship? Evidence from the United States and the United Kingdom. *Journal of Public Economics* **88**(9), 1667–1695.

Nie, N. H., Junn, J., and Stehlik-Barry, K. (1996). *Education and Democratic Citizenship in America*. Chicago, IL: University of Chicago Press.

Putnam, R. D. (2001). Tuning in, tuning out: The strange disappearance of social capital in America. In Niemi, R. G. and Weisberg, H. F. (eds.) *Controversies in Voting Behavior*, pp 38–68. Washington, DC: CQ Press.

Wolf, P. J., Greene, J. P., Kleitz, B., and Thalhammer, K. (2001). Private schooling and political tolerance. In Peterson, P. E. and Campbell, D. E. (eds.) *Charters, Vouchers, and Public Education*, pp 268–289. Washington, DC: Brookings Institution Press.

Wolfinger, R. E. and Rosenstone, S. J. (1980). *Who Votes?* New Haven, CT: Yale University Press.

Further Reading

Courant, P. N., McPherson, M. S., and Resch, A. M. (2006). The public role in higher education. *National Tax Journal* **59**(2), 291–318.

Glaeser, E. L., Ponzetto, G., and Shleifer, A. (2007). Why does democracy need education? *Journal of Economic Growth* **12**(2), 77–99.

Haveman, R. H. and Wolfe, B. L. (1984). Schooling and economic well-being: The role of nonmarket effects. *Journal of Human Resources* **19**(3), 377–407.

Lange, F. and Topel, R. (2006). The social value of education. In Hanushek, E. and Welch, F. (eds.) *Handbook of the Economics of Education*, vol. 1, pp 459–509. Amsterdam: North-Holland/Elsevier.

Markus, G. B., Howard, J. P. F., and King, D. C. (1993). Integrating community service and classroom instruction enhances learning: Results from an experiment. *Education Evaluation and Policy Analysis* **15**, 410–419.

Poterba, J. M. (1996). Government intervention in the markets for education and health care: How and why? In Fuchs, V. R. (ed.) *Individual and Social Responsibility: Child Care, Education, Medical Care and Long-Term Care in America*, pp 277–308. Chicago, IL: University of Chicago Press.

Taylor, L. L. (1999). Government's role in primary and secondary education. *Federal Reserve Bank of Dallas Economic Review* **1999**, 15–24.

Wolfe, B. L. (1994). External benefits of education. In Husen, T. and Postlethwaite, T. N. (eds.) *International Encyclopedia of Education*, pp 2208–2212. Oxford: Pergamon Press.

Education and Crime

L Lochner, University of Western Ontario, London, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Externality – An effect on others not considered by the individual decisionmaker.

General educational development (GED) – A test that may be taken in the U.S. by high-school dropouts that is meant to serve as an equivalent to high-school completion.

Human capital – An individual's skills or productivity as acquired through schooling, training, prior work, or other types of investment activities.

Social benefits – Benefits to society as a whole, including the individual decisionmaker and others.

Introduction

This article analyzes the relationship between education and crime. The article describes a number of important factors that may drive this correlation, and reviews the current empirical literature on the interrelationship between education and crime. The article concludes with a discussion of the social savings from crime reduction associated with increasing high-school graduation rates.

The Economics of Education and Crime

This section discusses a number of factors that may contribute to a correlation between education and crime. Many of these points are more formally analyzed in Lochner (2004), who takes a human capital approach to analyzing the relationship between education and crime. (Lochner also argues that the age–crime profile may be partially explained by the accumulation of human capital over the lifecycle.) This framework assumes that education (as well as job training) develops formal labor-market skills, which raises the opportunity costs of crime commission. Education may also develop criminal skills, although this is only likely to be important for certain white-collar crimes. Alternatively, education may socialize individuals such that they prefer not to engage in crime.

The Effects of Education on Subsequent Crime Among Adults

There is a relatively large literature linking wages and unemployment rates to criminal behavior. Recent studies

conclude that crime is increasing in local unemployment rates and decreasing in wage rates (e.g., Raphael and Winter-Ebmer, 2001; Gould *et al.*, 2002; Machin and Meghir, 2004). To the extent that education increases wage rates (and reduces the likelihood of unemployment), it increases the opportunity costs of crime and will tend to reduce postschool criminal activity. Higher wages raise the opportunity costs of crime in two distinct ways. First, since crime may require time to commit, that time cannot be used for other productive purposes like work. Here, it is useful to think of all of the time involved in planning a crime, locating a target, and, potentially, evading detection and arrest. Second, each crime committed entails an expected period of incarceration, which is more costly for individuals with better labor-market opportunities and wages.

On the one hand, property crimes such as burglary, auto theft, and drug dealing can involve significant planning or time spent on the actual activity itself. On the other hand, violent crimes such as assault would appear to require less time for planning and execution but are associated with higher expected probabilities of arrest, conviction, and incarceration as well as longer sentence lengths conditional on incarceration. For example, Lochner (2004) calculates that for each assault, the perpetrator can expect to spend 63 days incarcerated; however, the expected incarceration period for a burglary is only 13 days. These time costs would appear to exceed the direct time costs associated with committing most crimes. Thus, changes in wages or unemployment rates could have greater effects on violent crimes than on property crimes. (The estimated effects of low skill wages on violent crime are larger than on property crime for some specifications in Gould *et al.* (2002).)

Education may also affect the rewards from crime. This is most likely to be true for white-collar crimes such as fraud, forgery, and embezzlement. Education may actually increase these types of crime if it increases the rewards from crime more than it increases legitimate wages. Lochner (2004) finds some evidence that white-collar crime rates are increasing in average education levels as discussed below. To the extent that schools socialize students to become better citizens and to treat others better, education may also reduce the psychic returns to crime causing individuals to forego lucrative criminal opportunities.

Education may also teach individuals to be more patient. This will discourage crime, since forward-looking individuals place greater weight on any expected punishment associated with their criminal activities. Education may also affect preferences toward risk. To the extent that

schooling makes individuals more risk averse, it will tend to discourage crime.

Contemporaneous Crime and Education Decisions Among Youth

Youth crime will tend to be decreasing in both contemporaneous and future wage rates. Higher contemporaneous wages increase the direct opportunity cost of committing crime, while higher future wages increase the costs associated with potential incarceration. As education increases future wage rates, youth who are enrolled in school will be less likely to engage in crime than otherwise similar youth who are not in school. (Prison is costly for those who stand to earn a lot in the future — either legitimately or illegitimately. The extent to which education increases post-school criminal returns also tends to discourage youth crime. The potential to earn high rewards from white-collar crime after school may discourage youth criminal activity to avoid imprisonment.)

Schooling is not exogenously determined. Youth will choose to enrol in school if they receive a net benefit from doing so; otherwise, they will not. Not only does an increase in wages for high-school graduates or college attendees reduce crime for all youth who would have attained these schooling levels in the first place, but it also causes more youth to finish high school and attend college, lowering their lifetime criminal activity as well.

Since the benefits from schooling through higher lifetime earnings are delayed, youth who are more patient are more likely to attend school. More patient youth are also less likely to engage in crime, since the punishments tend to be delayed. Thus, differences in patience across the population will tend to lead to a negative relationship between education and crime. Population heterogeneity in preferences toward risk may also lead to a correlation between education and crime. If the rewards to school are risky as some economists suggest, more risk-averse youth will tend to quit school at earlier ages. Risk-averse youth are also more likely to engage in crime regardless of their schooling, generating a negative correlation between crime and education.

Youth who plan to commit more crime as adults will tend to benefit less from each year of school for two reasons. First, those who allocate more time to crime and less to work will benefit less from the increased wages associated with schooling (assuming schooling raises legitimate wages more than the returns from crime). Second, those who commit more crime can expect to spend more time in prison, a place which offers little reward to additional schooling. As a result, education decisions may depend on factors that affect the returns to or costs of crime. For example, a higher return to crime or more favorable treatment to prisoners will encourage criminal activity and, therefore, lower the returns from schooling.

The effects of a reduction in arrest or incarceration rates on schooling are less clear. On the one hand, a reduction in the probability of incarceration directly increases the value of schooling holding criminal activity constant. On the other hand, a lower incarceration probability encourages crime, which reduces the value of education. In general, criminal abilities/opportunities and law enforcement policies should affect the schooling decisions of marginal youth.

The role of peers or social networks may also be important determinants of crime and educational attainment. Youth who drop out of school may be influenced by a more negative set of peers, which may exacerbate any tendencies to engage in crime. Similarly, youth who join gangs or who otherwise engage in crime may be encouraged to leave school by their peers. Crime, or arrest and incarceration, may also come with a stigma which makes school more difficult, an issue we discuss briefly in the following section.

All of these factors suggest that youth will tend to make an early choice between little education and a life of street crime or a good education and a largely crime-free life. Education does not pay if one intends to participate in gang life, selling drugs, and engaging in other criminal activity. The fact that such a lifestyle is also likely to come with substantial periods of incarceration further reduces any potential benefits from schooling. At the same time, youth in neighborhoods with poor schools or that offer few jobs even to those who do finish high school are likely to find the street life relatively attractive. Social networks and peers are likely to strengthen these forces.

Empirical Evidence on Crime and Education

The empirical literature on education and crime has focused almost exclusively on the effects of educational attainment on postschool criminal activity; however, a few studies have attempted to estimate the effects of school enrolment on contemporaneous crime. Given the simultaneity of enrolment and crime choices (do youth drop out of school because they want to sell drugs all day or do they sell drugs because they dropped out of school?), this is a particularly daunting task and one that is often ill-defined. Recently, a few studies have attempted to estimate the effects of youth arrest and incarceration on educational outcomes. Studies that attempt to estimate the effects of law enforcement policies or criminal opportunities on educational decisions are virtually nonexistent. We briefly summarize the current state of empirical evidence on these issues.

The Effects of Education on Crime

We have discussed four primary reasons schooling might affect crime: (1) education raises wage rates, which raises

the opportunity costs of crime; (2) education may directly affect the financial or psychic rewards from crime; (3) education may alter preferences for risk taking or patience; and (4) schooling may affect the social networks or peers of individuals. For most crimes (except, possibly, white-collar crimes), one would expect these forces to induce a negative effect of schooling on crime.

Empirically, there is a strong negative correlation between educational attainment and various measures of crime. Freeman (1996) points out that more than two-thirds of all incarcerated men in 1993 had not graduated from high school. In the 1980 wave of the National Longitudinal Survey of Youth (NLSY), 34% of all men aged 20–23 with 10 or 11 years of completed schooling self-reported earning some income from crime, compared with 24% of those with a high-school degree, and only 17% of those with more than 12 years of school (Lochner, 2004). Similar differences are evident for other self-reported measures of both violent and property crime.

Early studies of the relationship between education and crime focused on their correlation conditional on measured individual and family characteristics using multivariate regression methods. For example, Witte and Tauchen (1994) find no significant relationship between educational attainment and crime after controlling for a number of individual characteristics. Grogger (1998) estimates a significant negative effect of wages on crime, but he finds no relationship between years of completed schooling and crime after controlling for individual wage rates. Of course, increased wages and earnings are important consequences of schooling. Thus, this study suggests that education may indirectly reduce crime through increased wage rates. (Gottfredson (1985), Farrington *et al.* (1986), and Witte and Tauchen (1994) explore the link between time spent in school and contemporaneous crime, concluding that time spent in school significantly reduces criminal activity. This type of analysis is particularly difficult to interpret given the simultaneous nature of the crime and schooling choices.)

These earlier studies must be interpreted with caution. A negative correlation between education and crime, even after controlling for measured family background and neighborhood characteristics, does not necessarily imply that education reduces crime. Standard regression studies are unlikely to estimate the causal effect of education on crime (i.e., the effect increasing someone's schooling on his criminal activity) for a number of reasons. First, unobserved individual characteristics such as patience or risk aversion are likely to directly affect both schooling and criminal decisions. Individuals who choose more schooling (even after conditioning on observable characteristics) might also choose less crime regardless of their education level, in which case regression-based estimates do not identify the causal effect of schooling on crime. Second, using variation in crime and education across states or

local communities may also produce biased estimates. Governments may face a choice between funding law enforcement and good public schools, which would tend to produce a spurious positive correlation between education and crime. Alternatively, unobserved characteristics about communities or their residents may directly affect the costs or benefits of both education and crime. For example, communities with few job opportunities that reward schooling may also be faced with severe gang problems. While it is often possible to account for permanent unobserved differences across communities by examining the relationship between changes in schooling and crime over time, such an approach cannot account for the effects of changing unobserved community characteristics. Third, reverse causality is another important concern, in which case traditional regression estimates may be confounded by the effect of criminal activity on schooling. Individuals who plan to heavily engage in crime (e.g., because they are particularly good at it, enjoy it, or live in areas with plenty of illicit opportunities) are likely to choose to leave school at a young age. Arrests or incarceration associated with juvenile crime may also cause some youth to drop out of school early. Finally, it is difficult to measure crime itself; instead, researchers are often forced to use measures of arrest or incarceration rather than actual crimes committed. It is possible that education reduces the probability of arrest and incarceration or the sentence lengths administered by judges. Estimates based on measures of arrest or incarceration will incorporate these effects in addition to any effects of education on actual crime.

Recently, economists have attempted to address these difficult issues through the use of instrumental variable (IV) estimation methods. In the context of estimating the effect of educational attainment on crime, an instrument is valid if it induces variation in schooling but is uncorrelated with other factors that directly affect criminal behavior (e.g., individual preferences or abilities). Intuitively, this approach exploits differences in educational attainment across individuals that arise in response to factors that have no direct effect on criminal decisions. An ideal instrument would randomly assign some youth to drop out of high school and others to finish high-school. Then, comparing the differences in crime rates across these groups would identify the causal effect of high-school completion on crime. In practice, we typically do not observe such perfect experiments, but researchers can sometimes come close.

Lochner and Moretti (2004) use changes in state-specific compulsory schooling laws over time as an IV for completed schooling to estimate the effects of education on arrest rates and the probability of incarceration among adult men. Intuitively, they measure the extent to which an increase in a state's compulsory schooling age leads to an immediate increase in educational attainment and reduction in subsequent crime rates for affected cohorts. This identifies the causal effect of schooling on

crime as long as the changes in compulsory schooling laws are not related to changes in the underlying propensity to commit crime. Lochner and Moretti's (2004) analysis suggests that changes in compulsory schooling laws are exogenous and not related to prior trends in schooling or state expenditures on law enforcement; so it appears to be a valid instrument.

Lochner and Moretti (2004) first use individual-level data on incarceration and schooling from the 1960, 1970, and 1980 US censuses to estimate the effects of educational attainment on the probability of imprisonment separately for black and white men. Their estimates control for age of the respondent (3 year age categories), state of birth, state of residence, cohort of birth, and state-specific year effects. Most importantly, controlling for state-specific year effects allows for the possibility that different states may have different time trends for law enforcement policies or may simply exhibit different trends in aggregate criminal activity. Identification comes from the fact that in any given state and year, different age cohorts will have faced different compulsory schooling laws during their high-school years, causing them to acquire different levels of schooling and to commit crime at different rates. Interestingly, both ordinary least squares (OLS) and IV estimates are very similar and suggest that, on average, an extra year of school reduces the probability of imprisonment by slightly more than 0.1 percentage point for whites and by about 0.4 percentage points for blacks. Given the probability of incarceration for male whites without a high-school degree averaged 0.83% across all three censuses and the incarceration rate for male black dropouts was 3.6%, these effects are sizable. OLS results suggest that completion of the 12th grade causes the greatest drop in incarceration, while their is little effect of schooling beyond high school.

In their analysis of male arrest rates, Lochner and Moretti (2004) use state-level arrest rates by criminal offense and age (5 year age categories beginning at ages 20–24 through 55–59) from the Federal Bureau of Investigation (FBI)'s Uniform Crime Reports (UCR) for 1960, 1970, 1980, and 1990. These data are linked to 1960–1990 US Census data on educational attainment and race to estimate regressions of the form:

$$\ln(A_{cast}) = \beta E_{ast} + \gamma B_{ast} + d_{st} + d_{sc} + d_{sa} + d_{ct} + d_{at} + d_{ac} + \varepsilon_{cast} \quad [1]$$

where $\ln(A_{cast})$ is the logarithm of the male arrest rate for crime c , age group a , in state s in year t (from UCR); E_{ast} is either average education or the high-school graduation rate for males in age group a in state s at time t (from Census); B_{ast} is the percent of males that are black in age group a in state s at time t (from Census). They analyze arrest rates for the following crimes: murder, rape, assault, robbery, burglary, larceny, auto theft, and arson. In using log arrest rates, the effect of education on arrest rates is assumed to be the same in percentage terms for all types of crime.

The d 's in eqn [1] represent indicator variables that account for unobserved differences across states, years, cohorts, and criminal offense types. The term d_{st} allows for state-specific time effects, which is more general than including time-varying observable state-level variables reflecting differences in public spending, economic conditions, or law enforcement. The inclusion of d_{sc} allows the distribution of crimes or arrests across states to differ. Some states may focus arrests more heavily on one type of crime, while others focus on other types. Furthermore, the age distribution of arrestees need not be the same across states – some age groups may be more prone to commit crimes in some states or the arrest policy with respect to age may differ across states. The term d_{sa} absorbs long-run differences in age–arrest patterns across states. Crime-specific and age-specific trends in arrest common to all states are accounted for by d_{ct} and d_{at} , respectively. Finally, d_{ac} accounts for long-term differences in age–crime profiles across different types of criminal offenses.

Using OLS, Lochner and Moretti (2004) estimate that a 1-year increase in average education levels in a state reduces state-level arrest rates by 11%. IV estimates suggest slightly larger effects, although they are not statistically different. These estimated effects are very similar to the predicted effects derived from multiplying the estimated increase in wages associated with an additional year of school by the estimated effects of higher wage rates on crime (Gould *et al.*, 2002). This suggests that much of the effect of schooling on crime may come through increased wage rates and opportunity costs. Using OLS, Lochner and Moretti (2004) also estimate separate effects of education for different types of crime. These results suggest similar effects across the broad categories of violent (murder, rape, robbery, and assault) and property (burglary, larceny, motor vehicle theft, and arson) crime — a 1-year increase in average years of schooling reduces both property and violent crime by about 11–12%. However, the effects vary considerably within these categories. A 1-year increase in average years of schooling reduces murder and assault by almost 30%, motor vehicle theft by 20%, arson by 13%, and burglary and larceny by about 6%. Estimated effects on robbery are negligible, while those for rape are significantly positive. This final result is surprising and not easily explained by standard economic models of crime. (However, it is consistent with evidence in Gould *et al.* (2002), which suggests that local wage rates are positively correlated with local rape crime rates.)

Lochner (2004) follows a similar approach to estimate the effects of average schooling levels on arrest rates for white-collar crime (forgery and counterfeiting, fraud, and embezzlement) using UCR and Census data from 1960, 1970, and 1980. In contrast to the results for violent and property crimes, he estimates a positive, though statistically insignificant, effect of schooling on white-collar arrest rates.

One obvious concern with these studies is their use of arrest and incarceration as measures of crime. It is possible that education improves the chances that someone evades arrest or conviction or that judges tend to give more educated defendants lighter prison sentences. While there is little direct evidence on these issues, Mustard (2001) finds negligible effects of defendant education levels on the sentence lengths they receive. Furthermore, results using self-reported measures of criminal activity in the National Longitudinal Survey of Youth (NLSY) support the case that education reduces actual violent and property crime and not just the probability of arrest or incarceration conditional on crime (Lochner, 2004, Lochner and Moretti, 2004).

A growing body of evidence suggests that early childhood interventions can also substantially reduce adult crime rates. Most famously, the High/Scope Perry Preschool Program for disadvantaged minority children measured lifetime arrests for randomly assigned participants and nonparticipants. While 55% of all nonparticipants were arrested five or more times through age 40, only 36% of the preschool participants had been arrested that often (Schweinhart *et al.*, 2005). The Syracuse Family Development Program also produced large reductions in delinquency (Lally *et al.*, 1988). These findings lead Donohue and Siegelman (1998) to conclude that small, rigorous early intervention programs may pay for themselves through reduced crime rates alone, if they can be targeted to high-crime groups.

A provocative recent study by Jacob and Lefgren (2003) explores the contemporaneous effects of school attendance on juvenile crime rates. To identify these effects, the study uses exogenous variation in teacher in-service days across jurisdictions over time, essentially comparing local juvenile crime rates on days when school is not in session to those when it is in session. Their findings suggest that school attendance reduces contemporaneous juvenile property crime while increasing juvenile violent crime. These results are consistent with an incapacitation effect of school that limits participation in property crime. However, the increased level of interaction among adolescents facilitated through schools may raise the likelihood of violent conflicts after school. It is important to distinguish between the contemporaneous effects of school attendance and crime estimated in this study from the effects of educational attainment on subsequent crime estimated by Lochner and Moretti (2004) or Lochner (2004). There is no logical inconsistency between the findings of these studies.

The Effects of Arrest and Incarceration on Education

Hjalmarsson (2006) empirically examines the effects of juvenile arrests and incarceration (through age 16) on

high-school completion by age 19. Her main specifications control for youth cognitive achievement, juvenile criminal activity, and family background. She also considers additional models that account for state or family fixed effects to account for differences in state-level juvenile law enforcement and education policies as well as differences in family (and, therefore, neighborhood) environments. Her regression-based estimates suggest substantial effects of both arrest and incarceration on subsequent schooling attainment; however, she finds that her estimated effects for arrest may be largely due to unobserved heterogeneity across youth. Her findings for juvenile incarceration are more robust and suggest that youth who become incarcerated, holding their juvenile criminal activity and arrest rates constant, are roughly 25 percentage points less likely to complete high school by age 19 than similar youth who are not arrested.

Hjalmarsson (2006) next explores a number of mechanisms that may generate an effect of incarceration on schooling. First, she finds that the effects are substantially larger for parochial school students than public school students. Second, she finds that incarceration has its greatest effects on high-school graduation when the sentence overlaps with the school year; however, the length of the sentence does not affect the probability of graduation. Finally, she finds that incarceration has substantially larger effects on high-school completion in states that require the justice system to notify schools of an arrest when compared with other states that do not require notification. All of this evidence suggests that teachers and/or administrators may treat students differently if they are known to have been incarcerated. Thus, juvenile incarceration may carry a negative stigma in schools, just as it appears to in the labor market.

Education and Training in Prison

Vast majority of US correctional facilities offer some form of education and training for prisoners, with General Educational Development (GED) preparation courses being the most common. To the extent that prison education programs help build valuable market skills (in the same way traditional schools do), we would expect them to increase postrelease earnings and reduce recidivism. Unfortunately, convincing empirical studies on this topic are scarce, primarily because prisoners who choose to enrol in prison education programs likely differ from those choosing not to enrol. Tyler and Kling (2006) attempt to account for these differences through a rich set of prisoner characteristics (e.g., sentence length, marital status and number of dependents, employment status prior to arrest, offense type, and a measure of cognitive ability), comparing the postrelease earnings of prisoners who received a GED in prison with similar high-school dropout prisoners who did not. They further account for prisoner differences by

controlling for preprison earnings. Their findings suggest that a GED earned in prison offers no postrelease earnings benefit for white offenders, but it does increase the earnings of minority offenders for the first 2 years after release (by about \$800 per year). The earnings benefits for minorities fade quickly after the second year and are no longer statistically significant.

Conclusions

To the extent that education reduces crime, schooling may have important externalities and social benefits that are not taken into account by individuals. When making their schooling decisions, youth may not consider the important negative costs they impose on society if they choose to drop out of high school in favor of a life of crime. Thus, policies to promote schooling may benefit society through reduced crime, in addition to the more obvious gains from increased productivity. Lochner and Moretti (2004) estimate that a 1% increase in high-school graduation rates would save the US economy nearly \$2 billion from reduced costs associated with criminal activity. (All social savings figures are reported in year 2006 dollars.) The social savings per additional male graduate from crime reduction alone amounts to \$1600–2900, or 14–26% of the private return to individuals from increased earnings.

See also: Human Capital; The External Benefits of Education.

Bibliography

- Donohue, J., III and Siegelman, P. (1998). Allocating resources among prisons and social programs in the battle against crime. *Journal of Legal Studies* **27**, 1–44.
- Farrington, D., Gallagher, B., Morley, L., St. Ledger, R., and West, D. (1986). Unemployment, school leaving and crime. *British Journal of Criminology* **26**, 335–356.
- Freeman, R. (1996). Why do so many young american men commit crimes and what might we do about it? *Journal of Economic Perspectives* **10**, 25–42.
- Gottfredson, M. (1985). Youth employment, crime, and schooling. *Developmental Psychology* **21**, 419–432.
- Gould, E., Mustard, D., and Weinberg, B. (2002). Crime rates and local labor market opportunities in the United States: 1977–1997. *Review of Economics and Statistics* **84**, 45–61.
- Grogger, J. (1998). Market wages and youth crime. *Journal of Labor Economics* **16**, 756–791.
- Hjalmarsson, R. (2006). Criminal justice involvement and high school completion. *Working Paper*. University of Maryland.
- Jacob, B. and Lefgren, L. (2003). Are idle hands the devil's workshop? Incapacitation, concentration and juvenile crime. *American Economic Review* **93**, 1560–1577.
- Lally, J., Mangione, P., and Honig, A. (1988). The Syracuse University family development research program: Long-range impact of an early intervention with low-income children and their families. In Powell, D. (ed.) *Parent Education as Early Childhood Intervention: Emerging Directions in Theory, Research, and Practice*, pp 79–104. Norwood: Ablex.
- Lochner, L. (2004). Education, work, and crime: A human capital approach. *International Economic Review* **45**, 811–843.
- Lochner, L. and Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American Economic Review* **94**, 155–189.
- Machin, S. and Meghir, C. (2004). Crime and economic incentives. *Journal of Human Resources* **39**, 958–999.
- Mustard, D. (2001). Racial, ethnic, and gender disparities in sentencing: Evidence from the US federal courts. *Journal of Law and Economics* **44**, 285–314.
- Raphael, S. and Winter-Ebmer, R. (2001). Identifying the Effect of Unemployment on Crime. *Journal of Law and Economics* **44**, 259–283.
- Schweinhart, L. J., Montie, J., Xiang, Z., et al. (2005). *Lifetime Effects: The High/Scope Perry Preschool Study through age 40 (Monographs of the High/Scope Educational Research Foundation, 14)*. Ypsilanti: High/Scope Press.
- Tyler, J. and Kling, J. (2006). Prison-based education and re-entry into the mainstream labor market. *National Bureau of Economic Research Working Paper No. 12114*.
- Witte, A. D. and Tauchen, H. (1994). Work and crime: An exploration using panel data. *National Bureau of Economic Research Working Paper No. 4794*.
- Behrman, J. and Stacey, N. (eds.) (1997). *The Social Benefits of Education*. Ann Arbor, MI: University of Michigan Press.
- Ehrlich, I. (1975). On the relation between education and crime. In Juster, F. T. (ed.) *Education, Income, and Human Behavior*, pp 313–338. New York: McGraw-Hill.
- Gallipoli, G. and Fella, G. (2006). Education and crime over the life cycle. *Working Paper*. University of British Columbia.
- Huang, C., Laing, D., and Wang, P. (2004). Crime and poverty: A search-theoretic approach. *International Economic Review* **45**, 909–938.
- Moretti, E. (2007). Crime and the costs of criminal justice. In Belfield, C. and Levin, H. (eds.) *The Price We Pay: Economic and Social Consequences of Inadequate Education*, pp 142–159. Washington, DC: Brookings Institution Press.
- Steurer, S., Smith, L., and Tracy, A. (2001). Three state recidivism study. *Report for the Office of Correctional Education*. Washington, DC: U.S. Department of Education.
- Usher, D. (1997). Education as a deterrent to crime. *Canadian Journal of Economics* **30**, 367–384.
- Waldfogel, J. (1994). The effect of criminal conviction on income and the trust 'reposed in the workmen'. *Journal of Human Resources* **29**, 62–81.

Further Reading

- http://www.fbi.gov – Federal Bureau of Investigation, Uniform Crime Reports.
- http://www.icpsr.umich.edu – ICPSR, National Archive of Criminal Justice.
- http://www.ojp.usdoj.gov – Office of Justice Programs, Bureau of Justice Statistics.
- http://www.albany.edu – University at Albany, Sourcebook of Criminal Justice Statistics.

Education and Economic Growth

E A Hanushek, Stanford University, Stanford, CA, USA

L Wößmann, University of Munich, Munich, Germany

© 2010 Elsevier Ltd. All rights reserved.

Education has long been viewed as an important determinant of economic well-being. The theoretical growth literature emphasizes at least three mechanisms through which education may affect economic growth. First, education can increase the human capital inherent in the labor force, which increases labor productivity and thus transitional growth toward a higher equilibrium level of output (as in augmented neoclassical growth theories, cf. Mankiw *et al.* (1992)). Second, education can increase the innovative capacity of the economy, and the new knowledge on new technologies, products, and processes promotes growth (as in theories of endogenous growth, cf., e.g., Lucas (1988), Romer (1990), Aghion and Howitt (1998)). Third, education can facilitate the diffusion and transmission of knowledge needed to understand and process new information and to successfully implement new technologies devised by others, which again promotes economic growth (cf., e.g., Nelson and Phelps, 1966; Benhabib and Spiegel, 1994).

Despite these theoretical predictions, the empirical evidence on the impact of education on economic growth has long been mixed. In large part, this seems to reflect measurement problems. Most people would acknowledge that a year of schooling does not produce the same cognitive skills everywhere. They would also agree that families and peers contribute to education. Health and nutrition further impact cognitive skills. Yet, until recently, research on the economic impact of education – largely due to expedience – has almost uniformly ignored these aspects. Recent research shows that ignoring differences in the quality of education significantly distorts the picture of how educational and economic outcomes are related.

Early Studies of Schooling Quantity and Economic Growth

The majority of the macroeconomic literature on economic returns to education employs measures of the quantity of schooling. The most common measure is years of schooling, averaged across the working-age population. (Woessmann (2003) surveys issues of measuring and specifying human capital from early growth accounting to current cross-country growth regressions.) The standard method of estimating the effect of education on economic growth is to estimate cross-country growth regressions where average annual growth in gross domestic product

(GDP) per capita over several decades is expressed as a function of measures of schooling and a set of other variables deemed important for economic growth.

Following the classical contributions by Barro (1991, 1997) and Mankiw *et al.* (1992), a vast early literature of cross-country growth regressions tended to find a significant positive association between quantitative measures of schooling and economic growth. Extensive reviews of the literature are found in Topel (1999), Temple (2001), Krueger and Lindahl (2001), and Sianesi and Van Reenen (2003). To provide an idea of the robustness of the basic association, primary schooling turns out to be the most robust influence factor (after an East Asian dummy) on growth in GDP per capita in 1960–1996 in the extensive robustness analysis of 67 explanatory variables in growth regressions on a sample of 88 countries by Sala-i-Martin *et al.* (2004).

Figure 1 provides a basic representation of the association between years of schooling and economic growth on the most recent version of available data. This research suggests that each year of schooling is associated with long-run growth that is 0.58 percentage points higher.

Yet, questions persist about the interpretation of such relationships. A substantial controversy addresses whether it is the level of years of schooling (as would be predicted by several models of endogenous growth) or the change in years of schooling (as would be predicted by basic neoclassical models) that is the more important driver of economic growth (e.g., Krueger and Lindahl, 2001). It seems beyond the scope of current data to draw strong conclusions about the relative importance of different mechanisms for schooling quantity to affect economic growth. Even so, several recent studies suggest that education is important both as an investment in human capital and in facilitating research and development and the diffusion of technologies, with initial phases of education more important for imitation and higher education for innovation (Vandenbussche *et al.*, 2006).

Two more skeptical studies raise caveats. Bills and Klenow (2000) raise the issue of causality, suggesting that reverse causation running from higher economic growth to additional education may be at least as important as the causal effect of education on growth in the cross-country association. Pritchett (2001, 2006) raises questions about the plausibility of simple growth models with years of schooling and stresses that it is important for economic growth to get other things right as well,

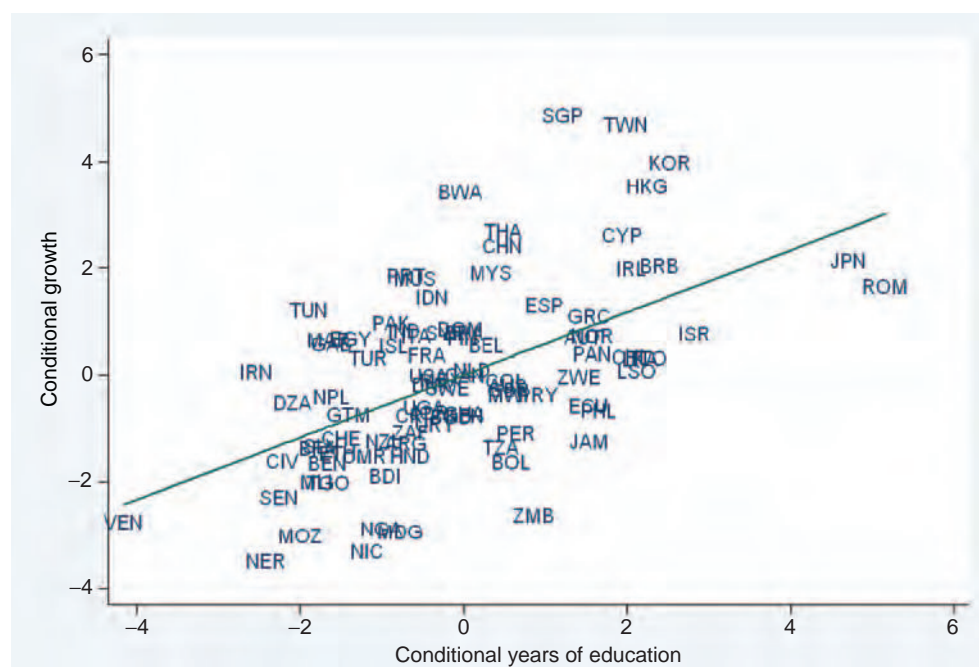


Figure 1 Association between years of schooling and long-run economic growth. Added-variable plot of a regression of the average annual rate of growth (in percent) of real GDP per capita in 1960–2000 on average years of schooling in 1960 and the initial level of real GDP per capita in 1960.

in particular the institutional framework of the economy. Both issues are discussed further below.

However, most importantly, using average years of schooling as an education measure implicitly assumes that a year of schooling delivers the same increase in knowledge and skills regardless of the education system. This measure also assumes that formal schooling is the primary source of education and that variations in the quality of nonschool factors affecting learning have a negligible effect on education outcomes. This neglect of cross-country differences in the quality of education is the major drawback of such a quantitative measure.

Initial Evidence on the Quality of Education and Economic Growth

Quite clearly, the average student in Ghana or Peru does not gain the same amount of knowledge in any year of schooling as the average student in Finland or Korea; however, using measures of years of schooling assumes that they are equivalent. In addition, using years of schooling implicitly assumes that all skills and human capital come from formal schooling. Yet, extensive evidence on knowledge development and cognitive skills indicates that a variety of factors outside of school – family, peers, and others – have a direct and powerful influence. Ignoring these nonschool factors introduces another element of measurement error into the growth analyses.

Since the mid-1960s, international agencies, such as the International Association for the Evaluation of Educational Achievement (IEA) and the Organization for Economic Co-operation and Development (OECD), have conducted many international tests – such as the Trends in International Mathematics and Science Study (TIMSS), the Programme for International Student Assessment (PISA), and their predecessors – of student performance in cognitive skills mathematics, involving science, and other subjects. In order to make performance on a total of 36 international tests from 12 testing occasions comparable, Hanushek and Woessmann (2009) develop a common metric to adjust both the level of test performance and its variation through two data transformations. First, each of the separate international tests is benchmarked to a comparable level by calibrating the US international performance over time to the external standard of the available US longitudinal test (the National Assessment of Educational Progress, NAEP). Second, the dispersion of the tests is standardized by holding the score variance constant within a group of 13 OECD countries with relatively stable secondary school attendance rates over time.

Figure 2 reports average performance at the standardized tests, which serves as a proxy for the quality of education. The variation in the quality of education that exists among OECD countries is already substantial, but the difference from developing countries in the average amount of learning acquired after a given amount of

schooling dwarfs any within-OECD difference. Outside of East Asia, nearly every developing country that participated in one of the tests performed dramatically lower than any OECD country (except Mexico).

Over the past 10 years, growth research demonstrates that considering the quality of education, measured by the cognitive skills learned, dramatically alters the assessment of the role of education in economic development. Using

the data from the international student achievement tests through 1991 to build a measure of educational quality, Hanushek and Kimko (2000) find a statistically and economically significant positive effect of the quality of education on economic growth in 1960–1990 that is far larger than the association between the quantity of schooling and growth. Ignoring quality differences very significantly misses the true importance of education for economic

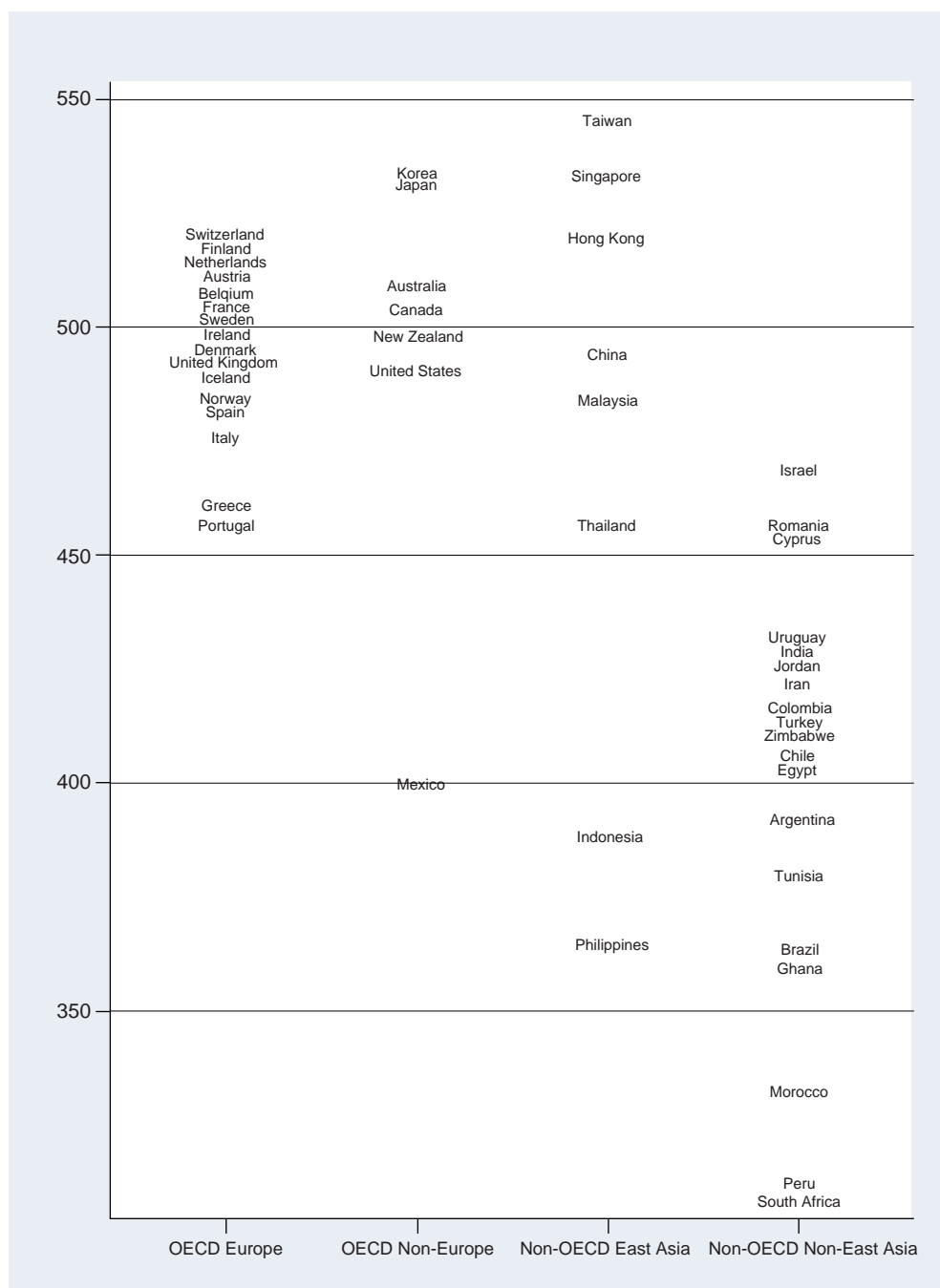


Figure 2 Performance on international student achievement tests. Simple average of the mathematics and science scores over all available international tests, using the rescaled data by Hanushek and Woessmann (2009) that puts performance at different international tests on a common scale.

growth. Their estimate suggest that one country-level standard deviation higher test performance (equivalent to 47-test-score points on the scale used in (Figure 2) would yield about one percentage point higher annual growth.

This estimate stems from a statistical model that relates annual growth rates of real GDP per capita to the measure of educational quality, years of schooling, the initial level of income, and several other control variables (including, in different specifications, the population growth rates, political measures, openness of the economies, and the like). Adding educational quality to a base specification including only initial income and educational quantity boosts the variance in GDP per capita among the 31 countries in Hanushek and Kimko's sample that can be explained by the model from 33% to 73%. The effect of years of schooling is greatly reduced by including quality, leaving it mostly insignificant. At the same time, adding the other factors leaves the effects of cognitive skills basically unchanged.

Several studies have since found very similar results, including Barro (2001), Woessmann (2002, 2003), Bosworth and Collins (2003), Coulombe and Tremblay (2006), and Jamison *et al.* (2007). In sum, the evidence suggests that the quality of education, measured by the knowledge that students gain as depicted in tests of cognitive skills, is substantially more important for economic growth than the mere quantity of schooling.

Recent Evidence on the Importance of Cognitive Skills for Economic Growth

The most recent evidence by Hanushek and Woessmann (2008, 2009) adds international student achievement tests not previously available and uses the most recent data on economic growth to analyze an even longer period (1960–2000). It extends the sample of countries with available test-score and growth information to 50 countries. These data are also used to analyze effects of the distribution of educational quality at the bottom and at the top on economic growth, as well as interactions between educational quality and the institutional infrastructure of an economy.

The measure of the quality of education is a simple average of the mathematics and science scores over international tests, interpreted as a proxy for the average educational performance of the whole labor force. This measure encompasses overall cognitive skills, not just those developed in schools. Thus, whether skills are developed at home, in schools, or elsewhere, they are included in the growth analyses.

After controlling for the initial level of GDP per capita and for years of schooling, the test-score measure features a statistically significant effect on the growth of real GDP per capita in 1960–2000 (Figure 3). According to this simple specification, test scores that are larger by 1 SD

(measured at the student level across all OECD countries in PISA) are associated with an average annual growth rate in GDP per capita that is two percentage points higher over the whole 40-year period which is almost identical to the prior estimates in Hanushek and Kimko (2000).

Adding educational quality to a model that just includes initial income and years of schooling increases the share of variation in economic growth explained from 25% to 73%. As reported above, the quantity of schooling is statistically significantly related to economic growth in a specification that neglects educational quality, but the association between years of schooling and growth turns insignificant and is reduced to close to zero once the quality of education is included in the model (Figure 4). In addition, considering the variation just within each of five world regions, educational quality is significantly related to economic growth, indicating that it does not simply reflect economic differences across regions.

Recent literature on the determinants of economic growth emphasizes the importance of the institutional framework of the economy (e.g., Acemoglu *et al.*, 2001, 2002). The most common and powerful measures of the institutional framework used in empirical work are the openness of the economy to international trade and the security of property rights. These two institutional variables are jointly highly significant when added to the model. However, the positive effect of educational quality on economic growth is very robust to the inclusion of these controls, albeit slightly reduced in magnitude to 1.26 percentage points annual growth per standard deviation of cognitive skills.

Other possible determinants of economic growth often discussed in the literature are fertility and geography. However, when the total fertility rate and common geographical proxies, such as latitude or the fraction of the land area located within the geographic tropics, are added to the model, neither is statistically significantly associated with economic growth.

The results are remarkably similar when comparing the sample of OECD countries to the sample of non-OECD countries, with the point estimate of the effect of educational quality slightly larger in non-OECD countries. When the sample is separated based on whether a country was below or above the median of GDP per capita in 1960, the effect of educational quality is statistically significantly larger in low-income countries than in high-income countries (cf. Hanushek and Woessmann, 2009).

Among the developing countries, the conclusion is that once there is a high-quality school system, it pays to keep children in school longer – but it does not pay if the school system does not produce skills.

The results are very robust to alternative specifications of the growth relationships. For example, the impact of

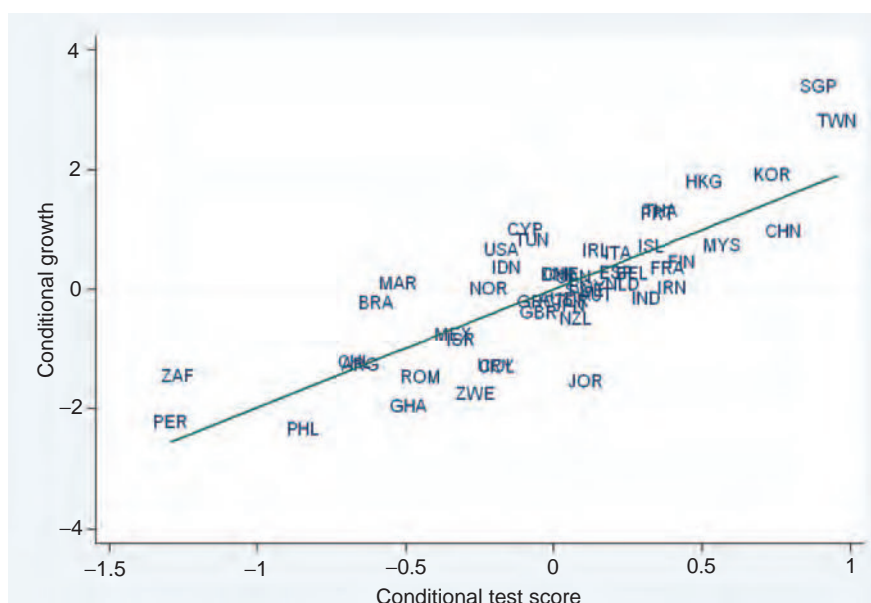


Figure 3 Test scores and long-run economic growth. Added-variable plots of a regression of the average annual rate of growth (in percent) of real GDP per capita in 1960–2000 on the initial level of real GDP per capita in 1960, average test scores on international student achievement tests, and average years of schooling in 1960. Calculations from Hanushek, E. A. and Woessmann, L. (2008).

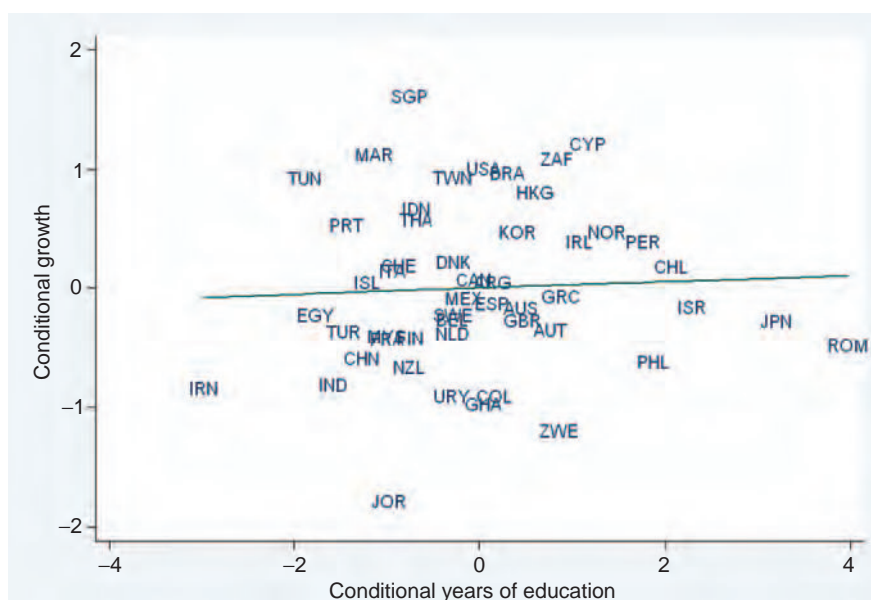


Figure 4 Years of schooling and economic growth after controlling for test scores. Added-variable plots of a regression of the average annual rate of growth (in percent) of real GDP per capita in 1960–2000 on the initial level of real GDP per capita in 1960, average test scores on international student achievement tests, and average years of schooling in 1960. Calculations from Hanushek, E. A. and Woessmann, L. (2008).

cognitive skills remains qualitatively the same when measured just by the tests performed at the level of lower secondary education, which seems the most readily comparable level. The results are also robust to performing the analyses in two subperiods, 1960–1980 and 1980–2000, and to dropping East Asian countries (which have both high levels of cognitive skill and rapidly growing economies).

All in all, the results do not appear to be an artifact of the specific time period, set of countries, or achievement measurement decisions.

Results are also confirmed when looking at whether a country's estimated cognitive skills affect the earnings of immigrants working in the United States. Higher home-country cognitive skills translate into higher earnings

if the immigrants were educated in their homeland, but not if educated in the United States (cf. Hanushek and Woessmann, 2009, for details). By looking at workers on the same labor market, this analysis addresses several issues of endogeneity, excluding the possibility that the results might be driven by other attributes of the countries that affect growth, such as efficient market organizations, which may also be associated with efficient and productive schools.

The Interaction of Educational Quality with Economic Institutions

Economic institutions appear to interact with the effect of educational quality on economic growth. The institutional framework of a country affects the relative profitability of piracy and productive activity. If the available knowledge and skills are used in the former activity rather than the latter, the effect on economic growth may be very different, perhaps even turning negative (North, 1990). The allocation of talent between rent-seeking and entrepreneurship matters for growth: countries with more engineering students grow faster and countries with more law students grow more slowly (Murphy *et al.*, 1991). Education may not have much impact in less-developed countries that lack other facilitating factors such as functioning institutions for markets and legal systems (Easterly, 2001). Besides, due to deficiencies in the institutional environment, cognitive skills might be applied to socially unproductive activities in many developing countries (Pritchett, 2001).

Adding the interaction of educational quality and one institutional measure – openness to international trade – to the growth specification indicates not only that both have significant individual effects on economic growth but also that there is a significant positive interaction. The effect of educational quality on economic growth is indeed significantly higher in countries that have been fully open to international trade than in countries that have been fully closed. The effect of educational quality on economic growth is significantly positive, albeit relatively low at 0.9 per SD in closed economies but increases to 2.5 per SD in open economies. When using protection against expropriation rather than openness to trade as the measure of institutional quality, there is similarly a positive interaction term with educational quality, although it lacks statistical significance.

In sum, both the quality of the institutional environment and the quality of education seem important for economic development. Furthermore, the effect of educational quality on growth seems significantly larger in countries with a productive institutional framework, so that good institutional quality and good educational quality can reinforce each other. Thus, the macroeconomic

effect of education depends on other complementary growth-enhancing policies and institutions. However, cognitive skills have a significant positive growth effect even in countries with a poor institutional environment.

Finally one issue related to institutions remains unresolved, Glaeser *et al.* (2004) suggests that human capital may itself lead to better institutions: as, such, the estimates of cognitive skills without consideration of economic institutions would be appropriate.

Simulating the Impact of Educational Reform on Economic Growth

It is important to understand the implications of policies designed to improve educational outcomes. The previous estimates provide information about the long-run economic implications of improvements in educational quality. For a better understanding of the impact of improved achievement, it is useful to relate policy reforms directly to the pattern of economic outcomes consistent with feasible improvements.

As a benchmark, consider a reform that yields a 0.5 standard deviation (SD) improvement in average achievement of school completers. We can put this metric in the context of the previous estimates. Consider, for example, a developing country – such as Brazil, Indonesia, Mexico, or Thailand in PISA 2003 – with average performance at roughly 400 test-score points, approximately minimal literacy. An aggressive reform plan would be to close half the gap with the average OECD student, an improvement of one half SD. Alternatively, consider what it would mean if a country currently performing near the mean of OECD countries in PISA at 500 test-score points (e.g., Norway or the United States in PISA 2000, or Germany in PISA 2003) managed to increase its educational quality to the level of top performers in PISA at roughly 540 test-score points (e.g., Finland or Korea). Such an increase amounts to 0.4 SD.

The timing of the reform is also important in two ways. First, such movement of student performance cannot be achieved instantaneously but requires changes in schools that will be accomplished over time (say, through systematic replacement of teachers through retirement and subsequent hiring). The time frame of any reform is difficult to specify, but achieving the change of 0.5 standard deviation (SD) described above for an entire nation may take 20–30 years. Second, if the reforms succeed, their impact on the economy will not be immediate – initially the new graduates will be a small part of the labor force. It will be some time after the reform of the schools before the impact on the economy is realized. In other words, the prior estimates are best thought of as the long-run, or equilibrium, outcomes of a labor force with a given educational quality.

Faster reforms will have larger impacts on the economy, simply because the better workers become a dominant part of the workforce sooner (**Figure 5**). However, even a 20- or 30-year reform plan begun in 2005 has a powerful impact. For example, a 20-year plan would yield a GDP 5% greater in 2037 (compared with an economy with no increase in educational quality). The figure also plots 3.5% of GDP, an aggressive spending level for education in many countries of the world. A growth dividend of 5% of GDP would more than cover all primary and secondary school spending. But even a 30-year reform program (not fully accomplished until 2035) would still yield more than 5% higher real GDP by 2041.

Projecting these net gains from improved educational quality further past the reform period vividly shows the long-run impacts of reform. Over a 75-year horizon, a 20-year reform yields a real GDP 36% higher than without a change in educational quality.

It must nonetheless be clear that these effects represent the result from actual gains in educational outcomes. There have been many attempts around the world to improve student outcomes, and many of these have failed to yield gains in student performance. Bad reforms – those without impacts on students – will not have these growth effects.

This simulation shows that the previous estimates of the effects of educational quality on growth have large impacts on national economies. At the same time, while the rewards are large, they also imply that policies must be considered across long periods, requiring patience – patience that is not always clear in national policymaking. These reforms must also be put in a broader perspective

because other kinds of institutional changes and investments will also take time. Changing basic economic institutions, for example, seldom happens overnight, and the economy needs time to adjust.

Summary

The accumulated evidence from analyses of economic outcomes is that the quality of education – measured on an outcome basis of cognitive skills – has powerful economic effects. Economic growth is strongly affected by the skills of workers. What people know matters.

This message is important in developed and developing countries alike. In the latter, much of the discussion of development policy today simplifies and distorts this message. It recognizes that education matters, but focuses most attention on ensuring that everybody is in school – regardless of the learning that goes on. As a recent report by the World Bank Independent Evaluation Group (2006) documents, high priority was accorded to increasing primary school enrolment in developing countries over the past 15 years. Whether children were learning garnered much less attention. International testing indicates that, even among those attaining lower secondary schooling, literacy rates (by international standards) are very low in many developing countries. By reasonable calculations, many countries have fewer than 10% of their youth currently reaching minimal literacy and numeracy levels, even when school attainment data look considerably better (cf. Hanushek and Woessmann (2008) for details).

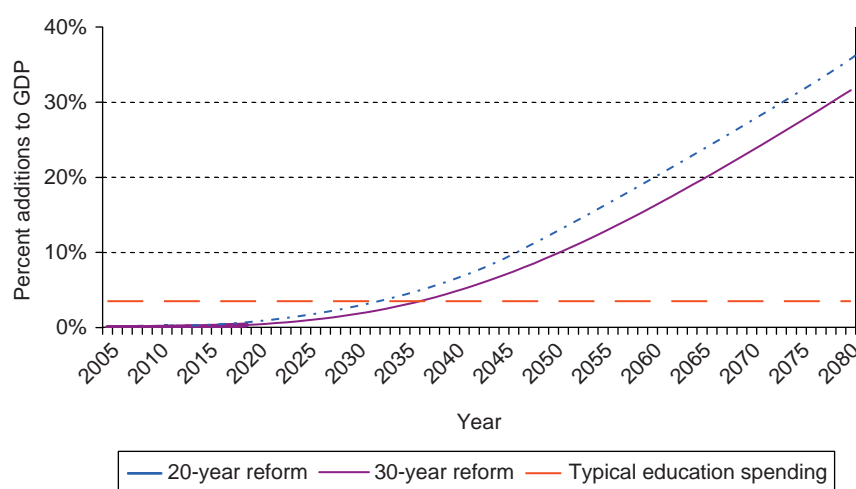


Figure 5 Simulation of the impact on GDP of moderately strong knowledge improvement. Simulation of the impact on the economy of reform policies beginning in 2005 and taking 20 or 30 years for a 0.5 SD improvement in student outcomes at the end of upper secondary schooling. The figure indicates how much larger the level of GDP is at any point after the reform policy is begun as compared to that with no reform; that is, the estimates suggest the increase in GDP expected over and above any growth from other factors. The figure also plots 3.5% of GDP, an aggressive spending level for education in many countries of the world. Even a 30-year reform program would yield a growth dividend covering the whole of this spending level by 2036. See Hanushek and Woessmann (2007) for details. From Hanushek, E. A. and Woessmann, L. (2007). The role of school improvement for economic development. *NBER Working Paper 12832*. Cambridge, MA: National Bureau of Economic Research.

Because of the reported findings – that knowledge rather than just time in school is what counts for economic growth – policies must pay more attention to the quality of schools.

See also: Education and Inequality; Returns to Education in Developed Countries; Returns to Education in Developing Countries; School Quality and Earnings; The External Benefits of Education.

Bibliography

- Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review* **91**(5), 1369–1401.
- Acemoglu, D., Johnson, S., and Robinson, J. A. (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *Quarterly Journal of Economics* **117**(4), 1231–1294.
- Aghion, P. and Howitt, P. (1998). *Endogenous Growth Theory*. Cambridge, MA: MIT Press.
- Barro, R. J. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics* **106**(2), 407–443.
- Barro, R. J. (1997). *Determinants of Economic Growth: A Cross-Country Empirical Study*. Cambridge, MA: MIT Press.
- Barro, R. J. (2001). Human capital and growth. *American Economic Review* **91**(2), 12–17.
- Benhabib, J. and Spiegel, M. M. (1994). The role of human capital in economic development: Evidence from aggregate cross-country data. *Journal of Monetary Economics* **34**(2), 143–174.
- Bils, M. and Klenow, P. J. (2000). Does schooling cause growth? *American Economic Review* **90**(5), 1160–1183.
- Bosworth, B. P. and Collins, S. M. (2003). The empirics of growth: An update. *Brookings Papers on Economic Activity* **2003**(2), 113–206.
- Coulombe, S. and Tremblay, J. -F. (2006). Literacy and growth. *Topics in Macroeconomics* **6**(2), Article 4.
- Easterly, W. (2001). *The Elusive Quest for Growth: An Economist's Adventures and Misadventures in the Tropics*. Cambridge, MA: MIT Press.
- Glaeser, E. L., et al. (2004). “Do institutions cause growth?” *Journal of Economic Growth* **9**(3), 271–303.
- Hanushek, E. A. and Kimko, D. D. (2000). Schooling, labor force quality, and the growth of nations. *American Economic Review* **90**(5), 1184–1208.
- Hanushek, E. A. and Woessmann, L. (2008). “The role of cognitive skills in economic development” *Journal of Economic Literature* **46**(3), 607–668.
- Hanushek, E. A. and Woessmann, L. (2009). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *NBER Working Paper No. 14633*. Cambridge, MA, National Bureau of Economic Research.
- Jamison, E. A., Jamison, D. T., and Hanushek, E. A. (2007). The effects of education quality on mortality decline and income growth. *Journal of Economic Literature* **46**(3), 607–668.
- Krueger, A. B. and Lindahl, M. (2001). Education for growth: Why and for whom? *Journal of Economic Literature* **39**(4), 1101–1136.
- Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics* **22**, 3–42.
- Mankiw, N. G., Romer, D., and Weil, D. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics* **107**(2), 407–437.
- Murphy, K. M., Shleifer, A., and Vishny, R. W. (1991). The allocation of talent: Implications for growth. *Quarterly Journal of Economics* **106**(2), 503–530.
- Nelson, R. R. and Phelps, E. (1966). Investment in humans, technology diffusion and economic growth. *American Economic Review* **56**(2), 69–75.
- North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.
- Pritchett, L. (2001). Where has all the education gone? *World Bank Economic Review* **15**(3), 367–391.
- Pritchett, L. (2006). Does learning to add up add up? The returns to schooling in aggregate data. In Hanushek, E. A. and Welch, F. (eds.) *Handbook of the Economics of Education*, pp 635–695. Amsterdam: North-Holland.
- Romer, P. (1990). Endogenous technological change. *Journal of Political Economy* **99**(5, pt. II), S71–S102.
- Sala-i-Martin, X., Doppelhofer, G., and Miller, R. I. (2004). Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *American Economic Review* **94**(4), 813–835.
- Sianesi, B. and Van Reenen, J. (2003). The returns to education: Macroeconomics. *Journal of Economic Surveys* **17**(2), 157–200.
- Temple, J. (2001). Growth effects of education and social capital in the OECD countries. *OECD Economic Studies* **33**, 57–101.
- Topel, R. (1999). Labor markets and economic growth. In Ashenfelter, O. and Card, D. (eds.) *Handbook of Labor Economics*, pp 2943–2984. Amsterdam: Elsevier.
- Vandenbussche, J., Aghion, P., and Meghir, C. (2006). Growth, distance to frontier and composition of human capital. *Journal of Economic Growth* **11**(2), 97–127.
- Woessmann, L. (2002). *Schooling and the Quality of Human Capital*. Berlin: Springer.
- Woessmann, L. (2003). Specifying human capital. *Journal of Economic Surveys* **17**(3), 239–270.
- World Bank Independent Evaluation Group (2006). *From Schooling Access to Learning Outcomes: An Unfinished Agenda*. Washington, DC: World Bank.

Further Reading

- Acemoglu, D., Johnson, S., and Robinson, J. A. (2005). Institutions as a fundamental cause of long-run growth. In Aghion, P. and Durlauf, S. N. (eds.) *Handbook of Economic Growth*, pp 385–472. Amsterdam: North-Holland.
- Benhabib, J. and Spiegel, M. M. (2005). Human capital and technology diffusion. In Aghion, P. and Durlauf, S. N. (eds.) *Handbook of Economic Growth*, pp 935–966. Amsterdam: North-Holland.
- Hanushek, E. A., Jamison, D. T., Jamison, E. A., and Woessmann, L. (2008). Education and economic growth: It's not just going to school but learning that matters. *Education Next* **8**(2), 62–70.
- Hanushek, E. A. and Woßmann, L. (2007). *Education Quality and Economic Growth*. Washington, DC: World Bank.
- Pritchett, L. (2006). Does learning to add up add up? The returns to schooling in aggregate data. In Hanushek, E. A. and Welch, F. (eds.) *Handbook of the Economics of Education*, pp 635–695. Amsterdam: North-Holland.

Relevant Website

<http://www.hanushek.net>

Parental Socioeconomic Status, Child Health, and Human Capital

J Currie, Columbia University, New York, NY, USA

J Goodman, Harvard University, Cambridge, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Fetal origins hypothesis – A biological theory suggesting that the environment an individual is exposed to as a fetus can have long-term effects on his or her health.

Grossman health model – A theoretical model describing how an individual's health relates to his or her previous health and choices of medical inputs, such as medical care, food, and housing.

Sibling fixed effects – A regression analysis technique that compares siblings to each other in order to remove the statistical bias due to family background.

Introduction

Investments in education pay off in the form of higher future earnings, and differences in educational attainments explain a significant fraction of the adult variation in wages, incomes, and other outcomes. But what determines a child's educational success? Most studies point to family background as the primary factor. But why does background matter? While many aspects are no doubt important, research increasingly implicates health as a potentially major factor. The importance of health for education and earnings suggests that if family background affects child health, then poor child health may in turn affect education and future economic status.

What evidence exists about the effect of parental socioeconomic status (SES) on child health? What evidence exists about the effect of child health on future outcomes, such as education? A great deal of evidence shows that low SES in childhood is related to poorer future adult health (Smith *et al.*, 1998). The specific question at the heart of this article is whether low parental SES affects future outcomes through its effects on child health. In most of the studies cited, SES is defined by parental income or poverty status, although some measure SES through residential neighborhood or parental schooling attainment. The article focuses primarily on children from developed countries because it is more obvious why the common and severe health problems of children

in many developing countries might impede human capital development.

Does Parental Socioeconomic Status Affect Child Health?

External Benefits of Parental SES

Parental SES may impact both parents' own health and the health of their children. Schooling attainment, in particular, has a stronger correlation with a parent's own good health than does income or other measures of SES (Grossman, 2007). The association between education and parents' own health is only partially explained by better health knowledge and may be better explained by the fact that more highly educated parents tend to exhibit better health behaviors. Even after controlling for income, parents with more schooling smoke less, drink less alcohol, exercise more, and work less often in dangerous occupations. They also adhere more carefully to prescribed medical therapies and are more likely to use newer medical technologies to address health problems. These tendencies may be caused by education, or they may indicate that people who plan for the future better tend both to pursue more schooling and to behave in healthier ways.

This article focuses, however, not on the internal benefits of parental SES (i.e., for parents' own health) but on the external benefits of parental SES for children's health. Through what channels might these benefits flow? In the health model presented by Grossman (2000), a health production function Q_t describes how a child's current health depends on health inputs such as medical care, food, and housing, as well as previous health levels Q_{t-1} , Q_{t-2} , etc. This is similar to an education production function that models how a child's test score depends on inputs such as teachers and textbooks as well as previous test scores. Grossman's health model yields several insights into how parental SES might affect child health. First, and perhaps most obviously, budget constraints bind more in poorer families, preventing them from buying more or better material health inputs such as better-quality medical care and food, as well as safer housing and neighborhoods.

Second, SES affects what parents choose to do with the health inputs they can afford, as parents of lower SES may have different past experiences with the healthcare system, or different health preferences, or different health beliefs (e.g., whether it is normal for a child to wheeze).

Parental education may play a particularly important role in this regard. Maternal schooling is strongly correlated with neonatal mortality rates and children's overall health, which may indicate the importance of health knowledge but is more likely explained by the association between schooling and various health behaviors (Grossman, 2007). More highly educated mothers smoke less, drink less, take more vitamins, and receive more prenatal medical care. In other words, they treat health inputs that impact their children, like cigarettes and alcohol, differently than do less-educated mothers. Once these inputs are controlled for, maternal schooling has little additional association with child health, suggesting that parental education may affect child health largely through the use of such health inputs.

Finally, children of lower SES families are likely to have lower health status at birth. This is not necessarily due to a worse genetic endowment but may stem from differing environmental triggers that activate certain genes (Rutter, 2006). Thus, a low SES child may have poor health at birth because of the circumstances surrounding gestation and birth, rather than because of worse genetic endowments. All of the above may be mechanisms through which SES affects child health.

Evidence

Correlation

Differences in the health of high and low SES children are apparent at birth. Data from Britain and California show that low SES children are more likely to have low birth weight than high SES children. Maternal reports of overall child health from the US, Britain, and Canada all show that the health gap between high and low SES children continues through early childhood and beyond (Currie *et al.*, 2007). The health gaps are smaller in Britain and Canada than in the US, perhaps due to universal health insurance coverage, but are still present.

Variations in the incidence of health insults (such as hospitalizations or new diagnoses of chronic conditions) may be of particular importance in explaining the gap in health status between rich and poor. Evidence from the US, Britain, and Canada suggests that poor children are more likely to receive health insults and suffer from chronic conditions than rich children (Currie and Lin, 2007). More than twice as many poor children than nonpoor children are reported by their mothers to be in less than very good health, a gap that increases as children age. Further, 32.4% of poor children suffer from a chronic condition, compared to 26.5% of nonpoor children, a gap that would likely be even larger if differences in diagnosis probabilities were accounted for. Such chronic conditions also limit poor children more than nonpoor children. The percentage of poor children report being limited by their chronic conditions is 11.4% as compared to 7.0% of nonpoor children. The fraction of children with a limitation

due to a chronic condition rises with age, and rises more sharply for poor children than for others. By their teenage years, poor children have almost double the probability of being limited by their chronic condition: 14.1% compared to 7.8% of nonpoor children.

Theoretical models suggest that persistent poverty is likely to have worse effects on health than transitory poverty. Although more research is needed, evidence from several studies suggests that persistent poverty affects child mental health, particularly aggressive behavior, more than current poverty (Strohschein, 2005).

Causation

The fact that children of low SES parents are less healthy on average than other children does not necessarily imply that low SES causes poor child health. A third factor such as poor parental health may, for example, cause both poverty and poor child health. Alternatively, poor child health may cause low SES by reducing parental earnings. Identifying causal effects matters very much because interventions to improve parental SES will not necessarily improve child health if parental SES does not directly affect child health. Unfortunately, relatively little literature attempts to identify causal impacts of parental SES on child health in a developed country context, perhaps because of the difficulty of finding interventions that affect parental SES but that do not also directly affect children's health. Research in this area uses one of two approaches. The first approach is to ask whether the correlation between SES and child health remains once other variables are controlled for. The second is to examine the effect of natural experiments that randomly change some parents' SES relative to a control group.

Mother's education, one measure of SES, seems to have a positive impact on child health. In the US, the great expansion of higher education in the 1960s and 1970s raised women's education levels, which in turn improved infant health as measured by birth weight and gestational age (Currie and Moretti, 2003). The effect may have occurred through increased rates of marriage and prenatal care, as well as through substantial reductions in smoking.

Income itself, as a measure of SES, seems to have relatively little effect on child health. Welfare-to-work experiments, for example, have had little impact on child health, either positive or negative. Income may, however, matter more in a developing country context. For example, black South African girls increased their height-for-age when their grandmothers started receiving old-age pensions, suggesting increased investment in nutrition (Duflo, 2000). Finally, the state of the economy may impact child health. Dutch citizens born during recessions have higher mortality rates at all ages compared to those born just prior to the recession, though the precise pathway for this effect is unclear (Van den Berg *et al.*, 2006).

Studies of American and British families find that the apparent effect of income on child mental health is considerably lessened once other factors, such as parenting skills and physical home environment, are controlled for (Berger *et al.*, 2006). Estimates from the American study suggest that even cash subsidies to bring every family up to the poverty line would not eliminate the observed gaps in child outcomes.

Neighborhoods are often said to be an important pathway for SES to affect outcomes. Some evidence on this point comes from a US social experiment that randomly assisted some public housing residents to move to low-poverty neighborhoods. This moving-to-opportunities experiment improved the mental health of girls through reductions in generalized anxiety disorders and psychological distress (Orr *et al.*, 2003). Curiously, there was no such positive effect for boys.

Some recent research has attempted to control for unobserved family background characteristics by examining children born to the same mother (i.e., estimating models with sibling fixed effects). Some studies of American mothers suggest that on average, maternal income during pregnancy does not affect the probability of having a child with low birth weight, but that it does matter if the mother herself had a low birth weight. Evidence from California birth records suggests that, even among women with the same mother, being born in a poor area increased the probability of being low birth weight and of later delivering a baby with low birth weight (Currie and Moretti, 2007). One final piece of evidence that maternal SES affects infant health comes from examination of the health improvements black women experienced as a result of the US Civil Rights movement. Infants of black women who themselves had healthier infancies as a result of the Civil Rights movement (which improved hospital access for blacks in southern states) show large gains in birth weight relative to the infants of black women born just a few years earlier (Almond *et al.*, forthcoming).

In summary, it is difficult to prove that the strong and exceedingly robust correlation between parental SES and child health is a causal relationship. The literature attempting to do so is underdeveloped. There is, however, evidence that maternal SES early in the child's life matters, and that child mental health may be particularly susceptible to the effects of early deprivation.

Does Child Health Affect Future Outcomes?

Possible Channels

Poor child health may impact adult labor supply and productivity through two channels. First, it may damage adult health. Cohorts that suffer high death rates in childhood

may also show high death rates in adulthood, in part because of the direct effects of childhood health conditions on future morbidity. In rich countries, cohorts that suffer a higher disease burden in childhood have higher adult death rates, though in poor countries the relation is reversed because only relatively healthy people survive to adulthood (Bozzoli *et al.*, 2007). In the US, adults' reports about their overall childhood health are highly correlated with current adult outcomes, a pattern that continues to hold even once family background is controlled for by comparing siblings to each other. Thus, adult siblings who had better health in childhood have 24% higher incomes, higher wealth, more weeks worked per year, and a higher growth rate of income (Smith, 2007). Comparing siblings to each other reduces the apparent effect of childhood health on future education, suggesting that childhood health may affect future income through mechanisms other than educational attainment. Sickly children may, for example, be less able to work hard as adults.

Second, poor child health may impair children's educational attainment and thus skill acquisition. Among older children, school absences may be a mechanism for health to affect education, although overall absenteeism is quite small for both poor and nonpoor children. It is more likely that poor health impacts skill acquisition by impairing children's ability to learn while they are in school. Conditions such as anemia and lead poisoning have this effect, though today they are relatively rare in developed countries. Conditions such as tooth decay and ear infections are much more common and might therefore have a greater overall impact. Mental health conditions may be a particularly important mechanism because they are common and have worse effects on schooling attainment than most physical chronic conditions.

Evidence

In developing countries, children in poor health tend to have lower educational attainments, but surprisingly little examination of this relationship has occurred in developed countries. Data on older Americans show that the apparent effect of a retrospective measure of childhood SES on future health, education, and income shrinks when child health measures are included (Luo and Waite, 2005). This result implies that child health may explain some of the impact of low childhood SES on future outcomes.

The primary deficiency of this literature is that correlations between child health and future outcomes, including those mentioned above, may be due to other characteristics of households that are associated both with poor child health and worse outcomes. Until the last decade, most studies claiming a causal connection between child health and future educational attainment suffered from methodological weaknesses, but in the past decade an outpouring of research on this topic has paid careful attention to the

causal question. The remainder of this section examines specific child health problems that may work through the two channels described above, starting with conditions *in utero* and low birth weight, for which there is much causal evidence, and continuing with nutrition, mental health, asthma, acute conditions, and environmental toxins, for which there are fewer causal studies.

Conditions in utero

Increasing numbers of studies have focused on the hypothesis that fetal conditions are related to adult risk of disease, an idea that has become known as the fetal origins or the Barker hypothesis (Barker, 1998; Gluckman and Hanson, 2005). This literature strongly suggests that conditions *in utero* affect not only birth weight but features such as basic metabolism, which in turn affect future health outcomes. Fetuses starved *in utero* may develop more efficient metabolisms that raise the risk of future obesity, heart disease, and diabetes. As adult health is strongly linked to adult economic well-being, this suggests a relationship between health *in utero* and future outcomes.

The most compelling tests of the hypothesis look for sharp exogenous shocks in fetal health caused by conditions outside the mother's control. Dutch adults who were *in utero* during the 1944–1945 famine caused by Nazi occupation were more likely to suffer various health impairments including nervous disorders, heart disease, and antisocial personality disorders. Swedes who were *in utero* when the 1986 Chernobyl disaster exposed their mothers to low-dose fallout were less likely to qualify for high school and had lower grades (Almond *et al.*, 2007). Americans who were *in utero* during the 1918 influenza epidemic were much less likely to graduate from high school, had lower wages, were more likely to be poor and receiving transfer payments, and as adults suffered more from schizophrenia, diabetes, and stroke (Almond, 2006). In general, health shocks in early life due to wars, famines, and other crises can have large, lasting effects on health.

Cognitive functioning can also be directly affected by conditions *in utero* and in infancy. For example, maternal alcohol consumption can lead to permanent brain damage, as can trauma during the birth itself. Extreme deprivation in early childhood, such as that experienced by some Romanian orphans in state-run nurseries, demonstrably impairs cognitive functioning (O'Connor *et al.*, 2000). Severe health insults *in utero* or in early childhood clearly can cause permanent cognitive impairments, but questions remain about how sensitive these sensitive periods are and whether damage due to less-extreme deprivation is noteworthy or widespread.

Birth weight

More direct evidence is provided by recent literature linking low birth weight to negative future outcomes. In the US, low-birth-weight babies have a much higher

infant mortality rate than their heavier counterparts (Conley *et al.*, 2003). They also have lower average scores on a variety of tests of intellectual and social development. British children with low birth weight have lower test scores, educational attainments, wages, and probabilities of being employed as adults, even conditional on many measures of family background (Case *et al.*, 2005).

Many of the studies exploring the effect of birth weight on future outcomes compare siblings or twins in an attempt to control for unobserved family characteristics that might otherwise bias the results. Some small-sample studies that do this in an American context conclude that siblings with lower birth weight tend to attain less education. More recently, several studies have employed individual-level national vital statistics (birth certificate) data in Canada, Norway, and Scotland to examine this question. All of these studies show a link between low birth weight and lower educational attainment, and some show a negative effect on height and intelligence, even among siblings or twins (Black *et al.*, 2007). Similar findings occur in the US, where a number of studies confirm that siblings (or twins) with lower birth weight attain less education than their higher birth-weight counterparts (Currie and Moretti, 2007). Data from the US also suggests that lower birth weight is associated with a higher probability of living in a poor area, a lower probability of being married, lower earnings, worse health, and worse cognitive abilities.

Nutrition

Nutrition may play a significant role in the child's cognitive development. Randomized trials in developing countries such as Guatemala, for example, indicate that poor nutrition can harm cognition (Maluccio *et al.*, 2006). It is less obvious that nutritional supplementation should have a large effect on the cognitive achievement of children in richer countries. Several US studies have, however, found positive effects of prenatal participation in the special supplemental nutrition program for women, infants, and children (WIC), which provides coupons that can be redeemed for specific foods to women, infants, and children who are deemed to be nutritionally at risk (Kowaleski-Jones and Duncan, 2002). Children of mothers participating in WIC had better outcomes on cognitive tests even when compared to a control group of higher income, better-educated women also receiving prenatal care in clinic settings. Children born while their mothers participated in WIC show better temperament, although not better motor or social skills, than their siblings born while their mothers were not participating. These studies underline the importance of the prenatal period.

Further evidence on the importance of nutrition comes from the fact that height is a good measure of a population's average health. Interestingly, the well-established relationship between adult height and earnings disappears when early childhood cognitive test scores are controlled for

(Case and Paxson, 2006). Since much of the variation in adult height is due to childhood nutrition, this suggests that poor childhood nutrition likely affects both cognitive performance and adult height, leading to the observed correlation between height and earnings.

Mental health

The prevalence and importance of child mental health problems have been increasingly recognized. Approximately one in five children and adolescents in the US exhibit some impairment from a mental or behavioral disorder, 11% have significant functional impairments, and 5% suffer extreme functional impairment. Moreover, mental health problems are one of the leading causes of days lost in the workplace because they strike many people of working age. Retrospective questions asked to US adults suggest that those with early-onset psychiatric problems were less likely to have graduated from high school or attended college. Children's mental health problems are usually grouped into four categories: anxiety, depression, hyperactivity, and conduct disorders (aggressive or antisocial behavior). The evidence to date suggests that these last two externalizing problems have the greatest impact on outcomes.

Children with behavioral problems in Britain and New Zealand have poorer schooling, earnings, and employment outcomes as young adults than their counterparts without such problems. Hyperactivity and conduct disorders seem to cause these negative outcomes, while anxiety and depression have little effect. US data also suggest that children with behavior problems at young ages are less likely to graduate from high school or to attend college, even after conditioning on maternal characteristics (McLeod and Kaiser, 2004). American children with attention-deficit hyperactivity disorder (ADHD) complete less schooling and are more likely to have continuing mental health problems than a group of control children consisting either of children from the same school or nonpsychiatric patients in the same medical center (Mannuzza and Klein, 2000).

Beyond adding available controls to regression models, many of these studies do not address the possibility that the negative outcomes might be caused by other factors related to a diagnosis of mental health problems, such as poverty or the presence of other learning disabilities. Recent studies address these challenges by comparing siblings, thus eliminating any family background characteristics as a source of bias (Currie and Stabile, 2006). In both the US and Canada, siblings with high scores on an ADHD screener had lower math and reading scores and higher probabilities of being in special education or having repeated a grade than their siblings with low scores on the ADHD screener. ADHD appears to have larger effects on academic outcomes than childhood depression, conduct disorders, or other mental

problems, and the effects of ADHD are large relative to those of physical chronic conditions.

Asthma

Poor children are more likely to suffer from and be limited by asthma, the most prevalent childhood chronic condition, than are nonpoor children. Though siblings with controlled asthma show no difference in achievement scores than siblings without asthma, several studies indicate that asthmatic children are more likely than similar but nonasthmatic children to have behavior problems, even when the asthma is well controlled (Calam *et al.*, 2003). Asthmatic children are absent more frequently from school, have higher incidence of learning disabilities, and repeat grades more often. They also have lower scores on a test of school readiness skills and their parents were three times more likely to report that they needed extra help with learning, particularly if children reported that their asthma caused activity limitations. These studies suffer from the deficiency that the apparent connection between asthma and outcomes could reflect omitted third factors because asthma is more prevalent among poor and minority children. The fact that several of the studies do, however, use very homogeneous groups of children and still find behavioral differences suggests that uncontrolled asthma probably does have a causal effect on behavior.

Acute illnesses

Poor children are more likely to suffer from acute illnesses such as tooth decay and ear infections than their richer peers. Ear infections affect most young children at one time or another and are the most common reason children visit a doctor. Roughly 5% of 2–4-year-old children have hearing loss because of middle-ear effusion lasting 3 months or longer. Hearing loss can delay language development, but little research has been done to determine how important these effects might be in explaining disparities in cognitive or academic outcomes.

Environmental toxins

One final category of health problems that may explain disparities in outcomes between poor and nonpoor children is exposure to environmental toxins. The most obvious of these toxins is lead, as lead poisoning has been shown to significantly decrease IQ, and majority of affected children are from low-income families. Lead may also worsen children's mental health, making them more prone to antisocial behavior. Adoption of public health measures such as banning lead in paint and gasoline have, however, caused the number of US children with unsafe lead levels to decline from 13.5 million in 1988 to less than half a million in 2000. Relatively little research examines the health effects of exposure to other environmental toxins at the level now generally occurring

in the population. Data on possible human health effects generally come from either animal studies, or disastrous releases. Residents of areas near hazardous waste sites are more likely to be poor and have lower levels of education than people in the remainder of the country and therefore their children's health outcomes are likely to differ even in the absence of negative health effects from exposure. Some studies try to control for observable confounding factors, but unobservable characteristics of people who live near hazardous waste sites may tend to cause bad outcomes.

One approach that avoids this omitted variable problem uses variation in pollution levels stemming from implementation of the 1970 and 1977 Clean Air acts, which caused exogenous changes in air-pollution levels across counties. Countries that experienced larger air-pollution reduction also experienced decreased rates of infant mortality (Chay and Greenstone, 2003). The Clean Air acts also reduced prenatal exposure to lead, which in turn decreased infant mortality and the proportion of low-birth-weight babies (Reyes, 2005). Other papers account for omitted characteristics such as ground-water pollution and SES by examining changes in pollution over time within single zip-code areas. These studies reveal that reducing pollutants such as carbon monoxide lowers infant mortality rates, as well as hospitalization rates for childhood asthma (Currie and Neidell, 2005). These studies show that pollution can have causal effects on child health, but there has been little investigation of whether these effects have long-term consequences for children's outcomes. The National Children's Study attempts to remedy this by examining the effects of environmental exposures on 100,000 children from birth to age 21 years.

Can Health Account for Gaps in Children's Educational Outcomes?

In order for a given health problem to lead to a disparity in educational outcomes, the health problem must either be more prevalent among the poor or have a larger negative effect on the poor, and must also be associated with lower educational attainments. Few of the specific health problems mentioned above fit both these criteria. Mental health problems are much more prevalent among the poor and have large negative effects, but are still too rare to explain observed human capital disparities. Similarly, the long-term effects of low birth weight are statistically significant but relatively small. The same is true for many of the other specific conditions, while for some of the other large categories, such as injuries and exposure to environmental toxins, too little evidence currently exists to determine the likely long-term effects or the extent of the disparity in exposures. One exception to this generalization is the fetal injuries mentioned above, which have

very large effects on future outcomes. Children of US mothers infected during the flu epidemic were 15% less likely to graduate from high school, and Swedish children exposed to low-level radiation after Chernobyl were 5.6% less likely to qualify for high school. These results raise the provocative idea that one of the best ways to safeguard children's health and educational attainments may be to start with their pregnant (or prepregnant) mothers.

To summarize, this article surveys literature focusing on two questions: Do parental circumstances affect child health at early ages? Does child health matter for future educational attainments? The answer to both questions appears to be yes. It is too early to tell how important these feedbacks between health and more conventional measures of human capital may be. We know too little about the cumulative and interactive effects of health insults. The available evidence suggests that fetal health may be particularly important. We need to understand more about the reasons why poor children suffer a higher incidence of negative health events, even *in utero*, so that we can do more to prevent them. Much of the literature reviewed here is extremely recent, suggesting that this topic will continue to be a fruitful area of research.

See also: Education Production Functions: Concepts; Education Production Functions: Developed Country Evidence; Education Production Functions: Evidence from Developing Countries; Human Capital; The External Benefits of Education.

Bibliography

- Almond, D. (2006). Is the 1918 influenza pandemic over? Long-term effects of *in utero* influenza exposure in the post-1940 U.S. population. *Journal of Political Economy* **114**, 672–712.
- Almond, D., Chay, K., and Greenstone, M. (forthcoming). Civil rights, the war on poverty, and black-white convergence in infant mortality in Mississippi. *American Economic Review*.
- Almond, D., Edlund, L., and Palme, M. (2007). Chernobyl's subclinical legacy: Prenatal exposure to radioactive fallout and school outcomes in Sweden. *NBER Working Paper* 13347.
- Barker, D. J. P. (1998). *Mothers, Babies and Health in Later Life*, 2nd edn. Edinburgh: Churchill Livingstone.
- Berger, L., Paxson, C., and Waldfogel, J. (2006). Income and child development. *Working Paper*.
- Black, S. E., Devereux, P. J., and Salvanes, K. G. (2007). From the cradle to the labor market? The effect of birth weight on adult outcomes. *Quarterly Journal of Economics* **122**, 409–439.
- Bozzoli, C., Deaton, A., and Quintana-Domeque, C. (2007). Child mortality, income and adult height. *NBER Working Paper* 12966.
- Calam, R., Gregg, L., Simpson, B. M., et al. (2003). Childhood asthma, behaviour problems and family functioning. *Journal of Allergy and Clinical Immunology* **122**, 499–504.
- Case, A., Fertig, A., and Paxson, C. (2005). The lasting impact of childhood health and circumstance. *Journal of Health Economics* **24**, 365–389.
- Case, A. and Paxson, C. (2006). Stature and status: Height, ability, and labor market outcomes. *NBER Working Paper* 12466.
- Chay, K. Y. and Greenstone, M. (2003). Air quality, infant mortality, and the Clean Air Act of 1970. *NBER Working Paper* 10053.

- Conley, D., Strully, K., and Bennett, K. (2003). *The Starting Gate: Birth Weight and Life Chances*. Berkeley, CA: University of California Press.
- Currie, A., Shields, M. A., and Wheatley Price, S. (2007). The child health/family income gradient: Evidence from England. *Journal of Health Economics* **26**, 213–232.
- Currie, J. and Lin, W. (2007). Chipping away at health: More on the relationship between income and child health. *Health Affairs* **26**, 331–344.
- Currie, J. and Moretti, E. (2003). Mother's education and the intergenerational transmission of human capital: Evidence from college openings. *Quarterly Journal of Economics* **118**, 1495–1532.
- Currie, J. and Moretti, E. (2007). Biology as destiny? Short and long-run determinants of intergenerational transmission of birth weight. *Journal of Labor Economics* **25**, 231–264.
- Currie, J. and Neidell, M. (2005). Air pollution and infant health: What can we learn from California's recent experience? *Quarterly Journal of Economics* **120**, 1003–1030.
- Currie, J. and Stabile, M. (2006). Child mental health and human capital accumulation: The case of ADHD. *Journal of Health Economics* **25**, 1094–1118.
- Dufló, E. (2000). Child health and household resources in South Africa: Evidence from the old age pension program. *American Economic Review* **90**, 393–398.
- Gluckman, P. and Hanson, M. (2005). *The Fetal Matrix: Evolution, Development, and Disease*. New York: Cambridge University Press.
- Grossman, M. (2000). The human capital model. In Culver, A. and Newhouse, J. P. (eds.) *The Handbook of Health Economics*, pp 347–408. Amsterdam: North Holland.
- Grossman, M. (2007). Education and nonmarket outcomes. In Hanushek, E. and Welch, F. (eds.) *Handbook of the Economics of Education*, pp 577–633. Amsterdam: North-Holland.
- Kowaleski-Jones, L. and Duncan, G. J. (2002). Effects of participation in the WIC food assistance program on children's health and development: Evidence from NLSY children. *American Journal of Public Health* **92**, 799–804.
- Luo, Y. and Waite, L. J. (2005). The impact of childhood and adult SES on physical, mental and cognitive well-being in later life. *Journals of Gerontology* **60B**, S93–S101.
- Maluccio, J., Hoddinott, J., Behrman, J., et al. (2006). The impact of nutrition during early childhood on education among Guatemalan adults. University of Pennsylvania, Population Studies Center Working Paper 06-04.
- Mannuzza, S. and Klein, R. G. (2000). Long-term prognosis in attention-deficit/hyperactivity disorder. *Child and Adolescent Psychiatric Clinics of North America* **9**, 711–726.
- McLeod, J. and Kaiser, K. (2004). Childhood emotional and behavioral problems in educational attainment. *American Sociological Review* **69**, 636–658.
- O'Connor, T. G., Rutter, M., and the English and Romanian Adoptees Study Team (2000). Attachment disorder behavior following early severe deprivation: Extension and longitudinal follow-up. *Journal of the American Academy of Child and Adolescent Psychiatry* **39**, 703–712.
- Orr, L., Feins, J., Jacob, R., et al. (2003). Moving to opportunity: Interim impacts evaluation. Washington, DC: U.S. Department of Housing and Urban Development.
- Reyes, J. W. (2005). The impact of prenatal lead exposure on infant health. *Working Paper*.
- Rutter, M. (2006). *Genes and Behavior: Nature–Nurture Interplay Explained*. Oxford: Blackwell.
- Smith, G. D., Hart, C., Blane, D., and Hole, D. (1998). Adverse socioeconomic conditions in childhood and cause specific adult mortality: Prospective observational study. *British Medical Journal* **316**, 1631–1635.
- Smith, J. P. (2007). The impact of SES on health over the life-course. *Journal of Human Resources* **52**, 739–764.
- Strohschein, L. A. (2005). Household income histories and child mental health trajectories. *Journal of Health and Social Behavior* **46**, 359–375.
- Van den Berg, G., Lindeboom, M., and Portrait, F. (2006). Economic conditions early in life and individual mortality. *American Economic Review* **96**(1), 290–302.

Further Reading

- Costello, J. E., Compton, S. N., Keeler, G., and Angold, A. (2003). Relationships between poverty and psychopathology: A natural experiment. *Journal of the American Medical Association* **290**, 2023–2028.
- Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature* **47**, 87–122.
- Heckman, J. J. (2007). The technology and neuroscience of capacity formation. *Proceedings of the National Academy of Sciences* **104**, 13250–13255.

Relevant Websites

- <http://www.cdc.gov> – Centers for Disease Control and Prevention (CDC).
- <http://papers.nber.org> – National Bureau of Economic Research.
- <http://www.nationalchildrensstudy.gov> – National Children's Study.
- <http://www.who.int> – World Health Organization.

The External Benefits of Education

W W McMahon, University of Illinois at Urbana-Champaign, Urbana, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Education externalities are the public benefits of education that spillover to benefit others in the society, including others in future generations. External benefits are distinguished from private market benefits to earnings and from the private nonmarket benefits beyond earnings such as those to own health, longevity, and the quality of life. Education externalities can be either positive or negative. But with a few exceptions, the evidence is that they are overwhelmingly positive.

The external benefits of education include education's direct benefits to the development of civic institutions that contribute slowly over long periods of time to the rule of law, democracy, human rights, and political stability. Externalities also include direct benefits to longevity, reduced poverty, lower crime rates, lower public welfare and prison costs, environmental sustainability, contributions to happiness and social capital, and effects from new ideas and adaptation of the results of research.

External benefits of education also include indirect effects of education that are over and above these direct benefits. Indirect effects operate through other variables and feed back over time to increase the private market and nonmarket benefits. Examples include the contribution of education to better governance, political stability, and trade that then indirectly increases growth. Indirect effects aid productivity and set the stage for new rounds of growth in the future, benefiting others and future generations. The reverse side of the coin is that earnings and well-being today are larger due to external social benefits of education from prior generations. The analysis of the dynamic process involved provides a basis for their measurement. External social benefits today continue to set the stage for growth within families and within nations.

The social benefits of education are normally defined to include the total benefits of education, including the external benefits. Therefore, private market and private nonmarket benefits, which are part of this total, must be identified so that they can be distinguished from the externalities. That is, to arrive at the total value of the external social benefits the value of each must be estimated, including the indirect effects!

In what follows, each of these concepts will be further explained as will the method of estimating their value. This is followed by a review of the empirical evidence for each, and estimates of their monetary value. The method of valuing the indirect effects will be explained, followed by considering those studies that have sought to measure

the value of aggregate education externalities without identifying each.

Distinguishing Private Market, Private Nonmarket, and Social Benefits

The external social benefits of education are not the same as nonmarket benefits. Some nonmarket benefits are private benefits, such as better own health, and some public benefits raise market earnings, such as political stability. These concepts and the methods of measuring each need to be distinguished.

The Market Benefits of Education

The market benefits of education are the additions to earnings, or in the aggregate the additions to per capita economic growth, that are due to education. These are increments to earnings above the amount earned by those with less education, measured either by a Mincer earnings function or by the full method that computes a pure internal rate of return (see McMahon, 1991). Increments to aggregate per capita growth are measured using a growth equation based recently on Lucas' (1988: 18) endogenous growth theory. His production function, including investment in human and physical capital, is normally augmented to include trade, the rule of law or political stability, life expectancy, and shocks. The direct market benefits of education that this reveals are illustrated in **Figure 1**, panel A-1. The indirect effects in panel B-1 result from feedback effects as education contributes through trade openness, political stability, and longevity (see McMahon, 2009: ch. 3 and Appendix D). Difference equations that incorporate lagged effects reveal these feedback effects on per capita growth over time.

The Private Nonmarket Benefits of Education

The private nonmarket benefits of education are private benefits to the student and his or her family as shown in panel A-2. (The relevant decision-making unit is regarded here to be the household. Intrafamily benefits to children, therefore, are private benefits. When individual pupils are regarded as independent decision-making units, as in Haveman and Wolfe (1984), benefits to children are externalities and increase the size of the externalities

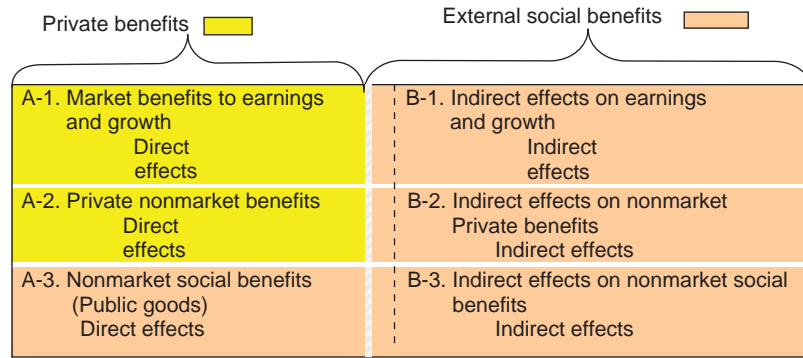


Figure 1 Total benefits of education. From McMahon, W. W. (2006). Education finance policy: Financing the non-market and social benefits. *Journal of Education Finance* 32(2), 264–284, with permission from University of Illinois Press.

reported here. Similarly, the firm is regarded as the relevant decision-making unit so intrafirm spillovers are also private benefits.) They are not external benefits but instead private benefits to own health, spousal health, longevity, lower infant mortality, child health, child education, child cognitive development, lower fertility leading to smaller families, and happiness or well-being. The value of each of these nonmarket private education has a monetary value that is estimated in McMahon (2009) by methods explained below and estimates of their value are just as legitimate as estimates of the value of the earnings due to only education. The nonmarket private benefits are measured using a household production function attributable to Becker (1965). It includes per capita income since market goods are also needed to produce final outcomes. The income term in the regression controls for the market benefits of education, which permits the value of the nonmarket benefits in addition to income to be valued. There are many studies in the literature that do not include this important control, although there are hundreds of studies that do use the household production function rationale, and therefore, include per capita income. Students, families, and policymakers generally have very poor information about these nonmarket benefits and their value, leading to significant market failure (see McMahon, 2009: ch. 4).

Some of these private nonmarket benefits are indirect. For example, education contributes to own health, which in turn, contributes to child education and child health (panel B-2).

The External Social Benefits of Education

The portion of direct social benefits that are public benefits are externalities because they benefit others in the society or in future generations. Public goods in public finance mean that consumption by one person does not reduce the consumption by others. Romer (1990) calls these public goods nonrivalrous. World events analysis offered by public broadcasting (PBS) is a public good,

for example, because consumption of it is not limited as others tune in. It is supported by donations but about 90% of all listeners do not contribute and are free riders. Hence, some contribution from these persons must be obtained through taxes or too little of the public service would be produced. Education is partly a public good, with external social benefits, and partly a private good, generating purely private benefits. Lucas (1988) proves analytically that with education externalities, the growth rate is higher and that this is required for overall economic efficiency (the optimal solution). He suggests that the individual's investment in education is small relative to the average level in the community so the individual takes these community benefits for granted. The individual or family will not invest more in education to receive external social benefits because of this and because they spillover to benefit others and future generations. Therefore, insufficient investment in human capital formation through education is the outcome. (This discussion assumes that government failure is limited, which is sometimes challenged. However, government failure in the United States and Organization for Economic Cooperation and Development (OECD) is relatively small in these democracies compared to massive government failure in providing basic education in authoritarian regimes, for example, in sub-Saharan Africa, Pakistan, Nepal, and Haiti.)

This has important public policy implications and is a second source of market failure. The value of the external benefits of education as a percent of the total benefits is a guide to the extent to which government must finance education through taxes if the system is to be economically efficient. However, efficiency is a term that is thrown around with wild abandon. Economic efficiency includes external efficiency, or how well education serves the needs and demands of society, as well as internal efficiency, and to achieve it requires providing for the production of the external social benefits of education. (Economic efficiency includes production efficiency and exchange efficiency. In the education literature, the former is usually referred to as internal efficiency, and the latter as external efficiency.

Overall economic efficiency requires both). A private for-profit college with no endowment and vocational offerings might be internally efficient, but not externally efficient, with few requirements in history, the humanities, or social sciences that encourage citizenship and public service. At the other extreme, government supported public universities with little tuition as in Europe may also not be very economically efficient. With lower resource recovery from parents and lower resources per student than in the United States, either quality or access must be sacrificed. In the latter case, there may be too few graduates to serve society's needs efficiently. A related policy implication involves how far private versus public financing of higher education should go if higher education is to remain economically efficient. That is, if the external benefits of education are to be realized, then some government investment will be necessary to encourage private families to invest more. Apart from equity, this is the primary economic rationale for public support of the schools, colleges, and public universities. (Equity is an important part of the rationale for public support, but distributional issues are not externalities and so are not considered further here.)

The Indirect Effects from Education

The indirect effects from education are those that operate through intervening variables. They are part of the total market benefits, of the private nonmarket benefits, and of the social nonmarket benefits in panels B-1, B-2, and B-3 above. For example, private earnings and the quality of life are higher in part due to the education of others and the education of prior generations. Looking forward, the education of current graduates benefits the earnings and quality of life of others in future generations. However, future benefits cannot be measured directly. Instead, they can be measured based on the benefits received from the past using simulations with difference equations that have been estimated by regression methods and contain parameters that then generate time paths into the future. Estimates and simulated outcomes for many countries worldwide and for internal US Deep South states are presented in McMahon (2002, 2007), respectively.

Why Are Indirect Effects Externalities?

The indirect effects from education are externalities not only because the effects of the individual's investment in education on intervening variables such as, say, the rule of law, is small and taken for granted, the Lucas (1988) rationale, but also because many of the benefits flow to others in the society and future generations and cannot be captured by the individual, the standard definition of externalities. Therefore, families have no incentive to invest more in education in order to receive either free public benefits or indirect benefits, which are also free.

Yet, these indirect effects are a significant portion of education's total benefits. The best estimate available so far is that they are about 42% of the total benefits of education (McMahon, 2002: ch. 13). Their value, shown in panels B-1, B-2, and B-3 of **Figure 1**, must be added to the value of the direct social benefit externalities shown in panel A-3 to arrive at the total value of the external benefits of education.

The Evidence: Identifying and Valuing the External Social Benefits

Studies quantifying and valuing the external social benefits of education will be discussed first, followed by studies that estimate the value of aggregate education externalities. Once the quantity of external benefits is estimated, there are four methods for estimating their monetary value:

- First is the income-equivalent method developed by Haveman and Wolfe (1984). This method finds the amount of income the typical household would be willing to pay for the increase in the social benefit that results from one more year, or one more level, of education. It is based on the standard economic proposition that individuals typically adjust until the ratios of the marginal products of each good including education to their respective prices are equal. For public benefits, the income-equivalent value is interpreted as the amount of income the typical household is willing to give up or vote for in support of civic institutions.
- Second is the dynamic simulation method in McMahon (2002: ch. 13, 2007). This is needed to value the indirect effects, which then can be added to the value of the direct effects to obtain total education externalities. This is explained following **Table 1** below where it is used, and more extensively in McMahon (2009: ch. IV). Then the difference equation model estimated by regression methods, which contains many control variables is used to generate the total benefits of education by means of dynamic simulations. These methods are widely used in astronomy, physics, meteorology, and macroeconomic forecasting. From this is subtracted the values at each point in time given by a simulation generating only the direct effects to obtain the indirect benefits as a residual. The latter can be expressed as a percentage of the total market benefits, thereby establishing their value.
- The third method is to value the aggregate external benefits of education as in Breton (2008). This is done by estimating a growth equation based on international data to arrive at the total social benefits, public and private, from which are subtracted the private benefits estimated using Mincer earnings functions based on micro data. Aggregate external benefits have also been estimated using panel data for US states or US cities,

Table 1 The direct external social benefits of education

<i>Specific social benefits: Dependent variable</i>	<i>Value of the social benefits of bachelors</i>	<i>Reported coefficient¹ of education</i>	<i>Reported coefficient² of income</i>	<i>Control variables (see footnote³)</i>	<i>Source</i>
Democratization & pol. institutions	1830				
Democratization	994	0.018***	0.372*	In Y, M	McMahon (2002)
Democratization	1726	0.0101*	0.05***	Y,M	McMahon (2009)Appendix.D,HE
Democratization	2771	0.0114***	0.05**	Y, M	ibid, Appendix.D,OECD,Sec.Ed.
Democratization	59 982	0.00917***	0.032	In Y, P, S	Keller (2006) ³
Democratization					Besley, Case (2003) ⁸
Human rights, civic Institutions	2865				
Human rights	2865	0.006*	0.194***	Y,M,D	McMahon (2002)
Political stability	5813				
Political stability	8625	0.0793***	0.00025***	Y,M,D	McMahon (2002; 107)
Political stability	4041	0.0423	4.7E-04***	Y,M,D	McMahon (2009)Appendix.D,HE
Political stability	3001	0.0849**	4.1E-04***		ibid. Appendix.D,OECD Sec.Ed.
Life expectancy	2308				
Positive benefits	3344	0.0504**	2.61E-04***	Y,P	ibid. Appendix.D, OECD, HE Coef.
Negative growth	590			I,T,PS,Y(70)	ibid.Appendix.D,OECD,LEXP
Positive benefits	2452	0.0483***	2.11E-04***	Y,P	ibid. Appendix.D.OECD, Sec Coef.
Negative growth	537			In Y,S,G,PS,t	Barro <i>et al.</i> (1995; 425.(2)) ⁹
Reduced inequality	3110				
Greater opportunity	+ US Only				Leslie & Brinkman (1989) ⁴
Reduced inequality	–(OECD)	0.0015**		S. T.	McMahon (2009)Appendix.D,HE ⁵
Poverty reduction, Sec	3110	–1.41***	–5.6*	Y, P, H	McMahon (2002; 115)Model
Lower crime	5647				
Homicide	719	–15.9***	1447***	In Y,U	McMahon (2002; 144)
All other crime	4928	–974***	22612***	Y, GI, PV	McMahon (2002; 148)
Lower public costs	544				
Lower health costs	544				Muenning (2000, p.28) ⁴
Lower prison costs					Lochner & Moretti (2002) ⁸
Higher Tax Receipts					A market social benefit
Environment: indirect	5609	Effects from less pop. growth & poverty, more democracy			
Cleaner water	136	–3 202**	7.79***	Y,y, P, PV,D	McMahon (2002)
Less air pollution	1482	–1.32**	–1E+00**	Y,S,D,p,PS	McMahon (2002; 137)HE, ⁶
Less deforestation	3991	9.9E-05*	6.7E-07**	Y,P,H	McMahon (2002) ⁶
Social capital					
Social capital	+	Education effects positive			Helliwell & Putnam (1999)
Happiness	+(?)	Effect above \$20 000		Many	Helliwell (2005)neg.effect ⁷
R&D dissemination	++				Non-mkt, apart from growth
Total soc.Benefits	27 726				Direct effect externalities

Significance level: *** = 0.01, ** = 0.05, * = 0.10

Contribution of education to economic growth in eight studies

<i>Growth equation estimates, macro data</i>	<i>In 2007 dollars</i>	<i>Education coefficient</i>	<i>Source</i>
	28 672	7.20E-03***	Barro (1998)
	18 919	0.05*	Barro & Martin (1995; 426)
	13 274	0.005*	Oliva & Rivera-Batiz (2002)
	28 379	0.075***	Keller (2006; 24), globally
	35 568	0.094***	Keller (2006; 30), HE, OECD
	9843	0.047***	McMahon App.D,HE,OECD
	0		Benhabib & Spiegel (1994)
	0		Pritchett (2006)
Average, All studies	16 832		

including a variant that uses the average level of education in the community to represent externalities based on Lucas (1988). This approach has been used by Rauch (1993), Acemoglu and Angrist (2000), Moretti (2003, 2004), and Ciccone and Peri (2006). But these have also been criticized by Lange and Topel (2006), although for reasons that Breton's (2008) newer study avoids.

- The fourth method is to examine the total social accounts in Eisner (1989). They estimate total income and total consumption that include the value of the nonmonetary qualities of life. They include the value of housewives' services, and build on earlier work by John Kendrick. However, the total accounts provide no method for separating the social benefits of education from the private benefits, or for isolating the value of the quality of life that is only due to education. For these reasons, this approach will not be pursued further here.

The Value of the Direct Nonmarket Public Benefits of Education

To place a monetary value on each of the direct external social benefits based on the Haveman and Wolfe (1984: 395) method, a standard proposition in economics is that households will tend to substitute among inputs until they find relatively cost-effective ways of producing each final satisfaction. Better health, for example, can be produced by earning a bachelors degree, which leads to more effective use of time in sustaining health, or it can be produced

by using time in the labor market to produce income and then to purchase doctors' services and drugs, or by some combination. As households balance these alternatives, the ratio of the marginal product of education for achieving health, $MP_{\text{education}}$ below, to its value, P_E , will approximately equal the ratio of the marginal product of income in purchasing medical services, $MP_{X \text{ market}}$, to their price, P_X :

$$\frac{MP_{\text{education}}}{P_E} = \frac{MP_{X(\text{market goods})}}{P_{X(\text{income})}} \quad [1]$$

By cross-multiplying and moving P_E to the left, this value of education for producing better health becomes equal to the ratio of the marginal product of education to the marginal product of market goods (i.e., income), times the marginal cost of the market goods, P_X , needed to produce an equivalent amount of better health. These marginal products, $MP_{\text{education}}$ and MP_X , are the regression coefficients for education and for income in a regression where the dependent variable is democratization, human rights, or some other external benefit shown in **Table 1** below. The education and income coefficients are shown in columns 3 and 4, control variables that are significant in are shown in column 5, and the source of each regression is shown in column 6, so the reader can interpret each estimate.

The estimates of the value of education in **Table 1** apply to the annual benefits generated by a college bachelor's degree, which is also interpreted as one more

³ Definitions of Control Variables: (For data sources see article or book cited)

Y = GDP Per Capita

M = Military Expenditure as % of Govt Budget

P = Primary Gross Enrollment Rate lag 10 years

S = Sec. Gross Enrollment Rate lagged 10 years

H = Higher Education Gross Enrollment Rate

D = Democratization, Freedom House (2007)

G = Government Consumption as % of GDP

U = Unemployment Rate lagged two years

GI = GINI Coefficient: inequality in the distribution of income

I = Investment in Phys.Cap.as % of GDP

T = Trade Openness; exports + imports as % of GDP

PS = Political Stability, International Risk Guide

Y(70) = Initial GDP per capita in 1970

lnY = log of GNP Per Capita

PV = Poverty Rate

p = Population Growth Rate

⁴ No regression in the survey.

⁵ Not included in average because income coefficient is not significant.

⁶ Not included in average because education coefficient is not significant.

⁷ Helliwell's income and other controls contribute to this. See McMahon (2009, Ch. IV).

⁸ To get the effects of only higher education when only a secondary education coefficient is available and when there is no control for higher education, the assumptions are made that this secondary education coefficient captures both, and that four years.

⁹ No income variable in the paper.

¹⁰ Government consumption (reflecting social security and aging) as a percent of GDP.

Source: Adapted from McMahon, W. W. (2009). Higher Learning, Greater Good: The Private and Social Benefits of Higher Education. Table 5.4. © 2009 The Johns Hopkins University Press. Reprinted with permission of The Johns Hopkins University Press.

¹ Gross Enrollment Rate includes replacement investment (65% of total)

² GDP Per Capita

percentage point on the per capita enrolment rate. Attention is given to this in part because, in the past, some have asserted that externalities at the college level are negligible since college graduates earn so much. But this assertion avoids analyzing the specific direct external benefits shown in **Table 1**, as well as the research on these, and also ignores consideration of the indirect effects from education that increase earnings. If most of the benefits listed have a roughly linear effect for one more year of schooling, whether it be at the high-school or college level, then the values in **Table 1** could be interpreted alternatively as also an approximation of the external benefits from high-school graduation. The valuation must relate to some level of education if they are to be relevant to policy. To include comparable tables for primary and junior secondary education and for masters and PhD levels would expand the scope beyond that of a short article.

To illustrate how these values are estimated, the ratio of the education coefficient to the income coefficient in columns 3 and 4 of **Table 1** is first calculated. This is the ratio of the marginal product of higher education in producing refinements in democracy, D , in the typical OECD country, for example, which is the $MP_{\text{education}}$ in eqn [1] above or 0.0101 empirically in **Table 1** below, to the marginal product of \$1 of income, $MP_{X(\text{income})}$ or 0.05, in producing the same amount of democratization. It is this ratio that is used to obtain the income-equivalent values of the net effects from more education. So using household income as the basis for the imputation, the average annual increase in the Freedom House (2007) index for D achieved in the OECD countries from 1975 through 2004 of 0.0017 is taken to be the amount of D to be produced, either through increased education enrolments or through purchases or votes in financial support for civic institutions. Using the income coefficient in the regression, it can be estimated how much income it would take to produce the 0.0017 typical increase in D based on past experience. This turns out to cost \$55 per capita. Taking this therefore to be P_X , the price or marginal cost of achieving the typical improvement in D in the OECD countries, the income equivalent value of achieving this same outcome through increased education is estimated to be \$1726 as shown. This is fairly close to the \$1830 average of the three studies that contain education and income coefficients that are both significant. The work of skeptics is considered later.

The imputations for the value of education in improving human rights, political stability, and almost all of the other social benefit outcomes in **Table 1** are made on the same basis. That is, the value of the social benefit of education is the income equivalent of achieving either by more education or by spending more income, the average annual improvement in each of the other indices in the OECD from 1975 through 2004. The details for

standardizing the studies in **Table 1** to convert everything to 2007 dollars and to make them comparable in other respects as well as further details for each imputation appear in McMahon (2009, Appendix E).

It must be stressed that these are the direct effects, calculated directly from the regressions, which usually apply to 5-year periods, and do not include the indirect effects that feed back and build up over time to make total education externalities larger. These are considered further below. The control variables that often remove these indirect effects that are listed in column 4 and footnote 3 of **Table 1** are only those that are significant at the 0.05 level. Other control variables thought to influence each dependent variable are often mentioned, but the data generally will not support their significance. There are also many studies not included in **Table 1** simply because both education and income are not in the regression, or if they are, both coefficients are not significant. Another constraint is that basic research does not exist in appropriate form to value all social benefits, so there are gaps and the total value in **Table 1** is conservative. Most studies have either checked for simultaneity or used instrumental variables or two-stage least squares methods. Nevertheless, this is the first effort to standardize the many studies and estimate the value of individual social benefit externalities comprehensively. Total precision is not claimed, and as gaps in the research pointed out here are filled further refinements will become possible. (All regressions control for heteroscedasticity. Those not using IV have been checked for simultaneous bias. For those from McMahon (2002), alternative specifications are reported there.)

Evidence Concerning Specific External Social Benefits

The empirical evidence on direct specific social benefits follows, with both indirect effects and efforts to estimate aggregate education externalities considered later.

Democratization

Democratization is the degree of development of political institutions at the national, state, and local levels as measured by the Freedom House (2007) index. Democratization worldwide is empirically determined as shown in line 1 of **Table 1** primarily by access to education, growing per capita income, and lower military expenditure as a fraction of public budgets (McMahon, 2002: 97–101; Diamond, 1992; Clague *et al.*, 1996). Other factors are empirically less significant. Clague *et al.* find that an additional variable for Muslim religion is negatively related to democracy. But it becomes insignificant whenever literacy is included, indicating that completion of basic education displaces Muslim's role. Secondary education enrolments are especially important, although for

the most developed OECD member nations, the effect is smaller (see rows 2 and 3 of **Table 1**). The reason is relative homogeneity; the variation in the democracy index is very small, the variation in enrolment rates is also small, and the result is statistically less dependable. Using worldwide data, Keller (2006) controls for per capita income and secondary education and finds investment in higher education lagged 10 years to be a highly significant determinant of democratization ($t = 3.22$). However, her income coefficient is not significant. In contrast to her result showing a large education impact, Acemoglu *et al.* (2005a) find no effects whatsoever from education on democratization. The methodology of their study, however, is not based on a dynamic view of the process since it includes no lags in education's impacts, includes year dummies which eliminate most of the effects of technology embodied in human capital, and uses school achievement that also eliminates effects from new technology embodied in replacement investment in human capital. It also uses lagged democracy as an explanatory variable that eliminates what little variation is left in 5-year movements in the very-slow-moving process of democratization. Therefore, with these problems, their study is not averaged in. The Beasley and Case (2003) study cannot be included because there is no control for income.

The best estimate therefore averaging the first three studies is the \$1830 value per year of a bachelor's degree to the development and operation of civic institutions shown on line 1 in **Table 1**. The evolution of democratic institutions is a long, slow process. The most difficult context for showing the relation of education to democracy is within the relatively homogeneous OECD or among states within the United States both because of this and because of the spatial equilibrium involved.

There are also contributions of education to behaviors such as voting that are important to democracy. The empirical evidence on this is strong and strengthens the case, but the value of these behaviors cannot be added to the value of education's contribution to overall democratization measures without double counting. Therefore, the important effects of education on contributing behaviors are not included in **Table 1**. To mention a few, those with more education voluntarily give twice as much of their time and twice as much of their money at each income level to civic institutions as do those with a high-school education or less, where only 12% give (Hodgkinson and Weitzman, 1988; NCES, 2005). There are also large positive effects on voter participation, support for free speech, and the quality of civic participation, the latter as indicated by the frequency of news readership (Dee, 2004). Tastes are shifted away from drag racing, dog fighting, and TV game shows, and toward world affairs. UK college graduates have also been shown in tracer studies of graduates to exhibit greater racial tolerance, less cynicism, and less unquestioned support

for authority than those with only high-school education (Byner *et al.*, 2003). These graduates engage in more life-long learning about public affairs' importance to good citizenship and social capital. More education also increases support for democracy in Muslim countries based on micro data by Shafiq (2009).

Human rights

Human rights are a public good that is very important to the quality of life. As measured by Freedom House's (2007) index of civil rights, it includes freedom of the press, freedom of speech, freedom of assembly, the legal protections of *habeas corpus*, trial by jury, freedom from unlawful searches, freedom from unlawful incarceration, protections from torture, reasonable equality of opportunity, and limited corruption. All require an effective criminal justice system and civic institutions. In worldwide data, human rights depend significantly on secondary education, higher per capita income, lower military expenditure as a percent of the government's budget, and on democracy (McMahon, 2002: 103). The highly significant effects from democratization mean that human rights are important benefits of democracy.

The value of education's contributions to human rights is estimated to be about \$2865 per year for each bachelor's (or high-school graduate) in **Table 1**. Other studies of education's contribution to human rights are very limited and also do not control for income (e.g., IHEP, 2005). Contributing behaviors discussed above also contribute to human rights. But their value cannot be added to the total without overlap.

Political stability

Political stability is measured by the comprehensive index from the International Country Risk Guide (2007) that includes both political and economic risk. Stability is generally agreed to be an important determinant of economic growth (Barro and Sala-I-Martin, 1995: 426; McMahon, 2002; Oliva and Rivera-Batiz, 2002). Political stability is significantly dependent on education, per capita income, lower military expenditure as a percent of government budgets, and democracy (McMahon, 2002: 105–110). However, although growth depends on political stability, stability is not 100% dependent on democracy in the regressions. Singapore, China, and Dubai, for example, are all authoritarian and yet growing rapidly. However, as growth continues and education spreads, the pattern may follow South Korea and Taiwan. Both were authoritarian after World War II; extended basic education and promoted growth; and both became full democracies in 1980. In contrast, Pakistan grossly underinvested in basic education to the point that a majority of the labor force and 95% of rural women are illiterate. From 1980 to 2008, it remained a military dictatorship, with low per capita income.

The value of each bachelor's contribution to political stability after controlling for income is estimated to be \$5813 per year in 2007 dollars in **Table 1**.

Life expectancy

Increased life expectancy is a positive private benefit of education in many studies (McMahon, 2009: ch. 4). However, the contribution of education to longevity in aggregate data can also be regarded as a social benefit. This is most obvious in the poor countries where life expectancy is very low, and many die before reaching the most productive years in their life cycle. The value of the net effects from more of education is estimated to be \$3110 per year in **Table 1**. But from this has been subtracted the negative effect on economic growth of increasing longevity in the OECD countries, estimated to be the average of the \$590 and \$539 estimates shown. The latter is a negative externality. The \$537 estimate uses Barro's government consumption, a proxy for life expectancy because it reflects the social security and health expenditures for an aging population. (The per year value is estimated by spreading out the value of the number of years of increased life expectancy over an average of 65 years remaining in the life cycle.)

Reduced inequality and lower poverty

Inequality in the distribution of income is increasing dramatically in the United States and in other OECD countries. It is related to inequality earlier in access to education as shown by Psacharopoulos. More recent studies also show that remarkable increases in earnings inequality in the United States, United Kingdom, and OECD since 1980 is linked with human capital skills in dealing with new technology (Faggio *et al.*, 2008).

Reducing high-school dropout rates also reduces inequality (Levin, 2006: 9). In higher education, Leslie and Brinkman (1988) conclude that increased access in the United States reduces inequality except in states where the tax systems are regressive as in Florida and Mississippi. In the OECD countries, however, higher education contributes to greater inequality as shown in **Table 1**. Admissions are restrictive in many European universities, 2-year associate degrees are far less widespread, and need-based aid is less available. Therefore, although increased access to higher education reduces inequality in most states in the United States, it probably does not in Europe because of policies related to less need-based aid and proportionally fewer in associate degree programs.

Poverty is reduced, however, by economic growth, increased high-school completion, and increased access to higher education. The value of this direct effect is estimated to be \$3110 per year for completion of college degrees in **Table 1**.

Lower crime rates

The effect of education in reducing crime rates and criminal justice system costs has received more attention than other externalities. Witte's (1997) review reveals that further education of those who have started on a life of crime is of limited effectiveness, whereas reducing high-school dropout rates and increasing 2-year college enrollments that cause young males to be under supervision in school (and in employment later) are effective. The value of high-school or college graduation in reducing murder rates (violent crime) and property crime (all other crime) after controlling for per capita income, lagged unemployment, inequality, and poverty is estimated to be \$719 per year per graduate for lower murder rates, and \$4928 per year per graduate for all of the many other kinds of crime. Higher education contributes to white-collar crime, a negative externality, but this has been netted out against education's positive benefits in reducing overall crime. Lochner and Moretti (2002) do not control for income.

Lower public health, welfare, and prison costs

Education reduces public welfare costs. Only 0.5% of all college graduates receive public assistance or Aid to Families with Dependent Children (AFDC) compared to 5.6% of all high-school graduates. The percentage of college graduates on welfare is 16.6 percentage points lower than those that stop short of high-school graduation (NCES, 1992). Muennig (2000: 28) estimates the lifetime savings in public costs of uninsured populations to be \$11077. When converted to 2007 dollars, his present value (he uses a 3.5% discount rate) is \$544 per year over 45 years after graduation as shown in **Table 1**. (His present value for each student completing 2 years of college is \$6317.) Muennig's estimate means that welfare costs to state governments could be reduced by about two-thirds if all students completed high school. They would be reduced by 91% if all high-school graduates completed a bachelor's degree. Education policy would dramatically augment welfare reform that has sought to move welfare recipients into often unskilled work.

The decrease in criminal justice system costs, and larger state and federal tax revenues from increased high-school completion are also found to be enormous in studies reported by Levin (2006). This evidence of external benefits from public support of schooling in poor districts benefits residents of wealthier districts by lowering their public welfare, prison, and health care system tax costs while also increasing tax revenue from others.

Increased tax receipts

Education dramatically increases income, sales, and property tax receipts in estimates by Rouse (see Levin, 2006). They are external social benefits, but not included in **Table 1**, which is confined to nonmarket benefits. Taxes

are already included in standard social rates of return based on earnings.

Environmental benefits

Many effects of education on the environment are indirect. There are direct effects of secondary education in reducing water pollution, but the indirect effects of education through lower poverty and increased democratization are greater. Together, they offset the adverse effect of income growth, so that education's direct effects plus indirect effects on balance reduce water pollution. The per graduate annual value of education's contribution to cleaner water is estimated to be \$136 in **Table 1**.

Similarly, although growth has adverse effects, there are net positive contributions of education to reduced air pollution and slower deforestation. The value of education's annual contribution to reduced air pollution, and probably, global warming is estimated to be \$1482 in **Table 1**. Education's contribution to reforestation and wildlife habitat is estimated to be \$3991. However, the coefficient for education's net effect has a lower level of significance, so this estimate may be too high. Most of education's net effects after controlling for income growth on the environment are indirect through lower fertility, slower population growth, and more democratization.

Happiness and social capital

There are many studies that show that education has a positive effect on happiness, known earlier as subjective well-being. But few of these control adequately for per capita income. Happiness not only is a private benefit of education, but it also is known to benefit others by contributing to greater social cohesion and social capital (Helliwell, 2005). The latter are external social benefits.

Happiness is now measured cardinally based on brain waves. It is known to increase with income up to about \$20 000 per capita, or \$80 000 for a family of four, but not beyond that (Layard, 2006). Education contributes to earnings and therefore indirectly to happiness up to this point. But beyond that, the basic research on education's effects on happiness and social capital at family incomes above \$80 000, where the income effect flattens out, is missing.

However, some informal inferences can be made. This is because education is well known to contribute to most of the major sources of happiness, and hence to social capital. These sources include genes; some persons are born to be happy, and others to be unhappy. However, beyond this, happiness increases with the better selection of a mate, less unemployment, lower divorce rates, better civic institutions, better human rights, less crime, better health, and a better work environment, all of which are positive functions of the level of education (Layard, 2006). The bottom line, however, is that there must be further research, especially on education's effects on happiness

and social capital through these intervening variables and also at family incomes over \$80 000.

Dissemination of technology

Probably the largest external social benefit of higher education is left to last. It is the benefit to the broader society from the embodiment in graduate students and undergraduates of new knowledge created by research and development (R&D) in all academic fields and disseminated by them as they take teaching jobs in other institutions or jobs engaged in creating new ideas (Jones and Romer, 2009). Utilization of most modern knowledge and technology in production is very complex, requires advanced education, and has little effect on the economy unless the means to use it are disseminated. Knowledge from throughout the world is accessed by faculty engaged in research and embodied in students at masters, PhD, professional, and undergraduate levels, who later disseminate it to other colleges, government agencies, and firms. This basic complementarity between human capital and R&D is stressed by Griliches (2000: 88) and fundamental to Romer's (1990) endogenous growth model. Unfortunately, it is too often overlooked in studies of the relation of education to growth. However, an important fraction of the external social benefits from research are also external social benefits from higher education. It is not possible to place a value on these, but the value of the knowledge embodied in the number and skill levels of postgraduate students may be a promising approach.

The Value of External Indirect Effects

The total value of the direct external social benefits of education in **Table 1** is estimated to be \$27 726 per year. These direct benefits per year per degree are not the total benefits from the average level of education in the United States, and also do not include the indirect effects.

The indirect effects of education operate through other variables to set the stage for growth of earnings and non-market benefits later. To estimate these, a dynamic simulation is first run using regression equations that control for other significant variables to generate the total benefits of education over the rest of the life cycle, say 40 years. Then a second simulation is run with the indirect effects suppressed, giving only the direct effects (i.e., the cross partial derivatives are set to zero). These direct effects then are subtracted at each point in time from the total benefits leaving only the indirect effects. Expressing the indirect effects as a percent of the market benefits (and as a percent of the value of the private nonmarket benefits as in McMahon, 2009: ch. 4) which establishes their value. Based on McMahon (2002: ch. 13, pp 228–242) where this is done, the value of the indirect benefits is about 42% of the value of the market benefits – 42% of the market benefits is \$10 540 in 2007 dollars (panel B-1), and 42% of the value of

the private nonmarket benefits is \$15 993 (panel B-2). Therefore, the total value of the market plus nonmarket indirect benefits is estimated to be \$26 533 per year.

The Total Value of Education's External Benefits

Adding the \$27 726 of direct public-good-type benefits from **Table 1** to the \$26 533 of indirect benefits (part of which are future values) gives a total estimated value of all of education's external benefits of \$54 259. This addition is possible because all regressions considered control for per capita income so the direct public good benefits do not overlap the market benefits. Indirect benefits, however, do overlap private benefits since some private benefits are due to the education of others and the education of earlier generations. (It is assumed that the indirect social benefits (panel B-3) are included in the other indirect effects.)

External social benefits, as a percent of the total benefits from education, is a matter of interest to public policy. If the total benefits are \$90 902 per year for a bachelor's degree in 2007 dollars, composed of \$25 096 in earnings increments (the average over the life cycle for males and females based on Census data from McMahon (2009: ch. 3)), \$38 080 in private nonmarket benefits (McMahon, 2009: ch. 4), and \$27 726 in direct social benefits from **Table 1**, then the external benefits of \$54 259 are 59% of this total. If the indirect effects simulated over a 40-year period were discounted back to their present value, they would be smaller. However, the total net effects of education would also have to be discounted back to the present. Therefore, the 59% would not be much affected.

Studies of Aggregate Externalities

The third method of estimating the value of the external benefits through studies of aggregate education externalities involves macro regressions of education's effects on growth. Either a social rate of return is estimated using a macro growth equation and the private Mincer return estimated from micro earnings data is subtracted to obtain the net externalities, or else a macro growth equation is estimated that contains the average education level in the community that represents the externalities. Obviously, if education is found to contribute nothing to growth in the macro growth equation, there are no externalities. Therefore, either of these aggregate approaches involves getting into the education and growth literature.

The results of various growth equation estimates are summarized in **Table 1** (continued). Benhabib and Spiegel (1994), Acemoglu *et al.* (2005a), and Pritchett

(2000, 2006) cannot find that education contributes anything to growth, and there are therefore no externalities. The details of these studies are reviewed elsewhere. (See McMahon (2009: chs. 3 and 4) and the cross-referenced growth entry below.) However, briefly, there are many other studies that show significant contributions of education to growth, six of which are summarized in the continuation to **Table 1**. The three zero-return studies mentioned in common use a static conceptual view, eliminating education's role in the dissemination of new technologies. They do this by using time dummies, using school attainment measures that exclude technology embodied in human capital replacement investment, which is about 68% of the total, and using controls eliminating education's indirect effects through democracy and political stability. The further severe problems with the Acemoglu *et al.* study were discussed above under democratization. Lange and Topel (2006: 479) in their thorough recent review conclude that there is "little evidence in favor of *negative* external returns to education," and that recent studies "cast doubt on the earlier studies by Benhabib and Spiegel and Pritchett who argue for small or even zero aggregate returns to schooling."

The aggregate approach to externalities has most frequently been applied using differences among US states or US cities. But questions can be raised as to whether externalities can ever be found using this type of data. There are two problems: there is little variation because of homogeneity within countries in slow-moving variables such as democratization, human rights, and stability. Second, Lange and Topel (2006: 505) conclude "that this type of evidence is inherently flawed as it does not sufficiently account for endogeneity issues implied by the spatial equilibrium in Rauch, 1993; Acemoglu and Angrist, 2000; and Moretti, 2003." In other words, worker mobility among localities responds to exogenously imposed differences in demand and earnings differences are not just due to prior human capital investments.

Studies based on differences among nations such as Breton (2008) or those as in **Table 1** are not subject to these problems. There is less worker mobility among nations and greater differences in democratization, human rights, political stability, and education since some countries are at earlier development stages. So with proper controls for other factors there is enough variation to estimate statistically meaningful coefficients. Using cross-country data, Topel (1999) and Heckman and Klenow (1997) estimate aggregate education externalities by first estimating a growth equation and then subtracting the private returns to schooling based on Mincer regressions from individual earnings data. Topel's estimate is a 23% social rate based only on market returns and Heckman and Klenow's is 30%. After the average private Mincer return of 8.3% is subtracted, this leaves 14.7% and 21.7%, respectively, for the external benefit. This offers useful insight but has two

problems. It is biased upward because it does not control other things that affect growth; and it is biased downward because it does not include the externalities related to nonmarket private and social benefits. They seek to correct the former by introducing life expectancy and time dummies as proxies for technology. The problems are that life expectancy subtracts from growth in the OECD countries because an aging population shrinks the labor force and increases social security costs. Moreover, time dummies again imply a static interpretation that rejects the role of education in disseminating technology. With these very debatable controls, no externalities are left.

In his later eclectic review of world data, Pritchett (2006) considers OECD data. He does not recognize this role of life expectancy, control for oil shocks, or include the role of technology through replacement investment in human capital. He again finds no positive contributions of education to growth and hence no externalities. Larger positive and roughly comparable effects of education on growth are found by Keller (2006), Barro (1998), Barro and Sala-I-Martin (1995: 426), Oliva and Rivera-Batiz (2002), and McMahon (2009: Appendix D), all in **Table 1** (Continued), and also by McMahon (1998), the World Bank (1993), and many others.

In a recent thorough study Breton (2008) using international data that avoids twin problems of spatial equilibrium and of homogeneity within a single country that plagued earlier studies estimates the market-based social rate of return to be 22% in the United States and 24% in the United Kingdom. He uses instrumental variables to avoid simultaneous bias, with the percent of the population that is Protestant as the somewhat debatable instrument for education. He uses cumulative investment expenditure on education to reflect quality and not just quantity that dominates attainment measures. Subtracting the private Mincer return based on micro earnings data, he finds the residual pure external social benefit rate to be 10.6% in the United States and 16.2% in the United Kingdom, close to Topel's 14.7% and Heckman and Klenow's 21% above. However, Breton's study is also confined to only earnings benefits. So if an imputation is made for the private nonmarket benefits, then his external social benefits as a percent of the total benefits is 57% for the United States and United Kingdom. This is almost identical to the 59% external benefits as a percent of the total that we have independently obtained above through analysis of the specific external social benefits and indirect effects.

Implications

With the cautions mentioned earlier, these estimates suggest that public support of education should be a little over 50% of the total costs on average to provide for these

external social benefits. Total investment in human capital formation includes foregone earnings costs borne primarily by parents as they support room and board costs and not just institutional costs. At the high school level, public investment is very close to this, a little over 50% on average, although the public share is lower for private schools and higher for public schools (where foregone earnings costs are less than half the total). At higher education levels, public support is less than 50% on average. This suggests overall that the trend toward privatization in financing higher education through higher tuition and fees may have gone far enough if overall economic efficiency that includes external benefits is to be maintained. Of course, there is wide variation. Public support of community colleges is above 50% when Pell Grants and Stafford Loans are considered. Some private universities have large endowments support external social benefits. But for most 4-year public colleges and universities and for the less-well-endowed private institutions, public support plus endowment support has fallen significantly below 50% in recent years.

In conclusion, education externalities are not the same as nonmarket benefits; some externalities raise monetary market benefits, and some nonmarket benefits are purely private and not social externalities. Many specific non-market external social benefit externalities are poorly understood and undervalued. When corrected social rates of return are calculated, they are much higher than conventional rates based only on earnings benefits. Poor information about what they are and their value is a significant source of market failure and underinvestment in education, both by households and by governments.

From a dynamic perspective, the external social benefits from education set the stage for each new round of growth and broader development within families and within nations over time. The process is cumulative so that the short-term effects of education, including the external benefits, are smaller. But they grow as interactions and feedbacks occur to the point that they can be extremely important to the future of both individual families and of nations.

See also: Education and Civic Engagement; Education and Crime; Education and Economic Growth; Education Production Functions: Concepts; Human Capital; The External Benefits of Education.

Bibliography

- Acemoglu, D., Johnson, S., Robinson, J., and Yared, P. (2005a). From education to democracy? *American Economic Review* **95**(2), 44–49.
- Becker, G. S. (1965). A theory of the allocation of time. *Economic Journal* **75**(299), 493–517.
- Breton, T. R. (2008). Schooling and national income: How large are the externalities? *Education Economics* **2008**, 1–26.

- Bynner, J., Dolton, P., Feinstein, L., Makepeace, G., Malmberg, L., and Woods, L. (2003). *Revisiting the Benefits of Higher Education*. Bedford Group for Lifecourse Statistical Studies, Institute of Education, London: University of London.
- Clague, C., Gleason, S., and Krack, S. (1996). *Determinants of Lasting Democracy in Poor Countries*, Mimeo, pp 35. San Diego: Department of Economics, University of California.
- Dee, T. (2004). Are there civic returns to education? *Journal of Public Economics* **88**, 1697–1720.
- Diamond, L. (1992). Economic development and democracy. *American Behavioral Scientist* **35**, 450–499.
- Eisner, R. (1989). *The Total Incomes System of Accounts*. Chicago, IL: University of Chicago Press.
- Faggio, G., Salvanas, K. G., and Nan Reenen, J. (2008). The evolution of Inequality in Productivity and Wages: Panel Data Evidence. *NBER Working Paper No. W13351*. New York: National Bureau of Economic Research.
- Freedom House (2007). *Freedom in the World*. New York: Freedom House.
- Griliches, Z. (2000). *R&D, Education and Productivity; A Retrospective*. Cambridge: Harvard University Press.
- Haveman, R. and Wolfe, B. (1984). Schooling and economic well being: The role of non-market effects. *Journal of Human Resources* **19**(3), 377–407.
- Heckman, J. and Klenow, P. (1997). *Human Capital Policy*. Working Paper. Chicago: Economics Department, University of Chicago.
- Helliwell, J. (2005). Well being, social capital, and public policy: What's new? NBER Working Paper No. W11807. New York: National Bureau of Economic Research.
- Hodginkson, V. and Weitzman, M. (1988). *Giving and Volunteering in the United States: Findings from a National Survey*. Washington: Independent Sector.
- IHEP (2005). *The Investment Payoff, a 50-State Analysis of the Public and Private Benefits of Higher Education*. Washington, DC: Institute for Higher Education Policy.
- Jones, C. I. and Romer, P. (2009). The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital. *NBER Working Paper 15094*. Cambridge MA: National Bureau of Economic Research.
- Lange, F. and Topel, R. (2006). The social value of education and human capital. In Hanushek, E. and Welch, F. (eds.) *Handbook of the Economics of Education*, pp 459–510. Amsterdam: Elsevier.
- Layard, R. (2006). Happiness and public policy: a challenge to the profession. *Economic Journal* **116**(March), C24–C33.
- Leslie, L. and Brinkman, P. (1988). *The Economic Value of Higher Education*. Phoenix, AZ: Oryx Press.
- Levin, H. (2006). The social costs of an inadequate education. *Symposium Summary*, 24–26 October, 2005. New York: Teachers College, Columbia University http://www.tc.columbia.edu/i/a/3082_socialcostsofinadequateEducation.pdf (accessed June 2009).
- Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics* **22**(1), 3–42.
- McMahon, W. W. (1991). Relative returns to human and physical capital in the US and efficient investment strategies. *Economics of Education Review* **10**(4), 283–296.
- McMahon, W. W. (1998). Education and Growth in East Asia. *Economics of Education Review* **17**(2), 159–172.
- McMahon, W. W. (2002). *Education and Development: Measuring the Social Benefits*. New York: Oxford University Press.
- McMahon, W. W. (2006). Education finance policy: Financing the non-market and social benefits. *Journal of Education Finance* **32**(2), 264–284.
- McMahon, W. W. (2007). An analysis of education externalities with applications to development in the Deep South. *Contemporary Economic Policy* **25**(3), 459–482.
- McMahon, W. W. (2009). *Higher Learning, Greater Good: The Private and Social Benefits of Higher Education*. Baltimore, MD: Johns Hopkins University Press.
- Moretti, E. (2002). Estimating the social returns to higher education: Evidence from longitudinal and repeated cross section data. *NBER Working Paper No. 9108*. Cambridge, MA: National Bureau of Economic Research.
- NCES (1992). *Education Pays*. USDE: National Center for Education Statistics. Washington, DC: USDE. http://www.collegeboard.com/prod_downloads/press/cost06/education_pays_06.pdf.
- NCES (2005). *Education Pays*. National Center for Education Statistics. Washington, DC: USDE. http://www.collegeboard.com/prod_downloads/press/cost06/education_pays_06.pdf.
- Pritchett, L. (2006). Does learning to add up add up? Returns to schooling in aggregate data. In Hanushek, E. and Welch, F. (eds.) *Handbook of the Economics of Education*, ch. 11, pp 637–695. Amsterdam: North-Holland/Elsevier.
- Romer, P. (1990). Endogenous technical change. *Journal of Political Economy* **98**(5), S71–S98.
- Shafiq, M. N. (2009). Do Education and Income Affect Support for Democracy in Muslim Countries? *Economics of Education Review*. accepted for publication.
- Topel, R. (1999). Labor markets and economic growth. In Ashenfelter, O. and Card, D. (eds.) *Handbook of Labor Economics*, vol. 3C, pp 2942–2984. Amsterdam: Elsevier.
- Witte, A. D. (1997). The social benefits of education: Crime. In Behrman and Stacey (eds.) *The Social Benefits of Education*, pp 219–246. Ann Arbor: University of Michigan Press.
- World Bank (1993). *The east Asian Miracle: Economic Growth and Public*. Policy Research Paper. Policy Research Department. Washington, DC: The World Bank.

Further Reading

- McMahon, W. W. (2004). The social and externality benefits of education. In Johnes, G. and Johnes, J. (eds.) *The International Handbook of the Economics of Education*. London: Edward Elgar pp 211–259.
- McMahon, W. W. (2007). An analysis of education externalities with applications to development in the Deep South. *Contemporary Economic Policy* **25**(3), 459–482.
- McMahon, Walter W. (2009). *Higher Learning, Greater Good: The Private and Social Benefits of Higher Education*. Baltimore: Johns Hopkins University Press.
- Wolfe, B. L. and Haveman, R. H. (2003). Social and non-market benefits from education in an advanced economy. In Kodrzycki, Y. (ed.) *Education in the 21st Century*, pp 97–131. Boston, MA: Federal Reserve Bank of Boston. Available at [NCES.ed.gov/pubs2003/digest/d412html](http://nces.ed.gov/pubs2003/digest/d412html).

Relevant Website

<http://netfiles.uiuc.edu/wcmahon/www> – The EXCEL spreadsheets used to estimate the value of the non-market social and private benefits of education can be downloaded.

ECONOMICS OF EDUCATION – LABOR MARKETS, EDUCATION AND EARNINGS

Contents

Education and Inequality
Human Capital
Race Earnings Differentials
Returns to Education in Developed Countries
Returns to Education in Developing Countries
School Quality and Earnings
Signaling in the Labor Market
The Economics of High School Dropouts

Education and Inequality*

J Blanden, University of Surrey, Guildford, UK

S Machin, University College London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Human capital theory – Investment in higher levels of educational attainment improves labor market outcomes through improved productivity.

Private returns to education – The benefits accruing to individuals from obtaining additional education. These include higher earnings and reduced likelihood of unemployment. 'Returns' to education are causal impacts rather than associations.

Skill-biased technical change – Changes in technology lead to increases in the productivity which differentially favors high-skilled workers. This is in part responsible for increases in inequality as earnings for high-skilled workers rise more quickly than the earnings of lower-skilled workers.

Social returns to education – The benefits of an individual obtaining additional education for other members of society. These can include higher productivity, reduced impacts from crime, a healthier workforce, and more civic engagement.

Introduction

This article considers the interplay between education and inequality. We have structured this into two-related-parts. First, we consider the extent to which educational achievement is unequally spread through the population. Second, we look at the implications of inequalities in education for economic and social well-being. The analysis we present studies several dimensions of inequality in educational experiences and achievement: social background; ethnicity and immigrant status; and gender. In doing so, we gauge the magnitudes of these inequalities at different educational levels in the main industrialized countries and comment, as far as possible, on their changes over time.

Attainment gaps are so important because education is crucial in determining individuals' prospects and life chances. Even the attainment of low-level qualifications can substantially reduce the probability of unemployment or worklessness (the major cause of poverty), and higher level qualifications increase individuals' earning power. The second part of the article reviews the size of these returns to education, and once again considers their evolution over time. We also briefly note the importance of education for individuals' noneconomic well-being, low education levels are closely linked to poorer health outcomes, a higher likelihood of committing crime, and even an early death.

* This article builds upon, and in parts reproduces, some of our related work in this area, including Machin (2005, 2008a, b).

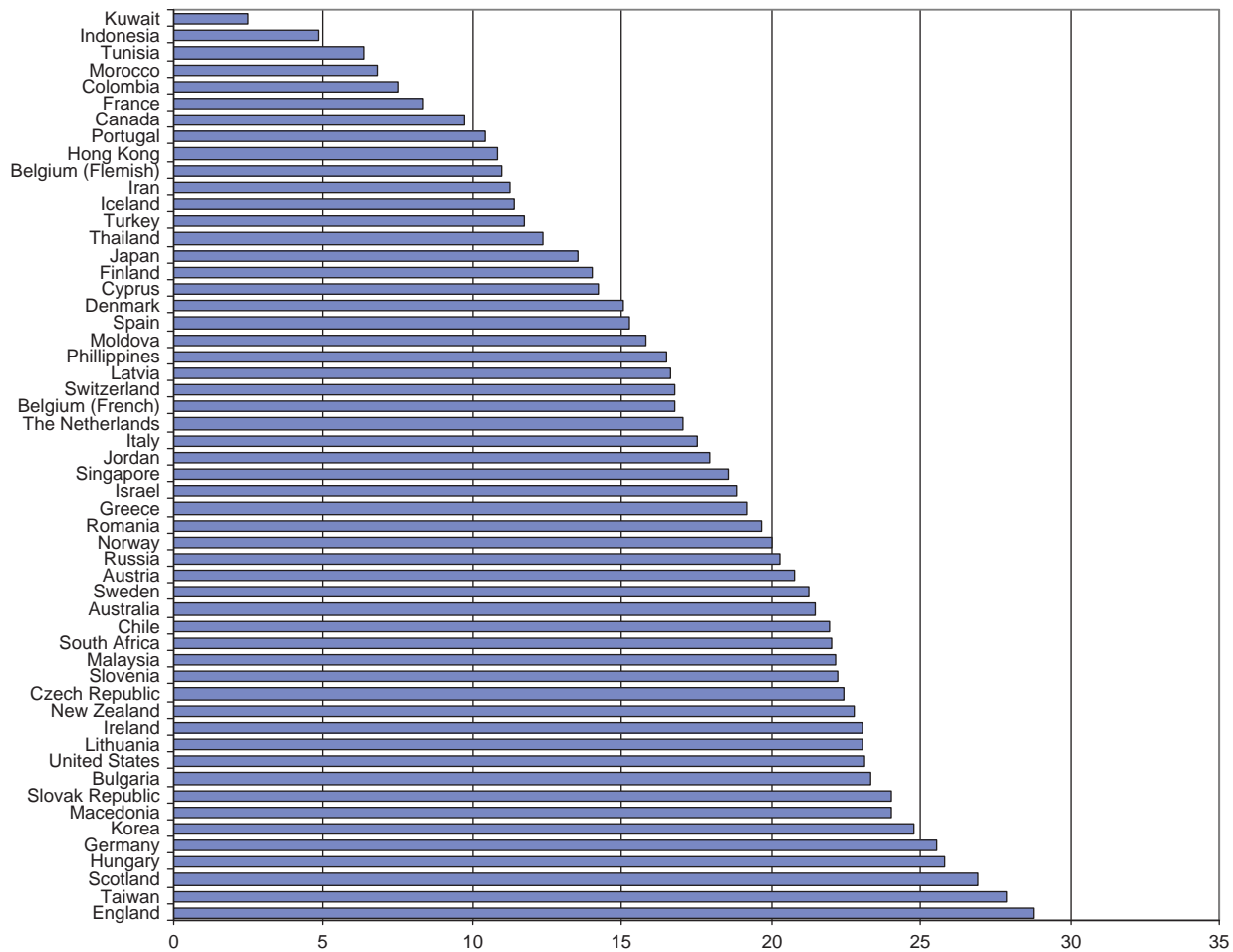


Figure 1 Estimated effects of family background on students' test scores across countries. Family background effects are based on reported measures of the number of books at home; test scores are average mathematics and science scores from TIMSS. The family background effects are estimated from statistical regressions explaining standardized test scores based on the number of books at home. As standardized test scores have an international standard deviation of 100, these effects can be interpreted as percentages of an international standard deviation by which test achievement increases if the number of books is raised by one category. The authors validate these estimates by also looking at other measures of family background from the 2001 Progress in International Reading Literacy Survey (PIRLS). From Schuetz, G., Ursprung, H., and Woessman, L. (2005). *Education Policy and Equality of Opportunity. CESifo Working Paper 1518*.

Inequalities in Educational Outcomes

Inequalities by Social Background

The observation that children from poorer backgrounds do worse in terms of educational outcomes has a long history, for example, being highlighted by Rowntree's investigation into poverty in York, England at the turn of the twentieth century (Rowntree, 1901). Gaps in educational attainment between children from richer and poorer backgrounds continue to be marked at the start of the twenty-first century.

A large literature indicates that gaps in attainment emerge very early in children's lives. Substantial gaps in test-score attainment are found by income group before children start school (see Carneiro and Heckman 2004, for the United States, and Blanden and Machin, 2008, for the United Kingdom).

Some evidence links these differences to the sizable disparities in preschool enrolment between children from high- and low-education parental backgrounds (Meyers *et al.*, 2004).

Evidence suggests that early gaps in test-score attainment tend to widen as children age (Carneiro and Heckman, 2004; Feinstein, 2003). Recent international surveys that test school-age children enable comparisons to be made of the strength of the influence of family background on achievement across many countries. For example, **Figure 1** shows family background effects on test scores from an interesting recent paper by Schuetz *et al.* (2005). This uses cross-country data from the Third International Mathematics and Science Study (TIMSS) from 1995 and its repeat survey from 1999. In 53 out of 54 countries, the family background effect (in this study measured by the number of books at home) is statistically

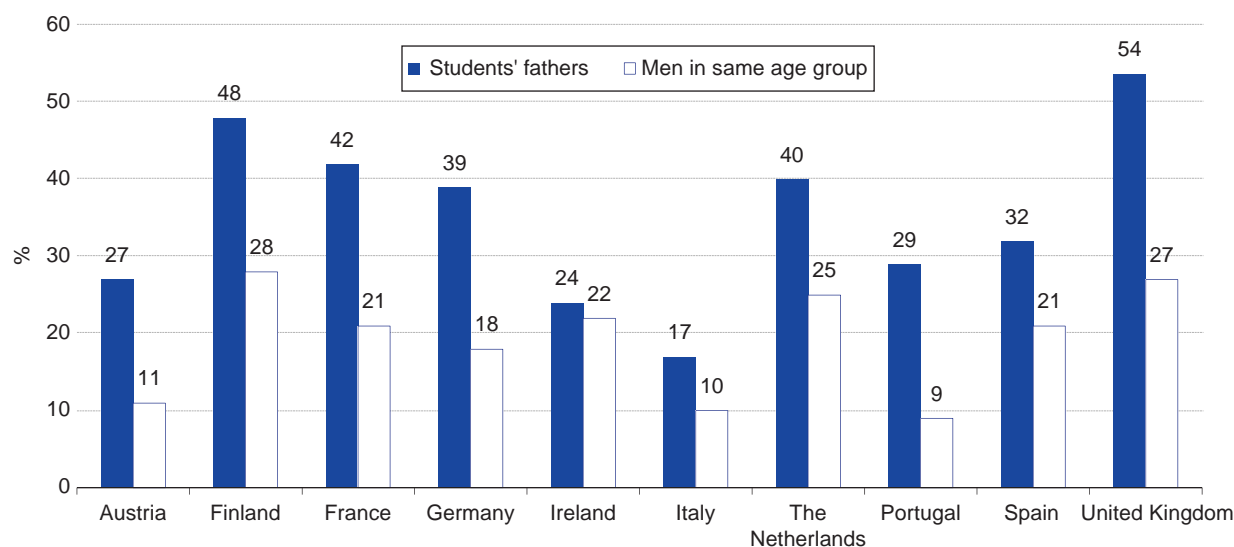


Figure 2 Comparison of students' fathers' educational level with other men of their generation. 1. England and Wales. Data refer to the parent (male or female) with the highest income. From figure 15, EUROSTUDENT 2005.

significant and the implied gaps in test scores are large. Moreover, the estimates are internationally comparable and show very large family background effects in some countries. The largest family background effects are in England and, while all show an important family background gradient, there is a fairly wide range of estimates. (Of course, some of the relative positions may well be sensitive to the use of TIMSS data as compared to other survey data on pupil achievement.)

Unsurprisingly, these substantial gaps in test scores lead to inequality in final educational attainments. This includes a higher probability of dropping out of school and lower qualification attainment. Indeed, there is evidence that inequalities continue to grow so that parental background influences final educational outcomes even once earlier achievements are taken into account.

Figure 2 compares the proportion of higher education (HE) students' fathers who have tertiary education with the proportion of all men of the same age with this level of education. Provided fathers are not substantively different from other men, this should provide a measure of the parental educational advantage of students in HE. In all of the countries that were included, fathers of university students are better educated. Once again, the United Kingdom stands out as among the most unequal countries.

Inequalities by Race, Ethnicity, and Immigrant Status

Gaps in educational performance between ethnicity, race, or immigration status are of interest because of the inequalities observed between different groups in the labor market. In the United States, the focus of this debate has tended to be around differences in performance between

blacks and whites (and to a lesser extent Hispanics). Cameron and Heckman (2001) find that minority groups are actually more likely to attend college, given their test scores; hence, it seems that the source of education inequality between whites and minorities in the United States is present in test scores (Jencks and Phillips, 1998).

In a recent contribution, Carneiro *et al.* (2005) have sought to investigate the time path of differences in cognitive and noncognitive skills by race. The authors find that differences open up as early as ages 1 or 2. There is evidence that this gap widens somewhat with age between blacks and whites, but with gaps appearing so early, it is difficult to state that the school system is the major source of racial inequality in the United States. Their conclusion is that adverse early environments have a long-term impact on the educational achievements of blacks and that early-age intervention is likely to be the most powerful tool in reducing inequality.

The debate on the educational differences between whites and minorities tends to reach rather country-specific conclusions, depending on the groups involved and the historical patterns of migration experienced. Platt (2007) finds substantial differences in educational attainment among the UK's 16- to 24-year-olds in 2001, with young people from Chinese and Indian backgrounds exceeding the performance of whites. Young black people perform more poorly than whites, although the gaps are not as stark as between whites and Bangladeshi and Pakistani groups. Evidence from the Millennium Cohort Study shows that all ethnic minorities are performing more poorly than whites in school readiness and vocabulary tests at age 3, with Bangladeshi and Pakistani children performing particularly badly (Hansen and Joshi, 2007). As these cohorts are separated by 20 years, it is not possible to combine this information to form an assessment of inequalities as children age.

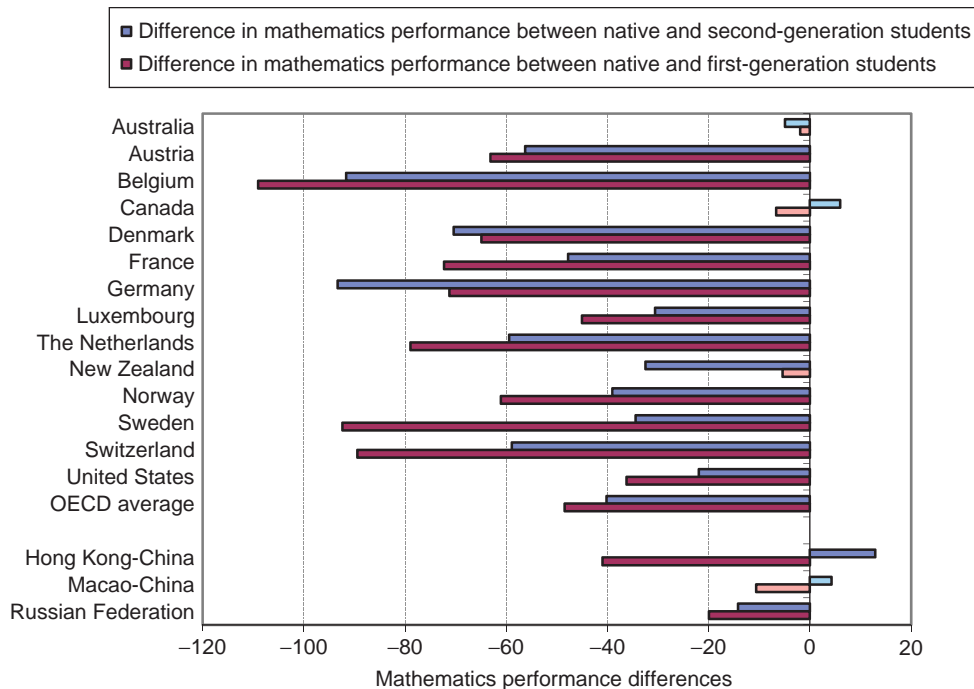


Figure 3 Differences in mathematics performance by immigrant status (2003). Statistically significant differences are marked in darker tones. From table A6.1a. in OECD PISA 2003.

Given the diversity of the racial and ethnic dimensions pertinent in different countries, it is difficult to provide a concise summary across countries. **Figure 3** attempts this by considering the differences between mathematics test scores in the Programme for International Student Assessment (PISA) between both first- and second-generation immigrants and native students (that is third generation or higher). In almost all countries (Canada being the exception), natives are outperforming immigrant groups at age 13 and, in most, gaps are larger for first-generation immigrants than second-generation immigrants, this suggesting assimilation. Card (2005) looks in detail at the performance of second-generation immigrants in the United States and finds that in many cases that they are outperforming whites. Worrying exceptions to this pattern in **Figure 3** are Germany and Denmark where second-generation groups appear particularly disadvantaged.

Inequalities by Gender

The relative performance of women compared to men in terms of qualifications gained has improved across the world in the past three or four decades. For example, in 1970, 42% of all US college undergraduates were women, whereas by 2000 this had risen to 60% (Bae *et al.*, 2000). Figures from the Organization for Economic Co-operation and Development (OECD) Education at a Glance show women's relative improvement in terms of

HE is widespread internationally, by comparing education rates for men and women in different age groups. Among the 30 countries studied all but six have higher graduation rates for men among the 55–64 age group; this reverses among the 25–30 age group so that women are outstripping men's HE performance in all but five countries. (Figures are from Education at a Glance 2007 Tables A1.2b and A1.2c.) Exemplar figures for the United States, United Kingdom, and Germany are given in **Figure 4**.

Jacob (2002) uses a decomposition approach to analyze why women in the United States are now more likely to go to college than men. He finds that 90% of the attendance gap can be explained by women's higher returns to degrees and women's greater noncognitive skills. Noncognitive skills are measured as middle school grades, grade retention, and the number of hours spent on homework – a measure of behavioral problems. The idea is that measures of school-based attainment will reflect effort and application once cognitive test scores are also taken into account. Jacob does not find any differences in cognitive scores between girls and boys conditional on other characteristics.

The PISA and TIMSS surveys reveal a fairly consistent pattern of differences in test-score achievement between girls and boys. As shown in **Table 1** (taken from Dustmann, 2005), figures from PISA reveal that in the majority of OECD countries, girls outperform boys in reading, while boys (although generally to a less extent) outperform girls in mathematics. Differences in science are much less pronounced.

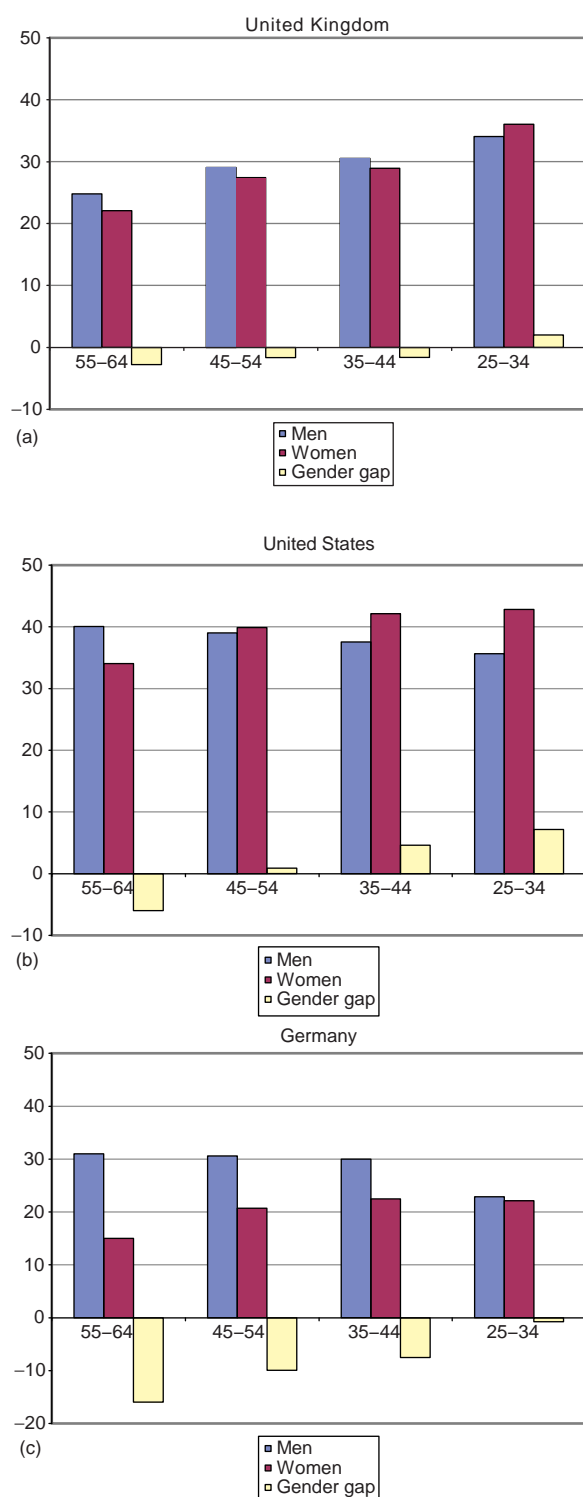


Figure 4 Proportion of population holding tertiary education qualifications by gender and age: United Kingdom, United States, and Germany. From tables A3.1b and A3.1c in OECD (2007). *Education at a Glance*. Paris.

Machin and McNally (2005) review the evidence on gender gaps in English schools from the 1970s to the turn of the century. They find evidence that while gaps in mathematics performance at the end of primary school

are slight, those for reading strongly favor girls and appear to have grown at least until the mid-1990s. Achievement at age 16 (the end of compulsory schooling) has an even larger gender bias with 10 percentage points more girls achieving 5 A–C grades at the General Certificate of Secondary Education (GCSE; the minimum grades required for progress into further academic education). The evolution of the gender gap at age 16 indicates that it is linked to the introduction of the GCSE assessment regime in 1988, which includes more continuous assessment, a change which appears to favor girls. This impression is confirmed by the most recent UK GCSE results, as coursework assessment ended and boys results in mathematics immediately outstripped that of girls.

Recent data from the Millennium Cohort in the United Kingdom allow us to consider gender gaps before school; as with gaps by socioeconomic status, the evidence suggests that they start early. Table 2 presents figures from a number of standardized test scores at ages 3 and 5. In all cases, girls tend to perform better in these test scores than boys do. In the vocabulary test, which is completed by children at both ages, the evidence is that the gap closes somewhat over the 2 years between the surveys. This might indicate that part of girls' better performance at age 3 is due to their higher level of maturity at that age.

To summarize this section on the dimensions of inequalities in education, we can say that inequalities by family background are large, persistent, and show little sign of reducing, while the picture for race/ethnicity and gender is more positive, at least in some countries.

Education and Economic Outcomes

Higher levels of educational attainment are strongly associated with higher earnings and better employment prospects. Much of the empirical work studying the labor market impact of education has its roots in the highly influential and pathbreaking work done in the economics of education field by American economists in the 1960s, especially Jacob Mincer (1958) who popularized the earnings function that relates wages to the number of years of schooling:

$$\log w = a + bS + c_1X + c_2X^2 + u$$

where w is earnings, S measures schooling, X denotes years of experience, and u is a random error term. Psacharopoulos and co-authors have written numerous papers comparing the coefficient b across the world and have found evidence that earnings returns to schooling are widespread and tend to be higher for primary education and in countries with lower levels of development (see Psacharopoulos and Patrinos, 2004, for the latest work).

Table 3 shows OECD evidence on educational wage differentials that accrue to people with tertiary education levels relative to post-secondary nontertiary levels in 15 countries. The existence of sizable gaps in earnings is seen

Table 1 Gender differences (males–females) in reading, mathematical, and scientific literacy in PISA 2000 results

	<i>Reading</i>		<i>Mathematics</i>		<i>Science</i>	
	<i>Mean score (total)</i>	<i>Difference in mean scores</i>	<i>Mean score (total)</i>	<i>Difference in mean scores</i>	<i>Mean score (total)</i>	<i>Difference in mean scores</i>
Australia	529	–33 ^a	533	12	527	3
France	505	–29 ^a	518	14 ^a	501	6
Germany	481	–34 ^a	486	15 ^a	484	2
Italy	485	–38 ^a	455	8	476	–9
Japan	522	–30 ^a	557	8	550	–7
Sweden	516	–37 ^a	509	7	511	–1
United Kingdom	524	–25 ^a	530	8	533	4
United States	504	–28 ^a	493	7	499	–5
OECD total	498	–29 ^a	497	11 ^a	501	–1
OECD average	499	–32 ^a	498	11 ^a	499	0

^aIndicates statistically significant gender gaps.

Figures taken from <http://www.pisa.oecd.org> and presented in Dustmann (2005).

Source: Schuetz, G., Ulsprung, H. W., and Woessmann, L. (2008). Education Policy and Equality of Opportunity. *Kyklos* 61(2): 279–308.

Table 2 Gender differences in early years test scores in the Millennium Cohort Study

<i>Test score</i>	<i>Range – all</i>	<i>Mean – boys</i>	<i>Mean – girls</i>	<i>Gender gap (standard error)</i>	<i>Sample</i>
<i>Age 3</i>					
School readiness	56–149	103.35	107.07	–3.72 (0.327)	14 037
Vocabulary	20–80	49.05	51.79	–2.74 (0.207)	14 773
<i>Age 5</i>					
Picture similarities	20–80	55.10	56.27	–1.17 (0.219)	15 131
Pattern construction	20–80	50.50	52.07	–1.57 (0.190)	14 897
Vocabulary	20–80	54.94	55.59	–0.644 (0.222)	15 144

Test scores are *t*-scores: these are adjusted for the difficulty of questions and the age of respondent. These are a more appropriate metric than percentiles as in most cases there are less than 100 possible scores.

Means are appropriately weighted according to guidance from the MCS data team. Calculations are authors' own.

Table 3 Earnings differentials between tertiary education and post-secondary nontertiary levels of education (aged 30–44, men and women, in 2005)

<i>Country</i>	<i>Earnings differentials</i>
Australia	0.34
Austria	0.48
Belgium	0.34
Denmark	0.22
Finland	0.38
France	0.48
Germany	0.50
Ireland	0.59
Italy	0.43
Korea	0.48
The Netherlands	0.47
Spain	0.30
Sweden	0.22
United Kingdom	0.61
United States	0.75

From table A1.9a in OECD (2007). *Education at a Glance*. Paris.

for all countries. According to these earnings differentials, acquisition of more education leads to significantly higher earnings.

As well as being associated with increased earnings, workers with educational qualifications also tend to correlate with improved employment probabilities. **Figure 5** compares the employment rates of those who do not complete upper secondary schools (equivalent to US high school) with employment rates for those who have high school education but no college education. It is very clear from these statistics that obtaining at least the typical level of education increases employment probabilities.

Causality

The empirical study of education-related earnings differentials was developed in tandem with human capital theory (see inter alia Becker, 1964); the idea is that educational attainment has a casual impact on labor market

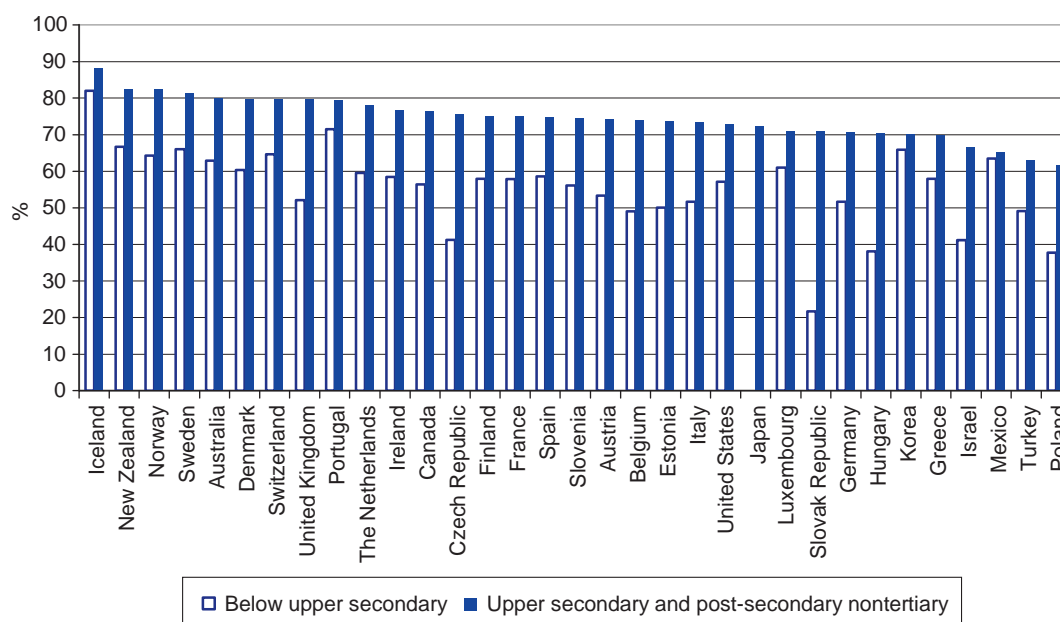


Figure 5 Employment rates by educational attainment (2005). From table A8.3a OECD (2007). *Education at a Glance*. Paris.

outcomes through improved productivity. Of course, the fact that more educated workers tend to be more productive does not prove that education is the cause of their higher productivity. Indeed, many commentators have noted that interpretation of positive earnings differentials for the more educated as a causal impact of education may not be correct due to differential selection into HE by more able individuals and those from higher-income families. As such, it may be that higher-ability/income people select into education more so that the positive coefficient on education in a wage equation is actually upward biased. The existence of ability bias of this form has been studied in detail, from the highly influential work of Griliches (1977) onward, and confirmed in research since. Empirically, it is the case that the education coefficient falls once ability proxies are included (this is because ability is positively correlated with both earnings and education).

Conversely, one may think that the education variable included in the earnings function may not be measured perfectly. This causes a downward bias in the estimated coefficient of education variables in earnings equations. Card (1999, 2001) considers the impact of ability bias and measurement error on estimated returns to education. In particular, Card carefully considers the interpretation of instrumental variables estimates, which are designed to purge endogeneity and measurement error bias.

In the recent literature addressing these issues, a number of what Card (2001) refers to as supply-side instruments are used to identify the causal impact of education, where researchers try to ensure that variations in education are driven by factors that do not

directly impact on wages. Estimated returns from this approach are above the basic uncorrected differentials. (Examples are changes in compulsory school leaving laws and differences in the accessibility of schools. Formally, these enter an education equation, but not the earnings equation, and so their impact on earnings is assumed to operate only through education acquisition.) **Table 4** summarizes some of the key papers in the field that identify the causal impacts. Moreover, differences between the basic and causal returns are of similar magnitudes for the studies considered in this table (they, of course, differ across studies owing to different data, different countries, and so on). Hence, there is robust cross-country evidence that the more educated get higher monetary rewards in the labor market.

Based on this, we can have a degree of confidence that education is causally associated with higher earnings. Moreover, it seems that the simple least squares regressions, which estimate human capital earnings functions provide a reasonably good idea of how big the education effects on earnings actually are. From this, it seems reasonable to conclude that there is a significant and sizable average rate of return in the labor market to HE.

Wider Benefits of Education

Social science researchers have considered the wider benefits of education by studying connections between education and a broader range of outcomes outside of the labor market. These are often thought of as picking up social returns to education and include health, crime, civic engagement, and intergenerational effects on chil-

Table 4 Evidence on the causal impact of education on earnings

Study	Data	Basic return (%)	Causal return (%)	Means to generate causal estimate
Angrist and Krueger (1991)	US Census	5–7	6–11	Variations in years of education generated by different quarter of birth
Card (1995)	1966 Cohort of Young Men, United States	7	13	Variation in years of education from proximity to college when growing up
Conneely and Uusitalo (1997)	Finnish men in the army in 1982	8	11	Variation in years of education from proximity to college when growing up
Harmon and Walker (1995)	Family Expenditure Survey, United Kingdom	6	15	Variation in education induced by raising of compulsory school leaving age
Ashenfelter and Rouse (1998)	1991–93 Princeton Twins Survey	7	9	Variation in education within twin pairs
Miller <i>et al.</i> (1995)	Australian Twins Register	3	5	Variation in education within twin pairs

Examples taken from Card's (1999) review.

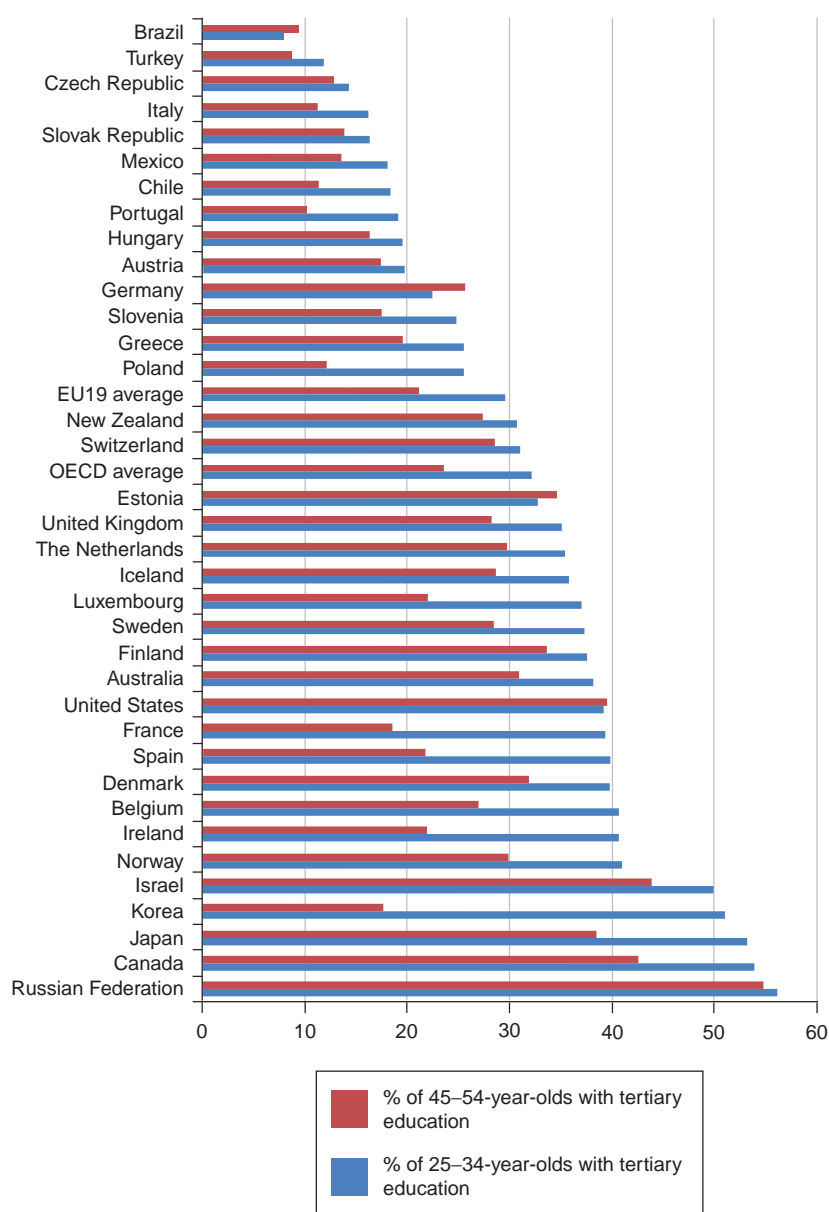


Figure 6 International education levels, 2005. The reference year is 2003 for the Russian Federation and 2004 for Chile and Brazil. From table A1.3a of OECD (2007). *Education at a Glance*. Paris.

dren's outcomes. From this work, there is evidence of important externalities, in that education significantly improves health outcomes (Grossman and Kaestner, 1997; Lleras-Muney, 2005), is associated with lower crime levels (Lochner and Moretti, 2004; Machin and Vujic, 2005), and enhances the extent of civic engagement and participation (Brehm and Rahn, 1997; Bynner and Egerton, 2001; Dee, 2004). Moreover, there are important intergenerational effects of education of adults on the education of their children (Black *et al.*, 2005).

Changes over Time

We have so far reviewed the evidence that differential educational attainments can lead to substantial differences in earnings, employment probabilities, and other outcomes that matter to individuals' welfare. It is clear that employers are prepared to pay higher wages to more educated workers; this reflects in part their scarcity value. As simple demand and supply analysis indicates that, *ceteris paribus*, as more workers become highly educated, the earnings returns to being educated will decline.

Figure 6 shows the proportion of the populations of 36 countries with tertiary education in two age groups, 25–34 and 45–54. In almost every case, the proportion of educated workers is higher in the younger group – attainment has risen over a generation. If demand had remained steady, we would expect this to have resulted in a fall in earnings (and indeed employment) differentials over time. However, the most recent survey by Psacharopoulos and Patrinos (2004) finds that while returns to each additional year of education have slightly declined worldwide, the returns to higher education have increased.

In the United Kingdom and United States, the rising educational wage differentials generated have made an important contribution to rising wage inequalities (Katz and Autor, 1999). Understanding why these changes have occurred is essential to understanding the relationship between education and inequality.

An increase in the supply of educated workers will lead to a decline in the wage premium they receive unless demand for them increases further. This seems to have been what has happened in practice. Large increases in the demand for graduates have occurred so that wage differentials related to education have stayed constant or increased in the face of the expansion of tertiary education in many countries. Researchers have studied what lies behind the increase in the demand for educated workers in the United States and the United Kingdom. In both countries, wage premia have risen despite a massive expansion in the supply of graduates with a tertiary education.

The leading explanation for the rise in demand for skilled workers is skill-biased technical change (SBTC).

This hypothesis states that the rise in demand for more skilled workers has been driven by new technologies in the workplace. The critical idea is that these new technologies lead to higher productivity, but that only some workers possess the necessary skills to use them. As such, employers are prepared to increase the wages of the skilled workforce who complement the new technology. However, at the same time, less-skilled workers do not possess enough skills to operate the new technologies and their wages are lowered or they lose their jobs. Relative wages and/or employment of the more skilled, therefore, rise.

The SBTC hypothesis has received substantial empirical support (see the survey of Machin and Van Reenen, 2008). At the risk of being somewhat simplistic, the approach generally taken to investigating SBTC is to model the demand for skilled workers (measured by their share in the wage bill) as a function of technology. A strong positive relationship is found, and the fact that similar industries are affected across the world, makes other competing explanations (such as increased international trade) less convincing (see Berman *et al.*, 1998, or Machin and Van Reenen, 1998, for international evidence in line with this).

Autor *et al.* (2003) have lately proposed a more sophisticated version of the SBTC hypothesis. Their argument is that computerization reduces the demand for routine tasks (for manual and nonmanual workers) but results in an increase in demand for analytic or nonroutine skills. Thus, routine nonmanual tasks (e.g., clerical work) may be replaced by computers, while some nonroutine tasks done by manual workers (like cleaning) are largely unaffected. Thus, one sees increased demand for workers with the skills and capabilities to do jobs involving nonroutine tasks. Again, this shows that education that confers these skills on workers is likely to have a bigger payoff in the labor market and generate earnings returns.

Conclusions

Education and inequality are closely related. It is evident that education acquisition is related to family background and that different demographic groups acquire different levels of education. Since education yields a private return in the labor market and there are social returns to education, it is clear that the uneven patterns of education acquisition have the potential to generate inequalities in economic and social outcomes. Depending on how these uneven patterns of acquisition are distributed across the population, it is evident that education can have an equalizing or disequalizing effect on outcomes. The recent experience of increased labor market inequality being linked to changing patterns of educational attainment suggests that it has, at least in this recent time period,

been disequalizing, and therefore, had a tendency to raise inequality.

See also: Human Capital; Race Earnings Differentials; Returns to Education in Developed Countries; Returns to Education in Developing Countries; School Quality and Earnings.

Bibliography

- Autor, D., Levy, F., and Murnane, R. (2003). The skill content of recent technological change: An empirical investigation. *Quarterly Journal of Economics* **118**, 1279–1333.
- Bae, Y., Choy, S., Geddes, C., Sable, J., and Synder, T. (2000). *Trends in Education Equity of Women and Girls*. Washington, DC: National Center for Education Statistics.
- Becker, G. (1964). *Human Capital: A Theoretical Analysis with Special Reference to Education*. New York: Columbia University Press.
- Berman, E., Bound, J., and Machin, S. (1998). Implications of skill-biased technological change: International evidence. *Quarterly Journal of Economics* **113**, 1245–1280.
- Black, S., Devereux, P., and Salvanes, K. (2005). Why the apple doesn't fall far: Understanding the intergenerational transmission of education. *American Economic Review* **95**, 437–449.
- Blanden, J. and Machin, S. (2008). Up and down the generational income ladder in Britain: Past changes and future prospects. *National Institute Economic Review* **205**, 101–116.
- Brehm, J. and Rahn, W. (1997). Individual-level evidence for the causes and consequences of social capital. *American Journal of Political Science* **41**, 999–1023.
- Bynner, J. and Egerton, M. (2001). *The Wider Benefits of Higher Education*. London: Higher Education Funding Council for England.
- Cameron, S. and Heckman, J. (2001). The dynamics of educational attainment for Black, Hispanic and White Males. *Journal of Political Economy* **109**, 455–499.
- Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling. In Christofides, L., Grant, E., and Swidinsky, R. (eds.) *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*, pp 201–222. Toronto, ON: University of Toronto Press.
- Card, D. (1999). The causal effect of education on earnings. In Ashenfelter, O. and Card, D. (eds.) *Handbook of Labor Economics*, vol. 3, pp 1801–1863. Amsterdam: North-Holland.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent problems. *Econometrica* **69**, 1127–1160.
- Card, D. (2005). Is the new immigration really so bad? *Economic Journal* **115**, F300–F323.
- Carneiro, P., Heckman, J., and Masterov, D. (2005). Labour market discrimination and racial differences in premarket factors. *Journal of Law and Economics* **48**, 1–39.
- Dee, T. (2004). Are there civic returns to education? *Journal of Public Economics* **88**, 1697–1720.
- Dustmann, C. (2005). The assessment: Gender and the life cycle. *Oxford Review of Economic Policy* **21**, 325–339.
- Feinstein, L. (2003). Inequality in the early cognitive development of British children in the 1970 Cohort. *Economica* **70**, 73–98.
- Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica* **45**, 1–22.
- Grossman, M. and Kaestner, R. (1997). Effects of education on health. In Behrman, J. and Stacey, N. (eds.) *The Social Benefits of Education*, pp 69–123. Ann Arbor, MI: University of Michigan Press.
- Hansen, K. and Joshi, H. (eds.) (2007). *Millennium Cohort Study Second Survey: A Users' Guide to Initial Findings*. London: Centre for Longitudinal Studies.
- Jacob, B. (2002). Where the boy's aren't: Non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review* **21**, 589–598.
- Jencks, C. and Phillips, M. (1998). *The Black-White Test Score Gap*. Washington, DC: Brookings Institution Press.
- Katz, L. and Autor, D. (1999). Changes in the wage structure and earnings inequality. In Ashenfelter, O. and Card, D. (eds.) *Handbook of Labor Economics*, vol. 3, pp 1463–1555. Amsterdam: North-Holland.
- Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *Review of Economic Studies* **72**, 189–221.
- Lochner, L. and Moretti, E. (2004). The effect of education on criminal activity: Evidence from prison inmates, arrests and self-reports. *American Economic Review* **94**, 155–189.
- Machin, S. (2005). Social disadvantage and education experiences. *Report for OECD*.
- Machin, S. (2008a). The new economics of education: Methods, evidence and policy. *Journal of Population Economics* **21**, 1–19.
- Machin, S. (2008b). Education and inequality. In Nolan, B., Salverda, W., and Smeeding, T. (eds.) *Handbook of Economic Inequality*. Oxford University Press.
- Machin, S. and McNally, S. (2005). Gender and student achievement in English schools. *Oxford Review of Economic Policy* **21**, 357–372.
- Machin, S. and Van Reenen, J. (1998). Technology and changes in skill structure: Evidence from seven OECD countries. *Quarterly Journal of Economics* **113**, 1215–1244.
- Machin, S. and Van Reenen, J. (2008). Changes in wage inequality. In Weinberg, B. and Taber, C. (eds.) *New Palgrave Dictionary of Economics*.
- Machin, S. and Vujic, S. (2005). Crime and education in the United Kingdom. *Draft Paper*. Centre for Economic Performance.
- Meyers, M., Rosenbaum, D., Ruhm, C., and Waldfogel, J. (2004). Inequality in early childhood care and education: What do we know? In Neckerman, K. (ed.) *Social Inequality*, pp 223–269. New York: Russell Sage Foundation.
- Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy* **66**, 281–302.
- Platt, L. (2007). Making education count: The effects of ethnicity and qualifications on intergenerational social class mobility. *Sociological Review* **55**, 485–508.
- Psacharopoulos, G. and Patrinos, H. A. (2004). Returns to investment in education: A further update. *Education Economics* **12**, 111–134.
- Rowntree, B. S. (1901). *Poverty: A Study of Town Life Centennial Edition*. Bristol: Policy Press.
- Schuetz, G., Ursprung, H. W., and Woessmann, L. (2008). Education Policy and Equality of Opportunity. *Kyklos* **61**(2): 279–308.

Further Reading

- Angrist, J. and Krueger, A. (1991). Does compulsory school attendance affect schooling and earnings. *American Economic Review* **106**, 979–1014.
- Ashenfelter, O. and Rouse, C. (1998). Income, schooling and ability: Evidence from a new sample of twins. *Quarterly Journal of Economics* **113**, 253–284.
- Carneiro, P. and Heckman, J. (2004). Human capital policy. In Heckman, J. and Krueger, A. (eds.) *Inequality in America: What Role for Human Capital Policies*, pp 77–240. Cambridge, MA: MIT Press.
- Conneely, K. and Uusitalo, R. (1997). Estimating heterogeneous treatment effects in the Becker schooling model. Mimeo.
- Harmon, C. and Walker, I. (1995). Estimates of the economic return to schooling for the United Kingdom. *American Economic Review* **85**, 1278–1286.
- Miller, P., Mulvey, C., and Martin, N. (1995). What do twins studies reveal about the economic returns to schooling? A comparison of Australian and US findings. *American Economic Review* **85**, 586–599.

Relevant Websites

- <http://www.eurostudent.eu> – EUROSTUDENT Report 2005.
- <http://www.oecd.org> – OECD: Education at a Glance 2007.

Human Capital

E R Eide and M H Showalter, Brigham Young University, Provo, UT, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Human capital refers to the intrinsic productive capabilities of human beings. These capabilities can be increased through investment in things such as education, on-the-job training, and health. Human capital is viewed as an asset that generates a flow of services, most often measured as earnings, although broader measures of output are also used.

Adam Smith in *The Wealth of Nations* (Smith, 1776) set the stage for the study of human capital. Although he does not use the phrase human capital, he identifies the acquired and useful abilities of individuals as a fundamental source of wealth and economic progress of a country. Writing more than a century later, Alfred Marshall notes in his *Principles of Economics* (Marshall, 1920) the long-term nature of investments in human capital and the family's role in undertaking them. Additionally, Marshall expands the notion of returns on human capital to include non-monetary considerations. However, Marshall is also credited with delaying the study of human capital due to his reluctance to put it on the same footing as physical capital.

The modern study of human capital coincides with two developments in economics. The first was a resurgent interest in understanding why economies grow. It was apparent to researchers that national output was growing at a much faster rate than the rates of growth of inputs – land, labor, and physical capital (Denison, 1962). A leading hypothesis to explain the anomaly was that labor was mismeasured: a day of work from a typical worker in the 1950s was substantively different than that of a worker in the 1920s.

The second development dovetailed with the first – the availability of large datasets that allowed exploration of worker productivity and earnings and how they related to characteristics such as the years of education and age. The data revealed that education levels were increasing dramatically and that the higher levels of education might explain rising productivity and wages.

Three authors deserve particular note amid the flurry of theoretical and empirical work in the late 1950s and early 1960s. Jacob Mincer (1958) stimulated a vast literature measuring the returns to education. Using census data, he documented the now-classic result that the years of education has an inverted U-shape on the rate of earnings growth. Theodore Shultz focused on the role of education and general investments in human capital in explaining the increased productivity of labor. His monograph

The Economic Value of Education focused attention on education as a subject worthy of concentrated study by economists, thus beginning the field of the economics of education. Gary Becker (1964) organized the emerging threads of empirical and theoretical work into a coherent framework that provided a guidebook for future research.

The study of human capital has proven to be an exceptionally fruitful vein of research with implications for individuals, firms, and nations. The general topic has resulted in an enormous quantity of research. Several comprehensive surveys have been done and are listed under the section titled 'Further reading'. This article will of necessity be very selective in its coverage. The primary focus is on education as an investment in human capital. In what follows, the private returns to education are first addressed. Returns are generally measured in monetary units, but some evidence on nonmonetary returns is also discussed. This is followed by a discussion of the social returns to education. As early as Adam Smith, the potential externalities of education were noted, and this section discusses some of the more prominent findings in the empirical literature. The final section summarizes the conclusions.

Private Returns to Education

Theoretical Issues and Modeling

This section begins with a brief overview of the theoretical underpinnings of the study of human capital. The terminology of capital is instructive: human capital and decisions involved with it are the same type as decisions involving physical capital, such as equipment and buildings. A business executive invests in new equipment if the expected benefit derived from the equipment exceeds the cost. Human capital theory uses the same construct: an individual invests in education with an expectation that the investment will provide a benefit in the form of higher earnings.

More generally, human capital theory assumes that individuals take actions that will likely increase their future earnings and overall well-being. Such investments are costly: they might involve direct costs such as tuition and fees for school, and indirect costs such as foregone earnings during the period spent in school. These investments result in some expected future benefits. The benefits might include a higher wage, but can also be anything that the individual values, for example, better working

conditions or a longer life. Human capital theory typically models investment decisions such as those resulting from an optimization process: an individual will invest in such activities in order to maximize well-being over the course of a lifetime. Observed outcomes in the marketplace will be the result of an equilibrium process where the demand for specific skills and abilities is balanced with its supply.

Human capital theory offers numerous useful insights and testable hypotheses about human behavior. One of the first to be explored is that observed differentials in earnings can be explained in part by differences in training costs. Adam Smith noted that professions that require years of training tend to have annual higher earnings than professions with no such restrictions. Using human capital theory, the explanation for this empirical observation is that a rational individual would only be willing to incur the costs of tuition and lost earnings during the training period if that investment has a sufficiently high rate of return in the form of higher earnings post training. This insight also suggests that lifetime earnings is a more appropriate measure when evaluating inequality: two individuals, one who invests in professional training and one who does not, can have identical lifetime earnings, but very different earnings at each point during their lives.

Human capital theory undergirds most of modern labor economics. It offers explanations and insights on topics as diverse as discrimination, inequality, unemployment, fertility, marriage markets, immigration, and productivity. This brief survey focuses on the relatively narrow slice of empirical work that touches on topics in education.

Empirical Estimates and Methodology

Although human capital theory suggests that individuals invest in education in anticipation of a wide range of benefits, most empirical work has focused on the monetary rewards of increased earnings. Such data have tended to show rates of return of about 10% for an additional year of education, with some variation by gender and race/ethnicity (Psacharopoulos and Patrinos 2004). These return measures ignore nonmonetary and consumption benefits of education, and they also ignore any externalities associated with education. There has been substantial work done on these latter topics which are discussed later in this article.

Several methodological issues arise in estimating the returns to education. The goal of most empirical work is to establish a causal link between education and measures of return, which tends to be much more difficult than simply measuring a correlation. Early work compared lifetime earnings profiles for groups with different levels of educational attainment. More advanced analysis based on multiple regression analysis has attempted to account for a host of technical estimation problems such as omitted variables,

nonrandom samples, incorrectly measured variables, and jointly determined outcomes. These issues are discussed briefly in what follows.

A key omitted variable in much empirical work is innate ability (e.g., natural intelligence, work ethic) which can also include unobserved effort. The fundamental difficulty is that high-ability individuals are likely to obtain more years of schooling than low-ability individuals, but high-ability individuals would also tend to earn more for any given amount of schooling than low-ability individuals. This makes it difficult to disentangle the effect of schooling versus innate ability.

A large literature is devoted to investigating and overcoming this particular problem. Ideally, a researcher would like to observe otherwise identical individuals who differ only in their level of education. Although this approach is impossible in practice, clever research strategies using advanced econometrics have been employed with varying degrees of success.

Nonrandom sampling is often a problem with estimating returns to education for individuals with weak attachment to the labor market. For example, in contrast to most college-educated males, a substantial fraction of women who graduate from college will exit the workforce for extended periods of time. This implies that estimating returns to college will rely on a subset of the available data (women who work) that might not be representative of the full population of women. As with accounting for unobserved ability, a large literature has developed to account for this type of problem, again, with varying degrees of success.

Measurement error can afflict several variables of interest in empirical models. One example is earnings: earnings can be misstated in surveys because of carelessness by either the responder or the person (or machine) coding the information; or, responder wariness toward the survey instrument might lead to inaccurate data. Another example is in the measurement of school quality: a year of schooling from an Ivy League college might be very different than a year of schooling from the local junior college. Such problems cloud the measurement of the causal relationship between schooling and earnings, and the relationship between human capital accumulation and outcomes more generally. Several approaches have been used to account for this problem, the simplest of which is the development of better data sources.

The problem of jointly determined variables, also known as endogeneity, arises in situations where multiple outcomes are simultaneously decided upon. For example, household fertility decisions are likely to be made jointly with decisions about the level and type of education to obtain. This makes it difficult to measure a causal impact of one endogenous variable on another, although such causal links are theoretically possible. For example, it seems likely that having additional children makes it

more costly and therefore less probable that a person will pursue additional years of education. But measuring the independent effect is confounded by the joint-decision process. The typical strategy to deal with endogeneity is to use an econometric technique known as instrumental variables estimation. Alternative complex estimation techniques have also been used in the empirical literature.

Topics in Empirical Estimation

Research on the development of human capital tends to have a three-pronged focus: (1) childhood and adolescence, with a heavy emphasis on kindergarten through high school completion; (2) young adulthood, emphasizing formal post-secondary education; and (3) adults, highlighting on-the-job training and general experience. Each is discussed in turn.

Primary and secondary education

Economic research on the effects of K-12 education is usually organized around the notion of an education production function. This approach posits that inputs such as teacher quality, class size, school policies, family demographics, and general aptitude of students determine observable outcomes such as test scores and high school completion.

A large empirical literature attempts to quantify the empirical relationship between inputs and outputs. One of the earliest and most influential studies was directed by James S. Coleman (Coleman *et al.*, 1966). This government-sponsored study documented a significant Black–White achievement gap in test scores, and it came to the controversial conclusion that family background characteristics were much more important in explaining outcomes than school-controlled inputs such as class size and teacher quality. Subsequent research critiqued both the methodology and the data of the Coleman report, but an acknowledgment of the weak association between school inputs and student outcomes has been the dominant view among economists since the 1980s (Hanushek, 1986; Burtless, 1996).

However, a small but growing set of research is reexamining the production function relationship using data generated by experimental and quasi-experimental designs. A notable example is the Tennessee student teacher achievement ratio (STAR) project that performed an experiment to measure the impact of class size on student outcomes (Word *et al.*, 1994; Krueger, 1999; Krueger and Whitmore, 2001). The results suggest a significant effect of class size on test scores and on the likelihood of taking a college-entrance exam.

An important set of related research employs quasi-experiments. A quasi-experiment refers to data generated without an explicit experiment such as the STAR project, but includes institutional details that allow researchers to

treat the data as if it were. An interesting example of this approach is a paper by Angrist and Lavy (1999) that uses the institutional rule of a 40-child maximum class size in Israeli public schools to examine the effects of class size on test scores. They find that smaller class size leads to higher test scores in some situations.

A variety of policy experiments designed to improve and evaluate student-learning outcomes are currently underway throughout the United States, and past policy innovations are currently being studied to determine their effects, if any. Examples include the following: (1) High-stakes testing where administrators and teachers can be punished if aggregate test scores are too low, or where individual students can be retained in a grade or be subject to remedial education if their test scores are not sufficiently high. (2) Vouchers where individuals receive government funding to be spent on the school of their choice, including private schooling. (3) Charter schools that are fully government funded, but have less-direct oversight than typical government schools. (4) The effects of peers on student outcomes. Peer effects are fundamental in evaluating policies like tracking where students are segregated by ability measures such as test scores, or combining children of different grades in the same classroom. (5) The role of teacher quality on student outcomes. Quality can refer to experience, education level, and type, or measures of teacher performance. (6) Direct incentives to students for performance on standardized tests, such as monetary payments for high test scores.

Many of these topics and research agendas are discussed in greater detail elsewhere in the encyclopedia.

Post-secondary education

A substantial amount of work has been done over the past two decades focusing on postsecondary formal education. In the early 1990s, researchers documented a substantial rise in the return to completing college relative to completing high school. A number of explanations have been given. The rising demand for skilled labor due to the forces of international trade and specialization has received some empirical and theoretical support. (Acemoglu, 2002)

Another area of research has focused on variations in quality of education. Noting that there are large discrepancies in college tuition across institutions, a natural question arises whether higher tuition results in correspondingly higher earnings. An interesting strand of research compares outcomes from individuals attending elite private institutions (e.g., Ivy League Schools) with less-prestigious universities. The empirical results are mixed. A widely cited work by Brewer *et al.* (1999) suggests earnings tend to be as much as 40% higher for students from elite institutions compared with their peers at less-selective universities. But using different data and different methodology, Dale

and Krueger (2002) find a much smaller effect on earnings from attending an elite university.

A related literature examines the returns to attending a community (junior) college. This literature notes that a majority of first-year college students enrol in a community college. The empirical evidence to date suggests that the returns to a year of community college are similar to the returns to a year of college at a 4-year institution (Kane and Rouse, 1995).

Returns to particular college majors has also been an active area of research. The literature documents substantial differences in the returns to various majors, with math-oriented and technical majors earning approximately 30% more than nontechnical majors. Large differences remain even after accounting for the ability bias that leads innately more capable individuals into technical majors.

On-the-job training and work experience

Becker's (1964) seminal book notes a difference between firm-specific and general human capital. In his model, firms would not invest in general human capital – general literacy, for example – because workers might leave the firm after they receive the training. But recent work suggests that there might be conditions under which firms do invest in general training (Acemoglu and Pischke, 1999). A growing body of empirical evidence supports this work. For example, Autor (2001) documents the workings of temporary help firms providing training in general skills.

A final area of inquiry deals with the depreciation of human capital. While formal education and work experience can build human capital, detachment from the labor force and general inactivity can deteriorate it. This is one of the major economic concerns arising from extended spells of unemployment. Work by Mincer and Ofek (1982) examines the implications of interruptions of labor-force participation by married women. They find that the longer the period of interruption, the lower the wages upon reentry into the labor force. However, they also find rapid growth in wages after these women return to work.

Nonmonetary Returns to Education

As noted by Marshall, human capital investments might have payoffs beyond increases in earnings. There is an expanding research agenda attempting to measure such returns. For example, recent work by Lleras-Muney (2005) examines the causal impact of education on longevity. Using variation in education level induced by compulsory education laws, she estimates that a year of schooling leads to an increased life expectancy at age 35 of 1.7 years.

A second example is the interaction of human capital investment and family interactions. Higher marketplace returns reduce the demand for children and lead to relatively higher labor-force participation by married women. Divorce rates, wage inequality, and occupational segregation by gender exemplify topics where human capital theory has extended its reach. It is often difficult to conclusively show causal relationships between educational attainment and these nonmonetary outcomes due to data constraints, but research continues to explore these topics and circumstantial evidence continues to mount.

Social Returns to Education

Research evidence suggests there exists substantial private gain to investment in education, but it is also an important question whether gains accrue to individuals other than the person actually investing in human capital: is there some social return to education over and above what any given individual obtains? In economics, such a phenomenon is known as a positive externality and its existence justifies government involvement in the promotion of education. Without such externalities, the rationale for government provision of education becomes much weaker.

Two general approaches have been used to measure the extent of externalities. First, macroeconomists have tried to identify aggregate human capital externalities. A recent focus of this literature has been to determine whether skilled (high human capital/education) workers tend to raise wages for unskilled workers (Ciccone and Peri, 2006). More generally, this aggregate approach attempts to measure whether there are monetary benefits to entire economies that are not fully reflected in wages. There is no consensus on the magnitude of this aggregate effect.

The second approach is a more narrowly focused search to identify specific externalities from education, usually involving nonmonetary outcomes. A few examples include the following:

- *Citizenship.* An educated citizenry is often argued to be necessary for a functioning democracy. Such fundamental tasks as voting and filing taxes require some basic level of cognitive function. Recent work has attempted to measure whether education has a discernable, causal effect on voting behavior. Milligan *et al.* (2004) find that schooling increases civic participation in the United States. They also find that higher levels of education are positively associated with individual awareness of campaign issues and that educated individuals are generally more involved in the political process.
- *Crime.* Criminal activity offers an interesting case study in externalities. Criminal acts almost by definition cause harm to people other than the person committing the crime. Hence the reduction in crime due to higher

education levels is *de facto* an externality of education. Human capital explanations for a link between education and crime usually involve opportunity-cost arguments. Individuals with few marketable skills have a low opportunity cost in the commission of crime and in the possible time cost of incarceration. Such arguments are used to explain the widely noted age profile of criminal activity: property and violent crimes increase through adolescence for males, peaks in the late teens, and declines markedly thereafter. On the other hand, white-collar crime, which tends to require somewhat more sophistication, peaks at a much later age and declines more slowly. An interesting study by Lochner and Moretti (2004) attempts to measure the causal effect of schooling on criminal activity. They estimate that an extra year of schooling results in a 0.37 percentage point reduction in the probability of incarceration for blacks, and a 0.10 percentage point reduction for whites. They estimate that the social savings per additional high school graduate is between US\$1170 and US\$2100. They also find that nearly a quarter of the difference in the black–white gap in incarceration rates can be accounted for by differences in years of schooling.

- *Health.* The case for externalities in health is theoretically compelling but empirically weak. A key example involves vaccinations. Vaccinating an individual against communicable diseases benefits the individual (a private return), but also lowers the incidence of the disease in the population (a social return). If education has a causal effect that increases vaccination rates, this would imply the existence of an externality. Although a strong positive correlation exists between education and vaccination rates, a definitive causal link has not been established.

Other possible health links include educational effects that lower rates of smoking and drunk driving. Both of these activities have the potential to harm individuals other than the one engaged in the activity. But as with vaccinations, solid empirical support is currently lacking.

Some recent work argues that individuals do not adequately account for the addictive nature of products like cigarettes and alcohol and that the resulting unintended harm to an individual from their own actions ought to be considered as an externality. This work is in its early stages and no definitive empirical work has yet been produced.

Finally, some authors have argued that insurance pooling results in an education externality on health (Manning *et al.*, 1991). The empirical evidence suggests that education leads to healthier lifestyles, but with insurance pooling, healthy individuals are not rewarded with lower health insurance premiums in employment-based insurance pools, despite the reduction in medical expenses resulting from better health. In not being rewarded,

healthy individuals confer a positive externality on the less-healthy individuals in the pool. The empirical magnitude of the potential effect is uncertain.

Conclusion

Human capital theory forms the basis for most of the empirical work in the economics of education. It has proved to be a powerful tool for conceptualizing how individuals make educational choices and for guiding how the implications of those choices are measured. Empirical work establishing causal relationships is often difficult, fraught with problems of limited and inadequate data. But exciting new work continues to refresh, refine, and occasionally refute previous theoretical and empirical conclusions.

See also: Education and Inequality; Education Production Functions: Concepts; Education Production Functions: Developed Country Evidence; Education Production Functions: Evidence from Developing Countries; Race Earnings Differentials; Returns to Education in Developed Countries; Returns to Education in Developing Countries; School Quality and Earnings; The External Benefits of Education.

Bibliography

- Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of Economic Literature* **40**, 7–72.
- Acemoglu, D. and Pischke, J. (1999). The structure of wages and investment in general training. *Journal of Political Economy* **107**, 539–572.
- Angrist, J. D. and Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* **114**, 533–575.
- Autor, D. H. (2001). Why do temporary help firms provide free general skills training? *Quarterly Journal of Economics* **116**, 1409–1448.
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, 3rd edn. Chicago, IL: University of Chicago Press.
- Brewer, D. J., Eide, E. R., and Ehrenberg, R. G. (1999). Does it pay to attend an elite private college? *Journal of Human Resources* **34**, 104–123.
- Burtless, G. (ed.) (1996). *Does Money Matter?: The Effect of School Resources on Student Achievement and Adult Success*. Washington, DC: The Brookings Institution.
- Ciccone, A. and Peri, G. (2006). Identifying human capital externalities: Theory and applications. *Review of Economic Studies* **73**, 381–412.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., *et al.* (1966). *Equality of Educational Opportunity*. Washington, DC: US Government Printing Office.
- Dale, S. B. and Krueger, A. B. (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *Quarterly Journal of Economics* **117**, 1491–1527.
- Denison, E. F. (1962). *The Sources of Economic Growth in the United States and the Alternatives before Us*. New York: Committee on Economic Development.
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature* **24**, 1141–1177.
- Kane, T. J. and Rouse, C. E. (1995). Labor-market returns to two- and four-year college. *The American Economic Review* **85**, 600–614.

- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics* **114**, 497–532.
- Krueger, A. B. and Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *Economic Journal, Royal Economic Society* **111**, 1–28.
- Lochner, L. and Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *The American Economic Review* **94**, 155–189.
- Lleras-Muney, A. (2005). The relationship between education and adult mortality in the United States. *Review of Economic Studies* **72**, 189–221.
- Manning, W. G., Keeler, E. B., Newhouse, J. P., Sloss, E. M., and Wasserman, J. (1991). *The Costs of Poor Health Habits*. Cambridge: Harvard University Press.
- Marshall, A. (1920). *Principles of Economics*, 8th edn. London: Macmillan [Reprinted 1997, Amherst, NY: Prometheus.].
- Milligan, K., Moretti, E., and Oreopoulos, P. (2004). Does education improve citizenship? Evidence from the United States and the United Kingdom. *Journal of Public Economics* **88**, 1667–1695.
- Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy* **66**, 281–302.
- Mincer, J. and Ofek, H. (1982). Interrupted work careers: Depreciation and restoration of human capital. *Journal of Human Resources* **17**, 3–24.
- Psacharopoulos, G. and Patrinos, H. A. (2004). Returns to investment in education: A further update. *Education Economics* **12**, 111–135.
- Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations* (Everyman's Library Edition). London: W. Strahan and T. Cadell. [Reprinted 1991, New York: Knopf.].
- Word, E. R., Johnston, J., Bain, H. P., et al. (1994). *The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Final Summary Report 1985–1990*. Nashville, TN: Tennessee State Department of Education.

Further Reading

- Acemoglu, D. and Angrist, J. (2000). How large are human-capital externalities? Evidence from compulsory schooling laws. *NBER Macroeconomics Annual* **15**, 9–59.
- Behrman, J. R. and Stacey, N. (eds.) (1997). *The Social Benefits of Education*. Ann Arbor, MI: University of Michigan Press.
- Cohn, E. and Terry, G. G. (2004). *The Economics of Education*, 3rd edn. Mason, OH: Thomson/South-Western.
- Davies, J. (2002). Empirical evidence on human capital externalities. London, ON: Mimeo.
- Desai, S. and Alva, S. (1998). Maternal education and child health: Is there a strong causal relationship? *Demography* **35**, 71–81.
- Hartog, J. and Van Den Brink, H. M. (2007). *Human Capital: Advances in Theory and Evidence*. Cambridge: Cambridge University Press.
- Grogger, J. and Eide, E. R. (1995). Changes in college skills and the rise in the college wage premium. *Journal of Human Resources* **30**, 280–310.
- Marshall, A. (1920). *Principles of Economics*, 8th edn. London: Macmillan.
- Rosen, S. H. (1987). *Human Capital. The New Palgrave: A Dictionary of Economics*. New York: Macmillan.
- Shultz, T. W. (1963). *The Economic Value of Education*. New York: Columbia University Press.
- Wolfe, B. and Haveman, R. (2001). Accounting for the social and non-market benefits of education, in the contribution of human and social capital to sustained economic growth and well-being. *International Symposium Report edited by Human Resources Development Canada and Organisation for Economic Co-operation and Development* September.

Race Earnings Differentials

M Carnoy, Stanford University, Stanford, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Ability differences – Refer to innate ability differences as measured by tests before children enter school, but in most economics literature is usually the measured test performance differences among individuals on a school-based test.

AFQT – Armed Forces Qualifying Test—a test developed to screen individuals for service in the military.

Cognitive skills – Mental skills that are used in the process of acquiring knowledge, such as attention, memory, reasoning, symbolic thinking, perception, and self regulation.

Cross-section data – Data collected at one point in time on a sample of individuals.

Educational quality – Refer to the quantity of schooling inputs per student, the quality of inputs per student, such as teacher content knowledge or teacher experience, or the measured outputs of schooling, such as test score gains from year to year.

Ethnic group – A group of people that shares a self-identity, which can be based on language, national heritage, culture (often mixed with religion), and usually a combination of all these.

Labor market discrimination – Differences among race, ethnic, and gender groups in employment, wages, and promotion due to differential treatment of individual characteristics associated with productivity, such as education, experience, and measured ability.

Longitudinal data – Data collected on a sample of a cohort of individuals over a number of years in a series of follow-up questionnaires.

Noncognitive skills – Skills such as persistence, reliability, self-discipline, the ability to work with others, and the capacity to listen.

Racial group – Although controversial for scientific and political reasons, this usually means a category of individuals based on heritable characteristics such as skin color and facial features, but also on self-identification and social construct.

Test score gap – Refers to the difference in average performance on national tests by identifiable social class, gender, ethnic, and race groups.

Earnings differentials between racial or ethnic groups exist in most societies. To the extent that they persist over time in democratic political conditions, they are of considerable concern, since they suggest that different groups are not getting either equal access to human capital investment opportunities or equal treatment in labor markets, or both.

The literature on such earnings differences (and more generally, unequal economic and educational opportunities) have been of interest to economists for almost 70 years, especially in the United States, where, in 1944, a Swedish economist, Gunnar Myrdal, published the path-breaking *An American Dilemma: The Negro Problem and Modern Democracy*, and where the civil rights movement in the 1950s and 1960s continued to draw attention to the unequal education and income of African-Americans and Latinos. The pressure exerted by the movement enabled the collection of detailed census statistics that helped researchers estimate the existence of wage and schooling differences and to posit explanations for them. In other countries where such data are collected, the results also show distinct differences in economic performance between ethnic and race groups, suggesting that discovering differences is largely a function of data-collection politics.

More recently, in the 1990s, a major debate was held over whether racial/ethnic discrimination in labor markets has disappeared, and whether the main problem is that African-Americans and Latinos earn lower incomes for a given level of education because their academic achievement is lower. This then raises the issue of why different race/ethnic groups achieve differently in school (e.g., the Black–White test-score gap), and how important those differences are in explaining earnings differences.

Although the discussion in this article focuses on the United States, this is primarily because of the richness of the data available and the highly developed debate about the explanation of differences. The methodology is applicable to other economies with minor changes of local conditions. Such a comparison is made with Brazil, with a large Black and mulatto population, and Israel, where European immigrants and those of Afro-Asian origin differ considerably in education and cultural background, and a large minority of Israeli Arabs, representing 19% of the population, that faces major barriers to entry into private and public sector jobs and are discriminated against for ideological reasons.

Race and Ethnic Earnings Differences in the United States

Data on race, ethnicity, and earnings in the United States are available for the period 1939–2006, and can be analyzed by native- and foreign-born percentage (see Chiswick, 1984 for the 1970 census breakdown). It is important to analyze the data by gender since earnings differ by gender across groups, and female labor force participation rates also differ among groups and over time.

The data suggest that certain ethnic groups, regardless of race, earn more than average and more than the dominant White Anglo (non-Latino) majority. Therefore, for example, Japanese Americans, Chinese-Americans, and Jewish-Americans all earn more than Anglo-Whites and this has been the case since at least the 1960s. African-Americans, Latinos, Filipinos, and Native-Americans are all examples of groups that earn considerably less, on average, than Anglo-Whites and have done so for as long as data have been available on such differences.

Among Latinos there is also variation, with Mexican and Puerto Rican Latinos earning much less than those of Cuban origin.

Table 1 presents earnings data for the period 1939–2006 according to ethnic group, race group, and gender group. The Asian-American and Latino groupings are broad and contain distinct subgroups that earn significantly more or significantly less than the group average. Estimates of earnings of Latinos of Mexican origin, Puerto Rican origin, and Cuban origin show that Cuban-origin Latinos earn more on an average, than Latinos of Mexican or Puerto Rican origin. Part of the subgroup variation is due to large differences in the socioeconomic origins of different national groups that immigrated to the United States. For example, the first wave of Cuban Americans who came to Florida in the late 1950s and early 1960s were professionals leaving Castro's Cuba. That group of immigrants was markedly different from the unskilled laborers from Mexico who immigrated throughout the same period, and especially after immigration laws changed in 1965.

Table 1 United States: Annual median earnings by education, race/ethnic group, and gender, 1939–2006 (full-time workers, current dollars, and percent)

Category	Year 1939	1949	1959	1969	1979	1989	1999	2006
All schooling (males, 25–34 years)								
White	1356	3001	5438	8633	17 389	28 578	35 603	40,964
Black	42	55	58	63	70	71	78	77
Latino	62	71	72	76	70	67	62	66
Asian-American	65	76	86	101	96	98	–	124
High school graduates (males, 25–34 years)								
White	1353	3026	5241	8082	14 830	22 288	30 496	34,387
Black	58	73	69	73	80	80	88	88
Latino	88	83	82	85	89	85	73	79
Asian-American	75	85	99	100	86	89	–	93
College graduates (males, 25–34 years)								
White	2719	3760	6788	10 549	18 394	31 279	42 173	50,471
Black	59	72	67	68	82	70	87	80
Latino	99	66	78	76	88	83	90	88
Asian-American	73	85	85	110	91	100	–	112
All schooling (females 25–34 years)								
White	816	2038	3032	4956	10 226	18 613	27 296	34,034
Black	39	57	63	77	91	90	84	81
Latino	61	87	75	90	87	82	76	77
Asian-American	83	89	99	113	109	109	–	119
High school graduates (females, 25–34 years)								
White	935	2212	3470	5121	9523	15 421	20 655	25,058
Black	51	75	73	84	95	90	88	88
Latino	94	103	95	103	93	95	91	92
Asian-American	85	69	100	94	92	119	–	–
College graduates (females, 25–34 years)								
White	1128	2491	4378	6336	12 228	23 732	34 377	40,614
Black	67	89	71	101	95	94	90	91
Latino	–	–	–	99	94	98	94	92
Asian-American	–	–	–	104	110	106	–	109

From US Department of Commerce, Bureau of the Census, Public Use Census Sample, 1940, 1950, 1960, 1970, 1980, 1990, 1996, and 2007.

Yet, despite this variation within broad groups, **Table 1** suggests that: (1) there are important and persistent ethnic/race differences in earnings in the United States labor market, (2) these differences cut across gender, and (3) they are subject to historical change. In the period 1939–79, non-Anglo men and women made large gains relative to Anglos, and after 1979, these gains leveled off, with some notable continued gains (Black male high school graduates), and notable reversals for Black and Latina women. Even with the large gains over four decades, African-Americans and Latinos have remained at a lower level of median income than either Anglos or Asian-Americans. It also should be noted that these data are reported for full-time, full-year workers, so they do not reflect the higher rates of unemployment among minority workers, and that, in real terms (adjusted for inflation), the average earnings reported for White male high school and college graduates, 25–34 years old declined, from US\$14 830 in 1979 to US\$12 369 in 2006 in the case of high school graduates, and from US\$18 394 in 1979 to US\$18 155 in 2006 for college graduates. Female high school graduates also had a 12% decline in real earnings in this period. Thus, except in the case of Black male high school graduates, after the 1970s, Blacks and Latinos generally maintained a constant or declining share of declining real earnings. It is also fairly evident from **Table 1** that once education and age are accounted for, Latinos earn relatively more than African-Americans with a college education, but males earn less at the high school graduate level.

The two important questions that the literature has addressed about these earnings differences, logically, are:

- (1) Why do some groups do better or worse than others?
- (2) What causes them to change over time?

A Model for Explaining Earnings Differences

The human capital model has typically been used to understand race and ethnic differences in earnings (Becker, 1957; Freeman, 1973, 1976; Hanushek, 1981; Welch, 1973; Chiswick, 1984). This model characterizes individual earnings as a function of education and experience in the labor force. The premise is that earnings of various groups would be explained largely, if not entirely, by their average education and experience. This still leaves the issue of why average education differs among different groups. Is the difference voluntary, or the result of discrimination in the supply of educational services? However, leaving this issue aside, if the human capital model is a correct representation of the market for labor, then race or ethnicity should play no significant role once

the human capital of individuals is controlled for, that is, the coefficients of education and experience in the earnings function should be equal of all ethnic groups. The model is as follows:

$$\ln Y_i = a + b_j \sum_{j=1}^m S_{ij} + C_k \sum_{k=m+1}^n E_{ik} + e_i$$

where Y_i = income of the individual i ; S_{ij} = dummy for schooling level j ; E_{ik} = dummy for experience level k ; and e_i = unexplained variance.

Other models are more elaborate, including the possibility that earnings for different groups vary because of employment in different industries (with significantly lower or higher pay), in public versus private employment, different parts of the country, and because of differences in civil status, and in native versus foreign parentage, which could represent English-language skill differences (Carnoy *et al.*, 1976; Farley, 1986; Bean and Tienda, 1988; Carnoy *et al.*, 1990). Time worked per week or per year is also an important potential factor affecting earnings differences, since some groups may have higher average levels of voluntary unemployment and part-time employment than others.

Most recently, in the 1990s, the argument on race/ethnic earnings differences focused on the achievement score differences between these groups even when they have the same number of years of schooling and similar experience in the labor force (Herrnstein and Murray, 1994; Murnane *et al.*, 1995; Neal and Johnson, 1996; Thernstrom and Thernstrom, 1997). Such achievement differences can reflect varying initial endowments, varying family and community investments in children's academic ability, and possible differences in the quality of schooling available to different race/ethnic groups. Some analysts are persuaded that the Black–White and Latino–White test gaps are the main obstruction to Black and Latino economic progress. In their view, the test gap indicates that the cognitive ability of many Blacks and Latinos is weak and therefore the main reason their wages and income lag behind whites. In the human capital explanation, academic ability among those with the same number of years of schooling needs to be added to the model to measure human capital more fully.

We take these arguments step by step; first, we use cross-section data over a long period of time to estimate how much school attainment and other variables (not including academic ability) may contribute to earnings differences between race/ethnic groups. Then we look at studies that use longitudinal data to test the contribution of ability differences to these earnings differences. We also comment on how the contribution of various factors to earnings differences may be changing over time.

Explaining Earnings Differences with Cross-Section Data

Empirical results of cross-sectional studies in a single year show that schooling level attained is a significant correlate of earnings differences among groups. African-Americans, Latinos, and Native-Americans take significantly less schooling than Anglo Whites and, particularly less than, Asian-Americans. The importance of education in explaining differences in income between various ethnic/gender groups can be simulated with log income equations, estimating how much equalizing education differences would equalize income between groups. For each year and each ethnic/gender group shown in **Table 2**, the figure in the table represents the percentage point increase in average income the group would have had if its education were equal to that of White males in the same year. For example, in 1939, Black males would have had 27 percentage points higher income if their education were equal to White males', taking account of age differences. White females, on the contrary, would have had 9 percentage points lower income if their education were equal to White males', implying that that working White females were more educated than males in the 1930s. Adding work-experience differences changes these percentages marginally, except for Latinos in the 1970s, when the Latino labor force became significantly younger than other groups. The additional effect of Latino experience decreased sharply in the 1980s.

Observable factors other than the quantity of schooling and age may explain part of the gap in the earnings of

Black and Latino workers compared to Whites. Some of these factors are industry and region worked in, English-language limitation (foreign birth/language spoken at home), and the quality of education taken by different groups, including differential investment in education at home.

Excluding measures of ability differences, which are not available in the census cross-section census samples, the graphs in **Figure 1** show the percentage points of income difference remaining when all the observable factors (school attainment, work experience, region of work, marital status, foreign/native birth, and industry of work) of each group are equalized by simulation to that of White full-time employed adult males (see Oaxaca (1973), Carnoy *et al.*, (1990), and Carnoy (1994), for the standard model of such a simulation). This residual fell significantly for full-time-employed Black males between 1939 and 1982 and for Latino and Asian-American males between 1939 and 1969. The residual shown in **Figure 1** is measured as percentage points of the earnings gap for each group that is not explained when minorities are made to look like White males educationally and in terms of experience and of place of work. In the 1980s, the residual rose for all three groups. The residual was essentially zero by the end of the 1940s for Asian-Americans, and has been, until the new Latino immigration after 1965, much lower for Latinos than for Blacks. In the 1980s, it represented about 16 percentage points of White male income for Blacks and 13 percentage points for Latinos. It is probable that the residual continued to rise for Latinos and may have fallen for Blacks between 1989 and 2006 (see **Table 1**). Part of the residual for

Table 2 Percentage points of income gain that would result from equalizing minority education to White male education, by ethnicity and gender, full-time employed, 1939–89^a

Year	Latino males	Latina females	Black males	Black females	White females
1939	29	18	27	22	-9
1949	17	12	18	11	-8
1959	15	10	15	9	-2
1969	12	9	14	7	0
1973	16	9	12	5	0
1979	16	8	11	6	2
1982	15	10	11	6	2
1985	15	10	11	6	2
1987	15	8	10	6	1
1989	20	14	10	7	0

^aThe education variable is measured in 1940, 1950, 1960, 1970, 1974, 1980, 1983, 1986, 1988, and 1990; incomes refer to the previous year – hence, the years in the table refer to the income year. The education gap is estimated from a simulation using a regression equation with human capital variables (years of schooling, labor-force experience, and, in census years, native or foreign born). The percentages in the table should be read as the number of percentage points that a given group would have gained just from getting the same distribution of education in its labor force as White males. A negative sign means that White females would receive lower incomes, all other variables equal, were education equalized with that of White males (White females in the labor market in those years had higher education than White males).

From Department of Commerce, United States Census, Public Use Sample. 1940, 1950, 1960, 1970, 1980 and Current Population Survey, 1974, 1983, 1986, 1988, 1990.

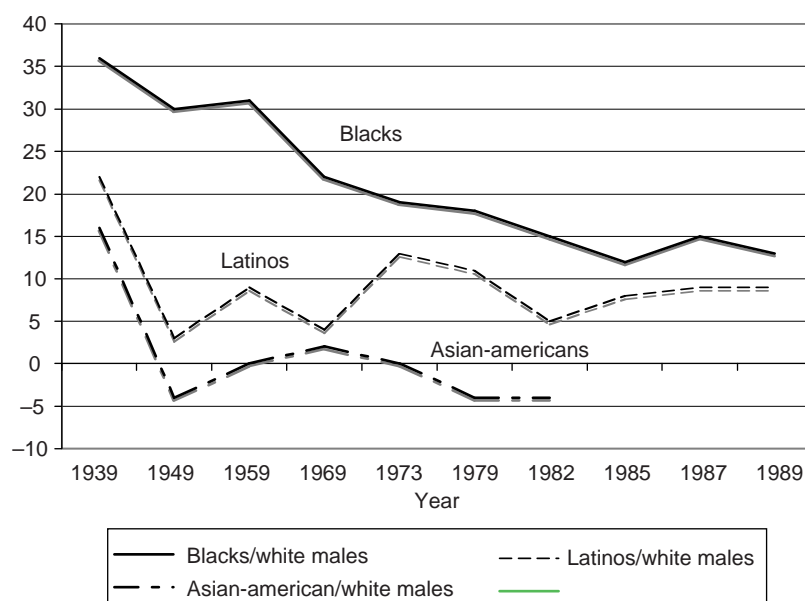


Figure 1 Minority-White income gap explained by wage discrimination (does not include adjustment for ability differences), 1939–89.

Latinos is associated with foreign birth, and this, in turn, may be a function of English-language capability. When Latinos of Mexican origin are separated into native- and foreign-born, the corresponding residuals for the native-born are 3 percentage points in 1960, 3 in 1970, and 6 in 1980, and for foreign-born, 27 percentage points in 1960, 14 in 1970, and 14 in 1980 (Carnoy *et al.*, 1990). The low residuals for native-born Latinos of Mexican origin suggests that the main explanation of lower incomes for Latinos, once English-language capacity is accounted for, is observable differences in educational attainment, experience, and place and type of work.

The residual can be interpreted as reflecting lower-quality education taken by minorities (Welch, 1973; Card and Krueger, 1992a, 1992b), lower returns to experience, such as in the choice of or access to jobs that have a smaller training component, hence flatter experience-earnings profiles, and possibly a strong interaction between the quality of schooling and access to training in jobs. The differences in quality of education received may not only reflect differences in the quality of schools attended by different groups, but also the amount invested in them by their families and the interaction between family background and school performance, which is discussed in detail below. Eliminating discriminatory practices in school access does not necessarily equalize outcomes (the quality of an individual's education) because of the cumulative nature of learning before and during school and also because of interaction effects. Differences in returns to education and in access to jobs with more training for different groups may also simply be the result of discriminatory practices in the labor market.

Why labor market discrimination occurs has been the subject of considerable controversy. In his pioneering work on racial discrimination, Myrdal (1944) argued that it was historically institutionalized in the fabric of society and required public intervention to change. Becker's (1957) model of discrimination focused on the taste for discrimination on the part of White workers and consumers, which drives the earnings of Black workers down relative to White workers. Reich (1982) refutes the claim by Becker that White workers profit when there is discrimination, showing that the lowest White earnings are in those states where the greatest discrimination exists. He argues that labor-market discrimination serves owners of capital, since it keeps all workers' wages lower.

Earnings Differences for Minority Women

Racial and ethnic earnings differentials vary by gender. **Table 1** suggests that full-time-employed minority women in the United States had, by the late 1960s/early 1970s, come to earn approximately the same income as full-time-employed White women when the level of schooling was the same, and, on an hourly basis, may actually have earned more than White women (Chiswick, 1987). By 2006, Black male high school and college graduates earned about 90% and 80% of the annual income of Whites, and Black women graduates earned about 90% of full-time White female incomes at both levels of schooling. Latino men high school graduates earned a much lower fraction of White male incomes than did Latina women relative to

White women, but only slightly lower at the college graduate level – this is the opposite of the pattern for Blacks.

If Black women and Latinas are subject to the same race/ethnic discrimination as men, why should they earn higher relative incomes with compared to their White gender counterparts? The answer lies partly in the gender segregation of the labor market and the willingness of White male employers to group minority and White women in the same category of labor. This was not always the case, and, recently, differences in pay are beginning to emerge again in favor of White women, perhaps for somewhat different reasons than in the past. Black (but not Latina) female earnings were much lower than those of Whites in the 1940s and 1950s, when Black women were highly concentrated in domestic service. As Black women moved into manufacturing and clerical jobs and received much more schooling, the differences between White and Black female earnings tended to disappear, at least until the 1970s. After that time, with the steady increase in the labor-force participation of White women, the ratio of Black and Latina women's earnings to White female earning has gradually declined. Even so, Neal (2004) argues that the measured wage gap between Black and White women continues to be too small because of the failure to correct for differential labor-force participation.

Besides the continued gender segregation of the labor market, one reason that Black women may continue to earn a higher ratio of White salaries than Black men, particularly among college graduates, is that a higher fraction of college graduate Black women work in the public sector, where female salaries are much higher than in the private sector. Public sector employment was especially important in the late 1960s and in the 1970s, when such employment was at its peak.

Do Ability/Educational Quality Differences Explain Race/Ethnic Earnings Differences?

The most recent debates center on whether discrimination has been eliminated or still exists in US labor markets (see, for example, Thernstrom and Thernstrom, 1997 vs. Brown *et al.*, 2003). The discussion centers in part on whether earnings differences not explained by education and experience differences are the result of continued discrimination or of race/ethnic differences in ability as measured by test-score differences (Neal and Johnson, 1996) – in part such test-score gaps are the result of family education differences (initial endowments) and, in part, of differences in the quality of schooling taken by Blacks/Latinos and Whites.

Estimates of the contribution of higher ability, as measured by scores on mathematics tests taken in high

school, to individual earnings later in life suggest that its value is significant even when years of schooling are accounted for (Murnane *et al.*, 1995), and that the return to academic skills may have increased in the 1980s (Blackburn and Neumark, 1993; Murnane *et al.*, 1995). (However, this is not a universally held view. Some economists, such as Samuel Bowles and Herbert Gintis have long argued that cognitive skills explain relatively little of the variance of earnings and productivity (Bowles and Gintis, 1975; Bowles *et al.*, 2001)).

Those estimates paid little attention to the role of race/ethnicity, but Carnoy and DeAngelis (1999) used the National Longitudinal Survey of Youth (NLSY) and High School and Beyond data to estimate the role of mathematics achievement and race/ethnicity in wages by gender group for the 1972 and 1982 high school senior cohorts in various years after graduation. They made these estimates separately for high school graduates, those who had completed some college, and for college graduates. They found that in the 1972 cohort, mathematics test score in high school had little direct effect on the wages of high school male graduates. Being Hispanic had no significant effect on earnings, but being Black had a strong negative effect; about 20–25% lower wages when education, mathematics achievement, and work experience were controlled for. The race effect was somewhat smaller but significant for those with some college, but was negligible for college graduates, whereas one standard deviation higher mathematics test score was worth 3–4% for those males with some college and 8–9% higher wages for college graduates. Higher mathematics score had a similar size effect for female college graduates, but neither race nor ethnicity had a significant relationship to wages.

For the 1982 cohort, only annual earnings (not wages) were available as the dependent variable. For male high school graduates, mathematics score is related to earnings in the 1991 follow-up, but so is race (being Black), although less so (10–18% lower earnings) than for the 1972 cohort. For those males with college education, race is unrelated to earnings when education and mathematics score is accounted for, but math score is related to earnings for college graduates to about the same degree as in the 1972 cohort. For females with college education, mathematics score is related to higher earnings, but race is generally not.

This analysis suggests that the labor market for high school graduates placed relatively less value on ability differences in the 1970s and 1980s and paid Black males considerably less even when all these human capital variables were accounted for. This did not appear to be the case for Hispanics with high school degrees or Hispanic/Black college graduates even in the 1970s, and by the 1980s and early 1990s, was not the case for even those males with some college. For females, race and ethnicity

do not seem to have been correlated with wages or earnings even in the 1970s.

Thus, race discrimination for Blacks seems to have been important in the labor market for males with lower levels of schooling even in the 1980s and early 1990s, whereas it seems to have played little role in the labor market for college-educated males by the late 1980s and for females with both high school and college education. For the latter groups, mathematics ability was related to earnings; so if Blacks and Latinos had lower mathematics ability, they would have received somewhat lower earnings (about 10% lower earnings for one standard deviation difference in mathematics test score). If we extrapolated these results to the earnings differences in **Table 1**, most of the differences observed in Black/White male high school graduate earnings in 1989 and 1999 could be attributed to racial discrimination, and most of the observed Black/White differences in college graduate earnings for males and females, to mathematics achievement differences.

An analysis of yet another data set, the NLSY, which includes the armed forces qualifying test (AFQT) as a measure of ability among a sample of 16–23-year-olds in 1979, Carneiro *et al.* (2005) explores wage differences between Blacks and Whites and Hispanic and Whites, males and females, in that sample followed over the period 1991–2000. When the AFQT scores are corrected for the level of schooling attained by the test taker in order to get a purer ability effect, one standard deviation higher AFQT score is associated with 9–10% higher male wages whereas being a Black male is associated with 14–19% percent lower wages than White males. Once AFQT scores are taken into account, Hispanic and White males do not have significantly different wages. Being a Black woman is associated with 7–8% lower wages than White women earn when ability is taken into account, and being a Hispanic woman is associated with a wage premium of 7–14% compared to White women. The payoff to ability is somewhat higher than for men, about 9–13% for one standard deviation higher AFQT score. These results are generally consistent with the results for the High School and Beyond Survey, but the test-score effect is somewhat larger in the NLSY, and being Black seems to have a negative effect on men and women's wages, although the coefficient is not large for women.

What Are the Sources of the Black/White Test-Score Gap?

Although it is beyond the scope of this article to provide a comprehensive review of the literature on the Black/White test-score gap, the overall question of race/ethnic difference in earnings merits at least some discussion of whether Blacks and Hispanics choose to invest less in

themselves academically, hence achieve less, or whether the schools they attend and the society around them are the main source of their lower achievement – hence, it is the civil society and state that makes them less productive (this would be akin to the discrimination argument in labor markets).

Earlier Jencks and Phillips (1998) and more recently Rothstein (2004) summarized the empirical research on this issue. Rothstein argues that much of the achievement gap between Blacks and Whites exists at age 5 years when children enter kindergarten. His main point is that the achievement gap is largely the result of differences in the social context in which most Blacks and Latinos live in the United States and that the potential for schools to overcome achievement gaps is relatively limited. Thus, he concludes that a more productive strategy to reduce the achievement gap would be improving minorities' social conditions – nutrition, healthcare, and sense of being full participants in American society.

Part of those conditions is also the degree to which Blacks feel discriminated against in labor markets. A number of opinion surveys suggest that Whites believe that the United States is free of racial discrimination, and Blacks believe that despite great improvement in race relations over the past 40 years, racial discrimination is still a major feature of their lives (Brown *et al.*, 2003). This could influence their desire to perform well in school and their test performance (Steele, 1997).

Evidence from national- and state-level data on whether the Black–White test score changes during schooling is contradictory:

Data from [longitudinal] ECLS-K and SECCYD suggest the gap is large at the start of kindergarten, and grows in the early elementary grades (particularly from first to third grade in ECLS-K), though the patterns differ somewhat depending on the gap metric used. Data from NAEP suggests that the gap continues to grow from age 9 to 13 (fourth to eighth grades, roughly), but [longitudinal] state-level data from Texas and North Carolina seem to contradict this finding, at least during the late 1990s and early 2000s, suggesting that the gap grows relatively little in standard deviation units over the latter half of elementary school. Finally, data from NAEP and NELS suggest the gaps change relatively little following eighth grade, though there is some uncertainty in these estimates, since most are based on analysis of repeated cross-sectional data (Reardon, 2007: 8).

Reardon (2007) does a careful analysis of changes in the Black–White test-score gap between kindergarten and fifth grade and concludes that the gap in both math and reading scores widens between kindergarten and fifth grade, and that the gap grows faster among higher-scoring students, suggesting that part of the widening gap may result from "... black students with high skill levels in kindergarten

being more likely than equally high-scoring white students to be in schools where the median skill level is far below their skills (because they are more likely to be in schools with predominantly black student populations)” (Reardon, 2007: 31). Thus, the increasing Black–White test-score gap might be the result of poorer schools, but schools whose quality is derived from the concentration of low-income students, in turn the result of residential segregation. Even so, the Black–White test-score gap at kindergarten entrance explains the bulk of the gap at grade five.

Carneiro *et al.* (2005) also argue that early environment rather than poor schooling is likely the most important reason that Blacks have lower test scores (academic skills) than Whites.

Minority deficits in cognitive and non-cognitive skills emerge early and widen. Unequal schooling, neighborhoods and peers may account for this differential growth in skills, but the main story in the data is not about growth rates but rather about the size of early deficits. . . . The failure of the Hispanic-white gap to widen with schooling or age casts doubt on the claim that poor schools and bad neighborhoods are the reasons for the slow growth rate in black test scores. (Carneiro *et al.*: 21).

Explaining Changes in Earnings Differences over Time

Why has the gap in earnings between Whites and minority groups changed over time? **Table 1** shows that full-time-employed Black males and females made large gains in relative earnings in the period 1939–79, especially in the 1940s and 1960s. The gains in the 1980s were negligible (the relative earnings of Black male college graduates fell), and Black males high school graduates made some gains in the 1990s, whereas Black female college graduates saw a decline in their earnings relative to White females. Latino males and females also made big gains in the period 1939–69, but then lost ground, particularly male high school graduates. Asian-Americans made large gains in the 1940s, 1950s, and 1960s, and college graduates continued to make gains in the 1990s and early 2000s.

The principal debate that has developed over these changes is whether they are the result of supply side forces, specifically changes in the relative investment in human capital made by the different groups, or demand side forces, specifically legal and direct employment intervention by government. Freeman (1976) maintained that the large gains in relative earnings of Blacks in the 1960s came as a result of the passage of the 1964 Civil Rights Act and subsequent federal employment legislation in 1965. Others (Heckman and Payner, 1989; Donohue and Heckman, 1991; Card and Krueger, 1993) have supported this explanation. Smith and Welch

(1989), on the contrary, have argued that the single best explanation for the gains in the period 1939–79 was the relative increase in Black education, and that federal intervention had only a minor effect, primarily in raising the earnings of young Black college graduates in the early 1970s. Card and Krueger (1992b) claim that about one-fourth of the gain in Black earnings attributed to government intervention in labor markets in the 1960s and 1970s is the result of improvements in Black education in the South in the 1930s and 1940s. Juhn *et al.* (1991) argues that the slowdown in Black–White convergence in the 1980s was due to unmeasured skill effects, namely that the payoff to educational quality increased in the 1980s, and Blacks still received lower-quality education than Whites.

If the analysis is done decade by decade in the 50-year period 1939–89, the supply-side argument is weakened. **Table 2** suggests that education gains by Blacks between 1939 and 1979 were correlated with a reduction in the Black–White earnings gap in those four decades. But other factors appear to have been much more important than education in that reduction. For example, **Table 3** corrects that contribution of education differences to the income gap estimated in **Table 2** for changes in the income distribution (specifically the income weights attached to different levels of schooling) in each decade. Wage compression in the 1940s was much more important than Black male educational gains in reducing the

Table 3 Change in education gap adjusted for changing income weights, by decade, full-time workers, 1939–89 (percentage points of income)

Year	Latino males	Latina females	Black males	Black females
1939	29	18	27	22
1949	23	16	24	15
1949	17	12	18	11
1959	16	10	16	9
1959	15	10	15	9
1969	12	9	14	7
1969	12	9	14	7
1979	16	8	11	6
1979	16	8	11	6
1989	19	13	9	7

From See **Table 2** for figures based on the education gap estimated using same-year White male income weights. Figures in this table compare the education gap in 1939 with the gap in 1949 figures based on average education differences between ethnic groups in 1949 weighted by White male coefficients estimated using 1939 census data; compare the 1949 figures from **Table 2** with 1959 figures based on average education differences in 1959 weighted by White male coefficients using 1949 census data, and so forth, for various census years. This permits comparison for each decade of the effect on income of the education gain net of changes in the distribution of income among education groups in that particular decade.

Black–White income gap, and was as important as education for Latinos. Although Black males' education did not increase much relative to Whites in the 1940s and racial discrimination declined only slightly, the enormous increases in incomes for all low-income earners relative to high-income earners – irrespective of race – during that decade lifted Black male relative incomes more than in any decade since. In the 1940s, there was also a job shift from agriculture to manufacturing that had a power effect on Black male incomes. As a point of comparison, in the 1950s – a period of significant educational gains for minorities but no reduction in overall income inequality, much-reduced sectoral job shift, and no reductions in job discrimination – gains in relative incomes for Blacks were minimal. In the 1960s and particularly, the 1970s, Blacks made large education gains, but the major gains in income for Blacks came from the reduction in wage discrimination. In the 1980s, when the political climate on affirmative action changed, Black male and female incomes stopped rising relative to Whites' (particularly for Black male college graduates), and in the 1990s, when the climate became more favorable again, Black male relative earnings rose again. Another possible influence on Blacks' relative incomes in the 1990s may have been the relative increase in fourth- and eighth-grade mathematics test scores in the late 1970s and 1980s, which could have been reflected in the higher relative test scores of both 25–34-year-old high school and college graduates in the 1990s and early 2000s.

A similar analysis for Latinos and Asian-American males and females suggests that education is a more important explainer of their relative income gains than for Black and White females, although changes in sector of work (shifts from agriculture to manufacturing) are also crucial in understanding their gains in the 1940s (Latinos and Asian-Americans) and 1950s (Latinos). Both groups also profited from reduced labor-market discrimination in the 1940s and 1960s.

The recent work focusing on the relation to relative wages of pre-market skill differences between minorities and Whites has been interpreted by many as a mandate to improve the quality of education taken by Blacks and Hispanics (e.g., by Thernstrom and Thernstrom, 2003). However, both Carneiro *et al.* (2005) and Rothstein (2004) make convincing cases that most of the ability differences between Blacks and Whites as measured by teenage test scores of those with similar levels of schooling occur before children enter school and persist through the schooling process. This is apparently not the same pattern for Hispanics, who are able to close the ability gap as they acquire more schooling. Furthermore, almost all studies with longitudinal data show that even in the 1990s, at least for Blacks, education and test-score differences compared to Whites do not explain even part of Black–White wage differences, particularly for males.

Comparing Results for Brazil and Israel

Brazil has a large Black (*preto*) and mulatto (*pardo*) population. For many years, the official ideology in Brazil was that its racial inequality was softened by the acceptance of intermarriage. In the past 10 years, however, research has identified large earnings differences between Whites and both *pretos* and *pardos* in the Brazilian labor market. The most recent of these studies uses 1996 household survey data for urban male workers and a rough proxy for the educational quality received by the individual worker (teacher–pupil ratios by state of birth and year of birth) to estimate wages of male household heads as a function of education, age, parents education, and industry worked in (Arias *et al.*, 2004). The study finds that the bulk of wage differences between Whites and workers of color is explained by productivity-related worker characteristics, but unexplained wage differences remain large at higher-income levels. In the fifth quintile of income, the residual wage gap is 15% for *pretos* and 12% for *pardos*, and this increases in the top-income decile to 25% for *pretos* and 16% for *pardos*. This is reflected in lower rates of return to education for males of color than for Whites. Nevertheless, whereas *pretos* face a larger gap in education returns at the top of the income distribution, the opposite is true for *pardos*. Apparently, the common belief in Brazil that in higher socioeconomic levels, interracial marriage softens racial discrimination probably holds true.

Similar estimates have been made for Israel, where ethnic origin (European-origin vs. North African-Asian-origin immigrants) and being an Israeli Arab are potentially a source of wage discrimination (Amirs, 1983, 1988; Margalioth, 2005). The results of such estimates show that in the 1970s, the wage differential between the foreign-born males of the two ethnic groups in the Jewish population fell substantially from 25% to 16% mainly because of demographic factors such as a change in the relative age structures and period of immigration of the two groups. Unexplained differentials (wage discrimination plus quality of schooling differences) also fell from 15% to 12%. When the same estimates are made for first generation Israeli-born of North African-Asian and European extraction, the wage differential is shown to have become larger than for the foreign-born in the 1970s, rising from 26% in 1970–72 to 30% in 1980–82. This is explained mainly by an increasing age difference over the decade, with European-origin Israelis becoming increasingly older in relative terms. Nevertheless, the results also show that educational differences (larger than between the foreign born of the two ethnic groups) continued to persist even though average levels in both groups rose significantly relative to their parents. The results also show that wage discrimination fell during the decade and was relatively low (about 6% in 1980–82). Other estimates of wage differences between Israeli Arabs and

Jews suggest that it is mainly the segregation of the labor market that contributes to the wage gap between the two groups. The gap is quite large: Israeli Arab employees earned about 28% less than Jewish employees. Arabs are highly overrepresented in low-paying industries such as agriculture and construction, whereas they are underrepresented in financial and business services (Margalioth, 2005). Although education differences, particularly a lower likelihood that Israeli Arabs attended university, contribute to wage differences, there is considerable evidence that an important source of wage differences is discrimination against Arabs in access to jobs on security threat and ideological grounds. Arabs also face a number of other barriers to entry, such as the requirement that potential employees have completed military service (Arabs normally do not serve in the Israeli defense forces).

See also: Desegregation, Academic Achievement and Earnings; Education and Inequality; Human Capital; Returns to Education in Developed Countries; Returns to Education in Developing Countries; School Quality and Earnings.

Bibliography

- Amirs, S. (1983). Educational structure and wage differentials of the Israeli labor force in the 1970s. *Discussion Paper No. 83.07*. Jerusalem: Bank of Israel.
- Amirs, S. (1988). Trends in wage differentials between Jewish males of different ethnic origin during the 1970s. *Bank of Israel Economic Review* 63, 52–75.
- Arias, O., Yamada, G., and Tejerina, L. (2004). Education, family background, and racial earnings inequality in Brazil. *International Journal of Manpower* 25(3/4), 355–374.
- Bean, F. and Tienda, M. (1988). *The Hispanic Population of the United States*. New York: Sage.
- Becker, G. S. (1957). *The Economics of Discrimination*. Chicago, IL: University of Chicago Press.
- Blackburn, M. and Neumark, D. (1993). Omitted-ability bias and the increase in the return to schooling. *Journal of Labour Economics* 11, 521–544.
- Bowles, S. and Gintis, H. (1975). *Schooling in Capitalist America*. New York: Basic Books.
- Bowles, S., Gintis, H., and Osborne, M. (2001). The determinants of earnings, a behavioral approach. *Journal of Economic Literature* 39, 1137–1176.
- Brown, M., Carnoy, M., Currie, E., et al. (2003). *Whitewashing Race: The Myth of a Color-Blind Society*. Berkeley, CA: University of California Press.
- Card, D. and Krueger, A. (1992a). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy* 100(1), 1–40.
- Card, D. and Krueger, A. (1992b). School quality and Black/White relative earnings: A direct assessment. *Quarterly Journal of Economics* 107(1), 151–200.
- Card, D. and Krueger, A. (1993). Trends in relative Black–White earnings revisited. *American Economic Review* 83(2), 85–91.
- Carneiro, P., Heckman, J., and Masterov, D. (2005). Labor market discrimination and racial differences in premarket factors. *Discussion Paper 1453*. Institute for the Study of Labor (IZA). Bonn, Germany.
- Carnoy, M. (1994). *Faded Dreams*. New York: Cambridge University Press.
- Carnoy, M., Daley, H., and Hinojosa, R. (1990). *Latinos in a Changing Economy*. Interuniversity Program for Latino Research. New York: City University of New York.
- Carnoy, M. and DeAngelis, K. (1999). Does ability influence individual earnings, and if so, by how much? School of Education Stanford University, Stanford, CA (mimeo).
- Carnoy, M., Girling, R., and Rumberger, R. (1976). *Education and Public Sector Employment*. Stanford, CA: Center for Economic Studies.
- Chiswick, B. R. (1984). Differences in education attainment among racial and ethnic groups: Patterns and preliminary hypotheses. *Paper Presented at the National Academy of Education Conference on the State of Education*. Washington, DC (mimeo).
- Chiswick, B. R. (1987). Race earnings differentials. In Psacharopoulos, G. (ed.) *Economics of Education: Research and Studies*, pp 232–237. Oxford: Pergamon.
- Donohue, J. III and Heckman, J. (1991). Continuous versus episodic change: The impact of civil rights policy on the economic status of Blacks. *Journal of Economic Literature* 29, 1603–1643.
- Farley, R. (1986). Assessing Black progress: Employment, occupations, earnings, income, poverty. *Economic Outlook USA* 13(3), 14–23.
- Freeman, R. (1973). Decline of labor market discrimination and economic analysis. *American Economic Review* 63(2), 280–286.
- Freeman, R. (1976). *Black Elite: The New Market for Higher Educated Black Americans*. New York: McGraw-Hill.
- Hanushek, E. A. (1981). *Sources of Black–White Earnings Differentials*. Stanford, CA: Stanford University. Institute for Educational Finance and Governance.
- Heckman, J. and Payner, B. (1989). Determining the impact of federal antidiscrimination policy in the economic status of Blacks: A study of South Carolina. *American Economic Review* 79(1), 138–177.
- Hernstein, R. and Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.
- Jencks, C. and Phillips, M. (1998). *The Black–White Test Score Gap*. Washington, DC: The Brookings Institution.
- Juhn, C., Murphy, K., and Pierce, B. (1991). Accounting for the slowdown in Black–White convergence. In Kisters, M. (ed.) *Workers and Their Wages: Changing Patterns in the United States*, pp 107–143. Washington, DC: American Enterprise Institute Press.
- Margalioth, S. R. (2005). Labor market discrimination against Arab Israeli citizens: Can something be done? *International Law and Politics* 36, 845–884.
- Murnane, R., Willet, J., and Levy, F. (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics* 77(2), 251–266.
- Myrdal, G. (1944). *American Dilemma: The Negro Problem and Modern Democracy*. New York: Harper and Row.
- Neal, D. (2004). The measured Black–White wage gap among women is too small. *Journal of Political Economy* 112, S1–S28.
- Neal, D. and Johnson, W. (1996). The role of premarket factors in Black–White wage differences. *Journal of Political Economy* 104, 869–895.
- Oaxaca, R. (1973). Male–female wage differentials in urban labor markets. *International Economic Review* 14(3), 693–709.
- Reardon, S. (2007). Thirteen ways of looking at the Black–White test score gap. Stanford, CA: Stanford University School of Education.
- Reich, M. (1982). *Racial Discrimination*. Princeton, NJ: Princeton University Press.
- Rothstein, R. (2004). *Class and Schools*. New York: Teachers College Press.
- Smith, J. and Welch, F. (1989). Black economic progress after Myrdal. *Journal of Economic Literature* 27(2), 519–564.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *The American Psychologist* 52(6), 613–629.
- Thernstrom, A. and Thernstrom, S. (2003). *No Excuses: Closing the Racial Gap in Learning*. New York: Simon and Schuster.
- Thernstrom, S. and Thernstrom, A. (1997). *America in Black and White: One Nation, Indivisible*. New York: Simon and Schuster.
- Welch, F. (1973). Black–White differences in returns to schooling. *American Economic Review* 63(5), 893–907.

Returns to Education in Developed Countries

M Gunderson and P Oreopoulos, University of Toronto, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Ability bias – The bias to the returns to schooling that can result from the fact that people who acquire more education may have greater innate skills that would allow them to earn more even without additional schooling.

Causal returns – The returns to education that are induced or caused by additional education rather than simply correlated or associated with additional education.

Endogeneity of education – The fact that education is a decision variable in that the amount of education acquired may be a function of factors such as ability, motivation, family background, income, proximity to school, and compulsory school laws.

Instrumental variables – In the context of education decision making, instrumental variables are variables that affect the amount of education acquired but do not affect the education outcomes or the returns to education (e.g., compulsory school laws or proximity to schools).

Measurement error – The possibility that collected data, like education, may be measured with error since people may not accurately report their education.

Returns to education – The financial rate of return to investing in an additional year of schooling, obtained by comparing the additional earnings from an additional year of education with the cost of acquiring the additional education; it shows how average earnings increase with added education.

Selection bias – In the context of returns to education, it is the bias that can be created by the fact that education may be a function of conventionally unobserved factors such as ability or motivation.

Sheepskin effect – The credential effect or additional returns associated with the credential of completing key phases of education like graduating from high school or university (sheepskin was used historically to make the parchment for diplomas).

importance. What are the private returns that individuals can expect from investing in education? How do those returns vary by factors such as level of education, field of study, and individual background characteristics? How have those returns varied over time and across different countries? Is there an extra effect from a year of education if that year provides the credential of completing a phase of study such as graduating from high school or university? If potential dropouts are compelled to stay in school longer by compulsory school laws, do they receive returns that are higher or lower than the average returns? Are the returns the result of education enhancing the productivity and skills of individuals or are they the result of signaling of such conventionally unobserved factors such as ability, motivation, and time-management skills? What are the appropriate methodologies for estimating the returns to education, especially for dealing with factors such as measurement error, ability bias, credential effects, and financial constraints?

The purpose of the article is to address these practical and methodological questions. The emphasis here is on the causal returns to education after controlling for other observable and unobservable factors such as innate ability or motivation that may affect the outcomes associated with higher education. Understanding the underlying causal relationship process is important for policy purposes so as to ascertain the effect of policy interventions, for example, to reallocate resources from fields of low returns to fields of high returns or raise the age of compulsory schooling or institute policies to deter dropping out. It can also be important for predicting future changes as the underlying causal factors change.

The article moves from the simple to the more complex. It starts with estimates of the return to education based on basic schooling equations where education is not exogenous but can be correlated with other factors that can affect outcomes. It then moves to a discussion of refinements to the basic model: the appropriate measure of earnings and the inclusion of nonwage benefits; measurement error in the schooling variable; corrections for ability bias, omitted variables, and selection bias; and the possibility of heterogeneous returns and credential or sheepskin effects.

Introduction

Understanding the causal relationship between education and the financial returns to such education is important for addressing a range of questions of practical and policy

Estimating Returns to Education via Schooling Equations

A wide range of methodological issues are associated with estimating the economic returns to education. (Reviews of

many of these issues include Card (1999, 2001); Chamberlain (1977); Chamberlain and Griliches (1975); Griliches (1979); and Lemieux (2002). Heckman *et al.* (2006) provide a critical review of much of the literature, emphasizing the heterogeneous returns to education and the importance of psychic costs in explaining such heterogeneous returns. Here, we have cited articles that illustrate the issues and that contain references to related articles.) In this section, the main methodological issues are outlined in a nontechnical fashion, generally referencing more technical treatments of the issues.

Basic Schooling Equation

Estimates of the private returns to education essentially build on the human capital earnings function of Mincer (1974) where the (natural) log of earnings is regressed on years of education and other control variables including years of labor-market experience. The latter is often entered in a quadratic form to capture the nonlinear relationship whereby earnings tend to advance rapidly for early years in the labor market, flatten in later years, and decline slightly thereafter. Higher-order polynomial functions for experience have also been recommended so as to better capture the more rapid earnings growth early in an individual's career and the slower decline in wages later in an individual's career, although Heckman *et al.* (2006: 333) indicate that such higher-order polynomials did not improve their estimates.

The estimated coefficient on the education variable has a convenient interpretation as the average percent increase in earnings from an additional year of schooling (e.g., a number like 0.10 or 10% which can be compared to the returns to other investments). For interpretation, Mincer and other social scientists often assume that tuition and psychological costs from schooling are negligible, individuals do not work much while in school, and schooling and years of experience have separate effects on earnings. Under these assumptions, the coefficient from the Mincer equation can be interpreted as the return to investing in the cost of an additional year of education and compared to alternative investments. In this case, the monetary benefits of an additional year of schooling are the additional earnings from such schooling, while the costs are the forgone income. Since both the benefits and the (opportunity) costs are in this way factored into the estimates of the earnings equation, the coefficient on the education variable yields an internal rate of return to investments in education. (Heckman *et al.* (2006) provide more detailed discussion on estimating internal rates of return from schooling and alternative approaches to assessing returns from education when the mentioned assumptions do not hold.) Since data sets often have different categories of highest level of education achieved, these are often entered in place of years of schooling so that the

returns can vary by different categories of completed education.

Estimates from a basic Mincer schooling equation tend to yield estimates around 0.07–0.10, being slightly higher for females and lower for males. The returns are slightly higher for general academic streams compared to technical vocational streams, and they are higher in the more professional fields such as engineering, medicine, business, and sciences and lower in social sciences and humanities and especially in fields such as fine arts. These can be thought of as the simple benchmark returns to education against which to gauge the effect of the myriad of procedures (discussed subsequently) to improve on those estimates and to consider how returns differ across particular groups of individuals.

Hourly Wages Versus Measures That Include Hours of Work

The appropriate measure of earnings is one that approximates the hourly wage so as to reflect the productivity effects of education and to control for differences in hours worked given that persons with higher education tend to work longer hours. To the extent, however, that higher education leads individuals to work more hours, the additional time is an endogenous part of the returns to education. Measuring increased earnings per hour may therefore underestimate the true returns to education over a fixed period of work. Card (1999: 1809), for example, estimates that slightly more than two-thirds of the returns to education based on annual earnings in the United States in the mid-1990s reflects higher wages while one-third reflects longer hours. More specifically, he estimated returns to education of about 10% for males and 11% for females based on hourly wages, and 14.2% for males and 16.5% for females based on annual earnings. The fact that the change was higher for females than males highlights the facts that higher education is also associated with longer hours of work (both hours per week and weeks per year in his data) and the effect on longer hours was greater for females than males.

Measurement Error in Schooling

Returns to schooling are typically estimated from survey data where individuals report their highest level of schooling. This reported schooling can be subject to measurement error or misreporting of education. Estimates indicate that about 10% of individuals misreport their level of education, and this is true in administrative data as well as survey data. (Misreporting is discussed, for example, in Ashenfelter and Rouse (1998) and Card (1999)). If the misreporting is random or unrelated to the level of education, then such classical measurement error leads to a downward bias in the estimated returns to

education. However, if the measurement error is systematically related to the level of education, then the bias can go in either direction. Persons with low levels of education may be prone to overstate their actual education and persons with higher education may be more accurate in their reporting, yielding an upward bias to the returns to education. However, there is also the possibility of higher-educated people having more opportunities to inflate their education given the multiplicity of different types of degree-granting institutions. In essence, the biases from measurement error in schooling can go in either direction. Overall, based on his assessment of the literature, Card (1999: 1834) concludes that measurement error in education leads to a downward bias in returns to education, with the estimated returns understating the true returns by about 10%. That is, if the estimated returns were 0.10, the true returns would be 0.11.

Ability Bias, Omitted Variables, and Selection Bias

The potentially most severe bias that can occur in estimating the causal returns to education occurs because educated people can have other characteristics that are associated with higher earnings, and those other characteristics are not controlled for in the estimating procedures. Indeed, models that attempt to explain differences in school attainment often do so by noting that costs and benefits from additional schooling are not the same for everyone. Individuals may differ by innate ability, motivation, organizational skills, entrepreneurship, time-management skills, and willingness to work hard. To the extent that these factors lead to higher earnings as well as higher education, and they are not accounted for in the statistical analysis, then omitting them from the estimating equation means that some of the higher returns to education may be reflecting the effect of these factors. That is, the estimated returns to education are biased because the higher education is capturing the economic returns to these omitted variables as well as the pure causal effect of education. Alternatively stated, higher-educated people may be a select group in terms of not only observable characteristics that can be controlled for in the regression analysis but also unobserved traits as indicated above that are not conventionally controlled for in the analysis. The returns to education can reflect a return to these traits as well as to education itself.

The literature on estimating the causal returns to education has been a growth industry in recent years based largely on devising ways to control for this ability or selection bias, often involving imaginative ways of obtaining exogenous variation in education that is independent of ability or selection bias. The following illustrate such procedures.

Include proxy measures of ability

A number of empirical studies have been able to include proxy measures of ability such as IQ scores or test scores designed to measure innate ability. (Studies that deal with test scores as a measure of ability are referenced in Card (1999, 2001) and Griliches (1977)). Such studies tend to find the ability bias to be small in that the estimated return to education drops only by about 10% (e.g., from 0.10 to 0.09) after controlling for the effect of ability.

Family characteristics such as the education of a parent or sibling are also sometimes included to control for factors that may help a person obtain more education and affect their earnings. Such studies also generally find the return to education to drop very slightly (by around 0.01) after controlling for such family characteristics. (Family background controls are used in Ashenfelter and Rouse (1998); Ashenfelter and Zimmerman (1997); and Card (1995b)). Studies that have also examined how the return to education varies by the ability of the individual or his/her family background have yielded inconclusive results (Ashenfelter and Rouse, 1998, and the literature cited therein).

Twin studies

Another way to control for ability bias, and perhaps some of the other potentially important omitted variables, is to use twins since they presumably have the same natural ability (especially if they are identical twins from the same egg as opposed to fraternal twins from two different eggs). Differences in their education are assumed to occur for random reasons (a possibly questionable assumption), and in this way this procedure approaches the ideal random-assignment procedure for estimating treatment effects (in this case the treatment being more education). Using same-sex twins also controls for the possibility that parents may favor one sex or the other in devoting family resources to them to improve their labor-market outcomes.

Studies that utilize differences in education between twins to identify education differences while controlling for ability and other differences generally lower the return to education slightly (suggesting a slight upward ability bias) but this tends to be offset by the measurement error bias from mis-measuring education so that on net the true returns are about the same as those estimated in the conventional regression of earnings on education without controlling for ability bias or measurement error. Estimates of returns to schooling using US twins generally range between 0.06 and 0.12. (Earlier twin studies are reviewed in Griliches (1977, 1979) with more recent twin studies reviewed in Ashenfelter and Rouse (1998) and Miller *et al.* (2006)).

Natural experiments based mainly on features of the education system

A number of empirical studies have used institutional features of the education system or the environment to generate differences in education that arise for reasons

beyond an individual's control. A policy change that lowers the cost of college in one state, for example, affects some individuals but not others depending on when and where they are born. Forces that cause differences in education for reasons outside individuals' control are called exogenous. Using exogenous forces to estimate returns to schooling helps address ability bias because differences in education that arise from exogenous forces are unlikely due to differences in individual ability. Returns-to-schooling estimates from this approach apply only to individuals affected by the exogenous force (e.g., policy change). (Such studies are reviewed in Ashenfelter *et al.* (1999); Ashenfelter and Rouse (1998); Card (1999, 2001); and Carnoy (1997) and critically assessed in Heckman *et al.* (2006)). Exogenous variation can be caused by various factors:

- Geographic proximity to educational institutions can generate exogenous variation in education in that those close to a university are more likely to attend university and hence acquire more education than those who are far away from a university (e.g., Cameron and Taber, 2004; Card, 1995).
- Differences across regions or over time in financial costs to attending school (e.g., through tuition) can also lead to differences in education attainment (e.g., Kane and Rouse, 1995; Chen, 2009)
- The Vietnam draft lottery generated exogenous increases in schooling because many persons enrolled in school in order to defer military service (Angrist and Krueger, 1994).
- Local labor-market earnings for persons at the age of 17 years have generated exogenous variation in education in that higher earnings may increase the opportunity cost of education and thereby induce dropping out (Cameron and Taber, 2004).
- The GI bill in Canada gave rise to exogenous variation in education in that the cohort of males from English-speaking Ontario received additional education due to the GI bill and not to the decision, say, of higher-ability people to acquire more education. The earnings of this treatment group were compared to the earnings of a control group from French-speaking Quebec who were less likely to have served or to have taken advantage of the bill (e.g., Lemieux and Card, 2001).
- Day of birth can interact with compulsory school laws. For example, persons born earlier in a year (e.g., January) tend to have less education than those born later in a year (e.g., December) because they are older when they start school (from missing the school entry age cutoff) and hence reach the compulsory school-leaving age with a lower level of education, many of whom then drop out (e.g., Angrist and Krueger, 1991).
- Differences in compulsory school laws and changes to these laws over time can also generate differences in

education attainment since some persons in jurisdictions with higher ages at which it is compulsory to remain in school will acquire more education because they cannot drop out until the compulsory age (e.g., Oreopoulos, 2006a, b).

- Exogenous variation in education has also been found in situations where an additional year of schooling has been added or subtracted to the high school or university curriculum (Webbink, 2007; Krashinsky, 2007).

Studies using the natural experiments from features of the education system or environment to obtain exogenous variation in education tend to find such causal returns to education to be in the neighborhood of 0.06–0.15, and sometimes higher. As emphasized in Heckman *et al.* (2006: 392), however, the returns are often imprecisely estimated, with large standard errors. These estimates are somewhat higher than the returns when conventional years of schooling are used.

Heterogeneous Returns

It is important to view returns to schooling as being individual and context specific. Gains from schooling depend upon the individual's background, motivation, and the quality and type of schooling. An individual's decision to take more schooling depends on both expected gains and costs. Estimates of returns to schooling using exogenous policy variation can often be interpreted as average returns to schooling among individuals affected by the policy variation. Sometimes the policy variation identifies particularly interesting parameters, such as the average gains to schooling among those forced to stay on in school because of more restrictive compulsory schooling laws or the average gains to schooling among those who entered college because tuition costs were lowered. As discussed in Card (1999, 2001), many of the natural experiments used to estimate returns to education identify average returns for more disadvantaged groups. The fact that increase in the education of such persons tends to generate higher-than-average returns suggests that increasing the education of such marginalized groups can have both desirable efficiency effects (high returns) as well as distributional effects (high returns to otherwise more-disadvantaged groups). This suggests the viability of increasing the education of such groups through policy initiatives such as increases in the compulsory school age, funding assistance, expansion of accessibility (e.g., by facilitating transfers from colleges to universities), and campaigns against dropping out.

A number of researchers have tried to model an individual's schooling decision by assuming more structure to the decision-making process (e.g., schooling only affects wealth, people have rational expectations, and discount the future geometrically). The models are simplified enough to

allow estimation of a few unknown parameters, such as an individual's time discount rate and return to schooling. A correctly specified model with enough structure permits estimation of a wide set of returns-to-schooling measures for individuals under different circumstances (e.g., faced with different costs or benefits or different schooling-level decisions). The advantage of this approach is that it emphasizes the economic content of what is being estimated and offers a potential approach to measuring individual rates of return in situations where an experimental approach cannot. Studies using this approach typically estimate returns around 0.04–0.07, which are lower than those from experimental approaches (Bezil, 2007, provides a review). If the assumptions of these models are incorrect, however, the returns-to-schooling estimates can be off significantly. Unfortunately, this is difficult to determine.

Annual Returns, Signaling, and Sheepskin Effects

An important issue to address is the extent to which the estimates of returns to schooling reflect not only the productivity-enhancing effect of schooling but also an effect on earnings of the underlying set of skills that schooling signals. There is a fundamental difficulty in unraveling the extent to which schooling is a signal of existing productivity as opposed to enhancing productivity: both theories are observationally equivalent – they both suggest that there is a positive correlation between earnings and schooling, but for very different reasons. If the rate at which employers learn an employee's correct set of skills is slow, or if early job placements influence long-term job opportunities, the effects of signaling can be long lasting.

The empirical literature generally finds evidence of such signaling or sheepskin effects (see Weiss, 1995; Chatterji *et al.*, 2003; for examples and reviews).

Ferrer and Riddell (2002) provide estimates of the credential effects and review much of the earlier literature, as do Heckman *et al.* (2006). The returns to an additional year of education that involves completion of a stage (e.g., graduating from high school, or university) is higher than the return to a year of education that does not involve the credential of the completion of a phase. For example, based on the 1996 Canadian census, Ferrer and Riddell (2002) estimate rates of return to an additional year of schooling to be 6% for males and 9% for females. These are averages of both the credential effects associated with milestones of completing various phases as well as the returns to years within the different phases. When the returns to completing the phases are calculated and annualized over the period of education necessary to complete the degree, the annual rates of return that are implied by completing university relative to high school are 9% for males and 11% for females.

While most researchers would agree that schooling affects earnings both by improving skills and by signaling skills, the relative importance of these effects is not well understood. One paper estimates that the contribution of signaling to the returns to schooling is less than 25% (Lange, 2007). Understanding the relative importance of signaling in explaining returns to schooling remains an important area for further research.

Trends and Some International Evidence

Returns to education in the United States have been increasing steadily in recent years likely reflecting the widening skill differential in wages (Card and Lemieux, 2001), in spite of the dramatic increases in education that would normally be expected to depress returns. Returns have also been increasing in Canada and European countries even though the wage structure there is more compressed when compared to the United States. (e.g., Bingley *et al.*, 2005). Obviously, the demand changes favoring higher-educated personnel have more than offset the supply changes.

International comparisons of the return to education are obviously difficult because of differences in the data sets and methodologies. In spite of this, reviews of the international evidence in general find similar results as those in the United States and Canada. (The international generalizations given here are based mainly on Psacharopoulos and Patrinos (2004), earlier studies are cited therein. Similar generalizations for some of the issues are given in Trostel *et al.* (2002) with further international evidence provided in OECD (1998)).

Summary

The following points sum up the returns to education:

- Returns to education tend to be in the neighborhood of 10%, typically ranging from 6% to 15%. **Table 1** illustrates the main approaches used to estimate these returns and their general findings.
- Returns to education tend to be in the 6–10% range when based on ordinary least-squares (OLS) estimates from conventional schooling equations and the 10–15% range (and sometimes higher) when based on instrumental variables (IV) and other procedures used to identify exogenous variation in education. As such, the 10% estimate is at the upper end of the OLS range and lower end of the IV range.
- The returns tend to be higher for
 1. females as opposed to males;
 2. obtaining the credentials associated with completing phases such as high school or university;

Table 1 Illustrative estimates of returns to education using alternative approaches

Author	General method	Sample	Returns to education
Griliches (1977)	Regress log median earnings of expected occupation at age 30 on schooling while using IQ score as additional proxy control for ability	17–27-year-old men in 1969 from US National Longitudinal Survey for Young Men, using log median earnings of expected occupation at age 30	0.059 (0.003)
Ashenfelter and Rouse (1998)	Regress difference in log earnings between identical twins on difference in schooling	1991–93 Princeton Twins Survey of identical twins	0.102 (0.010)
Card (1995)	Use indicator for whether living near a 4-year college as instrument for predicting schooling	US National Longitudinal Survey for Young Men	0.132 (0.049)
Chen (2008)	Use average tuition fees at the local college in county as instruments for predicting schooling	25–42-year-old men in 1979–2000 from US National Longitudinal Survey of Youth, 1979	0.133 (0.028)
Oreopoulos (2006b)	Use differences in state compulsory schooling laws as instruments for predicting schooling	25–64-year-old men and women in 1950–2000 US Censuses	0.142 (0.012)
Bezil and Hanson (2007)	Construct structural model on education-attainment decisions and estimate model along with returns to schooling	White males from the US National Longitudinal Survey of Youth	0.069 (average between grade 10 and 16)

Source: Bingley, P., Zhu, Y. and Walker, I. (2005). Education, work and wages in the UK. *German Economic Review* 6(3), 395–414.

3. general academic streams compared to technical vocational streams; and
 4. professional fields such as engineering, medicine, business, and sciences and lower in social sciences and humanities and especially fields such as fine arts.
- The returns tend to be increasing over time in spite of the large increases in the supply of educated persons, highlighting that the demand for education associated with the knowledge economy and the widening of the skilled–unskilled wage differential are outstripping the supply responses.

See also: Empirical Research Methods in the Economics of Education; Human Capital; Returns to Education in Developing Countries; Signaling in the Labor Market.

Bibliography

- Angrist, J. and Krueger, A. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* 106, 979–1014.
- Angrist, J. and Krueger, A. (1994). Why do World War II veterans earn more than non-veterans? *Journal of Labor Economics* 12, 74–97.
- Ashenfelter, O., Harmon, C., and Oosterbeek, H. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics* 6, 453–470.
- Ashenfelter, O. and Rouse, C. (1998). Income, schooling and ability: Evidence from a new sample of identical twins. *Quarterly Journal of Economics* 113, 253–284.
- Ashenfelter, O. and Zimmerman, D. (1997). Estimates of the return to schooling from sibling data. *Review of Economics and Statistics* 79, 1–9.
- Bezil, C. (2007). The return to schooling in structural dynamic models: A survey. *European Economic Review* 51(5), 1059–1105.
- Bingley, P., Christensen, K., and Walker, I. (2005). Twin-based estimates of the returns to education: Evidence from the population of Danish twins, unpublished Manuscript.
- Cameron, S. V. and Taber, C. (2004). Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy* 112, 132–182.
- Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling. In Christofides, L., Grant, K., and Swidinsky, R. (eds.) *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, pp 201–222. Toronto, ON: University of Toronto Press.
- Card, D. (1999). The causal effect of education on earnings. In Ashenfelter, O. and Card, D. (eds.) *Handbook of Labor Economics*, vol. 3, pp 1801–1863. Amsterdam: Elsevier Science.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent problems. *Econometrica* 69, 1137–1160.
- Card, D. and Lemieux, T. (2001). Can falling supply explain the rising return to college for younger men? A cohort-based analysis. *Quarterly Journal of Economics* 116, 705–746.
- Carnoy, M. (1997). Recent research on market returns to education. *Journal of Educational Research* 27, 483–490.
- Chamberlain, G. (1977). Omitted variable bias in panel data: Estimating the returns to schooling. *Annals de l'Insee* 30–31, 49–82.
- Chamberlain, G. and Griliches, Z. (1975). Unobservables with a variance components structure: Ability, schooling and the economic success of brothers. *International Economic Review* 16, 422–449.
- Chatterji, M., Seaman, P., and Singell, L. D., Jr. (2003). A test of the signalling hypothesis. *Oxford Economic Papers* 55, 191–215.
- Chen, S. (2008). Estimating the variance of wages in the presence of selection and unobserved heterogeneity. *Review of Economics and Statistics* 90(2), 275–289.
- Ferrer, A. and Riddell, W. C. (2002). The role of credentials in the Canadian labor market. *Canadian Journal of Economics* 35, 879–905.
- Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica* 45, 1–22.
- Griliches, Z. (1979). Sibling models and data in economics: Beginnings of a survey. *Journal of Political Economy* 87, S37–S65.
- Heckman, J., Lochner, L., and Todd, P. (2006). Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. In Hanushek, E. and Welch, F. (eds.) *Handbook of the Economics of Education*, vol. 1, pp 306–458. Amsterdam: Elsevier Science.
- Kane, T. and Rouse, C. (1995). Labor market returns to two-year and four-year college. *American Economic Review* 85, 600–614.
- Krashinsky, H. (2007). How would one extra year of high school affect labour market outcomes? Evidence from a unique policy change in Ontario, Canada, University of Toronto working paper.

- Lange, F. (2007). The speed of employer learning. *Journal of Labor Economics* **25**, 1–36.
- Lemieux, T. (2002). The causal effects of education on earnings in Canada. In deBroucker, P. and Sweetman, A. (eds.) *Towards Evidence-Based Policy for Canadian Education*, pp 105–116. Kingston, ON: McGill-Queens University Press.
- Lemieux, T. and Card, D. (2001). Education, earnings and the “Canadian G.I. bill.” *Canadian Journal of Economics* **34**, 313–344.
- Miller, P., Mulvey, C., and Martin, N. (2006). The return to schooling: Estimates from a sample of young Australian twins. *Labour Economics* **13**, 571–587.
- Mincer, J. (1974). *Schooling, Experience and Earnings*. New York: Columbia University Press.
- OECD (1998). *Human Capital Investment: An International Comparison*. Paris: Organisation for Economic Co-operation and Development.
- Oreopoulos, P. (2006a). The compelling effects of compulsory schooling: Evidence from Canada. *Canadian Journal of Economics* **39**, 22–52.
- Oreopoulos, P. (2006b). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review* **96**, 152–175.
- Psacharopoulos, G. and Patrinos, H. (2004). Returns to investment in education: A further update. *Education Economics* **12**, 111–134.
- Trostel, P., Walker, I., and Woolley, P. (2002). Estimates of the economic return to schooling in 28 countries. *Labour Economics* **9**, 1–16.
- Webbink, D. (2007). Returns to university education: Evidence from a Dutch institutional reform. *Economica* **74**, 13–134.
- Weiss, A. (1995). Human capital vs. signalling explanations of wages. *Journal of Economic Perspectives* **9**, 133–154.

Further Reading

- Carneiro, P., Heckman, J. J., and Vytlačil, E. (2005). Understanding what instrumental variables estimate: Estimating marginal and average returns to education. London: University College working paper.
- Harmon, C. and Walker, I. (1995). Estimates of the economic return to schooling for the United Kingdom. *American Economic Review* **85**, 1278–1286.

Returns to Education in Developing Countries

H A Patrinos, World Bank, Washington, DC, USA

G Psacharopoulos, London School of Economics, World Bank, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Earnings of workers classified by some dimension have been at the core of empirical economics, as well as of sociology, since the dawn of the discipline. Up to the middle of the twentieth century the main dimensions according to which earnings were classified were sector of economic activity, industry, occupation, and formal versus informal sector in developing countries. From about 1960 onward, following the human capital revolution in economic thought (Schultz, 1961; Becker, 1964; Mincer, 1974), the educational level of the worker was added as a classificatory variable.

Earnings differentials by level of education reflect the monetary incentives for someone to invest in education. Earnings differentials by education represent the intersection of supply and demand curves for educated labor. Differences in relative earnings between countries reflect a number of factors, for example, the demand for skills in the labor market, minimum wage legislation, the strength of unions, collective agreements, the supply of workers with various levels of educational attainment, the work experience of workers with high and low levels of schooling, the distribution of employment among occupations, and the relative incidence of part-time and seasonal work (OECD, 2007).

The study of earnings by schooling has led to a number of empirical works testing hypotheses on a great variety of social issues. These include, for example, racial and ethnic discrimination (McNabb and Psacharopoulos, 1981; Chiswick, 1988; Psacharopoulos and Patrinos, 1994); gender discrimination (Psacharopoulos and Tzannatos, 1992; Goldin and Polachek, 1987); income distribution (Mincer, 1958; Marin and Psacharopoulos, 1976); and the determinants of the demand for education (Freeman, 1976; Psacharopoulos and Soumelis, 1979; Psacharopoulos, 1982). Under certain assumptions, earnings differentials by level of education have been used to identify the sources of economic growth (e.g., Denison, 1967; Psacharopoulos, 1972). However, perhaps the application *par excellence* that has used earnings by level of education is the estimation of the rate of return to investment in schooling.

In what follows we give a taste of earnings differentials by education in developing countries, and compare them to those in advanced industrial countries. The section titled 'Returns to education' provides a review of the returns to education for developing countries. Empirical estimates of

the returns to education date from the late 1950s and thus range from one decade after the post-war years (after World War II) to the beginning of the twenty-first century. As the number of empirical studies increases, compilations of the rate of returns to education start to emerge in the early 1970s (Psacharopoulos, 1973) and have continued to the present (Psacharopoulos and Patrinos, 2004). The final section tries to draw policy conclusions from the empirical work based on earnings differentials by education.

Earnings Differentials by Education

It is a universal fact that, in all countries of the world, the more education one has the higher his/her earnings (Table 1). Age-earnings profiles by level of education behave as predicted by the seminal theoretical and empirical work of Mincer (1974), taking the general shape depicted in Figure 1.

To set a perspective, Table 2 presents earnings differentials in the Organization for Economic Co-operation and Development (OECD) countries. Note that the university premium amounts to a factor of 2 only in the case of Hungary. As shown in Figure 2, using readily available data from Latin America to serve as illustrative example, the earnings premium by level of education is steeper in developing countries. Out of the many factors responsible for the intersection of the supply and demand for educated labor taking place at different points in the two sets of countries, the main one must be that human capital is scarcer in developing, less-developed countries (LDCs), relative to industrial or developed countries (DCs).

Returns to Education

Using an appropriate methodology, the above earnings differentials can be used to estimate the returns to investment in education. Most published estimates of the rate of return to schooling rely on the Mincerian earnings function (Mincer, 1974), the most widely used estimation in economics (for a review of the Mincerian earnings function see Heckman *et al.*, 2006). However, the most correct methodology is what is known as the full discounting method that is based on the actual shape of the age-earnings profiles, rather than be smoothed out by an earnings function (Psacharopoulos and Mattson, 1998).

Based on this methodology, Tables 3 and 4 present the private and social rates of return stemming from the

earnings differentials presented above, for selected Latin American and OECD countries. A private rate of return takes into account only the costs and benefits realized by the individual, that is, earnings are after taxes and, because of the public subsidization of education, the main part of the cost is foregone earnings while studying. A social rate of return is based on pretax earnings and takes into account the full resource cost of education.

Table 1 Mean earnings of the labor force by level of education – Latin America and Caribbean countries (index)

	No education	Primary	Secondary	University
Argentina	35	62	100	171
Bolivia	74	77	100	219
Brazil	60	81	100	201
Chile	44	59	100	311
Colombia	42	64	100	253
Costa Rica	50	63	100	202
Dominican Rep.	56	76	100	251
El Salvador	42	63	100	208
Guatemala	30	55	100	247
Honduras	38	52	100	265
Jamaica	5	65	100	113
Mexico ^a	19	54	100	173
Panama	50	54	100	224
Paraguay	51	63	100	212
Peru	73	82	100	253
Uruguay	40	77	100	154
Venezuela	47	75	100	172
Average	44	66	100	213

^aAlso an OECD country

Based on table T-9 in Psacharopoulos, G. and Chu Ng, Y. (1992). Earnings and education in Latin America: Assessing priorities for schooling investments. *Policy Research Working Papers WPS 1056*, World Bank.

The returns to primary education are highest in the developing countries set, a finding giving a taste of diminishing returns. It should also be noted that in both sets of countries the private returns are higher than the social returns because of the public subsidization of education. In fact, the difference between the private and social rates of return offers a handy measure of the degree of such subsidization. As shown in **Table 5**, the highest level of education is subsidized more heavily relative to the lower ones, raising a major issue of perverse equity.

Table 6 reports results from a wider set of developing countries using the Mincerian earnings function. The average rate of return in developing economies is almost 11%, slightly higher than the 10% for the whole world. The returns are highest at the primary level, but the returns to higher education are also high. A typical pattern in developing countries is the dip in returns at the secondary level. This has been attributed not only to low quality of schooling (but without any empirical justification), but also to the option value of secondary schooling (Weisbrod, 1964). It could also be that the most successful secondary school students go on to higher education.

Among a few economies, there appears to be an increase in returns to education over time. With regard to educational attainment, there appears to be increasing returns to education in developed economies. While this conforms to the observations made in prior research, Schultz (2004) also mentioned that there are now increasing returns to education (by levels) in African economies. Developing transition countries, such as China, start from a low base (in terms of the rate of returns to education), and have experienced substantial increases in rates of return to education over the years.

In **Figure 3**, we break down the rate of returns to education by gender. We find that for developing

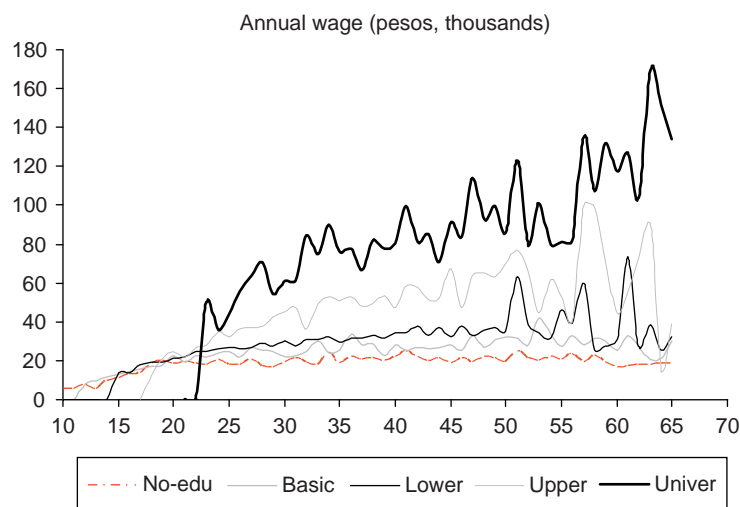


Figure 1 Based on age-earnings profiles by education, Mexico, 2004. From Patrinos, H. A. and Metzger, S. (2007). Returns to education in Mexico: An update. World Bank (processed).

Table 2 Relative earnings of the population with income from employment by level of education – OECD countries (index)

Country	Below upper secondary	Upper secondary (index base)	Tertiary
Australia	81	100	131
Austria	71	100	152
Belgium	90	100	134
Canada	78	100	138
Czech Republic	72	100	181
Denmark	82	100	126
Finland	94	100	149
France	86	100	144
Germany	88	100	156
Hungary	73	100	215
Ireland	86	100	164
Italy	79	100	160
Korea	67	100	141
Luxembourg	78	100	145
Netherlands	84	100	148
New Zealand	78	100	132
Norway	84	100	136
Poland	78	100	163
Portugal	57	100	179
Spain	85	100	132
Sweden	87	100	127
Switzerland	76	100	156
Turkey	65	100	141
United Kingdom	69	100	155
United States	67	100	175
Average	78	100	151

Based on table A9.1a in OECD (2007). *Education at a Glance 2007*. Paris: OECD.

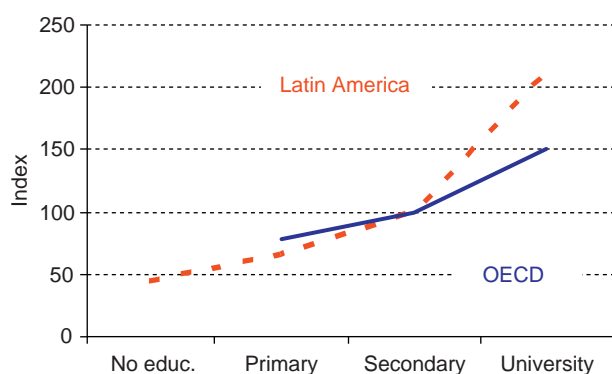


Figure 2 Based on education-earnings differentials are steeper in developing countries.

economies, the rate of return to education for men is highest for primary education (14.3%), decreases for secondary education (9.2%), and increases for university education (12.3%). For women, the rate of returns to education is lowest for primary education (7.2%), highest for secondary education (13.1%), and falls to 10.1% for university education. That is, in developing economies, on average, rate of return to primary education for women is lower than that of men but women overtake men by secondary education. For university education, on average, men's rate of returns to university education is 2.2 percentage points higher than that of women.

It is possible that there may be differential returns in education by gender due to the different job opportunities available to men and women at each educational level. For example, Hawley (2004) finds that in Thailand, women have higher returns to secondary education than men. He reasoned that this is because occupations, such as those in manufacturing, that pay more and require secondary education, have grown more rapidly for women than for men (Phananiramai, 1996; Kurian, 1999). Conversely, in jobs that may require only primary education (such as construction), the proportion of men may be higher because of tradition or labor intensity. Consequently, men may have higher returns to primary education.

A major policy concern, therefore, is the relatively lower returns to education for women at the primary education level in developing economies. It is a major concern because families in poor countries may decide to send sons rather than daughters to school. The lower returns to females at the primary level in developing countries is puzzling, given the fact that the rate of return to years of schooling is two percentage points greater for females than for males in developing countries, as well as transition countries. Dougherty (2005) considers various explanations. The most important involves the detrimental impact of discrimination and other factors that cause women to accept wage offers that undervalue

Table 3 Returns to education by level of education (%)

Country	Private			Social		
	Primary	Secondary	University	Primary	Secondary	University
Argentina	10.1	14.2	14.9	8.4	7.1	7.6
Bolivia	9.8	8.1	16.4	9.3	7.3	13.1
Brazil	36.6	5.1	28.2	35.6	5.1	21.4
Chile	9.7	12.9	20.7	8.1	11.1	14.0
Colombia	27.7	14.7	21.7	20.0	11.4	14.0
Costa Rica	12.2	17.6	12.9	11.2	14.4	9.0
Ecuador	17.1	17.2	12.7	14.7	12.7	9.9
El Salvador	18.9	14.5	9.5	16.4	13.3	8.0
Honduras	20.8	23.3	25.9	18.2	19.7	18.9
Mexico	21.6	15.1	21.7	19.0	9.6	12.9
Paraguay	23.7	14.6	13.7	20.3	12.7	10.8
Uruguay	27.8	10.3	12.8	21.6	8.1	10.3
Venezuela	36.3	14.6	11.0	23.4	10.2	6.2

Based on Psacharopoulos, G. and Chu Ng, Y. (1992). Earnings and education in Latin America: Assessing priorities for schooling investments. *Policy Research Working Papers WPS 1056*, World Bank.

Table 4 Returns to education – OECD countries (%)

Country	Private		Social	
	Upper secondary	University	Upper secondary	University
Belgium	14.8	10.7	11.4	12.2
Denmark	16.2	8.3	11.1	7.8
Finland	17.3	16.7	8.2	13.6
Hungary	12.0	22.6	8.3	18.8
Korea	14.0	12.2	6.7	14.2
New Zealand	14.1	9.3	8.3	9.9
Norway	9.0	12.1	5.5	9.5
Sweden	18.7	8.9	10.4	7.5
Switzerland	7.0	10.0	1.7	6.3
United Kingdom	21.3	16.8	13.4	13.7
United States	23.9	14.3	12.5	14.1

Based on tables A9.5 to A9.8, column 1, in OECD (2007). *Education at a Glance 2007*. Paris: OECD.

Table 5 Public subsidization of education (Index)

Country	Primary	Secondary	Higher
Argentina	17	50	49
Bolivia	6	10	20
Brazil	3	1	30
Chile	17	14	33
Colombia	28	22	35
Costa Rica	8	18	30
Ecuador	14	26	22
El Salvador	13	8	16
Honduras	13	17	27
Mexico	12	37	41
Paraguay	14	13	21
Uruguay	22	22	20
Venezuela	36	30	44

Based on table T-13 in Psacharopoulos, G. and Chu Ng, Y. (1992). Earnings and education in Latin America: Assessing priorities for schooling investments. *Policy Research Working Papers WPS 1056*, World Bank.

Table 6 Private returns to education and mean years of education, developing economies

	<i>Private returns to education (%)</i>			
	<i>An additional year of schooling</i>	<i>Primary Education</i>	<i>Secondary Education</i>	<i>Higher Education</i>
Mean	10.8	23.0	17.9	21.1
Number of countries	52	49	55	51

Based on Psacharopoulos, G. and Patrinos, H. A. (2004). Returns to investment in education: A further update. *Education Economics* 12 (2), 111–134.

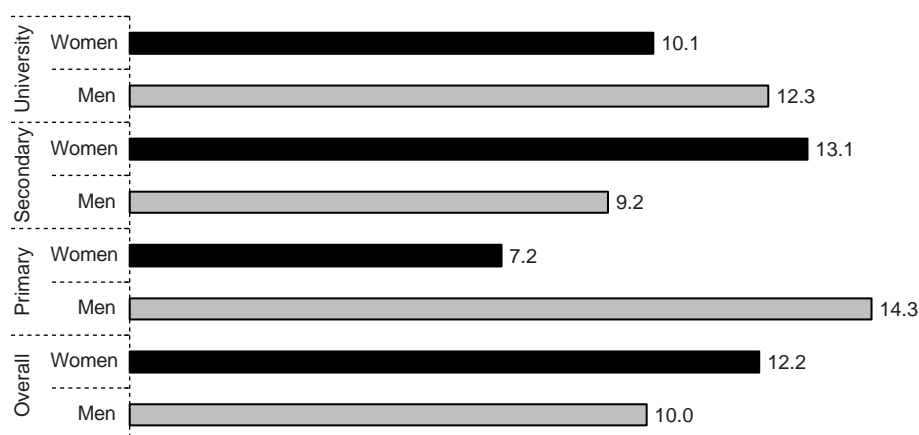


Figure 3 Based on private returns to education by gender and level. From Patrinos, H. A. and Psacharopoulos, G. (2007). Returns to education: An international update. World Bank, Working Paper Series (processed).

their characteristics. It is hypothesized that the better educated is a woman, the more able and willing she is to overcome the sex handicaps and compete with men in the labor market. He also considers the possibility that part of the differential could be attributable to male–female differences in the quality of educational attainment. Moreover, women may choose to work in sectors where education is relatively highly valued. The lower return to primary education in developing economies may be evidence of discrimination – such that women need to achieve more schooling in order to garner sufficient wages – or the option value of schooling – that is, since women are likely to perform better, stay in school longer, and experience higher returns for each year of schooling, the most able go beyond primary schooling, thus depressing the returns at that level.

Turning to the broader picture, the rate of return to education has fallen over the decades, while average years of schooling have risen. The returns to different educational attainments have fluctuated over the years, but it is also clear that amid the fluctuations, there has been a downward trend since the 1980s. The proportions of the population that are secondary educated and university educated have all risen markedly over the decades, while the proportion of the population that is primary educated has declined slightly. The rate of return to primary education has fluctuated over the decades. For secondary education, the rate of return has risen until the 1980s.

For university education, the private rate of return has declined as the proportion that is university educated increases (Patrinos and Psacharopoulos, 2007).

Causality

It is also worth noting that estimates of the returns on education based on advanced econometric techniques that control for different characteristics come to an average rate of return that is very similar to the global average presented in reviews: 10%. To estimate the true effect of education on earnings, some authors have employed instrumental variables (IVs) estimates. A useful instrument should be correlated with schooling, but uncorrelated with unexplained variation in earnings. Various variables on family background are frequently used. However, they are not expected to meet the requirement that they are uncorrelated with earnings (among other possible reasons, due to intergenerational effects).

Recently, information on the costs of schooling and supply-side sources of identifying information, such as various types of education reform, are increasingly sought after as instruments. Costs are important because people make decisions about investing in schooling on perceived costs and benefits. Therefore, loan policies, tuition changes, and distance are possible sources of instruments. Institutional constraints are also important. Therefore, supply-side

changes, such as the extension of compulsory school laws or making education facilities more accessible by reducing the distance to school, provide the researcher with the sort of natural experiment that allows one to say that the instrument is correlated with schooling but not earnings. One could argue, for example, that extending compulsory school age results in more people enrolling in school because of the legal change rather than their individual ability to generate more earnings. Such instruments will fail if they are corrected with earnings – as is sometimes the case with family background variables; in other words, the approach will fail when the researcher chooses bad instruments.

The IV estimates are often higher than ordinary least-squares (OLS) estimates, although it is unclear to what extent this is due to measurement error or inadequate instrumentation (see Trostel *et al.*, 2002). There are very few studies for developing countries dealing with the issues surrounding the endogeneity of education and the implications of estimating returns to education from IV. The rate of return estimates from IV is not only different from OLS, but it is also generally higher. Since the IV estimate is supposed to be the true return, this contradicts the standard ability bias intuition: OLS should be upward biased if higher ability individuals have more schooling. This approach uses sources for exogenous variation in educational attainment, such as institutional changes in the schooling system in the form of changes in compulsory schooling laws (Patrinos and Sakellariou, 2006), and reducing the distance to school (such as school construction projects; see Duflo, 2001) affecting the schooling decision, to estimate a causal return to education. The interpretation (Card, 2001) is that the returns to schooling vary across individuals. Institutional changes affect the schooling decision of a subset of individuals who, otherwise, would not have pursued a higher level of education and not the average individual. Furthermore, individuals affected by such reforms tend to have a higher return to education than the average individual. There is a distribution of returns, and OLS and IV correspond to different weighted averages of this distribution, and OLS can be below IV (see also Heckman and Li, 2004; Arabsheibani and Mussurav, 2007; Sakellariou, 2006).

Cognitive Ability

Beyond the returns to increasing the quantity of schooling, recent research has thrown light on the returns to improving the quality of schooling, the latter being measured by the cognitive skill of the student or graduate. Ability, or school quality, matters for earnings attainment (Leuven *et al.*, 2004). **Table 7** summarizes research results to increasing cognitive skills.

Conclusions and Policy Considerations

The empirical returns to schooling literature have proven to be a useful standard. The global average rate of return to schooling, estimated at 10%, is used as a global benchmark. Empirical evidence on returns on investment in education is a useful indicator of the productivity of education, and it serves as an incentive for individuals to invest in their own human capital.

More research is needed, however, on estimating the social benefits of schooling. After all these years, it is still an underdeveloped theme in the literature and remains a research priority. Still, existing evidence suggests that social returns and externalities are likely high.

Several important papers over the last decade on establishing causation between education and earnings have made the case for schooling as an investment. It is also clear that there is a need for more evidence on the impact of education on earnings using quasi-, and where possible, experimental design. Evaluation techniques are usefully applied in analyzing education and earnings, and providing evidence on the effectiveness of projects, programs, and reforms. Thus, more research along the lines of Oreopoulos (2006) and others' analysis of changes in compulsory schooling laws and other reforms is warranted. Such analyses allow one to provide not only estimates for those likely affected by reforms (i.e., the local average treatment effect, or LATE), but also estimates that come close to population average treatment effects (or ATEs).

Recent research shows that skills – ability and learning outcomes – matter for earnings attainment. Overall, in several countries, a one standard deviation increase in school

Table 7 Estimated returns to a standard deviation increase in cognitive skills

Country	Estimated effect	Source
Chile	0.17	Patrinos and Sakellariou (2007)
Ghana	0.14–0.30	Glewwe (1996)
Ghana	0.05–0.07	Jolliffe (1998)
Kenya	0.19–0.22	Boissiere <i>et al.</i> (1985), Knight and Sabot (1990)
Pakistan	0.12–0.28	Alderman <i>et al.</i> (1996)
Pakistan	0.25	Behrman <i>et al.</i> (forthcoming)
South Africa	0.34–0.48	Moll (1998)
Tanzania	0.07–0.13	Boissiere <i>et al.</i> (1985), Knight and Sabot (1990)
Average	0.17–0.22	

performance results in about a 12–15 increase in earnings. Although more work is needed in this area, clearly, learning outcomes are important determinants of earnings. Therefore, a policy priority is to empirically establish which policies and programs lead to learning outcome gains.

Evidence of heterogeneity in the returns to education exists. From the evidence available, in most high-income and middle-income countries, higher returns at the upper ends of the wage distribution have been observed. Overall, in low-income countries the returns tend to be higher at the lower ends of the wage distribution. Although more research is needed, such differences between countries could be due to: more job mobility in DCs allowing individuals to improve their position by changing jobs; scarcity of skills; differential exposure to market forces and the link between pay and productivity; or differential access to quality education or distribution of quality outcomes. What is clear is that average returns to the average individual will not suffice for policy purposes. We need to know more about which interventions are more likely to affect which parts of the distribution, and to establish at the same causality. In the above review, we noted that a major policy concern is the relatively low private rate of return to primary education for women in developing economies.

See also: Empirical Research Methods in the Economics of Education; Human Capital; Returns to Education in Developed Countries; Signaling in the Labor Market.

Bibliography

- Arabsheibani, G. R. and Mussurav, A. (2007). Returns to schooling in Kazakhstan: OLS and instrumental variables approach. *Economics of Transition* **15**(2), 341–364.
- Becker, G. S. (1964). *Human Capital*. Chicago, IL: Columbia University Press.
- Card, D. (2001). Estimating the return on schooling: Progress on some persistent econometric problems. *Econometrica* **69**(5), 1127–1160.
- Chiswick, B. R. (1988). Differences in education and earnings across racial and ethnic groups: Tastes, discrimination, and investments in child quality. *Quarterly Journal of Economics* **103**(3), 571–597.
- Denison, E. F. (1967). *Why Growth Rates Differ?* Washington, DC: The Brookings Institution.
- Dougherty, C. (2005). Why are returns to schooling higher for women than for men? *Journal of Human Resources* **40**(4), 969–988.
- Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review* **91**(4), 795–813.
- Freeman, R. B. (1976). *The Overeducated American*. New York: Basic Books.
- Goldin, C. and Polachek, S. (1987). Residual differences by sex: Perspectives on the gender gap in earnings. *American Economic Review* **77**(2), 143–151.
- Hawley, J. D. (2004). Changing returns to education in times of prosperity and crisis, Thailand 1985–1998. *Economics of Education Review* **23**, 273–286.
- Heckman, J. J. and Li, X. (2004). Selection bias, comparative advantage and heterogeneous returns to education: Evidence from China in 2000. *Pacific Economic Review* **9**(3), 155–171.
- Heckman, J. J., Lochner, L. J., and Todd, P. E. (2006). Earnings functions, rates of return, and treatment effects: The Mincer equation and beyond. In Hanushek, E. and Welch, F. (eds.) *Handbook of the Economics of Education*, pp 307–458. Amsterdam: Elsevier/North-Holland.
- Kurian, R. (1999). Women's work in changing labour markets: The case of Thailand in the 1980s. In Afshar, H. and Barrientos, S. (eds.) *Women, Globalisation and Fragmentation in the Developing World*, pp 174–195. New York: St. Martin's Press.
- Leuven, E., Oosterbeek, H., and van Ophem, H. (2004). Explaining international differences in male skill wage differentials by differences in demand and supply of skills. *Economic Journal* **114**, 466–486.
- Marin, A. and Psacharopoulos, G. (1976). Schooling and income distribution. *Review of Economics and Statistics* **58**(3), 332–338.
- McNabb, R. and Psacharopoulos, G. (1981). Racial earnings differentials in the U.K. *Oxford Economic Papers* **33**(3), 413–425.
- Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy* **66**(4), 281–302.
- Mincer, J. (1974). *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research.
- OECD (2007). *Education at a Glance 2007*. Paris: OECD.
- Oreopoulos, P. (2006). Estimating average and local average treatment effects when compulsory schooling laws really matter. *American Economic Review* **96**(1), 152–175.
- Patrinos, H. A. and Metzger, S. (2007). Returns to education in Mexico: An update. World Bank (processed).
- Patrinos, H. A. and Psacharopoulos, G. (2007). Returns to education: An international update. World Bank, Working Paper Series (processed).
- Patrinos, H. A. and Sakellariou, C. (2006). Economic volatility and returns to education in Venezuela: 1992–2002. *Applied Economics* **38**, 1991–2005.
- Phananiramai, M. (1996). Changes in women's economic role in Thailand. In Horton, S. (ed.) *Women and Industrialization in Asia*, pp 274–306. London: Routledge.
- Psacharopoulos, G. (1972). The marginal contribution of education to economic growth. *Economic Development and Cultural Change* **20**(4), 641–658.
- Psacharopoulos, G. (1973). *Returns to Education: An International Comparison*. Amsterdam/San Francisco, CA: Elsevier/Jossey-Bass.
- Psacharopoulos, G. (1982). An analysis of the determinants of the social demand for upper secondary education in Portugal. *Economics of Education Review* **2**(3), 233–251.
- Psacharopoulos, G. and Mattson, R. (1998). Estimating the returns to education: A sensitivity analysis of concepts and methods. *Journal of Educational Planning and Administration* **12**(3), 271–287.
- Psacharopoulos, G. and Patrinos, H. A. (eds.) (1994). *Indigenous People and Poverty in Latin America: An Empirical Analysis*. Washington, DC: World Bank.
- Psacharopoulos, G. and Patrinos, H. A. (2004). Returns to investment in education: A further update. *Education Economics* **12**(2), 111–134.
- Psacharopoulos, G. and Soumelis, C. (1979). A quantitative analysis of the demand for higher education. *Higher Education* **8**(2), 159–177.
- Psacharopoulos, G. and Tzannatos, Z. (1992). *Women's Employment and Pay in Latin America: Overview and Methodology*. Washington, DC: World Bank.
- Sakellariou, C. (2006). Education policy reform, local average treatment effect and returns to schooling from instrumental variables in the Philippines. *Applied Economics* **38**, 473–481.
- Schultz, T. P. (2004). Evidence of returns to schooling in Africa from household surveys: Monitoring and restructuring the market for education. *Journal of African Economies* **13**(2), 95–148.
- Schultz, T. W. (1961). Investment in human capital. *American Economic Review* **51**, 1–17.
- Trostel, P., Walker, I., and Woolley, P. (2002). Estimates of the economic return to schooling for 28 countries. *Labour Economics* **9**, 1–16.
- Weisbrod, B. (1964). *External Benefits of Public Education and Economic Analysis*. Princeton, NJ: Industrial Relations Section.

Further Reading

- Alderman, H., Behrman, J. R., Ross, D. R., and Sabot, R. (1996). The returns to endogenous human capital in Pakistan's rural wage labor market. *Oxford Bulletin of Economics and Statistics* **58**, 29–55.
- Arrow, K. (1973). Higher education as a filter. *Journal of Public Economics* **2**(3), 193–216.
- Ashenfelter, O. and Krueger, A. B. (1994). Estimates of the economic return to schooling from a new sample of twins. *American Economic Review* **84**(5), 223–265.
- Azariadis, C. and Drazen, A. (1990). Threshold externalities in economic development. *Quarterly Journal of Economics* **105**(2), 501–526.
- Bedi, A. S. and Edwards, J. H. Y. (2002). The impact of school quality on earnings and educational returns: Evidence from a low-income country. *Journal of Development Economics* **68**(1), 157–185.
- Bedi, A. S. and Gaston, N. (1999). Using variation in schooling availability to estimate educational returns for Honduras. *Economics of Education Review* **18**(1), 107–116.
- Behrman, J. R., Ross, D., and Sabot, R. (2008). Improving the quality versus increasing the quantity of schooling: Estimates of rates of return from rural Pakistan. *Journal of Development Economics* **85**, 94–104.
- Boissiere, M. X., Knight, J. B., and Sabot, R. H. (1985). Earnings, schooling, ability, and cognitive skills. *American Economic Review* **75**(5), 1016–1030.
- Card, D. (1999). The causal effect of education on earnings. In Ashenfelter, O. and Card, D. (eds.) *Handbook of Labor Economics*, vol. 3A, pp 1801–1863. Amsterdam: Elsevier.
- Dougherty, C. R. S. and Jimenez, E. (1991). The specification of earnings functions: Tests and implications. *Economics of Education Review* **10**(2), 85–98.
- Glewwe, P. (1996). The relevance of standard estimates of rates of return to schooling for education policy: A critical assessment. *Journal of Development Economics* **51**, 267–290.
- Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica* **45**(1), 1–22.
- Jolliffe, D. (1998). Skills, schooling, and household income in Ghana. *World Bank Economic Review* **12**, 81–104.
- Knight, J. B. and Sabot, R. H. (1990). *Education, Productivity, and Inequality: The East African Natural Experiment*. Oxford: Oxford University Press.
- Layard, R. and Psacharopoulos, G. (1974). The screening hypothesis and the returns to education. *Journal of Political Economy* **82**(5), 985–998.
- Moll, P. G. (1998). Primary schooling, cognitive skills, and wage in South Africa. *Economica* **65**, 263–284.
- Psacharopoulos, G. (1987a). Earnings functions. In Psacharopoulos, G. (ed.) *Economics of Education: Research and Studies*, pp 218–223. Oxford: Pergamon.
- Psacharopoulos, G. (1987b). The cost-benefit model. In Psacharopoulos, G. (ed.) *Economics of Education: Research and Studies*, pp 342–347. Oxford: Pergamon.
- Psacharopoulos, G. and Chu Ng, Y. (1992). Earnings and education in Latin America: Assessing priorities for schooling investments. *Policy Research Working Papers WPS 1056*, World Bank.
- Psacharopoulos, G. and Layard, R. (1979). Human capital and earnings: British evidence and a critique. *Review of Economic Studies* **46**(3), 485–503.
- Patrinos, H. A. and Sakellariou, C. (2007). Quality of schooling, returns to schooling and the 1981 vouchers reform in Chile. *Paper Provided by The World Bank in Its Series Policy Research Working Paper Series with Number 4617*.
- Schultz, T. P. (1988). Education investment and returns. In Chenery, H. and Srinivasan, T. N. (eds.) *Handbook of Development Economics*, vol. 1, pp 543–630. Amsterdam: North-Holland.
- Schultz, T. P. (2003). Evidence of returns to schooling in Africa from household surveys: Monitoring and restructuring the market for education. *Discussion Paper 875*. Yale University Economic Growth Center.
- Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics* **87**(3), 355–374.
- Weiss, A. (1995). Human capital vs. signaling explanations of wages. *Journal of Economic Perspectives* **9**(4), 133–154.
- Willis, R. J. (1986). Wage determinants: A survey and reinterpretation of human capital earnings functions. In Ashenfelter, O. and Layard, R. (eds.) *Handbook of Labor Economics*, pp 525–602. Amsterdam: North-Holland.

School Quality and Earnings

J R Betts, University of California, San Diego, La Jolla, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Endogeneity – A situation in which one variable, Y , is being modeled as a function of a second variable X , but variable X is determined by other variables that also influence Y . In such a situation, X is said to be endogenous, and standard statistical methods of fitting the best line between Y and X will tend to give the wrong answer because both X and Y are being determined by other variables.

What factors contribute to high-quality schooling? A large literature studies the relation between class size, teacher qualifications, spending per pupil, and other measures of school inputs with gains in student achievement. This literature is quite useful, in part because it uses an outcome close in time to when the student enjoys the benefits of additional resources, and because one could reasonably expect better school resources to boost test scores.

A shortcoming of this approach is that test scores are only weakly linked to adult outcomes that policymakers really care about, such as employment and earnings. For this reason, social scientists have attempted to estimate a link between school resources and earnings of students once they leave school and enter the labor market.

This article briefly summarizes the methods used in this literature, the findings, alternative explanations for the variation in findings across studies, and weaknesses of the methods used thus far in the literature.

Several patterns emerge:

1. Most of the studies that find no link or a weak link between school inputs and student outcomes measure school inputs at the level of the actual school attended; studies that do find a strong effect typically measure school resources at the level of the US state.
2. Studies that find that school inputs have a strong link tend to examine workers schooled before the 1960s. The opposite holds for studies that find no link.
3. Studies that find no significant link often examine workers who are in their early 30s or their 20s at the time earnings are observed.

Estimation Methods Typically Used

The typical study estimates a variation of the following log earnings equation for person i whose wage is observed

in year t , who attended school s in region b and who currently resides in region r :

$$\ln W_{isbtr} = \alpha + \beta \text{Qual}_{is} + X_{ist}\Gamma + V_{ib}\Phi + Z_{irt}\psi + u_{isbtr} \quad [1]$$

where the natural log of earnings (annual or hourly in most work) is assumed to depend on one or more measures of school characteristics Qual_{is} ; personal characteristics X_{ist} such as age, gender, race/ethnicity; characteristics of the area where the student grew up V_{ib} ; characteristics of the area where the person currently resides Z_{irt} ; u_{isbtr} is an error term that captures unobserved factors, and the Greek letters indicate coefficients or vectors of coefficients to be estimated.

Several important variants of this equation have been estimated. Some authors have included years of schooling ED_{it} on the right-hand side of this equation, and interacted it with the measure(s) of school resources Qual_{ist} , in the belief that school resources should have a bigger effect on earnings the longer one stays in school (see for instance Card and Krueger, 1992a, b). However, to the extent that years of schooling and earnings are jointly determined, such models will be biased because years of schooling ED_{it} is an endogenous variable, that is, a variable chosen by the person and influenced by other factors in eqn [1] as well as by unobserved factors. Such models will tend to be biased because of correlation between ED_{it} and u_{isbtr} .

Researchers have measured observed school resources Qual_{is} at different levels of aggregation. Equation [1] assumes that the researcher has data on the actual school attended, but many researchers have proxied for Qual_{is} with measures of spending per pupil, pupil–teacher ratios and other measures of school resources measured at the district level, or even the level of the US state in which the worker was born.

A third important source of variation derives from alternative approaches to modeling characteristics of the place b where a person attended school and the place r where the person resides at the time of the wage observation. Again, authors alternatively measure characteristics of local areas at a disaggregated level such as counties or cities, or more at a more aggregated level, typically US states. Further, other researchers who observe workers in a wide range of years add fixed effects for states in which students attended school and currently reside. These fixed effects control for anything observable and unchanging over time from the regions b and r .

Methodological Weaknesses

The literature as a whole shares numerous weaknesses which make it difficult to treat estimates of β from eqn [1] as estimates of the causal effect of the given measure of school resources on students' earnings years later. The first challenge to a causal interpretation is omitted variable bias: the error term u_{isbrr} likely contains many omitted personal, neighborhood, and regional characteristics that directly influence wages. If these are correlated with the explanatory variables such as school resources, then we will obtain biased estimates. Closely related to this concern is endogeneity bias. Economists have shown that people self-select into neighborhoods that offer the mix of taxation and public services that they most prefer. School quality, or perceptions of school quality, clearly is related to families' choice of where to live. This creates a strong positive link between average student achievement in a school and various measures of socioeconomic status of people living in a given neighborhood or city.

Two potential but partial solutions to this problem are to control for family and neighborhood demographics as much as possible, or to aggregate up beyond the neighborhood of residence to the school district or even the state. Neither of these methods is likely to yield causal estimates. The first method can at best only partially control for unobserved determinants of families' choices of school. The second method, involving aggregation, risks creating aggregation bias if the causal effects of school resources are nonlinear. Another potential bias in models that observe school resources at the state level is the possibly nonrandom-measurement error due to systematic differences between the location in which one was born (the most commonly used proxy for the state in which a student attended school in the many studies that use census data), the state in which one attended school, and the state in which one lived at the time of the wage observation.

A problem that potentially afflicts the school-level studies is random-measurement error in the measures of school resources, which will bias findings toward conclusions of a zero link to earnings. Small sample size is a second potential issue.

Summary of Findings

The literature review below, unless stated otherwise, discusses the relatively large US literature.

Testing for a Significant Relation between School Resources and Earnings

When a researcher asks "Does school quality matter for earnings of students?" this question can be interpreted in several ways. One response is to ask whether students who

attend certain high schools tend to earn more than students who attend other schools. Both Betts (1995) and Grogger (1996a) find that earnings of adults vary significantly by US high school attended, even after controlling for a host of personal characteristics.

However, the literature that instead tests for a relation between school resources and earnings of students typically finds either a zero or only a small link between earnings and measures of school resources.

The measure of school resources most typically used is school spending per pupil in the district attended, or in the worker's state of birth. Other measures of school resources include the pupil-teacher ratio, or its reciprocal, various measures of teacher education, teacher salary, and in a few studies, teacher salary, books per student, or length of the school year.

Betts (1996a) reviews the US evidence in detail. As the patterns in the literature are fairly similar across measures of school resources, below we focus mainly on the evidence regarding spending per pupil and the teacher-pupil ratio (or its reciprocal).

Most of the state-level estimates (83%) found a positive relation between spending per pupil and earnings, compared to 51% of the estimates based on district-level spending per pupil. The single existing estimate based on school-level spending per pupil found no significant link. This pattern, of more positive findings when one proxies school resources based on the worker's state of birth rather than based on the actual school or district attended, extends to most of the other measures of school resources. For example, Betts (1996a) reports that the percentage of estimates that found a positive association between earnings and the teacher-pupil ratio when the teacher-pupil ratio was measured at the state, district, and school level were 19, 0, and 0, respectively.

Figure 1 shows, for the US studies, the patterns of statistical significance across studies. Black bars indicate studies in which lesser than a third of results suggest a significant positive relation between school resources and earnings, while gray and white bars indicate cases where one-third to two-thirds and more than two-thirds, respectively, of estimates suggest a positive relation. The horizontal axis indicates the years in which the adults in the study were likely to have enrolled in grades 1 through 12. Vertically, the studies are arranged to display differences between studies that measure school resources at the school level, the district level, or by state of birth.

Figure 1 shows that school-level studies are much more likely than the district-level studies, and especially the state-level studies, to find no statistically significant effects of school resources on earnings.

Figure 1 also suggests that studies of workers who attended US public schools in the late 1950s onward are much more likely to find no significant effects of school resources than are studies of earlier cohorts.

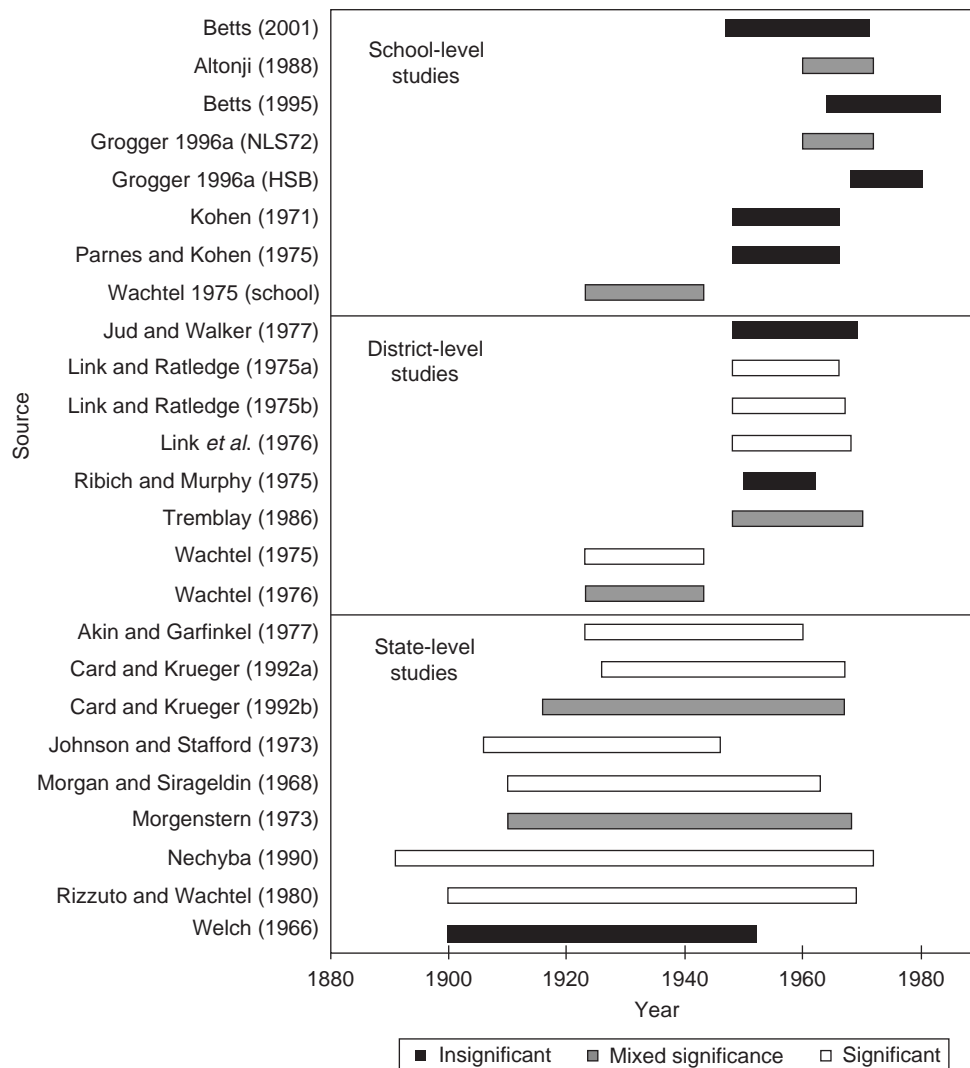


Figure 1 Patterns of statistical significance across years students were enrolled in grades 1–12, and across level of aggregation of school characteristics in the given study. The bars indicate the range of years in which adults in the given study would have attended grades 1–12. The shading indicates the percentage of estimates of the relation between school resources and earnings are significant. Dark bars (insignificant) indicate studies in which none to 32% of estimates were significant; gray bars (mixed significance) indicate that 33–65% of estimates were significant; and white bars (significant) indicate that 66–100% of estimates were significant.

Results for US women and workers in the United Kingdom

The literature that Betts (1996a) reviews includes only American studies. Remarkably, up to that point, the American literature had studied only men's earnings but not women's earnings. A number of studies since then partially fills in these voids.

For instance, Harmon and Walker (2000) and Dearden *et al.* (2002) test for an association between earnings and school resources in the United Kingdom. Harmon and Walker find that in England and Wales, there is no significant link between earnings and various measures of school resources such as pupil–teacher ratios, actual class sizes, teacher salaries, or spending on textbooks. Similarly, Dearden, Ferri, and Meghir find that in Britain

pupil–teacher ratios are not significantly related to adult men's wages.

Dearden *et al.* (2002) model outcomes for British women find some evidence that pupil–teacher ratios are significantly negatively related to women's earnings at age 33, but only for women of lower academic ability, as measured at age 11. Betts (2001) provides the first study of the relation between school resources and the earnings of women in the United States. For white women, no significant connection between school resources and wages emerges. But the pupil–teacher ratio and library books per student are significantly and positively related to black women's wages, even though overall spending per pupil in the district and teachers' salaries bear no relation to black female students' subsequent earnings.

Estimating the Size of the Relation Between School Resources and Earnings

As we have seen, the level of aggregation at which school resources are measured appears to be related to the likelihood that a statistically significant relation to wages emerges. A second and perhaps related pattern is that the size of the estimated relation between resources and earnings tends to be bigger when the researcher proxies the school resources that workers received based on their state of birth, than when resources are measured for the district or school actually attended. For example, the elasticity of earnings with respect to spending per pupil averages 0.128 in the state-level studies but only 0.096 in the studies that measure actual spending per pupil at the district level. (Elasticity is a measure of how much one variable is predicted to change given a change in a second variable. The state-level elasticity of 0.128 implies that a 1% increase in spending per pupil is associated with a 0.128% increase in earnings of workers once they leave school.)

This pattern applies to all of the school resources studied in Betts' review. For instance, the average elasticity of earnings with respect to the teacher–pupil ratio in state, district, and school studies is reported to be 0.099, 0.024, and –0.037, respectively.

In an absolute sense, are the elasticities discussed above big or small? To assess this, one can think of education as an investment. Just as a firm invests in equipment and worker training at a new plant with the expectation that the plant will generate revenues in future periods, one can think of an increase in spending per pupil today as an investment that will pay dividends in the future, as students who benefited from this spending graduate, enter the labor market, and earn more.

A difficulty with evaluating the returns to either type of investment is that all of the costs are upfront, while the benefits do not accrue until the future. A dollar today is worth more than a dollar a year from now because if the discount rate (which can be thought of as an interest rate) is 10% ($r = 0.1$), then a dollar saved today yields $(1 + r)$ dollars a year from now, or US\$1.10. The solution to this is to calculate the present discounted value of the investment project. The stream of benefits (higher wages less the costs of higher school spending) are expressed in today's money by discounting them to the base period. For example, consider a project which costs US\$1 today and produces benefits next year and the year after of US\$1.21 and US\$1.33. Then with a discount rate of $r = 0.1$, the present discounted value (PDV) is:

$$\text{PDV} = -\$1 + \$1.21/(1 + 0.1) + \$1.33/(1 + 0.1)^2 = \$1.20$$

The higher the PDV, the better is the investment project. A project with a negative PDV should not be undertaken.

Figure 2 shows estimates by Betts (1996a) of the present discounted value of increased spending per pupil, estimated at the state and district level, plotted against the discount rate. To provide a comparison, Figure 2 also shows estimates of the PDV to society of an individual staying in high school for 1 year longer or attending college for 1 year. The present value of staying in school or college is far higher than the estimates of the present value of increased spending per pupil. Indeed, as the discount rate increases, the PDV of increased spending per pupil quickly becomes negative.

One way of comparing the returns to spending of the given types is to calculate the internal rate of return, that is, the discount rate at which the present value equals zero. At discount rates beyond the internal rate of return, no rational investor would undertake the project because it would produce a negative rate of return. Thus, better investment projects will have a higher internal rate of return. But what is a reasonable internal rate of return? One reasonable point of comparison is the average real rate of interest, that is, the interest rate minus the rate of inflation.

Figure 3 plots the internal rate of return to increasing spending per pupil (based on both state- and district-level studies), with the average real interest rate overlaid as a horizontal line. Figure 3 makes clear that a student who stays in high school or college for an extra year is making a good investment. Conversely, even the more optimistic estimate of the effects of increasing spending per pupil suggests produces an internal rate of return below real interest rates. Betts (1996a) also estimates the internal rate of return to reducing the pupil–teacher ratio, and finds even lower rates of return.

One infers that policies to keep students in school longer, such as raising the school-leaving age, are likely to be more cost effective than increasing spending per pupil. This of course raises an important question: Are increases in educational resources associated with increases in years of schooling? This question is beyond the scope of this article. However, Betts (1996a, 2001) suggests either a weak or no relation between resources and educational attainment for men and women in the US, respectively.

Explanations for Variations in Results

Figure 3 suggests that the level of aggregation of the school resources matters, and that studies of American cohorts educated from the late 1950s onward are much less likely to exhibit a positive relation between school resources and earnings. Betts (1996a) also observes that many of the school-level studies focus on earnings of relatively young workers. What might explain these patterns?

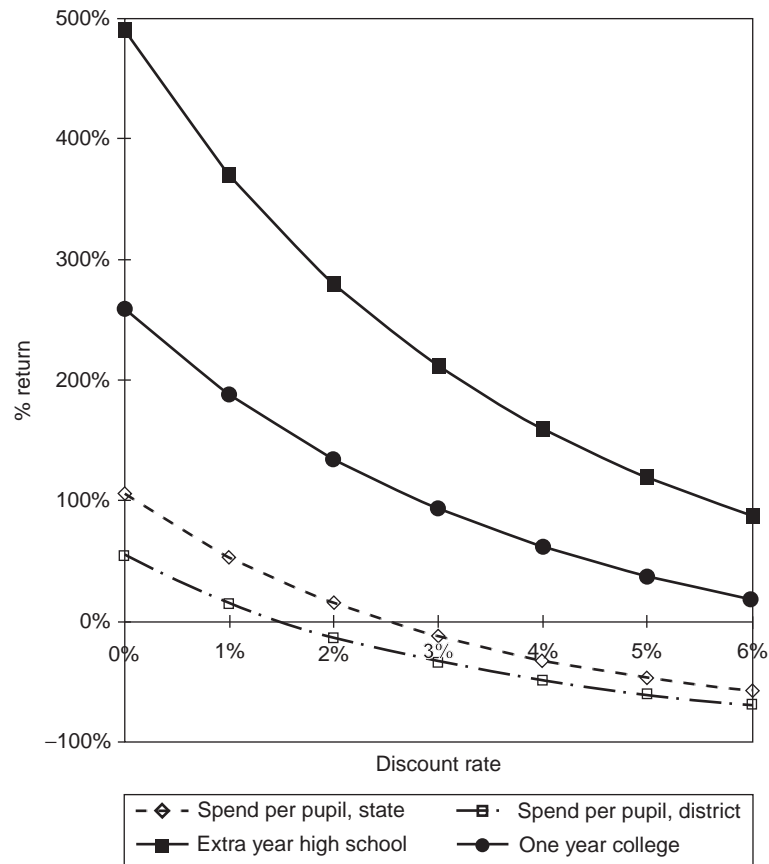


Figure 2 Net percentage return to various types of educational investments plotted against discount rate. Based on results from Betts, J. R. (1996a). Is there a link between school inputs and earnings? Fresh scrutiny of an old literature. In Burtless, G. (ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, pp 141–191. Washington, DC: Brookings Institution.

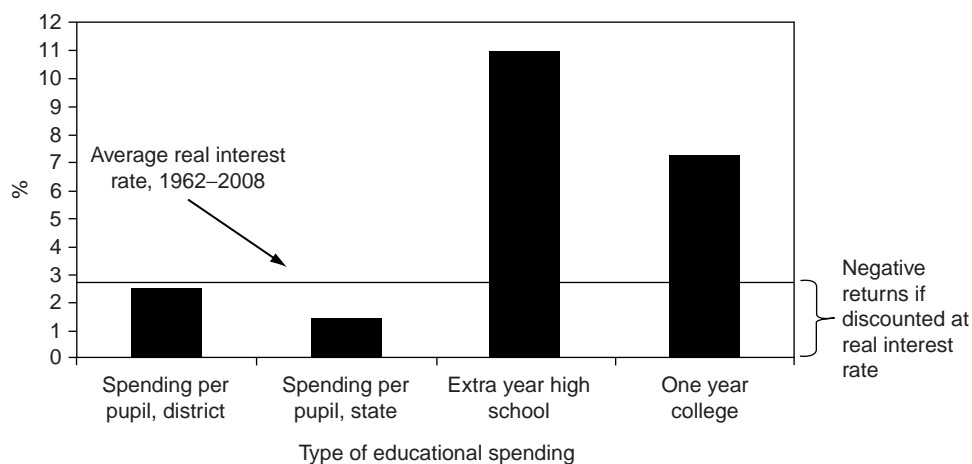


Figure 3 Comparison of internal rates of return to increased spending per pupil (from district- and state-level studies) and increased time spent enrolled in high school or college. The real interest rate was calculated as the average yield on 10-year US treasury bonds minus inflation in the US consumer price index, averaged over 1962–2008. It averaged 2.64% over this period. Historical data on treasury bond interest rates and the consumer price index were downloaded from, respectively, http://www.federalreserve.gov/releases/h15/data/Annual/H15_TCMNOM_Y10.txt and <ftp://ftp.bls.gov/pub/special.requests/cpi/cpiat.txt>.

Age Dependence

First, suppose that the effects on wages of attending schools with ample resources do not manifest themselves until workers are well into their careers. This alone could explain the general lack of significant results in the school-level studies. Betts (1996b) addresses this question using census data and also by projecting mid-career earnings of younger workers based on their occupation, and finds no evidence of positive age dependence. Betts (2001), which provides the first US evidence on school resources and women's earnings, uses a very long panel that observes women's wages from ages 18 to 50. This article finds no positive age dependence, and if anything, the effects of school resources appear to weaken as women become older.

Structural Changes?

The observation that studies focused on those educated in the late 1950s and later years are less likely to find positive relationships between school resources and earnings raises questions about whether some type(s) of structural change(s) might have occurred in American public schooling. Such changes could have arisen from diminishing returns, increasing bureaucratization and centralization of public schools, and rising teacher unionization. Betts (1996a) provides an overview of each of these ideas, in particular examining whether the returns to increased school resources may be diminishing as levels of school resources have risen over time.

Hoxby (2000) provides evidence that workers who attended large districts that face little competition from other districts tend to earn less than otherwise similar workers, and that these districts, by a variety of measures, are less productive than other districts. These findings could hold relevance for the structural change hypothesis because of a rapid consolidation of school districts that took place in the United States between 1945 and 1970 (Betts, 1996a).

Hoxby (1996) studies the relation between school resources and earnings of former students, and how the unionization of teachers mediates this relation. She replicates Grogger's (1996a) results from the same dataset that there is no significant link overall between school resources and earnings. But when she divides the sample into students who attended unionized versus nonunionized schools, she finds that some of the school-resource measures do become significant predictors of adult outcomes in the nonunionized schools. Given that the percentage of US public school teachers who were unionized tripled between 1960 and 1984, this finding could account for the apparent structural change that may have weakened the relation between school resources and earnings.

Hoxby and Leigh (2005) extend this teacher-union argument, focusing on the incentives for female graduates of selective universities to become teachers. They argue that unionization has compressed the variation in teachers' pay related to measures of teacher ability. This tendency, combined with the gradual opening of more career opportunities beyond education for women led to a sharp shift in the composition of the new female teacher labor force, toward women who graduated from the lowest tier universities.

Specific Problems Potentially Afflicting State-Level and School-Level Studies

One reason why the school-level studies may find no effects are that the sample sizes are small, in which case the studies might have little statistical power. Betts (1995) finds evidence against this hypothesis in the case of the pupil-teacher ratio and teacher education, but evidence in the case of teachers' relative salary that a lack of variation in this variable could explain the insignificant results.

A second potential problem with the school-level studies is that the school inputs are measured with error, so that their coefficients will be biased toward zero. State-level analyses might reduce this problem, since they use an average taken across schools. One can test this hypothesis directly by rerunning the school-level analyses using two-stage least squares (2SLS), with state-level school resources serving as instruments for the school-level variables. Both Grogger (1996a) and Betts (1995) take this step and find that the instrumented school inputs remain insignificant. This finding reduces the plausibility of the existence of significant measurement error. Grogger (1996b) tests for measurement error in a slightly different way and concludes that it is minor.

What issues may affect the validity of the state-level estimates?

Aggregation bias could be an issue if there is a nonlinear relationship between school resources and student outcomes. Hanushek *et al.* (1996) develop this argument in detail, and Betts (1996a) surveys evidence in other studies to this effect.

Omitted variable bias seems likely in studies that proxy the resources at the school a worker attended by using average resources in that worker's state of birth. Both Rizzuto and Wachtel (1980) and Akin and Garfinkel (1977) report that in some of their state-level models school resources become insignificant once one controls for income per capita. Betts (1995) finds that state-level resources do sometimes enter significantly in his sample, but whatever association there is does not work through correlation with the resources at the actual school attended. This raises concerns that state-wide averages of

school resources can sometimes proxy for other characteristics of a state.

Heckman *et al.* (1996) provide two key criticisms of the state-level literature. First, they show that once one allows for the returns to a year of schooling to be nonlinear, then the only workers whose earnings appear to remain correlated with state-level school resources is university graduates, who in the time frame they study constituted a distinct minority of the population. Second, these authors raise questions about nonrandom migration from state of birth, which could influence the coefficient estimates in studies that proxy school resources based on workers' state of birth.

Issues That Require Further Research

The literature suggests a fairly narrow range of estimates of the relation between school resources and earnings of students once they leave school. Some patterns emerge. In the United States, school resources may have become less strongly associated with earnings of workers over time. Moreover, studies that measure school resources at a finer level of geographic aggregation are less likely to find a positive association.

But the entire body of work appears to agree that the relation between school resources and earnings of adults ranges between none and small but positive. Even the most positive results, based on studies that measure spending per pupil based on each worker's state of birth, suggest an internal rate of return far below the rate of return to an extra year of high school or university, and below the real rate of interest.

Should we interpret these anemic effects as a sign that policymakers have been wrong to increase spending on schools, or that increased resources can never have an effect on adult earnings? The answer to both questions is probably no. There may be policy reasons for increasing school resources more generally that relate to student outcomes quite distinct from earnings. For example, suppose that increased school resources improve student attitudes along lines that society values. This could fruitfully be studied. Similarly, it would be inappropriate to extrapolate from the historical studies reviewed here to reach the conclusion that more spending can never matter. It could be that if policymakers increased spending in different ways than has been done historically, the benefits could be greater.

For example, the recent trend in the United States of coupling infusions of money with student testing and school accountability provides a distinct shift from that nation's past trends in educational spending.

Finally, note that none of the studies establishes a causal relationship between resources and students' subsequent

earnings as adults. Rather, they study correlation. Increasingly, researchers use actual experiments to examine whether an educational intervention affects test scores, or quasi-experiments that try to minimize the possibility that school resources and student outcomes are codetermined by some other set of variables or policy influences. In some experimental studies of various school interventions, it may soon become possible to study the long-term consequences of these interventions on adult outcomes including earnings.

See also: Cost-Benefit Analysis and Cost-Effectiveness Analysis; Education and Economic Growth; Education and Inequality; Education Production Functions: Concepts; Education Production Functions: Developed Country Evidence; Human Capital; Teacher Quality in Education Production; The Economics of Teachers' Unions in the United States.

Bibliography

- Akin, J. S. and Garfinkel, I. (1977). School expenditures and the economic returns to schooling. *Journal of Human Resources* **12**, 460–481.
- Betts, J. R. (1995). Does school quality matter? Evidence from the national longitudinal survey of youth. *Review of Economics and Statistics* **77**, 231–250.
- Betts, J. R. (1996a). Is there a link between school inputs and earnings? Fresh scrutiny of an old literature. In Burtless, G. (ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, pp 141–191. Washington, DC, Brookings Institution.
- Betts, J. R. (1996b). Do school resources matter only for older workers? *Review of Economics and Statistics* **78**, 638–652.
- Card, D. and Krueger, A. B. (1992a). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy* **100**, 1–40.
- Card, D. and Krueger, A. B. (1992b). School quality and black-white relative earnings: A direct assessment. *Quarterly Journal of Economics* **57**, 151–200.
- Dearden, L., Ferri, J., and Meghir, C. (2002). The effect of school quality on educational attainment and wages. *Review of Economics and Statistics* **84**, 1–20.
- Grogger, J. (1996a). Does school quality explain the recent black/white wage trend? *Journal of Labor Economics* **14**, 231–253.
- Grogger, J. (1996b). School expenditures and post-schooling earnings: Evidence from high school and beyond. *Review of Economics and Statistics* **78**, 628–637.
- Harmon, C. and Walker, I. (2000). The returns to the quantity and quality of education: Evidence for men in England and Wales. *Economica* **67**, 19–35.
- Heckman, J., Layne-Farrar, A., and Todd, P. (1996). The human capital pricing equations with an application to estimating the effect of schooling quality on earnings. *Review of Economics and Statistics* **78**, 562–610.
- Hoxby, C. M. (1996). How teachers' unions affect education production. *Quarterly Journal of Economics* **111**, 671–718.
- Hoxby, C. M. (2000). Does competition between public schools benefit students and taxpayers? *American Economic Review* **90**, 1209–1238.
- Hoxby, C. M. and Leigh, A. (2005). Wage distortion: Why America's top female college graduates aren't teaching. *Education Next* **4**, 51–56.
- Rizzuto, R. and Wachtel, P. (1980). Further evidence on the returns to school quality. *Journal of Human Resources* **15**, 240–254.

Further Reading

- Altonji, J. G. (1998). The effects of family background and school characteristics on education and labor market outcomes. Department of Economics, Northwestern University. Unpublished manuscript.
- Betts, J. R. (2001). The impact of school resources on women's earnings and educational attainment: Findings from the National Longitudinal Survey of young women. *Journal of Labor Economics* **19** (3), 635–657.
- Hanushek, E. A., Rivkin, S. G., and Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *Review of Economics and Statistics* **78**, 611–627.
- Johnson, G. and Stafford, F. P. (1973). Social returns to quantity and quality of schooling. *Journal of Human Resources* **8**, 139–155.
- Jud, G. D. and Walker, J. L. (1977). Discrimination by race and class and the impact of school quality. *Social Science Quarterly* **57**, 731–749.
- Kohen, A. I. (1971). Determinants of early labor market success among young men: Ability, quantity and quality of schooling – a preliminary report. *Research Report*. Columbus, OH: Center for Human Resource Research, The Ohio State University.
- Link, C. R. and Ratledge, E. C. (1975a). The influence of the quantity and quality of education on black–white earnings differentials: Some new evidence. *Review of Economics and Statistics* **57**, 346–350.
- Link, C. R. and Ratledge, E. C. (1975b). Social returns to quantity and quality of education: A further statement. *Journal of Human Resources* **10**, 78–89.
- Link, C., Ratledge, E., and Lewis, K. (1976). Black–white differences in returns to schooling: Some new evidence. *American Economic Review* **66**, 221–223.
- Morgan, J. and Sirageldin, I. (1968). A note on the quality dimension in education. *Journal of Political Economy* **76**, 1069–1077.
- Morgenstern, R. D. (1973). Direct and indirect effects on earnings of schooling and socio-economic background. *Review of Economics and Statistics* **55**, 225–233.
- Nechyba, T. J. (1990). The southern wage gap, human capital, and the quality of education. *Southern Economic Journal* **57**, 308–322.
- Parnes, H. S. and Kohen, A. I. (1975). Occupational information and labor market status: The case of young men. *Journal of Human Resources* **10**, 44–55.
- Ribich, T. I. and Murphy, J. L. (1975). The economic returns to increased educational spending. *Journal of Human Resources* **10**, 56–77.
- Tremblay, C. H. (1986). The impact of school and college expenditures on the wages of southern and non-southern workers. *Journal of Labor Research* **7**, 201–211.
- Wachtel, P. (1975). The effect of school quality on achievement, attainment levels, and lifetime earnings. *Explorations in Economic Research* **2**, 502–536.
- Wachtel, P. (1976). The effect on earnings of school and college investment expenditures. *Review of Economics and Statistics* **58**, 326–331.
- Welch, F. (1966). Measurement of the quality of schooling. *American Economic Review* **56**, 379–392.

Relevant Websites

- <http://www.federalreserve.gov> – Board of Governors of the Federal Reserve System.
- <ftp://ftp.bls.gov> – Bureau of Labor Statistics.

Signaling in the Labor Market

M E Page, University of California-Davis, Davis, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

The Signaling Model

The signaling model of education, usually attributed to Michael Spence (1973), is distinguished from the human capital theory of education by its premise that individual workers' innate productivity levels are identified by their years of schooling rather than enhanced by them. An implication of the model is that more-educated workers receive higher pay because education provides them with a credential, rather than because of acquired skills. Important variations on the signaling model include theories that have been developed by Arrow (1973), Layard and Psacharopoulos (1974), Riley (1975, 1979), Spence (1974, 1976), Stiglitz (1975), and Wolpin (1977). The terms signaling, screening, and sorting are often used interchangeably to describe variants of the same basic model.

Signaling theory is based on the following assumptions. (1) Individuals have different innate levels of productivity, which are not affected by their education. (2) Additional education incurs additional costs, which differ for high- and low-productivity workers. In particular, the psychic cost of schooling is higher for individuals with lower productivity levels. Those who learn easily can acquire the signal more cheaply than others. For example, they may need to spend less time studying. (3) There is asymmetric information with respect to workers' productivity: individual workers know their skill level, but potential employers do not. (4) Schooling levels can be observed without incurring a cost. As employers cannot observe potential workers' actual productivity, they instead use educational qualifications to predict productivity, make hiring decisions, and set wages, based on the assumption that individuals who have more years of education are more productive. In order for this assumption to be accurate, more productive workers must, in fact, choose more schooling. The model is based on the premise that individuals are rational and that they invest in education as long as the benefit of an additional year of schooling exceeds the cost. The benefit of an additional year of schooling is the same for high- and low-productivity workers, but the costs are higher for low-productivity workers. If the wage gain associated with education is sufficient for high-productivity workers to select into more schooling but not large enough for low-productivity workers to select more schooling, then education may sort differently skilled workers so that the employers' beliefs are ratified and equilibrium exists.

The signaling model hinges on the assumption that worker productivity is negatively related to the cost of acquiring the signal. As long as cost differences across workers result from differences in cognitive ability or tastes for learning, they may be indicative of differences in on-the-job productivity. However, if costs vary because of differences in family contributions toward tuition, etc. then individuals who face higher schooling costs may be no less productive than those who face lower schooling costs, and education will not allow employers to distinguish between high- and low-productivity workers.

An example to substantiate this is mentioned here which is based on Michael Spence (1973).

Suppose that there are two types of workers: more-productive workers have a productivity level equal to 2, and less-productive workers have a productivity level equal to 1. Suppose further that employers believe that job applicants with schooling levels equal to or greater than S^* will be type 2 workers and that those who have less than S^* years of education will be type 1. Firms pay workers according to their expected productivity level; so those with S^* or more years of schooling are paid a wage equal to 2, and those with less than S^* years of schooling are paid a wage equal to 1. Workers care about the (present discounted) value of lifetime earnings, which is E_1 for those with less than S^* years of schooling and E_2 for those with more than S^* years of schooling. The relationship between lifetime earnings and schooling is depicted in **Figure 1**.

Individuals will choose to invest in the level of education that maximizes their net benefits (total benefits minus total costs). For simplicity, assume that the only benefit that workers care about is their earnings. Then individuals will choose the level of education that produces the biggest difference between the (present) value of lifetime earnings and the cost of education. If education were costless then all workers would want to acquire the signal of S^* , but the signaling model hinges on the assumption that costs vary across individuals. Suppose as in **Figure 2**, that type 1 workers face a cost of C for each year of schooling but that type 2 workers, who find school easier, face a cost of $C/2$. It is easy to see that for type 1 workers the difference between lifetime earnings and C is maximized when they choose 0 years of education. On the other hand, for type 2 workers, the difference between lifetime earnings and $C/2$ is largest at S^* . Thus, type 1 and type 2 workers sort into different levels of schooling that

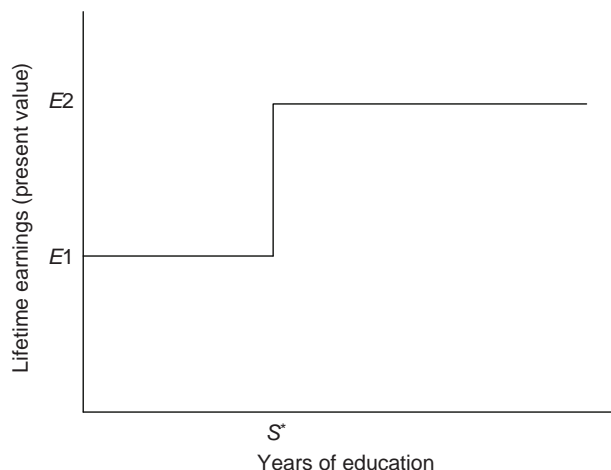


Figure 1 Lifetime benefits associated with education. Adapted from Spence, A. M. (1973). Job market signaling. *Quarterly Journal of Economics* 87, 355–374.

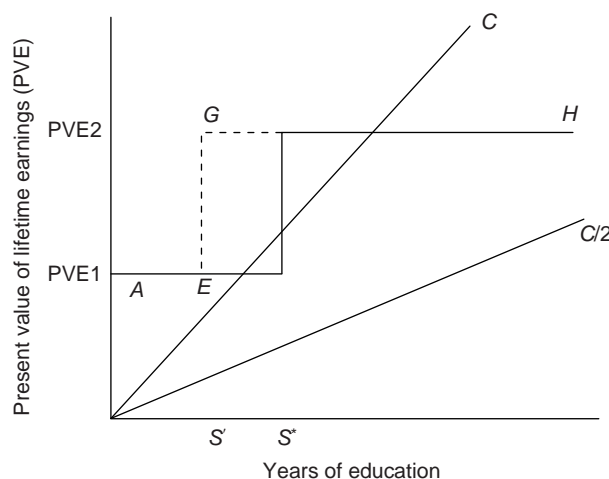


Figure 3 Benefits and costs associated with alternative signals. Adapted from Spence, A. M. (1973). Job market signaling. *Quarterly Journal of Economics* 87, 355–374.

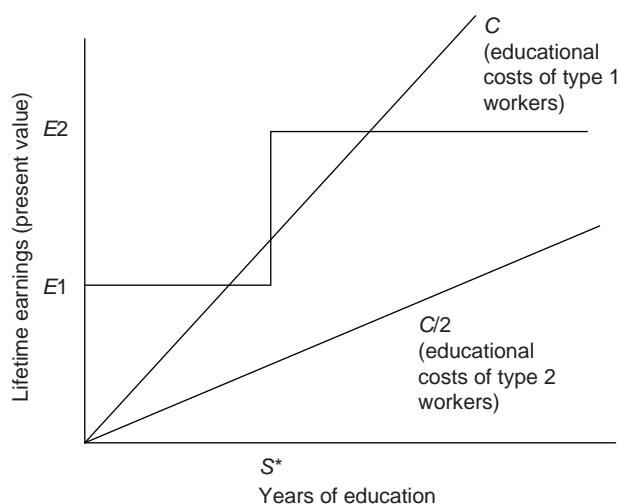


Figure 2 Lifetime benefits and costs associated with education. Adapted from Spence, A. M. (1973). Job market signaling. *Quarterly Journal of Economics* 87, 355–374.

are consistent with the employers beliefs, and S^* works as a valid signal.

Note that not all years of education necessarily provide valid signals. For example, in **Figure 3**, employers use S' years of education instead of S^* to differentiate between high- and low-productivity workers. Therefore, the earnings structure is now depicted as the distance AEGH. Here, both type 1 and type 2 workers maximize the difference between lifetime earnings and the cost of education by choosing S' . Since all job applicants choose the same level of education, S' does not signal anything about workers' productivity.

As stressed by Weiss (1995), the human capital and signaling models of education are not necessarily mutually exclusive. Education may simultaneously enhance

workers' productivity and act as a signal about their innate abilities. In such a case, signaling can be thought of as an extension of a human capital model in which some productivity differences that firms cannot observe are correlated with schooling costs.

Empirical Evidence

Empirically disentangling the relative importance of the human capital model and the signaling model has proven to be difficult. Both models predict that more-educated workers will earn higher wages, so the positive relationship that we observe between schooling and earnings does not provide useful information. There is no universally accepted method for separately identifying education's productivity-enhancing role from its signaling role.

One approach that has been used is to compare educational wage gaps for workers of different ages or years of work experience assuming that over time, it will become easier for employers to directly observe their employee's actual productivity (Cohn *et al.*, 1987; Layard and Psacharopoulos, 1974; Mendes de Oliveira *et al.*, 1989; Psacharopoulos, 1979; Rao and Datta, 1989; Wolpin, 1977). The idea is that if education is just a signaling device then it serves a useful function early in a worker's career, but as direct information on workers' actual productivity accumulates, the relationship between schooling and earnings should diminish. In fact, earnings differentials between more- and less-educated workers increase with age, which may be taken as evidence in favor of the human capital model. On the other hand, advocates of the signaling model argue that continued growth in earnings differentials occurs because the signal inherent in

educational attainment is an accurate predictor of individuals' productivity.

Another set of studies compare earnings across industries or professions where signaling is likely to be important, to industries or professions where it is not. Cohn *et al.* (1987), De Wit and Van Winden (1989), Katz and Ziderman (1980), Riley (1979), and Wolpin (1977), for example, compare the education and earnings of self-employed and salaried workers, based on the assumption that signaling should be irrelevant to the self-employed. Since self-employed workers do not have to invest in educational signals for potential employers, they will invest in less schooling. These studies have produced mixed evidence on the relative importance of the two hypotheses.

A third set of studies focuses on sheepskin effects, or wage returns to particular credentials and/or diploma years. The argument is that the productivity-enhancing part of education should be proportional to the time spent in school; therefore, a wage return to the diploma itself, controlling for years of education, may be interpreted as evidence in favor of the signaling hypothesis. Groot and Oosterbeek (1994), Hartog (1983), Hungerford and Solon (1987), Jaeger and Page (1996), Layard and Psacharopoulos (1974), and Weiss (1983) have all estimated variants of sheepskin models. The studies by Hungerford and Solon, and Jaeger and Page both find that relative to a year of high school or college schooling, there are higher wage returns associated with high school and college graduation. However, both of these studies note that the existence of sheepskin effects is not necessarily a corroboration of the signaling model: it is impossible to tell whether the observed differences result from a signal about completers' versus noncompleters' abilities or whether those who actually complete their degree increase their productivity more than those who do not.

Tyler *et al.* (2000) address this issue when estimating the signaling value of the general equivalency diploma (GED) by comparing the earnings of individuals who had the same GED test scores but lived in states with different passing standards. Individuals with the same test score are assumed to have acquired equal amounts of human capital and to be equally productive. They find that whites who obtain a GED have earnings that are 10–19% higher than whites with the same test score who do not get the credential because they live in a state with a higher standard. This suggests that the GED has an important signaling effect on earnings. On the other hand, Cameron and Heckman (1993) and Heckman and LaFontaine (2006) find that the wages of those with a GED are no higher than the wages of high school dropouts, which suggests that the GED is not a substitute for high school.

Lang and Kropp (1986) test the signaling model by comparing enrolment rates of age groups that are bound by compulsory schooling laws to the enrolment rates of age groups that are not directly affected. The idea is that

under the signaling model, a state compulsory school attendance law should increase the educational attainment of high-ability workers who are not directly affected by the law, because it lowers the average ability (and thus the wage) associated with the compulsory level of education. This gives high-ability individuals an incentive to get more schooling. Human capital theory predicts that such laws should only affect those individuals who are directly constrained. Lang and Kropp find that compulsory schooling laws do increase enrolment rates for age groups that are not directly affected, which supports the notion that education acts as a signaling device.

In a similar vein, Bedard (2001) compares high school dropout rates in labor markets with differential university access. The signaling model predicts that an environment in which more individuals are constrained from entering college will be characterized by higher high school graduation rates: when there are fewer barriers to college, then some previously constrained but relatively high-ability students will become university enrollees, and the average skill of high school graduates who do not go on to college will fall. There will be a commensurate decline in the wage associated with a high school diploma, which will reduce the incentive to obtain one. As a result, the least-able high school graduates will choose to drop out instead. Bedard finds evidence of such responses.

Kroch and Sjoblom (1994) suggest that if education acts as a signal, then within a cohort, the signal should be related to an individual's position in the distribution of education. They estimate models that include both absolute and relative years of education and find that while years of education are always positively associated with earnings, one's rank in the education distribution is not statistically significant. They interpret this as evidence that the value of schooling is not primarily due to signaling.

Weiss (1995) summarizes a number of additional studies that provide indirect evidence on the prevalence of signaling in education, including detailed discussions of Altonji's (1995) and Kang and Bishop's (1986) papers on the earnings effect of course work in school, and Card and Krueger's (1992) paper on the effects of school quality on wage returns to education. He concludes that labor economists have been unreasonably skeptical of signaling theory, given its ability to explain empirical regularities that are not explained by other models – such as why wage returns to education seem to be so much larger than the returns to courses taken in secondary school, and why the variance of wages increases with education.

In summary, empirical tests of the signaling model were most prevalent during the late 1970s and 1980s, but many of these tests were of questionable validity. Since the late 1980s, there have been fewer studies that have attempted to identify the role that signaling plays in the labor market, but the literature has laid more emphasis on finding external sources of identification. Recent studies that use

plausibly random variation in degree receipt while controlling for observable measures of productivity (such as the study by Tyler *et al.*, 2000) find evidence in support of signaling for certain degrees (such as the GED). There are also labor-market relationships such as those discussed by Weiss (1995) that seem to provide indirect evidence in support of the signaling model. However, the number of studies that use convincing empirical strategies to test the signaling model is short, and the evidence to date is mixed.

Conclusion

More than 30 years have passed since the signaling model first gained attention; yet, much is still unknown about the relative importance of signaling versus human capital theories. Many of the implications of the two models are the same, thus it has proven hard to empirically distinguish between them. The results of existing research do not conclusively reject the signaling theory, and more recent work has produced evidence that signaling does play a role in the labor market. Nevertheless, the extent to which education acts as a sorting device in addition to (or instead of) augmenting productivity, is still unknown.

See also: Human Capital.

Bibliography

- Altonji, J. (1995). The effects of high school curriculum on education and labor market outcomes. *Journal of Human Resources* **30**, 409–438.
- Arrow, K. J. (1973). Higher education as a filter. *Journal of Public Economics* **2**, 193–216.
- Bedard, K. (2001). Human capital versus signaling models: University access and high school dropouts. *Journal of Political Economy* **109**, 749–775.
- Cameron, S. V. and Heckman, J. J. (1993). The nonequivalence of high school equivalents. *Journal of Labor Economics* **11**, 1–47.
- Card, D. and Krueger, A. (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy* **100**, 1–40.
- Groot, W. and Oosterbeek, H. (1994). Earnings effects of different components of schooling: Human capital versus screening. *Review of Economics and Statistics* **76**, 317–321.

- Heckman, J. J. and LaFontaine, P. A. (2006). Bias-corrected estimates of GED returns. *Journal of Labor Economics* **24**, 661–700.
- Kang, S. and Bishop, J. (1986). Effects of curriculum on labor market success immediately after high school. *Journal of Industrial Teacher Education* **23**, 14–29.
- Kroch, E. A. and Sjoblom, K. (1994). Schooling as human capital or a signal: Some evidence. *Journal of Human Resources* **29**, 156–180.
- Lang, K. and Kropp, D. (1986). Human capital versus sorting: The effects of compulsory attendance laws. *Quarterly Journal of Economics* **101**, 609–624.
- Layard, P. R. G. and Psacharopoulos, G. (1974). The screening hypothesis and the returns to education. *Journal of Political Economy* **82**, 985–998.
- Riley, J. (1975). Competitive signaling. *Journal of Economic Theory* **10**, 175–186.
- Riley, J. (1979). Testing the educational screening hypothesis. *Journal of Political Economy* **87**, 227–252.
- Spence, A. M. (1973). Job market signaling. *Quarterly Journal of Economics* **87**, 355–374.
- Spence, A. M. (1974). *Market Signaling: Informational Transfer in Hiring and Related Screening Processes*. Cambridge, MA: Harvard University Press.
- Spence, A. M. (1976). Competition in salaries, credentials, and signaling prerequisites for jobs. *Quarterly Journal of Economics* **90**, 51–75.
- Stiglitz, J. (1975). The theory of screenings, education, and the distribution of income. *American Economic Review* **65**, 283–300.
- Tyler, J. H., Murnane, R. J., and Willett, J. B. (2000). Estimating the labor market signaling value of the GED. *Quarterly Journal of Economics* **115**, 431–468.
- Weiss, A. (1983). A sorting-cum-learning model of education. *Journal of Political Economy* **91**, 420–442.
- Weiss, A. (1995). Human capital vs. signaling explanations of wages. *Journal of Economic Perspectives* **9**, 133–154.
- Wolpin, K. I. (1977). Educational screening. *American Economic Review* **67**, 949–958.

Further Reading

- Hanushek, E. (1986). The economics of schooling. *Journal of Economic Literature* **24**(3), 1141–1177.
- Lang, K. (1994). Does the human capital/educational sorting debate matter for development policy? *American Economic Review* **94**, 353–358.
- Meyer, R. (1982). Job training in the schools. In Taylor, R., Rosen, H., and Pratzner, F. (eds.) *Job Training for Youth, the Contribution of the United States Employment Development System*, pp 307–344. Columbus, OH: The National Center for Research in Vocational Education/Ohio State University.
- Mincer, J. (1974). *Schooling, Experience, and Earnings*. New York: Columbia University Press.

The Economics of High School Dropouts

R W Rumberger, University of California, Santa Barbara, Santa Barbara, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

One of the major educational challenges in virtually all industrialized nations is ensuring that all students graduate from high school. Although many countries allow students to leave school prior to completing upper secondary school, a high school diploma is increasingly viewed as a minimal requirement for entry into the labor market and for further postsecondary education. In fact, with the economy generating an increasing number of jobs that require at least some postsecondary schooling, students who earn no more than a high school diploma will likely have diminishing economic prospects.

Despite the increasing importance of graduating from high school, a large segment of the student population in many countries leave high school prior to graduation. This article provides a brief review of the research on who drops out of high school, the economic and social consequences of dropping out, the causes of dropping out, and what can be done to improve high school graduation rates. The article relies primarily on literature in the United States.

Who Drops Out of High School?

Identifying the number of students who drop out of school and graduate each year is complicated by the difficulty of tracking students over time, and the sources of data currently available to track them. There are also a number of measures of dropout and graduation rates that are used to gauge the extent of the problem.

One measure is the annual dropout rate. According to US Census survey data, 414 000 students dropped out of grades 10–12 in 2004–05, or about 3.8% of the students enrolled (Laird *et al.*, 2007, Table 1). This is no doubt an undercount because it excludes students who dropped out before grade 10. Dropout rates are higher for some demographic groups: males have higher dropout rates than females; Blacks and Hispanics have higher dropout rates than Asians and Whites; and students from low-income families have higher dropout rates than students from high-income families. Annual dropout rates have been falling steadily over the last 25 years, from a high of 6.7% in 1979 to the current rate of 3.8% (Laird *et al.*, 2007, Table 4).

But annual dropout rates understate the likelihood that a student will drop out sometime during his or her

educational career. A better gauge is the proportion of dropouts in the population, which is referred to as the *status* dropout rate. Again according to US Census survey data, there were 3.5 million dropouts aged 16–24 years in October 2005, representing 9.4% of the population (Laird *et al.*, 2007, Table 6). These rates too are higher for the same demographic groups mentioned above, and they too have declined over the last 25 years, from a high of 14.6% to the current rate of 9.4% (Laird *et al.*, 2007, Table 7).

Yet, dropout rates alone may not be sufficient to reveal the extent of the problem. Census data have been criticized because they rely on respondents' self-reported educational status, which respondents may overstate. Moreover, the Census considers persons who earn equivalent high school credentials by taking the general education development (GED) examination as completing high school even though there is extensive evidence that alternative credentials do not provide the same economic benefits as a traditional diploma. Thus both federal and state governments are attempting to measure the proportion of entering ninth-grade students who earn a regular diploma within 4 years, which is known as the ninth-grade cohort graduation rate. Such a rate is particularly difficult to measure because it requires tracking students over several years. This is problematic, in part, because students often transfer from one educational setting to another during their high school careers. In addition, some students are retained, especially in the ninth grade, when they fail to earn enough credits to be promoted to the next grade level. A number of specific measures are currently used to estimate the ninth-grade graduation rate, and all have limitations. Recently, the federal government has adopted one measure known as the averaged freshman graduation rate, and has produced estimates for the nations' public schools and for the public schools in each state. These estimates show an averaged freshman graduation rate of 75% for all public schools nationally in 2003–04, with state rates varying from a low of 61% in South Carolina to a high of 88% in Nebraska (Laird *et al.*, 2007, Table 12). Other measures and data sources suggest the national high school graduation rate, especially when it includes students who earn a diploma after the normal 4 years of high school, exceeds 80%.

The Organization for Economic and Cooperative Development (OECD) computes the percentage of upper secondary graduates to the population at the typical age of graduation. The average graduation rate among OECD

countries in 2004 was 81%, ranging from 38% in Mexico to 100% in Norway, with the US rate at 75% (OECD, 2006, Table A2.1)

What Are the Consequences?

Dropping out of school has economic and social consequences both for dropouts themselves and for the country as a whole. First, dropouts have difficulty finding jobs. Government data show that more than 33% of 16- to 24-year-olds in October 2005 who dropped out of school the previous year were unemployed (Snyder *et al.*, 2007, Table 382). Among all 16- to 24-year-olds in October 2005, the unemployment rate for dropouts was 18.5%, compared to 12.1% for high school graduates and 5.0% for 4-year college graduates (Snyder *et al.*, 2007, Table 375). Second, even if they find a job, dropouts earn substantially less than high school graduates. In 2005, the median annual earnings of high school dropouts were almost 25% less than the income of high school graduates (Snyder *et al.*, 2007, Table 378). Over their working lives, the Census Bureau estimates that dropouts will earn about US\$200 000 less than high school graduates (Day and Newburger, 2002, Figure 3).

One reason for dropouts' poor economic outcomes is their low levels of education. Yet, dropouts can return to school. Almost half of 1988 eighth-grade students who dropped out of school completed either a regular high school diploma (16%) or a GED or alternative certificate (29%) within 2 years of their scheduled graduation date in 1992 (Berkthold *et al.*, 1998, Table 1). Dropouts who completed high school were more likely to enrol in postsecondary education than students who did not complete high school (42% vs. 14%) (Berkthold *et al.*, Table 15). Nonetheless, dropouts as a group are much less likely to enrol in postsecondary education than high school graduates, even though most states allow dropouts to enrol in community colleges without a high school diploma.

Dropouts experience other negative outcomes (Belfield and Levin, 2007). Dropouts have poorer health and higher rates of mortality than high school graduates; they are more likely to engage in criminal behavior and be incarcerated over their lifetimes compared to graduates. They are also more likely to require public assistance and less likely to vote. Although the observed relationship between dropping out and these economic and social outcomes does not necessarily imply a causal relationship, a growing body of research evidence has, in fact, demonstrated one. This suggests that efforts to reduce dropout rates would, in fact, reduce these negative economic and social outcomes.

The negative outcomes from dropouts generate huge social costs. Federal, state, and local governments collect fewer taxes from dropouts. The government also

subsidizes the poorer health, higher criminal activity, and increased public assistance of dropouts. One recent study estimated that each new high school graduate would generate more than US\$200 000 in government savings, and that cutting in half the dropout rate from a single cohort of dropouts would generate more than US\$45 billion in savings (Belfield and Levin, 2007).

A number of economic, demographic, and educational trends could exacerbate these problems in the future. As the US economy moves toward a higher-skilled labor force, high school dropouts will have an even harder time surviving economically. The numbers of students who are generally at greater risk of school failure – students from poor and low-income households, and racial, ethnic, and linguistic minorities – are increasing in the nation's schools. Finally, the growing push for accountability in the nation's public schools that has produced policies to end social promotion and to institute high school exit exams could increase the number of students who fail to complete high school.

Why Do Students Drop Out?

Understanding why students drop out of school is the key to addressing this major educational problem; yet, identifying the causes of dropping out is extremely difficult. Like other forms of educational achievement (e.g., test scores), the causes of dropping out are influenced by an array of proximal and distal factors related to both the individual student and to the family, school, and community settings in which the student lives.

Dropouts themselves report a variety of reasons for leaving school, including school-, family-, and work-related reasons (Rotermund, 2007). The most specific reasons cited by tenth-graders of 2002 who dropped out were “missed too many school days” (44%); “thought it would be easier to get a GED” (41%); “getting poor grades/failing school” (38%); “did not like school” (37%); and “could not keep up with schoolwork” (32%). But these reasons do not reveal the underlying causes of why students quit school, particularly those causes or factors in elementary or middle school that may have contributed to students' attitudes, behaviors, and school performance immediately preceding their decision to leave school. Moreover, if many factors contribute to this phenomenon over a long period of time, it is virtually impossible to demonstrate a causal connection between any single factor and the decision to quit school. Despite this difficulty, two types of factors have been identified that contribute to or increase the likelihood that students drop out of school: (1) individual factors associated with students' characteristics, attitudes, behaviors, and experiences; and (2) contextual factors associated with students' families, schools, communities, and peers.

Individual Factors

A variety of individual factors are associated with dropping out (Rumberger, 2004a), including several demographic factors. Generally, dropout rates are higher among males, Blacks and Hispanics, immigrants, and language-minority students. Attitudes also affect dropout rates. Dropout rates are also higher among students who have low educational and occupational aspirations. Several activities and behaviors also predict dropout rates, including absenteeism, misbehavior in school, and pregnancy. Finally, poor academic achievement is a strong predictor of dropping out. Together, these factors support the idea that dropping out is influenced by both the social and academic experiences of students.

In addition to these proximal factors, a number of distal factors are associated with dropping out, such as student mobility. Both residential mobility (changing residences) and school mobility (changing schools) increase the risk of dropping out of high school. Student mobility may represent a less-severe form of student disengagement or withdrawal from school. Another distal factor is grade retention. Although retention may have some positive impact on academic achievement in the short run, numerous studies have found that it greatly increases the likelihood that students will drop out of school. Finally, a number of long-term studies have found that lack of early academic achievement and engagement (e.g., attendance, misbehavior) in elementary and middle school predicts withdrawal from high school.

Institutional Factors

While individual factors clearly contribute to students' decisions to drop out of school, individual attitudes and behaviors are shaped by the various settings or contexts in which students live – families, schools, communities, and peer groups. The importance of context in shaping behavior, including dropping out, was acknowledged in a report by the National Research Council Panel on High-Risk Youth (1993) that argued that too much emphasis has been placed on high-risk youth and their families, and not enough on the high-risk settings in which they live and go to school. A number of factors within students' families, schools, and communities (and peer relationships) predict dropping out.

Families

Family background is widely recognized as the single most important contributor to success in school. Socio-economic status, most commonly measured by parental education and family income, is a powerful predictor of school achievement and dropout behavior. Parental education influences students' aspirations and educational support; while family income allows parents to provide more resources to support their children's education,

including access to better-quality schools, after-school and summer-school programs, and more support for learning within the home. In addition, students whose parents monitor and regulate their activities, provide emotional support, encourage independent decision-making (known as authoritative parenting style), and are generally more involved in their schooling are less likely to drop out of school. Additionally, students in single-parent and step-families are more likely to drop out of school than students in two-parent families.

Schools

It is widely acknowledged that schools exert powerful influences on student achievement, including dropout rates. Four types of school characteristics influence student performance: social composition of the schools, structural characteristics, school resources, and school policies and practices.

1. The social composition of schools – the characteristics of students attending the schools, particularly the socioeconomic composition of the student body – predicts dropping out even after controlling for the individual factors that influence dropping out.
2. The second characteristic has to do with the structural characteristics of schools, such as size, location, and school control (public vs. private). Dropout rates from Catholic and other private schools are lower than dropout rates from public schools, even after controlling for differences in the background characteristics of students. Yet, students from private schools typically transfer to public schools instead of or before dropping out, so that student turnover rates in private schools are not statistically different from turnover rates in public schools. Smaller schools also have lower dropout rates. What is less clear is whether structural characteristics themselves account for these differences or whether they are related to differences in student characteristics and school resources often associated with the structural features of schools.
3. The third type of characteristic concerns school resources. Resources, in particular student/teacher ratios and teacher quality, appear to influence dropout rates even after controlling for a host of individual and contextual factors that might also influence dropout rates.
4. The final type has to do with school policies and practices. In particular, academic and social climate – as measured by school attendance rates, students taking advanced courses, and student perceptions of a fair discipline policy – predict school dropout rates, even after controlling for the background characteristics of students as well as the resource and structural characteristics of schools.

School factors contribute to student withdrawal in two ways. One way is indirectly, by creating conditions that

influence student engagement and their voluntary withdrawal from school. Another way is directly, through explicit policies and conscious decisions by school personnel that cause students to involuntarily withdraw from school. These rules and actions may concern low grades, poor attendance, misbehavior (such as zero-tolerance policies), or being overage, and may lead to suspensions, expulsions, or forced transfers. This form of withdrawal is school initiated and contrasts with the student-initiated form mentioned above. Some schools, for example, contribute to students' involuntary departure from school by systematically excluding and discharging troublemakers and other problematic students.

Community and peers

In addition to families and schools, communities and peer groups can influence students' withdrawal from school. Differences in neighborhood characteristics can help explain differences in dropout rates among communities, apart from the influence of families. Some evidence suggests that there is a threshold or tipping point on the quality of neighborhoods that results in particularly high dropout rates in the most disadvantaged neighborhoods. Poor communities may influence child and adolescent development through the lack of resources (playgrounds and parks, after-school programs) or negative peer influences. Community residents may also influence parenting practices over and above parental education and income. Students living in poor communities may also be more likely to have dropouts as friends, which increases the likelihood of dropping out of school. Another way that communities can influence dropout rates is by providing employment opportunities both during or after school. Relatively favorable employment opportunities for high school dropouts, as evidenced by low neighborhood-unemployment rates, appears to increase the likelihood that students will drop out, while more favorable economic returns to graduating, as evidenced by higher salaries of high school graduates compared to dropouts, tend to lower dropout rates. Working long hours in high school can increase the likelihood of dropping out, although the impact of working in high school depends on the type of job held and on the student's gender.

What Can Be Done?

Knowledge about why students drop out suggests several things about what can be done to design effective dropout-intervention strategies (Rumberger, 2004b). First, because dropping out is influenced by both individual and institutional factors, intervention strategies can focus on either or both sets of factors. That is, intervention strategies can focus on addressing the individual values, attitudes, and behaviors that are associated with dropping out without

attempting to alter the characteristics of families, schools, and communities that may contribute to those individual factors. Many dropout-prevention programs pursue such programmatic strategies by providing would-be dropouts with additional resources and supports to help them stay in school. Alternatively, intervention strategies can focus on attempting to improve the environmental contexts of potential dropouts by providing resources and supports to strengthen or restructure their families, schools, and communities. Such systemic strategies are often part of larger efforts to improve the educational and social outcomes of at-risk students more generally.

Second, because dropping out is associated with both academic and social problems, effective prevention strategies must focus on both arenas. That is, if dropout-prevention strategies are going to be effective, they must be comprehensive by providing resources and supports in all areas of students' lives. As dropouts leave school for a variety of reasons, services provided them must be flexible and tailored to their individual needs.

Third, because the problematic attitudes and behaviors of students at risk of dropping out appear as early as elementary school, dropout-prevention strategies can and should begin early in a child's educational career. Dropout-prevention programs often target high school or middle school students who may have already experienced years of educational failure or unsolved problems. Dropout-recovery programs must also attempt to overcome longstanding problems in order to get dropouts to complete school; yet, such programs may be costly and ineffective. In contrast, early intervention may be the most powerful and cost-effective approach to dropout prevention.

The longstanding interest in the problem of school dropouts has spurred a great deal of effort in designing dropout interventions. Unfortunately, there is little scientific evidence on the effectiveness of these dropout interventions. One reason is that there have been relatively few rigorous evaluations of dropout-intervention programs. Another reason is that the scientifically rigorous evaluations that do exist, often fail to demonstrate program effectiveness. The US Department of Education established the What Works Clearinghouse in 2002 to review scientific evidence on the effectiveness of a variety of educational interventions (US Department of Education, 2007). To date, only 11 dropout-prevention programs have been identified that meet rigorous scientific standards. Although 8 of the 11 programs have demonstrated to be effective in reducing dropout rates, improving students' progress toward graduation, only one has been demonstrated to improve graduation rates. Despite the dearth of research evidence, evaluations of proven or at least promising programmatic and systemic approaches to dropout prevention suggest some important features that contribute to their apparent effectiveness.

Programmatic Approaches

Programmatic approaches to dropout prevention do not attempt to change existing institutions serving most students, but rather create alternative programs or institutions to target students who are somehow identified as at risk of dropping out. Although evaluations show that dropout-prevention programs differed in the types of students they enrolled, the curricula and services they provided, and the way they were structured, there appear to be several common features among effective programs (Rumberger, 2004b):

- a nonthreatening environment for learning;
- a caring and committed staff who accepted personal responsibility for student success;
- a school culture that encouraged staff risk-taking, self-governance, and professional collegiality;
- a school structure that provided for a low student-teacher ratio and a small size to promote student engagement.

Creating successful alternative programs presents two challenges. First, alternative programs can have difficulty in attracting students because of negative perceptions by students, parents, and educators that such schools are a dumping ground for bad students and that they symbolize the failure of the regular system. Second, because of their low regard, such programs often have a hard time competing for resources with regular school programs.

Systemic Solutions

Systemic solutions have the potential to reduce dropping out in a much large number of students by improving some of the environmental factors in families, schools, and communities that contribute to dropout behavior. Although the promise of systemic solutions to the dropout problem is great, the reality is not. The reason is simply that systemic changes are extremely difficult to achieve because they involve making fundamental changes in the way institutions work individually and within the system in which they are a part. For example, improving existing high schools requires changing two basic features of schools: the commitments and competencies of the people (teachers, administrators, and staff) and the organizational structure (size, staffing ratio, curriculum design, services, etc.). Yet, it is extremely difficult to identify the resources, technical support, and incentives to transform or restructure existing schools in order to create those features.

Despite the difficulties of designing and implementing systemic solutions, a number of whole-school reform models have been developed and shown to improve student achievement and high school graduation rates. A recent review of these programs identified a number of challenges that these programs have been able to successfully address (Herlihy and Quint, 2007):

- assisting students who enter high school with poor academic skills;
- improving instructional content and practice;
- creating a personalized and orderly learning environment;
- providing work-based learning opportunities and preparing students for the world beyond high school;
- stimulating change in overstressed high schools.

Efforts to reform other institutions that serve at-risk youth have proved more difficult. One ambitious systemic reform effort was the New Futures Initiative, promoted and funded by the Annie E. Casey foundation beginning in 1988 (White and Wehlage, 1995). New Futures was an attempt to build new collaborative structures among existing public and private institutions in five cities (Dayton, OH; Lawrence, MA; Little Rock, AR; Pittsburgh, PA; Savannah, GA) to address the problems of at-risk youth, including school dropout. The key strategy was to establish an oversight collaborative in each city with representation from public and private sector agencies to 'identify youth problems, develop strategies, and set timelines for addressing these problems, coordinate joint agency activities, and restructure educational and social services.' Evaluations of this ambitious, systemic-reform effort found that it did little to reduce dropout rates and other problems of at-risk youth. As with other systemic reform efforts, the evaluations found little incentive or support from the intervention for changing the fundamental functioning of schools.

Conclusions

Successfully addressing the dropout problem in the United States will require both capacity and will. Capacity requires technical expertise to develop and implement effective dropout-prevention and recovery programs, as well as more ambitious systemic school reforms. While some schools have such capacity, most require additional resources, technical expertise, and incentives to restructure existing schools. The development of such capacity will require political will; but even with the will to reform schools, it is unlikely that the United States will ever eliminate disparities in dropout rates among racial and ethnic groups without eliminating disparities in the resources of families, schools, and communities.

See *also*: Education and Inequality; Human Capital; Signaling in the Labor Market; Student Incentives.

Bibliography

- Belfield, C. and Levin, H. M. (eds.) (2007). *The Price We Pay: Economic and Social Consequences of Inadequate Education*. Washington, DC: Brookings Institution Press.
- Berkold, J., Geis, S., and Kaufman, P. (1998). *Subsequent Educational Attainment of High School Dropouts*. Washington, DC: U.S. Department of Education.

- Day, J. C. and Newburger, E. C. (2002). *The Big Payoff: Educational Attainment and Synthetic Estimates of Work-Life Earnings*. Washington, DC: U.S. Census Bureau.
- Herlihy, C. and Quint, J. (2007). *Emerging Evidence on Improving High School Student Achievement and Graduation Rates: The Effects of Four Popular Improvement Programs*. Washington, DC: National High School Center, American Institutes for Research.
- Laird, J., Kienzi, G., DeBell, M., and Chapman, C. (2007). *Dropout Rates in the United States: 2005*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- National Research Council Panel on High-Risk Youth (1993). *Losing Generations: Adolescents in High-Risk Settings*. Washington, DC: National Academies Press.
- Organization for Economic Co-operation and Development (OECD) (2006). *Education at a Glance: OECD Indicators 2006*. Paris: OECD.
- Rotermund, S. (2007). *Why Students Drop Out of High School: Comparisons from Three National Surveys*. Santa Barbara, CA: California Dropout Research Project, University of California, Santa Barbara.
- Rumberger, R. W. (2004a). Why students drop out of school. In Orfield, G. (ed.) *Dropouts in America: Confronting the Graduation Rate Crisis*, pp 131–155. Cambridge, MA: Harvard Education Press.
- Rumberger, R. W. (2004b). What can be done to reduce school dropouts? In Orfield, G. (ed.) *Dropouts in America: Confronting the Graduation Rate Crisis*, pp 243–254. Cambridge, MA: Harvard Education Press.
- Snyder, T. D., Dillow, S. A., and Hoffman, C. M. (2007). *Digest of Education Statistics, 2007*. (NCES 2007-017. U.S. Department of Education, National Center for Education Statistics). Washington, DC: U.S. Government Printing Office.
- U.S. Department of Education, Institute of Education Sciences (2007). *What Works Clearinghouse*. <http://ies.ed.gov/ncee/wwc/overview> (accessed June 2009).
- White, J. A. and Wehlage, G. (1995). Community collaboration: If it is such a good idea, why is it so hard to do? *Educational Evaluation and Policy Analysis* **17**, 23–38.
- Bridgeland, J. M., Dilulio, J. J., and Morison, K. B. (2006). *The Silent Epidemic: Perspectives on High School Dropouts*. Washington, DC: Civil Enterprises.
- Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research* **59**, 117–142.
- Jimerson, S. R., Anderson, G. E., and Whipple, A. D. (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. *Psychology in the Schools* **39**, 441–457.
- Mishel, L. and Roy, J. (2006). *Rethinking High School Graduation Rates and Trends*. Washington, DC: Economic Policy Institute.
- National Research Council, Committee on Increasing High School Students' Engagement and Motivation to Learn (2004). *Engaging Schools: Fostering High School Students' Motivation to Learn*. Washington, DC: The National Academies Press.
- Orfield, G. (ed.) (2004). *Dropouts in America: Confronting the Graduation Rate Crisis*. Cambridge, MA: Harvard Education Press.
- Romo, H. D. and Falbo, T. (1996). *Latino High School Graduation: Defying the Odds*. Austin, TX: University of Texas Press.
- Rumberger, R. W. (1987). High school dropouts: A review of issues and evidence. *Review of Educational Research* **57**, 101–121.
- Rumberger, R. W. (2008). Solving California's dropout crisis. *Report of the California Dropout Research Project Policy Committee*. Santa Barbara, CA: California Dropout Research Project, University of California, Santa Barbara.
- Rumberger, R. W. and Palardy, G. J. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Educational Research Journal* **41**, 3–42.
- Wehlage, G. G., Rutter, R. A., Smith, G. A., Lesko, N., and Fernandez, R. R. (1989). *Reducing the Risk: Schools as Communities of Support*. New York, NY: Falmer.

Relevant Websites

- <http://www.edweek.org> – Editorial Projects in Education Research Center, Diplomas Count Archive.
- <http://cdrp.ucsb.edu> – University of California, Santa Barbara, California Dropout Research Project.
- <http://ies.ed.gov> – U.S. Department of Education, Institute for Educational Sciences, What Works Clearinghouse, Dropout Prevention.

Further Reading

- Alexander, K. L., Entwisle, D. R., and Kabbini, N. S. (2001). The dropout process in life course perspective: Early risk factors at home and school. *Teachers College Record* **103**, 760–882.

ECONOMICS OF EDUCATION – MARKETS, INCENTIVES AND SCHOOL CHOICE

Contents

Competition and Student Performance
Educational Privatization
Student Incentives
The Economic Role of the State in Education
The Economics of Catholic Schools
The Economics of Charter Schools
The Economics of Parental Choice
The Economics of School Accountability
The Efficacy of Educational Vouchers
Tiebout Sorting and Competition

Competition and Student Performance

A A Payne, McMaster University, Hamilton, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Can we improve public schools by promoting competition between them? In 1955, Milton Friedman published an article that questioned the role played by government in the provision of education (Friedman, 1955). Friedman promotes the idea that while government involvement is important, it reduces competition. Without this competition, schools have an incentive to waste resources resulting in the provision of a less than optimal quality of education. The underlying concept is as follows: Assume a single community provides education via a single school. If the government requires that all children under a certain age attend this school and there are no other options for schooling, the school has a group of students that have no choice in schooling. Any choice depends on the willingness of parents to move neighborhoods. If parents are reluctant to move, the school has an incentive to behave like a monopoly resulting in the wasting of resources.

This scenario is more complicated in most school systems. Over the last century, single school districts have consolidated into very large school districts resulting in many schools providing similar levels of education. Historically, these school districts have given schools defined catchments, lessening any competition between schools affiliated with the same district. It is this threat of a potential monopoly that has been the basis of many arguments that encourage more choice in schooling, be it through the subsidization of private schools (e.g., vouchers or tax

credits) or through the provision of more choice in public school (e.g., charter schools). What is less clear is whether schools exert monopoly power and, if they do, the extent to which this exertion diminishes school quality.

The purpose of this article is to discuss the key features of the research that has been conducted to understand better the question of whether schools engage in this monopoly behavior and its effect on student achievement. This article focuses on the research that has studied competition within the public school system with some reference to the effects from competition from private schools (e.g., Catholic schools) and/or charter schools. Much has been written on this topic. This article highlights the research that helps to explain why understanding the effects from competition is difficult to measure.

Concepts and Empirical Background

The notion that competition could promote better achievement starts first with a consideration of the choice of schooling by a parent for his/her child. A parent is expected to take into account things such as the quality of the school, the distance from the parent's home to the school, and the teaching philosophy of the school. If a parent can choose from a set of schools, this choice should encourage school administrators to use their resources efficiently to provide a quality level of education. The school

administrators, however, must be motivated to react to this potential competition. What motivates the school administrators can be a range of things such as receiving more resources upon meeting some level of performance or having resources tied to the school's enrolment. Card *et al.* (2008) expound on this theoretical framework to illustrate that areas in which parents have choice in schooling should result in better student performance in all schools relative to areas that have less choice in schooling.

To assess whether competition affects school quality, we can study school resources or student outcomes. Competition can impact the allocation of resources to their use by schools. If resources are used more efficiently in areas where schools face greater competition, then we should observe stronger student performance by students in these areas, after controlling for things such as student ability and characteristics of the neighborhood in which the school is located. The basic idea behind the analysis of competition would be to use data to run a regression that looks something like

$$Y = \alpha + \beta C + \gamma Z + \epsilon$$

where Y represents a measure of school resources (e.g., funding per student) or student outcomes (e.g., high-school dropout rate, test scores), C (or competition) is a measure that reflects the degree to which a school or set of schools face competition from other schools, Z is a set of measures to control for factors that could explain variation in the Y measure such as socioeconomic neighborhood characteristics, and ϵ is a residual term to capture other factors that affect the Y measure. This analysis may be done at a student, school, neighborhood, or school district level. Ideally, the analysis is done with data that are measured over time, allowing the researcher to control for both time-varying changes within each level of analysis (student, neighborhood, or district) and time-invariant characteristics of the area under study.

The purpose of the analysis is to focus on the coefficient β to identify the effect of competition on the decisions taken by the school or school district and/or on student performance. In the ideal study, students are randomly assigned to the school and one can compare the outcomes for areas with greater competitive pressures (e.g., greater school choice) with less competitive pressures. For example, if we were to study the potential effects of school competition on student outcomes, we would want to capture the outcomes of two groups of students. Both groups of students would have similar characteristics in terms of their parental background, socioeconomic characteristics, and their motivation for learning. The groups would differ based on the degree of competition faced by the schools attended by the students. Unfortunately, constructing this type of experiment can be difficult. Because parents can choose where to live and competition might be driven by the presence of

a private school, in any study that seeks to study whether competition matters one has to be concerned with parents' house location and/or schooling decision being made in a nonrandom fashion.

Sorting associated with household location is commonly referred to as Tiebout choice based on the work of Tiebout (1956). Under this notion, parents may choose the neighborhoods based on their preferences or tastes for education and other characteristics. If some parents have stronger preferences for education over other parents, this could result in a sorting that is positive or negative. If parents are willing to move to switch schools if a school is not providing a level of service expected by the parent, schools may seek the best use of their resources because they can attract new students and reduce the risk of losing existing students. A more efficient use of resources should result in a higher quality of education and, thus, better student performance. Sorting, however, could have negative effects. If the sorting is based on taste for education, some school districts might end up with students who are less interested in educational pursuits than other districts. The districts with less interested parents may end up with fewer resources which can result in a more limited ability of the school to deliver a quality education. In addition, if students perform differently based on their peers or other spillover effects, the overall effect from promoting competition through Tiebout sorting could result in students underperforming in the areas with students that are less interested in school.

What can be difficult to disentangle is whether observations on variations in school resources are attributable to the sorting/taste for education, due to the constraints a given school or school faces because of a lack of resources, or associated with being in a more or less competitive environment. Similarly, it is difficult to rely on casual observation to understand whether districts that are spending more resources are doing so because of a sorting issue or because they are inefficiently wasting resources as a result of monopoly power. Barrow and Rouse (2004) study whether school expenditures are valued by potential residents. Their overall finding is that school expenditures are valued by residents and there is no evidence of overspending on education by school districts. They conclude that increased competition across school districts may account for the lack of a finding of overspending by school districts.

This sorting issue can bias the coefficient on a measure of competition. The coefficient can also be biased for other reasons. These reasons include endogenous changes in the measure of competition or the exclusion of measures in the analysis that are correlated with both the competition measure and the outcome measure (omitted variable bias). As discussed below, the competition measure could be endogenous if a change associated with competition is linked to the demand for quality

education, for example, if a new school opens in response to an increase interest in schooling. This new school opening would increase the competitive environment. But if the new school opening is in an area where there is an increase interest in education, any observed gains in student outcomes could be attributable to either this change in interest in education or increase in competition. Disentangling the degree to which a change in performance is attributable to a change in competition is difficult. Thus, when estimating the effect of competition on school resources or on student achievement, one should consider carefully whether there are sorting or endogeneity issues or omitted variables in the estimation that could bias the measured effect of competition.

A common method used to address this problem is to identify a set of measures that directly affect the competition measure but not the outcome measure (Y). These measures are then used to predict the portion of the competition measure that can be viewed as exogenous to the outcome measure using an estimation technique known as instrumental variables or two-stage least squares. This allows the researcher to identify an unbiased effect from competition (β) on the outcome (Y). A second method is to use a policy change that affects the competition measure and measure the difference in outcomes before and after the policy change. A third method is to identify stark differences (discontinuities) in the rules that can affect the level of competition and then compare the effects of competition on the outcome around the area of the discontinuity. Examples of these methodologies are provided below.

Competition Associated with the Presence of Private Schools

Hoxby (2000) points out that it is important to understand the role played by Tiebout choice as any type of reform should consider the nature of sorting that occurs across neighbors given that parents are free to choose where to live. One approach taken by researchers is to study the performance of students in public schools relative to the fraction of students in private schools within a school district. The underlying basis for this approach is based on the notion that greater private school presence provides parents with alternatives, encouraging public schools to use their resources efficiently. An early study by Couch *et al.* (1993) and a study by Dee (1998) found a positive correlation on test scores and high-school graduation rates, respectively. Studies, however, by Newmark (1995), Sander (1999), and Geller *et al.* (2006) failed to find significant effects.

Demand for private schools, however, is correlated with the quality of the public schools in the area. If the quality of public schooling is low, parents may opt to send their

children to private schools. Yet, only those parents with the financial resources will be able to opt out of the public schooling system. Using measures of the demand for private schools (e.g., private school enrollment) as a proxy for the degree of competition one faces could suffer from a downward bias given the potential endogeneity between demand for private schools and public school quality.

Hoxby (1994) seeks to correct for this potential endogeneity bias by using the share of Catholics in an area as an instrument for private enrollment. If Catholic parents are more likely to send their children to a private school for reasons other than school quality (e.g., to provide their children with an education within a religious environment), then one should expect to observe more private (Catholic) schools in areas where there is a greater share of Catholic residents. Using the share of Catholics in an area, thus, provides a potential source of exogenous variation that can be used to identify the extent to which private schools provide a competitive threat to public schools. Hoxby's analysis suggests a positive effect on test outcomes. More recent studies, however, suggests a weaker and fairly small effects when this type of variation is used to predict the degree of potential competition (e.g., Arum, 1996; Jepsen, 2002, 2003). Altonji *et al.* (2005) also raise into question the use of measures that reflect the share of Catholics as instruments for competition. There are several extensions that use similar empirical methodologies that compare the effect of mechanisms such as vouchers for private schooling, publicly funded charter schools, and a comparison of performance of students in both public and private schools.

Competition Among Public Schools

Another way to explore the potential effects of competition is to study competition between public school districts within a geographic area. Given that parents can choose to where to live, if the size of the district is small enough, a parent potentially can choose to live in one of a number of school districts and these districts will compete with each other for that parent. Borland and Howsen (1992) develop a Herfindahl index of enrolment shares at different school districts to compare the performance of students based on the degree of market concentration of the school districts within a US county. Assume, for example, there are three school districts within a given area. The total enrolment across all districts is summed and shares of total enrollment for each school district is calculated. The Herfindahl index is the sum of the squares of these shares. The index ranges between 0 and 1. A higher index suggests less competition and a lower index suggests greater competition. If each of the three districts had approximately one-third of the students in

the area, the index would be approximately 0.33. If, however, one district had 80% of the students in the area and the other two districts had 15% and 5%, respectively, the value of the index would be 0.665.

Borland and Howsen (1992) find that third-grade students perform worse in counties with more concentrated school districts. Millimet and Rangaprasad (2007) extend this research by testing for strategic interactions across the input choices of nearby school districts in Illinois. Their approach assumes that a given district will react to the decisions taken by neighboring school districts if these districts compete for students. They find positive effects from nearby choices on things such as pupil/teacher ratios, spending per pupil, and average school size.

The concentration of districts within a county, however, is also arguably endogenous and could bias results downward to not finding a competitive effect in areas that have more school districts. Hoxby (2000) illustrates the importance of identifying exogenous measures that control for the potential endogeneity associated with the degree of market concentration within a county. Our observation of any given school district represents a combination of decisions made about the district with respect to the level and use of resources and decisions by parents to locate within the geographical boundaries of the school district. For example, an area that has many rivers or mountains will constrain the size of a school district. Hoxby (2000) exploits the natural contours of the land to identify the number of school districts in a given area. She relied on the measures related to the number streams located within a metropolitan area to identify the level of concentration of school boards within the metropolitan area.

At the time the paper was written, Hoxby did not have access to machine readable information about the location of streams. She relied on maps provided from the US Geological Survey's 1/24 000 quadrangle maps. From these maps she developed two measures of streams: a count of the larger streams and a count of the smaller streams. Her results suggest a strong association between the stream measures and the degree of concentration of school districts. Hoxby (2000) represents a creative way to address the endogeneity associated with some US counties having higher levels of market concentration relative to other US counties.

Rothstein (2007) comments on his replication of the Hoxby (2000) results. Rothstein (and subsequently addressed by Hoxby (2007)) identified some coding and software program errors that affected the assignment of students to school districts. Using machine readable data on streams, Rothstein created a set of instruments that differ from Hoxby's instruments but seek to capture the same types of exogenous variation in school district competition. His results fail to find a significant effect from competition on student performance.

Another way to study the potential effects from competition is to focus on the differences in student performance between students with choice and students without choice. Cullen *et al.* (2006) undertake such an approach. Similar to the discussion above, any study that compares different options of choice available to students must use a choice variable that is exogenous. An example of this exogenous variation is the lottery system undertaken by the Chicago public school system to allocate students across different public high schools. Students were eligible to attend a non-neighborhood high school if space was available. Because the space for these schools was limited, the school system assigned interested students based on a lottery. Cullen, Jacob, and Levitt compare various outcomes for the winners and losers of these lotteries. In effect, the winners get to choose to attend their home school or another school within the school district and so have greater choice in schooling than do the losers. Thus, if choice matters, we should expect to observe stronger student outcomes by the winners.

Cullen *et al.* (2006) fail to find much evidence that supports the argument that greater school choice results in stronger student performance. Their results suggest that the argument that competition through more choice results in better student performance may reflect only part of the story on how best to improve the delivery of education. If parents are not well informed about the role that schools can play in promoting better outcomes, this can result in a mismatch between students and the schools they attend. Cullen *et al.* (2006) provide limited support for this hypothesis.

Lessons from Research Using Non-US Data

Most of the current research in the US has focused on understanding the degree to which contiguously situated school boards compete with each other and whether this competition affects resource allocation and/or student performance. (There is also a literature that considers how school quality and changes that affect access to public schools across neighborhoods can affect housing prices. Reback (2005) studies a policy change that allowed students from nearby school catchment areas to attend a school in a different catchment area in Minnesota. He shows that house prices change as a result of this policy change. House prices rose in areas where there was a tendency for students to attend a school in a different catchment area and house prices fell in areas that attracted students from the neighboring catchment areas.) In other countries, however, there are institutional features that can add to our understanding of the potential effects from competition. Relying on Canadian data,

Card *et al.* (2008) exploit a feature of the public schooling system that provides choice in schooling to Catholic parents and their children. In Ontario and other provinces, a student of Catholic descent can choose between schools operated by two different types of publicly funded school boards: one operated solely for Catholic students (separate school boards) and one operated for all students (public school boards). Approximately 40% of the Ontario population is Catholic. Thus, for any given area within the province, there are two distinct publicly funded schools from which to choose if one is Catholic. In their paper, the authors first demonstrate that the two systems can be expected to compete with each other. Both systems are publicly funded, with the bulk of their funding being driven by enrolment figures. If there is a shift in enrolment from one system to the other system, the system losing enrolment suffers financially and the system gaining enrolment gains financially. Card *et al.* (2008) demonstrate that if one system opens a new school, the enrolments in neighboring schools affiliated the opposite system decline. These enrolment shifts are closely linked to the share of Catholics residing in the neighborhood.

Card *et al.* (2008) then explore the effects of these two systems on student performance by studying the performance of grade 6 students in a neighborhood, relative to the performance of the students in the neighborhood 3 years earlier (grade 3 test scores) and relative to the proportion of Catholics in the neighborhood. Their research suggests that students in neighborhoods with a high proportion of Catholics perform better than students in neighborhoods with a low proportion of Catholics. Their results are strongest in neighborhoods that are growing, as measured by the share of new housing in the area.

Gibbons *et al.* (2006) use a set of primary schools in England to study how competition may affect student achievement. They exploit discontinuities in school district boundaries to identify the competition effect. Their study finds limited gains in student performance at schools facing more competition than at schools facing less competition. Burgess *et al.* (2004) study England's choice-based education system to understand better the extent to which students sort across schools and whether this sorting affects student outcomes. After demonstrating that most students have different options for schooling, the authors demonstrate that rules that do not rely on the residential location of students affect the degree to which students are sorted across schools based on their ability.

Lavy (2005) explores a policy change in Tel Aviv to study the effects on student performance. The policy change introduced a program that allowed students to choose schools in and out of their assigned district. Lavy finds the policy affected high-school dropout and graduation rates, especially among disadvantaged children. He, however, fails to find any significant differences in the

achievement levels of the students that took advantage of the program.

In 1981, Chile introduced a program that provided vouchers to any student attending a private school. Similarly, in 1992 Sweden introduced a program that provided equivalent public funding to private schools. Both of these programs resulted in a dramatic increase in the number of private schools. Hsieh and Urquiola (2006) study the effects of the Chilean program. They compiled a panel data set that covered 150 municipalities to study the effects of the program on test scores and other student outcomes. The program resulted in a dramatic increase in private school enrolment ($\sim 10\%$). Sandstrom and Bergstrom (2005) study the Swedish reform. This reform dramatically increased the number of private schools (from 90 to 400), although the share of students attending private schools was still small ($\sim 4\%$). Sandstrom and Bergstrom study the effects of the reform on student performance using mostly cross-sectional data from a small sample of Swedish municipalities (33 of 288).

For both studies, the authors had to address issues concerning endogenous location of private schools and the potential of a nonrandom switch by students from public to private schools. They both seek to study the effect of increased private school enrolment on public school performance. In both studies, the authors use as instruments for private school enrolment measures that capture different characteristics of the municipalities in which the schools are located. Hsieh and Urquiola (2005) and Sandstrom and Bergstrom (2005) find, at best, a small positive effect of an increased private school enrolment on public school student performance. Hsieh and Urquiola (2005) push their analysis further and find evidence of a nonrandom switching by students from public to private school which illustrates the difficulty associated with finding a strong effect from increasing potential competition.

Summary

The research on whether schools compete with each other and how this affects student outcomes and school resources teaches us many things. First, there are many potential ways to measure potential competition. Competition could come from the ownership of schools (private, public), the structure of public schools within a school district (choice in selection of schools or assignment to a neighborhood school), and/or the structure and geographic boundaries of school districts. With each method, however, there are identification issues that could result in a downward bias of the effect of competition on performance and resources. Dealing with these identification issues requires creativity, quality data, and a clear understanding of the statistical issues.

The existing research suggests that there is limited evidence of the effect of competition on either performance or school resources. More recent research, most using data from outside of the US, is demonstrating a modest effect on measures of student performance from competition. There is no evidence at this stage, alas, that schools or school districts waste resources or results in substantially lower student achievement given a monopolistic position.

See also: Educational Privatization; The Economic Role of the State in Education; The Economics of Catholic Schools; The Economics of Charter Schools; The Economics of Parental Choice; The Efficacy of Educational Vouchers; Tiebout Sorting and Competition.

Bibliography

- Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). An evaluation of instrumental variable strategies for estimating the effects of catholic schools. *Journal of Human Resources* **40**, 791–821.
- Arum, R. (1996). Do private schools force public schools to compete? *American Sociological Review* **61**(1), 29–46.
- Barrow, L. and Rouse, C. (2004). Using market valuation to assess public school spending. *Journal of Public Economics* **88**(9–10), 1747–1769.
- Borland, M. V. and Howsen, R. M. (1992). Student academic achievement and the degree of market concentration in education. *Economics of Education Review* **11**(1), 31–39.
- Burgess, S., McConnell, B., Propper, C., and Wilson, D. (2004). Sorting and choice in English secondary schools. *CMPO Working Paper Series No. 04/111*.
- Card, D., Dooley, M., and Payne, A. A. (2008). School competition and efficiency with publicly funded catholic schools. *NBER Working Paper #14176*.
- Couch, J. F., Shughart, W. F. II, and Williams, A. L. (1993). Private school enrollment and public school performance. *Public Choice* **76**(4), 301–312.
- Cullen, J. B., Jacob, B. A., and Levitt, S. (2006). The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica* **74**(5), 1191–1230.
- Dee, T. (1998). Competition and the quality of public schools. *Economics of Education Review* **17**(4), 419–427.
- Friedman, M. (1955). The role of government in education. In Solo, R. A. (ed.) *Economics and the Public Interest*, pp 123–144. New Brunswick, NJ: Rutgers University Press.
- Geller, C. R., Sjoquist, D. L., and Walker, M. B. (2006). The effect of private school competition on public school performance in Georgia. *Public Finance Review* **34**, 4–32.
- Gibbons, S., Machin, S., and Silva, O. (2006). Competition, choice, and public achievement. Mimeo: Centre for the Economics of Education.
- Hoxby, C. M. (1994). Do private schools provide competition for public schools? *National Bureau of Economic Research Working Paper No. 4978*. Cambridge, MA: NBER.
- Hoxby, C. M. (2000). Does competition among public schools benefit students and taxpayers? *American Economic Review* **90**(5), 1209–1238.
- Hoxby, C. M. (2007). Does competition among public schools benefit students and taxpayers? Reply. *American Economic Review* **97**(5), 2038–2055.

- Hsieh, C.-T. and Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program. *Journal of Public Economics* **90**(8–9), 1477–1503.
- Jepsen, C. (2002). The role of aggregation in estimating the effects of private school competition on student achievement. *Journal of Urban Economics* **52**, 477–500.
- Jepsen, C. (2003). The effectiveness of catholic primary schooling. *Journal of Human Resources* **38**, 928–941.
- Lavy, V. (2005). From forced busing to free choice in public schools: Individual and general equilibrium effects. Mimeo: Hebrew University.
- Millimet, D. L. and Rangaprasad, V. (2007). Strategic competition amongst public schools. *Regional Science and Urban Economics* **37**(2), 199–219.
- Newmark, C. M. (1995). Another look at whether private schools influence public school quality. *Public Choice* **82**(3/4), 365–373.
- Reback, R. (2005). House prices and the provision of local public services: Capitalization under school choice programs. *Journal of Urban Economics* **57**(2), 275–301.
- Rothstein, J. (2006). Good principles or good peers: Parental valuation of school characteristics, Tiebout equilibrium, and the incentive effects of competition among jurisdictions. *American Economic Review* **96**(4), 1333–1350.
- Rothstein, J. (2007). Does competition among public schools benefit students and taxpayers? A comment on Hoxby (2000). *American Economic Review* **97**(5), 2026–2037.
- Sander, W. (1999). Private schools and public school achievement. *Journal of Human Resources* **34**, 697–709.
- Sandstrom, F. M. and Bergstrom, F. (2005). School vouchers in practice: Competition will not hurt you. *Journal of Public Economics* **89**(2–3), 351–380.

Further Reading

- Bayer, P., Ferreira, F., and McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy* **115**(4), 588–638.
- Belfield, C. R. and Levine, H. M. (2002). The effects of competition on educational outcomes: A review of US evidence. *Review of Educational Research* **72**(2), 279–341.
- Besley, T. and Ghatak, M. (2005). Competition and incentives with motivated agents. *American Economic Review* **95**(3), 616–636.
- Bifulco, R. and Ladd, H. F. (2006). The impacts of charter schools on student achievement: Evidence from North Carolina. *Education Finance and Policy* **1**(1), 50–90.
- Brunner, E. J. and Imazeki, J. (2008). Tiebout choice and universal school vouchers. *Journal of Urban Economics* **63**(1), 253–279.
- Burgess, S. and Slater, H. (2006). Using boundary changes to estimate the impact of school competition on test scores.
- Clapp, J. M., Nanda, A., and Ross, S. T. (2007). Which school attributes matter? The influence of school district performance and demographic composition on property values. *Journal of Urban Economics* **63**(2), 451–466.
- Cullen, J. B., Jacob, B. A., and Levitt, S. D. (2005). The impact of school choice on student outcomes: An analysis of the Chicago public schools. *Journal of Public Economics* **89**(5–6), 729–760.
- Hanushek, E. A., Kain, J. F., Rivkin, S. G., and Branch, G. F. (2007). Charter school quality and parental decision making with school choice. *Journal of Public Economics* **91**(5–6), 823–848.
- Jacob, B. A. and Lefgren, L. (2007). What do parents value in education? An empirical investigation of parents' revealed preferences for teachers. *Quarterly Journal of Economics* **122**(4), 1603–1637.
- Lankford, H. and Wyckoff, J. (1992). Primary and secondary school choice among public and religious alternatives. *Economics of Education Review* **11**(4), 317–337.
- Sass, T. R. (2006). Charter schools and student achievement in Florida. *Education Finance and Policy* **1**(1), 91–122.

Educational Privatization

C Belfield, Queens College, City University of New York, Flushing, NY, USA

H M Levin, Teachers College, Columbia University, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Educational privatization is a term that has become widely used in the early part of the twentieth century. To those unfamiliar with the term, it may connote a shift to private schools. However, the term has come into wider usage largely because of a variety of forms of increased involvement of nongovernment entities in the sponsorship, financing, and provision of education. For example, one dimension of this trend is that of charter schools in the US, where some 4000 public schools have been given considerable autonomy in operation and management, including maintaining their own governing boards. However, there are many other examples of privatization including government schools operated under contract by private entities which include for-profit firms and universities or private schools that receive government funding. At the same time, many countries require private fees as well as other private expenditures by parents to attend public schools. All share in common some dimension of privatization in contrast to a purely public concept of education where schooling is provided and funded exclusively by public authorities.

In this article we delineate a range of different forms of privatization and set out criteria for evaluating their consequences. We begin by referring to four different forms of schooling: home schooling, private schools, charter schools, and public schools.

Privatization of Different School Types

Home schooling refers to the education that parents provide in the home, both informally and as an official form of schooling (Belfield, 2008). In some countries, such as the United States, home schooling that meets certain standards can satisfy compulsory attendance laws of the states, even when the children are not attending formal schooling institutions for any part of their education. In fact, in the US, about 2% of students are home schooled. In many respects, home schooling is the most privatized form of education with the least public intervention. The public involvement extends only to requirements that parents meet broad standards for curriculum, hours of instruction, and, in some cases, testing, but accountability for these standards is often lax.

Private schools are the most common form of educational privatization, schools that are founded and sponsored by private entities, including religious groups, independent organizations, and for-profit organizations. Such schools may be independent or operated by organizations with an educational or spiritual mission or operated as a business. In the US, private schools are mostly religiously affiliated and not for profit. In other parts of the world, particularly in Asia and Latin America, private schools were started by missionaries, but the major growth has been that of for-profit businesses. Private schools were founded in most countries long before the advent of universal and government-provided schooling, but only to serve particular religious or social groups.

In some countries, private schools are funded strictly by parental fees and philanthropy. In other countries, they are fully or partially funded by government funds, even though they are operated privately. Two of the best examples are Holland and Chile. In Holland, a school that meets specified standards in personnel, curriculum, class size, and other requirements can receive the same amount of funding whether it is a private or government-sponsored school. In Chile, the national voucher system also provides similar funding for religious, for-profit, or government public schools. In Europe, in general, it is also common to provide substantial government funding for private schools (McEwan and Carnoy, 2000).

Charter schools are a unique development in the US that mixes both public and private elements. They are public in receiving public funding and being required to meet some public requirements, but private in terms of governance and sponsorship and in being provided autonomy and release from many of the state and local laws (Bifulco and Bulkley, 2008). Charter schools, started in 1992, have proliferated among nearly 40 states to where there are about 4000 of them serving more than 1 million students in the elementary and secondary grades. Charter schools are expected to augment their public support by obtaining private funding through the philanthropic community, and a significant segment obtains a substantial amount of support through this source.

Public or government schools represent the most common form of schooling around the world. However, even these schools have many private components. For example, in many countries, public schools draw upon students in particular residential areas where students from wealthier families are segregated in their own government

schools and those from poorer families are segregated in schools with high concentrations of poor children. In other cases, parents must pay fees or provide other private costs for their children to attend public schools. In this respect, there is privatization of at least part of the costs and segregation of students that enables the schools to provide a different type of education as the public schools attended by students from good families are able to attract better teachers and other resources.

What can be demonstrated is that pristine categories of public versus private tend to oversimplify the educational privatization phenomenon. Most educational systems support schools that have both components (Belfield and Levin, 2005). Further, to the degree that all schools must produce not only individual learning and skills, but also preparation for citizenship in a society, even the most privatized schools in terms of finance or sponsorship are contributing to public goals.

Privatization of School Services

However, privatization can also take other forms beyond differences among major school types. For example, both public and private schools can produce their own services or make arrangements for those services to be provided by private suppliers. In many countries, construction, furnishings, text books, food services, and transportation are provided by outside contractors to public schools. In general, the view is that competitive markets for these types of school resources can provide greater choice and variety and lower costs than if governments produce these services. Of course, these decisions also have employment consequences as well which make them politically contentious where public employee unions are strong.

Nevertheless, even more controversial has been the shift to private educational management organizations (EMOs) to operate public schools (Miron, 2008). EMOs are firms that have arisen to contract with public school authorities to manage their schools. In general, there has been relatively little controversy over whether government or private entities should provide material goods or services to public schools. However, the expansion of such efforts to instructional services has brought great controversy, at least in the US where there has been the greatest development of this approach. In general, it takes the form of public entities such as school districts contracting out the management of schools to firms that hope to make a profit from such endeavors. These enterprises are given a set amount of revenue for each student and are required to provide educational services at least comparable to those of the government-managed schools.

School districts generally have contracted with EMOs for their poorest-performing schools with the hope that the private-sector effort can solve low-achievement problems

that have not been solved by government management. Typically, these schools enrol students from economically disadvantaged, minority, and immigrant families. The EMOs promise to outperform comparable public schools by hiring better teachers and administrators and using more efficient business practices. However, two decades of experience in the US have neither shown that EMOs can outperform public schools or be consistently profitable.

A similar story can be said about contracting for particular instructional services such as the tutoring of students who lag in academic achievement. In the US, under the No Child Left Behind legislation of the federal government, states are required to provide funding for students in failing public schools to receive tutoring and other instructional services from outside vendors. The rationale for opening these options to private vendors is that the public authorities have failed these students and should not be entrusted with additional funding for remedial services when the private sector has the capacity to outperform the public schools. However, after 7 years of experience there is virtually no evidence of private contractors providing better academic outcomes.

Financing Education and Privatization

Privatization of education is supported by many different financial approaches. The most general of these is that of educational vouchers.

Educational vouchers refer to a school financing approach where families are provided with a certificate or voucher that can be used to pay a certain level of tuition at any approved school, public or private (Levin, 2002). Schools compete for students and their vouchers and redeem them with the state to finance operations. Chile has had a national voucher scheme since 1980, and much has been written about it and many evaluations have been done with respect to performance (McEwan and Carnoy, 2000; Zimmer and Bettinger, 2008: 456–457). One of the major claims for vouchers is that by introducing competition into the schooling system in place of a government monopoly, schools will become more productive and efficient in order to attract enrolments. Thus far, there is no evidence of rising efficiency for the Chilean educational system, although it appears that the private schools receiving vouchers do slightly better in raising achievement than public schools, at least for middle-class students. However, Chile also permits schools to establish additional private family payments beyond the government voucher, and these higher costs may account for any higher achievement.

In the US, there is also a strong push for increasing educational privatization with educational vouchers. Both free-market advocates and those who simply see market competition as a stimulus to raising public school performance have supported vouchers in the US. Milwaukee has

had a voucher system since 1990 and Cleveland since 1995. Washington, DC, has had a voucher system since 2006, and Louisiana has just adopted one. In all of these cases, the number of students funded by vouchers is limited, and eligibility or preference is given to low-income families. In general, evaluations show that the families who seek and obtain the vouchers indicate a higher level of satisfaction with their schools of choice than comparable families in public schools, but there is little evidence that achievement of students using vouchers is higher (Zimmer and Bettinger, 2008).

A special version of vouchers has also been adopted by such states as Arizona, Florida, and Utah. In the US, there are laws that require public schools to provide appropriate educational services for students with disabilities in the least-restrictive environment. Often, the small student-teacher ratios and facilities requirements for such students necessitate high funding costs by school districts, sometimes 2–3 times the average pupil expenditure for nondisabled students. If a school district is unable to provide satisfactory services in the view of the parents, the parents can seek private school services at public expense through a hearing procedure. However, the more recent use of vouchers for students with disabilities makes them available to fund private student placement directly by giving the parent the option of accepting the public school's individual educational plan for their child or getting an equivalent amount of funding to send the child to a private school.

Tuition tax credits refer to a different mechanism for covering private costs of education (Huerta and d'Entremont, 2007; Zimmer and Bettinger, 2008). At least seven states in the US sponsor tuition tax credit plans. Under this approach, a portion of family tax liability is forgiven for paying tuition at a private school, or in some states, for other private educational expenses. For example, a family might be given a tax credit of \$1000 in reduced state tax liability for paying at least that amount toward tuition at a private school. Libertarians and many advocates of market competition prefer tuition tax credits to educational vouchers because they require less government involvement in the operations of schools. Educational vouchers require many detailed regulations in terms of what kinds of schools are eligible and what information they must report annually and other operating details to be able to redeem vouchers. With tuition tax credits it is only necessary for a parent to gain the tax benefit by providing a receipt of payment from a private school or from some other private educational service.

There are several variants of the tuition tax credit for parents. Several states allow nonparents and businesses to contribute to funds that will provide scholarships for students to attend private schools. Since the willingness to take a tax credit simply reduces one's tax burden, such a decision is virtually costless for the taxpayer, although tax

credits obviously have a cost to the state in terms of foregone tax revenues. Another variant is that of a tuition tax deduction. The tax deduction simply allows the taxpayer to reduce taxable income and save the tax on that reduction, a much smaller tax benefit than a tuition tax credit.

Private costs in public schools represent a prevalent form of privatization in the sense that families must contribute substantial resources to participate and ensure success for their offspring in the educational system. In some countries, such as Tanzania, the fees to attend public secondary schools are so substantial that they are an important reason for one of the lowest secondary school participation rates in East Africa. Many countries allow government schools to charge fees to students. In Bangladesh, India, and Pakistan, it is common for teachers to require students to pay for private tutoring sessions for curriculum topics that are deliberately not covered in class, but required for passing the examinations. In addition, many schools in the developing world require families to pay for textbooks, study guides, uniforms, writing supplies, transportation, and pay other special fees that are assessed during the year. The shifting of school finance from public sources to private ones is a form of privatization with the consequences that it may limit school enrolments or shift enrolments of the poor who cannot afford the private costs into schools with more limited learning resources.

Philanthropic resources are also important funding sources for privatizing education. Normally, these are provided by organizations or movements that are committed to a particular mission or population. Religious organizations, in particular, may subsidize the costs of private schools of their denomination in order to ensure that the costs to families are not a barrier to attendance. A different motive may be evident when philanthropic resources are applied to schools with relatively high-performing students or those from families of high status. By adding resources they can assure salaries and facilities that will attract the best teachers, increasing the appeal of the school and allowing the school to discriminate in admissions.

Evaluating Educational Privatization

Many advocates of educational privatization support it because they believe that it may be more efficient than public forms in promoting competition and choice. Opponents often argue against educational privatization because it leads to segregation by family income, race, ethnicity, and other distinctions that create inequities and inequalities. Although these are broad-stroke assumptions, the different forms of privatization must be evaluated according to their specific consequences, and different forms of privatization have different consequences. For example, in the case of educational vouchers, a high voucher (relative

to existing school costs) that is given to all students with supplements for special need students, such as students from immigrant and impoverished families or with disabilities, will tend to promote more equal outcomes than if the voucher is set at a low level with parental contributions supplementing it. In the latter case, wealthier families will be able to spend more on the education of their children, sequestering them in schools with other children from high-income families and better salaries and working conditions that will attract the best teachers away from the poor.

Evaluating Educational Privatization Approaches

Educational privatization is not just a single phenomenon, but comprises many different phenomena. Privatization can occur in the sponsorship, governance, funding, contracting, and many other aspects of schooling. Even what we call public schooling often has many dimensions of privatization. These conclusions suggest that being precise about what is meant by educational privatization is important in evaluating it.

Clearly, there is no perfect system as much as a quest for a best system in providing a balance among competing aims. In this context, one can denote four major criteria, which are detailed below, for addressing an effective educational system (Levin, 2002):

1. *Freedom to choose.* This criterion places a heavy emphasis on the private benefits of education and the liberty to ensure that schools are chosen by parents who are consistent with the child-rearing practices of families. Just as families wish to set the types of conditions that will influence their children's growth and development overall, they wish to choose schools that reinforce those values. Privatization advocates typically place great weight on this criterion.
2. *Productive efficiency.* This criterion refers to the maximization of educational results for any given resource constraint. Educational privatization advocates assume that competition among schools for students and the greater use of parental resources will create strong incentives for schools to meet student needs and to improve educational productivity. Privatization opponents believe that the assumptions that make competition effective will not be present in the educational marketplace.
3. *Equity.* This criterion refers to the quest for fairness in access to educational opportunities, resources, and outcomes by gender, social class, race, language origins, disability, and geographical location of students. Educational privatization advocates argue that many forms of privatization can be used to expand school choice and open up possibilities for students who are locked into inferior neighborhood schools and that a

competitive marketplace will have great incentives to meet the needs of all students more fully than existing schools. Challengers argue that educational privatization, and particular vouchers and tuition tax credits and increasing private costs, will create greater inequities because parents with education and income are better informed and have greater resources to cover school costs and to have access to transportation. In addition, they believe that the choices, themselves, will further segregate the poor and disenfranchised as those with power and status will select schools with students like themselves and schools will also select students by such criteria.

4. *Social cohesion.* This criterion refers to the provision of a common educational experience that will orient all students to grow to adulthood as full participants in the social, political, and economic institutions of our society. This is usually interpreted as necessitating common elements of schooling with regard to curriculum, social values, goals, language, and political institutions. Privatization advocates believe that social cohesion will develop in schools without making special provisions or that it will only require minimal regulations, such as common curriculum requirements. Opponents argue that it will take place only if schools adopt a common set of objectives and educational experiences through societal agreement and mandated policies.

More fully meeting some criteria may reduce the ability to meet others, so that any system of educational privatization must acknowledge trade-offs. Some goals cannot be attained without sacrificing others. Although some design provisions would improve outcomes along more than one criterion, almost all would also reduce outcomes on other criteria. Provision of information and transportation will improve choice options for all participants, especially for those from families with the least access to information and transportation, the poor. However, such provision would also raise the costs of the overall educational system and decrease resources available for instruction, probably reducing productive efficiency unless academic gains from competition due to better information and access offset the costs of the transportation and information. As another example, the establishment of extensive regulations on curriculum, admissions, and finance with continuous monitoring and enforcement could be used to increase equity and social cohesion, but only by creating more uniformity among schools which will undermine freedom of choice.

This means that there is no optimal system that provides maximal results among all four criteria. Ultimately, the choice of design features will depend upon specific preferences and values as transmitted through democratic institutions. Those who place a high value on freedom of choice will probably be willing to sacrifice some equity and social cohesion provisions by minimizing regulations,

stressing school autonomy, and giving parents the ability to increase private spending on the schooling of their offspring through permitting fees in government schools or private supplements to government vouchers. Conversely, those who place a high value on social cohesion will be willing to sacrifice some freedom of choice through establishing a common curriculum core and other standardized features of schools. Ultimately, much of the debate over the specifics of educational privatization plans revolves around the political power and preferences of the stakeholders.

Tools of Educational Privatization

How a society uses educational privatization will depend on its priorities in balancing the four criteria that were set out above. There are three tools for addressing these criteria, finance, regulation, and support services, which are described below:

1. *Finance.* How an educational system is financed will have strong implications for such goals as choice, efficiency, and cohesion. For example, if funding is kept low, whether for use at public or private schools, the range of choices will be limited for those who must depend upon public funding. As one finds in many countries of the developing world where government schools are funded minimally, middle- and upper-income families circumvent the government schools by using their own funds to send their children to private schools that charge additional tuition which the poor cannot afford. This means that the educational system will foster segregation of the poor from those who are financially more privileged. In contrast, if more funding is provided overall to make government schools more attractive to the middle and upper-middle class and additional funding is given for services for students from poverty backgrounds and those with disabilities, it is possible to provide a more equitable system of education with less stratification.
2. *Regulation.* Governments can use regulation of admissions, curriculum, and testing requirements as well as school performance criteria to promote educational goals. By requiring schools to accept students by lottery, establishing a common curriculum, and ensuring that school performance is measured through standard testing requirements, both an emphasis on equity and social cohesion can be sought. The use of a common rubric for measuring results may also increase efficiency of competition. However, regulation produces greater uniformity which reduces the range of schooling options and freedom of choice.
3. *Support services.* Governments can provide transportation, information, on school alternatives, and instructional support such as tutoring to expand educational choice

and equity. These services are especially likely to help those with the greatest economic disadvantage by providing access to more potential schools, information on school performance, and instructional assistance beyond the school. However, they may also reduce the instructional resources available to schools for providing educational services.

Summary

Educational privatization is not a singular phenomenon, but one that can take many different forms. Thus, simple generalization about the phenomenon is not useful. Different countries utilize educational privatization in different ways with enormously differential consequences with respect to the impact on choice, efficiency, equity, and social cohesion. Even in the same country there may be many different forms of educational privatization. The specific impacts of any form of educational privatization can be evaluated according to their consequences for different educational outcomes. Unfortunately, the goals, themselves, are often in conflict so that designing an approach that uses privatization may enhance one outcome at the expense of sacrificing others. In particular, the careful design of educational finance, regulation, and support services can be used as policy tools to obtain the mix of outcomes that is most desirable.

See also: Competition and Student Performance; The Economics of Catholic Schools; The Economics of Charter Schools; Tiebout Sorting and Competition.

Bibliography

- Belfield, C. R. (2008). Home schooling. In Ladd, H. F. and Fiske, E. B. (eds.) *Handbook of Research in Education Finance and Policy*, pp 467–474. Routledge: New York.
- Belfield, C. R. and Levin, H. M. (2005). *Privatizing Educational Choice: Consequences for Parents, Schools, and Public Policy*. Boulder, CO: Paradigm Publishers.
- Bifulco, R. and Bulkley, K. (2008). Charter schools. In Ladd, H. F. and Fiske, E. B. (eds.) *Handbook of Research in Education Finance and Policy*, pp 445–446. Routledge: New York.
- Huerta, L. A. and d'Entremont, C. (2007). Education tax credits in a post-Zelman era. *Educational Policy* **21**, 73–109.
- Levin, H. M. (2002). A comprehensive framework for evaluating educational vouchers. *Educational Evaluation and Policy Analysis* **24**, 159–174.
- McEwan, P. J. and Carnoy, M. (2000). The effectiveness and efficiency of private schools in Chile's voucher system. *Educational Evaluation and Policy Analysis* **22**, 213–239.
- Miron, G. (2008). Educational management organizations. In Ladd, H. F. and Fiske, E. B. (eds.) *Handbook of Research in Education Finance and Policy*, pp 475–494. Routledge: New York.
- Zimmer, R. and Bettinger, E. P. (2008). Beyond the rhetoric: Surveying the evidence on vouchers and tax credits. In Ladd, H. F. and Fiske, E. B. (eds.) *Handbook of Research in Education Finance and Policy*, pp 447–466. Routledge: New York.

Student Incentives

E R Eide and M H Showalter, Brigham Young University, Provo, UT, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

A fundamental principle of economics is that people respond to incentives. They change behavior in ways that promote their well-being when faced with new circumstances. Many education reform policies rely on incentives to change student, teacher, and principal behavior in ways that increase their effort and improve student outcomes. For example, some school districts in the United States have implemented high-stakes testing that determines which students are promoted to the next grade or allowed to graduate high school. In principle, students faced with the possibility of a disruption in their educational progress will work harder in order to meet the required standards.

A strand of research has recently emerged that assesses the ways, both intended and unintended, in which incentives have changed student achievement and educational progress. This article provides an overview of this research agenda. It includes a discussion of how education economists conceptualize and estimate the association between incentives and measures of student learning and educational progress, and it summarizes the research findings on this relationship.

Conceptual and Empirical Issues

Economists use a conceptual framework called an education production function to model the educational process that leads to improved student outcomes. In these models, a student's educational outputs, such as test scores, are a function of the quantity and quality of a variety of school and teacher inputs. Common educational input measures include class size, per-pupil expenditure, teacher education, and teacher experience. Education policy variables specific to a particular research objective, such as incentives to raise student performance, are also included as inputs.

Education production functions usually include non-school inputs that are theorized to affect student outcomes. Such measures include the students' family background (parents education and income), demographics (race/ethnicity, gender), and peers. Including nonschool inputs together with school inputs provides a deeper understanding of the factors that influence student outcomes.

While conceptually straightforward, statistically evaluating education policies, such as the effectiveness of incentives, is fraught with challenges. There are often multiple ongoing policies at the national, state, and local

level that simultaneously affect the educational process, making it difficult to assess the effectiveness of any one particular policy. Changes over time in the types of students who are being measured, for example, the demographic composition or the percentage of students who are classified under special education, pose another obstacle. Furthermore, if one of the research goals is to assess whether improved student outcomes are a result of increased student effort, a difficulty arises because student effort is not usually observed or measured. In most cases, researchers must take an indirect approach such as measuring the difference in student achievement before and after a policy is implemented.

Strategies for overcoming these estimation issues include the use of longitudinal data that follow students and school districts over times to cover both pre- and post-policy implementation, as well as the use of sophisticated statistical models. While these fixes are not perfect, they go a long way toward improving the validity of the estimates.

Incentives to Increase Student Effort and Performance

Numerous education policies incorporate incentives to raise student learning. For exposition purposes, here policies are grouped into three areas: standards-based accountability, direct monetary rewards to students, and incentives to teachers.

Standards-Based Accountability

The prominence of standards-based accountability in the education reform arena began in the United States in 2002 when President Bush signed into law the No Child Left Behind Act (NCLB). Standards-based accountability provides incentives to students, parents, teachers, and principals to improve student performance. Schools and school districts must develop curriculum and achievement standards, and measure student learning through tests that are aligned with the standards (Betts and Costrell, 2001). A key feature of standards-based accountability is that consequences are associated with how well students and schools progress toward their learning objectives. For students, consequences include high-stakes tests that determine whether or not they graduate from high school or are promoted to the next grade. For schools, consequences for failing to make adequate progress can include dismissal or reassignment of teachers and school administrators.

In principle, these incentives provide the motivation for students and schools to change behavior and improve. Betts and Costrell (2001) explore the possible incentive effects of high-stakes testing on student performance, and how student response to the test cutoff depends on where the student is in the distribution of ability and attitudes toward study. For example, for students who are just under the margin of passing, it would only take a small increase in their effort in order to raise performance enough to pass the test. Hence, in comparing the cost (i.e., effort) and the benefit (i.e., passing the test) of trying harder, students at the margin of passing would respond to the incentive and work harder to pass the test. Students who are far below or far above the cutoff have little incentive to change their effort.

While advocates of standards-based accountability believe these incentives will yield benefits to students and schools, critics argue that the incentives result in strategic behavior by teachers in ways that bring about unwanted changes. For example, teachers may shift resources away from low-stakes subjects toward high-stakes ones that will be explicitly tested, and they may teach to the test. There may also be a discouragement effect as students who fail high-stakes exams, such as high school graduation exams, choose to drop out of school rather than retake the exam in the future in an effort to pass (Martorell, 2004).

In addition to high-stakes testing, some schools elicit greater student effort and performance by raising grading standards. While the consequences of higher grading standards may not seem as direct or severe as high-stakes testing, the consequences can be considerable because grade point average is a main determinant of college admission and scholarships. Economic models of grading standards (Becker and Rosen, 1992; Betts, 1998) predict that, as with high-stakes testing, increasing grading standards may have uncertain effects on student achievement as more academically proficient students increase effort in order to reach the new standard, while less-proficient students, who must work harder in order to meet the higher grading standard, give up instead. Students at the extremes of the achievement distribution may be unaffected by increased grading standards.

Evidence

Studies examining the relation between standards-based accountability and student performance yield mixed conclusions. Results differ according to the city or state studied, and the type of policy that was implemented. For example, some studies relating high school exit exams to student achievement find a positive association, while studies with more extensive data do not find higher achievement (Jacob, 2001). Martorell (2004) used data from Texas to examine how high school exit exams affect a number of student

outcomes. He found that failing the exit exam reduces the probability of normal high school graduation and increases the likelihood of acquiring a general educational development (GED) degree. Martorell also found that failing the high school exit exam lowers earnings, but only among individuals whose work experience indicates a strong attachment to the Texas labor market.

Jacob (2005) studied the effectiveness of an accountability policy implemented in the Chicago public schools in 1996–97. The accountability policy consisted of two parts. One part was holding students accountable by ending social promotion, where students are advanced to the next grade regardless of their academic achievement level, and in its place instituting grade retention, where students repeat a grade if they are not achieving at grade level. Students were required to pass standardized tests in math and reading, or else be required to take a 6-week summer school program. They were tested at the end of summer school, and if they still failed the test, they would not be permitted to advance to the next grade.

The second part of the accountability policy was designed to hold teachers and schools accountable for student achievement. Schools in which fewer than 15% of students scored at or above national norms on the standardized reading exam were placed on probation. If they did not show adequate improvement, these schools could be reconstituted, which involved the dismissal or reassignment of teachers and school administrators.

Using longitudinal student-level and administrative data, the study found that math and reading achievement increased sharply following the introduction of the accountability policy, in comparison to both prior achievement trends in the Chicago district and to changes experienced by other large, urban school districts in the mid-Western United States. The results suggest that the observed achievement gains were driven by increases in test-specific skills and student effort. Moreover, the study found that teachers responded strategically to the incentives along a variety of dimensions – by increasing special education placements, preemptively retaining students, and substituting away from low-stakes subjects such as science and social studies.

A related study by Jacob and Lefgren (2004) examined the effect of the Chicago accountability policy on summer school and grade retention using a sophisticated statistical technique called regression discontinuity. The intuition behind regression discontinuity in this paper is that students who are just under the test cutoff must attend summer school, and those who are just past the cutoff do not have to attend, and these two groups of students who are just on either side of the cutoff are likely similar in their individual characteristics and background. Provided there is some randomness in the test outcomes, differences in future performance between these two groups of students can be attributed to the accountability policy.

Using this statistical technique, the study found that the Chicago accountability program substantially increased academic achievement among third graders, but not among sixth graders.

Other studies have examined the effects of accountability systems on student achievement using cross-state analyses, and have found positive effects. Research by Hanushek and Raymond (2005) and Carnoy and Loeb (2002) show that the introduction of accountability systems led to significantly higher student-achievement gains as measured by the National Assessment of Educational Progress (NAEP).

Research on the effects of increasing grading standards on student outcomes finds that students do indeed respond to higher grading standards. Betts and Grogger (2003) find that higher standards raise test scores across the distribution of achievement (low-performing, average, and high-performing students all improve achievement), but the greatest increase is found among students at the top of the distribution. They also find, however, that higher standards have no positive effect on educational attainment, and in fact have negative effects on high school graduation among some racial/ethnic groups.

Overall, research suggests that standards-based accountability can raise student effort and performance, depending on the context and type of accountability policy implemented. Moreover, students and teachers sometimes respond in unintended ways when faced with new incentive structures.

Student Financial Incentives

Monetary rewards to students for good academic performance are another incentive used to raise student effort and achievement. Financial incentives typically include scholarships to pursue higher education, and in less-frequent cases, direct monetary rewards to use for education purposes during primary grades. Critics of financial merit awards posit that rewarding students for educational performance may weaken the intrinsic motivation for learning, and is subject to gaming.

Evidence

In recent years, state governments have become increasingly involved in providing merit financial aid, allocating billions of dollars in merit-based college scholarships. States have undertaken these efforts in an attempt to promote academic achievement, increase college enrolments, and keep the most able high school graduates in state for college (Cornwell *et al.*, 2006).

One of the most influential of these state programs is Georgia's Helping Outstanding Pupils Educationally (HOPE) Scholarship, which began in 1993 and is funded by the state lottery. The scholarship covers major expenses for all eligible high school graduates attending Georgia postsecondary institutions. Eligibility includes graduating

from a Georgia high school since 1993 with at least a B-average, and being a resident of Georgia. Analysis of the HOPE scholarship program shows that it is effective in raising college enrolments. Cornwell *et al.* (2006) find that total college enrolment was about 6% higher in Georgia than other comparable southern states as a result of the HOPE scholarship program, which amounts to an increase of about 2889 freshmen per year increase in Georgia colleges. These figures suggest that more students are meeting the eligibility requirements in order to receive the scholarship.

Another state-based merit award for higher education is the Michigan Merit Award and the Michigan Promise Scholarship. The Michigan Merit Award was established by the Michigan legislature in 1999 for the high school classes of 2000 through 2006. The award provides up to US\$3000 to high school graduates who meet or exceed standards on reading, math, science, and writing tests. The Michigan Promise Scholarship, which was signed into law in 2006, provides up to US\$4000 to high school graduates for successfully completing 2 years of post-secondary education beginning with the high school graduating class of 2007. Analysis of the Merit Award program suggests that it does increase student effort and learning (Bishop, 2001). Not only did the number of students taking the exam go up after the introduction of the merit award, but also the number of students demonstrating that they met or exceeded Michigan's education goals rose dramatically.

The studies discussed so far have focused on the relation between monetary rewards and student outcomes in the US. There is also international evidence that students respond to monetary incentives. Kremer *et al.* (2007) report results from an evaluation of a merit scholarship program in Kenya in which girls who scored well on academic exams at the end of the sixth grade received scholarships to pay for their school fees and cash grants to cover the cost of school supplies for the next 2 years. The authors find that girls who are eligible for the scholarship show substantial gains in academic exam scores. There is also evidence of additional indirect benefits as girls with low pretest scores, who were unlikely to receive an award, showed test-score gains that appear to be comparable to the test-score gains for higher-scoring girls who were more likely to receive an award.

In sum, the evidence on financial incentives to students suggests that, as expected, students respond positively to financial merit rewards. Students perform at a higher level academically, and more students attend postsecondary education.

Teacher Financial Incentives

A relatively recent approach to raising student achievement is to offer financial awards to teachers whose students demonstrate improved academic performance.

The prospect of receiving a financial award leads teachers to change their effort and performance level, and to motivate their students to do the same.

Two of the most common teacher-incentive policies are individual teacher merit pay and school-based group bonuses. Merit pay is provided to individual teachers when their students demonstrate improved academic performance. Most merit-pay plans are a combination of time-based pay and output-based pay. Teachers continue to earn a base annual salary, but may receive special pay increases or bonuses that are dependent on the output of their students, typically measured by improved performance on standardized tests. A variation of individual teacher-merit pay is school-based rewards. School-based plans are similar to merit-pay plans in the sense that rewards are based on student performance. The difference is that the rewards in school-based plans are awarded on a school-wide basis and shared by all teachers in the school, rather than being awarded to individual teachers.

Evidence

Podgursky and Springer (2007) review the literature on the relation between teacher-incentive plans and student achievement. In most of the reviewed studies, the teacher-incentive policy yields positive student-achievement results.

Figlio and Kenny (2007) combine nationally representative data on schools, students, and their families with their own survey on the use of teacher incentives to document the relationship between individual teacher-performance incentives and student achievement. Their survey on teacher incentives has data on frequency and magnitude of merit raises and bonuses, teacher evaluation, and teacher termination. They find that test scores are higher in schools that offer individual financial incentives for good performance. Moreover, the estimated relationship between the presence of merit pay in teacher compensation and student test scores is strongest in schools that may have the least parental oversight.

Lavy (2002) studies the relation between school-based performance incentives and student performance. This article examines the effect of an Israeli program that provided monetary incentives to schools and to teachers as a group according to the relative improvement in their students' academic performance. The program offered incentives to schools in the form of performance awards, part of which were distributed to teachers and school staff as merit pay and the rest used for the well-being of teachers in the school. The competition was based on various achievement criteria, with the awards going to the top third of performers. The study finds that schools' and teachers' group monetary incentives caused significant gains in many dimensions of students' outcomes. Moreover, providing schools with more resources also led to improvement in student performance.

Muralidharan and Sundararaman (2006) report results from an experiment on performance pay in rural Indian schools. This study included both individual teacher-merit pay bonuses and school-wide group bonuses, both of which were linked to gains in student test scores. The authors found significant gains in both math and languages in response to the incentive programs, and found no significant difference in the effects of the incentive programs for individual teachers and the school-wide group bonuses.

The literature on teacher-incentive plans suggests that providing incentives to teachers leads to better student outcomes. Moreover, both individual teacher-merit plans and school-wide group bonuses are effective at raising student test scores.

Conclusion

Much of education policy relies on incentives to induce greater student effort and performance. The policies reviewed in this article suggest that students respond to changing incentive structures. When faced with high-stakes exams that determine their promotion to the next grade or graduation from high school, students tend to work harder in order to meet the test cutoff. When offered the possibility of a merit scholarship based on good grades to pay for further schooling, student outcomes tend to improve. When teachers are offered merit awards based on their students' academic performance, student test scores tend to rise.

Taken together, the evidence to date suggests that the incorporation of appropriate incentive structures within education policy holds promise as a driving force for improving student learning outcomes.

See also: Education Production Functions: Concepts; Teacher Incentives; The Economics of School Accountability.

Bibliography

- Becker, W. and Rosen, S. (1992). The learning effect of assessment and evaluation in high school. *Economics of Education Review* **11**, 107–118.
- Betts, J. R. (1998). The impact of educational standards on the level and distribution of earnings. *American Economic Review* **88**, 266–275.
- Betts, J. R. and Costrell, R. M. (2001). Incentives and equity under standards-based reform. In Ravitch, D. (ed.) *Brookings Papers on Education Policy 2001*, pp 9–74. Washington, DC: The Brookings Institution.
- Betts, J. R. and Grogger, J. (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review* **22**, 343–352.
- Bishop, J. H. (2001). Strengthening incentives for student effort and learning: Michigan's merit award program? *CAHRS Working Paper Series Working Paper 01–10*.
- Carnoy, M. and Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis* **24**, 305–331.

- Cornwell, C., Mustard, D. B., and Sridhar, D. J. (2006). The enrollment effects of merit-based financial aid: Evidence from Georgia's HOPE program. *Journal of Labor Economics* **24**, 761–786.
- Figlio, D. N. and Kenny, L. W. (2007). Individual teacher incentives and student performance. *Journal of Public Economics* **91**, 901–914.
- Hanushek, E. A. and Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management* **24**, 297–327.
- Jacob, B. A. (2001). Getting tough? The impact of mandatory high school graduation exams on student outcomes. *Educational Evaluation and Policy Analysis* **23**, 99–122.
- Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics* **89**, 761–796.
- Jacob, B. A. and Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics* **86**, 226–244.
- Kremer, M., Miguel, E., and Thornton, R. (2007). Incentives to learn. *Poverty Action Lab Paper No. 42*.
- Lavy, V. (2002). Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of Political Economy* **110**, 1286–1317.
- Martorell, F. (2004). *Do High School Graduation Exams Matter? A Regression Discontinuity Approach*. Department of Economics, UC Berkeley.
- Muralidharan, K. and Sundararaman, V. (2006). *Teacher Incentives in Developing Countries: Experimental Evidence from India*. Department of Economics, Harvard University.
- Podgursky, M. J. and Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management* **26**, 909–949.

Further Reading

- Bettinger, E. P. (2007). Paying to learn: The effect of financial incentives on elementary school test scores. *Case Western Reserve Working Paper*.
- Bishop, J. H. (2005). High school exit examinations: When do learning effects generalize? *CAHRS Working Paper Series Working Paper 05–04*.
- Cornwell, C., Kyung, H. L., and Mustard, D. B. (2005). The effects of merit-based financial aid on course enrollment, withdrawal, and completion in college. *Journal of Human Resources* **40**, 895–917.
- Dynarski, S. (2000). HOPE for whom? Financial aid for the middle class and its impact on college attendance. *National Tax Journal* **53**, 629–661.
- Eide, E. R. and Showalter, M. H. (2001). The Effect of grade retention on educational and labor market outcomes. *Economics of Education Review* **20**, 563–576.
- Jacob, B. and Lefgren, L. (2007). The effect of grade retention on high school completion. *National Bureau of Economic Research Working Paper No. 13514*.
- Manacorda, M. (2006). Grade failure, drop out and subsequent school outcomes: Quasi-experimental evidence from Uruguayan administrative data. *London School of Economics Working Paper*.
- McEwan, P. J. and Santibáñez, L. (2005). Teacher and principal incentives in Mexico. In Vegas, E. (ed.) *Incentives to Improve Teaching: Lessons from Latin America*, pp 213–253. Washington, DC: World Bank Press.
- Oettinger, G. S. (2002). The effect of nonlinear incentives on performance: Evidence from "Econ 101." *Review of Economics and Statistics* **84**, 509–517.

The Economic Role of the State in Education

D N Plank, Stanford University, Stanford, CA, USA

T E Davis, University of Maryland, College Park, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Actuarially fair insurance – An insurance policy is actuarially fair if the expected payout is equal to the premium.

Asymmetric information – When one party to a transaction has more or better information than another, there is asymmetric information. The field of information economics identifies various manifestations of asymmetric information, such as adverse selection, moral hazard, hidden action, screening, and signaling.

Division of labor – By breaking down industrial production into simple, repetitive tasks, each worker is able to specialize and become more productive. The division of labor increases economic efficiency.

Externalities – When the private actions of one individual or firm affect others, there exist either positive or negative externalities.

Government failure – When government action leads to an inefficient allocation of resources, there is a government failure. Government failure is a common focus of public-choice economists.

Market failure – There is a market failure when certain characteristics of the market lead to an inefficient allocation of resources. These characteristics may include the existence of a monopoly, asymmetric information, externalities, transactions costs, or poorly defined property rights.

Private – In economics, private refers to individual consumers and firms as opposed to government.

Public choice – Neoclassical economics is often criticized for giving short shrift to the role of government. Public choice is an area of economics that focuses on government by looking at the private motivations of bureaucrats and politicians in the public sector.

Public spending – Economists refer to government expenditure as public spending.

Rent-seeking – Rent-seeking occurs when economic actors manipulate the regulatory, economic, or political situation for their own benefit instead of earning a profit through production and trade.

Risk aversion – The unwillingness to accept an uncertain future payoff instead of a certain payoff with a lower value.

State monopoly – When a government agency is the sole provider of a good or service, economists call it a state monopoly.

The economic and political importance of education has increased dramatically over the course of the past century. Education is the largest item of public expenditure in countries around the world, and formal schooling consumes an ever-larger quantity of young people's time. The centrality of education in modern societies is mainly a consequence of state action. The state has built and expanded national education systems, encouraged and sometimes compelled young people to attend school, and fostered rewards systems that make adult success increasingly contingent on academic persistence and performance. In this article we question why this should be so, and discuss the economic factors that help to explain why the state finances and often provides schools.

Constructing Education Systems

Modern states constructed national education systems in the service of political, economic, and military goals (Archer, 1982). In France, for example, the state extended the public school system to all corners of the nation in order to foster a sense of national identity by encouraging fluency and literacy in French and familiarity with canonical knowledge and civic traditions (Weber, 1976). Following the opening of Japan in 1853, the state created a new education system, modeled on those in Prussia and the United States, in an effort to keep up economically and militarily with its Western rivals (Passin, 1965). In the US, state action supported the expansion, integration, and standardization of previously local educational systems (Tyack, 1974). More recently, the United Nations and the World Bank have encouraged and financed the construction and expansion of national education systems in countries around the world, in an effort to guarantee the right to education and to achieve the goal of education for all by 2015 (UNICEF, 2000).

National education systems grew inexorably in the nineteenth and twentieth centuries, in two distinct ways. On the one hand, the state worked systematically to extend educational opportunities both socially and geographically, to incorporate previously excluded groups

including rural children, girls, linguistic and ethnic minorities, and the disabled. On the other hand, the state sponsored and supported policies that required young people to spend an ever-increasing share of their time within the education system. These have evolved from the introduction and enforcement of child labor and compulsory education laws (which at first typically required 4 years of schooling, and now often require 12 or more) to current initiatives aimed at ensuring universal access to preschool and postsecondary education.

As in nineteenth-century Japan, state action to expand and improve national education systems has been and continues to be justified by reference to the imperatives of economic and military competitiveness. In the United States, for example, successive waves of state-sponsored educational reform have gained their impetus from public anxieties about keeping up with the Germans, the Russians, the Japanese, the Chinese, and the Finns (e.g., Reich, 1992; Marshall and Tucker, 1993). The competition among states for positional advantage in the global economy has spawned an educational arms race; the putative need for more and better education is called upon to justify increased state involvement in the education of its citizens.

Economics and the State's Role

Is Education a Public Good?

One generally acknowledged role for the state in the economy is the provision of public goods (Musgrave and Musgrave, 1980; Stiglitz, 2000). Consumption of these goods is nonrival: the amount consumed by one person has no effect on the amounts available for consumption by others. For example, adding more listeners to a radio broadcast does not diminish the value of the service to any of the existing listeners. Consumption of public goods is also nonexcludable: once a radio signal is broadcast there is no practical way to exclude additional listeners.

Private goods are both rival and excludable. The gallon of gas that a driver puts in his/her car is not available to other drivers, and the owner of the service station is readily able to exclude prospective consumers by demanding payment in advance. Private goods are efficiently provided through the familiar institutions of the market. Buyers and sellers have powerful incentives to reveal their true preferences, and their interactions determine how much of the good will be produced and sold.

Pure public goods must be financed by the state, as the price mechanism on which markets rely fails when goods are nonrival and nonexcludable. When consumption of a good is nonrival, it is not scarce to consumers; allocation of the good no longer depends on who has a stronger preference for it. When consumption of a good is nonexcludable, it is impossible to prevent consumers from

making use of the good, whether they have paid for it or not. Under these circumstances, consumers will not reveal their true preferences. The incentives they face instead push them to become free riders, consuming as much as they like of the good while paying little or nothing to support its provision (Olson, 1971). As a result, reliance on markets to produce public goods will result in too little (or no) production of these goods.

Stiglitz (2000) classifies education as a publicly provided private good because there is a large marginal cost associated with educating each additional child, which makes education rival. Education is also excludable, as can be readily observed in private schools or in tuition-funded colleges and universities. Since education does not satisfy the economist's definition of a public good, it could in principle be bought and sold in a market much the same as other private goods. The argument for public provision must therefore be sought elsewhere, in the failure of markets to produce an optimal level of educational output or an equitable distribution of educational opportunities and outcomes.

Market Failure in the Market for Education

Competitive markets may fail to deliver the optimal level of a good or service for a variety of reasons, including the presence of positive or negative externalities, information asymmetries between buyers and sellers, economies of scale, and risk aversion. Defenders of a strong state role in funding and providing schools argue that significant market failures in the market for education justify the state's involvement in the education system.

Externalities

Externalities exist when the private actions of one individual or firm affect others, either positively or negatively. For example, a farmer at the headwaters of a river decides how much water and fertilizer to use based on a private calculation of costs and benefits, without regard to the costs his/her decisions impose on fishermen and municipal water systems downstream. These external costs may be substantial, and state action may be the best way to ensure that they are taken into account (Coase, 1960).

Private decisions about education produce a number of mostly positive externalities for the broader society, above and beyond the benefits that the individual student receives. In considering how much education to consume, however, individuals base their decisions on their own private calculation of costs and benefits, taking no account of external benefits that may accrue to others. In consequence, leaving choices about education to individuals in a private market may result in a suboptimal level of educational investment for the society as a whole.

Economists have long recognized a number of positive externalities associated with an educated citizenry. In the

Wealth of Nations, for example, Smith (1937) described two external benefits of education. First, he argued that education is a necessary antidote to the mind-numbing repetition that results from the division of labor into the narrowly specialized tasks performed by each worker. Workers lacking in education would eventually become unable to converse, formulate emotional sentiments, or perform the normal duties of private life, rendering them unable to defend the country in a time of war. In addition, Smith argued that modern education, including science and mathematics, would provide a constant source of innovation in the production process.

More recent work has identified a wide variety of additional external benefits to educational investment. In general, a literate society functions more smoothly, with reduced communications costs, stronger democratic institutions, and a higher degree of social cohesion (Belfield, 2000; Stiglitz, 2000). Education also has a strong positive impact on health and fertility, both within and across generations. Educated mothers have fewer, healthier children, and the children of educated mothers are more likely to enrol and remain in school (Becker, 1991; Colclough, 1993; Schultz, 1988). Barnett (1985, 1996) has documented the effect of education on reducing crime and welfare payments.

There is a vast macroeconomic literature that suggests that education is a key input that drives technological change and economic growth (Barro, 1997; Hall and Jones, 1999; Krueger and Lindahl, 2001; Mankiw *et al.*, 1992; Romer, 1990). To the extent that this is so, state action to expand and improve the education that citizens receive and thereby raise the rate of economic growth may be fully justified (and even financed) by the increased productivity and gross domestic product (GDP) growth produced by educational investments (Hanushek and Kimko, 2000).

Realizing the external benefits of education may require state action; the economic and social gains produced by an educated citizenry and workforce may justify the state in encouraging individuals to acquire more education than they might otherwise prefer. In a contrasting view, however, increasingly powerful private incentives to invest in education may be sufficiently strong to produce high levels of external benefit even in the absence of state action. The fundamental question with regard to externalities is therefore the extent to which they are extra-marginal: Does state subsidy or state provision produce external benefits in addition to those that would be produced through private action alone?

Information asymmetry

One of the conditions for the efficient operation of markets is complete information. In a perfectly competitive market, buyers and sellers are on equal footing; both have full information on the quality and price of the product.

In some markets, however, including for instance the market for used cars, sellers have more and better information about the product than prospective buyers. Markets in which the distribution of information is asymmetric may produce inefficient outcomes (Akerlof, 1970; Stiglitz, 1996).

The education system is clearly characterized by asymmetric information between the consumers of education services (parents) and the producers (schools). Education is a complex bundle of services that involves a large commitment of time, a cumulative instructional process, and uncertain future payoffs. The production process is poorly understood and output measures are ambiguous, which makes interschool comparisons difficult. As a result, parents face great difficulties in accurately assessing the quality of the education services provided by the various schools available to their children.

There are two main ways to address the problem of asymmetric information in the education system. On the one hand, the state may provide schools itself, seeking to standardize educational services and guarantee minimum quality standards across schools. For example, the state may regulate curricula, require certification for teachers, or equalize funding across schools or school districts (Brown, 1992). On the other hand, the state may seek to increase the information available to parents by publishing data on the character and performance of different schools. Recent advances in assessment and information technology make the latter choice increasingly feasible (Gintis, 1995).

Uncertainty and risk aversion

Brown (1992) applies the economics of uncertainty and information to education to show why the public provision of education is so widespread and why schools are so similar. Parents face uncertainty about their children's abilities and the future payoffs to investments in their education. Schools are better able to shoulder risk than their more risk-averse students, enabling them to offer actuarially fair insurance (Mas-Colell *et al.*, 1995; Kreps, 1990). There are two components to this insurance. First, the vast majority of schools offer a comprehensive and broadly similar curriculum, giving all students access to a wide array of courses and programs. In addition, schools generally allow their students to accumulate a diversified portfolio of educational experiences, rather than requiring them to specialize in a specific course of study. Private and charter schools must compete with public schools in providing this form of insurance, which results in the observed similarity between curricula in state and other schools.

More fundamentally, parents and students may not fully know their own preferences for education due to the complex and cumulative process of schooling. Students' consumption of education services often relies critically on the goodwill and competence of the parents, which cannot be assumed (Gintis, 1995; Stiglitz, 2000).

As it is difficult to collateralize loans for primary and secondary schooling, parents may face credit constraints as well. In light of these circumstances, public provision may be warranted as a protection against the long-term private and social costs of bad choices or constrained resources.

Economies of scale

Economies of scale arise when producers' average total cost falls as output increases (Mankiw, 1998). In education, this suggests that larger schools and districts may face a lower per-pupil cost. For example, larger schools have a greater ability to provide science laboratories and libraries by spreading the cost over more tax-paying households. There are also potential scale economies in information gathering, organization, and in the development of a curriculum (Belfield, 2000). To the extent that there are economies of scale in the delivery of education services, market forces may result in a monopoly, as smaller schools are driven out of business by established state schools. Faced with this tendency for the market to drive out small schools, the state has two options: either run large schools as state monopolies, or actively encourage competition by leveling the playing field for smaller schools.

Equity and Equal Opportunity

The state may also have an interest in making the distribution of educational opportunities and attainments more equitable. State actions to advance equity goals may take either of two forms. On the one hand, the state may seek to ensure that all young people are provided with an education of sufficient duration and quality to equip them for productive citizenship and protect them from poverty (e.g., Colclough, 1993). On the other hand, the state may seek to alter the distribution of wealth and status in favor of previously disadvantaged groups through the provision of targeted subsidies or other forms of affirmative action (e.g., Fiske and Ladd, 2004).

The state can in principle pursue equity objectives through public financing rather than public provision, by targeting subsidies to specific groups. As Davis (1998) points out, however, the government may be more effective than private markets at ensuring the fair allocation and distribution of resources, equal access to services, nonprofit decision making in the best interest of consumers, appropriate personnel policies, and cooperative labor relations (Belfield, 2000).

In the US, education is primarily a local responsibility, which has led to large differences in the resources devoted to education. Where these differences solely related to consumer preferences for education, there would be very little need for government intervention. In fact, however, a number of state supreme courts have ruled that reliance on local property taxes to finance education violates the

provisions of state constitutions that guarantee equal access to public education (Odden and Picus, 2004; Stiglitz, 2000).

Critique of State Provision

The claim that the state should finance the education of its citizens is rarely subject to argument. Controversy arises over whether the state should provide schools itself, or underwrite the provision of schools by other actors including for-profit firms. The argument turns on questions of the relative efficiency and equity effects of state and private provision.

Friedman (1962) revived the libertarian argument against government provision of education, which dates back to Thomas Paine (West, 1964) and Smith (1937). Friedman acknowledged two justifications for government involvement in education. First, he described the positive externalities produced by schools, which include the basic skills and core values required for social stability. Second, he noted the state's paternalistic concern for the welfare of children, whose parents may not always act in their best interest. Friedman argued that these concerns warrant state funding and minimal regulation in the education system, but not a government monopoly in the provision of schooling.

Public Choice and Government Failure

In the decades since the publication of Friedman's seminal essay, economists and others have developed a comprehensive critique of state provision, closely associated with a call for greater reliance on markets in the education system (e.g., Chubb and Moe, 1990). Just as the case for state provision relies on claims of market failure, the case against state provision relies on claims of government failure (Tullock *et al.*, 2002), including inefficiency in production, inequity in provision, the institutionalization of low expectations, and rent-seeking by educators who exploit positions of authority and trust to increase their own utility at the expense of their students.

Inefficiency in production

State provision of schools may be less efficient than market provision for two main reasons. First, the absence of competition in the state education system undermines incentives for innovation and improvement. In a market where schools are obliged to compete with one another for the patronage of parents, in contrast, schools receive meaningful market signals about quality from consumers. The attendant challenge to compete for students should drive schools to lower costs, improve quality, and innovate (West, 1997). In the short run, competition might be expected to induce educators to work harder, to allocate

resources more efficiently, and to make better staffing decisions. In the long run, a competitive market for education could make schools more receptive to parental involvement, give them incentives to provide better student achievement information, reward more productive teachers with higher wages, lead schools to abandon unproductive pedagogical techniques, and ultimately affect the size and number of schools (Hoxby, 2003).

To the extent that state action in the provision of education replaces private production, consumers of education must rely on political institutions to voice their preferences. With the specification of the educational production function uncertain, the use of resources is determined through political interaction, according to criteria that bear no necessary connection to valued outcomes including student learning (Downs, 1957; Hanushek, 1986). The interplay of interests in the political system can easily lead to allocations of resources that are suboptimal from the standpoint of economic efficiency. For example, the free-rider problem may become a tax avoidance problem, as taxpayers seek to minimize the cost of educating other people's children (Tullock *et al.*, 2002). At the same time, the state's obligation to exercise strict control over the use of public resources may result in excessive regulation and high administrative costs, without commensurate gains in productivity (Hanushek, 1986; Stiglitz, 2000).

Inequity in opportunities and outcomes

Advocates of market-based policies in education argue that greater reliance on markets could also improve the overall equity of the education system (Coons *et al.*, 1970; Howell and Peterson, 2002). Wealthy parents are able to choose the schools that their children attend, either by moving to a desirable school district or by sending their children to private schools. It is only the children of the poor who are deprived of choice under the current system of state provision, and these children are often obliged to attend the least-salubrious and lowest-performing schools (Ladd, 2002; McEwan, 2000). Opening up the system to additional providers would serve to increase the number and variety of schools available to poor households (e.g., Tooley and Dixon, 2005), giving parents more and better choices about the schools their children attend and increasing the chances of educational and subsequent economic success for poor children (West, 1997).

Standardization and enforced mediocrity

State efforts to standardize educational services and guarantee minimum educational standards for all students may reduce the quality of instruction for young people attending state schools (Finn, 1993; Ravitch, 1985). The state's obligation to protect parents from risk and to ensure at least minimally equitable opportunities for students may prevent schools from accomplishing or even

seeking ambitious goals, preferring instead to institutionalize low expectations and low standards in an effort to ensure success for all of the children under their care. For example, the performance of American students on international assessments of reading and mathematics is often adduced as evidence of the disadvantages of state provision, to be remedied by the introduction of more choices and increased competition in the education system (e.g., Chubb and Moe, 1990; Walberg, 2007).

Rent-seeking and corruption

In the education system, as elsewhere in the public sector, individuals and groups may have opportunities to divert public resources to their own private benefit. Examples of rent-seeking may include widespread absenteeism among educators, who regard their jobs as sinecures and frequently fail to show up to teach their students or run their schools (Banerjee and Duflo, 2006; Castro and Fletcher, 1986). Transfer programs can be used to solicit political patronage rather than to improve the access to education of those in need (Plank, 1995). In similar fashion, unions may be more concerned with protecting the interests of their adult members than with the education of the children in their charge (Hess, 2004; Moe, 2001). Under some circumstances rent-seeking may give way to corruption, as educators employed by the state offer private instruction, passing grades, and diplomas for sale to students and households that are able to pay (Hallak and Poisson, 2007).

A Role for the State?

Under some circumstances, the state may supply education more efficiently than private markets (Belfield, 2000). First, it can bundle other services with education, such as health screening for vision and hearing. As the government offers unemployment insurance and welfare benefits, public provision of education that reduces these fiscal obligations could also be seen as bundling education and social services. Second, if people perceive the state as having a long-term commitment to the education of its citizens, then the state may be trusted more than private firms that come and go in a competitive market. As Shleifer (1998) has argued, if consumers have a difficult time identifying quality then producers will struggle to build their reputations and be unable to increase demand for their services. Third, economies of scale could lead to an education monopoly, which might be better run by the state than by a profit-seeking firm. Fourth, government providers can mediate the asymmetry of information and bargaining power between schools and individuals.

Gintis (1995) acknowledges that markets may be more efficient and more responsive to parents' preferences. He simultaneously argues that extensive state action may be required to ensure that markets work efficiently and fairly.

He proposes some regulatory interventions aimed at expanding school choice while mitigating market failures caused by violations of five fundamental assumptions:

Many producers. First, the state could subsidize student transportation costs. Second, it could either offer low-interest loans to new schools or construct the buildings and rent them to providers. Third, the state could force schools to share classrooms, athletic facilities, or other resources in order to decrease the minimum feasible school size.

Product quality is known. In order to be accredited to accept public tuition, schools could be required to provide standardized measures of school performance such as test scores, retention rates, and graduation rates. Schools could also be required to provide data on teacher accreditation, building safety and instructional techniques.

Consumers know their preferences. A guardian should be appointed when parents are found legally incapable of making sound decisions for their children. Unscrupulous schools that mislead parents should be sanctioned and schools should be prohibited from pandering to selfish parents who place their personal interests before the best interests of their children (e.g., by accepting kickbacks).

The price is set so that supply equals demand. The price of schooling is the tuition fee, which is set and paid by the government. Economists worry that because parents do not pay for education, they may consume more than they would in the absence of a government subsidy. In the presence of positive externalities, however, it is widely assumed that parents consume too little education on behalf of their children, rather than too much.

Education is a private good. As education confers both private and social benefits, some worry that parents will focus exclusively on the private benefits when making schooling decisions. Regulation can increase the social benefits to education by requiring schools to develop a curriculum that reflects social values and providing incentives to schools that encourage a diverse student body.

Education and the Shrinking State

After more than a century of steady growth, there are signs that the state's role in national education systems may have begun to shrink. As ever, there is little dissent from the idea that the state should finance the educational opportunities available to its citizens, but the benefits traditionally ascribed to state provision have been subjected to increasingly overt challenge. This is especially visible in higher education, where reduced public support for universities and their students has been matched by dramatic increases in the number and variety of private and for-profit higher education institutions (e.g., Altbach and Levy, 2005). Strong political pressures bolstered by arguments and evidence from economists

(e.g., Heckman and Masterov, 2007; Barnett, 1985) to expand the state's role in early childhood and preschool education have run into significant opposition, based in part on resistance to increased taxes and public spending, and in part on aversion to the idea that very young children should be placed under the care of the state rather than their families and communities (Fuller, 2007). In Europe and elsewhere, groups defined by communal, religious, and linguistic affinities have begun to demand increased control over the education of their own children, including the right to establish their own schools (Plank, 2006). Whether this marks a genuine turning point in the history of education or a brief pause in the steady expansion of the state's role in the education system remains to be seen.

See also: Competition and Student Performance; Educational Privatization; Human Capital; The Economics of Catholic Schools; The Economics of Charter Schools; The Economics of Parental Choice; The Economics of School Accountability; The Efficacy of Educational Vouchers; The External Benefits of Education; Tiebout Sorting and Competition.

Bibliography

- Akerlof, G. (1970). The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics* **84**(3), 488–500.
- Altbach, P. and Levy, D. (2005). *Private Higher Education: A Global Revolution*. Rotterdam: Sense Publishers.
- Archer, M. (1982). *The Sociology of Educational Expansion: Take-Off, Growth, and Inflation in Educational Systems*. New York: Sage.
- Banerjee, A. and Dufo, E. (2006). Addressing absence. *Journal of Economic Perspectives* **20**(1), 117–132.
- Barnett, S. (1985). Benefit-cost analysis of the Perry preschool program and its policy implications. *Educational Evaluation and Policy Analysis* **7**(4), 333–342.
- Barnett, S. (1996). *Lives in the Balance: Age-27 Benefit-Cost Analysis of the High/Scope Perry Preschool Program*. Ypsilanti, MI: The High/Scope Press.
- Barro, R. J. (1997). *Determinants of Economic Growth: A Cross-Country Empirical Study, Lionel Robbins Lectures*. Cambridge, MA: MIT Press.
- Becker, G. (1991). *A Treatise on the Family: Enlarged Edition*. Cambridge, MA: Harvard University Press.
- Belfield, C. (2000). *Economic Principles for Education: Theory and Evidence*. Northampton, MA: Edward Elgar Publishing.
- Brown, B. (1992). Why governments run schools. *Economics of Education Review* **11**(4), 287–300.
- Castro, C. and Fletcher, P. (1986). *A Escola que os Brasileiros Frequentaram em 1985*. Brasília: IPEA/IPLAN.
- Chubb, J. and Moe, T. (1990). *Politics, Markets, and America's Schools*. Washington, DC: Brookings Institution.
- Coase, R. (1960). The problem of social cost. *Journal of Law and Economics* **3**, 1–44.
- Colclough, C. (1993). *Educating All the Children: Strategies for Primary Schooling in the South*. Oxford: Clarendon Press.
- Coons, J., Clune, W., and Sugarman, S. (1970). *Private Wealth and Public Education*. Cambridge, MA: Harvard University Press.
- Davis, E. (1998). *Public Spending*. Harmondsworth, UK: Penguin.
- Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper Collins.

- Finn, C. (1993). *We Must Take Charge: Our Schools and Our Future*. New York: Free Press.
- Fiske, E. and Ladd, H. (2004). *Elusive Equity: Education Reform in Post-Apartheid South Africa*. Washington, DC: Brookings Institution Press.
- Friedman, M. (1962). *Capitalism and Freedom*. Chicago, IL: University of Chicago Press.
- Fuller, B. (2007). *Standardized Childhood: The Political and Cultural Struggle Over Early Education*. Stanford, CA: Stanford University Press.
- Gintis, H. (1995). The political economy of school choice. *Teachers College Record* **96**(3), 1–20.
- Hall, R. and Jones, C. (1999). Why do some countries produce so much more output per worker than others? *Quarterly Journal of Economics* **114**(1), 83–116.
- Hallak, J. and Poisson, M. (2007). *Corrupt Schools, Corrupt Universities: What Can Be Done?* Paris: UNESCO.
- Hanushek, E. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature* **24**, 1141–1177.
- Hanushek, E. and Kimko, D. (2000). Schooling, labor-force quality, and the growth of nations. *American Economic Review* **90**(5), 1184–1208.
- Heckman, J. and Masterov, D. (2007). *The Productivity Argument for Investing in Young Children*. NBER Working Paper No. 13016. Cambridge, MA: National Bureau of Economic Research.
- Hess, F. (2004). *Teacher Quality, Teacher Pay*. Washington, DC: American Enterprise Institute. http://www.aei.org/publications/filter_all.pubID.20326/pub_detail.asp (accessed September 2009).
- Howell, W. and Peterson, P. (2002). *The Education Gap: Vouchers and Urban Schools*. Washington, DC: Brookings Institution.
- Hoxby, C. (2003). School choice and school productivity: Could school choice be a tide that lifts all boats? In Hoxby, C. (ed.) *The Economics of School Choice*, pp 287–342. Chicago, IL: The University of Chicago Press.
- Kreps, D. (1990). *A Course in Microeconomic Theory*. Princeton, NJ: Princeton University Press.
- Krueger, A. and Lindahl, M. (2001). Education for growth: Why and for whom? *Journal of Economic Literature* **39**, 1101–1136.
- Ladd, H. (2002). School vouchers: A critical view. *Journal of Economic Perspectives* **16**(4), 3–24.
- Mankiw, N. (1998). *Principles of Microeconomics*. Orlando, FL: Dryden Press.
- Mankiw, N., Romer, D., and Weil, D. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics* **107**(2), 407–437.
- Marshall, R. and Tucker, M. (1993). *Thinking for a Living: Education and the Wealth of Nations*. New York: Basic Books.
- Mas-Colell, A., Whinston, M., and Green, J. (1995). *Microeconomic Theory*. New York: Oxford University Press.
- McEwan, P. (2000). The potential impact of large-scale voucher programs. *Occasional Paper No. 2, The Occasional Paper Series of the National Center for the Study of Privatization in Education*. New York: Teachers College, Columbia University.
- Moe, T. (2001). Teachers unions and the public schools. In Moe, T. (ed.) *A Primer on America's Schools*, pp 151–184. Stanford, CA: Hoover Institution Press.
- Musgrave, R. and Musgrave, P. (1980). *Public Finance in Theory and Practice*, 3rd edn. New York: McGraw-Hill.
- Odden, A. and Picus, L. (2004). *School Finance: A Policy Perspective*, 3rd edn. New York: McGraw-Hill.
- Olson, M. (1971). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.
- Passin, H. (1965). *Society and Education in Japan*. New York: Columbia University Press.
- Plank, D. (1995). *The Means of Our Salvation: Public Education in Brazil, 1930–1995*. Denver: Westview Press.
- Plank, D. (2006). Unsettling the state: How 'demand' challenges the education system in the US. *European Journal of Education* **41**(1), 13–27.
- Ravitch, D. (1985). *The Troubled Crusade: American Education, 1945–1980*. New York: Basic Books.
- Reich, R. (1992). *The Work of Nations: Preparing Ourselves for 21st Century Capitalism*. New York: Vintage Press.
- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy* **98**(5; part 2), S71–S102.
- Schultz, T. P. (1988). Educational investments and returns. In Chenery, H. and Srinivasan, T. N. (eds.) *Handbook of Development Economics*, vol. 1, Sections 2–S, pp 543–630. Amsterdam: North-Holland Press.
- Shleifer, A. (1998). State versus private ownership. *Journal of Economic Perspectives* **12**(4), 133–150.
- Smith, A. (1937). *An Inquiry into the Nature and Causes of the Wealth of Nations* (originally published in 1776). New York: The Modern Library.
- Stiglitz, J. (1996). *Whither Socialism?* Cambridge, MA: MIT Press.
- Stiglitz, J. (2000). *Economics of the Public Sector*, 3rd edn. New York: WW Norton.
- Tooley, J. and Dixon, P. (2005). *Private Education is Good for the Poor: A Study of Private Schools Serving the Poor in Low-Income Countries*. Washington, DC: Cato Institute.
- Tullock, G., Seldon, A., and Brady, G. (2002). *Government Failure: A Primer in Public Choice*. Washington, DC: Cato Institute.
- Tyack, D. (1974). *The One Best System: A History of American Urban Education*. Cambridge, MA: Harvard University Press.
- UNICEF (2000). *Education for All: No Excuses*. New York: UNICEF.
- Walberg, H. (2007). *School Choice: The Findings*. Washington, DC: Cato Institute.
- Weber, E. (1976). *Peasants into Frenchmen: The Modernization of Rural France, 1870–1914*. Stanford, CA: Stanford University Press.
- West, E. (1964). Private versus public education: A classical economic dispute. *Journal of Political Economy* **72**(5), 465–475.
- West, E. (1997). Education vouchers in principle and practice: A survey. *World Bank Research Observer* **12**, 83–103.

The Economics of Catholic Schools

W Sander, DePaul University, Chicago, IL, USA

D Cohen-Zada, Ben-Gurion University, Beer-Sheva, Israel

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Error term – A variable that is created when the model does not fully represent the actual relationship between the independent variable(s) and the dependent variable.

Exogenous variable – A factor in a causal model or causal system whose value is independent from the states of other variables in the system.

Instrumental variable – A method that is used to estimate causal relationships when controlled experiments are not feasible.

Self-selection – A term that is used to indicate that individuals select themselves into a group causing a biased sample.

Standard deviation – A measure of variability or dispersion.

Introduction

This article examines Catholic schooling in the United States. The article is organized as follows. First, an overview of Catholic schooling in the United States is given. Second, the determinants of the demand for Catholic schools are reviewed. Third, literature on the effects of Catholic schooling is presented. Although the focus of this article is on Catholic schooling in the United States, some attention is also given to Catholic schooling in other countries.

Overview

History and Enrolment

The establishment of Catholic schools in the United States was partly a result of the nature of public schooling in the nineteenth century. At that time, many public schools promoted Protestantism and some were overtly anti-Catholic. This resulted in Catholics establishing their own parochial schools. In 1884, Catholic bishops in the United States asked Catholic parishes to support at least one Catholic school and Catholic parents to send their children to them. By 1900, about 5000 Catholic schools had been established serving around 1 million students. The increase in enrolment in Catholic schools

was also due to the massive waves of immigration of Catholics to the United States. According to Finke and Stark (2005), between 1890 and 1906, the Catholic population increased from 7 million Catholics (12% of the population) to more than 14 million Catholics (17% of the population). Catholic schools continued to increase both in number and the percentage of the population served until about 1960. By 1960, there were almost 13 000 Catholic schools serving over 5 million students. Catholic schools at that time accounted for about 90% of private-school enrolment. Since 1960, Catholic schools have declined in number and the percentage of the population served. Many Catholic schools especially in big cities in the East and Midwest regions have closed. For example, about half of the Catholic schools in Chicago have closed over the past three decades. Enrolment has declined even more from about 132 000 in 1976–77 to 52 000 in 2004–05 (McDonald, 2007).

By 2000, about 11% of elementary and secondary school students attended private schools. Although the percentage attending private schools has not changed too much since 1960, only about half of this population is now in the Catholic school sector. Other religious schools and nonsectarian private schools account for the rest of the private school sector. The most recent data indicate that there are 7498 Catholic elementary and secondary schools that serve over 2 million students. Although Catholic schools have declined overall, there has been some growth in demand in the West and Southeast. This is a result of high rates of economic growth in these areas that has resulted in more Catholics living in these regions (McDonald, 2007).

Teachers

If one goes back in time, most of the teachers in Catholic schools were in religious orders. For example, in 1960 about three out of four Catholic school teachers were members of a Catholic religious order. Over time, this percentage has declined markedly. Recent data indicate that only about 5% of Catholic school teachers are members of a Catholic religious order (McDonald, 2007).

Students

Part of the decline in the Catholic school population is a result of a decline in the demand for a Catholic

education by Catholics. For Catholics born before 1950, at least half of them attended a Catholic school for at least 1 year. By 1991, only about one in five Catholic children of elementary school age attended a Catholic school (Youniss and Convey, 2000).

The decline in the demand for a Catholic education by Catholics has been partially offset by an increase in the demand for a Catholic education by non-Catholics. In 1970, only about 3% of Catholic school students were not Catholic. By 2007, the percentage of non-Catholics in Catholic schools increased to about 14%. Part of the increase in non-Catholics attending Catholic schools is a result of more non-Catholic African-Americans in big cities attending Catholic schools. For example, in Chicago about three out of four blacks in Catholic schools are not Catholic (McDonald, 2007; Sander, 2001).

Although most students in Catholic schools are non-Hispanic whites, the percentage of minorities in Catholic schools has increased over time to about one in four students in 2007. About 7% of Catholic school students are black, 6% are Hispanic, 3% are Asian, and 2% are multiracial (McDonald, 2007).

Location

Catholic schools are concentrated in urban areas and in the East and Midwest regions.

Over 40% of Catholic schools are located in big cities while about one in three are located in suburbs of cities. The rest are located in rural areas. Further, over 50% of Catholic schools are located in the East or Midwest regions where the Catholic population is more concentrated. The ten largest Catholic school systems are as follows: (1) New York, (2) Chicago, (3) Philadelphia, (4) Los Angeles, (5) Brooklyn, (6) Cleveland, (7) St. Louis, (8) Cincinnati, (9) Boston, and (10) Newark (McDonald, 2007).

Some metropolitan areas are served by more than one Catholic school system such as New York and Chicago. The reason for this is that the boundaries of the school system are defined by the Catholic archdioceses rather than census boundaries.

Tuition and Costs

The average tuition at Catholic schools in 2007 was US \$2607 at parish-based elementary schools and US\$6906 at secondary schools although the per-pupil cost was higher – US\$4268 at grade schools and US\$8743 at high schools. The difference between tuition and cost is a result of many factors such as parish subsidies and fund raising. This is below the average per-pupil expenditures at public schools. One of the reasons for relatively low costs at Catholic schools is the relatively low teacher salaries. Further, subsidies from parishes and elsewhere help keep tuition below costs. Recent data indicate that

the tuition at Catholic grades schools covers 61.7% of the actual cost per pupil while 80% of the cost per pupil is covered by tuition at Catholic high schools (Harris, 1996; McDonald, 2007).

Hoxby (1994) shows that higher Catholic-population densities have several important implications for tuition and per-pupil cost. First, the number of Catholic schools and Catholic school places per person increase as density goes up. Second, subsidies as a percentage of Catholic school income increase substantially as the Catholic share in the population grows. Third, tuition per pupil decreases as the Catholic share in the population increases.

Market Served

Although Catholic schools were established to serve all Catholics and did so for many decades, higher costs and tuition are resulting in Catholic schools serving fewer low-income students and more high-income students. Still, many Catholic school systems try to reach out to students who cannot afford them. For example, in Chicago many Catholic schools are designated as Big Shoulders schools – schools that mostly serve low-income minority students. The dropout rate in these schools is only 1%, while the high school graduation rate is 97%. The dropout rate for public schools in Chicago is markedly higher (Sander, 2001, 2006).

School Practices

It is possible that Catholic schools are more effective than public schools due to differences in school practices (Figlio and Ludwig, 2000). It has been reported that the proportion of students in Catholic schools who report that their school is excellent or good in effectiveness and of discipline is substantially higher than what is reported by public school students (72% vs. 42%). They report similar differences between Catholic and public schools in fairness and discipline (46% vs. 36%).

Vouchers

Vouchers that would enable more students to attend Catholic schools have been a controversial political issue in the United States. Several voucher programs are now ongoing in several states. In the New York voucher experiment, Howell and Peterson (2002) compare the achievement of voucher recipients relative to a similar control group, where the voucher amount was applied mainly to tuition at low-cost Catholic schools. They find that there are positive gains from using educational vouchers only for African-Americans. However, these gains are not consistent across the years of the trial. They have been challenged on methodological grounds by Krueger and Zhu (2004) who

find that the difference in test scores between the treatment group and the control group is not significant.

Catholic Schools in Other Countries

Catholic schools are also present in many other countries. In fact, the Catholic school system is the world's largest faith-based educational network with 120 000 schools and over 1000 colleges and universities (Grace and O'Keefe, 2007). In Australia, Catholic schools form the second largest sector after government schools (Wikipedia). As of the early 1990s, private schools accounted for about 25% of enrolment with approximately 75% attending Catholic schools (Vella, 1999). In Canada, seven of the 13 provinces and territories fund faith-based Catholic schools (Wikipedia).

In England and Wales, 2300 Catholic schools educate around 840 000 pupils each year. In Scotland, Catholic schools are all fully funded by the government. In Northern Ireland, about half of all students attend Catholic schools (Wikipedia).

In Italy, Catholic schools are relatively expensive and therefore tend to serve high-income families. In 1998, approximately 10.3% of high school graduates attended private schools in Italy. Of these, about one-third were Catholic (Di Pietro and Cutillo, 2006). In the Netherlands, about one in five primary schools are Catholic (Levin, 2002). In Chile, about one in ten grade school students is enrolled in a Catholic school (McEwan, 2001).

Demand

Many factors affect the probability that parents send their children to Catholic schools. The most recent study that estimates the relative importance of these factors finds that the probability of attending a Catholic school is higher for Catholics, especially more religious Catholics. Catholics who attend church at least weekly have about a 50% likelihood of sending their children to Catholic schools; Catholics who attend church once a month only have about a 25% likelihood of sending their children to Catholic schools; and Catholics who attend church less than once a month have an even lower probability of sending their children to Catholic schools. Non-Catholics have a very low probability of sending their children to Catholic schools regardless of their level of religiosity (Cohen-Zada and Sander, 2007).

Over time, Catholic religiosity as measured by church attendance has declined for Catholics aged less than 65 years. About 50% of Catholics 35–44 years attended church almost every week or more often during the 1970s. Regular church attendance declined for this age group to about 25% by the 1990s. If church attendance by adults is a determinant of the demand for a Catholic education,

declines in Catholic religiosity is a plausible reason for some of the decline in the demand for a Catholic education for children (Sander, 2005).

Other factors that have been found to affect the demand for Catholic schools include family income, parents' education, location, and race. Families who have very low incomes are less likely to send their children to Catholic schools. Parents with some college are more likely to send their children to Catholic schools relative to parents with a college degree and parents with a high school diploma. Parents with less than a high school education were the least likely to send their children to a Catholic school. Also, African-American has a positive effect on the demand for Catholic schools. Although blacks are disproportionately Protestant, a likely reason that they have a higher demand for Catholic schooling is that they are concentrated in big cities where the quality of public schooling is problematic (Cohen-Zada and Sander, 2007).

Another important factor that affects the demand for Catholic schooling is price. Large declines in the number of teachers in religious orders have forced Catholic schools to hire lay teachers who are more costly. Further, declines in subsidies from Catholic parishes have resulted in large increases in tuition. This is partly a result of declines in Catholic Church contributions over time. If one goes back to the 1960s, Catholics contributed over 2% of their income to the church. This declined to about 1% by the 1980s (Harris, 1996; Zech, 2000).

Further, the quality of public schools in a locality affects the demand for Catholic schools. In inner-city areas where the quality of public schooling is often perceived as low, the demand for a Catholic education is higher. For example, in Chicago about half of the children of non-Hispanic whites attend private schools. Most of these schools are Catholic. In the suburbs of Chicago, the demand for Catholic schools is much lower because the quality of public education is higher. Catholic parents are less willing to pay for both public schools and Catholic schools in such areas (Sander, 2006).

At the aggregate level, several factors affect the share of students in a county that attend Catholic schools (Cohen-Zada, 2006). The variable that has the most substantial effect is the Catholic share in the population. A one standard deviation increase in the Catholic share in the population increases the county Catholic-enrolment rate by more than 2 percentage points. This variable has a concave effect on the Catholic-enrolment rate. It has been shown that in places where Catholics are a lower share of the population, a higher percentage of them send their children to Catholic schools. One possible reason for this is that in these places where they are a minority, they have a stronger need to send their children to Catholic schools in order to preserve their religious identity. A higher Catholic-enrolment rate is also associated with a higher population density, higher mean income of the

population, higher share of African-Americans in the population, and a lower share living in a rural area. Finally, for a given Catholic share in the population, a higher share of Hispanics implies a lower-Catholic enrolment rate.

Effects

Academic Achievement and Educational Attainment

Students who attend Catholic high schools are more likely to graduate and go on to college relative to their public school counterparts. Recent data put the high school graduation rate in Catholic schools at 99.1%. The public high school graduation rate is only 71.0% if general educational development (GED) certificates are excluded. It is 86.5% if they are included. Further, students in Catholic schools have higher test scores on the average than students in public school. Data from the National Assessment of Education Progress (NAEP), the nation's report card, indicate that students in Catholic grade schools and high schools have higher test scores on the average than students in public schools (McDonald, 2007).

Higher graduation rates and test scores do not necessarily indicate that Catholic schools provide superior schooling. It could be the case that Catholic schools simply select better students. This so-called selection issue has received considerable attention by researchers.

There are a large number of studies on the effect of Catholic schools on various measures of educational attainment and academic achievement that try to control for selectivity bias. Important studies in the early 1980s by James Coleman and others at the University of Chicago indicated that Catholic high schools had positive effects on test scores and high school graduation rates (Coleman and Hoffer, 1987; Coleman *et al.*, 1982). The results in these studies were considered problematic by some because they did not control for selection into Catholic schools, especially on unobserved variables like unobserved aspects of family background.

A number of studies over the past two decades have tried to statistically control for the possibility of positive selection into Catholic schools (Neal, 1997; Perie and Goldstein, 2005; Sander, 2001). The idea is that if Catholic school attendance is correlated with any determinant of students' outcome that is missing in the estimation, one is likely to obtain biased estimates of Catholic school attendance on educational outcomes. The bias is a result of the estimated coefficient of Catholic school attendance capturing the effect of the missing variable on the outcome. For example, Cohen-Zada and Sander (2007) show that religiosity is positively correlated with Catholic school attendance and student outcomes. Thus, if religiosity is not taken into account, one tends to overestimate the effect of Catholic school attendance on educational outcomes.

Most studies try to control for selection into Catholic schools by using instrumental variables that causally affect Catholic school attendance, but do not have a direct effect on educational outcomes. Using this strategy, one estimates the effect of Catholic school attendance on outcomes using a two-stage approach. In the first stage, Catholic school attendance is regressed on the instruments and all of the other exogenous variables in the model. Assuming the instruments are exogenous, the predicted Catholic school attendance that one obtains from the first stage is also exogenous. This allows one to tease out the exogenous component of the variation in Catholic school attendance. In the second stage, the outcome equation is estimated by replacing Catholic school attendance with its exogenous approximation obtained from the first stage. This identification strategy is valid only if there is no unobserved determinant of the educational outcome that is correlated with the instrumental variable.

The results of many of the studies on the effect of Catholic schooling are somewhat problematic because the instruments that are used are not valid. That is, they have a direct effect on the outcome and are thus likely to be correlated with the error term. For example, Cohen-Zada and Sander (2007) show that Catholic religion and density of Catholic schools, which are often used as instruments for Catholic school attendance, are correlated with religiosity which is often missing in estimations of the treatment effect of Catholic schools. Altonji *et al.* (2005b) examined the validity of religious affiliation and the proximity to Catholic schools as instruments for Catholic school attendance and found that none of them were useful sources of identification of the Catholic school effect.

Perhaps the best study to date on Catholic high school effects finds that they have a positive effect on high school graduation rates and on attending college but no effect on test scores. This study acknowledges that none of the previously used instruments for Catholic school attendance is valid and thus develops a new estimation method for obtaining a lower bound for the causal effect of Catholic school attendance on outcomes based on the assumption that the amount of selection on the unobservables is not more than the amount of selection on the observables (Altonji *et al.*, 2005a). Recently, Cohen-Zada (2007) shows that the Catholic share in a county in 1890 is a valid instrument for private school competition and a potentially valid instrument for Catholic school attendance.

Another important study on Catholic high schools finds that the benefits of a Catholic education are particularly high for minorities in big cities where the quality of public schooling is low (Neal, 1997). Although most of the studies on Catholic schools have tended to focus on Catholic high schools, studies on Catholic grade schools indicate that the effects on academic achievement are modest at best (Jepsen, 2003; Sander, 2001).

Bad Behavior

A recent study uses a propensity score matching method to control for positive selection into Catholic schools and finds that Catholic school attendance reduces the propensity that female students use cocaine and have sex. However, it increases the propensity that male students use and sell drugs (Mocan and Tekin, 2006). Other studies provide evidence that Catholic schools are not more effective than public schools in reducing marijuana use, alcohol abuse, or smoking although Catholic school attendance may reduce some risky youth behaviors including sexual activity, criminal behavior, and the use of hard drugs (Figlio and Ludwig, 2000).

Civic Participation and Altruism

A number of studies have examined the effect of Catholic and other private schooling on civic participation and altruism. Some results from this literature are as follows. A study on the effect of vouchers on students' altruistic behavior finds that students attending private (mostly Catholic) schools in Ohio, after being randomly assigned to these schools through a lottery, gave more to charitable organizations but not more to their peers (other children attending the experiment) (Bettinger and Slonim, 2006). Another study using data from the National Household Education Survey of 1999 finds that Catholic school attendance is positively associated with community service participation, civic skills, political knowledge, and tolerance (Belfield, 2004). Greeley and Rossi (1966) found similar results regarding tolerance of civic liberties; on the other hand, they did not find that Catholic school attendance was related to community involvement. Other findings include Catholic schools increasing racial tolerance and voting (Dee, 2005; Greene, 1998).

Economic Effects

One of the few studies that considers the effects of Catholic schooling on economic outcomes finds that attending a Catholic high school increases earnings both directly and indirectly for young black and Hispanic men. Catholic schooling directly increases wages after controlling for educational attainment and indirectly increases wages by increasing the probability of graduating from high school and college. In this study, no effect is discerned for non-Hispanic white men (Neal, 1997).

Other Effects

Although studies on Catholic schools have focused on academic effects, other effects have been considered as well. One line of research focuses on the effect of Catholic

and other private schooling on the quality of public education. One of the issues of interest is whether competitive pressures from private schools could increase the quality of public schools. The findings from research on this topic are not conclusive (Hoxby, 1994).

Catholic schools exist in part to provide a religious value-based education. The data indicate that Catholics who attend Catholic schools are more religious Catholics. They tend to pray more, attend church more often, retain a Catholic identity as an adult, and have Catholic beliefs. Statistical analysis of the effects of Catholic schools on religiosity indicates that Catholic schools have a positive causal effect on all of these outcomes except church attendance (Sander, 2001).

The Effect of Catholic School Attendance in Other Countries

In the last decade, researchers have examined whether the substantial effect reported for attending Catholic schools in the United States, also exists for students in Catholic schools in other countries. A study on Catholic schooling in Australia indicates that attending a Catholic school increases the probability of completing high school, attending college, and being employed (Vella, 1999).

A recent study on Catholic schooling in Italy finds that attending a Catholic school increases the probability of attending a university, but it does not affect the probability of dropping out of college (Di Pietro and Cutillo, 2006).

A study on Catholic schooling in the Netherlands finds significantly higher achievement scores for fourth-, sixth-, and eighth-grade students in Catholic schools (Levin, 2002). The Dutch experience is of particular interest because Dutch law mandates equal public financing across all schools, regardless of their religious affiliation.

Conclusions

Catholic schools have made important contributions to education in the United States and elsewhere. However, the effects of Catholic schooling have been difficult to nail down conclusively. Further, little attention has been given to the changing nature of Catholic schooling by location and over time. Similarly, very little attention has been given to the effects of different types of Catholic schools such as parish-based schools or schools sponsored by religious orders. This is the grist for future research.

See also: Education and Civic Engagement; Educational Privatization; Empirical Research Methods in the Economics of Education; The Efficacy of Educational Vouchers.

Bibliography

- Altonji, J. G., Elder, T. E., and Taber, C. R. (2005a). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy* **113**, 151–184.
- Altonji, J., Elder, T., and Taber, C. (2005b). An evaluation of instrumental variable strategies for estimating the effects of Catholic schools. *Journal of Human Resources* **40**, 791–821.
- Belfield, C. R. (2004). Democratic education across school types: Evidence for the U.S. from NHES99. *Education Policy Analysis Archives* **12**(43). <http://epaa.asu.edu/epaa/v12n43> (accessed June 2009).
- Bettinger, E. and Slonim, R. (2006). Using experimental economics to measure the effects of a natural educational experiment on altruism. *Journal of Public Economics* **90**, 1625–1648.
- Cohen-Zada, D. (2006). Preserving religious values through education: Economic analysis and evidence from the US. *Journal of Urban Economics* **60**, 372–398.
- Cohen-Zada, D. (2007). An alternative instrument for private school competition. *Economics of Education Review* **28**, 29–37.
- Cohen-Zada, D. and Sander, W. (2007). Religion, religiosity, and private school choice: Implications for estimating the effectiveness of private schools. *Journal of Urban Economics* **64**, 85–100.
- Coleman, J. and Hoffer, T. (1987). *Public and Private High Schools*. New York: Basic Books.
- Coleman, J., Hoffer, T., and Kilgore, S. (1982). *High School Achievement: Public, Catholic and Private Schools Compared*. New York: Basic Books.
- Dee, T. (2005). The effects of Catholic schooling on civic participation. *International Tax and Public Finance* **12**, 1–21.
- Di Pietro, G. and Cuttillo, A. (2006). Does attending a Catholic school make a difference? Evidence from Italy. *Bulletin of Economic Research* **58**, 193–234.
- Figlio, D. and Ludwig, J. (2000). Sex, drugs and Catholic schools: Private schooling and non-market adolescent behaviors. *NBER Working Paper Number 7990*.
- Finke, R. and Stark, R. (2005). *The Churching of America, 1776–2005: Winners and Losers in Our Religious Economy*. Piscataway, NJ: Rutgers University Press.
- Grace, G. and O'Keefe, J. (eds.) (2007). *International Handbook of Catholic Education*. Dordrecht: Springer.
- Greeley, A. M. and Rossi, P. H. (1966). *The Education of American Catholics*. Chicago, IL: Aldine.
- Greene, J. P. (1998). Civic values in public and private schools. In Peterson, P. E. and Hassel, B. C. (eds.) *Learning from School Choice*, pp 335–356. Washington, DC: Brookings Institution Press.
- Harris, J. C. (1996). *The Cost of Catholic Parishes and Schools*. Kansas City, MO: Sheed and Ward.
- Howell, W. and Peterson, P. (2002). *The Education Gap*. Washington, DC: Brookings Institution.
- Hoxby, C. M. (1994). Do private schools provide competition for public schools? *National Bureau of Economic Research Working Paper No. 4978*.
- Jepsen, C. (2003). The effects of Catholic primary schools. *Journal of Human Resources* **38**, 928–941.
- Krueger, A. and Zhu, P. (2004). Another look at the New York City school voucher experiment. *American Behavioral Scientist* **47**, 658–698.
- Levin, J. (2002). *Essays in the Economics of Education*. Amsterdam: Tinbergen Institute.
- McDonald, D. (2007). *United States Catholic Elementary and Secondary Schools 2006–2007*. Washington, DC: National Catholic Educational Association.
- McEwan, P. (2001). The effectiveness of public, Catholic, and non-religion private schools in Chile's voucher system. *Education Economics* **9**, 103–128.
- Mocan, N. and Tekin, E. (2006). Catholic schools and bad behavior: A propensity score matching analysis. *Contributions to Economic Analysis and Policy* **5**(1), 1403.
- Neal, D. (1997). The effects of Catholic secondary schooling on educational attainment. *Journal of Labor Economics* **15**, 98–123.
- Perie, M. and Goldstein, A. (2005). *Student Achievement in Private Schools: Results from NAEP 2000–2005*. Washington, DC: US Government Printing Office.
- Sander, W. (2001). *Catholic Schools: Private and Social Effects*. Boston, MA: Kluwer Academic Press.
- Sander, W. (2005). Catholics and Catholic schooling. *Education Economics* **13**, 257–268.
- Sander, W. (2006). Private schools and school enrollment in Chicago. *Chicago Fed Letter Number 231*. The Federal Reserve Bank of Chicago.
- Vella, F. (1999). Do Catholic schools make a difference? Evidence from Australia. *Journal of Human Resources* **34**, 208–224.
- Wikipedia (2007). *The Free Encyclopedia*. Youniss, J. and Convey, J. J. (eds.) (2000). *Catholic Schools at the Crossroads*. New York: Teachers College, Columbia University.
- Zech, C. E. (2000). *Why Catholics Don't Give*. Huntington, IN: Our Sunday Visitor Publishing Division.

Further Reading

- Bryk, A. S., Lee, V. S., and Holland, P. B. (1993). *Catholic Schools and the Common Good*. Cambridge: Harvard University Press.
- Cohen-Zada, D. and Justman, M. (2005). The religious factor in private education. *Journal of Urban Economics* **57**, 391–418.
- Evans, W. N. and Schwab, R. M. (1995). Finishing high school and starting college: Do Catholic schools make a difference? *Quarterly Journal of Economics* **110**, 941–974.
- Greeley, A. M., McCready, W. C., and McCourt, D. (1976). *Catholic Schools in a Declining Church*. Kansas City, MO: Sheed and Ward.
- Grogger, J. and Neal, D. (2000). Further evidence on the effects of Catholic secondary schooling. In Gale, W. G. and Pack, J. R. (eds.) *Brookings-Wharton Papers on Urban Affairs*, pp 151–193. Washington, DC: The Brookings Institution.
- Sander, W. and Krautmann, A. (1995). Catholic schools, dropout rates, and educational attainment. *Economic Inquiry* **33**, 217–233.

Relevant Websites

- <http://www.ncea.org> – National Catholic Educational Association.
- <http://www.ncspe.org> – National Center for the Study of Privatization in Education.
- <http://nces.ed.gov> – The National Center for Education Statistics (NCES).

The Economics of Charter Schools

R Zimmer, Michigan State University, East Lansing, MI, USA

R Buddin, RAND Corporation, Santa Monica, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Charter schools – Are schools that receive public funding under a contract, or charter, with a public entity (e.g., school district, state, and universities), and are given greater autonomy than other public schools over curriculum, instruction, and operations.

External validity – An estimate of an effect that is valid beyond the sample of observations included in the analysis.

Fixed effects – Econometric method of estimating time-invariant factors in panel data analysis.

Internal validity – An estimate of an effect that is valid only to the sample of observations included in the analysis.

Selection bias – Estimated effect of programs could be biased because participants have chosen to participate, which is not captured by the econometric model.

Introduction

Charter schools are an educational reform in the United States that provides more school alternatives to parents and more local autonomy for individual schools. Individual schools receive public funding under a contract, or charter, with a public entity (e.g., school district, state, and universities), and are given greater autonomy than other public schools over curriculum, instruction, and operations. In exchange for greater autonomy, charter schools are held accountable for results.

The first US charter school opened in 1992, and the scale of the charter movement has grown to 4000 schools and over a million students in 40 states plus the District of Columbia. Charter enrolments constitute about 2% of annual public school enrolments in the United States with enrolment shares as high as 18% and 25% for Arizona and the District of Columbia, respectively. About 20 communities with more than 10 000 public school students have charter enrolment shares of 13% or more.

Charter reform has been a contentious education and public policy issue. Proponents argue that traditional public schools are burdened by excessive centralized control and regulation and greater school autonomy under charter

reform allows schools to innovate and improve (Finn *et al.*, 2000). School choice is heralded as a key component of charter success – charters must compete for students with traditional public schools, and their survival is literally predicated on their ability to implement programs that attract students.

Critics see charters as drawing resources away from traditional public schools and distracting them from internal reforms and innovation. These opponents argue that charter schools are no more effective than traditional public schools, that they may exacerbate racial segregation, that they create fiscal strains for school districts, and that too many of them are unreliable operations (Wells *et al.*, 1998). Charter competition is seen as inconsequential in many cases or as counterproductive when traditional public schools placate parents with programs that do little to enhance student learning.

The emergence of charter schools in the United States has paralleled movements in other countries that enacted legislation to create foundation or independent schools that operate much like charter schools in the United States – that is, they receive substantial public funding and have independent authority over curriculum and governance (Brewer and Hentschke, 2008). A key difference between these independent schools and traditional public schools is that enrolment is generally tied to parental choice. As with charter schools, the choice aspect of these schools means that their existence is directly tied to their ability to serve parents and their children.

Policy Questions

In the past several years, research has examined the comprehensive effects of charter schools on the educational system in the United States. Most of the current literature has narrowly focused on how charter schools affect achievement for students who attend these schools. However, it is important to consider the possible impact charter schools are having more broadly, including their cost effectiveness, systemic effects (e.g., effects on students who choose not to attend traditional public schools), distributional effects, both by ability and race, and any operational differences between charter and traditional public schools, which may lead to broader educational innovations. Other outcomes may be important as well, including parental satisfaction, behavioral outcomes, school safety, educational attainment

(measured by graduation rates or post-secondary attendance), and labor-market outcomes. However, the last set of outcomes has either not been addressed, or is so difficult to assess that it is hard to know what to make of the results. Therefore, this article primarily focuses on four issues:

1. What are the demographic and academic profiles of students entering charter schools, and how do charter schools affect the mix of students in traditional public schools by race/ethnicity and academic ability?
2. Are charter and traditional schools receiving comparable funding? Do charters have the same public resources to educate students as traditional schools?
3. What is the effect of charter schools on the achievement and educational attainment of their students? Do these effects differ across race/ethnicity? Are there operational features associated with achievement gains?
4. Do charters promote improved achievement in traditional public schools by introducing competition, or do they harm the performance of traditional public schools?

When summarizing the literature across these questions, we first highlight why these questions are important. Next, we establish a criterion for analyzing the quality of the research. Finally, we synthesize the findings across the literature, bearing in mind the quality of the research.

What Types of Students Do Charter Schools Serve?

As charter schools are schools of choice, it is important to examine whether they are serving the full range of the student population and whether they are doing so in integrated settings. Charter-school critics argue that charter success might be illusory if charter schools are simply recruiting the best students from traditional public schools or if they further stratify an already ethnically or racially stratified system (Cobb and Glass, 1999; Wells *et al.*, 1998). In general, these critics fear that charter schools may not only have negative consequences for the charter students who attend these schools, but if charter schools skim off high-achieving students, they may also have social and academic effects for students who remain in traditional public schools. However, proponents of charter schools argue that charter schools will improve racial integration by letting families choose schools outside of neighborhoods where housing is racially segregated (Finn *et al.*, 2000; Nathan, 1998).

Several studies have examined the racial/ethnic makeup of charter schools relative to the average racial/ethnic makeup of their surrounding districts and states (Powell *et al.*, 1997; RPP International, 2000; Miron and Nelson, 2002; Frankenberg and Lee, 2003). While this research has provided some insights, it does have some weaknesses. First, it has exclusively examined race/ethnicity, ignoring

altogether the issue of stratification by ability. Second, it has used cross-sectional snapshots of schools' enrolments, which do not permit examination of the movement of students between schools. Understanding how charter schools affect the mix of students requires a dynamic model that uses longitudinal data to examine the characteristics of students who migrate from a traditional public school to a charter and compares the students' characteristics with the distribution of students at the old and new schools.

Two recent studies examine the migration patterns of students of different race/ethnicity between traditional and charter schools. Bifulco and Ladd (2007) found that black students in North Carolina were likely to switch to charter schools with higher concentrations of black students than the traditional public schools they left. This charter-school migration increased the racial isolation of black students. Booker *et al.* (2005) also found that in California and Texas, black students generally transferred to charter schools with higher concentrations of black students. In contrast, Hispanic students in both states are likely to transfer to charter schools with a lower concentration of Hispanics than in the traditional public schools they left.

Charters may also attract students from middle-class families who would otherwise have attended private schools. Charter schools may be an important release valve for families frustrated with traditional public-school options. Toma *et al.* (2006) show that about 20% of Michigan charter school enrolments are drawn from private schools. This reengagement of parents in the public school system may improve prospects for improving public-education funding and reform efforts.

Are Charter and Traditional Schools Receiving Comparable Funding?

Charter schools are nominally entitled to funding comparable with traditional schools, but some evidence suggests that charter-school funding is lagging behind in many states (Osberg, 2006; Schneider and Tice, 2007). Average funding per pupil is about 22% lower for charter students, and some charters receive 30% less per pupil than do traditional public schools. The extent of these funding gaps is largely a function of state policies, including whether the state provides funding for facilities (or requires districts to provide facilities), whether state funds are supplemented by local resources to pay for education, and whether the state paperwork limited charter access to a large array of categorical funds.

In many states, charter schools are expected to pay for facilities out of the operating revenue, which creates large funding gaps (Krop and Zimmer, 2005). In addition, when states rely heavily upon local tax revenue to pay for education, charter schools are less likely to gain a share of these revenues – again creating large funding gaps

(Osberg, 2006). Finally, California and other states require education agencies to complete complex forms to receive categorical funding. In school districts, the cost of completing forms is spread over a large enrolment, but autonomous charter schools must complete comparable paperwork to receive funding for small numbers of students. This diseconomy of scale for charters reduces their access to categorical programs compared with traditional public schools and contributes to the reduced funding level of charter schools (Zimmer *et al.*, 2003).

The challenges of these funding gaps are gradually being recognized by states and some new policies have been introduced to address these issues. However, some argue that since one of the objectives of charter schools is greater efficiency, charter schools should prove they can provide similar or better results with fewer resources.

How Do Charter Schools Affect the Performance of Charter Students?

Over the last decade, a series of papers has emerged that has examined student achievement in charter schools in a number of states, including Arizona, Connecticut, Delaware, Florida, Illinois, Michigan, New York, Ohio, North Carolina, Texas, and Pennsylvania. In addition, a few recent studies have examined student achievement in charter schools nationally (Nelson *et al.*, 2004; Hoxby, 2004; Carnoy *et al.*, 2005). Some of this research has relied upon school-level data (Miron *et al.*, 2002; Rogosa, 2003; Greene *et al.*, 2003; LOEO, 2003; Russo, 2005) or cross-sectional comparisons of achievement in charter schools and traditional public schools at a single point in time.

A key weakness of a school-level analysis is the high degree of aggregation, which masks changes over time in the school's population of students, and variation of performance across different subjects and grades. In essence, school-level data may not pick up the nuances of school characteristics and can only provide an incomplete picture of why outcomes vary across schools. Meanwhile, point-in-time data, even at the student level, cannot account for the amount of time spent in different schools and cannot factor out the various nonschool forces at work.

Both school-level and point-in-time studies are likely to provide misleading indications of charter effectiveness due to a selection bias in charter-school enrolment. Parents choose charter alternatives because the traditional schools are not meeting their learning objectives. This parent selection process means that charter students are not representative of the pool of traditional school students, so the average difference in the performance of charter and traditional school students is a misleading indication of the effectiveness of charter schools. For example, if a charter school focuses on low-achieving students that are struggling in traditional schools, then the charter's average test score is likely to reflect the weak

preparation of its incoming students. Alternatively, if a high school charter's curriculum is built around college preparation, then the average charter test score may reflect the stronger preparation of its incoming students than that of a nearby traditional high school serving a broader mix of students. The danger is that charter effectiveness measures are easily confounded by the mix of student background and nonschool factors that lead parents to choose charter schools for their students.

Reliable measures of charter performance must assess the additional gain or value added from charter enrolment and not the historical background of students choosing charter schools. Researchers have dealt with the selection bias in charter enrolment in two ways: randomized experiments and longitudinal analyses. Both methods allow researchers to account for the amount of time a student has spent in a particular school, and both methods address differences among student populations served. Randomized experiments are often considered the gold standard in research, because, by assigning subjects randomly to the treatment condition or control condition, they ensure that differences observed later are the result of treatment rather than the result of background differences between the subject groups.

A few studies are beginning to examine oversubscribed charter schools that randomly admit students through lotteries. For instance, Hoxby and Rockoff (2004) examined four charter schools in Chicago, which provided evidence that this set of charter schools is outperforming traditional public schools. Later, Hoxby and Murarka (2007) used a similar design to evaluate 47 charter schools in New York City and found a small positive achievement effect for students attending charter schools. In addition, Mathematica Policy Research is currently in the midst of evaluating a small set of charter middle schools across a number of states for a project funded by the US Department of Education with the findings set to be released in 2008.

These studies have strong internal validity – in other words, the results have strong implications for schools included in the evaluation. One drawback of a randomized-design approach is that while it produces valid and reliable results for the set of charter schools examined, it may have limited implications for charter schools that are not part of the analysis. In other words, these studies have weak external validity. In order to conduct a randomized experiment of charter schools, researchers have relied upon randomly assigning students for schools that are oversubscribed with substantial wait lists. Therefore, these studies may have limited implications for those schools that do not have wait lists. In fact, one would expect schools with wait lists to be the best schools and it would be surprising if they had the same results as other charter schools.

When randomized designs have not been possible, or when researchers wanted to be more inclusive in their analysis of charter schools, researchers have often used a

fixed-effect approach to deal with the selection bias, which requires longitudinal student-level data. Fixed-effect approaches minimize the problem of selection bias by examining the academic gains made by individual students over time, factoring out those students' baseline achievement levels. Moreover, they permit within-student comparisons of achievement gains, examining changes in the achievement trajectories of individual students who move from traditional public schools to charter schools, or vice versa.

To date, a handful of studies have used this approach across a select number of states. In Arizona, Solmon *et al.* (2001) used the fixed-effect approach and found that students initially had lower scores, but students who remained 2–3 years in charter schools outperformed traditional public school students in reading and performed at least as well as public school students in math and possibly better (depending upon the model). Three separate studies using similar approaches found mixed results in Texas. Gronberg and Jansen (2001) found that charter schools that focus on at-risk students provided slightly more value adds than traditional public schools while non-at-risk charters provided slightly less value-adds than traditional schools. Hanushek *et al.* (2002) found negative achievement effects for Texas charter schools that were in the first few years of their operation, however, and no significant effect in reading and small positive effect in math in the long run. In a more recent study, Booker *et al.* (2007) found that in the first year of attending charter schools, student test scores drop, but the students' performance was comparable to students in traditional public schools by the second and third year in reading and math, respectively. In Florida, Sass (2006) found that charter schools initially had lower performance than traditional public schools, but eventually produced similar educational gains over time and by the fifth year, these schools were generally producing small positive effects. Bifulco and Ladd (2006) found negative reading and math effects for charter schools in North Carolina. In California, Zimmer *et al.* (2003) examined student-level data from six large districts and found that charter schools were performing on par with traditional public schools. In addition, Betts *et al.* (2006) examined charter schools exclusively in San Diego and similarly found that charter schools were performing on par with traditional public schools. In Milwaukee, Witte *et al.* (2007) found small positive effects for charter schools in math. Finally, in an anonymous district, Imberman (2007) found mixed results for charter schools with a small positive effect in math and a small negative effect in reading.

While the fixed-effect approach can be a strong control for selection bias and the papers using this approach are drawing significant attention, there are drawbacks to the fixed-effect approach as well. The fixed-effect approach provides an estimate of student achievement only for students who switch from a traditional public school to a

charter school or vice versa (switchers). Students who remain in either charter or traditional public schools for the duration of the analysis do not contribute to the estimate because the analysis does not compare their results in both contexts (charter schools and traditional public schools).

Recent work by Ballou *et al.* (2007) shows how the reliance on switchers can create two problems. First, charter schools that have high turnover rates (which may be associated with poor performance) will disproportionately contribute to the fixed-effect estimate. Second, Ballou *et al.* (2007) note that switchers may differ from nonswitchers in important ways, possibly biasing results that rely exclusively on switchers. They explore these issues by examining charter schools in Idaho using both a longitudinal student fixed-effect model, which only uses switchers, and by estimating the impact by using all tested charter and noncharter students. Exploring both approaches, they find conflicting results. The authors argue that the bias from only examining switching students in a student fixed-effect model may be greater than the self-selection bias the model attempts to correct and may imply that researchers should not exclusively rely upon fixed-effect models. However, in an interesting counterexample, Julian Betts *et al.* (2007), using data from San Diego, found that a fixed-effect approach produces math results much closer to the result of a randomized-design approach than other nonexperimental designs. In reading, random-effects model produced the best results. Overall, this suggests that researchers need to be explicit about the types of students the model is examining and what possible bias is remaining when examining charter schools through cross-sectional or longitudinal analyses, including the use of randomization.

Overall, the best of the current research has found that charter schools produce student-achievement results fairly similar to traditional public schools, but these effects vary across states. It is interesting to note that much of this research has shown that charter schools do improve over time, which implies that we may have to wait longer to draw definitive conclusions about the performance of charter schools.

Is Charter School Competition Improving the Performance of Traditional Public Schools?

While much of the existing research on charter schools has focused on student achievement effects for students who choose to attend charter schools, we argue that this focus may be too narrow. Supporters hope that charter schools can exert healthy competitive pressure on the existing K-12 educational system by giving families alternatives to traditional public schools. In fact, given that charter schools will probably never educate a substantial portion of the nation's student population, charter advocates argue that these schools may have their greatest impact through

systemic effects – the competitive effects of charter schools could improve the performance of traditional public schools and enhance the performance of students who do not attend charter schools.

The challenge in evaluating possible competitive effects is in knowing when district or school personnel will perceive a competitive threat. Do charter schools create competitive pressure when they are located near a traditional public school or when they first appear in a district? Do charter schools only create competitive pressure when they start recruiting students away from a particular school, or do they exert pressure when they capture a certain portion of students within a marketplace? In addition, the local environment may influence the competitive pressure that charter schools create. For instance, some districts may have well-developed, preexisting choice programs, including magnet schools or open enrolment policies. Moreover, some districts may be experiencing significant growth or already have overcrowded schools, in which case charter schools may act more like a release valve than a source of competitive pressure.

Nevertheless, research is developing around this issue and has made critical assumptions about the competitive process, which could affect conclusions made about the competitive effects. For example, Hoxby (2002) defines competition as where there is minimum market penetration of charter schools within a district and finds substantial positive competitive effects in Arizona and Michigan. However, Bettinger (2005), using an instrumental variable strategy, also examines competitive effects in Michigan as measured by distance and finds no effects. Using school-level data in North Carolina, Holmes *et al.* (2003) also used distance as a proxy for competition and found substantial competitive effects. In contrast, Bifulco and Ladd (2006) use student-level data in North Carolina and map out the distances of students exiting public schools to enter charter schools. Using this mapping, they analyze the effect charter schools have on traditional public schools within concentric distances of charter schools. Their analysis finds no competitive effects. Sass (2006) and Booker *et al.* (2004) also use student-level data in Florida and Texas, respectively, to examine competitive effects. Similar to Bifulco and Ladd, Sass uses concentric circles around public schools and measures whether a charter school is within these concentric circles and what proportion of total students are enrolled in charter schools. Using these approaches, Sass (2006) finds small positive competitive effects in Florida. Booker *et al.* (2004) uses two approaches, which find consistent competitive effects. First, like Hoxby, the authors use market-penetration measure at the district level. Second, they also use a campus-level market-penetration measure, which is defined by the percentage of students at a particular campus that exits the school to go to a charter school. They find competitive effects across both measures.

In a recent paper, Buddin and Zimmer (2005b) examine the competitive effects of charter schools using both survey and longitudinal student-level data in California. Examining survey responses from traditional public school principals about the changes in school operations, the analysis by the authors suggests that charter schools have no measurable impact on the performance or operation of traditional public schools. The authors go on to examine the student-achievement data, in which they use metrics of distance to the nearest charter schools along with the number of charter schools and the share of students attending a charter school within 2.5 miles of a traditional public school, while controlling for the level of preexisting competition and also using various measures of competition. Again, the authors found no evidence that charter schools create a competitive effect.

Across these studies, the authors generally found limited evidence of a competitive effect from charter schools. However, that could be explained by the fact that charter schools generally represent a low share of students in districts or states or by the fact that charter schools, at least in growing districts, may act as a release valve rather than a competitive pressure. It is possible that a broader implementation of charters would exert pressure on traditional public schools to improve their performance.

Conclusions

The charter movement grew out of a hope that by providing greater autonomy to schools, charter schools would be able to cut through bureaucratic frustrations and offer innovative, efficient, and effective educational programs, provide new options to families, and promote healthy competition for traditional public schools. However, critics feared that charter schools would drain scarce resources from the existing public education system, create greater stratification both by ability and race/ethnicity, and provide no real academic advantages to students attending these schools.

The current research has not shown charter schools to be a panacea for challenges facing the educational system that advocates hoped for, but expanding charter numbers and enrolments indicate that these schools may be meeting some other needs of parents, such as a safer educational environment or a wider array of instruction. Therefore, charter schools may be an initiative that should be given more time to show results, especially given that these schools are using fewer public resources.

See also: Competition and Student Performance; Educational Privatization; Empirical Research Methods in the Economics of Education; The Economics of Parental Choice.

Bibliography

- Alvarez, D. T. (2003). *Determinants of Charter Law Adoption and Law Strength across States*. Unpublished Honors Thesis University of Florida.
- Ballou, D., Teasley, B., and Zeidner, T. (2007). Charter schools in Idaho. In Berends, M., Springer, M. G., and Walberg, H. (eds.) *Charter School Outcomes*, pp 221–241. New York: Erlbaum.
- Bettinger, E. P. (2005). The effect of charter schools on charter students and public schools. *Economics of Education Review* 24, 133–147.
- Betts, J. and Hill, P. T. (2006). *Key Issues in Studying Charter Schools and Achievement: A Review and Suggestions for National Guidelines*. NCSRP White Paper Series No. 2. Seattle, WA: National Charter School Research Project.
- Betts, J. R., Rice, L. A., Zau, A. C., Tang, E., and Koedel, C. R. (2006). *Does School Choice Work? Effects on Student Integration and Achievement*. San Francisco, CA: Public Policy Institute of California.
- Betts, J. R., Tang, Y. E., and Zau, A. C. (2007). Madness in the methods? A critical analysis of popular methods of estimating the effect of charter schools on student achievement. *Paper Presented at the Annual Meetings of the American Educational Research Association*, Chicago, IL.
- Bifulco, R. and Ladd, H. (2006). The impact of charter schools on student achievement: Evidence from North Carolina. *Journal of Education Finance and Policy* 1, 778–820.
- Bifulco, R. and Ladd, H. (2007). Race and charter schools: Evidence from North Carolina. *Journal of Policy Analysis and Management* 26, 31–56.
- Booker, K., Gilpatric, S., Gronberg, T. J., and Jansen, D. W. (2007). Charter school performance in Texas. *Journal of Public Economics* 91, 849–876.
- Booker, K., Gukoatric, S., Gronberg, T. J., and Jansen, D. W. (2004). *The Effect of Charter Schools on Traditional Public Students in Texas: Are Children Who Stay behind Left behind?* College Station, TX: Private Enterprise Research Center Working Paper/Texas A & M University.
- Booker, K., Zimmer, R., and Buddin, R. (2005). *The Effect of Charter Schools on School Peer Composition*. Working Paper WR-306-EDU. Santa Monica, CA: RAND Corporation.
- Brewer, D. and Hentschke, G. (2008). The global phenomenon of publicly-financed, privately-operated schools. In Berends, M., Springer, M., Ballou, D., and Walberg, H. (eds.) *Handbook of Research on School Choice*, pp 227–246. New York: Erlbaum.
- Buddin, R. and Zimmer, R. W. (2005a). A closer look at charter school student achievement. *Journal of Policy Analysis and Management* 24, 351–372.
- Buddin, R. and Zimmer, R. W. (2005b). *Is Charter School Competition in California Improving the Performance of Traditional Public Schools?* Working Paper WR-297-EDU. Santa Monica, CA: RAND Corporation.
- Carnoy, M., Jacobsen, R., Mishel, L., and Rothstein, R. (2005). *The Charter School Dust-Up: Examining the Evidence on Enrollment and Achievement*. Washington, DC: Economic Policy Institute.
- Cobb, C. D. and Glass, G. V. (1999). Ethnic segregation in Arizona charter schools. *Education Policy Analysis Archives* 7, 1–36.
- Corcoran, S. and Stoddard, C. (2007). The political economy of school choice: Support for charter schools across states and districts. *Journal of Urban Economics* 62, 27–54.
- Fasarelli, L. D. (2003). *The Political Dynamics of School Choice*. New York: Palgrave MacMillan.
- Finn, C. E., Manno, B. V., and Vanourek, G. (2000). *Charter Schools in Action: Renewing Public Education*. Princeton, NJ: Princeton University Press.
- Finnigan, K., Adelman, N., Anderson, L., et al. (2004). Evaluations of the public charter schools program. *Final Report*. Washington, DC: SRI International.
- Florida Department of Education (2004). *Florida Charter Schools 2002–2003 Annual Report*. Tallahassee, FL: Florida Department of Education.
- Frankenberg, E. and Lee, C. (2003). Charter schools and race: A lost opportunity for integrated education. *Education Policy Analysis Archives* 11(32). <http://epaa.asu.edu/epaa/v11n32/> (accessed June 2009).
- Gleason, P. (2007). Evaluating charter school impacts using admission lotteries: Lessons from the field. *Paper presented at the annual meetings of the American Educational Research Association*, April 11, Chicago, IL.
- Greene, J. P., Forster, G., and Winters, M. A. (2003). *Apples to Apples: An Evaluation of Charter Schools Serving General Student Populations*. Education Working Paper No. 1. New York: Manhattan Institute.
- Gronberg, T. J. and Jansen, D. W. (2001). *Navigating Newly Chartered Waters: An Analysis of Texas Charter School Performance*. Austin, TX: Texas Public Policy Foundation.
- Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2002). *The Impact of Charter Schools on Academic Achievement*. Stanford, CA: Hoover Institute, unpublished manuscript.
- Hassel, B. C. (1999). *The Charter School Challenge: Avoiding the Pitfalls, Fulfilling the Promise*. Washington, DC: Brookings Institution Press.
- Henderson, V., Miezowski, P., and Suavageau, Y. (1978). Peer-group effects and educational production functions. *Journal of Public Economics* 10, 97–106.
- Holmes, G. M., DeSimone, J., and Rupp, N. (2003). Does school choice increase school quality? Working Paper 9683. Cambridge, MA: National Bureau of Economic Research.
- Hoxby, C. M. (2002). How school choice affects the achievement of public school students. In Paul, H. (ed.) *Choice with Equity*, pp 141–177. Stanford, CA: Hoover Press.
- Hoxby, C. M. (2004). *A Straightforward Comparison of Charter Schools and Regular Public Schools in the United States*. HIER Working Paper. Cambridge, MA: National Bureau of Economic Research.
- Hoxby, C. M. and Murarka, S. (2007). *Charter Schools in New York City: Who Enrolls and How They Affect Their Students' Achievement*. Cambridge, MA: National Bureau of Economic Research.
- Hoxby, C. M. and Rockoff, J. E. (2004). *The Impact of Charter Schools on Student Achievement*. Working Paper. Cambridge, MA: Harvard University.
- Imberman, S. (2007). *Achievement and Behavior in Charter Schools: Drawing a More Complete Picture*. Working Paper No. 142. New York, NY: National Center for the Study of Privatization in Education.
- Krop, C. and Zimmer, R. (2005). Charter school type matters when examining funding and facilities: Evidence from California. *Education Policy Analysis Archives* 13(50). <http://epaa.asu.edu/epaa/v13n50> (accessed June 2009).
- LOEO (2003). *Community Schools in Ohio: Final Report on Student Performance, Parent Satisfaction, and Accountability*. Columbus, OH: Legislative Office of Education Oversight.
- Miron, G., Cullen, A., Applegate, B., and Farrell, P. (2007). *Evaluation of the Delaware Charter School Reform: Final Report*. Kalamazoo, MI: The Evaluation Center of Western Michigan University.
- Miron, G. N. and Nelson, C. (2002). *What's Public about Charter Schools?* Thousand Oaks, CA: Corwin Press.
- Miron, G., Nelson, C., and Risley, J. (2002). *Strengthening Pennsylvania's Charter School Reform: Findings from the Statewide Evaluation and Discussion of Relevant Policy Issues*. Kalamazoo, MI: The Evaluation Center/Western Michigan University.
- Nathan, J. (1998). Controversy: Charters and choice. *The American Prospect* 9 November–December. <http://www.prospect.org> (accessed June 2009).
- Nelson, F. H., Rosenberg, B., and Van Meter, N. (2004). *Charter School Achievement on the 2003 National Assessment of Educational Progress*. Washington, DC: American Federation of Teachers.
- Osberg, E. (2006). Charter school funding. In Paul, T. H. (ed.) *Charter Schools against the Odds: An Assessment of the Koret Task Force on K-12 Education*, pp 45–69. Stanford, CA: Hoover Institution Press.
- Powell, J., Blackorby, J., Marsh, J., Finnegan, K., and Anderson, L. (1997). *Evaluation of Charter School Effectiveness*. Menlo Park, CA: SRI International.
- Rogosa, D. (2003). *Student Progress in California Charter Schools, 1999–2002*. Stanford, CA: Stanford University Working Paper.
- RPP International (2000). *The State of Charter Schools: National Study of Charter Schools Fourth-Year Report*. Washington, DC: Office of

- Educational Research and Improvement, US Department of Education.
- Russo, A. (2005). *A Tough Nut to Crack in Ohio: Charter Schooling in the Buckeye State*. Washington, DC: Progressive Policy Institute.
- Sass, T. R. (2006). Charter schools and student achievement in Florida. *Journal of Education Finance and Policy* **1**, 91–122.
- Schneider, M. and Tice, P. (2007). School choice: What the NCES can and cannot tell us. In Berends, M., Springer, M. G., and Walberg, H. (eds.) *Charter School Outcomes*, pp 267–281. New York: Erlbaum.
- Solmon, L., Paark, K., and Garcia, D. (2001). *Does Charter School Attendance Improve Test Scores? The Arizona Results*. Phoenix, AZ: Goldwater Institute Center for Market-Based Education.
- Summers, A. and Wolfe, B. (1977). Do schools make a difference? *American Economic Review* **67**, 639–652.
- Toma, E. F., Zimmer, R., and Jones, J. T. (2006). Beyond achievement: Enrollment consequences of charter schools in Michigan. In Gronberg, T. and Jansen, D. (eds.) *Improving School Accountability: Checks or Choice*, vol. 14, pp 241–255. Amsterdam: Elsevier Advances in Applied Economics.
- Vanourek, G. (2005). *State of the Charter School Movement: Trends, Issues, and Indicators*. Washington, DC: Charter School Leadership Council.
- Weiherr, G. R. and Tedin, K. L. (2002). Does choice lead to racially distinctive schools? Charter schools and household preferences. *Journal of Policy Analysis and Management* **21**, 79–92.
- Wells, A. S., Artilles, L., Carnochan, S., et al. (1998). *Beyond the Rhetoric of Charter School Reform: A Study of Ten California School Districts*. Los Angeles, CA: UCLA Charter School Study.
- Witte, J. F., Weimer, D., Shober, A., and Schlomer, P. (2007). The performance of charter schools in Wisconsin. *Journal of Policy Analysis and Management* **26**, 574–575.
- Wong, K. K. and Shen, F. X. (2004). Political economy of charter school funding formulas: Exploring state-to-state variation. In DeMoss, K. and Wong, K. (eds.) *Money, Politics, and the Law: Intersections and Conflicts in the Provision of Educational Opportunity*. American Education Finance Association 2004 Yearbook, pp 171–196. Larchmont, NY: Eye on Education.
- Zimmer, R. W. (2003). A new twist in the educational tracking debate. *Economics of Education Review* **22**, 307–315.
- Zimmer, R. W. and Buddin, R. (2006). Charter school performance in urban districts. *Journal of Urban Economics* **60**, 307–326.
- Zimmer, R. and Buddin, R. (2007). Getting inside the black box: Examining how the operation of charter schools affects performance. *Peabody Journal of Education* **82**, 231–273.
- Zimmer, R., Buddin, R., Chau, D., et al. (2003). *Charter School Operations and Performance: Evidence from California*. Santa Monica, CA: RAND Corporation, MR-1700.
- Zimmer, R. W. and Toma, E. F. (2000). Peer effects in private and public schools across countries. *Journal of Policy Analysis and Management* **19**, 75–92.

Further Reading

- Berends, M., Springer, M., and Walberg, H. (eds.) *Charter School Outcomes*. New York: Erlbaum.
- Betts, J. R. and Loveless, T. (eds.) (2005). *Getting Choice Right: Ensuring Equity and Efficiency in Education Policy*. Washington, DC: Brookings Institution Press.
- Finn, C. E., Hassel, B. C., and Speakman, S. (2005). *Charter School Funding: Inequity's Next Frontier*. Washington, DC: Thomas B. Fordham Institute.
- Gill, B., Timpane, P. M., Ross, K. E., and Brewer, D. J. (2001). *Rhetoric versus Reality: What We Know and What We Need to Know about Vouchers and Charter Schools*. Santa Monica, CA: RAND Corporation, MR-1118.
- Hoxby, C. (ed.) (2003). *The Economics of School Choice*. Chicago, IL: University of Chicago Press.
- Kolderie, T. (2004). *Creating the Capacity for Change: How and Why Governors and Legislatures Are Opening a New-Schools Sector in Public Education*. Bethesda, MD: Education Week Press.
- Zimmer, R. and Buddin, R. (2006). *Making Sense of Charter Schools: Evidence from California*. Santa Monica, CA: RAND Corporation, OP-157.

Relevant Websites

- <http://www.gatesfoundation.org> – Bill and Melinda Gates Foundation.
- <http://www.brookings.edu> – Brookings Institute, Charter Schools.
- <http://www.edreform.com> – Center for Education Reform.
- <http://evans.washington.edu> – Center for Reinventing Public Education.
- <http://www.edweek.org> – Education Week, Charters and Choice.
- <http://www.edexcellence.net> – Fordham Foundation.
- <http://www.hoover.org> – Hoover Institute.
- <http://www.ncspe.org> – National Center for the Study of Privatization in Education.
- <http://www.ncsc.info> – National Charter School Clearinghouse.
- <http://www.nea.org> – National Education Association.
- <http://www.rand.org> – RAND Corporation.
- <http://www.uscharterschools.org> – U.S. Charter Schools.
- <http://www.vanderbilt.edu> – Vanderbilt's National Center on School Choice.

The Economics of Parental Choice

R Chakrabarti, Federal Reserve Bank of New York, New York, NY, USA

J Roy, Georgetown University, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Charter schools – The publicly funded but privately managed and operated schools in the United States. They are free from many of the rules and regulations that apply to other public schools in exchange for accountability, and the exact rules governing this are set forth in each school's charter.

Sorting/cream-skimming – The phenomenon whereby private schools attract the most wealthy, best, and brightest students from the public schools, thus skimming the public schools off their cream.

Utility function – In economics, utility measures the satisfaction from consumption of goods. A utility function summarizes preferences of a consumer in terms of how much utility he or she gets from consuming the goods in the utility function.

Utility maximization – A concept in economics that, when making a decision, a consumer attempts to get the greatest value possible from the amount of money that he/she wishes to expend. This exercise yields the demand of the consumer for different goods.

Vouchers – A school voucher is a certificate by which parents are given the ability to pay for the education of their children at a private school of their choice rather than the public school to which they were assigned. This allows families to select the public or private schools of their choice and have all or part of the tuition paid.

One of the most important decisions that parents make regarding their children is the choice of their schools. A good education is often the cornerstone of a person's future life, being the foundation for his or her career. Most parents take care to choose a suitable school for their children, expending considerable resources in terms of time and money in the process. This short article reviews the economic aspects of parental choice of schools. We analyze the basic economic theory underlying a family's choice of schools for its children and then examine various dimensions of it, particularly as it relates to the organization of schools and provision of education in the real world. We briefly review the extensive literature that has come

up in recent years, partly fueled by a significant increase in school-choice options for students, and highlight contributions that have significantly increased our understanding of this important issue.

The Basic Economics of Parental Choice

It is instructive to begin with a review of the basic tenets underlying the economic theory of parental choice. This approach is rooted in the traditional paradigm of utility maximization, where school quality is considered a good, on par with other goods and services that families demand and consume.

In this framework, families with school-age children decide how to allocate their resources (wealth or income) among alternative goods, one of which is school quality. Formally, the family's preferences can be thought of as being represented by a utility function, where the family gets utility from school quality and a numeraire good (or money). School quality can be thought to be represented by a composite of school-related characteristics such as: academic performance, school demographics, school location/proximity, and school atmosphere. The family tries to maximize its utility given the income at its disposal (budget constraint) and various school types (represented by the above characteristics) and corresponding prices available to it. (For private schools, prices can be thought of as tuition. For various public school qualities, prices can be thought of as the opportunity cost of forgoing the next best option.) Under fairly standard assumptions regarding the shape and properties of the family's utility function, we can solve the maximization problem which yields the family's demand for school. (These assumptions are necessary to solve the model mathematically; however, the assumptions are fairly intuitive and likely to hold in real life. Formally, these include constraints on the shape of the utility function (nonsatiation and concavity).)

Given the utility function or preferences of the family, the demand for schooling depends on several factors, just like demand for other goods and services. These factors include the price or cost of different types of schooling or school quality, the family's income (or more broadly, resources at its disposal), and the prices of other goods that the family consumes (or might want to consume). In other words, the demand for school quality will change

with changes in these factors, holding the preferences of the family constant. For example, a rise in the price of other essential items that the family consumes may force it to cut back on its investments in schooling. (This may take different forms – e.g., a family might be forced to transfer its child to a school with lower costs (in terms of property taxes and/or tuition), or transfer to a school closer to home to save on transportation costs, or cut back on other investments complementary to schooling (such as computers at home), etc.) Of course, inherent preferences of families toward school quality are also likely to be different, and families who are more strongly committed to school quality (or a component of school quality) are more likely to forgo other expenses (or other components of school quality) in order to acquire it. Some recent studies have argued that such preferences might be systematically related to socioeconomic and demographic characteristics of households – these characteristics include parental education and incomes, race/ethnicity – this literature is reviewed later.

Choice among Alternative Types of Schools

One of the most important choices facing parents is whether to send their child to a public school or to a private school. The main distinguishing feature, from an economic point of view, is that public schools are free while private schools charge tuition and, hence, are costly to attend. Therefore, families are likely to send their children to private schools only if they believe that the benefits of private schooling outweigh the price differential. These benefits might be related to academic proficiency, or other noncognitive or even nontangible factors (discipline, safety, focus on moral values, etc.) In the United States, in particular, a significant percentage of the private schools are run by Catholic organizations or other religious denominations, and religion often plays an important role in family's choice of religious private schools.

Epplé and Romano (1998, 2002), Hoyt and Lee (1998), and Chakrabarti (2006, 2008) model a family's choice of private schools vis-a-vis public. In these models (with the exception of Hoyt and Lee (1998) where students differ only by income), students differ both in terms of family income – which affects their capacity to pay tuition and attend a private school – and in terms of inherent ability. The common finding in these theoretical models is that higher-income families and those with more able children will send their kids to private schools, and vice versa. Thus this leads to sorting by income and ability – children with higher family incomes and higher ability attend private schools, while children with lower family incomes and lower ability attend public schools.

Empirical Evidence on What Parents Care about

There is by now a large body of literature that investigates the characteristics of schools that parents prefer. These studies can be broadly divided into two parts: studies which survey individual families and ask them about their preferences for different school and schooling-related characteristics, and studies that look at actual schooling choices of parents and students. The findings from both strands of literature are discussed below. While the findings from the survey literature are interesting and informative, these suffer from an important caveat – the evidence is based on self-reporting by families. As a result, survey responses may often be associated with intended or unintended misrepresentation or parents not being completely honest with their preferences. Therefore, it is useful to compare findings of the survey literature with that of the observed choice literature. As can be seen from the discussion below, the findings from the two strands of literatures differ in important respects.

School Demographics

One of the most consistent results of the parental choice literature is that when surveyed, parents do not exhibit much preference toward student demographics and often display indifference toward the demographic composition of schools (Schneider *et al.*, 1998; Howell, 2004; Elacqua *et al.*, 2006; Weiher and Tedin, 2002; Williams *et al.*, 1983).

In contrast, observed choice behavior reveals significant parent preferences toward school demographics (Henig, 1990; Saporito and Lareau, 1999; Lankford and Wyckoff, 1992; Glazerman, 1997; Glazerman, 1998; Lankford and Wyckoff, 2000; Schneider and Buckley, 2002; Weiher and Tedin, 2002; Saporito, 2003; Hastings *et al.*, 2006; Elacqua *et al.*, 2006).

This finding is surprisingly consistent in this strand of literature in spite of the fact that the studies are based on different regions and time periods and the methodologies used to elicit choice behavior are also different. Schneider and Buckley (2002) study the search behavior of parents as they browsed information on the website <http://www.DCSchoolSearch.com>. This website provides detailed information (such as demographics, test scores, location, and academic programs) on public schools (both traditional and charter) in Washington DC. Elacqua *et al.* (2006) investigate search behavior of parents in the Metropolitan Region of Santiago, Chile and also compare these revealed preferences of parents with their survey responses relating to what they want from schools. Weiher and Tedin (2002) investigate the choice behavior of charter school parents in Texas and also compare them with expressed preferences of these same households. Hastings *et al.* (2006) analyze school-choice data from the Charlotte-Mecklenburg School

District in North Carolina; Henig studies enrolment patterns in magnet schools in Montgomery county, Maryland; Lankford and Wyckoff (1992) as well as Lankford and Wyckoff (2000) study school-choice decisions of parents in New York; Glazerman (1998) uses data from a public school choice program in Minneapolis, Minnesota; Saporito and Lareau (1999) use data from a large northeastern urban school district in the United States; and Saporito (2003) analyzes magnet school application data from Philadelphia.

In several studies in the US, researchers find that whites have a strong tendency to avoid minorities when choosing schools (Lankford and Wyckoff, 1992; Lankford and Wyckoff, 2000; Saporito and Lareau, 1999; Saporito, 2003). Some studies find this kind of behavior to be specific to whites (Lankford and Wyckoff, 2000; Saporito and Lareau, 1999), while others find that minorities also have own-race preferences (Weiher and Tedin, 2002; Glazerman, 1998; Hastings *et al.*, 2006). Schneider and Buckley (2002) find that parents generally seek out schools with fewer black students, with the result being stronger for more-educated parents. Research also shows parent preferences for school demographics other than race: Elacqua *et al.* (2006) find that 87% of parents in their sample only consider schools with student demographics similar to their own, as measured by parent education and socioeconomic status; Saporito (2003) finds that nonpoor parents are much more sensitive to poverty levels than poor parents when choosing schools; and Lankford and Wyckoff (1992) find that parents are very sensitive to income levels when choosing schools.

Academic Performance

Several studies of survey responses find academics to be the single most important factor for parents choosing schools (Williams *et al.*, 1983; Kleitz *et al.*, 2000; Howell, 2004; Elacqua *et al.*, 2006; some of these survey papers use criteria that likely encompass academics such as education quality (Kleitz *et al.*, 2000), and quality of teaching (Howell, 2004), thus, this list may be somewhat generous), or of generally high importance (Weiher and Tedin, 2002; Vanourek *et al.*, 1998; Greene *et al.*, 1998). Conversely, while some studies of observed behavior affirm the importance of academics (Saporito, 2003; Schneider and Buckley, 2002; Hastings *et al.*, 2006; Lankford and Wyckoff, 1992), others do not (Elacqua *et al.*, 2006; Saporito and Lareau, 1999; Weiher and Tedin, 2002; Glazerman, 1998).

Another important finding of studies relying on revealed preferences of parents is that parents give more weight to demographics than academics. Saporito (2003) finds that parent choices are significantly motivated by academic concerns, but that such concerns do not account for white families seeking schools with fewer nonwhite students. Schneider and Buckley (2002) find that parent Internet search behavior does reveal preferences for

better-performing schools; however, such preferences are less important than student demographics, both for parents with and without some college education. Hastings *et al.* (2006) find percent black to be a better predictor of school choice than school test scores. Elacqua *et al.* (2006) find that public school, private voucher, and private non-voucher parents, all construct school-choice sets with widely varying test scores; however, most parents do not choose the highest-performing school in their choice set, but nearly all parents only consider schools with demographics similar to their own. Saporito and Lareau (1999) “could not find evidence of white families’ decisions being linked to a range of school characteristics other than race” (p. 426). Weiher and Tedin (2002) find that most students transfer to schools with worse test scores, and that stating test scores as most important had no positive impact on the likelihood of transferring to a better-performing school. Lankford and Wyckoff (1992) find that parents do choose private schools based on academic performance, but that race may be a stronger predictor. Glazerman (1998) finds that test scores “were only important when middle and high ability choosers had a school with many low achieving potential classmates in their choice set” (p. 17).

Although it is certainly important to consider aggregate parent preferences, another question of significant interest is how parent preferences for academics vary across demographic groups. Some survey research finds little or no differences in preferences for academics across demographic groups (Kleitz *et al.*, 2000; Howell, 2004; Elacqua *et al.*, 2006), while other such research finds considerable differences. Williams *et al.* (1983) find that academic standards constitute the most important concern for private school parents and public school parents that considered other schools; however, it is somewhat more important for private school parents. Weiher and Tedin (2002) find that test scores are ranked most important for whites, second most important for blacks, and fourth most important for Hispanics – this holds true for parents with children in both at-risk and non-at-risk schools. Schneider *et al.* (1998) find that black parents and those with high school but no college education value test scores considerably more than parents of other racial groups and those that have attended college. Results from modeled and observed parent choices also show considerable differences, although fewer of these studies address the question. Schneider and Buckley (2002) find that parents with any level of college education value test scores much higher than those with no college. Elacqua *et al.* (2006) find the opposite: the chances of parents choosing the highest-performing school in their choice set decreases with education. Hastings *et al.* (2006) find that preference for higher-performing schools is strongly associated with higher income – both for whites and nonwhites – but overall, blacks seem to value test scores somewhat more than whites. Studying the same

choice program, Hastings *et al.* (2007) find that students whose parents had strong preference toward academic performance experienced gains in test scores as a result of going to the choice school, while students whose parents gave lower weights to academic performance experienced losses.

Location

Survey responses regarding the importance of location are mixed. Some show location to be very important with little differences across demographic groups, while others find it to be relatively unimportant but with considerable differences across demographic groups. Bridge and Blackman (1978) find that location was the single most important deciding factor for the majority of parents in their sample for all 5 years of the Alum Rock voucher demonstration. However, its importance decreased over the years as parents became more familiar with their school options. Elacqua *et al.* (2006) find location to be quite important to parents in choosing schools, being ranked first, second, or third out of eight factors for parents of all education levels and school types. Conversely, Kleitz *et al.* (2000) and Howell (2004) find location ranked fairly low across various demographic groups. However, both papers find more disadvantaged groups to value location higher. Williams *et al.* (1983) find distance to be of high interest to public school parents that did not consider other schools, moderate interest to public school parents that did consider other schools, and of very little interest to private school parents. Weiher and Tedin (2002) find that importance of location is ranked fifth out of six for whites and Hispanics and fourth for blacks.

Results from most of the literature relying on revealed preference show moderate to strong preferences for location. Saporito and Lareau (1999) find that both whites and blacks tend to choose schools close to their homes, but that whites are often willing to travel further to attend schools with higher proportions of white students. Hastings *et al.* (2006) find that families value location highly, but those with strong preferences for academics are generally willing to tolerate larger distances. Schneider and Buckley (2002) find that in parent Internet search behavior, location was the second most sought-after information after school demographics. However, they note that parents who know their city neighborhoods well may be using location as a proxy for other school characteristics. In contrast to the above literature, Glazerman (1998) finds that only 26% of the parents in his sample chose the school closest to their home, indicating a fairly low importance for location.

School Atmosphere: Safety, Discipline, Values, Class Size, and Friends

School atmosphere (i.e., safety, discipline, values, class size, and number of friends) is another factor that parents

consider when choosing schools. Since many variables associated with school atmosphere are relatively hard to observe (with the exceptions of safety and class size), most of the evidence on this factor comes from the survey literature.

Howell (2004) finds that discipline, safety, and order are ranked the second most important factor out of nine after quality of teaching and this factor is rated equally high by parents with children in performing schools and those with children in underperforming schools. Kleitz *et al.* (2000) find that safety is ranked third by all racial and socioeconomic groups, but is valued quite differently. Blacks are much more concerned with safety than Anglos, and Hispanics much more than blacks. Further, low- and moderate-income parents value safety similarly but consider it much more important than do high-income parents. They argue that variance in preferences for safety reflects differences in past school environments experienced by ethnic and income groups.

Williams *et al.* (1983) find that discipline and values/religion are most important for private school parents, less important for public school parents that considered other schools, and even less so for public school parents who did not consider other schools. Schneider *et al.* (1998) find that minority and less-educated parents value discipline much more than white and higher-educated parents, while white parents and those with higher education consider values and diversity much more than others.

Howell (2004) finds friends at school is ranked lowest both by parents with children in underperforming schools and those with children in performing schools and given nearly identical ratings. Of the five factors they consider, Kleitz *et al.* (2000) also find friends ranked lowest for Anglos, blacks, Hispanics, as well as for low-, moderate-, and high-income parents. However, they find friends are much more important for Hispanics than for blacks and Anglos, and much less important as income rises.

Kleitz *et al.* (2000) find that class size is ranked second and is of similar importance for all racial/ethnic and income groups. Howell (2004) finds class size is ranked third by all groups of parents, although those from underperforming schools rate it somewhat higher.

Does Parental Choice Vary by Income and Ability of Parents? Or Is There Evidence of Sorting by Income and Ability?

Apart from the above literature, there is also a rich empirical literature that investigates whether parental choice behavior varies with the income and ability of parents. Hsieh and Urquiola (2006) study the nationwide voucher program in Chile that was introduced in 1981 and find that wealthier families and families with higher education were more likely to avail of vouchers. In other words, they find that vouchers led to sorting by income and ability in

Chile. In the context of a national, means-tested school voucher program in the United States, Campbell *et al.* (2005) show that the chances of both application and actual voucher takeup increased with mother's education and decreased with family income. They argue that the pattern for income may be a reflection of the program feature that vouchers awarded to higher-income families were smaller. They also find that religious motivations (e.g., whether or not attended church services) played an important role in choice decisions. Peterson *et al.* (1999) find that voucher students in Cleveland had higher mother's education than public school students. Eppler *et al.* (2002) also find evidence of both sorting by income and ability. Howell *et al.* (2006) and Howell (2004) find that higher-income families were more likely to take a voucher in New York City, while Howell *et al.* (2006) find that low-income families were more likely to do so in Dayton, with no difference observed in Washington, DC. In a survey of previous literature, Levin (1998) finds that choosers were more advantaged both educationally and economically than nonchoosers.

Hoxby (2003) finds no evidence of cream-skimming in the context of charter school competition in the United States. Witte and Thorn (1996) and Chakrabarti (2006) study the Milwaukee voucher program. The former study finds that voucher applicants had lower income and higher mother's education than their counterparts in the Milwaukee public schools. Chakrabarti (2006) finds no difference in income between voucher applicants and nonapplicants in Milwaukee, but finds that the applicants had higher mother's education and parental involvement in education. Chakrabarti (2006) argues that the findings reflect the importance of voucher design – voucher design matters as far as student sorting is concerned. She argues that the program feature that vouchers could not be topped up (vouchers were to be taken as full payment of tuition and hence voucher-availing parents did not have to make any tuition payments) coupled with the absence of private school selection (private schools were required to accept students randomly if oversubscribed) likely resulted in the absence of sorting by income. Studying public school choice in New Zealand, Fiske and Ladd (2000) find that families who took advantage of choice tended to opt for higher socioeconomic status schools and the choice program resulted in minorities getting concentrated in low socioeconomic status schools. For a comprehensive review of this sorting literature, see McEwan (2000, 2004) and Howell *et al.* (2006).

The Role of Information

While it is important to consider, as the previous discussion does, what parents care about in schools, it is also

important to consider whether they have sufficient information about schools and school characteristics. Even strong preferences do not guarantee correct and educated parental choices unless and until they have adequate information. Two questions are relevant here. First, do parents have adequate information about their schools? Second, does better information matter? That is, does it entail different choices and better outcomes for families?

Drawing from a telephone survey of public school parents in the state of Massachusetts, Howell (2004) assesses the information that parents have about the statuses of their schools under the federal No Child Left Behind (NCLB) law, as well as general characteristics of their schools. Adequate knowledge of the statuses of their children's schools under NCLB is absolutely essential for parents to make informed choices. For example, low-performing statuses of schools make their students eligible for public school choice and supplemental services (such as tutoring), but parents can avail of these only if they are aware of their school's status. Howell (2004) finds that although parents assert that they are familiar with NCLB, most parents in low-performing schools are not aware of their schools' statuses and, consequently, the fact that they qualify for additional services. He also finds that minority, disadvantaged, and less-educated parents systematically had less information than white, more-advantaged, and more-educated parents.

In relation to the second question, Hastings and Weinstein (2007) analyze the effect of providing transparent information to parents (on school-level academic achievement) on their school choices and their child's subsequent academic outcomes. They find that providing transparent information resulted in parents making substantially better choices in terms of choosing higher-performing schools and attending these schools also resulted in significant increases in the children's test scores.

In contrast, Mizala and Urquiola (2007) find in the context of Chile that information on school effectiveness does not make much dent on school markets. They find that being identified as high-performing schools, where this information is widely disseminated, does not have much effect on the schools' enrolment, socioeconomic composition, or tuition. The differences in results in the two studies discussed above may be because they relate to very different schooling markets and school systems. Mizala and Urquiola (2007) argue that while parents may care about school effectiveness they might care about other factors more such as peer composition, or this information may not be new to them in the sense they may have already deduced the effectiveness of the schools on their own.

Conclusions

In this brief article, we have surveyed the evidence on how parents choose schools for their children.

The literature focuses on different aspects of this choice, highlighting features of schools such as composition of its student population, its academic performance, location or proximity, and school atmosphere. Somewhat surprisingly, and despite their professed indifference about it, many parents are observed to pick up schools based on similarities in demographic characteristics. This is in contrast to the role played by academic performance, which is often found to be of secondary importance (although some studies find performance to be the main concern). Parents are also found to be sensitive about the location of schools, and prefer schools with positive atmosphere as reflected in school safety, discipline, and class size.

The literature also finds that there is interesting heterogeneity in preferences of parents across income and racial groups. Several studies have also documented that more-educated and higher-income families disproportionately avail of school-choice options, thereby increasing the possibility of segregation. However, the design of these particular choice programs plays a crucial role in this respect. It is also instructive that better and more transparent information has the potential to play a significant role in families' choices, at least under certain circumstances. These findings have important implications for school-choice policy and its design, and also calls for better transmission of information to parents, as this might, at least in some circumstances, lead to better choices by them.

See also: Competition and Student Performance; Educational Privatization; The Economics of Catholic Schools; The Economics of Charter Schools; The Economics of School Accountability; The Efficacy of Educational Vouchers; Tiebout Sorting and Competition.

Bibliography

- Bridge, G. and Blackman, J. (1978). *A Study of Alternatives in American Education*, vol. 4. Santa Monica, CA: Rand Corporation.
- Campbell, D. E., West, M. R., and Peterson, P. E. (2005). Participation in a national, means-tested school voucher program. *Journal of Policy Analysis and Management* **24**(3), 523–541.
- Chakrabarti, R. (2006). Do vouchers lead to sorting under random private school selection? Evidence from the Milwaukee voucher program. *Working Paper*, Federal Reserve Bank of New York.
- Chakrabarti, R. (2008). Impact of voucher design on public school performance: Evidence from Florida and Milwaukee voucher programs. *Federal Reserve Bank of New York Staff Report No. 315*.
- Elacqua, G. and de Gobierno, E. (2006). Enrollment practices in response to vouchers: Evidence from Chile. *National Center for the Study of Privatization in Education*. Teachers College, Columbia University.
- Epple, D., Figlio, D., and Romano, R. (2004). Competition between private and public schools: Testing stratification and pricing predictions. *Journal of Public Economics* **88**(7–8), 1215–1245.
- Epple, D., Figlio, D., and Romano, R. (2004). Competition between private and public schools: Testing stratification and pricing predictions. *Journal of Public Economics* **88**(7–8), 1215–1245.
- Epple, D. and Romano, R. (1998). Competition between private and public schools, vouchers and peer group effects. *American Economic Review* **62**(1), 33–62.
- Epple, D. and Romano, R. (2002). Educational vouchers and cream skimming. *NBER Working Paper No. 9354*, November 2002.
- Fiske, E. B. and Ladd, H. F. (2000). *When Schools Compete: A Cautionary Tale*. Washington, DC: Brookings Institution Press.
- Glazer, S. (1997). *A Conditional Logit Model of Elementary School Choice: What Do Parents Value?* Chicago, IL: University of Chicago, Harris School of Public Policy.
- Glazer, S. (1998). School quality and social stratification: The determinants and consequences of parental school choice. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Diego, CA.
- Greene, J. P., Howell, W. G., and Peterson, P. E. (1998). Lessons from the Cleveland scholarship program. In Peterson, P. E. and Hassel, B. C. (eds.) *Learning from School Choice*, pp 357–394. Washington, DC: Brookings Institution Press.
- Hastings, J. S., Kane, T. J., and Staiger, D. O. (2006). Parental preferences and school competition: Evidence from a public school choice program. *NBER Working Paper No. 11805*, May 2006.
- Hastings, J. S., Kane, T. J., and Staiger, D. O. (2007). Preferences and heterogeneous treatment effects in a public school choice lottery. *NBER Working Paper No. 12145*.
- Hastings, J. S. and Weinstein, J. M. (2007). Information, school choice and academic achievement: Evidence from two experiments. *NBER Working Paper No. 13623*.
- Henig, J. (1990). *Choice in Public Schools: Limits of the Market Metaphor*. Princeton, NJ: Princeton University Press.
- Howell, W. G. (2004). Dynamic selection effects in a means-tested, urban school voucher program. *Journal of Policy Analysis and Management* **23**, 225–250.
- Howell, W. G. (2004). Parents, choice, and some foundations for education reform in Massachusetts, *White Paper No. 22*, Pioneer Institute for Public Policy Research.
- Howell, W. G. (2004). Fumbling for an exit key: Parents, choice and the future of NCLB. In Hess, F. M. and Finn, C. E., Jr. (eds.) *Leaving No Child Behind? Options for Kids in Failing Schools*, pp 161–190. New York, NY: Palgrave Macmillan.
- Howell, W. G., Peterson, P. E., Wolf, P. J., and Campbell, D. E. (2006). *The Education Gap: Vouchers and Urban Schools*, rev. edn. Washington, DC: Brookings Institution Press.
- Hoxby, C. (2003). School choice and school competition: Evidence from the United States. *Swedish Economic Policy Review* **10**(2), 9–65.
- Hoyt, W. H. and Lee, K. (1998). Educational vouchers, welfare effects and voting. *Journal of Public Economics* **69**, 211–228.
- Hsieh, C.-T. and Urquiola, M. (2006). When schools compete, how do they compete? An assessment of Chile's nationwide school voucher program. *Journal of Public Economics* **90**, 1477–1503.
- Kleit, B., Weiher, G., Tedin, K., and Matland, R. (2000). Choices, charter schools, and household preferences. *Social Science Quarterly* **81**(3), 846–854.
- Lankford, H. and Wyckoff, J. (1992). Primary and secondary school choice among public and religious alternatives. *Economics of Education Review* **11**(4), 317–337.
- Lankford, H. and Wyckoff, J. (2000). The effect of school choice and residential location on the racial segregation of k-12 students, *Working Paper*, SUNY-Albany.
- Levin, H. M. (1998). Educational vouchers: Effectiveness, choice, and costs. *Journal of Policy Analysis and Management* **17**(3), 373–392.
- Manno, B. V., Finn, C. E., Bierlein, L. A., and Vanourek, G. (1998). *How Charter Schools Are Different: Lessons and Implications from a National Study*. Manuscript. Phi Delta Kappa.
- McEwan, P. (2000). The potential impact of large scale voucher programs. *Review of Educational Research* **70**, 103–149.
- McEwan, P. (2004). The potential impact of vouchers. *Peabody Journal of Education* **79**(3), 57–80.
- Mizala, A. and Urquiola, M. S. (2007). School markets: The impact of information approximating schools' effectiveness. *NBER Working Paper No. 13676*, November 2007.
- Peterson, P. E., Howell, W. G., and Greene, J. P. (1999). Evaluation of the Cleveland voucher program after two years. Harvard University, *Program on Education Policy and Governance*, *PEPG No. 99-02*.

- Saporito, S. (2003). Private choices and public consequences: Magnet school choice and segregation by race and poverty. *Social Problems* **50**(2), 181–203.
- Saporito, S. and Lareau, A. (1999). School selection as a process: The multiple dimensions of race in framing educational choice. *Social Problems* **46**, 418–435.
- Schneider, M. and Buckley, J. (2002). What do parents want from schools? Evidence from the internet. *Educational Evaluation and Policy Analysis* **24**(2), 133–144.
- Schneider, M., Marschall, M., Teske, P., and Roch, C. (1998). School choice and culture wars in the classroom: What different parents seek from education. *Social Science Quarterly* **79**(3), 489–501.
- Weiher, G. R. and Tedin, K. L. (2002). Does choice lead to racially distinctive schools? Charter schools and household preferences. *Journal of Policy Analysis and Management* **21**(1), 79–92.
- Williams, M. F., Hancher, K. S., and Hutner, A. (1983). Parents and school choice: A household survey. *School Finance Project Working Paper*, Office of Educational Research and Improvement, US Department of Education, December.
- Witte, J. F. and Thorn, C. A. (1996). Who chooses? voucher and interdistrict choice programs in Milwaukee. *American Journal of Education* **104**, 186–217.
- Lankford, R. H., Lee, E. S., and Wyckoff, J. H. (1995). An analysis of elementary and secondary school choice. *Journal of Urban Economics* **38**, 236–251.
- Public Policy Forum (1997). School choice in Cleveland and Milwaukee: What parents look for. *Public Policy Forum Report*. Milwaukee, WI: Public Policy Forum.
- Rouse, C. (1998). Private school vouchers and student achievement: Evidence from the Milwaukee choice program. *Quarterly Journal of Economics* **113**(2), 553–602.
- Schneider, M., Elacqua, G., and Buckley, J. (2006). School choice in Chile: Is it class or the classroom? *Journal of Policy Analysis and Management* **25**(3), 577–601.
- Schneider, M., Teske, P., and Marschall, M. (2000). *Choosing Schools: Consumer Choice and the Quality of American Schools*. Princeton, NJ: Princeton University Press.
- Smith, K. and Meier, K. (1995). Public choice in education: Markets and the demand for quality education. *Political Research Quarterly* **48**, 329–343.
- Uchitelle, S. and Nault, R. L. (1977). The school choice behavior of parents afforded public school options: A case study of parental decision making in one suburban community. *Paper Presented to the American Educational Research Association*. San Francisco, CA.
- Wrinkle, R. D., Stewart, J., and Polinard, J. L. (1999). Public school quality, private schools, and race. *American Journal of Political Science* **43**, 1248–1253.

Further Reading

- Coldron, J. and Boulton, P. (1991). Happiness as a criterion of parents' choice of school. *Journal of Education Policy* **6**, 169–178.

Relevant Website

<http://www.DCSchoolSearch.com> – www.DCSchoolSearch.com.

The Economics of School Accountability

D N Figlio, Northwestern University, Evanston, IL, USA

H F Ladd, Duke University, Durham, NC, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Growth model – A measure of aggregate student test performance based on year-to-year changes in the test scores of individual students.

Principal-agent problem – A circumstance in which one individual or group is entrusted to make decisions on behalf of other people who may not share the same objectives.

Status model – A measure of aggregate student test performance based on the average level of student test scores.

Test-based school accountability – The practice of identifying, rewarding, and sanctioning schools on the basis of aggregated student test performance.

Demands for more accountability and results-based incentive systems in K-12 education come from many directions and currently dominate much of the education policy discussion at both the state and federal levels in the United States (Ladd, 1996; Ladd and Hansen, 1999) and abroad (Burgess *et al.*, 2005). Accountability in education is a broad concept that could be addressed in many ways: using political processes to assure democratic accountability, introducing market-based reforms to increase accountability to parents and children, developing peer-based accountability systems to increase the professional accountability of teachers, or using administrative accountability systems designed to drive the system toward higher student achievement. This article focuses on this last approach and pays particular attention to programs that focus on the individual school as the primary unit of accountability.

The use of school accountability programs in other countries notwithstanding, this article focuses exclusively on evidence and policies from the United States. The accountability systems of interest here operate within the traditional public school system and rely heavily on student testing (Hanushek and Raymond, 2003). Most emblematic is the federal No Child Left Behind Act (NCLB), which became law in 2002. NCLB requires states to test students in reading and mathematics in grades 3–8, as well as in one high school grade. In addition, it requires states to assess schools on the basis of whether their students (both in the aggregate and by subgroup) are making adequate yearly progress (AYP)

toward the ultimate goal of 100% proficiency by 2014, and it imposes consequences on schools and districts that fail to make AYP. This law is the most recent incarnation of a bipartisan standards-based reform movement that emerged from a historic 1989 summit in Charlottesville, Virginia, between President George H. W. Bush and the state governors. The meeting generated a set of national education goals that were subsequently embedded in the Clinton administration's 1994 Goals 2000: Educate America Act. NCLB differs from that act by its far heavier emphasis on accountability and its significantly greater federal intrusion into the operations of individual schools and districts.

School-based accountability programs preceded NCLB in many states. As of 2001, 45 states published report cards on schools, and 27 of them rated schools or identified low-performing ones (*Education Week, Quality Counts, 2001*). Several states (e.g., North and South Carolina, Texas, Kentucky, and Florida) as well as districts such as Chicago, Dallas, and Charlotte-Mecklenburg, also had full school-based accountability programs in which they rated schools based on their students' performance, provided rewards either to schools or to teachers for improved performance, and provided some combination of sanctions and assistance to low-performing schools.

The Rationale for School-Based Accountability

The current school-based accountability efforts emerged from the broader standards-based reform movement that began in the 1980s. Standards-based reform involves the setting of clear, measurable, and ambitious performance standards in a set of core academic subjects for students at various ages, aligning curriculum to these standards, and establishing high expectations for students to meet them. Assessment of students is a key component of standards-based reform. These assessments are used to measure student progress toward mastery of the standards and also the effectiveness of the schools they attend. The goal is to provide incentives for schools to raise student achievement.

In the context of standards-based reform, accountability is only one part of a larger policy package. Accountability – whether for schools, students, teachers, or districts – can also be viewed as a stand-alone policy. One rationale for such a policy comes from comparisons to the private sector

where business firms focus attention on results and rely on benchmarking procedures to measure progress. Another rationale is provided by the economists' model of the principal-agent problem. In such a model, school administrators and teachers may underperform because the state policymakers do not have a good means of monitoring them. It follows that student achievement would improve if state policymakers could monitor the teachers and school administrators more effectively.

School accountability systems serve as a mechanism for counteracting this principal-agent problem. By assessing schools against a common metric, policymakers generate independent information about the performance of schools and school districts. This common yardstick then allows policymakers to compare each school's performance to that of another school or to an external standard. By measuring, reporting, and, in many cases, attaching positive consequences to strong school performance and negative consequences to weak performance, policymakers provide incentives for schools and school districts to focus attention on what is being measured and possibly to change the way in which they deliver education. Such information may also facilitate improved monitoring by another important set of stakeholders in the education system, namely parents. Whether by complaining about poor performance or by threatening to withdraw their child from the school, parents could potentially use the publicly provided information on school performance to induce their children's schools to improve.

Most administrative accountability systems measure only a very small fraction of the educational outcomes that parents and policymakers value. The outcomes in such systems typically include student achievement as measured by test scores in only the core subjects of math and reading, supplemented in some cases (as under NCLB) with some nontest-based measures such as student attendance or graduation rates. As documented by Rothstein and Jacobson (2008), educational stakeholders value a much broader range of academic outcomes as well as other outcomes such as citizenship. Some of these, according to Hanushek and Raymond (2003), are more highly related to student achievement than are others. However, the narrow focus of administrative accountability systems on test results in math and reading clearly privileges one narrow set of outcomes over others.

School accountability systems can also provide administrators with incentives to game the system. Some of these responses, which tend to increase the measured achievement of a school's students but not necessarily their true achievement, are described later in this article. A number of other factors may also keep an accountability system from generating better student outcomes. Outcomes might not improve, for example, if the extrinsic motivations imposed by the accountability system crowd out powerful intrinsic motivations for educating children

(Frey, 2000). Perhaps, most important is that educators may lack the necessary skills, knowledge, or resources to meet the expectations of an accountability system. Thus, it could be that one of the major assumptions underlying stand-alone accountability programs – namely that teachers are underperforming because of insufficient monitoring of their behavior – is incorrect. In any case, greater learning gains associated with school accountability are by no means assured.

Designing School Accountability Systems

How a school accountability system is designed can have a significant impact on the nature of and the strength of the incentives that schools face to raise student achievement in the tested subjects. Moreover, the design can affect which students are likely to receive the most additional attention.

Broadly speaking, two main approaches have been used to measure school performance – status measures in which schools are judged based on their levels of performance, typically measured by average test scores or by the fraction of students attaining certain proficiency levels, and growth or value-added measures in which schools are rated on how much they improve individual students' performance from one year to the next (Hanushek and Raymond, 2003). The simplest way to measure achievement growth, or value added, is to use gains in student test scores from one year to the next. In some cases, those gains, averaged across all students in the school, are then compared to the gains in test scores that are predicted for that school given the achievement level of its students in the prior year. More refined measures of value added can be based on regression models, with or without additional statistical controls for student characteristics (Ladd and Walsh, 2002). States and local school districts have used variations of both the status and the growth approach (Hanushek and Raymond, 2003). Although NCLB requires that schools be evaluated in terms of their AYP toward the ultimate goal of 100% proficiency, such progress is measured based on comparisons of aggregate levels of student performance from one year to the next. Thus, the current iteration of NCLB is essentially a status model not a growth model, although the US Department of Education has recently given a few states the authority to introduce some growth elements into the basic approach.

The two types of measures have somewhat different goals and generate somewhat different incentives. Status-based systems that focus on the percent of students who achieve at proficient levels seek to encourage schools to raise performance at least to that level. This approach is appealing to many policymakers because it sets the same target for all groups of students and because it encourages schools to focus attention on the set of low-performing

students who in the past may have received little attention. Status-based systems also have the advantage of being transparent.

The goal of the growth-model approach is to encourage schools to improve the performance of their students independently of the absolute level of that achievement. Such an approach is appealing to many people because of its perceived fairness. It explicitly takes into account the fact that where students end up is heavily dependent on where they start and the fact that the starting points tend to be highly correlated on average with family background. At the same time, the use of the growth-model approach may raise political concerns, both because the public may find the approach less transparent than the status approach and because some see it as a way of letting schools with low average performance off the hook.

Systems using status and growth models generate different incentives in part because they lead to different rankings of schools. Many schools deemed ineffective based on their aggregate performance levels may actually have quite high value added and vice versa (Kane and Staiger, 2002). Some accountability systems (e.g., Florida's and North Carolina's) encourage both high levels of performance and high test score growth, by focusing on both levels and gains (Ladd and Zelli, 2002; Rouse *et al.*, 2007).

In addition, the two approaches send different signals about which students deserve more attention. Under a status-based system designed to encourage schools to raise student performance to some threshold level, the position of the threshold matters. A challenging performance threshold – for example, one that would be consistent with the high aspirations of the standards-based reform movement – would provide incentives for schools to focus attention on a larger group of students than would be the case with a lower threshold. Evaluating schools on the basis of value added, by contrast, provides a stronger incentive for schools to expend effort on the entire student body. In such a system, however, schools may have an incentive to focus attention on the more-advantaged students because the test score gains of those students are likely to exceed those of the less-advantaged students (Ladd and Walsh, 2002).

Under either approach, random errors in the measurement of student performance can generate inconsistent rankings of schools over time – a factor that weakens incentives for improvement. That is especially true for small schools because the smaller the number of students in the school, the larger the school-wide average measurement error, and hence less consistent the school's ranking is likely to be from 1 year to the next. Schools deemed to be improving at one point in time are often found to be declining the next year due to measurement error (Kane and Staiger, 2002). The problem of measurement error is exacerbated when schools are rated based on the growth model because it requires test scores for more

than 1 year. The danger is that personnel in such schools may receive such inconsistent signals from 1 year to the next that they have little incentive to respond in a constructive way.

Measurement-error issues in accountability are particularly pronounced when an accountability system focuses attention on specific subgroups within a school, as is mandated by NCLB and some state accountability systems. A potential alternative to the subgroup requirement would be to focus special attention on the segment of the school's students that performed at a low level in the previous year and to track that group's growth. This segment would likely include a large fraction of the economically disadvantaged and racial minority students, and so might capture the spirit of the NCLB law without exacerbating the problem of measurement error.

Neither approach to measuring school performance captures what economists call school efficiency – the effectiveness with which schools use their resources to maximize student outcomes, given the students they serve. According to the economists' model of the education production function, student achievement is determined by the characteristics of the student and his or her classmates, the school's resources (including the quantity and qualifications of the teachers), and the efficiency with which those resources are used. Efficiency cannot be observed directly and therefore, it must be inferred from statistical analysis that controls both for the resources available to the school and the characteristics of the students being served (Stiefel *et al.*, 2005). If the goal of an accountability system is to induce schools to use the resources they have more effectively, then, in principle, schools should be rated on their efficiency, not simply on the level or growth of their students' achievement. The problem is that the data requirements for such efficiency measures are often daunting and the statistical techniques can be complex (Ladd and Walsh, 2002; and Stiefel *et al.*, 2005).

The Evidence on Student Achievement

Measuring the effects of test-based accountability systems on student achievement is not a simple task. When such systems are part of a larger standards-based reform effort, it is difficult to separate the effects of the accountability system from those of other components of the reform package. In addition, researchers face the challenge of finding appropriate control groups to determine what would have happened to student achievement in the absence of the accountability system. In practice, researchers have used a variety of empirical strategies to address these challenges.

Due to the short time since NCLB was introduced and the lack of a counterfactual, the most compelling studies of how accountability affects student achievement are

based on the state and local accountability systems that preceded NCLB. This research includes district- or state-specific studies as well as cross-state studies that measure achievement using the results from the National Assessment of Education Progress (NAEP). Researchers conducting district- and state-specific studies (e.g., Ladd, 1999; Jacob, 2005; Figlio and Rouse, 2006; West and Peterson, 2006; Rouse *et al.*, 2007; Neal and Schanzenbach, 2007; Krieg, 2007) have used a combination of state- or district-wide trends in achievement, along with trends or patterns in school and student-level achievement in other comparable districts or states, to sort out how the specific accountability system in that district or state affected student achievement. The main advantages of district and state studies are that the analysis is firmly focused on a specific, well-defined accountability system. Some of the studies, particularly those for particular states, are hampered by the difficulty of predicting what would have happened to student achievement in the absence of the state's accountability system. However, it is possible to exploit idiosyncratic differences in state policy effects (e.g., Figlio and Rouse, 2006) or changes in accountability rules (e.g., Rouse *et al.*, 2007) in order to make apples-to-apples comparisons within a state.

Cross-state studies (e.g., Carnoy and Loeb (2002) and Hanushek and Raymond (2005) make use of the variation across states in the nature or timing of accountability systems. Although the conclusions of cross-state studies are sensitive to how accountability policies are defined as well as to methodological considerations such as the determination of control groups, the findings of cross-state studies are likely to be less idiosyncratic and more generalizable than those that emerge from the analysis of a specific program.

Some researchers have found modest but meaningful gains in test scores associated with school accountability, though many studies find little or no effect. Although a few methodologically rigorous studies, including Jacob (2005), Rouse *et al.* (2007), and Neal and Schanzenbach (2007), find positive impacts on reading scores, positive achievement effects emerge far less frequently in reading than in math. Larger and more consistent effects for math are intuitively plausible and are consistent with findings from other policy interventions such as voucher programs and tax and expenditure limitations. Compared to reading skills, math skills are more likely to be learned in the classroom, the curriculum is well defined and sequenced, and there is less opportunity for parents to substitute for what goes on in the classroom (Cronin, 2005).

Some evidence from the district- or state-specific studies suggests that the schools at the bottom of the performance distributions exhibit the greatest gains under an accountability system. This conclusion emerges from both Chicago (Jacob, 2005) and Florida (Figlio and Rouse, 2006). Working in the other direction is the finding

from Cronin *et al.*'s (2005) national study that the effects of high stakes are greater for the higher-scoring students. As predicted theoretically, school accountability systems often lead schools to target attention to students at particular points in the ability distribution: Neal and Schanzenbach (2007) and Krieg (2007) show that the gains in achievement associated with a status-based system tend to be most concentrated in the bubble students near the margin of proficiency, and Rouse *et al.* (2007) find that when Florida moved away from a status-based system to one that incorporates gains in student test scores, the students at the ends of the distribution (away from the bubble) appear to have benefited the most.

Evidence on Unintended Consequences

In any monitoring situation, those being monitored face incentives to appear as effective as possible against the metric being assessed. Thus, the concern arises that teachers might teach so narrowly to high-stakes tests that little or no generalizable learning would take place (Koretz and Barron, 1998). Typically, however, as long as the high-stakes tests reflect material that policymakers and the public consider important, teaching to the test would still be expected to improve student learning in the tested areas, at least to some extent. In a few cases, however, reported gains may be completely bogus. Jacob and Levitt (2004), convincingly demonstrate, for example, that a small fraction of Chicago teachers responded to accountability pressures in that city by engaging in outright cheating in order to boost measured student test performance.

Regardless of whether teacher cheating takes place, substantial evidence supports the conclusion that schools tend to concentrate their attention on the subjects tested and on the grades that have high-stakes tests (see, e.g., Deere and Strayer, 2001; Ladd and Zelli, 2002). Other studies (e.g., Hamilton *et al.*, 2005; Stecher *et al.*, 2000) show that teachers and schools tend to narrow the curriculum and shift their instructional emphasis from nontested to tested subjects, while earlier work by Shepard and Dougherty (1991) and Romberg *et al.* (1989) suggest that teachers focus more on tested content areas within specific subjects. In related work, Chakrabarti (2005) presents evidence that schools may concentrate their energies on the most easily improved areas of instruction, rather than on subjects across the board.

A common way to examine whether student learning has increased is to measure achievement using a low-stakes test, that is, one with no specific consequences for schools. A natural test for that purpose is the NAEP, which has been administered to a nationally representative random sample of students since the early 1970s and to representative samples of students in grades 4 and 8 in

most states since the 1990s. The downside to measuring achievement using scores from a test with no consequences for students or teachers is that students may not take the test sufficiently seriously to do their best work. Unless student effort differs from one administration of the low-stakes test to the next, however, changes in performance on the low-stakes test should provide a reasonable estimate of gains in student learning. In some cases, tests offer low stakes for schools but high stakes for students (e.g., for track placement or school choice); here, the low-stakes tests could potentially provide at least lower-bound estimates of the effects of accountability on genuine learning. In general, smaller estimates of accountability effects emerge when the researcher focuses on tests with lower stakes, although in some cases the lower-stakes test results approach the magnitudes of the high-stakes test results.

Schools may also engage in strategies that artificially improve test scores by changing the group of students subject to the test. The most widely studied behavior of this type is the selective assignment of students to special education programs. Many studies, including Cullen and Reback (2006) and Jacob (2005), show that schools tend to classify low achievers as learning disabled in the context of accountability systems. Though there may be some debate about whether the greater rates of classification are undesirable in all cases, nonetheless, they highlight the possibility that schools are manipulating the testing pool specifically to inflate measured school performance. Figlio's (2006) finding that some Florida schools changed their discipline and suspension patterns around the time of the testing in ways consistent with the goal of improving test-takers' average scores reinforces this concern.

Evidence suggests that schools engage in other types of strategic behavior as well. For example, Figlio and Winicki (2005) demonstrate that schools change their meal programs at the time of the tests in an apparent attempt to raise performance on high-stakes examinations, while Anderson and Butcher (2006) find that schools subject to accountability pressure are more apt than other schools to sell soft drinks and snacks through vending machines.

Many of these behaviors are less likely to occur in growth-based accountability systems than in status-based systems. The reason is that in the growth approach, the manipulative behavior that increases student achievement in one year would make it more difficult for the school to attain accountability goals the following year. No such trade-off arises in status-based accountability systems.

New evidence shows that Florida schools subject to accountability pressures in a system that includes both status and growth measures changed their instructional policies and practices in productive ways. Using detailed survey data collected from a census of public schools in Florida, Rouse *et al.* (2007) find that changes in educator practices account for a significant proportion of the realized

test-score gains associated with the change in accountability systems. The results indicate that accountability systems not only provide educators with incentives to change their behavior but also that their design affects the mix of productive and unproductive responses.

In conclusion, school accountability systems have the potential to influence school behaviors in tangible ways. School accountability systems in the United States appear to have modestly increased measured student test performance, at least in mathematics, and have likely done so through a combination of substantive and artificial means. There remains much to learn on the subject, and the research on the mechanisms through which accountability influences student achievement – and the macro-consequences of these mechanisms – is still in its infancy. More work is needed to fully understand the relative benefits and costs of school accountability systems.

See also: Competition and Student Performance; Economic Approaches to School Efficiency; Empirical Research Methods in the Economics of Education; Student Incentives; Teacher Incentives; Tiebout Sorting and Competition.

Bibliography

- Anderson, P. and Butcher, K. (2006). Reading, writing, and refreshments: Are school finances contributing to children's obesity? *Journal of Human Resources* **41**(3), 467–494.
- Burgess, S., Propper, C., Slater, H., and Wilson, D. (2005). Who wins and who loses from school accountability? The distribution of educational gain in English secondary schools. *Working Paper*, University of Bristol.
- Carnoy, M. and Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis* **24**(4), 305–331.
- Chakrabarti, R. (2005). Do public schools facing vouchers behave strategically? *Working Paper*, Harvard University.
- Cronin, J. (2005). *The Impact of the NCLB Act on Student Achievement and Growth: 2005 Edition*. Lake Oswego, OR: Northwest Evaluation Association.
- Cullen, J. and Reback, R. (2006). Tinkering towards accolades: School gaming under a performance accountability system. In Gronberg, T. and Jansen, D. (eds.) *Improving School Accountability: Check-Ups or Choice*, *Advances in Microeconomics*, vol. 14, pp 1–34. Amsterdam: Elsevier Science.
- Deere, D. and Strayer, W. (2001). Putting schools to the test: School accountability, incentives and behavior. *Working Paper*, Texas A&M University.
- Figlio, D. (2006). Testing, crime and punishment. *Journal of Public Economics* **90**(4–5), 837–851.
- Figlio, D. and Rouse, C. (2006). Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics* **90**(1–2), 239–255.
- Figlio, D. and Winicki, J. (2005). Food for thought? The effects of school accountability plans on school nutrition. *Journal of Public Economics* **89**(2–3), 381–394.
- Frey, B. (2000). Motivation and human behaviour. In Taylor-Gooby, P. (ed.) *Risk, Trust and Welfare*, pp 31–50. Basingstoke: Macmillan.
- Hamilton, L., Berends, M., and Stecher, B. (2005). Teachers' responses to standards-based accountability. *Working Paper*, RAND Corporation.

- Hanushek, E. and Raymond, M. (2003). Lessons about the design of state accountability systems. In Peterson, P. E. and West, M. R. (eds.) *NCLB? The Politics and Practice of School Accountability*, pp 127–151. Washington, DC: Brookings Institution Press.
- Hanushek, E. and Raymond, M. (2005). Does school accountability lead to improved school performance? *Journal of Policy Analysis and Management* **24**(2), 297–329.
- Jacob, B. (2005). Accountability, incentives and behavior. *Journal of Public Economics* **89**(5–6), 761–796.
- Jacob, B. and Levitt, S. (2004). Rotten apples. *Quarterly Journal of Economics* **118**(3), 843–878.
- Kane, T. and Staiger, D. (2002). Improving school accountability systems. *Working Paper*, National Bureau of Economic Research.
- Krieg, J. (2007). Are students left behind? The distributional effects of the No Child Left behind Act. *Education Finance and Policy* **3**(2), 250–281.
- Koretz, D. and Barron, S. (1998). The validity of gains on the Kentucky Instructional Results Information System (KIRIS). *Working Paper*, RAND Corporation.
- Ladd, H. (1996). *Holding Schools Accountable: Performance Based Reform in Education*. Washington, DC: Brookings Institution Press.
- Ladd, H. (1999). The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes. *Economics of Education Review* **18**(1), 1–18.
- Ladd, H. F. and Hansen, J. (1999). *Making Money Matter: Financing America's Schools*. Washington, DC: National Academy Press.
- Ladd, H. and Walsh, R. (2002). Implementing value-added measures of school effectiveness: Getting the incentives right. *Economics of Education Review* **21**(1), 1–17.
- Ladd, H. and Zelli, F. (2002). School-based accountability in North Carolina: The responses of school principals. *Education Administration Quarterly* **38**(4), 494–529.
- Neal, D. and Schanzenbach, D. (2007). Left behind by design: Proficiency counts and test-based accountability. *Working Paper*, National Bureau of Economic Research.
- Romberg, T., Zarinia, E., and Williams, S. (1989). *The Influence of Mandated Testing on Mathematics Instruction: Grade 8 Teachers' Perceptions*. Madison, WI: National Center for Research in Mathematical Science Education, University of Wisconsin–Madison.
- Rothstein, R. and Jacobson, R. (2008). Educational goals: A public perspective. In Ladd, H. F. and Fiske, E. B. (eds.) *Handbook of Research in Education Finance and Policy*, pp 78–86. New York: Routledge.
- Rouse, C., Hannaway, J., Goldhaber, D., and Figlio, D. (2007). Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure. *Working Paper*, National Bureau of Economic Research.
- Shepard, L. and Dougherty, K. (1991). Effects of high-stakes testing on instruction. *Working Paper*, ERIC.
- Stecher, B., Barron, S., Chun, T., and Ross, K. (2000). *The Effects of the Washington State Education Reform on Schools and Classrooms*. Los Angeles, CA: Center for Research on Evaluation, Standards and Student Testing.
- Stiefel, L., Schwartz, A. E., Rubinstein, R., and Zabel, J. (eds.) (2005). *Measuring School Performance and Efficiency: Implications for Practice and Research*. American Education Finance Association 2005 Yearbook. Larchmont, NY: Eye on Education.
- West, M. and Peterson, P. (2006). The efficacy of choice threats within school accountability systems: Results from legislatively induced experiments. *Economic Journal* **116**(510), C46–C62.

Further Reading

- Amrein, A. and Berliner, D. (2002). March 28. High states testing, uncertainty and student learning. *Education Policy Analysis Archives* **10**(18). <http://epaa.asu.edu/epaa/v10n18> (accessed June 2009).
- Clotfelter, C. and Ladd, H. (1996). Recognizing and rewarding success in public schools. In Ladd, H. (ed.) *Holding Schools Accountable: Performance-Based Reform in Education*, pp 23–64. Washington, DC: Brookings Institution Press.
- Elmore, R., Abelman, C. H., and Fuhrman, S. H. (1999). The new accountability in state education reform: From process to performance. In Ladd, H. F. (ed.) *Holding Schools Accountable: Performance-Based Reform in Education*, pp 65–98. Washington, DC: Brookings Institution Press.
- Figlio, D. and Getzler, L. (2006). Accountability, ability and disability: Gaming the system? In Gronberg, T. and Jansen, D. (eds.) *Improving School Accountability: Check-Ups or Choice, Advances in Microeconomics*, vol. 14, pp 35–50. Amsterdam: Elsevier.
- Haney, W. (2000). The myth of the Texas miracle in education. *Educational Policy Analysis Archives* **8**(41). <http://epaa.asu.edu/epaa/v8n41> (accessed June 2009).
- Jones, G., Jones, B., Hardin, B., et al. (1998). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan* **81**, 199–203.
- Ladd, H. (2001). School-based educational accountability systems: The promise and the pitfalls. *National Tax Journal* **54**(2), 385–400.
- O'Day, J. A. and Smith, M. S. (1993). Systemic reform and educational opportunity. In Fuhrman, S. (ed.) *Designing Coherent Education Policy: Improving the System*, pp 250–312. San Francisco, CA: Jossey-Bass.
- Peterson, P. E. and West, M. R. (2003). *NCLB? The Politics and Practice of School Accountability*. Washington, DC: Brookings Institution Press.
- Richards, C. and Sheu, T. (1992). The South Carolina School incentive reward program: A policy analysis. *Economics of Education Review* **11**(1), 71–86.

The Efficacy of Educational Vouchers*

R Zimmer, Michigan State University, East Lansing, MI, USA

E P Bettinger, Stanford University, Stanford, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Children Scholarship Fund – Scholarship program for students to attend private schools funded by the Walton Family Foundation.

Indirect effects – Effects on students not utilizing vouchers to attend private schools.

Means-tested – A criterion that uses the ability to pay as means of being eligible for a voucher.

PACES program – A Colombian program offered educational vouchers to over 144 000 low-income students throughout Colombia from 1992 to 1997.

Scholarships – While both vouchers and scholarships are entitlements to attend a private school, vouchers are publicly funded while scholarships are privately funded.

Sectarian schools – Schools affiliated with particular faith or religion.

Selection bias – Estimated effect of programs could be biased because participants have chosen to participate, which is not captured by the econometric model.

Self-selecting – Students choose to attend a private school.

Vouchers – A government-financed entitlement that students and their families could use to attend any school, including private schools.

Introduction

Advocates for vouchers argue that vouchers create incentives for schools to compete for students and that such competition can make all schools more effective and efficient. Milton Friedman originally proposed vouchers as a government-financed entitlement that students and their families could use to attend any school, including private schools (Friedman, 1955, 1962). Over time, support for voucher programs has come from diverse groups such as conservative pro-market advocates (including Friedman), Roman Catholic bishops hoping to reinvigorate parochial schools in urban cities, and minority advocacy groups.

Voucher opponents argue voucher programs cream skim the best students and that vouchers can increase

inequities if selection into voucher programs falls along racial or economic divides. They claim that by reducing public school enrolments, vouchers intensify the fiscal strains already felt by public schools.

Debates over vouchers have increased in the last decade or so with the advent of publicly funded voucher programs in Milwaukee, Cleveland, and Florida. Additionally, the increase in the number of foundations and philanthropists providing private money for students to attend private schools has also increased the public awareness of voucher programs. Outside of the United States, voucher-like programs have also increased in popularity; and while the motivation for these voucher programs differs from those in the US, debates over these programs rage worldwide. In this article, we examine many of the claims and counter-claims of the debate by surveying the research literature on both domestic and international programs.

In reviewing the existing voucher literature, we focus exclusively on empirical evidence on vouchers and do not include a review of other forms of school choice including charter schools, home schooling, or the effectiveness of private schools generally. While some research has extended beyond academic and behavioral outcomes or the distributional effects of voucher programs and examined the effects of voucher programs on residential housing markets (Nechyba, 2000), cost effectiveness of voucher programs (Levin, 2002), on subsequent public support for vouchers, or lack thereof (Brunner and Imazeki, 2006), and on who wins and loses under voucher programs (Epple and Romano, 1998, 2003), we only discuss these broader effects to the extent that the studies shed light on the specific voucher programs that we review.

This is not the first review of vouchers programs (e.g., Zimmer and Bettinger, 2007; Gill *et al.*, 2001; Levin, 2002; McEwan, 2000, 2004). These reviews include the conceptualization of, and rationale for, voucher programs as well as evidence on the efficacy of vouchers. This article builds on these reviews, but focuses on the achievement and behavioral effects of vouchers on students and the effect these programs have on access and integration.

Structure of Voucher Plans

Voucher programs differ greatly in their specific design. Much of this variation is related to the differences in the motivation for vouchers among the voucher advocates

* This review is an abbreviated version of our previous review (Zimmer and Bettinger, 2007).

(Levin, 2007 *Handbook*). For example, many conservatives believe that flat-rate, modest vouchers can increase efficiency through increased competition (e.g., Friedman, 1955, 1962). Other proponents argue that means-tested vouchers can promote equity by providing better educational opportunities to low-income families in inner-city schools. Programs also differ in their advocacy of subsidies for transportation, degree of information availability, regulation of admission policies, and accountability mechanisms (Jencks, 1970; Levin, 2002). In practice, voucher programs differ substantially in their level of support, the target population, and admission of sectarian schools.

The various programmatic details change with reference to who participates in the voucher program and are likely to influence the generalizability of individual programs and studies. For instance, programs that restrict participation to low-income students are likely to have different distributional effects than a universal voucher program. Programs that allow students to attend sectarian schools are also likely to attract a particular set of students—specifically students seeking religious instruction opportunities. Finally, the scale of the programs may affect whether these programs will generate general equilibrium effects that may be even larger than the localized effects of voucher programs (e.g., Nechyba, 2000; Epple and Romano, 1998, 2003).

As the programmatic design can affect outcomes this article, as in our previous review, highlights the details of each voucher program and the associated outcomes individually. We synthesize the results in the final section and provide a summary of the programs and outcomes in Table 1.

Domestic Voucher Programs

Domestic voucher programs have been funded either by public taxes or private donations, primarily through philanthropists or foundations. While the scale and regulation of public and privately programs may differ, they are similar in that they both provide a subsidy for students to attend private schools.

Publicly Funded Programs

The Milwaukee voucher program

The most visible domestic voucher program began in Milwaukee in 1990. The program focused on low-income students. The program started small being capped at 1% of the school district's population. In examining the program, Witte (2000) compared the performance of voucher students to that of all other students in the Milwaukee school system finding no differences in test scores. Greene *et al.* (1997, 1998), found increases in reading and math test scores when they compared outcomes for voucher recipients who had stayed in voucher schools for

multiple years to outcomes for a comparison group of students who had wanted to use the voucher but could not because of a lack of space. These divergent results led to an often heated debate over the results and methodologies (e.g., Mitgang and Connell, 2003; McEwan, 2000). Rouse (1998) attempted to resolve the controversy. Rouse took advantage of the fact that oversubscribed schools used randomization to assign vouchers. (McEwan (2007) discusses how randomization is the gold standard in evaluative research.) Rouse found slight positive, significant effects of the voucher program in math test scores, but none in reading.

Milwaukee's voucher experience also sheds light on the impact voucher programs can have on access and racial integration. African-Americans made up over 62% of the voucher population. Hispanics made up about 13% (WLAB, 2000). The program successfully targeted low-income families with the average family income for program participants being around US\$11 600 (Witte, 2000). Other studies showed that the voucher parents had slightly higher educational levels (Rouse, 1998; Witte, 1996; Witte and Thorn, 1996) than other parents in Milwaukee.

While the early Milwaukee program offered insights to policymakers, the small scale of the program limited the policy implications. In 1995, the state of Wisconsin greatly expanded the scholarship cap and the range of private schools that students could attend. These changes dramatically redefined the scope of the program. While only 341 students attended seven private schools in 1991, over 15 000 students attended dozens of sectarian and non-sectarian schools by the 2004–05 school year (Gill *et al.*, 2001). This expansion led some researchers to investigate whether Milwaukee's program resulted in changes in nearby public schools (e.g., Hoxby, 2005). Unfortunately, the expanded program did not require testing within the private schools, which resulted in no new studies of the program. A recently commissioned study (funded by the Bradley and Joyce Foundations among others) seeks to reevaluate the Milwaukee voucher program by having private schools voluntarily administer a standardized test.

The Cleveland voucher program

In 1995, the Cleveland Municipal School District began. The program was initially intended to allow vouchers to be used at sectarian schools; however, court decisions immediately blocked this provision. Not until many years later were students allowed to choose among both sectarian and nonsectarian schools. In 2007, Cleveland's program served over 6000 students.

Cleveland's voucher program has not yielded clear answers about the program's impact on academic achievement. As in the Milwaukee program, the problem with the evaluations was whether the researchers used an adequate comparison group. The most often cited evaluation of the program (Metcalf *et al.*, 2003) compared the

Table 1 Summary of international research on vouchers

<i>Voucher site</i>	<i>Authors</i>	<i>Measures effects on</i>	<i>Identification strategy</i>	<i>Key results</i>
Milwaukee WI	Witte (1990)	Voucher users	Regressions with controls	No substantial effect
	Greene, Peterson, Jiangtao (1997, 1998)	Voucher users	Quasi-experimental design	Significant effects in both reading and math
	Rouse (1998)	Voucher users	Quasi-experimental design with covariate controls	Significant effects in math only and the effects grew over time
Cleveland OH	Metcalfe (2003)	Voucher users	Observational with covariate controls	No effect
	Belfield (2006)	Voucher users	Quasi-experimental design with covariate controls	Little effect from the voucher program
Privately financed voucher programs (Children's Scholarship Fund)	Mayer <i>et al.</i> (2002), Howell and Peterson (2002)	Voucher users	Randomization of vouchers with baseline controls	Positive effects primarily among black students
	Kruger and Zhu (2004)	Voucher users	Randomization of Vouchers	No effect for any set of students, including black students
	Bettinger and Slonim (2006)	Voucher users	Randomization of vouchers	Little effects of the voucher program on student test scores; increased altruism among voucher recipients
Colombia	Angrist, Bettinger, Bloom, Kremer, and King (2002)	Voucher users	Randomization of vouchers	Positive effects on school years completed and tests
	Angrist, Bettinger, and Kremer (2006)	Voucher users	Randomization of vouchers	Positive effects on high school graduation
	Bettinger, Kremer, and Saavedra (2006)	Voucher users at vocational schools	Randomization of vouchers	Positive effects on high school graduation
Chile	Gallego (2005)	Overall system	Instrumental variable for voucher presence	System-wide increase in test scores
	Mizala and Romaguera (1998)	Voucher users	Regressions with controls	No effects on users
	Sappelli and Vial (2002)	Voucher users	Heckman selection model	Positive effects
	Rounds (1996)			
	Hsieh and Urquiola (2003)	Overall system	Regressions with controls; aggregation	No system-wide increase in test scores
	Bravo, Contreras, and Sanhueza (2000)	Voucher users	Regressions with controls	No effects on users; effects depend on controls
	McEwan (2001)	Voucher users	Regressions with controls	No effects at most voucher schools
	McEwan and Carnoy (2000)	Voucher users	Regressions with controls	Mixed effects but private schools more cost effective
	Tokman (2001)	Voucher users	Regressions with controls	Mixed effects depending on student background

academic achievement of voucher users to two groups of students: (1) students offered vouchers who did not use the vouchers, and (2) a matched set of students who were not offered vouchers. Other researchers, including Gill *et al.* (2001), argued that the study failed to create an adequate control for students self-selecting into the program. Belfield (2006) used a different comparison group focusing on applicants who were rejected in their application to private schools. Overall, he finds little effect on academic achievement.

Another key motivation for establishing the voucher programs was to improve students' access to better schools. Kim Metcalf (1999) found that the program successfully targeted minorities and students from low-income families. The mean income level of students utilizing the vouchers was US\$18 750 and these students were more likely to be African-American students than a random sample of Cleveland students. As in Milwaukee, parents of voucher applicants had slightly higher levels of education (Metcalf, 1999; Peterson *et al.*, 1999).

Other publicly funded voucher programs

Voucher programs are also in place in Vermont, Maine, Florida, and Washington, DC. Of these programs, only DC's Opportunity Scholarship program has any publicly released results, which are only the first-year impacts. The Institute of Education Sciences-funded evaluation of Washington's program is using a randomized design to evaluate the effectiveness of the program. To date, the results suggests that parents are more satisfied with the program, but students utilizing the vouchers do not report higher levels of satisfaction or feeling safer. In addition, test scores for these students are on par with nonvoucher students (Wolf *et al.*, 2007). However, caution is warranted in interpreting these results as they are noted only after the first year of implementing the voucher program.

Domestic, Privately Funded Programs

Roman Catholic dioceses have provided scholarships to students to attend their schools for years. The use of philanthropists and foundation money for private school scholarships was first popularized by J. Patrick Rooney, who formed the Educational Choice Charitable Trust in 1991. This trust allocated private scholarships to low-income students to use at private schools in Indianapolis. Other programs in Milwaukee, Atlanta, Denver, Detroit, Oklahoma City, and Washington, DC, soon followed.

Some of the programs were established and financed by conservatives who wanted to either open up sectarian school options to more students or to create competition between private and public schools. Other programs (e.g., Oklahoma City) were funded by liberal voucher proponents who wanted to provide additional schooling opportunities for inner-city students.

In 1994, the Walton Family Foundation helped create the Children's Educational Opportunity Foundation (CEO America) to support and create private scholarship programs. Shortly thereafter, the number of privately funded programs quickly expanded across the country. Of these programs, one of the most ambitious was a CEO America program in San Antonio funded at US\$5 million a year for at least 10 years. The program provided full scholarships to over 14 000 at-risk students making it the largest privately funded voucher program in existence at the time.

In 1998, the Walton Family Foundation helped create the Children's Scholarship Fund (CSF). CSF partnered with local funders to provide scholarships in cities across the nation. In April 1999, 1.25 million low-income students applied for 40 000 partial scholarships. Over the last decade, more and more of the programs developed, often receiving very little national attention, making it difficult to know exactly how many of the programs currently exist.

Of these programs, CSF has received the most national attention partially because of its relatively large scale and partially because CSF used randomization to assign scholarships. With randomization, researchers can compare students who participated in the lotteries to determine the effectiveness of these programs. (Zimmer and Bettinger (2007) discuss some of the challenges in these studies.) Other studies have been carried out in Charlotte, Dayton, New York, and Washington, DC among other locations. While Greene (2000) found a positive overall effect on test scores in Charlotte after 1 year, these studies have generally shown no effect. The one exception is among African-American students. Howell and Peterson (2002) and Myer *et al.* (2002) found that African-American students' test scores improved in New York's program. However, additional research by Krueger and Zhu (2002) showed that this research was sensitive to the definition of an African-American student. While the original analysis used the student's mother's identified race for the student's race, Krueger and Zhu define African-American students as those whose mother or father are identified as African-American. This more inclusive definition results in a scholarship effect size that is smaller than reported in the original study and is not statistically significant.

In a more recent study, Bettinger and Slonim (2006) focus on the effects of vouchers on nonacademic outcomes. Using surveys and new methods from experimental economics, they attempted to measure the effects of the vouchers on traditional outcomes such as test scores and on altruism. To test the effects of the voucher program on altruism, Bettinger and Slonim gave students US\$10 each and invited them to share some of their money with charities. In this context, they show that voucher recipients gave more to charities as a result of the voucher program.

Given that these programs were designed to target impoverished students, it is not surprising to find that

participants have relatively low income. Peterson (1998), for example, found that the average income level of families of students participating in the New York scholarship program was only US\$10 000. The parents' education level of the participants, however, was slightly higher than the average level of the eligible population. Similarly, Howell and Peterson (2000) and Wolf *et al.* (2000) found that the families of participating students were low income. In terms of race, 100% of students participating in the Washington, DC scholarship program were minorities, while 95% of the students in New York were minority, mostly Hispanic students. In Dayton, the percentage of minorities was lower but was still 66%.

Finally, these private voucher programs typically serve a low percentage of special education students. For example, in New York, 9% of participating students have disabilities compared to the district-wide average of 14% of special education students (Myers *et al.*, 2000). In Charlotte, 4% of participating students had disabilities, while the district-wide average of special education students was 11% (Greene, 2000).

International Voucher Programs

There are a number of educational voucher or voucher-like programs across the world, including programs in Chile, Colombia, Sweden, Netherlands, Belize, Japan, Canada, and Poland. These voucher programs often differ significantly from the US programs either in terms of the motivating factors or the policy context. For example, many of the non-US programs are motivated by the goal of increasing school attendance among girls or low-income students. The relationship between church-sponsored private schools and public schools is also less defined in other countries where, in many cases, religious groups operate public schools.

We focus our attention on two programs that have figured most prominently in debates on the efficacy of vouchers – Colombia's *Programa de Ampliación de Cobertura de la Educación Secundaria* (PACES) voucher program and Chile's national voucher program.

Colombia PACES Program

The PACES offered educational vouchers to over 144 000 low-income students throughout Colombia from 1992 to 1997. The voucher program awarded full private school tuition to secondary school students who wished to transfer from public to private school at the start of their secondary school. Over time, the voucher's value did not keep up with the cost of inflation.

In contrast to the US programs, the motivation for Colombia's voucher program was to expand the capacity of the public school system. In Colombia, most high

school buildings host multiple high school sessions per day. Since most private schools were not overcrowded, Colombia established the voucher program to take advantage of excess capacity in the private system.

In cities where PACES was oversubscribed, cities held lotteries to award the vouchers. Angrist *et al.* (2002) use this randomization to measure the effects of the program. They find that within 3 years, voucher students had completed about 0.1 years of schools more than their peers and had test scores about 0.2 standard deviations higher than those who did not receive vouchers. They also find that the incidence of child labor and teenage marriage was lower as a result of the educational voucher. Based on follow-up research on students' high school careers, Angrist *et al.* (2006) find that voucher lottery winners were 20% more likely to graduate from high school than voucher lottery losers.

In Colombia, the evidence on the impacts of stratification is less developed and conclusive than the effects on achievement. Colombia's educational vouchers targeted low-income families living in the poorest neighborhoods, and Ribero and Tenjo (1997) report that the vouchers were largely effective in reaching this population. Voucher applicants generally came from families with higher educational levels than other families in the same neighborhoods (Angrist *et al.*, 2002).

Chile

In 1980, Chile embarked on an ambitious series of educational reforms designed to decentralize and privatize education. At the urging of Milton Friedman, who along with other prominent economists advised the Pinochet government at the time (Rounds, 1996), Chile established perhaps the world's largest ongoing voucher program. The program offered tuition subsidies to private schools. Before the reform, the budgets of public schools were insensitive to enrolment; however, after the reform, local public schools lost money when students transferred to voucher schools.

Unlike Colombia, where vouchers targeted poor students, the Chilean system was available to all students. Additionally, unlike other programs that do not allow selective admission, voucher schools in Chile could admit the students they most preferred. As a result of these policies, Rounds (1996) found that the poorest families were less likely to attend voucher schools.

Research on the efficacy of Chile's voucher program has been much more difficult to interpret than the Colombian research because of the nature of this admission policy. Some of the early evidence suggested that voucher schools modestly outperformed public schools. This finding was common in many papers (e.g., Bravo *et al.*, 1999; McEwan and Carnoy, 2000) but was sensitive to the types of controls included in the empirical model,

the specific municipalities included in the sample, and the statistical methods used. McEwan (2001), for example, found that Catholic voucher schools tended to be more effective and productive than other schools, but only for certain specifications of the model. McEwan and Carnoy (2000) show that academic achievement is slightly lower in nonreligious voucher schools, particularly when located outside the capital. Given that these voucher schools have less funding, however, McEwan and Carnoy suggest that they could still be more cost effective than public schools.

As research on the Chilean system continued, many researchers took note of the fact that the voucher program altered the composition of both public and private schools. For example, Hsieh and Urquiola (2003) showed that more affluent families took advantage of the vouchers. Similar to McEwan and Carnoy (2000) and Carnoy (2000), Hsieh and Urquiola suggest that this shift in student populations could account for the finding that private schools appear to be more effective than public schools. In the early 1990s, many voucher schools began charging tuition in addition to the voucher, and the difference in parents' incomes and education levels between these tuition-charging voucher schools and the other voucher schools was significant (Anand *et al.*, 2006).

How this increased sorting across voucher and non-voucher schools affects student achievement depends on the nature of peer effects. If improvements in peer quality lead to better educational outcomes for voucher users, then the sorting associated with the voucher program could improve their outcomes. At the same time, given the exit of high-quality students from the public schools, the students left in those schools may have systematically worse outcomes because of the deterioration of their peers. The aggregate effect of the voucher depends on the strength of these two effects.

Hsieh and Urquiola (2003) argue that the only way to identify the overall effects of the voucher program is to focus on aggregate outcomes because it is difficult, if not impossible, to remove the selection bias inherent in comparisons of different schools. The change in aggregate test scores reflects both the direct achievement effects for voucher recipients and the indirect effects for students who remain in public schools. When the authors look at changes in aggregate test scores throughout Chile, they find no evidence that the voucher program increased overall test scores.

Gallego (2005) provides other evidence on the Chilean voucher program. Gallego uses an instrumental variable approach to estimate the effects of the program. His findings suggest positive effects of the voucher program on the academic outcomes of students throughout municipalities where the voucher program had more penetration. While the result may be indicative of competitive effects, it is driven in part by the effects of the program on

voucher recipients. It echoes earlier research (McEwan, 2001; McEwan and Carnoy, 2000) which suggested that voucher schools affiliated with Catholicism had better outcomes than other voucher schools, public or private. As McEwan and Carnoy (2000) show, Catholic schools produced better students at a lower cost than other, public or private, voucher schools.

Summary and Conclusion

Table 1 summarizes the achievement findings from both domestic and international voucher programs. The table highlights the location, program, methods used in the analysis, and their key findings.

Research on domestic voucher programs has produced inconsistent results. While advocates have often pointed to positive effects for African-American students in the CSF program as proof that vouchers can work, the strength of these findings has been questioned – leaving many skeptical about the purported benefits of vouchers. International studies have also found mixed results. The evidence from Colombia indicates that voucher users have better academic and nonacademic outcomes than they would have in the absence of the voucher. In Chile, the estimated effects differ across studies, and a number of confounding factors make it difficult to ascertain the true effects. The research generally shows that students who take advantage of vouchers are disadvantaged students, especially in voucher programs that are means tested and specifically target such students.

To date, most of the voucher studies focus on programs that are typically small in scale. Hence, it is not clear whether the results generalize to large-scale programs. For a program with massive movements of students, an important policy consideration is not only who is participating in these programs, but what happens to the racial/ethnic and ability distribution of students in public schools.

In sum, researchers have failed to come to consensus on the efficacy of vouchers as a reform effort. This murkiness may be clarified as new studies emerge, including studies of the current Milwaukee program and the federally funded Washington DC vouchers.

Future research needs to peer inside the black box and to examine the mechanism in which outcomes differ among schools. Voucher researchers should take the opportunity to look at private schools that may be doing things differently and how the variation in the operation affects outcomes. Additionally, future research can shed additional light on the effects of vouchers on other critical outcomes such as individual behaviors, educational attainment as measured by college attendance, and wages.

It is also interesting to consider the value of vouchers programs relative to other forms of choice. When the idea

of vouchers were first introduced, the only choices a family could make was sending their students to private schools without subsidies or choosing among various public schools based on residential location. However, over the last 50 years, a number of other alternatives have evolved including charter schools, magnet schools, and other inter- and intra-district choice programs. A relevant policy question has to do with the advantages and disadvantages of voucher programs relative to these other choice options, especially when many of these choice options are politically more feasible.

See also: Competition and Student Performance; Educational Privatization; The Economic Role of the State in Education; The Economics of Catholic Schools; The Economics of Parental Choice.

Bibliography

- Angrist, J., Bettinger, E., Bloom, E., Kremer, M., and King, E. (2002). The effects of school vouchers on students: Evidence from Colombia. *American Economic Review* **92**(5), 1535–1558.
- Angrist, J., Bettinger, E., and Kremer, M. (2006). Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia. *American Economic Review* **96**(3), 847–862.
- Belfield, C. R. (2006). The evidence on education vouchers: An application to the Cleveland scholarship and tutoring program. http://www.ncspe.org/publications_files/OP112.pdf (accessed June 2009).
- Bettinger, E. and Slonim, R. (2006). Using experimental economics to measure the effects of a natural educational experiment on altruism. *Journal of Public Economics* **90**(8–9), 1625–1648.
- Bravo, D., Contreras, D., and Sanhueza, C. (1999). *Rendimiento escolar, desigualdad y brecha de desempeño privado/público: Chile 1982–1997. Documento de Trabajo*, Departamento de Economía, U de Chile, 1999.
- Brunner, E. J. and Imazeki, J. (2006). Tiebout choice and the voucher. *Working Paper*. University of Connecticut Economics Department.
- Chubb, J. E. and Moe, T. M. (1990). *Politics, Markets, and America's Schools*. Washington, DC: Brookings Institution.
- Coons, J. E. and Sugarman, S. D. (1978). *Education by Choice: The Case for Family Control*. Berkeley, CA: University of California Press.
- Epple, D. and Romano, R. (1998). Competition between private and public schools, vouchers and peer group effects. *American Economic Review* **88**, 33–63.
- Epple, D. and Romano, R. (2003). Neighborhood schools, choice, and the distribution of educational benefits. In Hoxby, C. M. (ed.) *The Economics of School Choice*, pp 227–286. Chicago, IL: The University of Chicago Press.
- Friedman, M. (1955). The role of government in education. In Solo, R. A. (ed.) *Economics and the Public Interest*, pp 123–144. Piscataway, NJ: Rutgers University Press.
- Friedman, M. (1962). *Capitalism and Freedom*. Chicago, IL: University of Chicago Press.
- Gallego, F. A. (2005). Voucher-school competition, incentives, and outcomes. Evidence from Chile: MIT mimeo.
- Gill, B., Timpane, P. M., Ross, K. E., and Brewer, D. J. (2001). *Rhetoric versus Reality: What We Know and What We Need to Know about Vouchers and Charter Schools*. Santa Monica, CA: RAND Education.
- Greene, J. P. (2000). *The Effect of School Choice: An Evaluation of the Charlotte Children's Scholarship Fund*. New York: Center for Civic Innovation at the Manhattan Institute.
- Greene, J. P., Peterson, P. E., and Du, J. (1997). *Effectiveness of School Choice: The Milwaukee Experiment*. Cambridge, MA: Harvard University Program in Education Policy and Governance.
- Greene, J. P., Peterson, P. E., and Du, J. (1998). School choice in Milwaukee: A randomized experiment. In Peterson, P. E. and Hassel, B. C. (eds.) *Learning from School Choice*, pp 335–356. Washington, DC: Brookings Institution.
- Howell, W. G. and Peterson, P. E. (2000). *School Choice in Dayton, Ohio: An Evaluation after One Year*. Cambridge, MA: Program on Education Policy and Governance, Harvard University.
- Howell, W. G. and Peterson, P. E. (2002). *The Education Gap: Vouchers and Urban Schools*. Washington, DC: Brookings Institute.
- Hsieh, C. and Urquiola, M. (2003). When schools compete, how do they compete? An assessment of Chile's nationwide school voucher program. *NBER Working Papers* 10008.
- Jencks, C. (1970). Educational vouchers: A report on financing elementary education by grants to parents. *Report Prepared under Grant CG 8542 for the U.S. Office of Economic Opportunity*. Washington, DC: Centre for Policy Studies.
- Krueger, A. B. and Zhu, P. (2002). *Another Look at the New York City School Voucher Experiment, Working Paper 9418*. Cambridge, MA: National Bureau of Economic Research.
- Levin, H. (2002). A comprehensive framework for evaluating educational vouchers. *Educational Evaluation and Policy Analysis* **24**(3), 159–174.
- Levin, H. (2007). Issues in educational privatization. In Ladd, H. and Fiske, E. (eds.) *Handbook of Research in Education Finance and Policy*, pp 447–466. New York: American Education Finance Association/Erlbaum.
- McEwan, P. J. (2000). The potential impact of large-scale voucher programs. *Review of Educational Research* **70**(2), 103–149.
- McEwan, P. J. (2001). The effectiveness of public, Catholic, and non-religious private schools in Chile's voucher system. *Education Economics* **9**(2), 103–128.
- McEwan, P. J. (2004). The potential impact of vouchers. *Peabody Journal of Education* **79**(3), 57–80.
- McEwan, P. J. and Carnoy, M. (2000). The effectiveness and efficiency of private schools in Chile's voucher system. *Educational Evaluation and Policy Analysis* **22**(3), 213–239.
- Metcalfe, K. K. (1999). *Evaluation of the Cleveland Scholarship and Tutoring Grant Program: 1996–1999*. Bloomington, IN: The Indiana Center for Evaluation.
- Metcalfe, K. K., West, S. D., Legan, N. A., Paul, K. M., and Boone, W. J. (2003). Evaluation of the Cleveland scholarship and tutoring program. *Summary Report 1998–2002*. Bloomington, IN: Indiana University School of Education.
- Mitgang, L. D. and Connell, C. V. (2003). Educational vouchers and the media. In Levin, H. M. (ed.) *Privatizing Education: Can the Marketplace Deliver Choice, Efficiency, Equity, and Social Cohesion?* pp 20–38. Boulder, CO: Westview.
- Myers, D., Peterson, P., Mayer, D., Chou, J., and Howell, W. G. (2000). *School Choice in New York City after Two Years: An Evaluation of the School Choice Scholarship Program*. Washington, DC: Mathematica Policy Research.
- Myers, D. P., Peterson, P. E., Myers, D. E., Tuttle, C. C., and Howell, W. G. (2002). School choice in New York City after three years. An evaluation of the school choice scholarships program. *Rep. No. 8404-045*. Princeton, NJ: Mathematica Policy Research.
- Nechyba, T. (2000). Mobility, targeting and private school vouchers. *American Economic Review* **90**, 130–146.
- Peterson, P. E. (1998). *An Evaluation of the New York City School Choice Scholarships Program: The First Year*. Cambridge, MA: Program on Education Policy and Governance, Harvard University.
- Peterson, P. E., Howell, W. G., and Greene, J. P. (1999). *An Evaluation of the Cleveland Voucher Program after Two Years*. Cambridge, MA: Program on Education Policy and Governance, Harvard University.
- Ribero, R. and Tenjo, J. (1997). University de los Andes, Department of Economics Working Paper.
- Rounds, P. T. (1996). Will pursuit of higher quality sacrifice equal opportunity in education? An analysis of the educational voucher system in Santiago. *Social Science Quarterly* **77**(4), 821–841.
- Rouse, C. E. (1998). Private school vouchers and student achievement: An evaluation of the Milwaukee parental choice program. *Quarterly Journal of Economics* **113**, 553–602.

- Wisconsin Legislative Audit Bureau (WLAB) (2000). *Milwaukee Parental Choice Program: An Evaluation*. Madison, WI: Legislative Audit Bureau.
- Witte, J. F. (1996). Who benefits from the Milwaukee choice program? In Fuller, B., Elmore, R. F., and Orfield, G. (eds.) *Who Chooses, Who Loses? Culture, Institutions, and the Unequal Effects of School Choice*, pp 118–137. New York: Teachers College Press.
- Witte, J. F. (2000). *The Market Approach to Education*. Princeton, NJ: Princeton University Press.
- Witte, J. F. and Thorn, C. A. (1996). Who chooses? Voucher and interdistrict choice programs in Milwaukee. *American Journal of Education* **104**, 187–217.
- Wolf, P., Gutmann, B., Puma, M., Rizzo, L., and Eissa, N. (2007). Evaluation of D.C. opportunity scholarship program: Impacts after one year. Institute of Education Science, U.S: Department of Education, NCEE 2007-4009.
- Wolf, P., Howell, W. G., and Peterson, P. E. (2000). *School Choice in Washington D.C.: An Evaluation after One Year*. Cambridge, MA: Program on Education Policy and Governance, Harvard University.
- Zimmer, R. and Bettinger, E. (2007). Beyond the rhetoric: Surveying the evidence of vouchers and tax credits. In Ladd, H. and Fiske, E. (eds.) *Handbook of Research in Education Finance and Policy*, pp 447–466. New York: American Education Finance Association/Erlbaum.
- Gill, B. and Booker, K. (2007). School competition and student outcomes. In Ladd, H. F. and Fiske, E. B. (eds.) *Handbook of Research in Education Finance and Policy*, pp 183–202. New York: American Education Finance Association/Erlbaum.
- Greene, J. P. and Hall, D. (2001). *The CEO Horizon Scholarship Program: A Case Study of School Vouchers in the Edgewood Independent School District, San Antonio*. Washington, DC: Mathematica Policy Research.
- Howell, W. G. (2004). Dynamic selection effects in mean-tested, urban school voucher programs. *Journal of Policy Analysis and Management* **23**(2), 225–250.
- King, E., Rawlings, L., Gutierrez, M., Pardo, C., and Torres, C. (1997). Colombia's targeted education voucher program: Features, coverage and participation. *Working Paper No. 3*, Series on Impact Evaluation of Education Reforms, Development Economics Research Group, The World Bank.
- Ladd, H. F. and Hansen, J. S. (eds.) (1999). *Making Money Matter: Financing America's Schools*. Washington, DC: National Academy Press.
- Levin, H. (1998). Educational vouchers: Effectiveness, choice, and costs. *Journal of Policy Analysis and Management* **17**(3), 373–392.
- Levin, H. (1999). The public-private nexus in education. *American Behavioral Scientist* **43**(1), 124–137.
- Peterson, P. E., Campbell, D. E., and West, M. R. (2001). *An Evaluation of the Basic Fund Scholarship Program in the San Francisco Bay Area, California*. Cambridge, MA: Program on Education Policy and Governance.
- Toma, E. F. (1996). Public funding and private schooling across countries. *Journal of Law and Economics* **39**, 121–148.

Further Reading

- Belfield, C. R. and Wooten, A. L. (2003). Educational tax credits for America's schools. *NCSPE Policy Briefing 3*. http://www.ncspe.org/publications_files/PB03.pdf (accessed June 2009).
- Bettinger, E., Kremer, M., and Saavedra, J. (2006). How educational vouchers affect students – the role of vocational high schools in Colombia. *Case Western Reserve* mimeo.
- Carnoy, M. (1998). National voucher plans in Chile and Sweden: Did privatization reforms make for better education? *Comparative Education Review* **42**(3), 309–337.
- Coleman, J., Hoffer, T., and Kilgore, S. (1982). Cognitive outcomes in public and private schools. *Sociology of Education* **55**, 65–76.
- Evans, W. N. and Schwab, R. (1995). Finishing high school and starting college: Do Catholic schools make a difference? *Quarterly Journal of Economics* **110**(4), 941–974.
- Figlio, D. and Ludwig, J. (2000). Sex, drugs and Catholic schools: Private schooling and non-market adolescent behaviors. *NBER Working Paper No. 7990*, November 2000.

Relevant Websites

- <http://www.brookings.edu> – Brookings Institute.
- <http://www.edreform.com> – Center for Education Reform.
- <http://www.edweek.org> – Education Week.
- <http://www.friedmanfoundation.org> – Friedman Foundation for Educational Choice.
- <http://www.nea.org> – National Education Association.
- <http://www.ncspe.org> – National Center for the Study of Privatization in Education.
- <http://www.vanderbilt.edu> – Vanderbilt's National Center on School Choice.

Tiebout Sorting and Competition

T J Nechyba, Duke University, Durham, NC, USA; National Bureau of Economic Research, Cambridge, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Over 50 years ago, Charles Tiebout hypothesized that decentralized provision of public services (such as public schools) through local governments can result in efficient levels of such services (Tiebout, 1956). His key insight was that residential mobility of households might provide a disciplining force analogous to typical market forces under perfect competition in private goods markets. Local communities are viewed as analogous to competing shopping malls that offer different services – competing with one another while, at the same time, appealing to somewhat different clienteles. By considering the mix of local taxes and public services much as they consider the mix of services and goods at malls, households in effect vote with their feet, choosing communities (much like malls) that best meet their needs while making it difficult for local governments (much like mall owners) to engage in excessive rent-seeking.

Such Tiebout competition has three essential features. First, when households demand different mixes for public services, decentralization permits the emergence of different packages of such services (and taxes) across communities, providing opportunities for households to match to communities that are optimal from their perspective. Second, the resulting competition between communities for residents reduces the scope for local governments to focus on inefficient behavior that is not aimed directly at meeting the needs of their customers. It was hypothesized by Tiebout that this results in sorting of households across communities, with competition between communities providing a sufficient array of choices to satisfy different household needs in a productively efficient manner. At the same time, however, demand is determined in part by ability to pay – which implies a third feature of such Tiebout competition: sorting will occur at least in part on the basis of household income.

Over the past half century, it has become clear that Tiebout's insights have considerable empirical support. In the area of public education, however, it is far from clear that the results from Tiebout competition are as unambiguously desirable as might have been originally assumed. This is in large part due to the nature of education production that involves financial and non-financial scarce resources which are rationed through the same Tiebout forces that cause households to sort across communities. While such sorting indeed occurs along the lines suggested by Tiebout, much of it is related to household income and race – and this implies a public school system in which public school quality invariably becomes highly correlated

with these household characteristics even when higher-level governments equalize financial resources available to local public schools. The equity concerns that arise from residential Tiebout segregation (and the resulting sorting of nonfinancial inputs into schools) may then need to be balanced with any efficiency gains from sorting and from competitive pressures on school producers.

The following sections are discussed below. First, in the section titled 'Residential mobility, capitalization, and household preferences for education', we review the empirical literature that has validated important aspects of Tiebout's hypothesis while cautioning that this evidence implies that efficiency in education markets is not fully obtained through the type of competition envisioned by Tiebout. Second, in the section titled 'Tiebout sorting and the rationing of school inputs', we turn to the equity concerns that arise from the types of sorting that result from Tiebout competition. Third, we turn to a brief review of the portion of Tiebout's theory that relates to productive efficiency from residential competition (in the section titled 'Tiebout competition to enhance productive efficiency'). Finally, in the section titled 'A partial divorce between competition and Tiebout', we explore the various ways in which recent policy reforms introduce other forms of school competition alongside the type of competition envisioned by Tiebout.

Residential Mobility, Capitalization, and Household Preferences for Education

The earliest empirical tests of the Tiebout hypothesis (that decentralized provision results in efficiency) were, in many ways, only tests of the underlying Tiebout assumption of residential mobility. These tests, beginning with Oates (1969), took the form of tests for the extent to which inter-jurisdictional differences in local service and tax levels are captured in house prices. When higher service levels and lower taxes are accompanied by higher house prices – and when lower service levels and higher taxes are accompanied by lower house prices, economists refer to this as capitalization of service and tax levels. The only mechanism that would give rise to such capitalization is precisely the one that arises from Tiebout's assumption of residential mobility. If households indeed consider local service and tax levels as they determine their optimal residential location, it was argued, then identical houses in different jurisdictions will be priced differently depending on the differences in

these jurisdictional variables. At the same time, as was first pointed out by Edel and Sclar (1973) in a response to Oates, the presence of such capitalization can be interpreted as evidence against Tiebout's larger efficiency hypothesis. This is because capitalization of public service and tax levels suggests there is room in the market for additional communities to form – with increased supply of housing attached to such service/tax levels driving down capitalization as new communities emerge to meet demand. The presence of capitalization therefore suggests that there exist barriers to the emergence of such new communities – but a central feature assumed by Tiebout for his efficiency result was that such barriers do not play a significant role. A more careful theoretical exploration of capitalization under Tiebout competition, reviewed in more detail by Epple and Nechyba (2004), then clarified some of the initial confusion over the relationship between empirical findings of capitalization and the validity of the larger Tiebout hypothesis.

The empirical literature on capitalization of local services, and of school quality in particular, is substantial and has increased in sophistication since Oates' original publication. Black (1999) introduced a particularly useful innovation by considering housing price differences immediately around district boundaries rather than comparing housing across districts more generally – thus limiting the analysis to houses that arguably differ primarily in terms of the access they provide to local schools. This boundary discontinuity research design has since been generalized in more complete (or structural) models that attempt to directly estimate household tastes for school and neighborhood characteristics rather than merely identifying the magnitude of capitalization into housing prices (Bayer *et al.*, 2007). Although magnitudes differ, the qualitative finding of capitalization effects has been robust across all research designs. More importantly, however, the most recent literature directly confirms the importance of strong household tastes for both schools and neighbors, a result that has important implications for Tiebout sorting of households across communities when access to public schools is based on household residence.

Tiebout Sorting and the Rationing of School Inputs

Equity concerns that arise from Tiebout sorting are then closely linked to an understanding of the mechanism that gives rise to school quality. If financial inputs, such as per pupil spending, which, in turn, is closely linked to teacher–student ratios, fully characterized the education production process, such concerns could easily be addressed through higher-level government redistribution of financial resources. In many ways, this was precisely the focus of policy for a number of decades,

following a wave of state court decisions that found excessively decentralized public school funding mechanisms to violate the equal protection clauses in state constitutions. Beginning with the famous California Serrano decisions in California in the 1970s, states have responded with a variety of interventions ranging from mandating essentially equal per pupil spending across all districts to different state aid formulae that assist poorer districts in increasing per pupil spending levels. It is now well understood, however, that there is a limit to the extent to which such policies can address the inequities in a Tiebout-based system of public education. The reason for this is that financial inputs represent only one category of inputs essential for public school quality, with a number of different nonfinancial inputs playing a significantly more important role.

These nonfinancial inputs have often been modeled as peer effects. Researchers begin with the assumption that students have some innate ability level, and a given student's performance is dependent on the ability levels of students who attend the same school or class. When such ability levels are correlated with household income, the residential Tiebout sorting of households then implies a nonrandom sorting of ability levels into schools – with higher-income schools disproportionately benefiting from the nonfinancial input of good peer effects. This, in turn, implies that equalization of financial inputs cannot be expected to equalize educational quality so long as such nonfinancial inputs are sorted through Tiebout competition.

More generally, the same applies to other nonfinancial inputs that can be modeled in ways similar to such peer effects. For instance, teacher quality, long recognized as both an important input into school production and largely uncorrelated with measurable characteristics of teachers, is similarly allocated across schools in nonrandom ways – especially in systems that do not permit differential pricing of teacher salaries. In particular, when high-quality teachers cannot be adequately compensated through salary differentiation, they either leave the teaching profession or are compensated through assignments to more desirable schools in higher-income districts (Loeb and Page, 2001). This mechanism then gives rise to a differential sorting of teachers that is correlated with the residential sorting of higher-income households. It is similarly recognized that parents themselves are an important input into school production – both because of the home production of education (that results in different levels of child ability) and because of the role they play in monitoring schools. Both these features of parental inputs into education are similarly correlated with parental income – once again implying a differential allocation of nonfinancial inputs that arises from residential Tiebout sorting.

When incorporated explicitly into Tiebout models of residential location choices, this understanding of school production as emerging from both financial and

nonfinancial inputs has given rise to empirically grounded simulation models that can be used to investigate the potential for policy to address equity concerns from Tiebout sorting. Such models begin with a characterization of the underlying housing market within and across jurisdictional boundaries. (While early attempts modeled housing as differing primarily or solely between jurisdictions, it has been documented that housing quality varies both within and across districts – which, in turn, has led to different approaches for introducing within-jurisdiction differences in housing quality. In some such approaches, these differences emerge from the fact that households differ in terms of the relative weight they place on housing quality and school quality (Epple and Sieg, 1999). Alternatively, the distribution of housing quality can be modeled as historically determined within and across districts, with housing prices adjusting to equilibrate supply and demand at each location (Nechyba, 2000). These approaches are discussed in greater detail in Epple and Nechyba (2004) and Nechyba (2006).) School production is then introduced as a function of both financial and nonfinancial inputs, with the latter assumed to be correlated with aspects of the income distribution of the households whose children attend local public schools. Such Tiebout models that combine housing markets with realistic school production processes allow researchers to understand the emergence of school inputs from residential location choices that, in turn, are linked to housing markets. Nonfinancial inputs emerge from the residential sorting patterns that result from households voting with their feet, and financial inputs emerge both through a political process in which households vote in the more traditional sense and from the determination of local tax bases that is driven by supply and demand in housing markets.

The complexity of such theoretical models then precludes an analysis that is purely analytic – there are simply too many competing forces that depend on the underlying primitives in the models. These primitives include the characteristics of housing markets and the nature of jurisdictional boundaries, the relative weight that school production functions place on financial and nonfinancial inputs, the underlying income distribution and the parameters that characterize household tastes for consumption, schooling, and housing quality. It therefore becomes necessary to find ways of allowing data to inform these underlying primitives – in essence, to convert a rich theoretical model into one that is empirically grounded to the extent that it can replicate important features of local education and housing markets. Initial attempts relied heavily on the calibration of these primitives while more recent attempts rely on statistical and econometric estimation of underlying parameters of the models.

A series of papers emerging from Nechyba (2000), reviewed in greater detail by Epple and Nechyba (2004) and Nechyba (2006), relied solely on calibration techniques

to give rise to computational models that could replicate the salient features of local school and housing markets while providing a platform for simulating alternative policy regimes. Several insights have emerged from such simulations. First, while they suggest that the move toward less reliance on local sources of funding for public schools has resulted in a narrowing of education quality differences under Tiebout competition, they also provide evidence that substantial inequities persist precisely because of the rationing of nonfinancial resources that emerges from Tiebout sorting across housing markets. This sorting appears to be largely immune to state-level differences in school finance policy. Second, by introducing private school markets that do not explicitly rely on Tiebout sorting in admitting students, the simulations suggest that the decoupling of residential location from school choices has profound impacts on residential location patterns and that Tiebout sorting across public schools cannot be analyzed fully without taking the impact of private school markets into account. This becomes particularly important in the light of increasing efforts to decouple location from schooling within the public sector through nontraditional public schools (such as charter schools) and through explicit formulations of public school choice policies in many cities. Third, simulations of different types of private school competition – ranging from horizontal competition through targeted curricula to vertical competition through cream skimming of the best students from the public system – suggest that the systemic effect of new innovations on public school quality within a framework of residential Tiebout sorting differs widely depending on the precise nature of the competitive forces unleashed through such innovations.

Alongside this emergence of calibrated Tiebout models of education markets, recent attempts have focused on developing new methodologies for using econometric techniques to estimate (rather than calibrate) an increasingly rich set of underlying theoretical models. Three different strains have emerged in this new literature. First, Ferreyra (2007) builds an estimation framework directly on the Nechyba (2000) model while introducing explicit channels for horizontal differentiation of schools. Second, Epple and Sieg (1999) estimate an alternative framework based on the theoretical model of Epple and Platt (1998) that assumes the housing stock within and across jurisdictions adjusts more quickly to changing circumstances. Finally, Bayer and McMillan (2006) develop a more flexible and computationally tractable approach grounded in methodological innovations from industrial organization and focus more directly on the link between household income and race in Tiebout sorting.

Each of these empirical methodologies has provided independent evidence for the importance of Tiebout's insights that households indeed vote with their feet within local economies while uncovering ways in which this leads

to complex patterns of residential sorting which, in turn, forms the basis for admission to traditional public schools. Ferreyra's approach again highlights both the limits of traditional school finance policies as well as the potential role for policy to inhibit or foster horizontal competition through policy innovations that introduce non-residence-based competition. Bayer and McMillan have demonstrated that the inclusion of race into Tiebout models of residential and school choice adds an important and previously neglected level of complexity. Not only does ability to pay determine residential sorting patterns, but race also is independently important as members of different racial groups exhibit a desire to live in communities in which their racial group is represented. For middle-income black households, in particular, this results in a difficult set of trade-offs in many cities – with better public schools located in primarily non-black communities and large representations of black households primarily located in communities with poor public schools.

Tiebout Competition to Enhance Productive Efficiency

The literature highlighted previously has concentrated on sorting patterns that emerge from Tiebout competition linked to education markets. As such, it has focused heavily on the first of the two primary features of Tiebout competition – the sorting of households as households match with communities. At the same time, this literature has remained largely disconnected from the second primary feature of Tiebout competition – the potential for a disciplining market-like force that restrains rent-seeking and thus enhances efficiency. This force has been most notably highlighted by Hoxby (2000) who finds that increased public school competition (as represented by a higher concentration of local school districts within an education market) leads to lower costs and better educational outcomes. While the result has not been uncontroversial, it relates directly to one of the central arguments in favor of decentralized provision of public schooling envisioned by Tiebout.

The empirical question is whether allocating nonfinancial inputs that are linked to households is in essence a zero-sum game or whether more competitive mechanisms for achieving such an allocation result in more efficient production of school quality. Simulations emerging from the Nechyba (2000) model suggest that, if such competitive forces are sufficiently strong, the introduction of more competition through nontraditional school markets can in principle overcome equity concerns that emerge from increased vertical sorting across schools, but that – in the absence of sufficient efficiency gains, the reverse would hold as competition leads to lower nonfinancial inputs in some public schools. In other words, sufficiently strong disciplining forces through competition can give rise to

school reforms that result in a positive sum game, but whether they do depends ultimately on the nature of such reforms and the degree to which such forces are empirically important. Most recently, Bayer and McMillan (2005) provide evidence in favor of strong competitive effects among public schools within a framework that builds on the structural estimation techniques that have come to be prominent in the empirical Tiebout competition literature.

A Partial Divorce between Competition and Tiebout

When confined to the traditional Tiebout setting in which residential location patterns directly determine the allocation of students (and thus the distribution of nonfinancial inputs) into public schools, an inherent tension therefore emerges between the equity concerns from income- (and race)-based rationing of nonfinancial inputs and the potential for productive efficiency gains from decentralized competition. A major thrust in school finance policy over the past decade, however, introduces an added element that emerges from deliberate attempts on the part of policymakers to partially disconnect residential location choices from school assignments. Such policies, whether they take the form of non-residence-based public school competition or increased private school activity subsidized through some form of school voucher, essentially introduce new channels of competition that operate alongside the traditional Tiebout competition through housing markets. The challenge for researchers and policymakers alike lies in identifying potential channels through which such new forces can operate alongside traditional forces to address both efficiency and equity concerns that emerge under Tiebout competition.

The nature of the underlying issue can be put into focus by considering explicitly how different policies would ration access to the scarce resource of seats within good schools. In the traditional Tiebout mechanism, such rationing occurs solely through housing markets that indirectly price access to schools in the public sector. In traditional private schools, on the other hand, this rationing arises both from explicit tuition pricing and from admissions policies defined by them. As greater non-residence-based choice is introduced into the public sector, neither the implicit pricing through housing markets nor the explicit pricing through tuition levels is available as such public-choice schools explicitly reject residential location as a main determinant for admission while at the same time being constrained to offer seats for free to those who are admitted. This has resulted in some role for lotteries to determine allocations of students to public choice schools when such schools are oversubscribed. In charter schools, such lotteries are simply designed for each school independently, but an increasing number of cities and school districts have turned

toward a greater role for lotteries in their assignment mechanisms for all public schools – both those that emerge from the traditional Tiebout-based system and those that have formed more recently as charter schools.

Such districts typically do not rely solely on lotteries to determine where children are assigned. Rather, a Tiebout element is retained as walk-zones are defined around public schools – with households that reside within these zones given preference. Similarly, preference is typically given to siblings of children who are already enrolled in a particular school. Households provide school districts with rankings for schools to indicate their preferences, and assignments then proceed with both these rankings and the priority classes in mind. The resulting assignment mechanisms are therefore structured in tiers that define priority classes, with ties within classes broken through lotteries. Early incarnations of such mechanisms, however, had the unfortunate feature that they were not what economists call strategy proof in that they contained incentives for households to misrepresent their true preferences in attempts to alter the outcomes. For instance, households might not rank their most-preferred school if they viewed their chance of being assigned to that school as low, choosing instead to list less-preferred schools to which they were more likely to be assigned.

Economists have played an important role in refining these assignment mechanisms in large urban districts as such districts move toward increasing non-Tiebout choice within their public systems. In particular, strategy-proof mechanisms built on the same common priority classes have been developed and implemented, with such mechanisms designed to remove the incentives for households to strategically misrepresent their true preferences (Abdulkadiroglu and Sönmez, 2003). Such mechanisms can now be viewed as filling in the gap between the extremes of relying solely on Tiebout competition or relying solely on lotteries. They also provide potential opportunities for the public and private school systems to partially merge into a single system if public funding is extended to private schools to the extent to which private schools accept student assignments through the same mechanism (thereby foregoing the ability to explicitly cream skim students from public schools).

The challenge of such attempts to foster nontraditional channels of competition within an environment of Tiebout competition lies in finding ways of minimizing the undesirable features of competition from vertical sorting (or cream skimming) while maximizing the desirable features from horizontal sorting and from gains in productive efficiency. As the literature that we have reviewed illustrates, vertical sorting emerges naturally under Tiebout competition as ability to pay determines where households will locate in equilibrium and where the nonfinancial inputs that accompany households are located. Such vertical sorting plays a similar role in private schools that

rely on tuition pricing to ration access, with explicit attempts by private schools to cream-skin the best students from public schools reinforcing such sorting further. The introduction of nontraditional competition through mechanisms that reduce the role of residential location in school assignments reduces the vertical sorting that emerges in Tiebout-based school assignments, and, to the extent to which private schools are included in such assignment mechanisms in exchange for public funding, such vertical sorting into private schools is similarly ameliorated. At the same time, such new forms of competition open the door to more desirable channels of competition – that is, those that provide incentives for schools to compete by targeting students through pedagogical or curricula innovations and those that reward productive efficiency.

Conclusion

A rich literature on Tiebout competition in education markets has moved increasingly in the direction of developing empirical computational models that, in turn, have shed light on both the positive and negative aspects of a system that relies on residential location as the primary vehicle for assigning students to public schools. By incorporating nontraditional schools into the analysis, researchers have demonstrated alternative channels of school competition that can play a role within local economies in which Tiebout forces that are closely linked to household income and race play an important role. The challenge for researchers now emerges directly from new policy innovations that emphasize new channels through which competition operates. While the current literature offers hints as to how the resulting competitive forces might change the availability and distribution of educational opportunities, a new generation of models that explicitly analyzes different versions of such policies within the context of Tiebout competition will be needed for further progress to be made.

See also: Competition and Student Performance; Educational Privatization; School Finance: An Overview; The Economics of Catholic Schools; The Economics of Charter Schools; The Economics of Parental Choice; The Efficacy of Educational Vouchers.

Bibliography

- Abdulkadiroglu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review* **93**, 729–747.
- Bayer, P., Ferreira, F., and McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy* **115**(4), 588–638.
- Bayer, P. and McMillan, R. (2005). Choice and competition in local education markets. *NBER Working Paper 11802*.

- Bayer, P. and McMillan, R. (2006). Racial sorting and neighborhood quality. *NBER Working Paper 11813*.
- Black, S. (1999). Do better schools matter? Parental valuations of elementary education. *Quarterly Journal of Economics* **114**, 577–599.
- Epplé, D. and Nechyba, T. (2004). Fiscal decentralization. In Henderson, V. and Thisse, J. (eds.) *Handbook of Regional and Urban Economics*, vol. 4, pp 2423–2480. Amsterdam: North Holland.
- Epplé, D. and Sieg, H. (1999). Estimating equilibrium models of local jurisdictions. *Journal of Political Economy* **107**(4), 645–681.
- Edel, M. and Sclar, E. (1974). Taxes, spending and property values: Supply adjustment in a Tiebout-Oates Model. *Journal of Political Economy* **82**(5), 941–954.
- Ferreira, M. (2007). Estimating the effects of private school vouchers in multi-district economies. *American Economic Review* **97**, 789–817.
- Hoxby, C. (2000). Does competition among public schools benefit students and taxpayers? *American Economic Review* **90**, 1209–1238.
- Loeb, S. and Page, M. (2001). Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation. *Review of Economic Studies* **82**, 393–408.
- Nechyba, T. (2000). Mobility, targeting and private school vouchers. *American Economic Review* **90**, 130–146.
- Nechyba, T. (2006). Income and peer quality sorting in public and private schools. In Hanushek, E. and Welch, F. (eds.) *Handbook of Economics of Education*, vol. 2, pp 1329–1370. Amsterdam: North Holland.
- Oates, W. (1969). The effects of property taxes and local public spending on property values: An empirical study of tax capitalization and the Tiebout Hypothesis. *Journal of Political Economy* **77**, 957–971.
- Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of Political Economy* **64**, 416–424.

Further Reading

- Fischel, W. (2006). *The Tiebout Model at Fifty: Essays in Public Economics in Honor of Wallace Oates*. Cambridge, MA: Lincoln Institute Press.

ECONOMICS OF EDUCATION – PRODUCTION OF SCHOOLING

Contents

Desegregation, Academic Achievement and Earnings
Education Production Functions: Concepts
Education Production Functions: Developed Country Evidence
Education Production Functions: Evidence from Developing Countries
Family Environment in the Production of Schooling
Neighborhoods and Peers in the Production of Schooling
Teacher Quality in Education Production
The Economics of Class Size
The Economics of Early Childhood Interventions

Desegregation, Academic Achievement and Earnings

S Rivkin, Amherst College, Amherst, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Supreme Court Decisions and Desegregation Methods

Beginning with the decision in *Brown*, states lost the power to separate blacks and whites into two entirely separate school systems, a power affirmed by the 1896 ruling in *Plessey v. Ferguson*. As Welch and Light (1987) and Lutz (2005) summarize, there have been a number of Supreme Court decisions following *Brown* that have shaped the rules governing the allocation of students into schools. The early post-*Brown* decisions expanded the powers of those seeking to desegregate the schools and substantially broadened the arrangements classified as illegal. However, the 1974 ruling in *Milliken v. Bradley* made it difficult to seek a remedy for segregation across districts, and the 1975 ruling in *Morgan v. Kerrigan* sanctioning magnet schools as a valid desegregation method weakened the ability of the courts to mandate the implementation of coercive desegregation plans. Finally, the 1991 and 1995 decisions that require districts merely to take all practical steps to end the legacy of segregation (*Board of Education of Oklahoma v. Dowell*) and not to hold districts responsible for low achievement (*Missouri v. Jenkins*) reveal a movement away from mandatory desegregation as a primary means to increase school quality, at least at the federal level.

The recent Connecticut State Supreme Court ruling *Sheff v. O'Neill* (1996) that requires Connecticut officials to desegregate the schools may signal a shift of the desegregation battle to state courts, similar to the legal battle

over school finance equity. If this comes to pass, the standards for desegregation will vary by state, just as do those for funding inequality.

Districts have used a number of techniques to desegregate schools depending upon the legal precedents in effect, severity of segregation, residential housing patterns, availability of alternative public schools, and the extent of community resistance. Circumstances sometimes necessitated the use of more than one technique, while in other cases, a simpler construct was adequate.

Desegregation techniques can be classified as voluntary or involuntary depending upon whether students are permitted to choose the school they will attend; the distinction does not reflect whether or not the courts required desegregation. Welch and Light (1987) divide plans into six categories, three voluntary and three nonvoluntary. The voluntary techniques include open enrolment, magnet, and other transfer programs, while the involuntary plans include neighborhood-attendance zones, rezoning, and pairing and clustering (combining two or more schools into a single-attendance zone in which all students attend early grades at one school and later grades at another).

Trends in School Enrolment Patterns: 1968–2000

This section uses office of civil rights public elementary and secondary school enrolment data to describe school-enrolment patterns for non-Hispanic blacks (see Rivkin

Table 1 Racial composition of public elementary and secondary schools, by region and the United States as a whole: 1968, 1980, 1988, and 2000 (percentages)

Region	Percent White				Percent Black				Percent other ^a			
	1968	1980	1988	2000	1968	1980	1988	2000	1968	1980	1988	2000
Northeast	84.4	78.3	75.8	67.5	11.5	13.6	12.4	15.3	4.1	8.1	11.8	17.2
North Central	87.9	83.8	83.6	76.6	10.6	12.5	11.4	14.7	1.5	3.7	5.0	8.7
South	70.1	65.9	63.7	56.3	25.3	25.4	25.3	26.4	4.6	8.7	11.0	17.3
West	78.2	67.1	62.6	50.2	6.3	6.7	5.7	6.5	15.5	26.2	31.7	43.3
National	79.9	73.3	70.7	61.4	14.8	16.1	15.2	17.1	5.3	10.6	14.1	21.5

^aResidual category including Hispanic, Asian, and Native American.

Table 2 Average percentage of blacks' schoolmates who are white by region: 1968–2000

Percent of schoolmates Who are white	Distribution of enrolment (percentage)			
	1968	1980	1988	2000
United States	22.3	36.2	36.2	31.1
Region				
Northeast	33.7	27.8	26.9	24.9
North Central	23.3	30.9	32.1	29.1
South	17.9	40.8	40.1	33.7
West	26.7	34.6	36.0	30.7

and Welch (2006) for a detailed description of these trends across the school proportion black distribution) and whites for the years 1968, 1980, 1988, and 2000 (see Welch and Light (1987) for a comprehensive description of the data). Three interrelated pieces of information characterize school enrolment patterns, and I document changes over time in all three: demographic composition, segregation, and interracial contact. Note that segregation refers to the extent to which students are mixed conditional on the overall demographic shares of the district, state or other geographic entity, and interracial contact refers to the share of schoolmates who are white. Notice that overall demographic composition and degree of segregation jointly determine the level of interracial contact. (The focus here is exclusively on the enrolment patterns of non-Hispanic blacks and whites because the nation's integration policies have focused on the segregation of blacks.).

Table 1 shows the demographic composition of public schools by region. Between 1968 and 1988 the decline in white enrolment nationally as a percentage of the total was roughly offset by increases for Hispanics and Asians, while between 1988 and 2000 the continued decline in the white enrolment share was offset by increases in both the black (15.2–17.1%) and the Hispanic and Asian (14.1–21.5%) enrolment shares.

There are regional differences and similarities in both the magnitude and timing of demographic change. The decline in the white enrolment share was smaller in the

north central region than in the rest of the country, particularly prior to 1988. On the other hand, all regions experienced their largest white enrolment share decline following 1988, driven in part by the very high rate of migration into the US from Asian, Latin American, and West Indian countries.

Although demographic changes tempered the effectiveness of desegregation efforts to raise black/white contact in the schools, **Table 2** documents the substantial increase following 1968 in the average percentage of blacks' schoolmates who are white (exposure index). The average percentage of blacks' schoolmates who are white, also known as the exposure index, equals $\sum_{i=1}^n B_i^* PW_i/B$, where B_i equals the number of blacks in school i , PW_i equals the proportion of white students in school i , and B equals the number of black students in the region or nation. Nationally, exposure to whites increased by more than 50% between 1968 and 1980, rising from 22.3% to 36.2%. Between 1980 and 1988, exposure remained relatively stable prior to declining by 5 percentage points between 1988 and 2000. For the nation as a whole, the decline in segregation between 1968 and 1988 swamped the loss of whites on a national level, while following 1988, the demographic changes reduced exposure to whites.

The subsequent rows reveal regional differences in the timing of school enrolment changes. The region with the highest average exposure to whites in 1968, the northeast, had already become the region with the lowest average exposure to whites by 1980, while the south went from being the region with the lowest exposure to the region with the highest exposure during this same period.

The contrast between changes in overall demographic composition and the average exposure to whites provides clear evidence of a decline in segregation, and the dissimilarity index is used to document changes in both school and district segregation. School segregation comes from race differences in both the distribution of students among districts and school attendance patterns, while aggregation of the data to the district level isolates race differences in the distribution of students among districts. This is a crude decomposition of overall school segregation into housing and school attendance policy

components, as private school enrolment and housing patterns within districts also affect segregation levels conditional on district enrolment policies. Yet, since the evidence suggests that private school enrolment changes have not had a substantial impact on segregation at the regional or national level, this decomposition does provide a sense of the importance of housing segregation in determining school enrolment patterns. (Rivkin and Welch (2006) provide information on private school enrolment.)

The school (district) index measures the degree to which the distribution of students among schools (districts) deviates from a distribution in which all schools (districts) have identical enrolment shares. The dissimilarity index varies from zero to one, with a value of one indicating complete segregation and a value of zero indicating complete integration. The magnitude of the index reflects the share of blacks (or whites) that would have to switch schools (districts) to achieve complete integration. (Formally, the dissimilarity index measuring school segregation between blacks and whites is defined as, $\sum t_s |p_s - p| / 2Tp(1 - p)$, where the subscript s indicates school, t_s is total enrolment in school s , p_s is the black share of enrolment in school s , p is the black enrolment share in the nation (or region), and T is total enrolment in the nation (or region).)

The left panel of the top row of **Table 3** shows that the school-level dissimilarity index for the US as a whole declined from 81.2 to 71.0 between 1968 and 1980, remained almost constant during the 1980s and declined to 66.7 between 1988 and 2000, but the remaining rows reveal substantial differences in the timing and magnitude of changes in segregation that match up with the observed differences in the changes in interracial contact. The south made by far the most headway against segregation during the 1970s but experienced little subsequent change throughout the distribution, while the desegregation gains in the north central region were smaller but steadier. The west also experienced a substantial segregation decline during the 1970s and additional gains during the 1980s and 1990s. Consistent with the decline in exposure to whites, the northeast experienced virtually no change in segregation through 1988 and a modest decline during the 1990s.

The right panel presents district segregation curves that show that the remaining segregation in 2000 resulted primarily from the allocation of students among districts. Even if all schools had achieved complete integration without changing the distribution of students among districts, the school-level dissimilarity index would have declined from only 76 to 74 in the northeast, 76.5 to 74.5 in the north central, 58.8 to 49.2 in the south, and 64.3 to 63.7 in the west. Clotfelter (1999) documents a similar degree of segregation using metropolitan area rather than district as the unit of analysis.

In each region, there was substantial segregation at the district level throughout the time period. Each experienced some degree of additional segregation by district during the active desegregation period of the 1970s, but they differ in the extent to which segregation lessened in the subsequent decades. The south in particular was notable for its lack of progress following 1980.

Desegregation and White Enrolment Decline

Table 4 describes major studies of enrolment shifts in response to school desegregation, beginning with Coleman *et al.* (1975). Although the exact methods varied, the predominant approach was to compare enrolment changes around the time of program implementation with those several years before and several years after, allowing for differential effects by a number of factors. The pattern of findings up to 1985 provides no consensus on the magnitude or persistence of white flight from school desegregation (see Rivkin and Welch (2006) for additional detail). Although most studies reported sizeable enrolment changes surrounding program implementation, these changes were often not related to program characteristics in the expected ways. Welch and Light (1987) set out to resolve these uncertainties through an analysis of 125 large districts. They compiled 16 years of data on enrolments and desegregation program status and examined in detail, the changes in white enrolment surrounding the implementation of 116 major plans during the period of study.

Table 3 School and district dissimilarity indexes by region: 1968–2000

	School				District			
	1968	1980	1988	2000	1968	1980	1988	2000
<i>School districts</i>								
Northeast	76.9	78.8	78.5	76.0	70.9	76.7	76.7	74.0
North Central	85.7	80.1	78.9	76.4	74.5	77.4	76.9	74.5
South	80.1	57.3	57.1	58.8	44.2	48.9	49.9	49.2
West	81.4	70.6	66.9	64.3	65.7	66.5	63.4	59.6
National	81.2	71.0	70.4	68.7	63.8	66.2	66.2	63.7

Table 4 Summary of selected research on white flight

Article	Findings
Coleman <i>et al.</i> (1975)	Plans reduced segregation but increased the white enrolment decline
Clotfelter (1976)	A higher nonwhite student population share increased growth of private school enrolment for whites
Taeuber and Wilson (1978)	No lasting desegregation effect on rate of white enrolment decline; schools not often cited as a primary reason for moving
Rossell (1978)	Significant implementation-year decline in white enrolment, but small or nonexistent longer term effects
Armor (1978)	Enrolment fell relative to its projection, dropping sharply following plan implementation and continuing to fall thereafter
Farley <i>et al.</i> (1980)	Confirmed findings of large contemporaneous white enrolment decline following plan implementation
Wilson (1985)	Enrolment response was short lived and did not depend upon the attributes of the desegregation program
Welch and Light (1987)	White enrolment dropped below trend when desegregation programs were introduced, and the effect appeared to persist; effect magnitudes varied with plan type and district characteristics
Rivkin (1994)	White enrolment declined substantially in virtually all large central city districts regardless of the scope of desegregation efforts
Reber (2003)	Desegregation plans had a significant effect on white enrolment, and the type of program and number of nearby school districts were important predictors of white response
Lutz (2005)	Dismissal of desegregation orders moderately reduced black exposure to whites and did not reverse white flight in the short term
Clotfelter <i>et al.</i> (2005)	Dismissal of desegregation orders increased segregation in southern districts with a lag

Welch and Light find that white enrolment also declined much more in the year of plan implementation, although in contrast to the change in segregation the average enrolment decline picked up the year prior to implementation and remained above the preplan level in the year following implementation. The higher enrolment losses in the years following implementation are consistent with national enrolment data showing that losses were greater during the late 1970s than in earlier periods. But the fact that enrolment losses are much greater during implementation provides strong evidence of the existence of an enrolment response to segregation, just as the timing of the changes in the dissimilarity index document desegregation's substantial effect on district attendance patterns.

Not surprisingly, the most comprehensive plans led to larger declines in segregation. Both before and after the Swann (1970) decision, stating that racially identifiable schools must cease to exist and sanctioning the use of district-wide busing plans that combined rezoning with pairing and clustering, produced the largest decline in the dissimilarity index. A second clear pattern emerging in the post-Swann period is that plans using pairing and clustering produced larger average changes in enrolment (prior to Swann there is no obvious pattern in enrolment changes, likely because this was a period of trend reversal in total white enrolment that varied geographically).

The finding that pairing and clustering leads to greater departures from trend than rezoning, magnets, and other voluntary plans reflects qualitative differences among the desegregation techniques. Not only is pairing and clustering mandatory, but it typically requires that students travel greater distances than under rezoning, the other mandatory program type. All in all, it is not surprising that the mandatory plan type that is most disruptive for

students produces the largest average changes in both segregation and white enrolment.

Welch and Light also investigate differences in the white enrolment response on the basis of region and district urban status in order to allow for different responses in the south and in countywide districts. Countywide districts are particularly interesting, because the districts typically include suburban areas where white students are concentrated. These geographically larger districts tend to raise both the cost of moving out of the district in terms of the additional commute time to work and average distance students must be bused to achieve a given amount of desegregation. Consequently, it is unclear *a priori* whether countywide districts should experience a larger white enrolment response to desegregation-plan implementation, though families in these districts do have an unambiguously greater incentive to switch to a private school if one is available. In fact, Welch and Light (1987) show that departures from trends in white enrolment tended to be smaller for the countywide districts than for the urban southern and nonsouthern districts despite the fact that the countywide districts experienced much larger average reductions in segregation.

The discontinuous nature of the white enrolment changes following the implementation of desegregation programs provides strong evidence of a causal link between desegregation and white enrolment declines. Nonetheless, the possibility remains that other factors could produce a similar pattern. As an alternative to the approach used by Welch *et al.* (1987), Reber (2003) uses differences in the timing of implementation for the same set of school districts to estimate white enrolment changes around the time of implementation. The panel data method essentially compares changes in school districts around the time of

implementation to the national average white enrolment changes over corresponding years. Her results confirm the findings in Welch *et al.* (1987). Moreover, she finds that white flight appeared to increase with the number of districts in a metropolitan area, supporting the notion that the higher relocation costs of geographically larger districts more than offset larger increases in school commute times in terms of the incentive to switch districts.

Importantly, desegregation-induced enrolment declines may appear quite large at the time of program implementation but may have very little impact over the long term given the widespread suburban migration. The evidence in Rivkin (1994) suggests that desegregation programs may not have been a primary contributor to white enrolment loss over the longer term. Between 1968 and 1988, the distribution of white enrolment loss in the 10 large urban districts that experienced the largest decline in segregation is quite similar to the distribution for the 10 large urban districts that experienced the smallest decline in segregation. This holds despite the fact that the 10 nonintegrating districts began with much lower white enrolment shares and thus lower potential losses. It appears that as the time horizon increased, the deviations from trend caused by the implementation of a desegregation plan diminished in importance, relative to persistent long-term influences.

The ubiquitous white enrolment decline in large urban districts across the country also imposes a severe limit on district efforts to expand interracial contact. Even if whites in the 2000s were to respond much more favorably to expanded desegregation programs than did whites in the 1970s, the low white enrolment share in virtually all large districts precludes sizable increases in exposure to whites. As the district and school-level segregation curves and measures highlighted in the previous section show, it is the sorting of families among communities rather than district attendance policies that limit the extent of interracial contact in the schools as long as the Milliken (1974) finding, that there can be no interdistrict relief without an interdistrict violation, remains the law of the land.

Moreover, given recent Supreme Court decisions lowering barriers to the dismissal of desegregation orders, a contraction of desegregation efforts is much more likely than an expansion. Lutz (2005) finds that the number of years since the dismissal of a court-ordered plan is positively related to the change in segregation, and this in turn has led to a modest reduction in black exposure to whites. Moreover, Lutz finds no immediate white enrolment response to the dismissal in the form of reverse white flight. The findings in Clotfelter *et al.* (2005) are consistent with those of Lutz, particularly in the sense that resegregation is not immediate but increases over time. In the context of overall changes during this period, this evidence suggests that the dismissal of desegregation orders offset other factors that were tending to reduce

segregation. It is important to recognize that longer term effects may differ and the responses in the affected districts may not be representative of all districts with court-ordered plans currently in place.

Desegregation Program Effects on Academic and Economic Outcomes

Despite the movement of large numbers of whites out of central city school districts, school desegregation succeeded in dramatically increasing black exposure to whites. Moreover, many programs reallocated students among schools, exposing blacks and whites to different teachers, facilities, and other factors thought to affect school quality. Supporters of desegregation expected these changes to improve the quality of education for blacks and reduce the racial gap in school quality and future economic and social outcomes. (See Rivkin and Welch (2006) for additional detail.)

There are also reasons to believe that student or school efforts to neutralize desegregation, longer commute distances, and other adverse side effects might limit or even eliminate the benefits of school desegregation. Clotfelter *et al.* (2003) find evidence of substantial within-school segregation in North Carolina junior high and high schools but not in elementary schools. Clotfelter (2002) also finds some racial differences in the distributions of blacks and whites across extracurricular activities, although these activities do provide a vehicle for interracial contact. Such resegregation within schools not only reduces interracial contact, it may also lessen the improvement in teacher and school quality actually experienced by blacks bused to a non-neighborhood school. (Rivkin (2000) finds that busing and other desegregation programs appear to weaken the link between the quality of education experienced by blacks and nonblacks in the same school.) Finally, evidence suggests that busing students out of their neighborhoods decreases parental participation (Edwards, 1993; Leake and Faltz, 1993).

Consequently, the question of whether desegregation benefited blacks must be answered empirically, and that answer likely depends on characteristics of both programs and students. This section describes existing research on racial composition and desegregation effects on achievement and other outcomes. It begins with a discussion of relevant methodological issues prior to summarizing the findings.

Methodological Issues

Families and various entities including local school districts and the courts combine to determine the allocation of students among schools and districts. This nonrandom grouping of students in districts and schools provides a formidable impediment to the identification of the

relationship between academic, social, and economic outcomes on the one hand and desegregation efforts and school racial composition on the other. Manski (1993), Moffitt (2002), and others consider these issues in the more general context of the estimation of peer effects; here the focus is on those specific methods used in the study of racial composition and desegregation.

Equation [1] models outcome O for student i in grade G in school s as a function of family, desegregation program status, and peer influences:

$$O_{iGs} = X_{iGs}\beta + D_{Gs}\gamma + \bar{P}_{(-i)Gs}\lambda + D_G\bar{P}\theta + e_{iGs} \quad [1]$$

where D is an indicator variable for desegregation program status (or a vector of indicators for program type), \bar{P} is peer quality measured by average characteristics of schoolmates (individual i is omitted from the calculation); X is a vector of family background variables; and e is the error term. Note that the effects of peer composition are permitted to vary by desegregation program status.

The identification of γ and λ is complicated by the nonrandom sorting of children into schools and classrooms and the cumulative nature of schooling. Any correlation between \bar{P} and D on the one hand and current or past unobserved factors that affect outcomes on the other leads to biased estimates of γ and λ . Failure to account for the dynamic character of the learning process can also introduce bias and complicate the interpretation of the estimated effects. (Rivkin (2005) examines the identifying assumptions regarding the depreciation of knowledge for a number of education-production function specifications.) A number

of different methods have been used to identify desegregation program and racial composition effects including random-assignment experiments, ordinary least squares regression, value-added models, and panel data techniques.

A separate issue from identification is the relevance of the results for education policy. Most desegregation programs were implemented around 1970, and much of the research on desegregation effects on achievement comes from that period. Not only have demographic changes and residential movements dramatically altered the composition of schools, but also school-finance reform, shifts in student and teacher attitudes regarding race, expansion of school accountability systems, charter schools, and other public school reforms almost certainly have affected the benefits of both expanded interracial contact and attendance at a non-neighborhood public school. The heterogeneity of both desegregation program types and the environments that form the context of racial interactions introduce additional variation into the likely effects of specific interventions. Nevertheless, the literature provides important information on the contribution of desegregation to black academic and economic progress, the likely benefits of program expansion, and the potential harm from the dismissal of desegregation orders.

Summary of Results

The studies summarized in **Table 5** provide mixed evidence on the effects of school desegregation programs and

Table 5 Selected research on effects of desegregation and racial composition on student outcomes

Article	Findings
Coleman <i>et al.</i> (1966)	Higher percentage of black school-mates lowered achievement
Crain (1970)	Exposure to whites raised income and occupational achievement by leading to greater interaction with whites
Cook (1984)	Desegregation had at most a small effect on academic achievement
Armor (1992)	School desegregation contributed little to the closing of the black/white achievement gap
Boozer <i>et al.</i> (1992)	A higher proportion of high-school schoolmates who were black was associated with fewer years of schooling, a less-integrated work environment and lower wages
Grogger (1996)	A higher proportion of high-school schoolmates who were black reduced wages, and the size of the effect increased between 1979 and 1986
Rivkin (2000)	Little or no effect of exposure to whites on achievement, school attainment, or earnings regardless of type of desegregation plan
Hoxby (2000)	Higher black enrolment share lowered achievement, and the effect was larger in schools with a mid-range proportion black
Hanushek <i>et al.</i> (2004)	Higher black enrolment share lowered achievement, and peer average achievement accounts for little of the proportion black effect
Guryan (2004)	Desegregation accounted for roughly one half of the decline in the black high-school dropout rate between 1970 and 1980
Angrist and Lang (2004)	Little or no evidence that urban, black, and Hispanic students bused to a suburban district adversely affected achievement of non-black or Hispanic suburban students
Lutz (2005)	Desegregation program dismissal raises high-school dropout rates in northern but not southern school districts
Card and Rothstein (forthcoming)	School desegregation has little effect on SAT scores, possibly because of resegregation within schools

school racial composition. On the one hand, the random-assignment experiments of the benefits of busing reveal little evidence that desegregation substantially increased black student achievement. On the other hand, observational studies generally support the hypothesis that desegregation and interracial contact are beneficial, particularly for longer term outcomes including school attainment and earnings.

In a comprehensive summary of random-assignment studies of school desegregation effects on reading and mathematics achievement in early grades, Cook (1984) concludes that meta-analyses of the results of a number of studies supports the view that the effects were quite small or even zero. There are a number of factors, however, that raise doubts about this interpretation of the evidence. First, the small sample sizes common in studies of single desegregation programs limit the power of the estimates, and this makes it very difficult to identify small but potentially important effects on test scores. Second, any heterogeneity in the effects of different types of desegregation programs implies that the estimates depend in part on the composition of programs included in the studies. Third, it may take a few years to work out the glitches of new programs.

Nevertheless, the contrast between the generally small estimated effects of desegregation on elementary school achievement in random-assignment studies and the generally much larger effects on achievement and longer term outcomes in observational studies is striking. Beginning with the landmark legislatively mandated report, *Equality of Economic Opportunity* (Coleman *et al.*, 1966), the bulk of observational research finds that desegregation and interracial contact raised achievement, schooling, and earnings, and in some cases the estimated effects are quite large.

A number of alternative explanations could separately or together reconcile these disparate results. First is the aforementioned concern that the random-assignment studies failed to uncover the true benefits of desegregation. Second is the possibility that desegregation raises academic attainment but not test scores. This is supported by Guryan (2004), Lutz (2005), and Card and Rothstein (2006) but not Rivkin (2000). Third is the possibility that desegregation program induced interracial contact is less beneficial than contact resulting from residential integration, which is consistent with the positive effects of exposure to whites found in Hanushek *et al.* (2004) and Hoxby (2000) but not consistent with Guryan (2004) who finds that expanded interracial contact contributes to the positive desegregation effect on schooling.

A final and certainly not mutually exclusive explanation is that the observational studies fail to account for confounding factors and other sources of bias despite the use of a variety of empirical methods. Nonrandom sample selection, inadequate controls for confounding factors, retrospective information on school racial composition

that may be influenced by one's own experience, and the use of cross-sectional information on school attendance such as the US census that cannot identify recent movers into the district and often lacks geographic detail so that the allocation of students to districts tends to include surrounding suburbs not part of desegregation efforts all raise concerns. Some studies depend on the identification of appropriate counterfactuals (Guryan, 2004; Lutz, 2005), others rely on observed controls (Grogger, 1996), others use patterns of cohort differences in racial composition (Hanushek *et al.* 2006; Hoxby, 2000), some use aggregate information on racial composition (Card and Rothstein, 2006; Rivkin, 2000) and the study by Boozar *et al.* (1992) uses state of birth by cohort dummies as instruments for school racial composition. A final concern is that some of these empirical approaches are not consistent with a cumulative model of knowledge acquisition such as that developed by Todd and Wolpin (2003). Regardless of the exact approach, each makes strong assumptions that must be satisfied to produce consistent estimates.

An important issue for policy concerns the underlying determinants of any relationship between desegregation or interracial contact on the one hand and outcomes on the other. Are effects driven by peer influences, teacher quality, school factors, or some combination of the three? Do effect magnitudes vary by black student characteristics such as initial academic achievement? Do the reduced form estimates often based on specific states or experience over 20 years ago provide valuable information on the likely effects of additional desegregation efforts or the dismissal of plans currently in place?

See also: Education Production Functions: Concepts; Education Production Functions: Developed Country Evidence; Neighborhoods and Peers in the Production of Schooling; The Economics of Charter Schools; The Economics of Parental Choice; The Efficacy of Educational Vouchers; Tiebout Sorting and Competition.

Bibliography

- Ainsworth-Darnell, J. W. and Downey, D. B. (1998). Assessing the oppositional culture explanation for racial/ethnic differences in school performance. *American Sociological Review* **63**, 536–553.
- Angrist, J. and Lang, K. (2004). Does school integration generate peer effects? Evidence from Boston's Metco Program. *The American Economic Review* **94**, 1613–1634.
- Armor, D. (1978). White flight, demographic transition, and the future of school desegregation. *The Rand Paper Series* # 5931. Santa Monica, CA.
- Armor, D. (1992). Why is black educational achievement rising? *Public Interest* **108**, 65–80.
- Austen-Smith, D. and Fryer, R. G. Jr. (2003). The economics of 'acting white'. *National Bureau of Economic Research, Working Paper* # 9904.
- Bishop, J. H., Mane, F., Bishop, M., and Moriarty, J. (2001). The role of end-of-course exams and minimal competency exams in standards-

- based reforms. In Ravitch, D. (ed.) *Brookings Papers on Education Policy 2001*, pp 267–345. Washington, DC: Brookings.
- Boozer, M. A., Krueger, A. B., and Wolkon, S. (1992). Race and school quality since Brown v. Board of Education. In Baily, M. N. and Winston, C. (eds.) *Brookings Papers: Microeconomics*, pp 269–338. Washington, DC: Brookings.
- Brown v. Board of Education (1954). 347 U.S. 483.
- Card, D. and Rothstein, J. (2006). Racial segregation and the black–white test score gap. *Journal of Public Economics* **91**, 11–12.
- Clark, K. and Clark, M. (1939). The development of consciousness of self and the emergence of racial identity in Negro children. *Journal of Social Psychology* **10**, 591–599.
- Clotfelter, C. T. (1976). School desegregation, ‘tipping,’ and private school enrollment. *Journal of Human Resources* **22**, 29–50.
- Clotfelter, C. T. (1999). Public school segregation. *Land Economics* **75**, 487–504.
- Clotfelter, C. T. (2002). Interracial contact in high school activities. *Urban Review* **34**, 25–46.
- Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2002). Who teaches whom? Race and the distribution of novice teachers. *Paper Presented at the Annual Meetings of the American Economic Association*. Atlanta, GA.
- Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2005). Federal oversight, local control, and the specter of resegregation in southern schools. *NBER Working Paper 11086*.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., et al. (1966). *Equality of Educational Opportunity*. Washington, DC: US Government Printing Office.
- Coleman, J. S., Kelley, S. D., and Moore, J. A. (1975). Trends in school integration, 1968–73. *Urban Institute Paper #722-03-01*. Washington, DC.
- Cook, P. J. and Ludwig, J. (1997). Weighing the burden of ‘acting white’: Are there race differences in attitudes toward education? *Journal of Policy Analysis and Management* **16**, 256–278.
- Cook, T. D. (1984). What have black children gained academically from school desegregation? A review of the metaanalytic evidence. In *School Desegregation*. Washington, DC: National Institute of Education.
- Crain, R. (1970). School integration and occupational achievement of Negroes. *American Journal of Sociology* **75**, 593–606.
- Edwards, P. (1993). Before and after school desegregation: African American parents involvement in schools. *Educational Policy* **7**, 340–369.
- Farley, R., Richards, T., and Wurdock, C. (1980). School desegregation and white flight: An investigation of competing models and their discrepant findings. *Sociology of Education* **53**, 123–139.
- Ferguson, R. F. (1998a). Comment (on Cook and Ludwig). In Jencks, C. and Phillips, M. (eds.) *The Black–White Test Score Gap*, pp 394–397. Washington, DC: Brookings.
- Ferguson, R. F. (1998b). Teachers’ perceptions and expectations and the black–white test score gap. In Jencks, C. and Phillips, M. (eds.) *The Black–White Test Score Gap*, pp 273–317. Washington, DC: Brookings Institution.
- Fordham, S. and Ogbu, J. (1986). Black students’ school success: Coping with the burden of ‘acting white’. *Urban Review* **58**, 54–84.
- Fryer, R. G. Jr. and Levitt, S. D. (2003). The causes and consequences of distinctively black names. *National Bureau of Economic Research, Working Paper 9938*.
- Grogger, J. T. (1996). Does school quality explain the recent black/white wage trend? *Journal of Labor Economics* **14**, 231–253.
- Guryan, J. (2004). Desegregation and Black dropout rates. *American Economic Review* **94**, 919–943.
- Hanushek, E., Kain, J., and Rivkin, S. (2004). New Evidence about Brown versus Board of Education: The Complex Effects of School Racial Composition on Achievement. Unpublished manuscript.
- Hoxby, C. (2000). Peer effects in the classroom: Learning from Gender and race variation. *National Bureau of Economic Research, Working Paper #7867*.
- Iceland, J. and Weinberg, D. H. (2002). *Racial and Ethnic Residential Segregation in the United States: 1980–2000, Census 2000 Special Reports*. Washington, DC: US Bureau of the Census.
- Lazear, E. P. (2001). Educational production. *Quarterly Journal of Economics* **116**, 777–803.
- Leake, D. and Faltz, C. (1993). Do we need to desegregate all of our black schools? *Educational Policy* **7**, 370–387.
- Lutz, B. (2005). Post Brown vs. the Board of Education: The effects of the end of court-ordered desegregation. *FEDS Working Paper No. 2005-64*. Washington, DC: Federal Reserve Board – Research Division.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* **60**, 531–542.
- Massey, D. and Denton, N. (1993). *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.
- McWhorter, J. H. (2000). *Losing the Race: Self-Sabotage in Black America*. New York, NY: Free Press.
- Moffitt, R. A. (2001). Policy interventions, low-level equilibria, and social interactions. In Durlauf, S. and Young, H. P. (eds.) *Social Dynamics*, pp 45–82. Cambridge, MA: MIT Press.
- Ogbu, J. U. (2003). *Black American Students in an Affluent Suburb: A Study of Academic Disengagement*. Mahwah, NJ: Erlbaum.
- Orfield, G. and Monfort, F. (1992). Status of school desegregation: The next generation. *Report to the National School Boards Association*. Cambridge, MA: Harvard University Press.
- Plessey v. Ferguson (1896). 163 U.S. 537.
- Reber, S. J. (2003). *Court-Ordered Desegregation: Successes and Failures Integrating American Schools Since Brown*. Los Angeles, CA: School of Public Policy and Social Research, UCLA.
- Rivkin, S. G. (1994). Residential segregation and school integration. *Sociology of Education* **67**, 279–292.
- Rivkin, S. G. (2000). School desegregation, academic attainment, and earnings. *Journal of Human Resources* **35**, 333–346.
- Rivkin, S. G. (2005). Cumulative Nature of Learning and Specification Bias in Education Research. Unpublished manuscript.
- Rivkin, S. G. and Welch, F. (2006). Neighborhood segregation and school integration. In Hanushek, E. A. and Welch, F. (eds.) *Handbook of the Economics of Education*. Amsterdam: North Holland.
- Rossell, C. (1978). *The Unintended Impacts of Public Policy: School Desegregation and Resegregation*. Durham, NC: Institute of Political Science, Duke University.
- Steele, C. M. and Aronson, J. (1998). Stereotype threat and the test performance of academically successful African Americans. In Jencks, C. and Phillips, M. (eds.) *The Black–White Test Score Gap*, pp 401–427. Washington, DC: Brookings Institution.
- Symonds, W. (2004). Brown v Board of Ed: A Bittersweet birthday. *Business Week* May 17.
- Taeuber, K. and James, D. (1982). Racial segregation among public and private schools. *Sociology of Education* **55**, 133–143.
- Taeuber, K. and Wilson, F. (1978). The demographic impact of school desegregation policy. In Kraft, M. and Schneider, M. (eds.) *Population Policy Analysis*, pp 135–152. Lexington, MA: Lexington Books.
- Thernstrom, S. and Thernstrom, A. (2003). *No Excuses: Closing the Racial Gap in Learning*. New York, NY: Simon and Schuster.
- Weiler, J. (1998). Recent changes in school desegregation. *ERIC/CUE Digest* **133**, 1–8.
- Welch, F. (1985). A reconsideration of the impact of school desegregation programs on public school enrollment of White students, 1968–1976. *Sociology of Education* **60**, 215–221.
- Welch, F. and Light, A. (1987). *New Evidence on School Desegregation*. Los Angeles, CA: Unicon Research Corp.
- Wilson, F. (1985). The impact of school desegregation on White Public School enrollment, 1968–1976. *Sociology of Education* **58**, 137–153.

Education Production Functions: Concepts

D N Harris, University of Wisconsin–Madison, Madison, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Education production function – This constitutes all combinations of inputs that produce a given set of outputs. In education, outputs are typically measured by achievement scores and graduation rates. Inputs include both school and family factors that influence outcomes.

Value-added model – A method of estimating the impact of school programs and inputs that uses data on individual students linked over time so that each individual's achievement growth can be calculated. Value-added models can be used to estimate effects of programs and inputs (e.g., teacher experience) as well as the effects of individual teachers and schools. The concept of value-added is rooted in economics and refers to the difference in the economic value of outputs and inputs.

Introduction

The production function is one of the core concepts of microeconomics and its application to education has been one of the central research activities of quantitative education researchers, especially economists of education, for many decades. The education production function (EPF) is implicitly part of any research that attempts to establish a statistical relationship between education resources (e.g., class size) and measures of student outcomes (e.g., scores on standardized achievement tests). Results from these studies have had a significant impact on policy debates about how, and how much, public funding should be provided for education. The quasi-experimental research methods associated with EPFs have also become an increasingly common approach to education program evaluation, influencing a wide range of policies beyond education funding.

The importance of EPF research has also grown with the recent availability of rich data sources that include test scores for each student in each year and allow the scores of each student to be linked over time and to individual teachers, schools, and education programs. These data can be used to estimate value-added versions of EPFs that produce arguably more valid estimates of the (causal)

effects of programs and resources. In theory, value-added models can also be used to measure the effectiveness of individual teachers and schools, which can be used as the basis for accountability. In the next section, I describe the basic meaning, variations, and assumptions of the various types of EPFs, blending intuitive and technical explanations. This is followed by a discussion of the different uses of EPF estimates (e.g., program evaluation and accountability) and evidence of the validity of some of the assumptions that underlie production function research. Most of the research in this area, and therefore the examples given, relate to K-12 schools, although these ideas also extend to higher education.

Production Function Models and Assumptions

The EPF is the set of education outputs associated with all possible combinations of education inputs, under the assumption that inputs are utilized efficiently. As it is widely believed that schools, as well as many other organizations, use resources at less-than-optimum efficiency, this assumption is almost always violated and studies of the relationships between inputs and outputs are therefore not, strictly speaking, studies of EPFs (Hanushek, 1979). This is more than a semantic issue because the measured impact of any given input will depend on the efficiency of the input choice and use. For example, if different types of students were affected in different ways by the same inputs, then the measured effects of the inputs on student outcomes would depend on how well schools target particular inputs to the students who stand to gain the most from them. Nevertheless, the assumption about input efficiency does not invalidate the usefulness of EPF results and, for this reason, the issue is typically ignored and the EPF term, therefore, refers to any study of the quantitative relationship between education inputs and outputs.

Under certain assumptions, these input–output relationships can be plausibly interpreted as causal effects of inputs. The assumptions, which are necessary to draw causal inferences, and produce valid estimates for such inferences, are discussed throughout the remainder of this article. For a more detailed discussion of the issues associated with causal inference, see, for example, Shadish *et al.* (2002).

Basic Assumptions

The most basic assumptions have to do with the types of factors that influence student outcomes, which are informed by education theory and evidence. School inputs represent the first important category, commonly measured by funding levels, class size, teacher education, teacher experience, and availability of computers and textbooks. The second general category that affects education outcomes is family and other nonschool inputs. The importance of nonschool inputs initially became evident with the publication of the Coleman Report (1966) in the United States. Coleman found that the variation in education outcomes was explained primarily by the variation in family background rather than school inputs. While Coleman's methods were somewhat simplistic, the central importance of family background has been widely corroborated in more recent and sophisticated studies (see, e.g., Rothstein (2004) for a review).

The next step toward the specification of the EPF is to express this basic theory mathematically. Define education output for individual student i at time t as A_{it} which is a function $f(\cdot)$ of the school inputs S and family inputs F from current and all previous time periods, a fixed student contribution I_i and an error term, ε_{it} . This yields:

$$A_{it} = f(S_{it}, S_{it-1}, \dots, F_{it}, F_{it-1}, \dots, I_i, \varepsilon_{it}) \quad [1]$$

The fixed student contribution, I_i is often called innate ability by economists and is akin to what psychologists consider general intelligence, or g . The more general term, fixed student contribution, is used here because it is virtually impossible with education data sets to estimate anything like innate ability. No data sets include measures of student abilities at birth or, in their absence, sufficiently measure family and other environmental factors well enough to distinguish innate from environmental differences. Research by Lee and Burkham (2002), for example, suggests that there are wide differences in education achievement between different racial groups when students first enter kindergarten and these are partly explained by observed differences in the home environment.

As the home environment contributes substantially to student outcomes, but in ways difficult to observe and measure, researchers increasingly emphasize the importance of accounting for unobserved differences. Accounting for something that is unobserved may seem impossible, but it is possible when data are available on individual students over time so that each student can, in effect, serve as his or her own control group for the subsequent year. In other words, by observing changes in each student's entire growth trajectory over time, and being able to measure important school inputs, it is possible to isolate the school contribution from the student contribution.

Consider what would happen if the (unobserved) student contribution were not accounted for. Taking a simple

correlation between school inputs and achievement gains, one would be concerned that students with higher scores come from families with higher incomes who, not coincidentally, send their children to schools that have greater school inputs. In this case, how is one to determine to what degree the simple correlation between school inputs and achievement gains reflects a causal effect of school inputs or a causal effect of family background? In an experiment, these unobserved differences would be addressed by randomly assigning students to control and treatment groups. In secondary data analysis, and specifically in eqn [1], the unobserved differences are captured. This makes it reasonable to interpret the elements of φ as causal effects of school inputs in eqn [1] and, more generally, to call the EPF approach an example of quasi-experimental methods.

The fact that eqn [1] includes school inputs from previous periods ($t-1$ is shown and $t-2$, $t-3$, etc. are implied) is important and reflects the cumulative nature of education – that outcomes in each period are the result of the accumulation of all student experiences from birth up to the current time t . In the earliest generations of EPF research, the cumulative nature of education production was only occasionally mentioned and almost never really addressed because data were rarely available except at a single point in time. In effect, researchers were forced to make the questionable assumption that past inputs were irrelevant to current outputs.

The EPF also requires that there is a single education output, usually measured in terms of student standardized test scores. As Hanushek (1979) points out, the implications of this assumption depend on whether excluded outputs (e.g., students' socialization, civic-mindedness, and motivation) also influence student achievement. If they do not, then the exclusion of some outputs does not bias the estimates of the EPF for the included output measure (achievement). Student motivation and student achievement, for example, are almost certainly interconnected and this may bias estimates of the EPF estimates of the effects of inputs on achievement. While there are techniques for estimating models with multiple outputs, the data are rarely, if ever, available.

Additional Common Assumptions

By adding some additional assumptions, it is possible to make eqn [1] more concrete. Todd and Wolpin (2003) outline a now commonly used EPF in which they assume the following:

1. *Age independence.* The cumulative EPF does not vary with age so that, for example, the effect of an input applied in second-grade on third-grade student outcomes is the same as the effect of an input applied in third-grade on fourth-grade student outcomes.

2. *Additive separability.* The EPF is additively separable so that the effects of inputs do not interact with one another.
3. *Fixed family inputs.* All family inputs are fixed over time.

The last assumption implies further that the fixed family contribution can be combined with the fixed student contribution, so that the student contribution I_i becomes the student–family contribution γ_i . This assumption, like many others, is mainly one of convenience and reflects the lack of data available on students' families and home environments. These assumptions yield the following more concrete EPF:

$$A_{it} = \varphi_1 S_{it} + \varphi_2 S_{it-1} + \dots + \gamma_i + \varepsilon_{it} \quad [2]$$

where φ represents the set of contributions given by current and previous school inputs. Economists call these the marginal effects or marginal products of inputs, reflecting the change in output associated with a small (marginal) change in inputs. These are often, although sometimes misleadingly, referred to simply as effects. Whether this is a reasonable term depends on the potential violations of the assumptions and their influence on causal inference, although I adopt the term effect for simplicity.

Some additional simplification can be achieved by adding another assumption:

4. *Geometric decay.* Suppose the effects of all prior school inputs decline geometrically with the time between the application of the input and the measurement of achievement so that $\varphi_2 = \lambda\varphi_1$, etc., where λ is some constant.

Intuitively, this just formalizes the idea that the most recently received inputs have the greatest impact on current student achievement. For example, some estimates show that $\lambda = 0.8$, suggesting that a small increase in school inputs (class size, etc.) in the previous period ($t-1$) would only increase current achievement 80% as much as the same increase in inputs in the current period (t). If the geometric decay assumption is valid, then the achievement equation can be reduced to:

$$A_{it} = \varphi_1 S_{it} + \lambda A_{it-1} + \gamma_i + \eta_{it} \quad [3]$$

where $\eta_{it} = \varepsilon_{it} - \lambda\varepsilon_{it-1}$. This form of the error term has important implications for the computation of the value-added EPF, as discussed later. Equation [3] and similar models that account for unobserved differences across students and families are sometimes referred to as the value-added EPF specification. One reason for using this term is that accounting for unobserved student and family contributions makes it reasonably plausible that the estimated effects are causal. Moreover, unlike eqn [2], eqn [3] can be estimated with newly available administrative data sets.

Not surprisingly, estimating eqn [3] is difficult and involves a number of assumptions in addition to those mentioned above regarding the derivation of the equation:

1. *Interval-scaled tests.* The student achievement test scores are interval-scaled so that a one-unit change in score means the same thing on all parts of the scale.
2. *Uniform input effectiveness.* Each school input is equally effective for all types of students.
3. *Student assignment based on fixed characteristics.* Students are assigned to school (and nonschool) inputs in each period based on their fixed characteristics, not time-varying characteristics such as previous period achievement gains.

There are also two additional assumptions that relate to the computation of eqn [3] and to the properties of the data. First, recall that the error term in eqn [3] was defined as $\eta_{it} = \varepsilon_{it} - \lambda\varepsilon_{it-1}$. The usual estimates of eqn [3] (using ordinary least squares) are biased in this case because the error term contains a lagged value of achievement error and because lagged achievement is one of the inputs. It is possible that the lagged error will be correlated with lagged achievement and therefore biases the estimates. One way to address this is to assume that $\lambda = 1$ and lagged achievement can be subtracted from both sides and the dependent variable becomes $A_{it} - A_{it-1}$. This avoids the problem of the possible correlation between the lagged error and lagged achievement, but, as noted by Boardman and Murnane (1979) and Todd and Wolpin (2003), also imposes the false assumption that the effects of past school inputs are the same as the effects of current school inputs. This means, for example, that the quality of a child's kindergarten must have the same impact on their achievement at the end of age 5 as it does on their achievement at age 18. In contrast, recall that the earliest generation of EPFs assumed, perhaps even more unrealistically, that past inputs did not matter at all in producing current outcomes (i.e., $\lambda = 0$). Thus, it is necessary to make one of the following assumptions:

- h1. The lagged error is uncorrelated with lagged achievement.
- h2. No decay in school input effects.

In summary, this section discusses some general assumptions of EPF (e.g., regarding the school and nonschool influences on student outcomes) as well as three assumptions necessary to derive a general cumulative EPF (age independence, additive separability, and fixed family inputs), and assumptions required to obtain the value-added specification and to allow for causal interpretation of school input effects (geometric decay, interval-scaled tests, equal teacher effectiveness, and students assigned based on fixed characteristics). Many of these same assumptions are required in other forms of experimental and quasi-experimental methods. As discussed below, the implications of violations of the assumptions can be significant, depending on the nature of the conclusions the researcher is trying to draw.

EPFs and Accountability

The examples given so far have focused on the effects of school inputs, such as class size. Another possible application of eqn [3], however, is to estimate the contributions to student learning made by individual teachers and schools, which could be used to hold educators accountable for their performance. Harris (2008) describes this as the distinction between value-added for program evaluation (VAM-P) and value-added for accountability (VAM-A).

Estimating school effects is, in principle, simply a matter of adding to eqn [3] an indicator (dummy) variable for each school. This yields an estimated school effect that can be interpreted as the school's contribution to student learning (controlling, once again, for the student's contribution). This effect is identified from the movement of students across schools, including perhaps structural moves from lower-grade-level schools (e.g., elementary schools in the US) to high-grade-level schools (e.g., US middle schools). These school moves allow the researcher to observe and compare each student's learning trajectory in multiple school settings.

When including school effects, there is also a choice to be made regarding which measured school inputs (e.g., class size) should continue to be included in eqn [3]. If the objective of accountability is to hold schools responsible for aspects of school performance that they can control, then the general rule here should be to include those inputs that are outside the control of schools. What this means in practice depends on the larger policy context. In the United States, schools have very little control over the inputs they receive and, in these cases, the best approach is to include a wide variety of measurable school input measures.

The same logic applies when trying to estimate the effects of individual teachers. This can be accomplished by adding an indicator variable for each teacher. The teacher effects in this case are estimated from the movement of students across teachers, both across time and, in secondary schools, across classes within years. The shift from the school to the teacher level also raises issues about matching the specific education output being measured (e.g., math or reading) with the specific teacher who is responsible for that specific output. This is less of an issue at the school level because the school as a whole is considered responsible for all outputs.

The resulting teacher or school effects compare the contributions to student learning of each teacher/school relative to every other. That is, estimated teacher effects measure the value-added of each teacher compared with the other teachers in the sample, and likewise for schools. This is important because it highlights the fact that there is no absolute scale on which value-added measures can be compared – there is no way to determine whether any teacher or school is as efficient as possible, or above some objective standard.

Some researchers (e.g., Harris and Sass, 2005) have estimated multilevel models with student, teacher, and school effects together. This is not only computationally challenging, due to the very large number of variables (one per student, teacher, and/or school), but also changes the interpretations of the estimated effects. When school effects are excluded, teacher effects reflect each teacher's value-added compared with the average teacher in the database. When school effects are included, teacher effects reflect each teacher's value-added compared with the average teacher within the school. The interpretation of the school effects also changes depending on whether teacher effects are included. Without teacher effects, the school effects reflect the overall contribution of the school, including the teachers. When teacher effects are included, the school effect reflects only the effect of nonteacher factors, for example, some of the effect of school leadership.

An argument can also be made for including teacher and/or school individual effects even when the main purpose is to estimate the effects of inputs such as class size, an example VAM-P. Harris and Sass (2007) argue, for example, that including teacher effects helps to account for the nonrandom selection of teachers to teacher education, which is the focus of their study. There are also legitimate questions, however, about whether teacher and school effects should generally be included in VAM-P studies and whether VAM-A studies are really useful. Every combination of teacher and school effects changes the basis of comparison and it is not always clear which comparison is most useful, nor do researchers always make it clear what comparisons they are making. The inclusion of teacher and school effects also substantially reduces the degrees of freedom in the analysis and, consequently, the precision of the estimates. For this reason, there is considerably more agreement among researchers about the use of student effects than the use of teacher and school effects. More research is needed to confirm that teacher and/or school effects generally reduce bias and therefore whether the reduction in precision is offset by an even more important reduction in bias.

Evidence on the Assumptions of the EPF

There is relatively little research testing the validity of the assumptions discussed earlier, although interest in this is growing. In some cases, the assumptions are tested directly using standard tests of statistical significance. In other cases, where it is difficult to test the assumptions directly, researchers have focused on the sensitivity of the results of the various possible assumptions.

Harris and Sass (2005) test a variety of assumptions using an unusually rich database including nearly all students in grades 3–10 in the US state of Florida. First, they find that using lagged achievement, in conjunction

with the geometric decay assumption, is sufficient to capture the role past schooling inputs. They test this assumption by estimating an EPF that includes both actual previous period inputs (e.g., lagged class size) and previous achievement; in this case, previous inputs (jointly) have no statistically significant influence on current achievement, suggesting that these inputs can be dropped as in eqn [3].

Second, Harris and Sass (2005) find that the specific assumption about the degree of geometric decay has little effect on the estimated teacher effects, for realistic values. They find that conclusions about the effects of teacher experience and professional development are almost completely unaffected by the assumption about λ , except for very low values (0–0.2) that appear unrealistic. This suggests that the assumptions about geometric decay are fairly unimportant, which further implies that it is reasonable to assume $\lambda = 1$, move lagged achievement into the dependent variable, and thereby avoid the potential correlation between lagged achievement in the error term which produces biased estimates in ordinary least squares (OLS). McCaffrey *et al.* (2004) also test this assumption, comparing the teacher effects models in which $\lambda = 1$ with a model that leaves achievement on the right-hand side and leaves λ unrestricted. They, as well as Harris and Sass (2005), find a high correlation between the teacher effects using these different assumptions.

Harris and Sass (2005, 2007) find that the way in which family and home environment are accounted for matter significantly. They compare EPFs with student unobserved effects versus measured student characteristics such as student race and family income. This yields dramatically different estimates of the effects of teacher experience, for example. Given the apparent importance of family and home environments, this is perhaps not surprising, but the implications are significant: this means that the earlier generations of studies, which relied on student race and family-income-related measures generally produced biased estimates of school effects.

Some of the assumptions are addressed through more *ad hoc* adjustments to eqn [3] and associated statistical tests. For example, a modest number of studies have considered the possibility of nonadditive separability of a specific school input – teacher experience. In these cases, researchers include experience and experience squared as dependent variables. If the coefficient on the squared term is negative, it implies diminishing marginal returns in school inputs. This is the typical assumption made in economic models of production and an example of the more general possibility of interaction effects among inputs.

At least two studies have apparently focused on the general presence of additive separability. Figlio (1999) finds, using US data, that the class-size effect does not depend on starting teacher salary, but does depend on the number of school instructional hours. Harris (2007), using

international data, also finds evidence of interactions among a wide range of school and nonschool inputs, although these include a complex mix of diminishing and increasing returns. While the EPF models used in these studies are not based on a formally derived model, such as eqs [1]–[3], they do provide evidence against additive separability.

While this review is not comprehensive, it is reflective of the general lack of evidence available about the validity of the various EPF assumptions and, therefore, about the conclusions being drawn from EPF analyses. Because of the growing availability of student-level, longitudinally linked data, a large number of studies are currently in progress to address these issues. Access to these data will also undoubtedly add much to our knowledge of the relationship between education inputs and outputs, building on the insights of earlier generations of EPF research.

Bibliography

- Boardman, A. E. and Murnane, R. J. (1979). Using panel data to improve estimates of the determinants of educational achievement. *Sociology of Education* **52**, 113–121.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., *et al.* (1966). *Equality of Educational Opportunity*. Washington, DC: US Government Printing Office.
- Figlio, D. N. (1999). Functional form and the estimated effects of school resources. *Economics of Education Review* **18**, 241–252.
- Hanushek, E. A. (1979). Conceptual and empirical issues in estimating educational production function issues. *Journal of Human Resources* **14**(3), 351–388.
- Harris, D. (2007). Diminishing marginal returns and the production of education: An international analysis. *Education Economics* **15**(1), 31–45.
- Harris, D. (2008). The policy uses and policy validity of value-added and other teacher quality measures. In Gitomer, D. H. (ed.) *Measurement Issues and Assessment for Teacher Quality*, pp 99–130. Thousand Oaks, CA: SAGE Publications.
- Harris, D. and Sass, T. (2005). Value-added models and the measurement of teacher quality. *Paper Presented at the 2005 Conference of the American Education Finance Association*.
- Harris, D. and Sass, T. (2007). *Teacher Training, Teacher Quality, and Student Achievement. Working Paper #3*. Washington, DC: Urban Institute/National Center for the Analysis of Longitudinal Data in Education Research (CALDER).
- Lee, V. and Burkham, D. (2002). *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School*. Washington, DC: Economic Policy Institute.
- McCaffrey, D. F., Lockwood, J. R., Louis, T. A., and Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics* **29**(1), 67–101.
- Rothstein, R. (2004). *Class and Schools: Using Social, Economic, and Educational Reform to Close the Black–White Achievement Gap*. New York: Teachers College Press.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin.
- Todd, P. and Wolpin, K. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal* **113**(485), F3–F33.

Relevant Website

<http://www.finaidstudy.org> – Wisconsin Scholars Longitudinal Study.

Education Production Functions: Developed Country Evidence

E A Hanushek, Stanford University, Stanford, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Educational production function – A function that relates various inputs to education including those of families, peers, and schools to the maximum level of student achievement that can be obtained.

Fixed effects – A form of statistical analysis that removes the average effects of a factor (such as individual schools) from the analysis; in the case of teachers, the fixed effect in models of achievement growth is often interpreted as a measure of teacher quality.

Value added – In the context of education production functions, the value added of an input would be the separate contribution of learning after allowing for other inputs and the base level of knowledge of the students.

Overview

Much of the analysis in the economics of education flows from a simple model of production. The common inputs are things like school resources, teacher quality, and family attributes; and the outcome is some measure of student achievement – frequently but not always student test scores. Knowledge of the production function for schools can be used to assess policy alternatives and to judge the effectiveness and efficiency of public provided services. This area of research is, however, distinguished from many others because the results of analyses enter quite directly into the policy process.

Historically, the most frequently employed measure of schooling has been attainment, or simply years of schooling completed. The value of school attainment as a rough measure of individual skill has been verified by a wide variety of studies of labor-market outcomes (e.g., Mincer (1970) and Psacharopoulos and Patrinos (2004)). However, the difficulty with this common measure of outcomes is that it simply counts the time spent in schools without judging what happens in schools – thus, it does not provide a complete or accurate picture of outcomes. It assumes a year of schooling produces the same amount of student achievement, or skills, over time and in every country.

Recent direct investigations of cognitive achievement find significant labor-market returns to individual differences in cognitive achievement (e.g., Lazear (2003), Mulligan (1999), and Murnane *et al.* (2000)). Similarly, society appears to gain in terms of productivity; Hanushek and Kimko (2000) demonstrate that quality differences in schools have a dramatic impact on productivity and national growth rates. (A parallel line of research has employed school inputs to measure quality but has not been as successful. Specifically, school input measures have not proved to be good predictors of wages or growth.) The economic implications of differences in cognitive achievement appear to be very large (Hanushek, 2005).

As outcomes cannot be changed by fiat, much attention has been directed at inputs – particularly those perceived to be relevant for policy such as school resources or aspects of teachers.

Analysis of the role of school resources in determining achievement begins with the Coleman Report, the US government's monumental study on educational opportunity released in 1966 (Coleman *et al.*, 1966). While controversial, the study's greatest contribution was directing attention to the distribution of student performance – the outputs as opposed to various school inputs such as spending per pupil or characteristics of teachers (Bowles and Levin, 1968; Hanushek and Kain, 1972).

The underlying model that has evolved as a result of this research is very straightforward. The output of the educational process – the achievement of individual students – is directly related to both inputs – ones that are directly controlled by policymakers (e.g., the characteristics of schools, teachers, and curricula) and others that are not so controlled (such as families and friends and the innate endowments or learning capacities of the students). Further, while achievement may be measured at discrete points in time, the educational process is cumulative; inputs applied sometime in the past affect students' current levels of achievement.

Family background is usually characterized by socio-demographic characteristics such as parental education, income, and family size. Peer inputs, when included, are typically aggregates of student-sociodemographic characteristics or achievement for a school or classroom. School inputs typically include teacher background (education level, experience, sex, race, and so forth), school organization (class sizes, facilities, administrative expenditures, and so forth), and district or community factors (e.g., average expenditure levels). Except for the original Coleman

Report, most empirical work has relied on data constructed for other purposes, such as a school's standard administrative records. Statistical analysis (typically some form of regression analysis) is employed to infer what specifically determines achievement and what is the importance of the various inputs into student performance.

Measured School Inputs

The state of knowledge about the impacts of resources is best summarized by reviewing available empirical studies. Most analyses of education production functions have directed their attention at a relatively small set of resource measures, and this makes it easy to summarize the results (Hanushek, 2003). The 90 individual publications that appeared before 1995 contain 377 separate production-function estimates. For classroom resources, only 9% of estimates for teacher education and 14% for teacher-pupil ratios yielded a positive and statistically significant relationship between these factors and student performance. Moreover, these studies were offset by another set of studies that found a similarly negative correlation between those inputs and student achievement. Twenty-nine percent of the studies found a positive correlation between teacher experience and student performance; however, 71% still provided no support for increasing teacher experience (being either negative or statistically insignificant). Studies on the effect of financial resources provide a similar picture. These indicate that there is very weak support for the notion that simply providing higher teacher salaries or greater overall spending will lead to improved student performance. Per-pupil expenditure has received the most attention, but only 27% of studies showed a positive and significant effect. In fact, 7% even suggested that adding resources would harm student achievement. It is also important to note that studies involving pupil spending have tended to be the lowest quality studies as defined below, and thus there is substantial reason to believe that even the 27% figure overstates the true effect of added expenditure. More recent studies do not change this picture about how resources relate to student outcomes.

These studies make a clear case that resource usage in schools is subject to considerable inefficiency, because schools systematically pay for inputs that are not consistently related to outputs.

Study Quality

The previous discussions do not distinguish among studies on the basis of any quality differences. The available estimates can be categorized by a few objective components of quality. First, while education is cumulative, frequently only current input measures are available,

which results in analytical errors. Second, schools operate within a policy environment set almost always at higher levels of government. In the United States, state governments establish curricula, provide sources of funding, govern labor laws, determine rules for the certification and hiring of teachers, and the like. If these attributes are important – as much policy debate would suggest – they must be incorporated into any analysis of performance. The adequacy of dealing with these problems can thus be used as a simple index of study quality.

The details of these quality issues and approaches for dealing with them are discussed in detail elsewhere (Hanushek, 2003) and only summarized here. The first problem is ameliorated if one uses the value added versus level form in estimation. That is, if the achievement relationship holds across grades, it is possible to concentrate on the growth in achievement and on exactly what happens educationally between those points when outcomes are measured. This approach ameliorates problems of omitting prior inputs of schools and families, because they will be incorporated in the initial achievement levels that are measured (Hanushek, 1979). The latter problem of imprecise measurement of the policy environment can frequently be ameliorated by studying performance of schools operating within a consistent set of policies – for example, within individual states in the US. As all schools within a state operate within the same basic-policy environment, comparisons of their performance are not strongly affected by unmeasured policies (Hanushek *et al.*, 1996).

If the available studies are classified by whether or not they deal with these major quality issues, the prior conclusions about research usage are unchanged (Hanushek, 2003). The best quality studies indicate no consistent relationship between resources and student outcomes. The studies finding strong resource effects, particularly for expenditure per pupil, are heavily concentrated in the group of lowest quality studies.

An additional issue, which is particularly important for policy purposes, concerns whether this analytical approach accurately assesses the causal relationship between resources and performance. If, for example, school decision makers provide more resources to those they judge as most needy, higher resources could simply signal students known for having lower achievement. Ways of dealing with this include various regression-discontinuity or panel-data approaches. When done in the case of class sizes, the evidence has been mixed (Angrist and Lavy, 1999; Hoxby, 2000; Rivkin *et al.*, 2005).

The most significant innovation of recent years is the use of large administrative databases. These databases employ state or local records on individual student's performance and are most notable for tracking students across grades. Student performance is then related to those programs and personnel that each student is exposed to over time. These large-scale databases, often following all students

in a state over time, permit controlling for a wide range of influences on achievement through the introduction of fixed effects for schools, individuals, and time (see, e.g., Rivkin *et al.* (2005) or Boyd *et al.* (2006)). These fixed effects hold constant any systematic differences that are constant among the category (such as constant differences among the sampled schools in terms of the selection of schools by families and teachers) and obtain estimates of various inputs from their variation within each of the schools. By eliminating systematic selection and sorting of students and school personnel, they can concentrate on specific causal factors that determine individual student outcomes.

A final alternative involves the use of random-assignment experimentation rather than statistical analysis to break the influence of sample selection and other possible omitted factors. With one major exception, this approach nonetheless has not been applied to understand the impact of schools on student performance. The exception is the logislated Student/Teacher Achievement Ratio progs (or Project STAR), an experimental reduction in class sizes that was conducted in the US state of Tennessee in the mid-1980s (Word *et al.*, 1990). To date, the use of randomized experiments has not had much impact on research or our state of knowledge about the impacts of resources. While Project STAR has entered into a number of policy debates, the interpretation of the results remains controversial because of concerns about the quality of the experiment (Krueger, 1999; Hanushek, 1999). The results of this experiment suggested a significant but small impact of lower class size but that all of the impact was concentrated in the first year of schooling (kindergarten or grade one). Smaller class sizes in later years had no additional impact on student outcomes.

Benefits and Costs

Throughout most consideration of the impact of school resources, attention has focused almost exclusively on whether a factor has an effect on outcomes that is statistically different from zero. Of course, any policy consideration would also consider both the magnitude of the impacts and the costs of change. For magnitude of impact, even the most refined estimates of, say, class-size impacts do not give very clear guidance. The experimental effects from Project STAR indicate that average achievement from a reduction of eight students in a classroom would increase by about 0.2 standard deviations, but only in the first grade of attendance in smaller classes (kindergarten or first grade) (see Word *et al.* (1990) and Krueger (1999)). Hoxby (2000) in her regression-discontinuity estimation for Connecticut schools finds no systematic effect of class size. Rivkin *et al.* (2005), with their fixed-effects estimation, find effects half of Project STAR in grade 4 and declining to insignificance by grade 7.

From a policy perspective, the magnitude of alternative estimates is at best small. In order to be relevant for policy, it is necessary to compare the outcomes of any change with its costs. Most educational research ignores such comparisons and neglects any consideration of costs.

It is easy to see the importance of cost considerations when put in the context of the debates over class-size reduction. In economic terms, the potential impacts of class-size reduction are very small when contrasted with the costs of such large class-size reductions, which typically involve some of the most expensive policy changes currently contemplated (see the range of cost estimates in Brewer *et al.* (1999)). The relevant alternative policy would be to compare the gains from spending on class-size reduction with the potential gains from improving the quality of teachers, the subject of the next section.

Do Teachers and Schools Matter?

Due to the Coleman Report and subsequent studies discussed above, many have argued that schools do not matter and that only families and peers affect performance. Unfortunately, these interpretations have confused measurability with true effects.

Extensive research since the Coleman Report has made it clear that teachers do indeed matter when assessed in terms of student performance instead of the more typical input measures based on characteristics of the teacher and school. The alternative approach to the examination of teacher quality concentrates on pure outcome-based measures of teacher effectiveness. The general idea is to investigate total teacher effects by looking at differences in growth rates of student achievement across teachers. A good teacher would be one who consistently obtained high learning growth from students, while a poor teacher would be one who consistently produced low learning growth.

The general research design is to estimate models of the growth in individual student achievement that can be attributed to various measured school and family factors and to mean differences in learning across the students with different teachers. The differences in student-achievement growth across classrooms, which can be taken as a measure of teacher quality, appear to be very large. Hanushek (1992) estimates that the variation in student outcomes from a good to a bad teacher can be as much as a full year of knowledge per academic year; in other words, while a poor teacher gets gains of 0.5 grade-level equivalents during a school year, a good teacher gets gains of 1.5 grade-level equivalents. Clearly, with a string of good or bad teachers, the implications for student performance could be very large.

More modern research into state administrative databases have helped refine the understanding of the importance of differences in teacher quality. For example,

Rivkin *et al.* (2005) and Hanushek *et al.* (2005) are able to provide rough bounds on the variation in teacher quality as seen within Texas (the source of their administrative database). By these studies, one standard deviation in teacher quality implies around a 0.15 standard deviation in the growth of student achievement. By this, having a series of good teachers (teachers at the 84 percentile of the quality distribution) instead of average teachers would lead to substantially different learning after just a few years. For example, 4–5 years of a good teacher could close the average achievement gap between low-income and high-income students.

These results can also be reconciled with the prior ones. These differences among teachers are simply not closely correlated with commonly measured teacher characteristics (Hanushek, 1992; Rivkin *et al.*, 2005). Moreover, teacher credentials and teacher training do not make a consistent difference when assessed against student-achievement gains (Boyd *et al.*, 2006; Kane *et al.*, 2006). Finally, teacher quality does not appear to be closely related to salaries or to market decisions. In particular, teachers exiting for other schools or for jobs outside of teaching do not appear to be of higher quality than those who stay (Hanushek *et al.*, 2005).

Some Conclusions and Implications

The existing research suggests inefficiency in the provision of schooling. It does not indicate that schools do not matter. Nor does it indicate that money and resources never impact achievement. The accumulated research surrounding estimation of education production functions simply says there currently is no clear, systematic relationship between resources and student outcomes. At the same time, more modern research into the determinants of student achievement strongly indicates that teacher-quality differences are the most significant part of differences across schools.

See also: Cost–Benefit Analysis and Cost–Effectiveness Analysis; Data; Education Production Functions: Concepts; Education Production Functions: Evidence from Developing Countries; Empirical Research Methods in the Economics of Education; Human Capital; Teacher Quality in Education Production; The Economics of Class Size.

Bibliography

- Angrist, J. D. and Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* **114**(2), 533–575.
- Bowles, S. and Levin, H. M. (1968). The determinants of scholastic achievement – an appraisal of some recent evidence. *Journal of Human Resources* **3**(1), 3–24.

- Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy* **1**(2), 176–216.
- Brewer, D. J., Krop, C., Gill, B. P., and Reichardt, R. (1999). Estimating the cost of national class size reductions under different policy alternatives. *Educational Evaluation and Policy Analysis* **21**(2), 179–192.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., *et al.* (1966). *Equality of Educational Opportunity*. Washington, DC: US Government Printing Office.
- Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources* **14**(3), 351–388.
- Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy* **100**(1), 84–117.
- Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis* **21**(2), 143–163.
- Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal* **113**(485), F64–F98.
- Hanushek, E. A. (2005). The economics of school quality. *German Economic Review* **6**(3), 269–286.
- Hanushek, E. A. and Kain, J. F. (1972). On the value of 'equality of educational opportunity' as a guide to public policy. In Mosteller, F. and Moynihan, D. P. (eds.) *On Equality of Educational Opportunity*, pp 116–145. New York: Random House.
- Hanushek, E. A., Kain, J. F., O'Brien, D. M., and Rivkin, S. G. (2005). The market for teacher quality. *Working Paper No. 11154*, Cambridge, MA: National Bureau of Economic Research February.
- Hanushek, E. A. and Kimko, D. D. (2000). Schooling, labor force quality, and the growth of nations. *American Economic Review* **90**(5), 1184–1208.
- Hanushek, E. A., Rivkin, S. G., and Taylor, L. L. (1996). Aggregation and the estimated effects of school resources. *Review of Economics and Statistics* **78**(4), 611–627.
- Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics* **115**(3), 1239–1285.
- Kane, T. J., Rockoff, J. E., and Staiger, D. O. (2006). What does certification tell us about teacher effectiveness? *Evidence from New York City*. *Working Paper No. 12155*, Cambridge, MA: National Bureau of Economic Research (April).
- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics* **114**(2), 497–532.
- Lazear, E. P. (2003). Teacher incentives. *Swedish Economic Policy Review* **10**(3), 179–214.
- Mincer, J. (1970). The distribution of labor incomes: A survey with special reference to the human capital approach. *Journal of Economic Literature* **8**(1), 1–26.
- Mulligan, C. B. (1999). Galton versus the human capital approach to inheritance. *Journal of Political Economy* **107**(6, pt. 2), S184–S224.
- Murnane, R. J., Willett, J. B., Duhaldeborde, Y., and Tyler, J. H. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management* **19**(4), 547–568.
- Psacharopoulos, G. and Patrinos, H. A. (2004). Returns to investment in education: A further update. *Education Economics* **12**(2), 111–134.
- Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica* **73**(2), 417–458.
- Word, E., Johnston, J., Bain, H. P., *et al.* (1990). *Student/Teacher Achievement Ratio (STAR), Tennessee's K-3 Class Size Study: Final Summary Report, 1985–1990*. Nashville, TN: Tennessee State Department of Education.

Further Reading

- Burtless, G. (ed.) (1996). *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*. Washington, DC: Brookings.

- Ehrenberg, R. G., Brewer, D. J., Gamoran, A., and Willms, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest* 2(1), 1–30.
- Friedman, B. M. (ed.) (2003). *Inequality in America: What Role for Human Capital Policies?* Cambridge, MA: MIT Press.
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature* 24(3), 1141–1177.
- Hanushek, E. A. (2002). Publicly provided education. In Auerbach, A. J. and Feldstein, M. (eds.) *Handbook of Public Economics*, pp 2045–2141. Amsterdam: Elsevier.
- Hanushek, E. A. and Welch, F. (eds.) (2006). *Handbook of the Economics of Education*, vols. 1 and 2. Amsterdam: North-Holland.
- Mishel, L. and Rothstein, R. (eds.) (2002). *The Class Size Debate*. Washington, DC: Economic Policy Institute.
- Moe, T. M. (ed.) (2001). *A Primer on America's Schools*. Stanford, CA: Hoover Institution Press.
- Peterson, P. E. (ed.) (2003). *Our Schools and Our Future: Are We Still at Risk?* Stanford, CA: Hoover Press.

Education Production Functions: Evidence from Developing Countries

P Glewwe, University of Minnesota, St. Paul, MN, USA

S Lambert, Paris School of Economics, Paris, France

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Education production function – The process by which by cognitive and noncognitive skills are learned by students, which includes influences from both inside and outside of the schools they attend. Factors that influence the development of these schools can be thought of as inputs into the production of those skills.

Endogenous program placement bias – When using regression analysis to estimate the causal impact of factors that affect the variable of interest (in this case, cognitive or noncognitive skills), estimates of the impact of a variable that indicates which schools participated in some type of program may be biased if that program is deliberately targeted to schools where students are performing particularly well or particularly poorly.

Measurement error bias – When using regression analysis to estimate the causal impact of factors that affect the variable of interest (in this case, cognitive or noncognitive skills), the estimated impacts of explanatory variables that have errors in their measurements will be biased, often in the direction of underestimation of the causal effects.

Omitted variable bias – When using regression analysis to estimate the causal impact of factors that affect the variable of interest (in this case, cognitive or noncognitive skills), exclusion of one or more causal factors from the regression may lead to biased estimates of the impacts of the remaining variables in the regression.

Selection and attrition bias – When using regression analysis to estimate the causal impact of factors that affect the variable of interest (in this case, cognitive or noncognitive skills), the impact of a variable that also influences which observations (students) are in the data used for the regression may lead to biased estimates. For example, a school characteristic that leads to increased learning may also attract new students to a school. If the academic performance of those new students is higher (lower) than the students that were already in the school, the estimated impact of that school characteristic will be overestimated (underestimated).

Introduction

Economists and other social scientists have accumulated convincing evidence that education makes individuals more productive and thus increases their incomes. Education also provides nonpecuniary benefits, such as improved health and social integration. Much of the gap in living standards between developed and developing countries may be due to wide gaps in education. Thus, efforts to raise enrolment rates and to increase student learning could greatly improve living standards in developing countries.

This article addresses a central issue in education research: What education policies are most effective at raising enrolment and increasing learning in developing countries? While much research has been conducted, many estimated policy impacts may be biased, or may not apply to other countries. This article assesses what has been learned and, equally important, what remains unclear and thus merits further research. The article begins by introducing the education production function, a concept used by economists, and several other relevant relationships. The next section reviews estimation issues. The subsequent section summarizes research on enrolment and learning in developing countries, and the final section concludes the article.

The Education Production Function

Formal education increases individuals' well-being primarily through their acquisition of skills, both cognitive (e.g., literacy and numeracy) and noncognitive (e.g., social and organizational skills). Thus, an understanding of the process by which formal education produces those skills is crucial for crafting effective education policies. Economists characterize this process as the education production function.

Economists have studied factories, farms, and other productive organizations for two centuries. They have gradually developed a comprehensive, yet flexible, framework for thinking about production processes. At first, depicting education as production may seem strange, but upon further reflection this approach is useful because it provides a comprehensive framework for thinking about how formal education generates those skills. Most importantly, this framework provides crucial guidance on how to use

education data to estimate the impact of education policies (and other causal factors) on students' acquisition of skills.

The process by which both cognitive and noncognitive skills are learned is determined by many different factors. Production functions simply depict this process as a mathematical relationship. These relationships can be very flexible, allowing for almost any learning process. In this sense, an education production function always exists, although one must realize that its existence does not guarantee that one can estimate it.

Factors that determine learning, henceforth referred to as variables or inputs in the production process, can be divided into school, child, and household variables. A simple, yet flexible, learning production function is:

$$A = a(S, \mathbf{Q}, C, H, I) \quad [1]$$

where A is skills learned (achievement), S is years of schooling, \mathbf{Q} is the set of all school and teacher characteristics (quality) that affect learning, C is all child characteristics (including ability) and H is all household characteristics that affect learning, and I is educational inputs that households contribute, such as children's daily attendance and purchases of textbooks and other school supplies. While years of schooling (S) and educational inputs (I) can be grouped with child or household variables, eqn [1] separates them from C and H because they are (in part) under parents' control.

Equation [1] shows how each variable affects learning, holding other variables constant. This qualification is important. Consider an improvement in one school quality variable, call it Q_j , such as reducing class size. Equation [1] shows how changing Q_j affects learning for given values of the other variables. However, changing Q_j (or any school quality variable) could change household behavior, that is change S or one or more I variables. For example, parents may keep children in school longer (increase S) or reduce educational inputs (reduce I variables) in response to improved school quality. Thus, the full impact of changing Q_j on skills (A) is not captured by that variable's impact as depicted in eqn [1].

To obtain the full impact of changing school quality, one must know how changes in both the \mathbf{Q} variables and in other variables affect S and I in eqn [1]. These relationships can be expressed as:

$$S = f(\mathbf{Q}, C, H, \mathbf{P}) \quad [2]$$

$$I = g(\mathbf{Q}, C, H, \mathbf{P}) \quad [3]$$

where \mathbf{P} depicts the prices relevant for these household decisions, such as tuition, school supply prices, and even child wages (the price of children's time spent in school).

Inserting [2] and [3] into [1] gives another expression for A :

$$A = b(\mathbf{Q}, C, H, \mathbf{P}) \quad [4]$$

which economists call a reduced form relationship. It shows the full causal impact of school quality variables (and other variables) on learning. Equation [4] is not a production function because it depends on households' preferences (which guide households' decisions) and because it includes prices. While the production function in [1] shows the direct impacts of all variables that influence learning, when analyzing policy impacts one must estimate the full impact, which includes indirect impacts that work by changing variables that households control. Equation [4] shows this full relationship.

Should education policymakers focus on eqn [1] or eqn [4]? Equation [4] is useful because it shows actual changes in A after the \mathbf{Q} and \mathbf{P} variables change, and government policies primarily affect these two sets of variables. Yet, the impacts of \mathbf{Q} on A in eqn [1] are also important because they better capture overall welfare effects. Intuitively, if increases in Q_j induce parents to reduce educational inputs (I), household welfare increases because savings from these reduced purchases can be spent on other goods. Equation [4] shows the drop in A from reducing I , but not the increased household welfare from purchasing other goods. In contrast, the structural impact measured in [1] ignores both effects. Since they have opposing impacts on household welfare, they largely cancel out each other. Thus, overall welfare effects are better approximated by changes in A measured in [1]; see Glewwe *et al.* (2004) for details. Overall, estimates of both [1] and [4] are useful for policymakers.

Of course, some government policies cannot be described as changes in school quality (\mathbf{Q}) or schooling prices (\mathbf{P}). Examples are policies that decentralize decision-making processes or change teachers' contracts. Such policies affect schooling outcomes by changing what happens in classrooms, or changing prices for education goods and services. Glewwe and Kremer (2006) explain that one can depict \mathbf{Q} and \mathbf{P} as determined by (functions of) education policies, and perhaps by community characteristics. Ultimately, both skills (A) and years of schooling (S) are determined by child and household characteristics, education policies, and community characteristics. Knowledge of these ultimate relationships directly links education policies to these education outcomes.

Estimation of Production Functions

While knowledge of skills production functions and other education relationships is crucial for effective education policies, these relationships are very difficult to estimate. This section explains the methods used, potential problems, and possible solutions.

Consider estimating the production function in eqn [1]. (These estimation issues also apply to other education relationships, such as the determinants of years of

schooling, eqn [2].) This can be estimated using linear regression methods. Assuming linearity is not restrictive if one adds squared and interaction terms to the variables in [1]. A simple linear specification of [1] is:

$$A = \beta_0 + \beta_1 S + \beta_{Q1} Q_1 + \beta_{Q2} Q_2 + \dots + \beta_{C1} C_1 + \beta_{C2} C_2 + \dots \\ + \beta_{H1} H_1 + \beta_{H2} H_2 + \dots + \beta_{I1} I_1 + \beta_{I2} I_2 + \dots + u_A [1']$$

where each variable in Q , C , H , and I is shown explicitly. An error term, u_A , is added, for several reasons. First, data never exist for all variables in Q , C , H , and I , so u_A accounts for all variables in [1] that are not in the data. Second, u_A indicates that [1'] is only a linear approximation of [1]. Third, observed test scores (A) may measure actual skills with error, so u_A includes the difference between observed A and the true A . Finally, the right-hand-side variables in [1'] may also have measurement errors, so differences between their true and measured values are also in u_A .

While u_A may seem unimportant because it is unobserved, the causal impacts of the observed variables in [1'] on learning, the β coefficients, can be consistently estimated by ordinary least squares (OLS) only if u_A is uncorrelated with all the observed explanatory variables. Unfortunately, u_A is very likely to be correlated with those variables. The following subsections offer four reasons.

Omitted Variable Bias

The explanatory variables in eqn [1'] could be correlated with u_A because of omitted variable bias: no data set contains all variables in each set of variables (Q , C , H , and I), and many unobserved variables (which end up in u_A) may well be correlated with some observed variables. Difficult-to-observe variables include: teachers' motivation (a Q variable), school principals' management skills (Q), children's ability (C) and motivation (C), and parents' willingness (H) and capacity (H) to help, and the time they spend helping (I), their children with schoolwork. OLS estimates of the β 's in eqn [1'] may be biased because these variables, if unobserved, are probably correlated with some observed variables in [1']. For example, high-quality schools are usually better in many dimensions, both observed and unobserved. This produces positive correlation between u_A and observed school and teacher quality variables, leading to overestimation of the impacts of those variables. Similarly, parental tastes for children's education are rarely observed and probably positively correlated with parental education, causing overestimation of the latter's impact. Omitted variable bias can also induce underestimation of observed variables' impacts. For example, high school quality may lead parents to reduce time spent helping their children, generating negative correlation between school quality and u_A (assuming some parental efforts are unobserved, which is likely, and thus are in u_A). (When unobserved variables correlated with u_A are endogenous in the sense that

households choose them, researchers sometimes call the resulting bias endogeneity bias.) Omitted variable bias affects the estimated β terms not only for observed variables that are correlated with u_A but also for those uncorrelated with u_A .

Selection and Attrition Bias

Sometimes changes in school and teacher characteristics (Q) affect which children attend school. Improved school or teacher quality should increase enrolment, but new students may differ from original students in unobserved ways. New students' parents may care more about education and take (unobserved) actions to help their children beyond enrolling them in better schools. Alternatively, improved schools could attract students who would otherwise drop out (students with lower academic ability or less parental support). The former effect overestimates impacts of school quality variables on learning; the latter leads to underestimation. Bias due to changes in enrolment is called selection bias. A related phenomenon is attrition bias: current students are less likely to drop out if school quality improves. If weaker students would otherwise have left, the β 's on many school quality variables would be underestimated (due to their negative correlation with u_A).

Endogenous Program Placement Bias

School quality could also be correlated with u_A if governments improve schools with unobserved education problems (Pitt *et al.*, 1993). Governments may also raise school quality in areas with good education outcomes, if those areas have political influence (World Bank, 2001). The former causes underestimation of school-quality variables' impacts on learning, while the latter causes overestimation.

Measurement Error Bias

Anyone who has seen household or school survey data collection in developing countries understands that even the best data contain many errors. Data on school characteristics (including tuition fees) may be inaccurate or out of date. Child, household, and school input variables are prone to errors. Because measurement error is the difference between the true and observed values of a variable, it causes u_A to be correlated with the observed variable. Random measurement error typically causes underestimation of true impacts, while nonrandom errors could cause underestimation or overestimation.

Whatever the cause, the key point is: anything that induces correlation between u_A and the observed variables in [1'] will lead to biased estimates.

Methods to Reduce or Remove Bias

The above estimation problems are difficult to solve. When estimating the impacts of policy variables (P or Q) on learning, all difficulties arise because students facing different policy variables probably differ in unobserved ways. One approach to address omitted variable, measurement error, and endogenous program placement biases is instrumental variables, which uses other variables (that must be uncorrelated with u_A) to predict the variable(s) correlated with u_A . Unfortunately, it is difficult to find plausible instruments. Glewwe and Kremer (2006) provide further discussion. Another approach uses panel data to estimate the impact of changes (over time) in observed variables on changes in test scores; if the unobserved variables correlated with observed variables do not change, they are differenced out of the equation (their change over time is zero). However, many unobserved variables could change, and such differencing may exacerbate measurement error bias.

The ideal solution to these estimation problems is randomized trials: randomly divide the population into two groups, one is treated but not the other. Random assignment ensures that unobserved characteristics do not differ across groups, so they are uncorrelated with the policy treatment. Comparison of outcomes for these groups provides unbiased estimates of that policy's impact. Unfortunately, this approach might be costly or infeasible for other reasons (including ethical reasons). Yet, several studies discussed in the next section use this approach, which should raise one's confidence in their results.

Evidence from Developing Countries

This section presents basic facts concerning education in developing countries and then reviews the evidence from

those countries on the impact of various education policies on school attainment and learning. For detailed reviews, see Glewwe and Kremer (2006) and Glewwe and Miguel (2008).

Education in Developing Countries

Despite dramatic increases in school enrolment in most developing countries, they still lag behind developed countries. One widely available indicator is the gross enrolment rate, the number of children enrolled in a particular education level, regardless of age, divided by the population in the age group associated with that level. In 1960, primary school gross enrolment rates were 65% in low-income countries, 83% in middle-income countries, and over 100% in high-income countries (**Table 1**). By 2000, enrolment rates reached or exceeded 100% in low- and middle-income countries, and in all regions except sub-Saharan Africa (sub-Saharan Africa's gross enrolment rate increased to almost 100% by 2005). Secondary enrolment rates increased dramatically since 1960 (**Table 2**).

Gross enrolment rates above 100% do not imply that all children are enrolled. Grade repetition can cause gross rates to exceed 100% even when some children never enroll. Another measure of education progress is net enrolment rates, the number of children enrolled in a particular schooling level of the age associated with that level, divided by all children of that age. Net enrolment rates cannot exceed 100%; they remove the upward bias in gross rates caused by enrolment of overage children (due to repetition or delayed enrolment). **Table 1** shows that primary net enrolment rates are much lower than gross rates. Sub-Saharan Africa's net enrolment rates are particularly low, only 70% in 2005 (although this represents progress since 2000, when it was 56%). The educational outcomes in sub-Saharan Africa are low even

Table 1 Primary enrolment rates

	Gross enrolment rate				Net enrolment rate	
	1960	1980	2000	2005	2000	2005
<i>Area</i>						
World	80	97	104	109		89
<i>Country group</i>						
Low income	65	94	102	111	85	85
Middle income	83	101	110	106	88	93
High income	109	101	102	105	95	94
<i>Region</i>						
Sub-Saharan Africa	40	80	77	96	56	70
Middle East/North Africa	59	89	97	102	84	91
Latin America	91	105	127	110	97	94
South Asia	41	77	98	115	83	86
East Asia	87	111	111	112	93	94
East Europe/Formal Soviet Union	103	100	100	113	88	91
Organization for Economic Cooperation and Development (OECD)	109	102	102	102	97	96

compared to low-income non-African countries; its primary and secondary rates are well below those of South Asia, which is as poor as Africa. These averages hide some gender disparities, which are largest in South Asia (Table 3).

In 2004, the adult literacy rate in developing countries reached 77%, but was only 54% in least-developed

countries (UNESCO, 2006). Illiteracy arises because many adults never attend school and because attending school may not develop literacy. Sub-Saharan Africa exemplifies the latter. Figure 1 shows that the probability of being literate after 6 years of schooling in Africa is about 70%, while the same probability is attained after only 3 years in non-African low-income countries.

Problems in attaining literacy likely reflect low school quality in many developing countries; children learn much less per year of schooling than children in developed countries. This may reflect the rapid expansion of education in developing countries in recent decades (see Tables 1 and 2), which strained financial and human resources. Education quality comparisons across countries require internationally comparable data on academic performance. Two sources of such data are the Third International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS), administered by the International Association for the Evaluation of Educational Achievement.

The first two columns of Table 4 show the 1999 TIMSS mathematics test scores of grade 7 and 8 students. The two developed countries' scores were 579 (Japan) and 502 (US). Korean students scored even higher (587), and Malaysians also performed well (519). Other developing countries had much lower scores, ranging from 275 (South Africa) to 467 (Thailand).

The TIMSS data include few sub-Saharan African countries, but three comparable data sets exist for African countries: UNESCO Monitoring Learning Achievement (MLA) program, which covers 72 (African and non-African) countries; Programme d'Analyse des Systèmes Educatifs de la CONFEMEN (PASEC) program, which monitors 15 French-speaking African countries, and Southern African Consortium for Monitoring Educational Quality

Table 2 Secondary school enrolment rates (percent of students of secondary school age)

	<i>Gross enrolment rate</i>				<i>Net enrolment rate</i>	
	1960	1980	2000	2005	2000	2005
<i>Area</i>						
World	29	49	67	71		67
<i>Country group</i>						
Low income	14	34	54	60		44
Middle income	21	51	77	81		69
High income	63	87	101	99		91
<i>Region</i>						
Sub-Saharan Africa	5	15	27	31		26
Middle East/North Africa	13	42	66	75		71
Latin America	14	42	86	78		65
South Asia	10	27	47	52		21
East Asia	20	44	67	73		61
East Europe/FSU	55	93	88	92		85
OECD	65	87	107	103		93

Table 3 Gender disparities in gross primary and secondary enrolment rates (2000)

	<i>Primary</i>		<i>Secondary</i>	
	<i>Boys</i>	<i>Girls</i>	<i>Boys</i>	<i>Girls</i>
<i>Area</i>				
<i>Country group</i>				
Low income	107	98	60	47
Middle income	112	108	77	78
High income	102	101	100	102
<i>Region</i>				
Sub-Saharan Africa	83	71	29	24
Middle East/North Africa	101	92	71	61
Latin America	129	125	83	89
South Asia	107	90	53	39
East Asia	112	111	73	60
East Europe/FSU	100	99	88	89
OECD	102	102	106	108

Countries with populations of less than 1 million are excluded.
From World Bank (2003).

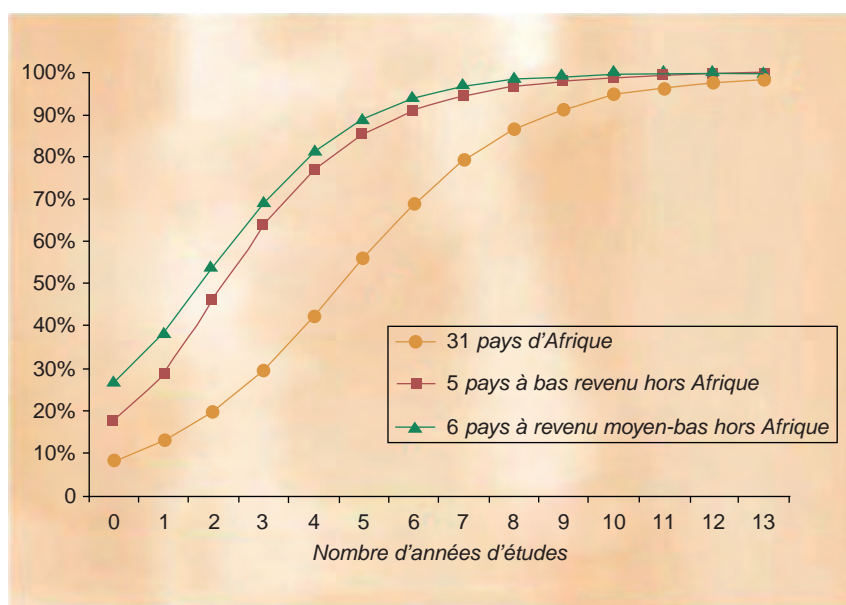


Figure 1 Probability of being literate according to the number of years of schooling (in the 2000s). *Other low-income countries: Moldavia, Uzbekistan, Laos, Tajikistan, and Vietnam. *Other low- and middle-income countries: Albania, Bolivia, Bosnia, Dominican Republic, Philippines, and Surinam. From Pôle de, D. (2007). *Les acquisitions scolaires et la production d'alphabétisation de l'école primaire en Afrique: approches comparatives. Note thématique n° 2.*

Table 4 Mean mathematics and reading achievement (TIMSS and PIRLS studies)

Country	Mathematics (TIMSS) 1999		Reading (PIRLS) 2001
	Grade 7	Grade 8	Grade 4
France	-	-	525
Japan	-	579	-
UK (England)	-	-	553
US	-	502	542
Argentina	-	-	420
Belize	-	-	327
Chile	-	392	-
Colombia	-	-	422
Indonesia	-	403	-
Iran	-	422	414
Jordan	-	428	-
Korea (South)	-	587	-
Kuwait	-	-	396
Malaysia	-	519	-
Morocco	337	-	350
Philippines	345	-	-
South Africa	-	275	-
Thailand	-	467	-
Tunisia	-	448	-
Turkey	-	429	449

From IAEEA (2000, 2003).

(SACMEQ), which conducted surveys in 15 southern and eastern African countries. These programs indicate that school quality varies widely across Africa. For example, in terms of correct answers on the PASEC test, the highest scoring country (Côte d'Ivoire) had an average score

2.5 times higher than the lowest (Mauritania). **Table 5** shows African countries' performance on these tests. Note that South Africa, the worst performer in the TIMSS, is about average for African countries.

These assessments imply that raising learning in developing countries requires not just increased enrolment but also more learning per year enrolled. Policy interventions that increase enrolment and/or learning can be classified into those that increase the demand for education and those that improve the supply of education services. The following subsections review both types. **Table 6** summarizes the findings. The discussion is limited to policy impacts on enrolment and learning; see Damon and Glewwe (2007) for comparisons of the costs and benefits of specific policies, focusing on Latin America.

Demand Interventions

Policies that raise demand for education increase learning by increasing time in school. This is done by increasing available household income and/or reducing education costs. (Raising school quality can also increase demand, but for convenience we treat school provision as a supply side intervention; see the next subsection.) Income transfers directly raise parental income. Unconditional transfers may increase households' education spending, but such effects may be small. Yet, two recent studies, based on pension reforms, find sizable (unconditional) income effects. Edmonds (2006) showed that, in South Africa, receiving a generous pension (about 125% of black households'

Table 5 Comparable skills test results for sub-Saharan African countries

	Mean scores on academic tests			Probability of being literate after 6 years of schooling	
	PASEC grade 5, average of reading and mathematics	MLA 1999 average of reading, science and mathematics	SACMEQ II average of reading and mathematics	Not adjusted	Adjusted to account for individuals who never enrolled
Angola				91	86
Bénin				87	87
Botswana		51.7	517		
Burkina Faso	45.6			73	69
Burundi				95	79
Cameroun	53.3			73	69
Central African Republic				62	60
Chad	31.0			51	51
Comores				68	48
Congo (Democratic Republic)				27	27
Côte d'Ivoire	46.1			85	79
Equatorial Guineau				86	56
Ethiopia				59	45
Gambia		40.4		47	44
Ghana				28	25
Guinée	37.5			88	76
Guinée Bissau				79	75
Kenya			555	57	46
Lesotho			449	89	71
Madagascar	50.6	56.8			
Malawi		51.7	431	68	52
Mali	35.0*	50.8		54	54
Mauritius		59.2	561		
Mauritania	19.8				
Morocco		62.8			
Mozambique			523	97	97
Namibia			440		
Niger	29.5*	42.0		60	59
Nigeria		30.0		41	33
Rwanda				98	95
Sao Tomé and Príncipe				79	48
Sénégal	35.9			71	61
Seychelles			568		
Sierra Leone				49	46
South Africa			489		
Swaziland			523	67	54
Tanzania			534	85	81
Togo	44.0*			81	79
Tunisia		71.0			
Uganda		58.0	494		
Zambia		43.3	438	44	43
Zanzibar			478		
Zimbabwe			505		

From Pôle de, D. (2007). *Les acquisitions scolaires et la production d'alphabétisation de l'école primaire en Afrique: approches comparatives. Note thématique n° 2.*

median income) increased enrolment among children living with pensioners. For example, male pension eligibility increased rural boys' school attendance by 18 percentage points. Carvalho (2000) finds similar effects from

Brazil's new pension scheme. For a benefit increase of R \$100 (about half the minimum wage), girls' enrolment increased by 4.5 percentage points. Boys' enrolment increases were much smaller.

Table 6 Summary of empirical results

			Impact on enrolment	Impact on learning
<i>Demand-side policies</i>				
Unconditional transfers (pensions)	Edmonds (2006)	South African pension scheme	Yes	Not estimated
	Carvalho (2000)	Brazilian rural pension scheme	Yes	Not estimated
Conditional cash transfers	Schultz (2004)	Progresa, Mexico	Yes	Not estimated
	Maluccio and Flores (2005)	Nicaragua	Yes	Not estimated
Conditional transfers in kind	Glewwe, Olinto and de Souza (2004)	PRAF, Honduras	Yes	Not estimated
	Ravallion and Wodon (2000)	FFE, Bangladesh	Yes	Not estimated
	Vermeersch and Kremer (2005)	Kenya, school breakfast	Yes	Yes
	Kremer <i>et al.</i> (2003)	Kenya, school uniforms	Yes	No
Abolition of school fees	School Fees Abolition Initiative (2006)	Kenya	Yes	Not estimated
Vouchers	Angrist <i>et al.</i> (2001)	Colombia	Yes	Yes
	Hsieh and Urquiola (2006)	Chile	No	No
Health: deworming	Miguel and Kremer (2004)	Kenya	Yes	No
<i>Supply-side policies</i>				
School construction	Duflo (2001)	Indonesia	Yes	Not estimated
Contractual teachers	PASEC (2004)	Mali	Yes	No
	Sy (2007)	Mali	Not estimated	No
Class size	Angrist and Lavy (1999)	Israel	Not estimated	Yes
	Urquiola (2001)	Bolivia	Not estimated	Yes
Teachers incentives	Glewwe, Ilias and Kremer (2003).	Kenya	No	No
Remedial education	Banerjee <i>et al.</i> (2005)	India	Not estimated	Yes
Traditional inputs (textbooks, etc.)	Glewwe <i>et al.</i> (2007)	Kenya	No	No
Computer-assisted education	Banerjee, Duflo and Linden (2003)	India	Not estimated	Yes

These findings imply that higher household income increases enrolment, which should increase learning. Thus, any economic policy that increases household incomes should increase enrolment and learning. For a detailed theoretical discussion, see Glewwe and Jacoby (2004).

To ensure that income transfers increase enrolment, one could condition them on education choices. Conditional cash transfer (CCT) programs provide transfers to households only if their children are enrolled in, and regularly attending, school. The best-known CCT program is Mexico's *Programa de Educación, Salud y Alimentación* (PROGRESA) program (now called *Oportunidades*). Schultz (2004), exploiting the randomized implementation of PROGRESA, estimates that after 3 years it raised children's years of schooling by 0.66 years. Honduras and Nicaragua also implemented CCT programs, and also randomized their implementation, which facilitates assessment of their impacts on schooling. Maluccio and Flores (2005) estimate that Nicaragua's program, which assisted poor children in grades 1–4, increased school enrolment rates by 13 percentage points, and increased attendance, conditional on enrolment, by 20 points. Glewwe *et al.* (2004) estimate that Honduras's CCT program raised children's completed years of schooling from 4.2 to 4.9. Bourguignon, Ferreira and Leite, using microsimulation techniques, show

that Brazil's CCT program, *Bolsa Escola*, induced 60% of poor, out-of-school 10- to 15-year-olds to enrol. They also simulated the impact of the transfer without conditionality, and found no impact on children's schooling.

While CCT programs appear to increase demand by raising incomes, they really work through prices: CCT programs decrease the price of education. Indeed, the price becomes negative; parents are paid for enrolling their children. Policies that directly reduce education prices have similar effects. Several African countries, including Ethiopia, Ghana, Kenya, Malawi, Mozambique, Tanzania, and Uganda, recently abolished school fees (School Fees Abolition Initiative, 2006). For example, Kenya did this in 2003, and nearly 1 million new pupils enrolled. Primary enrolment increased by 16%, overwhelming schools and ill-prepared teachers (World Bank, 2004).

Several Latin American countries have implemented vouchers, which provide poor households funds to enrol their children in private schools. Two programs have been rigorously evaluated. Colombia's *Programa de Ampliación de Cobertura de la Educación Secundaria* (PACES) awarded vouchers to over 125 000 poor urban students from 1992 to 1997. In most communities, demand for vouchers exceeded the supply, so eligibility was determined by lottery, creating a natural experiment. Angrist *et al.*

(2002) estimate that poor urban students given vouchers completed 0.12 additional years of schooling. The benefits may be even higher: voucher recipients' test scores were 0.2 standard deviations higher, equivalent to 1 year of schooling. In contrast, Hsieh and Urquiola (2006) found no effect of vouchers on Chilean students' test scores, repetition rates, or years of schooling.

Another policy that changes education prices is transfers in kind. Bangladesh's Food for Education program provides an in-kind subsidy equivalent to 5–10% of household income. Ravallion and Wodon (2000) estimate that this increased enrolment from 74% to nearly 100%. In Kenya, provision of school meals and uniforms were evaluated using randomized trials. Vermeersch and Kremer (2004) found that free breakfasts at preschools increased school participation (which combines enrolment with daily attendance) by 30%. Test scores also increased. Such programs reduce schooling costs and also improve children's nutritional status. Several studies show that better nutrition in the first years of life helps children stay in school and learn more per year of schooling (see Glewwe and Miguel, 2008). Providing uniforms also increased enrolment; Kremer *et al.* (2003) estimate that free uniforms increased pupils' participation by 15%.

Finally, some interventions to improve school-age children's health have strong impacts. Miguel and Kremer (2004) implemented a randomized trial in an area of western Kenya with very high helminth infection rates. The program, which provided students deworming medicine, increased school enrolment and attendance at very little cost. Surprisingly, no effect was found on learning.

In summary, evidence from many countries indicates that higher household incomes and especially lower schooling costs greatly increase school enrolment. This should increase children's acquisition of skills.

Supply Policies

Many government education policies operate not through increasing education demand but instead by improving the supply. Most supply-side policies take two forms: (1) increasing the quantity of schooling services offered and (2) increasing the quality of schooling services. The former involves building more schools and/or increasing the capacity of existing schools.

Building new schools reduces households' distance to the nearest school. Long distances reduce enrolment. Some regard distance as particularly harmful to girls' schooling (Foster and Rosenzweig, 2004; Bommier and Lambert, 2000). Filmer (2004) argues that decreasing distance to schools increases enrolment, but not by much. Duflo (2001), using a natural experiment, showed that building new schools increased years of schooling in Indonesia by about 0.15 years.

Another quantity policy is hiring more teachers to increase existing schools' capacity. In French-speaking West Africa, several countries have recently hired additional teachers, but on different terms. Contract teachers have short-term renewable contracts, receive little training, and are paid much less than permanent, civil service teachers. Senegal, Mali, and Niger have hired thousands of such teachers, permitting large enrolment increases. For example, Mali hired 11 400 contract teachers between 1992 and 2004; between 1998 and 2002 they constituted 86% of newly hired primary and secondary teachers. PASEC (2004) estimates that this allowed 100 000 additional pupils to be enrolled, a 5% increase.

The impact of contract teachers on education outcomes is widely debated, but few reliable evaluations exist. Results using PASEC data suggest that contract teachers raise student learning as much as regular teachers do (Sy, 2007). Theoretically, the impact is ambiguous. Lower wages could reduce incentives, but worries of losing one's job due to underperformance may offset this. Yet, designing performance-based teacher contracts is difficult. Glewwe *et al.* (2003) implemented a randomized trial in Kenya to assess the impact of awarding teachers prizes based on student learning. They found little impact of incentives on learning.

Consider next policies to increase education quality (increasing learning per year enrolled). Additional teachers may raise school quality as well as quantity. Banerjee *et al.* (2007) conducted a randomized evaluation of a program that hired young educated women from the community to teach basic literacy and numeracy skills to children falling behind in India's government schools. The program was very effective, increasing test scores by 0.14 standard deviations after 1 year, and 0.28 after 2 years. Similar results were found in Chile by Chay *et al.* (2005).

Increased enrolment from reduced tuition or other policies increases class size, which may reduce learning. Angrist and Lavy (1999) exploited an Israeli natural experiment arising from strict application of a rule that class size should not exceed 40 (which is low by sub-Saharan Africa's standards) to investigate how class size affects learning. They found that smaller classes improve learning; having ten fewer students increased reading scores by 2.5 points (the average score was 72, with a standard deviation of 8). Urquiola (2006) finds similar results for Bolivia.

Another way to improve education quality is to provide material inputs. Two randomized trials, one that provided textbooks and another that supplied flipcharts, were conducted in Kenya. Surprisingly, neither textbooks nor flipcharts increased learning, except that textbooks raised learning among the best students (Glewwe *et al.*, 2004, 2007). In India, Banerjee *et al.* (2005) conducted a randomized evaluation of computer-assisted education. Fourth graders were given 2 h of computer time per week to play games that reinforced mathematics skills.

The program increased math scores by 0.47 standard deviations after 2 years. These results persisted for at least 1 year after students finished the program.

Conclusions and Suggestions for Future Research

Developing countries have made great strides to ensure that all children complete primary school and most enrol in secondary school. Researchers have demonstrated that several policies do raise enrolment (see **Table 6**). Yet, school enrolment does not ensure that students learn fundamental cognitive (and noncognitive) skills. While hundreds of policies have been advocated to increase learning, few have been rigorously evaluated, and even fewer appear to be effective. Randomized evaluations of proposals to increase learning deserve the highest priority for future research.

In our opinion, six specific phenomena require further study. First, grade repetition is high in many countries (on average, 13% in sub-Saharan Africa and 12% in Latin America; see Glewwe and Kremer, 2006). Repetition is an inefficient policy for helping students who fall behind. The real issue is: How should schools accommodate students with very different backgrounds and abilities? Some advocate tracking (grouping students by academic performance). Others argue that tracking is inherently unequal. In any case, the impact of repetition on education outcomes and alternatives to that policy merit further research.

Second, to pass entrance exams to secondary or post-secondary education, many public school students in developing countries attend private tutoring classes (see Dang, 2007; Dang and Rogers, 2008). Tutors are often teachers from the students' own school, which may tempt teachers to reduce effort during the school day to generate demand for tutoring services. One policy that may avoid this is to forbid teachers from tutoring their own students. More research is needed on both possible problems and effective remedies.

A third issue, discussed above, is the costs and benefits of contract teachers. A fourth is the timing of education services: hours per school day, days per year, and vacation times. In some countries, schools operate during periods of high demand for agricultural labor.

Fifth, little is known about peer effects in developing countries. Peer effects may magnify the impacts of many policy interventions. For example, a primary school intervention that increases students' learning may increase learning among children those students are grouped with in secondary school.

A final question concerns language. In most sub-Saharan African countries, children in school face, even in first grade, a language (usually English or French) they do not know. Teaching in local languages, at least

during the first years, facilitates initial learning but may inhibit mastering the academic language used in higher education. Thus, teaching in local languages deserves further study.

Acknowledgment

We would like to thank Dominic Brewer and Patrick McEwan for comments on an earlier version of this article, and Meng Zhao for helping to prepare the tables. Some of the material in this article draws from Glewwe and Kremer (2006).

See *also*: Education Production Functions: Concepts; Education Production Functions: Developed Country Evidence; Empirical Research Methods in the Economics of Education; Human Capital; Returns to Education in Developing Countries; Teachers in Developing Countries.

Bibliography

- Angrist, J. D. and Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* **114**(2), 533–575.
- Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *American Economic Review* **92**(5), 1535–1558.
- Banerjee, A., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics* **122**(3), 1235–1264.
- Bommier, A. and Lambert, S. (2000). Education demand and age at school enrollment in Tanzania. *Journal of Human Resources* **35**(1), 177–203.
- Carvalho, I. E. (2000). *Income Effects on Child Labor and School Enrollment in Brazil*. PhD Thesis, Massachusetts Institute of Technology, Department of Economics, Cambridge, MA.
- Chay, K., McEwan, P., and Urquiola, M. (2005). The central role of noise in evaluating interventions that use test scores to rank schools. *American Economic Review* **95**(4), 1237–1258.
- Damon, A. and Glewwe, P. (2007). Three proposals to improve education in Latin American and the Caribbean: Estimates of the costs and benefits of each strategy. Department of Applied Economics, University of Minnesota.
- Dang, H.-A. (2007). *The Determinants and Impacts of Private Tutoring Classes in Vietnam*. PhD dissertation, University of Minnesota, Department of Applied Economics, St. Paul, MN.
- Dang, H.-A., and Rogers, H. (2008). The growing phenomenon of private tutoring: Does it deepen human capital, widen inequalities, or waste resources? *World Bank Research Observer* **23**(2), 161–200.
- Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review* September, 795–813.
- Edmonds, E. (2006). Child labor and schooling responses to anticipated income in South Africa. *Journal of Development Economics* **81**(2), 386–414.
- Filmer, D. (2004). If you build it, will they come? School availability and school enrollment in 21 poor countries. *Policy Research Working Paper Series No. 3340*. The World Bank.
- Foster, A. and Rosenzweig, M. (2004). Technological change and the distribution of schooling: Evidence from green-revolution India. *Journal of Development Economics* **74**(1), 87–111.

- Glewwe, P., Ilias, N., and Kremer, M. (2003). Teacher incentives. *NBER Working Papers* 9671. National Bureau of Economic Research.
- Glewwe, P. and Jacoby, H. (2004). Economic growth and the demand for education: Is there a wealth effect? *Journal of Development Economics* **74**(1), 33–51.
- Glewwe, P. and Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. In Hanushek, E. and Welch, F. (eds.) *Handbook of the Economics of Education*, pp 945–1017. Amsterdam: North-Holland.
- Glewwe, P., Kremer, M., and Moulin, S. (2007). Many children left behind? Textbooks and test scores in Kenya. *American Economic Journal: Applied Economics* **1**, 112–135.
- Glewwe, P., Kremer, M., Moulin, S., and Zitzewitz, E. (2004). Retrospective vs. prospective analyses of school inputs: The case of flip charts in Kenya. *Journal of Development Economics* **74**, 251–268.
- Glewwe, P. and Miguel, E. (2008). The impact of child health and nutrition on education in less developed countries. In Schultz, T. P. and Strauss, J. (eds.) *Handbook of Development Economics*, Vol. 4, pp 3561–3606. Amsterdam: North-Holland.
- Hsieh, C. -T. and Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's school voucher program. *Journal of Public Economics* **90**, 1477–1503.
- Kremer, M., Sylvie, M., and Robert, N. (2003). *Decentralization: A Cautionary Tale*. Mimeo: Harvard University, April 2003.
- Maluccio, J. and Flores, R. (2005). *Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan Red de Protección Social*. Research Report 141. International Food Policy Research Institute, Washington, DC.
- Miguel, T. and Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica* **72**(1), 159–217.
- PASEC (2004). Les enseignants contractuels et la qualité de l'école fondamentale publique au Mali: Quels enseignements? http://www.confemen.org/IMG/pdf/plaquette_Mali_2004.pdf (accessed June 2009).
- Pitt, M. M., Mark, R. R., and Donna, M. G. (1993). The determinants and consequences of the placement of government programs in Indonesia. *World Bank Economic Review* **7**(3), 319–348.
- Pôle de, D. (2007). Les acquisitions scolaires et la production d'alphabétisation de l'école primaire en Afrique: approches comparatives. Note thématique n° 2.
- Ravallion, M. and Wodon, Q. (2000). Does child labour displace schooling? Evidence on behavioural responses to an enrollment subsidy. *Economic Journal* **110**(462), C158–C175.
- School Fees Abolition Initiative (2006). Building on What We Know and Defining Sustained Support, UNICEF and the World Bank. <http://www.ungei.org/infobycountry/files/HighlightsSFAIWorkshopNairobiApril2006.pdf> (accessed June 2009).
- Schultz, T. P. (2004). School subsidies for the poor: Evaluating the Mexican PROGRESA poverty program. *Journal of Development Economics* **74**(1), 199–250.
- Sy, A. (2007). *L'impact des enseignants contractuels sur la performance des élèves du cycle primaire: les cas du Mali et du Niger*. Master Thesis EHESS, Paris, France.
- UNESCO (2006). *Education for All (EFA) in Least Developed Countries*. Paris: UNESCO.
- Urquiola, M. (2006). Identifying class size effects in developing countries: Evidence from rural Bolivia. *Review of Economics and Statistics* **88**(1), 171–177.
- Vermeersch, C. and Kremer, M. (2004). School meals, educational achievement and school competition: Evidence from a randomized evaluation. *Policy Research Working Paper* #3523. The World Bank.
- World Bank (2001). *World Development Report 2000/2001: Attacking Poverty*. Washington, DC: The World Bank.
- World Bank (2004). Kenya: Strengthening the foundation of education and training in Kenya. *Report No. 28064-KE*. Washington, DC: The World Bank.

Further Reading

- Bourguignon, F., Ferreira, F., and Leite, P. (2003). Conditional cash transfers, schooling and child labor: Micro-simulating Brazil's Bolsa Escola program. *World Bank Economic Review* **17**(2), 229–254.
- Duflo, E., Banerjee, A. V., Cole, S. A., and Leigh, L. L. (2005). Remedying education: Evidence from two randomized experiments in India. *CEPR Discussion Paper No. 5446*. *BREAD Working Paper No. 109*.
- Glewwe, P. (1999). *The Economics of School Quality Investments in Developing Countries: An Empirical Study of Ghana*. London: MacMillan.
- Glewwe, P. (2002). Schools and skills in developing countries: Education policies and socioeconomic outcomes. *Journal of Economic Literature* **40**(2), 436–482.
- Glewwe, P. and Meng, Z. (2006). Attaining universal primary completion by 2015: How much will it cost? In Cohen, J., Bloom, D., and Malin, M. (eds.) *Educating All Children: A Global Agenda*, pp 415–454. Cambridge, MA: MIT Press.
- Hanushek, E. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature* **24**(3), 1141–1177.
- Kremer, M. (2003). Randomized evaluations of educational programs in developing countries: Some lessons. *American Economic Review Papers and Proceedings* **93**(2), 102–115.
- Lazear, E. (2001). Educational production. *Quarterly Journal of Economics* **116**(3), 777–803.
- Todd, P. and Kenneth, W. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal* **113**(485), F3–F33.
- Vegas, E. and Jenny, P. (2008). *Raising Student Learning in Latin America: The Challenge for the 21st Century*. Washington, DC: The World Bank.

Family Environment in the Production of Schooling

R Rothstein, Economic Policy Institute, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

The association between social class and student achievement (measured by standardized test scores of basic skills) is well known. In 1966, James Coleman and colleagues concluded that “schools bring little influence to bear on a child’s achievement that is independent of his background and general social context” (Coleman *et al.*, 1966: 325). Shortly afterwards, Jencks *et al.* (1972) confirmed this finding. Researchers have replicated this relationship many times since. In 1998, social scientists, using more elaborate data, concluded that family environment accounted for more than half of the black–white achievement gap at the beginning of schooling, and if environmental measures could be more elaborate still, “environmental differences between blacks and whites may explain the entire test score gap” (Phillips *et al.*, 1998: 104).

Nonetheless, dispute persists regarding whether this association stems from disadvantaged children’s less adequate instruction in less adequately resourced schools; from background characteristics (such as family, community, social, and economic) that influence achievement after controlling for instructional quality; or from both school quality and background characteristics in concert.

The current (2001–2008) United States policy assumes that if school quality is not fully responsible for academic failure, schools can overcome negative influences of student background. Federal law requires schools to bring students from low-income and from racial and ethnic minority families to proficiency at challenging levels, without any prior narrowing of socioeconomic inequality. The law, however, has been ineffective, and claims that excellent schools can close (as opposed to narrow) the achievement gap between disadvantaged and middle-class students have not withstood careful scrutiny (Rothstein, 2004; Carnoy *et al.*, 2005).

Nonetheless, anecdotal accounts are plausible that high-quality schools get relatively better achievement from disadvantaged children, even if they do not fully close the gap. This article describes pathways by which family, community, as well as social and economic conditions also cause academic failure. Modest improvements in these would also yield better achievement.

There has been little experimentation with socioeconomic improvements to raise achievement. One experiment was ‘Moving to opportunity’, sponsored by the US Department of Housing and Urban Development in the 1990s. It found less impact than expected of offering lower class families housing in less distressed neighborhoods. However, families tended to choose

new neighborhoods not significantly less distressed than their home neighborhoods, making conclusions from the experiment difficult to draw (Sanbonmatsu *et al.*, 2006). A few experiments in the 1970s found improved achievement from a negative income tax for poor families (Maynard and Murnane, 1979). Other experiments (such as placing health clinics in schools) would be easy to implement, but have not been undertaken, although nonexperimental evidence of positive impact is abundant (Geierstanger *et al.*, 2004).

Without more experimentation, we cannot conclude whether reducing social and economic inequality would be more powerful ways to raise achievement than school improvement. Most plausible is that both are required.

This article describes specific pathways by which socioeconomic disadvantage plausibly causes lower achievement. As racial minority status overlaps with lower class status in the United States (33% of black children are from poor households, compared to 14% of all children (NCES, 2007, table 21)), and because data on black children are sometimes more available than on disadvantaged children generally, this article uses black, disadvantaged, and lower class as interchangeable terms. This does not imply that black children would more likely be lower class, in the absence of the specific racial inequalities of American society.

In all cases, lower class characteristics have average effects. Some disadvantaged children have higher achievement than typical middle-class children, and some middle-class children have lower achievement than typical lower class children. For any characteristic, outcome distributions are broad; social-class distributions overlap. Children’s varied school and socioeconomic experiences, and innate variability, interact to produce a range of outcomes. The fact that some lower class children have better outcomes is not evidence that socioeconomic factors are not powerful, or that superior school quality is the cause.

Social-Class-Based Child-Rearing Practices

Parents of different social classes raise children differently, on average. More-educated parents read to young children more consistently, and encourage them to read more for themselves when older (Bianchi and Robinson, 1997; Hofferth and Sandberg, 2001). Most children whose parents have college degrees are read to daily before

beginning kindergarten; children whose parents have less education benefit less from daily reading. Young children of more-educated parents are surrounded by books at home, while children of less-educated parents see fewer books (Denton and Germino-Hauskens, 2000, tables 19 and 20).

A child who enters school recognizing some words, used to turning pages of stories, will be easier to teach than one who has rarely held a book. The second child can be taught, but with equally effective teaching, the first will more likely have higher test scores.

Schools' inability to overcome disadvantages of less literate homes characterizes nations generally. The number of books in homes consistently predicts scores within almost every country (Lemke *et al.*, 2002, figure 17). Turkish immigrant students suffer from an achievement gap in Germany, as do Algerians in France, as do Caribbean, African, Pakistani, and Bangladeshi pupils in Great Britain, and so do Okinawans and low-caste Burakamin in Japan (Begag, 1990; Garner, 2004; Neuman and Peer, 2002; OFSTED, 1999; Ogbu, 1992; Sciolino, 2004; Ermisch *et al.*, 2001).

An international reading survey of 15-year-olds found strong relationships between socioeconomic status and student literacy in most countries. The literacy gap between children of highest status workers (e.g., doctors, professors, and lawyers) and lowest status workers (e.g., waitpersons, taxi drivers, and mechanics) was greatest in Germany and the United Kingdom, still great in France and the United States, somewhat lower in the Scandinavian countries and in Korea (OECD, 2001).

More-educated parents, often with professional occupations, read more themselves. When parents read to solve problems or for entertainment, they stimulate children's motivation to read. If parents do not read for themselves, they signal the unimportance of literacy (Snow and Tabors, 1996; Lareau, 1989).

How parents read to children matters; more-educated parents read aloud differently. Working-class parents reading aloud more likely tell children to pay attention without interruptions, sound-out words, or name letters. Parents' questions are more likely to be factual, asking for names of objects or recall of events. More-literate parents more likely ask creative, interpretive, or connective questions, such as "What do you think will happen next?" Middle-class parents are more likely to read aloud to have fun or to start conversations. Their children are then more motivated to read in school (Mikulecky, 1996).

When more-educated parents converse in children's presence, they use larger vocabularies and more complex sentences. Although middle-class preschoolers do not use such language themselves, they have an advantage when they hear college-educated teachers speak or when these words and constructions are encountered in books (Mikulecky, 1996; Snow and Tabors, 1996). Educated parents are more likely to give indirect instruction ("You

don't want to make so much noise, do you?"), inviting children to think through the reasoning. Children entering school with a background of indirect instruction have practiced what educators call critical thinking, upon which academic success relies.

Twenty years ago, University of Kansas researchers recorded home conversations of families from different social classes. On average, middle-class parents spoke over 2000 words per hour in the presence of very young children, working-class parents about 1300, and welfare mothers about 600. By age 4, children of professionals had vocabularies nearly 50% greater than those of working-class children and twice as large as those of welfare children.

The Kansas researchers also tracked how often parents verbally encouraged or reprimanded behavior. Toddlers of professionals got an average of six encouragements per reprimand. Working-class children had two. For welfare children, there was an average of one encouragement for two prohibitions (Hart and Risley, 1995). When these children later go to school, teachers cannot fully offset these differences. Children whose initiative was encouraged from early ages are more likely, on average, to take responsibility for their own learning.

These differences mirror those of adult roles. Parents whose professional occupations entail authority and responsibility believe more strongly that they can affect their environments and solve problems. At work, they explore alternatives and negotiate compromises. They naturally express these personality traits when giving their children indirect instruction or designing problem-solving activities. Middle-class parents are more likely to explain why rules are reasonable.

However, parents whose jobs entail following orders or doing routine tasks express a lesser sense of efficacy. Lower class parents are more likely to instruct children without extended discussion because following orders, after all, is how they themselves behave at work. Their children are more likely to be fatalistic about obstacles, in and out of school.

Middle-class children's self-assurance is enhanced in after-school activities that sometimes require large fees for enrolment and almost always require parents to have enough free time and resources to provide transportation. Organized sports, music, drama, and dance programs build self-confidence and discipline in middle-class children. Lower class parents find fees for such activities more daunting, and transportation less accessible – organized athletic and artistic activities are not available in many lower class neighborhoods. Therefore, lower class children's sports are more informal and less confidence building, with less opportunity to learn teamwork and self-discipline (Hofferth and Sandberg, 2001; Lareau, 2003).

Differences in role modeling also contribute to social-class achievement gaps. Where adults perform jobs requiring little academic skill, children's own academic

ambitions are influenced. As always, beware of deterministic simplification: some lower class children, despite few educated role models, succeed in school, perhaps as the first in their families to attend college. But on average, these children must struggle harder to motivate themselves to achieve than children who assume that, like their parents' social circle, appropriate roles are professional.

One survey found that 93% of middle-class parents had a friend or relative who was a teacher, compared to 43% of working-class parents and 36% of poor parents. Medical doctors were identified as friends by 70% of middle-class parents, 14% of working-class parents, and 18% of poor parents (Lareau, 2002). These adult friendships reinforce how children imagine future roles, and what they strive to achieve.

More affluent parents are more likely to seek to influence school policy (preferring their children assigned to a different teacher's classroom, wanting children admitted to a gifted program, or seeking special services for children having difficulties) and more confident about challenging administrators (Horvat *et al.*, 2003; Jacob and Lefgren, 2007). This adds to the academic achievement gap.

Educators attempt to overcome lower class parents' alienation from schools, because parents more involved in schooling may raise their children's comfort with academic ambition. Therefore, educators try to enlist lower class parents in classroom observation, helping teachers, ushering field trips, or raising supplementary funds (Rothstein *et al.*, 1999). However, parental involvement styles are also class-based. Parents whose own jobs are routine, expected to follow well-defined roles, assume schools should operate similarly. With less education themselves, they have neither qualifications nor confidence to question how teachers design instruction. Middle-class parents, in contrast, more easily assume a right to collaborate with teachers because these parents have comparable educations and their own professional roles often place a premium on collaboration.

In the United States, education policy has for two decades stressed raising disadvantaged children's achievement. From 1990 to 2007, mathematics proficiency of black children increased substantially. For fourth graders, average black student math achievement is now higher than white student achievement in 1991 (NCES, online). If white students' performance had been stagnant, there would now no longer be a fourth-grade mathematics achievement gap and only a small eighth-grade gap.

Black students have had no comparable reading gains, yet instructional improvement efforts have been as intense in reading as in math. It seems likely, and worthy of further investigation, that literacy is more dependent on the class-based family factors described in this section, while mathematics proficiency is more responsive to improved instruction in schools alone.

Health Influences on Achievement

Overall, lower class children are in poorer health.

Those with vision problems have difficulty reading. In the United States, 50% of poor children have vision impairment that interferes with academic work, twice the normal rate (Egbonu and Starfield, 1982; Starfield, 1982).

Some require corrective lenses but many also need eye-exercise therapy to correct focusing, converging, and tracking problems. Where therapy or lenses were provided to randomly selected fourth graders from low-income families, those receiving optometric services gained in reading achievement, while children without such services fell farther behind (Orfield, 2007; Harris, 2002).

Lower class children may be more likely to have vision problems because of less adequate prenatal development than middle-class children whose pregnant mothers had better medical care and nutrition. Visual deficits also arise from disadvantaged children being placed in inexpensive low-quality child-care settings where they watch too much television, activity that does not develop hand-eye coordination and depth perception – 42% of black fourth graders watch 6 h or more of television a day, compared to 13% of whites (NCES, 2003, table 117). Middle-class children more likely have manipulative toys that develop such coordination.

A longitudinal study of entering kindergarteners reveals that fine-motor-skill development at age 5 is a stronger predictor of later mathematics and reading performance than is kindergartners' preliteracy knowledge (of the alphabet, of numbers, and of phonemes) (Grissmer and Eiseman, 2008).

Lower class children also have more hearing difficulties, possibly because of untreated ear infections that occur in children whose overall health is less robust. Ear infections are easily treatable for children with access to regular pediatric care. But lower class children with less access to such treatment are less attentive, on average, in school (Egbonu and Starfield, 1982).

Children without dental care have more toothaches; untreated cavities are 2–3 times as prevalent among poor as among middle-class children (GAO, 2000, figure 1; GAO, 2008). Although only some cavities produce toothaches, children with toothaches pay less attention in class and are more distracted during tests, on average.

Children who live in older buildings have more lead dust exposure that harms cognitive functioning and behavior (Egbonu and Starfield, 1982; GAO, 1999; Neisser *et al.*, 1996; Neisser, 1997). High lead levels also contribute to hearing loss (Brooks-Gunn and Duncan, 1997). Low-income children have dangerously high blood lead levels at 5 times the rate of middle-class children (GAO, 1999).

Indeed, lead poisoning now exacerbates the achievement gap more than it once did. For middle-class children, leaded fumes from auto exhaust were once the chief

source of lead exposure. With automotive gasoline now unleaded, lead poisoning afflicts mostly low-income children who continue to suffer from lead in house paint to which they are exposed. They are also more likely to live in areas where leaded paint peels from fire escapes or steel beams of elevated trains (Johnson, 2003). Although lead-based paint was banned from residential construction in 1978, low-income children more likely live in buildings constructed prior to that date and in buildings not frequently repainted to prevent older layers from peeling. Urban children are also more likely to attend older schools, where drinking fountains draw water through pipes containing lead (Barton, 2003; Blum, 2004).

Lower class children, particularly those who live in densely populated city neighborhoods, are also more likely to contract asthma – the asthma rate is substantially higher for urban children, for those whose families are on welfare, and for those from single-parent or poor families (Forrest *et al.*, 1997). A New York City survey found that one of every four children in Harlem suffers from asthma, a rate 6 times as great as that for all children (Perez-Pena, 2003). A Chicago survey found a nearly identical rate for black children and a rate of one in three for Puerto Ricans (Ritter, 2004). Asthma is provoked partly from breathing fumes from low-grade heating oil, diesel trucks, and buses (school buses idling at schools are a serious problem); excessive dust and allergic reactions to mold, cockroaches, and secondhand smoke also contribute (Halfon and Newacheck, 1993). In the Chicago neighborhoods with high asthma rates, nearly 50% of the children suffering from the disease live in homes where adults smoke.

Asthma keeps children awake at night. If attending school, asthmatics are more likely to be drowsy and inattentive, more irritable and with more behavioral problems, and more likely to refrain from exercise and thus be less physically fit. Middle-class children typically get treatment for symptoms, while low-income children get treatment less often. Asthma has become the biggest cause of chronic school absence, with sufferers from low-income families more likely to miss school than those from middle-class families (Forrest *et al.*, 1997; Halfon and Newacheck, 1993).

Children without regular medical care are also more likely to contract other illnesses that keep them from school. Despite federal programs to make medical insurance available to low-income families, there remain gaps in access and utilization. Many eligible families are not enrolled because of ignorance, fear, or lack of conviction about medical care's importance.

Thus, 20% of poor children are without consistent health insurance, compared to 12% of all children; 13% of black children are without insurance, compared to 8% of white children (Mills and Bhandari, 2003, figure 4; DeNavas-Walt *et al.*, 2004). Yet, even with health insurance, parents' low-wage work interferes with medical-care

utilization. Parents who are paid hourly wages lose income when they take children to doctors. Parents with blue-collar jobs risk discharge for excessive absence, so are likely to skip well-baby and routine pediatric care, seeing doctors only in emergencies. Salaried middle-class parents have more flexibility to schedule doctor visits, for themselves and for their children, without loss of job or income.

Lower class families with health insurance also confront disparities in medical facilities. A survey of one low-income minority Los Angeles neighborhood found one primary care physician for every 13,000 residents. A nearby high-income neighborhood had one for every 200 residents (Brown *et al.*, 2003). Low-income families, with or without insurance, are more likely to use emergency rooms and less likely to use primary care doctors, even for routine care.

As a result, black preschoolers are one-third less likely than whites to get standard vaccinations for diphtheria, measles, and influenza (Hoffman *et al.*, 2003). Ongoing differences in regular pediatric care result in poor children losing many more days from school than the nonpoor, on average. School attendance differences, attributable to disparities in health-care access alone, cause differences between black and white children's average achievement. Good teaching cannot do much for children who are not in school.

Youngsters whose mothers consumed alcohol during pregnancy have more difficulty with academic subjects, are less able to focus attention, have poorer memory skills, less ability to reason, lower intelligence quotients (IQs), less social competence, and more aggression in the classroom (Astley, 2003; Simmons *et al.*, 2002). In adolescence, these children continue to have difficulty learning (Richardson *et al.*, 2002; Streissguth *et al.*, 1994).

Fetal alcohol syndrome, a collection of the most severe cognitive, physical, and behavioral difficulties experienced by children of prenatal drinkers, is 10 times more frequent for low-income black than for middle-class white children (Abel, 1995). Data are not available for disparities by social class for less severe symptoms of prenatal alcohol consumption; however, it can be presumed that the consequences are greater for the more disadvantaged. Although affluent women consume more alcohol than lower class women, the latter tend to drink in binges that apparently do more harm to developing fetuses (CDC, 2002a; Abel and Hannigan, 1995).

Smoking in pregnancy also contributes to lower achievement. Children of mothers who smoked prenatally do more poorly on cognitive tests, their language develops more poorly, and they have more serious behavioral problems, more hyperactivity, and more juvenile crime (Astley, 2003).

Thirty percent of poor women smoke, compared to 22% of nonpoor women. During pregnancy, 25% of high-school dropouts smoke, 50% more than the rate for high-school graduates, and 13 times more than that for college graduates (CDC, 2001).

Partly from prenatal smoking, low-income children are more likely to be born prematurely or with low birth weight and to suffer from cognitive problems; low-birth-weight babies, on average, have lower IQ scores and are more likely to have mild learning disabilities and attention disorders (Hack *et al.*, 1995). Thirteen percent of black children are born with low birth weight, double the rate for whites (Hoffman *et al.*, 2003). Even if all children benefited from equally high-quality instruction, this difference alone would ensure lower average achievement for blacks.

Low birth weight is only partly caused by inadequate prenatal care, exposure to urban pollutants, diet, smoking and drinking. The interaction of poor health habits with other stresses exacerbates children's adverse outcomes. Maternal stress has hormonal consequences that interfere with nutrient absorption on which healthy fetuses depend (Rich-Edwards *et al.*, 2001; Wadhwa *et al.*, 2001; Lu and Halfon, 2003). Thus, low birth weight, alcohol consumption, and smoking all have greater negative effects on poor children than on middle-class children who were exposed to similar risks. Poor women, with greater stress and less adequate nutrition, can tolerate less smoke and alcohol and still deliver healthy babies than women whose better overall health conditions protect their fetuses from the effects of alcohol or smoking (Abel and Hannigan, 1995). Middle-class children more easily overcome earlier health shocks, rebounding when they later experience healthier environments after exposure to risk (Egbonu and Starfield, 1982).

Poor nutrition also contributes to achievement gaps between lower and middle-class children. Even moderate undernutrition affects academic performance, particularly if sustained. Iron-deficiency anemia affects cognitive ability: 8% of all US children, but 20% of black children, are iron deficient (CDC, 2002b). Anemia also makes it more probable that children will absorb lead to which they are exposed (Brown and Sherman, 1995). Iron is but one example – compared to middle-class children, the poor also lack other vitamins and minerals (Koch, 2002). In experiments where pupils got inexpensive vitamin and mineral supplements, test scores rose from that treatment alone (Neisser *et al.*, 1996).

Cumulatively, differences in the health of children from different social classes explain a significant portion of the achievement gap. According to one recent calculation, a few measurable differences in children's health and in the health of their mothers (particularly, the frequency of depression) accounts for approximately 25% of the black-white achievement gap (Currie, 2005).

Housing and Student Mobility

Housing instability also contributes to poor academic achievement. In the US, urban rents have been rising faster than working-class incomes. Even low-income families with stable employment are more likely to move

when they miss rent payments. Family breakup and bouts of unemployment also cause mobility. In some schools in low-income neighborhoods, mobility rates are over 100% (for every seat, two children were enrolled at sometime during the year) (Kerbow, 1996; Bruno and Isken, 1996).

A 1994 survey found that 30% of the poorest children had attended at least three different schools by third grade, while only 10% of middle-class children did so. Black children were more than twice as likely as whites to change schools this much (GAO, 1994). High-mobility schools are frequently disrupted by classroom reconstitution, to avoid placing all newcomers together or because classes get too large or too small from arrivals and departures. Teachers with mobile students use more discrete instructional units, are less able to integrate instruction over time, more likely to review old than introduce new material, and less able to adjust instruction to unfamiliar students' individual needs (Bruno and Isken, 1996).

High mobility depresses achievement not only for movers, but also for stable children in schools whose classes are reconstituted.

Middle-class children usually have a quiet place at home, perhaps their own bedrooms, to read or do homework. Children in more crowded housing can less often escape television, conversation, or siblings.

Based on Texas student data, a recent analysis concluded that about 14% of the black-white achievement gap, and 8% of the gap between children from low-income families and others, could be explained by differences in the housing mobility rates of otherwise similar families (Hanushek *et al.*, 2004).

Income and Wealth Differences that Influence Achievement

When parents lose employment, children's achievement tends to suffer. Parents under economic stress are more likely to discipline children arbitrarily, causing more misbehavior. When parents lose work, adolescents are more likely to be delinquent, use drugs, lose faith in the future, and suffer from depression (McLoyd, 1990; McLoyd *et al.*, 1994; Flanagan and Eccles, 1993). Recovery is slow from these effects. Workers with lower incomes when employed (e.g., blacks and workers with only a high-school education or less) are more likely to experience periodic unemployment (Mishel *et al.*, 2009, tables 4.2 and 4.4). Between black and white families whose current incomes are similar, black workers are more likely to have suffered recent job loss than whites, even if now reemployed. This contributes to average achievement differences of their children.

Family assets, as well as income, influence achievement. Families with similar incomes may have different social-class positions because of differences in wealth. Median family income for blacks is about 60% of

whites', but family net worth for blacks is only 10% of whites' (Mishel *et al.*, 2009, tables 1.3 and 5.5). Thus, among middle-class families whose incomes are similar, whites are more likely than blacks to have spacious housing, not only because whites suffer no housing market discrimination but also because white parents are more likely to have received capital contributions from their own parents for a first home down payment. Black middle-class parents are more likely to be the first generation in their families to have middle-class status with their own parents less likely able to help in this way.

Asset differences also influence how much families of all races save for children's college educations. Once enrolled in college, family home equity predicts whether students graduate (Brown *et al.*, 2003). Secondary school students' awareness that their families have resources for college can influence the confidence with which those students assume that college attendance is within their grasp.

Conclusion

Considered separately, each of these social-class differences (in childrearing and literacy practices; health characteristics; housing stability; and in economic security) as well as others not discussed here, such as neighborhood crime and peer influences, may have very small influence on the test-score gap. But together, they sum up to cumulative disadvantage for lower class children, which depresses average performance, even with high-quality instruction.

None of the many background characteristics influencing achievement operates separately. Each interacts with others, and with children's unique genetic endowments. For any characteristic or group of characteristics predicting low achievement, some children possessing them will achieve at higher levels than those characteristics alone might predict.

But on average, with equally high-quality instruction, children from disadvantaged families will achieve at lower levels than those from families without these disadvantages. Strategies to improve lower class children's performance will be more effective if they combine school improvement with policies to narrow social and economic inequalities.

See *also*: Education Production Functions: Concepts; Education Production Functions: Developed Country Evidence; Education Production Functions: Evidence from Developing Countries; Parental Socioeconomic Status, Child Health, and Human Capital.

Bibliography

Abel, E. L. (1995). An update on incidence of FAS: FAS is not an equal opportunity birth defect. *Neurotoxicology and Teratology* **17**, 437–443.

- Abel, E. L. and Hannigan, J. H. (1995). Maternal risk factors in fetal alcohol syndrome: Provocative and permissive influences. *Neurotoxicology and Teratology* **17**(4), 445–462.
- Astley, S. (2003). FAS/FAE: Their impact on psychosocial child development with a view to diagnosis. In Tremblay, R. E., Barr, R. G., and Peters, R. D (eds.) *Encyclopedia on Early Childhood Development*. Montreal, QC: Centre of Excellence for Early Childhood Development, pp 1–6. <http://www.excellence-earlychildhood.com/documents/AstleyANGxp.pdf> (accessed September 2009).
- Barton, P. (2003). *Parsing the Achievement Gap. Baselines for Tracking Progress, October*. Princeton, NJ: Policy Information Center, Educational Testing Service.
- Begag, A. (1990). The 'Beurs,' children of North-African immigrants in France: The issue of integration. *Journal of Ethnic Studies* **18**(1), 1–14.
- Bianchi, S. M. and Robinson, J. (1997). What did you do today? Children's use of time, family composition, and the acquisition of social capital. *Journal of Marriage and the Family* **59**(2), 332–344.
- Blum, J. (2004). High lead levels found in water at 9 D.C. schools. *The Washington Post* February 25.
- Brooks-Gunn, J. and Duncan, G. J. (1997). The effects of poverty on children. *Future of Children* **7**(2), 55–71.
- Brown, L. J. and Sherman, L. P. (1995). Policy implications of new scientific knowledge. *Journal of Nutrition* **125**, 2281S–2284S.
- Brown, M. K., Carnoy, M., Currie, E., *et al.* (2003). *Whitewashing Race. The Myth of a Color-Blind Society*. Berkeley, CA: University of California Press.
- Bruno, J. and Isken, J. A. (1996). Inter and intraschool site student transiency: Practical and theoretical implications for instructional continuity at inner-city schools. *Journal of Research and Development in Education* **29**(4), 239–252.
- Carnoy, M., Jacobsen, R., Mishel, L., and Rothstein, R. (2005). *The Charter School Dust-Up. Examining the Evidence on Enrollment and Achievement*. Washington, DC: Economic Policy Institute.
- CDC (Centers for Disease Control and Prevention) (2001). *Women and Smoking. A Report of the Surgeon General-2001*. Washington, DC: US Department of Health and Human Services.
- CDC (Centers for Disease Control and Prevention) (2002a). *National Health Interview Survey, 1997–2000. Percent Distribution of the Drinking Levels of Females 18 Years of Age and Older According to Selected Characteristics: United States, NHIS, 1997–2000*. Washington, DC: US Department of Health and Human Services.
- CDC (Centers for Disease Control and Prevention). (2002b). *Pediatric Nutrition Surveillance 2001 Report*. Washington, DC: US Department of Health and Human Services.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., *et al.* (1966). *Equality of Educational Opportunity*. Washington, DC: US Department of Health, Education, and Welfare, Government Printing Office.
- Currie, J. (2005). Health disparities and gaps in school readiness. *Future of Children* **15**(1), 117–138.
- DeNavas-Walt, C., Proctor, B. D., and Mills, R. J. (2004). Income, poverty, and health insurance coverage in the United States: 2003. *U.S. Census Bureau, Current Population Reports*, P60-226. Washington, DC: US Government Printing Office.
- Denton, K. and Germino-Hauskens, E. (2000). *America's Kindergartner's. Findings From the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99, Fall 1998*. NCES 2000-070. Washington, DC: US Department of Education, National Center for Education Statistics.
- Egbonu, L. and Starfield, B. (1982). Child health and social status. *Pediatrics* **69**, 550–557.
- Ermisch, J., Francesconi, M., and Pevalin, D. J. (2001). Outcomes for children of poverty. *Research Report No. 158*. London: Department for Work and Pensions.
- Flanagan, C. A. and Eccles, J. S. (1993). Changes in parents' work status and adolescents' adjustment at school. *Child Development* **64**, 246–257.
- Forrest, C. B., Starfield, B., Riley, A. W., and Kang, M. (1997). The impact of asthma on the health status of adolescents. *Pediatrics* **99**(2), E1.
- GAO (US General Accounting Office) (1994). *Elementary School Children: Many Change Schools Frequently, Harming Their Education*. GAO/HEHS-94-45. Washington, DC: Author.

- GAO (US General Accounting Office) (1999). *Lead Poisoning: Federal Health Care Programs Are Not Effectively Reaching At-Risk Children*. GAO/HEHS-99-18. Washington, DC: Author.
- GAO (US General Accounting Office) (2000). *Oral Health in Low-Income Populations*. GAO/HEHS-00-72. Washington, DC: Author.
- GAO (US Government Accountability Office) (2008). *Medicaid. Extent of Dental Disease in Children Has Not Decreased, and Millions are Estimated to Have Untreated Tooth Decay*. GAO 08-1121. Washington, DC: Author.
- Garner, R. (2004). Chinese perform better in English than White children. *Independent Education* February 25.
- Geierstanger, S. P., Amaral, G., Mansour, M., and Walters, S. R. (2004). School-based health centers and academic performance: Research, challenges, and recommendations. *Journal of School Health* **74**(9), 347–352.
- Grissmer, D. and Eiseman, E. (2008). Can gaps in the quality of early environments and noncognitive skills help explain persisting Black–White achievement gaps? In Magnuson, K. and Waldfogel, J. (eds.) *Steady Gains and Stalled Progress. Inequality and the Black–White Test Score Gap*, pp 139–180. New York: Russell Sage.
- Hack, M., Klein, N. K., and Taylor, H. G. (1995). Long-term developmental outcomes of low birth weight infants. *Future of Children* **5**(1), 176–196.
- Halfon, N. and Newacheck, P. W. (1993). Childhood asthma and poverty: Differential impacts and utilization of health services. *Pediatrics* **91**, 56–61.
- Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2004). Disruption versus tiebout improvement: The costs and benefits of switching schools. *Journal of Public Economics* **88**(9–10), 1721–1746.
- Harris, P. (2002). Learning-related visual problems in Baltimore city: A long-term program. *Journal of Optometric Vision Development* **33**(2), 75–115.
- Hart, B. and Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore, MD: Paul H. Brooks.
- Hofferth, S. L. and Sandberg, J. F. (2001). How American children spend their time. *Journal of Marriage and the Family* **63**, 295–308.
- Hoffman, K., Llagas, C., and Snyder, T. (2003). *Status and Trends in the Education of Blacks*. Washington, DC: US Department of Education, Office of Educational Research and Improvement.
- Horvat, E. M., Weininger, E. B., and Lareau, A. (2003). From social ties to social capital: Class differences in the relations between schools and parent networks. *American Educational Research Journal* **40**(2), 319–351.
- Jacob, B. and Lefgren, L. (2007). What do parents value in education? An empirical investigation of parents' revealed preferences for teachers. *Quarterly Journal of Economics* **122**(4), 1603–1637.
- Jencks, C., Smith, M., Acland, H., et al. (1972). *Inequality. A Reassessment of the Effect of Family and Schooling in America*. New York: Basic Books.
- Johnson, K. (2003). Looking outside for lead danger. *New York Times*. November 2.
- Kerbow, D. (1996). Patterns of urban student mobility and local school reform. *Journal of Education for Students Placed at Risk* **12**, 147–169.
- Koch, K. (2002). Hunger in America. *CQ Researcher* **10**(44), 1034–1055.
- Lareau, A. (1989). *Home Advantage: Social Class and Parental Intervention in Elementary Education: Theory, Research, Practice*. Thousand Oaks, CA: Sage.
- Lareau, A. (2002). Invisible inequality: Social class and childrearing in Black families and White families. *American Sociological Review* **67**, 747–776.
- Lareau, A. (2003). *Unequal Childhoods*. Berkeley, CA: University of California Press.
- Lemke, M., Calsyn, C., Lippman, L., et al. (2002). *Outcomes of Learning. Results from the 2000 Program for International Student Assessment of 15-Year-Olds in Reading, Mathematics, and Science Literacy*. NCES 2002-115. Washington, DC: US Department of Education, Office of Educational Research and Improvement.
- Lu, M. C. and Halfon, N. (2003). Racial and ethnic disparities in birth outcomes: A life-course perspective. *Maternal and Child Health Journal* **7**, 13–30.
- Maynard, R. A. and Murnane, R. J. (1979). The effects of a negative income tax on school performance: Results of an experiment. *Journal of Human Resources* **14**(4), 463–476.
- McLoyd, V. C. (1990). The impact of economic hardship on Black families and children: Psychological distress, parenting and socioemotional development. *Child Development* **61**, 311–346.
- McLoyd, V. C., Jayaratne, T. E., Ceballo, R., and Borquez, J. (1994). Unemployment and work interruption among African American single mothers: Effects on parenting and adolescent socioemotional functioning. *Child Development* **65**, 562–589.
- Mikulecky, L. (1996). *Family Literacy: Parent and Child Interactions*. Washington, DC: US Department of Education. <http://www.ed.gov/pubs/FamLit/parent.html> (accessed September 2009).
- Mills, R. J. and Bhandari, S. (2003). *Health Insurance Coverage in the United States: 2002. U. S. Census Bureau. Current Population Reports*. Washington, DC: Government Printing Office.
- Mishel, L., Bernstein, J., and Shierholz, H. (2009). *The State of Working America 2008/2009*. Ithaca, NY: Cornell University Press.
- NCES (National Center for Education Statistics) (2003). *Digest of Education Statistics-2002*. NCES 2003-060. Washington, DC: US Department of Education, Office of Educational Research and Improvement.
- NCES (National Center for Education Statistics) (2007). *Digest of Education Statistics 2006*. NCES 2007-017. Washington, DC: US Department of Education, Office of Educational Research and Improvement.
- NCES (National Center for Education Statistics) (online). *National Assessment of Educational Progress, NAEP Data Explorer*. <http://nces.ed.gov/nationsreportcard/nde> (accessed September 2009).
- Neisser, U. (1997). Never a dull moment. *American Psychologist* **52**, 79–81.
- Neisser, U., Boodoo, G., Bouchard, T. J., Jr., et al. (1996). Intelligence: Knowns and Unknowns. *American Psychologist* **51**, 77–101.
- Neuman, M. J. and Peer, S. (2002). *Equal from the Start: Promoting Educational Opportunity for All Preschool Children – Learning from the French Experience*. New York: The French–American Foundation.
- OECD (Organization for Economic Cooperation and Development) (2001). *Knowledge and Skills for Life: First Results from the OECD Programme for International Student Assessment PISA 2000*. Paris, France: OECD.
- OFSTED (Office for Standards in Education) (1999). *Raising the Attainment of Minority Ethnic Pupils*. HMI 170. London: OFSTED Publications Centre/Author.
- Ogbu, J. (1992). Understanding cultural diversity and learning. *Educational Researcher* **21**(8), 5–14; 24.
- Orfield, A. (2007). *Eyes for Learning*. Lanham, MD: Rowman and Littlefield.
- Perez-Pena, R. (2003). Study finds asthma in 25% of children in Central Harlem. *The New York Times* April 19.
- Phillips, M., Brooks-Gunn, J., Duncan, G. J., Klebanov, P., and Crane, J. (1998). Family background, parenting practices, and the black–white test score gap. In Jencks, C. and Phillips, M. (eds.) *The Black–White Test Score Gap*, pp 103–148. Washington, DC: Brookings Institution Press.
- Richardson, G. A., Ryan, C., Willford, J., Day, N. L., and Goldschmidt, L. (2002). Prenatal alcohol and marijuana exposure: Effects on neuropsychological outcomes at 10 years. *Neurotoxicology and Teratology* **24**(3), 309–320.
- Rich-Edwards, J., Krieger, N., Majzoub, J., Zierler, S., Lieberman, E., and Gillman, M. (2001). Maternal experiences of racism and violence as predictors of preterm birth: Rationale and study design. *Paediatric and Perinatal Epidemiology* **15**(supplement 2), 124–135.
- Ritter, S. (2004). Asthma hits record rates among minority kids. *Chicago Sun-Times* January 8.
- Rothstein, R. (2004). *Class and Schools: Using Social, Economic and Educational Reform to Close the Black–White Achievement Gap*. New York: Teachers College.
- Rothstein, R., Carnoy, M., and Benveniste, L. (1999). *Can Public Schools Learn From Private Schools? Case Studies in the Public & Private Nonprofit Sectors*. Washington, DC: Economic Policy Institute.

- Sanbonmatsu, L., Kling, J. R., Duncan, G. J., and Brooks-Gunn, J. (2006). Neighborhoods and academic achievement: Results from the moving to opportunity experiment. *National Bureau of Economic Research Working Paper No. 11909*, January.
- Sciolino, E. (2004). France seems to try acting affirmatively on Muslims. *New York Times* January 15.
- Simmons, R. W., Wass, T., Thomas, J. D., and Riley, E. P. (2002). Fractionated simple and choice reaction time in children with prenatal exposure to alcohol. *Alcoholism: Clinical and Experimental Research* **26**(9), 1412–1419.
- Snow, C. and Tabors, P. (1996). *Intergenerational Transfer of Literacy*. Washington, DC: US Department of Education. <http://www.ed.gov/pubs/FamLit/transfer.html> (accessed September 2009).
- Starfield, B. (1982). Child health and socioeconomic status. *American Journal of Public Health* **72**, 532–534.
- Streissguth, A. P., Sampson, P. D., Olson, H. C., *et al.* (1994). Maternal drinking during pregnancy: Attention and short-term memory in 14-year old offspring – a longitudinal prospective study. *Alcoholism Clinical and Experimental Research* **18**(1), 202–218.
- Wadhwa, P. D., Culhane, J. F., Rauh, V., *et al.* (2001). Stress, infection, and pre-term birth: A biobehavioural perspective. *Paediatric and Perinatal Epidemiology* **15**(supplement 2), 17–29.

Relevant Websites

- www.bolderapproach.org – The Broader, Bolder Approach to Education.
- www.epi.org/publications/entry/books_grading_education/ – Grading Education: Getting Accountability Right (Teachers College Press and EPI, 2008).
- www.epi.org/publications/entry/books_class_and_schools/ – Class and Schools (Teacher College Press and EPI, 2004).

Neighborhoods and Peers in the Production of Schooling

J Vigdor, Duke University, Durham, NC, USA

J Ludwig, University of Chicago, Chicago, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

This article considers the role of peers and neighborhoods on educational outcomes. It begins by reviewing social scientific hypotheses linking peer and neighborhood characteristics to the amount of learning accomplished by youth in a given interval. It then summarizes recent empirical work testing these hypotheses, drawing upon earlier reviews to integrate earlier studies on the subject. While the issue of neighborhood or school segregation by social class or race/ethnic groups is of concern around the world, including in Europe, the former Soviet bloc, and South Africa, our article focuses primarily on empirical evidence drawn from the United States given the availability of high-quality longitudinal data sources and substantial changes over time in levels of segregation. In summary, recent experimental or quasi-experimental evidence suggests that neighborhood poverty, and associated characteristics have at most only modest impacts on educational outcomes at least over the medium run. Recent studies have more consistently found significant associations between school or classroom composition and student achievement. The effects are often small in magnitude, and in other cases they do not conclusively demonstrate that using policy to alter the assignment of students to schools or classrooms would necessarily lead to overall improvements in educational outcomes.

Why Peers and Neighborhoods Might Matter

Figure 1 presents a basic conceptual framework outlining the potential causal mechanisms linking schoolmates, classmates, and neighborhoods to academic outcomes (see also Jencks and Mayer, 1990 for an excellent discussion). Neighborhood factors could affect these outcomes directly or indirectly through the traditionally strong tendency for schools to serve particular neighborhoods or sets thereof. The strength of the link between neighborhood characteristics and schoolmate or classmate characteristics can be influenced by school attendance policies, including explicit desegregation or integration plans implemented by school districts (the use of which seems likely to decline in the wake of two 2007 Supreme Court decisions striking down school desegregation plans in Louisville and Seattle), or the use of school choice plans.

Even when there is no correlation between neighborhood and schoolmate or classmate characteristics,

neighborhood characteristics might still impact academic outcomes. Most directly, public schools often rely on local-source revenues to fund capital and operating expenses. Residence in a poor jurisdiction might therefore expose a student to inferior resources. School finance equalization policies enacted in virtually all states over the past three decades have undoubtedly ameliorated this concern to some extent. Nonetheless, evidence indicates that equalized spending across jurisdictions will not necessarily lead to an equalization of the quality of all relevant school inputs, since teachers exhibit clear preferences for working in lower-poverty environments.

Beyond this fiscal mechanism, the neighborhood social environment could shape youth expectations regarding appropriate forms of behavior and the return to schooling. Students in poor neighborhoods may perceive the return to schooling to be lower for several reasons. They may encounter few highly educated, successful neighbors, or may encounter a higher proportion of less educated yet relatively successful neighbors. Students may also perceive life expectancy to be shorter either because the incidence of chronic disease is higher in their neighborhood, or because they are exposed to a higher rate of violence.

The characteristics of schoolmates and classmates might associate with individual student achievement for any of several reasons. Given the importance of local financing for schools mentioned above, exposure to wealthier classmates might imply superior resources. In such a scenario, controlling for resource quality directly would reduce the magnitude of estimated schoolmate effects. However, education input quality is often difficult to measure. For example, it is widely acknowledged that teacher quality, defined as the ability to produce year-over-year increases in standardized test scores, varies substantially and is an important determinant of academic outcomes, but teacher quality correlates only modestly with commonly observed teacher characteristics and qualifications.

Schoolmates and classmates may contribute to the formation of students' expectations regarding the return to investments in human capital, but the relationship is not inherently clear. On the one hand, exposure to high-achieving peers may provide lower-achieving students with motivation to increase their effort. On the other hand, such exposure may do little more than convince lower-achieving peers that their ultimate schooling attainment is predetermined and that their own effort will have little impact on later outcomes. This sort of

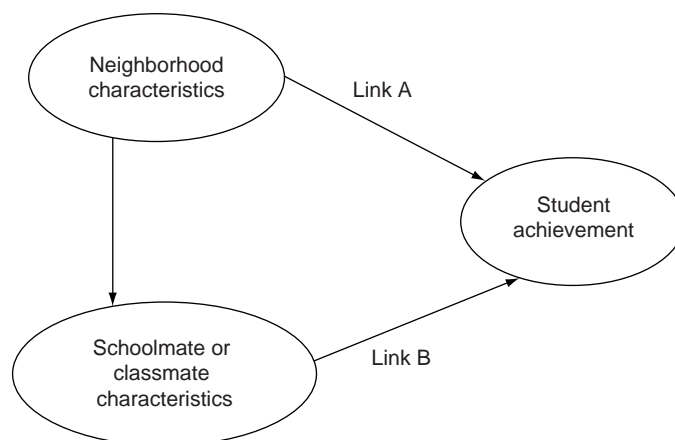


Figure 1 Link A mechanisms: community resources, student-expectation formation, violence or other disruptions to student learning; Link B mechanisms: parental resources, student-expectation formation, classroom disruptions, teacher expectations, and peer-to-peer education.

effect may be particularly acute in classrooms where achievement correlates with race, gender, or other easily observed characteristics associated with stereotypes. Exposure to lower-achieving peers has similarly ambiguous effects on the effort of higher-achieving students.

Even if schoolmates and classmates do not influence students' own expectations, they may influence the expectations of teachers, leading them to alter curricular choices in ways that may be beneficial or detrimental to any individual student. Teachers may tailor the overall difficulty level to the modal or median student, to the detriment of students who deviate strongly from that central tendency. Teachers may also be incentivized to improve the performance of small subgroups of students, to the detriment of others.

Peers, particularly classroom peers, may directly contribute to an individual's education, or detract from it. In the best-case scenario, students may learn through discussing topics with their classmates, or by directly asking their classmates with assistance in solving complex problems. In the worst-case scenario, classmates may interfere with the process of learning, by disrupting the entire classroom and preventing any instruction from occurring, or by disrupting individuals with off-topic conversation or other distractions. With these mechanisms, schoolmates and classmates may alter the actual return to investments in human capital, rather than the perceived return. Reading a book to prepare for a class discussion will generate higher rewards if other students also prepare; the return to paying attention in class may be diminished if the teacher cannot return that attention.

Understanding the distinct mechanisms that underlie peer and neighborhood effects is important for policy because policies like public or private school choice have the potential to change school but not neighborhood social compositions for children, while some housing-mobility interventions like the widely cited moving-to-opportunity

(MTO) experiment discussed below seem to generate larger changes in neighborhood than school characteristics. To the extent that peers or neighborhoods matter because they alter student perceptions, one could imagine interventions that directly target these perceptions, for example, by providing students more accurate information regarding the future returns to human capital investments. If peer and neighborhood effects simply reflect the impact of unmeasured inputs into education production, such as teacher quality, purely fiscal interventions might have a strong impact on the outcomes of disadvantaged students.

Evidence on Neighborhoods

Whether or how neighborhood context affects students' long-term life chances independently of school context remains somewhat unclear. The best evidence currently available suggests that moving from a poor or racially segregated neighborhood into a more socioeconomically or racially integrated one has at the most, very modest effects on children's test scores or other schooling outcomes, at least through the first 5 years or so following the neighborhood change.

Decades of research throughout the social sciences have documented substantial variation across neighborhoods in adult or child outcomes, even after conditioning on observable individual or family attributes (see e.g., Duncan *et al.*, 1997; Sampson *et al.*, 2002). Yet these studies, by and large, fail to provide convincing evidence in favor of any of these pathways, because they cannot separate the effects of neighborhoods *per se* from those of unmeasured individual attributes associated with residential selection.

The most compelling evidence for neighborhood effects on youth outcomes has come from quasi-experimental studies that exploit neighborhood changes induced by

housing programs. One of the best known of these quasi-experiments is Chicago's Gautreaux residential mobility program, which resulted from a 1976 Supreme Court decree finding that the Chicago Housing Authority and Housing and Urban Development (HUD) had engaged in systematic and illegal segregation. Follow-up interviews with Gautreaux families suggested improved outcomes for those who moved to the suburbs rather than other parts of Chicago, particularly for youth (Rubinowitz and Rosenbaum, 2001). For example, only 5% of suburban movers dropped out of high school compared to 20% of city movers, with an impressively large suburban advantage in college attendance as well (54% vs. 21%). However, more recent studies of Gautreaux show smaller impacts (Keels *et al.*, 2004). Moreover, Gautreaux has important limitations: the degree to which the program breaks the link between family preferences and mobility outcomes is unclear, since convincing documentation of the voucher offer and acceptance process remains difficult to reconstruct. In addition, Gautreaux studies cannot compare outcomes for program movers versus nonmovers or nonparticipants.

Motivated by the encouraging findings from Gautreaux, the US Department of HUD launched the MTO randomized housing-mobility intervention to identify the causal effects of residential-mobility interventions on families (Kling *et al.*, 2007). MTO has been in operation since 1994 in five cities (Baltimore, Boston, Chicago, Los Angeles, and New York). Eligibility for the voluntary MTO program was limited to low-income families with children living in public housing or Section 8 project-based housing located within designated high-poverty census tracts in these cities, with poverty rates of 40% or more.

By random lottery, some families were assigned to a Section 8 group that was offered unrestricted Section 8 subsidies (vouchers) that they could use to move to a new private-market apartment of their choice. (The Section 8 program provides housing assistance through rental certificates or vouchers. Income eligibility is usually set to 50% of local median income, with subsidies equal to the difference between 30% of tenant income and an area-wide threshold established by HUD. Since MTO started, the voucher program has been renamed the housing choice voucher (HCV) program, although we will use the terminology in effect at the start of MTO.) Other families were assigned to an experimental group that was offered housing-search assistance and Section 8 subsidies that could only be used to relocate to a low-poverty census tract (with a 1990 poverty rate below 10%). A third control group received no special assistance from MTO but remained eligible for public or project-based housing and other social programs.

The explicit goal of MTO was to help move families into less economically distressed communities, and as shown in **Table 1**, by this measure MTO was successful.

One year after random assignment families in the two MTO treatment groups live in census tracts with average poverty rates 11–13 percentage points (25–30%) below those of the control group. (Parentheses in **Table 1** show standard errors for the group means, which provides some sense for the statistical uncertainty surrounding our estimate.) The gap declines somewhat over time in part because of subsequent mobility among all groups. But even 6 years after, the treatment-control differences in tract poverty equal 7–8 percentage points (20% of the control mean), while the differences in cumulative exposure to neighborhood poverty (duration weighted averages) are 9–10 percentage points (20–25%). However, MTO treatment assignment had much more modest impacts on the neighborhood racial composition of participating families. Evidence also suggests that the treatment-control differences in the characteristics of public schools attended were also very modest. Thus, evidence of treatment-control differences in youth outcomes from a study such as MTO would point more clearly to causal pathways that did not involve school characteristics.

Table 2 summarizes results for MTO youth, drawn from analyses conducted 4–7 years after random assignment. **Table 2** presents mean outcomes for the control group together with the difference in average outcomes between those assigned to the MTO experimental group versus those assigned to the MTO control group (known in the program evaluation literature as the intent-to-treat effect, or ITT, which will understate the effects of actually relocating through the program because not all families assigned to the experimental group actually relocate). These data yield no evidence for statistically significant differences across MTO-assigned groups in reading or math-achievement scores measured using the Woodcock–Johnson revised tests (Sanbonmatsu *et al.*, 2006). While a growing body of research in developmental psychology, neuroscience, and economics suggest that children may be more sensitive to environmental changes early in life, there is no statistically significant evidence to date from MTO that treatment-control differences in test scores vary by age at random assignment, at least within the range of ages found among those MTO children who were administered tests. Those MTO participants who were very young children at baseline were not tested as part of the interim study, and so it remains possible that MTO effects on their achievement may be more pronounced.

There is a hint in the MTO data that female youth assigned to the experimental group may be somewhat more likely to graduate from high school than those assigned to the control group, although the point estimate is not statistically significant and the estimated experimental-treatment effect on graduation for male youth is negative (and also not statistically significant). More generally, the interim MTO evaluation finds that treatment-group assignment improves a variety of behavioral outcomes for female

Table 1 Mobility characteristics of MTO households^a

	<i>Controls</i>	<i>Experimental group</i>		<i>S8-only group</i>	
	<i>All</i>	<i>All</i>	<i>Program movers only</i>	<i>All</i>	<i>Program movers only</i>
# Moves since RA	1.345 (0.060)	1.741 (0.050)	2.290 (0.061)	1.864 (0.087)	2.416 (0.099)
<i>Tract share poor:</i>					
1 year post RA	0.454 (0.007)	0.328 (0.008)	0.194 (0.009)	0.344 (0.011)	0.286 (0.011)
6 years post RA	0.376 (0.008)	0.296 (0.007)	0.206 (0.007)	0.307 (0.010)	0.283 (0.010)
Average through 6 years post RA ^b	0.422 (0.006)	0.316 (0.007)	0.203 (0.005)	0.335 (0.008)	0.293 (0.008)
<i>Tract share college Ed.:</i>					
1 year post RA	0.147 (0.006)	0.211 (0.006)	0.278 (0.008)	0.169 (0.007)	0.178 (0.009)
6 years post RA	0.156 (0.006)	0.197 (0.005)	0.236 (0.007)	0.176 (0.007)	0.176 (0.009)
Average through 6 years post RA ^b	0.144 (0.004)	0.196 (0.004)	0.246 (0.006)	0.167 (0.006)	0.173 (0.007)
<i>Tract share minority:</i>					
1 year post RA	0.889 (0.009)	0.800 (0.011)	0.676 (0.017)	0.875 (0.013)	0.841 (0.017)
6 years post RA	0.885 (0.010)	0.821 (0.010)	0.744 (0.015)	0.859 (0.015)	0.859 (0.015)
Average through 6 years post RA ^b	0.886 (0.008)	0.801 (0.009)	0.700 (0.013)	0.855 (0.012)	0.832 (0.015)
<i>N</i>	426	701	361	263	172

^aRA – random assignment. Mean values shown with standard errors in parentheses. Sample consists of adults interviewed during the MTO interim evaluation and randomized by 12/31/1995 (results 1 and 4 years post RA are similar for full sample of MTO adults). Tract characteristics are from the 2000 census. Number of moves was calculated using the adult's address history.

^bAverage post RA neighborhood attributes calculated using address duration as weights (or duration-weighted postrandomization neighborhood averages).

Table 2 Impact of MTO on selected child and youth outcomes, interim evaluation^a

	<i>Females</i>		<i>Males</i>	
	<i>Control mean</i>	<i>Experimentals minus controls</i>	<i>Control mean</i>	<i>Experimentals minus controls</i>
<i>Experimental vs. control group</i>				
Reading test z-scores (6–20)	0.103	0.060 (0.038)	–0.096	–0.002 (0.045)
Graduated or still enrolled (15–20)	0.772	0.064 (0.036)	0.759	–0.044 (0.037)
Used pot last 30 days (15–20)	0.131	–0.059 ^b (0.028)	0.118	0.053 (0.030)
No. arrests violent crime (15–25)	0.241	–0.077 ^b (0.031)	0.537	–0.045 (0.051)
No. arrests property crime (15–25)	0.164	–0.057 ^b (0.026)	0.474	0.150 ^b (0.055)
Psych distress K6, z-score (15–20)	0.268	–0.246 ^b (0.091)	–0.162	0.069 (0.091)
Has fair or poor health (15–20)	0.101	0.021 (0.027)	0.045	0.033 (0.019)
<i>Section 8 vs. control group</i>				
Reading test z-scores (6–20)	0.103	0.012 (0.043)	–0.096	0.033 (0.046)
Graduated or still enrolled (15–20)	0.772	0.049 (0.037)	0.759	–0.040 (0.041)
Used pot last 30 days (15–20)	0.131	–0.052 (0.030)	0.118	0.075 ^b (0.035)
No. arrests violent crime (15–25)	0.241	–0.079 ^b (0.036)	0.537	0.024 (0.062)
No. arrests property crime (15–25)	0.164	0.031 (0.039)	0.474	0.072 (0.059)
Psych distress K6, z-score (15–20)	0.268	–0.133 (0.104)	–0.162	–0.027 (0.096)
Has fair or poor health (15–20)	0.101	–0.003 (0.027)	0.045	0.033 (0.023)

^aAge range of analytic sample as of 31 December, 2001 is reported in parentheses next to each variable. For more details see Kling, Ludwig and Katz (2005), Kling, Liebman and Katz (2007), and Sanbonmatsu *et al.* (2006).

^bAverage post RA neighborhood attributes calculated using address duration as weights (or duration-weighted postrandomization neighborhood averages).

youth but on balance has detrimental effects on behavior for male youth (Kling *et al.*, 2005, 2007).

One way to reconcile the MTO and Gautreaux findings is that the latter engendered more neighborhood racial segregation than the former, and so perhaps it is the racial rather than social class composition of a community that matters for youth schooling outcomes. This possibility can be tested by exploiting variation across MTO sites in treatment effects on mobility outcomes to use site-treatment group interactions as instruments for specific neighborhood characteristics. Census tract minority composition is a stronger predictor of youth violent criminal behavior than is either the tract poverty rate or local-area crime rate (Ludwig and Kling, 2007). However, in unpublished results, MTO researchers have found no evidence for an association between tract minority composition and children's achievement test scores.

Evidence, in some measure, of a small role for neighborhood socioeconomic characteristics from the MTO study is corroborated by recent quasi-experimental studies of families randomly assigned to public housing units in Toronto (Oreopoulos, 2003) and families displaced by public housing demolition in Chicago (Jacob, 2004). Studies using sibling comparisons or propensity score-matching methods (which are similar in spirit to methods that use statistical regression to control for observable differences across children) do report significant associations between neighborhood characteristics and youth outcomes. Studies have also documented a tendency for black-white disparities in educational outcomes to be more severe in more residentially segregated cities. But each of these methodologies is subject to criticism, and none can exclude the possibility that schools are the main causal pathway between neighborhood characteristics and outcomes. The next section focuses more specifically on arguments that racial disparities in school characteristics, made possible by school segregation, influence the magnitude of the black-white test-score gap.

Evidence on Schoolmates and Classmates

The methodological equivalent of conducting an MTO-style experiment to assess the importance of schoolmate or classmate characteristics would involve randomly assigning students to classes of varying composition, for potentially varying periods of time. To our knowledge, no such experiment has taken place at any point in history. There has been one noteworthy attempt to use a randomized trial, initially intended to study the impact of class size, to study the impact of assignment to peers of varying ability levels. By randomly assigning students to classes of different sizes, the Tennessee student teacher achievement ratio (STAR) experiment effectively assigned some students to higher

across-classroom variance in peer ability. If mean classmate ability is a predictor of achievement, then the wider variance in mean ability observed in smaller classes should produce a greater degree of variance in the ultimate outcome among students assigned to small classes. This exact pattern appears in the experimental data, indicating that students perform better when assigned to classes with higher mean ability (Graham, 2008).

Nonexperimental studies face the challenge of separating the influence of peers from the impact of unmeasured determinants of achievement that may correlate with peer characteristics. A common strategy for addressing this issue is the use of within-school variation in peer characteristics, across year-of-birth cohorts or classrooms within schools. This addresses the concern that individuals with strong unobservable characteristics sort into schools with observably more-able peers. The school fixed-effect strategy also eliminates at least one hypothesized mechanism linking peer ability to student outcomes – the tendency for more affluent parents to invest more resources in their children's school. Estimates derived from school fixed effects might therefore understate the benefit to some children from attending a school with more-able classmates, if these benefits accrue to the school at large and not to the specific classroom or cohort to which they belong.

Bearing this caveat in mind, several prominent studies that control for school fixed effects report patterns entirely consistent with experimental estimates. These studies use large administrative databases covering the population of students in Texas, North Carolina, or Florida. While not universally supportive of a positive association between peer ability and individual test scores, most of these studies report such a finding in at least some specifications.

Ability, as measured by prior year test scores, is not necessarily a sufficient statistic for all the potential peer characteristics that could influence student achievement. Researchers have also devoted considerable attention to the relationship between classroom racial or gender composition, or average socioeconomic status, and standardized test scores.

Some quasi-experimental studies have attempted to infer the impact of racial composition on outcomes by exploiting variation in the timing of school desegregation orders in the 1960s and 1970s, as well as unitary status declarations in a later time period. These studies rest on the assumption that variation in the timing of such court orders is idiosyncratic from the perspective of the individual student. Court-ordered school desegregation plans appear to have reduced black dropout rates by 2–3 percentage points, with no detectable effect on whites (Guryan, 2004). Qualitatively similar – but reversed – effects have been associated with the termination of many of these desegregation plans during the 1990s (Lutz, 2005). These court-ordered desegregation plans may also reduce violent crime-victimization rates among black and white youth, and

offending rates by blacks and perhaps whites as well (Weiner *et al.*, 2007).

Among studies examining school-level variation, which are subject to concerns regarding nonrandom sorting across schools, the Coleman report of 1966 is perhaps the best known (Coleman *et al.*, 1966). This study found that a school's socioeconomic class composition was among the more important predictors of individual student-achievement levels, certainly much more important than school racial composition. Recent studies that use cross-sectional variation in school segregation often generate similar findings – school racial composition seems to be weakly if at all related to student outcomes, while socioeconomic status (SES) of the school's student body may have a stronger association with student achievement (Rivkin, 2000; Cook and Evans, 2000; Card and Rothstein, 2006).

One study that used administrative data from Texas and controlled for school fixed effects reported that a 10% reduction in percentage of black in school would increase test scores for blacks by around 0.025 standard deviations and increase test scores for whites by around 0.01 standard deviations (the latter estimate is not statistically significant) (Hanushek *et al.*, 2004). These effects are about twice as large for black students at the top of the achievement distribution compared to those at the bottom of the distribution. Importantly, their finding holds even after controlling for either the average achievement level or social class composition of other students in the school. These authors control for a variety of potential confounding factors, including student-specific rates of change in achievement test scores and hard-to-measure factors that vary at the level of the school-by-grade or even the attendance zone-by-year.

A separate study of Texas data used plausibly random variation across cohorts in student demographic composition in Texas and found that a 10% increase in the share of one's classmates that are black reduces achievement test scores for blacks by around 0.1 standard deviations in reading and around 0.06 standard deviations in math (Hoxby, 2000). The effects of changes in segregation appear to be strongest in schools that are composed of at least one-third black students already. The effect of changes in racial segregation in the school on white students is of the same sign as for black students but only around one-quarter as large in magnitude. These effects may understate the impact of segregation because they exclude school-level causal mechanisms, or may overstate them if the restriction to cross-cohort variation in peer group composition fails to eliminate all potential sources of omitted variable bias.

However, more recent research using administrative data on North Carolina public schools to examine the relationship between peer composition and test scores at both the school and classroom levels fail to show any negative association between peer percent black and the

relative performance of black students (Vigdor and Nechyba, 2007). Relative to the Texas administrative data, the North Carolina data include more detailed family-background information and permit the matching of students to classrooms and teachers in elementary grades. The authors also analyze variation in peer composition associated with the opening of new schools in rapidly growing districts, by examining how achievement scores change for students that have a substantial change in peer-group composition (without moving schools themselves). The absence of peer effects could be attributed to differences in methodology, including the focus on elementary rather than middle-school grades.

Discussion

Overall, the most reliable evidence suggests that students learn more when they are placed in a school environment with more able peers, or more socioeconomically advantaged peers. Changing a student's neighborhood environment, without altering the school environment, appears to have at most only a modest impact on educational outcomes at least over the medium run. The frequently voiced concern that nonexperimental estimates of the relationship between neighborhood or schoolmate characteristics are contaminated by omitted-variable bias receives substantial support in the form of markedly attenuated effect estimates in more recent experimental or quasi-experimental literature.

The significant relationships estimated between classmate characteristics, most notably ability as measured by standardized test scores, and individual student achievement implies that the distribution of peers across classrooms has significant equity implications. The segregation of schools by factors that correlate with ability, or the stratification of classrooms by ability (commonly known as tracking) confers advantages on already-advantaged students and disadvantages on already-disadvantaged students. Integrating schools or classrooms by ability would amount to a progressive redistribution of resources across households.

While the equity implications of peer effects are clear, the implications for overall student learning (i.e., efficiency) are less clear. Would it be possible to rearrange students in a school, district, or metropolitan area in a manner that produced education gains that exceeded education losses? Would it be possible to rearrange pupils in a way that created gains without any losses whatsoever? For several reasons, existing literature has not provided details sufficient to answer these questions. While literature has provided fairly clear evidence that peer ability matters, it is much less clear whether the impacts of peer ability on individual students is nonlinear. Only in the presence of nonlinearities in peer effects, or in the presence

of some groups being completely immune to peer effects, would reallocation of students across settings improve the overall average level of learning for all students.

A second reason why existing literature is insufficient to address efficiency concerns is that standardized test scores, the most common outcome variable in the peer effects literature, have no obvious mapping into the ultimate goals that society holds for the development and well-being of children. A test-score gain for a student at the bottom end of the ability distribution may be much more socially valuable than a comparable gain to a student at the top end. The reverse might also be true. Ultimately, society's valuation of education improvements for individuals of differing inherent ability is a subjective matter; seen from this light, even the strongest evidence on the operation of nonlinear peer effects would be unlikely to settle debates regarding the most appropriate method of assigning students to classrooms within and across schools.

See also: Desegregation, Academic Achievement and Earnings; Education Production Functions: Concepts; The Economics of Charter Schools; The Economics of Parental Choice; The Efficacy of Educational Vouchers; Tiebout Sorting and Competition.

Bibliography

- Coleman, J. S., et al. (1966). *Equality of Educational Opportunity*. Washington, DC: Government Printing Office.
- Cook, M. D. and William, W. N. (2000). Families or schools? Explaining the convergence in white and black academic performance. *Journal of Labor Economics* 18(5), 729–754.
- Duncan, G., Brooks-Gunn, J., and Aber, J. L. (eds.) (1997). *Neighborhood Poverty: Context and Consequences for Children*, vols. I and II. New York: Russell Sage Foundation.
- Graham, B. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica* 76(3), 643–660.
- Guryan, J. (2004). Desegregation and black dropout rates. *American Economic Review* 94(4), 919–943.
- Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2004). New evidence about Brown vs. Board of Education: The complex effects of school racial composition on achievement. *Working Paper*, Hoover Institution, Stanford University.
- Hoxby, C. (2000). *Peer Effects in the Classroom: Learning from Gender and Race Variation*. Cambridge, MA: NBER. Working Paper 7867.
- Jacob, B. A. (2004). Public housing, housing vouchers, and student achievement: Evidence from public housing demolitions in Chicago. *American Economic Review* 94, 233–258.
- Jencks, C. and Mayer, S. E. (1990). The social consequences of growing up in a poor neighborhood. In Lynn, L. E. Jr. and McGeary, M. G. H. (eds.) *Inner-City Poverty in the United States*, pp 111–186. Washington, DC: National Academy Press.
- Keels, M., Rosenbaum, J., and Duncan, G. (2004). Second generation effects of Chicago's Gautreaux Residential Mobility Program on crime. *Paper Presented at October 28–30 Annual APPAM Research*

- Conference on Creating and Using Evidence in Public Policy Analysis and Management*. Atlanta, GA.
- Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica* 75(1), 83–119.
- Kling, J. R., Ludwig, J., and Katz, L. F. (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *Quarterly Journal of Economics* 120(1), 87–130.
- Ludwig, J. and Kling, J. R. (2007). Is crime contagious? *Journal of Law and Economics* 50(3), 491–518.
- Lutz, B. F. (2005). Post Brown vs. the Board of Education: The effects of the end of court-ordered desegregation. *Federal Reserve Board, Finance and Economics Discussion Series Working Paper 2005-64*.
- Oreopoulos, P. (2003). The long-run consequences of living in a poor neighborhood. *Quarterly Journal of Economics* 118, 1533–1575.
- Rivkin, S. G. (2000). School desegregation, academic attainment and earnings. *Journal of Human Resources* 35(2), 333–346.
- Rubinowitz, L. S. and Rosenbaum, J. E. (2001). *Crossing the Class and Color Lines: From Public Housing to White Suburbia*. Chicago, IL: University of Chicago Press.
- Sampson, R. J., Morenoff, J. D., and Gannon-Rowley, T. (2002). Assessing 'Neighborhood Effects': Social processes and new directions in research. *Annual Review of Sociology* 28, 443–478.
- Sanbonmatsu, L., Kling, J. R., Duncan, G. J., and Brooks-Gunn, J. (2006). Neighborhoods and academic achievement: Results from the moving to opportunity experiment. *Journal of Human Resources* 41(4), 649–691.
- Vigdor, J. L. and Nechyba, T. (2007). Peer effects in North Carolina public schools. In Peterson, P. E. and Woessmann, L. (eds.) *Schools and the Equal Opportunity Problem*, pp 73–102. Cambridge: MIT Press.
- Weiner, D. A., Lutz, B. F., and Ludwig, J. (2007). The effects of school desegregation on crime. *Paper Presented at the Annual Meetings of the American Economic Association*, Chicago, IL.

Further Reading

- Clotfelter, C. T. (2004). *After Brown: The Rise and Retreat of School Desegregation*. Princeton, NJ: Princeton University Press.
- Cooley, J. (2007). *Desegregation and the Achievement Gap: Do Diverse Peers Help?* Unpublished manuscript. Madison, WI: University of Wisconsin.
- Cutler, D. M. and Glaeser, E. L. (1997). Are ghettos good or bad? *Quarterly Journal of Economics* 112, 827–872.
- Hanushek, E. A. and Rivkin, S. G. (2006). School quality and the Black–White achievement gap. *NBER Working Paper #12651*.
- Keels, M., Duncan, G. J., Deluca, S., Mendenhall, R., and Rosenbaum, J. (2005). Fifteen years later: Can residential mobility programs provide a long-term escape from neighborhood segregation, crime and poverty? *Demography* 42(1), 51–73.
- Manski, C. F. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives* XIV, 115–136.
- Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica* 73(2), 417–458.
- Vigdor, J. L. (2006). Peer effects in neighborhoods and housing. In Dodge, K., Dishion, T., and Lansford, J. (eds.) *Deviant Peer Influences in Programs for Youth: Problems and Solutions*, pp 185–202. New York: Guilford Press.
- Wilson, W. J. (1987). *The Truly Disadvantaged: The Inner City, The Underclass, and Public Policy*. Chicago, IL: University of Chicago Press.

Teacher Quality in Education Production

I Pelayo and D J Brewer, University of Southern California, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

What is a good teacher? Do good teachers make a difference in improving student achievement? While these are simple questions, the answers are more complex. Policy-makers and educators are searching for strategies to improve student outcomes. In the US, the 2001 the No Child Left Behind Act (NCLB) requires that all classroom teachers be highly qualified. The assumption is that highly qualified teachers will produce higher measured student achievement. NCLB has set certain criteria for determining the credentials that such teachers must have, but does little to define the characteristics and skills a teacher must possess in order to be considered high quality. In fact, the research evidence on what observable teacher characteristics impact student achievement provides little guidance. In this article, we discuss empirical evidence on teacher quality in education production. Several comprehensive reviews of the literature have been conducted (see the sections titled ‘Bibliography’ and ‘Further reading’), so our focus here is on more recent studies.

The Education Production Function

The education production function is an economic formulation of an input–output model: measured input variables, such as teacher quality, are mapped to a certain measured output, usually student achievement test scores or test score gains. Although a few experimental studies exist (e.g., Glazerman *et al.*, 2006), the empirical estimation of education productions is typically via multivariate regression models in quasi-experimental or nonexperimental settings. Hence, the vast bulk of the research evidence is able to detect correlational relationships between various characteristics and student achievement, with no or weak inferences about causal effects. Further, early studies of education production were typically conducted at an aggregate level – for example, examining mean years of teaching experience or the percentage of teachers with certification, in a school or district. More recent research has been able to take advantage of student-level longitudinal data in which it is possible to identify the characteristics of the specific teacher that taught a particular student.

The landmark study by Coleman *et al.* (1966), *Equality of Educational Opportunity* (also known as the Coleman report), which was commissioned by the US federal government, is among one of the most widely known education production studies. Coleman found that measured

school resources explained a small portion of the variance in student achievement. This finding was (mis)interpreted to mean schools do not matter. However, the main issue is whether there exist systemic differences in the ability of schools and teachers to raise student achievement (Hanushek *et al.*, 2005). Many researchers have demonstrated that schools matter a great deal (Clotfelter *et al.*, 2006; Rivkin *et al.*, 2005) and that teacher quality is the single most important factor among all school-related components (Hanushek, 2002; Rivkin *et al.*, 1998; Sanders and Rivers, 1996). Not only can teachers positively impact their students (Goldhaber, 2007; Clotfelter *et al.*, 2006), a series of high-quality teachers can potentially mitigate socioeconomic inequalities for individual students. Hanushek (2002) notes that “schools have the ability to compensate for educational differences arising from family backgrounds [where a long] string of above average teachers can entirely close the achievement gap, by income level” (p. 27). However, there are large differences in teacher impact on student achievement and identifying the high-quality teachers is problematic.

What Is Teacher Quality?

The majority of production function research assumes that observable characteristics of teachers are important determinants of student achievement. Teacher quality is a combination of observable and unobservable characteristics (Goldhaber and Anthony, 2007; Rivkin *et al.*, 2005; Hanushek *et al.*, 2005). For example, two easily observable teacher characteristics are years of experience and degree level (Boyd *et al.*, 2006; Clotfelter *et al.*, 2006; Glazerman *et al.*, 2006; Hanushek *et al.*, 2005; Rivkin *et al.*, 2005; Rockoff, 2004). These two variables are typically the basis for paying teachers in the US and hence commonly collected information. More refined measures may also be utilized including the degree major or coursework, in an attempt to determine the importance of subject-matter knowledge or pedagogical training. Other characteristics include proxies for ability such as the quality of the college or university (indicated by the selectivity of the institution) attended by the teacher (Murnane and Phillips, 1981; Ehrenberg and Brewer, 1994; Clotfelter *et al.*, 2006; Glazerman *et al.*, 2006), or tests of teacher ability such as scores on the National Teacher Exam, Scholastic Aptitude Test (SAT) or American College Testing (ACT) scores, or

specially designed tests (Ehrenberg and Brewer, 1994, 1995; Summer and Wolfe, 1977, 1975; Ferguson and Ladd, 1996). Similarly, various studies have examined the relationship between teacher certification and student achievement (Goldhaber and Brewer, 2000; Goldhaber, 2007; Goldhaber and Anthony, 2007).

Unobservable characteristics refer to characteristics not easily measured. Some examples may include pedagogy used by the teacher, classroom management skills, philosophy of teaching, and interpersonal or social skills (Goldhaber and Anthony, 2007; Glazerman *et al.*, 2006; Rockoff, 2004). Goldhaber *et al.* (1999) estimate that as much as 97% of teacher effects are attributable to unobservable characteristics. Data are typically not available for most of these factors, and it is difficult to separate them out conceptually or design policy interventions that can affect them with much certainty. In the next section, we focus, therefore, on the variables most commonly studied: teacher experience, certification, subject-matter preparation, advanced degrees held, and ability.

Teacher Experience

Many education production studies have examined whether or not teacher experience matters. Given data limitations, research is typically able to simply count a teacher's number of years of experience, treating all years as identical; hence, there is usually no information on the quality of any given year of experience. While hundreds of education production studies have been conducted since the Coleman report, they have not necessarily reached similar conclusions. Hanushek (1986), one of the most widely cited reviews of education production literature, concluded that there was no strong evidence that a teacher's years of experience positively effects student achievement; for example, in his 1989 update on school expenditures and performance, he found that only 29% of the 140 studies that included teacher experience as an input had statistically positive coefficients (Hanushek, 1989). However, Greenwald *et al.* (1996), in their own review of similar studies found experience was related to student achievement. These divergent conclusions are explained largely by differences in weighting of the studies and methodologies used to aggregate the results (Eide *et al.*, 2004).

The consensus view is that the experience gained in the first few years of teacher's career are the most important. In other words, a teacher's learning curve is steep (Murnane, 1975; Boyd *et al.*, 2006; Clotfelter *et al.*, 2006; Hanushek *et al.*, 2005; Rivkin *et al.*, 2005; Rockoff, 2004). The issue is whether some years of experience matter more than others, and if so, which ones. Although there is disagreement about the number of years that makes a difference, studies show that most improvement occurs in the earlier years. Rivkin *et al.* (2001) find that students'

learning gains are small after their teacher's first few years of experience. Rockoff (2004) measures the impact of individual teachers on student achievement by using panel data that matches student test scores to teacher assignment, and finds that teaching experience significantly improves student outcomes, but only in the first few years. After teachers reach the cutoff point of about 10 years, teaching experience offers only marginal returns (Rockoff, 2004). According to Rockoff (2004), reading scores in particular differ by approximately 0.17 standard deviations on average between beginning teachers and teachers with 10 or more years of experience.

Similarly, Rivkin *et al.* (2005) suggest that beginning teachers perform worse than more experienced teachers and find important gains in teacher quality after the first year of experience. They suggest that teachers early in their careers engage in learning by doing, finding little evidence that improvement continues after the first 3 years of teaching (Rivkin *et al.*, 2005). Furthermore, they suggest that the effects of nonrandom selection – where lower-quality teachers stay in the profession and more talented teachers exit the profession – affect the average quality of the teaching pool (Goldhaber, 1997; Hanushek *et al.*, 2005). They also suggest that teachers “may vary [their] efforts systematically with experience in response to tenure decisions or other institutional and contractual issues,” and this could also affect student achievement (p. 17). Consistent with other studies, Clotfelter *et al.* (2006) examine teacher experience and find student achievement peaking in classrooms with teachers that have between 13 and 26 years of experience. The purpose of their study is to examine teacher effectiveness given that the matching of teacher and student is nonrandom (Clotfelter *et al.*, 2006). Taken together, these studies provide strong evidence that the early years of teaching matter a great deal, and observable effects of over 10 years of experience are quite small or negligible (Clotfelter *et al.*, 2006; Hanushek *et al.*, 2005; Rockoff, 2004).

Teacher Certification

The minimum requirement for teaching in most states in the US is some form of certification. Discussions of teacher certification and preparation programs are inherently complex because of the wide variation in requirements (Darling-Hammond and Youngs, 2002). Earlier studies looking at teacher education do not necessarily separate teacher certification from subject-matter preparation and degree (Hanushek, 1994, 1989, 1986). Although teacher certification is required, there is little evidence from which to determine whether or not teacher certification itself has a significant impact on student achievement (Boyd *et al.*, 2007).

Studies on teacher certification find conflicting evidence. Goldhaber and Brewer (2000) find that students of teachers with standard certification in mathematics had higher achievement than students of teachers who are not credentialed in their subject area. Yet, the authors also find that mathematics and science students with emergency credentialed teachers do no worse than students of teachers with standard credentials, suggesting that simply having a standard credential does not make for a high-quality teacher, rather, having teachers credentialed in their subject area is what makes a difference. Unfortunately, studies of certification are hampered by the lack of good data that actually describe the content, sequencing, and quality of certification programs, which differ widely across states and from program to program within a state. Data often do not even permit clear distinctions among emergency, noncredentialed, standard, or alternative certification or credentials, making it difficult to interpret findings in a way that can improve the outcomes of the credentialing process.

In recent years, various alternative teacher certification programs, including district-sponsored programs as well as national efforts such as Teach for America (TFA), have become more widespread. According to Glazerman *et al.* (2006), TFA novice teachers are as effective in producing student achievement as other (traditional or other alternative program) novice teachers and experienced teachers alike in producing gains in math. The effect size on math scores was about 15% of a standard deviation. They also find TFA teachers are no worse than other novice and experienced teachers with respect to student achievement gains in reading (Glazerman *et al.*, 2006). This study is important because it randomly assigned teachers to students within a school district. Boyd *et al.* (2006) find small differences in gains, typically about 2–5% of a standard deviation, among students who had TFA or New York City Teaching Fellows teachers and teachers from traditional college-based teacher-preparation programs and alternative programs; however, these small differences disappear within 3 years as the teachers mature. These findings are consistent with the findings on teacher experience, and suggest the effects of teacher characteristics interact with each other in ways that are not always separable.

Degree Level and Subject-Matter Preparation

Many teachers in the United States earn advanced degrees, a plausible indicator of teacher quality. However, the consensus view is that a teacher education *per se* does not significantly impact student achievement (Hanushek, 1994, 1989, 1986; Murnane, 1975; Murnane and Phillips,

1981; Summer and Wolfe, 1975, 1977). Hanushek (1994) found that 100 of the 113 studies, including teacher education as an input, have statistically insignificant results and of the remaining 13, the positive and negative results are divided. More recently, Hanushek *et al.* (2005) and Rivkin *et al.* (2005) suggest that while advanced degrees are important for teacher compensation, they are not important for student achievement, which is consistent with earlier studies.

However, there is some evidence that advanced degrees in the subject being taught do make a difference, at least in math and science. For example, Goldhaber and Brewer (2000, 1997b) find that teachers with advanced degrees in mathematics produced greater high school student achievement in math. Similarly, Monk and Rice (1994) studied eighth-grade students and found that teachers who took several math courses in college had a positive impact on student achievement, but only when these teachers were assigned to teach more advanced math courses.

Teacher Ability

Teacher ability is another indicator of teacher quality. According to Eide *et al.* (2004), teacher ability is a stronger predictor of student achievement than years of experience or advanced degrees held. In a reanalysis of the data from the Coleman report, Ehrenberg and Brewer (1994) find a positive relationship between teacher performance on a short verbal ability test and student achievement gains. Rowan *et al.* (1997) reach a similar conclusion using one question on the National Educational Longitudinal Study (NELS) of 1998 teacher survey: those who accurately answered a math quiz item show a 0.018 effect size on student achievement in math at schools with average-ability students (Rowan *et al.*, 1997). Ferguson (1991) finds a link between student achievement and teacher's language score on the Texas Examination of Current Administrators and Teachers (TECAT). In an earlier study, Strauss and Sawyer (1986) find that a 1% increase in teacher quality, as measured by teacher test scores, results in a 5% decline in student failure rate on high school competency exams. Ehrenberg and Brewer (1994) find that students learn more from teachers who attended more selective undergraduate institutions. While Greenwald *et al.* (1996) found that teaching experience and degree levels were statistically significant and positively related to student achievement in 15% and 29% of education production function studies, respectively, they report teacher academic proficiency was positive in 50% of the studies. Although not a complete measure of teacher quality, teacher ability is clearly a quantifiable predictor of teacher effectiveness in the classroom.

Conclusion

One common theme from studies of teacher quality in education production has been a growing interest in how teacher labor markets work. If teacher quality is important in affecting student achievement, then questions pertaining to teacher attrition, retention, and recruitment become critical (Glazerman *et al.*, 2006; Boyd *et al.*, 2006; Rockoff, 2004). Similarly, how teachers are distributed across students is important. Does the sorting of teachers evolve into patterns of teacher–student matching that affect student achievement? How does this occur? The empirical evidence suggests that teacher sorting and teacher–student matching is not random (Clotfelter *et al.*, 2006; Rivkin *et al.*, 2005; Lankford *et al.*, 2002). Generally, schools in urban and low-performing districts tend to have difficulties attracting and retaining teachers (Hanushek *et al.*, 2004). In a descriptive study of teacher sorting of New York city schools, Lankford *et al.* (2002) find that these schools lose teachers to schools with more appealing environments. Some school districts find tempered relief from alternative teacher-preparation programs such as TFA and New York City Teaching Fellows, which place teachers exclusively in hard-to-staff schools with the hope to improve the educational opportunities of less-advantaged students (Glazerman *et al.*, 2006; Boyd *et al.*, 2006). More experienced and highly qualified teachers tend to be in schools that are more high performing, affluent, white, and suburban; inexperienced, less-qualified teachers work in schools that are low performing, poor, minority, and urban (Clotfelter *et al.*, 2006; Lankford *et al.*, 2002; Jacob, 2007).

Teacher sorting also occurs within schools. For example, it may be the case that tenured teachers are more able to leverage their seniority to influence teaching assignments in their favor, leaving more difficult positions to teachers with less seniority. Clotfelter *et al.* (2006) suggest that experienced teachers are better able to resist being assigned to less-able students. Yet, low-performing students are in greatest need of more experienced and effective teachers (Rivkin *et al.*, 2005). These placements in turn contribute to high teacher attrition rates among new teachers (Murnane and Steele, 2007). Unfortunately, to respond to the large numbers of vacancies, many schools place inexperienced, noncredentialed substitute teachers in classrooms. While teacher assignment policies allow this sorting practice, the distribution of teachers will continue to be uneven, favoring more-advantaged students over less-advantaged students. These distributional effects make measuring the effects of particular teacher characteristics on student achievement difficult to estimate.

The empirical estimation of education production functions has been an active research endeavor for more than 30 years. From this work, we have some important

results: teaching experience in the first few years of teaching matters but not much thereafter; teacher ability (as measured by test scores or selectivity of educational institutions attended) makes a difference; advanced degrees, *per se*, are not significant but subject-matter preparation, at least in technical subjects such as math and science is probably important; and teacher certification is not likely critical at least as it currently operates in the US. Although still largely reliant on quasi- and non-experimental designs, improvements in data quality and econometric techniques have made recent results more convincing. Despite this, much remains unknown. Little progress has been made toward breaking down the largely unobserved aspects of teacher quality into more finely grained measurable characteristics, or observable behaviors that can be convincingly linked to student achievement and which can be reliably measured or even taught. Future research aiming for descriptive analysis from which more nuanced instruments of measurement could be developed and tested would further our understanding of teacher quality.

See also: Education Production Functions: Concepts; Teacher Supply; Teacher Training and Preparation in the United States.

Bibliography

- Boyd, D., Goldhaber, D., Lankford, H., and Wyckoff, J. (2007). The effect of certification and preparation on teacher quality. *Future of Children* 17(1), 45–68.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy* 1(2), 176–216.
- Clotfelter, C. T., Ladd, H. F., and Vigdor, J. (2006). Teacher–student matching and the assessment of teacher effectiveness. *Journal of Human Resources* 41(4), 778–820.
- Coleman, J., Campbell, E., Hobson, C., *et al.* (1966). *Equality of Educational Opportunity*. Washington, DC: Department of Health, Education, and Welfare.
- Darling-Hammond, L. and Youngs, P. (2002). Defining “highly qualified teachers”: What does “scientifically-based research” actually tell us? *Educational Researcher* 31(9), 13–25.
- Ehrenberg, R. G. and Brewer, D. J. (1994). Do school and teacher characteristics matter? Evidence from high school and beyond. *Economics of Education Review* 13(1), 1–17.
- Ehrenberg, R. G. and Brewer, D. J. (1995). Did teachers’ verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review* 14(1), 1–21.
- Eide, E., Goldhaber, D., and Brewer, D. (2004). The teacher labour market and teacher quality. *Oxford Review of Economic Policy* 20(20), 230–244.
- Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation* 28(2), 465–468.
- Ferguson, R. F. and Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In Ladd, H. (ed.) *Holding Schools Accountable*, pp 265–298. Washington, DC: Brookings Institution.
- Glazerman, S., Mayer, D., and Decker, P. (2006). Alternative routes to teaching: The impacts of teach for America on student achievement

- and other outcomes. *Journal of Policy Analysis and Management* **25**(1), 75–96.
- Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *The Journal of Human Resources* **42**(4), 765–794.
- Goldhaber, D. and Anthony, E. (2007). Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *Review of Economics and Statistics* **89**(1), 134–150.
- Goldhaber, D. and Brewer, D. (1997a). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources* **32**, 505–523.
- Goldhaber, D. and Brewer, D. (1997b). Evaluating the effect of teacher degree level on educational performance. In Fowler, W. J., Jr. (ed.) *Developments in School Finance, 1996*, pp 197–210. Washington, DC: National Center for Education Statistics, US Department of Education.
- Goldhaber, D. and Brewer, D. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis* **22**(2), 129–145.
- Goldhaber, D., Brewer, D., and Anderson, D. (1999). A three-way error components analysis of educational productivity. *Education Economics* **7**(3), 199–208.
- Greenwald, R., Hedges, L., and Laine, R. (1996). The effect of school resources on student achievement. *Review of Educational Research* **66**(3), 361–396.
- Hanushek, E. (2002). Publicly provided education. *Working Paper 8799*. National Bureau of Economic Research.
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature* **24**(3), 1141–1177.
- Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher* **18**(4), 45–51.
- Hanushek, E. A. (1994). Education production functions. In Husen, T. and Postlethwaite, N. T. (eds.) *The International Encyclopedia of Education*, 2nd edn., pp 1756–1772. Oxford: Pergamon.
- Hanushek, E. A., Kain, J. F., O'Brien, D. M., and Rivkin, S. G. (2005). The market for teacher quality. *Working Paper 11154*. National Bureau of Economic Research.
- Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources* **39**(2), 326–354.
- Jacob, B. (2007). The challenges of staffing urban schools with effective teachers. *Future of Children* **17**(1), 129–153.
- Lankford, H., Loeb, S., and Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis* **24**(1), 37–62.
- Monk, D. H. and Rice, J. K. (1994). Multi-level teacher resource effects on pupil achievement in secondary mathematics and science: The role of teacher subject matter preparation. In Ehemberg, R. G. (ed.) *Contemporary Policy Issues: Choices and Consequences in Education*, pp 29–58. Ithaca, NY: ILR Press.
- Murnane, R. (1975). *Impact of School Resources on the Learning of Inner City Children*. Cambridge, MA: Ballinger.
- Murnane, R. and Phillips, B. (1981). What do effective teachers of inner city children have in common? *Social Science Research* **10**(1), 83–100.
- Murnane, R. and Steele, J. (2007). What is the problem? The challenge of providing effective teachers for all children. *Future of Children* **17**(1), 15–43.
- Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (1998). Teachers, schools, and academic achievement. *Working Paper 6691*. National Bureau of Economic Research.
- Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2001). Teachers, schools, and academic achievement. Working Paper No. 6691, National Bureau of Economic Research.
- Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica* **73**(2), 417–458.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review* **94**(2), 247–252.
- Rowan, B., Chiang, F. S., and Miller, R. J. (1997). Using research on employees performance to study the effects of teachers on student achievement. *Sociology of Education* **70**(4), 256–284.
- Sanders, W. L. and Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. University of Tennessee Value-Added Research and Assessment Center, Knoxville, Tennessee.
- Strauss, R. P. and Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review* **5**(1), 41–48.
- Summer, A. and Wolfe, B. (1977). Do schools make a difference? *American Economic Review* **67**(4), 639–652.

Further Reading

- Darling-Hammond, L., Berry, B., and Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis* **23**(1), 57–77.
- Goldhaber, D. and Brewer, D. (2001). Evaluating the evidence on teacher certification: A rejoinder. *Educational Evaluation and Policy Analysis* **23**(1), 79–86.
- Hanushek, E. and Raymond, M. (2004). Does school accountability lead to improved student performance? *Working Paper 10591*. National Bureau of Economic Research.
- Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement data: Estimation using micro-data. *American Economics Review* **61**(2), 280–288.
- Hanushek, E. A. (1992). The tradeoff between child quantity and quality. *Journal of Political Economy* **100**, 84–117.
- Hanushek, E. A. (1997). Assessing the effects of school resources on student performances: An update. *Educational Evaluation and Policy Analysis* **19**(2), 141–164.
- Kane, T. J., Staiger, D. O., and Geppert, J. (2002). Randomly accountable. *Education Next* **2**(1), 57–62.
- Wayne, A. and Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research* **73**(1), 89–122.

The Economics of Class Size

D W Schanzenbach, University of Chicago, Chicago, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Class-size reduction is a politically popular but relatively expensive education reform. Understanding the causal relationship between class size and student achievement is critical for determining whether class-size reduction can be recommended as a policy to improve student outcomes. We begin with a review of the theory of why class size might matter, followed by a discussion of the empirical strategies for identifying the causal impact of class size on student achievement. Next, the empirical literature on class-size reduction is reviewed, focusing on studies using experimental and quasi-experimental techniques because these rely on the most credible strategies for identifying the true causal relationship between class size and student achievement.

Why Class Size Might Matter

For most readers, class size and academic outcomes are probably intuitively linked. Nonetheless, it may be helpful to formalize the idea somewhat. Lazear (2001) puts forth a useful theory of educational production. In it, reducing class size decreases the amount of time that the classroom is disrupted, increasing time devoted to productive tasks. A simple summary of the model is as follows: a child is behaving in class at a given moment with probability p , and misbehaving with probability $(1 - p)$. In the model, misbehavior is broadly defined, ranging from talking or fighting, to behaviors such as asking questions that slow down the class or monopolizing the teacher's time. When there are n children in the classroom, p^n is the probability that the entire class is behaving and learning is taking place (assuming that p is independent across children). Assuming a constant disruption rate, having fewer students in the class means that learning is taking place in a larger fraction of time.

In the model, the impact of reducing class size depends not only on the size of the class, but also on the behavior of the students in it. As a result, the Lazear theory predicts that class-size effects should be larger for classes with more poorly behaved students. For example, on average p may decrease with age, so the impact of class-size reduction might be smaller for high school students than elementary school students. The impact of class-size reduction – all else equal – is predicted to be larger for groups with lower propensities to behave.

Empirical Approaches to Studying the Impact of Class Size

Economists typically model the relationship between student achievement and class size for student i in school j as

$$Y_{ij} = aS_{ij} + bF_{ij} + \varepsilon_{ij} \quad [1]$$

where Y represents a measure of student achievement. S contains information on school-level inputs that impact achievement, such as class size, F contains family inputs, such as parental education, and ε is an error term. Both S and F measure inputs over the child's entire lifetime, and may contain inputs that are not observable to the econometrician. A negative coefficient on class size would suggest that student achievement declines as class size increases.

The problem with estimating eqn [1] (and similar versions of it) is that class size may be endogenous such that $E(\varepsilon | S, F) \neq 0$. For example, if students are assigned to small classes or better teachers in a compensatory manner – perhaps, because of low baseline test scores, or low levels of family inputs – but that information is not available to the researcher, the estimated impact of school resources will be biased. The most obvious such example is remedial or special-education courses, which tend to be small in size. Similarly, bias will result if parents who are more involved in their children's education are more likely to push for a smaller class or better teachers, and parental involvement is not measured in the dataset.

Due to these confounding factors, researchers have relied on strategies that use (plausibly) exogenous variation in class size in order to identify the causal impact of class size on student achievement. In other words, to identify the effect of class size, the variation in class size must come from factors that are more or less out of the control of decision makers such as parents and educators. The easiest example of this is where students are randomly assigned to classes of different size. The benefit of using a randomized experiment is that the treatment assignment is unrelated to any omitted characteristics. Such a design allows researchers to isolate the impact of the policy they are trying to test, without confounding factors such as parental pressure or compensating assignments. Thus, an experimental study typically offers more compelling evidence than a nonexperimental study, which simply observes the relationship between Y and S in the real world.

With a well-designed experimental assignment, a straightforward comparison of means by class type will

provide an unbiased estimate of the impact of class size on achievement. In the case of (an idealized version of) a class-size experiment in which students are randomized within schools, the equation to be estimated might be as follows:

$$Y_{ics} = \beta_0 + \beta_1 SMALL_{cs} + X_{ics}\gamma + \alpha_s + v_{ics} \quad [2]$$

where *SMALL* is an indicator variable for randomly assigned small-class treatment, and *c* indexes class *c* in school *s*. *X* is a vector of student-level characteristics. When treatments are randomized, student-level covariates are not related to class assignment and their inclusion should not change the estimated effect on class size, but should just contribute to the overall explanatory power of the model. A school-level fixed effect, α , is included, so that identification of small-class effects are identified off of within-school comparisons. Finally, the error term *v* contains class-level and individual-level components, reflecting random differences in teacher and student quality.

Nonexperimental Research

There have been volumes of research looking at the relationship between class size and student performance in nonexperimental settings. These are well summarized in a pair of influential meta-analyses by Hanushek (1986, 1997). In them, he argues that the lion's share of the economics of education literature finds no consistent relationship between class size and student performance. In the 277 estimates from 59 published studies included in the 1997 paper, only 15% showed a positive, statistically significant impact of class size on student outcomes. A re-analysis by Krueger (2003) raised questions about the validity of the meta-analysis, and argued that the literature largely supports a positive impact of reduced class size. Interested readers are encouraged to refer to Krueger (2001) and Hanushek (2003) for an overview of the debate.

The usefulness of the Hanushek meta-analyses is limited by the underlying quality of the studies included in them. Most – but certainly not all – of the underlying studies relied only on observational variation and did not have a research design that would allow the estimation of the causal impact of class size on student achievement. This is problematic for several reasons. For one, within-school variation in class size (in the absence of a true experiment) is seldom random. If, for example, there is compensatory assignment to smaller classes, then the coefficient on class size will not only pick up the true effect of being in a smaller class but also any correlated omitted variables such as special-education status or poor prior achievement as discussed above (see also Boozer and

Rouse, 2001). In addition, if there are only small differences in class size within school – for example, one classroom with 22 students and the other with 23 students – then one would need a large amount of data to precisely estimate the effect of such a small difference in class size. In general, it is inappropriate to base public policy on research that does not have a compelling identification strategy. As a result, much more weight should be placed on the experimental and quasi-experimental evidence outlined below.

Experimental Research

The most influential studies of class-size reduction are based on data from Project STAR, a large-scale randomized trial in the US state of Tennessee to test the impact of reducing class sizes in grades K-3. Mosteller (1995) described Project Steps to Achieving Resilience (STAR) as “one of the most important educational investigations ever carried out and illustrates the kind and magnitude of research needed in the field of education to strengthen schools.” In the experiment, students were randomly assigned within school to one of three treatment types: a small-size class (target of 13–17 students), a regular-size class (target of 22–25 students), or a regular-size class with a full-time teacher's aide. Importantly, teachers were also randomly assigned to class types. The experiment took place in 79 public schools across a variety of geographic locations (inner-city, suburban, and rural; predominantly low income, and middle class) for a single cohort of students in kindergarten through third grade in the years 1985–89. An eventual 11 600 students and 1330 teachers took part in the experiment. It is worth noting that the experiment made some students better off than they would have been, but of course did not otherwise increase class sizes beyond their normal range in the state at the time. Thus, no students were made worse off by being assigned to abnormally large classes.

In the ideal implementation of this experiment, students were to remain with the same randomly assigned class type from kindergarten through the end of third grade. In practice, though, there were several major sources of deviation from this model. Students who entered a participating school while the cohort was in first, second, or third grades were added to the experiment and randomly assigned to a class type. There were a substantial number of new entrants – 45% of eventual participants entered after kindergarten. A relatively large fraction of students also exited Project STAR schools (45% of overall participants), due to school moves, grade retention, or grade skipping, which also caused deviations from the original plan (see Krueger (1999) and Hanushek (1999) for further discussion). Fortunately, many of these exiting students are recaptured in the

follow-up analysis described below, which includes all students in Tennessee in grades 4–8, and a nationwide match of college-entrance-exam takers when the cohort is around twelfth grade. In addition, in response to parental concerns about fairness to students, all students in regular and regular-aide classes were re-randomized across the two treatment groups in first grade. This deviation is less problematic for evaluation of the program, because it involved new randomization. In general, studies have found no difference in performance between regular and regular-aide classes, so these two groups are sometimes combined as the control group.

Finally, a smaller number of students (about 10% of participants) were moved from one class type to another in a nonrandom manner. It has been reported that most of these moves were due to student misbehavior, and were not typically the result of parental requests for moves to small classes (Krueger, 1999). Of all transitions, 25% were into small (more desirable) classes. This weakness of the experiment can be addressed through use of an intent-to-treat setup – that is, to use the variation caused by initial randomly assigned class type instead of the actual (possibly nonrandom) class type attended.

In practice, the nonrandom transitions and new entrants described above complicate the approach somewhat. Due to nonrandom transitions after initial assignment, it would be inappropriate to use current-year class type; instead, initial class-type assignment (the intent-to-treat measure) is typically used in studies using Project STAR data. That is, all impacts are measured with regard to the class that students were assigned to, and not the class that they actually attended. The intent-to-treat measure used in this case likely understates the impact of small classes by up to 15% (Krueger, 1999). Nonetheless, the conservative intent-to-treat measure based on random assignment is typically considered preferable to models which measure the impact of actual class type attended in cases in which there is nonrandom movement between classes. A simple example may help illustrate this: if a child were moved from a regular class to a small class because his parents insisted on the move, it is also reasonable to assume that the parents are especially active in other aspects of the student's education. For example, they may monitor homework closely or provide other education-enhancing opportunities. Researchers cannot control for these characteristics because they do not have perfect measures of the home environment. In the ideal case in which class type is randomly assigned, these home-environment measures are uncorrelated with class type and their impacts are absorbed in the error term in eqn [2]. When the effect of actual (nonrandom) class attended is measured instead, some of the impacts of the active home environment also may be picked up because actual attendance may be correlated with this home-environment component of the error

term. Using the experimentally induced variation in this case means that not all students actually attend their assigned class type, and some students' test scores will count toward the regular-size class they were assigned to, even though they actually attended small classes. This approach circumvents the causation problem, but provides an understatement of the true impact. Krueger (1999) provides a more detailed discussion of this matter.

As described above, new entrants into the program were randomly assigned to class types. So, even though new entrants in first, second, and third grades on average are more disadvantaged than the kindergarten entrants, randomization allows us to compare new entrants in each grade to other new entrants in the same school across class types. In practice, then, the school-level fixed effect in eqn [2] is replaced with a fixed effect that combines school with a student's grade of entry (K, 1, 2, or 3) to the experiment, as this is the pool within which random assignment was determined (K representing kindergarten). In general, work on Project STAR has employed the following approach:

$$Y_{igs} = \beta_{0g} + \beta_{1g}SMALL_{is} + \beta_{2g}AIDE_{is} + \beta_{3g}X_{is} + \alpha_{rw} + \varepsilon_{igs} \quad [3]$$

Here g indexes the grade of the outcome measure. Both the *SMALL* and *AIDE* variables are measured as initial assignment, and not actual class attendance. The fixed effect varies by the randomization pool – school interacted with entry wave w . The coefficient on the control for classes with a teacher aide is sometimes omitted, as there appears to be no difference in outcomes between regular and regular-aide classes. As a result, the coefficient of interest measuring the small-class effect is similar whether or not aide classes are separately controlled. For precision, other student-level covariates such as gender, race, and free-lunch status are included in the vector X of control variables, but because of random assignment, including these controls does not change the magnitude of the small-class effect. The dependent variable is an outcome such as the mean math and reading score on the Stanford Achievement Test (SAT) for each grade, or whether a student took a college entrance exam.

Checks for Randomization

Due to the experimental design, impacts of reduced class size are straightforward to measure as the within-school (and entry wave) difference between class types, provided the randomization was done correctly. A compelling check of randomization is to examine a pretest to ensure that there are no measurable differences in the dependent variable between class types before the program begins. Unfortunately, no baseline test measure was collected in Project STAR. Another

way to investigate whether randomization was done properly is to compare student characteristics that are related to student achievement but cannot be manipulated in response to treatment, such as student race, gender, and age. If there are no systematic differences in observable characteristics across class types, this provides support that the randomization was done properly. A similar check should be done on observable teacher characteristics.

Table 1 presents estimates of differences in nonmutable characteristics across initial treatment assignment. This is similar to tables presented in Krueger (1999) and Krueger and Whitmore (2001). The estimating equation is similar to eqn [3] above, with student or teacher characteristic on the left-hand side and indicators for small-class assignment and school-by-entry-wave fixed effects. Standard errors are clustered at the school level. Each table entry represents a separate regression, and only the coefficient and standard error on small-class assignment are reported. Since none of the coefficients are large or statistically significant, this is evidence that the randomization was done correctly, at least with regard to observable characteristics. The single exception is that the teacher having a master's degree or higher is marginally significant ($p = 0.06$). This means that teachers with more education were slightly less likely to be assigned to small classes. The results below are virtually unchanged

if direct controls for teacher characteristics are included. Now that the randomization is validated, it is straightforward to turn to results of the experiment.

Achievement Results

Table 2 reports the impact of initial assignment to a small class on student test scores in grades K–3. Equation [3] is estimated, and each table entry reflects a separate regression. Test scores are normalized into z -scores based on the regular and regular-aide population. Average math and reading scores are reported in most cases, though if a student was missing a test score for one test but not both, the score for the non-missing test is used. The coefficient on the indicator variable for small class can be interpreted as the standard-deviation impact of the treatment. As many researchers have found (Word *et al.*, 1990; Krueger, 1999; Krueger and Whitmore, 2001), the table indicates that overall, students benefit about 0.15 standard deviations from assignment to a small class. When the results are disaggregated by race, it appears that black students benefited more from being assigned to a small class than the overall population, suggesting that reducing class size might be an effective strategy to reduce the black–white achievement gap. Krueger and Whitmore (2002) find that this result is largely driven by a larger treatment effect for all students regardless of race

Table 1 Testing whether covariates appear randomly assigned^a

	(1)	(2)	(3)	(4)
Panel A: Student characteristics	Female = 1 0.000 (0.012)	White = 1 –0.002 (0.006)	Free Lunch = 1 –0.014 (0.011)	Age in 1985 (in years) –0.012 (0.011)
Panel B: Teacher characteristics	Female = 1 –0.001 (0.006)	White = 1 –0.001 (0.018)	Master's degree or higher = 1 –0.051 (0.027)	Total experience (in years) –0.155 (0.470)

^aEach entry represents a separate regression. Only coefficients on initial assignment to small class are reported. Standard errors, in parentheses, are clustered by randomization pool. Other covariates include randomization-pool fixed effects.

Table 2 Small-class effects on test scores during the experiment^a

	(1)	(2)	(3)	(4)
Panel A: Overall	Kindergarten 0.187 (0.039)	Grade 1 0.189 (0.035)	Grade 2 0.141 (0.034)	Grade 3 0.152 (0.030)
Panel B: Black students only	Kindergarten 0.214 (0.074)	Grade 1 0.249 (0.063)	Grade 2 0.207 (0.054)	Grade 3 0.242 (0.060)
Panel C: Free-lunch students only	Kindergarten 0.188 (0.046)	Grade 1 0.195 (0.042)	Grade 2 0.174 (0.041)	Grade 3 0.174 (0.039)

^aEach entry represents a separate regression. Only coefficients on initial assignment to small class are reported. Standard errors are in parentheses, clustered by randomization pool. Other covariates include randomization-pool fixed effects and student demographic characteristics.

in predominantly black schools, suggesting that benefits from additional resources are higher in such schools. Benefits are also larger for students from low socioeconomic status families, measured by whether they receive free or reduced-price lunch.

In fourth grade, the class-size reduction experiment concluded and all students were returned to regular-sized classes. At the same time, the assessment test was changed from the SAT to the Comprehensive Test of Basic Skills (CTBS). Both tests are multiple-choice standardized tests that measure reading and math achievement, and are taken by students at the end of the school year. The CTBS results are scaled in the same manner as the SAT, in terms of standard deviation units. One important difference in the data is that all students in public schools statewide who had ever participated in Project STAR are included in the follow-up study, even if they had been retained a grade. It is estimated that 20% of students had been retained a grade by eighth grade, but this did not vary with initial class assignment. As a result, some students took the fourth-grade test in 1990, while others took it in later years or even took it more than once. In the analysis reported here, all scores from grade g – no matter what year a student was in that grade – are compared. In the event of multiple attempts at grade g 's test, the first available score is used. As in **Table 2**, all estimates are conditional on school-by-entry wave fixed effects and only the coefficient on small class is reported.

Results for grades 4–8 are reported in **Table 3**. Overall, there is a persistent positive impact of small-class assignment that is statistically significant (or borderline significant) through eighth grade, as has been found in previous studies (e.g., Krueger and Whitmore, 2001). The magnitude of the gain is one-third to one-half the size that was observed while the students were in the experimental classes. When the results are disaggregated, though, the impact appears to remain stronger with black and free-lunch students than with more advantaged students. There is also some evidence that nonacademic outcomes such as the rates of criminal behavior and teen pregnancy are improved (Krueger and Whitmore, 2002).

Another potential measure of student achievement is whether these students take the SAT or the American College Test (ACT) college-entrance exam, which can be used as an early proxy for college attendance. In order to measure this, Project STAR student data were matched to the national databases of college-entry test records, as described in Krueger and Whitmore (2001, 2002). To examine whether assignment to a small class influences the college-entrance exam test-taking rate, a binary variable indicating that a college-entrance exam was taken is the dependent variable in eqn [3]. The impact of small-class assignment on college test taking is included as the final column in **Table 3**. Overall, test-taking rates increase by about 2 percentage points. Black students were 5 percentage points more likely to take the SAT or ACT if they were assigned to a small rather than regular-size class. On average, 38% of black students assigned to small classes took at least one of the college-entrance exams, compared with 33% in regular classes. Such a striking difference in test-taking rates between the small and regular class students could occur by chance less than one in 10 000 tries. Krueger and Whitmore (2002) interpret the magnitude of these effects by reference to the resulting reduction in the black–white test-taking gap. In regular classes, the black–white gap in taking a college entrance exam was 12.9 percentage points, compared to 5.1 percentage points for students in small classes. Thus, assigning all students to a small class is estimated to reduce the black–white gap in the test-taking rate by an impressive 60%. After controlling for increased selection into the test among small-class students, the impact on test scores for blacks is 0.15 standard deviations – about the same as the test-score impact in third grade.

Additional Caveats

An important limitation to the experiment was nonrandom movement across class-type assignment, as well as sample attrition during the treatment phase. As discussed briefly above (and at much greater length in the referenced works, especially Krueger, 1999), nonrandom

Table 3 Small-class effects on long-term test scores^a

	Grade 4 (z-score) (1)	Grade 5 (z-score) (2)	Grade 6 (z-score) (3)	Grade 7 (z-score) (4)	Grade 8 (z-score) (5)	Took college entrance test (1 = yes) (6)
Panel A: Overall	0.035 (0.025)	0.048 (0.024)	0.060 (0.025)	0.040 (0.025)	0.036 (0.025)	0.024 (0.010)
Panel B: Black students only	0.078 (0.048)	0.080 (0.043)	0.105 (0.045)	0.066 (0.042)	0.063 (0.046)	0.050 (0.018)
Panel C: Free-lunch students only	0.029 (0.036)	0.058 (0.031)	0.080 (0.034)	0.067 (0.031)	0.064 (0.034)	0.031 (0.014)

^aEach entry represents a separate regression. Only coefficients on initial assignment to small class are reported. Standard errors are in parentheses, clustered by randomization pool. Other covariates include randomization pool-fixed effects and student demographic characteristics.

movement can be addressed through using initial class-type assignment and not the actual (nonrandom) class-type attended. The attrition problem is significantly addressed through the statewide and nationwide matches used for the follow-up analyses.

Another concern often raised about the results of randomized experiments generally is that the measured effect may be driven by Hawthorne effects and might not be generalized to nonexperimental settings. That is, people participating in the experiment might act differently than they normally do because they know they are being studied. Although one cannot directly test for Hawthorne effects, Krueger (1999) attempts to shed light on the issue by investigating differences in achievement using the variation across only regular-sized classes, as there is little reason to think that Hawthorne effects would cause some classes in the treatment group to behave differently relative to other treatment-group classes. Class size in regular-sized classes ranged from 16 to 30 students, but the bulk of the distribution was between 20 and 26 students. Whether or not school effects are controlled for, students in a regular class with slightly fewer members out-scored larger regular classes. The estimated magnitude of a one-student reduction in class size was consistent with the magnitude of the experimental results (which estimates the impact of a seven-student reduction).

Finally, another concern is whether the findings of this experiment may be generalized to other settings. Along many measures, Tennessee in the mid-1980s looks reasonably similar to other places that might be interested in implementing a class-size reduction policy, so it would be reasonable to expect similar effects as those in the experiment. On the other hand, the Tennessee sample has lower levels of education inputs than the United States overall at the time, as measured by spending per student and education level of teachers. If adding resources has a greater impact when the baseline levels are already low, this might mean that schools with higher levels of spending could experience a smaller impact of class-size reduction. In addition, in order to be eligible for the experiment, schools were required to have a large enough enrolment to support three classrooms per grade. As a result, Project STAR schools were about 30% larger than average schools in Tennessee or across the United States. If larger schools are somehow differently effective with additional resources, then the findings in Project STAR may not be generalizable to smaller school settings (see Schanzenbach (2007) for further discussion of these points).

Quasi-Experimental Research

As true randomized experiments are rare, researchers must also look for quasi-experimental approaches that allow isolation of the causal impact of class-size reduction.

One of the strengths of quasi-experimental approaches is that the participants are unaware that they are being studied, so Hawthorne effects are unlikely.

The most famous quasi-experimental approach to studying class-size reduction comes from Angrist and Lavy's (1999) use of a strict maximum class-size rule in Israel and a regression discontinuity (RD) approach. In Israel, maximum class size is dictated by Maimonides' rule, which specifies that no more than 40 students shall be in one class. As a result, if the school's total enrolment in a grade is 40 students or fewer, there will only be one classroom with a class size equal to the total enrolment. If the enrolment increases from 40 to 41 students, though, a second class must be added, and the average class size declines precipitously from 40 students to 20.5 students. Similarly, if enrolment increases from 80 to 81 students, a school must move from two to three classes and the average class size falls from 40 students to 27 students.

Using the local variation around the enrolment sizes that are multiples of 40 students, Angrist and Lavy isolate the causal impact of class-size reduction. They find strong improvements overall in both math and reading scores, of a magnitude that is consistent with Project STAR's experimental results. Like in Project STAR, they also find larger improvements among disadvantaged students.

Urquiola (2006) uses a similar RD approach in Bolivia and finds that a one standard-deviation reduction in class size (about eight students in his data) improves test-score performance by 0.2 to 0.3 standard deviations. Browning and Heinesen (2007) also find similar results on data from Denmark, even though average class size is much smaller in their study (20 pupils per classroom, compared to 31 students in Angrist and Lavy's Israeli data). Urquiola and Verhoogen (2008) provide a cautionary tale about possible endogenous responses of schools to class-size caps, and show that in Chilean data, these endogenous responses of schools lead to violations of the assumptions of the RD design. In addition, caveats about the external validity of these studies are required as was the case with Project STAR.

Another quasi-experimental approach comes from Hoxby's (2000) study of class size in the US state of Connecticut. In the study, Hoxby isolates the variation in enrolment that comes from random fluctuations in cohort sizes across adjacent years. That is, taking away any pre-existing trends in enrolment that might signify that a school district is waning or booming, the effect of class size is identified by variation in cohort size that reflects a temporary random shock in population size that may have been caused by an unusually small (or large) birth rate in a given year. Using this approach, Hoxby finds no positive effect of reduced class size, but has the statistical precision to rule out an effect as large as about one-fifth the size found in Project STAR. The discrepancy between these results and those of other well-identified experimental and quasi-experimental studies remains a puzzle.

Policy-Induced Variation

Another potentially promising approach to studying the effects of class-size reduction comes from sharp changes in policies regarding class size. The most famous recent example comes from the US state of California, where in 1996 a law was passed to give strong monetary incentives to schools to reduce class size in grades K–3 to 20 or fewer students. Unfortunately, from a research-design perspective, the take-up of the policy was nearly universal within a short period of time, so there was very little variation to exploit and evaluate its impact. In addition, test scores are only available starting in grade 4, so any evaluation of the policy was forced to use later test scores (instead of scores during the year that the reduced class size was experienced) as the outcome measure. Not surprisingly, the best evaluation of the policy found inconclusive results (Bohrnstedt and Stecher, 2002). It is unfortunate that this policy intervention, costing more than \$1 billion per year, did not yield useful information about the impact of class-size reduction on student outcomes.

Discussion

The bulk of the research using credible identification of the impacts of class-size reduction suggests that reducing class size will significantly improve test scores. In addition, the benefits appear to be larger for disadvantaged groups such as African American students and children from families with low socioeconomic status. The long-term follow-up of the Project STAR class-size experiment finds that the gains appear to persist even after students are returned to regular-sized classes. The prior research has been less able to credibly isolate potential nonlinear effects of class size, which is an important consideration for policymakers considering a potential class-size reduction.

An important question that policymakers must ask prior to embarking on class-size reduction is whether the projected benefits outweigh the costs. A cost-benefit study of Project STAR found that the overall benefits outweighed the costs (Krueger and Whitmore, 2001). The answer in other cases will depend on the school's situation. What is the current level of educational inputs? Are there many disadvantaged students? Do we want to put extra weight on questions of equity – for example, the potential for small classes to reduce the black–white achievement gap? Is there a ready supply of qualified individuals available to meet the increased demand for classroom teachers? And, of course, what is the next best use of the available funds?

See *also*: Education Production Functions: Concepts; Education Production Functions: Developed Country Evidence; Education Production Functions: Evidence

from Developing Countries; Empirical Research Methods in the Economics of Education.

Bibliography

- Angrist, J. D. and Pischke, J. V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* **114**(2), 533–575.
- Bohrnstedt, G. W. and Stecher, B. M. (2002). What we have learned about class size reduction in California. *CSR Research Consortium Capstone Report*.
- Boozer, M. and Rouse, C. (2001). Intraclass variation in class size: Patterns and implications. *Journal of Urban Economics* **50**(1), 163–189.
- Browning, M. and Heinesen, E. (2007). Class size, teacher hours and educational attainment. *Scandinavian Journal of Economics* **109**(2), 415–438.
- Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature* **24**, 1141–1177.
- Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis* **19**(2), 141–164.
- Hanushek, E. A. (1999). Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis* **21**, 154–164.
- Hanushek, E. A. (2003). The failure of inputs-based schooling policies. *The Economic Journal* **113**(485), F64–F98.
- Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics* **115**(4), 1239–1285.
- Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics* **114**(2), 497–532.
- Kureger, A. B. (2003). Economic considerations and class size. *The Economic Journal* **113**(485), F34–F63.
- Krueger, A. B. and Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project STAR. *Economic Journal* **111**, 1–28.
- Krueger, A. B. and Whitmore, D. (2002). Would smaller classes help close the black–white achievement gap? In Chubb, J. E. and Loveless, T. (eds.) *Bridging the Achievement Gap*, pp 11–46. Washington, DC: Brookings Institution Press.
- Lazear, E. P. (2001). Educational production. *Quarterly Journal of Economics* **116**, 777–803.
- Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *Future of Children* **5**(2), 113–127.
- Schanzenbach, D. W. (2007). What have researchers learned from project STAR? *Brookings Papers on Education Policy* **2007**, 205–228.
- Urquiola, M. (2006). Identifying class size effects in developing countries: Evidence from rural Bolivia. *Review of Economics and Statistics* **88**(1), 171–177.
- Urquiola, M. and Verhoogen, E. (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review* **99** (1), 179–215.
- Word, E., Johnston, J., Bain, H. P., et al. (1990). Student/teacher achievement ratio (STAR): Tennessee's K-3 class size study. *Final Summary Report 1985–1990*. Nashville, TN: Tennessee State Department of Education.

Further Reading

- Achilles, C. M., Nye, B. A., Zaharias, J. B., and DeWayne Fulton, B. (1993). The lasting benefits study (LBS) in grades 4 and 5 (1990–1991): A legacy from Tennessee's four-year (K-3) class-size study (1985–1989), project STAR. *Research Paper*, HEROS.
- Ehrenberg, R. G., Brewer, D. J., Gamoran, A., and Willms, J. D. (2001). Does class size matter? *Scientific American* **285**(5), 78–86.

The Economics of Early Childhood Interventions

M Nores, NIEER, Rutgers University, New Brunswick, NJ, USA

© 2010 Elsevier Ltd. All rights reserved.

The growing adoption of early childhood interventions derives from an increasing perception of the developmental importance of the early years for later educational, social, and economic outcomes and, more recently, the labor-market pressures on women. Early childhood education has not developed as fast as primary education did 50 years ago. Global pre-primary gross enrolment ratios are currently 37%. Enrolments in pre-primary education are under 20% in sub-Saharan Africa and the Arab States, between 20% and 40% in Asia, around 60% in Central and East Europe and Latin America, 72% in the Pacific, 78% in North America and West Europe, and 101% in the Caribbean (UNESCO, 2006). Although coverage has tripled in the last three decades, it remains extremely low in the developing world. While most growth has occurred in the public sector, child care continues to be provided mostly by private institutions. This is more the case in North America than in West Europe, consistent with their corresponding social spending traditions. Aid for early childhood policies has promoted early education interventions in many countries, but amounts to only 0.1–16% of aid to primary education.

The Economic Rationale

Why do governments provide early childhood education and child care? Why is early childhood education slightly increasing in the developing and first world country policy and aid agendas? From an economic perspective, there are two fundamental rationales for early childhood education and child care interventions (and education interventions in general): (1) market failures and (2) equity and redistribution (Gruber, 2007; Blau and Currie, 2005).

Market Failures

Market failures in education include (1) positive externalities (substantial positive effects associated with its consumption not directly observed by the consumer, and therefore not taken into account by him/her in the decision to consume); and (2) credit market failures (Gruber, 2007).

Externalities are benefits accrued to society (social benefits) when an individual chooses to consume a good or service, beyond those accrued to the individual (private benefits). Some private benefits, in particular earnings, are commonly known as private returns to education. Interventions that increase earnings produce private benefits.

Additional examples are our own health. On the other hand, social benefits are benefits that increase social welfare. Examples of social benefits are increased productivity of the labor force (contributing to economic growth and development), less welfare use of individuals with higher earnings, reduced crime, increased tax collection, and increased female productivity (by shifting women to the labor market). There are also a series of intangible social benefits that are harder to measure and relate to having a more educated society and less inequity (active participation and the quality of the democratic process).

Families do not take into account externalities when deciding about education (they are not willing to pay for them and these are not included in the price of schooling); therefore, they consume less than what is socially desirable (the social optimum). For example, a family knows that by investing in education his/her child will benefit through increased productivity and therefore earn higher wages as an adult. Therefore, they are willing to pay for education so long as the cost of educating their child is at least as much as the extra earnings the child will accrue in the future. However, society also benefits from an educated child and the family is not taking these benefits into account (only the additional earnings), therefore underinvesting in education.

Credit market failures relate to the inability of families to borrow to finance education (especially early childhood or basic education). When families cannot provide collateral (assets owned by the family that the bank can claim if the loan is not paid back), and because banks cannot ascertain whether individuals will earn enough in the future to pay for the loan (parents later on in their careers or the child when he/she grows up), they are not willing to supply the loan. This type of market failure can be addressed through loans to finance education which are absent for early childhood, unlike for higher education. Governments have chosen to provide early childhood services directly (e.g., preschool) or providing funds for it (e.g., vouchers and subsidized daycare).

Equity and Redistribution

The consumption of education, like the consumption of most goods, is dependent on the financial ability to pay for the service (i.e., the capacity of families to pay for education and the indirect costs of education, e.g., school uniforms, books, remedial classes, transportation, and parent association fees). Since higher family income renders

higher spending capacities, inequalities in income have the potential to translate into inequalities in educational attainment (and in the quality of education), and consequently, into differences in lifelong earning streams. Without government intervention, education has the potential to reproduce income inequalities. Through equality in access and educational opportunities, governments in democratic and nondemocratic societies have seen education as a way of redistributing income across generations. Public support for education based on distributional concerns is based on a belief that a children's access to education (and lifelong chances) should not be based on their parent's financial capital.

Similarly, whether or not most parents might be altruistic and prioritize expenditure on their children's education, the distributional argument also covers children whose parents might prioritize their consumption over their children's consumption (what is known as egotistic utility functions). Public provision of education and child care, as an alternative to money for child care or loans, ensures that all children are provided the same opportunities regardless of their parents' preferences.

Types of Early Childhood Interventions

Early childhood interventions have developed much later and at a lower speed than primary and secondary education, predominantly because governments and donor agencies have given a relatively lower priority to pre-primary education policies and expenditures. In the presence of market failures and a concern with redistribution, governments have three alternatives for intervention: regulation, funding, and provision. Although the reasons for government intervention described require a combination of regulation and funding, provision of early childhood education and child-care alternatives has prevailed in many countries.

Across countries, providers of early childhood education and care differ significantly. They support different aspects of childhood, from children's growth, development and learning (childhood education programs and preschool programs) to health, nutrition, hygiene, cognitive, social, physical, and emotional development. They take place in formal, informal, and nonformal settings, in arrangements ranging from center-based, to formal preschool education, to parent/community-based arrangements (UNESCO, 2006). In some countries, child care is provided at a parent's house serving children from the community and government subsidized (e.g., *Hogares Comunitarios* in Colombia); in these cases, the governments have chosen to increase access to early childhood care through increased funding and regulation predominantly.

In the last two decades, early childhood education has gained strength as a cost-effective way of reducing

disparities in educational outcomes and increasing schools' effectiveness. Because early environments (and disadvantages) define later outcomes, early childhood education has been promoted as a way of increasing educational attainment of the individual (productivity), as well as reducing social inefficiencies in the provision of education later on (reduce repetition, and the need for remedial or dropout prevention programs) (Carneiro and Heckman, 2003; Heckman and Masterov, 2007). Moreover, studies on cost effectiveness (measuring the cost of providing high-quality preschool education vs. the benefits produced by it in the form of higher productivity of the labor force, reduced welfare, reduced crime, and higher tax collection) have increased support for early childhood interventions.

In the international arena, as part of the millennium development goal of achieving universal primary education for all by 2015, early childhood education has been advocated to reduce disparities in primary educational attainment and increase attainment of the very poor. The overall argument on early childhood has moved from providing it as a basic human right to providing it because it is an effective way of reducing disparities in educational attainment that are the consequence of income inequalities. Because it leads to higher educational attainment, it increases individual welfare and reduces the probability of poverty in adulthood (the equity argument), while in parallel increasing efficiencies in the education system, reducing social costs, and increasing social welfare (the externality argument).

US and international evidence has shown that good-quality early childhood interventions have direct effects on cognitive and noncognitive development (Barnett, 2000; Blau and Currie, 2005; Boocock, 1995; Camilli, in press; Engle *et al.*, 2007; Heckman and Masterov, 2007). Important effects across different developmental dimensions have been present in several small controlled interventions, including random assignments, focused on disadvantaged populations. Among these are the Carolina Abecedarian Full-Day School Program, the Early Training Project, the Elmira Prenatal/Early Infancy Project, the Infant Health and Development Program, the Mauritius Preschool Study, the Milwaukee New Hope Project, and the Perry Preschool Project.

These studies are particularly valuable because randomization increases internal validity; namely, inferring causality between the program and the observed cognitive and noncognitive outcomes. On the other hand, as long as interventions remain small, external validity, or the degree to which conclusions can be generalized, remains small; that is, it is not clear whether the size of the results would remain the same when interventions are extended to a large-scale (i.e., national) level. This makes larger-scale interventions (e.g., evaluations of the Head Start program in the US, the Chicago Child-Parent Centers,

the Michigan School Readiness Program, the South Carolina Pre-K, the New York Pre-K, and the Effective Provision of Pre-School Education Project of the University of London, among others) substantiating the benefits of early childhood education, of fundamental importance (Blau and Currie, 2005; Heckman and Masterov, 2007).

Among the programs mentioned, the Perry Preschool Program is among the most-studied early childhood interventions, following a sample of 128 low-income children through age 40 (Barnett, 2000; Schweinhart, 2005). Children were randomly assigned to treatment and control groups, and the treatment group received half-day high-quality preschooling and weekly home visits for 2 years. The Abecedarian program consisted of a study following 111 children until the age of 21. Randomization was done at birth with the treatment group receiving center-based child care with an emphasis on language development, 8 h a day through the age of 5 (Heckman and Masterov, 2007). Similarly, the Chicago Child-Parent Center consisted on matching treatment individuals with similar children (of the same age, which would also be eligible for the program and of similar education and economic family background) rather than on random assignment of individuals. The program provided health and social services, free meals, and center-based half-day preschooling for 3- and 4-year-olds during the school year (Heckman and Masterov, 2007).

The different types of benefits observed from these interventions, as well as from international studies, are summarized below. These have been differentiated as cognitive (school progress and performance outcomes), noncognitive (social, behavioral, and labor-market outcomes), and indirect benefits to society.

Cognitive and Academic Outcomes

In the short and medium term, early childhood interventions, particularly those that focus on the developmental and cognitive aspects of childhood, are expected to impact participants' cognitive development, their school progress, and their educational attainment. These three types of indicators have been used as the basis for evaluating the success of most early childhood educational interventions.

Indicators, such as achievement tests and intelligence quotients (IQs), usually measure cognitive development. Several studies evidence positive effects of high-quality center-based care on children's cognitive growth. That is, children come into these programs extremely disadvantaged, and end up performing better than similar peers that were not granted access to the program. While in some cases IQ outcomes have been large, they have been short termed, sustained until school entry, and fading out through the first years of schooling except for the Abecedarian and Milwaukee programs. Not all types of

interventions have impacts on cognitive development as measured through IQ. Quality matters in early childhood (well-designed, intensive programs that include involvement with children and families). Moreover, children from impoverished homes benefit more from child care than children whose home environments are already supportive and stimulating (Engle *et al.*, 2007). The evidence from developing countries substantiates these findings (Boocock, 1995; Engle *et al.*, 2007; Schady, 2006).

Early childhood programs have also shown that they can have important medium- and long-term effects on school progress, that is, achievement, reducing repetition, reducing placement in remedial or special education, and increasing school outcomes (educational attainment). From random assignment studies, there is evidence of positive and lasting impacts of early child education on achievement (e.g., Milwaukee, Florida, Abecedarian, and Perry) (Barnett, 2000; Blau and Currie, 2005). Among the international research, there is evidence of increased achievement from various programs and evaluations; for example, carefully designed studies in Sweden (Boocock, 1995), a large program to construct preschool facilities in Argentina, a nutritional and early stimulation program in Jamaica (Schady, 2006), the Colombian *Hogares Comunitarios* program (Engle *et al.*, 2007), and the Cali Study (Boocock, 1995), integrating nutritional and child-care components.

Increased school readiness, achievement, and social development set the stage for improvements in school progress and educational attainment. That is, it reduces the chances that a child will repeat grades, need remedial education, be a special-needs child, and dropout. The evidence shows reduced grade retention, with the estimates for dropout rates across the Abecedarian, Chicago, and Perry interventions being quite similar (around 20–30%). There is also evidence of reduced special education rates or the number of years of special education (e.g., Abecedarian, Perry, Chicago, and Early Training Project). Center-based experiences in Nepal, Argentina, Burma, Mexico, and Colombia evidence improvements in enrolments, age of entry, retention, and performance (Engle *et al.*, 2007; Myers, 1992; Schady, 2006).

For the same reasons, since children come into school more prepared for it and progress through it more smoothly, they are also more likely to succeed in attaining a secondary education degree (and in consequence, more likely to transition into post-secondary education). In developing countries, transitioning into secondary education is an important concern. Only some few studies in the US and none in the international arena have followed children through high school and beyond. Abecedarian, Perry and Chicago, and the Florida studies, as well as evaluations on Head Start, have indicated effects on graduation rates of participants (Barnett, 2000; Blau and Currie, 2005).

Noncognitive Outcomes

Focusing on cognitive development indicators ignores the full array of social and economic noncognitive responses motivated by schools, families, and institutions. Short-term evidence on IQ or scores does not measure relevant nutritional, social, emotional, and overall developmental dimensions. Some examples of potential developmental outcomes of early childhood interventions are: improvements in the health and nutritional indicators of the child (whether because meal and nutrition programs are integrated into the intervention, or because of work being done with parents), effects on socialization (aggression and behavior indicators), which in the long term relate to delinquency and crime outcomes, and effects on motivation, self-discipline, and attentiveness that relate to long-term educational attainment.

Moreover, it is argued (Carneiro and Heckman, 2003) that while IQs might not be successfully affected by a variety of interventions, social and emotional skills are, leading to increased educational success, reduced criminal activity, integration of disadvantaged population into mainstream society, etc. A lifetime approach has addressed path dependency of circumstances and behaviors resulting from early education. Growing attention is being paid to how early cognitive development defines later attainments, life behaviors, opportunities, and experiences (Heckman and Masterov, 2007). When preschool programs deter children from transgressions, they may dissuade them from a life of crime. A similar argument goes for health-related behavior. While many programs are critically evaluated on cognitive development indicators, medium- and long-run behavioral effects translate into socioeconomic savings through life-long behavioral changes.

In the developing world, in the context of high child mortality rates, stunting, and serious child illnesses and hunger, health and nutrition take on added importance for child development (Barnett, 2000; Engle *et al.*, 2007). Health, nutritional, and developmental improvements have been evidenced across interventions in developing countries (e.g., Chile, Colombia, Egypt, Guinea, Jamaica, Kenya, Mexico, Mauritius, and Vietnam; Engle *et al.*, 2007; Myers, 1992; Raine *et al.*, 2003) and the US (Elmira Prenatal/Early Infancy Project), with improvements varying from improvements in iodine and iron deficiencies, to improvements in height/weight measures.

Substantial lifetime (long-term) effects are mediated through higher educational attainment. The effects of educational attainment on earnings are well established and long lasting. Higher educational attainment reflects cognitive advantages and enhanced noncognitive attributes, for example, self-discipline or diligence. Overall, higher attainment is associated with higher economic well-being over the long term, in particular, higher earnings and

employment status. The Perry Preschool Program followed children up to age 40, observing their earnings well into adulthood. Program males and females showed considerable higher earnings than the control group (Nores *et al.*, 2005). While increased earnings are a private benefit, these directly translate into higher amounts of tax payments (income taxes, through increased annual earnings, and sales taxes through increased consumption).

Reduced crime is another indication of social and emotional effects of early childhood. The Perry Preschool Project evidenced substantial lower rates of criminal activity for the treatment group. The Chicago Child-Parent Center (and an expansion program later on) showed reduced delinquency and crime. Evaluations of the Florida and Colorado Head Start (for females only), the Panel Study of Income Dynamics (PSID) Head Start study (comparing Head Start to non-Head Start participants between the ages of 18 and 31) and the Mauritius Preschool Study, support the findings from the Perry and Chicago random studies. Many of the benefits from reduced crime can be summarized into victims (some of these, such as pain, being intangible), policing, criminal justice system, and incarceration costs; all of which are different types of social benefits. Private benefits from reduced crime may vary from increased quality of life, to intergenerational benefits (discussed below), and a series of intangible benefits (such as freedom).

As individuals attain higher levels of education, subsequent increased earnings, and evidence reduced criminal behavior, the aggregation of such behavioral and economic changes are expected to reduce welfare use. The Perry Preschool Program indicates lower welfare reliance of the treatment group (the PSID Head Start evaluation also looked at welfare but found no difference across groups).

Other indicators of social and behavioral changes relate to improvements in the quality of life, health-related behavioral changes, and teenage pregnancy. Teenage pregnancy effects might be the consequences of postponing childbirth through increased educational attainment or through health-related behavioral changes. For youth and adults, the Abecedarian and Perry studies provided indication of health-related behavior in terms of drug usage and tobacco. These two programs also generated increases in the age when the participants had their first child (Schweinhart *et al.*, 2005; Blau and Currie, 2005).

Indirect Effects

Besides benefits reaped directly from the child's increased educational attainment and lifetime behavioral responses, early childhood interventions impact a child's family in two additional, important, yet indirect, ways: (1) by

freeing up women's time and (2) through the intergenerational links across earnings and across education.

Female Labor Supply

As an increasing number of mothers have entered the labor market, especially with global trends in urbanization and industrialization, the question has arisen on how to take care of their children. Even though child protection and the enhancement of child development and school readiness has been at the center of government interventions, increased female labor participation has increased awareness on the issues of early child care and education. Although female employment in developing countries is also increasing, females in these settings tend to concentrate on self-employment and unpaid family work.

Early childhood interventions may offer opportunities for maternal participation in the labor market as well as increased school participation of older siblings (usually females) freed from child-care responsibilities. Child care and labor-market choices are competing uses of the time of the mother and child-care prices affect labor-market choices, in particular, for married women who evidence a higher elasticity (option to decide depending on the price) in their labor supply with respect to child-care prices than single women who have no choice but to work, depend on welfare, or a relative. Mothers in non-wage jobs (e.g., agricultural work in their household) tend to combine work and family responsibilities more often than women in wage jobs (urban areas and developed countries). Moreover, in developing countries young daughters frequently act as providers of free child care, releasing mothers for free child care.

Evidence from the expansion of child-care policies and subsidies in the US and developing countries show a substantial relation between women's decision to work, in particular, single and low-earning mothers and child-care prices (e.g., see Gelbach (2002) and Bainbridge *et al.* (2003) for the US; Berlinski and Galiani (2007) for Argentina, and Lokshin *et al.* (2000) for Kenya). This decision to work is accompanied with increases in household earnings and reductions in welfare dependency.

Intergenerational Effects

Estimation of the intergenerational benefits of early childhood interventions has been done for the Abecedarian and Perry studies (Barnett and Belfield, 2006). However, these tend to be projections based on: (1) studies on the intergenerational links across earnings and across education, and (2) the literature on the determinants of educational attainment. Studies on the intergenerational links across earnings and education have estimated these links to be strong, although far from the unit (Bowles *et al.*, 2005). The literature on the determinants of educational attainment

has emphatically shown that parental education, earnings, and employment status (socioeconomic status) are the strongest determinant of a child's cognitive development and success. The combination of these two sets of literature leads to project at least educational and earning effects on subsequent generations. On the other hand, the effect of reduced crime and welfare behavior benefits across generations has proved harder to establish, although there is some evidence that health-related and social behaviors from children relate to parental behavior.

Conclusions

The economic rationale for early childhood interventions is similar to that of providing primary and secondary education. Early childhood interventions are argued based on the existence of externalities, family credit constraints, and equity. Expected benefits from good-quality early childhood interventions, on the basis of the evidence, vary from short-term impacts on cognitive measures such as IQ, to higher achievement, school progress, reduced repetition, reduced special education placement, higher educational attainment, higher earnings, more taxes paid, reduced crime, reduced welfare reliance, reduced teenage pregnancy, and improved health-related behavior. Moreover, early childhood interventions also free mothers (and in some contexts, older female siblings) for labor-market opportunities, thereby increasing female productivity.

Because many of the externalities of early childhood education have been quantified, it has gained support as a policy alternative to interventions later on. Early childhood education has gained strength as a cost-effective way of reducing disparities in educational outcomes and increasing schools' effectiveness. In parallel, while developed and developing contexts vary in terms of impact and needs (e.g., nutrition), the equity argument has gained strength as an outcome of early childhood provision over time. Early childhood interventions increase educational progress and attainment, increasing individual welfare and reducing the probability of disadvantages in adulthood.

See also: Cost-Benefit Analysis and Cost-Effectiveness Analysis; Education Production Functions: Concepts.

Bibliography

- Bainbridge, J., Meyers, M. K., and Waldfogel, J. (2003). Child care policy reform and the employment of single mothers. *Social Science Quarterly* **84**(4), 771–791.
- Barnett, W. S. (2000). Economics of early childhood intervention. In Shonkoff, J. P. and Meisels, S. J. (eds.) *Handbook of Early Childhood Intervention*, 2nd edn., pp 589–610. Cambridge: Cambridge University Press.
- Barnett, W. S. and Belfield, C. R. (2006). Early childhood development and social mobility. *Future of Children* **16**(2), 73–98.

- Berlinski, S. and Galiani, S. (2007). The effect of a large expansion of pre-primary school facilities on preschool attendance and maternal employment. *Labour Economics* **14**(3), 665–680.
- Blau, D. M. and Currie, J. (2005). Preschool, day care, and afterschool care: Who's minding the kids? In Welch, F. and Hanushek, E. A. (eds.) *The Handbook of Economics of Education*, pp 1163–1267. New York: North-Holland.
- Boocock, S. S. (1995). Early childhood programs in other nations: Goals and outcomes. *Future of Children* **5**(3), 94–114.
- Bowles, S., Gintis, H., and Osborne Groves, M. (2005). *Unequal Chances. Family Background and Economic Success*. Princeton, NJ: Princeton University Press.
- Camilli, G., Vargas, S., Ryan, S., and Barnett, W. S. (in press). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*.
- Cameiro, P. and Heckman, J. (2003). Human capital policy. *NBER Working Paper No. W9495*. Cambridge: National Bureau of Economic Research.
- Engle, P. L., Black, M. M., Behrman, J. R., et al. (2007). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world. *Lancet* **369**(9557), 229–242.
- Gelbach, J. B. (2002). Public schooling for young children and maternal labor supply. *American Economic Review* **92**(1), 307–322.
- Gruber, J. (2007). *Public Finance and Public Policy*, 2nd edn. New York: Worth Publishers.
- Heckman, J. J. and Masterov, D. V. (2007). The productivity argument for investing in young children. *NBER Working Paper No. W13016*. Cambridge: National Bureau of Economic Research.
- Lokshin, M. M., Glinskaya, E., and Garcia, M. (2000). The effect of early childhood development programs on women's labor force participation and older children's schooling in Kenya. *World Bank Policy Research Working Paper No. 2376*. Washington, DC: World Bank.
- Myers, R. (1992). *The Twelve Who Survive: Strengthening Programmes of Early Childhood Development in the Third World*. London: Routledge.
- Nores, M., Belfield, C. R., Barnett, W. S., and Schweinhart, L. (2005). Updating the economic impacts of the high/scope Perry preschool program. *Educational Evaluation and Policy Analysis* **27**(3), 245.
- Raine, A., Kjetil, M., Liu, J., Venables, P., and Mednick, S. A. (2003). Effects of environmental enrichment at ages 3–5 years on schizotypal personality and antisocial behavior at ages 17 and 23 years. *American Journal of Psychiatry* **160**(9), 1627–1635.
- Schady, N. R. (2006). Early childhood development in Latin America and the Caribbean. *Economia* **6**(2), 185–225.
- Schweinhart, L. J., Montie, J., Xiang, Z., et al. (2005). *Lifetime Effects: The High/Scope Perry Preschool Study through Age 40*. Ypsilanti, MI: High/Scope Press.
- UNESCO (2006). *EFA Global Monitoring Report. Early Childhood Care and Education*. Paris: UNESCO.
- Berlinski, S., Galiani, S., and Manacorda, M. (2007). Giving children a better start: Preschool attendance and school-age profiles. *World Bank Policy Research Working Paper No. 4240*. Washington, DC: The World Bank.
- Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., and Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian project. *Applied Developmental Science* **6**(1), 42–57.
- Connelly, R. and Kimmel, J. (2003). The effect of child care costs on the employment and welfare reciprocity of single mothers. *Southern Economic Journal* **69**(3), 498–520.
- Han, W. and Waldfogel, J. (2001). Child care costs and women's employment: A comparison of single and married mothers with pre-school-aged children. *Social Science Quarterly* **82**(3), 552–568.
- Kimmel, J. (2006). Child care, female employment and economic growth. *Community Development: Journal of the Community Development Society* **37**(2), 71–85.
- Lokshin, M. (2004). Household childcare choices and women's work behavior in Russia. *Journal of Human Resources* **39**(4), 1094–1115.
- Reynolds, A. J., Temple, J. A., Robertson, D., and Mann, E. (2001). Long-term effects of an early childhood intervention on educational attainment and juvenile arrest. *Journal of the American Medical Association* **285**(18), 2339–2346.

Relevant Websites

- <http://www.ecdgroup.com> – Early Childhood Counts, The Consultative Group on Early Childhood Care and Development. (Inter-agency consortium linked to regional networks, advocacy and knowledge generation and dissemination at the international level.).
- <http://www.dcsf.gov.uk/research/programmeofresearch/projectinformation.cfm?projectid=13144&resultspage=1> – Effective Provision of Pre-School Education Project. The Effective Provision of Pre-School Education (EPPE) Project is the first major longitudinal study by the University of London, on the effectiveness of the early years of education.
- <http://www.fpg.unc.edu> – Frank Porter Graham Child Development Institute.
- <http://www.highscope.org> – High Scope Educational Research Foundation.
- <http://www.nber.org> – National Bureau of Economic Research.
- <http://nieer.org> – National Institute for Early Education Research.
- <http://www.futureofchildren.org> – The Future of Children.
- <http://secc.rti.org> – The National Institute of Child Health and Human Development Study of Early Childhood and Youth Development.
- <http://www.unicef.org> – United Nations Children's Fund.
- <http://portal.unesco.org> – United Nations Education, Scientific and Cultural Organization (UNESCO). (UNESCO provides the leading source on the Education for All (EFA) Global Monitoring actions, including the in early childhood interventions as means for increasing primary and secondary educational attainment and equity.).
- <http://www.waisman.wisc.edu> – Waisman Center, Chicago Child-Parent Center Program. (The Chicago Longitudinal Study is a federally funded investigation of the effects of the Child-Parent Center (CPC) Program and is in its 16th year of operation.).

Further Reading

- Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *Future of Children* **5**(3), 25–50.

ECONOMICS OF EDUCATION – TEACHING LABOR MARKETS

Contents

An Overview of Teacher Labor Markets
Compensating Differentials in Teacher Labor Markets
Economic Approaches to Teacher Recruitment and Retention
Teacher Incentives
Teacher Supply
Teacher Training and Preparation in the United States
Teachers in Developing Countries
The Economics of Teachers' Unions in the United States

An Overview of Teacher Labor Markets

H Lankford, University at Albany, Albany, NY, USA

J Wyckoff, University of Virginia, Charlottesville, VA, USA

© 2010 Elsevier Ltd. All rights reserved.

The market for K-12 teachers is important both for its sheer size and because teachers can have large impacts on the life outcomes of students. The 3.6 million elementary and secondary school teachers in the United States make up 9.2% of all college-educated workers 25–64 years old (National Center for Education Statistics (2006); Digest of Education Statistics, (2006); US Census Bureau Educational Attainment in the United States (2005)). Research confirms the conventional wisdom that teachers can substantially contribute to improved student achievement (Sanders and Rivers, 1996; Aaronson *et al.*, 2003; Rockoff, 2004; Rivkin *et al.*, 2005; Kane *et al.*, in press). Yet, achievement levels often fall short of that expected by the public and policymakers. Despite improvements over the last 20 years, it remains common that US schools with largely poor, nonwhite students have fewer than half their students reach basic performance and fewer than 20% reach proficiency (e.g., on the 2007 National Assessment of Educational Progress (NAEP) eighth-grade mathematics achievement tests, 53% of all African Americans performed below the basic level and only 11% reached the proficient level), with subsequent high school graduation rates also below 50%. (Many states do not yet employ cohort-based graduation rates and thus estimates are notably inaccurate. New York State did employ a cohort-based approach to estimate that in 2002, fewer than half of all African Americans in ninth grade graduated from high school by 2006.) Thus, it is natural for policymakers and researchers to look to improvements in teaching as an important component in addressing long-standing

concerns about student-achievement gaps by socioeconomic status as well as the increasingly poor performance of US students in comparison to international peers.

The qualifications of college graduates who enter teaching in the US have declined over the last 40 years (Bacolod, 2007; Corcoran *et al.*, 2004) and the least qualified of those entering teaching typically teach poor, nonwhite, and low-performing students – the very students whose educational success is most dependent upon their teachers (Betts *et al.*, 2000; Lankford *et al.*, 2002; Peske and Haycock 2006; Clotfelter *et al.*, 2007c).

This is an extraordinarily exciting period for research exploring teacher labor markets. Policymakers appear intent on improving the quality of classroom instruction and are experimenting with a range of policies intended to improve teacher recruitment, preparation, retention, support, and evaluation – all in an effort to improve student outcomes. (As is increasingly common in the literature, we differentiate between teacher qualifications and teacher quality. Teacher qualifications are the measures of ability, knowledge, and skills of teachers such as test scores, certification status, and education level. Teacher quality is a measure of a teacher's ability to improve outcomes for students, such as value-added scores or principal evaluations.) In response, researchers interested in informing policy are employing a variety of methods and data in efforts to identify how policies can be leveraged to improve the educational achievement of students. An increasing number of states and districts have longitudinal data on teachers and students that provide a rich

understanding of the individuals who enter teaching, the nature of their preparation, their career choices, and the performance of their students on achievement tests. Designs that exploit natural experiments, regression discontinuities, or instrumental variables are increasingly commonplace. Finally, randomized control trials that evaluate treatments that isolate specific aspects of teacher labor markets have become more common as a result of research programs at the US Department of Education Institute of Educational Sciences (Angrist, 2004). While much work remains to be done, the methodological ante for applied research on teacher labor markets has risen, thereby providing policy-makers increasingly useful causal evidence.

The field is now better positioned to build a knowledge base that identifies what policies work in improving student outcomes. We highlight three important, unresolved research questions relating to teacher labor markets. We focus on these questions because addressing them will inform the most pressing educational-policy questions; these questions incorporate many more specific teacher-labor-market issues. (Many teacher-labor-market research questions fall under these three broad areas. To name but two: What is the role that teacher-retirement benefits play in defining teacher labor demand and supply? How have the No Child Left Behind Act (NCLB) and state accountability policies affected teacher demand and supply?) Besides, they illustrate some important conceptual and empirical issues:

1. Recruitment: What are the most efficient ways to attract individuals, who will be effective teachers, to the profession?
2. Preparation: Of those who enter teaching, what teacher preparation and professional-development policies best increase teachers' capacity to improve student outcomes?
3. Retention: Given the substantial concern over teacher retention, what policies ensure that effective teachers are retained, while ineffective teachers are encouraged to pursue other alternatives?

To answer these questions, researchers must develop causal understandings of a number of more traditional empirical questions pertaining to teacher labor markets, such as: how to define and measure teacher quality; how teacher quality varies with observable attributes of teachers; the elasticity of teacher quality with respect to wages and other working conditions; how transfer and quit decisions by teachers respond to changes in wages and other working conditions; and the relative marginal productivity of differing teacher-human-capital investments. The answers to these and other questions likely are context specific – the causal relationships differ across, and even within, teacher labor markets. For example, recruitment, preparation, and retention strategies to improve student outcomes may well differ between

schools having predominately low- and high-performing students and between schools in urban and rural settings. Ignoring such heterogeneity can result in misguided policy recommendations.

Much has been written on issues related to teacher demand and supply, with several recent papers providing excellent summaries of various aspects of this literature (e.g., Hanushek and Rivkin, 2006; Dolton, 2006; Murnane and Steele, 2007; Goldhaber, 2008). In addition, the articles following this overview provide excellent summaries of specific features of teacher labor markets. Our intent in this article is to provide a conceptual and empirical framework within which to situate the research questions outlined above. We begin by describing some features of teacher labor markets that both differ from many other labor markets and create challenges when analyzing the factors that contribute to poor student performance.

Teacher Labor Markets

Teachers are important to the education process because of their ability to contribute to student learning, both directly and indirectly. Equation [1] is a stylized version of the education-production function model that underlies most research linking teachers to student outcomes.

$$A_{it}=f(C_{it}, F_{it}, T_{it}, P_{it}, S_{it}) \quad [1]$$

A_{it} reflects educational outcomes for student i at time t , and C_{it} , F_{it} , T_{it} , P_{it} , and S_{it} represent all current and prior values of relevant attributes of the student (C) and his or her family (F), teachers (T), classroom, school, and neighborhood peers (P) and other inputs available in the school (S). (For a detailed discussion of models of student achievement and their estimation see Todd and Wolpin (2003)). Research estimating variants of eqn [1] finds that differences in teacher quality have substantial effects on student achievement. For example, Sanders and Rivers (1996) estimate that differences in teacher quality can result in a 50 percentile improvement in measures of student achievement and that these improvements are additive and cumulative over subsequent teachers. Aaronson *et al.*, (2003) find that a two-standard-deviation increase in teacher quality improves student achievement by 25–45% of a grade equivalent. Kane *et al.*, in press) estimate that the difference in effectiveness between the top and bottom quartile of teachers results in a third of a standard deviation difference in student test-score gains. However, teacher effects are estimated with substantial uncertainty, implying true values may differ from these estimates (McCaffrey *et al.*, 2003).

What determines the teacher quality a particular student receives? The allocation of teacher quality can be

viewed as a function of demand, supply, and market-specific institutional constraints.

$$T^d = g(A^*, w, r, C, X) \quad [2]$$

$$T^s = b(w, C, Y, Z) \quad [3]$$

Demand for teacher quality, T^d , is a function of the desired outcomes for students (A^*), teacher wages (w), prices of other inputs (r), the exogenous attributes of students (C), and restrictions placed on employers regarding the hiring of teachers (X), for example, regulations and a single-salary schedule. Teacher supply, T^s , is also a function of the wage (w), the characteristics of students (C), other working conditions (Y), and requirements associated with eligibility to teach (Z), for example, certification requirements.

Differences in teacher quality reflect the decisions of prospective teachers and school officials that determine initial job matches and subsequent decisions that affect job quits, transfers, and terminations. Most research on teacher labor markets has used models developed for the private sector. However, markets for public-school teachers typically differ in three ways from assumptions usually made in modeling private-sector labor markets – sorting of employees (teachers) with respect to clients (students), wage contracts determined almost entirely by experience and education, and, like other public-sector-hiring authorities, have limited incentives to operate efficiently. In addition, as in private professions, public-school teachers in the US typically are subject to rather stringent credentialing requirements. These aspects of the labor market for teachers likely have important implications for deciding who enters teaching, where they teach, and their career paths, all of which affect the overall quality of teaching and its distribution.

The allocation of poorly qualified teachers to schools with concentrations of poor, low-income, and low-achieving students ultimately reflects teacher preferences and the incentives they face. Policies have the potential of creating incentives to encourage more able individuals to become teachers and to allocate more effective teachers to low-performing schools. Unfortunately, many common policies and practices fail to account for basic realities of teacher supply and inefficiently allocate resources as a result.

The common practice of allocating students to schools by neighborhood-attendance zones, combined with residential sorting of families by socioeconomic status, race, and preferences for education, often leads to schools segregated by income, race, and achievement of students (Clotfelter, 2004; Lankford and Wyckoff, 2006). Research suggests that teachers prefer working in particular types of school environments as proxied by student-body attributes (Ingersoll, 2001; Lankford *et al.*, 2002; Scafidi *et al.*, 2003; Hanushek *et al.*, 2004; Boyd *et al.*, 2005) and other

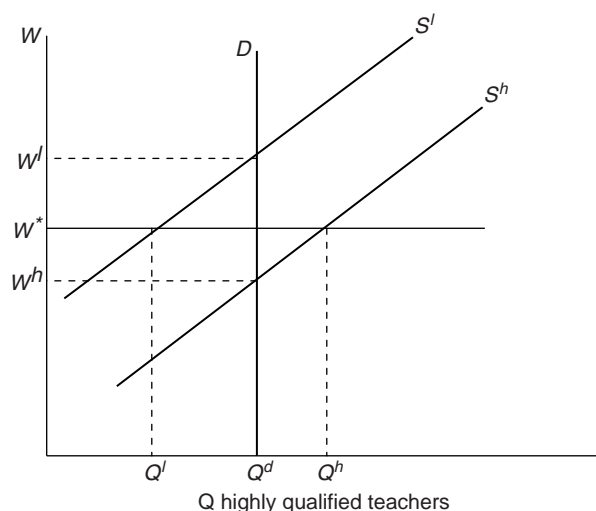


Figure 1 The demand and supply of highly qualified teachers.

working conditions (Johnson *et al.*, 2004; Luekens *et al.*, 2004; Buckley *et al.*, 2004; Carroll *et al.*, 2004; Loeb *et al.*, 2005), as well as salary (Baugh and Stone, 1982; Brewer, 1996; Clotfelter *et al.*, 2006a; Dolton, 1990; Dolton and van der Klaaw, 1999; Dolton and Makepeace, 1993; Hanushek and Pace, 1995; Manski, 1987; Mont and Rees, 1996; Murnane *et al.*, 1989; Stinebrickner, 1999, 2001a, 2001b; Theobald, 1990; Theobald and Gritz, 1996). Not surprisingly, teachers seek out preferable working conditions – both the attributes of students and other school working conditions that often are correlated with student composition.

However, the single-salary contract, typical in most school districts, rewards teachers primarily for education and school-district teaching experience, regardless of the particular school in the district where one teaches. Thus, differential salaries within districts rarely are used to provide compensating differentials for factors related to the attributes of students affecting teacher supply to specific schools (although several districts are experimenting with various forms of incentive compensation, including performance incentives and differential pay for difficult-to-staff subjects and schools). This results in some schools and subjects having chronic staffing problems and low-performing students being disproportionately taught by the least-qualified teachers.

This outcome is easily depicted in a simple teacher-labor-market diagram (Figure 1). Assume that based on constant enrolments and norms about class size, the demand for high-quality teachers is perfectly inelastic at Q^d . Due to accountability or other external forces, assume the demand for high-quality teachers is the same for schools within a district regardless of their student composition; schools with disproportionately low-achieving students have the same demand for teacher quality as schools with mostly high-achieving students. In addition, assume that schools prefer to hire high-quality teachers

but hire low-quality teachers if insufficient numbers of high-quality teachers are available at the prevailing wage. Finally, assume that at any given wage, rate high-quality teachers prefer to work with higher- rather than lower-achieving students. S^l and S^h in **Figure 1** represent the supply of high-quality teachers in schools having lower- and higher-scoring students, respectively. If teacher salaries could freely adjust to take into account this difference in supply, schools with low-achieving students could attract sufficient numbers of high-quality teachers by paying the salary W^l , which is higher than the salary W^h needed to attract sufficient numbers of high-quality teachers in schools with high-achieving students. Note that teachers view the schools as being equally attractive even though salaries differ between the schools. ($W^l - W^h$ is the compensating wage differential that results in high-quality teachers being indifferent between teaching in the two school types. A compensating differential for teaching low-achieving students also could be made with combinations of other endogenous school variables, such as better leadership or better facilities, rather than salaries. In fact, adjusting other working conditions may be a more cost-efficient means to induce sufficient supply of high-quality teachers to schools with low-achieving students.) Such an outcome is precluded by the single-salary schedule being the same across schools within districts. The typical case is depicted by a district-wide wage W^* , where there is excess supply of high-quality teachers in schools with high-achieving students ($Q^h - Q^d$) and a shortage of high-quality teachers in schools with low-achieving students ($Q^d - Q^l$). The shortage of high-quality teachers is especially apparent in schools with low-achieving students in subject areas such as math, science, or special education, where teacher supply is more limited.

The differential qualifications of teachers across schools and subjects that result from the interaction of wage inflexibility and teacher preferences for working conditions is made worse by four other practices found in some districts: late hiring that results in a diminished pool of highly qualified applicants (Levin and Quinn, 2003), hiring decisions that do not effectively distinguish the qualifications of applicants (Ballou, 1996; Ballou and Podgursky, 1997), seniority-based transfer policies (Levin and Quinn, 2003), and personnel allocations between schools based on full-time equivalent (FTE) lines rather than monetary budgets (Roza and Hill, 2004).

In sum, the analysis of teacher labor markets makes clear that some policies and practices of school districts systematically lead to low-achieving students disproportionately being taught by the least-qualified teachers. At one level, policies to address these issues may seem apparent – if wage inflexibility with respect to working conditions leads to differential qualifications across schools, alter compensation policies to allow between-school salary differences to compensate for differences in working

conditions. This, of course, begs the question as to the size of the appropriate compensating salary (or other working condition) differential needed to offset such differences. Going beyond issues of teacher compensation, what other policies are effective in improving the quality of teachers, either through their selection or preparation, broadly defined? Much of the research on teacher labor markets is focused on various aspects of these questions. Before discussing this effort, we briefly explore some of the methodological issues that researchers must confront when attempting to empirically estimate these relationships.

Methodological Issues

Researchers are interested in identifying how a wide variety of demand and supply factors affect the composition and distribution of the teacher workforce as well as how teachers, teacher attributes, and various teacher-workforce policies affect student outcomes, as generically described in eqn [1]. Identifying such causal effects requires researchers to adequately address a number of conceptual and methodological issues. We highlight a few of the more important ones.

As discussed above, who teaches and where they teach depend upon a wide range of decisions: individuals deciding whether to attain teacher certification; teacher candidates deciding whether and where to seek employment; hiring authorities deciding whom among their applicants to hire; those employed deciding whether to change employment, either by leaving the profession or transferring within or between districts; and school officials deciding whether to terminate or council out untenured teachers. Researchers attempting to sort out how various factors affect such choices and, in turn, the matching of teachers to jobs, face methodological and statistical challenges that result from the types of data that are available. Even with good longitudinal data, researchers typically only know who teaches, where, and for how long, not, for example, who applied for which jobs, which of the sets of all applicants received offers, or any information about job offers that were turned down. Thus, when a person teaches in one school rather than some other, we typically do not know whether the teacher preferred the first alternative or actually preferred the second school but was not competitive for job openings there. Thus, inferring the attributes of teachers and jobs that are valued is made more difficult. Very modest progress has been made in developing an empirical approach for using matched employer–employee (e.g., school–teacher) data to isolate demand and supply factors (Boyd *et al.*, 2007b). However, findings from the vast majority of empirical studies exploring how various factors affect the selection of individuals into teaching, the sorting of teachers to jobs, as

well as teacher transfers and quits must be interpreted as reduced-form results reflecting various combinations of demand and supply factors.

Isolating the causal effects of teacher attributes on student outcomes also places considerable demands on data and empirical estimation. Detailed theories do not exist as to how the knowledge, skills, and other attributes of teachers influence student achievement within the context of a classroom where students interact with each other and a school environment that reflects a variety of other school inputs and constraints, all interacting in complex ways. As a result, we have only an *ad hoc* understanding of what variables to include and how to specify empirical models. In general, the number and nature of the variables required to fully specify eqn [1] is staggering.

Identification of the causal effects of educational interventions relies on having a counterfactual that credibly controls for all other factors that could influence the outcome, while varying the treatment (Barrow and Rouse, 2005; Angrist, 2004). In education research on teachers, several factors complicate establishing a believable counterfactual. Among the most important is the selection or sorting of teachers and students to schools, as described above. For example, suppose policymakers want to know whether gains in math achievement will be higher if students are taught by teachers certified to teach math rather than certified in another area. If students taught by teachers certified in math have relatively higher math scores, compared to students whose teachers are not math certified, one might be tempted to conclude that being certified to teach math contributes to the higher student achievement. The causal relationship, however, may operate in the opposite direction; the more qualified teachers may be in schools where students perform well in math because they prefer to teach motivated students, who are higher performing, other things equal, and are able to do so because employers want to staff their courses with in-field certified teachers. Without research strategies that preclude or account for such selection, empirical results will overstate the contributions of being math certified. Selection could be taken into account if all the attributes of teachers and students that are both associated with the selection and the outcome of interest were observed. Unfortunately, there are a variety of variables typically not observed (e.g., teacher and student motivation) that are likely associated with the matching of teachers to students, teacher retention, and student achievement. Similar selection issues complicate many other aspects of teacher labor markets.

Research relating to teacher labor markets has made remarkable progress in addressing selection issues over the last 10 years, due in large part to the increasing availability of panel data and research designs that attempt to negate the influence of selection. When well designed and

implemented correctly, randomized control trials (RCTs) with teachers randomly assigned to students avoid the selection issue altogether. There is now a lively discussion concerning the circumstances under which RCTs can yield useful policy insights regarding teacher-workforce issues (see, e.g., Cook, 2007; Murnane and Nelson, 2007; Raudenbush, 2008). To date, there have been few RCTs focused on issues relevant to teacher labor markets. A notable exception is an RCT examining the effectiveness of Teach for America teachers relative to teachers entering teaching through other routes (Glazerman *et al.*, 2006). Several RCTs now are underway exploring various aspects of teacher labor markets, including a pay-for-performance experiment in Nashville public schools.

However, the vast majority of research analyzing teacher labor markets employs observational data where issues of selection are common. Researchers have drawn upon a variety of circumstances and methods in their efforts to identify credible counterfactuals. Panel data estimation, propensity-score matching, regression discontinuity, and instrumental variable designs all employ quantitative methods to attempt to identify the effects of labor-market treatments while controlling for selection.

In recent years, there has been a dramatic increase in the use of panel data methods, largely resulting from the availability of state and district data that link individual teachers and students over multiple years. Researchers employing such data and methods have examined several aspects of the research questions outlined above. A partial list includes:

1. Supply and quality of teachers (Hanushek *et al.*, 2005a, 2005b; Clotfelter *et al.*, 2006b; Boyd *et al.*, 2007b);
2. Teacher preparation and professional development (Boyd *et al.*, 2006, 2008; Darling Hammond *et al.*, 2005; Harris and Sass, 2007a; Kane *et al.*, in press);
3. Qualifications and credentialing of teachers (Boyd *et al.*, forthcoming; Clotfelter *et al.*, 2007a, 2007b; Goldhaber and Brewer, 1997, 2000; Goldhaber and Anthony, forthcoming; Goldhaber, 2007; Harris and Sass, 2007b; Kane *et al.*, in press); and
4. Teacher retention (Hanushek *et al.*, 2004; Hanushek *et al.*, 2005b; Boyd *et al.*, 2005; Boyd *et al.*, 2007a; Goldhaber *et al.*, 2007; Feng, 2008).

Others have employed regression-discontinuity designs (Jacob and Lefgren, 2004; Lavy, 2004), instrumental-variable estimation (Ehrenberg and Brewer, 1996; Dee, 2004), and propensity-score matching (Lavy, 2004) to examine various aspects of teacher labor markets.

In addition to complications associated with selection, data quality issues limit empirical research analyzing teacher labor markets and student outcomes. Three points are particularly relevant. First, rarely are outcome measures other than student test scores available, and increasingly the tests employed are the high-stakes tests states

have implemented to comply with the requirements of NCLB. There are well-documented limitations and perverse incentives associated with these tests that raise questions about their use as instruments for research on teacher policies (Koretz, 2002; Koretz and Hamilton, 2006; Jacob and Levitt, 2003). Nonetheless, they are useful measures precisely because they have been identified as policy outcomes, are widely available, and, when employed in value-added models, have been shown to distinguish effectiveness between the bottom and top quartiles in ways that are consistent with principals' perceptions of teachers' value added (Jacob and Lefgren, 2005; Harris and Sass, 2008). Additionally, although information is limited, it appears that the high-stakes assessments in some states rate well on validity and reliability (Ferrara and DeMauro, 2006). At the same time, reasonable questions have been raised about the technical adequacy (i.e., psychometric properties) of assessments in many states for value-added modeling (Ballou, 2008; Briggs *et al.*, 2008).

Second, administrative databases, designed by school administrators, almost always fail to include some of the variables of conceptual and empirical interest to researchers. For example, databases at best contain only weak measures reflecting teachers' skills and knowledge, even though these attributes are believed to be important in differentiating good and bad teaching. Weak measures of the underlying variables of interest significantly limit identification strategies.

Third, identifying the effects of various teacher attributes is difficult when the available data have relatively little variation in the relevant variables. For example, if schools are able to attract a given quality teacher because of the wage and nonwage benefits offered, then one would expect that the teachers in that school would be of similar quality. Teachers within a school who lack one attribute may have others that compensate. This realization should encourage caution when researchers conclude that attributes of teachers are unimportant when they employ observational data. Note also that the range of knowledge and skills of teachers observed is substantially reduced by a variety of factors, including state certification. For example, most states require that teachers of core subjects have a subject-matter major in those areas or otherwise demonstrate sufficient content knowledge. If this requirement is effective, we may not observe a sizable effect of content knowledge not because content knowledge is unimportant, but because the range of observed content knowledge is restricted to be above the point where its variation makes a difference. Thus, there is a premium on measuring the differential quality of knowledge and skills that the teacher brings to the classroom as opposed to some cruder measure of content knowledge such as whether she has taken courses in the field.

Looking Ahead

Over the last 10 years, research on teacher labor markets has dramatically increased in quantity and quality. Much of this growth is explained by the increased focus on teachers and teaching brought about by various accountability efforts and the availability of much better data that has been spawned by these efforts. Despite these advances, research on teacher labor markets has resulted in limited insights for policy to date. The pressing policy questions we identified above have received only partial answers:

1. Some research suggests that both selection and preparation of teachers can make a difference, but there is good evidence that the variation within any groups characterized by a preparation pathway or teachers having particular attributes substantially exceeds the differences between groups. Thus, there is little guidance regarding whether or how to structure entry requirements. Similarly, we know that prospective teachers value compensation and other working conditions, but know relatively little about how to structure policies to most effectively attract high-quality teachers to all schools.
2. Even though most believe that knowledge and skills can improve teacher effectiveness, there is only limited empirical evidence on which to base the design of preservice or in-service preparation or how this should vary across the range of teaching environments that teachers encounter.
3. Teacher retention policies suffer from similar ambiguities. Many good teachers leave the profession early in their careers, as do many weak teachers. Research has provided limited guidance to school administrators regarding how to structure performance evaluation and retention policies to insure a high-quality workforce across all schools.
4. We are only beginning to understand the effects that dramatic increases in accountability are having on who enters teaching, what preparation they receive, how long they persist, and most importantly how teachers may change the nature of their work in relation to student outcomes.

That said, the field is vibrant and in the midst of rapid change, with interesting debate regarding many of these issues. Some of the questions are methodological: Do the available measures of student achievement capture the actual contributions of teachers? Do the value-added models employed sufficiently account for the interactions between the attributes of teachers and students so as to capture the important contributions of teachers? Do these models provide a credible identification strategy in the face of formidable issues of teacher and student sorting? Also, how robust are results across alternative model specification and alternative estimation strategies? Other

questions revolve around data – developing variables, not typically collected in administrative data that have detail sufficient to help guide policy, for example, which working conditions are most influential in retaining effective teachers. Finally, some questions go beyond teacher-labor-market research *per se*, and involve the nexus of research and policy. Much remains to be learned about how researchers, educators, and policymakers can work together to improve educational outcomes for children.

See also: Compensating Differentials in Teacher Labor Markets; Economic Approaches to Teacher Recruitment and Retention; Empirical Research Methods in the Economics of Education; Teacher Incentives; Teacher Supply; Teacher Training and Preparation in the United States; Teachers in Developing Countries; The Economics of Teachers' Unions in the United States.

Bibliography

- Aaronson, D., Barrow, L., and Sander, W. (2003). Teachers and student achievement in the Chicago public schools. *Federal Reserve Bank of Chicago WP-2002-28*.
- Angrist, J. (2004). American education research changes track. *Oxford Review of Economic Policy* 20(2), 198–212.
- Bacolod, M. (2007). Do alternative opportunities matter? The role of female labor markets in the decline of teacher supply and teacher quality. *Review of Economics and Statistics* 89(4), 737–751.
- Ballou, D. (1996). Do public schools hire the best applicants? *Quarterly Journal of Economics* 111(1), 97–134.
- Ballou, D. (2008). Test scaling and value-added measurement. *Working Paper*.
- Ballou, D. and Podgursky, M. (1997). *Teacher Pay and Teacher Quality*. Kalamazoo, MI: E. Upjohn Institute for Employment Research.
- Barrow, L. and Rouse, C. (2005). Causality, causality, causality: The view of educational inputs and outputs from economics. *Working Paper 2005–2015*. Chicago, IL: Federal Reserve Bank of Chicago.
- Baugh, W. H. and Stone, J. A. (1982). Mobility and wage equilibration in the educator labor market. *Economics of Education Review* 2(3), 253–274.
- Betts, J., Rueben, K., and Danenberg, K. (2000). *Equal Resources, Equal Outcomes? The Distribution of School Resources and Student Achievement in California*. San Francisco, CA: Public Policy Institute of California.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy* 1(2), 176–216.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J. (2007a). Who leaves? Teacher attrition and student achievement. *Working Paper*.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J. (2008). Teacher preparation and student achievement. *Working Paper*.
- Boyd, D., Lankford, H., Loeb, S., Rockoff, J., and Wyckoff, S. (forthcoming). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. *Journal of Public Policy Analysis and Management*.
- Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J. (2005a). The draw of home: How teachers' preferences for proximity disadvantage urban schools. *Journal of Policy Analysis and Management* 24(1), 113–132.
- Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J. (2005b). Explaining the short careers of high-achieving teachers in schools with low-performing students. *American Economic Association Proceedings* 95(2), 166–171.
- Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J. (2007b). Analyzing the determinants of the matching of public school teachers to jobs: Estimating compensating differentials in imperfect labor markets. *National Bureau of Economic Research Working Paper 9878*, University at Albany.
- Brewer, D. J. (1996). Career paths and quit decisions: Evidence from teaching. *Journal of Labor Economics* 14(2), 313–339.
- Briggs, D., Weeks, J., and Wiley, E. (2008). Vertical scaling in value-added models for student learning. *Working Paper*.
- Buckley, J., Schneider, M., and Shang, Y. (2004). *LAUSD School Facilities and Academic Performance*. Washington, DC: National Clearinghouse for Educational Facilities.
- Clotfelter, C. (2004). *After Brown: The Rise and Retreat of School Desegregation*. Princeton, NJ: Princeton University Press.
- Clotfelter, C., Glennie, E., Ladd, H., and Vigdor, J. (2006a). Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. *NBER Working Paper #12285*.
- Clotfelter, C., Ladd, H., and Vigdor, J. (2006b). Teacher–student matching and the assessment of teacher effectiveness. *Journal of Human Resources* 41(4), 778–820.
- Clotfelter, C., Ladd, H., and Vigdor, J. (2007a). How and why do teacher credentials matter for student achievement. *Working Paper*. Duke University.
- Clotfelter, C., Ladd, H., and Vigdor, J. (2007b). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Working Paper*. Duke University.
- Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., and Wheeler, J. (2007c). High poverty schools and the distribution of teachers and principals. *North Carolina Law Review* 85, 1345–1379.
- Cook, T. (2007). Randomized experiments in education: Assessing the objections to doing them. *Economics of Innovation and New Technology* 16(5), 331–355.
- Corcoran, S., Evans, W. N., and Schwab, R. (2004). Changing labor market opportunities for women and the quality of teachers, 1957–2000. *American Economic Review* 94(2), 230–235.
- Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., and Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, teach for America, and teacher effectiveness. *Education Policy Analysis Archives* 13(42), 1–47.
- Dee, T. (2004). Teachers, race and student achievement in a randomized experiment. *Review of Economics and Statistics* 86(1), 195–210.
- Dolton, P. J. (1990). The economics of UK teacher supply: The graduate's decision. *Economic Journal* 100(400), 91–104.
- Dolton, P. (2006). Teacher supply. In Hanushek, E. and Welch, F. (eds.) *Handbook of the Economics of Education*, vol. 2, pp 1052–1080. Amsterdam: Elsevier.
- Dolton, P. J. and Makepeace, G. H. (1993). Female labour force participation and the choice of occupation. *European Economic Review* 37, 1393–1411.
- Dolton, P. J. and van der Klaaw, W. (1999). The turnover of teachers: A competing risks explanation. *Review of Economics and Statistics* 81(3), 543–552.
- Ehrenberg, R. and Brewer, D. (1996). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review* 14(1), 1–21.
- Feng, L. (2008). Hire today, gone tomorrow: The determinants of attrition among public school teachers. *Working Paper*.
- Ferrara, S. and DeMauro, G. (2006). Standardized assessment of individual achievement in K-12. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 531–578. Westport, CT: American Council on Education/Praeger.
- Glazerman, S., Meyer, D., and Decker, P. (2006). Alternative routes to teaching: The impacts of teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management* 25(1), 75–96.
- Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness? *Journal of Human Resources* 42(4), 765–794.
- Goldhaber, D. (2008). Teachers matter, but effective teacher quality policies are elusive. In Ladd, H. and Fiske, T. (eds.) *Handbook of Research in Education Finance and Policy*, pp 146–165. New York: Routledge.

- Goldhaber D. D., and Anthony E. (forthcoming). Can teacher quality be effectively assessed? National Board certification as a signal of effective teaching. *Review of Economics and Statistics*.
- Goldhaber, D., Gross, B., and Player, D. (2007). Are public schools really losing their "best"? Assessing the career transitions of teachers and their implications for the quality of the teacher workforce. *Working Paper*.
- Goldhaber, D. D. and Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *Journal of Human Resources* **32**(3), 505–523.
- Goldhaber, D. D. and Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis* **22**(2), 129–145.
- Hanushek, E., Kain, J., and Rivkin, S. (2004). Why public schools lose teachers. *Journal of Human Resources* **39**(2), 326–354.
- Hanushek, E., Kain, J., and Rivkin, S. (2005a). Teachers, schools, and academic achievement. *Econometrica* **73**(2), 417–458.
- Hanushek, E. A., Kain, J. F., O'Brien, D. M., and Rivkin, S. G. (2005b). The market for teacher quality. *National Bureau of Economic Research Working Paper* 11154.
- Hanushek, E. A. and Pace, R. R. (1995). Who chooses to teach (and why)? *Economics of Education Review* **14**(2), 101–117.
- Hanushek, E. A. and Rivkin, S. G. (2006). Teacher quality. In Hanushek, E. and Welch, F. (eds.) *Handbook of the Economics of Education*, vol. 2, pp 1052–1080. Amsterdam: Elsevier.
- Harris, D. and Sass, T. (2007a). Teacher training and teacher productivity. *Working Paper*. Florida State University.
- Harris, D. and Sass, T. (2007b). The effects of NBPTS-certified teachers on student achievement. *Working Paper*. Florida State University.
- Harris, D. and Sass, T. (2008). What makes for a good teacher and who can tell? *Working Paper*.
- Jacob, B. and Lefgren, L. (2004). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources* **39**(1), 50–79.
- Jacob, B. and Lefgren, L. (2005). Principals as agents: Subjective performance measurement in education. *Kennedy School of Government Working Paper No. RWP05-040*.
- Jacob, B. and Levitt, S. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics* **118**(3), 843–877.
- Johnson, S. M. The Project on the Next Generation of Teachers (2004). *Finders and Keepers: Helping New Teachers Survive and Thrive in Our Schools*. San Francisco, CA: Jossey-Bass.
- Kane T., Rockoff J., and Staiger D. (in press). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*.
- Koretz, D. (2002). Limitations in the use of achievement tests as measures of educators' productivity. In Hanushek, E., Heckman, J., and Neal, D. (eds.) *Special Issue: Designing Incentives to Promote Human Capital*. *Journal of Human Resources* **37**(4, Fall), 752–777.
- Koretz, D. and Hamilton, L. S. (2006). Testing for accountability in K-12. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 531–578. Westport, CT: American Council on Education/Praeger.
- Lankford, H., Loeb, S., and Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis* **24**(1), 38–62.
- Lankford, H. and Wyckoff, J. (2006). The effect of school choice and residential location on racial and economic segregation of students. In Gronberg, T. and Jansen, D. (eds.) *Advances in Applied Microeconomics, Volume 14: Improving School Accountability: Check-Ups or Choice*, pp 185–240. Amsterdam: Elsevier.
- Lavy, V. (2004). Paying for performance; the effect of individual financial incentives on teachers' productivity and students' scholastic outcomes. *Working Paper* 10622.
- Levin, J. and Quinn, M. (2003). *How We Keep High Quality Teachers out of Urban Classrooms*. New York: The New Teacher Project.
- Loeb, S., Darling-Hammond, L., and Luczak, J. (2005). How teaching conditions predict teacher turnover in California schools. *Peabody Journal of Education* **80**(3), 44–70.
- Manski, C. F. (1987). Academic ability, earnings, and the decision to become a teacher: Evidence from the national longitudinal study of the high school class of 1972. In Wise, D. A. (ed.) *Public Sector Payrolls*, pp 291–312. Chicago, IL: University of Chicago Press.
- McCaffrey, D., Lockwood, J., Koretz, D., and Hamilton, L. (2003). *Evaluating Value-Added Models for Teacher Accountability*. Santa Monica, CA: RAND Corporation.
- Mont, D. and Rees, D. I. (1996). The influence of classroom characteristics on high school teacher turnover. *Economic Inquiry* **34**, 152–167.
- Murnane, R. and Nelson, R. (2007). Improving the performance of the education sector: The valuable, challenging, and limited role of random assignment evaluations. *Economics of Innovation and New Technology* **16**(5), 307–322.
- Murnane, R. J., Singer, J. D., and Willett, J. B. (1989). The influences of salaries and "opportunity costs" on teachers' career choices: Evidence from North Carolina. *Harvard Educational Review* **59**(3), 325–346.
- Murnane, R. and Steele, J. (2007). What is the problem? The challenge of providing effective teachers for all children. In Loeb, S., Rouse, C., and Shorris, A. (eds.) *Excellence in the Classroom: Policies to Improve the Teacher Workforce*, **17**(1), pp 15–44.
- National Center for Education Statistics (2006). *Digest of Education Statistics, 2006*. http://nces.ed.gov/programs/digest/d06/tables/dt06_001.asp (accessed July 2009).
- Peske, H. and Haycock, K. (2006). *Teaching Inequality*. Washington, DC: The Education Trust.
- Raudenbush, S. (2008). Advancing educational policy by advancing research on instruction. *American Education Research Journal* **45**(1), 206–230.
- Rivkin, S., Hanushek, E., and Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica* **73**(2), 417–458.
- Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review* **94**(2), 247–252.
- Roza, M. and Hill, P. (2004). How within district spending inequities help some districts fail. In Ravitch, D. (ed.) *Brookings Papers on Education Policy*, pp 201–218. Washington, DC: Brookings Institution.
- Sanders, W. L. and Rivers, J. C. (1996). *Research Project Report: Cumulative and Residual Effects of Teachers on Future Student Academic Achievement*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.
- Scafidi, B., Sjoquist, D., and Stinebrickner, T. (2003). Do teachers really leave for higher paying jobs in alternative occupations? *Working Paper*. Georgia State University.
- Stinebrickner, T. (1999). Estimation of a duration model in the presence of missing data. *Review of Economics and Statistics* **81**(3), 529–542.
- Stinebrickner, T. (2001a). A dynamic model of teacher labor supply. *Journal of Labor Economics* **19**(1), 196–230.
- Stinebrickner, T. (2001b). Compensation policies and teacher decisions. *International Economic Review* **42**(3), 751–779.
- Theobald, N. D. (1990). An examination of the influences of personal, professional, and school district characteristics on public school teacher retention. *Economics of Education Review* **9**(3), 241–250.
- Theobald, N. D. and Gritz, R. M. (1996). The effects of school district spending priorities on the exit paths of beginning teachers leaving the district. *Economics of Education Review* **15**(1), 11–22.
- Todd, P. and Wolpin, K. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal* **113**, F3–F33.
- US Census Bureau Educational Attainment in the United States (2005). Detailed Tables. <http://www.census.gov/population/www/socdemo/education/cps2005.html> (accessed July 2009).

Further Reading

New York State Education Department (2006). *New Graduation Results Released for High Schools Statewide*. <http://www.oms.nysed.gov/press/Gradrates2006releaseFINAL.htm> (accessed July 2009).

Relevant Websites

<http://www.caldercenter.org> – CALDER: Longitudinal Data in Education Research.

<http://nces.ed.gov> – National Center for Education Statistics, NAEP Data Explorer.

<http://www.performanceincentives.org> – National Center on Performance Incentives.

<http://www.teacherpolicyresearch.org> – Teacher Research Policy.

<http://www.wcer.wisc.edu> – WCER, National Conference on Value-Added Modeling.

Compensating Differentials in Teacher Labor Markets

J Chambers, American Institutes for Research, Palo Alto, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cost factors (uncontrollable factors) – The factors outside of local control that reflect variations in the local cost of living.

Comparable wage index (CWI) – An index that is intended to capture regional compensating differences in the wages for workers in comparable occupations or industries.

Compensating differential – A wage difference that compensates for the relative advantage or disadvantage of a job or location in the eyes of the worker.

Controllable factors – The factors that are within the power of district decision makers to influence or change.

Geographic cost of education index – An index that takes into account geographical differences in the salaries necessary to attract and retain a comparable workforce of instructional staff across local school districts.

Hedonic wage model – A model based on the theory that salaries will reflect not only compensation paid for specific human capital characteristics (e.g., the specific skills and training an individual brings to the job), but also other characteristics of the job that influence the attractiveness of living and working in a particular geographic locale.

Price of living index – An index that compares the prices of a market basket of consumer goods over time or across geographic locations.

there was any employment evidently either more or less advantageous than the rest, so many people would crowd into it in the one case, and so many would desert it in the other, that its advantages would soon return to the level of other employments. (Smith, 1976: 111)

The advantages to which Smith refers are the combinations of compensation and working conditions associated with any given job or location. In other words, employers and employees would make adjustments in their employment decisions to reflect the relative advantages and disadvantages of different occupations or locations. These resulting adjustments would move the market toward wage levels that would equate the relative desirability of different occupations or locations, and this would be accomplished through what we refer to in the literature as compensating differentials. A compensating differential is, in effect, a wage difference that compensates for the relative advantage or disadvantage in the eyes of the worker.

Early Efforts

An early attempt to estimate compensating differentials in teacher markets by Antos and Rosen (1975) used econometric models to assess discrimination in the market for teachers. The purpose of this early paper was to estimate the compensating differentials associated with, for example, general neighborhood hazard indicators such as student mobility, nonwhite history, and student poverty. In other words, how much does it cost to recruit and employ comparable teachers to work in more challenging school environments? The assumption is that schools with a more challenging environment would be associated with a positive wage differential to compensate for these greater challenges. Antos and Rosen present econometric results which show that white males and black females accept lower wages to teach in schools with lower poverty levels.

With the movement to improve equity and access to school funding that arose out of the war on poverty of the 1960s, other additional work by economists was directed at a broader perspective on compensating differentials in teacher labor markets (Brazer, 1974; Grubb and Hyman, 1975; Chambers, 1978a; Kenny *et al.*, 1975; Wendling, 1979). The goal of this new generation of work on compensating differentials focused on developing indices that could be used to adjust state aid to local school districts to

Background

Over the past 30 years, the need for understanding and estimating the compensating differentials within teacher labor markets has generated numerous empirical and theoretical studies. The concept of compensating differentials dates back to Adam Smith's well-known volume *An Inquiry into the Nature and Causes of the Wealth of Nations*. Smith states the following:

The whole of the advantages and disadvantages of the different employments of labour and stock must, in the same neighborhood, be either perfectly equal or continually tending to equality. If in the same neighborhood,

equalize the purchasing power of the educational dollar. In the 1960s and early 1970s, few, if any, states made any attempt to adjust state aid for geographic variations in the costs of schooling inputs.

The state of Florida implemented the Florida Price Level Index (FPLI) in 1973 as one approach to develop a geographic cost adjustment (Wood *et al.*, 2001). The FPLI is essentially a price of living index that compares the prices of a market basket of consumer goods across the 63 counties in Florida. The FPLI was applied on the assumption that geographic differences in the price of living drive education cost differences.

Purchasing Power of the Educational Dollar and Hedonic Wage Model

The new generation of studies of teacher labor markets brought the recognition that while geographic differences in the price of living and cost of education are related, the geographic cost of education index, even if narrowed to focus on teacher labor markets, extends well beyond accounting for price of living differences. These studies focused on teachers because they are the largest single category of schooling inputs into educational production. Teachers' salaries and benefits account for about half of total school district budgets. (See table 161 in the 'Digest of education statistics' which shows that instruction expenditures account for about 50–60% of total current spending in the US. Far and away the majority of these direct instructional expenditures are accounted for by teachers' salaries and benefits, the remainder going to paraprofessionals, books, and supplies and materials.) Moreover, the same factors that affect teacher labor markets were also found to affect labor markets for other professional or paraprofessional school personnel.

These studies of teacher labor markets focused on answering the following question: How much more or less does it cost to recruit and employ comparable teachers in different geographic jurisdictions (e.g., school districts) across a state?

The authors of these studies of teacher labor markets began by identifying all (or as many as possible) of the factors that affect variations in wages and then attempted to run simulations to answer the question set out above.

These studies employed the hedonic wage model as the conceptual framework to support the empirical work on teacher labor markets. Goldhaber *et al.* (2007) describe this model as follows:

The theory behind the hedonic model is that salaries will reflect not only compensation paid for specific human capital characteristics (e.g., the specific skills and training an individual brings to the job), but also other characteristics of the job that influence the attractiveness of living

and working in a particular geographic locale. For instance, it should be more expensive to hire personnel into less-desirable jobs than it would be to hire personnel of comparable quality into jobs that are more attractive, all else equal. There are, of course, many different factors besides salary and benefits that influence the relative attractiveness of a job. For example, teachers are likely to favor jobs in areas that have a low cost of living and greater amenities and in well-led schools with students who arrive at school ready to learn. (pp. 4, 5)

As Chambers (1981a) outlined the cost simulation model, factors affecting wage variations included two categories of variables: those within local control and those outside of local control. Within local control in this context means within the control of school district decision makers: that is, all of those factors over which decision makers can exert choice. Using the analyses of all factors affecting wages, these studies of teacher labor markets simulated how much of the variations in wages were due to factors outside local control (e.g., the locational attributes of the district and generally the characteristics of the students they served). These factors outside local control were also commonly referred to as the cost factors.

Table 1 shows a list of the types of variables included in each of these two categories. The controllable factors include teacher experience, college degree level, certification level, the quality of college attended for undergraduate study, test scores, and class sizes. These factors are considered controllable because the district decision maker can, within some limits, exert choice in the long run over the characteristics of the teaching staff as well as the class sizes in the jobs to which they are assigned.

The uncontrollable or cost factors included proxy measures that were intended to reflect variations in the local cost of living such as housing costs, crime rates of the city or locale within which the district is located, distance from the central city, and population or population density of the county within which the district was located. Some of these county-level factors tend to be proxies for access to shopping and medical facilities, climatic conditions, traffic congestion, and access to certain cultural amenities.

Other cost factors included attempts to capture the climate of the school or district and the challenges of the student population being taught. These included factors such as race-ethnic composition of the students, the percent of students living in poverty or eligible for the free or reduced-price lunch program, the percent of students who are English learners, and factors reflecting the type of district – e.g., urban, suburban, or rural – or district size.

Although some of the early studies (e.g., Brazer, 1974; Grubb and Hyman, 1975; Kenny *et al.*, 1975) were done with district-level compensation and teacher characteristics

Table 1 List of independent variables used in hedonic wage analyses*Discretionary factors:**Personal characteristics of the individual teacher (T):*

Sex;
 Racial-ethnic background;
 Marital status;
 Membership in professional teacher or educational organization;
 Age,
 Total years since first began teaching (general experience),
 Total years in the present school (school-specific experience),
 and Number of breaks in service;
 Highest degree level;
 Undergraduate major.

Job assignment or classroom characteristics of the individual teacher (C):

Percentage of full time;
 Nature of assignment (itinerate or substitute teacher);
 Index of relative class size for teachers in similar subjects;
 Whether the teacher is a mentor;
 Percentage of time teaching out of field;
 Nonschool time spent on school-related activities;
 Whether the teacher assigned homework in the most recent week;
 Percentage of time teaching high- or low-achieving students.

Characteristics of the school in which the teacher is employed (S):

Indices of student behavior and problems;
 Indices of teachers' sense of support, control, influence, and overall satisfaction with their work environment;
 Racial-ethnic composition of the students at the school;
 Percentage students absent on a recent day;
 School type (i.e., elementary, secondary, special education, vocational, and alternative);
 an index of admission requirements.

*Cost factors:**Characteristics of the district in which the teacher is employed (D):*

Racial-ethnic composition of the students in the district;
 District size as measured by enrolment;
 Percentage growth in enrolment.

Characteristics of the region in which the teacher is employed (R):

Percentage of total county enrolment accounted for by the largest district in the county;
 Measures of the distances from the closest central city;
 Percentage change in county population over the past decade;
 Value per acre of farm land;
 Population and density of the county and metropolitan area;
 County unemployment rate;
 Measures of climatic conditions (mean temperatures and snowfall);
 County-level crime rates;
 Number of banks per 100 000 population.

data, most of the studies done in the late 1970s and later (e.g., Chambers, 1978b, 1995b, 1997a; Wendling, 1979) had access to individual data on teachers and other school personnel. Such data sets on individual school personnel are commonly available in some form or another from state departments of education and include compensation along with personal and job assignment characteristics (i.e., the controllable factors).

Data on the uncontrollable or cost factors are generally obtained from school data sets, the Census Bureau, National Center for Education Statistics (NCES), National Weather Bureau, or other government agencies (e.g., the Uniform Crime Statistics of the Federal Bureau of Investigation).

For most of these studies, standard econometric techniques (multivariate ordinary least squares regression) were applied for estimating the parameters of what economists refer to as hedonic wage models to carry out the analyses of wage variations and cost index simulations.

Empirical Estimates of Compensating Differentials across the US

Variations in the costs of comparable teachers vary significantly across the nation and the states. The two national studies conducted by Chambers (1995b, 1997) using national data gathered by the NCES through the Schools and Staffing Survey for teachers and principals and the Current Population Survey (CPS) data for classified (noncertified) staff shows an overall range of variation of about 3.4 to 1 across all districts with the highest versus lowest average cost of comparable teachers. The full range of variation across individual districts was from 55 to 187, with 100 representing the index value for the district attended by the average student. The map in **Figure 1** shows the patterns of the average variation in the overall costs of education, most of which is accounted for by school personnel, including teachers, principals, and classified personnel, across the US estimated by Chambers (1997) using the national data sources.

Correlations over Time

Based on the national study of geographic cost differences done by Chambers for NCES, the data show that the indices are highly correlated over time: that is, the factors that affect geographic differences do not change very much from one year to the next. Over the 6-year period, the correlations were 0.96. This high correlation results from a consistent set of cost factors (independent variables used to capture the compensating differentials) being used with little change in the parameter estimates of the hedonic wage equation used to combine them into a single index.

Analysis of Differences between Public and Private Schools

In two other reports by Chambers (1985, 1995a), hedonic wage models were used to analyze the differences in

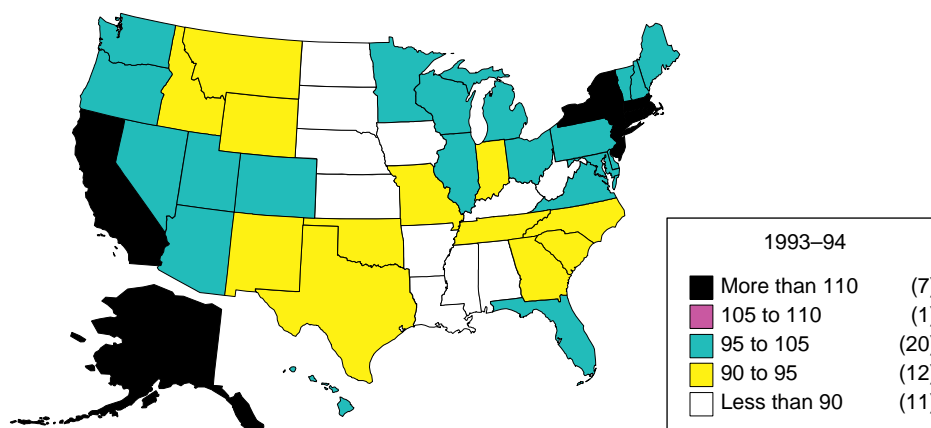


Figure 1 Geographic cost-of-education index by state, 1993–94. The geographic cost-of-education indexes are based on statistical analyses of the patterns of differences in the wages and prices of school inputs. Data sources include the following: (1) Bureau of the Census: Current Population Surveys, 1987–1994; 1990 Census of Governments, Survey of Local Government Finances; County level census files; (2) Bureau of Labor Statistics: Producer Price Indexes-1985–1994 data; Consumer Price Indexes. 1985–1994 data; (3) California Department of Education and Ohio Department of Education, databases on expenditures by object codes; (4) Geographic Names Information Systems (GNIS) CD-ROM (latitudes and longitudes for most U.S. cities, towns, and geographic locations); (5) Higher education Research Institute at the Graduate School of Education, electronic database on Scholastic Aptitude Test (SAT) scores for entering freshman, 1972, 1977, 1982 in approximately 2300 colleges in the United States; (6) National Climatic Data Center and the National Center for Atmospheric Research, The World Wealth Disc: Climate Data for the Planet Earth, CD-ROM; (7) US Department of Education, National Center for Education Statistics, Common Core of Data; Schools and Staffing Survey; 1990 Census School District Special Tabulation (summary file set I); and (8) US Federal Bureau of Investigation. (1995). The Uniform Crime Report (UCR), Return A, for the US Washington, DC.: US Department of Justice.

compensation between public and private schools. The notion here is that in the US, private schools tend to pay substantially less, on average, than public schools. In the 1996 study, Chambers reported that based on the Schools and Staffing Survey, which is a large random sample of more than 45 000 public and more than 8000 private school teachers, “public school teachers earn between about 25 to 119 percent more than private school teachers earn, depending upon the private subsector.” (p. 54). Private Catholic and nonsectarian private schools tend to be the highest paying, while Conservative Christian, Catholic parochial, and other religious-unaffiliated schools are the lowest-paying schools.

In his analysis, Chambers found that it was not possible to fully account for the differences in pay rates between public and private schools based on compensating differentials:

Between about 2 and 50 percent of this public–private difference can be accounted for by differences in teacher characteristics depending upon the private subsector. Controlling for differences in teacher and school characteristics between the public and private sectors, one observes a residual difference in the salaries of teachers that is simply associated with the sector in which the teacher is employed. (p. 54)

Basically, the model was not able to fully explain the differences in public–private teacher wage differences. One common perception is that the private school sector is a more desirable place to work or that there are unobservable

differences in teacher qualifications. To some degree, the differences may be a result of the lack of a free flow of teachers between the two sectors due to certification requirements in the public sector that do not exist in the private sector.

Problems with the Hedonic Wage Model

Most of the variations in geographic costs that have been measured in these studies represent broad regional or labor market differences rather than district-by-district differences. These are the factors that reflect the true cost of living as well as regional amenities that make one place more desirable to live than another. It has been difficult with the hedonic wage model to tease out the compensating differentials associated with unique district-level characteristics such as student poverty, English learner status, or other such variables that reflect the challenges associated with teaching.

Regional Factors and the Comparable Wage Index

Taylor (2006) has used various census and occupational-level data sources to estimate a comparable wage index (CWI) that is intended to capture these types of regional compensating differences. The notion here is that more or less the same factors that affect the wages paid to nurses,

business managers, lawyers, and other occupations requiring a college degree also affect teacher wages. Taylor includes only noneducation personnel in the analysis, and the advantage of using these data is that the index is independent of the actual levels of compensation being paid in local school districts. Rather, it captures the general context of the labor market within which local school districts must employ teachers or other school personnel.

The downside of the CWI is that it does not capture the tastes and preferences of teachers for these regional characteristics which may differ from those of individuals in other occupations. Perhaps more importantly, it does not capture any district unique factors (e.g., the challenges associated with the particular student population within a given district) that might make one district more desirable than another.

Measuring Teacher Quality

A more general problem that has been recognized with the hedonic wage models and the attempt to estimate compensating wage differentials results from the difficulty of obtaining adequate and comprehensive measures of teacher quality. This has been a problem acknowledged in the general literature on compensating differences (e.g., see Duncan and Holmlund, 1983) as well as the specific literature focused on teachers (e.g., see Goldhaber *et al.*, 2007; Boyd *et al.*, 2006). As Duncan and Holmlund state it:

If important but typically unmeasured characteristics such as “motivation” and “intelligence” lead to both higher pay and better working conditions, then the omission of measures of these characteristics may well bias the estimated relationship between wages and working conditions. (p. 367)

Goldhaber *et al.* (2007) point out that teacher quality is difficult to quantify because of the weak links between the commonly available teacher attributes such as experience, education, and certification and student outcomes (see Goldhaber and Brewer, 2000; Rivkin *et al.*, 2000). Other variables such as teacher intelligence, motivation, verbal ability, ability to relate to children, and classroom management skills are generally not easily accessible or measurable, and there is likely a correlation between these attributes and the working conditions to which teachers are subjected. The bottom line is that it may be difficult to obtain unbiased estimates of the cost factors to the extent that there is a correlation between unmeasured components of teacher quality and the included cost factors.

For example, consider the impact of student poverty on teacher wages. Wage differentials associated with student poverty may have many interpretations. Teachers may perceive a school with high percentages of students living in poverty as a challenging educational environment: that is,

higher poverty would require a compensating differential – i.e., high wages for teachers – all else equal. At the same time, poverty could be a proxy for the wealth of the community which one would anticipate to be associated with a greater ability to pay for and demand high-quality teachers: that is, lower poverty would be associated with higher wages reflecting the demand for unmeasured teacher quality.

Interpretation of Cost Factors

Another complicating issue with a variable like poverty is that it may be merely a proxy measure for something else that is difficult to measure. That is, what are the characteristics of the students that teachers would require a compensating differential? Is it poverty status, minority status, or status as an English learner *per se*, or is it the behaviors of students or the communities that may be correlated with these characteristics? For example, each of these characteristics could be associated with greater educational challenges or discrimination on the part of teachers of a different race-ethnic background. High-poverty communities may also tend to be associated with higher crime rates in the neighborhood or school and be regarded as threatening to someone considering the school or district as a place to teach. This is further complicated by the fact that the student behavior itself can be affected by district and school programs (for further discussion see Chambers (1981b)). Thus, it is difficult to measure and determine the impact of the job and assignment characteristics on wage differences.

Teacher Quality and Fixed-Effects Models

Chambers *et al.* (2006) used a teacher fixed-effects model to try and address the difficulty in measuring teacher quality. The fixed-effects model uniquely identifies each teacher in the estimation process using dichotomous variables and requires multiple years of data to capture these effects. The notion is that each teacher represents a bundle of qualifications which are essentially controlled for in the estimation technique. The difficulty with this approach is that the parameters used to estimate the regional- or district-level compensating differentials rely entirely on those teachers that move between schools and districts. With limited amounts of such movement across districts, it is difficult to estimate the parameters of this relationship.

Flexible Wages and the Nature of the Labor Market

Another flaw in the use of the hedonic wage models is that they essentially assume a competitive labor market similar to that observed in the private sector. That is, in such a labor market, wages are flexible and adjust to equate the

supply and demand for various employee and employer characteristics. While in most states there are generally many school districts competing for teacher services, there are some significant factors (e.g., lockstep salary scales, collective bargaining agreements, and state certification regulations) that create rigidity and reduce the applicability of the hedonic wage model. As Goldhaber *et al.* (2007) express it:

If teacher salaries are in fact ‘sticky’ in the sense that they don’t rapidly adjust to reflect preferences, then a key assumption of the hedonic model is violated and the distribution of teachers across schools does not convey the same information that it would in the context of a fully competitive labor market. In effect, there is likely to be an asymmetry in the information that we observe. If a salary offer for a particular teaching position does not meet the reservation wage of a teacher with a particular set of credentials, information is provided in that we will not see that teacher in that teaching position, and from this we can infer that a particular salary is not high enough to attract that teacher.

They go on to say:

Under a hedonic model it is assumed that there is no excess payment, so any excess leads to bias in the estimates of the factors influencing a teacher’s earnings. For example, if wealthier districts with low free/reduced-price lunch (FRL) student populations tend to pay teachers in excess of what is required to induce them to accept positions in schools, then the estimate of the amount that is required to pay teachers to teach in schools with few students on FRL would be biased upward, and conversely, the model would understate how much is required to induce teachers to teach in schools with large disadvantaged student populations. (pp. 6, 7)

Taylor *et al.* (2004) have used an alternative hedonic wage model to help account for this rigidity in wages that incorporates turnover behavior. This model requires multiple years of data to allow adjustment of the estimates to reflect the fact that certain wage payments to a teacher represent censored observations. For example, the wage of a teacher who moves in a subsequent year is effectively lower than the teacher reservation wage for the position and the turnover behavior in the following year is a flag that this wage value is a censored value and not a market clearing value. By accounting for this turnover in the econometric techniques, the procedure is intended to account for this lack of flexibility in wage levels.

Location as a Factor

Moreover, Boyd *et al.* (2006) indicate that one significant factor that teachers take into account in employment decisions appears to be the distance from where they

grew up. They also demonstrate that taking this factor into account and using an alternative econometric technique (a game theoretic approach to modeling the match between employers and employees in the labor market) can lead to very different parameter estimates on particular schooling attributes than the hedonic wage model. As Goldhaber *et al.* (2007) expressed it:

Unfortunately, the statistical methodology employed by Boyd *et al.* is quite novel (as far as we are aware it has not been used before in the context of estimating compensating differentials in teaching), complex and extremely computer resource intensive. Thus, at least for the time being, it is unlikely to be a practical approach for most states and localities. (pp. 6, 7)

Nevertheless, despite this practical consideration, Boyd *et al.* (2006) do identify what appear to be intuitive results regarding compensating differentials for teachers:

We find that employers demonstrate preferences for teachers with stronger academic achievement and for non-white teachers, while teachers show preferences for schools that are closer geographically, offer higher salaries, are suburban, and have a smaller proportion of students in poverty and of a different race. As predictable as these results are, they differ from wage-equation estimates that are often used to suggest, for example, that employers do not value teacher skills and that teachers do not value salary. (p. 3)

Concluding Remarks

Since the late 1960s with the movements to improve the equity of school funding, researchers and policymakers have explored various ways of measuring compensating differentials for teachers and other school personnel for the purpose of equalizing the purchasing power of the educational dollar distributed from states to local districts. It has even crept into the discussions of the rationale for adjusting federal Title I program allocations to states for differences in the statewide average per pupil expenditure.

The hedonic wage models have become a commonly used model for measuring compensating differentials and for developing geographic cost adjustments. While these models have been used since the 1970s in numerous studies, few states have adopted cost adjustments based on these types of studies.

Research monographs and published studies conducted since 2000 have raised a number of questions about the ability of the hedonic model to estimate unbiased parameter values for these compensating differentials. The problems with the hedonic wage model that have been cited include the rigidities that exist in the teacher labor market, the difficulties of measuring teacher quality, the measurement problems with the amenities,

and the tremendous complexity of modeling the labor market transaction between employers and employees.

The various econometric models that have been used in studies of teacher labor markets for the purpose of estimating compensating differentials represent black boxes to the typical policymaker and require a significant leap of faith. Ultimately, the value of these various methodologies that have been put forth must lie in the intuition of the estimates they generate. When policymakers, who are considering these models, look at the estimates, do they make intuitive sense? Do the variables and associated coefficients that reflect the challenges of teaching in certain districts or the factors that cause one district to be more or less attractive as a place to live and work jive with their perception of reality? Sometimes they do and sometimes they do not. When they do, the black box is a little easier to accept, but when they do not, it is impossible to accept.

One practical value of the concept of compensating differentials, whether they can be explicitly measured or not, is the recognition that there do exist factors that cause one job or job assignment to be more appealing than another and that teachers do and will respond to these differences (see Hanushek *et al.*, 1999). Despite the rigidities in the labor market and the commonly used salary scales for teachers, district decision makers can take advantage of recognizing these facts and use this knowledge to influence teacher decisions regarding employment. They can choose to provide additional incentives for teachers to take on job assignments in more challenging schools or classrooms. They can do this by providing additional compensation or better working conditions such as greater access to supplies and materials, smaller class sizes or caseloads, increased participation in school-level decision making, and improved opportunities for professional development.

Bibliography

- Antos, J. R. and Rosen, S. (1975). Discrimination in the market for public school teachers. *Journal of Econometrics* **3**, 123–150.
- Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J. (2006). Analyzing the determinants of the matching of public school teachers to jobs: Estimating compensating differentials in imperfect labor markets. *NBER working paper 9878*.
- Brazer, H. E. (1974). Adjusting for difference among school districts in the costs of educational inputs: *A Feasibility Report*. In *Selected Papers in School Finance*. Washington, DC: Department of Health, Education, and Welfare, Office of Education.
- Chambers, J. G. (1978a). An analysis of educational costs across local school districts in the state of Missouri, 1975–76. *Working Papers in Educational Finance*, 7, Denver, CO: Education Commission of the States.
- Chambers, J. G. (1978b). Educational cost differentials and the allocation of state aid for elementary/secondary education. *Journal of Human Resources* **13**(4), 459–481.
- Chambers, J. G. (1981a). Cost and price level adjustments to state aid for education: A theoretical and empirical review. In Jordan, K. F. (ed.) *Perspectives in State School Support Programs, Second Annual Year Book of the American Educational Finance Association*. Cambridge, MA: Ballinger.
- Chambers, J. G. (1985). Patterns of compensation of public and private school teachers. *Economics of Education Review* **4**(4).
- Chambers, J. G. (1995a). *The Patterns of Teacher Compensation*. Washington, DC: U.S. Department of Education, National Center for Education Statistics (NCES 95-829).
- Chambers, J. G. (1995b). *Public School Teacher Cost Differences across the United States*. Washington, DC: U.S. Department of Education, National Center for Education Statistics (NCES 95-758).
- Chambers, J. G. (1997). Measuring geographic variations in public schools' costs. Washington, DC: U.S. Department of Education, National Center for Education Statistics (NCES working paper).
- Chambers, J. G., Levin, J. D., and Parrish, T. B. (2006). Examining the relationship between educational outcomes and gaps in funding: An extension of the New York adequacy study. *Peabody Journal of Education* **81**(2), 1–32.
- Duncan, G. J. and Holmlund, B. (1983). Was Adam Smith right after all? Another test of the theory of compensating wage differentials. *Journal of Labor Economics* **1**(4), 366–379.
- Goldhaber, D. and Brewer, D. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis* **22**(2), 129–145.
- Goldhaber, D., Destler, K., and Player, D. (2007). Teacher labor markets and the perils of using hedonics to estimate compensating differentials in the public sector. *Working Paper 17*. Center on Reinventing Public Education.
- Grubb, W. N. and Hyman, J. (1975). Constructing teacher cost indexes: Methodological explorations with California unified school district. In *Selected Papers in School Finance*. Washington, DC: U. S. Department of Health, Office of Education, and Welfare.
- Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (1999). Do higher salaries buy better teachers? *Working Paper*.
- Kenny, L., Denslow, D., and Goffman, I. (1975). Determination of teacher cost differentials among school districts in the state of Florida. In Tron, E. (ed.) *Selected Papers in School Finance 1975*. Washington, DC: U.S. Department of Health, Education, and Welfare.
- Rivkin, S., Hanushek, E., and Kain, J. (2000). Teachers, schools, and academic achievement. *Econometrica* **73**(2), 417–458.
- Smith, A. (1976). *An Inquiry into the Nature and Causes of the Wealth of Nations* (Glasgow edition of The Works and Correspondence of Adam Smith), edited by Campbell, R. H. London: Oxford University Press.
- Taylor, L. L. (2006). *Comparable Wage Approach to Geographic Cost Adjustment Research and Development Report*. U.S. Department of Education Institute of Education Sciences National Center for Education Statistics (NCES 2006-321).
- Taylor, L. L., Chambers, J., and Robinson, J. P. (2004). A new geographic cost of education index for Alaska: Old approaches with some new twists. *Journal of Education Finance* **30**(1), 51–78.
- Wendling, W. (1979). *Cost-of-Education Indexes for New York State School Districts*. Denver, CO: Education Commission of the States.
- Wood, R. C., Chambers, M. L., Mendonca, S., and Birkett, K. F. (2001). Public school finance programs of the United States and Canada: 1998–99 – state of Florida. U.S. Department of Education, Office of Educational Research and Improvement National Center for Education Statistics, NCES 2001–309.

Further Reading

- Augenblick, J. and Adams, K. (1979). *An Analysis of the Impact of Changes in the Funding of Elementary/Secondary Education in Texas: 1974–75 to 1977–78*. Denver, CO: Education Commission of the States.
- Ballou, D. and Podgursky, M. (1993). Implicit markets for teacher and school characteristics. *Paper Presented at the American Statistical Association Annual meeting*, San Francisco, CA, 8–12, August.
- Ballou, D. and Podgursky, M. (1995). Education policy and teacher effort. *Industrial Relations* **34**(1), 21–39.

- Ballou, D. and Podgursky, M. (1997). *Teacher Pay and Teacher Quality*. Kalamazoo, MI: E. Upjohn Institute for Employment Research.
- Barro, S. M. (1992). Cost-of-education differentials across the states. *Revised Draft Report*. Washington, DC: SMB Economic Research, Inc.
- Baugh, W. H. and Stone, J. A. (1982). Mobility and wage equilibration in the educator labor market. *Economics of Education Review* **2**(3), 253–274.
- Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J. (2005a). The draw of home: How teachers' preferences for proximity disadvantage urban schools. *Journal of Policy Analysis and Management* **24**(1), 113–132.
- Brown, C. (1980). Equalizing differences in the labor market. *Quarterly Journal of Economics* **94**(1), 113–134.
- Ekeland, I., Heckman, J., and Nesheim, L. (2004). Identification and estimation of hedonic models. *Journal of Political Economy* **112**(1), S60–S109.
- Hanushek, E. A. (1996). School resources and student performance. In Burless, G. (ed.) *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success*, pp 43–73. Washington, DC: Brookings Institution.
- Hanushek, E. A. and Pace, R. R. (1995). Who chooses to teach (and why)? *Economics of Education Review* **14**(2), 101–117.
- Ingersoll, R. (1995). *Teacher Supply, Teacher Qualifications, and Teacher Turnover*. Washington, DC: U. S. Department of Education, National Center for Education Statistics (NCES 95-744).
- Loeb, S. and Page, M. (2000). Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation. *Review of Economics and Statistics* **82**(3), 393–408.
- McLean, H. M. (1978). Compensating wage differentials for hazardous work: An empirical analysis. *Quarterly Review of Economics and Business* **18**(3), 97–107.
- McMahon, W. and Chang, S. (1991). *Geographical Cost-of-Living Differences: Interstate and Intrastate, Update*. Normal, IL: Center for the Study of Educational Finance, MacArthur/Spencer Special series of November 20.
- Mont, D. and Rees, D. I. (1996). The influence of classroom characteristics on high school teacher turnover. *Economic Inquiry* **34**(1), 152–167.
- Nelson, F. H. (1991). An interstate cost-of-living index. *Educational Evaluation and Policy Analysis* **13**, 103–110.
- Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy* **82**, 34–55.
- Rosen, S. (1977). Wage-based indexes of urban quality of life. *Paper Presented at CUE Conference on Urban Economics, Academy for Contemporary Problems*, Columbus, Ohio, 5th May.
- Rosen, S. (1986). The theory of equalizing differences. In Ashenfelter, E. and Layard, R. (eds.) *Handbook of Labor Economics Volume I*, pp 642–692. Amsterdam: Elsevier Science.

Relevant Websites

- <http://nces.ed.gov> – National Center for Education Statistics: Digest of Education Statistics.
- <http://www.fsba.org> – The Florida School Boards Association.

Economic Approaches to Teacher Recruitment and Retention

S Loeb and J Myung, Stanford University, Stanford, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The quality of teaching in a school results from a range of factors, including available resources, curriculum, and instructional leadership, but it is also driven by the individuals who teach in each classroom. The staffing of teachers in schools, in turn, is a product of both recruitment and retention practices. This article describes how the choices of teachers and the actions of schools and districts influence who enters the profession and who stays. It then identifies common policy approaches for advancing recruitment and retention goals and summarizes the current research, discussing the effectiveness of these policies. The article focuses on teacher labor markets in the United States (for information on teacher labor markets outside of the US, see Ladd (2007) and Vegas (2007)).

The supply and demand model provides a simple framework for considering recruitment and retention. Wages and nonpecuniary job attributes combine to determine the supply of individuals interested in teaching in a given school, district, or state. A large body of research suggests that, like other workers, potential and current teachers respond to wage changes, although research on the degree of this response is not conclusive. Nonpecuniary components of teaching that influence the supply of teachers include working conditions, school location, and ease of entry into the occupation and the school. Feelings of success in the classroom also appear to be important for the retention of teachers already in the workforce.

The demand for teachers and the institutional constraints within which these demands are expressed also affect the teacher workforce. The number and characteristics of teachers demanded constitute a function of many factors, including student enrolment, teacher turnover, and the ability and willingness to pay for teachers. Institutional constraints, such as the skill and efficiency of hiring authorities, available information on the quality of individual teachers, budget timing, certification and licensure policies, tenure policies, and teacher contract provisions, can all affect the ability of districts to recruit and retain teachers.

In what follows, we address supply- and demand-side factors affecting this workforce in more detail, and conclude with a discussion of policies aimed at improving recruitment and retention.

The Supply of Teachers

The decisions of eligible individuals willing to teach aggregate to determine the teacher labor supply. Multiple factors affect the choices individual teachers make. Research has enumerated a few of these factors, including relative wages, working conditions, job location, and ease of entry into the occupation and into each job. Teacher self-efficacy also factors into teachers' decision of whether to remain in teaching.

Wages

A large body of literature suggests that individuals are more likely to choose to teach when starting teacher wages are high relative to wages in other occupations. Drawing upon multiple data sources, Bacolod (2007) found that highly qualified teachers are especially sensitive to changes in relative wages. Over the long run, trends in relative teacher pay have correlated with trends in teacher quality (Corcoran *et al.*, 2004). Wages may also affect retention. Murnane and Olsen (1990) found that teachers who are paid more stay longer in teaching, but that wages influence retention less for teachers with high test scores than for teachers with lower scores. Approximately 15% of public school teachers who decided to move to another school in 2004–2005 reported having done so for better wage or benefits (Marvel *et al.*, 2007). Wages are clearly associated with the retention decisions of teachers, although the causal analysis of this is less clear since high teacher wages in schools are often associated with a variety of other reasons for which teachers may choose to stay, such as better working conditions or higher student achievement.

While teachers respond to wages, much of the variation in teacher wages is between districts, reflecting differences in alternative wages, not within labor markets. Thus, the differences that we see across schools in the supply of teachers are likely driven by nonpecuniary characteristics of the jobs (Loeb and Page, 2000).

Working Conditions

Nonpecuniary job characteristics strongly affect the dynamics of the teacher labor market. While in some occupations additional wages compensate for adverse working conditions, in teaching, the single wage schedule

at the district level in the United States, and in many countries at the national level, can lead to great variation in the appeal of teaching in different schools, driven by variation in nonwage characteristics of the job.

Multiple studies have documented a relationship between teachers' career choices and the school's student population: teachers, on average, prefer schools with high-achieving, high-income, and white students. Whether these preferences are driven by direct preferences for particular types of students or by differences in working conditions in the schools these students attend is less clear. As an example, Georgia elementary teachers move from schools with higher proportions of minority students and from low-performing schools, but the latter appears to be explained by teacher preferences for fewer minority students (Scafidi *et al.*, 2003). Texas and New York data, on the other hand, found that teachers prefer higher achieving students even after controlling for student racial composition. Teachers, especially highly qualified teachers, are more likely to transfer or quit when teaching lower achieving students (Boyd *et al.*, 2005b; Hanushek *et al.*, 2004). As further evidence of the weight some teachers put on student-body characteristics, when class size reduction in California increased the demand for teachers across the state, many teachers in schools with low-achieving students switched to schools with higher achieving students (Betts *et al.*, 2000).

Student characteristics are not the only working condition that affects teachers' choices, in particular, school leadership also affect teachers' decisions. Approximately 37% of teachers move from their school due to dissatisfaction with their administrators (Marvel *et al.*, 2007). Weiss (1999) found perceived school leadership to be among the strongest variables associated with first-year teachers' feeling that it is worthwhile to exert their best effort, commitment to career path, and intentions to stay in teaching. An effective principal may have the ability to create a positive working environment for teachers, in spite of attributes of schools typically associated with high turnover. Other school factors are also important. A study of California teachers found that among the strongest predictors of turnover in a school are teachers' ratings of their tangible school conditions, such as physical facilities and availability of textbooks and technology, as well as the quality of professional development, involvement of parents, and quality and appropriateness of tests teachers are required to administer (Loeb *et al.*, 2005). Buckley *et al.* (2005) also found that facility quality is an important predictor of the decision of teachers to leave their current position, even after controlling for other contributing factors.

Psychic Benefits and Costs

Tangible working conditions are part of a job's appeal but teachers also respond to less-concrete job attributes.

Johnson and Birkeland (2003) found that new teachers who find that they cannot achieve a sense of success with students are less likely to find teaching a rewarding work and to remain in the classroom. Teachers who feel successful with students and whose schools were organized to support them in their teaching – providing collegial interaction, opportunities for growth, appropriate assignments, adequate resources, and school-wide structures supporting student learning – were less likely to leave their school than teachers in schools who were not organized to support them.

Farkas *et al.* (2000) similarly found the primary source of satisfaction among new teachers who planned to continue teaching was their confidence that they were making a difference in the lives of their students. In teachers' decisions to stay, leave, or transfer schools after the first year of teaching, more than anything else, teachers weighed whether they could be effective with their students. Difficult working conditions can affect a teacher's opportunity to teach well which, thus, affects his/her ability to succeed with students; however, it is this success that may ultimately determine whether or not the teacher chooses to stay.

School Location

School location has a strong influence on the distribution of teachers. Of all public school teachers who chose to move from one school to another between 2003–2004 and 2004–2005, 26% cited proximity to home as a very or extremely important factor in their decision to move; and of those who left teaching, 11% cited changing residence as very or extremely important (Marvel *et al.*, 2007). Most teachers prefer to teach close to where they grew up or in districts that are similar to the districts they attended as high-school students. Sixty-one percent of teachers who entered public school teaching in New York State between 1999 and 2002 started teaching in a school district located within 15 miles of the district where they went to high school, and 85% of teachers started in teaching in schools within 40 miles of their high school (Boyd *et al.*, 2005). Reininger (2006) found that these results are consistent nationwide in the US; in comparison to college graduates in nearly 40 other occupations, teachers were significantly more likely to reside in their hometown 8 years after high-school graduation.

Teachers' preferences to teach close to home or in similar settings pose serious concerns for urban districts, since urban areas produce a lower proportion of college graduates, and thus potential teachers, than do suburban areas. Rural areas also often have a smaller pool of college-educated workers from which to recruit teachers. Schools with large minority enrolments and large percentages of students receiving free and reduced-price lunch have significantly lower percentages of students earning bachelor's

degrees – a prerequisite for teaching. As a result, schools in these regions depend on hiring teachers from other regions. If they are unable to find qualified candidates, then they are forced to hire from a less-qualified pool of applicants or increase compensation.

Barriers to Entry

Traditionally, teaching in public schools in the United States required at least a bachelor's degree and certification, which in turn specifies coursework requirements, student teaching experiences, and a passing score on at least one standardized certification test. In theory, these requirements improve teaching by ensuring a minimum standard of quality on all teachers. However, these requirements also impose costs on qualified prospective teachers, which may deter them from entering the profession, effectively reducing the supply of teachers.

Until recently, while in theory teachers were required to be certified, in practice, many large urban areas employed a substantial number of uncertified teachers. Potentially, as a response to the Highly Qualified Teacher provision of the No Child Left Behind Act of 2001, schools and particularly schools serving a high proportion of students in poverty hire far fewer uncertified teachers. This tightened adherence to certification was accompanied by the creation of a number of alternative certification programs that reduced the entry requirements for teaching. Many states rely heavily on alternative routes for teachers. New Jersey, Texas, and California, for instance, obtain more than one-third of their new teachers from alternative routes. The reduced entry requirements in combination with substantial recruitment effort have substantially expanded the pool of individuals interested in becoming teachers. Furthermore, these new candidates often have stronger academic backgrounds than teachers entering from more traditional routes (Boyd *et al.*, 2006).

The Demand for Teachers

The supply of teachers determines the number of individuals willing to enter the profession and to teach in a given school, but the number of teachers actually hired and the characteristics of those teachers also depend on the demand. Important among demand factors are student enrolments, teacher retirement rates, class sizes, district hiring practices, and institutional constraints, which are described below.

Student Enrolment and Teacher Retirement

Due to the post-World War II baby boom, student enrolment increased in the United States in the 1950s and 1960s. Student enrolments declined by approximately

5 million between 1970 and 1990, but have since been steadily increasing. The baby boom era triggered a dramatic increase in the demand for teachers. Since the baby-boom generation moved through school, student enrolment changes have not driven as substantial an increased demand for teachers. However, currently, the teachers hired in the baby-boom era are reaching retirement age. Approximately 31% of public school teachers were aged 50 years or more in 2004–2005 (Marvel *et al.*, 2007). This segment of the teaching force is likely to retire over the next 10–15 years, which increases the demand for new teachers.

Reduction in Student-to-Teacher Ratios

Student-to-teacher ratios, which also affect the demand for teachers, have declined substantially during the past half century. In 1955, the ratio was 26.9; by the fall of 1985, it was 17.9, and in 2005, the average student/teacher ratio was 16.2 across all regular public schools (Marvel *et al.*, 2007). Federal policy has contributed to the decline of student–teacher ratios and the related increased demand in teachers since the 1970s. The Individuals with Disabilities Education Act (IDEA), implemented in 1975 and reauthorized in 2004, requires schools to provide accommodations for students with learning disabilities. Many schools have hired additional teachers to support students to comply with the act.

In a review of the research on class size effects, Hanushek (1998) attributed approximately a third of the decline in student–teacher ratios to special education accommodation. More recently, state policies such as the California Class Size Reduction Initiative of 1996, which paid schools to cap class sizes at 20 in grades K–3, have contributed to an increasing demand for teachers.

Hiring Processes

School and district hiring processes also affect demand and the resulting teacher workforce. In a study documenting district hiring practices across New York State, Balter and Duncombe (2008) found that most districts advertise openings in local newspapers and on the Internet; work with local colleges by supervising student teachers, posting job notices on campus, and contacting college faculty; attend at least one job fair; and use compensation for extracurricular activities and for outside teaching experience as recruitment incentives. Almost 90% of districts also use strategies to increase the local supply of teachers, such as recruiting substitute, alternatively certified, or retired teachers; or by providing assistance for paraprofessionals to become teachers. In spite of the efforts of districts in recruitment and hiring, however, it is difficult to tell who will be a good teacher. Jacob and Lefgren (2006) show that while principals are able to identify the

best and the worst teachers in their schools, they are not able to identify where the rest fall in the ability distribution. It is clearly even more difficult to tell who will be a good teacher during the hiring process. In a study of teacher hiring practices in New York State School Districts, districts most often chose candidates for interview on the basis of certification in the subject to be taught, major in the subject to be taught, and references or recommendations. A much smaller proportion of schools considered measures of a candidate's academic success such as his/her certification exam score, caliber of certifying institution, grade point average (GPA), and quality of teacher portfolio (Balter and Duncombe, 2008).

Institutional Constraints

The problem of suboptimal staffing is also driven by institutional constraints, outside the immediate control of schools and the district human resources department. In a study of district hiring patterns, the New Teacher Project uncovered three district-level policies contributing to the delays leading to suboptimal staffing patterns: lenient vacancy notification requirements, teachers' union transfer requirements, and late budget timetables (Levin *et al.*, 2005).

Lenient vacancy notification requirements do not require resigning or retiring teachers to provide notification of their intention to leave until late in the summer before the next school year. Such late notification deadlines make it very difficult for administrators to know which posts will be available when the school year starts, typically in September. By the time some districts extend offers, many of the applicants have already accepted other offers and have withdrawn their outstanding applications. Applicants who withdraw from the process early to accept other positions tend to be significantly better qualified than new hires in terms of undergraduate GPAs, a degree in their teaching field and completion of educational coursework.

Union contract provisions leave room for experienced teachers to request last-minute transfers, which excesses less senior incumbent teachers. In response, many principals delay advertising vacancies for fear of being required to hire a transferring teacher they do not want. Finally, as a result of late state budget deadlines, administrators are unaware of which positions will be funded in their schools. In 46 states, the fiscal deadline is not until 30th June, and even then, states can get extensions. Although stringent union contracts can decrease hiring effectiveness, in a study of the legal and policy structures designed to place high-quality teachers in high-minority schools, Koski and Horng (2007) did not find persuasive evidence that the seniority preference rules associated with union contracts independently affect the distribution of teachers across schools or exacerbate the negative relationship

between higher minority schools and noncredentialed and low-experience teachers.

Recruitment and Retention Policies to Date

Districts that face difficulty in hiring or retaining the teachers that they want aim to increase the supply of teachers and/or to remove institutional constraints to facilitate more effective hiring. This section looks at the following policies addressing recruitment and retention of teachers in the United States: partnerships between districts and local colleges, monetary incentives, changes in entry requirements, teacher induction and mentoring, performance-based pay, career differentiation, improving hiring practices, and modifying teacher due-process procedures. A review of extant literature reveals a lack of research that convincingly identifies the effects of most of these policy approaches.

Partnerships Between Districts and Local Colleges

To recruit potential teachers into the teaching pipeline, some districts have created partnerships with local colleges to encourage students to enter teaching. In New York State, for example, the most common college recruitment strategies used by districts are supervision of student teachers, posting of job notices at the colleges, and contacting college faculty in local colleges (Balter and Duncombe, 2008).

As a second example, the Urban Teacher Academy Program (UTAP) in Broward County Public Schools in Florida prepares high-school students for careers in urban education. This grow-your-own model provides successful program graduates with a scholarship at one of the district's higher education partners. While in college, these students major in education with opportunities for field experience in local schools. After finishing college, graduates are guaranteed a teaching job in the district. As of yet, no rigorous analyses of the effectiveness of such programs on teacher recruitment and retention have been conducted.

Monetary Incentives

In recent years, a number of states have experimented with various ways to offer higher compensation to prospective teachers to aid in recruitment and retention. Signing bonuses or crediting teachers for their years of experience teaching in other districts are examples of monetary incentive bonuses for recruitment. Some bonuses are paid in increments over time to promote retention.

Research on the effectiveness of monetary incentive programs for recruitment and retention is not conclusive. One such program, the Signing Bonus Program, implemented in Massachusetts in 1998 combined heavy recruiting, and a 7-week fast-track certification program, and a \$20,000 bonus paid in increments to all participants who continued to teach for 4 years in the state. The program did not succeed in retaining its participants – 20% of the first cohort of bonus recipients left teaching after 1 year, and attrition was particularly high in state-designated, high-need districts. Furthermore, over 50% of its second cohort ended up teaching in schools outside of the state-designated, high-need school districts for which the program was intended (Fowler, 2003).

Conversely, Clotfelter *et al.* (2006) found positive effects of North Carolina's program that provided yearly \$1,800 bonuses to teachers of math, science, and special education in middle and high schools serving low-income or low-performing students. The authors estimate that this program reduced teacher attrition by approximately 14%, though, perhaps because school eligibility for the bonus for a given academic year was not usually announced until the year had started, the program was not an effective recruitment tool.

Changes in Entry Requirements

Many states, in an attempt to increase the supply of teachers without the high cost of monetary incentives, are expanding the pool of potential teachers by reducing the cost of entry for academically competent individuals. Forty-seven states and the District of Columbia have some form of alternative-route program to recruit, train, and certify teachers. Many of these states rely heavily on alternative routes for teachers. Although alternative certification programs vary in size, scope, and competitiveness, the offer of alternative certification appears to be an effective recruitment strategy. Nearly 50% of those entering teaching through alternate routes say they would not have become a teacher if an alternate route to certification had not been available. Approximately one-third of entrants into teaching through alternate routes are nonwhite compared to 11% of the current teaching force. In terms of retention, nearly two-thirds of the survey respondents entering teaching through alternate routes expect to be teaching K-12 about 5 years from now. States with the highest percentage of alternatively certified teachers report that 87% of them are still teaching after 5 years. (Feistritzer, 2005).

Teacher Induction and Mentoring

Beginning teacher induction and mentoring have grown in prominence in school districts as methods to support new teachers' transition into the profession and to increase teacher retention. Induction programs typically involve

meetings, informal classes for new teachers, and the formation of new-teacher peer-support groups. The duration, intensity, and content of mentoring interactions can greatly vary across programs. Mentoring programs typically pair new teachers with experienced ones.

Studies of mentoring programs to date suggest that this may be a promising approach for increasing the retention of early career teachers. However, they are based on nonexperimental data and it is possible that districts or schools that implement high-quality mentoring differ from other districts, perhaps by being well run in other dimensions, and it is the other differences that drive the relationships that we see. In a synthesis of 10 empirical studies, Ingersoll and Kralik (2004) found empirical support for the claim that mentoring programs have a positive impact on teachers and their retention. Similarly, Smith and Ingersoll (2004) found that the turnover rates among new teachers decrease as the number of induction components in addition to mentoring increased – such as planning time with other teachers in the same subject, regularly scheduled collaboration with other teachers, and being part of an external network of teachers. In addition, schools that provided teachers with more autonomy and administrative support had lower levels of teacher attrition and migration. These studies suggest that mentoring may be a useful tool for retaining early career teachers. In one of the more convincing studies, Reed *et al.* (2006) found that in California, Beginning Teacher Support and Assessment (BTSA) programs in the early 1990s reduced the probability of transfer and exit among new teachers.

Performance-Based Pay

Some policymakers believe that the traditional single-wage schedule based on teacher's years of experience and number of university units provides no incentive for teachers to increase academic performance of students, and thus discourages particularly effective teachers from entering the classroom. Performance-based pay is a form of flexible compensation in which a portion of teachers' compensation is based on estimates of their effectiveness at raising student achievement. The unit of analysis can be individual teachers, groups of teachers, or schools, and payment can be based on student test performance or principal or peer evaluation.

Proponents of performance-based pay structures posit that rewarding teachers on the basis of an established set of goals would improve the motivation of teachers and assist in the recruitment and retention of high-quality staff. Critics of performance-based pay structures believe that teachers' output is too varied and difficult to observe. In addition, they worry that performance-based pay could distort incentives which could lead to suboptimal practices for long-term learning, such as teaching to the test. In addition, competition for merit awards could

result in competitive behavior among faculty at the same school and even reduce the appeal of teaching, particularly for individuals who are averse to risk.

There is little research on the effect of performance-pay on recruitment and retention, although the empirical research on the programs implemented to date has not found consistently positive effects from these reforms on student learning. Kelley (1999) examined the ways in which school-based performance award programs motivated teachers to modify or improve teaching practice in Kentucky, North Carolina, Colorado, and Maryland and concluded that such programs motivated teachers largely by creating conditions that increased intrinsic rewards and focused teacher efforts. Ballou and Podgursky (1993) found that teachers in districts that used performance-based pay did not seem demoralized by the system or hostile toward it, and that teachers of disadvantaged and low-achieving students were generally supportive of the system.

Career Differentiation Through Ladders

While the retention patterns of teachers are similar to that of other professions, such as nursing, social work, and accounting (Harris and Adams, 2007), some posit that teacher retention could be reduced by differentiating the profession, allowing paths for teacher promotion. Such promotions could provide the psychic benefits needed to improve retention. As an example, some career ladders divide the teaching career into stages by increasing responsibility and leadership, or by rewarding outstanding teaching practice. Career ladders have the potential to increase the job satisfaction of experienced teachers by diversifying their workday and skill set, thus increasing their likelihood of staying at the school, particularly because 20% of teachers leaving high-poverty urban schools report that more opportunities for advancement might induce them to stay (Ingersoll, 2004). Career ladders also have the built-in potential to increase retention among less-experienced teachers by presenting a challenging and rewarding future career prospect attainable without leaving the school. Brewer (1996) found evidence which suggests that later career opportunities affect quit decisions among teachers by examining the relationship between teaching and school administration. A study by Booker and Glazerman (2009) found that teachers in schools participating in the Missouri Career Ladder Program were less likely to leave the district as well as to leave teaching, as compared to those teachers in districts without career ladder programs, all else equal. However, the Missouri Career Ladder Program included bonuses with advancement, thus it is difficult to disentangle the impact of the monetary incentives on teacher retention from the impact of career differentiation itself.

Evidence of the effects of differentiation on teacher retention is mixed. Variations in the design and

implementation of career ladders influence teacher experiences with career ladders. Rosenblatt (2001) found that conditional on holding leadership roles that are well matched to individuals' skills and offer skill variety, career ladder programs can decrease the likelihood of burnout and increase teachers' intention to stay in their schools. However, career ladder programs that do not successfully match teachers skills to the position or offer variety can induce additional anxiety and stress for some teachers due to extra responsibilities (Henson and Hall, 1993). Without reasonable teacher assignment or without quality administrator support, the implementation of a career ladder policy is unlikely to have any positive effect on teacher satisfaction or retention. As with most retention and recruitment policies, there is little convincing causal evidence on either the advantages or disadvantages of career differentiation.

Improving Hiring Practices

Hiring practices have received attention from researchers, but relatively little attention from school leaders and policymakers. Given the contractual constraints placed on principals during the hiring process, principals are often forced to hire teachers late, by which time many higher qualified teachers may have already taken positions. Consequently, many teachers are hired late – more than one-third of new teachers in California and Florida were hired after the school year has already started (Liu and Johnson, 2006). Loosening institutional constraints on administrators and district personnel may increase efficiency in the hiring process. Jacob (2007) recommends that urban districts should streamline the administrative procedures associated with hiring so that they can make job offers more quickly; improve their ability to identify effective teachers from the pool of candidates; and implement a more decentralized process would likely result in better matches between teachers and schools. Furthermore, in their study of teacher hiring processes, the New Teacher Project formulated the following recommendations to facilitate more effective teacher hiring: ensure that transfer and excess placements are based on the mutual consent of teacher and receiving school, permit the timely hiring of new teachers, and better protect novice teachers who are contributing to their school.

Reform of Due Process

Teacher tenure policies were initially implemented to protect teachers who have successfully completed a probationary period from arbitrary dismissal. The job security tenure offers may attract prospective teachers in the teaching force and keep teachers already in the classroom. While little research has been conducted on the effect of teacher tenure on recruitment and retention, a study by

Brunner and Imazeki (2007) explored variation in probationary periods across districts and its relationship to variation in wage. The authors found evidence that districts compensate for longer probationary periods by offering higher wages. Wages for both beginning and experienced teachers are measurably higher in districts in states with longer probationary periods, which suggest that the offer of tenure may serve as a recruitment device, as teachers appear to value the prospect of an early tenure in their decision of where to teach. Clearly, tenure is a factor in a teacher's decision to remain teaching. After the probationary period, tenure creates a high level of job security and stability in the teaching profession, which could serve as an incentive for teachers to stay in the field, although no empirical work has been done to study the relationship between tenure with teacher recruitment or retention to date. Of course, tenure has the potentially negative effect of making it more difficult to dismiss less effective teachers and serves as a reminder that all teacher attrition may not be detrimental.

Conclusion

A growing body of research confirms the importance of teacher quality on student learning gains. These findings emerge at a time when policymakers and school leaders face growing concern about their ability to keep teachers currently in classrooms and how to replace teachers who leave. The teacher labor market is not all that different than other labor markets on average but the pool of available teachers is strikingly different across schools. Some schools, usually those with high proportions on non-white and low-achieving students, face a far more difficulty recruiting and retaining high-quality teachers. This article described how teachers' choices and their related preferences affect the supply of teachers and how the actions of schools and districts affect the demand for teachers and how supply and demand come together to create the workforce that we see. The article also summarized policy approaches to advancing recruitment and retention and the current research estimating the effectiveness of these policies. What stands out, as stands out in much of education policy research, is how little we know about the effectiveness of different policy approaches. Teachers respond to wage incentives, but nonwage aspects of jobs are at least as important in their decision to stay. Leadership plays a critical role both in working conditions and in the hiring process but, the market for school leadership faces similar issues of recruitment and retention and is an area in which we know even less.

See also: An Overview of Teacher Labor Markets; Teacher Incentives; Teacher Supply.

Bibliography

- Bacolod, M. P. (2007). Do alternative opportunities matter? The role of female labor markets in the decline of teacher quality. *Review of Economics and Statistics* **89**(4), 737–751.
- Balter, D. and Duncombe, W. (2008). Recruiting highly qualified teachers do district recruitment practices matter? *Public Finance Review* **36**(1), 33–62.
- Ballou, D. and Podgursky, M. (1993). Teacher attitudes towards merit pay: Examining the conventional wisdom. *Industrial and Labour Relations Review* **47**, 50–60.
- Betts, J. R., Rueben, K. S., and Danenberg, A. (2000). *Equal resources, equal outcomes? The distribution of school resources and student achievement in California*. San Francisco: Public Policy Institute of California.
- Booker, K. and Glazerman, S. (2009). The Effects of the Missouri Career Ladder Program on Teacher Mobility and Retention (MPR Reference No. 6333–400). Washington, DC: Mathematica Policy Research.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Journal of Education Finance and Policy* **1**(2), 176–216.
- Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J. (2005a). Explaining the short careers of high-achieving teachers in schools with low-performing students. *American Economic Review Proceedings* **95**(2), 166–171.
- Boyd, D., Lankford, H., Loeb, S., and Wyckoff, J. (2005b). The draw of home: How teachers' preferences for proximity disadvantage urban schools. *Journal of Policy Analysis and Management* **24**(1), 113–132.
- Brewer, D. J. (1996). Career paths and quit decisions: Evidence from teaching. *Journal of Labor Economics* **14**(2), 313–339.
- Brunner, E. J. and Imazeki, J. (2007). Teacher tenure: Does length of probation matter? *Paper Presented at American Education Finance Association Conference*.
- Buckley, J., Schneider, M., and Shang, Y. (2005). Fix it and they might stay: School facility quality and teacher retention in Washington, D.C. *Teachers College Record* **107**(5), 1107–1123.
- Clotfelter, C. T., Glennie, E., Ladd, H. F., and Vigdor, J. L. (2006). *Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence from a Policy Intervention in North Carolina*. Cambridge, MA: National Bureau of Economic Research.
- Corcoran, S. P., Evans, W. N., and Schwab, R. M. (2004). Changing labor-market opportunities for women and the quality of teachers, 1957–2000. *American Economic Review Papers and Proceedings* **94**(2), 230–235.
- Farkas, S., Johnson, J., and Foleno, T. (2000). *A Sense of Calling: Who Teaches and Why*. New York: Public Agenda.
- Feistritzer, C. E. (2005). *Profile of Alternate Route Teachers*. Washington, DC: National Center for Education Information.
- Fowler, R. C. (2003). The Massachusetts Signing Bonus Program for new teachers: A model of teacher preparation worth copying? *Education Policy Analysis Archives* **11**(13).
- Hanushek, E. A. (1998). *The Evidence on Class Size*. Rochester, NY: University of Rochester, W. Allen Wallis Institute of Political Economy.
- Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources* **39**(2), 326–354.
- Harris, D. N. and Adams, S. J. (2007). Understanding the level and causes of teacher turnover: A comparison with other professions. *Economics of Education Review* **26**(3), 325–337.
- Henson, B. E. and Hall, P. M. (1993). Linking performance evaluation and career ladder programs: Reactions of teachers and principals in one district. *Elementary School Journal* **93**(4), 323–353.
- Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal* **37**(3), 499–534.
- Ingersoll, R. M. (2004). *Why Do High-Poverty Schools Have Difficulty Staffing their Classrooms with Qualified Teachers?* Washington, DC: Center for American Progress.
- Ingersoll, R. M. and Kralik, J. (2004). *The Impact of Mentoring on Teacher Retention: What the Research Says*. Denver, CO: Education Commission of the States.
- Jacob, B. (2007). The challenges of staffing urban schools with effective teachers. *Future of Children* **17**(1), 129–153.

- Jacob, B. and Lefgren, L. (2006). When principals rate teachers. *Education Next* **6**(2), 59–69.
- Johnson, S. and Birkeland, S. (2003). Pursuing a “sense of success”: New teachers explain their career decisions. *American Educational Research Journal* **40**(3), 581–617.
- Kelley, C. (1999). The motivational impact of school-based performance rewards. *Journal of Personnel Evaluation in Education* **12**, 309–326.
- Koski, W. S. and Horng, E. L. (2006). Curbing or facilitating inequality? Law, collective bargaining, and teacher assignment among schools in California. Unpublished manuscript, *Prepared for California School Finance and Governance Project, Getting Down to Facts: A Research Project to Inform Solutions to California’s Education Problems*.
- Koski, W. S. and Horng, E. L. (2007). Facilitating the teacher quality gap? Collective bargaining agreements, teacher hiring and transfer rules, and teacher assignment among schools in California. *Education Finance and Policy* **2**(3), 262–300.
- Ladd, H. F. (2007). Teacher labor markets in developed countries. *Future of Children* **17**(1), 201–217.
- Levin, J., Mulhern, J., and Schunck, J. (2005). *Unintended Consequences: The Case for Reforming the Staffing Rules in Urban Teachers Union Contracts*. New York: The New Teacher Project.
- Liu, E. and Johnson, S. M. (2006). New teachers’ experiences of hiring: Late, rushed, and information-poor. *Educational Administration Quarterly* **42**(3), 324–360.
- Loeb, S., Darling-Hammond, L., and Luczak, J. (2005). How teaching conditions predict teacher turnover in California schools. *Peabody Journal of Education* **80**(3), 44–70.
- Loeb, S. and Page, M. E. (2000). Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation. *Review of Economics and Statistics* **82**(3), 393–408.
- Marvel, J., Lyter, D. M., Peltola, P., Strizek, G. A., and Morton, B. A. (2007). *Teacher Attrition and Mobility: Results from the 2004–05 Teacher Follow-Up Survey*. U.S. Department of Education, National Center for Education Statistics. Washington, DC: US Government Printing Office.
- Murnane, R. J. and Olsen, R. (1990). The effects of salaries and opportunity costs on length of stay in teaching: Evidence from North Carolina. *Journal of Human Resources* **25**(1), 106–124.
- Reed, D., Rueben, K. S., and Barbour, E. (2006). *Retention of New Teachers in California*. San Francisco, CA: Public Policy Institute of California.
- Reininger, M. (2006). Teachers’ location preferences and the implications for schools with different student populations. *Working Paper*.
- Rosenblatt, Z. (2001). Teachers’ multiple roles and skill flexibility: Effects on work attitudes. *Educational Administration Quarterly* **37**(5), 684–708.
- Scafidi, B., Stinebrickner, T., and Sjoquist, D. L. (2003). The relationship between school characteristics and teacher mobility. *Working Paper*. Atlanta, GA: Georgia State University.
- Smith, T. M. and Ingersoll, R. M. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal* **41**(3), 681–714.
- Vegas, E. (2007). Teacher labor markets in developing countries. *Future of Children* **17**(1), 219–230.
- Weiss, E. M. (1999). Perceived workplace conditions and first-year teachers’ morale, career choice commitment, and planned retention: A secondary analysis. *Teaching and Teacher Education* **15**(8), 861–879.

Further Reading

- Guarino, C. M., Santibanez, L., and Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research* **76**(2), 173–208.
- Hanushek, E. A., Kain, J. F., and Rivkin, S. G. (1999). Do higher salaries buy better teachers? *NBER Working Paper 7082*.
- Jacob, B. (2007). The challenges of staffing urban schools with effective teachers. *Future of Children* **17**(1), 129–153.
- Murnane, R. J. and Steele, J. L. (2007). What is the problem? The challenge of providing effective teachers for all children. *Future of Children* **17**(1), 15–43.
- Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica* **73**(2), 417–458.

Teacher Incentives

L Santibañez, Fundación IDEA, Mexico City, Mexico

© 2010 Elsevier Ltd. All rights reserved.

Teacher Incentives

Recent evidence suggests that teachers matter for student learning (OECD, 2005). Consequently, education systems around the world strive to recruit and retain high-quality teachers. To ensure qualified teachers get to the classroom (and remain there), salaries must be set accordingly. Because education is a labor-intensive endeavor, in many countries teacher salaries account for 60–95% of educational spending. The traditional way of paying teachers, however, using education- and seniority-based salary schedules is not designed to reward or encourage superior performance (Murnane and Cohen, 1985; Lavy, 2007; Hanushek, 1996). To tackle this problem, the use of targeted incentives or pay-for-performance has often been proposed.

The rationale behind using incentives to promote changes in individual behavior was first explored in the economics and business literature. Salary-incentive programs are designed to solve the employer's problem of motivating workers to perform well when individual effort and ability is not readily measured or observed (Asch, 2005). In other words, incentives are useful when a principal and an agent (i.e., an individual performing work for an employer) have differing objectives in a context of asymmetric information, when the employer cannot or does not want to dictate a particular procedure for attaining the outcome.

The school presents employers with a similar kind of principal-agent problem. School administrators, parents, and policymakers all want teachers to achieve certain outcomes (e.g., improve learning, develop well-adjusted adults, and build citizenship), but cannot monitor teachers' daily activities. To address this situation in a way that improves educational outcomes, incentive programs in education are being increasingly favored.

This article presents an overview of teacher incentive programs implemented in various countries. It begins with a brief background section on the kinds of incentive programs that have been put in practice. Second, it discusses the advantages of incentive programs and some of the positive results reported in the literature. These results are divided into individual and group incentives. Within these categories, the discussion is organized by type of outcome: efficiency and productivity, and the recruitment and retention of qualified teachers. Third, it explores some of the main disadvantages and criticisms of incentive programs as well as some of their unintended or adverse consequences. The last section concludes.

Background on Incentive Programs

Teacher incentive programs come in different flavors. They can be targeted to the individual or be designed to reward the group as a whole (a group of teachers, a school, or a school district). Incentives can be structured as rewards only, or as a system of rewards and sanctions. It can be a one-time event or an ongoing reward. They can be based on a relative criterion (value-added) or an absolute one (Lavy, 2007).

Most incentive programs reward teachers with salary bonuses, but some offer nonmonetary rewards. Incentives can be offered to teachers in the form of improved working conditions, in-kind contributions (e.g., educational materials), job stability, pensions, and the like (Vegas and Umansky, 2005). Other common examples of nonmonetary incentives are promotion-based (where promotions to higher grades in a career schedule are based on performance assessed over several time periods) and seniority-based incentives (when employers offer a reward later in an employee's career contingent on current levels of effort) (Asch, 2005). Both of these can be effective in improving worker productivity, but they are of limited use in educational systems. (Obviously, the first kind of incentives only works in an organization with a vertical hierarchy (e.g., workers promoted to managers). The second approach suffers from the problem of workers not wanting to retire when they are senior employees because they are being paid more than their productivity (Lazear, 1979, 1983). In addition, this approach is in some ways reflected by current salary schedules that are based on seniority. Research has found that more senior teachers are not necessarily more competent than junior teachers who have been teaching for at least 2–3 years (Hanushek, 1997). However, senior teachers receive higher salaries than novices do on all salary schedules.)

Murnane and Cohen (1985) discussed merit pay and concluded that most plans fail because the microeconomic framework under which most plans are based is not entirely appropriate for compensating teachers' work. For example, school administrators were rarely in a position to accurately explain why some teachers received merit pay and others did not, and what the latter group could do to receive merit pay in the future.

However, most recent incentive programs are merit-pay style programs focusing on student test scores as the main measure of teacher performance. Focusing on test scores is an attractive choice because they provide

observable measures of student learning. In addition, there is evidence to suggest that they are correlated with other longer-term outcomes such as lifelong earnings (Klerman, 2005). Next we discuss the main advantages of incentive programs. To illustrate these, a brief discussion of recent experiences is included. By design, most of the positive effects are on standardized student test scores. Whenever appropriate, however, we discuss effects on other outcomes of interest such as teacher attendance.

Advantages of Incentive Programs

This section discusses positive results reported by incentive programs implemented in various countries. It is organized in two main categories: incentives targeted to the individual teacher and group incentives. Within these two categories, the discussion is divided into efficiency and productivity results, and effects on recruitment and retention of qualified teachers.

Individual Incentives

Efficiency and productivity

In theory, the main advantage of incentive programs is that basing some part of a teacher's compensation (or offering incremental payments) on a common objective (learning, test scores, and graduation rates) enables the alignment of incentives directed to teachers or schools with those directed to students (Lavy, 2007). This could generate efficiency gains because teachers focus their efforts on those objectives that are valued most by parents and policymakers. Individual merit-pay schemes of this type help correct potential distortions in a teachers' effort that might result from gaps between her (work) preferences and those of her students (Lavy, 2007). In other words, incentives could produce better educational outcomes at lower or equal cost than traditional across-the-board salary raises.

There are some recent experiences that suggest the alignment of teacher and student objectives through incentives produces positive results. In India, Duflo *et al.* (2007) evaluated an intervention to reduce teacher absenteeism in rural areas through the use of teacher monitoring and incentives. An experiment of 57 randomly selected treatment schools (out of 113 schools) was conducted to explore if financial incentives would motivate teachers to have fewer absences (the baseline data suggested an absence rate of 42%). Teachers in the treatment schools were given a tamper-proof date and time camera to record their attendance at the beginning and end of the school day. Teacher salaries were calculated on the basis of valid school days. Tying salary to teacher attendance not only resulted in a drastic decline in teacher absence (to 21% in treatment schools over a period of two years) (the proper

evaluation phase lasted only 10 months; therefore, the positive effects on attendance lasted well beyond the experiment), but it also positively improved student achievement levels. A year after the program began, student test scores in treatment schools were higher (0.17 standard deviations) than test scores in comparison schools (Duflo *et al.*, 2007).

Other countries have experimented with incentives to improve teacher attendance in rural areas with less encouraging results. In Bolivia, teacher incentives were used to attract teachers to rural areas. The size of the incentive was small (between 0.3% and 1.1% of monthly salary) and was not successful in attracting teachers to these hard-to-staff areas. Furthermore, because of urbanization and demographic growth, some areas classified as rural were in fact borderline urban. The authors found that the test scores and other educational outcomes of students of urban- and rural-classified teachers with the same background and characteristics were not significantly different (Urquiola and Vegas, 2005).

In Mexico as part of the Programa para Abatir el Rezago Educativo (PARE) (this program was implemented in Mexico from 1992 to 1996 as part of a World Bank initiative to provide additional resources to very disadvantaged students, primarily those in rural areas and indigenous communities), teachers were offered monetary bonuses if they demonstrated they attended school regularly. A novel feature of PARE was that it was the parent associations who were in charge of teacher attendance records and monitoring. A qualitative evaluation of PARE, conducted in nine schools in two Mexican states, found incentives effectively reduced teacher absenteeism only in schools that had strong school principals and parent associations or had lower levels of teacher absenteeism (comparatively) before PARE was implemented (Ezpeleta and Weiss, 1996).

In Mexico, an evaluation of the national teacher incentive program *Carrera Magisterial* (teachers' career) showed slight positive effects of the monetary incentives on a small group of secondary school teachers, but virtually no effect on primary teachers (McEwan and Santibañez, 2005; Santibañez *et al.*, 2007). This program, unique in the world for its size and scope (implemented in 1992, the program is national in scope; hundreds of thousands of teachers and millions of students are tested every year as part of its yearly evaluations), awards bonuses beginning at 25% of base salary, up to 200% to teachers who consent to an evaluation of several factors including teacher knowledge (measured using a standardized teacher test) and student achievement (using a standardized student test). To evaluate the program and try to circumvent the fact that no natural comparison group exists, the authors used regression discontinuity design which compared teachers with strong incentives to those with weaker incentives during the evaluation year. (To obtain the bonus (get incorporated into the program) teachers must

obtain a minimum of 70 out of a total 100 points in the yearly evaluation. Up to 80 points of the total can be thought of as exogenous, because they either increase automatically every year or remain the same (the case of points given for seniority and highest degree earned, respectively), have virtually no variation (the case of the peer review component), or are within the control of the teacher (points given for obtaining professional development or the teacher test). Teachers with exogenous point scores below 50 or above 70 are in the weak incentive group. Teachers with exogenous point scores between 50 and 70 have a high incentive to improve student achievement, which is worth 20 points and could get them above the minimum cutoff for incorporation (McEwan and Santibanez, 2005). Results suggest that students of teachers in the strong incentive group who were competing for the bonus did exhibit slightly higher test results than teachers in the weak incentive group. This effect, however, was only observed for a very small group of secondary school teachers and not for primary teachers (McEwan and Santibañez, 2005; Santibañez *et al.*, 2007).

Using various data sources including the National Education Longitudinal Survey (NELS), Figlio and Kenny (2006) found that a strong positive association between the presence of merit pay in teacher compensation and student test scores in the United States. The authors were not able to conclude whether this positive association was the result of the incentives *per se* or of some unobserved measure of school quality not captured by the model.

Dee and Keys (2004) analyze the Career Ladder Evaluation System. This program was implemented in the state of Tennessee to improve student achievement through a blend of salary rewards with nonmonetary benefits such as more teacher autonomy and released time from teaching.

Using results from the Tennessee STAR class-size experiment, the authors found that the Career Ladder Evaluation System program had only mixed success in targeting rewards to the more meritorious teachers. They concluded that assignment to a career ladder teacher significantly increased mathematics scores by roughly three percentile points. However, most career-ladder teachers were not significantly more effective at promoting reading achievement. Furthermore, assignment to a teacher who had advanced further up the career ladder was not uniformly associated with significantly higher achievement (Dee and Keys, 2004). The authors acknowledge that one important caveat to their analysis stems from the fact that it uses an assessment system that was specifically designed to evaluate class size and not the effects of the incentive program.

Incentive programs could also encourage teachers to improve their practice or engage in professional development activities. One such study is the evaluation of the National Board Certification Pilot Project in Iowa. This project offers teachers' monetary incentives in exchange for obtaining advanced teacher certification. The evaluation

found that teachers involved in the certification process with the monetary incentives engaged in more professional development activities than teachers not targeted by the incentives (Dethlefs *et al.*, 2001). In Mexico, Ornelas (2002), in his study of CM, concluded that one of the major successes of the program's monetary incentives was to emphasize professional development. (However, it is not difficult to make the case that professional development should not be a goal in and of itself, but should be an intermediate objective to improving student learning. While it is important for teachers to receive continuous training, it is more important that training results in improvements in teaching that have ultimate beneficial consequences in the classroom.) A caveat to these programs is that teacher professional development or certification might not be an objective in and of itself. This is particularly true if these activities do not lead to better student learning. In this case, the incentive program could not lead to improvement in the bottom line goal of higher education quality.

Recruitment and retention of qualified teachers

If incentive programs correctly identify teacher productivity they will help recruit and retain more productive teachers. In theory, if the incentive program is well designed, it will tend to discourage teachers who are not as productive from remaining in the system and will tend to encourage more productive teachers to do so.

On the issue of how incentives affect teacher recruitment and retention, there is considerable US-based evidence suggesting that increasing teacher salaries affects who chooses to enter and remain in teaching. In the United States numerous studies suggest a link between higher teacher salaries and higher retention (Podgursky *et al.*, 2004; Stockard and Lehman, 2004; Kirby *et al.*, 1999; Gritz and Theobald, 1996; Brewer, 1996). Incentives used to attract teachers to less desirable schools may not work as well, however.

Group Incentives

Efficiency and productivity

Some of the gains in productivity discussed above are not only derived from individual incentive programs. Incentive programs designed to target a group of teachers or the school as a whole can help foster collaboration among teachers by getting them to work for a unified goal. This could have the dual positive effect of producing higher learning and overcome some of the potential negative effects of targeting individual teachers (Murnane and Cohen, 1986; Dee and Keys, 2004).

Glewwe *et al.* (2003) examine a teacher incentive program in Kenya which provided teachers with bonuses based on student performance and attendance. Teachers in 50 rural schools were randomly selected (out of 100)

for the incentive program based on the average test score of students already enrolled in school at the start of the program. Each year the program provided prizes valued at up to 43% of typical monthly salary to grade 4–8 teachers. The incentives were given to all teachers in the winning schools (regardless of their own individual performance), based on the performance of the school (relative to other schools) on the Kenyan government's districtwide exams.

The authors found that during the life of the program, students in treatment schools were more likely to score higher, at least on some exams, than students in the comparison group. This effect, however, tended to be short-lived. In addition, a closer examination of the channels through which this effect took place, suggested that teachers might have been teaching to the test. The authors did not find strong evidence to support that the program had induced greater teacher effort or that it would be able to reach longer term outcomes: student dropout rates did not fall, teacher attendance did not improve, the amount of homework assigned did not increase, and pedagogy did not change. There is, however, evidence that teachers' increased efforts to increase the number of pupils taking tests and to raise short-run test scores (by focusing instruction on the test). In addition, the authors found evidence that teachers adjusted to the program over time by offering more test preparation sessions (Glewwe *et al.*, 2003).

Lavy (2002) examined a small-scale program in Israel that provided incentives to teachers in 62 nonrandomly selected schools. He used a regression discontinuity approach to compare student outcomes in these schools with those of schools that just missed treatment because of eligibility rules. The program offered combined incentives to schools in the form of performance awards. These awards included merit pay for teachers and improvements in working conditions. Awards were based on student test scores and reduced dropout rates. This type of incentive program is of the rankorder type, which awards prizes based on the rank order of the winners.

An interesting feature of the Lavy (2002) paper is that it compares an incentive intervention with what he calls a resource intervention. That was a separate program awarding additional (and identical) resources, such as teacher training, to schools showing improvements. A comparison group of schools not admitted into the resource program serves as the basis for identification of programmatic effects. Lavy evaluates the effect of this parallel program and compares its effectiveness and cost to the teachers' incentives intervention.

Lavy's results suggest that monetary incentives had some effect in the first year of implementation in some schools and caused significant gains in many dimensions of students' outcomes in the second year in all schools. The program led to an increase in average test scores and a higher proportion of students who gained the high

school matriculation certificate (especially among those from a disadvantaged background). It also appears to have contributed to a decrease in the dropout rate in the transition from middle to high school. The results regarding the resource program suggest that it also led to significant improvements in student performance. However, the comparison of the programs based on cost equivalency suggests that the teachers' incentive intervention is more cost-effective (Lavy, 2002).

In Chile, the *Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educacionales* (SNED) or National System of School Performance Assessment, offers monetary bonuses to schools that show high student-achievement marks. This program was implemented in 1990 and preliminary evidence shows a cumulative positive effect on student achievement for schools which had a relatively high probability of winning the award (Mizala and Romaguera, 2003).

Summary of Key Findings

Table 1 provides a summary of key findings from more recent studies, divided by type of intervention (individual or group incentives).

As the table suggests, most of the teacher incentive interventions that showed positive results on student achievement and other outcomes, were carefully designed, small-scale interventions with sizeable benefits for teachers and schools.

Disadvantages and Criticisms

The theory behind incentive programs is appealing in its simplicity. It is intuitive that individuals will tend to work harder if they know they will be rewarded for their efforts. As the evidence from the business field shows, this logic often tends to work. In practice, however, many incentive programs in education often fail to reach their intended goals. Most authors attribute these failures to both design and implementation factors.

First, incentive program design assumes that there is a clear and known definition of performance (Klerman, 2005). This implies a common agreement about outputs. In most examples from the economics and business literature, the output is well defined and has a readily available measure for performance. For example, papers in the business literature have shown incentives improved the productivity of workers installing car windshields (Lazear, 2000) and planting trees (Paarsch and Shearer, 2000).

Working toward achieving or producing a well-defined output is not usually the case in education. Teachers and schools have multiple and complex objectives which include, but are not limited, to student learning. Self-esteem, citizenship building, development of core values,

Table 1 Summary of key findings from recent studies, by type of intervention

<i>Study</i>	<i>Country</i>	<i>Scale</i>	<i>Description</i>	<i>Intervention design</i>	<i>Results</i>
<i>Individual incentives</i>					
Santibañez <i>et al.</i> (2007)	Mexico	National scale	Provides monthly salary bonus (beginning at 25% of base salary) to teachers who pass a national evaluation including tests of their knowledge and their students' knowledge	Nonrandomized intervention, voluntary participation. Most teachers in the country enrolled in program. Evaluation used regression discontinuity to identify a strong incentive group	Students of teachers in strong incentive group had slightly higher results in secondary schools. No significant results in primary schools.
Duflo and Hanna (2007)	India	Small pilot program	Provide financial incentives to teachers who could demonstrate regular attendance through the use of tamper-proof cameras	Experiment randomly selected 120 treatment schools and 60 comparison schools	Tying salary to attendance and using rigorous monitoring (cameras) significantly and drastically decreased teacher absence and improved student achievement
Urquiola and Vegas (2005)	Bolivia	National	Offered salary bonuses to teachers who taught at less desirable (i.e., rural) schools	Nonrandomized intervention, used regression discontinuity to identify effect of bonus	Rural pay differential was not successful in attracting more effective teachers to rural areas. The bonus had no effect on student test scores.
<i>Group incentives</i>					
Glewwe <i>et al.</i> (2003)	Kenya	Small pilot program	Provide teachers with bonuses (prized at 43% monthly salary) based on student performance and attendance. All teachers in winning school get the incentive.	100 rural schools, teachers in 50 randomly selected schools get treatment	Students in treatment schools scored higher in exams, only during life of the program
Lavy (2002)	Israel	Small-scale	Offered schools combined awards (salary bonuses for teachers and improvements in school working conditions) for improvements in student test scores and reduced dropout rates	Used regression discontinuity approach to evaluate results in 62 nonrandomly selected schools	The program had positive and significant effects on student test scores and dropout rates during its 2-year duration
Mizala and Romaguera (2003)	Chile	National scale	Offers monetary bonuses to schools that show high student-achievement marks	Nonrandom selection of schools. Descriptive and qualitative analysis.	Preliminary evidence shows cumulative positive results on student achievement for schools with high probability of winning the award

and social skills are all outputs valued to varying degrees by teachers, parents, and policymakers. Even when all of us could agree that student learning is the top priority, it is not clear that a standardized test can capture learning in a way that satisfies a common definition. In addition, the

objectives of teachers and their employers (principals, school boards, and education authorities) could compete with one another and even be mutually exclusive. For example, government agents might be interested in ensuring an equitable education for all its citizens, while

teachers or school administrators might want to keep difficult or at-risk students out of the classroom (Vegas and Umansky, 2005).

Second, incentive schemes assume workers know the best way to improve their work. In other words, workers know the technology and have the means to implement it. Research on accountability in education, however, suggests that teachers could need assistance to figure out how to best improve student achievement (Hamilton, 2005). Hamilton (2005) argues that accountability reforms are often accompanied by technical assistance to schools, but that this technical assistance might not be enough to compensate for insufficient capacity problems (e.g., insufficient material and financial resources), or, for the broader context in which some schools must operate. Aside from some well-documented findings, there is a general lack of knowledge about how to improve teacher practice (Hamilton, 2005).

Third, incentive programs assume that the benefit of obtaining a desired outcome outweighs the costs of measuring performance. This is less of a problem when simple tasks are involved, but when performance is multidimensional and involves a series of complex tasks (as is the case in education), measurement costs might outweigh the benefits (Asch, 2005). This is related to the premise (on which incentive programs are based) that there is an operational (and commonly agreed to) definition of performance (Klerman, 2005). The more precise and operational this definition, the easier it will be to design a system of incentives to reach it.

Fourth, because education is a multidimensional task, focusing on a single dimension (e.g., student test scores) is problematic. Some have criticized these programs for curtailing creative thinking or teaching to the test (Hannaway, 1992; Holmstrom and Milgrom, 1991). Programs could design ways to reduce these unintended responses, for example, by combining the use of outcome measures (e.g., student tests) and practice measures (e.g., classroom observations, interviews, and supervisor ratings).

Because there is certain to be some level of noise affecting each period's evaluation, rewarding the worker for performance during a single period might expose him to considerable risk (Prendergast, 1999). For example, the program could reward teachers whose performance during the evaluation period improved due to a positive shock in working conditions rather than increased efforts in a given year. Using repeated incentives could reduce this noise and improve the quality of the performance measure. However, the costs and feasibility of using more precise measures of performance needs to be weighed against its potential benefits.

Fifth, some argue that explicit incentives (e.g., salary incentives) can reduce productivity by eliminating intrinsic desire. Most of the incentives literature in business and economics assumes that effort is costly for the worker, but

other research argues that in some cases the pride or sense of mission workers derive from their work makes carrying out their tasks an enjoyable activity. Teaching might just be such a job. Teachers often cite a desire to work with young people and contribute to society as important reasons to enter the teaching profession (Guarino *et al.*, 2006). Some authors have found that incentive programs demoralize teachers and lead to lower effort (Fehr and Schmidt, 2004 cited in Duflo *et al.*, 2007) or harm their motivation and sense of duty and enjoyment (Kreps, 1997 cited in Duflo *et al.*, 2007).

Adverse and unintended consequences of teacher incentive programs

Other critiques of using targeted incentives to improve teacher quality center on the unintended, even adverse consequences of these kinds of programs. Upon deeper analysis of why some incentive plans improved student achievement, for example, it was found that teacher cheating (Jacob and Levitt, 2003), exclusion of low achieving students (Cullen and Reback, 2002; Figlio and Getzler, 2002), focus on the tested subjects in detriment of others (Hamilton *et al.*, 2002), teaching to the test (Koretz, 2002), or even increasing students' caloric intake on the day of the test (Figlio and Winicki, 2005) were largely behind the observed results.

Group incentives can also have adverse consequences by exacerbating the potential for free-riding of teachers. Research has shown that if an employee's share of the reward is small relative to the difficulty of the work, and if the effort of all team members is difficult to observe by the employer, an individual on the team will have an incentive to shirk his or her work while benefiting from the teams' work or free-riding on the effort of others. This phenomenon weakens the power of group incentives (Asch, 2005; Prendergast, 1999; Asch and Karoly, 1993).

Conclusions

The evidence regarding the impact of programs that offer teachers monetary incentives to improve student achievement is mixed. Some small-scale programs have found that incentives are effective at improving student learning, although in most cases these results are short-lived and not always the product of increased teacher effort or improved teacher practice. Some programs (notably the India case) have found positive effects of financial teacher incentives on attendance, but this has not always been the case in other countries. Individual financial teacher incentives used to improve recruitment and retention or increasing professional development, seem to have positive effects, but those used for other purposes, such as

increasing teacher attendance, have demonstrated mixed results. Still, even if their positive results are modest, targeted incentives are more cost-effective than traditional salary increases (which have also not shown consistent positive results on student learning) and are thus more likely to remain an attractive policy tool.

Years of incentive program design and implementation have yielded programs of very different forms and shapes. All the programs reviewed here used monetary incentives to encourage teachers, but some of the programs showing positive results mix individual and group (or school-based) incentives. None of the programs discussed here had sanctions for poor performance. Some were national in scale, some targeted only a few schools. Some rewarded student achievement only, some included other dimensions of a teachers' work.

In all cases and to improve its potential for positive effects the program should have a well-defined measure of teacher performance covering all outcomes of interest and including short- and long-term measures (Lavy, 2007). Value-added measures are preferred to absolute measures that can confound context and other factors (McEwan and Santibañez, 2007; Lavy, 2007). Perhaps a mix of group and individual incentives in addition to subjective (peer or supervisor review) with objective measures can be used to counterbalance criticisms about hindering teacher collaboration and the limited scope of standardized tests, respectively (Lavy, 2007).

Bibliography

- Asch, B. J. (2005). The economic complexities of incentive reforms. In Klitgaard, R. and Light, P. C. (eds.) *High-Performance Government: Structure, Leadership, Incentives*, pp 309–342. Santa Monica, CA: RAND Corporation, MG-256-PRGS.
- Asch, B. J. and Karoly, L. (1993). *The Role of the Job Counselor in the Military Enlistment Process*. Santa Monica, CA: RAND National Security Research Division, RAND.
- Brewer, D. J. (1996). Career paths and quit decisions: Evidence from teaching. *Journal of Labor Economics* **14**(2), 313–339.
- Dee, T. S. and Keys, B. J. (2004). Does merit pay reward good teachers? Evidence from a randomized experiment. *Journal of Policy Analysis and Management* **23**(3), 471–488.
- Dethlefs, T. M., Trent, V., Boody, R. M., et al. (2001). Impact study of the national board certification pilot project in Iowa. *Report*. Des Moines: Iowa State Department of Education.
- Duflo, E., Hanna, R., and Ryan, S. (2007). *Monitoring Works: Getting Teachers to Come to School*. Cambridge, MA: Poverty Action Lab.
- Ezpeleta, J. and Weiss, E. (1996). Las escuelas rurales en zonas de pobreza y sus maestros: Tramas preexistentes y políticas innovadoras. *Revista Mexicana de Investigación Educativa* **1**(1), 53–69.
- Fehr, E. and Schmidt, K. (2004). Fairness and incentives in a multi-task principal-agent model. *Scandinavian Journal of Economics* **106**(3), 453–474.
- Figlio, D. N. and Getzler, L. S. (2002). Accountability, ability, and disability: Gaming the system. *Working Paper No. 9307*. Cambridge, MA: National Bureau of Economic Research.
- Figlio, D. N. and Winicki, J. (2005). Food for thought: The effects of school accountability plans on school nutrition. *Journal of Public Economics* **89**, 381–394.
- Figlio, D. N. and Kenny, L. (2006). Individual teacher incentives and student performance. *Working Paper No. 12627*. Cambridge, MA: National Bureau of Economic Research.
- Glewwe, P., Ilias, N., and Kremer, M. (2003). Teacher incentives. *Working Paper No. 9671*. Cambridge, MA: National Bureau of Economic Research.
- Gritz, R. and Theobald, N. (1996). The effects of school district spending priorities on length of stay in teaching. *Journal of Human Resources* **31**(3), 477–512.
- Guarino, C., Santibañez, L., and Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Education Research* **76**(2), 173–208.
- Hamilton, L. (2005). Lessons from performance measurement in education. In Klitgaard, R. and Light, P. C. (eds.) *High-Performance Government: Structure, Leadership, Incentives*, pp 381–406. Santa Monica, CA: RAND Corporation, MG-256-PRGS.
- Hamilton, L., Stecher, B., and Klein, S. (2002). *Making Sense of Test-Based Accountability in Education*. Santa Monica, CA: RAND Corporation.
- Hannaway, J. (1992). Higher order skills, job design, and incentives: An analysis and proposal. *American Educational Research Journal* **29**(1), 3–21.
- Hanushek, E. A. (1996). Outcomes, costs, and incentives in schools. In Hanushek, E. A. and Jorgenson, D. W. (eds.) *Improving America's Schools: The Role of Incentives*, pp 29–52. Washington, DC: National Academy Press.
- Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis* **19**(2), 141–164.
- Holmstrom, B. and Milgrom, P. (1991). *Special Issue: Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design*. *Journal of Law, Economics, and Organization* **7**, 24–52.
- Jacob, B. A. and Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics* **118**, 843–878.
- Kirby, S., Berends, M., and Naftel, S. (1999). Supply and demand of minority teachers in Texas: Problems and prospects. *Educational Evaluation and Policy Analysis* **21**(1), 47–66.
- Klerman, J. A. (2005). Measuring performance. In Klitgaard, R. and Light, P. C. (eds.) *High-Performance Government: Structure, Leadership, Incentives*, pp 343–380. Santa Monica, CA: RAND Corporation, MG-256-PRGS.
- Koretz, D. (2002). Limitations in the use of achievement tests as measures of educators' productivity. *Journal of Human Resources* **37**(4), 752–777.
- Lavy, V. (2002). Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of Political Economy* **110**, 1286–1317.
- Lavy, V. (2007). Using performance based pay to improve the quality of teachers. *Future of Children* **17**(1), 87–109.
- Lazear, E. (1979). Why is there mandatory retirement? *Journal of Political Economy* **87**, 1261–1264.
- Lazear, E. (1983). Pensions as severance pay. In Bodie, Z. and Shoven, J. (eds.) *Financial Aspects of the United States Pension System*, pp 57–89. Chicago, IL: University of Chicago Press.
- Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review* **90**(5), 1346–1361.
- McEwan, P. J. and Santibañez, L. (2005). Teacher incentives in Mexico. In Vegas, E. (ed.) *Beyond Pay: Motivating Teachers to Raise Student Learning. Lessons from Latin America*. Washington, DC: The World Bank.
- Mizala, A. and Romaguera, P. (2003). Regulación, incentivos y remuneraciones de los profesores en Chile. In Cox, C. (ed.) *Políticas educacionales en el cambio de siglo: La reforma del sistema escolar de Chile*. Santiago: Editorial Universitaria.
- Murnane, R. J. and Cohen, D. K. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. *Harvard Educational Review* **56**(1), 1–17.

- OECD (2005). *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*. Paris: Organisation for Economic Co-operation and Development.
- Ornelas, C. (2002). Incentivos a los maestros: la paradoja mexicana. In Ornelas, C. (ed.) *Valores, calidad y educación*, pp 137–161. Santillana/Aula XXI: México.
- Paarsch, H. J. and Shearer, B. (2000). Piece rates, fixed wages, and incentive effects: Statistical evidence from payroll records. *International Economic Review* **41**(1), 59–92. Department of Economics, University of Pennsylvania and Osaka University Institute of Social and Economic Research Association.
- Podgursky, M., Monroe, R., and Watson, D. (2004). The academic quality of public school teachers: An analysis of entry and exit behavior. *Economics of Education Review* **23**, 507–518.
- Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature* **37**(1), 7–63.
- Santibañez, L., Martínez, J. F., Datar, A., et al. (2007). *Breaking Ground: Analysis of the Assessments and Impact of the Carrera Magisterial Program in Mexico*. Santa Monica, CA: RAND/MG-141. (Also available in Spanish.)
- Stockard, J. and Lehman, M. B. (2004). Influences on the satisfaction and retention of 1st-year teachers: The importance of effective school management. *Educational Administration Quarterly* **40**(5), 742–771.
- Urquiola, M. and Vegas, E. (2005). Arbitrary variation in teacher salaries: An analysis of teacher pay in Bolivia. In Vegas, E. (ed.) *Incentives to Improve Teaching: Lessons from Latin America*, pp 187–211. Washington, DC: The World Bank.
- Vegas, E. and Umansky, I. (2005). Improving teaching and learning through effective incentives. Lessons from Latin America. In Vegas, E. (ed.) *Incentives to Improve Teaching: Lessons from Latin America*, pp 1–15. Washington, DC: The World Bank.

Further Reading

- Figlio, D. N. (2002). Can public schools buy better teachers? *Industrial and Labor Relations Review* **55**, 686–699.
- Hanushek, E. A. and Raymond, M. E. (2002). Improving educational quality: How best to evaluate our schools? In Kodrzycki, Y. K. (ed.) *Education in the 21st Century: Meeting the Challenges of a Changing World*, pp 193–224. Boston, MA: Federal Reserve Bank of Boston.
- Kane, T. J. and Douglas, O. S. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives* **16**(4), 91–114.
- Podgursky, M. J. and Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management* **26**(4), 909–949.
- Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica* **73**(2), 417–458.

Teacher Supply

P Dolton, University of London, Surrey, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

ITT (initial teacher training) – Those trainee teachers who have entered training but not yet qualified.

Nonpecuniary – Factors other than wages.

Opportunity wage – That wage that teachers could earn in the outside labor market in another job if they were not in teaching.

PIT (pool of inactive teachers) – The stock of qualified teachers in the labor market who do not teach.

PRT (pool of recoverable teachers) – The stock of teachers who have left the labor market or who are inactive but who could return to the profession.

PTR (pupil teacher ratio) – The number of pupils divided by the number of teachers.

Relative wage – The teacher wage compared to the opportunity wage (or another reference group – like the private sector).

Shortage – When demand exceeds supply.

Surplus – When supply exceeds demand.

Wastage – The flow of teachers who leave the profession.

A commitment to the education of the next generation means that the supply of high-quality school teachers must be one of the main education-policy priorities of any government. Any education system needs to recruit and retain high-quality teachers and understand the role of pay and other incentives intended both to attract people to the profession and to keep them.

Every country needs a relatively large number of teachers, as on average, in Organization for Economic Cooperation and Development (OECD) countries 2.6% of the total labor force are teachers (see OECD, 2004). In most countries, this means the education sector is the largest employer of university graduates. Accordingly, teachers pay is the largest component of a country's educational expenditure with an average of 63% of the educational budget.

In any consideration of the quality of teachers, we must be aware of the forces of supply and demand and the decisions governments choose to make about desirable pupil teacher ratios (PTRs), teacher working hours, the length and quality of teacher training, teacher pay and incentives, and the other nonpecuniary conditions in teaching

which must be affected by government educational-spending priorities.

Many countries have experienced recurrent crises in the recruitment and retention of teachers and in many there is a more or less continuous shortage of teachers, notably in secondary schools and in the most technical subjects. The shortage tends to be particularly acute in subjects such as math, science, and modern languages and in specific geographical areas, where the opportunity wage for would-be teachers is much higher.

The consequences of teacher-supply problem are inequitably felt across different socioeconomic groups in any society. It has been suggested (Darling-Hammond, 2000) that teacher quality is not only a major determinant of student achievement but also one of the most inequitably distributed resources as poor children invariably receive lower-quality teaching and poorer curricula.

The Labor Market for Teachers

The labor market for teachers can be thought of within a traditional supply-and-demand framework, with the additional complication that the government is virtually the sole hirer of labor. The demand for teachers can be determined by the number of children in the country of school age, and the government's desired PTR. For such a given ratio, the demand for teachers is therefore a constant, denoted by \bar{Q} in Figure 1. Under the reasonable assumption that the supply of teachers is a positive function of average teacher earnings, an upward-sloping labor-supply schedule can be drawn as S . In a perfectly competitive market, a teacher wage of W^* would therefore clear this labor market. However, the teachers' labor market is of course not competitive, and the government, in its role as (almost) exclusive purchaser of teaching labor, has other considerations, prime among which is the level of expenditure on teachers' salaries in total. For a given level of such expenditure, an inverse relationship can be plotted between teachers' earnings and the number of teachers hired, labeled E_1 in Figure 1; if the government wants to raise the salaries of teachers, it can afford to hire fewer of them, given a fixed budget. The number of teachers hired is therefore Q_g at average earnings of W_g , and the excess demand for teachers is $\bar{Q} - Q_g$. This can only be eradicated by a relaxing of the budget constraint leading to higher earnings, or other factors changing to

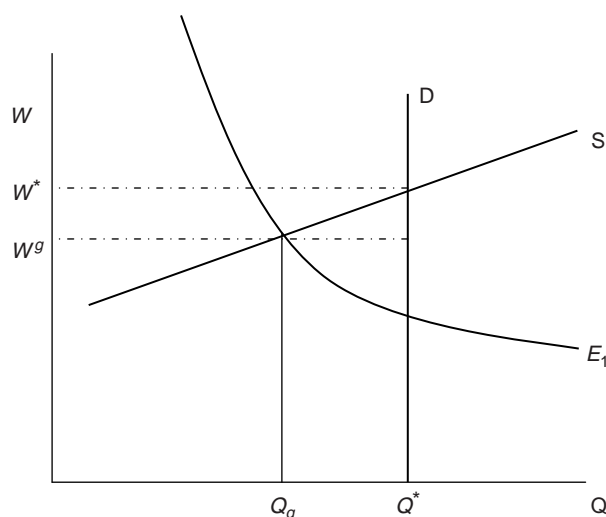


Figure 1 The labor market for teachers.

make teaching more attractive, so that more potential teachers supply their labor at any given wage.

Of course, the above analysis is simplistic in that it treats all teachers as being the same. In reality, within the same country, there may be teacher shortages in particular geographical locations or regions or in particular subjects, with an oversupply elsewhere. In addition, the real market position is very different for primary and secondary school teachers. We can amend **Figure 1** to allow for such possibilities by creating a simple distinction between different kinds of teachers. A simple analysis would suggest that the possibility of differential wages by subject, in different regions or between primary and secondary sectors could be adopted to solve the problems of short supply in particular areas. Whether this solution is actually viable, given the demands of teachers' unions and the rigid teacher-wage structures imposed in most countries, is another question.

Figure 2 shows the teacher demand-and-supply elements that may be used to determine if the teacher labor market is in shortage or in surplus. Demand is dependent of the number of pupils in the country and on the government's desired PTR. The higher the number of pupils enrolled in schools, or the lower is the PTR target set by the government, then higher will be the demand for teachers.

The supply of teachers as outlined in **Figure 2** can be divided into two: the current supply of teachers and the potential supply. The supply of teachers consists of those who are currently in service (*s*) which would include continuing teachers, new entrants (*e*) and reentrants (*rf*). New entrants are first timers teaching in public schools while reentrants are those with previous teaching experience, who left and are now returning to teaching. The number of students enrolled in the initial teacher training

(ITT) courses sustains the flow of new entrants as they can enter into teaching upon completion of their training. A shortage (*ex*) occurs when the demand for teachers is not matched by supply and a surplus occurs when the current supply of teachers exceeds the demand of teachers.

To complete the teacher supply-and-demand model, the outflow of teachers needs to be considered as well. Wastage makes up the outflow of teachers from the current supply. This group of leavers can be divided into those who leave at retirement age and those who leave for reasons other than retirement (i.e., those below the age of 60–65). When teachers (and those who are qualified to teach) leave the profession, they become inactive and enter the stock of potential teachers in the pool of inactive teachers (PIT). In addition to the leavers below retirement age, the PIT also contains the ITT graduates who do not enter into teaching. A second component in the potential supply of teachers is the pool of recoverable teachers (PRT) who are those who have left the profession but can be enticed to return to teaching and are therefore the main source of potential supply.

The Demand for Teachers

The first key element in the demand for teachers is the demographic pattern of pupil numbers which fluctuates all the time with changing fertility patterns. In most OECD countries, projected numbers of pupils is expected to fall over the next 5–10 years.

The second demographic trend affecting the demand for new teachers relates to the age distribution of the stock of existing teachers. **Figure 3** shows that in many OECD countries, the teacher stock is aging as an increasing proportion of teachers in the years 1992, 1996, and 2000 are over 50. This will lead to an increase in the demand for new teachers to replace those retiring over the next 10–15 years.

A third factor in the demand for teachers is the size of the class the government chooses for its pupils. This varies remarkably across countries as **Figure 4** shows that PTRs in primary and lower secondary schools are over 30:1 in some countries and as low as 16:1 in some countries. This will not only condition the demand for teachers but the quality of the teaching which is imparted to the children as larger classes may mitigate against individual pupil attention. In many OECD countries, PTRs have been falling. In the USA, average class size was 23 in 1965 – but this fell to 16 by 2000. This represents a dramatic growth in teacher supply over the last 35 years in the USA. Of course, what these aggregate figures hide are the increasing need for specialist teachers of subjects like information and communications technology (ICT).

A fourth factor in demand is the governments choice of the length of the working day for teachers and how many

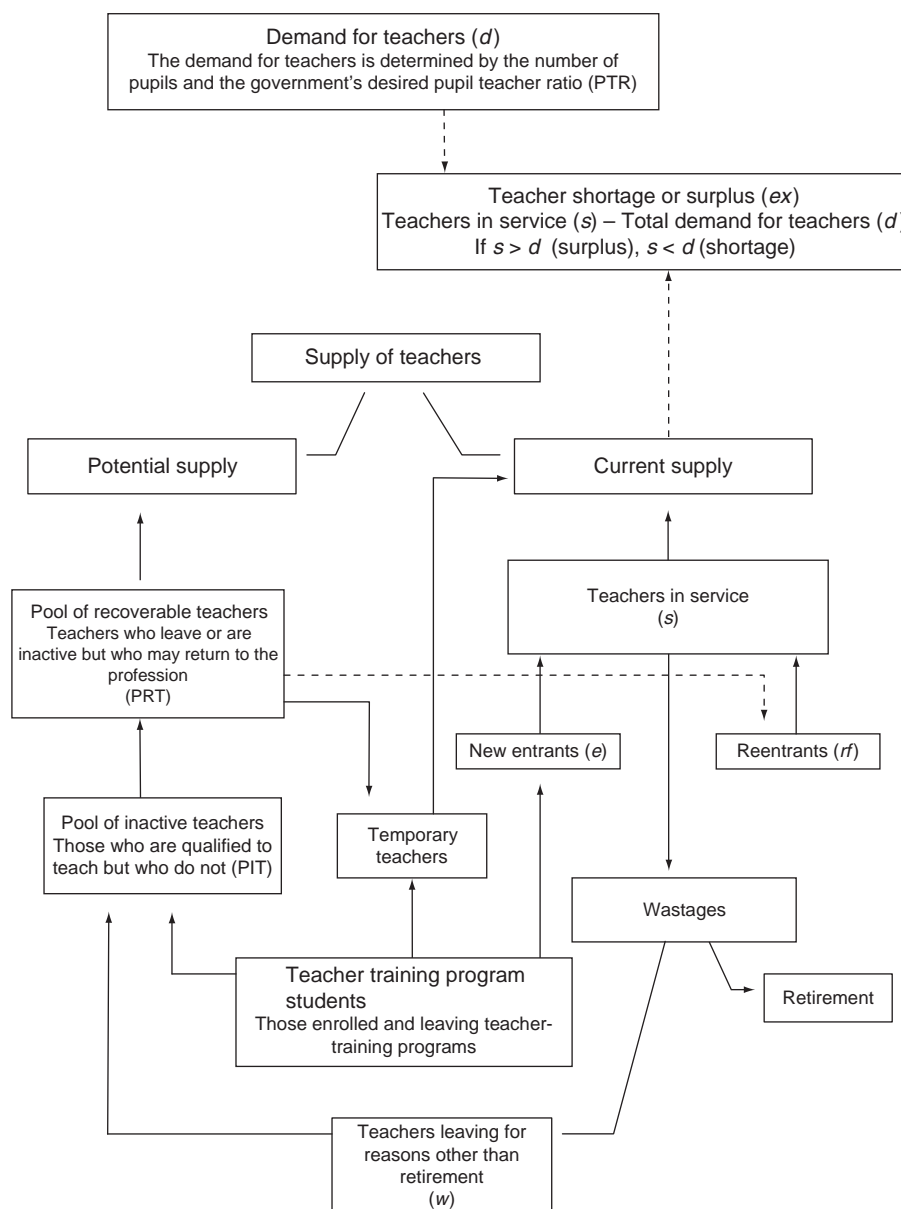


Figure 2 Teacher demand and supply.

teaching days there are in a year. Most countries make their elementary teachers teach between 650 and 800 h in the year but some teach a lot more – including those in the USA. **Figure 5** illustrates the variability of teacher hours in 2004 across countries and these differences must be reflected in the quality and intensity of the effort required of teachers in their job and hence the relative attractiveness of the job compared to alternatives.

Several other features may complicate the demand for teachers in different countries. First, if the financial administration of education is performed at a local level, resourcing decisions will depend on local and school-specific factors. Second, the determination of desired

PTRs and teacher recruitment may be influenced by educational criteria at the state or local level. Third, most governments control the nature of length of teacher training – this can and has been changed in times of crisis. Finally, different countries have different conventions about the extent to which school subjects, like mathematics may be taught by nonspecialists and principals have difficulty recruiting math and science teachers (**Figure 6**). Clearly, allowing nonspecialists to teach such shortage subjects will solve the short-run problem of having a teacher in front of a class – but at what cost to the quality of teaching? Ultimately, the demand for teachers will depend on the political will that creates the policy on

educational expenditure balanced with the importance on spending on health, welfare, defense, and other priorities.

The Supply of Teachers

The supply of teachers can be regarded as all those currently in teaching, plus those currently not teaching, but who are qualified to teach, and would consider teaching if the conditions were right. The supply issues at stake are therefore ones of recruitment and retention, as well as inducing the return of qualified individuals who have left

the profession. There are many factors that are likely to influence the supply of teachers, such as the relative earnings on offer in teaching and other careers, other labor-market opportunities, and varying relative non-pecuniary conditions of work. To a certain extent, some of these factors can be controlled by the government or federal authorities since it can determine how many places are provided on courses at universities to train teachers. In many countries, teacher training courses are not always filled, and attendance varies by subject. The measurement of teacher supply and most specifically the changes in teacher supply from year to year is problematic. Various studies have tried to measure teacher supply by changes in the pool of inactive teachers or the pool of recoverable teachers; changes in the stock of those teachers actually in service; the number of new entrants; or the numbers leaving teaching on teacher-training programs.

It is evident that the flow of newly qualified teachers does not necessarily indicate the level of overall supply. Focusing on those currently working as a teacher ignores individuals who are available for (and possibly seeking) work in teaching, but who are not currently employed as a teacher. Supply can be calculated as consisting of those entering the profession and those remaining in teaching from the previous year.

However, the difficulty is not just recruiting teachers but keeping them in the classroom. Some trainees drop out and others decide not to become teachers. Smithers and Robinson (2003) showed that only 88% of registered trainees passed the final examination, and only 59% were

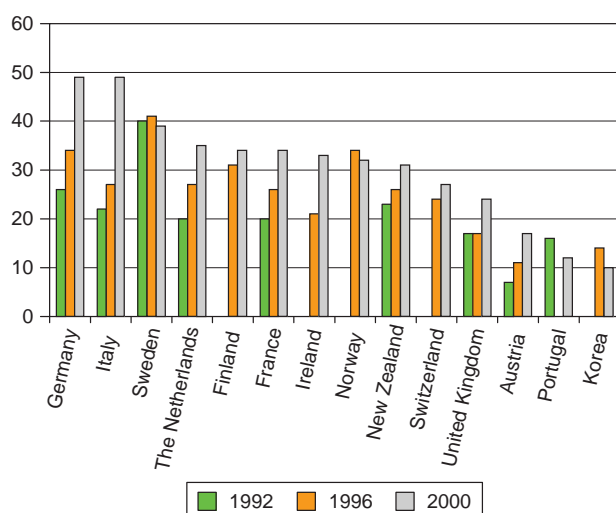


Figure 3 The percentage of teachers aged 50 years and over, lower secondary education.

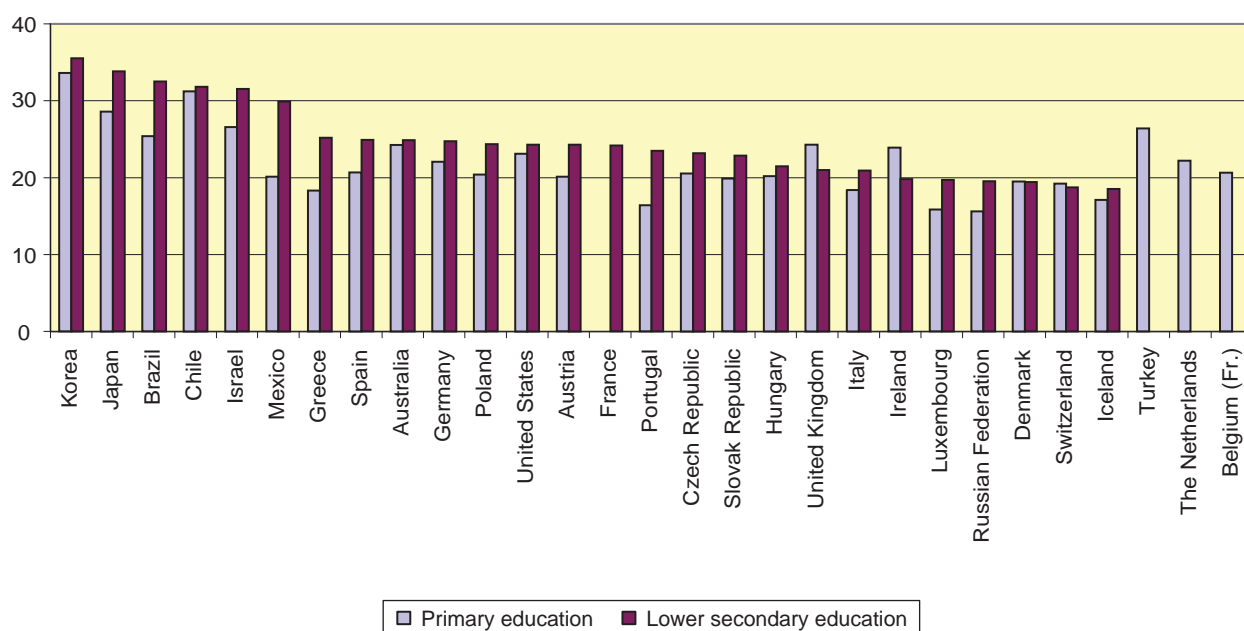


Figure 4 Average class size in educational institutions (public institutions only) by level of education (2004). Countries are ranked in descending order of average class size in lower secondary education.

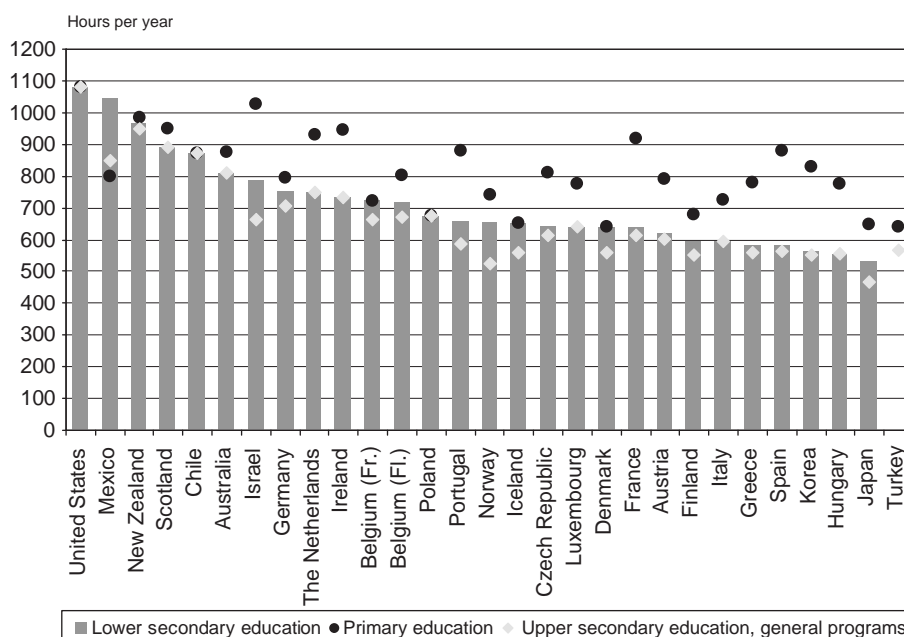


Figure 5 Number of teaching hours per year, by level of education 2004. Net contact time in hours per year in public institutions. From Education at a Glance (2006) Chart D4.2. Statlink <http://dx.doi.org/10.1787/421472785265D>.

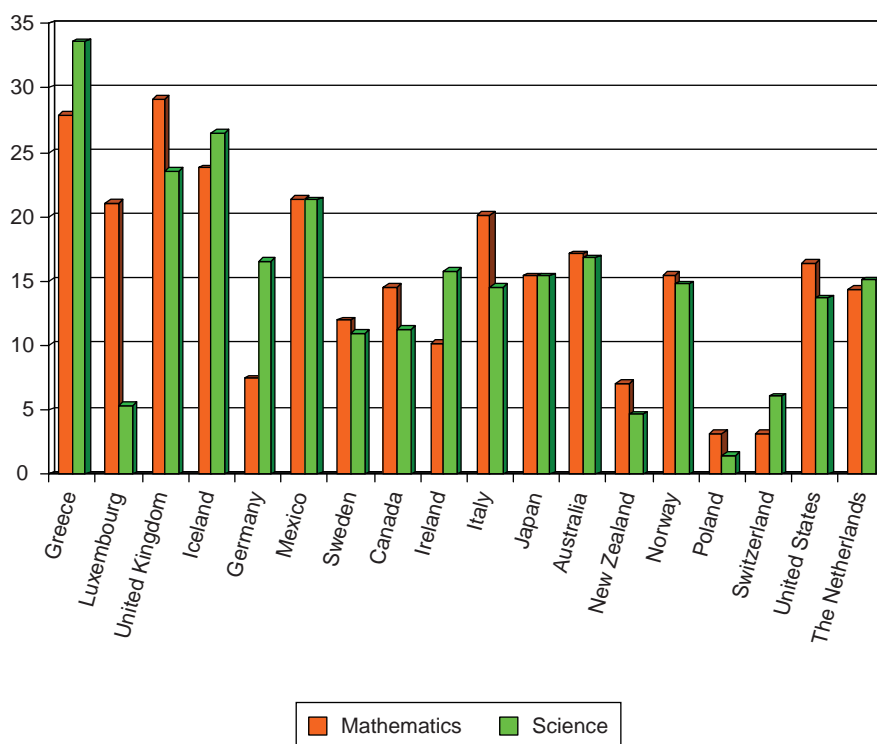


Figure 6 Percentage of 15-year-old students enrolled in schools where principals report learning is hindered to some extent or a lot by a shortage/inadequacy of teachers. From OECD PISA Database 2001.

teaching a year later. After 3 years, only 53% of the original trainees were still in the classroom. This wastage not only adds to the costs of providing teacher training but also has negative effects on child performance. The evidence is that

higher teacher turnover is associated with lower educational pupil outcomes (Dolton and Newson, 2003). This is of particular concern since we know that teacher turnover is highest in the most deprived areas.

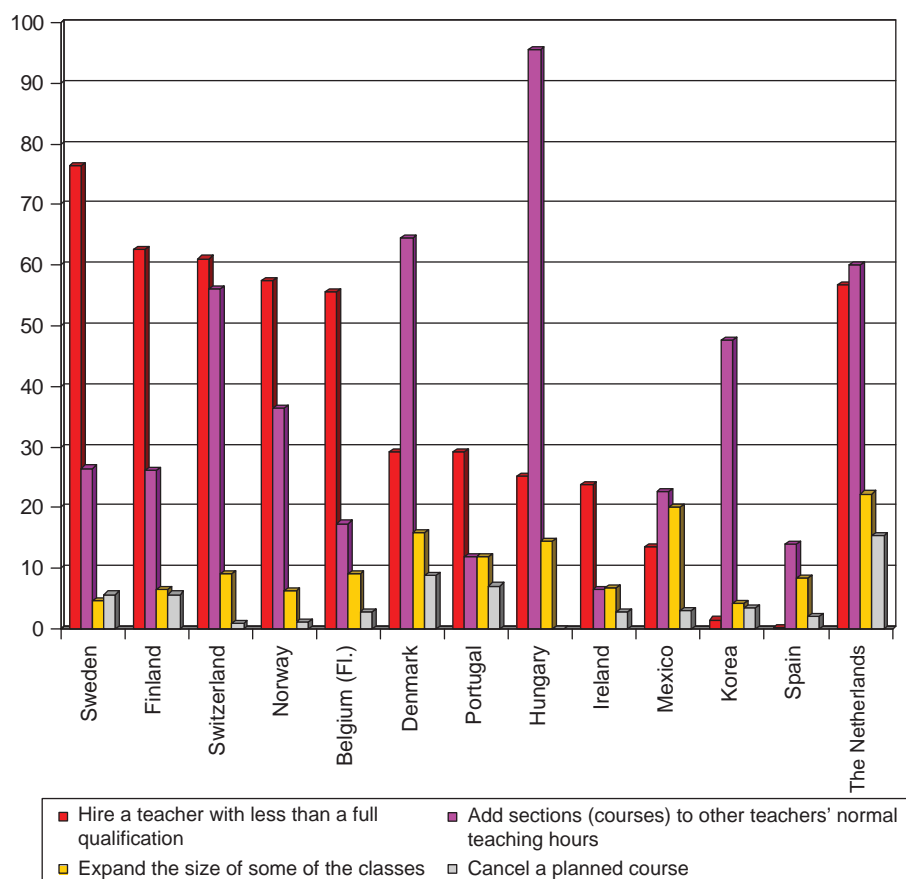


Figure 7 Percentage of upper secondary students attending schools which use the following methods to respond to teacher vacancies, as reported by school principals. From OECD International Survey of Upper Secondary Schools (ISUSS) database, 2003.

Policies to increase the retention of trainees and new teachers have been on the forefront of the political agenda on education. The most prominent measures are repayment of student loans for up to 10 years and a hardship allowance for students in shortage subjects committing to become teachers, bursaries for undertaking and completing teacher training, and cash golden hellos for new teachers in shortage subjects.

It is hard to find direct evidence of the interrelation between teacher supply and shortage and the quality of teachers. One limited piece of evidence is that reported by the OECD (2004). **Figure 7** comes from the OCED website. It shows how schools react when faced with a teacher shortage. In the usual course of events, we would expect a school to advertise and hire a fully qualified teacher – if teachers are in short supply the school may resort to hiring a teacher with less than a full qualification; expanding the size of the classes; add hours to other teachers' courses; or cancel the planned course. **Figure 7** shows that these practices are frequently adopted in different countries. This must reflect on the quality of the teachers appointed and so it provides some evidence of the link between teacher shortage and teacher quality.

What individuals can earn as teachers, relative to what they could earn in alternative occupations (the opportunity wage), is one of the key determinants of the decision to become a teacher. In particular, the lower the relative wages (or wage growth) are in teaching, the less likely a graduate is to choose that career. Relative earnings affect both initial career choices, as well as choices made later in an individual's career. Dolton (1990) also found that there is considerable inertia to remain in teaching, and suggested that this effect may be partially due to the different individuals' subjective evaluation of the relative pecuniary and nonpecuniary rewards to teaching. It is also likely that nonpecuniary factors such as workload, job stress, physical surroundings, materials and resources, parental support, administrative support, access to further training, and related factors also play an important role in the decision to enter teaching or to leave if these working conditions change (see Hanushek and Rivkin, 2007; Boyd *et al.*, 2007).

While relative pay affects the decision to become a teacher, it also affects the decision to remain a teacher. Analyzing the decision to leave teaching, Dolton and Van der Klaauw (1995, 1999) show that higher the relative earnings of teachers, the less likely they are to leave teaching. Work using US data suggests that raising teacher

pay could improve the quality of the stock of teachers. But attracting more able students to teaching is not the only difficulty for policymakers. Since individuals with higher ability generally command higher wages, high-ability teachers are at a higher risk of leaving the profession than less-talented teachers. To negate the lure of improved outside opportunities on able-teacher retention, some countries have introduced fast-track programs with the aim of recruiting and retaining the most able graduates by shortening payscales and providing them with additional training, support, and supervision.

Another important aspect of teacher supply is that teaching is a career that is relatively popular with female graduates. In nearly all OECD countries, the majority of teachers are women particularly in primary education, where women consist of 80% or over of the teacher labor force. A crucial aspect of the distinction between male and female occupational choice is that often women are simultaneously making decisions about starting a family and hence deciding whether to participate in the labor market. This is particularly true in teaching since a teaching career has complementarities with family formation and in particular, the ease with which one can return to teaching after a career interruption. Hence, for women, the choice of teaching as a career is intimately related to the decision to participate in the labor market (Dolton and Makepeace, 1993).

Labor-market conditions at the time the occupational choice is made are also important. The most recent evidence from Dolton *et al.* (2003) look at time series data in

the UK and finds that aggregate labor market conditions, particularly unemployment levels, are important determinants of teacher supply. Notably, they find that the supply of graduates to teaching is countercyclical with most graduates' perception of teaching (and willingness to enter the profession) improving when teacher pay is high compared to alternative occupations and when graduate unemployment is high.

Teacher quality is extremely difficult to observe. That is to say – it is very difficult to accurately measure the amount of effort exerted or output produced by any individual teacher. This is important because teacher effectiveness is an important determinant of pupil attainment. It is unclear whether teachers with better qualifications are necessarily better teachers. There is some evidence in the UK (see Chevalier *et al.*, 2007) that teachers are being drawn from a lower part of the educational achievement or ability distribution than they were in the past.

What we seek to do is recruit the individuals with the most appropriate personalities to be good teachers. To some extent, this does not necessarily mean the most able individuals – but those with the capacities to be good teachers. What makes a good teacher is not something which can be easily measured. Naturally governments and state agencies should try to maximize the effective use of the taxpayers' money spent on education – this means that there should be appropriate attempts to monitor spending on teachers. The problem with this is that we do not fully understand: how to measure teacher quality, how teacher

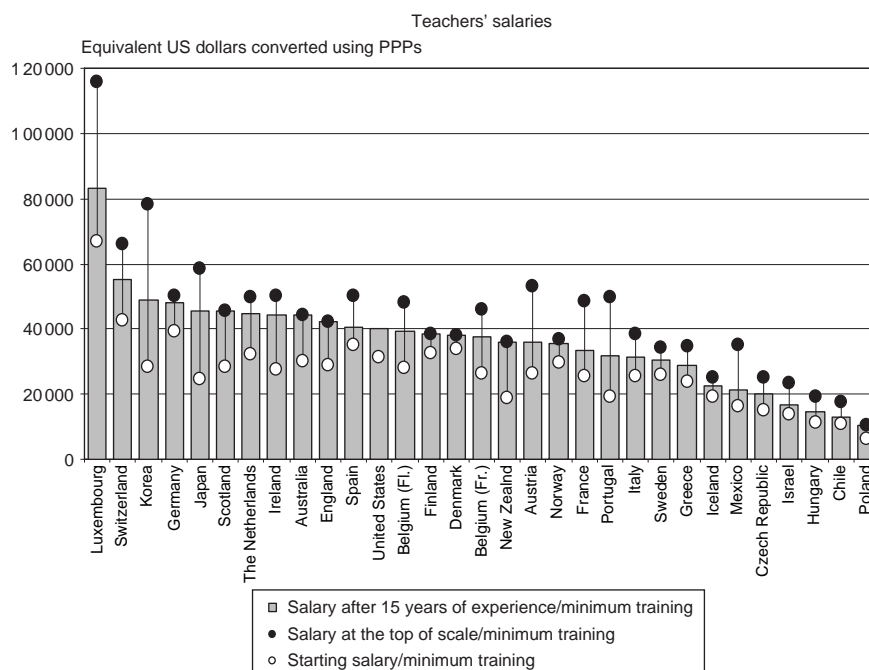


Figure 8 Teachers' salaries (minimum, after 15 years experience, and maximum) in lower secondary education (2004). From Education at a Glance, OECD (2006) Chart 3.2. Statlink. <http://dx.doi.org/10.1787/083407611234>.

quality affects pupil performance, and how to measure effective teacher input or effort. Under these circumstances it is tempting for governments to try to introduce various incentive mechanisms and even performance-related pay for teachers. See Lavy (2004) and Dolton *et al.* (2003b).

Teachers' Pay

The most important determinant of teacher supply is the relative wage on offer. There is a large body of econometric evidence which finds that relative earnings of teachers is a major factor in an individual's decisions to become a teacher or remain in the profession. **Figure 8** shows how variable teachers' wages are with 15-years experience, across countries by graphing average teacher wage expressed in US dollars using purchasing power parity to facilitate comparison. This shows that there are some countries where teachers are earning approximately 4–8 times (Luxembourg and Switzerland) that in other countries (Poland and Chile). Most countries pay their teachers between 1.3 and 1.7 times the gross domestic product (GDP) per head, including the UK, USA, and most European countries. Clearly, in some countries being a teacher accords a much higher-earning status than in others. This reflects on the caliber of the people doing the job in that country.

Figure 8 also shows the range of salaries between the minimum starting salary and that at the top of the scale. This gap is largest in countries which pay their teachers more.

It is also of concern that the relative wage of teachers in many OECD countries has been falling in recent years. **Figure 9** shows the ratio of an experienced teacher

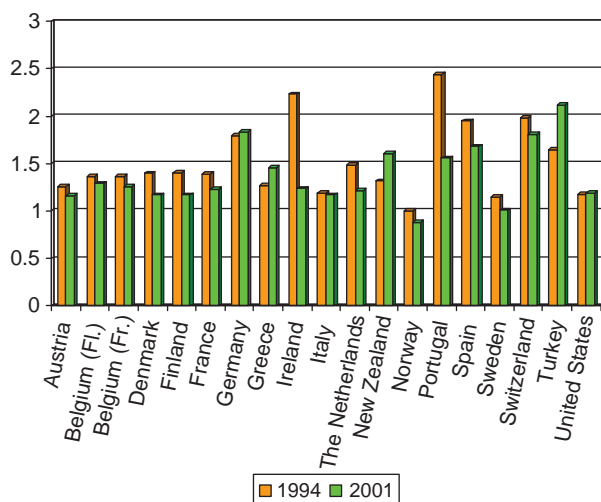


Figure 9 Ratio of salary after 15 years of experience to gross domestic product (GDP) per capita, public institutions, lower secondary education, 1994 and 2001.

with 15 years of service in 1994 and 2001. For many countries their wage – expressed as a proportion of GDP per head has been falling. Notable exceptions are Greece, New Zealand, and Turkey.

Tracking the relative pay of teachers over time in more detail is not straightforward in most countries. Dolton (2006) presents what has happened to relative teacher earnings over time in the UK and the USA. Often the process of public sector wage settlement is subject to delays and lags. This is reflected in the process of declining real wages in teaching which may then be followed, as was the case in the UK, with periods of catch up following a decline in relative earnings.

Another important but neglected aspect of remuneration is what individuals are paid over their working career lifetime. There is good evidence for the UK (Dolton and Chung, 2004) that the position of teachers has been getting worse. The earnings of male teachers were uniformly higher than earnings in the alternative occupation (for qualified teachers who choose not to teach) in 1975. However, over time, the earnings profile in the alternative occupation has been shifting up while that of teachers has been moving down.

There are many other areas in which our knowledge about teacher supply is scant. Although teacher shortage is widespread, we know from empirical evidence that many young people still choose to be teachers despite the low pay. Clearly, nonpecuniary factors matter in occupational choice but we know relatively little about this part of the process and how it operates. We know relatively little about the role of expected future lifetime earnings in the decision process and how people trade this off against a job which is rewarding. We also know little about young people's perceptions of teaching as a career or their real relative perceptions of earnings in teaching. In some countries, there are a sizable number of teachers who enter the profession in mid-career after working in the private sector for some time. It would be useful to know what motivates these individuals to make these choices and whether they are more effective teachers due to their outside perspective on the real world.

Summary

Examining the pattern of teacher supply across OECD countries, it is clear that there have been various trends. Teacher relative earnings have been declining and many countries have experienced teacher shortages particularly in subjects where graduates can earn a higher-opportunity wage. Countries can disguise these shortages in a number of ways including having more pupils taught by nonspecialist teachers. These factors might mitigate the quality of teachers.

Many governments have tackled their problems with teachers by reforming the training of new teachers or the retraining of existing teachers with professional-development schemes or trying to introduce teacher appraisal and various incentive schemes into teacher pay. Little is known about the effectiveness of these schemes compared to simply putting more government expenditure into education and teacher pay.

The overwhelming conclusion of research on the supply of teachers is that problems of teacher shortage could be alleviated with higher relative teacher salaries. If relative teacher salaries were higher, then the problems of teacher recruitment, retention, and duration could be alleviated.

It is clear that in most countries there are teacher-supply problems. There is evidence from different states in the USA and from other countries around the world that there are persistent episodic shortages of teachers. This manifests itself in terms of not enough recruits entering the profession and too many leaving it prematurely. However, it is also clear that there is not a universal shortage of all categories of teachers. Specifically, teachers are in short supply in difficult schools in areas with inner-city urban problems and in subjects which have a high-opportunity wages for those with specific technical skills. The problems seem to be worst in scientific and mathematical subjects. Hence, the real challenge of teacher supply is to get teachers into teaching the subjects and areas that are not appealing. The straightforward economic answer is to consider differential pay for those shortage categories. The problem in most countries is that there is a rigid wage structure in teaching. This limits the ability of school managers to effectively respond to relative wage differentials or target wage increases in difficult to recruit specialties. It is also unclear what the long-term consequences are of having children taught with temporary, nonspecialist, or low-quality teachers.

The perennial challenge for governments and education administrators is to establish a high-quality teacher labor force which is hard working and effective. Such problems could easily be solved by paying teachers higher wages. The problem, of course, is that governments are reluctant to throw more money at such problems and that even if they did it, is not clear what the outcome of increased expenditure is – precisely. Hence, more research on the relationship between educational-resource inputs, teacher quality, and outcomes must inevitably have important lessons for teacher supply. For example, if we knew categorically that the PTRs had no effect on pupil outcomes, we could simply solve teacher-shortage problems by having larger classes. In addition, we do not know in much detail, the effect of teacher working conditions on

their relative effectiveness. There are many unanswered questions relating to teacher supply.

Bibliography

- Boyd, D., Goldhaber, D., Lankford, H., and Wyckoff, J. (2007). The effect of certification and preparation on teacher quality. *Future of Children* **17**(1), 45–68. Princeton, NJ: Brookings Institution.
- Chevalier, A., Dolton, P., and McIntosh, S. (2007). Recruiting and retaining teachers in the UK: An analysis of graduate occupation choice from the 1960s to the 1990s. *Economica* **74**, 69–96.
- Darling-Hammond, L. (2000). Solving the dilemmas of teacher supply, demand and standards: How we can ensure a competent, caring and qualified teacher for every child. *National Commission on Teaching and America's Future*.
- Dolton, P. (1990). The economics of UK teacher supply: The graduate's decision. *Economic Journal* **100**, 91–104.
- Dolton, P. (2006). Teacher supply. In Hanushek, E. and Welch, F. (eds.) *Handbook of Education Economics*, ch. 18, pp 1053–1080. North Holland.
- Dolton, P. and Chung, T. -P. (2004). The rate of return to teaching: How does it compare to other graduate jobs? *National Institute Economic Review* **190**, 89–103.
- Dolton, P. and Makepeace, G. (1993). Female labour force participation and the choice of occupation: The supply of teachers. *European Economic Review* **37**, 1393–1411.
- Dolton, P., McIntosh, S., and Chevalier, A. (2003b). Teacher pay and performance: A review of the literature. *Bedford Way Papers*. London: Institute of Education.
- Dolton, P. and Newson, D. (2003). The relationship between teacher turnover and pupil performance. *London Review of Education* **1**(2), 133–140.
- Dolton, P., Tremayne, A., and Chung, T. -P. (2003). Teacher supply and the economic cycle. *Report to the OECD*.
- Dolton, P. and Van der Klaauw, W. (1995). Leaving teaching in the UK: A duration analysis. *Economic Journal* **105**, 431–444.
- Dolton, P. and Van der Klaauw, W. (1999). The turnover of UK teachers: A competing risks explanation. *Review of Economics and Statistics* **81**(3), 543–552.
- Hanushek, E. and Rivkin, S. (2007). *Pay, working conditions and teacher quality*. *Future of Children* **17**(1), 45–68. Princeton, NJ: Brookings Institution.
- Lavy, V. (2004). Performance pay and teacher's effort, productivity and grading ethics. *Working Paper 10622*, NBER.
- OECD (2004). The quality of the teacher workforce. *Policy Brief*. Paris: OECD Observer.
- Smithers, A. and Robinson, P. (2003). *Factors Affecting Teachers' Decisions to Leave the Profession*. Department for Education and Skills, RR430.

Further Reading

- Brookings Institution (2007). Excellence in the classroom. *The Future of Children* **17**(1). Princeton, NJ: Brookings.
- Duflo, E. and Hanna, R. (2005). Monitoring works: Getting teachers to come to school. *NBER Working Paper 11880*.
- Hanushek, E. and Rivkin, S. (2006). Teacher quality. In Hanushek, E. and Welch, F. (eds.) *Handbook of Education Economics*, ch. 19, pp 1081–1164. Amsterdam: North Holland.
- Murnane, R. J., Singer, J. D., Willett, J. B., Kemple, J. J., and Olsen, R. J. (1991). *Who Will Teach? The Policies That Matter*. Cambridge: Harvard University Press.
- OECD (2002). *Teacher Demand and Supply: Improving Teaching Quality and Addressing Teacher Shortages*. Paris: OECD.
- OECD (2005). *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*. Paris: OECD.

Teacher Training and Preparation in the United States

M Podgursky, University of Missouri – Columbia, Columbia, MO, USA

M G Springer and R Hutton, Peabody College of Vanderbilt University, Nashville, TN, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This article provides a brief overview of the current system of teacher preparation and training in the United States as well as a review of empirical research on the topic. It focuses on public elementary and secondary education in the United States, although some reference is made to international literature on teacher preparation and training. The review of relevant research focuses on study designs that employ randomization or control for prior achievement of students since studies without these features are likely unable to disentangle the impact of certification policies or programs from other factors that may be correlated with certification policies and programs.

To become a teacher in a public elementary or secondary school in the United States, it is usually necessary to obtain a state teaching license. This license is meant as an assurance to the public that a teacher has met a certain minimum standard of proficiency. Outside of K-12 education, licensing is commonly distinguished from certification. The former means that a practitioner cannot legally practice his profession or trade without a state-issued license. The latter refers to a situation in which a government agency, or, more commonly, a private body, certifies the skills of practitioner, but the government does not exclude noncertified practitioners from the market (Friedman, 1962). To further complicate matters, in K-12 education, endorsements refer to certification in a subspecialty. For purposes of this essay, we will treat license, certificate, and endorsement in K-12 education as synonymous.

When it is perceived that quality is low, some proponents have argued that teacher training and licensing should be the focus of reform efforts. The National Commission on Teaching and America's Future (NCTAF), the American Association of Colleges of Teacher Education (AACTE), and teacher unions have proposed reforms that would create national teacher testing standards, universal accreditation of teacher-training programs, creation of independent teacher licensing boards, and more entry restrictions into the teaching profession (NCTAF, 1996, 1997; Ballou and Podgursky, 2001). More recently, the federal No Child Left Behind (NCLB) act has devoted considerable attention to the topic of teacher training and licensing as a means of quality assurance, while further requiring states to phase out emergency, alternative, and other substandard types of teacher licenses so that all teachers are 'highly qualified' in their primary teaching

area. The US Department of Education has also funded development of a National Comprehensive Center for Teacher Quality to serve as a national resource on strategies for strengthening the quality of teaching and the teaching workforce.

Teacher training and licensing as a means of quality assurance assumes that how a teacher is trained and what type of license a teacher holds is an informative indicator of their classroom effectiveness and that higher standards in training and licensing will yield positive net benefits on student learning. Critics argue that teacher licensing, based on the currently proposed criteria, has not shown any significant impact in the short or long term on student achievement and other schooling outcomes. Moreover, these reforms may have significant costs in the form of expanded resources for teacher-training programs, more time spent fulfilling teacher-preparation program requirements, and, as a consequence, a reduction in the size of the teacher applicant pool.

Many critics further argue such reforms may also eliminate, or substantially downsize fast-track, alternative-route teacher-training programs, such as Teach for America and Troops to Teachers. Teach for America (TFA) recruits recent college graduates from top institutions of higher education across the country to commit 2 years to teach in urban and rural public schools and provides them with training and professional development throughout their 2 years of service. Troops to Teachers is a United States Department of Education and Department of Defense program that assists eligible military personnel with starting a career as a teacher in public elementary and secondary schools. If local decision makers have good information about the effectiveness of a teacher aside from their observed credentials (e.g., student teaching ratings, teaching portfolio, or recommendations), it may be that noncertified (or alternately certified) applicants are the more qualified candidate for a position (Goldhaber, 2004). Strict licensing requirements typically prevent a school district from ever hiring an uncertified candidate, if a fully certified candidate is available on the job market. Thus, an important part of the research agenda in this area is how well certification, or its component parts (e.g., tests or specialized teacher training), predict teacher effectiveness and whether more restrictive licensing requirements affect the teacher applicant pool and the quality of new hires.

The next section of this article provides a brief overview of the current system of teacher licensing in the

United States and then reviews the empirical literature on teacher training, preparation, and certification process.

Current Licensing Methods

There is a great deal of variation in licensing requirements from state to state, although there is a certain amount of commonality across particular components. All states require that new teachers possess a bachelor's degree; however, some states require an education degree from an accredited teacher-training program. In other states, teachers complete education courses while majoring in an academic discipline, or obtain a master's degree in education following completion of undergraduate coursework (NASDTEC, 2008; Education Week, 2008). Regardless of the path toward licensure, most teachers complete some education coursework in addition to student teaching before they have satisfied the professional requirements for permanent licensure.

In the majority of states, the power to license teachers is governed by the state board of education or state education agency. It has been proposed that this authority shift to independent professional boards such as those found in medicine and some states have enacted such certification reforms (NCTAE, 1997). However, unlike medicine or law, many approved teacher-training programs are not accredited by the principal private accrediting agency, the National Council for Accreditation of Teacher Education (NCATE). Few states mandate NCATE accreditation, although some have reached agreements whereby NCATE participates in the state's review of its training programs. In response to growing pressure from states for external accreditation of teacher-training programs, and concern by many private and research-oriented public universities with the NCATE approval process, a second accrediting body, the Teacher Education Accrediting Council (TEAC), was founded in 1997. TEAC's membership base is much smaller than that of NCATE, but is expanding.

Most states require teachers to pass at least one standardized assessment before they are admitted into a training program or granted a permanent license. These examinations fall into one of four categories: basic skills, general knowledge, pedagogical knowledge, and teaching field content knowledge. How many and which kind of examination a prospective teacher must take varies from state to state. NCLB seems to have stimulated the use of standardized teacher assessments through its requirement that teachers show evidence that they are highly qualified in their primary teaching field (ETS, 2007). Teacher performance on a standardized assessment is believed to offer proof of being highly qualified.

However, the emphasis that state and federal policies place on this supposition is far greater than one might expect given the current state of knowledge. The literature

on the relationship between a teacher's performance on a standardized test and their students' academic achievement detects a small, positive relationship of about 1–4% of a standard deviation (Goldhaber, 2007; Clotfelter *et al.*, 2007). It is unclear if standardized assessments offer a good screening tool for weeding out low-quality teachers or a sound signaling device for identifying high-quality teachers (Goldhaber, 2007; Podgursky, 2004).

One difference between licensing for teachers and other professions concerns the large number of teaching licenses. In medicine, law, dentistry, nursing, and other licensing professions, the state issues but one license. While many professionals, particularly in medicine, go on to earn certificates in highly specialized areas, these are not issued or required by states to practice. By contrast, the typical state education authority issues literally dozens of specialized teaching licenses (e.g., elementary education, secondary chemistry, secondary language arts, drama, various types of special education, and scores of different vocational licenses). This finely parsed system makes it difficult for a school, district, or state to make sure that every teacher in every classroom has the right certificate for that class. For example, if the certified chemistry instructor teaches an algebra class for which she is not certified, that teacher is out of compliance by teaching algebra. This problem has been exacerbated with the creation of middle schools which are typically defined by some combination of grades 5 through 8. Many middle school teachers hold elementary licenses and not newly created middle school licenses by specialty field, thus forcing them to complete additional training to remain eligible to instruct at the middle school level. Dee and Cohodes (2008) report that approximately 20% of mathematics and science teachers do not hold a certificate in their teaching field.

Research on Teacher Training and Licensing

Randomized experiments are widely recognized as the gold standard in social science research for determining causal relationships. By randomly assigning individuals or entire organizations to one or more treatment or a control group, unobserved factors that could lead to systematic differences between treatment and comparison conditions when estimating the relative effectiveness of the intervention are removed. The Institute for Education Sciences, the research funding arm of the United States Department of Education, is actively promoting randomized study designs in all areas of education research in light of the fact they have been underused in education research (Mosteller and Boruch, 2002).

With respect to teacher quality research, the most practical experimental design involves estimating the

effect of teachers with different credentials or training on an outcome of interest (typically student achievement) by randomly assigning students to classrooms of variously credentialed but otherwise comparable (e.g., experience) teachers within a school. Only one published study of teacher training employs a randomized design. Glazerman *et al.* (2006) compared the effect of TFA recruits with other new and experienced traditionally certified teachers within the same school at several sites around the United States. Students were randomly assigned to teachers participating in the study to maintain equivalence between TFA and non-TFA teachers on all factors believed to affect student achievement. TFA teachers were significantly more effective than other new recruits in math and equally effective in English. In substantive terms, Glazerman and colleagues concluded that the impact of TFA teachers on math scores was equivalent to about 1 month of instruction more than the impact of non-TFA teachers on math scores.

When randomization is not possible, and often it is not in education research, then researchers must rely on non-experimental or observational data to evaluate the effectiveness of an education policy intervention. In measuring the contribution of a classroom teacher on student achievement, it is necessary to control for prior achievement of the student before he or she enters the classroom. For example, researchers might pretest the students in the fall and test them again in the spring. The difference in fall and spring test scores, averaged over the classroom, would be a measure of a teacher's value-added. If students are not pretested in the fall, then it is also possible to use test scores from the previous spring, or for more than one previous year if longitudinal achievement data are available. Indeed, large longitudinal data files form the basis for the most sophisticated current research on teachers and teacher effects on student achievement (e.g., Sanders and Horn, 1994; Rivkin *et al.*, 2005; Boyd *et al.*, 2006; Aronson *et al.*, 2007). An excellent technical overview of the methodology of these education production function studies may be found in Todd and Wolpin (2003).

Even with longitudinal achievement data, there is still the potential for biased estimates of the effect of teacher training or licensing. For example, certain teachers may get assigned particularly difficult classrooms or students (Rothstein, 2008). These assignments may not be fully captured by prior student achievement or other covariates in large longitudinal data sets. Teachers who choose to participate in professional development programs may be more highly motivated and better teachers in the absence of the program. Again, these are variables that are not observed by the researcher or measured in data. Thus, researchers search for natural experiments, that is, situations in which assignment to a particular treatment is not subject to self-selection, and approximates a true experiment. Jacob and Lefgren (2004) exploited an arbitrary cutoff which assigned teachers to professional development to

estimate the effect of these training investments. They find that modest investments in teacher training do not increase the achievement of elementary school children in mathematics or reading.

Studies that do not have a rigorous study design, that is, with randomization or controls for prior student achievement, are likely to produce biased estimates of the effect of teacher certification or other teacher characteristics on student achievement. One reason is that these studies do not adequately control for the socioeconomic status (SES) of students in classrooms, and student SES is correlated with teacher credentials and strongly correlated with student achievement. In the language of econometrics, such cross-sectional studies of the effect of teacher credentials on student achievement suffer from omitted variable bias. Thus, our review of relevant research focuses on study designs that employ randomization or control for prior achievement of students.

It is also important to note that substandard certification tends to be relatively more common in schools with greater percentages of low SES students. Since SES has a very powerful effect on both student achievement levels and gains, unless the researcher has a very comprehensive set of controls for prior achievement and SES in a study of certification and student achievement, the resulting study is likely to yield an upward-biased estimate of the effect of certification. While this article discusses many studies that control for prior achievement, it is important that a reader remembers this study design is inferior to a study that carefully employs randomization.

The number of studies of teacher certification that meet the minimum methodological standards outlined above is limited. A survey of the literature by Wayne and Youngs (2003) found only two studies of teacher certification that were peer-reviewed, used longitudinal student-level achievement data, and controlled for student SES. Both studies were conducted by Goldhaber and Brewer (1997, 2000) and draw on the same nationally representative data provided by the National Education Longitudinal Study of 1988 (NELS:88). Their 1997 study concluded that students in tenth grade assigned to a teacher who held a mathematics teaching certificate, or earned a bachelor's or master's degree in mathematics, had higher test score gains than those students assigned to teachers with no mathematics certification.

The second study draws on a subset of students surveyed in the spring of tenth grade and then again in 12th grade to examine whether 12th-grade students assigned to teachers with probationary certification, emergency certification, private school certification, or no certification in their subject area have similar test score gains as those students assigned to teachers with traditional certification in their subject area. Goldhaber and Brewer (2000) conclude that 12th-grade students enrolled in a classroom taught by a traditionally certified teacher have larger test score gains

in mathematics than 12th graders taught by a teacher holding either a private school certification or no certification in mathematics. Interestingly, students enrolled in mathematics or science classrooms taught by a teacher holding emergency certification perform similar to those students taught by a traditionally certified teacher.

Darling-Hammond *et al.* (2001) criticize Goldhaber and Brewer's argument that standard certification should not be required of all teachers, since it does not appear to be a significant predictor of teacher effectiveness. However, in a separate survey of the literature focusing on high-quality studies that meet standards similar to those described above, Hanushek and Rivkin (2004) find little evidence linking teacher credentials to student achievement. For example, of nine estimates of the effect of teacher test scores on student achievement, six found no statistically significant effect. Of the three studies finding a significant effect, two were positive and one was negative.

Only a few studies that examine the impact of teacher certification on student achievement have appeared in peer-reviewed scholarly journals since Wayne and Youngs' (2003) and Hanushek and Rivkin's (2004) influential reviews. Boyd *et al.* (2006) examined whether teachers who enter the teaching profession through different pathways demonstrate greater effectiveness at improving student test scores when compared to those teachers who obtained a traditional teacher license. Alternative pathways included TFA, the New York City Teaching Fellows program, and the Teaching Opportunity Program. Using data on students and teachers in third through eighth grade from New York City school system, Boyd and colleagues find that students enrolled in classrooms taught by a TFA teacher had similar gains in mathematics and English language arts when compared to student test score gains of students enrolled in a classroom taught by a traditionally certified teacher. Moreover, students enrolled in a classroom taught by a New York City Teaching Fellow teacher or a teacher entering through the Teaching Opportunity Program had slightly lower test score gains. The authors note that the magnitude of these differences is relatively small. However, there is considerable variation in teacher effectiveness between classrooms – a point we return to later in this article.

Croninger *et al.* (2007) used data from the Early Childhood Longitudinal Study (ECLS), a national data set on children's status at birth and at various points after birth, to examine the relationship between characteristics of elementary school teachers and first-grade student achievement in mathematics and reading. Estimates do not indicate a statistically significant relationship between a teacher's certification status and their students' performance in reading or mathematics. Students taught by teachers with an elementary education degree, however, demonstrate greater reading test score gains than students taught by teachers with an early childhood education degree.

Dee and Cohodes (2008) draw on data from NELS:88 to extend earlier work by Goldhaber and Brewer. The authors emphasize that a student's assignment to a teacher with a particular teaching license may be correlated with the unobserved student effects that influence educational outcomes. They implement a student fixed-effect approach to condition on unobserved time-invariant characteristics of students which may be correlated with a student's ability and/or their probability of being enrolled in a classroom taught by a teacher with a particular teaching certificate. Findings suggest that assignment to a state-certified mathematics teacher increases an eighth-grade student's test score by about 10%, except for those students at the very bottom of the test score distribution where no effect is detected. Assignment to a state-certified social studies teacher for students across the entire ability distribution increased the probability an eighth-grade student had a high test score. Alternative specifications further indicated that state-certified teachers had the largest impact on male and high SES students and those enrolled in large, urban schools.

A recent, large-scale longitudinal study of Florida students by Harris and Sass (2007) was unusual in that the researchers had access to fairly extensive data on the educational and training background of Florida public school teachers. The sample included the census of teachers and students in grades 3 through 10 during the 1994–1995 to 2003–2004 school years. Harris and Sass find a consistent, positive correlation between content-focused in-service professional development and effectiveness of middle school math teachers. Nonetheless, they conclude that teacher training generally has little relationship to teacher productivity as most other relationships tested in the study were insignificant.

While the effect of measured teacher characteristics on student achievement is often small and statistically insignificant, one consistent finding is that there seems to be considerable variation in teacher effectiveness between classrooms. When we compare the effect on student learning of the top and bottom 20% of teachers ranked by performance, the effect is often quite substantial. For example, Hanushek and Rivkin (2004) report that if a student encountered five above-average teachers in a row, it could overcome the deficit typically reported between economically disadvantaged and higher-income students. However, these teacher effects are largely unrelated to traditional measures of teacher quality, such as licensing exam test scores, certification credentials, experience, or graduate degrees, a result highlighted in a survey of the literature by Goldhaber (2002).

A recent study of Chicago public school teachers by Aaronson *et al.* (2007) illustrates many of these points very well. Like other such studies, this work is based on a very large longitudinal data file of linked student achievement scores. This study is unique in that the authors also have extensive administrative data on teacher characteristics

that are unavailable in many other studies, including education, experience, types of teaching licenses, and selectivity of the teacher's undergraduate college. Aaronson and colleagues find that over 90% of teacher effects are not explained by any measured teacher characteristics. That is, characteristics of teachers that are not easily observable in their data explain the vast majority of teacher effectiveness.

While there are many descriptive studies of teacher training and licensing in other countries (e.g., conducted by the World Bank and Organization for Economic Cooperation and Development (OECD)), the number of rigorous evaluations using randomization or longitudinal designs with treatment and control groups is limited. However, given the growing evidence on student achievement and economic growth, and the availability of national and international data sets on student achievement in math and science, the limited number is likely to change (e.g., Angrist and Lavy, 2001; Woessman, 2003).

Conclusion

A new generation of studies based on large longitudinal data files on students and teachers finds large variation in the effectiveness of teachers as measured by student achievement gains. This variation is large in the sense that if an average low-achieving student were assigned to a teacher in the top quintile of effectiveness for several consecutive years, a substantial catch-up in achievement would occur. Unfortunately, little of this variation in teacher effectiveness is associated with traditional measures of teacher training or licensing, which makes identification of high-performance teachers using traditional credentials impossible. At the same time, there is growing recognition that the current system of training and licensing imposes costs on candidates and restricts entry into the teacher applicant pool. The challenge faced by education policymakers is to balance the benefits and costs of changing entry criteria into teaching. Growing use of large state longitudinal data files and more sophisticated scientifically valid studies are helping to shed light on the size and character of these benefits and costs.

Bibliography

- Aaronson, D., Barrow, L., and Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics* **25**(1), 95–136.
- Angrist, J. and Lavy, V. (2001). Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem public schools. *Journal of Labor Economics* **19**(2), 343–369.
- Ballou, D. and Podgursky, M. (2001). Reforming teacher preparation and licensing: What does the evidence show? *Teachers College Record* **101**(1), 5–26.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy* **1**(2), 176–216.
- Clotfelter, C. T., Hadd, H. F., and Vigdor, J. L. (2007). How and why do teacher credentials matter for student achievement? *Working Paper 12828*. Cambridge, MA: NBER.
- Croninger, R. G., Rice, J. K., Rathbun, A., and Nishio, M. (2007). Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review* **26**, 312–324.
- Darling-Hammond, L., Berry, B., and Thorenson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis* **23**(1), 57–77.
- Dee, T. S. and Cohodes, S. R. (2008). Out-of-field teachers and student achievement: Evidence from matched-pairs comparisons. *Public Finance Review* **36**(1), 7–32.
- Education Week (2008). *Quality Counts 2008*. <http://www.edweek.org/rc/articles/2004/10/15/qc-archive.html> (accessed July 2009).
- ETS (Educational Testing Service) (2007). *Teacher Quality in a Changing Policy Landscape: Improvements in the Teaching Pool*. Princeton, NJ: Educational Testing Service.
- Friedman, M. (1962). *Capitalism and Freedom*. Chicago, IL: University of Chicago Press.
- Glazerman, S., Mayer, D., and Decker, P. (2006). Alternative routes to teaching: The impacts of teach for America on student achievement and other outcomes. *Journal of Policy Analysis and Management* **25**(1), 75–96.
- Goldhaber, D. (2002). The mystery of good teaching. *Education Next* **2**, 50–55.
- Goldhaber, D. (2004). Why do we license teachers? In Hess, F., Rotherham, A., and Walsh, K. (eds.) *A Qualified Teacher in Every Classroom: Appraising Old Answers and New Ideas*, pp 81–100. Cambridge, MA: Harvard Education Press.
- Goldhaber, D. (2007). Everyone's doing it, but what does it tell us about teacher effectiveness. *Journal of Human Resources* **42**(4), 765–794.
- Goldhaber, D. and Brewer, D. (1997). Why don't school and teachers seem to matter? *Journal of Human Resources* **32**, 505–523.
- Goldhaber, D. and Brewer, D. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Education Evaluation and Policy Analysis* **22**(2), 129–145.
- Hanushek, E. A. and Rivkin, S. G. (2004). How to improve the supply of high quality teachers. *Brookings Papers in Education Policy: 2004*, pp 7–25. Washington, DC: The Brookings Institution.
- Harris, D. and Sass, T. (2007). Teacher training, teacher quality, and student achievement. *Working Paper #3*. Washington, DC: Urban Institute, CALDER. http://www.caldercenter.org/PDF/1001059_Teacher_Training.pdf (accessed July 2009).
- Jacob, B. A. and Lefgren, L. (2004). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources* **39**(1), 50–79.
- Mosteller, F. and Boruch, R. (eds.) (2002). *Randomized Trials in Education Research*. Washington, DC: The Brookings Institution.
- NASDTEC (National Associate on State Directors of Teacher Education and Certification) (2008). *NASDTEC Manual*. <http://www.nasdtc.org/directory/index.tpl> (accessed July 2009).
- NCTAF (National Commission on Teaching and America's Future) (1996). *What Matters Most*. Washington, DC: National Commission on Teaching and America's Future.
- NCTAF (National Commission on Teaching and America's Future) (1997). *Doing What Matters Most*. Washington, DC: National Commission on Teaching and America's Future.
- Podgursky, M. (2004). Improving academic performance in U.S. schools: Why teacher licensing is (almost) irrelevant. In Hess, F., Rotherham, A., and Walsh, K. (eds.) *A Qualified Teacher in Every Classroom: Appraising Old Answers and New Ideas*, pp 49–79. Cambridge, MA: Harvard Education Press.
- Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica* **73**(2), 417–458.
- Rothstein, J. (2008). *Teacher Quality in Education Production: Tracking, Decay, and Student Achievement*. Princeton, NJ: Princeton University.

- Sanders, W. and Horn, S. (1994). Mixed model methodology in educational assessment. *Journal of Personnel Evaluation in Education* **8**, 299–311.
- Todd, P. E. and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal* **113**(485), F3–F33.
- Wayne, A. J. and Youngs, P. (2003). Teacher characteristics and student achievement: A review. *Review of Education Research* **73**(1), 89–122.
- Woessman, L. (2003). Schooling, resources, and student performance: The international evidence. *Oxford Bulletin of Economics and Statistics* **65**(2), 117–170.
- OECD (2005). Attracting, developing and retraining effective teachers. *Final Report: Teachers Matter*. Paris: OECD.
- Podgursky, M. and Springer, M. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management* **26**(4), 909–949.
- US Department of Education, National Center for Education Statistics (2000). *Teacher Supply in the U.S.: Sources of Newly Hired Teachers in Public and Private Schools: 1987–88 to 1993–94*. Washington, DC: NCES 2000-309.
- Vegas, E. (2007). Teacher labor markets in developing countries. *Future of Children* **17**(1), 219–232.

Further Reading

- Ballou, D. and Podgursky, M. (1997). *Teacher Pay and Teacher Quality*. Kalamazoo, MI: W.E. Upjohn Institute.
- Cochran-Smith, M. and Zeichner, K. (2005). *Studying Teacher Education: Report of the AERA Panel on Research and Teacher Education*. Mahwah, NJ: Erlbaum.
- Hackel, M. D., Koenig, J. D., and Elliott, S. W. (eds.) (2004). *Assessing Accomplished Teaching: Advanced-Level Certification Programs*. Washington, DC: National Academies Press.
- Hanushek, E. A. (2003). The failure of input-based resource policies. *Economic Journal* **113**(485), F64–F98.
- Hanushek, E. A. and Rivkin, S. G. (2006). Teacher quality. In Hanushek, E. and Welch, F. (eds.) *Handbook of the Economics of Education*, vol. 2, pp 1051–1078. Boston, MA: Elsevier.
- Hess, F., Rotherham, A., and Walsh, K. (eds.) (2004). *A Qualified Teacher in Every Classroom: Appraising Old Answers and New Ideas*. Cambridge, MA: Harvard Education Press.

Relevant Websites

- <http://www.aacte.org> – American Association of Colleges of Teacher Education.
- <http://www.caldercenter.org> – Center for Analysis of Longitudinal Data in Education Research.
- <http://nces.ed.gov> – National Center for Education Statistics.
- <http://www.performanceincentives.org> – National Center on Performance Incentives.
- <http://www.nctaf.org> – National Commission on Teaching and America's Future.
- <http://www.ncate.org> – National Council for Accreditation of Teacher Education.
- <http://www.oecd.org> – Organization for Economic Cooperation and Development. Directorate for Education.
- <http://www.teachforamerica.org> – Teach for America.
- <http://www.teac.org> – Teacher Education Accrediting Council.
- <http://www.proudtoserveagain.com> – Troops to Teachers.

Teachers in Developing Countries

F H Rogers and E Vegas, The World Bank, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Why devote a separate article in this encyclopedia to the issue of teachers in developing countries? While there are many similarities between teacher labor markets in developed and developing countries, the marked differences in context and education systems between the two groups have important implications for teacher policy.

Both the education challenges and the resources available to meet them are of a far different scale in the developing world. First, the challenges: Unlike in OECD countries, expanding access to basic education remains a major objective in the developing world. Reflecting this objective, the international community has officially committed to a millennium development goal (MDG) of achieving universal primary education globally by 2015. The MDGs grew out of the agreements and resolutions of world conferences organized by the United Nations during the 1990s. The goals were accepted as a framework for measuring development progress, and for most of them, the self-imposed deadline is 2015. There are eight MDGs:

1. eradicate extreme poverty and hunger;
2. achieve universal primary education;
3. promote gender equality and empower women;
4. reduce child mortality;
5. improve maternal health;
6. combat acquired immune deficiency syndrome (AIDS), malaria, and other diseases;
7. ensure environmental sustainability; and
8. build a global partnership for development (The World Bank, "Achieving the Millennium Development Goals").

Yet in sub-Saharan Africa, only 61% of children complete primary school, and the primary completion rate remains below 90% in South Asia and the Middle East/North Africa region. Rates of secondary school completion are even lower. Even when children are in school, in many developing countries their learning rates are very low. For example, in Argentina, Colombia, and Mexico, less than 50% of fourth graders reach even the lowest level of proficiency on an OECD-normed international literacy test – even though these are middle-income countries with more advanced education systems than the average developing country. Continued population growth will exacerbate these quantity and quality challenges in education: despite falling fertility rates, the population of children aged 5–14 in developing countries is projected to grow 16% between 2005 and 2025. (data from the World Bank Development Data Platform).

While developing countries are under great pressure to expand access and improve quality, their investment in education is substantially lower than in developed countries, both in absolute levels and as a share of GDP (see **Table 1**). Students' parents are often poor and have little formal education themselves, making it harder for them to compensate for weak schools through household investments in education. In addition, teacher compensation accounts for a larger share of total investment in education in developing countries – about 78%, compared to around 64% in Organisation for Economic Co-operation and Development (OECD) countries. Given these facts and the well-documented empirical finding that teachers have important effects on student learning outcomes, it is clear that good teacher policies can pay even larger dividends in developing countries than in developed nations.

Teacher Labor Markets in Developing Countries: Stylized Facts

Meeting the education challenges discussed above depends on recruiting motivated teachers with the necessary skills and ensuring that they are in classrooms everywhere, including in underserved rural areas. As in developed countries, the supply of teachers to developing-country classrooms responds to teacher salaries, working conditions, and opportunity costs, as well as the relative characteristics of the teacher and nonteacher labor market, including entry requirements.

Teacher Salary Levels: Mixed Evidence on Adequacy

When teacher salaries are high relative to those of other employment options, teacher retention and qualifications also tend to be high. In countries such as Taiwan and Korea, where admission to teacher training colleges is highly competitive, teaching attracts candidates with high academic qualifications and pays about as well as most other professions. What is the experience in most developing countries?

Assessing the adequacy of teacher salaries is not straightforward. By the standards of wealthier countries, teacher salaries in low- and middle-income countries may seem unjustifiably low. However, relative to per capita incomes they often look considerably better. The salary/income ratio varies greatly from country to country, but

Table 1 Education expenditures in developed and developing countries, 2003

	<i>Annual expenditure on educational institutions per student (in equivalent US dollars (PPP), by level of education, based on full-time equivalents)</i>		<i>Annual expenditure on educational institutions per student relative to GDP per capita (by level of education, based on full-time equivalents)</i>	
	<i>Primary</i>	<i>Secondary</i>	<i>Primary</i>	<i>Secondary</i>
Developing Countries	\$1066	\$1183	13	16
OECD Countries	\$5450	\$6962	20	26

Sources: UNESCO/UIS WEI (www.uis.unesco.org/publications/wei2006); OECD countries: OECD 2006 (www.oecd.org/edu/eag2006)
 Authors' calculations; data from the World Bank Development Data Platform.

on average it is much higher in poorer than in richer countries: 3–4 times per capita GDP in poor countries, and higher in Francophone Africa, compared with 1.8 times in the OECD (Bruns *et al.*, 2003). It would be preferable to compare teacher salaries with those of comparably skilled workers elsewhere in the economy, but the shorter work week, greater job security, and other benefits of civil-service teachers may make even careful comparisons misleading. Perhaps the best indicator of teacher salaries is whether there are long queues of qualified applicants waiting for a position. In many developing countries, there are in fact substantial numbers of qualified applicants, sometimes even for lower-paid contract teaching positions.

Teacher Salary Characteristics: High Rigidity and Fiscal Cost

As in many developed countries, teacher salaries in developing countries are set according to a relatively rigid scale that maps the teacher's education, experience, and sometimes also location of work into a given salary level. With few exceptions, the teacher's performance does not affect his or her salary progression, unless s/he is able to be promoted to principal or deputy principal. Teacher salaries are rigid in another sense too: in most systems, teachers have very high job security, so that even in cases of malfeasance they can retain their jobs and their salaries.

At the economy-wide level, teacher salaries have major fiscal implications. For many developing countries, teacher salaries claim more than 80% of recurrent education costs, and sometimes the share reaches 95%. In countries that have recently expanded access to primary schools, such as Uganda, Kenya, and Tanzania, ministries cite the high proportion of spending on teacher salaries as the biggest constraint on educational expansion and improvements in the quality of teacher recruits.

Poor Working Conditions

In developing countries, many teachers work in schools that do not have adequate teaching materials or basic

infrastructure, with on average many more children per classroom than in advanced countries (see **Figure 1**).

Often, developing-country teachers cite a lack of resources as a serious obstacle to effective teaching (Futrell, 2004). Data from the 2003 Programme for International Student Assessment (PISA) support this notion. As **Figure 2** shows, schools in Thailand, Mexico, Brazil, Uruguay, and Turkey were more much likely to report that shortage in supplies hindered instruction than schools in OECD countries.

Recent experimental evidence from Kenya finds that providing additional textbooks improves the learning of the top two quintiles of students (presumably, those whose background allowed them to benefit from books that were too hard for other students). Part of the effect appears to be mediated by teachers: textbook provision increased their presence in the classroom and caused them to use the books more frequently in class (Glewwe *et al.*, 2007). Even when poor conditions do not directly hamper the teacher's effectiveness, they are likely to reduce motivation and make it harder to recruit teachers, particularly to serve in poorer areas.

Teacher Entry and Assignment: Challenges for Quality and Equity

In many developing countries, the number of credentials formally required for new civil-service teachers is increasing, in a reflection of what governments see as rising skill demands on teachers and students. At the same time, in their attempt to meet demands for education under tight budget constraints, governments are resorting to hiring part-time, noncredentialed, or contract teachers. Some countries, such as Tajikistan, have also reduced the length of teacher training programs (Chapman *et al.*, 2005).

Once teachers have been recruited into the service, teacher assignment and sorting have major effects on education quality and equity. Most teachers prefer to work in areas with better working conditions and better students; in developing countries, this tends to mean a preference for urban schools. As a result, schools in rural areas often

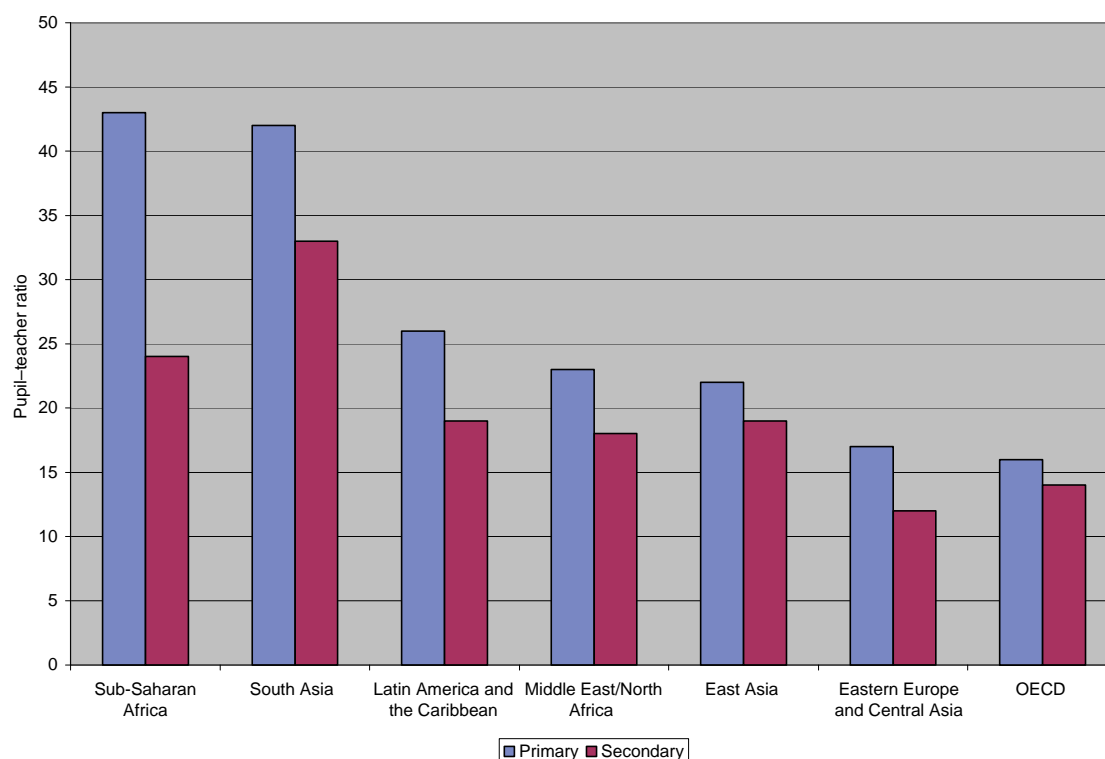


Figure 1 Average Primary Pupil-Teacher Ratio, by Region (2003). Countries with populations of less than 1 million are excluded. Source: UNESCO Institute of Statistics Data (2003). Available at <http://stats.uis.unesco.org/unesco/TableViewer/tableView.aspx?ReportId=165>.

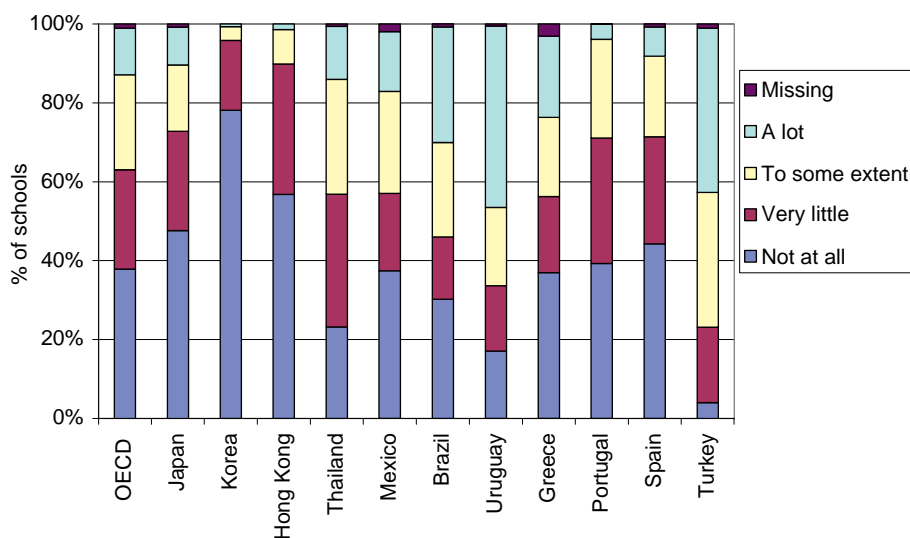


Figure 2 Percentage of schools reporting that shortage of supplies hindered instruction, by country. Source: Authors' graphic using data from the PISA 2003 database.

are staffed by less-qualified teachers. Rural students typically begin at a learning disadvantage, because their parents are less educated and school facilities are poorer than in urban areas; inequities in the distribution of skilled teachers exacerbate that disadvantage.

Inadequate Teacher Management

In many developing countries, public-school teachers are poorly managed, with little monitoring of their performance and few incentives for exerting effort. As a result, many

teachers perform at levels below their ability. Underperformance manifests itself most overtly in high levels of teacher absence: surprise school visits in 2002–03 found that 27% of primary-school teachers in Uganda and 25% in India were absent from their schools, with absence rates reaching roughly 40% in two Indian states (Chaudhury *et al.*, 2006). Even in countries with lower absence rates, teachers miss class at levels well above those that would be expected based on normal excused absence.

From a labor-market perspective, a lack of accountability reduces the opportunity cost of choosing teaching as a career. When out-of-class preparation and even school attendance are optional, teachers can supplement their incomes through outside jobs during school hours. This may make teaching more attractive as a career, for any given salary level, but it may not attract candidates with strong professional motivation to teach. If a teacher fails to give his or her best effort in the classroom, then the teacher-supply problem has been at best half-solved.

Teacher Policies and Their Effects

Developing-country governments face two major sets of challenges in teacher policy: attracting and retaining people with the necessary skills into the teaching profession, and managing them effectively so that their skills translate into more student learning. On the first challenge, budget constraints and a lack of capacity to train the numbers of teachers needed to fill the increasing number of classrooms have led to a range of reforms in developing countries. Some developing countries have introduced pay incentives to attract and retain qualified teachers, while also motivating them to perform better. Countries are also experimenting with devolution to the school of decision-making authority, including for teacher hiring and firing, in order to raise teacher quantity and quality. A third option is to rely more on hiring part-time, contractual, or assistant teachers. Finally, countries have historically used teacher training and professional development as a mechanism to raise teaching quality. The results of these policies on student achievement vary widely based on the context and country.

Teacher Pay

Although experts may disagree regarding just how important is the absolute level of teacher salaries in attracting and retaining qualified individuals to the profession, there is broad consensus that teacher salary level influences who enters the field and how long they remain in teaching. At the same time, other working conditions and regulations can counteract or amplify the influence of wages on teachers. In addition, the extent to which teacher

pay is linked to teacher behavior and student outcomes may affect the quality of those who choose to enter and remain in teaching.

Salary levels. Some evidence supports the intuitive notion that changes in mean teacher salary levels can affect the quality of teacher entrants and of individuals who remain in teaching. For example, a 2005 study documents that changes in teacher wage levels in Chile were accompanied by changes in the overall number as well as the quality of applicants to the teaching profession (Mizala and Romaguera, 2005).

Research also suggests that raising teacher salary levels can have beneficial impacts on student outcomes. In Brazil, a finance equalization reform that targeted redistributed funds to teachers (known as FUNDEF) resulted in smaller class sizes, fewer over-aged children in primary and secondary schools, and a diminishing gap between high- and low-performing students (Gordon and Vegas, 2005).

Salary structure: pay based on performance. Economic theory predicts that higher salaries alone may not improve teachers' effort, even if it improves the quality of new entrants, unless teachers are monitored and a higher level of performance is demanded of them. A few countries have recently experimented with nationwide performance-based pay mechanisms for teachers in an effort to raise teaching quality and student outcomes. Understanding how performance-based pay mechanisms are designed and linked to teacher performance is important in evaluating their impact.

For example, Chile and Mexico have both instituted very different performance-based incentives for teachers, with divergent results. In Chile, the SNED (National System of Performance Assessment) since 1996 has offered monetary bonuses to the schools that perform best in terms of student achievement, with the bonuses distributed among the teachers in the winning schools. A recent evaluation of the program's impact found preliminary evidence that the incentive has improved student achievement in those schools that face relatively good chances of winning the bonus, although the effect appears only cumulatively after a number of years (Mizala and Romaguera, 2005). By contrast, Mexico's *Carrera Magisterial* Program has apparently been less successful. *Carrera Magisterial* is a voluntary program that rewards participating teachers with higher pay based on assessments that include tests of their students and peer reviews of their teaching. Despite its promising design, the program does not appear to have improved student scores on the standardized exam, even though those scores directly affect teacher pay (McEwan and Santibañez, 2005).

An analysis of the characteristics of teacher incentive programs suggests that the size of the incentive matters for its impact on improving teaching quality and student learning. Often, a teacher's base salary accounts for a large share of her total compensation, and incentives for

desired behaviors (e.g., working in rural schools, serving children with special needs) account for only a small proportion of total pay. Even when countries have introduced pay-for-performance schemes for teachers, the majority of a teacher's pay is usually determined by years of service and education (Vegas and Umansky, 2005).

Recent randomized controlled experiments have yielded mixed but promising results of the impact of teacher incentives on improving student outcomes. An experiment in Rajasthan, India, monitored teacher attendance using cameras and then based part of the teachers' salary on their attendance rates. Not only did attendance improve, but so did student learning (Duflo *et al.*, 2009). An experiment in Kenya with incentive pay based on student test score gains found that the program increased learning, although perhaps primarily as a result of 'teaching to the test' (Glewwe *et al.*, 2003). By contrast, a larger-scale experiment in government schools in Andhra Pradesh, India, found that when teachers were paid bonuses (either individually or collectively) based on their students' learning rates, real learning accelerated significantly and cost effectively (Muralidharan and Sundararaman, 2009). (Surprisingly, the teachers' attendance did not improve.)

Rural-area incentives. As a mechanism to induce more and better-qualified teachers into rural and other disadvantaged areas, governments sometimes pay bonuses or supplements to teachers in rural areas. Little rigorous evaluation research has been done to assess the effects of these bonuses on teacher supply and student outcomes. One exception is a study of small location bonuses in Bolivia, which finds that they actually reduce labor supply somewhat and have no robustly positive or negative effect on student outcomes (Urquiola and Vegas, 2005). But despite the lack of evidence, governments believe that bonuses can pay off. In Indonesia, for example, the Parliament in 2005 passed a law providing for much larger bonuses – equal to the teacher's base salary, or roughly a 50–100% increase in compensation, depending on other qualifications – for teachers who agree to locate in special remote and underserved areas.

A policy challenge with rural pay bonuses, as with pay to poorly monitored rural teachers in general, is to ensure that the teachers receiving them do actually relocate to rural areas. Without good monitoring, some become ghost teachers who live in urban areas but draw rural bonuses, sometimes paying a less qualified substitute to take their place in the rural school. In addition, as the Bolivia case shows, many teachers who receive the rural bonus work in urban schools due to the expansion of urban areas.

An alternative strategy is to hire contract teachers from the local community. Even if these teachers are less qualified than their civil-servant counterparts, they may require less monitoring to be induced to stay in rural areas. A complementary approach is suggested by recent research on Pakistan: it finds that placing secondary schools in rural areas has had the effect of generating a supply of local

teachers, as many female secondary graduates go on to teach in their native villages (Andrabi *et al.*, 2005).

School-Based Management

Some developing countries have experimented with enabling schools to make teacher hiring and other administrative decisions, typically devolving power from the regional or central government level. The rationale is that by bringing these decisions closer to the school, and thus to parents and students, school-based management generates incentives and conditions that can improve teaching quality and student outcomes. A number of careful retrospective studies have tried to test the extent to which this works in practice.

In El Salvador, the EDUCO program appears to have improved management practices, teacher behavior, and student outcomes (Sawada and Ragatz, 2005). While many decisions continued to be made by central authorities, the power to hire and fire teachers had been devolved to the school level, with considerable power given to parents. Perhaps as a result, EDUCO schools had fewer school closings, less teacher absenteeism, more meetings between teachers and parents, and longer teacher work hours than did comparison schools. These changes, in turn, are related to higher student achievement in Spanish in EDUCO schools.

Honduras' PROHECO project, also a school-based management reform for rural primary schools, appears to have had beneficial effects as well (Marshall and di Gropello, 2005). Comparing PROHECO schools to similar schools in rural areas (using propensity score matching methods to construct a credible control group) reveals mixed program effects on attendance of PROHECO teachers, and it fails to identify differences between PROHECO and comparison teachers in terms of their classroom processes, planning, or motivation. Yet although PROHECO teachers are paid less and have fewer years of experience than comparison teachers, they teach more hours in an average week and assign more homework. Apparently as a result, PROHECO students score higher on math, science, and Spanish exams than students in similar non-PROHECO schools. These examples lend credence to the hypothesis that decentralized schools can induce greater teacher effort and efficiency.

While these evaluations have the strength of assessing reforms at a national level, they suffer from an inability to control for unobserved factors, such as dynamic school leadership or low teacher morale, that might influence both a school's selection into the program and its teacher behavior or student outcomes. Randomized controlled trials overcome this problem, although potentially at the cost of external validity, in that we cannot be certain the results apply beyond the subnational sample included in the experiment. A recent study in Kenya shows that

school-based management reforms can have significant payoffs, especially when they interact with other interventions. When school committees were given the power to monitor and review performance and to decide whether to re-hire contract teachers, test scores of students of civil-service teachers improved, apparently because the reform increased the teacher's effort.

Alternatives to Hiring Regular Teachers

Many developing countries are turning to contract teachers to address shortages. Contract teachers are typically graduates from regular teacher training institutes, but often receive much lower wages than do regular teachers and no benefits. In addition to being less costly to governments, contract teachers are often assumed to be more motivated. Unlike civil-service teachers, who have very high levels of job security, contract teachers know that performing poorly may prevent them from being re-hired in the future.

Only recently has rigorous evidence of the effects of contract teachers begun to appear. One approach uses retrospective data, as in the case of Togo, which recently reported that as much as 55% of its teaching force was contractual. The inception of contractual hiring at first reduced the supply of high-quality candidates, while also causing absenteeism and resentment over unfair pay. During this period, performance of students with contractual teachers lagged behind achievement of students taught by regular teachers, even after controlling for prior achievement, household characteristics, and school, classroom, and teacher variables in a retrospective evaluation. Not surprisingly, however, schools that can afford only contractual teachers (or can afford few regular teachers) also are disadvantaged in other ways: for example, they have less pedagogic supervision and poor facilities (De Laat and Vegas, 2005). Therefore, it is possible that estimates of contract teachers' effectiveness are biased downward by these variables that the researcher cannot observe. A more recent study of contract teachers in Togo, Mali, and Niger finds positive learning effects under certain circumstances – in lower grades and with less advanced students, and in cases (as in Togo) when teachers are hired and monitored locally rather than centrally (Bourdon *et al.*, 2007).

A recent randomized controlled experiment in Kenya overcomes problems with retrospective data by randomly assigning students to additional contract teachers provided to schools in the treatment group (Duflo *et al.*, 2009). Rates of teaching activity by contract teachers were 30 percentage points higher than those of civil service teachers, supporting the greater-effort hypothesis, and students of contract teachers showed more rapid learning gains. The addition of contract teachers reduced the effort of civil-service teachers, offsetting the beneficial effects on learning. This finding underlines the need to combine

contracting reforms with other reforms to increase accountability.

Countries are also experimenting with hiring assistant teachers, who often have fewer qualifications than do regular teachers and are paid substantially lower salaries. The Pratham community-based urban remedial education program in India reached over 15 000 students by hiring young women from the community without teaching certificates to teach basic literacy and numeracy skills to children who had not attained these basic skills by the third grade. A randomized evaluation found that after 2 years, the remedial education program increased learning by an amount equivalent to about half a year's worth of normal learning, with the largest gains among the least able students. This finding demonstrates that it is possible not only to retain poor-performing students but also to prevent them from falling behind. Because the program hires local high-school-educated women to teach classes of around 20 students, the average cost of the program is \$5 per child per year – far less than the average cost of outfitting classrooms with regular teachers (Banerjee *et al.*, 2005).

Teacher Training

The literature on the impact of teacher training programs on education quality is, at best, inconclusive. A survey of case studies in Latin America found consistently poor results from several methods of teacher preparation and training (Navarro and Verdisco, 2000).

To attract talented students to the teaching career, several Latin American countries have introduced scholarship programs. In Chile, a scholarship program for talented students covers 100% of tuition for up to 1 million pesos in exchange for a commitment to teach for 3 years. Priority has been given to candidates in natural sciences, mathematics, English, language, and basic education (Pogre and Lombardi, 2004). In Uruguay, where teacher education is free of charge, scholarships are provided to talented candidates from disadvantaged backgrounds to cover their living expenses during the 3 years of intensive training at regional teacher training centers (Vaillant, 1999).

Many developing countries have also made efforts to professionalize teacher education, gradually increasing the length of study from 3 to 5 years, and shifting teacher education from normal schools to higher education. However during the transition, multiple types of teacher training institutions operate simultaneously, creating teacher training systems that are fragmented and unsystematic. Another problem is that most teacher education programs in developing countries lack the flexibility to attract professionals from other fields, or candidates interested in a mid-career change. By contrast, most OECD countries offer multiple entry points to the teaching profession.

Some Lessons from Research on Teacher Policy in Developing Countries

From the stylized facts and empirical evidence presented above, we can draw some general lessons on teacher policy in developing countries. Many types of education reforms – and not just those usually considered as teacher policy – affect teaching quality and therefore student learning.

First, evidence from developing countries supports the intuitive notion that teaching quality is sensitive to the level and structure of compensation. Once teacher salaries are competitive, linking pay increases to teacher performance can improve teaching and accelerate learning. Both Chile's SNED program and smaller-scale experiments in developing countries have shown that this approach can work. The design of teacher incentive reforms is crucial, however. Even in nationwide performance-based pay reforms, often only a small proportion of teachers face greater incentives to improve learning in their classrooms. For an incentive scheme to work effectively, it must be targeted enough to recognize only teachers who truly exhibit the desired performance and results, but must also be generous enough to provide incentives for teachers who exert the extra effort.

The developing country evidence also suggests that changes in other aspects of teacher contracts can affect teaching quality and student learning, by influencing the characteristics of those who choose to enter and remain in teaching and, importantly, their work in classrooms. One important example is school-based management reforms, which devolve decision-making authority to the school. The Central American nationwide experiences and smaller-scale experiments have shown that these reforms can result in less teacher absenteeism, more teacher work hours, more homework assigned, closer parent–teacher relationships, and accelerated student learning. These are promising changes, especially in contexts of low educational quality.

A key lesson from research in developing and developed countries is that teachers do not always respond to incentives in predictable ways. Sometimes, programs that are specifically designed to reward teachers who adopt specific behaviors or achieve higher results fail to generate the intended behavioral response from teachers (Glewwe *et al.*, 2003; Clotfelter *et al.*, 2004; Koretz, 2002; Jacob and Levitt, 2002). For example, Bolivia's bonus for teaching in rural areas is not resulting in higher-quality rural teachers, and the apparently well-designed Carrera Magisterial is not improving student outcomes. These cases highlight the importance of design and implementation of teacher incentive reforms – and, finally, also the importance of rigorous evaluations that allow us to draw conclusions about their effectiveness.

See also: An Overview of Teacher Labor Markets; Compensating Differentials in Teacher Labor Markets; Economic Approaches to Teacher Recruitment and Retention; Teacher Incentives; Teacher Supply; Teacher Training and Preparation in the United States; The Economics of Teachers' Unions in the United States.

Bibliography

- Andrabi, T., Das, J., and Khwaja, A. I. (2005). Students today, teachers tomorrow? The rise of affordable private schools in Pakistan. Mimeo. Banerjee, A., Cole, S., Duflo, E., and Linden, L. (2005). Remedying education: Evidence from two randomized experiments in India. *NBER Working Paper No. 11904*. Cambridge: National Bureau of Economic Research.
- Bourdon, J., Frölich, M., and Michaelowa, K. (2007). Teacher shortages, teacher contracts and their impact on education in Africa. *IREDU Working Paper 2007/6*.
- Bruns, B., Mingat, A., and Rakotomalala, R. (2003). *A Chance for Every Child*. Washington, DC: World Bank.
- Chapman, D., Weidman, J., Cohen, M., and Mercer, M. (2005). The search for quality: A five country study of national strategies to improve educational quality in Central Asia. *International Journal of Educational Development* 25, 514–530.
- Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., and Rogers, F. H. (2006). Missing in action: Teacher and health worker absence in developing countries. *Journal of Economic Perspectives* 20(1), 91–116.
- Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., and Diaz, R. A. (2004). Do school accountability systems make it more difficult for low performing schools to attract and retain high quality teachers? *Journal of Policy Analysis and Management* 23, 251–271.
- De Laat, J. and Vegas, E. (2005). Do differences in teacher contracts affect student performance? Evidence from Togo. *Working Paper No. 26955*, Mimeo. Washington, DC: World Bank.
- Duflo, E., Dupas, P., and Kremer, M. (2009). Additional Resources versus Organizational Changes in Education: Experimental Evidence from Kenya. Unpublished paper, MIT.
- Duflo, E., Hanna, R., and Ryan, S. (2008). Monitoring Works: Getting Teachers to Come to School. CEPR Discussion Papers No. 6682.
- Futrell, M. H. (2004). *Speech by E.I. President to Fourth World Congress of Education International*. Porto Alegre, Brazil, 22 July 2004.
- Glewwe, P., Ilias, N., and Kremer, M. (2003). Teacher incentives. *National Bureau of Economic Research Working Paper* 9671.
- Glewwe, P., Kremer, M., and Moulin, S. (2007). Many children left behind? Textbooks and test scores in Kenya. *Poverty Action Lab Paper No. 44*.
- Gordon, N. and Vegas, E. (2005). Educational finance equalization, spending, teacher quality, and student outcomes: The case of Brazil's Fundef In Vegas, E. (ed.) *Incentives to Improve Teaching: Lessons from Latin America*. Washington, DC: World Bank.
- Jacob, B. and Levitt, S. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *NBER Working Paper w9413*, National Bureau of Economic Research, Cambridge, MA.
- Koretz, D. (2002). Limitations in the use of achievement tests as measures of educators' productivity *Journal of Human Resources* 37, 752–777.
- Marshall, J. and Di Gropello, E. (2005). Teacher effort and schooling outcomes in rural Honduras. In Vegas, E. (ed.) *Incentives to Improve Teaching: Lessons from Latin America*, pp 359–388. Washington, DC: World Bank.
- McEwan, P. and Santibáñez, L. (2005). Teacher and principal incentives in Mexico. In Vegas, E. (ed.) *Incentives to Improve Teaching: Lessons from Latin America*, pp 213–254 Washington, DC: World Bank.
- Mizala, A. and Romaguera, P. (2005). Teachers' salary structure and incentives in Chile. In Vegas, E. (ed.) *Incentives to Improve Teaching: Lessons from Latin America*, pp 307–358 Washington, DC: World Bank.

- Muralidharan, K. and Sundararaman, V. (2009). Teacher Performance Pay: Experimental Evidence from India. NBER Working Paper 15323.
- Navarro, J. C. and Verdisco, A. (2000). Teacher training in Latin America: Innovations and trends. *Sustainable Development Department Technical Paper Series*. Washington, DC: Inter-American Development Bank.
- Pogre, P. and Lombarda, G. (2004). *Escuelas que enseñan a pensar: Un marco teórico para la acción*. Buenos Aires: Consorcio de Editores.
- Sawada, Y. and Ragatz, A. (2005). Decentralization of education, teacher behavior, and outcomes. The case of El Salvador's EDUCO. In Vegas, E. (ed.) *Incentives to Improve Teaching: Lessons from Latin America*, pp 255–306. Washington, DC: World Bank.
- Urquiola, M. and Vegas, E. (2005). Arbitrary variation in teacher salaries. In Vegas, E. (ed.) *Incentives to Improve Teaching: Lessons from Latin America*, pp 187–212. Washington, DC: World Bank.
- Vaillant, D. (1999). Reforma del sistema formación inicial de docentes en Uruguay. *Paper Presented at the Los maestros en América Latina: Nueva perspectivas sobre su desarrollo y desempeño Conference* San Jose, Costa Rica.
- Vegas, E. and Umansky, I. (2005). Improving teaching and learning through effective incentives: What can we learn from education reforms in Latin America? In Vegas, E. (ed.) *Incentives to Improve Teaching: Lessons from Latin America*. Washington, DC: World Bank.

Further Reading

OECD (2003). Programme for International Student Assessment 2003 database. http://pisaweb.acer.edu.au/oecd_2003/oecd_pisa_data.html (accessed September 2009).

Relevant Websites

<http://www.worldbank.org> – The World Bank
<http://econ.worldbank.org> – Data and Research, The World Bank

The Economics of Teachers' Unions in the United States

K Strunk, University of Southern California, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Collective bargaining – The negotiations between the union and an employer (in this case, the school district or school board) over employees' working conditions and/or wages.

Collective bargaining agreement – The agreement that is reached through the process of collective bargaining; the employer–union contract.

Good faith – A term indicating that both parties in contract negotiations acted with honesty and good intentions during bargaining proceedings.

Monopsony – A market in which there is only a single buyer facing many sellers. In the case of school districts and teachers, a district can be a monopsonist employer.

collective bargaining (CB). Teachers' unions bargain over the teachers' contracts with their districts, helping to set the minutiae of districts' school policies that regulate teachers' workdays. In light of unions' influence and import in education policy and politics, surprisingly little empirical research has been conducted on teachers' unions and the effects of their influence and actions, and few general conclusions about the effects of teachers' unions have been reached.

This article provides an overview of teachers' unions in the United States, explains the theoretical ways in which teachers' unions may help or hurt school and district outcomes, and summarizes the empirical literature on teachers' unions and their effects. Due to the complex structures and diversity of teachers' unions in nations other than the United States, this article largely focuses on teachers' unions in the USA.

Introduction

Teachers' unions are important actors at all levels of education policy. While the organizational structure of teachers' unions varies across nations, most developed and many developing countries have at least one organization whose role it is to collectively bargain for public school teachers. In the United States, the National Education Association (NEA) and the American Federation of Teachers (AFT), the two main teachers' unions, together have active affiliates in every state and in virtually every school district in the country, with a combined membership of approximately 4.6 million teachers and school personnel. The influence of teachers' unions does not come from their size alone; they also command considerable monetary resources. Since 1990, United States teachers' unions and their members have contributed over \$50 million to federal election campaigns, including almost \$5 million during the 2006 election cycle alone (Center for Responsive Politics). Unions also lobby legislators, target media with support messages, provide endorsements, draft policy recommendations, and supply campaign volunteers. Unions' size, reach, wealth, and level of organization combine to give them profound power in education politics.

However, regardless of their national- and state-level political influence, the most profound day-to-day impact of United States teachers' union activity occurs in districts where unions are empowered to represent teachers in

A Brief History of Teachers' Unions in the United States

Teacher unionism in the United States began in 1857 with the founding of the National Teacher Association, the predecessor to today's NEA. The first antecedent to the AFT was formed about half a century later, in 1897, as the Chicago Teachers Federation. Around this time, teachers in many large cities joined together in local city unions. Four of these city-level unions combined in 1916 to form the AFT. The AFT quickly affiliated with the American Federation of Labor and joined the growing labor movement. Although the early purpose of these two unions was to assist public school teachers in their negotiations with school administrators and boards, the AFT did not start collectively bargaining with school boards until the 1940s and the first collective bargaining agreement (CBA) between school officials and teachers was not enacted until 1962 as a consequence of a 1-day strike 2 years prior in New York City (Education Commission of the States (ECS), Unions/Collective Bargaining). Although New York City had the first agreement, New York State was not the first state with a CB law for teachers. Wisconsin passed the first such law for public employees in 1959, but it was not until 1964 that the first teachers' bargaining agent was certified in Wisconsin. New York's law, the Taylor law, was passed in 1967. Even as labor union membership declined in the private sector, teachers' union membership climbed throughout the 1960s,

growing from approximately 826 300 members in 1961 to 1.06 million members in 1965. Today, the NEA and the AFT, the United States' two major teachers' unions, together claim over 4.6 million members.

An Overview of Laws Regulating Teachers' Unions in the United States

Although teachers' associations and unions exist in every state in the United States, they do not hold the same legal status or rights throughout the country. Union activity varies by state, in large part due to different state-level regulations concerning teachers' rights to organize, strike, and collectively bargain, as well as the extent of state-level policies mandating the operation of districts and schools. Thirty-six states have CB laws that outline the bargaining rights for teachers. CB laws simply guarantee that teachers have the right to negotiate a contract and that the union will have a place at the bargaining table. These laws can vary substantially across states, but typically each law has four key sections:

1. *Union representation.* This section of CB laws specifies how a bargaining agent (i.e., the union) is identified, certified, and decertified by the covered employee population. A bargaining agent is most often certified or decertified as the result of a majority vote by the union membership. Who exactly belongs to a union varies by district and by state. State laws can prohibit different school personnel from belonging to the same bargaining unit. The general rule is that personnel in the same union must share a community of interests. District administrations and union leaders can then negotiate over exactly who is covered within the state-imposed constraints.
2. *Good faith.* CB laws also require that both parties (the union and the school district) bargain in good faith. States most often interpret good faith as meaning a negotiation process guided by fairness and cooperation between the two parties, although they vary in the amount of guidance given within the law as to what constitutes such actions.
3. *Impasse and strikes.* When negotiations fail to lead to a final contract between parties, a legal impasse occurs and active bargaining is usually suspended. State laws dictate the process that the parties must follow at this point, including mediation and arbitration options to overcome the impasse. The CB laws address the legality of strikes as well. About half of states' CB laws allow teachers' unions to legally strike, and those that do not allow for strikes outline the consequences for unions that illegally coordinate a teachers' strike.
4. *Scope of bargaining.* CB laws outline the scope of representation in the bargaining process. What aspects of a teachers' work day can be negotiated in the CB process

varies across states, but can be as broad as allowing unions and school districts to bargain over wages, hours, and terms and conditions of employment, and as restrictive as mandating specific topics such as leave or benefits.

Twenty-five states have no CB laws for teachers. This means that teachers in these states are employed at-will and have no contract to govern their employment. In states without CB laws, many of the regulations that are traditionally set in CBAs are instead set by state statutes or in district policies.

There is also a distinction between states with right-to-work laws and those without. Twenty-three states have right-to-work laws, which prevent CBAs from including clauses that require teachers and other employees to join the union and/or require employees from contributing union membership fees. These laws do not necessarily prohibit employees from voluntarily joining and paying union fees. In fact, 11 of the 23 states with right-to-work laws also have CB laws, allowing unions to form and represent employees but prohibiting them from requiring those employees to join and pay dues if they do not wish to do so.

Five states have laws actually prohibiting teachers from collectively bargaining contracts. Two states simply have no laws on the books regarding CB for teachers.

Theoretical Ways in Which Teachers' Unions May Enhance or Detract from Education

Unionism might have both positive and negative impacts on school district outcomes. That is, union activity has the potential to simultaneously enhance some parts of the educational process and diminish other parts. Whether the net effect is positive or negative is an empirical question, although a difficult one to answer. Scholars and commentators have made several arguments for ways that teachers' unions can both positively and negatively impact school district processes and outcomes.

Ways in Which Unions May Positively Impact School Districts

Proponents of teachers' unions cite multiple ways in which these unions enhance student and/or school or district performance. This list often includes attracting higher-quality teachers, giving teachers time to focus on teaching, improving the flow of information between classrooms and administrators, stabilizing the school and district, and correcting market imperfections.

Unions improve teacher quality by increasing teachers' well-being

There is little doubt that better teachers lead to better-educated students. Insofar as workers in labor markets

choose jobs that maximize the utility (provide them with the maximum amount of happiness) that they receive from pecuniary and nonpecuniary benefits, unions may draw better teachers to the field by increasing teachers' wages and enhancing their working conditions and benefit.

Unions allow teachers to focus on teaching

By serving as a decision-making agent for teachers, unions free up valuable time and energy that teachers might otherwise spend on legal issues, bargaining with school system management over personnel contracts and administrative minutiae. Eliminating some noninstructional tasks allows teachers to focus resources on instruction.

Unions improve information flows

Because teachers work much more closely with students than administrators or policymakers, they may have better information about how resources should be best allocated to improve student learning. Unions can collect information from members and disseminate it, potentially improving policy. In addition, they may serve a mediating role, transmitting good information from administrators to teachers, thus easing the flow of information in both directions.

Unions increase efficiency of resource use

Proponents argue that unions have the same basic objective that parents have; increasing student achievement. However, because teachers have better information regarding the efficacy of specific resources, unions are in a position to bargain for optimal allocations. Efficiency-enhancing teachers' unions thus use their superior knowledge to bargain for improved school productivity.

Unions stabilize school districts

School district administrations turn over frequently, and district policies shift with each new superintendent or governing board. Union-enhanced job protection can help to insulate school districts against frequent and, possibly, disruptive change.

Unions correct a market failure in the education marketplace

Public school districts often hold monopsony power over the employment of teachers; they are often the only employer of individual teachers in a given geographic area. This imbalance of power can allow districts to provide low salaries to teachers because there are no other large employers for their skills. In the absence of labor competition, teacher pay and benefits are held below classical equilibrium levels, and since lower pay reduces the quality of the available workforce, students suffer. Unions help to correct this market failure by providing teachers with bargaining strength, which can compensate for schools' demand-side monopsony power.

Ways in Which Unions May Negatively Impact School Districts

Critics of teachers' unions offer an alternative set of mechanisms whereby unions decrease inputs to educational production and detract from student learning. Included among this list is the assertion that unions decrease the efficiency of inputs, skew teachers' performance incentives, hurt communities, and create divisive and hostile relations between teachers and districts.

Unions decrease efficient allocation of resources

As rational economic agents, teachers want to maximize their own happiness, which, for example, may be best served through increased salaries and decreased workloads. Under fixed budget constraints in which districts must work within a set budget, elevated teacher salaries may mean fewer dollars spent on other educational resources that may increase student achievement more efficiently.

Unions disrupt incentives for teacher performance

In private-sector labor markets, the opportunity to command higher pay gives employees incentives to increase their productivity, or, at the least, the desire to avoid termination induces them to maintain a minimum productivity level. Because unions negotiate for rules that protect teachers from dismissal and increase returns to seniority, they may reduce incentives for good teacher performance and make it more difficult for school districts to replace low-performing teachers. This interference with incentives may lead fewer teachers to perform at a high level despite being paid wages that are perhaps above market levels.

Unions reduce communities' social welfare

Unions inflate school budgets, thus increasing community tax rates. This will reduce overall social welfare if it is not accompanied by a corresponding increase in student performance. In this case, housing values will suffer, reducing the amount of wealth held by the community and lowering districts' capacity for future school investment. In addition, contentious negotiations and strike threats may hurt community morale and community members' personal investment in schools.

Unions drive wedges between teachers and districts

The preferences of union elites may not represent the preferences of the average teacher, leading unions to misrepresent teachers' views to administrators and the public as well as to impede the flow of information to union members. By setting up teachers and districts as opposing forces, unions would promote distrust between

teachers and administrators, making the accomplishment of day-to-day operations more difficult.

Unions threaten democratic processes

Teachers' unions are unique in their ability to help elect the very management with whom they bargain. Because teachers often teach in the district in which they vote, they help to elect the school board that bargains with the union that represents them. Unions thus have incentives to mobilize to elect school board candidates sympathetic to union interests. Because unions are well organized and command resources other groups are unlikely to be able to employ in local elections, unions may disproportionately crowd out the voices of other interests.

Teachers' unions thus have the potential both to help and to hurt educational processes through these different mechanisms. Perhaps the essential question for policy-makers, then, is one of net effects, which are an empirical matter.

The Impact of Unions on District Outcomes: Evidence from Empirical Literature

At the district level, unions and district administrations bargain over wages, employment levels (more employees can also be seen as reduced class sizes or reduced student–teacher ratios), and other working conditions. As such, much of the extant literature on teachers' unions has focused on their impacts on teachers' wages, class sizes, and other working conditions such as preparation time and employment protection. All of these factors affect district spending, leading to examinations of the relationship between unions and overall district spending. While all of these relationships are important, there is an overarching desire to know what effects all of these have on the bottom line – student achievement. A good deal of the research on teachers' unions attempts to parse out how unionization is related to student outcomes. This section outlines what research has shown about the impacts of teachers' unions. As important as what is known about teachers' unions is what is still left unknown – the gaps in the literature are wide and there is much room for continued research.

The Relationship between Teachers' Unions and Teachers' Pay

Economic theory predicts that unionism will elevate teachers' wages because teachers will be able to collectively bargain with monopsonistic employers rather than attempt to negotiate separately with a wage-setting organization. Research has generally found this to be the case. Hall and Carroll (1973) examine teacher salaries in Cook

County, Illinois, for the 1968–69 school year and estimate that the existence of a CBA in a district was worth approximately \$167 to teachers, nearly the same amount as an extra year of teaching experience. Baugh and Stone (1982) find a positive and significant wage premium for union members as well as for all teachers in unionized districts and a 14 percentage point growth in the union district wage premium from 1974 to 1977. Kleiner and Petree (1988) compare states with greater and lesser unionization and find that CB coverage is associated with higher salaries for public school teachers. Zwerling and Thomason (1995) estimate the effect of the state density of teachers' unions on teachers' salaries. Using the 1984 Administrator Teacher Survey from the High School and Beyond data set, they find that a 10% increase in the state density of teachers' unions increases the highest teachers' salaries by 2.6% and the lowest salaries by 0.2%. The authors also argue that the spread of teacher unionism in a state has spillover effects for nonunionized teachers within that state, raising their wages even if they are not themselves union members. Hoxby (1996), using a differences-in-differences regression analysis, finds a 5.1% increase in teachers' salaries stemming from teacher unionization. When she instruments for unionization with the timing of the passage of laws permitting teachers to unionize, Hoxby finds that teachers' unions cause a 5% increase in teachers' salaries.

The Relationship between Teachers' Unions and Student–Teacher Ratios

In keeping with the theory that predicts that unions bargain to increase their total membership, much of the research on teachers' unions finds that unionization decreases student–teacher ratios. Hoxby (1996), for example, finds that unionization significantly decreases the student–teacher ratio by 1.11–1.7 students per teacher. However, not all of the extant literature points in this direction: Kleiner and Petree (1988) find that unionism is associated with higher class sizes for college preparatory students (those who take the Scholastic Aptitude Test (SAT) and/or American College Test (ACT)).

The Relationship between Teachers' Unions and Other Working Conditions

Teachers and their union representatives care about enhancing other working conditions in addition to class sizes. For example, teachers and unions may desire fewer class preparations, increased preparation time, and more flexible working schedules and leave-time policies. While not as much research has focused on these other working conditions, what research does exist has found that unions increase time and funds spent on these factors. For example, Kleiner and Petree (1988), Eberts and

Stone (1984), and Eberts (1984) all find that unionism tends to be associated with more preparation time for teachers. Eberts (1984) also finds that unionized teachers have 4% more preparation time than do nonunionized teachers, an average of 3 min per day.

The Relationship between Teachers' Unions and School District Spending

Increased teachers' salaries, decreased student–teacher ratios, and increased preparation time all cost school districts money. As such, it comes as little surprise that the extant research points to a positive link between unionism and school district spending. Essentially, unionized districts tend to have higher per-student budgets than do nonunionized school districts. Hoxby (1996) shows that unionization induced district spending to increase by 2.9–12.3%. Berkman *et al.* (2004) find that districts with strong local unions spend roughly \$353 more per student per year and that stronger union states spend approximately \$135 more per student per year, on average.

The Relationship between Teachers' Unions and Student Achievement

The relationship between teachers' unions and student achievement is perhaps the most important question in the teachers' union evaluation literature, although it is also one of the most difficult to answer. The literature examining this relationship in the United States provides mixed results as to whether unions and unionization help or hurt student outcomes. Eberts and Stone (1987), comparing education production functions in union and non-union districts, find an average test score gain in union districts of just over 3% greater than in nonunion districts. However, they find a nonlinear relationship between students' pretest scores and productivity gains in union districts. Specifically, they find that students scoring at the sample mean see an even greater productivity increase in union districts, of 7%; however, students at the tails of the distribution (who score very low or very high on pretest scores) appear to perform better in nonunion districts. Milkman (1997) also finds that there are differences in how minority and nonminority students fare in unionized versus nonunionized school districts. He finds that while the average minority student benefits from being in a union school, he or she benefits less than the average nonminority student. Hoxby (1996) finds that teacher unionization caused dropout rates to increase significantly, by 1.8–2.3 percentage points. This finding is in line with the rest of the literature and indicates that students at the tail ends of the performance distribution (dropouts tend to fall in the far left-hand tail of the performance distribution) perform worse in union districts.

Research on the relationship between teachers' unions and student performance across the world is limited in scope. Wöessmann (2003), using data on students across 39 countries from the Third International Mathematics and Science Study (TIMSS), finds that students in schools where teachers' collective organizations or teachers' unions have substantial influence on curriculum perform significantly worse than their peers in schools in which teachers' organizations have less authority over curriculum choices.

Summary

Teachers' unions and the contracts that result from teachers' unions and districts bargaining are thought to have strong implications for education policy at all levels of government. Although there is much speculation and discussion of the role of teachers' unions in education and policy, little is known about the workings and affects of teachers unions. Questions of teachers' unions' impacts on district and student outcomes are largely unanswered. What is clear is that teachers' unions do impact the districts and states in which they work by raising teachers' salaries and district spending as a whole, and by affecting the outcomes of individual students. More research is needed on how and why teachers' unions and their contracts affect students and districts, and what can be done to augment the productive and diminish the obstructive aspects of teachers' unions.

See also: An Overview of Teacher Labor Markets; Economic Approaches to Teacher Recruitment and Retention; Teacher Incentives.

Bibliography

- Baugh, W. H. and Stone, J. A. (1982). Teachers, unions, and wages in the 1970s: Unionism now pays. *Industrial Labor Relations Review* **35**, 368–376.
- Berkman, M., Plutzer, E., and Jones-White, D. (2004). School budgets and local teachers' unions: The mediating role of institutions. *Paper Presented at the State Politics and Policy Conference*, Akron, OH: Kent State University.
- Eberts, R. W. (1984). Union effects on teacher productivity. *Industrial and Labor Relations Review* **37**(3), 346–348.
- Eberts, R. W. and Stone, J. A. (1984). *Unions and Public Schools: The Effect of Collective Bargaining on American Education*. Lexington, MA: Lexington Books.
- Eberts, R. W. and Stone, J. A. (1987). Teachers unions and the productivity of public schools. *Industrial and Labor Relations Review* **40**(3), 354–363.
- Hall, W. C. and Carroll, N. (1973). The effect of teachers' organizations on salaries and class size. *Industrial and Labor Relations Review* **26**, 834–841.
- Hoxby, C. M. (1996). How teachers unions affect education production. *Quarterly Journal of Economics* **111**(3), 617–718.
- Kleiner, M. and Petree, D. (1988). Unionism and licensing of public school teachers: Impact on wages and educational output. In

- Freeman, R. and Ichniowski, C. (eds.) *When Public Sector Workers Unionize*, pp 305–322. Chicago, IL: University of Chicago Press.
- Milkman, M. I. (1997). Teachers' unions, productivity and minority student achievement. *Journal of Labor Research* **18**, 247–250.
- Wöessmann, L. (2003). Schooling resources, educational institutions and student performance: The international evidence. *Oxford Bulletin of Economics and Statistics* **65**(2), 117–170.
- Zwerling, H. L. and Thomason, T. (1995). Collective bargaining and the determinants of teachers' salaries. *Journal of Labor Research* **16**, 467–484.

Further Reading

- Álvarez, J., Moreno, V. G., and Patrinos, H. A. (2007). *Institutional Effects as Determinants of Learning Outcomes: Exploring State Variations in Mexico*. Prepared for the Mexico/World Bank Study on the Quality of Education in Mexico.
- Cooper, B. S. (1992). *Labor Relations in Education: An International Perspective*. Westport, CT: Greenwood Press.
- Hannaway, J. (1993). Political pressure and decentralization in institutional organizations: The case of school districts. *Sociology of Education* **66**(3), 147–163.
- Hannaway, J. and Rotherham, A. (eds.) (2006). *Collective Bargaining in Education: Negotiating Change in Today's Schools*. Cambridge, MA: Harvard Education Press.
- Kerchner, C. T. and Koppich, J. E. (eds.) (1993). *A Union of Professionals: Labor Relations and Education Reform*. New York: Teachers College Press.
- Lawn, M. (ed.) (1985). *The Politics of Teacher Unionism: International Perspectives*. Dover, NH: Croom Helm.
- Loeb, S. and Miller, L. (2007). *A Review of State Teacher Policies: What Are They, What Are Their Effects, and What Are Their Implications for School Finance?* Stanford, CA: Institute for Research on Education Policy and Practice.
- Loveless, T. (ed.) (2000). *Conflicting Missions? Teachers Unions and Educational Reform*. Washington, DC: The Brookings Institution Press.
- McDonnell, L. M. and Pascal, A. (1988). *Teachers Unions and Educational Reform*. Santa Monica, CA: RAND.
- Pencavel, J. (1984). The tradeoff between wages and employment in trade union objectives. *Quarterly Journal of Economics* **99**(2), 215–231.
- Pencavel, J. (1991). *Labor Markets under Trade Unionism*. Cambridge, MA: Basil Blackwell.

Relevant Websites

- <http://www.opensecrets.org> – Center for Responsive Politics.
- <http://www.weac.org> – The Wisconsin Education Association Council.

EDUCATION OF CHILDREN WITH SPECIAL NEEDS

Contents

Adapted Physical Activity for Students with Special Needs
Assessment Accommodations for Children with Special Needs
Assistive Technology and Educational Practice
Augmentative–Alternative Communication
Community Integration and Employment of Youth with Special Needs
Cooperative Learning for Children with Special Needs
Co-Teaching
Cultural and Linguistic Diversity and Children with Special Needs
Curriculum-Based Assessment and Students with Special Needs
Diagnosis and Classification
Direct Instruction and the Education of Children with Special Needs
Disabilities and Gifted Education
Early Childhood Special Education: Birth to Eight Years Old
Early Intervention for Students with Special Needs
Early Transition of Children with Special Needs
Educating Children Who Are Deaf–Blind
Educating Children with Deafness and Hearing Impairments
Educating Children with Other Health Impairments (ADHD)
Educating Children with Physical Disabilities
Educating Children with Specific Learning Disabilities
Educating Children with Speech and Language Disorders
Educating Students with Autism and Related Disorders
Educating Students with Cognitive Disabilities
Educating Students with Emotional and Behavioral Disorders
Educating Students with Special Needs: An Overview
Educating Students with Traumatic Brain Injury
English Language Learners with Special Needs
Functional Behavioral Assessment
History of Special Education
Inclusion of Students with Special Needs in General Education Classrooms
Instructional Accommodations for Children with Special Needs In Inclusive Settings
Interagency Agreements
Interdisciplinary Collaboration in Early Childhood
Language Development in Children with Special Needs
Large Scale Assessment and Accountability and Students with Special Needs
Literacy Instruction for Students with Special Needs
Math Instruction for Children with Special Needs
Overrepresentation of Culturally and Linguistically Diverse Students in Special Education
Paraprofessionals in Special Education
Parent and Family Involvement in the Education of Children with Special Needs
Peer Relations and Socialization of Children and Adolescents with Special Needs
Peer-Tutoring
Postsecondary Participation of Students with Special Needs
Prereferral Strategies
Preventing Antisocial Behavior and Delinquency: A Comprehensive School-wide Approach
Programs and Instructional Strategies for Students with Gifts and Talents
Related Services for Children with Special Needs
Risk and Resilience Frameworks in Understanding Special Education
School Dropout Prevention

School-Based Services for Children with Special Needs

Self-Determination

Social Competence

The Identification of Students with Gifts and Talents

Transition from School to Adult Life

Visual Impairment

Workforce Issues in Special Education

Adapted Physical Activity for Students with Special Needs

S R Hodge, The Ohio State University, Columbus, OH, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Adapted physical activity – A comprehensive term encompassing, but not limited to, physical education, exercise, sport, recreation, dance, creative arts, medicine, rehabilitation, nutrition, and leisure to enhance physical activity goal achievement for persons of all ages with special needs due to various impairments, movement or activity limitations, and participation restrictions.

Adapted physical education – This describes educational services delivered to school-age children and youth with special needs, typically 3–21 years of age, in physical education contexts.

Culturally responsive (relevant) pedagogy – This promotes holistic student achievement while enabling and empowering students to accept and affirm their cultural identity as they develop critical understandings that challenge inequities in educational, legal, and social institutions.

Curricular adaptations – These denote any modification to the general education curriculum used to maximize student learning and performance.

Curriculum – The knowledge, skills, and learning experiences intentionally designed and provided to students within the school program.

Inclusion – This describes the many variations of educating children and youth with disabilities in general education settings with peers without disabilities. Specific to physical education, inclusive classes should provide all students access to high-quality physical activity experiences that promotes personal growth and achievement, teaches respect for difference, values diversity, and creates a sense of community in a supportive school culture.

Individualized education program (IEP) – As required by US law, the IEP document identifies a student's entire educational curriculum goals for each school year, and defines how the student will be

supported to reach those goals. An IEP sets the occasion for individualized instruction in general and inclusive settings. The IEP mandate was introduced in Public Law 94-142, the Education for All Handicapped Children Act of 1975 (now PL 108-446, the Individuals with Disabilities Education Improvement Act, 2004) for all students with disabilities to ensure a free and appropriate public education.

Individualized instruction – The personalized, student-centered instruction (e.g., teacher and student one-on-one interaction), which enables each student to progress at her or his own pace with consideration to the student's disability type and severity, abilities, learning style, culture, support needs, and any other relevant variables.

Instructional format – The ways in which teachers organize and deliver instruction and construct practice for students within various curricula.

Instructional modifications – The modifications made to the presentation of information, content, activity, or other instructional material.

Instructional task system – This refers to the subject-matter activity of physical education, the intended learnings students are to acquire by participating in the instructional activities.

Itinerant adapted physical education teacher – A teacher who travels from school to school to provide services to students with disabilities and special needs.

Managerial task system – The organizational and behavioral aspects of physical education – all the nonsubject-matter functions necessary for students and teachers to exist together over a period of time.

Student–social task system – The intentions for social interaction that students seek in physical education.

Task progressions – These refer to learning tasks that move students from less complex and sophisticated tasks to more difficult and complicated tasks by adding complexity and difficulty. Effective progressions require the blending of content knowledge and pedagogical knowledge such that teachers can communicate progressions to students through a series of informing, refining, extending, and applying instructional tasks within a lesson or unit, and across time.

Task system – A regularized pattern for accomplishing tasks. Common to physical education are three interactive task systems: (1) managerial task system, (2) instructional task system, and (3) student–social system. These task systems form the ecology of physical education.

According to the World Health Organization (WHO), there are an estimated 650 million people with disabilities worldwide. Disability is a complex term covering impairments, activity limitations, and participation restrictions. Impairment denotes a problem in an individual's body function or structure. Activity limitation denotes a difficulty encountered by an individual in executing a task or action. A participation restriction is a problem experienced by an individual involved in life situations (WHO, 2009). Disability represents interaction between features of a person's being and the systems of the society in which he or she lives (see WHO, 2009). In legal terms, for example, under Public Law 101-336, the Americans with Disabilities Act (1990), disability is defined as an impairment that substantially limits one or more major life activities.

Over the years, laws have been enacted in many countries that guarantee the rights of people with disabilities to education and other important areas of life. For example, in Ghana, West Africa, the Persons with Disability Act (Act 715) of 2006 was enacted. Act 715 covers areas such as civil rights, employment, education, transportation, healthcare and facilities, and more (Republic of Ghana, 2008). In the Asia-Pacific region, the Japanese government enacted Special Support Education policy in 2007 and guarantees an appropriate special support education to meet the unique needs of each youth, regardless of the type or severity of his or her disability (Ministry of Education, Culture, Sports, Science and Technology, 2002). In the United Kingdom, the Special Educational Needs and Disability Act 2001 addresses the education of persons with disabilities (Office of Public Sector Information, 2008). Likewise in the United States, Public Law 108-446, the 2004 Individuals with Disabilities Education Improvement

Act (IDEA) identifies 14 disability categories under which youth may be eligible for educational services. The disability labels and definitions guide how educational entities determine who is eligible for special education and related services. IDEA (2004) defines special education as “specially designed instruction, at no cost to parents, to meet the unique needs of a child with a disability, including (a) instruction conducted in the classroom, in the home, in hospitals and institutions, and in other settings; and (b) instruction in physical education” (p. 118). This is important because it means in essence that all students being served in special education must receive physical education instruction.

Importance of Physical Education

In 2005, the 2nd World Summit on Physical Education was held in Magglingen, Switzerland. The Summit organizers called for stronger international status of physical education in school politics and in everyday school life as essential and integral for education as well as for human and social development (International Council of Sport Science and Physical Education, 2005). Today in schools around the world, professionals work in collaboration to make school life meaningful for individuals with disabilities. Adapted physical education specialists and physical education generalists are among those who work with individuals with special needs in schools.

Physical education is a medium for guiding youth in the process of living a physically active life. To ensure high-quality physical education, teachers and service providers must develop instructional goals and design task progressions that lead to student achievement. High-quality physical education programs promote knowledge, positive social interactions, skill acquisition and improvement, regular moderate to vigorous physical activity (MVPA), and healthy sports. These activities are essential for children and youths' health and well-being and constitute important components of a healthy lifestyle, along with healthy dietary habits, tobacco-free living, and avoidance of other substances harmful to health (WHO, 2003). The WHO took the position that individuals with disabilities should be given ample opportunities and support to engage in physical activities and sports adapted to their conditions. Such opportunities are commonplace for students with special needs in physical education.

Physical Education for Students with Special Needs

Adapted physical education is the term commonly used to describe educational services delivered to school-age children and youth with special needs in physical education.

It is an individualized program of physical and motor fitness, fundamental motor skills and patterns, skills in aquatics and dance, and individual and group games and sports designed for students with special needs. These programs are usually designed and delivered by adapted physical education specialists or physical education generalists. Consistent with the least restrictive environment (LRE) concept, stated in IDEA, physical education may take place in integrated (e.g., general education) to segregated (e.g., self-contained) classes. Each school year, physical education goals must be identified in a student's individualized education program (IEP) in accord with the provisions of IDEA.

In the United States, physical education teachers are required to participate in the IEP process. A primary role is to evaluate, plan, and instruct students in the motor domain. Likewise in accord with Section 504 of Public Law 93-112, the Rehabilitation Act of 1973 (1973) and its amendments, an accommodation plan should be developed by a school-based assessment team for students with special needs (whether or not they have identified disabilities), but who are nonetheless in need of specialized services and accommodations. Sherrill (2004), a renowned scholar in adapted physical activity, asserted that the overarching aim of adapted physical education is to promote personal growth, self-actualization, and empowerment through meaningful physical activity experiences.

Historical Foundation

The history of the adapted physical education profession is a progression from medical gymnastics before 1905 through a transition to competitive sport, to an emphasis on corrective exercises to general physical education (Sherrill, 2004). In 1952, the American Association for Health, Physical Education, and Recreation gave official support to shifting curricular emphasis for students with disabilities from corrective exercises to adapted physical education. The common terminology changed from corrective exercises and gymnastics to physical education, and school-based curricula increasingly included sports, games, dance, and aquatics. In 1954, Arthur Daniels, professor at The Ohio State University (OSU) wrote the first textbook specific to adapted physical education. Since that time, many scholars worldwide have contributed to our understanding of teaching students with special needs in physical activity settings (Block, 2007; Doll-Tepper *et al.*, 1990; Dunn and Leitschuh, 2005; Hodge *et al.*, 2003; Jansma and French, 1994; Sherrill, 2004; Steadward *et al.*, 2003; Van Coppenolle *et al.*, 1993; Winnick, 2005).

During the 1950s, training clinics for professional preparation in working with children and youth with disabilities were introduced such as the OSU's Gym and Swim by Daniels in 1954 and the Children's Physical Development Clinic at the University of Maryland by

Johnson in 1957. In the 1960s, the Kennedy family facilitated professional preparation workshops on intellectual disabilities (cognitive disabilities) across the United States, founded the Special Olympics, and promoted the enactment of laws initiating professional preparation in physical education and recreation for children and youth with cognitive disabilities (Sherrill, 2004). In the 1970s, critical legislation was passed to support the education of students with disabilities. Of importance in the United States, Public Law 94-142, Education for All Handicapped Children Act (EHA) of 1975 was passed (now IDEA of 2004). Also during this era, leaders continued to emerge in adapted physical education teacher preparation. For example, Joseph P. Winnick, State University of New York (SUNY) at Brockport, founded the first master's level specialization in adapted physical education. In Canada, Patricia Austin, University of Alberta, pioneered in establishing teacher preparation in this field (Sherrill, 2004). In 1973, leaders from Canada and Belgium more broadly conceptualized the scope of adapted physical education to adapted physical activity and founded the International Federation of Adapted Physical Activity (IFAPA) in Quebec, Canada.

IFAPA is an international cross-disciplinary professional organization of individuals, institutions, and agencies supporting, promoting, and disseminating knowledge and information about adapted physical activity, disability sport, and all aspects of sport, movement, and exercise science for the benefit of persons who require adaptations to enable their participation (IFAPA, nd).

According to the bylaws of the IFAPA, adapted physical activity is a cross-disciplinary body of knowledge directed toward the identification and solution of individual differences in physical activity. It is a service-delivery profession and an academic field of study that supports an attitude of acceptance of individual differences, advocates access to active lifestyles and sport, and promotes innovation and cooperative service delivery programs and empowerment systems. Adapted physical activity includes, but is not limited to, physical education, sport, recreation, dance and creative arts, nutrition, medicine, and rehabilitation (IFAPA, nd).

Today, the IFAPA board has regional representatives from Africa, Asia, Europe, Middle East, Oceania, North America, and South/Central America. IFAPA's president Yeshayahu Hutzler and former president Claudine Sherrill asserted that in some countries adapted physical activity is not the preferred term, instead such descriptions as "sports for the disabled, sport therapy, and psychomotor therapy pertain to similar understandings" (Hutzler and Sherrill, 2007). They call for more communication and collaboration internationally across these areas in research and information sharing. Scholarly journals are an important medium for communication, research collaboration, and information sharing. In 1984, founded by Geoffrey D. Broadhead, the *Adapted Physical Activity*

Quarterly (APAQ) was first published. APAQ is now the premier international journal of adapted physical activity scholarship. It is the official publication of IFAPA and is published by Human Kinetics. Important also, information sharing on physical activity pedagogy for persons with disabilities has occurred across countries over the years. For example, textbooks such as *Adapted Physical Activity, Recreation and Sport: Crossdisciplinary and Lifespan* by Claudine Sherrill, and *Early Movement Experiences and Development* by Joseph P. Winnick have been translated from English to Chinese and Japanese, respectively. Since the 1990s, there has been advocacy and movement toward inclusive (integrated) programming where students with and without disabilities are educated together. Internationally, inclusion has become a progressive educational practice with support from educational and political advocates, policymakers, and legislators. Ideally, inclusion represents the placement of students who have various disabilities with proper accommodations and supports in integrated classes with classmates without disabilities for educational and social benefits. This philosophical shift from separate to integrated programs is about equitable opportunities, appropriate supports and accommodations, and the empowerment of students with disabilities. The emphasis is on student self-actualization and empowerment through meaningful physical activity experiences (Sherrill, 2004). Self-actualization and empowerment are situated in adaptation and humanistic theories.

Theoretical and Philosophical Underpinnings

Humanistic philosophy and adaptation theory have influenced our understanding of teaching children and youth with disabilities in physical education. Humanistic philosophy promotes individual uniqueness, morality, positive social relations, and defines success as individuals seeking to accomplish their personal best. In this philosophy, physical education teachers (specialists and generalists) typically endorse an abilities-based pedagogical approach, which emphasizes abilities, not disabilities and holistic and student-centered praxis. Adaptation theory suggests that teachers must develop instructional goals and design task progressions that create optimal learning opportunities within their particular class environments. This theory denotes the process by which teachers, students, and the environment reciprocally influence one another in dynamic and, oftentimes, in meaningful ways. In addition to inclusive, humanistic, and adaptation orientations, also supportive of student empowerment are culturally responsive (relevant) pedagogies.

In recent years, there has been increasing advocacy for culturally responsive pedagogy in physical education. Scholars have called for infusing diversity training (including disability content) across physical education teacher education (PETE) programs to ensure that

teacher candidates are culturally competent to teach a diversity of students. Gay (2000), a pioneer of culturally responsive pedagogy, explained that it has the following components:

- It acknowledges the legitimacy of the cultural heritages of different ethnic groups, both as legacies that affect students' dispositions, attitudes, and approaches to learning and as worthy content to be taught in the formal curriculum.
- It builds bridges of meaningfulness between home and school experiences as well as between academic abstractions and lived sociocultural realities.
- It uses a wide variety of instructional strategies that are connected to different learning styles.
- It teaches students to know and praise their own and each others' cultural heritages.
- It incorporates multicultural information, resources, and materials in all subjects and skills routinely taught in schools. (p. 29)

Earlier, Ladson-Billings (1995) stated that culturally relevant pedagogy empowers students through (1) academic success, (2) cultural competence, and (3) critical consciousness. Culturally responsive pedagogy uses the "cultural knowledge, prior experiences, frames of references, and performance styles" of culturally, ethnically, linguistically, and economically diverse students "to make learning encounters more relevant to and effective for them" (Gay, 2000: 29). Teachers and service providers, who implement culturally relevant teaching, empower students as they help them develop intellectually, socially, emotionally, and politically by using cultural referents to impart knowledge, skills, and attitudes (Ladson-Billings, 1992). Student empowerment translates into academic competence, confidence, and a will to act (Gay, 2000).

Effective Physical Education Pedagogy

The extant literature on teaching effectiveness describes specific teacher acts linked to achievement outcomes of students in physical education. To paraphrase, Siedentop and Tannehill (2000) noted that effective teachers should

1. have a belief in their own efficacy;
2. allocate sufficient time and opportunity to learn, and cover appropriate content;
3. communicate high, realistic expectations so that students can receive adequate instruction and practice time to learn their roles;
4. establish positive approaches to class management and student engagement;
5. design meaningful, success-oriented tasks;
6. create and sustain a brisk pace and maintain momentum;

7. communicate content with clear, brief demonstrations and explanations, followed by sufficient guided practice, and feedback for understanding;
8. actively supervise students' progress and practice;
9. hold students accountable for appropriate participation;
10. communicate with clarity and enthusiasm, and exhibit equitable support of all students; and
11. use students' inputs and ongoing assessment to inform their practice.

Clearly, effective teachers and service providers reflect to inform their practice. Moreover, educators must insist on teaching students with various disabilities with heightened emphasis on adaptations, modifications, and supports. Physical education teachers must have a sense of self-assurance that their teaching is effective.

Effective teaching means designing lessons to maximize the amount of time each student spends in direct practice at a level that ensures a continuing development of the skill compatible with a minimal number of errors (Webster, 1993). As Webster identified, a number of important indicators of effective teaching are applicable to teaching students with disabilities in physical education. These include:

1. finding ways to keep students appropriately engaged in subject matter for a high percentage of the time and do so without resorting to coercive, negative, or punitive behavioral techniques;
2. developing and maintaining positive class climate, whereby students have many opportunities to practice at levels appropriate to their abilities;
3. individualizing instruction for success-oriented and on-task behaviors of students, which includes adapting and adjusting strategies to match contextual variables;
4. using peer tutors to assist and give additional attention, which allows greater opportunities for students to practice;
5. using appropriate reinforcement strategies to motivate students;
6. providing congruent, specific skill and behavior feedback; and
7. organizing practices in a manner that promotes student learning and retention.

The time students spend in lesson activity should be directed at skill acquisition with clear accountability measures on learning outcomes through active participation. In lesson planning, teachers and service providers must design learning experiences that ensure their students actively participate in fun movement activities. It is important that teachers focus on students having fun while they move. However, they must also hold students accountable for correct forms of movement toward skill acquisition, development, or proficiency.

Instructional Strategies for Teaching Students with Special Needs

It is important that physical education teachers use curricular and instructional approaches that support inclusive and culturally responsive pedagogy. Many such curricular approaches are common to physical education using various instructional formats. These curricular approaches include the Fitness for Life model, I CAN and ABC models, Moving to Inclusion, Personal and Social Responsibility model, and Sport Education. In many countries, teacher educators encourage teachers to structure their curriculum and pedagogies around national standards for physical education. The National Association for Sport and Physical Education (NASPE, 2004) provided six national standards for physical education. These standards and essential components of a quality physical education program are presented in **Table 1**.

Physical education teachers can use different instructional formats to organize and deliver instruction and construct practice opportunities for students within various curriculum models. Some examples of instructional formats that could be used to support inclusive and culturally responsive pedagogy include active teaching, adventure education, teaching through questions, task teaching and learning stations, peer tutoring and reciprocal teaching, flexible groupings, and cooperative learning as well as self-instructional formats such as contract teaching and personalized systems of instruction to individualized instruction.

Active Teaching

Active teaching, also known as direct instruction, is an instructional format where the teacher controls the pace of the lesson through direct instruction and student practice opportunities. Direct instruction to the whole class, small groups, or individual students is followed by student guided practice and error correction. Next, students engage in independent practice where their activity is actively supervised. Important also, active teaching promotes a supportive climate where high, yet realistic, expectations are communicated to students and they are held accountable for performance and learning outcomes (Siedentop and Tannehill, 2000).

Adventure Education

Adventure education is an educational approach designed to teach students to solve problems and take responsibility for the outcomes of their decisions. In this model, the teacher serves as a facilitator of experiential learning. Its goal is to educate the whole child by developing his/her mental, physical, emotional, and social attributes.

Table 1 NASPE content standards and essential components of a quality physical education program*NASPE content standards^a*

Standard 1	Demonstrates competency in motor skills and movement patterns needed to perform a variety of physical activities
Standard 2	Demonstrates understanding of movement concepts, principles, strategies, and tactics as they apply to the learning and performance of physical activities
Standard 3	Participates regularly in physical activity
Standard 4	Achieves and maintains a health-enhancing level of physical fitness
Standard 5	Exhibits responsible personal and social behavior that respects self and others in physical activity settings
Standard 6	Values physical activity for health, enjoyment, challenge, self-expression, and/or social interaction

Components of a quality physical education (PE) program^b

Component 1	Is organized around content standards that offer direction and continuity to instruction and evaluation
Component 2	Is student centered and based on the developmental urges, cultures, tendencies, and interests of students
Component 3	Has physical activity and motor-skill development at its core
Component 4	Teaches management skills and self-discipline
Component 5	Emphasizes inclusion of all students
Component 6	Emphasizes instruction focused on the process of learning rather than performance outcomes
Component 7	Teaches lifetime activities that students can use to promote their health and personal values
Component 8	Teaches cooperative and responsibility skills and helps students develop sensitivity to multiple diversities

^aAdapted from NASPE (National Association for Sport and Physical Education) (2004). *Moving into the Future: National Standards for Physical Education*, 2nd edn. Reston, VA: National Association for Sport and Physical Education, p iv, with permission of NASPE Publications.

^bAdapted from Darst, P. W. and Pangrazi, R. P. (2006). *Dynamic Physical Education for Secondary School Students*, 5th edn. San Francisco, CA: Pearson Education, p 2.

Adventure education encourages risk taking based on holistic learning concepts to develop or improve self-awareness, confidence, self-esteem, emotional stability, and social skills through social interactions and group cooperation. This might include problem solving through group work and challenge activities (e.g., wall climbing, mountain biking, ropes course, and orienteering).

Teaching through Questions

Teaching through questions (Socratic method) is an instructional approach whereby tasks are communicated to students through questions that present scenarios that

guide student work toward specific goals or questions that pose problems to be solved. In this approach, teachers often use guided discovery or problem-solving activities. Teaching through questions is a variation of active teaching (Siedentop and Tannehill, 2000).

Task Teaching

Task teaching is an instructional format in which the class session is organized such that different students are engaged in different tasks at the same time, typically rotating among task or learning stations throughout the lesson. This format can accommodate student diversity in terms of skill level in particular, which holds well for students with disabilities as it sets the occasion for individualized instruction. Teachers and service providers should design tasks that have multiple levels of performance criteria to meet student skill and functional ability levels. They can use brief, simple descriptions on task cards, task posters (usually with pictures or drawings), and self-instructional materials to communicate the task for each learning station (Siedentop and Tannehill, 2000).

Peer Tutoring and Reciprocal Reaching

Peer tutoring and reciprocal reaching refer to instructional formats that use student pairings, triads of students, or small groups to provide basic instruction and guided practice. Students support one another, and engage in reciprocal teaching; and some will serve as tutors to classmates (e.g., higher-skilled students tutor lower-skilled students) in providing individualized attention and instruction, peer assessment, feedback, and social support (Siedentop and Tannehill, 2000). The use of peer tutoring helps improve achievement levels of students with disabilities.

Flexible Groupings

Flexible groupings, group variation, and small group instructional formats usually involve small groups of students working together on a common goal. The teacher identifies group member's responsibilities to shape the focus of their work to accomplish the assigned task (Siedentop and Tannehill, 2000). In this format, it is important that student work is structured such that there is interdependence in the achievement of group goals, as well as individual student accountability. In flexible groupings, students can be grouped or paired based on goals, interests, needs, or skills such that they have opportunities to share and learn from one another.

Cooperative Learning

Cooperative learning helps teachers put cooperation into daily practice in physical education. It is a teaching

approach where small groups or teams, each with students of different levels of ability (and potentially of different disabilities), use a variety of learning activities to improve their understanding of a skill, task, strategy, or content. Each team member is responsible for her/his own learning outcomes and for helping teammates learn, which creates an atmosphere of achievement. Students work cooperatively through the assigned tasks until all group members understand and complete them successfully.

Self-Instructional Format

Self-instructional formats set the occasion where students can progress through a sequence of learning experiences with purpose and direction. For students with disabilities, a teacher's knowledge, will, and ability to individualize instruction are essential. Individualized instruction is a personalized, student-centered instructional format, which enables each student to progress at her or his own pace with consideration to the student's disability type and severity, abilities, learning style, culture, support needs, and any other relevant variables. This involves task-analyzing skills and assessing performance, adapting lessons, modifying games and activities, accommodating the environment, and supporting each student with disabilities as needed. Individualized instruction results in improved physical activity engagement for students with disabilities. Individualizing instruction is recognized globally as beneficial to students with special needs. For instance, Japan's Special Support Education policy mandates the development of individual teaching plans for students with disabilities, which calls for individualized instruction (Ministry of Education, Culture, Sports, Science and Technology, 2002).

To individualize instruction, a physical education teacher must be able to task-analyze each student's motor skills by breaking his/her motor skills down into logical progressions, typically using sequential simple to complex skill progressions. In lesson planning, the teacher uses individualized task analysis to construct logical learning progressions based on the student's functional level, motor abilities, learning style and tendencies, and interests. I CAN is an individualized instructional model, which refers to individualized instruction, create social leisure competence, associate all learnings, and narrow the gap between theory and practice (Wessel, 1977). The Achievement-Based Curriculum (ABC) model is a refinement of Wessel's I CAN model (Wessel and Kelly, 1986). Other forms of self-instructional formats include contract teaching. This teaching format sets the occasion for individualized instruction whereby "students sign a learning contract to complete a sequence of learning tasks according to a predetermined set of criteria" (Siedentop and Tannehill, 2000: 299). Personalized system of instruction (PSI) is also a self-instructional format that can serve as a

medium for individualized instruction and student progress. Typically, the teacher or service provider will divide the unit content into small units of instruction wherein students are expected to meet specific performance criteria; and when successful, they then move on to additional units of instruction. In short, teachers (1) develop specific instructional tasks and explicit performance criteria, (2) allow students to practice the tasks until they meet the criteria, and (3) move then on to the next task (Siedentop and Tannehill, 2000). Clearly, effective physical education teachers who teach students with disabilities plan lessons, adapt, modify, and use supports to individualize instruction and promote student learning. In addition to exhibiting content knowledge and pedagogical skillfulness, a teacher must demonstrate skill as an efficient and effective manager of student behaviors.

Organization and Management

Class organization and management are integral aspects to teaching students with disabilities in physical education. A positive learning context typically (1) is well-planned and organized, (2) has rules and routines that promote on-task active participation, (3) has class activities that are set at the appropriate skill level for each student, and (4) has a class that promotes interesting activity experiences with smooth and short transition times. In creating a positive learning environment, teachers must establish rules and routines at the beginning of the school year. Establishing rules and routines helps students know what to do, what not to do, and when to do or not do something. At the start and periodically during the school year, the teacher discusses and posts the rules publically. In concert with establishing rules and routines to manage behavior, there are several other proactive organization and management strategies applicable to physical education. For example, setting up equipment before class starts is an important organizational strategy. Managing student behavior and addressing safety issues are usually handled by the teacher actively supervising students, emphasizing rules and routines, and enforcing established rules and routines. In addition to finding ways to create fun activities for students, teachers often use cooperation (teamwork) and promote socialization (good sports behaviors) as important lesson objectives. These objectives can be encouraged through thoughtful task structures, especially with station activities and team games. During their class sessions, teachers should regularly provide students with encouragement, and specific and corrective feedback. They should seek to establish good rapport with them by opening up to them and being approachable. Ensuring the safety of students is also a critical responsibility for all teachers. To avoid injuries and liability, teachers and service providers must be

mindful to ensure safety of the environment, equipment, and students (e.g., safety of students from collisions with one another as they participate in class activities).

Physical education environments can lead to challenges requiring the implementation of various behavior-management strategies. Behavior management must be taught to create a positive learning context that can help to prevent or minimize inappropriate student behaviors. There are a number of procedures used to influence student behavior. To increase desirable behaviors, teachers may use positive and negative reinforcement (Table 2). In addition, they may use a variety of reinforcement strategies to help maintain, strengthen, or teach new behaviors, which include shaping, chaining, token economy, behavioral contracting, and Premack principle. To decrease inappropriate or undesirable behaviors, teachers may use punishment (e.g., response cost, verbal reprimand, overcorrection, and time-out procedures), planned ignoring or extinction (Table 2), as well as reinforcement of other behavior, reinforcement of incompatible behavior, reinforcement of low response rates, and behavioral contracts or proclamations.

The psychodynamic and humanistic approaches to behavior management focus on teaching students the importance of respecting self and others and accepting responsibility through (1) understanding psychological causes of behavior, (2) developing a trusting relationship

between the teacher and student, and (3) teaching students self-control and social responsibility. Hellison's (1995) Personal and Social Responsibility model is a popular model in physical education. In this model, the student's responsibilities are to

1. recognize that his/her own behavior is a problem,
2. recognize why it is a problem,
3. understand what motivated the behavior, and
4. determine through discussion with the teacher (and schoolmates were appropriate) alternative ways to behave in similar situations.

Additional psychodynamic-humanistic approaches to behavior management include reality therapy, conflict resolution (e.g., talking bench), counseling methods, and positive coaching. These approaches to behavior management help students to analyze their own behaviors, respect others, plan strategies to change, and evaluate a behavior-change program. Ultimately, these approaches teach students to take responsibility for their own behaviors and promote positive social relationships.

Professional Preparation and Development

Internationally, professional qualification criteria for physical education teachers (specialists and generalists) vary by country. It is commonly held that adapted physical education specialists primary roles include (1) resource teachers who are assigned to a particular school, and team teach with generalist physical education teachers; (2) itinerant teachers who travel to and from different schools to provide direct instruction; and (3) consulting teachers who travel from school to school and do some direct service tasks but primarily consult with and provide support services to general education teachers and help them adapt instruction. In addition, these specialists may conduct in-service training (e.g., workshops for school district personnel). To properly carry out their job-related roles and responsibilities, physical education teachers need training whereby they can acquire skills and knowledge to meet the needs, interests, and abilities of their students.

Worldwide, PETE programs typically offer a 4-year course of study curriculum that culminates with student teaching (internships) at local elementary, middle, and/or high schools. In Ghana (West Africa), for example, two of the six public universities offer degree courses or programs in physical education. These are the University of Education at Winneba and the University of Cape-Coast. Both institutions are in the Central Region of Ghana and have oversight of the undergraduate and graduate programs in physical education. In addition, the Department of Health, Physical Education, Recreation and Sports at the University of Education at Winneba has a 2-year

Table 2 Descriptions of basic behavior management strategies

<i>Terminology</i>	<i>Description</i>
Stimulus control	A measurable event that might have an influence on behavior
Reinforcement	A stimulus event that increases or maintains the frequency of a behavioral response
Positive reinforcement	A behavioral strategy whereby a teacher offers a student something valued as a consequence of the student emitting a desired behavior, resulting in an increase in the frequency of the behavior by the student in the future
Negative reinforcement	A behavioral strategy defined as using reinforcement to increase a desired behavior by encouraging a student to perform that particular behavior in order to avoid something she or he dislikes
Punishment	A behavioral strategy carried out by either the presentation of an aversive stimulus or the removal of a positive stimulus The intent is to weaken or eliminate an undesirable behavior
Planned ignoring or extinction	A behavioral strategy where the teacher withholds reinforcement after a behavioral response that had previously been reinforced, which results in cessation of the particular behavior

diploma course of study program in coaching. In general, the physical education undergraduate course of study is a 4-year program typified with 4 weeks to 1 year of student teaching. Student teaching internship placements are at the basic education (elementary school), junior high school, or senior high school level. Prior to student teaching, physical education teacher candidates engage in on-campus teaching (micro-teaching) experiences at the university. At the University of Education at Winneba, students have three subject areas (i.e., major, minor, and cognate).

Scholars continue to raise issues of curriculum priorities of PETE programs. These revolve around professional preparation in training teachers to work effectively along with students with special needs. How teachers are prepared, and what happens during their professional careers in regard to their efficacy in teaching students with special needs are important questions. These questions must be considered in making decisions on policy, curriculum change, and teacher preparation and development. For example, PETE programs should ensure teacher candidates engage in micro-teaching episodes with small groups of students with disabilities (e.g., during practicum experiences) such that they develop a sense of personal mastery before teaching in larger classes as in their student teaching internships. Such strategies would build teachers' self-efficacy.

Obviously, teacher-preparation programs should follow competency-based guidelines. In 1993, the Adapted Physical Activity Council (APAC) of the American Alliance for Health, Physical Education, Recreation and Dance issued competency-related guidelines for adapted physical education teacher preparation. APAC's main purpose is advocacy for persons with disabilities through promotion of programs, policies, standards, and research (AAHPERD, 2008). There is pressing need and legal authority in special education to establish and maintain fully accredited teacher-education programs and to ensure ongoing professional development training to better prepare physical education teachers for working effectively with students with special needs. For practicing physical education teachers, district-wide professional development training focused on inclusive and culturally responsive practices in physical education should be made available on a regular basis. Such training would prepare teachers in a better manner to more effectively teach students with disabilities, give them an enhanced sense of confidence, and allow them to learn more about individualizing instruction to more effectively engage all students. Professional development activities should be offered and designed to equip teachers with advanced knowledge and skills necessary to teach students with various disabilities. Therefore, school districts should regularly engage teachers in relevant professional development training focused on teaching students with disabilities in physical education. Further, school districts should hold

teachers accountable for participating in professional development. Holding teachers accountable will promote a deeper appreciation for teaching students with disabilities as teachers' overall motivations and teaching effectiveness are likely to be enhanced.

Conclusion

Over the years, laws have been enacted in many countries that guarantee the rights of people with disabilities to education, including high-quality physical education. In schools worldwide, devoted physical educators (specialists and generalists) work to make school life meaningful for students with special needs. Physical education is a medium for guiding children and youth in the process of living a physically active life. In my opinion, effective teachers must reflect on their practice, thoughtfully plan lessons, adapt, modify, and individualize instruction to promote student learning. They must also exhibit skill as efficient and effective managers of student behavior. Physical education teachers may use different instructional formats to organize and deliver instruction and construct practice opportunities for students within various curriculum models. It is important that they use curricular and instructional approaches that support inclusive and culturally responsive pedagogy.

Unquestionably, there is a need to prepare future physical education teachers (specialists and generalists) to competently provide culturally responsive and informed instruction. Teacher candidates should have multiple practicums and field-based teaching experiences in a diversity of physical activity contexts. In our world's societies, teachers will have increased opportunities to teach culturally, ethnically, economically, and linguistically diverse students. They must acquire a strong experiential knowledge base to best serve all students. In the same vein, school districts should hold physical education teachers (both specialists and generalists) accountable for participating in professional development activities relevant to teaching students with special needs. In both policy and practice, accountability criteria should be built into annual teacher evaluations. This would help ensure that teachers appreciate the importance of regularly participating in professional development workshops, attending conferences, and engaging in continued learning to enhance their overall teaching efficacy. This would be a cost-effective means of helping to maintain a well-prepared teaching workforce.

See also: Cooperative Learning for Children with Special Needs; Direct Instruction and the Education of Children with Special Needs; Inclusion; Instructional Accommodations for Children with Special Needs In Inclusive Settings; Peer-Tutoring; Physical Education and Sports.

Bibliography

- AAHPERD (American Alliance for Health, Physical Education, Recreation and Dance) (2008). <http://www.aahperd.org> (accessed August 2009).
- Americans with Disabilities Act (1990). Pub. L. No. 101-336. <http://www.ada.gov/pubs/ada.htm> (accessed August 2009).
- Block, M. E. (2007). *A Teacher's Guide to Including Students with Disabilities in General Physical Education*, 3rd edn. Baltimore, MD: Paul H. Brookes.
- Doll-Tepper, G., Dahms, C., Doll, B., and von Selzam, H. (1990). *Adapted Physical Activity: An Interdisciplinary Approach*. Berlin: Springer.
- Dunn, J. M. and Leitschuh, C. (2005). *Special Physical Education*, 8th edn. Dubuque, IA: Kendall/Hunt.
- Education for All Handicapped Children Act of 1975 (1975). PL 94-142, 20 U. S. C. §§ 1400 et seq.
- Gay, G. (2000). *Culturally Responsive Teaching – Theory, Research and Practice*. New York: Teachers College Press.
- Hellison, D. (1995). *Teaching Responsibility through Physical Activity*. Champaign, IL: Human Kinetics.
- Hodge, S. R., Murata, N. M., Block, M. E., and Lieberman, L. J. (2003). *Case Studies in Adapted Physical Education: Empowering Critical Thinking*. Scottsdale, AZ: Holcomb Hathaway.
- Hutzler, Y. and Sherrill, C. (2007). Defining adapted physical activity: International perspectives. *Adapted Physical Activity Quarterly* **24**, 1–20.
- IFAPA (International Federation of Adapted Physical Activity) (nd). <http://www.ifapa.biz> (accessed August 2009).
- Individuals with Disabilities Education Improvement Act (2004). Pub. L. No. 108-446, Sec. 602, 118 Stat. 2657.
- International Council of Sport Science and Physical Education (2005). *Report on the 2nd World Summit on Physical Education and ICSP Meeting*. <http://www.icsspe.org> (accessed August 2009).
- Jansma, P. and French, R. (1994). *Special Physical Education, Physical Activity, Sports and Recreation*. Englewood Cliffs, NJ: Prentice-Hall.
- Ladson-Billings, G. (1992). Reading between the lines and beyond the pages: A culturally relevant approach to literacy teaching. *Theory into Practice* **31**, 312–320.
- Ladson-Billings, G. (1995). But that's just good teaching! The case for culturally relevant pedagogy. *Theory into Practice* **34**, 159–165.
- Ministry of Education, Culture, Sports, Science and Technology (2002). *Special Support Education in Japan – Education for Children with Special Needs*. http://www.mext.go.jp/a_menu/shotou/tokubetu/03110701/002.pdf (accessed August 2009).
- NASPE (National Association for Sport and Physical Education) (2004). *Moving into the Future: National Standards for Physical Education*, 2nd edn. Reston, VA: National Association for Sport and Physical Education.
- Office of Public Sector Information (2008). *Special Educational Needs and Disability Act 2001*. <http://www.opsi.gov.uk> (accessed August 2009).
- Rehabilitation Act of 1973 (1973). PL 93-112, 29 U. S. C. §§ 701 et seq.
- Republic of Ghana (2008). *Persons with Disability Act, 2006, Act 715*. <http://www.gdgh.org> (accessed August 2009).
- Sherrill, C. (2004). *Adapted Physical Activity, Recreation and Sport: Crossdisciplinary and Lifespan*, 6th edn. Dubuque, IA: McGraw-Hill.
- Siedentop, D. and Tannehill, D. (2000). *Developing Teaching Skills in Physical Education*, 4th edn. Mountain View, CA: Mayfield.
- Steadward, R. D., Wheeler, G. D., and Watkinson, E. J. (2003). *Adapted Physical Activity*. Edmonton, AB: The University of Alberta Press.
- Van Coppenolle, H., Simons, J., Neerincx, E., Vanlanfjewick, Y., and Verwilt, M. (1993). *European Master's Degree Adapted Physical Activity Textbook*. Leuven: Acco.
- Webster, G. E. (1993). Effective teaching in adapted physical education. *Palaestra* **9**(3), 25–27; 29–31.
- Wessel, J. (ed.) (1977). *Planning Individualized Education Programs in Special Education*. Northbrook, IL: Hubbard.
- Wessel, J. and Kelly, L. (1986). *Achievement-Based Curriculum in Physical Education*. Philadelphia, PA: Lea and Febiger.
- WHO (World Health Organization) (2003). *Health and Development through Physical Activity and Sport*. <http://www.who.int/en> (accessed August 2009).
- WHO (World Health Organization) (2009). <http://www.who.int> (accessed August 2009).
- Winnick, J. P. (2005). *Adapted Physical Education and Sport*, 4th edn. Champaign, IL: Human Kinetics.

Further Reading

- AADPERD (American Association for Health, Physical Education and Recreation) (1952). Guiding principles for adapted physical education. *Journal of Health, Physical Education and Recreation*, **23**, 15.
- Active Living Alliance for Canadians with a Disability (1994). *Moving to Inclusion*. Gloucester, ON: Active Living Alliance for Canadians with a Disability.
- Akuffo, P. B. and Hodge, S. R. (2008). The roles and responsibilities of adapted physical education teachers in an urban school district. *Education and Urban Society* **40**(2), 243–268.
- Ammah, J. O. A. and Hodge, S. R. (2005). Secondary physical education teachers' beliefs and practices in teaching students with severe disabilities: A descriptive analysis. *High School Journal* **89**(2), 40–54.
- Broadhead, G. D. (1984). Birth of a journal. *Adapted Physical Activity Quarterly* **1**, 1–2.
- Burden, J., Jr., Hodge, S. R., O'Bryant, C. P., and Harrison, L., Jr. (2004). From colorblindness to intercultural sensitivity: Infusing diversity training in PETE programs. *Quest* **56**, 173–189.
- Corbin, C. B. and Lindsey, R. (2005). *Fitness for Life*, 5th edn. Champaign, IL: Human Kinetics.
- Daniels, A. S. (1954). *Adapted Physical Education*. New York: Harper and Brothers.
- Darst, P. W. and Pangrazi, R. P. (2006). *Dynamic Physical Education for Secondary School Students*, 5th edn. San Francisco, CA: Pearson Education.
- DePauw, K. P. and Gavron, S. J. (1995). *Disability and Sport*. Champaign, IL: Human Kinetics.
- Grineski, S. (1996). *Cooperative Learning in Physical Education*. Champaign, IL: Human Kinetics.
- Hodge, S. R. and Akuffo, P. (2007). Adapted physical education teachers' concerns in teaching students with disabilities in an urban public school district. *International Journal of Disability, Development and Education* **54**(4), 399–416.
- Hodge, S. R., Ammah, J. O. A., Casebolt, K., LaMaster, K., and O'Sullivan, M. (2004). High school general physical education teachers' behaviors and beliefs associated with inclusion. *Sport, Education and Society* **9**(3), 395–419.
- Hodge, S. R., Davis, R., Woodard, R., and Sherrill, C. (2002). Comparison of practicum types in changing preservice teachers' attitudes and perceived competence. *Adapted Physical Activity Quarterly* **19**, 155–171.
- Hodge, S. R. and Jansma, P. (1999). Effects of contact time and location of practicum experiences on the attitudes of physical education majors. *Adapted Physical Activity Quarterly* **16**, 48–63.
- Hodge, S. R., Kozub, F. M., Robinson, L. E., and Hersman, B. L. (2007). Reporting gender, race, ethnicity, and sociometric status: Guidelines for research and professional practice. *Adapted Physical Activity Quarterly* **24**, 21–37.
- Hodge, S. R., Tannehill, D., and Kluge, M. A. (2003). Exploring the meaning of practicum experiences for PETE students. *Adapted Physical Activity Quarterly* **20**, 381–399.
- International Federation of Adapted Physical Activity (2009). *Countries*. <http://www.ifapa.biz> (accessed August 2009).
- Lavay, B. W., French, R., and Henderson, H. L. (2006). *Positive Behavior Management in Physical Activity Settings*, 2nd edn. Champaign, IL: Human Kinetics.
- Loovis, E. M. (2005). Behavioral disabilities. In Winnick, J. P. (ed.) *Adapted Physical Education and Sport*, 4th edn., pp 155–172. Champaign, IL: Human Kinetics.
- National Dissemination Center for Children with Disabilities (2008). *Categories of Disability under IDEA Law*. <http://www.nichcy.org> (accessed August 2009).

- Sato, T., Hodge, S. R., Murata, N. M., and Maeda, J. K. (2007). Japanese physical education teachers' beliefs about teaching students with disabilities. *Sport, Education and Society* 12(2), 211–230.
- Siedentop, D. (ed.) (1994). *Quality PE through Positive Sport Experiences – Sport Education*. Champaign, IL: Human Kinetics.
- Siedentop, D., Hastie, P. A., and Van der Mars, H. (2004). *Complete Guide to Sport Education*. Champaign, IL: Human Kinetics.
- Timken, G. L. (2005). Teaching all kids: Valuing students through culturally responsive and inclusive practice. In Lund, J. and Tannehill, D. (eds.) *Standard-Based Physical Education Curriculum Development*, pp 78–99. Sunbury, MA: Jones and Bartlett.

Relevant Websites

- <http://www.humankinetics.com> – Human Kinetics, Journal Mission.
- <http://www.icsspe.org> – International Council of Sport Science and Physical Education.
- <http://www.ifapa.biz> – International Federation of Adapted Physical Activity.
- <http://www.ncperid.org> – National Consortium for Physical Education and Recreation for Individuals with Disabilities.
- <http://www.who.int> – World Health Organization.

Assessment Accommodations for Children with Special Needs

R J Kettler and S N Elliott, Vanderbilt University, Nashville, TN, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Alternate assessment – A set of methods (e.g., portfolio assessment, performance assessment, and comprehensive rating scales) that can be used to assess students with special needs, severe disabilities, or with extremely limited English proficiency.

Assessment accommodations – The changes in the standard assessment process made because an individual's disability requires changes for the test to be a valid measure.

Differential effects or boost – An attribute of assessment accommodations and the concept that if an accommodation is working appropriately, then it should increase the scores of students with disabilities much more than it increases the scores of students without disabilities.

Individual need – An attribute of assessment accommodations stipulating that students are to be provided assessment accommodations based on individual functional impairments, rather than based on group disability status.

Modification – The changes to a test that may alter the construct being measured, or at least the degree of complexity of the construct being measured, and are useful for students who have disabilities severe enough to preclude taking the typical assessment.

Proficiency – The level of competency or understanding of knowledge in an academic area.

Sameness of inference – An attribute of assessment accommodations stipulating that test alterations result in scores being more comparable to make decisions.

Unchanged construct – An attribute of assessment accommodations stipulating that assessment accommodations should not change the construct being measured, but should only provide access to measurement of that construct for students with disabilities.

school districts across the United States to involve only students without disabilities in school-wide assessments. Reasons for excluding students with disabilities from school-wide assessments, likely, included fears that performances of students with disabilities would reflect poorly on the school, students with disabilities would be frustrated by the testing process and content, and the tests typically used to measure proficiency would not accurately reflect proficiency of students with disabilities. The concern with regard to unfavorable comparisons with other schools should be alleviated by current law (i.e., the Individuals with Disabilities Education Improvement Act) requiring all students with and without disabilities to be included in annual proficiency testing to the extent possible. However, concerns that tests used for students without disabilities will cause more frustration and be less valid when used for students with disabilities remain relevant.

To include students with disabilities in large-scale accountability systems, educators have a number of options, including (1) having students with disabilities complete the same achievement test that is taken by students without disabilities without any changes made to the test setting, format, or content; (2) having students with disabilities complete the same achievement test with some changes to the test setting and format; (3) having students with disabilities be evaluated by an alternate form of assessment; or (4) evaluating students with disabilities by some combination of the aforementioned methods. The focus of this article is the second option – having students complete the same achievement test but with changes to the setting and format (i.e., accommodations).

When considering the challenge of validly including students with disabilities in large-scale accountability systems, one should be familiar with the terms alternate assessment, modifications, and assessment accommodations. Alternate assessment is an umbrella term for a set of methods (e.g., portfolio assessment, performance assessment, and comprehensive rating scales) that can be used to measure the abilities of students with severe disabilities or with extremely limited English proficiency. Modifications are changes to a test that may alter the construct being measured or, at least, the degree of complexity of the construct being measured, and are useful for students who have disabilities severe enough to preclude taking the typical assessment. Assessment accommodations are departures from the typical assessment that are not as extreme as either alternate assessment or modifications.

Assessment Accommodations for Children with Special Needs

Until the passage of the Individuals with Disabilities Education Act, in 1997, it was regular practice in most

Assessment accommodations are defined as changes in the standard assessment process made because an individual's disability requires changes for the test to be a valid measure. Effective assessment accommodations remove barriers that would otherwise hinder students with disabilities from demonstrating their true abilities.

Elliott *et al.* (2002) used the analogy of eyeglasses or contact lenses, needed for a driving test, to clarify the concept of a testing accommodation. On a hypothetical driving test that does not allow eyeglasses, a perfectly capable driver with no visual impairment would likely pass, whereas a capable driver with a visual impairment would likely fail. In the case of the driver with no visual impairment, the driving test would yield a valid score indicating driving ability. In the case of the person with the visual impairment, the barrier created by that disability would prevent the test from yielding a valid score indicating the individual's driving ability. It is important to note that having the person with no visual impairment wear eyeglasses during the test would not help that person perform better (in fact, that person might perform worse), and would not make the test score a more valid indicator of that person's driving ability. Accommodations – similar to the eyeglasses in this example – are alterations to the assessment procedure that should be helpful only to the individuals for whom they are designed.

Assessment Accommodations and the Law

The 1997 reauthorization of the Individuals with Disabilities Education Act indicated that all students must be included in large-scale school-accountability systems – to the extent possible. The 2001 No Child Left Behind Act increased the number of grade levels at which the achievement of all students must be measured. These policies emphasized the issue of how to meaningfully include and accurately assess students with disabilities in large-scale assessment programs. In recent years, there has been steady growth in the number of research studies focusing on testing accommodations. The reauthorization of the 2004 Individuals with Disabilities Education Improvement Act reinforced that students with disabilities should be provided access to the general education curriculum and its associated assessment procedures to the maximum degree appropriate for each individual. In fact, inclusion within the general education-testing program, likely, influences access to the general education curriculum.

While national legislation mandates the availability of testing accommodations to help students with disabilities to meaningfully participate in statewide and district-wide assessments, decisions with regard to which accommodations are permissible and appropriate are made at the state

and local levels. For example, while federal law indicates that testing accommodations must be made available for students with disabilities, as of 2003, accommodations were actually made available for all students in 12 states (Arizona, California, Hawaii, Kansas, Minnesota, New Hampshire, Oregon, Pennsylvania, Rhode Island, Tennessee, Vermont, and Wyoming). Individual discrepancies in state allowance of accommodations may contribute to the wide range among states with regard to the percentage of students who receive accommodations (between 8% and 82%). Perhaps no assessment accommodation is more controversial – based on variations in law from state to state – than having the questions on a test read aloud to students. While three states (Massachusetts, Missouri, and Vermont) allow this accommodation with no restrictions, five other states (Hawaii, Illinois, Kansas, Nebraska, and Ohio) do not allow this accommodation under any circumstances. The remainder of the states allows questions on a test to be read aloud to students only under certain circumstances, and in 14 of those cases an allowance has implications for scoring. Further information with regard to the types of accommodations allowable in one's state is available from the National Center on Educational Outcomes (NCEO), which maintains a website (included in this article) dedicated to state policies on testing accommodations.

Types of Assessment Accommodations

Assessment accommodations are typically classified within four categories: (1) accommodations in the presentation format, (2) accommodations in the recording or response format, (3) accommodations in the assessment environment or setting, and (4) accommodations in the timing or scheduling. Accommodations in the presentation format are changes in the way that the directions or content of a test is presented to students, and include oral reading, large print, or sign language. Accommodations in the recording or response format of a test offer students different options to provide answers on a test, and include use of a computer scribe, writing answers directly into a test booklet, or use of a calculator. Accommodations in the assessment environment or setting are changes to the place in which the test is given, and include individual administration, testing in the student's home, or providing noise buffers. Accommodations in timing or scheduling make the timing of an assessment more flexible, and include extended time, frequent breaks, or testing over multiple days. **Table 1** contains a list of accommodations organized by the aforementioned categories.

Gibson *et al.* (2005) performed a study of the most-often-recommended accommodations, and, in doing so, utilized two helpful frameworks for identifying and categorizing accommodations. They used the Assessment

Table 1 Assessment accommodations by category*Accommodations in presentation format*

Use an enlarger to facilitate vision of material
 Assist the student to track test items by pointing or placing the student's finger on items
 Oral reading (either by an adult or an audio device)
 Use audio amplification devices (e.g., hearing aids)

Accommodations in the recording or response format

Use an adult to record the student's response
 Use a computer board, communication board, or tape recorder to record responses
 Have the student respond directly in the test booklet rather than on an answer sheet
 Use organizational devices, calculators, or spelling and grammar-assistive devices

Accommodations in the assessment environment or setting

Administer the test in a small group or individual session
 Place the student in a room or part of a room where he or she is most comfortable
 Allow a special education teacher or aide to administer the test
 Test in the student's home

Accommodations in timing or scheduling

Administer a test in shorter sessions with more breaks or rest periods
 Space testing sessions over several days
 Administer a test at a time most beneficial to the student
 Allow the student more time to complete a test
 Change the testing schedule or the order of subjects

Accommodations Checklist (AAC) to help teachers, guided by individualized education programs (IEPs), to recommend assessment accommodations for students with disabilities. The AAC is a list of 67 assessment accommodations organized into eight categories. The tool was designed to assist teachers and IEP teams in selecting accommodations for use during testing. The other framework used by Gibson and colleagues was CTB–McGraw Hill's taxonomy for characterizing accommodations by the likelihood of test-score interpretation being changed. According to the taxonomy, testing accommodations should be divided into three categories: Category-1 accommodations are not thought to impact the interpretation of a test; Category-2 accommodations may affect the interpretation of either norm- or criterion-referenced tests; and Category-3 accommodations may affect the interpretation of both norm- and criterion-referenced tests – by altering the construct being measured.

Gibson *et al.* (2005) found that five Category-1 accommodations (i.e., provide a distraction-free space, arrange for a special education teacher to administer the test, provide verbal encouragement of efforts, reread directions for each subtask, and administer practice activities) were among the ten most-recommended accommodations. Of these ten accommodations, the two most commonly recommended accommodations (i.e., provide extra time and read directions) were both classified in Category 2. Two accommodations from Category 3 (i.e., simplify

language in directions and restate question with more appropriate vocabulary) were also included among the ten accommodations. One accommodation (read questions and content to student) was classified as being in either Category 2 or Category 3 – perhaps reflecting the controversial nature of reading accommodations across states and tests. When considering these findings, it is critical to remember that accommodations are rarely made in isolation, and that the argument with regard to whether an accommodation or package of accommodations should actually affect the interpretation of a score is construct specific, and is best resolved via empirical investigation.

Assessment Accommodations and Validity

At its core, the use of assessment accommodations is a validity issue. Appropriate assessment accommodations can change a test that yields valid scores for students without disabilities into a test that yields equally valid scores both for students with disabilities and students without disabilities. The intent is not just to give the scores of students with disabilities a boost, but rather to provide access to a test that will accurately measure these students' abilities. For example, suppose that a math test with story problems is given to a classroom of students – most of whom do not have disabilities, but several of whom have disabilities that affect their reading. While the average score on the test for the class is an 85, the average score for the students who have disabilities in reading is 70. One might suspect that – due to the reading intensity of this test – students with reading disabilities were not able to show what they knew in mathematics. A simple fix that would make the scores for students with disabilities appear more similar to the scores for students without disabilities would be to add 15 points to the score of each student who has a disability, making the mean scores for each group equal. This would not be an appropriate assessment accommodation. The aforementioned attempt to make scores of students with disabilities comparable to those of students without disabilities would not be appropriate because it would not change the degree to which the scores of students without disabilities reflect those students' respective abilities in mathematics. While the maneuver would make the two sets of scores seem more similar, because their means would be equal, this change would be superficial. The scores for students with disabilities would still provide the same information provided prior to the conversion.

As with any test score, the scores for aforementioned students with disabilities would be influenced by two types of variance: construct-relevant variance and construct-irrelevant variance. Construct-relevant variance is the

degree to which the true variance of the construct being measured – mathematical ability in this case – contributes to the total variance in a set of scores. Construct-irrelevant variance is the degree to which variance in factors other than the variable being measured (e.g., reading ability, concentration, test-taking skills) contribute to variance within a set of scores. Test scores are likely to be more reliable and the validity argument is better served when construct-relevant variance is maximized in comparison to construct-irrelevant variance. In the current example, the higher the proportion of the variance in the test scores that can be attributed to actual differences in student math abilities, the more valid the test is for measuring this construct. When 15 points are added to the score of each student with a disability, neither the variance of the scores nor the portion that is construct relevant versus construct irrelevant is changed. This does not, likely, accomplish the goal of using assessment accommodations to level the playing field by providing access to a fair test. Perhaps, our subsample of students with reading disabilities is a group of high performers in math, and if given a mathematics test without a reading load, they may have scored an average of 95 points. In this case, the 15 bonus points would not have been sufficient compensation. It is also possible that the group would have scored an average of 75 on a valid mathematics test, and that 15 bonus points is overcompensation. The most likely scenario is that individuals within the group with reading disabilities were not all equally affected by the reading load on the test, and that no single, universal *post hoc* operation would make their scores comparable. Therefore, the best way to make the scores comparable between students with disabilities and those without disabilities is through assessment accommodations – selected for individuals to reduce the amount of construct-irrelevant variance associated with each test score. For one student with a disability, this might mean more time is necessary to read the test. Another student might need to be allowed to read the test aloud. A third student with a disability might be able to best show her mathematics ability if the test is read aloud to her. Any of these accommodations could reduce construct-irrelevant variance from a source (i.e., reading ability) that would likely affect students with disabilities in reading differently than it would affect those without disabilities.

One source of evidence for determining whether a set of accommodations is working to produce valid scores is whether a differential boost exists between how scores are enhanced for students with disabilities versus those without disabilities. A differential boost is the concept that indicates: if an accommodation is working appropriately, it should increase the scores of students with disabilities much more than it increases the scores of those without disabilities. **Figure 1** depicts the ideal differential boost – showing scores from students with and students without

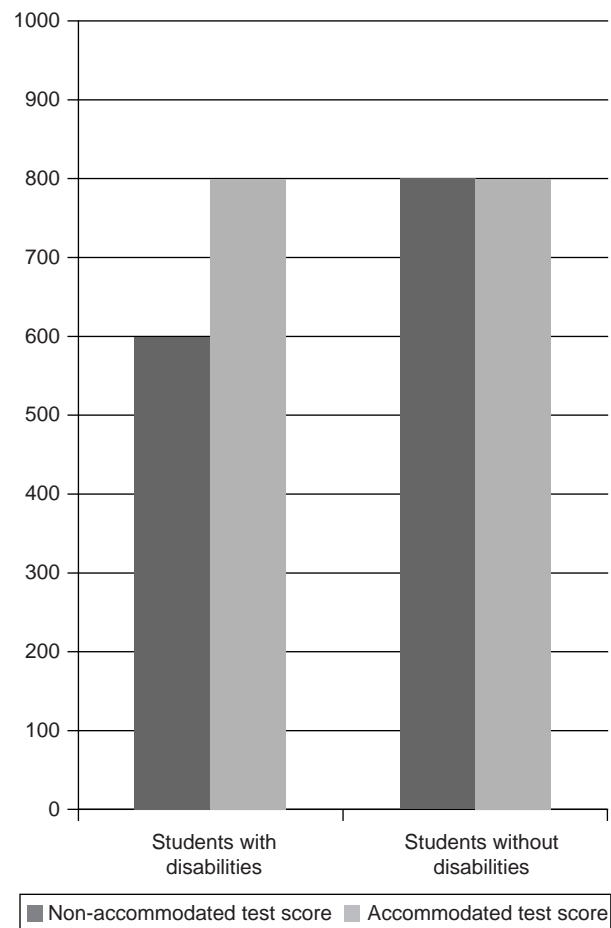


Figure 1 Ideal differential boost for valid assessment accommodations.

disabilities taking tests in accommodated and nonaccommodated conditions. Poor accommodations (e.g., adding 15 points to a score) would affect scores from persons without disabilities in the same way that they would affect scores from persons with disabilities. Effective accommodations are helpful to students with disabilities, and are unnecessary (e.g., reading aloud a mathematics test), or even harmful (e.g., wearing prescription eyeglasses) to persons who do not need them. A differential boost associated with an accommodation is one type of evidence that construct-irrelevant variance is being reduced.

Differential effects – or the aforementioned differential boost – are one of four attributes suggested by Hollenbeck *et al.* (2002) that help define appropriate accommodations. The other three attributes are unchanged constructs, individual need, and sameness of inferences. The attribute of unchanged constructs stipulates that assessment accommodations should not change the construct being measured, but should only provide access to measurement of that construct for students with disabilities. To evaluate this attribute, one must have a clear understanding of the

construct that the test is intended to measure. For example, having a reader may be a perfectly acceptable accommodation on a test designed to measure calculation, but it would not be an acceptable accommodation on a test designed to measure reading comprehension. The attribute of individual need stipulates that students are to be provided assessment accommodations based on individual functional impairments, rather than based on group disability status. While one student diagnosed with attention-deficit hyperactivity disorder (ADHD) might benefit from an extended time accommodation, because – after attending to distractions – she eventually returns to completing a test, another student with ADHD granted extra time may become increasingly frustrated with the length of the test session. Highlighted text and frequent breaks might be more appropriate for him. The sameness of inference attribute indicates that test alterations result in scores being more comparable to make decisions. A student with a disability – by receiving appropriate accommodations on a test – should obtain a score that is suitable for making inferences with regard to proficiency.

The Individualization of Accommodations

The Individuals with Disabilities Education Act requires that assessment accommodations be individualized, rather than recommended based on a disability category. IEPs are drafted – for all students with disabilities – by education teams that include regular and special education teachers, parents, test interpreters, and any service providers. Because of their involvement in the IEP process, teachers are often influential in determining which accommodations are appropriate for each individual student. However, current research indicates that teachers are not always accurate predictors of which students will benefit from testing accommodations or which testing accommodations will be helpful.

Fuchs and Fuchs (2001) developed a data-driven method to find individualized testing accommodations. The Dynamic Assessment of Testing Accommodations (DATA) is a brief instrument designed to help teachers determine whether an accommodation works for an individual student – by testing the student both with and without the accommodation. Available for both reading and mathematics tests, DATA is designed to test the effects of individual accommodations on individual students. The key determinant – with regard to whether an accommodation is considered valid for the student – is the presence of the differential boost. An accommodation is considered valid for the student with a disability if his or her score improves substantially more than the mean improvement found among students without disabilities in the normative sample. The system has been found to improve upon teacher selection of individualized accommodations.

State of the Field: Moving Forward

Thurlow *et al.* (2000) wrote a monograph for the NCEO to address the state of assessment accommodations research. The researchers indicated that future research should (1) focus on accommodations of most interest, (2) focus on students who comprise a large part of the population needing accommodations, (3) use students without disabilities as a group in any comparison designs, and (4) collect other measures to help clarify findings. Thurlow *et al.* (2000) identified four group research designs – all including participants taking tests under both accommodated and nonaccommodated conditions – to address the matter of whether assessment accommodations work. They also suggested item response theory, factor analysis, and criterion-related analyses, as well as a number of single-subject designs, as possible contributors to future assessment accommodations research. Their final suggestion was that work on assessment accommodations needs to be done as part of a program: “Even with the best research designs, we will not get nice answers unless we have a program of research to follow up with additional questions” (p. 28). Two such active programs of research – Fuchs, Fuchs, and colleagues and S. Elliott and colleagues – are discussed next.

Fuchs, Fuchs, and colleagues did a number of research studies on assessment accommodations – in addition to their development of DATA. In a meta-analytic study, Tindal and Fuchs (2000) found that the most effective accommodations included providing large print or Braille for the visually impaired, as well as reading problems aloud for students with disabilities in math. Rather than examining accommodations in packages, Fuchs *et al.* (2006) indicated that research should be done on one accommodation at a time – with work on packages only commencing after individual accommodations have been validated. Findings from their studies included that (1) extended time does not provide a differential boost, (2) reading a test orally helps students with disabilities more than it does students without disabilities, (3) heterogeneity among students with disabilities reduces the likelihood of one set of accommodations being appropriate for the entire group, (4) student demographic characteristics may influence teachers’ accommodation recommendations, and (5) students with severe disabilities and severe reading deficits may benefit from reading a test aloud. Based on a review of assessment-accommodations literature, Fuchs and colleagues concluded that decisions with regard to testing accommodations must be individualized, that the meaningfulness of test scores must be of primary importance when selecting accommodations, and that the mandated inclusion of students with disabilities in educational outcomes testing makes this area of research critical.

In the same vein, Elliott and colleagues did a number of studies utilizing both between-subjects and within-subjects designs – requiring participants to take assessments

in accommodated and nonaccommodated conditions. Accommodations packages used in these studies were selected using student IEPs and the AAC. Elliott *et al.* (2001) found an effect size for students with disabilities (0.88) much larger than the effect size for students without disabilities (0.44) on a performance assessment. Schulte *et al.* (2001) used the same design with a research version of TerraNova's standardized math test – finding that accommodations had a large differential effect on multiple-choice questions, but had no such effect on constructed-response questions. McKevitt and Elliott (2003) found that teacher-recommended accommodations from the AAC did not help students with disabilities or students without disabilities, and a read-aloud accommodation combined with teacher accommodations boosted the scores of both groups, when the outcome measure was a research version of TerraNova's standardized reading test. Kettler *et al.* (2005) also found a strong differential boost for fourth-grade students on the research version of the TerraNova reading test. The effect size for students with disabilities was 0.42, while the effect size for students without disabilities was 0.13. Lang *et al.* (2007) replicated this finding, and determined that students endorsed the fairness of testing accommodations, particularly when used among students with disabilities. Feldman *et al.* (in press) explored the relationship among the effects of testing accommodations, students' self-efficacy, and test performance. They determined that students with disabilities experience a differential boost in self-efficacy and motivation when provided accommodations, and that this boost corresponds with increases in test scores.

Sireci and Pitoniak (2007) published a review of assessment accommodations research that condensed the findings of previous reviews, recent studies, and studies not addressed in earlier meta-analyses. They concluded that assessment accommodations, sometimes help students with disabilities, help students without disabilities sometimes, and must be selected on an individual basis. They also emphasized the necessity of identifying the construct that a test is designed to measure to evaluate whether a specific accommodation is appropriate, and commended research in which students take tests in both accommodated and nonaccommodated conditions. With regard to the two most widely researched testing accommodations, Sireci and Pitoniak concluded that extra time is an appropriate accommodation when speed of response is not a construct being measured, and that having a test read aloud to a student is appropriate on tests that do not assess reading.

Conclusion

While recent years have featured a great deal of research activity on assessment accommodations, a number of

questions remain to be answered. Studies following the postpositive ideal of isolating and validating individual accommodations – coupled with studies following the pragmatic ideal of evaluating accommodations in packages and incorporating student feedback – will continue connecting pieces to solve from different sides the assessment accommodations puzzle. Perhaps, new techniques – such as the aforementioned item response theory, factor analysis, and criterion-related analyses (Thurlow *et al.*, 2000) – will provide breakthroughs in identifying individualized accommodations that provide access – for students with disabilities – to fair and valid assessment. Until such breakthroughs are made, teachers and IEP teams will rely on their knowledge of individual students and their familiarity with current assessment accommodations research and theory to select accommodations that make scores from students with disabilities comparable to those obtained from students without disabilities.

Bibliography

- Elliott, S. N., Kratochwill, T. R., and McKevitt, B. C. (2001). Experimental analysis of the effects of testing accommodations on the scores of students with and without disabilities. *Journal of School Psychology* **39**(1), 3–24.
- Elliott, S. N., McKevitt, B. C., and Kettler, R. J. (2002). Testing accommodations research and decision making: The case of "good" scores being highly valued but difficult to achieve for all students. *Measurement and Evaluation in Counseling and Development* **35**, 153–166.
- Feldman, E., Kim, J. S., and Elliott, S. N. (in press). The effects of accommodations on adolescents, self-efficacy and test performance. *Journal of Special Education*.
- Fuchs, L. S. and Fuchs, D. (2001). Helping teachers formulate sound test accommodation decisions for students with learning disabilities. *Learning Disabilities Research and Practice* **16**, 174–181.
- Fuchs, L. S., Fuchs, D., and Capizzi, A. M. (2006). Identifying appropriate test accommodations for students with learning disabilities. *Focus on Exceptional Children* **37**(6), 1–8.
- Gibson, D., Haeberli, F. B., Glover, T. A., and Witter, E. A. (2005). The use of recommended and provided testing accommodations. *Assessment for Effective Intervention* **31**, 19–36.
- Hollenbeck, K., Rozek-Tedesco, M., and Finzel, A. (2000). Defining valid accommodations as a function of setting, task, and response. *Paper Presented at the Annual Meeting of the Council for Exceptional Children*. Vancouver, BC, Canada.
- Kettler, R. J., Niebling, B. C., Mroch, A. A., *et al.* (2005). Effects of testing accommodations on math and reading scores: An experimental analysis of the performance of fourth- and eighth-grade students with and without disabilities. *Assessment for Effective Intervention* **31**(1), 37–48.
- Lang, S. C., Elliott, S. N., Bolt, D. M., and Kratochwill, T. R. (2007). The effects of testing accommodations on students' performances and reactions to testing. *School Psychology Quarterly* **14**(3), 1–28.
- McKevitt, B. C. and Elliott, S. N. (2003). The use of testing accommodations on a standardized reading test: Effects on scores and attitudes about testing. *School Psychology Review* **32**(4), 583–600.
- Schulte, A. G., Elliott, S. N., and Kratochwill, T. R. (2001). Effects of testing accommodations on students' standardized mathematics test scores: An experimental analysis. *School Psychology Review* **30**(4), 527–547.
- Sireci, S. G. and Pitoniak, M. J. (2007). Assessment accommodations: What have we learned from research? In Laitusis, C. C. and Cook, L. L.

- (eds.) *Large Scale Assessment and Accommodations: What Works?* pp 53–65. Arlington, VA: Council for Exceptional Children.
- Thurlow, M. L., McGrew, K. S., and Tindal, G., et al. (2000). Assessment accommodations research: Considerations for data analysis, *Tech. Rep. No. 26*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Tindal, G. and Fuchs, L. S. (2000). *A Summary of Research on Test Changes: An Empirical Basis for Defining Accommodations*. Lexington, KY: Mid-South Regional resource Center Interdisciplinary Human Development Institute.

Further Reading

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Barton, K. E. (2007). Validity and accommodations: The journey toward accessible assessments. In Laitusis, C. C. and Cook, L. L. (eds.) *Large-Scale Assessment and Accommodations: What Works?* Arlington, VA: Council for Exceptional Children.
- Clapper, A. T., Morse, A. B., Lazarus, S. S., Thompson, S. J., and Thurlow, M. S. (2005). 2003 state policies on assessment participation and accommodations for students with disabilities, *Synthesis Report 56*. Minneapolis, MN: University of Minnesota, National Center on Education Outcomes.
- CTB/McGraw-Hill (1997). *TerraNova Multiple Assessments Battery*. Monterey, CA: CTB/McGraw-Hill.
- CTB/McGraw-Hill (2000). *Guidelines for Using the Results of Standardized Tests Administered Under Nonstandard Conditions*. Monterey, CA: CTB/McGraw-Hill.
- Elliott, S. N., Braden, J. P., and White, J. (2001). *Assessing One and All: Educational Accountability and Students with Disabilities*. Alexandria, VA: Council for Exceptional Children.
- Elliott, S. N., Kratochwill, T. K., and Schulte, A. G. (1999). The assessment accommodations checklist: Facilitating decisions and documentation in the assessment of students with disabilities. *Teaching Exceptional Children* **31**(2), 10–14.
- Elliott, S. N. and Roach, A. T. (2007). Alternate assessments of students with significant disabilities: Alternative approaches, common technical challenges. *Applied Measurement in Education* **20**(3), 301–333.
- Fuchs, L. S., Fuchs, D., Eaton, S. B., et al. (2000). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children* **67**(2), 67–81.
- Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C., and Kams, K. (2000). Supplementing teacher judgments of test accommodations with objective data sources. *School Psychology Review* **29**, 65–85.
- Helwig, R. and Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Exceptional Children* **69**, 211–225.
- Hollenbeck, K. (2002). Determining when test alterations are valid accommodations or modifications for large-scale assessment. In Tindal, G. and Haladyna, T. (eds.) *Large Scale Assessment Programs for All Students*, pp 109–148. Mahwah, NJ: LEA.
- Individuals with Disabilities Education Act Amendments of (1997). Pub. L. No. 105-17, 111 Stat. 37 (codified as amended at 20 U.S.C. § 1400 et seq.).
- Lazarus, S. S., Thurlow, M. L., and Eisenbraun, K. D., et al. (2006). *State Accommodations Policies: Implications for the Assessment of Reading*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Luke, S. D. and Schwartz, A. (2007). Assessment and accommodations. *Evidence for Education* **1**(2), 1–12.
- Phillips, S. E. (1994). High-stakes testing accommodations: Validity versus disabled rights. *Applied Measurement in Education* **7**, 93–120.
- Thompson, S. J. and Thurlow, M. L. (1999). *1999 State Special Education Outcomes: A Report on State Activities at the End of the Century*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- US Department of Education (2001). *No Child Left Behind*. <http://www.ed.gov/nclb/landing.jhtml> (retrieved December, 2009).
- US Department of Education (2004). *Individuals with Disabilities Education Improvement Act of 2004*. <http://www.ed.gov/policy/speced/guid/idea/idea2004.html#law> (accessed August 2009).

Relevant Websites

- <http://www.wcer.wisc.edu> – Assessing One and All.
- <http://www.cehd.umn.edu/nceo> – National Center on Educational Outcomes.
- <http://www.specialconnections.ku.edu> – Special Connections.
- <http://www.ed.gov> – US Department of Education.
- <http://www.osepideasthatwork.org> – US Department of Education – IDEAs that Work.

Assistive Technology and Educational Practice

H P Parette and G R Peterson-Karlan, Illinois State University, Normal, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Assistive-technology device – Any item, piece of equipment, or product system, whether acquired commercially or off the shelf, modified, or customized, that is used to increase, maintain, or improve functional capabilities of individuals with disabilities.

Assistive-technology service – Any service that directly assists an individual with a disability in the selection, acquisition, or use of an assistive-technology device.

Cognitive demands – The amount of thinking required for performing tasks or for using a tool.

Compensation – To substitute for or perform a function (using a tool) that cannot be done by a person with a disability.

Consideration – A federally mandated team decision-making process focusing on selection, implementation, and outcomes monitoring of assistive technology for a student with disabilities.

Instructional technology – The technologies that are used to teach new skills, remediate deficiencies in skills learned, and/or to supplement or expand the curriculum for all students.

Outcomes – The effect or result of using an assistive technology tool.

Physical demands – The amount of muscle strength and movement needed to initiate, pursue, and complete a task or use a tool.

Social-linguistic demands – The amount of symbolic interpretation and processing required of a user to use a tool or complete a task.

Ubiquitous – Readily available or widely used.

Deeply embedded in the fabric of society is use of a wide array of technologies. Merriam-Webster, Inc. (2007) defined technology as both “a practical application of knowledge” and “a manner of accomplishing a task.” When viewed broadly, then, Peterson-Karlan and Parette (2008) observed that technology includes not only the things individuals use to accomplish tasks (e.g., overhead projectors, computers, cell phones, copy machines, dictionaries, and spreadsheet software), but also the procedures for how individuals use or operate them, and how they systematically proceed to accomplish tasks (i.e., use of strategies, such as rules for text messaging; dictionary

use rules; and writing steps of required tasks on a whiteboard). Acceptance of this broader context of what technology is – things, procedures, and strategies that help individuals complete tasks – is integral to understanding and effectively working with twenty-first-century students – both with and without disabilities – and education professionals (Peterson-Karlan and Parette, 2007).

Many technologies in society have become ubiquitous, that is, they are so prominent and accepted that they appear to be everywhere and used routinely by many people. Not many individuals would argue that ubiquitous technologies (e.g., light switches, cell phones, the Internet, information search software, electronic calculators, and word processors with spell-check features) are unimportant given their impact on the society’s quality of life, particularly with regard to how they increase effectiveness, efficiency, and comfort or ease of doing things, as well as productivity in the workplace. Sadly, however, the broad acceptance and inherent use of technology that permeates the broader world context, particularly in work settings, are much less frequently observed in school environments (Peterson-Karlan and Parette, 2008).

Presented in **Table 1** are three categories of primary technologies used in today’s schools accompanied by descriptions and examples. These categories, broadly, include (1) information and communication technology (ICT); (2) instructional technology (IT); and (3) assistive technology (AT).

Medical technologies, though a specific technology category by itself, are not included given the limited relationship to educational milieus. More detailed elaboration on the first two categories may be found elsewhere (cf., Grabe and Grabe, 2007; Lengel and Lengel, 2006; Roblyer, 2006; Smaldino *et al.*, 2005). In general, however, neither ICT nor IT by themselves enable all students to learn and participate in the general education curriculum (see **Figure 1**).

Students with disabilities are now included in general education classroom settings to a greater extent than ever before. Yet, these students often demonstrate gaps in their ability to perform at expected levels in various academic and life-skill areas (see **Figure 2**). Given the current pressures on public schools for all children to achieve (No Child Left Behind Act, 2001), the potential for technology to provide access to and support learning for students with disabilities and narrow the gap between abilities and expected educational performance levels has received considerable attention (International Society

Table 1 Categories of technology impacting schools

Category	Description	Examples
Information and communication	Helps to communicate and interact with others; produce work; problem-solve; find information; and manage ourselves, our homes and lives more efficiently and effectively across settings	Copy machines; word processing and graphics software; Microsoft® PowerPoint; cell phones; text messaging; blogs; Wikis; and the Internet
Instructional	Increases instructional <ul style="list-style-type: none"> • effectiveness (learning in a better way than without the experience) • efficiency (same amount of learning occurs but in a shorter time) • appeal (increases possibility that students will devote time and energy to the learning task) 	Computers; DVDs; projection devices; digital audio and video recording or editing devices or software used by teachers to prepare or present information; whiteboards; and educational software
Assistive	Compensates for difficulty in accomplishing functional tasks at an expected level of performance	Foam pencil grips; visual schedules; graphic organizers; electronic communication systems; wheelchairs; hearing devices; text-to-speech software; talking word processors; and seating and positioning systems

Instructional technology

Used to:

- Teach or learn functional skills
- Supplements or expands the curriculum
- Remediation of learning functional tasks



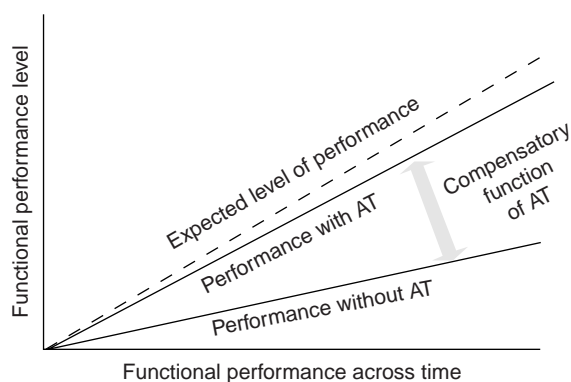
Assistive technology

Used to:

- Compensate for inability to perform functional tasks
- Help students do things they could not do at an expected performance level

Students can learn or perform functional tasks without IT

Students cannot learn or perform functional tasks without AT

Figure 1 Distinctions between IT and AT. From SEAT Center, ©2007. Understanding Assistive Technology.**Figure 2** Performance gap exhibited by students with disabilities across time, with AT compensating and closing the gap. From SEAT Center, ©2007.

for Technology in Education, 2005a, 2005b; SEAT Center, National Center for Technology Innovation (NCTI), and University of Kansas (KU), 2006; Thurlow *et al.*, 2007). For students with disabilities, AT (1) must be considered by education professionals when developing any individual education plan (IEP; Individuals with Disabilities

Education Improvement Act of 2004 (IDEIA, 2004); 20 U.S.C. 1401 § 614(B)(v)); and (2) is often required to ensure their inclusion and effective participation in the general education curriculum (Center for Technology in Education, Johns Hopkins University; and Technology and Media Division (TAM) of the Council for Exceptional Children, 2005).

Blackhurst (2005) noted that AT was developed for people with disabilities to (1) make the environment more accessible, (2) assist them in learning, (3) enable them to compete in the workplace, or (4) enhance their independence or otherwise improve their quality of life. The IDEIA 2004 defines an AT device as “any item, piece of equipment or product system, whether acquired commercially or off the shelf, modified, or customized, that is used to increase, maintain, or improve functional capabilities of individuals with disabilities” (20 U.S.C. § 1401(251)). There are now more than 25 000 AT items, equipment, software, and product services (Abledata, as cited in Edyburn, 2000) to consider for use with over 6 million students of ages 6–21 with disabilities. However, AT also includes services that support the acquisition and implementation of the AT. IDEIA 2004 defines AT services as “any service

that directly assists an individual with a disability in the selection, acquisition, or use of an assistive technology device” (20 U.S.C. 1401 § 602(2)). Such services would include (1) evaluation of the child’s needs, including a functional evaluation of the child in his or her customary environment; (2) purchasing, leasing, or otherwise providing for the acquisition of AT devices by the child; (3) selecting, designing, fitting, customizing, adapting, applying, maintaining, repairing, or replacing of AT devices; (4) coordinating and using other therapies, interventions, or services with AT devices, such as those associated with existing education and rehabilitation plans and programs; (5) training or technical assistance for the child, or, where appropriate, the child’s family; and (6) training or technical assistance (20 U.S.C. 1401§602(2)(A-F)).

The IDEIA 2004 definition places emphasis on the compensatory nature of AT, that is, it compensates for something a child cannot functionally do or perform. A more easily understood, or working definition of AT for practitioners is that it is a tool (or strategy) that allows a person to do a task he or she could not do without the tool (or strategy) at the expected performance level (Parette *et al.*, 2007). While IT and ICT may be helpful to teach new skills, remediate problems demonstrated by students with disabilities, supplement the curricular experiences of students, and facilitate access to and the manipulation of digital information, it is not individualized. AT is individually matched to and uniquely required for a student to participate in the curriculum or classroom and make educational progress (Lewis, 1993; Rose *et al.*, 2005).

Considering AT

Since AT is compensatory and its use in educational settings should culminate in positive outcomes (i.e., progress toward educational goals; Parette *et al.*, 2007), how it is considered by education professionals when developing and implementing program plans is of paramount importance. Numerous planning frameworks and models for making decisions about AT have been proposed over the past two decades (cf., Blackhurst, 2005; Bowser and Reed, 1995; Center for Technology in Education, Johns Hopkins University; and Technology and Media Division (TAM) of the Council for Exceptional Children, 2005; Chambers, 1997; Parette and VanBiervliet, 1990, 1991a, 1991b, 1991c, 1991d, 1991e; Parette *et al.*, 1991; Scherer, 2002; Zabala, 1993; Zabala and Carl, 2005). While providing some degree of guidance for AT decision making, educators and service providers may sometimes be uncertain about the practicality of these differing approaches and their relevance to the day-to-day activities in public school milieus. Of particular importance is that some educators currently do not have a broad understanding

of how seemingly different elements of AT consideration (e.g., student and school environment characteristics, educational demands and tasks within specific environmental settings, and AT tools and their features) interface and lead to effective decisions about what AT tools children with disabilities receive.

To facilitate a better understanding about what must be considered to effectively choose and implement AT with students having disabilities, **Figure 3** presents a framework that incorporates (1) what the field knows about human factors (Cook and Hussey, 2002; King, 1999); (2) the relationship between human factors and educational activities in which children participate (Cook and Hussey; King); and (3) understanding of the nature of tools that can be used by students with disabilities to make progress in the curriculum (Cook and Hussey). The following section describes three specific dimensions of AT decision making that are integrally related: (1) activities and their embedded tasks, (2) the demands created by activities and tasks, and (3) the manner in which individuals can appropriately respond to demands.

Activities and Embedded Tasks

Education professionals working with students with disabilities are keenly aware of typical daily activities in which students participate and in which they must be successful to make progress in the curriculum. For example, preschool children may participate in activities such as opening activity, choice time, reading time, free play, and lunch. Elementary-age children would participate in daily activities such as language arts (reading and writing), mathematics, social studies, science, or art. High school students’ activities would include specific content courses such as American history, geometry, English literature, biology, and physical education.

Regardless of the age level, each activity in which a student participates has embedded tasks. For example, to participate in free play, the preschool child may have to complete tasks such as (1) scanning the available activities and choosing an activity in which to engage, (2) engaging in the activity in a meaningful way, and (3) terminating the activity, often by putting materials away. To participate in language arts at the elementary level, a student might (1) read a text passage and then write a story about his/her own similar experience, (2) engage in writing to include completing tasks of planning the topic and making a content outline, (3) transcribe an initial draft, (4) edit and revise the composition, and (5) finally submit it to the teacher. At the high school level, to participate in history class, a student might (1) participate in class discussions, (2) listen to a presentation or view a video, (3) take notes, (4) read a text assignment, (5) write assignments in a planner, (6) complete and/or submit homework, and (7) take exams. Thus, participation may be viewed as

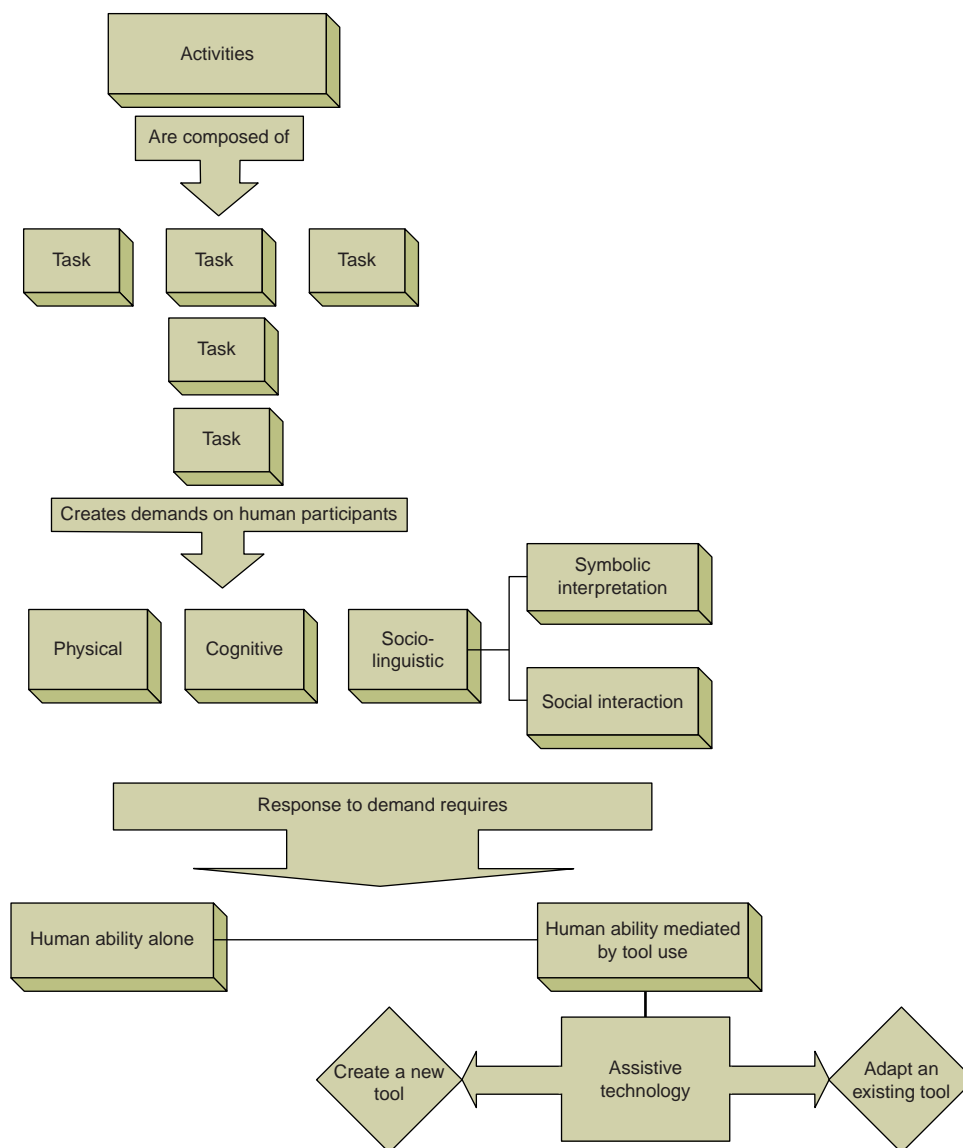


Figure 3 Relationship of tasks and their demands within activities to responses by humans to these demands. ©2009, G. R. Peterson-Karlan, Permission is granted to use my figure for this publication. GRP-K, 7/13/09.

a series of related tasks that culminate in successful completion of a specific activity by the student with a disability.

Demands on Human Users

Whatever the human activity in which students of any age and ability level participate, the embedded tasks present demands on students. These demands include (1) physical, (2) cognitive, and (3) sociolinguistic (King, 1999). Each is described in the following section.

Physical demands

Many tasks associated with school activities place physical demands on any human participant. King (1999)

described physical demands as the amount of muscle strength and movement “required to initiate, pursue, and complete a task” (p. 60). These demands would include such specific things as range of motion (i.e., ability to move joints, arms, and legs in various directions), resolution and repetition of movement, and sustained movement or position. For example, a preschool child participating in a lunch activity would be confronted with the task of sitting at a table. This demands that the student sit upright in chair at a table (which requires balance, hip stability, and flexion of the hip joints). To make a choice at the table, the child must coordinate muscles of the arms and hands to reach and grasp a preferred food item or manipulate a cup, plate, or utensil. These examples illustrate that many tasks within an activity require a student to do

something physically, and such demands may or may not present challenges to a particular student having a disability.

Cognitive demands

Generally speaking, most tasks that students are presented with in their school activities have varying degrees of cognitive demands (i.e., the amount of thinking required to complete a task). King (1999) noted that these demands may include (1) sensing (i.e., visual, auditory, and tactile-kinesthetic); (2) remembering (i.e., factual memory); (3) discriminating (i.e., differentiating); (4) analyzing (i.e., problem-solving); and (5) sequencing actions (i.e., sequential memory). For example, when a high school student participating in history takes notes (the task), he or she must be able to hear the teacher who is lecturing or see notes displayed with a liquid-crystal display (LCD) projector (sensing); must remember information from the reading homework; and must sort out unimportant information presented during the discussion from important content over which he or she might later be tested (discriminating). Finally, the student must write from left to right while note-taking using sequences of letters to spell words (sequencing).

Sociolinguistic demands

Tasks within school activities frequently present sociolinguistic demands on users. King (1999) stated that linguistic demands include “the amount of symbolic interpretation and processing that the user must invest” (p. 61). Linguistic demands posed by a particular task can also include interpretation of symbols and text. For example, when a young child is presented with a list of possible activities in which he or she may participate during choice time, both pictures and/or text representing the activities must be processed and interpreted. For some children, the text alone might not present an undue cognitive burden for interpretation, while for other children, both text and a picture (or picture alone) might be required for interpretation to occur. For a high school student presented with the task of note-taking, he or she would typically be presented with the linguistic demand of recognizing and understanding the meaning of text written by the teacher on a whiteboard.

Symbolic interpretation, however, often does not occur outside the context of a social interaction. For example, if a teacher uses a whiteboard to write a list of steps needed to complete a project, and then asks the class, “What do we do to complete this project?” There are social demands placed on all students to be able to participate in the task of responding to the teacher. First, the student must attend to the teacher who is asking the question. Second, the student must signal a desire to respond (e.g., raising one’s hand vs. blurting out an answer). Third, once the student’s signal is recognized by the teacher (e.g., “Johnny, can you tell us what we do next?”), the student must

appropriately communicate an answer to the question and terminate the communicative response appropriately. In other types of tasks, social skills such as turn-taking, civility, and use of manners are embedded in the activity and are important considerations to fully understand what a student must do to be successful in a targeted activity that has many tasks.

Response to Demands

When students with disabilities participate in school activities having a cadre of embedded tasks (with associated physical, cognitive, and/or sociolinguistic demands), they either do the task alone or use tools as a way to accomplish the task. In the former instance, many students with disabilities typically will exhibit some abilities to perform certain tasks within educational activities. For example, during math class in the elementary classroom, a student may be able to understand text in a story problem if it is read aloud and provide an answer, but has difficulty if asked to read the text in which the problem is detailed. In this instance, the student uses human ability alone to understand, process, and express an answer. However, because he has difficulty decoding (reading) the text – another task embedded in the activity – compensation is required to enable him to complete this more demanding task that requires a tool, such as digital text with speech output, which may not be typically used by others.

Human ability mediated by tool use

In many academic tasks, demands (whether physical, cognitive, and/or sociolinguistic) are typically met using a tool. Ubiquitous tools in school settings include things such as paper and pencils to record information for later use, the school planner, and a pencil or pen to record assignments and dates that must be remembered later, three-ring notebooks to organize materials and backpacks to help carry books and notebooks. Tools are used in school settings to (1) enhance effectiveness, (2) reduce effort, (3) reduce errors, (4) increase speed, (5) improve quality, (6) reduce physical effort, and (7) reduce cognitive effort. It is important to note that each tool created to help accomplish a task also imposes its own physical, cognitive, and sociolinguistic demands on the tool user.

When a tool must be used to compensate for functional inabilities to complete one or more specific tasks, it becomes an AT tool. Some AT tools are created specifically for persons with disabilities. For example, wheelchairs, hearing aids, and Braille are created specifically to compensate for the physical and sensory needs of persons with disabilities. Text-to-speech screen readers and web browsers are created to compensate for inabilities to differentiate between screen text by persons with visual impairments or to decode print by persons with learning disabilities. However, many existing tools can be adapted to provide

compensatory benefits for a student with a disability. In the case of a student with physical disabilities who has difficulty grasping standard eating utensils, oversized, padded handles or a change in the handle angle can provide the needed compensatory effect. For the student with a learning disability who has difficulty with spelling and word use, a spell checker and/or thesaurus with text-to-speech capability can provide needed compensatory effect.

AT Outcomes and Today's Classrooms

In today's school settings, ensuring the student's access to educational opportunities is not sufficient; successful participation of students with disabilities in the general education curriculum and monitoring that progress is of utmost importance (Thurlow *et al.*, 2007). More specifically, using research-evidenced and data-based practices in classrooms has become a singular focus in the field of special education (Gersten *et al.*, 2005; Horner *et al.*, 2005; Odom *et al.*, 2005; Parette and Peterson-Karlan, 2008). This is now true of AT consideration as well (Peterson-Karlan and Parette, 2007, 2008). There are three important phases in which data are crucial: (1) when identifying the problem for which AT is needed, (2) before making final a decision to acquire or purchase a particular AT solution, and (c) during long-term implementation of the AT solution (i.e., collecting data while the student uses the AT solution for the purpose intended; Parette *et al.*, 2007). Implicit in these phases are outcome questions for which data must be generated: Is AT needed? Is the AT effective? Over time, is the AT effective in supporting educational progress? Specific data-based strategies for assessing outcomes of AT consideration at each of the decision points have been reported elsewhere (see Parette *et al.*), but those strategies involve documenting the performance gap introduced in Figure 2, measuring the compensatory effect of the introduction of AT and monitoring the effect of continued AT use on educational progress. Concurrent time-series designs in which performance with and without AT tool use is measured at specific points (e.g., during trial implementation with the AT) and is repeatedly sampled over time (e.g., quarterly in the school year) are used to determine whether AT is effective, needed, and contributing to students' educational progress.

Conclusion

This article has emphasized the increasing role of technology in education settings, with particular emphasis on consideration of AT for students with disabilities. The legal mandate of IDEIA 2004 to consider AT when developing individual education programs, requires education professionals to better understand the compensatory nature

of AT within the context of human factors, educational activities, and the demands of tools on users (both physical, cognitive, and sociolinguistic). Of particular importance was the presentation of the concept of tasks being embedded within primary educational activities in which children participate. These various tasks present demands which require a student to use human ability alone or a tool.

The use of data in AT decision making has also been emphasized, typically occurring at three distinct phases. Each phase, in turn, has specific questions that education professionals must address to ensure that effective decision making for the student has occurred.

See also: Curriculum-Based Assessment and Students with Special Needs; Functional Behavioral Assessment; Instructional Accommodations for Children with Special Needs In Inclusive Settings; School-Based Services for Children with Special Needs.

Bibliography

- Blackhurst, A. E. (2005). Historical perspective about technology applications for people with disabilities. In Edyburn, D., Higgins, K., and Boone, R. (eds.) *Handbook of Special Education Technology Research and Practice*, pp 3–29. Whitefish Bay, WI: Knowledge by Design.
- Bowser, G. and Reed, P. (1995). Education TECH points for assistive technology planning. *Journal of Special Education Technology* **12**, 325–338.
- Center for Technology in Education, Johns Hopkins University, and Technology and Media Division (TAM) of the Council for Exceptional Children (2005). *Considering the Need for Assistive Technology within the Individualized Education Program*. Columbia, MD: Author.
- Chambers, A. C. (1997). *Has Technology Been Considered: A Guide for IEP Teams*. Reston, VA: Council for Exceptional Children.
- Cook, A. M. and Hussey, S. M. (2002). *Assistive Technology: Principles and Practices*, 2nd edn. St. Louis, MO: Mosby.
- Edyburn, D. L. (2000). Assistive technology and students with mild disabilities. *Focus on Exceptional Children* **32**(9), 1–24.
- Gersten, R., Fuchs, L. S., Compton, D., *et al.* (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children* **71**(2), 149–164.
- Grabe, M. and Grabe, C. (2007). *Integrating Technology for Meaningful Learning*, 5th edn. New York: Houghton Mifflin.
- Horner, R. H., Carr, E. G., Halle, J., *et al.* (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children* **71**(2), 165–180.
- Individuals with Disabilities Education Improvement Act (IDEIA) (2004). 20 U.S.C. § 1400 *et seq.*
- International Society for Technology in Education (2005a). *Connecting Curriculum and Technology*. http://www.cnets.iste.org/students/s_book.html (accessed Jun. 2009).
- International Society for Technology in Education (2005b). *Preparing Teachers to Use Technology*. http://www.cnets.iste.org/teachers/t_book.html (accessed Jun. 2009).
- King, T. W. (1999). *Assistive Technology: Essential Human Factors*. Boston, MA: Allyn and Bacon.
- Lengel, J. G. and Lengel, K. M. (2006). *Integrating Technology. A Practical Guide*. Boston, MA: Pearson.
- Lewis, R. B. (1993). *Special Education Technology: Classroom Applications*. Pacific Grove, CA: Brooks/Cole.
- Merriam-Webster, Inc. (2007). *Technology*. <http://www.merriam-webster.com/dictionary/technology> (accessed Jun. 2009).
- No Child Left Behind Act (2001). 20 U.S.C. 6301 *et seq.*

- Odom, S. L., Brantlinger, E., Gersten, R., et al. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children* **71**, 137–148.
- Parette, H. P. and Peterson-Karlan, G. R. (eds.) (2008). *Research-Based Practices in Developmental Disabilities*, 2nd edn. Austin, TX: Pro-Ed.
- Parette, H. P. and VanBiervliet, A. (1990). *Assistive Technology and Disabilities. A Guide for Parents and Students*. Little Rock, AR: University of Arkansas at Little Rock. (ERIC Document Reproduction Service No. ED364 026.)
- Parette, H. P. and VanBiervliet, A. (1991a). *Assistive Technology Curriculum: A Module of Instruction for Professionals (and) Instructor's Supplement*. Little Rock, AR: University of Arkansas at Little Rock. (ERIC Document Reproduction Service No. ED324 887.)
- Parette, H. P. and VanBiervliet, A. (1991b). *Assistive Technology Curriculum: A Module of Instruction for Students in Arkansas Colleges and Universities (and) Instructor's Supplement*. Little Rock, AR: University of Arkansas at Little Rock. (ERIC Document Reproduction Service No. ED324 884.)
- Parette, H. P. and VanBiervliet, A. (1991c). *Assistive Technology Curriculum for Arkansans with Disabilities*. Little Rock, AR: University of Arkansas at Little Rock. (ERIC Document Reproduction Service No. ED324 886.)
- Parette, H. P. and VanBiervliet, A. (1991d). *Assistive Technology Curriculum for Parents of Arkansans with Disabilities*. Little Rock, AR: University of Arkansas at Little Rock. (ERIC Document Reproduction Service No. ED324 885.)
- Parette, H. P. and VanBiervliet, A. (1991e). *Assistive Technology Guide for Young Children with Disabilities*. Little Rock, AR: University of Arkansas at Little Rock. (ERIC Document Reproduction Service No. ED324 888.)
- Parette, H. P., VanBiervliet, A., and Hughes, K. M. (1991). *Tools for Living. A Guide for Aging Arkansans*. Little Rock, AR: University of Arkansas at Little Rock. (ERIC Document Reproduction Service No. ED333 266.)
- Parette, H. P., Peterson-Karlan, G. R., Wojcik, B. W., and Bardi, N. (2007). Monitor that progress! Interpreting data trends for AT decision-making. *Teaching Exceptional Children* **39**(7), 22–29.
- Peterson-Karlan, G. R. and Parette, H. P. (2007). Evidence-based practice and the consideration of assistive technology effectiveness and outcomes. *Assistive Technology Outcomes and Benefits* **4**, 130–139.
- Peterson-Karlan, G. R. and Parette, H. P. (2008). Integrating assistive technology into the curriculum. In Parette, H. P. and Peterson-Karlan, G. R. (eds.) *Research-Based Practices in Developmental Disabilities*, 2nd edn., pp 183–214. Austin, TX: Pro-Ed.
- Roblyer, M. D. (2006). *Integrating Educational Technology into Teaching*, 4th edn. Upper Saddle River, NJ: Pearson Merrill Prentice Hall.
- Rose, D. H., Hasselbring, T. S., Stahl, S., and Zabala, J. (2005). Assistive technology and universal design for learning: Two sides of the same coin. In Edyburn, D., Higgins, K., and Boone, R. (eds.) *Handbook of Special Education Technology Research and Practice*, pp 507–518. Whitefish Bay, WI: Knowledge by Design.
- Scherer, M. J. (ed.) (2002). *Assistive Technology: Matching Device and Consumer for Successful Rehabilitation*. Washington, DC: American Psychological Association.
- SEAT Center, National Center for Technology Innovation, and University of Kansas (2006). *Assistive Technology Outcomes Summit. Assistive Technology and Educational Progress... Charting a New Direction. Executive Summary*. <http://www.nationaltechcenter.org/documents/ExecutiveSummaryFinal.pdf> (accessed Jun. 2009).
- Smaldino, S. E., Russell, J. D., Heinrich, R., and Molenda, M. (2005). *Instructional Technology and Media for Learning*, 8th edn. Upper Saddle River, NJ: Pearson Merrill Prentice Hall.
- Thurlow, M., Tindal, G., Powers, R., et al. (2007). Research on AT outcomes and large scale assessments. *Assistive Technology Outcomes and Benefits* **4**, 11–27.
- Zabala, J. (1993). *The SETT Framework: Critical Issues to Consider when Choosing and Using Assistive Technology*. Paper presented at the Closing the Gap Conference, Minneapolis, MN.
- Zabala, J. S. and Carl, D. F. (2005). Quality indicators for assistive technology services in schools. In Edyburn, D., Higgins, K., and Boone, R. (eds.) *Handbook of Special Education Technology Research and Practice*, pp 209–227. Whitefish Bay, WI: Knowledge by Design.

Further Reading

- Bryant, D. P. and Bryant, B. R. (2003). *Assistive Technology for People with Disabilities*. Boston, MA: Allyn and Bacon.
- Cook, A. M. and Hussey, S. M. (2002). *Assistive Technology: Principles and Practices*, 2nd edn. St. Louis, MO: Mosby.
- Dell, A. G., Newton, D. A., and Petroff, J. G. (2008). *Assistive Technology in the Classroom. Enhancing the Experiences of Students with Disabilities*. Upper Saddle River, NJ: Pearson Merrill Prentice Hall.
- Edyburn, D., Higgins, K., and Boone, R. (eds.) (2005). *Handbook of Assistive Technology Research and Practice*. Whitefish Bay, WI: Knowledge by Design.
- Johnston, L., Beard, L. A., and Carpenter, L. B. (2007). *Assistive Technology. Access for All Students*. Upper Saddle River, NJ: Pearson Merrill Prentice Hall.
- Peterson-Karlan, G. R. and Parette, H. P. (2005). Millennial students with mild disabilities and emerging assistive technology trends. *Journal of Special Education Technology* **20**(4), 27–38.

Augmentative–Alternative Communication

C Mason and K M Chinn, New Mexico State University, Las Cruces, NM, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Aided communication – Augmentative or alternative communication approaches that require a system or device to communicate.

Assistive technology – The use of any item, system, or equipment that is utilized to improve or maintain functional living skills and independence.

Digitized speech – Natural speech, which has been recorded and reproduced for use in voice output communication devices. However, these recordings are limited to recorded words or messages and limit the user's choice of messages.

Joint attention – When two or more individuals are focusing or attending to the same activity, object, topic, or thought process.

Switch access – When an individual is able to reach, activate, and utilize an external button or switch to create a cause-and-effect reaction.

Synthesized speech – Technologically generated speech which uses the linguistic patterns of digitized speech to create a speech system. This form of speech provides the user with much more choices and options in creating messages.

Unaided communication – Nonverbal communication approaches that utilize facial expressions and/or body movements, such as gestures, to communicate, for example, American Sign Language.

VOCA – The acronym for voice output communication aids, which is a group of devices that provides an alternative voice through recording or digitized speech.

Augmentative–alternative communication (AAC) provides individuals special needs with the means to express ideas, thoughts, concerns, and needs. Thus, the goals for AAC are twofold: (1) to make individuals more functionally independent by having the ability to express needs, wants, feelings, and preferences in a form that others can understand, and (2) to enhance communication skills and language skills, providing increased independence. Clearly, individuals with specific disabilities require AAC devices that provide accessibility specific to their needs. An individual who is deaf/hard of hearing not only may use a form of manual communication but may choose to use high-tech devices such as hearing aids and infrared listening devices available

in public settings such as theaters. In the educational setting, a child who is deaf may be provided with an FM unit to accompany the hearing aid. Cochlear implants are also a choice for individuals who are deaf. The individuals who are blind/visually impaired have Braille, print readers, and computers with large computer screens, and large font choices to aid in reading. Individuals who have cognitive deficits with communication effects may use more complicated to simplistic communication boards/switches. Individuals with cerebral palsy generally have communication boards/switches that are programmed to match the individual's physical abilities as well as cognitive level. Individuals with spinal injury at the level of the cervical vertebrae or advanced neuromuscular disorders such as amyotrophic lateral sclerosis (ALS) may have limited movement, and thus require AAC devices that use eye movement or head movement to signal letters with which words and sentences can be formed.

The American Speech-Language-Hearing Association defined AAC as a set of procedures and processes that maximize the individual's communication skills for functional communication. Separately defined, augmentative communication is any support system in a given situation that allows an individual to communicate independently, whereas alternative communication is specifically the use of nonvocal instruments and/or approaches.

The need for AAC spans socioeconomic, racial, ethnic, and age groups. According to the American Speech Hearing and Language Association, the prevalence of people with communication impairments severe enough to require AAC ranges from 8 to 12 in every 1000. Individuals that most frequently benefit from AAC include those who:

- are deaf or hard of hearing;
- have acquired neurological impairments such as TBI, cerebral palsy, or aphasia;
- have progressive neurological disorders such as ALS;
- have been diagnosed with an autism spectrum disorder (ASD);
- have motor speech disorders; or
- have been diagnosed with neurogenetic disorders.

Two distinct types of users include those with long-term need and those with temporary need for AAC. Individuals that have suffered a stroke and have developed aphasia may need temporary assistance through AAC until spontaneous recovery allows them to regain their speech, whereas individuals with ALS, otherwise

known as Lou Gehrig's disease, will need continued, long-term AAC support. These individuals may start with manual sign language or topic reference pictures to assist the listener understand the speaker's slurred speech more clearly. As the disease progresses and independence decreases, the individual with ALS will need a more sophisticated system.

Forms of AAC

AAC is found in many forms, from manual communication to low and high technology systems, and are further categorized as aided or unaided. Unaided systems, which require no external devices, might include, but are not limited to, gestures and sign language. Aided AAC devices might include picture systems and/or print (which are considered low tech) and computerized communication devices (which are considered high tech).

Manual Communication

Manual communication systems used for unaided communication are the most frequently used form of AAC. Manual communication consists of cued speech, American Indian Sign Language, American Sign Language, and various English signed systems such as Signing Essential English (SEE I), Signing Exact English (SEE II), Language of Visual English (LOVE), and pidgin forms of signed language. According to Mitchell (2006), there are approximately 11 000 000 individuals who are deaf or hard of hearing in the United States with 1 000 000 Deaf. ASL is the most frequently used version of manual communication as ASL is the language of the Deaf culture. Most other forms of manual communication, excluding American Indian Sign Language, are used in educational settings with children who are deaf or severely hard of hearing such that it precludes the use of speech and speech reading for communication purposes. SEE I, SEE II, and LOVE were specifically developed to assist deaf and severely hard-of-hearing children in learning English for the purposes of reading and writing.

Describing Low- and High-Level Technology

Lower technological devices, often referred to as low tech, are used with individuals that are unable to produce gestures and manual sign, due to motoric limitations. Forms of aided communication such as a communication board, picture schedule, eye gaze board, or other simple aids are not electronic, but offer a form of assistive technology. Assistive technology is a broad term which may include aids for daily living activities, vocational activities, or environmental controls. It may be used to refer to some forms of augmentative communication. A more detailed

list of low-tech, aided communication systems might include, but might not be limited to,

- Eye gaze board – where the individual moves his/her eyes in the direction of a specific picture or symbol to indicate a need.
- Alphabet boards – which provide the individual with the alphabet to spell out simple messages (example given the end of this article).
- Writing materials – to communicate wants, needs, thoughts, and desires.
- Communication book/dictionary – development of a gesture or communication dictionary provides the individual information on what the behaviors look like, what each gesture means, and how to respond to the message. The creation of the dictionary should include the primary caregivers input, and thus provide consistency in communicative attempts. These can then be paired with more conventional photos, objects, or symbols to help the individual expand his/her communication base.
- Picture exchange communication system (PECS) – a system that was designed to provide opportunities for individuals, primarily with autism, to begin initiating communication.
- Picture communication symbol systems – other systems, similar to PECS, which provide opportunities for communication and scheduling routines.
- Blissymbolics – another graphic meaning communication system originally designed for children who were deaf/hard of hearing that was applied to the general population.

Higher technological devices, referred to as high tech, are also forms of aided communication and assistive technology. These devices include different types of hand-held printers, word processors, and computers. Devices with voice output are referred to as VOCAs – voice output communication aids. These devices are not limited to voice output and may instead activate some other relevant objects in the environment. Examples of VOCAs used as high-tech-assistive devices may include some, but not limited to, of the following:

- Switches are various electronically controlled buttons of various sizes. These buttons are used to activate a device for speech or environmental controls. For example, the push of a button may turn on a television or personal fan, or it may ring a bell calling for someone's attention. These switches may be activated with various parts of the body from hand pressure to eye blink, depending on the abilities for movement of the individual.
- There are simple VOCA devices, in which, one single button activates a single message recorded on the system. This message is usually recorded by a peer or caregiver.

These systems are generally limited in the communication repertoire. This presents difficulties in the use of the system in all other communicative settings or situations.

- More sophisticated VOCA devices offer a variety of opportunities for communication, including keyboarding, word prediction, and programmed personalized information for ready-access. The devices may include direct selection opportunities, where an individual will push a single button for a message. This requires strength and the ability to access the buttons. The equipment will usually have the ability to scan different options with a single switch to hit as the item is highlighted on the system. These devices may also use recorded speech or synthesized speech, which is more robotic in nature. Although great advances have been made to the quality of the speech, we now have a variety of voices to provide a more appropriate vocal quality dependent upon age and gender.

Assessing the Need for AAC

Careful and thorough assessment of individuals, their needs, significant others, and lifestyle is necessary to determine the best mode of augmentative communication. Areas of evaluation would include all facets of the individual's medical history, premorbid state, current physical abilities, cognitive abilities, typical communication partners, lifestyles, and living environments. In assessing the individual, careful consideration should include the following: past and current medical history; cognitive abilities, including the ability to initiate conversation; as well as receptive and expressive language skills. The individual evaluating a client in need of an AAC should know the physical limitations of the individual and how these limitations would affect accessing a communication system. Consideration of the medical prognosis, including a progressive condition, may also impact the current and future levels of the individual's communicative needs. Further, what would be the medical prognosis of the condition? That is, is the condition a progressive disease, which may impact the current and future levels of communicative needs?

Determination regarding the permanence of the communicative disorder is important for choosing a device. If the disorder is progressive, how soon is the condition expected to worsen? As indicated above, if the individual has a progressive disease, this assessment process must take future needs into consideration. That is, what will the communication needs be at each level of the progression? In order to determine which AAC device is most appropriate, cognitive abilities and the ability to initiate communication must be evaluated. The therapist choosing and fitting the device must have information regarding the level of awareness of the individual who needs AAC. Further, the

ability to utilize joint attention with another individual must be assessed. Is this individual able to take turns in communication and gain the attention of others? Prior to choosing an AAC system, the individual's ability to recognize and understand at an object and photographic level, line drawing, or written word must be determined, as AAC devices vary in complexity based on individual abilities. For example, can the individual perform actions or use objects? Along with choice of complexity level, the number of items presented is also a consideration. Is the individual able to select one item out of how many choices?

Receptive and expressive language skills are part of the determination for an AAC device. The level of communication is an important factor for choice. What does the client understand? Is the client only able to use limited or telegraphic speech used by his/her communication partner? The ability to understand extended conversations and lengthy messages must be determined. Provided with opportunities through augmentative means, is the individual able to label, identify, or describe? Further, what types of messages is he/she able to relate? For example, can the individual make requests, negate information, refuse treatment, gain attention, report information, clarify messages, seek information, and/or express feelings and concerns? The range of physical motion and level of strength/endurance of the individual must be assessed to determine the appropriate type of AAC. The ability to move limbs, as well as the range of motion of limbs, fingers, head, feet, and eyes, are part of this assessment. Further, information regarding the ability to perform multiple tasks such as the ability to move while making decisions is important in the choice of AAC devices. Strength such as the ability to apply pressure on a switch or button must be determined. Does the individual being assessed for the device fatigue easily? The best time of day to utilize maximum strength to communicate must also be assessed. Some devices may require more dexterity along with the ability to make choices in a timely manner, as these scanning devices provide a limited time frame for item selection. Therefore, speed and accuracy may be factors in the selection of a device.

Positioning is important to ensure that individuals have sufficient support to allow them the maximum range of motion, provide clear contact with their device, and reduce fatigue during communication opportunities. An interdisciplinary team effort, including assistance and consultation with OT and/or PT, will provide input concerning the client's physical abilities and equipment needs, such as chair mounts, device portability and location on the walker or wheelchair, and optimal positioning for the client. Visual and auditory limitations should also be considered in selecting symbols for an AAC device. Regarding visual impairments, many devices are available

and assessment is needed to determine which system(s) best meets the individual's needs. In order to determine this, the diagnostician must assess how well he/she sees. If the individual can see print and line drawing, would a colored background assist in symbol recognition? How well can he/she scan multiple pictures for selection? How many choices can the individual maintain in his/her line of vision? The size of items on the communicative device also must be determined to meet his/her field of vision.

Since emotional status and level of motivation for communication affects the optimal use of the AAC device, the following questions must be addressed. Often individuals with particular disorders become depressed. Is this individual on depression medications, which may impact alertness? Does he/she have the desire and motivation to learn an augmentative communication system? What level of communication is important to him/her? Does he/she desire to carry on conversations, express only basic wants and needs, or be as independent as possible?

Assessing Communication Partners

Communication partners take on important roles in the life of the individual who requires an AAC device. The following questions must be answered regarding the communication partner. Who will be the primary conversation partners? Will the communication partner be willing to be involved in utilizing a new system? Even with the advancement in technology, conversational exchange using an AAC device remains much slower than natural speech, and will detract from the natural social exchange of information. Will the partner be able to display the patience in waiting for responses? Will the partner be committed to providing multiple communication opportunities for the individual? Will the communicative partner be able to understand and help the individual learn and program or change the device, dependent upon the situation? Lastly, during the evaluation, typical communication settings must be addressed. Where will the individual be communicating most frequently: in the home, school, or community? Will there be extraneous noise or will the setting be quiet? If there is extraneous noise, how loud will it be? To gather the most thorough and in-depth information in answering the aforementioned questions and in providing the optimum communication system for an individual, an assessment team should be organized. This multi-disciplinary assessment team may include: family and/or caregivers who are an integral part of the assessment, an audiologist, an educational specialist(s), an educational diagnostician, a physician, a psychologist, a speech-language pathologist, an occupational therapist, a physical therapist, a social worker, a rehabilitation engineer, a vision specialist, and/or a vocational counselor. Clearly,

the selection of the system will be determined by the observations and input from each of the evaluation team members, as well as the evaluation of the individual and interviews with the significant others. Information must be gathered by the diagnostic team; the potential user's feelings regarding the use of the equipment, such as possible embarrassment associated with the device, must be taken into account. Many of the high-tech devices are extremely expensive and require funding for affordability. If the individual does not have a payment source or personal funds, selection of a low-tech method may be more acceptable. Often, AAC vendors will assist with funding questions and paperwork. Along with this service, these individuals may also supply a variety of devices for trial and assessment. The user would have the opportunity to utilize the equipment in conversational trials to determine the best device, the number of selections available, and the best mode of access, such as scanning or direct selection as discussed above.

Initiating Communication

Many times as the individual is trained to utilize the alternative communication system, they become prompt dependent for responses. It is important that the educator or speech-language pathologist avoid prompt-dependent or cued communication interactions. The system must be as simple and unencumbered as possible to provide the easiest access for communication opportunities. The home, school, and other environments should be engineered for optimum communicative opportunities for the nonverbal individual and his/her communication partners. Vocabulary and access must be functional for the user and significant others. Lack of interest in communication may indicate that the individual is not being exposed to preferred activities or that the needs of the communicator are not being addressed with access of the system. The use of some AAC systems requires a myriad of skills and multi-tasking abilities. Skills necessary for responding to a simple question posed by the communication partner may require the user to (1) understand the question, (2) turn on and focus on the AAC system, (3) attend to symbols and perceive them accurately, (4) determine the locations and sequence to respond, (5) access the system, (6) determine if the communication partner has understood the message, and (7) repeat the communicative sequence in order to continue the conversation.

Early intervention makes a significant impact on the success of an alternative communication system. The earlier the child begins to use an AAC system, the greater the chance of acceptance and regular use. Current research documents the efficacy of communication services, reporting that services should begin before age 2 in children with severe disabilities. The use of AAC systems should be

considered as soon as the child demonstrates communicative needs. Although parents are sometimes concerned that introduction of AAC may preclude development of speech, numerous studies demonstrate that the opposite affect is found; this approach supports such development.

Communicating with Partners

Two different types of communication partners have been described. A partner-dependent user is one who must rely on another individual to manage the informational exchange. This situation requires highly familiar contexts. The user's partner may have to scaffold interactions by offering choices, provide visual input, or restructure the exchange for easier understanding. The second type of communication partner is an independent partner, who has relatively preserved or intact language skills and executive functions, and may be a candidate for more high-tech devices that will allow communication without the assistance of another individual. Teaching advocacy for the individual who uses the AAC device should be the responsibility of the entire evaluation/intervention team. Training should be shared with family members or caregivers.

Selecting Vocabulary for the Individual Using AAC

Vocabulary selection is imperative for the individual using the device. Without appropriate selection of terms, words, and labels, the individual will not utilize the system effectively. Vocabulary selection should focus on: frequently used words, culturally sensitive vocabulary that is used with significant others, commonly used names, phrases and needs, word prediction to make access faster; and personal data information, along with interests and experiences so that the individual may share personal information with others. Frequently used words would include those that are frequently found in the individual's environment and in the specific languages. These words are often pre-programmed in the device and are easily accessible through different menus, as well as in word-prediction keyboard access.

The vocabulary selection must be sensitive to cultural influences. A gesture may produce positive affect in one culture and an entirely different, perhaps negative message in another. There must also be sensitivity to language differences and environments. For example, a child may be Spanish dominant in the vocabulary utilized at home where no English is spoken, and may be English dominant regarding academic vocabulary spoken in the classroom setting. In order for a system to be utilized and effective, it must address concerns across cultures and environments. Personal information would include name, address, date of birth, physician's names, current medications, names of family members and friends, and other pertinent

information as desired by the individual. Most high-tech devices may also be programmed with jokes, personal stories, and even with speeches an individual may give during presentations. Along with their cultural style, personal information and language preferences, the individual's premorbid state must be taken into consideration when selecting vocabulary. The factors of educational status and experiences prior to the onset of the disease or accident are necessary in selecting items for the word bank. Based on the individual's language abilities, what parts of speech such as articles, conjunctions, and auxiliary verbs should be incorporated into the vocabulary, or would the individual benefit more from telegraphic speech such as isolated, simple nouns and verbs?

Using the AAC Device in the Classroom or Home and Community Settings

In order for a system to be effective in an academic setting, there must be consultation, education, and support for the classroom teacher, students, and the AAC user. Each setting must have the support and consultative opportunities for the system to work. In the home setting, the parents, siblings, and/or caretakers must be provided with training in how to utilize the system and must also be instructed in how to access additional information and resources. Training should be provided in functional settings. Acceptance of and use of an alternative method of communication is dependent upon all communication partners accepting and understanding the system. As stated previously, concerns by caregivers and parents that an augmentative communication system will deter the individual from using or developing speech are unfounded. The contrary is true. In fact, intervention approaches using AAC are found to facilitate natural speech. This information must be clearly stated and shared with all parties. It is important that the members of the AAC team who put a system in place continue to constantly monitor the user and communication partners for changes in the environment and the user's abilities and growth. Frequent modification and upkeep of equipment are also warranted.

Conclusion

Augmentative and alternative communication systems require in-depth assessment, taking many factors into consideration. The use of a selected system requires acceptance by all individuals working on the diagnostic and intervention teams, along with users, caregivers, and significant others. Careful support and maintenance of the AAC system should be ongoing to meet the changing needs of the individual and his/her environment.

A	B	C	D		
E	F	G	H		
I	J	K	L	M	N
O	P	Q	R	S	T
U	V	W	X	Y	Z
Y					

A vowel-based communication chart is a low tech AAC system utilized with individuals who can spell and are able to communicate answers to yes/no questions with head nods or eye blinks. The system is used to spell out words by having the communication partner guess the letters. The communication partner will start determining each letter by asking for the appropriate vowel in a vertical line and then moving across, horizontally to select the consonants. Although this process is slow and methodical, it does provide an avenue when other methods of AAC are not available.

Figure 1

Bibliography

Mitchell, R. (2006). How many deaf people are there in the United States? Estimates from the survey of income and program participation. *Journal of Deaf Studies and Deaf Education* **11**(1), 112–119.

Further Reading

American Speech-Language-Hearing Association (2002). *Augmentative and Alternative Communication: Knowledge and Skills for Service Delivery* (Knowledge and Skills). <http://www.asha.org/policy> (accessed August 2009).

American Speech-Language-Hearing Association (2005). *Roles and Responsibilities of Speech-Language Pathologists with Respect to Augmentative and Alternative Communication: Position Statement*. <http://www.asha.org/policy> (accessed August 2009).

Beukelman, D. and Mirenda, P. (2005). *Augmentative and Alternative Communication: Management of Severe Communication Disorders in Children and Adults*, 3rd edn. Baltimore, MD: Paul H. Brookes.

Castrogiovanni, A. (2008). Communication facts: Special populations: Augmentative and alternative communication. *American Speech-Language-Hearing Journal*. <http://www.asha.org/research/reports/aac.htm> (accessed August 2009).

Fallon, K. A. (2008). AAC in the schools: Current issues and future directions. *Augmentative and Alternative Communication* **17**(1), 48–59.

Fallon, K. A. and Light, J. C. (2001). Enhancing vocabulary selection for preschoolers who require augmentative and alternative communication (AAC). *American Journal of Speech-Language Pathology* **10**, 81–94.

Franklin Silverman (1999). *The Telecommunication Relay Service Handbook: Empowering the Hearing and Speech Impaired*. Newport, RI: Aegis Publishing Group.

Kemp, C. E. and Parette, H. P. (2000). Barriers to minority family involvement in assistive technology decision-making processes. *Education and Training in Mental Retardation and Developmental Disabilities* **35**(4), 384–392.

Lane, H. (1999). *The Mask of Benevolence: Disabling the Deaf Community*. New York: Randomhouse.

Lasker, J. P. (2005). AAC language assessment: Considerations for adults with aphasia. *Perspectives on Augmentative and Alternative Communication* **17**, 105–112.

Light, J. C., Beukelman, D. R., and Reichle, J. (2003). *Communicative Competence for Individuals Who Use AAC: From Research to Effective Practice (AAC Series)*. Baltimore, MD: Paul H. Brookes.

Lytoon, R., Hershberger, D., and Duggan, M. (2007). AAC users, families, clinicians, and vendors: Where are our boundaries? *Poster Session Presented at the American Speech Language and Hearing Association Convention*, Boston, MA.

McEwen, I. R. and Lloyd, L. L. (1990). Positioning students with cerebral palsy to use augmentative and alternative communication. *Language, Speech, and Hearing Services in Schools* **21**, 15–21.

Moore, D. (2002). *Education the Deaf: Psychology, Principles and Practices*. Boston, MA: Houghton Mifflin.

National Joint Committee for the Communication Needs of Persons with Severe Disabilities (n.d.). *FAQs: Communication Development and Concerns*. <http://www.asha.org/NJC/faqs-faqs-comm-dev.htm> (accessed August 2009).

Nicolosi, L., Harryman, E., and Kresheck, J. (2004). *Terminology of Communication Disorders*, 5th edn. Baltimore, MD: Lippincott, Williams and Wilkins.

Owens, R. E., Metz, D. E., and Haas, A. (1993). *Introduction to Communication Disorders – A Life Span Perspective*, 2nd edn. Boston, CA: Allyn and Bacon.

Proctor, L. A. and Oswald, J. (2008). Augmentative and alternative communication: Assessment in the schools. *Perspectives on Augmentative and Alternative Communication* **17**, 13–19.

Reichle, J., Light, J., and Beukelman, D. R. (2001). *Implementing an Augmentative Communication System: Exemplary Strategies for Beginning Communicators*. Baltimore, MD: Paul H. Brookes.

Seligman-Wine, J. (2007). Supporting families of children who use AAC. *The ASHA Leader* **12**(10), 17–19.

Siegel-Causey, E. and Bashinski, S. M. (1997). Enhancing initial communication and responsiveness of learners with multiple disabilities: A tri-focus framework for partners. *Focus on Autism and Other Developmental Disabilities* **12**(2), 105–120.

Tyler, R. S. and Schum, D. J. (1995). *Assistive Devices for Persons with Hearing Impairment*. Needham, MA: Allyn and Bacon.

Community Integration and Employment of Youth with Special Needs

B A Ford and A W Daviso III, The University of Akron, Akron, OH, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Career/job exploration – The process of finding a rewarding career path, as well as specific jobs within a particular career path.

Community experiences – Educational activities provided outside of the school building or in a community setting.

Culturally and/or linguistically diverse (CLD) – For this article, this includes populations from African-American, Hispanic American, Asian American, and American Indian backgrounds.

Self-determination – The act or power of making up one's own mind about what to think or do, without outside influence or compulsion.

Significant Multicultural Community Resources (SMCR) – Not-for-profit systems of informal support such as service or social organizations, sororities, fraternities, clubs or agencies, religious groups or churches, and individuals who are local community residents perceive as providing valuable services.

Summary of performance (SOP) – An overview of academic achievement and functional performance, which includes recommendations on how to assist the child in meeting postsecondary goals.

Community Integration and Employment of Youth with Special Needs

Given the recent emphasis on accountability for quality educational services, the issue of community integration of students with special needs has become even more critical. Despite legal mandates, significant numbers of young adults with disabilities continue to exit high school ill-prepared to effectively integrate into their respective communities or engage in postsecondary activities that promote productive lives. The provision of community-based experience during the transition-planning process is deemed an important strategy to address this problem (Geenen *et al.*, 2001; Hughes, 2001; Johnson, 2004; Kim and Morningstar, 2005; Neubert and Moon, 2006). The present literature supports the relationship between community integration, employment, and community-based experiences for high school-level students with disabilities.

Yet, too often, the transition-planning process does not include participation in community-based experiences. Storms *et al.* (2000) defined community experiences as educational activities provided outside of the school building or in a community setting. Community-based work experiences, job exploration, banking, shopping, job-site training, transportation, and recreational activities are examples of community experiences. When, and if, school personnel include community experiences during the transition process, the involvement, generally, has centered on the formal, traditional, or official organizations and agencies, rather than informal systems of support. The limited networking with informal systems of support by school personnel is particularly evident when working with families from culturally and/or linguistically diverse (CLD) backgrounds.

Community-based experiences are beneficial for all adolescents and young adults with disabilities and their families. These are especially critical for those youth who are at higher risks for future unemployment or underemployment or enrolment in postsecondary programs. This group encompasses students who manifest moderate-to-severe disabilities. Due to multifaceted racial, cultural, socioeconomic, and political reasons, youth from CLD populations with disabilities at all levels (e.g., mild-severe) are also at great risk for poor postsecondary outcomes.

Literature critiquing the transition-planning research (Ford, 2002, 2004, 2006; Geenen *et al.*, 2001; Hughes, 2001; Kohler, 1994) delineates several barriers to the systemic provision of relevant community-based experiences and the utilization of informal systems of support during the transition-planning process for young adults with disabilities. Three dominant barriers are: (1) the inconsistent use of research-based practices as a framework to guide the transition-planning process, (2) the lack of a person-centered approach, and (3) the lack of attention to important cultural and/or linguistic diversity in families and youth. This article discusses the benefits of community integration and employment of youth with special needs.

Frameworks that Inform Transition Planning and Postsecondary Education

The majority of the literature on transition and postschool outcomes for students with disabilities has focused on

employment and postsecondary education because the primary, mainstream determinant of a high-quality life is a good education and a secure job with a salary that fosters independence. Recognizing the importance of a person-centered approach when addressing the unique needs of youth with disabilities, this mainstream perspective must not serve as the ultimate framework that guides decision making during the transition process. For many people with disabilities – particularly those with severe disabilities – issues of employment and community participation must be related to other aspects of adult life. The place of residence, access to transportation, a variety of recreational opportunities, a network of relationships, and necessary specialized supports and services are all connected to an adequate quality of life for adults with disabilities. In addition, scholars advocating on behalf of youth from CLD backgrounds with disabilities highlight the importance of attending to cultural differences in determining what constitutes successful outcomes of the transition-planning process.

Educators have debated, for some time, that preparation for independent living and community participation are significant functions of the educational process. In spite of this, educational interventions for community participation are commonly left out of the transition-planning process. Consequently, adults with disabilities and their families are left unprepared for the challenges they will confront following graduation from secondary education – when they are no longer eligible for a free appropriate public education (FAPE) (Baer and Daviso, 2006). Furthermore, transition planning has, traditionally, been designed and implemented from a monocultural perspective (Geenen, *et al.*, 2001; Kim and Morningstar, 2005). To this end, community experiences that may be particularly relevant to CLD youth and their families – such as interactions with significant informal systems of support – have not been included in the transition-planning process.

The patterns of regular paid employment for some young adults with disabilities have improved (National Center on Secondary Education and Transition, 2003). Overall, however, the literature continues to reveal poor outcomes. Unemployment, financial dependence, and the lack of social relationships are outcomes faced by many students with disabilities when they leave high school. Webb (2003) noted that unemployment rate among persons with disabilities remains a staggering 70%. With regard to race/ethnicity, The National Center on Secondary Education and Transition Study-2 (National Center on Secondary Education and Transition, 2003) reported that employment rates continue to be higher for white youth (62%) than for African American (42%) or Hispanic youth (36%). When employed, African American youth are more likely to earn lower wages than white youth. In addition, African American and Hispanic students with disabilities

have lower rates of enrollment in postsecondary education programs than white youth. Discrepancies in geographic location are also noted; data outcome for youth with disabilities in inner cities are not positive. They are less likely than their peers in suburban or rural areas to graduate from high school and are more likely to drop out of school.

The phenomenon of poor outcomes for adolescents with disabilities is reported in the international arena. Hughes' (2001) discussion of the challenges surrounding transition from an international perspective found similar rates of unemployment and disengagement from the community. Until serious attention is given to the examination of the multidimensional factors influencing a quality transition-planning process for young adults with disabilities – including those from CLD backgrounds – their future lives will remain compromised and limited. Community-based participation by youth (and their families) in formal and informal systems of support should be deemed important strategies during the transition-planning process. During this process, school personnel, often, advocate for and support families receiving assistance from formal systems of support (e.g., medical and mental agencies and social services). However, generally, school personnel are not aware of – or choose not to utilize – meaningful informal systems of support operating within communities of CLD youth. The present student demographic changes necessitate serious attention to cultural aspects of the educational delivery process, including transition planning. Over one-third of the students in US public schools are from CLD backgrounds; and the three fastest-growing groups are Hispanic Americans, African Americans, and Southeast Asian Americans.

The expectation of community participation – including employment – anticipated by an individual with or without disabilities can vary greatly. Social integration is a complex process including a person's cultural background, family expectations, socioeconomic status, and individual expectations. According to Partington (2005), a "participatory approach" suggests that people with disabilities and their families become involved in "naturally occurring networks within their own communities." Earlier, Garcia (1991) remarked that public schools are a community affair – made up of children who tend to reflect their human communities. The school's community consists of varied social groups that interact with each other, developing cooperative and interdependent networks of relationships. In Ford's (2004, 2006) discussion of quality school–community partnerships within culturally diverse communities, she used the term, Significant Multicultural Community Resources (SMCR), to describe these potential systems of support for youth and families. According to Ford, they include not-for-profit service or social organizations, sororities, fraternities, clubs or agencies, religious groups or churches, and individuals who local community residents perceive as

providing valuable significant services. The following list some benefits derived from the SMCR:

- Increased parental involvement.
- Avenues for sensitivity toward culturally relevant transition programming through face-to-face encounters between administrators, teachers, and CLD families and community resource persons (e.g., informal systems of supports).
- Resilience-enhancing resources through accessible adult role models, mentors, and advocates.
- Exposure to a network of community-based job opportunities and job-related skills and experiences.
- Reinforcement of school-related academic, life, and social skills.
- Exposure to self-enhancing/affirming activities (e.g., development of values and cultural group identity; positive decision-making skills; goal setting; rites of passage).
- Forum for dissemination and collection of information.

Legal Mandates for Community Integration and Employment

Students with disabilities receiving specialized services for employment and community participation in secondary education are served under the Individuals with Disabilities Education Act (IDEA), passed in 1990 and 1997 and, later, passed as the Individuals with Disabilities Education Improvement Act (IDEIA) of 2004. This later law defined the term transition services as a coordinated set of activities for a child with a disability that is designed to be within a results-oriented process, and is focused on improving the academic and functional achievement of the child with a disability in order to facilitate the child's movement from school to postschool activities (including postsecondary education, vocational education, integrated and supported employment, continuing adult education, adult services, independent living, and community participation). This law focuses on the individual child's needs, taking into account the child's strengths, preferences, and interests and includes instruction, related services, community experiences, the development of employment, and other post-school adult living objectives. Under section 606, IDEIA is written to ensure that professionals serving students with disabilities make positive efforts to employ and advance employment for qualified individuals with disabilities in programs.

Storms *et al.* (2000) argued that community experiences are educational activities provided outside of the school building or in a community setting. Community-based work experiences, job exploration, banking, shopping, job-site training, transportation, and recreation activities are examples of community experiences. They,

further, noted that employment services are supposed to assist in the preparation of an individual for participation in a future career or job placement. Therefore, for individuals with disabilities to reach their desired level of community integration and employment, they will need to attain the critical skills needed for success. The skills required for integrated community participation and employment overlap considerably with communication skills, self-determination skills, independent living skills, and recreational skills (Sitlington and Clark, 2006). These skills should be addressed as needed through the students individualized education plan (IEP). In the United States' public school system, planning for community participation and employment primarily occurs in the transition section of the IEP. At age 16, a transition plan is required for students served under the law. The IEP transition team is required to collect data from many sources. Assessments, community and classroom observations, nondiscriminatory formal and informal testing, medical evaluation, and family history may be used to create transitional goals and determine eligibility for adult services upon school exit. Clearly, this information may be used to generate the student's summary of performance (SOP) – which is required under the IDEIA, which states that “for a child whose eligibility under special education terminates due to graduation with a regular diploma, or due to exceeding the age of eligibility, the local education agency shall provide the child with a summary of the child's academic achievement and functional performance, which shall include recommendations on how to assist the child in meeting the child's postsecondary goals.” The transition plan and SOP should be designed to bridge the gap between secondary education and adult life.

The Search for Promising Practices for Successful Community Integration

Secondary education professionals have become increasingly concerned with the employment and educational outcomes that students experience after leaving high school. Employment status is just one valued adult role, but it is a critical indicator of adult success in the United States (Eisenman, 2003). The skills needed by workers to compete in today's workforce are complex and continually changing (Harvey, 2002). Complicating matters further, many adolescents and young adults with disabilities not only have limited knowledge of the career decision-making process, but also lack awareness of career options, and the adequate skills for employment (Hitchings *et al.*, 2001). There are a number of empirical and nonempirical intervention strategies that are labeled as best or promising practices in the provision of transition services because they help to facilitate successful movements of youth with disabilities from school to adult life (Greene and Kochar-Byrant, 2003).

Earlier, Kohler (1993) found that only four of the 11 key components of transition services that are considered to be best practices were supported by empirical data. She added that research-based recommended practices, such as community-based instruction, parent involvement, vocational training, paid work, and social skills training, have been found to be inconsistently implemented in secondary transition programs.

Lehman *et al.* (2002) identified a set of generally agreed-upon skills that society expects youth to achieve by the time they leave school. Among these are the ability to (1) live independently, (2) determine an initial career path, (3) find and maintain meaningful employment and/or postsecondary education, (4) establish fulfilling relationships with family and friends, and (5) choose leisure activities that enhance a rewarding adult life. Although these skills have been identified, most youth with disabilities still do not have the ability to make choices without additional influence from stakeholders in the educational process. Later, Baer *et al.* (unpublished) noted the lack of consensus concerning the purpose of education, and the unclear definition of transition services. They identified common themes and concepts recognized as promising practices by policymakers, advocates, and researchers in transition. These concepts include self-determination (social skills training), ecological approaches (use of formal and informal supports), individualized backward planning, service coordination (interagency collaboration), community experiences (paid work experience), access and accommodation technologies (assistive technology), supports for postsecondary education, systems-change strategies (vocational career education, secondary curricular reform, and inclusion), and family involvement in the transition process. Apparently, many factors contribute to postschool success for students with disabilities (Eisenman, 2003). The reality is that secondary education should avoid a constricted academic focus and focus on a comprehensive curriculum of academic study, vocational experiences, and life skills in order to improve the rate of high school graduation and the transition to postsecondary education and employment.

Conclusion

One lives in a rapidly changing, multicultural, highly technological society. Youth – including those with disabilities – must be provided with educational services that prepare them to function within their immediate community and/or the community at-large. At present, an unacceptable number of these youth remain at risk for poor postsecondary outcomes. The current emphasis on educational accountability demands that we address the barriers that both define and limit their successful transitioning from high school into contributing citizenry.

The transition-planning process serves as the educational technique to help ensure that young adolescents with disabilities – upon completing high school – are prepared to integrate into their respective communities (and families) or engage in postsecondary activities that promote productive lives. This article emphasizes the provision of meaningful, community-based experiences as a best practice during transition planning. Schools networking with and utilizing informal systems of support via significant multicultural resources is offered as a best-viable practice strategy when working with all students, including CLD youth and their families.

See also: Cultural and Linguistic Diversity and Children with Special Needs; Early Transition of Children with Special Needs; Overrepresentation of Culturally and Linguistically Diverse Students in Special Education; Postsecondary Participation of Students with Special Needs; School Dropout Prevention; Transition from School to Adult Life; Workforce Issues in Special Education.

Bibliography

- Baer, R. and Daviso, A. (2006). Independent living and community participation. In Flexer, R., Simmons, T., Luft, P., and Baer, R. (eds.) *Transition Planning for Secondary Students with Disabilities*, 3rd edn, pp 290–314. Upper Saddle River, NJ: Prentice Hall.
- Baer, R. M., Flexer, R. W., and Dennis, L. (unpublished). *Using Cluster Analysis in Understanding the Career Paths of Students with Disabilities*.
- Eisenman, L. T. (2003). Theories in practice: School-to-work transition for youth with mild disabilities. *Exceptionality* **11**, 89–112.
- Ford, B. A. (2002). African American community resources: Essential education enhancers for African American children and youth. In Obiakor, F. E. and Ford, B. A. (eds.) *Creating Successful Learning Environments for African American Learners with Exceptionalities*, pp 159–173. Thousand Oaks, CA: Corwin Press.
- Ford, B. A. (2004). Preparing special educators for culturally responsive school-community partnerships. *Teacher Education and Special Education* **27**(3), 224–230.
- Ford, B. A. (2006). Culturally responsive school-community partnerships: Strategy for Success. In Landsman, J. and Chance, L. W. (eds.) *White Teachers/Diverse Classrooms*, pp 286–300. Sterling, VA: Stylus.
- Garcia, R. L. (1991). *Teaching in a Pluralistic Society: Concepts, Models, Strategies*. New York: Harper Collins.
- Geenen, S., Powers, L. E., and Lopez-Vasquez, A. (2001). Multicultural aspects of parent involvement in transition planning. *Exceptional Children* **67**(2), 265–282.
- Greene, G. and Kochhar-Bryant, C. A. (2003). *Pathways to Successful Transition for Youth with Disabilities*. Columbus, OH: Merrill Prentice Hall.
- Harvey, M. H. (2002). Comparison of postsecondary transitional outcomes between students with and without disabilities by secondary vocational education participation: Findings from the National Education Longitudinal Study. *Career Development for Exceptional Individuals* **25**, 99–121.
- Hitchings, W. E., Luzzo, D. A., Ristow, R., *et al.* (2001). The career development needs of college students with learning disabilities: In their own words. *Learning Disabilities Research and Practice* **16**, 8–17.
- Hughes, C. (2001). Transition to adulthood: Supporting young adults to access social, employment, and civic pursuits. *Mental Retardation and Developmental Disabilities Research Reviews* **7**, 84–90.

- Johnson, D. R. (2004). Supported employment trends: Implications for transition-age youth. *Research and Practice for Persons with Severe Disabilities* **29**(4), 243–247.
- Kim, K. H. and Morningstar, M. E. (2005). Transition planning involving culturally and linguistically diverse families. *Career Development for Exceptional Individuals* **28**(2), 92–103.
- Kohler, P. D. (1994). Best practices in transition: Substantiated or implied? *Career Development for Exceptional Individuals* **16**, 107–119.
- Lehman, C. M., Clark, H. B., Bullis, M., Rinkin, J., and Castellanos, L. A. (2002). Transition from school to adult life: Empowering youth through community ownership and accountability. *Journal of Child and Family Studies* **11**(1), 127–141.
- National Center on Secondary Education and Transition (2003). *NCSET NLTS2 Data Brief 2*(2). www.ncset.org/publications/printresource.asp?id=1310 (accessed October 2009).
- Neubert, D. A. and Moon, M. S. (2006). Postsecondary settings and transition services for students with intellectual disabilities: Models and research. *Focus on Exceptional Children* **39**(4), 1–8.
- Partington, K. (2005). What do we mean by our community? *Journal of Intellectual Disabilities* **9**(3), 241–251.
- Sitlington, P. L. and Clark, G. M. (2006). *Transition Education and Services for Students with Disabilities*, 4th edn. Boston, MA: Allyn and Bacon.
- Storms, J., O'Leary, E., and Williams, J. (2000). *Transition Requirements: A Guide for States, Districts, Schools, Universities, and Families*. Minneapolis, MN: University of Minnesota, Institute on Community Integration.
- Webb, S. (2003). Independent living and employment services: Equal paths to community integration. *Journal of Vocational Rehabilitation* **18**, 125–130.

Cooperative Learning for Children with Special Needs

E A Dooley, West Virginia University, Morgantown, WV, USA

K R Kossar, California University of Pennsylvania, California, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cooperative learning – A method where students are paired or grouped and work together to accomplish shared goals.

Cue – A prompt provide by the teacher to initiate a response from students.

Differential instruction – Teaching and assessing students according to readiness level, interests, and preferred mode of learning.

Inclusion – A process in which students with disabilities are educated with their same-age peers who are not disabled.

Reinforcement – An event following behavior that increases the likelihood of reoccurrence of that behavior.

Social interdependence theory – Premise that states that individuals in a group will work together to meet the group goal.

Socialization – A Process of inheriting or learning norms, customs; and ideologies to participate in society.

Cooperative Learning for Students with Special Needs

Cooperative learning is a concept that grew out of the field of social psychology. Because of its unique characteristics and effectiveness, it is commonly adopted as a teaching strategy in special education whereby students engage in collaboratively designed activities to achieve a common goal or task. Various theories have been associated with cooperative learning; and they have framed the context in which cooperative learning serves as an intragroup learning method that is outcome oriented. For instance, social interdependence theory has contributed to the development of cooperative learning theory. Kurt Levin – in the mid-1930s – was one of the first to introduce the concept of group dynamics and to show how each individual in the group contributes to a common goal, thus becoming a dynamic whole. This theory was fully conceptualized by Morton Deutsch, in 1949. He is recognized as the theorist

who fully conceptualized social interdependence theory and provided the fundamental concepts for cooperative learning. The primary principle supporting the theory is the notion that “humans are primarily concerned with developing organized and meaningful views of their world by perceiving events as integrated wholes rather than a summation of parts or properties” (Johnson and Johnson, 1989: 287). As a result, individuals within a group rely upon each other to achieve common goals. Deutsch penned two types of social interdependence: positive interdependence and negative interdependence. He referred to positive interdependence as the phenomena when individuals believe they can only achieve goals when there is an interactive group working toward a common goal. Negative interdependence exists when individuals believe they can accomplish goals only when their competitors fail to obtain their goals.

The cooperative learning strategy as used in education, generally – and, specifically, in special education – relies heavily upon the concept of positive interdependence. When operationalized, the social interdependence theory translates into a viable instructional method that requires the participation of several group members. The social identification theory is another theory that informs the cooperative learning-teaching strategy. This theory suggests that members of the group experience a collaborative awareness. Lau (1989) reported that group identification is essential to the social interdependence and social identification theories. During collaborative awareness, each member is aware of the group dynamics and responds as a contributor. To implement cooperative learning groups, group and social identifications are necessary. In group identification, individuals have a working relationship with each member in the group and there is a collective awareness of collective group membership; whereas, in social identification, individuals perceive themselves as being a part of a social group. Two key concepts come into play when considering cooperative learning: (1) task interdependence, and (2) outcome interdependence. Task interdependence implies that individuals collectively share the task, and all experience a mutual benefit. In order to complete the task, cooperation and input from several individuals are critical. Outcome interdependence – which is defined as “the degree to which the significant outcomes an individual receives depend on the performance of others” (Wagerman, 1995: 147) – encompasses the results achieved from efforts

contributed by each of the group members. In theory, there are a variety of dynamics during group interaction. When cooperative groups are convened, there is a specific job that needs to be done, and the function of that job requires the participation of a group wherein each member is assigned a specific task (Johnson and Johnson, 1989). In this article, the focus is on how cooperative learning can be utilized to help students with special needs.

Cooperative Learning: Definition, Types, and Fundamental Principles

Students work together to accomplish shared goals (Johnson and Johnson, 1989). Cooperative learning for students with special needs is, typically, a method to reinforce or supplement teacher instruction (Mercer and Mercer, 2005). In this age of inclusion, many general educators or co-teachers (i.e., general educators and special educators) can utilize this method to effectively instruct whole groups followed by individualization to reinforce skills. Generally, cooperative learning can be considered either formal or informal based upon the structure. Formal structures include a sequence of events, led by the teacher, which includes the following: (1) introduce a task or lesson; (2) assign students to groups based upon a pre-determination of members; (3) provide students with materials, resources, and clarity to complete tasks; (4) assign roles to group members; and (5) teach prerequisite skills. Based upon the task or roles, students then work as a group until the task is completed. It is important to note that the teacher is continuously monitoring and evaluating the progress of the groups. Once the task is completed, all group members – along with the teacher – process or reflect upon the strengths/weaknesses of group members. As opposed to formal structures that may take a measure of time to complete – thus strengthening the group dynamics – informal structures are generally shorter in scope, lasting only perhaps a class period/lecture. In this structure, a temporary group is formed for a short period of time to, perhaps, review lecture material (Johnson and Johnson, 1989). Whether formal or informal, cooperative learning has been represented in many different forms, types, and names, thus making an exhaustive list difficult. See **Table 1** for the most common types, characteristics, or uses of those types, and subtypes or common names of cooperative learning.

It is necessary to understand the five elements at the core of cooperative learning, irrespective of the type. The five elements include the following: positive goal interdependence, individual accountability, face-to-face promotive interaction, appropriate use of social skills, and group processing (Putnam *et al.*, 1996). The following sections discuss each element of cooperative learning:

Table 1 Types and characteristics of cooperative learning

Type	Characteristics/uses	Subtypes/common names
Tutoring	One-on-one student interaction with one serving as tutor and one as tutee Can be matched based on various criteria	Peer Cross-age Structured Team interviews
Class-wide peer tutoring (CWPT)	Classroom separated into dyads in which one-on-one tutoring takes place to reinforce teacher-led lesson	Peer-assistance learning strategies (PALS) Question answer relationships (QAR)
Class-wide peer tutoring (CWPT) with curriculum-based measurement (CBM)	Groups work on material that is geared toward individual needs	Class-wide student tutoring teams (CSTT)
Cooperative learning structures Student team learning model	Students assigned to heterogeneous groups/teams and given an assignment Teams divide tasks, approach task and are evaluated as whole group Group materials may be identical or individualized toward group members' abilities	Jigsaw Learning together Student-teams-achievement-divisions Teams-games-tournaments (TGT) Team-assisted individualization (TAI) – Mathematics, grades 3–6 Cooperative integrated reading and composition (CIRC) – Reading, writing, and language arts, grades 3–5 Academic controversy Group investigation

Element 1: Positive Goal Interdependence

This element follows the assumption that all individuals involved in a task recognize that, in order to succeed, all members must succeed. The failure of any one member leads to the failure of the group – thus, the term, interdependence. All students, including students with disabilities in cooperative groups must maximize their own productivity in order to maximize the productivity of the collaborative group (Johnson and Johnson, 1989).

For example, if teams are assigned to play a review game that consists of both students with special needs and those without, each student must play the game to his/her full potential or the team will not win the game.

Element 2: Individual Accountability

In addition to the interdependence among cooperative group members, there must also be individual accountability for performance. While all group members must work together toward the goal, each member (or student) is also accountable for his/her own contribution or lack thereof. Recognition of individual achievement is paramount. In the cooperative group, one is responsible for doing his/her best work, thus motivating other group members to do their best work. In the above example, the element would be recognized when each team member's role in the game is clear and success in that role is reinforced.

Element 3: Face-to-Face Promotive Interaction

Within the cooperative group task, each member must promote other team members. This element can take the form of encouragement, constructive feedback, and trust. In the above example, team members would cheer each other on – helping them to answer game questions when the need arises.

Element 4: Appropriate Use of Social Skills

Students with special needs, often, lack the appropriate social skill set to interact in cooperative learning arrangements (Mercer and Mercer, 2005). For this element to be recognized, all students must be taught the rules of socialization prior to placement in a cooperative learning group. In the above example, all students would learn the rules of the game and the appropriate reactions (verbal and nonverbal) to winning or losing the game, communicating, and turn-taking prior to team formation.

Element 5: Group Processing

During and following a cooperative learning task, all group members must reflect on and identify those actions and decisions that were effective or ineffective in a constructive manner. The purpose of this group processing is to continually improve the team; it is an ongoing process. In the above example – once the review game has ended – both the winning and losing teams would identify ways in which team members' efforts could be improved (e.g., studying, taking turns, encouraging players).

Using Cooperative Learning for Students with Special Needs

There has been an abundance of research conducted on cooperative learning as a best practice for students with special needs. In an era of evidence-based practice to meet federal mandates and differentiated instruction to meet the needs of a variety of abilities, many researchers and practitioners have identified methods based on the ease of use and on student gains. The majority of the research on cooperative learning centers on academic achievement and social impact.

The majority of research on cooperative learning has focused on its impact on the academic achievement of students with special needs. Researchers have conducted studies both at the elementary and secondary levels – with a focus on mild-to-severe disabilities, in a variety of content areas (although most focus on the elementary grades), and with regard to both those who have the disability and those who do not. Most types of cooperative learning arrangements have been studied. Interestingly, the results have been mixed; however, the predominant view is that cooperative learning as a best practice is effective since an increase in achievement is noted (even if there is no statistical significance). For example, Johnson *et al.* (2000) conducted a meta-analysis of 194 studies that involved impact of cooperative learning on academic achievement. The analysis summarized that all cooperative learning arrangements increased academic achievement to varying degrees. Further, the studies were coded as to the following: ease of learning, ease of initial use, ease of maintaining use over time, robustness, and adaptability, and the meta-analysis yielded a ranking of the most commonly used types. In a similar yet earlier meta-analysis of 65 studies, Cohen *et al.* (1982), examined – among other factors – academic performance on both tutors and tutees. Forty-five of 52 studies indicated that tutees had increased exam scores, six studies indicated decreased exam scores, and there was no difference in score in one study. Further, of the 65 studies, 33 indicated an increase in achievement for tutors; however, only ten were considered statistically significant. Perhaps, more interesting for practitioners were the following indirect findings that suggested that the most academic achievement was found under the following conditions: when the tutoring was more structured and shorter, when lower level skills were taught, with math achievement as the dependent variable, and on locally developed tests as opposed to standardized assessment.

Clearly, there are still many holes to be filled in the research on cooperative learning. When examining middle school science, Mastropieri *et al.* (2006) found increases in spelling, science, and social studies scores. Several trends and apparent limitations have emerged from the research on cooperative learning and academics. The first is that,

irrespective of method, all students must be taught the process. In other words, structure in planning and implementation are a must for success in academics. A second recommendation is that there must be frequent feedback, consistent expectations, cues, and reinforcement along with ongoing assessment. In many research studies, cooperative learning has been known to impact attitudes toward subject matter. Earlier, Cohen *et al.* (1982) found eight studies related to this, in which all eight reported an increase in positive attitudes toward the subject by those who were tutored; however, only one of those studies reported results that were statistically significant. Additionally, of five studies that reported on the effect of cooperative learning on the attitudes of the tutors, four reported more positive attitudes. Interestingly, one student in the control group also reported more positive attitudes toward the subject matter. Finally – in many other research studies – cooperative learning has been found to impact social adjustment including factors such as self-esteem, acceptance (by both students with and without disabilities), and social skills. As with the other two areas of research, there has been much variability in findings, and the same limitations apply. In the Cohen *et al.* meta-analysis, nine studies were found in which self-concept was studied, and none found statistical significance in relation to the tutees; however, four of 16 indicated statistical significant findings in relation to the tutors and self-esteem. The authors suggested that perhaps self-esteem could have been attributed to a variety of classroom variables and not necessarily the method.

Conclusion

Cooperative learning has proven to be an effective teaching and learning tool for students with special educational needs. The different types and uses of cooperative learning enable the teacher to select learning strategies based upon students' needs and styles of learning and the heterogeneous nature of the learning task. Clearly, task interdependence and outcome interdependence are critical to the success of cooperative learning groups, goals, and objectives. For those who would prefer a more comprehensive view of cooperative learning and children with special needs, please see the additional resources listed in the sections titled 'Bibliography' and 'Further reading.'

See also: Educating Students with Special Needs: An Overview; Peer-Tutoring; Prereferral Strategies; School-Based Services for Children with Special Needs; Social Competence.

Bibliography

Cohen, P. A., Kulick, J. A., and Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal* 19(2), 237–248.

Johnson, D. W. and Johnson, R. T. (1989). *Cooperation and Competition: Theory and Research*. Edina, MN: Interaction Book Company.

Johnson, D. W., Johnson, R. T., and Stanne, M. B. (2000). Cooperative learning methods: A meta-analysis. <http://www.co-operation.org/pages/cl-methods.html> (accessed August 2009).

Lau, R. (1989). Individual and contextual influences of group identification. *Social Psychology Quarterly* 52(3), 220–231.

Mastropieri, M. A., Scruggs, T. E., Norland, J. L., *et al.* (2006). Differentiated curriculum enhancement in inclusive middle school science: Effects on classroom and high-stakes tests. *Journal of Special Education* 40(3), 130–137.

Mercer, C. D. and Mercer, A. R. (2005). *Teaching Students with Learning Problems*, 7th edn. Upper Saddle River, NJ: Merrill.

Putnam, J., Markovchick, K., Johnson, D. W., and Johnson, R. T. (1996). Cooperative learning and peer acceptance of students with learning disabilities. *Journal of Social Psychology* 136(6), 741–752.

Waggener, R. (1995). Interdependence and group effectiveness. *Administrative Science Quarterly* 40, 145–180.

Further Reading

Avcioglu, H. (2007). Examining the effectiveness of a program developed for teaching social skills to hearing impaired students based on cooperative learning. *Educational Sciences: Theory and Practice* 7(1), 340–347.

Carter, E. W., Chushing, L. S., Clark, N. M., and Kennedy, C. H. (2005). Effects of peer support interventions on students' access to the general curriculum and social interactions. *Research and Practice for Persons with Severe Disabilities* 30(1), 15–25.

Fish, W. W. (2006). Cooperative learning: Why the reluctance? *Academic Exchange Quarterly*. <http://www.thefreelibrary.com/Cooperative+learning:+why+the+reluctance%3F-a0155567993> (accessed August 2009).

Fisher, J. B. (1995). Searching for validated inclusive practices: A review of the literature. *Focus on Exceptional Children* 28(4), 1–20.

Gisbert, D. D. and Font, C. M. (2008). The impact of peer tutoring on the improvement of linguistic competence, self-concept as a writer and pedagogical satisfaction. *School Psychology International* 29(4), 481–499.

Gnadinger, C. M. (2008). Peer-mediated instruction: Assisted performance in the primary classroom. *Teachers and Teaching* 14(2), 129–142.

Hancock, D. (2004). Cooperative learning and peer orientation effects on motivation and achievement. *Journal of Educational Research* 97(3), 159–166.

Harriman, N. and Schmidt, M. W. (1997). Strategies for inclusive teaching in rural schools. Unpublished manuscript.

Huss, J. A. (2006). Education and cooperative learning: A miss or a match? *Gifted Child Today* 4(29). http://www.masafloida.org/pdf/Gifted_Education.pdf (accessed August 2009).

Jacques, N., Townsend, M., and Wilton, K. (1998). Cooperative learning and social acceptance of children with mild intellectual disability. *Journal of Intellectual Disability Research* 42(1), 29–36.

Johnson, D. W. and Johnson, R. T. (2005). New developments in social interdependence theory. *Genetic, Social, and General Psychology Monographs* 131(4), 285–358.

Kagan, S. (1994). *Cooperative Learning*. San Clemente, CA: Kagan Publishing.

Leithwood, K. and Jantzi, D. (2008). Linking leadership to student learning: The contributions of leader efficacy. *Educational Administration Quarterly* 44(4), 496–528.

Malmgren, K. W. (1998). Cooperative learning as an academic intervention for students with mild disabilities. *Focus on Exceptional Children* 31(4), 1–8.

Menchaca, M. P. and Bekele, T. A. (2008). Learner and instructor identified success factors in distance education. *Distance Education* 29(3), 231–252.

Moriyama, J., Kato, Y., Aoki, Y., *et al.* (2008). Self-efficacy and learning experience of information education: In case of junior high school. *AI and Society* 23(2), 309–325.

- Offir, B., Lev, Y., and Bezalel, R. (2008). Surface and deep learning processes in distance education: Synchronous versus asynchronous systems. *Computers and Education* **51**(3), 1172–1183.
- Williams, P. (2007). Children teaching children. *Early Childhood Development and Care* **177**(1), 43–70.
- Yaoying, X., Gelfer, J. I., Sileo, N., Filler, J., and Perkins, P. G. (2008). Effects of peer tutoring on young children's social interactions. *Early Childhood Development and Care* **178**(6), 617–635.
- Yin-kum, L. (2008). Effects of cooperative learning on second graders' learning from text. *Educational Psychology* **28**(5), 567–582.

Relevant Websites

- <http://www.worksheetlibrary.com> – Worksheet Library, 10 Tips for Cooperative Learning.
- [http://edtech.kennesaw.edu/intech/cooperative learning.htm](http://edtech.kennesaw.edu/intech/cooperative%20learning.htm) – Cooperative Learning Websites.
- http://www.teach-nology.com/currenttrends/cooperative_learning – Teachnology: The online teacher resource.
- <http://www.thirteen.org/edonline/concept2class/coopcollab/index.html> – Concept to classroom workshop.
- http://www.vanderbilt.edu/cft/resources/teaching_resources/activities/cooperative.htm – Center for Teaching.
- <http://olc.spsd.sk.ca/DE/PD/instr/strats/coop/index/html> – Instructional strategies online.
- <http://www.studygs.net/cooplearn.htm> – Study guides and strategies.
- <http://www.sedl.org/scimath/compass/v01n02/> – Classroom compass.
- <http://www.ericdigests.org/1995-1/elements.htm> – The essential elements of cooperative learning in the classroom.
- <http://www.specialconnections.ku.edu/cgi-bin/cgiwrap/speconn/main.php?cat=instruction§ion=main&subsection=udl/cooperative> – Special connections-connecting teachers to strategies that help students with special needs successfully access the general education curriculum.

Co-Teaching

E D Cramer, Florida International University, Miami, FL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Co-teaching – When two or more professionals share instruction for a single group of students in the same class, it is known as co-teaching.

Parity – It refers to perceived equality.

Federal legislative changes, such as those described by the Individuals with Disabilities Education Improvement Act (IDEIA) reauthorized in 2004 (Pub. L. No. 108-466) and the No Child Left Behind (NCLB) Act of 2001 (Pub. L. No. 107-110), require that students with increasingly diverse learning characteristics have access to and achieve high academic performance in the general education curriculum. To illustrate the recent increase in student diversity, data from the US Department of Education (2005) indicated that between 1994 and 2004, the percentage of students with disabilities spending 80% or more of the school day in a regular classroom increased by nearly 49.9–50% in 2004. These proportions can be expected to increase given the national trends of the past three decades and IDEIA's requirement to include students with disabilities as full participants in rigorous academic and general education curriculum and assessment. This leads to a need for increased collaborative planning and teaching among school personnel attempting to comply with legal mandates. One increasingly popular vehicle to provide this service is co-teaching (Morocco and Aguilar, 2002; Murawski and Dieker, 2004; Villa *et al.*, 2004).

Many different definitions exist to describe co-teaching. A co-teaching team can be defined as two or more people sharing responsibility for teaching some or all of the students assigned to a classroom. It involves the distribution of responsibility among people for planning, instruction, and evaluation for a classroom of students (Villa *et al.*, 2008). Both educators on the co-teaching team are responsible for instructional planning and delivery, assessment of student achievement, and classroom management. Friend and Cook (2007) expanded this definition stating, “co-teaching occurs when two or more professionals jointly deliver substantive instruction to a diverse, blended group of students in a single, physical space” (p. 113). In their definition of co-teaching, they noted the importance of both parties being appropriately credentialed professionals. Salend (2008) referred to

cooperative or collaborative teaching as teachers who share responsibility and accountability for planning and delivering instruction, evaluating, grading, and disciplining students. The common theme among all of these definitions is that students are not removed from class to receive services but rather the services (i.e., academic instruction and supportive services) are provided in the general education classroom by two or more professionals who share responsibility for the students. This article focuses on this premise.

Rationale for and Elements of Co-Teaching Teams

Friend and Cook (2007) noted two rationales for co-teaching: to meet the educational needs of students with diverse learning abilities and to provide curricular access to students with disabilities. Co-teaching allows general and special education professionals to pool together their expertise. Most often, general education teachers have more in-depth knowledge about specific subject areas, whereas the special education teachers generally know more about differentiating instruction to meet the individual needs of students. This combined expertise provides both general and special education students with content knowledge taught effectively through group and individualized strategies. Although co-teaching professionals may not simply be general and special educators, (e.g., speech and language pathologists, occupational therapists, and literacy coaches), the fact that they are equal in credentialing and status ensures that they can, in fact, be partners in the classroom. In cases where a teacher and paraprofessional or classroom-volunteer work together to instruct students, supported instruction rather than co-teaching is occurring. While supported instruction can be a valid approach for instruction, it should not be confused with co-teaching. Similarly, when professionals are co-teaching, as both have equal and shared status, both should have equal and shared responsibility for instruction. It is a waste of resources to utilize a second licensed and trained professional in the room simply to have that person utilized as an assistant. Friend and Cook (2007) referred to situations such as this as “an inappropriate underuse of a qualified professional” (p. 115). In addition, when special education or related services personnel focus only on the student(s) with disabilities, more attention, stigma, or isolation of that student may occur therefore undoing some of

the possible social benefits of having the student included in a co-teaching situation in the first place. Under IDEIA 1997, any professional in the classroom can and should be working with all students, not simply those with disabilities.

Co-teaching is designed to minimize problems that typically occur with pull-out special education programs such as students missing academic instruction, lack of communication and coordination between professionals, scheduling problems, and fragmentation between general and special education curriculum (Wiggins and Damore, 2006). Co-teaching also potentially reduces the need for labeling students because all struggling students can benefit from the expertise of both teachers (Cramer *et al.*, under review; Wiggins and Damore, 2006). Villa *et al.* (2008) named five elements of co-teaching teams: (1) a common, publicly agreed-on goal, (2) a shared belief system, (3) parity, (4) distributed functions theory of leadership, and (5) cooperative process. A common goal may be for one particular outcome that teachers and service providers share for their students. A shared belief in the value of co-teaching is beneficial to the process. Parity refers to the perceived equality between true collaborators on a co-teaching team. In distributed leadership, both teachers make a conscious decision to share the leadership of and responsibility for the class. According to Villa *et al.*, the cooperative process finally is made up of (1) face-to-face interactions, (2) positive interdependence, (3) interpersonal skills, (4) monitoring co-teacher progress, and (5) individual accountability. Each of these elements can take place regardless of the model of co-teaching being utilized.

Friend and Cook (2007) described six approaches to co-teaching: (1) one-teaching-one-observe, (2) one teach-one-assist, (3) station teaching, (4) parallel teaching, (5) alternative teaching, and (6) team teaching. In the first and second approach, one teacher has the primary role of designing and delivering instruction while the other teacher moves around the room helping and observing individual students. In station teaching, the content of the lesson is divided between the teachers and every student moves to both teacher-led groups. Parallel teaching occurs when the class is divided and each teacher delivers the same content and instruction to his or her section of the class. Alternative teaching is employed when the special education teacher focuses on re-teaching, differentiating instruction, and making curricular accommodations and modifications for small groups of students with and without disabilities who may need extra assistance. This model can be flexible and permits any student who may need some additional help to receive additional instruction. Team teaching is used when teachers trade off roles and groups of students. For example, one teacher presents on an overhead the major events leading to the Civil War while the other teacher lists these events on a timeline on the whiteboard. In each of these models, even when

teachers are not both physically instructing students at all times, both should always be sharing the responsibility for planning and instructional decision making.

Benefits of Co-Teaching to Students and Teachers

Villa *et al.* (2008) listed some beneficial outcomes of co-teaching. They noted that:

- students develop better attitudes about themselves, show academic improvement, and enhanced social skills;
- teacher-to-student ratio is increased, leading to better teaching and learning conditions;
- teachers are able to use research-proven teaching strategies effectively;
- a greater sense of community is fostered in the classroom;
- co-teachers report professional growth, personal support, and enhanced motivation; and
- increased job satisfaction can be experienced because needs for survival, power, freedom of choice, a sense of belonging, and fun are met.

While co-teaching is an initiative that developed out of special education laws and with the needs of students with disabilities in mind, in co-teaching classrooms, all students have the benefit of two qualified professionals using a variety of teaching strategies to meet the needs of all learners. This can be particularly beneficial for students with special learning needs who may not have been identified as having a disability. This can include English language learners (ELLs), students who are socially or culturally diverse, gifted students, students at-risk for failure, or students with potential need for special education who simply have not yet been referred. Students of all abilities and levels should be integrated together in instruction. Classroom placement should not isolate 'these' students from 'those' students. If special education students are seated apart from the rest of the class, they are still, in a sense, socially isolated from their peers.

In addition, while at times, students will need to receive small homogenous group instruction, in inclusive classrooms, co-teachers should work with heterogeneous groups and allow students the opportunity to work cooperatively and learn from each other. Students who in the past would have been pulled out of the classroom can remain in one setting with their peers all day allowing them to have more smooth, uninterrupted instruction, and full access to general education curriculum. Co-teaching could be one way to increase special education students' time in and access to general education and to decrease reliance on special education placement to solve learning and behavior problems of culturally and linguistically

diverse students (e.g., Bahamonde and Friend, 1999; Cramer *et al.*, under review; Garrigan and Thousand, 2005; Salazar and Nevin, 2005).

Partnerships between highly qualified general educators who have demonstrated subject-area expertise and highly qualified special educators who have complementary expertise in specialized learning strategies could result in implementing the research-based curricular and instructional approaches required by the NCLB Act. All students need their teachers to learn and use the most effective teaching strategies, educational materials, and lesson formats currently known to educate all students. Teachers can accomplish this by exchanging such information and expertise through their co-teaching partnerships. Typically, the general educator is an expert of content and the special educator is an expert on adapting instruction. Each can learn and benefit from the expertise of the other. If other trained professionals are involved (such as a speech and language pathologist or a literacy coach), they bring even more specialized skills to the class. In addition, they have the emotional support of another colleague that they collaborate with on a daily basis and a partner to collaborate with on classroom decisions.

Meeting the Needs of Co-Teachers

While the majority of co-teachers see co-teaching as a beneficial experience, they are consistent and clear about what needs to be in place for co-teaching to work: planning time, compatibility of co-teachers, training, appropriate student skill levels, and ongoing administrative support (Cramer and Nevin, 2006; Cramer *et al.*, 2006; Scruggs *et al.*, 2007). Time is needed for planning, meetings, and professional development. Regularly scheduled meeting times should be built into the schedule for teachers to instruct students, plan, and reflect on what has occurred. The ideal would be reserved time each day through common planning time. In addition, teachers would like to have input on whether or not they co-teach and who their partner is. Co-teaching is often compared to a marriage and if the partners are not compatible with each other, the relationship will inevitably fail.

Most future teachers graduating from preservice teacher education programs today have little to no training in co-teaching; yet they are often expected to work at least part of the day in co-teaching teams. When asked, teachers consistently point out the need for more training. Another need that teachers often refer to is for students who are being included in co-taught classrooms to have appropriate skill levels to access the required curriculum. For example, teachers express frustration if students are expected to achieve the same state grade level standards and are functioning below grade level. Despite concerns over any of the aforementioned teacher needs, the perception of administrative

support (or lack thereof) can be the ultimate factor in determining whether or not teachers feel co-teaching is working. Administrators and school leaders need to believe in co-teaching as well as provide financial support. It is easy for a co-teaching class to become a special class (that is, where a disproportionate number of students in the class have special needs). When the administrator really understands what co-teaching is and buys into the concept, benefits occur for both special education and general education children (Cramer *et al.*, 2006).

There are many barriers to co-teaching. Some of the most common include (1) limited financial resources, (2) complicated logistics of assigning caseloads to service all students, (3) class size and ratios, and (4) maintaining a positive and collaborative relationship between professionals. Additional problems often arising at the secondary level (Salend, 2008) include (1) lack of time to plan and resources, (2) unclear roles of teaching team members, (3) resistance from colleagues, (4) scheduling, (5) increased workloads, and (6) increased responsibilities (Cook and Downing, 2005; Murawski, 2006). Lack of content-area expertise is also more prevalent at secondary level (Cramer and Nevin, 2006). To counteract co-teaching barriers, Salend recommended that co-teachers do the following:

- agree on goals and establish ground rules;
- learn about each other's beliefs and teaching styles;
- understand and coordinate responsibilities, expertise, and roles of others;
- utilize a variety of scheduling arrangements;
- understand cross-cultural perspectives;
- physically arrange the classroom to support collaboration;
- hold common expectations for grading academic, behavioral, and social performance;
- develop communication, problem-solving, and team-building skills;
- understand co-teaching is an ongoing process with stages;
- work toward equal status and shared workload;
- vary arrangements of teaching based on purpose, content, and student needs;
- include families in the process;
- communicate regularly to reexamine goals;
- address conflicts immediately;
- use a range of strategies;
- engage in self-evaluation and reflection; and
- acknowledge and celebrate success.

Co-Teaching and National Teacher Standards: Meeting Goals

Teacher education standards reflect the set of knowledge and skills of co-teaching. Standards from National Board for Professional Teacher Standards (NBPTS) (2005), the

Table 1 Analysis of standards from professional organizations

INTASC ^a	CEC ^b	NBPST ^c
Standard 3 requires teachers to understand how learners differ	Knowledge and skills in understanding characteristics of learners with different cognitive, physical, cultural, social and emotional needs	Teachers adjust their practice according to individual differences in their students
Standard 4 requires teachers to use a variety of instructional strategies	Competencies related to knowledge and skills for instructional content and practice	Teachers show multiple methods to engage student learning and to enable students to reach goals
Standard 10 asks teachers to collaborate and communicate with parents, families, colleagues to support student learning	Competencies related to communication and collaborative partnerships	Teachers collaboratively work with others and coordinate services

^ahttp://www.ccsso.org/projects/Interstate_New_Teacher_Assessment_and_Support_Consortium/

^b<http://www.cec.sped.org/ps/ps-entry.html>

^c<http://www.ets.org/nbpts>

Council for Exceptional Children (CEC) (2005), and Interstate New Teacher Assessment and Support Consortium (INTASC) (2005) were analyzed for content with respect to inclusive education and collaboration or co-teaching. As shown in **Table 1**, there seems to be substantial agreement among these diverse boards with respect to knowledge and skills for differentiating instruction, working collaboratively with others, and supporting the education of diverse learners. For example, INTASC Standard 3 requires teachers to understand how learners differ; Standard 4 requires teachers to use a variety of instructional strategies; and Standard 10 asks teachers to collaborate and communicate with parents, families, and colleagues to support student learning.

In comparison, CEC standards for entry into the profession includes competencies related to knowledge and skills in understanding characteristics of learners with different cognitive, physical, cultural, social, and emotional needs; competencies related to knowledge and skills for instructional content and practice; and competencies related to communication and collaborative partnerships. These are strongly correlated with NBPTS Standard 1 – teachers adjust their practice according to individual differences in their students; Standard 3 – teachers show multiple methods to engage student learning and to enable students to reach goals; and Standard 5 – teachers collaboratively work with others and coordinate services. In the results of a study of the relationship between teaching standards and co-teaching (Cramer *et al.*, under review), co-teacher attitudes, beliefs, and actions appear to be correlated with the standards shown in **Table 1**. Since co-teaching clearly aligns with recognized professional standards in the field, teacher training at both the pre-service and in-service levels should focus on preparing professionals to work collaboratively toward the common goal of meeting the needs of all learners. This will address the needs of increasingly diverse learners, the legal requirements of access to the general education for all students,

and the practical need for all teachers to feel well-prepared and supported in their efforts to do so.

Conclusion

Co-teaching is a widely researched method of instruction whereby two trained professionals work collaboratively to plan, implement, and evaluate instruction for students with and without disabilities. Any necessary services for students with disabilities are provided inside the general education classroom and the general education curriculum is made accessible for all students with appropriate accommodations. While there are many accepted models for co-teaching, all involve the sharing of the responsibility for planning and instructional decision making.

Overall, teachers and administrators perceive co-teaching to be beneficial for teachers and for students, both with and without disabilities. However, sufficient planning time, compatibility of co-teachers, training, appropriate student skill level, and ongoing administrative support are necessary. While research has consistently been able to provide information about what occurs in co-taught classrooms and perceptions of key stakeholders, ongoing research is still necessary regarding the true implementation of co-teaching and its efficacy as a service for students.

Bibliography

- Bahamonde, C. and Friend, M. (1999). Teaching English language learners: A proposal for effective service delivery through collaboration and co-teaching. *Journal of Educational and Psychological Consultation* **10**(1), 1–9.
- Cook, L. and Downing, J. (2005). The practicalities of collaboration in special education service delivery. *Intervention in School and Clinic* **40**(5), 296–300.
- Cramer, E. D. and Nevin, A. I. (2006). A mixed methodology analysis of co-teacher assessments: Implications for teacher education. *Teacher Education and Special Education* **29**(4), 261–274.

- Cramer, E., Nevin, A., Salazar, L., and Landa, K. (2006). Co-teaching in an urban, multicultural setting: Research report. *Florida Educational Leadership* 7(1), 43–50.
- Cramer, E.D., Salazar, M., and Marshall, D. (2008) Urban inclusive practices: A mixed methodological examination. *National Journal of Urban Education and Practice* 1(4), 385–397.
- Council for Exceptional Children (2005). *CEC International Standards for Entry into Professional Practice*. Arlington, VA: Author. Retrieved April 19, 2005: <http://www.cec.sped.org/ps/ps-entry.html>.
- Friend, M. and Cook, L. (2007). *Interactions: Collaboration Skills for School Professionals*, 5th edn. Boston, MA: Pearson Education.
- Garrigan, C. M. and Thousand, J. S. (2005). The effects of co-teaching on student achievement in the reading domain. *New Hampshire Journal of Education* 8, 56–60.
- Interstate New Teacher Assessment and Support Consortium (INTASC) (2005). *Interstate New Teacher Assessment and Support Consortium Standards*. Washington, DC: Council of Chief State Schools Officers.
- Morocco, C. and Aguilar, C. (2002). Co-teaching for content understanding: A school-wide model. *Journal of Educational and Psychological Consultation* 13(4), 315–347.
- Murawski, W. W. (2006). Student outcomes in co-taught secondary English classes: How can we improve? *Reading and Writing Quarterly* 22(3), 227–247.
- Murawski, W. W. and Dieker, L. A. (2004). Tips and strategies for co-teaching at the secondary level. *Teaching Exceptional Children* 36(5), 52–58.
- National Board for Professional Teacher Standards (NBPTS) (2005). *Performance-Based Teaching Assessments*. Princeton, NJ: Educational Testing Services.
- Salazar, L. and Nevin, A. (2005). Co-teachers in an urban multicultural school. *Florida Educational Leadership* 5(2), 15–20.
- Salend, S. J. (2008). *Creating Inclusive Classrooms: Effective and Reflective Practices*, 6th edn. Upper Saddle River, NJ: Pearson Education.
- Scruggs, T. E., Mastropieri, M. A., and McDuffie, K. A. (2007). Co-teaching in inclusive classrooms: A meta-synthesis of qualitative research. *Exceptional Children* 73(4), 392–416.
- US Department of Education, National Center for Education Statistics. (2005). *The Condition of Education 2005* (NCES 2005–094), Table 27-1.
- Villa, R. A., Thousand, J. S., and Nevin, A. I. (2004). *A Guide to Co-Teaching: Practical Tips for Facilitating Student Learning*. Thousand Oaks, CA: Corwin Press.
- Villa, R. A., Thousand, J. S., and Nevin, A. I. (2008). *A Guide to Co-Teaching: Practical Tips for Facilitating Student Learning*, 2nd edn. Thousand Oaks, CA: Sage.
- Wiggins, K. C. and Damore, S. J. (2006). “Survivors” or “Friends”? A framework for assessing effective collaboration. *Teaching Exceptional Children* 38(5), 49–56.

Further Reading

- Dieker, L. A. (1998). Rationale for co-teaching. *Social Studies Review* 37(2), 62–65.
- Nevin, A. I., Cramer, E. D., Voigt, J., and Salazar, L. (2008). Instructional modifications, adaptations, and accommodations of co-teachers who loop: A descriptive case study. *Teacher Education and Special Education* 31(4), 283–297.
- Swanson, H. L. (2000). What instruction works for students with learning disabilities? Summarizing the results from a meta-analysis of intervention studies. In Gersten, R., Schiller, E., and Vaughn, S. (eds.) *Contemporary Special Education Research: Synthesis of the Knowledge Based on Critical Instructional Issues*, pp 1–30. Mahwah, NJ: Erlbaum.
- Thousand, J., Villa, R., and Nevin, A. (2007). Collaborative teaching: Critique of the scientific evidence. In Florian, L. (ed.) *Handbook of Special Education Research*, pp 417–428. London: Sage.
- Trent, S. (1998). False starts and other dilemmas of a secondary general education collaborative teacher: A case study. *Journal of Learning Disabilities* 31, 503–513.

Relevant Websites

- <http://www.circleofinclusion.org> – Circle of Inclusion.
- <http://www.cec.sped.org> – Council for Exceptional Children.
- <http://www.ccsso.org> – Council of Chief State School Officers.
- <http://www.ets.org> – Educational Testing Service.
- <http://www.powerof2.org> – Power of 2.
- <http://hawbaker.pls.iowapages.org> – Research: Coteaching and Coplanning, Becky Wilson Hawbaker: Teaching, Research, and Outreach at Price Lab School, UNI.
- <http://www.geocities.com> – Special Educator's: Co-Teaching.
- <http://712educators.about.com> – The Co-Teach Model, About.com.
- <http://iris.peabody.vanderbilt.edu> – The IRIS Center.

Cultural and Linguistic Diversity and Children with Special Needs

J M Patton, The College of William and Mary, Williamsburg, VA, USA

N Day-Vines, Virginia Tech, Falls Church, VA, USA

A McSwain, Norfolk State University, Williamsburg, VA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Awareness competency – Constant examination of one's own attitudes, biases, and assumptions that may interfere with the ability to work effectively with culturally and linguistically diverse exceptional learners.

Cultural competency – A set of congruent attitudes, practices, policies, and structures that are integrated to enable professionals to work more effectively with culturally, class, and linguistically diverse exceptional learners.

Cultural reciprocity – Exchange of knowledge, values, and perspectives between two or more individuals of different backgrounds.

Culturally responsive – Using the cultural knowledge, prior experiences, and performance styles of diverse students to make learning more appropriate and effective for them.

Disproportionality – The over- or under-representation of a given population group, often defined by racial and ethnic backgrounds, but also defined by socioeconomic status, national origin, English proficiency, gender, and sexual orientation, in a specific population category.

Knowledge competency – The acquisition of factual information about people from culturally distinct groups.

NCLB – The No Child Left Behind Act of 2001 is a national law that aims to improve the standards of primary and secondary schools by increasing the accountability for schools, and the flexibility in parent's choice as to which schools their children will attend.

Skill competency – Synthesis of awareness and knowledge competencies with the aim of developing culturally relevant strategies that promote teaching effectiveness and enhanced outcomes.

Worldview – This is a consistent (to a varying degree) and integral sense of existence and provides a framework for generating, sustaining, and applying knowledge.

A substantial body of evidence documents the overrepresentation in special education of students from certain minority groups. This iniquitous arrangement was documented as early as the 1920s by a small group of researchers and scholars of that era. It has been recently reconfirmed by the 2002 National Research Council Report (Donovan and Cross). It is important to note that this overrepresentation phenomenon exists primarily in classes for students with emotional/behavioral problems and those with mild mental disabilities, both disability categories that rely on subjective judgments and which are considered soft disability categories. Although this problem has been a historical one for many Latinos, Native American students, and certain Asian American groups, today African Americans, especially males, continue to be the most disproportionately represented group in the disability categories previously cited. This phenomenon is manifested in the early grades and is particularly virulent from the third grade through middle schools. Further, the growing achievement and behavioral gaps for many ethnic minority learners, those from low socio-economic environments and English language learners, provide a compelling reason for restructuring and transforming the traditional curricula and pedagogy used to educate and prepare teachers, administrators, and related service professionals in ways that respond to the cultural, class, and language needs of these learners. As such, this article provides an overview, rationale, essential conceptual designs, and necessary elements of an evidence-based cultural competency training model for special educators.

Education in general and special education in particular cannot be separated from its relationship to the cultural, class, language, and cross-cultural dynamics that shape the context of society in general and public education in particular. Today, all Americans live and work in a society that is becoming increasingly multicultural and pluralistic. Given these explosive demographic changes among students and their families and the increasing cultural, ethnic, and class homogeneity of today's teaching, and the administrative, and related professional workforce, the potential for the creation of significant cultural, class, and language disconnections among students, their families, their teachers, and curriculum and instruction are inevitable. In fact, while African Americans and, in certain environments, Hispanics, Native and Americans

and some Asian American groups (e.g., Vietnamese and Hmong) are generally overrepresented in certain special education classrooms, they are, at the same time, severely underrepresented in the special education teaching, administration, and related service professionals. Given the fact that education and schooling cut across cultural, class, and linguistic borders, these differences represent critical challenges in the current No Child Left Behind Act (NCLB) school reform movement. As a consequence, those who educate special education teachers, administrators, and related service professionals are called upon to prepare these individuals to learn how to accommodate and respond to the emerging demands of an increasingly culturally, linguistically, and socio-economically diverse, exceptional student population. One has to wonder, however, where the source of the training to deal with this avalanche of cultural diversity will come from, given the fact that only a handful of these preparation programs provide training in cultural responsiveness, much less cultural competence. Additionally, the changing teacher and student demographics and the disproportionate representation of African American, Latino/Hispanic, and Native American populations in classrooms for students with mild mental, emotional, or behavioral disabilities have created several ethical challenges. The extent to which America achieves the ethical, educational, and cultural conditions for effective education for students with special needs depends on the degree to which these preparation programs educate practitioners to be culturally competent and effective in diverse school contexts. Further, it is readily known that in order for educators to be effective, cultural competency awareness, knowledge, skills, and dispositions normally cannot be externally imposed upon them. Rather, these realities must be viewed as being of worth and important to prospective special education teachers so that they are able to own, learn, and experience the training necessary to make them culturally competent. Cultural competence, then, can become an organizing construct that can be used to guide the development of culturally competent and responsive teacher education training curricula that can respond to the challenge of cultural and linguistic diversity of children with special needs.

Understanding Cultural Competence

Valuing cultural and linguistic diversity involves being culturally competent since we can not divorce somebody's language from his/her culture. Social scientists have, of late, established a cultural competency knowledge base that creates a paradigm shift in the way special educators conceptualize culture, class, and language differences. New ideas generally build upon the traditional cultural diversity knowledge and awareness knowledge bases, but

distinguish themselves by creating a new model that embraces notions of deep, structured culture and competency that can be routinized by educators in natural settings. In this regard, various cultural competency models that represent important progress have been developed. According to Hanley (1999), these cultural competency models have emerged as significant ones in the fields of social work, mental health service provision, and education. This new form of preparation of special education personnel is designed to transform the preparation of special education teachers, administrators, and related service professionals – many of whom are frequently unaware of or have, wittingly or unwittingly, not examined their own cultural or ethnic identities. While the concept of cultural competency has been variously defined, Hanley defined cultural competence as a set of congruent attitudes, practices, policies, and structures that are integrated in a system or agency to enable professionals to work more effectively with culturally and linguistically diverse learners. In order to reduce disproportionality and achievement and discipline gaps, structural changes in special education must occur. To value culture and diversity, Pedersen's model of cultural competence seems to be useful. This work is described in the following subsection.

Pedersen's Model of Cultural Competence

Pedersen (1994) outlined a relevant model of cultural competence that can serve as an underpinning of the design of a training curriculum for all educators. This model includes three particular and interrelated domains of cultural proficiency: (1) awareness, (2) knowledge, and (3) skills. As such, each domain must be mastered prior to proceeding to the next. Mastery in each of these domains provides the foundation and bases for a culturally competent training curriculum for special educators, as well as general educators, administrators, and related service professionals.

Awareness competencies. Awareness competencies include the recognition and constant examination of one's own attitudes, biases, and assumptions that may interfere with the ability to work effectively with culturally and linguistically diverse students with special needs. Awareness competencies require that special education teachers, general education teachers, administrators, and related service providers (1) critically and continuously explore their own attitudes, perceptions, assumptions, and any possible biases toward this population; (2) examine how these realities may impact the learners they serve; and (3) more importantly determine those strategies and practices needed to effectively work with culturally and linguistically diverse students with special needs.

A second feature of the awareness domain of cultural competency includes training that provides prospective

educators with the knowledge to recognize certain sociopolitical issues that often impede educational outcomes for children from disenfranchised groups. For instance, oppressive forces or systems of domination and control such as power, privilege, racism, sexism, and linguicism can result in diminished prospects for culturally and linguistically diverse students with special needs. Frequently, these youngsters experience single or multiple forms of oppression. For instance, an African American male receiving special education services may experience racism, as a result of his defined race; sexism, as a result of his gender status; linguicism, if he fails to speak Standard American English; and an internalized sense of shame and embarrassment because of the stigma associated with a diagnosed mental or emotional/behavioral disability. All educators who are striving to be culturally competent must recognize that these structural forces in society impact children from marginalized groups. Schools operate as a microcosm of the larger society, such that culturally different students and students with disabilities experience the pernicious effects of oppression that frequently contribute to their orientation toward education and school outcomes. Social mirroring, or the tendency of youngsters to internalize the widely negative perceptions that others maintain about their performance and abilities, can have adverse consequences on educational outcomes for these learners. Recognition of one's own biases, and the sociopolitical contexts that impact the educational experiences of minority group members, are important correlates of cultural competence. All of these cultural awareness elements can serve as the first theme to include in a culturally competent training curriculum that should help special educators reduce disproportionality and the achievement and behavior gaps.

Knowledge competencies. Knowledge competencies involve the acquisition of factual information about people from culturally distinct groups. Moreover, children's learning styles often correspond to their cultural preferences. Consideration of the demographic, historical, social, cultural, and educational experiences of culturally distinct students with special needs should allow educators to integrate salient cultural features into the educational process, thus moving toward cultural competence. For instance, among many African American and Latino youngsters, relationships are highly prized. Failure to develop warm, nurturing interpersonal connections with youngsters can have a detrimental impact on educational outcomes. Ladson-Billings (1994) maintained that the cultivation of a hospitable classroom climate remains indispensable when learning is the anticipated outcome. She noted that successful teachers of African American youngsters respect the personhood of students by demonstrating simple kindness through praise, listening, smiling, and showing respect for other students. Healthy classroom climates rest upon a child's ability to feel a sense of connectedness, support, and security. Classroom

climates that lack this culturally responsive dynamic seriously undermine a child's ability to profit from the learning enterprise.

Cultural misunderstanding and misattributions between culturally and linguistically diverse students with special needs and educators contribute greatly to the preponderance of discipline referrals, suspensions, and expulsions. Students with high levels of ethnic affiliation exhibit a distinctive set of communication styles that often do not conform readily to the norms and expectations required in schools. Many learners may communicate with one another and with school personnel in a manner characterized as loud, intense, and confrontational even without having accompanying feelings of anger. Educators may regard this conduct as volatile and assume that hostility is impending. In marked contrast, many Caucasians rely on more dispassionate, impersonal, and emotionally restrained communication styles. Nothing is inherently wrong with either form of communication, but when interpreted outside a particular cultural context, certain interpersonal styles may be regarded as rude and inappropriate and lead to an increase in disciplinary referrals for some learners. For instance, educators who have knowledge of the rhythmic orientation of many, though certainly not all, African American youngsters can introduce more culturally responsive kinesthetic approaches to learning into the repertoire of educators' behaviors. Gay (2000) underscored the importance of addressing multiple sensory stimulation among students who respond better to strategies that rely on visual stimuli, tactile, and kinesthetic learning, as well as personal reinforcement. Teachers and service providers who employ engaging, energetic, and performance-oriented teaching approaches are more likely to sustain the attention and interest of African American learners. Such instructional pedagogy correlates with the oratorical and communication styles that govern social interactions within two prominent African American institutions, the church and the community. Adopting culturally compatible instructional strategies holds considerable promise for fostering academic achievement. Blatant disregard of certain cultural features on the basis that students should perform academically despite the classroom climate or the particular cultural frame of reference to which educators adhere may damage any potential prospects for learning and further deepen the academic and behavior gaps between these learners and those from the majority population.

Skill competencies. The third component of Pedersen's conceptual framework involves skills competencies. This domain of competencies builds upon the previous two and should serve as a basis for the design of the third theme for a culturally competent training curriculum. Essentially, skills competencies refer to the synthesis of awareness and knowledge competencies with the ultimate aim of developing culturally relevant strategies that promote teaching effectiveness and enhanced educational

outcomes for all students, especially those with special needs. It is common knowledge that deficit thinking, or the presumption that children from marginalized groups are uneducable, impedes educational progress. Such a bias often guides the decision making of school personnel. Because the culture of the school and the culture of children from diverse groups often differ so markedly, culturally competent educators, on the other hand, develop a skill set to recognize that cultural differences do not automatically translate into deficits.

Culturally competent educators possess the skill set that enables them to maintain high expectations for students from marginalized groups, especially those with special needs. Steele's (1997) research indicated that when a negative stereotype exists about a particular group (e.g., minorities and women), members of that group internalize the stereotype in a manner that can interfere with academic performance. For this reason, the content expertise of educators must be balanced with a skill set that includes awareness and understanding that implicit and explicit attitudes about students can influence educational outcomes. Educators can no longer afford to consign minorities and low socio-economic youngsters to the educational fringe. This stake becomes even higher for students with special needs. In addition, educators must validate and affirm their students' rich cultural backgrounds and use students' funds of knowledge to facilitate new learning. Banks (2001) argued quite persuasively that the incorporation of diverse perspectives should not be relegated to the coordination of multicultural feasts and the presentation of culturally diverse heroes. He also explained that multicultural education at the ultimate level should aim toward heightening meaningful social consciousness, awareness, and social change.

The recognition of the cultural knowledge and learning styles of students is important and must be included in the skill set needed by culturally competent educators. Much has been written about cultural styles and preferences of culturally diverse youngsters and that their styles often differ markedly from instructional orientations from which learning is routinely conducted. Educators and service providers must work diligently to incorporate cultural values and styles that resonate with students from diverse groups with the ultimate aim of helping them operate effectively and learn code switching skills.

Clearly, culturally competent educators self-assess and self-monitor to guarantee that they are not behaving in biased ways. This skill requires that they possess a certain level of social, political, and historical knowledge and consciousness about children from marginalized groups. Moreover, this form of consciousness is a necessary, albeit insufficient, condition for cultural competence. Accordingly, educators must act on this conscious raising learning and serve as agents of social, political, and economic change who advocate for their students and their

families. As a result, they must work in integrated ways with families and communities. Such alliances may include partnerships that schools establish with religious, civic, and social organizations, as well as private industry in order to promote mutual and reciprocal relationships. Such relationships must be aimed at improving educational outcomes for culturally and linguistically diverse youngsters, especially those with special needs. Schools can no longer afford to work as isolated entities that ignore valuable community resources and partnerships. These partnerships are more apt to be effective when reciprocity exists across all partner stakeholder groups.

Community Connections and Students with Special Needs

Schools connect with families when there is cultural reciprocity. Cultural reciprocity functions as an important dynamic involved in adapting to cultural differences. For instance, Kalyanpur and Harry (1999) identified cultural reciprocity as an essential component for developing skill competencies. Cultural reciprocity involves a willingness to consider the assumptions that culturally diverse individuals and families have that could impact the educational process for children. Kalyanpur and Harry (1999) noted four critical steps for developing cultural reciprocity: (1) identification of cultural values inherent in the educational practice, (2) consideration of family perspectives regarding educational issues that may differ from educators' conceptions, (3) demonstration of respect toward varying values and viewpoints, and (4) development of effective strategies that integrate professional interpretations and the family's value system. This model promotes educators' recognition and incorporation of cultural adaptations and family values and viewpoints into the educational process. More specifically, cultural reciprocity involves a willingness to consider the assumptions that culturally diverse individuals and families hold that could impact the educational process of their children. In other words, cultural reciprocity challenges the often-heard stereotype of the educator as expert.

Historically, the church has served a significant role in the socialization of children, especially African American and Latino youngsters. For instance, the church provided educational support, religious guidance, mentoring, financial assistance, and political activism as a mechanism for enhancing the life chances of members in African American and Latino communities. Civic organizations have often served as self-help groups designed to enhance quality-of-life issues for many culturally diverse members and their communities. Service represents the central organizing principle around many organizations such as fraternities and sororities that often sponsor enrichment programs for youngsters. Given the prominence of the church in many minority communities and the commitment to service of

many civic organizations, school representatives can work closely with religious and civic institutions to promote mutual and reciprocal relationships aimed at improving educational outcomes for minority youngsters with special needs as well as enhancing partnerships between the school and community.

Conclusion

This article focuses on cultural and linguistic diversity and students with special needs. We believe educators and service providers must respond to the cultural and linguistic diversity that students bring to school. To do this, they must be culturally competent. It is essential for all educators and service providers to constantly examine and mediate their cultural and class knapsacks as a precondition to acquiring cultural competence. While this process is necessary and essential, it is often difficult as we are trained to look for specks of gravel in other people's eyes while we may have boulders in our own. This self-exploratory process should result in an increased awareness of one's own cultural and class identities, thus enabling one to develop a deeper awareness and understanding of one's non-self, especially those who are most unlike you. With this increased and deeper awareness of the contents of one's cultural and class knapsacks (i.e., one's stereotypes about self and others, assumptions about others, and the like), essential positive changes of behavior should occur. This transformation provides the basis from which deep levels of cultural awareness, knowledge, and skills can be built to educate culturally and linguistically diverse students with special needs. Clearly, no transformation will occur unless educators and service providers are prepared and trained to be culturally competent. The cultural competency awareness, knowledge, and skill dimensions outlined in this article can and should be used in the education and training of persons involved in the education of students, especially those with special needs. In the end, cultural competency will materialize when effective partnerships with key stakeholder groups are reinforced and embedded in all special education programming.

Bibliography

- Banks, J. (2001). Approaches to multicultural curriculum reform. In Banks, J. and Banks, C. (eds.) *Multicultural Education: Issues and Perspectives*, pp 225–246. Boston: Allyn and Bacon.
- Donovan, M. S. and Cross, C. T. (eds.) (2002). *Minority Students in Special and Gifted Education*. Washington, DC: National Academy Press.
- Gay, G. (2000). *Culturally Responsive Teaching: Theory, Research, and Practice*. New York: Teachers College Press.

- Hanley, J. (1999). Beyond the tip of the iceberg: Five stages toward cultural competence. *Reaching Today's Youth* 3(2), 9–12.
- Kalyanpur, M. and Harry, B. (1999). *Culture in Special Education: Building Reciprocal Family–Professional Relationships*. Baltimore, MD: Paul H. Brookes.
- Ladson-Billings, G. (1994). *The Dreamkeepers: Successful Teachers of African American Children*. San Francisco, CA: Jossey-Bass.
- Pedersen, P. (1994). *A Handbook for Developing Multicultural Awareness*, 2nd edn. Alexandria, VA: American Counseling Association.
- Steele, C. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist* 52(6), 613–629.

Further Reading

- Cross, T. L., Bazron, B. J., Dennis, K. W., and Issacs, M. R. (1998). *Toward a Culturally Competent System of Care* vol. I. Washington, DC: Georgetown University, Child Development Center.
- Day-Vines, N. (2000). Ethics, power, and privilege: Salient issues in the development of multicultural competencies for teachers serving African American children with disabilities. *Teacher Education and Special Education* 23, 3–18.
- Day-Vines, N. L. and Patton, J. M. (2003). The perils, pitfalls, and promises of the No Child Left Behind Act of 2001: Implications for the education of African American and other minority learners. *T/TAC Link Lines* 1–5.
- Day-Vines, N. L., Patton, J. M., and Baytops, J. L. (2003). Counseling African American adolescents: The impact of race, culture, and middle class status. *Professional School Counseling* 7, 40–51.
- Isaacs, M. R. and Benjamin, M. P. (1991). *Toward a Culturally Competent System of Care* vol. II. Washington, DC: Georgetown University, Child Development Center.
- Kallyanpur, M., Harry, B., Kearney, K., and LeBlanc, J. (1993). Forgotten pioneers in the study of gifted African Americans. *Roeper Review* 15, 192–199.
- Mason, J. L., Cross, T., Rider, M., and Friesen, B. (1998). Developing cultural competence for agencies. *Focal Point* 2(4), 5–7.
- McPhatter, A. R. (1997). Cultural competence in child welfare: What is it? How do we achieve it? What happens without it. *Child Welfare* 74(1), 255–279.
- McSwain, A. (2002). The effects of multicultural and bilingual training on preservice students' self-reported level of competency. *Multiple Voices* 5, 54–65.
- McSwain, A. and Patton, J. (2006). Early childhood special education and multicultural education. In Obiakor, F. E. (ed.) *Multicultural Special Education*, pp 196–205. Upper Saddle River, NJ: Merrill Prentice Hall.
- Meyer, G. and Patton, J. M. (2001). *On the Nexus of Race, Disability, and Overrepresentation: What Do We Know? Where Do We Go?* Newton, MA: National Institute for Urban School Improvement, Education Development Center.
- Miller, K. J. and Fuller, D. P. (2006). Developing cultural competency in early childhood preservice education through a cultural self-analysis project. *Journal of Early Childhood Teacher Education* 27, 1–11.
- Neal, L. I., McCray, A. D., and Webb-Johnson, G. (2001). Teachers' reactions to African American students' movement styles. *Intervention in School and Clinic* 36(3), 168–174.
- Nieto, S. (2000). *Affirming Diversity: The Sociopolitical Context of Multicultural Education*, 3rd edn. New York: Addison-Wesley.
- Patton, J. M. (1998). The disproportionate representation of African Americans in special education: Looking behind the curtain for understanding and solutions. *Journal of Special Education* 32, 25–31.
- Patton, J. M., Williams, B. T., Floyd, L. O., and Cobb, T. R. (2003). Recruiting and retaining culturally and linguistically diverse teachers in special education: Models for successful personnel preparation. *Teacher Education and Special Education* 26, 288–303.
- Rice, N. (2006). Critical pedagogy and African American males labeled "emotionally or behaviorally disordered" *Multiple Voices* 9, 64–76.

Trent, S. C. and Dixon, D. J. (2004). "My eyes were opened". Tracing the conceptual change of preservice teachers in special education/multicultural education class. *Teacher Education and Special Education* 27, 119–133.

Relevant Websites

<http://www.nativescience.org> – Alaska Native Science Commission.
<http://www.alliance2k.org/> – Alliance Project.
<http://www.colorado.edu> – BUENO Center for Multicultural Education at University of Colorado, Boulder.
<http://www.inclusivechildcare.org> – Center for Inclusive Childcare.
<http://www.co-operation.org> – Cooperative Learning, Values, and Culturally Plural Classrooms.
<http://cecp.air.org> – Cultural Competence Resources Homepage.
<http://www.intime.uni.edu> – Culturally Responsive Teaching.
<http://www.dredf.org> – Disability Rights Education and Defense Fund, Inc.
<http://www.decet.org> – Diversity in Early Childhood Education and Training.

<http://www.diversityuintl.com> – Diversity Training University International (DTUI).
<http://www.educationworld.com> – Education World.
<http://www.global-issues.org> – Global Issues-Links.
<http://www.iamme.org> – Illinois Association for Multilingual Multicultural Education (IAMME).
<http://implicit.harvard.edu> – Implicit Association Test.
<http://www.sietar.org> – International Society for Intercultural Education, Training and Research.
<http://www.edchange.org> – Multicultural Pavilion.
<http://www.nameorg.org> – National Association of Multicultural Education (NAME).
<http://www.ndpc-sd.org> – National Dropout Prevention Center for Students with Disabilities.
<http://www.nmci.org> – National Multicultural Institute.
<http://www.attc-ne.org> – New England Addiction Technology Transfer Center Network.
<http://www.ncpublicschools.org/racg/services> – Public Schools of North Carolina – Raising Achievement and Closing Gaps.
<http://www.tolerance.org> – Tolerance (Project of the Southern Poverty Law Center).

Curriculum-Based Assessment and Students with Special Needs

D J O'Shea and J D Katsafanas, O'Shea Learning Consultants, Media, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Appropriateness of assessment – Use of a variety of assessment tools and strategies to determine programming by relying on sound instruments that are non-discriminatory, valid, reliable, and administered by trained teachers.

Assessment – To measure and appraise students' instructional progress based on direct observation and data analysis.

Criterion-referenced measurement – Measuring an individual student's performance in terms of absolute mastery of particular skills.

Curriculum-based assessment – An assessment system used by teachers to identify, examine, analyze, and reflect on student progress and systematic instructional data in the local curriculum.

Curriculum-based measurement – A scientifically validated form of student progress monitoring in which teachers use standard methods for test development, administration, scoring, and data utilization with students involved in the curriculum, as teachers evaluate the effectiveness of instructional changes.

Curriculum measurement – Assessment involving direct or performance-based assessment linked directly to instruction.

Precision-teaching – A system for charting behavior frequency (or rate) against calendar days.

Problem solving process – Identifying a problem to be solved and alternative answers to the problem by implementing and testing alternative answers, revising unsuccessful answers, and terminating the problem-solving process.

Progress monitoring – Part of effective teachers' instructional decision-making process that uses data analysis to determine how students are learning based on predetermined instructional goals and objectives.

Assessment, a basic tenet of effective teaching, is fundamental to the instructional decision making of astute teachers. These teachers use multiple types of assessment data for a variety of educational decision making. This use of data, collected from assessment, tends to highlight important educational decisions collected about teaching-learning processes, on student or teachers' demographics,

and of relevant variables important to learning milieus in which instruction occurs. Accurate data help teachers to (1) become familiar with students' learning strengths, needs, and preferences; (2) design teaching-learning experiences that support students' growth and development; (3) adjust instruction in response to students' ongoing achievement; and (4) monitor students' progressions carefully.

In the United States, curriculum-based assessment (CBA) is known as a valuable assessment system that teachers employ to identify, examine, analyze, and reflect on student progress and systematic instructional data. CBA has a significant reputation, holding historical relevance to address student achievement, and teachers' careful consideration of instructional design and teaching-learning effectiveness. CBA retains continuing importance to students with special needs and effective instructional decision making.

CBA: Definition and Importance

CBA entails measurement that uses "direct observation and recording of a student's performance in the local curriculum as a basis for gathering information to make instructional decisions" (Deno, 1987: 41). CBA, with a history critical to, and highlighting, the special education movement in the United States, evolved during the last three decades into a system of ongoing measurement of student progress, and sound decision making about instructional practices. Today, it represents the collection of a solid research base that can be linked directly to teachers' instructional decisions. The application of CBA principles helps to ensure that individual students, in and outside of a classroom environment, achieve teachers' instructional objectives, regardless of the learning setting in which their instruction occurs. Clearly, teachers who implement CBA processes and procedures appropriately are able to design effective instruction. As they implement and manage instruction, they are able to gather assessment data, using frequent assessment of their students' progress across curricular areas in a systematic manner on critical curriculum objectives. These teachers, then, are able to reflect on their obtained data to make sound educational decisions about their own instructional effectiveness and classroom success.

Teachers employ ongoing assessment data to cover a range of skills in varied, academic core subject areas

(e.g., concept recognition, number recall, and rule differentiation in mathematics; letter naming, word recognition, and fluency in reading; and capitalization, punctuation, and subject–verb agreement in written expression). They view data over time (e.g., pre, during, or after instruction), and in a variety of settings (e.g., classroom, community, home, and work site). As teachers reflect on the usefulness of assessment linked to instruction, and as ongoing assessment continues in teachers' regular repertoire of reporting student academic progress over time, they, then, begin to involve students more directly and explicitly during teaching–learning processes. The use of reflective data, collected through instructionally relevant measurement by teachers and students, helps to inform whether or not students are making instructional progress. This helps to confirm whether or not students are demonstrating behaviors toward reaching levels of academic mastery. Such data use, also, helps to inform and confirm whether or not teachers are successful in educational decision making. Regardless of what skills teachers assess, the curricular standards that teachers employ, when they collect data, where they assess, how they measure, and/or the type of data teachers seek, instruction and learning analyzed by teachers lead to meaningful knowledge. Frequent analyses help teachers to (1) know critical skills for assessment that require teacher attention; (2) ensure valid and reliable data collection, (3) examine instructional data on an ongoing basis, (4) monitor student performance critical to student learning situations, (5) make reflective decisions about their own teacher effectiveness, and (6) improve students' academic achievement and self-instructional awareness.

CBA: Historical Context

Deno researched the use of assessment linked to instruction in the late 1970s with the goal of giving teachers of students with learning disabilities a simple set of evaluation procedures that would allow them to collect data on and graph an individual student's academic progress. Such measurement procedures sought direct observation and recording of a student's performance in a local curriculum as a basis for gathering information to make instructional decisions. In 1985, the Pine County Special Education Cooperative in Minnesota began field testing what was then referred to as curriculum measurement, a form of alternative assessment associated with special education involving direct or performance-based assessment typically using materials taken directly from instructional materials used with nondisabled students (Deno, 1985; Deno and Fuchs, 1987). Reading, mathematics, and written expression covered common applications. This research and testing showed assessment linked to instruction to be among the most reliable tools for accurate measurement

and evaluation of students' academic achievement, both to compare students to one another and chart individual student progress.

CBA expanded to refer to models of assessment that emphasized a direct relationship to an individual student, or group of students, within the curriculum. Curriculum-based measurement (CBM), a refinement of CBA, highlighted work by Deno and his colleagues by using repeated measures from students involved in the curriculum, to evaluate the effectiveness of teachers' instructional practices (Deno and Fuchs, 1987; Fuchs and Deno, 1991; Fuchs and Fuchs, 1984, 1985; Salvia and Hughes, 1990). Researchers collecting and analyzing data on atypical student learning reasoned that instructional changes by teachers may lead to more effective teaching methods and improved student achievement. CBM researchers, then, sought specific data on how achievement gains and academic progress change on generic tasks of constant difficulty (King-Sears, 1994; Salvia and Hughes, 1990; Salvia and Ysseldyke, 1991). Deno, his colleagues, and his team of graduate students extended their research by looking for and refining tasks that could serve as simple measures of student improvement. Academic tasks, such as reading words aloud or recalling rote multiplication facts, received research attention. These researchers began extensive testing on students with and without disabilities to make sure that their research provided consistent and comparable results with other forms of measurement over time (Deno and Fuchs, 1987; Fuchs and Deno, 1991; King-Sears, 1994). Extensions of assessment linked to instruction relied on principles of direct instruction (Becker and Carnine, 1980, 1981; Becker and Gersten, 1982) as refinements of observable and measurable student progress and teacher behaviors.

Researchers, also, began studying whether or not an increase in the behavior being measured on equivalent forms of the task would represent academic growth. Deno and Fuchs (1987) and Fuchs and Deno (1991) reported that CBM could be used in conjunction with a problem-solving process that included:

- identifying a problem to be solved (e.g., underachievement in reading or mathematics);
- identifying alternative answers to the problem (e.g., a new instructional method in reading or mathematics);
- implementing (e.g., teaching a new or different instructional strategy) and testing the alternative answers (e.g., assessing or evaluating the reading methods or mathematics methods);
- revising unsuccessful answers (e.g., continuing to make changes in the reading method or mathematics method); and
- terminating the problem-solving process (e.g., seeking to discontinue the special education services).

Other Systems of Curriculum Measurements

Application of assessment principles linked to curricular design became fundamental to progress reporting and academic monitoring, enjoying support from researchers and practitioners across the United States, including those in the Department of Education. Instructionally relevant measurement models, especially CBA and precision teaching, became the measurement and assessment tools of choice in numerous federally funded studies (Carpenter and King-Sears, 1998; Fuchs and Deno, 1991; Gersten *et al.*, 1995). Precision teaching, a system for charting behavior frequency (or rate) against calendar days has been used by measurement experts as a means to support practitioners, teachers, and students. Curriculum assessment linked directly to instruction spread into use around the United States and abroad as researchers began applying principles of instructionally relevant measurement models to other subject areas and age groups. Researchers began focusing on students of all ability levels with derivatives of CBA and CBM, such as in computer-based applications, and as measures of states' overall accountability programs (Blankenship, 1985; Carpenter and King-Sears, 1998, 2002; Cullen and Pratt, 1992; Fuchs and Deno, 1991; King-Sears, 1994).

The selections of critical skills for assessment and the use of CBA received researchers' attention (e.g., Blankenship, 1985; Salvia and Hughes, 1990). Researchers refined aspects of CBA, such as in the APPLY mnemonic, that synthesize steps for teachers' development and use of CBA in classrooms (King-Sears, 1994; King-Sears *et al.*, 1999) to analyze the curriculum, prepare items to meet curriculum objectives, probe frequently, load data using a graph format, and yield to results based on revisions and decisions (see King-Sears, 1994). Uses of various forms of instructionally relevant measurement, such as through implementation of the APPLY mnemonic, offered easy and reliable tools for measuring student progress in ways that allow for early intervention and assessment of teaching-learning effectiveness.

What CBA Means for Access to the General Curriculum

CBA can customize assessment for students in the general education curriculum. As a form of criterion-referenced measurement, CBA helps to pinpoint curricular objectives that act as the criteria for the identification of instructional targets and for the assessment of students' skills upon entry into an academic curriculum (Fuchs and Fuchs, 1984, 1985; King-Sears, 1994). Salvia and Ysseldyke (1991) identified criterion-referenced measurement as a system that measures a student's performance in terms of

absolute mastery of particular skills. Teachers use criterion-referenced tests to help when planning instructional programs and in monitoring students' mastery of skills. CBA acts as a guide for the development of individual goals, assessments, interventions, and accommodations. In addition, it allows for continual monitoring of students' developmental progress since it links assessment results to instructional interventions (Elliott, 1992).

Many assessment systems linked to instructional models, including CBA, CBM, and precision teaching, had their beginnings in the special-education discipline or in the educational-psychology domain. As it appears, alternative measurement tools provide an alternative to standardized norm-referenced measures. In many cases, the use of such standardized tools provides information more accurately to teachers at the eligibility and planning stages in the special education referral process. However, alternative measures, especially CBA and precision teaching, are more valuable tools for monitoring the progress of students in the curriculum of instruction, most often the general education curriculum (King-Sears, 1994; Salvia and Hughes, 1990). When using progress-monitoring devices in systematic ways, teachers evaluate student performance formatively in an academic skill area, specific to the curriculum used. Formative evaluation allows them to evaluate the adequacy of skills development. If progress is deemed inadequate, teachers then select and implement appropriate interventions. Conversely, if students perform beyond expectations, or criteria, teachers often have the information to make curriculum adjustments to challenge students at their appropriate level. For example, the fluency-assessment systems, the Dynamic Indicators of Basic Early Literacy Skills (DIBLES), developed by Good and peers, originated from monitoring students' accuracy and speed in curriculum materials typically used in general education literacy programs and has become widespread in current classrooms in the United States (Good and Kaminski, 2002; Good *et al.*, 2002).

In order to begin the process of designing instruction that is linked to assessment, teachers require an understanding of CBA's basic assumptions:

- observable behavior of critical student performances related to the curriculum being taught to underscore all data collection;
- data obtained represent frequent and brief measurements of student behavior prior to, during, and after instructional units;
- students and/or teachers display measurements graphically; and
- students and/or teachers use the graphic displays to make instructional decisions (Carpenter and King-Sears, 2002; Salvia and Hughes, 1990).

These assumptions view assessment as criterion referenced, curriculum embedded, ongoing, and programmatic.

As criterion referenced, teachers seeking to link instructional design and assessment measure students' performance in terms of absolute mastery of particular skills in a planned curriculum. Primary purposes of teachers' planning consider the link between instruction and documenting students' progress. These purposes entail: (1) identifying students' entry-level competencies, (2) monitoring student performance during instruction, (3) determining students' mastery of competencies at the conclusion of instruction, and (4) monitoring students' maintenance and generalization of competencies (Salvia and Hughes, 1990; Salvia and Ysseldyke, 1991). CBA assumptions are necessary prerequisites to ongoing processes of gathering information regarding students' strengths, interests, and emerging abilities related to important skills across content for the purpose of planning and monitoring instruction.

The construction of CBA occurs in the same way, regardless of the type of curricula that teachers employ when planning and implementing instruction (Idol *et al.*, 1996). First, the teacher (or CBA constructor) selects sample items from the curriculum or constructs sample items to match the curriculum. The teacher orders these items by difficulty, combining them within a single test. Consequently, the teacher provides the ordered items on the first day of student assessment (day 1). Then, the teacher constructs two or more forms of the same test, containing similar items and identical orders of difficulty. Next, the teacher administers these forms on the second and third days of student assessment (days 2 and 3). It is highly recommended that teachers construct CBA using three different forms on three separate occasions, to control for sporadic student response (Carpenter and King-Sears, 2002).

In administering the CBA, the teacher employing the CBA form tests the students across several levels of the curriculum. Ordinarily, the teacher measures their responses for speed or proficiency, as well as for accuracy. The teacher's use of continued and ongoing CBA assessment helps to record student responses over time. The teacher, then, focuses on performance criteria, established to determine acceptable levels of student performance or mastery. As it stands, normative sampling has been seen as a useful procedure for establishing mastery criteria in a general-education setting (Idol and Rutledge, 1993). This procedure involves taking samples of average and acceptable student performance in the general-education class as a basis for deciding what the absolute mastery criteria ought to be. Once this type of social validation occurs and the mastery criteria are established, the teacher administers the CBA, either to individual students or to groups of students.

Evaluating CBA

Researchers developed several characteristics to evaluate the effectiveness of CBA, common across models.

Embodied in these characteristics are the following strategic questions:

- Did teachers use measurement procedures to assess students directly using the material in which students received instruction? Were sampling items from the curriculum involved?
- Was teachers' administration of each measure generally brief in duration (typically 1–5 min)?
- Did the CBA design structure underscore frequent and repeated measurements as much as possible, as such measurements are sensitive to change?
- Did teachers and/or students display data graphically to allow monitoring of student performance?

Many teachers use CBA as a formative assessment to assist them during instruction. As they conduct formative assessment, many observe student results and are able to use the observational data to change their teaching. Graphically displaying those results, then, helps teachers in decision-making processes to increase their students' academic or behavioral achievement and to use the instructional time more productively. Various researchers compare procedures with those teachers not using any systematic data collection (Blankenship, 1985; Fuchs and Fuchs, 1984; Fuchs and Deno, 1991; King-Sears, 1994; Salvia and Hughes, 1990).

Motivating Students to Increase Self-Awareness of Learning and Progress

Instructionally relevant measurement helps to promote students' meaningful awareness of their instructional content and progress. Teachers can help students become part of their active learning by using direct and guided decisions regarding their progress. This can help students to become motivated about their own learning and achievement gains. Teachers and students, then, use the information gained from this awareness to (1) guide the teaching and learning process, (2) decide which instructional methodology is best suited to the instructional context, (3) examine and analyze instructional strategies and tactics, (4) consider and organize instructional materials, and (5) reflect on important instructional techniques used throughout instruction. Teachers support students as they guide students, fade their own presence, and help to establish students as independent and self-directed learners (See Blankenship, 1985; Deno, 1985; Fuchs and Fuchs, 1984; King-Sears, 1994, 1999; Salvia and Hughes, 1990).

By emphasizing CBA on a frequent basis, teachers can encourage student involvement in awareness and self-learning by teaching students how to set goals for learning and how to analyze data to self-evaluate actual performance. CBA has a long history as a driving force to help teachers develop meaningful ways throughout the

instructional process to assess students frequently and systematically on critical skills and to motivate them to become actively involved in the assessment process. Student motivation requires that teachers directly and explicitly instruct students on the use of observing, probing, charting, graphing, and analyzing data in order that students self-instruct with the teachers' direct or guided support. By using CBA techniques to inform and confirm students' knowledge of data-based decision making, teachers facilitate students' road to independence.

Steps for Using CBA

To use CBA effectively, general and special teachers should:

- analyze curriculum standards, identifying required academic skills and behaviors in a logical and sequential order;
- write performance objectives for each skill in measurable terms to enable data collection across time;
- determine mastery level for each skill;
- assess students' mastery of each skill before instruction;
- provide instruction on the identified skills;
- conduct frequent probes to assess students' mastery of each skill as it is taught;
- display the results in a simple graph form; and
- analyze performance trends to determine the effects of instruction on student learning.

Conclusion

Academic assessment has a long-standing tradition in education. Assessment historically involved measurement of student progress for the purpose of informing. One level of informing entails identification or eligibility decision making; a second level is that of informing instruction. Traditional assessment systems have limitations that restrict their application for instructional program planning (Blankenship, 1985; Fuchs *et al.*, 1990; Salvia and Hughes, 1990). Alternative assessment procedures appearing in educational literature in the past 20 years note the importance of CBA. Whereas traditional achievement tests tend to measure broad curriculum areas and/or skills, instructionally relevant measurement systems focus on specific skills currently being taught in the classroom, usually basic skills (Gersten *et al.*, 1995; Idol *et al.*, 1996). Interestingly, CBA has earned a reputation as an objective, empirically supported measurement system with characteristics of a technically sound and teacher-friendly measurement system. Throughout the last three decades, CBA has been found to be reliable, valid, time efficient, inexpensive, and easy to understand. It requires little training and uses multiple forms for repeated administration. Indicators of

student achievement help teachers and students to evaluate and graph baseline and intervention data. Astute teachers, then, use these data and predetermined data decision rules to decide when to make instructional changes. As they support students in learning and awareness, teachers help students to use the principles of data-driven assessment to inform and confirm students' academic achievement and progress over time. Such teachers become reflective as they link teaching and learning processes of what is assessed with what is taught.

See also: Peer-Tutoring.

Bibliography

- Becker, W. C. and Carnine, D. W. (1980). Direct instruction: An effective approach to educational intervention with the disadvantaged and low performers. In Lahey, B. B. and Kazdin, A. K. (eds.) *Advances in Clinical and Child Psychology*, pp 429–439. New York: Plenum.
- Becker, W. C. and Carnine, D. W. (1981). Direct instruction: A behavior theory model for comprehensive educational intervention with the disadvantaged. In Bijou, S. W. and Ruiz, P. R. (eds.) *Behavior Modification: Contributions to Education*, pp 44–65. Hillsdale, NJ: Erlbaum.
- Becker, W. C. and Gersten, R. (1982). A follow-up of follow through: Meta-analysis of the later effects of the direct instruction model. *American Educational Research Journal* **19**, 75–93.
- Blankenship, C. S. (1985). Using curriculum-based data to make instructional decisions. *Exceptional Children* **52**, 233–238.
- Carpenter, S. L. and King-Sears, M. E. (1998). Classroom assessment practices for instruction. In Rosenberg, M., O'Shea, L. J., and O'Shea, D. J. (eds.) *Student Teacher to Master Teacher*, 2nd edn., pp 89–122. New York: Prentice-Hall.
- Cullen, B. and Pratt, T. (1992). Measuring and reporting student progress. In Stainback, S. and Stainback, W. (eds.) *Curriculum Considerations in Inclusive Classrooms*, pp 175–196. Baltimore, MD: Paul H. Brookes.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children* **52**, 219–232.
- Deno, S. L. (1987). Curriculum-based measurement: An introduction. *Teaching Exceptional Children* **20**, 41.
- Deno, S. L. and Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children* **19**, 1–16.
- Elliott, S. N. (1992). Authentic assessment: An introduction to a neobehavioral approach to classroom assessment. *School Psychology Quarterly* **5**, 273–278.
- Fuchs, L. S., Fuchs, D., and Hamlett, C. L. (1990). Curriculum-based measurement: A standardized, long-term goal approach to monitoring student progress. *Academic Therapy* **25**, 615–632.
- Fuchs, L. S. and Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children* **58**, 488–500.
- Fuchs, L. S. and Fuchs, D. (1984). Criterion-referenced assessment without measurement: How accurate for special education? *Remedial and Special Education* **5**(4), 29–32.
- Fuchs, L. S. and Fuchs, D. (1985). Effectiveness of systematic formative evaluation: A meta-analysis. *Exceptional Children* **53**, 199–208.
- Gersten, R., Keating, T., and Irvin, L. K. (1995). The burden of proof: Validity as improvement of instructional practice. *Exceptional Children* **61**, 510–519.
- Good, R. H. and Kaminski, R. A. (2002). *DIBELS' Oral Reading Fluency Passages for First through Third Grades (Technical Report No. 10)*. Eugene, OR: University of Oregon.
- Good, R. H., Wallin, J. U., Simmons, D. C., Kame'enui, E. J., and Kaminski, R. A. (2002). *System-Wide Percentile Ranks for DIBELS*

- Benchmark Assessment (Technical Report No. 9)*. Eugene, OR: University of Oregon.
- Idol, L., Nevin, A., and Paolucci-Whitcomb, P. (1996). *Models of Curriculum-Based Assessment*, 2nd edn. Austin, TX: Pro-Ed.
- Idol, L. and Rutledge, M. (1993). Teaching phonics to poor readers: Direct instruction using sound sheets. *Teaching Exceptional Children* **25**(4), 58–61.
- King-Sears, M. E. (1994). *Curriculum-Based Assessment in Special Education*. San Diego, CA: Singular.
- King-Sears, M. E., Burgess, M., and Lawson, T. L. (1999). Applying curriculum-based assessment in inclusive settings. *Teaching Exceptional Children* **32**(1), 30–38.
- Salvia, J. and Hughes, C. (1990). *Curriculum-Based Assessment: Testing What Is Taught*. New York: MacMillan.
- Salvia, J. and Ysseldyke, J. E. (1991). *Assessment*, 5th edn. Boston, MA: Houghton Mifflin.
- Fuchs, L. S., Fuchs, D., Compton, D. L., et al. (2007). Mathematics screening and progress monitoring at first grade: Implications for responsiveness to intervention. *Exceptional Children* **73**, 311–330.
- Jones, E. D., Southern, W. T., and Brigham, F. J. (1998). Curriculum-based assessment: Testing what is taught and teaching what is tested. *Intervention in School and Clinic* **33**(4), 239–249.
- Lindsley, O. R. (1972). From skinner to precision teaching. In Jordan, J. B. and Robbins, L. S. (eds.) *Let's Try Doing Something Else Kind of Thing*, pp 1–12. Arlington, VA: Council on Exceptional Children.
- Madelaine, A. and Wheldall, K. (2004). Curriculum-based measurement of reading: Recent advances. *International Journal of Disability, Development and Education* **51**, 57–82.
- Notari, A. R. and Bricker, D. D. (1990). The utility of a curriculum-based assessment instrument in the development of individualized plans for infants and young children. *Journal of Early Intervention* **14**(2), 117–132.
- Roberts, M. L., Marshall, J., Nelson, J. R., and Albers, C. A. (2001). Curriculum-based assessment procedures embedded within functional behavioral assessments: Identifying escape-motivated behaviors in a general education classroom. *School Psychology Review* **30**(2), 264–277.
- Safer, N. and Fleischman, S. (2005). How student progress monitoring improves instruction. *Educational Leadership* **62**, 81–83.
- Stecker, P. M., Fuchs, L. S., and Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools* **42**, 795–819.
- West, R. P., Young, R., and Spooner, F. (1990). Precision teaching: An introduction. *Teaching Exceptional Children* **22**(3), 4–9.
- Witt, J. C., Elliot, S. N., Daly, E. J., Gresham, F. M., and Kramer, J. J. (1998). *Assessment of at Risk and Special Needs Children*, 2nd edn. Boston, MA: McGraw-Hill.

Further Reading

- Algozzine, B. and Ysseldyke, J. E. (2006). *Effective Instruction for Students with Special Needs*. Thousand Oaks, CA: Corwin Press.
- Anderson, J. A., Kutash, K., and Duchnowski, A. J. (2001). A comparison of the academic progress of students with EBD and students with LD. *Journal of Emotional and Behavioral Disorders* **9**(2), 106–115.
- Brown-Chidsey, R., Johnson, P., and Fernstrom, R. (2005). Comparison of grade-level controlled and literature-based maze CBM reading passages. *School Psychology Review* **34**, 387–394.
- Burns, M. K. (2002). Comprehensive system of assessment to intervention using curriculum-based assessments. *Intervention in School and Clinic* **38**(1), 8–13.
- Christ, T. J. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimating standard error of the slope to construct confidence intervals. *School Psychology Review* **35**, 128–133.
- Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review* **33**, 188–192.

Relevant Websites

- <http://www.studentprogress.org> – National Center on Student Progress Monitoring.
- <http://www.interventioncentral.org> – The CBM Warehouse.
- <http://flora.murray@vanderbilt.edu> – Vanderbilt University: Reading Tests.

Diagnosis and Classification

M Sapp, University of Wisconsin–Milwaukee, Milwaukee, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

BASIC-ID or IB-the assessment in acronym form of the following modalities of a student –

Behaviors, affect, sensation, imagery, cognition, interpersonal relationships, drugs/biology.

Multimodal Behavior Assessment (MMBA) – Is a method used within psychological assessment that can be adapted to students with developmental disabilities.

Educators use the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR, Fourth Edition, Text Revision) to diagnose students with special needs. The DSM-IV-TR allows one to make multiaxial diagnoses – one for each of five axes (DSM-IV-TR, 2000: 27–38). Axis I is used to diagnose the presenting problem for a student such as anxiety and depression. Axis II is used to diagnose personality disorders and mental retardation. Axis III is employed for general medical conditions, diagnosed by a medical practitioner, that are related to a student's performance. Axis IV is used for psychosocial and environmental concerns that may influence diagnoses on Axis I and Axis II. Negative life events that can produce interpersonal stress are diagnosed on Axis IV. Axis V is used to diagnose a student's overall level of functioning, and it is called the global assessment of functioning scale (GAF). This scale has values ranging from 0, inadequate information; to 10, persistent danger to self and others; 50, moderate symptoms; to 100, superior functioning in a variety of areas.

Multimodal Behavior Assessment (MMBA) is a method used within psychological assessment; it is a comprehensive intra-individually based assessment that can be adapted to students with developmental disabilities. Lazarus (1990) developed this assessment procedure. MMBA assesses seven modalities in acronym form called the BASIC-ID or IB, and this acronym assesses the following modalities: behaviors, affect, sensation, imagery, cognition, interpersonal relationships, and drugs/biology. It is common knowledge that diagnosis and assessment are never perfect, and they involve a collection of several sources of information on a student (Kendall and Hammen, 1998). For example, students will often fall within several diagnostic categories, and often they will have several comorbid factors that do not fit into clear categories. There are several factors to address when attempting to assess students with special needs (Sapp, 1999, 2002, 2006); however,

the two most important factors are reliability and validity, and these factors are not adequately addressed within the profession of special education.

Theories of Testing

The purpose of this article is to discuss the positive and negative roles of diagnosis and classification in special education. Before students with disabilities can be diagnosed and assessed, reliable and valid assessment measures must be employed. To illustrate, African American and Latino students are not adequately represented in most standardized assessments, and this is why great care must be used when attempting to diagnose and assess all students. In contrast to the general perception within education, reliability and validity are based on educational test score theories – not facts. Classical true-score theory, generalizability theory, and item response theory are the three dominant theories within educational assessment (Allen and Yen, 1979). Classical true-score theory is the dominant theory within education, and it states that a student's observed score (X) is an addition of the student's true score (T) and some error (E). The E is often misunderstood by educators, and it is not a mistake but a theoretical construct that takes into account the inability to measure concepts perfectly. Generalization theory extends classical true-score theory measurement by showing that it does not have to be restricted to the two-component linear model of a true score and error score – the true-score theory. Essentially, generalizability theory is concerned with the reliability of generalizing from a student's observed score on a test to his/her average measure that would occur under all possible conditions that are acceptable. Implicit in this assumption is that the student's measured attributes are in a steady state, and changes in the student's scores are not the results of maturation, learning, or development. Changes in the student's attributes are the result of multiple sources of error such as occasions, different test forms, different test administrators, and so on (Sapp, 2006). Item response theory is a form of item analysis, and unlike the group assessment model of the true-score theory, it is an individualistic model. For example, true-score theory is based on the number of students who pass a test item (item difficulty) and the extent to which test items discriminate among students (discrimination index), and all these factors are based on the norms of the standardization sample. In contrast, item

response theory makes no assumptions about the students involved and it does not rely on group norms; instead, a student's test score is the arithmetic multiplication of a student's ability and the student's item difficulty. As opposed to comparison to group norms, the greater a student's ability, the greater the probability that a client will get an item correct.

Test items are said to be valid when they measure what they purport to measure, and validity involves several forms such as face, content, criterion, and construct. Face validity is a nonstatistical aspect of validity that on face appearance items appear to measure what they purport to measure. Content validity is another nonstatistical aspect of validity, and it is the degree to which items assess relevant facets of a conceptual domain. For example, if an educator were interested in assessing word knowledge, he/she would want test items that sample the domain of word knowledge. Criterion validity is a statistical form of validity that has at least three forms such as predictive, concurrent, and retrospective, and criterion validity is also called empirical validity. Simply stated, criterion validity assesses the degree in which items from two tests correlate. Construct validity helps determine if a given educational measurement instrument actually measures the underlying conceptual construct that it was designed to measure. Construct validity must be constantly updated; and with minority students, often construct validity has not been assessed.

DSM-IV-TR Classification of Children with Special Needs

The DSM-IV-TR classifies children with special needs into the following categories:

1. *Mental Retardation (Cognitive Disabilities)*
 - 317 Mild Mental Retardation (Educable)
 - 318.0 Moderate Mental Retardation (Trainable)
 - 318.1 Severe Mental Retardation (Severe/Profound)
 - 318.2 Profound Mental Retardation (Severe/Profound)
 - 319 Mental Retardation, Severity Unspecified
2. *Learning Disorders*
 - 315.00 Reading Disorder
 - 315.1 Mathematics Disorder
 - 315.2 Disorder of Written Expression
3. *Motor Skills Disorders (Physical Disabilities)*
4. *Communication Disorders (Speech and Language Impairment)*
 - 315.31 Expressive Language Disorder
 - 315.32 Mixed Receptive-Expressive Language Disorder
 - 315.39 Phonological Disorder
 - 307.0 Stuttering
 - 307.9 Communication Disorder Not Otherwise Specified

5. *Pervasive Developmental Disorders (Autistic Spectrum Disorders)*
 - 299.00 Autistic Disorder
 - 299.80 Rett's Disorder
 - 299.10 Childhood Disintegrative Disorder
 - 299.80 Asperger's Disorder
 - 299.80 Pervasive Developmental Disorder
6. *Attention-Deficit and Disruptive Behavior Disorders (Emotional/Behavioral Disorders)*
 - 314.01 Attention-Deficit/Hyperactivity Disorder Combined Type
 - 314.00 Predominately Inattentive
 - 314.01 Predominately Hyperactive-Impulsive Type
 - 314.9 Attention-Deficit/Hyperactivity Disorder Not Otherwise Specified
7. *Conduct Disorder*
 - 312.81 Childhood-Onset Type
 - 312.82 Adolescent-Onset Type
 - 312.89 Unspecified Onset
 - 314.81 Oppositional Defiant Disorder
 - 312.9 Disruptive Behavior Disorder NOS
8. *Feeding and Eating Disorders of Infancy or Early Childhood*
 - 307.52 Pica
 - 307.53 Rumination Disorder
 - 307.59 Feeding Disorder of Infancy or Early Childhood
9. *Tic Disorders*
 - 307.23 Tourette's Disorder
 - 307.22 Chronic Motor or Vocal Tic Disorder
 - 307.21 Transient Tic Disorder
 - 37.20 Tic Disorder Not Otherwise Specified
 - Elimination Disorders
10. *Encopresis*
 - 787.6 With Constipation and Overflow Incontinence
 - 307.7 Without Constipation and Overflow Incontinence
 - 307.6 Enuresis (Not Due to a General Medical Condition)
11. *Other Disorders of Infancy, Childhood, or Adolescence*
 - 309.21 Separation Anxiety Disorder
 - 313.23 Selective Mutism
 - 313.89 Reactive Attachment Disorder of Infancy or Early Childhood
 - 307.3 Stereotypic Movement Disorder
 - 313.9 Disorder of Infancy, Childhood or Adolescence Not Otherwise Specified

Mental Retardation

The DSM-IV-TR defined mental retardation as significant below average intelligence with significant limitations in adaptive functioning in at least two of the following areas of skills: communication, self-care, use of

community resources, self-direction, functional academic skills, work, leisure, health, and safety. Mental retardation, known in many states and special education laws as cognitive disability, has a range from mild (50–55 to approximately 70 intelligence quotient (IQ)) to profound (below 20 or 25 IQ). Diagnostically, mental retardation can only be made with an IQ of about 70 and impairment in adaptive functioning. Adaptive functioning is a child's capacity to cope with life demands and how well a child meets standards of personal independence expected for his/her particular age group, sociocultural background, and community setting (Oakland and Houchins, 1985; Smalley, 1998).

Generally, intellectual functioning is an IQ or IQ equivalent assessed with one of the following individually administered standardized intelligence tests: Wechsler Intelligence Scales for children, 3rd Edition; Stanford-Binet 4th Edition; Kaufman Assessment Battery for Children. From a multicultural perspective, intelligence is a descriptive multivariate or multidimensional construct, and intelligence tests should be used to understand children and not to label them. In summary, mild mental retardation is an IQ level 50–55 to approximately 70; moderate mental retardation is an IQ level 35–40 to 50 to 55; severe mental retardation is an IQ level 20–25 to 35–40; finally, profound mental retardation is an IQ level below 20 or 25.

Learning Disorders

Learning disorders are diagnosed when a child's achievement on individually administered, standardized tests in reading, mathematics, or verbal expression is below his/her expected level for his/her age, schooling, or level of intelligence. To be diagnosable, learning disorders, known in many states and special education laws as learning disabilities, must significantly interfere with a child's academic achievement. Generally, a learning disability is defined by a discrepancy of two standard deviations (2 SD) or more between achievement measures in reading, mathematics, or written expression and IQ.

With children with limited English proficiency, such as African American and Latino children, the Test of Nonverbal Intelligence, 3rd Edition (TONI-3) can be useful (Brown *et al.*, 1997). The TONI-3 takes about 15–20 min to administer, and it measures nonverbally intelligence and visual-reasoning ability. Even though the TONI-3 is not based on any particular theory of intelligence, it has multicultural applications for assessing intelligence.

Motor Skills Disorders

Motor skills disorder, also called developmental coordination disorder and known in many states and special

education laws as physical disabilities, is the marked impairment in the development of motor coordination, and the diagnosis is made when impairment drastically affects academic achievement or activities of daily living. This diagnosis can be made if the coordination difficulties are not due to a general medical condition such as cerebral palsy, hemiplegia, or muscular dystrophy. With this disorder, younger children may display clumsiness and developmental delays in walking, crawling, sitting, tying shoelaces, buttoning shirts, and zipping pants. In contrast, older children may have difficulties with assembling puzzles, building toys, motor coordination while playing ball, and motor coordination difficulty with handwriting. The prevalence of this disorder is estimated to be about 6% of children in the age range of 5–11 years, and it is generally noticed when child attempts a motor task.

Communication Disorders

Communication disorders, known in many states and special education laws as speech and language impairments, are diagnosed through the assessment of receptive and expressive communication skills. Language refers to the verbal transactions that are used to receive and send messages for communication, and it is the essential way humans communicate thoughts and feelings. With children, expressive language refers to the ability to speak, write, or use sign language to communicate. In contrast, receptive language is when a child receives input from his/her senses and translates that into meaning.

Children with disabilities can display significant language deficits, and with school-age children communication disorder is a common disability. For example, with children under age 3, 10–15% experience language delays. With school-age children, 3–7% have language delays. Finally, developmental expressive language disorder is more common with boys than with girls (DSM-IV-TR). In terms of course of development, communication disorders can be noted by age 3, but expressive language disorder can occur at any age and can be due any form of organicity such as brain lesions, head traumas, or strokes. The assessment of communication disorders must take into account the child's culture and his/her language context, especially with children from bilingual homes.

Pervasive Developmental Disorders

The DSM-IV-TR classifies autistic spectrum disorders as pervasive developmental disorders, and these disorders are characterized by severe and ongoing impairment in reciprocal social interactions, communication skills, and stereotyped behavior. Under this category are autistic disorder, Rett's disorder, Asperger's disorder, and pervasive

developmental disorder not otherwise specified (American Psychiatric Association, 2000). As Dyches *et al.* (2004) pointed out, autistic disorder is an impairment in social and communication skills; and it can sometimes be referred to as infantile autism, childhood autism, or Kanner's autism.

Children with autistic spectrum disorders may have impairment with eye-to-eye contact, facial expressions, body postures, and gestures. Essentially, with this condition, children can have impairment in many nonverbal behaviors. Because autism is a developmental disorder, it has neurological origins that appear at birth or early development (Akshoomoff *et al.*, 2002; Dacey *et al.*, 1999; Filipek *et al.*, 2000).

The notion that autism is a broad spectrum disorder suggests that the skills and deficits vary from child to child. For example, some children with autism have normal or high IQs and others fall within the cognitive disability range. Moreover, some children with autism are highly verbal, while others do not have functional speech. Except for Rett's disorder, other pervasive developmental disorders are called autistic spectrum disorders (Atkinson, 1990; Atkinson *et al.*, 1992; Dyches *et al.*, 2004; Hviid *et al.*, 2003). Impairments in reciprocal social interaction, verbal and nonverbal communication, and restrictive, repetitive, and stereotyped behavior patterns are the major characteristics of autism (Autism Society of America, 2004). For instance, stereotyped behaviors include rocking, dipping, swaying, and clapping. Even a behavior as benign as a child walking on his/her tiptoes may contribute to the diagnosis of autism. The attachment to inanimate objects such as strings, rubber bands, and tin cans can also be indicators of this broad spectrum disorder.

The DSM-IV-TR defined adaptive functioning as the capacity a student has to cope with life demands and how well he/she meets the standards of personal independence expected for his or her particular age group, sociocultural background, and community setting (Oakland and Houchins, 1985; Smalley, 1998). The DSM-IV-TR reported the median rate of autism in epidemiological studies as 5 cases per 10 000 individuals and ranges from 2 to 20 cases per 10 000 individuals (Filipek *et al.*, 1999; Garnett and Attwood, 1997). The rate of autism is 4–5 times higher in males than females, and females with this disorder tend to have more severe cognitive disability than males. Diagnostically, the onset of autism is before the age of 3 years, and researchers are hoping to diagnose this spectrum disorder by 18 months so that interventions can occur (Ozonoff *et al.*, 2003). Familial patterns suggest that the probability of autistic spectrum disorder is increasing within families. For example, several individuals within a family may meet this diagnosis (Lord *et al.*, 2000).

The DSM-IV-TR described the diagnostic features of Rett's disorder as the development of multiple specific deficits following a period of normal functioning shortly after birth. At first, child development is normal through

the first 5 months of life. Even though at birth head circumference is normal, somewhere between 5 and 48 months, head growth decelerates. An associated feature of Rett's disorder is severe or profound cognitive disability, and in terms of prevalence, this disorder has only been reported with females. Finally, Rett's disorder has only been diagnosed in females.

The DSM-V-TR defined Asperger's disorder as impairment in social, occupational, and other areas of functioning, and unlike autistic disorder, there are no significant deficits in language acquisition. Very similar to autistic disorder, Asperger's disorder is defined as lack of social reciprocity, and eccentric and one-sided social interaction with others. Like other pervasive developmental disorders, the disorder is diagnosed at least 5 times more in males than in females. Even though there are no cures for autistic spectrum disorders, early assessment and interventions are critical for favorable outcomes. Experts recommend between 20 and 40 h of speech, occupational, and other interventions per week. Often, these interventions are not covered by private insurance and the cost can be as expensive as \$100 000 a year (Courchesne *et al.*, 2003; Volkmar, 2000; Wodrich and Barry, 1991; Yeargin-Allsopp *et al.*, 2003).

Developmental screening tests can be used to assess autistic spectrum disorders. Among the instruments that have been developed for screening are Checklist of Autism in Toddlers (CHAT; Baird *et al.*, 2000), the modified Checklist for Autism in Toddlers (M-CHAT; Robbins *et al.*, 2001), the Screening Tool for Autism in Two-Year olds (STAT; Stone *et al.*, 2000), and the Social Communication Questionnaire (SCQ; Berument *et al.*, 1999). The SCQ is used for children 4 years of age and older (Carter *et al.*, 1998; Couper and Sampson, 2003; De Bildt *et al.*, 2005; Ozonoff *et al.*, 2003). The first step for assessing an autistic spectrum disorder is to gather a developmental history from the parents of a child. Next, a professional must have direct interaction with a child. Finally, one has to review the child's records. Again, this level of assessment takes a team of professionals such as pediatricians, psychiatrists, psychologists, special educational teachers, audiologists, and parents (McEachin *et al.*, 1993; Stevens, 1986). In terms of research, the Autism Diagnostic Interview – Revised (ADI-R) is the gold standard for assessing autistic spectrum disorders, and it takes about 3 h (Lovaas, 1987; McDougle *et al.*, 2003; Ozonoff *et al.*, 2003). It is important that the SCQ (Western Psychological Services) is based on the ADI-R; but unlike the ADI-R, it is a parent-report questionnaire with empirically derived cutoff scores. In addition, a score of 15 distinguishes autistic spectrum disorders from nonautistic spectrum disorders. If a child receives a score of 11, additional screening is warranted. The SCQ has a sensitivity of 0.81 and a specificity of 0.73. A score of 22 distinguishes autism from other autistic spectrum disorders. Finally, a parent can complete the SCQ in about 10 min.

If one were to check the literature, it appears that there is a new Asperger's syndrome scale every month, but Ozonoff and her colleagues do not endorse any of these scales; however, she does recommend the ADI-R and SCQ. The difficulty with these scales is that they do not verify the diagnosis of Asperger's syndrome (Ehlers *et al.*, 1999; Scott *et al.*, 2002; Ottenbacher *et al.*, 1999; Tadevosyan-Leyfer *et al.*, 2003). Clearly, there are several instruments that can aid the diagnosis of autism spectrum disorders, and one may find a new one in the literature each month. The ADI-R is the gold standard for assessing autistic spectrum disorders, but it is a lengthy interview and takes about 3 h. Learners with autistic spectrum disorders tend to need very structured and specialized programs. It is critical to note that applied behavior analysis (ABA) is an evidence-based treatment that has been recommended for autistic spectrum disorders (Autism Society of America, 2004), and the ABA demonstrates the relationship between behavior and the environment (Sapp, 2004).

Attention-Deficit and Disruptive Behavior Disorders

In most states and special education laws, attention-deficit and disruptive disorders are partly the category, emotional/behavioral disorders (E/BD). This category encompasses conduct disorders, oppositional defiant disorders, and other emotionally and behaviorally related disorders. Specifically, attention deficit and disruptive disorders are characterized by inattention and/or hyperactivity-impulsivity, and children must display symptoms in school and home. Moreover, these symptoms must interfere with developmentally appropriate academic and social functioning. In terms of prevalence, 3–7% of school-age children are estimated to have these disorders, and these disorders are more dominant in boys than in girls. Even though behavior modification and stimulant medications are used to treat these disorders, within a school setting, these disorders can be enigmatic.

Conduct Disorders

The predominant features of conduct disorder are persistent patterns of behaviors that violate the rights of other children, and this disorder is broken down into two-types aggressive conduct that threatens the well-being of other children and nonaggressive conduct that results in property damage. The childhood-onset type of conduct disorder occurs before the age of 10, while the adolescent-onset type occurs after 10 years of age. The DSM-IV-TR reported prevalence rates ranging from less than 10% to more than 10%, and these rates are higher for boys than

girls. Likewise, boys tend to exhibit fighting, stealing, vandalism, and disruptions in school, and girls tend to display lying, truancy, running away, substance use, and prostitution. Finally, another pattern of this disorder is physical cruelty toward animals.

Oppositional Defiant Disorders

Even though conduct disorder and oppositional defiant disorder are both disruptive disorder in children, oppositional defiant disorder is a persistent pattern of negative, disobedient, or hostile behavior toward parents and teachers that last at least 6 months. A diagnosis of oppositional defiant disorder is made if a child has four or more of the following:

- Frequently loses temper.
- Frequently argues with parents or teachers.
- Frequently refuses to comply with rules or requests from adults.
- Frequently annoys others.
- Frequently blames other for his/her misbehavior.
- Frequently annoyed by others.
- Frequently anger and resentful.
- Frequently vindictive.

Feeding and Eating Disorders of Infancy or Early Childhood

Feeding and eating disorders of infancy or early childhood include Pica, rumination disorder, and feeding disorder of infancy or early childhood. Children or infants are diagnosed with Pica when they eat one or more nonnutritive substance on a regular period for at least 1 month. Younger children and infants tend to eat paint, plaster, cloth, and hair, while older children tend to eat sand, insects, animal droppings, leaves, and pebbles. In contrast, adolescents tend to eat mud or soil. Children with this disorder tend not to have aversion to food and the behavior has to be developmentally and culturally inappropriate. Before about 18 months, it is common for infants to eat non-nutritive substance and this would not warrant the diagnosis of Pica.

Ruminations disorder is the repetitive regurgitation and rechewing of food in a child and it lasts at least 1 month, but the condition cannot be due to an associated gastrointestinal condition such esophageal reflux. Fortunately, rumination disorder is rare, but when it happens, children can become malnourished. Feeding disorder of infancy or early childhood is characterized by the inability of an infant or a child to eat adequately for at least 1 month, and the result is significant weight loss; however, there cannot be medical conditions that account for the feeding disorder. Finally, children with this disorder tend to be irritable and withdrawn.

Tic Disorders

Tic disorders includes Tourette's disorder, chronic motor or vocal tic disorder, transient tic disorder, tic disorder not otherwise specified, and a tic is a sudden stereotyped motor movement or vocalization, and motor and vocal tics can be simple, involving a few muscles or a sound, or complex, involving many muscle groups, or many sounds. Simple vocal tics can be meaningless vocalizations such as grunting, sniffing, snorting, or chirpings; and complex vocal tics involve coprolalia (sudden vocalization of obscenities), palilalia (repeating one's own words), and echolalia (repeating the words of others). Complex motor tics include facial contortions, eye blinking, gestures, jumping, and twisting; and this disorder includes copropraxia (sudden tic-like vulgar gestures and echopraxia) tic-like imitation of another's movements.

Elimination Disorders

Encopresis and enuresis are the two most common elimination disorders with children that are not due exclusively to a medical condition. Encopresis is the elimination of feces into inappropriate places, while enuresis is the elimination of urine into inappropriate places. Encopresis cannot be diagnosed until a child is 4 years old and enuresis cannot be diagnosed until a child is 5 years old.

Other Disorders of Infancy, Childhood, or Adolescence

Often, disorders of infancy, childhood, or adolescence include separation anxiety disorder, selective mutism, reactive attachment disorder of infancy or early childhood, stereotypic movement disorder, and disorder of infancy, childhood, or adolescent not otherwise specified. separation anxiety disorder is excessive anxiety that a child has concerning his/her home or family, and this anxiety is not developmentally appropriate. Selective mutism, formerly elective mutism, is the failure of a child to speak in school, home, or with playmates. Reactive attachment disorder occurs when a child does not initiate or respond to social interactions with family members or friends, or the child shows a pattern of hypervigilance such as watchfulness or freezing in place.

Unlike pervasive developmentally disorders, reactive attachment disorder is not due only to developmentally delays. Stereotypic movement disorder, formerly stereotypy habit disorder, is repetitive nonfunctional motor behaviors that interfere with school or home activities, and these movements can include rocking, hand waving, head banging, self-biting, or a child hitting various parts of his/

her body. Finally, disorders of infancy, childhood, or adolescence not otherwise specified are disorders that do not meet criteria for a specific disorder with the classification.

Conclusion

From this article, scholars can conclude the following:

- Diagnosis and classification are complex multiaxial processes.
- MMBA is a comprehensive intra-individually based assessment that has multicultural appeal.
- Test score reliability and validity are based on educational test score theories, not facts.
- Prejudice can influence diagnosis and classification.
- Culture and language must be included within assessment processes.

This article discussed the intricacies of diagnosing and classifying children with special needs, and this is especially intriguing within a multicultural context. Moreover, minority children are seldom part of standardization samples for educational diagnosis and classification. Diagnosis and classification include processes of using tests and other measures to assess the educational needs of children with special needs. Unlike assessment in general education, assessment of children with special needs can be different for each child; and it must be an intra-individual appraisal of a child's needs. Finally, diagnosis and classification must be modified to fit the unique needs of children requiring special education (Venn, 2000).

See also: Educating Students with Autism and Related Disorders; Educating Students with Special Needs: An Overview.

Bibliography

- Akshoomoff, N., Pierce, K., and Courchesne, E. (2002). The neurobiological basis of autism from a developmental perspective. *Development and Psychopathology* **14**, 613–634.
- Allen, M. J. and Yen, W. M. (1979). *Introduction to Measurement Theory*. Belmont, CA: Wadsworth/Thomson Learning.
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR*, 4th edn., text revision. Washington, DC: American Psychiatric Association.
- Atkinson, L. (1990). Intellectual and adaptive functioning: Some tables for interpreting the Vineland in combination with intelligence tests. *American Journal on Mental Retardation* **95**(2), 198–203.
- Atkinson, L., Beve, I., Dickens, S., and Blackwell, J. (1992). Concurrent validities of the Stanford-Binet (fourth edition), Leiter, and Vineland with developmentally delayed children. *Journal of School Psychology* **30**, 165–173.
- Autism Society of America (2004). Biomedical and Dietary Treatments (Fact Sheet). Bethesda, MD: Autism Society of America. <http://www.autism-society.org/site/PageServer?pagename=BiomedicalDietaryTreatments> (accessed August 2009).
- Baird, G., Charman, T., Baron-Cohen, S., et al. (2000). A screening instrument for autism at 18 months of age: A 6-year follow-up study.

- Journal of the American Academy of Child and Adolescent Psychiatry* **39**, 694–702.
- Berument, S. K., Rutter, M., Lord, C., Pickles, A., and Bailey, A. (1999). Autism Screening Questionnaire: Diagnostic validity. *British Journal of Psychiatry* **175**, 444–451.
- Brown, L., Shebenou, R. J., and Johnson, S. K. (1997). *Test of Nonverbal Intelligence: A language-Free Measure of Cognitive Ability*, 3rd edn. Austin, TX: Pro-Ed.
- Carter, A. S., Volkmar, F. R., Sparrow, S. S., et al. (1998). The Vineland Adaptive Behavior Scales: Supplementary norms of the individuals with autism. *Journal of Autism and Developmental Disorders* **28**(4), 287–302.
- Couper, J. J. and Sampson, A. J. (2003). Children with autism deserve evidence-based intervention. *Medical Journal of Australia* **178**, 424–425.
- Courchesne, E., Carper, R., and Akshoomoff, N. (2003). Evidence of brain overgrowth in the first year of life in autism. *Journal of the American Medical Association* **290**(3), 337–344.
- Dacey, C. M., Nelson, W. M. III, and Stoekel, J. (1999). Reliability, criterion-related validity and qualitative comments of the fourth edition of the Stanford-Binet Intelligence Scale with a young adult population with intellectual disability. *Journal of Intellectual Disability Research* **43**(3), 179–184.
- De Bildt, A., Kraijer, D., Sytema, S., and Minderaa, R. (2005). The psychometric properties of the Vineland Adaptive Behavior Scales in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders* **35**(1), 53–62.
- Dykes, T. T., Wilder, L. K., Sudweeks, R., Obiakor, F. E., and Algozzine, B. (2004). Multicultural perspectives on autism. *Journal of Autism and Developmental Disorders* **43**(2), 211–222.
- Ehlers, S., Gillberg, C., and Wing, L. (1999). A screening questionnaire for Asperger syndrome and other high-functioning autism spectrum disorders in school age children. *Journal of Autism and Developmental Disorders* **29**(2), 129–141.
- Filipek, P. A., Accardo, P. J., Ashwal, S., et al. (2000). Practice parameter: Screening and diagnosis of autism. *Neurology* **55**, 468–479.
- Filipek, P. A., Accardo, P. J., Baranek, G. T., et al. (1999). The screening and diagnosis of autism spectrum disorders. *Journal of Autism and Developmental Disorders* **29**(2), 439–484.
- Garnett, M. S. and Attwood, A. J. (1997). The Australian scale for Asperger's syndrome. In Attwood, T. (ed.) *Asperger's Syndrome: A Guide for Parents and Professionals*. London: Jessica Kingsley.
- Hviid, A., Stellfeld, M., Wohlfahrt, J., and Melbye, M. (2003). Association between thimerosal-containing vaccine and autism. *Journal of the American Medical Association* **290**(13), 1763–1766.
- Kendall, P. C. and Hammen, C. (1998). *Abnormal psychology: Understanding human problems*, 2nd edn. Boston, MA: Houghton-Mifflin.
- Lazarus, A. A. (1990). Multimodal applications and research: A brief overview and update. *Elementary School Guidance and Counseling* **24**(4), 243–247.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., et al. (2000). The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders* **30**(3), 205–230.
- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology* **55**, 3–9.
- McDougle, C. J., Stigler, K. A., and Posey, D. J. (2003). Treatment of aggression in children and adolescents with autism and conduct disorder. *Journal of Clinical Psychiatry* **64**(supplement 4), 16–25.
- McEachin, J. J., Smith, T., and Lovaas, O. I. (1993). Long-term outcome for children with autism who received early intensive behavioral treatment. *American Journal on Mental Retardation* **97**, 359–372.
- Oakland, T. and Houchins, S. (1985). A review of the Vineland Adaptive Behavior Scales, survey form. *Journal of Counseling and Development* **63**, 585–586.
- Ottenbacher, K. J., Msall, M. E., Lyon, N., et al. (1999). Measuring developmental and functional status in children with disabilities. *Developmental Medicine and Child Neurology* **41**, 186–194.
- Ozonoff, S., Rogers, S. J., and Hendren, R. L. (2003). *Autism Spectrum Disorders: A Research Review for Practitioners*. Washington, DC: American Psychiatric.
- Robbins, D. I., Fein, D., Barton, M. I., and Green, J. A. (2001). The modified checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders* **31**(2), 149–151.
- Sapp, M. (1999). *Test Anxiety: Applied Research, Assessment, and Treatment Interventions*, 2nd edn. Lanham, MD: University Press of America.
- Sapp, M. (2002). *Psychological and Educational Test Scores: What Are They?* Springfield, IL: Charles C. Thomas.
- Sapp, M. (2004). *Cognitive-Behavioral Theories of Counseling: Traditional and Nontraditional Approaches to Counseling*. Springfield, IL: Charles C. Thomas.
- Sapp, M. (2006). *Basic Psychological Measurement, Research Designs, and Statistics without Math*. Springfield, IL: Charles C. Thomas.
- Scott, F. J., Baron-Cohen, S., Bolton, P., and Brayne, C. (2002). The CAST (Childhood Asperger Syndrome Test): Preliminary development of a UK screen for mainstream primary-school-age children. *Autism* **2**(1), 9–31.
- Smalley, S. I. (1998). Autism and tuberous sclerosis. *Journal of Autism and Developmental Disorders* **28**(5), 407–414.
- Stevens, F. I. (1986). Vineland Adaptive Behavior Scales: Classroom edition. *Journal of Counseling and Development* **65**, 112–113.
- Stone, W. L., Coonrod, E. E., and Ousley, O. Y. (2000). Brief report: Screening tool for autism in two-year-olds (STAT): Development and preliminary data. *Journal of Autism and Developmental Disorders* **30** (6), 607–612.
- Tadevosyan-Leyfer, O., Dowd, M., Mankoski, R., et al. (2003). A principal components analysis of the autism diagnostic interview-revised. *Journal of the American Academy of Child and Adolescent Psychiatry* **42**(7), 864–872.
- Venn, J. J. (2000). *Assessing Students with Special Needs*, 2nd edn. Upper Saddle River, NJ: Prentice-Hall.
- Volkmar, F. R. (2000). Medical problems, treatments, and professionals. In Powers, M. D. (ed.) *Children with Autism: A Parent's Guide*, 2nd edn. Bethesda, MD: Woodbine House.
- Wodrich, D. L. and Barry, C. T. (1991). A survey of school psychologists' practices for identifying mentally retarded students. *Psychology in the Schools* **28**, 165–171.
- Yeargin-Allsopp, M., Rice, C., Karapurkar, T., et al. (2003). Prevalence of autism in a US metropolitan area. *Journal of the American Medical Association* **289**(1), 49–55.

Direct Instruction and the Education of Children with Special Needs

T Mehring, Emporia State University, Emporia, KS, USA

© 2010 Elsevier Ltd. All rights reserved.

For most students with exceptionalities, a broad array of explicit and implicit instructional techniques and strategies has been known to be helpful. Clearly, this broad array of techniques has the potential to help a diversity of learners acquire and use academic skills and content. One approach that has proven to be successful with many students with mild cognitive and learning disabilities is direct instruction (Ormrod, 2008), which is the focus of this article. This model can be used in inclusive classrooms to benefit all learners as well as in special education classrooms with individuals or small groups of students with special needs. Historically, this model emerged from the work of Siegfried Engelmann and his colleagues in the mid-1960s. The underlying assumptions of the model are:

1. all children can be taught;
2. the learning of basic skills and their application in higher-order skills are essential to intelligent behavior and should be the focus of a compensatory education program; and
3. the student who is not achieving at grade level should be taught at a faster rate than typically occurs if they are to catch up with peers (Engelmann *et al.*, 1988: 303).

Stein *et al.*, (1997) emphasized three essential teaching components for the direct instruction model: instructional design, presentation techniques, and instructional organization. Each of these is considered essential for successful teaching and learning.

Direct instruction is an instructional model designed to teach well-defined knowledge and skills that are needed for later learning (Eggen and Kauchak, 2006; Rosenshine and Stevens, 1986). It is a form of instruction that encompasses the use of a systematic lesson structure as well as a set of key instructional behaviors (Carnine *et al.*, 1990; Becker *et al.*, 1982). As an instructional approach, it ranges from a highly structured, nearly scripted approach (Carnine *et al.*, 2006; Koziuff *et al.*, 2000) to one that is more flexible and cognitive in its direction (Eggen and Kauchak, 2006; Rosenshine and Stevens, 1986). It uses a variety of techniques, including questioning, cueing, and reinforcement of students. In addition, direct instruction is most suitable for teaching information and skills that are well defined and that should be taught in a step-by-step sequence (Rosenhine and Stevens, 1986). When using direct instruction, teachers proceed with instruction

at a brisk pace. They are organized and prepared to conduct the lesson as expeditiously as possible. Large amounts of content are dissected into short lessons so that most students are able to succeed. Every student has a chance to participate, few students answer questions incorrectly, and transitions between activities in the lesson are brief and smooth. This approach is generally implemented more easily with individual students and/or small groups rather than with an entire classroom. When implemented appropriately, direct instruction can lead to significant gains in achievement of basic skills and higher-level thinking processes (Ellis and Rock, 2002). High student interest and self-efficacy for the subject matter in question and low rates of student misbehavior have also been demonstrated (Weinert and Helmke, 1995; Tarver, 1992; Rosenshine and Stevens, 1986).

Direct instruction is an approach that incorporates elements of both expository instruction and mastery learning. Varied techniques are used to keep students continually and actively engaged in learning and applying classroom subject matter (Tarver, 1992; Gagne, 1985; Weinert and Helmke, 1995; Englemann and Carnine, 1982). To some extent, direct instruction is based on applications from behavioral and cognitive psychology. Incorporating behaviorism, direct instruction requires learners to make frequent responses and provides immediate reinforcement of correct responses through teacher feedback. Cognitive psychology is incorporated into this teaching approach through focusing instruction on the importance of attention and long-term memory-storage processes in learning, focusing on working memory and the value of learning basic skills to automaticity (Rosenhine and Stevens, 1986).

Theoretical and Practical Underpinnings

Different theorists (Carnine *et al.*, 1997; Lerner, 2006; Mainzer *et al.*, 2003) describe and implement direct instruction differently. However, in general, these theorists agree that direct instruction involves small and carefully sequenced steps, fast pacing, and significant teacher–student interaction. Specific teaching behaviors that have been shown to impact student achievement are key ingredients related to successful implementation of direct instruction. According to Rosenshine (1986) and Rosenshine and Stevens (1986), direct instruction:

- teaches academic skills directly;
- is teacher directed and controlled;
- uses carefully sequenced and structured materials;
- provides students mastery of basic skills;
- sets goals that are clear to students;
- allocates sufficient time for instruction;
- uses continuous monitoring of student performance;
- provides immediate feedback to students; and
- teaches a skill until mastery of that skill is achieved.

Each lesson typically involves most or all of the following components:

1. *Review of previously learned material.* The teacher or service provider reviews relevant content from previous lessons, checks homework assignments, involving that content, and reteaches any information or skills that students have not yet mastered.
2. *Statement of the objectives of the lesson.* The teacher or service provider describes one or more concepts or skills that students should master in the new lesson. Teachers explicitly tell students what will be covered during instruction, establish performance goals, and engage students in determining their daily learning goals to produce higher levels of achievement (Ellis and Rock, 2002), and describe how current learning relates to what has previously been learned.
3. *Presentation of new material in small, logically sequenced steps.* The teacher or service provider presents a small amount of information or a specific skill. An advanced organizer, question, or scaffolding of students' efforts may be used to enhance the ability of students to process and remember the material. Modeling involving teacher demonstration of how a task is completed correctly is provided to assist students in learning new information. Thinking aloud while completing the model also helps students make a cognitive connection to the knowledge or skill to be learned. Modeling should also include many examples of what the expected behavior looks like. Nonexamples should be presented so that students can clearly differentiate expected performance. Teacher-directed questioning throughout the lesson serves as a check to determine whether students are acquiring knowledge.
4. *Guided student practice and assessment after each step.* Guided practice involves student rehearsal while the teacher closely supervises student work. Students are provided numerous opportunities to practice what they are learning. Practice may include answering questions, solving problems, or performing modeled procedures. The teacher gives hints during students' early responses, provides immediate feedback about performance, makes suggestions about how to improve, and provides remedial instruction as needed. After students have completed guided practice, the teacher checks to be sure that mastery of the information or skill has occurred. Unison response using signals (e.g., finger snap, hand

drop, clap, touching a board, or presentation material) are used to cue students to demonstrate knowledge of information. Typically, the teacher gives directions or asks a question, provides a thinking pause, and then cues the response. Using a signal to cue the response gives slower students in the group extra time (i.e., 5–10 s) to think about the answer. The overall pacing of direct instruction lessons is relatively quick. The idea is to present numerous examples in rapid succession so that students can see how the examples are similar to and different from one another. If presentation of examples is too slow, students may forget the earlier examples and will not be able to compare them to later examples (Adams and Engelmann, 1996).

5. *Independent practice.* Once students have acquired a specified benchmark of mastery (e.g., 85% of questions answered correctly), they engage in additional activities (either independently or in small groups) to learn the material to automaticity. Independent practice involves student performance of the new task without assistance.
6. *Frequent follow-up reviews.* The teacher provides many opportunities for students to review previously learned material through homework assignments, writing tasks, paper-and-pencil quizzes, or other appropriate activities (Lerner, 2006). Daily review at the beginning, middle, or end of a lesson can alert teachers to less-than-acceptable levels of learning. When unacceptable learning is detected, reteaching occurs.

It is critical to provide frequent and immediate feedback and correction to students as they are engaged in guided and independent practice. Verbal praise or other types of reinforcement for correct performance assure students they are correctly demonstrating critical elements of the task to be learned. Likewise, alerting students when they are not on target redirects them to correct task performance.

Two direct instruction methods have been used extensively with students with severe reading problems and are not typically used in general education classrooms. The Direct Instruction Reading Program (Carnine *et al.*, 1990) teaches basic decoding and comprehension skills. Lessons are highly structured and are based on carefully sequenced skill hierarchies that emphasize letter–sound correspondence, word identification, and use of words in stories. The program uses a synthetic phonics approach and emphasizes auditory blending to combine isolated sounds into words. The Corrective Reading Program (Engelmann, 1999) is designed for students in grades 4 through 12 and consists of two strands: decoding and comprehension.

Future Perspectives

Recent developments in the field of education, such as changes made to the Individuals with Disabilities Education

Act (IDEA), the promulgation of standards-based education, and the implications of No Child Left Behind (NCLB) legislation, are having a profound effect on the programs of students with special needs. Demonstration of and accountability for student learning are critical elements of each of these directives.

Several studies have documented the effectiveness of direct instruction with students who have special needs (Mainzer *et al.*, 2003; Carnine *et al.*, 1990). Students with histories of problems in automaticity, metacognitive strategies, memory, attention, generalization, proactive learning, and motivation have demonstrated learning gains when engaged in direct-instruction teaching approaches (Howell and Nolet, 2000). Direct instruction is particularly beneficial for students who have a rigid motivational pattern, lack significant prior knowledge, or encounter repeated failure on a task (Howell and Nolet, 2000). Stanovich (1994) found that direct instruction and teacher-directed strategy training are more effective for teaching word-recognition skills, especially for at-risk children, children with learning disabilities, and children with special needs. Numerous researchers have demonstrated that direct instruction helps students at risk for learning problems acquire reading and spelling skills (Moats, 2000; and Lyon, 1995). Slavin *et al.* (1994) found that direct instruction improved reading skills in students at risk for reading failure. Mercer *et al.* (1994) found that direct-instruction methods, such as description of procedures, modeling of skills, use of cues and prompts, direct questioning of students to ensure understanding, and practice to mastery, result in improved mathematics performance for students with math disabilities. Leinhardt and Pallas (1982) and Rosenshine and Stevens (1986) documented the effectiveness of direct instruction with students who have learning disabilities. Direct instruction was found to be effective in working with low achievers and students with exceptionalities (Kroesberger *et al.*, 2004; Troia and Graham, 2002).

Although the direct instruction model has been around for several decades, new applications and improvements continue to emerge. Direct-instruction programs may be used with all students from whom high levels of academic achievement are expected and desired (Tarver and The Division for Learning Disabilities and Division for Research Alerts Editorial Committee, 1999). The repetition, reinforcement, careful sequencing of skills, and correction procedures are helpful for all students but are critical for students with disabilities. This instructional model sets up students for success and ensures mastery-level learning prior to moving on to new content. This must continue to be the vision of the future in the education of students with special needs.

Conclusion

Direct instruction allows teachers to carefully plan and execute specific strategies and presentations to enhance

learning for all students. Through using direct instruction, teachers ensure that students are well informed about what is expected, what is being learned, why it is being learned, and how it can be used (Ellis and Larkin, 1998). Effective instruction takes place in classrooms in which students have frequent, sustained, and consistent opportunities to read, write, listen, and talk about what they are learning (Schmidt *et al.*, 2002). Direct instruction provides an instructional approach that results in student learning and student success.

Bibliography

- Adams, G. L. and Engelmann, S. (1996). *Research on Direct Instruction: 25 Years Beyond DISTAR*. Seattle, WA: Educational Achievement Systems.
- Becker, W. C., Engelmann, S., Carnine, D. W., and Magge, A. (1982). Direct instruction technology: Making learning happen. In Karoly, P. and Steffen, J. J. (eds.) *Improving Children's Competence: Advances in Child Behavioral Analysis and Therapy*, vol. 1, pp 151–204. Lexington, MA: Heath.
- Carnine, D., Silbert, J., Kameenui, E., Tarver, S., and Jongjohann, K. (2006). *Teaching Struggling and At-Risk Readers: A Direct Instruction Approach*. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Carnine, D., Silbert, J., and Kameenui, E. J. (1990). *Direct Instruction Reading*, 2nd edn. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Carnine, D., Silbert, J., and Kameenui, E. J. (1997). *Direct Instruction Reading*, 3rd edn. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Ellis, E. S. and Rock, M. L. (2002). Strategic solutions to promoting incremental change. *Exceptionality* **10**, 223–247.
- Eggan, P. and Kauchak, D. (2006). *Strategies and Models for Teachers: Teaching Content and Thinking Skills*, 5th edn. Boston, MA: Allyn and Bacon.
- Engelmann, S. (1999). *Corrective Reading Program*. Chicago, IL: Science Research Associates.
- Engelmann, S., Becker, W. C., Carnine, D., and Gersten, R. (1988). The direct instruction Follow Through model: Design and outcomes. *Education and Treatment of Children* **11**(4), 303–317.
- Engelmann, S. and Carnine, D. (1982). *Theory of Instruction: Principles and Applications*. New York: Irvington.
- Gagne, E. D. (1985). *The Cognitive Psychology of School Learning*. Boston, MA: Little, Brown.
- Howell, K. W. and Nolet, J. V. (2000). *Curriculum-Based Evaluation: Teaching and Decision Making*, 3rd edn. Belmont, CA: Wadsworth/Thomson Learning.
- Koziuff, M. A., La Nunciata, L., and Cowardin, J. (2000). Direct instruction: Its contributions to high school achievement. *High School Journal* **84**, 54–71.
- Kroesberger, E. H., van Luit, E. H., and Maas, C. J. (2004). Effectiveness of explicit and constructivist mathematics instruction for low-achieving students in the Netherlands. *Elementary School Journal* **104**, 233–251.
- Leinhardt, G. and Pallas, A. (1982). Restrictive educational settings: Exile or haven? *Review of Educational Research* **52**, 557–578.
- Lerner, J. (2006). *Learning Disabilities and Related Disorders*, 10th edn. New York: Houghton Mifflin.
- Lyon, G. R. (1995). Research initiatives in learning disabilities: Contributions from scientists supported by the National Institute of Child Health and Human Development. *Journal of Child Neurology* **10**(supplement 1), 120–126.
- Mainzer, R., Deshler, D., Coleman, M., Kozleski, M., and Rodriguez-Walling, J. (2003). To ensure the learning of every child with a disability. *Focus on Exceptional Children* **35**(5), 1–12.
- Mercer, C. D., Jordan, L., and Miller, S. P. (1994). Implications of constructivism for teaching math to students with mild to moderate disabilities. *Journal of Special Education* **28**, 290–306.
- Moats, L. C. (2000). *Speech to Print: Language Essentials for Teachers*. Baltimore, MD: Paul H. Brooks.

- Ormrod, J. E. (2008). *Educational Psychology*, 6th edn. Upper Saddle River, NJ: Pearson Prentice Hall.
- Rosenshine, B. (1986). Synthesis of research on explicit teaching. *Educational Leadership* **43**, 60–69.
- Rosenshine, B. and Stevens, R. (1986). Teaching functions. In Wittrock, M. (ed.) *Handbook of Research on Teaching*, 3rd edn., pp 376–391. New York: Macmillan.
- Schmidt, R. J., Rozendal, M. S., and Greenman, G. G. (2002). Reading instruction in the inclusive classroom: Research-based practices. *Remedial and Special Education* **23**, 130–140.
- Slavin, R. E., Madden, N. A., Dolan, L. J., et al. (1994). Success for all: Longitudinal effects of systemic school-by-school reform in seven districts. *Paper Presented at the Meeting of the American Educational Research Association*, New Orleans, LA.
- Stanovich, K. E. (1994). Constructivism in reading education. *Journal of Special Education* **28**, 259–274.
- Stein, M., Silbert, J., and Carmine, D. (1997). *Designing Effective Mathematics Instruction: A Direct Instruction Approach*, 3rd edn. Upper Saddle River, NJ: Merrill.
- Tarver, S. G. (1992). Direct instruction. In Stainback, W. and Stainback, S. (eds.) *Controversial Issues Confronting Special Education*, pp 143–165. Boston, MA: Allyn and Bacon.
- Tarver, S. G. and The Division for Learning Disabilities and Division for Research Alerts Editorial Committee (1999). Focusing on direct instruction. *Current Practice Alerts* **2**, 1–4.
- Troia, G. A. and Graham, S. (2002). The effectiveness of a highly explicit, teacher directed strategy instruction routine: Changing the writing performance of students with learning disabilities. *Journal of Learning Disabilities* **35**, 290–305.
- Weinert, F. E. and Helmke, A. (1995). Learning from wise Mother Nature or Big Brother Instructor: The wrong choice as seen from an educational perspective. *Educational Psychologist* **30**, 135–142.

Further Reading

- Ellis, E. S. and Larkin, M. J. (1990). Strategic instruction for adolescents with learning disabilities. In Wong, B. Y. L. (ed.) *Learning About Learning Disabilities*, 2nd edn., pp 585–656. San Diego, CA: Academic Press.

Disabilities and Gifted Education

V I Daniels, Southern University and A & M College, Baton Rouge, LA, USA

M J McCollin, Slippery Rock University of Pennsylvania, Slippery Rock, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

The potential for giftedness exists among all racial, ethnic, cultural, and linguistic groups (Daniels, 2003, 2007), and for every segment of the population of students with disabilities (Bianco, 2005). However, prior to the 1970s, emphasis on students who are both gifted and disabled was virtually nonexistent (Grimm, 1998). It was during the 1970s that Maker (1977) increased educators' awareness of the needs of gifted students with disabilities. At first, the primary emphasis was almost exclusively on gifted students with learning disabilities, a cluster of students who soon became recognized as twice exceptional. Although it has long been realized that intellectual giftedness exists among many different populations of persons with disabilities, such as those with a learning disability (LD), physical disability, speech and language impairment, deafness or hearing impairment, blindness or visual impairment, autism, Asperger's syndrome, cerebral palsy, attention-deficit hyperactivity disorder, behavioral disorder, and other disabling conditions (Cash, 1999; Little, 2002; Hardman *et al.*, 2008; Karnes *et al.*, 2004; King, 2005; Neihart, 2000; Nielsen *et al.*, 1993; Reis and McCoach, 2002), the focus on gifted students with disabilities was not that widespread.

There are many prominent individuals with disabilities who have gifts and talents. Among them are actors, actresses, celebrities, singers, scholars, and world leaders. Helen Keller (author, political activist, and lecturer) had dual sensory impairment (deafness-blindness). Franklin D. Roosevelt (32nd US President) had polio and limited physical ability. Stephen Hawking (theoretical physicist) has amyotrophic lateral sclerosis (ALS), sometimes called Lou Gehrig's disease. Walt Disney (renowned cartoonist who produced the first full-length animated motion picture) was a slow learner and considered to be learning disabled. Albert Einstein (mathematician/physicist, provided the theory of relativity, and was also awarded the Nobel prize in physics) had (LD) and did not speak until age 3. Thomas Edison (American inventor) was deaf. Ray Charles (musician) was blind. Tom Cruise (actor, film producer, Golden Globe Award winner) and Whoopi Goldberg (Oscar, Grammy, Golden Globe, NAACP Image Award winning actress; comedienne, media personality, high school dropout) have dyslexia. Stevie Wonder (musician, singer) is blind.

Although the population of students with dual exceptionalities is a relatively new category among the gifted (Gallagher and Gallagher, 1994), they continue to remain underserved, underestimated, and underidentified. Problems

in identifying and determining the number of students with disabilities in gifted programs can be associated to a number of factors, including the differing definitions of giftedness, teacher biases, inadequate teacher training, the varying theoretical and political views, and assessment procedures. According to Karnes *et al.* (2004), the number of students with disabilities in gifted programs is estimated to "range from 120,000 to 180,000 (Davis and Rimm, 1998), to as many as 540,000 (Minner, 1990)" (p. 16). While there remains a sparsity of empirical data to support these estimates, it is well known that discrepancies in prevalence estimates do and can occur in all categories of exceptionality. In the case of students who are twice exceptional (gifted and disabled), these discrepancies seem more variable than most (Hardman *et al.*, 2008). This article focuses on how to identify gifted students with disabilities and how to help them to maximize their fullest potential.

Identifying Gifted Students with Disabilities

Giftedness is multifaceted, and identifying high-potential students who are twice exceptional can be a daunting and challenging task. Most identification strategies used for giftedness have focused primarily on cognitive abilities, with few considering creative behaviors (Clark, 1997). As giftedness is primarily associated with high intelligence quotient (IQ) scores and high scores on aptitude tests, these measures have not been particularly useful in identifying children with disabilities who are intellectually or otherwise gifted (Hardman *et al.*, 2008). However, what is known is that the assessment procedures typically used to assess giftedness in the general population are inadequate, without major modifications, for unmasking the hidden potential of gifted children with disabilities (Lovett and Lewandowski, 2006; Maker, 1977).

The best practices for identifying students with gifts and talents, including twice-exceptional students, are those that include multiple criteria. Multiple criteria, as explained by Coleman (2003), involve multiple types of information that provide indicators of the student's cognitive abilities, academic achievement, performance in a variety of settings as well as their interests, creativity, motivation, and learning characteristics/behaviors; multiple sources of information such as test scores, school

grades, and comments by classroom and specialty area teachers, counselors, parents, peers, and students themselves; and multiple time periods to ensure that students with gifts and talents are not missed, overlooked, or unidentified by one-shot identification procedures. Although attempts to identify twice-exceptional students (a term coined by James Gallagher) began during the 1970s, it was not until 2004 that federal law (the Individuals with Disabilities Education Improvement Act, P. L. 108–446) declared in Subpart 4, Section 681, (d), (3), (J) gifted and talented students with disabilities as a priority whose needs could be addressed with federal funding.

There are a number of barriers hindering the identification and referral of students with disabilities for gifted programs. Among the many barriers . . . are teachers' stereotypic beliefs (Cline and Hedgeman, 2001; Johnson *et al.*, 1997; Minner *et al.*, 1987; St. Jean, 1996) and inadequate teacher preparation (Davis and Rimm, 2004; Johnson *et al.*, 1997). According to Cline and Hedgeman, stereotypic expectations work against gifted students with disabilities in two ways: (1) misconceptions about the characteristics of gifted students with gifts and talents, and (2) low expectations for students with disabilities (see Bianco, 2005). Other factors that contribute to students with disabilities not being included in gifted education include:

1. the likelihood of not being identified as having an LD when the severity of the discrepancy between IQ and achievement is sufficiently high (MacMillan *et al.*, 1998);
2. students not being allowed sufficient (e.g., enough extra) time or appropriate (e.g., assistive) equipment during testing situations for them to demonstrate their talents in spite of their disabilities;
3. the inflexibility of states' regulations for education of the gifted (Grimm, 1998);
4. bias about some disabilities (e.g., cerebral palsy) that overshadow actual abilities and talents (Willard-Holt, 1998); and
5. the narrow views of some individuals that high abilities and learning problems cannot possibly coexist (Brody and Mills, 1997).

Thus, it is clear that identifying the hidden gifts and talents of twice-exceptional students requires the use of multiple assessment criteria, and the collaborative efforts of general-classroom teachers, special-education teachers, gifted-education teachers, and parents. Due to the complex nature of their dual exceptionalities, twice-exceptional students become at-risk youngsters because their diverse educational and social-emotional needs are often undetected; and they are overlooked or go unrecognized for long periods of time because teachers view them within the general population as underachievers or average learners (Nielsen, 2002).

Characteristics of Gifted Students with Disabilities

Twice-exceptional students are a highly diverse and very heterogeneous group. They face a number of challenges in the classroom and in their personal lives. Their unique characteristics, as seen by some (Nielsen and Higgins, 2005), thrust them into an emotional storm which become evident when they are expected to acquire specific academic skills that are underscored by prerequisite skills which they do not have or have not yet acquired; and, when they are expected to demonstrate appropriate social skills when these skills are beyond their reach. Initially, teachers view the abilities (e.g., high creativity and critical-thinking skills) of twice-exceptional learners as positive, exciting, and challenging, but their enthusiasm soon changes when they become frustrated with the inability of these youngsters to demonstrate academic skills, and with their recurring and sometimes extreme behavioral difficulties (see Nielsen and Higgins, 2005).

In the literature, twice-exceptional students are generally characterized as having gifts and talents in one or more areas of exceptionality (e.g., specific academics, creativity, and performing arts) and as having a disability, as defined by both federal and state law. Typical characteristics for this population are generally described as areas of giftedness and areas of challenge. In the areas of giftedness, twice-exceptional learners are characterized as having a wide range of interests that are often unrelated to school topics, an exceptional memory, a superior vocabulary, an unusual imagination, exceptional problem-solving and reasoning skills, an unrelenting sense of curiosity, an unusual sense of humor that may sometimes appear bizarre, an ability to manipulate situations to their own advantage as a means of compensating for their disability, a high level of creativity in their approach to tasks to mask or compensate for their disability, in-depth insight into complex issues, being resourceful, invoking lots of questions, and uninhibited in their expressions (Nielsen, 2002; Nielsen and Higgins, 2005). In the areas of challenge, twice-exceptional learners have uneven academic skill development, discrepant verbal and performance abilities, auditory and visual processing problems, a lack of organizational and study skills, low self-esteem, high levels of impulsivity and distractibility, poor social skills, deficits with long-term and short-term memory, and attention problems which are evident in areas of low interest (see Nielsen, 2002; Nielsen and Higgins, 2005).

Surprisingly, many twice-exceptional learners view themselves as primarily disabled because of the attention focused on things they cannot do well, which makes it difficult for them to believe they are bright, capable learners (Nielsen and Higgins, 2005). The continuous struggle between intellectual strengths and academic challenges has caused many of these youngsters to experience social

and emotional difficulties. Some develop low self-concepts (Swesson, 1994), have poor social skills, and experience difficulty fitting in with peers (Vespi and Yewchuk, 1992). They are easily frustrated and exhibit increased frustration from high expectations and standards for achievement that goes along with being gifted (Coleman, 2001). When faced with disappointments in the classroom, they may become disruptive or aggressive, argumentative, very opinionated, and inflexible (King, 2005; Nielsen and Higgins, 2005). When faced with difficult or challenging tasks, they are known to persevere when encouraged and supported (Vespi and Yewchuk, 1992). While it is evident that giftedness can, and does exist, across a broad spectrum of disabilities, only the distinguishing characteristics for students with learning disabilities, attention-deficit hyperactivity disorders (ADHDs), and emotional/behavioral disorders (EBDs) are discussed.

Gifted Students with Learning Disabilities

A student with an LD has a disorder in one or more of the basic psychological processes involved in understanding or in using spoken or written language. This disorder may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations, including conditions such as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. There are three types of gifted LD students—those with subtle learning disabilities who are less likely to be referred for special-education testing and who participate in gifted programs; those with severe LDs, who are rarely formally identified as gifted; and those who remain unidentified as either learning disabled or gifted because their disabilities conceal their gifts, and their gifts camouflage their disabilities (McCoach *et al.*, 2001).

Tannenbaum and Baldwin (1983) refer to gifted students, who also have LD, as paradoxical learners. These youngsters have high intellectual ability, but exhibit significant discrepancies in their level of performance in some academic areas (e.g., reading, math, spelling, and spoken or written language). This discrepancy, while not due to lack of educational opportunity or other health impairment, stems from differences in how the student's brain works and how it processes information. Although gifted students with LD exhibit patterns of behavior typical of other students who are gifted (e.g., high intellectual potential, excellent memory, and advanced vocabulary), they also exhibit characteristics common to students with ADHD such as distractibility, inattentiveness, inefficient learning strategies, and poor organization and study skills. Despite the high functioning level of gifted students with LD, most do not receive services for their gifts and talents because their giftedness masks their disability, or it provides them with the ability to compensate for their LD (Gargiulo, 2006). Unlike other gifted students, these

twice-exceptional learners have low self-knowledge and low self-esteem; and they often have conflicting thoughts concerning their capabilities and the expectations imposed upon them (Waldron, 1987). Therefore, it is not unusual for them to experience “frustration and have difficulty coping with the discrepancy between their giftedness and their learning disability” (King, 2005: 20).

Gifted Students with ADHD

ADHD is a common behavioral disorder of childhood. Children diagnosed with this disorder present a constellation of symptoms, including immaturity, impulsivity, inattention, and hyperactivity. Although gifted children with ADHD have intellectual characteristics similarly ascribed to gifted children without disabilities, little is known about their leadership ability, creativity, artistic talents, and musical abilities (Smith, 2004). What is known is that “many gifted children without ADHD exhibit symptoms similar to those seen in children with ADHD when they are bored or unchallenged” and that they “usually show variability in the quality of their performance on specific tasks, whereas gifted students are more consistent with their level of effort and performance . . .” (Reis and McCoach, 2002: 119).

Unlike their gifted counterparts, gifted students with ADHD differ in a number of ways. The most apparent being that of underachievement and high levels of non-conforming behaviors (e.g., inattention, impulsivity, distractibility, and high energy level), which mask their intellectual giftedness and make the probability of them being diagnosed as having giftedness and receiving services in a gifted program relatively low. It also creates the probability of them being diagnosed as either having ADHD or an LD. In terms of social and emotional maturity, these students lag behind their gifted peers. While some are able to compensate for their behaviors, others are seriously handicapped by them, thus adversely affecting their academic performance, causing challenges in their social interactions with others, and causing them to have low levels of self-concept or self-esteem. In addition, the low level of social and emotional maturity causes teachers to focus on their disruptive behaviors, subsequently creating situations that cause them to fail to see indicators of high ability.

Gifted Students with EBDs

Students with EBDs also exhibit behaviors that adversely affect their academic performance. They may display (1) an inability to learn that cannot be explained by intellectual, sensory, or health factors; (2) an inability to build or maintain satisfactory interpersonal relationships with peers and teachers; (3) inappropriate types of behavior or feelings under normal circumstances; (4) a general

pervasive mood of unhappiness or depression; and (5) a tendency to develop physical symptoms or fears associated with personal or school problems.

While some students with EBDs obtain measured intelligence in the gifted range of performance, they may not be referred or recommended for gifted programs; or they may be terminated from gifted programs because of their behavior (Reis and McCoach, 2002). The unpredictable and episodic behaviors and high dropout rates of gifted youngsters with EBDs are the defining features that differentiate them from the general population of students with gifts and talents. These behaviors usually result in inconsistent engagement and participation in learning opportunities, low levels of achievement, and a lack of opportunity for them to realize their potential. As the literature on gifted students with EBDs is sparse, more research is needed in this area to fully understand their diverse needs.

Educational Programming for Students Who Are Twice Exceptional

Academic and social needs of twice-exceptional learners can present a number of challenges for classroom teachers; and using teaching approaches not sanctioned by current research and evidence-based best practices can also present problems (Clark, 1997). According to the Individuals with Disabilities Education Act (IDEA), students with disabilities are entitled to a free and appropriate education in the least restrictive environment, which includes students with gifts and talents. However, in many instances, many twice-exceptional students are not receiving services that would allow for the full expression of their unique mixture of gifts, talents, and deficits. To ensure that dually diagnosed exceptional students develop their full potential, it will require the creation of appropriate curricula that addresses their gifts and talents as well as their disabilities. Thus, effective programming for the twice-exceptional learner requires collaboration between teachers of students with disabilities and teachers of children who are gifted (Clark, 1997). It also requires that teachers become aware of research and the current knowledge base that provide insight into strategies, interventions, and accommodations to meet the needs of these learners and that they gain access to enriched and accelerated instruction. Clearly, effective programming for twice-exceptional learners requires “a blending of instructional practices, such as cognitive training coupled with differential programming or curriculum compacting” (Gargiulo, 2006: 381). Earlier, Neu (2003) termed this form of instruction as a dually differentiated curriculum, and described it as a curriculum that “meets the needs of students who exhibit two contradictory sets of learning characteristics that creates “a balance

between nurturing the students’ strengths and compensating for their learning deficits” (p. 158).

Due to the unique characteristics of twice-exceptional learners, it is important that teachers and service providers provide them with appropriate curricular adaptations, accommodations, and interventions that respond to both their giftedness and their disabilities (Nielsen and Higgins, 2005). To succeed, these youngsters will need targeted services, programs, and interventions and a well-defined program model that ensures that services and interventions are tailored, well designed, integrated between special and gifted education, and consistent from year to year. The key elements that must be present in successful programs, according to Nielsen and Higgins are (1) implementation of an overarching program model, (2) use of interdisciplinary curricula, (3) intensive support for social, emotional, and behavioral needs, and (4) use of gifted education and special education strategies (p. 11). They also stated that the autonomous learner model (ALM) (Betts and Kerchner, 1999) and the Schoolwide Enrichment Model (SEM) (Renzulli and Reis, 1977) (p. 11) have been successful in maximizing the fullest potential of gifted students with special needs.

While there is no single best solution for meeting the needs of students with dual exceptionalities, Winebrenner (2003) noted that teachers and service providers should (1) teach students to appreciate, respect, and support individual differences; (2) understand that many students who have learning difficulties are global learners who prefer specific instructional formats for learning success; (3) when teaching content, always make sure that students see the big picture by introducing the concept first and the details second; (4) help students to become positively motivated by teaching them how to set realistic short-term goals and to take credit for teaching those goals, even if they represent the accomplishment of only part of the entire task; (5) teach students in ways that helps them to connect new learning with previously learned content or something they already know; (6) immerse all of the students’ sensory modalities in learning activities; (7) help students learn organizational skills by providing specific instruction in organizational techniques; (8) find and use all available technology to improve students’ productivity; and (9) allow students to take tests in diverse supervised environments that have the potential to enhance their concentration and performance. Moreover, effective teachers of twice-exceptional learners must understand how children learn. They must recognize that these children think differently from other children, and they must employ and use multifaceted and multidimensional approaches to assess learning. They must be knowledgeable of the subject matter they teach, the development of students, and the characteristics of learners. In addition, they must not adhere to a single theoretical perspective; but instead, they must adhere to both pedagogical frameworks in their teachings.

Clearly, effective teachers of twice-exceptional learners must be capable of transferring knowledge of subject matter and pedagogy into practice, not just focusing on students' disabilities but also on their abilities. They must use various technologies to support student learning as well as collaborate with teachers, parents, administrators, and other professionals. Effective teachers of twice-exceptional learners must develop personality traits that stimulate learning and understand that learning is a life-long process. These teachers must regularly monitor student learning, engage in ongoing assessments, and triangulate data from these assessments to seek and identify students' strengths and weaknesses. Moreover, they must be reflective practitioners who have a high sense of teaching efficacy.

Conclusion

The characteristics of students with gifts and talents are well documented in the literature. Among the common characteristics are (1) outstanding intellectual abilities, aptitude, and leadership; (2) extraordinary talents, creative abilities, and psychomotor skills; (3) abilities in the visual and performing arts; (4) advanced vocabularies, excellent memories, a great sense of humor, and vivid imagination; (5) highly verbal, quick learners, curious, and self-directed; (6) sensitivity to the needs of others; (7) good reasoning skills; (8) a remarkable ability to think logically, critically, and reflectively; and (9) high levels of independence, and a wide range of interests (Daniels, 2003, 2007). While we expect the occurrence of giftedness within the general population, it is often overlooked in the population of students with disabilities. Due to the complex nature in how giftedness masks or unmask itself among students with disabilities, educators need to look at these students holistically (in a variety of ways and in a variety of settings) and use a variety of approaches to provide opportunities for them to apply their strengths while ameliorating their deficits. For twice-exceptional children to maximize their potential, their intellectual strengths must be recognized and nurtured, and their disability must be simultaneously and appropriately accommodated using a dually differentiated curriculum that addresses both their disabilities and abilities to ensure that no child is truly left behind.

Bibliography

- Bianco, M. (2005). The effects of disability labels on special education and general education teachers' referrals for gifted programs. *Learning Disability Quarterly* **28**(4), 285–293.
- Cash, A. (1999). A profile of gifted individuals with autism: The twice-exceptional learner. *Roper Review* **22**(1), 22.
- Clark, B. (1997). *Growing up Gifted*, 5th edn. Upper Saddle River, NJ: Merrill Prentice-Hall.
- Coleman, M. R. (2001). Surviving or thriving? *Gifted Child Today* **24**(3), 56–64.
- Coleman, M. R. (2003). *The Identification of Children Who Are Gifted*. Arlington, VA: ERIC Clearinghouse on Disabilities and Gifted Education.
- Daniels, V. I. (2003). Students with gifts and talents. In Obiakor, F. E., Utley, C. A., and Rotatori, A. F. (eds.) *Effective Education for Learners with Exceptionalities: Advances in Special Education*, pp 325–348. Oxford, UK: Elsevier Science/JAI Press.
- Daniels, V. I. (2007). Working with multicultural learners with gifts and talents. In Obiakor, F. E. (ed.) *Multicultural Special Education: Culturally Responsive Teaching*, pp 110–125. Upper Saddle River, NJ: Pearson.
- Gallagher, J. J. and Gallagher, S. A. (1994). *Teaching the Gifted Child*. Needham Heights, MA: Allyn and Bacon.
- Gargiulo, R. M. (2006). *Special Education in Contemporary Society*, 2nd edn. Belmont, CA: Thomson Wadsworth.
- Grimm, J. (1998). The participation of gifted students with disabilities in gifted programs. *Roper Review* **20**(4), 285–286.
- Hardman, M. L., Drew, C. J., and Egan, M. W. (2008). *Human Exceptionality: School, Community, and Family*, 9th edn. Boston, MA: Houghton Mifflin.
- Individuals with Disabilities Education Act, Public Law 104–446.
- Karnes, F. A., Shaunessy, E., and Bisland, A. (2004). Gifted students with disabilities: Are we finding them? *Gifted Child Today* **27**(4), 16–21.
- King, E. W. (2005). Addressing the social and emotional needs of twice-exceptional students. *Teaching Exceptional Children* **38**(1), 16–20.
- Little, C. (2002). Which is it? Asperger's syndrome or giftedness? Defining the differences. *Gifted Child Today* **25**(1), 58–63.
- Lovett, B. J. and Lewandowski, L. J. (2006). Gifted students with learning disabilities: Who are they? *Journal of Learning Disabilities* **39**(6), 515–527.
- Maker, C. J. (1977). *Providing Programs for the Gifted Handicapped*. Reston, VA: Council for Exceptional Children.
- McCoach, D. B., Kehle, T. J., Bray, M. A., and Siegle, D. (2001). Best practices in the identification of gifted students with learning disabilities. *Psychology in the Schools* **38**(5), 403–411.
- Neihart, M. (2000). Children with Asperger's syndrome. *National Association for Gifted Children* **44**(4), 222–230.
- Neu, T. (2003). When the gifts are camouflaged by disability: Identifying and developing the talent in gifted students with disabilities. In Castellano, J. A. (ed.) *Special Populations in Gifted Education: Working with Diverse Gifted Learners*, pp 151–162. Boston, MA: Allyn and Bacon.
- Nielsen, M. E. (2002). Gifted students with learning disabilities: Recommendations for identification and programming. *Exceptionality* **10**(2), 93–111.
- Nielsen, M. E. and Higgins, L. D. (2005). The eye of the storm: Services and programs for twice-exceptional learners. *Teaching Exceptional Children* **38**(1), 8–15.
- Nielsen, M. E., Higgins, L. D., Hammond, A. E., and Williams, R. A. (1993). Gifted children with disabilities: The twice-exceptional child project. *Gifted Child Today* **16**(5), 9–12.
- Reis, S. M. and McCoach, D. B. (2002). Underachievement in gifted and talented students with special needs. *Exceptionality* **10**(2), 113–125.
- Smith, D. D. (2004). *Introduction to Special Education: Teaching in an Age of Opportunity*, 5th edn. Boston, MA: Allyn and Bacon.
- Swesson, K. (1994). Helping the gifted/learning disabled: Understanding the special needs of the 'twice-exceptional.' *Gifted Child Today* **17**(5), 24–26.
- Tannenbaum, A. J. and Baldwin, L. J. (1983). Giftedness and learning disability: A paradoxical combination. In Fox, L. H., Brody, L., and Tobin, D. (eds.) *Learning Disabled/Gifted Children: Identification and Programming*, pp 11–36. Baltimore, MD: University Park Press.
- Vespi, L. and Yewchuk, C. (1992). A phenomenological study of the social/emotional characteristics of gifted learning disabled children. *Journal for the Education of the Gifted* **16**(1), 55–72.

- Willard-Holt, C. (1999). *Dual Exceptionalities* (ERIC Document Reproduction Service ED 430 344). Reston, VA: ERIC Clearinghouse on Disabilities and Gifted Education.
- Winebrenner, S. (2003). Teaching strategies for twice-exceptional students. *Intervention in School and Clinic* **38**(3), 131–137.

Further Reading

- Assouline, S., Nicpon, M., and Doobay, A. (2009). Profoundly gifted girls and autism spectrum disorder. *Gifted Child Quarterly* **53**(2), 89–105.
- Bianco, M., Carothers, D., and Smiley, L. (2009). Gifted students with Asperger syndrome: Strategies for strength-based programming. *Intervention in School and Clinic* **44**(4), 206–215.
- Coleman, M. R., Harradine, C., and King, E. W. (2005). Meeting the needs of students who are twice exceptional. *Teaching Exceptional Children* **38**(1), 5–6.
- Daniels, V. I. (1998). Minority students in gifted and special education: A case for educational equity. *Journal of Special Education* **32**(1), 41–43.
- Krochak, L. A. and Ryan, T. G. (2007). The challenge of identifying gifted/learning disabled students. *International Journal of Special Education* **22**(3), 44–53.
- Moon, T., Brighton, C., Callahan, C., and Jarvis, J. (2008). Twice-exceptional students: Being gifted and learning disabled – implications of the IDEIA. In Grigorenko, E. L. (ed.) *Educating Individuals with Disabilities: IDEIA 2004 and beyond*, pp 295–317. New York: Springer.
- Rizza, M. and Morrison, W. (2002). Uncovering stereotypes and identifying characteristics of gifted students and students with emotional/behavioral disabilities. *Roeper Review* **25**(2), 73.
- Ruban, L. and Reis, S. (2005). Identification and assessment of gifted students with learning disabilities. *Theory into Practice* **44**(2), 115–124.
- Sanfran, S. P. (2001). Asperger syndrome: The emerging challenge to special education. *Exceptional Children* **67**(2), 151–160.
- Seeley, K. (2004). Gifted and talented students at risk. *Focus on Exceptional Children* **37**(4), 1–8.
- Webb, J. T., Amend, E. R., Webb, N. E., et al. (2005). *Misdiagnosis and Dual Diagnosis of Gifted Children and Adults: ADHD, Bipolar, OCD,*

- Aspergers, Depression, and other Disorders*. Scottsdale, AZ: Great Potential Press.
- Weinfeld, R., Barnes-Robinson, L., Jeweler, S., and Shevitz, B. (2005). What we have learned: Experiences in providing adaptations and accommodations for gifted and talented students with learning disabilities. *Teaching Exceptional Children* **38**(1), 48–52.

Relevant Websites

- <http://www.apa.org> – American Psychological Association, Center for Gifted Education Policy.
- <http://www.jhu.edu> – Center for Talented Youth at Johns Hopkins University.
- <http://www.ditd.org> – Davidson Institute for Talent Development.
- <http://www.ericdigests.org> – Eric Digests.
- <http://www.gifteddevelopment.com> – Gifted Development Center.
- <http://www.gt-cybersource.org> – GT-Cybersource.
- <http://www.hoagiesgifted.org> – Hoagies Gifted Education Page.
- <http://www.educationaladvancement.org> – Institute for Educational Advancement.
- <http://www.uniquelygifted.org> – Internet Resources for Gifted/Special Needs Children.
- <http://cty.jhu.edu> – Johns Hopkins University Center for Talented Youth (CTY).
- <http://www.nagc.org> – National Association for Gifted Children.
- <http://www.nfgcc.org> – National Foundation for Gifted and Creative Children.
- <http://www.nsgt.org> – National Society for the Gifted and Talented.
- <http://www.gifted.uconn.edu> – Neag Centre for Gifted Education, National Research Center on the Gifted and Talented (NRC/GT).
- <http://www.sedoparking.com> – Sedo Domain Parking, National Talent Network (NTN) at the Educational Information and Resource Center's (EIRC).
- <http://www.sengifted.org> – Supporting the Emotional Needs of the Gifted.
- <http://www.cec.sped.org> – The Council for Exceptional Children, The Association for the Gifted and Talented (CEC-TAG).
- <http://www.twicegifted.net> – Twice Gifted.
- <http://world-gifted.org> – World Council for Gifted and Talented Children.

Early Childhood Special Education: Birth to Eight Years Old

T S Lovelace, Duquesne University, Pittsburgh, PA, USA

R Gardner III, Ohio State University, Columbus, OH, USA

J B Axe, Simmons College, Boston, MA, USA

D M Sainato, Ohio State University, Columbus, OH, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cultural sensitivity – The ability to work with and provide effective intervention for children and families from a variety of cultural backgrounds (i.e., racial, ethnic, gender, sexuality, ability, and religious).

Early childhood special education – Refers to educational interventions for children ages 3–8.

Early intervention – Refers to educational interventions for children birth to age 3 and their families.

Empirically-validated interventions – Interventions that are derived from careful, direct observation and experimentation.

IFSP – Or the Individualized Family Service Plan is a legal document that outlines service delivery measures for the young child and their family.

Target behavior – The behavior designated for change or intervention.

This article provides an overview of early-childhood special education (ECSE). A definition of early-childhood special education is offered with regard to current practice and special education law. Second, we present information with regard to the children and families who are served by ECSE professionals. Third, we note the type and location of ECSE services (including transition services) and provide characteristics of effective programs. Finally, we provide suggestions for the future both in terms of professional practice and preparing ECSE professionals.

Birth to kindergarten is a critical period for all children. Most children learn a remarkable amount from the time of their birth until they begin kindergarten. The development of young children is typically monitored across six domains; fine motor, gross motor, self-help/adaptive, communication, cognition, and social/emotional development abilities (Raver, 2009). Children experiencing delays in maturation across these domains may find themselves at increased risk of being diagnosed with a disability and school failure. Since one of the goals of ECSE professionals is to position children for school success, it is important to identify markers essential for young children's

school achievement. Children best prepared for school success are those who have developed age-appropriate language and social skills. For example, children who begin kindergarten with greater language skills are better positioned for academic success (Hart and Risley, 1995). In comparison, those children who enter school with significantly lower levels of language and/or deficits in social skills are at increased risk for school failure and possible special education placement (Adams, 1990; Walker and Severson, 2002).

Keogh *et al.* (2004) reported, in their 20-year follow-up study, that 3-year-old children identified as having developmental delays often continued to experience problems even into adulthood. Helping children to develop cognitive and social abilities during their early years is essential to their later success in school and beyond. In fact, a foundational belief of ECSE professionals is that early intervention is more beneficial to children and families than remedial interventions (Hooper and Umansky, 2009; Wolery and Bailey, 2002). For children with – or who are at risk for – disabilities, the developmental process that happens during their first 8 years of life is even more critical to their future success or failure than it is for their typically developing peers (Raver, 2009). Therefore, it is important to promote interventions and programs that support families in helping children with developmental delays to develop language, social, physical, and academic skills for the purpose of improving their ability to achieve success in their schools and communities (Sandall *et al.*, 2005).

Early Intervention/Early Childhood Special Education (EI/ECSE) services are typically available for children with special needs from birth to age 8. This period involves multiple developmental stages (as the child progresses from infant to school-age child) and environment changes (e.g., home–community, home–preschool, preschool–kindergarten, kindergarten–early elementary school). Young children must be assessed carefully with sensitivity to determine their abilities, needs, and environmental influences (including culture) (Linder, 2008). Beyond assessments, EI/ECSE professionals must also develop cogent interventions that are individualized for the children and families they serve. The quality of the early-childhood services has a direct impact on children's outcomes (Bowman *et al.*, 2000). Professionals must be careful in matching children, families, and EI/ECSE services. What is an

appropriate intervention for an infant and family is not necessarily appropriate for a preschooler (Wolery and Bailey, 2002). This sensitivity to the differing requirements for children and families related to age is one reason ECSE has evolved into two distinct service categories.

What Is Early Intervention/Early-Childhood Special Education (EI/ECSE)?

The two categories that provide services to young children and their families are early intervention and ECSE. The term early intervention is generally used to define services provided to children from birth to age 3 and their families (Part C of Individuals with Disabilities Education Act (IDEA)) (Raver, 2009). Thus, the focus of early intervention is family-centered with an emphasis on providing services in natural environments (Bruder, 2001; Wolery and Bailey, 2002). Early intervention seeks to address the special needs of infants and toddlers who have disabilities, developmental delays, or are at risk for disabilities (Bowe, 2000). It provides support to families including assisting them in understanding medical and educational information pertinent to their child and their child's special needs (Cartledge *et al.*, 2009). In addition, it should (1) be individualized for each child and family, (2) be systematic and regularly assessed for effectiveness, and (3) provide for a smooth transition to the child's next developmental stage, which may involve full classroom inclusion. In other words, effective early intervention should facilitate a seamless transition for the child and family to ECSE (Odom and Wolery, 2003).

ECSE usually refers to educational interventions for exceptional children 3–8 years of age (Part B of IDEA) (Bowe, 2000). Young children are eligible for ECSE services based on developmental delays or lack of development in communication, cognition, social/emotional behaviors, physical skills, or adaptive behaviors; that is, there is an increased focus on those children who demonstrate delays rather than those who are only at risk for delays. Similar to early intervention, parents are fully included in the development and implementation of interventions. It is during this developmental period that children often make the transition to school environments. Odom and Wolery (2003) recommended a unified theory of practice for professionals working with young children merging EI and ECSE. EI/ECSE involves the implementation of purposeful and systematic practices that are based on professional evidence, when available (Odom and Wolery, 2003). An interdisciplinary service perspective is helpful so that a seamless system is in place for exceptional children and their families across the developmental years of birth to age 8 (Bowe, 2000; Hooper and Umansky, 2009).

Services for Children

In order to address the needs of infants, toddlers, and families, each state has a system of EI/ECSE services designed to meet the special needs of children with disabilities, developmental delays, or who are considered at risk (Odom and Wolery, 2003; Hebbeler *et al.*, 2007). The goal of EI/ECSE services is to prevent or lessen the effects of disabilities or developmental delays on the lives of children and their families; that is, through careful planning and use of empirically validated interventions, children with or at risk for disabilities can develop skills that will promote success in their homes and beyond. The definition of EI/ECSE is not only limited to descriptions of the recipients of services and the services provided, but also involves the impact of those services on children and their families. In other words, intervention outcomes for children and families are an essential consideration for EI/ECSE professionals (Bruder, 2001).

The reality is that the effectiveness of ECSE interventions is at least partially defined by the ability to minimize or prevent the impact of disabilities on young children and the ability to assist/support families. This characteristic of EI/ECSE interventions making a positive difference in the lives of child is not always easily captured in quantifiable ways in applied settings. For example, it is not always possible to determine if an intervention was the sole cause for a child, previously considered at risk, who is now demonstrating typical developmental behaviors (i.e., prevent the manifestation of a disability). Possibly, the child's positive behavior is primarily a result of physical maturation. Similarly, it is not a simple matter to determine how well an intervention has diminished the potential negative impact of identified disabilities or developmental delays. The fact that a positive outcome is the most important variable in EI/ECSE. Effective interventions build on and enhance the natural physical and mental maturation processes that occur in children. Most EI/ECSE professionals reject the premise of simply waiting and allowing children with special needs to develop without intervention as a viable option. The early years of children's development are just too critical to take the risk that improvement will occur on its own (i.e., as a result of time and natural maturation).

There is a firm belief among professionals that EI/ECSE is needed and required for young children with special needs and their families to achieve positive outcomes. It is generally believed by early-childhood specialists that the earlier the intervention begins, the more positive the outcome for children (Raver, 2009). Therefore, the initial timing of the intervention is important. Not only is it important that interventions begin as early as possible in the lives of children at risk for or with disabilities, but that interventions have an empirical foundation and are designed to meet the unique needs of

each child (Odom and Wolery, 2003). Further, it is essential that interventions be closely monitored to determine the effects of the intervention on each child. If the period from birth to age 8 is critical, then every day without intervention or without effective intervention is not just unhelpful but potentially detrimental. The best way to determine the effectiveness of an intervention is by the careful assessment of important target behaviors across time. Clearly, ECSE provides a wide range of services designed to assist infants and young children with disabilities, developmental delays, or who are considered at risk to maximize their potential. Further, ECSE provides services designed to support and equip families of children with special needs in order to promote successful transitions into schools and beyond. Finally, the belief in early intervention is a strongly held conviction by EI/ECSE professionals (Guralnick, 2001; Hooper and Umansky, 2009). This belief in early intervention is not simply a concept held only by EI/ECSE professionals and the families of special needs children. Our society has also embraced the necessity for early intervention and laws have subsequently been enacted to ensure and protect the rights of children and families to EI/ECSE services.

Early Education Laws

Over 35 years ago, the first landmark legislation that impacted children with special needs and their families – the Rehabilitation Act (PL 93-112) – was signed into law. Passed in 1973, the Rehabilitation Act prohibited discrimination on the basis of disabilities by programs receiving federal funds. This prohibition includes programs serving young children (e.g., Head Start Programs). Section 504 of the Rehabilitation Act allowed children with disabilities or who are at risk for academic failure to receive specialized education services. This law was important in setting a precedent for later special education legislation. In 1975, the Education of All Handicapped Children's Act (retroactively renamed Individuals with Disabilities Education Act (IDEA)) was enacted to acknowledge the right of all children in the age range 3–21 to a “free and appropriate education, the right for parental participation in educational decision making, individualized educational plan, multifaceted evaluation, and placement in the least restrictive educational environment.” Initially, this law focused on the educational rights of school-age children. The subsequent passing of PL 99-457, in 1986, both amended and expanded PL 94-142. PL 99-457 focused on the rights of infants and preschoolers to services that could enhance their future opportunities. The purpose of Congress in enacting PL 99-457 was:

- To enhance the development of infants and toddlers with disabilities and to minimize their potential for delay;

- To reduce educational costs by minimizing the need for special education and related services after handicapped infants and toddlers reach school age;
- To minimize the likelihood of institutionalization of handicapped individuals;
- To enhance the capacity of families to meet the special needs of infants and toddlers. (20 U.S.C. Sec. 1471)

The reauthorization of IDEA, in 2004, continued the affirmation of the need for early intervention from birth to age 8 (Raver, 2009). **Table 1** provides information with regard to the minimum components required under IDEA for statewide early-intervention services. As discussed earlier, Part C of IDEA or the Program for Infants and Toddlers with Disabilities is a federal grant program that provides funding for states to develop and operate a comprehensive statewide program of early-intervention services for infants and toddlers and toddlers with disabilities. As it stands, the reauthorization of IDEA, in 2004, included – for the first time – the concept of response to intervention (RTI; Burns and Coolong-Chaffin, 2006). RTI is based on a students' ability to benefit from empirically validated instruction (McLoughlin and Lewis, 2008). The inclusion of RTI in IDEA has increased interest in school accountability (Ysseldyke *et al.*, 2006). Response to intervention allows professionals to use students' responses to scientifically based interventions to diagnose and determine instructional strategies for those with special needs (Burns and Coolong-Chaffin, 2006). Response to intervention was originally intended for children with learning disabilities or those at risk for learning disabilities; however, it has been increasingly adapted to children with a variety of special needs (Buysse and Peisner-Feinber, 2008). There are numerous models of RTI, but they all seem to involve multiple tiers (McLoughlin and Lewis, 2008). Tier 1 involves the core curriculum in general education classrooms where all children are provided empirically validated instruction. Tier 2 provides specialized and remediation instruction for students who are not making adequate progress within the core curriculum. Tier 3 involves an intensive systematic intervention for students who are experiencing significant difficulty compared to their peers. The multiple-tiers model theoretically allows not only for the identification of students who require additional instruction, but also assists professionals in targeting how to efficiently use instructional resources (Burns and VanDerHeyden, 2006).

At present, preschool professionals are beginning to develop RTI models (Buysse and Peisner-Feinber, 2008). One Pre-K model – termed Recognition and Response (R&R) – is similar to RTI in that it includes a core curriculum that utilizes research-based instruction, interventions that are targeted to the individual needs of children who meet screening criteria, and universal screening and progress monitoring (Buysse and Peisner-Feinber, 2008).

Table 1 Minimum components under IDEA for a statewide, comprehensive system of early intervention services to infants and toddlers with special needs (Including American Indian and Homeless Infants and Toddlers)

1. A rigorous definition of the term 'developmental delay'
 2. Appropriate early-intervention services based on scientifically based research, to the extent practicable, are available to all infants and toddlers with disabilities and their families, including Indian and homeless infants and toddlers
 3. Timely and comprehensive multidisciplinary evaluation of needs of children and family-directed identification of the needs of each family
 4. Individualized family service plan and service coordination
 5. Comprehensive child-find and referral system
 6. Public awareness program including the preparation and dissemination of information to be given to parents, and disseminating such information to parents
 7. Central directory of services, resources, and research and demonstration projects
 8. Comprehensive system of personnel development, including the training of paraprofessionals and the training of primary referral sources
 9. Policies and procedures to ensure that personnel are appropriately and adequately prepared and trained
 10. Single line of authority in a lead agency designated or established by the governor for carrying out:
 - a. General administration and supervision
 - b. Identification and coordination of all available resources
 - c. Assignment of financial responsibility to the appropriate agencies
 - d. Development of procedures to ensure that services are provided in a timely manner pending resolution of any disputes
 - e. Resolution of intra- and interagency disputes
 - f. Development of formal interagency agreements
 11. Policy pertaining to contracting or otherwise arranging for services
 12. Procedure for securing timely reimbursement of funds
 13. Procedural safeguards
 14. System for compiling data on the early-intervention system
 15. State interagency coordinating council
 16. Policies and procedures to ensure that to the maximum extent appropriate, early-intervention services are provided in natural environments except when early intervention cannot be achieved satisfactorily in a natural environment
- Retrieved from the National Early Childhood Technical Assistance Center (June 6, 2008). <http://www.nectac.org/partc/componen.asp?text=1>

Source: The National Early Childhood Technical Assistance Center (NECTAC) [April 4, 2005]. Minimum components under IDEA for statewide, comprehensive system of early intervention services to infants and toddlers with special needs (including American Indian and homeless infants and toddlers). Retrieved June 6, 2008 from <http://www.nectac.org/partc/componen.asp>

The R&R model is a data-based model that allows professionals to make instructional decisions as seen in RTI models and is primarily designed for teaching reading and math. The Institute of Education Sciences has recently awarded a 10 million grant to University of Kansas to establish a national center for RTI for early childhood. For a more in-depth discussion of RTI in Pre-K, please see the special issue (vol. 35, no. 4) on early-childhood tier models in *School Psychology Review* (2006).

Who Receives ECSE Services?

Children and families come into contact with EI/ECSE professionals in a variety of ways. Some families will know before their child's birth that he/she has a disability. In this case, the physician or hospital often initiates contact with EI/ECSE professionals (Hebbeler *et al.*, 2007). Other children's need for early intervention becomes evident as they grow. In this case, a child's special needs may be first noticed by a physician, parent, child-care worker, pre-school teacher, or a family friend (Bowe, 2000). After a special need is suspected, each child is assessed and

evaluated to determine if there are disabilities or developmental delays that warrant intervention (Bowe, 2000; Hebbeler *et al.*, 2007). Hemmeter *et al.* (2005) recommended five professional guidelines when assessing young children: (1) actively solicit family and child's concerns and interests, (2) individualize assessments to address the unique needs of each child, (3) report assessment outcomes in a way that allows for the development of intervention objectives and permits parents to clearly understand the results, (4) use assessment materials that allow the child's authentic behaviors to be demonstrated during routine activities, and (5) follow all legal guidelines. Further, when assessing children, professionals should include multiple assessment sources in order to gain the most comprehensive information with regard to the needs of children (Neisworth and Bagnato, 2005).

Once a child is identified as needing EI/ECSE services, an individualized family service plan (IFSP) is created (Turbiville *et al.*, 1996). An essential component of any intervention for young children is the IFSP, which is integral in designing programs to address the unique needs of each child and family. The IFSP addresses the child's present level of performance in regard to physical,

motor, sensory, cognitive, communication, social-emotional, or adapted behaviors. This evaluation determines if the child is experiencing disabilities or developmental delays in one or more of the six domains. The IFSP involves a collaborative effort between families and EI/ECSE professionals to design and implement early interventions. The IFSP team must not only determine what services are to be provided but also where the services are to be provided (i.e., home, school, and community) (Bruder, 2001). IDEA advocates the delivery of services in natural environments (i.e., environments typical for children without special needs) whenever appropriate.

The family and team of professionals design an IFSP to assist the child in maximizing his/her ability to develop typical behavioral patterns based on professional knowledge and family desires. The selection of interventions is not only impacted by the type and degree of severity of the developmental delay but also by the family's circumstances (Hebbeler *et al.*, 2007); that is, some families may need more support for early-intervention services to be effective.

Families of children in need of EI/ECSE services are as diverse as the general population. Similar to the school-age population with disabilities, there are concerns of disproportionality with regard to who receives services. Boys tend to outnumber girls in early intervention. Approximately 61% of the children receiving early-intervention services are boys (Hebbeler *et al.*, 2007). Children from low-socioeconomic-status homes are also disproportionately represented among those receiving early-intervention services (Bowe, 2000). Ethnically diverse children receiving early intervention are more likely to be from African American families than white families (Hebbeler *et al.*, 2007). It is imperative that professionals are culturally sensitive to the unique needs of children and families. Hebbeler *et al.* (2007) reported that families overwhelmingly indicate that at least their first encounter with EI/ECSE professionals is positive. **Table 2** provides demographic information about children and their families who receive EI/ECSE services.

Service Delivery

Service delivery for exceptional children has always been characterized as meeting children and families at their point of need; this commitment does not waver with those that are the youngest members of the population. ECSE professionals have expanded their efforts over the past decade to include the needs of culturally and linguistically diverse families with various resources; ensuring that an individualized program of services is retained for all children (Bowe, 2000; Hooper and Umansky, 2009). As stated previously, public law mandates services to address the needs of young children with disabilities.

Table 2 Child's ethnicity, mother's education level, and household income for families of children receiving EI services and for the general population

Characteristic	EI Services Population, Percentage	General Population, Percentage
Child's race/ethnicity		
White	53	61
African American	21	14
Hispanic	16	18
Asian or Pacific Islander	4	2
American Indian or Alaska Native	1	1
Mixed or other	5	4
Mother's education level		
Less than high school degree	16	17
GED or high school degree	32	27
Some college	28	28
Bachelor degree or higher	24	27
Household income		
\$15 000 or less	27	21
\$15 001–\$25 000	16	16
\$25 001–\$50 000	29	31
\$50 001–\$75 000	16	16
Over \$75 000	13	16

General population data from National Household Education Survey (1999) for children up to age 3.

In all subsequent analyses, American Indians are included in the mixed race or other category.

* = $p < .05$, ** = $p < .01$

Source: Hebbeler, K., Spiker, D. Bailey, D., Scarborough, A. Mallik, S., Simeonsson, R., Singer, M., and Nelson, L. (January 2007). Early Intervention for infants and toddlers with disabilities and their families: Participants, services, and outcomes. Final report of the National Early Intervention Longitudinal Study (NEILS). Table 2–2 (p. 2–7). Menio Park, CA: SRI International. Retrieved from http://www.sri.com/neils/pdfs/Neils_Final_Report_02_07.pdf

The challenge is to make sure that appropriate resources are available to meet the unique needs of each child. This challenge can be exacerbated by the fact that early-intervention services are not available in all public schools. Additionally, eligibility and concerns about best practice and empirical evidence often dictate what services should and can be accessed by families (Bailey and Simeonsson, 1988). The task of identifying appropriate services for children is daunting for professionals and is even more difficult for families, who may be overwhelmed and struggling to accept the challenges their child faces. Professionals need to help families through the maze of regulations and services so that valuable time is not lost as families try to determine best services for their children.

Not only it is critical to identify types of EI/ECSE services needed for each child, but is also critical to identify the environment where the services are to be delivered.

Programming can occur in day-care centers, children's homes, schools, or local community agencies. For children whose health needs dictate more careful care, hospital-based programs for high-risk infant care are among the choices for families (Hooper and Umansky, 2009; Peterson *et al.*, 1994). Hospital-based programs have similar advantages to center-based programs with the ready availability of therapists and medical personnel to provide a wide range of services (Crocker and Porter, 2001; Peterson *et al.*, 1994). The variety of services makes choosing an effective option for service delivery a difficult early decision. Early-intervention services are often categorized as home-based services, center-based services, and combination programs. The diverse needs of families cannot only be met with these options, but the varieties of approaches showcase the commitment that EI/ECSE service providers and teachers have for this population. When choosing the type of service delivery method, it is important to remember the focus of each method of service delivery and how what is offered will impact children and families that access those services. **Table 3** provides information about service locations for children receiving services during their first 6 months.

Before discussing differences between service-delivery methods, it is important to look at the history of service delivery for this population and the basis for providing services to this age group. Odom and Wolery (2003) suggested that service delivery for this population has followed societal values, and, therefore, service should be based on evidence-based practice, systematic observation, and decision-making priorities for early-childhood professionals and families. **Table 4** provides information

Table 3 Locations of early intervention services received during the first 6 months after the initial IFSP as reported by service providers

	Percent
In the family's home	78
In a family day care/preschool/nursery school	10
In a specialized center-based early-intervention program	28
In a clinic or office (e.g., hospital-based clinic, therapist office)	29
Another setting (e.g., inpatient services in a hospital)	5

Percentages sum to more than 100% because children and families could receive more than one service.

Percentages exclude 2.3% of children and families who received no services in the first 6 months after signing the IFSP.

N = 2651.

From National Early Intervention Longitudinal Study.

Source: IDEA (2001). 23rd Annual report to Congress on the Implementation of Individuals with Disabilities Education Act (IDEA). Section III. Programs and Services received by children and families early intervention. Table III-14 (p. III-51). <http://www.ed.gov/about/reports/annual/osep/2001/section-iii.pdf>

about the types of services available for children and families. Parents may choose to participate in a variety of options to meet the needs of their exceptional children. As stated above, service delivery can be in the form of home-based options, center-based options, or inclusion options found in the general education arena.

Home-Based Options

Providing services in the homes has a long history in our society (Bailey and Simeonsson, 1988). The recent growth of home-based options has been attributed to the emergence of community-service agencies that focus on initiatives in child abuse, teenage pregnancy, children at-risk or disabled, as well as maternal and child health (Bailey and Simeonsson, 1988). Home-based options are based upon the belief that the most natural location for young

Table 4 Early-intervention services received by children and families during the first 6 months, as reported by service providers

Service	Percent
Assistive technology	4
Audiology	14
Behavior management services	6
Developmental monitoring	38
Family counseling/mental health counseling	4
Family training	20
Other family support	10
Genetic counseling/evaluation	3
Health services	7
Medical diagnosis/evaluation	11
Nursing services	7
Nutrition services	7
Occupational therapy	39
Physical therapy	38
Psychological or psychiatric services	4
Respite services	4
Service coordination	80
Social work services	12
Special instruction for the child	44
Speech/language therapy	53
Translation services (interpreter)	2
Transportation and/or related costs	7
Vision services	6
Other	2
Assistive technology	4

Percentages sum to more than 100% because children and families could receive more than one service.

Percentages exclude 2.3% of children and families who received no services in the first 6 months after signing the IFSP.

N = 2651.

From National Early Intervention Longitudinal Study.

Source: IDEA (2001). 23rd Annual report to Congress on the Implementation of the Individuals with Disabilities Education Act (IDEA). Section III. Programs and Services received by children and families early intervention. Table III-13 (p. III-49). <http://www.ed.gov/about/reports/annual/osep/2001/section-iii.pdf>

children with special needs, such as infants and toddlers, is the home (Gargiulo and Kilgo, 2000). Bailey and Simeonsson (1988) defined home-based services broadly – as an approach to early intervention in which services are delivered in the home. Hesitant to define this option more specifically, they stated that the diversity, philosophical orientations, and goals involved in early intervention prevent a precise definition. Others define home-based services environmentally, such as where the natural environment occurs for those needing services. The natural environment affords the opportunities for caregivers, parents, and service providers to work in the same setting and meet as an interdisciplinary team (Bowe, 2000; Gargiulo and Kilgo, 2000). The term natural environment was added to Part C and can be thought of as the place where the principle of the least restrictive environment can be enacted (see Bowe, 2000).

The definition of natural environment is under constant explanation and revision by the federal government (Noonan and McCormick, 2006). In Part C, it is referred to as inclusive environments for young children, birth to age 3. Currently, it is defined as settings that are natural or that typical young children utilize (Hooper and Umansky, 2009; Noonan and McCormick, 2006). Noonan and McCormick defined natural environments as those that provide children with access to activities generally found in the typical environment (e.g., the sandbox at playgrounds, story-time at the library, children's gyms). The decision to choose home-based services most often depends on two ideals: the level of parent/caregiver involvement and the possibility of effective program implementation (Bailey and Simeonsson, 1988). The home allows service providers access to pertinent information with regard to environmental stimuli, routines, and family structure that are part of the child's regular environment. The natural environment has proven to be a viable location for successful intervention. For example, Chiang (2008) found that children with autism tended to respond more frequently on communication tasks coded as high on the spontaneity continuum in naturalistic settings (e.g., natural cues and little prompting needed). Accessing home-based services can vary by community. Some communities choose to emphasize home-based services while others prefer to consolidate resources in a center. In the past, local governments have provided services through mental health, human resources, boards of mental retardation, and child-care and preschool programs (Bailey and Simeonsson, 1988).

Center-Based Programs

Center-based programs are those places away from the homes of children with special needs that provide intervention services (Gargiulo and Kilgo, 2004). A common center-based option is the child-care setting – in which

children with and without disabilities are enrolled. Center-based care can also extend to hospital-affiliated programs, preschools, and local programs serving populations diagnosed with mental retardation and/or developmental disabilities (MRDD). Characteristics of center-based programs are increased social interaction with peers, availability of various service providers in one setting, and parent support (see Gargiulo and Kilgo, 2004). Program make-up can vary; some programs operate several hours a week and some are more similar to traditional child-care settings that offer care consistent with the needs of working families (Heward, 2008).

Combination Programs

An alternative to choosing either a center- or home-based program is a combination program. Combination programs seek to combine the advantages of both of the previously discussed options, in addressing the needs of children with or at risk for disabilities. These programs can be characterized by a young child attending school for a few days a week while also receiving services through home visits from service providers. EI/ECSE has a variety of well-known programs that may fit in one or more of these categories (e.g., Head Start). While not thought of as a program that serves the needs of children with special needs, the federal Head Start program has served as an option for children with special needs for decades. Begun in 1965, Head Start is touted as the longest-running, most successful school-readiness program. In Fiscal Year 2007, 49,400 classrooms served just under 1 million children, with more than 25 million children having benefited from its services since its inception (Administration for Children and Families ACF, 2007). Early Head Start, the infant/toddler portion of the NHSA, provides programming to children from birth to age 3 and pregnant women. NHSA is committed to serving children who are at-risk for disabilities and children with identified disabilities by reserving 10% of enrolment in its programs for them (US Department of Health and Human Services, 2008; Wall *et al.*, 2005).

Home-based services offer a variety of advantages for families. Bailey and Simeonsson (1988) organized advantages and disadvantages into four useful categories of outcome: child development, child behavior, family functioning, and child health. Additionally, it is helpful to add a fifth category – operational concerns – which cover the monetary and logistical issues families may face. Utilizing center-based services or another natural environment does not preclude effective intervention, but utilizing the home as a focal point offers service providers and families with important (otherwise unavailable) information about child development and behavior. Clearly, early-intervention services are sought by parents as a way to maintain and augment child development as well as the

behavior and health of their exceptional child. Developmental concerns (i.e., social, motor, and cognitive) can respond positively to intervention in the home, but there may be more advantages to accessing services in the community (Cook *et al.*, 2004). Center-based programs involve trained teachers and support personnel who are available to facilitate the learning of exceptional children (Heward, 2008). Parents may also desire center-based options because of the availability of specialized therapists (see Bailey and Simeonsson, 1988). A concern with intervention is generalization to other environments and a disadvantage to center-based services may be the generalization of skills from classroom to home, and vice versa (see Bailey and Simeonsson, 1988). An example of this may be the ability of a child to have similar social and behavioral demands (such as participation or conversational skills) in both environments (see Bailey and Simeonsson, 1988).

Family Functioning

Services provided in the natural environment for children promote family members involvement that can be significant in the interventional process. It places the child in a setting that is familiar and allows instruction by the child's first teacher and closest supporters like parents and daily caregivers (Gargiulo and Kilgo, 2000). It is this advantage – the use of the parent as the primary teacher – that may also be a disadvantage concerning home-based services. Parents may not desire to be the primary provider of service for their children or believe they do not possess the skills necessary to aid in the effective development of their children. The availability, skill level, and commitment of the parent to intervene can have an immediate and pervasive impact on the effectiveness of service delivery.

Advantages in selecting home-based service delivery extend to less cost to families in terms of transportation to access services, tuition for center-based services, and finding services that may be more responsive to the particular needs of the family (see Gargiulo and Kilgo, 2000). These cost savings can also be found in programming. As Heward (2008) noted, home-learning activities and materials can be more natural and appropriate for young exceptional children. This may be difficult for families as they are placed in a role that they neither feel comfortable with nor wish to assume (Gargiulo and Kilgo, 2000). The raising number of single-parent homes also affects the willingness and appropriateness of parents to act as sole caregivers (Heward, 2008). Researchers have begun to analyze variables that may enhance home-based interventions (e.g., parental engagement; Wagner *et al.*, 2003).

Advantages of center-based options are that they allow for the development of social skills and opportunities for

social interaction between typical and exceptional peers. Center-based options often allow for a family to access several service providers at one site. This multidisciplinary approach allows for team members to work together in a common environment to provide the most effective intervention for the exceptional child. Parents are still part of the intervention team and their participation is needed and beneficial to their children's development. In addition, center-based options are preferred because of the availability of specialized equipment, exposure to various ability levels that support learning, and the ability for parents to access support services for themselves through networking with other parents (Gargiulo and Kilgo, 2000). On the other hand, center-based programs are limiting for families because the cost of transportation to and from the center can be costly in expenses and time. For centers, maintenance of the center and equipment is often costly. Although centers provide for a social network for parents with other parents and service providers, in comparison to home-based programming, the depth of those relationships may be shallow (Gargiulo and Kilgo, 2000).

Personnel Preparation

Preparing EI/ECSE professionals is critical to the effectiveness of the field (Miller and Stayton, 2005). EI/ECSE as a field traverses the benefits of early-childhood initiatives and the strengths of special education. Those working in the field of special education may or may not be required to be licensed in special education, early-childhood education, pre-kindergarten education, or a combination of the above. Irrespective of the license requirements, EI/ECSE professionals need knowledge concerning both general education and ECSE (Hyson, 2003). The training of early-childhood special educators focuses on areas of child development and provides specialized knowledge about children with disabilities and their families (Winton and McCollum, 1997). Typically, EI/ECSE professionals' responsibility goes beyond the classroom as they serve on multidisciplinary teams during IFSP meetings (Bowe, 2004). Finally, ECSEs serve as advocates or social service liaisons by being key professionals working with families (see Bowe, 2004). The Council for Exceptional Children, Division of Early Childhood, recommends five guidelines in preparing EI/ECSE professionals (Miller and Stayton, 2005):

- Greater participation of families in the planning and delivery of training activities;
- An increase in the crossing of discipline boundaries to access appropriate preservice and inservice education and training;
- Greater emphasis on preservice and inservice educational opportunities that include students from both

early childhood and early childhood special education together;

- Increasing interest in interdisciplinary collaboration preparation practices;
- Support of family-centered preparation practices. (pp. 190, 191)

The aforementioned guidelines affirm the field's commitment to family-centered services and bring increased focus on the need for collaboration by all professionals who work with children with disabilities. Professional guidelines make it clear that inclusive services are vital for serving young children (Odom and McLean, 1996). The effectiveness of inclusive sites for interventions is directly impacted by the preparations of the professionals across essential disciplines (Able-Boone *et al.*, 2003). In other words, if we expect EI/ECSE professionals to deliver quality programs to young children in inclusive settings, then we must equip them with the knowledge and skills required for inclusive settings.

Itinerant Professionals

Itinerant professionals are those who travel from one site to another providing services for children with disabilities (Cook *et al.*, 2004). Itinerant services can follow a direct-instruction role as well as a consultation role. The direct-instruction role is similar to that of an ECSE teacher. Itinerant professionals, or teachers – as they are sometimes referred to – may provide direct services that do not include the child with his or her peers. Conversely, the itinerant consultant model is a role that is characterized by providing support to students in inclusive settings (see Cook *et al.*, 2004). A key feature of this role is the reciprocity that must occur in order for effective consultation to take place. Itinerant professionals often provide high-quality technical expertise, guide systematic evaluation and interventions, and serve as a liaison between working parents and special education staff (see Cook *et al.*, 2004).

Transition in ECSE

Two major transitions occurring in the lives of young children with special needs are those from early intervention to preschool and preschool to kindergarten. These transitions may be stressful for families of young children as the nature of services in the three settings is markedly different. Types of personnel, number of children, and the involvement of the family differ, as do the federal regulations with regard to the three settings. IDEA (1997) mandates planning for seamless transitions. Effectively preparing for transitions is of critical importance for children's future success.

The 1997 IDEA defines transition as “a coordinated set of activities for a student, designed within an

outcome-oriented process, that promotes movement from [one educational setting to the next].” Ensuring continuity of services, minimizing disruptions to the family system by facilitating adaptation to change, preparing children to function in the receiving environment, and fulfilling legal requirements are four major goals of transition planning (Wolery, 1999). The IDEA regulations from 1999 (Fed. Reg. Sec. 300) require states to have transition policies to ensure a seamless system of service delivery (Brandes *et al.*, 2007; Lillie and Vakil, 2002; Sainato and Morrison, 2001). A minimum of 6 months is required for the transition period from early intervention to preschool (Johnson, 2001). The IDEA mandates early-intervention and early-childhood professionals meet with families to discuss goals and objectives, timelines, and how to mitigate potential barriers in the transition process. At these meetings, a transition plan is written. Although transition planning is critical at both the early intervention-to-preschool and the preschool-to-kindergarten transitions, the two transitions have important differences.

Transition from Early Intervention to Preschool

A fundamental difference between early intervention and preschool is the governing section of IDEA. In early intervention, children are served under Part C of IDEA and services are provided based on the IFSP. Once a child turns 3, services are provided through Part B of IDEA and services are established with an IEP. Rosenkoetter *et al.* (2007) outlined critical differences between the two service delivery models, including a transition from a focus on the family to a focus on the child, home-based to center-based services, individual to small-group intervention, play-oriented to structured teaching approaches, and a developmental to an academic orientation. Two potential sources of stress for families in this transition is a discontinuity of services and ineligibility for services; that is, services available under Part C – such as respite care and home therapies – may not be available under Part B (Fox *et al.*, 2002). Further, children who qualify for services under Part C are not automatically eligible for services under Part B, and parents must be made aware of this possibility (Wolery, 1999).

The significant discrepancy in service delivery between early intervention and preschool causes stress for families and increases the need to plan for transition (Howard *et al.*, 2005). Lovett and Haring (2003) interviewed 48 families participating in this transition. Approximately half of these families (46%) reported they were comfortable with the transition process and a little less than half (43%) reported feeling uncomfortable with the process. The families reported that factors contributing to successful transitions were the early-intervention staff preparing them for the transition by setting up meetings with the

receiving schools, being involved in the IEP process, having choices about placements and programmatic issues, and feeling satisfied with the placements. Families who were uncomfortable with the process reported feeling unprepared and anxious, abandoned by their early-intervention staff, not understanding the process of transition, not involved in the IEP process, and not given choices about placements. Communication has also been reported as a factor leading to involvement by families in the transition process (Hanson *et al.*, 2000). Feeling involved and given choices can help parents feel comfortable with the stressful process of transition from early intervention to preschool.

As a means of promoting smooth transitions from early intervention to preschool settings, some states have developed local interagency coordinating councils (LICC)s that are responsible for writing interagency agreements (Sainato and Morrison, 2001). Wischnowski *et al.* (2000) interviewed members from transition teams to identify supports and barriers to writing interagency agreements that help smooth the transition process. Participants included early-intervention administrators, public school administrators, local education agency (LEA) coordinators, and parents. These team members suggested many critical variables supporting interagency agreement plans. Having skilled, involved key people with positive attitudes on the interagency team was reported to be the most influential variable. Process variables – such as writing plans with jargon-free language – were also important. The ability to change policies based on the needs of the team and effective ways of handling conflict resolution were also rated as important. Finally, training and access to technical assistance were considered to be important in promoting the writing of interagency agreements. Another critical source of continuity in the transition from early intervention to preschool is the use of independent therapy providers. Myers (2007) surveyed 103 occupational therapists, physical therapists, and speech-language pathologists who were independent contractors in Kentucky. These providers reported supporting the transition process by attending transition planning meetings and communicating with receiving environments. Early-childhood service coordinators were considered vital in supporting independent therapists to aid in transition. As independent therapy providers spend many hours of direct contact with families in early intervention, they are in a position to provide critical support to families during transition.

Transition from Preschool to Kindergarten

Just as there are many differences between early intervention and preschool settings, there are many differences between preschool and kindergarten settings (Rosenkoetter *et al.*, 2007). For example, there is much wider diversity of preschool settings than kindergarten settings: preschool can occur in home settings or center-based settings whereas

kindergarten is typically delivered in a school setting. Class sizes are often larger in kindergarten than pre-kindergarten and there are usually fewer adults in a kindergarten classroom. There are more expectations related to self-care in kindergarten and less one-on-one support available for students with disabilities than in pre-kindergarten settings. Large-group instruction and requirements for independent work occur more in kindergarten than in preschool. Finally, expectations for exhibiting appropriate behavior are higher in kindergarten than preschool. This factor can be an especially important consideration for students with disabilities (Stormont *et al.*, 2005).

Two critical variables in promoting smooth transitions from preschool to kindergarten are the anticipation by the preschool teacher of the dramatically different receiving environment of the kindergarten classroom and the preparation of children for that new environment. Troup and Malone (2002) conducted an ecological assessment of 11 inclusive kindergarten programs and reported features that preschool teachers should be aware of when preparing children for the transition. Findings were that the majority of programs sat students at small tables, about half of the programs clustered children on the floor for group activities, and about a third of programs required children to remain seated throughout assigned activities. In addition, most programs used assigned activities rather than child-choice activities, expected children to raise their hands to talk, and used workbooks and skill sheets. Preschool teachers should consider these aspects of kindergarten classrooms when preparing their students for transition. Preschoolers should not be expected to do kindergarten-type work as it would not be age-appropriate. However, if kindergarten students are, for example, expected to complete worksheets identifying size and color, preschool teachers should address these concepts in developmentally appropriate ways.

Numerous activities are available for promoting smooth transitions from preschool to kindergarten. La Paro *et al.* (2003) followed 80 families with high-risk children through this transition documenting the transition activities with which they participated as well as supports and barriers to the transition process. Families were given choices about transition activities consisting of visiting kindergarten classrooms, meeting with a specific or non-specific kindergarten teacher, meeting with an elementary school principal, participating in elementary school-wide events, and talking to preschool staff about kindergarten. When given the opportunity, most parents participated in these activities and reported that they were helpful. However, parents' work schedules were a common barrier to this involvement. Families prepared for the transition for kindergarten with their children at home by talking about behavioral expectations and working on pre-academic skills, such as reciting their address and phone number. Preschool teachers were also involved in transition

activities such as visiting the kindergarten classrooms with the children, holding spring orientations for families, and sharing written documents with kindergarten teachers. Kindergarten teachers were less involved in the transition activities and reported unpaid summer work and being given class-lists too late in the year as barriers to their involvement. As schedules were a barrier to participating in transition activities, La Paro and colleagues recommended schools offer creative ways for participation, such as offering evening parent-teacher conferences and conducting meetings via the telephone.

Transition to school for children with disabilities can present more problems than for typically developing children. McIntyre *et al.* (2006) assessed the intelligence, adaptive functioning, self-regulation, social skills, and student-teacher relationships of 24 children with intellectual disabilities and 43 typically developing children when they were 60-months-old and starting school. Results indicated that children with intellectual disabilities exhibited more problem behavior, less self-regulation, and more delayed social skills than typically developing children. Findings also suggested that social skills in preschool years predicted strong academic, social, and behavioral outcomes in school. Implications of these results are that preschool teachers should focus on increasing social skills and decreasing maladaptive behavior in children with intellectual disabilities to ensure their success in school-age settings. Clearly, transitions from early intervention to preschool and preschool to kindergarten can be stressful for families as the nature of services is considerably different across these three settings. These differences range from the physical setting to the number of children in the setting to the federal regulations mandating the services. Educators and service providers must smooth the transitions for families by setting up meetings with receiving environments, offering options regarding placements, and involving families in the IEP process. Having skilled people participate in local interagency coordinating councils (LICC) can also prepare families for the transition from early intervention to preschool. Preschool teachers can facilitate the transition to kindergarten by anticipating increased expectations for self-help, academic, and social skills and preparing students for those changes. Just as educators must provide a free and appropriate education for all children, they must also provide planned and seamless transitions for young children with disabilities and their families.

Recommendations for the Future

ECSE is a relatively young field that has made tremendous progress in the past few decades. Educators provide services to children with special needs across many different settings using a variety-service models. Professionals should

be trained to collect data on target behaviors and use data to both determine and guide interventions. Future EI/ECSE professionals must be prepared to meet the unique and diverse needs of children from culturally and linguistically diverse families (DEC, 2002). Professionals must respect and value the home cultures of all children and actively seek to engage each family in their child's intervention program. In order to work effectively with families and other professionals, EI/ECSE professionals require skills in collaboration. In other words, they must have skills in team building and goal setting as these skills are especially important when assisting children and their families in making transitions to new environments.

In addition to skills required for future EI/ECSE professionals, there is the need to continue a rigorous research agenda in ECSE. Research should focus on the identification of comprehensive and coordinated early interventions that will promote long-term benefits for children and families (Guralnick, 2005). Research in early childhood is needed to connect assessment and interventions in meaningful ways (DEC, 2006). Research is also needed to examine the effectiveness of delivery models across culturally and linguistically diverse families. Finally, the results of child and family outcomes need to be evaluated carefully, examining areas such as physical health, developmental learning, and social-emotional development (DEC, 2006).

Conclusion

ECSE is an exciting and growing field that has demonstrated past success in improving the lives of infants, toddlers, and young children with or at risk for disabilities. Future success depends on continuing to expand the knowledge base and on training professionals to meet the varied and complex needs of young children with special needs. As we move forward, the ability to determine children who need specialized intervention as early as possible and how to target resources to address the needs of each child will be critical in improving outcomes for children. The future of EI/ECSE holds the promise that we will be able to more effectively serve the needs of all children.

Bibliography

- Able-Boone, H., Crais, E. R., and Downing, K. (2003). Preparation of early intervention practitioners for working with young children with low incidence disabilities. *Teacher Education and Special Education* 26(1), 79-82.
- Adams, M. J. (1990). *Beginning to Read: Thinking and Learning about Print*. Cambridge, MA: MIT Press.
- Bailey, J. D. and Simeonsson, R. J. (1988). Home-based early interventions. In Odom, S. L. and Karnes, M. B. (eds.) *Early Intervention for Infants and Children with Handicap: An Empirical Base*, pp 199-216. Baltimore, MD: Paul H. Brookes.

- Bowe, F. G. (2000). *Birth to Five: Early Childhood Special Education*, 2nd edn. Albany, NY: Delmar.
- Bowe, F. G. (2004). *Birth to Eight: Early Childhood Special Education*. Clifton Park, NY: Delmar Learning.
- Bowman, B. T., Donovan, M. S., and Burns, M. S. (2000). *Educating Our Preschoolers*. Washington, DC: National Academy Press.
- Brandes, J. A., Ormsbee, C. K., and Haring, K. A. (2007). From early intervention to early childhood programs: Timeline for early successful transitions (TEST). *Intervention in School and Clinic* **42**, 204–211.
- Bruder, M. B. (2001). Inclusion of infants and toddlers: Outcomes and ecology. In Guralnick, M. J. (ed.) *Early Childhood Inclusion: Focus on Change*, pp 203–228. Baltimore, MD: Paul H. Brookes.
- Burns, M. K. and Coolong-Chaffin, M. (2006). Response to intervention: The role of and effect on school psychology. *School Psychology Forum: Research in Practice* **1**(1), 3–15.
- Burns, M. K. and VanDerHeyden, A. M. (2006). Using response to intervention to assess learning disabilities: Introduction to the special series. *Assessment for Effective Intervention* **32**(1), 3–5.
- Buyse, V. and Peisner-Feinber, E. (2008). *Response to Intervention Goes PreK: Early Intervention System Called Recognition & Response*. <http://breeze.unc.edu/sec619call4> (accessed August 2009).
- Cartledge, G., Gardner, R. III, and Ford, D. Y. (2009). *Diverse Learners with Exceptionalities: Culturally Responsive Teaching in the Inclusive Classroom*. Upper Saddle River, NJ: Merrill/Pearson.
- Chiang, H. M. (2008). Communicative spontaneity of children with autism. *Autism: The International Journal of Research and Practice* **21**(1), 9–21.
- Cook, R. E., Klein, M. D., Tessier, A., and Daley, S. E. (2004). *Adapting Early Childhood Curricula for Children in Inclusive Settings*, 6th edn. Columbus, OH: Merrill/Pearson.
- Crocker, A. C. and Porter, S. M. (2001). Inclusion of young children with complex health care needs. In Guralnick, M. J. (ed.) *Early Childhood Inclusion: Focus on Change*, pp 309–412. Baltimore, MD: Paul H. Brookes.
- Division of Early Childhood (DEC) (2002). Responsiveness to family cultures, values, and languages. *Position Paper: Division of Early Childhood of the Council for Exceptional Children*. http://www.dec-sped.org/uploads/docs/about_dec/position_concept_papers/PositionPaper_ResP_FamCul.pdf.
- Division of Early Childhood (DEC) (2006). Research priorities for early intervention and early childhood special education. *Research Committee Division of Early Childhood of the Council for Exceptional Children*.
- Fox, L., Dunlap, G., and Cushing, L. (2002). Early intervention, positive behavior support, and transition to school. *Journal of Emotional and Behavioral Disorders* **10**, 149–157.
- Gargiulo, R. M. and Kilgo, J. (2000). *Young Children with Special Needs*. Albany, NY: Delmar.
- Gargiulo, R. M. and Kilgo, J. (2004). *Young Children with Special Needs*, 2nd edn. Albany, NY: Thomson/Delmar.
- Guralnick, M. J. (2001). *Early Childhood Inclusion: Focus on Change*. Baltimore, MD: Paul H. Brookes.
- Guralnick, M. J. (2005). Early intervention for children with intellectual disabilities: Current knowledge and future prospects. *Journal of Applied Research in Intellectual Disabilities* **18**, 313–324.
- Hart, B. and Risley, T. R. (1995). *Meaningful Differences*. Baltimore, MD: Paul H. Brookes.
- Hanson, M. J., Beckman, P. J., Horn, E., et al. (2000). Entering preschool: Family and professional experiences in this transition process. *Journal of Early Intervention* **23**, 279–293.
- Hebbeler, K., Spiker, D., Bailey, D., et al. (2007). Early intervention for infants and toddlers with disabilities and their families: Participants, services, and outcomes. *Final Report of the National Early Intervention Longitudinal Study (NEILS). Technical Report Prepared for the Office of Special Education Programs*. Menlo Park, CA: SRI International.
- Hemmeter, M., Smith, B., Sandall, S., and Askew, L. (2005). *DEC Recommended Practices Workbook*. Missoula, MT: Division of Early Childhood of the Council for Exceptional Children.
- Heward, W. L. (2008). *Exceptional Children: An Introduction to Special Education*, 9th edn. Upper Saddle River, NJ: Merrill/Pearson.
- Hooper, S. R. and Umansky, W. (2009). *Young Children with Special Needs*, 5th edn. Upper Saddle River, NJ: Merrill/Pearson.
- Howard, V. F., Williams, B. F., and Lepper, C. E. (2005). *Very Young Children with Special Needs: A Formative Approach for Today's Children*, 3rd edn. Upper Saddle River, NJ: Pearson.
- Hyson, H. (ed.) (2003). *Preparing Early Childhood Professionals: NAEYC's Standards for Programs*. Washington, DC: NAEYC.
- IDEA (1997). Individuals with Disabilities Education Act of 1997, P.L. 105–17, 20 U.S.C. para.1400 es seq.
- Johnson, C. (2001). *Supporting Families in Transition between Early Intervention and School Age Programs*. Denver, CO: Colorado Department of Education.
- Keogh, B. K., Bernheimer, L. P., and Guthrie, D. (2004). Children with developmental delays twenty years later: Where are they? How are they. *American Journal on Mental Retardation* **109**, 219–230.
- La Paro, K. M., Kraft-Sayre, M., and Pianta, R. C. (2003). Preschool to kindergarten transition activities: Involvement and satisfaction of families and teachers. *Journal of Research in Childhood Education* **17**, 147–158.
- Lillie, T. and Vakili, S. (2002). Transition in early childhood for students with disabilities: Law and best practice. *Early Childhood Education Journal* **30**, 53–58.
- Linder, T. (ed.) (2008). *Transdisciplinary Play-Based Assessment & Transdisciplinary Play-Based Interventions*, 2nd edn. Baltimore, MD: Brookes.
- Lovett, D. L. and Haring, K. A. (2003). Family perceptions of transitions in early intervention. *Education and Training in Developmental Disabilities* **38**, 370–377.
- McIntyre, L. L., Blacher, J., and Baker, B. L. (2006). The transition to school: Adaptation in young children with and without intellectual disability. *Journal of Intellectual Disability Research* **50**, 349–361.
- McLoughlin, J. A. and Lewis, R. B. (2008). *Assessing Students with Special Needs*, 7th edn. Upper Saddle River, NJ: Pearson/Merrill.
- Miller, P. S. and Stayton, V. D. (2005). DEC recommended practices: Personnel preparation. In Sandall, S., Hemmeter, M. L., Smith, B. J., and McLean, M. E. (eds.) *DEC Recommended Practices: A Comprehensive Guide for Practical Application in Early Intervention/Early Childhood Special Education*, pp 189–212. Longmont, CO: Sopris West.
- Myers, C. T. (2007). The role of independent therapy providers in the transition to preschool. *Journal of Early Intervention* **29**, 173–185.
- Neisworth, J. T. and Bagnato, S. J. (2005). DEC recommended practices: Assessment. In Sandall, S., Hemmeter, M. L., Smith, B. J., and McLean, M. E. (eds.) *DEC Recommended Practices: A Comprehensive Guide for Practical Application in Early Intervention/Early Childhood Special Education*, pp 45–69. Longmont, CO: Sopris West.
- NHSA (The National Head Start Association) (2007). *2007 Annual Report*. <http://www.acf.hhs.gov/programs/ohs/about/fy2008.html> (accessed august 2009).
- Noonan, M. J. and McCormick, L. (2006). *Young Children with Disabilities in Natural Environments: Methods and Procedures*. Baltimore, MD: Paul H. Brookes.
- Odom, S. L. and McLean, M. E. (1996). *Early Intervention/Early Childhood Special Education: Recommended Practices*. Austin, TX: Pro-Ed.
- Odom, S. and Wolery, M. (2003). A unified theory of practice in early intervention/early childhood special education: Evidence-based practices. *Journal of Special Education* **37**(3), 152–173.
- Peterson, N. L., Barber, P. A., and Ault, M. M. (1994). Young children with special health-care needs. In Spodek, B. and Saracho, O. N. (eds.) *Early Childhood Special Education: Yearbook in Early Childhood Education*, pp 165–189. New York: Teachers College.
- Power, T. J. (Ed.). (2006). Integrating frameworks from early childhood intervention and school psychology to accelerate growth for all children [Special Issue]. *School Psychology Review* **34**(4).
- Raver, S. A. (2009). *Early Childhood Special Education – 0 to 8 Years: Strategies for Positive Outcomes*. Upper Saddle River, NJ: Merrill/Pearson.
- Rosenkoetter, S. E., Hains, A. H., and Dogaru, C. (2007). Successful transitions for young children with disabilities and their families: Roles of school social workers. *Children and Schools* **29**, 25–34.
- Sainato, D. M. and Morrison, R. S. (2001). Transition to inclusive environments for young children with disabilities: Toward a seamless system of service delivery. In Guralnick, M. (ed.) *Early Childhood*

- Inclusion: Focus on Change*, pp 293–306. Baltimore, MD: Paul H. Brookes.
- Sandall, S., McLean, M. E., Santos, R. M., and Smith, B. J. (2005). DEC's recommended practices: The context for change. In Sandall, S., Hemmeter, M. L., Smith, B. J., and McLean, M. E. (eds.) *DEC Recommended Practices: A Comprehensive Guide for Practical Application in Early Intervention/Early Childhood Special Education*, pp 19–26. Longmont, CO: Sopris West.
- School Psychology Review (2006). **35**(4).
- Stormont, M., Beckner, R., Mitchell, B., and Richter, M. (2005). Supporting successful transition to kindergarten: General challenges and specific implications for students with problem behavior. *Psychology in the Schools* **42**, 765–778.
- Troup, K. S. and Malone, D. M. (2002). Transitioning preschool children with developmental concerns into kindergarten: Ecological characteristics of inclusive kindergarten programs. *Journal of Developmental and Physical Disabilities* **14**, 339–352.
- Turbiville, V. P., Turnbull, A. P., Garland, C. W., and Lee, I. M. (1996). Development and implementation of IFSPs and IEPs: Opportunities for empowerment. In Odom, S. L. and McLean, M. E. (eds.) *Early Intervention/Early Childhood Special Education: Recommended Practices*, pp 77–100. Austin: Pro-Ed.
- US Department of Health and Human Services, Administration for Children, Youth and Families (2008). *Head Start Program Performance Standards & Other Regulations*. http://www.access.gpo.gov/nara/cfr/waisidx_07/45cfrv4_07.html#1301 (accessed August 2009).
- Wagner, M., Spiker, D., Lin, M., Gerlach-Downie, S., and Hernandez, F. (2003). Dimensions of parental engagement: Exploratory study. *Topics in Early Childhood Special Education* **23**(4), 171–187.
- Walker, H. M. and Severson, H. H. (2002). Developmental prevention of at-risk outcomes for vulnerable antisocial children and youth. In Lane, K. L., Gresham, F. M., and O'Shaughnessy, T. F. (eds.) *Children with or at Risk for Emotional and Behavioral Disorders*, pp 177–194. Boston, MA: Allyn and Bacon.
- Wall, S., Taylor, N., Liebow, H., et al. (2005). Early head start and access to early intervention services: A qualitative investigation. *Topics in Early Childhood Special Education* **25**(4), 218–231.
- Winton, P. J. and McCollum, J. A. (1997). Ecological perspectives on personnel preparation: Rationale, framework, and guidelines for change. In Winton, P. J., McCollum, J. A., and Catlett, C. (eds.) *Reforming Personnel Preparation in Early Intervention*, pp 81–101. Baltimore, MD: Paul H. Brookes.
- Wischnowski, M. W., Fowler, S. A., and McCollum, J. A. (2000). Supports and barriers to writing an interagency agreement on the preschool transition. *Journal of Early Intervention* **23**, 294–307.
- Wolery, M. (1999). Children with disabilities in early elementary school. In Pianta, R. C. and Cox, M. J. (eds.) *The Transition to Kindergarten*, pp 253–280. Baltimore, MD: Paul H. Brookes.
- Wolery, M. and Bailey, D. B., Jr. (2002). Early childhood special education. *Journal of Early Childhood* **25**, 88–99.
- Ysseldyke, J., Burns, M., Dawson, P., et al. (2006). *School Psychology: A Blueprint for Training and Practice III*. Bethesda, MD: National Association of School Psychologists.
- Optometrists Network (2005). Parents' guide to children's normal visual development from infancy to preschool. http://www.children-special-needs.org/parenting/preschool/visual_child_development.html (accessed August 2009).
- US Department of Health and Human Services (2002). Making a difference in the lives of infants and toddlers and their families: Impacts of early head start. *Early Head Start Research: Executive Summary*. <http://www.mathematica-mpr.com/publications/pdfs/ehsfinalsumm.pdf> (accessed August 2009).

Relevant Websites

- <http://www.cerebralpalsy.org> – 4 My child: Help and Hope for Life – Cerebral Palsy.
- <http://www.allergykids.com> – Allergykids.com.
- <http://www.acaa.org/public/advice/chldr.htm> – American College of Allergy, Asthma and Immunity on-line: Children and Allergies.
- <http://www.diabetes.org> – American Diabetes Association.
- <http://www.lungusa.org> – American Lung Association: Asthma and Children.
- <http://www.autism.com> – Autism Research Institute.
- <http://www.autism-society.org> – Autism Society of America.
- <http://www.behavior.org> – Cambridge Center for Behavioral Studies: Autism and ABA.
- <http://childcareandearlyed.clasp.org> – Center for Law and Social Policy (CLASP), Child Care and Early Education: Policy Analysis, Research, and Technology Assistance.
- <http://www.comeunity.com/disability> – Children's Disabilities and Special Needs.
- <http://www.childrendisabilities.info> – Children's Disabilities Information.
- <http://www.cacf.org> – Coalition for Asian American Children and Families.
- <http://www.deaf-culture-online.com> – Deaf Culture: Signing with Your Deaf Baby.
- <http://www.cdc.gov> – Department of Health and Human Services Center for Disease Control and Prevention.
- <http://www.devdelay.org> – Developmental Delay Resources: Integrating Conventional and Alternative Approaches to Learning and Behavioral Problems.
- <http://www.ds-health.com> – Down Syndrome: Health Issues.
- <http://ccf.edc.org/latinos> – Education Development Center, Inc.: Improving Access and Opportunity for Latinos in Early Childhood.
- <http://www.epilepsyfoundation.org> – Epilepsy Foundation: Epilepsy in Children.
- <http://www.firstsigns.org> – First Signs.
- <http://www.childtrends.org> – Guide to Effective Programs for Children and Youth: Infant Health and Development Program (IHDP).
- <http://www.kidsspeech.com> – Kaufman Children's Center for Speech, Language, Sensory-Motor, and Social Connections, Inc.
- <http://www.KeepKidsHealthy.com> – Keep Kids Healthy.com.
- <http://www.lighthouse.org> – Lighthouse International: Causes of Blindness.
- <http://www.nads.org> – National Association for Down Syndrome.
- <http://www.charityadvantage.com> – National Association for the Education of African American Children with Learning Disabilities, Charity Advantage.
- <http://www.naeyc.org> – National Association for the Education of Young Children.
- <http://deafness.about.com> – National Association of Deaf: Deaf Babies, About.com.
- <http://www.policyforchildren.org> – National Center for Children and Families: Advancing policy, education, and Development. Teachers College Columbia University.
- <http://www.nccp.org> – National Center for Children of Poverty, Columbia University, Mailman School of Public Health.
- <http://www.zerotothree.org> – National Center for Infants, Toddlers and Families, Zero to Three.
- <http://www.nccic.org> – National Child Care Information and Technical Assistance Center (NCCIC).
- <http://ndep.nih.gov> – National Diabetes Education Program, US Department of Health and Human Services.

Further Reading

- Bailey, J. D. and Simeonsson, R. J. (1988). *Early Intervention for Infants & Children with Handicaps*. Odom, S. L. and Karnes, M. B. (eds.). Baltimore: Paul H. Brookes.
- Education Resource Information Center (ERIC) (1984). Child care: Beginning a national initiative. *Hearing before the Select Committee on Children, Youth, and Families. House of Representatives, Ninety-Eighth Congress, Second Session, April 4, 1984*. http://eric.ed.gov:80/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED261760&ERICExtSearch_SearchType_0=no&accno=ED261760 (accessed August 2009).

<http://www.ndss.org> – National Down Syndrome Society.
<http://www.nei.nih.gov> – National Eye Institute: Study Confirms Value of Treatment to Prevent Blindness in Premature Babies.
<http://www.nga.org> – National Governors Association Center for Best Practices: Resources for Early Childhood Advisory Councils.
<http://www.nhsa.org> – National Head Start Association.
<http://www.nidd.org> – National Institute of Developmental Delays.
<http://www.nidcd.nih.gov> – National Institute on Deafness and Other Communication Disorders, Auditory Processing Disorder in Children.
<http://www.ed.gov> – National Institute on Early Childhood Development and Education: Office of Educational Research and Improvement, US Department of Education.
<http://health.nih.gov> – National Institutes of Health.
<http://www.goldrich.com> – Retinopathy of Prematurity: Goldsmith Ctorides and Rodriguez.

<http://www.cwdfoundation.org> – The Children with Diabetes Foundation.
<http://www.fpg.unc.edu> – The Carolina Abecedarian Project. University of North Carolina, The Child Development Institute.
<http://www7.nationalacademies.org> – The National Academics, Division of Behavioral and Social Sciences and Education, Board on Children Youth and Families.
<http://researchforum.org> – The Research Forum at the National Center for Children in Poverty.
<http://www.kidscount.org/datacenter/rightstart.jsp> – The Right Start. The Annie E. Casey Foundation.
<http://www.ed.gov> – US Department of Education.
<http://www.acf.hhs.gov> – US Department of Health and Human Services: Administration for Children and Families.
<http://www.oiep.bia.edu> – US Department of the Interior: Bureau of Indian Education. Learn today, lead tomorrow.

Early Intervention for Students with Special Needs

P A Grant, Slippery Rock University of Pennsylvania, Slippery Rock, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Early Intervention for Students with Special Needs

Infants, typically, are identified at birth for special services when they are suspected to have problems or special needs. Gearheart *et al.* (1993) suggested that early intervention means reducing the effects of existing disabilities and preventing or reducing the effects of secondary disabilities. Additionally, Carter (1999) reported that professionals and families whose children have exceptionalities find that, during the early years of childhood, the family is likely to be intensely absorbed with the inner workings of care and support for children with disabilities. Turnbull *et al.* (2006) noted that direct early-intervention efforts have targeted two major groups of infants and children – those with identifiable disabilities and those considered to be at-risk. These infants and children are served through organized center-based and home-based programs and through a combination of these two service-delivery models. In all instances, parent education and continuing parent involvement are essential ingredients for early-intervention programs. This article focuses on how best to use early intervention to help young children to maximize their potential.

Importance and Benefits of Early Intervention

Early intervention is the centerpiece of any nation's effort to help vulnerable children and their families (Guralnick, 1998). It produces mechanisms for mitigating stress generated by factors associated with family and child-risk or child-disability status that can adversely affect family-interaction patterns and, hence, the child's development. Earlier, Gearheart *et al.* (1993) contended that early-intervention support can provide assistance to parents and other family members as they nurture the child with disabilities. It may reduce feelings of guilt or inadequacy, prevent emotional problems in both parents and siblings of the child, and help hold marriages together. These benefits to the family mean direct benefits to the child with a disability.

Olive and McEvoy (2004) note that a critical tenet of early intervention/early childhood special education (EI/ECSE) is the belief that services for young children are designed and implemented with close involvement of families. Families should be decision makers for children

with disabilities. Their priorities should be used to drive early-intervention services. However, changing parental roles directly affect how intervention services can be provided. As a result, professionals must respect family choices to work outside the home, to home-school their children, or to enroll their children in private programs. Given changes in family demographics, it seems likely that a major service delivery model should be one where EI/ECSE personnel consult with a variety of programs where childcare is provided.

The importance of early intervention for children with severe and multiple disabilities cannot be stated strongly enough, (Friend, 2006). These children – because of the complexity of their disability – enter preschool without the basic skills that most children acquire without effort. Their communication skills are, sometimes, severely delayed. When a child's problems are recognized early, school failure can – to a large extent – be prevented or reduced (Lerner, 2003). ECSE programs (1) identify children ages birth through 5 who have special needs and are likely to encounter difficulty in academic learning and (2) provide an immediate early-intervention program. According to Lerner, early-intervention benefits the family of young children with special needs. In the family-centered-intervention approach, the child is viewed as part of the family system. When parents are empowered to be an integral part of the intervention process, they become essential elements in the process of reaching the child and improving child–adult interactions. Early-intervention programs offer a substantial financial savings for the community by reducing the number of children requiring special education services in the future. Specifically, it helps to:

- enhance intelligence;
- promote substantial gains, in all developmental areas (physical, cognitive, language, psychosocial, and self-help);
- inhibit or prevent secondary disabilities;
- reduce family stress;
- reduce dependency and institutionalization;
- reduce the need for special education services at school age; and
- save the nation and society substantial healthcare and education costs.

Honig (2004) presented a further rationale for early intervention. He suggested that, at birth, a baby's brain contains 100 billion neurons (as many as there are stars in the

Milky Way). During his/her first year, he/she will grow trillions of brain-cell connections – called neural synapses. Although an infant's brain does have some neurological hardwiring (such as the ability to learn any language), it is more pliable and more vulnerable than an adult brain; therefore, when one provides loving, language-enrichment experiences for a baby, he/she is giving the child's brain neural connections and pathways more chances to become wired together. In turn, the child will acquire rich language, reasoning, and planning skills. Additionally, Lerner (2003) argued that, perhaps, the most promising success stories in education today are the reports of special education programs for young children who have disabilities or are at risk. The underlying premise of ECSE is that early intervention makes a significant difference in child growth and development (Steel, 2006/2007). Professionals in psychological, medical, scientific, and educational fields have documented the importance of the years between birth and 5 for learning.

Concerns with Regard to Early Intervention

Perhaps one of the most touted studies in the area of early intervention was the Milwaukee Project. This study made some assumptions as to the role of deprivation in the development of intellectual disability. This study began in 1960 and was set in the Milwaukee Housing Project (Garber, 1988). The children identified for the study were determined as at-risk for mental retardation because they lived in low socioeconomic status (SES) development and because their mothers were identified as having an intelligence quotient (IQ) of 75 or below. Fifty percent of the children received intervention, while the other half did not. At regular intervals, the children were given a battery of tests. The groups were then compared to determine whether the intervention led to decreased rate of mental retardation among the children. The intervention package consisted of two parts – family rehabilitation and child stimulation. In the family-rehabilitation component, mothers were provided with paid instruction in job skills, followed by assistance in finding a job. The mothers (and some fathers) were also given instructions in daily-life skills – such as counting money, paying bills, and so forth. In this Milwaukee study, assessment information indicated that the children who received the intervention had IQs that averaged 30 points higher than those who did not. Additionally, they were significantly better at problem solving, exhibited more enthusiasm for learning tasks, and had significantly more advanced language skills. Follow-up assessments of IQ indicated that the difference continued through age 10 but the point range was smaller. Perhaps more startling, 60% of the control group children were classified as mentally retarded by the end of the

fourth grade, while none of the children who received the intervention package were so classified. The Milwaukee Project generated a great deal of enthusiasm about the prospect of early intervention. However, the interventions were intensive. Some children were placed in foster homes. The infant stimulation and preschool programs were run for 7 h a day, 5 days a week.

Unfortunately, the enthusiasm generated by the Milwaukee project would be marred by suspicion. Criticisms were made of the scientific rigor (or lack thereof) employed. Additionally, interested researchers had difficulty accessing the data for reanalysis (Smith *et al.*, 2006).

In *Miseducation: Preschoolers At-Risk*, Elkind (1987) expressed dismay at the fact that the age-inappropriate approaches to early education have gained such momentum despite the undeniable evidence that pushing children into formal academics and organized activities before they are ready does more harm than good. As he lamented, “In a society that prides itself on its preference for facts over hearsay, on its openness to research, and on its respect for ‘expert’ opinion, ‘parents, educators, administrators, and legislators are ignoring the facts, the researchers, and expert opinions about how young children learn and how best to teach them . . . When we instruct children in academic subjects, or in swimming, gymnastics, or ballet, at too early an age, we miseducate them; we put them at risk for short term stress and long term personality damage for no useful purpose” (pp. 3–4). Elkind further observed that those advocating early intervention in skills and early-out-of-home education rely upon youngsters who are very disadvantaged to tout early-education advantages. “Accordingly, the image of the competent child introduced to remedy the under stimulation of low income children now serve as the rationale for the over stimulation of middle class children” (p. 70). Dr. Jack Westman of the Rockford Institute, renowned child psychologist Dr. Stanley Greenspan, and brain researcher Jane Healey warn of the implications of forcing the childhood-as-a-race approach upon young children. According to Westman, there is the “hot housing movement” for infants and toddlers devoted to expanding their development. This is occurring in spite of the evidence that the long-term outcomes of early didactic, authoritarian approaches with younger children relate negatively to intellectual development (Pierce, 2007/2008).

Guralnick (1998) suggested that some critics have suggested that there is a lack of strong evidence for the long term effectiveness of early intervention services. Moreover, Rodis *et al.* (2003) reported that the merits of early intervention – as opposed to letting children outgrow their disabilities – is still a topic debated among scholars, educators, physicians, and parents. Marfo and Cook (1991) noted that the political context with which the earliest efficacy studies on early-intervention programs for children with disabilities were spawned accounted, in part, for the

remarkable tolerance of poorly designed interventions. However, it is important to review what has gone on in the past to try to have a clear picture of what is happening today and what may happen in the future. Two objectives undergirding these early studies are clearly discernible. There was certainly the twin scientific objective of seeking to (1) demonstrate that the progress and outcomes of early-intervention programs could be investigated with objectivity and scientific rigor and (2) generate data that could lead to the development of more effective intervention models and strategies. However, there was also a political objective – born largely out of anxiety to demonstrate to legislatures and policy-makers at a critical point in the fiscal history of the early-intervention movement that continued funding and possible expansion of services were justified. It is not surprising, therefore, that the vast majority of the research proclaiming the effectiveness of early intervention prior to the mid-1980s is now generally deemed to be methodologically flawed to be interpretable (Dunst, 1986; Dunst *et al.*, 1987).

Studies conducted over time have made us aware that politicization of inquiry at the consumption level has been equally problematic. A study conducted in Canada compared a group of 9-month-old Down syndrome children in a center-based development intervention program with a control group following 6 months of intervention for the experiment (Piper and Pless, 1980). Finding no statistically significant difference between the two groups on subtest of the Griffiths Scale, the researchers concluded that the efficacy of the form of intervention evaluated in their study was doubtful. For some 15 years, the Piper and Pless (1980) study was perhaps the most frequently criticized efficacy study on record. Among other scholars, Sheehan and Keogh (1981) observed that, by drawing this sweeping conclusion, Piper and Pless had gone beyond their data. The key elements of the various critiques are:

- The intervention was relatively short (Bricker *et al.*, 1981; Sheehan and Keogh, 1981).
- The sample size was small (Sheehan and Keogh, 1981).
- Neither the intervention effort itself nor the treatment variables under evaluation were described in sufficient detail to warrant precise interpretation of the findings (Bricker *et al.*, 1981).
- The Griffiths scale may not have been sensitive enough to register changes triggered by the treatment (Bricker *et al.*, 1981).
- The statistical techniques used in the study were not particularly powerful (Sheehan and Keogh, 1981).
- Given that the researchers did not collect any process data, they erred in interpreting the absence of experimental effects as evidence of inefficacy; their outcome could very well have been the result of factors related to the quality of treatment implementation (Marfo and Kysela, 1985).

Additionally, between 1984 and 1986, a series of meta-analytic reviews of early-intervention-efficacy studies – carried out at the Utah State University Early Intervention Research Institute (EIRI) – appeared in several sources (Castro and Lewis, 1984; Casto and Mastropieri, 1986; White *et al.*, 1984). The findings from these studies challenged and contradicted deeply entrenched beliefs and practices within the field. Among them were the following:

- Although parents can be effective intervenors, “they are probably not essential to intervention success” because “those intervention programs which utilize parents are not more effective than those which do not” (Casto and Mastropieri, 1986: 421).
- Contrary to the conventional belief that earlier exposure to intervention is better, it appears that – for children with special needs – later exposure is associated with better outcomes.
- More structured programs are only marginally superior to less structured programs.

As expected, these findings triggered much debate, controversy, and sharp criticism within the field. Donovan and Cross (2005) reported that there is substantial evidence with regard to both behavior and achievement that early identification and intervention is more effective than later identification and intervention. They further concluded that intervention programs have been known to address early biological harms and injuries –these have increased the potential to substantially improve developmental outcomes in young children.

Professional Preparation and Early Intervention

Olive and McEvoy (2004) reported, in 1997, that 77.8% of services for young children with disabilities were provided in outpatient clinic settings or through home-visiting programs. This implies that less than 13% of services in 1997 were provided in settings such as preschool programs and childcare centers where many young children spend most of their time each day. It appears that the predominant model for service delivery relies on bringing children to the service rather than providing the service in natural environments where they spend their time.

It is important to recognize that early intervention is driven based on the needs of parents who are confused about taking care of a child with a disability at such an early age and based on laws. The mandates of Individuals with Disabilities Education Act (IDEA) for inclusive and natural environment suggest that many children with disabilities will be served in preschool and childcare programs. Consequently, early-childhood educators should have formal training in serving young children with disabilities and

their families (see Olive and McEvoy, 2004). According to Olive and McEvoy, teachers of young children with disabilities must (1) become familiar with the range of settings where early-intervention services may be provided and (2) provide a range of services in their new classrooms whether they are public libraries, city buses, or neighborhood recreation centers. It is imperative that teacher-preparation programs design practicum that provide experiences across the continuum of placement options and collaborate with other agencies. Mandlawitz (2007) concluded that a Local Education Agency (LEA), annually, may use not more than 15% of its part B funds (part B, in combination with other funds to develop and implement services for students in grades K–12 (focusing on K–3)) who are not, at present, identified as needing special education and related services, but need extra academic and behavioral support to succeed in the general education environment.

Conclusion

Friend and Cook (2007) argue that collaboration is now central to all of special education, but it has been a defining feature of programs that meet the needs of young children with disabilities almost since the inception of that field. Perhaps, the best illustration of collaboration at the core of early intervention and preschool services is found in the IDEA rules and regulations that specify who participates on the team to develop the individual family service program (IFSP). In addition to the professional serving the child, the team must include the parents, other family members as needed, the service coordinator who has been working with the family or who has been assigned responsibility by a public agency, professionals who have assessed the child and professionals who are likely to provide services to the child. Thus, it is important that early intervention is initiated in a collaborative and constructive manner to help the child maximize his/her fullest potential. While there have been concerns about early intervention, the consensus opinion is that early intervention is a preventive method to reduce the intensity of future problems that may confront young children.

Bibliography

- Bricker, D., Carlson, L., and Schwartz, R. (1981). A discussion of early intervention for infants with Down syndrome. *Pediatrics* **67**, 45–46.
- Carter, B. (1999). Becoming parents: The family with young children. In Carter, B. and McGoldrick, M. (eds.) *The Expanded Family Life Cycle: Individual, Family and Society Perspective*, 3rd edn, pp 249–273. Needham Heights, MA: Allyn and Bacon.
- Casto, G. and Mastropieri, M. A. (1986). The efficacy of early intervention programs: A meta-analysis. *Exceptional Children* **52**, 417–424.
- Castro, G. and Lewis, A. C. (1984). Parent involvement in infant and preschool programs. *Journal of the Division for Early Childhood* **9**, 49–54.

- Donovan, S. M. and Cross, C. T. (2005). Is eliminating overrepresentation beyond the scope of public schools? In Byrnes, M. A. (ed.) *Taking Sides: Clashing Views on Controversial Issues in Special Education*, 2nd edn, pp 26–39. Dubuque, IA: McGraw-Hill/Dushkin.
- Dunst, C. J. (1986). Overview of the efficacy of early intervention programs. In Bickman, L. and Weatherford, D. L. (eds.) *Evaluating Early Intervention Programs for Severely Handicapped Children and Their Families*, pp 79–147. Austin, TX: Pro-Ed.
- Dunst, C. J., Snyder, S. W., and Mankinen, M. (1987). Efficacy of early intervention. In Wang, M. C., Reynolds, M. C., and Walberg, H. J. (eds.) *Handbook of Special Education*, vol. 3, pp 259–294. New York: Pergamon Press.
- Elkind, D. (1987). *Miseducation: Preschoolers at-Risk*. New York: Alfred A. Knopf.
- Friend, M. (2006). *IDEA 1994 Updated Edition: Special Education Contemporary Perspectives for School Professionals*. Boston, MA: Pearson Education.
- Friend, M. and Cook, L. (2007). *Interactions Collaboration Skills for School Professionals*, 5th edn. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Garber, H. L. (1988). *The Milwaukee Project: Preventing Mental Retardation in Children at-Risk*. Washington, DC: American Association on Mental Retardation.
- Gearheart, B., Mullen, R. C., and Gearheart, C. (1993). *Exceptional Individuals: An Introduction*. Pacific Grove, CA: Brooks/Cole.
- Guralnick, M. (1998). Effectiveness of early intervention for vulnerable children: A developmental perspective. *American Journal on Mental Retardation* pp 337–340.
- Honig, A. S. (2004). 20 ways to boost your baby's brain power. *Scholastic Parent and Child* **11**(4), 55–56.
- Lerner, J. (2003). *Learning Disabilities Theories, Diagnosis and Teaching Strategies*, 9th edn. Boston, MA: Houghton Mifflin.
- Mandlawitz, M. (2007). *What Every Teacher Should Know about IDEA 2004: Laws and Regulations*. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Marfo, K. and Cook, C. (1991). Overview of trends and issues in early intervention theory and research. In Marfo, K. (ed.) *Early Intervention in Transition: Current Perspectives on Programs for Handicapped Children*, pp 3–40. Westport, CT: Praeger.
- Marfo, K. and Kysela, G. M. (1985). Early intervention with mentally handicapped children: A critical appraisal of applied research. *Journal of Pediatric Psychology* **10**, 305–324.
- Olive, M. L. and McEvoy, M. A. (2004). Issues, trends, and challenges in early intervention. In McCray-Sorrells, A., Rieth, H. J., and Sindelar, P. T. (eds.) *Critical Issues in Special Education: Access, Diversity, and Accountability*, pp 93–105. Boston, MA: Pearson Education.
- Pierce, A. R. (2007/2008). Who's raising baby? Challenges to modern-day parenting. In Freiberg, K. L. (ed.) *Annual Editions: Human Development*, 35th edn, pp 44–54. Dubuque, IA: McGraw Hill/Contemporary Learning Series.
- Piper, M. S. and Pless, I. B. (1980). Early intervention for infants with Down syndrome. *Pediatrics* **65**, 463–468.
- Rodis, P., Garrod, A., and Boscardin, M. L. (2003). *Learning Disabilities and Life Stories*. Boston, MA: Allyn and Bacon.
- Sheehan, R. and Keogh, B. K. (1981). Strategies for documenting progress of handicapped children in early education programs. *Educational Evaluation and Policy Analysis* **3**(6), 59–67.
- Smith, B. M., Patton, J. R., and Kim, S. H. (2006). *Mental Retardation: An Introduction to Intellectual Disabilities*. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Steel, M. M. (2006/2007). Making the case for early identification and intervention for young children at risk for learning disabilities. In Freiberg, K. L. (ed.) *from Annual Edition: Educating Exceptional Children*, 18th edn, pp 25–28. Guilford, CT: McGraw-Hill/Dushkin.
- Turnbull, A., Turnbull, R., Erwin, E., and Soodak, L. (2006). *Families, Professionals, and Exceptionality: Positive Outcomes Through Partnership and Trust*, 5th edn. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- White, K., Mastropieri, M., and Castro, G. (1984). An analysis of special education early childhood projects approved by the joint Dissemination review panel. *Journal of the Division for Early Childhood* **9**, 11–26.

Further Reading

- Drew, C. J. and Hardman, M. L. (2007). *Intellectual Disabilities Across the Lifespan*, 9th edn. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Krass, L. R. (2007/2008). The age of genetic technology arrives. In Freiberg, K. L. (ed.) *Annual Editions: Human Development*, 35th edn, pp 12–20. Dubuque, IA: McGraw Hill/Contemporary Learning Series.
- Lerner, J., Lowenthal, B., and Egan, R. (1998). *Preschool Children with Special Needs: Children at-Risk and Children with Disabilities*. Boston, MA: Allyn and Bacon.
- Ziegler, E. (1997). The promise if early intervention to enhance the life outcomes of children with poverty. In Paul, J. L., Churton, M., Rosselli-Kostoryz, H., et al. (eds.) *Foundations of Special Education: Basic Knowledge Informing Research and Practice in Special Education*, pp 142–156. Pacific Grove, CA: Brooks/Cole.

Early Transition of Children with Special Needs

S A Fowler, University of Illinois, Champaign, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Early childhood special education services – The services delivered to children between the ages of 3 and school entry (ages 5 or 6), with developmental delays or disabilities. These services are identified through the development of an individualized education plan (IEP) that identifies developmentally appropriate goals and objectives for the child and the services required to assist the child in meeting these goals. Services may include: access to a special education preschool teacher, paraprofessional, or therapists, such as speech and language therapy, occupational therapy, and physical therapy. Other related services include: audiology, vision therapist, behavioral consultant, transportation, and adaptive technology. Services may be delivered within a classroom designated for children with special needs or within community child care. Most prevalent models used in community child-care programs include co-teaching with a special education teacher or the use of itinerant services.

Early intervention services – The services delivered to children between birth and their third birthday, who have been identified as developmentally delayed or disabled and their family. The services are identified through the development of an individualized family service plan (IFSP) that identifies the strengths and needs of the child and family as well as priorities. Services may be delivered through a home-visiting model or in a clinic or community child care. Services may include developmental therapy focused on the parent–child relationship and overall development of the child or specific therapies, such as speech and language therapy, occupational therapy, and physical therapy. Other related services may include: audiology, vision therapy, behavioral therapy, nutritional support, or family support.

Individualized education plan (IEP) – A document developed by the special education, general education, and parents or guardians of a child who qualifies for special education and related services. It provides diagnostic information on the child's abilities and a plan for addressing the supports required to assist the child in learning and engaging in school-related activities. IEP addresses domains in which the disability has an impact and the goals and

objectives for addressing the child's learning needs. It also identifies the supports (e.g., paraprofessional support, functional behavior analysis, and extended time for assignments) that will assist the child in meeting the goals and objectives. The plan is updated annually and reevaluated every 3 years when eligibility is again determined.

Individualized family service plan (IFSP) –

A planning document in which the early intervention service providers and family identify the young child's and family's strengths, needs and priorities, as well as the outcomes expected from service delivery. IFSP is developed within 45 days of diagnosis and must be reevaluated every 6 months.

Interagency agreement – A document in which two or more agencies identify the ways in which they will cooperate and provide support to a family and child as they make the transition from one early service program to another. Typically, such documents address the roles and responsibilities of the sending program and receiving program to ensure that child and family services are not interrupted during the movement from one service agency to another.

Service coordination – This involves the identification of all services provided to a young child and their family and planning so that the services are delivered in a way that reduces conflicts around issues, such as scheduling services, the focus of services, and payment of services. The intent is also to ensure that information about all services is shared among the providers and family in a way that optimizes delivery of services and reduces redundancy.

Transition – A significant change in services that involves the child and family leaving one agency or primary provider and initiating services with another agency or primary provider.

In the world of early childhood special education, the term transition refers to the movement of young children and their families from one service delivery system to another. Transition is understood to be a process and not an event and one that requires planning and communication among the family, the sending programs, and the receiving programs as well as the agencies overseeing such programs. The process of navigating the transition

between programs may be simple or very complex, and depends on the local context, agreements among agencies, and family participation and expectations regarding child readiness for the change of services. The goal of a successful transition is to ensure the adjustment and success of the child and family in their new services or program.

What is known is that transitions between services can be very stressful for young children with disabilities and their families. For example, children who first enter preschool may show signs of distress, ranging from crying or social withdrawal to regressions in behavior and skills as they begin the adjustment and adaptation to new providers and settings. Most children show a return to prior levels of comfort and skills, but this return marks a process for many that may require days and temporary supports as they make the adjustment (Haymes *et al.*, 1994). Families often report a renewed sense of loss or grief about their child's disability as they recognize that changes that may appear fairly effortless for other families and children are not for their child or family. They may also find themselves in a position of needing to explain again their child's strengths and needs and advocate for appropriate services and supports. They may miss the relationships, which they developed over time with the sending program as they work to build rapport and trust with the staff of the receiving program.

The transition from one program to another may create discontinuity in a child's program due to differences in program philosophy, regulatory requirements, family or staff expectations, or simply the failure to share assessments and records about a child's current developmental status. This, in turn, can interrupt child progress. In the United States, young children, identified with disabilities or developmental delays are eligible for services from birth or shortly after the diagnosis has been made. As a result, many families begin receiving services for their child during the infancy and toddler years. Federal legislation in the United States has ensured services for infants and toddlers through the 1986 Early Intervention Program of the Education of All Handicapped Children Act (e.g., PL 99-457). The law calls for minimum components of a statewide intervention system for infants, toddlers, and their families but allows for state discretion in terms of organizing and structuring the services. As it stands, services may vary from state to state on a number of dimensions, including who is eligible for services, who provides services, and how services are coordinated across agencies within a community and within a state (Spiker *et al.*, 2000). In fact, research from a national longitudinal study on early intervention services, suggests that there is no typical early intervention service, but rather a range of services that may vary along different dimensions (e.g., setting for services, intensity of services, and focus of services) (Hebbeler *et al.*, 2008).

Because of the way in which the law was developed, most children with developmental delays or disabilities

and their families encounter a transition in services on the child's third birthday, when funding for early intervention ends and funding for preschool-age services begin (Malone and Gallagher, 2008). This involves the shift in services from early intervention providers to those provided through the local education agency for preschool. During this transition, children move from a service delivery model in which most services have been delivered through a developmental therapist and other specialists (e.g., speech therapy and physical therapy) in the context of the child's home, or in a clinic or child-care setting. New services are provided through the local school district and may take the form of a preschool program in which the child is enrolled or may be delivered through a consultation model to the child-care center attended by the child. The second transition, which typically occurs again during the child's fifth year of age, is entry to kindergarten and formal schooling. Again, this involves a change in services from preschool, which may or may not be located in the same place or under the same agency as the kindergarten.

This article addresses both the early transition to preschool and the later transition to formal schooling or kindergarten for children with disabilities as well as the evidence-based practices to increase the continuity of treatments and services for children and families across the first 6 years of life. Issues discussed include a conceptual framework for understanding the transition process, the role of agencies involved in the transitions, the role of the families, and the impact on the child.

Conceptual Framework

The current conceptual framework used to discuss transitions in the United States is one based on bioecological theory and organizational theory (Rous *et al.*, 2006). Bioecological theory focuses on the interaction between child development and the environmental context. This includes a consideration of the risk and resiliency factors evident in the developing child and the risk and resiliency aspects of the child's immediate environment: family, home, and community contexts (Bronfenbrenner and Evans, 2000). It also considers the timing of the transition process. **Figure 1** provides the ecological contextual factors first identified by Rous *et al.* (2006).

The underlying premise presented in **Figure 1** is that most services for very young children are delivered within the context of the family. That is, services are guided by the family's values, beliefs, and priorities, and are identified through a collaborative family and professional planning process. The child and family are a part of a larger community, and the availability and accessibility of services and resources within that community will impact what the family and child receive. For example, families

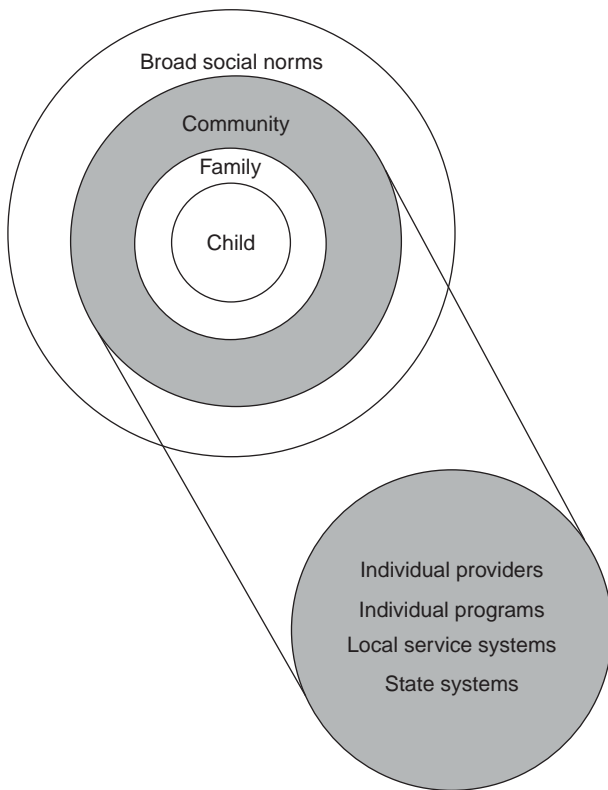


Figure 1 A conceptual framework for planning transitions in the early years.

who live in urban communities or densely populated areas are likely to have more choices regarding the programs that are available in their local service system as well as choices in the providers. In contrast, families in rural areas may have limited choice if few providers or only a single program is available.

In addition, the framework presented in **Figure 1** accounts for the influence of broad social norms, which dictate whether there are laws and adequate funding to provide early intervention services and to support the professional development of service providers. These norms also determine whether the community shares expectations about the process and outcomes for children and families who benefit from these services (Fowler *et al.*, 2006). As federal legislation established early intervention services in the United States more than 20 years ago, many citizens consider access to early intervention services as an entitlement and are likely to seek and accept services. However, research on service access indicates that not all families who are eligible participate. Their participation is influenced by cultural values, including the ways in which the families view disability, their beliefs about child rearing, and the role of individuals or agencies outside of the family and the family's community. Groups which are least likely to use early intervention services include immigrant families with limited English, families with limited education, and families

who live in isolated areas of the country (Crosby and Hatfield, 2008). The community, in turn, is impacted by the extent to which the larger governing bodies (states, provinces, and nations) provide an infrastructure to fund, support, and monitor the delivery of early intervention and special education services. In a transition, families navigate between providers and programs and sometimes even between service systems.

The transition process is depicted in **Figure 2** and illustrates the mediating variable of transition practices and activities to guide families and their children in preparing for transition and in adjusting to the transition. These practices and activities are determined through interagency agreements, that is agreements and plans made by agencies involved in programs from which the child is departing and programs to which the child is arriving (Rous and Hallam, 2006; Fowler *et al.*, 2000). This conceptual framework is further discussed within the role of the child, family, providers, and service system.

Early Service Transitions within the United States

Young children with disabilities or developmental delays, who are identified prior to age 3, are eligible for services, which are based on the development of an individualized family support plan (IFSP). The plan identifies strengths and resources of the family and child as well as needs of the family and child, which must be met to support the child's development. Children with significant disabilities are most likely to be identified early in life and receive early intervention for most of their first 3 years. Conversely, children with milder or less obvious disabilities or developmental delays may not be identified until they fail to reach normal developmental milestones, related to ambulation, speech and language, or social interaction. These children may be in early intervention for a year or less (Malone and Gallagher, 2008). Hebbeler *et al.* (2008) in their follow-up of over 3,000 children indicated that six services form the core of early intervention delivered in the United States. These include (1) speech-language therapy, (2) physical therapy, (3) occupational therapy, (4) special instruction for the child, (5) developmental monitoring, and (6) service coordination. The primary focus of the services may be the child and parent (or adult caregiver) or exclusively the child. They noted that the majority of families received services in the home. Within this group, more than half of the services focused on helping the parents to work with their child both during and beyond the therapist's visit. Oftentimes, families and early intervention providers form close relationships as they work together to foster optimal development in the child.

Transition planning is built into early intervention services and is a requirement in the United States. At least

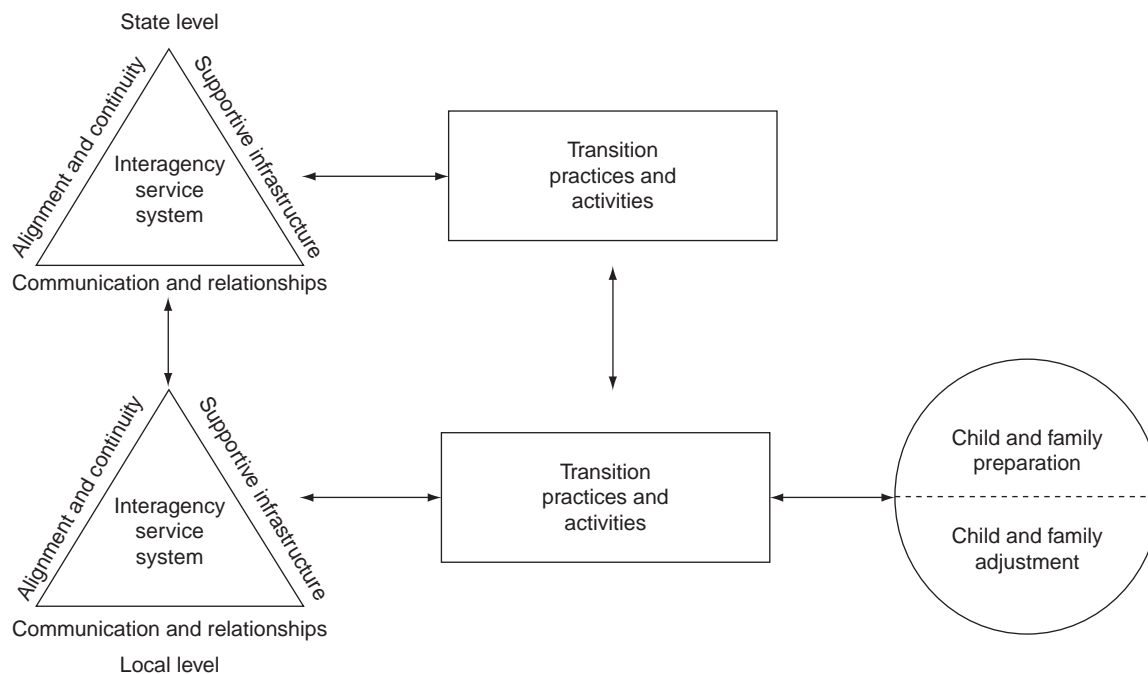


Figure 2 State and local factors that influence transition practices for young children and families.

90 days prior to the child's third birthday, the early intervention provider or agency, family, and local public school system must meet to plan the transition to preschool- or other public-school-funded services. In many states, the source of funding for services, as well as the method of delivering services, changes on the child's third birthday. As a result, the date of a child's transition does not reflect the child's developmental readiness or the family's readiness for change, but rather an administrative mandate tied to funding sources. For some children, the transition is seamless, if the early intervention agency and the public school program are either part of the same agency or have strong interagency agreements to ensure continuity of services. For many children, however, a disruption in service may occur. This disruption may be due to many variables, including a family decision to not initiate preschool, the time of year in which the child's birthday occurs, or the readiness of public schools to initiate services.

Factors that Influence the Age 3 Transition

The most common legal problems identified for age 3 transition include:

1. failure to develop or implement individualized education plans (IEPs) by age 3,
2. failure to establish eligibility for continued special education services,

3. lack of required assessments, and
4. absence of the required representative from the local education agency for the transition meeting (US Department of Education, 2004).

However, a range of child and family factors as well as local community and state factors should also be considered.

Child and Family Factors

Not all families may be ready for the change of services or willing to participate in the transition planning. In many communities, services for children aged 3–5 are provided in the context of a half-day preschool program located often in a public school. Some families may express concern that their child is too young or too vulnerable to begin a preschool program or to ride a school bus to such a program; or, they may not consider their child's developmental delay to warrant continued services, if these services mean a change in the family schedule, or a need to provide child-care services that can wrap around the half-day special education preschool program (Rosenkoetter *et al.*, 1994). In other cases, families who have been involved with services for child neglect or abuse may not want to further expose themselves to a system that monitors their child's development (Robinson and Rosenberg, 2004). Families, who are recent immigrants, may not understand or trust the referral process, which often

requires additional evaluations of the child and proof of residency.

Families who are supportive of the transition can play a critical role in easing their child's adjustment by participating in planning and by using recommended practices to introduce their child to the new program. Such recommended practices include:

1. providing opportunities for the child to learn to separate from the parent and play at a relative's or friend's home;
2. supporting the development of social skills;
3. establishing simple routines for the child to follow at home; and
4. reading to their child and expanding their child's language (e.g., Pianta and Cox, 1999).

Community Factors

In the United States, federal law requires that the early intervention program conducts a transition planning meeting with the family and a representative from the local education agency at least 90 days before the child's third birthday. At this meeting, they are required to include the following steps to support transition in the child's IFSP:

- discussion of future options with parents or guardians;
- procedures to prepare the child and to help the child adjust;
- transmitting information to the local education agency, with parents' consent; and
- explanation of how the families will be included in the transition plan.

The extent to which the community supports the transition is dependent on whether the early intervention programs and providers and local education agencies have developed agreements to cooperate or have a history of cooperation. Without a supportive interagency framework, it is difficult to ensure that relationships are built between programs that truly focus on serving the child and family. Communication is critical to ensure that a representative from the receiving program is at the transition planning meeting, or that appropriate records and assessments are both current and sent to the next program. If the eligibility criteria differ between the sending and receiving programs, then new assessments must be scheduled and conducted by the receiving or sending program. If assessments are not planned in advance and coordinated, then the child's entrance to new services may be delayed until evaluations are completed and eligibility established (Rous and Hallam, 2006). A national survey of public preschool teachers conducted by Rous *et al.* (2006) indicated that teachers engage in a variety of

Table 1 Frequently identified practices for supporting the entrance and adjustment to preschool

Talking with parents before school starts and/ or meeting with the parents after school starts
Sending a letter or informational brochure to parents before their child starts preschool
Conducting an open house before or after school starts
Reviewing written records about the child's prior experiences
Facilitating contacts between parents of children in the class

Adapted from Rous, B., McCormick, K., and Hallam, R. (2006). Use of transition practices by public preschool teachers. *Research Brief*. National Early Childhood Transition Center.

Table 2 Most common kindergarten transition practices

Sending parents information on kindergarten
Arranging for a preschool class to visit a kindergarten class
Meetings with kindergarten teachers to discuss curriculum

Adapted from Rous, B., McCormick, K., and Hallam, R. (2006). Use of transition practices by public preschool teachers. *Research Brief*. National Early Childhood Transition Center.

strategies to support a child's entrance to preschool services and subsequent adjustment. **Tables 1 and 2** contain the most frequently identified practices.

State, Provincial, or National Factors

Leadership at the national and regional level is essential to ensure that communities provide appropriate support and practices for families and children during the transition. In the United States, states provide such leadership through state policy, training, technical assistance, and funding.

Entrance to Preschool Services

Rous *et al.* (2007) have identified in their conceptual framework the concept of a "critical window of time" for children and families to adjust to a change in service delivery and programs. "This refers to the time it takes for the child to actively engage, adapt and continue along an appropriate developmental trajectory" (p. 144). They discussed the notion of three interdependent transition outcomes related to this critical window. These are:

1. the child's engagement in the program, as measured by his/her interaction with adults, peers, and materials;
2. the child's adaptation to the formal and informal rules of the program (e.g., schedule, routines, and behavioral expectations); and
3. the child's continued development socially, cognitively, and linguistically.

Children with these outcomes are most likely to have parents who view their program positively and are supportive of ensuring that their child attends services regularly.

Transition to Kindergarten

No federal regulations govern the transition to kindergarten in the United States. Children typically are eligible to enter kindergarten if they have turned 5 years of age before the school year begins or by a predetermined date. In many countries, entry to kindergarten is regarded as the first point of entry to formal schooling and may occur between age 5 and age 8. Attendance may be voluntary or compulsory, depending on local and national norms or laws. Many of the issues involved in the earlier transition to preschool services, are also common to kindergarten. The issues include recommended practices to prepare the child and family for kindergarten and to support adjustment in kindergarten. They also include the need for coordination and communication between the sending and receiving programs, which again, may be part of the same agency or represent different agencies. **Table 3** presents the most common practices identified in a survey of 3000 early childhood professionals by Rous (2008).

The issue of child readiness for school or the school's readiness for all children is often a debate in educational circles. The choice of child readiness versus school readiness may be determined by political, financial, social, and cultural factors. Many countries recognize a specific age

by which children are expected to start school. Readiness is an issue when schools expect children to enter with certain skills and when standardized assessments based on age norms are used. Children then are compared to their peers and not to their prior record or experiences. The debate in the United States was intense in the 1980s with many schools establishing pre-kindergarten programs or transitional kindergarten programs for children who did not meet certain age-expected norms. Recent trends have focused on the theory that schools should be ready for all children and the term readiness to a large extent has been modified to represent a "broad construct that incorporates all aspects of a child's life that contribute directly to that child's ability to learn" (Meisels, 1999: 62). Even so, across the United States, many states have implemented publicly funded pre-kindergarten programs for children identified as at risk for school learning problems based on risk factors such as poverty or English-language learning. The intent of these programs is to support children's readiness to learn.

Most children with disabilities are likely to enter kindergarten having already attended a preschool program in which special education services were provided or received services through their child-care program or at home. These children enter with an IEP, which is required for all children identified as having a developmental delay or disability. Children with IEPs are reevaluated every 3 years to determine the continued need for services and, if needed, may continue to receive special education services and support throughout the entire school career. Concerns most typically expressed by kindergarten teachers around the transition of children with developmental delays are (1) having time to plan and consult with special education and related services staff; (2) having support in the form of an assistant teacher or aide in the classroom as needed; and (3) receiving additional training to address the specific needs of the child (Rosenkoetter *et al.*, 1994). These needs are increased when a child's disability is significant or when several children with disabilities are enrolled in the same classroom. The need for communication between home and school and opportunities to discuss strategies with prior preschool teachers also have been identified as a need by some teachers in facilitating the adjustment of the child to kindergarten (Rous and Hallam, 2006). Transitions are not successful if supports are not provided to children with disabilities. This, then, can begin a cycle of failures and frustrating school experiences for the child, family, and teachers.

About two decades ago, Hains *et al.* (1989) also suggested that kindergarten teachers may have initially low expectations for children's entry skills to kindergarten, focusing primarily on self-help and social skills. These expectations grow significantly over the course of the year, as the curriculum changes and becomes more academic in nature with a focus on numeracy and literacy and greater

Table 3 Nationally validated transition practices for child and family preparation and adjustment

- Individual child and family transition meetings are conducted
- Staff follow-up on children after the transition to support their adjustment
- Transition team members share appropriate information about each child making a transition
- Transition plans are developed that include individual activities for each child and family
- Staff know key information about a broad array of agencies and services available within the community
- Children have opportunities to develop and practice skills they need to be successful in the next environment (program)
- Families are aware of the importance of transition planning and have information they need to actively participate in transition planning with their child
- Families' needs related to transition are assessed and addressed
- Families have information about and are linked with resources and services to help them meet their specific child and family needs

Adapted from Rous, B. (2008). Recommended transition practices for young children and families: Results from a national validation survey. *Technical Report #3*. Lexington, KY: University of Kentucky, Human Development Institute, National Early Childhood Transition Center, p 11.

self-regulation. The extent to which children with disabilities are able to maintain the same pace as their peers varies and may require additional supports, whether in the classroom or through the use of a resource room for specialized instruction. The issues then move from those of transition into kindergarten to issues of inclusive educational services and provision of special education in kindergarten and elementary school (Werts *et al.*, 1996). Transition is important at all levels of a young child's education. All hands must be on deck.

Conclusion

Transitions occur at significant points in a young child's life beginning with his/her diagnosis of a disability or developmental delay and entrance into early intervention services. The transition from early intervention to preschool services at age 3 marks a significant change for many children and families in the delivery of services. Children still eligible for services are most likely to enter a preschool program in which special education services are delivered. This program may be structured to enroll only children with delays or may be an inclusive program with typically developing children. Children then make a transition, universal for nearly all children, following their fifth birthday, to kindergarten and the beginning of a more academic focus in school. Transitions at each point in time require an assessment of the child's current performance and skills, planning for the expectations of the next program, implementing supports as needed to support the child's transition and adjustment to the new program, and an evaluation of how well the adjustment occurred. This cycle of assessment planning, implementing, and evaluating can provide important information if gathered over multiple transition points to ease a given child's transition throughout his/her school career. If systematically gathered for the entry of many children into preschool and kindergarten, then this information can also begin to identify patterns of successful adjustment and factors that contributed to or detracted from such adjustments.

Bibliography

- Bronfenbrenner, U. and Evans, G. (2000). Developmental science in the 21st century: Emerging questions, theoretical models, research designs and empirical findings. *Social Development* **9**, 115–125.
- Crosby, D. A. and Hatfield, B. E. (2008). The ALAS project: Removing barriers to mental health care for Latina immigrant mothers. *Zero to Three* **29**, 31–38.
- Fowler, S. A., Donnegan, M., Lueke, B., Hadden, D. S., and Phillips, B. (2000). Evaluating community collaboration in writing interagency agreements on the age 3 transition. *Exceptional Children* **67**, 35–50.
- Fowler, S. A., Ostrosky, M. M., and Yates, T. (2006). Teaching and learning in the early years. In Florian, L. (ed.) *The Sage Handbook of Special Education*, pp 349–359. London: Sage.

- Hains, A. H., Fowler, S. A., Schwartz, I., Kottwitz, J. E., and Rosenkoetter, S. E. (1989). A comparison of preschool and kindergarten teachers' expectation for school readiness. *Early Childhood Research Quarterly* **4**, 75–88.
- Haymes, L. K., Fowler, S. A., and Cooper, A. Y. (1994). Assessing the transition and adjustment of preschoolers with special needs to an integrated program. *Journal of Early Intervention* **18**, 184–198.
- Hebbeler, K., Spiker, D., Morrison, K., and Malik, S. (2008). A national look at the characteristics of Part C early intervention services. *Young Exceptional Children Monograph Series* **10**, 1–18.
- Malone, D. G. and Gallagher, P. A. (2008). Transition to preschool programs for young children with disabilities. *Journal of Early Intervention* **30**, 341–356.
- Meisels, S. J. (1999). Assessing readiness. In Pianta, R. C. and Cox, M. J. (eds.) *The Transition to Kindergarten*, pp 39–66. Baltimore, MD: Brookes.
- Pianta, R. C. and Cox, M. J. (1999). The changing nature of the transition to school: Trends for the next decade. In Pianta, R. C. and Cox, M. J. (eds.) *The Transition to Kindergarten*, pp 363–379. Baltimore, MD: Brookes.
- Robinson, C. C. and Rosenberg, S. A. (2004). Child welfare referrals to Part C. *Journal of Early Intervention* **26**, 284–291.
- Rosenkoetter, S. E., Hains, A. H., and Fowler, S. A. (1994). *Bridging Early Services for Children with Special Needs and Their Families: A Practical Guide for Transition Planning*. Baltimore, MD: Brookes.
- Rous, B. (2008). Recommended transition practices for young children and families: Results from a national validation survey. *Technical Report #3*. Lexington, KY: University of Kentucky, Human Development Institute, National Early Childhood Transition Center.
- Rous, B. and Hallam, R. (2006). *Tools for Transition in Early Childhood: A Step-by-Step Guide for Agencies, Teachers and Families*. Baltimore, MD: Brookes.
- Rous, B., Hallam, R., Harbin, G., McCormick, K., and Jung, L. (2007). The transition process for young children with disabilities: A conceptual framework. *Infants and Young Children* **20**, 135–149.
- Rous, B., McCormick, K., and Hallam, R. (2006). Use of transition practices by public preschool teachers. *Research Brief*. National Early Childhood Transition Center. <http://www.hdi.uky.edu/nectc> (accessed August 2009).
- Spiker, D., Hebbeler, K., Wagner, M., Cameto, R., and McKenna, P. (2000). A framework for describing variables in state early intervention systems. *Topics in Early Childhood Special Education* **20**, 195–207.
- US Department of Education (2004). *Part B Annual Performance Report Letters*. <http://www.ed.gov/fund/data/report/idea/partbapr/index.html> (accessed August 2009).
- Werts, M. G., Wolery, M., Snyder, E. D., Caldwell, N. K., and Salisbury, C. L. (1996). Supports and resources associated with inclusive schooling: Perceptions of elementary school teachers about need and availability. *Journal of Special Education* **30**, 187–203.

Further Reading

- Barnett, W. S., Hustedt, J. T., Friedman, A. H., Boyd, J. S., and Ainsworth, P. (2007). *The State of Preschool 2007: State Preschool Yearbook*. New Brunswick, NJ: National Institute for Early Education Research.
- Bruns, D. A. and Fowler, S. A. (1999). Facilitating early childhood transitions for children and families from culturally and linguistically diverse groups. *Teaching Exceptional Children* **31**, 26–30.
- Chandler, L. K. (1993). Steps in preparing for transition: preschool to kindergarten. *Teaching Exceptional Children* **25**(4), 52–55.
- Gallagher, J. (1999). Policy and the transition process. In Pianta, R. C. and Cox, M. J. (eds.) *The Transition to Kindergarten*, pp 351–362. Baltimore, MD: Brookes.
- Hanson, M. J. (2005). Ensuring effective transitions in early intervention. In Guralnick, M. (ed.) *The Developmental Systems Approach to Early Intervention*, pp 378–398. Baltimore, MD: Brookes.
- Harbin, G., Rous, B., and McLean, M. (2005). Issues in designing state accountability systems. *Journal of Early Intervention* **27**(3), 137–164.

- Rosenketter, S. E., Whaley, K., Hains, A. H., and Pierce, L. (2001). The evolution of transition policy for young children with special needs and their families: Past, present and future. *Topics in Early Childhood Special Education* **1**, 3–15.
- Rous, B., Schuster, J., and Hemmeter, M. L. (1994). Sequenced transition to education in the public schools: A system's approach to transition planning. *Topics in Early Childhood Special Education* **14**, 374–392.
- Scarborough, A. A., Spiker, D., Mallik, S., *et al.* (2004). A national look at children and families entering early intervention. *Exceptional Children* **70**, 469–483.
- US Department of Education (1999). Assistance to states for the education of children with disabilities and the early intervention program for infants and toddlers with disabilities: Final regulations (34 CFR Parts 300 and 303). *Federal Register* **64**(48), 330.132.
- Wischnowski, M. W., Fowler, S. A., and McCollum, J. A. (2000). Supports and barriers to writing an interagency agreement on the preschool transition. *Journal of Early Intervention* **23**, 294–307.
- Wolery, M. (1999). Children with disabilities in early elementary school. In Pianta, R. C. and Cox, M. J. (eds.) *The Transition to Kindergarten*, pp 253–280. Baltimore, MD: Brooks.

Relevant Websites

- <http://clas.uiuc.edu> – Culturally and Linguistically Appropriate Services for Young Children with Special Needs, University of Illinois.
- <http://facts.crc.uiuc.edu> – Family and Child Transitions into the Least Restrictive Environment, University of Illinois.
- <http://www.fpg.unc.edu> – FPG Child Development Institute, National Center for Early Development and Learning.
- <http://www.hfrp.org> – Harvard Family Research Project, The Transition to Kindergarten: A Review of Current Research and Promising Practices to Involve Families.
- <http://www.hdi.uky.edu> – Human Development Institute, National Early Childhood Transition Center, University of Kentucky.
- <http://www.nectac.org> – National Early Childhood Technical Assistance Center, University of North Carolina at Chapel Hill, Transition from Preschool Services to Kindergarten.
- <http://www.peels.org> – Pre-Elementary Education Longitudinal Study, National Center for Special Education Research, Institute for Educational Science, US Department of Education.

Educating Children Who Are Deaf-Blind

J Wood and K M Chinn, New Mexico State University, Las Cruces, NM, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Asphyxia – A condition in which the lungs are unable to function normally enough to provide a sufficient amount of oxygen to the body.

CHARGE – This term is made from the initial letters of some of the most common features of this condition. For example, C = Coloboma, an eye defect caused by failure of the eye to form properly so that part of the structure within the eye is missing, resulting in a range of defects: a so-called keyhole pupil at the front of the eye, to very large defects at the back of the eye. In some cases, the defect may be so large that it results in a smaller or even absent eye; and E = Ear anomalies that can affect the external, middle, and inner aspects of the ear. Some common problems include hearing loss, fusing of the middle-ear bones, chronic middle-ear effusion, stenosis, or atresia.

Close senses – The senses of smell, touch, and taste.

Congenital – A condition present at birth.

Distance senses – The senses of vision and hearing.

Encephalitis – An infection, either bacterial or viral, in the brain.

Etiology – Origin or cause.

Exacerbate – To make worse.

Functional blindness – A degree of blindness in which one may be able to distinguish shapes, colors, and shadows and may function within a normal lifestyle.

Hydrocephaly – Excessive cerebral spinal fluid in the cranium; also called water head.

Low vision – A degree of visual impairment that results in the need for larger print and better lighting for access to the printed word.

Meningitis – A form of encephalitis but limited to the membranes that line and protect the brain.

Microcephaly – Unusually small head.

Perinatal – During birth.

Postnatal – After birth.

Prenatal – Before birth.

Retinitis pigmentosa – This is also called tunnel vision; a condition in which the eye's retina slowly deteriorates from the peripheral vision to the center of vision over time, resulting in limited vision or blindness.

Rubella – German measles also called 3-day measles.

Usher's syndrome – A genetic condition in which the individual is born with hearing loss and gradually develops tunnel vision.

Educating Children Who Are Deaf-Blind

Deaf-blind is the presence of a dual sensory impairment involving the auditory and visual senses. According to the US Federal Government,

Deaf blindness means concomitant hearing and visual impairments, the combination of which causes such severe communication and other developmental and educational needs that they cannot be accommodated in special education programs solely for children with deafness or children with blindness. (Office of Special Education and Rehabilitation Services, 2007: 12)

According to the Colorado Department of Education (2007), deaf-blindness occurs in three out of 100 000 births. It should be noted that not all individuals who have deaf-blindness are completely deaf and blind; nor are all individuals who have deaf-blindness born with the condition (National Consortium on Deaf-Blindness; NCDB, 2006).

Knowledge of History

Although the first mention of deaf-blindness was in the literature at the time of the Crusades, the history of deaf-blind education began in the early 1800s in Paris, France, with a young lady named Victorine Morriseau. Laura Bridgman was the first deaf-blind person educated in the United States in 1837 (Collins, 1995). She was educated at Perkins School for the Blind. The technique used was finger spelling of the American deaf alphabet in the palm of her hand and it was developed by Samuel Howe. Using this hand alphabet technique, words and concepts about the student's environment and world were taught. Of interest, Anne Sullivan was Bridgman's teacher. Sullivan later went on to teach Helen Keller using the same technique developed at the Perkins School for the Blind. Helen Keller became an author and lecturer on the education of children who are deaf-blind. Anne Sullivan remained Helen Keller's

friend and teacher until her own death (Keller, 2003). From the 1900s to 1950, the United States opened multiple programs for individuals with deaf-blindness. These programs were established at Perkins's School for the Blind, the New York Institute for the Blind, Overbrook School for the Blind, Michigan School for the Blind, California School for the Blind, the Travis State School for the Deaf-Blind, and Illinois Braille and Sight Saving School (Collins, 1995). Famous schools that were established in Europe during the same time period were Condover Hall in the United Kingdom, Russia's Zagorsk School, and St. Michielsgestel in the Netherlands (see Collins).

The greatest incidence increase of deaf-blindness occurred during the 1963 worldwide Rubella epidemic. This large increase led to the term rubella bulge indicating a larger than normal prevalence of children who were deaf-blind in deaf schools, blind schools, and deaf-blind schools. An important fact as to the specialized education and placement of these children is that rubella caused not only deaf-blindness but multiple secondary disabilities. Many of these children were placed in classrooms and programs not for children with deaf-blindness, but for children with severe to profound multiple disabilities with mental retardation being the most common of the secondary disabilities. According to Collins, the result of the rubella bulge and the large increase in residential school populations in the United States and Europe led to increases in training and services provided to teachers for children who were deaf-blind. These services were mostly provided at the level of residential institutions. Since 1975, with the enactment of Public Law 94-142, the Education of All Handicapped Children's Act and the subsequent amendments, the number of individuals with deaf-blindness served in residential facilities has decreased (Hammer and Carlson, 1996). More than two decades ago, the National Consortium on Deaf-Blindness and National Information Clearinghouse for the Deaf-Blind (1988) provided services and information on deaf-blindness in the USA while worldwide the International Association for the Education of Deaf-Blind People provided services and education for individuals (Collins, 1995).

Understanding the Condition of Deaf-Blindness

Being blind or deaf alone presents many difficulties for anyone especially preschool and school-age children. Being born deaf-blind exacerbates the difficulties for normal language development and educational processes (Stremel and Schutz, 1995). During the first year and a half of life, infants gather information from sensory experiences/exploration and cognitively categorize this information into an internal language (Bower, 1982). All babies communicate; and through this, communication relationships are formed and sustained (Gleason, 2002). Around the age of 18 months, this internal knowledge of

the infant's world transfers into a categorical system that uses the language of the infant's environmental world whether spoken or signed by family members (Bower, 1982). According to Bruce (2005), within the world of the child who is congenitally deaf-blind, there may be no concept of distance, defined as the separation of self from others. Distance must be directly taught to the child through exploration with environmental objects via a hand-under-hand approach. The lack of visual and auditory inputs also affects a child's language development, motor development, cognitive development, and emotional and social development. Teaching parents and others how to overcome these challenges is the primary goal of early intervention (Gleason, 2002). While infants have five functioning senses with which to process information, the majority of this information is processed through visual and auditory processes. However, the deaf-blind child must process that same information through touch (Chen *et al.*, 2001; Chen and Haney, 1995; Miles, 2005). In essence, the infant is deprived of the majority of the sensory inputs that allows him/her to develop language concepts incidentally. Thus, it is not surprising that this infant will be linguistically and communicatively impaired (Stremel and Schutz, 1995).

The Office of Special Education Projects (OSEP) of the U.S. Department of Education annually collects demographic data on infants, children, and youth receiving special education services in the United States. Due to the high incidence of secondary disability associated with deaf-blindness, students who are deaf-blind may be categorized under multiple disabilities and thus may not be counted by OSEP under deaf-blind (NCDB, 2006). The NCDB (2006) conducted a census that counts children with deaf-blindness who have additional disabilities. Thus, the OSEP count is lower in number of students with deaf-blindness than the NCDB census data (see NCDB, 2006). These census data were collected on children and youth in the USA, ages from birth through 21 years, who were deaf-blind. The census included data on ethnicity, etiology, additional disabilities, type of degree of hearing and vision loss, special education classification and placement, as well as living setting (see NCDB). According to the NCDB, there were 9746 children from the ages of birth to 21 years who were deaf-blind and were receiving special education services in the USA. Of those, 5639 or 58% were white non-Hispanic/Latino; 1639 or 17% were Latino; 1413 or 14% were African-American; 389 or 4% were Asian or Pacific Islander; and 188 or less than 2% were American-Indian or Alaskan Native (see Table 1).

Regarding the OSEP policy of classifying children who are deaf-blind only in the deaf-blind category if there is no secondary disability, the secondary disability with deaf-blindness results in the student being placed under the category of multiple disability (NCDB, 2006). Under IDEA, legally, children are deaf-blind if they have such

Table 1 Ethnicity of the US school-age deaf-blind population

<i>Ethnic/cultural demographics</i>	<i>Census population</i>	<i>Percent of population (%)</i>
White non-Hispanic	5639	58
Latino	1639	17
African-American	1413	14
Asian or Pacific Islander	389	4
American-Indian or Alaskan Native	188	< 2
Other	478	5
Total	9746	95

Killoran, J. (2007). *The national deaf-blind child count: 1998–2005*. Manmouth, OR: National Technical Assistance Consortium for Children and Young Adults who are Deaf-Blind (NTAC), Teaching Research Institute, Western Oregon University. <http://nationaldb.org/NCDBProducts.php?prodID=57>.

severe communication and other developmental and learning needs that they cannot be appropriately educated in special education programs solely for children and youth with hearing impairments, visual impairments, or severe disabilities, without supplementary assistance to address their educational needs due to these dual, concurrent disabilities (IDEA, Sec.622, 1990). Diagnostically, to be classified under deaf-blind, students must be diagnosed with a visual acuity of 20/70 or worse in the better eye with correction and an auditory deficit of 30 dB or worse in the better ear (NCDB, 2006; National Eye Institute, 2008).

Understanding Hearing Loss

Hearing loss is documented by perception of loudness levels called decibels (dB) across frequency ranges or Hertz (Hz). The psychological perception or layman's term for frequency is pitch. The frequencies used in audiological testing are in octave levels of 250, 500, 1000, 2000, 4000, and 8000 Hz (Martin and Clark, 2000). Hearing loss is classified in degrees of hearing loss based on decibel levels of slight, ranging from 16 to 25 dB; mild, ranging from 26 to 40 dB; moderate, ranging from 41 to 50 dB; and moderate-to-severe, ranging from 50 to 70 dB; severe, ranging from 71 to 90 dB, and profound, ranging from 91 dB to worse (Martin and Clark, 2000). The 2006, NCDB deaf-blind census data reported on a range of hearing losses from mild (26–40 dB) to profound hearing loss (91 dB and above). Clearly, the NCDB (2006) data revealed that, of the 9746 students in US schools who were reported to have deaf-blindness, 1743 (18%) had profound hearing loss; 1401 (14%) had severe hearing loss; 1074 (11%) had moderate-to-severe hearing loss; 1332 (14%) had moderate hearing loss; and 1423 (15%) had mild hearing loss. A large number of children – 2978 (30%) – had either incomplete diagnosis or missing information.

Understanding Legal Blindness

Classifications of visual impairment are different from hearing impairment, in that, degrees of impairment are not reported. Instead, visual acuity and field of vision are the means of defining visual functioning (Corn and Koenig, 2000). Holbrook and Koenig (2006) argued that the criteria used to determine eligibility for government disability benefits do not necessarily indicate a person's ability to function. The criterion for legal blindness in the United States is having a visual acuity of 20/200 or worse in the better eye with corrective lenses (20/200 means that a person at 20 ft from an eye chart can see what a person with normal vision can see at 200 ft) or having a visual field restriction to 20° diameter or less (tunnel vision) in the better eye (National Eye Institute, 2008). The definition of legal blindness is for the US system; depending upon the country of student origin the definition may differ.

The NCDB (2006) census data for children who were deaf-blind revealed that 2084 (21%) had low vision; 2543 (26%) were legally blind; 213 (7%) had light perception only; and 658 (7%) were totally blind. Five percent were functionally blind and data on 1718 children were missing or incomplete. How deaf-blindness affects a child's development depends on several critical factors, according to Bennett *et al.* (1995). These include the age of onset, the degree and type of visual impairment and hearing loss, the stability of each sensory loss, and the educational intervention provided. Few individuals are actually completely deaf and/or blind (Ferrell *et al.*, 1990). There are varying degrees of residual hearing and vision. Most individuals who are deaf-blind rely on one of these senses more than the other, usually the sense with more residual ability. Many individuals are born with the loss of only one of the two distance senses and lose the other distance sense later in childhood (Gleason, 2002). Distance senses are those that bond a child with the world that extends beyond his/her individual body space. The young child who is deaf-blind does not have entry to opportunities for significant learning as do sighted and hearing children, and the information that the child does get from communication with others and the environment is often lacking continuity. Touch, taste, and the balance senses are close senses, giving information only about what is happening now, within arm's reach (Gleason, 2002).

Children born deaf-blind are more limited to these close senses than are individuals who are born with one of the distance senses then lose the other sense later. An example of this would be an individual who had either early hearing impairment and then later accompanying visual impairment, such as children who have Usher's syndrome, Cockayne syndrome, and Cornelia de Lange, Maternal Rubella, and STORCH. On the other hand, an

individual may have been born with or acquired early blindness with normal hearing but then acquires hearing loss through a progressive deafening condition (Karp and Santore, 1983; Mazzoni *et al.*, 2003).

Knowledge of Etiology

Eyes and ears develop during the first 12 weeks of pregnancy. They develop at the same time using the same types of embryonic cells and tissues and are autonomically similar in many ways. Therefore, it is easier to understand why many of the same conditions and genetic syndromes affect development of both organs causing deaf-blindness (Holbrook and Koenig, 2006). The NCDB (2006) reported the etiology data for children ages birth through 21 years with deaf-blindness who were receiving special education services. There were 3233 individuals who were deaf-blind due to nonsyndrome and syndrome heredity disorders, such as CHARGE, Down syndrome (Trisomy 21), and Usher's syndrome, I, II, III. Sixty-five percent or 2096 of the heredity classifications were nonsyndrome-type heredity disorders. There were 1853 children who were deaf-blind due to prenatal congenital conditions such as maternal rubella, cytomegalovirus, hydrocephaly, and microcephaly. There were 2715 children who were deaf-blind due to postnatal, noncongenital events such as asphyxia, encephalitis, meningitis, severe head injury, and complications of prematurity. One thousand fifty-four children had unknown etiologies (see **Table 2**). The most common causes of deaf-blindness are described below.

The NCDB (2006) census determined heredity as the most common cause of deaf-blindness. Deaf-blindness due to heredity falls under two categories: those genetic or chromosomal anomalies with which a specific set of physical characteristics are present and those in which the expression of the hereditary condition is deaf-blindness only. Some of the more common syndromes with deaf-blindness include Usher's syndrome, I, II, III, CHARGE syndrome, Trisomy 21, and III, Neurofibromatosis. Genetic codes for the effected gene for genetic types of deaf-blindness are specific to the gene and loci; for example, in Usher's syndrome, CDH23, MYO7A genes [28] and [29] and USH2A are the genes and loci (Shasty, 2000). Viruses can infect the fetus causing deaf-blindness. Cytomegalovirus causes damage to the fetus when the mother

is exposed to the virus. In addition, herpes virus can cause deaf-blindness. Finally, retinopathy of prematurity caused by scarring of the retina due to oxygen given to the premature infant is a well-known cause of deaf-blindness (Holbrook and Koenig, 2006).

Focusing on Communication

The greatest challenge for individuals with deaf-blindness is communication (Stremel and Schultz, 1995). Regarding deaf-blindness, Bjerkan (1996) related that communication is a form of contact in which meaning is conveyed by the use of signals that are perceived and interpreted by a partner. For some children, using vision, hearing, and touch all at one time is too confusing and causes the child to rely upon only one sense at a time. Communication often depends on which distance sense is more optimal. Tactile senses are often used to explore environments and communicate. Different communication modes are used depending on ability. These include sign language, body language, and gestures, as they are often the primary means of expressive language. Hand-over-hand or tactile sign language is used with those who have less visual efficiency. For example, someone who grew up deaf and experienced vision loss later in life is likely to use a tactile mode of a sign language; others who grew up blind and later acquired deafness are more likely to use a tactile mode for their spoken/written language.

Mode of communication is a term used for the choice or options for receptive and expressive communication in deaf education (Moore, 2001). Receptive communication refers to an individual's ability to receive and understand communicative intent from others, while expressive communication is the choice or option with which an individual chooses to convey his/her thoughts and ideas to others. Miles (2005) reported that some common forms of communication modes are:

- Use of residual hearing (speaking clearly and using hearing aids) or sight (signing within a restricted visual field and writing with large print).
- Tactile signing – sign language or a manual alphabet such as the American Manual Alphabet, or deaf-blind alphabet (also known as two-hand manual with tactile or visual modifications).
- Interpreting services (such as sign language interpreters or communication aides).
- Communication devices such as Tellatouch – a manual Braille writer.
- Braille reading.
- Large-print reading.
- Tactile communication cards.

These are used by individuals when other options are not available (Engleman *et al.*, 1992; Miles, 2005). They may be used in conjunction with other modes of communication

Table 2 Etiologies of US school-age deaf-blind population

<i>Etiology</i>	
Heredity	3233
Prenatal congenital	1853
Postnatal noncongenital	2715
Unknown	1054
Total	7845

depending on the abilities of the individual with whom the deaf-blind person needs to communicate. Tactile cards are unique with each card representing an individual meaning. They may be used in the classroom, at home, in school, and in public for shopping, going to work, daily routines, and other traveling (Bourguin and Sauerburger, 2005). An example of the use of these cards is as follows:

An individual who is deaf-blind wants to go shopping and will need to take the bus to the mall. He/she will have previously gone through training for traveling around town and shopping for himself or herself. Cards have been made and arranged so that the individual can exit his or her apartment building and walk to the bus stop. At the bus stop he/she takes a card from his or her pocket that has a line drawing on it that has been memorized to represent the bus stop. In this situation, the individual who is deaf-blind holds up a card in an attempt to get the attention of a person at the bus stop. Printed on the card are the words, "Please, let me know when bus # 2 arrives." Another card in the individual's set of cards is marked for the bus driver. Printed on that card are the words, "Please, let me know when we are at the bus stop for the Mall." The deaf-blind individual hands the card to the bus driver as he/she gets on the bus. Other cards have been prepared for each step of the adventure. The drawback of this method is that someone may not be available to help the individual who is using the cards or that strangers may not readily volunteer to assist the individual who is deaf-blind. For this reason the individual must have practiced this routine with an assistant many times before they feel comfortable with the process. (Bourguin and Sauerburger, 2005)

Early Intervention of Students Who Are Deaf-Blind

Gleason (2002) reported that parent-infant training is imperative for families with children who have deaf-blindness. Parent-infant training is usually performed in the home of the infant who is deaf-blind. A parent-infant training program, SKI-HI, initially designed for deaf infants to educate parents and promote language development in deaf infants provides parent-infant services for families with infants who are deaf-blind (HOPE, 2006). In SKI-HI training, a parent-infant trainer works with the parents and educates them on the disability itself, communication issues and choices, mobility and orientation, anatomy of vision and hearing, language development, amplification options, sight development/visual awareness, functional living skills (dressing, feeding, and hygiene), and hearing development/sound awareness.

As mentioned previously, most deaf-blind children have some residual vision/hearing/vision and hearing. Placement in schools must be consistent with the child's need. Some children who are deaf-blind will be placed in

a deaf school as their vision is their primary sensory system for communication. They will use American Sign Language or other manual form of communication, while some children will perform more as a blind child and will be educated using a curriculum for the blind as their auditory and possible oral verbal system will be used for communication. Finally, some children who are deaf-blind will have multiple impairments such that education in either schools/institutions for the deaf, blind, and deaf-blind are not appropriate for the best education. These children may be placed in schools/classrooms for children who have severe-to-profound multiple impairments (Fredericks and Baldwin, 1987).

The NCDB (2006) noted the placement of children in the USA, labeled under the category of deaf-blind as following: Regarding infants and toddlers, 75% of infants ages 0–2 receive parent-infant services in the home. The remaining 25% are served in specialized service settings, residential facilities, and hospitals. Seventy-two percent of preschool services for ages 3–5 are provided in special education classroom separate schools, or residential schools; while 20% of the 3–5 children with deaf-blindness are served in general preschool inclusive classrooms; and 5% are educated at home. Finally, for school-aged children 6–21 years old, only 15% of children who are deaf-blind are educated in general education classrooms. Thirty-nine percent of these children are educated in separate classrooms, while 16% are educated in separate public schools. Ten percent of the children who are deaf-blind are educated in public or private residential facilities while 8% are educated in separate private schools. The remaining 7% and 5% are educated in hospitals and at home, respectively.

Multiple disabilities are common for children who have deaf-blindness; according to the NCDB census, children with deaf-blindness have been reportedly classified under every disability category in IDEA Part B, with the majority under multidisabled. The most common secondary disability to the deaf-blindness was reported as cognitive impairment (NCDB, 2006).

Targeted Inside Classroom Techniques

Consistent orientation and classroom order are critical in the educational setting for students who are deaf-blind. The student begins his or her day in an order of tasks that are laid out in a specific place and in a specific order. For instance, as the student enters his/her classroom, the student approaches a specific shelf that contains objects that represent activities that he/she will do during the day. A spoon may be used to indicate eat breakfast. The next item might be a toothbrush to indicate that the student should brush his/her teeth after eating breakfast. These items may be more abstract for students with more

cognitive abilities. That is, sometimes cards with Braille are used, or cards with raised lines or images; and they can also be used to communicate with others. Regardless, this shelf routine is consistent across the literature in the field of deaf-blindness (Engleman *et al.*, 1998).

Transitioning Students Who Are Deaf-Blind

Transition is a necessary avenue for persons who are deaf-blind as they near their formal education years and prepare for finding suitable work and living accommodations. We know that over 70% of persons with disabilities are unemployed and of the 29% employed, many work part time (Harris, 1998). For a person who is deaf-blind, this new era will require more than just one person or one agency. Collaboration between specialists and agencies is necessary to assist with such things as housing, rehabilitation and vocational requirements, deafness, blindness, orientation and mobility, medical needs, and mental health (Ferrell, 2008). The individual's interest, goals, and abilities are key in the planning of this transition. The person that is deaf-blind must be given the opportunity to self-advocate for his/her interest and direction in adult living/learning that he/she wishes to pursue (Agran *et al.*, 2007). Family members as well as skilled interpreters and friends can assist in assuring that adults have voices in their future.

To facilitate the transition process, many assistive and augmentative communication devices are necessary and have been developed for those individuals who are deaf as well as those who are blind. These devices include communication boards, alerting devices such as vibrating alarms that can be placed under a mattress, flashing lights for those who have more vision, hearing aids, cochlear implants (Turnbull *et al.*, 2007a) computers with closed-circuit televisions that can display large print, computer programs with synthesized speech and input from the keyboard in Braille, a Braille embosser, and various brands of print readers (Turnbull *et al.*, 2007b). The deaf-blind individual may use a combination of these devices. Assistive devices and their descriptions are presented elsewhere in the encyclopedia.

Conclusion

Deaf-blind is the presence of a dual sensory impairment involving the auditory and visual senses. The condition of deaf-blindness consists of varying degrees of hearing loss and visual impairment. Federal laws and OSEP projects have provided programs for children with deaf-blindness. Unique communication modes are necessary for these children as are stated above. Education consists of early intervention, unique classroom techniques for childhood education, and assistive and augmentative devices which

are important for the successful transition of these children into the adult world.

Bibliography

- Agran, M., Hong, S., and Blankenship, K. (2007). Promoting the self-determination of students with visual impairments: Reducing the gap between knowledge and practice. *Journal of Visual Impairment* **101**, 453–464.
- Bennett, K., Griffin, H., Powers, J., *et al.* (1995). *Delivering Effective Instruction to Students with Deaf-Blindness and/or Other Severe Disabilities*. Raleigh, NC: State Department of Public Instruction.
- Bjerkman, B. (1996). When do congenital deafblinds communicate? On the distinction between communication and other types of social contact. In Laurent, M. (ed.) *Communication and Congenital Deafblindness: The Development of Communication. What Is New?* pp 179–195. Paris: Centre National de Suresnes.
- Bourguin, E. and Sauerburger, D. (2005). Teaching deaf-blind people to communicate and interact with the public: Critical issues for travelers who are deaf-blind. *Review* **37**, 109–116.
- Bower, T. G. R. (1982). *Development in Infancy*, 2nd edn. San Francisco, CA: W. H. Freeman.
- Bruce, S. M. (2005). The application of Werner and Kaplan's concept of 'Distancing' to children who are deaf-blind. *Journal of Visual Impairment and Blindness* **99**, 464–477.
- Chen, D., Downing, J., and Rodriguez-Gill, G. (2001). Tactile strategies for children who are deaf-blind: Consideration and concerns from Project SALUTE. *Deaf-Blind Perspective* **8**, 1–6.
- Chen, D. and Haney, M. (1995). An early intervention model for infants who are deaf-blind. *Journal of Visual Impairment and Blindness* **89**, 213–221.
- Collins, M. T. (1995). History of deaf-blind education. *Journal of Visual Impairment and Blindness* **89**, 210–212.
- Corn, A. and Koenig, A. (1996). *Foundations of Low Vision: Clinical and Functional Perspectives*. New York: AFB Press.
- Engleman, M. D., Griffin, H. C., and Wheeler, L. (1998). Deaf-blindness and communication: Practical knowledge and strategies. *Journal of Visual Impairment and Blindness* **92**, 783–798.
- Ferrell, K. (2008). *Transition of Youth with Visual Impairments, Multiple Impairments, or Deaf-Blindness: National Longitudinal Transition Study*. Baltimore, MD: American Foundation for the Blind. <http://www.afb.org/Section.asp?SectionID=43andDocumentID=3178> (accessed August 2009).
- Ferrell, K., Trief, E., Dietz, S. J., *et al.* (1990). Visually impaired infants research consortium (VIIRC): First-year results. *Journal of Visual Impairment and Blindness* **84**, 115–117.
- Fredericks, H. D. and Baldwin, V. L. (1987). Individuals with sensory impairments: Who are they? How are they educated? In Goetz, L., Guss, D., and Stremel-Campbell, K. (eds.) *Innovative Program Design for Individuals with Dual Sensory Impairment*, pp 3–12. Baltimore, MD: Paul H. Brookes.
- Gleason, D. (2002). *Early Interaction with Children Who Are Deaf-Blind*. The National Information Clearinghouse on Children Who Are Deaf-Blind, Monmouth, OR. www.dblink.org/lib/early2.htm.
- Hammer, E. and Carlson, R. (1996). Using the intervening model with adults who are deaf-blind. *Journal of Vocational Rehabilitation* **7**, 125–128.
- Harris, L. (1998). American with disabilities still face sharp gaps in securing jobs, education, transportation, and many areas of daily life. *See/Hear* **3**, 32–34.
- Holbrook, M. and Koenig, A. (2006). *Foundation of Education: History and Theory of Teaching Children and Youths with Visual Impairments*, 2nd edn. New York: AFB Press.
- HOPE (2006). *Resources and Materials*. <http://www.skihi.org/Research>.
- Karp, A. and Santore, F. (1983). Retinitis pigmentosa and progressive hearing loss. *Journal of Speech and Hearing Disorders* **48**, 308–314.
- Keller, H. (2003). *The Story of My Life*. Sewickley, PA: American Foundation of the Blind Press.
- Martin, F. N. and Clark, J. G. (2000). *Introduction to Audiology*, 7th edn. Des Moines, IA: Allyn and Bacon.

- Mazzoni, A., Dubey, S., Poletti, A., and Colombo, G. (2003). Sporadic acoustic neuroma in pediatric patients. *International Journal of Pediatric Otorhinolaryngology* **71**, 1569–1572.
- Miles, B. (2005). *Literacy for Persons Who Are Deaf-Blind*. Monmouth, OR: DB-LINK, The National Information Clearinghouse of Children Who Are Deaf-Blind.
- Moore, D. (2001). *Educating the Deaf: Psychology, Principles, and Practices*, 5th edn. Boston, MA: Houghton Mifflin.
- NCDB (National Consortium on Deaf-Blindness) (2006). Children who are deaf-blind. *Practice Perspectives – Highlighting Information on Deaf-Blindness* **2**, 1.
- National Eye Institute (2008). *Report of the Visual Impairment and its Rehabilitation Panel*. http://www.nei.nih.gov/resources/strategicplans/neiplan/frm_impairment.asp (accessed August 2009).
- Office of Special Education and Rehabilitation Services (2007). *Child with a Disability*. *Federal Register*, 34 CFR Part 300.8, Chapter III, pp 12. U.S. Department of Education.
- Shasty, B. S. (2000). Mammalian cochlear genes and hereditary deafness. *Microbial and Comparative Genomics* **5**(2), 61–69.
- Killoran, J. (2007). *The national deaf-blind child count: 1998–2005*. Monmouth, OR: National Technical Assistance Consortium for Children and Young Adults who are Deaf-Blind (NTAC), Teaching Research Institute, Western Oregon University. <http://nationaldb.org/NCDBProducts.php?prodID=57>.
- Stremel, K. and Schutz, R. (1995). Functional communication in inclusive settings for students who are deaf-blind. In Haring, N. and Romer, L. (eds.) *Welcoming Students Who Are Deaf-Blind into Typical Classrooms: Facilitating School Participation, Learning, and Friendships*, pp 197–229. Baltimore, MD: Paul H. Brookes.
- Turnbull, A., Turnbull, R., and Wehmeyer, M. L. (2007a). Understanding students with hearing loss. In *Exceptional Lives: Special Education in Today's Schools*, pp 336–367. Upper Saddle River, NJ: Pearson.
- Turnbull, A., Turnbull, R., and Wehmeyer, M. L. (2007b). *Understanding students with visual impairments*. In *Exceptional Lives: Special Education in Today's Schools*, pp 368–397. Upper Saddle River, NJ: Pearson.
- Programs for Deaf-Blind Children and Adults (1997). *American Annals of the Deaf* **142**, 227–232.
- Salem-Hartshorne, N. and Jacob, S. (2004). Characteristics and development of children with CHARGE Association-Syndrome. *Journal of Early Intervention* **26**, 292–301.
- Sauerburger, D. (1997). *Orientation and Mobility for Deaf-Blind People*. The Free Library, [http://www.thefreelibrary.com/Orientation and mobility for deaf-blind people](http://www.thefreelibrary.com/Orientation+and+mobility+for+deaf-blind+people).
- U.S. Department of Education, Office of Special Education Programs. (2006). Severe Disabilities Branch, Services for Children with Deaf-Blindness Program, IDEA, Part C; Section 622; CFDA 84.025 Washington DC.
- Wheeler, L. C. and Griffin, H. C. (1997). A movement-based approach to language development in children who are deaf-blind. *American Annals of the Deaf* **142**, 387–390.

Relevant Websites

- <http://www.azdb.net> – Azdb.net.
- <http://www.cadbs.org> – California Deaf-Blind Services.
- <http://www.cbss.umd.edu> – Connection Beyond Sight and Sound Maryland's Deafblind Project.
- <http://www.chsbs.cmich.edu> – DeafBlind Central: Michigan's Training and Resource Project.
- <http://www.esu3.org> – Educational Service Unit #3.
- <http://www.deafblind.ufl.edu> – Florida Outreach Project.
- <http://education.gsu.edu> – Georgia Sensory Assistance Project.
- <http://www.iadeafblind.k12.ia.us> – Iowa's Deafblind Services Project.
- <http://www.kansped.org> – Kansas State Department of Education Special Education Services.
- <http://www.dbproject.mn.org> – Minnesota DeafBlind Project.
- <http://www.usm.edu> – Mississippi Deaf-Blind Project.
- <http://www.msb.k12.mo.us> – Missouri School for the Blind.
- <http://kc.vanderbilt.edu> – Monroe Carell Jr. Children's Hospital at Vanderbilt TREDs Deafblind Project.
- <http://www.opi.state.mt.us> – Montana IDEA Deaf-Blind Services.
- <http://www.unr.edu> – Nevada Dual Sensory Impairment Project.
- <http://www.state.nd.us> – North Dakota Dual Sensory Project.
- <http://www.ou.edu> – Oklahoma Deaf-Blind Technical Assistance Project.
- <http://www.pattan.net> – Pennsylvania Training and Technical assistance Network.
- <http://www.usd.edu> – Sanford School of Medicine Center for Disabilities.
- <http://www.sesa.org> – Special Education Service Agency.
- <http://www.k12.wy.us> – Special Education.
- <http://www.tc.edu> – Teachers College Columbia University: New York State Technical Assistance Project (NYSTAP) Serving Children and Youth Who Are Deafblind.
- <http://www.tsbvi.edu> – Texas Deafblind Project.
- <http://www.sscs.org> – The Ohio Center for Deafblind Education.
- <http://cdd.unm.edu> – The University of New Mexico Health Sciences Center: Center for Development and Disability.
- <http://www.pathfindersofal.org> – Website Judge.
- <http://www.dpi.state.wi.us> – Wisconsin Department of Public Instruction.
- <http://www.educ.uidaho.edu/cdhd/CYWDB>.
- <http://www.usdb.org/departments/DeafBlind/default.asp>.
- <http://www.pdec.state.sc.us/deafblind/home.htm>.
- <http://www.ode.state.or.us/search/results/?id=185>.
- <http://www.uvm.edu/~cdci/programs/iteam/itdb.html>.
- <http://www.wsdsonline.org/db.html>.
- <http://web.indstate.edu/soe/blumberg/Deafblind.html>.
- <http://www.ncpublicschools.org/ec/exceptionality/deafblind>.

Further Reading

- Bakhurst, D. and Padden, C. (1991). The Meshcheryakov experiment: Soviet work on the education of deaf-blind children. *Learning and Instruction* **1**, 201–215.
- Bruce, S. M. (2002). Impact of a communication intervention model on teachers' practice with children who are congenitally deaf-blind. *Journal of Visual Impairment and Blindness* **96**, 154–168.
- Griffin, H. C. and Davis, M. L. (2004). Charge syndrome: Educational and technological interventions. *Review* **35**, 149–157.
- Janssen, M. J., Riksen-Walraven, J. M., Dijk, J. P. M. V., Ruijsenaars, W. A. J. J. M., and Vlas Kamp, C. (2007). Team interaction coaching with educators of adolescents who are deaf-blind: Applying the diagnostic intervention model. *Journal of Visual Impairment and Blindness* **101**, 677–689.
- McKenzie, A. R. and Davidson, R. (2007). The emergent literacy of preschool students who are deaf-blind: A case study. *Journal of Visual Impairment and Blindness* **101**, 720–725.
- Morgan, S., Bixler, E., and McNamara, J. (2002). Self-determination for children and young adults who are deaf-blind: *NTAC Briefing Paper*. Monmouth, OR: National Information Clearinghouse on Children Who Are Deaf-Blind. <http://www.tr.wou.edu/ntac> (accessed August 2009).
- Plann, S. (2008). 'Bad things': Child abuse and the nineteenth-century Spanish national school for the deaf and blind. *Sign Language Studies* **8**, 181–210.

Educating Children with Deafness and Hearing Impairments

G M Mukuria, Slippery Rock University, Slippery Rock, PA, USA

C J Eleweke, Radford University, Radford, VA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Audiologist – A professional trained in the testing of hearing, detecting hearing defects and involved in their treatment or management.

Deaf – One who has a severe or profound hearing loss.

Hard of hearing – One who has a mild or moderate hearing loss.

Hearing loss – A dysfunction of the auditory system.

Neuroplasticity – The brain's natural ability to form new connections in order to compensate for injury or changes in one's environment.

options and perspectives are continually provided to families and their children with hearing loss, their teachers, and other professionals in the field. Controversies regarding the best strategy for instructions and services to address the various developmental needs of children with hearing loss have continued to be prevalent. However, we believe these children can achieve success in education if such controversies are set aside, and efforts are focused on providing appropriate services that will foster their development. We also believe these services should provide (1) support for families, (2) appropriate and effective early intervention, (3) language development and communication skills, (4) development of literacy skills, and (5) educational needs for multicultural learners.

Educating Children with Deafness and Hearing Impairments

In this article, important factors for understanding and enhancing the educational experiences, and outcomes for learners who have hearing loss are highlighted. We find it prudent to use hearing impairments/loss to include deafness and hard of hearing as a broad generic term for any hearing dysfunction that precludes successful processing of linguistic information through the ears regardless of severity, etiology, and age of onset. A discussion of various causative factors of hearing loss is beyond the scope of this work. A review of these factors can be found in Cole and Flexer (2007) and Marschark (2007). We are cognizant of the fact that the different degrees of hearing loss: mild, moderate, severe, and profound clearly suggest that children who have hearing loss have different communication and educational needs.

In the United States, Canada, United Kingdom, and other Western countries, there have been concerted efforts toward developing strategies for addressing educational needs of individuals with hearing impairments over the past decades. This has been due to increased understanding of people with such impairments, their communication methods, and ongoing changes and progress in pedagogy, psychology, psycholinguistics, and technology. Evidence suggests that more progress has been made in the education of children with hearing loss in the United States during the last 30 years than in the previous 300 years (Marschark *et al.*, 2002). Consequently, new

Support for Families

Evidence is consistent that the provision of support to enable families to participate actively in the development and education of their children with hearing loss is imperative (Marschark, 2007). A major argument for providing support to ensure active involvement of the parents in the education of their children with hearing loss is that 95% of these children are born to hearing parents who have no previous experience and knowledge of raising a child with hearing loss (Byrne, 2000).

The diagnosis of hearing loss in a child can be emotionally devastating to parents (Scheetz, 2001). In addition to their varied emotional reactions to the diagnosis, these parents also are concerned as to whether their children will be able to use speech, speak normally, go to school, obtain employment, and have a family. These parents need to be educated regarding childhood hearing loss and its consequences, and about the special needs these children require to ensure optimum development. They also need to know the options in communication and education available to them and their children (Marschark, 2007). Furthermore, evidence indicates that the support provided to families could make the biggest differences in their lives by empowering them to take the initiative to seek services that would help them and their children (Turnbull *et al.*, 2004). The family-centered approach calls for collective empowerment rather than a more traditional relationship where the professional had power over the family. Parents and families must become part and parcel of the decision-making mechanism. The fundamental beliefs

underlying focus on family and community are based on the premise that families are interrelated social systems with unique characteristics and needs. Whatever affects one family member has ripple effects directly or indirectly on other family members and community. Consequently, teachers and other service providers should consider the family constellation as the appropriate focal point of professional attention. Turnbull *et al.* (2000) concluded that the study of family must take into account family dynamics that include family characteristics, family interactions, family functions, and family cycles.

Although the importance of providing support for families is recognized in the literature, evidence indicates that, in many instances, such support may not be provided (Lane *et al.*, 1996). Professionals tend not to be concerned about the family as a whole (Lane *et al.*, 1996). They lack the resources to adopt a family-centered or ecological approach in service delivery. Their focus is often on the child's disability and not on the needs and wishes of the parents and other family members. The failure to provide support to families has adverse consequences. For instance, a family's ability to adapt to the child's hearing loss may be hindered by the lack of information about available services. Evidence clearly indicates that many parents are still not receiving adequate information and support during the critical early development years, and many remain uninformed and unable to assist their children at home. Lacking information, support, and knowledge, many parents are unable to make appropriate choices about the communication and educational needs of their children (Davila, 2004). Providing quality information is the critical modality of supporting families with young children who have hearing loss. When clear and unbiased information is provided, they will make informed choices that will address the developmental needs of their children. However, concerns remain about the nature of the information professionals provide to families. Evidence indicates that many parents are aware that they are not getting all the information they need, and that some of the information they receive is patently biased (Marschark, 2007).

Evidence from the Colorado Home Intervention Program (CHIP) and studies in the United Kingdom indicate that families with young children with hearing loss could benefit immensely from programs provided within their homes (Yoshinaga-Itano, 2004). Parents could be visited by professionals such as teachers, speech and language pathologists, audiologists, early childhood special educators, bilingual educators, psychologists, and social workers. These professionals provide information about counseling, developmental assessment, communication, and educational options. Targets should be established for timing of initial visit by professionals to the family of a newly diagnosed child with hearing loss. The visits should be carefully planned and conducted with the goal of giving unbiased information and empowering parents so that they can make

informed choices. Some parents may be experiencing emotional reactions to the diagnosis. Therefore, having sensitive professionals around who are willing to listen and answer questions clearly and without bias could significantly impact how quickly parents adjust to challenges of raising a child with a hearing loss. It is critical that professionals give contact information about local and national organizations concerned with hearing loss to the parents. These professionals, especially those with hearing loss, should be included in the team that makes the initial visit to families since they have first-hand experience of living with a hearing loss. Consequently, apart from providing information and mentoring families, they could give hope and inspiration to families and their children. Additionally, it is imperative that parents with newly diagnosed children have the opportunity of meeting and chatting with other parents who have been or are currently working their way through similar issues. These parents provide mutual and social supports to each other to cope better in meeting the challenges of having a child with a hearing loss. Such parents manifest better behavioral interactions and greater sensitivity to their children's communication and other needs (Turnbull and Turnbull, 2006).

Appropriate and Effective Early Intervention

Hearing loss has been described as one of the most common childhood disabilities in the United States. Cole and Flexer (2007) estimated that due to technological advances in testing hearing in infants and children under Universal Newborn Hearing Screenings, between 16 000 and 18 000 babies and toddlers are diagnosed with hearing loss annually in the United States. Early identification and prompt initiation of early intervention services could prevent or greatly reduce the communication barriers posed by hearing loss. These services are necessary because of their focus on language development, parent-child communication, social skills development, and support for the utilization of any residual hearing children may have. Furthermore, they provide parents with strategies for enhancing their children's educational development through sign language instruction, speech training skills, or both depending on the particular program (Mauk and White, 1995). The goal of early intervention must be to foster effective parent-child communication starting soon after diagnosis of hearing loss. This kind of communication is the best single predictor of success in virtually all areas of development of children with hearing loss, including academic achievement (Marschark, 2007). Clearly, effective interactions, with the most appropriate means of communication, between caregivers and the child will provide opportunities for involvement in most normal activities of children in both social and academic areas.

Learning becomes possible in humans by consistent language input and models. Thus, an important goal of

early intervention should be to provide children with hearing loss with the opportunity to acquire language in the appropriate modality (sign or spoken) by having accessible and competent language models. Consistent means of communication, in the appropriate modality that is congruent for the child, will create opportunities for positive social interactions critical for the social-emotional development and educational success of all children including those with hearing loss. Evidence indicates that children with hearing loss who are identified early and participate in intervention programs achieve significantly better language, speech, social, and emotional developments than those who are not identified early and who receive intervention at a later time (Yoshinaga-Itano, 1998). To provide effective early intervention for children and their families, the program should not be influenced by the controversy regarding the best communication approach that has characterized the education for the hearing impaired for decades. There is a common agreement with Marschark (2007) that there is no single correct answer about the best or superior method of communication for children with hearing loss. The communication and other needs of children are different and therefore, no one method can be applied to all or even most of them. The important thing is to establish, as early as possible, an effective mode of parent-child communication, one that the child can access fully (Calderon, 2000). This is the foundation for successful social, emotional, and academic developments of children with hearing loss. Through objective assessment, open-mindedness, and willingness to be flexible, the mode of communication that would be appropriate for each child could be found.

Language and Communication Skills Development

How should children with hearing loss communicate? What language should be used in educating learners with hearing loss? These questions sum up the main thrust of the controversies over the past two centuries and presently in the education of learners with hearing loss.

The problem with hearing loss is that it prevents sound signals from reaching the brain (Cole and Flexer, 2007). With technological advances in the development of amplification devices (e.g., hearing aids and cochlear implants), the auditory neural connections throughout the brain could be accessed, stimulated, and enhanced for the development of spoken language, reading, and academic skills in children with hearing loss. According to Cole and Flexer, listening experiences in infancy, especially during the first three and half years of life when neuroplasticity is greatest are important. Similarly, lots of practices are necessary for the development of speech and language in children with hearing loss. Parents should be skilled in providing a consistent listening and spoken language-learning

environment to foster the development of speech and language skills in their children.

On the other side of the argument, there are those who insist that learning and using sign language can fulfill all the requirements for language development and education for children with hearing loss (Marschark, 2007). Sources (e.g., Marschark) have observed that only a minority of children with hearing loss are able, with intensive aural/oral approaches, to acquire fluent spoken language. This could be because the task of learning to speak may be easier for those children with lesser degrees of hearing loss. However, for those with profound hearing loss, Marschark argues that the process of acquiring speech that can be understood by others is difficult and frustrating. Thus, sign language could be a better alternative for these children even if they are fitted with amplification devices. So far, there is no evidence which shows that using sign language interferes with the development of spoken language. We agree with Marschark that children are different in their abilities. Thus, while some children with hearing loss could develop spoken language skills and could be educated in oral programs with intensive auditory-verbal support, other children with hearing loss may be unable to do so. Forcing the procedures to develop speech on them could be frustrating, destructive to their self-image, and depriving them of options that would have allowed them to experience a more normal childhood. The concern of professionals and parents should be to ensure early success in language development and communication for children with hearing loss in whichever modality that works for each individual child. As such, it must be restated that to achieve this goal, parents must be provided unbiased information by professionals about these communication methods in terms they can understand and remember. This will enable them to make informed choices that will benefit their children and themselves.

Development of Literacy Skills

It is common knowledge that enhancing the literacy and academic achievements of students who have hearing loss has remained a major concern of educators and professionals in the field (Rodda and Eleweke, 2000). Good literacy skills (i.e., the ability to read and write well) are essential for academic achievement. This is because so much information in texts and other educational materials is in the printed form of the spoken language (e.g., English). Although learners with hearing loss can utilize American sign language (ASL) for meaningful communication and education, there is no generally accepted written medium for ASL as yet. Consequently, learners with hearing loss must acquire good English literacy skills in order to perform successfully at school like their normally hearing peers. It remains the case, however, that acquiring English

literacy skills is difficult for many learners who have hearing loss. The acquisition of a high level of critical literacy (i.e., the literate use of language for problem solving and communication) is dependent upon the acquisition of the conversational forms (spoken/signs) of language at the earliest age possible.

Clearly, a child who has hearing impairment from birth or thereafter, is unable to hear the speech of other people and will not learn speech and language spontaneously, as do children with normal hearing. This ability is critical for language acquisition and development because it enables them to monitor and assess their voices and those of others. A child with hearing loss does not have that ability! All the challenges that hearing loss poses to learning the vocabulary, grammar, and syntax of the English aside, being unable to hear one's own voice makes it difficult to assess and monitor language. Consequently, children with hearing loss may speak too loudly or not loudly enough. They may speak in an abnormally high pitch or sound like they are mumbling because of poor stress, poor inflection, or poor rate of speaking.

Since reading and writing involve graphic representations of a phonologically based language, the child with hearing impaired must strive to decode and produce text based on language for which she/he may have little or no understanding (Hulit and Howard, 2006). Thus, it is argued that the reading and writing difficulties of learners with hearing loss could be attributed to the problems encountered in acquiring the conversational form (spoken and/or signs) of the language (English) in which they are attempting to read and express via writing (Paul, 1998). Nonetheless, literacy is important for academic success and effective functioning in the larger hearing society by those with hearing loss. However, the basic question remains unanswered: What approach/strategy could produce useful outcomes for enhancing the literacy skills and academic achievement of learners with hearing loss?

Recent trends in the education of learners with hearing loss are focusing on bilingual models of teaching literacy. Paul (1998) observed that no widely accepted definition for bilingualism in the education of learners with hearing loss exists. For the purpose of this entry the term will be taken to refer to proficiency in two languages (ASL and print English). As a result of hearing loss which may preclude affected learners from acquiring English as a first language, some workers have argued for the development and use of language/English bilingual programs (see Paul, 1998). Essentially, it is considered that the bilingual approach could enhance the knowledge and use of written/spoken language, and enrich the social, emotional, and academic experiences of learners with hearing loss. Strategies for the development and implementation of bilingual programs for learners with hearing loss vary and are influenced by various philosophical and theoretical underpinnings. However, the objective remains the same,

what is to facilitate the development of English literacy skills (L2) in learners with hearing loss using sign language as the first language (L1). We agree with the approaches and specifically recommend some of the essential characteristics of a bilingual approach for teaching English literacy skills to learners with hearing loss outlined by Paul. These include presenting English via print only, by using (1) reading and writing activities and (2) ASL to explain the grammar of English. In other words, students use their knowledge and skills in their first language (signs) only to learn literacy skills in the second language (English). Careful planning is necessary to develop bilingual programs that would meet the needs of different learners with hearing loss. This requires the establishment of (1) realistic diagnostic assessment of educational needs, (2) instruction that focuses on specific needs, (3) assessment of growth, and (4) evaluation of instruction. Furthermore, it is important to define precisely what the student is currently doing, what he/she needs to know, and how it is best taught. We suggest that in order for such literacy programs to be effective, teaching of reasoning and literacy should be closely intertwined. To achieve this goal will require paying appropriate attention to results of previous research, the methodologies of bridging and metacognition, alternative forms of assessment, and the selection of programs or resources that are tested and supported by valid research. Addressing these issues would enable educators provide appropriate opportunities to learners with hearing loss and help them to achieve their full potential regarding literacy in its broadest definition.

Multicultural Learners with Hearing Loss

The number of culturally and linguistically diverse (CLD) children with hearing loss has continued to increase in multiethnic countries like the United States, Canada, and United Kingdom. A child with hearing loss from CLD may be faced with formidable challenges to accessing appropriate education and other services that would enhance the development of their potentials (Chinn *et al.*, 2007). Due to the lack of early diagnosis and timely provision of intervention services, CLD learners start school with lower language skills than their white counterparts. This lack of basic language skills significantly affects these children's access to educational opportunities. At home, parents may not be English-speakers nor (American) sign language users which, decreases the child's opportunities for consistent language experiences at home. CLD parents may lack adequate understanding of how the system works, or where to go for assistance and support. It remains the case that many educators do not have the knowledge of diversity and skills to effectively work with CLD learners hearing loss and their families.

Educational authorities and personnel should set up programs that focus on providing equal opportunities for

CLD learners with hearing loss and their families. Teacher training programs should include courses on diversity, multiculturalism, and ethnicity in their curriculum. This, we believe, will ensure that preservice teachers have the competences to provide appropriate educational experiences for those CLD children with hearing loss. Furthermore, workshops and seminars on working with these learners should be conducted regularly for teachers and service providers in all educational programs. With increase need awareness and knowledge of the needs of CLD learners, teachers and service providers would be more effective in educating them. In addition, specific projects aimed at fostering the acquisition of English and sign language (e.g., ASL), should be established in the communities.

Conclusion

Despite the tremendous progress being made in educating children with hearing loss, controversies as to which is the best method of communication still exists. Addressing these controversies demands objective use of research findings in planning and implementing educational programs for children with hearing loss. Commitment to continually learn more from sound research would reduce these debates and ensure that valuable time and resources are channeled to enhancing the educational experiences and outcomes for these learners. Evidence is clear that children with hearing loss will achieve success in education when their parents or families have the knowledge and skills to support them and ensure that they have equal opportunities. To achieve this goal requires that parents receive unbiased information about the developmental needs of their children so that they can make informed choices on issues such as method of communication and educational placement. When parents are provided information that is not biased on the various options available, we believe they can make informed decisions. The challenge is for educators and professionals in the field to rethink their methods and explore strategies that will result in optimizing potential for all families. They can do this if they uphold the need for diverse approaches to language development and recognize individual differences and needs for all learners with hearing loss. If they do this, it will be possible that families of these children receive (1) supports they need, (2) appropriate early intervention services which are available, (3) the supports for effective communication and language skills, and (4) appropriate strategies to facilitate literacy skills and academic achievement. Clearly, the stakes are too high to permit anything less.

Bibliography

Byrne, M. R. (2000). Parents with severe to profound hearing loss. *Communication Disorders Quarterly* 21, 210–223.

- Calderon, R. (2000). Parents involvement in deaf children's education programs as predictors of child language, early reading, and social-emotion development. *Journal of Deaf Studies and Deaf Education* 5, 140–155.
- Chinn, K. M., Eleweke, C. J., and Lopez, E. (2007). Working with multicultural learners with hearing loss. In Obiakor, F. E. (ed.) *Multicultural Special Education: Culturally Responsive Teaching*, pp 166–179. Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.
- Cole, E. B. and Flexer, C. (2007). *Children with Hearing Loss Developing Listening and Talking: Birth to Six*. San Diego, CA: Plural.
- Davila, R. R. (2004). Reviewing the past, assessing the present, and projecting the future. In Powers, D. and Leigh, G. (eds.) *Educating Deaf Children: Global Perspective*, pp 3–12. Washington, DC: Gallaudet University Press.
- Hulit, L. M. and Howard, M. R. (2006). *Born to Talk: An Introduction to Speech and Language Development*, 4th edn. Boston, MA: Allyn and Bacon.
- Lane, H., Hoffmeister, R., and Bahan, B. (1996). *A Journey into the Deaf-World*. San Diego, CA: Dawn Sign.
- Marschark, M. (2007). *Raising and Educating a Deaf Child: A Comprehensive Guide to the Choices, Controversies, and Decisions Faced by Parents and Educators*. New York: Oxford University Press.
- Marschark, M., Lang, H. G., and Albertini, J. A. (2002). *Educating Deaf Students: From Research to Practice*. New York: Oxford University Press.
- Mauk, G. and White, K. (1995). Giving children a sound beginning: The promise of universal screening. *Volta Review* 97(1), 1–32.
- Paul, P. V. (1998). *Literacy and Deafness: The Development of Reading, Writing and Literate Thought*. Needham Heights, MA: Allyn and Bacon.
- Rodda, M. and Eleweke, C. J. (2000). Theories of literacy development in limited English proficiency for deaf people. *Deafness and Education International* 2, 101–113.
- Scheetz, N. A. (2001). *Orientations to Deafness*, 2nd edn. Boston, MA: Allyn and Bacon.
- Turnbull, A. P. and Turnbull, H. R. (2006). *Families and Exceptionalities*, 5th edn. Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.
- Turnbull, A. P., Turnbull, A., Shank, M., and Smith, S. J. (2004). *Exceptional Lives: Special Education in Today's Schools*, 4th edn. Upper Saddle River, NJ: Pearson/Merrill/PrenticeHall.
- Yoshinaga-Itano, C., Sedey, A., Coulter, D., and Mehl, A. (1998). Language of early- and later-identified children with hearing loss. *Pediatrics* 100, 135–164.

Further Reading

- Baynton, C. B., Gannon, J. R., and Bergey, J. L. (2007). *Through the Deaf Eyes: A Photographic History of an American Community*. Washington, DC: Gallaudet University Press.
- Brueggemann, B. (2004). *Literacy and Deaf People: Cultural and Contextual Perspectives*. Washington, DC: Gallaudet University Press.
- Easterbrooks, S. R. and Estes, E. L. (2007). *Helping Deaf and Hard of Hearing Students to use Spoken Language: A Guide for Educators and Families*. Thousand Oaks, CA: Corwin Press.
- Eleweke, C. J. and Rodda, M. (2000). Factors contributing to parent's selection of a communication mode to use with their deaf children. *American Annals of the Deaf* 145, 375–383.
- Marschark, M. and Spencer, E. (eds.) (2003). *Oxford Handbook of Deaf Studies, Language, and Education*. New York: Oxford University Press.
- Marschark, M. and Spencer, E. (2006). *Advances in the Spoken Language Development of Deaf and Hard-of-Hearing Children*. New York: Oxford University Press.
- Robinschaw, H. M. and Evans, R. (2001). Service provision for preschool children who are deaf. *Early Childhood Development and Care* 168, 63–69.

Relevant Websites

<http://www.betterhearing.org> – Better Hearing Institute.
<http://www.jtc.org> – John Tracy Clinic.

<http://www.napcse.org> – National Association of Parents with Children in Special Education.

<http://www.infanthearing.org> – NCHAM: National Center for Hearing assessment and Management Utah State University.

Educating Children with Other Health Impairments (ADHD)

G M Mukuria, Slippery Rock University, Slippery Rock, PA, USA

F E Obiakor, University of Wisconsin, Milwaukee, Milwaukee, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Asthma – A chronic pulmonary condition characterized by episodic inflammation and narrowing of the small airways.

Cancer – A large variety of disease in which the cell have the unique trait of unregulated excessive growth and have the ability to invade local or sometimes distant tissues.

Epilepsy – Also known as seizures disorders, a chronic condition in which seizures occur over time unprovoked from other conditions.

Neurological disorders – Impairments that occur in spinal cord, brain, or nerves.

Orientation and mobility – Problems students with physical disability encounter as they move inside and outside the classrooms.

Physical disabilities – A general term covering impairments activities limitations, and participation limitations.

Sickle Cell Anemia – One form of sickle cell disease that consists of anemia (fatigue), painful episodes, and complications.

Educating Students with Other Health Impairments

Some students have disabilities that are invisible and yet relate to health conditions. In Individuals with Disabilities Education Act (IDEA), students with these disabilities are categorized as having other health impairments, the term other differentiates this group from students with intellectual or other disabilities that also may affect health. Ideally, the definition of students with other health impairments includes characteristics related to having limited strength, vitality, or alertness, that result in limited educational opportunities (Zentall, 2006). Students in this category may have chronic or acute health problems such as asthma, attention deficit disorder or attention deficit hyperactivity disorder, diabetes, epilepsy, heart condition, hemophilia, cancer, leukemia, nephritis, rheumatic fever, sickle cell anemia, and HIV/AIDS. These conditions adversely affect children's educational performance and the ability to succeed

in other endeavors. This article focuses on how to help them to maximize their fullest potential.

Historical Perspectives

Students with special needs have been discriminated against, ostracized, labeled, and called demeaning names (e.g., stupid, imbecile, and retarded). In the United States, it is impossible to divorce the education of students with special needs from the Civil Rights Movement and the subsequent events that followed. To a great extent, the education of these students has been greatly influenced by social developments and court decisions in the 1950s and 1960s. For example, the landmark *Brown v. Board of Education of Topeka* (1954) case was a civil rights case that declared separate education as unequal education and unconstitutional (Obiakor, 2006). Consequently, this case opened doors of advocacy for students with special needs including those with other health impairments. The ruling of this case became a catalyst that prompted parents and professionals to lobby for equitable education for all their students.

In 1973, Section 504 of the Vocational Rehabilitation Act (PL 93-112) was passed to provide persons with special needs with (1) free and appropriate public education, (2) civil rights, (3) accessibility of programs, and (4) employability rights. In 1975, the Education of All Handicapped Children's Act (PL-94-142) was passed with the following fundamental ingredients: (1) education for students from 3 to 21 years, (2) free and appropriate public education, (3) identification of students, (4) nondiscriminatory assessments, (5) placement in the least restrictive environment (LRE), (6) confidentiality of information, (7) procedural safeguards, and (8) development of individualized education plans (IEPs). In 1986, PL 94-142 was amended to accommodate young children from birth to 3 years. This law (PL 99-457) was enacted to provide not just IEPs for children but also individual family support programs (IFSP) for parents and guardians. In 1990, PL-94-142 was renamed as the Individuals with Disabilities Education Act (IDEA; PL101-476). This Act involved funding for states to provide educational services to students from birth through 21 years, and ensured procedural safeguards for parents who guarantee meaningful participation in the evaluation process (Katsiyannis *et al.*, 2001).

Additionally, IDEA guaranteed improvement in the education of students with special needs through research training and technical support and transitional supports for student when they are 16 years. To challenge the private sectors, the American with Disabilities Act (1990) (ADA; PL101-336) was passed to provide more societal opportunities for persons with special needs. In 1997, IDEA was reauthorized as PL 105-17 to facilitate disciplinary procedures and reduce litigation costs. In 2001, the No Child Left Behind Act (NCLB; PL 107-110) was passed to educate all learners and quantifiably account for their progress at all levels. Later, in 2004, IDEA was again reauthorized as the Individuals with Disabilities Education Improvement Act (IDEIA; PL 108-446). Imbedded in IDEIA 2004 is the principle of accountability which indicates that schools should educate all students well enough that all children will demonstrate proficiency in certain core academic subjects (English, mathematics, and others). The technique for achieving this principle is to use the standardized state or local assessment of students' academic proficiency. The IDEIA makes provision for students with impairments and other disabilities to participate in these assessments. Those who may not participate due to severity of their condition are given alternative assessment based on what has been delineated in their individualized education programs. According to the provision of this law, teachers of students with special needs must be highly qualified, implying that they must be certified in the content areas that they are teaching (Turnbull *et al.*, 2004; Smith, 2005). In the following paragraphs we discuss seizure, asthma, ADHD, congenital hearts defects, cancer, and sickle cell anemia.

Students with Seizures

Seizure is a sudden, involuntary time-limited disruption of normal function of the central nervous system. Epilepsy refers to a chronic condition in which seizures occur over time and it occurs spontaneously. There are many types of seizures and epilepsy that can be characterized by altered consciousness, motor activity, sensory phenomena, or inappropriate behavior. Prior to seizure occurrence, the victim normally experiences an aura. Seizures are usually classified as partial or generalized seizures. Partial seizures begin in one area of brain, while generalized seizures occur in both hemispheres of the brain. Within each category, there are variations of seizures each with distinct characteristics, prognoses, and treatments. Seizures usually last from a few seconds to a few minutes and end spontaneously. It is critical for medical treatment to be provided once epileptic seizure occurs. Teachers and service providers should play an important role in identifying seizures in the classroom and taking the necessary steps to ensure that the child's safety and dignity

are maintained when they occur in the classroom. The important question is: How are teachers prepared to deal with seizures?

It is important that school personnel be acquainted with different types of seizures. In many instances, a teacher or service provider is the first person to identify a student as having a seizure, especially with more subtle seizures, such as absence seizures. He/she also may be the first person on hand to assist a student having seizure. If a student has generalized tonic-clonic seizure, which is more severe, the staff should be ready and need to know how to handle the situation. Information pertaining to seizure management should be provided to teachers and school staff. In-service training and provision of relevant literature should be available. In addition, teacher preparation programs have the obligation to educate those who can teach all children in spite of their health impairments.

Students with Asthma

The most common chronic illness among children is asthma, a lung disease that causes episodes of great difficulty in breathing (Gabe *et al.*, 2002). Asthma affects between 5 and 7% of all children in the United States. In recent times, the number of cases of asthma has been rising, and the severity of the disease is increasing. Parents and teachers should endeavor to protect children from known triggers to address and minimize the problem (Lemanek and Hood, 1999). Experts in the field advise parents to protect their children from known triggers such as tobacco smoke as one means of addressing this problem. For some children, asthma may be fairly benign while for others it can be chronic. For example, one child may have coughing spells when he/she laughs too hard. For another child, the condition can be much more serious. When exposed to certain triggers, these students' airways swell and they produce mucus that makes breathing difficult (Lemanek and Hood, 1999); and they may require emergency medical intervention. Asthma is hereditary in nature and in most cases if one of the parents has the disease, the child has 50% chances of inheriting it (see Lemanek and Hood, 1999). However, it is often triggered by allergens, including flowers, grass, pollen, dust, molds, animal dander, and food allergies such as eggs, seafood, and many other items, as well as by strenuous physical activity (Greiling *et al.*, 2005). However, asthma can be controlled through medication and the avoidance of environments that are likely to trigger it (Beers *et al.*, 2006; Taras and Potts-Datema, 2005).

There are two major treatment aims of asthma: (1) preventing the occurrence of an asthma attack and (2) treating an asthma attack when it occurs. Both require that child, family, and school personnel have a good understanding of this disease. Apparently, asthmas occur episodically, yet it is a chronic disease. Several medications are available, yet it is

often untreated. While it is perceived as a disease, it continues to be perceived by many as an emotional or hysterical condition. It is critical for teachers and service providers to have a clear understanding of the disease, its symptoms, and the action to take to manage asthma.

Students with Attention Deficit Hyperactivity Disorders

The IDEA does not represent attention-deficit hyperactivity disorder (ADHD) as a distinct category (Yell, 2004). For that reason, children with ADHD who do not qualify for eligibility for special education are served under Section 504 of the Vocational Rehabilitation Act of 1973. The IDEA is very specific while 504 is broad and subject to interpretation. Consequently, many administrators provide services for ADHD students who may not be eligible for special education services under the IDEA. Because of the condition of the students with ADHD, they require medication. Students with ADHD fall under the following classification:

- For some students with ADHD, inattention is the primary problem. These students might skip salient information of an assignment, may appear to be daydreaming during large group instruction, are disorganized, and generally seem forgetful both in school and at home. This variability of disorder is referred to as ADHD-predominantly inattentive type.
- For another group of students with ADHD, the salient symptom is a combination of hyperactivity—high amount of movement, and impulsivity—the ability to think before acting. Fidgeting is a common phenomenon for students with this type of ADHD. By looking at them one would think that they have an internal engine that does not stop and talks a lot. This type of ADHD is called ADHD-predominantly hyperactive-impulsive type.
- Some students with ADHD have symptoms suggesting that both inattentive and hyperactivity-impulsivity are part of this disorder. These students are referred to as having ADHD combined type.

As it stands, ADHD is considered neurological, chronic, long term, and acutely acquired. The primary trait of this condition is an inability to attend beyond what is typical for peers of comparable age. Significant impulsivity also may be a characteristic. The ADHD is not situational in that it affects children and adults who have it across all settings. However, the symptoms may be most conspicuous at school because of the structure and expectations there. Surprisingly, students with ADHD are more likely to have a productive deficit rather than an acquisition deficit. That is, they take information in and sometimes surprise their teachers by what they know. The greatest difficulty often lies in the production, that is, in completing their work. Apparently ADHD is not caused by

environmental situations or other disabilities, but may be present with them.

Since the goal of any educational program is to maximize the fullest potential of students, the use of medication, though controversial, has become the norm for students with ADHD. Some educators perceive the use of medication as a quick fix and argue that it is being overused. Other educators argue that medication is effective and calms down students with ADHD and thereby increases their productivity (Fornes *et al.*, 1999; Lloyd *et al.*, 2006; Molina and Pelham, 2001). According to the United States National Institute (2006), approximately 2.9% of all students aged 18 and younger in the United States take medication. Clearly, taking medication usually reduces significantly the amount of daydreaming or disruptive behavior. Stimulant medication such as Ritalin, Dexedrine, Cylert, Adderall, and Focalin are among the most common medication administered to students with ADHD. There are antidepressant medication such as Norpramin, Tofamil, Prozac, Zoloft, and many others that are administered for various reasons. As a result, it is critical for school personnel to observe their students daily and form opinions about the seriousness of their behaviors and what might be beneficial for them. However, school personnel do not have the expertise to recommend to parents what might help them. Teachers' preparation programs should endeavor to equip pre service teachers with techniques such as functional behavior assessment and analysis so that they can know how to positively deal with the disruptive behavior of ADHD students.

Students with Congenital Heart Defects

Congenital heart defects refer to a variety of conditions in which the heart is structurally impaired at birth. This might be as a result of hole in one of the chambers of the heart, a problematic valve, a misplacement of the vessel of the heart, or other abnormalities. Structural abnormalities of the heart occur during fetal development because of genetic factors, adverse environmental or maternal conditions, or other factors that are often unknown. Some defects are minor and often resolve on their own, while others require surgery. In normal heart, blood with oxygen enters the right side of the heart and is pumped through the pulmonary arteries to the lungs. From there, the new oxygenated blood goes into the left side of the heart, which pumps the oxygenated blood out of the body through the aorta.

Depending on where the heart defect is located, the defect falls into one of the two categories: acyanotic heart defects or cyanotic heart defects in which case oxygenated blood from the left side of the heart is incorrectly pumped to the unoxygenated right side of the heart due to an obstruction or uncorrected flow to the lungs. This causes a strain on the heart from pressure or volume overload

which can often lead to congestive heart failure. In cyanotic heart defects, there is a blue coloration of mucus membrane and skin because the defect causes some unoxxygenated blood from the right side of the heart to be pumped to the body (Cook and Higgins, 2004).

Inevitably, students who have not had their congenital heart defect repaired or who are not in the process of undergoing multiple surgeries to repair the defect will require close monitoring for signs of distress. Because of the lack of vitality, children will often limit activity on their own but can become easily fatigued and breathless if running, climbing stairs, playing, or walking for extended periods of time (Cook and Higgins, 2004). Adaptation such as using elevators instead of stairs may be a viable choice. Some students with cyanotic heart defects will periodically squat after motor activity to obtain relief and should be allowed to do so. Others may have cyanotic spells in which they become anxious and short of breath, hyperventilate, and have an increase in cyanosis due to their body having a sudden drop in oxygen. Depending on their condition, these students may need to rest, be given oxygen, or receive emergency care. Teachers and service providers should be alert and continuously monitor them. Clearly, children who have had repair of complex and cyanotic heart defects have shown deficits in exercise capacity with impaired exercise performance (Norozi *et al.*, 2005). Teachers and other professionals will need to verify with parents if there has been complete correction of the defect and if monitoring is required.

Students with Cancer

Cancer refers to a large variety of diseases in which the cells have the unique traits of unregulated, excessive growth, and the ability to invade local or, sometimes distant tissue. This can lead to tissue or organ damage that interferes with proper functioning of the tissue or organ being invaded. There are numerous types of cancers that vary depending on their location in the body and how they spread. Children often get cancer in the brain (brain tumor) and in the bone marrow (leukemia). Some other cancers are those affecting the lymph nodes, kidney, and bone. The type of cancer will determine how it affects the body, the form of treatment, and the prognosis. Although having a student with cancer raises fears about death and dying, it should be remembered that some cancers have a high cure rate and that others may go into long-term remission.

Students who have cancer that is treated by chemotherapy or radiation may experience side effects that the teacher or service provider will need to monitor and provide appropriate adaptation and support. For example, pain, stomach upset, sleep disturbance, and fatigue have been found in children with leukemia following chemotherapy (Gedaly-Duff *et al.*, 2006). Teachers and other professionals should carefully monitor and keep track of these symptoms,

provide breaks, or repeat instruction at a later time when the student is more alert and comfortable. Fatigue and other symptoms are often present in children who have end-stage cancer (Mooney-Doyle, 2006), which usually requires some type of adaptations such as rest breaks, shortened day, or homebound services. Further, children undergoing cancer treatment are usually more vulnerable to infections. Teachers and service providers need to be vigilant to sickness in other students and send them home so that the infection is not transmitted to the student who has cancer.

Students with Acquired Immune Deficiency Syndrome

Discussions on how to educate all children will be incomplete without mentioning the impact of acquired immune deficiency syndrome (AIDS) on children and their learning. AIDS is often thought to be a disease that merely makes one susceptible to fatal infections. However children with AIDS often acquire neurological problems as well, including intellectual disability, cerebral palsy, seizures, and emotional and behavioral disorders. AIDS is caused by human immunodeficiency virus (HIV). Although HIV infections can be spread in other ways, the primary means of transmission in adolescents and adults is sexual contact. Infants may contract the disease during the process of birth. Apparently, HIV does not often result in neurological damage and cognitive impairment. In some cases, the damage may be reversible with state-of-the-art drug therapies (Willen, 2006).

Sadly, most of the reported new cases of HIV are in the minority population and almost three-fourths of the new infections in the group aged 13–19 years are in the minority population (Rogers, 2001: 1). HIV/AIDS still carries considerable stigma in the society (Murphy *et al.*, 2002). As children and youth with AIDS and other virus and bacterial infections live longer as a result of improved medical treatment, there will be increasing need for special education and related services. Teachers and service providers should be cognizant of the fact that if reasonable procedures are followed for preventing infections, there is no serious concern regarding transmission of HIV in classrooms (Ainsa, 2000; Spiegel and Bonwin, 2002).

Students with Sickle Cell Anemia

Sickle cell anemia is one form of sickle cell disease that consists of anemia, painful episodes, and complications which lead to eye disease, spleen dysfunction, or stroke. This is a hereditary disease and mainly affects people of African descent. An abnormal type of hemoglobin (HBS) results in some red cells being shaped like a sickle instead of normal disk-shaped red cells with pinched-in sides. These poorly formed sickled red cells have a poor shorter life span, usually being broken down in approximately 20 days

instead of 120 days (Jakubik and Thompson, 2000) which results in chronic anemia and subsequent fatigue.

Apparently, teachers should try to prevent the triggers by events of vaso-occlusive crisis and monitor its occurrence. A vaso-occlusive crisis can be triggered by events that cause a decrease in oxygen, such as strenuous exercise, cold weather, dehydration, infection, or high altitude. When a vaso-occlusive crisis begins, some red blood cells change into a sickle shape. These deformed cells can plug the small blood vessels in the body, resulting in localized tissue hypoxia, which leads to further sickling of the blood cells. When blockage occurs in the blood vessel, it results in tissue death (Dorman, 2005). Intense pain ensues after a vaso-occlusive crisis. Coughing, shortness of breath, and severe abdominal pain follow. Teachers and service providers should be vigilant and monitor the situation carefully. Treatment for severe crisis may consist of rest, rehydration, pain medication, management of complication, and sometimes blood transfusion (Dorman, 2005).

Implications for Teacher Preparation

Clearly, educating students with other health impairments causes unique challenges. The big issues for teacher education programs is whether pre-service teachers are being adequately prepared to educate all learners. Special education in general is a multifaceted discipline; and teachers teach what and how they have been taught. Poorly prepared teachers cannot teach well. In many teacher programs, candidates are not exposed to courses that prepare them to meet the challenges presented by students with other health impairments. Collaboration with medical and health workers can be helpful to bridge the gap. General and special education educators should go beyond territorial protection and look for knowledge that would benefit their students. This is one way to move forward in this world of change.

Conclusion

While great strides have been made to provide education to all students including those with disabilities, there are many challenges that need to be met. Clearly many students with special needs are being educated, however, there are some gray areas when it comes to educating students with other health impairments. For example, students with HIV/AIDS (majority of whom are homeless), those with sickle cell anemia, and others in that category require more attention than they are currently receiving. Teacher education programs should adequately prepare teachers who can teach all students, including those in urban areas and other poor and disadvantaged communities. Education is a dynamic field and these programs must change with time.

See also: Assistive Technology and Educational Practice; Diagnosis and Classification; Educating Children with Physical Disabilities; Educating Students with Special Needs: An Overview; Gender Equity in Higher Education: Challenges and Celebrations; Inclusion of Students with Special Needs in General Education Classrooms; Related Services for Children with Special Needs; Transition from School to Adult Life.

Bibliography

- Ainsa, P. (2000). *Teaching Children with AIDS*. Wales, UK: Edwin Mellen Press.
- American with Disabilities Act (1990). Pub. L. No 101-336.
- Beers, M. H., Porter, R. S., Jones, T. V., Kaplan, J. L., and Berkwitz, M. (2006). *The Merck Manual of Diagnosis and Therapy*, 18th edn. Whitehouse Station, NJ: Merck Research Laboratories.
- Brown v. Board of Education of Topeka Kansas (1954). 347 U.S. 485, 745-ct 686, 98 L. ED. 873, 530.0. 326.
- Cook, E. H. and Higgins, S. S. (2004). Congenital heart disease. In Allen, P. J. and Vessey, J. A. (eds.) *Primary Care of Child with Chronic Condition*, 4th edn, pp 382-403. St. Louis, MO: Mosby.
- Dorman, K. (2005). Sickle cell crisis. *RN* **68**, 33-36.
- Fornes, S. R., Kavale, K. A., and Crenshaw, R. M. (1999). Stimulus medication revisited. Effective treatment of children with ADHD. *Reclaiming Children and Youth* **7**, 230-235.
- Gabe, J., Bury, M., and Ramsey, R. (2002). Living with asthma: Experiences of young people at home and in school. *Social Science and Medicine* **55**, 1619-1633.
- Gedaly-Duff, V., Lee, K. A., Nail, L. M., Nicholson, S., and Johnson, K. P. (2006). Pain, sleep disturbance, fatigue in children with leukemia and their parents: A pilot study. *Oncology Nursing Forum* **33**, 641-646.
- Greiling, A. K., Boss, L. P., and Wheeler, I. S. (2005). A preliminary investigation of asthma mortality in schools. *Journal of School Health* **75**, 286-290.
- Jakubik, L. D. and Thompson, M. (2000). Care of the child with sickle cell disease: Acute Complications. *Pediatric Nursing* **26**, 373-379.
- Katsiyannis, A., Yell, M. L., and Brandley, R. (2001). Reflection of the 25th anniversary of the individuals with exceptional education act. *Remedial and Special Education* **22**(6), 324-339.
- Lemanek, K. L. and Hood, C. (1999). Chronic asthma. In Brown, R. T. (ed.) *Cognitive Aspects of Illness in Children*, pp 78-104. New York: Guildford Press.
- Lloyd, G., Cohen, D., and Stead, J. (eds.) (2006). *Critical New Perspective on ADHD*. Oxford, UK: Routledge.
- Molina, B. S. G. and Pelham, W. E. (2001). Substance use, substance abuse, and LD among adolescence with a child history of ADHD. *Journal of Learning disabilities* **34**, 333-342.
- Mooney-Doyle, K. (2006). An examination of fatigue in advanced childhood cancer. *Journal of Pediatric Oncology Nursing* **23**, 305-310.
- Murphy, D. A., Roberts, K., and Hoffman, D. (2002). Stigma and ostracism associated with HIV/AIDS children carrying the secret of their mothers HIV+ serostatus. *Journal of Child and Family Studies* **11**, 191-202.
- Norazi, K., Gravenhorst, V., Hobbiebrunken, E., and Wessel, A. (2005). Normality of cardiopulmonary capacity in children operated on to correct congenital heart defects. *Archives of Pediatrics and Adolescents Medicine* **159**, 1063-1068.
- Obiakor, F. E. (2006). Educating African American urban learners: Brown in context. In Brown, M. C. II and Bartee, R. D. (eds.) *The Broken Cisterns of African American Education*, pp 61-72. Charlotte, NC: Information Age.
- Rogers, A. S. (2001). HIV research in American youth. *Journal of Adolescence Health* **29**, 1-4.
- Smith, T. E. (2005). IDEA 2004: Another round of the reauthorization process. *Remedial and Special Education* **26**(6), 314-323.
- Spiegel, H. M. L. and Bonwin, A. M. (2002). HIV infection in children. In Batshaw, M. L. (ed.) *Children with Disabilities*, 5th edn, pp 123-139. Baltimore, MD: Paul, H. Brooks.

- Taras, H. and Potts-Datema, W. (2005). Chronic health conditions and student performance at school. *Journal of School Health* **75**.
- Turnbull, R., Huerta, N., and Stove, M. (2004). *The Individual with Disabilities Education as Amended in 2004*. Upper Saddle River, NJ: Prentice Hall.
- Willen, E. (2006). Neurocognitive outcomes in pediatric HIV. *Mental Retardation and Developmental Disability Research Reviews* **12**, 223-228.
- Yell, M. L. (2004). *The Law and Special Education*, 2nd edn. Upper Saddle River, NJ: Prentice Hall.
- Zentall, S. S. (2006). *ADHD and Education*. Upper Saddle River, NJ: Prentice Hall.

Further Reading

- Mills v. Board of Education of the District of Columbia (1992). 348 f. Supp. 866 (D.D.C.1972).
- Pennsylvania Association for Retarded Children v. Commonwealth of Pennsylvania (1972). 343 F. Supp. 279 (D.C. PA 1972).

Educating Children with Physical Disabilities

Z P Nkabinde, New Jersey City University, Jersey City, NJ, USA

F E Obiakor, University of Wisconsin–Milwaukee, Milwaukee, WI, USA

M T Ofor, University of Wisconsin–Milwaukee, Milwaukee, WI, USA

D J Smith, Jackson State University, Jackson, MS, USA

© 2010 Elsevier Ltd. All rights reserved.

Physical disabilities can affect a person's ability to move about, to use arms and legs effectively, to swallow food, and to breathe unaided (Black and Pretes, 2007; Hardman *et al.*, 1999). Such limitations may also be evident in other areas such as vision, cognition, speech, language, hearing, and bowel movement. In recent years, due to advancement in medicine and technology, more and more students with mild to severe physical disabilities are now served in regular classrooms. Hill (1999) stated that students with chronic physical and medical conditions are being integrated into general public schools alongside their age- and grade-appropriate peers in far greater numbers than in the past. Some have attributed this change, particularly for students with orthopedic impairments, to the removal of physical barriers in the school environment and to changes in attitudes and beliefs (see Hill, 1999). Since students with physical disabilities face many challenges during their school years due to their precarious physical and health situations, their education requires careful planning involving administrators, teachers, parents, and students (Smith, 2007).

It is important for students with physical disabilities to feel accepted in a school environment. When they feel valued and respected by their peers, their self-image improves for the better. Stein (1989) as cited by Hill (1999) made the following statement:

It has been suggested that children are our most precious resources and that a society can be judged in part or in whole by how it treats its most vulnerable members. Children with chronic physical conditions represent a vulnerable population, and the time has come for us to go beyond the recognition of that fact to remedy the problems and address the needs. At present we lag behind many other nations in our willingness to assure a wide range of services to children and their families. The challenge for all of us is to continue to make our own individual and collective contributions to caring for children with chronic illness and to assume a leadership role in setting a high standard for the provision of comprehensive and humane services to these children and their families. (pp. 25–26)

Related Services for Students with Physical Disabilities

Hill (1999) contended that P.L. 94-142 and its amendments emphasized the importance of the provision of related services to assist in the removal of any potential barriers to the successful inclusion of a child with special healthcare needs into the general education classroom. According to Palfrey as cited by Hill (1999), the following related services were included in P.L. 101-476:

Transportation and such developmental, corrective and other supportive services as are required to assist a handicapped child [a child with disability] to benefit from special education, and includes speech pathology and audiology, psychological services, physical and occupational therapy, recreation, early identification and assessment of disabilities in children, counseling services, and medical services for diagnostic and evaluation purposes. The term also includes school health services, social work services in schools, and parent counseling and training. (p. 21)

In addition, Hill stated that the same educational legislation defined medical services as “services provided by a licensed physician to determine a child's medically related disability that results in the child's need for special education and related services” and school health services as “services provided by a school nurse or other qualified person.”

Students with physical disabilities may encounter serious orientation and mobility problems inside and outside of a classroom. These problems can have a negative impact in terms of classroom activities and overall performance in a school setting. Thus, students who are attending regular school often need support from varied professionals such as physical therapist, occupational therapist, therapeutic recreation specialist, and an adapted physical educator in order to participate in different school activities (Eriksson *et al.*, 2007). The role played by each professional in the student's life differs according to the need and severity of the disability. Shapiro and Sayers (2003) contended that physical therapists are health professionals with specific

training in the remediation of a disability, dysfunction, or pain. As they noted, physical therapists may do the following:

- Restore and preserve range of motion.
- Evaluate muscle length and perform stretching exercises.
- Perform muscle-strength evaluation.
- Offer exercises to increase strength, endurance, and coordination for specific muscle groups or the entire body.
- Evaluate muscle tone.
- Evaluate and train sitting and standing balance, transfers, and mobility, including wheelchair use for ambulation.
- Evaluate and train users of lower extremity orthosis and prostheses.
- Assess skin integrity and sensation.
- Offer various physical therapy modalities (e.g., heat, ultrasound, and massage) to relieve pain, prevent deformity, develop or improve muscle strength or motor skills, and maintain maximal functional capabilities. (p. 33)

In a school setting, these professionals offer techniques that correct, facilitate, or adapt the student's functional performance in motor development (e.g., a student's gross motor skills and related locomotor activities), motor control and coordination, posture and balance, functional mobility (e.g., transfer skills from a desk to a walker, from a wheelchair to a bench, gait training, use of walkers, wheelchairs, crutches, prosthetics, and/or braces (Shapiro and Sayers, 2003).

Another service provider for students with physical disabilities is an occupational therapist. The role of this professional is to provide services to people with specific performance difficulties in self-care, work, and play activities that limit their ability to deal with tasks of everyday (see Shapiro and Sayers, 2003). Services provided by occupational therapists include the following (see Shapiro and Sayers):

- Evaluation and training in self-care activities, such as dressing, eating, bathing, and personal hygiene.
- Training in home management skills.
- Exploration of vocational skills and interests.
- Maintaining and improving joint range of motion, muscle strength, endurance, coordination, and dexterity in upper extremities.
- Evaluation and training for compensation for sensory, perceptual, and cognitive deficits.
- Training in use of upper extremity prosthesis.
- Evaluation and training in use of assistive technology systems. (p. 35)

Occupational therapists render services aimed at improving student performance in postural stability, sensory processing, organization and integration, motor planning and coordination, fine-motor abilities (e.g., arm and hand), self-help abilities (e.g., eating and dressing),

environmental adaptations, use of assistive devices, and social and play abilities (see Shapiro and Sayers). In the school setting, occupational therapists are likely to consider the role of the hand by facilitating or teaching a student to use his/her hand to improve handwriting skills and other hand-manipulation skills related to areas of the curriculum, such as art, music and band, and transfer skills. Additionally, they may contribute to independence in school by helping a student who uses a wheelchair learn to carry a lunch tray on his/her lap, manage a ramp to and from the playground, and move through the halls.

Another service provider for students with physical disabilities is a speech and language pathologist. Bowe (2000) noted that this professional has the role of:

- Screening for articulation, breathing, feeding, and related problems.
- Assessing ability to express oneself in speech and in other ways (gestures, sign language, and communication boards).
- Assessing ability to use age-appropriate language (e.g., muteness, echolalia, and language delays).
- Identifying goals and interventions.
- Monitoring progress toward goals.
- Identifying, ordering, adjusting, and training the individual to use adaptive devices such as communication boards.
- Serving as a team member.

Art therapists play a major role in the lives of students with physical, sensory, and health disabilities because of the nature of physical disabilities. Bowe (2000) described art therapy as a service that offers benefits similar to counseling. It is particularly helpful to individuals who are coping with strong emotions, like those who have spinal cord injury (SCI) or traumatic brain injury (TBI) and many who have muscular dystrophy or other fatal conditions. As Bowe pointed out, art therapists:

- Assess physical and cognitive abilities, including perception of others and self-image, and problem-solving skills.
- Set goals for treatment.
- Identify appropriate individual and group therapy interventions, including not only artwork but also discussion of one's work.
- Serve as a team member. (p. 51)

There is no doubt that collaboration in meeting the needs of students with physical disabilities is a necessity. Such students have complex needs ranging from environmental to health needs.

Intervening Early in Physical Disabilities

Young children from birth to 3 years may receive early intervention services if they qualify under certain

conditions (Johnson *et al.*, 2007). These conditions include a developmental delay in one of five developmental domains: cognitive development, physical development, social or emotional development, communication development, and adaptive development. Another criterion that can lead to eligibility is established risk conditions such as young children who may have a diagnosed physical or medical condition that may lead to a developmental delay. The list may include premature babies, those children with low birth weight due to chronic illnesses, or exposure to drugs and alcohol. Early intervention sets the stage for better language, physical, social, and cognitive development. It may reduce the impact of physical disability on learning if an intensive and comprehensive therapy is put in place. It may reduce stress in family when a student learns to be less dependent on his/her caregivers. In addition, early intervention is crucial for people who are born with disabilities or acquire them during the early childhood years. According to Bowe (2000), the IDEA's Part C early-intervention services can be very helpful. Specifically, students with physical disabilities who are struggling with speech, communication, social, mobility, and other academic issues presented by their condition can benefit from early intervention. For example, students with spina bifida may receive physical and occupational therapy services. For young children with spina bifida who also have learning disabilities, special instruction can help them to develop effective learning strategies. Peer acceptance is critical in decreasing the impact of a physical disability.

Smith (2007) noted that whether disabilities are physical or health related, early-intervention programs provide a strong foundation for the child and family. Parents and other family members may benefit from counseling and information services that help them both to cope with the child's disability(ies) and to take advantage of public services that will help the family and the child. Secondary disabilities can be prevented through early screening and early intervention. Surgical procedures, which are common for spinal cord injuries, often reduce the severity of the injury and contribute to positive outcome (Hardman *et al.*, 1999). Motor development and positioning and the development of communication skills are often target for young children with physical or health disabilities (Smith, 2007). Most early-intervention services are provided in the home; and some are offered in early-intervention centers. Homebound children such as those who are medically fragile and technology dependent are most likely to receive their treatment in hospitals and/or special facilities such as respite care.

Clearly, some disorders are confined to childhood such as dyspraxia, a disorder characterized by an impairment in the ability to plan and carry out sensory and motor tasks (Stansell, 2007). Dyspraxia can be a problem largely confined to childhood; but as long as children receive therapy,

they can be cured of their disorder. In early childhood education, assistive technology is extremely important. It could lead to increased active participation in family, school, and community, successful play with peers and toys, increased ability to communicate needs and ideas, opportunities to make independent choices, and increased ability to move independently. With so many schools and districts now placing students with severe disabilities, including those with physical disabilities in inclusive settings, interventions that facilitate full participation in academic and social activities are becoming ever more important (Wolfe and Hall, 2003). The ultimate goal of early intervention is to help students with physical disabilities become as independent as possible.

Increasing Learning Potential of Students with Physical Disabilities

Turnbull *et al.* (2004) contended that 93% of students with physical disabilities attend regular schools. Approximately half of these students are educated outside the general education classroom for less than 21% of the time. One of the many challenges facing general education teachers, administrators, and other service providers is ensuring that these students access general education curriculum without any environmental barriers. Maximum educational participation can only be possible with appropriate curricular adaptations, accommodations, and modifications dictated by learning and physical needs. It must be noted that some students do not have conditions that limit their learning. However, in other cases, students may have conditions such as cognitive disabilities that limit their learning. Adaptations may range from stabilizing material on a slant board to using an augmentative communication device (Heller *et al.*, 2009). Earlier, Smith (2007) cited classroom tips for teaching students with physical disabilities and urged educators and service providers to:

- Adapt the physical environment such as removing hazards; creating more work space; providing storage space for equipment; making furniture accessible; widening aisles; using positioning devices; changing seating arrangements; and rearranging furniture.
- Change student response mode, such as allowing speaking instead of writing; using a speech synthesizer; allowing writing instead of speaking; and allowing computer-print output.
- Alter materials and equipment, for instance, giving hand-outs; adapting writing tools; using special eating utensils; and exploring assistive and adaptive technology.
- Modify the activity, for instance, allowing more time to complete the assignments; abbreviating assignments; and creating a flexible schedule.

- Provide extra assistance, such as arranging peer tutors; having parents or family members help; finding volunteers to assist; video or audio taping lessons; and using e-mail for help sessions.

Many students with physical disabilities may not be able to participate fully or perform to their full potential due to their severe physical conditions. The types of disability (e.g., orthopedic impairments, multiple disabilities, traumatic brain injury, or other health impairments) in many instances will determine the impact on school performance. According to Gargiulo (2006), students with orthopedic impairments often have problems accessing materials; students with other health impairments are more likely to have problems of endurance and stamina. The severity of a disability will also play a role in determining its impact on school performance. For instance, motor limitations have an impact on how students with physical disabilities participate in a school environment. Gargiulo reported that a student whose disabilities limit the movement of arms, hands, and fingers may be unable to do school tasks involving writing or manipulating material without modifications. Optimal school performance may not be reached even with modifications. As an example, a wheelchair-bound student may not be able to participate in all academic projects such as field trips. What takes students without disabilities a shorter period of time might be time consuming for students with disabilities – these students may be forced to use compensatory strategies such as orally presenting their work instead of writing them or using the Internet to visit museums instead of personally visiting them. Another strategy is that of using peers as note takers and/or parents or volunteers. Such strategies require prior planning and working around the routine and schedule that best suit the student's needs. Assignments may also need to be adapted to avoid fatigue and inattentiveness that may be common for students with physical disabilities. Wheelchair mounting allows wheelchair users to have a desktop computer attached to their wheelchair for easy use. In this manner, students are afforded the opportunity to have their work areas.

Students with physical disabilities can have their books adapted to fit their physical needs. For example, e-books can be used for students with motor limitations. Adapted books can increase the school performance of these students and make reading more pleasurable and meaningful. Page turners and use of color coding, sentence strips, and sticky notes are all low-tech strategies that can compensate for students' limitations. Restricted communication can be a factor on school performance. Effective communication is a critical component for successful education of students with physical disabilities, primarily because some of these students face communication challenges (Howell and Gengel, 2005).

Assistive technology such as voice-recognition technology offers great hope for this population. However, for students who cannot spell or write, no augmentative communication device will help. Students struggling with writing tasks can benefit from using a speech-recognition system, a device that provides an alternative strategy for students who experience writing difficulties using standard methods for computer access to write efficiently (see Gargiulo, 2006).

Students with physical disabilities experience fatigue and endurance limitations that can negatively influence their school performance. Keyser-Marcus *et al.* (2002) stated that because many students with traumatic brain injury experience difficulty with fatigue, a modified school day (e.g., regularly scheduled breaks and shortened length of school day) may be necessary for optimal performance. Workload for this population can be reduced without compromising quality. Students can be given extra time to complete their assignments. Task analysis where complex tasks are broken into teachable units is also recommended. Additionally, for students with health factors such as juvenile rheumatoid arthritis or for students with poorly fitted artificial legs, there might be discomfort which in turn can interfere with school performance (Gargiulo, 2006). Students with physical disabilities and other health impairments tend to miss classes a lot and that may have a negative effect on their school performance. According to Gargiulo, students with severe speech and physical impairments (as in severe cerebral palsy) tend to be further behind in reading level than is commensurate with intelligence quotient (IQ). Clearly, self-concept and self-esteem issues can affect school performance of students with physical disabilities. How these students perceive themselves and how they are perceived by others play a role in their overall social well-being. Unlike other disabilities, being physically impaired means being aware of your difference earlier on since the disability cannot be concealed. Gargiulo acknowledged that students with physical and health impairments may feel like outsiders in everyday life; they may feel different, deprived, or unable to live like other children. While other children may view life as being filled with possibilities, these students may see life as filled with limitations. In some instances, these students may feel worthless and burdensome to those who care for them. Such feelings of helplessness may contribute to lowered self-worth.

When students have physical disabilities, assistive technology can provide them with opportunities to control their environments in areas such as communication and independent daily living skills both at school and at home. Bowe (2000) stated that technology has revolutionized the lives of Americans with physical, health, and sensory limitations in school, at home, in the community, and in the workplace. What was almost impossible to

accomplish by these individuals has become possible. For example, with today's technologies a student with physical disability can and often does take courses in inclusive settings in elementary and secondary schools, performing the same assignments and taking the same tests as do children and youth without disabilities.

Bowe noted two basic kinds of technologies, namely: (1) low-tech devices which are products that have one or no moving parts, rarely have any electronic components, and usually are low in cost; and (2) high-tech products which tend to have numerous moving parts, feature complex electronics, and cost several hundreds to several thousands of dollars. Below are some examples of low-tech devices (Assistive Technology and Children with Special Needs, 2007):

- Picture exchange communication systems are useful to create communication boards for students with limited communication skills.
- Dry erase boards are good writing tools.
- Clip boards can be used to create communication boards for students with limited speech. Pictures and symbols may be used, printed, or cut out and pasted on a communication board for easy access.
- Laminated photographs can be used as labels for objects in the home and classroom.
- Manipulatives/objects can be used as teaching tools for counting and to stimulate communication across settings. (p. 6).

Varieties of low-tech products are now available for students with physical disabilities including a large number of switches. The Alliance for Technology Access (1994) agreed that the placement of the on/off switch might seem a trivial matter, but if one has limited mobility, one may want the switch to be accessible. Mouth-sticks, head-sticks, foot-operated rocker switches, and many other switches make it possible for an individual who can make at least one voluntary motion with their bodies to activate a switch to gain some control over their environment (Bowe, 2000). Switch-activated toys are also available. According to Best *et al.* (2005), to use a switch effectively, individuals should (1) be properly positioned if nonambulatory, (2) demonstrate nonreflexive stimulation, and (3) exhibit voluntary movement on cue. Other low-tech products include can grabbers, adapted spoon handles, Velcro fasteners (Turnbull *et al.*, 2004). Assistive-technology devices can also be beneficial to students with mobility limitations in their hands and arms, making it possible for them to write. Adapted writing tools include (1) grips, foam material, or balls added to pencils or pens to enhance their grip and (2) attachments, templates, splints, or weights that enable students to hold writing tools at the proper angle. Dell *et al.* (2008) stated that between sophisticated high-tech and nonelectronic low-tech devices are items classified as

mid-tech devices. Mid-tech devices are described as electronic in nature but are much less expensive and require less training than high-tech devices. Examples of mid-tech devices include, but are not limited to, tape recorders, digital recorders, hand-held electronic dictionaries, and spell-checkers. Oversized calculators and calculators that talk are additional examples of mid-tech devices (Dell *et al.*, 2008).

Examples of high-tech devices (Assistive Technology and Children with Special Needs, 2007) include:

- Video taping: Class lectures can be taped to allow students to work at their own pace and to allow more time in the completion of classroom tasks. Since many students with physical disabilities are absent from school for days or weeks at a time having lessons on tape can provide them with instructions in the comfort of their homes or in hospitals.
- Computers including adaptive hardware, touch window where the computer monitor has a built-in sensors that provide activating computer programs with a finger-tip, rather than using a regular mouse; intellikeys provides a keyboard with large print keys; big keys, trackballs, and software.
- Accessory equipment such as digital cameras and scanners. (pp. 7, 8).

Video taping can be used for homebound students to allow them to keep up with classroom expectations. Classroom work can also be posted online and such a tool can allow students with physical disabilities to participate in classroom discussions without being physically present. Phone hook-ups can be used as another avenue to reach homebound students. Both low-tech and high-tech devices increase independence of students and adults with physical disabilities. Different types of assistive technology may be required for students with physical disabilities in order to benefit from their education programs (Heller *et al.*, 2009). As Heller *et al.* pointed out, these include devices for augmentative and alternative communication, mobility, life management, computer access, academic content areas, and play and leisure. Dell *et al.* (2008) stated that one of the most powerful applications of computer technology has been the development and ongoing refinement of devices that can speak. An example of a software that facilitates communication is communication boardmaker, a versatile software program provides speech for people unable to communicate orally (The Alliance for Technology Access, 1994). Communication boardmaker features a searchable database of more than 3000 universal picture communication symbols. Another communication device is a touch talker, a portable device with a computer-generated voice (see Dell *et al.*, 2008). Speech-recognition programs allow the computer to translate spoken word to text. Such programs allow for hands-free control of the computer and are useful tools

for students with motor limitations. Some students with physical disabilities have normal IQs but may have problems expressing their thoughts using traditional methods of a pen and paper. Another important application for individuals with motor impairments is word prediction (Edyburn, 2003). Word-prediction word processors, such as Co:Writer, reduce the stress of writing for those individuals with deficit motor skills. Writing that requires extensive use of fine motor skills which may result in excessive physical effort and fatigue is reduced. Evidently, assistive technology is not only useful in reducing environmental barriers for students with physical disabilities but also helps them on many different levels in the content areas. The Power of Assistive Technology (2007) identified the following tasks that can be accomplished:

- *Master grade-level content.* Technology presents the material in different forms (visually, auditorily, etc.)
- *Improve writing and organizational skills.* Technology can enable students with learning disabilities to do such things as develop a concept map for a research paper and write using grade-level vocabulary words they would not use without a computer due to poor spelling skills.
- *Read grade-level text.* The computer either reads the text digitally or represents it at a lower grade level for students with reading disabilities or visual impairments.
- *Take notes.* Many students with disabilities have difficulty taking notes in longhand because of poor spelling, writing, and/or hand-coordination skills.
- *Master educational concepts that would have been beyond their reach.* Students can experience abstract concepts such as the metamorphosis of a flower through 3-D simulations.

Conclusion

This article has focused on how to maximize the potential of students with physical disabilities. Numerous studies have indicated that these students have lower participation both in structured and unstructured school activities. Removing all barriers for these students will expose them to unique opportunities that will help them reach their full potential. In a supportive environment, such students will have a more positive school experience, more social contacts, less dependence on others, and will enjoy and benefit from their autonomy. Though the inclusion of these students both with or without severe disabilities has become a norm, Wolfe and Hall (2003) concluded that special education laws strongly encourage appropriate placements in least-restrictive settings. Our goal as professionals must be to create environmental and instructional supports that will help students with physical disabilities to overcome various challenges presented by their physical limitations.

See also: Adapted Physical Activity for Students with Special Needs; Assistive Technology and Educational Practice; Educating Children with Other Health Impairments (ADHD); Instructional Accommodations for Children with Special Needs In Inclusive Settings; Related Services for Children with Special Needs; Risk and Resilience Frameworks in Understanding Special Education; Self-Determination; Workforce Issues in Special Education.

Bibliography

- Alliance for Technology Access (1994). *Computer Resources for People with Disabilities: A Guide to Exploring Today's Assistive Technology*. Alameda, CA: Hunter House.
- Black, R. S. and Pretes, L. (2007). Victims and victors: Representation of physical disability on the silver screen. *Research and Practice for Persons with Severe Disabilities* 32(1), 66–83.
- Bowe, F. (2000). *Physical, Sensory, and Health Disabilities: An Introduction*. Upper Saddle, NJ: Merrill/Prentice Hall.
- Dell, A. G., Newton, D. A., and Petroff, J. G. (2008). *Assistive Technology in the Classroom: Enhancing the School Experiences of Students with Disabilities*. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Edyburn, D. L. (2003). *What Every Teacher Should Know about Assistive Technology*. Boston, MA: Allyn and Bacon.
- Eriksson, L., Welander, J., and Granlund, M. (2007). Participation in everyday school activities for children with and without disabilities. *Journal of Developmental and Physical Disabilities* 19, 485–502.
- Gargiulo, R. M. (2006). *Special Education in Contemporary Society*, 6th edn. Belmont, CA: Thomson Wadsworth.
- Hardman, M. L., Drew, C. J., and Egan, M. W. (1999). *Human Exceptionality: Society, School, and Family*, 6th edn. Boston, MA: Allyn and Bacon.
- Heller, K. W., Fomey, P. E., Alberto, P. A., Best, S. J., and Schwartzman, M. N. (2009). *Understanding Physical, Health, and Multiple Disabilities*, 2nd edn. Upper Saddle River, NJ: Merrill/Pearson.
- Hill, J. L. (1999). *Meeting the Needs of Students with Special Physical and Health Care Needs*. Upper Saddle River, NJ: Merrill, Prentice Hall.
- Howell, J. J. and Gengel, S. (2005). Perspectives of effective teachers of students with low-incidence disabilities. *Teaching Exceptional Children Plus* 1(4), 1–13.
- Johnson, L., Beard, L. A., and Carpenter, L. B. (2007). *Assistive Technology: Access for All Students*. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Keyser-Marcus, L., Briel, L., Targett, P. S., et al. (2002). Enhancing the schooling of students with traumatic brain injury. *Teaching Exceptional Children* 34(4), 62–67.
- Shapiro, D. R. and Sayers, L. K. (2003). Who does what on the interdisciplinary team. *Exceptional Children* 35(6), 32–38.
- Smith, D. (2007). *Introduction to Special Education; Making a Difference*, 6th edn. Boston, MA: Pearson/Allyn and Bacon.
- Stansell, D. J. (2007). Giving a face to a hidden disorder: The impact of dyspraxia. *Teaching Exceptional Children Plus* 4(1), 1–8.
- Stein, E. K. (1989). *Caring for Children with Chronic Illness*. New York: Springer.
- The Power of Assistive Technology (2007). <http://www.connensebulletin.com/jendron.html> (accessed Jun. 2009).
- Turnbull, R., Turnbull, A., Shank, M., and Smith, S. J. (2004). *Exceptional Lives: Special Education in Today's Schools*. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Wolfe, P. S. and Hall, T. E. (2003). Making inclusion a reality for students with severe disabilities. *Teaching Exceptional Children* 35(4), 56–60.

Further Reading

- Male, M. (2003). *Technology for Inclusion: Meeting the Special Needs of All Students*, 4th edn. Boston, MA: Allyn and Bacon.

Educating Children with Specific Learning Disabilities

P A Grant, Slippery Rock University of Pennsylvania, Slippery Rock, PA, USA

P B Grant

© 2010 Elsevier Ltd. All rights reserved.

Educating Children with Specific Learning Disabilities

Children with specific learning disabilities (SLDs) are those children who have a disorder in one or more of the basic psychological processes involved in understanding or using language – spoken or written – which may manifest itself in imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations (Lloyd, 1996). Such disorders include conditions such as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. Such a term does not include children who have learning problems – which are, primarily, the result of visual, hearing, motor disabilities, mental retardation, emotional disturbance, or of environmental, cultural, or economic disadvantage (see Lloyd). Hunt and Marshall (2002) indicated that, according to the federal law, a student has a learning disability if he/she (1) does not achieve at the proper age level and ability level in one or more or several specific areas when provided with appropriate learning experiences; and (2) has a severe discrepancy between achievement and intellectual ability in one or more of the following areas: (a) oral expression, (b) listening comprehension, (c) written expression, (d) basic reading skills, (e) reading comprehension, (f) mathematics calculation, and (g) mathematics reasoning. The key element to the federal definition is the emphasis on the performance of students with LDs which is, often, tied to their ability to receive or express information (see Hunt and Marshall, 2002).

The SLD category is relatively new. It was first recognized by the federal government in 1969. An understanding of the history of SLD provides a context for understanding the many issues facing the field. This history has been well documented (e.g., Hallahan and Mercer, 2002; Wiedherhold, 1974). The LDs story actually began when Samuel Kirk introduced the term at a meeting of parents advocating special educational services for their children who were having difficulty in school but were not considered disabled by mental retardation or emotional disturbance (Kirk, 1975). In earlier decades, these children's difficulties had been variously categorized as mild exogenous mental retardation (mild mental retardation caused by brain injury), minimal brain dysfunction (behavior associated with brain injury, although brain damage cannot be verified), perceptual impairment (persistent difficulty in making sense of sensory stimulation), hyperactivity (excessive and socially

inappropriate behavior accompanied by problems in learning), and slow learning (a child whose achievement is not far enough below average to indicate mental retardation) (see Hallahan and Cruickshank, 1973; Hallahan and Kauffman, 1977; Mann, 1979; Wiedherhold, 1974). In the present day, people from nearly every walk of life – and every ethnicity – recognize the term learning disabled, which is often used generically by the general public to indicate that someone's behavior is highly unusual or inadequate for the circumstances. LD is a separate category in the special education literature – a disability defined by federal and state laws, and a specialization for which teachers in many states must obtain special certification. The term has gained almost universal acceptance among regular educators, special educators, and the general public in the United States and in many foreign countries (Mazurek and Winzer, 1994; Winzer, 1993).

Hallahan *et al.* (1996) concluded that the history of LDs can be summarized in relation to themes that have emerged as the field has developed. These themes are critical to understanding that LDs (1) represent an interdisciplinary field that is international and multicultural in scope; (2) have a variety of possible causes; (3) are, in part, a social construct; (4) are heterogeneous in nature; (5) vary in severity and pervasiveness; (6) are characterized by intra-individual differences; (7) affect a diverse group of people; (8) may coexist with other disabilities or giftedness; (9) require training in systematic approaches to tasks; (10) require educators to minimize the contribution of poor teaching; and (11) are developmental disorders persisting over the life span (p. 25). Apparently, LD is a perplexing disability that requires general and special educators dedicated to helping students to utilize their fullest potential. This is the focus of this article.

Definitions and Prevalence of SLD

The field of LDs is complicated by controversy in the areas of definition, prevalence, and demographics. Kirk (1962) defined LD as follows:

A learning disability refers to retardation, disorder, or delayed development in one or more of the processes of speech, language, reading, writing, arithmetic, or other school subject resulting from a psychological handicap caused by a possible cerebral dysfunction and/or emotional or

behavioral deprivation, or cultural and instructional factors. (p. 263)

There are five components of Kirk's definition that have appeared in many of the later definitions of LD: (1) subaverage achievement (reading, writing, or arithmetic) or achievement related (speech or language) behavior; (2) intraindividual differences – the possibility that the sub-average achievement or achievement-related behaviors occur(s) in only some areas or one area, with average or above-average achievement occurring in the other areas; (3) reference to psychological handicaps (often referred to as psychological processes by Kirk and others) as causal factors; (4) reference to cerebral dysfunction as a possible causal factor; and (5) exclusion of other disabling conditions (i.e., mental retardation) and environmental conditions as causal factors:

Bateman (1965), Kirk's student, offered the following definition:

Children who have learning disorders are those who manifest an educationally significant discrepancy between their estimated intellectual potential and actual level of performance related to basic disorders in the learning process, which may or may not be accompanied by demonstrable central nervous system dysfunction, and which are not secondary to generalized mental retardation, educational or cultural deprivation, severe emotional disturbance, or sensory loss. (p. 220)

Bateman's definition differed from Kirk's in at least two important respects. First, it did not include reference to emotional factors as a cause of LDs. In fact, it mentioned severe emotional disturbance as one of the disabilities that does not cause LDs. No major definitions since Bateman's have mentioned emotional disturbance as a possible causal factor. Second – and even more important – it included reference to a discrepancy between intellectual potential and actual performance.

Hallahan *et al.* (1996) summarized several important definitions of LDs over the years to include the following:

- Definitions proposed by task forces organized by the federal government, in 1966, that emphasized medical explanations and, in 1969, emphasized intra-individual differences and discrepancy between potential and performance.
- The definition suggested by the National Advisory Committee on Handicapped Children 1968 referred specifically to children, added thinking disorders to the list of examples of LDs, and did not refer to emotional disturbance as a causal factor.
- The authors of the definition proposed by the United States Office of Education (USOE), in 1977, considered the notion of developing a discrepancy formula; this

idea was met with an immediate and negative response, but the authors still referred to an ability-achievement discrepancy in their evaluation criteria.

- The National Joint Committee on Learning Disabilities (NJCLD), in 1981, reacted negatively to the USOE references to psychological processes.
- A major organization – the Learning Disabilities Association of America – did not endorse the NJCLD definition; instead, in 1986, it emphasized the lifelong nature of LDs, did not exclude other conditions as causal factors, and mentioned that adaptive behaviors can be affected.
- The Interagency Committee on Learning Disabilities 1987 definition was essentially the same as that of the NJCLD, but it added social skills deficits as a type of LD and attention deficit disorder as a potentially co-occurring condition. In 1988, the NJCLD revised its definition to acknowledge the lifelong nature of LDs and include socialization problems as potentially co-occurring, but not as an example of a type of LD.

Clearly, a new political and educational movement had been initiated. One of its first tasks was refining the above description to provide a more educationally focused definition (e.g., Bateman, 1965; Cruickshank, 1966; Kass and Myklebust, 1969). To bring order to the proliferation of SLD definitions, the National Advisory Committee on Handicapped Children (NACHC, 1968), headed by Samuel Kirk, was formed with the task of writing a definition of SLD that was to be the basis for Public Law 91-230 (Children with Specific Learning Disabilities Act). The definition, however, failed to clear delineation of LD parameters (Hammill, 1974), but when Public Law 94-142 – the Education for All Handicapped Children Act – was passed in 1975, SLD became an official category within the NACHC definition. With only minor wording, the NACHC definition was the one issued by the USOE. To this date, it remains unchanged as the definition of SLD used by the federal government. The definition reads as follows:

The term specific learning disability means a disorder in one or more of the psychological processes involved in understanding or in using language, spoken or written, which may manifest itself in an imperfect ability to listen, speak, read, write, spell, or to do mathematical calculations. The term includes such conditions as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. The term does not include children who have learning disabilities which are primarily the result of visual, hearing, or motor handicaps, or mental retardation, or emotional disturbance, or of environmental cultural, or economic disadvantage. (USOE, 1997: 65083)

As it appears, the above federal definition did not specify how states were to identify students as SLD. To remedy this situation, the USOE issued regulations intended to provide criteria for determining eligibility. These regulations represent an operational definition of SLD and center on the presence of an ability-achievement discrepancy:

A team may determine that a child has a specific learning disability if:

1. The child does not achieve commensurate with his or her age and ability levels in one or more of the areas listed in (a) (2) of this section, when provided with learning experiences appropriate to the child's age and ability levels; and
2. The team finds that the child has a severe discrepancy between achievement and intellectual ability in one or more of the following areas: (i) oral expression; (ii) listening comprehension; (iii) written expression; (iv) basic reading skill; (v) reading comprehension; (vi) mathematical calculation; or (vii) mathematical reasoning (USOE, 1997: 65083).

Lokerson (1992) noted that although the definition in federal law governs the identification of – and services to – children with LD, there are variations between states and among school systems. In an attempt to clarify the identification, some states specify an intelligence range. Others add a concept of a discrepancy between potential and achievements, sometimes quantifying the discrepancy using test scores. These slightly different yardsticks are indicative of a lack of a clear consensus regarding exactly what LDs are.

According to the figures maintained by the federal government, public schools have identified approximately 2.3 million students (~6–21 years) of age as learning disabled, which represents 4.09% of those in this age range (US Department of Education, 1994). However, this is a slight underestimate, because students in private schools identified with LDs have tripled. Critics have suggested that this rapid growth is evidence that LDs is an ill-defined category; and even defenders of the field are concerned that much of the growth is indicative of the confusion over definition and diagnostic criteria. Others have noted that the increase in the numbers of students identified with LDs is inversely related to the numbers identified for several reasons. Hallahan (1992) speculated that there may be many reasons for some of the growth. The field of LDs was quite new, in 1976, when the government started maintaining prevalence data, and it may have taken professionals several years to decide how to identify children. Second, he noted that many social and cultural changes have occurred in the last 30 years to heighten children's vulnerability to develop LDs. Baumeister *et al.* (1990) suggested that an increase in poverty has

placed more children at risk for biomedical problems, including central nervous system dysfunction. Hallahan, further, concluded that stress on parents may result in them being less able to provide the social support necessary to help their children, who themselves are experiencing an increasing amount of stress. The result may be that children who – in a previous time of less stressful lifestyle and more support – would have gotten by in their school work are now experiencing failure.

Characteristics of SLD

Early Childhood

The SLD category is reserved for children who experience academic failure. However, a variety of cognitive and behavioral characteristics place children at risk for academic failure. For example, poverty, poor nutrition, and environmental hazards are risk factors for many disabilities. Because SLD includes multiple and complex causes that manifest themselves in a variety of conditions varying in type and severity, children younger than 6 years of age are, often, given a noncategorical designation such as developmental delay (Stichter *et al.*, 2008). The 1997 reauthorization of Individuals with Disabilities Education Improvement Act (IDEA) incorporated early-childhood special education laws that ensure the same rights for younger children (age range: 3–5).

Younger children may possess delays in one or more of the following areas: (1) physical development, (2) cognitive development, (3) communication development, or (4) adaptive development. Specifically, precursors of SLD may be found in the following areas: (1) gross motor skills, (2) fine motor skills, (3) auditory processing, (4) visual processing, (5) linguistic processing, or (6) attention problems. The general term developmental delay is useful because children may be at risk for SLD, mental retardation (MR), or EBD. The goal is not to identify a specific disability but, rather, to provide early intervention to prevent future problems.

Elementary Grades

The federal definition describes SLD as

a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations.

An SLD is conceptualized as a deficit in one or more processes (e.g., listening, memory, perceptions, attention) assumed to be the underlying reason for academic difficulties (Stichter, 2008).

Academic difficulties may manifest themselves in reading, writing, and arithmetic. Severe difficulty in learning to read is alexia and minor is dyslexia; severe difficulty in learning to write is agraphia and minor is dysgraphia; severe difficulty in learning to solve mathematics is acalculia and minor is dyscalculia; and severe difficulty in learning to express oneself is aphasia and minor is dysphasia. Although possessing average or above-average intelligence, students with dyslexia have significant difficulties reading, spelling, understanding language, and expressing themselves in speaking and writing. Even though there is disagreement with regard to how dyslexia should be defined, there is agreement concerning the following (Hynd, 1992):

- dyslexia has a biological basis and is caused by a different anatomical brain structure;
- dyslexia persists in adolescence and adulthood; and
- dyslexia has linguistic, cognitive, and perceptual manifestations.

The National Reading Panel – in a comprehensive review of reading literature, in 2000 – determined that five areas in reading should receive primary instructional focus: phonemic awareness, phonics, fluency, vocabulary, and comprehension. These areas connect and build upon one another, leading to proficient reading. Regrettably, for nearly 90% of students with SLD (Kavale and Forness, 2000), reading problems are a significant cause of low performance. The basic nature of the problem is found at the linguistic level of phonology (Liberman and Shankweiler, 1985), where a critical aspect of reading involves manipulating phonemes – the smallest sound unit of language (Liberman and Shankweiler, 1991). The most basic deficit of students with SLD is the lack of phonemic awareness – the understanding that spoken language is composed of individual phonemes (Brady, 1997; Rack *et al.*, 1992). Without knowledge of the phonological structure of language, the student with SLD is likely to also lack phonological awareness, the ability to blend (connect sounds of individual phonemes into syllables and words), segment (dividing words into their individual phonemes), and rhyme (finding words that sound the same) (Blachman, 2001; Mutter, 1998; Wagner and Torgesen, 1987). In sum, a student's initial difficulties in phonemic awareness leads to difficulty with phonological awareness, which leads to decoding that is less than automatic. When students do not develop automaticity in decoding, they have to decode word-by-word, which reduces fluency. This reduction in fluency translates into inadequate comprehension, as students are expending so much cognitive energy on decoding that they have little left for comprehending what they are reading (Stichter *et al.*, 2008).

Students with an SLD perform poorly on written expression tasks, especially those involving vocabulary,

grammar, punctuation, and spelling (Newcomer, and Barenbaum, 1991). Students with SLD approach writing with minimal planning, effort, and strategic behavior (Thomas *et al.*, 1987). The consequences are found in difficulties in text production, limited knowledge of the writing process (e.g., drafting and revising), poor idea generation and organization, and improvised theme development (Graham *et al.*, 1991). Students with SLD may also experience difficulties with the mechanics of writing (handwriting) – as manifested in problems in letter formation and fluency. Additionally, students with SLD may experience math difficulties related to numerical reasoning and calculation. Compared to average-achieving peers, students with SLD perform more poorly in every aspects of arithmetic at every grade level (Crawley *et al.*, 2001). Notable are deficits in retrieving number facts and solving story problem (Geary *et al.*, 2000). Students with SLD progress slowly in math, and many plateau by age 12 (Crawley *et al.*, 1998).

Students with SLD demonstrate many processing deficits. In terms of cognitive styles (approaches to problem solving), students with SLD are, generally, more field dependent (i.e., influenced more by environment) and more impulsive (i.e., respond quickly with little reflection) (Blackman and Goldstein, 1982). Memory difficulties often occur in students with SLD. Short-term memory (ability to remember information immediately) tends to be more deficient in the auditory than in the visual realm (Hulme and Snowling, 1992). Reading and math problems appear to be more adversely influenced by deficits in working memory (ability to retain information while performing a mental operation) (Swanson and Ashbaker, 2000; Swanson and Sachse-Lee, 2001).

Motivation (i.e., activating, guiding, and maintaining behavior) is a problem area for students with SLD that results from three sources not limited to:

- An external locus of control in which the student with SLD waits for others to organize their behavior (Bryan and Pearl, 1979).
- Negative attributions (i.e., belief about the causes of success and failure) in which students with SLD are often not proud of success, minimize accomplishments, and are more prone to accept responsibility for failure. Such negative attributions tend to lower belief in self-efficacy (Tabassam and Grainger, 2002).
- Learned helplessness (i.e., belief that effort will not result in desired outcomes) which makes the student with SLD come to expect failure no matter how much effort he/she expends (Settle and Milich, 1999).

The presence of attention problems is an important component of SLD (Hallahan and Reeve, 1980). Behavioral problems appear at a higher-than-expected rate among students with SLD (Kavale and Forness, 1998).

The cause-and-effect relationship between behavioral problems and academic difficulties remains unclear. Additionally, many students with SLD demonstrate no behavioral problems. Students with SLD are more likely to manifest social deficits. Earlier, Kavale and Forness (1996) found that almost 75% of these students exhibited deficits in social skills. They appear less socially accepted because of difficulties in using language in social situations, in being sensitive to social cues, in correctly perceiving their social status, and in adapting to social situations (Sridhar and Vaughn, 2001). A fundamental difficulty contributing to poor social skills may be an inability to recognize emotions in others – in particular, nonverbal emotional expressions. Moreover, the same academic and social problems found in children with SLD are likely to continue into adolescence and adulthood (Kavale, 1988).

The most recent 2004 reauthorization of the IDEA, introduces a significantly different operational definition of SLD in which (1) the local education agency (school system) shall not be required to take into consideration whether a child has a severe discrepancy between achievement and intellectual ability, and (2) the local education agency may use a process that determines whether a child responds to scientific, research-based interventions. The proposed process is usually termed response to intervention (RTI), in which schools provide differing types and levels of instruction for low achieving students (usually in reading). The RTI model can be implemented in different ways that include variations in instructional interventions, levels of intervention before special education, and means by which students are deemed nonresponders (Fuchs *et al.*, 2003). Additionally, Graden (1989) suggested that the RTI model is best viewed as a pre-referral strategy in which instructional or organizational suggestions for enhanced achievement are provided prior to a referral for special education evaluation.

Assessing Students with SLD

The assessment of students with SLD remains as problematic as the definition itself. Ysseldyke *et al.* (1983) summarized several issues related to learners identified as learning disabled. In 5 years of research, they found no reliable psychometric differences between students labeled learning disabled and those who were perceived to be low achievers. In addition, identification as learning disabled depended on the criteria and definitions used. If that child moved to a different school district, he or she might no longer be identified as being learning disabled (Shea and Bauer, 1994). Algozzine (1991) argued that problems related to the identification of learners as learning disabled begin at the point of referral. The first problem is teacher reason for referral. For example, the

new move toward excellence in education may increase the likelihood that low-performing students will be referred to as potentially disabled and appropriate for special education. For the African American learner, it will, most often, be based on behavior; therefore, they will be referred for the emotionally disturbed or the mentally retarded classification. The second problem includes the high rate of referrals, with a shortage of individuals well prepared and certified to conduct the necessary evaluation. The third problem is that over 90% of students referred, and over 70% of those tested are then placed in special education programs.

The diagnosis of LD, as stated earlier, can be difficult, especially in the very early grades. However, if one follows some practical guidelines, the diagnosis can be easily substantiated. In order to diagnose a child with a suspected LD, the following procedures and criteria should be considered.

Clinical Interview

Clinical interview involves a series of interviews with the child to assess where the ultimate problem may lie. When interviewing the child with a suspected LD, one may want to look and listen for confusion over questions, poor use of vocabulary, problems expressing ideas and thoughts, awkward gait, poor memory, short attention span, lack of focus, poor fine-motor skills, and a history of academic difficulties.

Ecological Assessment

Ecological assessment involves observing the child in a variety of settings such as the classroom, playground, and other structured and nonstructural settings to determine where the student manifests the greatest difficulties. In the case of a child with a suspected LD, one may observe social withdrawal, alienation from peers, inability to focus in unstructured setting, and class-clown-type behavior as a means of being removed from academically stressful setting.

Parent Interview

Parent interview requires a personal meeting with parents to determine background history that may be essential for appropriate diagnosis. In the case of a child with a suspected LD, a parent might observe behaviors such as (1) has difficulty dressing him or herself, (2) avoids house work, (3) is disorganized, (4) has a short attention span, (5) forgets easily, (6) forgets to bring books home, (7) gets stomachaches in the morning before school, (8) gets frequent headaches, (9) has few friends, (10) is unwilling to try new things, and (11) gives up easily.

Teacher Interview

Teacher interviews may require several meetings with the classroom teacher to ascertain the child's basic intellectual, social, and academic performance. In the case of a child with potential LD, the interviewer should be aware of certain LD symptom clusters that may appear in the classroom. Some examples may include poor memory, gross-motor coordination difficulties, lack of focus, short attention span, procrastination, failure to hand in written work or homework, lack of confidence, self-derogatory statements such as "I'm so stupid," consistently low academic performance in certain subjects over time, social difficulties, lack of motivation for schoolwork, poor handwriting, and poor fine-motor skills.

Review of Cumulative Reports and Records

Review of report cards, attendance records, standardized tests, and so forth – to determine possible patterns of behavior – can be helpful. In the case of the child with potential LD, a review of cumulative records may reveal consistently low group of achievement scores in certain subjects over a period of years, past teacher comments showing a pattern to what the child's present teacher reports, a historical pattern of academic difficulties, frequent absences (which may occur when the child feels frustrated and overwhelmed by the work), and a discrepancy between ability and class performance as indicated by report card patterns.

Intelligence Testing

In the case of a child with a suspected LD, the psychologist will administer an individual intelligence test to look for an average to above average intellectual level. This does not mean that the child's IQ need be in the average range, but analysis of the profile indicates that greater potential be shown through significant uneven pattern between scores within the same test. Many children are commonly misdiagnosed as having a LD when, in actuality, they may be slow learners, children with emotional issues, or underachievers not performing for reasons other than a LD.

Achievement Testing

Children with LDs usually exhibit a severe discrepancy between potential ability (as measured on an individual IQ test) and academic achievement. This is a debatable criterion because it is possible that a child functioning on grade level – according to standardized achievement tests – may actually have a severe discrepancy if one takes into account ability levels (Pierangelo and Giuliani, 2002). Turnbull *et al.* (2002) reported that, in diagnosing LD, it is necessary to use a nondiscriminatory approach to

insure compliance with the mandates of the IDEA. An important aspect of nondiscriminatory evaluation for students with LDs is establishing a discrepancy between ability, as measured by an IQ test, and achievement – as measured by a standardized test.

Instructional Considerations for Students with SLD

One of the best ways for teachers to address the needs of students with LDs is to adopt instruction and materials. Many textbooks include teacher's manuals to assist in the presentation of instruction. Too often, however, the lessons in the manual are brief and potentially confusing. Teachers and service providers need to modify these lessons before presenting them to the class. (see Hunt and Marshall, 2002). They should understand that individuals with LDs have different learning styles – find how best a child learns, and teach to his/her strength. It is important to involve parents and extended family in the education of the child to increase the opportunity for the child to learn at the maximum level.

In addition, competent teachers and service providers should understand:

- Students with SLD are most likely to have reading difficulties, followed by math and writing problems.
- Students with SLD have learning strengths and weaknesses.
- Instructional goals must be comprehensive and unified.
- Intervention needs to be based on a diagnostic-prescriptive approach that includes assessment, planning, implementation, evaluation, and modification as needed.
- Students with SLD required intense and sustained interventional activities.
- There is a need for active learning to overcome the inactive learning of students with SLD.
- Best practice is predicated on using research-supported materials.
- Students with SLD will, likely, require specialized methods for teaching basic skills.
- Content-area instruction should be based on the use of learning strategies and study skills.
- There is a necessity for modifying the pace of instruction and providing organizational cues.
- There is a need for students with SLD to experience success.

Solutions to remedy problem confronted by students with SLD lies with good teachers, understanding administrators, and sound family, and extended family support. The key points to remember are:

- Family support is a must.
- Caring teachers are essential.

- Good teachers are crucial.
- In-school support systems are a necessity.
- Understanding how children learn is a major component.
- Adapting the classroom environment is necessary to meet the needs of the student.
- Adapting material to help students understand it.
- Using peer tutors to increase opportunity helps to increase chance of friendship in the classroom.
- Patience is a must.
- Allowing students to be creative in their presentation of information – to be practiced always.

Conclusion

This article explores many areas of specific LDs. It is understood that the judgment of whether a LD is a difficult one involves consideration of the students intelligence, achievement, past educational experience, age, rate of learning with various instructional approach, quality of present and past teaching, motivation and many other personal curriculum, and environmental factors. It is clearly understood that students with SLDs are bright enough to have learned the basic skills and information required by their school. The learning strategies and information-processing skills of most individuals with LDs are typically expressed, but they do so at a much slower pace.

The role of pollutants and teratogen (aspects of the environment that causes development mental malformation in humans) in causing LDs is well-known research; however, Murphy and Oei (1999) noted that we continue to expand our understanding of how exposure to chemicals during pregnancy and early developmental years contribute to LDs in a newborn or young children. This is a body of research that cannot be ignored based on the influences of our ever-expanding technological development today. With the passing of special education laws P.L. 94-142, in 1975 (The education of all handicapped children act), and amended to P.L. 101-76, in 1990 (ADA), the effectiveness of our teacher-training programs, regarding the treatment, care, understanding, and education of these children have made an impact on their lives as professionals have learned to work with and help them to overcome this disability. At present, the focus is on achieving the expectation of serving these children in what is presently considered to be the least restrictive environment – the regular education classrooms.

Bibliography

- Algozzine, B. (1991). Decision-making and curriculum-based assessment. In Wong, B. Y. L. (ed.) *Learning about Learning Disabilities*, pp 40–55. San Diego, CA: Academic Press.
- Bateman, B. D. (1965). An educator's view of a diagnostic approach to learning disorders. In Hellmuth, J. (ed.) *Learning Disorders*, vol. 1, pp 219–239. Seattle, WA: Special Child Publication.

- Baumeister, A. A., Kupstas, F., and Klindworth, L. M. (1990). New morbidity: Implications for prevention of children's disabilities. *Exceptionality* 1(1), 1–6.
- Blachman, B. A. (2001). Phonological awareness. In Pearson, P. D. (ed.) *Handbook of Reading Research*, pp 483–502. Mahwah, NJ: Erlbaum.
- Blackman, S. and Goldstein, K. M. (1982). Cognitive styles and learning disabilities. *Journal of Learning Disabilities* 15, 106–115.
- Brady, S. A. (1997). Ability to encode phonological representations: An underlying difficulty of poor readers. In Blachman, B. A. (ed.) *Foundations of Reading Acquisition and Dyslexia for Early Interventions*, pp 21–47. Mahwah, NJ: Erlbaum.
- Bryan, T. S. and Pearl, R. (1979). Self-concept and locus of control of learning disabled children. *Journal of Clinical Child Psychology* 8, 223–226.
- Crawley, J. F., Parmar, R. S., Foley, T. E., Salmon, S., and Roy, S. (2001). Arithmetic performance of students: Implications for standards and programming. *Exceptional Children* 67, 311–318.
- Crawley, J. F., Parmar, R. S., Yan, W., and Miller, J. H. (1998). Arithmetic computation performance of students with learning disabilities: Implications for the curriculum. *Learning Disabilities Research and Practice* 23, 68–74.
- Cruikshank, W. D. (ed.) (1966). *The Teacher of Brain-Injured Children*. Syracuse, NY: Syracuse University Press.
- Fuchs, D., Mock, D., Morgan, P. L., and Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence and implications for the learning disabilities construct. *Learning Disabilities Research and Practice* 18, 157–171.
- Geary, D. C., Hamson, C. O., and Hoard, M. K. (2000). Numerical and arithmetic cognition: A longitudinal study of process and concept deficits in children with learning disabilities. *Journal of Experimental Child Psychology* 77, 236–263.
- Graden, J. L. (1989). Redefining “pre-referral” intervention as intervention assistance: Collaboration between general and special education. *Exceptional Children* 56, 227–231.
- Graham, S., Harris, K. R., MacArthur, C. A., and Schwartz, S. (1991). Writing and writing instruction for students with learning disabilities: A review of a research program. *Learning Disability Quarterly* 14, 89–114.
- Hallahan, D. P. (1992). Some thoughts on why the prevalence of learning disabilities has increased. *Journal of Learning Disabilities* 25(8), 523–528.
- Hallahan, D. P. and Cruickshank, W. M. (1973). *Psycho-Educational Foundations of Learning Disabilities*. Englewood Cliffs, NJ: Prentice-Hall.
- Hallahan, D. P. and Kauffman, J. M. (1977). Labels, categories, behaviors: ED, LD, and EMR reconsidered. *Journal of Special Education* 11, 139–149.
- Hallahan, D. P., Kauffman, J. M., and Lloyd, J. W. (1996). *Introduction to Learning Disabilities*. Needham Heights, MA: Allyn and Bacon.
- Hallahan, D. P. and Mercer, C. D. (2002). Learning disabilities: Historical perspectives. In Bradley, R., Danielson, L., and Hallahan, D. P. (eds.) *Identification of Learning Disabilities: Research to Practice*, pp 1–67. Mahwah, NJ: Erlbaum.
- Hallahan, D. P. and Reeve, R. E. (2002). Selective attention and distractibility. In Keogh, B. K. (ed.) *Advances in Special Education*, vol. 1, pp 141–181. Greenwich, CT: JAI Press.
- Hammill, D. D. (1974). Learning disabilities: A problem in definition. *Divisions for Children with Learning Disabilities Newsletter* 4, 28–31.
- Hulme, C. and Snowling, M. (1992). Phonological deficit in dyslexia: A “sound” reappraisal of the verbal deficit hypothesis. In Singh, N. and Beale, I. (eds.) *Learning Disabilities: Nature, Theory and Treatment*, pp 270–301. New York: Springer.
- Hunt, N. and Marshall, K. (2002). *Exceptional Children and Youth: An Introduction to Special Education*, 3rd edn. Boston, MA: Houghton Mifflin.
- Hynd, G. (1992). Neurological aspects of dyslexia: Comments on the balance model. *Journal of Learning Disabilities* 25, 110–113.
- Individuals with Disabilities education improvement act (IDEA) (2003). Senate Bill 1248.
- Kass, C. E. and Myklebust, H. R. (1969). Learning disability: An educational definition. *Journal of Learning Disabilities* 2, 377–379.

- Kavale, K. A. (1988). The long-term consequence of learning disabilities. In Wang, M. C., Reynolds, M. C., and Walberg, H. J. (eds.) *Handbook of Special Education: Research and Practice*, Vol. 2: *Mildly Handicapping Conditions*. New York: Pergamon.
- Kavale, K. A. and Forness, S. R. (1996). Social skill deficits and learning disabilities: A meta-analysis. *Journal of Learning Disabilities* **29**, 226–237.
- Kavale, K. A. and Forness, S. R. (1998). Covariance in learning disabilities and behavior disorders: An examination of classification and placement. In Scruggs, T. and Mastropieri, M. (eds.) *Advances in Learning and Behavioral Disabilities*, vol. 12, pp 1–14. Greenwich, CT: JAI Press.
- Kavale, K. A. and Forness, S. R. (2000). What definitions of learning disability say and don't say: A critical analysis. *Journal of Learning Disabilities* **33**, 239–256.
- Kirk, S. A. (1962). *Educating Exceptional Children*. Boston, MA: Houghton-Mifflin.
- Kirk, S. A. (1975). Behavioral diagnosis and remediation of learning disabilities. In Kirk, S. A. and McCarthy, J. J. (eds.) *Learning Disabilities: Selected Papers*, pp 7–10. Boston, MA: Houghton-Mifflin.
- Liberman, I. Y. and Shankweiler, D. (1985). Phonology and the problems of learning to read and write. *Remedial and Special Education* **6**, 8–17.
- Liberman, I. Y. and Shankweiler, D. (1991). Phonology and beginning reading: A tutorial. In Reibent, L. and Perfetti, C. A. (eds.) *Learning to Read: Basic Research and its Implications*, pp 3–17. Hillsdale, NJ: Erlbaum.
- Lloyd, J. (1996). What is learning disability? (1 paragraph). *The Curry School of Education at the University of Virginia, Office of Special Education: A Web resource for Special Education*. <http://curry.edschool.virginia.edu/go/cise/ose/home.html> (accessed August 2009).
- Lokerson, J. (1992). *Understanding Learning Disabilities: A Parent Guide and Workbook*, 2nd edn. Richmond, VA: The Learning Disabilities Council.
- Mann, L. (1979). *On the Trail of Process*. New York: Grune and Stratton.
- Mazurek, K. and Winzer, M. A. (1994). *Comparative Studies in Special Education*. Washington, DC: Gallaudet University Press.
- Murphy, M. and Oei, P. S. (1999). Is there evidence to show that fetal alcohol syndrome can be prevented? *Journal of Drug Education* **29**, 5–24.
- Mutter, V. (1998). Phonological awareness: Its nature and its influence over early literacy development. In Hulme, C. and Joshi, R. M. (eds.) *Reading and Spelling: Development and Disorder*, pp 113–125. Mahwah, NJ: Erlbaum.
- NACHC (National Advisory Committee on Handicapped Children) (1968). *First Annual Report to Congress on Special Education for Handicapped Children*. Washington, DC: Department of Health, Education, and Welfare.
- Newcomer, P. L. and Barenbaum, E. M. (1991). The written composing ability of children with learning disabilities: A review of the literature from 1980 to 1990. *Journal of Learning Disabilities* **24**, 578–593.
- Pierangelo, R. and Giuliani, G. A. (2002). *Assessment in Special Education: A Practical Approach*. Boston, MA: Allyn and Bacon.
- Rack, J. P., Snowling, M. J., and Olson, R. K. (1992). The non-word reading deficit in developmental dyslexia: A review. *Reading Research Quarterly* **27**, 311–322.
- Settle, S. A. and Milich, R. (1999). Social persistence following failure in boys and girls with LD. *Journal of Learning Disabilities* **32**, 201–212.
- Shea, T. M. and Bauer, A. M. (1994). *Learnners with Disabilities: A Social Systems Perspective of Special Education*. Dubuque, IA: Brown and Benchmark Communications.
- Sridhar, D. and Vaughn, S. (2001). Social functioning of students with learning disabilities. In Hallahan, D. and Keogh, B. (eds.) *Research and Global Perspectives in Learning Disabilities*, pp 65–91. Marwah, NJ: Erlbaum.
- Stichter, J. P., Conroy, M. A., and Kauffman, J. M. (2008). *An Introduction to Students with High Incidence Disabilities*. Upper Saddle River, NJ: Prentice Hall.
- Swanson, H. L. and Ashbaker, M. H. (2000). Working memory, short-term memory, speech rate, word recognition, and reading comprehension in learning disabled readers: Does the executive system have a role? *Intelligence* **28**, 1–30.
- Swanson, H. L. and Sachse-Lee, C. (2001). Mathematical problem-solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Exceptional Child Psychology* **79**, 294–321.
- Tabassam, W. and Grainger, J. (2002). Self-concept, attributional style and self-efficacy beliefs of students with learning disabilities with and without attention deficit hyperactivity disorder. *Learning Disability Quarterly* **25**, 141–151.
- Thomas, C. C., Englert, C. S., and Gregg, S. (1987). An analysis of errors and strategies in the expository writing of learning disabled students. *Remedial and Special Education* **8**, 21–39/46.
- Turnbull, R., Turnbull, A., Shank, M., Smith, S., and Leal, D. (2002). *Exceptional Lives: Special Education in Today's Schools*, 3rd edn. Upper River Saddle, NJ: Merrill-Prentice Hall.
- U.S. Department of Education (1994). *Sixteenth Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act*. Washington, DC: U.S. Department of Education.
- USOE (U. S. Office of Education) (1997). Assistance to states for education of handicapped children: Procedures to evaluating specific learning disabilities. *Federal Register* **42**, 65082–65085.
- Wagner, R. K. and Torgesen, J. K. (1987). The nature of phonological processing and its casual role in the acquisition of reading skills. *Psychological Bulletin* **101**, 192–212.
- Wiederhold, J. L. (1974). Historical perspectives on the education of the learning disabled. In Mann, L. and Sabatino, D. (eds.) *The Second Review of Special Education*, pp 103–152. Philadelphia, PA: JSE Press.
- Winzer, M. A. (1993). *The History of Special Education: From Isolation to Integration*. Washington, DC: Gallaudet University Press.
- Ysseldyke, J. E., Thurlow, M. L., Graden, J. L., et al. (1983). Generalizations from five years of research on assessment and decision-making: The University of Minnesota Institute. *Exceptional Education Quarterly* **4**(1), 75–93.

Further Reading

- Bender, W. N. (1995). *Learning Disabilities: Characteristics, Identification, and Teaching Strategies*, 2nd edn. Boston, MA: Allyn and Bacon.
- Bredenkamp, S. and Copple, C. (eds.) (1997). *Developmentally Appropriate Practice in Early Childhood Programs*. Washington, DC: National Association for the Education of Young Children.
- Eiserman, W. D., Weber, C., and McCoun, M. (1995). Parent and professional roles in early interventions: A longitudinal comparison of the effects of two interventions. *Journal of Special Education* **29**, 20–44.
- Feldman, R. G. and White, R. F. (1992). Lead neurotoxicity and disorder learning. *Journal of Child Neurology* **7**, 354–359.
- Haring, K. A., Loret, D. L., Haney, K. F., et al. (1992). Labeling preschoolers as learning disabled: A cautionary position. *Topics in Early Childhood Special Education* **12**, 151–173.
- Minder, B., Das-Smaal, E. A., Brand, E. F., and Orlebeke, J. F. (1994). Exposure to lead and specific attention problems in schoolchildren. *Journal of Learning Disabilities* **27**(6), 393–399.
- Stanovich, K. E. (1989). Has the learning disabled field lost its intelligence? *Journal of Learning Disabilities* **22**(8), 487–492.
- Swanson, H. L. and Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research* **68**, 277–321.
- U. S. Department of Education (1999). Assistance to states for the education of children with disabilities and the early intervention programs for infants and toddlers with disabilities. Federal regulations. *Federal Register* **64**, 12406–12672.
- Vaughn, S., Gersten, R., and Chard, D. (2000). The underlying message in LD intervention research: Findings from research syntheses. *Exceptional Children* **67**, 99–114.

Educating Children with Speech and Language Disorders

S L Green and C D Qualls, State University of New York College at Buffalo, Buffalo, New York, USA

© 2010 Elsevier Ltd. All rights reserved.

The numbers of school-age children with speech and language disorders (SLD) in the US and Canada have continued to rise. Advances in medicine, genetics, and technology as well as recent changes in the laws affecting services to individuals with disabilities contribute to these increasing numbers, particularly as more students with disabilities attend school. According to the American Speech–Language–Hearing Association (ASHA) “communication disorders are among the most common disabilities in the United States.” (ASHA, 2008). A communication disorder can be manifested as a unitary speech, language, or hearing disorder, or as a combination disorder with one or more of these aspects of communication being affected. The National Information Center for Children and Youth with Disabilities (NICYD) reported that one in every ten people in the US have a communication disorder, and that over “one million of the students served in the public schools’ special education programs in the 1997–98 school year were categorized as having a speech or language impairment” (NICYD, 2008). The Canadian Association of Speech–Language Pathologists and Audiologists (CASLPA) reported that (CASLPA, 1996) the prevalence of SLD in children between birth and 19 years of age was as high as 10%. In some countries other than the US and Canada, for example, in China and Africa, the numbers may be even greater.

Approximately 50% of the 135 000 members of ASHA (2009) work as speech–language pathologists (SLPs) in an educational setting, many of them in the public schools. An SLP receives training on the nature, and causes, identification, prevention, assessment, and treatment of SLD. Also trained in interviewing and counseling, SLPs work closely with other educational and health professionals to ensure that individuals with SLD and their families receive the highest-quality services.

This article focuses on how those individuals can maximize their fullest potential. In addition, the roles of assessment, referral, intervention, and follow-up in the academic and social success are discussed.

Speech and Language Disorders

Speech is verbal expression for human communication. Students with good expressive abilities can produce or articulate the different sounds of speech, they can speak clearly and without difficulty (e.g., with ease and appropriate rhythm), and they are able to produce good voice

(e.g., adequate volume and no hoarseness/breathiness) for articulate/intelligible speech. A speech disorder may be present in the absence of one or more of these characteristics. Childhood speech disorders are categorized as childhood apraxia of speech (CAS), orofacial myofunctional disorders (OMD), speech sound disorders, articulation and phonological processes, stuttering, and voice (ASHA, 2008).

Childhood Apraxia of Speech

ASHA’s *ad hoc* committee on CAS defines CAS as a neurological childhood speech sound disorder that affects the precision and consistency of the movements necessary for smooth, articulate speech. A child with CAS has normal muscle tone and normal reflexes, yet has difficulty producing the sounds of speech. It is believed that the core impairment is in planning and/or programming the spatio-temporal movement sequences for the sounds and the rhythm of speech.

Orofacial Myofunctional Disorders

OMD affect the muscles (myo) of the mouth (oro) and face (facial). Children with OMD have difficulty with speech and swallowing because the tongue-thrust behaviors seen in the normal infancy fail to disappear as the child develops. Tongue thrust occurs when the tongue moves forward in an exaggerated way between the upper and lower teeth during speech and/or swallowing. School-age children with OMD tend to experience social isolation because they may be self-conscious about their physical appearance and their speech.

Speech Sound Disorders: Articulation and Phonological Processes

As children develop, many will typically make some mistakes in their speech, particularly in early childhood when they attempt to learn new and complex words. A speech sound disorder exists when such mistakes continue beyond the development age at which the child should correctly produce the sound(s). The prevalence of children with speech sound disorders range from 2% in older children (8 years old) to 24.6% in younger children (5 years old). The sound errors will render the child’s speech unintelligible to those unfamiliar with the child. Children with speech sound disorders can have problems with articulation

(making sounds) and/or phonological processing (sound patterns). An articulation disorder is characterized by speech sound production errors (e.g., rabbit is pronounced wabbit). A phonological disorder is characterized by the inability to follow the rules that govern production of sound combinations (e.g., sounds that require the tongue to be placed in back of the mouth are produced with the tongue in the front of the mouth (e.g., cat is pronounced tat) or vice versa (e.g., ball is pronounced gall). Early phonological disorders are related to later/subsequent reading, writing, spelling, and mathematical abilities.

Stuttering

Stuttering occurs when the easy onset and continuous flow of speech (termed fluency) are disrupted. Disruptions, or disfluencies (e.g., hesitations, or repetitions and prolongations of sounds or words), in the flow of speech occur in the speech of most people at times. However, when a person produces too many disfluencies, as seen in persons who stutter, communication often suffers. Stuttering begins in early childhood and, for some people, lasts a lifetime. Further, stuttering behaviors greatly vary across individuals and situations. The impact of stuttering for a child in the school setting is affected by his/her own speech and associated behaviors (e.g., avoidance of certain activities and rearranging the words in their sentence), coupled with others' reaction to their speech. Some children who stutter have other speech or language problems. In some cases, the coexisting speech or language problem will be the impairment that is treated first and, as time permits, the stuttering may be addressed (Blood *et al.*, 2003). These children often experience social isolation and are sometimes bullied by others (see Blood *et al.*, 2003).

Voice Disorders

Childhood or pediatric voice disorders are more common than one might expect. Voice disorders are characterized by aphonia (no voice), abnormal vocal quality (i.e., hoarseness, harshness, and breathiness), pitch that is too high or low, volume that is too loud or soft, impaired nasal resonance, and/or inadequate voice support to produce intelligible speech. A reported 6–23% of school-age children have hoarse voices; hence, the most common voice problem in school-age children is hoarseness related to vocal hyperfunction (i.e., yelling, screaming, crying, loud volume, and excessive talking). Children with voice disorders respond positively to treatment that addresses the source of the voice abuse/misuse and strategies to reduce those behaviors. Early identification and treatment of a childhood voice disorder is recommended to avoid a lifelong communication impairment, particularly because it is thought that the majority of adult voice disorders begins in childhood. Few school-age children receive direct

voice rehabilitation because their difficulty is not considered to be serious enough to be placed on the active caseload (Hooper, 2004). Children with voice disorders can experience difficulties in their academic performance and socialization.

Language Disorders

Language consists of socially shared rules and students who understand and appropriately use these shared rules will be successful not only academically, but also in their social interactions. Students with age-appropriate language:

1. have a good working knowledge of vocabulary;
2. understand that some words have multiple meanings (i.e., literal and figurative meanings); and
3. know how to make new words, put words together, and know what word combinations are best in what situations.

A language disorder impairs one's abilities to comprehend and/or use spoken or written language or other symbol systems (e.g., gestures or sign language). Language impairments may be manifest in the form (grammatical rules), content (e.g., understanding word meanings and creating words, and the ability to follow commands), and/or function (i.e., conversational skills, eye contact, turn-taking, and appropriate use of emphasis) of language in communication. A child's language can be delayed or disordered. A language delay means that the child will develop language skills at a later time than their age peers. A language disorder occurs when, as stated above, one or more language skills are impaired. Childhood language disorders are classified as language-based learning disabilities (LBLD), specific language impairment, and selective mutism.

Language-Based Learning Disabilities

Learning disabilities represent "a heterogeneous group of disorders manifested by significant disabilities in the acquisition and use of listening, speaking, reading, writing, reasoning, or mathematical abilities" (Hammill *et al.*, 1981: 336). Approximately 80% of all school-age children with learning disabilities have language impairments that substantially characterize their disorder (Wiig and Secord, 1998). Language and reading problems in children with LBLD result largely from impairments in vocabulary and comprehension (McCormick and Becker, 1996; Minskoff, 1982; Wright and Newhoff, 2001). Students with LBLD often show average to superior intelligence, but have difficulty in one or more aspects of language, such as spelling, reading, and/or writing. For instance, a child with dyslexia has trouble almost exclusively with reading, whereas a child with an expressive language disorder has

significant difficulty communicating their thoughts although they understand what has been said. These students also have difficulty with figurative (nonliteral) language, such as idioms and metaphors (Qualls *et al.*, 2004).

Children from Linguistically Diverse Backgrounds

Children from culturally and linguistically diverse (CLD) backgrounds are placed in special education at a much higher rate than their white counterparts. Majority of these children are referred for services related to speech, language, and/or reading disorders. African Americans represent more than 30% of the students in special education, yet they make up only 17% of the total school-age population (Obiakor, 2007; Rochon, 2007). Native Americans represent less than 1% of the school population and 1% of the special education population. Asian Americans represent less than 2% of all students in special education, while they make up 4% of the school-age population. White students are the most proportionate of all of the groups relative to the overall school-age population (67%), with placement in special education at 43.2% (National Center for Education Statistics, 2005; Obiakor, 2007). Traditionally, Latino American students have been overrepresented in all categories of special education, particularly SLD and learning disabilities; they are 16% of the national school population and 14% of the children in special education.

African-American English Dialect Speakers

Language is essential in the transmission of culture (Gollnick and Chinn, 2002; Hughes *et al.*, 2002). Through the process of enculturation, children learn their first (home) language and culture, which often differs from learning a second (school) language and culture. When methods for facilitating the learning of the school language do not take into consideration the child's home language/culture as a basis, as well as his/her interests and capabilities (Gay, 2002; Rickford, 2001), the psychological and social consequences can be profound because they can influence how they learn and interact with their peers (Saville-Troike, 1976).

Historically, African-American students' have been overidentified for placement in specific categories of special education, such as speech, language, and reading disorders (National Reading Center, 2002; Obiakor, 2007). This may be due, in part, to the dialect spoken by many African-American students, African-American English (AAE). Graham (1997) noted that AAE speakers comprise a significant percentage of US public school students, many of whom are placed in special education at a disproportionate rate (68% higher) compared to their white peers (National Center for

Education Statistics, 2005). AAE has its own system of sound, word and sentence structure, meaning, structural organization of vocabulary items, and lexical, phonological, and syntactic as well as semantic patterns intertwined with structures in Standard American English (SAE; Green, 2002). SLPs and special education teachers who have a thorough understanding of the distinctions between a language difference (vis-a'-vis, a dialect such as AAE) and a language disorder are key in reducing the numbers of inappropriate placement of AAE speakers in speech and language therapy. In particular, these professionals must have an understanding of the ways in which AAE differs systematically from SAE. Qualls (2007) stated that "speech-language pathologists, educators, and neuropsychologists who are responsible for testing and assessing the speech, language, and cognitive abilities of African Americans should have adequate knowledge about AAE. This is crucial for differentiating one's dialect from disordered speech" (p. 133). These professionals will also reduce the chance that AAE speakers, who are in need of speech and language services, will not be overlooked. SLPs and special education teachers must have (1) the ability to defer their own biases, (2) an appreciation for the range of linguistic and cultural variations in their students, and (3) the sensitivity to any inherent bias in standardized assessments. In their small sample study of 88 self-selected informants from five secondary schools in New York City, Blake and Culter (2003) reported that the majority of the teachers did not support bidialectal education and felt that federal funds should not be allocated for such programs (this is in contrast to their overwhelming support for bilingual education).

Bilingual Students

In 1974, California gave us bilingual education. The United States Supreme Court ruled that California had to provide some type of special educational assistance for linguistically diverse students. The original objective of bilingual education was to ensure students would not fall behind academically because of a poor command of English and to gradually teach them English as a second language. The Federal Bilingual Education Act (1968 and 1974) mandates and provides Title VII funding for local educational agencies to establish, implement, and sustain programs for the transition of children and youth of limited English proficiency (LEP; US Department of Education, 2007). Students from linguistically diverse backgrounds are a large percentage of school-age children in the United States and in California. Over the last 25 years, students from linguistically diverse backgrounds have increased rapidly comparative to mono-English speakers in California and in the rest of the US. In 2005 in the United States, more than 10 million children, between the ages of 5 and 17, spoke a language other than English. This population of students represented 20% of school-age children (US Census, 2005).

For example, in California, 44% or 3.9 million school-age children spoke a language other than English. This population increased by 187%, while mono-English speakers increased by only 8%. Similarly, national statistics for linguistically diverse students increased by 130%, while mono-English speakers decreased by 1.3 %. As a result, the majority of the 5 million additional school age children in the United States over the last 25 years were from linguistically diverse backgrounds (University of California Linguistic Minority Research Institute, 2005).

Disturbing numbers of these bilingual students underachieve in US public schools (Garcia and Domínguez, 1997) and are placed into special education programs. Largely due to language differences, English-language learners continue to be overidentified in special education. In an effort to compensate for such overidentification, there is a new trend of underidentification that has emerged for this population (Ovando and Collier, 1985). Because of the heightened awareness of cultural and linguistic diversity for SLPs and special education teachers, there seems to be more sensitivity when assessing speech and language skills of linguistically diverse students. However, due to personnel shortages, there still appears to be few trained skilled bilingual SLPs and certified bilingual special educators. Consequently mono-lingual SLPs and special educators are engaged in discussions that may not be appropriate for bilingual students. This may be due to the alternative placements of many English-language learners with disabilities in bilingual education instead of special education services (Baca and Cervantes, 1989, 1991). To ensure that this population of students receives a more appropriate placement, many districts implemented bilingual special education programs. Apparently, bilingual special education utilizes an instructional social system that involves active teaching of cognitive skills that takes into account the student's disabilities (e.g., an SLD), developing language skills at the same time as focusing on the acquisition of English (Baca and Cervantes, 1991). Regarding bilingual students with SLD, one assumption is that students learn best in their preferred language. It is therefore imperative that the SLP understands the most appropriate approach to assessment and treatment of the child and his/her family through interpreters, translators, and cultural brokers (ASHA, 2004). The goal is to provide all students with educational experiences that allow them to develop lifelong learning skills (Baca and Payon, 1989).

The Individuals with Disabilities Education Act and Students with SLD

The Individuals with Disabilities Education Act (IDEA) is the federal special education law (IDEA, 1997), that is an amendment to the 1975 Education of All Handicapped

Children's Act (PL 94-142). IDEA provides federal financial assistance to state and local education agencies to guarantee special education and related services to eligible children and youth ages 3 through 21. The IDEA '97 was reauthorized as The Individuals with Disabilities Education Improvement Act (IDEIA) of 2004. This reauthorization provides some guidance in planning and implementing assessment procedures for all students, including CLD students who may or may not have a disability. IDEA stipulates that testing and evaluation procedures should be nondiscriminatory and requires that all children be fairly assessed in their dominant (native) language (US Department of Education, 2007). It is important to note that existing special education referral practices consider primarily the culture and language of mainstream English speakers and do not take into account nonstandard dialect speakers and/or non-English speakers.

A large proportion of nonmainstream students have LEP or are speakers of nonstandard English dialects. These students may not fully benefit from classroom English instruction. Further, some classroom teachers may not be able to distinguish SLD from linguistic differences (Garcia and Domínguez, 1997).

It is well known that standardized assessments are used to determine placement or nonplacement into special education. It is also known that many, if not all, standardized assessments show some level of bias in favor of the culture, language, and cognitive learning styles of speakers of mainstream English. Hence, these assessments will disadvantage anyone who does not possess those characteristics of the mainstream population. For example, speakers of AAE perform more poorly on these assessments; besides, students with cognitive and learning styles that are field dependent and associative have similar outcomes. However, while it is not known to what extent AAE directly affects the placement of African Americans in special education for SLD, researchers assert that AAE among a multitude of variables impacts the attitudes, expectations, and perceptions that often lead to special education referral for SLD and placement (Anderson, 1992; Hilliard, 1992). In an effort to alleviate the number of non-mainstream English speakers' overidentification in the category of SLD, the Diagnostic Evaluation of Language Variance (DELV) was developed. The DELV is a norm-referenced test designed to assess children's language abilities by not penalizing those whose language may differ from mainstream English on test performance. The DELV is responsive to cultural and linguistic differences of many African-American children and is used frequently to assess this population. However, it is an effective assessment tool for children of all races and ethnicities because the DELV examines language structures common to different varieties of English (Harcourt Assessment, 2007).

Responding to Individualized Education Plans

According to law, an individualized education plan (IEP) is required for all students who have been identified as requiring special education services, including those receiving services for SLD. A multidisciplinary team (e.g., students, general education teachers, special education teachers, guardians, counselors, principals, social workers, and SLPs) determines students' eligibility in one or more of the 13 disability categories and the need for special education and/or related services (Figure 1).

Clearly, an IEP assists the school in meeting the student's learning needs as well as helps plan educational goals for him/her. Where a student has been diagnosed as having a speech and/or language disorder, the SLP would lead the IEP team in setting objectives and educational goals for him/her and help to determine any adaptations on approaches to services (e.g., where the student would receive his/her therapy and whether it is in the classroom with the other students or individually). The teacher, the speech pathologist, and parents/guardians must work collaboratively to provide necessary services for the student to successfully meet their annual IEP goals. Parents are considered as equal partners to school professionals. In addition, they provide pertinent information to the team about the child's development, including social and medical history, assist with ensuring he/she is practicing strategies at home, and report on his/her progress from the perspective of the home environment.

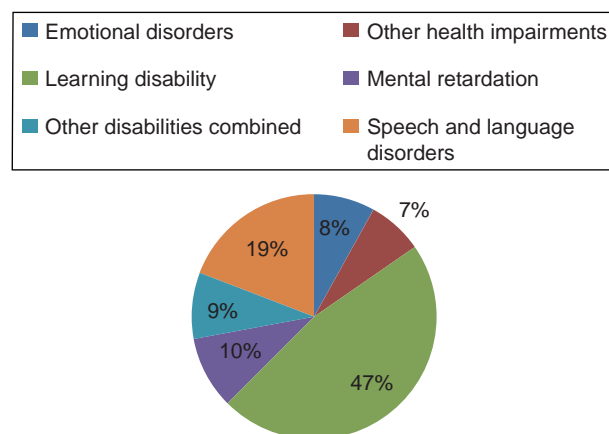


Figure 1 Percentage of students in special education by disability category. Other disabilities include multiple disabilities (2.2 %), hearing impairments (1.2 %), orthopedic impairments (1.1%), visual impairments (0.4 %), autism (2.3%), deaf-blindness (0.03%), traumatic brain injury (0.4%), and developmental delay (1.1%). From US Department of Education, Office of Special Education Programs, Data Analysis System (DANS), OMB #1820-0043: Children with Disabilities Receiving Special Education Under Part B of the Individuals with Disabilities Education Act, 2003. Data updated as of 31 July 2004. Also tables 1–3 in vol. 2 of this report. These data are for the 50 states, District of Columbia, BIA schools, Puerto Rico and the outlying areas.

Future Perspectives

School-based SLPs and special education teachers provide invaluable services to children with SLD. IDEA, in conjunction with the school-based curriculum, guides legal and ethical decisions regarding assessment for and assignment into the different categories of special education that include SLD. SLPs are highly trained in the acquisition and development of speech and language; they are supposed to know the nature, causes, and best practices of SLD. In addition, they work with classroom teachers, special educators, social workers, psychologists, interpreters, cultural brokers, and families to ensure that children with SLDs from a range of social, linguistic, and cultural backgrounds can be academically and socially successful.

There are a number of future trends and needs that require the attention of professionals working with school-age children with SLD, three of which are summarized as follows: testing/assessment; technology, and evidence-based practice.

Testing/Assessment

Evidence is incomplete in the area of predicting future functioning:

- large-scale studies are needed to increase the predictive validity of instruments used to assess SLD in children;
- normative samples should be representative of the populations of interest and of sufficient size to demonstrate the instrument's ability for valid, reliable interpretation;
- appropriate instruments are needed for reliable and valid assessment of language in population subgroups, defined in terms of language, dialect, and cultural differences (e.g., English/Spanish proficiency scales in oral and written language, and acculturation and assimilation scales);
- resources are needed to encourage research among teams of professionals with expertise in SLD, cultural experts/brokers, professionals with expertise in disorders that often co-occur with speech and language impairment, and psychometric experts;
- research is needed relative to the applicability of speech and language assessment instruments for individuals with different disorders such as autism, severe physical impairment, mental retardation, CAS, learning disabilities, and hearing impairment.

Technology

Technology is very important in working with students with speech and language problems. Efforts must be made to:

- Assess the efficacy of using technology in the assessment and treatment of children with SLD. There must be research on the efficacy of incorporating technological applications such as, personal digital assistant (PDA) texting, and blogging. It is important to understand the impact of attitudes, perspectives, and beliefs relative to technology.
- Expand the classroom teachers' knowledge base about technology transfer for those children with SLD.
- Refine the use of augmentative-alternative (AAC) devices for children with SLD, including funding, training and education, use in the classroom/at home, and security.

Evidence-Based Practice

As a result of personnel shortages, many SLPs working in the schools are seeing larger numbers of children with SLD. They are consequently spending less time per student per therapy session and feeling the pressure of increased accountability (see IDEA and No Child Left Behind Act of 2001) and consumer demand for high-quality services. Further, rising costs to provide services to children with SLD have led to the need for school-based SLPs to present evidence that their services are efficacious, effective, and efficient.

Evidence-based practice represents the nexus between the best available evidence, clinical expertise, and client/patient perspectives. There is a major movement in the field of communication sciences and disorders to define, understand, adopt, and employ evidence-based practice. However, to fully implement this practice requires a paradigm shift for how clinical decisions are made. To date, the complexities of this practice are yet to be fully understood, although significant resources are being devoted to this important approach to assessing and treating individuals with communications disorders (ASHA, 2008).

Johnson (2006) noted that much of the current research in speech-language pathology is that of basic research, addressing "theoretical or etiological questions and [employing] designs and methods that do not permit direct applications to clinical issues" (p. 21). She further states "to inform practice, research must address clinically relevant questions using particular designs and methods" (p. 21). On the other hand, Brackenbury *et al.* (2008) argue that the particular designs and methods necessary for evidence-based practice may be problematic for a child's language because most treatment approaches for language disorders do not work quickly. Consequently, the child's life circumstances may influence treatment outcomes (e.g., general development). A second conundrum for treatment research pointed out by Brackenbury *et al.* (2008) deals with the trade-off between external validity and internal validity. Increased controls increase internal validity, but decreases external validity, the penultimate goal for intervention research – real-world applicability.

Conclusion

The numbers of school-age children with SLD in the US and Canada have continued to rise and many of these children are from CLD backgrounds. CLD students are placed in special education at a much higher rate than their white counterparts and the majority of these children are referred for services related to speech, language, and/or reading disorders. For example, English-language learners were traditionally overidentified in special education, but are presently underidentified (Ovando and Collier, 1985). It is imperative that SLPs and special educators have a heightened awareness and adequate knowledge of what cultural and linguistic diversity entails. Moreover, they must employ culturally appropriate practices when identifying, referring, assessing, and treating the speech and language skills of CLD students. School-based SLPs and special education teachers are responsible for providing services to children with SLD that are curriculum based, evidence based, and that are guided by legal and ethical tenets that include consideration of the student's cultural and linguistic experiences. Future research should inform, more specifically, best practices for working with CLD children with SLDs and their families. Further, as the change in demographics of school-age children continues to evolve in the US, educational professionals will need to participate in professional development activities that will enable them to more fully engage in culturally competent practices that ensure the success of their students.

Bibliography

- Anderson, M. G. (1992). The use of selected theatre rehearsal technique activities with African American adolescents labeled behavior disordered. *Exceptional Children* **59**, 132–140.
- Baca, L. M. and Cervantes, H. T. (eds.) (1989). *The Bilingual Special Education Interface*, 2nd edn. Columbus, OH: Merrill.
- Baca, L. M. and Cervantes, H. T. (1991). Bilingual special education. *Learning Disabilities Online* (document reproduced from ERIC Digest No. E496). http://www.ldonline.org/ld_indepth/bilingual_ld/esl_id_eric.html (accessed August 2009).
- Baca, L. M. and Payon, R. M. (1989). Development of the bilingual special education interface. In Baca, L. M. and Cervantes, H. T. (eds.) *The Bilingual Special Education Interface*, pp 79–99. Columbus, OH: Merrill.
- Blood, G. W., Ridenour, V. J., Qualls, C. D., and Hammer, C. S. (2003). Co-occurring speech, language and non speech-language problems in children who stutter. *Journal of Communication Disorders* **36**(6), 427–448.
- Brackenbury, T., Burroughs, E., and Hewitt, L. E. (2008). A qualitative examination of current guidelines for evidence-based practice in child language intervention. *Language, Speech, and Hearing Services in Schools* **39**, 78–88.
- Garcia, S. and Domínguez, L. (1997). Cultural contexts that influence learning and academic performance. *Child and Adolescent Psychiatric Clinics of North America* **6**, 621–655.
- Gay, G. (2002). Culturally responsive teaching in special education for ethnically diverse students: Setting the stage. *Qualitative Studies in Education* **15**(6), 613–629.

- Gollnick, D. M. and Chinn, P. C. (2002). *Multicultural Education in a Pluralistic Society*. Washington, DC: Merrill-Prentice Hall.
- Hilliard, A. G. (1992). The pitfalls and promise of special education practice. *Exceptional Children* **59**, 168–171.
- Hooper, C. R. (2004). Treatment of voice disorders in children. *Language, Speech, and Hearing Services in Schools* **35**, 320–326.
- Hughes, M., Kroehler, C. J., and Vander Zanden, J. W. (2002). *Sociology: The Core*, 6th edn. New York: McGraw-Hill.
- Johnson, C. J. (2006). Getting started in evidence-based practice for childhood speech language disorders. *American Journal Speech Language Pathology* **15**, 20–35.
- National Center for Education Statistics (2005). National assessment of educational progress (NAEP): 1992–2000. *Reading*. . <http://nces.ed.gov> (accessed August 2009).
- Obiakor, F. E. (2007). *Multicultural Special Education: Culturally Responsive Teaching*. Upper Saddle River, NJ: Pearson Merrill/Prentice-Hall.
- Ovando, C. J. and Collier, V. (1985). *Bilingual and ESL Classrooms: Teaching in Multicultural Contexts*. New York: McGraw-Hill.
- Qualls, C. D. (2007). Speech, language, and neuropsychological assessment: Implications for African Americans. In Uzzell, B. P., Ponton, M., and Ardila, A. (eds.) *International Handbook of Cross-Cultural Neuropsychology*, pp 127. New York: Routledge.
- Rickford, J. R. (2001). Sociolinguistics and the public: Digging and being bug in return. *American Speech* **75**(3), 273–275.
- Saville-Troike, M. (1976). *Foundations for Teaching English as a Second Language: Theory and Method for Multicultural education*. Newark, NJ: Prentice-Hall.
- US Department of Education (2001). *Twenty-Sixth Annual Report to Congress on the Implementation of Individuals with Disabilities Act*. Washington, DC: US Government Printing Office.
- report/technology assessment No. 52. Washington, DC: US Department of Health and Human Services, Public Health Service. <http://www.ahrq.gov> (accessed August 2009).
- ASHA (American Speech-Language-Hearing Association) Ad Hoc Committee on Service Delivery in the Schools (1993). Definitions of communication disorders and variations. *ASHA* **35**(supplement 10), 40–41.
- Boone, D. R., McFarlane, S. C., and Von Berg, S. L. (2005). *The Voice and Voice Therapy*, 7th edn. Boston, MA: Pearson Education.
- Fey, M. E. (2006). Commentary on “Making evidence-based decisions about child language intervention in schools” by Gillam and Gillam. *Language, Speech, and Hearing Services in Schools* **37**, 316–319.
- Gascoigne, M. (2006). Supporting children with speech, language and communication needs within integrated children’s services. *Royal College of Speech and Language Therapists Position Paper*. RCSLT, London. <http://www.rcslt.org> (accessed august 2009).
- Gillam, S. L. and Gillam, R. B. (2006). Making evidence-based decisions about child language intervention in schools. *Language, Speech, and Hearing Services in Schools* **37**, 304–315.

Relevant Websites

- <http://www.asha.org> – American Speech-Language-Hearing Association, Speech and Language Disorders and Diseases; Frequently Asked Questions: Speech and Language Disorders in the School Setting; Typical Speech and Language Development; What Is Language? What Is Speech? NJC, Communication Services and Supports for Individuals with Severe Disabilities: FAQs.
- <http://www.caslpa.ca> – Canadian Association of Speech-Language Pathologists and Audiologists, Fact Sheets.
- <http://www.kidsource.com> – Kid Source, Info about Speech and Language Disorders.
- <http://nichcy.org> – National Dissemination Center for Children with Disabilities, Speech and Language Impairments.
- <http://www.speechpathologyaustralia.org.au> – Speech Pathology Australia, Fact Sheets No. 24, 41 and 23.

Further Reading

- Agency for Healthcare Research and Quality (2002). Criteria for determining disability in speech-language disorders: Evidence

Educating Students with Autism and Related Disorders

T T Dyches, Brigham Young University, Provo, UT, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Asperger syndrome – It is a developmental disorder which significantly impacts the qualitative nature of social interaction, along with a restricted and repetitive repertoire of behavior, interests, and activities, with no general delay in language or cognitive development.

Autism – It is a developmental disorder which significantly impacts the qualitative nature of social interaction and communication, along with a restricted and repetitive repertoire of behavior, interests, and activities.

Pervasive developmental disorder – It is a general diagnostic umbrella term under which five disorders are present: autism, Rett syndrome, childhood disintegrative disorder, Asperger disorder, and pervasive development disorder – not otherwise specified.

Pervasive developmental disorder – not otherwise specified – It is a developmental disorder in which severe and pervasive impairments exist in social interaction or communication skills, or when restricted behaviors, interests, and activities are present, but criteria are not met for other PDDs.

Historical Perspective of Autism Spectrum Disorders

The term autism, taken from the Greek *autos* meaning self, was used in the early 1900s by Swiss psychiatrist Eugen Bleuler to describe idiosyncratic behavioral symptoms of patients with schizophrenia who exhibited fundamental disturbances in contact with the social world. However, it was not until 1943 that autism became recognized as a distinctive condition when Leo Kanner, an Austrian-American and John Hopkins University psychiatrist, published his seminal article *Autistic Disturbances of Affective Contact*. One year later, Austrian pediatrician Hans Asperger published his pioneering paper *Autistic Psychopathy in Children*, in which he described children with severe difficulties with social integration. He chose the label autism to characterize these children; however, the children he was describing clearly differed from those described by Kanner. Asperger's work was not widely

known until after his paper was translated from German to English in 1981 by Lorna Wing, and English psychiatrist who suggested Asperger syndrome to be a separate entity from autism, with less-severe symptoms and much better prognosis.

Prior to the identification of autism as a distinct disorder, children with autism were diagnosed as having mental retardation, childhood schizophrenia, or other psychiatric disorders. Accounts of children with characteristics now accepted as being associated with autism have been reported for hundreds of years, the most famous being Victor of Aveyron. Victor was found in 1797 and originally assumed to have been raised in the wild. In spite of the education and training efforts of Jean-Marc Gaspard Itard, Victor did not learn to speak, yet he did gain the ability to understand spoken language and to conform to some social norms. Scientific study in the 1950s and 1960s investigated possible psychopathological roots of autism. Child psychologist Bruno Bettelheim blamed 'cold, refrigerator mothers' for causing their children's autism and suggested removing these children from their mothers. This psychogenic view was in direct contrast to Kanner's description of inborn autistic disturbances. In the 1960s, strong evidence against the psychogenic theory of autism changed the focus of etiological study. In his book, *Infantile Autism* (1964), Rimland's denouncement of psychogenesis and support for biological etiology led to the widely accepted view that family interaction patterns do not contribute to the occurrence of autism. While this claim discounts psychogenic etiologies, it still neglects the genetic liability that is gaining widespread empirical support, as autism is considered to be a highly heritable condition. Also, beginning in the 1950s, research began to focus on the education and treatment of individuals with autism, with an emphasis on psychotherapy. However, in the 1960s when techniques based on the principles of applied behavior analysis were used successfully to teach children with autism, the educational landscape changed, and more children were educated in research studies and in schools. With the passage of the Education for All Handicapped Children Education Act in the USA in 1975, children with autism were guaranteed a free and appropriate public education. In 1990, this law was reauthorized as the Individuals with Disabilities Education Act (IDEA), and autism was added as a specific educational classification. Since then, the number of students classified under this category has increased dramatically. Similar laws govern the education of students with autism in many countries; however, in a

large number of places throughout the world, individuals with autism are not guaranteed a free and appropriate public education or other publicly funded services.

Characteristics of Autism Spectrum Disorders

The *Diagnostic and Statistical Manual of Mental Disorders IV-TR* provides diagnostic criteria for pervasive developmental disorders (PDDs), which are similar to those listed in the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10). PDDs are pervasive, meaning they are persistent and all-encompassing, affect normal development, and have disabling effects on the lives of individuals. While there is some dispute regarding symptom boundaries for these disorders, the most common standard is to view five distinct and related disorders under the umbrella of PDD (see **Figure 1**):

- Autistic disorder (known also as autism)
- Asperger disorder (Asperger syndrome)
- Rett's disorder (Rett syndrome)
- Childhood disintegrative disorder

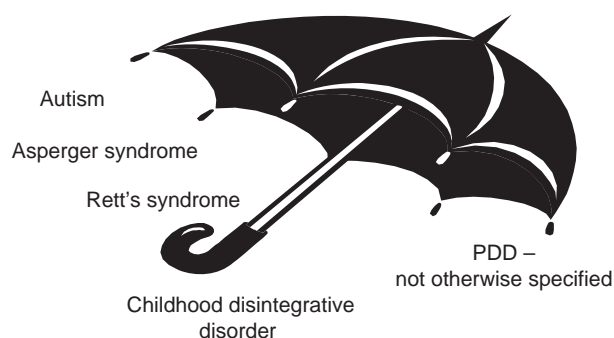


Figure 1 Pervasive developmental disorders.

- Pervasive developmental disorder – not otherwise specified (PDD-NOS; autistic-like, autistic tendencies, or atypical autism).

Autism spectrum disorders (ASDs) is a term which is generally accepted to include autism, Asperger syndrome, and PDD-NOS. These disorders have similar characteristics but vary in their expression of a triad of symptoms: social skill deficits, communication disorders, and restricted behaviors, interests, and activities. However, the spectrum of autism can also be manifested in various intellectual abilities, sensory sensitivity, and motor skills (see **Figure 2**).

Autism

Autism is a developmental disorder with neurological origins, present from birth or early development. Considered to be a lifelong disorder, autism impacts an individual's social interactions, communication skills, and behavior patterns. Delays and/or abnormal functioning are apparent by the age of 3 years, particularly in social interaction, social language, or symbolic or imaginative play. Autism is manifested along a broad spectrum, meaning the skills and deficits vary considerably among affected individuals. For example, while earlier research reported that approximately 70% of individuals with autism had intellectual disabilities, more recent epidemiological studies indicate rates as low as 40%. Social symptoms may vary from social quirkiness to severe self-absorption. Some children with autism are highly verbal, while approximately 50–75% have little or no functional speech. Also, some have mildly odd patterns of behavior or interests, while many have persistent destructive and disruptive behavioral challenges.

While neurological dysfunction is not a diagnostic marker, it is a common characteristic of autism. Many individuals with autism have abnormalities in chromosomes,

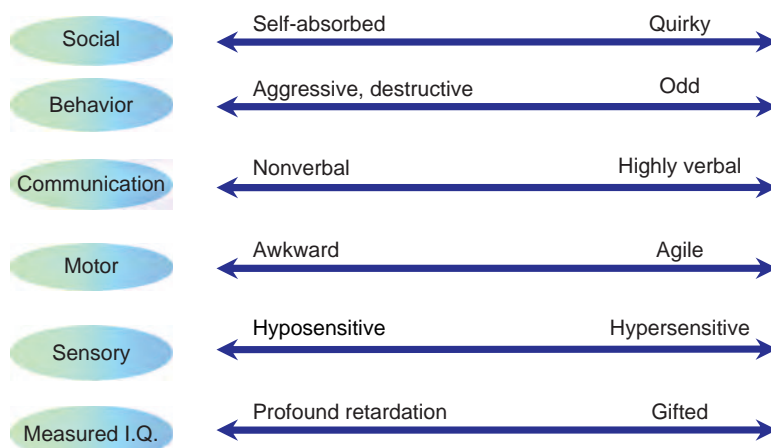


Figure 2 Continuum of symptoms in autism spectrum disorders.

brain structure, and brain chemistry. Approximately 25% of individuals with autism develop epilepsy, often during adolescence. Common symptoms of autism include a triad of impairments:

1. Social skills
 - Delayed development of appropriate peer relationships
 - Impaired spontaneous sharing (e.g., interests and achievements)
 - Lack of social/emotional reciprocity
 - Impaired nonverbal behavior (e.g., eye contact and facial expressions)
2. Behavioral repertoire
 - Inflexible, nonfunctional routines or rituals
 - Repetitive motor actions (e.g., hand flapping and body rocking)
 - Persistent preoccupation and restricted, intense patterns of interest
 - Abnormal mood/affect
 - Hyperactivity
 - Odd responses to sensory experiences
3. Communication
 - Delay or lack of spoken language
 - Conversational impairment
 - Repetitive and stereotyped language
 - Poor receptive language
 - Impaired play skills.

Asperger Syndrome

Individuals with Asperger syndrome have similar symptoms of autism in the areas of social interaction and behavior patterns; however, they show no clinically significant delays in language. According to diagnostic criteria, children with Asperger syndrome use single words by the age of 2 years and communicative phrases by the age 3 years. Further, children with Asperger syndrome have no clinically significant delay in the development of cognitive processes, self-help skills, adaptive behavior, or curiosity about the environment.

For an individual to be diagnosed with Asperger syndrome, the symptoms must significantly impair him or her in social, occupational, or other important areas of functioning. Associated features include mild motor clumsiness, hyperactivity, inattention, and other mental disorders such as depression. Individuals with Asperger syndrome may be overlooked because their high verbal skills may mask the severity of their social dysfunction, and their unusual behaviors may be interpreted as stubbornness. Prognosis for individuals with Asperger syndrome is generally much better than for those with autism. Many are able to complete college degrees, obtain meaningful employment, marry, and interact as contributing members of their communities. Some individuals with Asperger syndrome prefer to be

considered different rather than disabled, rejecting the notion that Asperger syndrome is a condition that needs to be remediated or cured.

Pervasive Developmental Disorder–Not Otherwise Specified

PDD-NOS is a diagnosis given by a physician when an individual exhibits a significant and pervasive impairment in social development, communication, or behavior, but does not meet the criteria for autism, Asperger Syndrome, or other pervasive developmental disorders. Failure to meet criteria includes a later age of apparent onset, atypical symptoms, or symptoms that do not significantly impair the individual quantitatively or qualitatively. A diagnosis of PDD-NOS is also referred to as atypical autism.

Screening and Diagnosis of Autism Spectrum Disorders

Early detection of autism is critical for obtaining early intervention services, yet ASDs are typically not diagnosed until after 3 years of age. However, physicians and parents of children as young as 2 years old should be alerted to the following developmental red flags: poor eye contact, diminished responsive smiling, low occurrence of babbling, reduced social responsiveness; and delayed or disordered development of language, play skills, and ability to initiate or sustain social interaction. At later ages, children who are suspected of having autism exhibit symptoms in the three areas of concern: communication, social interaction, and behavior. Often, when parents are concerned about their child's development, a pediatrician may conduct a screening evaluation. If warranted, the pediatrician refers the child to a child psychologist, child psychiatrist, pediatric neurologist, or other diagnostician. A full medical evaluation is conducted, during which information is gathered about the child's developmental history, along with many evaluations and diagnostic tests in areas such as medical and neurologic conditions, communication skills, cognitive development, adaptive behavior, academic abilities, behavior patterns, and fine and gross motor skills. Further, the family's strengths, resources, and needs are evaluated. This information is used not only to aid diagnosis of a possible disorder, but to guide recommendations for services that are responsive to values and cultures of the family.

Generally, children who have been diagnosed with an ASD are also evaluated by school professionals – not for diagnostic purposes, but for educational classification. Prior to 1990, special education law did not include an educational classification for autism. Children with these diagnoses were often educated under a classification of

mental retardation, severe emotional disturbance, or multiple disabilities. While school multidisciplinary teams determine how to best serve individual students with autism, most students with this diagnosis receive special education services under the autism classification; however, students with Asperger syndrome or PDD-NOS may be educated under the label of other health impaired or another classification which the team determines to be appropriate. If their disability does not warrant special education services because it does not significantly impact their progress in the general curriculum, these students are often served under Section 504 of the Rehabilitation Act of 1973. Students who qualify under Section 504 are served by general educators, who make adaptations and accommodations according to the plan designed by a multidisciplinary team.

Etiology of Autism Spectrum Disorders

Since Leo Kanner first brought the disorder to public attention in 1943, professionals have debated the extent to which autism is a biological condition or the result of environmental or psychosocial factors. There is agreement that autism is a complex, brain-based developmental disorder in which multiple areas of functioning are affected. The specific causes of ASD remain unknown. However, it is widely accepted that there is no single designated cause. Recent clinical and biological research has demonstrated that autism is a neurological disorder which may begin before birth. The strongest evidence to date supports a genetic component. However, investigations into environmental toxins, immune system dysfunction, and other biological factors are being conducted. Autism is not caused by poor parenting, failure of a mother to bond with her child, or a child's willful retreat into his own world. To a large measure, the genetic role in autism has been evident for decades, with the well-documented overrepresentation in males and the evidence since the 1950s of higher concordance rates in monozygotic than dizygotic twins. Further, family studies indicate higher rates of autism among siblings than expected, as well as higher rates of the broader autistic phenotype (e.g., social difficulties and depression) among first-degree relatives of individuals with autism. While there are studies which suggest a relationship between autism and a few select chromosomes, it is clear that genes are an influential but not the sole factor in the development of autism.

Prevalence

Precise prevalence rates for ASD are difficult to obtain due to the variation in the ways epidemiological research is conducted and data are analyzed. However, current research reports a range from 0.7 to 72.6 cases of autism

per 10 000, with a median rate of 13 per 10 000. Prevalence estimates have increased significantly since 1987. Epidemiological studies regarding the rate of PDD-NOS are difficult to interpret due to inconsistent labeling across studies; however, current evidence suggests a median prevalence estimate of 20.8 per 10 000, ranging from 1.9 to 36.1. Data regarding the prevalence of Asperger syndrome are sparse and imprecise, ranging from .3 to 48.4 per 10 000. Best estimates of the prevalence of Asperger syndrome is 2.6 cases per 10 000. The combined global estimate of prevalence for autism spectrum disorders is at least 36.4 per 10 000, yet when most recent prevalence surveys are isolated, this rate increases to 60 per 10 000. Clearly, proper epidemiological surveys are warranted to determine accurate prevalence estimates for each diagnostic category. The dramatic increase in prevalence estimates from the early estimates of four to five cases of autism per 10 000 reflects the impact of changes in diagnostic criteria for autism and other PDDs, improvements in societal awareness, and increases in service availability. No consensus has been reached in the professional community of whether there is a true increase in prevalence.

Treatment/Intervention Approaches for Individuals with Autism Spectrum Disorders

There are innumerable treatment/intervention approaches for individuals with autism spectrum disorders, and these can be educational, biomedical, therapeutic, or family centered. Individual approaches are varied and some are experimental, resulting in different outcomes for children with ASD. Some treatments/interventions claim to cure the child, while most are designed to reduce the effects of the disorder. The most common treatments/interventions include the following:

1. Educational approaches
 - applied behavioral analysis: breaking tasks into small, manageable parts and reinforcing success
 - structured teaching: providing visual and structural support to prevent failure
 - relationship development: developing a positive and nurturing relationship with the child with autism
 - social skills training: teaching appropriate social interactions by modeling, scripting, role-playing, and using personalized stories
2. Biomedical approaches
 - nutritional/dietary
 - wheat/milk-free diets
 - vitamin, herb, and mineral supplements
 - psychopharmacology (medication)
3. Therapeutic approaches
 - speech/language

1. speech training
2. augmentative and alternative communication
3. sign language
4. picture-based systems
5. computerized communication devices
- occupational
- Physical
- sensory integration
 1. visual
 2. tactile
 3. auditory
 4. relaxation techniques
4. Family support approaches
 - parent training
 - respite care
 - support groups.

The recommendations made by the National Research Council in 2001 along with meta-analytic research and syntheses indicate that several components of educational programs are supported by empirical research. Although various education approaches may be grounded in contradictory theories of child development, if the primary components of effective programs are present, it is likely that successful outcomes for children with autism will occur. The following recommendations for young children with autism are among those based on empirical evidence (See **Figure 3** for an evaluation tool which incorporates these recommendations):

- Early intervention into an educational program.
- Active and intense educational programming, up to 5 h per day for 5 days per week for a full year.

Evaluation of educational programs for students with autism

School/program: _____ Teacher: _____

Evaluator's name: _____ Date: _____

Early entry into the intervention program	Average age at which students enter this program:		
	Average age at which students begin receiving services:		
Individualized supports and services for students and families, based upon family and student preference	Support for students:		
	Support for families:		
Active engagement in effective instructional programming	0–2 Hours/day	Half-day	Full school day
	Various sites:		
	Days per week: 1	2	3 4 5
	Less than 9 months	9–10 months	Full year
Systematic instruction based upon valid educational goals, effective delivery of planned instruction (using principles of applied behavior analysis), and frequent evaluation and adjustment of instruction	Educationally and socially valid goals:		
	Type of instruction:		
	How learning is evaluated:		
	How often learning is evaluated:		
Comprehensible and/or structured environments , where students and staff can understand what is happening, predict what will happen, and know what is required of various settings	Curriculum is clear to students: Yes Sometimes No		
	Curriculum is clear to all personnel: Yes Sometimes No		
	How is curriculum communicated to students and staff?		

Figure 3 (Continued)

Specialized curriculum content that addresses core deficits and increases functional skills that can generalize across environments	Do IEPs include goals for core deficits of autism? Social: Yes No Behavioral: Yes No Functional Communication: Yes No Cognitive/Academic: Yes No How frequently are these goals addressed? Daily Weekly Monthly Are skills generalized across environments? Yes No
Adult attention sufficient to meet IEP goals	1:1 1:2–3 1:4–10 Large groups Are goals being met: Yes No
Functional approach to problem behaviors (i.e., positive behavior support), with a focus on replacing rather than merely eliminating problem behaviors, through teaching appropriate replacement behaviors	Evidence of functional assessment of behavior: Evidence of functional analysis of behavior: Evidence of appropriate behavior support plan:
Family involvement	Evidence of family involvement in aspects of educational program: Pre-referral process

Figure 3 Evaluation of educational programs for students with autism. Adapted from Dyches, T. T. (2008). Evaluating your educational program for students with autism. *The Utah Special Educator* 28(3), 80–81.

- Planned teaching opportunities organized around brief periods of time.
- Adult attention either individually or in small groups to facilitate mastery of educational goals.
- Multicomponent interventions that target individualized supports and services in the core deficit areas of social–emotional skills, cognition, language, and behavior.
- Comprehensible and/or structured environments.
- Intensive behavioral interventions using positive behavior support principles.
- Family involvement.

Clearly, essential components of effective programs for children and adolescents with autism include (1) emphasis on functional activities and skills needed to be successful in the real world, (2) activities that are appropriate for the individual's chronological age, (3) instructional decisions that are data based, (4) instruction in school and non-school environments, and (5) social integration to the maximum extent appropriate.

Currently, meta-analytic research has not found sufficient evidence for treatments that claim to produce

significant outcomes for students with autism (e.g., social stories and school-based social skills intervention); and furthermore, studies have not been examined empirically or they have produced unclear results (e.g., facilitated communication, chelation therapy, dolphin therapy, auditory integration therapy, dietary interventions, and magnets). As it appears, educational treatments/interventions occur across a wide variety of settings, as determined by the child's Individual and Family Service Plan (IFSP) team (ages 0–2 years) or Individualized Education Program (IEP) team (ages 3–21). Some children and youth with ASD are served in their homes, while others are served primarily in schools, ranging from full participation in general education classes to placement in self-contained schools for students with disabilities. The IFSP/IEP is a legal document which designates annual goals for the child with autism. The individualized goals, supports, and services offered are based on student and family needs. All IEP goals must be assessed and documented regularly in order to evaluate whether a child is benefiting from a particular intervention. If the child is not showing sufficient progress, environmental conditions and teaching strategies should be

evaluated for their potential to impact the student's learning. Student progress should be monitored frequently and learning objectives adjusted accordingly. In addition, parents and practitioners are advised to evaluate their programs for children with autism based on high-quality research.

Summary

Recognition of autism as a specific disorder occurred when Leo Kanner published his landmark work describing children with autistic disturbances in their affective contact. Since that time, hundreds of thousands of individuals throughout the world have been diagnosed as having autism or related conditions such as Asperger syndrome and PDD-NOS. These three conditions are generally accepted to be ASDs; however, there is no universal consensus regarding specific boundaries for each diagnostic category. Nevertheless, these ASDs have similar core features which classify them as PDDs. Screening for ASD begins when a child is young, often before age 2 years. Parents and physicians are counseled to attend to markers which may indicate a developmental delay or disorder. Early screening generally leads to early diagnosis, which in turn leads to early intervention services.

A single specific cause for ASD is not known. It is widely accepted that there is not one single cause, but that the etiology of the disorders derives from many sources, with a strong genetic component. Current research has begun to investigate the role of environmental factors in the development of ASD. According to most recent epidemiological studies, the prevalence of autism, Asperger syndrome, and PDD-NOS combined is 60 cases per 10 000 individuals. However, prevalence rates must be interpreted with caution due to imprecise methods of collecting and reporting diagnostic data from multiple studies. Recent increase in prevalence is likely due to increased awareness, broader diagnostic criteria, and increased service availability. Many treatment/intervention approaches are available for helping individuals with ASD, including educational, biomedical, therapeutic, and family-based supports. Meta-analytic research indicates several core features that should be in place for obtaining positive outcomes for students with ASD. Finally, parents and professionals are advised to support practices which are validated by rigorous research.

Further Reading

- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders (text rev.)*, 4th edn. Washington, DC: Author.
- Baron-Cohen, S. (2008). *Autism and Asperger Syndrome*. Oxford: Oxford University Press.
- Bashe, P. R., Kirby, B. L., Baron-Cohen, S., and Attwood, T. (2005). *The OASIS Guide to Asperger Syndrome: Completely Revised and Updated: Advice, Support, Insight, and Inspiration*. New York: Crown.
- Bauman, M. L. and Kemper, T. L. (1994). *The Neurobiology of Autism*. Baltimore, MD: Johns Hopkins University Press.
- Bettelheim, B. (1972). *The Empty Fortress: Infantile Autism and the Birth of the Self*. New York: Free Press.
- Donnellan, A. (ed.) (1985). *Classic Readings in Autism*. New York: Teachers College Press.
- Doyle, B. T. and Iland, E. D. (2004). *Autism Spectrum Disorders from A to Z: Assessment, Diagnosis and More*. Arlington, TX: Future Horizons.
- Freeman, S. and Dake, L. (1997). *Teach Me Language: A Language Manual for Children with Autism, Asperger's Syndrome and Related Developmental Disorders*. Langley, BC: SKF Books.
- Frith, U. (ed.) (1991). *Autism and Asperger Syndrome*. Cambridge: Cambridge University Press.
- Frith, U. (2003). *Autism: Explaining the Enigma*, 2nd edn. Oxford: Blackwell Publishers.
- Frith, U. and Hill, E. L. (eds.) (2004). *Autism: Mind and Brain*. Oxford: Oxford University Press.
- Gerlach, E. K. (2000). *Autism Treatment Guide*, 2nd edn. Eugene, OR: Four Leaf Press.
- Gillberg, C. and Coleman, M. (2000). *The Biology of the Autistic Syndromes*, 3rd edn. London: Mac Keith Press.
- Goldstein, S., Naglieri, J. A., and Ozonoff, S. (eds.) (2009). *Assessment of Autism Spectrum Disorders*. New York: Guilford Press.
- Gray, C. and White, A. L. (eds.) (2002). *My Social Stories Book*. London: Jessica Kingsley Publishers.
- Greenspan, S., Wieder, S., and Simons, R. (1998). *The Child with Special Needs: Encouraging Intellectual and Emotional Growth*. Reading, MA: Perseus Books Group.
- Handleman, J. S. and Harris, S. L. (2006). *School-Age Education Programs for Children with Autism*. Austin, TX: Pro-Ed.
- Harris, S. L. and Glasberg, B. A. (2003). *Siblings of Children with Autism: A Guide for Families*, 2nd edn. Bethesda, MD: Woodbine House.
- Harris, S. L. and Weiss, M. J. (1998). *Right from the Start – Behavioral Intervention for Young Children with Autism: A Guide for Parents and Professionals*. Bethesda, MD: Woodbine House.
- Hodgdon, L. (2001). *Visual Strategies for Improving Communication: Practical Supports for School and Home*. Troy, MI: QuirkRoberts.
- Hodgdon, L. A. (1999). *Solving Behavior Problems in Autism: Improving Communication with Visual Strategies*. Troy, MI: QuirkRoberts.
- Houston, R. and Frith, U. (2000). *Autism in History: The Case of Hugh Blair of Borgue*. Oxford: Blackwell Publishers.
- Howlin, P. (2002). *Children with Autism and Asperger Syndrome: A Guide for Practitioners and Carers*. West Sussex: Wiley.
- Iovannone, R., Dunlap, G., Huber, H., and Kincaid, D. (2003). Effective educational practices for students with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities* **18**(3), 150–165.
- Koegel, R. L. and Koegel, L. K. (2006). *Pivotal Response Treatments for Autism: Communication, Social and Academic Development*. Baltimore, MD: Paul H. Brookes.
- Levy, S., Kim, A., and Olive, M. (2006). Interventions for young children with autism: A synthesis of the literature. *Focus on Autism and Other Developmental Disabilities* **21**(1), 55–62.
- Lovaas, O. I. (1981). *Teaching Developmentally Disabled Children: The Me Book*. Austin, TX: Pro-Ed.
- Maurice, C., Green, G., and Luce, S. (eds.) (1996). *Behavioral Intervention for Young Children with Autism: A Manual for Parents and Professionals*. Austin, TX: Pro-Ed.
- Mesibov, G., Shea, V., and Schopler, E. (2004). *The TEACH Approach to Autism Spectrum Disorders*. New York: Springer.
- Myles, B. S., Trautman, M. L., and Schelvan, R. L. (2004). *The Hidden Curriculum: Practical Solutions for Understanding Unstated Rules in Social Situations*. Shawnee Mission, KS: Autism Asperger.
- National Research Council (2001). *Educating Children with Autism*. Washington, DC: National Academy Press.
- Quill, K. (2000). *Do–Watch–Listen–Say: Social and Communication Intervention for Children with Autism*. Baltimore, MD: Paul H. Brookes.
- Rimland, B. (1964). *Infantile Autism: The Syndrome and Its Implications for a Neural Theory of Behavior*. New York: Appleton-Century-Crofts.

- Smith-Myles, B. and Simpson, R. L. (2003). *Asperger Syndrome: A Guide for Educators and Parents*, 2nd edn. Austin, TX: Pro-Ed.
- Sundberg, M. L. and Partington, J. W. (1998). *Teaching Language to Children with Autism or Other Developmental Disabilities*. Pleasant Hill, CA: Behavior Analysts.
- Volkmar, F. R. (ed.) (2007). *Autism and Pervasive Developmental Disorders*, 2nd edn. New York: Cambridge University Press.
- Volkmar, F. R., Paul, R., Klin, A., and Cohen, D. (eds.) (2005). *Handbook of Autism and Pervasive Developmental Disorder, Vol. 1: Diagnosis, Development, Neurobiology and Behavior*, 3rd edn. Hoboken, NJ: Wiley.
- Wahlberg, T., Obiakor, F., Burkhardt, S., and Rotatori, A. F. (eds.) (2001). *Advances in Special Education, Vol. 14: Autism Spectrum Disorders: Educational and Clinical Interventions*. London: Elsevier Science.

Relevant Websites

- <http://www.asatonline.org> – Association for Science in Autism Treatment.
- <http://www.autism-society.org> – Autism Society of America.
- <http://www.cdc.gov/ncbddd/autism/index.htm> – Centers for Disease Control and Prevention: Autism Information Center.
- www.firstsigns.org – First Signs.
- <http://www.udel.edu/bkirby/asperger> – Online Asperger Syndrome Information and Support, University of Delaware.
- <http://www.nichd.nih.gov/autism> – National Institute of Child Health and Human Development: Autism Research at the NICHD.
- <http://www.nas.org.uk> – The National Autistic Society.

Educating Students with Cognitive Disabilities

J P Bakken, Illinois State University, Normal, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cognitive disability – An intelligence quotient (IQ) standard score of approximately 70 or below, with concurrent deficits in adaptive skill areas.

Community-based training – It provides functional-skills training in an integrated business setting, addressing the production, quality, and social demands of the natural work environment.

Image rehearsal – A form of visualization wherein the individual is taught to associate aspects of a task with pictures of events what will help to recall them.

Integrated settings – Settings which have students with and without disabilities.

Least prompts – These are also referred to as a response–prompt hierarchy, because the trainer progresses from the least amount of assistance (usually a verbal prompt) to the most intrusive (usually a physical prompt) until one prompt simulates correct responding.

Reinforcers – These strengthen the probability of (a response to a given stimulus) by giving or withholding a reward.

Self-determination – A person's ability to set and navigate his or her own life course.

Self-reliance – A person's ability to take care of himself or herself.

Systematic instruction – The organized and planned content that is directly taught to students.

Time delay – A procedure that allows the teacher to gradually fade assistance until the student performs without prompting.

Verbal rehearsal – It relates to the concept of self-instruction and refers to labeling aspects of a task, and verbalizing these labels aloud or silently while the task is being performed.

By definition, individuals who have cognitive disabilities are different from people who do not have cognitive disabilities based on intellectual functioning and adaptive skills. According to the most recent American Association on Mental Retardation (AAMR, 1992) definition, the assessed intellectual functioning of individuals who have cognitive disabilities is an intelligence quotient (IQ) standard score of approximately 70 or below, with concomitant deficits in adaptive skill areas. Traditionally, IQ has been described along a continuum of mild (IQ of 55–70),

moderate (IQ of 40–54), severe (IQ of 25–39), and profound (IQ less than 25), according to the degrees to which a person's measured general intelligence deviates from the average range. The term adaptive behavior historically has been used to convey the nature of one's personal independence and social responsibility.

For students with more severe cognitive disabilities, extensive supports are provided on a constant basis, across all environments, and are potentially life sustaining in nature. Supports may be provided to assist with basic and/or complex tasks of daily life, with management of emotional or psychological concerns, or with physical health needs. Persons with extensive support needs are likely to have difficulty acquiring new skills, applying the skills they have already learned to new situations, and communicating their wants, needs, thoughts, and feelings. However, these skills are helpful in functioning successfully in the social world. For this reason, persons with severe cognitive disabilities are likely to concentrate a great deal of time on learning adaptive skills so they can function in society.

Generally, characteristics of individuals with cognitive disabilities include short attention spans, problems with short-term memory, speech and language difficulties, generalizing information to new situations, motor difficulties, and frustration tolerance, especially with regard to academic learning and abstract instructional activities. For students with cognitive disabilities, their different characteristics may impact their learning. These characteristics are tied to attention, memory, speech and language, transfer/generalization, and strategies usage. See Table 1 for examples and descriptions of these.

Teaching Students with Cognitive Disabilities

To teach students with cognitive disabilities, instructional techniques that include drill and practice with flashcards and pictures and repetition of content and skills are often implemented. In addition, the transfer of knowledge learned in the classroom and on community trips where the focus is on generalization of skills is also important. Effective classroom teachers have direct control of instructional strategies they employ. These teachers not only have a large array of such strategies at their disposal, but are also adept at determining which strategies to use with specific students and content (Marzano, 2003). Interventions must be sufficiently

powerful to improve the performance of students with disabilities and lend themselves to integration with current teaching practices and conceptions. How do we know, as teachers, if an intervention is effective? Research indicates that effective interventions should be supported by outcome data (Gersten *et al.*, 1995). Guidelines for the evaluation and implementation of effective instructional practices include consistent and regular monitoring of students (daily and periodically), and implementing of informal and formal assessment strategies (as appropriate). All accountability measurements are usually aligned with content and processes involved in the practice, and trouble-shooting and/or interventions are immediately available in response to low performance.

Effective Teaching Behaviors

When presenting material to students with cognitive disabilities, it is more beneficial to the learner to present material in a group or cluster form than in random order. Grouping is perhaps the simplest method of organizing information (Spitz, 1973, 1979). Material may be grouped spatially (in different visual arrangements), temporally (with a pause or time lapse between items), perceptually (with certain items enclosed in a shape or configuration), or categorically (by content or commonality of items).

Verbal learning mediation refers to the process by which an individual connects a stimulus and a response. It can be used to work with students with cognitive disabilities. Verbalizing the connection between two stimulus words seems to enhance performance. The meaningfulness of the material and the use of stimulus words or objects familiar to the subjects also facilitate learning in paired associate tasks (Estes, 1970). Verbal rehearsal can also be used with such students. It relates to the concept of self-instruction and refers to labeling aspects of a task and verbalizing these labels aloud or silently while the task is being performed. This strategy allows the student to hear or think about all the steps needed to perform a task. In image rehearsal, a form of visualization, the individual is taught to associate aspects of a task with pictures of events that will help recall them. This strategy allows the learner to connect abstract information to concrete information to aid in the learner's recall of the material (Mercer and Snell, 1977).

Modeling, imitation, and learning through observation play fundamental roles in the education of students with cognitive disabilities. These are terms most often associated with observation learning, which refers to learning from the demonstration of others. Suggestions for using observational learning as a teaching tool include: (1) being aware that any behavior may serve as a model, (2) using prompts or cues to direct students' attention, (3) calling attention to students exhibiting desirable behaviors, (4) ignoring undesirable

Table 1 Learning related characteristics

Characteristics	Description
Attention variables	<ul style="list-style-type: none"> • Difficulties with attention span (length of time on task), focus (inhibition of distracting stimuli), and selective attention (discrimination of important stimulus characteristics) • Students need to be trained to be aware of the importance of attention and to learn how to actively monitor its use
Memory	<ul style="list-style-type: none"> • Short-term memory (STM) is more difficult for students • Students typically have a strong long-term memory (LTM) • STM problems involving nonacademic tasks are associated with the lack of spontaneous use of mediation strategies • Strategy production is difficult for students with cognitive disabilities, but improves in recall can be achieved when they are shown how to proceed in an organized, well-planned fashion
Speech and language	<ul style="list-style-type: none"> • Tend to be passive learners • Speech and language problems occur with greater frequency among individuals identified as cognitively disabled • Language disorders that commonly accompany cognitive disabilities include delayed language development and a restricted or limited vocabulary • The speech problems often seen are difficulties in articulation, voice, and stuttering • Children who do not have the verbal communication skills of their peers may withdraw from interpersonal relationships or seek attention in inappropriate ways. These children may misbehave because they cannot clearly distinguish between acceptable and unacceptable standards of behavior. Problem behavior can also result from the frustrations of scholastic failure or as an attempt to gain acceptance from other children, who might encourage inappropriate behavior
Transfer/generalization	<ul style="list-style-type: none"> • Tend to show deficiencies in the ability to apply knowledge or skills to new tasks, problems, or situations • In particular, they may fail to use previous experience to develop rules that will help solve future problems that are similar
Strategies	<ul style="list-style-type: none"> • Less likely than normal learner to employ effective techniques for organizing information for later recall • Mature learners implement effective strategies like verbal rehearsal and repetition, and imagery • Students with cognitive disabilities have difficulty producing effective strategies

behaviors so that others do not model it in an attempt to gain attention, and (5) rewarding imitation of appropriate behavior.

The majority of the literature on teaching vocational tasks to individuals with cognitive disabilities focuses on the use of least prompts (Cuvo *et al.*, 1978; Test, *et al.*, 1988). This strategy is also referred to as a response-prompt hierarchy, because the trainer progresses from the least amount of assistance (usually a verbal prompt) to the most intrusive (usually a physical prompt) until one prompt simulates correct response. The use of a least-prompting strategy can be very effective for teaching skills on community job sites and may also be effective for training students who have long been dependent on teachers for verbal instruction. In addition, gestures can be used instead of a full-model prompt or partial physical assistance such as touching the student's arm. Clearly, regardless of the types of prompts selected, the teacher should establish a latency period or time that he/she will wait for the student to respond before providing the next level of assistance. Usually a student should be given approximately 3–5 s to respond independently. Students with physical disabilities, however, may require longer latency periods based on their movement limitations, and this requirement should be determined on an individual basis (Inge, 1992; Sowers and Powers, 1992). Finally, the teacher is cautioned to deliver each prompt only once before moving to the next, more intrusive prompt. See **Table 2** for the guidelines for implementing a least-prompt hierarchy and **Figure 1** for an example of a data recording sheet. Before this is begun, a task analysis of the to-be-learned behavior must first be completed.

The use of time delay on vocational training sites is an option for teachers of students with cognitive disabilities (Inge *et al.*, 1993). There are several critical components to a time-delay procedure (Gast *et al.*, 1988; Snell and Gast, 1981). First, the teacher must select a prompt that will consistently assist the student to perform the task correctly. Initially, the prompt is given simultaneously with the request to perform the job duty. Gradually, increasing amounts of time (usually seconds) are awaited between giving the request to perform the task and providing the prompt to complete the task correctly. The number of trials at each delay level and the length of the delay should be determined prior to initiation of the program. By the teacher's pairing of the prompt with the request to perform a work task, the student is not allowed to make errors initially. The delay procedure allows the teacher to gradually fade assistance until the student performs without prompting. For example, trials are designated for 0-s delay, the next at 2 s, the next at 4 s, and so forth, until the student performs without assistance. Unlike the system of least prompts, time delay requires that the teacher select one prompt for use during the instructional program. Therefore, the procedure would be particularly useful if

Table 2 Least-prompt hierarchy guidelines for students with cognitive disabilities

Guidelines for using a least-prompt hierarchy

1. Have the student move to the appropriate work area unless movement is part of the task analysis (TA)
2. Stand behind or beside the individual so that you can quickly provide prompts when necessary
3. Provide the cue to begin the task (e.g., clean the white board, pass out the papers, etc.)
4. Wait 3 s for self-initiation of Step 1 of the TA
5. If the student completes the step independently, provide reinforcement and proceed to Step 2 of the TA. Score + or – on the data sheet
6. If the student is incorrect or does not respond within 3 s, provide a verbal prompt specific to Step 1 of the TA (e.g., pick up the eraser)
7. If the student completes the step with a verbal prompt, provide reinforcement and move to Step 2. Record V (for verbal) on the data sheet
8. If the student is incorrect or does not respond within 3 s, model the response (e.g., teacher picks up the eraser)
9. If the student completes the step with a model prompt, provide reinforcement and move to Step 2. Record M (for model) on the data sheet
10. If the student is incorrect or does not respond within 3 s, physically guide him or her through the response (e.g., teacher guides the student's hand to pick up the eraser). Record P (for physical) on the data sheet
11. Begin instruction on Step 2 of the TA
12. Repeat this procedure for each step in the TA until the task is completed. Always interrupt an error with the next prompt in the least-prompt system.

a student has consistently demonstrated a preference for one type of prompt. If an error occurs during time delay, the teacher should implement an error-correction procedure. Usually error correction consists of immediately interrupting the student's mistake and providing the prompt. If the student makes three or more errors in a row, the teacher may consider reverting to a number of trials at 0 s before again delaying the prompt. Monitoring of the training data is not constantly making errors during the procedure. See **Table 3** for a detailed description of time delay.

The selection of reinforcers and the systematic delivery of reinforcement are critical for student success on community-based vocational sites. The most effective reinforcers are those that arise as a natural consequence to a given task or situation within the work environment (Wilcox and Bellamy, 1982). Therefore, the teacher should begin by attempting to identify items that are available in a specific community-based setting. However, it should be remembered that not all individuals will be reinforced by the same items and that even the most preferred reinforcer used too frequently might lose its effectiveness (Falvey, 1989). Only after failing to identify a

Steps of activity (purchasing lunch at a fast food restaurant)	Behavioral recording	
1. Open door of restaurant	+	
2. Get in line	—	V
3. Decide what to order	—	M
4. Place order	+	
5. Pay for order	—	P
6. Get food and drink and sit down	+	
7. Eat meal	+	
8. Throw trash in garbage	—	V
9. Exit restaurant	+	
+ = independently performed step — = independently did not perform step V = verbal prompt required to perform step M = modeling required to perform step P = physical prompt required to perform step		

Figure 1 Sample data recording sheet.**Table 3** Time-delay explanation, types, and examples*What is time delay?*

Time delay is a prompt-fading strategy. With time-delay prompting, you fade instructional prompts by inserting a delay between giving an instruction and stepping in and prompting. Time delay is considered to be an error-minimizing prompting method. That is, we use this method to prevent a child from making a lot of errors while learning a new skill.

There are two types of time-delay prompting:

Progressive time delay:

The time between giving the instruction and giving a prompt is gradually increased

Constant time delay:

A fixed amount of time is always used between the instruction and the prompt

What types of skills can be taught using time delay?

Time delay can be used to teach almost any type of skill, including:

Communication skills

Imitation, academic skills (e.g., number identification or sight word reading)

Daily living skills (e.g., grocery shopping)

Leisure skills

natural reinforcer should the teacher select more artificial items. Teachers are also cautioned to select only age-appropriate material for use on community sites. See **Table 4** for steps to identify student reinforcers.

Table 4 Identifying potential student reinforcers

1. Survey people familiar with the student (i.e., parents, siblings, and grandparents) to determine likes and dislikes. Include leisure activities, tangible items, types of verbal reinforcement, and so forth.
2. Observe the student in several natural environments during his or her free time and record what he or she does.
3. Offer the student a chance to interact with several novel items and record what he or she does. Repeat the experience over several days and determine if there is a pattern to item selection.
4. Select an item and use it as a reinforcer for a behavior the student already performs independently. Observe to see if that behavior increases.

Strategic Educational Programming for Students with Cognitive Disabilities

Educational programming for persons with cognitive disabilities has made great strides in the last 25 years. Specifically, systematic instruction of students with severe cognitive disabilities has provided empirical data to illustrate that learning actually occurs (Snell, 1993). Optimal educational programs are generally community referenced, or related to actual incidents that naturally occur in the environment. Clearly, the most useful learning activities are both functional and age appropriate. Education is designed

to enhance the student's ability to function successfully in a world inhabited largely by people without disabilities and, therefore, it is important that instruction be delivered in integrated settings. Consider this example. There is a difference between a person from Illinois who took a class in German as a high school sophomore and a person from Illinois who spent his sophomore year as an exchange student in Germany. Although it is possible to learn about another culture without experiencing it firsthand, opportunities for learning are greatly enhanced through direct experience. There are four reasons that all school-aged children should be educated in the same settings (Brown *et al.*, 1989). First, students without disabilities who are educated alongside students with disabilities are more likely to function as responsible adults in a pluralistic society. Second, integrated schools provide more meaningful instructional environments. Third, families have greater access to activities in neighborhood schools. Finally, integrated schools offer more opportunities to develop a wide range of social relationships. Integrated learning settings are as important for adults as they are for children. When teaching a person who has pronounced difficulties learning new information, it is very important to give the person every advantage. For adults, this means that job skills are better taught on the job, and domestic skills are better taught in the person's home.

The determination of a curriculum is overwhelmingly important for learners with cognitive disabilities. Skills are typically found in the domains of communication, domestic, leisure, self-help, social/friendship, and vocational. Recently, self-determination skills have been added to the best-practice curricula. The need to teach self-determination was first documented in the literature related to employment (Moon *et al.*, 1990). Although persons with cognitive disabilities are able to complete requisite job skills, they are frequently dependent on their job coaches to cue them to begin and end tasks. The development of self-determination competencies challenges students to become actively involved in their learning and decision making. Ultimately, self-determination and overall student empowerment improve the quality of students' adult lives. Increasing students' self-determination may increase their success in transitioning from high school to adult living. See **Table 5** for some other effective research-based strategies.

The Use of Assistive Technology

Assistive technology is defined by the Technology-Related Assistance Act (Tech Act) of 1988 (P. L. 100-407), and the Individuals with Disabilities Act (IDEA) of 1990 (P. L. 101-476), as "any item, piece of equipment, or product system, whether acquired commercially off-the-shelf, modified, or customized, that is used to increase, maintain or improve the functional capabilities of individuals with disabilities."

It can describe both devices and services that aid an individual. Assistive technology allows a person with disabilities (in this case, a person with cognitive disabilities) to become an integral part of the school or community (Hasselbring and Peabody College of Vanderbilt University, 1998). Examples of commonly used assistive-technology devices that illustrate the formal definition include positioning equipment, mobility devices, computer applications, adaptive toys and games, adaptive environments, electronic interfaces, homemade battery-powered toys, medical equipment, prostheses, and alternative and augmentative communication aids (Parette *et al.*, 1996). There are two purposes of assistive technology. First, technology can augment an individual's strengths by counterbalancing the effects of a disability; and second, it can provide alternative methods for performing a task so that disabilities can be compensated for or bypassed entirely (Lewis, 1998). For example, an individual with difficulties in reading who possesses good listening skills can listen to books on tape rather than read the print version. Persons with poor computational skills but with good fine-motor skills may use a hand-held calculator. Those with poor spelling abilities but with a measure of computer literacy may write with a word processor that offers assistance in spelling. See **Table 6** for assistive-technology options.

To determine the type of technology or support a particular student may need, it is critical to address the issues of individual needs and differences. Applying technology involves matching the individual's exhibited needs with the potential benefits possible through the rise of the technology (Parette and Murdick, 1998). In addition, less emphasis should be placed on the categorization of the devices as being low or high-tech. Considering that students with disabilities possess individual interests, strengths, and weaknesses, one may conclude that a device appropriate for one person may be inappropriate for another. In a similar way, a device that may assist in one setting may be inappropriate in a different situation or environment (Bryant *et al.*, 1998). Once individual needs have been targeted and appropriate technology has been applied, then the technology has the potential to provide a variety of needs for a person with a disability such as a cognitive disability.

Instructional modifications include changes to teaching procedures, curricula, management materials and technology, and the physical environment to facilitate learning (Bryant and Bryant, 1998). Specifically, assistive technology may require simple, low-tech modifications. Off-the-shelf technologies can become adaptive when they are used to enhance the learning of a student with a cognitive disability (Lewis, 1998). For example, an audiotape becomes assistive when it is used to compensate for an individual's memory or note-taking problems. In addition, low-tech modifications such as sticky notes, flags, and highlighters can enhance a student's organizational skills. Such modifications require minimal time and

Table 5 Other strategies for students with cognitive disabilities

<i>Strategy</i>	<i>Description</i>
Mnemonic strategy	The keyword method is a mnemonic (memory-enhancing) technique used to increase the initial learning and retention of facts and fact systems of the kind often encountered in schools. The method incorporates both auditory and visual cues to enhance meaningfulness of the information to be learned and to promote strong associations between questions and answers.
Self-instruction	Self-instructional training has been employed to increase general behaviors required for academic success, such as student attention to task as well as to develop specific academic skills, such as handwriting, reading comprehension, and math performance. Specifically, the teacher first models what to say and do by verbalizing task-relevant strategies while performing a task. Responsibility is then gradually faded to the students by requiring them to perform the task while the teacher continues to verbally direct the student's performance. Next, the teacher fades responsibility further by requiring the students to verbally cue their behaviors while performing the task. The final steps of self-instruction training consist of the student being told to convertly cue their behaviors. At this point, teacher direction has been completely faded, and students are self-regulating their performance.
Visual supports	Visual supports are any visually perceived stimuli that assist us in comprehending environmental information and demands. In the classroom, visual supports help students understand direction, schedules, rules, and instructional materials. Pictures are universally understood and, therefore, can be generalized to most every functional setting. Combining visual cues with speech can help some children better understand what is verbally said to them.
Empowerment strategies	Choice boards give students a visual representation of their options at any given time. By pointing to or handing a picture or icon to another person, the student is able to communicate what he/she wants: what activity to do, what materials to use within an activity, whom to work with, and when to conclude the activity. Token boards serve as conditioned reinforcers for positive behaviors. This fundamental concept of a token board is that it serves as a visual display for students who can earn a specified number of tokens contingent on their behavior or performance.
Information-gathering strategies	Social stories give students precise information about what is happening in a given situation, concrete descriptions of relevant thoughts or feelings of other directly involved in the situation, and instruction to the student for behaviors that might come with those changes. A strategy similar to social stories is comic strip conversations. Comic strip conversations teach children with autism the art of conversation. This version is a conversation between two people that is illustrated on paper as the conversation evolves.
Simultaneous prompting	Visual schedules use pictures, photos, or words to create the order of events or activities. Simultaneous prompting is a systematic form of the antecedent prompt and test procedure because all trials are conducted with 0-s delays between the discriminative stimulus and the teacher prompts. A very effective strategy to teach grocery store sight words.

Table 6 Assistive-technology solutions

- *High-tech* solutions involve the use of sophisticated devices, such as computers and interactive multimedia systems.
- *Medium-tech* solutions use less-complicated electronic or mechanical devices, such as videocassette players and wheelchairs.
- *Low-tech* solutions are less sophisticated, such as adapted spoon handles, Velcro fasteners, or raised desks that can accommodate a wheelchair.
- *No-tech* solutions require no devices or equipment. These might involve the use of systematic teaching procedures or the services of related services personnel such as physical or occupational therapists.

training to be implemented in the classroom. Simple and uncomplicated modifications are sometimes all that are needed to allow a student to use a computer software program (Olson and Platt, 2000). Standard computer

keyboards pose a number of problems for some students with cognitive disabilities (Kincaid, 1999). Various modifications may need to be considered so that students with limited hand and finger mobility can access computer technology. Before a student can use computer technology for a given task, an appropriate method for inputting information must be available.

To a large measure, word processing has the capability of helping individuals improve their writing skills (MacArthur, 1996). Students with cognitive disabilities often possess limited conceptualization of editing and revising; thus, such students limit their revision to minor errors that fail to strengthen a written document as a whole. Therefore, using word processor may not only teach students who are cognitively delayed to edit their writing better but also can help them to make frequent revisions without labored rewriting. Word processing reduces resistance to revising as a whole and eliminates errors due to transcription (MacArthur *et al.*, 1991).

Table 7 Maximizing assistive technology for students with cognitive disabilities

- Use open-ended devices that permit customizing the user and/or task, are generally the most useful.
- Find the lowest technology solution that can provide a level of performance or function rather than a complex, high-technology device or system. Simply changing a student's angle of view of a computer monitor (e.g., placing the monitor at eye level as opposed to hairline level or above) could reduce strain and improve performance.
- Collaborate with other teachers. There are too many technologies developing too quickly for one teacher to monitor advancement in technology independently. Sharing expertise can help.
- Collaborate with parents to ensure that assistive-technology devices that go home are properly used and maintained. Parents can also be an excellent source of evaluative information on how effectively a device or piece of software is working.
- Do not believe that you have to master a device or software application completely before you begin using it. Many times an application can be used successfully early in the learning curve, and learning by doing can prompt eventual mastery.
- Assistive devices should match the age, gender, and preferences of the user to promote acceptance and use.
- The arrangement and separation of controls should be predictable and natural. Feedback to the user must be meaningful.
- Be sure that your school or school system has a comprehensive policy covering assistive technology, including the protection of student and teacher privacy, the repair and maintenance of equipment, and the home use of the school purchase equipment. Be sure also that the policy states who is the assistive-technology resource for the individualized education plan (IEP) team.
- Simply purchasing assistive-technology equipment will not ensure its use. Funds must also be allocated to ensure that teachers and other potential assistive-technology service providers receive training in the use of the equipment.
- Do not be afraid to experiment. Assistive technology is a very young field, and everybody is learning.

In addition, it has the potential to facilitate other operational revisions such as moving content and deleting material.

It has been proved that computers can help young children with cognitive disabilities develop language (MacArthur, 1996). The goal of any language development program is to provide young children with the tools for independent communication. Some children will learn to speak, some will learn sign language, and others will need the assistance of augmentative communication. Augmentative communication refers to a set of approaches used to improve the communication skills of persons who do not speak or whose speech is not intelligible (Olson and Platt, 2000). Aided systems require the use of a picture or word board, a notebook, or a computerized aid. An unaided system requires the individual to use only hand or body motions to communicate (e.g., sign language). Augmentative communication options can range from

high- to low-tech devices including aids such as a symbol system, manual communication boards, electronic communication devices, speech synthesizers, and communication-enhancement software. Communication boards, a low-tech alternative to augmentative communication, assist young children in language expression. Communication boards are usually made of cardboard or any other material used to display choices for children who cannot speak (Ysseldyke and Algozzine, 1990). See Table 7 for ways to maximize the use of assistive technology.

Moving Forward: For Persons with Cognitive Disabilities

Independence is a broad construct that encompasses self-reliance and self-determination of persons with cognitive disabilities. Self-reliance refers to a person's ability to take care of himself/herself. Skills that reflect self-reliance range from feeding oneself to living in a home alone. Educational programs offer many opportunities to encourage self-reliance. As described earlier, self-determination refers to a person's ability to set and navigate his/her own life course. Skills that educators may help to develop include making choices, communicating preferences, setting achievable goals, and self-advocating (Wehmeyer, 1993). In any society, independence is largely tied to productivity, which is generally defined as holding a job and being economically self-sustaining. Therefore, the most desirable outcome for all individuals, including those with cognitive disabilities, is competitive employment. For this to occur, education and training must focus on vocational instruction and on-the-job skill development. Research generated during the past 20 years demonstrates that competitive employment is a visible outcome for persons with cognitive disabilities (Hill *et al.*, 1987; Moon *et al.*, 1990). For successful school-to-work transitions, there must be individual transition plans, improved work opportunities, job placement in competitive, integrated settings, and documentation of progress in employment-related skills.

Good quality of life is an important educational outcome for persons with cognitive disabilities. Although it is important that persons with cognitive disabilities see and are seen by their nondisabled peers, interventions are sometimes necessary to ensure true social involvement and quality of life. Possibilities include teaching the person with cognitive disabilities where and how to find others with common interests and educating nondisabled peers about the person with the disability (Schoeller, 1997). In addition to focusing on quality of life, it is important to focus on person-centered planning. Although the two constructs are based on a philosophy that people with cognitive disabilities are entitled to live well-rounded lives, some differences do exist. Whereas the interest in quality of life has largely been

based on a desire to define and analyze its components, proponents of person-centered planning adhere to the notion that outcomes are diverse and individualized. In other words, the person should be enabled to express his/her own goals. Person-centered planning begins by asking the person what he/she wants life to be like. People who care about the individual are invited to join the process; their role is especially important when planning with a person who has severe cognitive disabilities. They are sometimes called on to speak for the person or to translate the person's attempts to communicate. The planning process involves service providers and educators who are able to identify resources for attaining goals (Gaylord, 1997). The goal is to make sure that outcomes are clear, individualized, and easily measured.

Conclusion

This article discussed ways that teachers can best educate children and youth with cognitive disabilities. The characteristics of effective interventions (e.g., least prompts, time delay, and identifying reinforcers) are discussed along with specific curriculum and teaching strategies (e.g., systematic instruction, integrated settings, and assistive technology). Additionally, educational outcomes (e.g., quality of life and personal outcomes) are highlighted. Clearly, persons with cognitive disabilities are humans; and as humans, every effort must be made to maximize their fullest potential.

See also: Assessment Accommodations for Children with Special Needs; Inclusion of Students with Special Needs in General Education Classrooms; Instructional Accommodations for Children with Special Needs In Inclusive Settings; Large Scale Assessment and Accountability and Students with Special Needs; Self-Determination; Social Competence.

Bibliography

- AAMR (American Association on Mental Retardation) (1992). *Mental Retardation: Definition, Classification, and Systems of Support*, 9th edn. Washington, DC: Author.
- Brown, L. E., Udavari-Solner, A., Davis, L., et al. (1989). The home school: Why students with severe intellectual disabilities must attend the schools of their brothers, sisters, friends and neighbors. *Journal for the Association of Persons with Severe Handicaps* **14**(1), 1–7.
- Bryant, D. P. and Bryant, B. R. (1998). Using assistive technology to enhance the skills of student with learning disabilities. *Intervention in School and Clinic* **34**(1), 53.
- Bryant, D. P., Erin, J., Lock, R., Allan, J. M., and Resta, P. E. (1998). Infusing a teacher preparation program in learning disabilities with assistive technology. *Journal of Learning Disabilities* **31**, 55–66.
- Cuvo, A. J., Leaf, R. B., and Borakove, L. S. (1978). Teaching janitorial skills to the mentally retarded: Acquisition, generalization and maintenance. *Journal of Applied Behavior Analysis* **11**, 345–355.

- Estes, W. K. (1970). *Learning Theory and Mental Development*. New York: Academic Press.
- Falvey, M. A. (1989). *Community-Based Curriculum: Instructional Strategies for Student with Severe Handicaps*, 2nd edn. Baltimore, MD: Brookes.
- Gast, D. L., Ault, M. F., Woler, M., Doyle, P. M., and Belanger, J. (1988). Comparison of constant time delay and the system of least prompts in teaching sight word reading to student with moderate retardation. *Education and Training in Mental Retardation* **23**, 117–128.
- Gaylord, V. (1997). Social inclusion of adults with developmental disabilities. *ERIC Document Reproduction Service no.* ED 412 704.
- Gersten, R., Morvant, M., and Brengelman, S. (1995). Closer to the classroom is close to the bone: Coaching as a means to translate research into classroom practice. *Exceptional Children* **62**(1), 52–66.
- Hasselbring, T. S. and Peabody College of Vanderbilt University (1998). *The Future of Special Education and the Role of Technology*. <http://peabody.vanderbilt.edu/lrc/hasselbring/future.html> (accessed Jun. 2009).
- Hill, M. L., Wehman, P. H., Kregel, J., Banks, P. D., and Metzler, H. M. (1987). Employment outcomes for people with moderate and severe disabilities: An eight-year longitudinal analysis of supported competitive employment. *Journal of the Association of Severe Disabilities* **12**(3), 182–189.
- Individuals with Disabilities Education Act Amendments of 1990 (P. L. 104–476). Washington, DC: US Government Printing Office.
- Inge, K. J. (1992). Cerebral palsy. In McLaughlin, N. P. and Wehman, P. (eds.) *Developmental Disabilities: A Handbook for Best Practices*, pp 30–53. Fort Washington, PA: Andover Medical Publishers.
- Inge, K. J., Moon, M. S., and Parent, W. (1993). Applied behavior analysis in supported employment settings. *Journal of Vocational Rehabilitation* **3**(3), 53–60.
- Kincaid, C. (1999). Alternative keyboards. *Exceptional Parent* **29**(2), 34–37.
- Lewis, R. B. (1998). Assistive technology and learning disabilities: Today's realities and tomorrow's promises. *Journal of Learning Disabilities* **31**(1), 16–26.
- MacArthur, C. A. (1996). Using technology to enhance the writing processes of student with learning disabilities. *Journal of Learning Disabilities* **29**(4), 344–354.
- MacArthur, C. A., Graham, S., and Schwartz, S. (1991). Knowledge of revision and revising behaviors among students with learning disabilities. *Learning Disability Quarterly* **14**, 61–73.
- Marzano, R. J. (2003). Using data: Two wrongs and a right. *Educational Leadership* **60**(5), 56–60.
- Mercer, C. D. and Snell, M. E. (1977). *Learning Theory Research in Mental Retardation: Implications for Teaching*. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Moon, M. S., Inge, K. J., Wehman, P., Brooke, V., and Barcus, J. M. (1990). *Helping Persons with Severe Mental Retardation Get and Keep Employment. Supported Employment Issues and Strategies*. Baltimore, MD: Brookes.
- Olson, J. L. and Platt, J. M. (2000). *Teaching Children and Adolescents with Special Needs*. Upper Saddle River, NJ: Merrill.
- Parette, H. P. and Murdick, N. L. (1998). Assistive technology and IEP's for young children with disabilities. *Early Childhood Education Journal* **25**(3), 193–197.
- Parette, H. P., Brotherson, M. J., Hourcade, J. J., and Bradley, R. H. (1996). Family-centered assistive technology assessment. *Interventions in School and Clinic* **32**, 104–112.
- Schoeller, K. (1997). Overcoming barriers to social inclusion. *Impact* **10**(3), 4–5.
- Snell, M. and Gast, D. L. (1981). Applying delay procedures to the instruction of the severely handicapped. *Journal of the Association of the Severely Handicapped* **5**(4), 3–14.
- Snell, M. E. (1993). *Instruction of Student with Severe Disabilities*, 4th edn. New York: Macmillan.
- Sowers, J. and Powers, L. (1991). *Vocational Preparation and Employment of Students with Physical and Multiple Disabilities*. Baltimore, MD: Brookes.
- Spitz, H. H. (1973). Consolidating facts into the schematized learning and more of mental retarded. *International Review in Mental Retardation* **6**, 149–168.

- Spitz, H. H. (1979). Beyond field theory in the study of mental deficiency. In Ellis, N. R. (ed.) *Handbook of Mental Deficiency: Psychological Theory and Research*, 2nd edn., pp 121–141. Hillsdale, NJ: Erlbaum.
- Technology-Related Assistance Act of 1988 (P. L. 100–407) Washington, DC: US Government Printing Office.
- Test, D. W., Grossi, T., and Keul, P. (1988). A functional analysis of the acquisition and maintenance of janitorial skills in a competitive work setting. *Journal of the Association for Persons with Severe Handicaps* **13**(1), 1–7.
- Wehmeyer, M. (1993). Self-determination as an educational outcome. *Impact* **6**(4), 6–7.
- Wilcox, B. and Bellamy, G. T. (1982). *Design of High School Programs for Severely Handicapped Students*. Baltimore, MD: Brookes.
- Ysseldyke, J. E. and Algozzine, B. (1990). *Introduction to Special Education*. Boston, MD: Houghton Mifflin.
- Mastropieri, M. A., Bakken, J. P., and Scruggs, T. E. (1991). Mathematics instruction for individuals with mental retardation: A perspective and research synthesis. *Education and Training in Mental Retardation* **26**(2), 115–129.
- Mastropieri, M. A., Scruggs, T. E., Whittaker, M. E. S., and Bakken, J. P. (1994). Applications of mnemonic strategies with students with mild mental disabilities. *Remedial and Special Education* **15**(1), 34–43.
- Odom, S. L., Horner, R. H., Snell, M. E., and Blacher, J. (eds.) (2007). *Handbook of Developmental Disabilities*. New York: Guilford Press.
- Skouger, J. R., Kelly, M. L., Roberts, K. D., Leake, D. W., and Stodden, R. A. (2007). Technologies for self-determination for youth with developmental disabilities. *Education and Training in Developmental Disabilities* **42**(4), 475–482.
- Wehmeyer, M. L. (1995). A career education approach: Self-determination for youth with mild cognitive disabilities. *Intervention in School and Clinic* **30**, 157–163.
- Wehmeyer, M. L. (2007). *Promoting Self-Determination in Students with Developmental Disabilities. What Works for Special-Needs Learners*. New York: Guilford Press.
- Whitman, T. L. (1987). Self-instruction, individual differences, and mental retardation. *American Journal of Mental Deficiency* **92**(2), 213–223.
- Bakken, J. P. (2003). Assistive technology for academics. In Wojcik, B. (ed.) *Assistive Technology: A Web-Based Introduction*, 1st edn. Normal, IL: SEAT Center, Illinois State University.
- Bakken, J. P. (2008). Data-based decision-making and students with developmental disabilities. In Parette, H. P. and Peterson-Karlan, G. R. (eds.) *Research-Based Practices in Developmental Disabilities*, 2nd edn., pp 431–450. Austin, TX: Pro-Ed.
- Cihak, D., Alberto, P. A., Taber-Doughty, T., and Gama, R. I. (2006). A comparison of static picture prompting and video prompting simulation strategies using group instructional procedures. *Focus on Autism and Other Developmental Disabilities* **21**(2), 89–99.
- Gardner, J. F. and Chapman, M. S. (1993). *Developing Staff Competencies for Supporting People with Developmental Disabilities: An Orientation Handbook*. Baltimore, MD: Brookes.
- Lee, S., Amos, B. A., Gragoudas, S., et al. (2006). Curriculum augmentation and adaptation strategies to promote access to the general curriculum for students with intellectual and developmental disabilities. *Education and Training in Developmental Disabilities* **41**(3), 199–212.
- Mastropieri, M. A. and Scruggs, T. E. (1998). Enhancing school success with mnemonic strategies. <http://www.idonline.org/article/5912> (accessed Jun. 2009).

Further Reading

Relevant Websites

- <http://www.abilityhub.com> – AbilityHub.
- <http://www.abledata.com> – ABLEDATA.
- <http://atto.buffalo.edu> – Creating Visual Environments, Assistive Technology Training Online Project: AT Basics.
- <http://www.pluk.org> – Family Guide to Assistive Technology, Parents, Let's Unite for Kids.
- <http://www.memory-key.com> – Mnemonics, About Memory.
- <http://www.ilr.cornell.edu> – Person-Centered Planning Education, Cornell University ILR School.
- <http://www.iidc.indiana.edu> – Visual Schedules and Choice Boards: Avoid Misinterpretation of Their Primary Functions, Indiana Institute on Disability and Community.

Educating Students with Emotional and Behavioral Disorders

S W Smith and G G Taylor, University of Florida, Gainesville, FL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cognitive behavioral interventions – Strategies used to teach the use of self-talk or inner speech to regulate overt behavior.

Externalizing behavior – Constant and destructive acting out behaviors such as verbal and physical aggression, arguing, chronic outbursts, and general unpleasant behaviors.

Functional behavioral assessment – A measure used to identify the purpose of a student's behavior by focusing on the function of the behavior in relation to its environment or context.

Internalizing behavior – Covert behaviors that are often concealed from plain view such as withdrawal, sad affect, excessive fears, and phobias.

Negative reinforcement – The removal of an aversive condition to serve as a form of reinforcement to increase desired behavior.

Overcorrection – Requiring the students to engage in repetitive action as a penalty for inappropriate behavior.

Positive reinforcement – A reward that is positive, or pleasing, to a student presented after an appropriate behavior.

Pre-referral – Systematic and collaborative efforts to prevent formal referral and placement of students in special education programs.

Punishment – The presentation of an aversive condition to decrease undesired behavior.

Response cost – The removal of a reinforcer when undesired behavior is displayed.

For professional educators, providing a quality public education for all children and youth has always been a challenge. Even in the best of circumstances, effectively delivering academic, social, and emotional learning can be complicated and frustrating, yet at the same time, it can be professionally rewarding. The educational enterprise, however, becomes especially difficult and perplexing when children and youth display behaviors that can be described as disruptive, deviant, violent, antisocial, aggressive, destructive, and/or noncompliant. Despite years of conceptual and empirical progress in understanding and successfully serving school-aged children and youth with emotional and behavioral disorders (EBD), there continues to be a host of perpetual problems that the field of special

education continues to work through and will for decades to come. This article focuses on how students with EBD are identified, the prevalence and characteristics of the disorder, and the complex issues related to where students with EBD are taught. Evidence-based prevention and intervention strategies that can maximize the potential of students with EBD are described and, finally, several special issues that confront the field of special education in the education of students with EBD are explored.

Defining EBD

Behavior exists on a continuum with no clear delineation between what is considered normal and what is abnormal or disordered. Indicators used to identify or diagnose disordered behavior include how frequently and severe undesired behavior is displayed. Defining disordered behavior is inherently subjective and the task to determine what intolerable and disordered behavior is often falls to authority figures such as medical personnel, teachers, or mental health professionals (Kauffman, 2005). While there is concern about the stigmatizing effects that assigning labels to children may have, the use of labels and accompanying terminology is a means from which to secure the necessary financial resources to diagnose and provide treatment. Terminology used across the education and related fields such as mental health and social services may vary to some extent, and, at times, they may signify the conceptual approach taken by a field. The legal term used in the special education federal definition of children with significant behavioral problems is emotionally disturbed. Various states use differing labels ranging from emotionally handicapped to socially impaired. Professionals in the field of special education commonly use the term EBD because they feel it more accurately describes the condition.

The Individuals with Disabilities Education Improvement Act (IDEA) of 2004 defines emotional disturbance as:

a condition exhibiting one or more of the following characteristics over a long period of time and to a marked degree, which adversely affects educational performance:

- A. An inability to learn which cannot be explained by intellectual, sensory, or health factors.
- B. An inability to build or maintain satisfactory interpersonal relationships with peers and teachers.
- C. Inappropriate types of behavior or feelings under normal circumstances.

- D. A general pervasive mood of unhappiness or depression.
- E. A tendency to develop physical symptoms or fears associated with personal or school problems.

The term includes schizophrenia. The term does not apply to children who are socially maladjusted, unless it is determined that they have an emotional disturbance. (US Department of Education, 2006)

Problems with the Federal Definition

Vague language and undefined terms included in the federal definition have led to dissatisfaction among professionals. Criticisms include the subjective nature of phrases such as to a marked degree or over a long period of time, the potential exclusion of children and youth who do not qualify because they are not academically delayed, or exclusion of children who are labeled socially maladjusted. Interpretation of the definition has been left to the states which some professionals describe as creating “at best a great potential for inconsistency across referred children and, at worst, conditions allowing unchecked bias, inequity, and prejudice” (Knoff, 1995: 852).

Years ago, the National Mental Health and Special Education Coalition attempted to improve the definition by including more up-to-date terminology, using language that was sensitive to ethnic and cultural differences, excluding problems due to temporary stress, acknowledging multiple disabilities, and not arbitrarily excluding the full range of emotional or behavioral disorders (Forness and Knitzer, 1992). While proponents are advocating for change, the proposed definition has not been adopted or incorporated into federal laws and regulations.

Characteristics of EBD

In general, students served in programs for EBD display highly visible and chronic externalizing maladaptive behaviors (e.g., verbal/physical aggression, noncompliance, excessive teasing, and classroom disruption) that are against universal social norms and behaviors expected for a safe learning environment. When engaging in aggression and disruption, peers may look elsewhere for companionship and positive social interactions and may elicit negative counter responses creating a negative spiral of aggression and continued rejection. At times, peers may reinforce a student's aggressive behavior unintentionally by attending to it or complying with the aggressor's demands. What is most problematic, however, is that students who exhibit aggressive, maladaptive behaviors associate with peers who act in similar ways, thus perpetuating their aggressive behavior and increasing their rejection by nonaggressive peers (Bierman *et al.*, 1996). For those students in EBD programs, who exhibit more covert maladaptive behaviors (e.g., depression, anxiety, excessive fears, and phobias),

their behaviors may also elicit social rejection that can contribute to withdrawal from lasting, positive social relationships.

Cognitive Characteristics

In general, students placed in special education programs for EBD have intelligence quotients (IQs) in the low average range, although scores can be below 70 (often considered mental retardation) and in the upper range that could be considered gifted. As might be expected, students with EBD in the low average IQ range, often experience academic difficulties. Since these students engage in classroom disruption and often lack motivation to complete tasks, they may experience increasing deficits in their academic performance as they progress through the school years. Students who are placed in EBD programs may have academic deficits in all areas, and may have more extreme deficits in reading and math (Bos *et al.*, 2002). Interestingly, researchers have found that 30–40% of students who exhibit EBD also have learning disabilities (LD; Fessler *et al.*, 1991).

Behavioral Characteristics

What most readily comes to mind about students with EBD is their externalizing behaviors, or constant and destructive acting-out behaviors such as verbal and physical aggression, arguing, chronic outbursts, and general unpleasant behaviors. What often goes unrecognized, however, are student behaviors such as withdrawal, sad affect, or excessive fears and phobias that are more covert or internalizing, and are concealed from plain view.

Externalizing behaviors

Externalizing or overt behaviors are characteristically interpersonal and are more easily identified by teachers and others professionals than the internalizing type. Externalizing behaviors are typically the acting-out and noncompliant type. Students who exhibit EBD are often described as aggressive and tend to experience continual negative interactions with authority figures at school and in the community. Externalizing behavior problems may be part of a student's escalating spiral of conflictual relationships with family members, peer group, and, in later years, with fellow workers and in marital situations. As these overt and undesired behaviors frequently elicit negative responses, they can limit the much-needed positive interactions with others and contribute to a deleterious pattern of opposition and negative social status.

Internalizing behaviors

Teachers and other professionals often overlook students who have significant internalizing or intrapersonal problems. As students with internalizing disorders do not

typically display more noticeable disruptive behaviors in the classroom, they become less frequently noticed as those with externalizing problems. Internalizing problems can take the form of childhood depression, anxiety, fears and phobias, and obsessive-compulsive disorder. Students' internalizing problems can sometimes negatively affect their academic success, socialization with peers and adults, and overall development.

Prevalence and Causes of EBD

The issue of prevalence of EBD is closely linked to its definition. The criteria used to judge the presence of any disorder is sometimes subject to a changing consensus. Therefore, it is difficult to make precise statements about the overall prevalence of children with behavioral disorders considering differing criteria across states and decades. Prevalence estimates, from the most recent Office of Special Education Programs, 25th Annual Report to Congress (US Department of Education, 2007), indicate that students ranging in age from 6 to 21, who received special education services for EBD, remained steady at about 0.7% from 1993 to 2003 and was approximately 8% of the total students in 2003 who received special education services. Identification of EBD is more likely for students who are racially and ethnically diverse. The US Department of Education (DOE) also reports that black students are 2.3 times more likely to receive special education services related to EBD than all other racial/ethnic groups combined.

The causes of EBD are complex and interrelated. Most behavioral problems occur in combination with other problems, thus causal influences can be viewed from both biological and environmental perspectives. According to Kauffman (2005), the causes of EBD can be arranged around four general categories: (1) the individual child, (2) family, (3) school, and (4) culture.

Individual Child

All behavior is a dynamic mix of biological dispositions and environmental contexts. Some biological factors suspected to contribute to emotional or behavioral disorders are: (1) genetics, (2) brain damage or dysfunction, (3) nutrition, (4) biochemistry, (5) physical illness or disability, and (6) temperament (Kauffman, 2005). Each of these factors may contribute individually or intermix with others to help explain the presence of EBD.

Family

Various influences, such as family structure, socioeconomic status, and styles of parental discipline, may serve as contributing factors to the development of EBD.

Complex interactions between conditions suggested by social demographic indicators and family dynamics have been shown to correlate with the presence of EBD. These indicators include poverty, absent fathers, divorce or separation, and hostile family relationships (Rutter, 1985). Researchers, however, suggest that while these factors may increase the risk for developing EBD, they may not be causes in, and of, themselves (Kauffman, 2005).

School

There are several factors found in schools that could contribute to the presence of EBD. Researchers indicate there is a strong correlation between poor academic performance and the presence of EBD. Disordered behavior and poor academic performance may result in social consequences that may cultivate more undesired behaviors. Kauffman (2005) suggested six ways schools may contribute to academic failure and disordered behavior, namely:

1. insensitivity to students' individuality,
2. inappropriate expectations for students,
3. inconsistent management of behavior,
4. instruction in nonfunctional and irrelevant skills,
5. poor reinforcement contingencies, and
6. undesirable models of school conduct.

In addition, the difficulty of finding qualified teachers for students with EBD places immense pressure on district personnel to provide quality instruction to those students who are often the most difficult to manage as well as the most difficult to teach. Chronic teacher shortages in special education and the fact that many teachers are ill-equipped and underqualified, especially in the area of EBD (Blake and Monahan, 2007; Hampton and Hess-Rice, 2003), can compromise the high-quality teaching necessary in academic instruction and social and emotional learning.

Culture

Aside from the cultural factors that influence a student's behavior in families and schools, standards and values of the larger cultures in which they live can have effect. The effects of popular culture with advertising, mass media, videogames, and violent movies may all play a role in increasing inappropriate behaviors (Coyne *et al.*, 2006). The culture of peer groups, communities, ethnicity, and social class are all factors that may also influence the development of a behavioral disorder. Rarely can the development of EBD be attributed to a single particular cause. Most significant behavioral problems, such as EBD, are due to a combination of factors that occur on many levels and locations. Physiology, parenting, teaching, and culture may all play a part in the development of a disorder but must be considered uniquely for each individual.

Assessment and Eligibility of Students with EBD

Regular education teachers who may experience significant problems with a student's behavior may think seriously about the referral of the student for special education services. Referral requires a significant assessment process designed to collect enough information by a team of professionals (including parent(s) and sometimes the student) to make an informed judgment about eligibility, determine an effective instructional program through an individualized education program (IEP), and agree upon the best placement option. To prevent formal referral, prereferral strategies are often generated as an attempt to keep students in the least-restrictive educational environment.

Prereferral

Prereferral strategies are systematic and collaborative efforts to prevent costly and time-consuming formal referral and placement of students in special education programs. For students with significant behavioral problems in the classroom, prereferral is a way for teachers to systematically approach behavioral issues to mollify their effect on the educational environment. Prereferral usually involves a team of education professionals who can approach carefully, in a collaborative manner, a student's behavioral issues. From this collaborative approach, specific action plans are developed to facilitate the student's success in the regular education environment, such as designing reinforcement programs for appropriate behavior, enlisting parent help, or enlisting other professionals' support such as school psychologist or school counselor. Before formal referral, prereferral efforts must be documented indicating that the student fails to respond positively to reasonable accommodations.

When Congress reauthorized IDEA in 2004, the law included a response to intervention (RTI) approach for the identification and eligibility of students with high-incidence disabilities. IDEA (2004) encourages schools to "... use a process that determines if the child responds to scientific, research-based intervention as part of the evaluation procedures..." (Section 1414(b)(6)). While this change in law was written specifically for determining eligibility of students with specific LD, the RTI model has found support for use across special education, including EBD.

RTI is a model used by school systems to systematically guide the prereferral process and ensure the provision of scientifically based intervention strategies. The three-tiered model provides increasingly intense services to students based on student outcome data. Tier-one universal interventions are based on core curriculum and address the needs of approximately 80% of the school population. Tier-two group interventions typically serve

approximately 15% of a school population and involve more intensive instruction. The third tier of intervention serves approximately 5% of a school population and provides services for the small number of students who need individualized instruction (Hawken *et al.*, 2008). While implementation differs across settings, the key components of RTI include a process that screens for at-risk students, monitors responsiveness to scientifically based instructional interventions, and uses outcome data to determine a course of action (Elliott, 2008).

Cheney *et al.* (2008) reported that tier-two instruction with at-risk students for EBD reduced the students' at-risk status and helped prevent the development of EBD. RTI data may also be taken into consideration when determining student eligibility for EBD services. The influence and implementation of the RTI model have had a positive impact on prereferral and other educational services for students with EBD.

Referral and Determining Eligibility

If prereferral strategies fail to provide a student the necessary supports for success, special education referral procedures become necessary. Referral for placement in special education, because of a suspected EBD, would include all of the typical evaluation procedures, such as measures of intelligence and achievement, but might also include assessment of social competence and peer relations through interviews, self-reports, anecdotal reports, behavior rating scales, and direct behavioral observations. For students being referred for EBD, direct observation, a method of recording specific behaviors, is typically conducted by someone other than the student's teacher, and the use of behavior rating scales insure that the assessment of students is based on numerous informational sources (i.e., people and behavioral instruments). As part of the assessment process, multidisciplinary team members must also assess the effect of a student's behaviors on peers, teachers, and family members and assess the effect of the teacher's behaviors. A teacher's interactions with a student may make a student's behavioral problems worse; hence, teacher and student interaction observations can significantly influence classification, placement, and treatment recommendations.

After all the necessary information is collected, students referred for EBD services become eligible if they exhibit behavioral problems as specified in the special education law to a marked degree and over a long period of time, and if their educational performance is affected. At the same time, the multidisciplinary team makes an eligibility decision and considers the need for related services (e.g., counseling and social services). For those students identified and requiring special education and related services, there exist a variety of placement options

along a continuum that ranges from least (e.g., regular class with supports) to most restrictive (e.g., hospital or homebound instruction). Other placements could include regular class placement with some programming in a special education resource room, a separate special education class for students with EBD for most of the school day, a separate school for students with EBD, or a residential setting. It is up to the multidisciplinary team to decide what setting is in the best interest of the student and what programming best meets their academic and social needs.

Functional Behavioral Assessment and Behavioral Intervention Plans

IDEA mandates that a functional behavioral assessment (FBA) be conducted for those students whose behaviors necessitate a change in school placement or represent a chronic pattern of misbehavior. Professionals use FBA to identify the purpose of a student's behavior (e.g., getting social approval from peers for inappropriate behavior and escaping a teacher's demand to complete work). FBA focuses on the function of the behavior (the why of the behavior) in relation to its environment or the context. When the function of a particular behavior problem is determined through a series of FBA steps, a behavioral intervention plan (BIP) is developed to assist a student in achieving the same or a similar goal (e.g., gain something or escape a demand) through more socially acceptable behaviors. Through FBA procedures, general and special education professionals can identify causes that may contribute to a student's inappropriate behaviors and design programs that can result in academic and social success.

Placement of Students with EBD

Nearly 70% of students with emotional disturbances spent more than 21% of their time outside regular education classes (US Department of Education, 2007). Nearly 47% of the students spent more than 60% of their day outside of regular education and some (16.9%) spent their entire day in a separate environment. Students with EBD are served in separate environments more often than any other categorical area except for deaf-blind and multiple disabilities (US Department of Education, 2007). The issue of where to best serve students with EBD has endured serious and long-term debate. The IDEA (2004) and other federal regulations mandate a continuum of alternative placements (CAPs), while some experts suggest that CAP segregates students from their more typical peers and access to the regular education curriculum. Kauffman and Smucker (1995) purported that specialized environments for students with EBD serve the following purposes:

1. protecting others (family, community, and schoolmates) from students' uncontrolled or intolerable behavior;
2. protecting students from themselves or others;
3. educating or training students in academics and other life skills and appropriate emotional responses, attitudes, and conduct;
4. educating or training children's families or teachers and peers to provide a more supportive environment;
5. keeping children available and amenable to therapies; and
6. providing opportunities for observation and assessment of children's behavior and its contexts.

There are, however, some special education advocates who disagree with Kauffman and Smucker (1995). Van Dyke *et al.* (1995), for instance, argued that all students with emotional and behavioral disorders are best served in the general education classroom. These advocates have noted that placing students in the general education classroom promotes a sense of belonging, provides opportunities to make friends, and more efficiently provides exposure to the regular education curriculum alongside nondisabled peers. They believe instead of providing separate environments, resources, and supports are better allocated for students with EBD to be successful in regular education classrooms. In addition, they argue that it prepares all students, with and without disabilities, to participate in a diverse world where interaction with people of differing abilities and behaviors is commonplace. In fact, this debate about the best placement model for students with EBD is ongoing. Professionals and advocates on both sides of the issue agree that children with disabilities should be educated in as typical an educational experience as possible. The discussion continues however, as to how this can best be accomplished.

Strategies to Prevent EBD

The prevention of problem behavior is an issue that receives much attention in politics and the national media. Traditionally, the procedures for identifying students with EBD are reactive and tend to occur after children's deviant behavioral repertoires are well established. Prevention of EBD is complicated by poor collaboration between service agencies, disjointed screening and eligibility criteria, and lack of funding for research and program implementation (Conroy *et al.*, 2004). This lack of coordination among agencies, such as schools, social welfare, and juvenile justice, is a fundamental obstacle toward prevention (Quinn and Poirier, 2004). Young children at risk for EBD are sometimes denied access to services due to eligibility restrictions and the lack of coordination is compounded by the fact that funding for prevention research is limited. Despite these barriers to

prevention, special education professionals and researchers are recognizing the value of prevention at the school, classroom, and individual student levels.

School-Wide Discipline

Media exposure and public perception of violence in schools have influenced administrators and policymakers to adopt rigid standards and a position of zero tolerance toward students who exhibit problem behavior in school (Quinn and Poirier, 2004). Researchers (e.g., Skiba, 2002) indicate that discipline practices that rely primarily on punishment and exclusion are ineffective and may actually contribute to an increase in problematic behavior. Recent prevention efforts, such as school-wide discipline strategies, use systematic processes to foster proactive solutions by implementing research based and validated practices. While there are several programs available to assist schools in the implementation of a unified school-wide supportive approach to discipline, many share similar elements such as:

1. a shared vision by the school staff on how to best prevent and minimize problematic behaviors based on evidence-based practices;
2. consistent and visible support by administrative leaders;
3. academic and social expectations are developed and implemented collaboratively by all staff; and
4. program change decisions are informed by evaluation of effectiveness data collected at the school (Colvin, 2007).

Prevention Strategies in the Classroom

Gresham (2002a) asserted that adopting an approach to prevention that begins with whole-class strategies and moves to addressing individual student needs may reduce over identification of children for special education. Witt *et al.* (2004) suggested the necessary components for prevention include solid academic instruction, explicit instruction of positive behavioral expectations to students, and consistent and effective teacher responses to inappropriate behavior. The emphasis should be to create a supportive classroom environment that promotes academic participation and achievement for all students. In addition to effective academic instruction, the promotion of positive behavioral expectations increases the length of academic engagement and decreases problem behaviors (Witt *et al.*, 2004). Positive behavioral expectations should be explicitly taught to students allowing adequate time for practice and feedback. Once students are aware of and can perform to expectations, careful monitoring of student behavior by teachers, coupled with the consistent delivery of effective contingent consequences, is central to maintaining behavioral compliance. Clearly, emphasis on proactive practice has led to procedures that identify environmental

contingencies that promote problem behaviors and the adaptation of those variables to reduce future misbehavior. A focus on class-wide effective academic instruction and proactive instruction of behavioral expectations are promising approaches for the prevention of behavioral problems. Sometimes, however, school-wide and classroom strategies are insufficient to address the needs of individual students who may require services that are more intensive.

Individual-Student-Centered Prevention Strategies

If school-wide and classroom attempts to promote desired behavior are ineffective, interventions based on a functional assessment of an individual student's behavior is the most effective evidence-based strategy currently available (Iwata *et al.*, 1993). Evidence in support of its use is so strong that the requirement to functionally assess chronic problem behaviors was added to the reauthorized IDEA in 1997. While functional assessment is used to identify the causes of behavior and to develop effective interventions for students with EBD, it can also be used as a strategy for prevention with individual students at risk for EBD. To a large measure, functional assessment examines the effects of specific environmental variables on a student's maladaptive behavior. Direct manipulation of these variables through intervention along with consistent monitoring of the effect on behavior reveals which contingencies may be supporting the enactment of the problem behavior. Interventions can then be adapted, based on the empirical evidence collected during the assessment. Once these interventions are implemented, progress is monitored and adaptations are made relevant to the performance of the individual as determined by ongoing data collection.

Intervening to Reduce Extreme Problematic Behaviors

While prevention efforts at school, classroom, and individual levels are foundational for effective management of student behavior, students who exhibit EBD may still require extensive behavioral programming to mollify the effects of their behavioral excesses and deficits. Reducing maladaptive behaviors and increasing appropriate ones can be achieved through the use of behavioral principles, such as reinforcement and punishment procedures, and appropriate replacement behaviors can be taught to students using social skill instruction, self-management techniques, and through the use of cognitive-behavioral interventions (CBIs).

Using Behavioral Principles

Teachers and service providers typically use behavioral strategies such as reinforcement to increase positive and

desirable behaviors and punishment to reduce unwanted and maladaptive behaviors. Positive reinforcement is a reward that is positive, or pleasing to a student presented after an appropriate behavior. Reinforcers can be of four types:

1. tangibles (e.g., toys, school supplies, posters, and magazines);
2. activities contingent on acceptable behavior (e.g., playing board games and listening to music);
3. social reinforcement (e.g., high five for work completion and verbal praise); and
4. token reinforcement (e.g., points to reach an established goal or imitation money exchangeable for some valued object or activity).

Conversely, negative reinforcement is the removal of an aversive condition and can also serve as a form of reinforcement to increase desired behavior. For example, a teacher may say to a student who is off task that if the work is not finished on time, he/she will have to finish it in detention. If the student completes the assignment on time, then he/she was negatively reinforced because he/she finished the work to avoid detention.

Punishment is used to reduce the future occurrence of problematic behaviors. Teachers and service providers may use reprimands, response cost, time out, and overcorrection. A reprimand is a short verbal scolding or correction that is designed to reduce inappropriate behavior. Response cost would be the removal of something the student has earned, such as points or privileges, when inappropriate behavior is displayed. Time out is usually associated with removing a student from a reinforcing activity or environment. Sending a student back to his/her desk as a result of inappropriate behavior during group work, sending a student into the hallway or to another room, or sending a student to the principal's office can be considered time out from reinforcement. Overcorrection typically happens when a teacher has a student engage in repetitive action as a penalty for inappropriate behavior. Although overcorrection has many forms, examples would include asking an unruly student to practice getting in line calmly and orderly a few times, and directing the pick up of refuse after a student kicks a trash can with additional cleanup of the rest of the floor. Using punishment, however, should always be a teacher's last resort to decrease unwanted behaviors from happening in the classroom. There are more positive techniques than punishment to help students decrease socially undesirable behaviors with more socially appropriate ones.

Teaching Appropriate Replacement Behaviors

Not only do teachers have to reduce problem behaviors of students with EBD, but they also need to explicitly teach more appropriate and positive replacement behaviors that

can lead to more successful long-term outcomes. Teaching replacement behaviors can be accomplished through social skills training, CBIs, and teaching self-control techniques to students.

Social skills training

There are a number of reasons to teach social skills to students who exhibit EBD. Social skills are arguably the most important skills students learn, not only in school, but also as they grow into adulthood. The ability to get along successfully with other people is a criterion for success for students with EBD whether at school, in the work world, family, or community. Before social skill instruction, assessment of a student's social strengths and weaknesses must occur. Assessment can help identify individual student needs, monitor progress, and evaluate the viability of a teaching strategy. Behavior rating scales or peer rating scales might be useful to assess student social skills. A teacher can identify a student's social strengths and determine areas that need some instructional intervention by combining direct observation of a student's social behavior and functional assessment of behavior (Gresham, 2002b).

Typically, social skills are taught by defining and modeling the skill to be learned, setting up role-plays for students to learn the skill steps, giving performance feedback, and, most importantly, providing opportunities for practice (see Goldstein and McGinnis, 1997). The teacher can also provide homework so that learned skills can be practiced and generalized to situations outside of the learning setting. Students have generalized or transferred their learning when they reliably perform the social skill in new and novel places, with a variety of people and situations, and with variations of the learned skill.

Cognitive-behavioral interventions

CBIs are designed to teach the use of self-talk or inner speech (verbal self-regulation) to regulate overt behavior. Simply, verbal self-regulation is talking to oneself to guide problem solving or some other behavior. CBIs are often described as stop-and-think strategies. Problem solving, one type of CBI, is a useful cognitive process that can help students with EBD analyze in a better manner their social encounters during the day and to exercise self-control. Problem solving consists of a series of mental steps designed to identify and define problems, generate a menu of appropriate responses or solutions to solve the problem, select an efficient and effective solution, and make a plan to carry it out successfully (Smith and Daunic, 2006).

Self-control

As students become older and enter into the world of work, the ability to successfully regulate or self-manage their own emotions and behavior becomes critical. Self-regulation begins when students take responsibility for their own behavior and learn to use internalizing or

cognitive processes that will enable success when external systems, such as reinforcement and punishment, are unavailable. Teaching self-regulation represents a critical step in reducing or fading a teacher's participation in controlling a student's behavior. Teachers and service providers can help students with EBD learn self-management strategies and use CBIs to increase self-regulatory functions (see Polsgrove and Smith, 2004). Self-regulation or self-management strategies include self-assessment of problematic behaviors and the necessary replacement behavior, self-monitoring of those behaviors, and self-reinforcement when the student experiences success. Self-management strategies are often recommended for students who are manipulative or oppositional when confronted and for facilitating the success of students who are mainstreamed into less-restrictive placements.

Special Issues: Looking at the Future

Providing a quality educational experience for students with EBD is and will continue to be problematic; however, despite the immense gains in providing successful programming for these students, there are complex issues that professionals still confront. Students with EBD exhibit such extreme behavioral deficits and excesses that issues related to suspension and expulsion and placement in alternative school settings differentially affect this population. Of special note is that minority students are disproportionately represented in EBD programs.

Suspension and Expulsion

Disciplining students with EBD who violate acceptable standards of student behavior is (and will continue to be) complex and often a contentious issue. Since students with disabilities are guaranteed a free and appropriate public education (FAPE), as outlined in case law and in IDEA (2004), long-term suspension (more than ten consecutive school days) or expulsion of students (removal from a current educational setting) with EBD can violate a student's right to FAPE. Any long-term suspension for violating acceptable standards of behavior or expulsion for egregious violation of the code of student conduct (e.g., weapons, drugs, and serious bodily injury) is considered a change of placement and can occur only after a meeting of the student's child study or IEP team. At that time, the team is to make a determination about the relationship between the student's misconduct and disability. Since students with EBD are served in special education because of their significant behavior problems, determining whether or not the misconduct is related to the disability is sometimes uncertain and confusing.

Placement in Alternative Educational Settings

Alternative educational settings are designed for students who are at risk for school failure. Alternative schools typically are small, personalized settings that emphasize individualized educational, vocational, and general living skills. Researchers have found that about 12% of students who are attending alternative schools are students with disabilities, mostly students with LD or EBD (Lehr, 2004). Students with EBD may be placed in alternative settings after a disciplinary removal or expulsion for violating the code of student conduct (e.g., weapons, drugs, and serious bodily injury). This placement option can only be considered when all procedural safeguards are followed as outlined in IDEA.

Overrepresentation of Minority Students in EBD Programs

Using data from the US Department of Education (2006), analyses suggest that minority students are disproportionately placed in EBD programs. For instance, black children are 1.92 times more likely than white students to be labeled EBD (Losen and Orfield, 2002). These researchers suggest that stereotypes, racial bias in implementing discipline policies, and education practices that are culturally non-responsive may contribute to the overidentification and placement of many minority students. Minority students who exhibit EBD are often subject to low expectations, are educated in separate settings, and excluded from many educational opportunities. While minority populations are at great risk of living in poverty, Losen and Orfield refute the myth that racial overrepresentation in special education can be explained away by factors associated with poverty. This undoubtedly will continue to influence future educational programming for students in EBD programs.

Conclusion

Educating students with EBD is a complex endeavor that requires careful consideration and deliberate action. Those professionals who are dedicated to serving children and youth with EBD are constantly trying to identify and develop strategies and techniques that will help this population of students maximize their potential in school and beyond. There are ongoing efforts in the field of special education to define and assess in a better manner the disorder to intervene effectively and efficiently on behalf of students. There is also an emerging technology to prevent student placement in EBD programs and, when required, there are powerful techniques to teach replacement behaviors for those students who exhibit antisocial and maladaptive behaviors that preclude school success.

Notwithstanding the tremendous progress that has been made over the past few decades, some issues remain, such as

where best to serve students with EBD, policies about suspension and expulsion, placement in alternative educational settings, and the overrepresentation of students from minority racial and ethnic backgrounds. Despite these unresolved problems, education professionals continue to work tirelessly in an effort to serve this unique population.

Bibliography

- Bierman, K. L., Greenberg, M. T., and CPMRG (1996). Social skills training in the fast track program. In Peters, R. D. and McMahon, R. J. (eds.) *Preventing Childhood Disorders, Substance Abuse, and Delinquency: Banff International Behavioral Science Series*, vol. 3, pp 65–89. Thousand Oaks, CA: Sage.
- Blake, C. and Monahan, E. C. (2007). Rethinking teacher preparation for EBD students: Toward a partnership model. *Support for Learning* 22(2), 60–65.
- Bos, C. S., Coleman, M., and Vaughn, S. (2002). Reading and students with E/ED: What do we know and recommend. In Lane, K. L., Gresham, F. M., and O'Shaughnessy, T. E. (eds.) *Interventions for Children with or at Risk for Emotional and Behavioral Disorders*, pp 87–103. Boston, MA: Allyn and Bacon.
- Cheney, D., Flower, A., and Templeton, T. (2008). Applying response to intervention metrics in the social domain for students at risk of developing emotional or behavioral disorders. *Journal of Special Education* 42(2), 108–126.
- Colvin, G. T. (2007). *7 Steps for Developing a Proactive Schoolwide Discipline Plan*. Thousand Oaks, CA: Corwin Press.
- Conroy, M. A., Hendrickson, J. M., and Hester, P. P. (2004). Early identification and prevention of emotional and behavioral disorders. In Rutherford, R. B., Quinn, M. M., and Mathur, S. R. (eds.) *Handbook of Research in Emotional and Behavioral Disorders*, pp 199–215. New York: Guilford Press.
- Coyne, S. M., Archer, J., and Eslea, M. (2006). 'We're not friends anymore! unless . . .': The frequency and harmfulness of indirect, relational, and social aggression. *Aggressive Behavior* 32(4), 294–307.
- Elliott, J. (2008). Response to intervention: What & why? *School Administrator* 65(8), 10–18.
- Fessler, M. A., Rosenberg, M. S., and Rosenberg, L. A. (1991). Concomitant learning disabilities and learning problems among students with behavioral/emotional disorders. *Behavioral Disorders* 16, 97–106.
- Forness, S. R. and Knitzer, J. (1992). A new proposed definition and terminology to replace 'serious emotional disturbance' in individuals with disabilities education act. *School Psychology Review* 21, 12–20.
- Goldstein, A. P. and McGinnis, E. (1997). *Skillstreaming the Adolescent: New Strategies and Perspectives for Teaching Prosocial Skills*. Champaign, IL: Research Press.
- Gresham, F. M. (2002a). Responsiveness to intervention: An alternative approach to the identification of learning disabilities. In Bradley, R., Danielson, L., and Hallahan, D. (eds.) *Identification of Learning Disabilities: Research to Practice*, pp 467–519. Mahwah, NJ: Erlbaum.
- Gresham, F. M. (2002b). Best practices in social skills training. In Thomas, A. and Grimes, J. (eds.) *Best Practices in School Psychology*, 4th edn., pp 1029–1040. Bethesda, MD: National Association of School Psychologists.
- Hampton, S. S. and Hess-Rice, E. K. (2003). Restructuring service delivery for students with emotional and behavioral disorders. In Obiakor, F. E., Utley, C. A., and Rotatori, A. F. (eds.) *Effective Education for Learners with Exceptionalities*, pp 119–138. Bingley: Emerald Group.
- Hawken, L. S., Vincent, C. G., and Schumann, J. (2008). Response to intervention for social behavior. *Journal of Emotional and Behavioral Disorders* 16(4), 213–225.
- Individuals with Disabilities Education Improvement Act of 2004 (IDEA) (2004). Pub. L. No. 108–446, 118 Stat 2647–2808.
- Iwata, B. A., Vollmer, T. R., Zarcone, J. B., and Rodgers, T. A. (1993). Treatment classification and selection based on behavioral function. In Van Houten, R. and Axelrod, S. (eds.) *Behavior Analysis and Treatment*, pp 101–125. New York: Plenum.
- Kauffman, J. M. (2005). *Characteristics of Emotional and Behavioral Disorders of Children and Youth*, 8th edn. Upper Saddle River, NJ: Prentice-Hall.
- Kauffman, J. M. and Smucker, K. (1995). The legacies of placement: A brief history of placement options and issues with commentary on their evolution. In Kauffman, J. M., Lloyd, J. W., Hallahan, D. P., and Astuto, T. A. (eds.) *Issues in Educational Placement: Students with Emotional and Behavioral Disorders*, pp 21–44. Hillsdale, NJ: Erlbaum.
- Knoff, H. M. (1995). Best practices in personality assessment. In Thomas, A. and Grimes, J. (eds.) *Best Practices in School Psychology*, pp 849–864. Washington, DC: National Association of School Psychologists.
- Lehr, C. (2004). *Alternative Schools and Students with Disabilities: Identifying and Understanding the Issues*. Minneapolis, MN: University of Minnesota, Institute on Community Integration.
- Losen, D. J. and Orfield, G. (2002). *Racial Inequity in Special Education*. Cambridge, MA: Harvard Education Press.
- Polsgrove, L. and Smith, S. W. (2004). Informed practice in teaching self-control to children with emotional and behavioral disorders. In Rutherford, R. B., Quinn, M. M., and Mathur, S. R. (eds.) *Handbook of Research in Emotional and Behavioral Disorders*, pp 399–425. New York: Guilford Press.
- Quinn, M. M. and Poirier, J. M. (2004). Linking prevention research with policy: Examining the costs and outcomes of the failure to prevent emotional and behavioral disorders. In Rutherford, R. B., Quinn, M. M., and Mathur, S. R. (eds.) *Handbook of Research in Emotional and Behavioral Disorders*, pp 78–97. New York: Guilford Press.
- Rutter, M. (1985). Family and social influences on behavioural development. *Journal of Child Psychology and Psychiatry* 26, 349–368.
- Skiba, R. J. (2002). Special education and school discipline: A precarious balance. *Behavioral Disorders* 27(2), 81–97.
- Smith, S. W. and Daunic, A. P. (2006). *Managing Difficult Behavior through Problem Solving Instruction: Strategies for the Elementary Classroom*. Boston, MA: Allyn and Bacon.
- US Department of Education (2006). Assistance to the states for the education of children with disabilities and preschool grants for children with disabilities: Final rule. *Federal Register* 71(156), CFR 300.8, p. 46756.
- US Department of Education (2007). *Twenty Seventh Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act*. Washington, DC: US Department of Education.
- Van Dyke, R., Stallings, M. A., and Colley, K. (1995). How to build an inclusive school community: A success story. *Phi Delta Kappan* 76, 475–479.
- Witt, J. C., VanDerHeyden, A. M., and Gilbertson, D. (2004). Instruction and classroom management. In Rutherford, R. B., Quinn, M. M., and Mathur, S. R. (eds.) *Handbook of Research in Emotional and Behavioral Disorders*, pp 426–445. New York: Guilford Press.

Further Reading

- Daunic, A. P., Smith, S. W., Brank, E. M., and Penfield, R. D. (2006). Classroom based cognitive-behavioral intervention to prevent aggression: Efficacy and social validity. *Journal of School Psychology* 44, 123–139.
- Downing, J. A. (2007). *Students with Emotional and Behavioral Problems: Assessment, Management, and Intervention Strategies*. Upper Saddle River, NJ: Pearson.
- Fuchs, D. and Fuchs, L. (1994). Inclusive schools movement and the radicalization of special education reform. *Exceptional Children* 60, 294–309.
- Kauffman, J. M. (1999). How we prevent the prevention of emotional and behavioral disorders. *Exceptional Children* 65, 448–468.
- Kauffman, J. M. and Landrum, T. J. (2006). *Children and Youth with Emotional and Behavioral Disorders: A History of Their Education*. Austin, TX: Pro-Ed.

- Landrum, T. J., Tankersley, M., and Kauffman, J. M. (2003). What is special about special education for students with emotional or behavioral disorders? *Journal of Special Education* **37**(3), 148–156.
- Marsh, D. T. and Fristad, M. A. (eds.) (2002). *Handbook of Serious Emotional Disturbance in Children and Adolescents*. New York: Wiley.
- Mooney, P., Epstein, M. H., Reid, R., and Nelson, J. R. (2003). Status of and trends in academic intervention research for students with emotional disturbance. *Remedial and Special Education* **24**(5), 273–287.
- Nelson, J. R., Benner, G. J., Lane, K., and Smith, B. W. (2004). Academic achievement of K-12 students with emotional and behavioral disorders. *Exceptional Children* **71**, 59–73.
- Robinson, T. R., Smith, S. W., Miller, M. D., and Brownell, M. T. (1999). Cognitive behavior modification of hyperactivity–impulsivity and aggression: A meta-analysis of school-based studies. *Journal of Educational Psychology* **91**, 195–203.
- Smith, S. W., Lochman, J. E., and Daunic, A. P. (2005). Managing aggression using cognitive–behavioral interventions: State of the practice and future directions. *Behavioral Disorders* **30**, 227–240.

Educating Students with Special Needs: An Overview

F E Obiakor, M K Harris, and P Harris-Obiakor, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
G M Mukuria, Slippery Rock University, Slippery Rock, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Educating Students with Special Needs: An Overview

From time immemorial, individuals with special needs have lived in our midst. They have existed, functioned, and participated in our societal functions in one way or another. For instance, the Jewish Talmud, Moslem Koran, and Christian Bible made particular references to persons who were atypical in nature. In fact, all societies have continued to focus on how to take care of their less fortunate, less powerful, disenfranchised, and disadvantaged (Obiakor and Algozzine, 1995). From these contexts, it appears that efforts have concentrated on how to tolerate and/or be nice to persons with special needs rather than on how to truly educate them with real pedagogical power. Hillard (1992) explained that real pedagogical power means that “all children who may have disabilities receive sophisticated, valid services that cause them to do better than they would have done if they had not received special services at all” (p. 168).

In the late eighteenth century, Jean Marc Itard, a French physician believed in real pedagogical power when he decided to take on the task of educating Victor, the wild boy of Aveyron, France. Even this wild boy was able to acquire some skills, an indication that special education works. In the early parts of the twentieth century, Dr. Alfred Binet, the brain behind the current Stanford-Binet Intelligence Scale, noted that human knowledge and/or intelligence can be improved. Binet (1909) warned against the over reliance in his intelligence quotient tool. In speaking about his special class, he remarked, “It is parochial sense, the only one accessible to us, that we say that intelligence of these children has been increased. We have increased what constitutes the intelligence of a pupil: the capacity to learn and to assimilate instruction” (p.104). More than two decades ago, Gould (1987) decried the mismeasure of people and argued that “if Binet’s principles had been followed, and his tests consistently used as he intended, we would have been spared a major misuse of science in our century” (p. 155). In addition, he warned against the blind following of the theory of biological determinism (i.e., the belief that human attributes are only genetically based) because it hampers human valuing and the ability to engage in real pedagogical power. Goodlad (1993) corroborated Gould’s premise and noted that:

We appear incapable of getting beyond individuals as the units of assessment with the accompanying allocation of

responsibility for success and failure. We must adopt as standard practice the kind of contextual appraisal that tells whether schools have in place the curriculum, materials, pedagogy, and other conditions necessary to the good education of individuals. The absence of these exposes and brings inequities that are the moral responsibility of a caring people in a just society to correct. (p. 20)

We strongly believe students with special needs can and must be educated. The entries in this section reveal that efforts have been made and continue to be made in this regard. As scholars and practitioners, our broad goal must be to provide education that is different from, additional to, and supplementary with those provided in the regular classroom with a systematic modification and adaptation of instructional techniques, materials, and equipment (Obiakor *et al.*, 2003a). To achieve this broad goal, we must know who these students with special needs are. Not surprisingly, these students fit into broadly or globally recognized categories of exceptionalities, namely: cognitive disabilities, learning disabilities, emotional/behavioral disorders, communication disorders/speech and language impairments, visual impairment/blindness, hearing impairment/deafness, autism, traumatic brain injury, gifted and talented, physical disabilities, and other health impairments. While some of these categories are fewer in some countries, every country wants to demonstrate its interest in helping students with special needs to maximize their potential. Clearly, the entries in this section reflect dynamic special education methods and techniques that mirror our changing times in our complex world. In addition, through their entries, the contributors have demonstrated their multidimensional interests and exceptional knowledge in the world of special education. To a large measure, they have shown global and specific characteristics of the field of special education.

Historical Impacts and Contexts

It is common knowledge that people like to associate with those who behave, look, speak, and act like themselves. Anyone who does not fall in that norm is perceived and treated differently (Obiakor, 2008, 2009). Clearly, students with special needs have been discriminated against, ostracized, labeled, and called demeaning names (e.g., stupid,

imbecile, and retarded). Today, it has become increasingly clear that differences are a part of life. Advocates of students with special needs have pressed for ways to positively respond to their needs in quantifiable ways (Obiakor *et al.*, in press). In the United States, it is impossible to divorce the education of students with special needs from the Civil Rights Movement and the subsequent events that followed. To a great extent, the education of these students has been greatly influenced by social developments and court decisions in the 1950s and 1960s. For example, the landmark *Brown v. Board of Education of Topeka* (1954) case was a civil rights case that declared separate education as unequal education and unconstitutional (Obiakor, 2009). This was significant because it had the goal of ending racial segregation in schools. Logically, this opened doors of advocacy for students with special needs. The ruling of this case became a catalyst that prompted parents and professionals to lobby for equitable education for their students.

The *Brown* ruling, in the United States, encouraged parent groups to petition the courts to allow students with special needs to be educated in public schools. The *Pennsylvania Association for Retarded Children v. Commonwealth of Pennsylvania* (1972) (PARC) case held that children could not be denied access to public schools, and entitled them to a free and appropriate public education (FAPE). In the *Mills v. Board of Education* (Mills, 1972) case, a class action lawsuit was filed on behalf of 18 000 children with varied exceptionalities in the Washington, DC schools. In this case, the court ordered the district to educate all children, including those with special needs and further clarified that specific procedures had to be followed to determine whether a student should receive special services. Generally, these cases and other subsequent cases formed the framework for the legislation that currently guides the field (Yell, 2004).

In the United States, the Civil Rights Act (1964) (PL 88-352) provided legal rights to equality in education and other sectors of human interactions. In 1973, Section 504 of the Vocational Rehabilitation Act (PL 93-112) was passed to provide persons with special needs with (1) free and appropriate public education, (2) civil rights, (3) accessibility of programs, and (4) employability rights. The Education of All Handicapped Children's Act (1975) (PL-94-142) was passed with the following fundamental ingredients: (1) education for students from 3 to 21 years, (2) free and appropriate public education, (3) identification of students, (4) nondiscriminatory assessments, (5) placement in the least restrictive environment (LRE), (6) confidentiality of information, (7) procedural safeguards, and (8) development of individualized education plans (IEP). In 1986, PL 94-142 was amended to accommodate young children from birth to 3 years. This law Education of All Handicapped Children's Act (1986)(PL 99-457) was enacted to provide not just IEPs for children but also individual family support programs (IFSPs) for parents and guardians. In 1990,

PL-94-142 was renamed as the Individuals with Disabilities Education Act (IDEA, 1990, PL101-476). This Act involved funding for states to provide educational services to students from birth through 21 years, and ensured procedural safeguards for parents that guarantee meaningful participation in the evaluation process (Katsiyannis *et al.*, 2001). Additionally, IDEA guaranteed improvement in the education of students with special needs through research training and technical support and transitional supports for students when they are 16 years of age. To challenge the private sectors, the 1990 Americans with Disabilities Act (ADA, 1990; PL101-336) was passed to provide more societal opportunities for persons with special needs. In 1997, IDEA was reauthorized as PL 105-17 (IDEA, 1997) to facilitate disciplinary procedures and reduce litigation costs. In 2001, the No Child Left Behind Act (NCLB, 2001; PL 107-110) was passed to educate all learners and quantifiably account for their progress at all levels. Later, in 2004, IDEA was again reauthorized as the Individuals with Disabilities Education Improvement Act (IDEIA, 2004; PL 108-446). This law mandated that teachers of students with special needs be highly qualified, meaning they must be certified in the content areas that they are teaching (Smith, 2005).

Staying the Course, Resisting Change, or Moving Forward

The goal of any educational program is to maximize the fullest potential of students. In other words, the goal must be to truly leave no child behind despite his or her ability or disability. The critical question remains: Do we stay the course, resist change, or move forward in the education of students with special needs? Without a doubt, the dream is to move forward! More than a decade ago, Schrag (1993) confirmed that "the proportion of students being served within special education programs today and in the future is changing, which requires closer integration and coordination of services within the educational system and with a broader array of health and social services" (p. 208). The response of the federal government with regard to these imperatives has been accountability without accountability (i.e., accountability that focuses narrowly on the exclusion of students via assessment). Sadly, some accountability measures are already hurting the spirit of special education. In his piece titled, 'The death of special education,' Lieberman (2001) remarked that:

Special education has been swallowed by the beast: the school system, with its mandated curriculum, mandated tests, and mandated standards. Now, children with disabilities are entitled – no, are practically required – to have the same education as every other child, regardless of whether or not that education is of high quality or is appropriate for a child with a disability. (p. 39)

While it is iconoclastic to believe special education has been swallowed by the beast because of accountability challenges that are forced upon it, it is equally unrealistic to assume that we should just stay the course in special education. Any field or profession that does not believe in positive change is dead. Clearly, recent demographic changes in our society have challenged general and special educators and leaders to look for innovative ways to maximize all students' potential in school programs (Obiakor, 2007; Rueda, 2007). As Rueda argued, "given the long-standing but continuing controversy over the issue of overrepresentation of diverse students in special education, the future implications for identification, referral, assessment, and instructions are abundant" (p. 292).

Of late, there have been some aggressive moves to silence critics of the current system of special education that overrepresents culturally diverse students (e.g., African American learners) in programs for children with emotional/behavioral disorders and underrepresents them in programs for students with gifts and talents. Kauffman (2002, 2003a, 2003b, 2004), Mostert *et al.* (2003), and Sasso (2003) agreed that it is wrong to criticize the current system of special education. In fact, in their works, they were less receptive on the issue of the disproportionate representation of culturally and linguistically diverse students in special education. For instance, Kauffman (2003b) argued that:

The assumption that special education, which is at its best the fair treatment of disability, *creates* stigma is not just wrong; it is perverse. It confuses treatment with cause, just as if we were to make the assumption that identifying and treating cancer caused the stigma that used to accompany having it. Without willingness to talk about disabilities in a simple and straightforward way, we cannot address the problem of stigma. Euphemisms are cloaks that hide nothing effectively. Always and inevitably, they are stumbling rags that trip up prevention. (p. 196)

There is no doubt that the current special education system works for some children. However, the question is: Do critics of the current system of special education believe in the spirit of special education? Sure, they do! The reality is that special education has become an important educational phenomenon that works well when it does not misidentify, misassess, miscategorize, misplace, and misinstruct students who are racially, culturally, linguistically, and socioeconomically different. It seems unprofessional and immoral to hide under the cloak of special education to get rid of students just because they exhibit different behavioral and learning styles. Again, while there is great need for evidence-based practice in special education, we strongly disagree with Kauffman's (2003a) assertion that "if you discount science as a way of finding things out and believe that special education is fundamentally flawed, second rate, ineffective, unfair, and oppressive, then you're not going to use it for prevention" (p. 206).

We believe science is necessary; however, the indiscriminate use of a scientifically proven medication to cure all illnesses is dangerous, unethical, and immoral (Obiakor, 2004). The heart or respect for humanity must be incorporated into whatever we do as professionals even though one's heart or spirituality cannot be measured. Science may not always be the only answer; feelings matter too! Even in the medical field, the touch of the doctor and the feelings of the patient can facilitate and advance the healing process. Why should special education be any different?

In their study titled, Do race of student and race of teacher influence ratings of emotional and behavioral problem characteristics of students with emotional disturbance?, Cullinan and Kauffman (2005) concluded that "results did not support the position that, among students with ED [emotional disturbance], overrepresentation of African Americans arises from racial bias in teacher perceptions of emotional and behavioral problems" (p. 393). Coupled with the study's limitations and weaknesses as identified by Cullinan and Kauffman, there is the presumption of innocence of teachers just because of their race or culture. In many urban schools in the United States, there are culturally and linguistically diverse professionals who through their actions have devastated the lives of students and their parents (Obiakor, 2001b, 2003). Also in the United States, there are many Black policemen or women who wrongfully arrest, brutalize, shoot, and kill fellow Blacks in strange attempts to maintain law and order. Their race or culture must never be an alibi that exonerates them from being criticized or sued for violating the civil rights of others (Prater, 2006). Clearly, on issues of misidentification, misassessment, miscategorization, misplacement, and misinstruction of students, a poorly prepared general and special educator will not advance the education of all students (Obiakor, 1999, 2001b, 2003, 2004, 2007, 2008, 2009; Obiakor and Beachum, 2005; Obiakor and Ford, 2002; Obiakor *et al.*, 2002; Utley and Obiakor, 2001).

We believe we must move forward to advance creative strategies for educating all students, their special needs notwithstanding. As it appears, we cannot help all students unless we tap on the energies of all stakeholders (Obiakor, 2003, 2007, 2008, 2009; Obiakor *et al.*, 2002). Though there are no magic solutions, we must explore the multidimensional nature of the Comprehensive Support Model (CSM) to tap into the contributions of individual students, families, schools, communities, and governments. The individual self is important because without personal responsibility or self-improvement, it will be difficult to manage behavior and learning problems. The family is important because it is the cornerstone of the special student and the bridge that connects the student with the school. The school is important because it has general and special education teachers

and professionals who have the power to shift their paradigms regarding demographic changes. The community is important because it provides a variety of opportunities and choices for our children and youth who have been labeled as trouble makers or problem students because of their different styles. Finally, the local, state, and federal governments are important because they generate equitable policies that strengthen the multiple voices of all students. Evidently, a responsible government must be worried about the civil rights of its people, even those of people who have special needs or exhibit nonproductive and antisocial behaviors (Obiakor *et al.*, 2002).

To move forward in the whole process of special education for all students, it is imperative that general and special educators and leaders:

- Develop and use identification, assessment, and instructional strategies that function within the context of cultural competence.
- Create a collaborative system of community support that focuses on eradicating social stereotyping based on race, ethnicity, national origin, gender, and socioeconomic status.
- Develop an awareness and appreciation for the many family forms that value individual differences and strengths.
- Thwart conditions that lead to violence in the home or community and cultivate a sense of safety for children and families.
- Advocate economic policies and human services that are pro-family by virtue of proven outcomes.
- Promote culturally competent practices in schools and in the larger society to respect differences in worldviews and learning styles among individuals.
- Advocate expanded services that provide for affordable quality childcare to meet the varied needs of all families and children.
- Develop collaborative community approaches to problem solving that involve students, parents, schools, and community leaders.
- Recognize that the focus of the problem in at-risk situations is not only in the individual but also in institutional barriers in the environment.
- Reconfigure curricula that incorporate culturally sensitive variables.
- Reestablish rites of passage and service opportunities that cultivate a sense of belonging and resiliency in youth.
- Broaden visions in educational reform that include economic reform and the investment in human capital.

Apparently, by moving forward in this age of change, general and special educators can assist all students in school programs. For instance, they can prevent and manage violent behaviors that have created psychological setbacks for students by shifting their own personal paradigms.

In contrast to the get-tough no-nonsense approaches (e.g., zero tolerance or three-strikes-you-are-out disciplinary models), school personnel can teach prosocial skills and educate children to manage interpersonal conflicts nonviolently (Goldstein, 1999; Long, 1997; Obiakor, 2001a). After a lifetime of experience with youth with emotional/behavioral disorders, Long simply suggested using kindness or what we call the heart. General and special educators employing caring transitional strategies must focus on a variety of communication skills that enable young people to manage their behaviors and respond to others in ways that do not provoke confrontations. They must revisit the traditional emphasis on intelligence or academic achievement that seems to downplay the emotional intelligence and resiliency needed to survive in a changing society (Gardner, 1993; Goleman, 1995; Obiakor *et al.*, 2004, 1997). As Goleman remarked, emotional intelligence entails “abilities such as being able to motivate oneself and persist in the face of frustrations; to control impulse and delay gratification; to regulate one’s moods and keep distress from swamping the ability to think; to empathize and to hope” (p. 34). He added:

Academic intelligence offers virtually no preparation for the turmoil – or opportunity – life’s vicissitudes bring. Yet even though a high IQ is no guarantee of prosperity, prestige, or happiness in life, our schools and our culture fixate on academic abilities, ignoring emotional intelligence, a set of traits – some might call it character – that also matters immensely for our personal destiny. (p. 36)

Some proactive measures have been found to foster emotional intelligence in all learners, including those with special needs! These measures incorporate partnership programs, prosocial skills instruction programs, and mentorship programs. How can students value differences if differences are not valued in their homes, schools, and communities? How can they work together if adults and communities fail to work together? People who have emotional intelligence skills can help dissipate some of the cultural forms of heartlessness that permeate schools, for example, putdowns based upon race, ethnicity, gender, or disability. Students must work together, their families must cooperate with each other, their schools must work collaboratively, and their communities must work together (Obiakor *et al.*, 2002, 2004). These collaborative and consultative behaviors frequently lead to cooperative resolutions of situations at all levels and help students, parents, and professionals to maximize their potential.

Implications for Teacher Preparation

Clearly, all students exhibit special and different learning and behavioral patterns. As a consequence, they are intentionally or unintentionally misidentified, misassessed,

miscategorized, misplaced, and misinstructed in school programs (Mukuria and Obiakor, 2004; Obiakor, 1999, 2001b, 2003a, 2003b, 2007; Obiakor and Beachum, 2005; Obiakor and Wilder, 2003; Utley and Obiakor, 2001). What then are the roles of teacher-preparation programs for learners who are different? Even with the best intentions, many colleges and universities have failed to satisfactorily prepare educators for today's classrooms. Earlier, Haberman (1995) asserted that upon completion of traditional teaching programs teachers and service providers are as prepared for today's classrooms as a swimmer who prepared for the English Channel by training in the university swimming pool. It is important that teacher educators and leaders take the bull by its horns! They must be professionally responsible – they must prepare general and special educators and leaders to respond to demographic changes. They must shift their own paradigms to prepare teachers and leaders who can shift their paradigms (Smith *et al.*, 2001; Winzer and Mazurek, 1998). For those engaged in research, they must broaden their horizons in their understanding of nature versus nurture and other human behaviors and attributes. As scholars, they must go beyond the archaic theory of biological determinism and the myth of socioeconomic dissonance to make sense of their research (Gould, 1981; Weikart, 1977). For instance, Weikart warned that the deficit model of thinking, when applied to a certain population, “seems to limit potential assistance to that group because it channels thinking in ways that emphasize weaknesses rather than strengths, and it interprets differences from the norm as individual deficits” (p 175). The logical extension is that:

We cannot limit ourselves to the identification of trait dimensions or typological classifications across individuals without also considering the characteristics of the environments within which individuals function. Nor can we limit ourselves to an analysis of the environmental determinants of human differences without also considering the hereditary determinants. Finally, we have to ask ourselves what kind of society is most desirable for the expression of human diversity – for the opportunity for each of us to grow as individuals and at the same time not infringe on the rights of others to develop their own individuality. (Minton and Schneider, 1985: 489)

Minton and Schneider's (1985) statements have far-reaching implications for research, policy, and practice in general and special education and leadership. For instance, first, research that focuses on behavior problems of children and youth needs to address measures that will help us to understand them. When we understand them, we assist them to be functional, goal-directed decision makers in our complex society. Put another way, research that focuses on underlying pathological attributes of students needs to be valued with caution because such a research is deficit-oriented and lacks measurable or

observable solution-based attributes. Second, research, policy, and practice ought to go hand in glove. Many years ago, Keogh (1990) noted that “from this perspective, policy should follow research, and change should be found in evidence” (p. 186). It is apparent that something is wrong with our intervention strategies for culturally diverse students with special needs. Third, research that divorces itself from the fundamental principles of individualized instructional programming fails to appreciate or value individual differences in people. In the end, we need redirection in research funding and projects to reflect culturally sensitive proactive measures. Research studies with skewed divisive, emotionally loaded, political underpinnings must be discouraged in special education. Any research that does not lend itself to common-sense problem-solving interpretation and practice must be viewed with caution. Fortunately today, most scholarly publications (e.g., *Exceptional Children*, *Journal of Special Education*, *Multicultural Learning and Teaching*, *Multiple Voices*, and *Teacher Education and Special Education*) are demanding practical implications to works. Consequently, scholars, educators, and leaders must begin to broaden their definitions, theories, and intervention models to reduce illusory conclusions, perceptual assumptions, and prejudicial generalizations (Obiakor, 2007). Special students need specialized training of professionals. Strategies that empower all students must be developed if education is to get its desired respect. For instance, teacher educators and leaders must begin to realign themselves with new ways of thinking that go beyond games and politics.

We are convinced that poorly prepared teachers teach poorly. It is important that teacher educators and leaders practice what they preach. In this age of change, they must use divergent techniques to prepare future educators who will, in turn, use divergent techniques to teach learners who exhibit different styles and express special needs. To look for the magic pill that can cure educational problems of all students is not realistic. However, the key is for teacher educators and leaders to prepare those who value individual differences and special needs (Ford *et al.*, 1995; Obiakor, 2001b, 2003, 2007; Obiakor and Beachum, 2005; Obiakor and Ford, 2002; Obiakor *et al.*, 1999, 2003b; Wilder *et al.*, 2003). By so doing, they become aware of emotional first-aids needed to address crises confronting their students (Obiakor *et al.*, 1997). From our perspective, it is nonproductive to bemoan new multicultural paradigms that incorporate quality and equity in educational programming (Price, 1991). We must avoid any kind of multiculturalism that tends to project goodness with underlying negative intentions, and phony sense of community that hampers ways to increase knowledge about the interactions between human behaviors and cultural styles. To this end, teacher educators and leaders must make efforts to recruit and retain culturally sensitive students, faculty, and staff to remain competitive in this age of change (Obiakor, 2001b,

2007; Obiakor and Beachum, 2005; Obiakor and Utley, 1997; Wald, 1996).

Conclusion

While the United States and other countries have done a commendable job of putting legal mandates to protect students with disabilities and provide them with equal public education, the interpretation and implementation of those laws have many loopholes that need to be sealed. Overrepresentation of students from culturally and linguistically diverse background in special education has caused great concern. If students have been misplaced, it means that their educational needs cannot be met. Appropriate placement should be in least restrictive environments in which students' cultures and language do not result in misidentification, misassessment, miscategorization, and misplacement. The heart and soul of quality service delivery for students with special needs must include nonrestrictive environments and settings that maximize their potential. Finally such environments must be culturally, linguistically, and socioeconomically accepting.

Bibliography

- ADA (Americans with Disabilities Act) (1980). Pub. L. No. 101-336.
- Binet, A. (1909). *Les idées modernes sur les enfants (Modern Ideas for Children)*. Paris: Hammarion.
- Brown v. Board of Education of Topeka Kansas (1954). 347 U.S. 483, 745-ct-686, 98 L. Ed. 873, 530. O. 326.
- Civil Rights Act (1964). Pub. L. No. 88-352.
- Cullinan, D. and Kauffman, J. M. (2005). Do race of student and race of teacher influence ratings of emotional and behavioral problem characteristics of students with emotional disturbance? *Behavioral Disorders* **30**, 393-402.
- Education of All Handicapped Children Act (1975). Pub. L. No. 94-142.
- Education of All Handicapped Children Act Amendments (1986). Pub. L. No 99-457.
- Effectiveness of special education. *Journal of Special Education*, 36(4), 186-194.
- Ford, B. A., Obiakor, F. E., and Patton, J. M. (1995). *Effective Education for African American Exceptional Learners: New perspectives*. Austin, TX: Pro-Ed.
- Gardner, H. (1993). *Multiple Intelligences: The Theory of Practice*. New York: Basic Books.
- Goldstein, A. P. (1999). *The Prepare Curriculum: Teaching Prosocial Competencies*. Champaign, IL: Research Press.
- Goleman, D. (1995). *Emotional Intelligence: Why it Matters More than IQ*. New York: Bantam Books.
- Goodlad, J. I. (1983). Access to knowledge. In Goodlad, J. I. and Lovitt, T. L. (eds.) *Integrating General and Special Education*, pp 1-22. New York: Merrill.
- Gould, S. J. (1981). *The Mismeasure of Men*. New York: W. W. Morton.
- Haberman, M. (1995). *Star Teachers of Children in Poverty*. West Lafayette, IN: KappaDelta Pi.
- Hillard, A. G. (1992). The pitfalls and practices of special education practice. *Exceptional Children* **59**, 168-172.
- IDEA (Individuals with Disabilities Education Act) (1990). Pub. L. No. 101-476.
- IDEA (Individuals with Disabilities Education Act) (1997). Pub L. No. 105-17.
- IDEIA (Individuals with Disabilities Education Improvement Act) (2004). Pub. L. No. 108-446.
- Katsiyannis, A., Yell, M. L., and Bradley, R. (2001). Reflections of the 25th anniversary of the individuals with exceptionalities education act. *Remedial and Special Education* **22**(6), 324-339.
- Kauffman, J. M. (2002). *Education Deform? Bright People Sometimes Say Stupid Things about Education*. Lanham, MD: Scarecrow Education.
- Kauffman, J. M. (2003a). Reflections on the field. *Behavioral Disorders* **28**, 205-208.
- Kauffman, J. M. (2003b). Appearances, stigma, and prevention. *Remedial and Special Education* **24**, 195-198.
- Kauffman, J. M. (2004). The president's commission and the devaluation of special education. *Education and Treatment of Children* **27**, 307-324.
- Keogh, B. K. (1990). Narrowing the gap between policy and practice. *Exceptional Children* **57**(2), 186-190.
- Lieberman, L. M. (2001). The death of special education. *Education Week* 39-41.
- Long, N. J. (1997). The therapeutic power of kindness. *Reclaiming Children and Youth* **5**, 242-246.
- Mills v. Board of Education of the District of Columbia (1992). 348 f. Supp. 866 (D.D.C.1972).
- Minton, H. and Schneider, F. (1985). *Differential Psychology*. Prospect Heights, IL: Waveland Press.
- Mostert, M. P., Kauffman, J. M., and Kavale, K. A. (2003). Truth and consequences. *Behavioral Disorders* **28**, 333-347.
- Mukuria, G. and Obiakor, F. E. (2004). Special education issues and African diaspora. *Journal of International Special Needs Education* **7**, 12-17.
- No Child Left Behind Act (2001). Pub. L. No. 107-110.
- Obiakor, F. E. (1999). Teacher expectations of minority exceptional learners: Impact on "accuracy" of self-concepts. *Exceptional Children* **66**, 39-53.
- Obiakor, F. E. (2001a). Developing emotional intelligence in learners with behavioral problems: Refocusing special education. *Behavior Disorders* **26**, 321-331.
- Obiakor, F. E. (2001b). *It Even Happens in Good Schools: Responding to Cultural Diversity In Today's Classrooms*. Thousand Oaks, CA: Corwin.
- Obiakor, F. E. (2003). To asses or not, Is that the question? Who benefits from the No Child Left Behind Act? *Invited Scholar Presentation Sponsored by the Institute on Multicultural Relations and the Milwaukee Urban League, University of Wisconsin-Milwaukee*. Milwaukee, WI.
- Obiakor, F. E. (2004). Impact of changing demographics on public education for culturally diverse learners with behavior problems: Implications for teacher preparation. In Bullock, L. M. and Gable, R. A. (eds.) *Quality Personnel Preparation in Emotional/Behavioral Disorders: Current Perspectives and Future Directions*, pp 51-63. Denton, TX: Institute for Behavioral and Learning Differences at the University of North Texas.
- Obiakor, F. E. (2007). *Multicultural Special Education: Culturally Responsive Teaching*. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Obiakor, F. E. (2008). *The Eight-Step Approach to Multicultural Learning and Teaching*, 3rd edn. Dubuque, IA: Kendall/Hunt.
- Obiakor, F. E. (2009). Educating African American urban learners: Brown in context. In Brown, M. C. II and Bartee, R. D. (eds.) *The Broken Cisterns of African American Education*, pp 61-72. Charlotte, NC: Information Age.
- Obiakor, F. E. and Algozzine, B. (eds.) (1995). Educating learners with problem behaviors: An unresolved issue for general and special educators. In *Managing Problem Behaviors: Perspectives for General and Special Educators*, pp 1-19. Dubuque, IA: Kendall/Hunt.
- Obiakor, F. E. and Beachum, F. D. (2005). *Urban Education for the 21st Century: Research, Issues, and Perspectives*. Springfield, IL: Charles C. Thomas.
- Obiakor, F. E., Enwefa, S., Utley, C., et al. (2004). *Serving Culturally and Linguistically Diverse Students with Emotional and Behavioral Disorders*. Arlington, VA: Council for Children with Behavioral Disorders, the Council for Exceptional Children.

- Obiakor, F. E. and Ford, B. A. (2002). *Creating Successful Learning Environments for African American Learners with Exceptionalities*. Thousand Oaks, CA: Corwin.
- Obiakor, F. E., Grant, P. A., and Dooley, E. A. (2002). *Educating all Learners: Refocusing the Comprehensive Support Model*. Springfield, IL: Charles C. Thomas.
- Obiakor, F. E., Harris, M. K., and Beachum, F. D. (in press). The state of special education for African American learners in Milwaukee. In Williams, G. (ed.) *The State of Black Children in Milwaukee* (Book).
- Obiakor, F. E., Mehring, T. A., and Schwenn, J. O. (1997). *Disruption, Disaster, and Death: Helping Students Deal with Crises*. Arlington, VA: Council for Exceptional Children.
- Obiakor, F. E., Schwenn, J. O., and Rotatori, A. F. (1999). *Advances in Special Education: Multicultural Education for Learners with Exceptionalities, Vol. 12*. Stanford, CT: JAI Press.
- Obiakor, F. E. and Utley, C. A. (1997). Rethinking preservice preparation for teachers in the learning disabilities field: Workable multicultural strategies. *Learning Disabilities Research and Practice* **12**, 100–106.
- Obiakor, F. E., Utley, C. A., and Rotatori, A. F. (2003a). *Advances in Special Education: Effective Education for Learners with Exceptionalities, Vol. 15*. Oxford, UK: Elsevier Science/JAI Press.
- Obiakor, F. E. and Wilder, L. K. (2003). Disproportionate representation in special education: What principals can do. *Principal Leadership* **4**, 17–21.
- Pennsylvania Association for Retarded Children v. Commonwealth of Pennsylvania 343 F. Supp. 279 (D. C. Pa 1971).
- Prater, L. P. (2006). Institutionalized terror: A social system's analysis of police brutality. *Paper Presented at the Annual Professional Development Conference of the National Social Science Association*, San Francisco, CA.
- Price, H. B. (1991). Multicultural education: The debate. *Humanities in the South* 1–8.
- Rueda, R. (2007). Multicultural special education: Future perspectives. In Obiakor, F. E. (ed.) *Multicultural Special Education: Culturally Responsive Teaching*, pp 290–297. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Sasso, G. M. (2003). An examined life: A response to James Kauffman's reflections on the field. *Behavioral Disorders* **28**, 209–211.
- Schrag, J. A. (1993). Restructuring schools for better alignment of general and special education. In Goolad, J. I. and Lovitt, T. C. (eds.) *Integrating General and Special Education*, pp 203–227. New York: Merrill.
- Smith, T. E. (2005). IDEA 2004: Another round in the reauthorization process. *Remedial and Special Education* **26**(6), 314–323.
- Smith, T. B., Richards, P. S., MacGranley, H., and Obiakor, F. E. (2004). Practicing multiculturalism: An introduction. In Smith, T. B. (ed.) *Practicing Multiculturalism: Affirming Diversity in Counseling and Psychology*, pp 3–16. Boston: Allyn and Bacon.
- Utley, C. A. and Obiakor, F. E. (2001). *Special Education, Multicultural Education, and School Reform: Components of Quality Education for Learners with Mild Disabilities*. Springfield, IL: Charles C. Thomas.
- Vocational Rehabilitation Act (1973). Pub. L. No. 93-112.
- Wald, J. L. (1996). Diversity in the special education training force. *NCPSE News* **1**, 1 and 6.
- Weikart, D. P. (1977). Preschool intervention for the disadvantaged child: A challenge for special education. In Spicker, H. H., Anastasiow, K. J., and Hodges, W. L. (eds.) *Children with Special Needs: Early Development and Education*, pp 73–89. Minneapolis, MN: Leadership Training Institute/Special Education, University of Minnesota.
- Wilder, L. K., Obiakor, F. E., and Algozzine, B. (2003). Homeless students in special education: Beyond the myth of socioeconomic dissonance. *Journal of At-Risk Issues* **9**, 9–16.
- Winzer, M. A. and Mazurek, K. (1998). *Special Education in Multicultural Contexts*. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Working with Children with Disabilities*, 2nd edn. Norwood, MA: Christopher Gordon.
- Yell, M. L. (2004). *The Law and Special Education*, 2nd edn. Upper Saddle River, NJ: Prentice Hall.

Further Reading

- Fordham, S. (1988). Racelessness as a factor in Black student's success. Pragmatic Strategy or Pyrrhic victory. *Harvard Educational Review* **58**, 54–84.

Educating Students with Traumatic Brain Injury

S Burkhardt and A Rotatori, Saint Xavier University, Chicago, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Injury factors – Common causes of traumatic brain injury in children, including accidents and assault.

Mental functions – Cognitive capabilities which support academic achievement and behavior, including attention, memory, executive functions and response time.

Multidisciplinary planning – Educational programming which includes goals and strategies from various learning specialists.

Positive behavioral support – Behavior management programming that relies on reinforcement and strengths as the means of increasing adaptability.

Recovery factors – Post-trauma elements of recovery from traumatic brain injury that affect prognosis, including severity of injury and environmental supports.

Secondary attention-deficit hyperactivity disorder (SADHD) – The development of attention deficit disorder following traumatic brain injury.

School readiness factors – Teacher and school preparedness for post-injury support of students.

Traumatic brain injury (TBI) – Open and/or closed head injury that is acquired after birth and is caused by external physical force.

Educating the Student with Traumatic Brain Injury

In 1990, traumatic brain injury (TBI) was added as a disability category within the Individuals with Disabilities Education Act (IDEA) in the United States (see Code of Federal Regulations, Title 34, Section 300.7(b) (12); (NICHY, nd)). TBI is defined, in the IDEA, as an acquired open or closed head injury to the brain, caused by external physical force. TBI results in total or partial functional disability and/or psychosocial impairment that have an adverse impact upon the student's educational performance. Specifically excluded from the term TBI are congenital or degenerative brain dysfunction as well as brain injuries caused by birth trauma (NECCYD, 1997).

Brain injury affects more than 1 million children every year, with about 30 000 of them sustaining a lifelong disability

due to the brain injury (NICHY, nd). Although children and teenagers are more likely to survive TBI than adults (US Department of Health and Human Services, 1999), it is the most common cause of death and disability among children in the United States (Keyser-Marcus *et al.*, 2002), with African Americans having the highest fatality rate (Arroyos-Jurado and Savage, 2008). A review of the National Trauma Data Base, by Haider *et al.* (2007), indicated that black children “experience worse clinical and functional outcomes following TBI” compared to white and Hispanic children. Socioeconomic status (SES) and availability of postinjury services were associated with these findings. In terms of gender, males are more likely to sustain TBI than females (Arroyos-Jurado and Savage, 2008).

TBI may affect the way a student thinks, speaks, remembers, pays attention, and behaves. **Table 1** identifies the mental functions that may be affected by TBI. The learning characteristics of students with TBI may differ significantly from their learning characteristics prior to TBI (NICHY, nd). Understanding their characteristics and methods to maximize the learning potential of students with TBI is very important. This is the focus of this article.

Categories of Factors Related to TBI

To educate students with TBI, it is important to consider a framework to meet their needs. Four general categories are considered. Injury factors are those aspects of TBI that constitute the basis of the educational difficulties experienced by a student with TBI. School-readiness factors refer to programs and personnel available in the general education setting who can accommodate the learner with TBI. Coordination of services refers to various intervention strategies and accommodations provided on behalf of a student with TBI. Lastly, recovery factors are discussed as the basis for addressing the continuing needs of students with TBI (**Figure 1**).

Injury Factors: Causes of TBI

The top ten categories of head injury among children, age 14 or younger, are the following: bicycle accidents; football injuries; baseball/softball injuries; powered skateboard and scooters; basketball injuries; unpowered skateboards and scooters; winter sports; powered recreational vehicles; and water sports and trampolines (American Association of Neurological Surgeons, 2007). Additional causes of injury

Table 1 Mental functions affected by traumatic brain injury

Cognition
Speech and language
Memory
Attention
Reasoning
Abstract thinking
Information processing
Judgment
Problem solving
Sensory, perceptual, and motor abilities
Psychosocial behavior
Physical functions

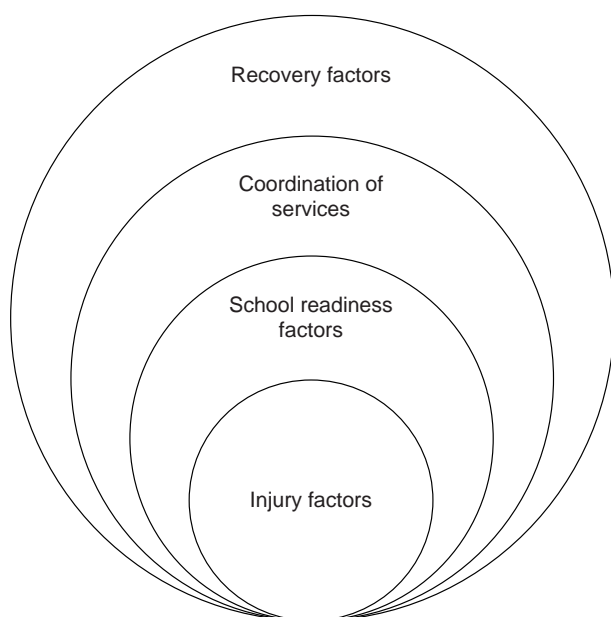


Figure 1 Correlates of student success following TBI. (Outermost) Injury factors, including severity of injury. (Outer) School-readiness factors, including teacher preparedness. (Inner) Coordination of services, including multidisciplinary assessment and intervention. (Innermost) Recovery factors, including positive behavior management.

include falls, assaults, other violence, child maltreatment, and child-passenger injury (CDC, 2008). Sports injuries are a common source of head injury for high school and college students, with football, ice hockey, and soccer being most frequently associated with concussion – a form of head injury that does not include loss of consciousness (Moser *et al.*, 2007; National Academy of Neuropsychology Planning and Policy Committee, 2007).

The National Academy of Neuropsychology Planning and Policy Committee (2007) highlighted concerns with regard to mild brain trauma and young athletes. School sporting events seldom have medical personnel in attendance and injured student players do not always receive immediate medical attention. Brooks (2004) discussed pertinent legal cases with regard to the liability for school districts related to

student-athlete head injuries, including differing policies for benching a player who complains of dizziness following a play. Clearly, schools are well placed to address brain-injury prevention for all students. The Centers for Disease Control and Prevention (CDC, 2007) made the following suggestions for prevention education: reminding students to wear seat belts at all times when in motor vehicles; encouraging parents to place children in the back seat to avoid air-bag injury; and promoting the wearing of helmets when riding a bike or horse, skateboarding, snowboarding, and playing sports.

School Readiness Factors

Two types of school readiness factors are identified. First, the role of the school in supporting the re-entry of students to the school environment following TBI is considered; and, second, the preparedness of the general education teacher for including a student with TBI is discussed.

The Role of the School in Re-Entry of Students with TBI

A report of best practices for the rehabilitation of TBI in children and adolescents was compiled a decade ago (US Department of Health and Human Services, 1999); however, the information is currently deemed archival. The 1999 report indicated that data with regard to special education placement, services, and outcomes for students with TBI was limited. The report offered analysis of several studies from the 1990s, including the estimate that between 9% and 38% of students with TBI were referred to special education after returning to school.

Keyser-Marcus *et al.* (2002) promote the creation of a student learning profile for a student with TBI upon return to school. Such a profile emphasizes student strengths while suggesting instructional interventions that are effective for the individual. Bowen (2005) identified instructional methods for students with TBI which include adapted school environment and special education supports. Direct instruction (i.e., Glang, 1993) and cognitive strategy intervention, were determined to assist students with and without disability (Glang *et al.*, 2008). Both methods have shown promise with students with learning and behavioral characteristics similar to those of students with TBI (Bowen, 2005; Glang *et al.*, 2008). Assessment emerges as an essential best practice for the support of students with TBI (see North Carolina State Department of Public Instruction, 1997). Bowen emphasized the need for functional assessments of learning and behavior, in addition to formal assessments of cognition and achievement, to capitalize on remaining strengths and to minimize the impact of diminished capabilities. Assessment measures for students with TBI, often, include neuropsychological evaluation, with assessment of numerous domains: information processing, sensory-motor

integration, reaction time, mental flexibility, memory and learning, language comprehension, speech fluency, and social-emotional responses. Interview of family members and review of pre-injury academic performance are necessary, in light of the findings that post-injury achievement is associated with pre-injury performance (Arroyos-Jurado *et al.*, 2006).

Teacher Preparedness for Students with TBI

Educators may struggle with school programming that meets educational and psychosocial needs of students with TBI and their families. A report by Mohr and Bullock (2005) based upon focus group discussion with educators identified potential difficulties regarding teacher preparedness for students with TBI: lack of teacher awareness with regard to the characteristics of TBI and the course of recovery, poorly defined educational goals, and few specific accommodations for students with TBI.

Useful information for the classroom teacher needs to be as real world as possible (Keyser-Marcus *et al.*, 2002). The types of problems experienced by students with TBI may include difficulty remembering the previous day's lecture or the location of a locker; disorganization; losing track of conversation; inflated estimate of academic performance; and difficulty naming familiar objects. Thus, it is critical to provide a classroom observation checklist to assess functional domains and problems experienced by the student in the classroom environment. The checklist includes functional assessment of the following domains: memory, attention and concentration, executive functioning, self-awareness, and language. The checklist helps to identify significant indicators associated with TBI, such as losing track of conversation, poor time management, and limited meta-cognition (i.e., overestimating knowledge or preparation for academic tasks). In the end, Glang *et al.* (2004) recommended a TBI Team Model that provides for peer consultants to assist schools with the education of students with TBI. This model involves the following components: an assessment of student needs, recruitment of an intervention team, training of the team members, and evaluation of implementation and outcomes. Team members serve as consultants to school staff and the classroom teacher.

Coordination of Services

Two aspects of coordination of educational services for students with TBI are identified. Multidisciplinary planning is presented as a needed component of educating the student with TBI who is, often, returning to school as a transition from rehabilitation services. Additionally, although TBI is a special education category of eligibility, it is a category that does not encompass specific educational programming. Thus, the role of the Individual Education Plan (IEP) for the student with TBI is considered.

Multidisciplinary Planning

Mohr and Bullock (2005) defined specific elements of a school prepared to support successful re-entry of students with TBI. Excellent communication among school, family, and rehabilitation staff should be established prior to the student's discharge from rehabilitation programming. Communication may be enhanced by identifying a liaison from the school knowledgeable in TBI and transition from rehabilitation. Involvement of parents in planning is needed, particularly utilizing the family's successful strategies at home to support transition back to school. Continued use of strategies from home and the rehabilitation facility provide continuity and avoid confusion for the student making post-injury adjustments.

Since 1990, students with TBI are eligible for special education services within the TBI category. Despite this, few unique, coordinated, and comprehensive services are designated specifically for TBI. Thus, students with TBI are likely to receive special education services by learning disabilities specialists, occupational therapists, speech pathologists, and school social workers (Arroyos-Jurado and Savage, 2008). Common accommodations for students with TBI include shortened school day or modified school schedule, with most rigorous classes attended during the student's alert and rested periods of the day. Continuous monitoring and assessment of not only the learner with TBI but also the suitability of the learning environment is needed. Clearly, instructional methods and materials that build on the student's strengths and avoid areas of deficit are particularly appropriate (Bowen, 2005). Harvey (2006) recommended coordinated services that are both variable and flexible.

The Individualized Education Plan

The individualized education plan (IEP) should reflect the best practices of each professional supporting students with TBI (Arroyos-Jurado and Savage, 2008). Reports of specific IEP accommodations are lacking in systematically reviewed outcome studies; however, special education services and an IEP are associated with increased success in transition to post-secondary education (Bowen, 2005; Todis and Glang, 2008). Adjusted schedule was a frequent accommodation for students with TBI following return to school (Gfroerer *et al.*, 2008). Strategies associated with best practices for students with TBI include positive behavioral support, interventions designed to compensate for memory loss, customized supports for organization, and social regulation (Bowen, 2005). A positive behavioral support approach focuses on the replacement of inappropriate behaviors with more appropriate behaviors that allow learners with TBI to achieve their social-emotional and academic needs. Positive behavioral support encompasses a variety of behavioral strategies such as: environmental

scans (Mukuria and Obiakor, 2008), functional behavioral assessment, positive reinforcement, behavioral momentum (Ylvisaker *et al.*, 2003), and structured environment (Bowen, 2005). The successful implementation of these strategies allows teachers and service providers to tailor a learner's environment based on his/her preferences, strengths, and needs so that more functional behavior occurs. (see Mukuria and Obiakor, 2008). Customized programming is dependent upon the special education staff's familiarity with TBI. Keyser-Marcus *et al.* (2002) recommended strategies for teaching the student, with TBI, in the general education setting. Strategies include multimodal instruction, frequent use of review and repetition, task analysis to support skill acquisition, and breaking down assignments into smaller components. Instructors are encouraged to patiently answer questions and encourage independent thinking, memory notebooks, tape recorders, color coding of instructional materials, and customized checklists may be needed to support the learner with TBI (Bowen, 2005).

NICHY (nd) provided a list of tips for teachers that target the residual effects of TBI with regard to education-related behaviors, such as information processing, response time, and psychosocial adjustment. Suggestions include allowing extra time for schoolwork and tests, the use of simple, short, and/or written instructions to clarify directives; the use of demonstration to teach how to perform new tasks; ample opportunity to practice new skills; establishing a routine and warning with regard to changes in the routine. Planned rest periods to address fatigue may be needed. Reduction of distractions in the learning environment is, often, helpful. Finally, educators are encouraged to have flexible expectations and patience in their efforts on behalf of students with TBI.

Recovery Factors

Two elements of recovery from TBI are discussed as they relate to educating the student with TBI. Although children survive TBI and make good progress, the risk of abnormal post-injury development remains. In a similar vein, problematic behaviors related to TBI may require continuous positive behavior management for students with TBI in order to support education in the least restrictive environment.

Post-Injury Development of Students with TBI

Although recovery continues following TBI, post-injury development of students who have sustained brain injury may be affected. Arroyos-Jurado *et al.* (2006) determined that severity of TBI was associated with post-injury loss of nonverbal intelligence quotient (IQ) performance, impaired encoding, and difficulty recalling information. Max *et al.* (2005a) determined that 22% of children in the age range of 5–14 years, who sustained TBI, exhibited personality changes in the 6 months following injury. Poor post-injury

regulation of emotion was determined to be related to the severity of the brain injury and not psychosocial factors. In addition, secondary attention-deficit hyperactivity disorder (SADHD) is a common complication following pediatric TBI (Max *et al.*, 2005b). SADHD includes personality changes, behavioral problems, cognitive deficits, and impairments in adaptive functioning. Up to 21% of children experiencing TBI develop SADHD by 24 months post-injury. The findings of Max *et al.* (2005b) suggest the development of SADHD occurs in students who, prior to injury, did not display ADHD.

Max *et al.* (2005a) found that, of 143 children with no pre-injury ADHD who sustained mild-to-severe TBI, 15% developed SADHD within 12 months of injury. Additionally, a minority of the children with SADHD also developed personality changes as well as other disorders including oppositional defiant disorder, obsessive-compulsive disorder, conduct disorder, and dysthymia. Analysis of risk factors associated with the development of SADHD identified not only biological features related to injury (severity and location), but psychosocial risk factors as well (Max *et al.*, 2005b). Pre-injury SES correlated with the development of SADHD, although the specific mechanisms underlying such correlations remain unknown (Max *et al.*, 2005a). Family history of ADHD was not predictive of SADHD – suggesting a genuine distinction between primary ADHD and SADHD. At 6 months post injury, lesion location and SES predicted SADHD. However, by 18 months post injury, family adversity and personal pre-injury functioning were the strongest predictors of the extent of SADHD. These findings suggest that TBI – combined with psychosocial disadvantage – predicts the development of SADHD. By 18 months post injury, severity of injury was not associated with SADHD. Thus, there is evidence to indicate that while severity of injury is an important determinant of post-injury functioning, post-injury environment may influence post-injury recovery. As return to school constitutes a major feature of the post-injury environment, special care must be taken to provide an optimal environment for a student who is experiencing learning and behavioral changes due to TBI. Additionally, demands of learning may need modification for these students as they transition back to the school environment.

Continuous Positive Behavior Supports

TBI encompasses the possibility that the student with TBI may continue to display difficulties with self-regulation and emotional expression (Dykeman, 2003). Constructive management of behavioral difficulties related to TBI, SADHD, and/or organic personality changes is necessary to support recovery and inclusion in the school setting. Behavioral contracts that emphasize positive reinforcement for appropriate behavior should serve as the central feature of a behavioral management plan for a student with

TBI. The need to accept the slow rate at which desired behavior can be shaped following TBI requires patience and persistence; and disciplinary approaches that include punishment and response cost for undesirable behavior may impede adjustment to the school setting. Reduced expectations for work load, stamina, and participation set the stage for the student with TBI to experience success for that which is accomplished rather than anxiety or discouragement for that which is not readily achieved.

Rules and routines may need to be directly taught and practiced. Activities that require prolonged concentration may need to be avoided. Verbal interactions may require longer processing time; the student with TBI should not be reprimanded for delayed response to directives. The IEP is likely to be a working document that is flexible and allows for unique and changing needs of a student with TBI. Clearly, the student may need direct and ongoing assistance with organizational skills.

Conclusion

Learning is supported by good mental functioning, thus, TBI in many cases has a significant adverse impact upon a student's ability to learn. Impairment in many of the mental processes, including attention, memory, language, and regulation of emotion, is frequently related to TBI. Educators and service providers may find themselves underprepared for the return to school of their student with TBI. Since students with TBI, often, differ greatly in terms of the severity and type of impairment, individualized assessment, and service plans are needed.

TBI is an acquired disability in which injury factors, such as severity, influence the degree of impairment, and course of recovery. In order to educate a student with TBI, the school must be ready to meet his/her needs, with instructional methods and educators who are knowledgeable with regard to learning and behavioral characteristics of TBI during recovery. Educational services for TBI include multidisciplinary teams who employ best practices in assessment and intervention. Emphasis upon a flexible IEP is needed in order to identify and utilize the student's remaining and emerging strengths. Finally, education of the student may require patience; by its very nature, TBI continues to affect development and functioning. Thus, students with TBI may need continuing services, including behavioral management plans to promote restoration of adaptive behavior while effectively managing residual post-injury dysfunction.

See also: Affect, Mood and Emotions; Anxiety; Educating Students with Emotional and Behavioral Disorders; Emotion in Educational Contexts; Functional Behavioral Assessment; Parent and Family Involvement in the Education of Children with Special Needs; Peer Relations

and Socialization of Children and Adolescents with Special Needs; School-Based Services for Children with Special Needs.

Bibliography

- American Association of Neurological Surgeons (2007). Sports related injuries. *Neurosurgery*. http://www.neurosurgerytoday.org/what/patient_e/sports.asp?link_id=5792 (accessed August 2009).
- Arroyos-Jurado, E., Paulsen, J., Ehly, S., and Max, J. (2006). Traumatic brain injury in children and adolescent: Academic and intellectual outcomes following injury. *Exceptionality* **14**(3), 125–140.
- Arroyos-Jurado, E. and Savage, T. (2008). Intervention strategies for serving students with traumatic brain injury. *Intervention in School and Clinic* **43**(4), 252–254.
- Bowen, J. (2005). Classroom interventions for students with traumatic brain injuries. *Preventing School Failure* **49**, 34–41.
- Brooks, M. (2004). Coach, I'm OK, can I go in? *Principal Leadership (High School Edition)* **5**, 63–64.
- CDC (Centers for Disease Control and Prevention) (2007). Traumatic brain injury: Prevention. <http://www.cdc.gov/ncipc/tbi/Prevention.htm> (accessed August 2009).
- CDC (Centers for Disease Control and Prevention) (2008). National Center for Injury Prevention and Control: What is traumatic brain injury. <http://www.cdc.gov/ncipc/tbi/TBI.htm> (accessed August 2009).
- Dykeman, B. (2003). School-based interventions for treating social adjustment difficulties in children with traumatic brain injury. *Journal of Instructional Psychology* **30**, 225–230.
- Groerer, S. D., Wade, S. L., and Wu, M. (2008). Parent perceptions of school-based supports for children with traumatic brain injury. *Brain Injury* **9**, 649–656.
- Glang, A. (1993). Using direct instruction with brain injured students. *Direct Instruction News* **Fall**, 23–28.
- Glang, A., Tyler, J., Pearson, S., Todis, B., and Morvant, M. (2004). Improving educational services for students with TBI through statewide consulting teams. *NeuroRehabilitation* **19**(3), 219–231.
- Glang, A., Ylvisaker, M., Stein, M., et al. (2008). Validated instructional practices: Application to students with traumatic brain injury. *Journal of Head Trauma Rehabilitation* **23**(4), 243–251.
- Haider, A. H., Efron, D. T., Haut, E. R., et al. (2007). Black children experience worse clinical and functional outcomes following traumatic brain injury: An analysis of the National Pediatric Trauma Registry. *Journal of Trauma* **62**, 1259–1263.
- Harvey, J. M. (2006). Best practices in working with students with traumatic brain injuries. In Thomas, A. and Grimes, J. (eds.) *Best Practices in School Psychology*, vol. 4, pp 1433–1446. Bethesda, MD: National Association of School Psychologists.
- Keyser-Marcus, L., Briel, L., Sherron-Targett, P., et al. (2002). Enhancing the schooling of students with traumatic brain injury. *Teaching Exceptional Children* **34**, 62–67.
- Max, J. E., Levin, H. S., Landis, J., et al. (2005a). Predictors of personality change due to traumatic brain injury in children and adolescents in the first six months after injury. *Journal of the American Academy of Child and Adolescent Psychiatry* **44**(5), 434–442.
- Max, J. E., Schachar, R. J., Levin, H. S., et al. (2005b). Predictors of secondary attention deficit/hyperactivity disorder in children and adolescents 6 to 24 months after traumatic brain injury. *Journal of the American Academy of Child and Adolescent Psychiatry* **44**(10), 1041–1049.
- Mohr, J. L. and Bullock, L. M. (2005). Traumatic brain injury: Perspectives for educational professionals. *Preventing School Failure* **49**(8), 53–57.
- Moser, R., Iverson, G., Echemendia, R., et al. (2007). Neuropsychological evaluation in the diagnosis and management of sports-related concussion. *Archives of Clinical Neuropsychology* **22**, 909–916.
- Mukuria, G. M. and Obiakor, F. E. (2008). Curriculum innovation to educate students with autism. In Rotatori, A. F., Obiakor, F. E., and Burkhardt, S. (eds.) *Autism and Developmental Disabilities: Current Practices and Issues*, vol. 19, pp 25–40. Bingley: Emerald Group.

- National Academy of Neuropsychology Planning and Policy Committee (2007). Neuropsychological evaluation in the diagnosis and management of sports-related concussion. *Archives of Clinical Neuropsychology* **22**(8), 909–916.
- NECCYD (National Education Center for Children and Youth with Disabilities) (1997). *General Information about Traumatic Brain Injury. Fact Sheet Number 4*. <http://www.kidsource.com/NICHCY/brain.html> (accessed August 2009).
- NICHY (National Dissemination Center for Children with Disabilities) (nd). Traumatic brain injury. <http://www.nichcy.org/Disabilities/Specific/Pages/TraumaticBrainInjury.aspx> (accessed August 2009).
- North Carolina State Department of Public Instruction (1997). Best practices in assessment and programming for students with traumatic brain injury. *ERIC Document Reproduction Service No. ED417541*.
- Todis, B. and Glang, A. (2008). Redefining success: Result of a qualitative study of postsecondary transition outcomes for youth with traumatic brain injury. *Journal of Head Trauma Rehabilitation* **23**(4), 252–263.
- US Department of Health and Human Services (1999). Rehabilitation for traumatic brain injury in children and adolescent: Summary. *Evidence Report/Technology Assessment, Number 2*. Rockville, MD: Agency for Health Care Policy and Research. <http://www.ahrq.gov/clinic/epcsums/tbisum2.htm> (accessed August 2009).
- Ylvisaker, M., Jacobs, H. E., and Feeney, T. (2003). Positive supports for people who experience behavioral and cognitive disability after brain injury: A review. *Journal of Head Trauma Rehabilitation* **18**, 7–32.
- Bullock, L. M., Gable, R. A., and Mohr, J. D. (2005). Traumatic brain injury: A challenge for educators. *Preventing School Failure* **49**(4), 6–10.
- CDC (Centers for Disease Control and Prevention) (nd). Injuries among children and adolescents. <http://www.cdc.gov/ncipc/factsheets/children.htm> (accessed August 2009).
- Hooper, S. (2006). Myths and misconceptions about traumatic brain injury: Endorsements by school psychologists. *Exceptionality* **14**(3), 171–182.
- Hux, K. (2003). *Assisting Survivors of Traumatic Brain Injury: The Role of Speech-Language Pathologists*. Austin, TX: PRO-ED.
- Mayfield, J. and Homack, S. (2005). Behavioral considerations associated with traumatic brain injury. *Preventing School Failure* **49**(4), 11–17.
- Ylvisaker, M., Harwick, P. L., and Stevens, M. (1991). School reentry following head injury: Managing the transition from hospital to school. *Journal of Head Trauma Rehabilitation* **6**(1), 10–22.
- Ylvisaker, M., Todis, B., Glang, A., et al. (2001). Educating students with TBI: Themes and recommendations. *Journal of Head Trauma Rehabilitation* **16**(1), 76–93.

Relevant Websites

- <http://kidshealth.org/parent> – Kids Health, Occupational Therapy.
- <http://learningdisabilities.about.com> – Learning Disabilities, Traumatic Brain Injuries – Traumatic Brain Injury Programs in Public School.
- <http://msktc.washington.edu> – Model Systems Knowledge Translation Center for Traumatic Brain Injury (TBI), Spinal Cord Injury (SCI) and Burn Injury.
- <http://www.neuro.pmr.vcu.edu> – National Resource Center for Traumatic Brain Injury.
- <http://specialed.about.com> – Special Education, Sample Behavior Contracts and Worksheets for Students.
- <http://specialed.about.com> – Special Education, Special Education for Students with TBI.
- <http://www.tbi.pmr.vcu.edu> – Traumatic Brain Injury Model System, Virginia Commonwealth University.

Further Reading

- Blosser, J. and DePompei, R. (2003). *Pediatric Traumatic Brain Injury: Proactive Intervention*, 2nd edn. Clifton Park, NY: Delmar/Thomson Learning.
- Bohmann, J. (2007). Traumatic brain injury and teens: Information for school administrators. *Principal Leadership* **3**, 12–15.

English Language Learners with Special Needs

R Rueda and G Ragusa, University of Southern California, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

English learner – English learners are those individuals whose primary language is other than English and who is in the process of acquiring English language skills in the areas of listening comprehension, speaking, reading, and writing necessary to succeed in the school and in the workplace.

IDEA – The Individuals with Disabilities Education Act (IDEA), which is the current version of the federal special education law that was first enacted in 1975 and was originally known as the Education for All Handicapped Children Act of 1975 or Public Law 94-142. It outlines the provisions to be followed in the education of students with disabilities.

Response to intervention – Response to intervention refers to a family of approaches which aim to prevent unnecessary assignment to special education. With RTI, low-performing children are offered intense, individualized academic intervention. Student progress is monitored to see if response to this intervention yields adequate academic growth. A key aspect of the approach is the provision of services to students with academic delays using one or more research-validated interventions. The student's academic progress is monitored frequently to see if those interventions are sufficient to help the student to catch up with his or her peers. If the student fails to show significantly improved academic skills despite several well-designed and implemented interventions, this failure to respond to intervention can be viewed as evidence of an underlying learning disability.

This article discusses the challenges and dilemmas associated with placement and service provision for English language learners (ELLs) in special education systems and provides recommendations for improved assessment practices and service provisions. These issues represent a range of factors, including inappropriate assessment and psychometric issues, overrepresentation in special education programs, ambiguities in special education categorization, questionable pre-referral intervention practices, inadequate practitioner preparation related to language and cultural factors, and ineffective instructional practices both within and outside of the special education system

(McCardle *et al.*, 2005). These challenges are not new to the special education field as they have been described in the literature since the 1970s.

The difficulties associated with meeting the needs of ELLs appear to remain in spite of new reauthorizations of federal special education legislation that was originally prescribed by Public 94-142 (IDEA, 2004). An extensive review of the literature reveals a continuing issue with practitioners' beliefs and discrepancies in their understanding of the needs of ELLs in public education that may lead to biased assessment practices, overreferral to special education systems, and ineffective instructional practices. In this article, we provide a brief review of these and related issues.

The English Language Learner Population

In order to fully understand the needs of ELLs in education systems, it is important to describe the unique characteristics of the population. Linguistic diversity has dramatically increased in the past decade in the United States. The most recent United States Census data reveals that nearly 47 million US residents comprising nearly 18% of the total US population speak languages other than English. This growth is reflected in both the elementary and secondary school population. According to the US Department of Education, National Center on Educational Statistics (NCES), services related to ELLs were provided to 3.8 million students (11% of all students) during the 2003–04 school years. California and Texas had the largest reported number of students receiving ELL services during this period. In California, there were 1.6 million students (26% of all students) who received ELL services, and in Texas, there were 0.7 million students (16% of all students) who benefited from ELL services.

Not only do ELLs face challenges related to language barriers, they are also adversely affected by persistently low teacher expectations and struggles associated with national and state achievement tests (McCardle *et al.*, 2005). According to the National Assessment of Educational Progress (NAEP) data, the academic achievement of ELL students has not kept pace with their English-speaking peers. In 2007, NAEP data revealed that 7% of ELLs read at or above the proficient level in fourth grade, in contrast to 36% of English-speaking students.

Research has revealed a variety of reasons for why ELLs do not fare well on NAEP and related types of

standardized assessments. Test biases, misinterpretation of student needs, inadequate teacher training related to language assessment and instruction, and inconsistent teacher beliefs about the needs and expectations for ELLs all affect the achievement of this group, in particular those who are referred for special education services (Artiles *et al.*, 2004). These difficulties and misperceptions have been cited as factors contributing to overrepresentation of ELLs in special education programs nation-wide. Overrepresentation of ELLs in special education presents significant challenges to families, students, and educational systems nationally (Rueda and Windmueller, 2006).

Range of Special Education Programs

Currently, special education offers a range of programs in the United States. This range has increased as a function of each reauthorization of the federal special education legislation. In particular, the federal mandate for placing students in the least restrictive environment (LRE) has been an important factor in increased number and type of special education program options. Current program options include services for children beginning at birth and extending through to the age of 22 years and ranging in service intensity from full inclusion in general education with minimal consultative special education support to residential, segregated special education services.

While access to special education services is expected to be equitable and universally available to all students who qualify, national statistics have illustrated that service choice, particularly when inclusion is warranted, is inequitable and racially divided. The most recent reauthorization of the Individuals with Disabilities Education and Information

Act (IDEIA) included legislative revision that increased the number of students included in general education classroom for a portion of their day.

While inclusive settings have become more common, inequities abound. Researchers recognize that inclusion has benefited many students, particularly students identified with specific learning disabilities (SLDs). In the 2003 US Department of Education, Office of Special Education Report to Congress, underrepresented groups including ethnic minorities were placed in special education at significantly higher rates than white students. Moreover, students of color have been found to spend less time in inclusive settings. While Hispanic and Black K-12 students make up the largest population of special education learners, it appears from the national statistics that these subgroups spend more of their educational experiences in more restrictive special education settings than their peers. According to the National Center on Educational Statistics, (2005) approximately 50% of all disabled students in 2003–04 spent 80% or more of their day in a regular classroom, up from 45% in 1994–95. In contrast, Black and Hispanic students with disabilities spent less time in a regular classroom on average than their peers of other races/ethnicities with disabilities. This inequity suggests that particular ethnic groups are not provided with the full range of service options available to students with special needs. **Table 1** provides an illustration of this representation by ethnic group and federally recognized disability category.

As indicated by **Table 1**, in 2003, the percentage of students receiving special education with SLDs (the largest reported category) is highest for Hispanics (many of whom are ELLs). Hispanics represent 57.3% of all students with SLDs served.

Table 1 Disability distribution, by race/ethnicity, of students aged 6 through 21 receiving special education and related services: Fall 2003

<i>Disability</i>	<i>American Indian/ Alaskan Native</i>	<i>Asian Pacific Islander</i>	<i>Black (not Hispanic)</i>	<i>Hispanic</i>	<i>White (not Hispanic)</i>
Specific learning disability	54.5	39.5	44.5	57.3	45.6
Speech language impairments	16.0	26.0	14.3	18.3	20.1
Mental retardation	7.5	8.9	16.1	8.1	7.9
Emotional disturbance	8.0	11.6	11.2	4.9	7.9
Multiple disabilities	2.0	2.7	2.2	1.9	2.3
Hearing impairments	1.0	2.8	1.0	1.5	1.1
Orthopedic impairments	0.7	1.7	0.8	1.2	1.2
Other health impairments	5.7	5.3	5.9	4.1	9.1
Visual impairments	0.4	0.8	0.4	0.5	0.4
Autism	1.1	5.7	1.8	1.5	2.7
Deaf-blind	0.0	0.1	0.0	0.0	0.0
Traumatic brain injury	0.4	0.4	0.3	0.4	1.4
Developmental delay	2.6	1.4	1.2	0.6	1.2
All disabilities	100	100	100	100	100

Total may not sum to 100 because of rounding.

Source: US Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs (2007).

27th Annual (2005) Report to Congress on the Implementation of the Individuals with Disabilities Education Act, vol. 1. Washington, DC.

Overrepresentation of ELLs in Special Education

For nearly four decades, minority students have been overrepresented in special education. As a recent example, Artiles *et al.* (2005) conducted a study of within-group diversity of disproportionate representation of ELL students in special education. They discovered that ELLs identified as having limited proficiency in both their primary language (L1) and English (L2) showed the highest rates of identification in the special education categories investigated. Moreover, these students were consistently overrepresented in programs for students with learning disabilities and language and speech disabilities, and had significantly greater chances of being placed in special education.

Because of the long-standing overrepresentation of certain groups of students in special education, federal legislation now requires states to collect and examine data on disproportionality related to race and ethnicity. Wherever, states identify disproportional representation, they are mandated to plan services and to develop policies and procedures to adjust proportionality (US Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs, 2007).

In addition to examining referral and identification data, the legislation requires that states adjust their identification criteria for students with SLDs and adopt a policy of pre-referral intervention with a process of response to intervention (RTI). RTI is defined as “the use of strategies used by classroom teachers to target instructional interventions to children’s areas of specific need as soon as those needs become apparent” (US Department of Education, 2007), and before referral to special education is formally commenced.

Inappropriate Referrals Based on Language Differences

NCLB and the massive demographic changes of the student population have placed a tremendous demand on school districts to educate all students to proficient levels. One challenge is that in response to these conditions, educators increasingly inappropriately refer ELL students for special education (Rhodes *et al.*, 2005). In addition to these pressures, biases in the referral process have been implicated in disproportionality. This bias impacts other steps in the process such as pre-referral interventions and assessment practices.

English Language Learners and Issues of Education Assessment

Researchers have suggested that inappropriate assessment practices account for much of the disproportionality of English language learners receiving special education

services. Various studies have demonstrated that standardized achievement tests present significant challenges for ELLs. Differentiating between typical second language acquisition and learning disabilities is an important and continuing educational challenge (Gonzalez *et al.*, 1997). Another stems from the difficulty that the educational field has with development of assessment measures and procedures that adequately determine when a child with a primary language other than English is ready to be tested in English only, the primary language currently used to assess learners for special education (Abedi, 2006). Students often appear to be orally proficient in English long before they have fully developed cognitive academic language proficiency (CALP), the linguistic proficiency required for comprehending and effectively utilizing academic informational text. Bailey and Butler (2004) propose that effective inclusion efforts must begin with equitable exposure to and learning of the academic language required in the classroom and on content examinations.

Macswan and Rolstad (2006) documented a significant discrepancy between ELLs’ performance on commonly used language assessment measures and qualitatively coded natural language samples completed by the study’s research team. This led the researchers to question the construct validity of widely used language tests, and they noted that practitioners should use them with caution when considering special education status.

In a related large-scale national study that included analyses of standardized assessment data from K-12 settings, Abedi (2006) found a significant performance gap between ELL and non-ELL students in reading and writing assessments. The gap was less for science assessment measures, and lowest for math problem solving, for which the test items were less linguistically challenging for ELL students. The performance gap virtually disappeared in math computation, where the language demands of the test items were minimal. These results suggest that language-rich tests have significant potential for misidentifying students into special education when they are factored into students’ special education eligibility profiles.

Research suggests that test biases and validity issues are highly plausible factors related to the overrepresentation of ELLs in special education. These findings suggest a need to develop assessment measures that fairly, validly, and reliably assess the needs, abilities, and achievements of ELLs with special needs. Additionally, the assessment research suggests that multiple measures that create comprehensive student assessment profiles should be used to assess ELL students for special education eligibility and placement so that the students’ educational future is not determined by a single assessment measure (Rueda and Windmueller, 2006). The practice of developing assessment profiles that include both standardized and holistic community specific measures is a viable option for decreasing misidentification of ELLs for special education placement.

Language and Cultural Considerations in Pre-referral and RTI

Assessment is not the only step in the special education referral process that may contribute to inappropriate placement and overrepresentation of English learners in special education services. In response to the ineffectiveness of the discrepancy model for special education eligibility, the RTI process has been established. RTI focuses on pre-referral interventions to address the overrepresentation of minority groups in special education groups (Fuchs *et al.*, 2003). RTI is defined as the practice of providing high-quality instruction/intervention matched to student needs and using learning rate over time and level of performance in advance of formal referral to make important educational decisions about special education eligibility and service provision. That is, instead of relying on a score as an index of or predictor of learning potential, RTI assesses actual response to high-quality treatment or intervention as a pre-referral metric.

Recognizing the potential of pre-referral intervention, researchers, practitioners, and policymakers established RTI to systematically improve and articulate levels of support and intervention provided to students in at-risk conditions (Artiles and Klinger, 2003). The IDEIA legislation requires educators to measure students' RTI before referring the students for special education assessment and services.

A widely used model for RTI includes a three-tiered process. The first tier consists of quality instruction in a general education classroom based on evidence-based practices; the second tier is implemented only with students who do not reach expected benchmarks using an assessment instrument; and the third tier refers to students that are going to engage in the special education referral and assessment processes. While RTI is designed to alleviate disproportionate representation of students of color and English language learners in special education by increasing opportunities to learn, it may at the same time be subject to the same challenges related to cultural and linguistic differences that have plagued the special education system historically (Klinger *et al.*, 2005).

The National Center for Culturally Responsive Educational Systems suggests that RTI implementers consider the following four critical areas in designing and delivering educational intervention: (1) culturally meaningful task criteria, (2) teacher–student shared understandings of the purpose of tasks and activities, (3) culturally inclusive participation frameworks in classroom discourse, and (4) deficit-based ideologies about low-income racial minority students used in referral and placement practices. These considerations are sparsely evident in many RTI programs or in the RTI research. It has been noted that practitioners and researchers are in conflict around effective interventions in RTI (NCCRESt, 2005). In general, practitioners

conducting RTI use a problem-solving approach to intervention, while researchers often favor the use of standardized treatment protocols.

Intervention Strategies and Cultural Appropriateness

Not only are many special education assessment practices inappropriate for ELLs, there are also problems with intervention and instructional practices. In an effort to alleviate inequity in RTI-related services, Gersten and Dimino (2006) recommend that the three RTI tiers be fully implemented with the focus on preventing referral to special education by identifying at-risk factors in students early on in the process. They propose using a combination of national benchmarks and local and classroom norms to identify students at risk for failure. Gersten and Dimino describe the need for immediate, specific one-on-one small group intervention aligned with data that teachers consistently monitor in general education classrooms.

In addition, appropriate intervention with ELLs requires training nonspecial education teachers to recognize risk factors in students and to have proficiency in a range of strategies for providing remediation for at-risk learners. Often, teachers' beliefs, expectations for students, skills sets, and efficacy are challenged at levels that they are not prepared for in their teacher preservice preparation program (Tyler *et al.*, 2004). This situation may contribute to the disproportionality in special education noted earlier in this article.

In addition to the potential problems associated with pre-referral (RTI) intervention, researchers note that intervention strategies in full service special education are often culturally and linguistically inappropriate for ELLs. Inflexible implementation of commercial and undifferentiated instruction is of primary concern to researchers in their quests for finding effective intervention strategies for ELLs in intervention programs. Unfortunately, there are relatively few intervention programs specifically designed for ELLs with special needs. Yet even these efforts to deliver early and individualized help may provide benefit to these students.

As an example, Williams (2005) conducted a study using a reading-related intervention often found in RTI Tier 3 (as described earlier in this article), focused on text structure and gradual introduction of reading comprehension strategies. The content was introduced to learners in small increments, moving from simpler to complex while providing (1) modeling by the teacher, (2) scaffolding that faded as instruction progressed, (3) substantial guided followed by unguided practice with texts specifically structured for instruction with feedback, and (4) a thematic along with feedback instructional approach. The goal for this intervention was the development of automaticity of

the comprehension strategies in addition to transfer of the strategies to real life situations. This study revealed significant improvements in reading scores for both Hispanic and African American students, with the resulting treatment effect sizes between 1.64 and 4.06. This finding is consistent with other studies on this issue.

While reading has been identified as a primary intervention need for ELLs, many ELLs also receive mathematics instruction through RTI or through ongoing special education service provision. Fuchs and Fuchs (2005) identified mathematical problem solving (MPS) as a primary instructional need for students at risk for special education identification. They identify skill transfer as a primary difficulty and suggest that many math interventions have focused on basic skills rather than real life problem solving. Fuchs and Fuchs designed an intervention drawing on schema-based instruction and self-regulation strategies to address the transfer problems in mathematics problem solving. These math intervention strategies have yielded promising results.

As previously stated, intervention studies do not have to focus specifically on the needs of ELLs. This is quite unfortunate as the recent NCES statistics (2005) reveal that ELLs continue to be overrepresented in special education and in the special education pre-referral process. While some small-scale studies reveal promising intervention results for ELLs, few studies have focused on the unique cultural and linguistic needs of ELLs in designing interventions. Additionally, at the pre-referral stage in the special education system, the bulk of responsibility for identification and intervention is placed on general education teachers who have little, if any, training at preservice or in-service levels in working with and assessing learners with special needs who are ELLs.

Teacher Qualification, Beliefs, Expectations, and Practices Associated with ELLs in Special Education

Teacher education research suggests that teachers' beliefs influence instructional practices and influence the expectations that they have for students. Other factors such as teachers' self-efficacy may be important in the quality of instruction provided to students. Research has revealed that many teachers are often uncertain about how to determine whether ELLs are having difficulties due to language differences or due to learning challenges. For example, Obiakor (1999) identified low teacher expectation for ELL students as a significant factor in inappropriate referral to special education. Obiakor suggested that teachers' beliefs about the abilities and needs of ELL students inform their expectations and in turn inform their instructional practices. These beliefs and assumptions are informed both by personal experiences as well as by formal preservice and

in-service teacher education. A significant limitation of teacher training programs nationally is that they overemphasize knowledge acquisition to the detriment of equipping teachers with practical skills for teaching to a diverse range of students, including those with disabilities (Edelen-Smith *et al.*, 1993). Additionally, in most states, preservice teacher preparation programs require only one course in special education (NCATE, 2002), and special educators are often not required to take courses on meeting the needs of ELLs (NCATE, 2002). The shortcomings of teacher preparation programs contribute to the challenges faced by teachers in their professional educational practices. As an example, Richardson *et al.* (1989) studied two elementary schools and investigated the reasons for teachers referring students for special education assessment. They concluded that often referrals are more a reflection of teacher stress and inadequate training than a result of the students they teach having an identified learning disability.

Tyler *et al.* (2004) suggest that general education teachers often interpret culturally diverse students' performance through European American middle-income normative parameters of competence. Because culturally diverse students' performance does not always align with such parameters, it is often regarded as deficient. In many cases, a deficit perspective perpetuates low expectations for ELL students by teachers, and likely contributes to the overrepresentation of ELLs in special education.

Summary and Conclusions

While progress has been made in reducing ELLs' overrepresentation in special education, it remains a problem. Few interventions have been designed specifically to address the linguistic and cultural needs of ELLs at both the pre-referral and service provision levels. Also problematic are teacher preparation programs and teachers' misunderstanding and misattribution of the needs of ELLs. These challenges have existed for nearly four decades and have not been comprehensively addressed. It appears that without significant changes in the special education system, the needs of ELLs will continue to be unmet.

Recommendations for Research and Practice

Significant changes are needed in order to meet the needs of ELLs. Particularly needed are strategies to address the needs of ELLs at the pre-referral stages and in special education. However, the issues are multidimensional (Rueda and Windmueller, 2006) and must be addressed from various avenues. It is critical to address the training needs of general education teachers at both in-service and preservice levels

so that they fully understand the needs of ELLs in their classroom. The language, cultural, and socioeconomic-related dimensions of students' experiences need to be considered in special education referral processes. This requires a significant overhaul of accreditation standards for teacher preparation programs so that colleges and universities are required to include comprehensive coursework and meaningful supervised preservice practice for all general and special education teacher candidates. This may help address the challenges associated with over-referral to the special education system. Training related to instructional strategies should be provided to effectively teach ELLs using culturally, linguistic, and developmentally appropriate practices. Training of special education teachers is needed in this area because statistics reveal that ELLs in special education are less likely to exit special education once placed in this system.

Additionally, assessments that are psychometrically appropriate for validly assessing ELLs are needed. Culturally and linguistically appropriate pre-referral strategies must be developed to address the RTI requirements of the new special education legislation so that ELLs can fully benefit from RTI approaches. The legislative mandates associated with monitoring disproportionality statistics and action plans associated with changing the disproportionality must be fully implemented (IDEA, 2004). These recommendations are critical for ELLs to be insured educational equity.

See also: English Language Learners with Special Needs; Overrepresentation of Culturally and Linguistically Diverse Students in Special Education; Prereferral Strategies.

Bibliography

- Abedi, J. (2006). Psychometric issues in ELL assessment and special education eligibility. *Teachers College Record* **108**(11), 2282–2303.
- Artiles, A. J. and Klingner, J. K. (2003). When should bilingual students be in special education? *Educational Leadership* **61**(2), 5–12.
- Artiles, A. J., Rueda, R., Salazar, J. J., and Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Council for Exceptional Children* **71**(3), 283–300.
- Artiles, A. J., Trent, S. C., and Palmer, J. (2004). Culturally diverse students in special education: Legacies and prospects. In Banks, J. A. and Banks, C. M. (eds.) *Handbook of Research on Multicultural Education*, 2nd edn, pp 716–735. San Francisco, CA: Jossey-Bass.
- Bailey, A. L. and Butler, F. A. (2004). Ethical considerations in the assessment of the language and content knowledge of U.S. school-age English learners. *Language Assessment Quarterly: An International Journal* **1**(2), 177–193.
- Edelen-Smith, P., Prater, M., and Sileo, T. (1993). The impact of current issues in teacher education on the preparation of special educators. *Issues in Special Education and Rehabilitation* **5**(1), 7–15.
- Fuchs, D. and Fuchs, L. S. (2005). Enhancing mathematical problem solving for students with disabilities. *Journal of Special Education* **39**(1), 45–57.
- Fuchs, D., Mock, D., Morgan, P. L., and Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research and Practice* **18**(3), 157–171.
- Gersten, R. and Dimino, J. A. (2006). RTI (response to intervention): Rethinking special education for students with reading difficulties (yet again). *Reading Research Quarterly* **41**(1), 99–108.
- Gonzalez, V., Brusca-Vega, R., and Yawkey, T. (1997). *Assessment and Instruction of Culturally and Linguistically Diverse Students with or at Risk of Learning Problems: From Research to Practice*. Needham Height, MA: Allyn and Bacon.
- IDEA (Individuals with Disabilities Education Act) (2004). *20 U.S.C. 1400 et. Seq. (Statute). 34 C.F.R. 300 (Regulations)*. <http://www.wrightslaw.com/idea/law.htm>.
- Klinger, J. K., Artiles, A. J., Kozleski, E., et al. (2005). Addressing the disproportionate representation of culturally and linguistically diverse students in special education through culturally responsive educational systems. *Educational Policy Analysis Archives* **13**(38), 1–42.
- Macswan, J. and Rolstad, K. (2006). How language proficiency tests mislead us about ability: Implications for English language learner placement in special education. *Teacher's College Record* **108**(11), 2304–2328.
- McCardle, P., Mele-McCarthy, J., Cutting, L., Leos, K., and D'Emilio, T. (2005). Learning disabilities in English language learners: Identifying the issues. *Learning Disabilities. Research and Practice* **20**(1), 1–5.
- NCATE (National Council for Accreditation of Teacher Education) (2002). *Professional Standards for the Accreditation of Schools, Colleges and Departments of Education: 2002 Edition*. Washington, DC: NCATE – The Standards of Excellence in Teacher Preparation.
- NCCRESt (National Center for Culturally Responsive Educational Systems) (2005). *Cultural Considerations and Challenges in Response-to-Intervention Models: An NCCRESt Position Statement*. Tempe, AZ: NCCRESt.
- Obiakor, F. E. (1999). Teacher expectations of minority exceptional learners: Input on "accuracy" of self-concepts. *Exceptional Children* **66**(1), 39–53.
- Rhodes, R. L., Ochoa, S. H., and Ortiz, S. O. (2005). *Assessing Culturally and Linguistically Diverse Students: A Practical Guide*, pp 104–127. New York: Guilford.
- Richardson, V., Casanova, U., Placier, P., and Guilfoyle, K. (1989). *School Children at-Risk*. New York: Falmer.
- Rueda, R. and Windmueller, M. P. (2006). English language learners, LD, and overrepresentation: A multiple-level analysis. *Journal of Learning Disabilities* **39**(2), 99–107.
- Tyler, N. C., Yzquierdo, Z., Lopez-Reyna, N., and Flippin, S. S. (2004). Cultural and linguistic diversity and the special education workforce: A critical overview. *Journal of Special Education* **38**(1), 22–38.
- US Department of Education, National Center for Education Statistics (2005). *The Condition of Education 2005 (NCES 2005–094)*. Washington, DC: US Government Printing Office.
- US Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs (2007). *Annual Report to Congress on the Individuals with Disabilities Education Act, Part D*, Washington, DC.
- Williams, J. P. (2005). Instruction in reading comprehension for primary grade students: A focus on text structure. *Journal of Special Education* **39**(1), 6–18.

Further Reading

- Artiles, A. J., Rueda, R., Salazar, J. J., and Higareda, I. (2002). English-language learner representation in special education in California urban school districts. In Losen, D. J. and Orfield, G. (eds.) *Racial Inequality in Special Education*, pp 117–136. Cambridge, MA: Harvard Education Press.
- Artiles, A. J. and Trent, S. C. (1994). Overrepresentation of minority students in special education: A continuing debate. *Journal of Special Education* **27**, 410–437.
- Artiles, A. J. and Zamora-Duran, G. (1997). Disproportionate representation: A contentious and unresolved predicament. In Artiles, A. J. and Zamora-Duran, G. (eds.) *Reducing Disproportionate*

- Representation of Culturally Diverse Students in Special and Gifted Education*, pp 1–6. Reston, VA: Council for Exceptional Children.
- August, D. and Shanahan, T. (eds.) (2006). *Developing Literacy in Second-Language Learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Erlbaum.
- Donovan, S. and Cross, C. (2002). *Minority Students in Special and Gifted Education*. Washington, DC: National Academy Press.
- Harry, B. (1994). *The Disproportionate Representation of Minority Students in Special Education: Theories and Recommendations*. Alexandria, VA: National Association of State Directors of Special Education.
- Hosp, J. L. and Reschly, D. J. (2004). Disproportionate representation of minority students in special education: Academic, demographic, and economic predictors. *Exceptional Children* **70**(2), 185–199.
- Klingner, J. K. and Harry, B. (2006). The special education referral and decision-making process for English language learners: Child study team meetings and placement conferences. *Teachers College Record* **108**(11), 2247–2281.
- Lesaux, N. L. (2006). Building consensus: Future directions for research on English language learners at risk for learning disabilities. *Teachers College Record* **106**(11), 2406–2438.
- Markowitz, J., Garcia, S. B., and Eichelberger, J. (1997). *Addressing the Disproportionate Representation of Students from Racial and Ethnic Minority Groups in Special Education: A Resource Document*. Alexandria, VA: National Association of State Directors of Special Education.
- National Center for Educational Statistics (2007). *NAEP Data Explorer. Reading Assessments*. Washington, DC: US Department of Education, Institute of Educational Sciences.
- NCLB (2002). No Child Left Behind Act of 2001. Pub. L. No.107-110.
- Ortiz, A. A. (1997). Learning disabilities occurring concomitantly with linguistic differences. *Journal of Learning Disabilities* **30**(4), 321–332.
- US Census Bureau (2003). American Fact Finder. http://factfinder.census.gov/servlet/BasicFactsServlet?_lang=en (accessed August 2009).
- US Department of Education (2003). *US Department of Education 27th Annual Report to Congress*.
- Vaughn, S., Linan-Thompson, S., and Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Exceptional Children* **69**(4), 391–409.

Relevant Websites

- <http://www.cec.sped.org> – Council for Exceptional Children.
- <http://www.nccrest.org> – National Center for Culturally Responsive Educational Systems.
- <http://www.ed.gov> – US Department of Education, Office of Special Education and Rehabilitative Services.

Functional Behavioral Assessment

J P Bakken, Illinois State University, Normal, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Antecedent – An event that precedes the behavior.

Assessment – The act of judging or assessing a person or situation or event.

Behavior – The manner in which one behaves; the actions or reactions of a person or animal in response to external or internal stimuli.

Consequence – An event that follows a behavior.

Discipline – Punishment inflicted by way of correction and training.

Intervention – This is provided to improve an academic or behavioral situation; it can also be referred to as strategies.

Observable – Capable of being or liable to be observed; noticeable; visible.

Functional behavior assessment (FBA) is a systematic method of “generating information on events preceding and following behavior in an attempt to determine which antecedents and consequences are reliably associated with the occurrence of the behavior” (Miltenberger, 1997: 563). FBA is the process of identifying the events that reliably predict and maintain problem behaviors before an intervention is determined (Scott *et al.*, 2003). The purpose of functional assessment information is to improve the effectiveness and efficiency of behavior support (Horner, 2000). Functional assessment is a key aspect of behavior intervention strategies mandated for IEPs by the 1997 and 2004 reauthorizations of IDEA. Though its primary research base comes from studies and clinical use with individuals with severe disabilities, such as mental retardation (cognitive disability) and autism, more recent uses of FBA have focused on a broader range of individuals with problem behavior, including individuals with mild or moderate disabilities or those at risk for such disabilities.

Because FBA produces no numerical score or other measures for comparison or ranking, the major purpose of FBA is to inform practitioners regarding effective intervention options. At the individual student level, this connection between FBA and effective intervention has been well documented in both general and special education settings (e.g., Blair *et al.*, 1999; Jones *et al.*, 2000; Lee *et al.*, 1999; Meyer, 1999). FBA typically involves first defining a student’s problem behavior, followed by identifying predictable antecedent and consequence events, and, finally, developing a testable statement or hypothesis

of the function that problem behavior serves for him/her (O’Neill *et al.*, 1997).

FBA is generally considered to be a problem-solving process for addressing student problem behavior. It relies on a variety of techniques and strategies to identify the purposes of specific behavior and to help IEP teams select interventions to directly address the problem behavior. FBA is likely to be more simplified and team-based, making use of questionnaires, checklists, and team-based discussions (Conroy *et al.*, 1999). In either case, analyses focus effectively on relationships between individual student’s behavior and the environment, including the time of day, location, context, curricula, teacher, peers, and all other potentially relevant stimuli (Ryan *et al.*, 2003). FBA gathers broad and specific information in order to better understand the precise reason for a student’s problem behavior. A component of the FBA is an analysis of the information that has been gathered that results in predictions or hypotheses of what maintains the behavior. FBA looks beyond the behavior itself. The focus when conducting a FBA is on identifying significant, pupil-specific social, affective, cognitive, and/or environmental factors associated with the occurrence (and nonoccurrence) of specific behaviors. This broader perspective offers a better understanding of the function or purpose behind student behavior.

According to IDEA, a student with a disability who has an Individualized Education Program (IEP) can be disciplined in the same manner as any other student for 10 consecutive school days or less if he/she violates the school’s code of conduct. If he/she is disciplined for more than 10 consecutive school days within the same school year, school or district staff must conduct an FBA and implement a behavior intervention plan (BIP) before the 10th day, or before moving him/her to an interim alternative educational placement. In many cases, a student with an IEP will already have a BIP in place as part of his/her IEP in order to support learning. In fact, a student’s IEP must include a BIP whenever his/her behavior impedes his/her own learning or the learning of others. For students at-risk of suspension or those suspended for less than 10 days, school districts should complete an FBA and develop a behavior support plan to prevent future suspensions. An FBA provides the IEP team with critical information, including (1) interventions which have been tried and their effects, (2) possible motivation(s) underlying the student’s behavior, and (3) whether instructional and behavioral supports being provided to the student are

working. Functional assessment helps educators to understand what function the problem behavior serves for the student. This enables them to determine interventions that reduce or eliminate specific problem behavior by replacing it with acceptable behavior that serves the same purpose or function for the student (e.g., teaching a student more acceptable ways to gain peer attention). This is the broad goal of this article.

Phases of FBA

There are two phases of FBA. Phase 1 is generating hypotheses about the purposes of a child's misbehavior. To generate hypotheses (or best guesses) about what purpose a child's misbehavior serves, a teacher or consultant collects information using data collection tools. The more information collected, and the more diverse the types of information collected, the more informed the hypothesis will be. The goal is to identify the emergence of any patterns in behavior. Therefore, teachers, parents, and the student may be interviewed as well as the teacher may observe the student at various times in the school day in various classes. Phase 2 is testing those hypotheses. To test the hypotheses (or confirm our best guesses) about the purpose served by the child's misbehavior, we must make some alteration to the environment or curriculum. This requires four steps: (1) defining the misbehavior using objective words (does not finish work, rather than, lazy), (2) counting the behavior over a period of days, (3) altering some aspect of the environment or curriculum, and (4) continuing to count the behavior to see if it decreases, stays the same, or increases during alteration. If it decreases like we thought it would, then our hypothesis is confirmed. If there is no change in the behavior, then we develop another hypothesis and test it.

Focusing on Behavior

The goal of functional assessment is to deal with inappropriate behavior. Educators and service providers want to (1) rearrange events that occur before a child misbehaves (antecedents) so that they now prompt the occurrence of appropriate behavior, (2) change the consequence that comes after a behavior occurs so that the consequences are more likely to reinforce a child for performing appropriate behavior, and (3) most importantly, teach the child a replacement behavior. A replacement behavior is an appropriate behavior that the child can perform that accomplishes the same goal as the inappropriate behavior. Without teaching a child a replacement behavior, meaningful, positive changes in behavior will be difficult, if not impossible, to obtain. Students need to be taught what to do and how to behave appropriately. It is the

responsibility of the teacher (and parent) to diagnose why the problem behavior is occurring and to teach a more appropriate behavior in its place.

Problem behaviors typically fall into one or more of three general categories: (1) behavior that produces attention and other desired events (e.g., access to toys, the computer, and other desired activities), (2) behavior that allows the person to avoid or escape demands or other undesired events/activities (e.g., getting out of homework, an oral presentation, or reading out loud), and (3) behavior that occurs because of its sensory consequences (e.g., feels good or relieves pain). Students will change their behavior only when it is clear that a different response will more effectively and efficiently result in a desired outcome. Identifying the purpose of problem behaviors (what the student gains, controls, or avoids through those behaviors) can provide information that is essential to developing instructional strategies and supports to reduce or eliminate behaviors that interfere with successful classroom performance or participation. When investigating student behaviors the team will need to judge the impact of the behaviors being exhibited (see **Table 1** for questions that can be used as a way for a team to judge the significance of the behavior exhibited by the student). The team needs to decide why the behavior is occurring. For example, if a student acts out when given an assignment in math, it may not be the assignment that caused the acting-out behavior, but the fact that the student does not know what is required (a skill deficit) and thus anticipates failure or ridicule. Information of this type may be gleaned through a discussion with the student or a review of information gathered in the FBA process. This is crucial because problem behavior can have an adverse effect on learning and physical safety of students and teachers. For this reason, the actions the student displays in response to problem behavior are critical. Whether these problem behaviors disrupt instructional routines (e.g., leaving class without permission, being loud, and talking out) or pose threats to the safety of teachers and students (e.g., destroying property and physical altercation), educators are responsible for implementing interventions to reduce these behaviors (Dukes *et al.*, 2007).

Table 1 Questions to Guide team members about their behaviors

Does the student's behavior significantly differ from that of his/her classmates?
Does the behavior lessen the possibility of successful learning for the student and/or others?
Have past efforts to address the behavior using standard interventions been unsuccessful?
Does the behavior represent a skill or performance deficit, rather than a cultural difference?
Is the behavior serious, persistent, chronic, or a threat to the safety of the student or others?
If the behavior persists, is some disciplinary action likely to result?

Collecting Data

Before an FBA can be conducted, it is necessary to pinpoint the behavior causing learning or discipline problems, and to define that behavior in concrete terms that are easy to communicate and simple to measure and record. If descriptions of behaviors are vague (e.g., bad attitude), it is difficult to determine appropriate interventions. By collecting and analyzing various kinds of information about behavior that significantly disrupts the teaching and learning process, school personnel are better able to select the most appropriate interventions. It is important to obtain information on the antecedent and consequent events (i.e., events preceding or following the behavior), and past events that may influence present behavior. The information gathered in this process will be utilized to develop interventions to change and/or modify behaviors of concern and to teach new behavior patterns.

An FBA is not only for the purpose of defining and eliminating undesirable behaviors but to understand the structure and function of those behaviors in order to teach and promote effective alternatives. To a large measure, an FBA is a flexible process that is implemented when a student continues to exhibit challenging behaviors after school, and class-wide supports have been implemented. In order to complete the FBA, a clear, observable, and measureable description of the behavior of concern must be developed. The description of the behavior should include the following items: frequency (number of times), duration (how long), intensity (how often), the time between the trigger event and the behavior of concern (latency), and a description of the context in which the behavior occurs (e.g., people, setting events, activity, time, level of student participation, place, and expectations; see **Table 2** for some sample definitions).

Assessing Behaviors

It is crucial that multiple means are used to collect information about a behavior. This might include a review of student records as well as an evaluation of a sample of

the student's academic products (e.g., homework, in class assignments, and exams). In addition, the team will want to use various observation procedures, questionnaires, and interviews (e.g., teacher, parents, student, and bus driver). The team needs to decide what tools will help to better understand the causes of a specific behavior problem. Several different rating scales have been developed to try and get to the function of the behavior; however, reliability is usually poor. These scales should only be used as a starting point for systematic and direct observation of a person's behavior.

Relying strictly on interviews and rating scales should never be considered a functional assessment. A more reliable method involves directly observing the person's behavior in his/her natural environment and analyzing the behavior's antecedents (environmental events that immediately precede the problem behavior) and consequences (environmental events that immediately follow the problem behavior). Specific techniques include the record review, the scatterplot, and the ABC Observation Form. The record review is a process that can be used to gather information about a student's developmental and educational history (e.g., medical, social/emotional, communication, academic, career, and cognitive), history of the behaviors of concern, and the impact of previous interventions and supports. The purpose of a scatterplot is to identify patterns of behavior that relate to specific contextual conditions. A scatterplot is a chart or grid on which the observer records single events (e.g., number of student call outs) or a series of events (e.g., teacher requests and student responses) that occur within a given time frame (e.g., at lunch, at recess on the playground, and during teacher led reading instruction). Another way to observe student behavior is with an ABC Observation Form (Antecedent-Behavior-Consequence). This approach allows an observer to organize anecdotal or descriptive information on the student's interactions with other students and adults in such a way that patterns of behavior become clear. Possible tools to use for interviews or informant data collection are listed in **Table 3**. This is not an exhaustive list; therefore, one may have or use additional tools that would be appropriate.

Analyzing Data

Following data collection, data analysis is the next component of the FBA process. This analysis of information results in the formulation of predictions of what maintains the behavior and when and where the behavior will and will not occur (see **Table 4** for the questions the analysis should answer).

Clearly, one way to reach the conclusion about a behavior is to use a graphic tool. Using a graphic tool can help the team analyze all the compiled information.

Table 2 Example definitions of behavior

<i>Problem behavior</i>	<i>Concrete definition</i>
Doug is aggressive	Doug hits others during recess when he doesn't get his way.
Juan is disruptive	Juan makes irrelevant and inappropriate comments during class discussions.
Felipe is hyperactive	Felipe leaves his assigned area without permission, completes only parts of his independent work, and calls out answers without raising his hand.

Table 3 Data collection examples

<i>Tool</i>	<i>What it is</i>	<i>What it does</i>	<i>Uses</i>
Antecedent Behavior Consequence (ABC)	Direct observation tool. Specific event analysis. Can be analyzed from a videotape segment.	Determines what happens before a behavior occurs, what the behavior looks like during the event, and what happens after the behavior occurs.	Use when analyzing specific conditions that may be triggering or maintaining behavior.
Functional Assessment Observation Form	Direct observation tool. Multiple event recording log.	Provides information regarding predictor events, consequences, frequency, perception about how the function of the behavior is maintained, the time the behavior occurs, and patterns of behavioral occurrences.	Use when behavior occurs at a moderate to high frequency (25 times per day). Can be used with lower frequency behaviors when used as a time sampling tool.
Functional Assessment Interview	Informant interview tool.	Provides information about events that influence problem behavior.	Use when identifying settings, events, and activities.
Motivational Assessment Scale	Informant tool-analysis of specific behavior which can be used for multiple events.	Determine desired outcome or function behavior serves, interactions that are occurring, behavior patterns.	Use when developing hypotheses about behavior.
Scatterplot Analysis	Direct environment observation recording tool.	Charts baseline of information about the occurrence of the behavior within specific time increments and activities.	Use when determining when the behavior does and does not occur.
The Instructional Environmental System II	Informant/ecological observation tool multiple checklists and interview formats. Emphasis on general education classroom environment and curriculum.	Provides parent, teacher, and student perceptions about learning environment interactions and expectations. Additionally provides a direct observation format for collecting academic and behavioral data.	Use when there is a need to understand the multiple influences on a student's performance.

Table 4 Questions to answer related to the data

In which settings does the behavior occur and not occur (playground, classroom)?
What are the times when the behavior does/does not occur (during English, just prior to reading)?
Who are the individuals present when the behavior is most/least likely to occur (substitute teacher, certain students)?
What are the conditions when the behavior does/does not occur (during free play, unstructured time, in small groups)?
What activities take place prior to the behavior of concern?
What are the events or conditions that typically occur after the behavior occurs (student sent to the office)?
Who interacts with the student following a behavioral incident?
Is the student recognized when behaving appropriately?
Is there communication of expectations to the student?
Are instructions and directions for activities clear and concise?
Are the task requirements appropriate?
Does the student like or dislike the task?
Are tasks too difficult? Or too easy?
Is the physical structure of the environment appropriate?
Is an adequate level of assistance provided in order for the student to complete tasks?
Can the student communicate his/her wants and needs?
Are there medication/health concerns or needs?
Do others inadvertently reinforce the problem behavior?

This tool, called a data triangulation chart or a data triangle, provides a framework on which the team can pull together and visually compare information collected from various sources (e.g., interviews, scatterplots, and ABC charts).

Behavior Interventions

An FBA provides critical information to identify the most appropriate and relevant possible behavior interventions. This is the process of determining the cause (or function) of behavior before developing an intervention. The intervention must be based on the hypothesized cause (function) of behavior. Failure to base the intervention on the specific cause (function) very often results in ineffective and unnecessarily restrictive interventions that do not improve the problem behavior. As part of creating success, the hallmark of effective function-based behavior intervention is effective instruction (Scott *et al.*, 2001; Scott and Nelson, 1999). Effective instruction begins to identify an appropriate intervention to reduce the problem behavior. Instruction must teach the student a more appropriate way to achieve the same desired outcome.

Instruction begins with a rationale for the skill and includes clear expectations, modeling with auditory and visual examples, guided practice, independent practice, and constructive feedback (Greenwood, 2001; Lloyd *et al.*, 1997; Swanson and Hoskyn, 2001). FBA is not a simplistic event or activity. To date, questions remain about the feasibility of school-based personnel acquiring the knowledge in the FBA process and subsequently using the process to develop behavioral interventions (Scott *et al.*, 2005). Many research teams have directly compared the effects of interventions derived from FBA results to interventions selected without this information in general education settings. Results in all five studies showed little to no improvement in behavior for the non-FBA interventions and powerful effects for plans designed based on FBAs (Filter and Horner, 2006; Ingram *et al.*, 2005; Newcomer and Lewis, 2004). This research supports the use of interventions designed based on FBA results and should be developed this way.

Although the interventions teachers choose to change problem behavior are varied, they are not always effective. Ishii-Jordan (2000) found that teachers frequently select punitive interventions (e.g., threats and punishment) for students when they display behaviors that interfere with teacher routines, regardless of whether the interventions actually change the problem behavior. Unfortunately, teachers react without thinking (many times) and actions to decrease problem behaviors may actually promote or reinforce those behaviors. Critics argue that simple contingency-based interventions (rewarding desired behavior and punishing problem behavior) are unlikely to be effective with sophisticated students who engage in complex problem behavior. The state of technology for supporting students has moved away from simple consequence manipulation to more comprehensive plans that also include antecedent-based interventions designed to prevent problem behaviors (Kern *et al.*, 2002), teach functional alternative skills (O'Neill *et al.*, 1997), and promote strategies that generalize to contexts where there is no possibility of adult-delivered punishers or reinforcers (Kern and Dunlap, 1999).

Conclusion

This article began by introducing and defining FBA. Next, it discussed phases of FBA. Behavior-related data collection and analysis are discussed; and finally, interventions developed as a result of FBA are highlighted. Clearly, today's complex society calls for new ways of doing things. FBA is one critical way to discover behavior problems and intervene as needed and appropriately.

See also: Assessment Accommodations for Children with Special Needs; Diagnosis and Classification; Educating

Students with Emotional and Behavioral Disorders; Overrepresentation of Culturally and Linguistically Diverse Students in Special Education.

Bibliography

- Blair, K. C., Umbreit, J., and Bos, C. S. (1999). Using functional assessment and children's preferences to improve the behavior of young children with behavior disorders. *Behavioral Disorders* **24**, 151–166.
- Conroy, M. A., Clark, D., Gable, R. A., and Fox, J. J. (1999). Building competence in the use of functional behavior assessment. *Preventing School Failure* **43**(4), 140–144.
- Dukes, C., Rosenberg, H., and Brady, M. (2007). Effects of training in functional behavior assessment. *International Journal of Special Education* **22**(3), 163–173.
- Filter, K. J. and Horner, R. H. (2006). Functionally-indicated academic interventions for problem behavior.
- Greenwood, C. R. (2001). Science and students with learning and behavioral problems. *Behavioral Disorders* **27**, 37–52.
- Horner, R. H. (2000). Positive behavior supports. *Focus on Autism and Other Developmental Disabilities* **15**(2), 97–105.
- Ingram, K., Lewis-Palmer, T., and Sugai, G. (2005). Function-based intervention planning: Comparing the effectiveness of FBA function-based and non-function-based intervention plans. *Journal of Positive Behavior Interventions* **7**, 224–236.
- Ishii-Jordan, S. R. (2000). Behavioral interventions used with diverse students. *Behavioral Disorders* **25**(4), 299–309.
- Jones, K. M., Drew, H. A., and Weber, N. L. (2000). Noncontingent peer attention as treatment for disruptive classroom behavior. *Journal of Applied Behavior Analysis* **33**, 343–346.
- Kern, L., Choutka, C. M., and Sokol, N. G. (2002). Assessment-based antecedent interventions used in natural settings to reduce challenging behavior: An analysis of the literature. *Education and Treatment of Children* **25**, 113–130.
- Kern, L. and Dunlap, G. (1999). Assessment-based interventions for children with emotional and behavior disorders. In Repp, A. C. and Horner, R. H. (eds.) *Functional Analysis of Problem Behavior: From Functional Assessment to Effective Support*. Monterey, CA: Brooks/Cole Publishing.
- Lee, Y., Sugai, G., and Horner, R. H. (1999). Using an instructional intervention to reduce problem and off-task behavior. *Journal of Positive Behavior Interventions* **1**, 195–204.
- Lloyd, J. W., Weintraub, F. J., and Safer, N. D. (1997). A bridge between research and practice: Building consensus. *Exceptional Children* **63**, 535–538.
- Meyer, K. A. (1999). Functional analysis and treatment of problem behavior exhibited by elementary school children. *Journal of Applied Behavior Analysis* **32**, 229–232.
- Miltenberger, R. G. (1997). *Behavior Modification: Principles and Procedures*, 2nd edn. Florence, KY: Wadsworth.
- Newcomer, L. L. and Lewis, T. J. (2004). Functional behavioral assessment: An investigation of assessment reliability and effectiveness of function-based interventions. *Journal of Emotional and Behavioral Disorders* **12**, 168–181.
- O'Neill, R. E., Horner, R. H., Albin, R. W., *et al.* (1997). *Functional Assessment and Program Development for Problem Behavior: A Practical Handbook*, 2nd edn. Pacific Grove, CA: Brooks/Cole.
- Ryan, A. L., Halsey, H. N., and Matthews, W. J. (2003). Using functional assessment to promote desirable student behavior in schools. *Teaching Exceptional Children* **35**(5), 8–15.
- Scott, T. M., Liaupsin, C., and Nelson, C. M. (2001). *Behavior Intervention Planning: Using the Outcomes of Functional Behavioral Assessment*. Longmont, CO: Sopris West.
- Scott, T. M. and Nelson, C. M. (1999). Using functional assessment with challenging behaviors: Practical school applications. *Journal of Positive Behavior Interventions* **1**, 242–251.
- Scott, T. M., Nelson, C. M., and Zabala, J. (2003). Functional behavior assessment training in public schools: Facilitating systematic change. *Journal of Positive Behavior Interventions* **5**(4), 216–224.

- Scott, T. M., McIntyre, J., Liaupsin, C., *et al.* (2005). An examination of the relation between functional behavior assessment and selected intervention strategies with school-based teams. *Journal of Positive Behavior Interventions* **7**(4), 205–215.
- Swanson, H. L. and Hoskyn, M. (2001). Instructing adolescents with learning disabilities: A component and composite analysis. *Learning Disabilities Research and Practice* **16**, 109–119.

Further Reading

- Conroy, M. A. and Davis, C. A. (2000). Early elementary-aged children with challenging behaviors: Legal and educational issues related to IDEA and assessment. *Preventing School Failure* **44**(6), 163–168.
- Crone, D. A., Hawken, L. S., and Bergstrom, M. K. (2007). A demonstration of training, implementing, and functional behavior assessment in 10 elementary and middle school settings. *Journal of Positive Behavioral Interventions* **9**(1), 15–29.
- Fitzsimmons, M. K. (1998). Functional Behavior Assessment and Behavior Intervention Plans. <http://www.cec.sped.org/AM/Template.cfm?Section=Home&CONTENTID=1742&TEMPLATE=/CM/ContentDisplay.cfm> (accessed August 2009).
- IDEA (1998). *An IEP Team's Introduction to Functional Behavioral Assessment and Behavior Intervention Plans*, 2nd edn.

- http://www.fape.org/idea/what_idea_is/osher/main.htm (accessed August 2009).
- Quinn, M. M. (2000). Functional behavioral assessment: The letter and spirit of the law. *Preventing School Failure* **44**(4), 147–151.

Relevant Websites

- <http://causeonline.org> – Functional Behavior Assessment, CAUSE.
- <http://www.usu.edu> – Functional Behavior Assessment, Utah State University.
- <http://www.tcnj.edu> – Functional Behavior Assessments, The College of New Jersey.
- <http://www.behavioradvisor.com> – Functional Behavioral Assessment (FBA), BehaviorAdvisor.com.
- <http://cecp.air.org/fba> – Functional Behavioral Assessment, Center for Effective Collaboration and Practice.
- <http://www.nls.net> – Overview of Functional Behavioral Assessment, Netlink.
- <http://www.specialconnections.ku.edu> – Special Connections: An Introduction to Functional Behavioral Assessment.
- <http://www.wrightslaw.com> – Wrightslaw: Functional Behavioral Assessments: What? Why? How? Who? by Stephen Starin, PhD.

History of Special Education

L M Garrick Duhaney and S J Salend, State University of New York at New Paltz, New Paltz, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

The term special education appears to have been first used publicly by Alexander Graham Bell at a National Education Association meeting in 1884 (Winzer, 1998). However, the field of special education emerged before that event and has a long and dynamic history that has been influenced by an amalgam of international, philosophical, political, economic, legal, and sociocultural factors (Fleischer and Zames, 2001; Giordano, 2007). Although the history of special education reveals many changes (Osgood, 2007; Reynolds, 1989), it has been shaped by ongoing challenges, successes, and debates related to: (1) the purposes of special education (custodial care vs. education and segregation vs. integration); (2) the groups to be served (individuals with biologically based disabilities vs. individuals with socially constructed disabilities); (3) the identification of effective pedagogical practices and related services; and (4) the impact of special education on individuals with disabilities. While distinct from the history of people with disabilities, the history of special education also has been impacted by the ways individuals with disabilities have been treated and perceived by the larger society and culture (Smith, 1998; Winzer, 1993).

Special education today has evolved into the delivery of research-based instructional and assessment practices and related services specifically designed to promote educational, social, behavioral, and physical developments of persons with disabilities (Heward, 2006). This article outlines the main events, factors, and movements that have influenced the development of special education from its philosophical origins as a care-taking enterprise focused on deaf and blind individuals to its transformation to serving individuals with a wide range of disabilities and being an integral part of the educational system in many parts of the world.

The Origins of Special Education

Although very little record existed about persons with disabilities prior to the 1700s, by all accounts deviations from what was considered normal by society were subject to harsh and inhumane treatment, ostracism, and even death (D'Antonio, 2004; Winzer, 1993, 1998). This, however, began to change in the sixteenth and seventeenth centuries as rational philosophical beliefs about human dignity started to impact societal perceptions of individuals with disabilities and ways to educate them (Winzer, 1998). Earliest among those who tried to improve the lives of

people with disabilities were pioneer special educators and advocates who wrote about their students, their approaches, and their failures and successes (Winzer, 1993).

Philosophical Underpinnings

In the 1500s and 1600s, the education of individuals with disabilities was mainly driven by those in pursuit of philosophical knowledge (Winzer, 1993). France, where many phenomenal efforts to educate the disabled began, experienced increased philanthropic efforts related to beliefs about the essence of human nature, human reason, human rights and dignity, and self-sufficiency. This intellectual movement referred to as the Enlightenment, raised questions about conventional notions regarding sociopolitical structures and called for an egalitarian and just society that acknowledged the rights of all of its citizens (Knight, 1968; Safford and Safford, 1996; Winzer, 1993).

In England, emerging philosophical beliefs also motivated better treatment of individuals with disabilities and increased efforts to educate them (Winzer, 1993). Pioneering work on the origin and development of language led to further research by members of the Royal Society of London, a group of intellectuals and scholars (Winzer, 1998). These efforts culminated in the publication of a book by John Wallis that described the origins of language and served as an important resource guiding the education of deaf individuals (Hoolihan, 1985; Winzer, 1993).

Initial Focus on the Deaf

The earliest efforts to provide special education focused on deaf individuals (Best, 1930; Winzer, 1998). In mid-sixteenth-century Spain, Pedro Ponce de Leon, a Benedictine monk, developed oralism, which emphasized the teaching of lip-reading and speech to communicate rather than the use of signed communication, to teach wealthy deaf individuals to speak in order to obtain their inheritance (Buchanan, 1999; Burch and Sutherland, 2006; Lane, 1989; Winzer, 1998). The growth in oralism, which became the prevailing method used in deaf schools from the 1890s to the 1920s, also was supported by the original pedagogical work of Jacob Rodrigue Péreire who was considered one of the first deaf educators in France (Burch and Sutherland, 2006; Winzer, 1998). Later, the French priest Michel Charles de l'Épée who opposed Péreire's pedagogical efforts, promoted the belief that

the use of written characters and sign language was the best way to teach the deaf, which also contributed to sign language being the dominant vehicle for deaf education during the first-half of the 1800s (Winzer, 1998). Encouraged by European success in educating deaf persons, Thomas H. Gallaudet studied these methods in Europe and in 1817 established the first institution for the deaf in Hartford, Connecticut, United States of America (Osgood, 2005).

Focus on the Blind

Building on the successful efforts to educate deaf individuals, special education began to focus on the plight of the blind (Winzer, 1998). In 1784, Valentin Haüy founded the Paris Blind School and employed raised print and embossed books to teach blind students (see Winzer, 1998). In 1829, Louis Braille, who attended the Paris Blind School earlier in his life, developed his raised-dot method for reading and stylus for writing. Braille's method led to the establishment of a tactile alphabet that enabled blind individuals to read and to become more integrated members of French society (Koestler, 1976).

Effective efforts to teach blind individuals in Europe were exported to North America by advocates of the disabled (Winzer, 1993). After studying in Paris, Dr. John D. Fischer, started the New England Asylum for the Blind in 1829, which was later renamed the Perkins Institute for the Blind and is now called the Perkins School for the Blind (Fleischer and Zames, 2001; Winzer, 1993). Samuel Gridley Howe, who was appointed as the first director of the institution, worked with Laura Bridgman, a deaf-blind student. Howe successfully taught Laura Bridgman using a method that built on her ability to identify letters by distinguishing shapes, which challenged conventional beliefs that deaf-blind persons were uneducable. The successes of Howe and Bridgman also served to set the stage for the subsequent achievements of Helen Keller, a girl who was deaf and blind, and her teacher Anne Mansfield Sullivan, herself a former student at the Perkins Institute (Osgood, 2005; Smith, 1998).

The Rise of Specialized Interventions, Schools, and Institutions

There were dramatic changes in the field of special education that were related to competing beliefs about the extent to which individuals with disabilities should be feared, segregated, categorized, and educated. Whereas earlier efforts focused initially on the deaf and then the blind, the field of special education began to address issues related to individuals with cognitive disabilities. As a result, many institutions and specialized schools began to crop up and directed placement and institutional efforts.

Institutionalization

Driven by society's restrictive perceptions and fears of individuals with disabilities, particularly those persons with cognitive and emotional and behavioral difficulties, the mid-nineteenth century saw the rise of institutions and asylums for individuals with disabilities (Armstrong, 2002). Although some institutions attempted to offer educational and vocational training programs and moral and religious training (Giordano, 2007), most of these settings viewed special education as mainly having medical, vocational, and custodial purpose and serving as a means to isolate, rehabilitate, and control disabled and defective individuals who were viewed as threats to society (Armstrong, 2002; Humphries and Gordon, 1992; Winzer, 1998). In response to the humanitarian, legal, and fiscal concerns associated with institutional settings, community-based day-care and occupation centers that provided care and minimal manual and vocational training also were established (Giordano, 2007; Read and Walmsley, 2006).

Specialized Interventions, Schools, and Classes

Although institutional settings played an important role until the 1970s, the early and mid-nineteenth century and the beginning part of the twentieth century also saw the rise of specialized interventions, schools, and classes for students with disabilities in North America and Western Europe (Read and Walmsley, 2006). French physician, Jean-Marc-Gaspard Itard's work in the early 1800s with Victor, the wild boy of Aveyron, France, gained much attention. Itard described this boy as "indifferent to everything and attentive to nothing" (see Safford and Safford, 1996). This work became considered as a watershed in the development of the field of special education (Safford and Safford, 1996). Itard implemented a specialized instructional program with Victor, that fostered his cognitive and language learning and showed that individuals previously considerable uneducable could learn (Safford and Safford, 1996). Itard's successes also led other European scholars and educators (Hinshelwood, 1900; Ireland, 1877; Morgan, 1896) to publish works that presented a set of pedagogical practices for special education. The first and most influential of these publications was Édouard Seguin's *Treatise on Idiocy*, which presented unique instructional principles, strategies, and devices that served as a pedagogical model for educating individuals with disabilities (Giordano, 2007).

The writings of Itard, Seguin, and other European educators served to change society's perception of the degree to which people with disabilities could be educated and fostered, and informed efforts to provide special education via specialized schools and classes (Giordano, 2007). In addition, their work helped promote the passage of several laws related to education of students with disabilities. In the

early 1900s, France passed a law calling on local schools to institute special improvement classes for students with learning difficulties that were connected to schools serving children without disabilities (Armstrong, 2002). In 1913, Great Britain adopted the Mental Deficiency Act which established policies for defining and educating individuals with disabilities and identified the educational and governmental agencies responsible for meeting their needs (Giordano, 2007). In the long run, these laws motivated advocates of special education throughout the world and served as the forerunners of the 1970s legislation that would guarantee the educational rights of students with disabilities.

The Establishment of Advocacy Groups

The establishment of specialized schools and classes and legal initiatives in Europe, prompted families and professionals to form groups to advocate for increased educational opportunities for individuals with disabilities (Yell *et al.*, 1998). These groups included the Council for Exceptional Children, a professional organization that was founded in 1922, and the Cuyahoga County Ohio Council for the Retarded Child, one of the first coalitions of families established in 1933. Advocacy groups were successful because (1) parents got involved in the education of their children, (2) there was an increase in research on deviant behaviors, and (3) changes began to occur in teacher-preparation programs.

Intelligence Testing

The progress in educating individuals with disabilities was thwarted by the advent of intelligence testing in the early twentieth century as it led to an emphasis on more rigid notions of normality (Armstrong, 2002; Safford and Safford, 1998). Thus, intelligence testing helped promote not only the labeling of students with lower IQs as feeble minded, mentally defective, and deviant, but also the belief that these individuals were uneducable and should be segregated from others and exempted from compulsory education laws (Read and Walmsley, 2006; Yell, *et al.*, 1998). The view of intelligence as providing a genetic basis for the school success of different groups coupled with the beliefs that individuals with inferior intelligence and morality would undermine society also contributed to placement of individuals with disabilities in institutions and to the promulgation of the Eugenics movement (Bursztyn, 2007; Humphries and Gordon, 1992).

Skeels' and Dye's Research

Views of individuals with disabilities as uneducable were questioned when Skeels and Dye (1939) published landmark research that showed that children with mental retardation benefited from having a stimulating environment.

This study, coupled with related longitudinal studies, served to establish education rather than custodial care as an important purpose of special education programs. In addition, the work of Skeels and Dye demonstrated the potential of early intervention and the need to document empirically educational interventions used with students with disabilities (Morse, 2000).

The Legalization and Transformation of Special Education

The contemporary period has been marked by significant changes and growth in the field of special education (Winzer, 1993). This period has been characterized by political advocacy, which has led to the legalization of the field of special education. In turn, special education has been transformed into an integral part of the educational system and a vehicle for promoting the inclusion of the individual with disabilities into the larger society.

The Civil Rights Movement and Brown v. Topeka Board of Education

In the early 1950s, the legal precedent for much of the special education-related litigation and legislation in the United States that fueled the growth of special education was established by the civil rights movement and the 1954 case of *Brown v. Topeka Board of Education* (Blanchett *et al.*, 2005). The Supreme Court decision in this landmark civil rights case refuted the doctrine of separate but equal. The decision served as the basis for both the subsequent legal actions brought about by families to ensure that their children with disabilities received a free appropriate public education (FAPE) and the movement toward educating students with disabilities in public schools with their classmates without disabilities (Blanchett *et al.*, 2005; Morse, 2000).

Special-Education-Related Court Cases

Following the legal precedents established in *Brown v. Topeka Board of Education*, court decisions have affirmed and expanded the educational rights of individuals with disabilities. The historic outcomes in *Pennsylvania Association for Retarded Children (PARC) v. Commonwealth of Pennsylvania* 1972 established the right of individuals with disabilities to an appropriate education and gave parents of children with disabilities procedural safeguards regarding the education of their children (Hulett, 2009; Yell, 2006).

The courts also addressed issues that advanced equal educational opportunities for individuals with disabilities from culturally and linguistically diverse backgrounds in

cases such as *Diana v. California State Board of Education* (1970), *Lau v. Nichols* (1974), and *Larry P. v. Riles* (1979) (Baca and Cervantes, 2004). The consent decree in *Diana v. California State Board of Education* established that identifying Mexican-American students for placement in special education on the basis of culturally biased intelligence tests was inappropriate and should be modified to include the use of nonverbal tests, the elimination of culturally biased testing, and the administration of tests in students' primary languages. *Lau v. Nichols* extended the right to equal educational opportunity to include bilingual special education services and language programs for speakers of languages other than English and in *Larry P. v. Riles*, the court questioned the use culturally biased intelligence tests that resulted in a disproportionate number of African American students being placed in special education.

Several court cases focused on the placement of individuals with disabilities in the least-restrictive educational environment including *Daniel R. R. v. State Board of Education* (1989), *Sacramento City Unified School District Board of Education v. Rachel H.* (1994), and *Oberti v. Board of Education of the Borough of Clementon School District* (1993) (Hulett, 2009). *Daniel R. R. v. State Board of Education* established a two-prong test for determining students' placement in the general education classrooms. These are: (1) whether students can be educated in the general education classroom with supplementary aids, and (2) whether students have been included in the general education classroom to the maximum extent appropriate (Lipton, 1994). Similarly, *Oberti v. Board of Education* set forth the principle that placement in general education classrooms with instructional accommodations must be tried prior to placing students in segregated settings (Murdick *et al.*, 2007; Yell, 2006).

Advocacy and the Disability Rights Movement

In the 1960s and 1970s, the successes of the civil rights movement fostered a period of acceptance and possibilities, which led groups of individuals with disabilities, family members, and professionals to continue to form advocacy groups to overcome discrimination, segregation, and marginalization, and seek equity, opportunity, and inclusion (Giordano, 2007). Many of these groups were guided by the principle of normalization, which originated in Scandinavia, and called for the provision of educational, housing, employment, social, and leisure opportunities, for individuals with disabilities, that resemble as closely as possible the opportunities and activities enjoyed by individuals without disabilities (Wolfensberger, 1972). Advocacy efforts of these groups served as the foundation for the disability rights movement and the creation of a disability culture and disabilities studies, which affirmed and celebrated disability, and challenged society's

conventional notions of disability and the best ways to educate individuals with disabilities (Fleischer and Zames, 2001; Burch and Sutherland, 2006).

Special-Education-Related Legislation

Advocacy groups pushed for public policies that led to legislative actions intended to guide the delivery of a comprehensive set of special education services and provide individuals with disabilities with greater access to society and the public schools (Giordano, 2007; Yell, *et al.*, 1998). The enactment of the 1970 Education (handicapped children) Act in England, the *Loi d'Orientation en Faveur des Personnes Handicapées* in France in 1975, and the Education for All Handicapped Children Act in the United States in 1975 (which was renamed the Individuals with Disabilities Education Act (IDEA)) provided students with disabilities, access to public schools (Armstrong, 2002). The IDEA also mandated that students with disabilities be educated in the least-restrictive environment (LRE) and have an individualized educational plan (IEP) that guides the delivery of special education services.

International Actions

The work of advocacy groups and the passage of special education related legislation also prompted international actions that reflected a global commitment to disability rights and inclusive education (Forlin, 2008). At the World Conference on Special Needs in 1994, 92 countries and 25 international organizations endorsed the Salamanca statement, which called for the development and implementation of general education programs that educate all children and youth together. Since the adoption of the Salamanca statement, countries throughout the world have implemented various inclusive education initiatives that have been based on each country's educational philosophy and history as well as a range of social, political, cultural, and economic factors (Alur and Bach, 2008; Brown, 2005; Fletcher and Artiles, 2005; Heng and Tam, 2006; Mitchell, 2005; Mitchell and Desai, 2005). In many countries social justice and multicultural education are viewed as being inextricably linked to inclusive education, which has broadened the focus of inclusive education beyond disability to include issues of race, linguistic ability, economic status, gender, learning style, ethnicity, cultural, and religious background, family structure, and sexual orientation (Mitchell, 2005; Slee, 2005; Verma *et al.*, 2007). Building on the Salamanca statement, in 2008, the United Nations adopted a landmark pact that promotes the inclusion of individuals with disabilities throughout the world by providing them with equal access to educational, employment, and social opportunities.

Socially Constructed Disabilities and Overidentification Concerns

The passage of laws mandating the education of students with disabilities contributed to a significant increase in the numbers and noteworthy changes in the types of students served by special education. These developments were fueled by the creation of socially constructed disability categories such as emotionally disturbed and learning disabilities (Armstrong, 2002), a term that was initially used by Kirk and Bateman (1962) to describe students without sensory, physical, or severe cognitive disabilities who performed poorly in school. The creation of socially constructed disabilities and the use of unreliable tests and procedures to identify students with disabilities led to growing concerns about the overidentification of students, the disproportionate representation of students from culturally and linguistically diverse backgrounds (Artiles *et al.*, 2005; Obiakor, 2007; Ortiz, 1997), and the need for new models for identifying, assessing, categorizing, placing, and teaching individuals in need of special education (Fuchs and Fuchs, 2006; Mercer, 1973).

Inclusion in a Least Restrictive Environment

Concerns about the growth and segregated nature of special education were voiced by Lloyd Dunn (1968) in the late 1960s when he argued that special education classes for students with mild disabilities were unjustifiable because they served as a form of homogeneous grouping and tracking and limited these students' academic, social, and behavioral developments. Ongoing research questioning the efficacy of segregated special education programs beginning in the 1960s (McLeskey, 2007), legislative and judicial actions, and advocacy groups also prompted the field of special education to initially focus on mainstreaming, the regular education initiative (REI), and then the implementation of inclusion programs that educate all students together in the general education classroom (Osgood, 2005; Salend, 2008). Inclusion is a relatively recent movement and therefore, research that examines its implementation, effectiveness, and long-term impact continues to be a challenge for the field (Sindelar *et al.*, 2006). Clearly, the driving force behind inclusion is the LRE, the mandate of the law.

Distinctive and Empirical Pedagogy

Consistent with the field's legacy of searching for effective practices, this contemporary period has experienced a concerted effort to identify a distinctive and empirically based pedagogy (i.e., evidence-based method) associated with the field of special education (Giordano, 2007). The 1960s and 1970s saw controversies over effective models (e.g., the medical model vs. the behavioral model) and pedagogical practices (e.g., perceptual and modality training,

dietary changes, and motor patterning) (Mostert and Crockett, 2000; Van Acker, 2006). The inclusive education movement led special education researchers to debate the efficacy of general education placements for students with disabilities (McLeskey, 2007) and to examine culturally responsive instructional practices (Obiakor, 2007). The technological advances of the late twentieth and early twenty-first centuries also resulted in the development of a wide range of assistive and instructional technologies that have become integral parts of special education practices and transformed notions of disability (Blackhurst, 2005; Edyburn, 2003). The growing body of research on effective pedagogical practices have inevitably led to the rise of special education teacher education programs designed to increase the quantity and quality of educational personnel who work with students with disabilities (Kleinhammer-Tramill and Fiore, 2003).

Conclusion

Viewed through historical lenses in Western Europe and North America, the field of special education has an extensive and vibrant past that has been fashioned by a myriad of philosophical, sociopolitical, economic, and legal factors intricately tied to society's treatment of individuals with disabilities. Marked by ongoing debates about purposes, groups served, effective practices, and outcomes, the field of special education has experienced periods of rapid expansion and has evolved from merely care and segregation to education and inclusion. However, since much of the recorded history related to special education has focused on developments in Western Europe and North America and has been written from the perspectives of professionals, historical accounts from other parts of the world and from the viewpoints of individuals with disabilities would enrich the field of special education by broadening our collective understanding of the field's past roots, connections to present practices, and implications for future development.

See also: Early Intervention for Students with Special Needs; Educating Students with Special Needs: An Overview; Inclusion of Students with Special Needs in General Education Classrooms; Overrepresentation of Culturally and Linguistically Diverse Students in Special Education; Parent and Family Involvement in the Education of Children with Special Needs.

Bibliography

- Alur, M. and Bach, M. (2008). *The Journey for Inclusive Education in the Indian Sub-Continent*. London: Routledge.
- Armstrong, F. (2002). The historical development of special education: Humanitarian rationality or wild profusion of entangled events? *History of Education* 31(5), 437–456.

- Artiles, A. J., Rueda, R., Salazar, J. J., and Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children* **71**, 283–300.
- Baca, L. M. and Cervantes, H. T. (2004). *The Bilingual Special Education Interface*, 4th edn. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Best, H. (1930). Educational provisions for the deaf, the blind, and feeble-minded compared. *American Annals of the Deaf* **75**, 239–240.
- Blackhurst, A. E. (2005). Historical perspective about technology applications for people with disabilities. In Edyburn, D., Higgins, K., and Boone, R. (eds.) *Handbook on Special Education Technology Research and Practice*, pp 3–29. Whitefish Bay, WI: Knowledge by Design.
- Blanchett, W. J., Brantlinger, E., and Shealey, M. W. (2005). Brown 50 years later – exclusion, segregation and inclusion. *Remedial and Special Education* **26**(2), 66–69.
- Blanchett, W. J., Mumford, V., and Beachum, F. (2005). Urban school failure and disproportionality in a post-Brown era: Benign neglect of the constitutional rights of students of color. *Remedial and Special Education* **26**, 70–81.
- Brown, R. C. (2005). Inclusive education in Middle Eastern cultures: The challenge of tradition. In Mitchell, D. (ed.) *Contextualizing Inclusive Education: Evaluating Old and New*, pp 253–278. London: Routledge.
- Buchanan, R. (1999). *Illusions of Equality*. Washington, DC: Gallaudet University Press.
- Burch, S. and Sutherland, I. (2006). Who's not here yet? American disability history. *Radical History Review* **94**, 127–147.
- Bursztyn, A. M. (2007). A brief history of intelligence testing. In Bursztyn, A. M. (ed.) *The Praeger Handbook of Special Education*, pp 3–5. Westport, CT: Praeger.
- Daniel R. R. v. State Board of Education (1989). 874 F. 2d 1036, 5th Circuit.
- D'Antonio, M. (2004). *The State Boys Rebellion*. New York: Simon and Schuster.
- Diana v. California State Board of Education (1970). No. C-70-37, RFP, (N.D. Cal., 1970).
- Dunn, L. (1968). Special education for the mildly retarded: Is much of it justifiable? *Exceptional Children* **35**, 5–22.
- Edyburn, D. L. (2003). Technology in special education. In Roblyer, M. D. (ed.) *Integrating Educational Technology into Teaching*, 3rd edn., pp 315–333. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Fleischer, D. Z. and Zames, F. (2001). *The Disability Rights Movement: From Charity to Confrontation*. Philadelphia, PA: Temple University Press.
- Fletcher, T. and Artiles, A. A. (2005). Inclusive and equity in Latin America. In Mitchell, D. (ed.) *Contextualizing Inclusive Education: Evaluating Old and New*, pp 166–201. London: Routledge.
- Forlin, C. (2008). Educational reform to include all learners in the Asia-Pacific region. In Forlin, C. (ed.) *Catering for Learners with Diverse Needs: An Asia-Pacific Focus*, pp 3–10. Hong Kong: Hong Kong Institute of Education.
- Fuchs, D. and Fuchs, L. S. (2006). Introduction to response to intervention: What, why and how valid is it? *Reading Research Quarterly* **41**(1), 92–99.
- Giordano, G. (2007). *American Special Education: A History of Early Political Advocacy*. New York: P. Lang.
- Heng, M. A. and Tam, K. Y. (2006). Special education in general teacher education programs in Singapore. *Teacher Education and Special Education* **29**, 149–156.
- Heward, W. L. (2006). *Exceptional Children: An Introduction to Special Education*, 8th edn. Upper Saddle River, NJ: Merrill/Prentice-Hall.
- Hinshelwood, J. (1900). Congenital word-blindness. *Lancet* **1**, 1506–1508.
- Hoolihan, C. (1985). Too little, too soon: The literature of deaf education in 17th century Britain. *Volta Review* **87**, 28–44.
- Hulett, K. E. (2009). *Legal Aspects of Special Education*. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Humphries, S. and Gordon, P. (1992). *Out of Sight: The Experience of Disability, 1900–1950*. Plymouth: Northcote House.
- Ireland, W. W. (1877). *On Idiocy and Imbecility*. London: Churchill.
- Kirk, S. and Bateman, B. (1962). Diagnosis and remediation of learning disabilities. *Exceptional Children* **29**, 73–78.
- Kleinhammer-Tramill, J. and Fiore, T. A. (2003). A history of federal support for preparing special educators and related services personnel to serve children and youth with disabilities. *Teacher Education and Special Education* **26**(3), 217–229.
- Knight, I. (1968). *The Geometric Spirit: The Abbé Condillac and the French Enlightenment*. New Haven, CT: Yale University Press.
- Koestler, F. A. (1976). *The Unseen Minority: A Social History of Blindness in America*. New York: David McKay.
- Lane, H. (1989). *When the Mind Hears: A History of the Deaf*. New York: Vintage.
- Larry P. v. Riles (1979). 495 F. Supp. 926 (N.D. Cal. 1979).
- Lau v. Nichols (1974). 414 U.S. 563 (1974).
- Lipton, D. (1994). The “full inclusion” court cases: 1989–1994. *National Center on Education Restructuring and Inclusion Bulletin* **1**(2), 1–8.
- McLeskey, J. (2007). *Reflections on Inclusion: Classic Articles That Shaped Our Thinking*. Arlington, VA: Council for Exceptional Children.
- Mercur, J. R. (1973). *Labeling the Mentally Retarded*. Berkeley, CA: University of California Press.
- Mitchell, D. (ed.) (2005). *Contextualizing Inclusive Education: Evaluating Old and New International Paradigms*. London: Routledge.
- Mitchell, D. and Desai, I. (2005). Diverse socio-cultural contexts for inclusive education in Asia. In Mitchell, D. (ed.) *Contextualizing Inclusive Education: Evaluating Old and New*, pp 202–232. London: Routledge.
- Morgan, W. P. (1896). A case of congenital word blindness. *British Medical Journal* **2**, 1378.
- Morse, T. E. (2000). Ten events that shaped special education's century of dramatic change. *International Journal of Educational Reform* **9**(1), 32–38.
- Mostert, M. P. and Crockett, J. B. (2000). Reclaiming the history of special education for more effective practice. *Exceptionality* **8**, 133–143.
- Murdick, N. L., Gartin, B. C., and Crabtree, T. L. (2007). *Special Education Law*, 2nd edn. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Oberti v. Board of Education of the Borough of Clementon School District (1993). 995 F.2d, 1009, 3rd Circuit.
- Obiakor, F. E. (2007). *Multicultural Special Education: Culturally Responsive Teaching*. Columbus, OH: Merrill/Prentice Hall.
- Ortiz, A. A. (1997). Learning disabilities occurring concomitantly with linguistic differences. *Journal of Learning Disabilities* **30**, 321–332.
- Osgood, R. L. (2005). *The History of Inclusion in the United States*. Washington, DC: Gallaudet University Press.
- Osgood, R. L. (2007). *The History of Special Education: A Struggle for Equality in American Public Schools (Growing up: History of Children and Youth)*. Westport, CT: Praeger.
- Pennsylvania Association for Retarded Children v. Commonwealth of Pennsylvania (1972). 343 F. Supp. 279 (E.D. Pa., 1972).
- Read, J. and Walmsley, J. (2006). Historical perspectives on special education, 1890–1970. *Disability and Society* **21**(5), 455–469.
- Reynolds, M. C. (1989). An historical perspective: The delivery of special education to mildly disabled and at-risk students. *Remedial and Special Education* **10**(6), 7–11.
- Sacramento City Unified School District, Board of Education v. Rachel H. (1994). 14F.3d, 1398, 9th Circuit.
- Safford, P. S. and Safford, E. J. (1996). *A History of Childhood and Disability*. New York: Teachers College Press.
- Safford, P. L. and Safford, E. J. (1998). Visions of the special class. *Remedial and Special Education* **19**(4), 229–238.
- Salend, S. J. (2008). *Creating Inclusive Classrooms: Effective and Reflective Practices*, 6th edn. Columbus, OH: Merrill/Prentice Hall.
- Sindelar, S. D. K., Yendol-Hoppey, D., and Liebert, T. W. (2006). The sustainability of inclusive school reform. *Exceptional Children* **72**, 317–331.
- Skeels, H. and Dye, H. A. (1939). A study of the effects of differential stimulation on mentally retarded children. *Proceedings of the American Association on Mental Deficiency* **44**, 114–136.
- Slee, R. (2005). Education and the politics of recognition: Inclusive education—An Australian snapshot. In Mitchell, D. (ed.) *Contextualizing Inclusive Education: Evaluating Old and New*, pp 139–165. London: Routledge.

- Smith, J. D. (1998). Histories of special education: Stories from our past, insights for our future introduction to the special series. *Remedial and Special Education* **19**(4), 196–200.
- Van Acker, R. (2006). Outlook on special education practice. *Focus on Exceptional Children* **38**(6), 8–18.
- Verma, G. K., Bagley, C., and Jha, M. (eds.) (2007). *International Perspectives on Educational Diversity and Inclusion*. London: Routledge.
- Winzer, M. A. (1993). *The History of Special Education: From Isolation to Integration*. Washington, DC: Gallaudet University Press.
- Winzer, M. A. (1998). A tale often told: The early progression of special education. *Remedial and Special Education* **19**(4), 212–218.
- Wolfensberger, W. (1972). *The Principle of Normalization in Human Services*. National Institute on Mental Health.
- Yell, M. L. (2006). *The Law and Special Education*, 2nd edn. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Yell, M. L., Rodgers, D., and Rodgers, E. L. (1998). The legal history of special education. *Remedial and Special Education* **19**(4), 219–229.

Further Reading

- Winzer, M. A. (1986). Early developments in special education: Some aspects of enlightenment thought. *Remedial and Special Education* **7**(5), 42–49.

Relevant Websites

- <http://www.disabilitymuseum.org> – Disability History Museum.
- <http://www.americanhistory.si.edu> – Disability Rights Movement Virtual Museum.
- <http://www.disabilityhistory.co.uk> – The Disability History Group.
- <http://www.disabilityhistory.org> – The Disability Social History Project.

Inclusion of Students with Special Needs in General Education Classrooms

M A Prater, Brigham Young University, Provo, UT, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Access to the general-education curriculum –

The meaningful interaction with instruction and curriculum used in general education by all students. It is also interpreted to include access to school activities such as lunch and recess breaks, as well as extracurricular activities.

Inclusion – In the context of schooling, a belief or philosophy that all students should be included in the general-education classroom and have access to the general-education curriculum.

Least-restrictive environment – A legal principle that students with disabilities are to be educated with students without disabilities to the maximum extent possible.

Mainstreaming – A common term for including students with disabilities in the general-education classroom to the extent possible.

In the context of schooling, inclusion is a belief or philosophy that all students should be included in the general-education classroom and have access to the general-education curriculum. Some professionals differ, however, in what is meant by all students; some add the phrase, all students to the extent possible. This allows for those with very specialized needs, such as those with chronic and significant behavioral problems, to receive services outside of the general classroom.

Some professionals distinguish between the terms inclusion and full inclusion. Full inclusion means literally all students are included with their typically developing peers regardless of services they may need; and inclusion means all students are included in general education to the extent possible. In full inclusion, services are brought to the student in the general classroom, rather than the student being removed from the classroom to receive such services. Therefore, if the student cannot meet expectations of the general classroom, these expectations are altered rather than the student removed to a different setting. Inclusive schools welcome all students and view the education of all students as the responsibility of all educators.

Historical Perspective

In the United States, prior to the passage of P.L. 94-142 in 1975 mandating education for students with disabilities, students could be denied access to public education due to their disabling condition. In 1990, the law was re-authorized as the Individuals with Disabilities Education Act or IDEA. To help compensate for previous neglect, the principle of least-restrictive environment (LRE) was included in the legislation. LRE meant that students with disabilities were to be educated with students without disabilities to the maximum extent possible. LRE is considered to be a placement issue. Removing students with disabilities from the general educational environment was to occur only when the nature or severity of the disability was so great that supplementary aids and services in general classrooms were insufficient. Putting LRE into practice implied that school districts must provide a continuum of services for students with disabilities. This continuum of services included teaching students with disabilities solely in general-education classrooms to students confined to their home or a hospital and everything in between. Students with disabilities, therefore, could be pulled out of general classrooms for portions of the day or educated in separate classrooms or special schools. The decision about where to educate a student was designed to be an individual decision based on where the needs of the student could best be met.

Although the term, mainstreaming, was never adopted in the US laws governing special education, soon after P.L. 94-142 was enacted, the term became synonymous with educating students with disabilities with their nondisabled peers whenever possible. Often such mainstreaming occurred in nonacademic classes such as art, music, and physical education while students were still segregated for academic or life-skills instruction. Even though students with disabilities were now entitled to a free, appropriate, public education, many were segregated from their typical peers all day or for a portion of the day. Many professionals and parents began to question the assumption that students who need more intensive services should be segregated to receive those services. They believed that if students' needs cannot be met in a traditional academic classroom, then the classroom should change, not the student removed from that setting. Earlier in 1986, Madeline Will, then assistant secretary of the Office of Special

Education and Rehabilitative Services in the US Department of Education, called for general and special educators to share responsibility for teaching students with disabilities. This was referred to as the regular-education initiative (REI). The REI was based on several arguments, namely: (1) students with disabilities are more alike than different from their typical peers; (2) all good teachers can teach all students, and (3) segregation is inherently inequitable and discriminatory. In all, the REI helped set the stage for the inclusion movement.

The reauthorization of IDEA in 1997 also moved inclusion forward. The reauthorization required that students with disabilities have access to the general-education curriculum. Access, as defined here, meant that students had to have meaningful interaction with the curriculum in order to learn. Clearly, access to the general curriculum required barriers be removed that may inhibit students' access, as well as additional supports provided to help them actively engage with the curriculum. The general-education curriculum was also interpreted as moving beyond instruction and materials to full school day's events including extra-curricular activities. The meaning of inclusion went beyond a placement of students with disabilities into general classroom. It was now defined as an educational outcome for students with disabilities. The most recent legislation to impact inclusion in the US was the No Child Left Behind Act of 2001 and the reauthorization of IDEA in 2004 as the Individuals with Disabilities Education Improvement Act (IDEIA). These laws mandated alignment of general-education standards with academic content for students with disabilities. This created underlying tension between being able to meet general-education standards and at the same time accommodate the individualized needs of students with disabilities. Under these laws, general education also became accountable for the academic gains made by all students, including those with disabilities.

Other education reform and social movements, such as normalization and deinstitutionalization, helped lay the foundation for the inclusion movement. The principle of normalization began in Scandinavia and was later brought to the United States. Under the normalization principle, opportunities, interactions, and experiences that parallel those of society were promoted for individuals with disabilities. This principle extended beyond educational opportunities. Individuals with disabilities were thought to be entitled to the same, for example, recreational, leisure, transportation, employment, and housing opportunities. As indicated, deinstitutionalization also impacted the inclusion movement. Reports of abusive conditions in institutions that housed individuals with disabilities resulted in their closure; and smaller, community-based living arrangements were developed. Inclusion in these contexts goes beyond education to mean inclusion of individuals with disabilities in society at large.

Pros and Cons of Inclusion

The arguments in favor of inclusive education have been based on legal, moral, educational/social, and political grounds. Proponents using the legal or civil rights' perspectives argue that segregating students is antithetical to basic human rights. All individuals, including those with disabilities, have the right to receive the same educational opportunities. Others take an ethical and moral perspective and contend that professionals serving students with disabilities, regardless of the students' needs, have a moral and ethical obligation to keep them with their typical peers. Proponents using the moral argument contend that the practice of inclusion does not require empirical support. The educational/social argument is based on the claim that those with disabilities experience developmental, educational, and social benefits. The educational/social argument also asserts that inclusion is a positive experience for teachers and typically developing peers. Teachers' professional skills improve as a result of teaching in inclusive classrooms; and peers develop more positive attitudes toward those with disabilities which increases the likelihood that those with disabilities will be accepted as equal members of society. The educational/social argument can and has been examined empirically. A summary of the results of these studies are summarized later. The last argument is political. Educating students with disabilities is expensive and costs continue to rise. Some policymakers and administrators view special education as financially draining the rest of the educational system and from their perspective one solution is to discontinue educating students with disabilities in segregated placements requiring expensive equipment and personnel. This last argument, however, goes counter to the legal mandates and experience shows that inclusion, if done properly, is not less expensive.

Opponents of inclusive education are generally concerned about the impact that inclusion has on teachers, students, and/or schooling in general. For example, they contend that general-education teachers have neither been adequately prepared nor should they be expected to work with such a wide range of student needs. Under the current wave of standards reform, teachers are increasingly accountable for the academic gains of all students for whom they are a teacher on record. Including students with disabilities in their classrooms increases general educators' teaching loads and stress levels. Many are concerned that general educators will be reprimanded if students in their classroom do not meet, for example, annual yearly progress as mandated by the No Child Left Behind Act. Concern about the role of the special educator in an inclusive model has also been raised. Traditionally, special educators were prepared to teach students with disabilities in segregated settings. With inclusion, the special educator's

role has evolved to a consultant and/or co-teacher; and, opponents argue, co-teaching often results in the special educator becoming an overly paid instructional aide to the general educator. Others argue against inclusion because they believe it will have a negative impact on students without disabilities. They believe general educators' attention to those with disabilities will leave other students' needs neglected. In addition, they believe inclusion can never meet individual needs of those who require extensive services. Opponents express concern that schools have jumped on the inclusion bandwagon before proper systemic changes have taken place; and some argue that school-wide support and systemic change required to implement inclusion are unrealistic expectations.

Implementation of Inclusion

To be considered inclusive, schools must be physically, socially, and instructionally integrating students with disabilities. That is, students with disabilities are physically placed in the same classroom as their typical peers, social relationships between students with and without disabilities are fostered, and all students are taught using the same curriculum. Students with disabilities are supported by teachers modifying instruction and how learning is measured. For students with significant disabilities, instruction is anchored in the general curriculum but expectations are appropriately adjusted (Friend and Bursuck, 2006). Typically, how schools implement inclusion varies. If full inclusion is endorsed by a school, then students with disabilities would be dispersed across all classrooms to best represent the proportion of disabilities found in the general population. Some schools, however, prefer to designate specific general-education classrooms as inclusion classrooms. In these schools, students with disabilities may be clustered within several general-education classrooms. In some inclusion models, general educators and special educators co-teach in the inclusive general-education classrooms. In other models, the special educator acts as a consultant to the general-education teacher. Regardless of the special educators' role, the following instructional strategies are generally used to implement inclusion: (1) differentiated instruction, (2) interdisciplinary curriculum, (3) technology, (4) student collaboration/peer-mediated instruction, (5) cooperative learning groups, (6) partner learning/peer tutoring, and (7) personal supports and accommodations (Villa *et al.*, 2005).

Inclusion is most prominent in preschool and early elementary grades. As children age and skill levels widen between those with disabilities and their typically developing peers, implementing inclusion becomes more difficult. Thus, the challenge of full inclusion is perhaps greatest at the high school level. In high school, academic

courses are highly academic within specialized disciplines, the number of classes in which students are enrolled increases, and the student to teacher ratios are higher. Clearly, implementing inclusion is not easy. The entire school must commit to major changes and these changes must be supported by all stakeholders, including district personnel, school boards, and family members. Some resistance is to be expected. Schools must be empowered to manage their own change and find the model that best fits their needs. No one inclusion model fits all schools, and the development of an inclusive school is never complete (McLeskey and Waldron, 2002).

Impact of Inclusion and Attitudes About Inclusion

Many studies have examined the academic and social impact of inclusion on students with and without disabilities. Researchers have also studied the perception of inclusion across various groups of stakeholders. Others have synthesized these studies into reviews of literature (e.g., Avramidis and Norwich, 2002; Duhaney and Salend, 2000; Salend and Duhaney, 1999; Schmidt *et al.*, 2002; Scott *et al.*, 1998). Findings are generally varied, complex and impacted by a multitude of variables.

Students with and Without Disabilities

The results of studies examining the impact of inclusion on students with disabilities have received mixed results. Some studies conclude that inclusion programs improve educational outcomes such as standardized test scores, grades, on-task behaviors, attitudes about school, and reading skills. Other studies indicate that certain students are not given specially designed instruction to meet their needs in the inclusive classroom, and thus, students who receive instruction in pull-out programs (e.g., resource rooms) achieve higher academic gains. Reports of social gains are also varied. Generally speaking, however, students in inclusive settings improve socially, behaviorally, and in self-concept to a greater degree than those educated in noninclusive settings. Studies examining the inclusion of students with severe disabilities found that they interacted more frequently with their nondisabled peers, but do so in a nonreciprocal manner. Those with mild disabilities, on the other hand, do develop reciprocal relationships, but some are rejected by their peers and have lower self-perceptions than those without disabilities. In the end, students with disabilities have expressed both benefits and concerns about inclusive practices. Some prefer pull-out programs because of the individualized academic help they receive. At the same time, they do not like the stigma of being pulled out of their general-education classroom which

causes them to miss general-classroom activities. At the secondary level, students report negative experiences in both general and special-education classrooms. In general education, they often feel lost and confused; and in the special education classroom, they often feel stigmatized and unchallenged.

Some studies have indicated that placing students with disabilities in inclusive classrooms does not interfere with (and to some degree may enhance) the academic progress of students without disabilities. For example, some research discovered that academic performance of typical elementary students in inclusive classrooms was equal to or better than those in noninclusive classrooms. Studies also found that general-education teachers in inclusive elementary classrooms do not significantly reduce teaching time for students without disabilities. At the secondary level, students with disabilities have been found to receive more attention from the teacher, but that attention does not have a negative impact on the academic engagement or achievement of students without disabilities. Moreover, typical peers' academic achievement was enhanced when given an opportunity to provide peer support to those with more severe disabilities. Some studies have discovered that students without disabilities generally support inclusive practices. Research has found that inclusion programs help them become more accepting and tolerant of each other. For example, fears toward those with differences were reduced and attitudes toward individuals with disabilities were improved. Typical peers view inclusion as an opportunity to gain greater appreciation for human diversity and as an opportunity to have friendships with those with disabilities.

General and Special Education Teachers/Service Providers

General-education teachers and service providers have varying attitudes about inclusion which seem to be shaped by their experiences, student characteristics, and availability of supports. Some of them support inclusion only if they are required to make minimal accommodations for students with disabilities. Others prefer pull-out programs, such as resource rooms, for students with mild disabilities. One study found, for example, that teachers were more likely to reject students with mild disabilities than those with severe disabilities. In addition, general educators who work in inclusive classrooms have been found to have more positive views of inclusion than those who have not. Elementary teachers are more likely to support inclusion than secondary teachers. Some general educators view inclusion as having helped them develop skills at meeting a wider variety of student needs including improved confidence in their teaching abilities. They also indicate more favorable attitudes toward those with disabilities. Other general educators

and service providers are concerned about whether they can truly meet the needs of students with disabilities; many of them feel they have not been adequately prepared to do so.

With the practice of full inclusion, many special educators and service providers have questioned how their job duties have and will change. Most inclusive schools create co-taught classrooms where special educators work directly with general educators in delivering instruction to all students, including those with disabilities. In this model, special educators have expressed concerns about loss of control over their own classroom environment and their role as a subordinate in the general-education classroom, particularly at the secondary level. They have also expressed concern about whether special-education students are receiving appropriate services in inclusive classrooms. Studies have indicated, however, that special educators and service providers working in inclusion classrooms report feeling more accepted by general-education teachers and students without disabilities. In general, they feel more integrated into the school community.

Family Members

Attitudes and perceptions of families of children with disabilities have been consistently explored by researchers. Results show that they have mixed, but generally positive feelings about inclusion. Family members view inclusion as a way to teach their children about needs of others and to accept human diversity. On the other hand, parents also have concerns, such as assurances that their child will receive the necessary help, and that they may emulate inappropriate behaviors of those with disabilities. In the end, researchers have concluded that the majority of parents of students with disabilities support inclusion. Some of the benefits they identify include acceptance of their children by nondisabled peers, improving their child's self-esteem, providing appropriate role models, and preparing their children for the real world. Parents also express concerns about the lack of qualified personnel and specialized services for their children being educated in inclusive classrooms and about their children being ignored, isolated, or mistreated.

International Perspective

The movement toward inclusion is a global phenomenon. Countries and regions within countries, however, are at different points along the road to implementation. Some countries have implemented inclusion more universally and for a longer period of time than the United States. Italy is one such country. In 1977, an Italian law was passed requiring that all students with disabilities be included in general-education classrooms. Unfortunately,

there exists limited research supporting positive student outcomes of inclusion in this country. Nevertheless, Italy has been identified as a model country of inclusion that others could replicate (Begeny and Martens, 2007). Denmark, Sweden, and the United Kingdom are also among the countries that have implemented some form of inclusion for a number of years. Other countries have not moved as swiftly as these countries. In 1995, for example, about 16% of students with disabilities were segregated into Australian special schools as compared to 5% in the US (Anderson *et al.*, 2007). In South Africa, students with disabilities were not guaranteed access to educational institutions until 1996 (Engelbrecht *et al.*, 2005). In the same vein, Vietnamese laws and regulations changed as late as in the 1990s allowing students with disabilities to attend either special or general-education schools (Villa *et al.*, 2003).

Inclusion is still considered a Western concept. Before it could be implemented in many places, changes in philosophical thinking and culturally embedded beliefs about those with disabilities and the role of schooling would need to change. For example, some societies still view those with disabilities as deviant or broken and that others should be protected from them. Students and their parents prefer not to be identified as having a disability given the stigma attached. Schools may have high academic standards, promote strong competition, and focus only on excellence, leaving little room for addressing individuals' needs. Additionally, individual civil rights and the idea that schools should be opportunities for improving one's quality of life are not accepted principles by all societies. Many cultures, however, are beginning to accept individuals with disabilities as potential contributing members of society. In the last few decades, such changes are reflected in grassroots movements positively impacting laws and regulations for students with disabilities across the world.

Conclusion

Inclusion is a philosophy or belief that all students are included in the general-education classroom and have access to the general-education curriculum. All students' learning is the responsibility of all educators. Full inclusion means that literally all students remain in the general-education classroom with needed services brought to them. If the student cannot meet expectations of the general classroom, then those expectations should be altered. The inclusion movement has been influenced by legal mandates, as well as social and educational reforms. Proponents of inclusion have based their arguments on legal, moral, educational/social, and political grounds. Opponents express concern about the impact of inclusion on teachers, students, and schooling in general. Empirical studies indicate that students with and without disabilities

can benefit from inclusion and that most general educators, special educators, students with and without disabilities, and family members of these students support inclusive practices. While inclusion may be found worldwide and many countries and regions have moved closer to inclusive practices in the last few decades, belief systems about disabilities continue to inhibit such practices.

Bibliography

- Anderson, C. J. K., Klassen, R. M., and Georgiou, G. K. (2007). Inclusion in Australia: What teachers say they need and what school psychologists can offer. *School Psychology International* **28**, 131–147.
- Aramidis, E. and Norwich, B. (2002). Teachers' attitudes toward integration/inclusion: A review of the literature. *European Journal of Special Needs Education* **17**, 129–147.
- Begeny, J. C. and Martens, B. K. (2007). Inclusionary education in Italy: A literature review and call for more empirical research. *Remedial and Special Education* **28**, 80–94.
- Duhaney, L. M. G. and Salend, S. J. (2000). Parental perceptions of inclusive educational placements. *Remedial and Special Education* **21**, 121–128.
- Engelbrecht, P., Oswald, M., Swart, E., Kitching, A., and Eloff, I. (2005). Parents' experiences of their rights in the implementation of inclusive education in South Africa. *School Psychology International* **26**, 459–477.
- Friend, M. and Bursuck, W. D. (2006). *Including Students with Special Needs: A Practical Guide for Classroom Teachers*, 4th edn. Boston, MA: Allyn and Bacon.
- McLeskey, J. and Waldron, N. L. (2002). School change and inclusive schools: Lessons learned from practice. *Phi Delta Kappan* **84**(1), 65–72.
- Salend, S. J. and Duhaney, L. M. G. (1999). The impact of inclusion on students with and without disabilities and their educators. *Remedial and Special Education* **20**, 114–126.
- Schmidt, R. J., Rozendal, M. S., and Greenman, G. G. (2002). Reading instruction in the inclusion classroom: Research-based practices. *Remedial and Special Education* **23**, 130–140.
- Scott, B. J., Vitale, M. R., and Masten, W. G. (1998). Implementing instructional adaptations for students with disabilities in inclusive classrooms: A literature review. *Remedial and Special Education* **19**, 106–119.
- Villa, R. A., Tac, L. V., Muc, P. M., *et al.* (2003). Inclusion in Viet Nam: More than a decade of implementation. *Research and Practice for People with Severe Disabilities* **28**, 23–32.
- Villa, R. A., Thousand, J. S., Nevin, A., and Liston, A. (2005). Successful inclusive practices in middle and secondary schools. *American Secondary Education* **33**(3), 33–50.

Further Reading

- Carter, N., Prater, M. A., and Dyches, T. T. (2009). *Adaptations and Accommodations for Students with Mild to Moderate Disabilities*. Boston, MA: Allyn and Bacon.
- Cushing, L. S., Clark, N. M., Carter, E. W., and Kennedy, C. H. (2005). Access to the general education curriculum for students with significant cognitive disabilities. *Teaching Exceptional Children* **38**(2), 6–13.
- Kavale, K. A. (2002). Mainstreaming to full inclusion: From orthogenesis to pathogenesis of an idea. *International Journal of Disability, Development and Education* **49**, 201–214.
- King-Sears, M. E. (2001). Three steps for gaining access to the general education curriculum for learners with disabilities. *Intervention in School and Clinic* **37**(2), 67–76.

- McLeskey, J. and Waldron, N. L. (2007). Making differences ordinary in inclusive classrooms. *Remedial and Special Education* **42**, 162–168.
- Salend, S. J. (2008). *Creating Inclusive Classrooms: Effective and Reflective Practices*, 6th edn. Upper Saddle River, NJ: Merrill Prentice Hall.
- Scherer, M. M. (ed.) (2003). Teaching all students. *Educational Leadership* **61**(2).
- Villa, R. A. and Thousand, J. S. (2005). *Creating an Inclusive School*, 2nd edn. Alexandria, VA: Association for Supervision and Curriculum Development.
- Winzer, M. A. and Mazurek, K. (eds.) (2000). *Special Education in the 21st Century: Issues of Inclusion and Reform*. Washington, DC: Gallaudet University.

- Wolfe, P. S. and Hall, T. E. (2003). Making inclusion a reality for students with severe disabilities. *Teaching Exceptional Children* **35**(4), 56–60.

Relevant Websites

- <http://www.ascd.org> – Association for Supervision and Curriculum Development.
- <http://www.cec.sped.org> – Council for Exceptional Children.
- <http://www.kidstogether.org> – Kids Together, Inc.
- <http://www.ldonline.org> – Learning Disabilities On-Line.
- <http://www.tash.org> – TASH.

Instructional Accommodations for Children with Special Needs in Inclusive Settings

A L Yurick, Cleveland State University, Cleveland, OH, USA

S Musti-Rao, The University of Cincinnati, Cincinnati, OH, USA

G Cartledge, The Ohio State University, Columbus, OH, USA

© 2010 Elsevier Ltd. All rights reserved.

Many educators can attest to the fact that classrooms in the United States are becoming increasingly diverse. In fact, some estimates indicate that up to 42% of students enrolled in public schools in 2005 reported minority status and at least 20% speak a language other than English at home (NAEP, 2008). However, the nature of diversity in the classroom extends beyond cultural and linguistic diversity; it also includes diversity in academic functioning and programming. Currently 13% of the school enrolment is being served under Individuals with Disabilities Education Act (IDEA) and up to 96% of general education teachers have reported having a student with learning disabilities in their classroom (US Department of Education, 2005). It is certainly understandable how a teacher faced with such a collage of student characteristics and needs can become overwhelmed by the task of providing appropriate instruction for all of his or her students. Now more than ever, all teachers need to be able to modify and adapt their instruction to meet their students' individual benchmarks.

Teaching within classrooms with highly diverse populations needs to be flexible both for instructional and assessment purposes. Instructional accommodations are covered in federal legislation such as Section 504 of the Rehabilitation Act of 1973 (PL 94-112) and the reauthorization of IDEA (IDEIA, 2004). Under Individuals with Disabilities Education Improvement Act (IDEIA), students with one or more of the 13 specific disability diagnoses are guaranteed reasonable accommodations necessary to ensure a free and appropriate public education. After an evaluation by a multidisciplinary team, if a student is found to have a disability, he or she should begin receiving instructional accommodations as outlined in his or her individualized education plan. However, students with other types of impairments or disabilities who are not eligible for funding under IDEIA may still secure assistance through Section 504 of the Rehabilitation Act of 1973, a civil rights law that protects individuals who have a physical or mental impairment that substantially impedes daily life activities. The definition of disability is much broader under Section 504 and may apply to students with more temporary conditions such as communicable diseases, accident injuries, short-term hospitalizations, and environmental or attention difficulties. For either law, the nature of the accommodations may function to provide both additional

services and accommodations or may be as simple as just eliminating existing barriers to optimal functioning. The nature and extent of the instructional accommodation is, and should be, individually determined according to specific needs.

The terms accommodation and adaptation are used widely in general and special education, sometimes with an element of confusion. Therefore, it is important to note the difference between an instructional accommodation and an instructional adaptation. With an instructional accommodation, the content being taught or standard performance criterion is not altered. Accommodations do not reduce learning expectations; they are intended to provide greater access to the content. All students in the classroom, regardless of disability, pursue the same curriculum and are subject to the same grading scales and evaluation procedures. Changes are made to the delivery of instruction without any change to the curricular content or conceptual difficulty. Examples of instructional accommodations might include using multiple formats to present information, using completed work samples for demonstration, providing material in large print for those with visual impairments, or enlisting an interpreter for students with hearing impairments. Unlike instructional accommodations, making an instructional adaptation includes making changes to the curricular content, physical environment, or methods while assisting children to learn in the least-restrictive environment (Bauer and Brown, 2001). Students with more significant cognitive disabilities may possess limited understanding of the general education curriculum and require modifications such as fewer assignments or material revised for lower readability level. The decision regarding whether a student requires an accommodation or an adaptation is largely an individual one and should be determined based on optimizing the outcomes for that particular student.

With the dynamics of the student populations in schools changing dramatically, there is an increased emphasis on access to the general curriculum for all students. This means that teachers need to deliver instruction with same curricular content to both students with and without special needs. In this era of multicultural classrooms and standards-based education, teachers and service providers are asking important questions about how to accommodate the individual needs of students in their

classrooms. One approach to this issue is to introduce greater flexibility through differentiating the instruction. Although the concept of differentiated instruction lies in general education and the concept of making instructional accommodations is from the field of special education, the end results are basically similar (Prater, 2007). It would be useful to view making instructional accommodations for students with special needs in the context of differentiated instruction in the inclusive classroom.

In this article, we discuss instructional accommodations in the inclusive classroom using differentiated instruction as our conceptual base. We provide a model for making accommodations plausible, and suggest some ways for using this model at various educational levels. The focus in this article is on instructional accommodations, since assessment accommodations are addressed elsewhere in the encyclopedia.

The Inclusive Classroom

The ideal inclusive classroom is one in which learners of variable readiness, learning styles, and interests can optimally benefit from the curriculum being taught. In order to achieve this goal, accommodations must be made. However, the reality is that multiple curricular areas and content must be covered. Factor this multiplicity by the number of students (and individual needs) in the classroom, and making appropriate accommodations becomes a daunting task indeed! It may be helpful to point out that not all students will require accommodations, and many will not need accommodations in multiple areas. However, for students who do need some curricular adjustments, identifying the specific aspects of instruction which can be adapted may help clarify the task.

Tomlinson (1999) makes the distinction that instructional accommodations can take at least three forms: content, process, or product. Content accommodations refer to altering the materials used in the lesson. For example, a student with a reading disability who displays off-task behavior or incomplete work samples may be able to fully participate in the lesson if the reading material is revised to reflect a more appropriate difficulty level. Likewise, students with special needs sometimes become overwhelmed by task complexity. Therefore, presenting the same content in smaller steps can make the difference between a completed product and an abandoned assignment due to frustration. In a process-type accommodation, the teacher varies instructional activities according to the learner's preferences and strengths. Rather than having students take independent notes on a chapter, the teacher might choose a more scaffolded approach such as lecturing with the assistance of guided notes. This provision ensures that the student retains a complete and accurate product without becoming lost in a lecture-style format, which is

sometimes difficult for students with learning or attention deficits. Finally, a product accommodation allows the student to illustrate his/her mastery of the lesson in a differentiated format. The culmination of most lessons is a student-generated product reflecting the student's understanding of the material. However, even if students are proficient, those who need accommodations may fail to demonstrate competencies similar to peers simply due to the nature of the assignment. Perhaps the most common example is the student who is a weak test-taker. Allowing the student more time for the exam or taking oral responses rather than written responses may ameliorate this limitation. Breaking down the classroom elements in terms of content, process, and product has its roots in the literature on differentiated instruction and therefore a clear understanding of differentiated instruction is warranted

Differentiated Instruction

In each of the aforementioned examples, the teacher is making accommodations by differentiating instruction. According to Tomlinson (2003), "differentiated instruction is responsive instruction" (p. 2). As stated, due to the increasing diversity in classrooms, teachers and service providers need approaches that are flexible enough to be responsive to their students' learning needs. In fact, in this age of accountability, it is crucial that general and special educators come to their classrooms armed with the knowledge and expertise to teach students in the most effective and efficient multidimensional manner (Meo, 2008). This means that they need to be responsive to the ever-changing dynamics of student learning and demonstrate the ability to adjust their instruction accordingly. Central to the educational reform movement is the commitment to closing the achievement gap and contributing to the overall improvement of student outcomes. These goals may be achieved by establishing the use of differentiated instructional practices as the rule, rather than the exception, for the entire school. In fact, some schools have implemented differentiated instruction at the school level and succeeded in meeting these goals (Cusumano and Mueller, 2007). However, the benefits to this type of approach are not limited to just satisfying recent educational legislation and reform. This approach is collaborative in nature and provides a real opportunity for general and special education teachers to learn from one another and streamline their teaching into a best-practice model. Differentiated instruction can serve as a platform for collaborative efforts, such as employing co-teaching approaches in the classroom (Hawkins, 2007). Most importantly, however, when teachers and service providers work collaboratively to bring their students the best instruction, student outcomes improve and students begin to feel more competent, confident, and successful in the classroom.

Choosing appropriate accommodations for individual students such that they are able to perform at their maximum potential is an important task. Unfortunately, in the past, individualized educational plan (IEP) teams have had little to no guidance on choosing accommodations. Most accommodations listed have used assessment accommodations with no emphasis on instructional accommodations. The need for a close match between instructional accommodations and assessment accommodations cannot be overstated. It is only logical that students receiving a specific type of assessment accommodation be given the same accommodation during instruction. That is, for example, if a student is given the opportunity to make oral responses during assessment, the same type of responses should have been required during instruction. By the same token, it would be inappropriate to deny students accommodations during test-taking that they used during instruction. Certain guidelines and a model can assist teachers to make decisions about designing possible accommodations for students in their classrooms.

A Model for Instructional Accommodations

Before deciding on which accommodations can facilitate students' learning, teachers and service providers must first familiarize themselves with the different types of accommodations that are possible. Viewing accommodations in terms of content, process, and product can simplify the task for teaching professionals as they begin

to plan for instruction. Although many accommodations that fall under product pertain to assessment or test accommodations, there may be some overlap with content and process accommodations, and nevertheless, will help with planning for assessment accommodations. **Table 1** lists the different accommodations under each category along with examples.

Educators can use **Table 1** to generate a list of accommodations that can be used in their classrooms. Soliciting students' input in providing accommodations may enhance its use and effectiveness. **Figure 1** shows steps that educators can use and questions to ask when making decisions on accommodations.

SM3 Model

Another way of conceptualizing instructional accommodations in the inclusive classroom is the setting, methods, motivation, and materials (SM3) model, which incorporates changes to content, process, product, and learning environment (Tomlinson, 2007).

Setting

Setting refers to the physical organization of the classroom that needs to be factored in during lesson planning. This includes the arrangement of desks and a planned seating chart based on students and their individual needs, schedule and length of activities, and other environmental elements such as lighting, noise levels, room temperature, and time of day.

Table 1 Types of instructional accommodations

Categories	Accommodations	Examples
Content	Size/quantity – vary the number of items the student is expected to learn	Reduce (or increase) number of worksheets to be completed at one time
	Difficulty – vary the level of difficulty based on students' needs	Provide partially completed class or lecture notes; list multi-step directions
	Alternative goals – vary lesson objective for students while using same materials and content	Write definition of vocabulary words vs. matching the word to its definition vs. orally defining the words
	Substitute curriculum – work on curriculum different from regular curriculum	One student with special needs works on adding with manipulatives, while rest of the class work on more complex math
Process	Input – delivery of instruction	Small group instruction vs. lecture format; motivation and reinforcement systems
	Level of support – varying the degree of assistance student receives	Provide learning aids, such as guided notes, graphic organizers, and concept maps
	Time – the actual time allotted for learning or task completion and lesson structure	Decrease (or increase) pace of instruction
	Participation – level of active involvement with the task	Active student responding (ASR), choral response, discovery learning
Product	Time – the period of time allowed to complete assignments or tests	Extended time on tests and assessments
	Output – the way a student responds to instruction	Oral responses; using word processors to write essays

Adapted from Prater, M. A. (2007). *Teaching Strategies for Students with Mild to Moderate Disabilities*. Boston, MA: Pearson Education.

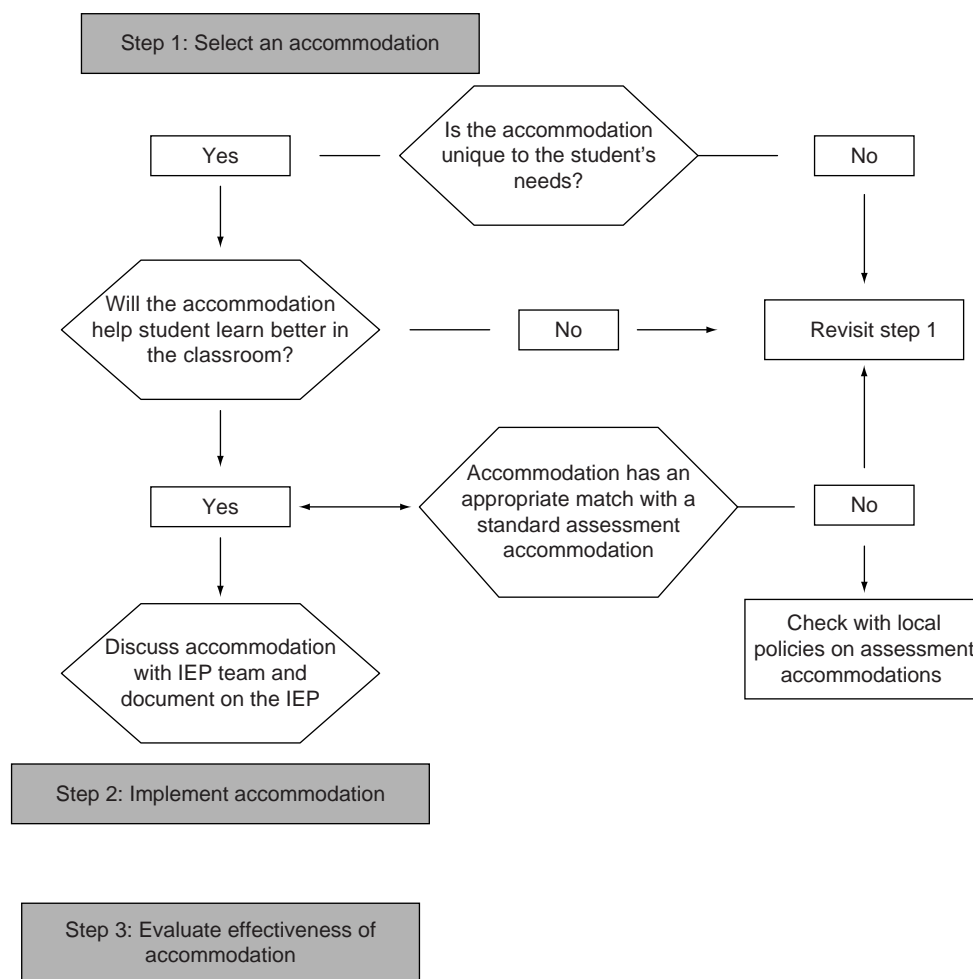


Figure 1 Accommodation selection decision tree. Adapted from Prater, M. A. (2007). *Teaching Strategies for Students with Mild to Moderate Disabilities*. Boston, MA: Pearson Education.

Methods

Methods refer to the instructional strategies that teachers employ in their classroom. Depending on the diversity in terms of ability levels, instruction can take different forms. For example, teachers can use direct instruction, cooperative learning, lecturing, or small group discussions to cover the content of the subject area. Similarly, the teachers and service providers can employ different modalities (e.g., auditory, visual, and kinesthetic) to help students with learning. Methods also refer to the type of response that students make and also the type of assessment used to evaluate their learning.

Motivation

Motivation refers to accommodations made to the social environment of the classroom to facilitate student learning. Smaller groups and peer-tutoring formats can be set up to motivate students of all ability levels. Use of various types of reinforcement systems and contingency contracts can further motivate students to learn.

Materials

Materials include tools and equipment that teachers use for instruction and assessment. Some examples of materials include desks, study carrels, guided notes, self-monitoring checklists, calculators, and computer software. Here again a close link between instructional and assessment accommodations is critical. If a student reads below grade level and receives materials at a lower reading level during assessments, then teachers and service providers will have to design the same type of accommodations during instruction. Tape-recorded materials, online resources, and simplified texts are some examples of accommodations made for students with reading problems. Any type of assistive-technology devices that are both low-tech and high-tech would fall under this category.

Assistive technology is another example of materials accommodations that enhance student learning. Assistive technology includes any device, low-tech and high-tech, that helps people with disabilities function with greater independence. Using assistive technology in the classroom

to accommodate students with disabilities can range from providing students with guided notes or partially completed notes during lectures to providing students with laptop computers with voice-recognition software to record a complete lecture. Depending on the type of disability, assistive-technology devices can be more specific. For example, students with physical disabilities often have difficulty taking notes, turning pages of a book, or taking written assessments. Alternative input devices such as alternative keyboards, joysticks, sip-and-puff systems, and touch screens assist students that have low muscle tone or functionality in their upper limbs. Students with learning disabilities may find the reading level in content area subjects (e.g., science and social studies) difficult and frustrating. Text-to-speech software programs convert printed text to auditory feedback with a scanner.

Classroom Accommodations at Elementary and Secondary Levels

It may be helpful to put this information back into context by illustrating how a classroom teacher puts content, process, or product accommodations into practice for a student with special needs in an inclusive setting. The first example describes how these accommodations can be incorporated into an elementary setting. Next, we outline the process for a student at the secondary level.

Elementary level

In Mr. Donaldson's 12 years of teaching, he learned to be flexible and responsive to the individual learning styles of each of his students. He had students with disabilities fully included in his classroom before, and learned a few techniques over the years that helped him make adjustments for these students. However, sometimes he tried several different approaches before finding one that truly suited the student's needs. Instead of making isolated and somewhat arbitrary alterations, he frequently wondered if he could streamline his approach to more effectively and efficiently make instructional accommodations that would more likely work the first time. With three students with mild disabilities fully included in his classroom this year, he would have liked to save time and energy by prescriptively matching the accommodations to the students' needs. On the other hand, Sasha was a student with a mild reading disability who sometimes became easily distracted when it was time to work independently on assignments in class. When presented with a reading or writing assignment, she made repeated trips to the pencil sharpener, asked to use the restroom, or began whispering to the student next to her. She seemed to do just about everything she could to avoid doing her work. After a discussion with the school's intervention specialist,

Mr. Donaldson hypothesized that Sasha's poor persistence to task may be exacerbated by her reading disability. That is, if parts of the assignments require reading and writing that are too overwhelming for her, she may be using her off-task behavior to avoid the assignments altogether. Mr. Donaldson reviewed the parts of his assignments that have reading or writing requirements and decided to align a content, process, and product accommodation to the assignment for Sasha in order to help her benefit more from the instruction.

For the next class assignment, a unit about the upcoming presidential election, Mr. Donaldson had his class select three or more books from the library to write a comprehensive report on each candidate. The required reading and writing for this assignment was more cumbersome for Sasha and would delay her progress through the unit. Therefore, Mr. Donaldson suggested that Sasha rely on more multimedia resources, such as television news programs, the Internet, blogs, and webcasts for the content of the unit. By making this content accommodation, Sasha was not limited in resources and she could use materials that were at her instructional level. For the research process, Sasha created a webquest related to researching for information on presidential candidates. The interactive nature of this process accommodation significantly reduced her levels of avoidance and distraction that she typically displayed when required to read independently for an extended period of time. Finally, instead of turning in a written product for the outcome measure in the unit, Mr. Donaldson suggested that Sasha's final product be an oral presentation and demonstration of her presidential candidate webquest for the class. In spite of her reading disability, she demonstrated that she gained knowledge related to the unit, while also showing her peers new resources for their further investigations. Mr. Donaldson was pleased with her overall performance and her level of understanding, and Sasha remained excited and engaged in her schoolwork.

Secondary level

Andre was a friendly, outgoing, hard-working high school sophomore. He enjoyed many sports and was an accomplished junior varsity basketball player. Unfortunately, Andre's learning disability impeded his performance in his geometry class to the degree that he risked being cut from the basketball team if his grades did not improve. Andre, his parents, the school intervention specialist, and Ms. Morgan, the geometry teacher, decided to develop a plan that would help Andre be more successful in the class. As with most geometry classes, the units were introduced in a lecture format, students practiced problems from the book independently, and then were tested on the material at the conclusion of the unit. After some discussion, it became apparent that Andre was not taking effective notes in class, and therefore, could not complete

his independent practice assignments accurately. After practicing errors on his homework, it was not surprising that his test scores suffered as well.

In order to provide more guided practice and promote better accuracy in Andre's homework, Ms. Morgan agreed upon some accommodations they could make to the assignments. First, as a content accommodation, Ms. Morgan suggested that it was not the quantity of homework problems completed that was most critical, but the accuracy of the ones he had completed. Therefore, for each assignment given, she selected a subset of problems that Andre completed as homework. To scaffold his learning further, she used this exact subset of problems for the in-class lecture and demonstration when introducing and demonstrating the material to the class. Unfortunately, Andre's poor note-taking skills did not produce an accurate model for him to refer to when doing his independent homework activities. As a process accommodation, Ms. Morgan provided Andre with a structured worksheet and guided notes, which were partially completed notes and examples that required him to write less, and therefore, focus more on the lecture being given. Additionally, Andre found it extremely helpful to audiotape the lecture and demonstration to replay while he was at home completing his assignment. Listening to the recording as he went going through his reduced assignment ensured that he performed each step correctly; and since the problems Ms. Morgan demonstrated in class were the ones Andre practiced at home, the repeated practice increased his accuracy substantially. Finally, Ms. Morgan permitted a product accommodation to be made to the final exam. Again, she reduced the length of the test and allowed Andre to take the test with her after school. In this setting, he was permitted to talk through the answers with her, just as he did during his homework assignments. An additional benefit of this accommodation was that it provided valuable insights for Ms. Morgan to understand how Andre was thinking when he made an error while solving a problem. With these relatively conspicuous accommodations, Andre's geometry performance

improved one letter grade and he was more confident in his ability to be successful in class.

Conclusion

The practice of making instructional accommodations is very consistent with the belief that students are individuals who each learn in their own unique ways (King-Shaver and Hunter, 2003). However, adapting instruction to these many individuals may be a daunting task for teachers. In this article, we have outlined a streamlined approach to incorporating these accommodations in a prescriptive fashion, rather than a haphazard application of strategies which may or may not be sufficient to assist the student. The combination of instructional flexibility and needs-based accommodations can truly result in a superior experience for both the teacher and the student. These approaches are the essence of effective individualized instruction that can make a substantial difference in the education of all students, especially those with special needs.

Bibliography

- Bauer, A. M. and Brown, G. M. (2001). *Adolescents and Inclusion: Transforming Secondary Schools*. Baltimore, MD: Paul Brookes.
- Cusumano, C. and Mueller, J. (2007). How differentiated instruction helps struggling students. *Leadership* 36(4), 8–10.
- Hawkins, V. J. (2007). Narrowing gaps for special needs students. *Educational Leadership* 64(5), 61–63.
- King-Shaver, B. and Hunter, A. (2003). *Differentiated Instruction in the English Classroom*. Portsmouth, NH: Heinemann.
- Meo, G. (2008). Curriculum planning for all learners: Applying universal design for learning (UDL) to a high school reading comprehension program. *Preventing School Failure* 52(2), 21–30.
- Prater, M. A. (2007). *Teaching Strategies for Students with Mild to Moderate Disabilities*. Boston, MA: Pearson Education.
- Tomlinson, C. A. (1999). *The Differentiated Classroom: Responding to the Needs of All Learners*. Alexandria, VA: Association for Supervision and Curriculum Development.
- US Department of Education (2005). *Twenty-Sixth Annual Report to Congress on IDEA*. Washington, DC: Author.

Interagency Agreements

L A Owens, University of Wisconsin–Milwaukee, Milwaukee, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

A major barrier to postschool employment success for young adults with disabilities is the lack of access to appropriate adult services and supports. The Workforce Investment Act (WIA), Rehabilitation Act, and the Individuals with Disabilities Education Act (IDEA) mandate interagency agreements to facilitate collaborative planning and services for youth as they leave the educational system to enter the adult service system. IDEA clearly stated that providing transition services is not solely the responsibility of the school system (Cozzens *et al.*, 1999). While the reauthorization of IDEA in 2004 as the Individuals with Disabilities Education Improvement Act (IDEIA) eliminated the mandate to include a statement of interagency linkages and responsibilities, it did retain the provision that if a participating agency does not provide the transition service described in the individualized education plan (IEP), an IEP meeting must be reconvened to identify alternative strategies (i.e., Sec. 614). If possible, systematic planning for collaboration between school systems and adult service systems should begin with a state plan. This plan should include a description of roles and responsibilities between schools and adult services agencies and provide a clear alignment of goals, priorities, and resources needed to successfully collaborate. Typically, state vocational rehabilitation agencies, state departments of education, and other state agencies (i.e., State Department of Health and Social Services, and State Corrections) are involved in establishing a state plan to: (1) ensure collaboration of services, (2) expect coordinated services, (3) maximize resources, and (4) increase the effectiveness of transition efforts for students with disabilities.

Interagency agreements are a tool that can facilitate the collaborative efforts between the educational and rehabilitation systems. Interagency agreements, also known as memorandums of understanding (or MOUs), are required under the WIA and the Rehabilitation Act. Interagency agreements can be formalized written documents signed by senior administrators of all participating agencies. Agreements can also be developed at an operational or local level and be much less formal in nature. Good interagency agreements promote actions that directly or indirectly improve personal outcomes for those receiving services and promote systems change. This is the focus of the article.

Interagency Agreements and Collaboration

Schools and adult service agencies (e.g., state vocational rehabilitation, local service providers, and independent

living centers) typically operate in isolation with uncoordinated agendas. Many organizations do not have the information or experience in working with one another and this has caused difficulties in a smooth transition from school to adult life. Often, when students transition from school to adult life, there are gaps in services, duplication of services, cost ineffectiveness, and/or inaccessibility of services which can be resolved through interagency collaboration (Metzel *et al.*, 2005). Over the past 10 years, interagency agreements have been utilized to facilitate a smoother transition from one system to another. These agreements bring organizations together to focus on collective expertise and resources to improve transitioning from one system to the next. Interagency agreements promote collaborations that create systems change by outlining in a more formal way, strategies to implement services and effectively utilize resources and funding. A written interagency agreement outlines roles and responsibilities and resources to be committed to reach common goals. Interagency agreements are only as effective as the integrity with which organizations follow through on their roles and responsibilities described in the agreement.

Effective interagency agreements require collaboration between all stakeholders. These agreements are typically implemented when students move from one system to another. For example, those moving from early childhood services to the educational system or students moving from the educational system to the adult service system all benefit from interagency collaboration and the implementation of interagency agreements.

Developing a Good Interagency Agreement

Effective interagency agreements should promote actions that improve personal outcomes for individuals with disabilities and promote changes within systems providing those services. According to Butterworth *et al.* (2001), four important questions should be considered when developing interagency agreements: (1) What is the purpose of the collaboration?; (2) Who owns the interagency agreement?; (3) Who are the allies?; and (4) Will the agreement lead to action? All agencies responsible for transition planning should be engaged in a discussion revolving around these questions to ensure clarity in roles and responsibilities to enhance transition outcomes for students with disabilities.

Purpose of Collaboration

Collaboration is a mutually beneficial relationship with clear roles and responsibilities between two or more organizations to achieve a common goal. For collaboration to be most effective, there must be a commitment to the: (1) definition of mutual relationships and goals; (2) jointly developed structure and shared responsibility; (3) mutual authority and accountability; and (4d) sharing of resources and rewards (Mattessich and Monsey, 1992). When developing interagency agreements, it is important to determine the purpose of such agreements and ensure that this purpose is clear to everyone who is trying to accomplish the identified goals. Butterworth and colleagues suggest that an interagency agreement can be developed for a one-time event that is more political than service delivery focused. This type of agreement may set the stage for a large systems change initiative or bring together appropriate stakeholders to lay the groundwork for subsequent interagency collaboration. A more important purpose of an interagency agreement would be to spell out the procedures for a more mature interagency collaboration, one that has been in place over an extended period of time. This will ensure continuity of the relationship and activities involved such as planning for transition. Finally, an interagency agreement can be an effective tool that enables two or more agencies to accomplish a specific goal such as improving service delivery or implementing systems change.

Allies of Interagency Agreement

It is critical that there is a strong commitment within agencies and organizations, and other stakeholders who may be impacted by the interagency agreement. It is important to find allies internally, within organizations and systems and externally, outside the organizations or systems (e.g., advocates, political, and business).

Interagency Agreement Ownership

Interagency agreements are only effective if there are people involved who will implement the services within the agreement. According to Butterworth *et al.* (2001), owners of the interagency agreement are at the heart and soul of implementation and have a commitment to getting things done. Owners of the agreement are responsible for providing oversight and vision, leading the implementation of the agreement, and monitoring the outcomes of the agreement. In order for interagency agreements to lead to change, owners must ensure that the activities are implemented based on the goals agreed upon.

Agreement Leading to Action

Effective interagency agreements must be concrete and action oriented. To be effective, interagency agreements

must include clear goals with specific action plans for the owners to implement. While interagency agreements can be powerful tools in expanding services to improve student outcomes, it is critical that the goals and action steps are defined to increase student postschool outcomes. **Figure 1** provides a checklist to ensure that interagency agreements address each of the questions.

Strategies for Integration of Services

One impact of interagency agreements is that it provides a tool to assist professionals and individuals in crossing traditional professional boundaries. Interagency agreements must include: (1) collaboration through collective problem solving and sharing of resources; (2) cooperation where agencies interact and communicate on a regular basis with shared accountability; (3) service coordination to assist in the movement from one system to another; and (4) networking where professionals and individuals identify a range of services available in the community to support the individual in adult life (Dunst and Bruder, 2002). There are inherent challenges when students move from one system to another and it is important to consider unique approaches to collaboration. Many times, self-imposed barriers such as turf issues, competition for clients, differences in philosophies, and financial/resource constraints cause difficulties in effectively implementing interagency agreements (Woodside and McClam, 2003). An umbrella agency structure may remove some of the typical constraints involved with issues of turf (Harley, 2003) which may strengthen the impact of interagency agreements. Another alternative would be an agency broker who had knowledge of both the educational and adult service system to provide the needed services to individuals with disabilities. This broker would be part of the interagency agreement and would have a working knowledge of the services provided by each signer of the agreement.

It is critical for interagency agreements to have an impact on student outcomes and has local input and ongoing support. Service delivery and service coordination are two collaborative activities in the interagency agreements. These activities should be clearly defined to spell out exactly what services are to be delivered and exactly what constitutes coordination in order to better inform the activities and expectations that ensure accountability. Federal legislation and state policies continue to identify the expectation of interagency agreements among agencies. As a result, it is important to identify successful elements in the creation and implementation of interagency agreements that result in increasing positive postschool outcomes for individuals with disabilities. Interagency agreements provide an understanding about services provided by state agencies, school systems, and adult service agencies. An interagency agreement provides guidelines

1. The agreement has a clearly defined purpose
 - Outcomes have been defined: Outcome measures emphasize quality of post school outcomes on an based on individual needs.
In Place - Partial - Needed - Comments
 - Outcomes are being measured: A data collection system is in place, and outcome summaries are provided to key stakeholders.
In Place - Partial - Needed - Comments
2. The agreement is well supported by internal and external allies.
 - Internal allies are in place: Allies from within the partner agencies have been identified who support the agreement. Stakeholders are involved in planning and implementation.
In Place - Partial - Needed - Comments
 - External advocates are involved and committed: Support from allies outside of the partner agencies is in place to ensure support for the agreement.
In Place - Partial - Needed - Comments
 - Part of a greater whole: Partners share a common interest and mutually defined goals.
In Place - Partial - Needed - Comments
3. Ownership of the agreement is clearly defined
 - Ownership is clear: An individual is identified in each participating agency who is responsible for implementation and oversight of the agreement.
In Place - Partial - Needed - Comments
 - Linkages/relationships at an operational level: Mechanisms are in place to support ongoing communication among agency personnel who are involved in implementation.
In Place - Partial - Needed - Comments
4. The agreement is action oriented
 - Specific action plans are in place (Activities are well defined. Agreements use action-oriented terms).
In Place - Partial - Needed - Comments
 - Money or resources are committed: Funds, staff time, or other resources have been committed to the agreement by participating agencies.
In Place - Partial - Needed - Comments
5. Mechanisms for communicating values and resolving differences are included
 - Nonnegotiable organizational values have been identified and addressed (All partner agencies provide accessible and welcoming services to individuals disabilities).
In Place - Partial - Needed - Comments
 - A mechanism for resolving agency differences or disagreements during implementation of the agreement is defined.
In Place - Partial - Needed - Comments

Figure 1 Interagency agreement planning checklist.

and procedures for each agency to follow, clarifies roles and responsibilities, and identifies clear methods for communication between agencies. Developing an interagency agreement is not always easy but the benefits outweigh the time and effort involved. **Table 1** highlights essential features of effective interagency agreements (see Crane *et al.*, 2004). While state-level interagency agreements may provide a broad framework to guide transition practices, local interagency agreements allow

community organizations to clarify policies, procedures, roles, and responsibilities to be followed within and among organizations (Fowler *et al.*, 1990).

Examples of Interagency Agreements

State-level interagency agreements are important to set the stage for collaborative efforts between state and local

Table 1 Features of effective interagency agreements

<ul style="list-style-type: none"> ● Agency staff are responsible for design, revision, and implementation of the agreement. ● Agency directors have commitment to the development and implementation of the agreement. ● Input is received from direct service staff in the design, revision, and implementation of the agreement.
Opportunities are available for agency staff and directors to meet, discuss ideas, and develop relationships.
<ul style="list-style-type: none"> ● Opportunities for agency directors and direct staff to learn from each other and see how each can benefit from the mission of other organizations. ● Active involvement in strategic planning by participating agency representatives.
Collection and analysis of data are important to determine the impact and outcomes resulting from the agreement.
<ul style="list-style-type: none"> ● Utilization of data is necessary for strategic planning and continuous improvement
Dissemination of the agreement is important to all stakeholders. Training and technical assistance are needed for direct service staff regarding implementation of the agreement.

stakeholders. Generally, these agreements are developed and signed by state and local agencies that impact transition services for students with disabilities. Some interagency agreements combine birth-to-3 transition goals with the school-to-adult world transition, while others have separate agreements for each stage of transition. For instance, the Wisconsin Interagency Agreement (June 2007) was developed by the Department of Education, Division of Vocational Rehabilitation, and the Department of Health and Family Services. Included in the agreement are: (1) statement of need, (2) purpose and goals, (3) roles and responsibilities, and (4) signatures of all parties involved.

In the same vein, local-level interagency agreements are used to ensure that collaboration continues in communities. Hobbs (1993) developed an interagency agreement that can be used to define roles and responsibilities, as well as a tool for sharing information with all stakeholders. See **Figure 2** for an interagency agreement checklist (Butterworth *et al.*, 2001).

This agreement is between the center and _____ (agency name) _____.

Purpose

[Instructions: This statement must specify the principal expected outcomes of the collaborative.]

The center is an interagency collaboration program whose purpose is to: (1) develop an integrated-services approach based on a shared philosophy and a collaborative leadership structure; (2) provide services to children and families with agency staff who have increased authority to solve problems; (3) develop a cross-agency training institute to build commitment to interagency collaboration and provide technical assistance to managers and line staff; (4) develop an automated information network system that facilitates interagency personal information sharing, referrals, and collection of case characteristic and outcome data.

Much of the interagency casework will be carried out through members of a multidisciplinary team. Members of the team will continue as employees of their current agency and retain their current job classifications while taking on new caseload responsibilities for families with children who reside in the service area.

Service area

[Instructions: This section must state the geographic/jurisdictional boundaries of the program.]

The boundaries of the service area are defined as:

Responsibilities

[Instructions: This section must specify the responsibilities and roles of the parties to this agreement.]

Participating agency

Participant services

- Provide assessment, information, and referral to all children and families who live in the service area.
- Provide case-management services to eligible children and families.

Coordination

- Agree to meet with _____ (agency name) _____ staff to develop appropriate referral procedures.

Figure 2 (Continued)

- Provide regular reports and consultation to ____ (agency name) ____ staff regarding mutual clients under approved confidentiality procedures.
- Provide in-service collaboration skill-development workshops for ____ (agency name) ____ staff.

Confidentiality

- Provide training on Center confidentiality policy and procedure.

(Participating agency)

Multidisciplinary Team Members

____ (Agency name) ____ will designate a multidisciplinary team liaison/staff person to:

- Work as part of the team and help center staff coordinate services to families and children in the service area.
- Meet with center staff to develop appropriate referral and communication channels between parties.
- Attend in-service training and work shops offered at the center and elsewhere.

Confidentiality

- Agree to amend the agency's confidential procedures, as necessary, to conform with center policy; accept the center's release forms for mutually served families; and instruct and supervise members of the team on any special provisions with respect to exchange of confidential information and/or records.
- Agree to enter into any necessary memorandums of agreement to access automated confidential files.

LIFE OF AGREEMENT

This agreement shall continue to ____ (date) _____. It may be terminated earlier by a 30-day written notice from either party for cause. Cause includes, but is not limited to, a change in state, federal, or local directive. The effective date is the date this agreement is signed by both parties.

SIGNATURES

These responsibilities are agreed to by the following authorized signatories.

Participating agency:

Name (Print) _____ Title _____

Signature of center coordinator or designee _____

Date _____

Participating agency:

Name (print) _____ Title _____

Signature of director or designee _____

Date _____

Figure 2 Interagency program memorandum of agreement: linkage agreement between the center and participating agencies (see Hobbs, 1993).

Conclusion

Interagency agreements are effective tools to ensure collaboration among stakeholders to enhance outcomes for

individuals with disabilities. Interagency agreements are only as effective as the partners make them. It is critical that agreements be monitored and amended as needed to provide continual improvement in the systems that

provide the services to ensure the successful outcomes of individuals with disabilities. Interagency agreements will better affect the outcomes of individuals with disabilities if collaboration is seen as a necessary component of service delivery instead of a separate service or goal (Johnson *et al.*, 2003). Interagency collaboration strengthens professional linkages (e.g., schools and vocational rehabilitation centers) to ensure that students are provided with the supports and services needed to reach their full potential after graduating from high school.

Bibliography

- Butterworth, J., Foley, S., and Metzel, D. (2001). *Developing Interagency Agreements: Four Questions to Consider*. Boston, MA: Institute for Community Inclusion.
- Crane, K., Gramlich, M., and Peterson, K. (2004). *Putting Interagency Agreements into Action*. Minneapolis, MN: National Center on Secondary Education and Transition.
- Dunst, C. J. and Bruder, M. B. (2002). Valued outcomes of service coordination, early intervention, and natural environments. *Exceptional Children* **68**, 361–375.
- Fowler, S. A., Hains, A. H., and Rosenkoetter, S. E. (1990). The transition between early intervention services and preschool services: Administrative and policy issues. *Topics in Early Childhood Special Education* **9**(4), 55–65.
- Harley, D. A. (2003). Interagency collaboration: Reinforcing professional bridges to serve aging populations with multiple service needs. *Journal of Rehabilitation* **69**(2), 32–38.
- Hobbs, L. J. (1993). *Tackling the Confidentiality Barrier: A Practical Guide for Integrated Family Services (Special Report for New Beginnings)*, 2nd edn. San Diego, CA: San Diego County Department of Social Services.
- Johnson, L., Zorn, D., Tam, B. Y. T., LaMontagne, M., and Johnson, S. (2003). Stakeholders' views of factors that impact successful interagency collaboration. *Exceptional Children* **69**(2), 195–210.
- Mattessich, P. and Monsey, B. (1992). *Collaboration: What Makes It Work*. Saint Paul, MN: Amherst H. Wilder Foundation.
- Metzel, D. S., Foley, S. M., and Butterworth, J. (2005). State level interagency agreements for supported employment of people with disabilities. *Journal of Disability Policy Studies* **14**, 163–173.
- Woodside, M. and McClam, T. (2003). *Generalist Case Management: A Method of Human Service Delivery*, 2nd edn. Pacific Groves, CA: Thomson/Brooks-Cole.

Relevant Website

<http://dwd.wisconsin.gov> – Wisconsin Department of Workforce Development.

Interdisciplinary Collaboration in Early Childhood

N File, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Collaboration – Accomplishing the work of service-delivery to young children and their families through a model of professionals working closely together across disciplines and settings. The concept involves multiple layers of the work and in practice may be actualized in various manners.

Co-teaching – A model of service delivery in which the classroom is staffed by both a general education teacher and a special education teacher who plan and implement curriculum as a team.

Direct service delivery – Services are provided by the professional directly to the child, in a classroom, pull-out room, or clinic setting.

Indirect service delivery – Services are provided by the professional who consults with the direct provider (e.g., classroom teacher), providing direction, assistance, joint problem-solving, and/or coaching to the direct service provider.

Interdisciplinary team collaboration – A model of service delivery in which team members from various disciplines work together, in their roles as direct service providers, to coordinate services and in which they have equal and shared input into decision-making.

Multi-disciplinary team collaboration – A model of service delivery in which team members from various disciplines work independently and decision-making is directed independently by each member via the pertinent discipline approach.

Trandisciplinary team collaboration – A model of service delivery in which team members from various disciplines work together in both direct and indirect service roles, thereby requiring that some individuals change roles. Services are provided in a coordinated manner with equal and shared input into decision-making from all involved team members.

The work of those who serve young children and their families has, traditionally, taken place in varied settings, including childcare centers, preschools, agencies focused on disabilities services, and public schools. These settings often exist in very different spheres, with little to no connection. Similarly, the professions involved in these services have maintained their own preservice preparation structures, professional associations, and standards.

Even in a single field – early-childhood general education – job-entry requirements for professional preparation vary greatly depending upon the work setting – childcare or public school. Finally, the professions draw variously on the historic root disciplines, such as developmental psychology, education, and social work, to mention a few. When young children with disabilities are served, the picture becomes more complicated, as professionals from the allied health and medical fields become involved. Given the complexity of these settings, collaboration is a necessary but challenging goal. This article aims to explain the nature and practice of interdisciplinary collaboration in early-childhood programs. The focus is on collaboration among professionals, although, in family-centered practice, families may indeed be considered members of interdisciplinary teams.

Models of Collaborative Work

Traditional ways of working, involving tight boundaries, disconnected service systems, and limited teaming have been slow to give way to the more comprehensive and interwoven systems needed to serve all young children well in inclusive settings. Winton (2000) told a story of her own experience in convening an interdisciplinary panel for a course presentation, noting, with regret, that most of those on the panel had not met previously, although they worked within one community and, sometimes, served the same children. Yet, she believed that there are research-based knowledge and promising initiatives that form the base for preparing personnel to work in strongly collaborative, effective ways. In her estimation, the gap between what is known about serving young children and their families and what practice actually looks like can be traced back to resources and policies.

One issue within early childhood is the lack of clear consensus on the meaning of interdisciplinary collaboration (note that the term early childhood is used to indicate both general and special education). Various forms of service delivery and practices might be described as collaborative by those involved. Even those operating within the same teams might quibble about what collaboration means and the degree to which their work is true to those terms. Ogletree *et al.* (2001) explicated three forms of teaming, each with a unique set of advantages and disadvantages. In the multidisciplinary-team model, services are delivered independently by team members, and

members do not participate equally in decision making. Rather than operating interdependently, the team members contribute input with little connection among themselves. While efficient in nature, the risks to resultant service delivery are a lack of coordination and, potentially, the buy-in that can result via joint decision making. Interdisciplinary teams provide for more direct connections and interdependence among the team members, with more equal input into decision making. In essence, advantages and disadvantages of this model are the mirror image of the multidisciplinary model. Interventions may be more coordinated, with whole-group ownership of the plans. On the other hand, time and resources are needed for this model to be effective. In addition, there is some concern that boundary issues related to the various disciplines involved may still negatively affect the work if not effectively met (see Ogletree *et al.*, 2001). Lastly, transdisciplinary teams are noted by greater release from traditional roles and disciplinary boundaries. Some members of the team act as consultants, while others deliver direct services. One advantage of this model is that the team focus can be broadened considerably. A primary disadvantage is that it is risky; its success depends upon team members valuing role change (see Ogletree *et al.*, 2001).

Each collaborative model comes with both advantages and disadvantages. Families may have varying preferences for their own team roles, as well as what they want from the various professionals involved. Professionals have their own preferences, flexibility thresholds, and resources (e.g., time). Therefore, shared commitment to the model of working together, flexibility for various situational factors, and an awareness of what is required for the teaming model to work are all important to the effectiveness of any teaming work (Ogletree *et al.*, 2001). What is clarified is that collaboration (or teaming) is not a singular concept. The complexity is increased by the multiple individual perspectives coming together, as well as the constraints and resources each individual perceives are operating.

Following Professional Standards and Recommended Practice

The major professional associations in early-childhood education are the National Association for the Education of Young Children (NAEYC), with a membership of general educators, and the Division for Early Childhood (DEC) of the Council for Exceptional Children, with a membership of special educators. The standards for initial preparation of personnel in both associations include a focus on collaboration. For example, NAEYC's standards note that at initial licensure, candidates should "demonstrate understanding of and essential skills in interdisciplinary collaboration" (NAEYC, 2001: 25). There is little

specification of what this entails, except that candidates should be able to participate in interdisciplinary teams, particularly in relation to serving children with special needs. Likewise, the standards for DEC include collaboration; the constituent parts include practices with both families and other professionals and agencies. In addition, skills for providing consultation are named, but not further delineated (DEC, 2008). Thus, it is clear that both of the major professional organizations for early-childhood educators place value on collaborative practice.

It remains up to those preparing personnel to appropriately define the skills and attitudes that lead to and allow for collaboration. West and Cannon (1988) provided early work on the essential competencies required of consultants – a role that is, often, viewed as collaborative (although expert models may be utilized as well). Based upon this work, File and Kontos (1998) conducted a needs assessment of special educators moving from a direct service role to consultation in early-childhood programs. They found that special educators' identified needs included knowledge of consultation processes (e.g., understanding stages/phases, understanding a match between consultation approaches and situational needs, implementing strategies to empower change), as well as strategies for both conflict management and collaborative problem solving. These areas overlap in large part with the heart of interdisciplinary collaborative work. Meanwhile, both the special educators and general educators believed they brought strengths in their willingness to learn and seek professional development and in the form of their enthusiastic attitudes.

Rich (2002) offered a list of benchmarks of early-childhood collaborative practices for inclusion. These benchmarked key components of File and Kontos's work (e.g., conflict-resolution processes, problem solving, continuous improvement). In addition, these indicators identified several procedural points – including communicating roles and responsibilities, developing communication strategies, and developing plans for working together. As well, Rich identified some essential stances toward collaborative work, including sharing skills, talents, and resources; and being responsive to new ideas and open to change. While it seems that the field has formed some sense of what collaboration means, implies, and requires, new aspects to collaboration continue to emerge. As Rich explained it also entails, "collaborating across disciplines and settings to plan, use, and evaluate complex or technological intervention" and "collaborating/consulting with each other to integrate specialized interventions and services into daily routines and natural settings" (p. 8). Collaborative teams are fraught with challenges, including what may be multiple perspectives on just how collaborative teams actually operate. While it is difficult to boil human interaction down to a list of skills and traits, continued work in this research area can help to specify what is essential to achieve collaboration.

Understanding Different Collaborative Practices

An illustration of the lack of clarity surrounding collaboration can be found in a case study reported by Hinojosa *et al.* (2001). This team of researchers from various disciplines (e.g., allied health, education, and psychology) commented specifically upon the challenges and the time required as they operated as an interdisciplinary research team. They collected data from an early-intervention team housed in a hospital. While referring to this team as interdisciplinary, it seemed much more aligned with Ogletree *et al.*'s (2001) definition of a multidisciplinary team. The researchers noted that roles and responsibilities of team members were individually defined and that they were confident in and comfortable with the relative autonomy of the team members' work. The research team noted that there seemed to be two programs serving the child and family – one, the therapy team, and the other, the classroom team. These two teams had different perspectives in their approach to practice, with different underlying philosophies guiding their decisions. Because therapies were provided via a pull-out model, one theme identified by the researchers was that the classroom tended to function as a “way station” (p. 212). The educational services being provided by team members were relegated to a subordinate role to the therapy services. These researchers concluded that the team – although functional, effective, and generally self-satisfied – did not operate collaboratively. While noting several constraints on the team's functioning in their particular context, Hinojosa and his colleagues concluded that “team collaboration may be hard to realize” (p. 219).

Interdisciplinary teams, as defined by Ogletree *et al.* (2001) can be found in the research literature on team teaching. In this model, teachers from both general education and special education work together in shared classrooms to meet the needs of all children. Citing the complicated relationships between co-teachers, Noonan *et al.* (2003) queried early-childhood teaching teams with regard to how similar they believed they were to each other. Factor analysis during instrument development resulted in one factor with a very high internal consistency. Interestingly, these items ranged from viewpoints toward practice (e.g., classroom arrangement, scheduling, and activity structure) to professional values (e.g., dedication to teaching, beliefs about inclusion, and confidence as an educator) to more personal characteristics (e.g., flexibility in dealing with unforeseen events, sense of humor, and ability to be supportive to colleagues and other staff). What is left to explore is the concept of threshold. In this regard, these critical questions remain pertinent. Are there areas in which team members, absolutely, must be aligned? How much interpersonal difference can be tolerated, and in what areas of belief and practice? What factors might

influence the process of negotiating similarities and differences?

A model for co-teaching collaboration in early-childhood settings was designed, implemented, and evaluated by Hunt *et al.* (2004). In addition to regular and special education staff, the professional team members included a speech–language therapist and family members. These researchers found evidence that this intervention (a well-defined structure for team meeting, planning, and assessment) was effective in producing positive child outcomes. In addition, they examined the viewpoints of team members, who generally believed the processes allowed for the sharing of individual perspectives and expertise. In addition, team members noted that their commitment was increased as they shared ownership in the planning process. In the end, the researchers noted that their grant-funded project helped to create resources needed for the group process, the lack of which could prove to be a barrier (Hunt *et al.*, 2004). Based on this study, collaborative processes require time and effort beyond teaching together in the same classroom. Collaboration is achieved through the work undertaken apart from the daily direct service to children and their families.

The change in roles – including both direct and indirect service delivery – described by Ogletree *et al.* (2001) as marking transdisciplinary collaboration is evident in early-childhood research on consultation or itinerant teacher models. Dinnebeil *et al.* (2006) conducted observations of five itinerant early-childhood special education teachers across the school year and found that teachers did not conduct their activities in classrooms in the same manner. Despite this, Dinnebeil *et al.* found that most of the time in classrooms across the sample was spent in direct service with children on teachers' caseloads. There was little conversation among adults, as well as little activity that could be viewed as consultation with early-childhood general educators. Similarly, in the activity reports, teachers were most likely to not choose tasks relative to consultation and teaming to describe their work. Some reported collaboration with other professionals – such as those from allied health – but this was still not a major job role. In sum, these special education teachers had not readily adopted role change to a more indirect service-delivery model, facilitating the work of early-childhood general educators working daily in classrooms. Instead, they periodically provided direct service on an itinerant basis and performed case-management tasks.

In another study, Wesley *et al.* (2000) explored the comfort level of special education consultants with their work. Using a profile template that was also used to indicate the severity level of children's disabilities in several areas of functioning, consultants indicated their level of comfort in consulting in specific areas. They were asked to designate how severely involved the child's disability could be before they were no longer comfortable in the

consultant role. Wesley *et al.* found that the greatest areas of discomfort were in behavioral issues, communication, and multiple disabilities. There was some indication that consultants were more comfortable serving more severely involved children in a direct service role than in a consultant role. In addition, there were differences among the consultants. Participants expressed higher comfort levels when they had more years of job experience as a consultant and when they had experienced more training in consultation processes.

Based on the literature, transdisciplinary collaboration, sometimes, provides the greatest challenges to personnel. Role change may be even more demanding than other skills that are involved in collaborative teaming processes. Wesley and Buysse (2004) found that consultants "appeared to be operating from the belief that providing consultation is parallel to providing direct early intervention services" (p. 133). Additionally, they found little discussion among the consultants they interviewed with regard to basic concepts and processes of consultation. In other words, role change at a fundamental level did not seem to have occurred.

Effectiveness of Collaborative Models

As it appears, meaningful collaboration among professionals from different disciplines is neither easy nor typical. Barriers to collaboration include time constraints, lack of commitment among team members, difficulty in establishing regular communication channels, lack of naturally occurring opportunities for frequent contact, personality differences, lack of training in roles and processes, and differences in philosophy and approach (Hinojosa *et al.*, 2001; Malone *et al.*, 2001). Wesley and Buysse (2004) noted that consultants reported they were more comfortable with their jobs when they had access to experts, technology, and training. Training in team strategies seems to be important as well for early-childhood educators operating in collaborative models (Malone *et al.*, 2001). Clearly, collaboration can be made more difficult by structural differences between the fields involved. Noonan *et al.* (2003) noted that education level varied systematically among their participants. Most of the special educators in their teaching teams held Master's degrees. Early-childhood educators were equally split between those with a bachelor's degree and those with an associate's degree or the CDA credential. They speculated that these disparities in professional development might have played a part in how team members perceived their similarities and differences with each other. These systematic differences in preservice preparation are quite common between special education and general education at the early-childhood level. In addition, the emphases of the preparation programs are dissimilar. Wesley and Buysse (2004) found that consultants were frustrated and

reported more discomfort when they believed the knowledge, skills, and attitudes of the classroom general education teachers with whom they were working differed from their own.

Malone *et al.* (2001) surveyed general education teachers who participated in teams with a variety of self-described structures, including interdisciplinary and multidisciplinary structures. They found that these early-childhood teachers held generally positive attitudes on teamwork scales. Teachers' viewpoints with regard to the performance characteristics of their teams (e.g., positive communication, role clarity, cooperation, conflict resolution, equal power) were, generally, positive. Malone *et al.* concluded that, "general education teachers are well aware of effective strategies to promote teamwork to support children with developmental concerns" (p. 588). Of course, awareness does not guarantee that collaboration will result among team members. In an interesting study, Campbell and Halbert (2002) asked service coordinators and providers in early-childhood special education to describe up to three wishes that would serve to make early-intervention systems high quality. Almost 15% of the responses related to teaming. Respondents' wishes for more communication ranged from the very concrete desire for more time to meet to changes in organizational infrastructures that would facilitate communication. Campbell and Halbert noted that wishes formed around teaming were more common among allied health professionals and among professionals with less experience. Clearly, while there exists a general literature across fields with regard to collaborative practices, there are also practice issues unique to fields. Wesley and Buysse (2004) carefully described the challenges and issues unique to early-childhood settings that may arise during each stage of the consultation process. Additionally, they proposed responses to these challenges. Expansion of this type of literature across various forms of collaborative practice in early-childhood settings would be beneficial for the field.

A key area to consider in assessing the state of the art in interdisciplinary collaboration is the experiences provided in preservice education. To what degree are professionals prepared to collaborate effectively with professionals from other fields? Research has been conducted to examine both the emphases of programs and the functioning of higher education programs. In regard to the latter, Mellin and Winton (2003) conducted surveys among faculty from higher education programs in education, medicine/nursing, allied health, and the social sciences. They found that only a small percentage of faculty activities related to preservice personnel preparation were conducted from an interdisciplinary point. Most of the variance among the participants, on this point, was predicted by work-environment variables, and these were more likely to pose barriers rather than facilitate interdisciplinary efforts. Bruder and Dunst (2005) surveyed faculty in higher education from a variety of fields. Their queries focused on course coverage

of topic areas deemed indicative of best practices. These included family-centered practice, natural environments, individualized family service plan (IFSP) processes, teaming, and service coordination. The latter two areas – both related to collaboration – were reported as receiving less coverage than the other topics. In addition, less emphasis was reported on teaming in allied health programs, compared to early-childhood special education and multidisciplinary programs. Not long ago, Kamens (2007) explored a more unique form of preservice preparation in a small study of general and special education students placed in a co-teaching student-teaching experience. She noted that there were both positive and less beneficial impacts of this experience. For instance, student teachers appeared to benefit from the peer support available through co-teaching. However, they expressed concerns with regard to being well-prepared to manage a classroom on their own. Being forced to negotiate relationships with their partner-teacher furthered students' understanding of the impact of both their own and other personalities. While student-teachers were placed with team of co-teachers, Kamens noted that their relationships as co-teachers did not necessarily mimic the relationship and style of their cooperating teacher models. In other words, the student-teacher teams developed unique interaction patterns through their work together.

Clearly, the aforementioned studies reflect that there is more work to be done within higher education and other levels to make collaboration effective. Undergraduate students are unlikely to directly experience models of collaborative practice, as faculty were less likely to report interdisciplinary teaming in teaching activities compared to research (Mellin and Winton, 2003). Furthermore, questions remain regarding course coverage. When and where should students gain competence in collaborative practices? Can this material, effectively, be embedded in a course? Should it be the focus of a stand-alone course? Can students be effectively prepared to work across multiple collaborative models, for example, interdisciplinary and transdisciplinary paradigms? In her study, Kamens (2007) noted that general education students had no specific coursework focused on collaboration, while for the special education students, this topic was covered in "some class sessions" (p. 157).

Promising Directions on Collaboration

Throughout this article, directions for future research efforts have been presented. While there is a fairly well-developed literature on training for and implementing collaborative models (see, for instance, the section titled 'Further reading'), more remains to be done.

Noonan *et al.* (2003) raised the point that there has been little examination of the role of culture in co-teaching relationships. Indeed, among the variables explored in research, this one has not been directly named. Differences

in educational preparation, personality, and style have been included in the literature, but there has been no study of how cultural background (including variations in race, ethnicity, class, language, and sexuality) may more systematically impact teaming attempts. In addition, separate systems from which professionals work could be examined from the perspective of culture. Barrera and Corso (2002) focused cross-cultural communication among professionals and families and provided a model for equalizing power differentials among groups. In addition, they stressed on accessing and integrating the strengths of different perspectives into the work of family-professional teams. Cultural differences are not restricted to only parent-professional teams but may also be found among professionals themselves. This line of scholarship can further strengthen the practice of collaborative teams while research can be conducted to help professionals better understand how these differences might play out among those involved in interdisciplinary collaboration.

The field would benefit from further development of effective ways to cross traditional disciplinary boundaries. Wesley and Buyse (2001) proposed that a communities-of-practice approach could enhance collaboration and provide great promise for the field. Through communities of practice, a group of individuals can "formalize, broaden, and deepen" the manner in which they collaborate around ideas and problem-solving (Wesley and Buyse, 2001: 121).

Conclusion

As noted earlier, the barriers operating against effective collaboration across disciplines in early-childhood programs are both many and varied. Despite this, there is much to be found in the existing literature to guide practice. Collaborative teaming requires two-pronged attention from group members, attention to issues and challenges at hand in service delivery, as well as attention to group process. Effective collaborative groups must develop a shared awareness of their expectations, roles, and strategies for operating together. While a multidisciplinary model is efficient, it no longer represents the most effective marshalling of resources toward serving young children and their families most effectively. It is necessary to draw upon the advantages of interdisciplinary and transdisciplinary models to truly achieve this goal.

See also: Co-Teaching; Inclusion; Inclusion of Students with Special Needs in General Education Classrooms.

Bibliography

- Barrera, I. and Corso, R. M. (2002). Cultural competency as skilled dialogue. *Topics in Early Childhood Special Education* 22, 103–113.

- Bruder, M. B. and Dunst, C. J. (2005). Personnel preparation in recommended early intervention practices: Degree of emphasis across disciplines. *Topics in Early Childhood Special Education* **25**, 25–33.
- Campbell, P. H. and Halbert, J. (2002). Between research and practice: Provider perspectives on early intervention. *Topics in Early Childhood Special Education* **22**, 213–226.
- DEC (2008). Early childhood special education/early intervention (birth to age 8) professional standards with CEC common core. http://www.dec-sped.org/About_DEC/PositionConcept_Papers/Personnel_Standards.
- Dinnebeil, L. A., McInerney, W. F., and Hale, L. (2006). "Shadowing" itinerant ECSE teachers: A descriptive study of itinerant teacher activities. *Journal of Research in Childhood Education* **21**, 41–52.
- File, N. and Kontos, S. (1998). Preparing personnel for integration: Needs and strengths of early intervention and early childhood education professionals. *Journal of Early Childhood Teacher Education* **19**, 181–191.
- Hinojosa, J., Bedell, G., Buchholz, E. S., et al. (2001). Team collaboration: A case study of an early intervention team. *Qualitative Health Research* **11**, 206–220.
- Hunt, P., Soto, G., Maier, J., Liboiron, N., and Bae, S. (2004). Collaborative teaming to support preschoolers with severe disabilities who are placed in general education early childhood programs. *Topics in Early Childhood Special Education* **24**, 123–142.
- Kamens, M. W. (2007). Learning about co-teaching: A collaborative student teaching experience for preservice teachers. *Teacher Education and Special Education* **30**, 155–166.
- Malone, D. M., Gallagher, P. A., and Long, S. R. (2001). General education teachers' attitudes and perceptions of teamwork supporting children with developmental concerns. *Early Education and Development* **12**, 577–592.
- Mellin, A. E. and Winton, P. J. (2003). Interdisciplinary collaboration among early intervention faculty members. *Journal of Early Intervention* **25**, 173–188.
- NAEYC (2001). NAEYC standards for early childhood professional preparation: Initial licensure programs. <http://www.naeyc.org/files/ncate/file/initiallicensure.pdf>.
- Noonan, M. J., McCormick, L., and Heck, R. H. (2003). The co-teacher relationship scale: Applications for professional development. *Education and Training in Developmental Disabilities* **38**, 113–120.
- Ogletree, B. T., Bull, J., Drew, R., and Lunnen, K. Y. (2001). Team-based service delivery for students with disabilities: Practice options and guidelines for success. *Intervention in School and Clinic* **36**, 138–145.
- Rich, A. (ed.) (2002). Benchmarks of recommended practice: Needs assessment for successful inclusion of infants and toddlers with disabilities in natural settings. (ERIC Document Reproduction Service No. ED474360). Norge, VA: Child Development Resources.
- Wesley, P. W. and Buysse, V. (2001). Communities of practice: Expanding professional roles to promote reflection and shared inquiry. *Topics in Early Childhood Special Education* **21**, 114–123.
- Wesley, P. W. and Buysse, V. (2004). Consultation as a framework for productive collaboration in early intervention. *Journal of Educational and Psychological Consultation* **15**, 127–150.
- Wesley, P. W., Buysse, V., and Keyes, L. (2000). Comfort zone revisited: Child characteristics and professional comfort with consultation. *Journal of Early Intervention* **23**, 106–115.
- West, J. F. and Cannon, G. S. (1988). Essential collaborative consultation competencies for regular and special educators. *Journal of Learning Disabilities* **21**, 56–63.
- Winton, P. J. (2000). Early childhood intervention personnel preparation: Backward mapping for future planning. *Topics in Early Childhood Special Education* **20**, 87–94.

Further Reading

- Barrera, I., Corso, R. M., and Macpherson, D. (2003). *Skilled Dialogue: Strategies for Responding to Cultural Diversity in Early Childhood*. Baltimore, MD: Brookes.
- Buysse, V. and Wesley, P. W. (2001). Models of collaboration for early intervention: Laying the groundwork. In Blasco, P. (ed.) *Early Intervention Services for Infants, Toddlers, and their Families*, pp 259–293. Boston, MA: Allyn & Bacon.
- Buysse, V. and Wesley, P. W. (2004). *Consultation in Early Childhood Settings*. Baltimore, MD: Brookes.
- Dinnebeil, L., Buysse, V., Rush, D., and Eggbeer, L. (2007). Becoming effective collaborators and change agents. In Winton, P. J., Catlett, C., and McCollum, J. A. (eds.) *Practical Approaches to Early Childhood Professional Development: Evidence, Strategies, and Resources*, pp 227–246. Washington, DC: Zero to Three Press.
- Dinnebeil, L., McInerney, W., and Hale, L. (2006). Understanding the roles and responsibilities of itinerant ECSE teachers through Delphi research. *Topics in Early Childhood Special Education* **26**, 153–166.
- Fleming, J. L. and Monda-Amaya, L. E. (2001). Process variables critical for team effectiveness: A Delphi study of wraparound team members. *Remedial and Special Education* **22**, 158–171.
- Kloo, A. and Zigmond, N. (2008). Coteaching revisited: Redrawing the blueprint. *Preventing School Failure* **52**, 12–20.
- Miller, P. S., Fader, L., and Vincent, L. J. (2000). Preparing early childhood educators to work with children who have exceptional needs. In Horm-Wingerd, D. and Hyson, M. (eds.) *New Teachers for a New Century: The Future of Early Childhood Professional Preparation*, pp 89–122. Jessup, MD: U.S. Department of Education.
- Miller, P. S. and Stayton, V. D. (2006). Interdisciplinary teaming in teacher preparation. *Teacher Education and Special Education* **29**, 56–68.
- Pugach, M. C. and Johnson, L. J. (2002). *Collaborative Practitioners, Collaborative Schools*, 2nd edn. Denver, CO: Love.
- Wesley, P. W. and Buysse, V. (2006). Ethics and evidence in consultation. *Topics in Early Childhood Special Education* **26**, 131–141.

Language Development in Children with Special Needs

M S L Stewart, University of Central Florida, Orlando, FL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Accents – The sound differences of spoken languages. They may be derived from one's geographic region and/or the influences of foreign languages.

Articulation disorder – A common speech disorder that results when individuals are not able to produce the various sounds and sound combinations of language at the developmentally appropriate age.

Collaboration – A process whereby professionals, including parents and primary caregivers, work together for a shared purpose in a mutually beneficial, supportive relationship.

Communication – A written or spoken transmission and reception of information, which has at least three components: sender, receiver, and message.

Communicative difference/dialect – A variation of a symbol system and reflects shared regional, social, or cultural/ethnic factors.

Fluency disorder – A communication disorder that occurs when the individual has difficulty with the flow and rate of speech, which may be accompanied by struggle behavior.

Language disorder – A communication disorder that involves understanding or producing meaningful conversation.

Morphemes – The smallest linguistic unit of meaning in a language; forms a bridge between phonology and syntax.

Phonology – The system of rules that governs sounds and their combination.

Pragmatics – The actual use of language within a social context.

Semantics – The content of language.

Speech disorder – A term used when the individual's communication is unintelligible, unpleasant, or interferes with communication.

Speech-language pathologists – The professionals who treat speech and language disorders.

Syntax – The arrangement of words combined by following the rule governance and patterns of the specific language to form meaningful sentences.

Voice disorder – A communication disorder wherein there is difficulty with the quality, resonance, pitch, or intensity of one's voice.

One's language is personal, interactive, and original. It is the most human characteristic that helps to shape one's cultural and personal identity, and to a large extent, demonstrates what separates humans from other beings. It celebrates who we are and from whence we have come. It then should come as no surprise that as the nation's schools become more diverse, more challenging issues related to communication, language acquisition, language differences, and language disorders arise. To the extent possible, all students should develop language and communication skills that will aid them in being successful in both school and community settings. Learners who experience difficulty in absorbing information through listening and reading, and who are unable to express themselves in spoken words are almost certain to experience difficulties, not only in their schools, but in their communities as well. This difficulty may have an impact on the individual's ability to form satisfying relationships with others (Table 1).

The ability to effectively communicate is one of the most important aspects of an individual's development. In any setting, there may be a variety of individuals present who do not speak at all, those requiring assistive/augmentative communications, those who must learn to speech read and operate assistive listening devices, as well as those who display socially inappropriate verbal and non-verbal communication skills. An understanding of these disabilities, as well as a global understanding of language and speech development, is critical to our understanding of human beings. Children identified as special needs have many characteristics similar to those identified in typically developing children. These children may take longer to develop their speech, while others may have disabilities that are not recognized at birth but may become more apparent over time. Others may experience difficulties in school, ranging from problems with concentration, learning, language, and perception to problems with behavior. (American Academy of Child and Adolescent Psychiatry, 2008). This article provides a framework for understanding terms/concepts concerning normal speech and language development, communication, language disorders, and language differences in students with special needs. Further, it presents information on how professionals may collaborate with families/primary caregivers as a way of sharing information that should lead to expanded information for all parties involved. In addition, specific information regarding students with disabilities that may impact language development is discussed.

Table 1 Identifying children with language problems

Directions: The following behaviors may indicate that a child in your classroom has a language impairment that is in need of language intervention. Please check the appropriate items.

- ☐ Child mispronounces sounds and words.
- ☐ Child omits words, endings, such as plural-s and past tense-ed.
- ☐ Child omits small, unemphasized words, such as auxiliary verbs or prepositions.
- ☐ Child uses an immature vocabulary, overuses empty words, such as 'one' and 'thing', or seems to have difficulty recalling or finding the right word.
- ☐ Child had difficulty comprehending new words and concerns.
- ☐ Child's sentence structure seems immature or over-reliant on forms, such as subject-verb-object. It's unoriginal, dull.
- ☐ Child has difficulty with one of the following:
 - ☐ Verb tensing ☐ Articles ☐ Auxiliary verbs
 - ☐ Pronouns ☐ Irreg. verbs ☐ Prepositions
 - ☐ Word Order ☐ Irregular plurals
- ☐ Child has difficulty relating sequential events.
- ☐ Child has difficulty following directions.
- ☐ Child's questions often poorly formed.
- ☐ Child has difficulty answering questions.
- ☐ Child's comments often off topic or inappropriate for the conversation.
- ☐ There are long pauses between a remark and the child's reply or between successive remarks by the child. It's as if the child is searching for a response or is confused.
- ☐ Child appears to be attending to communication but remembers little of what is said.

Language: Definition and Attributes

It is generally agreed that a child's language skills are acquired in a predictable order. Children, however, may differ at the rate in which they acquire a given language skill. This can even occur with children in the same family group. Language, as defined by the American Speech-Language-Hearing Association (ASHA), is a "complex and dynamic system of conventional symbols that is used in various modes for thought and communication." The development of language, often hailed as the hallmark accomplishment of humans, is rule-governed with specific conventions that are learned. In humans, language is dependent upon one having a society in which to learn it, other human beings to speak to, and the intelligence to make it possible.

The development of language is a highly complex process. It begins during infancy and continues throughout one's life. Young children normally progress through several stages in developing language, and as previously highlighted, normally exhibit considerable variability in their stages of development. These differences may be due, in part, to general health, environmental influences, and heredity. Specific causes may include hearing and visual impairments, difficulty with oral-motor skills as found in children with specific syndromes, such as Down syndrome, fetal alcohol syndrome, cerebral palsy, cleft lip/palate, and Prader-Willi syndrome. Often the cause can be unknown.

It is generally agreed that today's youngsters come from more diverse cultural and linguistic backgrounds. These learners exhibit a wide range of language, learning, and behavioral characteristics that present challenges to

educators. Some are at risk of academic failure and placed in special education settings because of their limited English proficiency, behavioral characteristics, and/or socioeconomic status. Nearly 4 million of our children under the age of 18 are being reared by their grandparents or great-grandparents; many live in, or near, poverty. Factors for this occurrence include parents not being able to care for their own children because of incarceration and drug and/or alcohol abuse.

The United States is a society made up of many cultures and languages. Each person uses language differently. For one to fully appreciate or understand a language, its use, diversity, and disorders, an understanding of the broader context of language is required. Following are the attributes of language:

- language evolves within specific historical, social, and cultural contexts;
- language is rule-governed;
- language learning and use are determined by the interaction of biological, cognitive, psychological, and environmental factors; and
- effective use of language for communication requires a broad understanding of human interactions, including associated factors such as nonverbal cues, motivation, and sociocultural role.

Oftentimes, language is described along dimensions such as form (phonology, morphology, and syntax), content (semantics), and use (pragmatics). Phonology is the system of rules that governs sounds and their combination. Each language has specific sounds, or phonemes, that are characteristic of that language. A phoneme is further defined as the smallest unit of sound in a word that makes

a difference in its meaning. Morphology is the system of rules for combining sounds into meaningful units, such as words, suffixes, and prefixes. Simply stated, morphemes are the smallest linguistic unit with meaning and form a bridge between phonology and syntax. Syntax refers to the arrangement of words combined by following the rule governance and patterns of the specific language to form meaningful sentences. The content of language is referred to as the semantics of the language. The concept of semantics refers to the ability to distinguish word meaning, including multiple meanings and subtle nuances, and to understand the language. The actual use of language within a social context is referred to as pragmatics; moreover, it is the awareness of socially appropriate behaviors in communication interactions.

Developmental Milestones: Communication Skills in Young Children

Researchers such as Bernstein and Tiegerman-Farber (2002) and Owens (1988) presented detailed summaries of commonly agreed-upon communication developmental milestones in young children. They are:

- *At 12 months.* Infants recognize their own name and can follow simple motor instructions; they produce the first word and express vocabulary of one or more words.
- *At 15 months.* Infants have a vocabulary of four to six words.
- *At 16 months.* Infants develop the ability to point to toys, persons, and animals. They use jargon and words in conversation with a four-to-six-word vocabulary.
- *At 18 months.* Infants begin to use two-word utterances and have a 20-word vocabulary.
- *At 21 months.* Infants understand personal pronouns; they use 'I' and mime.
- *At 24 months.* Infants have a 200–300-word vocabulary and name common everyday objects.
- *At 36 months.* Infants have a vocabulary of 900–1000 words; create three- to four-word sentences and follow two-step commands.
- *At 48 months.* Infants have a vocabulary of 1500–1600 words and use increasingly complex sentences.
- *At 60 months.* Infants have a vocabulary of 2100–2200 words, follow three-step commands, and have 90% grammar acquisition.

Note that children's receptive vocabulary, oftentimes, exceed their expressive vocabulary. This simply means that they understand far more than they are able to verbally express. However, there are children who may not have normal developments. They may have special needs and speech/language impairment. According to the 27th Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act (IDEA)

by the US Department of Special Education, in 2003, the largest disability category was specific learning disabilities (47.4%). The next most common disability category was speech or language impairments (18.7%), followed by mental retardation (9.6%), emotional disturbance (8.0%), and other health impairments (7.5%).

Language Difference versus Language Disorder

In view of the demographic shifts in our nation's schools, it is important to understand communication in the context of students. Communication is considered to be disordered when it deviates from the standards of the community, enough to interfere with the transmission of messages and stands out as being unusually different, or it produces negative feelings within the communicator. The disorder may either be developmental or acquired through injuries or diseases that affect the brain. Conversely, speech disorders result when the individual's communication is unintelligible, unpleasant, or interferes with communication.

Communication involves languages. How then, can communication and/or language disorders be further explained? There are different kinds of communication disorders. Articulation disorders, perhaps the most common speech disorders, result when individuals are not able to produce the various sounds and sound combinations of language at the developmentally appropriate age. A fluency disorder occurs when the individual has difficulty with the flow and rate of speech. The two most common types of fluency disorders are stuttering (i.e., an interruption of the forward flow of speech) and cluttering (i.e., speech that is overly rapid, disorganized, and occasionally filled with unnecessary words and unrelated insertions). A voice disorder is one wherein there is difficulty with the quality, resonance, pitch, or intensity of one's voice. Students with language disorders may have delays in the development of comprehensive or expressive language. The question continues to be, when is a language difference misinterpreted to mean language disorder?

ASHA, the national professional, scientific, and credentialing association for speech, language, and hearing professionals, notes that English language is composed of many linguistic varieties. This may include Ebonics, Spanish-influenced English, Appalachian English, and Standard English. Accents are sound differences of spoken languages. They may be derived from one's geographic region and may be influenced by foreign languages. A communicative difference or dialect, on the other hand, can be a variation of the form and content of a language. Accents and dialects are types of language differences. A language difference, in and of itself, may not require intervention and remediation; however, those students who are found to be language disordered by a team of professionals will require

intervention and remediation. Language can be described as being disordered when it interferes with communication, calls unfavorable attention to itself, or causes its user to be maladjusted. In earlier studies, ASHA noted that an individual with a language disorder may have problems in sentence processing and in abstracting information meaningful for storage and retrieval from short- and long-term memory.

A language difference is not necessarily a language disorder. As previously explained, language contains a variety of forms, called dialects. Reiterating, dialects are a variation of a symbol system and reflect shared regional, social, or cultural/ethnic factors. Each dialect represents a functional form of English, permitting speakers to communicate adequately, while at the same time maintaining a symbolic representation of the historical, social, and cultural background of the speakers. Although some dialect speakers may have a speech or language disorder, the dialect itself is not the disorder. The presence of a dialect should not necessarily rule out the possibility that a language disorder might exist. A language disorder is a problem in understanding or producing meaningful conversation and might involve

1. a serious disruption of the language acquisition process;
2. difficulty with following rules of grammar;
3. the ability to understand or use words in correct context;
4. not choosing appropriate language for different situations; and
5. problems of written or spoken language, and/or other symbol systems.

Language disorders may vary in severity, ranging from mild to severe.

Specific Disabilities and Language Development

Some individuals may have specific disabilities that may impact language development. This may include the categories of individuals with mental disabilities, cognitive disabilities, learning disabilities, or those categorized with behavioral disorders. The existence of cognitive disabilities usually is identified early in the young child's life, due to the developmental delay evidenced by the child's overall functioning in gross and fine motor skills and usually confirmed with intellectual and behavioral assessments. Cognitive disability has been defined as intellectual functioning that is significantly below average, observed during the developmental years, and accompanied by deficits in adaptive behavior. The regulations for IDEA provide the following technical definition for cognitive disability:

Students with cognitive disability may follow the developmental sequence of language acquisition, use, and

comprehension, but they will not progress at the same rate of their intellectual functioning. Therefore, the extent and rate of acquisition of their total development of language will not be that for "typical" learners. Therefore, it is critical to clearly delineate the needed language and communication goals for the individual student considering his or her long-term life outcomes through the Individualized Education Planning (IEP) process.

Students with learning disabilities may exhibit a combination of characteristics, which may mildly, moderately, or severely impair the learning process. Learning disabilities are characterized by a significant difference in the child's achievement in some areas, as compared to the overall intelligence. Academically, students with learning disabilities may exhibit a wide range of traits, including problems with language as evidenced in reading comprehension, spoken language, written language, and reasoning abilities. Receptive and expressive language, both in the written and oral forms, can be a significant area for remediation. Difficulties will be evidenced in reading, writing, spelling, and comprehending oral and written directions across each of the academic settings. The discrepancies between expectations and achievement for students with learning disabilities increase without intense remediation in these areas.

As the exact numbers of students with learning disabilities are difficult to accurately establish, so also are the terms and defining characteristics as elusive when describing and identifying students who have been described with emotional, behavioral, or mental disorders. The federal definition of serious emotional disturbances (SEDs) includes:

A condition exhibiting one or more of the following characteristics over a long period of time and to a marked degree that adversely affects educational performance:

- An inability to learn that cannot be explained by intellectual, sensory, or health factors;
- An inability to build or maintain satisfactory interpersonal relationships with peers or teachers;
- Inappropriate types of behavior or feelings under normal circumstances;
- A general pervasive mood of unhappiness or depression; or;
- A tendency to develop physical symptoms or fears associated with personal or school problems. (Code of Federal Regulations, Title 34, 300.7(b)(9))

The causes of SED have not been adequately determined. Although various factors, such as family functioning, heredity, diet, brain trauma or disorder, and stress, have been suggested as possible causes (NICHCY, 1996), research has not shown any of these factors to be a direct cause. Given the social and emotional nature of the defining characteristics of this disability, the ramifications in the areas of speech, language, and communication are

evident, especially with communication. Social interactions, as defined by the appropriate use of pragmatics and vocabulary, are the greatest areas of needs evidenced by students with serious emotional disturbances. Because language and communication are essential components of interpersonal relationships, children with SED may experience difficulties giving and receiving information from others. Consequently, language skills seem to impact upon children's social skills.

If a professional/service provider is unsure about whether the child has a speech or language problem, he/she might refer the child for testing by a speech, language, or hearing professional. Speech-language pathologists are also specially trained to serve as consultants to teachers and other professionals on dialectal variations and modifications. Further, they are able to provide individual therapy for the child; consult with the child's teacher about the most effective ways to facilitate the child's communication in the class setting; and work closely with the family to develop goals and techniques for effective therapy in the classroom as well as at home. The use of assistive technology may help children whose physical conditions make communication difficult.

Proactive Strategies for Working with Students with Communication Disorders

To work with exceptional learners with speech and language disorders or communications disorders, it is critical to form and maintain partnerships that effectively facilitate the community of learning that occurs in the home, school, and community environments. The types of parent-teacher-community partnerships that are formed and maintained are related to an understanding of both multicultural students with exceptionalities and their diverse cultural and linguistic backgrounds. As a matter of fact, ASHA noted that:

In order to effectively meet the needs of an increasingly diverse multicultural clientele, it is very helpful for speech-language-hearing professionals to be aware of the historical, social and political factors that have contributed to the development of various speech communities throughout the United States. (American Speech-Language-Hearing Association, 2005)

Several critical factors concerning families of children with special needs and their involvement with schools are important to note. The participation of parents in an educational partnership with school staff may be affected by their comfort level. For some parents/family members of children with special needs, schools are not often seen as very user-friendly places. If parents/family members feel uncomfortable with the school's conceptualization of their parent involvement, they may choose to abstain from

any of the roles made available to them by school personnel. Many parents with limited verbal skills and little experience in negotiating the educational system may find it difficult to identify and access appropriate educational resources for their children. Dealing with any bureaucracies, including school systems, can be a frustrating and overwhelming process. These parents may rarely have the capacity to participate in educational partnerships, especially when they sense that they are dealing with a system that does not recognize that they suffer from the double-whammy of dealing with both race and disability.

For parents who have had negative experiences themselves with schools, this may present a barrier to their participation as partners in the educational process. Some of these negative experiences may be more subtle; other experiences may be more overt. For example, when parents feel that there is little or no attempt to accurately represent the historical contributions of African Americans in the school curriculum, they may be reluctant or even unwilling to form partnerships with teachers and others involved in the education of their children. Parents, including those who have children with exceptionalities, usually understand that knowledge of one's racial identity and cultural heritage are critical factors in promoting positive self-esteem in some children and in socializing them to effectively cope with racism.

Conclusion

Most would agree that the United States is a nation comprised of persons from different backgrounds, experiences, ethnic and cultural groups, and expectations. As a result, it is the responsibility of teachers and service providers to provide academic and social-emotional support that empowers both students and families with special needs. For many students with special needs, especially those that are culturally and linguistically diverse, problems of misidentification, labeling, and inappropriate schooling still remain. Specifically, educators must value community and family involvement. In addition, they must understand when students have language disorders or language differences as they explore the different types of communication disorders. Finally, they must understand that collaboration with the child's parents/primary caregivers can have a major impact on the child's communicative development and skills. This is true for all children, but especially for students whose first language is not the majority language.

Bibliography

American Academy of Child & Adolescent Psychiatry (2008). *Services in School for Children with Special Needs: What Parents Need to Know*. 89. Retrieved May 22, 2008, from <http://www.aacap.org/cs/root/>

- facts_for_families/services_in_school_for_children_with_special_needs_what_parents_need_to_know (accessed August 2009).
- American Speech-Language-Hearing Association (2005). *Multicultural History and Demographic Profile of the United States*. http://www.asha.org/about/leadership-projects/multicultural/readings/reading_1.htm (accessed August 2009).
- Bernstein, D. K. and Tiegerman-Farber, E. (2002). *Language and Communication Disorders in Children*, 5th edn. Boston, MA: Allyn and Bacon.
- NICHCY (National Dissemination Center for Children with Disabilities) (1996). *Educating Students with Emotional/Behavioral Disorders*. <http://www.nichcy.org/pubs/bibliog/bib10txt.htm> (accessed August 2009).
- Owens, R. E., Jr (1988). Child development, In *Language Development: An Introduction*, 2nd edn., pp 63–103. Columbus, OH: Merrill.

Further Reading

- Adger, C. T., Wolfram, W., and Detwyler, J. (1993). Language differences: A new approach for special educators. *Teaching Exceptional Children* **26**, 44–48.
- American Speech-Language-Hearing Association (1982). Definitions: Communication disorders and variations. *ASHA* **24**, 949–950.
- American Speech-Language-Hearing Association (1983). Position paper: Social dialects. *ASHA* **25**, 23–24.
- American Speech-Language-Hearing Association (2001). *Communicating in a Diverse Society*. http://www.asha.org/about/news/tipsheets/Diverse_society.htm (accessed August 2009).
- Anderson, N. B. and Shames, G. H. (2006). *Human Communication Disorders: An Introduction*, 7th edn. Boston, MA: Allyn and Bacon.
- Bank Street College (2008). English language learners: Working with students with whom English is a new language. *Bank Street's Guide to Literacy*. <http://www.bankstreet.edu/literacyguide/ell.html> (accessed August 2009).
- Berk, L. (2000). *Child Development*, 5th edn. Boston, MA: Allyn and Bacon.
- Carmack, F. and Thomas, P. J. (1990). *Speech and Language: Detecting and Correcting Special Needs*. Boston, MA: Allyn and Bacon.
- Craig, H. K. and Washington, J. A. (2000). An assessment battery for identifying language impairments in African American children. *Journal of Speech, Language and Hearing Research* **43**, 366–379.
- Craig, H. K., Washington, J., and Thompson-Porter, C. (1998). Performances of young African American children on two comprehension tasks. *Journal of Speech, Language, and Hearing Research* **41**, 445–457.
- Culatta, R., Tompkins, J., and Werts, M. (2003). *Fundamentals of Special Education: What Every Teacher Needs to Know*. Boston, MA: Allyn and Bacon.
- Delpit, L. (2002). Introduction. In Delpit, L. and Dowdy, J. (eds.) *The Skin that We Speak: Thoughts on Language and Culture in the Classroom*, p xvii. New York: New Press.
- Delpit, L. (2006). *Other People's Children. Cultural Conflict in the Classroom*. New York: New Press.
- Dev, P., Doyle, B., and Valente, B. (2002). Labels needn't stick: "At risk" first graders rescued with appropriate intervention. *Journal of Education for Students Placed at Risk* **7**(3), 327–332.
- Dockrell, J. E. (2001). Assessing language skills in preschool children. *Child and Adolescent Mental Health* **6**(2), 74–85.
- Hardman, M. L., Drew, C. J., and Egan, M. W. (2008). *Human Exceptionality: School, Community, and Family*, 9th edn. Boston, MA: Houghton Mifflin.
- Hedge, M. N. (1995). *Introduction to Communication Disorders*. Austin, TX: Pro-Ed.
- Heward, W. L. (2005). *Exceptional Children*, 8th edn. Upper Saddle River, NJ: Pearson Education.
- Johansson, I. (1994). *Language Development in Children with Special Needs: Performative Communication*. London: Jessica Kingsley.
- Kuder, S. J. (2007). *Teaching Students with Language and Communication Disabilities*, 2nd edn. Boston, MA: Allyn and Bacon.
- Lindsay, G. and Dockrell, J. E. (2004).
- Lue, M. S. (2001). *A Survey of Communication Disorders for the Classroom Teacher*. Boston, MA: Allyn and Bacon.
- Mercer, C. D. and Pullen, P. C. (2005). *Students with Learning Disabilities*, 6th edn. Upper Saddle River, NJ: Pearson/Merrill/Prentice-Hall.
- Moeller, M. (2000). Early intervention and language development in children who are deaf and hard of hearing. *Pediatrics* **106**(3). <http://pediatrics.aappublications.org/cgi/content/abstract/106/3/e43> (accessed August 2009).
- Morton, R., Sharma, V., Nicholson, J., Broderick, M., and Poyser, J. (2002). Disability in children from different ethnic populations. *Child: Care Health and Development* **28**(1), 87–93.
- National Center on Educational Statistics (2005). Language minority school age children: Participation in education. *Condition of Education*. <http://nces.ed.gov/programs/coe/2005/section1/indicator05.asp> (accessed August 2009).
- Nieto, S. and Bode, P. (2008). *Affirming Diversity. The Sociopolitical Context of Multicultural Education*, 5th edn. Boston, MA: Pearson.
- Obiakor, F. E. (ed.) (2007). *Multicultural Special Education. Culturally Responsive Teaching*. Upper Saddle River, NJ: Pearson Merrill/Prentice-Hall.
- Owens, R. E., Jr., Metz, D., and Haas, A. (2007). *Introduction to Communication Disorders: A Lifespan Perspective*, 3rd edn. Boston, MA: Allyn and Bacon.
- Payne, K. T. and Taylor, O. L. (1998). Communication differences and disorders. In Shames, G. H., Wiig, E. H., and Secord, W. A. (eds.) *Human Communication Disorders: An Introduction*, 5th edn., pp 118–154. Boston, MA: Allyn and Bacon.
- Pierangelo, R. (2003). *The Special Educator's Book of Lists*. San Francisco, CA: Jossey-Bass.
- Polloway, E., Patton, J., and Serna, L. (2001). *Strategies for Teaching Learners with Special Needs*, 7th edn. Upper Saddle River, NJ: Merrill/Prentice-Hall.
- Porter, L. (2002). *Educating Young Children with Special Needs*. Thousand Oaks, CA: Sage.
- Roseberry-McKibbin, C. (1995). Distinguishing language differences from language disorders in linguistically and culturally diverse students. *Multicultural Education* **2**, 12–16.
- Sileo, T. W., Sileo, A. P., and Prater, M. A. (1996). Parent and professional partnerships in special education: Multicultural considerations. *Intervention in School and Clinic* **31**, 145–153.
- Stone, C. A., Silliman, E. R., Ehren, B. J., and Apel, K. (eds.) (2004). *Handbook of Language and Literacy*. New York: Guilford.
- Turnbull, R., Turnbull, A., Shank, M., Smith, S., and Leal, D. (2002). *Exceptional Lives: Special Education in Today's Schools*, 3rd edn. Upper Saddle River, NJ: Merrill/Prentice-Hall.
- Van Riper, C. and Erickson, R. (1996). *Speech Correction: An Introduction to Speech Pathology and Audiology*, 9th edn. Boston, MA: Allyn and Bacon.
- Vaughn, S., Bos, C., and Schumm, J. (2000). *Teaching Exceptional, Diverse, and at-Risk Students in the General Education Classroom*, 2nd edn. Boston, MA: Allyn and Bacon.

Relevant Websites

- <http://www.ataccess.org> – Alliance for Technology Access.
- <http://www.asha.org> – American Speech–Language–Hearing Association.
- <http://www.childrensdisabilities.info> – Children's Disabilities Information.
- <http://www.cec.sped.org> – Council for Exceptional Children.
- <http://www.easterseals.com> – Easter Seals Disability Services.
- <http://lshss.asha.org> – Language, Speech, and Hearing Services in Schools.
- <http://www.ldanatl.org> – Learning Disabilities Association of America.
- <http://www.ldonline.org> – LD Online.

<http://www.nabe.org> – National Association for Bilingual Education.
<http://www.nads.org> – National Association for Down Syndrome.
<http://www.napcse.org> – National Association of Parents with Children
in Special Education, Speech and Language Impairments.
<http://www.nidcd.nih.gov> – National Institute on Deafness and Other
Communication Disorders.
<http://www.nncc.org> – National Network for Childcare-Children with
Disabilities or Special Needs.

<http://www.nsastutter.org> – National Stuttering Association.
<http://www.cfw.tufts.edu> – Tufts University, Child and Family.
<http://www.ed.gov> – US Department of Education, Twenty-Seventh
Annual Report to Congress on the Implementation of the Individuals
with Disabilities Education Act, Parts B and C.
<http://www.zerotothree.org> – Zero to three: Early Language and
Literacy.

Large Scale Assessment and Accountability and Students with Special Needs

M L Thurlow, University of Minnesota, Minneapolis, MN, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

1% assessment – Another name given to alternate assessments based on alternate achievement standards; this term is used because of the Elementary and Secondary Education Act accountability system, which limits the percentage of students who may be counted as proficient on this type of assessment to 1% of the total student population.

2% assessment – Another name given to alternate assessments based on modified achievement standards; this term is used because of the Elementary and Secondary Education Act accountability system, which limits the percentage of students who may be counted as proficient on this type of assessment to 2% of the total student population.

504 plan – A plan for accommodations that is developed for students with disabilities who do not require special education services.

Accessibility – A quality of those assessments that reveal the knowledge and skills of students whose disability, limited language, or other specific characteristics create barriers to accurate measurement through large-scale assessments of the students' knowledge and skills.

Accommodations – The changes in the materials or procedures used during large-scale assessments that do not change the intended content to be measured; these changes include adjustments to the presentation of materials, responses that students make, setting of the assessments, or timing or scheduling of the assessment.

Accountability – A system for holding a person or group of people responsible for a process or product. Large-scale assessments are used to hold school systems, educators, or students responsible for the performance of the students.

Assessment – A method for collecting data; large-scale assessments are tests or other data collection procedures given to large numbers of students at the same time for the purpose of documenting student achievement.

Disability – An impairment in the sensory, intellectual, learning, or behavioral aspects of a

student's learning and performance; in the United States, specific categories of disability are identified to group students with similar characteristics.

Growth models – An approach to system accountability that looks at the performance of cohorts of students, or that looks at the performance of students longitudinally, to determine whether there is improvement, or growth, in performance over time; these models are in contrast to status models, which compare different groups of students over time.

High stakes – The consequences that are significant to either the educational system or individuals within it; assessments themselves may be designated as high stakes based on the consequences assigned to results of the assessment.

Individualized education program (IEP) –

A written document that defines the goals and directs the provision of special education services to students with disabilities who need them; identifying the way in which students participate in large-scale assessments is part of the IEP.

Modifications – The changes in the materials or procedures used during large-scale assessments that change the intended content to be measured; these changes include adjustments to the presentation of materials, responses that students make, setting of the assessments, or timing or scheduling of the assessment.

Special education – The services provided to students whose disability creates barriers to learning or demonstrating learning in assessments.

Universal design – An approach to making instruction or assessments appropriate for the widest range of students; for large-scale assessments, emphasis is given to maintaining the intended content to be measured while attending to aspects of the assessment such as item comprehensibility and readability, legibility, and similar characteristics.

Testing is an integral part of education in the United States. A significant portion of the assessment activity is devoted to administering large-scale assessments, which are given to large numbers of students at the same time for the purpose of documenting student achievement.

Other testing is conducted to diagnose student eligibility for specific programs such as gifted education or special education. In the United States, all students, including those with special education needs, participate in large-scale assessment systems. Accountability is one purpose for which large-scale assessments are used. The assessments provide all or part of the information that is used to hold schools or individuals responsible for student achievement. Accountability consequences include reporting the performance of students to the public, altering the resources directed to individuals or schools, or imposing specific sanctions for poor performance.

The inclusion of students with special needs in large-scale assessments and accountability systems is important for promoting the positive consequences that they are intended to produce, and for avoiding unintended consequences that result from exclusion. Students with disabilities, English-language learners (ELLs), and ELLs with disabilities benefit from participation because the students are held to higher expectations and provided the same educational opportunities as other students, including increased access to the curriculum, additional resources for learning, and better outcomes. Unintended consequences that are avoided include lower expectations, exclusion from instructional opportunities and resources, and poor outcomes.

Large-Scale Assessments

Large-scale assessments vary on a number of dimensions, including the purpose of the assessments, the types of items that are included in the assessment, and the consequences of their use. Decisions about these dimensions affect the inclusion of students with special needs in the assessments.

Purpose of Assessments

Large-scale assessments are used for many purposes. They provide information used to determine whether students graduate or are promoted from one grade to the next. Assessment results help determine whether students are admitted to college or have the competency to be a teacher, a dentist, or another professional or worker. They are used to make decisions about whether schools, districts, or states have produced the levels of student achievement considered necessary for them to progress through school and become successful adults.

Many large-scale assessments are used for accountability purposes. They provide information that is used to hold a student, educator, school, or district responsible for student achievement. Responsibility has consequences that may range from public reporting of assessment results

to determining that a school needs to be restructured or closed so that students have a better opportunity to achieve at higher levels.

Assessment Options for Students with Disabilities

Large-scale assessments frequently are multiple choice tests with one or more extended response items that require students to write short answers or longer essays. Occasionally, large-scale assessments may be performance events, in which the student is asked to perform a task such as a science experiment. They also may be portfolios in which the student or the student's teacher gathers evidence of learning and achievement in specific content areas.

For students with disabilities, participation in large-scale assessment systems occurs in three primary ways: in the general assessment the same way as other students take the assessment, in the general assessment with accommodations, or in an alternate assessment. Individualized education program (IEP) teams make the decision about the way in which each individual student with disabilities participates in large-scale assessments. IEP team members also determine the specific accommodations that students may need to participate in the assessment.

General assessments

Many students with disabilities take the same large-scale assessments in the same way as their peers without disabilities. Most test developers ensure that the participation of students with special needs in these assessments is appropriate by complying with accessibility principles and incorporating elements of universal design.

Accessibility and universal design include the concepts of considering all students' needs at the beginning of item and test design, clearly defining the content to be measured, ensuring that all items are comprehensible, readable, and legible, and that assessments have clear, simple, and intuitive instructions. Some universal design advocates argue that assessments must allow many ways to present items and many ways for students to respond to assessments. Both accessibility and universal design address ways to allow the students with varying characteristics and needs to show what they know and can do. The notion that accessible and universally designed assessments will help produce more valid results for students with disabilities is often not understood. Some test developers assume that an accessible or universally designed assessment is just an easier assessment.

General assessments with accommodations

The majority of students with disabilities participate in large-scale assessment systems by using accommodations when they take the assessments their peers do.

Accommodations are changes in materials or procedures that enable students with disabilities to avoid the barriers that their disabilities create. Accommodated assessment performance is a more accurate reflection of the knowledge and skills of those students needing accommodations. Accommodations do not change what the test is intended to measure; they are generally distinguished from modifications, which are changes in materials or procedures that alter what the test is intended to measure.

Accommodations and modifications frequently are categorized according to the nature of the changes that are made. A common categorization scheme for accommodations and modifications is: presentation, response, timing/scheduling, and setting. Equipment or technology accommodations are included in these categories according to the function of the equipment or technology, or may be considered as a separate category of accommodations.

There are literally hundreds of accommodations or modifications identified in state policies that determine which changes are acceptable for the state's specific content standards and the purpose of the assessment. The policies are frequently updated, with variations occurring among states and over time for the same state. States rely on existing research, logical rationales, and input from stakeholder groups to determine whether changes to standard procedures should be considered accommodations or modifications for each particular assessment content area, purpose, and grade level. Controversy over which accommodations produce valid results for students with disabilities continues despite considerable research on the effects of accommodations.

Presentation accommodations or modifications are changes in how the testing materials are presented. Examples are braille and large-print editions of the assessment, reading part or all of the assessment to the student, using a screen reader, presenting the assessment in sign language, and allowing students or teachers to highlight specific words or phrases in the assessment. For this category, the accommodations or modifications most frequently used tend to not be those that are most often allowed. For example, reading part or all of the assessment is used more often than braille editions, but braille editions are allowed by more states than reading the assessment to the student.

Response accommodations or modifications are changes in how the student responds to the assessment. Examples are using a braille or a scribe to input responses, using a calculator to help determine answers, responding in sign language, and marking in the test booklet rather than on a bubble sheet that is scanned to record students' responses.

Timing or scheduling accommodations or modifications are changes in the duration of time that a student participates in testing, or in when the student participates. Examples of timing accommodations or modifications are extended time, providing extra breaks during the assessment, and allowing

students to take the test across multiple days when other students take the test during one day. Scheduling accommodations or modifications include conducting assessments for a student only during a certain part of the day, regardless of when other students are tested, or scheduling testing for times when it is best for the individual student, regardless of the schedule for all other students.

Setting accommodations or modifications are changes in where the assessment occurs. Examples of setting accommodations or modifications include administrations to a small group or individuals, testing in the special education classroom rather than the lunchroom with other students, in a carrel to avoid distractions, or in a hospital.

IEP teams making decisions about accommodations should consider the goals of the assessment, the nature of the student's characteristics and needs, and the policies that states or districts have set to guide accommodation policies. Numerous issues surround the implementation of accommodations. IEP teams should receive training on making accommodation decisions, and states and districts should monitor the use of accommodations on assessments. Taking these steps reduces documented inconsistencies between the accommodations identified by the IEP team and the accommodations used in the classroom and during assessments.

Alternate assessments

Alternate assessments are designed for students unable to participate in typical large-scale assessments. These assessments may be designed to hold students to the same grade-level achievement standards as other students, but in a different way such as by using a portfolio assessment rather than a multiple choice assessment.

Alternate assessments may also be designed to hold students to different performance or achievement standards. The different standards may be either alternate achievement standards or modified achievement standards. The exact meaning of these alternate and modified achievement standards is not always clear.

Most assessment developers consider alternate achievement standards to reflect truly different expectations that are consistent with the significant intellectual disabilities of the students typically participating in them. Modified achievement standards, in contrast to alternate achievement standards, are directly related to the achievement standards held for all students, but represent reduced difficulty. Even though the performance or achievement standards are adjusted in these two types of alternate assessments, the assessments are still linked or aligned to the grade level of the content of the assessment. Students held to alternate or modified achievement standards have expectations for their performance that are different in some way from those for other students.

The most common type of alternate assessment is the alternate assessment based on alternate achievement

standards. Some educators refer to these as 1% assessments, due to the percentage of students in the total population who can be counted as proficient on these assessments for the accountability requirements of the Elementary and Secondary Education Act. Alternate assessments based on alternate achievement standards generally look different from traditional assessments. They tend to be portfolio assessments, performance assessments, or rating scales.

When alternate assessments based on alternate achievement standards are portfolios, or bodies of evidence as they are sometimes called, they generally involve the teacher collecting samples of student work, perhaps over a specific period of time, such as a year. These are scored against a rubric that describes performance at several levels of achievement, and may be scored by more than one person to ensure reliability of the score. Teachers select entries based on guidelines provided to them, such as a certain number of entries in each targeted content area, some entries that show the results of participating in the general education classroom, or entries that show use of technology. The criteria in the rubrics vary considerably.

Alternate assessments based on alternate achievement standards that are performance based generally involve specific tasks to which the student is to respond. The student's response to the task is judged against a rubric to produce a score. Independent observers or teachers may score each student's performance. In some cases, records of performance are developed and these are sent to a scoring center to produce scores for each student's performance.

Rating scales provide the teacher or other raters with specific criteria to judge the past performance of students. In some states or districts, raters are directed to evaluate the performance of students during a specific window of time, such as 2 weeks before other students are taking the general assessment, and in some states they are directed to rate the change in performance from the beginning to the end of the school year.

Alternate assessments based on modified achievement standards reflect a new approach to including special education students in large-scale assessment systems. Some educators refer to these assessments as 2% assessments, due to the percentage of students in the total population who can be counted as proficient on these assessments for the accountability requirements of the Elementary and Secondary Education Act. Alternate assessments based on modified achievement standards look more like general assessments than not, and are intended for students with disabilities who are unable to meet proficiency on the state assessment within the period of a student's IEP.

Accountability

Accountability refers to holding a person or group of people responsible for a process or product. When

coupled with large-scale assessment, accountability generally indicates that schools and districts, or the students and personnel within them, are held responsible for performance on assessments. Accountability systems have goals such as improved performance, ways to measure the goals such as large-scale assessments, criteria for determining when goals have been met, and consequences for meeting or not meeting the goal. When the consequences become significant, they are referred to as high stakes. High-stakes assessments can be either those where the responsibility lies with the school system, or where the responsibility lies with the individual educator or student.

The inclusion of special needs students in accountability systems creates challenges for several reasons. One reason is that there is increased concern about the potential unintended consequences that may follow when students with disabilities or ELLs are included in accountability systems. This concern exists whether the school system or educators within it are held responsible for the performance of the students, or the students themselves are held responsible through promotion or exit exams. Unintended consequences include students dropping out of school, being exposed to inappropriate instructional efforts to produce increased test scores, and losing their special education status due to the level of their performance. There are also concerns about how to best include students who may be in very different educational settings such as residential schools and how to account for students who are with some teachers for very small amounts of time or who move in and out of special education services.

System accountability

System accountability refers to holding schools, districts, and states responsible for the performance of students within them. Large-scale assessments are an integral part of most system-accountability measures, but not the only measures. Others include indicators of attendance, graduation rates, and the extent to which all students participate in large-scale assessments. In the United States, the Elementary and Secondary Education Act includes requirements for system accountability. It also requires the inclusion of all students in assessments and the accountability system.

The inclusion of students with disabilities in system accountability requires decisions to be made about how their participation in large-scale assessments will factor into accountability calculations. Students with significant cognitive disabilities often are identified as one group that does not fit the traditional assessment approach because their disabilities are intellectual and multiple in nature. If these students were to be held to the same performance standards as other students, they would always be at the bottom of the performance continuum, simply because of the nature of their disability. Determining that these

students can have their own set of performance standards based on unique assessment approaches is a solution to the challenge of how to include them.

Another challenge occurs when considering the inclusion of students who use accommodations or modifications during testing. If accommodations are considered to not change the intent of the measurement and to produce valid scores, then the use of these is not considered to inappropriately change the student's performance. Their inclusion in accountability systems then is usually considered to be appropriate without challenges, other than those logistic challenges associated with making appropriate decisions about which students need which accommodations and making sure that students who are supposed to use accommodations during testing actually do receive them.

The challenge comes from determining how to include those students who use modifications. These are changes in testing materials or procedures that alter the intent of the measurement and therefore produce invalid scores. Determining ways to represent these students in accountability systems, while not inaccurately indicating that students have valid scores, is a major challenge when including students with disabilities in accountability systems. A common approach for the inclusion of those students who use modifications is to count them as participating in the assessment, but instead of reflecting the scores that they obtained the students are assigned the lowest possible score or a zero. The Elementary and Secondary Education Act has determined, however, that for purposes of its accountability system, students using modifications cannot be counted as participants in the assessment because they do not receive valid scores. Adequate numbers of students participating in the assessment is an additional factor in the Elementary and Secondary Education Act accountability system.

Different accountability models

System accountability may take different forms depending on the model that is used. Common models are status models, cohort models, and longitudinal models. Sometimes cohort and longitudinal models are referred to as growth models. In status models, the performance of students in one grade in a year is compared to the performance of students in the same grade the next year. Status accountability measures are based on different groups of students, but the system is held accountable for showing improvement over time.

Cohort-accountability measures are based on primarily the same group of students. The performance of students in one grade in a year is compared to the performance of students in the next grade in the next year. This model results in primarily the same set of students being compared from one year to the next. Some students move out of the system, and some students move in, so there is actually some variation in the specific students included in the

comparisons. To the extent that mobility or dropout rates are higher, there is decreased consistency in the students who are included in the cohort-accountability system.

Longitudinal accountability is based on the same specific students being compared from one year to the next. This approach is highly dependent on good data management so that students can be followed across years. It is more reflective of the student population when mobility rates are small.

Growth models are a variation of these models of accountability. They use either a cohort or longitudinal approach. Sometimes they factor in the effects of students' socioeconomic, ethnic, or disability status. Value-added approaches attempt to demonstrate the added effect of the education system or teachers within it to where the student started out, including adjustments for the student's poverty status or disability status. Value-added models often are controversial when consideration is given to the inclusion of students with disabilities in them. Concerns are raised about whether these students are held to lower expectations as a result and whether they are excluded too often from these systems because of other factors in the calculations that determine value added.

Educator accountability

Educator accountability refers to holding administrators, teachers, or other school personnel responsible for the performance of students, and assigning consequences to them based on the performance of their students. Pay-for-performance systems financially reward educators for the performance of their students. Students with disabilities typically are included in calculating student performance, and special educators are also the target of the accountability systems. The inclusion of students with disabilities often causes challenges for the accountability system. Accountability system developers must determine ways to include students who may spend only part of their time or none of their time in the general education classroom. In some cases, students may receive special education services in a residential or alternative setting. All of these unique situations create challenges for the inclusion of the performance of students with disabilities in an educational accountability system. Including special educators in the accountability systems is also challenging. Unlike general education teachers, special educators may work with many more students for much shorter periods of time.

Student accountability

Student accountability refers to holding students responsible for their own performance. Promotion exams and exit or graduation exams are an integral part of most student-accountability measures. Additional measures are typically added to student-accountability decisions. For promotion exams, consideration may also be given to the amount of progress the student made during the

school year or the participation of the student in a remedial program. For exit or graduation exams, completion of course work is often another factor in determining whether the student is able to graduate.

The inclusion of students with disabilities in student-accountability measures is complicated by a history of excluding them and holding them to different standards. Regardless of whether excluded or held to different standards, the same results such as promotion to the next grade or graduation with a standard diploma often were allowed for these students. Increasingly, however, different results, such as receipt of a certificate of attendance or a special diploma rather than a standard diploma, are being imposed for those students with specific needs who do not meet the same requirements as their peers.

For students with disabilities, there may be alternative routes to the standard diploma that are not available to other students. For example, they may be allowed to demonstrate that they have met graduation performance standards by providing a portfolio of their work when they repeatedly fail the graduation exam, or they may be able to show that their classroom performance and grades are equal to those of their peers who have passed the graduation exam. A variety of approaches have been developed in some places to allow students with disabilities the opportunity to show that they have met the same student-accountability requirements as their peers. Other places have no alternative ways for students with special needs to demonstrate that they have met the requirements even though they have not been able to pass the large-scale assessment.

Summary

Large-scale assessments used for accountability purposes are a prevalent part of the educational system in states and schools in the United States. They vary in appearance and purpose. A commonality across them is that they include students with disabilities. This inclusion has required the careful consideration of how all students with special needs can participate in the assessments so that the assessment results are valid reflections of the students' knowledge and skills.

The inclusion of students with disabilities requires test developers and test users to think carefully about the nature of the assessment, its purpose, and the kinds of flexibility that can be introduced without changing what the original assessment intends to measure. When students with disabilities participate with accommodations or in an alternate assessment, careful consideration must be given to the reasons that the student is not able to participate in the general assessment in the same way as students without disabilities. When participation in the assessment other than the general assessment without

accommodations is the best option, the assessment should produce valid results for students with disabilities so that they can be combined with the performance of other students.

Further Reading

- Downing, S. M. and Haladyna, T. M. (eds.) (2006). *Handbook of Test Development*. Mahwah, NJ: Erlbaum.
- Elmore, R. and Rothman, R. (1999). *Testing, Teaching, and Learning*. Washington, DC: National Academy Press.
- Fuhrman, S. H. and Elmore, R. F. (eds.) (2006). *Redesigning Accountability Systems for Education*. New York: College Press.
- Haladyna, T. M. and Downing, S. M. (2004). Construct irrelevant variance in high stakes testing. *Educational Measurement: Issues and Practice* 23(1), 17–27.
- Heubert, J. and Hauser, R. (1999). *High Stakes: Testing for Tracking, Promotion, and Graduation*. Washington, DC: National Academy Press.
- Johnstone, C. J., Thompson, S. J., Bottsford-Miller, N. A., and Thurlow, M. L. (2008). Universal design and multimethod approaches to item review. *Educational Measurement: Issues and Practice* 27(1), 25–36.
- Kleinert, H. L. and Kearns, J. L. (eds.) (2001). *Alternate Assessment: Measuring Outcomes and Supports for Students with Disabilities*. Baltimore, MD: Brookes.
- Kleinert, H. L. and Thurlow, M. L. (2001). An introduction to alternate assessment. In Kleinert, H. L. and Kearns, J. L. (eds.) *Alternate Assessment: Measuring Outcomes and Supports for Students with Disabilities*, pp 1–15. Baltimore, MD: Brookes.
- Koenig, J. A. and Bachman, L. F. (2004). *Keeping Score for All: The Effects of Inclusion and Accommodation Policies on Large-scale Educational Assessments*. Washington, DC: National Academies Press.
- Krentz, J., Thurlow, M., Shyyan, V., and Scott, D. (2005). *Alternative Routes to the Standard Diploma*, Synthesis Report 54. Minneapolis, MN: University of Minnesota/National Center on Educational Outcomes.
- Lazarus, S. S., Thurlow, M. L., Lail, K. E., and Christensen, L. (2008). A longitudinal analysis of state accommodations policies: Twelve years of change 1993–2005. *Journal of Special Education*. doi:10.1177/0022466907313524.
- Linn, R. L., Baker, E. L., and Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher* 31, 3–16.
- McDonnell, L. M. and McLaughlin, M. J. (eds.) (1997). *Educating One and All: Students with Disabilities and Standards-Based Reform*. Washington, DC: National Academy Press.
- National Academies Press (2007). *Lessons learned about testing: Ten years of work at the National Research Council*. Washington, DC: Author.
- Rose, D. H., Meyer, A., and Hitchcock, C. (eds.) (2005). *The Universally Designed Classroom: Accessible Curriculum and Digital Technologies*. Cambridge, MA: Harvard Education Press.
- Salvia, J. E., Ysseldyke, J. E., and Bolt, S. (2007). *Assessment: In Special and Inclusive Education*, 9th edn. Boston, MA: Houghton Mifflin.
- Sireci, S. G., Scarpatti, S. E., and Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research* 75(4), 457–490.
- Thompson, S. J., Johnstone, C. J., and Thurlow, M. L. (2002). *Universal Design Applied to Large Scale Assessments*, Synthesis Report 44. Minneapolis, MN: University of Minnesota/National Center on Educational Outcomes.
- Thurlow, M. L. (2006). Biting the bullet: Including special-needs students in accountability systems. In Fuhrman, S. H. and Elmore, R. F. (eds.) *Redesigning Accountability Systems for Education*, pp 115–137. New York: College Press.

- Thurlow, M. L. (2007). State policies and accommodations: Issues and implications. In Laitusis, C. C. and Cook, L. L. (eds.) *Large-scale Assessment and Accommodations: What Works?*, pp 13–22. Arlington, VA: Council for Exceptional Children.
- Thurlow, M., Johnstone, C., Thompson, S., and Case, B. (2008). Using universal design research and perspectives to increase the validity of scores on large-scale assessments. In Johnson, R. and Mitchell, R. (eds.) *Assessing Deaf Students' Academic Achievement in an Age of Accountability*, pp 63–75. Washington, DC: Gallaudet University Press.
- Thurlow, M., Thompson, S., and Johnstone, C. (2007). Policy, legal, and implementation issues surrounding assessment accommodations for students with disabilities. In Florian, L. (ed.) *The Sage Handbook of Special Education*, pp 331–346. Thousand Oaks, CA: Sage.
- Thurlow, M. L., Thompson, S. J., and Lazarus, S. S. (2005). Considerations for the administration of tests to special needs students: Accommodations, modifications, and more. In Downing, S. M. and Haladyna, T. M. (eds.) *Handbook of Test Development*, pp 653–673. Mahwah, NJ: Erlbaum.
- Thurlow, M. L., Elliott, J. E., and Ysseldyke, J. E. (2003). *Testing Students with Disabilities: Practical Strategies for Complying with District and State Requirements*, 2nd edn. Thousand Oaks, CA: Corwin.
- Zabala, D., Minnici, A., McMurrer, J., et al. (2007). *State High School Exit Exams: Working to Raise Test Scores*. Washington, DC: Center on Education Policy.

Relevant Websites

- <http://www.cep-dc.org> – Center on Education Policy.
- <http://www.ccsso.org> – Council of Chief State School Officers, Assessing Special Education Students State Collaborative.
- <http://www.narap.info> – National Accessible Reading Assessment Projects.
- <http://www.nasdse.org> – National Association of State Directors of Special Education.
- <http://www.cse.ucla.edu> – National Center for Research on Evaluation, Standards, and Student Testing.
- <http://www.nceo.info> – National Center on Educational Outcomes.

Literacy Instruction for Students with Special Needs

L H Mason, The Pennsylvania State University, University Park, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Children with difficulties in reading, vocabulary, and writing constitute the largest group of students identified for special education services and are most frequently included in the general education curriculum with minimal supports. A large proportion of students with special needs, perhaps over 70%, are at risk for academic failure in the general education curriculum due to a lack of foundational skills in literacy. When their achievement is compared to that of other struggling learners, students with special needs may appear similar; however, they experience reading, vocabulary, and writing problems that are more severe than those of their peers.

The National Reading Panel (NRP) report in 2000, fortunately, concluded that students' performance in reading could be improved by (1) promoting phonemic awareness that teaches students to break apart and manipulate the sounds in words; (2) promoting phonics that teaches students that sounds are represented by letters which can be blended together to make words; (3) having students practice what they have learned with guidance and feedback; and (4) teaching reading comprehension strategies (Carnine *et al.*, 2004). Correspondingly, students' vocabularies can be enhanced by directly and indirectly teaching strategies for learning word meanings. Writing performance, furthermore, can also be improved with explicit instruction in teaching students with special needs to be strategic while planning, drafting, and revising (Graham and Harris, 2003).

Instructional Approaches for Literacy

Students with special needs may have a number of problems that interfere with their learning by means of conventional general education instruction. Students with special needs, for example, often have difficulties with memory and poor language abilities. They may have interfering behaviors such as attention problems, hyperactivity, aggressive behavior, withdrawn behavior, or unusual behavior (Bos and Vaughn, 2006). Effective instruction for these students, therefore, should include careful progress monitoring; instructional planning that considers flexible groupings, adaptation, and scaffolding; and instruction that is paced well and allows for student responding and error correction. Two instructional approaches that meet this criterion, direct instruction (DI) and cognitive strategy instruction (CSI), are highlighted in this synopsis of literacy instruction. Although these approaches, often implemented to address the

diverse array of reading, vocabulary, and writing competencies, are by no means the only documented effective approaches for teaching literacy to students with special needs, both DI and CSI have been used broadly in special education and have proven effective in improving literacy outcomes for these students.

Principles for both DI and CSI are presented in **Table 1**. Although these approaches appear to be quite different, both contain congruent principles implemented in varying degrees, depending on the skill to be taught. For example, teacher-led modeling is a critical component of both DI and CSI. Instruction is criterion-based as opposed to time based. In other words, students meet a pre-established goal for learning prior to moving to the next instructional stage or lesson. Student motivation is supported in both approaches through student success as noted through demonstration and/or self-reinforcement. Although neither approach is limited in scope of skills to be taught, DI has been generally established in research as most effective for developing fluency in basic skills such as word recognition and spelling, while CSI has been established in research as most effective with higher-level skills such as reading comprehension and written expression. DI and CSI, and variations of the two approaches, will be described further in the context of specific reading, vocabulary, or writing skills to be taught. In addition to DI and CSI models, examples of other models of instruction, taken from the research literature, will be employed to illustrate the range of effective literacy practices for students with special needs.

Reading

According to the NRP (2000) report, effective reading instruction includes phonemic awareness instruction, phonics instruction, fluency instruction, vocabulary instruction, and text comprehension instruction. There is little disagreement among researchers that these components are essential for students with special needs who are struggling to learn to read (Carnine *et al.*, 2004). Although researchers and practitioners may disagree on how to teach reading, for these low-achieving students, engagement in appropriate and meaningful skill-level reading-related activities is generally deemed crucial.

Beginning reading

Although many students come to school prepared to read, students with special needs often come to school needing

Table 1 Effective instructional principles for students with special needs

<i>Direct instruction (DI)</i>	<i>Cognitive strategy instruction (CSI)</i>
<i>Unison oral responding.</i> All students in a group respond at the same time to a teacher's question or prompt.	<i>Preskill development.</i> Preskills for the task are developed. Students' prior knowledge about the task and strategy are assessed and remediation for individual students is provided as needed.
<i>Signaling.</i> Cues that are developed by the teacher to cue responses from students.	<i>Discussion.</i> The teacher and students discuss the strategy to be learned. A purpose for using the strategy is established and the benefits of using the strategy for the task, is presented.
<i>Pacing.</i> The teacher effectively paces instruction to avoid student downtime during instruction.	<i>Modeling.</i> The teacher cognitively models (models while thinking out loud) how to use and apply the strategy for the task.
<i>Monitoring.</i> Teacher systematically plans the use of elicited responses from individual students to monitor learning.	<i>Memorization.</i> Students are provided time to memorize the strategy steps until fluent in understanding any mnemonics and meanings and each strategy step.
<i>Error correction and teaching to mastery.</i> Teachers include five instructional steps: model, lead, test, firm up, and delayed test.	<i>Guided practice.</i> Instruction is scaffolded from student–teacher collaborative practice to independence based on individual student needs. Systematic feedback is given to individual students.
<i>Diagnosing.</i> Teacher works with individual students to determine areas of weaknesses.	<i>Independent practice.</i> The teacher provides independent practice across task and settings to foster generalization and maintenance.
<i>Motivation.</i> Student motivation is fostered through demonstration of student success and rewards for learning.	^a <i>Self-regulation.</i> Student takes control of learning through goal setting, self-instruction, self-evaluation, and self-reinforcement.

^aIncluded in self-regulated strategy development instruction (SRSD).

6 months to 1 year of instruction prior to beginning reading. These students will require intensive instruction in:

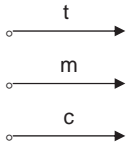
1. Phonological awareness: the identification and manipulation of spoken language (e.g., words, syllables, onsets, and rimes). This includes phonemic awareness, the identification of manipulation of sounds (phonemes) in words.
2. Letter–sound correspondence: knowledge of letters of the alphabet and the corresponding sounds they represent.
3. The alphabetic principle: the combination of phonological awareness and letter–sound correspondence.

The NRP (2000) report noted that phonological awareness is best supported by instruction that uses alphabet letters for phoneme manipulation; limits manipulation to one or two phonemes; includes phonemes in words for blending and segmenting activities; and provides explicit description about the application of phonemic skills to both reading and writing (i.e., spelling). Phonological awareness can include skills such as identification of rhyming words, listening for words with same beginning sound, identifying the number of words by clapping, and blending and segmenting words by sounds and syllables (Torgesen, 1999). Carnine *et al.* (2004) recommended DI for teaching telescoping (blending), segmenting, and rhyming skills. Telescoping and segmenting tasks teaches phonemic awareness, and provides students with practice in saying sounds and blending sounds without pausing. Rhyming assists students in understanding the relation of same-sound letter clusters and provides practice in sounding out words that begin with stop sounds (e.g., the p in pop as opposed to the r in rat). Rhyming is introduced after students have mastered telescoping and segmenting skills. **Table 2** illustrates how DI is used to teach these beginning reading skills. While a developmental model is generally accepted for teaching skills in each of these areas, it is not recommended that instruction focus on one skill at the expense of others. Simply stated, some of these students may never master a skill. Rhyming, for example, may be quite difficult for some students with special needs, whereas a higher-level skill such as syllable blending may be learned easily.

Given that phonological awareness instruction supports auditory skill development as opposed to specific skills for reading letters and words, it should be taught in conjunction with letter–sound correspondence, or instruction known as the alphabetic principle, through phonic instruction. Phonic instruction focuses on teaching students to recognize familiar words and to decode unfamiliar words. It includes teaching letter–sound associations, letter–sound correspondences, sound–spelling, written language (grapheme)–phoneme correspondences, and graphophonemic relationships. Carnine *et al.* (2004) identified six approaches to phonic instruction:

1. *Synthetic phonics.* Students convert letters into sounds, and then blend the sounds to form words.
2. *Analytic phonics.* Letter–sound relationships in learned words are analyzed; sounds are not pronounced in isolation.
3. *Analogy-based phonics.* Word families are used to identify unknown words.
4. *Phonics through spelling.* Words are segmented into phonemes; letters are written for phonemes.
5. *Embedded phonics.* Letter–sound relationships are taught during reading of connected text. (Note: this is not considered an explicit instructional approach.)
6. *Onset–rime phonics.* Common spelling patterns are used to decode words.

Table 2 Direct instruction for telescoping, segmenting, and rhyming sounds

<i>Teacher (teacher says words in quotations)</i>	<i>Unison choral responding</i>
<i>Telescoping^a</i>	
1. "Listen, we are going to play a say-the-word game. I will say the word slow. You say the word fast"(pacing).	"Go."
2. "Listen – ggggoooo-What word?" Signal (signaling).	
3. Repeat step 2 with four more words.	
4. Repeat the set of words until students can respond with no errors (error correction and teaching to mastery; motivation).	
5. Provide individual turns for students (monitoring and diagnosing).	
<i>Segmenting</i>	
1. "We are going to say words slowly."	"Ggggoooo."
2. "First word: go. I will say it slowly. Listen. Ggggoooo. You say it slowly. Get ready." Signal.	
3. Repeat procedure in step 2 with three more words.	
4. Repeat the set of words until students can respond with no errors.	
5. Provide individual turns for students.	
<i>Rhyming</i> (Write on board)	
	
1. "Listen. I am going to rhyme with – an. What am I going to rhyme with?" Signal.	
2. Model by putting finger on the dot of the first arrow and say, "My turn, rhymes with an." Pause for 1 s, move finger across arrow and say, "tan."	"An."
3. Repeat step 2 with remaining arrows.	
4. Put finger on dot of first arrow. Say, "You are going to rhyme with an. What are you going to rhyme with?" Signal. "Rhymes with an." Pause for 1 s, move finger across arrow.	"An." "Tan."
5. Repeat step 4 with remaining arrows.	
6. Provide individual turns for students.	

^aDI principles in parenthesis.

Adapted from Carmine, D. W., Silbert, J., Kame'enui, E. J., and Tarver, S. G. (2004). *Direct Instruction Reading*, 4th edn. Columbus: Merrill Prentice Hall.

Several modifications to these instructional approaches are recommended for students with special needs. First, when teaching letter–sound correspondence, a core set of frequently used consonants and short vowel sounds should be taught with consistent use of key words. Students should

learn that some letters have more than one sound, different letters can make the same sound, and more than one letter can represent some sounds. Kinesthetic (trace letters as saying sound) and visual (color-coding) approaches can be used to reinforce learning. In addition to these recommendations, explicit code instruction (instruction that is systematic and scaffolded to include modeling, guided practice, corrective feedback, and review) is recommended. Several methods have been recommended by researchers and are used widely for phonic instruction for students with special needs: (1) the linguistic approach for onset–rime and word families (e.g., *Merrill Reading Program*); (2) DI approaches (e.g., *Reading Mastery*); and (3) multisensory structured language instruction (e.g., *Wilson Reading System*) (see the section titled 'Further reading').

Fluency

Fluent readers read text quickly, accurately, with ease, and with expression. Fluency is a critical skill that is required to move the student from an over-focus on word recognition to attending to understanding what is being read. Several methods for improving reading fluency for students with special needs have been identified (Bos and Vaughn, 2006). First, fluency (e.g., number of words correctly read in 1 min minus number of errors) should be assessed and monitored. Students can plot their scores on a graph for goal setting and to support motivation. Next, teachers should model fluent reading by reading aloud to students, teaching them to preview books prior to reading, and providing practice in reading at the students' instructional to independent work-recognition level. Clearly, fluency can also be supported in the classroom by providing students opportunities for reading performance. In addition, it can be supported, for example, by asking older students to read to younger students. Teachers can model the acceptability of reading an easy book and make a difficult book more accessible by tape recording, providing partner reading, and using series books.

Repeated reading as an instructional approach for fluency has proven effective for improving reading speed, accuracy, expression, and comprehension. In repeated reading, the student reads a 50–200-word passage at their instructional level 3–5 times until a more fluent rate is achieved. The teacher provides direct support by pronouncing any unknown words while the student reads. This procedure has been modified effectively with the use of choral repeated reading, taped books, or computer-assisted instruction. Direct instruction for repeated reading should occur in a couple of short (10–15-min) sessions weekly.

Another instructional approach, peer-supported reading (i.e., classwide peer tutoring and partner reading), supports reading fluency and comprehension by providing students with special needs, multiple opportunities for reading in the general education classroom. One elementary school

Table 3 Peer-assisted learning strategies (PALSs) for reading fluency and comprehension*Features of PALS*

1. High-performing students are paired with low-performing students.
2. Students are trained in specific prompting, correction, and feedback procedures.
3. Verbal interactions are frequent to support student's opportunities to respond.
4. Student roles are reciprocal; each serves as a tutor and tutee, but the stronger reader always reads first.
5. Activities are structured.
6. Students are trained to implement activities independently.
7. Following training, PALS instruction is conducted weekly in three 35-min sessions.
8. Student pairs are assigned to teams. Points are earned by each pair by correctly responding during the activities.

PALS activities

1. Partner reading with retell
Each student reads a passage aloud for 5 min. If an error is made the tutor (the student who is listening) asks for correction. After both students have read, the weaker student states what was read in a 2-min retell. Students earn one point for each correctly read sentence, retelling earns 10 points.
2. Paragraph shrinking
During this activity students read orally, stopping at the end of each paragraph to identify the main idea. The tutor prompts the reader by asking the reader to identify who or what is most important. The reader must supply an answer in 10 words or less to earn points.
3. Prediction relay
Students complete four steps in this activity:
 - the reader predicts what will happen in next half of page
 - the remaining half of the page is read out loud
 - the reader confirms or disconfirms their prediction
 - the reader summarizes the main idea in the passage
 Students earn points for identifying the main idea in 10 words or less.

From McMaster, K. L., Fuchs, D., and Fuchs, L. S. (2006). Research on peer-assisted learning strategies: The promise and limitations of peer-mediated instruction. *Reading and Writing Quarterly*, 22, 5–25.

peer-supported method that includes a reading comprehension component, peer-assisted learning strategies (PALSs), is described in **Table 3** (McMaster *et al.*, 2006). A version of PALS, K-PALS instruction for kindergarten and first-grade students, supports young students by including more teacher support through extended coaching, and adds activities for a sound play (five games for phonological awareness) and sound and words (four activities for reading words and sentences). Additionally, PALS has been modified for high school students by changing partners more frequently than is done with younger students, by adding a system for managing and spending PALS dollars for points earned, and by focusing on reading with expository text.

Reading comprehension

Good readers maintain conscious control of their reading process by implementing strategies to support

comprehension as they deconstruct text (decode) and reconstruct meaning. Students with special needs in reading, in contrast to good readers, rely on simpler, less-efficient strategies and fail to implement what strategies they do know in a fluid, controlled fashion. Researchers, however, have clearly established that when students who struggle with reading, including students with special needs, receive explicit instruction combined with practice in comprehension strategies, reading comprehension gains are among the highest (Mastropieri and Scruggs, 1997). These strategies include reading comprehension approaches such as prior knowledge acquisition, self-questioning, main idea prompts, story grammar and text structures, summarization, and comprehension monitoring.

For students with special needs, however, successful reading comprehension is dependent on multiple elements. For these students, gains in improving reading comprehension have been more consistent and significant when multiple strategies are taught. Story grammar, for example, can be directly taught through a self-questioning strategy (Carnine *et al.*, 2004):

1. What is the story about?
2. What is she or he trying to do?
3. What happens when she or he tries to do it?
4. What happens in the end?

As students begin to read more complex expository text, as is often found in science and social studies textbooks, multifaceted approaches are best for supporting reading comprehension. Researchers (e.g., Mastropieri and Scruggs, 1997) have noted that multifaceted approaches (i.e., reciprocal teaching; collaborative strategic reading; predict, organize, search, summarize, evaluate (POSSE); and think before reading, think while reading, think after reading (TWA)) that include procedural facilitators for learning (i.e., questions, prompts, or simple outline of learning structures) with cognitive strategies to support student learning are effective in improving students' acquisition and facilitation of linking new knowledge to prior knowledge. These approaches present strategies for prior knowledge acquisition, comprehension monitoring, and summarization of information throughout reading (Baker *et al.*, 2002).

The TWA expository reading comprehension strategy, for illustration, is one approach that integrates multiple strategies to support active text engagement throughout the reading process. Individual strategies imbedded in TWA are described in **Table 4**.

In the first component of TWA, think before reading, three steps for prior knowledge acquisition are taught. Students examine text structure while completing the first step. They then develop statements and questions for what you know and think about what you want to learn. Three steps are taught in the think while reading component: reading speed, rereading parts, and linking knowledge. Students are taught to (1) monitor their

Table 4 TWA strategy

<i>TWA strategy step</i>	<i>Strategy base</i>	<i>Sources^a</i>
Think before reading: Authors' purpose	Expository text structure (description, sequential, compare/contrast, cause/effect, problem/solution)	Englert and Thomas (1984)
What I know	K-W-L What I know, what I want to learn, what I learned	Ogle (1989)
What I want to know	K-W-L	
While reading think about: Reading speed Rereading parts	Comprehension monitoring and rereading parts	Graves and Levin (1989)
Linking knowledge	Discuss students' previous experience, connect this experience with the text, and share ideas for expanding knowledge	Hansen and Pearson (1983)
After reading think about: Main idea	RAP strategy (1) read the paragraph, (2) find the sentence in the paragraph that tells the main idea of the paragraph, and (3) put the main idea into your own words	Ellis and Graves (1990)
Summarize	Five steps for summarization (delete trivial information, delete redundant information, substitute super-ordinate terms for a list of terms or actions, select a topic sentence, and invent a topic sentence)	Brown and Day (1983)
What you learned	K-W-L	

^aSee the section titled 'Further reading'.

reading speed, (2) monitor their understanding by rereading when something is not understood, and (3) make connections between their prior knowledge and the current text. Think after reading has two strategic steps developed at the paragraph level (main idea and summarizing information), followed by the final step, an oral retelling of the complete passage (what you have learned). To reinforce locating the main idea in each paragraph, students use yellow highlighters (student use of these is faded as students master the strategies) to mark sentences and phrases related to the main idea in each paragraph. Five rules are used for developing paragraph summaries; students highlight the important details in blue and trivial details in pink for

each paragraph. Once students have developed main ideas and summaries, they orally retell what had been read and learned in the passage. TWA is taught using the principles of CSI within the self-regulated strategy development (SRSD) instructional model (see Table 1). Instruction for TWA is carefully scaffolded from teacher–student collaborative practice to student–student collaborative practice to student independent practice.

Vocabulary

A student's understanding of word meanings, or their vocabulary, is often considered the glue that binds word recognition and comprehension. For students with special needs, gaps in vocabulary development cause reading delay and affect reading growth (Bos and Vaughn, 2006). There are two types of vocabulary instruction, oral vocabulary and reading vocabulary. Instruction in reading vocabulary (the words a reader recognizes in print or uses in writing) is the focus of this synopsis. Although teacher-directed instructional approaches such as oral language instruction (i.e., teaching vocabulary during reading by modeling how to use context; using synonyms and definitions) work well for average achieving students, students with special needs often require additional explicit instructional approaches.

DI for word-learning strategies, morphemic and contextual analysis, is an approach often used to support student independence in developing vocabulary. In morphemic analysis, students are taught to use word parts to interpret word meanings. For example, the teacher introduces and teaches the meaning of a morpheme (e.g., Pre usually means before, what does pre mean?). The teacher then teaches new words that include the morpheme through signaling and unison oral responding (e.g., So what does pretest mean?). Carnine *et al.* (2004) cautioned that this approach is limited by the difficulties in translating some words into functional definitions, the fact that many words have dual meanings, and the difficulty in selecting appropriate morphemes for instruction. Contextual analysis, in contrast, fosters students' independence in learning word meanings by using surrounding words in text. In contextual analysis instruction, the teacher points out the unknown word in the text, asks the students to find the words that tell what the word means, and prompts the student to restate the sentence substituting the known word for the unknown word. Students are taught to look in text for an imbedded word-definition, a synonym, a description, a contrast, or a comparison.

Mnemonics with modified CSI has been used effectively for teaching vocabulary and associated concepts to students with special needs (Hughes, 1996). The keyword picture strategy, for example, uses visualization in the following steps taught to students: (1) select a word or term; (2) state the definition; (3) select another similar

sounding word; (4) create a picture using the definition, the word to be learned, and the key word; and (5) think about and study the picture. When developing mnemonic strategies for vocabularies and concepts, Mastropieri and Scruggs (1989) recommended that the amount of elaboration required is dependent on the students' familiarity with both the word and the concept abstraction.

Writing

Good writers spend time in planning, revising, monitoring, evaluating, and regulating the writing process. Students with special needs, on the other hand, spend little time in these processes and tend to focus on low-level transcription skills (i.e., handwriting, spelling, capitalization, and punctuation). A process approach to writing (prewriting – planning and organizing, drafting, revising, editing, and publishing) combined with CSI, therefore, is generally accepted as the best instructional approach to foster students' performance in higher-level skills for written expression. For students with special needs, Graham and Harris (2003) recommended that teachers:

1. provide time for writing, a minimum of 30 min daily;
2. provide a range of writing tasks, starting with what the student knows best;
3. create a social climate that promotes writing through peer and teacher conferencing;
4. provide opportunities for writing in all academic content classes;
5. focus on the writing process;
6. focus on high-order tasks, attending to mechanics during the revision process;
7. explicitly teach the characteristics of good writing, providing examples of and modeling good writing;
8. promote higher-level composing by teaching skills for strategy use;
9. teach students to identify and set goals for their own writing; and
10. minimize grammar instruction (i.e., sentence diagramming) and overemphasis on students' errors.

Written expression

One approach for teaching written expression to students with special needs, SRSD instruction, blends explicit instruction in self-regulation procedures with CSI (see **Table 1**). Theory in students' cognitive development and learning, student behavior, and the role of affect are all utilized in SRSD (Graham and Harris, 2003).

SRSD instruction has been used effectively for teaching planning and revision strategies across narrative, persuasive, and informative genres (see **Table 5** for planning strategy examples). As illustrated in the following sample lesson sequence for teaching the 'pick my idea, organize my notes, write and say more' (POW) + 'topic sentence, reasons, explain reasons, ending' (TREE) persuasive writing

Table 5 Strategies for written expression

Narrative writing	POW + W-W-W, What = 2, How = 2 Pick my idea. Organize my notes – use W-W-W, What = 2, How = 2! Write and say more. + Who is the main character? Where does the story take place? When does the story take place? What does the main character do or want to do? What happens next? How does the story end? How do the characters feel?
Persuasive writing	POW + TREE Pick my idea. Organize my notes – use TREE! Write and say more. + Topic sentence Tell what you believe! Reasons – 3 or more Why do I believe this? Will my readers believe this? Explain reasons Say more about each reason. Ending Wrap it up right!
Informative writing	Plan and write ^a P = Pay attention to the prompt L = List main ideas to develop your essay A = Add supporting ideas (details, examples, etc.) N = Number major points in the order you will use them and W = Work from your plan to develop a thesis statement R = Remember your goals I = Include transition words for each paragraph T = Try to use different kinds of sentences E = Exciting, interesting, \$1 000 000 words

^aDeveloped by S. De La Paz. See Harris *et al.* (2008) for complete lesson plans and materials.

strategy, instruction is recursive rather than linear. In other words, lessons and instructional stages and principles are repeated and revisited based on individual student needs:

1. *Lesson one.* The purpose of the first lesson is to introduce POW + TREE and to discuss the steps for each strategy. The teacher leads a discussion of what good writers do when writing an essay. The strategic parts are introduced and each step is described. The teacher then guides students in locating essay parts in an essay example. The students and the teacher both sign a learning contract: the student setting a goal to learn and use POW + TREE, the instructor committing to

do his/her best in teaching the strategy. In this first lesson and all subsequent lessons, each student is provided time to practice memorizing the POW and TREE steps; then students are told that they will be asked to share the steps from memory in the next session.

2. *Lesson two.* This lesson begins with the initial activity of all lessons: asking students to write out the POW + TREE and state the steps orally. If needed, the teacher asks students to locate essay parts in another persuasive essay. Students are asked to locate essay parts in one of their previously written essays. To reduce the possibility of an inappropriate reaction from the student, the teacher stresses, "Of course, the essay does not have all the parts; you did not know the POW + TREE strategy!" Students mark a graphing sheet to correspond with their essay parts. The teacher and students then discuss what could have been done to make their essay better.
3. *Lesson three.* Following memorization practice, the teacher cognitively models using the strategy by demonstrating how and when to use any support materials (e.g., graphic organizer). Following the teacher's modeling, students develop and record personal self-statements to be used while thinking of good ideas (relax, ask for help), while working (use POW or use TREE, I can do this), and to check work (examine, "perfect").
4. *Lesson four.* The teacher and students collaboratively write a persuasive essay, students provide ideas for the essay while the teacher writes the notes and essay. Support materials are used, self-instructions are reinforced, and the number of essay parts are examined and recorded on the graphing sheet.
5. *Lesson five.* Students write their own notes and essay with teacher guidance and supporting materials. If a student is having difficulty with any of the strategy processes, the teacher can return to modeling or collaborative practice.
6. *Guided practice.* The focus of all remaining lessons is to scaffold instruction until each student can independently implement all strategic steps without teacher or material support. Instructional materials, for example, are gradually replaced with student-written products (i.e., students write their planning outline on blank paper).
7. *Independent practice.* Once students can independently write a persuasive essay with all essay parts, the teacher can provide opportunities to support generalization and maintenance. The prompt can be revised to a different format, or the teacher can ask the student to write in a different academic class with a different teacher.

Transcription skills

Students with special needs generally have difficulty with both spelling and handwriting and these difficulties can interrupt composition processes, limit writing development, and mark a student as a poor writer (Graham, 1999). Students with special needs should be taught how to spell

words they use frequently in writing; generate reasonable spellings for unknown words; check and correct spelling errors; and develop a desire to spell correctly. A modified DI approach, using phonemic awareness and phonic rules, is often best for teaching spelling to students with special needs. DI for spelling includes mastery of daily words (no more than three a day or 12 a week), individualized instruction, and continual review. In addition, word-study approaches have proven effective for students with special needs. For example, the five-step word-study strategy (say the word; write and say the word, check the word; trace and say the word; write the word from memory and check; repeat the five steps) can be effectively used within a test-study-test (test; provide word study for misspelled words; and retest) spelling approach (Graham, 1999). Maintenance and generalization should be supported by continual review of previously learned words and integration of spelling words with students' reading and writing.

The goal of handwriting instruction is to teach students an efficient way to write letters and words legibly and fluently. DI is recommended for teaching handwriting to students with disabilities. Effective handwriting instruction includes (1) individualization to meet a student's needs; (2) frequently taught short lessons; (3) explicit modeling; (4) comparing and contrasting letter features; (5) facilitative supports such as marks for paper placement on student desks and tripod grip molds for pencils; (6) teaching students independence for evaluating and improving their handwriting; (7) handwriting fluency; and (8) mastery practice in isolation followed by application in context. In the end, spelling and handwriting instruction should be explicit and systematic. Graham (1999) suggested that teachers include less-formal methods of instruction (i.e., using teachable moments to model) to cultivate correct spelling and handwriting development.

Conclusion

Literacy instruction is best supported when reading, vocabulary, and writing are integrated. For example, students with special needs should be taught to read and spell words simultaneously. Word meanings should be taught throughout both the reading and writing processes. Written expression should be applied as a method for extending and fostering reading comprehension. By integrating these processes, students with special needs have multiple opportunities to practice and improve skills in literacy.

Bibliography

- Baker, S., Gertsen, R., and Scanlon, D. (2002). Procedural facilitators and cognitive strategies: Tools for unraveling the mysteries of comprehension and the writing process, and for providing meaningful access to the general curriculum. *Learning Disabilities Research and Practice* 17, 65-77.

- Bos, C. S. and Vaughn, S. (2006). *Strategies for Teaching Students with Learning and Behavior Problems*, 6th edn. Boston, MA: Pearson Allyn and Bacon.
- Camrine, D. W., Silbert, J., Kame'enui, E. J., and Tarver, S. G. (2004). *Direct Instruction Reading*, 4th edn. pp 53–35. Columbus, OH: Merrill Prentice Hall.
- Graham, S. (1999). Handwriting and spelling instruction for students with learning disabilities. A review. *Learning Disabilities Quarterly* **22**, 78–98.
- Graham, S. and Harris, K. R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In Swanson, H. L., Harris, K. R., and Graham, S. (eds.) *Handbook of Learning Disabilities*, pp 323–344. New York: Guilford Press.
- Hughes, C. A. (1996). Memory and test-taking strategies. In Deshler, D. D., Ellis, S., and Lenz, B. K. (eds.) *Teaching Adolescents with Learning Disabilities*, 2nd edn, pp 579–616. Denver, CO: Love.
- Mastropieri, M. A. and Scruggs, T. E. (1989). Constructing more meaningful relationship in the classroom: Mnemonics research into practice. *Learning Disabilities Research and Practice* **13**, 138–145.
- Mastropieri, M. A. and Scruggs, T. E. (1997). Best practices in reading comprehension in students with learning disabilities: 1976–1996. *Remedial and Special Education* **18**, 197–214.
- McMaster, K. L., Fuchs, D., and Fuchs, L. S. (2006). Research on peer-assisted learning strategies: The promise and limitations of peer-mediated instruction. *Reading and Writing Quarterly* **22**, 5–25.
- NRP(National Reading Panel) (2000). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. Bethesda, MD: National Institute of Child Health and Human Development, National Institutes of Health.
- Torgesen, J. K. (1999). Assessment and instruction for phonemic awareness and word recognition skills. In Catts, H. W. and Kamhi, A. (eds.) *Language and Reading Disabilities*, pp 128–153. Boston, MA: Allyn and Bacon.
- Brown, A. L. and Day, J. D. (1983). Macro rules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior* **22**, 1–14.
- Ellis, E. S. and Graves, A. W. (1990). Teaching rural students with learning disabilities: A paraphrasing strategy to increase comprehension of main ideas. *Rural Special Education Quarterly* **10**, 2–10.
- Engelmann, S. and Bruner, E. C. (1995). *Reading Mastery*. Columbus, OH: SRA/McGraw Hill.
- Englert, C. S. and Thomas, C. C. (1987). Sensitivity to text structure in reading and writing: A comparison between learning disabled and non-learning disabled students. *Learning Disabilities Quarterly* **10**, 93–105.
- Graham, S. and Harris, K. R. (2005). *Writing Better: Effective Strategies for Teaching Students with Learning Difficulties*. Baltimore, MD: Brookes.
- Graves, A. W. and Levin, J. R. (1989). Comparison of monitoring and mnemonic text-processing strategies in learning disabled students. *Learning Disabilities Quarterly* **12**, 232–236.
- Hansen, J. and Pearson, P. D. (1983). An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers. *Journal of Educational Psychology* **75**, 821–829.
- Harris, K. R., Graham, S., Mason, L. H., and Friedlander, B. (2008). *Powerful Writing Strategies for All Students*. Baltimore, MD: Brooks.
- Mason, L. H. (2004). Explicit self-regulated strategy development versus reciprocal questioning: Effects on expository reading comprehension among struggling readers. *Journal of Educational Psychology* **96**, 283–296.
- O'Connor, R., Notari-Syversen, A., and Vadasy, P. R. (1998). *Ladders to Literacy: A Kindergarten Activity Book*. Baltimore, MD: Brookes.
- Ogle, D. M. (1989). The know, want to know, learn strategy. In Muth, K. D. (ed.) *Children's Comprehension of Text*, pp 205–223. Newark, DE: International Reading Association.
- Swanson, H. L., Hoskyn, M., and Lee, C. (1999). *Interventions for Students with Learning Disabilities: A Meta-Analysis of Treatment Outcomes*. New York: Guilford.
- Torgesen, J. K. and Mathes, P. G. (2000). *A Basic Guide to Understanding, Assessing, and Teaching Phonological Awareness*. Austin, TX: Pro-Ed.
- Wilson, B. A. (1996). *Wilson Reading System*. Millbury, MA: Wilson Language Training Corporation.

Further Reading

- Bertin, P., Mercer, C. D., and Perlman, E. (1999). *Merrill Reading Program*. Columbus, OH: SRA/McGraw-Hill.

Math Instruction for Children with Special Needs

B A Bottge, University of Kentucky, Lexington, KY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Anchored instruction – Portrays or anchors problems in realistic contexts from which students identify embedded data that leads to a plausible solution.

Cognitive strategy instruction – Teaches students to monitor their own learning activities with methods such as with modeling, verbal rehearsal, and feedback.

Concrete-to-representational-to-abstract (CRA) sequence – Teaches students math concepts by representing them with concrete materials followed by pictures and drawings and finally by math numbers and operations.

Curriculum-based measurement – Is an example of formative assessment that measures student academic progress with frequent tests for the purpose of managing instruction.

Direct instruction – Refers to teaching number facts, operations, and problem solving in explicit and systematic ways.

Enhanced anchored instruction – Is a form of anchored instruction specifically designed for developing the problem-solving skills of adolescents with learning disabilities.

Graphic organizers – Use visual displays and graphic diagrams to help students develop meaningful representations of problem types.

Learning disability in math (MD) – Refers to an array of difficulties some students experience in computing and problem solving in math.

Metacognition – Refers to the ability to monitor and self-regulate one's own thinking.

Peer-mediated instruction – Includes a variety of interventions in which students without MD provide students with MD additional learning and practice opportunities.

Schema-based instruction – Helps to teach students the underlying structure of problems thereby helping them identify procedures for solving specific kinds of problems.

Self-regulated learning – Refers to helping students monitor and record their own activities related to academic and behavior.

Introduction

By most accounts, up to 8% of the school-age population has a disability in learning math. The difficulties experienced by students with math disabilities (denoted MD) center on two main areas: computation and problem solving. Within each area, many of these students exhibit other characteristics, such as confusion over sign operations and difficulty aligning numbers. To develop their computation skills, school instruction focuses primarily on number sense, automaticity with number combinations, and procedural fluency (i.e., adding, subtracting, multiplying, and dividing). Instruction on developing students' problem-solving skills is aimed primarily at comprehending the meaning of problems and helping students devise appropriate strategies for solving them. National assessments and research reports have highlighted the seriousness of the problem in both areas. For example, results from the large-scale testing efforts such as the National Assessment of Educational Progress revealed that approximately two-thirds of eighth-grade students with MD scored below the basic level, far below what is necessary for success in secondary education and job settings. Students scoring at the basic level should complete problems correctly using diagrams, charts, and graphs as well as understand informal algebraic and geometric concepts in problem solving. Thus, what is now called basic skills, as defined by contemporary standards, extend far beyond procedural competency and reflect the high levels of proficiency advocated by many math organizations such as the National Council of Teachers of Mathematics (NCTM).

In addition to point-in-time assessments, longitudinal studies have shown that students with MD fall even further behind their peers as they progress through high school. The National Longitudinal Transition Study-2, for example, reported that 14–27% of secondary youth with MD scored more than two standard deviations below the mean of average-achieving students across math subtests. These academic difficulties often lead to other problems such as inappropriate school behaviors. It is common knowledge that difficulties in math begin at an early age. At the elementary level, research indicates students with MD gain only 1 year of achievement on procedural skills for every 2 years they are in school. In particular, students lag far behind their classmates in whole-number operations, fractions and decimals computation, and problem solving.

One of the most obvious characteristics of students with MD is their inability to (1) develop conceptual understanding of number, (2) use this knowledge to help them retrieve basic math facts from memory, and (3) achieve fluency with math facts. Clearly, in solving problems, students with MD face challenges identifying relevant information from text due in part to reading and comprehension difficulties. These students often experience problems such as focusing attention on key task variables, self-monitoring during problem solving, and self-management. At the secondary level, students with MD experience both long-term and short-term memory deficits, have difficulty organizing information, and make more procedural errors than students without MD. The purpose of this article is to describe strategies that have been developed for remediating the aforementioned learning difficulties in math.

Historical Backdrop

In the late 1980s and early 1990s, there emerged a growing concern in the general mathematics community that students were not developing sufficient understanding of math concepts to be successful problem solvers. The publication of the NCTM standards in 1989 recommended teachers place more emphasis on helping students develop sophisticated understanding of math concepts and less importance on rote memorization. Not only the ability to compute accurately remained a legitimate goal, but it was also recognized that postsecondary settings (i.e., job and education) wanted students and workers who could compute and use math as a tool for solving important problems. In an effort to improve the math achievement of all students, including students with MD, the US Department of Education set six strategic goals. They were most likely fueled by:

1. a series of reports urging educators to help all students acquire essential skills for successful employment in the twenty-first century;
2. the low skills demonstrated by large numbers of students on national tests; and
3. the lackluster showing of students in international comparisons.

Coinciding with these recommendations, the No Child Left Behind Act of 2001 imposed an increased level of external accountability on schools and states to demonstrate higher student achievement in reading and math. The Individuals with Disabilities Education Act required students with special needs to participate in assessments with accommodations as needed.

Most recently, the National Mathematics Advisory Panel, commissioned by the US Department of Education, emphasized the need to provide all children with a sound education in math. Among its recommendations, the panel emphasized the need to teach students to proficiency, which

means developing automaticity in number facts, accurate and automatic execution of algorithms, and the use of these competencies to solve important problems. The report also suggested that curriculum should simultaneously develop students' conceptual understanding, computation fluency, and problem-solving skills. Of particular interest to teachers of students with MD is their finding that educators should not be too quick in making assumptions about the ability of children to learn math based solely on their age or presumed readiness to learn. When provided with appropriate teaching strategies, based on frequent formative assessments, all students can learn math. Researchers and teachers over the past several decades have developed these strategies, which range from explicit instruction on basic facts to constructivist approaches aimed at developing students' ability to solve problems.

Math Instruction for Students with MD

Early pioneers in math education made important contributions to the development of teaching strategies and curricular materials for students with MD. Much of this work was characterized by teacher-led explicit instruction on well-defined skills. Based primarily on behavioral orientations, teachers used direct instruction that incorporated tightly controlled and scripted lessons to teach small chunks of math skills in sequential steps. As new concepts were introduced, teachers used prompts to scaffold understanding, which were gradually phased out when students showed they no longer needed them. Frequent assessments, called mastery checks, were used to monitor student progress. When mastery was not achieved, teachers followed prescribed procedures to remediate their students' skills. As instruction for students with MD at that time typically took place in resource rooms, teachers taught students in small groups, which the teachers were encouraged to reshuffle based on student performance. Direct instruction approaches were shown to be effective for improving certain kinds of math skills of particular students with MD. However, as students with MD were more fully included in general-education classrooms, their math instruction became more aligned with that of their classmates. Both the math content and the instructional strategies used to teach the content had to be modified and adapted to enable students with MD to use the general-education curriculum. Although it remained important for all students to develop basic skills, students with MD faced the additional challenge of solving problems, promoting higher levels of conceptual understanding.

Following are some strategies developed by researchers and teachers to improve both the computation and problem-solving skills of students with MD. In most other reviews of this kind, instruction has been grouped according to instructional orientation (e.g., behavioral or cognitive);

however, these distinctions are not useful because they do not adequately account the instructional goals or the subtleties of intervention procedures. For that reason, a representative sample of methods is described, which has improved the math performance of students with MD without trying to categorize them according to orientation. It is important to note that not all worthy instructional strategies could be included in this article because of space limitations.

Schema-Based Instruction and Graphic Organizers

The most common way of teaching students problem-solving skills is with text-based problems. Test developers also use word problems to assess student performance. Unfortunately for many students with MD, they experience difficulty in math and reading, which further limits their comprehension of the problems presented in text. For students to access these representations, they must first understand the structure of the problem and then match it to knowledge networks stored in memory. These knowledge networks help students choose the strategies and mathematical operations that most likely will lead to a successful solution of the problem. Apparently, schema-based instruction helps to reduce the reading comprehension difficulties by using graphic organizers, or schemas, in the form of visual displays and graphic diagrams to help students develop meaningful representations of problem types. First, teachers explicitly teach students the structure of several problem types and how to distinguish between them. Then, teachers supply students with graphic representations of different problem types. Once students are able to identify key features of problem types, they are taught to find the solution to the problem by matching the graphic to the problem type.

The primary difference between schema-based instruction and other forms of strategy instruction is the depth at which the mathematical relations are taught in the diagrams. General strategy instruction leads students through a series of steps in solving word problems (e.g., reading the problem, drawing a picture, solving the problem, and checking the work). Often, the underlying structures of problems are not explicitly taught and problem schemata are not categorized. During schema-based instruction, however, students are taught deep structures of problems and how math procedures can be used to solve each type of problem.

Example

Yesterday, John earned \$27.50 for mowing the lawn. John spent \$5.50 for lunch. Now John has \$22.00 (Figure 1).

Graduated Guidance Sequence

Sequencing math instruction has traditionally attracted much attention and for good reason. As with many students, the abstractions posed in math problems are especially difficult for students with MD to comprehend. Often,

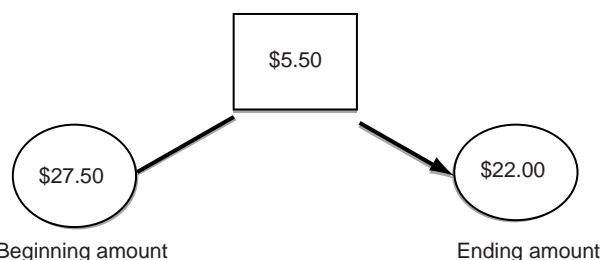


Figure 1 Graphic representation of the word problem.

students need to relate the abstract to something tangible that they already know. The sequence recommended for students with and without MD generally follows a concrete-to-representational-to-abstract (CRA) sequence. At the concrete level, the abstraction is modeled with materials (e.g., unifix cubes, blocks, and beans). At this stage, students are actively engaged in manipulating materials that show the concept they will eventually learn as an abstraction. When they are able to demonstrate their knowledge with manipulatives, teachers transfer the concepts to semiconcrete representations in the form of pictures or drawings. As they move to the abstract stage, the pictures are replaced with math symbols such as numbers and operations. At all three stages, students are afforded many opportunities to practice and demonstrate their understanding.

One way of showing the CRA sequence among middle school students with MD is with fraction strips, which are useful for demonstrating equivalence of fractions. For example, students can be given four long strips of paper and asked to imagine the first strip is a long candy bar shared by two people. Students decide to share the candy bar by folding the paper in half. The teacher then directs them to label the fold to indicate the relative size of the candy bar to that point ($1/2$). The students repeat the process with three more strips – representing candy bars that needed to be shared among 4, 8, and 16 people – using repeated folding. Each fold is labeled with its appropriate fraction. Once all four fraction strips are labeled, students are asked to offer any observations. The teachers' questions encourage students to focus on fraction equivalence and the relationship between a fraction's size and its numerator and denominator. For example, a teacher asks students to answer the following questions as they compare fractions indicated on their fraction strips:

- What is the relationship between the size of the denominator and the number of pieces into which the fraction strip is divided?
- What fractions on your strips are equivalent to $1/4$?
- What is another name for $2/8$?
- What is another name for $1/2$?
- On our fraction strips, the numerator is always smaller than or equal to the denominator. Does it always have to be this way? Could you have a fraction like $5/4$? How would you represent it?

On subsequent days, students can use their fraction strips to solve problems such as “If you have $\frac{3}{4}$ of a candy bar and your partner has $\frac{1}{2}$ of a candy bar, who has more? How much more? How much do you have together?” Teachers can ask students to show how they could use their fraction strips to solve addition and subtraction problems such as $\frac{3}{8} + \frac{1}{8}$ and $\frac{1}{4} + \frac{5}{16}$. At the beginning of this sequence, teachers encourage students to focus on their understanding of fraction equivalency rather than on highly procedural sequences or memorized rules to find common denominators. As students demonstrate their understanding of fractions and equivalence, teachers can move to instruction that emphasizes more procedural knowledge. It should be noted that the CRA strategy is recommended for use with all students in the primary grades. However, teachers may need to continue teaching older students with MD with this sequence because they lack the ability to form mental models of abstract concepts without concrete and representational scaffolds.

Peer-Mediated Instruction

Peer-mediated instruction encompasses a variety of interventions in which a student without MD assists a student with MD in learning and practicing math tasks. The most popular version of this practice is called Peer-Assisted Learning Strategies (PALS). With PALS, every student in the class is paired. Each student in the pair takes turns as a tutor and tutee while working on classroom activities provided by the teacher. Typically, PALS is implemented in elementary classrooms for several 35-min sessions per week. The underlying theory to this approach is that peer interaction, if carefully planned, can have powerful effects on academic motivation and achievement. Some of this motivation can be attributed to the increased socialization experiences that peer-mediated instruction affords both the tutor and tutee. Students tend to stay on the task longer and engage in productive practice on math facts, computational skills, and problem solving. In addition to increasing student motivation and academic progress, peer-mediated instruction enables teachers to accommodate students with a wide range of abilities. Teachers monitor the discussions of student pairs and provide additional help as needed.

To implement peer-mediated instruction effectively, teachers must first teach students how to be effective tutors and receptive tutees. Students have to learn both roles because they switch roles during each tutoring session. As a tutee, students are taught appropriate methods for seeking help when they do not understand a concept. In some instances, the pair develops a signal to indicate when the tutee needs help. These may include taps on the desk or pointing to the tutor. Tutors need to pay close

attention to tutees so they know when to offer help. They also need to describe ways of finding the answer rather than simply giving the answer. Teachers are responsible for providing tutors with objectives for the lessons and prompts that can be used strategically to reinforce skills.

Cognitive Strategy Instruction and Self-Regulation

Students with MD display difficulty with memory, attention, and self-regulation. These characteristics are due, in part, to their immature problem-solving strategies and their failure to monitor their own performance. Cognitive strategy instruction attempts to explicitly teach students to monitor their own learning activities with methods such as modeling, verbal rehearsal, and feedback. To help students solve problems, teachers provide students with strategies such as paraphrasing, visualization, and planning. Self-regulation teaches students to take control of their own learning and move to independence. Training in self-monitoring often includes ways that may help students engage in self-assessment, self-recording, and self-instruction. Students are taught to monitor their own activities related to academic and behavior. For example, students can be taught to monitor and record the number of problems they have completed or computed correctly during study time. They can record their data in a concrete way, such as on a card in their folder or attached to their desktop. When students first use the self-monitoring system, an external source may cue them to record the data. As they become more accustomed to the procedures, they eventually monitor their own activities.

One of the important features of self-regulation is that teachers can develop self-monitoring checklists and procedures based on individual needs of their students. For example, when given problems that require finding equivalent fractions, a checklist could be written this way:

1. I looked at denominators of the fractions.
2. I saw that the denominators were not the same.
3. I asked myself if the smaller denominator could be multiplied by a number to make it the same as the larger denominator.
4. If the answer to #3 is yes, I multiply the numerator and the denominator by that number.
5. If the answer to #4 is no, I need to find the least common multiple for both denominators.

As students go through their mental checklist, they talk themselves through the problem-solving routine. They may also ask themselves additional questions such as the function of denominators compared to numerators in the example provided above. To buttress self-regulation, teachers may teach students acronyms to help them remember the sequence of more generic problem-solving steps. For example, students are taught to remember

RPV-HECC (read, paraphrase, visualize, hypothesize, estimate, compute, and check) for solving word problems. Students are explicitly taught how to use each procedure in the sequence. Montague (2006) suggested the following instructional methods:

- Teachers model self-regulation strategies during math instruction.
- Students practice these strategies before applying them.
- Teachers provide students with cue cards or prompt sheets to remind them of the instruction they should follow when completing the task.
- Teachers provide students with a record of progress either in chart or graph format.
- Teachers fade cues and prompts as students become more skilled and independent, applying appropriate strategies.

It should be emphasized that students with MD may not use the strategies unless they have had ample opportunities to practice using the strategy. Ways of getting them to use these strategies independently is through explicit practice that includes (1) verbal practice designed to help students talk through the intent of the strategy; (2) controlled practice and feedback on materials written at their own instructional level; (3) advanced practice on materials with grade level materials; and (4) generalization on new but related materials.

Formative Assessment

Formative assessment is a way of monitoring the effects of instruction on the math performance of students with MD. It differs from summative assessment in the frequency of administration, the relative length of the tests, and instructional purpose. Formative tests are usually made up of small number of items that measure specifically what students were expected to have learned over a period of days or weeks. Curriculum-based measurement (CBM) is one kind of formative assessment that indexes student achievement and provides teachers with important information about how to manage instruction. From these frequent measures, growth rates (or slopes) are graphed for each student. These slopes are then compared to typical growth rates of students to assess how well students are progressing.

The use of frequent, ongoing assessment is included in this article because appropriate instruction depends on the validity and reliability of student evaluation. CBM strategies are the best known of these types of assessment strategies, but teachers can also develop their own monitoring strategies. What is most important is that they identify the math difficulties that students have and then devise methods for remediating them in a timely manner. Frequent and ongoing assessment of student skills is also a critical element of Response to Intervention (RTI), a way

of identifying patterns of achievement among students who may be at risk of developing serious problems in math. The RTI approach may help teachers make more informed decisions about how to align instruction with individual learning needs of children.

Enhanced Anchored Instruction

Enhanced Anchored Instruction (EAI) is a method specially designed for developing problem-solving skills of adolescents with MD. EAI is based on the concept of anchored instruction, popularized by the Cognition and Technology Group at Vanderbilt. In anchored instruction classrooms, learning activities are designed around an anchor, usually presented in an 8–12-min video. The video portrays actors who are attempting to solve a realistic problem. All the information for solving the problem, and its subproblems, are presented in motivating contexts in the video. To solve the problem, students have to identify what information is relevant and then formulate a plan that will eventually lead to a successful solution. A primary objective of anchored lessons is to help eliminate the inert knowledge problem that exists when typical school-based learning materials do not tap what many low-achieving students know. Students may have relevant knowledge but do not know when and under what conditions to apply it.

EAI extends anchored instruction by affording students additional opportunities to practice their skills as they solve analogous problems in applied contexts. In studies involving students with MD, students first solve a problem in a multimedia format and then apply what they learn in related hands-on problems. One important finding in these studies is the ability of EAI to directly immerse students in problem contexts, thus helping to eliminate the comprehension difficulties those with low skills in math and reading often experience with complex text-based problems. An example of a multimedia-based anchor is called *Fraction of the Cost*, which stars three middle school students. The video opens in a skateboard store and rink where students are shown discussing how they can afford to buy materials for building a skateboard ramp. To figure out the most economical use of 2×4 s ($2'' \times 4''$) lumber in *Fraction of the Cost*, students first need to identify the lengths of boards shown in the plan. Then they convert these dimensions from feet and inches to inches and try out several combinations of lengths to maximize their use of available wood. Once they compute what additional wood is needed, they consult the store advertisement to decide what to buy. Their final task is to show (1) how much money they had to spend, (2) construct a materials list, and (3) calculate sales tax and the total cost.

After students solve the problems posed in *Fraction of the Cost*, they work on solving a related problem called the *Hovercraft Challenge*, in which students plan and construct a rollover cage for a hovercraft out of polyvinyl

chloride (PVC) pipe. Small groups of students plan how they can make the cage in the most economical way. When the cages are complete, students lift them onto a 4 × 4-ft plywood platform (i.e., hovercraft). A leaf blower inserted into a hole in the plywood powers the hovercraft, which inflates the plastic attached to its underside and elevates it slightly above the floor. On the last day of the project, students ride on their hovercrafts in relay races up and down the halls of the school. Clearly, each anchor requires a mix of math skills. In the example above, to solve the problem, students need to:

1. calculate percent of money in a savings account and sales tax,
2. read a tape measure,
3. convert feet to inches,
3. visualize building plans,
4. build a table of materials,
5. compute whole numbers and mixed fractions,
6. estimate and compute combinations, and
7. compute total cost.

Conclusion

Over the past two decades, considerable progress has been made in developing and testing strategies and curricula to enhance the computation and problem-solving skills of children with MD. As more students with MD were included in general-education classrooms, instruction for these students had to be crafted in ways that could be used alongside their peers without MD. Some of these methods are summarized in this article. It is important to note that explicit instruction still benefits a large number of students with MD and should be continued. However, math instruction should take a balanced approach that includes both student-centered and teacher-directed practices.

Bibliography

Montague, M. (2006). Self-regulation strategies for better math performance in middle school. In Montague, M. and Jitendra, A. (eds.) *Teaching Mathematics to Middle School Students with Learning Difficulties*, pp 89–107. New York: Guilford Press.

Further Reading

- Allsopp, D. H., Kyger, M. M., Lovin, L., et al. (2008). Mathematics dynamic assessment: Informal assessment that responds to the needs of struggling learners in mathematics. *Teaching Exceptional Children* **40**(3), 6–16.
- Botte, B. A., Rueda, E., LaRoque, P. T., Serlin, R. C., and Kwon, J. (2007). Integrating reform-oriented math instruction in special education settings. *Learning Disabilities Research and Practice* **22**, 96–109.
- Botte, B. A., Rueda, E., Serlin, R. C., Hung, Y., and Kwon, J. (2007). Shrinking achievement differences with anchored math problems: Challenges and Possibilities. *Journal of Special Education* **41**, 31–49.

- Calhoon, M. B. and Fuchs, L. S. (2003). The effects of peer-assisted learning strategies and curriculum-based measurement on the mathematics performance of secondary students with disabilities. *Remedial and Special Education* **24**, 235–245.
- Carnine, D. (1997). Instructional design in mathematics for students with learning disabilities. *Journal of Learning Disabilities* **30**, 130–141.
- Cawley, J., Parmar, R., Foley, T. E., Salmon, S., and Roy, S. (2001). Arithmetic performance of students: Implications for standards and programming. *Exceptional Children* **67**, 311–328.
- Deshler, D. D., Schumaker, J. B., Lenz, B. K., et al. (2001). Insuring content-area learning by secondary students with learning disabilities. *Learning Disabilities Research and Practice* **16**, 96–108.
- Fuchs, L. S. and Fuchs, D. (2002). Mathematical problem-solving profiles of students with mathematics disabilities with and without comorbid reading difficulties. *Journal of Learning Disabilities* **35**, 563–573.
- Fuchs, L. S., Fuchs, D., Finelli, R., Courey, S. J., and Hamlett, C. L. (2004). Expanding schema-based transfer instruction to help third graders solve real-life mathematical problems. *American Educational Research Journal* **41**, 419–445.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., and Bentz, J. (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children* **60**, 518–537.
- Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities* **37**, 4–15.
- Gersten, R. and Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education* **33**, 18–28.
- Ives, B. and Hoy, C. (2003). Graphic organizers applied to higher-level secondary mathematics. *Learning Disabilities Research and Practice* **18**, 36–51.
- Jitendra, A., DiPipi, C. M., and Perron-Jones, N. (2002). An exploratory study of schema-based word-problem-solving instruction for middle school students with learning disabilities: An emphasis on conceptual and procedural understanding. *Journal of Special Education* **36**, 23–38.
- Lee, J., Grigg, W., and Dion, G. (2007). *The Nation's Report Card: Mathematics 2007 (NCES 2007-494)*. Washington, DC: National Center for Education Statistics/Institute of Education Sciences/U.S. Department of Education.
- Ma, L. (1999). *Knowing and Teaching Elementary Mathematics: Teachers' Understanding of Fundamental Mathematics in China and the United States*. Mahwah, NJ: Erlbaum.
- Maccini, P., Mulcahy, C. A., and Wilson, M. G. (2007). A follow-up of mathematics interventions for secondary students with learning disabilities. *Learning Disabilities Research and Practice* **22**, 58–74.
- Miller, S. P. and Hudson, P. J. (2007). Using evidence-based practices to build mathematics competence related to conceptual, procedural, and declarative knowledge. *Learning Disabilities Research and Practice* **22**, 47–57.
- Montague, M. (2007). Self-regulation and mathematics instruction. *Learning Disabilities Research and Practice* **22**, 75–83.
- National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- National Mathematics Advisory Panel (2008). *Foundations for Success: The Final Report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.
- Schoenfeld, A. H. (1989). Teaching mathematical thinking and problem solving. In Resnick, L. B. and Klopfer, L. E. (eds.) *Toward the Thinking Curriculum: Current Cognitive Research*, pp 83–103. Alexandria, VA: Association for Supervision and Curriculum Development.
- Swanson, H. L. and Deshler, D. (2003). Instructing adolescents with learning disabilities: Converting a meta-analysis to practice. *Journal of Learning Disabilities* **36**, 124–135.
- Wagner, M., Newman, L., Cameto, R., and Levine, P. (2006). *An Overview of Findings from Wave 2 of the National Longitudinal Transition Study-2 (NLTS2)*. (NCSE 2006-3004). Menlo Park, CA: SRI International.
- Woodward, J. (2004). Mathematics education in the United States: Past to present. *Journal of Learning Disabilities* **37**, 16–31.

Relevant Websites

<http://www.cited.org> – Center for Implementing Technology in Education.
<http://www.cec.sped.org> – Council for Exceptional Children, CEC News.

<http://www.ldonline.org> – LD Online, Math.
<http://www.ncld.org> – National Center for Learning Disabilities.
<http://ies.ed.gov> – Organizing Instruction and Study to Improve Student Learning: A Practice Guide.
<http://www.ed.gov> – US Department of Education.

Overrepresentation of Culturally and Linguistically Diverse Students in Special Education

S C Trent, University of Virginia, Charlottesville, VA, USA

© 2010 Elsevier Ltd. All rights reserved.

Overrepresentation of Culturally and Linguistically Diverse Students in Special Education

The overrepresentation of historically underserved students (historically underserved students are students from diverse racial/ethnic, linguistic, and socioeconomically disadvantaged backgrounds who have traditionally experienced significant school failure; this term is used instead of minority students or culturally and linguistically diverse learners (CLD) to illuminate the responsibility of educational agencies and school districts to address historical, political, and systemic constraints that impede learning rather than adopting a traditional deficit view which places the blame for school failure primarily on students and their families) in special education in the high incidence categories of learning disabilities (LD), intellectual or developmental disabilities (IDD) (IDD has replaced the label formally referred to as mild mental retardation), and emotional and/or behavioral disorders (E/BD) has been a controversial issue in education for several decades. Overrepresentation is defined “as the representation of a group in a category that exceeds our expectations for that group, or differs substantially from the representation of others in that category” (Skiba *et al.*, 2008: 266).

Generally, two methods are used to determine if students are overrepresented in special education: the composition index and the risk index or risk ratio method. The composition index determines “the extent to which a group is over- or underrepresented in a category compared to its proportion in the broader population” (Skiba *et al.*, 2008: 266). The risk index and risk ratio determines “the extent to which a group is found eligible for service at a rate differing from that of other groups” (see Skiba *et al.*, 2008: 266). Widespread use of these measures over three decades reveals that overrepresentation does indeed exist (Artiles and Trent, 1994; Chinn and Hughes, 1987; Donovan and Cross, 2002; Heller *et al.*, 1982; Parrish, 2002). For example, over time, researchers have determined that African-American students continue to be overrepresented in programs for students with IDD and E/BD (Donovan and Cross, 2002). Researchers have also found that proportionally, some historically underserved students with disabilities, particularly African-American students, are placed in more restrictive settings than their White counterparts.

Moreover, parallels exist between poor school performance, high-incidence disabilities, and time served in the juvenile justice system (Drakeford and Staples, 2006). To a lesser degree, Native American, Latina/o, and some Asian or Pacific Islanders subgroups continue to be overrepresented in some of the high-incidence categories (Osher *et al.*, 2004; Parrish, 2002).

Some scholars purport that overrepresentation is unfortunate yet often necessary because many historically underserved students come from economically disadvantaged backgrounds (MacMillan and Reschly, 1998). Risk factors used to support this argument include, but are not limited to, (1) improper prenatal care, (2) teenage parenthood, (3) single parent households, (4) low birth weight, (5) poor nutrition, and (6) limited opportunities to engage in school readiness activities in the home and community prior to school enrolment. While there is ample evidence that documents the adverse effects of risk factors on student performance, results from large-scale studies reveal that only about 33% of the racial gap in education is explained by socioeconomic status (Hedges and Nowell, 1999; Mickelson, 2003). In addition, Skiba *et al.* (2005) found in their research that, “poverty makes a weak and inconsistent contribution to the prediction of disproportionality across a number of disability categories” (p. 130). These findings raise the questions: Can overrepresentation of these students in special education be attributed predominantly to deficits that reside within them? Are there school factors that contribute to overrepresentation and underachievement?

Of course, it is known that significant numbers of historically underserved students attend poorly funded schools with inadequate human and material resources (Children’s Defense Fund, 2004). Obviously, these school factors adversely affect outcomes for these learners. But, the question remains, are there other school factors that contribute to dismal outcomes that have been reported for decades? Emerging research supports the hypothesis that there are, and that throughout our nation’s history, we have placed preeminent focus on identifying deficits that reside within children, their families, and their communities. In addition, we have placed far too little emphasis on understanding how schools can create and sustain environments that provide culturally responsive, evidence-based practices that engage all students affectively, educationally, and developmentally (Cartledge and Kourea, 2008; Deschenes

et al., 2001; Dorn, 2007). One can hypothesize further that the overrepresentation of these students in high-incidence special education programs and their subsequent low academic achievement may be attributed to a preeminent focus on deficit thinking. The under-acknowledgement of these truths has had a profound adverse affect on attempts to meet the needs of historically underserved students more effectively, even when reform initiatives emanate from genuine concern and good intentions (Anderson, 1998; Spencer 2008a, 2008b).

In order to chart a more successful course for the present and the future, it is necessary to understand more completely how deficit thinking has influenced past educational practices for historically underserved students. This effort might lead to a better understanding of the patterns in behavior and thinking that have contributed and continue to contribute to the sustainment of overrepresentation, poor academic outcomes, and limited postsecondary opportunities for these learners (Obiakor and Ford, 2002; Trent and Artiles, 2007). Understanding of this history, if acknowledged, valued, and incorporated into current and future educational reforms, can also contribute to the creation of more effective school environments for these students. Hence, in the remainder of this article, the author summarizes the traditional deficit thinking that still guides educational policy and practices, presents under-acknowledged theories and research that challenge this thinking, and elucidates how these theories can be used to create and sustain school environments where more students experience success. Note that the content herein is not an indictment of special education for historically underserved students. Instead, it is a call for equitable educational opportunities for all children, especially those who historically have not experienced sustained access to effective and equitable educational opportunities in general and/or special education (Edgar *et al.*, 2002; Lee, 2008).

Traditional Explanations for School Failure among Historically Underserved Students

For many years, researchers have developed explanations for the sustained school failure and overrepresentation in special education in the high-incidence categories among historically underserved students. Causes include legal and administrative requirements, characteristics of students, quality of the instruction received, possible biases in the assessment process, characteristics of the home and family environment, and broader historical and cultural contexts (i.e., overemphasis on a deficit perspective and conflicts between minority and dominant cultures) (Chinn and Hughes, 1987; Heller *et al.*, 2002). Other factors identified by Donovan and Cross (2002)

include “education personnel, class size, and school funding” (p. 172). Categorization of these factors reveals positions that explain school failure for these students – cultural deficits (deficit thinking) and cultural differences.

Cultural Deficits: It's Something Within

Deficit thinking posits that, “the student who fails in school does so because of internal deficits or deficiencies. Such deficits manifest, it is alleged, in limited intellectual abilities, linguistic shortcomings, lack of motivation to learn and immoral behavior” (Valencia, 1997: 2). These beliefs have resulted in the emergence of explanations for school failure that include (1) the genetic pathology model, (2) the cultural disadvantage model, and (3) the cultural and accumulated environmental deficit model.

Genetic pathology. Proponents of the genetic pathology model argued that genetic make-up played a more substantial role in determining cognitive ability than did the environment. Moreover, proponents of this model insisted that the racial achievement gap was mainly the result of inherent racial differences in intellectual capabilities (McCray *et al.*, 2003; Menchaca, 1997).

Cultural disadvantage. The liberal orientation embraced by this country during the 1960s diminished the popularity of the genetic pathology model and ushered in ideals of equality and social justice for all US citizens. Foley (1997) and Pearl (1997) asserted that during this period, even though a desire to elevate the status of historically underserved citizens was the stated intent of participants of this movement, one form of pathology was replaced by another to explain the school failure of students from these backgrounds. More specifically, Pearl argued:

“[t]he *cultural deprivation model* singled out the family unit (rather than genes) as the transmitter of deficiencies. Although culture (or the alleged lack of it, as indicated by the notion of deprivation) was central to the cultural deprivation model, the family was key to this deficit framework. The family unit – mother, father, home environment – was pegged as the carrier of the pathology.” (p. 133)

Cultural and accumulated environmental deficits. Supporters of the cultural and accumulated environmental deficit model argued that deficiencies within the family unit that are sustained over time will lead to cognitive deficits that are irreversible. Due to the inadequate socialization of parents and communities, historically underserved children did not receive the adequate stimulation needed to prepare them to be successful in school. These enduring environmental deficits made it virtually impossible to break the cycle of poverty and limited the intellectual capacity of these children; it was highly unlikely that the cycle could be broken (Foley, 1997).

In sum, this deficit focus contributed to school failure for many historically underserved children; and they, their parents, and their communities were blamed primarily for this failure. These beliefs were sustained in part because educators rarely entertained the notion that poor school performance might also be linked to other factors such as their beliefs, poorly funded schools, and inappropriate instruction. These factors, coupled with Darwinian theory, the introduction of Jim Crow legislation, and immigration laws made it morally and legally justifiable to segregate these students from mainstream students (Deschenes *et al.*, 2000; Span, 2003; Valencia and Suzuki, 2001). Also, standardized intelligence tests administered in English became very popular in the US during the first three decades of the twentieth century and provided scientific justification for segregated schooling practices. Many historians lauded the work of those who developed special education programs largely based on results from IQ tests (e.g., Kode, 2002). Armed with this new technology, educators believed that they would be able to sort students into programs that were more aligned with their abilities and provide them with the tools they needed to be successful in life. However, those who have studied this era through a sociocultural lens contend that the use of IQ tests resulted in segregated classrooms and schools for recent immigrants, African-Americans, and Native Americans. The primary goal was to separate these children from the true Americans, those of Western European descent, who had been fully assimilated into the majority culture (Dorn, 2007; Valencia and Suzuki, 2001). Furthermore, the curriculum was standard and focused largely on remediation and vocational skills (Safford and Safford, 1998). Despite evidence to the contrary, these beliefs have persisted over time and remain with us today (Herrstein and Murray, 1994; Heubert, 2002; Spencer, 2008b).

Alternative Explanations for School Failure: Beyond Blaming the Victim

Several researchers and theorists disagree with models derived from deficit thinking. Banks (2008) provided a reason for the rejection of this argument, *to wit*:

Unlike cultural deprivation theorists, cultural difference theorists reject the idea that low-income students and students of color have cultural deficits. They believe that ethnic groups such as African-Americans, Mexican Americans, Asian Americans, and Native Americans have strong, rich, and diverse cultures (Gay, 2000; Alim and Baugh, 2007). These cultures consist of languages, values, behavioral styles, and the perspectives that can enrich the lives of all Americans The school must change in ways that will allow it to respect and reflect the culture of low-income students and students of color and at the

same time use teaching strategies that are consistent with their cultural characteristics. (p. 54)

Models derived from cultural difference theory include: (1) cultural incongruity, (2) the oppositional cultural framework, and (3) stereotype threat.

Cultural Incongruity Model

Those who embrace this model purport that historically underserved students experience persistent school failure not because of deficits that reside within them, but because aspects of their cultures are incompatible with the thinking, behaviors, and values of mainstream culture (Ladson-Billings, 1994). These incompatibilities result in lowered expectations, intolerance, and inequitable decision-making processes. For example, research conducted during the school desegregation era as well as current research reveals that historically underserved students are suspended from school at significantly higher rates than White students (Yudof, 1975). In addition, researchers have found that justifications for suspensions for the former group are based more on subjective (e.g., overlapping speech seen as disrespectful) versus objective criteria (e.g., physical altercations) and are influenced by factors in and outside of school. These factors include criminalization of African-Americans by the media, race and class privilege emanating from cultural norms that dictate what is considered normal and abnormal, and zero tolerance policies (Monroe, 2005).

Oppositional Cultural Framework

The underlying thesis of this model is that involuntary immigrants (e.g. African-American students), due to stereotypical beliefs and sustained discrimination on the part of the majority culture, take on reactionary thinking and behaviors (e.g., variant dialects, unpreparedness, brash or disrespectful interactions with authority figures) that are oppositional to school success. In part, these behaviors are attributed to students' fears of being labeled as a wannabe or as acting White if they overtly engage in teacher pleasing behaviors and experience school success. These behaviors threaten their designation as valued members of their peer groups (Mickelson, 2003; Ogbu, 2003).

Stereotype Threat

Proponents of this model argue that long-standing stereotypes about people from historically underserved groups can have an adverse effect on the academic performance of students from these groups. Steele (1977) demonstrated empirically the existence of this model. As Mickelson (2003) summarized, Steele found that if, "academically able black students are cued about race before engaging in

an intellectual task, their performance will be lower than that of comparable blacks who did not receive such a cue” (p. 1069). This anxiety suppresses motivation to engage in intellectually challenging activities because if students fail, they are validating the stereotypes (Artiles *et al.*, 2009).

Models based on cultural difference theory afford schools and educators the opportunity to consider factors beyond within-group deficits to explain school failure. By viewing teaching and learning through this lens, one can see that there are multiple factors that contribute to school failure including the role of history in shaping beliefs about people and their children and how these beliefs influence the identity, development, and performance of these children. Furthermore, these models elucidate the importance of instructional relevance and responsiveness. Also, this theory reinforces the belief that all students come to school with funds of knowledge that can be used to inform instruction (Gonzalez and Moll, 2002). In this vein, educators need to examine more closely their frames of reference about others’ children, develop ways to access information that might mitigate biases and stereotypes, and from this inquiry develop instructional practices that are more culturally responsive, evidence-based, and effective (Berry and McClain, 2009).

Still, just as deficit theory alone is inadequate to explain school failure, cultural difference theory, if not operationalized contextually, will not bring about the changes in education that teacher educators and practitioners seek to achieve. Instead, an ecological/contextual conceptualization that accounts for the interplay between deficits and differences, risk and protective factors, and the political nature of schooling must be used to frame current problems and create more far-reaching and efficacious solutions. An example of how this framework can be used to address overrepresentation is provided below.

Toward an Ecological and Contextual View of Overrepresentation

Within an ecological/contextual view of overrepresentation, the unit of analysis is not students from historically underrepresented groups and the deficits that reside within them, but these students within the context of an extremely complex and interactive activity system charged to meet their educational needs. The interconnected, interactive layers of the activity system include the experiences of these children within multiple contexts such as the larger society (outer layer), their families and communities, and their school districts, schools, and classrooms (inner layer). In other words, learning does not occur in a vacuum, but within situated contexts; and these contexts shape how students view themselves and how others view them. If others’ acknowledged or unacknowledged beliefs about students are adversely influenced by culturally based

constructs grounded in pathology, their experiences in school may lead to less than desired results. In fact, results may be injurious.

Spencer (2008b) provided an example of how the interplay between the layers of the activity system might result in such undesirable outcomes. While acknowledging that special education has resulted in many benefits for many children, she explained that practices that lead to special education placement may put many historically underserved students at risk. Using an ecological/contextual framework, she argued that the study of overrepresentation not only requires more than acknowledgement of deficits or cultural mismatches, but requires a deeper understanding of how societal factors (e.g., prevailing social norms) may influence educators’ views about historically underserved students. Furthermore, these norms will subsequently mediate these educators’ thinking and actions in schools and classrooms. For example, educators who adopt prevailing social norms about low-income African-American males may categorize these students as would-be criminals who, due to their disrespect for norms and rule-governed behaviors set forth by the larger society, will likely end up in the juvenile justice system (Monroe, 2005). Internalization of these beliefs adversely affects teacher efficacy and consequently, the performance of students who have been marginalized. This marginalization is exacerbated when these same educators do not acknowledge that many children from historically underserved backgrounds are expected to make the same gains as more privileged children despite the obvious inequities faced by the former. Failure to acknowledge that protective factors are provided in more abundance for privileged students leads to the false assumption that the contexts for both groups of students are virtually the same. This assumption may result in feelings of superiority and entitlement on the part of privileged students and detract from the identity development and motivation of historically underserved students (e.g., stereotype threat).

Because failure is attributed to pathology, members of the educational enterprise tend to conclude that some students need instruction that must be delivered by trained special education teachers. This pathological status does not acknowledge contexts that influence cognitive and emotional development and how risk factors in schools such as teacher beliefs and quality of instruction may affect students’ social and academic performance. In this situation, significant numbers of students from historically underserved backgrounds will be placed in special education and many of them will be served in more restrictive settings than White students. Consistent with the progressive education movement of the early twentieth century, there is a high likelihood that instruction in these settings will be inadequate and unresponsive to the needs of students.

Implications and Conclusions

Given the ecological/contextual theory outlined above, stakeholders who are truly committed to equitable educational practices for all children must ask questions such as: What are the political, sociocultural, and economic factors that have cumulatively contributed to the overrepresentation of historically underserved students in special education? What theories and models have been used to frame the discourse on overrepresentation? Whose theories and models have been ignored? What contextual factors must change if heretofore unacknowledged voices are to be heard, valued, and incorporated? How might power, racism, and/or classism contribute to overrepresentation? What processes and rules are needed to sustain authentic participation among all stakeholders including parents? Finally, the most important question we need to ask is: How do we create and sustain safe educational environments where the desired collective outcome is not higher scores on state-wide assessments but, "more equal levels of student achievement and improved social outcomes for all students" (Anderson, 1998: 575)? Otherwise, we will continue to implement policies and practices that historically have not addressed the root of the problem, and the persistent failure and overrepresentation in special education of significant numbers of students from historically underserved groups will be unabated.

Bibliography

- Alim, H. S. and Baugh, J. (eds.) (2007). *Talkin Black Talk: Language, Education, and Social Change*. New York: Teachers College Press.
- Anderson, G. L. (1998). Toward authentic participation: Deconstructing the discourses of participatory reforms in education. *American Educational Research Journal* 35(4), 571–603.
- Artiles, A. J., Kozleski, E. B., Trent, S. C., Osher, D., and Ortiz, A. (2009). Justifying and explaining disproportionality, 1968–2008: A critique of underlying visions of culture. Unpublished manuscript submitted for publication.
- Artiles, A. J. and Trent, S. C. (1994). Overrepresentation of minority students in special education: A continuing debate. *Journal of Special Education* 27(4), 410–437.
- Banks, J. A. (2008). *An Introduction to Multicultural Education*, 4th edn. New York: Pearson.
- Berry, R. Q., III and McClain, O. L. (2009). Contrasting pedagogical styles and their impact on African American students. In Schoenfeld, A. H. (ed.) *Mathematics Teaching, Learning, and Liberation in the Lives of Black Children*, pp 123–144. New York: Routledge.
- Children's Defense Fund (2004). *The State of America's Children 2004*. Washington, DC: Author.
- Chinn, P. C. and Hughes, S. (1987). Representation of minority students in special education classes. *Remedial and Special Education* 8, 41–46.
- Deschenes, S., Cuban, L., and Tyack, D. (2001). Mismatch: Historical perspectives on schools and students who don't fit them. *Teachers College Record* 103, 525–547.
- Donovan, S. and Cross, C. (eds.) (2002). *Minority Students in Special and Gifted Education*. Washington, DC: National Academy Press.
- Dorn, S. (2007). *Accountability Frankenstein: Understand and Taming the Monster*. Charlotte, NC: Information Age.
- Drakeford, W. and Staples, J. M. (2006). Minority confinement in the juvenile justice system: Legal, social, and racial factors. *Teaching Exceptional Children* 39(1), 52–58.
- Edgar, E., Patton, J. M., and Day-Vines, N. (2002). Democratic dispositions and cultural competency: Ingredients for school renewal. *Remedial and Special Education* 23, 231–241.
- Foley, E. E. (1997). Deficit thinking models based on culture: The anthropological protest. In Valencia, R. R. (ed.) *The Evolution of Deficit Thinking: Educational Thought and Practice*, pp 113–131. Bristol, PA: Falmer Press.
- Gay, G. (2000). *Culturally Responsive Teaching: Theory, Research and Practice*. New York: Teachers College Press.
- Gonzalez, N. and Moll, L. (2002). Cruzando el puente: Building bridges to funds of knowledge. *Educational Policy* 16(4), 623–641.
- Hedges, L. V. and Nowell, A. (1999). Changes in the Black–White gap in test scores. *Sociology of Education* 72, 111–135.
- Heller, K. A., Holtzman, W., and Messick, S. (eds.) (1982). *Placing Children in Special Education: A Strategy of Equity*. Washington, DC: National Academy Press.
- Herrstein, R. J. and Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.
- Heubert, J. P. (2002). Disability, race, and high-stakes testing of students. In Olsen, D. J. and Orfield, G. (eds.) *Racial Inequity in Special Education*, pp xv–xxxvii. Cambridge, MA: Harvard Education Press.
- Kode, K. (2002). *Elizabeth Farrell and the History of Special Education*. Arlington, VA: Council for Exceptional Children.
- Ladson-Billings, G. (1994). *The Dreamkeepers: Successful Teachers of African American Children*. San Francisco, CA: Jossey-Bass.
- Lee, C. D. (2008). The centrality of culture to the scientific study of learning and development: How an ecological framework in education research facilitates civic responsibility. *Educational Researcher* 37, 267–279.
- MacMillan, D. and Reschly, D. (1998). Overrepresentation of minority students: The case for greater specificity or reconsideration of the variables examined. *Journal of Special Education* 32, 15–24.
- McCray, A. D., Webb-Johnson, G., and Neal, L. I. (2003). The disproportionality of African American in special education: An enduring threat to equality and opportunity. In Yeakey, C. C. and Henderson, R. D. (eds.) *Surmounting All Odds: Education, Opportunity, and Society in the New Millennium*, pp 455–485. Greenwich, CT: Information Age.
- Menchaca, M. (1997). Early racist discourses: Roots of deficit thinking. In Valencia, R. R. (ed.) *The Evolution of Deficit Thinking: Educational Thought and Practice*, pp 13–40. Bristol, PA: Falmer Press.
- Mickelson, R. A. (2003). When are racial disparities in education the result of racial discrimination? A social science perspective. *Teachers College Record* 105, 1052–1086.
- Monroe, C. L. (2005). Why are 'bad boys' always Black? *Clearing House* 79, 45–50.
- Obiakor, F. E. and Ford, B. A. (2002). Educational reform and accountability: Implications for African Americans with exceptionalities. *Multiple Voices for Ethnically Diverse Exceptionalities* 5, 83–93.
- Ogbu, J. (2003). *Black American Students in an Affluent Suburb: A Study of Academic Disengagement*. Mahwah, NJ: Erlbaum.
- Osher, D., Cartledge, G., Oswald, D., Artiles, A. J., and Coutinho, M. (2004). Issues of cultural and linguistic competency and disproportionate representation. In Rutherford, R., Quinn, M., and Mather, S. (eds.) *Handbook of Research in Behavioral Disorders*, pp 54–77. New York: Guilford.
- Parrish, T. (2002). Racial disparities in the identification, funding, and provision of special education. In Losen, D. J. and Orfield, G. (eds.) *Racial Inequities in Special Education*, pp 15–37. Cambridge, MA: Harvard Education Press.
- Pearl, A. (1997). Cultural and accumulated environmental deficit models. In Valencia, R. R. (ed.) *The Evolution of Deficit Thinking: Educational Thought and Practice*, pp 132–159. Bristol, PA: Falmer Press.
- Safford, P. L. and Safford, E. H. (1998). Visions of the special class. *Remedial and Special Education* 19, 229–238.
- Skiba, R. J., Poloni-Staudinger, L., Simmons, A. B., Feggins-Azziz, L. R., and Chung, C. G. (2005). Unproven links: Can poverty

- explain ethnic disproportionality in special education? *Journal of Special Education* **39**(3), 130–144.
- Skiba, R. J., Simmons, A. B., Ritter, S., et al. (2008). Achieving equity in special education: History, status, and current challenges. *Exceptional Children* **74**, 264–288.
- Span, C. M. (2003). 'Knowledge is light, knowledge is power': African American education in antebellum America. In Yeakey, C. C. and Henderson, R. D. (eds.) *Surmounting All Odds: Education, Opportunity, and Society in the New Millennium*, pp 3–29. Greenwich, CT: Information Age.
- Spencer, M. B. (2008a). Lessons learned and opportunities ignored since Brown v. Board of Education: Youth development and the myth of a color-blind society. *Educational Researcher* **37**, 253–266.
- Spencer, M. B. (2008b). Phenomenology and ecological systems theory: Development of diverse groups. In Damon, W. and Lerner, R. M. (eds.) *Child and Adolescent Development: An Advanced Course*, pp 696–740. Hoboken, NJ: Wiley.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist* **52**, 613–629.
- Trent, S. and Artiles, A. J. (2007). Today's multicultural, bilingual, and diverse schools. In Turnbull, R., Turnbull, A., Shank, M., and Smith, S. J. (eds.) *Exceptional Lives: Special Education in Today's Schools*, 5th edn., pp 56–79. Upper Saddle River, NJ: Pearson.
- Valencia, R. R. (1997). Conceptualizing the notion of deficit thinking. In Valencia, R. R. (ed.) *The Evolution of Deficit Thinking: Educational Thought and Practice*, pp 1–12. Bristol, PA: Falmer Press.
- Valencia, R. R. and Suzuki, L. A. (2001). *Intelligence Testing and Minority Students: Foundations, Performance Factors, and Assessment Issues*. Thousand Oaks, CA: Sage.
- Yudof, M. G. (1975). Suspension and expulsion of Black students from the public schools: Academic capital punishment and the constitution. *Law and Contemporary Problems* **39**(2), 374–411.

Further Reading

- Artiles, A. J. (2003). Paradoxes and dilemmas in special education's changing identity: Culture and space in inclusion and racial representation discourses. *Harvard Educational Review* **73**(2), 164–202.
- Hall, S. (2002). Race, articulation, and societies structured in dominance. In Essed, P. and Goldberg, D. G. (eds.) *Race Critical Theories. Text and Context*, pp 38–68. Malden, MA: Blackwell.
- Lee, C. D. (2007). *Culture, Literacy, and Learning: Taking Bloom in the Midst of the Whirlwind*. New York: Teachers College Press.
- Lee, J. (2006). *Tracking Achievement Gaps and Assessing the Impact of NCLB on the GAPS: An In-Depth Look into National and State Reading and Math Outcome Trends*. Cambridge, MA: The Civil Rights Project at Harvard University.
- Losen, D. J. and Orfield, G. (2002). *Racial Inequity in Special Education*. Cambridge, MA: Harvard Education Press.
- Spencer, M. B. and Harpalani, V. (2004). Nature, nurture, and the question of 'How?': A phenomenological variant of ecological systems theory. In Garcia-Coll, C., Kearer, K., and Lerner, R. (eds.) *Nature and Nurture: The Complex Interplay of Genetic and Environmental Influences on Human Behavior and Development*, pp 53–77. Mahwah, NJ: Erlbaum.

Paraprofessionals in Special Education

B Y Ashbaker and S C Dunn, Brigham Young University, Provo, UT, USA

J Morgan, Swansea Metropolitan University, Wales, UK

© 2010 Elsevier Ltd. All rights reserved.

Prior to 1975, the US Congress found that the educational needs of children with disabilities were not being fully met. In fact, Congress found that more than 50% of children with disabilities in the United States were not receiving appropriate educational services, and one million children with disabilities were entirely excluded from the public school system and from being educated with their peers. Further, there were many children with disabilities in regular school programs who were prevented from having successful educational experiences because their disabilities remained undiagnosed PL 94-142. Congress enacted the Education for All Handicapped Children Act (EAHCA) of 1975 to ensure that all children with disabilities have available to them a free appropriate public education that emphasizes special education and related services designed to meet their unique needs and prepare them for further education, employment, and independent living. The EAHCA was reauthorized as the Individuals with Disabilities Education Act (IDEA) and again as the Individuals with Disabilities Education Improvement Act (IDEIA) of 2004. This later law tried to set standards for school districts in educating students with disabilities and in providing federal funding to support appropriate services for these students.

In 2001 the No Child Left Behind (NCLB) Act added new emphasis to the legislation as it required paraprofessionals working in Title I programs to have training beyond the high school requirement that most schools held. Likewise, the changes in the IDEIA requirement were intended to more closely match the NCLB requirements for instructional paraprofessionals, but continue to remain essentially unchanged.

What did change, however, were training requirements for paraprofessionals. By incorporating principles of NCLB, the IDEIA included a requirement that the personnel who were needed to carry out the services (including paraprofessionals) must be appropriately and adequately trained, and supervised. But, with increased costs and shrinking budgets, school administration had become more creative in utilizing the services of paraprofessionals in place of higher-paid professional staff. Federal regulations require states and districts to ensure that paraprofessionals are only assigned appropriate duties and that they are properly trained and supervised (Ashbaker and Morgan, 2006). Assigning paraprofessionals to duties that should be performed by teachers or other more highly qualified credentialed staff violated the

law. Moreover, failing to provide them with proper training appears to compromise the student's right to a free and appropriate public education. This article focuses on the changing roles of paraprofessionals in the field of special education.

Paraprofessionals have become increasingly important in the US public education system. Indeed, in the 2003–04 school year, 91% of public schools reported employing at least one paraprofessional (Hampden-Thompson *et al.*, 2007); yet, there has been confusion on what to call the personnel who provide student support through tutoring, assisting, or working as a student and teacher aide in the instructional process (Morgan and Ashbaker, 2001). A national change began when the NCLB Act specifically defined paraprofessionals and specified their responsibilities and limitations with regard to their employment by the local educational agency (LEA). A portion of IDEIA federal money is spent each year to provide services of paraprofessionals to assist children with disabilities; and today, most public schools have instructional paraprofessionals employed as support staff responsible for assisting in the delivery of instruction (Hoffman and Sable, 2006).

When paraprofessionals were first introduced into classrooms, their main role was one of support to the teacher in the form of clerical work. Typically, their duties consisted of taking attendance, checking papers, preparing materials and bulletin boards, organizing paperwork, collecting money, and other clerical or housekeeping duties (Blalock, 1991; French, 1999). Over time, this workforce of people has come to be utilized for many other important tasks within classrooms. Paraprofessionals are often found in self-contained classes, resource rooms, and bridging inclusion activities in general-education classrooms. Assigned classrooms will determine the types of roles performed by paraprofessional. Broer *et al.* (2005) found that students with intellectual disabilities viewed paraprofessionals as having various roles such as mother, friend, protector, and primary teacher. Chopra *et al.* (2004) suggested that paraprofessionals also play the role of connector, especially since they connect parents to teachers, parents to community services, students to teachers, students to parents, students to peers, and students to the curriculum. In many instances, paraprofessionals work one-on-one with students or conduct small group activities that have been prepared by the classroom teacher (Daniels and McBride, 2001; Minondo *et al.*, 2001). Rueda and Monzo (2002) identified the major

roles of the paraprofessional to include (1) instruction, (2) school support, (3) liaison, (4) personal support, and (5) one-to-one class support. With each of these roles, come specific responsibilities. Oftentimes, children with disabilities need additional supports or related services in order to receive a free and appropriate public education. Paraprofessionals may also serve under the direction of speech–language pathologists, physical therapists, and occupational therapist (often called para-therapists) as they deliver related services while being trained and supervised by professionals (IDEA Public Law, Section 26(a)). Their services have been found to be important and needed.

Paraprofessionals in the United Kingdom

In the United Kingdom, the term used for paraprofessionals in government documentation is teaching assistants (TAs). In 2000, the Department for Education and Science (DfES) published *Supporting the TA: A Good Practice Guide*, which suggested that TAs provide four different types of support noted below:

1. For pupils – all pupils they encounter during their working day, not just those they have responsibility for.
2. For teachers – by carrying out routine tasks and more specialist teaching roles.
3. For the curriculum – particularly literacy and numeracy.
4. For the whole school – as part of a team which promotes the well-being of pupils.

According to the UK's Department for Education and Science (DfES, 2000), there were over 146 000 TAs in schools in England in 2005. **Table 1** shows the proportion of TAs working in elementary, secondary, and special schools, and also provides a comparison with the numbers and proportions of teachers working in those settings. The number of TAs has been steadily increasing since 1997, when there were only about 61 000 in England, with the largest increase around 2004–05 with the UK government's workforce-remodeling initiative. Apparently, this initiative gave teachers a half-day each week for

preparation and planning, and allowed TAs (who were appropriately qualified and experienced) to cover for them (Morgan, 2007). In addition to these figures for English schools, there were more than 13 000 TAs working in schools in Wales, and in Scotland (where the title is Classroom Assistants) more than 7500 TAs working in primary schools alone.

The Training and Development Agency for Schools (TDA) – the UK government department responsible for standards in teaching – in 2002, produced the National Occupational Standards (NOS) for teaching/classroom assistants. These standards contain competencies not only considered necessary for effectiveness, but also appropriate for developing job descriptions, recruiting new staff, and conducting staff appraisals. **Table 2** lists the basic concepts underpinning the standards.

Additionally, there are professional standards for higher-level teaching assistants (HLTAs) – a relatively new qualification represented as level 4 on **Table 5**. The HLTA program was conceived in response to concerns over teacher workloads and over who could provide classroom support at a level to help teachers with their responsibilities while preserving teachers' status. In 2004, the government document: *The Education (Specified Work and Registration) Regulations* provided guidance on who could appropriately undertake tasks such as preparing, planning, and delivering lessons and assessing pupil performance, tasks formerly considered appropriate only for teachers. The document authorized schools to use TAs for this type of work, but preferably only those with higher-level status. Standards for HLTAs are listed in **Table 3**.

As it appears, the standards cover a very broad range of activities relating closely to a teacher's role. However, there is a clear indication that HLTAs are still required to work under the direction of a teacher.

Paraprofessionals and Training Programs

An in-depth review of the literature on the use of paraprofessionals in special education concludes that there is insufficient evidence – which is minimal, at best – to support the practice of relying on paraprofessionals in

Table 1 Numbers of teaching assistants (TAs) and teachers in England in 2005

Stage	Number (and %)	
	TAs	Teachers
Primary schools	95 460 (65%)	204 840 (47%)
Secondary schools	29 980 (21%)	220 760 (50%)
Special schools	18 540 (13%)	14 900 (3%)
Pupil referral units (PRUs)	2 070 (1%)	6 520 (1%)
Total	146 050	447 020

Table 2 National occupational standards (NOS) for TAs: Basic concepts

- Working in partnership with the teacher
- Working within statutory and organizational frameworks
- Supporting inclusion
- Equal opportunities
- Anti-discrimination
- Celebrating diversity
- Promoting pupils' independence
- Confidentiality
- Continuous professional development

Table 3 Professional standards for higher-level teaching assistants (HLTAs)

1. <i>Professional values and practice</i>	
•	Have high expectations of all pupils; respect social, cultural, linguistic, religious, and ethnic backgrounds; commit to raising educational achievement
•	Build and maintain successful relationships with pupils; treat them consistently, with respect and consideration, with concern for their development as learners
•	Demonstrate and promote the positive values, attitudes, and behavior they expect from pupils
•	Work collaboratively with colleagues, knowing when to seek help and advice
•	Liaise sensitively and effectively with parents and carers
•	Improve their own practice, through observation, evaluation, and discussion with colleagues
2. <i>Knowledge and understanding</i>	
•	Have sufficient understanding of their specialist area to support pupils' learning; acquire further knowledge to contribute effectively and with confidence
•	Be familiar with the school curriculum, age-related expectations of pupils, and the relevant teaching methods and testing/examination frameworks
•	Understand the aims, content, teaching strategies, and intended outcomes for lessons, and understand the place of these in the overall teaching program
•	Know how to use information and communication technology (ICT) to advance pupils' learning
•	Know the key factors that affect the way pupils learn
•	Have a qualification in English/literacy and mathematics/numeracy equivalent to at least level 2 of the National Qualifications Framework.
•	Be aware of the statutory frameworks relevant to their role
•	Know the legal definition of special-educational needs (SEN), familiar with the SEN Code of Practice, and know key factors that affect the way SEN pupils learn
•	Know a range of strategies to promote good behavior and establish a purposeful and disciplined learning environment

the delivery of instruction (Giangreco *et al.*, 2001). The 2004 IDEIA and its alignment with the NCLB of 2002 compelled states and school districts to look at personnel-development programs to support their paraprofessionals. They could not use lack of funds as an excuse to avoid offering training. The NCLB Act stipulated that a "local educational agency receiving funds under this part may use such funds to support ongoing training and professional development to assist teachers and paraprofessionals in satisfying the requirements of this section." It directed that "each local educational agency receiving assistance under this part shall ensure that a paraprofessional working in a program supported with funds under this part is not assigned a duty inconsistent with this subsection." These NCLB regulations have given school districts specific obligations to provide paraprofessionals with adequate training and supervision. When an

accident, improper disclosure, or other adverse incident involving a paraprofessional leads to a lawsuit, the district may significantly reduce its potential liability if it can prove that its paraprofessionals received adequate training and supervision (Ashbaker and Minney, 2007).

There is no nationally accepted training program for newly hired paraprofessionals, leaving school districts responsible to properly train and orient their paraprofessionals. Effective instruction is the key to the use of paraprofessionals in schools. Hofmeister *et al.* (1996) suggested that the increased training and use of paraprofessionals could offset limited resources in rural schools. Morgan and Ashbaker (2001) recommended specific supervision for paraprofessionals, where they are given guided practice under the direction of an educator.

Further, Ashbaker and Wilder (2007) encouraged rural-school administrators to tap into the pool of potential teachers from among the paraprofessionals. Citing research, they argued that paraprofessionals would be good candidates for teacher training programs because they usually live in the community – suggesting staying power – they are native speakers of students' languages, and they provide a sorely needed language resource. They bring with them a great deal of classroom experience and a sense of how children learn based on how they themselves learn (Ashbaker and Wilder, 2007). Moreover, as the number of immigrant children in United States classrooms continues to grow each year, paraprofessionals expand the pool of potential teachers from underrepresented groups because they represent minority populations in greater percentages than do teachers (Haselkorn and Fidelier, 1996). Paraprofessional-to-teacher programs save school districts funds associated with lower attrition rates in teacher-education programs and among newly hired teachers (Hentschke, 1995).

Training for TAs in the United Kingdom is the responsibility of the School Workforce Development Board (SWDB). The support staff's link on the TDA website provides details of the various qualifications now available to TAs. These include national vocational qualifications (NVQs), modern apprenticeships, HLTA status, and the foundation degree. In addition, although this does not lead to a formal qualification, the United Kingdom requires (and provides materials for) induction training for TAs. **Table 4** provides a brief overview of each of these programs.

Required Skills for Paraprofessionals

For paraprofessionals to work effectively with special populations, they need the necessary skills. Due to the changing roles within the profession, the skill set needed to work effectively is also changing. In the past, skills

Table 4 Training and qualifications for TAs in the United Kingdom

All new TAs should undergo basic induction/orientation training carried out by someone in the school. In addition, the UK government now requires local education authorities (LEAs) to administer a set induction program for new TAs – ideally 4 days of training, with a strong emphasis on supporting the national literacy and numeracy strategies from reception to year 6 (K–6). The government has provided materials for the training, which is considered appropriate for both elementary and secondary levels, and all subject areas, as materials provided can be adapted to local needs, and expanded as necessary. The four modules cover the (1) TA's role in supporting teaching and learning; (2) national literacy strategy; (3) national numeracy strategy; and (4) how TAs can promote high standards of child behavior. TAs should be assigned a school-based mentor who should supervise activities undertaken during and after the induction training.

National vocational qualifications (NVQs)

NVQs (in England and Wales, Scottish vocational qualifications (SVQs) in Scotland) were developed in response to the national occupational skills, and are available at levels 2 and 3, with mandatory and optional items at each level.

For level 2, four units of competence from the NOS are mandatory (help with classroom resources and records, help with care and support of pupils, provide support for learning activities, and provide effective support for your colleagues) and three more units must be chosen from the remaining five (literacy/numeracy, management of behavior, pupil safety and security, pupil health and well-being, and ICT).

At level 3, TAs must pass ten units of competence from the NOS, four mandatory courses (management of behavior, establishing and maintaining relationships with individuals and groups, supporting pupils during learning activities, and reviewing and developing your own professional practice) and six electives.

NVQs are a work-based qualification – a TA must be in employment so that assignments and assessments can be based on current responsibilities and performance.

Modern apprenticeships

This is a qualification that has been available to TAs only since 2004. The modern apprenticeship differs from traditional apprenticeships in that there is no age limit. Apprenticeships are also a form of work-based learning, and provide training through a combination of college courses and work experience, with government funding often available to cover tuition fees and a small weekly wage paid to the apprentice by the employer. Many apprenticeships can be completed within 2 years. This combination of college and practical application on the job is intended to provide a high-level technical qualification so that new knowledge can immediately be applied to a work situation.

Modern apprenticeships use the traditional model of a partnership between a training organization and industry, with college instruction supplemented by on-the-job training. However, the modern apprenticeship is not a set course of study or specific qualification, but rather a package of qualifications appropriate to the skills needed for the chosen career. Qualifications can be gained singly and independently and through a variety of routes.

Table 4 Continued

Modern apprenticeships are available at either foundation or advanced level—foundation for TAs with limited responsibilities working under the close supervision of a teacher; advanced for TAs working under the direction of a teacher, but contributing to planning, implementing, and evaluating learning activities. The Scottish modern apprenticeship has some differences in the components because of the differences in the Scottish education system. There is no fixed time period for the apprenticeship. Typically, a foundation modern apprenticeship takes 12–15 months, an advanced modern apprenticeship about 2 years.

The higher-level teaching assistant (HLTA)

The high-level teaching assistant program came out of the National Agreement, Raising Standards and Tackling Workload of January 2003. The purpose of the program was to recognize the high level of work that many TAs were already undertaking, and to provide opportunities for TAs to take on additional roles and responsibilities to reduce teacher workloads. A TA must already be in employment to gain HLTA status; however the training component is minimal – a matter of hours only. The program has two major components:

- A portfolio of written evidence that the TA meets the standards established for HLTA status, and
- An assessment (through interview and observation) by an outside moderator or examiner.

The foundation degree (FdA)

The FdA is also a work-based training option of recent date. Most often offered on a part-time basis, with some daytime sessions but mostly twilight sessions held in the evenings or after school hours, courses are usually modular and cover such topics as learning theory, effective teaching strategies, and behavior management. Course assignments are related directly to students' current classroom situation. A part-time FdA program typically takes 3 years to complete. A major advantage of the FdA is that it can usually be counted toward a full teaching qualification. The Foundation Forward website provides further information.

associated with clerical work were needed but now skills associated with working one-on-one with children and collaboration skills are more the focus. Killoran *et al.* (2001) identified competencies that must be exhibited by paraprofessionals, namely:

- knowledge and range of child development and the impact of disabilities on children and their families;
- assessment-related confidentiality issues;
- working in partnerships;
- respect of families' culture and diversity;
- sharing information with families and assisting them to embed goals and activities;
- effective communication skills;
- ability to create an appropriate environment;
- positive social interactions;

Continued

- ability to implement and evaluate progress, to make program-data-based changes, and to use adaptive equipment/techniques;
- ability to manage/schedule daily operations and maintain safe environments;
- responsive leadership skills, collaborative teaming, ability to facilitate interpersonal and social skills, and to problem solve;
- develop and implement effective transition plans;
- use of self-evaluation for professional growth and ability to receive professional feedback; and
- adhere to identified codes of ethics.

Paraprofessionals and Instructional Process

The NCLB Act specified that a paraprofessional may be assigned to:

- provide one-on-one tutoring if such tutoring is scheduled at a time when a student would not otherwise receive instruction from a teacher;
- assist with classroom management, such as organizing instructional and other materials;
- provide assistance in a computer laboratory;
- conduct parental involvement activities;
- provide support in a library or media center;
- act as a translator; or
- provide instructional support services to students under the direct supervision of a highly qualified teacher (NCLB, Section 1119(g)(2)).

Furthermore, the NCLB Act specified the limitations of a paraprofessional's duties, stating a paraprofessional may not provide any instructional service to a student unless the paraprofessional is working under the direct supervision of a teacher consistent with Section 1119; but they are allowed to:

Assume limited duties that are assigned to similar personnel who are not working in a program supported with funds under this part, including duties beyond classroom instruction or that do not benefit participating children, so long as the amount of time spent on such duties is the same proportion of total work time as prevails with respect to similar personnel at the same school. (NCLB, 2002)

In the United Kingdom, job assignments and responsibilities are distinguished based on training. In the 2003 document, *School Support Staff: The Way Forward*, the United Kingdom distinguished four levels of TAs, according to qualifications, knowledge and skills, and the responsibilities they would therefore undertake. At levels 1 and 2, the requirements are designated as general. At levels 3 and 4, a distinction is made between those working to support and deliver learning, and those providing

behavior guidance or support. (See Table 5 for a comparison of roles, qualifications, and knowledge/skills at each of the four levels for TAs working to support and deliver learning.) As Table 5 shows, the gradual increase in knowledge/skills corresponding to increased levels of responsibility is accompanied by a decrease in intensity of supervision from a professional. At level 4, TAs not only take considerable responsibility for teaching, but also work quite independently, even supervising other TAs.

Although IDEIA stated that paraprofessionals who are providing special-education services must be adequately supervised and the NCLB Act requires that Title I paraprofessionals must work under the supervision of a teacher, there is as yet, no federal standard for what constitutes adequate supervision. Moreover, the NCLB Act and IDEIA required paraprofessionals to be appropriately trained and supervised. These are new roles for both the teacher as the supervisor and the paraprofessional as a support instructor. Oftentimes, the paraprofessional has not received prior training for instructional support and thus needs additional supervision and assistance in order to feel adequate to the task (Rueda and Monzo, 2002). This is especially important because many paraprofessionals may not know that they are supervised. Likewise, teachers may feel that they do not know how to formally supervise their paraprofessionals (Salzberg and Morgan, 1995).

The consequence of not providing supervised preparation could be disastrous; children with disabilities failing to receive a free appropriate public education (FAPE), paraprofessionals overstepping the limits of their training and authority, and teachers delegating curricular and instructional decisions – all of these actions can give ground to abounding lawsuits. By formalizing and documenting individual supervisory routines and procedures, school personnel and the district can meet its legal obligation to ensure that all personnel are adequately supervised and that properly trained staff support the educational services provided to special-education students. However, adequate supervision requires a large investment of time and energy devoted to this process. Schedules must be adjusted and resources allocated as teachers and paraprofessionals often do not have the allotted time to sit down together due to their other roles and responsibilities (Ashbaker and Morgan, 2001). Leblanc *et al.* (2005) suggested that verbal review, videotaped demonstration, performance feedback, rehearsal, and modeling can be used as methods for the teacher to provide instructional supervision and help paraprofessionals to improve their instructional skills. Some school districts and undergraduate institutions have begun to incorporate courses for teachers to learn how to manage and supervise their paraprofessionals (Trautman, 2004); and much of the supervision can be done on the job as teachers and paraprofessionals work together to support student learning (Ashbaker and Morgan, 1999).

Table 5 Four levels of roles, qualifications, knowledge, and skills required of TAs in the United Kingdom

Level 1	Level 2	Level 3	Level 4
<p>Basic role</p> <p>Work under the direct instruction of teaching/senior staff, usually in the classroom with the teacher, to support access to learning for pupils and provide general support to the teacher in the management of pupils and the classroom</p>	<p>Work under the instruction/ guidance of teaching/senior staff to undertake work/care/ support programs, to enable access to learning for pupils and to assist the teacher in the management of pupils and the classroom. Work may be carried out in the classroom or outside the main teaching area</p>	<p>Work under the guidance of teaching/senior staff and within an agreed system of supervision, to implement agreed work programs with individuals/groups, in or out of the classroom</p> <p>This could include those requiring detailed and specialist knowledge in particular areas and will involve assisting the teacher in the whole planning cycle and the management/ preparation of resources</p> <p>May also occasionally supervise whole classes during short-term absence of teacher. The primary focus will be to maintain good order and keep pupils on task. Will need to respond to questions and generally assist pupils to undertake set activities</p>	<p>Complement the professional work of teachers by taking responsibility for agreed learning activities under an agreed system of supervision</p> <p>This may involve planning, preparing, and delivering learning activities for individuals/groups or short term for whole classes, and monitoring pupils and assessing, recording, and reporting on pupils achievement, progress and development</p> <p>Responsible for the management and development of a specialist area within the school and/or management of other teaching assistants including allocation and monitoring of work, appraisal, and training</p>
<p>Qualifications</p> <p>-Good numeracy/literacy skills</p>	<p>-Good numeracy/literacy skills</p>	<p>-Very good numeracy/literacy skills</p>	<p>-Excellent numeracy/literacy skills (maths and English equivalent to NVQ Level 2)</p>
<p>-Completion of TA induction program</p>	<p>-Completion of TA induction program</p>	<p>-NVQ 3 for TAs or equivalent qualification or experience</p>	<p>-Meet HLTA standards or equivalent qualification or experience</p>
<p>-Participate in development and training opportunities</p>	<p>-NVQ 2 for TAs or equivalent qualifications/experience</p> <p>-Training in the relevant learning strategies e.g., literacy</p> <p>-First aid as appropriate</p>	<p>-Training in relevant strategies e.g., literacy, and/or in particular curriculum or learning area e.g., bilingual, sign language, dyslexia, ICT, math, English, CACHE</p> <p>-Appropriate first aid training</p>	<p>-Training in relevant learning strategies e.g., literacy</p> <p>-Specialist skills/training in curriculum or learning area e.g., bilingual, sign language, ICT</p>
<p>Knowledge and skills</p> <p>-Appropriate knowledge of first aid</p>	<p>-Effective use of ICT to support learning. Use of other equipment technology – video, photocopier</p>	<p>-Use ICT effectively to support learning. Use of other equipment technology – video, photocopier</p>	<p>-Use ICT effectively to support learning</p>
<p>-Use basic technology – computer, video, photocopier</p>	<p>-Understanding of relevant policies/codes of practice and awareness of relevant legislation</p>	<p>-Full working knowledge of relevant policies/codes of practice and awareness of relevant legislation</p>	<p>-Full working knowledge of relevant policies/codes of practice/legislation</p>
<p>-Ability to relate well to children and adults</p>	<p>-General understanding of national/foundation stage curriculum and other basic learning programs/strategies</p>	<p>-Working knowledge of national/foundation stage curriculum and other relevant learning programs/strategies</p>	<p>-Working knowledge and experience of implementing national/foundation stage curriculum and other relevant learning programs/ strategies. Understanding of statutory frameworks relating to teaching</p>

Continued

Table 5 Continued

Level 1	Level 2	Level 3	Level 4
-Work constructively as part of a team, understanding classroom roles/responsibilities and own position within these	-Basic understanding of child development and learning -Ability to self-evaluate learning needs and actively seek learning opportunities -Ability to relate well to children and adults -Work constructively as part of a team, understanding classroom roles and responsibilities and own position within these	-Understanding of principles of child development and learning processes -Ability to self-evaluate learning needs and actively seek learning opportunities -Ability to relate well to children and adults -Work constructively as part of a team, understanding classroom roles and responsibilities and own position within these	-Good understanding of child development and learning processes -Constantly improve own practice/ knowledge through self-evaluation and learning from others -Ability to relate well to children and adults -Work constructively as part of a team, understanding classroom roles and responsibilities and own position within these -Ability to organize, lead, and motivate a team

CACHE – Council for Awards in Children's Care and Education.

Conclusion

Paraprofessional-to-teacher programs save school districts funds associated with lower attrition rates in teacher-education programs and among newly hired teachers (Hentschke, 1995). Paraprofessional training programs must be research-based programs that produce a highly skilled paraprofessional workforce. Articulation agreements must be forged at the local and state levels, as well as reciprocal agreements across states that recognize the quality of the training and the skills gained. Pickett *et al.* (2006) at the National Resource Center for Paraprofessionals (NRCPP) noted that although there has been some progress in developing policies and standards, in far too many situations, these efforts have been piecemeal and have not led to infrastructures and policies that are integral parts of statewide systems of personnel development. Similar conditions are visible in the United Kingdom where there are now moves to standardize pay and training for paraprofessionals. It is when good initiatives are developed along with a strong infrastructure that paraprofessionals will receive the recognition and value that they deserve. This must happen with the overarching goal of improving the quality of educational services for students with special needs and their families.

See also: Apprenticeship Approach to Learning; Continuing Professional Development of Teachers; Co-Teaching; Early Childhood Education and Care Staff Support; Mentoring and Professional Development in Finland; Evaluating Schools as Learning Communities; Professional Development of Teacher Educators; Professional Learning Community; Scotland; Situated View of Learning; Supervision in Teacher Education; Teacher Training and Preparation in the United States; USA; Workforce Issues in Special Education.

Bibliography

- Ashbaker, B. Y. and Minney, R. B. (2007). *Planning Your Paraprofessionals' Path: An Administrator's Legal Compliance and Training Guide*. Horsham, PN: LRP Publications.
- Ashbaker, B. Y. and Morgan, J. (1999). The "S" in ASCD: Teachers supervising paraeducators for professional development. *Paper Presented at the Annual Meeting of the Association for Supervision and Curriculum Development*, San Francisco, CA.
- Ashbaker, B. Y. and Morgan, J. (2001). Paraeducators: A powerful human resource. *Streamlined Seminar* 19(2), 1–4. (Quarterly newsletter of the National Association of Elementary School Principals - NAESP.)
- Ashbaker, B. Y. and Morgan, J. (2006). *Paraprofessionals in the Classroom*. Boston, MA: Allyn and Bacon.
- Ashbaker, B. Y. and Wilder, L. K. (2007). Changes in administering special education programs. In Obiakor, F. E., Rotatori, A. F., and Burkhardt, S. (eds.) *Current Perspectives in Special Education Administration*, pp 71–86. Oxford: Elsevier.
- Blalock, G. (1991). Paraprofessionals: Critical team members in our special education programs. *Intervention in School and Clinic* 36, 200–214.
- Broer, S. M., Doyle, M. B., and Giangreco, M. F. (2005). Perspectives of students with intellectual disabilities about their experiences with paraprofessional support. *Exceptional Children* 71(4), 415–430.
- Chopra, R. V., Sandoval-Lucero, E., Aragon, L., *et al.* (2004). The paraprofessional role of connector. *Remedial and Special Education* 25(4), 219–231.
- Daniels, V. I. and McBride, A. (2001). Paraeducators as critical team members: Redefining roles and responsibilities. *National Association of Secondary School Principals Bulletin* 85(623), 66–74.
- DfES (Department for Education and Employment) (2000). *Supporting the TA: A Good Practice Guide*. London: DfES.
- Education for All Handicapped Children Act (EAHCA) of 1975, Pub. L. No. 94-142.
- French, N. K. (1999). Paraeducators and teachers: Shifting roles. *Teaching Exceptional Children* 32(2), 69–73.
- Giangreco, M. F., Edelman, S. W., Broer, S. M., and Doyle, M. B. (2001). Paraprofessional support of students with disabilities: Literature from the past decade. *Exceptional Children* 68(1), 45–63.
- Hampden-Thompson, G., Diehl, J., and Kinukawa, A. (2007). *Description and Employment Criteria of Instructional Paraprofessionals*. <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007008> (accessed August 2009).

- Haselkorn, D. and Fidler, E. (1996). *Breaking the Glass Ceiling: Paraeducator Pathways to Teaching*. Belmont, MA: Recruiting New Teachers.
- Hentschke, G. C. (1995). *Increasing Minority Representation in Teacher Education Programs: Funding Issues for Us to Consider (Minority Teacher Policy Forum)*. Los Angeles, CA: University of Southern California, Latino Teacher Project.
- Hoffman, L. and Sable, J. (2006). *Public Elementary and Secondary Students, Staff, Schools, and School Districts: School Year 2003-04, E.D. TAB. NCES 2006-307*. Washington, DC: National Center for Education Statistics/Education Statistics Services Institute.
- Hofmeister, A. M., Ashbaker, B. Y., and Morgan, J. (1996). Paraeducators: Critical members of the rural education team. In rural goals 2000: Building programs that work. *16th Annual National Conference Proceedings of the American Council on Rural Special Education (ACRES) (ERIC Document Reproduction Services No. ED 394 757)*, Baltimore, MD.
- Individuals with Disabilities Education Improvement Act (IDEIA) of 2004, Pub. L. No. 108-446.
- Killoran, J., Templeman, T. P., Peters, J., and Udell, T. (2001). Identifying paraprofessional competencies for early intervention and early childhood special education. *Teaching Exceptional Children* **34**(1), 68-73.
- Leblanc, M., Ricciardi, J. N., and Luiselli, J. K. (2005). Improving discrete trial instruction by paraprofessional staff through an abbreviated performance feedback intervention. *Education and Treatment of Children* **28**(1), 76-82.
- Minondo, S., Meyer, L. H., and Xin, J. F. (2001). The role and responsibilities of teaching assistants in inclusive education: What's appropriate? *Journal of the Association for Persons with Severe Handicaps (JASH)* **26**(2), 114-119.
- Morgan, J. (2007). *How to Be a Successful Teaching Assistant*. London: Continuum International.
- Morgan, J. and Ashbaker, B. Y. (2001). *A Teacher's Guide to Working with Paraeducators and Other Classroom Aides*. Alexandria, VA: Association for Supervision and Curriculum Development.
- National Joint Council for Local Government Services (2003). *School Support Staff: The Way Forward*. <http://www.unison.org.uk/file/A4406.pdf> (accessed August 2009).
- NCLB (2002). No Child Left Behind Act (NCLB) of 2001, Pub. L. No. 107-110 section 6319.
- Pickett, A. L., Likins, M., and Wallace, T. (2006). *Developing Strategies*. The National Resource Center for Paraprofessionals. <http://www.nrcpara.org> (accessed August 2009).
- Rueda, R. and Monzo, L. D. (2002). Apprenticeship for teaching: Professional development issues surrounding the collaborative relationship between teachers and paraeducators. *Teaching and Teacher Education* **18**(5), 503-521.
- Salzberg, C. L. and Morgan, J. (1995). Preparing teachers to work with paraeducators. *Teacher Education and Special Education* **18**, 49-55.
- Trautman, M. L. (2004). Preparing and managing paraprofessionals. *Intervention in School and Clinic* **39**(3), 131-138.

Relevant Websites

- <http://www.aft.org> – American Federation of Teachers (AFT), Paraprofessionals and School Related Personnel (PSRP).
- <http://www.asha.org> – American Speech-Language-Hearing Association (ASHA).
- <http://www.cec.sped.org> – Council for Exceptional Children, Performance Based Standards for Paraeducators.
- <http://www.reading.org> – International Reading Association.
- <http://www.lge.gov.uk> – Local Government Employers (LGE), School Support Staff – The Way Forward, 2006.
- <http://www.nea.org> – National Education Association (NEA), Education Support Personnel (ESP).
- <http://www.nrcpara.org> – National Resource Center for Paraprofessionals (NRCP).
- www.tda.gov.uk – Training and Development Agency for Schools.

Parent and Family Involvement in the Education of Children with Special Needs

L P Prater, Southeast Missouri State University, Cape Girardeau, MO, USA

© 2010 Elsevier Ltd. All rights reserved.

The education of any child is a very complex issue, comprising interconnected internal and external relationships (Garrett and Morgan, 2002). This is especially true when the child has been identified as one with special needs, primarily because of the wide variance of those needs and abilities. Educating these children is much more than molding a piece of clay, or drawing on a blank canvas, as some would describe child socialization. There are so many variables that impact how knowledge is imparted from one human to another, even if there are no special needs identified. In fact, the phenomenon called education is rather mysterious. When does learning begin? How is knowledge gained? By definition, who is an educated person? Who is responsible for educating children? Can all children learn? These are merely a few of the questions one might ask. In attempting to respond to these questions, it must be recognized that all children are unique. There are some children who are more different than others and have been identified as having special needs, constituting a wide array of behaviors, ability, and performance levels, as well as challenges.

In an effort to match the special needs of children with appropriate services, labels have been devised to facilitate this process. There are advantages and disadvantages to this strategy, especially if children are labeled inaccurately or if there are political motivations involved. For example, there are instances in which children with special needs are labeled differently depending on where they live. In more affluent areas, more children are labeled as learning disabled, rather than mentally retarded. These labels matter because they are connected to specified services, sometimes including fiscal support. For instance, there is currently much attention and grant opportunities focused on autism. The critical question could be asked, "Are there more students with autism or are more students, who would have received other labels a few years earlier, being labeled as autistic?" This is why teachers, service providers, and parents should be knowledgeable about the values of parent and family involvement. Children, and especially those with special needs, are educated within a societal system and not in isolation. This is a three-prong system composed of family, educators, and, collectively, other societal constituents such as social service agencies and policymakers. As with any three-legged unit, all pieces are necessary for the structure to stand. The family, in concert with educators and

others who impact this process, is a significant factor in educating children with special needs.

The late social scientist Urie Bronfenbrenner (1977) developed a classic systems model that aligns with internal and external variables which impact family functioning. In that model, the family ecosystem is divided into subsystems of micro-, meso-, exo-, and macro-levels. The family unit and school are at the micro- and meso-levels, respectively, and have regular interaction with each other. Social service agencies and policymakers are at the exo-level and attitudes, beliefs, and culture operate at the macro-level. This family systems approach should be used in educating students with exceptionalities, because these children and their families are very much interconnected with other groups. In this article, the focus is on the significant micro- and meso-subsystems as the systems components within which the family and school interact to accomplish the goal of educating children with special needs.

Identity of Family Members

In contemporary times, many forms of relationships are recognized as representative of a family. Traditionally, in some cultures, a family unit consisted of a legal union between a man and a woman and their biological children. That definition has been greatly expanded. Persons may identify themselves as family members if they are related by blood, marriage, or self-imposed affiliation. They may cooperate economically and may or may not share a common household, such as extended family members, and may not have children. If children are in the household, they may be their biological children, step-children, foster children, or adopted. In fact, the least common family structure today mirrors the traditional model. For example, in America, the fastest growing family structure is single females with children. When children with special needs are part of a single-parent household, parenting responsibilities can become overwhelming. Family structure may, in part, explain the educational outcomes of these children. There may not be time and energy for one person to work, care for the household, assist with educational programming at home, and actively participate in school functions. In addition, many of these single parents, who are mostly mothers, have parenting responsibilities with other children as well.

The list of ways that families of children with special needs differ is unlimited. To name a few, families differ in family size, socioeconomic status, number and educational level of parents as primary caregivers, ethnicity, and in the primary language spoken in the household. All of these factors influence the likelihood of family involvement with schools. Increasingly, schools are asked to serve these diverse families and accomplish parity of educational outcomes, regardless of the frame of reference with which the child enters.

Unique Circumstances Impacting Parental Participation in Schooling

It is assumed that all potential parents want the perfect child, the one who is perfectly formed, perfectly behaved, and perfectly functioning in every way. In reality, some children are born with mild, moderate, and severe congenital disabilities. Circumstances may cause children born, without these differences, to acquire physical or mentally limiting conditions. Illness or accidents are most likely the culprits in those cases. Parents may experience chronic sorrow, grief, and mourning for the loss of their ideal child (Winkler *et al.*, 1981). These feelings can recur at transitional periods such as the child's entrance into school or onset of puberty. Parents may become engulfed with bitterness or anger and wonder why this happened to them. Although unintentional, they may become consumed with jealousy of families with children identified as normal. Does society prepare families to parent a special needs child? No, there is no proactive socialization process to prepare parents for this situation. Possibly, this is one reason that the divorce rate is high for couples with a special-needs child. The caregiving demands are high and stressful, and can lead to increased feelings of depression, helplessness, guilt and blame, and decreased feelings of parental competence.

Parenting is a tough and important role, maybe the most challenging in all of society. When this responsibility is coupled with parenting a child with special needs, involvement in discretionary school activities may be less of a priority. Other than the legislated involvement and meetings, or school contact initiated as a result of reported problems, these parents tend to be uninvolved. The time within their 24 h in a day, 7 days in a week, and 12 months in a year are occupied with the daily challenges of financing the needs of their child; managing a myriad of unpredictable and atypical circumstances; navigating the related legal and social systems; and handling the resulting family-relationship issues. Families experience stressful situations unknown to other families. The anticipated daily family functioning of their dreams remains a fantasy. It is amazing that they function as well as they do. For

example, some families, who parent one of the children diagnosed with autism, have experienced financial ruin in trying to secure and finance services for their child. If it is a two-parent household, one parent may have to remain outside of the labor force to care for the child's needs. This further reduces the family income! The possible genetic influence may cause some families to have more than one child with autism, especially if the siblings are twins. Government assistance is very specified and limited, if it exists at all for children with autism in some communities. In countries with socialized medicine programs, this may not be a problem. In the United States, there is no entitlement for universal healthcare. For some families, there may be government-subsidized care; in other situations, care is dependent upon the family's ability to fiscally support the health needs of the child. It must also be recognized that many of these families have typically developing children to support as well. Rather than criticize these families for the low participation in school activities, educators, social agency representatives, and policymakers should be more sensitive to the inadequacy of support that is needed for these families, and contribute to advocacy initiatives to provide more assistance.

The typical school day conflicts with the work schedule of many parents, including parents of special needs children. This fact is significantly a barrier for families of low income and those with inflexible positions that prohibit leave time to visit school settings. For these parents, this is a double disadvantage. They cannot afford absence from work to participate in many of the school functions and they are judged as being apathetic, uncaring, or uninterested with regard to their child's schooling. There are fewer high-income families and, therefore, as compared to others, many of these families are within the range of poverty, which presents a multitude of circumstances which serves as a barrier to their participation in school activities.

Children with special needs are members of families from all social classes. Parents at all economic levels should perceive that they have the opportunity for participation in some form of school activity, but social class is an issue. The social stratification results in many of the parents of the same income having children clustered in the same school. In schools where a large majority of the children are from families with low income, there is low parental participation. Conversely, children from families with high income, tend to attend the same schools. As their parents are more likely to have professional careers with flexibility of schedules, their participation in school events is higher. The attention that these parents give to the child and school interaction is not a signal that they love their child any more than parents with inflexible work schedules. It does signify that there is more of an opportunity for them to be involved. Being concerned about a child's education is not a function of income level.

Family members of children with special needs, like everyone else, are impacted by past experiences. Memories of the school experiences of family members influence their attitudes and perceptions about school. Those with positive experiences tend to be more open to school interaction, than family members with memories of negative events during their own school years. Similar to any other situation, persons tend to avoid unpleasant situations and embrace positive ones. There are parents, especially those with less education than school officials, who feel intimidated by school personnel. Insecurity, inferiority, and embarrassment are emotions they feel when in the presence of teachers or school administrators. Their response to avoiding these feelings is merely to limit school visits to mandated appearances. Overall, parental involvement is a result of the family's perceived skills and abilities, their employment and other obligations, and opportunities provided by the school.

Family Participation in Schooling Children with Special Needs

A primary responsibility of schools is to educate students, but students are connected to some part of a family unit, in which most of their time is spent. Family members are actually the first teachers. Collectively, considering the child's total being of time and influence, school is dwarfed by home and family impacts. There is a role in educating students that is occupied only by family members and cannot be replaced by school faculty, administrators and staff, school board members, or legislators. School personnel and others making important decisions about schooling may have degrees from the best colleges and universities, but lack crucial information known only to family members.

The school needs to take the lead on whether the parent will be afforded only a passive role, or encouraged to assume an active role. The participation has to go beyond the traditional model of parent-teacher association (PTA), established in 1924, and the school climate must be welcoming and encouraging of parental involvement. To accommodate this involvement, schools can structure opportunities for parents to participate in schooling (Epstein, 1986). There are some situations in magnet schools in which family members are required to volunteer a specified number of hours at the school. As mentioned earlier, some level of involvement is mandated for parents of children with special needs. In the United States, the Individuals with Disabilities Education Act (IDEA) of 1990, Public Law 101-476 provided the opportunity for the parents of children with special needs to influence the education of their children. Parents of children in special-education programs are legally required to

participate in the child's individualized educational plan. Public Law 101-476 and its 1997 amended version, especially addressed parents and educational resources for minority children (Prater and Ivarie, 1999).

For many parents, they are invited to the school to assist the teacher in the classroom, raise funds, chaperone field trips, or possibly help coach a ball team. For parents of special-needs children, the involvement tends to be more focused. The educational process can be a 7-days-a-week, 24-hours-a-day process, depending on the nature of the birth or acquired differences. The parent and teacher partnership is much more significant than baking cookies for a class party. Of course, parents often help with homework, but parental assistance with homework for a special-needs child is more intense, more focused, and more than likely, more time consuming.

Parents who can navigate the political and educational system are empowered to get the best services for their children. Schools that seek to empower parents assume that parents are capable of influencing the outcome of their child's educational process.

However, sometimes there is a disconnect between families and school personnel. Society is becoming more diverse, not less diverse. Considering that families reside within a global society, this diversity includes racial, ethnic, language, and cultural differences, but must also include differences in cognitive, emotional, and physical functioning. When there were more homogeneous school populations and neighborhood schools, empowerment of parents did not seem to be a concern. The incidence of children in need of special education was not much of an issue, because children with special needs were not sent to school. They were mostly kept at home and shielded from public view.

Family involvement is a certainty for parents of children in special-education programs, if only indicated by parents signing a consent form, rather than actively participating in policymaking. This formal, one-way communication, from school to home is limiting and does not empower parents. When families are empowered, they feel that their views have value. When families are not empowered, there is a feeling of powerlessness, followed by intimidation that prohibits them from asking important questions on their child's behalf. To empower parents, parental involvement should include roles such as advocates and policymakers.

School personnel must view each child and family within a framework that encompasses the entire political, social, economic, cultural, and spiritual experiences that shape the identity and behavior of the families and children with special needs. For more than any other population of school-age children, the one-size-fits-all approach is inappropriate, which is exactly why the individualized educational program emerged. For these children and families, there are endless variances.

Teachers, like everyone else, may tend to avoid situations in which they perceive rejection as the outcome. Sensing negative reactions from parents, there could be instances in which teachers are more passive in seeking parental involvement because they fear an unsuccessful result. Some of this fear could be based on biased pre-judgments, myths, or stereotypes couched within Urie Bronfenbrenner's macro-level inclusive of attitudes, beliefs, and cultures. In these circumstances, parents can sense the feelings of teachers and reciprocate with feelings of isolation, alienation, and disengagement.

In establishing a support system for parent involvement, school personnel should become familiar with the child's uniqueness, the family, and the school's community to determine those activities that are more likely to solicit the desired outcome of parental involvement. As parents are afforded a greater role in defining involvement, they will gain greater confidence leading to increased participation in educational decision making focused on their child with special needs. Aside from meetings to discuss the individualized educational plan, family members have few avenues by which they can challenge curriculum choices, instructional strategies, or additional educational resources for their child.

In too many situations, educators continue to exclude families, devalue their suggestions, primarily use the ideas of experts, and value more the opinions of those who have possibly never even met the students. Prater and Tanner (1995) completed a case study report that detailed the tragedy of such practices, and the subsequent result of sabotaging parental aspirations and curtailing educational opportunities for youth in special-education programs. It is a mistake to rely solely on the ideas of experts and ignore the real people most knowledgeable about the child, the families. It is vital that parents are empowered and parental involvement is integrated into the instructional plans for students in special education, beginning with early childhood special education and continuing throughout the educational process.

Creating a School Environment to Empower Families

There is a mistaken notion that involvement and empowerment are synonymous terms. Involvement may merely imply some form of contact, but empowerment is more of a reciprocal partnership that includes decision making and influences outcomes. In order for parents to be empowered in special-education programs, they must assume roles as assessors, presenters of reports, policy-makers, advocates, and peer supporters. Educators will not be a vehicle for empowering parents unless they respect the fact that parents are themselves empowered

with an understanding of the system within which families of children with special needs exist and function. Coupled with educators' need to understand, is the need to avoid stereotyping families based on a presumption of uniformity of experience, culture, and knowledge. Often, there is cultural and educational dissonance between school governance professionals and the families they serve. If there are no deliberate attempts to eliminate, or at least bridge these poles, a barrier develops that will cause families to avoid school participation. This cultural divide is very evident in America, where school-governing bodies tend to be male, White, and middle class. By contrast, minorities comprise the majority of children enrolled in special-education programs.

The school environment can facilitate or impede successful relationships between the school and families. Perceptions based on gender, sexism, and social stratification can reinforce dissonance between parents and teachers, causing some educators to view parents as the cause of problems with schooling their children. These perceptions marginalize parents and inhibit parental empowerment, and discourage involvement. There must be trust between families and school personnel. Without trust, little will be accomplished. To benefit the educational goals for the child, this trust must be mutual and must be earned to be acquired.

One might ask, How can educators earn this trust? For parents to trust school personnel, there must be a consistent pattern of positive experiences, resulting in a strong foundation of credibility rooted in honesty and practices targeted to the sincere well-being of the child. For whatever the reason, this does not always occur. If a child is misdiagnosed, misplaced, mislabeled, or miseducated, trust is compromised. If there are unanswered questions as to why certain groups are overrepresented in remedial programs and rarely in gifted programs, it may be difficult to accomplish a comfortable level of trust. One might assume that because a child is assigned to a special-education classroom that he or she is appropriately placed. This might not necessarily be the situation. As described previously in Bronfenbrenner's systems model, the macro-level comprises attitudes and beliefs. Despite their academic credentials, educators have personal biases which serve to influence behavior in their professional activities.

Benefits of Family and School Partnerships

Regardless of the challenges, schools must seek to engage parents. There are positive outcomes involved in reaching out to parents. By any measure, students in special-education programs benefit from the partnership. The academic outcome is more positive and the availability of the joint resources provides additional assistance. Time and

energy lost in disruptions between the family and school are costly, with the child being the biggest loser. With the help of school personnel, parents who can navigate the political, educational, and social systems are empowered to get the best services for their children. Unfortunately, some parents do not even know their rights. When parents know the laws and have political expertise, they can create happenings. For example, in one mid-Western town in the United States, parents lobbied state legislators to get millions of state money allocated to serve children with autism in their district. This occurrence would not have happened without parental collaboration and intervention.

Summary

An effective relationship between family and school impacts school adjustment and progress. As there are so many variables in the education of children with special needs, all opportunities must be explored to create and sustain effective partnerships between families and the schools, even those options that have never been used before, or may seem atypical. To address the fact that families are different, there should be more development of projects targeted to specific groups, such as grandparents raising grandchildren, single fathers, foster parents, and other groups of identifiable populations parenting children with special needs.

Institutions preparing teachers must include sensitivity to the role of parents in their teacher-training curriculum, because current and future educators must empower parents. This sensitivity must include knowledge of multicultural factors. Otherwise, it may be assumed that a culture different from the educator implies a lower status of cognitive functioning and educational deficits for the children and families. Currently, school professionals are not prepared for these diverse families.

If parents are not empowered, they cannot and will not advocate for the child. Intimidation may prohibit them from asking important questions or seeking social service support. More importantly, a systemic network of negative attitudes will serve to victimize parents and retard the likelihood of parents establishing a partnership in the education of their children. Empowered parents ask questions and insist on accountability.

The model used by schools for involving parents is outdated. It has not changed, but the family structures and lifestyles have changed. Rather than expecting parents and the child to modify their behavior, it is imperative that educators modify their behavior and attitudes toward students in special education and their families. The complexity of family functioning must be considered in order for families to be repositioned as active, equal partners in schooling. Educational systems nurturing special-education programs must establish a web of support

for its students. Students, parents, educators, employers, lawmakers, and other community entities must be totally focused on an integrated goal of educating all of the students with special needs.

See also: Child Rearing and Early Education: Parents and Professionals: Theoretical and Historical Influences; Educating Students with Special Needs: An Overview; Parent Support in Early Childhood – Approaches and Outcomes.

Bibliography

- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist* **32**, 513–531.
- Epstein, J. (1986). Parents' reactions to teacher practices of parent involvement. *Elementary School Journal* **86**, 277–294.
- Garrett, J. E. and Morgan, D. E. (2002). Celebrating diversity by educating all students: Elementary teacher and principal collaboration. *Education* **123**, 268.
- Prater, L. and Ivarie, J. (1999). Empowering culturally diverse parents in special education programs. In Obiakor, F. E., Schwenn, J. O., and Rotatori, A. F. (eds.) *Advances in Special Education: Multicultural Education for Learners with Exceptionalities*, pp 149–166. Stamford, CT: JAI Press.
- Prater, L. P. and Tanner, M. P. (1995). Collaboration with families: An imperative for managing problem behaviors. In Obiakor, F. E. and Algozzine, B. (eds.) *Managing Problem Behaviors: Perspectives for General and Special Educators*, pp 178–206. Dubuque, IA: Kendall/Hunt.
- Winkler, L., Wasow, M., and Hatfield, E. (1981). Chronic sorrow revisited: Parent vs. professional depiction of the adjustment of parents to mentally retarded children. *Journal of Orthopsychiatry* **51**, 63–70.

Further Reading

- Acock, A. C. and Kiecolt, J. K. (1989). Is it family structure or socioeconomic status? Family structure during adolescence and social adjustment. *Social Forces* **68**, 553–571.
- Becker, H. and Epstein, J. (1982). Parent involvement: A survey of teacher practices. *Elementary School Journal* **83**, 86–102.
- Bianchi, S. M. (1995). The changing demographic and socioeconomic characteristics of single parent families. *Marriage and Family Review* **20**, 71–97.
- Bracey, G. W. (2001). School involvement and the working poor. *Phi Delta Kappan* **82**(10), 795–797.
- Bright, J. A. (1996). Partners: An urban black community's perspective on the school and home working together. *New Schools, New Communities* **12**, 32–37.
- Bristol, M. M., Gallagher, J. J., and Schopler, E. (1988). Mothers and fathers of young developmentally disabled and nondisabled boys: Adaptation and spousal support. *Developmental Psychology* **60**(3), 441–451.
- Dunst, J. (2002). Family-centered practices: Birth through high school. *Journal of Special Education* **36**(3), 139–147.
- Epstein, J. (1991). Effects on student achievement of teachers' practices of parent involvement. In Silvern, S. (ed.) *Literacy through Family, Community and School Interaction*, pp 24–36. Greenwich, CT: JAI Press.
- Friedrich, W., Wiltner, L., and Cohen, D. (1985). Coping resources and parenting mentally retarded children. *American Journal of Mental Deficiency* **90**, 130–139.

- Gartland, D. and Strosnider, R. (2007). Learning disabilities and young children: Identification and intervention. *National Joint Committee on Learning Disabilities Quarterly* **30**, 63–72.
- Gavin, K. M. and Greenfield, D. B. (1998). A comparison of levels of involvement for parents with at-risk African American kindergarten children in classrooms with high versus low teacher encouragement. *Journal of Black Psychology* **24**, 403–417.
- Harry, B. (1992). Restructuring the participation of African-American parents in special education. *Exceptional Children* **59**, 123–131.
- Harry, B., Allen, N., and McLaughlin, M. (1995). Communication versus compliance: African-American parents' involvement in special education. *Exceptional Children* **61**, 364–377.
- Henderson, A., Marburger, C., and Ooms, T. (1986). *Beyond the Bake Sale: An Educator's Guide to Working with Parents*. Columbia, MD: National Committee for Citizens in Education (ED 270 508).
- Kirk, S. A. and Gallagher, J. J. (1986). *Educating Exceptional Children*. Boston, MO: Houghton Mifflin.
- Mulkey, L. M., Crain, R. L., and Harrington, A. J. (1992). One-parent households and achievement: Economic and behavioral explanations of a small effect. *Sociology of Education* **65**, 48–65.
- Muller, C. (1993). Parent involvement and academic achievement: An analysis of family resources available to the child. In Schneider, B. and Coleman, J. S. (eds.) *Parents, Their Children, and Schools*, pp 77–114. Boulder, CO: Westview.
- Obiakor, F. E. (1994). *The Eight-Step Multicultural Approach: Learning and Teaching with a Smile*. Dubuque, IA: Kendall/Hunt.
- Oswald, D. P., Coutinho, M., and Best, A. M. (1999). Ethnic representation in special education: The influence of school-related economic and demographic variables. *Journal of Special Education* **32**, 194–206.
- Reynolds, A. J. (1992). Comparing measures of parental involvement and their effect on academic achievement. *Early Childhood Research Quarterly* **7**, 441–462.
- Spann, S. J., Kohler, F. W., and Soenksen, D. (2003). Examining parents' involvement in and perceptions of special education services: An interview with families in a parent support group. *Focus on Autism and Other Developmental Disabilities* **18**(4), 228–237.
- Taylor, A. R. and Machida, S. (1994). The contribution of parent and peer support to head start children's early adjustment. *Early Childhood Research Quarterly* **9**, 387–405.
- Townsend, B. L. and Patton, J. M. (2000). Reflecting on ethics, power, and privilege. *Teacher Education and Special Education* **23**, 32–33.
- Turnbull, A. P. and Turnbull, H. R. (1997). *Families, Professionals, and Exceptionality: A Special Partnership*, 3rd edn. Columbus, OH: Merrill.

Peer Relations and Socialization of Children and Adolescents with Special Needs

L M Bullock, University of North Texas, Denton, TX, USA

M Wong-Lo, University of North Texas, Denton, TX, USA

R A Gable, Old Dominion University, Norfolk, VA, USA

© 2010 Elsevier Ltd. All rights reserved.

According to the literature, children who are socially competent, experience positive social interpersonal relationships, peer acceptance, satisfactory school adjustment, and cope with and adapt to the demands of various social settings (Gresham *et al.*, 2003). In the absence of these prosocial skills, students suffer – both socially and academically (Bullock and Gable, 2001; Kerr and Nelson, 2006; Schoenfield *et al.*, in press). Indeed, a defining characteristic of children with emotional/behavioral disorders is the inability to build or maintain satisfactory interpersonal relationships (Kauffman, 2003; Kerr and Nelson, 2006). Between 35% and 75% of children with learning disabilities exhibit problems establishing and maintaining friendships (Kavale and Forness, 1996). In addition, a significant number of individuals with cognitive impairments show social skill deficits (Strain and Kerr, 1981). With the disillusionment of a two-box system of public education – one box for general education and the other for special education students, there has been mounting attention given to peer relations and social development of students with special needs.

We have long recognized the critical role that peer relationships play in children's overall socialization. To better understand that developmental process and how that knowledge can improve prevention/intervention efforts, researchers have examined the role peer relationships play in the development of both adaptive and maladaptive patterns of behavior (e.g., Deater-Deckard, 2001; Santos and Winegar, 1999). This growing body of research has increased substantially our understanding of the complex nature of social interpersonal relationships among children and adolescents – with and without special needs. Notwithstanding the contributions of these investigations, a myriad of questions remain regarding the nuances of children's interpersonal relationships (Ellis and Zbaratany, 2007) and the socialization process as it relates to individuals with special needs (Larrivee and Horne, 1991).

In this article, we examine peer relations and socialization of children/adolescents with special needs. To accomplish our goal, we conducted a review of the accumulated literature by means of the search engine ERIC and Academic Search Premier using the keywords: peer relations, socialization, and special education. Furthermore, we used the Internet search engine Google and the same keywords to obtain additional information.

In addition, we conducted a manual search of textbooks and professional journals. Finally, we made use of various reference lists to complete our literature review on peer relations and socialization of students with disabilities.

Past-to-Present Perspectives on Peer Socialization

Beginning in the 1920s, child development specialists were among the first to focus attention on social development and the role of peers in the socialization process (Haynie and Osgood, 2005). For example, Piaget (1932) posed that peer experiences contribute uniquely to children's social development and what they might reasonably expect from peer relationships. Hartup (1977, 1979) asserted that children who did not experience positive encounters with peers are developmentally at risk. Similarly, Piaget (1978) maintained that children who fail to learn from early social interactions must either learn to alter their behavior or risk peer rejection. Subsequently, a series of longitudinal studies confirmed these assumptions and documented the fact that social interpersonal problems from childhood to late adolescence are inextricably linked to early manifestations of peer-group problems and to later social maladjustment (McCord *et al.*, 1963; Roff *et al.*, 1972). In one such study, Roff *et al.* found that poor relations in the primary grades were highly predictive of adolescent delinquency and mental health problems in adulthood. Others also have documented the long-term deleterious effects of isolation and social withdrawal during childhood (e.g., Kauffman 2003; Kerr and Nelson, 2006; Strain and Kerr, 1981).

Although the underpinning for peer relationships starts with parents and other caregivers, the task of adjusting to the peer group begins in preschool (Read, 1960; Ross *et al.*, 1990). Not surprisingly, early-childhood educators attach major significance to social development and the socialization experiences that occur in these early peer-group interactions (Forness *et al.*, 1981). For instance, Read looked closely at nursery school education and its contributions to understanding the basics of human behavior and the socialization process. Read noted that early-childhood education is a place where adults learn much from their observations of children, as they interact with them and, most importantly, provide them with a setting that nurtures

interpersonal relationships. Fortunately, the majority of preschoolers thrive in such an environment and make gains in such disparate areas as critical thinking, expressive and receptive language, problem solving, and prosocialization. However, from the beginning, some children struggle in one or more of these areas, to the degree that they stand apart from their normally developing classmates.

Cook and Wollersheim (1976) explored individual differences among children to the degree to which some are labeled as disabled. They highlighted the potential negative consequences of labeling students, but noted the fact that positive social contacts can engender more positive attitudes and increase the level of peer acceptance. Earlier in their seminal work, Hartup *et al.* (1967) demonstrated that positive social interactions could ameliorate many emergent social interpersonal problems. In the same dimension, Morrison and Borthwick (1983) investigated the relationship between patterns of classroom behavior and the relative social status of children with special needs. They found that classmate's perspectives on cognitive and behavioral performance contribute differentially to social acceptance versus peer rejection. Across time, researchers have examined numerous factors that relate to the overall social adjustment of children and adolescents. Among the more significant findings is the fact that the development of peer relationships and academic achievement are interrelated and of equal importance to peer acceptance (Quicke, 1986). In addition, Quicke asserted that peer interactions were a critical part of the learning process for all students – with or without disabilities. Adults influence not only academics, but also socialization. Studies on the relationship between teachers' perception and socially appropriate behavior indicate that students who fail to meet teacher expectations are at increased risk for poor social interactions with teachers and peers, diminished academic performance, and a disproportionately higher number of disciplinary problems (Bereiter and Engelmann, 1967; Driscoll and Carter, 2004; Stainback and Stainback, 1992). The net result is that students who do not perform well academically risk disapproval of both classmates and adults and, if those problems go untreated, they increase the prospect of future school-related problems.

Chronological Age and the Socialization Process

There is ample evidence that peer interactions represent an important context within which children reach various social-interpersonal milestones (Hartup, 1979; Hartup and Stevens, 1997; Hodges *et al.*, 1999; Kao, 2000; Parker and Asher, 1987). From the beginning, peer attitudes play a pivotal role in the overall socialization of children and youth with and without special needs (McDougall

et al., 2004). At the elementary-school level, children engage in a wide range of social exchanges in various settings, such as the classroom, cafeteria, and the playground. Research suggests that children's peer interactions occur on several different levels: (1) general interactions, (2) interpersonal relationships, and (3) peer groups (Rubin *et al.*, 1998). Research on children's general interactions and relationships with peers suggest that children place a great deal of emphasis on social competence (Diamond and Carpenter, 2000; Ellis and Zarbatany, 2007; Hodges *et al.*, 1999) and popularity (Parker and Asher, 1987; Rubin *et al.*, 1998). In both respects, children at risk or with disabilities are more likely to be marginalized than their classmates.

By early adolescence, the majority of social interactions occur within a group context (Crockett *et al.*, 1984). In a study of 1872, high schoolers, McDougall *et al.* (2004) found that 61% of the participants held attitudes toward peers with special needs that ranged from slightly above neutral to very positive. However, 21% of the students held slightly below neutral to very negative attitudes toward peers with special needs. The researchers also found that when school personnel put emphasis on understanding and acceptance of students with individual differences, students had a more positive attitude toward classmates with special needs (McDougall *et al.*, 2004). In a related investigation, Hartup and Stevens (1997) looked at the influence of children's dyadic friendships on friends' behavior and found that attitudes generally become more alike over time. One conclusion is that, although adults have a powerful influence on children, routine exchanges with other children appear to play a critical role in overall social development (Musser and Graziano, 1991). Peer-peer interactions afford children experiences that are not endemic to child-adult interactions and have a profound impact on the social development of students with and without special needs (Hartup, 1983; Parker and Asher, 1987). At the same time, research on adolescent peer relationships indicates that popularity in this age group represents a defining characteristic of adaptive social development (e.g., Musser and Graziano, 1991). Interestingly, adolescence popularity within their peer group does not immunize them from succumbing to anti-social, delinquent behavior, and substance abuse (Allen *et al.*, 2005). Furthermore, Allen and his colleagues found that, in early adolescence, peer popularity plays a dual role. That is, adolescents who are popular within their wider peer group also are likely to be viewed as being more competent within their closest friendships.

Gender-Related Differences and Socialization

Our review revealed substantial gender-related differences, peer relationships, and the socialization process

(Rose and Rudolph, 2006). Studies on the frequency of group interactions among young children have resulted in conflicting outcomes (e.g., Benenson, 1993; Fabes *et al.*, 2003; Martin and Fabes, 2001). Studies conducted on playgroups revealed larger playgroup sizes for boys than girls, and that there were gender-related differences in the organization of children's playtime activities (Lever, 1976; 1978). Specifically, Lever (1976) found six differences of peer socialization between boys and girls including the fact that boys play outdoors far more than girls. Rather than participate in outdoor activities, girls preferred to play board games or with dolls. Even though boys and girls spent essentially the same amount of time playing alone, boys more often played in larger and more age-heterogeneous groups. Boys engaged in more competitive games than girls, with boys' games lasting longer than girls' games. In addition, researchers who assessed the frequency of dyadic interaction among young children found that boys interact in dyads as frequently, or more frequently, than do girls (Benenson, 1993; Benenson *et al.*, 1997; Martin and Fabes, 2001). Although it is difficult to draw any definitive conclusions, it is reasonable to conclude that the disparate nature of female/male play behavior results in different cognitive, emotional, and behavioral lessons learned.

Rose and Rudolph (2006) argued that girls have a greater tendency than boys to experience stress within the context of a dyadic friendship and to experience vicariously the emotional stress experienced by friends who are part of their social network. Purportedly, girls experience higher levels of stress than boys, particularly during adolescence. The only kind of peer stress that boys experience more than girls was physical and direct verbal victimization. One possible conclusion is that these gender-related differences may make girls more vulnerable to problems of emotional adjustment (Rose and Rudolph, 2006). Repeated exposure to peer stress may lead to diminished perceptions of social competence, concern about friends' welfare, and a sense of hopelessness, all of which can put youth at risk for emotional difficulties such as anxiety and depression (Rudolph, 2002). Further, girls tend to be exposed more to peer stress than boys – especially during adolescence (Rudolph and Hammen, 1999) which may explain the differences in age of onset and intensification of internalizing versus externalizing problems of adolescence (Rose and Rudolph, 2006). Males are diagnosed earlier and manifest more externalizing problems, such as verbal and physical aggression and antisocial acts; whereas, females are guilty of engaging in the so-called status offenses – behavior that is chronically unacceptable, such as running away and sexual activity. In sum, the differing social and emotional nature of male-versus female-dominated peer groups underscore the significance of individual cognitive, emotional, and behavioral coping skills and the influence

peers exert over one another's behavior. Finally, it seems logical to consider age- and gender-related differences in the socialization process, for both students with and without special needs.

Research on behavioral adjustment between boys and girls suggests that boys are more often exposed to overt or physical victimization and that exposure may contribute to their propensity for developing more externalizing behavioral problems such as aggression and antisocial behavior (Rose and Rudolph, 2006). In addition, boys are more likely to be targets of overt victimization which is highly predictive of an increase in externalizing problems over time (Hodges *et al.*, 1999; Hodges and Perry, 1999). There have been relatively few studies on peer stressors; those that have been conducted have produced inconsistent outcomes. Some indicate that the effects of peer stressors on adjustment are similar for boys and girls; while, others suggest that the effects may vary by sex (Rose and Rudolph, 2006). Peer-related stress among girls can pose a serious threat to their emotional well-being (Rose and Rudolph, 2006). Specifically, girls find that interpersonal stress (Rudolph, 2002) and relational or social victimization more emotionally challenging than their male counterparts (Crick *et al.*, 1996). Girls also report experiencing more negative emotions within the context of peer-peer interactions (Larson and Asmussen, 1991). The significance of these investigations is that interpersonal and peer stress are highly correlated with emotional difficulties for both female and male students, but the impact of that stress may manifest itself in different gender-specific ways (Rudolph, 2002).

Rose and Rudolph (2006) maintained that in responding to behavioral problems, girls tend to be more active in seeking support, whereas, boys are less likely to seek outside support and instead try to internalize negative feelings. Unresolved interpersonal conflict among peers may lead to future misunderstandings and hard feelings which, in turn, may precipitate bouts of physical aggression calculated to satisfy a desire for revenge (Rose and Rudolph, 2006). Additionally, Pollack (1998) suggested that because boys have fewer outlets for expressing emotions, such as disappointment or hurt feelings, it is more likely that they will express their emotions through anger and/or overt aggression.

As Rose and Rudolph (2006) pointed out, the social exchanges that occur within a relationship among children/adolescents are inextricably linked to positive emotional adjustment. Oldenburg and Kerns (1997) found that the perception of friendship as validation of an individual's self-worth can translate into lower levels of depression. In addition, positive interpersonal peer relationships appear to decrease the risk of school-related behavior problems (see Rose and Rudolph). In contrast, a negative emotional adjustment can have broad implications with regard to an individual's assessment of self and feelings of

low self-worth, emotional depression, and/or physical anxiety. In sum, gender-related differences appear to account for at least some of variance in children's emotional and behavioral adjustment versus nonadjustment (see Rose and Rudolph).

Culture and Socialization

Some researchers believe that culturally related factors play a key role in how society responds to the behavior of males versus females (e.g., Cartledge and Loe, 2001). However, Hastings *et al.* (2007) noted that gender differences in prosocial behavior may be more attributable to perception than to reality. They believe adults may be more likely to notice and to remember the prosocial behavior of girls because they conform to a cultural stereotype of appropriate feminine behavior. However, Cartledge and Loe (2001) discussed ways in which culture plays a role in shaping the school environment and how it impacts differently among diverse ethnic groups. Cartledge and Loe further explained that collectivistic and individualist orientations were the major distinctions between the dominant group in this society and those of other culturally diverse groups. Earlier, Kao (2000) summarized the stigma attached to diverse groups where students' self-images are shaped on the basis of what defines success among ethnic youth; particularly, some students from culturally diverse backgrounds speak of academic goals in terms of avoiding failure while Asian students focus on keeping up with high expectations on their academic pursuits. Additionally, Kao explored the effects of group image in relation to racial and ethnic stereotypes and expectations and further found that (1) groups defined their goals primarily in terms of the stereotypical images attached to their ethnic group, (2) minority youth focused on avoiding failure defined by prevalent group stereotypes, and these images maintain racially and ethnically segregated extracurricular activities that reinforced segregated peer groups, and (3) socialization with same-race peers promotes comparable conceptions of success within racial groups.

Aside from cultural influences on peer relations, researchers have examined economics and its effect on the development of interpersonal relationships. For example, Garner *et al.* (1994) investigated social competence among low-income populations and its effect on emotional development and socialization. They found that there is a positive correlation between emotional socialization practices and emotional knowledge in the development of low-income children's interpersonal skills. Garner (1996) found that the affective or moral attributions of low-income school-age children were similar to those of middle-income children. Middle- and low-income children reported more empathy, altruism, and

more aggression than guilt or denial, in response to a hypothetical emotion-eliciting situation (see Garner, 1996). Garner *et al.* (1994) stressed that family socialization practices are linked to the social competence of the child, the child's social and cognitive knowledge, and that the child's social and cognitive knowledge contributed to the child's overall social competence. Notwithstanding the obvious influence of families, there is compelling evidence that situational or contextual factors influence the socialization among peers with or without special needs as well (Favazza and Odom, 1997; Hampson, 1984; Haynie and Osgood, 2005).

Program Integration and Social Development

The long-standing practice of serving young children with special needs in settings that include normally developing peers is predicated on moral, legal, and educational grounds (Bailey *et al.*, 1998; Strain and Kerr, 1981). Various authors have offered a moral rationale to support of inclusive practices, namely, that children with special needs have a right to a life that is as normal as possible (Stainback and Stainback, 1992). Some (e.g., Diamond and Carpenter, 2000) have interpreted this perspective to include participation in the same high-quality programs afforded to children without special needs, along with participation in class activities and in positive relationships with individual class members. An educational justification for inclusion typically has focused on the developmental benefits of inclusion for children with special needs, including providing a more challenging and socially responsive learning environment in which they have an opportunity to learn from more competent peers.

Recent federal legislation, specifically the Individual with Disabilities Education Improvement Act (IDEIA) (IDEIA, 2004) and the Americans with Disabilities Act (ADA) (1990), lend legislative and legal support for inclusion (US Department of Education, 2007). IDEIA requires that children with special needs receive education services in the least-restrictive environment and, to the maximum extent appropriate, are educated with children who are not disabled (Buysse and Bailey, 1993). Research on social influences that stem from the integration of children with and without special needs has focused on various approaches and has yielded support for a mix of program options (Buysse and Bailey). Advocates of early integration offer a three-fold argument: (1) children have not yet formed impressions or biases about groups of individuals; therefore, it is important to minimize the probability of overt teasing or rejection, (2) early interactions with persons who have special needs should increase the future probability of acceptance by normally developing peers, and (3) early placement of children in real-world settings creates an expectation on the part of professionals and

parents that integration is the norm that more adequately prepares children with special needs to succeed – academically and socially – in a mainstream environment. In a closely related study, Wentzel (1993) explored the relationship between academic outcomes, measured against grades and standardized achievement test scores and students' social and academic behavior. Teacher ratings of students' prosocial behavior were a significant predictor of standardized achievement test scores. Finally, prosocial, antisocial, and academic behavior was highly predictive of students' grade-point averages. As others before, Gresham *et al.* (2001) claimed that the ability to interact successfully with peers and significant adults is one of the most important aspects of students' development. Other researchers (e.g., Parker and Asher, 1987) have emphasized that maintaining social competence and social adjustments depends on the extent to which students are able to (1) establish and maintain satisfactory interpersonal relationships, (2) gain peer acceptance, (3) establish and maintain friendships, and (4) terminate negative or pernicious interpersonal relationships.

Promoting Peer Relations and Socialization through Programs and Interventions

There is significant overlap between specific programs and the actual interventions conducted within those environments designed to promote peer relations and socialization in schools for students with special needs (Brown *et al.*, 2001; King-Sears, 2001; Mathur and Rutherford, 1991). Although the original concept of head start stressed the social and emotional aspects of a child's development (Zigler and Trickett, 1978), more prominent early-childhood programs emphasized curricular activities that would facilitate cognitive and language developments (Gray *et al.*, 1966; Weikart *et al.*, 1971), as well as reading and math (Bereiter and Engelmann, 1967). Social development was addressed in global, unspecified terms that related to a child's self-concept and sense of personal achievement (Brown *et al.*, 2001).

Diamond (2001) postulated that if children's experiences in inclusive programs are positive, these experiences will support the development of positive attitudes toward people with special needs, both during the early and in later years. In contrast, negative experiences in inclusive programs may lead to the development of prejudices toward people with special needs (Stoneman, 1993). In addition, research with preschool children has shown that peer-mediated intervention can be successful in facilitating children's successful participation in preschool activities (e.g., Robertson *et al.*, 2003). Research conducted on mid-childhood and adolescents indicated enduring positive

outcomes of participation in programs with peers with special needs (Mervis, 1998; Ryan, 2000).

Other researchers (e.g., Moroz and Jones, 2002) have examined the effectiveness of positive peer reporting on children's social involvement. Positive peer reporting is a technique that consisted of rewarding peers for publicly praising the social behavior of their peers. Moroz and Jones' findings supported the use of peers as sources of positive reinforcement for the prosocial behavior of at-risk children. In addition, Ang and Hughes (2001) conducted a meta-analytic investigation that explored the potential benefits of skills training with antisocial youth based on group settings. They found that in 38 studies of social skills training interventions and treatment with anti-social youth, skills-training interventions delivered in the context of groups consisting of only anti-social peers produced smaller benefits than did skills-training interventions that avoided aggregating antisocial peers. In their follow-up, 18 studies reported that treatments provided in the context of either mixed or individual treatment produced larger effect sizes than did deviant-only group interventions. Further, their findings suggested an implication for the recruitment of prosocial peers for inclusion in skills training interventions for aggressive and antisocial youth.

Conclusion

Both researchers and practitioners have long shown interest in peer relations and the socialization process. Despite a growing body of research, gaps between research and practice have become increasingly problematic and pose a major obstacle to optimizing positive student outcomes (Cook *et al.*, 2003). In reflecting on the progress within special education, Kauffman (2003) offered the following observations: (1) the scientific basis for special education is not yet universally accepted; (2) too many in the field discount science as a way of finding things out and believe that special education is fundamentally flawed, second rate, and ineffective as it relates to prevention, (3) we have yet to attach enough importance to academic instruction, (4) we have done too little to address the needs families who suffer the vicissitudes of poverty or the needs of children who have a variety of undesirable life outcomes, and (5) we must strive to achieve the elusive goal of greater harmony, mutual understanding, and equality of opportunity regardless of ethnicity or gender. In the past several decades, we have made real progress in promoting the prosocialization and peer acceptance of students with special needs, but much is yet to be accomplished. Given the life-altering significance of that challenge, now is not the time for modest ambitions.

Bibliography

- Allen, J. P., Porter, M. R., McFarland, F. C., McElhaney, K. B., and Marsh, P. (2005). The two faces of adolescents' success with peers: Adolescent popularity, social adaptation, and deviant behavior. *Child Development* **76**(3), 747-760.
- American Disabilities Act (1990). American Disabilities Act Information. <http://www.usdoj.gov/crt/ada/ada.htm> (accessed Jun. 2009).
- Ang, R. P. and Hughes, J. N. (2001). Differential benefits of skills training with antisocial youth based on group composition: A meta-analytic investigation. *School Psychology Review* **31**(2), 164-185.
- Bailey, D. B., McWilliam, R. A., Buysse, V., and Wesley, P. W. (1998). Inclusion in the context of competing values in early childhood education. *Early Childhood Research Quarterly* **13**, 27-47.
- Benenson, J. F. (1993). Greater preference among females than males for dyadic interaction in early childhood. *Child Development* **64**, 544-555.
- Benenson, J. F., Apostoleris, N. H., and Parnass, J. (1997). Age and sex differences in dyadic and group interaction. *Developmental Psychology* **33**, 538-543.
- Bereiter, C. and Engelmann, S. (1967). *Teaching Disadvantaged Children in the Preschool*. Englewood Cliffs, NJ: Prentice-Hall.
- Brown, W. H., Odom, S. L., and Conroy, M. A. (2001). An intervention hierarchy for promoting young children's peer interactions in natural environments. *Topics in Early Childhood Special Education* **21**(3), 162-175.
- Bullock, L. M. and Gable, R. A. (2001). *Addressing the Social, Academic, and Behavioral Needs of Students with Challenging Behaviors in Inclusive Settings*. Reston, VA: Council for Children with Behavioral Disorders.
- Buysse, V. and Bailey, D. (1993). Behavioral and developmental outcome in young children with disabilities in integrated and segregated settings: A review of comparative studies. *Journal of Special Education* **26**(4), 434-461.
- Cartledge, G. and Loe, S. A. (2001). Cultural diversity and social skill instruction. *Exceptionality* **9**(1), 33-46.
- Cook, B., Landrum, T. J., Tankersley, M., and Kauffman, J. M. (2003). Bringing research to bear on practice: Effecting evidence-based instruction for students with emotional and behavioral disorders. *Education and Treatment of Children* **26**(4), 345-361.
- Cook, J. W. and Wollersheim, J. P. (1976). The effect of labeling of special education students on the perceptions of contact versus noncontact normal peers. *Journal of Special Education* **10**(2), 187-198.
- Crick, N. R., Bigbee, M. A., and Howes, C. (1996). Gender differences in children's normative beliefs about aggression: How do I hurt thee? Let me count the ways. *Child Development* **67**, 1003-1014.
- Crockett, L., Losoff, M., and Petersen, A. C. (1984). Perceptions of the peer group and friendship in early adolescence. *Journal of Early Adolescence* **4**, 155-181.
- Deater-Deckard, K. (2001). Annotation: Recent research examining the role of peer relationships in the development of psychopathology. *Journal of Child Psychological Psychiatry* **42**(5), 565-579.
- Diamond, K. E. (2001). Relationships among young children's ideas, emotional understanding, and social contact with classmates with disabilities. *Topics in Early Childhood Special Education* **21**(2), 104-113.
- Diamond, K. E. and Carpenter, E. S. (2000). Participation in inclusive preschool programs and sensitivity to the needs of others. *Journal of Early Intervention* **23**(2), 81-91.
- Driscoll, C. and Carter, M. (2004). Spatial density as a setting event for the social interaction of preschool children. *International Journal of Disability, Development, and Education* **51**(1), 7-37.
- Ellis, W. E. and Zarbatany, L. (2007). Peer group status as a moderator of group influence on children's deviant, aggressive, and prosocial behavior. *Child Development* **78**(4), 1240-1254.
- Fabes, R. A., Martin, C. L., and Hanish, L. D. (2003). Young children's play qualities in same, other, and mixed sex peer groups. *Child Development* **74**, 921-932.
- Favazza, P. C. and Odom, S. L. (1997). Promoting positive attitudes of kindergarten-age children toward people with disabilities. *Exceptional Children* **63**(3), 405-418.
- Forness, S. R., Guthrie, D., and MacMillan, D. L. (1981). Classroom behavior of mentally retarded children across different classroom settings. *Journal of Special Education* **15**(4), 498-509.
- Garner, P. W. (1996). The relations of emotional role taking, affective/moral attributions, and emotional display rule: Knowledge to low-income school-age children's social competence. *Journal of Applied Developmental Psychology* **17**, 9-36.
- Garner, P. W., Jones, D. C., and Miner, J. L. (1994). Social competence among low-income preschoolers: Emotion socialization practices and social cognitive correlates. *Child Development* **65**, 622-637.
- Gray, S. W., Klaus, R. A., Miller, J. O., and Forrester, B. J. (1966). *Before First Grade: The Early Training Project for Culturally Disadvantaged Children*. New York, NY: Teachers College Press.
- Gresham, F., Sugai, G., and Horner, R. (2001). Interpreting outcomes of social skills training for students with high-incidence disabilities. *Exceptional Children* **67**(3), 331-344.
- Gresham, F., Van, M. B., and Cook, C. R. (2003). Social skills training for teaching replacement behavior: Remediating acquisition deficits in at-risk students. *Behavioral Disorders* **31**, 363-377.
- Hampson, R. B. (1984). Adolescent prosocial behavior: Peer-group and situational factors associated with helping. *Journal of Personality and Social Psychology* **46**(1), 153-162.
- Hartup, W. W. (1977). Peer interactions and the process of socialization. In Guralnick, M. J. (ed.) *Early Intervention and the Integration of Handicapped and Nonhandicapped Children*, pp 27-51. Baltimore, MD: University Park Press.
- Hartup, W. W. (1979). The social worlds of childhood. *American Psychologist* **34**, 944-950.
- Hartup, W. W. (1983). The peer system. In Mussen, P. H. (ed.) *Handbook of Child Psychology: Socialization, Personality, and Social Development*, pp 103-196. New York: Wiley.
- Hartup, W. W. and Stevens, N. (1997). Friendships and adaptations in the life course. *Psychological Bulletin* **121**(30), 35-270.
- Hartup, W. W., Glazer, J. S., and Charlesworth, R. (1967). Peer reinforcement and sociometric status. *Child Development* **38**, 1017-1024.
- Hastings, P. D., McShane, K. E., Parker, R., and Ladha, F. (2007). Ready to make nice: Parental socialization of young sons' and daughters' prosocial behaviors with peers. *Journal of Genetic Psychology* **168**(2), 177-200.
- Haynie, D. L. and Osgood, D. W. (2005). Reconsidering peers and delinquency: How do peers matter? *Social Forces* **84**(2), 1109-1130.
- Hodges, E. E. and Perry, D. G. (1999). Personal and interpersonal antecedents and consequences of victimization by peers. *Journal of Personality and Social Psychology* **76**, 677-685.
- Hodges, E. E., Boivin, M., Vitaro, F., and Bukowski, W. M. (1999). The power of friendship: Protection against an escalating cycle of peer victimization. *Developmental Psychology* **35**, 94-101.
- Individuals with Disabilities Education Improvement Act (IDEIA) of 2004, 20 U.S.C. § 1400 et seq. (2004).
- Kao, G. (2000). Group image and possible selves among adolescents: Linking stereotypes to expectations by race and ethnicity. *Sociological Forum* **15**(3), 407-430.
- Kauffman, J. M. (2003). Reflections on the field. *Education and Treatment of Children* **26**(4), 325-329.
- Kavale, K. A. and Forness, S. R. (1996). Social skill deficits and training: A meta-analysis. *Journal of Learning Disabilities* **29**, 226-237.
- Kerr, M. M. and Nelson, C. M. (2006). *Strategies for Managing Behavior Problems in the Classroom*, 5th edn. Upper Saddle River, NJ: Pearson Education.
- King-Sears, M. E. (2001). Institutionalizing peer-mediated instruction and interventions in schools. *Remedial and Special Education* **22**(2), 89-101.
- Larrieve, B. and Home, M. D. (1991). Social status: A comparison of mainstreamed students with peers of different ability levels. *Journal of Special Education* **25**(1), 90-101.
- Larson, R. and Asmussen, L. (1991). Anger, worry, and hurt in early adolescence: An enlarging world of negative emotions. In Colten, M. E. and Gore, S. (eds.) *Adolescent Stress: Causes and Consequences*, pp 21-41. New York: Aldine de Gruyter.
- Lever, J. (1976). Sex differences in the games children play. *Social Problems* **23**, 478-487.

- Lever, J. (1978). Sex differences in the complexity of children's play and games. *American Sociological Review* **43**, 471–483.
- Martin, C. L. and Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology* **37**, 431–446.
- Mathur, S. R. and Rutherford, R. B. (1991). Peer-mediated interventions for promoting social skills for children and youth with behavioral disorders. *Education and Treatment of Children* **14**, 227–242.
- McCord, J., McCord, W., and Howard, A. (1963). Family interaction as antecedent to the direction of male aggressiveness. *Journal of Abnormal and Social Psychology* **66**, 239–242.
- McDougall, J., DeWit, D. J., King, G., Miller, L. T., and Killip, S. (2004). High school-aged youths' attitudes toward their peers with disabilities: The role of school and student interpersonal factors. *International Journal of Disability, Development, and Education* **51**(3), 287–313.
- Mervis, B. (1998). The use of peer-pairing in schools to improve socialization. *Child and Adolescent Social Work Journal* **15**(6), 467–477.
- Morrison, G. and Borthwick, S. (1983). Patterns of behavior, cognitive competence, and social status for educable mentally retarded children. *Journal of Special Education* **17**(4), 441–452.
- Moroz, K. and Jones, K. (2002). Research into practice: The effects of positive peer reporting on children's social involvement. *School Psychology Review* **31**(2), 235–245.
- Musser, L. M. and Graziano, W. G. (1991). Behavioral confirmation in children's interactions with peers. *Basic and Applied Social Psychology* **12**(4), 444–456.
- Oldenburg, C. M. and Kerns, K. A. (1997). Associations between peer relationships and depressive symptoms: Testing moderator effects of gender and age. *Journal of Early Adolescence* **17**, 319–337.
- Parker, J. G. and Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin* **102**, 357–389.
- Piaget, J. (1932). *The Moral Judgment of the Child*. New York: Free Press.
- Piaget, J. (1978). *Behavior and Evolution*. New York: Pantheon Books.
- Pollack, W. (1998). *Real Boys*. New York: Henry Holt.
- Quicke, J. C. (1986). Pupil culture, peer tutoring, and special educational needs. *Disability, Handicap, & Society* **1**(2), 147–164.
- Read, D. L. (1960). *The Nursery School: A Human Relations Laboratory*, 3rd edn. Philadelphia, PA: W. B. Saunders.
- Robertson, J., Green, K., Alper, S., Schloss, P., and Kohler, F. (2003). Using a peer-mediated intervention to facilitate children's participation in inclusive childcare activities. *Education and Treatment of Children* **26**(2), 182–197.
- Roff, M., Sells, S. B., and Golden, M. M. (1972). *Social Adjustment and Personality Development in Children*. Minneapolis, MN: University of Minnesota Press.
- Rose, A. J. and Rudolph, K. D. (2006). A review of sex differences in peer relationship processes: Potential trade-offs for the emotional and behavioral development of girls and boys. *Psychological Bulletin* **132**(1), 98–131.
- Ross, H., Tesla, C., Kenyon, B., and Lollis, S. (1990). Maternal intervention in toddler peer conflict: The socialization of principles of justice. *Developmental Psychology* **26**(6), 994–1003.
- Rubin, K. H., Bukowski, W., and Parker, J. G. (1998). Peer interactions, relationships, and groups. In Damon, W. and Eisenberg, N. (eds.) *Handbook of Child Psychology: Vol. 3. Social, Emotional, and Personality Development*, pp 619–700. New York: Wiley.
- Rudolph, K. D. (2002). Gender differences in emotional responses to interpersonal stress during adolescence. *Journal of Adolescent Health* **30**, 3–13.
- Rudolph, K. D. and Hammen, C. (1999). Age and gender as determinants of stress exposure, generation, and reactions in youngsters: A transactional perspective. *Child Development* **70**, 660–667.
- Ryan, A. (2000). Peer groups as a context for the socialization of adolescents' motivation, engagement, and achievement in school. *Educational Psychologist* **35**(2), 101–111.
- Santos, A. and Winegar, L. T. (1999). Child social ethology and peer relations: A developmental review of methodology and findings. *Acta Ethol* **2**, 1–11.
- Schoenfield, N. A., Rutherford, R. B., Jr., Gable, R. A., and Rock, M. L. (in press). ENGAGE: A blueprint for incorporating social skills training into daily classroom academic instruction. *Preventing School Failure*.
- Stainback, W. and Stainback, S. (1992). *Curriculum Considerations in Inclusive Classrooms: Facilitating Learning for All Students*. Baltimore, MD: Paul H. Brookes.
- Stoneman, Z. (1993). Attitudes toward young children with disabilities: Cognition, affect, and behavioral intent. In Peck, C., Odom, S., and Bricker, D. (eds.) *Integrating Young Children with Disabilities in Community Programs: From Research to Implementation*, pp 223–248. Baltimore, MD: Brookes.
- Strain, P. S. and Kerr, M. M. (1981). *Mainstreaming of Children in Schools: Research and Programmatic Issues*. New York: Academic Press.
- US Department of Education (2007). IDEIA Information. <http://www.ed.gov/index.jhtml> (accessed Jun. 2009).
- Weikart, D. P., Rodgers, L., Adcock, C., and McClelland, D. (1971). *The Cognitively Oriented Curriculum*. Washington, DC: ERIC-NAEYC Publications.
- Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology* **85**(2), 357–364.
- Zigler, E. and Trickett, P. K. (1978). IQ, social competence, and evaluation of early childhood intervention programs. *American Psychologist* **33**(9), 789–798.

Further Reading

- Bhavnagri, N. and Samuels, B. (1996). Making and keeping friends: A thematic unit to promote understanding of peer relationships in young children. *Childhood Education* **72**, 219–224.
- Farmer, T. and Hollowell, J. (1994). Social networks in mainstream classrooms: Social affiliations and behavioral characteristics of students with E/BD. *Journal of Emotional and Behavioral Disorders* **2**, 143–155.
- Gresham, F. (1998). Social skills training: Should we raze, remodel, or rebuild? *Behavioral Disorders* **24**, 19–25.
- Hamre-Nietupski, S., Hendrickson, J., Nietupski, J., and Sasso, G. (1993). Perceptions of teachers of students with moderate, severe, or profound disabilities on facilitating friendships with nondisabled peers. *Education and Training in Mental Retardation* **28**, 111–127.
- Hemphill, L. and Siperstein, G. N. (1990). Conversational competence and peer response to mildly retarded children. *Journal of Educational Psychology* **82**(1), 128–134.
- Hyatt, K. and Filler, J. (2007). A comparison of the effects of two social skill training approaches on teacher and child behavior. *Journal of Research in Childhood Education* **22**(1), 85–96.
- Mathur, S. and Rutherford, R. (1994). Teaching conversational social skills to delinquent youth. *Behavioral Disorders* **19**(4), 294–300.
- Mathur, S. and Rutherford, R. (1996). Is social skills training effective for students with emotional or behavioral disorders? Research issues and needs. *Behavioral Disorders* **22**(1), 21–28.
- Mathur, S. and Rutherford, R. (1991). Peer-mediated interventions promoting social skills of children and youth with behavioral disorders. *Education and Treatment of Children* **14**(3), 227–235.
- Mathur, S., Kavale, K., Quinn, M., Forness, S., and Rutherford, R. (1998). Social skills interventions with students with emotional and behavioral problems: A quantitative study. *Behavioral Disorders* **23**(3), 193–200.
- Overton, S. and Rausch, J. (2002). Peer relationships as support for children with disabilities: An analysis of mothers' goals and indicators for friendship. *Focus on Autism and Other Developmental Disabilities* **17**(1), 11–29.
- Quinn, M., Kavale, K., Mathur, S., Rutherford, R., and Forness, S. (1999). A meta-analysis of social skill interventions for students with

- emotional or behavioral disorders. *Journal of Emotional and Behavioral Disorders* 7(1), 54–60.
- Rutherford, R., Mathur, S., and Quinn, M. (1998). Promoting social communication skills through cooperative learning and direct instruction. *Education and Treatment of Children* 21(3), 354–360.
- Sabornie, E. J., Kauffman, J. M., Ellis, E. S., Marshall, K. J., and Elksnin, L. K. (1988). Bi-directional and cross-categorical social status of learning disabled, behaviorally disordered, and nonhandicapped adolescents. *Journal of Special Education* 21(4), 39–56.

Peer-Tutoring

L J O'Shea, Delaware County Intermediate Unit, Morton, PA, USA

D J O'Shea, O'Shea Learning Consultants, Media, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Researchers, policymakers, and teachers currently focus on identifying educational practices that are research based. Effective educational practices must (1) have evidence that is supported by rigorous and scientific data, and (2) have a body of studies that demonstrate positive outcomes. The No Child Left Behind Act (NCLB) of (2001) (Public Law 107-110), and the Individuals with Disabilities Education Improvement Act (IDEIA) of 2004 (Public Law 108-446) have directed teachers – in the United States – to use scientifically based research to drive their decisions with regard to educational interventions.

This type of research is objective, empirical, and replicable. It has valid and reliable data, employs particular research designs, and undergoes rigorous data analysis (see *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User-Friendly Guide*, 2007 (US Department of Education, 2007)). Increasing exposure to scientifically based practices using proven instructional methods, materials, technologies, and supports will help students with disabilities to engage in acquiring and generalizing functional content from the general education curriculum. One such program that has a strong scientific basis – and has a body of studies that demonstrate positive outcomes – is peer tutoring (Baker *et al.*, 2004; Calhoun and Fuchs, 2003).

Due partly to NCLB and IDEIA 2004 mandates, teachers and service providers must hold their students with disabilities to the same instructional standards as normally developing peers. This push toward increased inclusion and standards in general education classes creates many obstacles. To cope with varied instructional and social demands on the increased inclusion, teachers often rely on peer tutoring to help diverse students survive and succeed. This article defines peer tutoring and discusses its importance to students with special needs.

Peer Tutoring: Definition and Rationale

A component under the direct-instruction practice of teaching to master is the practice of peer tutoring. Peer tutoring allows students to receive a substantial amount of practice in a specific skill without demanding a great deal of individual teacher attention, while still receiving immediate feedback to determine mastery of the skill

(Maheady *et al.*, 1994; O'Shea *et al.*, 1998). It has been in vogue for a number of years because of the scientific rigor it has undergone to support its effectiveness and efficacy. It has been at the center of scientifically based studies concerning students with and without disabilities – centering on students serving as powerful instructional resources to themselves, while providing positive social interactions to others (Maheady *et al.*, 2001; Mastropieri *et al.*, 2001, 2007). In fact, it is built on one major premise: students helping students to learn. Peer-tutoring programs contain components for active student response, opportunity to respond, feedback, and reinforcement. Important to the successful implementation of peer tutoring is the belief that teachers have confidence in students to instruct and support other students through practice. Other teacher beliefs that are critical to the successful use of peer tutoring include (1) peer tutoring is one means that allows students to receive a substantial amount of practice of a specific skill, (2) peer tutoring – as a teaching approach to assessment and intervention – is effective, convenient to classroom practice, and replete with instructional value to all students involved in the practice, and (3) through instruction in peer tutoring that reinforces the practice of appropriate instructional or behavioral responses, students can gain the ability to control their own actions covertly and maintain more control over situations.

Designing Peer Tutoring

The actual design of peer tutoring is simple: once the teacher determines the specific skill that needs to be mastered, the peer tutor provides the stimulus and the tutee (i.e., the receiver of the tutoring services) responds appropriately to that stimulus. The stimulus provided by the peer tutor can be from almost any subject area, but it must be a discreet variable that cues the tutee to respond appropriately. In addition, the peer-tutoring session that the teacher designs can be across students' ability and achievement levels if the teacher has clearly delineated the stimulus-response chain (O'Shea *et al.*, 1998). Researchers have examined types of peer-tutoring groups. The traditional tutor group involves two students; one student less skilled in a particular area tutored by a more capable student. However, tutoring can also involve a team of three members: one tutor and two tutees. A grouping of as many as six students of heterogeneous ability, also, may

compose the structure. It is imperative that teachers design and arrange groups heterogeneously – by skill ability – to ensure that each tutoring group has a student within the group whose ability for the particular skill being studied can assist the less capable student(s) when necessary (Maheady *et al.*, 1988, 1994). Teachers focusing on peer tutoring underscore the use of pairing appropriate student combinations. Teachers can assign dyads, small groups, or class-wide group based on their knowledge of student achievements, ability levels, social skills, interests, and preferences for selected topics. Their implementation of peer tutoring through the use of dyads or groups has allowed students to practice previously taught skills that include a wide range from academic concepts to specific rule differentiation. Academic and social benefits of peer tutoring have been clearly demonstrated in various settings and with a myriad of academic concepts. Research has clearly demonstrated the effectiveness of peer tutoring (Carlson *et al.*, 1985; Scruggs and Richter, 1985; King, 1982; Maher, 1984; Mathes *et al.*, 1994) as a method for teaching specific skills. To a large group peer tutoring also has resulted in increased academic performance by both the tutor and the tutee (Greer and Polirstok, 1982; Mathes *et al.*, 1994).

Peer-tutoring groups may involve a student with a disability in the role of tutee and the general education student as the tutor. In one study, results indicated that students with disabilities involved in peer-tutoring experiences scored significantly higher on posttesting in social studies than those students studying independently for the same test. The peer-tutoring session for students with disabilities and nondisabled peers resulted in all students reaching mastery (Campbell *et al.*, 1991). Other studies have viewed the role of students with disabilities as the tutor as well as the tutee, also reporting on positive academic gains made by students with and without disabilities (Dineen *et al.*, 1977). Peer tutoring resulting in increased academic performance – both by the tutor and the tutee – was the focus of studies that confirmed positive results for both members of the dyad (Greer and Polirstok, 1982; Mathes *et al.*, 1994). As it appears, peer tutoring has a built-in means for structuring interactions among and across students who normally do not socialize (Valcante and Stoddard, 1990). For example, it can be a valid means for increasing the socialization of isolated students and has proven to be beneficial for improving social skills of the tutor as well as the tutee (Cohen *et al.*, 1982; Maher, 1984). Jenkins and Jenkins (1985) confirmed that peer tutoring also allows students to practice appropriate social skills, while improving students' social standing within the class. Clearly, peer-tutoring programs can be used to promote and increase confidence in students because – through practicing content-area skills – many students report that they become the experts in the area and, in return, have greater confidence in content knowledge. Peer tutoring can be used effectively in diverse populations, especially inclusive classrooms

(Medcalf *et al.*, 2004; Nixon, *et al.*, 2001; Wright and Cleary, 2006).

Processes and Procedures Involved in Peer Tutoring

The success of a peer-tutoring program relies on a clear understanding of expectations and procedures for the tutoring session. It is critical that teachers, or other knowledgeable adults, provide training in skills of tutoring to peer tutors. Niedermeyer (1970) found increased mastery of content when tutors received structured training that included:

- Teaching the tutor a structured plan for each session.
- Teaching the tutor appropriate academic responses.
- Providing practice in giving appropriate corrective feedback.
- Providing practice in giving appropriate social reinforcements for correct responses.
- Developing a practical means for evaluating the tutor and tutees' progress in each session.

It is imperative that the tutor be knowledgeable of the content knowledge and skill being taught in the tutoring session. This may require that teachers hold a pretutoring session addressing the content and attributes of information to the tutor, as well as holding ongoing training sessions as warranted throughout the year (Brown *et al.*, 1971).

The Process of Structuring Specific Tasks

In order to ensure classroom success when implementing peer-tutoring opportunities, the tutor must demonstrate commitment to the tutoring session – approaching each session with the same enthusiasm a classroom teacher needs to approach each lesson. For this reason, teachers work to ensure that tutoring sessions provide structure to each student dyad with specific tasks to complete. Those class-wide programs designing and implementing a contingency plan for the positive learning and interaction of the tutor–tutee grouping also reported higher success rates when specific, structured tasks are apparent prior to the tutoring-session implementation (Delquardi *et al.*, 1986; Greenwood *et al.*, 1991).

Teachers can design and implement a point system for those tutoring groups who complete the daily assignments or reach mastery of material. To determine mastery of materials, evaluative tools – such as curriculum-based assessment (Deno and Fuchs, 1987; Fuchs and Deno, 1991; King-Sears, 1994) or precision teaching (Lindsley, 1972, 1990) – that detect daily progress are most effective for providing quick reinforcement for the students' efforts. Teachers can target ongoing data collection using a daily check of progress and on-task behavior(s). As teachers

or students use graphing charts to display data, students visually see progress and can receive opportunities to realize that hard work pays off. In addition, teachers can plan and implement an intermittent bonus point system every time positive phrases emanate from a student dyad or a tutoring group.

Initial training may impede the start of tutoring sessions, but it is imperative that the tutor is cognizant of his/her responsibilities and is capable of completing the task. The teacher needs to model appropriate behaviors throughout the teaching day – such that, in large-group formats, the teacher continues to model explicitly and directly – and guide the use of important teacher behaviors of structuring, demonstration, prompt, practice, and reinforcement for the tutor (Stoddard and Valcante, 1991).

The Process of Preparing for Peer Tutoring

Teachers can facilitate student success through the use of structured peer-tutoring sessions when they offer to students structured sessions and opportunities to practice (O'Shea *et al.*, 1998). The critical aspect of peer tutoring through dyads, small groups, or class-wide peer tutoring is that teachers control instruction, but use peer tutoring to support students' practice. Teachers must determine what and how subject-matter content will be presented (i.e., they observe and evaluate students, and demonstrate target responses, and lead students in making responses prior to having students practice in peer opportunities). As teachers gradually fade cues, students, increasingly, are able to work with peers. They are expected to give extensive practice so that students have sufficient opportunities to make responses and, thus, master the target skills. Practice activities are done in a fast-paced mode so that there is a high density of responses during instructional periods.

Teachers employing peer tutoring must monitor the accuracy of student responses, ensuring that dyads, groups, or the entire class engage in academic tasks. When tutors err, teachers must respond to provide corrective feedback. There are critical teacher behaviors that correlate positively with student academic gains. Teachers using direct observation and instruction include not only specific tutoring procedures, but, often, use highly structured peer-tutoring materials. The purpose of these materials is to help teachers systematically direct tutors–tutees in making accurate responses. Precise communication must occur between the tutor and tutee. During instruction but prior to tutoring sessions, teachers model target responses, lead students in making responses, and periodically test students' skills to respond without teacher cues.

Students with disabilities profit from classroom experiences in which teachers execute observation, assessment, and instructional behaviors in a standard, systematic manner. If teachers use predictable steps in peer-tutoring opportunities, such students are able to participate in dyads,

small groups, or class-wide opportunities, and perform responses in a more consistent manner. For example, when teachers execute the demonstration step of concept development, they model target responses for tutor and tutees. These responses relate to the performance of an academic rule (e.g., sets of procedures for performing academic tasks such as solving computations with regrouping) or the formation of concepts (e.g., well-defined categories of information used for classification – such as names of state capitals). Teachers are, then, in a better position to observe and assess what is being practiced when peer-tutoring opportunities are initiated. Teachers need to offer guided practice that is supervised prior to (and during) peer-tutoring sessions, as warranted. Guided practice entails teachers using cues and prompts to emit target responses. Teachers lead groups of students in making the desired response (i.e., the teacher and students perform the response simultaneously) prior to dyad implementation. As students demonstrate accurate responses, the teacher gradually fades his/her degree of leading. Instead of making the response with the dyads, the teacher allows the tutor to present the stimulus and has the tutee or group of tutees respond. Using unison responding, class-wide tutees respond – making responses simultaneously, allowing more opportunities for responding during a period of time. Teachers must monitor the response of individual dyads to determine whether all are responding accurately. As students become more accurate in their responding, they shift the performance criterion to fast and accurate responding and concentrate on asking for as many responses as possible during a guided-practice session. This increases opportunities to respond and help to ensure skill mastery. Effective teachers ensure that responses are demonstrated and that students have numerous opportunities to practice responses while the teacher leads.

Finally, once students have performed responses accurately without cues or prompts, teachers using peer-tutoring assign dyads in class-wide grouping and provide dyads the opportunity to practice response accuracy and speed. Independent practice, therefore, is minimally supervised practice in which dyads respond in class-wide opportunities without the teachers' presence. Two important and closely related features of independent practice assist in skill development and efficient management of class-wide activities. First, teachers can work with other groups of students while dyads practice independently; and second, dyads must be making accurate responses so they can truly work independently of the teacher (i.e., without a constant parade of dyads moving over to the teacher to ask questions about their seatwork). When students engage in class-wide peer-tutoring activities, the teacher should prepare them and monitor their progress in completing the assignment. As a sequence typically, the teacher should: (1) give directions to complete the peer-tutoring tasks, (2) lead class-wide activities in completing example tasks, (3) indicate

Table 1 Skills facilitating active presentation–response of both the tutor and tutee

<i>Tutor presentation</i>	<i>Tutee response</i>
1. Orally states spelling word	Writes spelling word orally states spelling word
2. Card with vocabulary word from core academic domains: science, reading, mathematics, social studies, geography, computers, health, world history, economics	Orally defines word
3. Card with math fact on front	Mentally computes and orally states the answer
4. Orally states a numeral	Writes the numeral stated
5. Presentation of a color on a card	Orally states the color

deadline for completing activity, (4) circulate and assist dyads, if warranted, and (5) review class-wide responses to activity tasks during – and at the culmination of – peer tutoring. Clearly, ongoing observation better ensures that dyads will not proceed through an extended series of tutoring sessions with high-error rates before instruction is modified to improve response accuracy. Monitoring can take the form of simply tallying the number of correct and error responses dyads make during an oral practice session or on a paper-and-pencil tutoring assignment. These frequencies can be kept in a daily log or a bulletin-board chart, or can be plotted on a graph. Teachers continue to collect instructional data based on educational decisions and by linking their evaluations and instruction.

Procedures to Ensure Appropriate Peer-Tutoring Content

Teachers must ensure that the content of material to be mastered within the tutor session is appropriate for members of a particular dyad. They can work actively to adapt the presentation of content to fit a tutoring session. For example, Heron *et al.* (1983) reported that most successful dyads are those in which there is active participation by both members of the group in a tutor–tutee interaction mode. In addition, those practice areas that allow student dyads to use a predeveloped correct answer key enable students of equal abilities to pair. The key to correct responding can prevent arguments by overzealous partners who are certain that their incorrect answer is correct. The skills identified by O'Shea, *et al.* (1998) lend themselves to an active presentation–response of both the tutor and tutee (see **Table 1**).

Greenwood *et al.* (1991) reported on the benefits and practice effects of continually introducing new material to class-wide groups once the previous information has been mastered. If the expectation of the peer-tutoring experience is to increase interaction between two students, the dyad must meet for a consistent period of time. Those peer-tutoring sessions resulting in quick mastery of content information ending the tutoring session may be successful for teaching content mastery, but the limited social interactions of the tutor and tutee may not promote long-term reciprocal

benefits in partner relationships. Teachers' monitoring of social interactions, task completion, and student engagement become critical antecedents to material introduction and future class-wide assignments.

What are the Benefits of Peer-Tutoring Opportunities?

It has become increasingly evident that peer tutoring has a sound theoretical basis. It has a consensus with existing literature and produces desired outcomes (Heron *et al.*, 2006; Wright and Cleary, 2006). Peer tutoring offers a practical solution to meet the demands of inclusive learning milieus with diverse learners of varying abilities and levels (Copeland *et al.*, 2004; Maheady *et al.*, 2001; Mastropieri *et al.*, 2007). It has been shown to increase students' confidence and achievement through practice in core subject areas – such as mathematics, reading, and written expression. In addition, it holds justification beyond the purpose of increasing students' opportunities for practice and academic achievement. Class-wide programs can support teachers' inclusion opportunities as they work in different instructional groupings of students with varying needs.

One of the most recognized and heavily researched peer-tutoring programs is entitled, Peer-Assisted Learning Strategies (PALS) (Baker *et al.*, 2004; Calhoun and Fuchs, 2003; Fuchs *et al.*, 2000). This program is typically used for mathematics and starts with the teacher designing appropriate student dyads to practice mathematical problems. In this program, the teacher must instruct students to tutor each other through specific guidelines. Typically, a student with low achievement or ability can be paired with a peer without disabilities. PALS has been found to promote and improve computational as well as comprehension skills for students with disabilities. Teachers favor PALS over other programs because teachers consistently use it to assist with the task of class-wide progress monitoring. Students report that they actually favor the program over receiving tangible rewards and report that seeing their progress helps to motivate them (either as tutors or tutees), irrespective of ability level (Baker *et al.*, 2004; Fuchs *et al.*, 2000).

Strategies can be taught and learned from the use of peer-tutoring programs. For example, Calhoun and Fuchs (2003) reported that the use of *PALS* can be an extremely viable instructional technique to use in mathematic classes at the high school level. Baker *et al.* (2004) confirmed the use of *PALS*—finding that students can improve cognitively and socially from working with their peers, as students can explain and clarify issues for one another, while developing solid strategies and skills for solving problems. Reading and writing skills can also be improved in inclusive classrooms through the use of peer-tutoring programs. Through various studies, it was found that both tutors and tutees improved their writing skills and benefited overall from participating in the peer-tutoring sessions. Both groups were found to be on task a majority of the time and also expressed enjoyment about working in dyads with their peers (Medcalf *et al.*, 2004; Nixon *et al.*, 2001; Wright and Cleary, 2006). Clearly, peer-tutoring programs can be optimized throughout inclusive education and can yield many possible benefits for all students. McDonnell *et al.* (2001) reported that a class-wide peer-tutoring program—combined with curriculum and accommodations—can improve levels of academic responding and decrease levels of competitive behaviors for students with disabilities. They also concluded that peer-tutoring programs can increase meaningful participation among all diverse, classroom learners. In addition, adult benefits derived from the use of peer-tutoring programs have been confirmed associated with cost, accommodation, and duration. Peer tutoring is highly cost effective to schools and is user friendly. Teachers can use these programs effectively and correctly while helping to produce significant gains in academic achievement for students, while not costing school entities much, if any, money. Classroom teachers can easily establish peer tutoring to accommodate students' social needs, in accordance with student demographics and classroom composition. Simply put, peer tutoring is durable across districts (Nugent, 2001).

Conclusion

Peer tutoring is a very successful and effective practice methodology to help students increase their academic and social skills. Peer-tutoring programs are important in today's classrooms. They have proven to be objective, empirical, and replicable; in addition, they have been substantiated by valid and reliable data, across research designs and undergoing rigorous data analysis. Many teachers working with students with disabilities use teacher-guided instruction in class-wide programs directed toward the mastery of specific skills. By using peer tutoring, they can include all students in inclusion programs with high rates of engagement. These authors believe peer-tutoring programs can be successfully utilized and maximized. Teachers have the

burden of trying to meet the needs of all students and the task of creating a positive working environment where all students can collaborate and share in learning. With the use of a peer-tutoring program, competency levels of both students with and without disabilities can benefit. A class-wide peer-tutoring program can improve levels of academic responding and decrease levels of competing behaviors. Peer-tutoring programs are effective to use across all ages and abilities. They help students to make marked improvements throughout various disciplines while increasing their socialization skills and accelerating their confidence and motivation. Programs are positive and effective to use with diverse populations, across all age levels, and for different types of subject matter. Additionally, peer tutoring enables teachers to teach strategies and give students time to practice these strategies through tutoring sessions (Copeland *et al.*, 2004; Davenport *et al.*, 2004; Mastropieri *et al.*, 2001; McDonnell *et al.*, 2001). Finally, peer tutoring makes so much sense in this age of inclusion. It has been research proven and its pedagogical benefits are impressive.

Bibliography

- Baker, S., Gersten, R., Dimino, J. A., and Griffiths, R. (2004). The sustained use of research-based instructional practice: A case study of peer-assisted learning strategies in mathematics. *Remedial and Special Education* 25(1), 5–24.
- Calhoun, M. B. and Fuchs, L. S. (2003). The effects of peer-assisted learning strategies and curriculum-based measurement on the mathematics performance of secondary students with disabilities. *Remedial and Special Education* 24(4), 235–245.
- Campbell, B. J., Brady, M. P., and Linehan, S. (1991). Effects of peer-mediated instruction on the acquisition and generalization of written capitalization skills. *Journal of Learning Disabilities* 24, 6–14.
- Cohen, P. A., Kulik, J. A., and Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal* 19, 237–248.
- Copeland, S. R., Hughes, C., Carter, E. W., *et al.* (2004). Increasing access to general education: Perspectives of participants in a high school peer support program. *Remedial and Special Education* 25(6), 342–352.
- Davenport, S. V., Arnold, M., and Lassmann, M. (2004). The impact of cross-age tutoring on reading attitudes and reading achievement. *Reading Improvement* 41(1), 3–12.
- Delquardi, J., Greenwood, C. R., Whorton, D., Carta, J. J., and Hall, (1986). Classwide peer tutoring. *Exceptional Children* 52, 535–542.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children* 52, 219–232.
- Deno, S. L. and Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children* 19, 1–16.
- Dineen, J. P., Clark, H. B., and Risley, T. R. (1977). Peer tutoring among elementary students: Educational benefits to the tutor. *Journal of Applied Behavior Analysis* 10, 231–238.
- Fuchs, L. S. and Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children* 58, 488–500.
- Fuchs, D., Fuchs, L. S., and Burish, P. (2000). Peer-assisted learning strategies: An evidence-based practice to promote reading achievement. *Learning Disabilities Research and Practice* 15(2), 85–91.
- Greenwood, C. R., Carta, J. J., and Maheady, L. (1991). Peer tutoring programs in the regular education classroom. In Stoner, G., Shinn, M. R., and Walker, H. M. (eds.) *Interventions for Achievement and Behavior*

- Problems*, pp 179–200. Washington, DC: National Association of School Psychologists (NASP).
- Greer, R. D. and Polirstok, S. R. (1982). Collateral gains and short-term maintenance in reading and on-task responses by some inner-city adolescents as a function of their use of social reinforcement while tutoring. *Journal of Applied Behavior Analysis* **15**, 123–139.
- Heron, T. E., Heward, W. L., Cook, N. L., and Hill, S. (1983). Evaluation of class wide peer tutoring systems: First graders teach each other sight words. *Education and Treatment of Students* **6**, 137–152.
- Heron, T. E., Villareal, D. M., Yao, M., Christianson, R. J., and Heron, K. M. (2006). Peer tutoring systems: Applications in classroom and specialized environments. *Reading and Writing Quarterly* **22**(1), 27–45.
- Individuals with Disabilities Education Improvement Act of (2004), *H.R. 1350*.
- Jenkins, J. J. and Jenkins, L. M. (1985). Peer tutoring in elementary and secondary programs. *Focus on Exceptional Children* **17**(6), 1–12.
- King, R. T. (1982). Learning from a PAL. *Reading Teacher* **35**, 682–685.
- King-Sears, M. E. (1994). *Curriculum-Based Assessment in Special Education*. San Diego, CA: Singular.
- Lindsley, O. R. (1972). From skinner to precision teaching. In Jordan, J. B. and Robbins, L. S. (eds.) *Let's Try Doing Something Else Kind of Thing*, pp 1–12. Arlington, VA: Council on Exceptional Children.
- Lindsley, O. R. (1990). Precision teaching: By teachers for students. *Teaching Exceptional Students* **22**(3), 10–15.
- Maheady, L., Harper, G. F., and Mallette, B. (2001). Peer-mediated instruction and interventions and students with mild disabilities. *Remedial and Special Education* **22**(1), 4–14.
- Maheady, L., Harper, G. F., and Sacca, M. K. (1988). Peer-mediated instruction: A promising approach to meeting the diverse needs of LD adolescents. *Learning Disabilities Quarterly* **11**, 108–113.
- Maheady, L., Mallette, B., Harper, G. F., Sacca, K., and Pomerantz, D. (1994). Peer-mediated instruction of high-risk students. In Wood, K. D. and Algozzine, B. (eds.) *Teaching Reading to High Risk Learners*, pp 269–290. Boston, MA: Allyn and Bacon.
- Maier, C. A. (1984). Handicapped adolescents as cross-age peer tutors: Program description and evaluation. *Exceptional Children* **51**, 56–63.
- Mastropieri, M. A., Scruggs, T. E., and Berkeley, S. L. (2007). Peers helping peers. *Educational Leadership* **64**(5), 54–58.
- Mastropieri, M. A., Scruggs, T., Mohler, L., et al. (2001). Can middle school students with serious reading difficulties help each other and learn anything? *Learning Disabilities Research and Practice* **16**(1), 18–27.
- Mathes, P. G., Fuchs, D., Fuchs, L. S., Henley, A. M., and Sanders, A. (1994). Increasing reading practice with Peabody class wide peer tutoring. *Learning Disabilities Research and Practice* **9**, 44–48.
- McDonnell, J. M., Mathot-Buckner, C., Thorson, N., and Fister, S. (2001). Supporting the inclusion of students with moderate and severe disabilities in junior high school general education classes: The effects of class wide peer tutoring, multi-element curriculum, and accommodations. *Education and Treatment of Students* **24**(2), 141–160.
- Medcalf, J., Glynn, T., and Moore, D. (2004). Peer tutoring in writing: A school systems approach. *Educational Psychology in Practice* **20**(2), 157–178.
- Niedermeyer, F. C. (1970). Effects of training on the instructional behaviors of student tutors. *Journal of Educational Research* **64**, 119–123.
- Nixon, J., Dundee, U., and Topping, K. (2001). Emergent writing: The impact of structured peer interaction. *Educational Psychology* **21**(1), 41–58.
- No Child Left Behind Act of (2001). Public Law 107–110. www.nclb.gov (accessed August 2009).
- Nugent, M. (2001). Raising reading standards – the reading partners approach: Cross-age peer tutoring in a special school. *British Journal of Special Education* **28**(2), 71–79.
- O'Shea, L. J., O'Shea, D. J., and Algozzine, B. (1998). *Learning Disabilities: From Theory Toward Practice*. New York: Prentice-Hall.
- Scruggs, T. E. and Richter, L. (1985). Tutoring learning disabled students: A critical review. *Learning Disability Quarterly* **8**, 286–289.
- Stoddard, K. and Valcante, G. (1991). "Classroom buddies" methods and effects of mainstreaming on special needs preschoolers and kindergarten regular education students. *Paper Presented at the International Convention of the Council for Exceptional Children* Atlanta, GA.
- U.S. Department of Education (2007). *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User-Friendly Guide*. www.ed.gov/rschstat/research/pubs/rigoroussevid/guide_pg3.html (accessed August 2009).
- Valcante, G. and Stoddard, K. (1990). Peer tutoring in natural community settings: Strategies for teacher training. *Paper Presented at the American Association on Mental Retardation Conference*. Atlanta, GA.
- Wright, J. and Cleary, K. (2006). Kids in the tutor seat: Building schools' capacity to help struggling readers through a cross-age-peer tutoring program. *Psychology in the Schools* **43**(1), 99–107.

Further Reading

- Algozzine, B. and Ysseldyke, J. E. (2006). *Effective Instruction for Students with Special Needs*. Thousand Oaks, CA: Corwin Press.
- Biancarosa, G. and Snow, C. (2004). Reading next: A vision for action and research in middle and high school literacy – A Report to Carnegie Corporation of New York. Washington, DC: Alliance for Excellent Education.
- Cartledge, G. and Lo, Y. (2006). *Teaching Urban Learners: Culturally Responsive Strategies for Developing Academic and Behavioral Competence*. Champaign, IL: Research Press.
- Cawelti, G. (ed.) (1995). *Handbook of Research on Improving Students Achievement*. Arlington, VA: Educational Research Service.
- Dale, E. (2005). *Learning Pyramid*. Bethel, ME: National Training Laboratories.
- Flood, J. and Anders, P. L. (2005). *Literacy Development of Students in Urban Schools: Research and Policy*. Newark, DE: International Reading Association.
- Fuchs, L. S. and Fuchs, D. (1991). Curriculum-based measurements current applications and future directions. *Preventing School Failure* **35**(3), 6–11.
- Grossen, B. (2004). Success of a direct instruction model at a secondary level school with high-risk students. *Reading and Writing Quarterly* **20**, 161–178.
- National Reading Panel (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction; *Reports of the Subgroups*. Bethesda, MD: National Institute of Child Health and Human Development. http://www.nationalreadingpanel.org/Publications/publications.htm.
- Padrón, Y. N. and Waxman, H. C. (1999). Effective instructional practices for English language learners. In Waxman, H. C. and Walberg, H. J. (eds.) *New Directions for Teaching Practice and Research*, pp 171–203. Berkeley, CA: McCutchan.
- Stevens, R. J. and Durkin, S. (1992). *Using Student Team Reading and Student Team Writing in Middle School*. Baltimore, MD: Johns Hopkins University, Center for Research on Effective Schooling for Disadvantaged Students.
- Topping, K. J. (2005). Trends in peer learning. *Educational Psychology* **25**(6), 631–645.
- Vacca, R. T. and Vacca, J. L. (2005). *Content Area Reading*, 8th edn. Boston, MA: Allyn and Bacon.

Relevant Websites

- http://www.nsu.edu – North Carolina State University Undergraduate Tutorial Center.
- http://lynx.csusm.edu/capi/owl.html – The Collaborative Academic Preparation Initiative.
- http://www.crla.net – The College Reading and Learning Association.
- http://www.writingcenter.tamu.edu – The International Writing Centers Association.
- www.writing.msu.edu – The Writing Center at Michigan State University.
- http://coldfusion.louisville.edu – The Writing Center Research Project.

Postsecondary Participation of Students with Special Needs

R A Stodden and K D Roberts, University of Hawaii at Manoa, Honolulu, HI, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Education is the key factor in achieving meaningful employment and an enhanced quality of life for persons with special needs. As the US strives to compete in the global economy, leaders in the nation's business, information, and commerce sectors have indicated the need for highly educated workers who are competent in higher-order thinking and possess specialized technical skills. This need, coupled with gradual changes in the US labor market, functionally mandates a post-secondary education for those seeking professional success and financial stability. Completion of post-secondary education, including vocational-technical training, significantly improves the chances of those with and without special needs to secure gainful and satisfying employment and achieve financial independence. With this reality, it is encouraging to see that students with disabilities who earn bachelor of arts degrees attain subsequent employment nearly on par with students without disabilities (National Council on Disability, 2000; Office of Special Education and Rehabilitative Services, 2000).

Policy and curricular changes supported by the Individuals with Disabilities Education Act (IDEA) and Title II of the Higher Education Act (HEA) under the No Child Left Behind Act (NCLB) have focused on improving academic achievement and post-high school expectations for students with special needs, with the aim of facilitating access to higher education. Although they have support in secondary education under IDEA, students with special needs who successfully complete secondary education are too often inadequately prepared to accept responsibility for their own life decisions during the post-high school years (Brinckerhoff, 1994; Izzo *et al.*, 2001). In post-secondary settings, students who may have previously received a prescriptive web of support in secondary school, often find themselves unprepared to face the necessities of making complex financial decisions, learning about their legal rights, and self-advocating for their support needs. Quite suddenly, the provision of assistance is no longer automatic or standardized under one federal rubric. The Rehabilitation Act and the Americans with Disabilities Act (ADA) do not mandate specific accommodations. Individual institutions have considerable discretion to interpret the parameters of the reasonable accommodations required by law. Resources are often inadequate and disconnected. The type, range, and availability of services as well as the terms related to these services are often widely discrepant

and poorly integrated while access to mentors or technological training is either limited or nonexistent (Stodden *et al.*, 2002). This article provides an overview of issues currently faced by individuals with special needs in terms of access to, retention in, and completion of post-secondary education programs. Possible solutions and suggestions are also provided.

Gaining Access to Post-Secondary Education

Youth with special needs seeking the right of access to post-secondary education may encounter formidable obstacles in their pursuit. Two of the biggest challenges students will face are the preparation for, and the transition to, these programs.

K-12 Preparation for and Transition to Post-Secondary Education

For individuals with special needs transition to post-secondary education is generally smoother if, during their secondary education, they are engaged in rigorous learning through curricula focused on post-secondary-education goals. However, often during secondary school, emphasis is instead placed on providing youth with special needs with prescriptive, specialized services and supports focused specifically upon remediating learning or behavior deficits experienced by them. Often the determination of these supports is made without the student's involvement in the decision-making process (Abery and Stancliffe, 1996). This process, in which professionals and parents often make the decisions, limits the opportunities for students with special needs to develop and practice their self-determination and self-advocacy skills (Izzo and Lamb, 2002). Not surprisingly, these students often leave secondary school with neither these skills, nor the knowledge of the impact that their special need has upon their learning or of the related modes of assistance which can help mitigate this impact. Furthermore, they may lack the understanding of how to negotiate post-secondary settings, where the provision of reasonable accommodations rather than detailed services designed to meet individual needs is the focus (National Center for the Study of Postsecondary Education Supports, 2002; Stodden *et al.*, 2003a). Without the essential skills of access to higher education namely self-determination and

self-advocacy, students with special needs face an uphill task in transitioning to post-secondary education.

Efforts have been made over the past 20 years to teach youth with special needs the skills of self-determination and self-advocacy (i.e., decision making, problem solving, goal setting and attainment, and leadership) based on teacher interpretations of self-determination. However, the literature suggests that these efforts would be more effective if they were based instead on providing real opportunities for students to make their own decisions and accept their consequences (Brinckerhoff, 1994; Izzo *et al.*, 2001). Building the skills of self-advocacy and self-determination is essential for empowering students with special needs seeking post-secondary education, as they have to face the transition from services provided under IDEA to the self-advocacy required in post-secondary education under the ADA and the Rehabilitation Act.

Another barrier faced by students with special needs is the difficulty of achieving high academic standards, which is often exacerbated by the failure of secondary institutions to provide adequate or appropriate curricula (Berliner and Biddle, 1996; Fuchs and Fuchs, 1998; Hatch, 1998). According to Stodden *et al.* (2003), there is an all too common tendency for secondary schools to place students with special needs in separate classrooms where they may receive substandard secondary curricular content (Stodden *et al.*, 2003a, 2003b). This problem is magnified by other institutional factors including problems with disability identification in secondary institutions (Thurlow, 2001), poor use and application of promising technology (Burgstahler, 2002), poor coordination and management of supports and services (Whelley *et al.*, 2002), and lack of clarity among professionals and families about the necessary supports and accommodations to be provided (Stodden *et al.*, 2003a).

Other issues affecting academic performance include state-standard-based curricula and assessment measures designed without input from special educators, which needlessly increase the challenges for students with special needs and their teachers (Stodden *et al.*, 2003b); teacher concerns such as inadequate professional development, excessive paperwork, and attrition specifically resulting from these problems (American Youth Policy Forum and Center on Education Policy, 2002); academic and career counselors who lack the necessary skills to provide guidance to students with special needs; and inadequate direction and counsel due to a lack of coordination among teachers and counseling staff. Perhaps the most troubling barrier to academic achievement for students with special needs is the one constructed by teachers, career counselors, administrators, family members, and the students themselves. Shackling students with special needs with low expectations and a limited sense of opportunity makes them develop a sense of failure before they have even begun to explore their interests and aspirations (Stodden *et al.*, 2002).

In May 2000, The National Council on Disability stated that endeavors to promote a smooth transition from secondary to post-secondary education have not met the goals of federal laws and initiatives, such as IDEA and Section 504 of the Rehabilitation Act. The Council attributed this to a number of factors including the methods of transition planning, which do not involve the students themselves, and the inadequate allocation of resources (NCD, 2000). Adding to the difficulty of the task, state and local education agencies across the United States are experiencing a chronic shortage of qualified personnel to serve children and youth with special needs (American Youth Policy Forum and Center on Education Policy, 2002). When supports and services are available, too often they are primarily focused on the students achieving a single academic outcome rather than on a holistic approach focused upon a continuum of outcomes leading to a successful transition (Izzo and Lamb, 2002; Stodden *et al.*, 2003). Part of the problem is the lack of consensus on the definition of a successful outcomes. For example, if obtaining a general educational development (GED) or high school diploma is viewed as a successful outcome, then the preparation process may be viewed as finished with the completion of high school. This premature and short-sighted view of outcome may provide justification that special-education services, offered in secondary school, are no longer necessary and therefore, be retracted in post-secondary education institutions. Hence, the selective process of developing certain skills or goals is unlikely to ameliorate the ability of students with special needs to face upcoming challenges in their transition and participation in post-secondary education.

Complicating current secondary-school transition efforts further is the lack of awareness among educators and parents regarding the policy contrasts between IDEA at the secondary level and ADA and Section 504 at the post-secondary level. Many secondary schools lack a formal structure to assist students in adjusting to the highly discrepant laws governing secondary and post-secondary education (Stodden *et al.*, 2003b). As a result, they do not tailor the delivery of services and instruction toward strengthening the links between secondary and post-secondary education. The result is that the students themselves, parents, and other natural supports are often caught unaware when the level of service provision drops and/or is not automatically extended following high school (Stodden *et al.*, 2003a). The lack of knowledge about differences in their rights, services, and funding has the effect of discouraging or possibly even barring students with special needs from higher education.

As students with special needs endeavor to access post-secondary education, they may find that circumstances vary significantly from college to college, and from state to state. Each college provides a different level and type of support. Some institutions employ a single counselor to

take responsibility for disability issues. In such institutions, staff members provide advice and letters to professors verifying that a student request for accommodations is justified. Little else may be provided. At other schools, multiple staff members coordinate services and accommodations for students with special needs so that the educational environment provides supplementary support, and additional staff is prepared to teach the students about self-advocacy (Youth Advisory Committee to the National Council on Disability, 2003).

Clearly, the challenge to locate and advocate for services and accommodations can be quite frustrating. The various systems feature limited resources, inconsistent terminology, disconnected agencies, inconsistent laws, and conflicting eligibility requirements. This is difficult to manage, even for the most self-determined student (Whelley *et al.*, 2002). Ultimately, without a functioning, successful transition program from secondary to post-secondary education, youth with special needs find themselves burdened with additional disadvantages.

Recommendations: Preparation for and Transition to Post-Secondary Education

Reform efforts that work for all students will also work for students with special needs. The same strategies that recent research has found to improve the outcomes for many students with special needs can easily be employed to better serve all at-risk youth. Five effective research-based recommendations are presented here:

- *Facilitate a personalized atmosphere and closer relationships between students and high school staff.* As this is best accomplished through smaller schools, provide incentives for schools to dramatically restructure schedules, create schools within a school, or build new facilities. Use evidence-based strategies to improve one-on-one relationships with youth at risk. Encourage families to develop collaborative and mutually beneficial relationships with high school staff, employers, and college personnel who can assist students to transition. Provide incentives to role models, those successful individuals with special needs who can mentor youth facing similar challenges. Fund follow-up studies on students who are educated through these reforms, 1 year and 3 years after leaving school, to record their outcomes and refine future efforts. Continue to call for improvements until adult outcomes for youth with disabilities match those of their peers without disabilities.
- *Provide opportunities for students to apply their knowledge in realistic settings.* Give employers incentives to offer students a variety of work experiences (i.e., job shadow and job tryout, or internships) that are closely linked to the academic studies and future career goals of youth at risk. Enable more effective collaboration among employers,

career guidance staff, and students. Increase vocational rehabilitation agency involvement by increasing the number of students (ages 16–22) they serve and promoting their use of post-secondary education services. Make individualized education plans (IEPs) to identify career prep opportunities for students with special needs, and use these IEPs to regularly record students' progress as they explore careers. Promote self-determination education to encourage students to actively participate in, and take responsibility for, their career preparation and/or transition to college.

- *Prepare teacher and school personnel to have the highest quality and capacity.* Ensure that all educators are highly qualified to teach students at risk through licensure requirements, mentoring of new teachers, and incentives. Absorb teachers in training and in the field to hone and update their knowledge and document this with individual professional-development plans. It is vital to encourage stakeholder involvement to design and implement proven professional development activities. This can be done economically if districts are given targeted finances to fund long-term support and incentives to master teachers and other professionals.
- *Incorporate multiple uses of technology as a tool to raise student achievement.* Provide up-to-date computer technology to reform efforts that link technology use to career preparation and the future goals of youth with special needs. Improve and support competency in the use of technology by all students, and staff who serve them. Provide personal assistive technology (AT) such as computer software and hardware for audio/visual/learning impairments, and mobility devices for physical impairments, to students with special needs. Enable students with special needs to improve functions of daily living that present barriers to their pursuit of a challenging education and career.
- *Align educational standards at all levels and across districts and colleges.* Ensure that state policy aligns with national reform efforts that include evidence-based strategies. Preparing youth at risk to succeed in any state college requires the creation of common and multiple measures of readiness, and not just single high-stakes academic tests, for advancement to higher education and/or professional careers. Insist on the participation and accountability of stakeholders at all levels, from students and families, to teachers, to all administrators. Encourage stakeholder commitment to the design, implementation, and outcome evaluation of all standards alignment efforts.

Having the aforementioned elements in place and implemented in the K-12 educational system would greatly increase the entry of youth with special needs into post-secondary education. However, after post-secondary entrance, the provision of services and support possible in

K-12 schools is not matched in the new educational landscape. In order for individuals to access services in post-secondary settings, they must (1) disclose their disability, (2) provide documentation of their disability, and (3) request specific support, services, and accommodations necessary to succeed in their classes. Accomplishing these steps requires a skill set that many individuals with special needs do not have. Thus, the necessity of requesting for services becomes another barrier that may impede post-secondary education success. The issues around these steps arise during secondary education due to inadequate transition preparation. In addition to the difficulty of requesting for these services, primary gaps exist in the fragmented and inconsistent support services that are available in post-secondary-education settings. Suddenly, students with special needs may find themselves burdened with the necessity of essentially becoming their own case managers, a responsibility that may, for them, be insurmountable.

Participation, Retention, and Persistence in Post-Secondary Education

For most students, participation in post-secondary education goes beyond merely being physically present in a lecture hall. It is the opportunity to ask questions, to discuss ideas with classmates, to have a critical conversation with professors, to reflect upon readings, to explore the library, to have access to information, to work on research projects, to have coffee with friends, and be able to participate at campus social and cultural events that really completes the college experience. A quality education is about coming away from each campus experience with knowledge gained about, and insight into, a wide variety of human experiences and disciplines. Most critically for students with special needs, it is about being able to do these things without hardship that exceeds that of the typical student during the post-secondary educational years. For many of these students, however, issues surrounding self-advocacy and determination, social life needs, availability of educational assistance, differences in academic requirements, and limited preparation for post-secondary education become a primary preoccupation, and a distraction that affects academic performance (Burgstahler *et al.*, 2001). These problems are amplified in the presence of institutional inadequacies in providing services that would accommodate persons with special needs. Additional factors that come into play such as finances, availability of supports, socioeconomic status, culture, parental background, and school-commute distances often compound these challenges.

One factor affecting participation of persons in post-secondary education is the limited availability of educational and related supports within academic institutions. Once a student enters higher education, the lack of

student input and selective emphasis in servicing students may leave students with special needs at a disadvantage when it comes to exercising self-advocacy skills. Most supports are not directly individualized for the specific needs of individual students in post-secondary education. For example, teacher-centered, instead of student-centered, curricula and instructional methods have been the dominant approaches at the post-secondary level for a very long time. In this situation, the entrenched educational method reduces opportunities for students with special needs or less-traditional learning styles to thrive. Implementation of approaches that work for all learners becomes a critical issue in imparting knowledge that students need to thrive in the post-secondary setting and later in life.

For students who disclose their special needs and present appropriate documentation to support service offices, Section 504 of the Rehabilitation of 1973 mandates that post-secondary institutions provide reasonable accommodations to ensure full access to program offerings (Frank and Wade, 1993; West *et al.*, 1993). Post-secondary schools have no legal obligation to help students transition into their institution. Educators and institutions typically define their role with students more clearly in preparing them to succeed in future education and employment settings (Siegel and Sleeter, 1991). Furthermore, disability service staff members are often caught between the need to protect the rights of students and the needs of the post-secondary institution. Post-secondary administrations require consultation by staff members who also must adhere to particular state and federal guidelines. The lack of consensus on the nature of the supports to be provided can be difficult for staff as well as for students seeking accurate and complete information to determine which institutions and organizations are best prepared to meet their needs.

Another factor affecting participation in post-secondary education is the absence of any minimum standards of disability support provision. Many of the challenges students with special needs face are connected to necessary services and accommodations related to their specific needs. The lack of standardization of support services among academic institutions tends to result in differing levels and types of services and supports, and students are often left to manage complicated and unguided procedures on their own. Some institutions have single academic or counseling service staff members who take on disability issues as a small portion of their job description, and basically provide advice and a letter to professors verifying that a student's request for accommodation is justified. In contrast, in other schools, multiple staff persons, coordinated services, and accommodations are in place for students with special needs (Youth Advisory Committee to the National Council on Disability, 2003).

The lack of, or limited access to, support and its availability is a major factor that eventually discourages or excludes many students with special needs from continuing

their schooling. For instance, a national survey developed and distributed to post-secondary students with disabilities by the National Center for the Study of Postsecondary Educational Supports (2000) found that while supports such as testing accommodations, note-takers, personal counseling, and advocacy assistance were requested and extended with some regularity, disability specific scholarships, assessments and evaluations, real-time captioning, AT, and study-abroad opportunities were rarely offered to students with special needs (Stodden *et al.*, 2001). Furthermore, this study found that 50% of the respondents indicated that their institutions did not offer accessible transportation on campus for students with disabilities and many campuses still featured architectural barriers (Stodden *et al.*, 2001). The survey revealed that equal access and reasonable accommodations are still an issue for individuals with special needs attempting to continue in higher education, and often the most basic needs pertaining to their activities of daily living, including physical access, are unmet.

Several studies have confirmed these findings of institutional inadequacies and found that most post-secondary institutions are not at par in assisting youth with special needs (Stodden *et al.*, 2003a; Whelley *et al.*, 2002). A major problem encountered is the insufficient number of staff members handling huge caseloads to accommodate students. According to Gajar (1998), there are insufficient resources they are to serve the large numbers of youth with special needs who require case-management assistance and related support needs. Post-secondary educational institutions often do not provide per capita funding for disability support services, so as increased numbers of youth with special needs enter the various institutions, budgetary and faculty resources for support become thin and impact decisions.

Another factor affecting persistence and retention is that faculty members and other academic personnel in post-secondary education settings are often unaware of disability needs and supports. Limited awareness about their students' special needs prevents staff members and other academic personnel from providing the most suitable approach to enhancing the access and ability of students to learn. Moreover, the lack of proper background in managing students with special needs may invite further misunderstanding and conflict, and eventually lead these students to drop out. Finally, when accommodations and services are not provided by universities or colleges, students with special needs pursue post-secondary education at a higher risk for failure. Consequently, their grades and scholarship access are reduced, or they may have to pay for services themselves, thus increasing their cost of attendance in post-secondary education (the sum of which is the starting point for calculating student need for financial aid) (Youth Advisory Committee to the National Council on Disability, 2003).

Recommendations: Participation, Retention, and Persistence in Post-Secondary Education

The following are steps to achieve the important objectives of increasing participation, retention and persistence of students with special needs.

1. *Personnel development.* Improved faculty and staff (e.g., instructional faculty, financial-aid personnel, student-service personnel, and disability-support personnel) knowledge, and skills on a range of disability related topics such as:

- rights and responsibilities
- range of special needs (e.g., learning disability (LD), other hidden disabilities, and medical conditions)
- AT
- accommodations
- types of appropriate accommodation
- accommodations intended to level the playing field and not provide an unfair advantage; and
- universally designed instruction focused on content accessible for all students.

It is necessary for faculty and individuals who interact with individuals with special needs at the post-secondary education level to become knowledgeable in the areas mentioned above as relevant to each faculty and staff member's role and responsibilities.

2. *Increased connectedness with peers and other supportive individuals.* Feeling and being connected to others is an important consideration for everyone. Research has shown that a sense of belonging is a primary indicator of success in post-secondary education. Some strategies to improve post-secondary education institutions in this area include:

- Smaller learning communities that provide for cohorts of students to, at a minimum, take their prerequisite courses together. Results include increased retention and grade-point averages (GPAs) and higher levels of completion. Smaller learning communities are also shown to make the transition from secondary to post-secondary education easier.
- Disability support groups that may assist in helping individuals better understand their unique needs and provide peer support.
- Mentoring programs including reciprocal mentoring where students with special needs serve as mentors for post-secondary education faculty and staff and vice versa. In addition, peer-mentoring programs are also valuable in providing support for students as they maneuver through their post-secondary education.

3. *Increased access to technology.* The range of technology, including AT, on the market today has the potential to

increase access to and success in post-secondary education. However, individuals with special needs do not always have access to these tools. Many technologies that could specifically benefit individuals with special needs, such as electronic augmentative communication devices and voice-recognition software, are quite costly. Sometimes individuals or institutions invest in technology but these tools go unused because of a lack of adequate training. Finally, individuals with special needs can gain increased access if an effort is made to link them with agencies that provide financial support for AT purchase and training.

4. *Increased access to appropriate accommodations.* There is a substantial need to provide appropriate accommodations to students with special needs in post-secondary education and increase faculty and support staff knowledge and awareness regarding these accommodations. The most widely used accommodations include: extended time on tests, quiet rooms for testing, note takers in class, access to a computer to take exams, multi-modality instruction, flexible assignments, flexible deadlines, accessible classrooms, approval to turn in assignments late as appropriate to the disability and situation, sign-language interpreters, and materials available electronically for greater access.

Conclusion

Although gains have been made to support individuals with special needs in attending and complete post-secondary education, there are still numerous challenges to address. Positive changes will sustain only through concerted efforts by students, families, educators, administrators, and support staffers. First, we must address the realization that our expectations for students with special needs are too low and serve to artificially limit these students' horizons. In an increasingly competitive global economy, these limited horizons come at a great cost. Second, students need to be challenged and encouraged to view their education as continuing beyond secondary education. Third, special effort must be made to prepare students for transition from high school to postsecondary settings, and institutional capacity must be increased to support this. Specifically, students must be given training and opportunities to develop their self-determination and self-advocacy skills at the secondary level. Fourth, increased knowledge and awareness is necessary for students, faculty, and support staff regarding the provision of appropriate accommodations and a supportive and connected learning environment. By working toward these objectives and building on previous successes, we can re-envision post-secondary education as a realistic and productive option for persons with special needs,

and recreate our campuses as true models for inclusion and democracy.

Bibliography

- Abery, B. and Stancliffe, R. (1996). The ecology of self-determination. In Sands, D. J. and Wehmeyer, M. L. (eds.) *Self-Determination across the Life Span: Independence and Choice for People with Disabilities*, pp 111–145. Baltimore, MD: Paul H. Brookes.
- American Youth Policy Forum and Center on Education Policy (2002). Twenty-five years of educating children with disabilities: The good news and the work ahead. http://www.aypf.org/publications/special_ed.pdf (accessed August 2009).
- Berliner, D. and Biddle, B. (1996). Standards amidst uncertainty and inequality. *School Administrator* **53**(5), 42–47.
- Brinckerhoff, L. C. (1994). Developing effective self-advocacy skills in college-bound students with learning disabilities. *Intervention in School and Clinic* **29**(4), 229–237.
- Burgstahler, S. (2002). Universal design of distance learning. *Journal of Information Technology and Disabilities* **VIII**(1). <http://www.rit.edu/~easi/itd/itdv08n1/burgstahler.htm> (accessed August 2009).
- Burgstahler, S., Crawford, L., and Acosta, J. (2001). Transition from two-year to four-year institutions for students with disabilities. *Disability Studies Quarterly* **21**(1). <http://www.ds-qds.org/issue/archive>.
- Frank, K. and Wade, P. (1993). Disabled student services in postsecondary education: Who's responsible for what? *Journal of College Student Development* **34**, 26–34.
- Fuchs, D. and Fuchs, L. S. (1998). Peer-assisted learning strategies (PALS). <http://www.vanderbilt.edu/kennedy/pals> (accessed August 2009).
- Gajar, A. (1998). Postsecondary education. In Rusch, F. and Chadsley, J. (eds.) *Beyond High School: Transition from School to Work* pp 383–405. Belmont, CA: Wadsworth.
- Hatch, T. (1998). How comprehensive can comprehensive reform be? *Phi Delta Kappan* **79**, 518–522.
- Izzo, M. V., Hertzfeld, J. E., and Aaron, J. H. (2001). Raising the bar: Student self-determination + good teaching = success. *The Journal of Vocational Special Needs Education* **24**, 26–36.
- Izzo, M. V. and Lamb, P. (2002). Self-determination and career development: Skills for successful transition to postsecondary education and employment. http://www.ncset.hawaii.edu/publications/pdf/self_determination.pdf.
- National Center for the Study of Postsecondary Educational Supports (2000). *Quarterly Summary Report: October 1, 2000 to December 31, 2000*. Honolulu, HI: University of Hawai'i at Manoa, Center on Disability Studies.
- National Center for the Study of Postsecondary Educational Supports (2002). Preparation for and support of youth with disabilities in postsecondary education and employment: Implications for policy, priorities and practice. *Proceedings and Briefing Book Presented at the 2002 National Summit on Postsecondary Education for Persons with Disabilities*. <http://www.ncset.hawaii.edu/summits/july2002/briefs/docs/B12AccomDiffs.html>.
- National Council on Disability (2000). *National Disability Policy: A Progress Report: November 1, 1998 – November 19, 1999*. <http://www.ncd.gov/newsroom/publications/policy98-99.html> (accessed August 2009).
- Office of Special Education and Rehabilitative Services (2000). *IDEA Lessons for All!* <http://www.ed.gov/offices/OSERS/OSEP/IDEA25th/> (accessed November 2009).
- Siegel, S. and Sleeter, C. E. (1991). Transforming transition: Next stages for the school-to-work transition movement. *Career Development for Exceptional Individuals* **14**(1), 27–41.
- Stodden, R. A., Conway, M. A., and Chang, K. (2003). Professional employment for individuals with disabilities. <http://www.ncset.hawaii.edu/institutes/feb2003/papers/txt/PROFESSIONAL%20EMPLOYMENT%20.txt> (accessed November 2009).

- Stodden, R. A., Galloway, L. M., and Stodden, N. J. (2003b). Secondary school curricula issues: Impact on postsecondary students with disabilities. *Exceptional Children* **70**(1), 9–25.
- Stodden, R. A., Jones, M. A., and Chang, K. (2002). *Services, Supports and Accommodations for Individuals with Disabilities: An Analysis across Secondary Education, Postsecondary Education, and Employment*. University of Hawaii-Manoa, Center on Disability Studies. <http://www.rtc.hawaii.edu/documents/products/phase3/01.pdf> (accessed August 2009).
- Stodden, R., Whelley, T., Harding, T., and Chang, C. (2001). Current status of educational support provision to students with disabilities in postsecondary education. *Journal of Vocational Rehabilitation* **16**, 1–10.
- Thurlow, M. L. (2001). *Secondary Education Standards and Academic Preparation*. Minneapolis, MN: National Center on Education and Transition. ERIC Document Reproduction Service No. ED456584.
- West, M., Kregel, J., Getzel, E., et al. (1993). Beyond section 504: Satisfaction and empowerment of students with disabilities in higher education. *Exceptional Children* **59**, 456–467.
- Whelley, T., Hart, D., and Zafft, C. (2002). Coordination and management of services for individuals with disabilities in the transition from secondary to participation in postsecondary education and employment. *Paper Presented at the Meeting of the Capacity Building Institute*, Honolulu, HI.
- Youth Advisory Committee to the National Council on Disability (2003). *Record of Personal Meeting (Teleconference)*. National Council on Disability. <http://www.ncd.gov> (accessed August 2009).
- Stodden, R. A. and Roberts, K. D. (2006). Beyond high school: Preparing youth for employment. In Rusch, F. R. (ed.) *Policy Supporting Preparation and Transition to Post-School Environments: Past and Present*, pp 24–53. Belmont, CA: Wadsworth.
- Stodden, R. A., Roberts, K. D., Picklesimer, T., Jackson, D., and Chang, C. (2005). An analysis of assistive technology supports and services offered in postsecondary education institutions. *Journal of Vocational Rehabilitation* **22**(2), 1–10.
- Stodden, R. A. and Zucker, S. H. (2004). *Postsecondary Education and Youth with Cognitive Disabilities: Education and Training of the Developmentally Disabled (ETDD)*. Arlington, VA: Council for Exceptional Children.

Relevant Websites

- <http://www.abledata.com> – Abledata, AT Information.
- <http://www.aypf.org> – American Youth Policy Forum.
- <http://www.ahead.org> – Association on Higher Education and Disability (AHEAD).
- <http://www.cast.org> – Center for Applied Special Technology (CAST).
- <http://www.cep-dc.org> – Center on Education Policy.
- <http://www.cldinternational.org> – Council for Learning Disabilities.
- <http://www.eeonline.org> – Equity and Excellence Project.
- <http://www.heath.gwu.edu> – HEATH Resource Center, Online Clearinghouse on Postsecondary Education for Individuals with Disabilities.
- <http://www.ldonline.com> – LD Online.
- <http://www.ldanatl.org> – Learning Disabilities Association of America.
- <http://www.usd.edu> – Learning Style Inventories.
- <http://www.ncld.org> – National Center for Learning Disabilities.
- <http://www.engr.ncsu.edu> – NC State University, Index of Learning Styles Questionnaire.
- <http://www.nau.edu> – Northern Arizona University, Arizona Technology Access Program, The Assistive Technology Act of 2004.
- <http://www.olin.org> – Ohio Learning Network, FAME Teaching Workshops.
- <http://telr.osu.edu> – Ohio State University, Fast Facts for Faculty.
- <http://www.rtc.edu> – Renton Technical College.
- <http://www.samhsa.gov> – SAMHSA, Fact Sheet on Mental Illness and Violence.
- <http://www.ahead.org> – The Association on Higher Education and Disability (AHEAD).
- <http://www.rtc.hawaii.edu> – The Rehabilitation Training and Research Center at the University of Hawaii at Manoa.
- <http://www.ist.hawaii.edu> – UH OPE Project.
- <http://www.washington.edu> – University of Washington, DO-IT Project.
- <http://www.washington.edu> – University of Washington, Universal Design in Postsecondary Education.
- <http://www.w3.org> – W3C, Checklist for Accessible Websites.

Prereferral Strategies

C G Simpson, Sam Houston State University, Huntsville, TX, USA

J P Bakken, Illinois State University, Normal, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Alternative intervention – The strategies different from what are already being implemented to help student functioning in the general education classroom.

Accommodations – These provide access for students with disabilities to participate in a course, standard, or test. They do not fundamentally lower or alter the standard or expectation of the course or test.

Behavior interventionist – A specialist who develops interventions to help students with behavioral issues.

Instructional strategies – The attempts to remedy the identified educational concerns prior to the evaluation of a student for special education services.

Intervention – A teaching technique aimed at improving an academic or behavioral concern.

Intervention assistance teams – The teams made up of professionals with different areas of expertise who help analyze student data and suggest possible interventions to be implemented by the classroom teacher.

Modifications – The changes made to standards, test location, timing, courses, scheduling, expectations, and/or other attributes that provide access for a student with a disability to participate in a standard, course, or test, which does fundamentally alter or lower the standard or expectation of the test, course, or standard.

Pre-referral – A process that happens prior to the referral of a student to special education.

Referral – A process, to evaluate a student for special education services, which happens when pre-referral strategies are ineffective.

Strategies – Synonymous with interventions, these teaching techniques are aimed at improving an academic or behavioral concern.

are afforded the necessary accommodations and modifications in the general education setting prior to determining eligibility for special education services. Accommodations provide access for students with disabilities to participate in a course, standard, or test. They do not fundamentally lower or alter the standard or expectation of the course or test. Some accommodations include:

1. varying of activities, frequent breaks, and extended time;
2. preferential seating, physical arrangement of the room, and change in classroom setup;
3. shortening test length, and test format options (matching vs. fill in the blank); and
4. highlighting material, notes provided by the teacher, computer, calculator, and books on tape.

Modifications are changes made to standards, test location, timing, courses, scheduling, expectations, and/or other attributes that provide access for a student with a disability to participate in a standard, course, or test, which does fundamentally alter or lower the standard or expectation of the test, course, or standard. Some modifications include:

1. lowering the reading level of the test;
2. simplifying reading level and vocabulary on worksheets and assignments;
3. grading based on individualized education plan (IEP) goals; and
4. curriculum modified to meet user needs.

As it appears, accommodations level the playing field, whereas modifications change the playing field. Such accommodations/modifications are typically referred to as pre-referral intervention or instructional strategies and are recommended by the pre-referral assistance team (referred to as a pre-referral team, student assistant team, teacher assistance team, school-based intervention team, or child study team). This team is comprised of individuals with the knowledge and expertise to provide guidance and assistance to teachers. Members of the team might include classroom teachers, the special education teacher, school psychologist, student counselor, behavior interventionist, and principal. Regardless of the specific name given to the team, the purpose is to derive pre-intervention strategies or alternative intervention strategies. In research settings, pre-referral intervention teams have proven effective and worthwhile. Specifically, they have “(a) reduced special education referrals, (b) reduced unnecessary special

According to Truscott *et al.* (2005), “Contemporary special education programming and classification proposals place effective prereferral intervention at the center of special education delivery and diagnosis” (p. 130). The use of a pre-referral intervention process ensures that students

education testing, (c) resulted in more appropriate special education referral referrals, (d) resulted in improved student performance, and (e) resulted in improved teachers attitudes and handling difficult-to-teach students” (Truscott *et al.*, 2005: 131). Hammond and Ingalls (1999) summarized several critical factors obtained from various research projects that assure the effectiveness of such teams. These pre-referral intervention team’s success factors are listed below:

- administrative support for actions and activities;
- defined purpose exists which team members are aware;
- adequate training is provided in the implementation of the interventions;
- team members learn effective team skills;
- support from professional colleagues;
- effective communication among members;
- efficient system of operations; and
- efficacy of teacher time.

“Preintervention strategies are attempts to remedy the identified educational concerns prior to the evaluation of a student for special education services. Any intervention or combination of preintervention strategies should be used for a significant length of time; and objective, measureable data should be collected continuously in order to adequately assess the results” (Murdock and Petch-Hogan, 1996: 173). Pre-referral interventions are planned and systematic efforts to resolve apparent learning or behavioral problems initiated by the problem-solving team, and the design and outcomes of these interventions must be documented. The concept of the pre-referral assistance team is not a new one; however, “the intervention assistance reform effort was designed to be proactive and focus on providing more effective instruction to meet students needs in the general education classrooms” (Rock and Zigmond, 2001: 1). Clearly, the use of pre-referral assistance teams within a school can dramatically reduce the numbers of referrals into special education. In addition, “the teams have been extremely effective in helping to maintain a student, who is encountering academic or behavioral problems, within the general education setting” (Hammond and Ingalls, 1999: 2). The use of pre-referral teams has recently been given a greater emphasis in the referral process with the recent legislation of the Individuals with Disabilities Education Act (IDEA) 2004 in regard to response to intervention (RTI) in the identification of children with learning disabilities. RTI is an approach for sorting out whether a struggling student really is a student with a disability as defined by IDEA or just needs more intensive regular education strategies to succeed in school. It usually consists of three levels of assistance and may be used to see how students respond to research-based interventions and other direct supports.

Pre-referral intervention, as a critical instructional practice, can be implemented to reduce the overidentification

of students with disabilities (Kovaleski *et al.*, 1999). The use of a pre-referral intervention model “is an indirect delivery model of services that focuses on interventions occurring at the point of teacher referral” (Murdock and Petch-Hogan, 1996: 172). Clearly, “such a model allows not only for the prevention of inappropriate placements, but also for the provision of assistance to the general education teacher to enhance his/her interactions with students with disabilities” (Murdock and Petch-Hogan, 1996: 172). Research has indicated that everyone makes progress when best practices are followed by practitioners. When investigating the effectiveness of pre-referral interventions, the implications are very clear. First, assistance is provided to students in their home classrooms before special education is considered. Second, through the work of the pre-referral team, collaborative relationships are developed to enhance the effectiveness of programs. Third, effective pre-referral interventions will decrease the numbers of referrals to special education and can save money by having fewer inappropriate placements. Finally, educational programs can be enhanced to create a more inclusive school environment for all students (Buck *et al.*, 2003).

As previously mentioned, the pre-referral team focuses on the reasons a child may be experiencing difficulty related to behavioral or academic skills. Pre-referral interventions or strategies are derived from a brainstorming session in which the identified concerns lay the foundation. During this session, previous interventions employed within the classroom setting are analyzed along with student success using the strategies. Pre-referral strategies are then developed along with a specific plan as to how the strategies will be carried out. This plan would also involve timelines in which interventions should take place and how the success of the strategy will be monitored. The use of the pre-referral instructional strategies or interventions in the general education setting to target the needs of students with specific learning needs or behavioral problems are mandated by Part 200 Regulations. Section 200.2 (b) (7) and Part 200.4 (d) of the Commissioner of Education specifically state, “General education support services, instructional modifications, alternative instructional approaches, or alternative program options have been attempted to address the students performance prior to a referral to the committee on Special Education.” Therefore, not only do the pre-intervention strategies need to be developed and implemented, but the process must also be documented along with the impact the recommended strategies had on student learning.

The use of an intervention assistance team will assist the classroom teacher in developing instructional strategies that focus on the specific needs of the student. Well-developed intervention teams recognize the need for strong interventional strategies. “Research based practices must be incorporated into the teams’ recommendations to

increase the likelihood of intervention effectiveness” (Papalia-Berardi and Hall, 2007: 107). Questions that the teams should consider are listed below:

- When considering data collected on the student issue, what intervention should be tried to improve the situation?
- What interventions are most natural, least intrusive, and most effective?
- What training is required for the person who needs to implement the intervention?
- How might other students be impacted by the intervention?
- How will the proposed intervention improve the current situation?
- How will the team know if the intervention is effective?
- What resources are needed and how easily can this intervention be implemented?

Clearly, answering these questions will aid the team with action plan development, which is vital for the successful execution of intervention strategies. The action plan should describe, in detail, who is responsible for each intervention step, the best setting for implementation, dates for initiation of the intervention, and methods for assessing progress toward intervention goals. Data should be collected frequently and consistently to document progress or lack of it. These data will help the team to decide whether the intervention (1) was working, (2) should be continued, and, ultimately, (3) was successful.

Effective Pre-Intervention Strategies

Several studies have been conducted to determine the most effective and most frequently used intervention strategies. However, each of the suggested strategies should be viewed with caution as strategies should not be recommended based on the success rate of their use with other students as a sole reason of selection. Suggested strategies should only be employed after careful analyses of the student’s needs and strengths. Individualizing strategies to meet the needs of students will increase the success of the strategy. It is also imperative that strategies are based on the premise that “extra help for struggling learners must be more than additional practice” (Burns, 2007: 17). Whitten and Dieker (1995) surveyed 312 elementary schools in Illinois and found that only 178 schools had intervention assistance teams. Questionnaires completed by 83 schools identified the eight most frequently implemented strategies. These strategies respectively include behavior management, curricular modification, individualized instruction, small-group instruction, peer tutoring, consultation with professionals, teacher observations, and cooperative learning. Although these eight strategies were the most frequently occurring strategies, the order of

effectiveness of the strategies did not align to those strategies most frequently used. In order, the most successful strategies reported were behavioral management, peer tutoring, individualized instruction, small-group instruction, consultation with professionals, teacher/student conferences, teacher observations, and cooperative learning. Below are brief descriptions of these strategies:

- *Behavior management.* This entails implementing interventions to manage behavior. Examples include a behavior contract, token economy, reinforcement schedule, and choice menus.
- *Peer tutoring.* This involves having students who know strategies and content and they teach other students how to complete an assignment or task implementing the strategy. (Note: this does not replace classroom instruction by the teacher; it is intended to reinforce and supplement classroom instruction.) Students must also be trained (tutor and tutee) for this intervention to be effective.
- *Individualized instruction.* This type of instruction meets students’ needs by considering student’s knowledge and skill base before designing instruction. Each student gets what he/she needs to be successful in the classroom.
- *Small-group instruction.* This involves teaching three to four students together in a small group. Reinforcement of previously learned skills or new skill development can be emphasized. This decreases the teacher–student ratio and can better meet student academic and behavioral needs.
- *Consultation with professionals.* In this case, the teacher works with other professionals to capitalize on their expertise to help them teach content and skills and improve student behavior.
- *Teacher/student conferences.* These conferences entail frequently meeting with students to discuss progress, strengths, and areas in need of improvement; they can also be used to assess how instruction/learning is progressing in the classroom. It is important to remember that students have valuable information and insight in the instructional process.
- *Teacher observations.* These observations are conducted to see how progress is resulting and whether the intervention is effective. One must remember that when conducting observations it is important to operationally define the behavior to be observed as well as choose an accurate recording procedure (i.e., frequency, duration, time sampling, and interval) to collect data on the behavior.
- *Cooperative learning.* These learning groups consist of three to four students working toward a common goal or outcome. Typically, students are placed in groups by the teacher; and they are commonly assigned roles and responsibilities commensurate with their skills and

abilities. (Note: this does not replace classroom instruction, but is intended to reinforce and supplement classroom instruction.) Students must also be trained on the roles and responsibilities of each member and how to work collaboratively with each other.

Implications for Effective Pre-Referral Implementation

The above information is particularly interesting, considering the data that Lane *et al.* (2004) obtained in their investigation. In their study, they obtained 354 elementary teachers' views of the interventions generated by the pre-referral teams. Of the 799 interventions reported, 485 or 61.79% focused on academic issues, 102 or 12.77% dealt exclusively with behavior concerns, and others were a combination of academic and behavior issues. It is clearly evidenced that a majority of interventions are currently focusing on both academic and behavioral issues. Despite the types of interventions that are implemented, "there is little question that everyone gains when professionals follow best practices in their work. In the case of prereferral interventions, the implications for effective implementation are clear" (Buck *et al.*, 2003: 358). Below are implications for effective implementation of strategies:

- First, students receive assistance in their own classrooms prior to considering a special education placement as an option. For example, a teacher may give copies of PowerPoint presentations to the students so that they can follow along during the lecture.
- Administrators carefully and purposefully plan times when teachers can work together to develop a collaborative relationship which hopefully allows them to develop effective programs beyond concerns for pre-referral strategies. For example, teachers are given planning times simultaneously so they can actually meet and discuss student issues and plan for success.
- If teachers have time to plan and develop appropriate pre-referral strategies that are effective, the result should be a reduction of the number of inappropriate referrals to special education which will help schools and districts save money due to fewer inappropriate placements. For example, having pre-referral teams in place can help teachers to not refer students for special education, thus saving the district money if the student's needs are met in the general education classroom.
- A more inclusive school environment can be created by having an effective pre-referral strategy program in place because it can enhance educational programs for all students. For example, if the pre-referral strategy program is in place and effective, it can result in providing acceptable strategies for students to stay in the general

education classroom which can result in more effective programs for all students.

To a large measure, effectively implementing a pre-referral intervention process will allow the following factors, which may result in misidentification of students as needing special education services, to be more closely examined: lack of cultural background experiences, low socioeconomic levels, limited English-proficient, bilingual, students lacking educational experiences, and those students whose educational foundation or experiences were experienced with a teacher who may have a teaching style that did not match the students learning style.

Conclusion

Current research in the field reports that implementation of the pre-referral model has directly impacted assessment and referral processes. Specifically, pre-referral implementation has reduced special education referrals, decreased unnecessary special education referrals, provided more appropriate special education referrals, increased overall student performance, and improved specific teachers' attitudes and skills in handling difficult-to-teach students. However, without clear federal mandates regulating state guidelines in reference to the pre-referral process, a clear consensus will not be met. Pugach and Johnson (1989) stated that "the outcome of special education's current involvement with prereferral will likely be a good indicator of the path the field chooses to pursue." (p. 224). It is apparent that this outcome is still under investigation. Although the use of pre-referral interventions is common practice, the act or process of the implementation of such a model is still varied in nature. A systematic approach is necessary to measure the effectiveness of pre-referral interventions with regard to quality, accountability, and equity.

Bibliography

- Buck, G. H., Polloway, E. A., Smith-Thomas, A., and Cook, K. W. (2003). Prereferral intervention processes: A survey of state practices. *Exceptional Children* 69(3), 349-360.
- Burns, M. (2007). Nine ways to catch kids up. *Educational Leadership* 65, 16-21.
- Hammond, H. and Ingalls, L. (1999). Maintaining school based prereferral teams: An eight year study. *Rural Special Education Quarterly* 18(2), 17-21.
- Kovaleski, J. F., Gickling, E. E., Morrow, H., and Swank, P. R. (1999). High versus low implementation of instructional support teams: A case for maintaining program fidelity. *Remedial and Special Education* 20(3), 173-183.
- Lane, K. L., Pierson, M. R., Robertson, E. J., and Little, A. (2004). Teachers' views of prereferral interventions: Perceptions of and recommendations for implementation support. *Education and Treatment of Children* 27(4), 420-439.
- Murdock, N. and Petch-Hogan, B. (1996). Inclusive classroom management: Using preintervention strategies. *Intervention in School and Clinic* 31(3), 172-176.

- Papalia-Berardi, A. and Hall, T. (2007). Teacher assistance team social validity: A perspective from general education teachers. *Education and Treatment of Children* **30**(2), 89–110.
- Pugach, M. C. and Johnson, L. J. (1989). Prereferral interventions: Progress, problems, and challenges. *Exceptional Children* **56**(3), 217–226.
- Rock, M. and Zigmond, N. (2001). Intervention assistance: Is it substance or symbolism? *Preventing School Failure* **45**(4), 153.
- Truscott, S. D., Cohen, C. E., Smas, D. P., Sanborn, K. J., and Frank, A. J. (2005). The current state(s) of prereferral intervention teams. *Remedial and Special Education* **26**, 130–140.
- Whitten, E. and Dieker, L. (1995). Intervention assistance teams: A broader vision. *Preventing School Failure* **40**(1), 41–45.
- Rathvon, N. (2008). *Effective School Interventions: Evidence-Based Strategies for Improving Student Outcomes*, 2nd edn. New York: Guilford.
- Richards, C. and Leafstedt, J. (2009). *Early Reading Intervention: Strategies and Methods for Struggling Readers*. Upper Saddle River, NJ: Allyn and Bacon.
- Safran, S. P. and Safran, J. S. (1996). Intervention assistance programs and prereferral teams. *Remedial and Special Education* **17**(6), 363–369.
- Weishaar, M. C., Weishaar, P. M., and Budt, C. (2002). Prereferral intervention: Problem-solving method and curriculum-based measurement. *Rural Special Education Quarterly*. http://findarticles.com/p/articles/mi_qa4052/is_200207/ai_n9129463 (accessed August 2009).

Further Reading

- Coleman, J. G. (2006). *The Early Intervention Dictionary: A Multidisciplinary Guide to Terminology*. Bethesda, MD: Woodbine House.
- Garcia, S. B. and Ortiz, A. A. (2006). *Preventing Disproportionate Representation: Culturally and Linguistically Responsive Prereferral Interventions*. National Center for Culturally Responsive Educational Systems. <http://www.schoolsmovingup.net/cs/smu/view/rs/12347> (accessed August 2009).
- McCarney, S. B. (2006). *Pre-Referral Intervention Manual*. Columbia, MO: Hawthorne Educational Services.

Relevant Websites

- <http://www.fsd.k12.ca.us> – Fullerton School District, Pre-Referral Interventions for General Education.
- <http://www.greatschools.net> – Great Schools, Before Special Ed: How Pre-Referral Works.
- <http://www.familyvillage.wisc.edu> – The Family Village, Early Intervention Resources.
- <http://www.atsweb.neu.edu/cp/ei/>
- <http://earlyinterventionsupport.com/>
- <http://www.ncld.org/content/view/921/456099>

Preventing Antisocial Behavior and Delinquency: A Comprehensive School-wide Approach

P R Hallam, K R Young, P Caldarella, D G Wall, and L Christensen, Brigham Young University, Provo, UT, USA

© 2010 Elsevier Ltd. All rights reserved.

Antisocial behavior can be conceptualized as any conduct that is hostile to the well-being of an individual, group, system, or society (Walker *et al.*, 2004). When the antisocial behavior of youth leads to illegal or criminal acts, it is labeled as juvenile delinquency (Kauffman, 2005). To specifically define the problems regarding juvenile delinquency, we refer to the key findings of a report released by the United States Office of Juvenile Justice and Delinquency Prevention (Loeber *et al.*, 2003). The number of child delinquents (between the ages of 7 and 12) handled in the nation's juvenile courts has increased by 33% over the last decade (Synder, 2001; as cited in Loeber *et al.*, 2003: 1). Delinquent youth have a specific offense profile. They make up one-third of all juvenile arrests for arson, one-fifth of juvenile arrests for vandalism and sex offenses, one-eighth of juvenile arrests for forcible rape and burglary, and one-twelfth of juvenile arrests for violent crime. This report concluded that pre-teen youth with juvenile court records are "far more likely to become chronic juvenile offenders than youth whose initial contact occurs at a later age, [thus] there is reason for concern about the growing number of child delinquents" (Loeber *et al.*, 2003: 1).

Intervening with delinquent youth is one of the most difficult challenges for families, schools, and communities, primarily because there are no guaranteed solutions for preventing and treating delinquency, especially for youth with violent behaviors (Kauffman *et al.*, 2004). Historically, schools have struggled with educating pre-delinquent and delinquent youth. For example, incarcerated youth often have histories of school failure (Center on Crime, Communities, and Culture, 1997) and struggle with severe behavioral and learning difficulties not sufficiently identified or treated in schools (Nelson, 2000). Research also suggests that the earlier a child receives effective interventions, the greater the probability of success (Walker *et al.*, 2004). The goal should be to keep youth off the pathway to delinquency or reverse the direction as early as possible. Educational researchers strongly suggest a need for both prevention and intervention strategies that have been empirically demonstrated to be effective, as many current practices lack such validation (Nelson *et al.*, 2004). No single practice or approach has been proven to be highly effective in either preventing or intervening in juvenile delinquency. Therefore, a comprehensive school-wide approach is the preferred model. A comprehensive approach includes multiple components and requires the involvement and commitment of many individuals.

To be successful in addressing the needs of at-risk youth, schools must invest more than ever in implementing systems to properly intervene with students who are on the pathway to juvenile delinquency. We agree with Walker *et al.* (2004) who stated,

The problem is not one of not knowing what to do, but rather one of forging a marriage between research-based intervention programs and practices that are known to work, and also reforming school structures and systems in ways that support their effective use and sustainability over the long term (p. 33).

This article describes a prevention approach designed to produce system improvement by training school-based teams to gather and respond to data to prevent antisocial behavior and delinquency. This comprehensive approach includes the following components: (1) ensuring that the school has supportive administrative leadership, along with commitment and buy-in from faculty and staff; (2) providing high-quality needs-based professional development; (3) using academic and behavioral screening to identify at-risk youth; (4) using data systems and data-based decision-making practices; (5) using evidence-based strategies such as response to intervention or PBS; (6) integrating programs and planning into regular school operations through PLCs; and (8) implementing continuous program evaluation and modification. This approach combines research on best practices in delinquency prevention with the goal of increasing the capacity of school personnel to provide sustainable benefits to the organization and its students. Before further describing this approach, the needs and the challenges facing schools in implementing such a program are discussed.

Needs and Challenges Facing Schools

The United States has experienced a dramatic change in its demographics over the last 5 years. Educational leaders are required to prioritize decisions to emphasize helping students who are failing to meet rising academic standards. Increases in student populations and in the number of students who are at risk of failure, occurring in an environment of dwindling resources, are also challenging schools. The results include a growing trend for school failure and other issues that often accompany juvenile delinquency. Disaggregated student-achievement data

have revealed large gaps in achievement scores for students of poverty and minority ethnic groups (Ray-Taylor, 2003). For low-performing students, school failure often includes or evolves into emotional or behavioral problems, including antisocial behaviors, the precursors of delinquency (Office of Juvenile Justice and Delinquency Prevention, 2008).

School Failure, Juvenile Delinquency, and Antisocial Behavior

There is a strong association between school failure and juvenile delinquency. In 2003, there were over 96 000 youth in the United States incarcerated in juvenile correctional facilities (Office of Juvenile Justice and Delinquency Prevention, 2008). Large numbers of these youth struggle with literacy and have experienced academic failure in school (Nelson *et al.*, 2004). They are often male, socioeconomically disadvantaged, and members of minority groups (Quinn *et al.*, 2005). Many wrestle with significant behavioral and learning difficulties which may not have been adequately identified or treated in schools, leading some to call juvenile justice, the default system for students (1) who cannot read or write, (2) have mental health problems, and (3) drop out or are expelled from school (Nelson, 2000). In fact, 3 years after leaving school, 70% of antisocial youth are arrested (Walker *et al.*, 2004). The reality is that youth who struggle, grow up to be adults who struggle. People who drop out of school commit 82% of all crimes in the United States (American Psychological Association Commission on Violence and Youth, 1993). Walker *et al.* (1995) explained that the pathway to delinquency, criminality, and eventually prison begins very early in a child's life, often with early antisocial behavior.

Antisocial behavior is disruptive to the perpetrators, peers, families, schools, and communities; and it can take numerous forms – from noncompliance in early childhood to fighting in middle childhood and vandalism in adolescence. Although the typography varies, all forms of antisocial behavior share some common characteristics; they are unpleasant, destructive, or aversive to those victimized (Dishion and Patterson, 2006). Once established, antisocial behavior has proven to be highly resistant to intervention efforts in home, school, and community settings. According to Walker *et al.* (2004),

Kazdin (1987) has argued that, after about age 8, antisocial behavior and conduct disorder should be viewed as a chronic disorder (like diabetes) that has no cure but that can be controlled and managed. . . . Having said this, it is important to note that it is never too late to intervene with behaviorally at-risk children and youth (p. 10).

Therefore, early student identification and intervention are essential to the success of an effective prevention program. Antisocial behavior can thus be conceptualized

as the primary pathway to delinquency; and understanding this pathway is a requisite step in explaining delinquency.

From an abundance of research, four key factors contributing to antisocial behavior and delinquency have been identified (Jimerson *et al.*, 2006). First are individual factors, including problem-solving abilities, language development, cognitive processing, self-regulation skills, and mental-health symptomology. Second are peer factors, such as normative group behavior and affiliation with youth involved in antisocial conduct. Third are family factors, including the nature of social and disciplinary transactions, as well as communication patterns. Finally, there are school factors, such as the prevalence and nature of social supports and disciplinary programs. Considering the number and complexity of these factors, as well as their idiosyncratic interplay in the lives of youth, it is improbable that a single developmental pathway will ever be identified. Nevertheless, enough is currently known to organize and employ effective prevention and intervention efforts targeting antisocial behavior and delinquency in schools.

Schools are the most salient context for prevention and early-intervention efforts because they provide more access to more youth than any other setting. Although these targeted efforts could be implemented in various settings (e.g., home, clinic, hospital, or juvenile justice system), they are most commonly found within the school context (Sprague and Walker, 2000). Numerous school-based programs have been developed to reduce students' antisocial and delinquent behaviors by helping students gain prosocial skills. For example, a recent meta-analysis examining 83 school-based studies revealed that if such programs were comprehensive, they significantly reduced students' antisocial behavior, hostile attribution bias, hostile interpersonal-negotiation strategies, and aggressive fantasies (Derzon, 2006). Early intervention in the home, school, and community is the single best way of diverting children from the path of antisocial behavior (Walker *et al.*, 2004). Effective school-based prevention programs train faculty to identify and respond to student needs earlier than they presently do.

Prior Fragmented Efforts

For years, schools have been challenged with the question of how best to serve students who struggle academically and behaviorally. Unfortunately, fragmented initiatives have been implemented to resolve problems without examining the culture of schools and fully understanding their mission and priorities (Sarason, 1996). Greenberg *et al.* (2003) argued that these programs are unsuccessful for several reasons. First, programs promoting worthy prevention and lifestyle goals (e.g., drug and alcohol prevention, conflict resolution, and violence prevention) are often shortlived

because they are disconnected from the overall academic mission of the school. Second, these programs typically do not have strong administrative support; and personnel responsible for implementation are not professionally trained. In addition, these programs are rarely efficiently monitored and assessed for their effectiveness.

As it appears, schools have begun to prioritize needs and implement research-based programs that are designed to use data in making decisions on academic and behavioral matters. While schools have prioritized their resources to support such programs, many are either in the early stages of implementation or do not have an efficient system for tying these practices together in a comprehensive and coordinated approach. Comprehensive prevention programs are designed to enable organizers to prioritize and tie these services together through pooling community, higher education, and public-school resources to improve school success for at-risk students.

A Comprehensive School-Wide Approach

Greenberg *et al.* (2003) lamented that, “rather than integrating these segments and seeing the contributions of prevention programming to academic as well as social and emotional development, educators often make the false choice to emphasize academics only” (p. 472). In addition, they suggested designing and evaluating new programs that improve health and social–emotional skills as well as academic achievement of students. An effective prevention program includes seven critical components that help school-based teams screen students in both academic and behavioral dimensions and then use data

to make decisions that help prevent juvenile delinquency. The seven goals and objectives of successful school-wide comprehensive prevention programs are highlighted in **Table 1**.

Administrative Leadership and Faculty Buy-In

Administrative leadership (e.g., principal or district-level leadership) is essential for the success of any school-wide prevention effort (Kam *et al.*, 2003). Researchers at the Mid-Continent Research for Education and Learning (McREL) Institute (2005) conducted a meta-analysis of school-level leadership and its impact on student achievement. They found that school leadership has an indirect but significant impact on schools and on student learning (Marzano *et al.*, 2005). This and other evidence lead Killion (2000) to state that improving school leadership is important in resolving problems facing schools today.

Administrative leadership ensures that schools have access to resources and ongoing professional development necessary to implement programs with fidelity. Effective administrators communicate the vision and need for the program, inform their faculty what is required of them, discuss how the program will benefit the school, and then monitor progress by holding people accountable. Without such administrative leadership, any attempts at school-wide prevention will likely fail. Faculty and staff must also support school-wide prevention efforts. Faculty buy-in should be formally assessed via informational meetings and commitment from the majority of the faculty before implementing any school-wide prevention

Table 1 The seven goals and objectives of a comprehensive school-wide prevention program

Goals	Objectives
1. Ensure administrative support and faculty buy-in	1. Conduct informational meetings and obtain signed agreements from a majority of the faculty
2. Ensure high-quality professional development	2. Conduct an assessment to determine faculty readiness and professional-development needs
3. Enhance early identification of students who are academically and/or behaviorally at risk for school failure and juvenile delinquency	3. Ensure that faculty are trained in and implement academic and behavioral screening procedures, including a procedure that classifies students into low, medium, or high risk
4. Improve faculty access to students' academic and behavioral information via the district data systems	4. Ensure that faculty in schools are trained how to input and access students' academic and behavioral information via the district data systems
a. Improve faculty data-based decision-making (DBDM) skills to choose evidence-based interventions	a. Provide faculty training in academic and behavioral DBDM
	b. Enable weekly meeting of faculty-grade-level teams to apply DBDM practices
5. Implement evidence-based strategies	5. Provide faculty training in evidence-based strategies including response to intervention (RTI) and positive-behavior support (PBS)
6. Integrate program and planning into the regular school operations	6. Structure regular and consistent collaborative-grade-level professional learning communities (PLCs)
7. Provide continuous program evaluation	7. Conduct ongoing evaluation and modify practices accordingly

effort. Otherwise, the lack of a critical mass of faculty will make implementation weak and potentially ineffective (Fashola and Slavin, 1996).

Needs-Based Professional Development

Professional development is critical to the success of any prevention effort because advances in student learning cannot be expected without proportionate improvements in the quality of teaching (Guskey, 2002). Professional development offerings are most successful when they have (1) sufficient funding for implementation, (2) buy-in from participants, (3) recognition of beliefs and practices of participating faculty, (4) outside resources, and (5) common goals among participants (Richardson, 2003). Consideration of the school context is essential in designing training that meets the goals of creating teacher change and improving student outcomes (Guskey, 2003). An ongoing needs assessment based on both student data and faculty input should be conducted to ascertain what types of professional development activities are needed throughout the year.

When determining which types of professional-development programs are needed, some preliminary steps must be followed. First, data should be available to determine the needs of the school. The following might be considered:

- juvenile crime index of the school attendance area;
- student mobility rate;
- percentage of students who apply for free school lunch at each school;
- percentage of ethnic minority students at each school;
- percentage of limited-English-proficiency students at each school;
- percentage of students at each school from a single-parent family;
- percentage of faculty turnover;
- percentage of students with disabilities;
- daily attendance rate;
- proficiency levels of students in language, arts, math, and science;
- adequate yearly progress of students;
- achievement gap; and
- parental involvement.

The second step is to determine the school's strengths and needs by assessing the following variables:

- student-achievement data known and readily available;
- behavioral problems (who, when, and where) tracked and consequences followed;
- whether academically and behaviorally at-risk youths have been identified;
- whether school leadership values the need to use data to make decisions;

- whether school staff values the use of data in making decisions;
- collaboration teams that are in place and time available for collaboration and data-based decision making;
- whether a majority of the faculty are committed to the project; and
- the amount of support for the program that exists at the district level.

The results of these two steps help determine the types of professional development that schools need as a starting point.

A train-the-trainer model (Anderson *et al.*, 1996), ideally supported by district trainers who in turn train school trainers, is recommended in order to build sustainability and expansion to other schools within the district. This model is based on adult learning theory, which states that people who train others remember 90% of the material they teach, and diffusion of innovation theory, which states that people adopt new information through their trusted social networks (UCLA Center for Health Policy Research, 2008). For schools, it is suggested that the process begin by developing a cadre of skilled school-based trainers, who in turn train others in their school on how to plan and conduct the varied components of the prevention program. For example, we have found that school administrators, grade-level team leaders, and staff developers are ideal trainers who can provide onsite, ongoing support and follow-up. School-based trainers will also enhance the capacity of schools to use their existing database, screen for the early identification of students at risk for juvenile delinquency, and effectively use data to make intervention decisions.

Academic and Behavioral Screening

Screening is intended to alleviate the wait-to-fail approach often inherent in the educational process (Glover and Albers, 2007). Through systematic screening procedures, it is possible to assess, remediate, or prevent many of the factors that contribute to the development of antisocial behavior and delinquency (Sprague and Walker, 2000). With the most recent reauthorization of IDEA, there is a targeted emphasis on prevention and early intervention through the RTI and PBS models (US Department of Education, 2004). Early screening and identification are essential components of a successful prevention program, as they avoid waiting for students to fail before addressing their needs (Albers *et al.*, 2007).

Academic and behavioral screening make use of the data found in schools' data systems, including grades, criterion-referenced test scores, curriculum-based assessment results, office discipline referrals, suspensions, expulsions, safe school violations, and attendance data. Other screening criteria may be added to the data systems

as appropriate. For example, the dynamic indicators of basic early literacy skills (DIBELS), an academic screening instrument for reading performance originally developed by researchers at the University of Oregon, uses curriculum-based measurement procedures and functions as a general outcome measure (Deno and Mirkin, 1977; Fuchs and Deno, 1994). The tool is designed to assess and monitor the acquisition of early-literacy skills from kindergarten through sixth grade (Kaminski *et al.*, 2008). Research-based benchmark goals are used as part of DIBELS to provide educators and service providers with standards for assessing the progress of all students, including those at risk. Teachers administer the first assessment at the beginning of the year and use the results as a screening tool. The DIBELS assessments are then given two additional times during the school year, allowing educators to determine each student's progress on benchmark goals and make instructional changes as appropriate. Students' scores are assigned a category of progress that aids in planning appropriate evidence-based strategies to help them meet early-literacy goals.

Schools can use the systematic screening for behavior disorders (SSBD) (Walker and Severson, 1992) to help identify students who are behaviorally or emotionally at risk. The SSBD is a multistage process whereby teachers nominate, rank order, and then rate students from their classes who display internalizing or externalizing behavior problems. The SSBD has been shown to be useful to both elementary and secondary schools when attempting to identify students and behaviors that need attention from school personnel (Caldarella *et al.*, 2008; Marchant *et al.*, in press).

School administrators and program coordinators should collect both academic and behavioral screening data and monitor the percentage of students who are classified into high, medium, or low risk. Screening should at a minimum be conducted once in the beginning of the school year to identify students in need of extra supports. However, some types of screening data could be analyzed quarterly (e.g., grades and results of curriculum-based assessments) or even daily (attendance and office discipline referrals). The essence of effective screening is to use data as they become available to help identify at-risk students and make data-based decisions to prevent further decline in academic and behavioral performance.

Data Systems and Data-Based Decision Making

Schools need to have a feasible way to track student academic and behavioral data, as well as aggregate and disaggregate them easily. These functions can be accomplished through using an effective data collection and reporting system. Technology and data are among the most important supports needed for collaborative school-based teams.

In the absence of proper technological support, even the best-trained team will encounter barriers that continually impact the improvement of students' educational outcomes (Wayman *et al.*, 2006). It is suggested that schools use their existing database management system, assuming it is reliable and of high quality (e.g., efficient, accessible, and comprehensive). Doing so increases buy-in and reduces the perception that the school is always adding something new. Such systems must provide student data directly to classroom teachers to inform their daily instructional decisions.

However, the use of data at the school level is incredibly difficult because school personnel often lack proper systemic supports for data use (Stringfield *et al.*, 2001). For example, many teachers do not understand how to access, interpret, and use data to design appropriate interventions. Typically, each district has a data system, but the skill and understanding teachers have on how to access and use the system is often deficient. In fact, Holcomb (2004) identified the lack of proper training as one of the major reasons why data are not widely used by teachers and administrators. Thus, an important component of an effective prevention program is training on the collaborative use and analysis of data to determine and address individual student academic and behavioral needs.

School stakeholders need to share the belief that data are essential to instructional and behavioral decision making for students. An increasing amount of evidence supports the impact of access and use of academic and behavioral data on student outcomes (Stringfield *et al.*, 2001). When teachers and principals track student data systematically, they can make adjustments in educational practice that result in measurable improvements in student success. "Good data are as much a resource as staff, books, and computers" (National Forum on Education Statistics, 2004: 3). Through collection, analysis, and response to data, needs are recognized and addressed. When this information is incorporated into an improvement plan, impact on student outcomes is not only possible but likely (Jones and Mulvenon, 2003). Increasing the skill and capacity of teachers and administrators to understand and use data cannot be accomplished through the traditional one-time workshop because, as Holcomb (2004) suggested, "Understanding how to use data to guide decision making is far more complex than individual knowledge and skill acquisition, and it must be embedded in the ongoing work of groups of professionals" (p. 30). Thus, an important component of an effective prevention program is training on the collaborative analysis and use of data from which individual student needs can be determined. Such a program incorporates training in data-based decision making, based on a problem-solving model consisting of a set of sequential and repeatable steps: for example, (1) define and analyze the problem within context, (2) develop a hypothesis and determine possible interventions, (3) monitor progress and evaluate the effectiveness of

interventions, and (4) repeat the cycle as needed (Lau *et al.*, 2006). Utilizing a problem-solving model such as this has been demonstrated to effectively facilitate students' academic achievement and behavioral improvement (Shinn, 2007).

Evidence-Based Strategies

In order to facilitate the most efficient and effective prevention or intervention effort possible, schools should be taught and encouraged to use evidence-based strategies and ongoing progress monitoring (Ikeda *et al.*, 1996). When school faculty find that evidence-based strategies benefit their students, they are more likely to implement these practices (Biglan *et al.*, 2003). Two evidence-based strategies that schools should consider using to aid in prevention and intervention efforts are RTI and PBS.

RTI is an integrated approach that encompasses general, remedial, and special education through a multi-level service-delivery model. It utilizes a problem-solving framework to identify and address academic and behavioral difficulties for all students via evidence-based interventions. RTI is the practice of (1) providing high-quality instruction and intervention that is matched to each students' needs, and (2) using both learning rate over time and student-performance level to make essential educational decisions that guide classroom instruction (Batsche *et al.*, 2005). RTI practices are proactive, incorporating both prevention and intervention. As it stands, RTI is intended to reduce the number of instructional casualties by providing high-quality instruction and interventions matched to student needs (Pieranelo and Giuliani, 2008: 60). Instructional interventions are monitored frequently to make decisions about changes in instruction (National Association of State Directors of Special Education, 2006). By using RTI, districts and school teams can provide interventions to students as soon as a need is identified, in contrast to the traditional approach associated with the

aptitude-achievement discrepancy models used for identifying specific learning disabilities – sometimes criticized as a wait-to-fail approach (Glover and Albers, 2007).

PBS is another comprehensive multilevel school-wide approach. The primary level of PBS, also called universal support, consists of a wide range of strategies designed to achieve positive social and learning outcomes and to prevent problem behavior (Lewis and Sugai, 1999). These strategies include clearly articulated rules, opportunities for teaching and rehearsing expected behaviors and procedures, pre-correction reminders to alert students to expectations, and high rates of positive recognition – all of which are procedures for preventing problem behavior (Sugai and Horner, 2002). The secondary level comprises selected interventions designed to address the needs of students identified as emotionally or behaviorally at risk. These interventions may include small group instruction or individualized support such as anger management, social and emotional learning, or mentoring. For students identified as highly at risk, a tertiary level of PBS is designed to provide more intensive interventions such as individualized function-based assessment and intervention or wrap-around services (e.g., family and community resources).

As described in the previous section, the three-tiered model on which RTI and PBS are based uses preventative approaches to create positive school environments (Sugai *et al.*, 2002) (see **Figure 1**). This model is based on research indicating that approximately 80% of students respond to universal interventions. Targeted interventions can provide specific services and support to an estimated 10–15% of students who may have been labeled as at risk and projected to benefit from services such as small group instruction. The more intensive individual interventions target approximately 1–5% of students who require highly focused assessment and intervention services, including those with educational disabilities (Horner and Sugai, 2002; Sugai and Horner, 2002). RTI has typically focused on academics, while PBS has

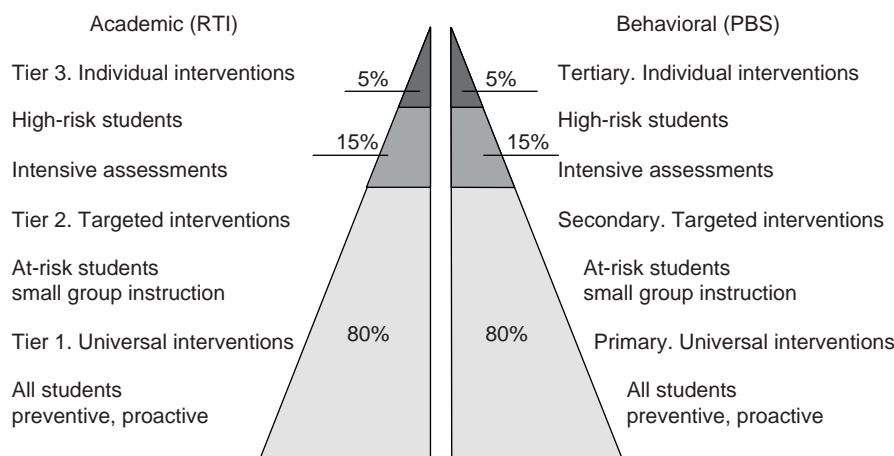


Figure 1 The three-tiered model on which response to intervention (RTI) and positive-behavior support (PBS) are based.

focused on behavior. However, in any prevention or intervention program targeting juvenile delinquency, both should be addressed. For example, at-risk students often struggle behaviorally and academically; thus grade-level teams may realize they need additional expertise for choosing, implementing, and evaluating evidenced-based interventions to address both needs. This could prompt the administrator or program coordinator to arrange for the use of additional school or district resources such as PBS or RTI.

Integration with Regular School Operations

For any prevention model to be successful, the critical components must be integrated into the regular operations of the school. Such integration requires teachers and administrators to continually focus on students they have identified as at risk and then monitor and adjust their prevention strategies as necessary. This process is very different from the one-shot workshop-driven approach common in schools. The National Staff Development Council (2009) suggested that the most powerful forms of staff development occur in regularly scheduled collaborative teams that meet to learn, plan, and problem solve together for the purpose of improving student learning. Collaboration is supported by many researchers as an efficient way to improve education and recommended as an effective method of placing a student-centered focus on school data (Schmoker, 2004). Earlier, DuFour (2002) reported that in the Chicago Public School system, at-risk schools with strong PLCs were four times more likely to improve academic outcomes than were schools with weaker faculty collaboration. As Schmoker (2004) noted, "Though such terms as 'learning communities' and 'lesson study' are heard more than ever, we hardly acknowledge their central importance in actual practice . . . [as a] time for teachers to create, test, and refine their lessons and strategies together" (p. 431).

PLCs provide embedded professional development for teachers and administrators involved in school-improvement efforts such as delinquency prevention. PLCs are composed of collaborative teams of teachers and other school professionals who meet regularly (ideally weekly) to examine the academic and behavioral progress of their students (DuFour and Eaker, 1998). They concentrate on the needs of students by seeking out and implementing effective practices with the goal of increasing student achievement. Teachers and administrators work together toward common objectives creating a results orientation by which learning for all students is expected and their progress is celebrated. Schools working as PLCs have a shared vision, engage in collective inquiry, are action oriented, and expect continuous improvement and results. While successful PLCs focus on achieving these characteristics, DuFour (2004) suggested they also

continually ask these three crucial questions: (1) What do we want each student to learn? (2) How will we know when each student has learned it? and (3) How will we respond when a student experiences difficulty in learning?

DuFour (2004) summarized how to create a PLC by advising participants to "focus on learning rather than on teaching, work collaboratively, and hold yourself accountable for results" (p. 1). Proponents believe PLCs offer the most hope for sustained substantive school improvement (DuFour, 2007; McLaughlin, 1995). Fragmentation and incoherence are cited by the National Staff Development Council (2009) as the most common threats to effective staff-development approaches. Thus, PLC teams that have a clear vision of student learning and are focused on continuous improvement and experimentation eliminate these threats; and school improvement can move forward at a quicker pace.

Program Evaluation and Modification

The continuous evaluation of student outcomes is critical to the success of any delinquency-prevention program. Both formative and summative assessments are important in program evaluation, as illustrated in the following examples. Formative assessment could be accomplished as a part of weekly PLCs where team members track academic and behavioral changes in students classified in each of the three at-risk categories and make instructional changes based on this feedback. Summative assessment could be accomplished by comparing data from the beginning of the year with data from the end of the year (pre/post comparison) to determine whether the program is meeting the goals of decreasing risk (antisocial behaviors and office discipline referrals) and increasing resiliency (academics and social skills). If the program is not achieving its goals, an important issue to assess as part of the evaluation is treatment integrity, or whether the intervention is implemented as intended (Lane and Beebe-Frankenberger, 2004). Interventions which are not implemented properly are less likely to achieve the desired outcomes. In addition, social validity – the social significance, acceptability, and importance of the intervention – must also be assessed, as interventions with low social validity also tend to have low treatment fidelity (Lane and Beebe-Frankenberger). This is especially critical for students who come from socioeconomically disadvantaged or culturally and linguistically diverse backgrounds.

Conclusion

This article has explored problems associated with juvenile delinquency and challenges facing schools as they attempt to implement early-intervention strategies for students identified as at risk. Research regarding the

prevention of antisocial behavior and delinquency has been reviewed to synthesize best practices to develop a comprehensive school-wide approach. We have laid out goals and objectives of such a program and described each of its seven components in detail. Surely, we agree with other experts who have concluded that prevention is a better approach than remediation for addressing juvenile delinquency, and the earlier the better (Kauffman, 2005; Loeber *et al.*, 2003; Walker *et al.*, 2004). The prevention program outlined in the article seeks to target the antecedents of delinquency by implementing a comprehensive school-wide approach that includes (1) ensuring that the school has supportive administrative leadership, with commitment and buy-in from faculty and staff; (2) providing high-quality needs-based professional development; (3) using academic and behavioral screening for the identification of at-risk youth; (4) using data systems and data-based decision-making practices; (5) using evidence-based strategies such as RTI and PBS; (6) integrating programs and planning into regular school operations through PLCs; and (7) implementing continuous program evaluation and modification. As at-risk students are identified early, evidence-based strategies implemented, and student responses monitored and tracked regularly, incidents of juvenile delinquency may be dramatically reduced.

Bibliography

- Albers, C. A., Glover, T. A., and Kratochwill, T. R. (2007). Introduction to the special issue: How can universal screening enhance educational and mental health outcomes? *Journal of School Psychology* **45**(2), 113–116.
- American Psychological Association, Commission on Violence and Youth (1993). *Violence & Youth: Psychology's Response, Vol. 1: Summary Report*. Washington, DC: Author.
- Anderson, J. L., Russo, A., Dunlap, G., and Albin, R. W. (1996). A team training model for building the capacity to provide positive behavioral supports in inclusive settings. In Koegel, L., Koegel, R., and Dunlap, G. (eds.) *Positive Behavioral Support: Including People with Difficult Behavior in the Community*, pp 467–490. Baltimore, MD: Brookes.
- Batsche, G., Elliot, J., Graden, J. L., *et al.* (2005). *Response to Intervention: Policy Considerations and Implementation*. Alexandria, VA: National Association of State Directors of Special Education.
- Biglan, A., Mrazek, P. J., Carnine, D., and Flay, B. R. (2003). The integration of research and practice in the prevention of youth problem behaviors. *American Psychologist* **58**, 433–440.
- Caldarella, P., Young, E. L., Richardson, M. J., Young, B. J., and Young, K. R. (2008). Validation of the systematic screening for behavior disorders in middle and junior high school. *Journal of Emotional and Behavioral Disorders* **16**, 105–117.
- Center on Crime, Communities, and Culture (1997). *Education as Crime Prevention*. (Occasional Paper Series No. 2). New York: Author.
- Deno, S. L. and Mirkin, P. K. (1977). *Data-Based Program Modification: A Manual*. Reston, VA: Council for Exception Children.
- Derzon, J. (2006). How effective are school-based violence prevention programs in preventing and reducing violence and other antisocial behaviors? A meta-analysis. In Jimerson, S. R. and Furlong, M. J. (eds.) *Handbook of School Violence and School Safety: From Research to Practice*, pp 429–441. Mahwah, NJ: Erlbaum.
- Dishion, T. J. and Patterson, G. R. (2006). The development and ecology of antisocial behavior in children and adolescents. In Cicchetti, D. and Cohen, D. J. (eds.) *Developmental Psychopathology, Vol. 3: Risk, Disorder, and Adaptation*, pp 503–541. New York: Wiley.
- DuFour, R. (2002). The learning principal. *Educational Leadership* **59**(8), 12–15.
- DuFour, R. (2004). Schools as learning communities. *Educational Leadership* **61**(8), 6–11.
- DuFour, R. (2007). Professional learning communities: A bandwagon, an idea worth considering, or our best hope for high levels of learning? *Middle School Journal* **39**(1), 4–8.
- DuFour, R. and Eaker, R. (1998). *Professional Learning Communities at Work: Best Practices for Enhancing Student Achievement*. Bloomington, IN: Solution Tree.
- Fashola, O. S. and Slavin, R. E. (1997). Promising programs for elementary and middle schools: Evidence of effectiveness and replicability. *Journal of Education for Students Placed at Risk* **2**(3), 251–307.
- Fuchs, L. S. and Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children* **61**, 15–24.
- Glover, T. A. and Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology* **45**(1), 117–135.
- Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., *et al.* (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist* **58**, 466–474.
- Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice* **8**(3/4), 381–391.
- Guskey, T. R. (2003). The characteristics of effective professional development: A synthesis of lists. *Paper Presented at the 84th Annual Meeting of the American Educational Research Association* (ERIC Document Reproduction Service No. ED 478 380), Chicago, IL.
- Holcomb, E. L. (2004). *Getting Excited About Data*, 2nd edn. Thousand Oaks, CA: Corwin Press.
- Horner, R. H. and Sugai, G. (2002). *School-Wide Positive Behavior Support: Implementer Blueprint and Self-Assessment*. Eugene, OR: OSEP Center on Positive Behavior Support.
- Ikedda, M., Tilly, W. D. III, Stumme, J., Volmer, L., and Allison, R. (1996). Agency-wide implementation of problem-solving consultation. *School Psychology Quarterly* **11**, 228–243.
- Jones, M. and Mulvenon, S. (2003). *Leaving No Child Left Behind: Using Data Driven Decision Making in School Systems*. Phoenix, AZ: Allstar Publishing.
- Jimerson, S. R., Morrison, G. M., Pletcher, S. W., and Furlong, M. J. (2006). Youth engaged in antisocial and aggressive behaviors: Who are they? In Jimerson, S. R. and Furlong, M. J. (eds.) *Handbook of School Violence and School Safety: From Research to Practice*, pp 3–19. Mahwah, NJ: Erlbaum.
- Kam, C. M., Greenberg, M. T., and Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science* **4**(1), 55–63.
- Kaminski, R., Cummings, K. D., Powell-Smith, K. A., and Good, R. H. (2008). Best practices in using dynamic indicators of basic early literacy skills for formative assessment and evaluation. In Thomas, A. and Grimes, J. (eds.) *Best Practices in School Psychology V*, pp 1881–1884. Bethesda, MD: National Association of School Psychologists.
- Kauffman, J. (2005). *Characteristics of Children with Emotional and Behavioral Disorders*, 8th edn. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Killion, J. (2004). Building for success: State challenge grants for leadership development. *Report of a Study by the National Staff Development Council*, Oxford, OH.
- Lane, K. L. and Beebe-Frankenberger, M. (2004). *School-Based Interventions: The Tools You Need to Succeed*. Boston, MA: Allyn and Bacon.
- Lau, M. Y., Sieler, J. D., Muyskens, P., *et al.* (2006). Perspectives on the use of the problem-solving model from the viewpoint of a school psychologist, administrator, and teacher from a large Midwest urban school district. *Psychology in the Schools* **43**(1), 117–127.

- Lewis, T. and Sugai, G. (1999). Effective behavior support: A systems approach to proactive school-wide management. *Focus on Exceptional Children* **31**(6), 1–24.
- Loeber, R., Farrington, D. P., and Petechuk, D. (2003). *Child Delinquency: Early Intervention and Prevention (Child Delinquency Bulletin Series)*. Washington, DC: US Department of Justice, Office of Juvenile Justice and Delinquency Prevention.
- Marchant, M. R., Anderson, D., Caldarella, P., et al. (in press). School-wide screening and programs of positive behavior support: Informing the intervention process. *Preventing School Failure*.
- Marzano, R. J., Waters, J. T., and McNulty, B. A. (2005). *School Leadership That Works: From Research to Results*. Alexandria, VA: Association for Supervision and Curriculum Development.
- McLaughlin, M. (1995). Creating professional learning communities. *Keynote Address at the Annual Conference of the National Staff Development Council*, Chicago, IL.
- National Association of State Directors of Special Education (2006). *Response to Intervention: Policy Considerations and Implementation*. Alexandria, VA: Author.
- National Forum on Education Statistics (2004). *Forum Guide to Building a Culture of Quality Data: A School and District Resource*. (NFES 2004–800). Washington, DC: National Center for Education Statistics.
- National Staff Development Council (2009). *Learning Communities*. <http://www.nsdc.org/standards/learningcommunities.cfm> (accessed August 2009).
- Nelson, C. M. (2000). Educating students with emotional and behavioral disabilities in the 21st century: Looking through windows, opening doors. *Education and Treatment of Children* **23**, 204–222.
- Nelson, C. M., Leone, P. E., and Rutherford, R. B. (2004). Youth delinquency: Prevention and intervention research. In Rutherford, R. B. Jr., Quinn, M. M., and Mathur, S. R. (eds.) *Handbook of Research in Emotional and Behavioral Disorders*, pp 282–301. New York: Guilford.
- Office of Juvenile Justice and Delinquency Prevention (2008). How many juveniles are held in residential placement on a given day in the U.S. <http://ojjdp.ncjrs.gov/ojstatbb/corrections>.
- Pieranelo, R. and Giuliani, G. (2008). *Classroom Management for Students with Emotional and Behavioral Disorders: A Step-by-Step Guide for Educators*. Thousand Oaks, CA: Corwin Press.
- Quinn, M. M., Rutherford, R. B., Leone, P. E., Osher, D. M., and Poirier, J. M. (2005). Students with disabilities in detention and correctional settings. *Exceptional Children* **71**(3), 339–345.
- Ray-Taylor, R. (2003). NCLB offers challenges and opportunities for MSAN districts. *Network News: Newsletter of the Minority Student Achievement Network* **12**, 1–5.
- Richardson, V. (2003). The dilemmas of professional development. *Phi Delta Kappan* **84**(5), 401–406.
- Sarason, S. B. (1996). *Revisiting the Culture of the School and the Problem of Change*. New York: Teachers College Press.
- Schmoker, M. (2004). Tipping point: From feckless reform to substantive instructional improvement. *Phi Delta Kappan* **85**, 424–432.
- Shinn, M. R. (2007). Identifying students at risk, monitoring performance, and determining eligibility with response to intervention: Research one educational need and benefit from academic intervention. *School Psychology Review* **36**, 601–617.
- Sprague, J. and Walker, H. M. (2000). Early identification and intervention for youth with antisocial and violent behavior. *Exceptional Children* **66**, 367–379.
- Stringfield, S., Reynolds, D., and Schaffer, E. (2001). *The High Reliability Schools Project*. Jolimont: IARTV.
- Sugai, G. and Horner, R. H. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child and Family Behavior Therapy* **24**(1/2), 23–50.
- Sugai, G., Horner, R., and Gresham, F. M. (2002). Behaviorally effective school environments. In Shinn, M., Walker, H., and Stoner, G. (eds.) *Interventions for Academic and Behavior Problems II: Preventive and Remedial Approaches*, pp 315–350. Bethesda, MD: National Association of School Psychologists.
- UCLA Center for Health Policy Research (2008). *Health Data Train the Trainer Project*. <http://www.healthpolicy.ucla.edu/healthdata> (accessed August 2009).
- Walker, H. M., Colvin, G., and Ramsey, E. (1995). *Antisocial Behavior in School: Strategies and Best Practices*. Pacific Grove, CA: Brooks/Cole.
- Walker, H. M., Ramsey, E., and Gresham, F. (2004). *Antisocial Behavior in School: Evidence-Based Practices*, 2nd edn. Pacific Grove, CA: Brooks/Cole.
- Walker, H. M. and Severson, H. H. (1992). *Systematic Screening for Behavior Disorders*, 2nd edn. Longmont, CO: Sopris West.
- Wayman, J. C., Midgley, S., and Stringfield, S. (2006). Leadership for data-based decision-making: Collaborative data teams. In Danzig, A., Borman, K., Jones, B., and Wright, B. (eds.) *New Models of Professional Development for Learner Centered Leadership*, pp 189–206. New York: Erlbaum.

Further Reading

- Danzig, A., Borman, K., Jones, B., and Wright, B. (2006). *New Models of Professional Development for Learner Centered Leadership*. New York: Erlbaum.
- DuFour, R., DuFour, R., Eaker, R., and Many, T. (2006). *Learning by Doing: A Handbook for Professional Learning Communities at Work*. Bloomington, IN: Solution Tree.
- IDEA (2004). *Individuals with Disabilities Education Improvement Act of 2004*. <http://idea.ed.gov> (accessed August 2009).
- Kurns, S. and Tilly, W. D. (2008). *Response to Intervention Blueprints: School Building Level Edition*. Alexandria, VA: National Association of State Directors of Special Education.
- Rutherford, R. B., Quinn, M. M., and Mathur, S. R. (2004). *Handbook of Research in Emotional and Behavioral Disorders*. New York: Guilford.

Programs and Instructional Strategies for Students with Gifts and Talents

V I Daniels, Southern University and A & M College, Baton Rouge, LA, USA

M J McCollin, Slippery Rock University of Pennsylvania, Slippery Rock, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Accelerated cognitive and intuitive abilities –

The ability to easily and efficiently acquire, retain, conceptualize, synthesize, and learn new information; as well as process and manipulate large amounts of information at an accelerated pace.

Accelerated instruction – Students with above average academic and intellectual aptitude receive modified/adapted instruction that provides them the opportunity to advance more rapidly through their schooling or to finish schooling at an earlier age than their peers. The curriculum is clearly adapted to the higher aptitude level of students in these programs.

Augmented curriculum – A wide-ranging scope of resources (including out of grade level print and non-print materials) is made available to enhance the curriculum and supplement independent learning opportunities for students with gifts and talents.

Curriculum models – A set of planned cognitive, social, cultural, and emotional experiences that meet the needs of students who are gifted and talented.

Compacted curriculum – Learning experiences that are accelerated in a developmentally appropriate (not redundant) method to meet the needs, interests, and abilities of students who are gifted and talented students.

Differentiated instruction – Differentiated instruction provides for the accelerated development of critical, creative, problem solving and research skills, advanced content, and authentic and appropriate products for gifted students.

Gifted education – Special practices, procedures and theories used to teach children who are gifted and talented.

Gifted and talented – Children and youth with outstanding talent or intellectual ability who perform or show the potential for performing at remarkably high levels of accomplishment or achievement.

Individuals with exceptional abilities have been celebrated throughout history. Our understanding and description of giftedness, however, has changed over time (Smith *et al.*, 2004; VanTassel-Baska and Brown, 2009). Early twentieth-century giftedness studies emanated from research on

the inheritance of mental incompetence, below normal academic achievement among children, development of assessment instruments to measure both the below- and above-average intelligence, and the realization that public schools could not adequately meet the needs of all children (Coleman and Cross, 2005). Terman and Hollingworth, two of the earliest pioneers in the area of gifted research, conducted some of the earliest and most widely published analyses on gifted children. The discipline of gifted education has evolved in relation to the emerging scientific and technological needs of the country, particularly those emanating after the Soviet Union's launch of Sputnik in the late 1950s (Turnbull *et al.*, 2004).

National interests in the education of students with gifts and talents led to the commission of several highly publicized reports, namely the Marland (1972) Report, *A Nation at Risk* (1983), *National Excellence* (1993), and *A Nation Deceived* (2004). The highly acclaimed Marland (1972) Report, defined giftedness as capacity for high performance, demonstrated achievement, and/or potential ability in any of the areas of general intellectual ability, specific academic aptitude, creative or productive thinking, leadership, visual and performing arts, and psychomotor abilities. The report also provided assessments of the needs of students with gifts and resulted in direct assistance to gifted education and the establishment of the US Office of the Gifted and Talented. This definition of giftedness continues to be the foundation of many gifted programs in most districts and states (Karnes and Beane, 2009). Earlier in 1983, *A Nation At Risk* presented startling data that indicated that students in the United States were no longer receiving a superior education, and could not compete globally (Coleman and Cross, 2005). Several of the many report recommendations included but were not limited to: (1) standards for identification and servicing of students with gifts, (2) increasing services to gifted-education programs, (3) improving curriculum enrichment, and (4) creating accelerated instructional approaches (VanTassel-Baska, 1998; VanTassel-Baska and Brown, 2009).

Consequently, in 1988, Congress passed the Jacob K. Javits Gifted and Talented Student Education Act. This Act provided financial support and resources for the Office of Gifted and Talented Education, a national research center focusing on children with gifts, and demonstration projects in gifted education. In 1990, a National Center on Gifted and Talented was established (Turnbull

et al., 2004). The No Child Left Behind Act of 2001 (NCLB, P. L. 107-110), formerly known as the Elementary and Secondary Education Act (ESEA), was enacted to eliminate the achievement gap that exists among students in US schools and ensure that children of all abilities and racial, ethnic, cultural, and socioeconomic backgrounds have a fair, equal, and significant opportunity to obtain a high-quality education, and be taught by highly qualified teachers. In *A Nation Deceived* (2004), Colangelo, Assouline, and Gross described the advantages of accelerated programs for students with gifts and talents and further highlighted America's inability to effectively address the academic, psychosocial, and emotional needs of its most able students in spite of the overwhelming research supporting accelerated programming and differentiated instructional practice. The twenty-first century provides an opportunity of limitless possibilities and hopefully a national focus on the education of children who are gifted and talented.

Characteristics of Gifted and Talented

The study of giftedness encompasses both children and adults and it involves the investigation of cognitive, affective, intuitive, and psychomotor attributes. Much of the research on giftedness (e.g., Clark, 2008; Hallahan and Kauffman, 2003; Smith *et al.*, 2004) describes children with gifts and talents as manifesting a wide range of abilities, including accelerated cognitive and intuitive abilities, a phenomenal memory, excellent recall, visual and performing abilities, and psychomotor intelligence. These children also have the ability to quickly acquire, retain, conceptualize, synthesize, and learn new information as well as the ability to easily process and manipulate large amounts of information at an accelerated pace.

These children are also highly inquisitive. Not only do they enjoy learning, they also have a strong desire to learn and an intense need for mental stimulation. They have the innate ability to learn faster and are more intent and focused on pursuing their interests, a capacity for seeing unusual and diverse relationships, as well as an unusual ability to think abstractly. To a large measure, these children possess superior problem-solving skills, can easily apply and transfer knowledge to new situations, and show an interest in a wide range of topics and disciplines (Smith *et al.*, 2004).

Other characteristics manifested by children with gifts and talents include (1) the ability to expand their understanding of concepts in a number of ways, including their ability to connect with other ideas (e.g., see the relationship between concepts or through their understanding of the perspectives of others); (2) the willingness to be persistent and goal directed and show the capacity to learn at a much earlier age; (3) the ability to learn to speak early

and talk in sentences using complex sentence structure before the age of 2 or 3; and (4) the ability to have highly receptive vocabularies and come to school reading significantly beyond their chronological age-mates. Many of these children possess advanced verbal abilities, extensive vocabularies, and vivid imaginations. Due to their unique abilities, they often require accelerated instruction and sophisticated instructional materials. They are highly motivated and extremely independent. Many are recognized by their teachers for their superior achievement in one or more academic domains (e.g., reading, math, and science). These children like to take risks, and have a high tolerance for ambiguity. They have a strong dislike for skill and drill activities; and find conventional instructional practices less motivating, less challenging, and less rewarding. Like students with disabilities, there exist both inter- and intra-individual differences in their performance. In addition to their high levels of intelligence, some possess high levels of creativity particularly in the visual and performing arts, while others demonstrate outstanding intuitive, leadership, and psychomotor abilities. No doubt, these children display unique social and emotional characteristics (Clark, 2008). They hold high expectations of themselves, possess a keen (sometimes subtle) sense of humor, and have a strong sense of self-awareness. They strive for perfection, are highly focused, and experience enormous stress from failure. They possess high morals, have a strong sense of justice, and are sensitive to the feelings and needs of others. Many of these children display high energy levels in their work and play which can sometimes be misinterpreted and misdiagnosed by psychologists and other health professionals as a form of hyperactivity or an emotional or behavioral disorder. The most common misconception about these children is that they lack social and emotional competence (Clark, 2008). While there are some of them who feel different, have low self-esteem, are misunderstood or socially isolated, most are well adjusted, sensitive, healthy, emotionally independent, and stable individuals who are well liked by their peers (Clark, 2008) and who possess emotional traits that are well balanced. Clearly, the aforementioned characteristics are only examples of attributes of children with gifts and talents and should not be considered as exhaustive. Not every child identified as gifted or talented will display all of these traits.

Beliefs About Intelligence and Gifted Education

Given what we know about the nature versus nurture argument as related to human development and intelligence, it is not surprising that giftedness is considered to result from both genetic and environmental influences (Friend, 2005; Daniels, 2003). The correlation between

intelligence and gifted education continues to impact our decision making regarding curriculum models implemented in the school district. Traditionally, intelligence theories have affected the way in which we identify and assess students, our attitudes toward giftedness and gifted students, the curriculum models upon which we base our programs and interventions, and many other aspects of gifted education.

Research in the areas of gifted and talented has been closely linked with intelligence studies (Plucker, 2001). Many scholars concerned with matters of intelligence also focused on manifestations of talent and genius: Terman (1925) initiated long-term studies in giftedness; and Hollingworth (1942) pioneered work with exceptional children and women; Renzulli (1979) concentrated on distinguishing between real world and academic giftedness; Feldman (1980) looked at developmental trajectories unique to each intellectual domain; Gardner (1983) implemented multiple intelligences as a model; Gagne (1985) focused on the differences between giftedness as potential and talent as fulfillment of potential; and Sternberg (1985) examined the types of gifted abilities, as well as the strengths and weaknesses of intelligence, to name just a few. Many of these theories dominate and bridge the chasm between intelligence and gifted education.

Curricular and Programmatic Models

Connecting intelligence theories to curricular models requires a juxtapositioning of the theories with specific curricular models. VanTassel-Baska and Brown (2009) defined curricular models as a curriculum design and development framework that provides a systematic approach to developing an appropriate curriculum for the target population. These curricular models must include in their frameworks the ability to identify curriculum-product elements. In order for a model to be effective, it must possess differentiation, flexibility, transferability, and usability across content areas, age/grade span, and flexible grouping models. Additionally, it must delineate ways in which it responds to the particular needs of the individual with gifts. The models are descriptive and do not validate the most effective method for maximizing student development in a specific time or space. Each of the curricular models has more than 10 years of supportive research, development, and implementation undergirding its use, endurance, and attention (Coleman and Cross, 2005; VanTassel-Baska and Brown, 2009). The following are subsections that discuss these models.

The Stanley Model of Talent Identification

The Stanley model focuses on lifelong education for the individual. Model principles include: (1) utilization of assessment instruments that encourage high-level verbal and mathematical reasoning; (2) diagnostic instructional

methodologies; (3) accelerated and fast-paced core content; and (4) flexible curriculum across the school environment. This content-based model is aligned with national standards, is highly sustainable, and demonstrates the benefits of acceleration for advancement.

The Renzulli Schoolwide Enrichment Triad Model

The major principles of the Schoolwide Enrichment Triad Model (SEM) model are supported by the (1) use of interest and learning style inventories to assess inter and extracurricular abilities; (2) provision of compacting curriculum (i.e., the regular curriculum is modified by eliminating mastered content, and substituting alternative work); and (3) access to the appropriate triad level based on students' abilities, interest, and task commitment. Enrichment level 1 consists of general exploratory experiences (e.g., guest speakers, field trips, and interest centers); enrichment level 2 includes instructional strategies designed to promote thinking; and enrichment level 3 utilizes analytical activities and creative productions that support primary inquiry and thinking.

Gardner's Multiple Intelligences (MI)

The Gardner model is a core-curricula approach that employs the multidimensional concept of intelligence (Gardner, 1983). The eight types of intelligence are defined as (1) verbal/linguistic, (2) logical/mathematical, (3) visual/spatial, (4) musical/rhythmic, (5) bodily/kinesthetic, (6) interpersonal, (7) intrapersonal, and (8) naturalistic. The multiple-intelligences curricula model has been used as the base curriculum for the identification of individual differences and multidimensional evaluation.

The Schlichter Model for Talents Unlimited Inc. (TU)

Talents Unlimited was based upon Guilford's (1967) inquiry into the nature of intelligence and it features four major components which include (1) a demonstration of talents, productive thinking, decision making, and academic abilities; (2) instructional materials; (3) in-service professional development for teachers; and (4) assessment of students' thinking-skills development (Schlichter, 1981). The model has been noted for developing student's creative and critical thinking.

Sternberg's Triarchic Componential Model

The Sternberg's Triarchic Componential Model (STTCM) is based upon information-processing intelligence theory

(Sternberg, 1985). The model incorporates components of the mental processes used in thinking. These components include (1) the executive process component is involved in planning, decision making, and monitoring performance; (2) the performance component processes facilitate executive problem-solving strategies; and (3) the knowledge-acquisition component facilitates the acquisition, retention, and transference of new information. This model supports the concept that interaction and feedback between the individual and his/her environment within any given context allows cognitive development to occur.

VanTassel-Baska's Integrated Curriculum Model

The Van-Tassel-Baska Integrated Curriculum model (ICM) was developed for high-ability learners. It is based upon three dimensions: (1) advanced content; (2) high-level process and product work; and (3) divergent and convergent thought development and understanding. The use of advanced curricula and accelerated rates of instruction in core subject areas is supported by an extensive body of research. Over 30 years of research supports the implementation of more enrichment-oriented models for students with gifts and talents (VanTassel-Baska and Brown, 2009).

On the whole, decisions as to which curricular model to implement should be consistent with general premises of general school policy. It is important that the model supports students' understanding of reality; recognize curriculum alternatives based on various gifts and talents; and be responsive to the effect of student choices (Coleman and Cross, 2005).

Differentiated Instruction for Students with Gifts and Talents

The most powerful models of instruction are accelerated, interactive, and differentiated. Differentiation is defined by the (1) individualized learning needs of children with gifts and talents; (2) specific content and skill sets of the curriculum; (3) pedagogy used to convey the content; and (4) flexible settings needed to effectively implement the curriculum (Kaplan, 2005; VanTassel-Baska and Brown, 2009). Differentiated instruction encourages the learner to construct and produce knowledge in meaningful ways. Within the construct of differentiated instruction, students engage in generating original products, collaborative construction of knowledge, and problem-based learning. Differentiated strategies include (1) individual and group summarizing; (2) investigation of divergent perspectives, higher-order critical thinking; (3) brainstorming; (4) Socratic dialog; (5) problem-solving processes; and (6) team teaching.

Differentiated curriculum for students with gifts is based on the strengths of the students and the inability of the core curriculum to meet the needs of this population. The analysis of the skill sets of these students with respect to the traditional core-curriculum content, processes, and product components justifies the rationale for a differentiated curriculum.

Differentiated core-curriculum programs should stimulate the development of students' actual and potential abilities as well as their skills, creativity, and talents. Instructional strategies for students with gifts and talents utilize several curriculum modifications to meet the challenges faced by inadequacy of the core curriculum. These modifications include (1) higher levels of thinking, (2) open-discovery learning, (3) evidence of reasoning, (4) freedom of choice, (5) group interaction, (6) pacing/acceleration, and (7) variety of processes. Implementing these elements to modify the core curriculum helps teachers to (1) present content related to comprehensive; (2) broad-based issues; (3) integrate multiple disciplines; (4) allow for in-depth self-selected learning; (5) develop independent study skills; (6) focus on open-ended tasks; (7) develop research methods; (8) encourage the development of new products and techniques; (9) self-understanding; and (10) self-directedness.

Conclusion

While there are a number of strategies and approaches that can be used to deliver instructional services to students with gifts and talents, this article shares information on some of the strategies and approaches commonly used by school districts. It is important to keep in mind that all programs, regardless of grade level or curriculum model used, must enhance (1) flexible teaching strategies, (2) multiple venues for students to reach their learning and creative potentials, (3) continuous measures to assess progress, (4) collaborative construction of knowledge through peer and teacher interactions, and (5) continuity. From our perspective, our goal is to have highly qualified teachers with specialized training to teach students with gifts and talents using creative multidimensional techniques.

See also: Educating Students with Special Needs: An Overview; The Identification of Students with Gifts and Talents.

Bibliography

- Clark, B. (2008). *Growing up Gifted: Developing the Potential of Children at Home and at School*, 7th edn. Upper Saddle River, NJ: Prentice-Hall.
- Coleman, L. and Cross, T. (2005). *Being Gifted*, 2nd edn. Waco, TX: Prufrock Press.

- Daniels, V. I. (2003). Students with gifts and talents. In Obiakor, F. E., Utley, C. A., and Rotatori, A. F. (eds.) *Effective Education for Learners with Exceptionalities*, vol. 15, pp 325–348. Kidlington: Elsevier Science.
- Feldman, D. (1980). *Beyond Universals in Cognitive Development*. Norwood, NJ: Ablex.
- Friend, M. (2005). Special education: Contemporary perspectives for school professionals. Boston, MA: Pearson.
- Gagne, F. (1985). Giftedness and talent: Reexamining a reexamination of the definition. *Gifted Child Quarterly* **29**, 103–112.
- Gardner, H. (1983). *Multiple Intelligences: The Theory in Practice*. New York: Basic Books.
- Guilford, J. (1967). *The Nature of Human Intelligence*. New York: McGraw-Hill.
- Hallahan, D. and Kauffman, J. (2003). *Exceptional Learners: Introduction to Special Education*, 9th edn. Boston, MA: Allyn and Bacon.
- Hollingsworth, L. (1942). *Children above 180 IQ (Stanford Binet): Origin and Development*. Yonkers-on-Hudson, NY: World Book.
- Kaplan, S. (2005). Layering differentiated curricula for the gifted and talented. In Karnes, F. A. and Bean, S. M. (eds.) *Methods and Materials for Teaching the Gifted*, pp 107–135. Waco, TX: Prufrock Press.
- Karnes, F. A. and Beane, S. (eds.) (2009). *Methods and Materials for Teaching the Gifted*. Waco, TX: Prufrock Press.
- Marland, S. P. (1972). Education of the gifted and talented. *Report to the Congress of the United States by the US Commissioner of Education*. Washington, DC: US Government Printing Office.
- Plucker, J. (2001). Looking back, looking around, looking forward: The impact of intelligence theories on gifted education. *Roeper Review* **23**, 124–125.
- Renzulli, J. S. (1979). Some concerns about educational acceleration for intellectually talented youth, or are treadmills really different if we run them at a faster rate? In George, W. C., Cohn, S. J., and Stanley, J. (eds.) *Educating the Gifted: Acceleration and Enrichment*, pp 190–191. Baltimore, MD: John Hopkins University Press.
- Schlichter, C. (1981). The multiple talent approach in mainstream and gifted programs. *Exceptional Children* **48**, 144–150.
- Smith, T., Polloway, E., Patton, J., and Dowdy, C. (2004). *Teaching Students with Special Needs in Inclusive Setting*, 4th edn. Boston, MA: Pearson.
- Sternberg, J. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence*. Cambridge, UK: Cambridge University Press.
- Terman, L. (1925). *Mental and Physical Traits of a Thousand Gifted Children: Genetic Studies of Genius*, vol. 1. Stanford, CA: Stanford University Press.
- Turnbull, R., Turnbull, A., Shank, M., and Smith, S. (2004). *Exceptional Lives: Special Education in Today's Schools*, 4th edn. Boston, MA: Pearson.
- VanTassel-Baska, J. (1998). Appropriate curriculum for the talented learner. In Van-Tassel-Baska, J. (ed.) *Excellence in Educating Gifted and Talented Learners*, pp 339–361. Denver, CO: Love.
- VanTassel-Baska, J. and Brown, E. (2009). An analysis of gifted education curriculum models. In Karnes, F. A. and Beane, S. M. (eds.) *Methods and Materials for Teaching the Gifted*, pp 75–106. Waco, TX: Prufrock Press.

Further Reading

- Colangelo, N., Assouline, S. G., and Gross, M. U. M. (2004). *A Nation Deceived: How America's Schools Hold back America's Brightest Students*. Iowa City, IA: Belin Blank Center.
- Gagné, F. (1991). Toward a differentiated model of giftedness and talent. In Colangelo, N. and Davis, G. A. (eds.) *Handbook of Gifted Education*, pp 65–80. Boston, MA: Allyn and Bacon.
- National Association for Gifted Children and Council of State Directors of Programs for the Gifted (2007). *State of the States in Gifted Education 2006–2007*. Washington, DC: National Association for Gifted Children and Council of State Directors of Programs for the Gifted.
- National Commission on Excellence in Education (NCEE) (1983). *A nation at risk: The imperative educational reform. Report No. 065-000-00177-2*. Washington, DC: Government Printing.
- Obiakor, F. E. (ed.) (2007). *Multicultural Special Education: Culturally Responsive Teaching*. Upper Saddle River, NJ: Pearson.
- US Department of Education (2002). No Child Left Behind Legislation: Public Law 107 – 110th Congress. <http://www.ed.gov/legislation/ESEA02> (accessed August 2009).
- US Department of Education Office of Research and Evaluation (1993). *National Excellence: A Case for Developing the Nation's Talent*. Washington, DC: Government Printing.

Relevant Websites

- <http://www.apa.org> – American Psychological Association, Center for Gifted Education Policy.
- <http://www.cfge.wm.edu> – Center for Gifted Education.
- <http://www.ericdigests.org> – Eric Digests.
- <http://www.pz.harvard.edu> – Gardner's Multiple Intelligence Teacher Training Project Zero.
- <http://www.gt-cybersource.org> – GT-Cybersource.
- <http://www.uniquelygifted.org> – Internet Resources for Gifted/Special Needs Children.
- <http://cty.jhu.edu> – Johns Hopkins University Center for Talented Youth (CTY).
- <http://www.nagc.org> – National Association for Gifted Children.
- <http://www.nfgcc.org> – National Foundation for Gifted and Creative Children.
- <http://www.nsrgt.org> – National Society for the Gifted and Talented.
- <http://www.gifted.uconn.edu> – Neag Center for Gifted and Talented Development.
- <http://www.sedoparking.com> – Sedo Domain Parking, National Talent Network (NTN) at the Educational Information and Resource Center's (EIRC).
- <http://www.sengifted.org> – Supporting the Emotional Needs of the Gifted.
- <http://www.cec.sped.org> – The Council for Exceptional Children, The Association for the Gifted and Talented (CEC-TAG).
- <http://www.gifted.uconn.edu> – University of Connecticut, National Research Center on the Gifted and Talented (NRC/GT).
- <http://world-gifted.org> – World Council for Gifted and Talented Children.

Related Services for Children with Special Needs

S O Obi, Kentucky State University, Frankfort, KY, USA

© 2010 Elsevier Ltd. All rights reserved.

Special education means specially designed instruction, provided at no cost to parents, to meet the unusual needs of a student with exceptionalities. These include special materials, teaching techniques, equipment, and/or facilities. The regular classroom or special education teacher who is aware of the individual needs of students and skilled at meeting them may need consultation with special education, transportation, audiology, psychological services, physical therapy, occupational therapy, recreation, early identification, medical services, school health services, counseling, social work services, and parent counseling and training. This article focuses on how students' needs can be met with these related services.

Children with disabilities have sometimes been prevented from attending their neighborhood schools or benefiting from educational activities by circumstances that impeded their access or participation: a child who uses a wheelchair, for example, may require a specially equipped school bus; one with special health needs may require medication several times a day; and a child with an orthopedic impairment may need physical therapy to maintain sufficient strength and flexibility in his/her arms and legs. Individuals with Disabilities Education Improvement Act (IDEIA) requires that schools provide any related services that a child with a disability may need to access and benefit from special education (Heward, 2009). More specifically, if special education is to be effective, related services will be necessary. A child who is between the ages of 3 and 21 years with a verified disability and also demonstrates the need of a related service, may be eligible to receive them. The receipt of a related service is based on the documented and assessed needs of the child. In determining whether or not assessed educational needs should be met by a specific related service, the following should be apparent. In other words, related services are required for a student with disabilities when:

- the services are necessary for a student to gain access to a special education program or the services outlined in an individual education plan (IEP);
- the services are necessary for a student to remain physically in an educational program;
- a student cannot make meaningful progress toward the goals and objectives of an IEP without the services;
- a student's various needs are so intertwined that an integrated program of a special education and related services is needed; and

- progress toward the goals and objectives of the IEP is dependent on the resolution of other needs.

The IEP team is necessary to accomplish the need for related services. However, it should consider terminating a specific service if the IEP goals have been met and no additional services are necessary to meet the child's educational needs. Termination should also be considered when the potential for further change appears unlikely based on previous documented intervention attempts (Schafer, 1998).

Service delivery, the intensity of the service, and the frequency of service are determined by the IEP team based on the child's educationally related needs. The critical determinants are the individual, unique educational needs of the child, and the most appropriate manner in which the special education needs can be supported by the related services in question. In developing goals and objectives for related services, IEP team members should consider the following factors:

- the least-restrictive environment needed to accomplish the goals and objectives for the related service;
- the type of skills to be learned and the methods and strategies of intervention anticipated;
- the level of expertise required to provide the service; and
- the need for and availability of others to carry out the child's IEP.

In general, service delivery will vary in frequency depending on the needs of the child. The need for sharing of information among IEP team members and joint program planning and evaluation must be done throughout the program plan. Monitoring of student performance is necessary to determine if the amount of service is appropriate to promote progress toward attainment of the child's special education goals and objectives (Nebraska Department of Education, 2000).

Related Service Professionals

Many types of professionals are needed to provide multidisciplinary services required by individuals with disabilities. A special educator might be a paraprofessional (teacher's aide), a resource teacher, a consultant, an itinerant teacher, a special education classroom teacher, a job coach, an assistive technology specialist, a home or

hospital teacher, a diagnostician, or an administrator. Furthermore, a teacher may assume several of these professional roles during his/her career. In addition, a special education professional might work in a related service as a school psychologist, a speech/language pathologist, an audiologist, an occupational therapist, a physical therapist, a counselor, a nurse or physician, a transportation specialist, a recreational therapist, a supported living worker, a personal-care attendant, a vocational rehabilitation worker, or a lawyer (Smith, 1998).

Special education teachers and others who work with individuals with disabilities must be able to collaborate and cooperate with professionals from a variety of disciplines. They work together in multidisciplinary teams (MDTs) comprised of those professionals each individual student needs. It is through collaboration and working together as a team that is ultimately the key to successful integration for individual students, given that they require various combinations of services over their school careers (Smith, 1998). Special education teachers collaborate in many areas that touch a child's life. They use collaborative skills when working with parents and families, performing multidisciplinary assessments, working with a team to develop individualized program plans, coordinating the components of students' individualized plans, and helping them make transitions through the school years and from school to work. For instance, the Nebraska Department of Education (2000) categorized services into three types, which are detailed below.

Direct Service

In direct service, the related service provider works with a child individually or in a small group on a regularly scheduled basis to assist the child to develop skills relevant to the child's educational performance. The emphasis of direct service is performance and skill acquisition to benefit from instruction. Direct service may also be indicated to maintain newly acquired skills or to slow down the rate of regression. Collaboration with parents, teachers, and paraeducators is necessary to achieve goals and carry-over activities into regular routine. Consultation is needed to maximize the educational benefit of the intervention.

Integrated Service

Integrated service combines direct, hands-on child contact and consultation with others directly involved with the child. There is an emphasis placed on the need for practice of skills in the child's daily routine. The process of goal achievement is shared among those involved with the education of the child. Those involved may include the related service provider, teachers, parents, and classroom paraeducators. Intervention typically includes adapting activities occurring in the child's routine and creating opportunities

for him/her to practice new skills with others. The related service is provided within his/her daily environment, and should always include others involved with him/her to carry out the delegated activities or services.

Consultative Service

The related service provider in consultative services focuses on providing input to teacher, staff, and parents regarding the child's specific needs rather than providing direct services. This does not preclude the need for the related service provider to work directly with the child to identify needs or solve problems. The related service provider supports the educational program, but is not the primary provider of services. The child's needs may not be rapidly changing or only minimal input is needed from the related service provider. Consequently, only periodic direct contact is needed. The related service provider's involvement includes providing suggestion to the teacher, staff, and parents for modifications of educational materials and the environment as well as infrequent monitoring of the child's progress and any adaptive devices that may be used.

Types and Definitions of Related Services

The IDEIA requires school boards to provide related, or supportive, services to students with disabilities, when children need such services to assist them in benefiting from their special education programs. Other services may be deemed related services if they assist students with disabilities in benefiting from special education. For example, services such as artistic and cultural programs or art, music, and dance therapy could be related services under the appropriate circumstances. Persons of varying professional backgrounds who have a variety of occupational titles may provide related services. The detailed types, definitions, and professionals of related services as indicated in IDEIA (2004) are discussed below.

Audiology

Audiology is provided by audiologists and speech/language pathologists to children with known or suspected hearing or listening disorders, and/or those with speech or language delays. It is the scope of practice for speech/language pathology (SLP) to provide the communication assessment and rehabilitation services to children who are deaf or hearing impaired, as described in the American Speech-Language-Hearing Association (ASHA, 1996) Scope of Practice in Speech-Language Pathology. Deconde *et al.* (1997) suggested that children with central auditory processing disorders (CAPDs) may be eligible for special education or 504 services, depending on the severity

and implications of the problem – usually students with significant CAPD qualify for eligibility consideration with either a hearing, speech/language, or learning disability. The audiologist is the only professional who is qualified to perform an audiological evaluation for verification purposes. For purposes of the MDT evaluation, the audiologist provides a written report that documents the hearing loss.

Counseling

Counseling means assisting parents in understanding the special needs of their children and providing them with information about child development. Services include social skills, peer relationships, crisis intervention, and classroom guidance, including topics such as self-concept, decision making, problem solving, and system support. System support includes referring children and families to community agencies or providers, support groups, and law enforcement. These services are provided by qualified social workers, psychologists, guidance counselors, rehabilitation counselors, or other qualified personnel defined by the professional scope of practices.

Early Identification and Assessment Services

These services mean the implementation of a formal plan for identifying a disability as early as possible in a child's life. As stated by the American Academy of Pediatrics Committee on Children with Disabilities (1994), early identification of children leads to effective therapy of conditions for which a definitive treatment is available. In conditions that cannot be fully reversed, early intervention improves the child's outcomes and enables families to develop strategies and obtain resources for successful family function. In order to engage in the practice of medicine, a physician, surgeon, or obstetrician, must be duly licensed by the state in which he/she practices. Specifically, the varieties of providers that are involved in screening young children include pediatricians, family practitioners, physician assistants and nurse practitioners in pediatrics and family practice, and public health nurses. Any medical provider may make a referral to the local school district.

Medical Service

These services are provided by a licensed physician to determine a child's medically related disability that results in his/her need for special education and related services. In order to receive services, the school-age child must have a verified disability that has the potential to interfere with his/her educational process and school performance. An accurate medical diagnosis provides the family and school personnel information about the child's condition as well as particular precautions that

may pertain to him/her and impact the educational program. The attending physician can also be of assistance in the development of expected goals and provide opinion about the provision of related services for the child's special education program. It is also appropriate for the treating physician to assist the evaluation team, including the parent, in determining whether the child's disability would be eligible for services under Section 504 of the Rehabilitation Act of 1973. It is the responsibility of the school district's MDT to evaluate his/her eligibility for special education and related services.

Occupational Therapy

This kind of therapy is provided by qualified occupational therapists. It includes:

- improving, developing, or restoring functions impaired or lost through illness, injury, or deprivation;
- improving ability to perform tasks for independent functioning if functions are impaired or lost; and
- preventing, through early intervention, initial or further impairment or loss of function.

Occupational therapy is intended to serve children from 3 to 21 years of age, who have a verified disability. This therapy may include feeding and self-help skills, fine and visual motor skills, visual processing skills, sensory processing, and positioning and adaptive services or equipment. Those who are qualified to provide this therapy include registered and licensed occupational therapists, certified occupational therapy assistants, and occupational therapy aides.

Orientation and Mobility Service

These services are provided to students with visual impairments and blindness by qualified personnel to enable them to attain systematic orientation and safe movement within their environments in school, home, and community. They include teaching students the following:

- spatial and environmental concepts and use of information received by the senses (such as sound, temperature, and vibrations) to establish, maintain, or regain orientation and line of travel (e.g., using sound at a traffic light to cross the street);
- the use of long cane to supplement visual travel skills or as a tool for safely negotiating the environment for students with no available travel vision;
- understand and use remaining vision and distance low vision aids, as appropriate; and
- other concepts, techniques, and tools as determined appropriate.

Hazekamp and Huebner (1989) noted that the ability to understand, interact with, and move within one's physical

and spatial environment is a fundamental developmental skill. This ability is one of the milestones indicative of maturation for students with sights or visually impairments. The development of orientation and mobility skills is essential for students to travel independently in various community settings. Services in this area are provided by Certified Orientation and Mobility Specialists (COMS) and Certified Orientation and Mobility Assistants (COMA).

Parent Counseling Services

These services assist parents in understanding the special needs of their child and in providing them with information about child development. They also help parents to acquire the necessary skills that will allow them to support the implementation of their child's IEP (Wright and Wright, 2007). Parent counseling as a related service may be provided by school guidance counselors, social workers, psychologists, or other qualified personnel.

Physical Therapy

This kind of therapy is provided by a qualified physical therapist. Physical therapy serves children aged 3–21 with a verified disability; and it involves gross motor skills, orthopedic concerns, impaired mobility, and adaptive equipment/positioning needs that interfere with the student's educational performance. The focus of educational services is to allow the student to benefit from his/her special education program by promoting functional independence or participation within the educational environment (Wright and Wright, 2007).

Psychological Services

These services include administering and interpreting assessments, consulting with other staff members, planning psychological services and counseling, assessing behaviors, and developing positive behavioral intervention strategies. School psychologists are required to have valid certification in education and psychology to serve children from birth to age 21 and families within school districts. As a related service, school psychologists work closely with classroom teachers and paraprofessional providers on such issues as behavior management and instructional practices. In providing the full range of psychological services, they also work closely with school counselors and, when available, with school social workers. The decision as to who specifically provides a particular service is usually made by the MDT or during the IEP meeting. A variety of factors can be considered when making this decision. These include availability, expertise in the specific need area, and previous experience with the individual student and/or the family. (Stuberg and Schafer, 2000).

Recreational Therapy

This kind of therapy provides assistance, instruction, and strategies to facilitate the short-term objectives of special education. This therapy is also important to the long-range life goals of the student with disabilities. It facilitates social involvement and friendships and encourages and provides skills associated with opportunities for community integration. As an educational tool, recreational therapy can be used to achieve cognitive, social, emotional, and physical objectives identified as learning needs. Recreation includes:

- assessment of leisure function;
- therapeutic recreation services;
- recreation programs in schools and community agencies; and
- leisure education.

The recreational therapist is not a teacher, but rather is a supportive trained certified therapeutic recreational specialist (CTRS), who works and consults with other education personnel to extend and enhance the students' total education.

Rehabilitation Counseling

This means services provided by qualified personnel in individual or group session that focus specifically on:

- career development;
- employment preparation;
- achieving independence; and
- integration in the workplace and community of a student with a disability.

The term also includes vocational rehabilitation services provided to a student with disabilities by vocational rehabilitation programs funded under the Rehabilitation Act of 1973, as amended.

Rehabilitation counseling is intended to serve students aged 14–21 with a verified disability and a current IEP with transition plan in place as mandated by Individuals with Disabilities Education Act (IDEA). Rehabilitation counseling can be provided by any special education endorsement, guidance and counseling, vocational special needs, or diversified occupations endorsement, or special services certificate for school rehabilitation counselor.

School Health and School Nurse Services

These health services are designed to enable a child with a disability to receive free appropriate public education as described in the child's IEP. School nurse services are provided by a qualified school nurse. School health services may be provided by either a qualified school nurse or other qualified person. Health services provided by a

registered nurse include services designed to help a student protect, improve, and maintain physical, emotional, and social well-being. Students with special health conditions may require the provision of a safe and healthy environment, emergency health care procedures, communicable disease prevention and control, administration of medication, administration of invasive nursing procedures, individualized health assessment and care planning, and health care record keeping.

Social Work

Social work services in schools include:

- preparing a social developmental history on a child with a disability;
- group and individual counseling with the child and family;
- working with problems in a child's living situation (e.g., home, school, and community) that affect the child's adjustment in school;
- mobilizing school and community resources to enable the child to learn as effectively as possible in his/her educational program; and
- assisting in developing positive behavior intervention strategies.

Social work services in the school are provided by an individual with a baccalaureate or master's degree in social work from an approved educational program. School social workers develop skills for effective service to children, families, personnel of the local education agency, and the community (NASW, 1992).

Speech/Language Pathology

This means providing services that include:

- identification of children with speech or language impairments;
- diagnosis and appraisal of specific speech or language impairments as well as referral for medication or other professional attention necessary for the habilitation of speech or language impairments;
- provision of speech and language services for habilitation or prevention of communicative impairments;
- counseling and guidance of parents, children, and teacher regarding speech and language impairments; and
- referral for medical or other professional attention necessary for the habilitation of speech or language impairments.

Speech/language pathologists serve children with speech, voice, language, communication, swallowing, and related disabilities. It is provided by a person who holds a valid and current credential as a speech or hearing specialist from an academic program approved by the board. In

order for a student to receive SLP as a related service from his/her resident public school district, he/she must have a verified disability as determined by a multidisciplinary evaluation team.

Transportation

This includes:

- travel to and from school and between schools;
- travel in and around school buildings;
- specialized equipment (such as special or adapted buses, lifts, and ramps), if required to provide special transportation for child with a disability.

A student cannot benefit from an educational program if he/she cannot get to the program. For this reason, transportation is probably the most common related service that school boards provide. The transportation personnel will, with training provided by the district, be responsible for taking appropriate safety precautions and care of children with a wide range of ages and disabilities.

Recommendations for Practice

Clearly, school boards must provide students with disabilities with related services when such services are necessary for them to benefit from their special education programs. The only limitation that has been placed on what may be considered a related service is that medical services are exempted unless they are for diagnostic or evaluative purposes. Much litigation has ensued from the IDEA's related services provision. This litigation has given school officials considerable guidance. The following recommendations have been developed on that litigation. Schools officials should:

- Provide related services to students who are receiving special education if they are necessary for the student to receive educational benefit.
- Consider whether a student with a disability who is not receiving special education may be entitled to related services under Section 504.
- Provide services of a life-support nature, if necessary, during the school day.
- Provide medical services, if needed, for diagnostic or evaluative purposes.
- Provide all necessary school health services that can be performed by a school nurse, health aide, or other trained layperson.
- Provide a full-time nurse to help guarantee that students with significant medical needs are integrated into public schools.
- Check to see if students who require extensive health-related services are eligible for Medicaid benefits that

could offset the costs of many expensive health-related services.

- Investigate the option of using parents' health insurance, particularly for a diagnostic and evaluative services (however, parents may not be required to use their health insurance if it will incur a cost to them in the form of a reduction in benefits, a cap on benefits, or an increased premium).
- Provide psychotherapy, social work services, or counseling when the resolution of emotional concern is a prerequisite to a student's successful progress toward the goals of the IEP.
- Be diligent in ascertaining exactly what is the primary reason for a residential placement, as placements that are required for reason that are not education may not be the responsibility of the school board.
- Provide special transportation when a student is unable to access a standard mode of transportation.
- Avoid excessively long bus or van rides to school. All transportation arrangements must be reasonable.
- Make provisions for any child whose disability prevents him/her from getting to and from the transportation vehicle.
- Provide transportation to and from a residential facility.
- Establish natural policies regarding transportation of all students to day care centers or after-school caretakers.
- Ensure that students with disabilities are given the same considerations as students who are disabled regarding transportation outside of the school's attendance boundaries.
- Provide an aide on the transportation vehicle, if necessary, to ensure safe passage for medically fragile students.
- Make alteration to the physical plant to allow a student with a disability to participate fully in and benefit from an education program (however, another option may be to bus the student to a nearby school that is accessible if the required alterations would be excessively costly).
- Provide parents with training so that there is consistency between the techniques used in school and at home (this option may be a viable alternative to residential placement).
- Pay a student's room and board expenses if the only facility that can provide an appropriate education is not within commuting distance from the child's home.
- Provide full access to sports or other extracurricular activities whenever a student qualifies for participation.
- Include participation in sports or other extracurricular activities in a student's IEP if these activities may assist

him/her in benefiting from his or her educational programming.

- Provide assistive technology to students with disabilities when needed (however, school boards are not required to provide personal devices that a student would require regardless of whether the child attended school).
- Consider instruction in daily living skills when developing transition services for students who are close to exiting the educational system.

Conclusion

Students with disabilities need related services to maximize their fullest potential. However, these services must be provided by qualified personnel to effectively help students to benefit from special education programs. This article has addressed related services, related professionals, and basic issues surrounding related services. Basic qualifications of the related professionals are necessary to prevent litigations on schools. The overall goal is for professionals and related professionals to collaborate, consult, and cooperate to facilitate services for students who need them.

Bibliography

- American Academy of Pediatrics Committee on Children with Disabilities Screening (1994). Infants and young children with developmental disabilities. *Pediatrics* **93**, 863–865.
- ASHA (American Speech–Language–Hearing Association) (1996). *Scope of practice in audiology*. ASHA **38**, 16–20.
- DeConde, J. C., Benson, P. V., and Seaton, J. B. (1997). *Scope of Practice in Speech-Language Pathology*. San Diego, CA: Singular.
- Hazekamp, J. and Huebner, K. M. (eds.) (1989). *Program Planning and Evaluation for Blind and Visually Impaired Students: National Guidelines for Educational Excellence*. New York: American Foundation for the Blind.
- Heward, W. L. (2009). *Exceptional Children: Introduction to Special Education*, 9th edn. Upper Saddle River, NJ: Merrill/Prentice-Hall.
- Schafer, L. (1998). *Special Education Related Services: Volume One. A Guide for Educators and Parents in Nebraska Public Schools*. Lincoln, NE: Nebraska Department of Education.
- Smith, D. D. (1998). *Introduction to Special Education: Teaching on an Age of Challenge*, 2nd edn. Boston, MA: Allyn and Bacon.
- Stuberg, W. and Schafer, L. (2000). *Special Education Services: A Guide for Educators and Parents in Nebraska Schools*. Lincoln, NE: Nebraska Department of Education.
- Wright, P. W. and Wright, P. D. (2007). *Wrightslaw: Special Education Law*, 2nd edn. Hartfield, VA: Harbor House Law Press.

Risk and Resilience Frameworks in Understanding Special Education

W-Y Hsieh and M L Donahue, University of Illinois at Chicago, Chicago, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Culturally responsive pedagogy – Culturally responsive pedagogy comprises three dimensions: (1) institutional (the administration and its policies and values), (2) personal (the cognitive and emotional processes teachers must engage in to become culturally responsive), and (3) instructional (materials, strategies, and activities that form the basis of instruction). Teaching and learning occur in a culturally supported, learner-centered context, whereby the strengths students bring to school are identified, nurtured, and utilized to promote student achievement in a culturally responsive classroom (The National Center for Culturally Responsive Educational Systems).

High-incidence disabilities – Learning disabilities, emotional/behavioral disorders, mild intellectual disabilities or speech/language disabilities.

Inclusive settings – Inclusion as a value supports the right of all children regardless of their diverse abilities to participate actively in natural settings within their communities. A natural setting is one in which the child would spend time if he or she had not had a disability (The Division for Early Childhood of the Council for Exceptional Children).

Protective factors – Positive factors that help to mitigate adversity in development.

Risk factors – Adverse factors and stressors that tend to result in negative outcomes in development.

Special education – Individualized education programs for students with special needs that address the students' individual differences and respond to their needs.

The ecological model – A model proposed by Urie Bronfenbrenner, proposing that a child's development must be understood at the center of a nested system of interactive influences. These consist of a microsystem (a child's immediate surroundings, e.g., family), a mesosystem (the relationship between the structures of the child's microsystem, e.g., interaction between the child's teacher and his parents), an exosystem (the larger social system that does not impact the child directly, e.g., parent workplace schedule), and a macrosystem (the outermost layer in the child's

environment, e.g., cultural values). This model illustrates how a child's development is affected by the relationships and interactions of the four systems.

The risk-resilience model – Children raised with multiple adverse influences, including disabilities, may still achieve positive outcomes, given the right combination and timing of protective factors.

Historically, the field of special education has focused on identifying hallmarks or common characteristics of particular disabilities, in a search for the definition and refinement of various disability categories (e.g., learning disabilities, autism, and hearing impairment). However, the foundation of special education lies in the recognition of individual differences in human development (Osgood, 2008). Researchers and practitioners in special education have long been aware that assessment and intervention policies for students with disabilities depend on the understanding of the developmental trajectory of a variety of individual, family, and community factors and their interactions. These professionals affirm that individual differences in physical, cognitive, and emotional outcomes are not simply noise or measurement error in the data, but must be taken into account by some viable theory and intervention model.

It is no surprise then that the first conceptualization of the risk and resilience framework was derived from research by Werner and Smith (1982) on students with developmental disabilities. As one of the first comprehensive and longitudinal developmental studies with a large sample, Werner *et al.* (1971) studied the entire cohort of infants born on the island of Kauai in 1955. The influence of a wide range of biological and environmental factors on the cognitive and emotional outcomes of these individuals was then investigated periodically for the next 40 years (Werner and Smith, 1982, 1992, 2001). This rich source of data helped identify the power of adverse factors and stressors in explaining negative outcomes (risk factors), as well as the positive factors that helped to mitigate this adversity (protective factors). One compelling finding was a subset of children who beat the odds (i.e., children exposed to significant biological risk, poverty, and poor parenting who still managed to achieve positive life

experiences and achievements). This group of invincible or resilient children triggered a dramatic search for risk and protective factors and their transactions that could explain positive developmental outcomes in the face of adversity (e.g., Reiff *et al.*, 1997). Since then researchers who study disability have been eager to invoke risk/resilience models (e.g., King *et al.*, 2003; Murray, 2003; Spekman *et al.*, 1993; Werner and Smith, 2001; Williamson, 2007; Wong, 2003). In fact, the principles of this framework affirm the most strongly held beliefs of special educators. The risk–resilience model recruits the metaphor that children raised with multiple adverse influences, including disabilities, may still achieve positive outcomes, given the right combination and timing of protective factors. This success-story perspective laid the groundwork for the special education field.

Another reason for the appeal of the risk/resilience model in explaining the unpredictable outcomes of children with special needs is its developmental and transactional focus. What constitutes a risk or protective factor may differ in different contexts, and at different ages. Further, factors that appear to have no influence at one point may have a sleeper effect later. For example, Werner (1993) found that positive temperament did not buffer negative outcomes in a group of 18-year-olds with learning disabilities, but by the time these individuals reached the age of 32, early measures of temperament did predict positive adjustment, including marital and vocational satisfaction. Further, risk and protective factors are not likely to have equal impact, and so determining their relative importance is desirable. For example, for some outcomes, the particular risks to which a child is subjected are less important than the sheer quantity of them (e.g., Sameroff *et al.*, 1993). Most importantly, risk and protective factors can work together in different ways to produce outcomes. Protective factors embedded in a child's family and educational contexts may lead naturally to intervention efforts. Identifying the transaction of these factors can also guide intervention efforts, in that some risk factors may be more amenable to intervention at some points than at others. Another question is whether there is more of a payoff if efforts are directed to lowering risks or to adding or supporting protective factors. In this article, we describe and illustrate the use of the risk and resilience frameworks in guiding research and policy in special education. Risk and protective factors that arise from the interaction of individual, family, and community characteristics are illustrated; and implications for assessment, intervention, and teacher preparation are highlighted.

Risk and Protective Factors

Research on resilience in individuals without special needs has revealed that children who are exposed to adverse

conditions or multiple risks do not necessarily experience negative outcomes (Rutter, 2000). Of course, the dynamics of individual characteristics and environmental factors make it sometimes difficult to define resilience (Kraemer *et al.*, 2001) as risk and protective factors that mediate an individual's vulnerability to negative life outcomes may be biological, psychological, cognitive, or environmental. The ecological model by Bronfenbrenner (1986) illustrates how a child's development is affected by the transactions of individual, family, and community factors.

Individual Factors

Of course, disability itself is a risk factor that makes individuals more vulnerable to adverse conditions; these effects can persist across the lifespan. Research on risk/resilience on different types of disabilities shows similar pathways. For example, individuals with learning disabilities (LDs) typically show discrepancies between intellectual abilities and academic performance. This gap can lead to social and emotional problems in addition to academic difficulties (e.g., Wong and Donahue, 2002). Higher school dropout rates, unemployment, job difficulties, and lower rates of postsecondary school attendance are more likely to occur in individuals with LDs (Blackorby and Wagner, 1996; Goldstein *et al.*, 1998; Murray *et al.*, 2000; Spekman *et al.*, 1993). Students with emotional and behavioral disorders (EBDs) experience similar negative outcomes, as well as higher incarceration rates (e.g., Frank *et al.*, 1995; Malmgren *et al.*, 1998; Wagner, 1995). While students with mild cognitive disabilities have many similar problems following high school, these youth are even more likely to experience a constellation of poor outcomes compared to youth with LDs and EBDs (Affleck *et al.*, 1990; Blackorby and Wagner, 1996).

For many students, however, individual characteristics may lessen the impact of their disabilities. The identification of a disability itself may become a protective factor for some students and their families. For example, a student with LD stated that placement in special education helped him succeed in college (Miller, 1997). Cosden *et al.* (2002) confirmed that even factors such as physical attractiveness, athletic abilities, and social skills can be protective factors for individuals with LD. Regarding social acceptance, individual variables such as race, gender, severity of disability, and oral language ability have all been found to influence positive peer relationships (Pearl and Donahue, 2004). Other mediating factors include the individual's temperament and self-understanding. Even within samples of people with special needs, temperament has been found to be an important mediator of later development, as it elicits different responses from a variety of caring people around the individual (e.g., parents, teachers, and friends) (Cosden *et al.*, 2002; Garmezy, 1991; Werner and Smith, 2001). Chess and Thomas (1991) classified behaviors

that demonstrate temperamental characteristics in young children into three clusters: easy, difficult, and slow to warm-up temperaments. They found infants with a constellation of difficult temperamental traits were prone to behavioral problems in early and middle childhood. Negative traits of these temperament types increase the likelihood of developing behavioral problems or reduced social and academic competence (e.g., Bates and Pettit, 2007; Guerin *et al.*, 2003). Another factor that mitigates the effects of disability and enhances an individual's resilience is self-understanding. Adolescents and adults with LDs who understood and accepted their disability were more likely to have values and skills to assess their cognitive strengths and weaknesses, set up realistic goals, and strive to attain their goals. As children and adolescents, they were found to have regular responsibilities, and later established realistic education and vocational plans (Cosden *et al.*, 2002; Spekman *et al.*, 1993; Werner and Smith, 2001). Other strengths related to beliefs about one's self (e.g., self-esteem, self-efficacy, problem-solving skills, positive identity, and social competence) have also been found to be resources that help children thrive under adverse environments (Daniel and Wassell, 2002; Hagborg, 1996).

Family Factors

Perhaps the most compelling research evidence for the interactive effects of risk and protective factors on child development comes from studies of the family environment. In particular, a family's efforts and adaptability in supporting a child with special needs may be a lifetime challenge (Spekman *et al.*, 1993). In searching for resources, the family may have more contact with different professionals, which can itself be stressful (Waggoner and Wilgosh, 1990). Murray (2003) summarized a list of risk and protective factors at the family level. Risk factors include low socioeconomic status, inconsistent, harsh, and disorganized parenting style, family history of mental illness, and history of child maltreatment. In fact, growing up in an abusive home environment is a risk factor for emotional and behavioral problems across one's life span (Cicchetti and Carlson, 1989). Protective factors include secure child-caregiver attachment relationship, warm but demanding parenting style, family composition, parent level of education, and stable parent employment. For example, in one study of health-compromising behaviors among youth with LDs, EBDs, mobility impairments, and youth without disabilities, protective factors that reduced students' use of health-compromising behaviors included family connectedness, parental presence, parental expectations, and activities with parents (Blum *et al.*, 2001). Other researchers also highlight the personal characteristics of parents and child, including the fit between the child's temperament and the parenting style, and the external support available to

the family, such as a caring and supportive extended family member (Morrison and Cosden, 1997).

In addition to the aforementioned tangible factors, how parents perceive their child's academic and behavioral disabilities and strengths may lessen the stress created by the disability (Murray, 2003). Those who adjust well to having a child with special needs have been termed 'resilient families', who can appreciate the positive contributions of disability to their family life and well-being (Hastings *et al.*, 2005). These families not only understand the nature of the child's disability, but also recognize the importance of not generalizing that disability to the whole child (Switzer, 1985). For example, families of children with autism reported that family members became more connected as they pulled resources together to support their child with autism (Bayat, 2007). Additionally, parenting style mediates children's beliefs about their abilities and their own goals. For example, children's self-esteem is promoted when raised in a structured, supportive, and emotionally stable home context (Cosden *et al.*, 2002; Werner and Smith, 2001) with at least one caring parent. Resilient children had mothers who provided good role models, with education beyond high school, and good jobs (Werner and Smith, 2001). In contrast, findings of research on the nature of parental expectations for children's academic performance vary. Some studies suggest that lower parental expectations result in lower anxiety and higher overall student achievement in their children (Tollison *et al.*, 1987). In contrast, other evidence suggests that parents' high expectations facilitated resilience in their children with disabilities (Reynolds, 1999).

School and Community Factors

As schools are often the settings in which disabilities are first identified and in which special education services are first provided, characteristics of school and classroom structure, teacher beliefs and skills, and peer relationships are critical to fostering resilience (Pianta and Walsh, 1998; Werner and Smith, 2001). For example, even in schools with few resources, values and practices of teachers have been shown to be important factors in the successful inclusion of students with special needs in general education classrooms, resulting in growth in both social domains (Hutchinson *et al.*, 2002) and academic domains (Scruggs and Mastropieri, 1994).

Frustration experienced in school environments can make school itself a risk factor for individuals with special needs (Pianta and Walsh, 1998). In a study of self-perception and achievement (Lackaye *et al.*, 2006), adolescents with LDs reported more negative moods, lower self-efficacy in academic and social areas, lower levels of hope, and less investment of effort in their academic work. Importantly, a subgroup of students with LDs who were successful in their studies reported low levels of hope even though they

had similar academic performance as their peers without LDs. In addition, placement in special education impacts individuals with LDs, but results vary. One successful student with LD appreciated special education placement (Miller, 1997). Conversely, in another case study, the participant reported a dislike of special education classes and felt he missed out on many academic opportunities due to the teacher's low expectations (Miller and Fritz, 1998). The acute shortage of qualified special education teachers in urban schools may exacerbate students' feelings of frustration and experiences of ineffective teaching strategies.

The role of community factors in creating resilience in students with disabilities is still not well understood. However, some evidence suggests that supportive adults outside of family and school contexts (e.g., youth leaders and members of church groups) also serve as protective factors, through providing sustained emotional support or serving as role models (Cosden *et al.*, 2002; Garnezy, 1991). Other environmental factors such as having access to opportunities at crucial life transition points were instrumental in successful adult adaptation (Werner and Smith, 2001). For example, opportunities to take college preparation courses, to have a job, and to make a career choice created avenues to transition to adulthood. Community factors such as exposure to violence (Vig, 1996), access to playgrounds and park districts, and availability of medical care are also likely to interact with family and school efforts.

Conclusion

Risk and resilience models for explaining the developmental trajectories of individuals with disabilities hold great promise for improving special education, especially for broadening and enhancing assessment practices. First, this framework moves the field away from the prevailing deficit model that emphasizes the narrow developmental milestones that the child has not achieved. For example, evidence that social competence and positive self-perceptions explain long-term outcomes at least as well as cognitive and academic abilities suggests that more emphasis is needed on the assessment of individuals' personal-social characteristics. Similarly, professionals using a risk/resilience lens are reminded to assess a child within a constellation of individual, family, and school/community characteristics. The identification of protective factors in each of these domains and contexts is at least as important as uncovering the risk factors.

A corollary to this notion is that professionals should avoid unexamined assumptions about what personal or family factors in fact promote risk or resilience. For example, a child with a disability from a low-income environment may be a member of a close extended family whose support counteracts the effects of limited financial

resources. Similarly, the recognition that each student is embedded in multiple contexts opens up the possibility that risk factors in one context may emerge as protective factors in another setting. For example, reticence in a classroom where oral contributions are valued may look like a risk factor, but may emerge as a protective factor in family or community cultural contexts where being quiet is considered a sign of respect to elders.

Reframing the assessment of risk and protective factors surrounding a child with a special need broadens the intervention process. Examining curricula goals through the lens of family and community resources and expectations is likely to help the teacher design more culturally and contextually appropriate tasks and strategies. Similarly, recruiting intervention agents from outside the school opens up other positive human resources, recognizing the multiple contexts that support a student. For example, if there is stress in the parent-child relationship, a supportive neighbor or sibling or coach may be a better choice for helping the child with homework. Importantly, revising teacher preparation programs within risk resilience frameworks will resonate with efforts to enhance culturally responsive pedagogy in today's increasingly diverse society.

See *also*: Educating Students with Special Needs: An Overview; Inclusion of Students with Special Needs in General Education Classrooms; Parent and Family Involvement in the Education of Children with Special Needs; Self-Determination.

Bibliography

- Affleck, J. Q., Edgar, E., Levine, P., and Kortering, L. (1990). Postschool status of students classified as mildly mentally retarded, learning disabled, or nonhandicapped: Does it get better with time? *Education and Training in Mental Retardation* **25**, 315-324.
- Bates, J. and Pettit, G. (2007). Temperament, parenting and socialization. In Grusec, J. and Hastings, P. (eds.) *Handbook of Socialization*, pp 153-180. New York: Guilford.
- Bayat, M. (2007). Evidence of resilience in families of children with autism. *Journal of Intellectual Disability Research* **51**, 702-714.
- Blackorby, J. and Wagner, M. (1996). Longitudinal postschool outcomes of youth with disabilities: Findings from the National Longitudinal Transition Study. *Exceptional Children* **62**, 399-413.
- Blum, R. W., Kelly, A., and Ireland, M. (2001). Health risk behaviors and protective factors among adolescents with mobility impairments and learning and emotional disabilities. *Journal of Adolescent Health* **28**, 481-490.
- Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology* **22**, 723-742.
- Chess, S. and Thomas, A. (1991). Temperament and the concepts of goodness of fit. In Bates, J. and Rothbart, M. (eds.) *Exploration in Temperament*, pp 15-28. New York: Plenum Press.
- Cicchetti, D. and Carlson, V. (eds.) (1989). *Child Maltreatment: Theory and Research on the Causes and Consequences of Child Abuse and Neglect*. New York: Cambridge University Press.
- Cosden, M., Brown, C., and Elliott, K. (2002). Development of self-understanding and self-esteem in children and adults with learning disabilities. In Wong, B. Y. L. and Donahue, M. (eds.) *Social Dimensions of Learning Disabilities*, pp 33-51. Mahwah, NJ: Erlbaum.

- Daniel, B. and Wassell, S. (2002). *The School Years: Assessing and Promoting Resilience in Vulnerable Children II*. London: Jessica Kingsley.
- Frank, A. R., Sitlington, P. L., and Carson, R. R. (1995). Young adults with behavioral disorders: A comparison with peers with mild disabilities. *Journal of Emotional and Behavioral Disorders* **3**, 156–164.
- Garnezy, N. (1991). Resiliency and vulnerability to adverse developmental outcomes associated with poverty. *American Behavioral Scientist* **34**, 416–430.
- Goldstein, D. E., Murray, C., and Edgar, E. (1998). Employment earnings and hours of high school graduates with learning disabilities through the first decade after graduation. *Learning Disabilities Research and Practice* **13**, 53–64.
- Guerin, D. W., Gottfried, A. E., Oliver, P. H., and Thomas, C. W. (2003). *Temperament: Infancy through Adolescence*. New York: Kluwer Academic/Plenum Press.
- Hagborg, W. L. (1996). Self-concept and middle school students with learning disabilities: A comparison of scholastic competence subgroups. *Learning Disability Quarterly* **19**, 117–126.
- Hastings, R. P., Beck, A., and Hill, C. (2005). Positive contributions made by children with an intellectual disability in the family: Mothers' and fathers' perceptions. *Journal of Intellectual Disabilities* **9**, 155–165.
- Hutchinson, H., Freeman, J., and Bell, K. S. (2002). Children and adolescents with learning disabilities: Case studies of social relations in inclusive classrooms. In Wong, B. Y. L. and Donahue, M. (eds.) *Social Dimensions of Learning Disabilities*, pp 189–214. Mahwah, NJ: Erlbaum.
- King, G., Brown, E., and Smith, L. (eds.) (2003). *Resilience: Learning from People with Disabilities and the Turning Points in Their Lives*. Westport, CT: Praeger.
- Kraemer, H. C., Stice, E., Kazdin, A., Offord, D., and Kupfer, D. (2001). How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry* **158**, 848–856.
- Lackaye, T., Margalit, M., Ziv, O., and Ziman, T. (2006). Comparisons of self-efficacy, mood, effort, and hope between students with learning disabilities and their non-LD-matched peers. *Learning Disabilities Research and Practice* **21**(2), 111–121.
- Malmgren, K., Edgar, E., and Neel, R. S. (1998). Postschool status of youths with behavioral disorders. *Behavioral Disorders* **23**, 257–263.
- Miller, M. (1997). Resilience in university students who have learning disabilities. *Learning Disabilities: A Multidisciplinary Journal* **8**(2), 89–97.
- Miller, M. and Fritz, M. (1998). A demonstration of resilience. *Intervention in School and Clinic* **33**(5), 265–271.
- Morrison, G. M. and Cosden, M. A. (1997). Risk, resilience, and adjustment of individuals with learning disabilities. *Learning Disability Quarterly* **20**, 43–60.
- Murray, C. (2003). Risk factors, protective factors, vulnerability, and resilience: A framework for understanding and supporting the adult transitions of youth with high-incidence disabilities. *Remedial and Special Education* **24**(1), 16–26.
- Murray, C., Goldstein, D. E., Nourse, S., and Edgar, E. (2000). The postsecondary school attendance and completion rates of high school graduates with learning disabilities. *Learning Disabilities Research and Practice* **15**, 119–127.
- Osgood, R. L. (2008). *The History of Special Education: A Struggle for Equality in American Public Schools*. Westport, CT: Praeger.
- Pearl, R. and Donahue, M. (2004). Peer relationships. In Wong, B. Y. L. (ed.) *Learning about Learning Disabilities*, pp 131–163. New York: Academic Press.
- Pianta, R. C. and Walsh, D. J. (1998). Applying the construct of resilience in schools: Cautions from a developmental systems perspective. *School Psychology Review* **27**(3), 407–417.
- Reiff, H., Gerber, P., and Ginsberg, R. (1997). *Exceeding Expectations: Successful Adults with Learning Disabilities*. Austin, TX: Pro-Ed.
- Reynolds, A. J. (1999). Themes and issues in the Chicago Longitudinal Study. *Journal of School Psychology* **37**(4), 481–484.
- Rutter, M. (2000). Resilience reconsidered: Conceptual considerations, empirical findings and policy implications. In Shonkoff, J. P. and Meisels, S. J. (eds.) *Handbook of Early Intervention*, 2nd edn, pp 651–682. New York: Cambridge University Press.
- Sameroff, A. J., Seifer, R., Baldwin, A., and Baldwin, C. (1993). Stability of intelligence from preschool to adolescence: The influence of social and family risk factors. *Child Development* **64**, 80–97.
- Scruggs, T. and Mastropieri, M. (1994). Successful mainstreaming in elementary science classes: A qualitative investigation of three reputable cases. *American Educational Research Association Journal* **31**, 785–811.
- Spekman, N. J., Goldberg, R. J., and Herman, K. L. (1993). An exploration of risk and resilience in the lives of individuals with learning disabilities. *Learning Disabilities Research and Practice* **8**(1), 1–18.
- Switzer, L. (1985). Accepting the diagnosis: An educational intervention for parents of children with learning disabilities. *Journal of Learning Disabilities* **18**, 151–153.
- Tollison, P., Palmer, D., and Stowe, M. (1987). Mothers' expectations, interactions, and achievement attributions for their learning disabled or normally achieving sons. *Journal of Special Education* **21**, 84–93.
- Vig, S. (1996). Young children's exposure to community violence. *Journal of Early Intervention* **20**(4), 319–328.
- Waggoner, K. and Wilgosh, L. (1990). Concerns of families of children with learning disabilities. *Journal of Learning Disabilities* **23**, 97–98.
- Wagner, M. (1995). Outcomes for youths with serious emotional disturbance in secondary school and early adulthood. *Future of Children* **5**, 90–112.
- Werner, E. (1993). Risk and resilience in individuals with learning disabilities: Lessons learned from the Kauai longitudinal study. *Learning Disabilities Research and Practice* **8**, 28–34.
- Werner, E., Bierman, J., and French, F. (1971). *The Children of Kauai: A Longitudinal Study from the Prenatal Period to Age Ten*. Honolulu, HI: University of Hawaii Press.
- Werner, E. and Smith, R. (1982). *Vulnerable but Invincible: A Longitudinal Study of Resilient Children and Youth*. New York: McGraw-Hill.
- Werner, E. and Smith, R. (1992). *Overcoming the Odds: High Risk Children from Birth and Adulthood*. Ithaca, NY: Cornell University Press.
- Werner, E. and Smith, R. (2001). *Journeys from Childhood to Midlife: Risk, Resilience, and Recovery*. New York: Cornell University Press.
- Williamson, C. (2007). *Black Deaf Students: A Model for Educational Success*. Washington, DC: Gallaudet University Press.
- Wong, B. Y. L. (ed.) (2003). *Special Issue: General and Specific Issues for Researchers' Consideration in Applying the Risk and Resilience Framework to the Social Domain of Learning Disabilities*. *Learning Disabilities Research and Practice* **18**, 68–76.
- Wong, B. Y. L. and Donahue, M. (eds.) (2002). *Social Dimensions of Learning Disabilities*. Mahwah, NJ: Erlbaum.

School Dropout Prevention

F E Obiakor, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

In an ideal world, people would always benefit from a difficult situation and turn it to some personal gain. However, to categorically assume that a person will emerge renewed from a crisis is naïve and ignores the reality of what life crises can do to a person. I do not believe that every life crisis offers a challenge or an opportunity for positive growth experience, a chance to emerge a better and stronger person. In the real world, an event may leave a person changed in ways that could only be construed as adverse. A person's failure to gain strength from such an event reflects the reality of personal experience, not some problem in the attitude or perspective of the person. (Holmes, 1994, p. ix)

Holmes' (1994) statement reflects intriguing realities that human beings endure. Understandably, bad stuff happens to good or bad people – that is the reality of personal experience. Since personal experience is a part of life's reality, blaming the victim can have far-reaching negative psychological and educational consequences (Obiakor *et al.*, 1997). One such consequence is learned helplessness that results in the self-fulfilling prophecy that eats deeper at the resiliency quotient or the inner power of people with and without special needs (Rosenthal and Jacobson, 1968).

Holmes (1994) explained that real-life problems can be temporal or lasting. He indicated:

To efficiently address a temporal real-life problem, the person experiencing the problem and those who are trying to help must deal with the reality of what the person is experiencing. Neither the person nor the helper should approach efforts to resolve distress as if the origin of the problem were psychopathological. It is not. Moreover, life situations outside the person's control may change, the person may learn new ways to cope, or the person may learn ways to change the situation itself. Thus, the possibility of change offers hope of eventual resolution of the problem. Nonetheless, the distress must be endured, and the person's efforts to cope with that distress require support. (p. 4)

He added:

One must keep in mind when considering lasting real-life problems that the psychological and emotional damage of the event has been done; there is no possibility of resolution through any action that might be taken by the person or the helper. However, to say that resolution is not

possible is not the same as saying that there is nothing one can do to help the person deal with conflict and distress. (p. 5)

Clearly, real-life problems affect real people, whether they are young, youthful, or old. However, how we deal with these problems has continued to matter in our families, schools, communities, and nation (Holmes, 1994; Obiakor *et al.*, 1997). Just as grown-ups quit their jobs, divorce their husbands or wives, or engage in strange actions when they feel overwhelmed by life's crises, our children and youth drop out of school and fail to complete their education when confronted with their own life's crises. Rather than make negative judgments and draw illusory conclusions that are based on biological determinism and pathological interpretations (Gould, 1981), general and special educators can creatively help our youth to stay in school, dream big dreams, graduate from school, and become productive citizens. In this article, commonsense approaches to preventing dropout and increasing school completion are discussed.

School Dropout in Context

School dropout and school completion have been burning issues confronting general and special educators and school leaders for a long time. Schools and government agencies have made some efforts to deal with these issues. Consider the following recent examples. The promulgation of the 2001 No Child Left Behind Act (Public Law 107-110) presented an excellent rationale for leaving no child behind and dealing with all educational problems confronting our children and youth. It was supposed to reduce and eliminate the existing achievement gaps between disadvantaged students and, to a large extent, increase institutional accountability for adequate yearly student progress. In 2004, the Individuals with Disabilities Education Improvement Act (2004; Public Law 108-446) was passed to deal with educational problems confronting learners with special needs. While these legislative efforts have created some form of accountability in educational programming, they appear to have "scotched the snake, but not killed it." The goal of educating all learners is still far from reach (Obiakor *et al.*, 2002b).

At present, dropout rates among students with special needs are still high and pervasive. According to

Williams Bost and Riccomini (2006), dropout rates “vary by characteristics such as ethnicity, socioeconomic status, geographical location, and type of disability” (p. 301). Earlier, the US Department of Education (2003) noted that students with emotional and behavioral disorders and students with learning disabilities drop out more than others with disabilities. The intriguing question continues to be, what actually causes school drop-out? In her study, Williams Bost (2007) found some factors that cause school dropout. As she pointed out:

- Problem behaviors coupled with academic difficulties or prior academic failures are key risk factors that are predictive of school dropout.
- The repeated use of exclusionary discipline practices, such as suspension, has been identified as one of the major factors contributing to dropout. Exclusion from class due to disciplinary action also leads to lost instructional time and increased academic difficulties.
- Academic progress and school completion are not equally distributed across disability, income, or ethnicity. Almost half of youth with emotional disturbances drop out. Youth with disabilities from low-income households continue to experience high dropout rates, and Hispanic youth have experienced the smallest improvement in school completion over time.
- High absenteeism and being held back a grade are serious risk factors for dropping out that can be monitored by schools.
- Feelings of isolation and alienation often lead to social and psychological disengagement that result in school dropout. (p. 1)

In addition to the above points, many disenfranchised students (e.g., culturally and linguistically diverse (CLD) students), with and without exceptionalities, drop out of school because of endemic school-related and teacher-related factors such as (1) misidentification, (2) misassessment, (3) miscategorization/mislabeling, (4) misplacement, and (5) misinstruction/misintervention (Obiakor, 2001, 2007, 2008; Obiakor and Beachum, 2005a, 2005b; Obiakor *et al.*, 2002a). There is nothing more frustrating than to be educated in an environment where one is expected to be ‘invisible’ and see himself/herself as a ‘silenced voice’. When CLD students feel unwanted, they perform less and fail to maximize their fullest potential. When expectations are low, inaccurate, or prejudicial, the potential of students is not maximized, and productivity suffers (Obiakor, 1999; Obiakor *et al.*, 2002a). Additionally, when assumptions are negative and based on biological determinism (i.e., genetic superiority or inferiority), underlying pathological deficit-oriented technique is prescribed; and students suffer (see Gould, 1981). It is in the interest of the society to deal with issues of school dropout and school completion.

Consider the following negative consequences of school dropout:

- Dropouts have fewer options for postsecondary education than do students who remain in school. Additionally, only a few dropouts complete a general educational development (GED) within 2 years of leaving high school.
- Dropouts are more likely to be unemployed or employed in low-skilled, lower-paying positions. While initial earning may be comparable between dropouts and graduates, dropouts tend to work more jobs and earn less per hour than graduates.
- Dropouts also experience a ceiling effect in earning power much sooner than graduates.
- Dropouts are more likely to commit crimes as compared to students who complete school. Three to 5 years after dropping out, the cumulative arrest rate for youth with serious emotional disturbance (SED) is 73%.
- Dropouts are more likely than high school graduates to need the support of living with parents in early adulthood, experience health problems, engage in criminal activities, and become dependent on welfare and other government programs. (Williams Bost, 2007, p. 2)

Dropout Prevention to Increase School Completion

It is imperative that school completion be the goal of students, families, schools, communities, and government. To reduce student dropout and increase school completion, all stakeholders must collaborate, consult, and cooperate with each other. In other words, the whole village must work together to enhance school completion. As a result, the CSM is prescribed to connect all stake-holding entities in helping students to complete school (Obiakor, 2001; Obiakor *et al.*, 2002b). The CSM has mutually inclusive elements that are operational when:

- The development and use of identification, assessment, and instructional strategies function within multidimensional and cultural contexts.
- The creation of a collaborative system of community support for families has its guiding principle in the eradication of social stereotyping based on race, ethnicity, national origin, gender, and socioeconomic status.
- The development of an awareness and appreciation for the many family forms values individual differences and strengths.
- The thwarting of conditions leading to violence in the home or the community cultivates a sense of safety for children and families.
- The advocacy for economic policies and human services attests to being pro-family by virtue of proven outcomes.
- The promotion of culturally competent practices in schools and in the larger society respects differences in worldviews and learning styles among individuals.

- The advocacy for expanded services provides for affordable quality childcare to meet the varied needs of all families and children (e.g., infant and adolescent 24-h care and weekend care).
- The development of collaborative community approaches to problem solving involves students, parents, schools, community leaders, and government agencies.
- The recognition that the problem in at-risk situations is not only in the individual but also in institutional barriers in the environment.
- The reconfiguration of curricula eliminates the hidden curriculum and other culturally insensitive curricula variables.
- The reinstitution of rites of passage and service opportunities cultivates a sense of belonging and resiliency in youth.
- The broadening of visions in educational reform includes economic reform and the investment in human capital.

These aforementioned elements must be functional in nature and lead to goal-directed decisions of stakeholders (i.e., students, families, schools, communities, and government agencies). Surely, these stakeholders have to play specific and interrelated roles to maximize the potential for school completion in this age of change.

The Student's Role in Preventing School Dropout

Based on the CSM, the student has roles to play in increasing his/her school completion. This is not the traditional blame-the-victim idea; it is the individual's power and ability to be involved in his/her destiny (Obiakor and Weaver, 1995). There are success stories of persons who have pulled themselves up by their own boot straps, even though some had boots without straps and straps without boots. Since the self is so important in reducing school dropout, students must be taught to be resilient and believe in themselves (Obiakor and Beachum, 2005a). Specifically, they can be taught to:

- develop self-talks and individual plans;
- relax and not jump to conclusions;
- learn to work collaboratively and consultatively with others;
- engage in positive thinking;
- talk with counselors about personal and school problems;
- be a part of school conflict-resolution teams;
- inform adults and parents when situations are not going right;
- manage their time properly;
- respect school regulations and society's laws; and
- utilize mentors from the school and community.

The Family's Role in Preventing School Dropout

Family functioning and parent-child relationships have greater influence in career development than family structure or parents' educational and occupational status (Kerka, 2000). Parents must be proactive, involved, and supportive. Negative home circumstances can affect school performance; and when parents are discouraged in the education of their children, they become unaware of how their children perform. Kerka concluded that proactive families:

- are well-organized, cohesive, and expressive;
- are extroverted and manage conflict positively;
- seek out ways to grow;
- make decisions through the democratic process;
- are sociable;
- encourage individual development;
- are emotionally engaged; and
- are willing to work with their child, school authorities, and community and government agencies.

The School's Role in Preventing School Dropout

Schools have the power to uplift humanity when teachers and service providers are well prepared. Poorly prepared, ill-prepared, or unprepared professionals negatively impact their students. Renschler (1992) agreed that schools can increase students' motivation by implementing policies that promote:

- goal-setting and self-regulation;
- student choices;
- student achievements;
- teamwork and cooperative learning; and
- self-assessment models rather than social comparisons.

Apparently, schools can reduce the failure syndrome (Brophy, 1998) if they are to increase school-completion strategies for students. The failure syndrome can be reversed when schools value their students, collaborate with families, work with community members, and consult with government agencies (King, 2003; Obiakor *et al.*, 2002a; 2002b). For instance, King concluded that teachers and service providers must arrange and modify their classrooms and programs to:

- facilitate on-task behaviors;
- facilitate listening and attending skills;
- facilitate academic performance;
- make implementation of a behavior management system easy;
- allow for large, small, and cooperative grouping and one-on-one instruction;
- have a place for students to relax; and
- provide students with private space. (p. 12)

In addition to the above points, schools must have prudent professionals who can use common sense approaches to solve problems (Algozzine, 1995). These professionals must engage all students with realistic expectations and avoid the myth of socioeconomic dissonance (i.e., when poverty is viewed as the ultimate cause of all students' malaise). Poverty does not mean that students and their parents have poor intelligence, poor self-concept, and poor zest for success (Obiakor, 1999, 2001, 2007, 2008; Obiakor and Beachum, 2005a, 2005b; Utley and Obiakor, 2001). To avoid all forms of prejudicial, xenophobic, and racist actions that force some students to hate school or drop out of school, general and special educators and school leaders must:

- know who they are;
- learn the facts when they are in doubt;
- change their thinking;
- use resource persons;
- build self-concepts;
- teach with divergent techniques;
- make the right choices; and
- continue to learn.

The Community's Role in Preventing School Dropout

The community traditionally houses a wealth of resources that students can access for academic and social development (Ford, 2002). Additionally, students can take advantage of the many learning opportunities that are available within the community (e.g., libraries, museums, schools, jobs, and entrepreneurial offers). Some communities are more forward-looking than others, and some are extremely destructive to their children and youth. Dooley and Toscano-Nixon (2002) concluded that some communities are:

- Dysfunctional and struggling – in such communities, “the problem can be traced back to either the role that community members are playing or the direction that community members are following” (p. 103).
- Borderline – in such communities, “there is universal community participation and the citizens rely heavily on the government to impose changes” (p. 104).
- Conscientious – in such communities, “all members assume their social and moral responsibilities to their community's social and economic growth” (p. 106).

To reduce dropout rates and increase school-completion rates, the goal is to have conscientious communities that:

- develop cutting-edge programs that build capacity for change for students at risk;
- help to build responsible citizens through churches, mosques, synagogues, and community agencies (e.g., YMCA,

YWCA, Boy Scouts, Girl Scouts, Boys and Girls Club, Urban League, to name a few);

- discover softer ways to manage behavior problems and not build jails/prisons to replace schools; and
- have as their slogan, “Together we can make a difference,” and as their principle, “It takes a whole village to raise a child.”

The Government's Role in Preventing School Dropout

Government initiatives (e.g., No Child Left Behind Act and the Individuals with Disabilities Education Improvement Act) have been instrumental in buttressing some levels of accountability in schools and communities. In many cases, they have provided funding to various institutions, school districts, and community organizations to create innovative school-completion programs such as charter, voucher, and choice schools. However, to effectively increase school completion of students, the government must support programs by:

- establishing policies that buttress positive changes and advancements;
- instituting laws that protect its citizens; for instance, due process of students, parents, and teachers must be maintained;
- imposing penalties on institutions that violate the civil rights of students;
- funding research to discover new ways of doing things;
- coordinating conferences that bring professionals and agencies together;
- making programs accountable to their consumers/students;
- assisting institutions in shifting their paradigms and powers; and
- rewarding visionary leaders and programs that do what they are supposed to do.

Conclusion

In this article, it has been argued that everyone goes through some form of crisis. There are apparent and underlying stressors that impinge upon all human activities. While grown-ups lose it and quit their jobs or divorce their husbands/wives, students drop out of school and become societal problems. It is imperative that general and special education teachers, service providers, and leaders build proactive programs that can reduce dropout and increase school completion. Students, parents, teachers, community members, and policymakers must work together for the common good. It is in this spirit that the CSM is prescribed to take advantage of the energies of all stakeholders.

We must avoid unwarranted labels, derogatory categories, and illusory generalizations that disempower students. We strongly believe that disempowered students drop out of school, do not graduate, and get involved in anti-social behaviors. Finally, we must practicalize the Native American proverb, "all the flowers of all the tomorrows are in the seeds of today." In the words of Brooks (1991):

While all students deserve to have their islands of competence displayed and built upon, there is a more urgent need to do so for those students who lack confidence in their ability to learn. If we can reinforce the areas of strength these students possess, my experience has been that we can open the way for a 'ripple effect', where students may be willing to venture forth and confront tasks that have been problematic for them. (p. 32)

See also: Community Integration and Employment of Youth with Special Needs; Cooperative Learning for Children with Special Needs; Co-Teaching; Cultural and Linguistic Diversity and Children with Special Needs; Early Intervention for Students with Special Needs; Early Transition of Children with Special Needs; Functional Behavioral Assessment; Inclusion of Students with Special Needs in General Education Classrooms; Interdisciplinary Collaboration in Early Childhood; Overrepresentation of Culturally and Linguistically Diverse Students in Special Education; Parent and Family Involvement in the Education of Children with Special Needs; Peer Relations and Socialization of Children and Adolescents with Special Needs; Peer-Tutoring; Postsecondary Participation of Students with Special Needs; Preventing Antisocial Behavior and Delinquency: A Comprehensive School-wide Approach; Risk and Resilience Frameworks in Understanding Special Education; School-Based Services for Children with Special Needs; Self-Determination; Social Competence; The Identification of Students with Gifts and Talents; Transition from School to Adult Life; Workforce Issues in Special Education.

Bibliography

- Algozzine, B. (1995). *Teachers Little Book of Wisdom*. Merrillville, IN: ICS Books.
- Brooks, R. (1991). *The Self-Esteem Teacher*. Circle Pines, MN: American Guidance.
- Brophy, J. (1998). Failure syndrome students. *Elementary and Childhood Education* May, 1-3.
- Dooley, E. A. and Toscano-Nixon, C. M. (2002). Community involvement in education. In Obiakor, F. E., Grant, P. A., and Dooley, E. A. (eds.) *Educating All Learners: Refocusing the Comprehensive Support Model*, pp 99-112. Springfield, IL: Charles C. Thomas.
- Ford, B. A. (2002). African American community resources: Essential education enhancers for African American children and youth. In Obiakor, F. E. and Ford, B. A. (eds.) *Creating Successful Learning Environments for African American Learners with Exceptionalities*, pp 159-173. Thousand Oaks, CA: Corwin Press.
- Gould, A. J. (1981). *The Mismeasure of Man*. New York: Norton.
- Holmes, B. C. (1994). *Like a Lasting Storm: Helping with Real Life Problems*. Brandon, VT: Clinical Psychology Publishing.
- Individuals with Disabilities Education Improvement Act. (2004), Pub. L. no. 108-446.
- Kerka, S. (2000). Parenting and career development. *Career Education* 214, 1-5.
- King, D. (2003). Tips to help new teachers with classroom management. *CEC Today* September/October, 12.
- No Child Left Behind Act. (2001), Pub. L. no. 107-110.
- Obiakor, F. E. (1999). Teacher expectations of minority exceptional learners: Impact on "accuracy" of self-concepts. *Exceptional Children* 65, 39-53.
- Obiakor, F. E. (2001). *It Even Happens in "Good" Schools: Responding to Cultural Diversity in Today's Classrooms*. Thousand Oaks, CA: Corwin Press.
- Obiakor, F. E. (2007). *Multicultural Special Education: Culturally Responsive Teaching*. Upper Saddle River, NJ: Pearson Merrill/Prentice Hall.
- Obiakor, F. E. (2008). *The Eight-Step Approach to Multicultural Learning and Teaching*, 3rd edn. Dubuque, IA: Kendall/Hunt.
- Obiakor, F. E. and Beachum, F. D. (2005a). Developing self-empowerment in African American students using the Comprehensive Support Mode. *Journal of Negro Education* 74, 18-29.
- Obiakor, F. E. and Beachum, F. D. (2005b). *Urban Education for the 21st Century: Research, Issues, and Perspectives*. Springfield, IL: Charles C. Thomas.
- Obiakor, F. E. and Weaver, K. A. (1995). Has the U.S. become a nation of victims? *Emporia Gazette* May 17, 7.
- Obiakor, F. E., Mehling, T. A., and Schwenn, J. O. (1997). *Disruption, Disaster, and Death: Helping Students Deal with Crises*. Reston, VA: The Council for Exceptional Children.
- Obiakor, F. E., Algozzine, B., Thurlow, M., et al. (2002a). *Addressing the Issue of Disproportionate Representation: Identification and Assessment of Culturally Diverse Students with Emotional or Behavioral Disorders*. Arlington, VA: Council for Children with Behavioral Disorders, the Council for Exceptional Children.
- Obiakor, F. E., Grant, P. A., and Dooley, E. A. (2002b). *Educating All Learners: Refocusing the Comprehensive Support Model*. Springfield, IL: Charles C. Thomas.
- Renchler, R. (1992). School leadership and student motivation. *Educational Management* 71, 1-5.
- Rosenthal, R. and Jacobson, L. (1968). *Pygmalion in the Classroom*. New York: Holt, Rinehart, and Winston.
- US Department of Education (2003). *Twenty-Fourth Annual Report to Congress on the Implementation of IDEA*. Washington, DC: US Government Printing Office.
- Utley, C. A. and Obiakor, F. E. (2001). *Special Education, Multicultural Education, and School Reform: Components of Quality Education for Learners with Mild Disabilities*. Springfield, IL: Charles C. Thomas.
- Williams Bost, L. (2007). *Building Effective Dropout Prevention Programs - Some Practical Strategies from Research and Practice*. Clemson, SC: National Dropout Prevention Center for Students with Disabilities, Clemson University.
- Williams Bost, L. and Riccomini, P. J. (2006). Effective instruction: An inconspicuous strategy for dropout prevention. *Remedial and Special Education* 27, 301-311.

School-Based Services for Children with Special Needs

S O Obi, Kentucky State University, Frankfort, KY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Assistive technology – Devices or services to help compensate for an individual's disability.

Cascade of educational services – A range of placement and instructional options for children with disabilities with placement ranging from the general education classroom to special school, residential, home bound, or hospital placement.

Collaboration – Professionals working in partnerships to provide educational services.

Coteaching – General and special education teachers' team teaching.

Due process – Noncourt proceeding before an impartial hearing officer, used when parents and school personnel disagree on a special education issue.

Inclusion – Educating students with disabilities in general education classrooms.

Individualized education program – Management tool to identify and organize needed services.

Individualized family service plan – Identifies and organizes services and resources for infants and toddlers (birth to age three) and their families.

Related services – Special education services from a wide range of disciplines and professions.

Special education – Individualized education and services for students with disabilities sometimes including gifted and talented students.

The Education of All Handicapped Children's Act of 1975 and its reauthorizations and amendments guarantee a free and appropriate public education (FAPE) for all youngsters with disabilities through age 21. Although, some of these laws (e.g., PL 99-457 and PL 101-476) include incentives and mandates for states to implement services for infants, toddlers, and their families, they provide guidelines for assisting high-school students make the transition from school to postschool activities (Smith and Luckasson, 1995). It is common knowledge that special education is based on the individual needs of the student; and it is also clear that individualized instruction is delivered in a wide range of settings, ranging from itinerant teachers, who travel from school to school, to center schools. To a large measure, school-based services are necessary to help students with special needs to maximize their potential. This is the thrust of this article.

School-Based Educational Options

The laws require that students with special needs be educated in the least restrictive environment (LRE). These environments may be:

- *Regular class*. Student remains in the regular class. The teacher and/or student receive assistance from a specialist – itinerant or consultative.
- *Resource class*. Student attends a regular class most of the day but goes to a special education class several hours per day or for blocks of time each week.
- *Special education class (partially self-contained)*. Student attends a special class but is integrated into regular education classes for a considerable amount of time each day.
- *Special education class (self-contained)*. Student attends a special class most of the school day and is inducted into regular education activities minimally.
- *Special education schools (center schools)*. Some private separate school facility, public residential facility serve only students with a specific category of disability. Some offer residential services; others do not.

Apparently, special education services should be flexible and responsive to the needs of the student. Students with special needs should have access to a variety of services, varying by type, intensity, location, personnel, and length of time. For example, in some cases, the regular education classroom can meet the needs of the student with some support from a consulting teacher or specialist (e.g., a speech/language pathologist). At other times, more intense services are required. As long as an array of services is available, students should not be forced to prove themselves at every step of the ladder before finally entering a regular classroom program (see **Table 1**).

Special education is meant for infants, preschoolers, and elementary through high-school students with disabilities, and (in some cases) individuals with disabilities up through the age of 21. In other words, the laws (see the Individuals with Disabilities Education Improvement Act of 2004; IDEIA 2004) require that individualized education plan be provided to all students who are receiving special education services. Clearly, at the heart of individualized programs are individualized education plans (IEPs) for school children ages 3–21 and individualized family service plans (IFSPs) for infants and toddlers (birth through age 2) with disabilities and their families. In some states, the guarantee of an individualized education is

Table 1 Cascade of educational services for students with special needs

Homebound or hospital
Eligible student receives services at home or in a hospital program for a minimum of four consecutive weeks as documented by a physician.
Separate classroom
Student receives special education and related services for most of the day under the direction of a special education teacher.
Student is integrated for some specials on an individual basis.
Separate school
Student receives special education and related services under a specially trained staff in a nonintegrated setting. Student is transported every day.
Residential school
Student receives special education and related services under a specially trained staff. The student actually lives at the school and receives care or services 24 h a day.
Resource room
Student is in a general education setting for the majority of the school day but goes to the special education resource room for specialized instruction for part of each school day.
Regular classroom with supplementary instruction and services
Student receives a prescribed program under the direction of the regular classroom teacher and also receives instruction and related services within the regular classroom from the special educator and/or a parent educator.
Regular classroom with consultation
Student receives a prescribed program under the direction of the regular classroom teacher, who is supported by ongoing consultation from the special educator(s).
Regular classroom
Student receives a prescribed program under the direction of the regular classroom teacher.

extended to gifted students as well, but because gifted students' education is not addressed by federal law, this is not a requirement. IEPs and IFSPs are cornerstones that guarantee appropriate education to each student with a disability. An appropriate education is tailor-made, individually designed, and complete with supportive (related) services. In addition, parents of students with disabilities have important roles to play. For example, they are expected to participate in the development of their children's IEPs. One idea behind the IEP is for parents to become partners with teachers and schools – as a result, many parents and families participate actively in decision making about their child's education. All parents of children with disabilities enjoy the right to due process, procedures to follow when they do not agree with schools about the education planned for or being delivered to their children. Moreover, they are entitled to services not usually offered to parents of typical learners. For example, for infants and toddlers with disabilities (ages birth to 2), parents and their children receive intensive instruction through special education (Smith, 2007).

An important difference between special education and general education is the array of services offered to

students and their families. Special education provides additional services to help students with disabilities profit from instruction. It includes direct services from special education teachers, as well as instruction and therapy from related services professionals – experts from a broad array of disciplines other than education. These multidisciplinary services, in many cases, are what make inclusion possible for many students with special needs, because they provide individualized assistance to them for extended periods of the school day. Three commonly used related services are speech therapy, physical therapy, and assistive technology (AT). Therefore, special education is a comprehensive set of services designed to support the education of students with disabilities. Specifically, special education involves many different professions and specialty areas. When some or all of these related services are enlisted, the result is multidisciplinary teams (i.e., groups of professionals with different areas of expertise who work together to meet the educational needs of each student with a disability) (Smith, 2007).

The overarching purpose of special education is to make it possible for all children with special needs to achieve to their fullest potential so that as adults, they can attain full community presence by holding meaningful jobs and living independently. Inclusive education – that is, participating exclusively in the general education classroom using the general education curriculum – is not necessarily a goal of special education (Kavale and Mostert, 2003). However, too few students with disabilities leave school with a standard diploma; and parents, policymakers, and advocates insist that these students should participate in the general education curriculum and be a part of accountability measures (e.g., state- and district-wide tests) that monitor all students' progress (US Department of Education, 2002). Increased participation in the general education is expected to lead to increased graduation rates. Therefore, beginning with Individuals with Disabilities Education Act (IDEA) '97, extended through No Child Left Behind Act (NCLB) of 2001, and reinforced in Individuals with Disabilities Education Improvement Act (IDEIA) 2004, IEPs must address students' access and participation in the general education curriculum and justify any limitations (Wehmeyer *et al.*, 2003). If a student is removed from the typical general education curriculum, the IEP must specifically explain why the student cannot participate at this particular time (US Department of Education, 2005). As indicated in IDEIA 2004, general education curriculum is not appropriate for all students with disabilities. Some require an alternative curriculum, intensive treatment, or supplemental instruction on topics not available or suitable for instruction in the general education classroom. For example, students with blindness and visual impairments need to master orientation and mobility training and learn job skills in community placements; students with cognitive

disabilities need to learn how to use public skills training students with cerebral palsy might need physical therapy; student with stuttering problem might need speech therapy; students with learning disabilities might need phonics instruction, and so on (Smith, 2007).

School-Based Service Professionals

Many different types of professionals are needed to provide services required by children with special needs. A special education professional might be a paraprofessional (teacher's aide), a resource teacher, a consultant, an itinerant teacher, a special education classroom teacher, a job coach, an AT specialist, a diagnostician, or an administrator. Furthermore, a teacher may assume several of these professional roles during his/her career. In addition, a special education professional might work in a related service as a school psychologist, a speech/language pathologist, an audiologist, an occupational therapist, a physical therapist, a counselor, a nurse or physician, a transportation specialist, or a recreational therapist. The detailed roles of the professionals as indicated in IDEIA 2004 are as follows:

- *Special educator*: One who might be a resource specialist, a consultant, an itinerant teacher, a special education classroom teacher, a job coach, a home or hospital teacher, or an administrator. The skills needed by special educators are many. They must have in-depth knowledge about making accommodations, differentiating instruction, implementing practices validated through rigorous research, monitoring students' progress, and ensuring that every student with a disability receives an appropriate education. Because of the requirements that teachers be highly qualified, co-teaching is gaining in popularity, particularly at the middle- and high-school levels, where meeting the requirements of every core subject area that special educators teach is not possible. Blending the expertise of general education professionals (e.g., math, science, history, and English) and special educators through co-teaching arrangements can make services received truly special (Magiera *et al.*, 2005). Special educators (1) assist with data collection, (2) test effectiveness of educational modifications and accommodations, (3) individualize instruction, (4) apply instruction to small groups, (5) adapt material and instruction, (6) consult with and provide assistance to other educators, monitor progress frequently, (7) determine appropriate adaptations and accommodations, (8) foster parent partnerships, (9) communicate with school personnel about needed accommodations, (10) train and supervise paraprofessionals, (11) advocate for each student with a disability, (12) coordinate students' related services, (13) conduct in-service training sessions about access to the general education curriculum, (14) develop the IEP, and (15) maintain records of accommodations and IEP progress.
- *Assistive technologist*. One who offers high-tech equipment, low-tech aids, and other devices that help compensate for the needs of children with special needs (e.g., speech-to-text translators, holders to keep books open, and wheelchairs).
- *Audiologist*. One who examines hearing losses and auditory problems.
- *Social workers, psychologists, and guidance counselors*. School counselors are related professionals who provide individual counseling, group counseling, guidance, and consultation.
- *Interpreters*. They translate auditory communications into sign language for deaf students.
- *School nurses*. They provide and assist with special health care needs.
- *Occupational therapists (OTs)*. They direct activities that help improve muscular control and develop self-help skills.
- *Orientation mobility specialists*. They teach blind students and students with low vision how to move through environments, and travel safely and efficiently from place to place.
- *Social counselors, parent counselors, and psychologists*. These are service providers who assist parents in preparation for IEP meetings and in becoming more involved in their child's special education programs.
- *Physical therapists (PTs)*. They treat physical disabilities and motor problems through many nonmedical means.
- *School psychologists, clinical psychologist, educational diagnosticians, and psychometricians*. They are service providers who administer psychological and educational tests and interpret results.
- *Speech/language pathologists (SLPs)*. They diagnose and treat problems to the area of speech and language development.
- *Therapeutic physical therapists and recreation specialists*. They assess and provide instruction in leisure skills and recreational activities.
- *Transportation specialists*. They provide travel to and from school, as well as in and around school.
- *School principal*. A key person in the collaborative effort at every school (Praisner, 2003). As principals often coordinate management efforts at their site, they can be most helpful in developing and ensuring the delivery of accommodations and modifications, in monitoring the array of services indicated on every student's IEP, and in ensuring the coordination of services throughout the school and across the district.
- *General educators*. They conduct prereferral assessments, instruction using tactics of varying intensity, instruction

under different conditions, apply instruction to whole class and small groups, ensure maximal access to the general education curriculum by implementing adaptations and accommodations, train and supervise peer tutors, develop and maintain portfolios, implement adaptations and accommodations for testing situations, communicate with parents and families, work in partnership with special education personnel, work with paraprofessionals, participate in IEP meetings, facilitate the scheduling and delivery of related services, maintain anecdotal records, and keep records of accommodations use and effectiveness.

- *Special needs students.* When these students are of age they are involved in the development of their own IEPs. They contribute to their IEP meetings by (1) describing their strengths, weaknesses, and needs; (2) evaluating their progress toward accomplishing their goals; (3) bringing a list of accommodations, and explaining how each is helpful; (4) communicating their preferences and interests; and (5) articulating their long-term goals and desires for life, work, and post-secondary schooling.

As it stands, special education teachers and other professionals and service providers, who work with students with special needs, must collaborate and work cooperatively with professionals from a variety of disciplines, since students with special needs require various combinations of services over their school careers. These teachers, professionals, and service providers collaborate in many areas that touch a child's life. They use collaboration skills when working with parents and families, performing multidisciplinary assessments, working with a team to develop individualized plans, and helping them make the transitions through the school years and from school to work or work to school (Smith and Luckasson, 1995).

Special Education Categories

Through IDEIA 2004 the federal government describes 13 disability categories that can be used to qualify infants, toddlers, preschoolers, and students eligible to receive special education services. Within these categories are many conditions, such as stuttering, included as a speech impairment; or attention-deficit hyperactivity disorder (ADHD), included under the other health impairment category; and Tourette's syndrome, in the government's emotional disturbance category. Moreover, in an attempt to avoid either incorrectly labeling young children as having a disability when they do not or identifying them with the wrong disability, the federal government provides the option of using a general category (non-disability-specific group) for children under the age of

8 (Muller and Markowitz, 2004). Each area of disability, included in the federal regulations governing programs and services for children with special needs, is defined in the following sections.

Mental Retardation (Cognitive Disability)

This is a disability that involves school-age children who are clearly and consistently behind their peers in all areas, including academic, social, language, and self-care skills that it is obvious to anyone who interacts with them that they require special education and related services. Mental retardation (cognitive disability) ranges from mild to severe, but often overlaps with low-incidence disabilities. This condition is manifested during the developmental period. An interdisciplinary team develops a profile of the types and intensity of needed supports.

Specific Learning Disabilities

This is a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written. Academic deficits are the defining feature of learning disabilities. The academic area that poses the most difficulty for most children with learning disabilities is reading. Children with learning disabilities may also have problems with written or spoken language and with math. Some may exhibit visual and/or auditory perceptual disabilities in that they are unable to organize and interpret visual and/or auditory stimuli. Children with learning disabilities are typically inactive learners who lack strategies for attacking academic problems. Two major approaches to the academic problems of students with learning disabilities are cognitive training and direct instruction. For these students who have ADHD, educational approaches should focus on structure and stimulus reduction, behavior modification, cognitive training, and modification.

Emotional and Behavioral Disorders

Emotional/behavioral disorders (E/BD) is a disability characterized by behavioral or emotional responses in school programs so different from age, cultural, or ethnic norms that they adversely affect educational performance. The inappropriate behaviors must be long-standing patterns that occur regularly and often enough as to interfere consistently with the student's own learning process. The two conceptual models that guide most educational programs are the psycho-educational model and the behavioral model. Regardless of the model that guides education, successful strategies must include (1) systematic data-based interventions, (2) continuous assessment and monitoring of progress, (3) prevision for practice

of new skills, (4) treatment matched to the problem, (5) multicomponent treatment, (6) programming for transfer and maintenance, and (7) commitment to sustained intervention.

Autism

Autism is a developmental disorder characterized by significant and pervasive impairment in several areas of development. Autism spectrum disorders (ASD) include five developmental disabilities of childhood: autistic disorder, Asperger syndrome, childhood disintegrative disorder, pervasive developmental disorder, and Rett's syndrome. Children with autism require carefully planned, meticulously delivered, and continually evaluated and analyzed instruction. Successful educational approach must include (1) applied behavior analysis, (2) discrete trial planning (DTT), an important part of applied behavior analysis, (3) picture exchange communication system (PECS), (4) social stories that explain social situations and expected behaviors, and (5) picture activity schedules – a series of images, photos, icons, or video clips depicting activities a child can perform.

Communication Disorders (Speech and Language Impairments)

Speech and language are tools used for communication; speech is the behavior of forming and sequencing the sounds of oral language and language is the communication of ideas. Communication disorders may involve speech or language or both. A speech disorder is an impairment of voice, articulation or speech sounds, and/or fluency. A language disorder is the impairment of comprehension and/or use of a spoken, written, and/or other symbol system and may involve the form, content, and/or function of language. SLPs employ a wide range of techniques for identifying, evaluating, and providing therapeutic services to children. These include (1) structured exercises and drills as well as individual and group therapy sessions, (2) helping children speak as clearly as possible, (3) treating voice disorders medically if the cause is organic, (4) encouraging precommunication activities explore expressive language, and (5) aided or unaided augmentative and alternative communication.

Deafness and Hearing Loss

Children who cannot hear sounds at or above a certain intensity level, 90 db, are classified as deaf and others are classified as hard of hearing. They may require various instructional modifications and related services in order to make full use of their learning opportunities. The three major approaches to teaching deaf and hard-of-hearing students are oral/aural approach, total communication, and

bilingual–bicultural approach. Regardless of instructional approach, the primary objective of all teachers and service providers of students with hearing loss is the development and use of language and communication skills (Heward, 2006).

Blindness and Low Vision

Students, who must learn to read Braille or use aural methods, are classified as blind and those who can read print, even though they may need magnification, have low vision. Many students who are legally blind are not educationally blind because they can read print. Numerous specialized teaching methods and curriculum materials have been developed in an effort to overcome the obstacles to learning presented by blindness and low vision. Advances in instructional methodology and, in particular, technology have greatly increased access to the general education curriculum and academic success among students with visual impairments. Educational approaches must include (1) Braille, (2) standard print prerecorded materials, (3) visual efficiency, (4) approach magnification, (5) optical devices, (6) large print, (7) instruction in orientation and mobility, (8) development of listening skills, and (9) instruction in functional living skills.

Physical Disabilities, Health Impairment, and ADHD

Children and youth with physical disabilities are those whose physical limitations or health problems interfere with school attendance or learning to such an extent that special services, training, equipment, materials, or facilities are required. Some children with ADHD are served under the other health impairments category of IDEA, with the reasoning that their condition results in a heightened alertness that adversely affects their educational performance. However, many children with ADHD are served under other learning disabilities or emotional disturbance. For children whose disability is only physical, curriculum and educational goals should be the same as for their nondisabled peers. The intensive health and learning needs of students require a complex and coordinated array of specialized instruction, therapy, and related services. Many students with physical and health impairments need (1) intensive instruction in a parallel curriculum (Bowe, 2000: 75); and (2) adoptive methods and ATs for mobility, communication, and daily-living tasks.

Developmentally Delayed

Children identified as preschool delayed are those who are ages 3 and 4 or those 5-year-olds who are ineligible for kindergarten and whose development and/or behavior is so significantly delayed or atypical that special education and related services are required.

Low-Incidence Disabilities (Severe/Multiple Disabilities, Deaf-Blindness, and Traumatic Brain Injury)

Multiple disabilities make up concomitant impairments (mental retardation–blindness, mental retardation–orthopedic impairment, etc.), the combination of which causes such severe educational problems that they cannot be accommodated in special education programs solely for one of the impairments. The term does not include deaf–blindness. Most students with severe disabilities exhibit extreme deficits in intellectual functioning. Many need special services and supports because of motor impediments; communication, visual, and auditory impairments; and medical conditions such as seizure disorders. Contemporary curriculum content for students with several disabilities is characterized by its focus on functional skills that can be used in immediate and future domestic, vocational, community, and recreational/leisure environments (Heward, 2006). Some students with severe disabilities use augmentative and alternative systems of communication (AAC), such as gestures, pictorial communication boards, and electronic communication aids.

Transitioning Students to Adulthood: Beyond Narrow Confines

Transition from school to life in the community has become the most challenging issue in special education today. Models for school-to-adult-life transition stress the importance of a functional secondary school curriculum that provides work experience in (1) integrated community job sites, (2) systematic coordination between school and adult service agencies, (3) parental involvement and support, and (4) a written individualized transition plan (ITP) to guide the entire process. Secondary school programs can enhance the competitive employment prospects for young adults with disabilities by (1) stressing functional, vocational skills; (2) conducting school-based instruction in integrated settings as much as possible; and (3) beginning community-based instruction as early as age 12 for students with severe disabilities.

AT places an important part in the transition of students with special needs. AT is any item, piece of equipment, or product that is used to increase, maintain, or improve the functional capabilities of children with special needs. It is common knowledge that many students need AT to transition from school to work, and vice versa.

Conclusion

Special education comes into consideration when a student is suspected of or has been identified as having

behavioral, emotional, sensory, or other disability that requires school-based services. This article has addressed the cornerstone for appropriate school-based services and school-based professional for children with special needs.

Bibliography

- Bowe, F. (2000). *Teaching Individuals with Physical and Multiple Disabilities*, 4th edn. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Heward, W. L. (2006). *Exceptional Children: An Introduction to Special Education*, 8th edn. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Individuals with Disabilities Education Improvement Act (2004). Pub. L. No. 108-446. 118 STAT 2647.
- Kavale, K. A. and Mostert, M. P. (2003). *River of ideology, islands of evidence*. *Exceptionality* **11**, 191–208.
- Magiera, K., Smith, C., Zimond, N., and Gebauer, K. (2005). Benefits of co-teaching in secondary mathematics classes. *Teaching Exceptional Children* **37**, 20–24.
- Muller, E. and Markowitz, J. (2004). *Disability Categories: State Terminology, Definitions and Eligibility Criteria*. Washington, DC: Project Forum, National Association of State Directors of Special Education (WASDSE).
- Praisner, C. L. (2003). Attitudes of elementary principals toward the inclusion of students with disabilities. *Exceptional Children* **69**, 135–145.
- Smith, D. D. (2007). *Introduction to Special: Making a Difference*, 6th edn. Boston: Allyn and Bacon.
- Smith, D. D. and Luckasson, R. (1995). *Introduction to Special Education: Teaching in Age of Challenge*, 2nd edn. Boston, MA: Allyn and Bacon.
- US Department of Education (2002). *Twenty-Fourth Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act*. Washington, DC: US Government Printing Office.
- U.S. Department of Education (2005). Assistance to states for the education of children with disabilities program and the early intervention program for infants and toddlers with disabilities; Proposed Regulation (Federal) Register, 34, CRF Parts 300, 301, and 304. Washington, DC: U.S. Government Printing Office.
- Wehmeyer, M. L., Lattin, D. L., Lapp-Rincker, G., and Agran, M. (2003). Access to the general education curriculum of middle school students with mental retardation: An observational study. *Remedial and Special Education* **24**, 262–272.

Further Reading

- Allen, K. E. and Cowdery, G. E. (2005). *The Exceptional Child: Inclusion in Early Childhood Education*. New York: Thomson/Delmar Learning.
- CASE (1999). *Section 504 and the ADA: Promoting Student Access – a Resource Guide*, 2nd edn. Arlington, VA: Council of Administrators of Special Education.
- Corbett, J. (2001). *Supporting Inclusive Education: A Connective Pedagogy*. London: Falmer Press.
- Couchenour, D. and Chrisman, K. (2004). *Families, Schools, and Communities: Together for Young Children*. New York: Thomson Delmar Learning.
- Friend, M. and Bursuck, W. D. (1999). *Including Students with Special Needs: A Practice Guide for Classroom Teachers*. Boston, MA: Allyn and Bacon.
- Kochhar, C. A., West, L. L., and Taymans, J. A. (2000). *Successful Inclusion: Practical Strategies for a Shared Responsibility*. Upper Saddle River, NJ: Prentice Hall.
- Peters, W. L. (2005). The special education process: Services and accommodations. In Schwartz, P. (ed.) *Including Children with Special Needs*. Westport, CT: Greenwood Press.
- Thomas, C. C., Correa, V. I., and Morsink, C. V. (2001). *Interactive Teaming: Enhancing Programs for Students with Special Needs*. Upper Saddle River, NJ: Merrill/Prentice Hall.

Turnbull, A. and Turnbull, R. (1997). *Families, Professionals and Exceptionality: A Special Partnership*. Upper Saddle River, NJ: Merrill.

Zirkel, P. (1999). Legal workshop on inclusion. *Paper Presented at Nassau BOCES Sliver GrantInclusion Training*. Rockville Centre: New York.

<http://www.greenville.k12.sc.us> – Greenville County Schools.

<http://www.nichcy.org> – NICHCY.

<http://www.olrs.ohio.gov> – Ohio Legal Rights Service (OLRS).

<http://www.senecacenter.org> – Seneca Center.

<http://www.ed.sc.gov> – South Carolina Department of Education.

<http://www.ed.gov> – US Department of Education.

Relevant Websites

<http://www.cec.sped.org> – Council for Exceptional Children (CEC).

<http://www.eric.ed.gov> – Education Resources Information Center (ERIC).

Self-Determination

M L Wehmeyer, University of Kansas, Lawrence, KS, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Autonomous functioning – Autonomy means, literally, self (*auto*) rule or law (*nomos*). A behavior is autonomous if the person acts (1) according to his or her own preferences, interests, and/or abilities and (2) independently, free from undue external influence or interference.

Decision making – The process of making a judgment about which of several options is the best option given one's circumstances.

Problem solving – The act of identifying and explicating a problem (e.g., a situation or event for which one does not know a solution), and selecting alternatives that address the event or circumstance.

Psychological empowerment – The multiple dimensions of perceived control, including cognitive, personality, and motivational domains of perceived control. People acting in a psychologically empowered manner do so on the basis of a belief that (1) they have control over circumstances that are important to them (i.e., internal locus of control); (2) they possess the requisite skills to achieve desired outcomes (i.e., self-efficacy); and (3) if they choose to apply those skills, the identified outcomes will result (i.e., outcome expectations).

Self-determination – The volitional actions that enable one to act as the primary causal agent in one's life and to maintain or improve one's quality of life.

Self-realization – A process whereby people use a comprehensive, and reasonably accurate, knowledge of themselves and their strengths and limitations to act in such a manner as to capitalize on this self-knowledge and self-awareness.

Self-regulation – The process of setting goals, developing action plans to achieve those goals, implementing and following action plans, evaluating outcomes of action plans, and changing action plans, if the goal is not achieved.

The international literature in special-needs education documents that an effective education for students with disabilities must include instruction to promote student self-determination. Research has linked higher self-determination to positive adult outcomes, including employment and independent living, for youth with special

educational needs (Emerson *et al.*, 2001; Wehmeyer and Palmer, 2003; Wehmeyer and Schwartz, 1997), as well as to a higher quality of life (Lachapelle *et al.*, 2005; Nota *et al.*, 2007; Wehmeyer and Schwartz, 1998). Further, most school standards for all students include a focus on skills leading to enhanced self-determination (e.g., goal setting, problem solving, decision making, self-advocacy, self-management, and others) and when instruction is available school-wide to address these component elements, all students benefit (Wehmeyer *et al.*, 2004). Finally, there is evidence that students with special educational needs can acquire the knowledge and skills to become more self-determined if provided such instruction (Algozzine *et al.*, 2001). The purpose of this article is to introduce the self-determination construct to educators working with students with special educational needs and to identify instructional and assessment strategies that enable teachers to promote self-determination.

What is Self-Determination?

There are now numerous frameworks that serve as a basis for instructional design to promote self-determination (Wehmeyer *et al.*, 2003), as well as specially designed instructional methods, materials, strategies, and assessments to promote and measure self-determination (Wehmeyer *et al.*, 2007; Wehmeyer and Field, 2007). The framework that guides the presentation of relevant interventions in this article is based on a model of self-determination (Wehmeyer *et al.*, 2003) in which self-determined behavior refers to "volitional actions that enable one to act as the primary causal agent in one's life and to maintain or improve one's quality of life" (Wehmeyer, 2005: 117). An act or event is self-determined if the individual's action reflects four essential characteristics: (1) the individual acted autonomously, (2) the behaviors were self-regulated, (3) the person initiated and responded to event(s) in a psychologically empowered manner, and (4) the person acted in a self-realizing manner. Self-determination refers to self- (vs. other) caused action to people acting volitionally, based upon their own will. The word volitional is defined as the act or instance of making a conscious choice or decision. Conscious means intentionally conceived or done – deliberate. Volitional behavior, then, implies that one acts consciously – with intent. Self-determined behavior is volitional and intentional, not simply random and nonpurposeful.

The concept of causal agency is central to this perspective. Broadly defined, causal agency implies that it is the person who makes or causes things to happen in his or her life. One frequent misinterpretation of self-determination is that it means 'doing it yourself'. When self-determination is interpreted this way, however, there is an obvious problem for most students with special educational needs, who frequently have limits to the number and types of activities they can perform independently. However, the capacity to perform specific behaviors is secondary in importance to whether one is the causal agent (e.g., caused in some way to happen) over outcomes those specific behaviors are implemented to achieve. Thus, students who may not be able to independently make a complex decision or solve a difficult problem may be able, with support, to participate in the decision-making process, and thus have the opportunity to be the causal agents in the decision-making process, and consequently, act in a self-determined manner.

Wehmeyer *et al.* (2003) have argued that self-determination emerges across the life span as children and adolescents learn skills and develop attitudes and beliefs that enable them to be causal agents in their lives. These skills and attitudes are referred to in this model as component elements of self-determined behavior, and include choice making, problem solving, decision making, goal setting and attainment, self-advocacy, and self-management skills.

Factors That Contribute to Self-Determination

Self-determination can be impacted by environmental and intra-individual factors. The following section discusses these factors.

Environmental Factors Related to Self-Determination

Research internationally has consistently and repeatedly shown that environments in which adults and youth with disabilities live, learn, work, and play affect opportunities for them to become self-determined. This research has shown that adults with disabilities who live or work in more congregate or restrictive settings have fewer opportunities to make choices and decisions, to self-advocate, and to set goals (Duvdevany *et al.*, 2002; Emerson *et al.*, 2001; Tossebro, 1995; Stancliffe *et al.*, 2000; Wehmeyer and Bolding, 1999).

Intra-Individual Factors Contributing to Self-Determination

There are some data pertaining to the contribution of individual characteristics, including intelligence level,

age, gender, and adaptive behavior, on self-determination. Stancliffe *et al.* (2000) found that self-determination competencies (skills) and levels of adaptive or challenging behaviors were significant predictors of the expression of self-determination, with the former predicting higher self-determination and the latter (challenging behavior) lower self-determination. Like other factors discussed in this section, the relationship between self-determination and adaptive or challenging behaviors seemed to be mediated by the degree to which these factors related to variables such as where one lives or works and/or the choice opportunities one has. For example, in a UK sample, Perry and Felce (2005) found that objective measures of choice were positively correlated with adaptive-behavior scores. In addition, data pertaining to differences in self-determination by gender are found to be limited and mixed. Soresi *et al.* (2004) found, in an Italian sample, that adolescent males tended to have higher self-determination scores than females. Alternatively, Wehmeyer (1996) and Wehmeyer and Garner (2003) examined the self-determination of large samples of adolescents with cognitive disabilities, and found no significant differences between males and females on overall self-determination scores, although females scored slightly higher than their male counterparts. The relationship between gender and self-determination is, almost certainly, a function of societal norms and gender roles and expectations and not a function of any true intra-individual factor. As to the impact of age on self-determination, Wehmeyer and Garner found that age did not predict the membership of adults with developmental disabilities in a high or low self-determination group. In a sample of students aged 15–18 years, though, Wehmeyer (1996) found a consistent trend for higher self-determination scores as a function of increased age. In general, as children age and have increased opportunities to assume responsibility for their lives, they will have increased opportunity to become more self-determined. Although people continue to become more self-determined as they grow older, at some point that steady growth in self-determination levels off.

Research has consistently documented a statistically significant relationship between self-determination and intelligence level, as measured by intelligence quotient (IQ) scores (Wehmeyer *et al.*, 2003), with that relationship between measured self-determination and IQ scores ranging between $r = 0.15$ and 0.20 . It stands to reason that if one has greater capacity to solve complex problems, make decisions, and figure out the means to achieve goals, one will have greater capacity to act in a self-determined manner. What is surprising is that the relationship between IQ and self-determination, while statistically significant in large sample studies, remains practically insignificant. Once again, though, the relationship between self-determination and intelligence is complex. Wehmeyer and Garner (2003) conducted a series of

analyses to determine predictors of self-determination, autonomy, work setting (competitive vs. congregate), and living arrangement (independent vs. congregate) for individuals with intellectual disability. Only choice opportunity (from among four variables, including IQ score) predicted self-determination, while everything except IQ scores predicted autonomy level, with higher perceptions of choice opportunity being the most powerful predictor. Level of IQ was, however, the most significant contributor to employment and living outcomes, with IQ the only significant predictor for the former (work) and self-determination and autonomy also significant predictors of more positive living outcomes. In other words, IQ level predicted where people would live or work, but was not a predictor of whether they would be self-determined. Only choice opportunities predicted the latter, a finding consistent with the previously reviewed literature on the relationship between choice and environment.

Finally, self-determination has been linked by research to a more positive individual quality of life. Wehmeyer and Schwartz (1998) found that self-determination status predicted membership in a high quality of life group for adults with developmental disabilities. Lachapelle *et al.* (2005) examined the relationship between self-determination and quality of life for a larger sample of adults with disabilities in Canada, United States, Belgium, and France, and found that self-determination predicted membership in the high self-determination group, and overall, self-determination and quality of life were significantly correlated ($r = 0.49$). Similarly, Neely-Barnes *et al.* (2008) found that community-based living arrangements and greater opportunities to make choices were associated in a sample of adults with developmental disabilities, and these were, in turn, associated with greater quality of life.

Importance of Self-Determination to Students with Special Educational Needs

Self-determination is important to students with special needs because instruction to promote self-determination is a means to greater access to the general education curriculum and because students who are more self-determined achieve more positive adult outcomes. These points are discussed in the following section.

Improving Access to General Education

An important benefit of promoting self-determination is that it provides a means to ensure that students receiving special educational services in general education classroom settings can become involved with the general education curriculum. Wehmeyer *et al.* (2004) identified two ways that this occurs. First, by identifying where in the general education curriculum all students are expected to

learn skills and knowledge related to the component elements of self-determined behavior, teachers can promote self-determination and promote access to the general education curriculum. Second, teaching students with disabilities the skills they need to become more self-determined – problem solving, goal setting, decision making, and self-directed learning skills – provides students with valuable skills to enhance their academic performance. Students who are more self-determined – who learn to effectively set learning goals and objectives to reach those goals and then use problem-solving and self-regulation skills to tackle the activities to achieve those goals – will ultimately be able to perform, academically, more successfully.

There are now multiple studies confirming that student-directed learning strategies can be taught to enable students with special educational needs to self-direct learning in the general education classroom (Agran *et al.*, 2003; Wehmeyer *et al.*, 2007). Studies have also tested the impact of interventions to promote self-determination on student progress on goals derived from the general education curriculum. For example, Palmer *et al.* (2004) examined the attainment of goals linked to science, social studies, or language arts standards for middle school students with intellectual disability. Students were taught to self-direct learning to address a goal derived from the state standard in each content area that emphasized a self-determination focus. Students were able to achieve educationally relevant goals tied to district-level standards at expected or greater than expected levels, thus supporting the suggestion that instruction in self-determination can serve as an entry point to the general curriculum for students with disabilities.

Increasing Positive Adult Outcomes

The impact of student self-determination on more positive adult outcomes, including employment, independent living, and community inclusion, has been a principal factor in the growing emphasis within special needs education on promoting self-determination. There are a number of studies that provide evidence of the relationship between self-determination and more positive student outcomes. Wehmeyer and Schwartz (1997) and Wehmeyer and Palmer (2003) measured the self-determination of students with cognitive disabilities and then examined adult outcomes both at 1 year and 3 years after high school. Students in the high self-determination group were twice as likely as youth in the low self-determination group to be employed 1 year out of school and earned, on average, more than students in the low self-determination group who were employed. One year after high school, students in the high self-determination group were disproportionately likely to have moved from where they were living during high school, and by the third year they were still disproportionately likely to live somewhere other than their high school

home and were significantly more likely to live independently. For employed students, by the third year, those scoring higher in self-determination made statistically significant advances in obtaining job benefits, including vacation, sick leave, and health insurance, an outcome not shared by their peers in the low self-determination group. Earlier, Sowers and Powers (1995) showed that students with disabilities involved in instruction to promote self-determination increased their participation and independence in performing community activities. In summary, there is an expanding base of evidence suggesting that higher self-determination and increased capacity in the component elements of self-determined behavior results in better educational and adult outcomes for youth and young adults with special educational needs.

Strategies for Promoting Self-Determination

As noted previously, Algozzine *et al.* (2001) revealed that there are a number of empirically validated methods and strategies that enable students with special educational needs to acquire skills and knowledge pertaining to component elements of self-determined behavior. These behaviors help these students to self-regulate learning and improve overall self-determination. Following are sections that discuss these strategies.

Goal Setting

Having the skills to set and attain goals is central to one's self-determination. The process of promoting goal-setting skills involves working with students to help them learn to: (1) identify and define a goal clearly and concretely, (2) develop a series of objectives or tasks to achieve the goal, and (3) specify the actions necessary to achieve the desired outcome. At each step, students must make choices and decisions about what goals they wish to pursue and what actions they wish to take to achieve their goals. Goal-setting activities can be easily incorporated into a variety of educational activities and instructional domains, as well as in educational planning, as discussed subsequently. Research has suggested some strategies to make goals both meaningful and attainable for students with disabilities. First, goals should be challenging. If they are too easy, there is no motivation to do the work to attain them, nor is there a feeling of accomplishment after achieving them. Second, while it is preferable for students to set their own goals, if this is not possible and goals need to be set by teachers, then the students' preferences and interests should be incorporated into the goal to increase the students' motivation to pursue the goal. In the end, goals with personal meaning are more likely to be attained.

Choice Making

Choice making is simply the expression of a preference between two or more options. Opportunities to make choices should be infused throughout a student's day, as experiences with making choices teach students that they can exert control over their environment. Students can be provided opportunities to choose within or between instructional activities, choose with whom they engage in a task, where they engage in an activity, and if they complete an activity at all. Some students may need to learn how to make choices, particularly if the students' previous opportunities to do so have been restricted. Further, choices can be made more meaningful for students by involving them in decisions about what, how, and why they learn. Research has found that when students with disabilities are provided opportunities to make choices, reductions in problem behaviors and increases in adaptive behaviors are observed (Shogren *et al.*, 2004).

Problem Solving

A problem is an activity or task for which a solution is not known or readily apparent. The process of solving a problem involves: (1) identifying and defining the problem, (2) listing possible solutions, (3) identifying the impact of each solution, (4) making a judgment about a preferred solution, and (5) evaluating the efficacy of the judgment. Research not only shows that students with disabilities generate fewer potential solutions to problems and have greater difficulty solving more complex problems (Gumpel *et al.*, 2000), but also that students benefit from instructional efforts to promote this outcome (Crites and Dunn, 2004). Developing skills associated with social problem solving may be particularly important for students with disabilities. Research suggests that students with autism, for example, may have difficulty understanding social and emotional cues, which limit their ability to interact with others, but, again, interventions have been shown to be effective in improving problem-solving skills for this population (Bauminger, 2002).

Decision Making

Decision making involves coming to a judgment about which of a number of potential options is best at a given time. Making effective decisions involves: (1) identifying alternative courses of action, (2) identifying possible consequences of each action, (3) assessing the probability of each consequence occurring, (4) choosing the best alternative, and (5) implementing the decision. Working to promote systematic decision-making skills is best addressed at the secondary level, while at the elementary level a focus on choice making and problem solving can support the development of effective decision-making

skills later in life. When teaching decision-making skills, opportunities to make decisions should be embedded in the curriculum. By supporting students to make decisions in real-world situations, they will better develop their ability to conceptualize and generalize the decision-making process. The process of evaluating alternatives is an area in which direct instruction should occur; students can be provided support to (1) develop lists of decision options, (2) evaluate the risk and benefit associated with a given alternative, and (3) evaluate biases in their decision making. Students often evaluate risk somewhat differently than adults, perhaps because they see the excitement of risk as positive, rather than negative. However, by teaching students how to evaluate and conceptualize risk, both in terms of short-term and long-term consequences, these biases can be reduced.

Self-Regulation

Self-regulation involves student-directed learning. It is the process of setting goals, developing action plans to achieve those goals, implementing and following action plans, evaluating outcomes of action plans, and changing actions plans, if the goal was not achieved (Mithaug, 1993). Since self-regulation involves the use of student-directed learning strategies, students can be taught to self-regulate, self-direct, or self-manage their behaviors through external operants, such as visual or audio/verbal prompts or self-recording devices, or through internal operants, such as self-instruction, self-monitoring, self-evaluation, or self-reinforcement (Agran *et al.*, 2003). When students are taught and provided opportunities to use self-directed strategies, it gives them more control over initiating tasks, moving through task sequences, evaluating their success on the task and then reinforcing themselves for task completion. These strategies can be applied to academic, vocational, community participation, and functional curriculum goals and objectives.

One of the first steps in teaching a student a new skill or behavior is to have a clear plan of the steps and sequence of the task, as well as the cues and stimuli that occur naturally within the task or learning situation. Built within the context of a task analysis, this plan should map out the “instructional cues or directions, setting, teaching and task materials, opportunities for choice making and prompts and procedures for fading prompts, and consequences” (Snell and Brown, 2000: 151). Such strategies are called antecedent cue-regulation strategies. These are frequently combined with self-instruction strategies, in which the goal is to move toward self-instruction where the student is taught to verbalize cues that direct or maintain his or her own behavior (Agran *et al.*, 2003). Another strategy, self-monitoring, involves teaching a student to track his or her progress toward a goal, followed by a self-evaluation of that progress. Typically, a student uses

a checklist of steps or set of criteria that have been provided in advance to observe and monitor progress. For example, Sands and Doll (2005) recommended the use of self-monitoring as a support to goal setting and attainment and described several methods where students can track their progress through the use of rubrics, graphs or charts, a series of audio prompts, and checklists. These same tools can be used to teach students to self-record progress. When a desired behavior or skill involves a series of steps, each step may serve as a natural cue, which then serves as a catalyst for the student to take the subsequent step. Finally, self-evaluation is a summative process whereby students judge the progress toward their goal and the quality of their work. If they determine they have achieved their goal, they can use self-reinforcement strategies. With self-reinforcement, they are involved in identifying and administering their own consequences for having successfully completed a target skill or behavior. In other words, a student is taught how to reward herself for a job well done or to recognize errors that need to be addressed and corrected.

Self-Advocacy

Students with disabilities need to learn to advocate on their own behalf. To be an effective self-advocate, students have to learn both how to advocate and what to advocate for. There are ample opportunities for students to practice and learn self-advocacy skills within the context of the educational planning process. Too often, students’ perspectives have been lost, because students have not had opportunities or skills to express their perspectives within educational planning and decision-making meetings. A first step to enabling them to express their wants and needs during these meetings is educating them about their rights and responsibilities in these areas. Instructional strategies have been developed for teaching such knowledge to students with disabilities (Wehmeyer and Field, 2007). When teaching students how to advocate for themselves, the focus should be on teaching them how to be assertive, how to effectively communicate their perspective (either verbally or in written or pictorial form), how to negotiate, how to compromise, and how to deal with systems and bureaucracies. Students need to be provided real-world opportunities to practice these skills. This can be done by embedding opportunities for self-advocacy within the school day, by allowing students to set up a class schedule, work out their supports with a resource room teacher or other support provider, or participate in individual education plan (IEP) and transition meetings.

Self-Awareness and Self-Knowledge

The final component elements of self-determined behavior focus on the attitudes, beliefs, and perceptions that

enable people to act in a self-determined manner. Students need to develop perceptions of efficacy, along with self-awareness and self-knowledge, to give them the motivation and confidence to practice the skills discussed above. Research has shown that students with developmental disabilities tend to have less-adaptive perceptions of efficacy and outcome expectations than do students without disabilities (Wehmeyer *et al.*, 2003). For students to become more self-determined, they must possess a reasonably accurate understanding of their strengths, abilities, unique learning and support needs, and limitations. Further, they must know how to utilize this understanding to maximize success and progress.

Instructional Models, Curricular Materials, and Assessments to Promote Self-Determination

The Self-Determined Learning Model of Instruction (SDLMI) (Wehmeyer *et al.*, 2003) is an instructional model that has been shown to be effective in promoting self-determination and student-directed learning. The SDLMI is a model of teaching based on the component elements of self-determination, the process of self-regulated problem solving, and research on student-directed learning. Implementation of the model consists of a three-phase instructional process. Each instructional phase presents a problem to be solved by the student. The student solves each problem by posing and answering a series of four student questions per phase that students learn, modify to make their own, and apply to reach self-selected goals. Each question is linked to a set of teacher objectives. Each instructional phase includes a list of educational supports that teachers can use to enable students to self-direct learning. In each instructional phase, the student is the primary agent for choices, decisions, and actions, even when eventual actions are teacher directed. The student questions in the model are constructed to direct the student through a problem-solving sequence in each instructional phase. To answer questions in this sequence, students must regulate their own problem solving by (1) setting goals to meet needs, (2) constructing plans to meet goals, and (3) adjusting actions to complete plans. As noted, each instructional phase poses a problem the student must solve (What is my goal? What is my plan? What have I learned?), by solving a series of problems posed by questions in each phase. The four questions differ in each phase, but represent identical steps in the problem-solving sequence: (1) identify the problem, (2) identify potential solutions to the problem, (3) identify barriers to solving the problem, and (4) identify consequences of each solution. The solutions to the problems in each phase lead to the problem-solving sequence in the next phase.

The teacher objectives within the model are objectives that teachers intend to accomplish by implementing the model. In each instructional phase, the objectives are linked directly to the student questions. These objectives can be met by utilizing strategies provided in the educational supports section of the model. The emphasis in the model on the use of educational supports that are student directed provides another means of teaching students to teach themselves. As important as this is, however, not every instructional strategy implemented will be student directed. There are circumstances in which the most effective instructional method or strategy to achieve a particular educational outcome will be a teacher-directed strategy. Students who are considering what plan of action to implement to achieve a self-selected goal can recognize that teachers have expertise in instructional strategies and take full advantage of that expertise.

As indicated, the aforementioned strategies are to be infused into instruction across the curriculum. There are, however, several curricular materials and assessments available that provide direction for teachers seeking to promote student self-determination. A comprehensive examination of these curricular materials is beyond the scope of this article. Test *et al.* (2000), however, conducted an extensive review of existing curricular materials to promote self-determination and located 60 such products. Reviews and information about most of these procedures are available online and, as such, readers are referred to that resource for greater detail on the array of materials available. Further, a number of assessments exist to enable teachers to assess student needs with regard to instruction in self-determination. Most of these are available online through the Zarrow Center at the University of Oklahoma.

Another instructional focus that has emerged as important to promoting self-determination entails encouraging student involvement in educational planning and decision making, as discussed previously. Involvement in education planning, decision making, and instruction can take many forms, from students generating their own educational goals and objectives, to tracking their progress on self-selected goals or objectives, to running their own educational planning meeting. It is important to emphasize that it is not what the student does in the meeting that is critical, but, instead, the degree to which the student is an equal partner in and, to the greatest extent possible, in control of his or her planning. Students with severe disabilities can be involved in their educational program every bit as much as students with less-severe disabilities. Student involvement may look very different in these cases, and students with more severe disabilities may not be able to make independent decisions or solve problems, but that is not the criteria by which we should judge student involvement. It is, instead, the degree to which the student is actively engaged and empowered in his or her planning and education program. Clearly, there are

multiple advantages to student involvement. Test *et al.* (2004) reviewed studies examining efforts to promote student involvement, and concluded that there was clear evidence that such efforts enhanced student involvement in educational planning. In addition, student involvement in the educational process is a good way to teach and allow students to practice skills important to self-determination (i.e., goal setting, decision making, and/or negotiation), self-advocacy, leadership, and teamwork. There are several programs designed to promote student involvement, and information about them can be found in the section titled 'Relevant websites'.

Conclusion

Promoting self-determination has become best practice in the education of students with disabilities. Those students who leave school as self-determined young people achieve more positive adult outcomes. Moreover, promoting student self-determination provides, as it were, a gateway to the general education for students with disabilities, and can result in enhanced learning skills. This article overviewed the methods, materials, and strategies to achieve the outcome that enables students with and without disabilities to become more self-determined. If educators are to achieve the outcomes most envisioned for their students, they will need to ensure that students with disabilities are provided sufficient opportunities to learn these skills and strategies, and to use them to play a meaningful role in their educational programs, from planning to implementation.

See also: Risk and Resilience Frameworks in Understanding Special Education; Social Competence; Transition from School to Adult Life.

Bibliography

- Agran, M., King-Sears, M., Wehmeyer, M. L., and Copeland, S. R. (2003). *Teachers' Guides to Inclusive Practices: Student-Directed Learning Strategies*. Baltimore, MD: Paul H. Brookes.
- Algozzine, B., Browder, D., Karvonen, M., Test, D. W., and Wood, W. M. (2001). Effects of interventions to promote self-determination for individuals with disabilities. *Review of Educational Research* **71**, 219–277.
- Bauminger, N. (2002). The facilitation of social-emotional understanding and social interaction in high-functioning children with autism: Intervention outcomes. *Journal of Autism and Developmental Disorders* **32**, 283–298.
- Crites, S. A. and Dunn, C. (2004). Teaching social problem solving to individuals with mental retardation. *Education and Training in Developmental Disabilities* **39**, 301–309.
- Duvdevany, I., Ben-Zur, H., and Ambar, A. (2002). Self-determination and mental retardation: Is there an association with living arrangement and lifestyle satisfaction? *Mental Retardation* **40**, 379–389.
- Emerson, E., Robertson, J., Gregory, N., *et al.* (2001). Quality and costs of supported living residences and group homes in the United Kingdom. *American Journal on Mental Retardation* **106**, 401–415.
- Gumpel, T. P., Tappe, P., and Araki, C. (2000). Comparison of social problem-solving abilities among adults with and without developmental disabilities. *Education and Training in Mental Retardation and Developmental Disabilities* **35**, 259–268.
- Lachapelle, Y., Wehmeyer, M. L., Haelewyck, M. C., *et al.* (2005). The relationship between quality of life and self-determination: An international study. *Journal of Intellectual Disability Research* **49**, 740–744.
- Mithaug, D. E. (1993). *Self-Regulation Theory: How Optimal Adjustment Maximizes Gain*. Westport, CT: Praeger/Greenwood.
- Neely-Barnes, S., Marcenko, M., and Weber, L. (2008). Does choice influence quality of life for people with mild intellectual disabilities. *Intellectual and Developmental Disabilities*, **46**(1), 12–16.
- Nota, L., Ferrari, L., Soresi, S., and Wehmeyer, M. L. (2007). Self-determination, social abilities, and the quality of life of people with intellectual disabilities. *Journal of Intellectual Disability Research* **51**, 850–865.
- Palmer, S. B., Wehmeyer, M. L., Gibson, K., and Agran, M. (2004). Promoting access to the general curriculum by teaching self-determination skills. *Exceptional Children* **70**, 427–439.
- Perry, J. and Felce, D. (2005). Correlation between subjective and objective measures of outcome in staffed community housing. *Journal of Intellectual Disability Research* **49**, 278–287.
- Sands, D. J. and Doll, B. (2005). Teaching goal setting and decision making to students with developmental disabilities. In Wehmeyer, M. and Agran, M. (eds.) *Mental Retardation and Intellectual Disabilities: Teaching Students with Innovative and Research-Based Strategies*, pp 320–345. Merrill/Prentice Hall: Columbus, OH.
- Shogren, K., Faggella-Luby, M., Bae, S. J., and Wehmeyer, M. L. (2004). The effect of choice-making as an intervention for problem behavior: A meta-analysis. *Journal of Positive Behavior Interventions* **6**, 228–237.
- Snell, M. and Brown, F. (2000). Development and implementation of educational programs. In Snell, M. and Brown, F. (eds.) *Instruction of Students with Severe Disabilities*, 5th edn., pp 115–172. Upper Saddle River, NJ: Prentice-Hall.
- Soresi, S., Nota, L., and Ferrari, L. (2004). Autodeterminazione e scelte scolastico-professionali: Uno strumento per l'assessment (Self-determination and school-career choices: An instrument for the assessment). *Giornale Italiano di Psicologia dell'Orientamento* **5**, 26–42.
- Sowers, J. and Powers, L. (1995). Enhancing the participation and independence of students with severe physical and multiple disabilities in performing community activities. *Mental Retardation* **33**, 209–220.
- Standliffe, R., Abery, B., and Smith, J. (2000). Personal control and the ecology of community living settings: Beyond living-unit size and type. *Mental Retardation* **105**, 431–454.
- Test, D. W., Karvonen, M., Wood, W. M., Browder, D., and Algozzine, B. (2000). Choosing a self-determination curriculum: Plan for the future. *Teaching Exceptional Children* **33**(2), 48–54.
- Test, D. W., Mason, C., Hughes, C., *et al.* (2004). Student involvement in individualized education program meetings. *Exceptional Children* **70**, 391–412.
- Tossebro, J. (1995). Impact of size revisited: Relation of number of residents to self-determination and privatization. *American Journal on Mental Retardation* **100**, 59–67.
- Wehmeyer, M. L. (1996). A self-report measure of self-determination for adolescents with cognitive disabilities. *Education and Training in Mental Retardation and Developmental Disabilities* **31**, 282–293.
- Wehmeyer, M. L. (2005). Self-determination and individuals with severe disabilities: Reexamining meanings and misinterpretations. *Research and Practice in Severe Disabilities* **30**, 113–120.
- Wehmeyer, M. L., Abery, B., Mithaug, D. E., and Standliffe, R. J. (2003). *Theory in Self-Determination: Foundations for Educational Practice*. Springfield, IL: Charles C Thomas.
- Wehmeyer, M. L., Agran, M., Hughes, C., *et al.* (2007). *Promoting Self-Determination in Students with Intellectual and Developmental Disabilities*. New York: Guilford.

- Wehmeyer, M. L. and Bolding, N. (1999). Self-determination across living and working environments: A matched-samples study of adults with mental retardation. *Mental Retardation* **37**, 353–363.
- Wehmeyer, M. L. and Field, S. (2007). *Self-Determination: Instructional and Assessment Strategies*. Thousand Oaks, CA: Corwin Press.
- Wehmeyer, M. L., Field, S., Doren, B., Jones, B., and Mason, C. (2004). Self-determination and student involvement in standards-based reform. *Exceptional Children* **70**, 413–425.
- Wehmeyer, M. L. and Garner, N. W. (2003). The impact of personal characteristics of people with intellectual and developmental disability on self-determination and autonomous functioning. *Journal of Applied Research in Intellectual Disabilities* **16**, 255–265.
- Wehmeyer, M. L. and Palmer, S. B. (2003). Adult outcomes for students with cognitive disabilities three years after high school: The impact of self-determination. *Education and Training in Developmental Disabilities* **38**, 131–144.
- Wehmeyer, M. L. and Schwartz, M. (1997). Self-determination and positive adult outcomes: A follow-up study of youth with mental retardation or learning disabilities. *Exceptional Children* **63**, 245–255.
- Wehmeyer, M. L. and Schwartz, M. (1998). The relationship between self-determination, quality of life, and life satisfaction for adults with mental retardation. *Education and Training in Mental Retardation and Developmental Disabilities* **33**, 3–12.
- self-directed IEP as an evidence-based practice. *Exceptional Children* **72**, 299–316.
- Martin, J. E., Van Dycke, J. L., Greene, B. A., et al. (2006). Direct observation of teacher-directed IEP meetings: Establishing the need for student IEP meeting instruction. *Exceptional Children* **72**, 187–200.
- Mithaug, D. E., Mithaug, D., Agran, M., Martin, J., and Wehmeyer, M. L. (eds.) (2003). *Self-Determined Learning Theory: Construction, Verification, and Evaluation*. Mahwah, NJ: Erlbaum.
- Mithaug, D. E., Mithaug, D., Agran, M., Martin, J., and Wehmeyer, M. L. (2007). *Self-Instruction Pedagogy: How to Teach Self-Determined Learning*. Springfield, IL: Charles C Thomas.
- Test, D. W., Karvonen, M., Wood, W. M., Browder, D., and Algozzine, B. (2000). Choosing a self-determination curriculum: Plan for the future. *Teaching Exceptional Children* **33**(2), 48–54.
- Wehmeyer, M. L. (2005). Self-determination and individuals with severe disabilities: Reexamining meanings and misinterpretations. *Research and Practice in Severe Disabilities* **30**, 113–120.
- Wehmeyer, M. L., Abery, B., Mithaug, D. E., and Standcliffe, R. J. (2003). *Theory in Self-Determination: Foundations for Educational Practice*. Springfield, IL: Charles C Thomas.
- Wehmeyer, M. L., Agran, M., Hughes, C., et al. (2007). *Promoting Self-Determination in Students with Intellectual and Developmental Disabilities*. New York: Guilford Press.
- Wehmeyer, M. L. and Field, S. (2007). *Self-Determination: Instructional and Assessment Strategies*. Thousand Oaks, CA: Corwin Press.
- Wehmeyer, M. L., Field, S., Doren, B., Jones, B., and Mason, C. (2004). Self-determination and student involvement in standards-based reform. *Exceptional Children* **70**, 413–425.

Further Reading

- Agran, M., King-Sears, M., Wehmeyer, M. L., and Copeland, S. R. (2003). *Teachers' Guides to Inclusive Practices: Student-Directed Learning Strategies*. Baltimore, MD: Paul H. Brookes.
- Karvonen, M., Test, D. W., Wood, W. M., Browder, D., and Algozzine, B. (2004). Putting self-determination into practice. *Exceptional Children* **71**, 23–41.
- Martin, J. E., Van Dycke, J. L., Christensen, W. R., et al. (2006). Increasing student participation in IEP meetings: Establishing the

Relevant Websites

- <http://www.beachcenter.org> – Beach Center on Disability.
- <http://www.uncc.edu> – UNC Charlotte.
- <http://www.education.ou.edu> – University of Oklahoma, Zarrow Center.

Social Competence

L A Owens, University of Wisconsin – Milwaukee, Milwaukee, WI, USA

S Johnston-Rodriguez, Northern Illinois University, DeKalb, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Social competence is a topic of increasing importance to many professionals in the field of education and rehabilitation. Social competence refers to the social, emotional, and cognitive skills and behaviors that young adults need for successful social adjustment in school and beyond. Over the years, research has suggested that an individual's social and emotional adjustment, academic and cognitive development, and sense of citizenship are enhanced by opportunities to strengthen social competence during his/her youth (Hartup and Moore, 1990; Kinsey, 2000; Ladd and Profilet, 1996; McClellan and Kinsey, 1999; Parker and Asher, 1987; Rogoff, 1990). The ability to successfully engage socially with peers and significant others (i.e., parents, teachers, and counselors) is an essential part of a student's development (Gresham *et al.*, 2001). Further, school success and academic success have been related to levels of social competence. Simply stated, social competence plays a critical role in a student's ability to flourish and navigate the complexities of the school, family, and work (Brown and Heath, 1998). This is the focus of this article.

For many students, particularly those with high-incidence disabilities, the area of social development presents ongoing challenges (Gresham *et al.*, 1997). It has been suggested that children with mild disabilities experience lower levels of social skills, adaptive behavior, and peer connections. As noted by Elliott and Gresham (1991), young persons lacking social skills have fewer interactions with their typical peers resulting in negative relationships. Elksnin and Elksnin (1998) asserted that students who are unsuccessful in learning suitable social skills have a lower probability of finishing their education, obtaining employment, and attaining stable adulthood. Social problems have also been cited as key factors in job loss for persons with disabilities (Hagner *et al.*, 1992).

While the issue of social competence is indeed a crucial one, there is no single factor that can sufficiently describe a person's social competence (Brown and Heath, 1998). Social competence involves the ability to apply social aptitudes and behaviors across a variety of settings and involves social skills, adaptive behavior, motivation, and interpretation of the social situation (Elksnin and Elksnin, 1998). The socially competent student possesses a range of skills that enables him/her to interpret his/her social setting, react to social cues and contexts, and achieve positive outcomes in situations. Elksnin and Elksnin noted social competence entails more than social skills; the interpretation of the social context and recognition of the expectations

and norms of the situation are integral and critical components of social competence.

Social Competence versus Social Skills

In the past, the distinctions between the constructs of social skills and social competence have been blurred, as they have been used interchangeably. While both social skills and social competence are interrelated, they are not indistinguishable (McFall, 1982). The term social competence, in contrast to social skills, refers to a broad construct that includes both adaptive behaviors (social skills) and an evaluative component that judges those specific behaviors in a given context (Gresham, 1982). Evaluation of social competence, as recommended by Gresham (1986), should involve:

1. pertinent judgment of a student's social actions by significant persons (e.g., peers, teachers, and family);
2. assessment of social behaviors according to a defined set of conditions; and
3. social actions compared to specific norms (e.g., social skills rating scale).

While social competence is connected to the concept of social skills, the distinctions are significant. Social skills refer to specific behaviors utilized within a specific social activity or interaction (e.g., greeting another person and taking turns). Social skills are behaviors that can be acquired through instruction and then performed in completing a task. In contrast, social competency centers on the evaluation of social outcomes, and based on a judgment of how well the individual has performed the social behavior in a specific situation (McFall, 1982). Thus, the success of social performance depends on not only the behavioral aspect (what the person did) and the environmental aspect (when, where, and with whom did it occur) but includes the evaluation by others. Thus, the success of social interaction results not only from knowledge of specific social skills but also from the understanding of the norms, roles, demands, and expectations of a given context. Social competence encompasses a person's ability to (1) possess specific social skills in his/her repertoire, (2) evaluate a given situation as to the social criterion and expectations, and (3) know when and how to use those social skills within a given context. In sum, social competence reflects the ability to use social skills based on a person's perceptions and understanding of the social environment (Smith *et al.*, 2006).

Perspectives on Social Competence

Historically, numerous conceptualizations of social competence have been described throughout the literature. Greenspan's (1981) characterization of personal competence asserts that the construct of personal competence is characterized by three domains:

1. academic competence,
2. social competence, and
3. physical competence.

For children with mild disabilities, issues related to academic and social abilities have the most salience, while students with more significant disabilities (e.g., severe and profound mental retardation) or health-related impairments (e.g., cerebral palsy and traumatic brain injury) experience more challenges in physical competence (Gresham *et al.*, 1997). Earlier, McFall (1982) noted that social competence, while closely related to social skills, has distinguishable differences. Social skills are discrete behaviors employed by an individual in completing social tasks, whereas social competence derives from the evaluation of an individual's performance based on a social agent's judgments. These judgments may be based on opinions of significant others (e.g., teacher, parents, and peers), as to whether the social task was performed adequately in comparison to explicit criteria (e.g., number of social interactions initiated based on an expected criterion). McFall's view of social competence considers social skills to be specific behaviors which result in judgments of social competence. Thus, social skills are behaviors, and social competence represents a judgment about those behaviors.

Social competence has been viewed from the perspective of social validity. A person can be seen as socially competent if exhibiting socially acceptable skills that achieve the desired, valid social outcomes in a given context (Gresham, 1982). Within the framework of social validity, social competency can be characterized as socially significant behaviors exhibited in specific situations which predict important social outcomes for children and youth (Gresham, 1983; Gresham and Elliott, 1990). Socially significant behaviors signify behaviors that significant persons in the situation (e.g., parents, teachers, and peers) consider valuable and desirable and which predict socially important outcomes – these outcomes represent outcomes that others consider valuable, adaptive, and useful. They are also outcomes which improve a person's ability to function in social settings. Outcomes considered socially important may include peer acceptance (McConnell and Odom, 1986; Parker and Asher, 1987) and teacher and parental acceptance (Gresham and Elliott, 1990). Gresham and Reschly (1988) conceptualized social competence as a multidimensional construct that includes adaptive behavior, social skills, and peer-relationship variables (e.g., peer acceptance, friendship, and peer rejection). This conceptualization is similar to

Gresham's earlier (Gresham, 1983) social validity definition of social skills. According to this definition, social skills are those behaviors which, within certain situations, predict important social outcomes for children and youth. In school settings, important social outcomes might include:

1. peer acceptance,
2. significant others' positive judgments of social competence,
3. academic achievement,
4. adequate self-concept,
5. positive attitudes toward school, and
6. freedom from loneliness.

This definition has the advantage of specifying behaviors in which children and youth may be deficient and relating these deficiencies to socially important outcomes in school settings. Vaughn and Hogan (1990) proffered an additional model of social competence. While similar to previous interpretations, they consider the role of higher-order thinking within the concept of social competence. Within their model, social competence is a complexity of four factors: (1) pro-social interactions with peers, (2) correct social awareness/perceptions; (3) lack of maladaptive behavior, and (4) successful social skills. Under this model, socially competent individuals would experience peer status, friendships, and positive family interactions. Individuals would demonstrate age-appropriate levels of interpersonal problem solving, self-monitoring, and self appraisal of their social capacity. Social competence results from the interaction and correlation between the four components, thus no one element by itself can accurately describe a student's social competency (see Vaughn and Hogan, 1990).

The Role of Social Competence in Disability Classification

As stated earlier, making distinctions between social competence and social skills is vital as the two constructs play a pivotal role in defining specific disabilities, such as serious emotional disabilities (SEDs) or learning disabilities. Increasingly, the classification of students with high-incidence disabilities has relied on the construct of social competence as a decisive factor in the determination of disability. Current criteria used in classifying mental retardation (cognitive disability) consider the significance of both cognitive and social competence as having similar influence (Gresham *et al.*, 1995). Deficits in social competence are viewed by some as key determinants in classifying students with emotional disorders (Forness and Knitzer, 1992) as well as students with specific learning disabilities (ICLD, 1987) and attention-deficit hyperactivity disorder (ADHD; Barkley, 1990). According to criteria listed in the Individuals with Disabilities Act (IDEIA), students with SED lack the ability to develop or sustain acceptable social connections with

their peers or adults and in typical situations experience uncharacteristic emotions. Recently, the position among many professionals, government groups (e.g., Interagency Committee on Learning Disabilities), and professional organizations (e.g., NJCLD), is that deficits in social competence and social skills constitute a specific learning disability (Gresham *et al.*, 1997).

Students with mild disabilities have frequently been recommended for evaluation for special education services due to the teacher's perceptions that they do not meet acceptable expectations or standards for social behavior (Gresham *et al.*, 1997). Teachers typically expect and tolerate a level of behavior that conforms to their notions of teachability (Hersh and Walker, 1983). A teachability profile signifies those patterns of social behavior that include absence of disruptive behavior, compliance, as well as academic related (e.g., attention to task and listening). In addition, schools may view such students as resistant to intervention in the course of determining their eligibility for special education (Gresham, 1991). Since the goal of any intervention is to create a positive change in a target behavior, resistance to intervention would indicate that a given intervention has failed to produce such change.

Socially Competent Behavior

In reviewing the literature on social-behavioral competence in school settings, students with disabilities are characterized as facing challenges as they attempt to navigate social behavior expectations in two domains: adult/teacher relation and student/peer relation. Regarding adult/teacher interactions, students with disabilities must understand and meet the social standards of the school and classroom environment. In terms of peer relationships, students with disabilities must adjust their social behavior so that it meets the standards for peer acceptance and results in friendship (Walker *et al.*, 2004). Social competence includes pro-social behaviors that promote the student's ability to interact and develop positive relationships with others. Pro-social behaviors include flexibility, empathy, caring, communication, and humor (Bernard, 1991). Earlier, Hazel *et al.* (1981) assert that prosocial competencies demand that a person possess prosocial skills and in a given situation, understand what behavior is required.

There is a difference between being popular and having friendships. According to Welsch and Bierman (1998), while some individuals are outgoing and have many friends, others may be shy and are happy with one or two good friends. They suggest that three characteristics be considered when assessing social competence:

1. the nature of the individual's social behavior (e.g., are they aggressive, bossy, disruptive, or impulsive);
2. the way in which students are perceived by peers (e.g., are they disliked, teased, ostracized, or ignored); and

3. the stability and duration of peer problems (e.g., are the problems with peer relationships short term).

Assessment of Social Competence

The value of any assessment process lies in its validity. In turn, valid assessment should generate accurate decisions about a student's behavior, and in turn produce beneficial results (Walker *et al.*, 2004). A principle purpose of assessment is to assist in developing appropriate intervention strategies that will improve social behavior. According to Gresham (1982), assessment of social competence examines three primary factors: (1) assessment of person's social actions based on views of significant others, (2) appraisal of social behavior according to set standards, and (3) actual performance of social behaviors. Underlying every valid assessment process are key assumptions that serve as the basis for the process. Depending upon the context of the situation, expectations, and a student's acculturation, students will frequently differ from their peers in terms of their behavior. In addition, assessments represent samples of a student's behavior and no singular assessment can provide adequate information about the student's status (Witt *et al.*, 1998). Decisions should be based on a collection of information from multiple sources about behavior patterns, and include environmental information. As Walker *et al.* (2004) noted, assessment information should be "multimethod, multi-agent, and multisetting in nature" (p. 85).

Different assessment methods have been used to identify and classify social skill deficits. Gresham and Lambros (1998) described assessment approaches as either direct or indirect. Direct assessment techniques evaluate behavior in real time in the actual, natural setting. Structured observation of a student's behavior in his/her natural surroundings at school or home would be a typical type of direct assessment. Indirect approaches examine behavior out of context and away from the actual time of occurrence. These approaches include the use of behavioral rating scales or interviews by teachers or parents. Social skills assessment should be approached in a systematic, thoughtful manner. While different sources have suggested a range of stages and goals for social skills assessment, three major goals emerge as a guide for the process:

1. problem identification – this examines strengths, competing problems, deficit acquisition, performance, and fluency;
2. problem analysis includes functional behavioral analysis, targeting behaviors, and determining social validity; and
3. treatment evaluation involves assessment of outcome measures and generalization and maintenance of intervention outcomes (Gresham, 2001; Walker *et al.*, 2004).

Techniques for evaluating social behaviors may involve a range of commonly used approaches that vary according

to the purpose and desired outcomes of the assessment. Major categories of assessment procedures include (1) behavior rating scales, (2) social skills rating scales, and (3) functional behavioral assessment (FBA). Behavior rating scales are one common method frequently used by parents and teachers to evaluate a student's behavior. They are considered by some as more cost effective, reliable, and objective when compared with other forms of assessment (Gresham and Elliott, 1990). Social skills rating scales (SSRSs) examine specific social skills behavior that measure the rater's perceptions of the student's skills across various social domains. While a number of rating scales are available, some are considered more user friendly and based on standardized samples (Walker *et al.*, 2004). Two frequently used tools include the SSRS (Gresham and Elliott, 1990); and the Walker–McConnell Scales of Social Competence and School Adjustment (SSCSA) (Walker and McConnell, 1995). The central purpose of an FBA is to identify the function or cause of a student's target behavior and to subsequently teach him/her an acceptable replacement behavior that will meet the student's needs. The primary goal is to determine the specific communicative intent of a behavior for a student through observation, interviews, and review of relevant information (O'Neill *et al.*, 1997). FBA emphasizes examination of the environmental factors that may elicit unacceptable behavior (target behaviors) and may reveal what outcomes student receives as a result of exhibiting those behaviors. FBA would then provide support by teaching positive alternatives that would achieve the same results.

While a variety of assessment methods are utilized to evaluate social competency in terms of social skills and social literacy, Strichter *et al.* (2007), in their review of recommended practices, noted that an array of formal and informal tools should be used.

Employing multiple measures of social functioning can provide the types of information needed to make both data-based decisions as well as provide a complete description of intervention efficacy (Bryan, 1997).

Interventions to Increase Social Competence

Before interventions can be identified, it is important to conduct a social autopsy where an adult assists an individual to improve social skills by jointly analyzing social errors he/she makes and designs alternative strategies (Lavoie, 2005). Once the skill of social competence has been identified, strategies for remediation can be identified. These strategies for increasing social competence differ depending upon the age of the individual. One strategy involves social skill training (SST) to teach individuals verbal and nonverbal ways of interacting in social

situations. This training can be done either with a small group of students or using video modeling. Clearly, it helps the individual to learn to interpret the subtle cues involved in social interaction. Finding a good social niche for the individual is critical in SST. Often large, unstructured settings (e.g., recess, study hall, and lunch) are particularly difficult situations for students who have difficulty with social skills. More structured, smaller peer-interaction setting where adults can support and guide positive peer interaction generally work best.

SST is considered an effective approach to increasing social competence for students who are socially at risk (Bullis *et al.*, 2001). SST, if implemented in a structured, systematic, and ongoing manner, can assist students who demonstrate social skills acquisition and/or performance deficits. It is important to remember that not all children demonstrate social competence or incompetence in all settings. This is why it is essential to complete an accurate social skills assessment that will identify specific social problems. SST may be conducted at a universal level (all students in a group, classroom, or school) or at an individual level. Universal SST strategies may include class-wide social skills programs or school-wide positive discipline approaches where all students are taught the same skills in the same manner. Such methods strive to promote positive behaviors, minimize unacceptable behaviors from developing, and create positive school climates.

Conclusion

It is important to note that social skills vary as do the strategies depending upon the age of the student. Developmental changes occur in the structure and quality of peer interactions which affect the complexity of skills contributing to social competence. The ways in which children spend their time changes with age. As children grow, their preferences for play change, as well as the thinking and language skills that provide a foundation for social competence also change (Welsch and Bierman, 1998). As teachers and service providers, we need to understand each child and what he/she needs to interact socially and competently in our changing society.

Bibliography

- Brown, A. and Heath, N. (1998). Social competence in peer-accepted children with learning disabilities. *Paper Presented at the National Association of School Psychologists*. Orlando: FL.
- Bryan, T. (1997). Assessing the personal and social status of students with learning disabilities. *Learning Disabilities Research and Practice* 12, 63–76.
- Bullis, M., Walker, H. M., and Sprague, J. R. (2001). A promise unfulfilled: Social skills training with at-risk children and youth. *Exceptionality* 9, 67–90.

- Elksnin, L. K. and Elksnin, N. (1998). Teaching social skills to students with learning and behavioral problems. *Interventions in School and Clinic* **33**, 131–140.
- Elliott, S. N. and Gresham, F. M. (1991). *Social Skills Intervention Guide: Practical Strategies for Social Skills Training*. Circle Pines, MN: American Guidance Service.
- Forness, S. and Knitzer, J. (1992). A new proposed definition and terminology to replace “serious emotional disturbance” in individuals with disabilities education act. *School Psychology Review* **21**, 12–20.
- Greenspan, S. (1981). Social competence and handicapped individuals. Practical implications and a proposed model. *Advances in Special Education* **3**, 41–82.
- Gresham, F. M. (1982). Misguided mainstreaming: The case for social skills training with handicapped children. *Exceptional Children* **48**, 422–433.
- Gresham, F. M. (1983). Social validity in the assessment of children’s social skills. Establishing standards for social competency. *Journal of Psychoeducational Assessment* **1**, 297–307.
- Gresham, F. M. (1986). Conceptual issues in the assessment of social competence in children. In Strain, P., Gualnick, M., and Walker, H. (eds.) *Children’s Social Behavior: Development, Assessment, and Modifications*, pp 143–179. New York: Academic Press.
- Gresham, F. M. (1991). Conceptualizing behavior disorders in terms of resistance to intervention. *School Psychology Review* **21**, 23–36.
- Gresham, F. M. (2001). Assessment of social skills in children and adolescents. In Andrews, J., Safloske, D., and Janzen, H. (eds.) *Handbook of Psychoeducational Assessment*, pp 325–355. Orlando, FL: Academic Press.
- Gresham, F. M. and Elliott, S. N. (1990). *Social Skills Rating System*. Circle Pines, MN: American Guidance Service.
- Gresham, F. M. and Lambros, K. (1998). Behavioral and functional assessment. In Watson, T. S. and Gresham, F. M. (eds.) *Handbook of Child Behavior Therapy*, pp 3–22. New York: Plenum.
- Gresham, F. M. and Reschly, D. J. (1988). Issues in the conceptualization, classification, and assessment of social skills in the mildly handicapped. In Kratochwill, T. (ed.) *Advances in School Psychology*, pp 203–247. Hillsdale, NJ: Erlbaum.
- Gresham, F. M., Sugai, G., and Horner, R. H. (2001). Interpreting outcomes of social skills training for students with high-incidence disabilities. *Exceptional Children* **67**, 331–344.
- Gresham, F. M., MacMillan, D. L., and Siperstein, G. (1995). Critical analysis of the 1992 AAMR definition: Implications for school psychology. *School Psychology Quarterly* **10**, 1–19.
- Hagner, D., Rogan, P., and Murphy, S. T. (1992). Facilitating natural supports in the workplace: Strategies for support consultants. *Journal of Rehabilitation* **58**, 29–34.
- Hartup, W. W. and Moore, S. G. (1990). Early peer relations: Developmental significance and prognostic implications. *Early Childhood Research Quarterly* **5**(1), 1–18.
- Hazel, J. S., Schumaker, J. B., Sherman, J. A., and Sheldon-Wildgen, J. (1981). *ASSET: A Social Skills Program for Adolescents (Video and Leader’s Manual)*. Champaign, IL: Research Press.
- Hersh, R. and Walker, H. (1983). Great expectations: Making schools effective for all students. *Policy Studies Review* **2**, 147–188.
- ICLD (Interagency Committee on Learning Disabilities) (1987). *Learning Disabilities: A Report to the U.S. Congress*. Bethesda, MD: National Institute of Health.
- Kinsey, S. J. (2000). *The Relationship between Prosocial Behaviors and Academic Achievement in the Primary Multiage Classroom*. Unpublished doctoral dissertation. Loyola University, Chicago.
- Ladd, G. W. and Profilet, S. M. (1996). The child behavior scale: A teacher-report measure of young children’s aggressive, withdrawn, and prosocial behaviors. *Developmental Psychology* **32**(6), 1008–1024.
- LaVoie, R. (2005). *Social competence and the child with learning disabilities*. <http://www.idonline.org/article/6169> (accessed October 2009).
- McClellan, D. E. and Kinsey, S. (1999). Children’s social behavior in relation to participation in mixed-age or same-age classrooms. *Early Childhood Research & Practice* [online] **1**(1). <http://ecrp.uiuc.edu/v1n1/v1n1.html> (accessed October 2009).
- McConnell, S. and Odom, S. (1986). Sociometrics: Peer-referenced measures and the assessment of social competence. In Strain, P., Gualnick, M., and Walker, H. (eds.) *Children’s Social Behavior: Development, Assessment, and Modifications*, pp 143–179. New York: Academic Press.
- McFall, R. (1982). A review and reformulation of the concept of social skills. *Behavioral Assessment* **4**, 1–35.
- O’Neill, R. E., Horner, R. H., Albin, R. W., Storey, K., and Newton, S. (1997). *Functional Analysis of Problem Behavior: A Practical Assessment Guide*. Pacific Grove, CA: Brooks/Cole.
- Parker, J. G. and Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin* **102**(3), 357–389.
- Rogoff, B. M. (1990). *Apprenticeship in Thinking: Cognitive Development in Social Context*. New York: Oxford University Press.
- Smith, T. E. C., Polloway, E. A., Patton, J. R., and Dowdy, C. A. (2006). *Teaching Students with Special Needs in Inclusive Settings*, 4th edn. New York: Pearson.
- Strichter, J. P., Randolph, J., Gage, N., and Schmidt, C. (2007). A review of recommended social competency programs for students with autism spectrum disorders. *Exceptionality* **15**, 219–232.
- Vaughn, S. and Hogan, A. (1990). The social status of children with learning disabilities: A reexamination. *Journal of Learning Disabilities* **23**(1), 32–37.
- Walker, H. and McConnell, S. (1995). *Walker-McConnell Scale of Social Competence and School Adjustment: Elementary Version*. Florence, KY: Thomson Learning.
- Walker, H. M., Ramsey, E., and Gresham, F. M. (2004). *Antisocial Behavior in School: Evidence-Based Practices*, 2nd edn. Belmont, CA: Thomson Learning.
- Welsch, J. A. and Bierman, K. L. (1998). Social competence. In Gayle *Encyclopedia of Childhood and Adolescence*. http://findarticles.com/p/articles/mi_g2602/is_0004/ai_2602000487 (accessed October 2009).
- Witt, J. C., Elliott, S. N., Daly, E., Gresham, F. M., and Kramer, J. (1998). *Assessment of At-Risk and Special Needs Children*, 2nd edn. Boston: McGraw-Hill.

Further Reading

Individuals with Disabilities Education Act of 2004 (2004). 20 U.S.C. § 1400 et seq.

The Identification of Students with Gifts and Talents

V I Daniels, Southern University and A&M College, Baton Rouge, LA, USA

M J McCollin, Slippery Rock University-PA, Slippery Rock, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Societal interest in individuals with high aptitudes, gifts, talents, and extraordinary abilities dates back thousands of years – perhaps, as early as 3000 BC. In Western cultures, interest in superior abilities of individuals was stimulated by the works of Charles Darwin and Sir Francis Galton in the 1800s. Lewis M. Terman – a noted psychologist and influential pioneer in the field of gifted education – conducted a comprehensive research study of gifted individuals, in the 1920s, that drew attention to the identification, education, and nurturing of students with gifts and talents. Leta Stetter Hollingworth – another influential pioneer – taught the first course for gifted students and published the first textbook in the field of gifted education.

Concepts such as gifted and talented have been closely linked with intelligence. This thinking went into high gear, during the early 1900s, when French Psychologists Alfred Binet and Theodore Simon were commissioned to construct a test that provides a means for schools to determine the educability of students on the basis of intellectual ability. While the terms gifted and talented are, often, used interchangeably throughout the literature, there are some conceptual differences between these two. Gagné (1985; 1991) differentiated the concepts by defining giftedness as above-average competence in natural, untrained, spontaneous, and human ability (aptitude, intellectual, or creative abilities) and talent as above-average competence in an area or field of human activity (e.g., mathematics or music) which manifests itself when the individual engages in systematic learning, training, and practicing.

It was the 1957 launch of the space capsule Sputnik by the Soviet Union that stimulated momentum in the gifted-education movement in the United States. This historical event was viewed as a risk to national security. It not only provoked the United States to rethink the quality of American schooling, but also led to the passage of the National Defense Education Act in 1958, the appropriation of federal funding for the establishment of programs for gifted students, the development of ways to identify students with superior abilities and high academic achievements (particularly in math and science), and research to identify effective methods for providing quality educational experiences to gifted students.

During the late 1960s, emphasis on the education of gifted students shifted and took a back seat to civil rights and antipoverty movement concerns on educational equity and the improvement of schools for at-risk students. In 1970, the United States Congress commissioned

a study to examine the needs of gifted students – which resulted in the highly acclaimed 1972 Marland Report. This report provided impetus for the first direct assistance to gifted education and the establishment of the US Office of the Gifted and Talented. National interests in the education of students with gifts and talents led to the establishment of the World Council for Gifted Children, which was formed in 1975. In 1981, the US Department of Education created a National Commission on Excellence in Education to examine the quality of education in American Schools. In its 1983 report, national concerns for the problems afflicting American education were once again renewed. This groundbreaking study recommended the development of a set of standards for the identification and provision of enriched and accelerated programs for students who are gifted and talented. Consequently, in 1988, Congress passed the Jacob K. Javits Gifted and Talented Student Education Act, which provided financial support and resources for the Office of Gifted and Talented Education – a national research center focusing on gifted children, and demonstration projects in gifted education. In addition, in 1990, a National Center on Gifted and Talented was established. This center – comprised of a consortium of universities – addresses research needs in the field.

In 1993, the second national report on the status of educating gifted and talented students was issued by the US Department of Education, Office of Educational Research and Improvement. This report described the state of gifted education as being a quite crisis and identified a number of indicators that pointed to the need of America to change the way it educates gifted and talented youth and those with the potential for high abilities. In addition, the report found that gifted education still suffered from (1) a lack of highly qualified teachers, (2) questionable methods for identifying gifted students, (3) weak curriculum standards, (4) limited learning opportunities for disadvantaged and minority children, and (5) a failure of the educational system to challenge gifted and talented youngsters to work to their full potential. In 1994, the Jacob K. Javits Act was reauthorized, but the funding appropriated was significantly reduced as a result of Congressional party changes and differing political and philosophical views of its constituents. In 2001, the Javits Act was again reauthorized. This reauthorization was contained in the No Child Left Behind Act of 2001 (NCLB, P. L. 107–110) – formerly known as the Elementary and Secondary Education Act

(ESEA) – and signed into law by President George Bush on 8 January 2002. NCLB is a historic law that was enacted to eliminate the achievement gap that exists among students in US schools and ensure that children of all racial, ethnic, cultural, and socioeconomic backgrounds, and children with exceptional learning needs (e.g., children with disabilities and those with gifts and talents) have a fair, equal, and significant opportunity to obtain a high-quality education, and be taught by highly qualified teachers.

Definitions of Gifted and Talented

Over the years, many theorists, psychologists, researchers, and practitioners not only grappled with the concept of giftedness but also proposed numerous definitions of giftedness, yet, none has been universally accepted as a standard definitional criterion to describe the extraordinary gifts, remarkable talents, outstandingly high level of abilities and potentialities of children and youth. Problems associated with the definition of giftedness stem from multiple sources, including (1) differing theoretical views on giftedness; (2) the wide range of characteristics associated with children having extraordinary skills, abilities, aptitudes, and talents; (3) the varying definitions of giftedness (e.g., psychometric definitions, trait definitions, educationally oriented definitions); (4) the variance across cultures as to what is considered giftedness; (5) the various terminologies used; and (6) ongoing research in the field (Daniels, 2003).

The first federal definition of giftedness was introduced in the Marland Report. This report, commissioned by Congress in 1972, identified six categories of giftedness, encouraged schools to define giftedness broadly (i.e., to extend beyond intellectual ability), and provided the catalyst for the development of a US Office of Gifted and Talented. The report also served as the foundation upon which gifted educational programs are built. According to the Marland Report

Gifted and talented children are defined as those identified by professionally qualified persons who by virtue of outstanding abilities, are capable of high performance. These are children who require differentiated educational programs and/or services beyond those normally provided by the regular school program in order to realize their contribution to self and society. (p. 2)

The report added

Children capable of high performance are those who demonstrate any of the following abilities or aptitudes, singly or in combination in any of the following areas: (1) general intellectual ability, (2) specific academic aptitude, (3) creative or productive thinking, (4) leadership ability, (5) visual and performing arts aptitude, and (6) psychomotor ability. (p. 2)

The most current federal definition of giftedness can be found in the NCLB Act. Under this law,

The term “gifted and talented”, when used with respect to students, children, or youth, means students, children, or youth who give evidence of high achievement capability in areas such as intellectual, creative, artistic, or leadership capacity, or in specific academic fields, and who need services or activities not ordinarily provided by the school in order to fully develop those capabilities (Title IX, Part A, Section 9101(22), p. 1959).

While federal definitions of giftedness have been criticized over the years (e.g., not providing school districts with the specificity necessary to guide professionals in the identification of students with gifts and talents; emphasizing performance as the defining characteristic of giftedness; not placing an emphasis on observable behaviors), many school districts still base their definition of giftedness on federal definitions (Daniels, 2003). Perhaps, the most influential contemporary definition of gifted is that proposed by the National Center for Gifted Children (NCGC) – an organization composed of parents, teachers, educators, other professionals, and community leaders. As the NCGC pointed out

A gifted person is someone who shows, or has the potential for showing, an exceptional level of performance in one or more areas of expression. . . . Some of these abilities are very general and can affect a broad spectrum of the person's life, such as leadership skills or the ability to think creatively. Some are very specific talents and are only evident in particular circumstances, such as a special aptitude in mathematics, science, or music. The term *giftedness* provides a general reference to this spectrum of abilities without being specific or dependent on a single measure or index

The NCGC further stated

A person's giftedness should not be confused with the means by which giftedness is observed or assessed. Parent, teacher, or student recommendations, a high mark on an examination, or a high IQ score are not giftedness; they may be a signal that giftedness exists. Some of these indices of giftedness are more sensitive than others to differences in the person's environment.

Thus, definitions of giftedness can reflect a broad array of purposes. They may be used to (1) identify and count children to establish prevalence estimates; (2) diagnose, label, and place students in specially designed accelerated or enrichment programs; (3) enable professionals, parents, and others to communicate effectively with regard to students who have high abilities or talents; (4) serve as a medium to establish priorities for funding educational or support programs; and (5) facilitate legislative efforts on behalf of gifted and talented individuals (Daniels, 2003).

Prevalence, Characteristics, and Etiology of Giftedness

Determination of prevalence estimates for children with gifts and talents can be quite challenging. The complexity is due, in part, to variations among state definitions of giftedness in terms of the children to whom they apply, state laws and local policies governing the identification of students for gifted and talented programs, and the availability of funding for such programs. Problems with determining prevalence are, further, exacerbated by the forms of giftedness and talents being considered, the traits of giftedness being assessed, the minimum threshold for intellectual quotient (IQ), and the political, educational, and societal landscape of the times. Although prevalence estimates for children with gifts and talents are highly variable, the NCGTC estimated that approximately 5% of the student population (or 3 million children) in the United States are gifted.

The study of giftedness encompasses both children and adults and it involves the investigation of both cognitive and affective attributes. Much of the research on giftedness (e.g., Clark, 2008; Daniels, 2003; Friend, 2005; Glass, 2004; Hallahan and Kauffman, 2003; Smith, 2004) describes gifted and talented children as manifesting a wide range of abilities – including accelerated cognitive abilities, a phenomenal memory, and excellent recall. These children also have the ability to quickly acquire, retain, conceptualize, synthesize, learn new information, and easily process and manipulate large amounts of information. In addition, children with gifts and talents are very curious and highly inquisitive. Not only do they enjoy learning, but they also have a strong desire to learn and an intense need for mental stimulation. They have a strong need to pursue their own interests, a capacity for seeing unusual and diverse relationships, and an unusual ability to think abstractly. Clearly, they possess superior problem-solving skills, can easily apply and transfer knowledge to new situations, and show an interest in a wide range of topics and disciplines. They are able to expand their understanding of concepts in a number of ways, including their ability to connect with other ideas (e.g., see the relationship between concepts or through their understanding of the perspectives of others). They are also persistent and goal directed. They are capable of learning more rapidly than their peers and are knowledgeable concerning things at a much earlier age. They learn to speak early and can talk in sentences using complex sentence structures before the age of 2 or 3.

Apparently, children with gifts and talents have highly receptive vocabularies and come to school reading significantly beyond their chronological age-mates. Many possess advance verbal abilities, extensive vocabularies, and vivid imaginations. Because of their unique abilities, they often require accelerated instruction and sophisticated instructional materials. In addition, they are highly motivated and

extremely independent. Many are recognized by their teachers for their superior achievement in one or more academic domains (e.g., reading, math, science). These children like to take risks and have a high tolerance for ambiguity. They have a strong dislike for drill-and-practice activities and are easily bored and impatient with routine tasks. In addition, they find conventional instructional practices less motivating, less challenging, and less rewarding. Similar to students with disabilities, there exist both inter- and intraindividual differences in their performance. Despite their high levels of intelligence, some possess high levels of creativity – particularly in the visual and performing arts – while others demonstrate outstanding leadership and psychomotor abilities.

Children who are gifted and talented display unique social and emotional characteristics (Clark, 2008; Daniels, 2003; Friend, 2005). They hold high expectations of themselves, possess a keen (sometimes subtle) sense of humor, and have a strong sense of self-awareness. These children strive for perfection, are highly focused, and experience enormous stress from failure. They possess high morals, have a strong sense of justice, and are sensitive to the feelings and needs of others. Additionally, they display high energy levels in their work and play which can, sometimes, be misinterpreted and misdiagnosed by psychologists and other health professionals as a form of hyperactivity or an emotional or behavioral disorder. The most common misconception with regard to gifted children is that they lack social and emotional competence (Clark, 2008). While there are some gifted and talented children who feel different, have low self-esteem, or are misunderstood or socially isolated, most are well-adjusted, sensitive, healthy, emotionally independent, and stable individuals who are well liked by their peers (Clark, 2008) and who possess emotional traits that are well balanced.

It should be noted that the characteristics presented, in this article, are only examples of attributes of children with gifts and talents and should not be considered as exhaustive. Not every child identified as gifted or talented will display all of these traits. Given what is known about human development and the role of the environment, it is not surprising that giftedness is considered to result from genetic and environmental influences (Daniels, 2003; Friend, 2005). While it is, generally, believed that genetic makeup plays a more important role than the environment in the development of gifts and talents, the effects of environmental influences (e.g., family, school, home, community) on the development of one's potential should not be underestimated.

Identifying Gifts and Talents

Students with gifts and talents can be found across all racial, ethnic, cultural, linguistic, socioeconomic, and

even disability groups (Clark, 2008; Daniels, 2002). What remains a salient and prominent concern is how to accurately identify these students, particularly those from underrepresented populations (i.e., at risk students, racial and ethnic minority students, students from economically depressed or low-socioeconomic backgrounds, students with limited English proficiency; Callahan, 2005; Daniels, 2002; Ford, 1998; Ford *et al.*, 2008). Although there are procedures required by federal law for identifying students with disabilities, such procedures are nonexistent for students who are gifted and talented (Friend, 2005).

Despite the fact that culturally, linguistically, and ethnically diverse students make up a substantial percentage of the school-age population in US schools, they still constitute a small proportion of the students in gifted and talented programs. Data by the National Academy of Sciences (Donovan and Cross, 2002) on the representation of minority students in gifted and special education report that, while there is considerable variation among states in how they identify and classify students for giftedness, students that comprise some of the racial/ethnic groups still continue to be less likely to be in gifted programs compared to their white counterparts. Possible causes of racial/ethnic disparities among students in gifted programs include (1) lack of incentives or opportunities, (2) lack of teacher recruitment and referrals, (3) poor IQ-test performance by students from cultural and linguistically diverse backgrounds, and (4) deficit thinking orientations (Ford *et al.*, 2008). In addition, disparities occur because of (1) inadequate opportunities for talent development, (2) inadequate or one-shot paper-and-pencil assessments, and (3) policies and procedures that interfere with the funding of gifted programs and the identification of students from underserved populations (Callahan, 2005). To a larger degree, disparities occur because of (1) inadequate screening procedures and identification tools, (2) the lack of authentic and multidimensional assessments, and (3) the lack of a universally accepted definition of giftedness (Daniels, 2002). They also occur because of (1) differences in states criteria for giftedness, (2) limited language development, (3) learning-style differences, and (4) teacher quality (Donovan and Cross, 2002). Gender is another area in which disparities occur in gifted and talented programs. Girls – compared to boys – are less likely to be recognized or identified as gifted. While there has been a gradual increase in the number of girls identified as gifted in recent years, they are still underrepresented in gifted and talented programs. Factors that contribute to the under identification of girls in gifted and talented programs can be attributed to a number of variables, including societal and cultural stereotyping of the roles of girls, biases related to sex roles, and teacher biases about the areas of academic achievement for boys (e.g., math, science, technology) versus that of girls (e.g., English, reading, writing, the arts; Friend, 2005; Smith, 2004).

Because of wide-ranging attributes of giftedness and biases inherent in the identification process, the assessment of children for gifts and talents should be conducted by a group of professionals representing different areas of expertise, and be comprised of multiple criteria, multiple assessments, multiple approaches, and multiple sources of information derived from both quantitative and qualitative sources. In addition, they should include carefully selected formal and informal measures that do not discriminate against students who differ from the mainstream (i.e., students from economically disadvantaged backgrounds, at-risk students, and students with disabilities and gifts). Procedures for identifying children for gifted and talented programs, generally, comprise five basic steps – screening, referral, assessment, classification, and placement. Among assessment tools commonly used to identify gifted children (or children with potential gifts and talents) are intelligence tests; achievement tests; divergent thinking and creativity tests; aptitude tests; nonverbal ability tests; personality inventories; scales on leadership; alternative or performance-based assessments; teacher, self, parent, peer, and community nominations; students' grades, portfolios, work samples, and products; in-school and outside-of-school observations; checklists and rating scales; and recommendations of professional artists (Clark, 2008; Daniels, 2003; Friend, 2005).

A number of factors can inhibit or impede the identification of gifted children and children with the potential for giftedness (e.g., Callahan, 2005; Clark, 2008; Daniels, 2003). Among school-related factors are boredom, poor teaching, poor educational experiences, limited learning opportunities, the teacher's inability to recognize gifted behaviors of students, inappropriate curricula, teacher quality, lack of use of culturally responsiveness practices in teaching, limited use of positive reinforcement, and lack of consideration of students' learning styles. Among the environmentally related factors are unfavorable environmental conditions, poverty, limited-to-no access to community resources, and substance abuse. Family-related factors that can adversely affect the identification of gifted children include poor healthcare, family disorganization, single-parent families, insufficient infant stimulation, lack of exposure to books and other educational experiences, abuse, neglect, and drugs.

Educational Considerations for Students with Gifts and Talents

Educational placement alternatives for students with gifts and talents – and those with the potential for giftedness – can range from very restrictive (e.g., special schools, self-contained classrooms, home schooling), partially restricted (e.g., resource rooms and pull-out programs), to very inclusive settings (e.g., traditional schools and general education

classrooms). The types of placement alternatives for the delivery of educational services can vary from state to state as well as within individual school districts. Educational programs should stimulate the development of students' actual and potential abilities as well as their skills, creativity, and talents. The two approaches most widely used to deliver instruction to these students are enrichment and acceleration. Enriching the content of instruction (enrichment) incorporates many creative and innovative strategies. It includes additional topics and activities, expanding the study of a specific topic in more depth than is typically pursued in the traditional curriculum, using new curriculum strategies and techniques, having the student engage in independent study, providing the student with access to a mentor who shares similar interests, and participating in internship programs that allow for firsthand experiences in the profession or area of interest. Acceleration enables students to learn and progress at a rate commensurate with their abilities. This approach can take on a variety of forms. It includes early entrance to kindergarten and early entrance to college, grade-level advancement (i.e., grade-skipping), taking high school classes while in middle school, taking honors and advanced placement (AP) classes while in high school, taking advanced courses by correspondence or via the Internet, enrolling in extra courses for early graduation, using technology resources, and enrolling in intensive summer programs or summer institutes.

Ability grouping – another approach to teaching students with gifts and talents – has been used in schools for over a decade. This approach involves the placement of two or more students with comparable abilities (who can keep pace with each other) into small informal groups for subject-matter instruction. Curriculum compacting is another way in which teachers can meet the needs of these students. It involves streamlining lessons by eliminating the coverage of content, concepts, or topics the student already knows, has mastered, or is capable of mastering in a fraction of the time that is required by peers, and the reallocation of that time to other enrichment activities. It is important to keep in mind that all programs for students with gifts and talents, irrespective of grade level or how they are structured, must provide what Clark (2008) described as differentiated instruction to allow for (1) flexibility in teaching, (2) multiple venues for students to reach their learning potential, (3) continuous measures to assess progress, (4) activities that lead to intellectually stimulating peer interactions, (5) continuity, and (6) teachers with specialized training on gifted learners.

Conclusion

Societal interest in persons with gifts and talents – including those with extraordinary abilities or aptitudes – dates back thousands of years, and it is unlikely that

interest among this population will change anytime soon. Although children with gifts and talents exist across all racial, ethnic, cultural, linguistic, socioeconomic, and disability groups, what remains a salient concern is how to accurately identify them – particularly those from at-risk and underrepresented populations. Despite the fact that children with gifts and talents share many characteristics (e.g., cognitive, academic, social/emotional characteristics) commonly cited throughout the literature, much discussion still exists with regard to definitions of giftedness and the representation of minority students in gifted education programs. While there are a number of factors that underscore minority student representation, researchers are still grappling with factors that are known to influence their identification and visibility within the schools. Factors that appear most prevalent are test bias, gender bias, teacher quality, the lack of culturally responsive teaching practices, a lack of use of instructional practices directed toward students' learning styles, and limited learning opportunities or opportunities to demonstrate giftedness. Although educational placement alternatives and enrichment programs for children with gifts and talents vary widely, discussions about these types of programs tend to be relatively well received and less controversial in nature.

See also: Assessment Accommodations for Children with Special Needs; Cultural and Linguistic Diversity and Children with Special Needs; Disabilities and Gifted Education; History of Special Education; Large Scale Assessment and Accountability and Students with Special Needs; Overrepresentation of Culturally and Linguistically Diverse Students in Special Education; Peer Relations and Socialization of Children and Adolescents with Special Needs; Programs and Instructional Strategies for Students with Gifts and Talents; Social Competence.

Bibliography

- Callahan, C. M. (2005). Identifying gifted students from underrepresented populations. *Theory into Practice* 44(2), 98–104.
- Clark, B. (2008). *Growing up Gifted: Developing the Potential of Children at Home and at School*, 7th edn. Upper Saddle River, NJ: Prentice-Hall.
- Daniels, V. I. (2002). Maximizing the learning potential of African American learners with gifts and talents. In Obiakor, F. E. and Ford, B. A. (eds.) *Creating Successful Learning Environments for African American Learners with Exceptionalities*, pp 95–105. Thousand Oaks, CA: Corwin Press.
- Daniels, V. I. (2003). Students with gifts and talents. In Obiakor, F. E., Utley, C. A., and Rotatori, A. F. (eds.) *Effective Education for Learners with Exceptionalities*, vol. 15, pp 325–348. Oxford, UK: Elsevier Science.
- Donovan, M. S. and Cross, C. T. (eds.) (2002). *Minority Students in Special and Gifted Education*. Washington, DC: National Academy Press.
- Ford, D. Y. (1998). The underrepresentation of minority students in gifted education: Problems and promises in recruitment and retention. *Journal of Special Education* 32(1), 4–14.

- Ford, D. Y., Grantham, T. C., and Whiting, G. W. (2008). Culturally and linguistically diverse students in gifted education: Recruitment and retention issues. *Exceptional Children* **74**(3), 289–306.
- Friend, M. (2005). *Special Education: Contemporary Perspectives for School Professionals*. Boston, MA: Person Education.
- Gagné, F. (1985). Giftedness and talent: Reexamining a reexamination of the definitions. *Gifted Child Quarterly* **29**(3), 103–112.
- Gagné, F. (1991). Toward a differentiated model of giftedness and talent. In Colangelo, N. and Davis, G. A. (eds.) *Handbook of Gifted Education*, pp 65–80. Boston, MA: Allyn & Bacon.
- Glass, T. F. (2004). What Gift? The reality of the student who is gifted and talented in public school classrooms. *Gifted Child Today* **27**(4), 25–29.
- Hallahan, D. P. and Kauffman, J. M. (2003). *Exceptional Learners: Introduction to Special Education*. Upper Saddle River, NJ: Merrill.
- Smith, D. D. (2004). *Introduction to Special Education: Teaching in an Age of Opportunity*, 5th edn. Boston, MA: Pearson Education.

Further Reading

- Bélanger, J. and Gagné, F. (2006). Estimating the size of the gifted/ talented population from multiple identification criteria. *Journal for the Education of the Gifted* **30**(2), 131–163.
- Davis, G. A. and Rimm, S. B. (2004). *Education of the Gifted and Talented*, 5th edn. Upper Saddle River, NJ: Prentice-Hall.
- Elhoweris, H., Mutua, K., Alsheikh, N., and Holloway, P. (2005). Effect of children's ethnicity on teacher's referral and recommendation decisions in gifted and talented programs. *Remedial and Special Education* **26**(1), 25–31.
- Frasier, M. M., Garcia, J. H., and Passow, A. H. (1995). *A Review of Assessment Issues in Gifted Education and their Implications for Identifying Gifted Minority Students*. Research Monograph 95204. (ERIC Document Reproduction Service No. ED 388 024). Storrs, CT: National Research Center on the Gifted and Talented.
- Frasier, M. M., Hunsaker, S. L., Lee, J., et al. (1995). *Core Attributes of Giftedness: A Foundation for Recognizing the Gifted Potential of Minority and Economically Disadvantaged Students*. (ERIC Document Reproduction Service No. ED 402 703). Storrs, CT: National Research Center on the Gifted and Talented.
- Gardner, H. (1993). *Multiple Intelligences: The Theory in Practice*. New York: Basic Books.
- Gargiulo, R. M. (2006). *Special Education in Temporary Society: An Introduction to Exceptionality*, 2nd edn. Belmont, CA: Thomson Wadsworth.
- Grantham, T. C. (2002). Underrepresentation in gifted education: How did we get here and what needs to change? *Roeper Review* **24**, 50–51.
- Harmon, D. (2002). They won't teach me: The voices of gifted African American inner-city students. *Roeper Review* **24**(2), 68–75.
- Karnes, F. A. and Stephens, K. R. (2008). *Achieving Excellence: Educating the Gifted and Talented*. Upper Saddle River, NJ: Prentice Hall.
- Manning, S. (2006). Recognizing gifted students: A practical guide for teachers. *Kappa Delta Pi Record* **42**(2), 64–68.
- Marland, S. P. (1972). *Education of the Gifted and Talented. Report to the Congress of the United States by the U.S. Commissioner of Education*. Washington, DC: U.S. Government Printing Office.
- National Association for Gifted Children. *What is giftedness?* <http://www.nagc.org/index.aspx?id=574&an> (accessed October 2009).
- National Association for Gifted Children and Council of State Directors of Programs for the Gifted (2007). *State of the States in Gifted Education 2006–2007*. Washington, DC: National Association for Gifted Children and Council of State Directors of Programs for the Gifted.
- Oakland, T. and Rossen, E. (2005). A 21st-century model for identifying students for gifted and talented programs in light of national conditions: An emphasis on race and ethnicity. *Gifted Child Today* **28**(4), 56–63.
- Obiakor, F. E. (ed.) (2007). *Multicultural Special Education: Culturally Responsive Teaching*. Upper Saddle River, NJ: Pearson Education.
- Renzulli, J. S. and Park, S. (2000). Gifted dropouts: The who and the why. *Gifted Child Quarterly* **44**, 261–271.
- Seeley, K. (2004). Gifted and talented students at risk. *Focus on Exceptional Children* **37**(4), 1–8.
- Silverman, L. K. (ed.) (2000). *Counseling the Gifted and Talented*. Denver, CO: Love.
- Webb, J. T. (2000). *Mis-Diagnosis and Dual Diagnosis of Gifted Children: Gifted and LD, ADHD, OCD, Oppositional Defiant Disorder*. (ERIC Document Reproduction Service No. ED448 382).

Relevant Websites

- <http://www.apa.org> – Center for Gifted Education Policy.
- <http://www.ditd.org> – Davidson Institute for Talent Development.
- <http://www.ericdigests.org> – Eric Digests.
- <http://www.gt-cybersource.org> – GT-Cybersource.
- <http://www.uniquelygifted.org> – Internet Resources for Gifted/Special Needs Children.
- <http://cty.jhu.edu> – Johns Hopkins University Center for Talented Youth (CTY).
- <http://www.nagc.org> – National Association for Gifted Children.
- <http://www.nfgcc.org> – National Foundation for Gifted and Creative Children.
- <http://www.gifted.uconn.edu/NRCGT.html> – National Research Center on the Gifted and Talented (NRC/GT).
- <http://www.nsrgt.org> – National Society for the Gifted and Talented.
- <http://www.sedoparking.com> – National Talent Network (NTN) at the Educational Information & Resource Center's (EIRC).
- <http://www.sengifted.org> – Supporting the Emotional Needs of the Gifted.
- <http://world-gifted.org> – World Council for Gifted and Talented Children.

Transition from School to Adult Life

M J McCollin, Slippery Rock University of Pennsylvania, Slippery Rock, PA, USA

F E Obiakor, University of Wisconsin–Milwaukee, Milwaukee, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

The transition from birth to childhood, through adolescence, and then to adulthood is a turbulent and sometimes unconventional evolution all people experience with varying degrees of achievements and challenges. This transition involves several interrelated events, role changes, and behavioral adjustments linked with many person-centered life occurrences (e.g., beginning school, puberty, search for independence, search for autonomy, graduating from secondary school, employment, family, social relationships, and growing older) that occur throughout adult life. Transition to the complexities of adulthood can be challenging for most young adults.

Many students with disabilities have difficulty in successfully making the transition from adolescence to adulthood and then from school to work. Sometimes these students (1) do not graduate; (2) encounter great difficulty in finding and maintaining suited employment; (3) exhibit high levels of unemployment or underemployment; (4) are socioeconomically unstable; (5) have limited access to community resources; (6) are involved with the justice system; (7) do not receive postsecondary training; and (8) become dependent on their families or government programs (Hallahan and Kauffman, 2003; Halpern, 1994; Rusch and Chadsey, 1998; Sands and Wehmeyer, 1996; Wehman, 2001). Overall postschool outcomes for students with disabilities are substandard. Research has found that students with severe and multiple disabilities and those with emotional behavioral disorders typically have poorer outcomes compared to those with specific learning disabilities or speech and language impairments (Rampey *et al.*, 2009). Therefore, in preparing students with disabilities for continued education, adult responsibilities, independence, and employment, transition planning must also include strategies (e.g., self-determination and self-advocacy), for successfully negotiating the challenges, concerns, and uncertainties of adulthood. To ensure successful functioning in adulthood, these strategies should be acknowledged, considered, and addressed within the context of the transition plan so that students with disabilities are able to competently deal with the challenges of life (Patton and Dunn, 1998; Smith *et al.*, 2004; Wehmeyer *et al.*, 1998). This is the focus of this article.

Historical Overview of Federal Transition Services Legislation

In 1976, the Vocational Education Amendments (P.L. 94-482) established the need for collaboration among

vocational education and other job-training programs as well as increased funding for vocational-education programs and youth with disabilities. During the mid-1980s, the Carl D. Perkins Vocational and Technical Education Act (P.L. 98-524) further provided resources for young people with disabilities to successfully prepare for transition from secondary education to the world of adulthood. The act was passed to

assure that individuals who are inadequately served under vocational education programs are assured access to quality vocational educational programs, especially individuals who are disadvantaged, who are handicapped, men and women who are entering non-traditional occupations, adults who are in need of training and retraining, individuals who are single parents or homemakers, individuals with limited English proficiency, and individuals who are incarcerated in correctional institutions.

Since 1983, there has been a legislative prioritization regarding transition services for students with disabilities. P.L. 94-142 was amended with a new Section 626, entitled Secondary Education and Transition Services for Handicapped Youth to authorize federal grants to support the coordination of educational and adult transitional service programs for youth with disabilities. The 1990 amendments of IDEA, continued federal support for transition-related activities, refined transition services, and mandated that student interests, preferences, and needs be considered when developing transition plans. Moreover, IDEA required that the Individualized Education Program (IEP) for students ages 16 and older take into consideration specific transition components (i.e., transition services and inter-agency collaborations). The 1997 amendment of IDEA produced the most significant mandates related to special education legislation: (1) the expansion of the transition requirements regarding age at which services were to be introduced on the IEP, and more importantly (2) the significance of postschool outcomes for students with IEPs. The most recent reauthorized IDEA 2004 further defines transition services as a coordinated set of activities for a student with disability that:

- is designed to be within an outcome-oriented process, which promotes both academic and functional achievement in an effort to facilitate the movement from school to postschool activities, including postsecondary education, vocational education, integrated employment (including supported employment), continuing

and adult education, adult services, independent living, or community participation;

- is based on the individual student's needs, taking into account the student's capacity, access to support and resources, preferences, and interests; and
- includes instruction, related services, community experiences, the development of employment and other postschool adult living objectives, and when appropriate, acquisition of daily living skills and functional vocational evaluation. (H.R. 1350, Sec 602[34]).

As a result of the significant prioritization of transition services for students with IEPs, there has been considerable improvement of postschool outcomes and student skill development (Wehmeyer, 1998).

Transition Planning

The skill sets needed to successfully navigate adulthood in the twenty-first century is much more complex and comprehensive than those required by previous generations. Today's young adults must not only contend with challenges of employment, family life, leisure, and recreation, health and quality of life, but also with access to technology and community resources, interagency collaboration, and self-determination. Adulthood means being an active, contributing, and independent member of the community and greater society. Independence for an adult includes the ability to participate in society, work, have a home, autonomy, self-respect, raise a family, and share the joys and responsibilities of community life (Neubert, 2003). Adults with disabilities face numerous obstacles within the context of their daily living that affects not only their ability to access and use community resources but also limits the opportunities they have for social interaction and self-fulfillment (Heward, 2008). Therefore, transition planning is a critical aspect for successful navigation of adult life. Paramount to transition planning are the concepts of collaboration and interagency communication. These concepts are important and comparatively new to special-education service providers (Heward, 2008). Transition planning is conceptualized as the coordination, delivery, and transfer of services from the secondary school program to receiving agencies (e.g., employers, postsecondary education and vocational training programs, and residential service providers).

Effective transition planning must have a student focus, involve parents and families, and foster interagency and interdisciplinary collaboration. Within the context of student-focused planning, students have an opportunity to strengthen and develop self-determination and self-advocacy skills through practice and implementation. Paramount to student-focused planning is the development of a student's self-awareness which directly impacts

educational decisions. A self-aware person is able to take into consideration his/her own goals, vision, and interests when making educational decisions (Kohler and Field, 2003). An important strategy to help students identify their interests and preferences is to provide them with interdisciplinary curricular experiences so that they are able to (1) reflect on their experiences, (2) derive meaning particular to their context, and (3) use that information for future decision making (Kohler and Field, 2003).

Parental and Family Involvement

Parental or family involvement is a critical aspect of successful transition planning. It involves the development of participation and role, empowerment, and training of family members; hence allowing families the ability to actively participate in the decision-making processes for their family members is critical. Family involvement has been shown to (1) improve school attendance, (2) improve student's self-esteem and confidence, (3) reduce dropout rates, and (4) improve student autonomy – a critical aspect of self-determination (Sands and Wehmeyer, 1996). Student-focused planning is extended and supported through parent or family involvement.

Collaboration and Consultation

Collaboration and communication between and among professionals and families are key to effective transition planning. In special education, collaboration and consultation are important in the planning and delivery of transition services to secondary students with disabilities. Although interagency cooperation is critical to the success of transition, the amendments to IDEA made it clear that the initial and most significant transition responsibilities lie with schools (Heward, 2008). As it appears, every state has a history of federally maintained work-study and vocational training programs; however, the systematic coordination of and communication between schools and community-based adult services have not regularly occurred (Sands and Wehmeyer, 1996). As a result of the limited collaboration, many young adults in transition are not able to access the necessary community supports and resources.

Collaborative interagency and interdisciplinary collaboration are fostered by agreements that clearly articulate goals, roles, communication strategies, and other collaborative actions that enhance the curriculum and service delivery. Interagency collaboration includes activities between family, special educators, and adult service providers such as working as a team, sharing information, attending transition planning meetings, accessing community resources, and establishing and utilizing effective lines of communication to benefit students with disabilities as they transition from high school to the adult world.

The purpose of these collaborative activities is to implement an integrated system that addresses lifelong learning and supports the needs of a community member. Addressing barriers to effective collaboration such as ineffective use of transition meetings, intimidating language, and complex agency procedures may be an important goal of the interagency collaboration, thereby affording students with disabilities the opportunity to have successful entry into adult life (Oertle and Trach, 2007). To be successful in adult life, young people urgently need an interdisciplinary high school curriculum that has a strong focus on their transitional needs (Dowdy *et al.*, 1990).

Significance of Self-Determination and Self-Advocacy

Self-determination is a set of skills or competencies that assist students in (1) determining their wants and needs, (2) deciding on the best actions to take, (3) taking the action, (4) evaluating, and (5) improving on the outcome (Repetto, 2003). Self-determination is a highly critical skill necessary for successful navigation of adult life; without this skill, others will be responsible for making decisions in the lives of individuals with disabilities. Specific skills in self-determination include decision making, coping with failure, self-assessment, action planning, and option exploration (Carter *et al.*, 2008; Repetto, 2003; Wehmeyer and Palmer, 2003). Engaging in self-determination strategies during the formal transition planning process and throughout adult life requires students to complete tasks they may find challenging. These strategies may include the ability to actively participate in transition meetings, self-disclose strengths and weaknesses, and request services and accommodations (Repetto, 2003).

Conclusion

Often, students with disabilities experience even greater challenges when transitioning from school to the adult world. For the purposes of this article, the developmental imperative of transitioning from school to adulthood (e.g., choosing a career; moving out of one's parents' home; establishing social networks, transportation, and recreational activities independent of one's parents) is a challenge for most adolescents. For students with disabilities, transition may be considerably more difficult if adequate planning, support, and training are not provided. Successful participation in adulthood requires active long-range planning by students, their parents or families, school staff, and adult service providers. It is essential that students have a thorough understanding of the consequences and options for their transition plans. The attainment of positive postschool outcomes for adolescents and adults with disabilities is dependent upon professionals becoming more knowledgeable about

the changing demands of education. Creative solutions are essential to helping this population gain access to tools and information so that they can add to the world knowledge and economy, thereby increasing their own potential worth.

Bibliography

- Carter, E., Lane, K., Pierson, M., and Stang, K. (2008). Promoting self-determination for transition-age youth: Views of high school general and special educators. *Exceptional Children* **75**, 55–70.
- Dowdy, C., Carter, T., and Smith, T. (1990). Differences in transitional needs of high school students with and without learning disabilities. *Journal of Learning Disabilities* **23**, 343–348.
- Hallahan, D. and Kauffman, J. (2003). *Exceptional Learners: Introduction to Special Education*. Boston, CA: Allyn and Bacon.
- Halpern, A. (1994). The transition to youth with disabilities to adult life: A position statement of the Division on Career Development and Transition. *Career Development for Exceptional Individuals* **21**, 113–128.
- Heward, W. (2008). *Exceptional Children: An Introduction to Special Education*. Upper Saddle River, NJ: Prentice Hall.
- Kohler, P. and Field, S. (2003). Transition focused education: Foundation for the future. *Journal of Special Education* **37**, 175–183.
- Neubert, D. (2003). The role of assessment in the transition to adult life process for students with disabilities. *Exceptionality* **11**, 3–75.
- Oertle, K. and Trach, J. (2007). Interagency collaboration: The importance of rehabilitation professionals' involvement in transition. *Journal of Rehabilitation* **73**, 36–44.
- Patton, J. and Dunn, C. (1998). *Transition from School to Young Adulthood: Basic Concepts and Recommended Practices*. Austin, TX: Pro-Ed.
- Ramsey, B., Dion, G., and Donahue, P. (2009). *NAEP 2008 Trends in Academic Progress (NCES 2009–479)*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education.
- Repetto, J. (2003). Transition to living. *Exceptionality* **11**, 77–87.
- Rusch, F. and Chadsey, J. (1998). *Beyond High School: Transition from School to Work*. New York: Wadsworth.
- Sands, D. and Wehmeyer, M. (1996). *Self-Determination across the Life Span: Independence and Choice for People with Disabilities*. Baltimore, MD: Paul H. Brookes.
- Smith, T., Polloway, E., Patton, J., and Dowdy, C. (2006). *Teaching Children with Special Needs in Inclusive Setting*, 4th edn. Boston, MA: Pearson.
- Wehmeyer, M., Agran, M., and Hughes, C. (1998). *Teaching Self-Determination to Students with Disabilities: Basic Skills for Successful Transition*. Baltimore, MD: Paul H. Brookes.
- Wehmeyer, M. L. and Palmer, S. B. (2003). Adult outcomes for students with cognitive disabilities three years after high school: The impact of self-determination. *Education and Training in Developmental Disabilities* **38**, 131–144.

Further Reading

- Carter, E., Lane, K., Pierson, M., and Glaeser, B. (2006). Self-determination skills and opportunities of transition-age youth with emotional disturbance and learning disabilities. *Exceptional Children* **72**, 333–346.
- Field, S. and Hoffman, A. (2002). Preparing youth to exercise self-determination: Quality indicators of school environments that promote the acquisition of knowledge, skills, and beliefs related to self-determination. *Journal of Disability Policy Studies* **13**, 113–118.
- Individuals with Disabilities Education Improvement Act Amendments of 2004, 20 U. S. C. 1400 et seq. H.R. 1350, Sec 602(34).
- The Carl D. Perkins Vocational and Technical Education Act of 1998, 20 U. S. C. 2301 et seq.
- The Rehabilitation Act 1973 and its amendments, 29 U. S. C. 794.

- Sands, D. and Doll, B. (1996). Fostering self-determination is a developmental task. *Journal of Special Education* **30**, 58–76.
- US Department of Education (2006). *Twenty-Sixth Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act, 2004*. Washington, DC: Author.
- Wehmeyer, M. and Lawrence, M. (1995). Whose future is it anyway? Promoting student involvement in transition planning. *Career Development for Exceptional Individuals* **18**, 69–83.

Relevant Websites

<http://www.dcdt.org> – Division on Career Development and Transition (DCDT).

<http://www.ihdi.uky.edu> – Community-Based Work Transition Program (CBWTP).

<http://www.ahead.org> – High School to the Community Transition for Students with Disabilities Association on Higher Education and Disability (AHEAD).

<http://www.transad.pop.upenn.edu> – MacArthur Foundation: The Network on Transitions to Adulthood.

<http://tinyurl.com> – National Center for Learning Disabilities (NCLD).

<http://www.ncwd-youth.info> – National Collaborative on Workforce and Disability for Youth.

<http://ici2.umn.edu> – National Transition Network, University of Minnesota.

<http://www.pepnet.org> – Postsecondary Education Programs Network (PEPNet).

<http://www.transitioncoalition.org> – Transition Coalition.

Visual Impairment

W Sapp, Visual Impairment Education Services, Cohutta, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Blindness – It technically refers to a total absence of vision, although the term is often used to refer to severe visual impairments that result in a need for primarily using nonvisual sensory information.

Low vision – It refers to visual impairments that are less severe than blindness but still impact a person's ability to complete daily activities to some degree.

Visual impairments – These include low vision and blindness and refer to any degree of impairment to a person's ability to see, which affects his or her daily life.

Visual impairments, including low vision and blindness, occur in a small percentage of children in all countries around the globe. The percentage of children with visual impairments and the causes of visual impairment vary greatly between developing countries and wealthier nations; and access to clean water and medical care are major influences on the rates and types of visual impairments. Regardless, all children with visual impairments deserve an appropriate education that gives them “the opportunity to be equal and the right to be different” (Hatlen, 1996: 25). In this article, the author advances this premise.

Legislative and Policy Support for Students with Visual Impairments

According to the United Nations' Universal Declaration of Human Rights, article 26, all people have the right to a free and compulsory primary education and parents have the right to choose the kind of education that will be given to their children. The United Nations' International Covenant on Economic, Social, and Cultural Rights, which has been ratified by 155 nations, further expounds upon the right to a free education stating that education shall encourage the development of dignity and “shall enable all persons to participate effectively in a free society.” The basic human right to receive an education applies to all children and youth, including those with visual impairments.

Many countries have passed federal legislation to codify the aforementioned educational rights for people with visual impairments. For example, in the US, the 1990 Individuals with Disabilities Education Act outlines the

educational rights for all students with disabilities and makes specific reference to some of the unique needs of students who are blind. This law requires that materials be provided in appropriate formats for students who are blind and that Braille instruction be provided for students who are blind unless their educational team determines that another media (e.g., print with optical devices) is more appropriate for the student. In Great Britain, the Disability Discrimination Act protects the rights of children and adults to receive a basic and higher education without discrimination due to their disabilities, including visual impairments. In Canada, the Charter of Rights and Freedoms and the Education Act guarantee appropriate access and educational services for people with visual impairments and other disabilities. Even with these worldwide efforts, people with visual impairments still fail to receive appropriate educational programming to maximize their full potential.

Etiologies of Visual Impairments and Their Impact

Before discussing the implications of visual impairments, it is important to understand several key concepts: definitions of key terms related to visual impairments, the most prevalent etiologies of visual impairments, and the functional implications of different types of visual impairments.

Definitions

Visual impairments include low vision and blindness and refer to any degree of impairment to a person's ability to see that affects his or her daily life. Blindness technically refers to a total absence of vision, although the term is often used to refer to severe visual impairments that result in a need for primarily using nonvisual sensory information. Low vision refers to visual impairments that are less severe than blindness but still impact a person's ability to complete daily activities to some degree. People with low vision may need to use tools and techniques to enhance their ability to use their limited vision, or they may need to use nonvisual means for completing tasks.

Etiologies and Prevalence

The rate of visual impairments in children in wealthy nations is approximately 1–2 per 1000 children with

only approximately 1 in 10 000 having a severe visual impairment or total blindness. In developing countries, the rates of severe visual impairments are estimated to be ten times higher than in wealthier nations, occurring at a rate of one in 1000 children (Gilbert *et al.*, 1999).

The most prevalent types of visual impairments vary depending on age and geographic location. For children in developing countries, the most prevalent forms of visual impairments are vitamin A deficiency, corneal scarring, trachoma (an eye infection), inherited conditions, and onchocerciasis (river blindness caused by a parasitic worm that lives in contaminated waterways) (Gilbert *et al.*, 1999; Roodhooft, 2002). With the exception of inherited conditions, all of these causes of visual impairments are preventable through proper nutrition, hygiene, and medical care. Unfortunately, many children in developing countries do not have access to these basic needs that would prevent their visual impairments.

In wealthier nations where basic nutrition, sanitation, and healthcare are more readily available, different causes result in most childhood visual impairments. In the US, the most common causes of visual impairments in children are cortical vision impairment, optic nerve hypoplasia, retinopathy of prematurity, albinism, and abnormalities in the structures of the eye (Hatton and Model Registry of Early Childhood Visual Impairment Collaboration Group, 2001). Some of these causes of visual impairment are often associated with the presence of additional disabilities (e.g., cortical visual impairment) and the rates of these etiologies have increased as medical care allows children with more severe disabilities and children born with extremely low birth weights to survive.

Functional Implications

Visual impairments impact many aspects of seeing. Most visual impairments are the result of damage to one or more structures in the eye. When the eye is damaged, there can be reduced acuity (clarity of what is seen), reduced visual fields (e.g., tunnel vision or blind spots), reduced ability to see color or contrast, or difficulty with lighting and glare, or some combination of these problems. Visual impairments, such as cortical vision impairment and optic nerve hypoplasia, are the result of a poorly developed or damaged optic nerve or visual cortex. These conditions can have any or all of the problems associated with damage to the eye; and additionally, there is difficulty in the brain receiving and processing visual information.

Coupled with the causes of visual impairments and the differing impacts on vision that can occur, a major factor in the needs of a person with visual impairments is based on when the visual impairment first occurred. Adventitious visual impairments occur after a child or teen has lived with typical vision for many years, benefiting from

visual information to understand and interact with his or her world. Children who experience an adventitious visual impairment have to adjust psychologically to the vision loss and learn new ways to do things they already knew how to do (e.g., dressing, reading, and eating). Congenital visual impairments occur at birth or when a child is very young. Children with congenital visual impairments develop their concepts about the world with limited or no visual input. Vision typically helps children unify the information they receive through their other senses and helps them gather vast amounts of information more quickly than they can do with other senses. When vision is absent or impaired in the early years, children need intentional interactions with their world to develop logical and accurate concepts.

Educational Needs of Students with Visual Impairments

When parents and teachers learn that a child has a visual impairment, they are often at a loss as to what that student needs. Each student has unique needs but there are some general guidelines that apply to all children with visual impairments. When selecting adaptations for a child, it is important that only the adaptations that are appropriate for him/her are used. Using adaptations that are not appropriate for a child may result in no benefit for him/her, may be harmful for the child's education, and may promote unhealthy attitudes about functioning with a visual impairment for the child, the family, and the community.

Due to the vast range of visual etiologies, visual functioning, and visual experiences, it is impossible to predict the educational needs of students just by knowing that they have a visual impairment. It is critical to know how they functionally use their vision and other senses in their usual daily routines. It is also important to know if their visual conditions are stable or fluctuating and if they are expected to improve or worsen. Additionally, if other disabilities are present, their impact on learning must also be considered. Due to the complexity of these factors, it is best if a teacher knowledgeable about educating students with visual impairments takes an active role in planning and implementing an educational plan for students with visual impairments. Most countries have teachers who specialize in teaching these students and their expertise should be relied upon when working with them. General information on medical care, expectations for achievement, and typical adaptations must be provided; this information is applicable to many, but not all, children with visual impairments.

Medical care is obviously essential in monitoring the eye health of children with visual impairments. Proper medical care can prevent many of the visual impairments

and minimize the severity of others. Even when people are totally blind, they should still receive regular eye examinations to ensure that there are no additional problems with the eyes and optic nerve such as infections or tumors. The frequency of such examinations should be determined by the child's ophthalmologist. When working with children with visual impairments, adults must maintain high standards and high expectations for them. They should be (1) expected to learn the same skills and knowledge that other children the same age are learning, (2) held to the same standard in completing work and chores at home and at school, and (3) expected to interact appropriately with adults and peers demonstrating positive social skills. To a large measure, when providing instruction to children with visual impairments, there are some general guidelines that may be useful. Teachers and service providers must always introduce themselves when approaching a child who is blind; they must not ask the child to guess who they are from their voice. They must describe information that other students are receiving visually. In addition, they must use precise language rather than general terms such as "over there" that cannot be understood without accompanying gestures. Finally, they must realize that children will have different levels of comfort with their visual impairment and, therefore, they must be discreet in discussing their visual impairments and offering adaptations.

For children with low vision, there are five adaptations to the environment that can enhance their ability to use their vision (Corn, 1989). Depending on the type of visual impairment, increasing or decreasing lighting can improve visual performance. For almost all children with low vision, positioning the child so that there is minimal glare from lights is beneficial. Allowing the child to position himself/herself at the best distance for viewing is helpful. While most children with visual impairments see better the closer they are to material, some children see best from a distance. Adjusting the color and contrast of an object can also allow a child to see the object better. Some children will see certain colors more easily than others and most children will benefit if there is high contrast between the color of the object or print and the color of the background. The final adaptation is to allow the child more time to process the information received visually or to take breaks to prevent visual fatigue. Each child's visual needs are different and, therefore, it is important to provide the appropriate adaptations and only adaptations that are needed. Many children with visual impairments can be taught in their local schools with support from specialists who can direct adaptations, share modified materials and resources, and provide specialized instruction as needed. Other children may require more extensive support from specialized classrooms or schools for children with visual impairments.

Accessing Materials

To receive an appropriate education, students with visual impairments must be able to access the same learning materials (books, pictures, maps, and graphs) as other students. Lack of access to written material is one of the greatest challenges faced by people with visual impairments. A variety of ways of accessing written materials are available to people with visual impairments; these ways are referred to as learning media. Determining the appropriate media requires knowledge and understanding of visual impairments, their impact on learning, and the individual needs of the student. There is no hierarchy of preference in which media is used, rather the best media for a student is the one that allows the student to gain knowledge and information as independently and efficiently as possible (Lusk and Corn, 2006). Most frequently, students use a combination of media to read different materials depending on the time, location, and purpose of reading. Some students find that using a combination of Braille, print, and taped materials allows them to learn most effectively. For instance, when students with visual impairments read print, there are several options available to them: regular print, regular print with optical devices, large print, and electronic magnification. Students with low vision often experience eye fatigue when reading and writing, resulting in headaches, blurred vision (beyond the child's typical visual impairment), and slow reading speeds. The goal of using adapted reading materials is to allow students access to visual materials that cannot be easily seen so that they can read more quickly and with more comfort. With all types of visual learning, adjusting lighting can be extremely helpful. Some students will benefit from having brighter light, while others will function better with dimmer lighting. All students benefit from light that produces minimal glare.

Some students with mild low vision are able to access some standard print (e.g., text in elementary level textbooks). To read standard size texts, students may hold the text very close to their eyes; this will not harm their eyes. Other students may use optical devices to access print (Lussenhop and Corn, 2002). Optical devices use lenses to enlarge the size of the image received by the eye (Zimmerman, 1996). Optical devices for near vision, known as magnifiers (hand-held, stand, or spectacle-mounted), enlarge the size of print or other material that is viewed at close range, within a few inches. These devices allow students to access print that is too small for them to comfortably read without magnification and allows students to view other learning materials such as maps, diagrams, charts, and pictures. Optical devices known as monoculars or telescopes allow students to view print and other visual information at a distance such as material printed on a chalkboard, teacher demonstrations, and

audiovisual materials. Optical devices come in many configurations and must be individually prescribed for each student with low vision. Giving a student a device with too much or too little magnification will not be beneficial. Instruction in using the devices greatly increases the ability of students to use the devices for learning and recreational purposes. Clearly, optical devices are not appropriate for all students with low vision who read print, and in these instances, large-print books and materials can be purchased. Large-print materials are simply enlarged versions of standard print books. The font size of texts is increased and usually accompanying graphics are enlarged as well. Some students find large-print textbooks helpful, but for others, the amount of enlargement is not sufficient to allow them to read for extended periods of time. For students who require more magnification than can be achieved with optical devices or large print, electronic magnification systems such as closed circuit televisions (CCTVs) can be used. CCTVs use a camera to capture an image and display the enlarged image on the screen. They can be adjusted from low levels of magnification to extremely high levels. Some CCTVs are able to use a movable camera to capture both near images, such as a textbook, and distance images, such as the chalkboard. However, CCTVs are bulky, require electricity to run, and are expensive, but for some students, they allow access to materials that would otherwise be unavailable.

Many students with visual impairments find Braille to be the most efficient reading media. Braille is a system of embossed dots arranged in a six-dot cell used for reading and writing. Although other embossed systems of reading have been used, Braille is the most frequently used system. Braille, originally developed in the 1700s in France, has been adopted by most countries as the system for tactual reading and writing. It employs different combinations of dots to represent letters, punctuation marks, and accents that are applicable to any language. Standard Braille can be used with any language using the Latin alphabet; and many other written systems, including but not limited to Hebrew, Japanese, and Chinese, also have Braille versions. In English, there are additional Braille symbols to represent certain words and combinations of letters; and there are additional codes and symbols to handle specialized writing such as math, computer code, and music. Official rules for Braille usage in the US and Canada are determined by the Braille Authority of North America. The International Council of English Braille promotes the unified English Braille code that provides a single code for literature, math, and computers. This code has been adopted by Australia, New Zealand, Nigeria, and South Africa.

Regardless of which code is used, students must have a means for writing using Braille so that they can read their own work. The slate and stylus is the original and least expensive tool for writing Braille. The slate is a metal or

plastic hinged frame in which paper is inserted. The stylus is an instrument with a metal point used to press each dot into the paper from the back. Manual Braille writers, such as the Perkins Braille writer, are similar to manual typewriters, but rather than having standard letters on the keyboard, they have six keys for the six dots in the Braille code. When one or more keys are pressed, the corresponding dots are embossed onto the paper. Electronic Braille writers are similar to manual Braille writers but they require less finger strength to use and provide options for more easily correcting errors. Electronic Braille note takers are small computers employing a six-eight button key pad for input. These devices typically have word processing programs, calculators, calendars, address books, and other common computing programs. They allow users to read back their material using voice output, refreshable Braille (a series of metal pins that raise and lower to form Braille cells), printing in Braille when connected to a Braille embosser, and printing in print when connected to a printer. Computers with specialized software can also be used to write in Braille either using the standard keyboard or using part of the home row as a standard six-key Braille keyboard.

Though students need a written form of communication, many students find auditory materials to be good supplements to print and/or Braille. Occasionally, a student will have a visual impairment that does not allow print reading and a physical impairment that prevents Braille reading; this student will, by necessity, be an auditory reader. Auditory materials are available in four-track tapes that are coded with sections and pages so students can more easily find specific material. Auditory materials are also available in digital formats that allow even greater flexibility in use. One of the benefits of taped and digitized material is that it can be played at faster-than-average speeds, thus decreasing the amount of time necessary to listen to the material. The use of live readers (a person who reads the material aloud) is also an effective tool for students to access materials that are not available in any other format.

Unique Instructional Needs

Students with visual impairments are limited in the visual information they receive and, therefore, they often require instruction in areas that go beyond standard school curricula. One formulation of the instruction needed by students with visual impairments is the expanded core curriculum (Hatlen, 1996), a curriculum that proposes nine areas in which students with visual impairments should be assessed. For any area in which deficits are noted, instruction should be provided enabling the student to learn skills commensurate with their typically sighted peers and to fully participate in society. When a student

has a visual impairment and additional disabilities, they may require more assistance in the following areas and may never reach full independence, but they should still receive instruction in these areas appropriate to their potential:

1. *Academic and compensatory skills.* These refer to the standard academic skills and adapted skills that allow students to participate in standard academics. Compensatory skills could include instruction in Braille reading and writing or the use of an abacus to compute mathematical problems.
2. *Career skills.* These are the skills necessary to participate in the work force and include everything from knowledge of possible jobs to an ability to work independently and complete tasks. It includes more advanced skills such as completing a job application, knowing how to act in an interview, and gaining the skills needed for a specific job.
3. *Independent living skills.* These are the basic tasks that people complete each day to care for themselves. Independent living skills include skills necessary for hygiene, grooming, dressing, eating, shopping for necessities, etc.
4. *Orientation and mobility.* These refer to the systematic skills and concepts necessary for people with visual impairments to know their location in space and to travel safely, independently, gracefully, and efficiently in all environments (Blasch *et al.*, 1997). Instruction may involve learning to use all senses to gather information about one's environment, developing skills using a long cane to detect obstacles, and determining when to request assistance from others. In many countries, instructors specialize in orientation and mobility so that instruction can be provided to maximize independent travel while minimizing the dangers inherent in traveling independently with limited or no sight.
5. *Recreation and leisure skills.* These are skills for spending one's free time in enjoyable and healthy ways. These are important skills for maintaining one's physical and emotional health.
6. *Self-advocacy skills.* These are skills that allow a student to request assistance when needed and reject assistance when not required. Learned helplessness frequently occurs in children with visual impairments when well-meaning adults and peers do everything for the child. Self-advocacy skills allow children to learn to be independent and feel confident in their own abilities.
7. *Social skills.* These are the skills necessary to interact with family, friends, and community members. Skills as basic as knowing how to position yourself when talking with someone can be difficult for a child with a severe visual impairment who cannot observe others. Social skills are important for the person with the visual impairment to develop friendships, find a mate,

interact with people in their communities, and maintain employment.

8. *Technology skills.* These are critical for allowing access to the wider world of knowledge and information. Advances in technology have changed the world for everyone and adapted software and hardware allows people with visual impairments to also benefit from the new technology. Knowing what technology is appropriate and how to use that technology requires instruction and practice.
9. *Visual and sensory skills.* These refer to learning how to best use one's remaining vision and other senses to gain information about the world.

Not all students will need instruction in every area of the expanded core curriculum, but most will have deficits in at least one area. Recognizing that these are potential areas of need and addressing concerns that are identified can greatly enhance the educational outcomes for students with visual impairments.

Conclusion

Students with visual impairments are found in every country in the world. Their degree of visual impairment ranges from very mild to total blindness, meaning that there is not a single set of adaptations or skills that will be appropriate for all children. A teacher knowledgeable about educating children with visual impairments should evaluate each child to determine the best way to meet his/her educational needs. Based on this assessment, instructional adaptations, specialized materials, and instruction in unique skills must be provided. When adequate instruction is provided and a child is expected to meet high standards, there are no limits to what the child can accomplish.

See also: Assessment Accommodations for Children with Special Needs; Assistive Technology and Educational Practice; Educating Children Who Are Deaf-Blind; Parent and Family Involvement in the Education of Children with Special Needs.

Bibliography

- Blasch, B. B., Wiener, W. R., and Welsh, R. L. (1997). *Foundations of Orientation and Mobility*, 2nd edn. New York: AFB Press.
- Corn, A. L. (1989). Instruction in the use of vision for children and adults with low vision: A proposed program model. *Review* 21, 26–38.
- Gilbert, C. E., Anderton, L., Dandona, L., and Foster, A. (1999). Prevalence of visual impairment in children: A review of available data. *Ophthalmic Epidemiology* 6(1), 73–82.
- Hatlen, P. (1996). The core curriculum for blind and visually impaired students, including those with additional disabilities. *Review* 28, 25–34.

- Hatton, D. D. Model Registry of Early Childhood Visual Impairment Collaborative Group (2001). Model registry of early childhood visual impairment: First year results. *Journal of Visual Impairment and Blindness* **95**(7), 418–433.
- Lusk, K. E. and Corn, A. L. (2006). Learning and using print and braille: A study of dual-media learners: Part 1. *Journal of Visual Impairment and Blindness* **100**, 606–619.
- Lussenhop, K. and Corn, A. L. (2002). Comparative studies of the reading performance of students with low vision. *Review* **34**, 57–69.
- Roodhooft, J. M. J. (2002). Leading causes of blindness worldwide. *Bulletin of the Belgian Societies of Ophthalmology* **283**, 19–25.
- Zimmerman, G. J. (1996). Optics and low vision devices. In Corn, A. L. and Koenig, A. J. (eds.) *Foundations of Low Vision: Clinical and Functional Perspectives*, pp 115–142. New York: AFB Press.
- Joshi, M. R., Yoshitaka, Y., Akura, J., and Shakya, S. (2008). The efficacy of low vision devices for students in specialized schools for students who are blind in Kathmandy Valley, Nepal. *Journal of Visual Impairment and Blindness* **102**, 430–435.
- Lohmeier, K. (2009). Aligning state standards and the Expanded Core Curriculum: Balancing the Impact of the No Child Left Behind Act. *Journal of Visual Impairment and Blindness* **103**, 44–47.
- Lohmeier, K., Blankenship, K., and Hatlen, P. (2009). Expanded Core Curriculum: 12 Years Later. *Journal of Visual Impairment and Blindness* **103**, 103–112.
- Lueck, A. H., Erin, J. N., Corn, A. L., and Sacks, S. Z. (2006). *Facilitating Visual Efficiency and Access to Learning for Students with Low Vision*. Council for Exceptional Children, Division of Visual Impairments Position Paper, retrieved. <http://www.cecdivi.org/positionpapers.html>.
- Sacks, S. Z., Lueck, A. H., Corn, A. L., and Erin, J. N. (2006). *Supporting the Social and Emotional Needs of Students with Low Vision to Promote Academic and Social Success*. Council for Exceptional Children, Division of Visual Impairments Position Paper, retrieved. <http://www.cecdivi.org/positionpapers.html>.
- Spungin, S. J. and Ferrell, K. A. (2007). *Role and Function of the Teacher of Students with Visual Impairments*. Council for Exceptional Children, Division of Visual Impairments Position Paper, retrieved. <http://www.cecdivi.org/positionpapers.html>.
- Toelle, N. M. and Blankenship, K. E. (2008). Program accountability for students who are visually impaired. *Journal of Visual Impairment and Blindness* **102**, 97–102.
- Erin, J. N., Holbrook, M. C., Sanspre, M. J., and Swallow, R. M. (2006). *Professional Preparation and Certification of Teachers of Students with Visual Impairments*. Council for Exceptional Children, Division of Visual Impairments Position Paper, retrieved. <http://www.cecdivi.org/positionpapers.html>.
- Griffin-Shirley, N., Kelley, P., and Lawrence, B. (2006). *The Role of Orientation and Mobility Specialists in Public Schools*. Council for Exceptional Children, Division of Visual Impairments Position Paper, retrieved. <http://www.cecdivi.org/positionpapers.html>.
- Hatton, E., et al. (2003). *Family-Centered Practices for Infants and Young Children with Visual Impairments*. Council for Exceptional Children, Division of Visual Impairments Position Paper, retrieved. <http://www.cecdivi.org/positionpapers.html>.
- Holbrook, M. C. (2009). Supporting students' literacy through data-driven decision-making and ongoing assessment achievement. *Journal of Visual Impairment and Blindness* **103**, 133–136.
- Huebner, K. M., Garber, M., and Wormsley, D. P. (2006). *Student-Centered Educational Placement Decisions: The Meaning, Interpretation, and Application of Least Restrictive Environment for Students with Visual Impairments*. Council for Exceptional Children, Division of Visual Impairments Position Paper, retrieved. <http://www.cecdivi.org/positionpapers.html>.

Further Readings

Relevant Websites

- <http://www.acb.org> – American Council of the Blind.
- <http://www.afb.org> – American Foundation for the Blind.
- <http://www.aerbvi.org> – Association for Education and Rehabilitation of the Blind and Visually Impaired.
- <http://www.once.es> – *Organizacin Nacional de Ciegos Españoles* (National Organization of the Spanish Blind).
- <http://www.mib.org.uk> – Royal National Institute of Blind People (RNIB).
- <http://www.cnib.ca> – The Canadian National Institute for the Blind.
- <http://www.visionaustralia.org.au> – Vision Australia.
- <http://www.worldblindunion.org> – World Blind Union.

Workforce Issues in Special Education

E Carlson, Westat, Rockville, MD, USA

B Billingsley, Virginia Tech, Blackburg, VA, USA

© 2010 Elsevier Ltd. All rights reserved.

Maintaining an adequate supply of qualified personnel to educate children with special needs is an ongoing challenge that is embedded within a cultural, political, and economic context. This context helps to shape the workforce through its influence on personnel shortages/surpluses; expectations for educators, in terms of professional knowledge and skills; and the conditions of service that educators encounter in their work. There is no consensus on what constitutes personnel quality in general or special education. As such, it may be more appropriate to discuss a qualified workforce as opposed to a high-quality workforce, since qualifications can be observed, measured, or documented more easily than quality, recognizing that the one is an imperfect substitute for the other. Credentialing agencies set minimal standards for entry into the profession, and as such, define workforce qualifications. However, it is important to note that those standards differ considerably across jurisdictions.

In any field, members of a qualified workforce are expected to possess the requisite professional knowledge and skills. Over time and across locales, those may differ. Educators typically acquire their professional knowledge and skills through initial preparation, induction, and continuing professional development, including work experience. As such, institutions that provide education and training to teachers and other service providers help shape the knowledge and skills they possess. In addition, conditions of service constitute the third factor promoting or impeding development of a qualified workforce. These include work contexts, such as school climate, roles and responsibilities, and levels of support, as well as overall compensation (e.g., salary, benefits, and rewards of teaching). Conditions of service play a critical role, not only affecting the performance of the workforce directly but also affecting it indirectly through its influence on supply/demand and turnover/retention.

It is important to note that in some countries, a qualified and committed special education workforce includes not only special education teachers but also high-quality general education teachers and related service personnel, for example, speech-language pathologists, physical therapists, occupational therapists, and paraprofessionals. In other countries, related services are provided through healthcare systems. While the myriad of professionals involved in providing services to children with special needs makes workforce issues complex (see **Figure 1**), the underlying constructs driving workforce quality are

similar for each type of service provider. Thus, it is critical to understand workforce issues in special education, the focus of this article.

Personnel Shortage/Surplus

In traditional microeconomic models, shortages of goods or services exist when demand exceeds supply. Many geographic regions have experienced long-standing shortages of personnel to meet the educational needs of children with disabilities. During 1997 to 2003, between 7 and 13% of special education teachers in the US were not fully certified for their positions (Boe, 2006). In education, workforce shortages can be quantified in two primary ways: the number of unfilled vacancies and the number of individuals employed who do not meet minimum job qualifications (e.g., certification requirements). Counts of unfilled vacancies are typically much lower than counts of unqualified staff, as education agencies assume that even an unqualified educator is better than none. To properly discuss personnel shortage or surplus, it is critical to define workforce supply, demand, and retention.

Workforce Supply

Understanding long-term shortages of personnel in special education requires familiarity with workforce supply as well as workforce demand. On the supply side, qualified special education personnel are available from a variety of sources:

- current workforce of qualified and unqualified personnel;
- new, traditionally prepared personnel;
- new, alternatively prepared personnel;
- reserve pool (i.e., personnel who are fully qualified but not currently working in the field); and
- underqualified individuals already working in the field who may complete the necessary preparation to become fully qualified (McLeskey *et al.*, 2004).

Graduates of traditional teacher education programs account for a sizable portion of new special education teachers, and in the US, the production of teachers from these traditional programs increased by over 20% from 1993 to 1998 and now exceeds 20 000 graduates per year (US Department of Education, National Center for Education Statistics (NCES), 2001; McLeskey *et al.*, 2004). Of course,

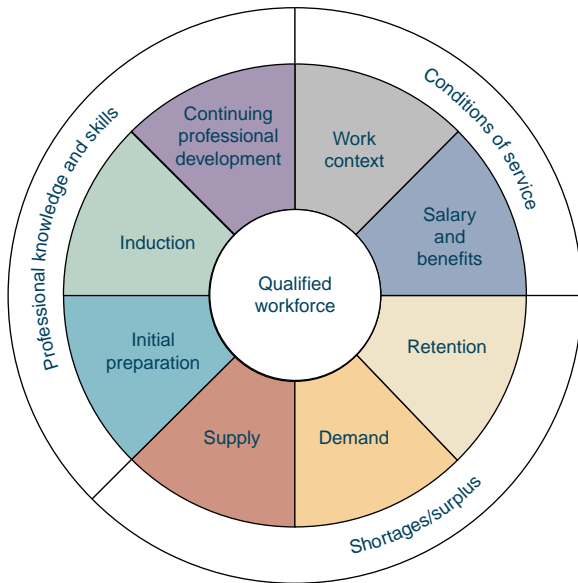


Figure 1 Factors in securing a qualified workforce.

not everyone completing a teacher-preparation program will take up a teaching position. In the US, the percentage of graduates taking up teaching positions in the year after graduation may be as low as 40% (Boe *et al.*, 1999).

Newly prepared special educators and related service providers who do not immediately take a special education position enter the reserve pool of those qualified but not employed in the field. The reserve pool can be a significant source of qualified personnel, since many members of the reserve pool eventually enter or return to teaching.

In reaction to chronic shortages of education professionals, alternative routes to certification have rapidly expanded. These preparation programs provide access to teaching credentials that circumvent conventional college and university preparation programs. They vary considerably in length and structure, delivery mode, and candidate population. Alternative routes to certification now represent a small but growing source of new special education professionals (Feistritzer, 2001). Another group of special educators are those teaching but not yet fully qualified for their positions. Very little is known about this group of teachers, except that research suggests they are likely to become a turnover liability (Miller *et al.*, 1999).

Workforce Demand

Two primary factors drive the demand for special education teachers and related services personnel. The first is the increasing student enrolment in special education. Rapid enrolment growth in special education has had a dramatic effect on demand, particularly in light of the numbers of

teachers prepared yearly for this field. For example, between 1992 and 1999, the population of school-age children in the US grew 7% nationally. In comparison, the special education enrolment increased 20% over the same period, fueling the demand for special educators and related services personnel (McLeskey *et al.*, 2004).

Caseloads for special educators and related service personnel are the second factor driving demand. If jurisdictions have no caseload limits, educators may serve larger and larger numbers of children, undoubtedly affecting the quality of services, but without increasing the number of vacancies or positions filled by uncertified personnel. However, because of concerns about the quality of services, many jurisdictions have limits on the number of children that can be served by any one individual or in one class. Increases in special education caseloads and waivers to bypass caseload limits may be a byproduct of personnel shortages and, in fact, may mask shortages in some locales.

Workforce Retention

Retention or its inverse, turnover, is a mediating variable in the supply–demand equation. How long special education teachers stay in teaching or in a district influences the demand for new personnel. Turnover is a broad term that refers to teacher attrition and teacher migration to other districts and positions. For example, turnover may refer to leaving the profession, transferring to other districts or states, or changing fields within education. Depending on the definition of leaving, rates will vary widely from 7% for those leaving the teaching profession to roughly 22% for all types of turnover combined (Boe *et al.*, 2007). Although turnover is sometimes positive, because recently hired teachers are a source of new ideas and enthusiasm, high levels of attrition are expensive. Turnover requires that limited resources be spent on recruitment, hiring, and induction, and can interfere with the development of coherent programs. In periods of teacher shortage, high levels of turnover are particularly problematic, since it is often difficult to find qualified replacements for those who leave. Ingersoll (2001) used the image of a revolving door to illustrate the difficulty of building a qualified workforce when many leave after only a few years in the classroom. Some school systems have a more difficult time both attracting and retaining teachers. For example, urban and rural systems have higher turnover than suburban systems. Not surprisingly, high-poverty school systems also tend to have more turnover than wealthier ones (Guarino *et al.*, 2006). When turnover is high, students are more likely to be taught by less-qualified and less-experienced teachers.

Retention and turnover are influenced by a wide range of complex variables: economic conditions, the age distribution of staff, and the attractiveness and compensation of teaching compared with other occupations. Efforts to

reduce turnover need to be based on an understanding of factors that contribute to leaving, such as teacher characteristics, teacher ability and preparation, and working conditions. Although several researchers have investigated the relationship of teacher characteristics, such as age, gender, and race, few conclusions can be drawn. Age is the only demographic variable that is consistently linked to turnover in the special education literature. Early career leaving is often presented as part of a U-shaped curve, where teachers are more likely to leave early in their careers, less likely during mid-career, and then more likely as they get closer to retirement. Reasons for higher levels of leaving among early career teachers have been attributed to fewer accrued investments (e.g., home and retirement system) (Grissmer and Kirby, 1987) as well as difficulties some new teachers have adjusting to teaching (Ingersoll, 2001). The relationship between gender and attrition has been included in only a few special education studies, and the findings are mixed (Billingsley, 2004). Race as a factor in special education-leaving studies has also been inconclusive (Singer, 1992).

Teacher qualifications have received less attention in the turnover literature than any other set of variables. Researchers have demonstrated higher levels of attrition among uncertified teachers than certified teachers (Boe *et al.*, 1999). Only a few researchers have considered the association between special educators' academic abilities and their decisions to stay or leave, and one large-scale study has shown that teachers with higher tested ability are more likely to leave the field (Singer, 1992). The most researched area is the influence of working conditions on teacher turnover. Special educators leave teaching for two types of reasons: personal (e.g., staying home with children or relocating) and professional. Overall, special education teachers are more likely to leave because of dissatisfaction with their work than general educators. The reasons special education teachers give for leaving include problematic work circumstances, such as school climate, isolation, high caseloads, lack of administrative support, and excessive bureaucratic responsibilities (Billingsley, 2004).

Professional Knowledge and Skills

Teacher educators, policymakers, and researchers agree that expert teachers are knowledgeable about their subject. For example, teacher knowledge of language has been correlated with observed classroom practices, and teachers with subject-matter preparation are able to secure better gains in student achievement than their counterparts without such preparation. However, in addition to subject-matter knowledge, expert teachers may have knowledge that is unique to teaching content, and this knowledge enables them to represent subject matter for the novice learner. This is the knowledge that teachers use for instructional purposes,

including the decisions made during planning and instruction. For example, research suggests that teachers with both subject-matter expertise and education coursework in methods for teaching a particular subject secure larger student-achievement gains than teachers with subject-matter expertise alone (Monk, 1994).

It is important to recognize that the knowledge and skills required for a specific teaching assignment may differ over time and across locations. The curriculum is driven by the culture in which children are being educated, and so are expectations for achievement. For example, as children with less-severe disabilities are identified, the special education curriculum may focus increasingly on academic content and away from a life-skills orientation. Additionally, it is important to note that special educators' responsibilities extend beyond instruction. The Council for Exceptional Children (2003), an international professional organization, developed standards for professional practice for special educators. These include instructional responsibilities as well as management of behavior, support procedures (such as delivering therapies or administering medications), and parent relationships.

Initial Preparation

Initial preparation is designed to equip educators with the basic knowledge and skills necessary to assist children in achieving learning standards established by their education agencies. Fully certified teachers produce higher achievement scores in mathematics and reading than less-than-fully-certified teachers; teachers with subject-matter preparation in mathematics and science accompanied by related pedagogical courses influence student achievement more than teachers with only subject-matter preparation; and teachers with pedagogical and content preparation are better able to engage students in the learning process. However, research on teacher-preparation programs lacks empirical evidence linking program features with classroom practices and student performance (Wilson *et al.*, 2001). There are a number of reasons for this research deficit: teacher preparation is a long and complex process; the direct outcome (i.e., teacher quality) is ill-defined; and working conditions present a strong moderating effect on teaching performance, which makes measures of the ultimate outcome of interest (i.e., child outcomes) difficult to attribute to teacher preparation.

Expert opinion has been used to identify some features considered exemplary in initial preparation of teachers. These features include:

- coherent program vision;
- conscious blending of theory, disciplinary knowledge, and subject-specific pedagogical knowledge and practice;
- carefully crafted field experiences;
- standards for ensuring quality graduates;

- active faculty pedagogy that employs modeling, connects theory and practice, and promotes reflection;
- focus on meeting the needs of a diverse student population; and
- collaboration as a vehicle for building professional community (Brownell *et al.*, 2003).

Graduates of less-intensive special education training programs, particularly brief alternative routes, feel less well-prepared and are less likely to stay in the field (Boe *et al.*, 2006). However, no consistent differences have been documented in traditionally and alternatively certified teachers in terms of teacher efficacy or teacher competence as measured by classroom observations or principal or supervisor ratings of teacher competence (Hawley, 1992; Rosenberg and Sindelar, 2005).

Induction

New special educators often enter the workforce with unrealistic expectations and report that their first years in the classroom are difficult and stressful. These new teachers struggle to apply what they have learned in their teacher-preparation programs at the same time that they must establish routines, collaborate with general education teachers, and manage a range of varied tasks in limited time. New teachers with difficult assignments and little support often leave teaching prematurely, making it difficult to build a qualified workforce. To support teachers through these intense early years, researchers and policymakers have called for responsive induction programs. Induction programs are built around the belief that regardless of the quality of initial preparation, new teachers need support and guidance as they learn to apply their knowledge and skills in complex school settings. Although there is some evidence that induction makes a difference in teacher retention and quality, many new special education teachers receive minimal support and assistance (Billingsley *et al.*, 2004).

Induction programs need to be tailored to the qualifications of the individual teacher and the context in which they work. Mentoring is frequently recommended as a component of induction. New teachers benefit from experienced special education mentors who are available to help them with the problems they confront, to observe and provide feedback, and to help them navigate their complex roles in schools. Induction programs may also include orientation seminars, peer meetings, professional development activities, and ongoing opportunities to reflect about teaching practice. Clearly, responsive induction programs require substantial commitments. Some countries and states commit significant resources to prepare and reward mentors, schedule release time for mentors to work with new teachers, and provide ongoing professional development. Others create university-school partnerships

and professional development schools to support induction. Although these model programs are expensive, the investment in induction may be offset by improved teacher quality, increased retention, and reduced attrition costs (e.g., for recruiting and hiring replacements).

Continuing Professional Development

Continuing professional development, including on-the-job training can make an important contribution to educators' professional knowledge and skills, and ultimately to improved outcomes for students with disabilities. The content of staff development needs to focus on research-based practices shown to be effective with students with disabilities. Unfortunately, there is evidence that many of these practices are often not used (Cook and Schirmer, 2003; Gersten and Dimino, 2001). Another area of need is the development of collaborative knowledge and skills, so that special and general education teachers, paraprofessionals, and related service personnel can work together to address the diverse needs of students with disabilities (Friend and Cook, 2003). Surely, professional development plans require attention to not only the content, but also how the content is delivered and supported. Professional development is most likely to be successful in school environments where teacher learning is part of the daily culture. In learning communities, teachers are partners in their own professional development — they plan and work together around instructional challenges, use student data to make modifications in their practices, and discuss what is working and what needs improvement.

It has been suggested that teachers need to have a conceptual understanding of new practices so they can better use and adapt these practices in their classrooms. Teachers are more likely to be successful with new instructional processes when they are given clear and specific guidelines for implementation and when the practices can be used within the structure of their classrooms. As a result, professional development is also encouraged through a range of opportunities, such as learning to be a mentor or teacher leader, participating in an external teacher network or professional organization, or earning advanced credentials, such as a graduate degree or national board certification.

Conditions of Service

Conditions of service refer to compensation and work context as they influence teacher quality. Teacher quality is influenced by the ability to attract and retain the best special educators and to structure the work environment in ways that encourage continuous development of

teacher expertise and promote productive use of teachers' knowledge and skills. Following are subsections that discuss two conditions of service: compensation and work context.

Compensation

The labor economic theory of supply and demand suggests that teachers will enter and stay in teaching (or a particular position) if it is the most attractive activity available based on overall compensation (Guarino *et al.*, 2006). Compensation in this context is not limited to salary and benefits; rather, it includes any aspect of work that influences one's desire to enter, stay or leave, such as location, work schedule, or the intrinsic rewards that one derives from teaching. Conditions of service are important in that they have an influence on the desirability of teaching as a career and, once employed, whether teaching remains an attractive alternative among all of those available.

Work Context

All teachers need certain conditions to do their work well, including strong leadership; a positive school climate; teaching resources; and the time, schedule, and supports needed to carry out their varied responsibilities. Teachers who participate in a learning culture have opportunities to continue to acquire knowledge and skills and to refine their teaching practices, thus improving teacher quality. Knowledgeable teachers in supportive work environments are more likely to be successful and find teaching rewarding. Although many special education teachers are satisfied with their work, not all of them find their work contexts supportive. The turnover literature links a broad range of work problems to teacher dissatisfaction and turnover, such as inadequate or uncompetitive salaries, poor school climates, lack of administrative support, problems with collaboration and inclusion, role overload and paperwork, high caseloads, and isolation. New teachers are particularly vulnerable to these working conditions. Many of these problems prevent teachers from functioning effectively in their roles as teachers. Over time, problematic working conditions lead to stress, dissatisfaction, and reduced commitment. Roughly, one-third of teacher attrition in special education is to general education positions, and teachers' dissatisfaction with special education teaching is a contributor to this loss.

District and school leaders are in a powerful position to shape the work context and the ways in which students with disabilities are served in schools. Leaders who help facilitate shared understanding about these students can do much to reduce the chronic problems reported by special education teachers and other service providers.

Leadership is also needed for the development of induction programs and the creation of reasonable teacher roles. These actions should help not only to alleviate some of the major contributors to turnover, but also to provide teachers with greater opportunities to meet the needs of their students (Littrell *et al.*, 1994).

Conclusions

In summary, although it is widely acknowledged that a knowledgeable and skilled workforce is essential to enhance student achievement, the chronic and persistent shortage of qualified special education teachers and related service personnel threatens the education that students with disabilities receive. Comprehensive efforts by policymakers and administrators are needed to attract and prepare promising personnel, support them in their early years through responsive induction programs, and provide them with ongoing professional development opportunities throughout their careers. At the same time, conditions of service, including salary, benefits, and the work context have an impact on teacher quality. Teachers are more likely to thrive in environments that are supportive, collegial, and include well-designed work assignments. Conditions of work influence not only the extent to which teachers are successful with their students, but also their job satisfaction and career decisions as well.

See also: An Overview of Teacher Labor Markets; Characteristics, Scholarship and Research of Teacher Educators; Determining Long Term Effects of Teacher Education; Economic Approaches to Teacher Recruitment and Retention; Professional Development of Teacher Educators; Teacher Incentives; Teacher Induction; Teacher Quality in Education Production; Teacher Supply; Teacher Training and Preparation in the United States; The Economics of Class Size.

Bibliography

- Billingsley, B. (2004). Special education teacher retention and attrition: A critical analysis of the research literature. *Journal of Special Education* 38(1), 39–55.
- Billingsley, B., Carlson, E., and Klein, S. (2004). *Exceptional Children* 70(3), 333–347.
- Boe, E. E. (2006). Long-term trends in the national demand, supply, and shortage of special education teachers. *Journal of Special Education* 40(3), 138–150.
- Boe, E., Bobbitt, S., Cook, L., Barkanic, G., and Maislin, G. (1999a). Teacher turnover in eight cognate areas: National trends and predictors. *Data Analysis Report No. 1998 DAR3*. Philadelphia, PA: University of Pennsylvania/Graduate School of Education/Center for Research and Evaluation in Social Policy.
- Boe, E., Cook, L., Paulsen, C., Barkanic, G., and Leow, C. (1999b). Productivity of teacher preparation programs: Surplus or shortage in quantity and quality of degree graduates. *Data Analysis Report*

- No. 1999-DAR2. Philadelphia, PA: University of Pennsylvania/Graduate School of Education/Center for Research and Evaluation in Social Policy.
- Boe, E. E., Cook, L., and Sunderland, R. J. (2007). *Teacher Turnover in General Education: Attrition, Teaching Area Transfer, and School Migration*. No. Research Report No. 2006-TSDQ3. Philadelphia, PA: Center for Research and Evaluation in Social Policy/Graduate School of Education/University of Pennsylvania.
- Brownell, M., Ross, D. R., Colon, E. P., and McCallum, C. L. (2003). Critical features of special education teacher preparation: A comparison with exemplary practices in general teacher education. *COPSE Document RS-4*. Gainesville, FL: University of Florida/Center on Personnel Studies in Special Education.
- Cook, B. and Schirmer, B. (2003). *What is special about special education? Overview and analysis*. *Journal of Special Education* 37(3), 200–205.
- Council for Exceptional Children (2003). *What Every Special Educator Must Know: Ethics, Standards, and Guidelines for Special Educators*. Arlington, VA: Author.
- Feistritz, E. (2001). *Alternative Teacher Certification: An Overview 2001*. Washington, DC: National Center for Education Information. http://www.ncei.com/2001_Alt_Teacher_cert.htm (accessed Jun 2009).
- Friend, M. and Cook, L. (2003). *Interactions: Collaboration Skills for School Professionals*, 4th edn. Boston, MA: Allyn and Bacon.
- Gersten, R. and Dimino, J. (2001). The realities of translating research into classroom practice. *Learning Disabilities Research and Practice* 16, 120–130.
- Grissmer, D. W. and Kirby, S. N. (1987). *Teacher Attrition: The Uphill Climb to Staff the Nation's Schools*. Santa Monica, CA: RAND.
- Guarino, C. M., Santibanez, L., and Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research* 76(2), 173–208.
- Hawley, W. D. (1992). The theory and practice of alternative certification: Implications for the improvement of teaching. In Hawley, W. D. (ed.) *The Alternative Certification of Teachers*, pp 3–34. Washington, DC: ERIC Clearinghouse on Teacher Education.
- Ingersoll, R. (2002). *Teacher Turnover, Teacher Shortages, and the Organization of Schools*. Seattle, WA: University of Washington/Center for the Study of Teaching and Policy.
- Littrell, P., Billingsley, B., and Cross, L. (1994). The effects of principal support on special and general educators' stress, job satisfaction, school commitment, health, and intent to stay in teaching, 15. *Remedial and Special Education* 5, 297–310.
- McLeskey, J., Tyler, N. C., and Flippin, S. S. (2004). The supply and demand for special education teachers: A review of research regarding the chronic shortage of special education teachers. *Journal of Special Education* 38(1), 5–21.
- Miller, M. D., Brownell, M., and Smith, S. (1999). Factors that predict teachers staying in, leaving, or transferring from the special education classroom. *Exceptional Children* 65, 201–218.
- Monk, D. H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review* 13, 125–145.
- Rosenberg, M. S. and Sindelar, P. T. (2005). The proliferation of alternative routes to certification in special education: A critical review of the literature. *Journal of Special Education* 39(2), 117–129.
- Singer, J. D. (1992). Are special educators' career paths special? Results from a 13-year longitudinal study. *Exceptional Children* 59(3), 262–279.
- US Department of Education, National Center for Education Statistics (2001). *Digest of Education Statistics*. Washington, DC: Author.
- Wilson, S. M., Floden, R. E., and Ferrini-Mundy, J. (2001). *Teacher Preparation Research: Current Knowledge, Gaps, and Recommendations*. Seattle, WA: Center for the Study of Teaching and Policy/University of Washington.

Further Reading

- Billingsley, B. S. (2005). *Cultivating and Keeping Committed Special Education Teachers: What Principals and District Leaders Can Do*. Thousand Oaks, CA: Corwin Press.
- Billingsley, B. S. and McLeskey, J. (2004). Critical issues in special education teacher supply and demand. *Journal of Special Education* (special issue) 38, 2–4.
- Blanton, L., Sindelar, P., Correa, V., et al. (2002). *Research Summary: Conceptions of Beginning Teacher Quality: Models for Conducting Research*. Gainesville, FL: University of Florida/Center on Personnel Studies in Special Education (COPSSE).
- Carlson, E., Lee, H., and Schroll, K. (2004). Identifying attributes of high quality special education teachers. *Teacher Education and Special Education* 27(4), 350–359.
- Eurydice European Unit (2002–2004). Key topics in education: The teaching profession in Europe (4 volumes). <http://www.eurydice.org> (accessed Jun. 2009).
- Kozleski, E., Mainzer, R., and Deshler, D. (2000). *Bright Futures for Exceptional Learners: An Action Agenda to Achieve Quality Conditions for Teaching and Learning*. Alexandria, VA: Council for Exceptional Children.
- Moore, J. S. and Birkeland, S. E. (2003). Pursuing a sense of success: New teachers explain their career decisions. *American Educational Research Journal* 40(3), 581–617.
- OECD (2005). *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*. Paris: Author.
- Smith, T. M. and Ingersoll, R. M. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal* 41(3), 681–714.
- Tyler, N., Yzquierdo, Z., Lopez-Reyna, N., and Flippin, S. S. (2004). Cultural and linguistic diversity and the special education workforce: A critical overview. *Journal of Special Education* 38(1), 22–38.
- Wayne, A. J. and Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research* 73(1), 89–122.

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

This page intentionally left blank

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

EDITORS-IN-CHIEF

PENELOPE PETERSON
School of Education and Social Policy
Northwestern University
Evanston, IL
USA

EVA BAKER
UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

BARRY MCGAW
McGaw Group Pty Ltd
Mont Albert Nth, VIC
Australia



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
The Boulevard Langford Lane, Kidlington, Oxford OX5 1GB, UK

Copyright © 2010 Elsevier Ltd. All rights reserved

LATVIA
Copyright © 2010 A Joma

STUDENT DATA SYSTEMS AND THEIR USE FOR EDUCATIONAL IMPROVEMENT
Copyright © 2010 J C Wayman, V Cho, and M P Richards

QUALIFICATIONS FRAMEWORKS AND THEIR ROLE IN THE REFORM OF EDUCATION AND TRAINING
Copyright © 2010 M Young

LANGUAGE OF VOCATIONAL EDUCATION AND TRAINING
Copyright © 2010 J Misko, J MacKenzie, and R-A Polvere

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44)(0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at (<http://elsevier.com/locate/permissions>), and selecting *Obtaining permissions to use Elsevier material*

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Catalog Number
A catalogue record for this book is available from the Library of Congress

ISBN: 978-0-08-044893-0

For information on all Elsevier publications
visit our website at books.elsevier.com

PRINTED AND BOUND IN THE UNITED KINGDOM
09 10 11 12 13 10 9 8 7 6 5 4 3 2 1



PREFACE

A preface usually provides a brief introduction to a work, intended to set the stage, provide some background insight, and whet the appetite of the reader. In our case, however, this preface has to address a fundamental question, one that was in our minds at the time we were recruited as Editors-in-Chief for the International Encyclopedia of Education. The question was “Why do we need an Encyclopedia? Its subtext was inspired by the ever-growing, ever-popular Internet. We believe that *this* Encyclopedia is desperately needed and will become a valued resource in education and associated social sciences and arts. The reasons are intellectual and procedural. Anyone with a modicum of knowledge knows that finding and trusting information gleaned from the Internet are two separate actions. The reliance on browsers to help discover references and comments result in resources based on popularity not quality. Pithy titles catch the eye and references rise in the ranks of browser searchers. Related to this is the “editing” in the Internet realm of populist efforts at encyclopedia, references, and other compilations. Once again, after removing offensive material, the accuracy, completeness, lack of bias, and other provenance for entries simply do not exist. Experienced researchers in education can sort through and make intelligent choices. Novices and many journeyman, or practitioners, parents, and policy makers cannot. Contrast how this Encyclopedia was built. Key domains of educational research were identified, and a tentative list of sub-domains or useful applied areas was posited. Then the Editors-in-Chief (apologies for the awkwardness of the term) identified the leading researcher in a particular domain, and with surprisingly little effort, recruited them to participate. They in turn identified the two best researchers in a sub-domain, such as formative assessment or the training of pre-school teachers. The authors of the sections of the Encyclopedia do not represent a collective group of friends and acquaintances, although friendships have been made. Rather they embody a deep and broad scholarly community. The difference from compiled Internet resources is the built-expertise and intellectual engagement of the authors. The summary of the developments and futures in their personal areas of scholarship have been filtered through their years of experience, both as scholars and communicators. Quality, then, is endemic to each piece, developed through this top-down identification of expertise, and made indelible by the bottom-up application of high standards from people leading the sub-domains – the authors, and the domains themselves, the section editors.

On a procedural level, the publishers early committed to the notion that this Encyclopedia would also be an online resource, and access would be available through print, for those with strong bookcases and the persisting love of turning real pages. The Internet version will allow multiple prisms through which the reader may access articles and provide, as it were, an emulation of the Internet in our field, albeit bounded by expertise and high quality.

What must be underscored in the assessment of this effort are the Editors-in-Chief and the publishers’ commitment to find excellence worldwide. We tried very hard to persuade notable scholars from all parts of the world to make contributions. Less than to fulfill the title of “International,” we were on the hunt for perspectives that would enrich the scope and depth of the sections. Our section editors put in enormous time attempting to find the best in the field, wherever they resided. Yet, not everyone is in the volume. Some were overcommitted. Many were not fully confident of their English, and the automated translation software has not yet met standards for technical writing. We believe that such writing and editing tools will make the outreach to an even broader International group of scholars possible in future revisions, or online updates. Furthermore, the birth of the World Educational Research Association (in 2009) will provide a better set of interlocking networks to find and evaluate scholarship from any place on the globe.

Finally, the scope of the effort must be acknowledged: 28 section editors, 926 articles were commissioned, drafted, reviewed, redrafted, edited, and put together in the space of four years. The publishers underwent some internal changes, and alterations in management. We as Editors-in-Chief, changed roles, moved, and also had to keep our own research and development enterprises afloat. Deadlines wobbled; authors dropped from view and had to be replaced.

Yet, at times frustrating as all development is, we find the final product exhilarating. We are enthusiastic not simply because it came into being at all, but because the collective light of the minds that wrote have left a bright resource for the future, one that will impact the way our colleagues understand and experience the educational knowledge, improvement, and impact in the future.

Eva Baker
Penelope Peterson

EDITORS

EDITORS-IN-CHIEF

Penelope Peterson

School of Education and Social Policy
Northwestern University
Evanston, IL
USA

Eva Baker

UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

Barry McGaw

McGaw Group Pty Ltd
Mont Albert Nth
Melbourne, VIC
Australia

SECTION EDITORS

Walter Allen

HEOC Division
University of California, Los Angeles
Los Angeles, CA
USA

Vibeke Grøver Aukrust

Institute for Educational Research
University of Oslo
Oslo
Norway

Dominic J Brewer

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

Kai-Ming Cheng

Faculty of Education
The University of Hong Kong
Hong Kong

Gregory Chung

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Penny Enslin

Department of Educational Studies
University of Glasgow
Glasgow
UK

Noelle Griffin

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Dennis Hocevar

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

David Hopkins

Institute of Education
University of London
London
UK

Mary James

Faculty of Education
University of Cambridge
Cambridge
UK

Sanna Järvelä

Department of Educational Sciences and
Teacher Education
University of Oulu
Oulu
Finland

Thomas Karmel

National Centre for Vocational Education Research
Adelaide, SA
Australia

Harold G Levine

School of Education
University of California, Davis
Davis, CA
USA

Robert L Linn

University of Colorado at Boulder
Ouray, CO
USA

Allan Luke

Faculty of Education
Queensland University of Technology
Brisbane, QLD
Australia

Ann M Mastergeorge

Human Development and Family Studies
University of California, Davis
Davis, CA
USA

Patrick J McEwan

Department of Economics
Wellesley College
Wellesley, MA
USA

Barbara Means

Center for Technology in Learning
SRI International
Menlo Park, CA
USA

Ingrid Moses

University of Canberra
Canberra, ACT
Australia

Bill Mulford

Faculty of Education
University of Tasmania
Hobart, TAS
Australia

Yu Jing Ni

Faculty of Education
The Chinese University of Hong Kong
Hong Kong

Festus Obiakor

Department of Exceptional Education
University of Wisconsin-Milwaukee
Milwaukee, WI
USA

P David Pearson

Graduate School of Education
University of California, Berkeley
Berkeley, CA
USA

Colin Power

Graduate School of Education
University of Queensland
Brisbane, QLD
Australia

Kjell Rubenson

Faculty of Education
The University of British Columbia
Vancouver, BC
Canada

Sandip Sinharay

Education Testing Services
Princeton, NJ
USA

Collette Tayler

Queensland University of Technology
Brisbane, QLD
Australia

Nico Verloop

ICLON, Graduate School of Education
Leiden University
Leiden
Netherlands

HOW TO USE THE ENCYCLOPEDIA

The International Encyclopedia of Education is intended for use by students, research professionals, and interested others. Articles have been chosen to reflect major disciplines in the study of education and common topics of research by academics in this domain. Each article serves as a comprehensive overview of a given area, providing both breadth of coverage for students, and depth of coverage for research professionals. We have designed the encyclopedia with the following features for maximum accessibility for all readers.

The contents of the encyclopedia are arranged alphabetically by section, and within sections, alphabetically by article. The Subject Index is located in Volume 8. Some topics are covered in a multitude of articles from differing perspectives, while other topics may have only one entry. We encourage use of the index for access to a subject area, rather than use of the Contents list alone, so that a reader has a full notion of the coverage of that topic.

The articles include cross-references to other related encyclopedia articles, suggested further readings where applicable, and many contain relevant websites for additional information. We encourage readers to use the cross-references to locate other encyclopedia articles that will provide more detailed information about a subject.

The Further Reading sections include recent secondary sources to aid the reader in locating more detailed or technical information. Review articles and research articles that are considered of primary importance to the understanding of a given subject area are also listed. These suggested further readings are not intended to provide a full reference listing of all material covered in the context of a given article, but are provided as next steps for a reader looking for additional information.

This page intentionally left blank

CONTENTS

Preface	v–vi
Editors	vii–viii
How to use the Encyclopedia	ix

VOLUME 1

A

Adult Education

Adult Education Overview	<i>K Rubenson</i>	1
--------------------------	-------------------	---

Adult Education – Adult Learning, Instruction and Program Planning

Adult Learning	<i>S B Merriam</i>	12
Adult Learning and Instruction: Transformative Learning Perspectives	<i>P Cranton</i>	18
Adult Learning in a Biographic Perspective	<i>L West</i>	25
Adult Learning, Instruction and Programme Planning: Insights from Freire	<i>P Mayo</i>	31
Characteristics of Adult Learning	<i>K Illeris</i>	36
Informal Learning: A Contested Concept	<i>P Hodgkinson</i>	42
Organizational Learning	<i>P-E Ellström</i>	47
Program Planning	<i>A L Wilson and R M Cervero</i>	53
Workplace Learning Frameworks	<i>S Billett</i>	58

Adult Education – as a Field of Study

Class Analysis in Adult Education	<i>T Nesbit</i>	64
Gender Analysis	<i>J Stalker</i>	70
Race and Ethnicity in the Field of Adult Education	<i>J Johnson-Bailey and D Drake-Clark</i>	77
Trends in Workplace Learning Research	<i>T Fenwick</i>	83

Adult Education – Concepts

Lifelong Learning	<i>J Field</i>	89
Rewriting the History of Adult Education: The Search for Narrative Structures	<i>B J Hake</i>	96

Adult Education – Domains and Provision

Adult Basic Education: A Challenge for Vocational Based Learning	<i>R Höghelm</i>	102
Adult Literacy Education	<i>L Tett and R St.Clair</i>	107
Citizenship and Immigrant Education	<i>D Schugurensky</i>	113
Community Based Adult Education	<i>B Connolly</i>	120
Continuing Professional Education: Multiple Stakeholders and Agendas	<i>R M Cervero and B J Daley</i>	127
Financing of Adult and Lifelong Learning	<i>Å Sohlman</i>	133
Labor Education	<i>K Forrester</i>	139
Learning Cities and Regions	<i>C Duke</i>	144
Museums as Sites of Adult Learning	<i>R S Grenier</i>	150
Overview of Lifelong Learning Policies and Systems	<i>D-B Kwon and C H O Daeyeon</i>	156
Popular Adult Education	<i>S Tøsse</i>	162
Provision of Prior Learning Assessment	<i>P Andersson</i>	169
The Age of Learning: Seniors Learning	<i>P Jarvis</i>	175
University Adult Continuing Education: The Extra-Mural Tradition Revisited	<i>M Slowey</i>	183

Adult Education – Economy and Society

Adult Education and Civil Society	<i>J Crowther and I Martin</i>	191
Adult Education and Nation-Building	<i>S Walters</i>	198
Economic Outcomes of Adult Education and Training	<i>A Ferrer and W C Riddell</i>	204
Health and Adult Learning	<i>P Bélanger and M Robitaille</i>	211
Modernization Processes and the Changing Function of Adult Learning	<i>H S Olesen</i>	217
The Political Economy of Adult Education	<i>G Rees</i>	223
Wider Benefits of Adult Education	<i>R Desjardins and T Schuller</i>	229

Adult Education – Participation

Barriers to Participation in Adult Education	<i>K Rubenson</i>	234
Participation in Adult Learning	<i>R Desjardins</i>	240

C**Comparative Education**

Comparative and International Education: English-Language Scholarly Publications	<i>D Post and M Fedri</i>	251
Comparative Education: Philosophical Issues and Concepts	<i>M Mason and C W Evers</i>	257
Comparative Education: Societies and Associations	<i>M Bray</i>	266

Curriculum Development

Curriculum and Religion	<i>C Kapitzke and P Graham</i>	272
Gender and Curriculum	<i>J Marsh</i>	277
Local Teacher-Based Curriculum Development	<i>M D Osborne and H B Barker</i>	283

Curriculum Development – Contexts

Curriculum and Human Rights	<i>C Lohrenscheit</i>	287
Curriculum and the Education of Cultural and Linguistic Minorities	<i>S May</i>	293
Curriculum in Postcolonial Contexts	<i>A Hickling-Hudson</i>	299
Curriculum, Economic and Cultural Development	<i>A Rampal</i>	307
Globalization and Curriculum	<i>R Dale</i>	312
Popular Culture	<i>K Schultz and R Throop</i>	318

Curriculum Development – Evaluation and Research

An Overview of Research in Curriculum Inquiry	<i>F M Connelly and S J Xu</i>	324
Curriculum Evaluation: Approaches and Methodologies	<i>V Klenowski</i>	335
Curriculum Studies, Discourse Analysis, and the Construction of Historical Time	<i>J Laspina</i>	342
Curriculum Use in the Classroom	<i>M Ben-Peretz and B Eilam</i>	348
The Roots and Routes of Teacher-Based Action Research and Curriculum Inquiry: An Historical Perspective	<i>L G Putney and J L Green</i>	355

Curriculum Development – Planning and Development

Curriculum and Syllabus Design	<i>A Woods, A Luke, and K Weir</i>	362
Curriculum and the Publishing Industry	<i>J Larson, A-R Allen, and D Osborn</i>	368
Curriculum and Teacher Change	<i>J G Ladwig</i>	374
Curriculum Governance and Planning	<i>B Levin</i>	379
Curriculum Planning and Systems Change	<i>Z Deng</i>	384
Curriculum, Digital Resources and Delivery	<i>Y Zhao, G Zhang, and C Lai</i>	390
Textbook Development and Selection	<i>P Shannon</i>	397

Curriculum Development – Subject Matter

Curriculum and Globalization: Higher Education	<i>K H Mok</i>	403
Curriculum Development as Subject Matter: Social Studies	<i>J Brophy and J Alleman</i>	411
Curriculum Development in the Area of Reading	<i>C L Madda and T E Raphael</i>	415
Curriculum in the Arts	<i>C M Thompson, L Bresler, and T E Costantino</i>	423
Curriculum: Foreign Language Learning	<i>Angel M Y Lin</i>	428
First Language Learning	<i>T Rogers and L Moffatt</i>	435
Learning a Second Language in First Language Environments	<i>E Bernhardt and D J Tedick</i>	441
Literature	<i>S Hynds</i>	447
Mathematics	<i>B Kaur and C P Vistro-Yu</i>	453
Physical Education and Sports	<i>D Kirk</i>	459
Primary and Elementary/Middle Grades Reading	<i>D E Alvermann and C A Mallozzi</i>	464
Science	<i>Y Kali and M Linn</i>	468

Curriculum Development – Theory

Bilingual Learning (Learning L1 and L2 in an L1 and L2 Environment)	<i>D Schwinge</i>	475
Curriculum and Complex Systems Theory	<i>W-M Roth and J Thom</i>	481

Curriculum and Constructivism	<i>B Davis and D Sumara</i>	488
Curriculum and Critical Theory	<i>L Yates</i>	494
Curriculum and Poststructuralist Theory	<i>J L Miller</i>	499
Curriculum and Structuralist Sociology: The Theory of Codes and Knowledge Structures	<i>J Muller and J Gamble</i>	505
Curriculum Reform	<i>R Gilbert</i>	510
Early Childhood Curriculum and Developmental Theory	<i>S M-Y Lim and C Genishi</i>	514

D

Demography and Social Change

An Evaluation Framework for E-Learning Effectiveness in the Arab World	<i>W AbuSneineh and M Zairi</i>	521
Globalization and Social Justice in Higher Education	<i>E Kimura-Walsh</i>	536
Racial Inequality and the Economic Payoffs of Higher Education	<i>S A Lachtman</i>	542
School Health Education	<i>J S Yang</i>	547

Demography and Social Change – Demography

A Paradigm of Contradictions: Racism and Science Education	<i>B A Brown and J W Mutegi</i>	554
Africana Studies: Past, Present And Future	<i>B R Hare, A Young, J B Stewart, and D P Aldridge</i>	565
Beyond Black Nationalism to Black Internationalism: The Compelling Case for Examining Black Educational Challenges Globally	<i>K Freeman</i>	573
Beyond Racial, Ethnic, and Gender Bias in Education Statistics	<i>J T Lott and C Bennett</i>	581
Caste and Education in India	<i>G G Wankhede</i>	589
Demography, Community, and the Education of Asian American and Pacific Islander Students	<i>R T Teranishi</i>	595
Educational Attainment and Mortality Differentials	<i>H D Horton, L Hixson, and S M Monnat</i>	602
Educational Opportunity and Latino/Chicano College Choice: New Findings, and Theoretical Perspectives	<i>M A Ceja and P A Pérez</i>	610
Ethnic Minority Identity and Educational Outcomes in a Rising China	<i>G A Postiglione</i>	616
Ethnic Studies and Women's Studies: From the Past to the Present, Providing a Space in the Academy for Scholarship Focused on the Histories and Experiences of Marginalized People	<i>C A Brown</i>	623
Gender Equity in Higher Education: Challenges and Celebrations	<i>L Morley</i>	629
Locating Space and Place in the College Access Debate: New Tools for Mapping and Understanding Educational Inequity and Stratification	<i>B Cantwell and J F Milem</i>	636
Public Policy and Inequality in Postsecondary Opportunity: Educational Statistics and the Failure of Education Reform	<i>E P St. John, K Williams, and K Moronski</i>	649
The Education of Indigenous Students	<i>J W Tippeconnic and S C Faircloth</i>	661
The Health Advantages of Educational Attainment	<i>G Y Dinwiddie</i>	667
To Go or Not to Go: Access and Barriers to Ethnic Minority Participation in Higher Education within the United Kingdom	<i>P Brug and T L Oliver</i>	673
Universal Secondary Education and the Two-Tiered School System in the English-Speaking Caribbean	<i>P Anderson and J Devonish</i>	679
Women of Color in Science, Technology, Engineering and Mathematics (STEM): Refining the Concepts, Reframing the Issues	<i>C B Leggon</i>	686

Demography and Social Change – Social Change

Educational Diversity in U.S. Law Schools	<i>C E Daye, A T Panter, W R Allen, M E Deo, and L F Wightman</i>	691
Educational Reform	<i>F Lopez-Segrera</i>	698
Emergence of For-profit Higher Education	<i>X Li</i>	705
Engaged Education: Experiential Learning, Intensive Field Experiences, and Social Change	<i>T R Rone</i>	711
Higher Education and Social Change in Mexico	<i>A Mungaray, G Estrella, P Moctezuma, and M Ramírez-Urquidy</i>	719
The Changing Role of the State in Higher Education	<i>K Gürüz</i>	728
The Role of the State in School Reform: Responding to Increasing Ethnic Diversity in Britain	<i>P Ratcliffe</i>	737
Transforming Mindsets Through Education for Sustainable Development	<i>V M Segovia</i>	746

Demography and Social Change – Sociology

Achievement Motivation in Ethnic Minority Youth	<i>S Graham and E Bryant</i>	753
Affirmative Action and Higher Education in Brazil	<i>J T dos Santos and D M Queiroz</i>	760
Black in White: Black Students at White and Black Colleges	<i>L Evans and J R Feagin</i>	767
Bologna Process: On the way to a Common European Higher Education Area	<i>A V Rezaev</i>	772
Critical Race Theory in Education Research	<i>R Vue and C B Newman</i>	779
Cultural Diversification and Japanese Education: Social Constructions of the New Diversity	<i>R Tsuneyoshi</i>	787
Education in War Countries: A Disempowerment of Society	<i>J L Nkata</i>	793
Educational Attainments of U.S. Black Males and Females: 1971 to 2003	<i>F B Bonner, M-C Jipguep-Akhtar, and R J Harrison</i>	799
Educational Preparation: Fostering the Self-Efficacy and Resilience of Urban Adolescent Youth	<i>M B Spencer and B Tinsley</i>	806
HIV Education for Low-Literate People: Transforming Students and Communities through Paulo Freire's Praxis and the Pedagogy of Action	<i>N Z Haniff</i>	814
How Are Discussions of Interdisciplinary Studies Linked to Diversity Discourse?	<i>C N Spence</i>	823
Identifying Patterns of Doctoral Attrition Across Academic Fields	<i>A James, K J Fosnacht, E B Stolzenberg, and D Thirumalai</i>	827
Male Underachievement in Education Across the Globe: A Shift in Paradigm for Gender Disparities Regarding Academic Achievement	<i>J F L Jackson, J L Moore III, and R A Leon</i>	838
Noncognitive Measures for Higher Education Admissions	<i>W E Sedlacek</i>	845
Peace Education	<i>J Page</i>	850
Remote Control: Africa's Development and Outflow to Developed Nations of Technical Talent	<i>F Haniff</i>	855
Scholar-Baller®: Student Athlete Socialization, Motivation and Academic Performance in American Society	<i>C K Harrison, S Bukstein, J Mottley, E Comeaux, J Boyd, C Parks, and D Heikkinen</i>	860
Service-Learning and Diversity as Mediums to Foster Civic and Community Engagement in Undergraduates	<i>L Oseguera</i>	866
Success and Failure in Education and Criminal Justice: Identifying Common Mechanisms	<i>S Ramirez and A Harris</i>	874
The History of Education: Race and Education	<i>J O Jewell</i>	881

VOLUME 2**E****Early Childhood Care and Education Theory**

- Early Childhood Care and Education: The Family, The Market, and The State *S B Kamerman and M J Neuman* 1
- Gender Issues in Early Childhood Education and Care *B Pocock* 11

Early Childhood Education and Care Governance

- Early Childhood Risk, Protection and Abuse Prevention *E Khoo* 17
- Evaluating Early Childhood Education and Care Programs *M Janus and S Brinkman* 25

Early Childhood Education and Care Investment

- Investing in Early Childhood Education and Care: The Health and Wellbeing Case *T Moore and F Oberklaid* 32
- Cost-effective Early Childhood Programs from Preschool to Third Grade *A J Reynolds, J A Temple, and B White* 38
- Early Childhood in Post-Modern Cultures: Thoughts and Some Concerns *P Gammage* 49
- Investing in Early Childhood Education and Care: Some Policy Implications *J Bennett* 55
- Investing in Early Childhood Education and Care: The Economic Case *G Cleveland and M Krashinsky* 63

Early Childhood Education and Care Programs

- Early Care and Education Programs for Infants and Toddlers *B Nyland* 69
- Literacies in Early Childhood: The Preschool Period *M Flear and B Raban* 75
- Participation in Early Childhood Education and Care Programs: Equity, Diversity and Educational Disadvantage *M Vandenbroeck* 81
- Teaching in Early Childhood Centers Instructional Methods and Child Outcomes *I Siraj-Blatchford* 86

Early Childhood Education and Care Provision

- Early Childhood Education and Care in the People's Republic of China *Zeng Xiaodong* 93
- Home-Based and Institutional Early-Childhood Education and Care Services *R Fukkink and H Blok* 98
- Integrated Early Childhood Education and Care Services – Care, Upbringing, Education and Health *S Colley* 104
- Issues of Access and Program Quality *J T Hustedt and W S Barnett* 110
- Kindergarten Transitions and Linkages to Primary School – Readiness Re-conceptualized *A Petriwskyj* 120
- Parent Support in Early Childhood – Approaches and Outcomes *K S Howard, K A Beckmann, and J Brooks-Gunn* 126

Early Childhood Education and Care Research

- Development and Implementation of Early Learning Standards in the United States *C Scott-Little* 132
- Global Status of Early Learning and Development Standards *P R Britto and S L Kagan* 138
- Rethinking Early Childhood Education and Care: Implications for Research and Evaluation *K Thorpe, D Cloney, and C Tayler* 144

Early Childhood Education and Care Theory

- Child Rearing and Early Education: Parents and Professionals: Theoretical and Historical Influences *H Penn* 151

Early Childhood Education and Care Workforce

- Early Childhood Education and Care Professional Profiles in the United Kingdom: Multiagency Teamwork and Transdisciplinary Challenges *A Anning and P Calder* 157
- Early Childhood Education and Care Staff Support: Mentoring and Professional Development in Finland *P Lindberg* 162

Economics of Education

- Education and Health *P Muennig* 169
- Data *M Perez and M Socias* 178
- Empirical Research Methods in the Economics of Education *P J McEwan* 187
- Theoretical Concepts in the Economics of Education *D J Brewer, G C Hentschke, and E R Eide* 193

Economics of Education – Costs and Financing of Schooling

- Cost–Benefit Analysis and Cost–Effectiveness Analysis *C Belfield and H M Levin* 199
- Economic Approaches to Adequacy *J Imazeki* 204
- Economic Approaches to School Efficiency *L L Taylor* 210
- School Finance Reform *T Downes* 216
- School Finance: An Overview *J K Rice, D H Monk, and J Zhang* 222
- The Economics of Tuition and Fees in Higher Education *R G Ehrenberg* 229

Economics of Education – External Benefits of Schooling

- Education and Civic Engagement *T S Dee* 235
- Education and Crime *L Lochner* 239
- Education and Economic Growth *E A Hanushek and L Wößmann* 245
- Parental Socioeconomic Status, Child Health, and Human Capital *J Currie and J Goodman* 253
- The External Benefits of Education *W W McMahon* 260

Economics of Education – Labor Markets, Education and Earnings

- Education and Inequality *J Blanden and S Machin* 272
- Human Capital *E R Eide and M H Showalter* 282
- Race Earnings Differentials *M Carnoy* 288
- Returns to Education in Developed Countries *M Gunderson and P Oreopoulos* 298
- Returns to Education in Developing Countries *H A Patrinos and G Psacharopoulos* 305
- School Quality and Earnings *J R Betts* 313
- Signaling in the Labor Market *M E Page* 321
- The Economics of High School Dropouts *R W Rumberger* 325

Economics of Education – Markets, Incentives and School Choice

- Competition and Student Performance *A A Payne* 331
- Educational Privatization *C Belfield and H M Levin* 337

Student Incentives	<i>E R Eide and M H Showalter</i>	342
The Economic Role of the State in Education	<i>D N Plank and T E Davis</i>	347
The Economics of Catholic Schools	<i>W Sander and D Cohen-Zada</i>	354
The Economics of Charter Schools	<i>R Zimmer and R Buddin</i>	360
The Economics of Parental Choice	<i>R Chakrabarti and J Roy</i>	367
The Economics of School Accountability	<i>D N Figlio and H F Ladd</i>	374
The Efficacy of Educational Vouchers	<i>R Zimmer and E P Bettinger</i>	380
Tiebout Sorting and Competition	<i>T J Nechyba</i>	388

Economics of Education – Production of Schooling

Desegregation, Academic Achievement and Earnings	<i>S Rivkin</i>	394
Education Production Functions: Concepts	<i>D N Harris</i>	402
Education Production Functions: Developed Country Evidence	<i>E A Hanushek</i>	407
Education Production Functions: Evidence from Developing Countries	<i>P Glewwe and S Lambert</i>	412
Family Environment in the Production of Schooling	<i>R Rothstein</i>	423
Neighborhoods and Peers in the Production of Schooling	<i>J Vigdor and J Ludwig</i>	431
Teacher Quality in Education Production	<i>I Pelayo and D J Brewer</i>	438
The Economics of Class Size	<i>D W Schanzenbach</i>	443
The Economics of Early Childhood Interventions	<i>M Nores</i>	450

Economics of Education – Teaching Labor Markets

An Overview of Teacher Labor Markets	<i>H Lankford and J Wyckoff</i>	456
Compensating Differentials in Teacher Labor Markets	<i>J Chambers</i>	465
Economic Approaches to Teacher Recruitment and Retention	<i>S Loeb and J Myung</i>	473
Teacher Incentives	<i>L Santibañez</i>	481
Teacher Supply	<i>P Dolton</i>	489
Teacher Training and Preparation in the United States	<i>M Podgursky, M G Springer, and R Hutton</i>	498
Teachers in Developing Countries	<i>F H Rogers and E Vegas</i>	504
The Economics of Teachers' Unions in the United States	<i>K Strunk</i>	512

Education of Children with Special Needs

Adapted Physical Activity for Students with Special Needs	<i>S R Hodge</i>	519
Assessment Accommodations for Children with Special Needs	<i>R J Kettler and S N Elliott</i>	530
Assistive Technology and Educational Practice	<i>H P Parette and G R Peterson-Karlan</i>	537
Augmentative–Alternative Communication	<i>C Mason and K M Chinn</i>	544
Community Integration and Employment of Youth with Special Needs	<i>B A Ford and A W Daviso III</i>	550
Cooperative Learning for Children with Special Needs	<i>E A Dooley and K R Kossar</i>	555
Co-Teaching	<i>E D Cramer</i>	560
Cultural and Linguistic Diversity and Children with Special Needs	<i>J M Patton, N Day-Vines, and A McSwain</i>	565
Curriculum-Based Assessment and Students with Special Needs	<i>D J O'Shea and J D Katsafanas</i>	571

Diagnosis and Classification	<i>M Sapp</i>	577
Direct Instruction and the Education of Children with Special Needs	<i>T Mehring</i>	584
Disabilities and Gifted Education	<i>V I Daniels and M J McCollin</i>	588
Early Childhood Special Education: Birth to Eight Years Old	<i>T S Lovelace, R Gardner III, J B Axe, and D M Sainato</i>	594
Early Intervention for Students with Special Needs	<i>P A Grant</i>	608
Early Transition of Children with Special Needs	<i>S A Fowler</i>	613
Educating Children Who are Deaf-Blind	<i>J Wood and K M Chinn</i>	621
Educating Children with Deafness and Hearing Impairments	<i>G M Mukuria and C J Eleweke</i>	628
Educating Children with Other Health Impairments (ADHD)	<i>G M Mukuria and F E Obiakor</i>	634
Educating Children with Physical Disabilities	<i>Z P Nkabinde, F E Obiakor, M T Ofor, and D J Smith</i>	640
Educating Children with Specific Learning Disabilities	<i>P A Grant and P B Grant</i>	646
Educating Children with Speech and Language Disorders	<i>S L Green and C D Qualls</i>	654
Educating Students with Autism and Related Disorders	<i>T T Dyches</i>	661
Educating Students with Cognitive Disabilities	<i>J P Bakken</i>	669
Educating Students with Emotional and Behavioral Disorders	<i>S W Smith and G G Taylor</i>	678
Educating Students with Special Needs: An Overview	<i>F E Obiakor, M K Harris, P Harris-Obiakor, and G M Mukuria</i>	688
Educating Students with Traumatic Brain Injury	<i>S Burkhardt and A Rotatori</i>	695
English Language Learners with Special Needs	<i>R Rueda and G Ragusa</i>	701
Functional Behavioral Assessment	<i>J P Bakken</i>	708
History of Special Education	<i>L M Garrick Duhaney and S J Salend</i>	714
Inclusion of Students with Special Needs in General Education Classrooms	<i>M A Prater</i>	721
Instructional Accommodations for Children with Special Needs in Inclusive Settings	<i>A L Yurick, S Musti-Rao, and G Cartledge</i>	727
Interagency Agreements	<i>L A Owens</i>	733
Interdisciplinary Collaboration in Early Childhood	<i>N File</i>	739
Language Development in Children with Special Needs	<i>M S L Stewart</i>	745
Large Scale Assessment and Accountability and Students with Special Needs	<i>M L Thurlow</i>	752
Literacy Instruction for Students with Special Needs	<i>L H Mason</i>	759
Math Instruction for Children with Special Needs	<i>B A Bottge</i>	767
Overrepresentation of Culturally and Linguistically Diverse Students in Special Education	<i>S C Trent</i>	774
Paraprofessionals in Special Education	<i>B Y Ashbaker, S C Dunn, and J Morgan</i>	780
Parent and Family Involvement in the Education of Children with Special Needs	<i>L P Prater</i>	788
Peer Relations and Socialization of Children and Adolescents with Special Needs	<i>L M Bullock, M Wong-Lo, and R A Gable</i>	794
Peer-Tutoring	<i>L J O'Shea and D J O'Shea</i>	802
Postsecondary Participation of Students with Special Needs	<i>R A Stodden and K D Roberts</i>	808
Prereferral Strategies	<i>C G Simpson and J P Bakken</i>	815

Preventing Antisocial Behavior and Delinquency: A Comprehensive School-wide Approach <i>P R Hallam, K R Young, P Caldarella, D G Wall, and L Christensen</i>	820
Programs and Instructional Strategies for Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	829
Related Services for Children with Special Needs <i>S O Obi</i>	834
Risk and Resilience Frameworks in Understanding Special Education <i>W-Y Hsieh and M L Donahue</i>	840
School Dropout Prevention <i>F E Obiakor</i>	845
School-Based Services for Children with Special Needs <i>S O Obi</i>	850
Self-Determination <i>M L Wehmeyer</i>	857
Social Competence <i>L A Owens and S Johnston-Rodriguez</i>	865
The Identification of Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	870
Transition from School to Adult Life <i>M J McCollin and F E Obiakor</i>	876
Visual Impairment <i>W Sapp</i>	880
Workforce Issues in Special Education <i>E Carlson and B Billingsley</i>	886

VOLUME 3

Education of Professionals

An Overview of Accreditation, Certification, and Licensure Processes <i>J J Norcini and M van Zanten</i>	1
Educating America's Military Officers <i>M H Adamshick</i>	7
Education for Architecture in the United States and Canada <i>S Schreiber</i>	13
Education for Medicine <i>M B Anderson</i>	19
Education for the Legal Profession <i>J W Wegner</i>	25
Education of Clergy <i>D O Aleshire</i>	31
Framing Professions Education <i>P S O'Sullivan</i>	37
Issues in Accounting/Business Education <i>W H Gijsselaers and R G Milter</i>	39
Professions – Engineering <i>J Heywood</i>	45
Professions – Health-Related <i>G M Jensen and C B Royeen</i>	51
Professions – Pharmacy <i>D E Beck</i>	58
Teaching for the Complex Practice of Nursing <i>P Benner and M Sutphen</i>	65
Veterinary Education <i>K M Edmondson</i>	71

Education Research Methodology: Quantitative Methods and Research

Abductive Research Methods <i>B D Haig</i>	77
Field Experimentation in Education <i>M A Constanas</i>	83
Internal Validity <i>M H Clark and S C Middleton</i>	90
Methods for Approximating Random Assignment <i>B Schneider and S-K McDonald</i>	97
Methods for Approximating Random Assignment: Regression Discontinuity and Propensity Scores <i>L M Scheier</i>	104

Multitrait–Multimethod Designs	<i>M Bong and S Hong</i>	111
Multivariate Longitudinal Data Analysis	<i>H Liu and L Chang</i>	120
Quasi-Experimentation: Two Group Design	<i>E Hong</i>	128
Regression Discontinuity Designs	<i>A Kelava, S Rohrmann, and V Hodapp</i>	134
Sampling	<i>R S Brown</i>	142
Statistical Conclusion Validity	<i>P L Busk</i>	147
Survey Research Methods	<i>A Fink</i>	152

Educational Assessment

An Overview of Educational Assessment	<i>M James</i>	161
Assessment and the Evaluation of Institutional Effectiveness	<i>S M Thomas</i>	172
Assessment in Schools – Dispositions	<i>R D Crick</i>	181
Assessment in Schools Related To Literacy: Reading	<i>K J Topping</i>	189
Impact of Assessment on Students' Learning Strategies and Implications for Judging Assessment Quality	<i>M Segers, F Dochy, and D Gijbels</i>	196

Educational Assessment – Assessment Activities

Alternative Assessment	<i>G Stobart and C Gipps</i>	202
Classroom Assessment Tasks and Tests	<i>J Phelan and J Phelan</i>	209
Dynamic Assessment	<i>J G Elliott, E L Grigorenko, and W C M Resing</i>	220
Instructional System Provided Feedback	<i>H F O'Neil, S Chuang, and T Huang</i>	226
Peer and Self-assessment	<i>C Harrison</i>	231
Portfolio Assessment	<i>V Klenowski</i>	236
Records of Achievement: Beyond Traditional Tests	<i>P Broadfoot</i>	243

Educational Assessment – Assessment in Domains

Assessment in Higher Education	<i>D R Sadler</i>	249
Assessment in Schools – Affective Domain	<i>M Ventura, R D Roberts, and M Zeidner</i>	256
Assessment in Schools – Creative Subjects	<i>T E Costantino and L Bresler</i>	262
Assessment in Schools – Learning to Learn	<i>A Hautamäki, J Hautamäki, and S Kupiainen</i>	268
Assessment in Schools – Literacy Writing (Extended)	<i>J M Parr and L Earl</i>	273
Assessment in Schools – Mathematics	<i>M Brown and J Hodgen</i>	279
Assessment in Schools – Oracy	<i>A L Bailey</i>	285
Assessment in Schools – Primary Science	<i>W Harlen</i>	293
Assessment in Schools – Secondary Science	<i>J Shemwell, A Fu, M Figueroa, R Davis, and R Shavelson</i>	300
Assessment in Schools – Technology Education and ICT	<i>A Jones, B Cowie, and J Moreland</i>	311
Assessment in the Early Years	<i>M J Drummond</i>	316
Assessment in the History, Civics and Social Studies Domains	<i>P G Avery</i>	323
Assessment in the Workplace of Performance, Developing Expertise and Competence	<i>A J Brown</i>	330
Assessment in Vocational Education	<i>K Ecclestone</i>	337

Educational Assessment – Concepts and Issues

Assessing Group Work	<i>M Galton</i>	342
Assessment and the Regulation of Learning	<i>L Allal</i>	348
Challenges of Developing and Implementing Formative Assessment Practices in Schools	<i>R Sutton</i>	353
Formative Assessment	<i>P Black</i>	359
Formative Assessment and Instructional Planning	<i>B Marshall</i>	365
Progression and Assessment: Developmental Assessment	<i>M Forster and G Masters</i>	369
Students' Rights in Assessment and Evaluation	<i>J Gardner and B Holmes</i>	378
Summative Assessment by Teachers	<i>R Daugherty</i>	384
The Multiple Purposes of Assessment	<i>P E Newton</i>	392
The Relationship between Assessment and the Organization and Practice of Teaching	<i>L B Resnick, D William, R Apodaca, and E S Rangel</i>	397

Educational Assessment – Enhancing Quality and Use

Assessment Conversations: Reading and Writing Conferences with Students and with Parents	<i>M Barrs</i>	403
Assessment: Pre-service and In-service Teacher Education	<i>A Davies</i>	410
Classroom Assessment in Policy Context (Australia)	<i>J J Cumming</i>	417
Classroom Assessment in Policy Context (England and France)	<i>M Osborn and M Raveaud</i>	425
Classroom Assessment in Policy Context (French Sources)	<i>D Laveault</i>	432
Classroom Assessment in Policy Context (Hong Kong)	<i>D Carless</i>	438
Classroom Assessment in Policy Context (New Zealand)	<i>T J Crooks</i>	443
Classroom Assessment in Policy Context (Sub-Saharan Africa)	<i>S J Howie, L Zimmerman, and K Draper</i>	449
Moderation of Student Work by Teachers	<i>G S Maxwell</i>	457
School Policies and Practices to Support Effective Classroom Assessment for Learning	<i>D Pedder</i>	464
Student Assessment: Policy and Practice in Eastern Europe	<i>G Bethell</i>	472

Educational Assessment – Key Relationships

Assessment Practice in Policy Context: Latin American Countries	<i>F M Rizo</i>	479
Impact of Assessment on Learner Groups (Boys/Girls)	<i>J Elwood</i>	486
Impact of Assessment on Learner Groups (Disabilities)	<i>M Rouse</i>	492
Impact of Assessment on Students' Test Anxiety	<i>P Buchwald and C Schwarzer</i>	498
Impact of Assessments on Classroom Practice	<i>J L Herman</i>	506
Social Practices in School Assessment and their Impact on Learner Identities	<i>A Pollard and A Filer</i>	512

Educational Evaluation – Concepts, Practice, and Future Directions

Educational Evaluation: Concepts, Practice, and Future Directions	<i>Y J Ni</i>	518
---	---------------	-----

Educational Evaluation – Contemporary Themes in Evaluation

Evaluating E-Learning	<i>D D Williams and C R Graham</i>	530
Evaluating Schools as Learning Communities	<i>S M Thomas</i>	539
Evaluation of Adult Education and Training Programs	<i>G McNamara, P Joyce, and J O'Hara</i>	548
Evaluation of Integrated Health Programs in School	<i>T M Franke</i>	555

Formative Assessment in Teacher Education and Teacher Professional Development	<i>H Tillema</i>	563
Informal Education and Evaluation	<i>T Jeffs</i>	572
Methods to Evaluate Technology	<i>W L Bewley</i>	578

Educational Evaluation – Contexts of Educational Evaluation

Cost Analysis in Evaluation Studies	<i>J D Fletcher</i>	585
Cultural Issues that Can Affect the Validity of Educational Evaluations	<i>R P Durán</i>	592
Evaluation and Accountability	<i>E Sato and S N Rabinowitz</i>	600
Evaluation Governance and Planning	<i>R Picciotto</i>	607
Evaluation Use	<i>V G Thomas</i>	614
Moral and Ethical Issues in Evaluation	<i>M Morris</i>	622
The Role of Stakeholders in Educational Evaluation	<i>S Taut and M C Alkin</i>	629

Educational Evaluation – Evaluation Domains

Curriculum Evaluation	<i>T Levin</i>	636
Evaluating Education in Three Policy Eras	<i>M A Peters, R Britez, and S M Weber</i>	645
Evaluation of Mathematics Education Programs	<i>J Cai</i>	653
Evaluation of Science Education Programs	<i>G Berberoğlu</i>	660
Evaluation of Social Studies and Civic Education Programs	<i>S G Grant</i>	667
Evaluation of Teacher Quality and Practice	<i>W Robinson and J Campbell</i>	674
International Evaluations	<i>A Schleicher</i>	681
National Assessment Programs: The Example of the U.S. National Assessment of Educational Progress	<i>M Schneider</i>	690
Principal Evaluation: Concepts, Needs and Realities	<i>N S Glasman and L D Glasman</i>	698
Program Evaluation	<i>C A Christie and L A Fierro</i>	706
School Based Evaluation: Purposes, Protocols and Processes	<i>J MacBeath</i>	713
School Inspection/External School Evaluation	<i>C Chapman and P Earley</i>	719
The International System for Teacher Observation and Feedback: A Theoretical Framework for Developing International Instruments	<i>L Kyriakides, B P M Creemers, C Teddlie, and D Muijs</i>	726

Educational Evaluation – Evaluation Methods

Evaluation Methodology	<i>M Chatterji</i>	735
Evaluation Reporting and Communicating	<i>R T Torres</i>	746
Longitudinal Evaluation Designs	<i>X Ma</i>	757
Meta-evaluation: Purpose, Prescription, and Practice	<i>L A Wingate</i>	765

Educational Evaluation – Purposes and Models of Educational Evaluation

Defining Quality in Evaluation	<i>P Dahler-Larsen</i>	775
Internal and External Evaluation	<i>D Nevo</i>	781
Needs Assessment in Education	<i>J W Altschuld and T L Lepicki</i>	786
The Purpose of Educational Evaluation	<i>S Mathison</i>	792
Understanding Approaches to Evaluation	<i>A J Love</i>	798

VOLUME 4**Educational Measurement**

Ability Testing	<i>P B Tymms</i>	1
Admissions Testing	<i>R Zwick</i>	7
Appropriateness Measurement: Person Fit	<i>R R Meijer and A M Weekers</i>	15
Automated Essay Scoring: Writing Assessment and Instruction	<i>M D Shermis, J Burstein, D Higgins, and K Zechner</i>	20
Classical Test Theory Reliability	<i>M D Miller</i>	27
Criterion-Referenced Measurement	<i>S A Shrock and W C Coscarelli</i>	31
Differential Item Functioning	<i>M D Hidalgo and J Gómez-Benito</i>	36
Educational Measurement: Overview	<i>R L Linn</i>	45
Equating and Scaling	<i>A A von Davier</i>	50
Fairness	<i>C Gipps and G Stobart</i>	56
Generalizability Theory	<i>R L Brennan</i>	61
Impact of Assessment on Classroom Practice	<i>J L Herman</i>	69
Intelligent Systems	<i>V J Shute and D Zapata-Rivera</i>	75
Item Response Theory	<i>W J van der Linden</i>	81
National Assessments	<i>T J Crooks</i>	89
Norm-Referenced Measurement	<i>H D Hoover</i>	95
Progress Monitoring	<i>L S Fuchs and D Fuchs</i>	102
Rasch Models	<i>D Andrich</i>	111
Scoring Rubrics	<i>J A Arter</i>	123
Second Language Assessment	<i>L F Bachman</i>	140
Self Adaptive Testing	<i>J Revuelta</i>	148
Student Test Results in School Accountability	<i>L S Hamilton</i>	153
Test Development	<i>S M Downing</i>	159
Test Translation and Adaptation	<i>A Allalouf and P Hanani</i>	166
Testing Creativity	<i>M A Runco</i>	170
Testing in History	<i>S Wineburg and B Fogo</i>	175
Validity	<i>R L Linn</i>	181
Validity of Achievement Gains on High-Stakes Tests	<i>D Koretz</i>	186
Validity of Educational Indicators	<i>B P M Creemers and L Kyriakides</i>	193

G**Globalization of Universities and Institutions**

Globalization, Universities and Medium of Instruction	<i>J Lo Bianco</i>	201
The Washington Accord on Engineering Curriculum	<i>E Ko</i>	209

H

Higher Education

Higher Education: An Overview	<i>P Scott</i>	217
-------------------------------	----------------	-----

Higher Education – Management, Leadership and Governance

Academic Managers – Heads of Departments/Schools	<i>S Marshall</i>	229
Leading and Managing the University – Presidents and their Senior Management Team	<i>R Middlehurst, T Kennie, and S Woodfield</i>	238
Management of and in Higher Education Institutions	<i>B Sporn</i>	245
Managerialism and Collegialism in Higher Education Institutions	<i>M Shattock</i>	251
Strategic Planning in Higher Education	<i>L Zechlin</i>	256
The Changing Role of University Presidents, Vice-Chancellors and Rectors	<i>M F Green and P D Eckel</i>	264
The Changing Roles of University Governing Boards and Councils	<i>J Fielden</i>	273

Higher Education – Research and Scholarship

Funding of University Research	<i>G Harman</i>	279
Mobility of PhD Students and Scientists	<i>S Avveduto</i>	286
National Science Policy and Universities	<i>M Henkel and U Kogan</i>	294
Research Commercialization	<i>S Slaughter</i>	300
Research Quality Assessment	<i>I McNay</i>	307
The Management of University Research	<i>J Kirkland</i>	316
The Training of New Scientists	<i>R Neumann</i>	322

Higher Education – Society

Academic Freedom	<i>S F Macintyre</i>	328
Access and Equity in Higher Education	<i>G Baldwin and R James</i>	334
Boundaries of Institutional Autonomy and their Impact on Higher Education	<i>V L Meek</i>	341
Diversity of Higher Education	<i>U Teichler</i>	347
Employability of University Graduates and Graduate Outcomes	<i>J Brennan</i>	354
Higher Education and the Knowledge Society	<i>J Välimaa</i>	360
Higher Education and the Labor Market	<i>J Enders</i>	365
Higher Education and the Transformation of Society	<i>P Scott</i>	370
Quality Assurance in Higher Education – Practices and Issues	<i>D Dill</i>	377
The Role of Accreditation of Higher Education Institutions	<i>J S Eaton</i>	384
Universities and Regional Development	<i>J Goddard and J Puukka</i>	390
Universities' Engagement with Society	<i>D Watson</i>	398

Higher Education – Teaching and Learning in Higher Education

Distance Education and Open Universities	<i>A Kanwar and J Daniel</i>	404
Education Development in Higher Education	<i>Y Ryan and K Fraser</i>	411
Effective Use of Technology in Teaching and Learning in HE	<i>D Laurillard</i>	419

Facilitating Students' Success – Student Support and Learning Resources in Universities	<i>E Woodberry and C Barnes</i>	427
Faculty Values and Expectations and Their Implications for Teaching, Learning and the Organization of Higher Education Institutions	<i>K Harman</i>	433
Flexible Learning in Higher Education	<i>S Alexander</i>	441
Innovation in Teaching and Curriculum Design	<i>J M Dearn</i>	448
Productivity of University Faculty Staff	<i>S Kyvik</i>	455
Promoting Effective Student Learning in Higher Education	<i>K Trigwell</i>	461
Student Attrition from Higher Education Institutions	<i>M Hayden</i>	467
University Students' Experiences of Higher Education	<i>C McInnis</i>	473
What Do University Students Learn?	<i>G D Kuh and B K Bridges</i>	479

Higher Education – Tertiary Education in a Globalized World

Associations of Universities	<i>G F Moodie</i>	485
Collective Bargaining in Higher Education	<i>C Allport, K McAlpine, and S Roberts</i>	491
Financing of Higher Education	<i>B Chapman and P Tulip</i>	499
Higher Education Crossing Borders	<i>J Knight</i>	507
Higher Education in Post-Conflict Societies	<i>M Daxner</i>	514
Higher Education in the New Economies	<i>J Fielden</i>	520
Human Resource Issues in Higher Education	<i>E El-Khawas</i>	527
International Accreditation in Higher Education	<i>A Stella and D Woodhouse</i>	533
Internationalization of Higher Education	<i>M van der Wende</i>	540
National and International Rankings of Higher Education	<i>S Marginson</i>	546
Steering of Higher Education Systems – The Role of the State	<i>K Yokoyama and V L Meek</i>	554
Student and Faculty Transnational Mobility in Higher Education	<i>U Lanzendorf and B M Kehm</i>	559
The Bologna Process in European Higher Education	<i>D Rich</i>	566
The Economics and Finance of Higher Education	<i>A Wolf</i>	573
The Role of the OECD in the Development of Higher Education in a Globalized World	<i>R Yelland</i>	584
Transforming Higher Education in Developing Countries: The Role of the World Bank	<i>J Salmi and R M Bassett</i>	590
UNESCO's Role in the Development of Higher Education in a Globalized World	<i>S Uvalić-Trumbić</i>	597

Higher Education – The Organization and Financing of Higher Education

Higher Education in Federal Systems	<i>M A M Cortés and U Teichler</i>	603
Pathways and Articulation into Higher Education	<i>G F Moodie</i>	609
Unitary, Binary and other National Systems of Higher Education	<i>L Goedegebuure</i>	616
Viewing Private Higher Education: How Much, Where, Why, and What?	<i>D C Levy</i>	622

I

Immigration

Children of Migrant Populations	<i>M M Suárez-Orozco and C Suárez-Orozco</i>	629
Third Culture Kids	<i>R E Van Reken</i>	636

International Comparisons

IEA Studies in ICT	<i>N Law</i>	644
IEA Studies in Mathematics and Science	<i>F K S Leung</i>	650
IEA Study in Civic Education	<i>J Torney-Purta, J-A Amadeo, and J Schwillie</i>	656
IEA: Globalization and Assessment	<i>H Wagemaker</i>	663

International Organizations in Education

Education and the European Union	<i>S Hegarty</i>	669
National University on Educational Planning and Administration, New Delhi	<i>J B G Tilak</i>	676
On School Management – International Council on School Effectiveness and Improvement, Commonwealth Council on Educational Administration and Management	<i>A Walker and T Townsend</i>	681
UNESCO's International Institute for Educational Planning	<i>S J Hite and M Bray</i>	688

L**Leadership and Management**

Leadership and Management Overview	<i>B Mulford</i>	695
------------------------------------	------------------	-----

Leadership and Management – as a Field of Study

Contexts for Leadership at the Beginning of the 21st Century	<i>M Bottery</i>	704
Educational Leadership: Philosophical Issues	<i>C W Evers</i>	710
Ethical/Moral Issues in Educational Leadership	<i>R J Starratt, L Langlois, and P Duignan</i>	717
History of Educational Leadership/Management	<i>R Bates</i>	724
Educational Leadership as a Field of Study: Professional Journals 1994–2006	<i>A R Thomas</i>	731
Research on Educational Leadership – Approaches/Promising Directions	<i>W A Firestone and V M J Robinson</i>	740

Leadership and Management – Leader Preparation

Leaders for Productive Schools	<i>J Murphy, S N Elliott, E B Goldring, and A C Porter</i>	746
New Approaches in Preparing School Leaders	<i>S G Huber</i>	752
Research for Leaders: The National College for School Leadership	<i>G Southworth</i>	762
School Leadership Standards	<i>L Ingvarson</i>	768

Leadership and Management – Leadership Issues

Effective Leadership in Challenging Schools	<i>A Harris</i>	777
Esthetics, Educational Leadership and Management	<i>R Bates</i>	783
Leader Mistreatment of Teachers	<i>J Blase and J Blase</i>	790
Leadership and Gender	<i>J Blackmore</i>	797
Leadership and Student Voice	<i>C Lanskey and J Rudduck</i>	803
Leadership and Urban Education	<i>K A Riley</i>	809
Leadership for Learning	<i>J MacBeath</i>	817
Leadership in Diverse Cultures	<i>A Walker and P K Yu-kwong</i>	824

Leading Diverse Schools	<i>J Ryan</i>	832
Supervision and Personnel Administration	<i>P D Tucker and D G Pounder</i>	839
The Big Challenge: Leadership for Inclusion	<i>M Ainscow and A Sandill</i>	846
The Educational Leader's New Work	<i>P Gronn</i>	852

VOLUME 5

Leadership and Management – Leadership Types

A Distributed Perspective on School Leadership and Management	<i>J P Spillane</i>	1
Leadership: Authentic	<i>P T Begley</i>	7
Leadership: Democratic	<i>J Møller</i>	12
Leadership: Instructional	<i>P Hallinger, K Leithwood, and R H Heck</i>	18
Leadership: Transformative	<i>C M Shields</i>	26
Strategic Leadership	<i>B Davies and B J Davies</i>	34
Teacher Leadership and Organizational Development	<i>A Harris</i>	40
Understanding How Leadership Influences Student Learning	<i>K Leithwood and B Levin</i>	45

Leadership and Management – Politics and Governance

Equity and Educational Effectiveness	<i>P Sammons</i>	51
Leading/Managing Schools in Communities Made Poor	<i>P Thomson</i>	58
Principals Role in Restructuring Schools	<i>G M Crow and K D Peterson</i>	66
School Reform and Restructuring: Self Managing School	<i>B J Caldwell</i>	72
System Leadership	<i>D Hopkins and R Higham</i>	78

Leadership and Management – School and Community

Educational Leadership and Social Capital	<i>J West-Burnham</i>	85
Networked Learning Communities School-to-School Collaboration as an Essential Component of a System Reform Strategy	<i>D Jackson</i>	92
Networks and Communities of Knowledge	<i>T Schuller and H Theisens</i>	100
School Community Relations	<i>R L Crowson and E B Goldring</i>	106
Social Capital, Educational Institutions and Leadership	<i>S Kilpatrick, S Johns, and B Mulford</i>	113

Leadership and Management – School Effectiveness and Improvement

Leadership and Technology	<i>C F Webber</i>	120
Leadership in the Implementation of Innovations	<i>I E Bogotch, T Townsend, and M Acker-Hocevar</i>	128
Leadership: School Improvement	<i>R H Heck and P Hallinger</i>	135
Organizational Learning in Schools	<i>B Mulford and H Silins</i>	143
Professional Learning Community	<i>L Stoll</i>	151
Transformational School Leadership	<i>K Leithwood</i>	158

Learning and Cognition – Issues and Concepts

Attention in Cognition and Early Learning	<i>M Heimann, T Tjus, and K Strid</i>	165
Concept Learning	<i>S J Unsworth and D L Medin</i>	172
Cooperative Learning	<i>R E Slavin</i>	177
Intelligence	<i>R J Sternberg</i>	184
Learning from Multiple Information Sources	<i>H I Strømsø and I Bråten</i>	191
Memory	<i>S Magnussen and T Brennen</i>	197
Metacognition	<i>L Baker</i>	204
Personal Epistemology in Education: Concepts, Issues, and Implications	<i>I Bråten</i>	211
Piaget: Recent Work	<i>L Smith</i>	218
The Adult Development of Cognition and Learning	<i>E A L Stine-Morrow and J M Parisi</i>	225
Vygotsky and Recent Developments	<i>J V Wertsch</i>	231

Learning and Cognition – Issues and Concepts – Focus on Cognition

Cognition and Emotion	<i>F Pons, M de Rosnay, and F Cuisinier</i>	237
Cognition: Overview and Recent Trends	<i>J Wiley and B D Jee</i>	245
Development of Creativity	<i>M A Runco</i>	251
Knowledge Domains and Domain Learning	<i>L Maggioni and P A Alexander</i>	255
Problem Solving and Human Expertise	<i>T J Nokes, C D Schunn, and M T H Chi</i>	265
Problem Solving and Reasoning	<i>R E Mayer</i>	273
Reasoning, Moral and Social	<i>N S Jampol, C B Richardson, and M Killen</i>	279
Remembering as Social Activity	<i>D Middleton</i>	285

Learning and Cognition – Issues, Concepts, Types

Computer-Supported Collaborative Learning: Basic Concepts, Multiple Perspectives, and Emerging Trends	<i>S R Ludvigsen and A I Mørch</i>	290
Learning as Inquiry	<i>S R Goldman, J Radinsky, S Tozer, and D Wink</i>	297
Learning Through Narrative Socialization	<i>A Fasulo</i>	303
Neuroscience Bases of Learning	<i>M H Immordino-Yang and K W Fischer</i>	310
The Neuroscience of Reading	<i>U Goswami</i>	317

Learning and Cognition – Issues, Concepts, Types – Focus on Learning

Learning Strategies	<i>C E Weinstein, J Jung, and T W Acee</i>	323
Learning Through Play	<i>K Aronsson</i>	330
Situated View of Learning	<i>A Collins and J G Greeno</i>	335
The Neuroscience of Aging and Cognition	<i>W J Hoyer and B Bessette-Symons</i>	340

Learning and Cognition – Language, Literacy and Subject-Related

An Overview of Language and Literacy in Educational Settings	<i>V G Aukrust</i>	345
Bilingualism and Learning	<i>L S Verplaetse and E Schmitt</i>	355
Classroom Discourse and Student Learning	<i>L Hemphill</i>	361
Early Writing	<i>B E Hagtvet</i>	367

First Language Acquisition	<i>N B Ratner</i>	375
Foreign Language Learning	<i>C Kinginger</i>	382
Learning Science	<i>J Leach and P Scott</i>	387
Learning to Read	<i>M S Burns and J K Kidd</i>	394
Mathematics Learning	<i>L Verschaffel, B Greer, and E De Corte</i>	401
Music Learning	<i>S Hallam</i>	407
Reading Comprehension: Reading for Learning	<i>C E Snow</i>	413
Second Language Learning	<i>J S Arnfast, J N Jørgensen, and A Holmen</i>	419
Writing: Advanced	<i>R Andrews</i>	426

Learning and Cognition – Other Issues

Gender and Learning	<i>L Yates and J McLeod</i>	432
Parent–Child Relationships in Early Learning	<i>C P Edwards, S M Sheridan, and L L Knoche</i>	438
Peer Interaction and Learning	<i>S Blum-Kulka and S Dvir-Gvirsman</i>	444

Learning and Cognition – Places of Learning

Learning in a Cross-Cultural Perspective	<i>L-F Zhang and R J Sternberg</i>	450
Learning Outside of School	<i>T Nunes</i>	457
Reasoning and Explanatory Talk: Learning in Everyday Settings	<i>D E Beals</i>	464

Learning and Cognition – Theoretical Perspectives – Learning

Apprenticeship Approach to Learning	<i>K Nielsen</i>	469
Connectionism and Learning	<i>T R Shultz</i>	476
Constructivism and Learning	<i>S Sjøberg</i>	485
Cultural–Historical Activity Theory	<i>A Havnes</i>	491
Learning in a Sociocultural Perspective	<i>R Säljö</i>	498
Self-Regulated Learning and Socio-Cognitive Theory	<i>P H Winne and A F Hadwin</i>	503

N

National Systems of Education

Australia	<i>C N Power</i>	510
Austria	<i>M J Gaertner</i>	517
Bangladesh	<i>R Z Dewan</i>	524
Barbados	<i>R Rudder</i>	534
Bulgaria	<i>A Chavdarova and D Vesselinov</i>	540
England	<i>C Whatford</i>	550
Estonia	<i>M Keränen and J Kõrgesaar</i>	555
Fiji	<i>I Senibulu</i>	563
Finland	<i>J E Tuovinen</i>	569
France	<i>M Sellier</i>	576

Germany	<i>A Hüfner and K Hüfner</i>	582
Ghana	<i>R Palmer</i>	590
Greece	<i>K Toura and A Karagiannis</i>	599
Iceland	<i>M Hardardottir and M Gunnlaugsdóttir</i>	605
Indonesia	<i>A C Alwasilah and Furqon</i>	611
Israel	<i>A Volansky</i>	618
Jamaica	<i>E Hannam</i>	627
Latvia	<i>A Joma</i>	630
Lebanon	<i>N A Mneimneh</i>	636
Lithuania	<i>J Navickaitė, L Paurienė, R Zablackė, A Tijūnelytė, and V Kožemiakina</i>	656
Malawi	<i>F R Mkandawire and D Mulera</i>	666
Malta	<i>C Farrugia</i>	673
Mongolia	<i>B Yembuu</i>	681
Namibia	<i>J Ellis</i>	687
New Zealand	<i>R Baker and C Wylie</i>	692
Niger	<i>A A I Daouda</i>	699
Norway	<i>H Farstad</i>	709
Oman	<i>S A Issan</i>	718
Pakistan	<i>N Mahmood and R Malik</i>	727
Poland	<i>A Janowski</i>	737
Qatar	<i>S R Salman</i>	744
Russia	<i>V M Filippov and G F Tkach</i>	750
Rwanda	<i>E Bahizi</i>	758
Scotland	<i>D Raffe</i>	770
Singapore	<i>P T Ng</i>	776
Slovenia	<i>T Plevnik and A B Lakota</i>	782
Solomon Islands	<i>D P Malasa</i>	790
South Africa	<i>P Kgobe and J Pampallis</i>	799
Spain	<i>A Tiana</i>	806
Sri Lanka	<i>B M J Balasooriya, M M Wehella, and S U Wijeratne</i>	813
Suriname	<i>H R Ori</i>	826
Taiwan	<i>C-M Lee</i>	834
Tanzania	<i>R Wedgwood</i>	841
Thailand	<i>C Rukspollmuang</i>	847
The Gambia	<i>Y A-M Jobe</i>	858
The Philippines	<i>E A P Valenzuela</i>	865
Uruguay	<i>B Guinovart and C Troncoso</i>	873
USA	<i>B Pusser and M E R Lloyd</i>	888
Zanzibar	<i>I M Hija</i>	896

VOLUME 6**P****Philosophy of Education**

Philosophy of Education: Overview	<i>P Enslin</i>	1
-----------------------------------	-----------------	---

Philosophy of Education – Contemporary Issues

Neoliberalism, the Market and Performativity	<i>M A Peters</i>	11
Parents and Children's Rights	<i>M Clayton</i>	17
Peace Education	<i>I Gur-Ze'ev</i>	22
Philosophy and Educational Research	<i>D Bridges</i>	29
Race, Critical Race Theory and Whiteness	<i>B Applebaum</i>	36
School Choice and the Common School	<i>G Haydon</i>	44

Philosophy of Education – Philosophical Perspective

African Philosophy of Education	<i>K Horsthemke</i>	50
Feminism	<i>L Stone</i>	56
Hermeneutics	<i>C Bingham</i>	63
Islamic Education	<i>Y Waghid</i>	69
Liberalism and Education	<i>W Feinberg</i>	74
Phenomenology and Existentialism	<i>B Levering</i>	80
Pragmatism	<i>B Thayer-Bacon</i>	86
Psychoanalysis and Education	<i>P Smeyers</i>	92
Social and Cultural Capital in Education	<i>M Print</i>	97
The Aims of Education	<i>R Marples</i>	103
The Analytical Tradition	<i>D Carr</i>	108

Philosophy of Education – Philosophical Themes

Adult Learning: Philosophical Issues	<i>D Beckett</i>	114
Autonomy	<i>C A Wringe</i>	121
Bildung	<i>J Masschelein and N Ricken</i>	127
Critical Theory and Pedagogy	<i>M Mason</i>	133
Critical Thinking	<i>H Siegel</i>	141
Environmental Education	<i>M Bonnett</i>	146
Identity	<i>M S Merry</i>	152
Inclusion	<i>R Cigman</i>	158
Justice and Care	<i>M Tjiattas</i>	164
Lifelong Learning	<i>J Halliday</i>	170
Post-Structuralism and Education	<i>M Mason and M Clarke</i>	175
Reason and Rationality	<i>P Hogan</i>	183

Religious and Spiritual Education	<i>J C Conroy and S J McKinney</i>	189
The Capabilities Approach	<i>L Terzi</i>	194

Primary and Secondary Education – Learning and Teaching in School Age Education

Class Size – Arguments and Evidence	<i>P Blatchford and K C Lai</i>	200
Continuing Professional Development of Teachers	<i>P Earley</i>	207
Homeschooling	<i>P Farenga</i>	214
Innovations and Early Intervention in the Teaching of Literacy	<i>R E Slavin</i>	221
Personalized Learning in School Age Education	<i>D Hopkins</i>	227
Perspectives on Schooling in the Middle Years	<i>V J Russell</i>	233
Reform of the School Workforce	<i>H M Gunter and G Forrester</i>	242
Secondary Education in Developing Countries	<i>P Hughes and A Walker</i>	249

Primary and Secondary Education – Social Justice Issues in School Age Education

Community Focused Schools	<i>A Dyson</i>	255
---------------------------	----------------	-----

Primary and Secondary Education – The Effectiveness of School Systems

Reforming the American School System	<i>M Barber</i>	261
School Effectiveness in Developed Societies	<i>D Reynolds</i>	269
School to School Collaboration: Innovation and Improvement	<i>R Higham</i>	274
Studies of School Improvement in Developing Countries	<i>S E Anderson</i>	281
Sustainable Educational Reform	<i>A Hargreaves</i>	289
The Role of the District in Tri Level Reform	<i>M Fullan</i>	295
Whole School Designs for Enhancing Student Achievement	<i>S Stringfield and J Nunnery</i>	303

Q

Qualitative Research

Action Research in Education	<i>E Stringer</i>	311
Classroom Ethnography	<i>F Erickson</i>	320
Computer Assisted Qualitative Data Analysis	<i>C Silver and A Lewins</i>	326
Conversational Analysis	<i>H G Levine</i>	335
Critical Ethnography	<i>D Foley</i>	337
Critical Race Theory	<i>D Gillborn and G Ladson-Billings</i>	341
Critical Theory	<i>J J Ryoo and P McLaren</i>	348
Cultural Studies	<i>R Gaztambide-Fernández and K Cairns</i>	354
Cultural–Historical Activity Theory	<i>Laboratory for Comparative Human Cognition</i>	360
Design Experiments	<i>A Collins</i>	367
Discourse Analysis	<i>P Nikander and S Vehviläinen</i>	373
Document Analysis	<i>T Rapley and K N Jenkins</i>	380
Ethnography	<i>M Hammersley</i>	386

Ethnomethodology in Education Research	<i>D Macbeth</i>	392
Focus Groups	<i>D L Morgan and K Hoffman</i>	401
Grounded Theory	<i>K Charmaz and A Bryant</i>	406
Hermeneutics	<i>J Dyer</i>	413
Interpretive Research	<i>E Rauscher and M E Graue</i>	419
Interviews and Interviewing	<i>A Marvasti</i>	424
Life History	<i>J A Hatch and S K Newsom</i>	430
Narrative Inquiry	<i>D J Clandinin and J Huber</i>	436
Participant Observation	<i>J Schostak</i>	442
Phenomenology	<i>M van Manen and C A Adams</i>	449
Qualitative Case Studies	<i>S B Merriam</i>	456
Qualitative Data Management	<i>C Grbich</i>	463
Semiotics	<i>M Danesi</i>	475
Symbolic Interactionism, Naturalistic Inquiry, and Education	<i>J A Forte</i>	481
The Quality of Evidence in Qualitative Research	<i>S K Biklen</i>	488
Visual Data in Education Research	<i>J Wagner</i>	498

Qualitative Research – Standards of Proof

Validity: Mapping Diverse Perspectives	<i>G B Rossman, S F Rallis, and A M Kuntz</i>	505
--	---	-----

Quantitative and Qualitative Approaches to Research – Integration

Mixed Methods	<i>A Tashakkori and I Newman</i>	514
---------------	----------------------------------	-----

Quantitative Methods and Research Design

External Validity	<i>G E Matt, A Brewer, and M Sklar</i>	521
-------------------	--	-----

R

Role of NGOs in Globalization

NGOs and Globalization of Education	<i>M Ahmed</i>	529
-------------------------------------	----------------	-----

S

Social and Emotional Aspects of Learning

Achievement Goal Theory: Definitions, Correlates, and Unresolved Questions	<i>T Urdan</i>	537
Affect, Mood and Emotions	<i>S Schnall</i>	544
Anxiety	<i>M Zeidner</i>	549
Attribution Theory	<i>B Weiner</i>	558
Children's Friendship	<i>S R Asher and K L McDonald</i>	564
Coping with Stressful Situations: An Important Aspect of Self-Regulation	<i>M Boekaerts</i>	570
Culture in Motivation Research: A Challenging and Enriching Contribution	<i>K Kimmel and S Volet</i>	576
Early Social Development and Schooling	<i>G M van der Aalsvoort</i>	585

Emotion in Educational Contexts	<i>P A Schutz, P D Quijada, S de Vries, and M Lynde</i>	591
Empathy	<i>R Arnold</i>	597
Flow in Education	<i>J A Schmidt</i>	605
Interest	<i>M Ainley</i>	612
Intrinsic and Extrinsic Motivation	<i>C Levesque, K J Copeland, M D Pattie, and E L Deci</i>	618
Motivating Students in Classrooms	<i>J Brophy</i>	624
Motivation Regulation	<i>C A Wolters and S A Mueller</i>	631
Peer Learning in the Classroom	<i>N M Webb</i>	636
Poverty, Effects of on Social and Emotional Development	<i>N Philipsen Hetzner, A D Johnson, and J Brooks-Gunn</i>	643
Seeking Help as an Adaptive Response to Learning Difficulties: Person, Situation, and Developmental Influences	<i>S A Karabenick and R S Newman</i>	653
Self-Concept in Learning: Reciprocal effects model between academic self-concept and academic achievement	<i>H W Marsh and L F Scalas</i>	660
Self-Efficacy Beliefs	<i>D H Schunk and F Pajares</i>	668
Social and Emotional Outcomes of Learning	<i>J Vorhaus</i>	673
Social Aspects of Collaborative Learning	<i>M Arvaja and P Häkkinen</i>	685
Social Development and Schooling	<i>G W Ladd</i>	691
Social Interaction and Learning	<i>K Littleton</i>	698
Social Networks and the Education of Children and Youth	<i>B Schneider, T G Ford, and L Perez-Felkner</i>	705
Sociocultural Issues in Motivation	<i>R A Walker</i>	712
Socioemotional Aspects of Technology-Supported Learning	<i>S Järvelä</i>	718
Volitional Control of Learning	<i>J Husman and L Corno</i>	724
Wellbeing	<i>T Hascher</i>	732

VOLUME 7

Statistics

An Overview of Statistics in Education	<i>S Sinharay</i>	2
Analysis and Interpretation of Multivariate Data	<i>D J Bartholomew</i>	12
Analysis of Covariance	<i>N T Longford</i>	18
Analysis of Extreme Values in Education	<i>J Beirlant, G Dierckx, K Luwel, and P Onghena</i>	25
Analysis of Variance	<i>B M King</i>	32
Bayesian Statistical Analysis	<i>G Kokolakis</i>	37
Bootstrap Method	<i>K Singh and M Xie</i>	46
Canonical Correlation	<i>A Basu and A Mandal</i>	52
Categorical Data Analysis	<i>S E Fienberg and B W Junker</i>	58
Causal Inference	<i>D B Rubin</i>	66
Cluster Analysis: Overview	<i>S Landau and I Chis Ster</i>	72

Cognitive Psychology and Educational Statistics	<i>H S Kim and S Embretson</i>	84
Computational Statistics	<i>J E Gentle</i>	93
Continuous Probability Distributions	<i>S Sinharay</i>	98
Correspondence Analysis	<i>M Greenacre</i>	103
Data Mining	<i>R S J d Baker</i>	112
Decision Theory	<i>M Mendoza and E Gutiérrez-Peña</i>	119
Design of Experiments	<i>C A Albers and T R Kratochwill</i>	125
Discrete Probability Distributions	<i>S Sinharay</i>	132
Discrimination and Classification	<i>R C Neath and M S Johnson</i>	135
Empirical Bayes Methods	<i>E Ip and G Molenberghs</i>	142
Evaluation Research	<i>T Kellaghan</i>	150
Exploratory Data Analysis	<i>A Unwin</i>	156
Factor Analysis: An Overview and Some Contemporary Advances	<i>F A N Alhija</i>	162
Generalized Linear Mixed Models	<i>S Rabe-Hesketh and A Skrondal</i>	171
Generalized Linear Models	<i>J J Faraway</i>	178
Generating Random Numbers	<i>P J Pashley and A Amodeo</i>	184
Goodness-of-Fit Testing	<i>A Maydeu-Olivares and C García-Forero</i>	190
Graphical Models	<i>R G Almond</i>	197
Growth Modeling	<i>I Plewis</i>	203
Hierarchical Linear Models	<i>D Rindskopf</i>	210
Hypothesis Testing and Confidence Intervals	<i>T R Konold and X Fan</i>	216
Instrumental Variables	<i>E P Bettinger</i>	223
Jackknife Methods	<i>S Sinharay</i>	229
Large-sample Statistical Methods	<i>J M Singer and D F Andrade</i>	232
Latent Class Models	<i>J K Vermunt</i>	238
Markov Chain Monte Carlo	<i>A M Johansen</i>	245
Matrix Algebra	<i>H Yanai and H Ishii</i>	253
Measure of Association	<i>G-H Huang</i>	260
Measures of Central Tendency	<i>D Clark-Carter</i>	264
Measures of Dispersion, Skewness and Kurtosis	<i>D Rindskopf and M Shiyko</i>	267
Meta Analysis	<i>J Sánchez-Meca and F Marín-Martínez</i>	274
Missing Data	<i>C A W Glas</i>	283
Model Selection	<i>S J Haberman</i>	289
Monte Carlo Methods	<i>A M Johansen</i>	296
Multidimensional Scaling	<i>Z Zhang and Y Takane</i>	304
Multiple Comparisons	<i>C Lewis</i>	312
Multivariate Analysis of Variance	<i>N T Longford</i>	319
Multivariate Linear Regression	<i>S Ganesh</i>	324
Multivariate Normal Distribution	<i>J Tacq</i>	332

Nonlinear Regression Analysis	<i>H-H Huang, C K Hsiao, and S-Y Huang</i>	339
Nonparametric Statistical Methods	<i>K Sijtsma and W H M Emons</i>	347
Observational Studies	<i>P R Rosenbaum</i>	354
Order Statistics	<i>I Bairamov</i>	360
Point Estimation Methods with Applications to Item Response Theory Models	<i>F Bartolucci and L Scrucca</i>	366
Principal Components Analysis	<i>D J Bartholomew</i>	374
Probability Theory	<i>T Rudas</i>	378
Recursive Partitioning	<i>S Landau and S Barthel</i>	383
Sampling	<i>J Qian</i>	390
Sequential Probability Ratio Test	<i>T J H M Eggen</i>	396
Sequential Testing	<i>H J Vos</i>	401
Signal Detection Theory	<i>H Abdi</i>	407
Small Area Estimation	<i>A Noble</i>	411
Statistical Analysis of Functional Data	<i>F Ferraty, P Sarda, and P Vieu</i>	417
Statistical Inequalities	<i>H M Mahmoud and J L Gastwirth</i>	426
Statistical Paradoxes	<i>Y-K Tu</i>	429
Statistical Power Analysis	<i>L V Hedges and C Rhoads</i>	436
Statistical Significance Versus Effect Size	<i>X Fan and T R Konold</i>	444
Stochastic Processes	<i>F Baudoin</i>	451
Structural Equation Models	<i>S-Y Lee and X-Y Song</i>	453
Survival Data Analysis	<i>A F De Champlain</i>	459
The Normal Distribution and its Applications	<i>J Tacq</i>	467
Time Series Analysis	<i>P J Brockwell</i>	474
Univariate Linear Regression	<i>M L Hazelton</i>	482
Value-Added Models	<i>D F McCaffrey</i>	489

T

Teacher Education – Evaluation in the Teacher Education Program

Accreditation and Standards in Teacher Education	<i>R J Yinger and K L Daniel</i>	495
Determining Long Term Effects of Teacher Education	<i>C N Brouwer</i>	503

Teacher Education – Evaluation of the Prospective Teacher

The Evaluation of Prospective Teachers in Teacher Education	<i>D E H Tigelaar and J van Tartwijk</i>	511
---	--	-----

Teacher Education – Macro Factors

Diversity in Teacher Education	<i>T Wubbels</i>	518
Social and Cultural Influences on Teacher Education	<i>M Ben-Peretz and R Lotan</i>	525
The Role of Governance in Teacher Education	<i>G E Karlsen</i>	532

Teacher Education – Post-Initial Professional Development

Action Research as a Tool for Teachers' Professional Development	<i>P Ponte</i>	540
Contemporary Approaches to Teacher Professional Development	<i>H Borko, J Jacobs, and K Koellner</i>	548
School Development for Teacher Learning and Change	<i>P Slegers and K Leithwood</i>	557
Teacher Induction	<i>D Beijaard, J Buitink, and C Kessels</i>	563
Teacher Learning as Workplace Learning	<i>J Imants and K van Veen</i>	569
Teacher Learning with Lesson Study	<i>A Murata</i>	575
Teachers Career Stages and Professional Development	<i>R Fessler and E Rice</i>	582

Teacher Education – Preservice Teacher Education

A Pedagogy of Teacher Education	<i>J Loughran</i>	587
Case Methods in Teacher Education	<i>M Sato and C Rogers</i>	592
Conceptions of Teacher Education	<i>J A Whitcomb</i>	598
Inquiry-Oriented Teacher Education	<i>R Schulz</i>	604
Narratives and Biography in Teacher Education	<i>G Kelchtermans</i>	610
Supervision in Teacher Education	<i>J McCarthy and L F Quinn</i>	615
Taking Prospective Teachers' Beliefs into Account in Teacher Education	<i>M Valcke, G Sang, I Rots, and R Hermans</i>	622
Teacher Education and Models of Teacher Reflection	<i>V K LaBoskey</i>	629
Teacher Education as Teaching for Understanding with New Technologies	<i>M S Wiske and D E Spicer</i>	635

Teacher Education – Preservice: The Knowledge Base

Experienced Teachers' Craft Knowledge	<i>P C Meijer</i>	642
Moral Values in Teacher Education	<i>W Veugelers</i>	650
Pedagogical Content Knowledge	<i>J H Van Driel and A Berry</i>	656
Teacher Education and the Educational Foundations Knowledge Base	<i>F B Murray</i>	662
The Relationship Between Theory and Practice in Teacher Education	<i>F A J Korthagen</i>	669

Teacher Education – Teacher Educators

Characteristics, Scholarship and Research of Teacher Educators	<i>M Lunenberg</i>	676
Professional Development of Teacher Educators	<i>P K Smith</i>	681
Self-Study by Teacher Educators	<i>T Russell</i>	689

Teacher Education – Teaching Specific Domains

Arts Teacher Education	<i>L Beudert</i>	695
Mathematics Teacher Education	<i>K Krainer and S Llinares</i>	702
Music Teacher Education	<i>M W-y Wong</i>	706
Science Teacher Education	<i>J H Van Driel and S K Abell</i>	712
Social Studies Teacher Education	<i>A-L Halvorsen and S M Wilson</i>	719
Teacher Education for Elementary Education	<i>K D Brown and G N Carnes</i>	725

Teacher Education for Teaching the Gifted	<i>A Y Baldwin</i>	732
Teacher Preparation in Mother Tongue Education	<i>M Kooy and H Hulshof</i>	738
Vocational Teacher Education	<i>S Johansson</i>	744

Teacher Education – The Cooperating School

Mentoring in Teacher Education	<i>P D Tomlinson, A J Hobson, and A Malderez</i>	749
Partnerships Between Schools and Higher Education	<i>C Day and L Smethem</i>	757

VOLUME 8

Technology and Learning

An Overview of Technology and Learning	<i>B Means and J Roschelle</i>	1
Educational Data Modeling	<i>S Iksal</i>	11
Student Data Systems and Their Use for Educational Improvement	<i>J C Wayman, V Cho, and M P Richards</i>	14
Technology and Learning: Access in Schools Around the World	<i>R Anderson and J Ainley</i>	21

Technology and Learning – as a Subject of Instruction

Conceptions of Technology Literacy and Fluency	<i>O Erstad</i>	34
--	-----------------	----

Technology and Learning – Assessment

Technology and Formative Assessment	<i>J W Pellegrino</i>	42
Technology for Large-Scale Assessment	<i>R E Bennett</i>	48
Technology Supports for Assessment Design	<i>R J Mislevy, F I Winters, I I Bejar, R E Bennett, and G D Haertel</i>	56

Technology and Learning – Education Reform and Economic Development

National Strategies to Build a Technology Workforce	<i>H-M Cheah</i>	66
Public–Private Partnerships for Educational Reform	<i>V Quah, C P Lim, and C Brook</i>	73
Relating Technology, Education Reform and Economic Development	<i>R B Kozma</i>	81

Technology and Learning – Evaluating Technology

Examining the Effects of Educational Technology Programs: Challenges and Strategies	<i>M Russell</i>	88
---	------------------	----

Technology and Learning – Global Trends

Equity in Technology Access and Opportunities	<i>T Matuchniak and M Warschauer</i>	95
Internet-Based Education	<i>M Bakia</i>	102

Technology and Learning – Issues

Media Use and School Achievement	<i>K Roe</i>	109
Technology and Physical and Social Health	<i>K Subrahmanyam</i>	112

Technology and Learning – Meeting Special Students’ Needs

Universal Design for Learning *D H Rose and J W Gravel* 119

Technology and Learning – School Leadership for Technology Integration

Technology as a Support for School–Community Connections *K M Culp* 125

Technology and Learning – Supports for Managing Instruction

Classroom Uses of Technology to Manage Instruction *W R Penuel* 133

Technology and Learning – Supports for Skill Learning

New Media, Learning from *W Schnotz and H Horz* 140

Reading and Technology *N DeFrance, D Khasnabis, and A S Palincsar* 150

Technological Supports for Acquiring Twenty-First-Century Skills *C Dede* 158

Technology Supports for Acquiring Inquiry Skills *T de Jong* 167

Technology Supports for Acquiring Mathematics *M J Nathan* 172

Technology Supports for Lifelong Learning *A Collins and R Halverson* 184

Technology and Learning – Supports for Student Collaboration

Knowledge Communities in the Classroom *J D Slotta and H Najafi* 189

Technology and Learning – Supports for Subject Matter Learning

Technology Supports for Science Learning *C Delgado and J Krajcik* 197

Technology Supports for Second Language Learning *D Nunan* 204

Technology and Learning – Teacher Learning

Teacher Skills and Knowledge for Technology Integration *N Law* 211

Technology in Preservice Teacher Education *N Davis* 217

Technology Resources for Teacher Learning *J E Hinojosa, C Labbé, and L López* 222

V

Vocational Education and Training

An Overview of Vocational Education and Training *T Karmel* 229

Informal and Non-Formal Learning in Vocational Education and Training *F Beddie and S Halliday-Wynes* 240

School-Based Vocational Education and Training *B Knight and R Sweet* 247

The Economics of Vocational Education and Training *P Carneiro, L Dearden, and A Vignoles* 255

Vocational Education and Training – Individuals and Students

Age Equality in Education and Training *P Taylor* 262

Career Guidance in Vocational Education and Training and in the Workplace *P Plant* 269

Displaced Workers, Unemployed and Vocational Education and Training *B Kriechel* 275

Immigrant Investment in Host Country Vocational Education and Training *I Kogan* 280

Indigenous People and Vocational Education and Training	<i>R G Schwab</i>	289
Persons with a Disability and Vocational Education and Training	<i>F Ferrier and C S Smith</i>	294
Prisoner Education and Training	<i>S Dawe, S H Wynes, and J Misko</i>	300
Vocational Education and Training and the School-to-Work Transition	<i>G Hayward</i>	306

Vocational Education and Training – Industry and Employers

Apprenticeships	<i>E Smith</i>	312
Skills Shortages: Concepts, Measurement and Policy Responses	<i>C Shah and G Burke</i>	320
Training and Learning in the Workplace	<i>R L Jacobs</i>	328
Training in Small Businesses	<i>H Matlay</i>	337

Vocational Education and Training – Research on VET

Language of Vocational Education and Training	<i>J Misko, J MacKenzie, and R-A Polvere</i>	343
Research on Vocational Education and Training	<i>J Robson and B Bailey</i>	352
Vocational Education and Training Information Resources	<i>L-A Harris, R Naidu, and M Saunders</i>	358

Vocational Education and Training – Teaching and Learning

Employability Skills	<i>E Smith</i>	368
Learning Styles in Vocational Education and Training	<i>J Vorhaus</i>	376
Modularization in Vocational Education and Training	<i>H Ertl and G Hayward</i>	383
Personalized Learning and Vocational Education and Training	<i>I Duckett</i>	391
Recognition of Prior Learning and Vocational Education and Training	<i>J Keating</i>	397
The Competency Model	<i>A Gonczi and P Hager</i>	403

Vocational Education and Training – VET in Context

Globalization and Vocational Education and Training	<i>P Loveder</i>	411
Learning to Work for the Future	<i>J Fien, R Maclean, and D N Wilson</i>	416
The Relationship between Technical and Vocational Education and Training and Innovation	<i>K Ruth and L Deitmer</i>	423
The Status of Vocational Education and Training	<i>J Lasonen</i>	429
Vocational Education and Training in Post-Conflict Countries	<i>E Lyby</i>	441

Vocational Education and Training – VET System

Dual System	<i>T Deissinger</i>	448
Industry Involvement in the Vocational Education and Training System	<i>M Hughes and C Hughes</i>	455
Non-Completion in Vocational Education and Training: Issues, Patterns and Possibilities	<i>S Lamb and A Walstab</i>	463
Planning and Policy Development for Technical Vocational Education and Training Systems	<i>R Maclean and M Pavlova</i>	469
Qualifications Frameworks and their Role in the Reform of Education and Training	<i>M Young</i>	476
Quality Assurance in Vocational Education and Training	<i>D Gunning</i>	482

Training Finance	<i>A Ziderman</i>	489
Training Markets	<i>P Ryan</i>	496
Vocational Education and Training Teacher Training	<i>S Nielsen</i>	503
Vocational Education and Training Workforce	<i>D L Wallin</i>	513
Subject Index		519
Contributors		747

EDUCATION OF PROFESSIONALS

Contents

An Overview of Accreditation, Certification, and Licensure Processes

Educating America's Military Officers

Education for Architecture in the United States and Canada

Education for Medicine

Education for the Legal Profession

Education of Clergy

Framing Professions Education

Issues in Accounting/Business Education

Professions – Engineering

Professions – Health-Related

Professions – Pharmacy

Teaching for the Complex Practice of Nursing

Veterinary Education

An Overview of Accreditation, Certification, and Licensure Processes

J J Norcini and M van Zanten, Foundation for Advancement of International Medical Education and Research (FAIMER), Philadelphia, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The qualities or stages of development of a profession are: (1) the activity becomes a full-time job, (2) formal training is available, (3) the job becomes the subject of university study, (4) professional associations are established, (5) practitioners are licensed, and (6) a code of ethical practice is adopted. In describing the characteristics of a profession, some authors have included both a formal training mechanism and a means for testing the competence of practitioners. Although the definition of a profession has evolved steadily from these early efforts, some form of recognition, both of education and the individual professional, remains central. Accreditation, licensure, and certification are the most common forms of these recognitions.

The Centre for Educational Research and Innovation of the Organization for Economic Cooperation and Development (OECD) has recently reviewed the status of quality assurance and professional recognition around the world. The Centre's research demonstrates that most of the OECD member nations have these types of quality-assurance mechanisms in place and that many non-OECD members do as well, or are in the process of developing such systems. However, there is great diversity in the

nature and rigor of the processes employed and of the scope and governance of the regulatory bodies that oversee the recognition.

Given the diversity across countries, the purpose of this article is to broadly describe accreditation, licensure, and certification. Specifically, we define the terms, explain the benefits to a variety of different stakeholders, provide descriptions of some of the models currently in use, and review the evidence in favor of their effectiveness.

Accreditation

In the field of education in the professions, accreditation can be defined as a process by which a designated authority reviews and evaluates an educational institution or specific program using a set of clearly defined criteria and procedures. Accreditation of educational programs can be mandatory or voluntary, depending on the scope and authority of the organization. While there is diversity in the methods of implementation, most accreditation processes share the following approach:

1. a guided self-study is conducted by the institution;
2. an external review and site visit is conducted by a peer committee;

3. based on the self-review data and information gathered on site, an accreditation decision is made by the organization, or a recommendation is given to a governmental ministry for final approval;
4. decisions and reports are published and disseminated; and
5. the process is repeated at specified intervals, usually every 4–8 years.

Benefits of Accreditation

Independent third-party certifiers, such as an accreditation authority that reviews educational institutions, provide market signals to the public concerning the value of the product under review. Through a system of accreditation, educational institutions and programs can demonstrate objective evidence of quality. Within the professions, sometimes a certain degree of uniformity of study programs is required by governmental or guild-specific organizations. In settings where accreditation is mandatory, the public is ensured that all educational institutions are meeting a minimal level of required quality. Where accreditation is voluntary, benefits to an educational institution could include increased student applications, higher profits, better retention of current students, ease of student mobility, and an enhanced reputation. Individual students, especially those entering specialized professions, often carry a stronger legitimacy in the workforce if they are graduates of programs that have been successfully assessed against rigorous standards of quality.

In addition to promoting educational quality within a specific country, accreditation processes can allow for the direct comparison of professional training experiences and the mutual acceptance of degrees. A standardized process or recognition of transparent systems of education evaluation is increasingly useful as cross-border movement of professionals continues to rise worldwide. For example, the Washington Accord of 1989 set generic criteria, policies, and procedures for accrediting engineering academic programs. Countries that sign the Washington Accord formally agree to accept each other's engineering program accreditation decisions. In the field of medical education, regional cooperation of accreditation activities in various areas in the world, including the Caribbean, Central and South America, and the Gulf region, is evidence of a growing need to pool resources and facilitate the evaluation of foreign degrees. Some international accreditation systems have no legal standing, either nationally or internationally, but are used for comparison and benchmarking of similar professional education programs. An example of such international cooperation is the European Quality Improvement System (EQUIS), which has evolved to become a global system of quality assessment, improvement, and accreditation of higher education institutions in the field of management and business administration.

Models of Accreditation

Accreditation organizations can accredit educational institutions as a whole, or accredit specific programs, that is, law, business, nursing, etc. Some countries use both models, such as in the United States (US), where 19 institutional accrediting organizations accredited approximately 6300 institutions as a whole, and more than 60 programmatic accrediting organizations accredit approximately 17 500 programs. Some countries have chosen to adopt only one of these models. For example, in the Philippines, a voluntary accreditation system utilizes four separate accrediting organizations recognized by the Philippine Department of Education to evaluate only higher education programs, not institutions. Some countries use one organization employing a hybrid system, such as The Norwegian Agency of Quality Assurance in Education (NOKUT), which accredits both educational institutions and/or study programs.

Accreditation systems can vary in scope and nature of the authority charged with the quality review. In many countries around the world, the government is directly responsible for designing and implementing a process of educational institution quality assurance. For example, most accrediting bodies in Central and South America, such as those in Argentina, Colombia, and Ecuador, are governmental entities. In contrast, in other countries, autonomous, independently run organizations are accountable for ensuring education quality. The Japanese government requires all universities to undergo accreditation by a certified body, and the government accepts the accreditation decision granted by various independent agencies. The Japan University Accreditation Association (JUAA) and the National Institute for Academic Degrees and University Evaluation (NIAD-UE) both accredit Japanese educational institutions. In other instances, an independent agency is responsible for designing and implementing a quality-assurance process and making recommendations to a ministry to grant the official accreditation status. In the Czech Republic, the Czech Republic Accreditation Commission (CRAC) is an advisory body that evaluates higher education institutions and programs, and the Ministry of Education, Youth and Sports makes the final accreditation decisions. In the case of medical education programs, the country's Ministry of Health must also approve of the recommendation of the CRAC as a prerequisite to the final decision by the Ministry of Education. In India, the National Assessment and Accreditation Council (NAAC) voluntarily assesses institutions of higher education. While the NAAC process uses generic criteria and is voluntary, many professions in India require a review of educational programs by specific organizations. For example, the National Board of Accreditation (NBA) was created by the All India Council for Technical Education (AICTE) to evaluate technical programs, such as engineering and technology, architecture, and town and country planning.

Accreditation systems can also focus on either the process of the education, or on outcomes, or a combination. A process-focused approach measures an educational program's potential to effectively train students by determining compliance with accreditation standards. For example, an accreditation system utilizing a process model considers if the educational program has established a mission and objectives, employees suitably qualified instructors, uses an approved curriculum, and complies with resource requirements. In contrast, rather than focusing on a program's potential to educate professionals, an accreditation approach based on outcomes focuses on the students' or graduates' actual accomplishments. An outcomes-based accreditation system strives to measure the students' level of achievement of the learning objectives established by the program.

Development of Accreditation Systems

Accreditation systems are developed for various reasons and can be the result of influences from a variety of sources. In many countries, the development of accreditation systems has been the result of government laws aimed at ensuring quality in domestic educational institutions. In other instances, the government was not directly responsible for the evolution of accreditation systems. For example, in the US, at the end of the nineteenth century there was virtually no oversight of higher education, and quality and curriculum content varied significantly between institutions. Among the professions, knowledge and learning were becoming more specialized and numerous educational institutions were operating with wide variation in quality. In addition, the population was becoming more mobile and institutions needed a mechanism for evaluating the education of transfer students. This situation led to accreditation activities beginning simultaneously from professional organizations outside of the educational arena, and from within the educational institutions themselves. The advancement of applicable national or international standards has also been a key factor influencing the development of an accreditation system for various professions. Standards authorities provide sampling and testing protocols that can be used by accreditation authorities in their review process, and supply structure and guidance on appropriate benchmarks.

Evidence of Effectiveness

The general purpose of accreditation systems in professions education is to ensure that a minimum level of standards has been met, and frequently also to promote quality improvement activities within the educational institutions or programs. Despite the recent trends toward establishment and enhancement of these systems in the

global realm, much of the support for these activities tends to be limited to numerous positive anecdotal accounts regarding improvement. Traditionally, accreditation has been based on qualitative rather than quantitative assessments of educational program criteria, and therefore isolating the linkage between accreditation variables and appropriate outcomes can be difficult. As a result, it appears that few rigorous investigations have been conducted and published evaluating the effectiveness of accreditation processes. One study measuring the impact of a change in accreditation standards on engineering programs noted significantly higher levels in students' learning outcomes as a result of the new system. Evaluations of various criteria to determine the effectiveness of an accreditation system on postgraduate medical training programs in the field of cardiology have demonstrated positive findings, and also some inconclusive results. A recent systematic review of health-sector accreditation, which included some studies of educational programs, also noted both positive and inconsistent findings.

Recently, there has been an increased call for extended research in a variety of fields to link educational institution accreditation to quality indicators, and a request that standards be based on data which clearly demonstrates that schools which adhere to these benchmarks produce students who are more knowledgeable and skilled than their peers from nonaccredited institutions. In addition, within established evaluation systems, it is often not known which specific aspects of the accreditation protocols and procedures can be linked to improved quality. Therefore, while accreditation systems continue to become more numerous, and have well-established and publicly accepted face validity as a means of ensuring a minimum standard of quality, further research is necessary to specifically describe the value of these systems in terms of improved workplace skills of the graduates. For example, as is the case within many specialized healthcare fields, little research has been done to evaluate if accreditation processes improve the quality of nursing education, and most importantly, ultimately the longer-term outcome of improvement in the quality of patient care delivered by the nurse graduates.

Licensure and Certification

A license is formal permission from a government for an individual to engage in a particular occupation. Certification is the same, except that the recognition is usually granted by a nongovernmental entity. Registration, a term used extensively outside of the US, is comparable in function to both of these in that individuals must register with an organization for a particular occupation. For purposes of this article, we will use the terms recognition or professional recognition to refer to all of these concepts.

According to the Centre for Educational Research and Innovation, the nature of the institutions that grant professional recognition varies considerably from country to country. In some, they are governmental, in some they are independent, and in some both governmental and independent agencies (sometimes professional associations) operate, with the latter often holding higher standards. In addition to variation in the nature of the regulatory institutions, there is also diversity in both the definitions of particular professions and the standards held for them. This wide range of practice and concepts limits the degree to which generalizations can be drawn about professional recognition, thereby limiting the mobility of the professionals themselves.

Although there are exceptions, the regulatory institutions typically require that an individual completes an accredited or recognized educational process and passes an examination before acquiring a license or certificate. In some instances, a period of professional practice is also required. Historically, the license/certificate was granted in perpetuity, but in recent years a system of maintenance of certification is becoming increasingly common across a variety of professions.

Benefits of Licensure and Certification

The four chief stakeholders in professional recognition are clients (or patients), the professionals themselves, employers, and the profession as a whole. With a recognized professional, individual clients should experience improved quality of services, reduced risks of adverse events, and increased satisfaction. For the individual professionals, licensure and certification establish their credentials, improve their career opportunities, often increase their earnings, offer peer recognition, and reflect achievement and commitment.

Licensure and certification also have benefits for institutions. For employers, hiring recognized professionals serves as recruitment tool, demonstrates staff credibility, and anchors the hiring decision in an objective marker of competence. Moreover, it improves the quality and efficiency of the services provided to clients and enhances their satisfaction. For the profession itself, recognition defines the scope of competence and practice, acts as a barrier to entry, ensures high standards, and promotes professionalism.

Models of Licensure and Certification

There are several models of licensure and certification ranging from general recognition in a broad area to recognition in more focused aspects of practice. Often, licensure and registration convey privilege in a general field of activity and the institutions conferring it have

wide-ranging power over the profession. For example, the Institute of Chartered Accountants of India (ICAI) was created by statute and it (1) regulates the profession in India, (2) is responsible for professional development, and (3) acts in an advisory capacity to government and other institutions. The ICAI has multiple functions including setting the curriculum, providing the examination, taking disciplinary action against individuals when warranted, creating accounting standards, generating ethical standards, peer review, and post-qualification course development.

In contrast, US institutions that grant licenses or registrations often have a much more focused set of activities. This more restrictive role is intended to eliminate the potential conflicts of interest inherent in having multiple responsibilities (i.e., being involved in education, regulation, and professional advocacy all at the same time). For example, the National Council of Architectural Registration Boards (NCARB) is composed of the 50 jurisdictions in the US that grant licenses in the field. For certification, it requires a degree from an accredited institution, successful completion of an intern development program, and successful performance on a registration examination. It does not regulate the profession nor does it represent either the educational process or concerns of practicing architects.

Within broad disciplines, there is frequently a need to recognize focused areas of practice. For example, in teaching, the National Board for Professional Teaching Standards (NBPTS) in the US offers certification in 25 different areas, such as mathematics in early adolescence, mathematics in adolescence and young adulthood, and reading in early and middle childhood. Similarly, in the UK, the different medical specialties are recognized through separate royal colleges (e.g., the Royal College of Physicians and the Royal College of Surgeons).

Recognition at the level of specialties within broad disciplines is more frequently, although not exclusively, voluntary and carries with it no legal restrictions on the scope of practice. For example, NBPTS teacher certification was created as a voluntary process aimed at raising standards in the field. Likewise, certification in the medical specialties in the US is voluntary. Neither acts as a barrier to entry to practice for those who do not hold the certificate, although diplomates of these boards have advantage by virtue of the fact that they have demonstrated their expertise. In some instances, such as the UK Royal College of Physicians, membership is required to progress in training.

As fields mature and are driven by advances in science and practice, it is not unusual for subspecialties to develop. For instance, in the US, cardiovascular diseases constitute a subspecialty of internal medicine. To become certified in this field, physicians must first train and acquire certification in internal medicine before undergoing training and certification in cardiovascular diseases.

It is also not unusual for these fields to spawn their own subspecialties. Therefore, for example, cardiovascular diseases in the US now have their own subspecialties, including interventional cardiology and clinical cardiac electrophysiology. Further, different specialties sometime share the same subspecialty. In the US, sleep medicine is a subspecialty of family medicine, internal medicine, psychiatry and neurology, pediatrics, and otolaryngology.

Within a broad discipline, the development of specialties and subspecialties is influenced by a variety of scientific, social, political, and cultural factors. Consequently, the final results may vary across country and may not even be consistent within the same country. For instance, in the US, the medical specialties have tended to remain integrated within internal medicine, while the surgical specialties have tended to separate. This makes international comparisons difficult and it also has implications for the portability of these credentials.

Requirements

Although a variety of different requirements are used for professional recognition, most regulatory entities often require two or more of the following: (1) an approved educational experience, (2) a practice experience, and (3) successful examination performance. For example, NCARB requires the applicant to have earned a professional degree from an accredited program, satisfy intern training requirements, pass all parts of the examination, acquire licenses from a jurisdiction in the US, and then apply for certification. Similarly, the National Council of Examiners for Engineers and Surveyors (NCEES) has four requirements that must be satisfied to obtain a state license: (1) graduation from an accredited engineering program, (2) successful performance on a test of fundamentals, (3) a period of approved work experience, and (4) successful performance on a discipline-specific examination.

Evidence of Effectiveness

The development of evidence for the effectiveness of licensure and certification programs is a relatively recent event. Moreover, most of it focuses on the examination since it is typically the most discriminating step in the process. The evidence most often seeks to establish the relationship of certification test scores to the quality of the educational process, scores on other tests, and practice performance.

One example of this type of work is the research on examination scores used as part of certification in internal medicine and its subspecialties. These scores have positive relationships with the quality of undergraduate medical education. More importantly, they are sensitive to the presence and amount of formal training in a discipline and how well trainees perform during that training. They are

also responsive to the educational characteristics of the training experience, such as faculty–trainee ratio.

There are correlations between the certifying examination in internal medicine, which is composed of multiple-choice questions, and other methods of assessment including other written examination formats, oral examinations, essay examinations, video simulations, written simulations, computer-based simulations, and faculty ratings of competence. Naturally, these correlations are greater when the two measures focus on the same aspects of competence.

It is well established in medicine that larger practice volume is associated with better patient outcomes. There is evidence that scores on the certifying examinations in critical care medicine are correlated with practice volume as well. Most importantly, certification status and examination scores have reasonable relationships with actual outcomes in practice, such as mortality following acute myocardial infarction.

This relationship with practice performance has been found with other professions as well. Students of teachers certified by the NBPTS had higher test scores, greater learning gains, and they demonstrated deeper learning. Of note, these effects were of particular benefit to special needs and minority children.

Summary

This article has provided an overview of a number of issues pertaining to accreditation of educational programs and licensure and certification of individual professionals. Examples of current practices, various models, development of systems, evidence of effectiveness, and the need for further research have been described.

The report by OECD highlights the fact that most quality-assurance mechanisms in accreditation, licensure, and certification focus on national concerns and public institutions and practices. As a consequence, there is frequently no framework for formally coordinating efforts across borders. The effect of this lack of legally sanctioned regional or global cooperation frequently leaves the burgeoning e-learning offerings and the rapidly increasing numbers of for-profit schools unregulated, which are also often not under the jurisdiction of the currently operating national accreditation systems. In addition, it limits the mobility of professionals whose credentials neither have common meaning nor reflect comparable standards across borders.

While the creation of an international framework may have positive effects on the quality of education in the developed countries, it may also have unintended consequences for less-developed countries. For example, the issue of brain drain among numerous professions, such as physicians, nurses, and other health care providers, is of significance in those countries with the highest burden of

disease. Anything that enhances the mobility of professionals is likely to exacerbate this problem. It is important that the creation of regional or international frameworks of educational quality assurance and/or foreign-degree recognition also be accompanied by policies that ensure equitable access critical professional services.

Further Reading

- Bashore, T. M. and Wang, A. (2008). Recent accreditation council for graduate medical education initiatives have not been shown to improve the education of cardiology fellows. *Circulation* **118**, 532–537.
- Brennan, T. A., Horwitz, R. I., Duffy, F. D., *et al.* (2004). The role of physician specialty board certification status in the quality movement. *Journal of the American Medical Association* **292**, 1038–1043.
- Centre for Educational Research and Innovation (2004). *Quality and Recognition in Higher Education: The Cross-Border Challenge*. Paris: Organisation for Economic Co-Operation and Development.
- Committee on Evaluation of Teacher Certification by the National Board for Professional Teaching Standards, National Research Council (2008). *Assessing Accomplished Teaching: Advanced-Level Certification Programs*. Washington, DC: National Academies Press.
- Davis, D. J. and Ringsted, C. (2006). Accreditation of undergraduate and graduate medical education: How do the standards contribute to quality? *Advances in Health Sciences Education Theory and Practice* **11**, 305–313.
- Eaton, J. S. (2003). *The Value of Accreditation: Four Pivotal Roles*. Washington, DC: Council for Higher Education Accreditation.
- Greenfield, D. and Braithwaite, J. (2008). Health sector accreditation research: A systematic review. *International Journal for Quality in Health Care* **20**(3), 172–183.
- Johnson, T. J. (1972). *Professions and Power*. London: Macmillan.
- Joly, B. M., Polyak, G., Davis, M. V., *et al.* (2007). Linking accreditation and public health outcomes: A logic model approach. *Journal of Public Health Management and Practice* **13**, 349–356.
- Kuehn, B. M. (2007). Global shortage of health workers, brain drain stress developing countries. *Journal of the American Medical Association* **298**, 1853–1855.
- Kuvin, J. T. (2008). Accreditation Council for Graduate Medical Education initiatives improve the education of cardiology fellows. *Circulation* **118**, 525–531.
- Matthiesen, V. and Wilhelm, C. (2006). Quality outcomes and program evaluation in nursing education: An overview of the journey. *Quality Management in Health Care* **15**, 279–284.
- Newble, D., Jolly, B., and Wakeford, R. (1994). *The Certification and Recertification of Doctors: Issues in the Assessment of Clinical Competence*. Cambridge, UK: Cambridge University Press.
- Pinkham, F. O. (1955). The accreditation problem. *Annals of the American Academy of Political and Social Science* **301**, 65–74.
- Volkwein, J. F., Lattuca, L. R., Harper, B. J., and Domingo, R. J. (2007). Measuring the impact of professional accreditation on student experiences and learning outcomes. *Research in Higher Education* **48**, 251–282.
- Wilensky, H. L. (1964). The professionalization of everyone? *American Journal of Sociology* **70**, 137–158.

Educating America's Military Officers

M H Adamshick, US Naval Academy, Annapolis, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

The Professional Military Officer

From as early as the seventeenth century, the formal education of a professional corps of military officers in Western society can be attributed to two significant factors. One was the transformation of military leadership from members of the noble class to those non-noble people selected because of their specialized education and training. The other was the rise of the nation-state and the subsequent formation of large armies requiring military leaders with specialized education and training to accommodate the parallel advance in military technology and tactics. The professionalization of the military officer necessitated the creation of formal institutions of higher learning to equip military leaders to accommodate this technical and doctrinal transformation. This gave rise to the military schools and academies from the eighteenth century until the beginning of the nineteenth century in most important Western countries. The Russian Military Academy was founded in St. Petersburg in 1723; the Royal Military Academy of the United Kingdom was founded in 1741; the *Ecole Royale Militaire* was founded in France in 1751; and the Prussian *Kriegs-Akademie* in Potsdam was established in 1765. The nineteenth century saw the establishment of Sandhurst in Great Britain and West Point in the United States in 1802; Saint-Cyr in France in 1808; and the *Kriegs-Akademie* in Königsberg (Prussia) in 1810. Initially dedicated to professional indoctrination and training of military commanders, the education of military officers evolved in concert with the technological and scientific pace of the nineteenth century. The focus on leadership and military strategy gave way to an educational curriculum focused on scientific and engineering matters and eventually aligned with the traditional format of the university. In twentieth century Europe and in the United States, the education of military officers took on a very standardized pattern, even though there were some differences from country to country. The pattern exhibited the following:

1. The basic educational program was conducted in typically 2–4 years and primarily focused on doctrinal study and professional training.
2. The location for educating military professionals was segregated from similar educational places for nonmilitary students and was generally a stand-alone institution.
3. Admission to the institution was generally restricted to young males who had already earned a secondary

education in a civilian school. Most students were in their late teens or early 20s.

4. Selection criteria were highly formalized, based upon an assessment of a variety of factors, and were intended to preserve the quality and nature of the military professional.
5. Entrance to the institution gave members a military status granting the institution some social control over their new members.

Today, all US military officers begin their service by receiving a common commission from the president by swearing their allegiance to support and defend the constitution of the United States. The commission, which is similar for all military branches (the Army, Navy, Air Force, and Marine Corps), is both a letter of instruction and a grant of authority that commands the officer to “carefully and diligently discharge the duties of the office to which appointed . . .” The commission delineates that all commissioned officers, like all officials of the executive branch of government, serve at the pleasure of the president.

The profession of the military officer displays particular characteristics some of which, as suggested by Huntington, are expertise, responsibility, and corporateness, while for others, such as Janowitz, they are a core of skills that is difficult to master and of considerable social importance, autonomy or self-rule, an ethical code, and a system of compulsion. These two authors, considered the fathers of contemporary sociology of the military, believe that the military officer corps most closely represent the ideal type of profession. The professionalization of the US military officer corps begins at the candidate's commissioning source.

The US military commissions about 20 000 officers each year. Almost all officers are college graduates. About 40% of officers received their commissions through participation in Reserve Officer Training (ROTC) programs at civilian colleges and universities, about 22% through officer candidate schools (OCSs) or officer training schools (OTSs), and about 11% – primarily people with medical or legal training or the clergy – receive direct commissions. A significant minority (about 15%) are commissioned through three federal military academies: the US Military Academy at West Point, NY; the US Air Force Academy at Colorado Springs, CO; and the US Naval Academy at Annapolis, MD. Some Naval Academy graduates are commissioned to serve in the Marine Corps rather than the Navy.

Military Officer Development (Leader Education)

The traditional higher education academic experience for the officer candidate student is similar to their civilian counterparts in most aspects. Candidates choose their academic course of study and select a major within the guidelines of the ROTC scholarship program funding their enrollment. All must comply with the academic guidelines enforced by the institution as directed by accreditation standards. From an educational standpoint, what remains unique is the military leader development model students experience in each commissioning program. For example, the US military academies invest a considerable amount of their cadets' and midshipmen's years in education for leadership. Each academy endeavors to synthesize the entire experience of student activity into a leadership component that prepares the young adult for their subsequent military service. The experience of officer formation is different for each cadet and midshipman, although each institution endeavors to control this variation through standardized training and adherence to strict performance requirements. Variance occurs due to differences in student background, culture, trait personalities, intelligence, motivation, interests, activities, and a whole host of other variables that cannot be controlled for in the officer development model. Simply, students have different traits and experiences that may profoundly effect how they ultimately lead. What does remain constant for each cohort group, however, is a core academic program aligned with the broad leader development model. Core leadership instruction at the federal military service academies forms the academic bedrock of the leadership development continuum for each school. Generally, each program acknowledges that certain individual characteristics, attributes, or skills are associated with effective leader's performance under various conditions, and that these military institutions develop or cultivate these characteristics in future leaders. While the officer development experience is uniquely different at each commissioning source, the military services have a standard set of officer attributes that must be met by every candidate upon the completion of their educational program.

The Professional Military Educational Continuum

Pre-Commissioning

Military education received at institutions and through programs producing commissioned officers upon graduation is referred to as pre-commissioning, which focuses on preparing officer candidates to become commissioned officers within the Military Department that administers the pre-commissioning program. The curricula are

oriented toward providing candidates with a basic grounding in the US defense establishment and their chosen military service, as well as a foundation in leadership, management, ethics, and other subjects necessary in preparing them to serve as commissioned officers.

Primary

Education typically received at grades (officer) O-1 through O-3 subsequent to commissioning is considered as primary education, which focuses on preparing junior officers to serve in their assigned branch or warfare or staff specialty. The curricula are predominantly service oriented, primarily addressing the tactical level of war. Service schools that have programs centered on pay grade O-3 officers will foster an understanding of joint war-fighting necessary for success at this level. Joint learning areas are embedded in service professional military education (PME) instruction.

Intermediate

Education typically received at grade O-4 approximately 9–11 years after commissioning is referred to as intermediate education, which focuses on war-fighting within the context of operational art. Officers expand their understanding of joint force deployment and employment at the operational and tactical levels of war. They gain a better understanding of joint and service perspectives. Inherent in this level is the development of an officer's analytic capabilities and creative thought processes. In addition to continuing development of their joint war-fighting expertise, they are introduced to theater strategy and plans, national military strategy, and national security strategy and policy.

Senior

Senior education is typically received at grades O-5 or O-6, approximately 15–17 years after commissioning. Here, the focus is to prepare officers for positions of strategic leadership. Senior education focuses on strategy, theater campaign planning, the art and science of developing, and integrating and applying the instruments of national power (diplomatic, informational, military, and economic) during peace and war. Studies at these colleges should emphasize analysis, foster critical examination, encourage creativity, and provide a progressively broader educational experience.

General/Flag Officer

This constitutes the education received as a general/flag officer (G/FO). The number of commissioned officers

eligible to serve at this rank is determined by federal law and is a small percentage of the total force strength. The percentage of commissioned officers on active duty serving at the G/FO rank is approximately 4.3% of the total commissioned officer force (875 G/FOs out of approximately 203 000 commissioned officers). The promotion rate from O-6 to G/FO is about 2% across each military service. Courses within the G/FO level of the Joint Professional Military Education (JPME) continuum prepare senior officers of the US Armed Forces for high-level joint, interagency, and multinational responsibilities. These courses may address grand strategy, national security strategy, national military strategy, theater strategy, and the conduct of campaigns and military operations in a joint, interagency, and multinational environment to achieve US national interests and objectives. G/FO JPME is tiered to ensure the progressive and continuous development of executive-level officers.

Officer/Leader Development at the United States Military Service Academies

Broadly defined, the approach to studying leadership/officership at all three military service academies is based upon an experiential learning model comprised of conceptualization, experimentation, reinforcement, and reflection over a period of 4 years in residence. This is accomplished through a fairly common process of personal learning, classroom instruction, and interaction, complemented by the unique professional experiences and opportunities for reflection during the academic year and through summer training. The leadership education program consists of formal instruction by military and civilian professionals in leadership, management, philosophy and ethics, human behavior, and law; complemented by the practical knowledge and real-time experiences with operational service components. This relevant and effective combination of academic and professional expertise profoundly enhances the learning environment. This approach intends to transform the learning environment from an abstract study into a more pertinent and timely application of knowledge.

Core Officer/Leader Development Curriculum

All three military service academies enrol students in core academic courses throughout their 4 years of study in addition to the courses specific to their chosen academic major. While the academic core curriculum at each institution varies, the officer/leader development curriculum remains consistently aligned with academic courses in psychology, leadership, management, ethics, philosophy, and law. As major requirements and specializations allow, many students are afforded the opportunity to take

specialized electives in addition to the core requirements. The core provides a consistent 4-year program that allows officer candidates to study leadership, human behavior, ethics, law, and character, as well as individual, group, and organizational behavior. All will gain an understanding and appreciation of the values, culture, identity, and specialized knowledge encompassed by the military profession.

Leadership education

West Point requires all cadets to take general or advanced psychology in their freshman year and military leadership in their junior year. The focus of the introductory psychology course is the development of an awareness and understanding of one's own behavior and the behavior of others. Emphasis is placed on applying the behavioral principles learned to the cadets' current lives and their functioning as future officers. Military leadership is a multidisciplinary study of leadership in an organizational context which focuses on the integration of theory and practice. The cadet studies the leader's direct influence on individual motivation and group processes through the application of leadership theories, skills, and attributes. In addition, the cadet learns how to influence subordinates indirectly through organizational systems and procedures, organizational culture, and ethical climate.

Midshipmen at Annapolis are required to take principles of self-leadership and organizational dynamics in their freshmen year and theory and applications of leadership in their junior year. They begin the study of leadership in the context of theories and principles of individual and group behavior. The introductory course emphasizes the development and understanding of personal strengths, values, and opportunities for growth. The theory and applications course builds on the concepts introduced in the first year by examining the theory and research of the contingent and dynamic process of leadership. The course combines literature from the fields of social psychology, organizational behavior, and group dynamics to help students understand the factors that influence leadership in a military context.

At the US Air Force Academy, all cadets are required to take introduction to behavioral sciences in their freshman year, foundations for leadership development in their junior year, and management and command in their final year before graduation. The introductory course provides an introduction to the scientific study of behavior and mental processes across diverse levels of analyses. The leadership development course explores leadership development as a scientific study. Specifically, the course examines principles that will set students on a lifelong path of becoming a leader of character who treats others with respect and dignity. Management focuses on the successful techniques that allow people to understand and influence their environment.

Philosophy and ethics education

Sophomore cadets at the US Military Academy take a core course in philosophy designed to develop capacities to think clearly and critically. Sophomore Naval Academy Midshipmen take ethics and moral reasoning for the military leaders. This course is structured around classical and contemporary writings in moral philosophy. At the Air Force Academy, cadets take ethics in their junior year. Students study several major moral theories and their application to contemporary problems with special emphasis on the moral problems of the profession of arms.

All three schools use classical writings and contemporary case studies to illuminate the moral and philosophical complexities of the modern military professional. The students enjoy a faculty comprised of military officers and civilian philosophers and the curriculum differs on matters of emphasis regarding philosophical theory over moral reasoning and applied ethics.

Law education

West Point seniors enrol in constitutional/military law, a course that studies the US Constitution and the military justice system. Cadets acquire information and skills in order to recognize and resolve constitutional and legal problems. The course provides analytical models for dealing with problems regarding societal and military order. Midshipmen take law for the junior officer in their senior year. This course provides a survey of relevant legal topics applicable to the role of the future junior officer as a leader, manager, and decision maker. Students examine operational law concepts, including the law of armed conflict, rules of engagement, and the law of the sea. They also study the various types of military investigations, as well as the different types of disciplinary venues, such as nonjudicial punishment and courts-martial. Students are provided with an exposure to the various crimes enumerated in the Uniform Code of Military Justice, and the administrative discharge process. Cadets at the Air Force Academy are required to take law for Air Force officers in their sophomore year. This course introduces cadets to the legal knowledge and skills they will need as Air Force officers and educated citizens. The course examines the nature of law and its role in American society and the military and introduces substantive areas of the law that military officers likely will encounter in their personal and official capacities.

Professional Military Education

PME is the critical element in officer development and is the foundation of a joint learning continuum that ensures that the US Armed Forces are intrinsically

learning organizations. The PME vision understands that young officers join their particular service, receive training and education in a joint (interservice) context, gain experience, pursue self-development, and, over the breadth of their careers, become the senior leaders of the joint force. Performance and potential are the alchemy of this growth, but nothing ensures that they are properly prepared leaders more than the care given to the content of their training, education, experience, and self-development opportunities. Modern PME guidelines entail ensuring that military officers are properly prepared for their leadership roles at every level of activity and employment, thus also ensuring that the US armed forces remain capable of defeating present-day and future threats.

PME intends to build an officer who understands the strategic implications of tactical actions and the consequences that strategic actions have on the tactical environment. Service delivery of PME, taught in a joint context, instills basic service core competencies; joint PME instills joint core competencies. JPME should position an officer to recognize and operate in tactical, operational, and strategic levels of national security.

The Ronald W. Reagan National Defense Authorization Act of 2005 expanded the opportunities to receive JPME and established a link between joint officer development and JPME. The future joint force requires knowledgeable, empowered, innovative, and decisive leaders capable of succeeding in fluid and perhaps chaotic operating environments with more comprehensive knowledge of interagency and multinational cultures and capabilities. Officer Professional Military Education Policy (OPMEP) is at the heart of building those officers. This policy intends to maintain a force of dedicated, highly educated, and well-trained men and women capable of leveraging new ideas that will succeed in the complex and fast-paced environment of future military operations.

The military officer education continuum reflects the dynamic system of officer career education. It identifies areas of emphasis at each educational level and provides joint curriculum guidance for PME institutions. It is a comprehensive frame of reference depicting the progressive nature of PME and JPME, guiding an officer's individual development over time. The continuum structures the development of service and joint officers by organizing the PME continuum into five military educational levels: pre-commissioning, primary, intermediate, senior, and G/FO. It defines the focus of each educational level in terms of the major levels of war (tactical, operational, and strategic) and links the educational levels such that each builds upon the knowledge and values gained in previous levels. The continuum also recognizes both the distinctiveness and interdependence of joint and service schools in officer education. Service schools, in keeping with their role of developing service specialists, place emphasis on

education primarily from a service perspective in accordance with joint learning areas and objectives. Joint schools emphasize joint education from a joint perspective.

International Military Education and Training

The International Military Education and Training (IMET) program is a component of US security assistance that provides training on a grant basis to students from allied and friendly nations. Funding is appropriated from the International Affairs Budget of the Department of State. It is a key component of US security assistance that provides US training on a grant basis to students from allied and friendly nations. Students are exposed to US military procedures and the manner in which the US military functions under civilian control. They may receive as examples of types of training: PME as well as technical training. Overall objectives of the IMET program are:

1. to further the goal of regional stability through effective, mutually beneficial military-to-military relations, which culminate in increased understanding and defense cooperation between the United States and foreign countries; and
2. to increase the ability of foreign national military and civilian personnel to absorb and maintain basic democratic values and protect internationally recognized human rights.

The IMET objectives are achieved through a variety of military education and training activities conducted by the US Department of Defense for foreign military and civilian officials. These include: formal instruction involving over 2000 courses taught at approximately 150 military schools and installations; on-the-job training; orientation tours for key senior military and civilian officials; and mobile education teams which take the curriculum to the host country.

Further Reading

- Bass, B. M. (1990). *Bass and Stogdill's Handbook of Leadership: Theory, Research & Managerial Applications*. New York: Free Press.
- Caforio, G. (ed.) (2001). *The Flexible Officer*. Rome: Artistic and Publishing.
- Caforio, G. (2003). Military officer education. In Caforio, G. (ed.) *Handbook of the Sociology of the Military*, pp 255–278. New York: Kluwer Academic/Plenum Publishers.
- Chappell, K. (2008). ROTC on campus: Leadership development for a lifetime. *Ebony* **63**, 170.
- Day, D. V. (2001). Leadership development: A review in context. *Leadership Quarterly* **11**, 581–613.
- Holder, L. D. and Williamson, M. (1998). Prospects for military education. *Joint Force Quarterly* **18**, 81–90.

- Huntington, S. P. (1957). *The Soldier and the State*. Cambridge: Harvard University Press.
- Janowitz, M. (1960). *The Professional Soldier: A Social and Political Portrait*. Glencoe, IL: Free Press.
- Korb, L. J. (1976). *The System for Educating Military Officers in the U.S.* Pittsburgh, PA: International Studies Association.
- McCausland, J. D. (2008). *Developing Strategic Leaders for the 21st Century*. Carlisle, PA: Strategic Studies Institute.
- Moskos, C. C., Williams, J. A., and Segal, D. R. (2000). *The Postmodern Military: Armed Forces After the Cold War*. New York: Oxford University Press.
- Neiberg, M. S. (2001). *Making Citizen Soldiers: ROTC and the Ideology of American Military Service*. Cambridge, MA: Harvard University Press.
- Nuciari, M. (2003). Military academies, cadets and officer training. In Callaghan, J. and Kernic, F. (eds.) *Armed Forces and International Security: Global Trends and Issues*, pp 321–328. Piscataway, NJ: Transaction Publishers.
- Packard, G. A. (1999). *Longitudinal Study of the Social Network Influences on the Leadership and Military Development of Cadets at the U.S. Air Force Academy*. Chapel Hill, NC: University of North Carolina.
- Reed, G., Bullis, C., Collins, R., and Paparone, C. (2004). Mapping the route of leadership education: Caution ahead. *Parameters Autumn 2004*, 44–60. <http://carlisle-www.army.mil/usawc/Parameters/04autumn/reed.htm> (accessed August 2009).
- Segal, D. R. and Segal, M. W. (2004). America's military population. *Population Bulletin* **59**, 8–9.
- Snider, D. M. and Watkins, G. L. (eds.) (2002). *The Future of the Army Profession*. New York: McGraw-Hill.
- Sweatt, O. (1997). *Leadership as Teachership*. Carlisle Barracks, PA: US Army War College.
- US Department of Defense (2007). *The Armed Forces Officer*. Washington, DC: National Defense University Press.
- US Department of Defense. Office of the Inspector General (1993). *Joint Professional Military Education: Inspection Report*. Arlington, VA: Office of the Inspector General.
- US Department of Defense. Office of the Inspector General (1998). *Joint Professional Military Education: Phase II, JPME Evaluation Report*. Arlington, VA: Office of the Inspector General.
- Watson, C. A. (2007). *Military Education: A Reference Handbook (Contemporary Military, Strategic, and Security Issues)*. Westport, CT: Praeger Security International.

Relevant Websites

- <http://www.strategicstudiesinstitute.army.mil> – Agile Leaders, Agile Institutions: Educating Adaptive and Innovative Leaders for Today and Tomorrow, Strategic Studies Institute of the US Army War College.
- <http://www.usafa.af.mil> – Air Force Academy.
- <http://www.cpms.osd.mil> – CPMS Joint Leader Development Division.
- <http://www.strategicstudiesinstitute.army.mil> – Developing Strategic Leaders for the 21st Century, by McCausland, Strategic Studies Institute (SSI).
- <http://www.dtic.mil> – Doctrine Networked Education and Training (DOCNET).
- <http://c21.maxwell.af.mil> – Global War on Terrorism: Understanding the Long-Term Strategy – Why Education is Key, by Caslen, MECC Briefing, 3 Feb 05.
- <http://www.dtic.mil> – JCS Library and Research, Defense Technical Information Center (DTIC).
- http://www.jfsc.ndu.edu/current_students/documents_policies/documents/jsogpub_1_2000.pdf – JFSC Pub 1, The Joint Staff Officer's Guide 2000.
- <http://www.jdtc.jfcom.mil> – Joint Deployment Training Center (JDTCC).
- <http://www.dtic.mil> – Military Education Coordination Council (MECC), Defense Technical Information Center (DTIC).

<http://www.dtic.mil> – Military Education Home Page, J-7 of JCS, Defense Technical Information Center (DTIC).

<http://www.usna.edu> – Naval Academy Preparatory School, United States Naval Academy.

<http://www.au.af.mil> – Naval War College Library bibliography on Professional Military Education (PME), USAF Air University Public Portal.

<http://www.fas.org> – PME 2020, Future of Military Education, Federation of American Scientists (FAS).

<http://www.nwc.navy.mil> – United States Naval War College.

<http://www.nadn.navy.mil> – United States Naval Academy.

<http://www.dtic.mil> – US and International Intermediate and Senior JPME Equivalent Institutions, Defense Technical Information Center (DTIC).

<http://carlisle-www.army.mil> – US Army War College.

<http://www4.law.cornell.edu> – US Code Collection: Title 10, Chapter 107, Professional Military Education, Legal Information Institute at Cornell Law School.

<http://www.jwfc.jfcom.mil> – US Joint Forces Command, Joint Warfighting Center a Joint Training Directorate.

<http://www.usma.edu> – US Military Academy.

Education for Architecture in the United States and Canada

S Schreiber, University of Massachusetts, Amherst, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Architecture is one of the key professions involved in shaping the built environment and urban space. Architectural education has the primary purposes of producing competent, creative, critically minded, and ethical professional designers/builders who contribute to the social, economic, and cultural development of society – both nationally and globally. To achieve those purposes, it is an interdisciplinary field that comprises several major components: humanities, social and physical sciences, technology, and the creative arts.

Institutionalized architecture education developed in the second-half of the nineteenth century in response to broadening objectives for higher education, increasingly complex building systems requiring the expertise of specialized professionals, and new knowledge from related fields. Since then, architecture programs have flourished in a wide range of academic settings.

Regulation of the profession of architecture, including the licensing of practitioners, is a function of each US state/territory or Canadian province/territory exercising its power to protect the health, safety, and welfare of the people. Each jurisdiction registers architects by its own set of requirements. The requirements are generally consistent from jurisdiction to jurisdiction, but there are variations. The requirements to become an architect are often described as a three-legged stool, which includes:

- education (usually an accredited degree),
- experience (structured internship), and
- examination (the architect registration exam).

These requirements can often be completed concurrently. For example, some of the internship can be completed during school, and in many jurisdictions, some of the examinations can be completed during the internship.

All jurisdictions have an education requirement which may be satisfied by earning a professional degree in architecture accredited by the National Architectural Accrediting Board (NAAB) in the United States or the Canadian Architectural Certification Board (CACB) in Canada. Many registration boards requiring a professional degree in architecture from an NAAB/CACB-accredited program also accept other education assessed as equivalent (e.g., a professional degree in architecture from a foreign institution). Not all boards require a professional degree from an accredited program to satisfy their education requirements. Some boards require a pre-professional

degree in architecture, while others require a bachelor's degree in any subject.

The entire path to licensure typically requires 8–10 years – 5–7 years in school plus a 3-year internship. The internship years are spent as a salaried employee in an architectural or related practice working under the supervision of registered professionals.

The National Council of Architectural Registration Boards (NCARB) in the United States and the Committee of Canadian Architectural Councils (CCAC) have established standards and criteria that most licensing boards have adopted as their standard for admission to licensing examinations. Satisfaction of these requirements facilitates initial and subsequent reciprocal licensure.

The three-step path to architecture licensure – education, experience, and examination – is similar in many countries. However, in some countries, including many in Latin America and Europe, the completion of the professional architecture degree is equivalent to earning an architecture license. In this system, passing the capstone project is the same as passing a licensing exam.

Discussions regarding reciprocity between countries with similar licensing systems (such as the one that exists between the United States and Canada) have been promising and may lead to an easy transfer between countries in the near future. In 2008, accreditation and validation agencies from Australia, Canada, China, Korea, Mexico, the United States, and the Commonwealth Association of Architects announced the ratification of an agreement (the Canberra Accord) declaring substantial equivalency of professional degrees in architecture covered by their accreditation/validation systems. It is anticipated that the Canberra Accord will facilitate international mobility of graduates in architecture and contribute to improving the quality of architectural education through benchmarking.

Education

Architecture is recognized as a broadly based academic subject studied in its own right and as a professional discipline that leads to licensure. In most schools of architecture, these two concepts are richly interwoven and are reflected in the degree options for students.

Architecture programs fit into one of the four following categories:

1. *Pre-professional architecture degree.* This term refers to architecturally focused 4-year degrees that are not

professional degrees (and thus cannot be accredited). These degrees have titles such as BS in architecture, BS in architectural studies, BA in architecture, bachelor of environmental design, bachelor of architectural studies, etc. The amount of architecture work in the program varies and will determine the length of time required to complete further professional architecture studies.

2. *Professional (accredited) architecture degree.* A professional architecture degree is one that is accredited by NAAB or CACB. Accredited degrees are required by most jurisdictions for licensure as an architect. In the United States, these degrees are either the bachelor of architecture (BArch) degree or the master of architecture (MArch) degree, or the doctor of architecture (DArch). In Canada, the only accredited degree is the master of architecture.

The bachelor of architecture normally requires at least 5 years. The master of architecture requires from 1 to 5 years depending on the individual student's previous education. When the master's degree follows a 4-year, pre-professional architecture degree, it represents the two in the term four-plus-two program, and is the final portion of the professional phase of the study program. The doctor of architecture may require 7 years to complete.

3. *Post-professional architecture degree.* This is a graduate degree offered to students who already have a professional degree in architecture. Such degree programs are not accredited by NAAB and CACB. The degrees may be in highly specialized areas of study such as design theory, healthcare facilities, preservation, interior design, solar design, etc. This type of degree can be either a master's degree or, in a few cases, a PhD or other doctorate.
4. *Nonprofessional graduate architecture degree.* This title refers to graduate degrees in architecture offered to students who do not have a professional degree and wish to pursue nonprofessional graduate work in architecture. These programs vary widely in duration and degree title.

Professional (Accredited) Architecture Degrees

The NAAB and the CACB are the national architecture accrediting bodies in the United States and Canada, respectively. The board of directors of the NAAB and the CACB include representatives from architecture schools, professional societies, licensing authorities, and students. The NAAB and CACB operating under similar procedures determine whether architecture programs meet certain minimum education criteria including faculty qualifications, physical facilities, budget, curricula, course content, and program goals.

Only specific professional degree programs, not architectural schools, are accredited. A professional degree program is accredited by the same standards whether it leads to a bachelor of architecture, a master of architecture degree, or a doctor of architecture. Most schools offer only one or two accredited architectural degrees, but may have other related degree programs. For example, a school of architecture may offer a program in historic preservation or architectural engineering as well as the accredited professional architecture degree program. There are about 130 schools offering professional architecture degree programs in the United States and Canada.

Every architecture program has its special features and unique philosophy. Some schools offer many options while others have a more defined program. A single school may offer several accredited professional degree programs – for example, it may offer both a 5-year undergraduate degree for high school graduates, and a 3.5-year graduate degree intended for people who already have a degree in another field.

There are several usual paths to obtaining the professional degree in architecture:

- *Bachelor of architecture.* The 5-year bachelor of architecture degree requires a minimum of 5 years of study (equivalent to 150 credit hours minimum) and a 3-year internship (some of which can be completed in the fourth and fifth years of the program).

A 5-year program provides the quickest route to satisfy the academic requirements for licensure. Many bachelor of architecture programs begin with a concentration of architecture courses in a fairly prescribed manner, although some schools begin with a general course of instruction. Electives tend to be few and exposure to other fields limited.

Many bachelor of architecture programs, however, have devised curriculum structures which allow for considerable flexibility. These programs are broken into yearly components of 2 + 3, 1 + 4, 3 + 2, and 4 + 1, with entry and exit points from the various phases of the full 5-year program. In most schools, the student's work is carefully reviewed before advancement to the next phase. Such points provide a relatively easy means of transfer into an architecture program; transfer to another institution, particularly between degrees; or transfer to another academic discipline. The logical break points may also be used by the faculty to transfer out students who have demonstrated little progress. The early segments of the curriculum mix pre-professional design courses with general education and often serve as a common base for several different environmental design disciplines such as architecture, landscape architecture, industrial design, graphic design, etc.

Most professional degree programs also accept transfer students at the designated break points, but transfer

credit is usually evaluated on an individual course-by-course and case-by-case basis. While the structure of some programs makes it relatively easy for a professional school to accept transfer students from both junior colleges and other colleges, many feeder programs cannot match the academic depth, library and faculty resources, and extracurricular activities of the professional schools.

The CACB does not recognize the bachelor of architecture.

- *Master of architecture (5.5 years).* The 5.5-year master of architecture degree path may be entered directly from high school requiring about 168 credits and following a curriculum that is similar to that of a bachelor of architecture program. Students must be expressly accepted into the university's graduate school in order to pursue the final phase of this degree. Those who do not qualify for graduate study have the option of completing a nonprofessional undergraduate degree, with the possibility of pursuing the master's degree in one of the more traditional paths at another institution. Several schools are planning to convert their BArch programs over the coming years to MArch programs.
- *Pre-professional degree plus master of architecture.* The pre-professional degree plus master of architecture degree normally requires 6 years to complete and a 3-year internship (some of which can be completed during the fourth of the undergraduate program and during the graduate program). The flexibility in the program is readily apparent. At the end of 4 years, the student has a college degree and may decide to continue in architecture and get the professional master's degree, spend 1 or 2 years working for an architect, or change disciplines and pursue study in other design-related fields. Or the graduate may decide at this point to shift careers completely and/or seek an advanced degree outside the design field.

Pre-professional programs are not professionally accredited and vary widely with respect to title, emphasis, electives, requirements, and specific architecture offerings. They are, however, preparatory for advanced architectural or other environmental design fields. The 4-year pre-professional program may be subdivided into two phases, such as 2 + 2, or 1 + 4. The pre-professional programs provide time for experiencing a wider range of subjects, allowing the maturing student a better opportunity to make career choices. Ideally, the extra courses in the humanities and social sciences will give students a broader background from which to start their professional education.

For those who ultimately receive advanced degrees in design areas other than architecture or in nonarchitecture subjects, the 4-year degree may be preferable to the 5-year professional program in minimizing coursework and time. One other advantage of a four-plus-two program is the potential for earning the professional degree

at an institution other than where the undergraduate work was completed.

The graduate degree component of the four-plus-two path is the professional NAAB- or CACB-accredited degree. This degree is most appropriate for students who have a 4-year, pre-professional undergraduate degree in architecture. The course of study generally takes 2 years; however, at some schools, up to 3 years is required. These programs are designed to provide the professional education of the student as well as provide an opportunity for independent and creative exploration.

- *Nonarchitecture degree plus master of architecture.* The non-architecture degree plus master of architecture degree requires 7–8 years of study (a 4-year undergraduate degree plus a 3.5-year master of architecture degree), and a 3-year internship (some of which can be completed during the graduate program). The immersion into architecture is quick and very intense in this program. Some schools provide all of the education at the graduate level, while others will admit degree holders into their professional master's program with deficiencies. This means that preparatory undergraduate coursework must first be successfully completed before formal admittance to the graduate program. However, like the 2-year master of architecture degree, the 3.5-year MArch may be a fully NAAB- or CACB-accredited professional degree program, and successful graduates have professional education credentials equal to those with a BArch or other MArch.
- *Doctor of architecture.* The doctor of architecture was recently recognized by NAAB as a professional degree. It requires either an undergraduate baccalaureate degree or a minimum of 120 undergraduate semester credit hours, or the equivalent, and a minimum of 90 graduate-level semester credit hours, or the equivalent, in academic coursework in professional studies and electives.

Characteristics of Professional (Accredited) Architecture Programs

Students and faculty

In 2007, approximately 30 000 students were enrolled in NAAB- and CACB-accredited programs. That year, about 2600 students earned NAAB-accredited bachelor of architecture degrees, 2640 earned NAAB-accredited master of architecture degrees, 330 earned CACB-accredited master of architecture degrees, and 11 earned NAAB-accredited doctor of architecture degrees.

There are approximately 2200 full-time architecture faculty in the United States and Canada, and about 3300 part-time faculty. About 73% of the faculty members are women and 79% are white. About 22% of faculty members have PhDs with most holding the master of architecture (or equivalent) which is considered the terminal degree in the discipline.

Admissions

Admissions standards for undergraduate architecture programs vary widely. At many universities, students must be admitted to both the university and the architecture program. Some programs will admit students only as sophomores or juniors. Many programs, particularly bachelor of architecture programs, require a portfolio for admission. Most master of architecture programs require the Graduate Record Exam, a portfolio, completion of undergraduate calculus and physics, and architecture history.

Coursework

The structure and content of curricula of architecture programs in the United States and Canada are reasonably consistent. As accreditation does not distinguish between undergraduate and graduate degrees, bachelors programs can appear to be similar to masters and doctoral programs. The curriculum of NAAB- and CACB-accredited degree programs includes professional studies, general studies, and electives. To gain and retain accreditation of its degree program, each institution must both develop a program specific to its mission and also educate students to be knowledgeable and capable of producing work that can be measured by, and satisfy, specific performance criteria. The content of the professional studies can be divided into five main areas: design and graphics, technology, history and theory, and practice.

Design and graphics

The core of professional architecture programs, in both credit hours and time, is design which is almost always taught as a studio course. In most programs, design is a required course every semester, with an emphasis on actual or hypothetical architectural problems. Professional programs generally endeavor to educate students in the contextual, technical, and pragmatic aspects of architectural design.

Students work individually or collaboratively with a faculty critic. Design solutions are graphically (and often verbally) presented to juries of faculty and professionals who discuss and evaluate the student solutions. Ideally, knowledge from other courses is applied in the design studio.

Typically, there are 12–18 students for each design-studio critic. The studio critic may operate independently, doing his or her own projects and following his or her own schedule. The studio may be part of a coordinated effort involving several sections, an entire level, or even the entire school. The assignment typically includes a building type; a list of requirements that the building must satisfy; a description of the site on which the building will be located; and a statement and/or reading list that either explicitly describes or implicitly suggests the pedagogic motives and goals of the critic.

Studio courses have the greatest amount of student–teacher contact in the curriculum, and more than most other courses in any university. They are ordinarily four to nine credit courses and meet 3 or 4 days each week for 2–4 h per day. Part of this time is spent working independently, part is spent getting critiques of work at the desk, and part is spent evaluating work in group reviews or pinups. Conversation, not lecture, is the prevailing pedagogical mode, and the students must make decisions about what kinds of information they will use to address the problem.

In the first year of design (foundation), projects are usually abstract and conceptual. Students are introduced to visual composition in two and three dimensions, concepts of spatial and functional organization, circulation, and structural behavior. Beginning studios often focus on nonbuilding spatial context; beginning-user consciousness with a familiarity of spatial analysis, design-process methodology, and development of communication skills; and design literacy. The foundation sequence continues with greater emphasis on the environment, user-space study, and further skill development; introduction of qualitative technical materials; a minimum proficiency in the design and communication of simple buildings with an introductory understanding of construction and structural systems; and data analysis, programming, site analysis, and design.

Students then advance into higher-level design studios, where they explore design projects with more complicated site and program requirements and opportunities for more sophisticated investigations. Intermediate studios might focus on simple and complex building case studies with qualitative technical input and a general proficiency in the complete design of simple buildings with a minimum ability to deal with complex buildings and multi-building complexes, site analysis, and design. Advanced studios might focus on the synthesis of complex building and multi-building complexes within the urban context; integration of technical information; general proficiency in the total synthesis of complex buildings and related systems; and transportation, communication, life-safety systems, and social ramifications of planning and architecture.

The last semester of design frequently involves an independent capstone project. This may require research and preparation in the preceding semester, during which the student selects a topic or project and gathers the data needed to undertake the design. Most capstone work is done independently, unlike the preceding design studios. These projects often emphasize complex building design, planning, and urban design, involving mastery of data collection, analysis, programming, planning, building design, structures, building systems, landscape design, and other related knowledge. Students in the capstone year normally have an adviser or advisory committee of faculty who may review the student's progress.

Almost all architectural schools have adopted minimum standards of competency and performance that students must meet before advancing to the next level of design studio or before graduation. Since the design studio sequence is continuous throughout most programs, it is not unusual to find studio sections with students who did not all begin architectural school at the same time.

Typically, several nonarchitectural disciplines play key roles in determining architectural solutions – the behavioral sciences, engineering (structural and mechanical), and economics, to name a few of the more obvious. While some schools have made concerted efforts to teach these and other disciplines in an integrated studio situation, in many programs the actual instruction is still provided in discrete courses – sometimes in the architecture department, sometimes in other colleges.

A typical architecture program will recognize the importance of graphic and communication skills – instruction is given in freehand drawing, digital media, and graphic delineation.

Technology

Architecture students must have an understanding of building science and the accumulated experience of construction. Courses typically focus on structural systems, environmental control systems (building elements that pertain to the modification of the microclimate for purposes of human use and comfort), construction materials, and assemblies (the characteristics of building materials and how they are used, made, and applied in a building project).

History/theory

All architecture programs require the study of the history of architecture to, including Western and non-Western, national and regional traditions. Topics include historical movements in architecture, history of architecture, history of art and architecture, history of building technology, and theory of architecture. Architecture students must learn to see how buildings function, and how the future is built on the basis of the current understanding of the past. Analyzing architectural precedents, coupled with reading the many texts on architecture theory written by historians and architects, helps situate architectural design in an historical context.

Practice

Practice courses are concerned with the protocols of the profession including project process (the entire range of activities involved in a typical architectural design project as it moves from inception through completion of construction), project economics (the financial aspects of building, including the economics of development), business management (the concepts, ethics, and procedures

related to different forms of organization for architectural practice, including private and corporate offices as well as public-sector organizations and agencies), and laws and regulations (the body of common law, legislation, and regulation, including rules of professional conduct, that affect architectural practice).

General education and electives

Interspersed with architectural or technical support courses are general education and elective courses. A program may require the student to select a directed sequence of courses in one field as well as attempt to provide a diverse exposure to some of the many studies important to architecture, such as psychology, philosophy, history, geography, economics, urban studies, literature, sociology, and political science.

Assessment

Since 1984, the NAAB began requiring student-outcome assessments for all programs of architecture degrees. Accredited programs in the United States and Canada must now demonstrate student proficiencies in a wide range of areas. The most significant is comprehensive design, in which each student must demonstrate an ability to produce a comprehensive architectural project based on a building program and site that includes development of programmed spaces demonstrating an understanding of technical systems, life-safety provisions, and the principles of sustainability.

Experience

One must complete training requirements in structured internship programs to become licensed architects in the United States and Canada. Comprehensive internship programs are designed reinforce the knowledge, integrity, judgment, skills, discipline, and quest for learning that must serve the architect for a lifetime. Interns can gain experience from working in small architecture firms, large firms, engineers, contractors, and other professionals in firms across the country and around the world. The programs are designed to take about 3 years and can begin prior to graduation when the appropriate requirements are fulfilled.

In most US states, this training requirement is fulfilled by completing the intern development program (IDP). The IDP is offered by the NCARB. In Canada, the internship in architecture program (IAP) is a comprehensive program administered by the ten provincial associations of architects. Many schools have integrated internships into their curricula, while others require some work experience to graduate.

Examination

The Architect Registration Examination (ARE) is designed to determine whether applicants for architectural licensure possess sufficient knowledge, skills, and abilities to provide professional services while protecting the health, safety, and welfare of the general public. Each of the divisions of the ARE is designed to test for minimum competency in a specific area important to the protection of the public. NCARB develops the ARE specifications through extensive practice analyses. The process is further refined by cut-score studies and regular monitoring. Pass rates on the exam are regularly posted. The ARE, which is offered in French and English, and in metric and feet/inches, is accepted by all United States and Canadian licensing jurisdictions.

Continuing Education

Most states and provinces require continuing education for architects to maintain licenses. A typical requirement is 12 contact hours per year in subjects beyond the basic knowledge of the discipline, and particularly in areas focusing on the health, safety, and welfare of the public.

Further Reading

- AIAS Studio Culture Task Force (2002). *The Redesign of Studio Culture*. Washington, DC: American Institute of Architecture Students.
- Anthony, K. (1991). *Design Juries on Trial*. New York: Van Nostrand Reinhold.
- Anthony, K. (2002). Designing for diversity: Implications for architectural education in the twenty-first century. *Journal of Architectural Education* 55, 257–267.
- Association of Collegiate Schools of Architecture Press (2004). *Guide to Architecture Schools*. Washington, DC: Association of Collegiate Schools of Architecture Press.
- Bloomer, J. (1998). Teaching by design: The intimacies of architectural study. *The Chronicle of Higher Education – Chronicle Review*, cover story, **September 11**.
- Boyer, E. and Mitgang, L. (1996). *Building Community: A New Future for Architecture Education and Practice*. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.

- Cuff, D. (1991). *Architecture: The Story of Practice*. Cambridge: MIT Press.
- Goldhagen, S. (2003). Our degraded public Realm: The multiple failures of architecture education. *The Chronicle of Higher Education – Chronicle Review* 5.
- Lewis, R. K. (1998). *Architect? A Candid Guide to the Profession*. Boston, MA: MIT Press.
- National Architecture Accrediting Board (2004). *Conditions for Accreditation*. Washington, DC: NAAB.
- Nicol, D. and Pilling, S. (eds.) (2000). *Changing Architectural Education: Towards a New Professionalism*. London: Spon.
- Piotrowski, A. and Williams Robinson, J. (2001). *Discipline of Architecture*. Minneapolis, MN: University of Minnesota Press.
- Pollak, M. (1997). *Education of the Architect: Historiography, Urbanism and the Growth of Architectural Knowledge*. Cambridge: MIT Press.
- Schon, D. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.
- Schon, D. (1984). The architectural studio as an exemplar of education for reflection-in-action. *Journal of Architectural Education* 38, 2–9.
- Schreiber, S., Dulaney, R., and Lyn, F. (eds.) (2003). *Finishing School: Inquiries into the Completion of Architecture Education*. Tampa, FL: University of South Florida.
- Solomon, N. (2008). *Architecture: Celebrating the Past, Designing the Future*. New York: Visual Reference Publications.
- UIA Education Workgroup (2005). *UNESCO/UIA Charter for Architectural Education*. Paris: International Union of Architects (UIA).
- Waldrep, L. (2006). *Becoming an Architect: A Guide to Careers in Design*. New York: Wiley.
- Wasserman, B., Sullivan, P., and Palermo, G. (2000). *Ethics and the Practice of Architecture*. New York: Wiley.
- Weatherhead, A. (1941). *The History of Collegiate Education in Architecture in the United States*. PhD Dissertation New York: Columbia University.
- Wilkes, J. (ed.) (1988). *Encyclopedia of Architecture: Design, Engineering and Construction*. New York: Wiley.
- Wright, G. and Parks, J. (eds.) (1990). *The History of History in American Schools of Architecture, 1865–1975*. New York: Temple Hoyne Buell Center for the Study of American Architecture.

Relevant Websites

- <http://www.aia.org> – American Institute of Architects.
- <http://www.aias.org> – American Institute of Architecture Students.
- <http://www.acsa-arch.org> – Association of Collegiate Schools of Architecture.
- <http://www.cacb-ccca.ca> – Canadian Architecture Certification Board.
- <http://www.carnegiefoundation.org> – Carnegie Foundation for the Advancement of Teaching.
- <http://www.uia-architectes.org> – International Union of Architects (UIA).
- <http://www.naab.org> – National Architecture Accrediting Board.
- <http://www.ncarb.org> – National Council of Architecture Registration Boards.
- <http://www.raic.org> – Royal Architecture Institute of Canada.

Education for Medicine

M B Anderson, Association of American Medical Colleges, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The goal of medical education is to produce physicians who are prepared to serve the fundamental purposes of medicine and who have a profound sense of service and commitment to their patients. Physicians must possess the attributes that are necessary to meet their individual and collective responsibilities to society, to their patients, their patients' families, and to the profession. The goal of medical education is universal and the issues faced by medical educators throughout the world are the same. While the goals are the same, there are many differences in the approaches taken to achieve those goals. Medical education is inextricably intertwined with the culture of the country, the expectations of physicians in each country, the economy, the health care system, and the structure of the educational programs in each country. In addition to differences in curriculum and pedagogy, governmental regulations for certification and licensure of physicians as well as the accreditation of educational programs are different in every country.

Medical schools teach medical students, of course. But they do much more. Medical schools and teaching hospitals also train medical residents during the 3–7-year period following the MD degree (postgraduate education), which leads to eligibility for licensure and board certification in many countries. They are involved in continuing medical-education programs, too, which allow physicians to stay current with the fast-moving advances in diagnostic capability and therapeutic techniques and with the expanding knowledge on medical-intervention outcomes. Medical schools also are committed to the education of biomedical scientists through PhD training programs.

Medical education is undergoing significant changes worldwide, particularly in the United States. These changes are a response to increased scientific knowledge, differences in the way that students learn, increasing regulatory and governmental pressures, and changing expectations of society and the patients cared for by physicians.

Each of these elements is considered in the following pages, with the recognition that the goal of medical education is the same, while the mechanisms to educate and license physicians to practice differs according to the country and the system in which the education occurs. The changes underway in US medical education are both reflected in changes underway in other countries as well as informed by a network of international medical educators.

Admission to Medical School

Admission to medical school in the United States and Canada is a selective process, which contrasts with the open enrolment policies of many other countries. Selective admission allows medical schools to admit women and men who, in the faculty's opinion, have the academic abilities and personal qualities requisite for a profession based on high standards of competence and service to others. By retaining the prerogative to select their students, medical school faculty also can ensure that the number of enrolling students matches available resources. In countries with an open enrolment, the class size is reduced gradually during the first and second years of medical school through attrition and knowledge examinations.

With the exception of one US medical school, (University of Missouri-Kansas City) all students entering medical school are required to have a baccalaureate degree, while most countries outside of the United States and Canada admit students directly from secondary school (US high school) and follow a 6–8-year program of medical education. There are several BA–MD programs where students enter the program from high school and earn their bachelor's degree and MD in 6–8 years.

In the United States, most schools require that entering students have taken and successfully passed certain science courses. The faculty at each medical school use broad-based selection criteria in their admission process, including prior academic achievement, assessments of the candidate's academic abilities and personal qualities by college faculty and advisors, and evidence of values and attitudes commensurate with a career of service. All medical schools conduct interviews to assess the personal qualities, values, and attitudes of applicants, a practice less common in business, law, and other professional schools. The final responsibility for selecting students to an MD-granting educational program rests with the faculty admissions committee at each medical school. Most US medical schools use the Association of American Medical Colleges-sponsored Medical College Admission Test (MCAT) to assist admission committees evaluate students' academic abilities. This is a standardized examination that includes questions in the biological sciences, physical sciences, and verbal reasoning, and composition of a writing sample. A few countries (e.g., Japan and Korea) have begun to experiment with the application of the MCAT to their admissions process.

The Medical School Curriculum

Medical school curricula, although quite varied, have as their goal the preparation of students to enter a period of graduate medical education. The 4 years of medical school in the United States and Canada include preclinical (basic science) education and clinical education. Medical schools outside of the US and Canada incorporate basic science and clinical education as well though it is structured differently due to the longer medical-school experience.

Each medical school's faculty is responsible for determining the curriculum of the medical school and for establishing the criteria for assessing student performance, promotion, and graduation from that medical school.

The US medical-education system, while undergoing significant change currently, was based on the German medical-education system, promoted by Abraham Flexner in his report to the Carnegie Foundation in 1910. In this model, the medical school was based in the university with its attendant access to research and a solid grounding in science. The traditional curriculum in US medical schools has been a highly compressed, 2-year block of didactic, discipline-oriented basic sciences followed by 2 years of largely hospital-based clinical activity. Most US and Canadian medical students spend their first 2 years in the study of anatomy and cell biology, biochemistry and molecular biology, genetics, physiology, pharmacology, microbiology, neuroscience, and pathology, with a transitional course on physical diagnosis, medical interviewing, and an introduction to the practice of clinical medicine. These core courses are often enriched by other specific medical-science courses or topics such as medical ethics, biostatistics, humanities in medicine, and medical sociology.

These first 2 years of basic-science preparation are generally followed by a year of required clerkships in clinical disciplines that include internal medicine, pediatrics, psychiatry, obstetrics and gynecology, surgery, and family medicine. Increasingly, clinical clerkships in ambulatory-based medicine are part of the required experience. The length of these clerkships may vary from 5–12 or even 16 weeks. Individual schools may also have other required clinical experiences of a shorter duration than major clerkships. The clerkship is a period where the student works either on the hospital wards or in a physician's clinic, in supervised direct contact with patients.

The fourth year of medical school provides additional clinical experiences, generally by means of a mixture of elective and required course work, in disciplines such as radiology, anesthesiology, and the medical and surgical subspecialties.

Change

While this has been the dominant method of instruction and structure of the educational program, there are

significant changes underway in medical schools in the United States and Canada, as well as globally. The changes being implemented stem from medical educators' consensus that education and training must be closely aligned with society's needs and expectations; parallel developments in a healthcare system that is growing increasingly more complex; prepare students for a lifetime of continuous learning and self-improvement; and provide a training ground for future leaders.

The objective of most of the change is to develop the medical student as an adult learner, one who is a self-starter, a problem solver, and a critical thinker, and to link more closely the content of the coursework with the treatment of patients. There are numerous and complex reasons for the changes, but among the key issues are the following:

- There has been an exponential growth in biomedical knowledge. In particular, advances in molecular and cell biology have refined the understanding of mechanisms governing health and disease.
- Recent discoveries in biochemistry, immunology, and genetics that explain fundamental biological processes mandate a closer integration of the study of the basic sciences with clinical medicine. Such integration can only be achieved through a more interdisciplinary approach to medical education.
- A number of recent studies of medical education emphasize the desirability of making education more student and learning oriented, rather than faculty and teaching oriented. This requires that medical faculties offer educational experiences in which students are active, independent learners and problem solvers, rather than passive recipients of information.
- Greater attention is being given to the patient–physician relationship and on providing students with the skills and attitudes to make this relationship more meaningful. There is also increased emphasis on a physician's responsibility to work with individual patients and their communities to promote health and prevent disease.
- Heretofore, most clinical education has taken place in the teaching hospital. However, in recent years, medical educators have come to recognize that the teaching hospital has significant deficiencies as an educational setting.

Clinical Education

Perhaps the most significant changes have occurred in the ways in which medical students learn to interact with, diagnose, and treat patients. Experience in clinical settings occurs early in students' medical education. Some first-year medical students begin to see patients immediately; others are assigned a patient or family in their first year and follow those patients throughout their 4 years.

Traditionally, the inpatient services of the teaching hospitals provided an ideal educational setting for the clinical education of medical students. Medical students were able to observe, discuss, and participate in diagnostic and therapeutic activities in the company of and supervised by residents and faculty physicians. As patients were hospitalized for many common ailments for sufficient lengths of stay, students witnessed the whole course of admission, workup, diagnosis, treatment, and follow-up care.

While medical schools continue to use hospital inpatient services, these sites have developed some limitations as clinical-training settings. Technological advances and the financial incentives inherent in modern healthcare systems in the United States mean patients are hospitalized less frequently. Patients admitted to teaching hospitals today have very complex, highly specialized, acute illnesses that represent a skewed distribution of the medical conditions students will confront later in the office or clinic setting. Those who do receive care in a teaching hospital tend to stay for a shorter period, with much of the initial diagnostic workup and posttreatment care occurring in the ambulatory setting. As a result, medical students have little time to get to know hospitalized patients, to study their medical conditions, and to follow the course of treatment. They also see little of the early stages of illness and disease.

Although major teaching hospitals remain an important component of clinical education, other sites have gained prominence in medical student training. US medical students interact with patients in ambulatory clinics, physician offices, nursing homes, hospices, community clinics, and even prisons.

US medical schools rely on more than 1000 community hospitals as inpatient sites for one or more required clerkships. About 30 medical schools operate clinical campuses at great distances from the main medical school campus where a portion of the third- and fourth-year class receives clinical training.

The focus on communities is new for medical schools in the United States, though more a part of the curriculum in other countries. In the United States, public health and medicine have been taught separately, in different disciplines and different schools since the beginning of the twentieth century. There is now far more attention to the health of populations. The outpatient and clinic settings provide more representative patient populations and enable a strong educational focus on patient- and family-centered care; communication skills; health literacy; understanding the social, psychological, and cultural aspects of disease and disability and their implications for care; and public health and population-based medicine.

Responding to demographic, cultural and social changes, and the explosion of new medical knowledge, medical schools have dramatically expanded their curricula to include and integrate new topics such as: pain management,

and palliative care; emerging public health threats such as biological and chemical terrorism; epidemiology, preventive health, personalized health and population-based care; domestic violence and other sociobehavioral issues; and chemo-preventive methods, and medical genetics.

How Medical Students Learn

Simulations

One outcome of the changing demographics and location of the patient population has been to include the use of simulations, both human and mechanical (including computers, mannequins, and virtual reality) to introduce students to patient care. In addition to the changes to hospital-based care, the technology for simulations has improved to the extent that a good simulation can be created that is virtually indistinguishable from a real patient encounter. Second, patient care has moved more to the outpatient, ambulatory setting in the United States due to payment regulations, and innovations such as laparoscopic surgeries that can be done without a hospital admission. The typical hospital-ward-teaching environment has become more challenging for medical-student education because, in the United States, patients who are in the hospital are there because they are too ill to be cared for in an outpatient setting and therefore it is more difficult for the student to conduct routine histories and physicals on the patient; the patients who are in the hospital are there for a shorter duration, which means that it is more difficult for the student to participate in a longitudinal experience from admission, through work up, through care and treatment plan, to follow-up care. Medical students often are assigned a family or a group of patients who they follow through the course of their medical-school experience, in lieu of the hospital-based patient. Medical educators have expressed long-standing concern that the medical school experience does not adequately prepare physicians to attend to the personal/psychological needs of patients or to the healthcare needs of communities. The longitudinal experience allows the student to participate in the healthcare of an individual over an extended period of time.

The changes in the demographics of the patient population and the need to increase clinical encounters has led to the use of human simulations, called standardized patients, a pedagogy in use throughout the world. The idea of the standardized patient in Canada and the United States is attributed to Drs. Howard Barrows and Paula Stillman. The standardized patient is someone who presents with either a chronic condition or is trained to simulate a medical condition (e.g., one can be trained to simulate a collapsed lung, or the symptoms of multiple sclerosis) and the person presents with the same medical history and medical situation each time for each student, hence the term standardized patients. Standardized

patients are used for both teaching and assessment and are a component of the physician licensure examination in the United States.

Problem-based learning

Another prevalent pedagogy in medical schools throughout the world is a concept adapted from the case-based learning originated in law schools that is usually called problem-based learning (PBL). This pedagogy was created by Dr. Howard Barrows when he was at McMaster University in Canada. PBL involves a small group of students engaged in self-directed learning aimed at gaining an understanding of the ways that biological science and clinical disciplines can elucidate a specific scientific or clinical problem. Research examining the effectiveness of PBL as a pedagogy has shown that students possess increased lifelong learning skills, work better in teams, and enjoy their education more than learning from a strictly lecture-based curricula.

Competency-based education

Competency-based education and learning is increasing in United States medical schools, stimulated by reports such as, Objectives for Medical Student Education and the Accreditation Council for Graduate Medical Education (ACGME) Core Competencies. Competency-based education allows students to learn at their own pace, to focus on mastering a competency rather than focusing on passing a course, and promotes lifelong learning. A competency is a complex set of behaviors built on the components of knowledge, attitudes, skills, and competence as personal ability and it must be demonstrated. The ACGME competencies are domains in which a physician must ultimately demonstrate competence.

The principles of clinical research are another important component in modern medical education. In 1910, Abraham Flexner stated that “research can no more be divorced from medical education than can medical education be divorced from research.” That sentiment is vital in today’s world of tremendous breakthroughs in the knowledge and understanding of the biologic basis of human health. Therefore, medical schools introduce students to the basic principles of clinical and translational research, including how such research is conducted, evaluated, explained to patients, and applied to patient care.

These changes in educational venue and objectives have been essential for medical students to prepare properly for practice in the twenty-first century, but have not been without challenges. Modern-day health care, for example, places a premium on doctors’ efficiency and productivity in patient care. Teaching in ambulatory settings may incur other costs, such as capital and infrastructure required to accommodate students in clinics, transportation, and living expenses when the site is distant from the medical school, and the effort and

expense of recruiting and training community physicians who are not full-time faculty. Geographically dispersed sites that use large cadres of community physicians also challenge medical schools to ensure that students have a comparable educational experience and are assessed with similar standards.

Residency Education

Following the award of the MD degree, in the United States and Canada, young physicians begin an intensive period of hospital-based clinical training, the residency, which is designed to prepare them for the practice of a particular medical specialty. This period is called graduate or postgraduate medical education and is characterized in the United States by the internship and residency. Residency as a component of medical education began at Johns Hopkins Hospital in the 1890s, modeled on the assistantships prevalent at the time in Germany. Residents were called house officers because they literally lived in the hospital. Even today, residents are sometimes referred to as house officers or house staff.

All medical-school graduates who seek full medical licensure and board certification in a medical specialty or subspecialty in the United States must complete a period of residency training. Residency programs vary in length depending upon specialty but generally last 3–5 years for initial board certification and subspecialty training may extend the period of education to as long as 11 years following the award of the MD degree.

Graduate medical education is a period of education in which the physician is still a student in training even though they perform procedures, make diagnoses, and are involved in all aspects of patient care. The difference between the attending physician and the physician in graduate training is that the resident still lacks the experience to practice independently in their specialty. In fact, in the US system, residents are both graduates in training and often the primary teachers of students on hospital wards. Residents assume responsibility for patients under supervision of physician faculty. Their clinical experiences are organized as a series of rotations that may include assignments to inpatient services, hospital-outpatient clinics, and sites in the community such as community health centers and physician offices. Each rotation has a set of defined educational objectives for which the resident is evaluated. Residents also participate in conferences, seminars, and other nonpatient learning experiences as well as independent, self-directed learning, and scholarly activities.

The accreditation process for residency programs emphasizes the actual outcomes of what residents learn rather than just the program’s potential to educate. Residents are expected to demonstrate competence in

six domains of knowledge, skills, and attitudes deemed necessary to be effective and compassionate physicians (the ACGME competencies).

Licensure and Certification

Licensure and certification of physicians is approached individually by each country. The United States and Canada have different licensing bodies; in the United Kingdom, physicians are certified by the General Medical Council and in some countries there is no national licensing and certifying process for physicians in practice.

The external examination required of all candidates for licensure that will permit them to practice medicine in the United States, whether they are graduates of accredited US schools or of foreign medical schools, is the three-part United States Medical Licensing Examination (USMLE), co-sponsored by the National Board of Medical Examiners (NBME) and the Federation of State Medical Boards (FSMB), which represents the various state and jurisdictional licensing authorities. The three steps of the examination include:

- step 1, which focuses on knowledge of the basic sciences and is often taken at the conclusion of the second year of medical school;
- step 2 focuses on the knowledge and understanding necessary to give clinical care under supervision and includes both a written and a clinical-skills examination and is usually taken before graduation from medical school; and
- step 3, which focuses on the knowledge and understanding necessary for the unsupervised practice of medicine and is usually taken during or after the first year of residency training.

International medical graduates seeking entry to accredited residency programs must first obtain a certificate awarded by the Educational Commission for Foreign Medical Graduates (ECFMG). This certificate is based upon satisfactory passing scores of the USMLE basic science and clinical science components (including a clinical-skills exam that requires spoken-English-language proficiency) and on documentation of a completed education program for, and receipt of, a medical diploma.

The USMLE examination is constructed by committees comprised of faculty of US medical schools, expert in the subjects being examined. The USMLE is a high-stakes examination because the majority of US medical schools uses passage of the USMLE step 1 and step 2 examinations as one criterion for graduation from medical school. As such, the examination has a significant impact on medical-school curricula.

There are efforts underway in several countries to develop national certifying examinations. The United

Kingdom has created the Foundation Programme, a 2-year training program required of all graduates in order to practice medicine in the UK. The program is conducted by 14 medical schools and deaneries in the UK. Competencies have been identified and foundation doctors must complete all competencies to be certified.

Accreditation

Medical schools in the United States and Canada have a unified accreditation process, conducted by the Liaison Committee on Medical Education (LCME) for US schools and the LCME and the Association of Faculties of Medicine of Canada (AFMC) for Canadian schools. Many countries do not have any sort of nationally sanctioned accreditation process though the standards set forth by the World Federation for Medical Education (WFME) and the Bologna Declaration are being instituted throughout the world. In 2005, the Association for Medical Education in Europe (AMEE) and the WFME issued a joint statement on the Bologna Process and medical education that endorsed the Bologna Process and urged governmental bodies to implement the elements outlined in the Bologna document. The WFME has established standards for medical-student education that are being implemented in many countries and they are promoting accreditation standards.

Summary

The goals of medical education are universal – to educate physicians who are caring professionals, who will bring a humanistic, patient-centered approach to care, who are lifelong learners, and who recognize the importance of research to the advancement of medicine as well as the fundamental importance of the basic sciences to medicine. These goals are met using different pedagogical approaches and different curricula, but the outcome is the same.

There are many questions facing medical education as it advances in the twenty-first century. These include how the educational program can keep up with the rapid increase in scientific knowledge; what impact the new technologies will have on patient care; what the understanding of the human genome will mean for patient-centered care; and how to balance the increases in technology with the importance of the patient's story.

Further Reading

Anderson, M. B. (ed.) (2000). A snapshot of medical students' education at the beginning of the 21st century: Reports from 130 Schools. *Academic Medicine* 75(supplement 9), 475.

- Blumenthal, D. (2002). *Training Tomorrow's Doctors. The Medical Education Mission of Academic Health Centers. A Report of the Commonwealth Fund Task Force on Academic Health Centers*. New York: The Commonwealth Fund.
- Blumenthal, D., Gokhale, M., Campbell, E. G., and Weissman, J. S. (2001). Preparedness for clinical practice. Reports of graduating residents at academic health centers. *Journal of the American Medical Association* **286**, 1027–1034.
- Candler, C. (2007). *Effective Use of Educational Technology in Medical Education*. Washington, DC: Association of American Medical Colleges.
- Flexner, A. (1910). Medical education in the United States and Canada. *A Report to the Carnegie Foundation for the Advancement of Teaching*. New York: Carnegie Foundation for the Advancement of Teaching.
- General Medical Council (1993). *Tomorrow's Doctors. Recommendations on Undergraduate Medical Education*. London: General Medical Council.
- Institute of Medicine (2001). *Crossing the Quality Chasm A New Health System for the 21st Century*. Washington, DC: National Academy Press.
- Institute of Medicine (2003). *Health Professions Education. A Bridge to Quality*. Washington, DC: National Academy Press.
- Institute of Medicine (2003). *Academic Health Centers: Leading Change in the 21st Century*. Washington, DC: National Academy Press.
- Inui, T. S. (2003). *A Flag in the Wind: Educating for Professionalism in Medicine*. Association of American Medical Colleges. <https://services.aamc.org/publications> (accessed August 2009).
- Kassebaum, D. G. and Anderson, M. B. (1993). Proceedings of the consensus conference on the use of standardized patients in the teaching and evaluation of clinical skills. *Academic Medicine* **68**(6), 437–483.
- Ludmerer, K. M. (1999). *Time to Heal*. New York: Oxford University Press.
- McGlynn, E. A., Asch, S. M., Adams, J., et al. (2003). The quality of health care delivered to adults in the United States. *New England Journal of Medicine* **348**, 2635–2645.
- Medical School Objectives Writing Group (1999). Learning objectives for medical student education – guidelines for medical schools. *Academic Medicine* **74**(1), 13–18.
- Schoen, D. A. (1987). *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*. San Francisco, CA: Jossey-Bass.
- Shugars, D. A., O'Neil, E. H., and Bader, J. D. (eds.) (1991). *Healthy America: Practitioners for 2005: An agenda for action for U.S. health professional schools. Report of the Pew Health Professions Commission*. Durham, NC: The Pew Health Professions Commission.

Relevant Websites

- <http://www.acgme.org> – Accreditation Council for Graduate Medical Education.
- www.aamc.org – Association of American Medical Colleges.
- www.amee.org – Association for Medical Education in Europe.
- www.ecfmg.org – Educational Commission on Foreign Medical Graduates.
- www.lcme.org – Liaison Committee on Medical Education.
- www.nbme.org – National Board of Medical Examiners.
- www.foundationprogramme.nhs.uk – The Foundation Programme.
- www.wfme.org – World Federation for Medical Education.

Education for the Legal Profession

J W Wegner, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction and Comparative Context

Lawyers and Their Roles

Legal education occurs within a professional, educational, and regulatory context that is best considered at the outset. Different societies cast lawyers in a variety of roles, and professional employment patterns differ. For example, many common law countries, including the United States, have a unitary bar (meaning that there is not a distinction in licensure for lawyers who advise clients in office settings and those who appear as advocates in the courts. England and Canada, among others, are common law countries but distinguish in licensure between solicitors (who advise clients outside the courtroom and draft documents) and barristers (who appear before the courts). Civil law jurisdictions (drawing upon Roman traditions and emphasizing legal principles embodied in codes) typically employ a preponderance of lawyers in government roles including those of notary (responsible for verifying facts, drafting documents, and managing land registries) and judge (who are involved in the inquisitorial system for investigating, finding facts, charging, and adjudicating criminal conduct). Lawyers in such systems often specialize in one or another role rather than being free to shift from position to position and role to role as in many common law jurisdictions. In many common law jurisdictions, lawyers are more significantly employed in the private sector, increasingly in larger firms. Although many American jurisdictions provide opportunities for certification of lawyer specialists in substantive areas, most rely on requirements for experience and special licensure examinations to determine eligibility for such recognition, rather than concentrations of academic courses during law school.

Student Preparation

Significant differences among legal-education systems also result from the prior educational experiences of students, and differences in the background and maturity of law students play an important role in the goals and design of legal education around the world. For example, many countries allow students to study law as part of their undergraduate educations. Within the European Union (and other interested countries), the Bologna process and related initiatives have driven important changes in higher-educational processes. These initiatives are designed to distinguish between bachelors', masters', and doctoral levels

of education; recognize the work of students who have studied in other countries; promote mobility; develop quality-assurance standards; and foster common curricular standards. It is expected that 3-year undergraduate degrees will become common in Europe, and curricula will be modified to reflect a European (rather than only national) focus. Nonetheless, students may be required to complete an additional year of articling or intensive practical training before licensure, offered either by the established bar or by intensive programs affiliated with certain law schools.

In contrast, American legal education requires postgraduate preparation rather than concentration during undergraduate programs. Such education is open to those with prior experience in any number of undergraduate fields. No articling is required prior to licensure, and graduates need only pass a bar examination in specific jurisdictions to receive a general license to practice law. Japan, among others, has moved toward a similar system of postgraduate education for lawyers in recent years.

Accreditation and Licensure

Countries differ in their approaches to accreditation and licensure. Some, such as Japan and Korea, have historically used very stringent licensing examinations to limit significantly the number of lawyers and judges compared to those who are educated in the law in the nation as a whole.

American legal education, on the other hand, operates within a more balkanized regulatory system through which each state regulates candidates seeking licensure to practice law. State supreme courts derive power from state constitutions that afford them ultimate authority to determine who is to be admitted to practice and how a license to practice is to be maintained, while state legislation generally establishes a state-specific state bar entity charged with disciplinary oversight of licensed lawyers. Most states have a board of bar examiners charged with reviewing candidates for admission. Candidates must generally attend an accredited law school, take requisite courses, satisfy requirements for character and fitness (such as those relating to honesty, criminal conduct, and financial propriety), and perform satisfactorily on mandatory licensing examinations. The American Bar Association (ABA), through its section on legal education and admission to the Bar (ABA), serves as the principal accrediting body for American legal education. State high courts regard the ABA as an appropriate entity to

develop minimum standards for assessing law-school programs. Once a school has received ABA accreditation, the schools' graduates are allowed to sit for the bar in a state other than the one in which they attended law school. The ABA's accreditation power also derives from the United States Department of Education, which certifies accrediting agencies before students enrolled in those programs are eligible for federal financial aid.

Educational Institutions and Their Communities

The number of educational institutions offering law programs varies from country to country. The International Association of Law Schools publishes information on the number of law programs in many countries and the differences in legal systems that shape educational programs. In the United States, approximately 200 law schools offer programs approved by the ABA and designed to permit graduates to sit for bar examinations within the United States, generally after receiving a Juris Doctor (JD) degree. The number of law students receiving JD degrees has increased from 35 604 in 1980–81 to 42 518 in 2006–07, and now face significantly increased tuition (in 2007 private schools charged US\$32 367 on average, while public schools charged US\$26 691 to nonresidents and US\$15 455 to state residents). American law students typically fund their education through substantial educational loans, resulting in significant debt upon graduation. Admission to American law schools depends on students' performance on the law school admissions test (LSAT), undergraduate performance, and other indicators. The Law School Admissions Council (LSAC) and the ABA have cautioned against excessive reliance on the LSAT in admission decisions, but law schools' increasing sensitivity to rankings by US News and World Report have led many to increase reliance on this measure reaching admissions decisions. The United States Supreme Court's decision in *Grutter v. Bollinger* has permitted law schools to continue efforts to diversify the legal profession by weighing students' diverse backgrounds and experiences, at least for now.

Pedagogy, Curriculum, and Delivery of Instruction

Signature Pedagogy: The Case-Dialog Method and the Cognitive Apprenticeship

In many countries where students study law as undergraduates, traditional strategies such as lectures are often employed. American legal education on the other hand has adopted the case-dialog method as its signature

pedagogy since that approach pioneered at Harvard Law School beginning in the 1870s. This approach focuses on cases drawn from appellate court decisions, and collected in case book anthologies. Sometimes referred to as the Socratic method, this approach does involve dialog, but does not reflect the disinterested inquiry after the good employed by the Greek sage. Instead, the method involves faculty–student interchanges focused on questions implicit in the subject matter (the dialectical quality of legal disputes and the dynamic character of analogical reasoning across jurisdictions and time) and is driven by the instructional goal of developing students' capacity for independent professional thought. The case-dialog method provides a platform for instruction in legal reasoning (developing higher-order thinking skills and an epistemology suitable to the legal context). It helps beginning students understand the law by developing legal literacy, depicting the legal landscape, and introduces Anglo-American common law jurisprudence. It is especially useful in teaching students to navigate the zone of uncertainty that is central to the exercise of professional judgment in a rapidly changing world. Students are called upon to recite the facts, holdings, and reasoning of cases, requiring them to develop content knowledge, comprehend the sometimes arcane cases they must read, tease apart key analytical elements, synthesize cases, apply key principles to novel fact patterns, evaluate judicial reasoning, and compare judicial and legislative approaches to significant problems. This work is not undertaken in the abstract, since students must simultaneously master a new vocabulary of legal language, become more adept critical readers, and imagine themselves in the roles of lawyers serving as attorneys for varying parties, counselors, judges, or legislators.

The case-dialog method is particularly powerful in helping students face the necessity of wading into unknown waters. It forces them to grapple with carefully posed questions, treat questions as a routine part of professional inquiry, reconstruct knowledge in working with cases, and tame uncertainty by asking and answering questions. The case-dialog method fits well with principles of situated learning and cognitive apprenticeships, and provides instructors with a ready means to draw out students by modeling, coaching, scaffolding, and fading while insisting that novices articulate, reflect, and explore their insights through dialog.

Despite its power, the case-dialog method has important drawbacks. It is especially effective in dealing with the cognitive apprenticeship, but its shadow side tends to limit attention to the apprenticeship of skills and practice and the apprenticeship of identity and values unless instructors deliberately compensate for its narrow focus on appellate cases and emphasis on lawyers in litigation roles. The case-dialog method emphasizes only particular kinds of doctrinally oriented principles that

function as doctrinal baselines, ingrains caution about overgeneralization, and introduces legislative, transactional, and cross-cultural perspectives only if instructors go out of their way to do so. The method tends to discourage critiques of legal principles except as to procedural justice, clarity, and efficiency, and provides no ready platform for conversations about substantive justice in the classroom.

In contrast to some other fields, American legal education's signature pedagogy is employed full force during the first year of a 3-year program of professional education, rather than as a capstone that positions students to move into the profession as is true in fields such as medicine and nursing. It tends to limit the extent to which students can be introduced to the broader range of skills and characteristics of professional expertise, particularly if their instructors have had little actual professional experience.

Skills and Practice

Evolving Expectations

The development of clinical educational initiatives has drawn considerable attention from around the world. Even as the case-dialog method and attention to the cognitive apprenticeship dominated American legal education, some American faculty members and students began efforts to deliver needed legal services to the poor through programs associated with scattered law schools. The Ford Foundation fostered these developments through the Council on Legal Education for Professional Responsibility (CLEPR) and clinical offerings in American law schools expanded substantially as a result. The ABA subsequently undertook a number of initiatives designed to encourage American law schools to integrate professional skills programs in their educational programs. ABA president Robert MacCrate chaired the Task Force on Law Schools and the Profession – Narrowing the Gap, working with practitioners, judges, and legal educators to map professional skills and values deserving of considered attention during law school and beyond.

In the years following the MacCrate report, the ABA adopted more rigorous accreditation standards mandating that schools provide substantial instruction designed to provide graduates with basic competence in legal analysis and reasoning, legal research, problem solving, and oral and written communication, at least one rigorous writing experience, adequate instruction in professional skills, and live-client or other real-life practice experiences for credit . . . through clinics or externships. Law schools were required to offer substantial opportunities for live-client or other real-life practice experiences, appropriately supervised and designed to encourage reflection by students on their experiences and on the values

and responsibilities of the legal profession, and the development of one's ability to assess his or her performance and level of competence.

Common Models

Four principal models for providing instruction designed to develop practice-oriented professional skills have emerged in American law schools and have served as models elsewhere.

Clinical programs generally involve live clients, that is, those who seek legal assistance in connection with matters that licensed lawyers would otherwise handle. In many states, third-year practice rules limit clients who may be assisted to those of limited means, and may require that only carefully supervised students who have completed two full years of law school can provide representation. Most clinical programs have a specific focus, for example, representation of indigent clients in civil matters, juvenile or appellate defense, assistance economic development or small businesses, domestic violence, representation of immigrants, or policy development in specified areas. While individual programs differ, most involve students in specific matters that allow them to learn how to interview and counsel clients, comply with ethical obligations, gather facts, develop case theories, prepare documents, make oral presentations, and collaborate with others.

Several pedagogical strategies quite unlike the case-dialog method are commonly employed in clinical settings. Faculty supervisors engage in one-on-one coaching to assist students in developing and demonstrating pertinent professional skills. Students are also encouraged to reflect on their work in order to gain deeper insights. Students in clinical programs are often exposed to distinctive pedagogical strategies involving case rounds. Many clinical programs also require classroom components that introduce theory or basic skills instruction either concurrently or as a prerequisite of clinical enrolment.

Externships (postings in which law students are assigned to work under practitioners or judges in field settings) are also common in American law schools. The types of placements vary by location, but often encompass an array of settings including the chambers of trial and appellate judges, public defenders' or prosecutors' offices, government agencies, legal services, nonprofit organizations, and corporate counsel offices. Under ABA accreditation requirements, externships are required to have clearly stated objectives and evaluation strategies, well-trained field supervisors, close oversight and periodic site visits by law school faculty supervisors, and integrated opportunities for students to reflect on the work undertaken in the field. Students receiving academic credit for externships may not receive compensation.

Simulation courses have also become quite popular in law schools. Simulation courses often involve extensive

case files or similar materials designed to present specific lawyering tasks (such as cross-examination or interviewing). Students are provided with overviews of key content in lecture settings, and then asked to engage in role plays in which they have the opportunity to practice key professional skills. Practitioners are often involved in critiques of student performance. While many courses in specific areas (such as trial advocacy, pretrial practice, interviewing, counseling, and negotiation) have been established to develop specific types of practice skills, there has been a growing interest by some faculty members in creating more complex problems and instructional modules. Few of these courses approach the sophistication of the Simulated Professional Learning Environment (SIMPLE) developed by Prof. Paul Mahrag of Glasgow Graduate School of Law, University of Strathclyde, Scotland, and his colleagues as a means of preparing students with advanced professional skills for entry into the practicing legal profession in the United Kingdom.

Practicum courses are gradually emerging as a distinctive approach to providing education in professional skills. Such offerings often involve team teaching by law faculty members with colleagues from the practicing bar, or efforts by nonclinical faculty members to develop courses focused on engaging law students in projects with real-world application.

Legal writing offerings in the first year of law school also incorporate distinctive pedagogical strategies. Increasingly, such offerings situate beginning law students in professional contexts and call upon them to prepare legal memoranda, opinion letters, briefs, and oral arguments focused on legal problems that might arise in practice. Many schools employ practicing lawyers and judges as adjunct faculty members responsible for providing such instruction. A number of law schools have also begun to expand their legal writing courses to address a broader array of professional skills, including counseling, interviewing, fact-finding, problem solving and negotiation, and have stretched such courses into the second year of law school.

The Clinical Legal Education Association (CLEA), under the leadership of Prof. Roy Stuckey, has initiated a best-practices project designed to emphasize the importance of setting goals, organizing and delivering varied types of instruction, assessing students learning, and evaluating the success of the program of instruction. The best-practices project seeks to improve all facets of legal education by bringing to bear important insights about pedagogy and instructional priorities that are relevant not only to clinical legal education but also to legal education more generally.

Professional Identity and Values

Intensified attention was devoted to the role of American law schools in developing students' sense of professional

identity and values in the aftermath of the Watergate scandal in the early 1970s when lawyers serving as presidential advisers were widely viewed as having failed to abide by professional obligations. The ABA subsequently incorporated an accreditation requirement requiring students to receive instruction in the legal profession, including the history, goals, structure, and responsibilities of the legal profession and its members. This requirement was subsequently revised to refer explicitly to the values and rules of the legal profession. States have reinforced this requirement by incorporating requirements that law graduates take a special licensing examination on professional responsibility before they may be admitted to the bar.

Two major approaches to instruction in professional responsibility and ethics have emerged to address this requirement. American law schools typically offer free-standing courses on professional responsibility using instructional strategies focused on cases and rules in order to meet this requirement. More recently, instructional strategies have put greater emphasis on contextualized problems and case studies. Typically, free-standing courses cover the role of lawyers as advocates and counselors, regulatory powers relating to discipline and bar admission, the attorney–client relationship, confidentiality, conflicts of interest, and special requirements applicable in criminal and civil settings.

Leading scholars in the field have urged that law schools incorporate a pervasive approach, so that students receive instruction not only in a stand-alone course, but also in contexts across the full range of substantive fields. Clinical legal educators generally give special consideration to issues of professional responsibility in connection with students' work with clients. Efforts by other faculty members to incorporate ethics and professional responsibility into substantive courses have tended to be sporadic at best.

Increasingly, American law schools have used the informal curriculum to address concerns about a decline in professionalism (standards of civility, mutual respect, service, and integrity that exceed the minimums set by disciplinary standards). Professionalism initiatives often focus on orientation programs, acceptance of professionalism codes, speakers, mentoring programs, and specialized courses (e.g., in law office management). Some schools have also encouraged students develop work/life balance in keeping with growing interest in humanizing legal education. Many law schools have developed affiliations with local Inns of Court (local groups including law students, young and more senior lawyers, judges and law professors) who meet on a recurring basis for educational programs and informal networking.

American law schools have also worked closely with the bar in developing robust *pro bono* programs that invite or require students to offer legal services at no cost to those of individuals and organizations serving those of limited means. The ABA's accreditation standards state

that law schools “should encourage (their) students to participate in *pro bono* activities and provide opportunities for them to do so.”

Curriculum

Legal education in the United States is a 3-year postgraduate program open to students who have already completed undergraduate programs and have majored in a variety of fields. Apart from the accreditation requirement that instruction in professional responsibility be provided, American law schools are generally not required to provide instruction in particular substantive areas. Nonetheless, clear patterns are widely evident across most American law schools. Curricular offerings in other countries vary depending on undergraduate or graduate emphasis.

American first-year curricula typically include common law subjects that provide substantive grounding, facilitate development of analytical abilities through use of the case-dialog method, and cover subjects generally tested by bar examinations. Most law schools include criminal law, torts, contracts, property, civil procedure, and constitutional law as part of the first-year curriculum. Legal research and writing is also a staple of the first year. Some schools incorporate a broader range of professional skills (e.g., negotiation and counseling) within the scope of core offerings, but many extend such instruction into the second year. In recent years, some elite law schools have moved to include courses in legislation or the regulatory state as part of the first-year curriculum in order to provide more balanced coverage. Others have included coverage of international or comparative law, at times through intensive intersession programs.

American law schools typically offer a cafeteria array of options for students in the second and third year of their programs. Generally, students are advised about the coverage of bar examinations in jurisdictions in which they may seek licensure. The upper division curriculum typically includes coverage in the areas of litigation, business and tax law; employment law; family law; constitutional and civil rights law; international and comparative law; environmental law; property and intellectual property; litigation; and perspectives on the law. Some law schools have developed concentrations of courses in substantive areas as an option for students who wish to devote focused attention to particular fields, while others offer joint degrees in fields such as business, planning, public administration, health, and social work. A substantial number offer LLM degrees targeted to international students or those seeking advanced preparation in specialized fields such as tax or international law.

Delivery Strategies

American law schools have been relatively slow in exploring educational strategies using advanced technology.

The computer-assisted legal instruction (CALI) initiative provides several hundred web-based tutorials authored by law faculty members, sponsors conferences, facilitates use of pod-casting, explores other means of developing electronic instructional materials, and hosts a consortium fostering development of online courses. Pioneering professors have developed specialized courses and online LLM degrees, while Concord Law School has developed a comprehensive degree program that offers all courses online.

The ABA regulates distance education and programs reliant upon technological transmission through its accreditation standards. Offerings must be carefully planned and separately authorized as a part of the curriculum. Students must be supervised, involved in ample interaction with instructors, and carefully monitored and assessed. Distance-education courses may not be offered to students until they have completed their first year, and are limited to four units per semester and 12 units toward a standard law degree.

Assessment

American law schools are quite weak when it comes to assessment, according to a recent Carnegie Foundation study. ABA accreditation requirements require schools to adhere to sound academic standards and monitor students' academic achievement. Schools must also periodically review their mission and curricula, but most lack the will or expertise to link educational programs to outcome measures or specific goals. As large student/faculty ratios characterize many core courses, formative assessments are used infrequently. Traditional examinations typically include problems based on complex hypothetical fact patterns requiring students to analyze and apply legal principles. Many instructors also use multiple-choice questions to assess the breadth of student knowledge and provide students with practice in responding to the types of questions included in multi-state bar examinations. Most courses in American law schools are graded on a curve so that grades in larger courses reflect a standard distribution.

Recent developments reflect growing attention to assessment as a formative tool. Most American law schools have developed academic-support programs designed to assist students who may be at risk of performing below their potential (e.g., those who are first-generation college graduates), participants in admission-by-performance programs, and students who fare poorly at the outset of their law school careers. The LSAC has provided important support for academic support professionals. Those serving in these capacities have enriched educational programs by virtue of their backgrounds in educational theory as well as law. Academic-support programs have extended

their reach beyond the first year, particularly since the ABA has adopted a requirement that law schools demonstrate that the bulk of their students pass state bar examinations (or that they have programs in place to assist them to do so). Bar preparation courses designed to strengthen the ability of low-performing law students to succeed on bar examinations may now be offered for credit, and many schools have begun to make them available on an optional or required basis.

Other countries have adopted much more sophisticated assessment practices, for example, as part of the Bologna process discussed above.

Further Reading

- ABA Task Force on Lawyer Competency: The Role of Law Schools (1979).
- American Bar Association, Section of Legal Education and Admissions to the Bar (1979). *Report on Recommendations of the Task Force on Lawyer Competency: The Role of Law Schools, Cramton Report*. Chicago: American Bar Association.
- American Bar Association, Section of Legal Education and Admissions to the Bar (1983). Task force on professional competence. *Final Report and Recommendations of the Task Force on Professional Competence*. Chicago: American Bar Association.
- American Bar Association Section on Legal Education and Admissions to the Bar (1992). Legal Education and Professional Development – An Educational Continuum, *Report of the Task Force on Law Schools and the Profession: Narrowing the Gap, MacCrate Report*.
- American Bar Association, Section of Legal Education and Admissions to the Bar (1996). Teaching and learning professionalism: *Report of the Professionalism Committee*. Chicago: American Bar Association.
- Amsterdam, A. J. (1984). Clinical legal education – A 21st-century perspective. *Journal of Legal Education* **34**, 612.
- Boyer, B. B. and Cramton, R. C. (1974). American legal education: An agenda for research and reform. *Cornell Law Review* **59**, 221.
- Frank, J. (1933). Why not a clinical law school? *University of Pennsylvania Law Review* **81**, 907.
- Glendon, M. A. (1994). *A Nation under Lawyers: How the Crisis in the Legal Profession is Transforming American Society*. Cambridge: Harvard University Press.
- Grutter v., Bollinger, 539 U.S. 306 (2003).
- Kissam, P. (2003). *The Discipline of American Law Schools*. Durham, NC: Carolina Academic Press.
- LaPiana, W. D. (1994). *Logic and Experience: The Origin of Modern American Legal Education*. New York: Oxford University Press.
- Llewellyn, K. N. (2008). *The Bramble Bush: The Classic Lectures on the Law and Law School*. New York: Oxford University Press.
- Luca, L. D. (2008). Enriching the law school curriculum in an increasingly interrelated world – learning from each other. *Pennsylvania State International Law Review* **26**, 831.
- MacCrate, R. (1997). Educating a changing profession: From clinic to continuum. *Tennessee Law Review* **64**, 1099.
- Mahrag, P. (2007). *Transforming Legal Education*. Aldershot, UK: Ashgate.
- Mertz, E. (2007). *The Language of Law School: Learning to “Think Like a Lawyer”*. Oxford, UK: Oxford University Press.
- Milstein, E. and Bryant, S. (2007). Rounds: A ‘signature pedagogy’ for clinical education? *Clinical Law Review* **14**, 195.
- Munro, G. (2001). *Outcomes Assessment for Law Schools*. Topeka, KS: Institute for Law School Teaching.
- Packer, H. L., Ehrlich, T., and Pepper, S. (eds.) (1972). New directions in legal education: *A Report Prepared for the Carnegie Commission on Higher Education*. New York: McGraw-Hill.
- Rhode, D. (2003). *In the Interests of Justice: Reforming the Legal Profession*. New York: Oxford University Press.
- Schrag, P. G. and Meltsner, M. (1998). *Reflections on Clinical Legal Education*. Boston, MA: Northeastern University Press.
- Shephard, S. (2007). *The History of Legal Education in the United States: Commentaries and Primary Sources*. Clark, NJ: Law Book Exchange.
- Shulman, L. (2005). Signature pedagogy in the professions. *Daedalus* **134**(3), 52–59.
- Sonsteng, J. O. (2008). *A Legal Education Renaissance: A Practical Approach to the Twenty-First Century*. Lake Mary, FL: Vandeplas.
- Stevens, R. (1987). *Law School: Legal Education from the 1850’s to the 1980’s*. Chapel Hill, NC: UNC Press.
- Stuckey, R., et al. (2007). *Best Practices for Legal Education: A Vision and a Road Map*. Columbia, SC: Clinical Legal Education Association.
- Sullivan, W. M., Colby, A., Wegner, J. W., Bond, L., and Shulman, L. S. (2007). *Educating Lawyers: Preparation for the Profession of Law*. San Francisco, CA: Jossey-Bass.
- Terry, L. S. (2007). The Bologna Process and Its Implications for U.S. Legal Education. *Journal of Legal Education* **57**, 237.
- Terry, L. S. (2008). The Bologna process and its impact in Europe: It’s so much more than degree changes. *Vanderbilt Journal of Transactional Law* **41**, 107.

Relevant Websites

- <http://www.americanbarfoundation.org> – American Bar Foundation.
- <http://www.abanet.org> – American Bar Association Pro Bono Programs.
- <http://www.abanet.org> – American Bar Association Section on Legal Education and Admission to the Bar.
- <http://www.aals.org> – Association of American Law Schools.
- <http://www.aallnet.org> – Association of American Law Librarians.
- <http://www.ialsnet.org> – Association of International Law Schools.
- <http://www.alwd.org> – Association of Legal Writing Directors.
- www.carnegiefoundation.org – Carnegie Foundation for the Advancement of Teaching.
- <http://www.cleaweb.org> – Clinical Legal Education Association.
- <http://www.lsac.org> – Law School Admissions Council.
- <http://www.nalp.org> – National Association for Legal Career Professionals.
- <http://www.ncbex.org> – National Conference of Bar Examiners.
- <http://www.saltlaw.org> – Society for Law School Teachers.
- <http://www.ukcle.ac.uk> – UK Centre for Legal Education.

Education of Clergy

D O Aleshire, The Association of Theological Schools in the United States and Canada, Pittsburgh, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

The education of clergy varies by religious tradition (i.e., Christianity or Judaism); by subgroups or denominations within the same tradition (i.e., Lutheran, Baptist, Presbyterian within Christianity or Orthodox, Conservative, and Reformed within Judaism); and across continents. Education varies by tradition and denomination because each authorizes clergy for ministry in different ways. Clerical education varies by continent because theological education typically follows the dominant educational pattern of the continent, and those patterns vary. In the context of these many variations, however, educational goals tend to be similar because clerical work has common aspects. Across traditions, denominations, and continents, clergy (1) preside at worship and liturgical celebrations, (2) teach, preach, and otherwise seek to communicate religious understanding; (3) provide guidance and leadership to congregations or parishes where the faithful gather for worship, instruction, mission, and fellowship; (4) provide spiritual care in the context of major life events such as birth, marriage, illness, and death; and (5) lead religious communities to address human need and moral vision. This article explores influences on the education of clergy and describes characteristics of the normative pattern of theological education in North America.

Influences on Clergy Education

The education of clergy is influenced by patterns of identifying qualifications for ministerial work and systems for providing authorization to do that work. Three patterns are particularly instructive because of their implications for contemporary aspects of the education of clergy.

Three Patterns of Clerical Qualification and Authorization

For ancient Israelites, the members of the tribe of Levi were responsible for overseeing religious practices. Priesthood was the tribal community's family business. Little is known about the education of these early priests, but they must have learned first from a parent, then from the structures that the tribe maintained to pass on its unique work. No contemporary religious community identifies clergy by family of origin; however, some contemporary patterns of clerical education may be similar to those ancient practices. They were evident in early America,

where a person reads with an experienced clergyman as preparation for ministry. While this model was not a unique form of professional education in the nineteenth century, it is unique in that it continues as a form of clergy education for some religious groups. In some religious communities, an experienced pastor or religious leader takes under his or her tutelage younger persons who learn from the senior clergy and serve as assistants in the senior minister's congregation until they have one of their own.

History is full of the emergence of new religious movements. Some became world religions, others lingered as sectarian groups on the margin, but all of them had clergy. In a new religious movement, how do the clergy attain clerical authority? It cannot be bestowed by an external source or be based on the leader's mastery of traditional religious teaching. It emerges from the leader's ability to articulate new religious teaching or convey a unique relationship with the divine. While this kind of religious authority is historically evident in new religious movements, it also functions in many established religious communities. In either, education tends to take a back seat. The fastest growing segment of global Christianity in the twentieth century was Pentecostalism, and as a new pattern of religious expression within Christianity, it was inclined toward this pattern of ministerial qualification and authorization. For some, authenticity in religious leadership is determined by gifts and abilities that are God-given rather than human-learned.

The early Puritans in North America (and others in many branches of Christianity throughout its history) vested a central qualification for ministry in theological knowledge. They advocated a learned ministry and, as a result, founded Harvard, Yale, and other institutions for the education of ministers. A learned ministry involved religious fidelity and conviction, to be sure, but also valued knowing religious texts and tradition and being able to communicate them. Advanced education for ministry in colonial America often resulted in ministers being among the most-educated members of the community. This pattern of ministerial qualification continues, and most established Christian churches or denominations in North America and elsewhere in the world have institutions for the advanced education of their respective clergy. The hundreds of theological schools in North America, with their curricula, disciplinary specializations, and qualities characteristic of higher education institutions, are the legacy of the learned ministry model.

Implications for Contemporary Education of Clergy

These patterns of clerical qualification and educational practice have a palpable influence on contemporary education of clergy. All three of them are present, both as unique models and in various combinations. No other profession in the developed West is subject to the variations inherent in clerical education. One can be considered a qualified clergy person, and duly authorized for ministry, on the basis of a graduate professional degree or no formal education at all. Few other professions have this kind of free-market form and flexibility. In the midst of these all but endless variations, a normative pattern has emerged in North America. It is graduate professional education, and while bent in many ways, it is the dominant model and the focus of this article.

Characteristics of Education of Clergy

The education of clergy in North America has several salient characteristics, including differing levels of education, a common curricular design, a perceptible educational culture, signature pedagogies, accreditation of degree-granting institutions, and the unique pattern of licensure following graduation.

Levels of Clerical Education

During the twentieth century, postbaccalaureate degree education emerged as the normative pattern of clergy education in the United States and Canada for Protestants, Roman Catholics, and most Orthodox Christians, as well as for the education of rabbis. Alongside this dominant model, two other models have continued with a sizable number of educational providers and students enrolled: baccalaureate degree education with a concentration in theological studies and nondegree certificate education.

Graduate professional theological education

Graduate professional theological education involves the equivalent of either 3 or 4 years of full-time study that results in the master of divinity degree (MDiv). Schools offering the master of divinity are either free-standing theological schools (most often called seminaries) or divinity schools related to a larger educational institution, such as a university. In general, these degree programs include classroom-based education, context-based education, and often some form of clinical-based education in hospitals or other institutions. Four-year MDiv programs typically require the equivalent of an entire academic year in congregational or parish-based work – either during 1 academic year or in the three summers between 4 academic years. Faculty in schools offering the MDiv typically hold research doctorates or, in pastoral arts disciplines,

either a research or professional doctorate. In the United States and Canada, the median enrolment of the more than 250 schools that are members of the Association of Theological Schools (ATS) is about 200, and the typical faculty numbers are 12–15, with disciplinary expertise in a variety of fields of study. Rabbinic programs of study for several Jewish denominations require 5 years of postbaccalaureate study, and most institutions require at least some residential study in Israel. Successful completion of the 5-year program results in the master of arts in Hebrew letters and/or ordination to the rabbinate, which in Judaism is an action of the rabbinic faculty. Buddhist institutions offer a master of Buddhist studies as preparation for ministry or chaplaincy, and Islamic schools offer a master of arts in Islamic studies, although this degree may or may not lead to the recipient becoming an imam.

Baccalaureate and certificate education

Another model for the education of Christian clergy in North America uses baccalaureate and nondegree programs of study. Some baccalaureate programs are offered by church-related or Christian colleges, and others are offered by Bible colleges. The Bible college movement began in the late nineteenth century among conservative Protestants who thought that ministry addressed so many urgent issues that training should be focused and limited to essential skills needed for the work. Bible colleges continue to offer 4-year programs of study that include the equivalent of 3 years of courses in Bible, theological disciplines, and ministry studies, and 1 year of courses in general education. Successful completion of the program of study results in degrees such as a bachelor of ministry, bachelor of theology, or bachelor of arts in biblical studies. Compared to liberal arts colleges, Bible colleges typically have much smaller enrolments and fewer degree programs. Nondegree programs are abundant, but because of their small size and local focus, statistics about them are difficult to obtain. These programs are typically offered by individual congregations or local denominational agencies for persons who are serving in ministry but do not have a baccalaureate degree. In North America, these programs have often been the means of education for bivocational clergy (persons who work in a church but have a job in an unrelated area as primary means of income) or persons who belong to immigrant communities or communities that have limited access to post-secondary educational opportunities.

Educational levels outside North America

These three levels of theological education – graduate professional, baccalaureate professional, and nondegree certificate programs – are characteristic of Christian theological education worldwide. In countries with advanced educational systems, such as many Asian countries, the graduate professional model is dominant. In countries with modest educational infrastructure, certificate training

programs are most typical, as they are for certain groups throughout the world. Europe and the United Kingdom continue an older sophisticated pattern of 5-year professional baccalaureate programs or university degrees from a faculty of theology followed by church-related nondegree programs for pastoral formation.

Theological Curriculum

Across these educational models, and perhaps most dominant in the graduate professional model of clergy education, the central curricular goal is “the development of theological understanding, that is, aptitude for theological reflection and wisdom pertaining to responsible life in faith” (in the language the Association of Theological Schools (ATS) accrediting standards). Within this overarching goal, clergy education typically addresses four major areas of study:

1. the text and tradition of the religious community served by the school;
2. issues related to the community and context in which ministry is undertaken;
3. skills and abilities needed for ministerial leadership; and
4. spiritual and personal development.

Schools vary in the ways they approach these four areas and the amount of attention they give to each; however, most programs of ministerial education include some educational attention to all four.

Theological understanding

The overarching goal of education for religious leadership is not specialized knowledge or technical professional skill. Ministerial work certainly requires this knowledge and these skills, but they are not the ultimate educational goal of the theological curriculum. The ultimate goal is a theological wisdom that provides an undergirding and overarching frame for religious life and work. A religious leader is, first and foremost, a religious person, and religious leadership emerges from religious identity. Ministry begins with a pervasive religious vision, however defined, that orders perspective on life. Excellence begins with theological wisdom pertaining to responsible life in faith – however that may be understood or construed. A curriculum for clergy education begins with this affirmation and continues with the more typical professional school emphasis on areas of study.

Text and tradition

Religious vocation requires an intellectual knowledge of the texts and tradition of a faith community. The theological curriculum includes: study of sacred texts and how to interpret them, theology and how doctrine has emerged historically, and the history of the long tradition in which

religious visions are located. Excellent ministerial leadership requires an intellectual grasp of the tradition of the faith community. Ministers should know something, and they should know it not as information that can be recalled but as constructs that can be argued, analogies that can be sustained, and images that have intellectual credibility.

Ministry contexts and broader culture

Religious leadership entails attending to the public voice and work of the religious tradition. The curriculum typically requires some classroom study and contextual experiences that help students understand the social dimensions of religious communities and the role of religion in broader social contexts. This may involve ministries of mercy as congregations mount programs to the poor, homeless, hungry, and others in need; or, it may involve public witness as the religious community brings its commitments to public debate and policy.

Abilities requisite to ministry

Religious vocation also requires a number of professional skills and abilities. These abilities include preaching, liturgical arts, teaching, counseling, administration, and other aspects of ministerial practice. Religious leadership involves the exercise of a wide range of activities, and the theological curriculum seeks to maximize the quality with which these practices are performed. These activities are not the applied version of what is learned more theoretically in theological or philosophical studies. Activities of professional practice, while theologically grounded, have their antecedent theoretical constructions in areas such as communications theory, educational theory and practice, organizational development, and counseling psychology.

Spiritual awareness and moral sensibility

Ministerial leadership involves spiritual awareness and moral sensibility and character. Somehow, the pervasive religious vision must take root in the way the minister orders his or her life. The minister is not spiritual or moral on behalf of parishioners or congregants, but pursues this kind of integrity because he or she has a theological understanding pertaining to responsible life in faith. The presence of spiritual awareness and moral responsibility in ministerial lives contributes to the congregation's confidence in the priest's or minister's capacity to lead. Seminary students do not graduate spiritually or morally mature – these are lifelong tasks for people of faith. They are not easily learned from formal educational strategies; they form over time as people deal with texts, human beings, ministry contexts, and the struggles of human existence.

The Educational Culture of Theological Schools

Theological schools tend to be educationally intense environments because students are grappling with what they believe and hold to be true. Their subject matter extends to perceptions of ultimate meaning and truth. Theological education invites students to discern a religious vision of right human relationships as much as it instructs them in how to interpret texts and traditions. The result of this intensity is the ability of theological schools to form a particularly powerful form of educational culture. An educational culture is not the result of structured educational processes; the faculty never votes to decide what the school's culture will be. Rather, the culture exists alongside formal curricular structures as a powerful educational force with observable impact on both professional socialization and personal formation.

One study of the culture in two theological schools explored the process by which culture functions as an educational agent. It begins with a dominant cultural message that identifies a vision of right and wrong, truth and untruth, as well as of acceptable and unacceptable patterns of human relating. In a variety of ways, from a variety of voices, this message is presented and represented. Theological schools, however, are not single-idea environments, and there are alternative cultural messages that compete with the dominant cultural message. In one study of students in two very different theological schools, students could articulate the central cultural message as well as the alternative messages that compete with it. In addition, students bring religious convictions and perceptions with them to seminary, and these perceptions are often at odds with the central and alternative cultural messages. The result is a struggle between competing and compelling visions. The struggle often leads to assimilation of the dominant cultural message, but it does not end there. The struggle continues as students find themselves in ministry contexts where the theological school's more rarified message is tested against the muddled realities of ministerial practice. In the end, students appear to come to a resolution in which they affirm some, but not all, of the religious perceptions that they brought with them to the seminary; affirm much, but not all, of the central message the seminary conveyed; and come to a resolution by the end of their seminary study that differs both from what they brought to theological study and from the central message of the school.

Signature Pedagogies

The Carnegie Foundation for the Advancement of Teaching included clergy education in its study of education for the professions, and that study, more than any other, has identified the signature pedagogy that is present in theological education. Theological education, the study noted,

focuses pedagogically on four activities: interpretation, formation, contextualization, and performance. Together, these four pedagogical intentions constitute a signature pedagogy in theological education – an educational strategy that signals the most central elements of ministerial practice. The signature pedagogy in theological education is ubiquitous. It is evident in the work of the faculty as a whole, in course after course, in different ways, and in different combinations, as they teach with a focus on interpretation, formation, contextualization, and performance. Students do not experience one of these pedagogical strategies in their study of one discipline and another strategy in another discipline. They experience many of them, in different combinations, in most of their classes.

Interpretation

Theological faculty members spend a great deal of time on the interpretation of texts. Most theological learning is textually based, and interpretation is crucial to theological learning and pastoral practice. Interpretation is a complex activity. Professors use different processes, but most of them guide students in learning how to interpret a text. Interpreting religious texts involves close reading and analysis, but that is not all. Interpretation is never only about a text; it is a process of using a text in the context of situations and relationships. These characteristics of interpretation are closely associated with how theological educators understand critical thinking.

Formation

Formational pedagogy aims at students learning the “dispositions, habits, knowledge, and skills that cohere in professional identity and practice, commitments and integrity.” Formational learning is critical to theological studies, and it is central to the deepest intentions in professional service that are present in other professions as well. It seeks to cultivate personal character, intellectual capacity, and professional abilities into a human whole. The result is a habit, a disposition, which undergirds the life and work of the religious leader. Formational pedagogy is more than religious language for professional formation. Ministry is a profession that depends on the religious leader affirming for himself or herself what religion holds true for all others; it is a form of professional practice in which the personhood and religious commitments of the minister are as crucial to effective practice as the knowledge and skills these leaders employ in their work.

Contextualizing

The third pedagogy, contextualization, is a function of the way in which all religious knowledge is situated in social, historical, relational, or other contexts. To function faithfully in ministry, students need to learn the contexts of

texts, historical events, religious practices, and ministerial work. Contexts are not just backgrounds that serve as settings for texts or religious practices; they involve patterns of relationship, in the Carnegie study, that range from social structures to personal dispositions and habits. Contextualization involves the pedagogical efforts by which religion's long and ancient traditions are brought into in dialog with current realities.

Performance

A final pedagogical focus is on performance. Ministry has its public leadership dimensions, and students need to learn the performance skills involved in preaching, liturgy, public leadership, and other ministerial tasks. Competence in the performance of these public dimensions of ministry is crucial to the leadership of communities of faith, but in ministry, performance has a unique character. The minister or priest performs as a person of faith, and this performance guides the community into its own shared and corporate faith. When a minister preaches a sermon, the goal is not applause for the performance but a deepened or more-informed faith in the listener. Ministerial performance is not like an actor who takes on the role of a character. Rather, it is an outward, public representation of an inward disposition.

Accreditation

In the United States, institutions that provide education for clergy are accredited in one of four ways, depending on tradition and level of theological education that they provide.

Graduate professional theological education

Graduate professional schools of theology are accredited by the Commission on Accrediting of the Association of Theological Schools in the United States and Canada (ATS). ATS accredits about 235 seminaries, divinity schools, and schools of theology, all of which have as one of their educational tasks the education of clergy. The ATS Commission on Accrediting accredits institutions according to institutional standards and approves each degree program offered by the school on the basis of degree program standards. ATS member schools currently reflect the broad range of Christian churches and denominations in North America. Jewish institutions that provide postbaccalaureate programs of study are accredited by one of the six regional accrediting associations in the United States and are also eligible to pursue accreditation with ATS. The regional agencies accredit a wide range of schools in one geographical area. Because regional agencies do not specialize in one kind of educational institution, their standards are broadly institutional and can be applied to colleges, universities, and special-purpose institutions such as seminaries. In addition

to their accreditation by ATS, about 80% of free-standing Christian seminaries are also accredited by a regional accrediting body.

Baccalaureate and certificate theological education

Bible colleges are the primary providers of baccalaureate-level theological education for conservative Christians, and they are accredited by the Association for Biblical Higher Education (ABHE), which has about 120 accredited institutions. While some of these schools are also accredited by a regional agency, ABHE serves as their primary agency for accreditation, affiliation, and articulating the aims and purposes of baccalaureate theological education. Several of these schools also have graduate professional seminaries, most of which are accredited by ATS. Bible colleges with seminaries provide both baccalaureate and graduate, professional forms of education for clergy. The Rabbinic and Talmudic schools that train Orthodox rabbis are accredited by the Association of Advanced Rabbinic and Talmudic Schools. Their degree programs are generally 4 or 5 years in duration and involve concentrated study of Hebrew texts and the Talmud. While no agencies accredit the nondegree institute and certificate programs of theological education, some denominations approve certificate or nondegree programs for clerical certification.

Licensure and Continuing Education

Unlike most professions in North America, clergy are not certified or licensed by a governmental entity. They are certified – ordained – by a bishop in some church bodies, by an individual denomination in other church bodies, and by individual congregations in still others. Each of these religious bodies determines the criteria for ordination or other forms of licensing, and each one of them may have a dominant educational qualification; however, alternative sets of qualifications also lead to authorization for ministry. Because of these multiple variations, ministerial credentials have limited public usefulness other than that the clergy person has been duly authorized or certified by the procedures of that person's religious community. When clergy participate in certain government-related functions, such as presiding at a marriage ceremony, a governmental authority may request demonstration that they are credentialed clergy, but the governmental authority typically does not certify clergy to perform their religious duties. This varies significantly on the basis of whether a country has an established state religion or not. In a similar way, the criteria for continued certification, including requirements for continuing professional education, are determined by the religious community authorizing the clergyperson, and these range from regular annual requirements to none at all.

Further Reading

- Aleshire, D. O. (2008). *Earthen Vessels*. Grand Rapids, MI: Eerdmans.
- Bendroth, M. L. (2008). *A School for the Church*. Grand Rapids, MI: Eerdmans.
- Brereton, V. L. (1990). *Training God's Army: The American Bible School*. Bloomington, IN: Indiana University Press.
- Carroll, J. W., Wheeler, B. G., Aleshire, D. O., and Marler, P. L. (1997). *Being There: Culture and Formation in Two Theological Schools*. New York: Oxford University Press.
- Cherry, C. (1995). *Hurrying Toward Zion: Universities, Divinity Schools, and American Protestantism*. Bloomington, IN: Indiana University Press.
- Farley, E. (1983). *Theologia: The Fragmentation and Unity of Theological Education*. Philadelphia, PA: Fortress Press.
- Foster, C. R., Dahill, L. E., Golemon, L. A., and Tolentino, B. W. (2006). *Educating Clergy: Teaching Practices and Pastoral Imagination*. San Francisco, CA: Jossey-Bass.
- Holifield, E. B. (2007). *God's Ambassadors: A History of the Christian Clergy in America*. Grand Rapids, MI: Eerdmans.
- Johnson, D. A. (ed.) (2001). *Vanderbilt Divinity School: Education, Contest, and Change*. Nashville, TN: Vanderbilt University Press.
- Kelsey, D. H. (1992). *To Understand God Truly: What's Theological about a Theological School*. Louisville, KY: Westminster/John Knox Press.
- Kelsey, D. H. (1993). *Between Athens and Berlin: The Theological Education Debate*. Grand Rapids, MI: Eerdmans.
- Marsden, G. M. (1987). *Reforming Fundamentalism: Fuller Seminary and the New Evangelicalism*. Grand Rapids, MI: Eerdmans.
- Miller, G. T. (1990). *Piety and Intellect*. Atlanta, GA: Scholars Press.
- Miller, G. T. (2007). *Piety and Profession*. Grand Rapids, MI: Eerdmans.
- Schuth, K. (1999). *Seminaries, Theologates, and the Future of Church Ministry*. Collegeville, MN: Liturgical Press.
- United States Conference of Catholic Bishops (2006). *Program of Priestly Formation*, 5th edn. Washington DC: United States Conference of Catholic Bishops.
- White, J. M. (1989). *The Diocesan Seminary in the United States: A History from the 1780s to the Present*. Notre Dame, IN: University of Notre Dame Press.
- Wheeler, B. G. (ed.) (1993). *Auburn Studies*. New York: Auburn Center for the Study of Theological Education, Auburn Theological Seminary.

Relevant Websites

- <http://www.auburnsem.org> – Auburn Theological Seminary Center for the Study of Theological Education.
- <http://www.abhe.org> – The Association for Biblical Higher Education.
- <http://www.ats.edu> – The Association of Theological Schools in the United States and Canada.
- <http://www.wabashcenter.wabash.edu> – Wabash Center for Teaching and Learning in Theology and Religion.
- <http://www.oikoumene.org> – World Council of Churches, Ecumenical Theological Education.

Framing Professions Education

P S O'Sullivan, UCSF School of Medicine, San Francisco, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

A profession is generally defined as requiring advanced knowledge. This definition of a profession does not tell us much about professionals nor about how to educate them. Here we introduce a number of professions and the specific approach to their education, including curriculum, pedagogy, accreditation, and licensure. The list of professions is not exhaustive, but these articles provide insights into the many issues, changes, and challenges that the professions are facing and how these influences are addressed in education. A number of our authors have provided context for this overview. We have drawn on their writings to frame the introduction, including the development of a profession (Heywood) and the work of the Carnegie Foundation for the Advancement of Teaching (Benner, Sutphen, and Wegner).

Historically, professions have developed because of a certain mystique that surrounds their knowledge which is needed and valued by diverse people. These practitioners develop formal associations and devise ways to control entry to the profession. The profession establishes a system of education and defines standards for acceptance of a professional into the practice. Importantly, the profession adopts codes of professional ethic that provides assurance to the public that their interests are protected. Education is the primary means by which entry into the profession is controlled. This process can lead to a scarcity of members of the profession, thus enhancing its prestige. Moreover, since professions have distinct practices, the education is a balance of the tension between practical and theoretical knowledge.

The education of a professional has been studied primarily by sociologists. In a series of studies on education for the professions, the Carnegie Foundation for the Advancement of Teaching proposed a threefold framework of metaphorical apprenticeships useful in conceptualizing the dimensions of professional education and comparison of professional education across disciplinary fields. The first, cognitive, apprenticeship focuses on developing students' thinking skills in the specific context and content related to the profession. The second apprenticeship, of skill and practice, focuses on developing students' abilities to understand and intervene in particular contexts and to perform as expert professionals responsible for the well-being of others. The third apprenticeship, of identity and purpose, concerns the development of students' appreciation for professional roles, possible conflicting dimensions of those roles, ethical obligations, and individual meaning derived by professionals from the

work they do. Taken together, these dimensions result in professional formation that brings together epistemology, knowledge, skill in action, and professional responsibilities and beliefs.

The Carnegie Foundation studies also highlight much about domain-specific educational practices and pedagogies. Each professional discipline needs to design curricula and pedagogies in relation to integrative teaching of the three professional apprenticeships specific to their particular discipline. Discipline-specific teaching and learning are based upon a deep understanding of the practice and structure of the discipline. While much can be learned from the general body of research on teaching and learning, each discipline, particularly practice disciplines, needs to teach in ways that give students the best access to the practice so that they understand the notions of good and practical goals and boundaries of the discipline.

The framework from the Carnegie Foundation for the Advancement of Teaching emphasizes the power of signature pedagogies, that is, characteristic approaches to teaching that are widely adopted by instructors and programs across a field of professional education, generally reflecting an alignment of theory and practice in the given field, and possessing unusual power to shape understanding of the nature of knowledge and professional roles. Shulman has argued that signature pedagogies have three dimensions: a surface structure (the action of teaching and learning), a deep structure (based on assumptions about how best to teach), and an implicit structure (reflecting judgments about attitudes, values, and dispositions in the field). In addition, signature pedagogies tend to have a shadow side, reflecting what it leaves out. Shulman has attributed the power of signature pedagogies to several factors. Typically, they involve pervasive repetition and routine, resulting in habits of mind that can be employed, almost automatically, when engaging in complex problem solving. They generally require students to perform in a role, necessitating activity, interaction, and visibility within a public setting before others that fosters accountability. They require students to grapple with uncertainty in order to develop professional judgment. Often, the emotional stakes are high (coupling excitement with anxiety), resulting in experiences that shape students in profound ways affecting their values and dispositions as members of a particular profession.

Research by scholars such as Dunne, Eraut, Grossman, McDonald, Shulman, and Sullivan has expanded the line of inquiry on teaching and learning in the professions.

All professional disciplines share the need for teaching and learning in practice situations, where students learn to use practical reasoning across time, as situations unfold. All must use practical reasoning in addition to the formal analytical thinking that they are taught in their practice discipline, whether it be diagnostic 'rule in' and 'rule out' thinking or standard guidelines for public practices. They also learn the appropriate use of different types of critical and creative, imaginative thinking. As all practice disciplines use practical reasoning, practitioners typically must learn to recognize and characterize the nature of the practical situation as it presents itself. The logic of practical reasoning lies in recognizing the nature of the practical situation, staying open to feedback about one's practical grasp and moving from there. Practical reasoning occurs through engagement in the situation rather than standing outside the situation in an abstract, completely detached style.

The entries in this section will describe these issues as they apply to a number of traditional and emerging professions. There are several descriptions that merit highlighting as a preview. The description of the professions emerging from the allied health sciences captures the various struggles of control and recognition to obtain the stature as a profession. A number of professions describe the development of competencies that provide challenges to the curriculum, pedagogy, and assessment that goes well beyond the traditional emphasis on knowledge. Assessment, particularly of skills beyond knowledge, is increasing prevalent in every profession. Performance assessments are particularly important since learners are expected to perform in a context similar to the practice arena. This is a higher fidelity of assessment and an indication as to how someone may practice than was available from traditional paper-and-pencil examinations and, perhaps, observations. Technology has highly influenced the

nature of both the educational and assessment processes. In general, licensure and accreditation characterize most professions and what is notable in the context of these topics is the degree to which these efforts are tending to expand globally, facilitating the ability to practice the profession in other than the jurisdiction where one was educated.

Further Reading

- Benner, P., Hooper-Kyriakidis, P., and Stannard, D. (1999). *Clinical Wisdom and Interventions in Critical Care, a Thinking-in-Action Approach*. Philadelphia, PA: Saunders.
- Bourdieu, P. (1990). *The Logic of Practice*. Stanford, CA: Stanford University Press.
- Dreyfus, H. and Dreyfus, S. E. (1986). *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.
- Dunne, J. (1993). *Back to the Rough Ground: Practical Judgment and the Lure of Technique*. Notre Dame: University of Notre Dame Press.
- Eraut, M. (1994). *Developing Professional Knowledge and Competence*. London: Falmer Press.
- Grossman, P. and McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal* **45**, 184–205.
- Heywood, J. (2002). Autonomy in the classroom: Toward the professionalization of teaching. In Heywood, J. and Taylor, P. (eds.) *School Autonomy*, pp 69–78. Düsseldorf: European Forum on Education Administration.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive and Computational Perspectives)*. Cambridge: Cambridge University Press.
- Shulman, L. (2005). Signature pedagogy in the professions. *Daedalus* **134**, 2–59.
- Shulman, L. S. and Wilson, S. M. (2004). *The Wisdom of Practice: Essays on Teaching, Learning, and Learning to Teach*. San Francisco, CA: Jossey-Bass.
- Sullivan, W. M. (2004). *Work and Integrity: The Crisis and Promise of Professionalism in America*. San Francisco, CA: Jossey-Bass.
- Sullivan, W. M. and Rosin, M. S. (2008). *A New Agenda for Higher Education: Shaping a Life of the Mind for Practice*. San Francisco, CA: Jossey-Bass.

Issues in Accounting/Business Education

W H Gijsselaers, Maastricht University, Maastricht, The Netherlands

R G Milter, Johns Hopkins University, Baltimore, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Attestation – The evidence by which something is affirmed to be correct, true, or genuine.

Certification programs – Require some amount of business education at higher levels as a baseline to earn designation as business professional.

Content-based curriculum notions – Puts focus on questions of What should we teach, and not on pedagogy (how should we teach?).

Criticasters – A petty or inferior critic.

Functional silos business school curricula – Business taught around key disciplines (accounting, finance, strategic management, and marketing) with strong foundations in scientific approaches to problems as they occur in business practice.

Handelshoyskole and Handelshochschule – Business Schools founded in the European tradition to prepare business professionals for the trade profession with a strong orientation on bookkeeping, commerce, and economics. Courses are typically taught with a strong orientation toward business practice.

International accreditation bodies – International organizations that accredit business schools to promote and improve higher education in business education.

MBA degree – Master of Business Administration degree pursued for career enhancement for business professionals with at least 2 years of work experience in various backgrounds (engineering, law, healthcare, etc.).

MSc degree in business – Master of Science degree for students without initial work experience.

National accreditation bodies – Deal with keeping standards for academic education.

a program was developed that included courses in accounting, mercantile law, economics, finance, and statistics. Such courses are still considered as cornerstones of any modern business-education program. Over the past century, business education has held a unique position with business schools serving as knowledge-creation institutions through research and by delivering substantial amounts of bachelor and master graduates.

Accounting education went through a similar development as business education. Students who wanted more practical business training went to commercial schools. These schools frequently evolved into separate schools for business training only. Again, it was the Wharton School leading the way by introducing an accounting course in 1883. It can be easily understood that in those days, academics considered the newly developing programs in business and accounting as too practical or too much aligned to the needs of business practice. Yet, accounting became very quickly the strongest specialization within business education in terms of student numbers and disciplinary development. Today, accounting is considered to be indispensable for our complex modern society, because accountants judge whether companies' annual reports are in accordance to US generally accepted accounting principles (GAAP) or, for businesses located in other countries, to their accounting standards. As such, accountants play a major role in society, and can be considered as true professionals when it comes to licensure to practice, registration requirements for professional association, and professional courts of appeal. In many Western business schools, accounting is a required portion of their basic-degree programs. A few institutions of higher education offer stand-alone accounting programs (without offering other business degrees).

Available figures about undergraduate enrolments and degrees granted demonstrate that the market share of business degrees fluctuates around 20% in the United States, and 10% in the United Kingdom. During the second-half of the 1990s, business education worldwide saw an erosion of its enrolment. Despite these fluctuations, it was demonstrated that market share of business degrees fluctuated between 15% in the 1970s in the US and reached a high of 24% in the late 1980s.

Graduate business education shows similar developments. Western-style business education in Europe and North America deliver professional degrees – the master of business administration (MBA) degree and the master of science degree in business (MSc) for students without initial

Definition of a Profession

Overview

Business education has demonstrated unprecedented growth since the first bachelors program in 1881 was offered at the University of Pennsylvania in 1881. It is said that through the efforts of businessman Joseph Wharton,

work experience. The MBA degree is pursued for career enhancement for business professionals with at least 2 years of work experience in various backgrounds (engineering, law, healthcare, etc.). MBA education has become a big industry. In the 1990s, about 75,000 MBAs graduated each year in the US. The numbers of MBA programs and students has continued to grow since that time and growth is prominent across the US, Europe, and Asia.

The fields and industries served by business education providers are very diverse. There is no standard model for business-education programs, nor is there a standard higher education platform compared to other professions. Providers range from traditional university-based business schools, for-profit institutions, and independent business colleges sponsored or governed by chambers of commerce. Next, hybrid faculties or institutions can be found offering business degrees within a setting of business and economics-degree programs, or colleges for business and engineering within technical universities. Finally, business itself provides programs through in-company training, or within the framework of corporate universities, which can also result in business degrees in cooperation with universities or other accredited institutions.

Typical business school curricula hold a strong functional orientation. Business schools have become highly specialized places that teach business around key disciplines (accounting, finance, strategic management, and marketing) with strong foundations in scientific approaches to problems as they occur in business practice. While on one hand business research has developed strongly along these disciplinary lines, passionate calls have been voiced to organize teaching and research around themes, projects, or problem situations as they occur in business practice. The arguments in favor of this call are similar to those pronounced in many other professions (e.g., medicine and healthcare): people in real jobs do not encounter carefully formulated disciplinary problems or neat stylized prepackaged cases, but they work within places with messy data, in environments of calculated chaos, or in situations where the impending problems are not even recognized. The evident answer to these criticisms is that business scholars work in academic institutions that seek generalizable theories by using analytic tools to examine data – data that are preferably quantitative.

One of the key concerns – and debates among business scholars – is the currency and relevance of business curricula. While core courses and electives follow the organization of business schools in functional silos, the issue is whether boundary-spanning courses or cross-disciplinary courses might better address the problem of keeping business curricula up to date. Another concern is whether the educational methods of many business programs provide true understanding of business practice. This debate goes further than the traditional questions as to whether lectures are a place of learning, and whether they inspire

and help students gain an understanding of course contents. The issue is whether one can understand business practice without having the experience of business practice. Again, this issue resembles previous debates in other professions such as medicine or healthcare.

Historical Perspective

Mintzberg provides an excellent brief history of business education in his book about MBA education in the US. He describes how the first master degree in business was offered in 1900 by Dartmouth College. Harvard (1908) and Stanford (1925) followed with master degrees, while bachelor programs in business became more and more *en vogue*. The major input for these master programs came not from business but from the established academic disciplines, for example, economics, sociology, psychology, and political science.

The establishment of bachelor and master degree programs in business was a response to the increasing need for trained managers due to progressive industrial developments. Through research and education, business schools sought to introduce scientific methods in the analysis of business processes. However, pretty much from the outset, doubt was raised about the academic credibility of business education. In the early 1950s, concerns were expressed about the academic standards at both undergraduate and postgraduate levels. Reports (Carnegie Report and Ford Report) were commissioned by the US government to recommend changes in business-education programs. In the 1960s, similar developments took place in Europe.

Schools in Europe had already developed a strong tradition in vocational education focused on training commercial and management skills such as bookkeeping. For example, several European countries established vocational high schools with an explicit focus on education in trade. Norway had the tradition of *Handelsbøyskole*, and Germany its *Handelshochschule*. During the 1960s and 1970s, universities and independent schools emerged out of these schools dedicated to academic research and training in business. Another typical European progression is that many business-education degrees developed out of economics programs. Universities in several West European countries offered economics programs with specializations or electives in business economics and accounting. Around the 1980s, these programs attracted more and more students and turned into business-administration programs at the (under)graduate level.

With respect to MBA education in Europe, tremendous growth can be observed. It is said that in 1988, only 1200 students graduated with an MBA degree in the UK. In 2000, more than 200 MBA programs were offered by 100 universities and business schools delivering about 10,000 MBA graduates. Substantial increases in market demands for MBA graduates also caused another development. More

and more MBA students do not fit within the traditional college populations (18–24-year-olds). Data about student enrolment in MBA programs can be found at the Association to Advance Collegiate Schools of Business (AACSB)'s website. This site shows that substantial shifts can be noted in the demographic characteristics of the business degree student populations with respect to the demand for part-time education, the amount of work experience before they enter the program, and the increased average age of students.

Specialized Technical Knowledge, Habits of Mind (Academic Training)

Criticisms on the nature of business education seem to be fueled by two leading arguments. First, the question is raised whether business school curricula overemphasize that management is a science and not a profession based on best-evidence practices. One of the critical issues is whether business education is overly scientific and out of touch with business realities, and whether business schools concentrate on research which has little to do with the needs of the business world. Second, it has been argued that business graduates do not have a realistic understanding of the business world. The issue is whether graduates are sufficiently prepared to appropriately respond to current work situations. Attention is needed to insure that graduates in management possess adequate problem-solving abilities, decision-making skills, and leadership skills.

Clearly, these issues are not exclusive to business education. Other domains in professional education deal with similar concerns. In medical education, calls for change have been expressed since the publication of the early Flexner Report in 1910. As medicine took on professional status in the early twentieth century, there was a definite need to establish clinical education in the basic laboratory sciences. Due to a lack of scientific foundations, medical training was in need of educational reform. Flexner was one of the first to recognize that, due to the advancement of medicine, clinical practice should be connected to theoretical laboratory sciences.

Tacit Body of Skills Shared by Competent Practitioners (Apprenticeships)

Before cognitive-science approaches became fashionable to analyze managerial practice, data collection was based on interviews or survey data from human-resource executives, management practitioners, members of executive advisory boards, program directors, and graduates. A consistent research finding is that about four required competencies may be identified that are needed to function adequately in managerial practice. Graduates are expected to possess (1) functional competencies (discipline specific), (2) systemic competencies (cross-disciplinary knowledge

and skills), (3) personal competencies (self-management), and (4) organizational competencies (managing others). It has been found that employers put a growing emphasis on interpersonal skills such as interpersonal communication, team-building, and cognitive skills such as problem solving.

Current cognitive research on expertise development in business, management, and accounting suggest that students, graduates, and experienced business professionals go through similar stages in expertise development as has been implemented in medicine. For example, it has been demonstrated that graduates and students possess significant amounts of factual knowledge, but are not yet able to transform their school knowledge into dynamic knowledge. This finding suggests that the significance of acquiring formal knowledge and formal learning is commonly over-emphasized. Graduates typically possess large amounts of formal knowledge but this is not enough for solving practical problems. This finding confirms claims from critics that formal business education produces only limited conditions for the development of managerial expertise. The absence of developing practical know-how within business schools is a point of concern for educational reform, because it seems to be an extremely powerful learning tool for development of managerial expertise.

Values and Attitudes Shared by Professional Community (Ethical and Moral Dimension)

The values and attitudes shared by the professional business community across industry boundaries are perhaps best formulated by the Ethics and Compliance Officers Association in their standards of conduct for business ethics and compliance professionals as displayed below:

1. Responsibilities to my employing organization:
 - be a role model in adhering to my employer's code of conduct;
 - proactively advocate the integration of ethical business practices and a commitment to compliance into all aspects of my employer's business;
 - ensure to the best of my abilities that my employer upholds all relevant laws and regulations wherever business is conducted;
 - be a leader in the formation of ethical business practices in support of evolving business strategies and opportunities, taking into consideration legal requirements, customs, and best practices;
 - raise and escalate, as necessary, significant business ethics and compliance issues;
 - protect confidential information obtained in the course of my professional activities unless disclosure of such information is required by law, applicable regulation, or company policy, or if maintaining the confidentiality of such information would create an appreciable health or safety risk; and

- avoid any actual, potential, or perceived conflicts between my personal and business responsibilities, and promptly disclose and resolve any issues that may arise.
2. Responsibilities to the profession:
- maintain exemplary standards of personal and professional integrity;
 - strive to continually advance my knowledge of business ethics and compliance;
 - work both individually and collectively with other members of the business ethics and compliance profession to advance the development of business ethics and compliance; and
 - take advantage of opportunities to improve public understanding of business ethics and compliance and their importance to sound business management.

Cross-Cutting Issues

Accreditation of Programs

Next to national accreditation bodies that deal with keeping standards for academic education, business education faces the rather unique situation of international accreditation bodies in the US and Europe. In the US, the AACSB has developed from a national organization (a not-for-profit corporation of educational institutions) that accredits US business schools to an international organization to promote and improve higher education in business education. It was founded in 1916, and established the first set of accreditation standards for business schools in 1919. To date, AACSB has accredited about 560 member institutions of which about 30% have additional specialized accreditation for their accounting programs. About 100 accreditation council members are at colleges and universities outside the United States. It is important to note that accreditation by AACSB is for schools not programs with the exception of accreditation for accounting programs.

In Europe, there are two leading accreditation organizations: the European Foundation for Management Development (EFMD) and the Association of MBAs (AMBA). EFMD is a global organization devoted to the continuous improvement of management development. It has accredited more than 100 business schools across 33 countries. The brand name for its accreditation activities is European Quality Improvement System (EQUIS). EQUIS is not primarily focused on the MBA or any other specific program. Its scope covers all programs offered by an institution from the first degree to the PhD. The institution gets accredited, not an individual degree program. In addition to EQUIS, EFMD has developed other accreditation services for individual programs within institutions. While EFMD mainly focuses on accreditation of institutions,

AMBA's accreditation service is dedicated to all MBA, doctor of business administration (DBA), and master of brand management (MBM) programs. AMBA, established in 1967, is considered to be the international impartial authority on postgraduate business education. AMBA is a global organization and has accredited MBA, DBA, or MBM programs at more than 150 business schools.

Of the approximate 3000 business schools worldwide, about 1% of these schools hold the so-called triple-crown accreditation. This connotes that the business school or its programs is accredited by AACSB, EFMD, and AMBA. Within the community of business schools, receiving the triple crown holds an element of prestige.

Licensure and Certification of Individual

Degrees, licenses, and certifications are many when we look at business professionals. Most of the licensures and certification programs require some amount of education at higher levels as a baseline to earning their designation. Many business-school programs provide instruction specifically geared to equip an individual with the necessary knowledge and skills needed to perform the requirements for earning licenses and certifications. All of the programs that provide instruction and certification provide a code of ethics or some form of professional standards for those professionals who hold their designation. To follow is an overview of some of the most widely supported license and certificate programs.

Certified public accountant

One certification that is also tied to state/federal regulations in the United States is the certified public accountant (CPA). The CPA is a license that is required to serve in public accounting as an independent auditor. State laws govern the specific authority and responsibilities of the CPA. Although specific requirements may vary by state, most require an undergraduate degree with 120 semester hours. There is also a requirement for 24–30 semester hours of specific coursework in accounting that could be earned via an undergraduate or graduate degree program. Many states also mandate an experience criterion that includes a minimum of 1–2 years experience in audit or accounting work. As the CPA license typically takes 18 months to complete beyond the requirements listed above, many students opt to pursue a master's degree that fulfills these additional requirements. Beyond the classroom and experience requirements, individuals must sit for a 14-h computerized exam administered by the American Institute of Certified Public Accountants, which serves as the governing body for CPAs in the US. The exam consists of four key sections: (1) auditing and attestation, (2) financial accounting and reporting, (3) regulation, and (4) business environment and concepts. For more on this license, see the AICPA website.

Chartered financial analyst

Another popular designation among students at business schools is the chartered financial analyst (CFA). The CFA designation requires a broad understanding of finance implications across a number of other disciplines including accounting, economics, statistical analysis, and portfolio management. Unlike the CPA, the CFA is not regulated by state or federal governments and there is no legal requirement to hold the designation in order to serve as a financial analyst. Being a CFA charterholder provides recognition of financial expertise that is often accepted as parallel to holding an MBA with a concentration in finance. The CFA designation is granted by the CFA Institute that provides a rather extensive routine in which to earn the charter. For more, see the website of the CFA institute.

Certificate in investment performance measurement

The certificate in investment performance measurement (CIPM) targets ethical excellence and technical expertise in the application of investment performance evaluation and presentation, analytical techniques, and standards. In order to earn the CIPM designation, one must have 2 years professional experience with calculating, analyzing, evaluating, or presenting investment results; providing consulting services related to such activities; verifying global investment performance standards (GIPS) compliance; supervising persons who perform such activities; or teaching such activities. Alternatively, one can substitute 4 years experience in the investment industry. Candidates must agree to comply with the CIPM Association code of ethics and standards of professional conduct and pass both a principles and expert exam. For more, see website of the CFA institute.

Financial risk management

A widely accepted professional certification for senior risk managers is the financial risk management (FRM) that benchmarks knowledge regarding strategic areas of FRM including: credit, market, operational, and investment-management risks. Certification requires a passing score on the FRM exam, active fellow membership in the Global Association of Risk Professionals (GARP), and a minimum of 2 years experience in the area of financial risk management or another related field. For more on FRM, see the website of GARP.

Earning licenses and certifications are more and more required for business professionals who provide advice about financial investments, or work as personal financial advisor. Depending on their activities, different certificates and licenses are required.

Professional Code of Conduct (Ethics)

As mentioned earlier, all the university, licensure, and certification programs contain some statement of purpose

that includes a code of conduct, ethics code, or other statements of expected professional behavior. An increasing number of corporations and businesses are joining the Ethics and Compliance Officer Association (ECOA), formerly called the Ethics Officer Association. Membership in that organization has grown significantly since its inception in 1991 (current membership is over 1300 ethics and compliance professionals around the globe). The association provides a platform for knowledge expansion and thought leadership across all industries for business leaders holding compliance and/or ethics positions. The association's vision is to be the recognized authority on business ethics, compliance, and corporate integrity. More information is available at ECOA's site.

Globalization and Diversity: Progressive Responsibilities of the Professional

There is an absolute resounding call around the globe for greater accountability by business professionals. More than that, business organizations are reaching out to demonstrate value to customers, shareholders, employees, and society. Narrow focus on short-term profits has been demonstrated as a path to destruction as businesses, both large and small, have been dissolved over the past decade. Today, business models offering sustainability must include much larger vision statements and much larger landscapes that hold all stakeholders, not just shareholders. Since the turn of this century, we have experienced a shrinking, flatter world where technology has made it possible to collaborate and share information instantaneously across time and space. The ease at which consumers can access information adds to the urgency and importance of accurate and legitimate information dissemination by business leaders. Business futurists like Peter Drucker and Charles Handy have been providing accurate, albeit provocative, forecasts on the look and feel of the future of organizational life for sometime now. The future of management and leadership has been proclaimed by such thinkers as Warren Bennis, Gary Hamel, and Peter Senge. Although taking different paths, each of these scholars has highlighted an urgency for business leaders to begin to think and act as professionals. It is clearly time to accept the fact that future organizational life will be very different from how things worked in the past. This difference must be accepted and promoted by a new breed of leaders: A breed that must take on increasing progressive responsibilities.

Signature Pedagogy

In order to respond to these progressive responsibilities of business leaders, business education must accept the challenge to help in the development of requisite capabilities to lead the organizations of the future. Business professions have consistently called for more dynamic and

responsive learning platforms to prepare the next generation of leaders, yet universities have been unresponsive to such calls. Some valiant models have been provided that target a more vibrant, responsive, and integrated approach to developing business leaders and industry leaders have also weighed in on the importance of upgrading the educational platforms to better address current and future business needs.

Business schools have sought many ways to bridge the gap between theory and practice by incorporating business practice into their curricula through cases (Harvard case method) and projects, or through learning methods that heavily rely on understanding practice (e.g., action learning, project-based learning, and problem-based learning). However, curricular-reform visions seem to be driven by content-based notions (what should we teach?), and less by pedagogy (how should we teach?). In this way, business education is lagging behind by developments in other professions (especially medicine and healthcare) that went through profound instructional changes based on educational-research findings and input from professional practice. On the other hand, business education is beginning to close the gap with other domains in professional education as is reflected in its emphasis on accreditation, and its strong links with business practice and societal needs.

See also: Engaged Education: Experiential Learning, Intensive Field Experiences, and Social Change; Knowledge Domains and Domain Learning; Problem Solving and Human Expertise; Professions – Health-Related.

Further Reading

- AACSB (2002). Management education at risk. *Report of the Management Education Task Force to the AACSB International Board of Directors*. AACSB International, Tampa, FL, August.
- Arts, J. A., Gijssels, W. H., and Boshuizen, H. P. A. (2006). Managerial expertise development: Diagnostic reasoning and problem-solving. *Contemporary Educational Psychology* **31**, 387–410.

- Barsky, N. P., Clements, M., Ravn, J., and Smith, K. (eds.) (2008). *The Power of Technology for Learning. Advances in Business Education and Training*, Vol. 1. Berlin: Springer.
- Bennis, W. G. and O'Toole, J. (2005). How business schools lost their way. *Harvard Business Review* May, 1–9.
- Boyatzis, R., Cowen, S., and Kolb, D. (1995). *Innovation in Professional Education*. San Francisco, CA: Jossey-Bass.
- Drucker, P. F. (2003). *Peter Drucker on the Professional of Management*. Boston, MA: Harvard Business Press.
- Ericsson, K. A., Charness, N., Feltovich, P. J., and Hoffman, R. R. (eds.) (2006). *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press.
- Ghoshal, S. (2005). Bad management theories are destroying good management practices. *Academy of Management, Learning and Education* **4**, 75–91.
- Hamel, G. and Breen, W. (2007). *The Future of Management*. Cambridge, MA: Harvard Business School Press.
- Handy, C. (2002). *Elephant and the Flea: Reflections of a Reluctant Capitalist*. Boston, MA: Harvard Business School Press.
- Milner, R. G. (2000). Innovation in learning methodologies for adult learners: Implications for theory and practice. *Virtual University Journal* **3**, 228–236.
- Mintzberg, H. (2004). *Managers not MBAs. A Hard Look at the Soft Practice of Managing and Management Development*. London: Prentice Hall, Financial Times.
- Van Wyhe, G. (1994). *The Struggle for Status: A History of Accounting Education*. New York: Garland Publishing.
- Wagner, R. K. (2002). Smart people doing dumb things: The case of managerial incompetence. In Sternberg, R. J. (ed.) *Why Smart People Can Be So Stupid*, pp 42–63. Yale, CT: Yale University Press.
- Watson, S. R. (2006). Management education as if both matter: A response to Jonathan Gosling and Henry Mintzberg. *Management Learning* **37**, 429–431.

Relevant Websites

- <http://www.aacsb.edu> – AACSB International (Association to Advance Collegiate Schools of Business).
- <http://www.aicpa.org> – AICPA (American Institute of Certified Public Accountants).
- <http://www.allbusinessschools.com> – AllBusinessSchools.
- <http://www.mbaworld.com> – AMBA (Association of MBAs).
- <http://www.cfainstitute.org> – CFA Institute.
- <http://www.theecoa.org> – ECOA (Ethics and Compliance Officer Association).
- <http://www.efmd.org> – EFMD.
- <http://www.garp.com> – Global Association of Risk Professionals (GARP).

Professions – Engineering

J Heywood, Trinity College Dublin, Dublin, Republic of Ireland

© 2010 Elsevier Ltd. All rights reserved.

Introduction

A striking feature of social change in the nineteenth and twentieth centuries was the rise of the number of occupations into the ranks of the professions. Engineering was among the earliest to seek this status. This has not been without controversy, for what are the characteristics of a profession?

The occupations that concern us here are ones based on extensive academic and practical knowledge and/or social skills and, in time, they generate a special mystique of their own. The clients of these occupations constitute a varied and heterogeneous population that offers considerable potential for exploitation by members of the occupation. As a result, the moral practitioners establish associations to create and protect a monopoly and to control entry. Education and training have been and are used as the prime means of selection. This discriminatory role would seem to be as important as their role in the provision and improvement of knowledge for it seems that sometimes an education is provided that is not suited to the work task. They practice self-regulation and, if possible, seek to have it placed on a statutory footing. In the UK, in order to become a profession an occupation builds up a system of education and training and defines methods and regulations for acceptance; and in order to establish a position of responsible guarantors of the public interest they adopt a code of practice or professional ethic. Taken together, all these serve to provide the professional person with status and British society clearly acquiesces in the hierarchy that has evolved. As there is a potential for conflict between public and professional interest, public support for these restrictive practices will depend on the perception that it has of the effectiveness with which a professional organization deals with malpractice. In this respect, there has been some criticism of regulatory bodies, for example, law and medicine.

It is against this general context that the engineering profession is considered but primarily in respect of education and training, codes of practice, and professional autonomy with a focus on identity and status. Developments in Europe, North America, and the UK are considered within their historical context. While there are many similarities, there are also considerable differences.

Learned Societies and the Development of the Professional Associations

When the Royal Society was founded in 1660, its members met on a weekly basis to discuss experiments and other topics of scientific interest. Its purpose is expressed in its title in the Royal Charter it received in 1663, namely, *The Royal Society for Improving Natural Knowledge*. It was a society of amateurs and persons actively engaged in scientific work. It was a learned society and it published proceedings, and great importance was attached to the development of a library. In 1847, it imposed membership requirements and Fellows (FRS) now had to be elected on the basis of their scientific work. A very small number of Fellows are elected each year. Fellowship is a scarce commodity that is highly prized by the scientific community and generally valued by society. It is an élite. A Royal Charter provides it with a legal framework of governance and carries with it considerable prestige.

Engineers in Britain have long wanted to emulate the Royal Society and in so doing substantially improve the status of the profession. In 1976, the Royal Academy Engineering was founded. It is representative of every branch of engineering. In that sense it provides engineering with an identity. Sixty Fellows are elected each year and that makes fellowship a scarce commodity. Its equivalent in the USA is the National Academy of Engineering (NAE) which has had a more positive role in policymaking. The NAE was founded in 1964, 101 years after the National Science Foundation (NSF), by the same act of incorporation as the NSF. Its advisory role to government is spelt out in its charter thus “whenever called upon by any department or agency of government to investigate, examine, experiment and report upon any subject of science or art.” Like the Academy in the UK, whose brief does not extend that far, it is an independent, nonprofit institution and proclaims to “provide engineering leadership in service to the nation.” Nevertheless, distinguished American engineers continue to decry the lack of engineers in prominent positions as they do in the UK, and hence the ability to influence policy. However, it is the professional associations in the UK and USA that are of interest here rather than the academies.

James Smeaton, an FRS who coined the term civil engineer to distinguish between his kind of work

(e.g., constructing lighthouses) and that of military engineer, founded the Society of Civil Engineers in 1771. It was a learned society. Later in 1818, the Institution of Civil Engineers was founded to become both a learned society and the first professional association of engineers. It obtained a Royal Charter in 1828 with the purpose of fostering and promoting the art of civil engineering that was seen at that time to embrace all engineering. However, The Institutions of Mechanical and Electrical Engineers were founded in 1847 and 1871, respectively. These institutions greatly valued their learned function and in particular their libraries. Numerous other institutions were created in the twentieth century representing various specializations in engineering.

In 1887, irrespective of the fact that there were university departments of engineering, the Institution of Civil Engineers set its own qualifying examinations. By 1889, they had raised the standards to correspond with that of a general university degree. The other institutions were to follow suit. In 1921, the government initiated a system of national certificates for work done by technical college students on part-time courses. Assessors of the higher stages of these courses were appointed by a joint committee that was representative of the professional institutions and the Ministry of Education. The assessors reviewed the examination papers and marking. The complete course to degree level took about 7 years. Until the 1970s this was a major route for training electrical and mechanical engineers. By that time it had been agreed that all engineers should be educated in full-time degree programs.

The Chartered Engineer and Professional Engineer Qualifications

By the 1960s a need was felt to bring some coherence to the educational activities of these numerous institutions. To meet this need they formed a Council of Engineering Institutions (CEI) and this became their examining body for the award of the chartered engineer (C. Eng) qualification. Subsequently, it was replaced by an Engineering Council UK. Its Royal Charter empowered it to regulate the engineering profession in the UK. It functions through 36 institutions that are licensed by the Council to place members on its register in one of three grades – chartered engineer, incorporated engineer (previously technician engineer), and engineering technician. Placement on these registers requires skill in a range of competencies and practical experience.

Membership of the American professional societies does not entitle a person to practice engineering in the sense that they can take legal responsibility for the documents they sign. This is particularly important in civil and mechanical engineering. Each state requires such persons to be licensed and they are recognized by the letters PE

(professional engineer) after their names. Each state has its own requirements for licensure but there is also an examination set by a National Council of Examiners for Engineering and Surveying (NCEES). There is a three-stage process after graduation that begins with a ‘Fundamentals of engineering’ examination: this is followed by work experience and completed by an examination in the ‘Principles and practice of engineering’ (PE). After 2015, candidates for the PE will require a master’s degree in engineering which would seem to align it with the Bologna process (see below). A distinction is made between graduate engineers and professional engineers. A university teacher is not required to be a PE. Engineers employed in interstate commerce are not regulated and some industrial exemptions are given. In some states the title engineer is protected.

An Innovative Merger

The most recent development in the UK is the merger of the Institution of Electrical Engineers (IEE) with the Institution of Incorporated Engineers (IIE) to become the Institution of Engineering and Technology. The use of the term technology clearly responds to popular usage and may help resolve the identity issue. The IEE was a society that granted chartered status to electrical engineers. The IIE granted incorporated status to a group that had been called engineering technicians. They are broadly equivalent academically to holders of technology degrees in the USA. The IIE embraced technicians in mechanical and general engineering and had merged with other technician institutions in the electrical and radio fields to become the IIE. The new institution thus embraces both professional and technician engineers as do the American Societies.

The Origins of Validation (Accreditation) in the UK

In the UK, these developments were paralleled between 1956 and 1992 with a search for status and identity by those educational institutions in the public sector allowed to do degree-level work; in short, they sought university status. The overt intention was to improve the quality of technological education. In 1956, a National Council for Technological Awards (NCTA) created a system of validation of diplomas awarded by ten of these colleges, and that is, arguably, the origin of current systems of accreditation in the UK. The professional engineering institutions participated in that practice. These colleges became universities in 1967. NCTA was replaced by the Council for National Academic Awards (CNAA) which was created for all other colleges in the public sector with degree programs irrespective of subject, and by 1992 all those institutions had become universities.

France and the Origins of the US System of Engineering Education

No problems of identity and status were experienced in France. From the beginning there was a deliberate attempt to create *corps d'élite* of military engineers through the development of a college – the *École Polytechnique* that would provide a high-level education with emphasis on science together with studies in practical subjects and the humanities. A similar *Grands École* to provide a similar kind of élite for industry was created in 1829. Their mission is the same today. Engineering acquired status from the beginning and in continental Europe engineers are held in much higher esteem, particularly in Germany than in Britain and North America.

In the USA, Rensselaer Polytechnic was founded in 1823 and its curriculum was subsequently remodeled so that by 1850 it mirrored that of the *École Centrale des Arts and Manufactures* in Paris. That model of parallel studies in the physical sciences, mathematics, technical subjects, and the humanities remains the basis of engineering curricular in North America to this day.

The American Societies and the American Society for Engineering Education

In 1852, the American Society of Civil Engineers was founded for all engineers. Had it remained for all (and likewise the Institution of Civil Engineers) the profession may have had a stronger identity. However, as in England, societies were soon formed for mechanical and electrical engineers. They never provided a qualifications route that was left to the universities. One outcome was that in 1893 some teachers in the US organized the Society for the Promotion of Engineering Education that made some influential recommendations. It later became the American Society for Engineering Education (ASEE) and again in that guise it was responsible for a number of very influential reports. It is represented in the accreditation process that began in 1932 as the Engineers Council for Professional Development (ECPD) now Accreditation Board for Engineering and Technology (ABET). ASEE has enabled a continuing discussion to take place about matters concerning engineering education in a way that the professional institutions in Europe and the UK have not. During the last 2 or 3 years academics have begun to talk of engineering education coming of age. Thus, it is possible to conceive of an engineering education profession. Given precedents from school teaching it is possible to distinguish between restricted and extended professionalism among engineering educators. For example, restricted professionalism regards teaching as an intuitive activity, whereas extended professionalism regards it as a rational activity supported by research and evaluation. Thus, extended

professionals will not only be conversant with research and development (R&D) in their own fields of engineering, but also in that of engineering education. They will keep abreast of both subject areas. The perspective of restricted professional is limited to the immediate time and place, whereas that of extended professional embraces the broader social context of both engineering and engineering education. Whereas a restricted professional believes skills are derived from experience, the extended professional believes that skills are derived from a mediation between experience and theory, etc. The Centre for Scholarship in Engineering Education (CASEE) of National Academy of Engineering (NAE) which is encouraging research and best practice in engineering education is encouraging extended professionalism among engineering educators in the USA.

The Washington Accord and Developments in Europe

In 1989 an international agreement called the Washington Accord was signed. Its purpose was to develop an instrument that would enable the mutual recognition of professional degree programs across the world. At the heart of this task is the mutual recognition of systems of accreditation. Support for this move came from a pan-European organization representative of the engineering institutions: *Fédération Européenne d'Associations Nationales d'Ingénieurs* (FEANI). This would require agreement on approaches to program accreditation. In Britain, modern principles of accreditation were established in 1956 by the NCTA. Procedures were established in France in 2007, Germany in 2002, Portugal in 2008, and in Russia much earlier in 1992. Within Europe, FEANI brought together a group who proposed a European accredited engineer. There followed a European Union (EU)-funded project to establish an agency for accreditation (the European Network for Accreditation of Engineering Education, ENAEE) that was incorporated in 2006. It has established accreditation for bachelor and masters degrees. This work relates to the Bologna Declaration signed in 1999 by 45 countries with the intention of creating a European system of higher education that facilitated student mobility. The pattern of education that has been agreed follows two cycles. The first cycle is of 3 years duration for a basic degree and a second cycle of 2 years for a master's degree. A broad interpretation of the Declaration enables universities to apply it to engineering programs. There is a debate about whether graduates from the engineering science program are employable after 3 years, although there is a 3-year program that leads to an engineering technology qualification. Whereas in some universities in Germany the master degree programs are designed to allow engineering technology graduates to enter

these programs directly, this is not generally the case. Normally, they would be expected to pursue an additional program of study prior to entry to the master's program.

Education, Occupation, and Professionalization

The foregoing remarks illustrate the importance attached to educational qualifications by organizations seeking to become professions and professions seeking to enhance their status. However, the content also contributes to perceptions of identity and content is very much a product of history. There is little doubt that when such persons as James Smeaton set up these fledgling societies they were trying to escape from the image of engineering as something done by craftsmen. They wanted to promote its scientific dimension (theory) and this has had consequences that may or may not have been beneficial. Therefore, it was that universities on both sides of the Atlantic have tended to emphasize the scientific dimension at the expense of design and the academic at the expense of the practical. Promotion depends on research in some aspect of engineering science and design can count for little. Not only do university departments of engineering have to provide courses and assessments that meet these requirements, their selection criteria have to ensure that students are capable of success in the courses they offer. In Britain, this affects the choice of subjects (A levels) that pupils make in high school, and in the highly specialized curriculum of the high school this means physics, mathematics, and generally another science. First, this puts them in competition with physics as a subject of study at university because physics departments want the same qualifications and in the past physics has had higher status. Second, it prevents them from taking other subjects that might be valuable. Cambridge University has refused to accept a diploma in engineering developed with the support of the Engineering Council UK unless the candidate has physics at A level. For example, very few universities regard metal work as suitable, yet it is an ideal subject for developing the spatial ability that is essential not only for design but also for mathematics. Third, like physics, engineering gives the perception of it being an extremely difficult subject. In both subjects, there has been a decline in the number of applicants and this has been a matter of considerable concern.

There has been particular concern in the English-speaking world to increase the number of women entering engineering, and the number obtaining high positions in engineering and engineering education. Some have argued that the curriculum is insufficiently humane to attract women. Programs that might, such as service (engineering in the community) courses, have low status as does the low-tech approach required when the community in which students work is in a developing nation. High

status is attached to work in high tech. Yet, according to an American report from the Carnegie Foundation "the widespread emphasis of theory over practice . . . discourages many potential students while leaving graduates with too little exposure to real-world problems and ethical dilemmas."

Codes of Conduct and Professionalization

As important as education is to occupations, seeking to become a profession is the code of conduct. Unlike the other professions that have a unitary code of conduct, various associations (e.g., ABET, the National Society of Professional Engineers (NSPE) and the Institute of Electrical and Electronic Engineers (IEEE)) in the USA have developed codes of conduct all of which differ. The same is true of the UK. However, most recently, the Royal Academy has published a code. Michael Davis in his critique of engineering ethics makes the point that in the USA an engineer could be subject to different codes and this could make it difficult for him/her to know what to do. It is not far fetched to equate the code with identity and having to choose between codes his/her identity might become confused. Davis supports the explanation that codes of conduct are important because they are a contract between individuals that protects them from certain consequences of competition. "They provide a guide to what engineers might reasonably expect of each other. [...] it should also provide a guide to what we may expect other members of our profession to help us to do." Inherent in that view is the responsibility for the development of new members of the profession. The more important points are that (1) the code provides unifying framework of values and (2) if all engineers adhere to the same code, then an engineer is in a position to explain to other engineers why a particular course of action is unjustified and not be in the position of putting his/her job at risk. Davis who uses the story of the *Challenger* disaster to make his point concludes that "in practice the ethics of engineers is as important to the success of engineering as good design or testing is." That suitably translated could apply to any human endeavor. The implications for the teaching of ethics are profound because it makes it much more than a discussion of codes of conduct. It would also seem that a unitary code goes some way to providing an underlying bond to the fragmentation that is perceived by some to have lost engineering its professional identity, provided that there is common skill (competency – ability base).

Professional Autonomy

One of the complaints that this approach overcomes is the view that engineers are employees and, have therefore, to

do what their employers tell them. The problems of whistle blowing are much discussed. Without detailing Davis's discussion of these matters he does raise an issue that is common to professions that are made up mainly of employees and that is the question of autonomy. This applies in particular to teachers where micromanagement has been a feature of some governments' attempts to improve schooling. It has been argued that these strictures have led to the deprofessionalization of teaching. The question has been how in a situation of serious constraints on teaching, teachers' professional autonomy can be made consistent with employer requirements? In respect of teaching, it has been argued that the way in which teachers are socialized into teaching coupled with limitations in their training leads it to a form of restricted professionalism. Moreover, it is argued that extending teacher professionalism would enhance rather than limit the possibility of providing the better education that employers seek. Bowen argues that caring, which is an important characteristic of the teaching and medical professions, should also be fundamental to an engineering ethic: "The special capabilities for promoting wellbeing that engineers possess bring special responsibilities." Davis suggests that research needs to be implemented on professional autonomy. One approach might be to examine the attitudes of professionals at work to restricted and extended professionalism and how organizational structure and the values associated with it influence professional behavior.

Emerging Issues

The academies in China, the UK, the USA, and similar organizations have become interested in philosophy and engineering. Much of what is being argued in workshops and seminars in recent years relates to the identity of engineering. Is there a philosophy of engineering that is distinct from a philosophy of science? The question arises from the fact that the emphasis in engineering courses is on science subjects and their application and the consequent view that engineering is but applied science. Very often, industry will employ physicists to do what engineers do, which suggests that some industrialists see little difference between engineering and physics. An important subquestion about which there has been much discussion is: In what ways do the thought patterns (processes, ways of thinking) of engineers differ from those of scientists? Questions such as these raise important epistemological issues and have a bearing on the curriculum. They are directly concerned with the questions: What is engineering? Also, what is an engineer? Questions are also being asked about content and approach to the teaching of ethics, its relation to and provision in the curriculum for moral development. One of the potential challenges that

might arise is that should the answers come up with an alternative to the present curriculum, a substantial paradigm shift in attitudes and values would be required. As indicated in many reports, design has very little prestige in spite of the fact that it is central to engineering. Liberal studies have little or no place in the hierarchy. If Rosalind Williams' thesis that the fragmentation of engineering has caused engineering to lose its identity and that in the future it will have to work hand in glove with the liberal arts, then the shift is unlikely to be made. Others have argued that this will leave engineering where it is now and engineers will be perceived as technicians for the purpose of doing practical, technical work.

Distinguished engineers have sought to improve the status of engineering at the highest levels of political distinction. Other researchers from within and without engineering argue that engineering education has a key role to play and that if it is to do this it has to change. Davis considers a profession to be a "number of individuals sharing an occupation voluntarily organized to earn a living by serving some moral ideal in a morally permissible way beyond what law, market and ordinary morality require" in which case engineering educators are a sub-profession within the engineering profession. They are interdependent and for there to be change in the moral ideal of engineering education, if change is judged to be desirable, it will only be achieved if the profession as a whole regards it to be necessary.

Further Reading

- Accreditation Board for Engineering and Technology (ABET) (1977). *Code of Ethics*. Baltimore, Maryland.
- Bowen, W. R. (2009). *Engineering Ethics. Outline of an Aspirational Approach*. London: Springer.
- Bucciarelli, L. L., Coyle, E., and McGrath, D. (2009). Engineering education in the US and EU. In Christensen, S. H. (ed.) *Engineering in Context*, ch. 5. Aarhus, Denmark: Academica.
- Davis, M. (1998). *Thinking Like an Engineer. Studies in the Ethics of a Profession*. New York: Oxford University Press.
- Heywood, J. (2005). *Engineering Education. Research and Development in Curriculum and Instruction*. New York: IEEE/Wiley.
- Heywood, J. (2009). Ch 8 From restricted practice to professionalism of *Managing and Leading Schools as Learning Organizations: Adaptability and Change*. Dublin: National Association of Principals and Deputies/Original Writing.
- IEEE (2006). *Code of Ethics*. Institute of Electrical and Electronic Engineers. (The code is printed on the back of their leaflet Memberships and Subscriptions 2008.)
- Institution of Civil Engineers, London (1870). *The Education and Status of Civil Engineers in the United Kingdom and Foreign Countries*. (1891) *Engineering Education in the British Dominions*. (1928) Centenary Publication pp 48–51.
- Janausch, K. H. (1990). *The Unfree Professions. German Lawyers, Teachers and Engineers, 1900–1950*. New York: Oxford University Press.
- Macdonald, K. M. (1995). *The Sociology of the Professions*. London: Sage.
- Payne, G. L. (1960). *Britain's Scientific and Technological Manpower*. Stanford, CA: Stanford University Press.

- Royal Academy of Engineering and Engineering Council UK (2005). *Statement of Ethical Principles*. London. Revised 2007 and available on website (accessed 02:03:09).
- Sheppard, S. D., Mcatangay, K., Colby, A., and Sullivan, W. M. (2008). *Educating Engineers. Designing for the Future of the Field*. San Fransisco, CA: Jossey-Bass.
- Silver, H. (1990). *A Higher Education. The Council for National Academic Awards and British Higher Education 1964–1989*. London: Falmer Press.
- WEP (2007). *Workshop on Philosophy and Engineering Abstracts*. University of Delfte. <http://www.illigal.uiuc.edu> (accessed August 2009).
- WEP (2008). *Workshop on Philosophy and Engineering*. Abstracts. Royal Academy of Engineering, London. <http://www.raeng.org.uk> (accessed August 2009).

- Williams, R. (2003). Education for a profession formerly known as engineering. *The Chronicle of Higher Education*. Issue 24th January.
- Wulf, W. A. (2002). The Urgency of Engineering Education Reform. Excerpts from the Laboratory for Innovative Technology and Engineering Education. 2002 Distinguished Lecture 3/3 and July–Dec. 2002.

Relevant Websites

- <http://ethics.iit.edu> – Center for the Study of Ethics in the Professions. Illinois Institute of Technology.
- <http://www.washingtonaccord.org> – Washington Accord (1989).

Professions – Health-Related

G M Jensen, Creighton University, Omaha, NE, USA

C B Royeen, Saint Louis University, St. Louis, MO, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Accreditation – The process of external, quality review process used by higher education to scrutinize colleges, universities, or higher education programs for both quality assurance and enhancement; the external review and approval of an institution or school is done to see if the entity meets specific standards or requirements.

Allied health – Allied health professionals are involved with the delivery of health or related services pertaining to the identification, evaluation, and prevention of diseases and disorders; dietary and nutrition services; rehabilitation and health systems management, among others. Allied health professionals, to name a few, include dental hygienists, diagnostic medical sonographers, dietitians, medical technologists, occupational therapists, physical therapists, radiographers, respiratory therapists, and speech language pathologists.

International Classification of Functioning, Disability and Health – The International Classification of Functioning, Disability and Health, known more commonly as ICF, is a classification of health and health-related domains. These domains are classified from body, individual and societal perspectives by means of two lists: a list of body functions and structure, and a list of domains of activity and participation. Since an individual's functioning and disability occurs in a context, the ICF also includes a list of environmental factors. The ICF is WHO's framework for measuring health and disability at both individual and population levels.

Professional Doctorates – These are degrees sometimes called professional or clinical doctoral degrees; they are practice-based degrees in line with the 'traditional professional degrees' (medical doctor or lawyer); graduates of professional doctoral degree programs demonstrate a level of professional practice appropriate to the purpose of the degree and an expected and demonstrated level of scholarship contributing to the advancement of the profession.

Specialized accreditation – The process of external, quality review process used often in professional education to review programs or professional schools for both quality assurance and enhancement.

Introduction

As a broad-based term, allied health refers to a collection of fields educating practitioners at the associate, baccalaureate, masters, or doctoral levels. In the United States health care system, they represent more than 60% of the health care workforce providing diagnostic, rehabilitation, other therapeutic services, and access to or management of healthcare. For a sense of the scope of these fields, 7 of the 70 (or more) of them are detailed in **Table 1**, which reveals that these fields are high growth areas through 2016, all likely to have a disproportionate impact on education in the United States due to their high demand. Among these allied health professions, we focus on three as prototypical that have experienced rapid growth and change as emerging professions (occupational therapy, physical therapy, and speech-language pathology), which are identified in their own right.

How occupations emerge to become professions is fundamentally a social-political process that is strongly related to the relationship between the occupation and society that includes public recognition for those services that are essential, exclusive, and complex. Sociologists have often argued that medical dominance makes it difficult for other health professions to emerge as full professions. For the purposes of this article, we define emerging health professions among the multitude of allied health professions as those that are demonstrating increasing public recognition for their services, including direct access for those services; professional education at the graduate level in university, a growing research community, and proactive health policy efforts aimed at improving access and quality of healthcare. This article begins with a brief analysis of the genesis and evolution of the term allied health, followed by sections on regulatory issues such as accreditation, licensure, practice domains, and self-regulation; as well as educational institutions and educational trends.

Table 1 Listing of major professions in allied health with number of educational programs, number of existing practitioners, and projected growth based upon Bureau of Labor Statistics

<i>Profession</i>	<i>Number of accredited educational programs</i>	<i>Number of existing practitioners and growth projections from 2006–16</i>
Clinical laboratory science	470 programs	319 000 practitioners 14% growth
Health informatics and information management	NA	170 000 practitioners 18% growth
Nuclear imaging including nuclear medicine technology and radiation therapy	100 programs	20 000 15% growth
Nutrition and dietetics	281 baccalaureate programs 22 masters programs	57 000 9%
Occupational therapy	196 masters programs 5 doctoral programs	99 000 23%
Physical therapy	43 masters 166 doctoral programs	173 000 27%
Speech–language pathology	260 programs	110 000 11%

History of Allied Health

A simple definition of allied health is ‘all health care professionals other than doctors or nurses.’ The term is viewed as pejorative by some since it implies subjugation to medicine, but for practical purposes we will use the term in this article. The role of the American Medical Association (AMA) in allied health history is a critical element for understanding past and current challenges regarding cohesion and identity. Historically, while both occupational therapy and physical therapy had military origins and the need for reconstruction aides during and following World War I served to promote both of these disciplines, the financial difficulties of several early allied health professions caused them to turn to the AMA for support to develop and maintain their educational programs. For several decades, the AMA continued to be involved in helping define and regulate allied health as well as supervise and control its related practice. In 1966, the US Congress passed the Allied Health Personnel Training Act (P.L. 89-751). This legislation led to more focused attention on allied health fields essential for modern health services and was responsible for rapid growth in the number of allied health education programs from 2500 in 1966 to more than 8000 in 1980. During this period, educational programs in rehabilitation (occupational therapy and physical therapy) which were postbaccalaureate, graduate certificate programs resisted the movement of their programs into schools of allied health where the educational degree would be the baccalaureate. These graduate certificate programs were usually hospital based and affiliated with elite, private institutions of higher education.

There has been resistance from the emerging professions to the allied health designation and they have, as independent professional organizations, left this designation behind,

but not without a larger, continued political battle. The AMA still strongly influences who and what are considered real health professions compared to those in allied health professions. The AMA publishes an annual directory of accredited allied health education programs that includes 71 professions. Excluded from the AMA allied health list are medicine (allopathic), osteopathic medicine, dentistry, veterinary medicine, optometry, podiatry, pharmacy, chiropractic, clinical psychology, all levels of nursing education, and graduate degrees in health administration or public health.

Regulatory Issues

Accreditation

Accreditation is a process of external quality review that is used in higher education to scrutinize institutions and educational programs for quality assurance and quality improvement. There are six regional accreditation agencies for institutions of higher learning in the United States which are themselves regulated or recognized (not accredited) by the US Department of Education and by the Council for Higher Education Accreditation (CHEA). One recent challenge with regional accrediting agencies has been the rapid move and multiple requests for a change in degree offering at the institutional level. Several health professions have moved or are in discussion regarding a clinical doctorate degree (e.g., pharmacy, nursing, physical therapy, occupational therapy, and audiology). Given that a large number of health professions educational programs may not be in research intensive, doctoral institutions, but in primarily master's degree institutions, an accreditation review for the institution is required. This movement continues to generate dialog and controversy among

Table 2 Listing of accreditation agencies in the major fields of allied health

<i>Profession</i>	<i>Accrediting organization</i>	<i>Accrediting standards</i>
Clinical laboratory science	National Accrediting Agency for Clinical Laboratory Science	2006 National Accrediting Agency for CLS
Health informatics and information management	Commission on Accreditation of Health Informatics and Information management Education	AHIIM 2005 Standards
Nuclear imaging including	Joint Review Committee on Educational Programs in Nuclear Medicine Technology	JRCNMT Essentials, revised 2003
Nutrition and dietetics	Commission on Accreditation for Dietetics Education	2006 CADE
Occupational therapy	Accreditation Council for Occupational Therapy Education	January 2006 Final Draft of Standards for Accredited master's level Educational Program for the Occupational Therapist and Standards for an Accredited Educational Program for the Occupational Therapy Assistant
Physical therapy	Commission on Accreditation in Physical Therapy Education	2006 Evaluative Criteria for the Accreditation of Educational Programs for the Preparation of Physical Therapists
Speech/language pathology	Council on Accreditation in Audiology and Speech Language Pathology	2008 Standards for Accreditation of Graduate Programs in Audiology and Speech–Language Pathology

deans of graduate schools, schools of allied health or health sciences, and regional accrediting bodies.

In allied health education, a wide range of agencies accredit the education of allied health professionals. This means that an institution has not only a regional accreditor for the entire institution, but also specialty accreditation for many of the programs.

Historical Perspective

Central to all, however, is the development as an independent field of study or profession, moving from an allied health profession to an independent health profession. For the most part, accreditation agencies within the fields of allied health did not originally exist as separate or independent agencies. When some of the first programs in allied health were implemented, they were accredited by the Committee on Allied Health Education and Accreditation (CAHEA), which was part of the AMA. At one time, the AMA accredited 3000 educational programs working with 50 different national professional associations and 22 review committees. At that time, this made sense since most all of these fields were relatively young and developing professions, dependent upon the validity and power of medicine to support them – and hence the origination of the concept that these professions as allied with medicine. Within each of these allied health fields, there is a unique history of how development of a body of knowledge, sociology of professionalism in the profession, and educational methods and actions developed.

As these professions have matured in this manner, the move toward establishing independent accreditation agencies has evolved and, currently, all of the emerging health professions discussed in this article have their own

accreditation agency presented in **Table 2**. The result is that each profession has control of its educational standards, independent of medicine. In 1994, with the dissolution of CAHEA, the AMA ended their direct involvement with accreditation of allied health programs. The Commission on Accreditation of Allied Health Education Programs was formed and reviews and accredits programs across 21 health occupations.

Historically, accreditation educational standards in allied health have been most dominant in the United States. There is some movement, however, to expand the scope of some accreditation agencies in the United States elsewhere in the world. This is likely a future trend to watch as a form of globalization in action.

Licensure and Practice Domains

The development of independent accreditation agencies for these professions is integrally related to the related development of practice domains and licensure within these professions. Along with this growth of emerging professions practice domains has come development of paraprofessional or assistants who carry out the technical aspects of practice, allowing evaluation and decision making to be done by the professional. This section addresses these core practice issues: state regulation of the practice, professional turf issues, and the role of technical personnel.

Practice Acts

Accreditation agencies pertaining to a particular allied health profession depend, in part, upon data about the actual practice of the practitioners in the field to assist

in determining what should be included in educational programs, and therefore influencing accreditation standards and assessment of qualification for practice. The practice of any profession within a state in the United States is based upon the regulatory status in that state, usually designated as licensure or certification. State regulations pertaining to what a professional may and may not do in practice is typically based upon practice acts generated by the profession, groups of members within the profession, or federations of state boards across the profession. This means that in each of the states in the United States, there is a licensure or practice act that is based upon a model or scope of practice. The emerging professions (occupational therapy, physical therapy, and speech-language pathology) have licensure and established practice acts in every state.

Professional Turfs

For the most part, health care professionals are not educated to work in a team – and, hence, there is a lack of connection across health professionals. This disconnect in the United States is supported and reinforced by our insurance and billing systems, and our educational programs. The dominant educational paradigm is that every health care practitioner is educated in a single profession to be the best practitioner in that particular field. Very little, if any, education and experience is devoted to learning about other health care professions, how the health care system itself operates, how patients respond to and participate in the health care system, and how to work with other health care providers to improve patient health and wellness.

The net result of this approach is turf wars. At any given time, we might find dissension across health professions. Cross-purposes exist at the practice level and the educational level which is isolated from other health professionals. The issues will become more complicated with the consideration of paraprofessionals or technical personnel and respective educational settings.

Role of Assistants/Technical Personnel

Most of the allied health professions we are addressing in this article are educated in postbaccalaureate educational programs in 4-year universities or academic health centers. The three emerging professions upon which we focus (occupational therapy, physical therapy, and speech-language pathology) have some form of technical level of education that produces assistants to the therapist. The most structured and consistent technical programs are for occupational therapy assistants and physical therapist assistants. These assistants are similarly licensed or registered in each state, and similarly graduate from specialty certified programs. Most of these programs, however, are at 2-year

institutions and the graduates are – in fact – technicians and not professionals in the truest sense. This is another source of tension.

Ethics and Professional Responsibility

The concept of self-regulation of ethical standards is an important characteristic of professions. Codes of ethics are commonplace across allied health professions. Declaring a code of ethics, however, does not result in self-regulation. This requires the profession's obligation to serve the best interests of their patients, meet societal needs, and enforce the code. If we look at the historical evolution of the various codes of ethics in allied health professions, we can see a changing professional identity. For example, one of the first codes for physical therapists specifically cited is never criticize or disagree with the physician. The critical role of defining and creating a professional identity for these professions is further developed later in this article.

Evolving Professional Responsibilities

In the previous section that mentioned turf wars among all of the health professions, a key component of what all health care professions should focus on was neglected. The first and foremost responsibility of every health care provider is not to their respective profession or beliefs, but to their patient.

Access and Referral

A key element in the development of these emerging professions is direct access and referral to the health professional. Does the health professional need a referral from a physician or other designated health care providers? In physical therapy, there continues to be a strong professional association focus for legislative changes across states that allow direct access to physical therapists. In a different historical pathway, occupational therapy assumed they had direct access in the 1950s as the field began to work in communities, at that point believing they could work without a physician prescription. Thus, occupational therapy has always had direct access, while physical therapy is now working for it. Speech-language pathology also has direct access to clients. The exception for all three of these emerging health professions is the federal Medicare program. All three of these professions have been working together to make changes to the Medicare legislation that would eliminate the need for a medical referral for older patients seeking speech, occupational, or physical therapy services. Physician requirements for referral remain in place as a perceived cost-saving measure ensuring that Medicare beneficiaries receive medically appropriate care.

Educational Trends in Emerging Health Professions

Educational Institutions and Degrees

In addition to seeking control over setting educational standards and forming a profession-specific recognized accrediting body, there continues to be rapid changes in educational degrees in the emerging allied health fields. The primary rationale is that the expanding knowledge base and societal needs require professional education programs of sufficient depth and breadth. Starting as hospital-based programs, many of these fields have moved to baccalaureate programs and most consider master's and clinical doctoral degrees to be adequately prepared.

Physical therapy was one of the first allied health professions to move to the clinical doctorate (doctor of physical therapy, DPT). Now there are several health professions in active discussion or planning of advancing the level of the first professional degree, or the advanced practice degree. For example, occupational therapy has set a postbaccalaureate entry-level requirement; audiology now only accredits doctoral programs and will no longer recognize master's level programs; clinical laboratory science is working on implementing a clinical doctorate as the postprofessional degree; and nursing is moving to a clinical doctorate for the advanced nurse practitioner. Facilitating change for these health professions continues to be an arduous struggle and raises the question whether the predominant feminization of these professions plays a role in the tensions surrounding policy changes within organizations and institutions.

Faculty

The predominant model for faculty across allied health professions continues to be recruitment of an adequate number of clinicians to educational programs. The transition of health professions programs to professional education at the graduate level also places increased emphasis on scholarship. The rapid transition of programs has led many programs to a model of growing their own faculty and actively supporting doctoral-level education for their faculty.

Curriculum

Professional identity

Past curricular discussions often focused on the practical logistics of implementation such as how much content and number of laboratory hours. Currently, focus on professionalism and what it means to be a professional is perhaps the most visible and consistent theme underlying discussions and planning in professional education curricula in the maturing health professions. As these fields have converted their entry-level degrees to a postbaccalaureate

(master's or clinical doctorate), there continues to be an internal dialog and debate as to what this means in terms of depth and breadth of educational preparation. Much of this dialog is centered on role of professional identity and collaborative work in making professional judgments. As other mature professions, such as medicine and law, have refocused and emphasized curricular elements on formation of professional identity and professional commitments to meeting societal need, there is a commitment to focus on self-identity in the emerging health professions. Accreditation requirements, practice guidelines, and other professional documents have all moved well beyond teaching the code of ethics to include a broader and explicit focus on development or formation of a professional and professional responsibility.

Curriculum shift from technical application to knowledge, skill, and decision making

There is a dynamic shift in curricula in emerging health professions. A vast majority of professional textbooks are now written by faculty in the health profession. The foundation knowledge dimension of the professional curricula has grown. This foundation knowledge supports the primary components of the field. For example, occupational therapy has a strong base in the social and human sciences with multiple theory bases in the field, whereas physical therapy and speech-language pathology have more emphasis in the physical and biological sciences. Clinical courses have been redesigned with a clear focus on evaluation or examination of the patient/client as well as treatment interventions and evidence-based practice. Application of theoretical and conceptual models to practice situations is a routine part of the curriculum. All three of these emerging professions (occupational therapists, physical therapists, and speech-language pathologists) have adopted and use the International Classification of Functioning, Disability and Health. This conceptual model emphasizes prevention, function, activity, and participation along with consideration of the environmental and personal factors that may affect function, and is perhaps the single most unifying force for these professions on an international level.

An integrated model of clinical education or field experiences within the academic program has long been the standard in emerging professions. Historically, the educational environments for many of these programs were the hospital where the focus was on a tightly integrated model of knowledge, technical skill, and clinical practice. For most of the emerging professions, when they moved to a university setting, the model of an integrated model of clinical experiences (prelicensure), within the academic program remained. With increased emphasis on developing the thinking, decision making and judgment skills, and the technical skills, there is more discussion and models of

clinical education that would include a postlicensure internship.

Over the last decade, there has been rapid, continued growth of the central importance of clinical reasoning and the decision-making process for these emerging professions. There is no one shared model of clinical reasoning among these professions. While there may be common elements of application of core knowledge, patient-provider interaction, and a reflective inquiry process, the evolving models perhaps provide the best representation of the unique context and contributions of these professions. For example, the clinical reasoning models in occupational therapy have a strong emphasis on multiple forms of reasoning, including narrative, pragmatic, and ethical reasoning throughout the process of client interaction, while the reasoning models in physical therapy are more focused on the diagnostic process and patient self-management.

Pedagogy

Since their inception, the allied health professions have been focused on the development of proficient skills. In the emerging professions, the laboratory, skill-based component of teaching remains strong. While one of the continued challenges has been finding the balance between classroom teaching and laboratory teaching, there remains an essential for students to demonstrate fundamental, safe, clinical skills prior to entering a clinical situation. Skilled laboratory teaching frequently uses a case-based approach that integrates evaluation and examination process with development and performance of realistic interventions. The laboratory portion of the curriculum is tightly sequenced with the didactic component and requires ongoing and extensive collaborative faculty planning. The laboratory component of the curriculum also requires adequate space and supervision of students. It is not uncommon for faculty to have to divide a class into laboratory sections, thus creating a large number of faculty-student contact hours.

If one were to identify a signature pedagogy that represents the characteristic approach to teaching in the emerging professions that is adopted across educational programs, it is likely to be teaching psychomotor skills from multiple perspectives with a reflective component. Each skill requires performance, application, analysis, and teaching the skill across settings. Students must first learn how to perform the skill whether this is the performance of a daily activity or walking up stairs with an assistive device. The student must learn how to apply the skill and analyze the performance of the skill as they teach a client, patient, or family member. The analysis of what is needed for successful teaching or transfer of the skill goes well beyond the physical performance. Students then need to develop reflective insight and problem-solving ability as

they work to design an intervention that is well suited to the environmental context as well as the patients' beliefs, values, and expectations. They must reflectively monitor and assess ongoing intervention and family or societal participation.

Assessment

The model of practical, performance-based examination has been a long-standing assessment strategy in allied health education for clinical skills. Historically, the predominant model has a one-to-one testing situation where performance of clinical procedures was done directly on the faculty or on another student and then scored by the faculty member using some form of scoring rubric. Over time, the availability of facilities and technology for standardized assessments, particularly for programs that are part of academic health science centers, has contributed to the growth of standardized simulations such as the use of standardized patients and objective structured clinical examinations. Formative assessments of both performance and knowledge retention are common in these programs. A variety of test constructions are used across programs and there is more emphasis on case-based scenarios embedded in a standard multiple-choice test format as the national examinations have become increasingly more clinically based.

Clinical Specialization

All the three emerging health professions discussed in this article have some form of advanced specialization in place. The fields vary in the range and intensity of advanced specializations. However, for the most part, these advanced specializations currently occur outside of the academic settings. Typically, advanced specialization is based upon some degree of time and practice with particular areas of study and some sort of national certifying examination or process. These processes may be operated by the professional association or by private agencies.

Interprofessional Education

The international community of health professions in Canada, the United Kingdom, and Australia has established working models of interprofessional education facilitated by a national health care system that supports shared team-based competencies. Several of the US Institute of Medicine reports, including *Health Professions Education: A Bridge to Quality*, conclude that health professionals are neither adequately prepared in either academic or clinical settings to address changes in patient populations nor trained in team-based skills. In the United States, institutions that have schools of allied health or health professions are better able to implement

interprofessional sharing of core courses and learning experiences, although many logistic and professional challenges remain. The future belongs to those who can surmount these challenges and truly implement interprofessional education across health and human science professions.

See also: Curriculum and Critical Theory; Portfolio Assessment; Scoring Rubrics; The Role of Accreditation of Higher Education Institutions.

Further Reading

- Association of Specialized and Professional Accreditors (2008). Statement on Professional doctorates. September 2008. <http://www.aspa-usa.org/documents/ASPA-Statement-Professional-Docs.pdf>
- Boyce, R. (2006). Emerging from the shadow of medicine: Allied health as a 'profession community' health. *Sociology Review* **15**(5), 520–534.
- Cooper, R. A. (2003). The expanding scope of practice and nonphysicians clinicians and implications for medical practice. *Journal of Medicine Licensure Disciplines* **89**, 75–93.
- Council of Graduate Schools (2008). *Task Force Report on the Professional Doctorate*. Washington, DC.
- Donini-Lenhoff, F. G. (2008). Coming together, moving apart: A history of the term allied health in education, accreditation and practice. *Journal of Allied Health* **37**(1), 45–52.
- Dower, C., Christian, S., and O'Neil, E. (2007). *Promising Scope of Practice Models for the Health Professions*. Center for the Health Professions, University of California, San Francisco, CA, USA.
- Golde, C. (2007). Signature pedagogies in doctoral education: Are they adaptable for the preparation of education researchers? *Educational Researcher* **36**(6), 344–351.
- Higher Learning Commission (2006). *A Report to the Board of Trustees from the Task Force on the Professional Doctorate*. <http://www.ncahlc.org/download/TaskForceProfDocFinal0606.pdf> (accessed August 2009).
- Institute of Medicine (2003). *Health Professions Education: A Bridge to Quality*. Washington, DC: National Academy Press.
- Lecca, P. J., Valentine, P. A., and Lyons, K. J. (eds.) (2003). *Allied Health: Practice Issues and Trends in the New Millenium*. Binghamton, NY: Haworth Press.
- Purtilo, R., Jensen, G., and Royeen, C. (eds.) (2005). *Educating for Moral Action: A Sourcebook in Health and Rehabilitation Ethics*. Philadelphia, PA: FA Davis.

- Royeen, C. B., Jensen, G. M., and Harvan, R. (eds.) (2009). *Leadership in Interprofessional Health Education and Practice*. Boston, MA: Jones and Bartlett.
- Shulman, L., Golde, C., Bueschel, A. C., and Garabedian, K. (2006). Reclaiming education's doctorates: A critique and a proposal. *Educational Researcher* **35**, 25–31.
- Sullivan, W. (2005). *Work and Integrity: The Crisis and Promise of Professionalism in America*, 2nd edn. San Francisco, CA: Jossey-Bass.
- Threlkeld, J., Jensen, G., and Royeen, C. (1999). The clinical doctorate: A framework for analysis in physical therapy education. *Physical Therapy* **79**(6), 567–581.

Relevant Websites

- <http://www.aota.org> – Accreditation Council for Occupational Therapy Education (ACOTE), ACOTE Accreditation.
- <http://www.aota.org> – American Occupational Therapy Association (AOTA).
- <http://www.apta.org> – American Physical Therapy Association (APTA).
- <http://www.asha.org> – American Speech and Language-Hearing Association (ASHA).
- <http://www.asahp.org> – Association of Schools of Allied Health Professions (ASAHP).
- <http://www.aspa-usa.org> – Association of Specialized and Professional (ASPA).
- <http://www.caipe.org.uk> – Center for the Advancement of Interprofessional Education (CAIPE).
- <http://www.apta.org> – Committee on Accreditation in Physical Therapy Education (CAPTE).
- <http://www.caahep.org> – Commission on Accreditation of Allied Health Education (CAAHEP).
- <http://www.asha.org> – Council on Academic Accreditation in Audiology and Speech-Language Pathology.
- <http://www.ches.org> – Council for Higher Education Accreditation (CHEA).
- <http://www.fsbpt.org> – Federation of State Boards of Physical Therapy (FSBPT).
- <http://www.isca-speech.org> – International Speech Communication Association (IPCA).
- <http://www.nbcot.org> – National Board for Certification in Occupational Therapy (NBCOT).
- <http://www.wfot.org> – World Federation of Occupational Therapists (WFOT).
- <http://www.wcpt.org> – World Confederation for Physical Therapists (WCPT).

Professions – Pharmacy

D E Beck, University of Florida, Gainesville, FL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Clinical pharmacy – A health science discipline in which pharmacists provide patient care, leading to optimized medication therapy and the promotion of health, wellness, and disease prevention.

Experiential – The learning that occurs by accomplishing tasks and activities in the real-world or practice setting.

Fellowship – An organized and directed postgraduate training program that prepares an individual to become an independent researcher within a pharmacy practice specialty area.

Foreign Pharmacy Graduate Equivalency Committee (FPGEC) Certification Program – A program offered by the National Association of Boards of Pharmacy that allows foreign pharmacy graduates a mechanism for documenting that their foreign pharmacy degree is equivalent to one that is accredited by the American Council for Pharmacy Education.

Internship – A practical experience where an unlicensed individual, who is eligible for licensure as a pharmacist, works under the direct supervision of a licensed pharmacist.

Interprofessional – A reciprocal interaction and collaboration among several different health professionals who are from different health professions.

Medication therapy management – The services that are provided to a patient to optimize therapeutic outcomes and these services are either independent of or occur in conjunction with the provision of the drug product.

Patient-centered care – The direct care of the patient that involves establishing a covenantal relationship with the patient.

Pharmaceutical care – A patient-centered pharmacy practice in which the practitioner assumes responsibility for a patient's medication-related needs and is held accountable for this commitment.

Residency – An organized and directed postgraduate training program that focuses on a specific areas of pharmacy practice and the individual is under the direct supervision of a pharmacist-preceptor.

Introduction

Pharmacists have a unique role in assuring the safe use of medications. Over the years, the pharmacist's role has transitioned from a product focus, which centered on dispensing medications, to a patient-centered focus where the pharmacist shares knowledge with patients and other health professionals and, in the process, optimizes therapeutic outcomes for individual patients. Specifically, the pharmacist's role is evolving beyond dispensing the correct medication and providing counseling to responsibilities – such as providing comprehensive patient evaluations for preventing, identifying, and managing medication-related problems, collaborating with physicians in managing patients with chronic disease, and providing care that promotes health and wellness.

The change from a product focus to a patient-centered focus is still underway within the profession and the extent to which this change has occurred varies at a global level. Many countries refer to these more patient-centered care responsibilities as clinical pharmacy. Although more time is still needed for the profession to fully implement this change, it has already stimulated significant transformations in the licensure, professional code of conduct, the educational requirements, and pedagogy used in pharmacy education.

The goals of this article are to first provide an overview of pharmacy education at a global level and then provide a case study using the United States as an example to understand how changes in the pharmacist's role has impacted licensure and credentialing, the pharmacy code of conduct, progressive practice responsibilities, and pharmacy curricula. Finally, emerging issues related to the academic preparation of pharmacists are highlighted.

Pharmacy Education at a Global Level

At a global level, countries vary in the extent that this transformation has occurred among practicing pharmacists. Almost all countries require a pharmacy degree as a prerequisite for licensure and have a licensure process. Moreover, globally, a growing number of countries are placing greater emphasis on experiential learning and/or internships that prepare graduates for providing patient-centered care rather than just acquisition of knowledge related to pharmaceutical sciences. However, pharmacy

curricula and the degrees conferred greatly vary among academic institutions at a global level.

For example, in China, the entry-level degree into the profession is a 4–5-year bachelor of science (BS) degree. The focus of this degree is on pharmaceutical sciences and the traditional responsibilities of pharmacists. Most graduates are employed in research and development or as representatives for pharmaceutical manufacturers. However, in the last 10 years an increasing number of universities are offering a 5-year BS degree or a 6-year BS/MS degree in clinical pharmacy. Clinical pharmacy is growing in China and there is great demand for graduates of these programs.

In the United Kingdom, the 4-year MPharm followed by 1 year of supervised practical experience (preregistration) is required to become licensed as a pharmacist. Although not required, most programs include experiential coursework in the 4-year course of study. Most graduates enter the profession and practice in the community setting. With additional education and training, pharmacists can assume prescribing responsibilities. This role involves patient-centered care.

In Australia, a 4-year bachelor of pharmacy degree serves as entry into the profession. However, some universities now offer a graduate-entry master of pharmacy degree that requires 2 years of study and is an accelerated pathway for individuals who already have a bachelor's degree in a related discipline. Both bachelor and master degree graduates are required to complete approximately 12 months of internship which is called preregistration training. In Australia, most graduates practice in community pharmacies and hospital pharmacy departments and there is increasing emphasis on patient-centered care.

Across the Middle East, the 5-year BS pharmacy degree typically is required for entry into the profession. Jordan, Lebanon, and Saudi Arabia have moved to the doctor of pharmacy (PharmD) degree, which is typically a 6-year course of study. Advancement of practice has been slower in these countries. Hospital pharmacy typically requires a higher level of practice than that required in the community setting. Most pharmacy schools in the Middle East are revising their curriculums to include a greater focus on patient care and include more experiential education.

In Canada, a minimum of 5 years of postsecondary education is required to attain the BS in pharmacy degree. However, over the last decade there has been considerable discussion about changing the entry degree to the PharmD. Several pharmacy schools in Quebec now offer the PharmD as the entry-level degree. Most graduates in Canada work in community pharmacy settings and patient-centered care is a guiding principle for the profession. Pharmacists in Alberta have achieved prescribing privileges which requires a patient-centered focus of care.

To practice as a pharmacist in France, the *diploma d'état de Docteur* or doctor of pharmacy degree is required and it is awarded after 6 or 9 years of study. In the fifth year of this study, students complete a 1-year internship at a university hospital. During the fifth or sixth year, students specialize in both hospital and organizations pharmacy, and industrial and biomedical pharmacy; besides, they are exposed to specialized topics such as biotechnology, or medical biology (biological testing in analytical laboratories). Discussions are underway to revise these specializations so that one of them will place emphasis on clinical pharmacy.

In the United States, the entry-level degree changed from the 5-year BS pharmacy degree to the 6-year PharmD degree in 2000. For individuals who attended pharmacy school in the United States, District of Columbia, or Puerto Rico, the pharmacy degree must be from an institution that is accredited by the Accreditation Council for Pharmacy Education (ACPE). Individuals who received a pharmacy degree outside of the United States, the District of Columbia, or Puerto Rico may document their educational equivalency by completing the Foreign Pharmacy Graduate Equivalency Committee (FPGEC) Certification Program that is administered by the National Association of Boards of Pharmacy (NABP). The practice transformation and degree change have led to changes in licensure requirements, increased certification opportunities, the professional code of conduct, pharmacist's progressive responsibilities, and pharmacy curricula. The following case study highlights these points.

Case Study – United States

Licensure and Certification

To become licensed as a pharmacist in the United States, an individual must meet the multiple eligibility requirements established by the board of pharmacy in the state where the intended practice site is located. Since transition to the PharmD degree as the degree required for entry into the profession and the movement in the profession to a practice that is patient-centered, foreign graduates desiring to document educational equivalency are now required to have completed a BS degree that required at least 5 years of coursework compared to 4 years of coursework in the past.

In addition to attainment of a pharmacy degree, other licensure requirements include completion of internship hours. The type of practice activities and length of the internship vary significantly among the states.

Once individuals meet the eligibility requirements established by their board of pharmacy, they are permitted to sit for the two national examinations that are required for licensure. To become licensed in a state, individuals must successfully pass the North American Pharmacist

Licensure Examination (NAPLEX) and the Multistate Pharmacy Jurisprudence Examination (MPJE). Both of these examinations consist of multiple-choice test questions and are administered as a computer-adaptive examination. They are developed, administered, and scored by the NABP.

Once an individual satisfies all requirements for licensure within a state, the individual is eligible to practice in any setting. Most often, pharmacists in the United States practice in community, ambulatory, or hospital/institutional patient-care settings.

In the United States, certification is granted by the Board of Pharmaceutical Specialties (BPS) to pharmacists who demonstrate the abilities needed to accomplish evolving, advanced levels of practice. Individuals who attain certification typically assume positions that involve provision of patient-centered care and they also assume leadership roles in advancing practice within their setting.

Certification is not required to practice as a pharmacist. Instead, it is a voluntary process that confirms a practitioner has education, experiences, knowledge, and skills in a specific practice area that is well beyond that required for licensure. In addition to a written examination, some specialties require submission of a portfolio that documents practice knowledge and skills. There are currently five specialty practice areas recognized by BPS and others are planned to keep pace with the movement to patient-centered care. As certificates recognize evolving knowledge and skills, pharmacists who achieve BPS certification must recertify every 7 years.

Other credentials that infer advanced training beyond licensure are pharmacy residencies and fellowships. A pharmacy residency is an organized, directed postgraduate training program which focuses on a specific area of pharmacy practice. The demand for residencies has grown as the profession has moved to a patient-centered practice. Individuals complete a 1-year general residency and, if they desire additional expertise, may also complete a specialized residency that is also of 1-year duration. In all residencies, the resident is directly supervised by a qualified practitioner-preceptor. Residencies typically prepare pharmacists to implement and offer innovative pharmacy services. Pharmacy residents are also prepared to become involved in pharmacy management activities. Majority of pharmacy residencies are accredited by the American Society of Health-System Pharmacists (ASHP).

A pharmacy fellowship prepares an individual to become an independent researcher and it is a highly individualized program that is usually directed by a faculty member who is a recognized researcher. For example, a fellowship may prepare individuals for independent research in areas such as cardiology, critical care, neurology, psychiatry, or pediatrics. Individuals who enter a fellowship must also have prerequisite basic practice skills in the area of practice that is the focus of the fellowship.

Fellowships require a predetermined time period which is typically longer than 12 months.

Over the years, pharmacy fellowships have decreased in number. Increasingly, individuals who want to develop expertise in research complete a combined PharmD–PhD degree. This credential prepares them as clinical scientists for positions in academia and industry.

Professional Code of Conduct

The American Pharmaceutical Association (APhA) led the pharmacy profession in establishing the Code of Ethics in 1852. Since 1852, there have been multiple revisions with the most recent version adopted in 1994. The Code of Ethics describes the moral obligations and virtues that a pharmacist must demonstrate in order to accomplish the roles and responsibilities expected of a pharmacist. The most recent changes in the Code of Ethics align with transformation from a product- to a patient-centered practice.

The Code of Ethics provides pharmacists with guidance in establishing and maintaining relationships with patients, health professionals, and other members of society. The most recent Code conveys that the pharmacist is expected to have a covenantal relationship with a patient and promote the well-being of all patients in a confidential manner. In accomplishing this, the pharmacist is expected to be caring and compassionate while maintaining the autonomy and dignity of each patient. The Code also conveys the expectation that a pharmacist respects the values and abilities of other health professionals and also maintains interprofessional relationships with honesty and integrity. In addition, pharmacists are expected to maintain professional competence. Several sections of the Code also convey that a pharmacist must also serve community and societal needs and that sometimes it is necessary to balance the needs of patients and society.

The Oath of a Pharmacist was developed in 1994 and conveys many of the expectations outlined in the Code of Ethics. In addition, the profession has a pledge of professionalism. This pledge is frequently used in white coat ceremonies held early in a pharmacy education program to instill the conduct expected of a pharmacist. This pledge gives reference to the Oath of a Pharmacist and the Code of Ethics.

Progressive Responsibilities of the Professional

In the 1960s, a group of pharmacists were very dissatisfied with the traditional practice norms and also realized that healthcare needed pharmacists who could provide patient-centered care, such as evaluating a patient's medication regimen and individualizing medication therapy. As a result, the clinical pharmacy movement began.

In the 1970s and 1980s, the practice of clinical pharmacy grew as an increasing number of pharmacy schools offered the PharmD degree. Increasingly, this new generation of pharmacists developed innovative clinical services that included optimizing medication therapy of patients and becoming integral members of interprofessional healthcare teams. Many of these clinical pharmacists focused their practice in specialty areas such as critical care, pediatrics, internal medicine, and oncology. In institutional settings, these pharmacists led development of drug information centers, pharmacokinetic consultation services, and nutrition consultation services. Clinical pharmacy continued to evolve and is now considered to be a health science discipline.

In the 1990s, the pharmacy profession refined these responsibilities and services into a new concept of practice that embraced pharmaceutical care as a philosophy of practice. Pharmaceutical care is a patient-centered practice in which the practitioner assumes responsibility and is also accountable for a patient's medication-related needs.

By the beginning of the twenty-first century, pharmaceutical care had become an established philosophy within the profession. In addition, the academic requirements for entry into the profession in the United States changed from a baccalaureate to a doctor of pharmacy degree that included extensive preparation in the discipline of clinical pharmacy. However, the offering of patient-centered care by every pharmacist did not evolve because reimbursement for pharmacy services continued to be for the drug product and not for the provision of cognitive services.

In the early 2000s the US government initiated discussions with pharmacy leaders about the needs of the elderly population and medication therapy management (MTM) services were identified as a strategy to meet these needs. MTM includes services that are provided to a patient to optimize therapeutic outcomes and are either independent of or occur in conjunction with the provision of drug product. This strategy now provides opportunity for reimbursement of pharmacy services that include the provision of patient-centered care. The provision of MTM requires the pharmacist to exhibit pharmaceutical care as a philosophy of practice and apply knowledge and skills from the discipline of clinical pharmacy.

Increasingly, pharmacists are now providing MTM through contracts with self-insured employers, traditional insured groups, government programs, and individual payers. Not all patients require MTM services. Patients who are ideal candidates for MTM services include those who are receiving medications from multiple prescribers, taking more than four chronic medications, are not taking their prescribed medications, have high monthly medication expenses, recently been discharged from a hospital, and/or have at least one chronic disease (e.g., diabetes, asthma, or heart failure).

When providing MTM, the pharmacist conducts a medication therapy review in consultation with the patient or caregiver. Then, the pharmacist develops or updates a personal medication record for the patient. Next, the pharmacist develops a medication action plan that includes a list of medication-related issues (e.g., dosage change is indicated, another medication would be more appropriate, and adverse drug reaction), a plan for resolving each issue, and the individual responsible for action. Following development of a medication action plan, the pharmacist refers the patient to another health care provider if it is indicated. Finally, MTM services are documented in a consistent manner and a follow-up MTM visit is scheduled with the patient or caregiver if indicated.

Although community pharmacy settings were the initial setting for provision of MTM, pharmacists are now encouraged to expand and provide this service in all patient-care settings, including hospital admission, transition from the hospital to home, or admission to a long-term care facility. MTM is a new innovative practice for pharmacists and the MTM innovation is clearly in the early adopter stage and moving into the early majority stage. Although MTM is changing pharmacy practice, greater advocacy at all levels is needed to promote further growth.

In 2004, the Joint Commission of Pharmacy practitioners established a future vision of pharmacy practice statement that envisions pharmacy practice in 2015. Similar to the MTM innovation, this vision states that "pharmacists will be the health care professionals responsible for providing patient care that ensures optimal medication therapy outcomes." This statement defines:

1. the educational preparation needed by pharmacists to practice in 2015,
2. how pharmacists will practice, and
3. how pharmacy practice will benefit society.

With respect to the educational preparation needed by pharmacists, the statement emphasizes that pharmacists must be prepared to provide patient-centered and population-based care learning to optimized medication therapy. Pharmacists must also be able to manage health care system resources so they can improve therapeutic outcomes. In addition, they must be prepared to help patients promote their wellness, prevent diseases, and improve their health. Specific abilities that pharmacists must develop and maintain are:

1. patient caring;
2. an in-depth knowledge of medications and the pharmaceutical, biomedical, sociobehavioral, and clinical sciences; and
3. the ability to apply evidence-based therapeutic principles and guidelines, evolving sciences and technologies, and the legal, ethical, social, cultural, economic, and professional issues that are relevant to contemporary pharmacy practice.

Signature Pedagogy

In the United States, the PharmD degree requires a minimum of 2 years of defined preprofessional coursework and then acceptance into a pharmacy school where a professional curriculum, which is typically 4 academic years in length, is completed. A few pharmacy schools offer the professional curriculum in 3 calendar years.

The doctor of pharmacy curriculum offered by ACPE-accredited pharmacy schools enables students to achieve competencies needed to provide pharmacist-delivered patient care including MTM services. Accreditation standards further state that students must be prepared to ensure patient safety and optimal medication therapy outcomes of patients.

The professional curriculum develops in students the knowledge, skills, attitudes, and values needed to provide pharmacist-delivered care. Pharmacy schools are also required to prepare graduates so that they can identify changes that are needed in pharmacy practice and the health care delivery system and successfully implement improvements.

The ACPE accreditation standards have established professional competencies that must be achieved by graduates and these competencies align with the Joint Commission on Community Pharmacy Practice (JCCP) Vision 2015 statement for the pharmacy profession. Specifically, graduates must be able to:

1. provide patient care in cooperation with patients, prescribers, and other members of the interprofessional health care team,
2. manage and use resources of the health care system, and
3. promote health improvement, wellness, and disease prevention in collaboration with patients and other interprofessional team members.

In addition, the professional curriculum must prepare graduates to be self-directed, lifelong learners so that they can individually achieve evolving practice competencies across the span of their career. The Center for the Advancement of Pharmacy Education (CAPE) within the American Association of Colleges of Pharmacy has established a set of outcomes that enable achievement of these core competencies. These learning outcomes also encompass core competencies that the Institute of Medicine has established for the health professions.

The professional curriculum consists of didactic coursework that provides the necessary scientific and clinical foundation and experiential coursework. In the first 2.5–3 years of most professional curricula, the didactic coursework is offered along with experiential coursework called Introductory Pharmacy Practice Experiences (IPPEs). The remaining 1–1.5 years of the curriculum consist of experiential coursework which are called Advanced Pharmacy Practice Experiences (APPEs).

The didactic coursework offered in a professional curriculum must meet the criteria of good science, which infers that the coursework is evidence based, logical, convincing, explanatory, honest, testable, and systematic. This foundational coursework is drawn from the biomedical sciences, pharmaceutical sciences, social/behavioral/administrative sciences, and the clinical sciences.

The teaching and learning methods used in a curriculum must be appropriate for the stated learning outcomes. In addition, they must promote the development of critical thinking and problem-solving skills and also transition the student from a dependent to an active, self-directed, lifelong learner. During the first 1–2 years of the curriculum, the typical teaching and learning methods are lectures and large group discussions. Faculty members are expected to use active learning strategies as much as possible and encourage students to ask questions. However, during years 2–3 of the curriculum, the focus is on case-based learning and small group discussions. Other learning methods typically used in these years include guided group discussions, problem-based learning, laboratory experiences, simulations, and other practice-based exercises. Computer and instructional technologies are being increasingly used to promote the development of critical thinking and problem-solving skills. The fourth year consists of the practice experiences called APPEs.

Some pharmacy schools are using synchronous or asynchronous distance-learning technologies to deliver the curriculum across several campuses. However, outcomes that are not appropriately achieved through distance-learning technology must be taught using other teaching and learning methods.

Special Topics and Emerging Issues

Preprofessional Coursework

Some pharmacy schools are moving to requiring 3–4 years of preprofessional coursework so that the professional curriculum can focus on more advanced science coursework and students enter the professional curriculum with a higher level of maturity. However, evidence is needed to determine whether a longer preprofessional education leads to enhanced educational outcomes.

Experiential Coursework

The doctor of pharmacy degree requires an increased depth and breath of experiential learning coursework compared to the previous baccalaureate in pharmacy degree. Although this curricular change was implemented in 2000, maintaining a quality experiential program continues to be a challenge due to a shortage of preceptors and sites and other resources. To promote the development of faculty, who oversee experiential learning

programs and to promote quality experiential education, AACP has established the Academic-Practice Partnership Initiative. In addition, AACP and the American Society of Health-System Pharmacists collaborated on a study to assess the current status and future capacity of hospitals and health systems as experiential learning sites. The findings of this study call for the following:

1. standardization and coordination of experiential programs among pharmacy schools,
2. better preparation of practitioner preceptors to be effective teachers, and
3. establishment of partnerships that provide mutual value to the sites and pharmacy schools.

Professionalism

Transition of the pharmacy profession from a product-centered to a patient-centered practice and the economic issues related to healthcare and drug therapy have prompted pharmacy educators to give more attention to ethics and professionalism. To promote pharmacy student professionalism, a task force that involved the collaboration of the American Pharmaceutical Association of Students of Pharmacy (APhA-ASP) and the American Association of Colleges of Pharmacy Council of Deans (AACP-COD) published a white paper and cited research that the professional socialization of students is often inconsistent and this hinders their professional growth. The task force also provided recommendations related to recruitment, admissions, educational programs, and practice experiences. It also emphasized the important role played by practicing pharmacists in the professional socialization of students. In addition, other approaches that address the cognitive–moral development of students have been advocated.

Outcomes Assessment

Although pharmacy schools have created positions so that individuals can lead outcomes assessment and some schools have developed innovative models, most schools are struggling to meet the outcomes requirements for accreditation. With respect to measurement of learning outcomes, pharmacy schools currently have data from graduate performance on the licensure examination and data measuring student and graduation perceptions. However, pharmacy schools need assessment tools that will more effectively measure achievement of the core competencies.

Residencies as a Requirement for Practice

Some in pharmacy practice and pharmacy education believe that a 1-year residency should be required following graduation from a doctor of pharmacy program so that not only graduates are more fully prepared to provide

direct patient care but also the profession can achieve the Vision 2015 statement. However, others believe that a residency requirement should not be used to address needs such as a higher quality experiential curriculum and enhanced assessment of student performance. One challenge is that there are an insufficient number of residency sites to place all doctor of pharmacy graduates.

Curricular Delivery Methods

Technology continues to rapidly expand opportunities for improving learning and assessment. Pharmacy schools are now using a variety of synchronous and asynchronous technologies to deliver instruction across several campuses. Some pharmacy schools are now exploring how to enhance problem-solving skills and critical thinking through use of simulations and virtual learning. Research is needed to evaluate the effectiveness of these new delivery methods.

Further Reading

- Abate, M. A., Stamatakis, M. K., and Haggett, R. R. (2003). Excellence in curriculum development and assessment. *American Journal of Pharmaceutical Education* **67**, Article 89.
- Accreditation Council for Pharmacy Education (2006). *Accreditation Standards and Guidelines for the Professional Program in Pharmacy Leading to the Doctor of Pharmacy Degree*. Chicago, IL: Accreditation Council for Pharmacy Education. <http://www.acpe-accredit.org> (accessed August 2009).
- American Association of Colleges of Pharmacy (2004). *Educational Outcomes*. Alexandria, VA: Center for the Advancement of Pharmacy Education Outcomes. <http://www.aacp.org/resources/education/Documents/CAPE2004.pdf> (accessed August 2009).
- American College of Clinical Pharmacy (2008). The definition of clinical pharmacy. *Pharmacotherapy* **28**, 816–817.
- American Pharmaceutical Association (1994). Code of ethics adopted. *American Pharmacist* **NS34**(11), 72.
- American Pharmaceutical Association (2008). *Medication Therapy Management Digest – Perspectives on MTM Service Implementation*. Washington, DC: American Pharmacists Association.
- American Pharmaceutical Association, Academy of Students of Pharmacy, American Association of Colleges of Pharmacy Council of Deans Task Force on Professionalism (2000). *White Paper on Pharmacy Student Professionalism*. *Journal of the American Pharmaceutical Association* **40**, 96–102.
- American Pharmacists Association and the National Association of Chain Drug Stores Foundation (2008). Medication therapy management in pharmacy practice: Core elements of an MTM service model (version 2.0). *Journal of the American Pharmacists Association* **48**, 341–353.
- American Society of Health-System Pharmacists (1987). Definitions of pharmacy residencies and fellowships. *American Journal of Hospital Pharmacy* **44**, 1142–1144.
- American Society of Health-System Pharmacists and American Association of Colleges of Pharmacy (2008). Capacity of hospitals to partner with academia to meet experiential education requirements for pharmacy students. *American Journal of Pharmaceutical Education* **72**, Article 117.
- Bradberry, J. C., Droege, M., Evans, R. L., et al. (2007). Curricula then and now – an environmental scan and recommendations since the commission to implement change in pharmaceutical education:

- Report of the 2006–7 academic affairs committee. *American Journal of Pharmaceutical Education* **71**, Article S10.
- Cipolle, R. J., Strand, L. M., and Morley, P. C. (1998). *Pharmaceutical Care Practice*. New York: McGraw-Hill.
- Clark, R. (1999). *A primer in diffusion of innovations theory*. <http://www.anu.edu.au/people/Roger.Clarke/SOS/InnDiff.html> (accessed August 2009).
- Duncan-Hewitt, W. (2005). The development of a professional: Reinterpretation of the professionalization problem from the perspective of cognitive-moral development. *American Journal of Pharmaceutical Education* **69**, Article 6.
- Greiner, A. C. and Knebel, E. (2003). *Health Professions Education: A Bridge to Quality*. Washington, DC: The National Academies Press.
- Hammer, D. P., Berger, B. A., Beardsley, R. S., and Easton, M. R. (2003). Student professionalism. *American Journal of Pharmaceutical Education* **67**, Article 96.
- Hunter, T. S., Deziel-Evans, L., and Marsh, W. A. (2003). Assuring excellence in distance pharmaceutical education. *American Journal of Pharmaceutical Education* **67**, Article 94.
- Kroboth, P., Crismon, L. M., Daniels, C., et al. (2007). Getting to solutions in interprofessional education. *Report of the 2006–7 Professional Affairs Committee*. *American Journal of Pharmaceutical Education* **71**, Article S19.
- Lee, M., Bennett, M., Chase, P., et al. (2004). Final report and recommendations of the 2002 AACP task force on the role of colleges and schools in residency training. *American Journal of Pharmaceutical Education* **68**, Article S2.
- McCarthy, R. L. (2004). Health care ethics. In McCarthy, R. L. and Schafermeyer, K. W. (eds.) *Introduction to Health Care Delivery. A Primer for Pharmacists*, 3rd edn., pp 199–224. Sudbury, MA: Jones and Bartlett.
- Miller, R. L. (1981). History of clinical pharmacy and clinical pharmacology. *Journal of Clinical Pharmacology* **21**, 195–197.
- National Association of Boards of Pharmacy (2006). *Foreign Pharmacy Graduate Examination Committee (FPGEC) Certification Program and Application Bulletin*. Mount Prospect, IL: National Association of Boards of Pharmacy. <http://www.nabp.net> (accessed August 2009).
- National Association of Boards of Pharmacy (2008). *NAPLEX/MJPE Registration Bulletin*. Mount Prospect, IL: National Association of Boards of Pharmacy. <http://www.nabp.net> (accessed August 2009).
- Roche, V. F., Nahata, M. C., Wells, B. G., et al. (2006). Roadmap to 2015: Preparing competent pharmacists and pharmacy faculty for the future. *Combined Report of the 2005–6 Argus Commission and the Academic Affairs, Professional Affairs, and Research and Graduate Affairs Committees*. *American Journal of Pharmaceutical Education* **70**, Article S5. <http://www.ajpe.org> (accessed August 2009).
- The Council on Credentialing in Pharmacy (2001). Credentialing in pharmacy. *American Journal of Health-System Pharmacists* **58**, 69–76.

Relevant Websites

- <http://www.acpe-accredit.org> – Accreditation Council for Pharmacy Education.
- <http://www.aacp.org> – American Association of Colleges of Pharmacy.
- <http://www.pharmacist.com> – American Pharmacists Association.
- <http://www.nabp.net> – National Association of Boards of Pharmacy.
- Foreign Pharmacy Graduate Equivalency Examination Study Guide. 2007.

Teaching for the Complex Practice of Nursing

P Benner, Carnegie Foundation for the Advancement of Teaching, Berkeley, CA, USA

M Sutphen, Carnegie Foundation for the Advancement of Teaching, San Francisco, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Clinical reasoning and skilled know-how – One of three professional apprenticeships required for a practice discipline. Clinical reasoning and skilled know-how are enacted in clinical situations. All of clinical reasoning is an example of practical reasoning and a broader form of rationality compared to a narrow technical rationality in professional education as illustrated by such strategies such as trying to apply theory to practice with little or no interpretation or translation. Rational calculation is a formal and narrow approach to practical situations and is as illustrated by strategies such as cost–benefit analysis; formal standardized plans such as critical pathways; decision analysis; and formal criterial or snap shot reasoning, as used in scientific experiments.

Ethical comportment – The everyday ethical practice within a practice discipline. It includes actions, communication, respect, and fulfillment of the notions of good internal to the practice discipline.

Integration and integrative teaching – Teaching can be integrative in multiple ways but we focus on integrating the three professional apprenticeships, and integrating classroom and clinical teaching and learning. However, teaching for a sense of salience and understanding the nature of a particular clinical situation are also integrative, as is articulating and reflecting on experiential learning and situated coaching.

Situated cognition – The ability of a practitioner to stand inside a situation and reason about recent trends and trajectories in the patient and the situation. Clinicians take into consideration the context of the situation and take into consideration changing relevance over time.

Three professional apprenticeships of professional education – (1) The cognitive apprenticeship includes learning to think like a professional practitioner, and includes learning all the requisite science and theoretical knowledge in order to practice well in a professional discipline; (2) the skilled know-how and clinical reasoning apprenticeship is required to use knowledge in

practice and this includes situated cognition in particular unstructured practice situations; and (3) the ethical comportment and formation apprenticeship includes learning well, the first two apprenticeships, and enacting or instantiating the notions of good within the practice discipline.

Introduction

As do other professionals, nurses learn to use practical reasoning through experiential learning. They must learn clinical reasoning across time through transitions in the patient's clinical condition or concerns and/or through transitions in the clinician's understanding of the situation. Modern medicine depends upon the nurse's monitoring the patient's condition and adjusting therapies to the patient's physiological responses across time. This calls for situated teaching and learning across time, and the ability to recognize relevant changes in the patient's physiological responses.

As a practice discipline nursing education requires teaching and learning in three domains, or apprenticeships. One apprenticeship requires teaching and learning the knowledge base for nursing or how to think like a nurse. In another apprenticeship, nursing students learn skilled know-how and clinical reasoning that they will use in practical situations. In still another apprenticeship, students learn ethical comportment in the particular practice discipline. It is in this last apprenticeship that students undergo formation, or where they learn to see as and act in the capacity of, to be a practitioner in their profession. These apprenticeships are best taught as an integrated whole; once they are separated, students find it difficult to bring them back together. For example, teachers with their doctoral training in physiology may accept their domain-specific assumption of the discipline of physiology that pathophysiology should always be taught separately from normal physiology, but for nursing students, a focus on both in the same course might be more advantageous because nurses must recognize deviations from the normal and how to support normal physiology. While separating the two is probably useful

for teaching physiologists, it is not necessarily true for nursing where using physiological knowledge in actual situations is crucial.

Learning and performance requirements must be established for each discipline, as well as the best ways of integrating teaching in and out of classrooms and all three professional apprenticeships described above.

Nursing stands out in professional-educational programs because it has multiple educational pathways to practice. Nurses may enter through 2–3-year diploma programs, though there are only about 80 of these programs in existence in the US. They may also enter through community college programs where they must take at least 1 year of prerequisites for the nursing program and then at least 2 years of nursing. Due to oversubscription that requires waiting to get space in these programs, a community college education in nursing takes a minimum of 3 years and often as much as 5 years. The profession's preferred programs are the generic baccalaureate programs, and the second baccalaureate degree programs or the master's entry program for students with a baccalaureate degree from another field. All prelicensure student nurses are required to take the National Council Licensure Examination-Registered Nurse (NCLEX-RN) licensing exam. There is an increasing shortage and demand for nurses with a master's degree for advanced practice, and a doctoral degree for teaching. Nursing also has doctorate in nursing practice, which is an advanced doctoral degree that focuses on nursing practices. However, this article focuses on the education for the initial preparation as a nurse.

Domain and Practice-Specific Teaching and Learning: Nursing

Recent scholarship in nursing education has highlighted domain-specific teaching and the need to better understand the discipline's commonly used pedagogies. Nursing relies heavily on situated instruction and coaching in actual clinical practice situations. As the basis for instruction, faculty use situations students encounter, helping them interpret in new ways what they experience and act in new ways. They also rely on formative education where the clinician's perceptual and thinking capacities as well as skilled know-how are transformed by the demands and notions of good in the practice. Notions of good in the practice for nursing are the goals and aspirations generally agreed upon by nurses and include an ethos that emphasizes advocating for patients and their families. Nursing's signature pedagogy is coaching for situated understanding and action. As do clergy, nurses must develop their clinical imagination because they enter into unstructured, open-ended situations that require them to understand the situation and take action.

Formation in nursing requires practitioners to adopt role expectations as well as learn skilled know-how, new perceptual capacities, and situated thinking in action. The practicing professional literally sees new possibilities and concerns in practical situations in ways that never solicited their attention or action before. A nurse develops over time through experiential learning his perceptive ability to notice when things are not quite right or are not going according to usual expectations. A nurse develops also a sense of salience about what things in a situation must be addressed now and what can wait. This sense of salience in underdetermined practical situations is also experientially learned over time, and depends upon the teacher helping the student grasp and define the nature of clinical situations through coaching in the early phases of learning, and questioning during more advanced stages of learning. More advanced students have enough experiential learning to develop a beginning sense of salience and recognize recurring signs and symptoms and common meanings and concerns in practical situations. In learning, a practice the practitioner is formed to enact skilled know-how and their ethical comportment based upon their understanding of the situations and its demands for action. This, like all experiential learning, requires domain- and practice-specific teaching and learning.

Students learn the practice of nursing in the three apprenticeships of knowledge, clinical reasoning, and formation. They require the following:

1. Students must learn a sense of salience, situated cognition, and action in particular clinical situations.
2. Students must learn to use clinical reasoning and to think in multiple ways that include critical thinking.
3. Students must integrate what is presented in many different settings, including classroom, clinical, and skills or simulation laboratories.
4. Students must learn everyday ethical comportment through professional formation.

Teaching for a Sense of Salience, Situated Cognition, and Action

Nurse educators and students tend to valorize a narrow form of reasoning, where they try to teach students to stand outside of a given situation to take a snapshot view of a situation. However, nurses work in complex relatively unstructured clinical situations where they must learn to quickly recognize and assess what are most and less important in particular clinical situations. Students need to develop a sense of salience about what is relatively important and unimportant in any particular clinical situation. The heart of practical reasoning is understanding the nature of the situation. Faculty must help students to

recognize what is most urgent, most important, and less urgent in particular practice situations that are under-determined and change over time. Practitioners must first grasp the nature of the situation before they can act intelligently and prudently. For the first-year nursing student (the novice), the instructor must coach him to recognize what must be addressed during his care of a particular patient/family. The nurse educator must teach and coach students to develop their own situation-recognition capacities. It is not possible in actual complex practice situations for the student to build up a holistic grasp of the situation element by element. The most reduction and simplification that can be arranged for the student nurse is for the instructor to assign patients who are not in crisis, and who require relatively straightforward and simple interventions. Continued situated coaching is required in order for the student to grasp the changing relevance, and demands, resources, and constraints in particular situations. Eraut has called this a productive form of knowledge use and Lave and Wenger called it situated cognition.

In their classroom teaching, teachers often use models of how students should think about problems, such as models of critical thinking. They describe these as frameworks of thinking, such as decontextualized and mentalistic representations of critical thinking or the nursing problem-solving process. A framework of thinking though is not the same as thinking. This problem is not unique to nursing in academia. The critical-thinking agenda in college seeks to help students disengage from context bound, concrete knowledge to what is considered a higher form of knowledge that is made explicit, formal, operational, and more abstract and general. In doing so, educational programs may overlook tacit and embodied knowledge that is part of skillful judgment.

We advocate a robust respect for practical reasoning and approaches that teach students the habits of thought they need to exercise situated clinical judgments. Astute, engaged practical clinical reasoning requires well-educated graduates who take up a broader view of rationality, one that takes into account the particulars of a situation and values reasoning through transitions across time.

Teaching Clinical Reasoning and Thinking in Multiple Ways

Most academic settings emphasize the importance of critical thinking; however, in professional practices, critical thinking is bounded by practical reasoning and performance is based upon the best-available scientific evidence in the case of nursing. Critical thinking is best reserved for situations where the practitioner needs to reflect upon a received tradition or practice that is no longer relevant, or needs to be improved. It is also needed in situations of

extreme breakdown in understanding and actions, so that deliberative engaged reasoning no longer makes sense. Critical thinking is an essential, but not a sufficient thinking strategy for the practicing professional. Practitioners also need to be able to imagine future eventualities and to perform according to the best evidence available in under-determined open-ended situations.

Currently, critical thinking has become a catch-all phrase for all forms of thinking required in nursing practice. But this use of critical thinking is actually a misnomer that needs clarification so that nurse educators focus on multiple ways of thinking, with a much greater emphasis on clinical reasoning.

Thinking like a nurse requires clinical reasoning as well as critical, creative, scientific, and formal critical reasoning. Cynicism and excessive doubt, often the byproduct of overuse of reflective critical thinking, will not help the professional nurse or physician draw on appropriate knowledge and act in particular situations. Nor does critical thinking alone develop students' perceptual acuity or clinical imagination about using science, skilled know-how, and practical knowledge in particular situations. Clinical imagination is required for students to grasp the nature of patients' needs as they change over time. Likewise, narrative understanding and interpretation of clinical situations help to enrich the student's clinical imagination and reasoning about changes in the patient's condition over time. The need for teaching critical reflection, analysis, and thinking are essential to any education; however, they are best taught when teachers and students understand their goals and functions in relation to practice.

Integrative Teaching

Most schools of nursing radically separate classroom and clinical teaching. It is as if teachers assume that students acquire knowledge in the classroom and apply it to clinical practice, rather than draw on appropriate knowledge and interpret it for a given situation. A formal separation of clinical and classroom teaching makes learning unwieldy and uncoordinated. To make matters worse, many teachers present information about nursing theory, physiology, disease categories, signs, symptoms, interventions, and outcomes as layered taxonomies to be memorized. They try to force too much information into their lectures and often worry that they present is at too superficial a level. Many teachers presented theories and clinical knowledge as taxonomic naming systems are the most abstract and difficult for students to imagine how to use the knowledge in caring for patients. Few students can imagine how the classification systems can be used in their actual direct patient care. No one abstract theory can be directly applied, except the most technical physiological theories such as measuring blood gasses, or circulatory pressures and dynamics.

Ubiquitous is the use of elemental competencies, by which we mean when teachers expect students to perform, demonstrate, or present a given skill that is stripped of clinical context. The most common example of this approach occurs in skills laboratories. For example, it is useful to teach first-year nursing students how to take vital signs outside of a clinical situation. But the pedagogy of teaching from simple stripped-down examples to use in complex situations poses inherent limitations in any practice discipline. In order for a student to make a clinical assessment about, for example, a hypertensive patient in labor and delivery, the student very quickly needs to move on to interpreting information from the patient's vital signs and taking account of the particulars in the situation. Students must integrate what they learn from different settings – clinical, skills or simulation laboratory, and classroom – as a seamless whole.

Professional Formation

Nursing education engages theories about socialization, roles and role taking, role making, and role performance. While such ideas are useful in any practice discipline that must coordinate and link role functions with other disciplines and team members, we suggest a more robust view of changes in identity and self-understanding of the practitioner. Formation of student nurses occurs within their skilled practices and highly relational work that literally transform their ways of perceiving and acting in situations.

Many nursing students describe profound changes in their sense of identity and how the experience of skilled practice and highly relational work transforms the way they perceive and act. Students often identify particular learning experiences as affecting their capacity to act in future similar situations, increasing their self-knowledge and understanding, or reassuring them of their chosen profession. Student nurses talk readily about the interpersonal challenges they face in nursing.

Nursing students learn the skills of involvement – or the skills they will need to be involved with patients in appropriate and therapeutic ways – in particular assignments, usually in situations where patient well-being is at stake. Even before the student begins to develop the professional identity of nurse, he may find that the patient and family treat him as a nurse and automatically expect him to respond like a nurse. The following account from our own work illustrates how one student talked about learning nursing as a practice in this way:

Jennifer Student Nurse: I had this experience on our first day in the Mother/Baby, Labor and Delivery course. My professor said, 'Okay, *Jennifer*, you're going to have two couplets today. You're going to have four patients,' and,

back then, that was a big deal. So I was running around, head down. I did the assessments and I was really getting nervous because it was getting late and I still hadn't assessed the last baby yet. So I walked into the nursery and there was a nurse there who said: 'Oh, don't worry about that baby. I just assessed her and everything's fine.' And I remember standing there thinking, 'oh, good.' But then I said, 'I can't do that,' you know, 'I can't do that,' *and you can't do that*. . . . I opened up the blanket and looked in and sure enough, the baby was tachycardic, flaccid, and cold. That's the day I realized the awesome responsibility to do our job because that baby was then placed in the incubator and Intensive Care and was there for at least a week. And I'm thinking how many hours would it have been 'til somebody had actually, you know, discovered the baby's condition— Had I just listened to the staff nurse, as a student, and said 'okay, well the RN said she assessed the baby now, and she had documented her assessment.' But just because it was my patient assignment, as a student, I still had to go and do it [the assessment]. It was a good feeling to realize hey, I really can actually assess an infant and find something wrong. And I think that's the day I realized that you never—, you are responsible. I don't care if somebody else has that assignment. If you have anything to do with that patient you're responsible. . . . It was a day I realized that *even that early on* you're very important. . . . I still remember that day. So every day I go in to do assessment I think, nope, I'm doing everything. I don't care, head to toe, you know. We're not skipping over anything. So that was an important day. . . . They give you the tools and you really just have to go in there and do it and even though there are time constraints and there are always time constraints as a student, you know, it's worth getting in trouble for being behind rather than to put your patients' care in jeopardy. So that was an important lesson.

Jennifer's emphasis and tone of voice in telling the story point to the significance it holds for her as a student. One might imagine an expert nurse dismissing the incident in one sentence by invoking a rule about the necessity of frequently checking the infant's condition herself. The tone and emphasis of the seasoned nurse's voice would likely reflect the solidity of a well-formed practice assumption.

The clinical experiential learning Jennifer describes is formative for her. She locates the dawning of a new awareness and new agency within her choices and actions involving a particular infant, complete with all the attendant pressures and contingencies. No nursing faculty member intentionally plans such a risky learning experience. Even so, similar situations will likely repeat in different settings and with different patients in most students' clinical experience. Jennifer is awed by the infant's fragility and dependence on her vigilance and skill, and by the

responsibility this carries. She does not feel prepared to handle the situation, but she is pleased to find that she can do the right thing and make a significant contribution to an infant's life. In this situation, in response to what is demanded of a nurse, Jennifer moves from acting like a nurse to being a responsible nurse.

Jennifer identifies this experience as pivotal for her. It forms part of her habitus, in that it creates a way to act, a way of being in the world, and a new set of perceptual capacities associated with ways to respond both to the routines a nurse is expected to enact (always assess your own patients) and to unexpected clinical concerns (finding the baby flaccid and cold). The stakes are high. The infant's life is in her hands. Although the staff nurse may have been correct in her earlier assessment of the infant, infants can change rapidly. The import of the event is that Jennifer has learned, first hand, that she must take responsibility for her patients, and that she now knows that she can notice significant life-saving changes in an infant.

This experiential learning in the clinical setting integrates all three professional apprenticeships for Jennifer. In the course of her narrative, the student illustrates how she draws on her learning in the first and second apprenticeships. She notes the baby was tachycardic – knowledge gained from physiology – and flaccid – a clinical judgment learned in the physical-assessment skills laboratory. She also grasps immediately that it was imperative to act quickly, with her focus being on doing all that she could for the baby, thus illustrating her ethical comportment in the situation. Many such experiences shape the character and skill formation of a nurse over time.

Like Jennifer, many other students' narratives depicted their sense of responsibility being formed by confronting challenges where their own actions impacted the patient's survival.

Transformation and formation terms that borrowed from Carnegie's study of clergy education address the development of the senses, esthetics, perceptual acuities, relational skills, knowledge, and dispositions that takes place as student nurses form professional identity. A shift to a student-focused sense of transformation and reflection on formative experiences in nursing school enriches the student's sense of identity and self-understanding as a nurse.

Further Reading

- Bain, K. (2004). *What the Best College Teachers Do*. Cambridge, MA: Harvard University Press.
- Benner, P., Hooper-Kyriakidis, P., and Stannard, D. (1999). *Clinical Wisdom and Interventions in Critical Care, a Thinking-in-Action Approach*. Philadelphia, PA: Saunders.
- Benner, P. and Sutphen, M. (2007). Learning across the professions: The clergy, a case in point. *Journal of Nursing Education* **46**, 103–108.

- Benner, P., Sutphen, M., Leonard, V., and Day, L. (2008). Formation and everyday ethical comportment. *American Journal of Critical Care* **17**, 473–476.
- Benner, P., Tanner, C., and Chesla, C. (1996). *Expertise in Nursing Practice: Caring, Clinical Judgment and Ethics*. New York: Springer.
- Bourdieu, P. (1990). *The Logic of Practice*. Stanford, CA: Stanford University Press.
- Cronenwett, L., Sherwood, G., Barnsteiner, J., et al. (2007). Quality and safety education for nurses. *Nursing Outlook* **55**, 122–131.
- Dewey, J. (1933). *How We Think – A Restatement of the Relation of Reflective Thinking to the Educative Process*. Boston, MA: Heath.
- Dewey, J. (1960). *Theory of the Moral Life*. New York: Holt, Rinehart and Winston.
- Dewey, J. (1973). Experience and thinking. In J. J. McDermott, J. J. (ed.) *The Philosophy of John Dewey*, (Original work published 1916) vol. 2, pp 494–506. New York: Putnam's.
- Dreyfus, H. and Dreyfus, S. E. (1986). *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.
- Dunne, J. (1993). *Back to the Rough Ground: Practical Judgment and the Lure of Technique*. Notre Dame: University of Notre Dame Press.
- Eraut, M. (1994). *Developing Professional Knowledge and Competence*. London: Washington, DC: Falmer Press.
- Foster, C. R., Dahill, L. E., Golemon, L. A., and Tolentino, B. W. (2005). *Educating Clergy: Teaching Practices and Pastoral Imagination*. San Francisco, CA: Jossey-Bass.
- Grossman, P. and McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal* **45**, 184–205.
- Huber, M. T. and Hutchings, P. (2005). *The Advancement of Learning: Building the Teaching Commons*. San Francisco, CA: Jossey-Bass.
- Hurwitz, B. (2000). Narrative and the practice of medicine. *Lancet* **356**(9247), 2086–2089.
- Hurwitz, B. (2002). Erring and learning in clinical practice. *British Journal of General Practice* **52**(30), S26–S30.
- Ironsides, P. M. (1999). Thinking in nursing education part I: A student's experience learning to think. *Nursing and Health Care Perspectives* **20**, 238–242.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation (Learning in Doing: Social, Cognitive and Computational Perspectives)*. Cambridge: Cambridge University Press.
- Mohrmann, M. E. (2006). On being true to form. In Taylor, C. and Dell'Oro, R. (eds.) *Health and Human Flourishing, Religion, Medicine, and Moral Anthropology*, pp 9–102. Washington, DC: Georgetown University Press.
- Montgomery, K. (2005). *How Doctors Think: Clinical Judgment and the Practice of Medicine*. New York: Oxford University Press.
- Schon, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.
- Schon, D. A. (1987). *Educating the Reflective Practitioner*. San Francisco, CA: Jossey-Bass.
- Shulman, L. S. and Wilson, S. M. (2004). *The Wisdom of Practice: Essays on Teaching, Learning, and Learning to Teach*. San Francisco, CA: Jossey-Bass.
- Sullivan, W. M. (2004). *Work and Integrity: The Crisis and Promise of Professionalism in America*, 2nd edn. San Francisco, CA: Jossey-Bass.
- Sullivan, W. M., Colby, A., Wenger, J., Bond, L., and Shulman, L. (2007). *Educating Lawyers: Preparation for the Profession of Law*. San Francisco, CA: Jossey-Bass.
- Sullivan, W. M. and Rosin, M. S. (2008). *A New Agenda for Higher Education: Shaping a Life of the Mind for Practice*. San Francisco, CA: Jossey-Bass.
- Tanner, C. A. (1994). On clinical teaching. *Journal of Nursing Education* **33**, 387–388.
- Tanner, C. A. (1997). Teaching is a practice. *Journal of Nursing Education* **36**, 306–307.
- Taylor, C. (1993). Explanation and practical reason. In Nussbaum, M. and Sen, A. (eds.) *The Quality of Life*, pp 208–231. Oxford: Clarendon Press.

Van Maren, M. (1991). *The Tact of Teaching: The Meaning of Pedagogical Thoughtfulness*. Albany, NY: SUNY Press.

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. New York: Cambridge University Press.

<http://www.nln.org> – National League for Nursing.

<http://www.nsna.org> – National Student Nurses' Association.

<http://www.bls.gov> – United States Department of Labor.

Relevant Websites

<http://www.aacn.nche.edu> – American Association of Colleges of Nursing.

Veterinary Education

K M Edmondson, Cornell University, Ithaca, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Theriogenology – The branch of veterinary medicine concerned with the physiology and pathology of reproduction in animals, and with the clinical practice of obstetrics and gynecology.

Xenotransplantation – The transplantation of living cells, tissues, or organs from one species to another, such as from pigs to humans.

Introduction and Context

Veterinary medicine is the medical profession at the interface between animals and people. Although a majority of veterinarians provide medical care for pets and food animals, professional activities of veterinarians also include preventing the transmission of zoonotic diseases from animals to humans, wildlife preservation and conservation, and public health. The broad, comparative training veterinarians receive in the biological sciences prepares them to care for a wide range of animal species, and uniquely positions them to serve as valuable team members in the promotion of human health and protection of the food supply.

Within the global community, public awareness of the interrelatedness of animal and human health is increasing, as illustrated by emerging programs, such as the One Health Initiative, which recognize the synergy and potential benefits of promoting collaboration among physicians, veterinarians, and other scientific health professionals to develop global strategies to advance healthcare. Changing trends in pet ownership, the increasing value placed upon the bond between humans and animals, and the role of that bond in promoting human health, including mental health and well being, are also significant.

Veterinary medicine as a profession is in a state of transition. Not only has a significant gender shift occurred, but there is a trend toward increased specialization as well. According to the American Veterinary Medical Association (AVMA), 75% of today's veterinary students are female (in contrast to 55% in 1988), and more than one-third of all veterinary graduates seek advanced training in a wide range of specialty areas (more than twice the number seeking additional training in 1988). Consequently, faculty at veterinary colleges must consider how

to balance the need to train generalists with more specialized knowledge, and how to best prepare future professionals for an ever-changing world. As scientific knowledge increases, and the important roles veterinarians play in service to society become more visible, the expectations of new veterinary graduates also expand. The high costs associated with veterinary education, combined with significant student debt, heighten the challenges veterinary colleges face. This article presents the current status of veterinary education, discusses inherent themes that are common among veterinary colleges, and outlines future challenges.

Requirements, Accreditation, and Licensure

In the United States, a typical veterinary curriculum is 4 years in length and leads to the degree of doctor of veterinary medicine, or DVM. In the United Kingdom, Europe, and Australasia, students enter veterinary college following secondary school. Programs of study are usually 5 years in length, and lead to the degree of bachelor of veterinary science (BVSc).

Students in the United States and Canada have typically completed a baccalaureate degree prior to entering veterinary school. Some have completed advanced degrees. The academic requirements that applicants must fulfill are very similar to those required by schools of allopathic medicine or dentistry. In addition to substantial prerequisite coursework in the biological sciences and scores from standardized tests, prospective students are required to provide evidence of veterinary-related (and animal) experience, submit a statement of intent, and provide information related to their other activities and interests.

The number of colleges of veterinary medicine is relatively small (28 in the United States and five in Canada), and the average number of students per class is just under 100. Veterinary curricula follow the traditional Flexner model of medical education, with emphasis on basic science disciplines in the first 2 years, followed by instruction in clinical disciplines in the second half of the curriculum, including a year of practical clinical training. The majority of preclinical instruction is didactic, with supporting laboratory exercises. Clinical instruction follows the apprenticeship model of education: usually within the context of a teaching hospital, students rotate through a series of clinical services, supervised by clinicians, and are directly involved in medical case management. Preclinical disciplines common to veterinary

curricula include gross anatomy, histology, molecular genetics, physiology, pathology, microbiology, bacteriology, virology, parasitology, epidemiology, pharmacology, and embryology. Clinical disciplines include medicine and surgery, anesthesiology, imaging, population medicine, behavior, nutrition, and other specialty areas such as oncology, ophthalmology, neurology, cardiology, toxicology, dentistry, laboratory animal medicine, theriogenology, emergency and critical care medicine, and exotic animal medicine. Most schools in the US educate generalist veterinarians, but some allow students to focus on particular species.

Veterinary students, therefore, are expected to master a significant amount of material that includes a strong foundation of basic science and clinical knowledge, clinical skills, and professional behaviors. Their knowledge and skills are assessed through a range of approaches that include traditional written examinations (with essay, short answer, or multiple-choice questions), oral and practical exams, demonstrations, and, occasionally, student presentations. As there is no requirement for additional training, new veterinary graduates can be licensed to practice medicine following graduation. Skills labs are incorporated into the educational program with the expectation that students will become competent in performing many standardized procedures prior to graduation. Those who wish to pursue additional training may choose internships and/or residency training in one of the disciplines listed above. Internships are usually 1 year in length. Residency training lasts 2–3 years, and qualifies the individual to sit for a Board Certification examination in that discipline.

Since 1953, the Council on Education (COE) of the American Veterinary Medical Association has been the accrediting body for veterinary colleges in the United States and Canada. Although the COE makes all accreditation decisions, it is accountable through the AVMA to a governmental regulating body (the US Department of Education) and to a nongovernmental agency (the Commission for Higher Education Accreditation). In recent years, several veterinary colleges from other countries have sought and received accreditation from the AVMA. Currently, in addition to those in the United States and Canada, there are AVMA-accredited colleges of veterinary medicine in the United Kingdom, Australia and New Zealand, and in the Netherlands. To receive accreditation from the AVMA, each college must satisfy standards in 11 areas, and demonstrate the ability to provide students with the necessary resources for success within a high-quality program that meets the designated standards.

Once accredited, colleges receive full reviews every 7 years. Veterinary colleges in the United Kingdom are accredited by the Royal College of Veterinary Surgeons (RCVS), which conducts periodic visits every 7–10 years.

Other countries have their own systems for accrediting veterinary colleges. Those that are members of the European Union (EU) have considered a move toward

consistency or uniformity of their standards and curricula. The European Association of Establishments for Veterinary Education (EAEVE) and the Advisory Committee on Veterinary Training (ACVT) have developed a system for evaluating veterinary schools that applies to institutions within the EU. Beyond this, and based in part upon an effort to facilitate the globalization of animal disease control, a number of schools are supportive of recent initiatives to increase compatibility among veterinary education programs in different parts of the world. This is a complex undertaking, however, and is complicated by differences in organization, governance, financial structure, curricula, and overall costs.

A license to practice veterinary medicine signifies that the holder is competent to treat all animal species. For graduates of AVMA-accredited institutions, the North American Veterinary Licensing Exam, NAVLE, is required for licensure in all of the licensing jurisdictions in North America. In addition, some states require a State Board Examination. To maintain certification, some states require practicing veterinarians to participate in continuing education activities, while others do not. Graduates of non-accredited schools can seek licenses in the United States through two certification programs: the Educational Commission for Foreign Veterinary Graduates (ECFVG) which is operated by the AVMA, and the Program for the Assessment of Veterinary Education Equivalence (PAVE), which is operated by the American Association of Veterinary State Boards. Canada has its own certification requirements for graduates of international veterinary colleges, administered through the National Examining Board.

New veterinary graduates are expected to function independently, and to possess a level of competence in a wide range of clinical domains. They are expected to demonstrate attributes such as ethical behavior, compassion, respect, honesty, and responsibility, as well as knowledge of animal disease and wellness, diagnosis, management and treatment, and business practice. They are also expected to reason deductively, exhibit problem-solving skills, communicate effectively with others, and perform core clinical procedures (including standard surgical procedures). Graduates should have a continual thirst for new knowledge, the ability to critically examine new knowledge, and the ability to recognize the limits of their knowledge and skill.

Throughout their careers, veterinarians must adhere to a code of professional ethics. For veterinarians who belong to the AVMA, these are outlined in the Principles of Veterinary Medical Ethics. Local and state veterinary associations and licensing boards have additional rules governing professional conduct. The principles lay out expectations in the following categories: professional behavior; the relationship between the veterinarian, client, and patient; attending, consulting, and referring; influences on judgment; therapies; genetic defects; medical records; fees and remuneration; advertising; and euthanasia. The Veterinarian's Oath

comprises these elements, and includes additional language related to conservation of animal resources, promotion of public health, and the advancement of medical knowledge. In general, veterinarians are expected to relieve disease, suffering, or disability, minimize pain and fear, and to be honest, law abiding, and fair. Veterinarians should not engage in fraud, misrepresentation or deceit, and they should strive to enhance their image with respect to the public. They must be committed to improving their veterinary knowledge and skills, and be courteous, considerate, and compassionate.

Impetus for Change in Veterinary Education

In the past two decades, three major reports published within the veterinary profession contained recommendations for veterinary education. The earliest report, part of a national trend of curriculum reform efforts in medical education, was the Pew Report published in 1989 calling for sweeping changes for veterinary schools that redirected the emphasis on accumulating factual information toward a focus on how to find and use information, solve problems, and acquire essential skills and behaviors. It encouraged colleges of veterinary medicine to identify mechanisms for increasing students' responsibility for their education, for strengthening curriculum content in the basic biological sciences, and for providing students with the opportunity for in-depth clinical instruction. To better meet the needs of the public with regard to food safety and public health, the Pew Report also called for the inclusion of additional curriculum content. It encouraged veterinary colleges to make research activities a higher priority, and to place more emphasis on post-DVM education programs. Colleges of veterinary medicine were also directed to decrease enrollments, and to focus their programs on a smaller number of specialty areas, creating centers of excellence. The Report projected changes in the organization of veterinary education, particularly as it related to clinical experiences, and predicted that veterinary practitioners would contribute more to educational programs, and to the direct instruction of veterinary students.

In response, many veterinary colleges in the United States and Canada engaged in processes of curriculum review, and implemented a number of the Pew recommendations in the 1990s. Some schools experimented with and implemented different approaches to teaching. Following the lead of many medical schools, several veterinary colleges considered problem-based learning as an instructional approach. In an attempt to better integrate the basic science and clinical disciplines, to foster the development of clinical problem-solving skills, and to engage students in the learning process, many schools incorporated some aspects of this approach into existing

courses, or made structural changes to their curriculum to include parallel or capstone problem-based courses. A small number of schools implemented entirely new curricula that used problem-based learning as the predominant instructional approach in the preclinical years. Still others restructured their curricula and incorporated other forms of small group learning.

A common form of educational innovation among many schools was an increased reliance on various forms of computer-assisted instruction. Schools developed a wide range of educational resources during this period that supplemented or replaced material that had previously been taught didactically, and reduced the number of live animals there were used for teaching. Some schools developed tracking curricula that allowed students to focus on particular species or aspects of professional practice. Others changed approaches to clinical instruction to involve practitioners in innovative ways.

Ten years after the Pew Report, the influential KPMG Study was published and focused primarily on the future job market for veterinarians and veterinary services. Contrary to the Pew Report, the KPMG Study projected a shortage of veterinarians in some areas, calling for colleges of veterinary medicine to expand curricular offerings relative to the professional knowledge, skills, and behaviors that contribute to economic success in veterinary practice. These included communication skills, technical skills, information related to business and practice management, critical thinking, and lifelong learning skills. While primarily focused on the veterinary profession and not on education (only one of the six critical issues identified in the report related to education), the implications of this report for veterinary colleges were clear, and it served as the impetus for additional curricular changes.

In the time that has elapsed since the publication of the KPMG Study in 1999, a number of veterinary colleges in the United States, Canada, and around the world have taken significant steps to develop courses and curriculum materials focusing on client relationships and communication skills. In many instances, these initiatives have also led to a new consideration of clinical competency and outcomes assessment.

The third and most recent major report to impact veterinary education is the Foresight Report, published in 2007. The goal of the report was to determine a vision for the future of academic veterinary medicine, and to provide a basis for strategic planning and curriculum development. The Report contained a series of recommendations that included acceptance of the ideas that veterinary licensure in the future would be limited to selected focus areas, that colleges of veterinary medicine should collaborate, and that they should identify areas of focus or program emphasis. These recommendations were based upon the notion that the needs of society were increasing, and that veterinary colleges could only respond

to the changing needs of society through expansion. By encouraging colleges to work together, the recommendation to create centers of emphasis would allow veterinary medical education at large to respond, without requiring each college to expand its programs. With this, veterinary colleges were also asked to increase enrollments, so that the profession would be better able to respond to new demands and roles. In turn, accreditation standards would be modified, and limited to the requirements to teach a core program plus the areas of professional focus offered by an individual college.

The Report also encouraged colleges to capitalize on emerging technology to provide distance education and web-based instruction, and to develop opportunities for virtual training through the use of simulators. These would be used to train students, and for continuing education programs for practicing veterinarians wishing to further develop their skills, or to retrain for other areas within the profession. The Report also directed veterinary colleges to address the problem of student debt, particularly in segments of the profession or regions of the country that have the greatest need for veterinary care.

The Foresight Report generated lively discussion among veterinary educators, and the impact of its recommendations has not yet been fully realized. However, a key point in the Report focused on a national strategic plan for veterinary education that would be uniformly implemented, and it outlined possible models. Whether or not this is achievable in the coming decades remains to be seen.

Themes in Veterinary Education

Within the broad framework described above, a number of themes span veterinary education as a whole, and are also common to other professions. Perhaps the most familiar themes are the tension between breadth and depth, the need to integrate theoretical knowledge with practice, and finding ways to help students become comfortable with ambiguity in solving novel, ill-structured problems. All of these contribute to the development of professional expertise, and present challenges for teaching and learning. Many veterinary educators recognize the importance of integrating theory with practice, but few curricular alternatives have been offered for improving upon (or moving beyond) the apprenticeship model of educating. Educators have long incorporated various forms of real-world experience in their teaching to present new information, underscore the practical application of curricular material, and to enhance and extend student learning. For students at the university level, apprenticeships, field experiences, internships, and practicum experiences represent examples of the types of practical experiences many encounter as extensions (or even requirements) of their

formal education. However, the Flexner model for medical education (formal theoretical training followed by a period of supervised practical application) dominates veterinary education, at the expense of alternatives that may be effective for achieving these goals.

The research on meaningful learning supports a view of the development of expertise in veterinary medicine that helps students make meaningful and progressively elaborate connections between the knowledge they gain in a classroom and that acquired through field and clinical experience. The need to develop elaborate, valid propositional knowledge structures that provide a scaffold for subsequent knowledge gained through experience is consistent with research on learning in other domains, and with the Flexner model for medical education: acquisition of basic concepts, followed by practical application. While the characteristics of novices and experts (and their respective abilities) are now more clearly understood, models of medical expertise have lacked an account of the processes by which novices become experts. However, three factors are key:

1. the information to be learned must be appropriately structured,
2. the learning context must be clearly established, and
3. the learner needs to be actively engaged in the learning process.

The structure of most veterinary curricula inhibits, rather than facilitates, meaningful learning and the integration of theoretical with practical knowledge. The disciplinary boundaries imposed in most traditional curricular structures encourage students to compartmentalize their learning, and most traditional examination formats focus on the retention of factual information rather than on more complex cognitive processes. This problem is exacerbated by exhaustive content coverage, and the need for veterinary curricula to address comparative aspects of many topics, such as important species differences. Increased specialization in the clinical domain has also led to the addition of significant amounts of content, which are now identified as core and required of all students. Topics such as oncology, dentistry, and emergency and critical care medicine are relatively new additions to the veterinary professional curriculum, and require time within course syllabi that are already very full. As mentioned previously, the number of veterinary graduates who are electing to pursue additional postgraduate training through internships and residencies is also increasing. A shift away from the education of generalist veterinarians toward a more specialized model could provide increased flexibility in curricular structure that would also bring important benefits to student learning.

Another theme in veterinary education, which is common to medical educators, is the development and assessment of clinical competency. New accreditation

standards relating to the assessment of clinical competence have increased accountability for the acquisition of a broad range of skills, and in recent years, veterinary colleges have implemented initiatives to document the development of students' technical skills. The focus on clinical competency (which in its extreme form might be interpreted as promoting the view of veterinarian as technician) is in tension with the view of the veterinarian as scientist, and raises questions within the colleges about the appropriate balance between theoretical and practical knowledge. To meet accreditation requirements and continue to deliver curricula of sufficient breadth and depth that are sustainable over time, colleges must find ways to incorporate the assessment of multiple attributes (knowledge, skills, and behaviors) within existing courses and frameworks.

A common solution to a too-full curriculum is to shift the need for certain information or content downstream, by increasing program prerequisites. While this strategy can provide some relief in the short term, it could negatively impact admissions by increasing the requirements applicants are asked to fulfill. While the prerequisite courses required of applicants to veterinary colleges are not uniform, there are a number of requirements that are common. Large numbers of required courses, or variation among admissions requirements, can create undue impediments or obstacles for prospective students interested in preparing for a career in veterinary medicine. Moreover, for any profession seeking to attract the greatest number of applicants with diverse backgrounds, this is an important issue that must be considered.

Similar to medical education, there is an emphasis on the need to ensure diversity within the student body and, ultimately, the profession. Certainly, in order for the profession to more closely resemble the society it serves, the number of underrepresented minorities who pursue a veterinary career must increase. Diversity is important not only in terms of an individual's ethnic or cultural diversity, but also in terms of her or his area of interest within the profession. The profession needs applicants who are also interested in research, in rural practice, in food supply veterinary medicine, and in public health. The gender shift that has occurred in the past 30 years also has implications for the profession, related to the declining number of male applicants. The decline of any group of applicants is troubling, and is at least partially attributed to the salary levels of new veterinary graduates and student indebtedness. Financial considerations create additional challenges for attracting a large, diverse, and talented group of prospective students.

Another important theme in veterinary education is the use of animals in teaching. Animals are an essential component of veterinary education, but associated costs, availability of healthy animals and of cadavers, concerns for animal welfare, and the availability of technological

alternatives, such as models and simulators, can impact the number and type of experiences offered. Even for noninvasive procedures, such as animal handling and restraint, the availability of animals for teaching can be an issue. In addition, as veterinary medical teaching hospitals provide increasing levels of tertiary care, students' exposure to routine cases that are typical of primary care practice is becoming limited. Animal patients in veterinary teaching hospitals are seriously ill, and students are often relegated to the role of observer. Creative solutions, such as spay/neuter programs between animal shelters and veterinary colleges, can help address societal problems such as pet overpopulation, while improving a shelter animal's chances of being adopted and providing opportunities for students to develop important skills. However, programs such as this will not meet all needs, including the need for case material in certain subject areas.

Many of the veterinary schools in the United Kingdom and Europe have made strides in the use of synthetic models, providing cost-effective opportunities for students to learn important skills. Other schools have forged partnerships with veterinary practitioners, who serve as adjunct faculty while educating veterinary students in private practices. Particularly for food animals, this model for clinical instruction has been used successfully in many parts of the world. As the US population becomes increasingly urbanized and veterinary colleges open referral clinics in metropolitan areas, this decentralized model of clinical instruction may become more common.

Challenges for Veterinary Education

Additional challenges to those noted above include the costs associated with clinical instruction and the economic future of the teaching hospital; demographic trends that project large numbers of faculty members will retire from veterinary colleges in the next two decades; the educational impact of increased enrollments in colleges of veterinary medicine to meet anticipated workforce needs; implications of limited licensure on educational programs, including continuing education; and ethical challenges resulting from new scientific advancements. Other developments that result from discoveries in cloning, synthetic animal-cell-based protein foods, organ farming, and xenotransplantation not only hold promise for improving human health and food safety, but also pose ethical challenges that veterinarians will need to help society address.

The opportunity for veterinary educators to find creative and innovative approaches for meeting these challenges is large and growing, but will require bold initiatives and vision. Professional schools are responsible for responding to the needs of their stakeholders, but must also take the initiative to continually define and advance their professions. Veterinary colleges must be leaders in ensuring

that the profession continues to serve society, promoting human and animal health, and in preparing flexible, skilled graduates who are capable of solving problems many of us cannot anticipate today.

See also: Education for Medicine; Professions – Health-Related.

Further Reading

- Adams, C. L. and Kurtz, S. M. (2006). Building on existing models from human medical education to develop a communication curriculum in veterinary medicine. *Journal of Veterinary Medical Education* **33**, 28–37.
- Blumberg, P. (2005). Why self-directed learning is not learned and practiced in veterinary education. *Journal of Veterinary Medical Education* **32**, 290–295.
- Brown, J. P. and Silverman, J. D. (1999). The current and future market for veterinarians and veterinary medical services in the United States: Executive summary. *Journal of the American Veterinary Medical Association* **215**, 161–183.
- Chieffo, C., Kelly, A. M., and Ferguson, J. (2008). Trends in gender, employment, salary, and debt of graduates of US veterinary medical schools and colleges. *Journal of the American Veterinary Medical Association* **233**, 910–917.
- Edmondson, K. M. (2001). Applying what we know about learning to veterinary education. *Journal of Veterinary Medical Education* **28**, 54–55.
- Farnsworth, C. C. (1997). Measuring the effects of problem-based learning on the development of veterinary students' clinical expertise. *Academic Medicine* **72**, 552–554.
- Fernandes, T. H. (2004). European veterinary education: A bridge to quality. *Veterinary Journal* **169**, 210–215.
- Halliwell, R. E. W. (2004). Accreditation of veterinary schools in the United Kingdom and the European Union: The process, current issues and trends, and future concerns. *Journal of Veterinary Medical Education* **31**, 105–110.
- Hubbell, J. A. E. (2008). Veterinary teaching hospitals: Current challenges and pathways for the future. *Journal of Veterinary Medical Education* **35**, 62–65.
- Kogan, L. R., Butler, C. L., Lagoni, L. K., et al. (2004). Training in client relations and communication skills in veterinary medical curricula and usage after graduation. *Journal of the American Veterinary Medical Association* **224**, 504–507.
- McGreevy, P. (2007). Firm but gentle: Learning to handle with care. *Journal of Veterinary Medical Education* **34**, 539–541.
- Michell, A. R. (2004). European veterinary education: What does harmonization mean? *Veterinary Journal* **169**, 157–158.
- Regehr, G. and Norman, G. R. (1996). Issues in cognitive psychology: Implications for professional education. *Academic Medicine* **71**, 988–1001.
- Walsh, D. A., Osburn, B. I., and Christopher, M. M. (2001). Defining the attributes expected of graduating veterinary medical students. *Journal of the American Veterinary Medical Association* **219**, 1358–1365.
- Willis, N. G., Monroe, F. A., Potworowski, J. A., et al. (2007). The Foresight Report. *Journal of Veterinary Medical Education* **34**, 3–25.

Relevant Websites

- <http://www.aavsb.org> – American Association of Veterinary State Boards.
- <http://www.avma.org> – American Veterinary Medical Association.
- <http://www.aavmc.org> – Association of American Veterinary Medical Colleges.
- <http://www.nbvme.org> – National Board of Veterinary Medical Examiners.
- <http://canadianveterinarians.net> – National Examining Board (Canada).
- <http://www.nbvme.org> – North American Veterinary Licensing Examination.
- <http://www.onehealthinitiative.com> – One Health Initiative.

EDUCATION RESEARCH METHODOLOGY: QUANTITATIVE METHODS AND RESEARCH

Contents

Abductive Research Methods

Field Experimentation in Education

Internal Validity

Methods for Approximating Random Assignment

Methods for Approximating Random Assignment: Regression Discontinuity and Propensity Scores

Multitrait–Multimethod Designs

Multivariate Longitudinal Data Analysis

Quasi-Experimentation: Two Group Design

Regression Discontinuity Designs

Sampling

Statistical Conclusion Validity

Survey Research Methods

Abductive Research Methods

B D Haig, University of Canterbury, Christchurch, New Zealand

© 2010 Elsevier Ltd. All rights reserved.

Educational and social science researchers make use of different kinds of research methods and strategies. Some of them are inductive in nature. For example, the well-known triangulation strategy of constructive replication is designed to establish the generalizability of empirical relationships. It is, therefore, a strategy of enumerative induction. Other methods are hypothetico-deductive in nature. Structural equation modeling, for example, conforms to the hypothetico-deductive strategy of testing certain classes of latent variable models in terms of their empirical adequacy.

Less well known is a third class of methods that is abductive in nature. These methods differ from inductive and hypothetico-deductive methods in that they make use of explanatory considerations in order to judge the worth of research hypotheses and theories.

Although researchers frequently engage in explanatory reasoning, there is a dearth of codified abductive methods available for ready use in the behavioral sciences. Scientists use abductive methods to a limited extent, although they seldom refer to them as abductive methods. This article discusses three different abductive research methods that are of relevance to educational and social science research. The first of these – exploratory factor analysis – has been widely used in education, although it is not generally recognized as an abductive method. The second abductive method – known as the theory of explanatory coherence

(TEC) – is a promising approach to theory appraisal that can be used to evaluate empirically based educational theories. The third method – grounded theory – is a well-known approach to conducting qualitative research, but it tends to be regarded as an inductive method rather than as an abductive method.

Exploratory Factor Analysis

Exploratory factor analysis (EFA) is a multivariate statistical method designed to facilitate the postulation of latent variables that are thought to underlie – and give rise to – patterns of correlations in new domains of manifest variables. Intellectual abilities, personality traits, and social attitudes are well-known classes of latent variables that are the product of factor-analytic research. EFA uses multiple regression and partial correlation theory to model sets of manifest or observed variables in terms of linear functions of other sets of latent – or unobserved – variables. Despite being regarded by many as a latent variable method, its inferential nature is usually described as inductive. This is odd, because inductive inference – being descriptive inference – cannot take the researcher to theoretical entities that are different in kind from their manifest effects. Therefore, EFA is best understood as an abductive method of theory generation (Haig, 2005a) – a

characterization that coheres well with its general acceptance as a latent variable method.

EFA is well positioned to facilitate the drawing of explanatory inferences that can be described as existential abductions. Existential abductions enable researchers to hypothesize the existence – but not the nature – of entities previously unknown to them. The innumerable examples of existential abduction in science include the initial postulation of hidden entities such as atoms, genes, tectonic plates, and personality traits. In cases such as these, the primary thrust of the initial abductive inferences is to claims concerning the existence of theoretical entities in order to explain empirical facts or phenomena. Similarly, the hypotheses given to us through the use of EFA postulate the existence of latent variables such as Spearman's g and extraversion. It remains for further research to elaborate on the first rudimentary conception of these variables and their interrelation.

The factor-analytic use of existential abduction to infer the existence of, say, the theoretical entity g can be coarsely reconstructed in accordance with the following argument schema:

The surprising empirical phenomenon of positive correlations among different tests of ability is identified. If g exists, and it is validly and reliably measured by a Weschler intelligence scale (and/or some other objective test), then the positive manifold would follow as a matter of course.

Hence, there are grounds for judging the hypothesis of g to be initially plausible and worthy of further pursuit.

Note that the schema for abductive inference – and its application to the generation of the hypothesis of g – are concerned with the form of the arguments involved, not with the actual generation of the explanatory hypotheses. In each case, the explanatory hypothesis is given in the second premise of the argument. An account of the genesis of the explanatory hypothesis must, therefore, be furnished by some other means. It is plausible to suggest that reasoning to explanatory hypotheses trades on our evolved cognitive ability to abductively generate such hypotheses. Whatever its origin, an informative methodological characterization of the abductive nature of factor-analytic inference must appeal to the scientist's own psychological resources as well as those of logic.

Although EFA exemplifies well the character of existential abduction, it is, clearly, not an all-purpose method for abductively generating explanatory hypotheses and theories. With its focus on common factors, it can appropriately serve as a generator of elementary theories only in those multivariate domains that have common causal structures.

It is well known that EFA is a common factor-analytic model in which the latent factors it postulates are referred to as common factors. There is an important principle of

scientific inference – known as the principle of the common cause – which drives the nature and shape of the existential abductive inferences involved in EFA. The principle can be formulated concisely as follows: “Whenever two or more events are improbably, or significantly, correlated infer one or more common causes, unless there is good reason not to.” Clearly, the principle should not be taken as a hard-and-fast rule, for – in many cases – proper inferences with regard to correlated events will not be of the common causal kind. The qualifier, “unless there is good reason not to,” should be understood as an injunction to consider causal interpretations of the correlated events other than the common causal kind. For example, the correlated events might be related as direct causes, or they might be mediated by a third variable in a causal sequence. Understood in the context of theory generation, methods of existential abduction such as EFA should not be expected to achieve highly developed and well-validated scientific theories. At best, they deliver rudimentary theories that have initial plausibility. It is important to realize that these abductive methods enable us to justify the initial plausibility of the theories they spawn. The very process of the abductive generation of theories has a bearing on the first determinations of their worth, in that one appeals to the soundness of the abductive arguments employed in the introduction of theories in order to evaluate their early epistemic promise.

Relatedly, the nascent theories bequeathed us by methods like EFA postulate the existence of hidden causal mechanisms. They have the status of dispositional theories in that they provide us with oblique characterizations of the properties we attribute to things by way of their presumed effects under specified conditions. A move beyond the rudimentary nature of their dispositional characterization will require subsequent elaboration by other means.

The Theory of Explanatory Coherence

Inference to the best explanation is the name given to an important, but neglected, approach to scientific theory evaluation. This approach is founded on the belief that much of what we know about the world is based on considerations of explanatory worth. Being concerned with explanatory reasoning, inference to the best explanation is a form of abduction. It involves accepting a theory when it is judged to provide a better explanation of the evidence than its rivals do (Haig, 2009; Capaldi and Proctor, 2008). In science, inference to the best explanation is, often, used to adjudicate between well-developed, competing theories. The major point of inference to the best explanation is that the explanatory merits of explanatory theories should count in their favor.

Thagard (1992) has developed a detailed account of inference to the best explanation as a scientific method – one which helps a researcher to reliably appraise competing theories in terms of their explanatory goodness. This method is known as the *theory of explanatory coherence* (TEC). The theory comprises an account of explanatory coherence that includes a number of evaluative criteria and constituent principles, a computer program for implementing the principles, and various simulation studies that demonstrate its promise as a method of inference to the best explanation.

According to TEC, inference to the best explanation is centrally concerned with establishing relations of explanatory coherence. To infer that a theory is the best explanation is to judge it as more explanatorily coherent than its rivals. TEC is not a general theory of coherence that subsumes different forms of coherence, such as logical and probabilistic coherence. Rather, it is a theory of explanatory coherence where the propositions hold together because of their explanatory relations.

Relations of explanatory coherence are established through the operation of seven principles. These principles are: symmetry, explanation, analogy, data priority, contradiction, competition, and acceptance. The determination of the explanatory coherence of a theory is made in terms of three criteria. Within the TEC, each of these criteria is embedded in one or more of the seven principles. In this instance, the criteria are considered first.

Thagard determined that explanatory breadth is the most important criterion for choosing the best explanation. This criterion captures the idea that a theory is more explanatorily powerful than its rivals if it explains a greater range of facts.

The notion of simplicity that Thagard deems most appropriate for theory choice is a pragmatic criterion that is closely related to explanation; it is captured by the idea that preference should be given to theories that make fewer special or *ad hoc* assumptions. Thagard regards simplicity as the most important constraint on explanatory breadth; one should not sacrifice simplicity through *ad hoc* adjustments to a theory in order to enhance its explanatory breadth.

Finally, Thagard found that analogy is an important criterion of inference to the best explanation because it can improve the explanation offered by a theory. Explanations are judged more coherent if they are supported by analogy to theories that scientists already find credible.

Within TEC, each criterion is embedded in one or more of the seven principles. These principles are stated here informally in Thagard's words (Thagard, 2000: 43), and the accompanying comment on the principles closely follows his (1992) discussion of a more formal statement of those principles.

1. *Symmetry*. Explanatory coherence is a symmetric relation, unlike, say, conditional probability. That is, two propositions p and q cohere with each other equally. The principle of symmetry maintains that both coherence and incoherence are symmetric relations, unlike the nonsymmetric relations of entailment and conditional probability. The sense of coherence conforms to the ordinary sense of coherence as holding together.
2. *Explanation*. (a) A hypothesis coheres with what it explains, which can either be evidence or another hypothesis. (b) Hypotheses that together explain some other proposition cohere with each other. (c) The more hypotheses it takes to explain something, the lower the degree of coherence. Because the principle of explanation establishes most of the coherence relations, it is the most important principle in determining explanatory coherence. Principle 2a, with principle 7, acceptance, subsumes the criterion of explanatory breadth, which is central to determining the best explanation. Principle 2c accommodates the notion of simplicity which is an important criterion in theory choice.
3. *Analogy*. Similar hypotheses that explain similar pieces of evidence cohere. The principle of analogy is the same as the criterion of analogy just mentioned. It states that if similar propositions explain similar pieces of evidence, then they cohere with each other. The analogy must be explanatory in nature.
4. *Data priority*. Propositions that describe the results of observations have a degree of acceptability on their own. The principle of data priority maintains that claims concerning observations and empirical generalizations can stand on their own more successfully than explanatory hypotheses. Of course, they can be doubted, but the reliability of their production will often be sufficient grounds for their initial acceptance.
5. *Contradiction*. Contradictory propositions are incoherent with each other. This principle straightforwardly includes syntactic contradictions involving logical inconsistency and semantic contradictions involving inconsistency of meaning. The principle covers the negative relations that hold between contradictory propositions that actively resist cohering, and that are said to incohere.
6. *Competition*. If p and q both explain a proposition, and if p and q are not explanatorily connected, then p and q are incoherent with each other (p and q are explanatorily connected if one explains the other or if together they explain something). This principle claims that theories that explain the same evidence should normally be treated as competitors. In such cases, theories are regarded as competing if they are not explanatorily related. Noncontradictory theories may compete with each other.
7. *Acceptance*. The acceptability of a proposition in a system of propositions depends on its coherence with

them. This last principle asserts that propositions are accepted or rejected based on their degree of coherence with other propositions. The overall coherence of a system of propositions, or a theory, is obtained by considering the pair-wise coherence relations through use of principles 1–6.

The principles of TEC combine in a computer program, Explanatory Coherence by Harmany Optimization (ECHO), to provide judgments of the explanatory coherence of competing theories. This computer program is connectionist in nature and uses parallel constraint satisfaction to accept and reject theories based on their explanatory coherence.

The TEC has a number of virtues which make it an attractive theory of inference to the best explanation: It satisfies the demand for justification by appeal to explanatory considerations rather than predictive success; it takes theory evaluation to be a comparative matter; it can be readily implemented by – and, indeed, is instantiated in – the computer program, ECHO, while still leaving an important place for judgment by the researcher; and, it effectively accounts for a number of important episodes of theory assessment in the history of science. In short, TEC and ECHO combine in a successful method of explanatory coherence that enables researchers to make judgments of the best of competing explanatory theories. Thagard (1992) is the definitive source for a detailed explication of the TEC.

The behavioral and social sciences are replete with competing theories that might usefully be evaluated in respect of their explanatory coherence. However, their use of TEC to appraise the best of competing explanatory theories is yet to be undertaken.

Grounded Theory

Grounded theory (GT) is probably the most widely known methodological perspective on how to conduct qualitative research in the social sciences. Originally introduced by sociologists Barney Glaser and Anselm Strauss (Glaser and Strauss, 1967), GT is used extensively in education and related fields.

GT comprises a distinctive methodology, a particular view of scientific method, and a set of specific procedures for analyzing qualitative data and constructing theories from those data. The methodology provides a justification for regarding qualitative research as a legitimate – indeed, rigorous – form of inquiry. The view of scientific method adopted by GT is generally taken to be inductive in nature, although this is a contested matter. GT researchers gather non-numeric data from a variety of sources, including interviews and field observations. Once gathered, the data are analyzed using coding and theoretical

sampling procedures. A set of interpretative procedure are then used to assist in the construction of theory that emerges from, and is grounded in, the data.

In efforts to identify empirical social phenomena, and construct theories that are constrained by those phenomena, almost all accounts of GT adopt the three major strategies of data coding, memo writing, and theoretical sampling.

In GT, data gathering and data analysis are interactive. From the time data collection begins, grounded theorists engage in data analysis, which leads to further data collection, subsequent data analysis, and so on.

The first data analytic phase of GT begins with the coding of the data. This is undertaken to conceptualize the data by discovering categories into which they fit. The coding process has three phases: open coding, axial coding, and selective coding. In open coding, researchers describe the data by looking at it line-by-line. This strategy of focusing on small units of data, and their interpretation, encourages the development of a theoretical sensitivity to new ideas with regard to the data, and helps prevent the forcing of data into existing categories. Strauss (1987) maintains that when a full array of categories have been identified, one should undertake axial coding – whereby one puts the data back together again in new ways by making connections between the numerous categories. Following that, a selective coding step is implemented in which the researcher looks to systematically identify those categories that relate closely to the core category. The core category lies at the heart of the emerging theory and is central to its integration.

Although memo writing can occur at any stage of the research process, it frequently takes place between the coding of data and the writing of the initial draft of the research report. Memos are written to identify, develop, and keep track of theoretical ideas. Where relevant, they are recorded, recalled, and reworked to produce new theoretical memos. Memo writing becomes more systematic, focused, and intense as theory of greater density and coherence is produced.

Memos written with regard to data codes and theoretical ideas enable the researcher to identify gaps that require the collection of further data. For this, theoretical sampling is undertaken. With theoretical sampling – in contrast with traditional representative sampling – decisions concerning what data to collect, code, analyze, and interpret are directed by the emerging GT. Theoretically relevant events, activities, and populations are all sampled, and the comparisons between these are aimed at increasing the conceptual density and integration of the emerging theory. Thinking effectively with regard to data in theoretical terms requires an adequate degree of theoretical sensitivity. When the additional gathering and analysis of data no longer contribute to the understanding of a concept or category, a point of theoretical saturation

is reached. At this point, one stops collecting data in respect of a category and moves to consider another category or concept.

Consistent with the pragmatist influences on GT methodology, Strauss (1987) characterizes scientific method as a sequence of induction, deduction, and induction: grounded theories emerge inductively from the data, test predictions are then deduced from the theories, and, finally, the theories are inductively confirmed or disconfirmed.

Despite the considerable attention given to the exposition of data analysis in GT, it is difficult to fathom just how – and in what sense – GT is said to inductively emerge from, and be grounded in, the data. This is because the nature of the inductive reasoning involved is not described.

Glaser and Strauss have been criticized on the grounds that they advocate a return to a simple and unacceptable Baconian inductivism. On this interpretation, GT is depicted as a *tabula rasa* view of inquiry which maintains that observations are not theory or concept dependent. However, this is not Glaser and Strauss's position. In *The Discovery of Grounded Theory* (Glaser and Strauss, 1967), they explicitly disavow this view of inquiry – noting that the researcher requires a theoretical perspective in order to see and abstract from data. It is in the interest of obtaining emergent, diverse categories at different levels of abstraction that Glaser and Strauss would have the researcher hold all potentially relevant facts and theories in the background for some time. Clearly, this is a form of bracketing, not a *tabula rasa* conception of inquiry.

Although it is clear that Glaser and Strauss are not naive inductivists, the actual nature of the inductive relation that – for them – grounds emergent theories in their data is difficult to fathom. For Glaser and Strauss, GT is said to emerge inductively from its data source in accordance with the method of constant comparison. As a method of discovery, the constant comparative method is an amalgam of systematic coding, data analysis, and theoretical sampling procedures which enables the researcher to make interpretive sense of much of the diverse patterning in the data by developing theoretical ideas at a higher level of abstraction than the initial data descriptions. However, the notion of constant comparison is of little help in figuring out whether the inductive inference in question is enumerative, eliminative, or of some other form.

Given the pragmatist influence on GT methodology, it is not surprising that Strauss (1987) mentions the notion of abduction in his brief discussion of induction. Unfortunately, however, he refrains from including it in his discussion of the inductive generation of theory. In his mature writing, the American pragmatist Charles Sanders Peirce clearly distinguished between these two forms of inference. Both inductive and abductive arguments are ampliative, or content-increasing – in that their

conclusions contain more information than is contained in their premises. However, the type of ampliation is different for each. Inductive arguments are descriptive in character because they reach conclusions concerning the same type of manifest attributes mentioned in their premises. By contrast, abductive arguments reason from factual premises to explanatory conclusions, as when we reason from presumed effects to underlying causes.

A growing number of authors have characterized the creative inference involved in the generation of GT as abductive in nature (e.g., Haig, 1996; Reichertz, 2007) – that is, rather than viewing a GT as an inductive abstraction from data analysis, it is thought of as the result of explanatory inference to factors that transcend the data in a more fundamental way. On this view, the data analytic dimension of GT can reasonably be construed as inductive in nature. However, in order to explain the abstracted data patterns, the construction of GT needs to be thought of as abductive in nature.

Haig has gone further and suggested that the entire process of theory construction in GT can be cast in an abductive light (Haig, 1996, 2005b). On his account, the abductive nature of GT extends beyond theory generation to include theory development and theory appraisal. A strategy of analogical modeling is used to develop GT. Because analogical modeling increases the content of explanatory theories, the reasoning it embodies is referred to as analogical abduction. This reconstruction of GT adopts inference to the best explanation as the preferred approach to the evaluation of mature theories. Specifically, TEC – outlined earlier – is adopted, and the better of competing GTs is judged to be the one that is more explanatorily coherent.

Conclusion

The fundamental purpose of scientific inquiry is to detect and explain empirical phenomena. Social science methodology is replete with methods that help in the descriptive task of identifying empirical phenomena. However, in order to help researchers explain these phenomena, abductive methods such as those described in this article should be part of their methodological armamentarium.

See also: Critical Thinking; Grounded Theory; Learning as Inquiry.

Bibliography

- Capaldi, J. and Proctor, R. (2008). Are theories to be evaluated in isolation or relative to alternatives? An abductive view. *American Journal of Psychology* **121**, 617–624.

- Glaser, B. and Strauss, A. L. (1967). *The Discovery of Grounded Theory*. Chicago, IL: Aldine.
- Haig, B. D. (1996). Grounded theory as scientific method. In Neiman, A. (ed.) *Philosophy of Education 1995*, pp 281–290. Urbana, IL: University of Urbana Press.
- Haig, B. D. (2005a). Exploratory factor analysis, theory generation, and scientific method. *Multivariate Behavioral Research* **40**, 303–329.
- Haig, B. D. (2005b). An abductive theory of scientific method. *Psychological Methods* **10**, 371–388.
- Haig, B. D. (2009). Inference to the best explanation: A neglected approach to theory appraisal in psychology. *American Journal of Psychology* **123**, 219–234.
- Reichert, J. (2007). Abduction: The logic of discovery of grounded theory. In Bryant, A. and Charmaz, K. (eds.) *The Sage Handbook of Grounded Theory*, pp 215–228. Los Angeles, CA: Sage.
- Strauss, A. L. (1987). *Qualitative Analysis for Social Scientists*. New York: Cambridge University Press.
- Thagard, P. (1992). *Conceptual Revolutions*. Princeton, NJ: Princeton University Press.
- Thagard, P. (2000). *Coherence in Thought and Action*. Cambridge: MIT Press.

Further Reading

- Bryant, A. and Charmaz, K. (eds.) (2007). *The Sage Handbook of Grounded Theory*. Los Angeles, CA: Sage.
- Lipton, P. (2004). *Inference to the Best Explanation*, 2nd edn. London: Routledge.
- Mulaik, S. A. (2009). *The Foundations of Factor Analysis*, 2nd edn. London: Chapman/Hall/CRC.

Relevant Websites

- <http://www.unc.edu> – Robert C MacCallum's, The University of North Carolina at Chapel Hill.
- <http://cogsci.uwaterloo.ca> – JavaECHO, Paul Thagard's Computational Epistemology Laboratory, University of Waterloo.
- <http://dmoz.org> – Open Directory Project (ODP), Science: Social Sciences: Methodology: Grounded Theory.

Field Experimentation in Education

M A Conostas, Cornell University, Ithaca, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Between groups – Feature of research design that involves the administration of measures to different groups of participants as way to compare the effect of one study condition against another study condition.

Causal inference – The cognitive processes, rules of logic, empirical conditions, and theoretical assumptions that lead one to make assertions about the existence of cause and effect relations.

Counterfactual – The existence or construction of a set of investigative circumstances organized to satisfy the basic empirical conditions needed build an argument for cause and effect relations.

Implementation measure – An indicator that reflects the degree to which a given intervention was administered as intended.

Instrumentation – Any one of a variety of tools used to collect data.

Power – The probability that a statistical test will yield significant results to support the rejection of the null hypothesis.

Randomization – Research procedure in which individuals or investigative units (e.g., classroom, schools, school districts) have an equal chance of being allocated to one of several.

Treatment – An intervention selected to produce a desired set of outcomes.

Units – Entities of varying scale, from individuals to collections of individuals, that serve as study samples.

Within groups designs – Feature of research that involves the administration of measures to a single group of participants (on different occasions) as way to compare the effect of one study condition against another study condition.

All questions about research design, when correctly conceived, take the formulation of problems as their points of departure. As Karl Popper has argued, “[t]he progress of science lies, essentially, in the evolution of its problems” (Popper, 1994: 155–156). A concern for the advance of science must item from an understanding of the general form of the problem at hand, rather than from an allegiance to a particular design. With this in mind, the justification for field experimentation can be constructed

as a response to a foundational question. How might education be defined and what is the argument for the application of field experiments? Consistent with the perspective of pragmatism (see Dewey, 1938; Peirce, 1905), education may be defined as the intended and unintended configuration of cause and effect relations. The causes may be viewed as a set of characteristics held by individuals (e.g., students, teachers, and administrators) and groups of individuals with shared attributes, as a set of traits or practices followed by those individuals and groups, as the set of systemic variables such as school size school setting (e.g., urban vs. rural), school organization (e.g., decentralized governance), or school finance (e.g., funding models). The effects may be viewed as a set of valued educational outcomes, such as achievement scores, conceptual understanding, motivation, psychological well-being, or a host of other academic and nonacademic outcomes.

If evidence is to play a role in education by solving the problems of education, decisions about which aspects of education need to be retained, eliminated, or more widely disseminated depend on one’s inclination to view education as essentially causal. Within the research enterprise, the goodness of evidence depends on one’s ability to produce relatively unbiased estimates of the effect that selected characteristics and conditions of education have had on outcomes of interest. When education is defined in terms of cause-and-effect relations, field experiments are the best design option. No other design option has been shown to produce estimates of effect more accurately than field experiments. The process of randomly assigning educational units (e.g., students, classrooms, and schools) to different educational conditions offers the important benefit of randomly distributing the array of known and unknown characteristics evenly across study conditions. This is important because any one of such characteristics (and probably more than one) could interfere with the desire to accurately estimate effects of a given intervention or set of interventions. In the absence of random assignment, the possibility of systematic error linked to group differences becomes more pronounced. The production of reasonably unbiased estimates of effect, in combination with a well-developed theory of causal relations for a particular phenomenon, is what permits one to make the evidence-based assertion that a given set of educational conditions brought about a particular set of outcomes. Identifying the most productive configuration of cause–effect relations is of great interest, both to educational professionals who are in a position to select and

implement programmatic initiatives and to researchers who attempt to produce trustworthy estimates of the effects of those same initiatives.

At least since the work of Fisher (1926, 1935), experimental design has been regarded as the best option for producing reasonably unbiased estimates of the effects of a given intervention. Education, perhaps more than any field, has been slow to explore the inferential benefits afforded by experimental design. Until quite recently, it has been very difficult to locate significant numbers of experiments in the field of education (see Boruch *et al.*, 2002; Cook, 2001). In the United States, the passage of key legislation for both education program and for educational research produced a shift in the kind of research designs employed in the field of education (see Constanas, 2007). The passage of the Elementary and Secondary Education Act of 2000 and the passage of the Educational Sciences Reform Act of 2001 both highlighted the importance of randomized field trials. In the past few years, a new professional organization was formed, the Society for Research on Educational Effectiveness and a new peer-reviewed journal was founded, the *Journal of Research on Educational Effectiveness* (JREE). The new organization, along with its flagship journal JREE, places randomized field trials at the center of its work. The landscape of educational research has undergone a significant change over the past few years. The use of field experiments, which were relatively rare over the course of recent history, has been applied more regularly in the field of education. The purpose of this article is to describe the most critical design features of field experiments, to discuss a limited number of statistical issues, and to highlight the importance of planning and implementation. It should be noted that the article summarizes the design features of studies that randomly assign educational entities of varying scale (e.g., individuals, groups, and locations) to study conditions. There are a range of other methods that are also used to produce estimates of the effects of educational intervention that do not randomly assign educational units to study conditions. Such methods include, but are not limited to, quasi-experimental design (e.g., Cook and Campbell, 1979), interrupted time series (e.g., Bloom, 2003), regression discontinuity designs (e.g., Thistlewaite and Campbell, 1960), and propensity score matching (e.g., Rosenbaum and Rubin, 1984). Shadish *et al.* (2002) provide a well-detailed treatment of the full range of experimental and quasi-experimental design options that can be used to generate estimates of effect, with and without randomization.

Field Experimentation: An Overview

Field experiments are based on the inferential logic of laboratory experiments. There are two key characteristics of the laboratory-based experiments. One characteristic is

the construction of a controlled environment, one in which the conditions meant to produce a particular effect are carefully manipulated. All other conditions theorized to have an effect are held constant, to the extent they can be. Such environments are designed to be as free as possible from influences that would disturb the attempt to establish cause-and-effect connections between features of the manipulated environment and the observed outcomes of interest. A second key characteristic of laboratory experiments is that they employ the method of randomly assigning units to either a treatment condition or to a control condition. The units may be individuals, groups of individuals, or locations where individuals cluster (e.g., classrooms, schools, and school districts) (see Donner, 1988). The practice of random assignment is critical because it establishes a counterfactual condition, where a counterfactual represents an all-else-equals comparison (see Lewis, 1973; Mackie, 1974). This creation of a plausible counterfactual means that any observed difference of significant magnitude should be attributable to differences in the conditions that were intentionally manipulated. Outside influences (e.g., individual differences or group differences) that could interfere with the attempt to make such attributions should, in principle, be equally distributed across study conditions.

Markedly different from laboratory or controlled conditions under which laboratory studies are conducted, field experiments are conducted in natural settings where influences that are not integral to the explicit purpose of the investigation must be reckoned with. While field experiments are not characterized by controlled environments, the element of random assignment represents the foundational feature of this design option. Shadish *et al.* (2002: 12) define an experiment simply as "... a study in which an intervention is deliberately introduced to observe its effects." As a specific type of experiment, a randomized field experiment is a study which has two additional features. First, the feature of randomization requires that each of the entities that serve as the primary analytic entities (e.g., students, teachers, classrooms, and schools) has an equal chance of being part of the group that receives the treatment as it does of being part of the group that does not receive the treatment. The unbiased, random establishment of an intervention (or treatment) group and a nonintervention (or control) group is assured by randomly assigning the entities to their respective groups. Second, the reference to the idea of the field means that the study is conducted in everyday educational settings. The second feature places a particular set of demands on the investigator as he/she contemplates ways to deal with the inferential hazards wrought by an uncontrolled set of study conditions. Randomization, as a key feature of field experiments, addresses some of these challenges but not all. The way in which randomization is deployed along with other components of design provides a fuller response to the factors that cloud

the investigator's ability to minimize sources of bias, beyond selection bias. At least since the work of Fisher (1926, 1935), experimental designs that employ random assignment have been regarded as the best option for producing reasonably unbiased estimates of the effects of a given intervention. In their assessment of designs used within the field of education Schneider *et al.* (2007: 11) noted that "when correctly implemented, the randomized controlled experiment is the most powerful design for detecting treatment effects." Boruch and Foley (2000: 194) describe a randomized field trial as a "... study that attempts to produce a fair comparison of the relative effectiveness of two or more approaches so as to understand what works, for whom, and for how long."

While randomization is a necessary component of field experimentation, it is one of many design components to consider in the planning of a field experiment. If the aim is to increase precision and to minimize bias, the potential inequalities produced by differences between or among study samples represent only a single source of error. As Shadish *et al.* (2002: 256) noted "... random assignment is just one part of experimental design" and range of other important issues such as "scheduling of observations, the choice of treatments and comparisons, the selection of observations and measures, the determination of who should be the respondents, and the manner of assigning units to treatments."

Types of the Randomized Field Experiments Through the Lens of Internal Validity

In investigations of educational effectiveness, well thought-out research design is what allows one to separate, with a reasonable degree of confidence, treatment effects from a range of unwanted influences. Any one design type has an associated set of strengths and weaknesses that inhibit or facilitate an investigator's ability to draw inferences about the effect of a given intervention. With reference to research design, the ability to draw such inferences is best explained by the notion of internal validity (see Campbell, 1957; Campbell and Stanley, 1963; Cook and Campbell, 1979). Shadish *et al.* (2002: 55) describe nine different threats to the internal validity of a study:

1. ambiguous temporal precedence that indicates a confusion between the ordering of hypothesized cause and effects relations;
2. selection effects that produce differences in the characteristics of study samples (i.e., treatment vs. control);
3. history effects that admit a set of influences concurrent with the intervention;
4. maturation effects that account for change as a natural consequence of development of the entity (e.g., individual and group of individuals) under study;

5. regression effects that highlight the consequences of selecting entities at extreme points in a distribution on an outcome of interest (e.g., achievement and self-esteem);
6. attrition effects that exist when differential, nonrandom factors account for attrition observed between study groups;
7. testing exposure that produces change (pretest to posttest) as a consequence of being sensitized to attributes of interest experienced in a pretest;
8. instrumentation effects that produce changes in outcomes that may be incorrectly attributed to the intervention; and
9. additive and interactive effects that result from the more than one of the eight main threats to internal validity.

For a more comprehensive discussion of each of the threats to internal validity, as well as other components of validity that influence the strength of inferences (e.g., construct validity, statistical conclusion validity, and external validity) the reader is directed to Shadish *et al.* (2002: 33–102). The concept of internal validity is introduced here because it provides a useful point of access to consider the consequences of different design options.

There are a large number of designs that one may employ under the heading of randomized field experiments. Differences among the designs can be described in terms of the four features, expressed here in the form of four questions. First, was a pretest employed as part of the design? Second, was a control group employed and, if so, how many control groups were used to construct a counterfactual? Third, how many treatments were compared? Fourth, at what points were data collected and were data collected to draw inferences about discrete effects between two points in time or about growth over more than two periods of time? Using five common designs, **Table 1** illustrates how such questions are answered. The table also includes details related to the relative strengths and weaknesses of each design in terms of internal validity. It is important to note the designs and the details on internal validity are intended to be illustrative and not comprehensive.

In an educational setting, the most basic form of a randomized field experiment involves two groups, where one group receives the intervention that is under examination and the second group continues under the standard set of educational conditions (the control group, denoted as Ctrl). More complicated variations on this basic design are accomplished by adding more treatment groups, more control groups, and by varying the placement of data collections before, during, and after treatments. Following a design framework offered by Shadish *et al.* (2002), **Table 1** shows how the use of Tx groups, control groups, pretests, and posttests, vary across the four illustrative design types. What is important to notice in **Table 1** is that the addition

Table 1 A sample of design types: Observed effects, strengths, and weaknesses*

<i>Design type and Comparisons</i>	<i>Example of differences observed to assess various effects</i>	<i>Strengths and weaknesses of design type</i>
1. Tx with no control group Single group compared against itself over time with a single Tx, control group	Pretest–posttest = Tx^1 effects	+ Controls for between-group difference via repeated measures + Measures change – No counterfactual condition – No measure of maturation – Instrumentation effects unmeasured
2. Basic randomized design comparing two Tx Two Tx groups compared against each other at posttest, no pretest administered and no control group	Pretest (Tx^1)–posttest(Tx^2) = Comparison of two Tx effects	+ Controls for between-group difference via randomization – Does not measure change or maturation – No counterfactual condition – Instrumentation effects unmeasured
3. Basic randomized design comparing two Tx and a control Three groups compared against each other, no pretest administered	Posttest(Tx^1) – posttest(Ctrl) = Tx^1 effects Posttest(Tx^2) – posttest(Ctrl) = Tx^2 effects	+ Controls for between-group difference via repeated measures + Includes a counterfactual Condition – does not measure change or maturation – Instrumentation effects unmeasured
4. Pretest, posttest control group Two groups compared, one with Tx and one without	Posttest(Tx^1) – posttest (Tx^2) = comparative effects of Tx^1 vs. Tx^2 Pretest(Tx^1) – pretest(Ctrl) = initial group differences Posttest(Tx^1) – posttest(Ctrl) = Tx^1 effects Posttest(Ctrl) – posttest(Ctrl) = Maturation effects (Pretest () – posttest (Tx^1)) – maturation effects = Tx^1 taking maturation into account	+ Controls for between-group difference via randomization – Includes a counterfactual condition + Measures maturation/growth – Instrumentation effects unmeasured
5. Longitudinal design with pre-test multiple posttests and single Tx Two groups (or more) compared, one or more with Tx, one without control, with pretests and multiple posttests	Posttest(Tx^1) – posttest(Ctrl) = Tx^1 effects Pretest(Tx^1) – posttest(Ctrl) = initial group differences Posttest(Ctrl) – posttest Ctr ¹) = maturation effects Pretest (Tx^1) – posttest one Tx^1 = effects Posttest one (Tx^1) – posttest two Tx^1 = growth effects	+ Includes counterfactual + Instrumentation effects measured + Measures growth over time + Measure maturation + Measures instrumentation – Costly/time-consuming design

of control groups, the judicious placement of measures, and the opportunity to track change over time influence the strength of inferences that may be drawn from a given study.

The ability to draw inferences on the relative effectiveness of a given intervention requires that one has comparative data. The second column of **Table 1** shows, in simple notation, the kinds of comparisons afforded by the five designs. Some designs only allow the investigator to assess the effect of the intervention without assessing the effect of other factors that could cloud the clarity of the inferences drawn. The more comparisons afforded by a given design type, the better prepared the investigator is to separate intervention effects from nonintervention effects or effects introduced by threats to internal validity.

To provide a more specific illustration of the strengths and weaknesses of each design presented in **Table 1**, consider a field experiment on the effects of an intervention meant to improve early language skills among a group of students over the course of their first year of primary school. For such a study, one would be interested in determining how effective the new intervention is in comparison to the standard set of practices. The development of a design that will best support the investigator's attempt to separate the effects of the intervention from other effects is confronted with a number of immediate challenges. For the sake of illustration, let us consider three sources of interference connected to internal validity. First, knowledge of the language development for

primary school children suggests that, even in the absence of any intervention, language skills are likely to develop over the time (maturation effects). Second, exposure to a pretest is likely to have an impact on the development of language skills (instrumentation effect). Moreover, assume that this impact is amplified when offered in combination with the intervention. Third, the conclusions that may be drawn about growth depend on assumptions made about trajectory of language development over the course of the year. Consideration of these three sources of potential interference demonstrates the necessity of design that includes a first control group (to address maturation effects), a second control group with no pretest (to address the instrumentation), and the collection of data in a longitudinal format (to examine growth). A simple pretest, posttest design with no control group (design number one **Table 1**) would not permit the investigator to separate intervention effects from maturation effects. A design that included a control group in addition to the Tx group would help the investigator assess maturation effects, but the potential effects of instrumentation would not be known. Finally, a design that collects data at only two points in time (pretest and posttest) would not allow the investigator to measure growth in the most appropriate manner. The demands of inference (internal validity) combined with knowledge of language acquisition (construct validity) suggest that a pretest, posttest longitudinal design with two control groups would be the best option.

As previously noted, the description of design types offered here is not comprehensive as there are obviously wide variety of ways to organize a field experiment. Ultimately, decisions related to the use of a control group (or control groups), pretests, and posttests are made on the joint consideration of two sets of design issues, one of which is based in the logic of inference, while the other is based on the realities of the field settings. Design issues connected to the logic of inference reflect an ambition to design and conduct studies that can best anticipate threats to internal validity. Design issues connected to the realities of the field reflect an understanding of local contexts and a concern for the ethics of random assignment. The concern for ethics, while critical, is often overstated. There are many educational interventions, for example, in which enrolment in an educational program operates on a lottery. On other occasions, the implementation of an intervention is phased in over time, across a district or across a state.

Planning and Implementation Issues

On the surface, the steps required to initiate and complete a randomized field trial are no different than the steps followed for any form of education inquiry. For field

experiments, the critical difference is that attention is paid to the identification of factors that interfere with the ability to produce trustworthy knowledge claims about educationally important cause and effect relations.

The field of education is almost always politicized and the various constituents interested in education will be differently invested in the intervention that is under investigation. When working outside of laboratory settings, and across multiple sites, the range of logistical and political complexities (and these two set of challenges are often intertwined) encountered as investigators aim to conduct studies of the unpredictable, and often uncontrollable settings present as a special set of challenges. The existence of such issues, and others, means that investigators must take special care in the planning and implementation of randomized of field experiments.

Highlighting the complexity of field experiments, Boruch and Wothke (1985) described seven types of plans that investigators should consider as they aim to implement a randomized field trial, noting that “[n]o single plan or strategy will suffice given the diversity of programs and stakeholders involved in experimentation” (Boruch and Wothke, 1985: 99). Their planning schema specified that plans should be developed to:

1. incentivize organizational entities to participate in the study;
2. conduct a pilot study to test and improve upon all aspects of the design and implementation of the study;
3. ensure effective implementation of the design and monitor the quality of study procedures over the course of an investigation;
4. negotiate randomization plans and collaborate with the organizations;
5. develop alternative plans as a back-up plan in the event that study plans do not proceed as expected;
6. specify the exact conditions under which Tx and control groups will function and identify preexisting conditions that facilitate randomization; and
7. ascertain how the technical aspects of local conditions (e.g., school organization and school district administrative structures) may influence design decisions such as the level of randomization (e.g., individual level, classroom, and school level).

For a more extensive discussion of planning issues in randomized field experiments the reader is directed to Boruch and Wortke (1985).

Implementation Measures, Blocking, Design Effects, and Cluster Sampling

The task of randomly assigned entities to study conditions is an important strategy of field experimentation. It is, however, not the only strategy that may be used to enhance the

quality of inferences drawn from a field experiment. There are additional tactics and which, when used as part of randomization, can enhance the quality of inferences. Second, the tactic of blocking, which involves matching prior to randomization, will be introduced. Third, the existence of amplified levels of homogeneity observed within comparison groups (Tx or control) can distort estimates of effect (and decrease power; see Cohen, 1988). Fourth, settling questions about the most appropriate level of randomization introduces a special set of considerations about power and about having unit of assignment aligned with the unit of analysis. Each of three topics is reviewed briefly below.

Implementation Measures and the Fidelity of Interventions

The validity of claims made about the comparative effect of an intervention condition and a control condition depends largely on features of study design. There is, however, a basic assumption that must be questioned, one that is foundational and chronologically prior to the empirical study of intervention effects. The assumption concerns the distinctiveness and consistency of conditions features that allow one to characterize and distinguish between the intervention and control condition. The fact that one set of classroom conditions, for example, may be designated as the intervention and another set may be designated as the control provides no guarantee that the two conditions will be as different as the investigator hoped. In most studies that aim to compare one set of educational conditions to another, a good deal of variation can be found not just between conditions but also within. The fact that such variation exists should compel investigators to include implementation measures, both for the intervention condition and the control conditions. The inclusion of an implementation measure will help the investigator explain sources of variation within conditions, across sites, and or over time.

Blocking Prior to Randomization

The most important feature of field experiments involves the random assignment at some specified scale (e.g., individuals, classrooms, schools, and school districts) to a particular study condition. A second feature involves the use of a strategy to construct groups that possess similar characteristics of an attribute of interest, where an attribute of interest is known to, or hypothesized to, exert an influence on outcome variables. This useful strategy, referred to as blocking or matching, is important given that random assignment always provides an imperfect guarantee of true random allocation. Blocking involves the creation of subgroups constructed along dimensions that are thought to be of consequence for a particular outcome variable or set of outcome variables. Rather than leaving the distribution of

such characteristics, between study groups, blocking attempts to create reasonably equivalent groups (prior to randomization), at least for the set of characteristics that are believed to be connected to the outcome. Random assignment is still used to distribute other, unknown sets of impactful characteristics evenly between groups.

Cluster Sampling and Unit Congruence

Randomization may be applied at varying scales. Within educational settings, individuals, classrooms, schools, school districts, or clusters of larger sizes may be assigned to different study conditions. The decision about the appropriate scale of randomization is a function of three issues. The first issue, theoretical rationale, concerns the explanatory framework upon which the study and its questions are based – the question to consider here relates to the level at which the causal agent of the intervention is meant to function. The second issue, implementation constraints, concerns the conditions in the field. Here, one must confront feasibility issues and consider the issues related to contamination between study conditions.

There are two key points to bear in mind as one designs field experiments with the notion of clustering in mind. First, the decision about the level of randomization has a direct bearing on the unit of analysis. To make inferences at the group level when units were randomized individuals produces faulty inferences. This illustrates what Robinson (1950) referred to as the ecological fallacy. Valid analysis of field experiments requires that the unit of randomization be matched with unit of analysis. Second, the decision to sample clusters or groups, rather than individuals, introduces a special set of considerations that affect decisions related to power analysis (i.e., sample size requirements related to expected effects). It is important to know that the number of clusters has a more important bearing on power than does the number of individuals contained within groups (see Raudenbush and Spybrook, 1997).

Intraclass Correlations and Design Effects

It is long been postulated that individuals or entities that reside within particular locations (e.g., schools and school districts) are more likely to be similar to one another than are the groups or individuals observed between groups (see Kish, 1965). The problem of differences found between groups versus within groups is referred to as the intraclass cluster correlation (ICC). When the ICC is not taken into account, a sufficiently large ICC may produce two problems. First, a large ICC can produce inflated estimates of an intervention's effect. The problem of an inflated estimate is what Kish (1965) referred to as a design effect. Second, an inflated ICC introduces requirements related to sample size. The larger the ICC,

the lower the power. Understanding of the consequences of ICC into sampling decision can ensure that field experiments are not underpowered because of design effects. As part of the planning process of field experiments, it is helpful to consider the range of ICCs that are typically found in connection with a particular class of outcomes (see Hedges and Hedberg, 2007).

Conclusion

In 1971, the Social Science Research Council assembled a committee to examine the potential benefits and design feature of social experiments in the field of education (see Reicken and Boruch, 1974). The foreword to the report offered “a comprehensive statement of the promise and problem of social experimentation” (Sheldon, 1974: ix). Field experimentation is part of a larger evolving set of design decisions, analytic options connected to the process of inquiry as Sheldon noted, “experimentation is viewed as a cycle that begins with problem analysis; proceeds through the planning of an intervention, its development, experimental trial, and evaluation; and ends in either program implementation or in re-planning the intervention” (Sheldon, 1974: x). In the conclusion to her foreword to the report, Sheldon observed that “[s]ocial experimentation may develop into an extremely valuable tool for decision making or it may prove to be too unwieldy in dealing with complex social problems” (Sheldon, 1974: xi). Lipsey (1990: 175) argued that experimental design “. . . is one of our most probing and cogent tools for identifying the means by which people problems can be ameliorated and our collective quality of life improved.” With an emerging corpus of work that employs randomized field experiments the field of education may now be in the position to examine the potential benefits noted by Reicken and Boruch nearly four decades ago.

Bibliography

- Bloom, H. (2003). Using “short” interrupted time series analysis to measure impacts of whole school reform. *Evaluation Review* **27**, 3–49.
- Boruch, R., DeMoya, D., and Snyder, B. (2002). The importance of randomized field trials in education and related areas. In Mosteller, F. and Boruch, R. (eds.) *Evidence Matters: Randomized Trials in Education Research*, pp 50–79. Washington, DC: Brookings Institution Press.
- Boruch, R. and Foley, E. (2000). The honestly experimental society: Sites and other entities as units of allocation and analysis in randomized trials. In Bickman, L. (ed.) *Validity and Social Experimentation: Donald Campbell's Legacy*, vol. 1, pp 193–238. Thousand Oak, CA: Sage.
- Boruch, R. and Wothke, W. (eds.) (1985). Seven kinds of randomization plans for designing field experiments. In *Randomization and Field Experimentation: New Directions for Program Evaluation*, vol. 28, No 5, pp 95–113. San Francisco, CA: Jossey-Bass.

- Campbell, D. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin* **54**, 297–312.
- Campbell, D. and Stanley, J. (1963). *Experimental and quasi-experimental designs for Research*. Belmont, CA: Wadsworth.
- Cohen, J. (1988). *Statistical power for the behavioral sciences*, 2nd edn. Mahwah, NJ: Erlbaum.
- Constat, M. (2007). Reshaping the methodological identity of education research: Early signs of the impact of federal policy. *Evaluation Review* **31**, 391–399.
- Cook, T. (2001). Sciencephobia. *Education Next* (Fall), pp 63–68.
- Cook, T. and Campbell, D. (1979). *Quasi-Experimental Design: Design and Analysis Issues for Field Settings*. Chicago, IL: Rand McNally.
- Dewey, J. (1938). *Logic: The Theory of Inquiry*. New York: Henry Holt and Company.
- Donner, A. (1988). Some aspects of the design and analysis of cluster randomization trials. *Applied Statistics* **47**, 95–113.
- Fisher, R. A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture* **33**, 503–515.
- Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.
- Hedges, L. and Hedberg, E. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis* **29**, 60–87.
- Kish, L. (1965). *Survey Sampling*. New York: Wiley.
- Lewis, D. (1973). Causation. *Journal of Philosophy* **70**, 556–567.
- Lipsey, M. (1990). *Design Sensitivity: Statistical Power for Experimental Research*. Newbury Park, CA: Sage.
- Mackie, J. (1974). *The Cement of the Universe: A Study of Causation*. Oxford: Oxford University Press.
- Peirce, C. (1905). What pragmatism is. *Monist* **15**, 161–181.
- Popper, K. (1996). *The Myth of the Framework: In Defence of Science and Rationality*. New York: Routledge.
- Raudenbush, S. and Spybrook, J. (1997). Strategies for improving precision in group-randomized experiments. *Educational Evaluation and Policy Analysis* **29**, 5–29.
- Reicken, H. and Boruch, R. (eds.) (1974). *Social Experimentation: A Method for Planning and Evaluating Social Programs*. New York: Academic Press.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review* **15**, 351–357.
- Rosenbaum, P. and Rubin, D. (1984). Reducing bias in observational studies using sub-classification in the propensity score. *Journal of the American Statistical Association* **79**, 516–524.
- Schneider, B., Carnoy, M., Schmidt, W., and Shavelson, R. (2007). *Estimating Causal Effects Using Experimental and Observational Designs*. Washington, DC: American Educational Research Association.
- Shadish, W., Cook, T., and Campbell, D. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin.
- Sheldon, E. (1974). Forward. In Reicken, H. and Boruch, R. (eds.) *Social Experimentation: A Method for Planning and Evaluating Social Programs*, pp ix–xi. New York: Academic Press.
- Thistlewaite, D. and Campbell, D. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology* **51**, 309–317.

Further Reading

- Boruch, R. (2005). Better evaluation for evidence-based policy: Place randomized trials in education, criminology, and social welfare. *Annals of the American Academy of Political and Social Science* **599**, 6–18.
- Cook, T. (2001). Why education researchers reject randomized experiments. *Next Fall* 63–68.
- Cook, T. (2005). Emergent design principles for the design implementation, and analysis of cluster-based experiments on social sciences. *Annals of the American Academy of Political and Social Science* **599**, 176–198.

Internal Validity

M H Clark and S C Middleton, Southern Illinois University, Carbondale, Carbondale, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Additive and interactive threats to internal validity – When the impact of a threat can be compounded by – or may depend on the level of – another threat.

Ambiguous temporal precedence – Inability to determine which variable occurred first, thereby, preventing the researcher from knowing which variable is the cause and which is the effect.

Attrition – When units drop out of the experiment after being assigned to treatment conditions.

Causal inference – Making a conclusion which claims that one variable or event causes another.

Construct – Concept or broad characteristic (i.e., intelligence or achievement).

Construct validity – How well researchers can infer that the research operations sufficiently represent constructs of interest.

Control group – Group in an experiment that has not been manipulated or did not receive an intervention.

Counterfactual – What would have occurred to the treatment group if they had not received the intervention.

Covariate – Extraneous variable that is related to the effect, but is not the cause of interest.

Covariate adjustment – Statistical correction that uses either an analysis of covariance or a multiple regression to account for a covariate to provide a clearer relationship between the cause and effect.

Cross-sectional – When data used in a study are measured all at once on a single occasion.

External validity – How well researchers can generalize causal relationships across units, treatments, observations, and settings.

History – When events occur simultaneously with the treatment and could affect the outcome.

Hot-decking – Data-imputation procedure that matches participants with missing data with those without missing data and uses the same available data value from the matched participant to substitute the missing value.

Instrumentation – When changes in the instrument used to measure responses provide different results for the outcome, appearing as if an effect occurred.

Intent-to-treat analysis – When people in the treatment condition who drop out of a study are assessed as if they had received the treatment.

Internal validity – How well researchers can make causal inferences with regard to the relationship between presumed causes and effects.

Last value forward – Procedure that uses the last available measure to replace a missing data value.

Matching – Pairing participants together based on similarity of a continuous covariate related to the outcome (i.e., pairing a control participant with a treatment participant of the same age).

Maturation – When a natural development over time occurs during the same time as the treatment and could affect the outcome.

Multiple imputation – Data-imputation procedure that predicts missing data values from the available data multiple times, statistical analyses are conducted on all data for each set of predicted values, and the multiple analyses are averaged to find a single estimate.

Operations – Specific methods used to measure variables of interest (i.e., using the Stanford–Binet to measure intelligence).

Posttest – Measure of an outcome (presumably an effect) made after a cause or intervention.

Pretest – Measure of an outcome made before a cause or intervention.

Propensity score – Conditional probability that a unit will be in a treatment condition given a set of observed covariates.

Random assignment – Assigning participants to experimental conditions (i.e., treatment or control groups) so that each participant has a known (and sometimes equal) probability of getting into each condition.

Regression to the mean – When participants are selected from their extreme scores, but have less extreme scores on other measures, appearing as if an effect occurred.

Selection – When systematic differences between unit characteristics in each treatment condition could affect the effect due to a nonrandom assignment procedure.

Single regression imputation – Data imputation procedure that predicts missing data values from the

available data using a regression equation; in a stochastic regression model, random error is added to the predicted value.

Statistical conclusion validity – How well researchers can conclude that cause and effect variables covary.

Stratifying – Grouping participants together based on similarity on a categorical covariate related to the outcome (i.e., grouping control participants with treatment participants in the same grade level).

Testing – When repeatedly exposing units to a test may permit them to learn the test, appearing as if a treatment effect occurred.

Threats to internal validity – Common reasons why researchers may be incorrect about the causal inferences they draw from experiments.

Treatment group – Group in an experiment that has received some sort of manipulation or intervention.

Weighting – Multiplying values (or weights) to individual outcome observations so that some observations contribute more or less to the overall effect; this is usually done as a statistical correction to adjust a flawed design.

Internal validity addresses how well researchers can make causal inferences about the relationship between predictive variables, presumed causes, and outcomes, their effects. Campbell and Stanley (1963) created a validity typology to provide a logical and objective way to evaluate the quality of causal inferences made when using randomized and quasi-experimental designs. Originally, they described only two types of validity: internal validity and external validity. The latter addresses how well causal relationships generalize across units (people), treatments, observations, and settings. Cook and Campbell (1979) later added statistical conclusion validity, which addresses the ability to establish that the cause and effect covary, and construct validity, which addresses the ability to infer that the research operations sufficiently represent constructs of interest.

Threats to Internal Validity

As part of their validity typology, Campbell and colleagues included threats to validity, common reasons within each type of validity that explain why researchers may be incorrect about the causal inferences they draw from experiments. Although Campbell and Stanley were the first to describe the threats, since then, Shadish *et al.* (2002) have added to and refined the threats to each of the validity typologies. Of the four types of validity,

threats to internal validity are most crucial in making causal claims from experiments. The threats to internal validity are described below.

Ambiguous Temporal Precedence

This is the inability to determine which variable occurred first, thereby preventing the researcher from knowing which variable is the cause and which is the effect. This commonly occurs when the cause is not manipulated and researchers measure the cause and effect on a single occasion. For example, Benner and Mistry (2007) used a longitudinal study to examine the influence of adult (i.e., mother and teacher) educational expectations on the academic achievement of low-socioeconomic-status students. According to the proposed model, adult expectations will cause the youth to form beliefs about expectations and competency in school which will lead to the youth's outcomes. Although the authors found direct relationships between both mother and teacher expectations and youth outcomes, they conceded that their study might suffer from ambiguous temporal precedence with the teacher-assessed variable. The data were cross sectional; therefore, it was not possible to know which variable, teacher influence or academic motivation, occurred first. Ambiguous temporal precedence can be avoided by using (1) cause and effect variables in which the order of occurrence is known, (2) designs that require that the cause be manipulated, or (3) a design that measures the cause and effect on two separate occasions – such as in a Cross Lagged Panel design described later.

Selection

This is when systematic differences between unit characteristics in each treatment condition could affect the effect due to a nonrandom assignment procedure – due to either self-selection or a criterion-based assignment. Selection most commonly occurs when it is not feasible or ethical to randomly assign participants to treatment conditions. For example, Gormley *et al.* (2005) investigated the effect of a pre-kindergarten program on academic success in the first grade. Rather than randomly assigning children to conditions, parents chose whether or not to enrol their children into the program. However, it is unlikely that children whose parents enrolled them in the program were identical to the children who were not enrolled. If the children in the treatment group were systematically different (i.e., raised by parents with high educational values) than children in the control group before the intervention, it is difficult to conclude that the significant effects found by the researchers can be attributed to the prekindergarten program. When selection is a plausible threat to validity, researchers can assess

the likelihood of its influence on an effect by using a pretest–posttest nonequivalent control-group design described below. To correct for the bias attributed to selection, a variety of statistical corrections can be made to account for the bias – assuming that the factors that influence the selection bias are known and measured.

History

This is when events occur simultaneously with the treatment and could affect the outcome. This is a concern in any field research study using a within-subjects design, since people cannot usually be isolated from outside influences that may affect the outcome. One way to determine whether or not history is a plausible threat is to add a control group to a within-subjects design, such as in a pretest–posttest control-group design. For example, Trautwein *et al.* (2008) investigated the relationship between physical self-concept and physical activity in children. Although all students showed a decrease in physical concept over a 15-month period, students in classes with high class-averages of physical ability tended to demonstrate a greater decline in physical self-concept over time. However, they could not rule out the possibility that other events or situations, such as a competitive sports season or parental demands, could have contributed to self-concept. Any additional, unmeasured variables that occurred concurrently with the physical education classes provide plausible evidence for history. To their credit, the researchers used a variety of methods to reduce threats to internal validity. They compared several levels of the treatment (class-average of physical activity), allowing for comparison groups; and they statistically controlled for several covariates that might have affected self-concept.

Maturation

This is when a natural development over time occurs during the same timeframe as the treatment and could affect the outcome. This is a similar problem as history, but is most likely to occur among any group of participants that will likely mature relatively quickly over the time that within-group measures are taken. Typically, maturation can be assessed employing the same methods used for history. However, an excellent example of controlling for maturation can be seen in the Santor *et al.* (2007) study on school-based interventions designed to help boys and girls with behavioral problems. When using different age groups to compare treatment and control groups, it can be difficult to know whether the results are derived from the intervention or are normal developmental changes. In this case, the researchers attempted to account for potential maturation by giving the intervention to

eighth graders and using seventh and ninth graders as the control groups. If they had found a reduction in behavioral problems for both eighth- and ninth-grade students, then maturation would have been a likely threat. However, because the effect was only found for the eighth graders, maturation could be ruled out as a threat.

Regression to the Mean (also Referred to as Regression or Statistical Regression)

This occurs when participants are selected based on their extreme scores, but have less extreme scores on other measures, appearing as if an effect occurred. This most often occurs when researchers use unreliable measures or heterogeneous units to match participants' treatment and control scores. This can occur either when the control score is a pretest measure taken from the same participant or if it is an outcome measure taken from another participant in a control group who is matched to the treatment participant using a matching variable. For example, Lau-Barraco and Dunn (2008) examined alcohol-treatment interventions among college students. They compared a positive expectancy challenge intervention to a traditional alcohol educational program and a control group with no intervention using expectancies in alcohol consumption and drinking behaviors as the outcomes. Individuals with extremely high pretest scores were most likely to demonstrate a difference in the outcomes. Since there were several individuals with drinking problems, the results were biased in favor of the intervention. In this case, regression to the mean could have been reduced by using participants whose outcome scores were more stable. In other cases, regression to the mean can be avoided by using more reliable measures or an aggregate of several measures to match treatment observations to control observations.

Attrition

Attrition occurs when units drop out of the experiment after being assigned to treatment conditions. This is less problematic if they drop out randomly (i.e., the reason they drop out is not related to the experiment) than nonrandomly. However, if participants who drop out of the experiment are systematically different in their responses than those who remain, this type of attrition – known as differential attrition – is a considerable threat to validity. An example of attrition is seen in DeNeui's (2003) study investigating the relationships between personality and participation in campus activities and sense of community. Although 314 individuals completed the first set of measures at the beginning of the school year, only 120 individuals completed the second set of measures at the end of the school year. Because those who

dropped out scored differently on two of the personality measures than those who remained in the study, differential attrition was evident. These differences might have led to his conclusion that students' sense of community did not increase over the school year.

Testing

Testing occurs when repeatedly exposing units to a test may permit them to learn the test, appearing as if a treatment effect occurred. This is likely to occur when an outcome is measured more than once from the same participants. For example, Powers Rock (1999) assessed the effects of coaching on the SAT I: Reasoning test. For the participants who had taken the Preliminary SAT/National Merit Scholarship Qualifying Test (PSAT/NMSQT) or the SAT more than once, Powers assessed the change in scores. Even after accounting for selection into conditions, they noted that many of those in the control group demonstrated an increase in SAT scores from previous testing. Therefore, it is quite possible that some students in both the treatment and control groups had benefited from taking previous or practice tests. Although testing can be reduced by increasing the amount of time between measures, a Solomon four-group design provides evidence as to whether or not testing has occurred.

Instrumentation

This occurs when changes in the instrument used to measure responses provide different results for the outcome, appearing as if an effect occurred. This often occurs when observations are taken over an extended period of time and a measure is revised. For example, Pettapiece (2007) used data that had been collected from a clinical sample over several years to assess growing behavioral problems in adolescents. When data collection began, clinicians were using the Personality Inventory for Children (PIC) to measure behavior. However, in 2001, the PIC-2 was published, and the clinics from which she obtained her sample replaced the PIC with the PIC-2. Therefore, part of her sample was measured using the PIC and part was measured using the PIC-2. If she had found an effect by comparing participants measured with the PIC to those measured with the PIC-2, the apparent effect may have been due to the change in the instrument, not due to an intervention. Although instrumentation can be avoided by using the same measure over time, this increases the likelihood of testing. One way to determine if instrumentation may be a threat to internal validity is to use the same methods used to test concurrent validity. Both the current and previous measures are given to the same participants at the same time (in a counterbalanced

order to reduce testing). The correlation between the measures should be close to one if the measures are giving the same results.

Additive and Interactive Threats to Internal Validity

These occur when the impact of a threat can be compounded by, or may depend on the level of, another threat. For example, Moss and Yeaton (2006) evaluated a program intended to improve college students' English-composition skills. To serve students who needed help the most, only those with low scores on a placement test were assigned to the treatment program. As a result of being in the treatment group, these students delayed enrolling in the required English courses. Students with high test scores were assigned to the control group and could enroll in the required English courses immediately. The effectiveness of the program was determined by comparing the grades from the required English courses between the treatment and control groups. Because those in the treatment group would have earned their grades a semester later than those in the control group, it is possible that other factors could have influenced the grades the treatment group received. For instance, having had an additional year in college could have permitted the treatment group to acclimate to college expectations and testing situations, giving the treatment group an advantage. This could have created an interaction between selection and maturation. History could have also added to these threats if students also learned how to improve their compositions skills through writing term papers for other classes.

Design Features

To reduce the risk of a threat occurring or to diagnose its presence and impact on internal validity, researchers can add certain features to basic designs to improve the validity of causal inferences made from experiments. These design features include (1) adding observations before (pretests) or after (posttests) a treatment to examine trends over time; (2) adding more than one treatment or comparison group to serve as a source of inference about the counterfactual (what would have occurred to the treatment group if they had not received the treatment); (3) varying the type of treatment, such as removing or varying a treatment; and (4) in the event that random assignment to conditions is not possible, using nonrandomized assignment methods that the researcher can control, such as a regression discontinuity design. Most experiments use one or more of these design features. However, the more elements that are included in designs, the more likely that researchers can diagnose or minimize the plausibility of threats to validity.

Designs that Improve Internal Validity

A pretest–posttest control group design assesses the likelihood that history and maturation are plausible threats to internal validity. Participants are randomly assigned to treatment conditions and all participants are measured at pretest, in which the outcome (or effect) is observed prior to an intervention (cause), and again at posttest, in which the outcome is measured after the intervention. The design can be diagrammed as:

$$\begin{array}{c} RO_1 XO_2 \\ RO_1 O_2 \end{array}$$

in which *R* indicates that participants were randomly assigned to conditions, *O*₁ is the pretest observation, *X* indicates that an intervention was administered, and *O*₂ is the posttest observation.

A two-factor, mixed-design analysis of variance (ANOVA) can be used to assess a treatment effect. A treatment effect is evident when there is a significant interaction between time of observation and treatment condition. If the intervention is responsible for the outcome, there will be a change in scores from pretest to posttest for the treatment group, but not for the control group. If there are changes in scores from pretest to posttest for both groups, then history and maturation are likely threats to internal validity.

Olusi (2008) used this design to study secondary school students' use of computers to solve mathematic problems. Participants were randomly assigned to either the experimental or the control group. The experimental group was taught mathematics using a computer-aided learning intervention and the control group was taught mathematics using their traditional method of instruction. All participants took a pretest and posttest of mathematic ability, allowing the researcher to assess how both groups changed over time. The treatment group had a greater increase in scores from pretest to posttest than the control group, indicating a treatment effect for the intervention.

A pretest–posttest non-equivalent control group design assesses the likelihood that selection is a plausible threat to validity. In this design, participants are not randomly assigned to treatment conditions nor do researchers know the mechanism for assignment; in many cases, participants self-select into conditions. Like in the previous design, participants are measured at both pretest and posttest. The design can be diagrammed as:

$$\begin{array}{c} NRO_1 XO_2 \\ NRO_1 O_2 \end{array}$$

in which *NR* indicates that participants were not randomly assigned to conditions.

The ANOVA used with the previous design is also used with this design. A significant difference between pretest scores indicates that selection is a likely threat.

If treatment and control groups are not equivalent before the intervention, any differences at posttest cannot be interpreted as an effect. To adjust for the threat to selection, a simple – but not always effective – solution is to use an analysis of covariance (ANCOVA), in which the dependent variable is the posttest scores and the covariate is the pretest scores. This analysis controls for the bias of the pretest scores; however, it does not account for any unobserved bias, variance that is not measured by the pretest scores, but which may contribute to the posttest scores.

McGarvey *et al.* (2004) used this design to study obese children of preschool age. The investigators had access to two similar state clinics that provided nutritional information and services to preschoolers. Rather than being randomly assigned, all of the participants in one clinic were assigned to the treatment condition and all of the participants from the other clinic were assigned to the control group. The treatment group received an intervention that focused on increasing healthy activities and the control group received standard clinical services. Both groups were administered batteries of pretests and posttests that measured daily activities, fruit and vegetable consumption, and physical activity. Because the treatment and control groups did not differ on any of the outcome measures (or most of the covariates) at pretest, selection was not likely to have threatened internal validity at posttest.

A solomon four-group design assesses the likelihood that testing is a threat to validity. In this design, participants are randomly assigned to one of four conditions: (1) treatment with pretest, (2) control with pretest, (3) treatment without pretest, and (4) control without pretest. The design can be diagrammed as:

$$\begin{array}{c} RO_1 XO_2 \\ RO_1 O_2 \\ RXO_2 \\ RO_2 \end{array}$$

A two-factor ANOVA is used to assess (1) a treatment effect – whether posttest scores from the treatment group were different from those in the control group; (2) a testing effect, whether the groups who received the pretest had posttest scores that were different from those who did not; and (3) whether taking a pretest interacts with treatment effects. A significant interaction between treatment and pretest suggests that exposure to the pretest may influence the treatment effect. If there is not an interaction, a significant main effect for treatment indicates that the treatment had an effect, and a significant main effect for the pretest indicates that pretest may influence posttest.

Kvalem *et al.* (1996) used this design to study condom use among high school students. Researchers randomly assigned 124 classes of students to one of the four groups described above. One of the treatment groups and one of the control groups were administered a pretest consisting

of a measure assessing behavioral activities. Two groups received an intervention in which the teens formed solutions to cope with negative behavior, and all of the participants were administered two posttests, 6 and 12 months after the intervention. At the 6-month posttest, researchers found that the group that received both the pretest and the intervention differed from the other three groups. Because there was no difference between the two control groups, the pretest alone did not contribute to the effect. However, because there was no difference between the two posttest-only groups, the treatment alone did not cause an effect.

A regression discontinuity design is a design that may be used to avoid selection, although it does not control or test for it. Participants are assigned to conditions by an assignment variable, which does not need to be related to the effect, but it must be continuous. A cut-off value, usually the mean of the assignment variable, is determined. Those who score above the cut-off are assigned to one treatment condition and those scoring below the cut-off are assigned to the other condition.

An ANCOVA is used to account for the relationship between the assignment variable and the outcome and provides an unbiased estimate of the treatment effect. If the treatment effect is statistically significant after accounting for the assignment variable, we can safely assume that the treatment is making its own contribution to the outcome without being confounded by other variables. Although this design still requires that the researchers control assignment, it is easier to account for variables that may influence participants' selection into conditions.

Moss and Yeaton (2006) used this design to assess the impact of developmental English programs within a college setting. Because their intervention was meant to improve English composition, researchers believed that students with deficiencies in composition would benefit most from this program. Therefore, they used English placement-test scores as the assignment variable. Those who scored below the university's score requirement were placed in the treatment group and those who scored at or above the requirement were placed in the control group. The treatment group took a developmental course in English composition before enrolling in the required English

courses, and those in the control group enrolled in the required English courses without additional assistance. The effectiveness of the program was determined by comparing the course grades between the treatment and control group, while accounting for the placement test scores.

A cross lag panel design assesses the likelihood that ambiguous temporal precedence is a threat to validity. This design is used when participants are not randomly assigned to conditions and the cause is not controlled by the researchers; therefore, it is not clear whether or not the cause preceded the effect. Both the cause and the effect are measured at two different time periods. In some cases, there is not a clear treatment or control group; instead, the cause may be the extent or strength of a characteristic and is measured by a continuous variable rather than a dichotomous one (Figure 1).

The relationships designated by the arrows can be measured by a path analysis or a series of correlations. If the cause did precede the effect, then the path coefficient (or correlation) between the cause at Time 1 and effect at Time 2 should be larger than the path coefficient between the effect at Time 1 and cause at Time 2.

Van de gaer *et al.* (2008) used this design to assess how class participation affected achievement in mathematics. Students were measured on four occasions as they progressed from the seventh grade through the twelfth grade. Path analyses (structural equation modeling) were used to predict achievement from participation and vice versa. As expected, researchers found a stronger relationship between the two variables when participation was modeled as the cause and achievement was the effect than when achievement was modeled as the cause and participation was the effect. Therefore, the researchers could reasonably conclude that participation preceded achievement in mathematics.

Statistical Adjustments

The best way to ensure internal validity is by using good research designs. However, there are cases in which using good designs may be unfeasible, either because certain assignment methods are not ethical or participants are

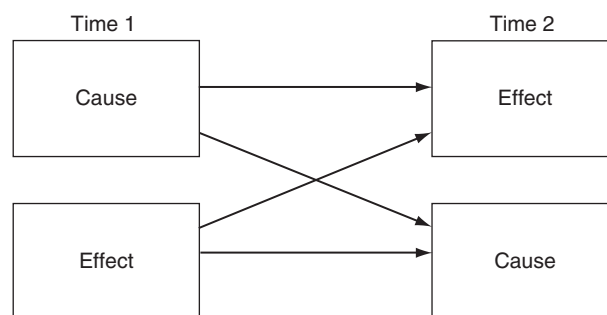


Figure 1 A Cross-Lag Panel Design.

unwilling to provide additional measures. One alternative when forced with a weak design is to statistically adjust outcomes. Common adjustments when selection is a threat include matching, stratifying, covariate adjustment, or weighting by a single variable, multiple variables, or an aggregate score of several variables, such as a propensity score. Ideally, the variables used for the adjustment should be related to both selection into conditions and the outcome of interest.

Because attrition can be difficult to control through designs, researchers must often correct for attrition using statistical adjustments. Many researchers simply exclude participants who fail to complete the treatment or the measures from analyses. While this is the simplest method and makes the fewest assumptions about unknown outcome values, dropping participants reduces statistical conclusion validity (i.e., statistical power) and external validity. Some researchers prefer to keep the participants with missing data, but assume that outcomes are either zero or do not change from pretest to posttest. A last-value-forward procedure uses the last available outcome measure for the subsequent, unobserved measure. In an intent-to-treat analysis, researchers analyze those assigned to the treatment conditions, but who dropped out, as if they had actually received the treatment. Both of these methods tend to underestimate true treatment effects, thereby affecting statistical conclusion validity. Another way to retain subjects with missing values is to impute missing data, in which researchers attempt to predict the missing values based on the available data. There are several methods for doing this (i.e., hot-decking, multiple imputation, and single regression imputation); however, all of these methods assume that data are missing at random – that is the values that are missing are not systematically different from the values that are present.

While statistical adjustments have been shown to reduce some of the threats to internal validity, each of these methods have their limitations and some may even reduce other types of validity in an attempt to control for internal validity. Therefore, researchers should always be encouraged to use good designs over elaborate statistics whenever possible.

Bibliography

- Benner, A. D. and Mistry, R. S. (2007). Congruence of mother and teacher educational expectations and low-income youth's academic competence. *Journal of Educational Psychology* **99**, 140–153.
- Campbell, D. T. and Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand-McNally.
- Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago, IL: Rand-McNally.
- DeNeui, D. L. C. (2003). An investigation of first-year college students' psychological sense of community on campus. *College Student Journal* **37**, 224–234.

- Gormley, W. T. Jr., Gayer, T., Phillips, D., and Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology* **41**, 872–884.
- Kvalem, I. L., Sundet, J. M., Rivø, K. I., Elertsen, D. E., and Bakketeig, L. S. (1996). The effect of sex education of adolescents' use of condoms: Applying the Solomon four-group design. *Health Education Behavior* **23**, 34–47.
- Lau-Barraco, C. and Dunn, M. E. (2008). Evaluation of a single-session expectancy challenge intervention to reduce alcohol use among college students. *Psychology of Addictive Behaviors* **22**, 168–175.
- McGarvey, E., Keller, A., Forrester, M., et al. (2004). Feasibility and benefits of a parent-focused preschool child obesity intervention. *American Journal of Public Health* **94**, 1490–1495.
- Moss, B. G. and Yeaton, W. H. (2006). Shaping policies related to developmental education: An evaluation using the regression-discontinuity design. *Educational Evaluation and Policy Analysis* **28**, 215–229.
- Olusi, F. I. (2008). Using computers to solve mathematics by junior secondary school students in Edo state Nigeria. *College Student Journal* **42**, 748–755.
- Pettapiece, J. A. (2007). Evaluation of assessment methods in the detection of over-reporting of ADHD symptoms in children. (Doctoral dissertation, Southern Illinois University, 2007). *Dissertation Abstracts International: Section B: Sciences and Engineering* **67**, 4719.
- Powers, D. E. and Rock, D. A. (1999). Effects of coaching on SAT I: Reasoning test scores. *Journal of Educational Measurement* **36**, 93–118.
- Santor, D. A., Poulin, C., LeBlanc, J. C., and Kusumakar, V. (2007). Facilitating help seeking behavior and referrals for mental health difficulties in school aged boys and girls: A school-based intervention. *Journal of Youth Adolescence* **36**, 741–752.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton-Mifflin.
- Trautwein, U., Gerlach, E., and Lüdtke, O. (2008). Athletic classmates, physical self-concept, and free-time physical activity: A longitudinal study of frame of reference effects. *Journal of Educational Psychology* **100**, 988–1001.
- Van de gaer, E., Pustjens, H., Van Damme, J., and De Munter, A. (2008). Mathematics participation and mathematics achievement across secondary school: The role of gender. *Sex Roles* **59**, 568–585.

Further Reading

- Burns, P. A. (1997). *Securing Internal Validity – How to Avoid the Threats*. <http://www.socialresearchmethods.net/tutorial/Burns/int.html> (accessed August 2009).
- Campbell, D. T. and Kenny, D. A. (1999). *A Primer on Regression Artifacts*. New York: Guilford Press.
- Garson, G. D. (2008). *Research Designs*. <http://faculty.chass.ncsu.edu/garson/PA765/design.htm> (accessed August 2009).
- Reichardt, C. S. (2008). An alternative to the Campbellian conceptualization of validity. *Technical Report*. Denver, CO: University of Denver.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.
- Shadish, W. R. and Luellen, J. K. (2006). Quasi-experimental design. In Green, J. L., Camilli, G., and Elmore, P. B. (eds.) *Complementary Methods for Research in Education*, 3rd edn., pp 539–550. Washington, DC: American Educational Research Association.
- Sinharay, S., Stern, H. S., and Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods* **6**, 317–329.
- Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin* **46**, 137–150.
- Trochim, W. M. K. (2006). Internal validity. In *Research Methods Knowledge Base*. <http://www.socialresearchmethods.net/kb/intval.php> (accessed August 2009).
- Trochim, W. M. K. and Donnelly, J. P. (2007). *The Research Methods Knowledge Base*, 3rd edn. Mason, OH: Atomic Dog Publishing.

Methods for Approximating Random Assignment

B Schneider, Michigan State University, Lansing, MI, USA

S-K McDonald, NORC at the University of Chicago, Chicago, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Counterfactual – What would have happened to the treated in the absence of the intervention.

Fixed – Time invariant.

Quasi-experiment – Investigation with experimental treatments or manipulations but without random assignment of subjects to the comparison conditions from which the impact of the intervention is inferred.

Selection bias – Probability that differences among treatment conditions could plausibly be attributed to preexisting differences among the treated.

Random Assignment and the Search for ‘What Works’

Evidentiary claims warranting action (e.g., the adoption of an educational policy, the enactment of a curriculum, the implementation of a school reform, the provision of funds to support the development of an intervention) typically require the strongest possible indication that the action taken will result in the desired outcome. Thus, the movements to promote evidence-informed policies and to establish what does and does not work to improve educational outcomes are closely linked to efforts to strengthen the scientific basis of educational research.

Random assignment to treatment conditions is an important technique for establishing the credibility of inferences regarding causal relationships among variables. Accordingly, many view random controlled trials (RCTs) as a necessary step in forging the chain of evidence that justifies the commitment of resources beyond proof-of-concept and efficacy studies to establish effectiveness and the potential scalability of an innovation. Random assignment is not, however, the only method for providing robust evidence of the probability that some effect was produced by some cause. Reviewed here are four other methods for approximating a random assignment condition: (1) regression discontinuity (RD) designs, (2) propensity score matching (PSM) techniques, (3) instrumental variables, and (4) fixed effects models.

The Role of Random Assignment in Investigating Relationships Among Variables

Evidence regarding relationships among variables may be descriptive (recording observations of the characteristics of the phenomenon or variables of interest), associative (describing the relationship between one set of changes and another), primarily by correlational analyses (that test the strength of that relationship), or causal (indicating which produces – and which results from – the other). The ability to assert a causal relationship between an intervention (such as a reading program) and an outcome (achievement) – in a more extended form, an intervention and variations in populations of interest (e.g., race, ethnicity, and gender) and other contextual variables (e.g., average classroom academic performance) – requires robust evidence of a strong probability that the former will produce an effect (a change) in the latter.

To produce such evidence, one must simultaneously rule out or control for factors (e.g., unexpected independent variables, or unexpected differences or differential changes in the study's subjects) that might otherwise contribute to observed changes in the dependent variable. Experimental trials, in which subjects and/or their contexts are deliberately manipulated, are designed to determine the probability that the independent variable (intervention or treatment) produces (is the cause of) an observed effect.

In ruling out alternative explanations for an observed effect, it is critical to ensure that those whose futures the investigator expects to alter when he/she intervenes by administering the intervention do not differ in important ways from those in the comparison (control) condition. The latter group provides an estimate of the counterfactual – what would have happened to the treated in the absence of the intervention. For this untreated group to serve as a reasonable proxy of the counterfactual in the treatment group, the two groups must be as similar to each other as possible on the full range of (observed and unobserved) characteristics that might alter the effect of the intervention. Investigators who select dissimilar subjects into their experimental and control groups introduce selection bias into their studies, thus constraining the internal validity of their findings.

Threats to the internal validity of a study's findings undermine the experimenter's ability to rule out (or reduce to a very low probability) the possibility that factor(s) other

than the intervention account for changes in the dependent variable. When internal validity is in question, a correlation between variables may be well documented, but assumptions regarding the causal nature of the relationship cannot be adequately supported and potential confounds cannot be excluded. (Threats to external validity also affect the confidence that can be placed in the intervention – in this case, confidence that the intervention will have similar effects in other settings, e.g., when introduced with different subjects, in different contexts.)

Random assignment of subjects to treatment and control conditions by definition addresses this problem by using an unbiased selection mechanism (chance) to reduce to zero the probability that preexisting differences among those who do and do not receive the intervention will affect the outcome. This is not the same as eliminating differences between treatment and control groups (sampling imbalances); instead, randomization yields comparably unbiased groups, such that with enough replications sampling differences can be identified and effects more accurately estimated. On average, any observed differences among treatment conditions will not be attributable to any preexisting differences across the groups, provided enough subjects are randomly assigned to the various comparison conditions.

Challenges and Costs of Employing Random Assignment in Field Settings

The circumstances which make it challenging to employ and/or maintain randomization in field studies are widely acknowledged and documented in the literature. The most forceful arguments against random assignment are those with a moral or ethical dimension. Examples include (1) the absence of proof of concept (i.e., instances in which the chain of evidence does not yet establish the likely benefit of the new alternative), so that moving to an experiment is too risky; and (2) situations in which the treatment is having extreme positive or negative effects on the population, calling for early termination of the experiment (e.g., a retention program that diminishes students' attendance).

Other challenges in achieving and maintaining random assignment include: (1) problems of implementation (e.g., treatment noncompliance and attrition); (2) the likelihood that spillover, bleeding, or contamination (when features of interventions become known and reproduced and/or their impacts affect those in the control group) will result in atypical responses; (3) the ability to document any non-intervention resources available to the control group that might not have been available to the treatment group, absent the experiment (to establish accurately the true comparison condition, and how well it actually approximates the counterfactual); and (4) the resources necessary

to implement and maintain random assignment in social settings. Careful consideration of the investigation's capacity to address these challenges helps to ensure resources channeled to educational research are likely to generate an appropriate rate of return (in the form of sufficiently robust evidence). Other challenges, not unique yet important to consider in planning RCTs, include (1) the capacity to detect treatment effects (e.g., sufficient information to estimate the number of subjects required to detect meaningful intervention effects – i.e., to conduct a statistical power analysis); and (2) the ability, in longitudinal studies, to track members of the treatment and control groups over time.

The RCT is an ideal mechanism for avoiding the challenge to internal validity posed by selection bias, and while there are acknowledged challenges to implementing RCTs in educational and other settings, there are also many resources available to those committed to random assignment; see, for example, Shadish *et al.* (2002). It is important to underscore, however, that random assignment does not address other potentially confounding factors. For example, history can only be satisfactorily ruled out as the potential cause of an observed change when other conditions are systematically controlled – for example, isolating subjects from other events that could produce (or affect the extent of) an observed effect. Thus, even an experiment with random assignment may not yield results warranting causal inferences. The question is whether causal inferences can be justified based on evidence from investigations without random assignment.

Four Methods for Approximating Random Assignment

Where random assignment is not feasible or possible, alternative techniques for assigning subjects to treatment conditions should be employed to reduce selection bias, thus increase the confidence that can be placed in research evidence.

Regression Discontinuity Designs

The RD design is generally considered as the best alternative to random assignment as this is the only other design capable of yielding unbiased estimates of average treatment effects (see Cook, 2008). Importantly, RD designs can be used to assess observational data in a way that mimics pretreatment assignment to unbiased groups. For example, students apply for scholarship aid based on their grades, test scores, etc. Then a committee decides the cut-point and selects those above the mean for the scholarship. Using an RD design, one outcome of interest would be if students not selected for scholarships continued in their program of study at the same level of success as those who received the scholarship.

First applied in education by Thistlethwaite and Campbell (1960), the RD design is one of a class of quasi-experiments – that is, investigations with experimental treatments or manipulations but without random assignment of subjects to the comparison conditions from which the impact of the intervention is inferred. Lord noted “If the individuals are not assigned to the treatments at random, then it is not too helpful to demonstrate statistically that the groups after treatment show more difference than would have been expected from random assignment – unless, of course, the experimenter has special information showing that the nonrandom assignment was nevertheless random in effect” (Lord, 1967: 38; cited by Tallmadge and Horst, 1976). In the RD, this special information is that individuals at the point of discontinuity in the predicted regression line are in effect randomly assigned to the treatment and control condition.

The logic of the RD design is that subjects immediately adjacent to a cutoff score on a continuous pretest assignment variable are sufficiently similar in other respects, that those assigned to the control group provide a reasonable proxy of the counterfactual for those assigned to the treatment group. In essence, the RD design assumes such cut-points to be arbitrary, so that the populations of individuals at, immediately above, and immediately below the cut-point are fundamentally similar in all other respects than their scores on the assignment variable.

Since groups thus comprised are probabilistically equivalent, comparing the groups’ posttreatment outcomes yields an unbiased estimate of the effects of an intervention (an observed discontinuity in outcomes warranting causal inference of a treatment’s impact). Such discontinuities are observed by controlling for prior achievement so that any jump in separate regression lines for the two groups at the point of assignment (seen when posttest scores are plotted on the vertical against pretest scores on the horizontal) is interpreted as a (simple, main) treatment effect. For example, Jacob and Lefgren (2004) compare students just above with those just below the criteria for promotion from third to fourth, and from sixth to seventh grades used by the Chicago Public Schools to investigate the impacts of summer school and grade retention on subsequent academic achievement. Ludwig and Miller (2007) use variations in funding levels across areas (counties) to examine impacts of the Head Start program.

Used prospectively, RD designs assign subjects to treatment conditions on the basis of cut-scores. Used retrospectively, RD designs mine observational data sets that record outcome variables of interest (e.g., student achievement on standardized test scores) for those previously assigned to various treatment conditions (e.g., participation in a summer bridge program, in a remedial instructional program, or in an accelerated instructional program) on the basis of a specified cut-score. In either case, the treatment assignment mimics a natural experiment.

The connection between preintervention status on outcomes of interest (scores on assignment variables) and treatment status – a defining feature of RD designs (and not characteristic of random assignment) – does have potentially negative consequences. Chief here are the larger sample sizes required to establish genuine intervention effects (i.e., to distinguish chance from actual differences between the treatment and control conditions). The minimum detectable impact (commonly expressed in effect size units) that a particular design has a high probability of documenting is calculated in a statistical power analysis. Such analyses have demonstrated that the sample size required to reliably detect these real (albeit minimal) impacts is substantially larger for RD designs than for studies in which subjects are randomly assigned to treatment status. Schochet concludes, in the case of clustered (e.g., school-based) designs frequently employed in educational impact evaluations, that “[s]chool sample sizes typically need to be about three to four times larger under RD than RA [random assignment] designs to achieve impact estimates with the same level of precision,” (Schochet, 2008: 33), a consideration which may render RD designs prohibitively expensive or otherwise impractical.

Other important questions to be considered in weighing the appropriateness of RD designs include:

- Can additional evidence on the sample just at, above, and below the cut-score be provided, or can additional selection criteria (e.g., matching procedures) be employed, to ensure their comparability?
- Are those near the cutoff sufficiently similar to others on the distribution to be considered representative of a (somewhat) larger group, or sufficiently different that they should be construed as a separate population (a key issue in considering the potential generalizability of the RD study findings)?

Cook (2008: 651) observes that “causal inference is most warranted near to the cutoff,” yet notes that “if the repertoire of sensitivity tests for functional form can be routinely extended to include a pretest, a matched but nonequivalent group or, at last resort, a nonequivalent dependent variable, then better counterfactual estimates of functional form will be possible and reliance on unobserved extrapolations will be less. By moving to a difference in differences strategy for RDD, we may achieve reasonable estimates of whether a causal conclusion should be limited only to the cutoff area.”

RD designs lack the statistical power of studies with random assignment, and are generally considered theoretically inferior to the latter. Such considerations notwithstanding, RD designs are widely acknowledged to be “superior to all other known causal methods” in the “circumscribed set of circumstances where an experiment might not be feasible” (Cook, 2008: 652) described above.

Propensity Score Matching

Another way to control for potential selection bias is to examine the outcomes of those whose tendency to receive a treatment is described in terms of a propensity score that balances analytic samples of the untreated against the treated on a range of observable characteristics that predict but are unaffected by participation. Developed by Rosenbaum and Rubin, the PSM technique acknowledges that the comparison condition could contain individuals never intended to receive the treatment. Thus, population average treatment effects distort understandings of the likely impact of the intervention on its intended audience. Propensity scores establish the tendency to receive the intervention (i.e., to be included in the treatment condition) given particular – observed – characteristics. These scores can then be used to compare to the treated a subset of the untreated whose outcomes more reasonably represent the counterfactual.

The capacity to isolate comparable groups with equivalent propensities to be selected into the treatment and control conditions requires: a comprehensive assessment of relevant pretreatment characteristics; extensive information on participants and nonparticipants; and confidence that a variety of assumptions can reasonably be made. For example, the data available to the investigator must be sufficient to warrant the assumption that all variables influencing both assignment and outcomes can be observed (“unconfoundedness,” see Rosenbaum and Rubin, 1983). Outcomes “must be independent of treatment conditional on the propensity score,” the conditional independence assumption (CIA), (Caliendo and Kopeinig, 2008: 38). In addition, every subject’s (potential) outcomes must be independent of others’ treatment status – the stable unit treatment value assumption (SUTVA, see Rubin, 1980; Holland, 1986) – including peer and general equilibrium effects (see Caliendo and Kopeinig, 2008: 66).

Caliendo and Kopeinig, who provide a helpful overview of the steps required, conduct a study employing PSM, including:

1. selecting the model and specifying the variables to be employed in estimating the probability of assignment;
2. choosing an algorithm for matching based on that propensity score (e.g., nearest-neighbor matching with or without replacement);
3. checking for overlap and common support;
4. checking the quality of the match; and
5. assessing the sensitivity of the estimates of treatment effects.

They also highlight practical considerations (e.g., opportunities to participate in multiple, heterogeneous programs or treatment states simultaneously) that can bedevil efforts to implement PSM. Luellen, Shadish, and Clark provide a general introduction to (and practical

advice for undertaking) propensity score analysis, and a comparison of “three different methods for computing propensity scores on a data set” documenting “the change in observable bias associated with each method relative to a randomized experiment consisting of participants from the same population as those in the quasi-experiment” (Luellen *et al.*, 2005: 549).

PSM is ideally suited for use in situations where random assignment to a treatment condition (e.g., retention in one or promotion to another grade, see Hong and Raudenbush, 2005) would be problematic, and where sufficient information is available to construct matched groups of treated and untreated subjects with equivalent propensities to be assigned to the treatment condition. The quality of the data available to the investigator is critical for PSM to succeed in approximating random assignment, making this technique particularly suitable in analyzing large-scale observational data sets. PSM is not a desirable approach, however, to approximating random assignment where credible information supports likely relationships of unobservables to outcomes.

Instrumental Variables

When propensity score matching’s strong assumption of unconfoundedness is not justified by the data available, steps to reduce selection bias must account for unobservables (see Caliendo and Kopeinig, 2008: 35). One way to do so is to introduce into the model a variable which prior knowledge suggests (1) has no independent relationship with the outcome variable of interest, (2) does have a (at minimum, associative) relationship with the independent variable, and (3) is not correlated with other (unobservable, omitted) variables that may directly effect the outcome (dependent) variable. In other words, the new – instrumental – variable is related to the dependent variable only indirectly, through its relationship to the independent variable.

Originally developed to address problems other than selection bias, instrumental variables “solve the omitted variables problem by using only part of the variability” in the hypothesized cause “– specifically, a part that is uncorrelated with the omitted variables – to estimate the relationship between” the hypothesized cause and the effect of interest (Angrist and Krueger, 2001: 73). In the example, this quote from Angrist and Krueger references, the hypothesized cause of an effect on earnings is schooling. The two instrumental variables employed to estimate a rate of return (the dependent variable, expressed in earnings) of schooling (the exogenous independent variable, operationalized as the highest grade of school completed) were birth date and compulsory schooling laws. Birth date (expressed by quarter of birth) was judged a valid instrument for schooling as it “is probably unrelated to the person’s innate ability, motivation or family connections” – other factors that might reasonably be hypothesized to

condition the effect of the treatment on the outcome of interest (Angrist and Krueger, 2001: 74). Compulsory schooling laws determine legal dropout ages. Such policies often produce natural experiments suitable for inclusion as instrumental variables in analyses of observational data.

Analytically, instrumental variables rescale (downward) the impact of the treatment on the outcome of interest to only that portion of the potential treatment effect associated with the instrument. Reducing the estimate of the effect of the treatment by focusing only on the variability in the intervention associated with the instrument controls for the possibility that other, unobserved variables might interact with the intervention to enhance/inflate its impact on the outcome of interest. When this scaled-down estimate concurs with an analysis (e.g., ordinary least squares (OLS) regression) of the relationship between the treatment and the dependent variable, one can more confidently infer that the magnitude of the effect is genuine, (i.e., not affected by omitted unobserved variables). Conversely, when the (scaled-down) instrumental variables estimates are substantially lower than those obtained from an OLS analysis, it would appear that unobserved variables do have important impacts on the dependent variable.

As demonstrated by Angrist *et al.* (1996), the instrumental variables approach can also provide “an alternative to a more conventional intention-to-treat analysis” – for example, when random (thus ignorable) assignment does not map directly to the (therefore nonignorable) treatment received. In such cases, treatment assignment can itself be used as an instrumental variable (see, e.g., Nye *et al.*, 2004: 247). Consider the case in which implementation challenges (e.g., noncompliance) mean randomization was not achieved as intended. In such cases, the investment in the planned RCT may be at least partially recovered if the assignment (e.g., families assigned to use vouchers) is associated with the treatment (e.g., voucher use) but independent of that treatment is not associated with the outcome of interest (e.g., school achievement), the conditions necessary to undertake an instrumental variables analysis (see Hedges, 2008).

The use of instrumental variables to rule out or confirm the likely effects of unobservables is not without criticism, for example, concerns that underlying theoretical relationships need not be specified in advance; that subjects may react differentially to the instrument, making it difficult to interpret estimates with heterogeneous responses (see, e.g., Angrist and Krueger, 2001; Imbens and Angrist, 1994). These concerns notwithstanding, Angrist and Krueger conclude that “instrumental variables methods often solve the first-order problem of eliminating omitted variables bias for a well-defined population,” (Angrist and Krueger, 2001: 78). Importantly, instrumental variables can be employed to address potential selection bias arising from unobservables with cross-sectional

(e.g., posttreatment) data only. For longitudinal and other dependent group designs (e.g., pre-postintervention studies), it may be desirable to employ fixed effects models.

Fixed Effects Models

A second approach to correcting bias in estimated treatment effects associated with unobservables is to adjust for the presence of time invariant (i.e., fixed) individual characteristics “that may be associated with selection in to the treatment group,” (Schneider *et al.*, 2007: 42). Examples of such invariant characteristics include the household characteristics of siblings who do and do not receive specific interventions, for example, attend Head Start in Currie and Thomas’ (1995) analysis of the impacts of that program. By examining the impacts of siblings from the same household who did and did not attend Head Start Currie and Thomas were able to account and control for the household effects which were treated as fixed (the same) for each sibling in the household. In this way, they were able to demonstrate that those who attended Head Start outperformed those who did not, despite the overall lower-than-average achievement of children in Head Start programs (see Schneider *et al.*, 2007: 43–44).

Fixed effects or “difference-in-differences” estimators assess first the differences between pre- and postintervention outcome measures “for each participant and non-participant,” then “the difference between the average pre-versus post-program change for participants and . . . non-participants.” The resulting “estimator of program impact will be free of selection bias if selection into the program depends on unobserved person-specific ‘fixed effects.’” (Human Resources and Social Development Canada (HRSDC), 1998). Another way in which data on individuals can be used to control for unobserved characteristics is if the same individuals participate in two treatment conditions. For example, Bifulco and Ladd (2006) analyzed longitudinal data on student achievement that included data on the type of school (regular public or charter) the student attended. As the data set captured results for students who at one point attended charter, at another point a regular public school, Bifulco and Ladd were able to attribute to the type of school outcomes that might otherwise have been attributed to fixed (time-invariant) student characteristics, ultimately concluding that “students in charter schools scored significantly lower in both reading and mathematics” (see Schneider *et al.*, 2007: 44).

Successful use of fixed effects approaches to address selection bias clearly place substantial demands on the data available to the investigator. In contrast to the instrumental variables approach, modeling fixed effects requires longitudinal data (a minimum of one pretreatment and one post-treatment observation; see Human Resources and Social Development Canada (HRSDC), 1998). In addition to the data required to conduct fixed effects analyses, the suitability

of this approach is critically dependent upon the assumption that the unobservables in question are, in fact, appropriately considered fixed (time invariant) over the course of the study period. Clearly, this is an assumption that will not always hold (e.g., the career orientation and ambitions that affect selection into and success in educational programs may change over time; the maternal personality that affects “the likelihood both of being employed and of having good child outcomes” (Schneider *et al.*, 2007: 42) may similarly shift from time to time, affecting assessments of the impact of mother’s employment on their children, see Human Resources and Social Development Canada (HRSDC), 1998; Currie, 2003). In such cases, variations on the approach (e.g., the difference-in-differences in growth rates estimator, see Human Resources and Social Development Canada (HRSDC), 1998; Moffitt, 1991) may be appropriate.

Conclusion

Both philosophical and practical considerations argue against the possibility of unequivocally establishing the true case of many of the effects educational policies and practices are designed to bring about. The appropriate standard for experimental studies in education, as in many other fields of inquiry, may thus be the capacity to generate compelling evidence that there is a sufficiently high probability the intervention will produce the desired outcome – that is, to warrant causal inference.

Investigations which incorporate random assignment to treatment conditions reduce the probability that differences among treatment conditions could plausibly be attributed to preexisting differences among the treated (selection bias). Random assignment assists investigators in inferring treatment effects, by ruling out one critically important alternative interpretation of the cause of an effect. For this reason, experimental trials incorporating random assignment (RCTs) have a critically important role to play in the production of robust evidence of the likely causal relationships among variables.

RCTs are not, however, the only – indeed at times they are arguably not the optimal – method of estimating causal effects; absent other controls, random assignment to treatment conditions may not be sufficient to warrant causal inference. To acknowledge this is not to argue against RCTs but rather to underscore the importance of matching method to research question and circumstance. Whether research designs that approximate or ensure random assignment of subjects to treatment conditions are preferable to each other in studies which seek to establish causal relationships among variables is contingent upon numerous factors. Considerations include:

- Is it more appropriate/desirable to generalize to (narrower) populations at or near the value used to assign

subjects to treatments? (In these cases, the RD design could be preferable to random assignment; see Schochet (2008) and Heckman and Vytlačil (2005) regarding evidence warranting marginal expansion of the intervention.)

- What resources are available to devote to recruitment of subjects to an intervention, and how receptive are potential subjects likely to be to assignment based on status on a specific variable (e.g., achievement)? (Where resources are limited and subjects perceive a logical relationship between assignment variables and interventions, RD designs may be easier to implement than random assignment; see Schochet, 2008.)

Proponents of RCTs are not wrong to insist this design be employed as often as research questions and resources combine to make this an achievable ideal. Studies that employ RD designs, propensity score matching, instrumental variables, or fixed effects models can be useful when RCTs are not possible. These four methods are particularly helpful in leveraging the knowledge in large-scale, representative, longitudinal observational data sets, advancing the generation of robust evidence regarding causal relationships. Our understanding of the factors that improve educational attainment would grow at a much slower pace, and our ability to leverage extant data would be severely curtailed if we do not explore multiple methods for estimating causal relationships among populations and outcomes of interest, such as instructional techniques and student achievement.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants No. 0129365 and 0815295. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

See also: Causal Inference; Evaluation Research; External Validity; Regression Discontinuity Designs; Sampling.

Bibliography

- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**, 444–455.
- Angrist, J. D. and Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives* **15**, 69–85.
- Bifulco, E. and Ladd, H. F. (2006). The impacts of charter schools on student achievement: Evidence from North Carolina. *Education Finance and Policy* **1**, 50–99.
- Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* **22**, 31–72.

- Cook, T. D. (2008). "Waiting for life to arrive": A history of the regression-discontinuity design in psychology, statistics, and economics. *Journal of Econometrics* **142**, 636–654.
- Currie, J. and Thomas, D. (1995). Does Head Start make a difference? *American Economic Review* **85**, 341–364.
- Currie, J. (2003, June). *When do we really know what we think we know? Determining causality*. Invited paper presented at Work, Family, Health and Well-Being conference, NICHD Administration for Children and Families, Washington, D.C.
- Heckman, J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* **73**, 669–738.
- Hedges, L. V. (2008). Using the instrumental variables technique in educational research. Presented at *Education Research: What's Now? What's Next?* Institute of Education Sciences 3rd annual research conference. Washington, D.C. http://ies.ed.gov/director/conferences/08ies_conference/slides.asp?ppt=hedges (accessed September 2009).
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association* **81**, 945–970.
- Hong, G. and Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis* **27**, 205–224.
- Human Resources and Social Development Canada (HRSDC), (1998). Quasi-experimental evaluation. Evaluation and data development strategic policy human resources development Canada. SP-AH053E-01-98. <http://www.hrsdc.gc.ca/en/cs/sp/sdc/evaluation/sp-ah053e/page10.shtml> (accessed September 2009).
- Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica* **62**, 467–475.
- Jacob, B. and Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics* **86**, 226–244.
- Lord, F. M. (1967). Elementary models for measuring change. In Harris, C. W. (ed.) *Problems in Measuring Change*, pp 21–38. Madison, WI: University of Wisconsin Press.
- Ludwig, J. and Miller, D. (2007). Does head start improve children's life chances? Evidence from a regression discontinuity design. *Quarterly Journal of Economics* **122**, 159–208.
- Luellen, J. K., Shadish, W. R., and Clark, M. H. (2005). Propensity scores: An introduction and experimental test. *Evaluation Review* **29**, 530–558.
- Moffitt, R. (1991). Program evaluation with nonexperimental data. *Evaluation Review* **15**, 291–314.
- Nye, B., Konstantopoulos, S., and Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis* **26**, 237–257.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.
- Rubin, D. B. (1980). Comment on Basau, D. – randomization analysis of experimental data: The Fisher randomization test. *Journal of the American Statistical Association* **75**, 591–593.
- Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., and Shavelson, R. J. (2007). Estimating causal effects using experimental and observational designs. *Report from the Governing Board of the American Educational Research Association Grants Program*. Washington, DC: American Educational Research Association.
- Schochet, P. Z. (2008). Technical methods report: Statistical power for regression discontinuity designs in education evaluations (NCES 2008-4026). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <http://ies.ed.gov/ncee/pdf/20084026.pdf>.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin Company.
- Tallmadge, G. K. and Horst, D. P. (1976). A procedural guide for validating achievement gains in educational projects (Revised). *RMC Research Corporation Report Number RMC-UR-240 for the U.S. Department of Health, Education, and Welfare; Office of Education/Office of Planning, Budgeting, & Evaluation*, Los Altos, California: RMC Research Corporation.
- Thistlethwaite, D. L. and Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *Journal of Educational Psychology* **51**, 309–317.

Further Reading

- Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* **106**, 979–1014.
- Campbell, D. T. and Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Boston, MA: Houghton Mifflin Company.
- Dignam, J. (2007). From efficacy to effectiveness: Translating randomized controlled trial findings into treatment standards. In Schneider, B. and McDonald, S. -K. (eds.) *Scale-Up in Education: Vol. 1, Ideas in Principle*, pp 123–186. Lanham, MD: Rowman & Littlefield.
- Goldberger, A. S. (1972). Selection bias in evaluating treatment effects: Some formal illustrations, *Working Paper*. Madison: University of Wisconsin, Economics Department.
- Heckman, J., Ichimura, H., and Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training program. *Review of Economic Studies* **64**, 605–654.
- Jurs, S. G. and Glass, G. V. (1971). Experimental mortality. *Journal of Experimental Education* **40**, 62–66.
- Myers, D. and Dynarski, M. (2003). *Random Assignment in Program Evaluation and Intervention Research: Questions and Answers*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- National Research Council (2004). Implementing randomized field trials in education: *Report of a Workshop*. Washington, DC: National Academy Press.
- Rosenbaum, P. R. (1995). *Observational Studies*. New York: Springer.
- Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources* **33**, 127–169.

Methods for Approximating Random Assignment: Regression Discontinuity and Propensity Scores

L M Scheier, LARS Research Institute, Las Vegas, NV, USA; Washington University, School of Medicine, St. Louis, MO, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bias – A term usually referring to an unwanted or extraneous source of systematic variance affecting the standard error of a treatment effect that arises from several sources (see the definition for the glossary term ‘Internal validity’). Usually, minimized or controlled through methods of statistical adjustment or stringent experimental design procedures. Selection bias occurs when there is a different probability of a unit or individual being chosen to participate in a study or assigned to a treatment condition and the characteristics of that individual are confounded with treatment outcomes (e.g., skills, motivation, or preexisting conditions are not independent of the anticipated treatment outcome).

Causal inference – Also called causal propositions, but not the same as causal explanation or causal mechanism, it concerns the expected relationship between some event B that is anticipated to always occur when A happens. The inference or supposition of existence is based on expectation that if all other possible factors that induce a relationship between A and B have been controlled, then the absolute reason for B is always A. In terms of a manipulative account of causation based on experimental work defined as using *prima facie* and, often, probabilistic evidence to infer causal relations (A causes B when the experimenter induces A always).

Confounding – A means of describing relationships between events, where A is the manipulation and B the expected outcome from A. Confounding arises because some measure C is on the path from A to B and mixes its variability so as to “falsely obscure or accentuate the relationship between them” (MacKinnon *et al.*, 2000: 174). When these variances cannot be unambiguously tested, it is usually a sign of confounding. Controlling for the confounder, usually, provides an undistorted estimate of the true relationship between the predictor and outcome in question. Unlike mediators or intervening variables – which carry a causal connotation – confounders are not necessarily posited to cause the linkage between

predictor and outcome, (i.e., spuriousness) but are merely related to one or both.

Counterfactual reasoning – A term borrowed from eighteenth-century Scottish philosopher David Hume – a staunch advocate of empiricism in experimental methods – to indicate something that is contrary to fact. In experimental work, a treatment causes an observed effect or alteration on some designated outcome; the counterfactual is what would have happened to the subjects in the absence of a treatment effect (the subjects were designated controls and did not receive the treatment).

A counterfactual inference (based on qualitative distinction) is made experimentally when a treatment condition is pitted against a control group and all other potential factors that might contribute to group differences are made identical between the two groups through statistical control.

Effect – Defined as the difference between what did happen to a treated individual (unit) and what would have happened had the treatment not worked. Requires a counterfactual model in which comparisons exist between a group of individuals exposed to the treatment (observed result) compared to a reasonably similar group held back from receiving the same treatment (unobserved result). A stable phenomenon that scientists seek to explain in terms of behavioral regularities brought about through manipulation.

Internal validity – A system of checks and balances (deductive reasoning) that enables a researcher to make causal inferences about whether a treatment or intervention affects designated outcomes while at the same time eliminating potential sources of bias (i.e., third-variable alternatives). The benchmark criteria vary but have, traditionally, included unambiguous temporal precedence, contemporary history, testing, maturation, pretesting procedures, measurement (reliability), statistical regression, differential selection, mortality, contamination (diffusion of treatment), and compensatory responses (equalization, rivalry, demoralization, etc.).

Propensity score – The probability that an individual is assigned to an experimental treatment group conditional on the individual’s covariate information.

Derived as a single scalar function (like a discriminant function) to model a cluster of individual differences variability.

Randomization – A sampling design procedure to ensure that preexisting conditions of the experimental units or observations do not influence variability in the outcomes in question. Will also assure independence of errors at the level of unit of assignment and distribute any possible bias evenly across experimental treatment conditions. Best achieved when subjects are randomly selected, randomly assigned to groups, which are then further assigned randomly to the experimental treatment conditions.

Selection – The process of picking, choosing, or assigning individuals (units) from a larger reservoir to form a homogeneous sample with regard to experimental condition (i.e., treatment vs. control). When done to create two or more probabilistically equivalent groups using strategies of randomization, serves to eliminate bias that might interfere with rendering causal interpretations.

Treatment – A declaration or demarcation of an event manipulation that is oftentimes associated with a structured program, intervention, application, or specific exposure constructed to induce behavioral change defined in terms of a measured endpoint.

Introduction

Randomization is one of many tools researchers have made available to provide experimental control. The English statistician and geneticist Sir Ronald A. Fisher (1935) was one of the first to formally introduce the need for rigorous control in experimental research in order to attribute some element of causation. The emphasis on cause and the need for control arises from a focus on manipulation to induce change in behavior. The term intervention can be used interchangeably with program, treatment, or instructional method. Most common types of interventions in education consist of curricular instruction, programs intended to change behavior (in students, the teacher, or both), improve learning and/or achievement, enrich academic success, and so forth. In psychology, the term intervention is often synonymous with therapeutic engagement (individual or group) or some specified treatment modality (Freudian vs. cognitive-behavioral). In the medical sciences, the term treatment indicates a desired physical change (i.e., reduction in symptoms) through visceral manipulation (removing or adding tissue or bone) or physiological change obtained by invoking a pharmacological drug agent. In some cases,

for instance, with psychiatry, symptom reduction can be based on both pharmacological treatment and conjunctive psychotherapy with the goal of symptom amelioration. Regardless of discipline, all references to intervention seem to accentuate a manipulation that is meant to cause something to happen. When this occurs, a behavioral scientist wants to make attributions that the change in behavior was a direct result of some manipulation controlling for other possible sources of variability that may exert an influence. The most common form of a causal statement suggests that some form of a treatment designated as a factor in an experiment results in behavioral change. In its simplest form, the factor is given two levels – one designating experimentally treated participants and the other level connoting those individuals who were not exposed to the treatment. (Conceivably, we could apply the same treatment to the same physical entity requiring only one level; however, there are major drawbacks to this process. Concerns that need to be addressed in this type of design include temporal stability (consideration of constancy of response over time), causal transience (the effect of treatment A does not linger and influence treatment B), and unit homogeneity (all of the physical entities in the study are identical, and thus order of implementation would not be of concern). Were these assumptions valid, the method of differences would apply and randomization not required). For descriptive purposes, the term treated is used to indicate those individuals or experimental units exposed to the intervention (the nomenclature E is often used) and the term control or untreated used to capture individuals assigned to a comparison condition (here C is frequently used). Randomization speaks to the process of how individuals are assigned to the different treatment levels (E and C) and the techniques used to equilibrate both measured and unmeasured pretreatment characteristics during assignment. When a randomized field trial is implemented and all other factors considered (i.e., fidelity of implementation, contamination, and differential attrition, to name a few), it is considered the gold standard for asserting causal inference.

In the case of many educational interventions, a researcher is largely invested in finding out whether a select intervention or program of instruction works by improving academic performance or boosting achievement scores in a designated sample. Here then, the focus is on derivation of a meaningful effect that comes about by contrasting treatment conditions that perform in a specific time frame on a common benchmark. The level of bias in estimating this effect is thus, largely, determined by the method of assignment of individuals or units to the various experimental treatment conditions. All things considered, if participants assigned to either the treatment group (receiving the intervention) or a comparison control group has an equal probability of being assigned to their

respective treatment condition, then the study meets necessary – but not sufficient – conditions to assert some level of causal inference (and can be labeled true experimental or randomized). Should this occur, there is much less concern with selection bias or extraneous confounding nuisance factors that may influence the outcomes in question. One of the major assumptions of randomization is that treatment assignment method is independent of any systematic pretreatment differences that could be used to characterize participants. Maintaining this stringent condition reduces the influence of potential confounds and allows a researcher to make causal assertions about the program of instruction or intervention in question (all factors considered the program or instructional method identified as the treatment is the only variable that can cause the observed differences between the groups). It is easy to see then that the absence of randomization serves as a threat to the internal validity of a study and affects the processes through which scientists make inductive inferences.

This article outlines the requirements for randomization paying particular attention to the underlying philosophical arguments surrounding cause and effect. The article briefly describes the most commonly used methods for randomization, noting what these techniques net in terms of experimental control. In addition, the article offers several alternatives to randomized assignment in the event that such manipulation is not possible or would disturb the scientific evidence. Each strategy is discussed with regard to how it stands up against the gold standard of randomization as well as noting certain pitfalls that might be encountered with inappropriate use.

Purpose of Randomization

Philosophically speaking, one can never really know why something happens nor can one predict events with certainty. To an ever-mindful child, flipping a light switch is the sole reason why the light glows and no other ulterior explanation concerning voltage, impedance, and electric current is required. However, to the more sensitive and knowing adult mind, there is considerable more to the light burning incandescent than the flip of a switch. Even though scientists operate knowing full well that they cannot assert ultimate cause, there is still the desire to learn as much about the world and our human experience as possible. As part and parcel of this epistemological pursuit, scientists engage a logic of causation using an investigative system of reasoning (i.e., inductive, hypothetico-deductive, and abductive) that provides systematic methods to clarify knowledge. This system has undergone considerable change since early Greek philosophers argued about the necessary and sufficient conditions of causation – primarily because modern physical sciences (referring

mainly to quantum physics) have cast doubt on whether one will ever observe certain events to occur, but with available mathematical precision will know with certainty (or regularity) they do occur. In an axiomatic system, scientists postulate that things will occur and then through logical deduction gather observational data to strengthen the veracity of original postulates. In subatomic or elementary particle physics, the human mind can rarely truly see events as they unfold, but the same human biocomputing machine is capable of proceeding from axioms and postulates to gather data or evidence of patterns. As a result, the changing veneer of science has stimulated a more modern treatment of the philosophy of causation; one that has fabricated a reasoned stance from which we gather evidence of stable phenomenon.

As philosophical discourse on theories of causation unfolded, philosophers became more engaged in fashioning the particulars of causal inferences and backed off the need to equate causation with explanation. Greater emphasis was placed on the methods scientists use – both in the laboratory as well as the nature of experimental reasoning used to explain their findings. One outgrowth of this emphasis is critical rationalism, which provides epistemic criteria for a scientific method emphasizing not what one knows, but rather how one ascertains what one knows. According to critical rationalism, the goals of science are to rule out false impressions and eliminate inadequate or weak theories. Philosopher Karl Popper – responsible for explicating the tenets of critical rationalism – termed this emphasis on paring away improper explanations the falsification or refutability of a theory. No matter the level of experimental sophistication even that obtained in the hard physical sciences, a theory can never be proved as true; one merely probes the value of our theories as explanations of events. Good theories are the ones that withstand tests attempting to refute them (this has been termed the logic of falsification).

This stance makes all experimental work especially in the behavioral sciences really a statement of probability articulated in terms of how likely something can be attributed to a set of events. As part of the logic of scientific method an investigator frames the null hypothesis (the outcome Y is the same regardless of whether E vs. C was administered as a treatment) by a stated conditional significance level to minimize the probability that we would observe this event ($E - C = 0$) by chance alone. Stated differently, if an investigator sets the significance level $p < 0.05$, one could expect the event expressing no difference in treatment conditions (i.e., null hypothesis) to occur 5 times in 100 trials (i.e., chance findings); whereas, the other 95 times, the investigator rejects the null in favor of an alternative hypothesis ($E - C > 0$). The alternative is part of the success of science and occurs because of some systematic reason – likely an effect produced by the treatment in question. This logic of inference helps to frame the

importance of randomization because a researcher wants to avoid chance findings and also eliminate the potential that some systematic but unknown cause was responsible for the desired outcomes.

With this hypothesis-testing framework in mind, a scientist interested in predicting a specific outcome (effect) from an experimental manipulation must rule out that extraneous or unwanted sources of variation (also termed rival hypotheses, alternative explanations, or threats to the interpretation), and also present valid explanations for the outcome – even though these sources of variance were not, directly, part of the manipulation. If the connection between two variables is causal and all other possible contributing events are known, the scientist has taken great strides to ensure the internal validity of the study. If there are plausible threats to internal validity whatever they may be – with random assignment – they are distributed equally over conditions (the extraneous variable is equally like to occur among participants assigned to receive an experimental treatment as in those assigned to the control condition). Based on the principles of randomization, if a researcher were to sample a unit or individual from one level and compare this to a randomly selected unit obtained from another level (here we expressly state that the experimental treatment has two levels), all other factors considered these units would be comparable (this is even more true with repeated samplings). In essence, randomization grapples with differences in people or resources by selecting units or individuals irrespective of any prior existing conditions that may be confounded with the outcome. While randomization cannot offset many plausible and real threats to internal validity, it is a surefire method to eliminate selection bias as a counterfactual position.

Methods of Randomization

Random assignment occurs when the procedure for assigning units or individuals to conditions is based on chance. The most ubiquitous form of randomization involves a simple coin toss. Given the equal probability of obtaining a head or tail from any single independent toss of a fair coin, this represents the simplest way to randomize subjects to experimental treatment conditions. Rolling a fair die can also achieve the same result given the probability of obtaining any number on the face of a die is one-sixth and the probability of obtaining any particular number is governed by chance alone. Simple random sampling can also be accomplished using the lottery method – which involves placing an equal number of distinctly colored small objects into a receptacle of some sort. The process of assignment for simple random sampling involves reaching into the receptacle without looking and removing the objects one by one (without replacement). There is no set or defined way to remove

each object; they are just grabbed one by one. After each object is removed, it is assigned to an experimental group in some systematic fashion. For example, the first object pulled out can assign a subject to receive an experimental treatment and the second object pulled out assigns the next subject to the control group with each successive draw assigned in the same prescribed manner. The objects can be colored pieces of paper, paper with numbers scribbled on them that match up to student or patient identifiers, or colored marbles. If a study warrants three experimental conditions with two experimentally treated groups and one comparison control group, then three colors would be used to distinguish assignment procedures. In any of the examples provided above, a simple random number generator using computing resources can be used to create numeric lists. Subjects can then be assigned on the basis of whether their identification number matches the one appearing on a list. In the event that numbers are generated that do not match, these are discarded or ignored. A researcher would move down the list until a match is found and then assign the matched number to experimental condition. To avoid tedium, multiple numbers or ranges of numbers can be used to find suitable matches for personal identifiers (all numbers ending with 1 as in 2001, 3001, and so forth would be assigned to one condition while numbers ending in 2 as in 2562, 3472, and so forth would be assigned to a different condition until all participant's numbers have been exhausted).

In order to truly benefit from random assignment, there must be sufficient numbers of units (individuals) “relative to the variability between units” (Cook and Campbell, 1979: 5). Again, this reinforces the goal of random assignment to equilibrate any potential differences that might exist before a treatment is applied. Designing a process of random assignment around two participants – one male and one female – is unlikely to achieve the desired results, particularly if gender masks other meaningful causal attributes (i.e., sex-linked differences), or has direct relevance to the endpoint. While the need for sufficient sampling speaks directly to the issue of power, it also reinforces the emphasis on average behavior across a representative sampling.

Is Randomization a Panacea?

Despite the overall emphasis on randomization as a means to control extraneous and systematic variance in an experimental framework, there are some concerns that it is not the ultimate panacea. For instance, Cook and Campbell (1979) suggest “it is indeed false to claim that randomization controls for all threats to internal validity” (p. 85). Even with randomization, threats to internal validity can arise from study attrition (i.e., loss of subjects over time), differential attrition (unequal loss of subjects from treatment conditions), and compensatory reactions. The latter

situation, often, occurs if experimental assignment procedures are meant to remove any systematic differences that may result from some privileged resources such as those obtained through status or power (socioeconomically valued factors). In the case of compensatory rivalry, participants in the untreated control condition who do not receive the benefits of an intervention begin to make claims that they too should receive the intervention (and its benefits), and thus disrupt the experiment by creating their own desirable treatment to improve the designated outcomes. Compensatory equalization, often, arises when someone in a position of authority (i.e., management or school administration) decides that participants assigned to the control condition should benefit from an intervention, and thus contaminate the assignment procedure by distributing intervention materials. Compensatory rivalry can be attenuated by promising control participants that they will receive a delayed intervention or providing something of equal value that is unrelated to the target endpoint.

In many instances, randomization is not desirable, feasible, or permissible, and this has led to proliferation of quasi-experimental designs that use various methods to approximate equivalence between experimental units. The nonexperimental group design, interrupted time series design, and other remedies (matching, stratifying, masking or blinding, cutoff-based methods, and regression adjustment with covariates) are used to contrast a treatment condition and comparison group on some outcome of interest. While these remedies afford some element of control, they do not equilibrate preexisting conditions by design, but rather by statistical analysis procedures. Even with these refinements, it is better to actively deliberate about plausible threats to validity before designing a study rather than hoping some laundry list of experimental design procedures would eliminate entirely all threats to validity. The following section explores several commonly used methods that make adjustments for sample selection bias and create the specter of randomization: propensity scores and regression discontinuity (RD).

Propensity Score Method

Propensity scores offer an alternative to randomization with observational studies or quasi-experimental designs when a researcher wants to make causal statements. Without covariate adjustments, a researcher runs the risk of obtaining biased estimates and arriving at erroneous conclusions regarding treatment effectiveness. Proper covariate adjustments for propensity or balancing scores produce unbiased estimates for treatment effects even in the absence of randomization. A propensity score is a derived measure indicating the likelihood of a subject being assigned to the treatment group based solely on that subject's covariate information. In the case of two

experimental conditions, where individuals are assigned to either receive an intervention or assigned to some type of control comparison group, the propensity score method replaces the confounding covariate measures with a single coarse function that represents the conditional probability of treatment assignment. Derivation of a conditional probability (scalar propensity) score helps adjust any differences between groups (produces an unbiased estimate of average treatment effect) based on a known set of observed covariates (i.e., pretreatment measures).

In many cases, stratification (i.e., subclassification) of the sample based on at most a few covariates would be ideal. Comparisons could then be made between the different cells with the assumption that treatment and control participants are equally represented within each cell. However, oftentimes, the number of possible confounding covariates is large, leading to the likelihood of sparse cells containing either only treatment or control participants. This would truncate the possibility of estimating a treatment effect within the particular strata. In addition, proliferation of strata from excessive covariates creates problems with cross-classification and the inability to obtain proper diagnostics on joint distributions of regressors. When this occurs, propensity score methods provide an ideal means to achieve parsimonious representation of all the different observed or known characteristics that can differentiate treatment and control participants (i.e., this multivariate difference is akin to a linear discriminant function).

Under the stable unit-treatment value assumption, scalar propensity scores are akin to a missing data problem and will satisfy this condition if the treatment assignment is strongly ignorable given the balancing score (the scalar propensity based on a vector of covariates) – that is, the relationship between treatment assignment and the designated endpoint are conditionally independent of the covariates. With a propensity score adjustment, everyone in the population has an equal chance to be assigned to either the treatment or control group, irrespective of any predisposing characteristics (covariates). Once propensity scores are created – and using the *t*-statistic as a gauge – participants can then be iteratively balanced into homogeneous bins or subclasses based on the distribution of estimated propensity scores (five subclasses or quintiles appear sufficient). The end result is that individuals located within a bin have equal probabilities of being selected into either experimental condition (treatment vs. control), and can be considered to have been randomly assigned to their respective conditions. (The propensity score method is based on the notion of equality of variance/covariances within subclasses. If balancing is not achieved, then the model must be reformulated so that the likelihood of being assigned to treatment or control based on propensity scores is equal within a subclass.) Estimates of treatment effects can then be generated within each strata

(quintile) and averaged over the number of strata created. This approach works best with large samples given the expected distributional balance on the covariates in different strata or subclasses. With smaller sample sizes, the anticipated balance may come into question as would be expected with any sample shrinkage and corresponding inflation of standard errors.

Regression Discontinuity

The term regression discontinuity (RD) connotes a type of quasi-experimental design that relies on standard regression methods to obtain unbiased causal estimates of a treatment effect. In order to create boundaries for delineating assignment to experimental condition, an investigator will create a cutoff point using an assignment variable. The assignment variable cannot be caused by the treatment, but does not have to be a pretest score equivalent in form to the posttest (a posttest score is required). It is best – as a means of increasing power – that the assignment variable has minimal overlap with treatment and often multiple assignment measures in the form of an index will improve power. The sample is then divided into two components based on this cutoff score with those scoring on one side of the cutoff assigned to one experimental condition and those scoring on the other side assigned to a different condition. For example, students scoring below the cutoff point on the assignment variable (and, perhaps, showing a deficit in performance) can be designated to participate in a remedial program emphasizing reading or mathematical skills, or some type of instructional curriculum that can enhance their academic aptitude. Those students scoring above the cutoff point are considered the control or comparison group. A scatterplot depicting the assignment scores plotted on the X or horizontal axis and the posttest scores on the Y or vertical axis might show a vertical displacement at the cutoff score if the remedial program has an effect (i.e., the treatment improves participating students' performance), whereas no change or displacement would be expected if there was no effect. The displacement can be witnessed as either a shift in means reflecting treatment benefits measured in units of the outcome or a change in slope of the regression line at the designated cutoff point. Shifts – both in means or slopes – give the name RD and the size or magnitude of the discontinuity is the size of the effect resulting from the treatment. When there is no program or treatment effect, the functional forms of the slopes will be equivalent – holding all other factors constant.

Greater confidence is gained that the remedial program was responsible for the overall treatment effect if all students below (or above) the cutoff benefit equally from the treatment. It is essential to rule out model misspecification (nonlinearity) and to ensure the performance shift in the posttest scores is clearly demarcated by the cutoff; that is,

the same treatment benefit is not observed at other random places along the plot that could support rival interpretations. Moreover, rival hypotheses would have to achieve the precise same effect, using the same markers of performance, and under the same conditions without the benefit of the cutoff. Possible threats to internal validity would have to cause a discontinuity at the precise point of the cutoff and this is highly unlikely, although not implausible. Attrition (from one side of the cutoff) could reduce power and bias the sample as can history, which can influence participants on one side of the cutoff at the exclusion of others. One caution to using this method is the effect of restricted range on variance estimates commonly known as shrinkage. The way to offset shrinkage involves the use of multiple discontinuities to retain the original variances based on a broader range of values for the predictor measure. It is also unlikely that the students in the two boundary conditions would be matched on any aggregate characteristics as encountered with propensity score methods using observed measures. These gray areas make it unlikely that this method will produce efficient and stable estimates of treatment effectiveness without large samples, precise and meaningful cutoff points for assignment, and efforts at cross-validation.

Summary

Randomization affords researchers a clear method to ensure that any differences between units or individuals assigned to experimental conditions are by chance alone. It is an optimal design strategy to fend off certain threats to internal validity and provides the foundation for true experiments seeking to make valid causal inferences. When subjects have an equal probability of being selected and assigned to an experimental condition the unique characteristics they bring to the laboratory or field experiment are equated across treatment conditions. Simply put, the goal of randomization is to seek a level of equality of subjects prior to their assignment to experimental condition. When this occurs, researchers can then make more confident assertions whether a specific manipulation results in the anticipated treatment effect by design as opposed to happening by chance alone.

In the realm of educational studies, it is not always possible to assign students (or teachers) to treatment conditions using random assignment methods. In many cases, well-known studies examining the role of school vouchers, private versus public school education, efficacy of school-based drug-prevention, grade retention, and evaluations of many remedial instructional modalities to improve learning and achievement take shape as observational, quasi-experimental studies lacking the precision afforded by random assignment. These studies tend to be more economical and less cumbersome than randomized trials;

however, the absence of complete randomization hampers scientists' abilities to make valid causal inferences. In recent years, several alternative approaches have been proposed to accommodate the necessity of controlling nuisance factors that may diminish the authority of causal attributions. With these tools in hand, a researcher is much closer to being able to state that a certain manipulation resulted in a specific effect and thus reinforce the cause–effect relationship that is the backbone of all scientific effort.

Although several viable alternatives to randomization exist, two in particular are covered in this article (instrumental variables and fixed-effect methods as possible remedies to randomization are discussed elsewhere in the encyclopedia). Propensity scores provide a parsimonious and efficient remedy to the problem of obtaining unbiased estimates of treatment effectiveness by adjusting or balancing treatment group differences based on a single composite characteristic. Any bias associated with treatment condition assignment is controlled statistically through the covariate adjustment, and subjects are balanced on their propensity for selection. Importantly, models using this approach are only as valid as the model selection process used to include covariates in the scalar function. Hidden or missing covariates can differentiate participants in ways not considered and alter the statistical outcomes or at the very least undermine confidence in any causal interpretation.

Another remedy discussed in this article involves dichotomization of samples in a way that permits investigators to make meaningful comparisons or treatment contrasts as if the subjects had been randomly assigned to discrete experimental groups. While these variants on randomization have less stringent requirements for balancing preexisting differences, they still afford a means of comparison not available with subclassification or other covariate-adjustment methods. The technique of RD works on the assumption that oftentimes natural boundary conditions mimic random assignment to treatment conditions. With RD methods, an investigator formulates the designation (assignment) of experimental versus control groups based on students achieving just above (or below) a certain threshold on some benchmark performance criteria. As a special feature, the selection mechanism using the assignment variable cutoff point is fully known (i.e., there is no measurement error), and there is no hidden bias that can influence the estimate of the treatment effect. The elegance of this approach is that it can be mixed with randomization techniques, utilize multiple cutoff points, and incorporate more than one treatment. When certain assumptions regarding internal validity are met, the resulting estimate of treatment effect is virtually unbiased.

All told, despite differences in these two popular methods to approximate randomization, both are similarly geared toward removing extraneous variance or controlling for preexisting treatment group differences that might bias

the estimation of treatment effects. Random assignment is a major canon of experimental design but not the *sine qua non* that defines experimentation. Other important components include selection of the treatment, dosage, method of dispensing treatment, measuring and timing of effects, choosing participants, the nature of comparison, and assignment protocols. Even with these additional requirements, the goal of any experimental design or sampling strategy is still to ensure that causal assertions are not mistaken, implausible, or falsifiable.

Bibliography

- Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin.
- Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.
- MacKinnon, D. P., Krull, J. L., and Lockwood, C. M. (2000). Equivalence of the mediation, confounding, and suppression effect. *Prevention Science* 1, 173–181.

Further Reading

- Blackwell, E. and Hodges, J. L. (1957). Design for the control of selection bias. *Annual Mathematical Statistics* 28, 449–460.
- Chalmers, A. (2002). Experiment and the growth of experimental knowledge. In Gärdenfors, P., Wölenski, J., and Kijania-Placek, K. (eds.) *In the Scope of Logic, Methodology and Philosophy of Science*, pp 157–169. Dordrecht: Kluwer Academic Publishers.
- D'Agostino, R. B. (1998). Propensity score methods for bias reduction in the comparison of treatment to a non-randomized control group. *Statistics in Medicine* 17, 2265–2281.
- Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* 58, 403–417.
- Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69, 201–209.
- Hirano, K., Imbens, G. W., and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* 71, 1161–1189.
- Indurkha, A., Mitra, N., and Scrag, D. (2006). Using propensity scores to estimate the cost-effectiveness of medical therapies. *Statistics in Medicine* 25, 1561–1576.
- Jacob, B. and Lefgren, L. (2004). Remedial education and student achievement: A regression discontinuity analysis. *Review of Economics and Statistics* 86, 226–244.
- Popper, K. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge Classics.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41–55.
- Rubin, D. B. (1979). Using multivariate matched sampling and regression adjustments to control bias in observational studies. *Journal of the American Statistical Association* 74, 318–324.
- Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine* 127, 757–763.
- Rubin, D. B. and Thomas, N. (1996). Matching using estimated propensity scores: Relating theory to practice. *Biometrics* 52, 249–264.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Casual Inference*. New York: Houghton Mifflin.
- Sobel, M. E. (1995). Causal inference in the social and behavioral sciences. In Arminger, G., Clogg, C. C., and Sobel, M. E. (eds.) *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, pp 1–38. New York: Plenum Press.

Multitrait–Multimethod Designs

M Bong and S Hong, Korea University, Seoul, South Korea

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Confirmatory factor analysis – A restricted form of factor analysis that allows testing validity of psychological measures by imposing theory-driven parameters in terms of factor loadings, factor variance and covariance, as well as error variance and covariance.

Convergent validity – The degree to which multiple measurements of the same psychological construct correlate with one another.

Correlated uniquenesses – Consistent portions of measurement errors that are assumed to embody method effects in confirmatory factor analysis models that do not explicitly specify method factors.

Discriminant validity – The degree to which measurements of different psychological constructs are distinct from each other.

Halo – Cognitive bias with which perceptions of particular characteristics, personality, or attributes influence perceptions of other characteristics, personality, or attributes.

Methods – Particular ways of measuring characteristics, personality, or attributes of individuals.

Random error – Errors in measurements that occur due to chance alone.

Reliability – The degree to which multiple measurements to measuring instruments yield consistent scores.

Social desirability – Tendency to respond to survey items in ways that are deemed more socially acceptable and to be judged favorably by others.

Traits – Particular characteristics, personality, or attributes of individuals.

Validity – The degree to which a scale or an instrument actually measures the Psychological construct that it purports to measure.

Multitrait–multimethod (MTMM) designs refer to an approach proposed by Campbell and Fiske in 1959, which can be employed for validation of psychological measures (Campbell and Fiske, 1959). The MTMM designs rest on an assumption that valid yet independent measurement procedures used for assessing the same psychological construct would produce scores that necessarily correlate highly with each other. Stated differently, if any instrument

or assessment technique under consideration is a valid measurement tool for the psychological construct that it purports to measure, then the scores so produced will not only be reliable but also (1) correlate highly with scores of the same construct assessed with different measurement procedures and (2) not correlate highly with scores of different constructs, even if they are assessed with the same measurement procedures. The MTMM designs thus offer a convenient way of simultaneously examining convergent as well as discriminant validity of a given measure.

To apply MTMM designs, multiple traits (i.e., psychological constructs) need to be assessed using multiple methods – hence multitrait–multimethod. First, researchers assess the trait or the construct of interest for a group of individuals, using at least two or more methods among various instruments and assessment techniques available. It is recommended that maximally different measurement procedures be used to more clearly establish convergent validity. Second, the researchers assess at least one more construct for the same group of individuals with the same set of multiple methods. These additional constructs should be sufficiently distinct from the construct of interest to more accurately address the issue of discriminant validity. They must also be variables that can be adequately assessed by the set of measurement procedures selected. Third, correlation coefficients among the multiple constructs, each assessed with the multiple methods, are computed to generate an MTMM matrix. This MTMM matrix forms the basis for determining whether sufficient evidence exists for the convergent and discriminant validity of the measures under consideration.

Figure 1 shows a schematic representation of an MTMM matrix, showing three traits, traits A, B, and C, each of which is assessed with three methods, methods 1, 2, and 3.

Campbell and Fiske's Approach

As can be seen in Figure 1, the coefficients in an MTMM matrix can be divided into four distinct types. The first type of coefficients are the monotrait–monomethod or reliability coefficients presented in the three reliability diagonals enclosed in clear rectangular boxes. Theoretically, these coefficients are believed to represent correlations between the same construct assessed with the same method and hence, indicate reliability of each of the nine measures. Thus, r_{A1A1} , for example, represents the reliability coefficient for trait A assessed with method 1. The second type refers to the

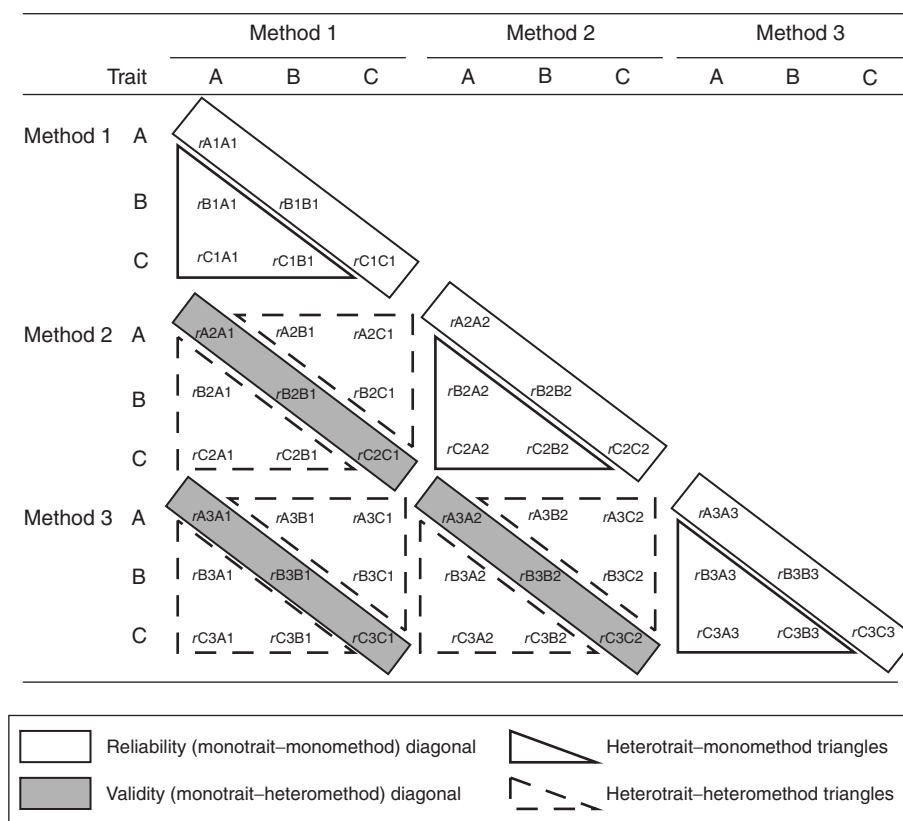


Figure 1 An MTMM matrix of three traits (i.e., traits A, B, and C) assessed with three methods (i.e., methods 1, 2, and 3). From Pedhazur, E. J. and Schmelkin, L. P. (1991). *Measurement, Design, and Analysis: An Integrated Approach*, p 76. Hillsdale, NJ: Erlbaum.

monotrait-heteromethod (MTHM) or validity coefficients displayed in the three validity diagonals presented in shaded rectangular boxes. These coefficients represent correlations between the same construct assessed with different methods. For example, r_{B2B1} represents the correlation between measures of trait B, each assessed with method 1 and method 2, respectively. The MTHM coefficients are also called convergent validity coefficients.

The third type of coefficients are the heterotrait-monomethod (HTMM) coefficients presented in the three solid triangles. These coefficients represent correlations between different constructs assessed with the same method and, as such, can be viewed to embody method effects. For example, r_{C2A2} and r_{C2B2} are the correlations between trait A and trait C and between trait B and trait C, respectively, all assessed with method 2. The fourth type includes the heterotrait-heteromethod (HTHM) coefficients encircled in the six dashed triangles, which indicate correlations between different constructs assessed with different methods. The coefficient, r_{B3C2} , indicates the correlation between trait C assessed with method 2 and trait B assessed with method 3. The HTMM and HTHM coefficients are also called discriminant validity coefficients. Campbell and Fiske proposed a set of criteria that can be applied when comparing the four types of coefficients in an MTMM matrix for assessing construct validity of psychological measures.

Campbell and Fiske regarded reliability as the correlation between measures of the same trait assessed with maximally similar methods and validity as the correlation between measures of the same trait assessed with maximally different methods. The coefficients in the monotrait-monomethod or the reliability diagonal, therefore, must be high to begin with because reliability is a necessary prerequisite for construct validity. As these measures share the same trait and the same method, values of the coefficients in the reliability diagonals are often the largest of all coefficients included in an MTMM matrix.

Provided that the measures are judged to be reliable, Campbell and Fiske's first criterion is that values of the coefficients in the MTHM or the validity diagonals must be sufficiently large to warrant further inspection of construct validity. Correlations between measures of the same construct assessed with different methods constitute primary evidence of convergent validity. Therefore, not only these validity coefficients must be significantly different from zero and sizeable in absolute magnitude but they should also be considerably higher than the coefficients in the HTHM or the HTMM triangles. These latter requirements constitute the second and third criteria, respectively.

The criterion that coefficients in the MTHM or the validity diagonals should be higher than those in the HTHM triangles represents one test of convergent validity. Theoretically, this should be the easiest requirement

to fulfill because the coefficients in the HTHM triangles share neither a common construct nor a common method. Because there is no theoretical reason to predict two distinct constructs to be correlated, especially when each of them is separately assessed with an independent measurement procedure, the HTHM coefficients are expected to be the lowest on average among the four types of coefficients presented in an MTMM matrix. Each of the coefficients in the validity diagonals must be higher than the coefficients in the HTHM triangles that are in the same row or the same column to the validity coefficient under examination. For example, the coefficient, r_{B2B1} , represents the correlation between measures of trait B assessed with two distinct methods, methods 1 and 2. This coefficient, r_{B2B1} , must be higher than the coefficients in the same row, r_{B2A1} and r_{B2C1} , as well as those in the same column, r_{A2B1} , r_{C2B1} , r_{A3B1} , and r_{C3B1} . If any of the coefficients in the HTHM triangles is as high as or higher than the coefficients in the validity diagonals, it indicates poor convergent validity.

The next criterion states that coefficients in the validity diagonals must also be higher than those in the HTMM triangles. It implies that the common variance due to traits must be larger than the common variance due to methods. Stated differently, measures of the same construct assessed with different methods should correlate more strongly with each other than measures of different constructs assessed with the same method. Using the coefficient, r_{B2B1} , again as an example, this coefficient must be higher than the coefficient such as r_{C1B1} , which denotes the correlation of trait B with trait C, both assessed with method 1, or r_{B2A2} , the correlation of trait A with trait B, both assessed with method 2. If coefficients in the HTMM triangles are as high as or higher than those in the validity diagonals, it suggests that a greater amount of variance is due to a common method rather than a common construct. For example, if the coefficient, r_{C1B1} , is higher than the validity coefficient, r_{B2B1} , then it implies stronger effects of method 1 compared to those of trait B. When the trait effects are even weaker than the method effects, it is difficult to claim there exists a sufficient degree of convergent validity.

Discriminant validity states that measures of different constructs, theoretically unrelated, should not be correlated. The criterion that the validity coefficients must be higher than the coefficients in the HTHM triangles can thus function as a test of discriminant validity as well. Discriminant validity is judged to be in jeopardy when measures of distinct constructs correlate too highly with each other. This principle applies equally to situations where the constructs are assessed with the same method or different methods. Assuming the same constructs are being assessed and hence the measures contain an equal amount of variance due to traits, the remaining variance in each measure can be thought of as being due to some factors related to methods and random error. When correlations between measures of distinct traits assessed with different

methods (i.e., the correlations in the HTHM triangles) are too high, appearing as high as or even higher than those between measures of the same trait assessed with an identical set of heteromethods (i.e., the correlations in the validity diagonals), discriminant validity is deemed not to be present. Similarly, when correlations between measures of distinct traits assessed with the same method (i.e., the correlations in the HTMM triangles) are as high as or higher than those between measures of the same trait assessed with the same method (i.e., the correlations in the reliability diagonals), discriminant validity of the measures is again called into question.

Finally, Campbell and Fiske proposed that the pattern of correlations among the traits should stay similar across methods. This criterion applies equally to all heterotrait correlations regardless of whether the traits are assessed with the same or different methods. For example, if r_{B1A1} is higher than r_{C1A1} , which is in turn higher than r_{C1B1} in the top HTMM triangle, then this pattern should hold in all other HTMM as well as HTHM triangles. This means that traits A and B should be most strongly correlated with each other, traits A and C should be next strongly correlated, and traits B and C should be least strongly correlated, irrespective of the methods used to assess these traits. Therefore, in the two HTHM triangles produced by methods 1 and 2, for example, r_{B2A1} and r_{A2B1} should be the highest correlations among the three coefficients within the same triangle, r_{C2A1} and r_{A2C1} the next highest, and r_{C2B1} and r_{B2C1} the lowest correlations, respectively.

The Campbell and Fiske criteria provided useful guidelines for judging the quality of measures of psychological constructs and had been widely employed in construct validation attempts. However, researchers have since pointed out several inherent problems associated with the criteria. The first such problem is an analytical one. The criteria require that one type of correlations (e.g., MTMM correlations) be higher than the other types of correlations and if certain type of correlations (e.g., HTHM correlations) are too high, then it poses a threat to construct validity. However, it is not clear how much higher the desired correlations must be compared to others to be acceptable or how high the unwanted correlation might be to be considered problematic. Visual inspection of MTMM matrices cannot provide helpful answers to these questions. The situation can become even more ambiguous when only some of the correlations in the respective diagonal or triangle are consistent with the criteria or when the particular MTMM matrix as a whole only meets some, not all, of the criteria stipulated by Campbell and Fiske.

The second problem is both conceptual and analytical in nature. The operational definition of maximally different methods to be employed in MTMM designs is vague. For example, if researchers ask a group of individuals to rate some aspect of their own personality and then ask their parents, siblings, or friends to rate the same aspect of those individuals' personality, do these ratings by the self

versus others comprise different methods, let alone maximally different methods? Further, even when sufficiently different methods are used, it is more likely than not that scores produced by these independent methods show some degree of correlation. Depending on the specific traits, subjects, and settings of the measurement, factors such as halo or social desirability might be in operation, rendering the correlation between any two measures to be much higher than warranted purely on the basis of the methods themselves. However, these effects can be neither controlled nor partitioned into different sources by simply inspecting the correlation coefficients visually, as suggested by Campbell and Fiske.

Alternative Approaches

To overcome the aforementioned problems associated with Campbell and Fiske's approach to MTMM analyses, several alternative approaches have been proposed. These alternative models include a confirmatory factor analysis model with trait and method factors, a confirmatory factor analysis model with correlated uniquenesses, and a direct product model.

Confirmatory Factor Analysis Model with Trait and Method Factors

Confirmatory factor analysis (CFA) of MTMM matrices has many important advantages to traditional MTMM analyses that rely upon less-than-systematic comparison of zero-order correlations among the measures as assessed. In general, CFA is a useful tool for testing theoretically driven relationships between variables. It is particularly useful in MTMM designs because, unlike measures in natural sciences, most measures in social and behavioral sciences contain sizeable measurement errors. CFA takes these measurement errors into account by correcting for possible attenuation due to imperfect measurements in estimating relationships among the measures. The CFA-TM (CFA-TM, confirmatory factor analysis model with trait and method factors) approaches further allow partitioning of the variance in each measure into trait, method, and random error components.

The general confirmatory factor analysis model is:

$$\Sigma = \Lambda\Phi\Lambda' + \Psi$$

where Σ is a population correlation matrix, Λ is a matrix of factor loadings, Φ is a matrix of factor correlations, and Ψ is a diagonal matrix of uniqueness variances. For the analysis of MTMM data, the CFA model is extended as shown below:

$$\Sigma = \Lambda_T\Phi_T\Lambda_T' + \Lambda_M\Phi_M\Lambda_M' + \Psi$$

where Λ_T and Λ_M are matrices of factor loadings for traits and methods, respectively. Similarly, Φ_T and Φ_M are matrices of trait and method correlations, respectively. As the correlations among observed variables (i.e., measures) are sum of trait and method factors, the model is an additive model (see Figure 2).

More specifically, the CFA-TM has the following advantages over Campbell and Fiske's traditional approach:

1. it explains the MTMM matrix by introducing latent constructs, rather than making an unrealistic assumption of perfect measurements;
2. it provides statistical testing of global model fit (i.e., whether or not the particular model with multiple traits and multiple methods, as specified, can be regarded as a valid representation of the relationships among the measures);
3. it provides statistical testing of convergent as well as discriminant validity; and
4. it decomposes the variance in each measure as being due to different traits, different methods, and error.

As such, researchers now consider CFA a far more defensible and informative approach than is Campbell and Fiske's original approach to the analyses of MTMM matrices.

Unfortunately, as the CFA-TM approach to MTMM matrices has become widespread, the frequency of problems associated with it has also increased. These problems

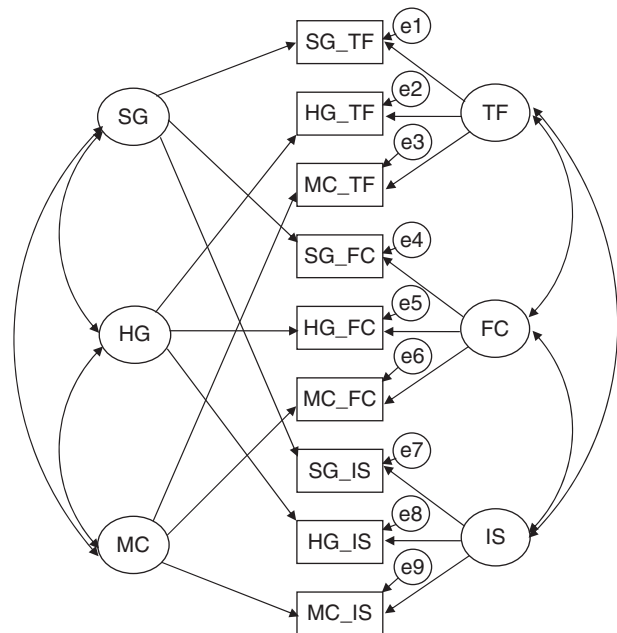


Figure 2 The CFA-TM approach to the MTMM matrix of three traits (i.e., traits SG, HG, and MC) assessed with three methods (i.e., methods TF, FC, and IS).

include nonconvergence, negative variances, nonpositive definite correlation matrices, and other ill-defined solutions.

Confirmatory Factor Analysis Model with Correlated Uniquenesses

To overcome the problem of ill-defined solutions frequently resulted from the CFA-TM applications, Marsh proposed CFA-CU (CFA-CU, confirmatory factor analysis model with correlated uniquenesses) approaches to MTMM matrices. In CFA-CU, method factors are eliminated from the model and correlated uniquenesses are specified instead (see **Figure 3**). The CFA-CU hence specifies correlations among the uniqueness terms of the measures that are assessed with the same method. The assumption is that the variance in each measure not attributable to a trait is due to either a method or random error and the systematic portion of this remaining variance (i.e., uniqueness) is attributable to the method used. By specifying uniquenesses of the measures assessed by the same method to correlate, it becomes possible to account for the shared method effects while avoiding the empirical difficulties commonly encountered in the CFA-TM approaches to MTMM matrices.

The mathematical model of the CFA-CU is identical to the general CFA:

$$\Sigma = \Lambda\Phi\Lambda' + \Psi$$

Specifically, the CFA-CU model differs from the CFA-TM model in the following two points: (1) the CFA-CU model does not include method factors and (2) Ψ contains uniqueness variances on the diagonal and covariances off the diagonal. Covariances among the uniquenesses represent method effects.

Based on extensive evaluation of the CFA-CU using empirical and simulated data, Marsh has reported that this approach converged to proper solutions in most applications. In particular, the CFA-CU approach can represent MTMM data more accurately when the effects of methods are not unidimensional. However, this approach is not without its limitations. The first shortcoming is that correlated method effects cannot be modeled. Because correlations are specified among the uniqueness terms of the measures assessed with a common method instead of latent variables representing the methods, correlations between different methods cannot be specified. The second shortcoming is that method effects can be confounded with random error.

Direct Product Model

Both the CFA-TM and the CFA-CU approaches assume an additive method effect on the trait factor. However, it has been pointed out that these CFA approaches ignore possible

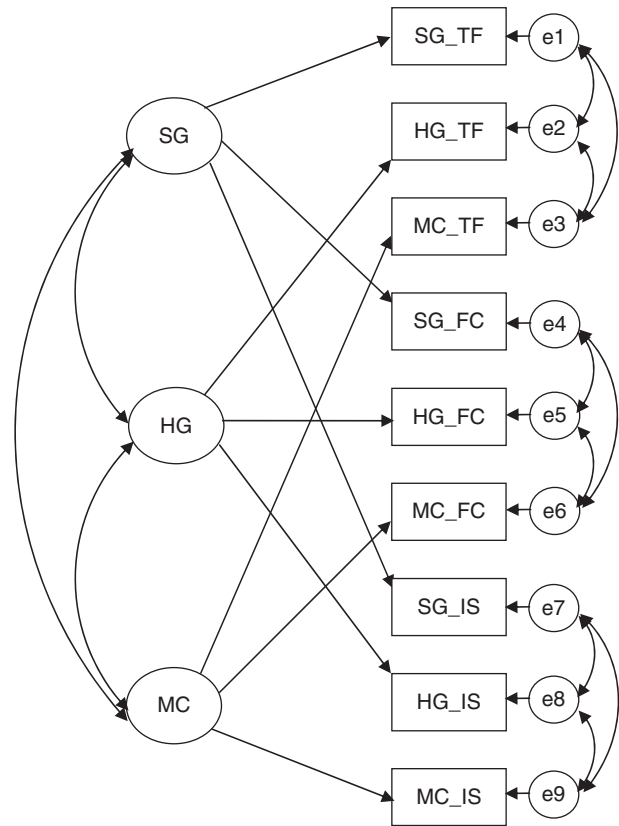


Figure 3 The CFA-CU approach to the MTMM matrix of three traits (i.e., traits SG, HG, and MC) assessed with three methods (i.e., methods TF, FC, and IS).

multiplicative trait–method interactions. In response to this issue, Browne proposed the direct product (DP) model, where the correlation matrix of measured variables can be defined as the direct product of a correlation matrix of traits and a correlation matrix of methods. The mathematical expression of the DP model is as follows:

$$\Sigma = D(P_M \otimes P_T + E)D$$

where Σ is a population correlation matrix, D is a diagonal matrix of communality, P_T and P_M are trait and method correlation matrices, respectively. Here, E is a diagonal matrix of uniqueness and \otimes is right direct Kronecker product. The DP model is less prone to the problem of ill-defined solutions because parameterization of trait and method variance is more parsimonious than it is for the CFA approaches.

Illustration of Data Analyses

The MTMM matrix from a study by Mosher, published in 1968, was re-analyzed here to illustrate the CFA-TM, CFA-CU, and DP approaches. **Table 1** presents an MTMM matrix of three types of guilt as measured by three types of self-reports. The three types of guilt are sex-guilt (SG),

Table 1 Multitrait-multimethod matrix from Mosher (1968)

		TF			FC			IS		
		SG	HG	MC	SG	HG	MC	SG	HG	MC
TF	SG	(0.95)								
	HG	0.28	(0.86)							
	MC	0.58	0.39	(0.92)						
FC	SG	0.86	0.32	0.57	(0.95)					
	HG	0.30	0.90	0.40	0.39	(0.76)				
	MC	0.52	0.31	0.86	0.55	0.26	(0.84)			
IS	SG	0.73	0.10	0.43	0.64	0.17	0.37	(0.48)		
	HG	0.10	0.63	0.17	0.22	0.67	0.19	0.15	(0.41)	
	MC	0.35	0.16	0.52	0.31	0.17	0.56	0.41	0.30	(0.58)

Note: $N = 62$. Numbers in parentheses are reliability coefficients.

TF, true-false; FC, forced-choice; IS, incomplete sentences; SG, sex-guilt; HG, hostility-guilt; MC, morality-conscience.

From Mosher, D. L. (1968). Measurement of Guilt in Females by Self-Report Inventories by, *Journal of Consulting and Clinical Psychology*, 32, 693.

hostility-guilt (HG), and morality-conscience (MC) and the three types of self-reports are true-false (TF), forced-choice (FC), and incomplete sentences (IS).

Campbell and Fiske's Approach

In the MTMM matrix presented in **Table 1**, the bold-typed coefficients presented in shaded cells represent convergent validity (i.e., MTHM) coefficients. The coefficients boxed together by dotted lines (i.e., HTHM) and those by solid lines (i.e., HTMM) represent discriminant validity coefficients. Reliability (i.e., MTMM) coefficients are italicized and put in parentheses in the main diagonal.

The first criterion suggested by Campbell and Fiske for ensuring construct validity is that values of the coefficients in the MTHM or the validity diagonals should be significantly different from zero and sufficiently large. Visual inspection of the coefficients presented in shaded cells indicates this is indeed the case with the coefficients ranging from 0.52 to 0.90, although some of the coefficients are substantially smaller than others. The second criterion of Campbell and Fiske is that the validity coefficients should be larger than the HTHM coefficients. Again, visually comparing each validity coefficient with the coefficients in the same column or the same row within the HTHM boxes reveals this criterion is successfully met. The third criterion states that the validity coefficients should be larger than the HTMM coefficients, which also holds.

Campbell and Fiske's fourth criterion for validity dictates that the same pattern of correlations should emerge between traits, irrespective of whether they are assessed by the same or different methods. In the MTMM matrix presented in **Table 1**, the same pattern of correlations is indeed observed among the three traits across the

HTHM and HTMM triangles. Specifically, the highest correlation coefficient is observed between SG and MC, the next highest correlation coefficient between HG and MC, and the lowest correlation coefficient between SG and HG within each triangle. Out of the 27 correlation coefficients in the six HTHM and the three HTMM triangles, only two are exceptions to this trend. The correlation coefficient between SG and HG (0.39) is higher than that between HG and MC (0.26), when all of them are assessed with the forced-choice tests. The correlation coefficient between SG assessed with the force-choice test and HG assessed with the incomplete sentence test (0.22) is also higher than that between HG assessed with FC and MC assessed with IS (0.17). On the whole, one must conclude that validity of the measures is supported according to the Campbell and Fiske's criteria.

Confirmatory Factor Analysis Model with Trait and Method Factors

The CFA-TM was applied to the MTMM matrix using AMOS 7.0. As standard deviations are not reported in Mosher's original article, arbitrary values were chosen with one condition – similar values should be assigned to the variables assessed with the same method. We developed three nested models to demonstrate how to evaluate the convergent and discriminant validity of the trait factors.

Model 1 is a basic model in which three correlated trait and three correlated method factors are specified, as presented in **Figure 2**. Model 2 specified method factors only, without any trait factor. Model 2 is thus nested within model 1 with greater χ^2 values ($\Delta\chi^2 > 0$) and greater degrees of freedom ($\Delta df > 0$) than those of model 1. A significant difference in χ^2 values at the difference in

degrees of freedom between the two models can be taken as evidence of convergent validity because deterioration in model fit (e.g., greater χ^2 values of model 2) when trait factors are removed indicates the necessity of three types of guilt factors for adequately illustrating the response patterns to the three types of self-reports.

Similarly, we developed model 3 by fixing trait correlations in model 1 at unity. Therefore, model 3 is also nested within model 1 with greater χ^2 values ($\Delta\chi^2 > 0$) and greater degrees of freedom ($\Delta df > 0$) than those of model 1. Again, a significant difference in χ^2 values at the difference in degrees of freedom between the two models indicates evidence of discriminant validity. Significant deterioration in model fit (e.g., greater χ^2 values of model 3) when trait factors are treated as identical ($\phi = 1$) means that the trait factors are independent from each other. Stated differently, significantly poorer model fit with three identical trait factors demonstrates the necessity of treating the trait factors as distinct from each other for proper model definition.

Results show that the fit of Model 1 ($\chi^2 = 25.123$, $df = 18$) was reasonable to satisfactory. In evaluating the fit of CFA models, root mean square error of approximation (RMSEA) values less than 0.10 and Tucker–Lewis

index (TLI) and comparative fit index (CFI) values greater than 0.90 are typically regarded as indicating acceptable model fit. The values of RMSEA, TLI, and CFI for model 1 were 0.081, 0.966, and 0.983, respectively (note that error variances of the variables measured by a common method were fixed to be equal to solve the problem of improper solutions). In contrast, the analyses using models 2 and 3 produced improper solutions, indicating that the models could not be considered as reasonable representation of the empirical data. Together, these results argue for convergent as well as discriminant validity.

The magnitude of trait factor loadings also supports convergent validity. As **Table 2** shows, all trait factor loadings were relatively strong and statistically significant, ranging from 0.582 to 0.979.

Discriminant validity was evaluated by examining the trait and method factor correlation matrices according to Goffin's guideline. The guideline states that correlations greater than 0.71 in magnitude are of concern because values exceeding 0.71 indicate that the two variables share more than 50% of their variances. As **Table 3** demonstrates, none of the trait correlations were greater than 0.71. Therefore, discriminant validity for the trait factors was supported.

Table 2 Factor loadings from the CFA-TM model

		<i>Traits</i>			<i>Methods</i>		
		<i>SG</i>	<i>HG</i>	<i>MC</i>	<i>TF</i>	<i>FC</i>	<i>IS</i>
TF	SG	0.952			0.244		
	HG		0.959		0.215		
	MC			0.979	−0.075		
FC	SG	0.920				0.192	
	HG		0.936			−0.083	
	MC			0.879		0.339	
IS	SG	0.756					0.278
	HG		0.701				0.418
	MC			0.582			0.577

Table 3 Factor correlations from the CFA-TM model

		<i>Traits</i>			<i>Methods</i>		
		<i>SG</i>	<i>HG</i>	<i>MC</i>	<i>SG</i>	<i>HG</i>	<i>MC</i>
SG		1.000					
HG		0.314	1.000				
MC		0.622	0.395	1.000			
TF					1.000		
FC					0.022	1.000	
IS					0.040	0.268	1.000

Confirmatory Factor Analysis Model with Correlated Uniquenesses

Here we demonstrate application of CFA-CU to the same MTMM matrix. We developed two nested models with correlated uniquenesses. Model 4 is a CFA-CU model with only trait factors. Instead of specifying separate method factors, uniquenesses of the variables assessed with the same method were correlated with each other (see **Figure 3**). We next developed model 5 by fixing trait correlations in model 4 at unity. As model 5 is nested within model 4, we were able to evaluate discriminant validity of the trait factors by comparing the two models.

As was the case in the CFA-TM application presented earlier, a significant difference in χ^2 between the two models indicates evidence of discriminant validity. Results show that the fit of Model 4 ($\chi^2 = 21.808$, $df = 21$) with three correlated yet independent trait factors was excellent. The values of RMSEA, TLI, and CFI were 0.025, 0.997, and 0.998, respectively (note that error variances of the variables measured by a common method were fixed to be equal to solve the problem of improper solutions). However, the analysis using models 5 produced improper solutions, indicating that the model with identical trait factors was not reasonable. This result supports discriminant validity of the three trait factors.

Evidence of convergent validity comes from the magnitude of trait factor loadings. As shown in **Table 4**, all trait factor loadings were strong and statistically significant, ranging from 0.639 to 0.948. Discriminant validity for the trait factors was supported, given that none of the trait correlations was greater than the 0.71 limit suggested by Goffin's guideline (see **Table 5**).

DP Model

In the DP model, convergent and discriminant validity can be evaluated by examining the correlations among

method factors and those among trait factors, respectively. Correlations among method factors that are substantially different from zero argue for convergent validity. Correlations among trait factors that are substantially lower than the upper bound of unity can be taken as evidence of discriminant validity. In addition, correlations among method factors should be greater than those among trait factors to further ensure discriminant validity.

Using a computer program called MUTMUM, the DP model was analyzed. The fit of the DP model was mediocre (RMSEA = 0.092). **Table 6** provides MUTMUM outputs.

To evaluate convergent validity, correlations among the method factors were examined first. They ranged from 0.755 to 0.963 and were substantially different from zero. A computer program called CONFID, included in the MUTMUM package, was used to construct a 95% confidence interval around each of these correlations to evaluate convergent validity statistically. The lower and upper limits of the confidence intervals were 0.884 and 0.989 for the TF-FC, 0.579 and 0.864 for the TF-IS, and 0.617 and 0.899 for the FC-IS pairs, respectively. None of the 95% confidence intervals contained zero, indicating that these correlations are all statistically significant at $\alpha = 0.05$. These results can be taken as evidence of convergent validity.

Similarly, 95% confidence intervals were constructed around correlations among the trait factors to examine discriminant validity statistically. The lower and upper limits of these 95% confidence intervals were 0.131 and

Table 4 Factor loadings from the CFA-CU model

		<i>Traits</i>		
		<i>SG</i>	<i>HG</i>	<i>MC</i>
TF	SG	0.948		
	HG		0.948	
	MC			0.948
FC	SG	0.918		
	HG		0.918	
	MC			0.927
IS	SG	0.718		
	HG		0.705	
	MC			0.639

Table 5 Factor correlations from the CFA-CU model

	<i>Traits</i>		
	<i>SG</i>	<i>HG</i>	<i>MC</i>
SG	1.000		
HG	0.303	1.000	
MC	0.639	0.379	1.000

Table 6 Final estimates (MWL)

<i>Correlations among traits ($p \downarrow T$)</i>			
	<i>SG</i>	<i>HG</i>	<i>MC</i>
SG	1.000		
HG	0.353 (0.106)	1.000	
MC	0.610 (0.087)	0.461 (0.098)	1.000
<i>Correlations among methods ($p \downarrow M$)</i>			
	<i>TF</i>	<i>FC</i>	<i>IS</i>
TF	1.000		
FC	0.963 (0.022)	1.000	
IS	0.755 (0.071)	0.798 (0.069)	1.000

0.541 for the SG-HG, 0.411 and 0.753 for the SG-MC, and 0.249 and 0.631 for the HG-MC pairs, respectively. None of these confidence intervals contained unity and hence discriminant validity was supported. The correlational pattern that shows greater correlations among the method factors than those among the trait factors also argues for the discriminant validity of the traits.

See also: Classical Test Theory Reliability; Educational Measurement: Overview; Validity.

Further Reading

- Bagozzi, R. P. and Yi, Y. (1990). Assessing method variance in multitrait–multimethod matrices: The case of self-reported affect and perceptions at work. *Journal of Applied Psychology* **75**, 547–560.
- Bagozzi, R. P. and Yi, Y. (1991). Multitrait–multimethod matrices in consumer research. *Journal of Consumer Research* **17**, 426–439.
- Bagozzi, R. P. and Yi, Y. (1992). Testing hypotheses about methods, traits, and commonalities in the direct-product model. *Applied Psychological Measurement* **16**, 373–380.
- Bong, M. and Hocevar, D. (2002). Measuring self-efficacy: Multitrait–multimethod comparison of scaling procedures. *Applied Measurement in Education* **15**, 143–171.
- Browne, M. W. (1984). The decomposition of multitrait–multimethod matrices. *British Journal of Mathematical and Statistical Psychology* **37**, 1–21.
- Campbell, D. T. and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin* **56**, 81–105.
- Coover, M. D., Craiger, J. P., and Teachout, M. S. (1997). Effectiveness of the direct product versus confirmatory factor model for reflecting the structure of multimethod–multitrait job performance data. *Journal of Applied Psychology* **82**, 271–280.
- Crocker, L. and Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Orlando, FL: Holt, Rinehart and Winston.
- Kenny, D. A. and Kashy, D. A. (1992). Analysis of the multitrait–multimethod matrix by confirmatory factor analysis. *Psychological Bulletin* **112**, 165–172.
- Marsh, H. W. (1989). Confirmatory factor analysis of multitrait–multimethod data: Many problems and a few solutions. *Applied Psychological Measurement* **13**, 335–361.
- Marsh, H. W. (1993). Multitrait–multimethod analyses: Inferring each trait–method combination with multiple indicators. *Applied Measurement in Education* **6**, 49–81.
- Marsh, H. W. and Bailey, M. (1991). Confirmatory factor analysis of multitrait–multimethod data: A comparison of alternative models. *Applied Psychological Measurement* **15**, 47–70.
- Marsh, H. W. and Byrne, B. M. (1993). Confirmatory factor analysis of multitrait–multimethod self-concept data: Between-group and within-group invariance constraints. *Multivariate Behavioral Research* **28**, 313–349.
- Marsh, H. W., Byrne, B. M., and Craven, R. (1992). Overcoming problems in confirmatory factor analyses of MTMM data: The correlated uniqueness model and factorial invariance. *Multivariate Behavioral Research* **27**, 489–507.
- Marsh, H. W. and Hocevar, D. (1983). Confirmatory factor analysis of multitrait–multimethod matrices. *Journal of Educational Measurement* **20**, 231–248.
- Marsh, H. W. and Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin* **97**, 562–582.
- Marsh, H. W. and Hocevar, D. (1988). A new, more powerful approach to multitrait–multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology* **73**, 107–117.
- Millsap, R. E. (1990). A cautionary note on the detection of method variance in multitrait–multimethod data. *Journal of Applied Psychology* **75**, 350–353.
- Millsap, R. E. (1992). Sufficient conditions for rotational uniqueness in the additive MTMM model. *British Journal of Mathematical and Statistical Psychology* **45**, 125–138.
- Mosher, D. L. (1968). Measurement of guilt in females by self-report inventories. *Journal of Consulting and Clinical Psychology* **32**, 690–695.
- Pedhazur, E. J. and Schmelkin, L. P. (1991). *Measurement, Design, and Analysis: An Integrated Approach*. Hillsdale, NJ: Erlbaum.
- Schmitt, N. and Stults, D. M. (1986). Methodology review: Analysis of multitrait–multimethod matrices. *Applied Psychological Measurement* **10**, 1–22.
- Scullen, S. E. (1999). Using confirmatory factor analysis of correlated uniqueness to estimate method variance in multitrait–multimethod matrices. *Organizational Research Methods* **2**(3), 275–292.
- Tomás, J. M., Hontangas, P. M., and Oliver, A. (2000). Linear confirmatory factor models to evaluate multitrait–multimethod matrices: The effects of number of indicators and correlation among methods. *Multivariate Behavioral Research* **35**, 469–499.
- Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait–multimethod data. *Applied Psychological Measurement* **9**, 1–26.
- Wothke, W. (1996). Models for multitrait–multimethod matrix analysis. In Marcoulides, G. A. and Schumacker, R. E. (eds.) *Advanced Structural Equation Modeling*, pp 7–56. Mahwah, NJ: Erlbaum.
- Wothke, W. and Browne, M. W. (1990). The direct product model for the MTMM matrix parameterized as a second order factor analysis model. *Psychometrika* **55**, 255–262.

Multivariate Longitudinal Data Analysis

H Liu, Beijing Normal University, Beijing, People's Republic of China

L Chang, The Chinese University of Hong Kong, Hong Kong People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

Analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) have traditionally been used to analyze longitudinal or repeated measures data. However, these traditional methods are limited by the strict assumptions concerning missing data across time and the variance–covariance structure of the repeated measures. These methods are also limited because they focus on estimating developmental trends more than identifying and accounting for the individual variations in their development or change over time. For these and other reasons, longitudinal data analysis has in recent years been treated not as a special topic within the analysis of variance framework but as a modeling methodology within the larger multivariate data modeling framework. Key developments in this transition have resulted in three major longitudinal data analysis methods: hierarchical linear models (HLMs; Bryk and Raudenbush, 1992; Goldstein, 1995), latent growth modeling (LGM), and cross-lagged regression model (Jöreskog and Sörbom, 1979). These longitudinal modeling tools help researchers to interpret repeated measures data and formulate research questions that go beyond simply estimating growth trends. Researchers are interested in understanding how attributes of an individual change over time, how individuals differ in their attribute changes, and what accounts for the individual variations of the attribute change (Singer and Willett, 2003; Collins and Sayer, 2001). In this article, we review these three longitudinal data modeling methods by discussing the conceptual framework of each method, presenting empirical examples using the method and by discussing the relationships among the methods.

Multilevel Models for Multivariate Change

In general, multilevel models are used for analyzing hierarchically organized data. Longitudinal data can be considered to be hierarchical in that measurement occasions are nested within subjects. Raudenbush and Bryk (2002) and Goldstein (1995) provide detailed treatment of multilevel models. We consider the case that includes two response variables, X_1 and X_2 , each measured on four occasions. In the HLM framework, the response variables do not need to be measured at the same time points or for the same number of times. In specifying a multivariate multilevel model, the data are treated as if only a single outcome variable is measured. Let X_{itk}^* be a score for

individual i ($i = 1, 2, \dots, n$) at occasion t ($t = 0, 1, 2, 3$) on outcome variable k ($k = 1, 2$), and consider X_{itk}^* as if it were a single variable with aspects of its measurement defined by subscripts. We then define two dummy variables, δ_1 and δ_2 , one for variable X_1 and the other for X_2 . That is, where $\delta_1 = 1$ if a given measure is on X_1 , and $\delta_1 = 0$ otherwise; $\delta_2 = 1$ if a given measure is on X_2 , and $\delta_2 = 0$ otherwise. The data can be arranged as shown in Table 1.

A general representation of the multivariate multilevel model for change could be given as $X_{itk} = \sum_k \delta_k(\text{model})$. The model refers to univariate change,

$$X_{itk} = \sum_k \delta_k(\beta_{0ik} + \beta_{1ik}t + e_{itk}).$$

In this model, the trajectory trend (e.g., linear model) is assumed to be the same for different response variables. β_{0ik} is the intercept for individual i on outcome variable, X_k , $k = 1, 2$. β_{1ik} is the slope for individual i on outcome variable, X_k , $k = 1, 2$. t is the measurement of time (the values do not have to be the same for each individual or on each outcome variable). e_{itk} is the residual for individual i on occasion t for outcome variable, X_k , $k = 1, 2$. The intercepts and slopes for the k outcome variables are random variables, with their variations across individuals being modeled as the following:

$$\beta_{0ik} = \beta_{00k} + u_{0ik}$$

$$\beta_{1ik} = \beta_{10k} + u_{1ik}$$

Here β_{00k} and β_{10k} represent the fixed effects for the intercept and slope on the outcome variable, X_k , $k = 1, 2$, and u_{0ik} and u_{1ik} represent random variation of individuals around the mean intercept and slope, respectively, for the outcome variable, X_k , $k = 1, 2$.

The key to analyzing multivariate longitudinal data within the multilevel framework is to treat the data as if there is only one outcome variable. For each outcome variable X_k , a fixed intercept, β_{0k} , and a fixed slope, β_{1k} , are estimated. Random effects are estimated as the following: The residual variance, $\sigma_{ek}^2 = \text{var}(e_{itk})$; the intercept variance, $\sigma_{u0k}^2 = \text{var}(u_{0ik})$; the slope variance, $\sigma_{u1k}^2 = \text{var}(u_{1ik})$; the covariance between the intercept and slope, $\sigma_{u01k} = \text{cov}(u_{0ik}, u_{1ik})$. In addition, random effects are estimated for the covariance matrices of intercepts and slopes for X_1 and X_2 as the following: $\sigma_{u0102} = \text{cov}(u_{0i1}, u_{0i2})$ is the covariance of intercept on outcome variable, X_1 , and the intercept on outcome variable, X_2 ; $\sigma_{u0112} = \text{cov}(u_{0i1}, u_{1i2})$ is

the covariance of the intercept on outcome variable, X_1 , and the slope on outcome variable, X_2 ; $\sigma_{u1102} = \text{cov}(u_{1i1}, u_{0i2})$ is the covariance of the slope on outcome variable, X_1 , and the intercept on outcome variable, X_2 ; $\sigma_{u1112} = \text{cov}(u_{1i1}, u_{1i2})$ is the covariance of slope on outcome variable, X_1 , and the slope on outcome variable, X_2 .

The above describes the unconditional multivariate growth model, which can be extended to the conditional models that include individual-level variables as predictors of individual differences in the growth coefficients representing developmental change. For example, we can include the individual predictor, Z , into the equations involving the intercept and the linear slope to account for the individual-level variations in these two growth coefficients. The conditional model can be stated as the following:

$$\beta_{0ik} = \beta_{00k} + \beta_{01k}Z + u_{0ik}$$

$$\beta_{1ik} = \beta_{10k} + \beta_{11k}Z + u_{1ik}$$

where β_{01k} and β_{11k} represent the fixed effects of the individual variable Z on the random intercept and slope of outcome variable, X_k , $k = 1, 2$; u_{0ik} and u_{1ik} represent random residual variation of individuals of the intercept and slope, respectively, for outcome variable, X_k , $k = 1, 2$.

Table 1 The hierarchical structure of two variables measured on four occasions

Variables	δ_1	δ_2	Time
X_{i11}	1	0	0
X_{i21}	1	0	1
X_{i31}	1	0	2
X_{i41}	1	0	3
X_{i12}	0	1	0
X_{i22}	0	1	1
X_{i32}	0	1	2
X_{i42}	0	1	3

Data Illustration

The following data are used to illustrate the multivariate multilevel longitudinal data analysis procedures. Cognitive and general self-concept of primary school students in Hong Kong were measured four times at approximately 1-year intervals over the period from grade 3 to grade 6. The sample size was 264 (149 boys and 115 girls). The Perceived Competence Scale for Children (Harter, 1982) was used to obtain these self-concept measures. The scale has been used with other Chinese children of similar ages (Chang, 2003). Internal consistency reliability based on the present sample was 0.71, 0.73, 0.70, and 0.74 for times 1, 2, 3 and 4, respectively. The results from the unconditional model without the individual-level predictors are presented in **Table 2**.

The fixed effects in the multilevel model include mean intercepts and slopes of the two outcome variables, cognitive self-concept (CC) and general self-concept (GC). Because the four measurement occasions were coded as 0, 1, 2, and 3 for grade 3, 4, 5, and 6, the intercepts (2.689 for CC and 2.654 for GC) are interpreted as the mean level of the two self-concepts for grade 3. It can also be seen from **Table 2**, the linear slopes for cognitive self-concept and general self-concept are significant (−0.268 for CC and −0.333 for GC). The negative slope estimates suggest that, across all students, the mean level of cognitive self-concept and general self-concept decrease linearly from grade 3 to grade 6. For the random parameter estimates, all variances are significant, as shown by the variance estimates being at least several times their standard errors in magnitude. Thus, individuals showed statistically significant variations in their start levels (0.263 for CC and 0.251 for GC) and in the rates of linear change on the two outcome variables (0.031 for CC and 0.033 for GC). The correlation between the two intercepts is significant (0.419), suggesting that the level of cognitive self-concept at grade 3 is significantly correlated with the level of

Table 2 Fixed and random effects results for cognitive self-concept (CC) and general self-concept (GC)

Fixed parameters	Estimate	SE	t-Value	
CC intercept	2.689	0.034	79.088	
CC slope	−0.268	0.012	−22.333	
GC intercept	2.654	0.034	78.059	
GC slope	−0.333	0.013	−25.615	
Variance (diagonal), covariance (above diagonal), standard error (in bracket) and correlation (below diagonal) of the random intercepts and slopes				
Random parameters	CC intercept	CC slope	GC intercept	GC slope
CC intercept	0.263 (0.025)	−0.019 (0.007)	0.108 (0.019)	0.003 (0.007)
CC slope	−0.206	0.031 (0.004)	−0.010 (0.007)	0.003 (0.003)
GC intercept	0.419	−0.098	0.251 (0.025)	−0.026 (0.008)
GC slope	0.027	−0.092	−0.280	0.033 (0.004)

general self-concept in grade 3. The correlations between intercepts and slopes are negative (-0.206 for CC and -0.280 for GC), indicating that higher initial levels of self-concept are associated with faster decline over the 4 years from grade 3 to grade 6. The correlation between the two slopes is not significant, suggesting that the rate of decline is independent of these two domains of self-concept.

To account for some of these intercept and slope variations, we included individual-level predictor, gender. The results of this conditional model are presented in **Table 3**. The effects of gender on the intercept of CC (-0.307) and that of GC (-0.441) are statistically significant. Because gender was coded as 1 = male and 0 = female, the negative estimates suggest that the initial self-concept is lower for male than female students. Also shown in **Table 3**, gender does not have a significant effect on the linear slopes of the two self-concept variables (0.009 for CC and 0.048 for GC), suggesting that the linear decline in self-concept across 4 years does not differ between the two genders.

Latent Growth Curves Model

The latent growth curve model (LGCM) is built on the pioneering work by Rao (1958), Tucker (1966), Meredith and Tisak (1990), McArdle (1988), and McArdle and

Epstein (1987). The model is best depicted graphically as shown in **Figure 1**.

Figure 1 represents growth curves of two variables, X_1 and X_2 , measured on four occasions which are represented as 0, 1, 2, and 3. For each variable, we include an intercept and a slope (I_1, S_1 and I_2, S_2 for variables X_1 and X_2 , respectively). Parameter estimates of interest include slope loadings for each of the two variables ($\lambda_{10}, \lambda_{11}, \lambda_{12}, \lambda_{13}$ and $\lambda_{20}, \lambda_{21}, \lambda_{22}, \lambda_{23}$), two intercept means (M_{I1} and M_{I2}), two intercept variances (D_{I1} and D_{I2}), two slope means (M_{S1} and M_{S2}), two slope variances (D_{S1} and D_{S2}), correlations between intercepts and slopes ($r_{I1S1}, r_{I2S2}, r_{I1S2}, r_{I2S1}$), the correlation between the two intercepts (r_{I1I2}) and between the two slopes (r_{S1S2}). The corresponding equations are as follows:

$$X_{1t} = I_1 + \lambda_{1t}S_1 + e_{1t} \quad t = 0, 1, 2, 3$$

and

$$X_{2t} = I_2 + \lambda_{2t}S_2 + e_{2t} \quad t = 0, 1, 2, 3$$

This model implies that a single trajectory underlies growth in each variable across all measurement occasions. Thus, the model allows the identification and estimation of growth in each variable as well as the estimation of the covariance between the two growth slopes. However, the covariance or correlation is not time dependent, thus, ignoring possible interrelations among variables over time. This model is unconditional without individual-level predictors of change. Like the multivariate multilevel model discussed in the previous section, this unconditional model can be extended by including individual-level predictors, which can either be directly observed or latent constructs derived from multiple indicators. A conditional model using directly observed individual-level predictors is presented in **Figure 2**.

Table 3 Results of the effects of gender on the growth parameters for bivariate multilevel model

Dependent variable	Estimate	SE	t
CC intercept	-0.307	0.063	-4.873
GC intercept	-0.441	0.060	-7.350
CC slope	0.009	0.025	0.360
GC slope	0.048	0.026	1.846

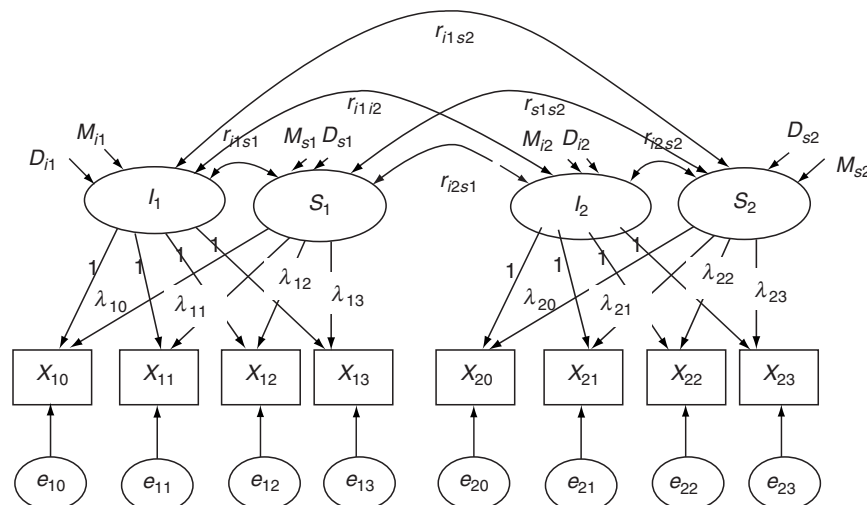


Figure 1 Path diagram of an unconditional bivariate latent growth curve model.

In this model, variable Z (in the SEM framework, variable Z can also be a latent variable measured by multiple indicators) is specified as a predictor of developmental trends, which are the intercepts and slopes of variables X_1 and X_2 , and variable Y is predicted by the same developmental trends, that is, intercepts and slopes of variables X_1 and X_2 . That is, individual-level variables can serve both as predictors of growth curves and as the outcome variables accounted for by the growth trends. This modeling feature is not available in the multilevel method of longitudinal analysis discussed earlier.

Data Illustration

The same data presented in the previous section were used to illustrate the use of multivariate growth curve

modeling. Table 4 presents numerical results. As shown in Table 4, the linear growth model yields a relatively good fit (chi-square = 61.504, $df = 22$; RMSEA = 0.082; CFI = 0.976; TLI = 0.969. Please note that the HLM model discussed earlier do not have these goodness-of-fit statistics). Some important parameter estimates of this model include initial mean levels of intercept (2.689 and 2.655 for cognitive self-concept (CC) and general self-concept (GC), respectively), deviations from the mean intercept (0.262 and 0.250 for CC and GC, respectively), slope means (-0.269 and -0.334 for CC and GC, respectively), slope deviations (0.030 and 0.032), the correlation between the two intercepts (0.418), and the correlation between the two slopes (0.093). The results of latent linear growth model are essentially identical to those obtained using multivariate multilevel method that is presented in

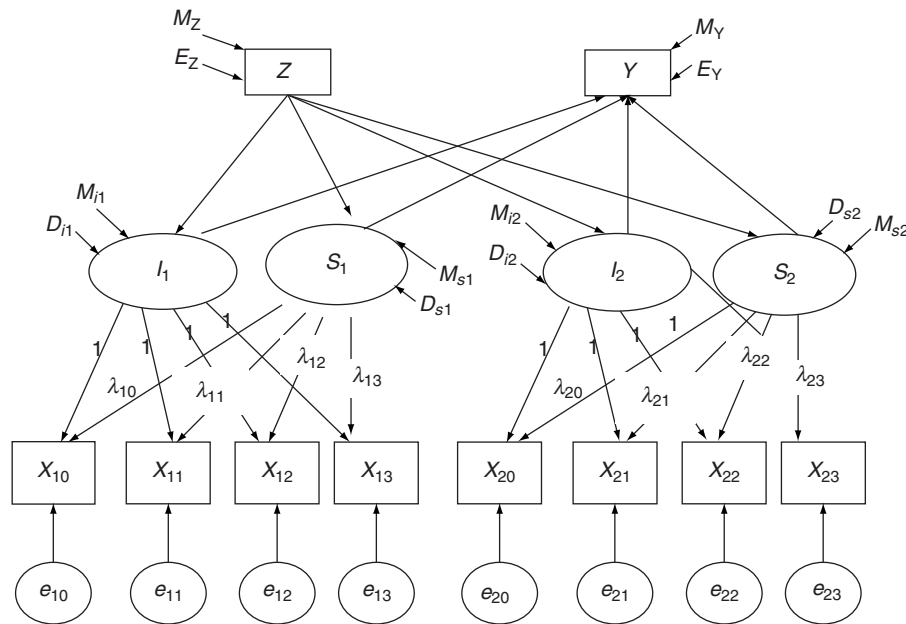


Figure 2 Bivariate LGM with individual-level predictor and outcome variables.

Table 4 Results from bivariate latent linear growth models for cognitive self-concept (CC) and general self-concept (GC)

<i>Fixed part parameters</i>	<i>Estimate</i>	<i>SE</i>	<i>t</i>	
Latent means				
CC Intercept	2.689	0.033	81.725	
CC Slope	−0.269	0.012	−21.885	
GC Intercept	2.655	0.033	80.695	
GC Slope	−0.334	0.013	−26.485	
<i>Variance (diagonal), covariance (above diagonal), standard error (in bracket) and correlation (below diagonal) of the random intercepts and slopes</i>				
<i>Random part parameters</i>	<i>CC intercept</i>	<i>CC slope</i>	<i>GC intercept</i>	<i>GC slope</i>
CC intercept	0.262 (0.025)	−0.018 (0.007)	0.107 (0.019)	0.002 (0.007)
CC slope	−0.206	0.030 (0.004)	−0.009 (0.007)	0.003 (0.003)
GC intercept	0.418	−0.098	0.250 (0.025)	−0.025 (0.008)
GC slope	0.027	0.093	−0.280	0.032 (0.004)

Table 2. The mirror differences are attributable to the use of the different software and fitting algorithms.

Gender was then included as an individual-level predictor of change. The results are presented in **Table 5**. Again, the results of gender as a predictor of the intercept and slope of growth are identical to those obtained earlier using the multilevel model. In this example, we did not include an outcome variable on the individual level to be predicted by the growth models. As shown in **Figure 2**, such variables can also be included in the growth curves model. Finally, this conditional multivariate latent growth model showed sound goodness-of-fit indicators (chi-square = 66.493, $df = 26$; RMSEA = 0.077; CFI = 0.976; TLI = 0.967).

Structural Cross-Lagged Regression Model

The structural cross-lagged longitudinal model based on the pioneering work by Jöreskog and Sörbom (1979) is depicted as a path diagram in **Figure 3**.

Table 5 Results from including gender as a predictor of growth trends of cognitive (CC) and general self-concept (GC) using multivariate latent growth model

Dependent variable	Estimate	SE	<i>t</i>
CC intercept	−0.308	0.063	−4.867
GC intercept	−0.442	0.060	−7.392
CC slope	0.009	0.025	0.359
GC slope	0.049	0.025	1.925

The model in **Figure 3** includes two variables, X_1 and X_2 , measured on four occasions. The observed scores, X_{1t} , $t = 0, 1, 2, 3$ and X_{2t} , $t = 0, 1, 2, 3$, are linear functions of latent scores (true scores), ξ_{1t} , $t = 0, 1, 2, 3$ and ξ_{2t} , $t = 0, 1, 2, 3$, plus measurement errors, e_{1t} , $t = 0, 1, 2, 3$ and e_{2t} , $t = 0, 1, 2, 3$. These relations are defined as

$$X_{1t} = \xi_{1t} + e_{1t} \quad t = 0, 1, 2, 3$$

and

$$X_{2t} = \xi_{2t} + e_{2t} \quad t = 0, 1, 2, 3$$

Each latent variable at time t is a function of three parts: (1) an autoregression (β_{x1} and β_{x2}), which represents the effect of the same variable on itself across two consecutive time points and constrained to be equal across times; (2) a cross-lagged regression (γ_{12} and γ_{21}), which represents the effect on one variable at a later time from the other variable at a previous time, and is also constrained to be equal across times; and (3) a residual which is allowed to correlate with the residual of the other variable at the same time point and is constrained to be equal across times. These relations are defined as

$$X_{1t} = \beta_{x1} X_{1t-1} + \gamma_{21} X_{2t-1} + \xi_{1t} \quad t = 1, 2, 3$$

and

$$X_{2t} = \beta_{x2} X_{2t-1} + \gamma_{12} X_{1t-1} + \xi_{2t} \quad t = 1, 2, 3$$

According to these equations, the autoregression and cross-lagged regression effects are parameters affecting the observed variables but not the growth functions derived from the observed variables over time. This modeling

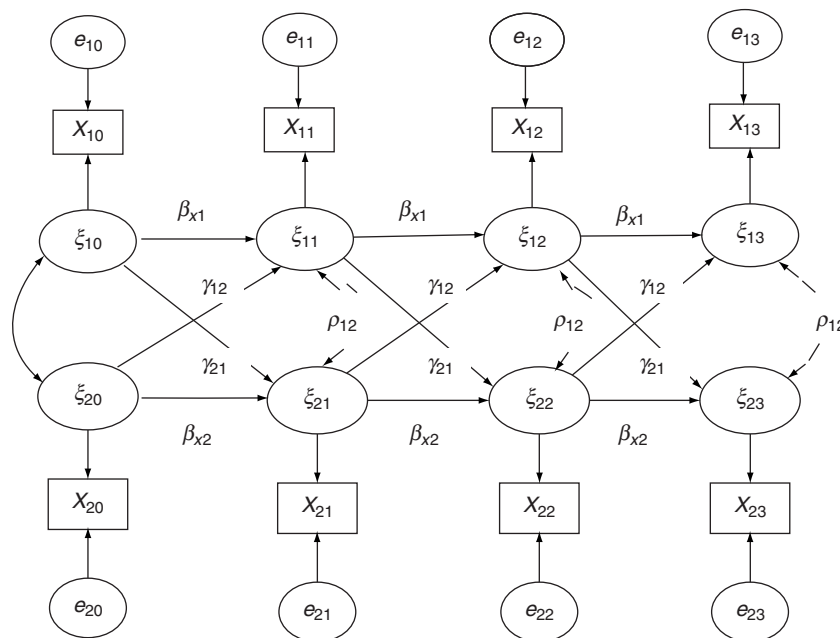


Figure 3 Path diagram of a bivariate cross-lagged regression model.

characteristic distinguishes the cross-lagged regression model from the other two growth models presented earlier. That is, we cannot directly estimate a growth function and directly account for the growth variations among individuals. The cross-lagged regression model approaches the growth or change question by examining each individual cross-lagged change. This cross-lagged change approach can be seen by re-writing the above equations into the first difference equations as the following:

$$\Delta X_{1t} = \beta_{x1}(\Delta X_{1t-1}) + \gamma_{21}(\Delta X_{2t-1}) + \Delta \zeta_{1t} \quad t = 1, 2, 3$$

and

$$\Delta X_{2t} = \beta_{x2}(\Delta X_{2t-1}) + \gamma_{12}(\Delta X_{1t-1}) + \Delta \zeta_{2t} \quad t = 1, 2, 3$$

where $\Delta X_{1t} = X_{1t} - X_{1t-1}$ represents the change of variable X_1 from time $t-1$ to time t , and $\Delta X_{2t} = X_{2t} - X_{2t-1}$ represents the change of variable X_2 from time $t-1$ to time t .

Again, the cross-lagged model defines growth by computing a series of intercorrelations to represent consecutive time change in observations but not by estimating a single statistical function to summarize the potential data trend from repeated observations. This characteristic makes the cross-lagged model not directly comparable with the other two models.

The structural cross-lagged regression model in **Figure 3** is the so-called unconditional model because we do not specify individual-level predictors to account for the series of cross-lagged intercorrelations. In fact, the model in **Figure 3** specifies only one variable that has repeated measurements over four time points. In such data structure, there is not even the distinction between predictor and outcome variables. They are both predictor and outcome variables depending on the time of measurement.

Like the other two growth models previously discussed, it is possible to formulate a conditional model by specifying individual-level predictors of the time series of intercorrelations. Different from the other two previously discussed models, the predictor variables can be time varying or time invariant. They can also be directly observed variables or latent variables derived from multiple indicators. **Figure 4** presents the simplest conditional model by including one directly observed and time-invariant predictor, Z . In the figure, Z is specified to predict the latent variables ξ_{10} and ξ_{20} but we can use Z to predict latent variables derived from all the time observations such as ξ_{11} , ξ_{21} , ξ_{12} , ξ_{22} , ξ_{13} , and ξ_{23} . In the latent cross-lagged regression model, the variable Z is defined to predict the measurements of different times but not to predict changes of measurements over time. The cross-lagged model is thus different from the other two growth models discussed earlier.

Data Illustration

The numerical results for multivariable longitudinal responses cross-lagged regression model is presented in **Table 6**. The model has satisfactory goodness of fit (chi-square = 30.112, $df = 26$; RMSEA = 0.024; CFI = 0.997; TLI = 0.997). As shown in **Table 6**, the autoregression effects for both variables are high (0.080 and 0.813 for CC and GC, respectively), suggesting strong stability across time. The cross-lagged regression from cognitive self-concept to general self-concept (0.080) is statistical significant, suggesting that the level of one self-concept in the previous year affects change in the other self-concept in the next year. The cross-lagged regression from general self-concept to cognitive self-concept is not significant

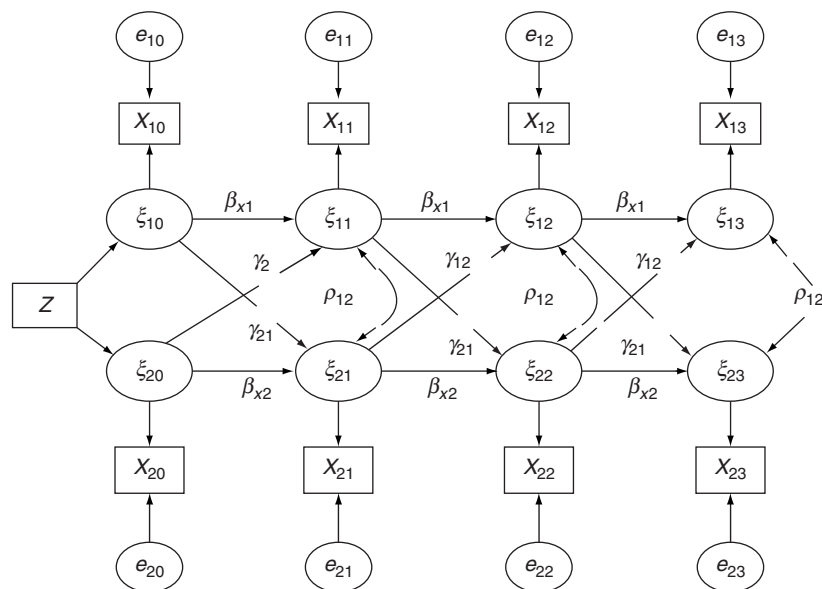


Figure 4 Path diagram of a bivariate cross-lagged regression model with one predictor variable.

Table 6 Results from cross-lagged bivariate model for cognitive self-concept (CC) and general self-concept (GC)

	<i>Estimate</i>	<i>SE</i>	<i>t</i>
Regression coefficient			
Auto-regression (CC)	0.866	0.024	36.509
Auto-regression (GC)	0.813	0.025	32.702
Cross-lag (CC → GC)	0.080	0.024	3.349
Cross-lag (GC → CC)	0.026	0.025	1.047
Variance			
σ_0^2 (CC)	0.279	0.024	11.489
σ_0^2 (GC)	0.285	0.025	11.489
σ_d^2 (CC)	0.141	0.007	19.900
σ_d^2 (GC)	0.143	0.007	19.900
Var_{dxy} (CC with GC)	0.002	0.005	0.492

Table 7 Results from including gender as an individual level predictor variable of cognitive self-concept (CC) and general self-concept (GC) using cross-lagged bivariate model

<i>Dependent variable</i>	<i>Estimate</i>	<i>SE</i>	<i>t</i>
CC at time 1	-0.294	0.062	-4.710
CC at time 2	-0.064	0.048	-1.345
CC at time 3	-0.040	0.047	-0.838
CC at time 4	0.052	0.047	1.107
GC at time 1	-0.454	0.059	-7.632
GC at time 2	0.001	0.048	0.025
GC at time 3	0.055	0.048	1.146
GC at time 4	-0.068	0.047	-1.433

(0.026). The variance of the two self-concept at the initial observation is high (0.279 and 0.285), whereas the residual variance is low (0.141 and 0.143), suggesting strong predictions across consecutive times. These results indicate that the changes in both variables are strongly influenced by their own changes, and there is a cross-lagged regression from cognitive self-concept to general self-concept.

Table 7 reports the results from the conditional model that includes gender as a directly observed individual-level predictor. Both the CC and GC are higher for girls, which was coded as 0, than for boys, which was coded as 1, (-0.294 and -0.454 for CC and GC, respectively). The direct effect of gender on cognitive self-concept and general self-concept from time 2 to time 4 are not significant (-0.064, -0.040, and 0.052 for CC; 0.001, 0.054, and -0.068 for GC). The cross-lagged regression model yields satisfactory goodness of fit (chi-square = 43.407, $df = 27$; RMSEA = 0.048; CFI = 0.991; TLI = 0.988).

Conclusion

In this article, we introduced three multivariate longitudinal methods namely the multilevel longitudinal model, latent growth curve model, and cross-lagged regression

model. A real data set was used to illustrate the operations of these methods, the interpretations of their results, and the differences and similarities among the three methods. Multilevel model analyzes longitudinal data within the HLM framework and the latent growth curve model analyzes longitudinal data within the SEM framework. However, these two methods use the same analytical approach to investigate developmental questions by estimating statistical functions of growth and by accounting for the variations across individuals in these growth functions. In our examples, we specified linear growth function by estimating the intercept and the linear slope. Other growth functions such as quadratic functions can be similarly estimated and modeled. These two methods yield identical results which are also presented in similar forms by the two programs. The cross-lagged method approaches growth differently by estimating a series of regression coefficients across times. This approach represents growth and development by focusing on changes over time and inter-correlations between variables across time. One advantage of the cross-lagged regression method that is absent from the other two methods is that individual-level variables may serve both as predictor and the outcome variables of the various regression coefficients representing change across time.

We presented the basic models of the three methods, whereas these models may be extended to include more advanced data analytic features (McArdle and Hamagami, 2001; Muthén *et al.*, 2002, 2003). One advanced data feature that all three methods have is to impute missing data. For the growth curve and cross-lagged method, missing data are usually handled by multiple imputations or multiple group analyses. The treatment of missing data is more complicated with the multilevel growth modeling method because of the nested nature of the data structure. However, the method can handle such unbalanced data where the space between repeated assessments is different among individuals. Finally, the examples we used are based on directly observed variables. To a different extent, all three methods can treat latent constructs measured by multiple indicators.

Bibliography

- Bryk, A. S. and Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage.
- Chang, L. (2003). Variable effects of children's aggression, social withdrawal, and prosocial leadership as functions of teacher beliefs and behaviors. *Child Development* **74**, 535-548.
- Collins, L. M. and Sayer, R. M. (2001). *New Methods for the Analysis of Change*. Washington, DC: American Psychological Association.
- Goldstein, H. (1995). *Multilevel Statistics Models*, 2nd edn. New York: Halsted.
- Harter, S. (1982). The perceived competence scale for children. *Child Development* **53**, 87-97.

- Jöreskog, K. G. and Sörbom, D. (1979). *Advanced in Factor Analysis and Structural Equation Models*. Cambridge, MA: Abt Books.
- McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In Cattell, R. B. and Nesselroade, J. (eds.) *Handbook of Multivariate Experimental Psychology*, 2nd edn, pp 561–614. New York: Plenum Press.
- McArdle, J. J. and Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development* **58**, 110–133.
- McArdle, J. J. and Hamagami, F. (2001). Linear dynamic analyses of incomplete longitudinal data. In Collins, L. and Sayer, A. (eds.) *Methods for the Analysis of Change*. Washington, DC: APA Press.
- Meredith, W. and Tisak, J. (1990). Latent curve analysis. *Psychometrika* **55**, 107–122.
- Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods* **8**, 369–377.
- Muthén, B., Brown, C. H., Masyn, K., et al. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics* **3**, 459–475.
- Muthén, B., Khoo, S. T., Francis, D., and Kim Boscardin, C. (2003). Analysis of reading skills development from Kindergarten through first grade: An application of growth mixture modeling to sequential processes. In Reise, S. R. and Duan, N. (eds.) *Multilevel Modeling: Methodological Advances, Issues, and Applications*, pp 71–89. Mahaw, NJ: Lawrence Erlbaum Associates.
- Rao, C. R. (1958). Some statistical methods for the comparison of growth curves. *Biometrics* **14**, 1–17.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edn. Newbury Park, CA: Sage Publications.
- Singer, J. D. and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.
- Tucker, L. R. (1966). Learning theory and multivariate experiment: Illustration by determination of parameters of generalized learning curves. In Cattell, R. B. (ed.) *The Handbook of Multivariate Experimental Psychology*, pp 476–501. Chicago, IL: Rand McNally.

Further Reading

- Bryk, A. S. and Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin* **101**, 147–158.
- Liu, H. and Chang, L. (2005). *Longitudinal Data Analysis Methods and Applications* (in Chinese). Beijing: Education and Science Press.
- McArdle, J. J. and Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development* **58**, 110–133.
- McArdle, J. J. and Nesselroade, J. R. (1994). Structuring data to study development and change. In Cohen, S. H. and Reese, H. W. (eds.) *Life-Span Developmental Psychology: Methodological Innovations*, pp 223–267. Hillsdale, NJ: Erlbaum.

Quasi-Experimentation: Two Group Design

E Hong, University of Nevada, Las Vegas, NV, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Causal inference – The conclusion made based on the findings of a data analysis that an event or events caused a subsequent event.

Control group – A group of research units that does not receive a treatment in an experiment.

Experimental group – A group of research units that receives a treatment in an experiment.

Matching – A research procedure that equates two or more groups on one or multiple extraneous variables prior to experimentation in order to rule out their effects on the relationship between treatment and outcome.

Nonequivalent control group design – A quasi-experimental design in which research units are not randomly assigned to experimental and control groups.

Quasi-experiment – An experiment in which research units are not randomly assigned to experimental conditions.

Randomized experiment – An experiment in which research units are randomly assigned to experimental conditions.

The three main objectives of this article are to: (1) illustrate the major types of two-group quasi-experimental designs, (2) identify major threats to the validity of causal inferences involved in each design, and (3) illustrate how some of these threats are controlled by specific features of the designs. Although the title of this article reads ‘Two-group designs’ – reflecting that research engages one experimental group and one control group – quasi-experimentations can involve more than two groups. For example, one treated group and multiple untreated groups with different conditions, two treated groups and one untreated group, other combinations of groups are possibilities.

As with randomized experiment (also called true experiment; Rosenthal and Rosnow, 1991), quasi-experiment includes interventions or treatments manipulated by the investigator, pretests and/or posttests, and a number of groups in the study. Similar to randomized experiments, quasi-experiments are used to examine cause-and-effect relationships by testing hypotheses about manipulated variables – that is, the researcher manipulates the cause, and the effect is measured following the manipulation.

The most distinguishing feature between quasi-experimentation and randomized experimentation is that, in the former, participants (units) are not randomly selected and, often, they are not randomly assigned to treatment conditions. In educational research, randomized experimentation is limited for investigating problems encountered in classrooms and schools. It is not unusual that randomization is impossible to employ because randomization procedures can be disruptive to groups of individuals in existing conditions (e.g., classes or schools). Thus, experiments are conducted on volunteers, with some units selecting treatment for themselves (self-selection) or teachers, principals, or others deciding which units receive which treatment (administrator-selection) (Shadish *et al.*, 2002, 2007).

Although researchers have some controls over their scheduling of data-collection procedure, such as selecting measures or scheduling treatment, units in the control condition in quasi-experimentation may be different from those in the treatment condition in many nonrandom ways beyond the intervention provided to the experimental group (Campbell and Stanley, 1966; Shadish *et al.*, 2007). Therefore, manipulation or control is possible only on some of the relevant variables in quasi-experimentation.

In these situations, observed treatment effects may be caused by some sources – other than the treatment itself – that differ systematically. More threats to the validity of a causal inference between treatment and outcomes exist in quasi-experimentation than in randomized experimentation. Some systematic variations generated by sources other than the treatment could be alternative explanations for the observed effect. Due to these threats, findings from quasi-experimentations are not as dependable as those from randomized experimentations. Thus, the researcher using quasi-experimental designs must document potential sources for alternative explanations as closely as possible (e.g., students’ ability or previous achievement, students’ socioeconomic level, teachers’ experiences) and take measures to assess whether any of the alternative explanations account for the observed effect. To rule out possible alternative or rival hypotheses, the researcher may use a modified design that can detect alternative explanations.

In brief, educational researchers conducting quasi-experimentation should endeavor to build as much control as possible into the design, use available information to assess the similarity between experimental and control groups, and interpret outcomes in the context of the

information employed and the experimental conditions (Wiersma and Jurs, 2005).

This article introduces two-group designs (treatment and control groups) with or without pretests, their sources of invalidity, and ways to control some of the threats these designs incur. This article is divided into two sections: (1) two-group designs without pretest (i.e., posttest-only) and (2) two-group designs with pretest and posttest.

Two-Group Designs with Posttest-Only

As this article only includes quasi-experimentation, all groups are assumed to be formed nonrandomly. To indicate features of quasi-experimental designs diagrammed throughout this article, the following notations are used. The temporal order is indicated by the position of entities – from left to right. The broken line in the diagrams separates nonequivalent groups (experimental group vs. control group) – that is, since randomization has not been employed, the equivalence of units in the two groups is not assured.

E Experimental group

C Control group

X Treatment

O Observation (pretest or posttest)

Posttest-Only Design with Nonequivalent Groups

E X O₁

C O₂

In this design, participants in only the experimental group receive treatment (*X*). This design is weak and may not be used unless some data are available that can be used to assess the extent of pretreatment similarity between the two groups. Participants' initial differences in some important variables may have differential effect on observations (*O₁* and *O₂*). For example, participants in one group may be more motivated, more intelligent, or different from those of the other group in some systematic ways. Thus, selection bias is the most serious threat to the validity of a causal inference made from research results of this design and may be the most plausible alternative explanation to the observed effect. This design may best be employed when participants in the two groups are comparable and similar on the trait being measured before the treatment is given to the group, with documented data to examine the trait in question. Other potential threats to validity include history, instrumentation, maturation, subject attrition, diffusion of treatment, experimenter effects, subject effects, and statistical conclusion.

One reason for using this weak design over pretest–posttest design could be that pretest may sensitize participants and thus affect their responses to the posttest. For example, racial prejudice measured prior to a treatment (e.g., positive behavioral messages) may sensitize participants, thus affecting their responses to the racial prejudice measure observed following the treatment. If a treatment effect is found in this instance, the extent of the testing effect is difficult to isolate from the treatment effect.

However, this testing effect may be reduced when different treatment groups are compared, because the threat to internal validity (testing effect) is constant across groups. Shadish *et al.* (2002) summarize how the testing effects can be reduced, including examples of studies that employed a variety of control approaches (p. 116): administering alternative forms of a test (one at pretest and the other at posttest), using item-response theory (IRT) to calibrate different tests to the same scale, lengthening the time interval between pretest and posttest, using a Solomon Four Group Design to assess the presence and impact of such effect, using unobtrusive measures that are less reactive than self-report, using techniques such as the bogus pipeline, using retrospective pretests, and using explicit reference groups or behavioral criteria to anchor responding.

Shadish *et al.* (2002, 2007) suggest several approaches to improving posttest-only design with nonequivalent groups. This article illustrates three of the approaches that are, largely, applicable to educational research: (1) posttest-only design using proxy pretests; (2) posttest-only design using matching; and (3) posttest-only design using multiple control groups.

Posttest-only design using proxy pretests

E O_{A1} X O_{B2}

C O_{A1} O_{B2}

Instead of gathering pretest and posttest data on the same test, one can measure proxies for pretests. Proxies are variables that are conceptually related to – and correlated with – the posttest (Cook and Campbell, 1979; Shadish *et al.*, 2002, 2007). In the diagram above, *A* represents the proxy pretest and *B* the posttest. A misconception, often, found in the use of proxies is utilizing measures that may be related to an outcome but are not conceptually related to the outcome specific to the study. At times, proxy pretests are more relevant than a pretest on the outcome variable itself. For instance, when evaluating a statistics course with a medium-level difficulty for students who have not studied such statistics, a pretest on the medium-level statistics may not produce much variability. Instead, a proxy pretest of introductory-level statistics or

mathematical aptitude may be a better choice. Indexing selection bias and attrition bias will depend on the extent of correlation between proxies and the posttest.

Posttest-only design using matching

$E_{\text{matched}} X O_1$

$C_{\text{matched}} O_2$

Another way to reduce selection bias due to having no pretests is employing a matching approach to forming treatment and control groups. Matching is performed on correlates of the posttest. In the diagram above, E_{matched} and C_{matched} represent the two groups (E and C) that were matched on some correlates of the posttest.

The purpose of matching (or blocking) is to form the two groups with participants whose characteristics on the matching variable are similar, thus yielding similar scores on the matching variable. As with proxy pretests, matching variables should be correlated with the posttest. The matching technique has been used frequently in twin studies (Ashenfelter and Krueger, 1994).

Shadish *et al.* (2002) list various methods used for matching (p. 119). Exact matching yields exactly the same score for units within a match – which is difficult to achieve when samples are small or if the distribution of participants in each group on the matching variable is unbalanced. When matching variables are measured in very fine gradations, exact matching may not be a good choice. Caliper matching uses scores within a defined distance of each other. The distance can be measured in various ways, yielding information for Mahalanobis distance matching or nearest-neighbor matching (Hill *et al.*, 2000; Rosenbaum, 1995). Other matching methods include index matching, cluster group matching, benchmark group matching, and optimal matching (Shadish *et al.*, 2002, 2007).

Although matching is used to decrease selection bias, in quasi-experimentation its use can be quite problematic. Matching can be a very useful addition to random assignment. However, matching in quasi-experimentation is, often, not effective when it is not conducted carefully. For example, matching performed on an unstable or unreliable variable can be more harmful than helpful (Shadish *et al.*, 2002). When the populations being matched do not overlap completely on the matching variable, the outcome may result in an unpredicted direction (Campbell and Erlebacher, 1970). This occurs more often when the matching variable has a large amount of measurement error and has an imperfect correlation with outcome. In addition, with an incomplete overlapping population for matching variable, statistical regression can occur. Marsh (1998; cited in Shadish *et al.*, 2002) compared a gifted and talented population with a control population for the evaluation of programs for gifted and talented children.

Matches were obtained from the overlap between the lower end of the gifted and talented population and the upper end of the control population (i.e., match did not overlap completely). Statistical regression occurred at the posttest – with children attending the gifted and talented program regressing upward toward their population mean and those in the control group regressing downward – making the gifted and talented programs look more effective than they might otherwise have been. Using as many relevant variables as possible is also an important item to consider for matching.

It is important to form experimental and control groups, prior to matching, that are as similar as possible (D'Agostino and Kwan, 1995). As indicated, matching on unstable and unreliable variables is more harmful than useful. One can improve reliability of matching variables by creating a composite of many variables taken at the same time. For example, the researcher may use an average of several test scores measured prior to the treatment as a matching variable. When using more than one variable, a multivariate composite may be used for matching. The importance of using better matching procedure cannot be stressed enough. As Shadish *et al.* (2002) conclude, “The days of simple matching on single variables that are not reliably measured should be left completely behind” (p. 122).

Posttest-only design using multiple control groups

$E X O_1$

$C_A O_{A2}$

$C_B O_{B2}$

... ..

$C_N O_{N2}$

In the diagram above, A through N represent multiple control or comparison groups. The control groups do not receive treatment, whereas the experimental group does (X). However, these control groups are selected from populations with similar characteristics of the experimental populations except for certain conditions that separate them from the experimental group. For example, Bell *et al.* (1995; cited in Shadish *et al.*) compared an experimental group that received job training with four comparison groups. Research participants in the four groups had distinct characteristics that separate them from those of the experimental group: those who failed to apply for the program; rejected applicants; accepted applicants who failed to show; and those who started treatment but dropped out. Including such groups in a study enables the researcher to index the magnitude of hidden biases that may present (Rosenbaum, 1995). In the study by Bell *et al.* (1995), if the four comparison groups are as different from one another as they are from the treatment group, the result indicates that the treatment effect is not significant.

In many research situations, the direction of bias in each control group is known when compared to the treatment group. In such cases, the researcher may be able to bracket the treatment effect within a range of known biases. Bracketing controls – advanced by Campbell (1969) and Shadish *et al.* (2002, 2007) – involves selecting a group predicted to outperform the treatment group and another group predicted to underperform it. The finding on the treatment effect becomes strengthened if the treatment group outperforms both groups.

Two-Group Designs with Pretest and Posttest

Although including control groups in quasi-experiments (designs described above) is useful in making causal inferences, designs with a control group – as well as pretest – increase chances to detect the cause-and-effect relationship, if, in fact, there is such a relationship. Pretest measures can be taken on the same outcome variables as the posttest or on variables that are highly related to the outcome. Pretests provide the researcher with information that can be used to detect threats to validity (e.g., selection bias) by revealing whether and to what extent the two groups differ initially.

Pretest–Posttest Control Group Design

$E O_1 X O_2$

$C O_1 O_2$

This is the most common design widely used in quasi-experimentation. The design has been previously termed the “nonequivalent control-group” design by Campbell and Stanley (1966), the “untreated control-group design with pretest and posttest” by Cook and Campbell (1979), and “untreated control-group design with dependent pretest and posttest samples” by Shadish *et al.* (2002, 2007). As indicated in the diagram, the experimental group (E) receives treatment (X), whereas the control group (C) does not. Because data on both pretest and posttest are collected on the same participants, samples become dependent. Although selection bias exists due to using nonequivalent (nonrandomized) groups, the researcher can explore the direction and magnitude of the selection bias using pretest measures. Pretest scores can be used for statistical control or to generate gain scores (Wiersma and Jurs, 2005). Threats to the validity of this design include selection, statistical regression, and instrumentation. In addition, the existence of pretest differences indicates that selection may combine with other treatments additively or interactively. Such interactive threats include selection-maturation,

selection-instrument, selection-regression, and selection-history (Shadish *et al.*, 2002).

A selection-maturation interaction-threat exists if participants in one group learn faster or grow more tired than participants in another group. Shadish *et al.* illustrate patterns of research outcomes that may reveal the plausibility of threats. A fan-spread model of maturation occurs when the nonequivalent groups grow apart over time – from the center out to the edges at different average rates in the same direction. Fan-spread happens because measured variances grow systematically over time. Standardization of scores may reduce the fan-spread effect (Shadish *et al.*, 2002). However, when self-selection or administrator-selection into the treatment group represents selection bias in such a way that treatment condition is selected for participants who are, for example, more motivated to improve, it is likely that the treatment group improves at a faster rate for reasons that are not related to treatment. In such instances, score transformation will not be able to differentiate the treatment effect from this alternative explanation (i.e., selection-maturation interaction). Shadish *et al.* offer a few analytic clues for detecting a fan-spread type of maturation (p. 140). First, if the selection-maturation interaction is the reason for the mean difference between groups, then the differential growth shown between the experimental and control groups should also take place within groups. If higher-performing members of the group with the higher pretest scores grow faster than lower-performing members of the same group, then the selection-maturation threat may be the alternative explanation. Second, plot pretest scores against possible maturational variables (e.g., years of teaching experience) for the two groups separately. If the slope of the regression equation is not homogeneous, growth rates of the two groups differ. This differential growth is not due to treatment, obviously, since the treatment has not taken place yet.

Another outcome of possible selection-maturation interaction is demonstrated when change or growth does not occur in the control group (Shadish *et al.*, 2002). One might question why is it that behavioral change occurred only in the experimental group while even a slower rate of behavioral change did not take place in the control group? Again, analyses performed within groups may help detect the plausibility of selection-maturation threat. If, for example, years of teaching experience in a study with teachers may be the possible variable causing the differential growth between two groups (i.e., the treatment group has more teachers with longer careers in education), then the number of years can be used to split the data. On the other hand, if participants in the treatment group demonstrate behavioral changes whether they are novice or experienced teachers, the plausibility of selection-maturation explanation decreases.

Another pattern of outcome that involves selection-maturation threat is when the control group outperforms

the treatment group. Although the causal relationship from this type of outcome is mostly interpretable, the reasons for the initial difference should be carefully examined for possible selection-maturation threat. Using the same example, teachers in the control group had more years of teaching experience as compared to teachers in the experimental group.

As mentioned earlier, additional threats caused by selection bias combined with other sources of invalidity include selection-instrumentation, selection-regression, and selection-history. A selection-instrumentation threat occurs when the instrument (e.g., rater) changes differently for the experimental and control groups. For example, two observers rate classroom behavior of children prior to and following the introduction of a behavior-modification intervention to the experimental group. The differential change in the instrument occurs when the observer in one group becomes bored or tired over time. In this case, the mean difference found in the posttest could be the result of selection-instrumentation effect. The selection-instrumentation threat can be increased when the pretest scores have different beginning points – especially when the extent of initial group difference is large, the pretest-posttest change is great, and floor or ceiling effects occur (i.e., group means are close to one end of the scale).

A selection-regression threat occurs when one group has more extreme scores on the pretest than the other group. For example, in a study that experiments with a new instructional approach to reading, a group gets a disproportionate number of children with low reading ability. This group will likely have a posttest mean that will regress toward the population mean, thus giving the appearance that this group gained more than did the other group, although the gain may not be from the treatment.

Lastly, a selection-history threat occurs when an event or events occur between pretest and posttest that affect one group more than another. Participants in the two groups who differ initially (i.e., selection bias) react differently to an event (i.e., history). In the study introduced above – a new instructional method for reading – the experimental group happens to be composed of quite a few children whose parents tutor for reading. If the mean posttest score of the experimental group were higher than the control group, it could be misleading since one cannot isolate the effect by the treatment.

Shadish *et al.* (2002, 2007) describe several approaches to improving pretest-posttest design with nonequivalent groups. This article illustrates three designs that are largely applicable to educational research: (1) pretest-posttest control group design with multiple pretests; (2) pretest-posttest control group design using switching replications; and (3) pretest-posttest control group design using a reversed-treatment control group.

Pretest-posttest control group design with multiple pretests

$E\ O_1\ O_2\ X\ O_3$

$C\ O_1\ O_2\ O_3$

A two-pretest design is illustrated in this section. O_1 and O_2 are the same pretest administered at two different time points. Note, in the diagram, that the intervals between O_1 and O_2 and between O_2 and O_3 (posttest) are the same – which is a preferred method in analyzing the treatment effect as it provides information to address possible threats to validity. A selection-maturation threat can be examined in this design for the control group by comparing the growth rates between O_1 and O_2 and between O_2 and O_3 (posttest). Whether the rate of growth from O_1 to O_2 continues between O_2 and O_3 without the treatment can be assessed, although changes in the instrument between the two intervals make the comparison less ideal. In addition, the group difference in the pretreatment growth can be examined with the addition of the second pretest in order to evaluate the plausibility of selection-maturation threat. Presence of regression effects can also be assessed with the double pretest if the O_2 score is uncharacteristically low or high compared to O_1 , in either group (Cook and Campbell, 1979; Shadish *et al.*, 2002).

Pretest-posttest control group design using switching replications

$E/C\ O_1\ X\ O_2\ O_3$

$C/EO_1\ O_2\ XO_3$

With this design, both groups receive treatment but at different time points. The experiment is replicated in two phases with the role of group switching as experimental or control group (i.e., switching replications). Note, however, that the first phase and the second phase do not provide the same experimental condition. Treatments are rendered to two groups at different time points – with the first group receiving treatment following one pretest, whereas the second group receives treatment following two pretests. Likewise, the first group is administered the posttest twice, whereas one posttest is used for the second group. These different conditions create contextual differences for the two groups for both treatment and testing conditions. However, this modified replication still helps the evaluation of threats to validity. Potential threats include history and compensatory rivalry when the treatment is removed from the group that the treatment was first introduced to.

Pretest–posttest control group design using a reversed-treatment control group or groups

$E O_1 X_+ O_2 \quad E O_1 X_- O_2$

$C O_1 X_- O_2 \quad C_A O_{A1} X_- O_{A2}$

$C_B O_{B1} O_{B2}$

In both diagrams, X_- represents a conceptually opposite treatment expected to reverse the effect from the original treatment (X_+) predicted by the research to produce an effect in a direction (Shadish *et al.*, 2002) – that is, a treatment effect is supported by a statistical interaction, with posttest scores of the two groups pointing at different directions. With this design, the causal construct should be defined and manipulated carefully to develop a test that captures the outcomes that run opposite ways. As the control group receives a reversed treatment, some plausible threats to validity can be ruled out. For example, a pretest–posttest difference can be attributed to the treatment and not to participants' reactivity to the experimental situation (i.e., participants' perceptions of the experimental situation are part of the treatment tested) or novelty effects (participants respond well to a novel experiment). A reversed treatment is not feasible for situations where such treatment causes ethical and practical concerns.

The second design (the diagram on the right-hand side) that includes an untreated control group is useful when X_+ and X_- do produce differential outcomes in two groups but in the same direction. Inclusion of an untreated control group will shed more light on the direction of the outcomes.

Conclusions

Quasi-experimentation is an important research approach – especially when randomization is not reasonable. Thus, experimental research conducted in field settings (classrooms, schools, or districts) employs mostly quasi-experimental design. Due to the nonrandomization in quasi-experimentation, threats to the validity of causal inferences are numerous. Apparently, for quasi-experimental research to be meaningful and useful for improving educational practices and providing support for various educational theories, substantial improvements of the commonly used quasi-experimental design are necessary. This article illustrates only two basic

designs for two-group situations and provides improved designs that may help researchers assess various threats involved in quasi-experimentation.

See also: Validity.

Bibliography

- Ashenfelter, O. and Krueger, A. B. (1994). Estimates of the economic returns to schooling from a new sample of twins. *American Economic Review* **84**, 1157–1173.
- Bell, S. H., Orr, L. L., Blomquist, J. D., and Cain, G. G. (1995). *Program Applicants as a Comparison Group in Evaluating Training Programs*. Kalamazoo, MI: Upjohn Institute for Employment Research.
- Campbell, D. T. (1969). Prospective: Artifact and control. In Rosenthal, R. and Rosnow, R. L. (eds.) *Artifact in Behavioral Research*, pp 351–382. New York: Academic Press.
- Campbell, D. T. and Erlebacher, A. E. (1970). How regression artifacts can mistakenly make compensatory education program look harmful. In Hellmuth, J. (ed.) *The Disadvantaged Child: Volume 3. Compensatory Education: A National Debate*, pp 185–210. New York: Brunner/Mazel.
- Campbell, D. T. and Stanley, J. C. (1966). *Experimental and Quasi-Experimental Design for Research*. Chicago, IL: Rand McNally.
- Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago, IL: Rand McNally.
- D'Agostino, R. B. and Kwan, H. (1995). Measuring effectiveness: What to expect without a randomized control group. *Medical Care* **33** (supplement), AS95–AS105.
- Hill, J. L., Rubin, D. B., and Thomas, N. (2000). The design of the New York school choice scholarship program evaluation. In Bickman, L. (ed.) *Validity and Social Experimentation: Donald Campbell's Legacy*, vol. 1, pp 155–180. Thousand Oaks, CA: Sage.
- Marsh, H. W. (1998). Simulation study of non-equivalent group-matching and regression-discontinuity designs: Evaluation of gifted and talented program. *Journal of Experimental Education* **66**, 163–192.
- Rosenbaum, P. R. (1995). *Observational Studies*. New York: Springer.
- Rosenthal, R. and Rosnow, R. L. (1991). *Essentials of Behavioral Research: Methods and Data Analysis*. New York: McGraw-Hill.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2007). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, 2nd edn. New York: Houghton Mifflin.
- Wiersma, W. and Jurs, S. G. (2005). *Research Methods in Education: An Introduction*, 8th edn. New York: Pearson.

Further Reading

- Kirk, R. E. (1995). *Experimental Design: Procedures for the Behavioral Sciences*, 3rd edn. Pacific Grove, CA: Brooks/Cole.
- Maxwell, S. E. and Delaney, H. D. (2004). *Designing Experiments and Analyzing Data: A Model Comparison Perspective*, 2nd edn. Mahwah, NJ: Erlbaum.

Regression Discontinuity Designs

A Kelava and S Rohrmann, Technical University Darmstadt, Darmstadt, Germany
V Hodapp, Goethe University Frankfurt am Main, Frankfurt, Germany

© 2010 Elsevier Ltd. All rights reserved.

Sometimes there are research situations when the realization of a randomized experiment is practically impossible or ethically not justifiable. For example, in educational research, a reading training program should be given to children who have lower scores in a reading achievement test instead of giving the training program to those children who do not need it. Therefore, programs are given to children below a cutoff score and not given to children above the cutoff score.

In such research situations, regression discontinuity (RD) designs allow for an unbiased estimation of treatment effects when participants of a quasi-experimental study are assigned to conditions on the basis of a cutoff score (for an excellent introduction see Shadish *et al.*, 2002).

The Principle of RD Designs

The key idea of the RD design is that the experimenter controls the assignment of participants to two or more conditions. The experimenter assigns participants to conditions on the basis of a cutoff score of an assignment variable that was measured before the treatment. Participants scoring below a given cutoff score are assigned to one condition and participants scoring equal or greater than the cutoff score are assigned to another condition. In contrast, in a randomized experiment, the assignment is determined by a coin toss or another random drawing which is uncontrolled by the experimenter.

The basic RD design can be summarized as a simple pre-post design (see **Table 1**).

In the RD design, the assignment variable (pretest) and the outcome variable (posttest) need not be the same measure. They can be completely different and independent measures. However, the assignment variable O_{assign} has to be at least an ordinal measure. Nominal variables, for example, dichotomous variables, cannot be used as assignment variables.

Figures 1 and **2** illustrate a hypothetical RD study. The assignment variable could be a pretest on a fictive outcome, for example, a reading test. Children scoring less than a cutoff score, here set at 50, are assigned to the treatment group and children scoring equal or greater than the cutoff are assigned to the control group. In **Figures 1** and **2**, the assignment variable (pretest) is given as the abscissa and the outcome variable (posttest) is given as the ordinate. These two figures graph a scatter

plot of the assignment variable against the outcome variable. A vertical line separates the treatment group from the control group.

Suppose that the training was not effective (see **Figure 1**). In this case, the assignment variable scores (pretest) and outcome scores (posttest) would be linearly related across both conditions with a given amount of correlation, here set at 0.90. There would be no substantive change of the scores except for differences that are due to unreliable measures (changes due to maturation, etc. are excluded).

However, if there is a (constant) effect of the treatment, participants in the treatment condition show increased outcome scores. In **Figure 2**, the treatment effect is set to ten score points. This means that the children's reading achievement improves by ten points. This change is equivalent to a ten-point score shift and results in a vertical displacement (discontinuity) at the cutoff. The slopes of the two regression lines are identical (parallel). The discontinuity (the vertical displacement) gives the RD design its name.

A true treatment effect will cause an upward or downward displacement in the regression line of the treatment group that relates the assignment to outcome. The displacement can be either a change in mean or a change in the slope of the treatment group's regression line. A change in mean indicates a constant change of the outcome score for all participants in the treatment group. A change in the slope indicates that the treatment is more effective for some values of the assignment variable than for others. **Figure 3** shows a change of slope in the context of the reading training program example presented above. As can be seen, there are two effects. First, there is a treatment effect that shifts the outcome score values. This means that children receiving the training program improve in reading ability, regardless of their pretest score. Second, there is an interaction between the assignment variable and treatment. This means that the effect of the training is stronger for the children who had higher assignment scores.

Model Specification within the RD Design

In order to estimate the treatment effect within the RD design, an ANCOVA-like regression model has to be specified. In the simplest (linear) case, when a constant treatment effect is expected, eqn [1] can be estimated (see **Figures 1** and **2**):

$$O_{\text{post}} = \beta_0 + \beta_1 \text{Condition} + \beta_2(O_{\text{assign}} - O_{\text{cutoff}}) + \varepsilon \quad [1]$$

where β_0 is an intercept, β_1 is the treatment effect of the binary variable Condition that records which treatment the participant received (0 = control, 1 = treatment), β_2 is the effect of the assignment variable O_{assign} (centered at the cutoff O_{cutoff}), and ε is the error of prediction.

As can be seen in eqn [1], the selection bias, which is due to the nonrandom assignment to conditions, is controlled by specifying the effect of the assignment variable O_{assign} .

In **Figure 1**, there is no treatment effect, which is equivalent to $\beta_1 = 0$. In **Figure 2**, there is a ten score points treatment effect, which corresponds to $\beta_1 = 10$ in eqn [1]. The intercept and the effect of the assignment variable are $\beta_0 = 50$ and $\beta_2 = 1$, respectively.

In the nonlinear case (see **Figure 3**), when an interaction between treatment variable and assignment variable is expected, eqn (2) can be estimated:

$$O_{\text{post}} = \beta_0 + \beta_1 \text{Condition} + \beta_2(O_{\text{assign}} - O_{\text{cutoff}}) + \beta_3 \text{Condition}(O_{\text{assign}} - O_{\text{cutoff}}) + \varepsilon \quad [2]$$

where β_3 is the coefficient of the interaction term. In **Figure 3**, the interaction effect corresponds to $\beta_3 = 0.5$. The intercept is $\beta_0 = 50$, the treatment effect is $\beta_1 = 35$, and the effect of the assignment variable is $\beta_2 = 1$.

Requirements for the RD Design

Assignment Variable

The RD design requires that the assignment to the conditions is based only on the cutoff score and not on any other variables or influences. The assignment rule (based on the cutoff) must not be overridden by removing a participant from one condition and assigning him to another. Otherwise, the validity of the experiment is impaired.

In order to have a valid design, the assignment variable must not be caused by treatment. Therefore, the assignment variable has to be measured before the start of the treatment. The assignment can be a pretest of the outcome variable, but does not have to be. It can even be totally unrelated to the outcome and have no substantive meaning within the research situation. For example, the assignment could be based on the date of application to a program (e.g., Cain, 1975). In this case, the RD design functions like a randomized experiment in which the assignment process is unrelated to the outcome (like a coin toss). Therefore, the RD design works whether the assignment variable is related or not to the outcome variable.

The assignment variable has to be a continuous variable or at least a (monotonically increasing) ordinal variable,

Table 1 Basic RD design

Pretest	Cutoff	Conditions	Posttest
O_{assign}	X	T	O_{post}
O_{assign}	X	C	O_{post}

O_{assign} = pretest measure of the assignment variable.

X = participants are assigned to conditions T or C based on a cutoff score.

T = treatment condition; C second treatment (e.g., control condition).

O_{post} = posttest measure of the outcome variable.

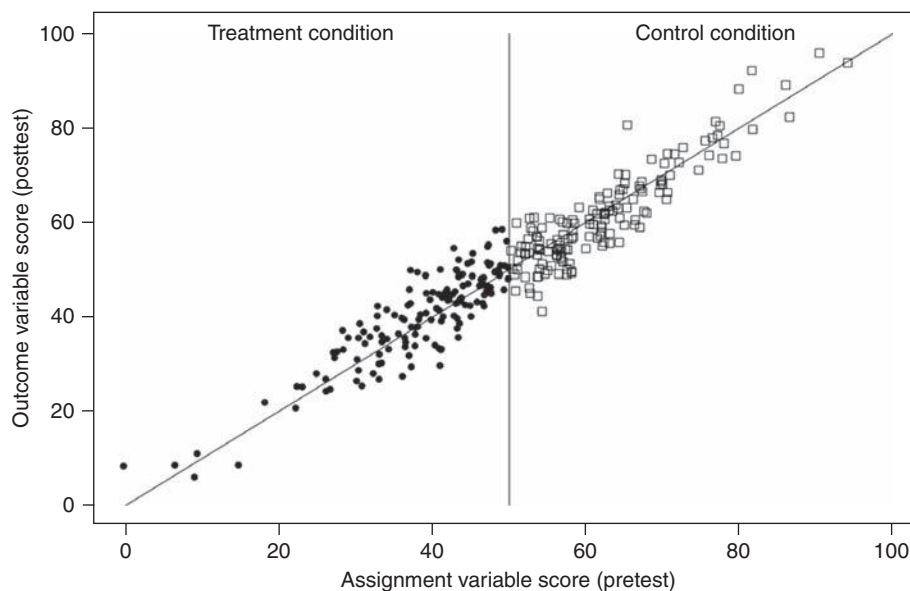


Figure 1 Regression discontinuity study with no treatment effect; no effect of the training program.

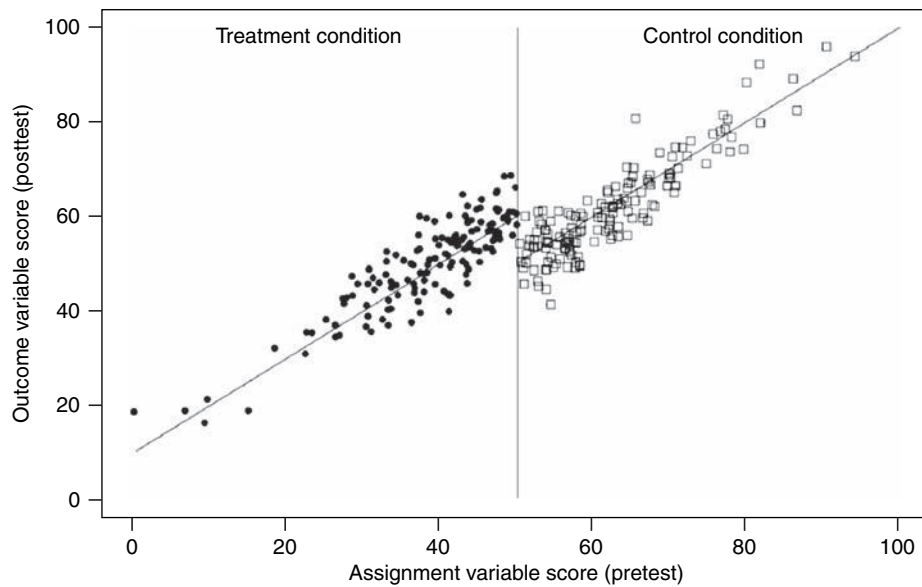


Figure 2 Regression discontinuity study with treatment effect; effective training program.

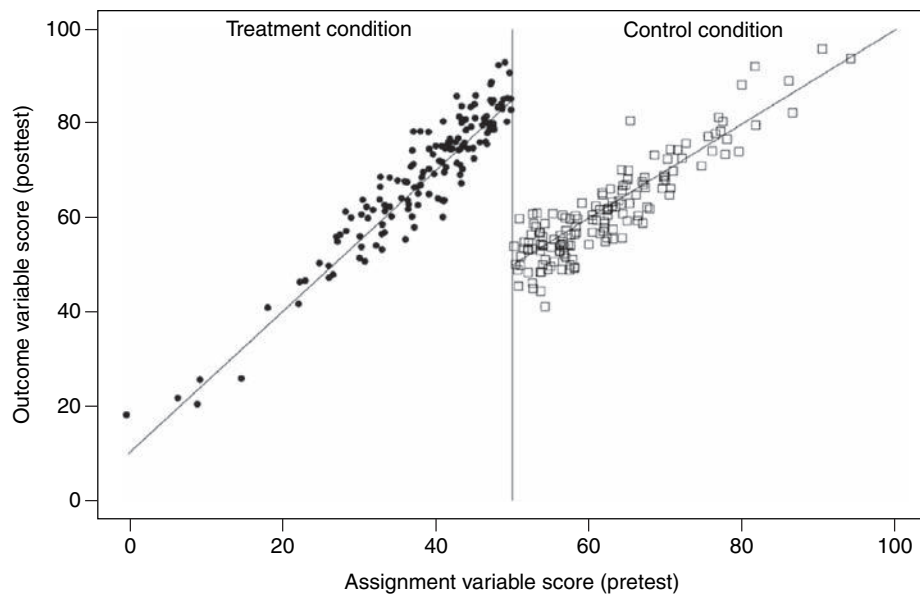


Figure 3 Regression discontinuity study with treatment and interaction effect.

such as income or test achievement scores. Dichotomous variables, such as gender, are not allowed, because they do not permit the estimation of a regression (line) within each group. There would be no variation within each group which makes the estimation of a regression line impossible. Therefore, the assignment variable should be continuously distributed with sufficient variance.

Cutoff Point and Additional Requirements

The selection of a cutoff point depends on several considerations. It should be chosen on substantive consideration,

such as professional opinion about who needs a medical treatment or which children need compensatory education. As shown by Cohen (1983), the statistical power and the estimation of interaction effects are both facilitated if the cutoff is set at the mean of the distribution of the assignment variable scores. However, there can be difficulties locating the cutoff at the mean when participants are arriving one after the other in the study so that the mean cannot be computed until all participants have arrived. Difficulties also arise when cost or need limits treatment to those participants who have higher scores on the assignment variable. In each case, placement of

the cutoff at extreme values reduces the power to detect a treatment effect.

If several assignment variables are available with different metrics, a total score can be formed by z-transforming the variables and averaging them (Judd and Kenny, 1981). By this strategy, unreliable assignment variables and variables with poor distributional properties can be summarized to assignment variables with desirable properties.

It is important to know the overall functional form that relates the assignment and outcome variable (i.e., whether it is linear, curvilinear, logistic, etc.). If the functional form is misspecified in the analysis, biased treatment effect estimates will result.

All participants must belong to one population prior to being assigned to the conditions, such that prior to assignment, it must have been possible for all participants in the study to receive the treatment if the cutoff had been set differently. For example, suppose a reading training (treatment) was implemented in school A for children scoring below a cutoff and that the control group contained children in school B who scored above that cutoff. As the treatment was not implemented in school B, it would not have been possible for school B children to receive the training even if their scores were lower than the cutoff score. Therefore, this selection bias would also lead to biased inferences on the treatment effect in the RD design.

Like in a randomized experiment, the participants in the treatment group should all receive the same amount of treatment, and those in the control group no treatment at all. When some treatment group participants receive less treatment than others, the validity of the study will be impaired. Treatment crossover participants also weaken the integrity of assignment. Therefore, it is important to retain all participants in the conditions to which they were originally assigned.

Extension of the Simple Basic RD Design

The basic RD design can be extended in several ways. Instead of comparing a treatment condition with a control condition, two treatment conditions could be compared with each other. The analysis remains the same.

The design can be extended for more than two groups, namely three or more conditions. For example, there could be two different treatment conditions and one control condition. Then, the assignment variable score has to be divided into three intervals by defining two cutoffs.

Treatments can be administered at different doses to different treatment groups so that those participants with higher addiction scores receive higher doses. For example, Fagerstrom (1978) applied an RD design in a study using nicotine gums to prevent relapse among participants who quit smoking. Participants with higher addiction scores

received doses of 4 mg nicotine and participants with lower addiction scores received 2 mg nicotine.

By using two cutoff scores the participants could be separated into three groups. The two outer groups would receive exactly the same treatments, say a training program, and the inner group would be the control group. If the treatment is effective and the control group shows no effect, the relationship between the outcome, on the one side, and the assignment variable and the dichotomous variable representing receiving of treatment, on the other side, would be quadratic. This design has an increased power by a reduced multicollinearity between these two predictors of the outcome variable in a linear model (Kelava *et al.*, 2008).

Theory of the RD Design

In order to show that the RD design yields unbiased estimates of treatment effects that are comparable to a randomized experiment, we will explain that the randomized experiment also uses RDs to estimate effects and that RDs serve as complete models of the selection process.

Link to Randomized Experiments

Suppose that in a reading achievement study a randomized experiment with a pre-post design is applied. Participants are pretested with a reading achievement measure, randomly assigned to training or no training (treatment vs. control condition), and then posttested with the outcome measure reading achievement again. If the treatment has no effect, the experiment would show a scatter plot of the pretest against posttest scores as shown in **Figure 4**. This figure would have a positive slope of the regression line, but no cutoff, because the assignment to groups was random across all pretest scores. The participants are randomly intermixed because random assignment ensures that there is no systematic relationship between treatment group membership and any pretest scores.

Figure 5 shows a scatter plot that would result if the treatment in the randomized experiment had the same 10 score points effect on the reading achievement as the RD design in **Figure 2**. For each group, a separate regression line is graphed. If we add a vertical line at a pretest score measure of 50, the displacement of the two parallel regression lines at this score would be an unbiased estimate of the treatment effect; and if we delete all treatment participants to the right and all control condition participants to the left, the resulting graph would be identical to the RD design in **Figure 2**.

Figure 5 illustrates another important difference between the RD design and a randomized experiment.

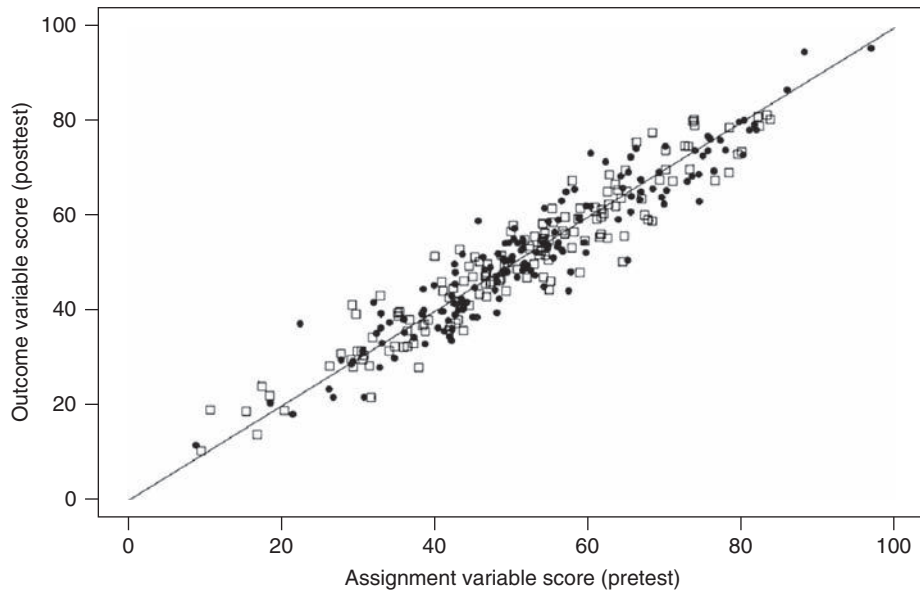


Figure 4 Randomized experiment with no effect of the treatment.

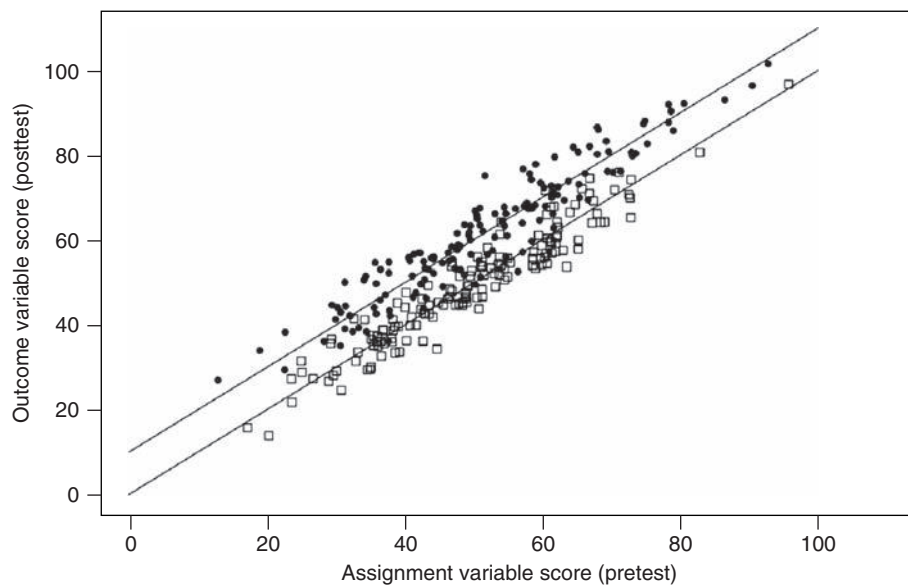


Figure 5 Randomized experiment with a ten-point treatment effect.

In the randomized experiment, the pretest means of the groups are identical at about 50 score points by random assignment. In the RD design, the cutoff-based assignment creates groups with maximally different pretest means and with nonoverlapping distributions. Nevertheless, the RD design yields unbiased estimates of the treatment effect.

In randomized experiments, the treatment effect is inferred by comparing the treatment group posttest mean with the control group posttest mean. In RD, the comparison is not between means but between regression lines. Their intercepts and slopes are compared with each other.

RD as a Complete Model of the Selection Process

In contrast to other quasi-experiments in which assignment to conditions is less controlled, the selection process in RD and in randomized experiments is completely known and perfectly measured. In a randomized experiment, the assignment mechanism is perfectly known, because it is the result of a perfectly measured coin toss. The experimenter can record the result of the coin toss and the assignment that follows from the coin toss. In an RD design, the assignment mechanism is also completely

known and depends on whether the score on the assignment variable is above or below the cutoff. Therefore, the RD design can adjust for the differences at pretest (which is in fact the perfect modeling of the selection bias) and obtain an unbiased estimate of treatment effect. In both cases, in the random experiment and in the RD design, there are no unknown variables that determine to which condition participants are assigned, whereas in quasi-experimental studies, we do not know what the selection process into conditions is and cannot measure it perfectly (Lord, 1967). Therefore, we are not able to correct for selection bias.

Problems that May Arise

The biggest problem in the implementation of the RD design is the adherence to the cutoff. In RD, the treatment assignment must follow the cutoff, but this can be violated by the expectations of treatment professionals that their personal judgment should decide whether a participant gets a treatment. If their judgment is not quantified, recorded, and made part of the assignment variable, the use of such judgmental strategies violates the assumptions of the RD design and the statistical modeling process will lead to false inferences. Cases that are admitted to a treatment, although their scores indicate an assignment to another condition, should be eliminated prior to the treatment and the data analysis.

Another problem complicates the adherence to the cutoff. If participants trickle in too slowly or too quickly, the cutoff score has to be adjusted to keep the program filled. However, if sample sizes are large enough, one can adjust the cutoff successively and analyze each group admitted using a different cutoff in separate RD designs.

Severe problems can result when cutoffs are public. This can foster manipulations by those participants who want to be assigned to the treatment condition. For example, in Berk and Rauma's (1983) RD study, prisoners were provided with unemployment compensation if they had worked more than 652 h over the previous 12 months while in prison, if they had not, no compensation was paid at all. However, those prisoners that were close to the cutoff might work more hours so they can get into treatment, whereas others do not try if they know they have already made it or have little chance of doing so. This affects the distribution of the assignment variable. In the worst case, leaving a gap at the cutoff can result in bimodal data.

Treatment crossovers and attrition are also problems that may arise. Biased treatment effect estimates result when those assigned to treatment do not take it or those assigned to control end up in treatment. Attrition is a problem when it is not random at all. If attrition is related to assignment, biased estimates result again. For example, if students who are on the dean's list are admitted to a

program and subsequently change to a better college due to the good results, the estimates of the treatment effect will be biased (Seaver and Quarton, 1976).

A summarized description of other potential threats on the validity, such as over- and mis-specification of the regression model, nonlinearity, interactions, maturation processes, and events that coincide with the cut-off, is given by Shadish *et al.* (2002).

The Combination of the RD Design with Randomized Experiments and Quasi-Experimental Designs

It is an advantage, if an RD design can be combined with a randomized experiment. For example, this can be achieved, when two cutoffs are used which define a cutoff interval lying between the two scores. Above the upper score, all participants are assigned to one condition and those below to another one. In between, participants are assigned randomly to both conditions. This results in a randomized experiment that is embedded in an RD design. In **Figure 6** for example, children scoring on the assignment variable below 35 are assigned to the treatment condition (reading training program) and scoring above 65 are assigned to the control condition. Children scoring in between 35 and 65 are assigned randomly to both conditions.

There are many different ways to combine a randomized experiment with the RD (for an overview: see Cappelleri, 1991; Trochim, 1984, 1990). First, all participants below a cutoff that are in need could be assigned to a treatment and participants above a cutoff could be randomly assigned to treatment or control condition. Second, an RD design could be added to an existing randomized experiment. For example, those that usually would not pass a cutoff for entering the randomized experiment could be kept as an RD control group rather than just being discarded. Third, one could use multiple cutoff intervals with randomization in some intervals and RD in others. Fourth, multiple cutoffs with increasing proportions of participants randomly assigned to treatment over intervals could be used. For example, five intervals with probabilities of 0:100, 25:75, 50:50, 75:25, and 100:0 for receiving treatment could be used. This approach is very useful when there are concerns that marginal cases might be inappropriately assigned, because the greater the participant's need for treatment the greater the chance of getting treatment. Fifth, participants within a cutoff interval could receive randomly a treatment, but participants outside the interval could be assigned to just one condition. This design might be used in situations when effects of dangerous drugs are examined to minimize risk for the participants. Participants that do not receive the treatment could receive a placebo.

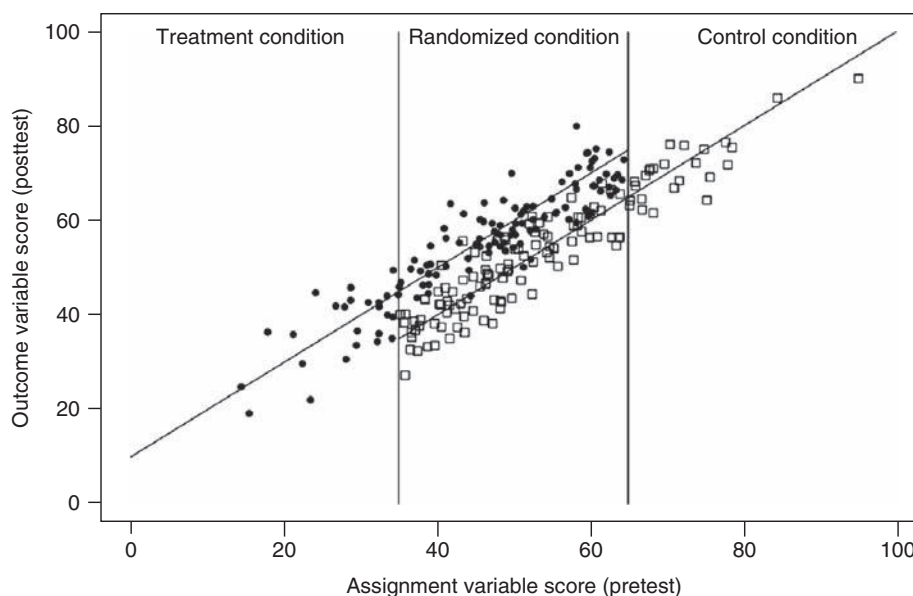


Figure 6 Extended RD design with a randomized assignment cutoff interval.

The advantage of the combined design is that randomization increases the power of detecting treatment effects (Cappelleri, 1991). In addition to this, the combination of RD and a randomized experiment allows the estimation of the regression lines for both treatment and control participants over the same range of assignment scores within the randomization interval. In a basic RD design, a projection (extension) of the lines is needed to compare the conditions. Another advantage is that the ethical problem of where to set the cutoff can be avoided. By defining an interval in the middle of the assignment variable that assigns randomly to treatment, a justifiable solution could be found.

There are also different ways to combine an RD with a quasi-experimental design. One could use a cutoff interval with RD outside the interval and self-selection of the treatment inside the interval. This design should be most accepted by the participants in medical settings.

If there are enough resources, another way to combine RD with a quasi-experimental design is to use the RD design, but at the end of a study give treatment to all the control participants. In psychotherapy, for example, every participant could get the treatment at the end of a study.

In order to find the correct overall functional form of the relationship between predictors and outcome variables, a control cohort could be added that models the functional form without receiving treatment.

feature. In addition to this, there are several issues that disallow a status equal to the randomized experiment. First, it is a new design and the problems described above are too serious and not yet as clear as with the randomized experiment (Shadish *et al.*, 2002; Stanley and Robinson, 1990). Second, the RD design heavily depends on correctly modeling the functional form of the relationship between the assignment variable and the outcome variable. Therefore, in order to obtain unbiased estimates, more sophisticated statistical analyses are required than is usual in experimental settings. Third, the RD design has less power than the randomized experiment (Cappelleri *et al.*, 1994). This is mainly due to multicollinearity between the assignment and treatment variable. The loss of power is greater the more the cutoff deviates from the mean of the assignment variable. When nonlinear effects (e.g., interaction and quadratic effects) are present, multicollinearity increases as well as reduces the power of the design (Kelava *et al.*, 2008). In order to increase the power to 0.80, the RD design needs up to 3 times as many participants as a randomized experiment (Cappelleri *et al.*, 1994).

As a conclusion, the RD is more than a quasi-experiment. It is somewhere in between a quasi-experiment and a randomized experiment with some pitfalls. However, the RD design yields unbiased treatment effects when its assumptions are met.

Final Considerations

The RD design has to be classified as a quasi-experimental design. Although the RD design has a clear structure of an experimental design, it lacks the random assignment

Bibliography

- Berk, R. A. and Rauma, D. (1983). Capitalizing on nonrandom assignment to treatment: A regression discontinuity evaluation of a crime control program. *Journal of the American Statistical Association* 78, 21–27.

- Cain, G. G. (1975). Regression and selection models to improve nonexperimental comparisons. In Bennett, C. A. and Lumsdaine, A. A. (eds.) *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs*, pp 297–317. New York: Academic Press.
- Cappelleri, J. C. (1991). *Cutoff-Based Design in Comparison and Combination with Randomized Clinical Trials*. Cornell University, Ithaca: New York.
- Cappelleri, J. C., Darlington, R. B., and Trochim, W. M. K. (1994). Power analysis of cutoff-based randomized clinical trials. *Evaluation Review* **18**, 141–152.
- Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement* **7**, 249–253.
- Fagerstrom, D. O. (1978). Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. *Addictive Behaviors* **3**, 235–241.
- Judd, C. M. and Kenny, D. A. (1981). *Estimating the Effects of Social Interventions*. Cambridge, UK: Cambridge University Press.
- Kelava, A., Moosbrugger, H., Dimitruk, P., and Schermelleh-Engel, K. (2008). Multicollinearity and missing constraints: A comparison of three approaches for the analysis of latent nonlinear effects. *Methodology* **4**, 51–66.
- Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin* **68**, 304–305.
- Seaver, W. B. and Quarton, R. J. (1976). Regression–discontinuity analysis of dean’s list effects. *Journal of Educational Psychology* **66**, 459–465.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton-Mifflin.
- Stanley, T. D. and Robinson, A. (1990). Shifting statistical significance from the artifact of regression–discontinuity design. *Evaluation Review* **14**, 166–181.
- Trochim, W. M. K. (1984). *Research Design for Program Evaluation. The Regression–Discontinuity Approach*. Newbury Park, CA: Sage.
- Trochim, W. M. K. (1990). The regression–discontinuity design. In Sechrest, L., Perrin, E., and Bunker, J. (eds.) *Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data*, pp 119–140. Rockville, MD: Public Health Service, Agency for Health Care Policy and Research.

Further Reading

- Cappelleri, J. C. and Trochim, W. M. K. (1994). An illustrative statistical analysis of cutoff-based randomized clinical trials. *Journal of Clinical Epidemiology* **47**, 261–270.

Sampling

R S Brown, University of Southern California, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Throughout the course of our everyday existence, we make judgments, inferences, and decisions based on less than the full collection of potential data. We rate restaurants and hotels based on a few meals or several nights of occupancy. We judge athletes based on samples of athletic performance, appraise students based on relatively few assignments, assessments, or tests, and form impressions of new acquaintances based on limited interactions. Sampling is a pervasive and important part of how we interact with and form judgments about the world in which we live. Sampling occurs both informally – as in the instances stated above – and formally as in virtually all governmental surveys, market research polls, and research studies. This article attempts to provide an overview of sampling for research purposes, including terms and definitions, along with various sampling designs.

The basic premise behind sampling is that we seek to gather information from a subset of a larger group of potential respondents in order to make judgments or inferences back to the larger group. As Pedhazur and Schmelkin (1991) state, “Formal sampling is a process aimed at obtaining a representative portion of some whole, thereby affording valid inferences and generalizations to it” (p. 318). Thus, the goal is to identify a group that can adequately represent the whole – so that inferences and generalizations can be made from their responses to survey items or other measures back to the larger population, with some known, or estimable, amount of imprecision. Sampling theory seeks to develop sample-selection strategies and estimation methods that generate sufficiently precise estimates at minimum cost, thereby making sampling more efficient.

There are a number of advantages to sampling compared to collecting data from an entire population. These include reduced cost, greater speed, enhanced feasibility, and greater accuracy. Clearly, cost can be reduced by limiting the amount of data collected to a sample rather than the full population of potential respondents. Likewise, collecting less data would require less time. This consideration for reducing the amount of time necessary for data collection becomes important in situations in which the outcome measure being collected may be influenced by changing events. In volatile social circumstances, early respondents’ reactions may differ from later respondents’ reactions as influences on the outcomes change over an extended data-collection period. Reducing the timeframe for data collection by employing a sampling

strategy can minimize this impact. It is also possible to enhance the feasibility of a study by sampling. In some cases, a complete census is simply impractical and without sampling the study could not be accomplished. In other cases, a census approach would be infeasible for other reasons – such as lack of knowledge of the consequences of a treatment or intervention. It would not be advisable or appropriate to administer a new drug to the entire population, or to compel all students to undertake a given curriculum without first knowing whether the drug is safe and effective or that the curriculum generates enhanced knowledge acquisition (Pedhazur and Schmelkin, 1991). Sampling can also enhance data accuracy by providing more control over the training and supervision of the collection and recording of data, as well as following up on nonresponses. In addition, sampling frees up valuable resources that can be used in the instrument-development and -refinement stages of an investigation.

Sampling Concepts and Terminology

Population

The population refers to the “aggregate from which the sample is chosen” (Cochran, p. 5). It generally refers to the collection or larger grouping to which one wishes to infer or generalize. In some cases, this can be quite straightforward. In a factory setting, the population may be the total collection of widgets produced in a given month or year. A sample of those widgets may be useful for providing sample statistics that estimate population parameters of interest, such as size, weight, or density. In other cases, the population must be defined explicitly in terms of content, units, extent, and time. An example of a given population could be all children attending preschool in California during the 2009–10 academic years.

Sample Units

The sample units are the elements of the population to be sampled. In aggregate, they define the population to which inferences are to be drawn. The sample units may be individuals, classrooms, schools, or a variety of possible elements, but they are the units from which data are collected and serve as the unit of analysis. Sample units may include single elements from the population or clusters of elements from the population, such as when a dwelling may be the sample unit and individual persons

would be the elements from the population. In the example-population given above, the sample units would be the individual children attending preschool in California during the 2009–10 academic years.

Sampling Frame

The sampling frame is the list of sample units from which the sample is drawn. A perfect frame is one in which each element in the population is separately listed once, only once, and no other population irrelevant or extraneous elements are listed. However, not all sampling frames are perfect, and it takes effort and attention to review potential sampling frames to ensure that they are free from error or that the errors in the frames can be addressed. Kish (1965) identifies four primary problems with sampling frames. The first problem is that some sampling frames are incomplete; thus, they do not include all available elements or sample units from the population. Another problem involves clustering of elements within a single listing, which violates the notion that each element be listed separately. A third problem deals with blanks listing or foreign listings in a sampling frame. This violates the rule that each listing should contain a single element. The fourth primary problem with sampling frames is that occasionally duplicate elements appear in the listing. This violates the edict that each element should be listed only once in a sampling frame.

Selection Process

Following the establishment of an appropriate sampling frame, a selection process must be identified, of which there are numerous considerations depending on the study purpose, resources, timeframe, and sample structure. The process depends upon the list from the sampling frame. As Kish states, “The sampling frame or list is the keystone around which the selection process must be designed” (Kish, p. 53). The selection process may involve probability or nonprobability approaches, which are described in more detail below.

Estimation Process

Samples provide data from which sample statistics can be calculated. These sample statistics are used to estimate population values. A common sample statistic of interest is the mean value for the sample. This mean value is a random variable dependent upon the sample of n units from which it is derived. Another sample of size n from the population may result in a different sample mean. The distribution of sample means (the sampling distribution of the mean) allows for the calculation of the population mean and variance (or standard deviation). In practice, not all samples are drawn, so not all sample means are

available, thus a sample design is used. A sample design is unbiased if the expected value of the sample mean (i.e., the mean value of the sample means) is equal to the population mean. Estimates of the population variance can be obtained from the sample statistics and used to develop standard errors of the mean and confidence intervals around the population mean estimates. The magnitude of the sampling error is reflected in the standard errors of the mean and the confidence intervals provide a range of scores that can be interpreted to provide confidence in the population mean estimate. A much more detailed treatment of estimation formulas and estimates is presented in Cochran (1977).

Probability Sampling

There are a number of probability-sampling approaches. All of these approaches require that each element in the population has a known, nonzero probability of being selected into the sample. Probability-sampling approaches rely heavily on random selection. In fact, Kish states emphatically, “Probability selection demands randomized selection. When randomization is both simple and important, disregarding it amounts to carelessness or ignorance” (Kish, p. 28). Utilization of random-selection procedures such as random-number generators and tables of random numbers is important because human beings have shown they are incapable of identifying a random sample independent of such protocols. As Kish states, “Personal judgment has been shown inadequate for selecting random samples of integers, or stones from a pile, or plants from a field, or people on streets or in homes” (Kish p. 29).

Simple Random Sampling

Simple random sampling is a method of selecting a sample size, n , from the population of N elements such that each of the ${}_NC_n$ distinct samples has an equal probability of being selected. This is usually done by selecting at the unit level, with all units being numbered from 1 to N . A table of random numbers or a computer-generated list of random numbers can then be applied to each unit. A total of

$$N! / [(N - n)!n!]$$

distinct sample of size n can be drawn. For example, with a population of size 10, and sample size of 3, a total of 120 unique, nonoverlapping samples can be drawn. From a practical standpoint, simple random sampling is not frequently utilized in large-scale studies. In addition, many studies are interested in more than just the effects on the general populations, for instance many studies are concerned with the effects on subpopulations. Thus, additional sampling strategies have been developed.

Systematic Sampling

Systematic sampling is similar to simple random sampling in that all N elements of the population are numbered and ordered from 1 to N . However, unlike in simple random sampling in which each of n elements is then randomly chosen, in systematic sampling a element is chosen from the first k elements (k is often defined as N/n), then every k th element is chosen until n elements are selected. For instance, if k is 11 and the first element selected is 7, then elements 18, 29, 40, and 51 would be chosen to complete the sample of $n = 5$. Systematic sampling has the advantage of being easier to execute in some instances than simple random sampling, and may be more precise than both simple random sampling and stratified random sampling.

Stratified Random Sampling

Stratified random sampling utilizes known information about the population elements to separate the sample units into nonoverlapping groups, or strata, from which they are then randomly selected. For example, a population of fourth-grade school children may be stratified into various geographic regions, or types of schools attended. Therefore, the full population, N , is subdivided into k strata, such that

$$N = N_1 + N_2 + N_3 + \dots N_k.$$

Once the strata have been determined, a simple random sample is selected from within each stratum. There are a number of advantages to stratified random sampling. One such advantage is that it provides more precise estimates of the subpopulation parameters than does simple random sampling. Stratified sampling can reduce sampling variability by creating relatively homogeneous subsets of sample units with respect to the outcome variable when the stratifying variable is correlated with the outcome measure, although the gains in precision are often minimal. Another potential advantage to stratified sampling is that sampling frames for naturally occurring strata may be readily available, such as student rosters within schools as opposed to all second-grade students in a given country.

There are disadvantages to stratified sampling procedures as well as potential advantages. Stratified sampling designs are more complex both in the selection process and in the estimation process. In addition, some investigations are concerned with multiple outcomes measures, not all of which will necessarily be well correlated with the stratifying variable. In this case, using a stratified sampling design might be beneficial for one outcome measure, but less beneficial for others. Another consideration is the determination of which variables should be used for establishing the strata, and how many strata are optimal. Kish (1965) suggests that the initial advantage obtained by

stratification diminishes with increasing numbers of strata, and that from three to ten strata should be sufficient for any single outcome variable.

Proportionate Sampling

In proportionate sampling, strata sample sizes are held proportional to their existence in the population. This is often used to show the representativeness of a sample relative to the population's characteristics. The sampling fraction for each stratum is held constant and equivalent to the sampling fraction for the overall population. For example, assume that a study seeks a sample of 200 from a population of 1000 (thus the sampling fraction, f , for the population is $200/1000$ or $1/5$). The population is divided into three strata of differing sizes: $N_1 = 500$, $N_2 = 200$, and $N_3 = 300$. The overall sampling fraction, $f = 0.2$, is applied to each stratum to yield samples of $n_1 = 100$, $n_2 = 40$, and $n_3 = 60$. As a result, the proportions of selected sample units in each stratum match the proportions of the strata units relative to the population (e.g., $N_1 = 500/1000 = 0.5 = n_1/n = 100/200 = 0.5$). The advantage of this approach is that it generates a self-weighting sample and the population mean can be estimated from the simple mean of the sample elements.

Cluster Sampling

Cluster sampling is a probability-sampling design that capitalizes on naturally occurring groups, or clusters, in the population. Examples of such naturally occurring groups are students within a classroom or school, residents of a city block, or patients at a given medical facility. In cluster sampling, the groups or clusters are first randomly selected; then, all members of that group are included in the study. Thus, if the clusters are schools and school A is chosen, all the students (or all sixth-graders, depending upon the sampling frame and study intent) in school A would participate. Cluster sampling has the advantage of reducing cost and time associated with sampling and data collection. However, the selected clusters need to represent the population of clusters. When the clusters are more heterogeneously composed internally, this is less problematic. But when the clusters are more internally homogeneous with respect to the outcome measures, a considerable loss of precision in estimation can occur. This leads Kalton (1983) to conclude that unless the economy of using a cluster-sampling design allows for a sufficient increase in sample size to override the associated loss in estimator precision, cluster-sampling designs should be avoided. In addition to the heterogeneity within a cluster, the variability of the size of clusters makes a difference in estimation using cluster sampling. Since most naturally occurring clusters are of unequal size, this variation in cluster size should be taken

into account. Large variability of cluster sizes leads to less precise estimation, unless adequately accounted for. However, additional sampling designs, such as selection with probabilities proportionate to size provides approximate control over the influence of unequal cluster sizes.

Multistage Sampling

Multistage sampling is an extension of cluster sampling in that, first, clusters are randomly selected and, second, sample units within the selected clusters are randomly selected. In this design, random selection occurs at both the cluster or group level and at the sample unit level. Multistage sampling also may be useful when naturally occurring cluster sizes are rather large, resulting in reduced precision when compared to the stratified random-sampling approach. In this event, smaller clusters can be created and sampled. For example, rather than sampling all sixth graders at a selected school, a secondary cluster, such as classes, could be utilized and sampled. In this case, only students in sampled classes at sampled schools would be included. This has the added advantage of reducing the cluster size, thereby enhancing estimator precision.

Double Sampling

Double sampling refers to the practice of selecting an initial sample in order to obtain information on a potential stratifying variable, then using that information to develop a stratified random sampling plan to obtain information on the outcome measure of interest. This is frequently done when the distributional characteristics of the stratifying variable, x , are unknown in the population. The advantage of this approach is that it provides information that may be used to create strata for the sampling design for the collection of data on the outcome variable, y , of primary interest. The disadvantage is that resources that might otherwise be utilized in data collection and analysis for information on y must be spent to collect information on x . Thus, it is important that the increase in precision of estimation due to stratification offsets the loss of precision due to reduced sample size in the data collection for variable y . This technique is also, sometimes, referred to as two-stage sampling.

Nonprobability Sampling

There are a number of nonprobability sampling techniques that have been identified in the sampling literature. While these approaches may yield subjects from whom data may be collected, these design do not benefit from the main advantage of probability approaches, namely,

that probability designs allow for the development of statistical theory to examine the properties of sample estimators (Kalton, 1983). Nevertheless, these approaches are used widely in social science research and they include convenience sampling, purposive sampling, or quota sampling. In convenience sampling, a sample is selected based on ready availability – such as students in a given classroom, or passers-by on a street corner or in front of a busy market. Other convenience samples may include respondents to an advertisement in a magazine or dial-in number for a reality television program. Purposive sampling differs from convenience sampling in that certain characteristics of the sample are sought out *a priori*; that is, a sample that possesses certain characteristics, often to be seen as representative of some larger population is sought. One example of a purposive sample may involve a researcher selecting communities from across a state to ensure geographic diversity.

Another nonprobability sampling approach is quota sampling. In quota sampling, the researcher attempts to collect data on a specified number of respondents in each of a number of groups of potential respondents. For example, a researcher may seek responses from 20 third-grade teachers, 20 fourth-grade teachers, 20 eighth-grade teachers, and 20 tenth-grade teachers. One advantage of quota sampling is that it can reduce data-collection time and associated costs. It is somewhat similar to stratified sampling in that it seeks to obtain responses from more homogeneous subsets of the total populations. However, the major and pivotal distinction is that with stratified random sampling, the sample units within the strata are randomly selected, whereas in quota sampling, the sample units within the quota groups are not. Further, while some may argue that quota sampling reduces nonresponse, the reality is that quota sampling merely replaces nonrespondents with other respondents in the quota group, thereby underestimating the responses of hard-to-find or unwilling sample units.

The fundamental problem with nonprobability sampling designs is that they are potentially biased in their sample estimators, and the magnitude of this bias is unknown. What is known is that the concern with regard to the bias in sample estimators increases with sample size, since probability-sampling designs become more precise with larger samples. Thus, whereas bias in sample estimators may be of less concern when deploying nonprobability sampling approaches for small-scale studies, more care should be taken to avoid nonprobability designs with larger research efforts. These sampling approaches may be utilized widely, but as Pedhazur and Schmelkin state, “...the incontrovertible fact is that, in nonprobability sampling, it is not possible to estimate sampling errors. Therefore, validity of inferences to a population cannot be ascertained” (p. 321).

Sample Size and Power

One of the most common questions that surfaces when dealing with discussions on sampling deals with how large a sample is required for a given purpose. Samples should be large enough to provide the representativeness desired, but the precision of estimation is also a critical component. Since sample size is a primary component in determining estimation precision, once an acceptable level of precision is identified, the optimum samples for such precision can be readily calculated (see Cochran, 1977, for a more thorough treatment of optimal sample sizes for various sampling designs). The determination of what constitutes an acceptable amount of precision is not easily made, and is determined, in part, by the purpose of the research, practical and economic constraints, and the estimated effect size (ES) for a given treatment or intervention. Sample size also affects the likelihood of identifying statistically significant effects or differences among groups.

Of significant importance is the requisite sample size to provide a research investigation with sufficient power to validly and statistically explore the research questions of interest. Sociobehavioral research is full of analytical overviews of research indicating that, in many disciplines, much of the published research fails to establish high levels of statistical power – primarily due to inadequate sample sizes. The statistical power of an analysis is determined by three components. These include: the ES, the probability of rejecting the null hypothesis when it should not be rejected (α), and sample size (N). The power of a statistical test is defined as $1 - \beta$, or 1 minus the probability of failing to reject the null hypothesis when it should be rejected (β). Thus, for a given or expected ES with an *a priori* determined α -level – given statistical analysis, and desired power level – the appropriate sample size can be calculated. Tables of such values can be found in Cohen (1988). In addition, computer programs are widely available (e.g., G*Power 3, Power and Precision, Power Analysis and Sample Size (PASS) are but a few of the many

programs available) to assist the researcher in estimating necessary sample sizes given a desired power level, expected effect size, and α -level.

Conclusion

Sampling is a necessary and important component of much of the data-collection efforts we undertake both in informal and formal settings. Understanding the importance of sampling, the various approaches to collecting adequate samples, the advantages and disadvantages of different sampling procedures, and the relationship of sample size to statistical power enables researchers to make more valid inferences and generalizations from sample data to the population of interest, and thereby increases our knowledge of the domain in which they operate.

Bibliography

- Cochran, W. G. (1977). *Sampling Techniques*, 3rd edn. New York: Wiley.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, NJ: Erlbaum.
- Kalton, G. (1983). *Introduction to Survey Sampling*. Newbury Park, CA: Sage.
- Kish, L. (1965). *Survey Sampling*. New York: Wiley.
- Pedhazur, E. J. and Schmelkin, L. P. (1991). *Measurement, Design, and Analysis: An Integrated Approach*. Hillsdale, NJ: Erlbaum.

Further Reading

- Babbie, E. (1990). *Survey Research Methods*, 2nd edn. Belmont, CA: Wadsworth.
- Fink, A. (1995). *How to Sample in Surveys*. Thousand Oaks, CA: Sage.
- Fowler, F. J. (2002). *Survey Research Methods*, 3rd edn. Newbury Park, CA: Sage.
- Henry, G. T. (1990). *Practical Sampling*. Newbury Park, CA: Sage.
- Jaeger, R. M. (1984). *Sampling in Education and the Social Sciences*. New York: Longman.
- Konjin, H. S. (1973). *Statistical Theory of Sample Survey Design and Analysis*. New York: Elsevier.
- Lohr, S. (1999). *Sampling Design and Analysis*. Pacific Grove, CA: Duxbury Press.

Statistical Conclusion Validity

P L Busk, University of San Francisco, San Francisco, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Error rate – The error rate refers to the type I error. When only a single statistical test is performed, then the error rate is that set by the researcher, typically $p = 0.05$. If more than one test is conducted, then the error rate increases with the number of tests performed.

Robustness of a statistical test – Indicates that the results of a statistical test can be interpreted accurately even when the conditions for the test violate certain of its assumptions.

Statistical power – Is the probability of finding an effect or relationship when the effect or relationship actually exists. The power of a statistical test is computed by subtracting the type II error from 1.

Type I error – A type I error is the probability of rejecting the null hypothesis when the null hypothesis is true – that is, rejecting the null hypothesis and concluding that there is an effect when in fact there really is no effect.

Type II error – A type II error is the probability of not rejecting the null hypothesis when it is false or the alternative hypothesis is true – that is, a type II error occurs when the researcher concludes that there is no effect when there really is an effect. (See also the definition for the glossary term ‘Statistical power’.)

Prior to undertaking data analyses, the researcher must determine how assumptions for the statistical test and for the selection of appropriate tests affect the interpretations of results. In other words, the researcher must determine the statistical conclusion validity. The classic definition of this essential was given by Cook and Campbell (1979: 37) as the truthfulness of “conclusions about covariation” between treatment and outcome – that is, the degree of accuracy with which researchers can conclude that two variables are related. Only one of two conclusions apply to the covariation between treatment and outcome: rejecting the null hypothesis or not rejecting the null hypothesis. If the treatment is ineffective, then no evidence exists to reject the null hypothesis (although a type II error could have occurred). If the treatment is effective, then the null hypothesis is rejected (although a type I error could have occurred).

Shadish *et al.* (2002) have listed nine threats to statistical conclusion validity: low statistical power, violated

assumptions of statistical tests, fishing and the error rate problem, unreliability of measures, restriction of range, unreliability of treatment implementation, extraneous variance in the experimental setting, heterogeneity of respondents, and inaccurate effect-size estimation. These threats constitute the reasons why researchers would be incorrect in drawing valid inferences concerning the covariation between treatment and outcome – that is, that two variables are related or correlated.

An Internet search on the phrase “statistical conclusion validity and books” reveals that it is mainly the threats to validity that are mentioned in research methods books. Authors of statistics books, however, would benefit from including details of these validity issues. Although many specifically address certain aspects of statistical validity – statistical power (see, e.g., Cohen, 1988; Lipsey, 1990; Maxwell and Delaney, 2004), violation of assumptions of statistical tests (Wilcox, 1996), and the error-rate problem (Maxwell and Delaney, 2004) – they do not connect these problems to statistical conclusion validity. What is meant by addressing specifically aspects of statistical validity is that there are sections in chapters devoted to these topics. Many of the statistics textbooks will discuss type I and II errors and robustness of the statistical tests, for example, but in general and without separate sections for each statistical procedure presented. The exceptions are rare. Some of the authors have addressed the topic of statistical conclusion validity in detail (see, e.g., De Gruijter and van der Kam, 2007; Kirk, 1995; Saxe and Fine, 1981). Girden (1996) discussed six of the threats in her book on evaluating research articles, and Kirk focused on seven of the threats listed by Shadish *et al.* (2002), although the names for some of the threats do not correspond to those used by Shadish *et al.*

What follows is an explanation of each of the nine threats, as proposed by Shadish *et al.* (2002), to statistical conclusion validity. Included are additional resources and references that will assist the researcher and the statistician in understanding the threats in order to avoid or, at least, to meet them effectively.

Low Statistical Power

Most statistical textbooks focus on the issue of low statistical power, although the articles reporting research results rarely mention power. Low statistical power, often, results in type II errors and is explained by small

sample sizes or by extraneous sources of variation that are not controlled. Analyses of statistical power in published research in the 1970s and 1980s found that the power of statistical tests was 0.50 (see, e.g., Sedlmeier and Gigerenzer, 1989; Lipsey, 1990), and – in a study conducted in the twenty-first century in engineering – low power levels remain weak (Kampenes *et al.*, 2007). Steps can be taken to improve statistical power by performing a priori power analysis to estimate sample-size requirements prior to conducting a research study. Researchers need to specify power size, level of significance, and effect size to find the sample size. Cohen (1988) recommended that one should plan for power of at least 0.80. His book contains power tables that researchers can use to find the sample size required to achieve the desired power with specified level of significance (or type I error) and effect size. Determining the effect size is not straightforward, of course, as it is, generally, unknown and difficult to guess – because basing the effect size estimate on sample data is problematic as the sample estimate is susceptible to influence of the treatment implementation, measures, samples, and research design of a particular study. Lipsey has recommended using results of previous research to create a base for the effect-size estimate. He had provided from research in psychology and in education specifics on 186 meta-analyses that were used to estimate the effect size for other researchers' investigations.

The Internet is a source of websites that provide power-analysis calculators or sample-size calculators. The latter source is especially useful to researchers at the planning stage when they are determining effective sample size.

Extraneous sources of variation that are uncontrolled affect the variance of the sample and hence the standard error or mean square error used to test the hypothesis of interest. Eliminating extraneous variables at the implementation stage reduces the variances, increasing the power. Shadish *et al.* (2002: 46–47) provided additional methods for increasing the power of statistical tests, as well as comments on feasibility, application, exceptions to their use, and the disadvantages. Researchers will also find comprehensive references that are valuable for directing them in designing studies with power to detect the effects that they are investigating.

Violated Assumptions of the Test Statistics

All statistical tests require that certain conditions be met for valid inferences. If these conditions or assumptions are not met, incorrect inferences or conclusions with regard to covariation or treatment effectiveness often result. In-depth presentations on the violations of assumptions can be found in most statistics textbooks (see, e.g., Kirk,

1995; Wilcox, 1996). Research-methods books usually devote only a chapter to inferential statistics – that is, analyzing and interpreting quantitative data – which means that, at most, a few paragraphs or a few pages will be used to address the violation of the assumptions of statistical tests. Often, the authors of research books do not have a background in statistics and may not be aware of the consequences of violating the assumptions. For example, compare Krathwohl's (1998: 486–488) with Shadish *et al.*'s (2002: 48) discussion of violation of assumptions.

As a recommendation to researchers with regard to violations of assumptions, Wampold and Drew (1990) emphasize that “the researcher should never conduct an experiment if he or she is blind to clues about the distributional characteristics of the population under investigation” (p. 152). In order to understand the distributional characteristics of the population under investigation, researchers may derive clues from previous research, from theory, or from the nature of the experiment or the instruments used to obtain the observations. If one is not able to obtain sufficient information from these sources, a pilot study must be conducted.

Fishing and the Error-Rate Problem

When a large number of statistical tests are performed on the same set of data, then – by chance alone – one or more tests may result in a type I error. If a researcher conducts multiple tests but only reports those that are statistically significant, the research conclusions can be misleading. Authors of statistical books (see, e.g., Glass and Hopkins, 1995; Kirk, 1995; Maxwell and Delaney, 2004) have warned against the practice of fishing or multiple comparisons and recommended the Bonferroni correction – which divides the overall error rate by the number of tests conducted, ensuring that the error rate over all tests will not exceed the nominal level of significance. Kirk (1995) and Maxwell and Delaney (2004) have discussed the correct conceptual unit for type I error in light of three different type I error rates – family-wise error rate, per-contrast error rate, and experiment error rate – and the techniques that have been developed to control them. Maxwell and Delaney provided a flowchart for choosing an appropriate procedure. The difficulty arises when researchers only report statistically significant results when they have conducted multiple tests. If results of all statistical tests were reported in published articles, then readers could assess the chances of spurious statistically significant results. Given space limitations in printed journal articles, this recommendation is not likely to occur. With Internet-based publications, however, space is no longer a limitation, and such reporting should occur.

Unreliability of Measures

If the measuring instrument does not obtain consistent results, then the error term for the statistical tests will be inflated – which will result in low power and possibly a type II error. The problem of unreliability will affect the assessment of covariation because bivariate relationships will be attenuated – that is, the correlation coefficient will be systematically biased downward. Attenuation is a topic typically found in many measurement books rather than in research or statistics texts.

Fan (2003) has provided two approaches for correcting attenuation when measurement error is known – that is, when reliability data are available for the measured variables being correlated. Another source of attenuation may result from assessing change in longitudinal studies and in consequences for analysis of covariance (Maxwell and Delaney, 2004: ch. 9). For longitudinal studies, recommendations include assessing and reporting reliability for each measure as well as increasing the number of measurements, improving the quality of measures, and using techniques such as latent variable modeling. For analysis of covariance, Huitema (1980: ch. 14) had proposed a number of corrections for measurement error in the covariate that involve estimating the reliability of the covariate.

Restriction of Range

Restriction of range affects both correlation coefficients and the power of statistical tests. If the sample on which a correlation is based is homogeneous with respect to the variables being measured, then the correlation between the variables will be much less than if the sample were more heterogeneous. This phenomenon of range restriction is problematic for researchers trying to predict performance based on a sample that has a restricted range, for example, predicting performance for the first year in college based on applicants to that college or predicting performance for first year in college based on high-school students who take the Scholastic Aptitude Test (SAT) or American College Testing (ACT). Given that not all high-school students have college aspirations or will be attending a college that requires the SAT or ACT for admission and, further, not all students who take the test are admitted to colleges or universities, the range of scores is restricted for both variables and, hence, the correlation on which the prediction is based is restricted. Glass and Hopkins (1995) provided a formula for correcting the correlation coefficient for restricted variability in one of the samples. This equation should be used only for large samples and only where the conditions of linearity and homoscedasticity are met.

The range restriction on the dependent variable lowers the power of statistical tests. Researchers may be studying restricted samples that will result in floor or ceiling effects for the dependent variable. For example, if severely depressed individuals are studied, their data will result in floor effects as scores will cluster near the highest scores on a measure of depression, and, if highly anxious individuals are studied, their data will result in ceiling effects as their scores will cluster near the highest scores on a measure of anxiety. Researchers conducting studies with homogeneous samples should compensate for low power with larger sample sizes.

Unreliability of Treatment Implementation

When the researcher fails to standardize the administration of treatment levels, unreliability of treatment implementation can occur. Such variability in treatment implementation may result in inflated error variance and in a type II error. Such a threat to statistical conclusion validity is typically addressed in research books. The usual recommendation is that all procedures for the treatment implementation be standardized. Such standardization may be difficult when different instructors are implementing the treatment in their respective classrooms or different observers are collecting data in a research study. Unless trained thoroughly and monitored closely, they may not follow the treatment protocol or may alter the control condition. Such a lack of standard implementation could result in a smaller effect size, which would be more difficult to detect statistically. Again, the researcher could increase the sample size to increase power.

In some situations (e.g., single-case studies), the treatment is tailored to certain participants in order to increase treatment effectiveness. In these cases, the components of the treatment should be measured, and these variations related to changes in outcomes. Shadish *et al.* (2002) have devoted two chapters to methods for improving, measuring, and analyzing treatments that are not implemented reliably. Studies where the treatments are tailored to participants, often, occur in single-case investigations.

Extraneous Variance in the Experimental Setting

Extraneous variance in the experimental setting or random irrelevancies in the experimental setting result when variation in the environment in which the treatment is administered may inflate the error variance and result in a type II error. Variation in environment might be physical or social and could include distracting noises, fluctuations

in temperature or lighting, faulty equipment, or changes requested by site administrators or other nonparticipants in the research – such as parents. Control, usually, can be achieved in a laboratory setting, but, in the field, variances are more likely to occur. The researcher must minimize such distractions. If this is not possible, then the researcher must measure the source of the extraneous variance and factor the variation into the statistical analysis so that the error variance will be reduced and be less likely to result in not rejecting a false null hypothesis.

Heterogeneity of Respondents

Although researchers desire heterogeneity in their samples for detecting correlation among variables, homogeneity of participants on the outcome variable within conditions in experimental or quasi-experimental research will reduce error variance and assist in detecting statistical significance for the treatment – that is, heterogeneity of respondents within conditions on the outcome variable increases the variance and reduces the power of the statistical test. The obvious solution is to obtain a homogeneous sample, but this solution reduces the external validity and may even result in restriction in range. Another solution would be to measure the respondent characteristics that are heterogeneous and use these measurements for blocking or as covariates in the analysis procedures.

Inaccurate Effect-Size Estimation

There are several factors that affect accurate effect-size information. Cohen (1988) defined his effect-size measure as the difference between treatment and control means divided by the pooled standard deviation – which assumes that the two groups have equal or approximately equal sample sizes and standard deviations. If either of these two conditions is not met, then the effect-size estimate is inaccurate. To address this problem, Glass *et al.* (1981) defined the effect-size estimate as the difference between means divided by the control standard deviation, thus eliminating the issue of equal or approximately equal sample sizes and standard deviations. Glass *et al.* made an additional assumption – as did Cohen – that both groups have a normal distribution. In particular, Cohen's nonoverlap is based on a normal distribution as is Glass *et al.*'s graphical illustration of effect-size differences (see p. 29). Problems arise when the distributions are not normal and the means are unequal, because the variances might be unequal as well, which means that the magnitude of the effect size can be misleading and difficult to interpret. Wilcox (1996) proposed alternative approaches for effect-size estimation methods along with a Minitab macro to perform the calculations.

Statistical Conclusion Validity in Meta-Analysis

When conducting a meta-analysis, Cooper (1998) and Wortman (1994) have recommended to judge research quality. Many meta-analysts judge the quality of research articles on whether the designs were randomized, non-randomized, or quasi-experimental. Other design features should be considered in the coding of research studies. Wortman included Cook and Campbell's (1979) statistical conclusion-validity categories in the coding of research qualities. Wortman referenced an early meta-analysis by Devine and Cook (1983) that examined six alternative explanations that could account for the observed effects. In the meta-analysis of effects of psychoeducational interventions on the length of postsurgical hospital stay, Devine and Cook considered six threats to validity including one for statistical conclusion validity: the reliability of treatment implementation. They found that treatments that included content from only one domain had lower effect sizes (0.06) than treatments that included content from more than one domain (0.51 for two domains and 0.32 for three domains). The three domains of psychoeducational care were information, skills training, and social support. This early example is one of the few that include statistical conclusion validity as a moderator variable (note that the second author is Cook of Cook and Campbell 1979; Shadish *et al.*, 2002).

Meta-analysts may find that investigating the threats to statistical conclusion validity will provide relevant moderator variables for investigating differences in effect sizes. Consideration of low statistical power may provide insights valuable both to studies that are published and those that are not. Recommendations for future research would aid researchers in publishing their findings.

Conclusion

Statistical conclusion validity is an integral part of a validity typology proposed by Cook and Campbell (1979) that is composed of four related parts: in addition to statistical conclusion validity, the three other parts are internal validity, construct validity, and external validity. Cook and Campbell (1979) introduced statistical conclusion validity for quasi-experimental research designs, which Shadish *et al.* (2002) broadened to include experimental designs. With this pedigree in research design and methodology, statistical conclusion validity demands a place in the body of statistical literature, rather than a nod in the footnotes – especially given the emphasis in research on interpreting findings. Without the support of statistical conclusion validity and the three other validity components, accurate decisions could be made on the basis of unjustified statistical analysis. Statistical literature

focuses exclusively on procedures for analyzing data; the research design and methodology literature includes statistical analyses, but the focus is not comprehensive. Authors of research-design books, usually, include sections on threats to validity that only include internal, construct, and external validity. It is rare that educational-research books focus on statistical validity threats. One of the few exceptions is nursing-research texts.

Even though studies have been published in the literature, beginning in 1962, with regard to violations of statistical conclusion validity, researchers still have been slow to employ its benefits – despite constant criticism by reviewers of low statistical power in research investigations. These reviews are not restricted to one field, as reports appear in the medical, educational, and psychological disciplines. Reviews of statistical conclusion validity have been conducted in criminal justice and criminology, nursing, business research, and early-interventions research. These evaluations consistently urge researchers in their various disciplines to be aware that the interpretation of the results depends on the assumptions for the statistical test and on the selection of proper tests, as Cook and Campbell (1979) detailed.

As pointed out above, some authors of books have chosen to include a subset of the threats to statistical conclusion validity. This article has detailed the threats which researchers need to address. Although some of the threats are more serious than others, they all affect the validity of the researcher's conclusions. Given the time and energy that a researcher devotes to designing studies and the role these studies have in decision making, researchers should take full account of the, often, readily solvable threats to statistical conclusion validity.

See also: Internal Validity; Meta Analysis; Quasi-Experimentation: Two Group Design; Statistical Power Analysis.

Bibliography

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, NJ: Erlbaum.
- Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago, IL: Rand-McNally.
- Cooper, H. (1998). *Synthesizing Research: A Guide for Literature Reviews*, 3rd edn. Thousand Oaks, CA: Sage.
- De Gruijter, D. N. and van der Kam, L. J. T. (2007). *Statistical Test Theory for Behavioral Sciences*. Cleveland, OH: CRC Press.

- Devine, E. C. and Cook, T. C. (1983). A meta-analytic analysis of effects of psychoeducational interventions on length of postsurgical hospital stay. *Nursing Research* **32**, 267–274.
- Fan, X. (2003). Two approaches for correcting correlation attenuation caused by measurement error: Implications for research practice. *Educational and Psychological Measurement* **63**, 915–930.
- Glass, G. V. and Hopkins, K. D. (1995). *Statistical Methods in Education and Psychology*, 3rd edn. New York: Allyn and Bacon.
- Glass, G. V., McGaw, B., and Smith, M. L. (1981). *Meta-Analysis in Social Research*. Newbury Park, CA: Sage.
- Huitema, B. E. (1980). *The Analysis of Covariance and Alternatives*. New York: Wiley.
- Kampenes, V. B., Dyna, T., Hanney, J. E., and Sjöberg, D. I. J. (2007). Systematic review: A systematic review of effect size in software engineering experiments. *Information and Software Technology* **49**, 1073–1086.
- Kirk, R. E. (1995). *Experimental Design: Procedures for Behavioral Sciences*, 3rd edn. Pacific Grove, CA: Brooks/Cole Publishing.
- Krathwohl, D. R. (1998). *Methods of Educational and Social Science Research: An Integrated Approach*, 2nd edn. New York: Longman.
- Lipsey, M. W. (1990). *Design Sensitivity: Statistical Power for Experimental Research*. Thousand Oaks, CA: Sage.
- Maxwell, S. E. and Delaney, H. D. (2004). *Designing Experiments and Analyzing Data: A Model Comparison Approach*, 2nd edn. Mahwah, NJ: Erlbaum.
- Saxe, L. and Fine, M. (1981). *Social Experiments: Methods for Design and Evaluation*. Beverly Hills, CA: Sage.
- Sedlmeier, P. and Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin* **105**, 309–316.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.
- Wampold, B. E. and Drew, C. J. (1990). *Theory and Applications of Statistics*. New York: McGraw-Hill.
- Wilcox, R. R. (1996). *Statistics for Social Sciences*. San Diego, CA: Academic Press.
- Wortman, P. M. (1994). Judging research quality. In Cooper, H. and Hedges, L. V. (eds.) *The Handbook of Research Synthesis* pp 97–109. New York: Russell Sage Foundation.

Further Reading

- Girden, E. R. (1996). *Evaluating Research Articles from Start to Finish*. Thousand Oaks, CA: Sage.
- Kirk, R. E. (1995). *Experimental Design: Procedures for Behavioral Sciences*, 3rd edn. Pacific Grove, CA: Brooks/Cole Publishing.
- Lane, D. M. (2000). *Rice Virtual Lab in Statistics: Power Analysis*. <http://davidmlane.com/hyperstat/power.html> (accessed August 2009).
- Lipsey, M. W. (1990). *Design Sensitivity: Statistical Power for Experimental Research*. Thousand Oaks, CA: Sage.
- Mark, M. A. and Reichardt, C. S. (2009). Quasi-experimentation. In Bickman, L. and Rog, D. J. (eds.) *The SAGE Handbook of Applied Social Research Methods*, 2nd edn., pp 193–228. Thousand Oaks, CA: Sage.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.
- Trochim, W. M. (2006). *The Research Methods Knowledge Base: Statistical Conclusion Validity*, 2nd edn. <http://www.socialresearchmethods.net/kb> (accessed August 2009).

Survey Research Methods

A Fink, University of California, Los Angeles, Los Angeles, CA, USA; The Langley Research Institute, Pacific Palisades, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Categorical or nominal – Categorical or nominal response choices have no numerical or preferential values.

Concurrent validity – Demonstrated when two assessments agree or a new measure is compared favorably with one that is already considered valid.

Construct validity – Established experimentally to demonstrate that a survey distinguishes between people who do and do not have certain characteristics. It is usually established experimentally.

Content validity – Refers to the extent to which a measure thoroughly and appropriately assesses the skills or characteristics it is intended to measure.

Face validity – Refers to how a measure appears on the surface: Does it seem to ask all the needed questions?

Survey's Internal Consistency or homogeneity – A survey's internal consistency or homogeneity. Refers to the extent to which all the items or questions assess the same skill, characteristic, or quality.

Nonprobability or convenience samples – They are chosen based on judgment regarding the characteristics of the target population and the needs of the survey. With nonprobability sampling, some members of the eligible target population have a chance of being chosen and others do not.

Numerical – Response choices call for numbers such as age (e.g., number of years) or height (e.g., number of meters).

Ordinal – Response choices require rate or order choices, say, from very positive to very negative.

Predictive validity – Refers to the extent to which a survey measure forecasts future performance.

Probability sampling – It provides a statistical basis for saying that a sample is representative of the study or target population. In probability sampling, every member of the target population has a known, nonzero probability of being included in the sample.⁷ Probability sampling implies the use of random selection.

Reliable survey instrument – A reliable survey instrument is one that is relatively free from "measurement error." Because of this "error,"

individuals' obtained scores are different from their true scores, which can only be obtained from perfect measures.

Survey – A system for collecting valid information from or about people to describe, compare, or explain their knowledge, attitudes, and behavior (Fink, 2008). The system consists of interrelated activities starting with defining precise survey objectives, choosing respondents, preparing a reliable and valid survey instrument, testing the survey with respondents, and conducting all activities in an ethical manner.

Test-retest reliability or stability – A measure of a survey's stability over time. A survey is stable if the correlation between scores from one time to another is high.

Survey Research: Definitions and Uses

A survey is a system for collecting valid information from or about people to describe, compare, or explain their knowledge, attitudes, and behavior (Fink, 2008a). The system consists of interrelated activities starting with defining precise survey objectives, choosing respondents, preparing a reliable and valid survey instrument, testing the survey with respondents, and conducting all activities in an ethical manner. The survey researcher is responsible for implementing all survey activities and ensuring their quality.

Survey information can be collected directly by asking people to answer questions, or indirectly, by reviewing written, oral, and visual records of people's thoughts and actions. Survey data can also be obtained by observing people in natural or experimental settings.

Surveys are taken of political and consumer choices, use of health services, numbers of people in the labor force, and opinions on just about everything from aardvarks to zyzyvas. Individuals, communities, schools, businesses, and researchers use surveys to find out about people by asking questions about feelings, motivations, plans, beliefs, and personal backgrounds. The questions in survey instruments are typically arranged into mailed or self-administered questionnaires (on paper or online) and in-person (face-to-face) or telephone interviews.

Surveys are a prominent part of life in many major industrialized nations, particularly the United States. US elections are always accompanied by a flood of polls. The US Census, which is administered to the entire nation every 10 years, is a survey. However, most people encounter surveys more frequently in their daily lives in medical, educational, and social settings.

Survey Objectives

A survey's objectives are general statements of the survey's outcomes and provide the direction for selecting questions (**Case study 1**).

When planning a survey, the survey researcher must define all potentially imprecise or ambiguous terms in the objectives (Fink, 2004a). For the objectives above, the imprecise terms are needs; educational services; characteristics; and benefits. No standard definition exists for any of them. What are needs, for example, and of the very long list that the survey researcher can create, which are so important that they should be included on the survey? Definitions can come from the literature and from consultation with knowledgeable individuals.

Survey objectives may be independent of an existing study or related to it. For instance, suppose a school district decides to investigate the causes of a measurable increase in smoking among students between 12 and 16 years of age. The district could then ask the survey research department to design and implement a survey of students who smoke to find out why they do. On the other hand, the school district may be part of a study to prevent

smoking in students, and a component of that study may be a survey of students' smoking habits.

A survey's objectives can be derived from reviews of the literature and other surveys. The literature refers to all published and unpublished public reports on a topic. Systematic reviews of the literature describe current knowledge and reveal gaps. For instance, a review of the literature can provide information on best practices in teaching reading but may not provide sufficient information about the ease of implementing such programs in classrooms. A survey of reading teachers may provide data on implementation.

Survey objectives can also come from experts. Experts are individuals who are knowledgeable about a topic, will be affected by the survey's outcomes, or are influential in implementing its findings. Experts can be asked about objectives by mail or telephone or brought together in meetings. Two types of meetings that are sometimes used to help survey researchers identify objectives, research questions, and research hypotheses are focus groups (Krueger and Casey, 2000) and consensus panels (Jones, 1995).

Straightforward Questions and Responses

Survey questions take two primary forms. When they require respondents to use their own words, they are called open ended. When the answers or responses are preselected for the respondent, the question is termed closed or forced choice. Both types of questions have advantages and limitations.

An open question is useful when the intricacies of an issue are still unknown, for eliciting unanticipated answers, and for describing the world as the respondent sees it – rather than as the questioner does. Some respondents also prefer to state their views in their own words. Sometimes, when left to their own devices, respondents provide quotable material. The disadvantage is that unless you are a trained anthropologist or qualitative researcher, responses to open questions are often difficult to compare and interpret.

Some respondents prefer closed questions because they are either unwilling or unable to express themselves while being surveyed. Closed questions are more difficult to write than open ones, however, because the answers or response choices must be known in advance; however, the results lend themselves more readily to statistical analysis and interpretation, and this is particularly important in large surveys because of the number of responses and respondents. Moreover, as the respondent's expectations are more clearly spelled out in closed questions (or the survey researcher's interpretations of them), the answers have a better chance of being more reliable or consistent over time.

Case study 1

Illustrative objectives for a survey of vocational educational needs

- Objective 1: Identify the most common needs for educational services for three professions: teaching, nursing, and computer programming.
Sample question:
How proficient would you say you are in performing the following job-related activities?
- Objective 2: Compare the educational needs of men and women
Question:
Are you male or female?
- Objective 3: After participation in a vocational education program, identify the characteristics of participants who receive the most benefits.
Questions about the characteristics of survey participants:
What is your occupation? What was your household income last year?
Questions about benefits:
To what extent has this program helped you improve your job skills? How long did you wait to get a job in your preferred profession?

Survey Question Stems

Open-ended survey questions have only a stem: Why were you dissatisfied with the workshop? Close-ended questions have a stem and one or more responses: Which of the following describes the reasons for your dissatisfaction? (The question is followed by a list of potential reasons for dissatisfaction.)

A straightforward survey question asks what it needs in an unambiguous, concrete way and extracts accurate and consistent information (Sudman and Bradburn, 1982; Fink, 2004b). Straightforward questions are purposeful and use correct grammar (**Case study 2**).

Question stems should contain fewer than 20 words and be easy to understand, nonjudgmental, and unbiased. They should be phrased in a socially and culturally sensitive manner and at a reading level that is appropriate for the respondents. The survey researcher should avoid abbreviations and absolute terms (never, always). Question writing is complex, and requires ability to write simply. For many topics (e.g., perceptions, values, feelings, and beliefs) question writing also requires the researcher to understand psychological, social, and health theories and research. The difficulty in question-writing makes survey researchers often adopt or adapt already-existing and commonly used questions and even entire survey instruments to avoid the painstaking and often painful process of writing questions on their own. The Web has numerous sources for questions. For instance, questions on reading and math in schools can be found on the website of the National Assessment of Educational

Case study 2

Survey questions

- Purposeful questions: The relationship between the question and the survey's objective
- Survey objective: To find out about reading habits in the community
- Survey question: In what year were you born?
- Comment: The relationship between the question and the objective is far from clear. An explanation is needed.

Complete sentences:

- Poor: Place of birth?
- Comment: Place of birth means different things to different people. I might give the city in which I was born, but you might tell the name of the hospital.
- Better: Name the city and state in which you were born.

Concrete questions:

- Less concrete: How would you describe your mood?
- More concrete: In the past 3 weeks, would you say you were generally happy or generally unhappy?
- Comment: Adding a time period (3 weeks) and defining mood (generally happy and generally unhappy) adds precision to the question.

Progress. For demographic questions, researchers often search the literature using common article databases of Education Resources Information Center (ERIC), US Census Bureau, or the Web of Science.

Responses

The choices given to respondents for their answers can take three forms. The first is called nominal or categorical. (The two terms are often used interchangeably.) Categorical or nominal choices have no numerical or preferential values. For example, asking respondents if they are male or female asks them to name themselves as belonging to one of two categories: male or female. Any number associated with either category (e.g., 1 = male) has no inherent importance. The second response form is called ordinal. When respondents are asked to rate or order choices, say, from very positive to very negative, they are given ordinal choices. The third response form, numerical, asks for numbers such as age (e.g., number of years) or height (e.g., number of meters). Numerical responses produce data that are continuous (e.g., age) or discrete (e.g., number of books read in the past month). Categorical, ordinal, and numerical response choices are illustrated in **Case study 3**.

Survey Instruments

Self-Administered Questionnaires

A self-administered questionnaire consists of questions that an individual completes by oneself. These questionnaires may be completed on site (in a classroom or clinic) or at a distant location (home). Respondents receive self-administered questionnaires in the mail or are directed (via email) to a website. Many survey software vendors provide easily accessible and relatively inexpensive programs that enable the novice to design easy-to-use surveys that analyze survey responses in real time and provide user-friendly reports.

Interviews

An interview requires at least two people: one to ask the questions (the interviewer) and another to respond (the interviewee). (Group interviews are possible.) Interviews can take place on the telephone, face to face, or via teleconference or web cam. Telephone interviews may be relatively simple, involving just an interviewer and interviewee, or they may be relatively complex, relying on advanced technology to identify and query respondents (Bourque and Fielder, 2004).

Structured Record Reviews

A structured record review is a survey that uses a specially created questionnaire to guide the collection of data from financial, medical, school, and other records including

Case study 3

Categorical, ordinal, and numerical responses

1. Nominal response: Which of these books have you read? Please circle yes or no for each choice.

Book title	1. Yes	2. No
<i>Tom Sawyer</i>	1	2
<i>Tale of Two Cities</i>	1	2
<i>Etc. (six books are listed)</i>		

2. Ordinal: How important has each of the following books been in helping form your image of modern life? Please use the following scale to make your rating:

- 1 = Definitely unimportant
2 = Probably unimportant
3 = Probably important
4 = Definitely important
5 = No opinion/Don't know

Book title	Please circle one for each book				
<i>Tom Sawyer</i>	1	2	3	4	5
<i>Tale of Two Cities</i>	1	2	3	4	5
<i>Etc. (six books are listed)</i>					

3. Numerical: How many books did you read in the past month?

The response choice format dictates the analysis and report of results. For example, a report of the responses to the first question yielding categorical data could take this form:

Nominal or categorical data: More than 75% of the respondents read at least one book on the list, but no one read all six. Of 75 respondents, 46 (61.3%) indicated they had read *Tom Sawyer*, the most frequently read book.

A report for the second question (ordinal data) might look like this:

Ordinal data: Of 75 respondents completing this question, 43 (57.3%) rated each book as definitely or probably important. The average ratings ranged from 3.7 for *Tom Sawyer* to 2.0 for *A Tale of Two Cities*.

Typical survey findings for the third question (numerical) might appear as follows:

On average, respondents read about one book each month.

electronic, written, and filmed documents. An example of a structured record review is the use of a questionnaire to collect information from school attendance records on the number and characteristics (e.g., age, reading level) of students who are absent 4 or more weeks each semester.

Structured Observations

A structured observation is designed to guide the observer in focusing on specific actions or characteristics.

For example, two visitors to school are participants in a structured observation if both are asked to count and record the number of computers they see, look for the presence or absence of air conditioning, and measure the room's area in square feet.

Survey Sampling

A sample is a portion or subset of a larger group called a population. Surveys often use samples rather than populations. A good sample is a miniature version of the population – just like it, only smaller. The best sample is representative, or a model of the population. A sample is representative of the population if important characteristics (e.g., age, gender, and reading level) are distributed similarly in both groups. Suppose the population of interest consists of 1000 people, 50% of whom are male, with 45% over 13 years of age and older. A representative sample will have fewer people, say, 500, but it must also consist of 50% males, with 45% over age 13 years and older.

No sample is perfect. Usually, it has some degree of bias or error. To ensure a sample whose characteristics and degree of representation can be described accurately, the survey researcher must start with very specific and precise survey objectives. The researcher also must have clear and definite eligibility criteria, apply sampling methods rigorously, justify the sample size, and have an adequate response rate.

Eligibility Criteria

The criteria for inclusion in a survey refer to the characteristics of respondents who are eligible for participation in the survey; the exclusion criteria consist of characteristics that rule out certain people. The survey researcher applies the inclusion and exclusion criteria to the target population. Once the researcher removes from the target population all those who fail to meet the inclusion criteria and all those who succeed in meeting the exclusion criteria, the remaining study population, is eligible for participation. A major reason for setting eligibility criteria is that to do otherwise is simply not practical. Setting inclusion and exclusion criteria is an efficient way of focusing the survey on only those people from whom you are equipped to get the most accurate information or about whom you want to learn something.

Sampling Methods

Sampling methods are often divided into two types. The first is called probability sampling. Probability sampling provides a statistical basis for saying that a sample is

representative of the study or target population, and every member of the target population has a known, nonzero probability of being included in the sample. Probability sampling implies the use of random selection. Random sampling eliminates subjectivity in choosing a sample. It is a fair way of getting a sample.

The second type of sampling is nonprobability or convenience sampling. Nonprobability samples are chosen based on judgment regarding the characteristics of the target population and the needs of the survey. With nonprobability sampling, some members of the eligible target population have a chance of being chosen and others do not. By chance, the survey's findings may not be applicable to the target group at all. True probability sampling is extraordinarily complex. Nevertheless, many survey researchers aim to come as close to a probability sample as possible. One approach is simple random sampling.

In simple random sampling, every subject or unit has an equal chance of being selected. Members of the target population are selected one at a time and independently. Once they have been selected, they are not eligible for a second chance and are not returned to the pool. Because of this equality of opportunity, random samples are considered relatively unbiased. Typical ways of selecting a simple random sample are using a table of random numbers or a computer-generated list of random numbers and applying it to lists of prospective participants.

The advantage of simple random sampling is that the survey researcher can get an unbiased sample without much technical difficulty. Unfortunately, random sampling may not pick up all the elements of interest in a population. Suppose a researcher is conducting a survey of teacher satisfaction. Consider also that the researcher has evidence from a previous study that older and younger teachers usually differ substantially in their satisfaction. If the researcher chooses a simple random sample for the new survey, he or she might not pick up a large-enough proportion of younger teachers to detect any differences that matter. In fact, with really bad luck, by chance, the entire sample may consist of older teachers. To be sure that the sample consists of adequate proportions of people with certain characteristics, the researcher can use stratified random sampling.

A stratified random sample is one in which the population is divided into subgroups or strata, and a random sample is then selected from each subgroup. For example, suppose a survey researcher wants to determine the effectiveness of a program to care for the health of homeless families. The researcher plans to survey a sample of 1800 of the 3000 family members who have participated in the program. The researcher also intends to divide the family members into groups according to their general health status (as indicated by scores on a 32-item test) and age. Health status and age are the strata. The strata or subgroups are chosen because evidence is available that they

are related to the outcome – in this case, care for health needs of homeless families. The justification for the selection of the strata can come from the literature and expert opinion.

Stratified random sampling is more complicated than simple random sampling. The strata must be identified and justified, and using many subgroups can lead to large, unwieldy, and expensive surveys.

Systematic sampling is another sampling method used by survey researchers. Suppose a researcher has a list with the names of 3000 students, from which a sample of 500 is to be selected. Dividing 3000 by 500 yields 6, which means that one of every six persons will be in the sample. To systematically sample from the list, a random start is needed. To obtain this, a die can be tossed. Suppose a toss comes up with the number 5. This means the fifth name on the list would be selected first, then the 11th, 17th, 23rd, and so on until 500 names are selected.

To obtain a valid sample, the researcher must obtain a list of all eligible participants or members of the population. This is called the sampling frame. Systematic sampling should not be used if repetition is a natural component of the sampling frame. For example, if the frame is a list of names, systematic sampling can result in the loss of names that appear infrequently (e.g., names beginning with X). If the data are arranged by months and the interval is 12, the same months will be selected for each year. Infrequently appearing names and ordered data (January is always month 1 and December month 12) prevents each sampling unit (names or months) from having an equal chance of selection. If systematic sampling is used without the guarantee that all units have an equal chance of selection, the resultant sample will not be a probability sample. When the sampling frame has no inherently recurring order, or you can reorder the list or adjust the sampling intervals, systematic sampling resembles simple random sampling.

A cluster is a naturally occurring unit (e.g., a school, which has many classrooms, students, and teachers). Other clusters are universities, hospitals, cities, states, and so on. The clusters are randomly selected (called cluster sampling), and all members of the selected cluster are included in the sample. For example, suppose that California's counties are trying out a new program to improve physical education for teens. A researcher who wanted to use cluster sampling can consider each county as a cluster and select and assign counties at random to the new physical education program or to the traditional one. The programs in the selected counties would then be the focus of the survey.

Cluster sampling is used in large surveys. It differs from stratified sampling in that with cluster sampling you start with a naturally occurring constituency. The researcher then selects from among the clusters and either surveys all members of the selection or randomly selects from among them. The resulting sample may not be

representative of areas not covered by the cluster, nor does one cluster necessarily represent another.

Nonprobability samples do not guarantee that all eligible units have an equal chance of being included in a sample. Their main advantage is that they are relatively convenient, economical, and appropriate for many surveys. Their main disadvantage is that they are vulnerable to selection biases. Convenience sampling is a type of nonprobability sampling in which the researcher surveys individuals who are ready and available. For example, a survey that relies on people in a shopping mall is using a convenience sample.

Snowball sampling relies on previously identified members of a group to identify other members of the population. As newly identified members name others, the sample snowballs. This technique is used when a population listing is unavailable and cannot be compiled. For example, surveys of teenage gang members and undocumented workers might be asked to participate in snowball sampling because no membership list is available.

Quota sampling divides the population being studied into subgroups such as male and female and younger and older. Then you estimate the proportion of people in each subgroup (e.g., younger and older males and younger and older females).

Sample Size

The size of the sample refers to the number of units that that will be surveyed to get precise and reliable findings. The units can be people (e.g., boys and girls over and under 16 years of age), places (e.g., counties, hospitals, and schools), and things (e.g., medical or school records). The number of needed units is influenced by a number of factors, including the purpose of the study, population size, the risk of selecting a bad sample, and the allowable sampling error (Kraemer and Thiemann, 1987; Cohen, 1988). When you increase the sample size, you increase the survey's cost. Larger samples mean increased costs to provide participants with financial or other incentives and to follow up with nonresponders.

The most appropriate way to produce the right sample size is to use statistical calculations. These can be relatively complex, depending on the needs of the survey. Some surveys have just one sample, and others have several. Formulas for calculating survey samples can be found on the Internet. Type in the words, "sample size calculator."

Response Rate

All surveys hope for a high response rate. The response rate is the number that responds (numerator) divided by

the number of eligible respondents (denominator). Practically all surveys are accompanied by a loss of information because of nonresponse. These nonresponses may introduce error into the survey's results because of differences between respondents and nonrespondents.

Reliable and Valid Survey Instruments

A reliable survey instrument is consistent; a valid one is correct (Litwin, 2004). For example, an instrument is reliable if each time you use it (and assuming no intervention), you get the same information. But it may not be correct! Reliability or the consistency of information can be seriously imperiled by poorly worded and imprecise questions and directions. If an instrument is unreliable, it is also invalid because inconsistent data are incorrect. Valid survey instruments serve the purpose they were intended to and provide correct information. For example, if a survey's objective is to find out about mental health, the results should be consistent with other measures of mental health and inconsistent with measures of mental instability. Valid instruments are always reliable, too.

Reliability

A reliable survey instrument is one that is relatively free from measurement error. This error, caused individuals to obtain scores that are different from their true scores, which can only be obtained from perfect measures. What causes this error? In some cases, the error results from the measure itself—it may be difficult to understand or poorly administered. For example, a self-administered questionnaire on the value of preventive healthcare might produce unreliable results if its reading level is too high for the teen mothers who are to use it. If the reading level is on target but the directions are unclear, the measure will be unreliable. Of course, the survey researcher can simplify the language and clarify the directions and still find measurement error. This is because measurement error can also come directly from the examinees. For example, if teen mothers are asked to complete a questionnaire and they are especially anxious or fatigued, their obtained scores could differ from their true scores.

Four kinds of reliability are often discussed: stability, internal consistency, and inter- and intrarater reliability.

Test-Retest Reliability or Stability

A measure is stable if the correlation between scores from one time to another is high. Suppose a survey of students' attitudes was administered to the same group of students at School A in April and again in October. If the survey

was reliable and no special program or intervention was introduced, then, on average, we would expect attitudes to remain the same. The major conceptual difficulty in establishing test–retest reliability is in determining how much time is permissible between the first and second administration. If too much time elapses, external events might influence responses for the second administration; if too little time passes, the respondents may remember and simply repeat their answers from the first administration.

When testing alternate-form reliability, the different forms may be administered at separate time points to the same population. Alternatively, if the sample is large enough, it can be divided in half and each alternate form administered to half the group. This technique, called the split-halves method, is generally accepted as being as good as administering the different forms to the same sample at different time points. When using the split-halves method, you must make sure to select the half-samples randomly.

Internal Consistency or Homogeneity

A survey’s internal consistency or homogeneity refers to the extent to which all the items or questions assess the same skill, characteristic, or quality. Cronbach’s coefficient alpha, the average of all the correlations between each item and the total score, is often calculated to determine the extent of homogeneity. For example, suppose a survey researcher created a questionnaire to find out about students’ satisfaction with Textbook A. An analysis of homogeneity will tell the extent to which all items on the questionnaire focus on satisfaction.

Equivalence or alternate-form reliability is a type of internal consistency. If two items measure the same concepts at the same level of difficulty, they are equivalent. Suppose students were asked a question about their views toward technology before participating in a new computer skills class and again 2 months after completing it. Unless the survey researcher was certain that the items on the surveys were equal, more favorable views on technology after the second administration could reflect the survey’s language level, for example, rather than improved views.

Some variables do not have a single dimension. Student satisfaction, for example, may consist of satisfaction with school in general, their school in particular, teachers, classes, extracurricular activities, and so on. If you are unsure of the number of dimensions expressed in an instrument, a factor analysis can be performed. This statistical procedure identifies factors or relationships among the items or questions.

Inter- and Intrarater Reliability

Interrater reliability refers to the extent to which two or more individuals agree. Suppose two individuals were

sent to a clinic to observe waiting times, the appearance of the waiting and examination rooms, and the general atmosphere. If the observers agreed perfectly on all items, then interrater reliability would be perfect. Interrater reliability is enhanced by training data collectors, providing them with a guide for recording their observations, monitoring the quality of the data collection over time to see that people are not burning out, and offering a chance to discuss difficult issues or problems. Intrarater reliability refers to a single individual’s consistency of measurement, and this, too, can be enhanced by training, monitoring, and continuous education.

Validity

Validity refers to the degree to which a survey instrument assesses what it purports to measure. For example, a survey of student attitude toward technological careers would be an invalid measure if the survey only asked about their knowledge of the newest advances in space technology. Similarly, an attitude survey will not be considered valid unless you can demonstrate that people who are identified as having a good attitude on the basis of their responses to the survey are different in some observable way from people who are identified as dissatisfied.

Four types of validity are often discussed: content, face, criterion, and construct.

Content Validity

Content validity refers to the extent to which a measure thoroughly and appropriately assesses the skills or characteristics it is intended to measure. For example, a survey researcher who is interested in developing a measure of mental health has to first define the concept (“What is mental health?” “How is health distinguished from disease?”) and then write items that adequately contain all aspects of the definition. The literature is often consulted either for a model or for a conceptual framework from which a definition can be derived because of the complexity of the task. It is not uncommon in establishing content validity to see a statement like “We used XYZ cognitive theory to select items on mental health, and we adapted the ABC role model paradigm for questions about social relations.”

Face Validity

Face validity refers to how a measure appears on the surface: Does it seem to ask all the needed questions? Does it use the appropriate language and language level to do so? Face validity, unlike content validity, does not rely on established theory for support.

Criterion Validity

Criterion validity compares responses to future performance or to those obtained from other, more well-established surveys. Criterion validity is made up two subcategories: predictive and concurrent. Predictive validity refers to the extent to which a survey measure forecasts future performance. A graduate school entry examination that predicts who will do well in graduate school has predictive validity. Concurrent validity is demonstrated when two assessments agree or a new measure is compared favorably with one that is already considered valid. For example, to establish the concurrent validity of a new survey, the survey researcher can either administer the new and validated measure to the same group of respondents and compare the responses, or administer the new instrument to the respondents and compare the responses to experts' judgment. A high correlation between the new survey and the criterion means concurrent validity. Establishing concurrent validity is useful when a new measure is created that claims to be better (shorter, cheaper, and fairer).

Construct Validity

Construct validity is established experimentally to demonstrate that a survey distinguishes between people who do and do not have certain characteristics. For example, a survey researcher who claims constructive validity for a measure of satisfaction will have to demonstrate in a scientific manner that satisfied respondents behave differently from dissatisfied respondents. Construct validity is commonly established in at least two ways:

1. The survey researcher hypothesizes that the new measure correlates with one or more measures of a similar characteristic (convergent validity) and does not correlate with measures of dissimilar characteristics (discriminant validity). For example, a survey researcher who is validating a new quality-of-life survey might posit that it is highly correlated with another quality-of-life instrument, a measure of functioning, and a measure of health status. At the same time, the survey researcher would hypothesize that the new measure does not correlate with selected measures of social desirability (the tendency to answer questions so as to present yourself in a more positive light) and of hostility.
2. The survey researcher hypothesizes that the measure can distinguish one group from the other on some important variable. For example, a measure of compassion should be able to demonstrate that people who are high scorers are compassionate but that people who are low scorers are unfeeling. This requires translating a theory of compassionate behavior into measurable

terms, identifying people who are compassionate and those who are unfeeling (according to the theory), and proving that the measure consistently and correctly distinguishes between the two groups.

Quality of Survey Data

The quality of survey data depends on how well respondents understand the survey's questions. Their comprehension may be affected by language skills, education, and culture. A pretest of the survey enables the survey researcher to evaluate whether respondents reliably interpret questions and that they do so as they survey researcher intends. Usually small groups (five to ten people) who are similar in demographic and languages skills to prospective respondents evaluate each survey questions individually or in a group. The idea is to decide whether to accept the original question and meaning, to change the question but keep the meaning, or to eliminate the question or write a new one. Pretest participants are also asked to examine the survey with regard to its flow and administrative ease, identifying unusual, redundant, irrelevant, or poorly worded question stems and responses. They are also asked to record the time required to complete the survey.

Pilot testing is a miniature version of an experiment in which a sample of respondents is chosen to complete the survey for the purpose of evaluating its effectiveness in practice. In theory all the technical problems that can be anticipated have already been resolved in the pretest. A pilot test is, however, the reality test. If the respondents in the real setting are not as motivated as those in the pretest environment, differ in their language or other skills, or are unfamiliar with surveys, the researcher may have to go back to the drawing board.

Bibliography

- Bourque, L. B. and Fielder, E. P. (2004). *How to Conduct Telephone Surveys*. Thousand Oaks, CA: Sage.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. New York: Academic Press.
- Fink, A. (ed.) (2004a). *The Survey Kit*. Thousand Oaks, CA: Sage.
- Fink, A. (2004b). *How to Ask Survey Questions*. Thousand Oaks, CA: Sage.
- Fink, A. (2008a). *How to Conduct Surveys*, 3rd edn. New Delhi: Sage.
- Fink, A. (2008b). *Practicing Research: Discovering Evidence that Matters*. New Delhi: Sage.
- Jones, J. (1995). Qualitative research: Consensus methods for medical and health services research. *British Medical Journal* **311**, 376–380.
- Kraemer, H. C. and Thieman, S. (1987). *How Many Subjects? Statistical Power Analysis in Research*. Thousand Oaks, CA: Sage.
- Krueger, R. A. and Casey, M. A. (2000). *Focus Groups: A Practical Guide for Applied Research*. Thousand Oaks, CA: Sage.

- Litwin, M. (2004). *How to Assess and Interpret Survey Psychometrics*. Thousand Oaks, CA: Sage.
- Sudman, S. and Bradburn, N. M. (1982). *Asking Questions*. San Francisco, CA: Jossey-Bass.

Further Reading

- Campbell, D. T. and Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally.

Relevant Websites

- <http://www.bmj.com> – BMJ Publishing Group Ltd.
- <http://www.eric.gov> – Education Resources Information Center.
- <http://nces.ed.gov> – The National Center for Education Statistics.
- <http://scientific.thomson.com> – The Thomson Scientific Database.
- <http://www.census.gov> – US Census Bureau.

EDUCATIONAL ASSESSMENT

Contents

An Overview of Educational Assessment

Assessment and the Evaluation of Institutional Effectiveness

Assessment in Schools – Dispositions

Assessment in Schools Related To Literacy: Reading

Impact of Assessment on Students' Learning Strategies and Implications for Judging Assessment Quality

An Overview of Educational Assessment

M James, University of Cambridge, Cambridge, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This section on educational assessment is new for this encyclopedia. There is another, on educational measurement, that shares some of its themes but there are important respects in which the two sections differ. Readers who are interested in the general field of assessment and testing are advised to peruse both sections.

As Robert Linn makes clear in his overview of the 'Educational measurement' section, measurement in education is closely associated with formal testing for purposes associated with the placement of students, the award of qualifications, monitoring achievement and progress, and holding teachers, schools, school districts, states, and nations accountable for the quality of the public services they provide. The validity and reliability of such measures is paramount in order to make useful and fair comparisons between individuals, groups, and organizations. The basis for this is provided by the theory and techniques of the well-developed science of psychometrics, which offers criteria for the evaluation of tests, their results, and their uses. These are sophisticated procedures and require considerable expertise, time, and other resources to develop and implement them properly. For this reason, educational measurement is often large scale and carried out, at least in part, by professional agencies external to the schools in which tests are administered. The examples that Linn quotes illustrate this.

Such educational measurement, in the form of tests, is often an element of educational assessment but the latter generally has a wider scope in terms of purpose, form, agency and use. Indeed, the word assessment has its roots in the Latin verb *ad sedere* meaning to sit beside, a notion somewhat remote from familiar images of examination halls with students writing silently at separated desks. In recent years the idea that assessment might have something important to do with the teacher sitting beside

the student (literally or metaphorically), and gaining an understanding of what the student knows and can do in order to help them move on in their learning, has begun to gain new ground. Linn mentioned this formative function of assessment in his overview although it was not dealt with fully because it has more to do with good instructional practices than with technical measurement characteristics. In this section, the relationship between educational assessment, especially formative assessment, and instructional practices in classrooms is a prime concern and a central focus for many of the articles. Some of the subject matter looks similar in both sections, for example, there is an article on 'Achievement testing in mathematics' in the 'Educational measurement' section and one on 'Mathematics assessment in schools' in this 'Educational assessment' section. However, the emphasis shifts from the concerns of test developers to the issues for educators, from the technical to the professional, from tests as the dominant form of assessment to a consideration of alternative activities, and from reliability as perhaps the most powerful criterion of quality to concerns with validity in terms of the consequences of assessments for the experience of students and schools. This is not to imply that the concerns of test developers are unimportant but to offer a somewhat different, and equally important, perspective on these complex phenomena.

There is one more thing that should be said about the meanings of words. English is a rich language with many synonyms for broadly similar ideas although there are subtle variations in meaning and usage in different Anglophone countries. The term assessment is a case in point. In the UK it is used widely for all those activities that involve eliciting evidence of student learning and drawing inferences as the basis for decisions. In the US, these processes are often referred to as student evaluation and we could equally well have used this as the title of this section.

However, in the UK, and elsewhere, the term evaluation is more often used for procedures for collecting evidence and judging the worth of educational programs and institutions. The 'Evaluation' section in this encyclopedia has this particular focus. For all these reasons, we chose 'Educational assessment' for this section.

The articles in this section have been clustered under five loosely defined topic headings but anyone who has the interest and stamina to read them all will note a number of recurring themes and issues. The remainder of this overview article introduces the topics and draws out some of the most prominent cross-cutting themes and issues.

Topics

Key Relationships: Impact of Assessment

Making the impact of assessments a first consideration may seem to be putting the cart before the horse, but the consequences of assessment practices, for learners, for curriculum, for instructional practices and for institutions, are a crucial concern and the source of much debate. Educational practice in classrooms involves interactions among curriculum, assessment, and instruction (key elements), and between teachers and students (key actors). In this complex system of classroom activity, assessment cannot easily be abstracted and examined without reference to the other components because change in assessment practice will almost certainly have knock-on effects – for good or ill. An early illustration of this can be found in the very first annual report, in 1859, of the examiners of the University of Cambridge Local Examinations Syndicate. On the performance of candidates in English history they opined:

...forty-six failed to pass, but this number of failures does not measure the full extent of the general deficiency. The questions were of an elementary character, and the Candidates had received notice of the range of reading within which their knowledge would be tested. But their answers, even when accurate, shewed (sic) a general uniformity of expression which seemed to imply that meager handbooks had been placed before the students to be 'got up', and that little attempt had been made by their instructors to excite the interest of their pupils by questionings or remarks of their own. (reproduced in Watts, 2008)

So history suggests that the consequences of assessment practice, some of them unintended, can be profound and need to be considered in the planning, development, implementation, use, and evaluation of any assessment policy. An examination of current evidence of impact is therefore an important basis for deliberation about future alternatives.

The articles in this cluster focus particularly on impact on classroom practice and on learners. Pollard and Filer

provide an account, drawn mainly from ethnographic research, of the powerful influence of students' experiences of assessment through their school careers on the formation of their identities as learners. Buchwald and Schwarzer focus specifically on the impact of test anxiety. Using conservation of resources (COR) theory, they conceptualize test anxiety as a loss of resources (self-worth, motivation, and productive cognitive processing) that deleteriously affects learning and performance for individuals and groups. Other articles examine the impact on specific learner groups: boys and girls and learners with disabilities. We had wanted to have articles in this collection about the impact of assessment on students from different ethnic groups and on different socioeconomic status groups but lack of sufficient contemporary research with these specific foci, possibly because relevant background data have not been routinely collected, made this unfeasible. Perhaps this gap will be filled in the next edition of this encyclopedia.

Assessment Activities

Teacher-administered tests continue to be a prominent feature of assessment activity in many classrooms. Indeed, so dominant is this experience that the popular press, at least in the UK, tends to substitute the monosyllabic test for assessment even where there is reference to a greater range of activities. In some instances tests are developed by testing agencies or awarding bodies and are mandatory as part of national, state, or school district monitoring or qualifications systems, or they are bought in by schools for their own purposes. Sometimes they are constructed by the teachers themselves but usually without the technical expertise necessary to ensure their quality, which limits the validity of the inferences drawn from their results. For example, Martinez Rizo describes how, in Latin American countries in the mid-twentieth century, teachers were asked to prepare multiple-choice tests and standardize their scores by grading on the (bell) curve without the statistical training that is necessary (see the articles in the section on 'Educational measurement'). This poor practice is not confined to one time or place; the present author confesses to doing it herself as a newly qualified teacher in England in the 1970s, mistakenly believing that she was being scientific.

In character, tests are usually pencil and paper, time-limited, administered to individual students under controlled conditions, and composed of multiple-choice or constructed response items. Multiple-choice tests dominate in the USA but in the UK and elsewhere, constructed response items, essays, and coursework projects are more familiar. Historically, these latter are associated with the different tradition of examinations derived less from the science of psychometrics and more from practices in universities, which, in turn, adapted forms of induction practiced by the ancient guilds. The craft apprentice

traditionally completed his training by producing a masterpiece that is, a piece of work for the master, in which he attempted to demonstrate all that he had learnt in a single project of the highest quality. This resonates with current practice in modern universities where theses or dissertations are submitted by students for examination on completion of masters or doctoral programs.

Within schools, concerns about the value of conventional tests have stimulated debate about alternative forms of assessment. This has arisen especially when there is interest in purposes other than making comparisons between the abilities, aptitudes, or attainments of students in rather narrowly defined domains, some of which are only loosely related to what is actually taught and learned in the classroom. Many of the articles in this cluster focus on these alternative approaches. For interested readers, the article by Stobart and Gipps is a good place to start because it distinguishes alternative forms or formats of assessment from alternative models of assessment and from alternative purposes of assessment. This classification provides a useful framework for locating the discussion in other articles, which have more specific foci. For example, there are articles on computer-based instructional systems that provide feedback, dynamic assessment, diagnostic dialog, peer and self-assessment, portfolio assessment, and records of achievement. On the surface it may appear that these are about alternative forms of assessment but the justification for these alternative approaches is based on arguments about their fitness for different purposes, especially instructional purposes, and their congruence with more powerful models of how assessment practice can contribute to better learning. (This theme is expanded below.)

Assessment in Domains

Large-scale testing for system monitoring at international, national, or state level tends to focus on a limited number of constructs. They typically assess attainments in subdomains of core subjects, notably mathematics and national language and science, or generalized aptitudes for further study (see the section on 'Educational measurement'). There are two main reasons for these domain restrictions.

First, robust and dependable tests are expensive to develop, trial, and administer, especially constructed response tests which demand armies of trained markers. Such testing of all students for all the learning outcomes of potential interest would not be manageable or affordable. Large-scale national testing in even limited subject areas has proved difficult, as the fiasco over national testing in England in 2008 clearly demonstrated. The failure to deliver, mark, and report test results, for around one million 11- and 14-year-olds in English, math, and science, led the Government to cancel the contract with

the testing agency (Educational Testing Service (ETS), Europe), abolish the National Assessment Agency, and suspend the chief executive of the Qualifications and Curriculum Authority.

Second, unless countries have a mandatory national curriculum that applies to all maintained schools, as in England, it is not possible to assume that all students will have the same opportunity to learn and achieve across a whole curriculum, or even that there will be consensus about what curriculum content should be. Therefore what is tested tends to be the basics that most students are expected to encounter. International surveys of achievement are sometimes criticized for focusing on the lowest common denominator, that is, what is covered in most countries, rather than what might best reflect aspirations for future world citizens.

When one moves away from concern with large-scale, whole cohort, or large sample testing, for high stakes accountability purposes, the picture changes. When the focus is on assessment of the achievements of particular groups of students in particular subjects, courses, or phases of their education, for the award of qualifications, for reporting to parents or for diagnostic or formative purposes, then there is more scope for debate and innovation in assessment according to the particular characteristics of the domain. The articles in this cluster therefore focus on these debates and innovations within specific domains.

A large group focuses upon subject domains, especially those of the school curriculum. There are articles on what is widely regarded as the core of reading, writing, oracy, mathematics, and science (for science in elementary schools see the article 'Assessment in schools: primary science' and 'Assessment in schools: secondary science' for science in high schools). There are also contributions on areas of the broader curriculum: creative subjects, history, civics, and social studies, information and communications technology (ICT) and technology. Another small group focuses upon the assessment of what might be called cross-curricular or generic skills and attitudes such as learning to learn, learning dispositions, and the affective domain more broadly. Many countries, for example, Hong Kong and the whole of Europe (stimulated by the Lisbon Strategy 2000), are moving to giving more prominence to such skills and dispositions on the grounds that these are essential preparation for lifelong learning in a rapidly changing world. However, a key issue is the extent to which these can be taught, learned, and assessed in a way that is detached from learning something in a cognate field and whether the expression and application of such skills and dispositions varies with context. The stronger argument is for the importance of context although this creates challenges for curriculum planning, instruction, and assessment. The implication is that skills, dispositions, and knowledge are highly integrated and this needs to be recognized in assessment practice.

Mainstream education in schools is well covered by the aforementioned articles. However, another group of articles within this cluster focuses more particularly on assessment within different phases or sectors of education and training: early years, vocational training, higher education, and the adult workplace. All these articles challenge assumptions about what constitutes good assessment practice, derived as many of these assumptions are from the dominance of experience in schools. They provide thoughtful alternatives to most of the why, what, who, where, and how questions; as such they should cause those who work predominantly in schools to reflect further on the formats, models, and purposes of assessment.

Quality and Use: Classroom Assessment in Policy Contexts

Even in this International Encyclopedia, space does not permit a thorough examination of all the different expressions of assessment policy in all countries of the world. For this reason the articles in this cluster are an indicative selection chosen because they represent interesting and/or important trends in policy and practice, and where evidence is available. Readers should be aware that there is a bias towards accounts of Anglophone countries although some redress is attempted by inclusion of an article on French sources, a comparison between England and France, the article on Latin America mentioned earlier, and others exploring developments in Eastern Europe and Africa. In terms of research, analysis, and commentary, great attention has been paid internationally to the development of assessment systems in Australia, and New Zealand, and articles are included on these. Although a small country, Hong Kong's recent experience also warranted an article because, soon after the former British colony was handed back to China in 1997, this Special Administrative Region embarked on a 10-year program of educational reform that included new senior secondary examinations, basic competency testing, school-based assessment and a commitment to assessment for learning (AfL). Hong Kong has self-consciously struggled to create a system in which assessment of learning for reporting, selection, and accountability is in balance with AfL for educational growth. In this sense it represents, in microcosm, the struggles that are implicit in many other policy contexts although the ways in which the problems are formulated, and tackled, may differ.

In reviewing this cluster of articles, it is worth reflecting on the diverse ways in which assessment systems evolve over time. Where there is little history of assessment policy, the instinct may be to adopt objective, scientific testing as the principal instrument of policy. But as countries become more sophisticated in the use of tests, they also become more aware of limitations in the way they are used and the challenges they present to policy formation, professional

practice, and public understanding. Thus, more mature systems show a shift in focus towards enhancing the quality of everyday classroom assessments, especially for formative purposes, and attempting to find the best balance with summative assessment and testing to meet the legitimate demands for accreditation and accountability. More research would be needed to test this hypothesis but the articles in this group provide some evidence for this proposition.

Concepts and Issues

This final cluster of articles focuses on a number of key ideas and discussions. Many of these ideas have already been mentioned in this overview and some are expanded a little more in the final part of this article.

At the heart of debates surrounding assessment and testing are issues of purpose and use, that is, what assessment practice is intended to do and the functions it actually performs, which may or may not coincide. Newton identifies 22 purposes although he emphasizes that a single assessment can serve a range of purposes, although not all. A number of concepts related to purpose have come to particular prominence in recent years, partly because there are continuing debates about what they mean and how they should be distinguished. Black, who is particularly associated with research in the area, provides an article on formative assessment (or AfL) and Daugherty provides one on summative assessment (or AoL) by teachers. Both were members of the influential UK Assessment Reform Group (1989–2010) who distinguished AfL from AoL (ARG, 1999) and defined AfL in a way that has been widely used and variously reinterpreted (ARG, 2002a, quoted below).

Daugherty makes the point that the quality of teachers' summative assessments and the credibility ascribed to the results are dependent on procedures introduced to increase accuracy, consistency, and fairness in their judgments. Moderation by teachers is one such procedure, discussed by Maxwell. This represents an important and valuable form of assessment training for teachers, which is crucial for effective classroom assessment practice whether targeted at formative or summative purposes. Davies discusses this further. Pedder expands the discussion to focus on the need for teachers to be supported to develop or change their practices, their corresponding beliefs about learning, and their conceptions of classroom roles for teachers and learners. Both he and Sutton, also highlight the implications and challenges for the leadership and management of schools.

At organizational level, in most systems, school managers have to balance the formative and summative uses of assessment data, which sometimes appear in tension. The high stakes attached to school accountability in centralized systems where severe consequences flow from failure to perform well can encourage games-playing and

data manipulation in order to show institutions in the most favorable light. The article by Thomas argues that the publication of raw statistics cannot give an accurate picture of how effective a school and its teachers are at raising and maintaining the achievement of all students, but that value-added methodologies are available for understanding the mechanisms of school improvement.

A final group of articles in this cluster explores the question of what makes assessment educational and some of the deeper issues surrounding assessment practice. In particular, these articles examine the relationship of assessment to progression in learning, instructional planning, the regulation of learning in the activity of the classroom, assessment of group work, and the role of three-way assessment conversations involving students, teachers, and parents. Underpinning all of these concepts and issues are some fundamental debates about the models best able to explain how assessment practice can contribute to better learning.

So far this overview has been very broad brush. The remainder of this article discusses some of the themes, introduced above, in a little more detail because they recur in many of the contributions collected in this section. The aim is to draw out what seem to be the dominant trends in research and scholarship at the current time.

Cross-Cutting Themes and Issues

What Makes Assessment Educational?

Assessment for grading, selection, and certification (summative purposes) has a two thousand year history, stretching back to the Han dynasty in China. In contrast, the stated aspiration that assessments should fulfill formative purposes is more recent. Although the distinction between formative and summative has been used for several decades, a turning point came when Paul Black and Dylan Wiliam, of King's College London, published a pamphlet, *Inside the black box*, derived from an extensive review of research on assessment and classroom learning (Black and Wiliam 1998). (This is referred to in many articles in this section.) The review was commissioned by the UK Assessment Reform Group (ARG), which perceived a need to update an earlier review by Crooks (1988). The central thesis of Black and Wiliam's review was that there is a body of firm evidence that formative assessment is an essential feature of effective pedagogy and its development can raise standards. The authors put a figure on the size of measured gains and pointed to effect sizes in the range of 0.4–0.7, amongst the largest for any educational intervention. It was probably these figures, and extrapolations that indicated what they might mean in terms of scores on national tests, examinations, and international surveys of achievement, which provoked policy makers to take notice. However, some

confusion about what the term formative assessment actually meant encouraged the ARG to adopt a distinction between AfL and assessment of learning as a more accessible version of the formative/summative distinction (ARG, 1999). Both pairs of terms are now widely used, within and beyond the UK, as the articles in this section testify.

All assessments involve certain processes associated with making observations of performance, interpreting the evidence, and making judgments that can inform decisions. For formative purposes these processes have been construed as:

...the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there. (ARG, 2002a)

However, no assessment can properly be described as formative unless assessment information is actually used to bring about improved learning. It is only by so doing that assessment in education becomes educational assessment.

Practices Associated with Formative Assessment

Black and Wiliam's review of research provided important guidance about which particular classroom practices, as part of everyday pedagogy, would help teachers to fulfill the aims of formative assessment (see also the article 'Formative assessment'). These are:

- Sharing learning intentions and success criteria with learners so that they understand criteria of quality (standards). This can be done at any point in a lesson and through various devices, including discussion of exemplars.
- Developing classroom talk and questioning to elicit evidence of knowledge and understanding but also to identify difficulties and opportunities, including those associated with prior learning and experiences.
- Giving appropriate feedback that identifies strengths and weaknesses but, most importantly, helps learners know what to try next in order to improve. Sometimes this is referred to as feed forward.
- Providing opportunities for peer- and self-assessment so that learners can eventually carry out the above practices for themselves and become autonomous learners.

In a follow-up development project, Black *et al.* (2003) also investigated the possibilities of using summative assessments formatively but this tends, because of a time lag, to be formative for teachers' future instructional practices rather than formative for the students who sat the tests. This is a useful reminder that formative for 'what' and 'whom' is a pertinent question to ask.

Relationships between Formative and Summative Assessment

Many of the articles in this section deal with both summative and formative assessment and it is important to emphasize that these labels attach only to the purposes and uses to which assessments are put, and not to their formats. In other words, a pencil-and-paper test, or quiz, can be used formatively in the classroom, provided that the teachers plan time in their lessons for discussions of results that go deeper than scores. Analysis of responses to particular items can lead into productive discussions about what students did well or poorly, what this indicates about their learning, and how they might improve their understanding and performance. Similarly, it is possible that assessments carried out primarily for formative purposes in the classroom, even those of an informal nature such as dialog with a group of students engaged in a learning activity, can be reviewed subsequently for evidence to sum up what individuals or groups have achieved over a period.

What is of concern, certainly in England where AfL has become enshrined in national policy, is that understanding of the formative dimension is unclear and in danger of being lost. For example the Primary and Secondary National Strategies of 2008 made reference to definitions of AfL and research-based accounts of good practice. However, these materials also reveal tensions with researchers' definitions of AfL. For example, they imply that AfL can be formative, or summative, or both. The government in England has invested a great deal in the development of pupil tracking and planning tools, to help teachers and principals use the results of statutory national tests for monitoring, prediction, and target setting. It is politically expedient therefore to promote frequent mini-summative assessment, to secure higher performance on tests to meet prescribed numerical targets, rather than use scarce resources on less-tangible approaches to formative assessment for the purpose of enhancing deep and lasting learning.

The distinction between learning and performance is a subtle one and not well understood. Measured performance should indeed be an indicator of underlying learning (or what Dweck, 2000, calls mastery) but debates about the validity of assessments underscore the difficulties of making such assumptions. It is quite possible to drill pupils to perform well on tests without enhancing learning. Given the high stakes consequences for schools that perform badly, there is increasing evidence that this is happening, at least in England (ARG, 2002b).

Despite a decision, following the debacle in 2008, to abolish national tests for 14-year olds in England, the underlying rationale did not change significantly. The drive was still to raise standards as measured by national curriculum assessment levels and sublevels, and

assessment for learning continued to be seen as an instrument for this purpose. More nuanced ideas, about the role of AfL in pedagogy to enhance the learning of capable, resourceful, and autonomous citizens in a changing world, seemed almost entirely absent. One possible explanation for the mixed messages that appeared in policy documents is that the authors have tried to finesse competing claims between those who are convinced by research that formative assessment is the key to improved learning and achievement, and those who believe that the pressure of regular testing raises standards. Or the mixed messages may simply be indicative of some confused thinking that has elided learning and performance.

Models of Assessment and Models of Learning

If assessment is to become educational, when it is integrated into pedagogy, then there needs to be a considerable degree of alignment or congruence between the learning outcomes sought, learning opportunities and instructional practices, and the assessment processes used to detect outcomes and promote learning. This includes concerns for consistency in the theoretical foundations of practice. A relevant question to ask, as Stobart and Gipps do, is "What models of assessment are consonant with alternative perspectives on learning?" (Alternative models, perspectives or theories of learning are themes in other sections of this encyclopedia; for example, see the article 'Learning, cognition and language: an overview', which provides an overview of learning and cognition.)

Greeno *et al.* (1996) argue that four perspectives on the nature of human mind should be taken into account when constructing assessments: the differential, the behaviorist, the cognitive, and the situative (or what Europeans tend to describe as sociocultural). Pellegrino *et al.* (2001) also note:

Most current tests, and indeed many aspects of the science of educational measurement, have theoretical roots in the differential and behaviorist traditions. The more recent perspectives – the cognitive and situative – are not well reflected in traditional assessments but have influenced several recent innovations in the design and use of educational assessments. (p. 60)

Pellegrino *et al.*'s key text provides many examples from mathematics and science of how the cognitive (sometimes labeled constructivist) perspective has been incorporated into the recent design of assessments. Particular emphasis is given to ways of detecting how people organize information in schema in long-term memory, how they use this knowledge to solve problems, how they develop metacognitive skills, and how they engage in

communicative practices appropriate to subject domains. However, the treatment of situative or sociocultural approaches is slight in that the social dimension is largely interpreted as a context for individual learning, rather than integral and inseparable.

Helpfully, Pellegrino *et al.* advise that assessment methods can be expanded beyond traditional tests by drawing on the ways in which research is conducted to investigate learning according to these different theories in domains. They also suggest ways in which assessment can be integrated with learning. For cognitive approaches, computerized systems are mentioned including intelligent tutoring, time-reaction monitoring, and simulations. For situative approaches they mention possibilities drawing on ethnographic analysis (p. 101) to study social practices *in vivo*. However, this account is given only a single paragraph. There is still work to be done, therefore, to find models of assessment, and develop practice, in ways that are more congruent with the situative or sociocultural theories of learning that are exciting a great deal of interest among educators and researchers in the fields of curriculum and pedagogy across domains, sectors, and continents. Some of the articles in this section record steps taken in this direction.

The roots of most situative or sociocultural approaches to the understanding of learning can be traced to the Soviet psychologist, Lev Vygotsky, although some commentators connect his thinking with earlier transactional psychologists in the US: William James, G.H. Mead, and John Dewey. All of these theorists were interested in the activity of learning and the development of mind but believed that there is dynamic interaction, or transaction, between the individual agent and the structured world, which is mediated by tools and systems of signs and symbols, especially language – the tool of tools. In American sociology this line of enquiry developed into symbolic interactionism; in Russia, Vygotsky initiated activity theory as a psychologically relevant application of dialectical and historical materialism. These theoretical developments blur the boundaries between what have subsequently become separated into distinct disciplines of sociology and psychology. Psychology, cognitive science, and psychometrics have tended to dominate the discourse surrounding learning outcomes and their assessment, although the suggestion by Pellegrino *et al.* that ethnographic methods in research might offer some solutions to assessing learning within a situative or sociocultural perspective encourages a reexamination of these earlier influential approaches and their more recent manifestations.

Vygotsky studied the development of the higher mental faculties of the individual mind, but, in keeping with his identity as a Soviet psychologist, he was also interested in the way that collective consciousness was brought into the individual mind. Central to his theory was the assumption that social interaction is central, that is, much more than a

context for learning. In a much-quoted passage he describes how an interpersonal process is transformed into an intrapersonal process:

Every function in the child's cultural development appears twice: first, on the social level, and later, on the individual level; first *between* people (*interpsychological*), and then *inside* the child (*intrapsychological*). This applies equally to voluntary attention, to logical memory, and to the formation of concepts. All the higher functions originate as actual relations between human individuals. (author's emphases) (Vygotsky, 1978: 57)

The transfer and translation of cultural knowledge involves both externalization and internalization through shared activity (interaction) and individual learning activity (action). For example, in a life-drawing class the object of an activity (the problem space) may be to make a convincing drawing of a model lying horizontal on a couch with her arms and hands stretched above her head toward the novice artist. Knowing, that the hand is physically about the size of a face, the instinct of the novice is to draw the hand too small: as it would be if the model stands upright. The teacher, or more expert other, can help the novice by introducing a tool, known as the artist's measure. This is simply a pencil that the artist holds at arm's length, with one eye closed, to measure the relationships between, say, the width of the head and the width of the hand. This enables the novice to see what the eye actually sees and not what the mind interprets to be the case based on previously constructed mental models about the proportions of the body. Eventually this technique is internalized and the novice can look more carefully at proportions without using the tool. This achieves the goal, or outcome, of the activity, that is, to understand the concept of foreshortening and apply it to practice. The novice artist therefore learns to look in a different way. In Vygotsky's terms, the object (the learning activity that the subject is engaged with) works back to the subject (the learner) to create new forms of subjectivity.

This theory also predicts creativity or knowledge creation in learning because, once the individual has internalized the use of tools (artifacts, practices, and language), he or she can adapt them or create new ones to tackle new problems with new results. These can be shared with others, thus creating a cycle of externalization, internalization, and externalization. This provides a response to problems with both of Sfard's (1998) two metaphors of learning: the acquisition metaphor (AM) and the participation metaphor (PM). The pervasive acquisition metaphor implies that units of knowledge are possessed, accumulated, refined, combined, and passed on, whilst the linguistic turn of the participation metaphor shifts the emphasis from having to doing. Learning, according to PM, is embedded in the social processes of becoming a member of a community of practice. Apprenticeship is a

familiar analogy. However, these two metaphors provide only partial explanations of innovation and creativity as important learning outcomes. In contrast, Vygotsky's theory of goal-oriented, tool-mediated learning activity can encompass learning outcomes associated with both of these metaphors, but also with outcomes associated with knowledge creation, because it provides a description of how knowledge and practices can be transformed. In other words, it can encompass a very wide range of outcomes: higher and lower mental processes; attitudinal, cognitive, and behavioral outcomes; individual and shared activity; problem-solving processes and products; the acquisition of existing knowledge and the creation of new knowledge. Indeed it accommodates most, if not all, of the learning outcomes discussed in the articles in this section of the encyclopedia.

Vygotsky also has something to offer on the issue of progression. The structures of grades, scales, and attainment levels have accustomed us to regard progression as step by step and linear. According to Vygotsky, the mastery of tools of the mind takes place in the zone of proximal development (ZPD). As Grigorenko (1998) points out:

The word *zone* refers to the nonlinearity of children's development. In this zone, the child might move forward or backward, to the left or to the right. . . . The main characteristic of the ZPD is its sensitivity to the child's individuality – its responsiveness to the unique profile of each child's skills. The ZPD is of tremendous importance educationally. Generally, when educators evaluate a child's skills, they focus on what the child demonstrates in his or her independent performance. . . . Vygotsky stated that the level of independent performance is important, but not the only, index of development. To account for the dynamic process of development, we should consider the level of the child's assisted performance. (pp. 210–211)

There are two points to be made here. First, there is value in plotting a learners' development as a profile within a zone and encouraging them to expand, deepen, and enrich their knowledge, skills, and understanding. Second, there is value in assessing how learners respond to assistance and the introduction of new tools by others. Grigorenko argues that by doing this we are more likely to determine students' true level, than by administering tests of unassisted performance.

Very often, teachers wait for children to demonstrate completely formed functions; teachers think that they can lead a child to the next step only after they have seen evidence that the child has successfully acquired the function taught at the previous step. As a result, children are limited in their learning opportunities to those that correspond to their level of independent performance. In other words, teachers who follow this standard practice

minimize the student's ZPD by almost closing it. On the contrary, ideally, the ZPD should be wide open, so that it can expand developmentally appropriate practices up to the level of assisted performance. (Grigorenko, 1998: 212)

Vygotsky's view that what learners can do at lower levels should not limit opportunities for their development of higher psychological functions was almost certainly influenced by his experience of working at a boarding school for children who were both deaf and blind. He claimed that education targeted on the formation of higher mental processes could help people overcome, or compensate for, deficits because it equipped them with adaptive strategies that might be unique for each person.

All of these ideas have implications for assessment. Practical applications of Vygotsky's theories have been developed in versions of dynamic assessment. Some forms of dynamic assessment have been described as little more than superior intelligence tests, but there are clinical versions (see the article 'Dynamic assessment (two types)') that use hint structures in scaffolded instruction to maximize feelings of competence and efficacy. This affects test reliability, in a psychometric sense, although advocates claim that these concerns are less important than the quality of the insights that result and the likely magnitude of change in a learner's performance (i.e., their validity). Dynamic assessments of this nature are currently used in one-to-one situations, usually with students with learning difficulties; therefore they are often felt to be impractical for wider application. However, the principles could well be adapted for everyday use by teachers if they were to be integrated with instructional strategies. Moreover they could serve both formative and summative assessment purposes although the formative would take priority. In this sense, versions of dynamic assessment could provide tools for formative assessment integrated into pedagogy.

As noted earlier, much current assessment practice derives from behaviorist or differentialist approaches, many of which are now rejected as learning theories except in very limited senses. Pellegrino *et al.*'s (2001) important work provides both the underpinning theory and practical examples of more valuable forms of assessment underpinned by sound cognitive science. James (2008: 31) offers the following pointers to what might constitute 'third generation' (sociocultural) assessment practices:

- If learning cannot be separated from the actions in which it is embodied, then assessment too must be situated.
- Assessment alongside learning implies that it needs to be done by the community rather than by external assessors.
- Assessment of group learning is as important as the learning of the individual.
- *In vivo* studies of complex problem solving may be the most appropriate form for assessments to take. (Some

ethnographic methods could be used, including methods for assuring quality of inferences from evidence.)

- The focus should be on how well people exercise agency in their use of the resources or tools (intellectual, human, material) to formulate problems, work productively, and evaluate their efforts. This would be a proper justification for coursework assignments with students having access to source materials, because it is the way that these are used that is of most significance.
- Learning outcomes can be captured and reported through various forms of recording, including narrative accounts and audio–visual media. The portfolio has an important role here.
- Evaluation needs to be more holistic and qualitative, not atomized and quantified as in measurement approaches.

The main obstacle to implementing this sociocultural approach to assessment is likely to arise from a perceived need to ensure the trustworthiness of such assessments when large numbers of students are involved, and when those who are interested in the outcomes of such learning cannot participate in the activities that generate them. Within vocational education, which often aspires to some aspects of this model, systems of internal and external assessors and verifiers have attempted to assure quality, although such large-scale systems almost inevitably become bureaucratic, unwieldy, and restrictive. Clearly, more work needs to be done to develop approaches to assessment that are more coherent with valued sociocultural perspectives on learning.

Support for the Development of Educational Assessment Practice

Another challenge to the effective implementation of educational assessment relates to the demands it makes on the knowledge and skills of educators. Traditional forms of testing and examinations are usually the product of extensive piloting and trialing of instruments by experts in educational measurement who work in test agencies or awarding bodies. Everyday classroom assessment, however, is the responsibility of teachers and those who manage them. Rarely are educators adequately equipped for this task in their initial training, yet they need to develop sufficient assessment literacy to allow them to examine, and make critical sense of, all forms of evidence of students' learning and performance. Development of an understanding of the use and misuse of statistical data generated by large-scale assessment systems is particularly urgent, but the knowledge and skills of effective classroom assessment practices for formative purposes is equally challenging. This constitutes a specific case of educational transformation to which more generalized insights from research on educational change, professional development, and organizational learning may apply (see the overview in the

section 'Leadership and management'). However, there is also a need to investigate the specific conditions that enable educators to implement effective assessment practice because it is usually a failure of implementation, rather than a failure of knowledge, that stands in the way of success (see the articles '(School) Organizational policies and practices to support effective classroom assessment' and 'Challenges of developing and implementing formative assessment practices in schools').

Although the studies that Black and Wiliam (1998) reviewed provided information about the kinds of interventions that were most likely to be profitable, these were tested in small-scale, intensively supported experiments. If these innovations are to be scaled-up and sustained across a system, they need to be able to grow with much less support. The right conditions for the creation and spread of knowledge and practice are crucial to their successful implementation. These conditions include the professional development of teachers and organizational structures and cultural processes that support and enhance educators' own learning. The development study by Black *et al.* (2003) investigated some of these conditions at individual teacher level. Subsequently, a larger study (James *et al.*, 2007) investigated conditions for the creation and spread of knowledge and practices within and across classrooms, schools, and professional networks. On the basis of their research, the latter identified a number of key conditions for successfully embedding and sustaining effective assessment for learning:

1. The ultimate goal of educational assessment is to promote learning as well as to record achievement. The AfL practices are a useful starting point. However, these practices should serve underlying principles, such as making learning aims, processes, and outcomes explicit to learners and promoting learning autonomy through encouraging mindfulness and strategic approaches.
2. Although teachers appreciate practical advice, AfL practices can become ritualized and mechanistic if teachers are not stimulated to think about the principles of learning that underpin them. The project team made a distinction between those teachers who implemented the letter of AfL by injecting AfL practices into what they usually do without changing anything more fundamentally, and those who captured the spirit of AfL by integrating practices into the flow of lessons to regulate the learning process itself. The latter requires some understanding of underlying principles. Beliefs and practices are interrelated and need to be developed together; it is not sufficient simply to tell teachers what to do.
3. Those teachers (only about 20% of the sample at the beginning of the study) who had most success with implementing AfL in their classrooms were those who demonstrated a capacity for strategic

and reflective thinking and took responsibility for what happened in their classrooms. They were not inclined to blame external circumstances (such as initiative overload) or pupil characteristics (such as innate ability or attitudes) but concentrated on the ways in which they could improve the learning experience for pupils.

4. Although most teachers held clear and positive educational values, the majority of teachers struggled to bring practice in line with their stated values. These values–practice gaps closed somewhat over the course of the project but many teachers felt constrained by a policy context that encouraged rushed curriculum coverage, teaching to the test and a tick-box culture.
5. Values, beliefs, and practices were not uniform over all groups of teachers. In secondary schools, there were differences in the beliefs and practices of teachers with different subject specialisms. The highest percentage of teachers with classroom practices consistent with AfL were teachers of English. In contrast, mathematics teachers seemed to struggle with these ideas. There were also differences between schools. These differences indicate the level of challenge for leadership and support, and particularly for development of differentiated strategies for professional development.
6. Classroom-based collaborative inquiry practices for teacher learning emerged as the key influence on teachers' capacity to promote learning autonomy with their pupils. These include learning from research and also working together to plan, try out, and evaluate new ideas.
7. Such knowledge creation among teachers can extend beyond the classroom through networking across the school and with teachers in other schools. Networking through face-to-face meetings of various kinds builds the social capital (mutual support and trust) that supports the exchange of intellectual capital (ideas and practices).
8. Opportunities for teachers to learn in these ways, through classroom inquiry and networking, depend significantly on organizational structures, cultures, and leadership. Particularly important is a school's knowledge of the expertise in its midst, or available to it, and its capacity to tap into this expertise, grow it, and spread it through professional development activities and networking. The quality of leadership at every level is crucial to making this a reality.
9. The key challenge for leadership is therefore to create the space and climate for school staff to reflect on and share aspects of their practice. This includes encouraging and stimulating dialog and risk taking. In this way, innovations in assessment for learning can be tested, embedded, and sustained. Without it, they remain surface changes that decay and disappear when the next initiative comes along. (adapted from James *et al.*, 2007: 214–217).

Conclusion

This article has traversed a wide territory, starting with reference to testing practice, ending with issues of educational transformation and change, and dealing with assessment in different contexts for different purposes along the way. What it sought to do, in addition to introducing the other articles in this section, was to encourage a thoughtful consideration of the primacy of purpose and to unpack some of the choices that flow from decisions about means and ends. These not only involve choices about techniques but about implementation and use, about relationships between constituent parts of a system, and about the training and support that will enable the goals of educational assessment to be achieved. These are complex, multidimensional issues that require a wide range of skills, knowledge, and understanding: technical, professional, ethical, and political. It may be too much to expect that any one educator will possess all of these but, in that educators work in professional communities, we might hope that collective professional knowledge and expertise will expand in the future in order to promote the educational goals of assessment.

See also: Alternative Assessment; An Overview of Educational Assessment; An Overview of Language and Literacy in Educational Settings; Assessing Group Work; Assessment and the Evaluation of Institutional Effectiveness; Assessment and the Regulation of Learning; Assessment Conversations: Reading and Writing Conferences with Students and with Parents; Assessment in Higher Education; Assessment in Schools – Affective Domain; Assessment in Schools – Creative Subjects; Assessment in Schools – Dispositions; Assessment in Schools – Learning to Learn; Assessment in Schools – Literacy Writing (Extended); Assessment in Schools – Mathematics; Assessment in Schools – Oracy; Assessment in Schools – Primary Science; Assessment in Schools – Secondary Science; Assessment in Schools – Technology Education and ICT; Assessment in Schools Related To Literacy: Reading; Assessment in the Early Years; Assessment in the History, Civics and Social Studies Domains; Assessment in the Workplace of Performance, Developing Expertise and Competence; Assessment in Vocational Education; Assessment Practice in Policy Context: Latin American Countries; Assessment: Pre-service and In-service Teacher Education; Challenges of Developing and Implementing Formative Assessment Practices in Schools; Classroom Assessment in Policy Context (Australia); Classroom Assessment in Policy Context (England and France); Classroom Assessment in Policy Context (French Sources); Classroom Assessment in Policy Context (Hong Kong); Classroom Assessment in Policy Context (New Zealand); Classroom Assessment in Policy Context (Sub-Saharan Africa); Classroom Assessment Tasks and Tests; Dynamic Assessment; Formative Assessment and Instructional

Planning; Formative Assessment; Impact of Assessment on Learner Groups (Boys/Girls); Impact of Assessment on Learner Groups (Disabilities); Impact of Assessment on Students' Learning Strategies and Implications for Judging Assessment Quality; Impact of Assessment on Students' Test Anxiety; Impact of Assessments on Classroom Practice; Instructional System Provided Feedback; Leadership and Management Overview; Moderation of Student Work by Teachers; Peer and Self-assessment; Portfolio Assessment; Progression and Assessment; Developmental Assessment; Records of Achievement: Beyond Traditional Tests; School Policies and Practices to Support Effective Classroom Assessment for Learning; Social Practices in School Assessment and their Impact on Learner Identities; Student Assessment: Policy and Practice in Eastern Europe; Students' Rights in Assessment and Evaluation; Summative Assessment by Teachers; The Multiple Purposes of Assessment; The Relationship Between Assessment and the Organization and Practice of Teaching.

Bibliography

- ARG (1999). *Assessment for Learning: Beyond the Black Box*. Cambridge: University of Cambridge School of Education (Assessment Reform Group).
- ARG (2002a). *Assessment for Learning: 10 Principles*. Cambridge: University of Cambridge Faculty of Education (Assessment Reform Group).
- ARG (2002b). *Testing, Motivation and Learning*. Cambridge: University of Cambridge Faculty of Education (Assessment Reform Group).
- Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2003). *Assessment for Learning: Putting It into Practice*. Maidenhead: Open University Books.
- Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice* 5(1), 7–75.
- Crooks, T. (1988). The impact of classroom assessment practices on students. *Review of Educational Research* 58, 438–481.
- Dweck, C. S. (2000). *Self-Theories: Their Role in Motivation, Personality and Development*. Philadelphia, PA: Psychology Press.
- Greeno, J. G., Pearson, P. D., and Schoenfeld, A. H. (1996). Implications for NAEP of research on learning and cognition. *Report of a Study Commissioned by the National Academy of Education. Panel on the NAEP Trial State Assessment, Conducted by the Institute of Research on Learning*. Stanford, CA: National Academy of Education.
- Grigorenko, E. (1998). Mastering tools of the mind in school (trying out Vygotsky's ideas in classrooms). In Sternberg, R. and Williams, W. (eds.) *Intelligence, Instruction and Assessment: Theory into Practice*, pp 201–244. Mahwah, NJ: Erlbaum.
- James, M. (2008). Assessment and learning. In Swaffield, S. (ed.) *Unlocking Assessment: Understanding for Reflection and Application*, pp 20–35. Abingdon: Routledge.
- James, M., McCormick, R., Black, P., et al. (2007). *Improving Learning How to Learn – Classrooms, Schools and Networks*. London: Routledge.
- Pellegrino, J., Chudowsky, N., and Glaser, R. (eds.) (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.
- Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher* 27, 4–13.
- Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Watts, A. (ed.) (2008). *1858 Question Papers: 150 Years of Examinations from Cambridge Assessment*. Cambridge: Cambridge Assessment.

Relevant Websites

- <http://www.assessment-reform-group.org> – Assessment Reform Group, UK.
- <http://www.learn2learn.ac.uk> – Learning How to Learn: A Project of the ESRC Teaching and Learning Research Programme, (See also www.tlrp.org).

Assessment and the Evaluation of Institutional Effectiveness

S M Thomas, University of Bristol, Bristol, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

DCSF – UK's Department for Children, Schools and Families.

FSM – Eligible for free schools meals.

GCSE – General Certificate of Secondary Education.

KS4 – National Curriculum Key Stage Four (England).

NCLB – No Child Left Behind Act of 2001 (USA), Pub. L. No. 107-110, 115 Stat. 1425 (2002).

OECD – Organization for Economic Cooperation and Development.

PISA – OECD Programme for International Student Assessment.

VA – Value-added measure.

Introduction

In the last two decades, there has been an increasingly worldwide trend to reform national assessment and examination systems with the primary aim of enhancing educational quality. The rationale underlying these reforms is often based on the premise that increasing monitoring, evaluation, and accountability systems will be a positive lever for improvements in student and school performance. For example, Hamilton *et al.* (2002) have reported the development and use of standardized tests for school accountability purposes in USA, particularly in response to the No Child Left Behind (NCLB) legislation. Similarly Kellaghan and Greaney (2004) and Yu and Thomas (2008) in the African context and Peng *et al.* (2006) in the Chinese context have examined developments in assessing student learning and how national examinations are used for institutional accountability. Moreover, the most recent edition of OECD (OECD, Organization for Economic Cooperation and Development) *Education at a Glance* (2007) provides up-to-date evidence of the number of countries that now have national examination systems and which countries provide feedback data to policy-makers, teachers, and parents. In 2005, just over half of the OECD countries (16) had national high-stakes examinations that are completed by lower secondary school students; of these almost all (14) require schools to conduct regular self-evaluation and three-quarters (12) make information on school evaluation available to the local

school community or general public for one or more, sometimes overlapping, purposes including parental school choice, accountability, funding decisions, and school improvement. At the international level, the demand for comparative and monitoring information to inform policy development and evaluation has also led to the increased use of international surveys of educational achievement, such as Programme for International Student Assessment (PISA) introduced for the first time in 2000 (Scheerens, 2004; Goldstein and Thomas, 2008).

Given this apparent global shift toward a stronger emphasis on student testing and assessment, it is more important than ever to highlight and continue to evaluate the impact of these policy developments on educational quality and standards. For example, previous research has explored some of the negative consequences of high-stakes assessment at the level of individual students (Harlen and Deakin Crick, 2002) and similarly at the school level, in relation to accountability, evaluation, and publication of examination results (Volante, 2007; Van Petegem *et al.*, 2005). Further analysis and critique of assessment policy at the student level is vital but nevertheless beyond the scope of this article. However, with regard to either positive or negative consequences for school quality, it is of course particularly crucial to examine in detail how the data are actually used to measure and evaluate institutional effectiveness. Therefore, this article explores the different ways that national assessments are used to measure institutional effectiveness as well as the validity of different methods for different purposes. The article focuses on policy developments and research that has been conducted at the level of statutory schooling and examples will be provided from a large-scale study conducted in England (see Thomas *et al.*, 2007; there has also been equivalent research at other educational phases, for pre-primary see Sammons *et al.*, 2004, and for higher education see MacLeod, 2007.)

The Problem with Raw League Tables

The issue of appropriate and valid ways of reporting institutional performance is of vital importance whatever the purpose, be it public accountability, funding, parental choice, or school improvement. In each case, it is important that the information is fit for purpose, accurate, and the context and limitations of any measures are emphasized, particularly when the data may potentially contribute to

life-changing conclusions or decisions for students or teachers or both. Unfortunately, in terms of both policy and practice, this is not always what happens. For example, in many countries student assessment data are commonly used to evaluate institutional performance through publication of schools' raw examination scores taking no account of the context of the school or the characteristics of the student intake (OECD, 2007). However, schools with high-achieving intakes will tend to do well for that reason alone. Neither the initially high-achieving nor the initially low-achieving school is assisted by accountability in the form of raw league tables. In the former, the need for improvement may not be understood; in the latter, serious demoralization of staff may occur through no fault of their own. It has been strongly argued by many educational practitioners as well as by academic researchers that, taken on its own, information about schools' raw examination results will always be a very inadequate measure of performance and, without any knowledge of the context of that school, can be misleading. For example, Nuttall (1990) argued that "... natural justice demands that schools are held accountable only for those things that they can influence (for good or ill) and not for the pre-existing differences between their intakes" (p. 25).

However, in contrast to raw examination results, what are commonly called value-added measures provide the starting point for a way of evaluating school performance that takes account of intake factors outside the control of the school but which have a considerable impact on pupil attainment.

The Value-Added Approach

The value-added approach stems from the assumption that schools add value to the achievement of their students. It is based on the idea of measuring student progress, usually in cognitive outcomes, such as reading or mathematics attainment, during a given period of time. However, the concept may also be applied to other quantifiable student outcomes, such as measures of vocational competence or citizenship. In order to measure progress baseline and outcome, measures are required at the beginning and end of a particular time period (e.g., covering all or part of the primary or secondary phases of education). Of course, as students grow older, progress or improvement would be expected and average attainment levels to rise. Value-added measures thus seek to establish whether students in some schools make relatively greater or lesser progress than those in other schools over a specified period of time. The most effective schools would be those in which student progress exceeds expectations. For the purpose of internal school self-evaluation, value-added measures provide teachers with more meaningful, valid, and accurate evidence of the relative progress of

their own students, in comparison to raw results, and this information can be used to reflect on their professional practice. Similarly, in the high-stakes context of external school evaluation and accountability, value-added measures provide a more valid source of comparative information about a school's effectiveness, in comparison to raw results. Nevertheless, it is important to remember that the value of schools' educational quality is broader than what can be measured by attainment or progress in a few specific areas of student activity. A comprehensive value-added evaluation framework might also encompass measures related to numerous other aspects of a school's mission, processes, and outcomes.

The Assessment of Institutional Effectiveness

In the last 20 years or so, the significant growth of large national and international educational assessment data sets and associated research findings on school effectiveness as well as significant advances in methodological techniques have stimulated international debate and government thinking on educational policy and practice. As a consequence, national (and regional) policymakers in various European countries and worldwide have focused their attention on the possibilities for improving educational practice and, indirectly, educational standards, competitiveness, and performance, by encouraging more systematic and sophisticated approaches to student assessment and institutional evaluation and self-evaluation (Reynolds, 2007). For example, the UK governments have emphasized the need for schools and teachers to use evidence and data to inform their own internal evaluations of the education they provide. This approach involves an ongoing and systematic self-evaluation of a school's educational practice and improvement processes using information drawn from a variety of sources including, for example, the detailed pupil assessment and tracking information provided confidentially through RAISEonline website by the Department of Children, Schools and Families (DCSF). At the national level in England, the system of external school inspections introduced in 1992 has been developed extensively and the new 2005 framework requires school self-evaluation to be the central element of the inspection process (OFSTED, 2004). Moreover, as noted above in relation to over half of OECD countries, school examination performance tables continue to be published by the DCSF as a mechanism for educational accountability. However, since 2004, the English league tables now also include value-added school performance as well as raw examination performance for schools.

The use of value-added measures provides an important innovation in the assessment of institutional effectiveness – in terms of both external evaluation and

internal self-evaluation. Nevertheless, the methodology of measuring value added is not perfect and of course the typical limitations of measurement still apply and need to be understood by practitioners and policymakers, an issue that will be returned to subsequently.

Data Collection Issues

Calculating the effect a school, or any educational institution, has on student progress is complex. In part, this is because the educational experiences of any individual student and the wide variety of factors influencing her or his progress can be viewed as unique and almost impossible to quantify. However, the more information it is possible to have about all individual students, as well as comparative data across a whole population (or representative sample) of schools, the more valid, reliable, and informative any subsequent analysis is likely to be. The key data required to measure a student's academic progress are baseline and outcome attainment over a specific period of time. Other background and contextual information about individual students and schools is also needed to provide statistical controls for those factors – outside the control of the school – that have a significant impact on a student's attainment, relative progress, or both (Thomas, 2001).

The different types of quantitative data required to create school performance measures are described below.

Student Outcomes

Ideally, student outcomes need to be measured using valid and reliable assessment instruments relevant to the academic curriculum taught in schools or other aspects of the curriculum (such as vocational training and qualifications). This is because it is important to measure only those aspects of a student's education that the school has a clear aim and statutory role to provide and develop. It is also necessary that the assessment methodology employed to measure student outcomes is unbiased and sufficiently reliable to ensure that students at the same level of attainment will be assessed – as far as possible – at the same level irrespective of where the assessment takes place. For example, it may be argued that teacher-assessed outcomes (such as portfolio assessment) are open to unequal standards being applied across schools and are therefore less appropriate for value-added techniques than standardized and externally marked tests such as those from national examination systems and standardized attainment tests.

Student Baselines

In order to measure a student's relative progress overall, or in any particular curriculum area, it is necessary to

obtain information about their previous – or baseline – attainments, either at entry to a phase of education or some alternative and predefined starting point. Ideally, as a minimum, data are required concerning students' baseline attainment in the core curriculum subjects of language or numeracy or both, and if possible other relevant aspects of the curriculum. Again, as noted in relation to outcome assessments, it is necessary that the baseline assessments are valid, reliable, and fit for purpose. Moreover, baseline and outcome measures need to be collected and recorded in such a way that individual student records can be matched accurately over time. The use of unique student, class, and school identification codes are vital in order to facilitate this matching process. Once a longitudinal data set of individual student attainment records has been created, it is then possible to proceed with a statistical analysis to examine both absolute attainment at one point in time and – most importantly – the relative progress of students in a particular school over two or more time points in comparison to students in the wider population (or a representative sample) of schools.

Background and Contextual Information

Information about student baseline and outcome attainment is absolutely essential to provide an accurate and direct measure of students' relative progress in attainment over time and this approach is the only method of calculating valid value-added measures of school effectiveness. However, the collection of other student and school data is also needed in order to examine the impact of various background and contextual factors that are outside the control of the school and also appear to influence the rate of student progress. Previous research has shown that student background characteristics such as socioeconomic status, gender, ethnicity, first language, and mobility are often statistically significantly related to student attainment and progress and therefore are able to provide a means of fine-tuning value-added measures of school effectiveness. Similarly, some previous studies have shown that school context data, such as the percentage of disadvantaged students in a school, have a statistically significant impact on student progress (Thomas, 2001). In reviewing the evidence of institutional effectiveness, it is therefore important to consider whether a straightforward progress measure or, alternatively, a more sensitive value-added measure that also controls for other factors that are outside the control of the school is used (e.g., in England this latter approach is referred to by DCSF as contextualized value-added measures).

Of course, in the absence of progress data, the impact of socioeconomic and other background factors on student attainment is considerable at any one point in time (e.g., at either baseline or outcome or both). This finding is well documented and not surprising given the cumulative

effects of such factors on children's educational experiences since birth. Therefore, student and school background data may also be usefully employed to provide contextualized attainment measures at any one time point as an additional, or an alternative (but conceptually different and arguably less accurate and valid), approach to measuring school performance.

Educational and School Process Information

Educational and school process information is not usually employed in the calculation of value-added measures of school performance. This is because in contrast to the background and contextual data described above, measures of educational and school processes aim to quantify aspects of school life that are defined to be within the control of the school. Measures of this kind include, for example, teacher attitudes, leadership, experience and supply, class size, organization and grouping strategies, school ethos, and organization. However, the rationale of collecting additional information about educational and school processes is to enable school and classroom process variables to be contrasted and evaluated against measures of students' relative progress (see Sammons *et al.*, (1997) for examples and Willms and Raudenbush (1989), for a discussion of this issue). In this case, the aim is to provide evidence to explore and illuminate the reasons underlying any differences in school effectiveness.

Of course, whatever type of quantitative data is collected, it is crucial that quality-assurance procedures and accuracy checks are put in place as an important part of the data-collection exercise. Data errors or missing data, or both, will mean that the results of any analysis will be difficult, if not impossible, to interpret. In order to implement a value-added system of school evaluation and/or self-evaluation, it is important that schools are willing to collaborate with other schools at the local, regional, and national level in order to provide comparative data in an identical format. This approach usually involves the creation of national or regional examination databases and the central organization of collecting, analyzing, and presenting feedback results to schools in a common format for the purpose of supporting and stimulating school evaluation and improvement activities. As an exemplar, at the regional level in England, Lancashire local education authority value-added project illustrates how the ongoing collection of examination and other data from over 130 secondary schools has been organized successfully over a 14-year period (Thomas *et al.*, 2007). Moreover, the English government now provides data to all schools nationally through extensive online access to a wide variety of student and school assessment and evaluation data via the RAISEonline website.

How is Institutional Effectiveness Measured?

Taking into Account of Different Types of Data

Table 1 provides a summary of the data requirements for different value-added and raw school performance measures. The requirements shown relate in each case to a single cohort of students and must include data from a representative sample of schools, if not all schools, in the region or system. Of course, in order to examine trends over time or improvement in value-added performance, it is necessary to collect equivalent outcome and baseline data for at least three consecutive student cohorts (for an example of a value-added analysis across ten cohorts, see Thomas *et al.*, 2007). Similarly, data for consecutive student cohorts would also be necessary to examine trends in contextualized or raw performance measures.

Using Different Methods of Statistical Analysis

A key challenge for researchers has been to develop approaches that allow the statistical analysis to separate out the effect of the school experience on individual student outcomes (what students achieve) and the extent to which student intake characteristics (those things the students arrive at school with) affects student outcomes. A variety of statistical techniques can be employed for this purpose that vary in the sensitivity and sophistication of analysis. There are three main approaches: summary statistics, multiple regression, and multilevel modeling.

The first approach, summary statistics, including measures such as the mean and standard deviation, can be used to calculate aggregate school-level measures from student-level data. These measures provide a crude picture of

Table 1 Data requirements for value added and raw school performance measures

	<i>Variables required/controlled for in analysis</i>
<i>Value-added measures</i>	
Progress only value-added measure	Outcome and baseline assessment measures at two (or more) time points
Contextualized value-added measure	Outcome and baseline assessment measures at two (or more) time points, student and school background and context measures
<i>Raw performance measures</i>	
Contextualized raw attainment measure	Outcome assessment measures at any one time point, student and school background and context measures
Raw attainment measure	Assessment measures at any one time point

school performance through estimates of raw levels of student attainment for each school in a sample (i.e., unadjusted for any other factors) and may be used, for example, to produce simple league tables of average school performance at one point in time. As mentioned previously raw measures are generally considered invalid for the purpose of evaluating institutional effectiveness, given student progress in attainment cannot be estimated. A further disadvantage of this method is that the school is the unit of analysis and therefore detailed information about individual students is lost in the analysis.

The second approach, multiple regression analysis, is the standard statistical technique for calculating the residual difference between a student's observed and expected assessment score. The observed score is a student's actual level of attainment at the end of an educational phase (such as primary or secondary) and the expected score is the level that would be predicted on the basis of his or her previous – baseline – attainment. Consequently, the residual score is interpreted in terms of whether a student is performing above or below expectation on the basis of the overall statistical relationship between baseline and outcome attainment of all students in all schools in the sample. In essence, the residual score – which ranges from a positive to negative value – provides the key statistical measure of student's relative progress, that is, the value added. One advantage of this approach is that several factors, such as baseline attainment and other student characteristics like gender and socioeconomic status, can be employed in the analysis to provide a more sensitive estimate of value added than would result from employing a single baseline predictor. However, the disadvantage of this approach is that the unit of analysis has to be either at the level of the student (i.e., student residual scores are calculated) or the

school (i.e., school residual scores are calculated). In the former case, important information about the clustering of students within a particular school is lost, and in the latter case, detailed information about individual students is lost. **Figure 1** demonstrates this issue using a simple example of aggregated school level data from over 80 English secondary schools to illustrate how value added is estimated. Note that some research has explored alternative regression approaches, such regression-discontinuity analysis that compares the raw outcomes of consecutive school cohorts to estimate institutional effects. However, these methods are limited by a lack of longitudinal data and individual student records matched over time. Each point represents one school and the regression line represents the expected 2001 General Certificate of Secondary Education (GCSE) mean score, on the basis of the mean Cognitive Abilities Test (CAT) score of the same students in 1997. Schools with points above the regression line will have a positive value-added score (i.e., residual), suggesting that performance is above expectation. Schools below the regression line will have a negative value-added score, suggesting performance below expectation. Of course, information is also needed about the level of confidence that is associated with any value-added score (e.g., 95% confidence interval) to indicate whether an individual school's performance is likely to have occurred by chance or, alternatively, is statistically significantly different from what would be expected.

The third approach, multilevel modeling, is a relatively recent generalization of multiple regression that involves the same principle of calculating a residual value-added score. However, this technique takes account of the clustering of students within schools and allows the unit of analysis to include both the student and the school

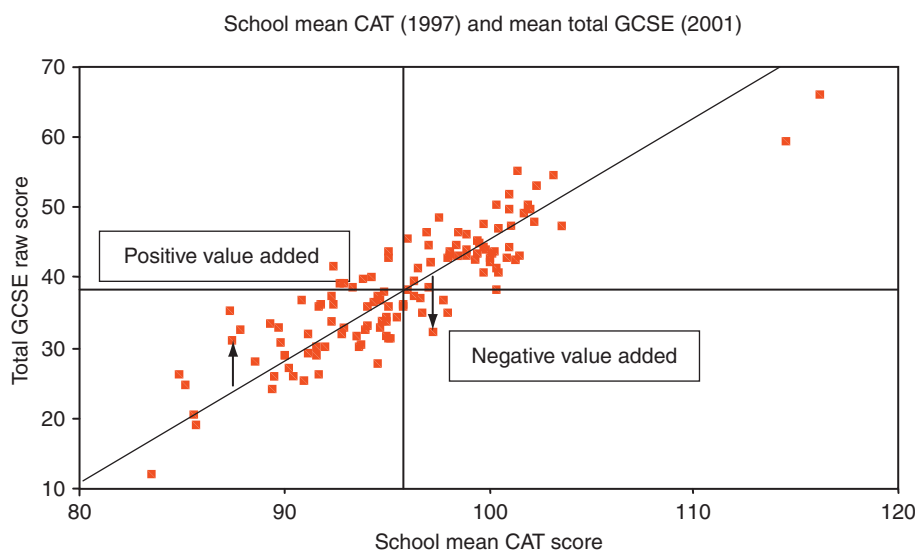


Figure 1 Estimating value added using aggregated data – schools total GCSE (2001) and mean CAT (1997) performance.

level. Other clustering factors, such as teacher, class, or region, can also be included in the analysis as appropriate. Thus, multilevel modeling is a far more sophisticated approach than both summary statistics and standard multiple regression when the aim is to disentangle the complexity of schools' effectiveness. It is now widely recognized as the most flexible tool for examining the hierarchical nature of student attainment data. This approach can be used to calculate unbiased and accurate estimates of school residuals for all students (or particular groups of students such as boys or girls) as well as, crucially, the statistical significance of an individual school's results. If data are available for consecutive student cohorts, this technique can also be employed to model trends in value-added results over time. New developments in the methodology of calculating value-added measures are ongoing and, for example, have examined the continuity of previous school effects on later attainment and the need explicitly to take into account pupil mobility (Goldstein, 1997).

Validity of Institutional Effectiveness Measures

Previous research has indicated that in order to provide a realistic picture of a school's performance a range of different assessment outcomes and value-added measures is required. This approach is necessary to reveal the internal variations in school effectiveness across one or more dimensions – such as in different aspects of the curriculum, for different student groups, and different time periods or cohorts (Thomas, 2001; Scheerens *et al.*, 2003).

Research has shown that schools doing well with students in one aspect of the curriculum are not necessarily effective in all aspects. This evidence strongly suggests the

need to examine school effectiveness measures across a range of academic outcomes in order to reveal the pattern of departmental or subject-area performance. Clearly, using a single-measure effectiveness may conceal important within-school differences not only across academic aspects of the curriculum but also across other aspects such as vocational or attitudinal outcomes. Researchers have also examined the issue of differential school and departmental effects for different groups of students (such as high and low attainers, boys and girls, or different ethnic groups) and found that not all schools are equally effective for all student groups. Thus, it has been established that using an overall measure of school (or departmental) performance may be misleading or even invalid by masking important differences in the relative progress made by different student groups, particularly those categorized by prior attainment (Thomas, 2001). The issue of teacher, or classroom, effects also impacts on the validity of institutional effectiveness measures. Previous research in UK and USA has identified significant variations in student progress attributable to differences between teachers (e.g., McCaffrey *et al.*, 2004). Therefore again, overall performance measures may conceal important within-school differences in teacher effectiveness that needs to be considered when judging institutional effectiveness.

Recent evidence in England also shows that generally for most schools the linear trend in value-added performance is fairly stable over time, although this conceals both considerable year-to-year variation in performance and an underlying time trend across all schools of increasing raw GCSE performance at KS4 (Thomas *et al.*, 2007). For an example of nonlinear value-added time trends over 14 cohorts from the Lancashire project see **Figure 2**. Moreover, for some schools, results can vary substantially

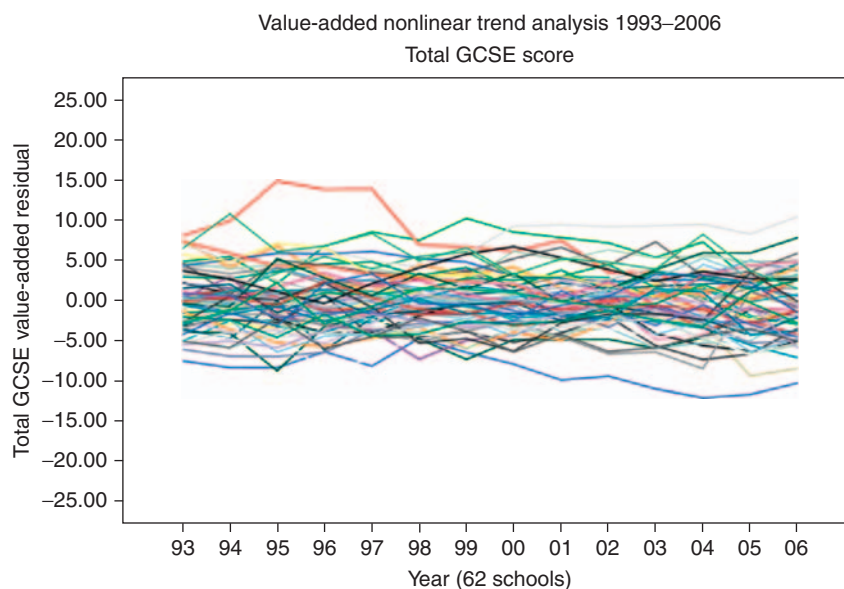


Figure 2 Schools value-added performance – nonlinear time trends over 14 consecutive student cohorts.

from the average linear trend, indicating either improvement or decline in performance. In this context, it is important to emphasize that real improvement (or decline) in performance, resulting perhaps from a shift in school policy or practice, can only be identified by examining long-term changes in results over time (over three cohorts as a minimum). Therefore, validity of institutional effectiveness measures will be enhanced by examining in detail equivalent outcome measures for consecutive student cohorts, alongside the educational processes associated with different patterns of improvement. The statistical power of the analysis is also improved by the increased sample size when successive student cohorts are modeled simultaneously.

Also relevant to validity of measuring institutional effectiveness is the issue of regional or national differences in the size, extent, and consistency of school effects or the differential impact of pupil and school background characteristics in different regional, socioeconomic, and educational policy contexts – issues that have been addressed by relatively few studies. Although further analysis is required using national databases, Thomas (2001) reports tentative findings which suggest that the difference or similarity in schools' departmental results can and do vary across UK regions, indicating that local context measures and, possibly, region-specific outcomes also need to be considered in order to reflect the possible influence of local area, regional, and national policy and practice on institutional effects.

Thus, the validity of institutional effectiveness measures is dependent on and limited by a number of factors including:

1. Data

- Data employed – are longitudinal data employed including prior attainment data and what statistical adjustments are made for factors outside the control of the school?
- Data accuracy and measurement error – have data and measurement errors been minimized?
- Representative sample – does the sample of schools and students reflect the maximum range and/or variability in performance?

2. Method of analysis

- Statistical method – are descriptive statistics, multiple regression, or more complex data modeling (e.g., multilevel modeling) used?
- Unit of analysis – are the significance of variations in performance at the level of students, classes/teachers, and regions considered, as well as at the institution level?
- Time period – what time period is examined, and are data from multiple/consecutive student cohorts used to examine changes in school effects and improve the power of analysis?

- Statistical uncertainty – are statistical confidence intervals presented alongside all performance measures? Note that this uncertainty prevents any fine distinctions to be made between the performance of most schools.

3. Performance outcomes

- Student groups – are separate measures used for different student groups?
- Subject areas – are separate measures used for different academic areas of curriculum?
- Other outcomes – are separate measures used for other areas of curriculum such as vocational, health, and citizenship education?
- Regional outcomes – are separate measures used for different regions that vary in school policy, practice, or context?

Clearly, the validity of institutional effectiveness measures is influenced to a large extent, by the quality, reliability, and validity of the data analyzed. The methodology is also not perfect in the sense that it is retrospective and, as for any data modeling, today's results cannot necessarily predict future performance. Moreover, it is impossible to measure all relevant factors outside the control of the school. For example, the indicator of student disadvantage eligible for free schools meals (FSM) may be inaccurate because some parents do not apply for an eligibility means test. Furthermore, the system does not completely cover all students likely to suffer from social and economic disadvantage. Other relevant measures of socioeconomic status, such as private tuition and level of parental education, occupation, and income are difficult and costly to collect and, therefore, are often not examined by researchers in sufficient detail. Therefore, it is important to remember that value-added estimates reflect the school effect as well as any other relevant factors not accounted for in the analysis. All the above caveats and limitations point to the critical need to consider the statistical significance of individual school results as well as other relevant data or evidence that may be available in a school in order to avoid making unjustified claims or misinterpreting the results. Value-added approaches can provide a powerful and sophisticated tool for institutional evaluation and self-evaluation; however, crucially, validity issues and limitations need to be well understood by teachers, policymakers, and researchers.

Using Value-Added Measures to Evaluate Institutional Effectiveness

Given the validity issues outlined above, a cautious approach should be employed whereby performance data are seen as a means of raising questions about any particular school results, and a way of identifying where further

evidence is required, rather than a definitive measure of institutional effectiveness. Wherever possible, a broad range of measures are required to reflect more fully the aims of schooling and contrast the results against other types of evaluation data available about schools, such as information about the views of key groups obtained using, for example, student, teacher and parent questionnaires. As a summary, and with these caveats in mind, some of the key uses of value-added measures are listed below:

1. Research

- examine overall trends in pupil and school performance over time;
- examine local, regional, or national differences in value-added results between schools and the implications for local, regional, or national education policy;
- examine case studies of more/less effective schools; and
- examine the school processes and factors related to improvement or decline in school effects.

2. School self-evaluation

- Provide teachers with information that will allow them to reflect on, evaluate and improve their educational practice. For example, examine;
 - (1) departmental, subject, and/or teacher effectiveness versus summary measures of school effectiveness and the implications for whole school policies;
 - (2) differential effectiveness for different groups of students (e.g., boys/girls, high/low attainers, and different ethnic groups) and implications for equal opportunities; and
 - (3) effectiveness for different year groups or age cohorts, such as students at different curriculum stages, and implications for differing rates of progress/curriculum coverage.

3. External evaluation

- Provide school inspectors with information about relative school performance that can be used alongside their own judgments and other evidence of quality and standards.
- Track changes in results over time to examine real improvements, or random fluctuations in performance, or both, in relation to school improvement initiatives.
- Examine the way teachers use data to reflect on past performance and to inform, evaluate, and improve their current policy and practice. Note that this approach is now a key aspect of some national inspection systems, for example, in England.

4. Accountability

- Publication of schools' value-added results. However, note that league tables or ranks will always be problematic due to the need to consider statistical significance of each school's results.
- Inform parental choice of schools.

5. Funding or rewards

- Inform funding review or as one type of evidence in allocating school awards/rewards.

Conclusions

In summary, the general consensus of research evidence from several countries is that raw statistics of student performance alone cannot give an accurate picture of how effective a school is at raising and maintaining the achievement of all its students, or how capable it is of sustaining its standards over time. Moreover, some schools that may appear to be effective overall in terms of value-added performance may not be so effective in terms of individual departments, different groups of students, different year groups, or over different periods of time. Research findings indicate that internal variations in effectiveness need to be monitored at all stages of statutory education. Value-added approaches also provide a powerful methodology for understanding the mechanisms and levers of school improvement – for example, by comparing value-added and school process data at any particular stage or phase of schooling. The analysis and feedback of value-added data are often viewed as an integral part of the development and improvement work carried out with schools by external consultants, school inspectors, and academic researchers. Examining the way teachers and other practitioners use data to reflect on past performance and to inform and evaluate their current policy and practice is a crucial aspect in understanding how schools improve.

The complex issues related to evaluating institutional effectiveness are likely to be particularly relevant in countries that are not currently employing value-added methods, and where the focus is still on raw examination data to judge school performance. In these cases, the availability and analysis of longitudinal individual student-level data are essential to allow schools and teachers to examine different aspects of their school's effectiveness in detail. Research has demonstrated overwhelmingly that value-added approaches provide a powerful innovation to compliment and develop methods of assessing students and schools. As a final point, it is also relevant to note that comparisons of educational quality between countries are often based on international surveys of student attainments that are typically cross-sectional and do not employ value-added approaches (Goldstein and Thomas, 2008). This situation similarly highlights the urgent need for longitudinal data sets to improve the validity of internationally comparative measures of institutional and educational effectiveness.

See also: Evaluation and Accountability; Summative Assessment by Teachers; Validity: Mapping Diverse Perspectives.

Bibliography

- Goldstein, H. (1997). Methods in school effectiveness research. *School Effectiveness and School Improvement* **8**(4), 369–395.
- Goldstein, H. and Thomas, S. M. (2008). International surveys of educational achievement – reflections on the debate. *Special Issue: International Comparative Studies of Achievement. Assessment in Education* **15**(3), 215–222.
- Hamilton, L. S., Stecher, B. M., and Klein, S. P. (2002). Introduction. In Hamilton, L. S., Stecher, B. M., and Klein, S. P. (eds.) *Making Sense of Test-Based Accountability in Education* (MR-1554-EDU), ch. 1. Santa Monica, CA: RAND.
- Harlen, W. and Deakin Crick, R. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning. *EPPI Review*. <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=108> (accessed August 2009).
- Kellaghan, T. and Greaney, V. (2004). *Assessing Student Learning. Directions in Development 30115*. Washington, DC: The World Bank.
- MacLeod, D. (2007). Countdown to the guardian's university league tables. *Guardian*. <http://education.guardian.co.uk/higher/news/story/0,,2067182,00.html> (accessed August 2009).
- McCaffrey, D. F., Koretz, D., Lockwood, J. R., and Hamilton, L. S. (2004). *Evaluating Value-Added Models for Teacher Accountability*. Santa Monica, CA: RAND.
- Nuttall, D. (1990). *Differences in Examination Performance RS1277/90: 25*. London: Research and Statistics Branch, IEA.
- OECD (2007). *Education at a Glance*. Paris: OECD.
- OFSTED (2004). *A New Relationship with Schools: Improving Performance through School Self-Evaluation*. Nottingham: The Department for Education and Skills.
- Peng, W. -J., Thomas, S., Yang, X., and Li, J. (2006). School evaluation methods to improving the quality of schooling in China. *Special Issue: Asian-Pacific Region. Assessment in Education* **13**(2), 135–154.
- Reynolds, D. (2007). School effectiveness, school improvement and accountability – a European survey. In Townsend, T. (ed.) *International Handbook of School Effectiveness and Improvement*. ch. 26, pp 471–484. Dordrecht: Springer.
- Sammons, P., Elliot, K., Sylva, K., et al. (2004). The impact of pre-school on young children's cognitive attainments at entry to reception. *British Education Research Journal* **30**(5), 691–712.
- Sammons, P., Thomas, S., and Mortimore, P. (1997). *Forging Links: Effective Schools and Effective Departments*. London: Paul Chapman.
- Scheerens, J. (2004). Benchmarking the quality of education. *European Journal of Educational Research* **3**(1), 177–212.
- Scheerens, J., Glas, C., and Thomas, S. (2003). *Educational Evaluation, Assessment and Monitoring: A Systemic Approach*. Lisse, Swets and Zeitlinger.
- Thomas, S. (2001). Dimensions of secondary school effectiveness: Comparative analyses across regions. *School Effectiveness and School Improvement Journal* **12**(3), 285–322.
- Thomas, S. M., Peng, W. -J., and Gray, J. (2007). Value added trends in English secondary school performance over ten years. *Oxford Review of Education* **33**(3), 261–295.
- Van Petegem, P., Vanhoof, J., Daems, F., and Mahieu, P. (2005). Publishing information on individual schools? *Educational Research and Evaluation: An International Journal on Theory and Practice* **11**(1), 45–60.
- Volante, L. (2007). Evaluating test-based accountability systems: An international perspective. *Paper Presented at the Association for Educational Assessment – Europe*, Stockholm, Sweden, November 2007. <http://www.aea-europe.net/userfiles/Volante%20paper%20AEA-Europe.pdf> (accessed August 2009).
- Willms, J. D. and Raudenbush, S. W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement* **26**(3), 209–232.
- Yu, G. and Thomas, S. M. (2008). Exploring school effects across Southern and Eastern African school systems and in Tanzania. *Special Issue: International Comparative Studies of Achievement. Assessment in Education* **15**(3), 283–306.

Further Reading

- Goldstein, H. (1993). Assessment and accountability. *Education* **183**(3), 33–34.
- Goldstein, H. and Spiegelhalter, D. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society A* **159**(3), 385–443.
- Gray, J., Goldstein, H., and Jesson, D. (1996). Changes and improvements in schools' effectiveness: Trends over five years. *Research Papers in Education* **11**(1), 35–51.
- Karsten, S., Visscher, A., and De Jong, T. (2001). Another side of the coin: The unintended effects of the publication of school performance data in England and France. *Comparative Education* **37**(3), 231–242.
- McGeevor, P., Giles, C., Little, B., Head, P., and Brennen, J. (1990). *The Measurement of Value Added in Higher Education*. London: PCFC/CNA.
- Thomas, S. and Mortimore, P. (1996). Comparison of value added models for secondary school effectiveness. *Research Papers in Education* **11**(1), 5–33.
- Visscher, A. J. (2001). Public school performance indicators: Problems and recommendations. *Studies in Educational Evaluation* **27**(2), 199–214.

Relevant Websites

- <http://www.bris.ac.uk> – Bristol University, Creating and Sustaining Effective Professional Learning Communities (EPLC) Project.
- <http://www.cmm.bristol.ac.uk> – Centre for Multilevel Modeling (CMM).
- <http://www.dcsf.gov.uk> – Department for Children, Schools and Families (DCSF).
- <http://www.gtce.org.uk> – General Teaching Council (GTC) for England.
- <http://ieeqc.bristol.ac.uk> – Improving Educational Evaluation and Quality in China (IEEQC) Project.
- <http://www.innovation-unit.co.uk> – Innovation Unit UK.
- <http://www.ncsl.org.uk> – National College for School Leadership (NCSL).
- <http://www.ed.gov> – No Child Left Behind (NCLB) USA's ED.
- <http://www.raiseonline.org> – RAISEonline.

Assessment in Schools – Dispositions

R D Crick, University of Bristol, Bristol, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Autogenic – Originating in the self.

Disposition – A tendency to behave in a certain way.

Habitus – A system of cognitive and somatic dispositions developed in a sociohistorical context over time.

Hermeneutic – A particular interpretation of phenomena.

Hexis – A habit resulting from behaviour repeated over time.

Telos – An overarching goal, or desired end point.

Virtue – An acquired human quality the possession of which leads to a desired good.

Introduction

The notion of dispositions as educational outcomes has emerged as an important but contested area in educational theory and practice in the last 10 years. In the changing world of the twenty-first century, it is widely recognized that the acquisition of knowledge, skills, and understanding is not enough. On the one hand, proliferation and accessibility of information render simple accumulation impractical, even if there were agreement as to what was important; on the other hand, a sustainable social world requires a range of competences on the part of individuals and communities which enable successful functioning in real situations: such competences include values, attitudes, and dispositions as well as cognitive resources. This article deals first with contemporary psychological definitions of the term dispositions and second with how dispositions are located and operate in social contexts, drawing on the work of Bourdieu in particular. It then identifies two widely accepted candidates for dispositions as educational outcomes and goes on to explore the professional and theoretical issues encountered in their assessment in classrooms. Finally, it summarizes a range of assessment practices drawn from the literature.

Conceptual Issues

Although the term disposition as an educational outcome is imprecise, both theoretically and in practice, it is widely

agreed that it refers to a relatively enduring tendency to behave in a certain way. It is a construct linked to motivation, affect, and valuing as well as to cognitive resources. It is acquired and manifested in a sociohistorical setting. Dispositions may be culture specific as well as a relatively enduring feature of personality. A disposition may be recognizable in the action a person takes in a particular situation – for example, someone who is disposed to be curious will demonstrate it in generating questions and investigating problems. In practice, in education, the term is often used interchangeably with competence, or style, or capability and it is frequently subsumed within the concept of personal development as distinct from academic development or attainment. There are many dispositions relevant to education – ranging from the specific to the very general. In this article, the focus is on two broad types of disposition widely perceived as of educational value: dispositions for learning and for active citizenship.

One of the earliest uses of the term as an educational outcome was Katz who defined dispositions as relatively enduring habits of mind or characteristic ways of responding to experience across types of situations. She went on, in a later review of literature, to define a disposition as a frequently exhibited pattern of behavior which is voluntary, intentional, and related to a particular goal.

A disposition addresses the gap between the ability to do something and actually doing it, between what people can do and what they actually do in real-life contexts. Dewey also recognized this gap and linked it to desire, suggesting that knowledge of methods alone is not sufficient: there must be the desire and the will to use them. This desire, he argued, is an affair of personal disposition.

The concept of motivation, or desire, is importantly related to dispositions, but it is a notoriously complex area in which there is little consensus. There are drive theories – in which motivation arises from basic drives, instincts, or emotions in predictable ways; qualitative conceptions of motivation, such as achievement motivation – attribution theory and motivational style theory. Ames describes motivation as the systematic, qualitative response that people make to the various challenges and threats arising from situations in which either success or failure is possible.

A disposition arises from desire, or motivation, which provides the energy necessary for action. Ritchhart and Perkins proposed a theory of dispositions that incorporates this motivational and agentic role and includes attention

to sensitivity, inclination, and abilities. Sensitivity is an awareness of occasions to engage in certain behavior, inclination is the motivation to do so, and ability is the capacity to do it.

Thus, they argue that in order to nurture a disposition in the classroom, first an opportunity to engage in certain behavior needs to be recognized or noticed (intuitively or consciously), then the person needs to be inclined to engage in that behavior, and then be able to carry it out.

Carr and Claxton describe these elements of dispositions as being ready, willing, and able to engage profitably with a particular goal, whereas Dweck uses the term orientations and Costa uses the term habits of mind. There is merit in understanding disposition as a verb, with qualifying adverbs, instead of a noun, a thing to be acquired; Carr and Claxton argue that a disposition is not acquired, rather one becomes more or less disposed to respond in such-and-such a way.

The Sociohistorical Location of Dispositions

The term disposition has a long heritage in ancient and medieval philosophy. For Aristotle, people had the capacity to act in certain ways by virtue of possessing a disposition to do so. He argued that people become strong or brave by doing strong or brave things and that by consistently choosing such actions, they become better able to act with strength or courage. One who has performed a brave act must have acquired the disposition to act in that way. Aristotle described a disposition as a habit (*hexis*), which is produced by similar acts and inclines to similar acts. He was primarily concerned with the development of character and its relationship to moral behavior, although he applied the same ideas to bodies of knowledge – for example, one who has trained as a scientist is disposed to act in a scientific way.

Building on this tradition, Bourdieu's concept of *habitus* offers a sociological understanding of dispositions. He was concerned with mechanisms of social domination and reproduction and his focus was on bodily know-how and competent practices in the social world. He defined the term *habitus* as a system of cognitive and somatic dispositions internalized through sociohistorical experience. Bourdieu sees dispositions as a form of embodied cultural capital, whereas objective cultural capital may include books, materials, or machines and institutional cultural capital may include certificates and examination results. Accordingly, dispositions and habits are inculcated through childhood experiences and the cultural practices and values of the classroom, which in turn are shaped by the structures and practices of the schooling system. The formation of dispositions is the site for the development of agency in the learner within a limited arena of choice.

Thus, the question of the psychological stability of particular dispositions in an individual over time and in different contexts is debatable since scholars, such as Bourdieu, argue that they are shaped by culture and history. Sociocultural theorists have convincingly argued that human learning and performance are highly situation specific, originating and developing in a context of interaction. Thus, formation of dispositions is influenced by a range of factors: psychological, cultural, historical, and social – and there is an ethical element, since a disposition leads to action in pursuit of a desired goal. In the ecology of the classroom, specifically, there are many factors which are shown to influence students' achievement and hence their dispositions; these include, but are not limited to, quality of relationships and teacher practices; the classroom climate; assessment practices and pedagogic discourse.

Related Concepts

Certain concepts, used as educational outcomes, which are related to dispositions should be discussed and differentiated. In particular, the terms competence and learning styles have been prominent in both policy and practice. A competence is a complex combination of knowledge, skills, understanding, values, attitudes, dispositions, and desire that lead to effective, embodied human action in the world, in a particular domain. Achievement at work, in personal relationships, or in civil society is not simply successful accumulation of received knowledge, but a combination of this knowledge with skills, values, attitudes, desires, and motivation and its application in a particular human setting at a particular point in a trajectory in time. Competence implies a sense of agency, action, and value.

The learning styles field is complex and confusing, with a host of conceptual and empirical problems. A learning style is an individual's preferred way of learning and some scholars have developed a classification of the field based on the degree to which the underlying assumptions about learning styles are fixed. At one end of the spectrum is the assumption that learning styles and preferences are largely fixed and constitutionally based and, at the other, is the assumption that they are changeable and better described as orientations, learning approaches, or learning strategies.

From this discussion, it is possible to develop a limited and an elaborated description of a disposition (**Figure 1**). A limited description focuses on a tendency to behave in a certain way in a particular context. At the other end of the spectrum an elaborated description is of a disposition embedded in sociocultural, historical, and ethical narrative and includes a sense of agency, intention, and capability in real-life contexts of achievement, lifelong learning, and citizenship. The latter locates dispositions

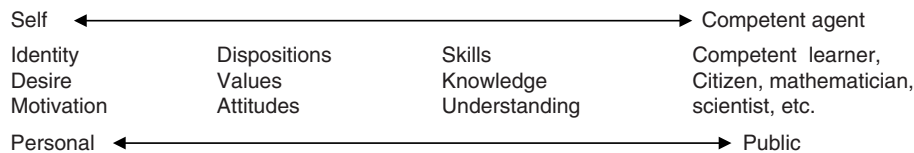


Figure 1 Elaborated description of dispositions. From Dr. Ruth Deakin Crick. Published by Elsevier Limited. All rights reserved.

and their assessment as part of a complex process of sustainable, human learning and change over time. In the elaborated description, the telos of the process is a person competent in a particular domain – for example, a competent citizen, a competent learner, a competent mathematician, designer, or ergonomist.

Potential Candidates for Education

There is no shortage of lists of desirable dispositions for educational outcomes. Carr, for example, derives a set of learning dispositions from the New Zealand national early-childhood curriculum: courage, curiosity, playfulness, perseverance, confidence, and responsibility. Deakin Crick *et al.* identified a set of seven dispositions for learning: changing and learning; critical curiosity; meaning making; creativity; learning relationships; strategic awareness; and resilience. Crick described learning outcomes for citizenship education as a composite including knowledge and understanding, skills and aptitude, and values and dispositions. Among the latter were acting responsibly, caring for others, courage to defend a point of view, and openness to changing one's mind.

One of the challenges for schools is to identify dispositions that are key to successful life in the twenty-first century. Dispositions for learning to learn and citizenship are widely accepted candidates, described as elements of key competencies which are as indispensable for the good life – reminiscent of an Aristotelian notion of virtue.

Issues in the Assessment of Dispositions

The assessment of dispositions raises several important philosophical and methodological challenges. What we assess is, or should be, a function of what we value and there can be little doubt that contemporary educational assessment should take into account the importance of dispositions, particularly learning dispositions. The global accessibility of traditional, factual knowledge means that the question of its value is a local, personal one, concerned with its relevance to a context and purpose brought to it by a learner. What I learn cannot be separated from what matters to me – it is a product of intention and desire. Therefore, knowledge has become no longer simply what I learn, but also how and why I learn. While intentional

learning begins with desire, which is autogenic and personal, the outcomes of formal learning are mostly assessable by publicly agreed criteria set by a particular community of practice and are thus formal, external, and publicly valued.

This integration of the personal with the public opens space for a fresh approach to continuous formative assessment and evaluation. How can what we value, such as learning dispositions, be identified and assessed within each stage of the journey from desire to competence, rather than simply via set pieces of performance at the end? Assessment could then be linked with and derived from the formation of learning dispositions, which begin with desire and move through self awareness, ownership, and responsibility toward the acquisition of formal knowledge, skills, and understanding. Assessing learning dispositions establishes the quality of learning power, recognized, owned, and assessed by the learner in the first instance, along a sequential movement from the autogenic to publicly agreed and formally valued funds of knowledge, skills, and competence.

Desire, as Dewey said, is an affair of personal disposition. To desire to learn, or to be courageous, leads to an inclination to learn or be brave and to make a personal choice rooted in experience, expressing a lived narrative. Narrative is a key concept in relation to the assessment of dispositions: forming dispositions for learning, or citizenship, involves the person who is learning and requires desire, a sense of direction and the hope, agency, and self-regulation to set out in that direction.

An important assessment question in relation to what is personal and intuitive is who has the authority to authenticate and assess this sense of agency and intentionality? How valid can an outsider's judgment of it be? The learner herself must be the arbiter, since desire and dispositions are embodied and unique to the person. Thus, the most valid starting point in assessing dispositions is self-assessment. What can be done from the outside is suggest criteria for it.

The critical questions for the learner at this stage are what really matters to me? Also, what do I really want to know about it? The readiness with which a learner takes up the challenge and identifies the object of a new enquiry, the energy galvanized by his/her perplexity and fascination, and the acuteness of his/her discrimination and selectivity might all ultimately be reflected in the quality and quantity of his/her learning and its outcomes. At the time, however, their power can only be assessed by

the learner himself/herself, perhaps using criteria such as strength of engagement, extent of commitment, degree of critical curiosity, quality of self-awareness, and potentiality of relationship between the self and the object of interest.

Once a desire is stimulated and activated and a learner is engaged and acting on the desire to learn, or doing something about an injustice perhaps, then they are embodying learning or citizenship dispositions – such as, say, curiosity or meaning making – and utilizing them to describe, organize, and sort material and data into their chosen representation of it. Here, the mode of assessment must still therefore be anchored in self-assessment, since only the learner can assess the extent to which his/her representation captures the engagement, strength of purpose, and personal elements that arise from desire and choice. The assessment criteria will move beyond being purely personal at this stage, however, because already the enquiry is beginning to acquire a communicative purpose. A thread between the internal, personal starting point of the learning and the inevitable, ultimate arbiter of an external judgment starts to be visible. Assessment is still grounded in the learner's self-evaluation but with some more objective criteria, such as clarity of observation, accuracy of response, harmony or appropriateness of organization, and breadth and depth of analysis. It is also appropriate, with some sensitivity at this stage, for teacher assessment and peer assessment to be introduced for formative purposes: helping the learner to see how he/she might improve the learning, particularly its capacity to achieve its communicative purpose.

Dispositions, according to Bourdieu, are embodied and historical, linked to individual history and belonging to a genetic mode of thought. They are formed by habit, and reenactment, such that they become durably incorporated into the body. It follows that narrative is an important means of understanding, articulating, and assessing dispositions because narrative enables, captures, and illuminates experiences as human beings live them in time, in space, in person, and in relationship. Narrative attends to a three-dimensional inquiry space – the temporal, the spatial, and the personal-social. Human lives are woven of stories. Individuals construct their identities through their own and others' stories. They experience daily encounters and interactions as stories. Every present moment has a storied past and a storied future possibility.

The use of narrative in the self-assessment of dispositions is therefore important and may provide a key link between the personal and the public, between the autogenic and the formal. Describing, or telling, a learning process as a story enables the construction and expression of a learning identity and the formation of a learning disposition through meaning making and identification. When this story intersects with narratives embedded in the material world, which are the focus of learning, then this is a powerful site for the formation of dispositions for citizenship.

The capacity to tell a story can be assessed and validated against yet-more objective and external criteria, concerned with the achievement of communicative purpose. These may even include technical and esthetic considerations, such as appropriateness of form and language, breadth and depth of view, continuity and consistency of style, range and coherence of narrative or plot between arousal or surprise at the new and resolution, or satisfaction in its relationship with the known. In formulating, recounting, and evaluating the story of the learning itself, the learner is enacting and embodying the disposition of reflectiveness in the context of an even stronger emphasis on learning relationships or reciprocity and, thus, engaging in self-evaluation.

In tracking the formation of dispositions, as the learner moves toward the public end of the journey, the issues in the assessment of dispositions come closer to what has been traditional in the assessment of knowledge, skills, and understanding: recognition and interpretation of public or imposed problems, questions, and issues; selection and use of information/evidence to address them; relevance and breadth of reference; and reasoning/logic to justify opinions/solutions. The knowledge, skills, and understanding outcomes of the learning process can be assessed summatively against externally validated criteria and assessments can be demonstrated to be more or less reliable in the traditional sense.

However, since dispositions are enacted in the process of learning – whether that is learning to become an active citizen or learning how to become a better mathematician – any summative assessment of dispositions at the end of the process can only engage with what the learner asserts or the teacher observes to have been achieved, either directly or deductively on the basis of the evidence presented in the product of learning. Such evidence may be identified in learners' narrative accounts of the real-life issues they have explored in the formulation of their learning outcome, or in metacognitive reflection on the process of learning. Polanyi argued that criteria for validating knowledge can be subjective while offering a valid basis for objective judgment – suggesting criteria such as beauty, simplicity, and coherence. The point is that assessment criteria for summative purposes, at the end of a process, need not only be technical and subject specific, but also be ethical and esthetic, taking account of ephemeral evidence of the process as well as of tangible evidence in its more durable products.

Validity and Reliability – or Authenticity and Trustworthiness?

At the heart of the debate about the assessment of dispositions is the question of reliability. Traditionally, reliability is operationalized by examining consistency, defined

quantitatively, across independent observations, which are intended as interchangeable. Thus, measurements can be standardized and generalized across populations. Reliability is a necessary condition for validity.

However, while we should not abandon reliability in assessment, we should consider it as just one means of serving important epistemological and ethical purposes. The purpose of assessment should be to improve learning and teaching and, in addition to the post-positivist notion of reliability, there is a need for a more hermeneutic, or interpretive approach which honors the purposes and lived experiences of students and the collaborative judgments and discernment of teachers.

Since dispositions are embodied in people, and people are idiosyncratic, the choice of the traditional reliability criterion, couched within a psychometric, post-positivist paradigm, is inappropriate for their assessment. Even if it were possible to achieve a consistent, quantitative measure across observations of students' dispositions for, say truth seeking, or reflectiveness, what educative purpose would that serve? People are not the same, nor should they be. The purpose of the assessment of dispositions is to enable and strengthen the individual learner – as a learner, or an active citizen, mathematician, or an artisan, and to provide teachers and mentors with diagnostic information so that they can target their pedagogical practices more precisely.

In the assessment of dispositions, authenticity and trustworthiness, rather than reliability, are the necessary conditions for validity. Does this particular assessment represent an authentic picture of the individual at a particular point in time and in a particular domain? Is it true to the experience of formation, as well as to the enactment or product, of his/her disposition? While there may be public agreement that curiosity or courage is manifest in certain behaviors, such as asking questions or standing up for one's beliefs, their manifestation *in situ* cannot be generalized, nor can they be quantitatively compared between individuals or across cultures, because of the presence of so many confounding variables, both internal and external to the learners. At most, a range of possible criteria for curiosity might be drawn upon in making judgments in relation to an individual and a context. Then, ultimately, the face validity of the assessment, for the individual and the authoritative community around him/her, must be the final arbiter.

One exception to generalization of assessment data in relation to dispositions is in self-report questionnaires such as the Effective Lifelong Learning Inventory, in which students respond to questions designed to elicit information about their levels of strategic awareness, or creativity. Here, the assessment data are fed back to individuals in graphic form, but the raw data can be captured across whole populations and can be subject to statistical analysis. The European Union (EU) Commission is currently exploring

the possibility of assessing learning to learn dispositions through self-report as a means of evaluating the development of learning-to-learn competencies across the EU. Here, however, what is being measured, and thus possibly generalized across a population, is what individuals say about themselves in relation to the characteristics of a disposition, rather than an objective measure of the actual performance of a particular disposition. The former is what matters, and the latter is not possible or desirable.

Dispositions as Learning Outcomes

So far the discussion has been about the formation of dispositions in the process of learning how to learn or how to be an active citizen, rather than as an outcome. There is a growing body of evidence that attending to the process of learning to learn, particularly through the development of metacognitive strategies and habits, self-regulation, self-efficacy, and motivation, has a positive impact on learning outcomes and overall achievement.

However, the accumulation and practice of dispositions over time, such that they become relatively permanent features of a person's character or identity, may be an educational goal or outcome as much as the accumulation of standards of excellence in a particular discipline or skill. As an outcome this is akin to MacIntyre's notion of a virtue defined as an acquired human quality, which enables the individual to achieve those goods which are internal to particular practices. Learning to learn and active citizenship are types of human social practice that have their own internal goods, standards of excellence, and, in themselves, lead to a certain kind of good life – that of the lifelong learner or active citizen.

Modes of Assessment of Dispositions

The mode of assessment for dispositions can be divided into three types:

- self-assessment by the learner;
- analysis of self-report questionnaires or interviews; and
- observation of behavior and outcomes by teachers, mentors, or peers.

The information derived from these assessments can be used formatively – in order to stimulate awareness, ownership, and responsibility in becoming a better learner or active citizen, or summatively – in order to summarize learners' achievements at a particular point in time. Classroom assessments for dispositions need to be valid and authentic: Do they measure what they claim to measure and include the learner as well as the teacher in their validation? They need to be flexible and pedagogically

useful: Can learners and their teachers adapt them to particular contexts, domains, and individuals? They also need to be relational – since there is an implicit movement within the assessment of dispositions between the self of the learner and the formal text of what is being learned, requiring trust in pedagogical relationship(s), which affirm and challenge the learner to take responsibility for his/her own journey.

A holistic approach to reflecting learners' achievements, which recognizes the complexity of development of skills and competencies, might use criteria of depth and quality of contribution in an identified context and range of application in a diversity of contexts. Similar criteria include robustness and sophistication in the use of dispositions.

Assessment Practices for the Formation of Dispositions

Figure 2 builds on an elaborated description of dispositions to locate their assessment in the formation of competence(s) over time. It shows how assessment practices and relationships are complex and dynamic and it is impossible to separate questions of assessment of personal capabilities from curriculum, pedagogy, and culture. Assessment should be integrated into the process of learning; it should be sensitive to context, and promote self-worth; be meaningful and owned by learners; act as a bridge between learners and the community; and be flexible, evolutionary, and responsive to diversity. Their proposal for assessment in this domain is a relational one: that

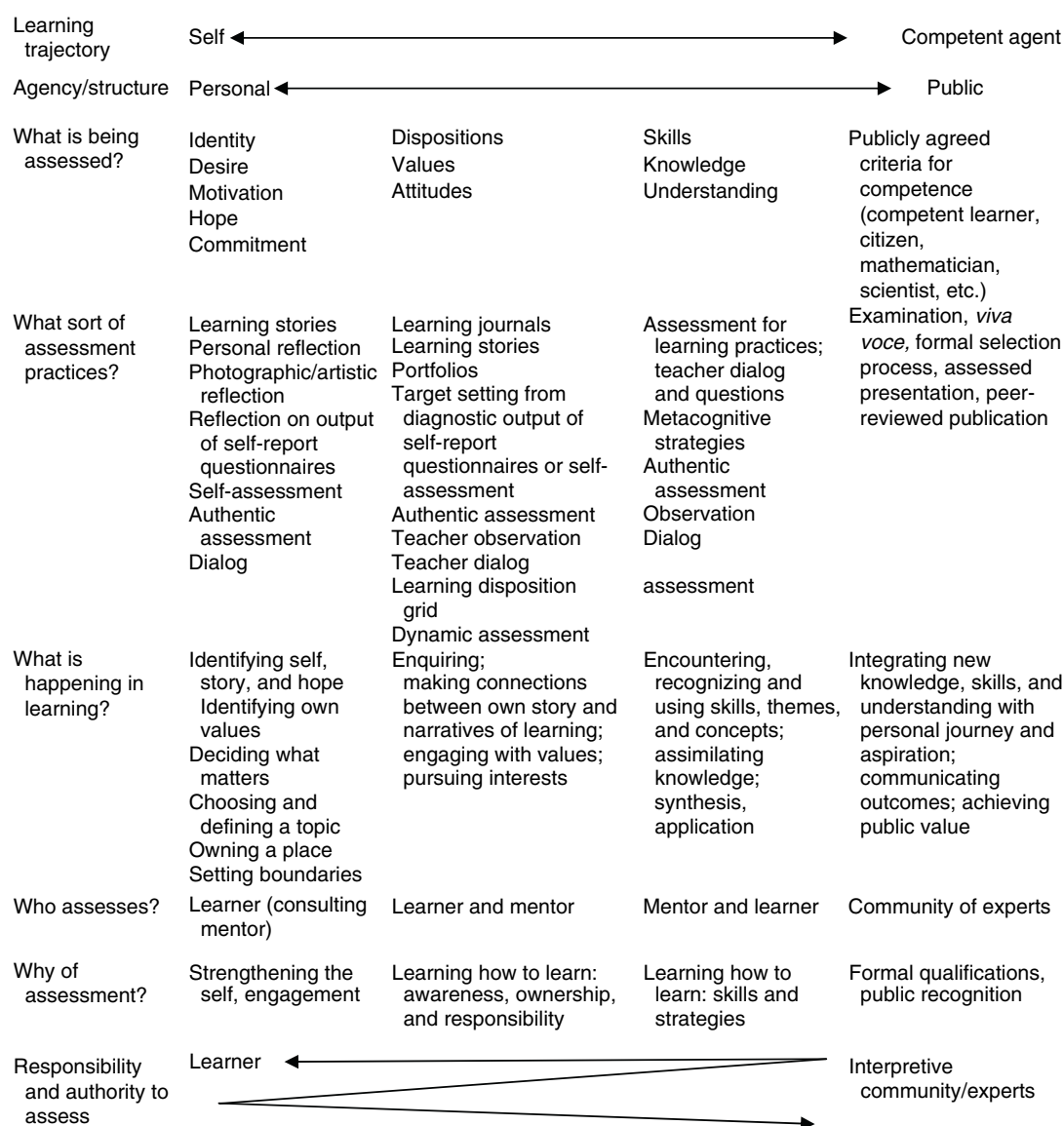


Figure 2 Assessment practices and purposes for dispositions in the trajectory from self to competent agent. From Dr. Ruth Deakin Crick. Published by Elsevier Limited. All rights reserved.

is, it is based on key questions to be explored through dialog between learners and their teachers, together with forms of modeling through examples of work and role models. The dialog and modeling is based on agreed characteristics of the dispositions, values, and attitudes and skills deemed to be necessary for becoming competent and capable as learners. **Figure 2** illustrates the idea of assessment as dynamic and relational, between learner and mentor and between the personal and public domains. The assessment of dispositions can be described as the mentored movement of selective attention between the self and the formal texts of the expert community.

Thus, there are several genres of assessment practice which are relevant to dispositions. The models of dynamic assessment developed by Feuerstein, which draw explicitly on the Vygotskian notion of the zone of proximal development and identify a person's learning power through the way they negotiate scaffolded interaction with the assessor, are relevant. Although perhaps not easily manageable in regular classrooms, they can have a place in the assessment of dispositions. Assessment for learning practices is centrally important and well discussed as a vehicle for developing learning strategies, skills, and dispositions. The use of stories in early childhood assessment has been developed substantially by Carr, while Mahn and John Steiner have shown how the use of interactive journals facilitates the development of second-language learning. Carr and Claxton suggest that portfolios are a powerful tool in tracking dispositions because they can potentially include data from the full range of assessment methods, and demonstrate both progress and achievement over time. They also recommend a learning disposition grid as a vehicle for capturing samples of evidence of a learner's dispositions over time and *in situ*. In terms of self-report questionnaires, there is a plethora of learning styles questionnaires and, more recently, the concept of learning power and its assessment has emerged as a means of reflecting back to learners what they say about themselves in relation to learning-to-learn dispositions and using that information formatively to strengthen learning power.

Conclusions

The concept of dispositions is a contested, but important area in education, particularly at a time when there is increasing international concern for social sustainability, through the development of a range of competences on the part of individuals and communities which enables successful functioning in real-world situations: such competences include values, attitudes, and dispositions, as well as cognitive resources. A limited description of a disposition focuses on an individual's tendency to behave in particular ways over time, but a more complex, elaborated

understanding locates dispositions as part of an embedded and embodied journey over time, from personal desire and motivation to the achievement of competence in a particular public domain. Competence in this sense incorporates dispositions, knowledge, and know-how as well as a contemporary reading of virtue. The focus of this article has necessarily been restricted to dispositions as part of competence for learning and for citizenship, since these are widely accepted educational outcomes. The challenge of assessing dispositions, in this more complex reading, is rooted in their relationship both to the deeply personal and to learning-to-learn skills, strategies, and achievement. Thus, assessing dispositions can be described as the mentored movement of selective attention between both self and text and between the self and the expert, competent community. Dispositions both scaffold the process of the journey toward competence and form part of the outcome of the journey. Validity of assessment is demonstrated through interpretation – primarily by the learner himself/herself, using the criteria of authenticity and trustworthiness. The purpose of assessment of dispositions is to enable learning and growth. Aggregated data from student self-report questionnaires can also be used for organizational learning. Finally, a range of types of assessment practices can be identified, providing an indication of the sorts of resources which could be drawn upon in developing local and situated assessment practices.

Further Reading

- Ames, C. (1987). *The Enhancement of Student Motivation. Advances in Motivation and Achievement*. Greenwich, CT: JAI Press.
- Bourdieu, P. (1993). *Sociology in Question*. London: Sage.
- Carr, M. and Claxton, G. (2004). A framework for teaching learning: The dynamics of disposition. *Early Years* **24**, 87–97.
- Coffield, F., Moseley, D., Hall, E., and Ecclestone, K. (2004). *Learning Styles and Pedagogy in Post-16 Learning: A Systematic and Critical Review*. London: Learning and Skills Research Centre.
- Costa, A. (2000). Describing the habits of mind. In Costa, A. and Kallick, B. (eds.) *Habits of Mind: Discovering and Exploring*, pp 21–40. Alexandria, VA: Association for Supervision and Curriculum Development.
- Crick, B. (1998). *Education for Citizenship and the Teaching of Democracy in Schools. Final Report of the Advisory Group on Citizenship*. London: Qualifications and Curriculum Authority.
- Crick, R. D. (2007). Learning how to learn: The dynamic assessment of learning power. *Curriculum Journal* **18**, 135–153.
- Crick, R. D., Tew, M., Taylor, M., Durant, K., and Samuel, E. (2005). *A Systematic Review of the Impact of Citizenship Education on Learning and Achievement. Research Evidence in Education Library*. London: Evidence for Policy and Practice Co-ordinating Centre, Social Science Research Unit, Institute of Education.
- Dewey, J. (1933). *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. Boston, MA: Heath.
- Dweck, C. (1999). *Self Theories: Their Role in Motivation, Personality and Development*. Philadelphia, PA: Psychology Press.
- Facer, K. and Pykett, J. (2007). *Developing and accrediting personal skills and competencies*. Bristol: Futurelab.
- Feuerstein, R. (1979). *The Dynamic Assessment of Retarded Performers: The Learning Potential Assessment Device*. Baltimore, MD: University Park Press.

- Grigorenko, E. and Sternberg, R. (1998). Dynamic testing. *Psychological Bulletin* **124**, 75–111.
- Hattie, J., Biggs, J., and Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta analysis. *Review of Educational Research* **66**, 99–136.
- Higgins, S., Baumfield, V., and Hall, E. (2007). *Learning Skills and the Development of Learning Capabilities, Technical Report. Research Evidence in Education Library*. London: EPPI Centre, Social Science Research Unit, Institute of Education, University of London.
- Katz, L. G. (1993). *Dispositions: Definitions and Implications for Early Childhood Practices. Perspectives from ERIC/EECE: A Monograph Series, No. 4*. Urbana, IL: ERIC Clearinghouse on Elementary and Early Childhood Education.
- Macintyre, A. (1981). *After Virtue*. London: Duckworth.
- Mahn, H. and John-Steiner, V. (2002). Developing the affective zone of proximal development. In Wells, G. and Claxton, C. (eds.) *Learning for Life in the 21st Century, Socio-Cultural Perspectives on the Future of Education*, pp 46–58. Oxford: Blackwell.
- Moss, P. (1994). Can there be validity without reliability? *Education Researcher* **23**, 5–12.
- Ritchhart, R. and Perkins, D. (2000). Life in the mindful classroom: Nurturing the disposition of mindfulness. *Journal of Social Issues* **56**, 27–47.
- Rychen, S. and Salagnick, L. (2003). A holistic model of competence. In Rychen, S. and Salagnick, L. (eds.) *Key Competencies for a Successful Life and a Well-Functioning Society*, pp 41–62. Gottingen: Hogrefe and Huber.
- Tishman, S., Jay, E., and Perkins, D. (1993). Teaching thinking dispositions: From transmission to enculturation. *Theory into Practice* **32**, 147–153.
- Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Assessment in Schools Related To Literacy: Reading

K J Topping, University of Dundee, Dundee, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Adaptive – A computer test which responds to student performance to shorten test.

Auditory discrimination – The ability to tell the difference between similar sounds.

Auditory sequential memory – The ability to remember sounds in sequences of increasing length.

Authentic assessment – An assessment in the context of real reading.

Informal Reading Inventory – Graded reading paragraphs at independent, instructional, and frustration levels.

Item-banked – A computer test with many items, of which a few are presented at each testing.

Phonic/phonetic/phonological – The sign–sound correspondence in reading.

Prosodic demands – The requirements for expression and audience communication when reading.

Reading Miscue Inventory – The analysis of nature of reading errors.

Reliability – The extent to which the same thing is measured on repeated occasions.

Validity – The extent to which measure focuses on what is intended to be measured and only that.

Visual discrimination – An ability to tell the difference among similar visual stimuli.

Visual sequential memory – An ability to remember visual stimuli in sequences of increasing length.

The purpose of this article is to explore those reading assessments which take place in the classroom under the complete control of the class teacher. Other kinds of reading assessments are high-stakes or psychometric tests, both dealt with in another section of this book. High-stakes tests are, typically, scored and interpreted by stakeholders at the school, district, state, or national/federal level. Psychometric tests are typically carried out by professionals other than the class teacher, who may or may not obtain some insight into their results. However, this article focuses on class-teacher assessment.

Teacher Use of Tests

Teachers may, of course, use different types of tests to help them understand student reading progress – or lack

of it. However, these tests represent only one aspect of a total reading assessment, and excessive credence should not be given to an individual test score in isolation. A key issue is whether the test or tests are being used for screening (to indicate which students might be in need of help with reading or at least more intensive assessment) or progression (cumulating year-on-year scores to indicate how a student is developing as a reader). Screening tests are, usually, shorter than progression tests.

Group norm-referenced tests are given to a group of students at a time, and compare each student's performance to a much larger group of students whose overall scores indicate what is normal for all students and how far from normal each newly tested student is. They may be quite brief (and, therefore, of less reliability) or they may be longer and involve weaker students struggling with many items that are too hard for them. They may be embedded in a suite of tests of other areas or be free-standing. Examples in the US include the Iowa Test of Basic Skills and the Metropolitan Achievement Test (including other areas) and Gates-MacGinitie Reading Tests and Woodcock Reading Mastery Tests (free-standing), and – in the UK – the Bristol Achievement Test (including other areas) and the Primary Reading Test and the Group Reading Test (free-standing). Countries with a less widely used language than English as the mother tongue, may have fewer tests, and countries which are also relatively economically underdeveloped, may have very few or none at all.

Computer tests of reading may be merely a computer version of a hard-copy test (replicating all the problems of the hard copy) or may have been specifically designed to be item-banked and adaptive. In this latter case, the student can take the test frequently (because different items from the bank will be presented each time), and because the test adapts to each student (considering at each question whether the answer was right or wrong and choosing to offer an easier or harder question), the testing time is much shorter and students are not stressed by being presented with too many questions they cannot answer. However, all group norm-referenced tests will indicate the quality of student performance by a number (or maybe three similar numbers) – this only suggests the level of the reading, it does not give any deep insight into the nature of the student's reading. The latter is necessary if any kind of relevant intervention program is to be considered.

Some schools subscribe to a testing service. The service provides tests (often, short and in a number of different areas) which are administered by class teachers. The papers

then go off to the service to be scored and the teacher receives a detailed report by way of feedback – which indicates the strengths and weaknesses of each child, often in relation to local as well as national norms (e.g., nearby schools with similar socioeconomically disadvantaged catchment areas). While this is more complex, it still indicates only levels, so the teacher is still left not knowing how to intervene to accelerate learning.

Beyond group tests, there are also individual tests of reading – so called because they are given individually to students, although still norm-referenced. Obviously, this requires much more time per student, so only a few students in the class are likely to receive such assessment. Individual tests are likely to be much more complex – perhaps, involving not only the assessment of accuracy of reading (ability to decode and correctly say a word) but also the comprehension of reading (the ability to understand the meaning implied) and, indeed, also the rate of reading (the speed at which words are correctly decoded). Such tests are also more likely to look more like real reading – they might include passages of some length to read (with an accompanying picture), rather than a list of separate sentences to be completed such as one might find in a group test.

The rate of reading is, however, tricky to interpret – Is a slower but more accurate reader better than a faster but less accurate one? Individual tests might also have a miscue- or error-analysis protocol, encouraging the tester to note down the nature of all individual errors while the student is completing the test, with a view to subsequently allocating to the categories of errors of a more diagnostic nature (see below). These diagnostic indications are much more closely linked with implications for intervention. Computerized diagnostic tests are also available – which require no immediate interaction between the teacher and the student – and these may be useful with somewhat larger numbers of students.

Authentic Assessment

While group-reading tests usually have information with regard to their reliability and validity published in their user manuals, this information is, of course, aggregated – often, reflecting very large numbers. However, the teacher is often most interested in the scores of those students who are not doing well at reading, and these scores are often of atypically low reliability and validity. Thus, there is also a need (and, maybe, primarily a need) for an authentic assessment – which is rooted in the actual processes of real reading in the classroom and beyond. Such an assessment may have a higher reliability and validity for the individual student, especially the one with difficulties. Teachers may be used to making comments concerning their students at the end of the year, and these comments are passed to the next class (or school). However, the length of these comments will not be great, and they are subject to

interpretation by the next teacher, who might – in some cases – decline to read them.

Of course, the teacher will frequently – during the course of teaching – observe students reading, and may, indeed, have sessions where students read aloud to her or him. Observations made during these sessions might be very useful diagnostically – but will the teacher have ways to formalize these observations? They include the cognitive, affective, social, and cultural contexts in which children read. Unfortunately, they all involve time – and the time spent on assessing, is the time not spent on teaching. On the other hand, teaching without assessment is unlikely to be effective, so a balance needs to be struck between teaching and assessment. Teachers need to make strategic choices with regard to the kind of authentic assessments they employ and with how many students each might be employed.

Group Discussions

An excellent way of enhancing class interactivity is to engage students in discussions relating to reading. If working with the whole class seems unlikely to be productive, start with a brief teacher-led session, then break the class into cooperative small groups, and then have the groups report back to a plenary. The focus can be simple to start with – what books have you read in the last 3 months which you really like? This can lead to many ancillary activities – displays of favorite books, charts of most liked books, deeper discussions of the most read books, suggestions for related reading, and so on and so forth. Once the class is somewhat practiced in the skills of working in this way, more contentious topics can be introduced. What things would really improve our reading enjoyment? Or ability? Helping students to realize that there are nearly as many different views as there are students – and that the teacher's job is not an easy one – are some of the side benefits.

Affective and Motivational Aspects

The emotional aspects of reading are enormously important if students are to become readers in adult life. Students may have preferences for nonfiction (information) books rather than fiction (story) books which are usually more common in classrooms (and more often used by elementary school class teachers). They may have idiosyncratic interests in particular areas, where they prove able to read and comprehend to a far higher level than in other books. They may prefer easy books which make no demands on their reading capability, or they may choose some books that are far too hard because they can get some sense from the pictures and it looks good to have them. Of course, they may be much more interested in newspapers, magazines, and the Internet than in books. Various interest and attitude surveys attempt to explore students'

feelings with regard to reading and books. These can be of written questions soliciting written answers, but there are also versions for emergent readers which require them to indicate on smiley and grumpy faces where their feelings lie. Of course, the extent to which students complete the survey honestly all the way through, might be questionable, and again teachers need to correlate the responses with observations of the students' demeanor when actually reading different kinds of texts.

Pre-Reading Sub-Skills

With emergent readers, teachers, sometimes, try to assess the component sub-skills which are assumed to underlie and contribute towards the development of reading skills. The most common of these are visual discrimination (the ability to discriminate between similar visual stimuli), visual sequential memory (the ability to remember visual stimuli in sequences of increasing length), and similarly, auditory discrimination and auditory sequential memory. The stimuli can be pictures, letter-like symbols or sounds, letter symbols or sounds, or written or spoken words (according to the abilities of the student). These assessments might indicate if a student is weak auditorily while strong visually, or vice versa – with implications for intervention. Or that discrimination is strong while memory is weak, likewise. However, a difficulty is that correspondence between the ability to complete the subtest and actually use these sub-skills in the much more complex act of reading is, often, quite weak. Consequently, teachers should also be preoccupied with the examination of the early reading attempts of these students, to see whether these sub-skills actually appear (or not) in real life.

Phonic Skills

Nearly half of the words in the English language are phonically regular, so phonic (or phonetic or phonological) skills are important. These are skills in sign-sound correspondence – the signs representing the elementary sounds of speech or expressing the pronunciation of words. Checking the phonic skills of readers may be important long after they are assumed to be independently capable. Student knowledge of lower-case letter-sound correspondences comes first, but the upper case (capital) letters must also be learned. Of course, student knowledge of the individual letter sounds needs to be checked, but also their awareness of phonic digraphs (two or more letters and sounds put together with effects different from what might be expected). These will include ee and oo; the long a, i and o with a final e; ar, er, or and all; sh, st, th, ss and ch; wh, bl, tr, fr, gr, fl, cl, gh, final y, ea; ck, ai, ay, oi, oy, ou, ow, oa, au, aw; igh, er, ur, ir, ew, ge, dge, air, are, ie, nk, ng, mb, ll, ff, tion, ation, ection, sion, ph, qu, kn, gn, wr, zz, and hard and soft c. As with other sub-skills (above), the extent

to which students can actually use phonic skills in the context of reading is potentially rather different from their ability to chant to a checklist; so teachers need to be aware when searching for examples of student use of phonic skills while reading.

Systematic Observation

It is difficult to capture observations that are made informally. Some teachers always have a clipboard or self-stick labels to hand out for this purpose. However, recording can easily get in the way of actively exploring the literacy incident further. Checklists of relevant reading skills might be helpful, but the long ones are very long (and thus, not very usable), while the short ones probably do not mention all the skills one is observing. It is, perhaps, better to look at checklists before observing – in order to refresh concepts of the wide range of reading skills – rather than actually use them while observing. Then, maybe, one can tick off skills following the event or simply record observations continuously and less formally. Remember that the teacher is only one person – others may have valid (and different) observations to make – classroom assistants, parents, and so on and so forth. Training classroom assistants to complete systematic reading observations may be a good use of their time.

Informal Reading Inventories

Informal reading inventories (IRIs) have been used for many decades. Teachers can devise their own or use commercially available IRIs. They have three graded reading paragraphs of different levels of difficulty: one at the level at which the student can read independently, one at the instructional level at which a few mistakes are made, and one at the frustration level – where many mistakes are made and the student cannot extract meaning from the text. The paragraphs for reading aloud are accompanied by comprehension questions. The independent level features 99% accurate word recognition and 90% correct comprehension, the instructional level 95–98% recognition and 75–89% comprehension, and the frustration level 90% and below recognition and 50% and below comprehension. On each passage, the student's performance is expressed in percentages and compared to these benchmarks. The grade level at which the student should be reading individually and with instruction becomes evident, together with any discrepancies between accuracy and comprehension.

Reading Miscue Inventories

Reading miscue inventories (RMIs) have less emphasis on determining the appropriate level for student reading and have a more elaborated analysis of miscues (as errors are now more kindly known, as something which is informative).

When teachers learn to do miscue analysis, it is, often, helpful if they audiotape or videotape an assessment session, and then compare their written analysis to the tape to make sure that they observed everything correctly.

Teachers make a copy of the passage chosen and write above each miscued word the nature of the miscue. Correct words should not be ticked as the child may see this and become preoccupied with teacher behavior. **Table 1**

Table 1 Miscue running record

Name of Child David L Name of Teacher Mrs P

Date _____ Words _____ Words correct _____ Words correct per minute _____

Title: The Adventures of Natividad and Toledo

SC= Self-Correction LAP=Looked at picture A=Asked T=Told . =pointed

Nat... T . . . includes SC . omitted to explore LAP mountain

Natividad has five children, including two sons who also like exploring in the mountains. His

Melk... LAP T Sal.... T

sons are called Melkiades (Melkie for short) and Salsedo (Salsie for short).

daughter SC Reena Lucy SC

Two of Nati's daughters (Reina and Lucinda) work in the family fields. Here they keep cows

SC

and sheep, and make excellent cheese.

Cla... A T bring SC oph... T

As well as his own children, Nati and his wife Claudia also brought up six other orphan

who omitted dead SC great... T

children whose own mother and father had died. Claudia and Nati now have 18 grandchildren

to keep them busy as well.

. . . omitted How... SC

When he was young, Nati first worked as a farmer. However, when he was about 35 years

counties SC omitted came

old, people from other countries began to come to explore the mountains. Nati helped them,

at addition lots equ... T poter gui T

first by carrying loads of equipment as a porter, and later as a mountain guide.

Tol... T company... omitted

Toledo the horse has been Nati's companion in the mountains for many years. At the end of a

omitted Toledo sid, sad...T horse swipey

long day, Nati takes off Toledo's saddle. The horse's back is hot and sweaty from the hard

shake....SC sal.. SC

work. Then Toledo likes to roll in the grass, to scratch his back and rub off the salt.

es..spec... T ic..itch...T omitted mountain ta..tas..T

Toledo especially likes to eat ichu grass, that grows only in the high mountains. It tastes very

for omitted itch.. SC to..T

good to horses, but it makes people all itchy, so Nati takes care not to touch it. When he is

usual omitted cal... ob..obdi.. T omitted stop

with Nati, Toledo is usually very calm and obedient. But when they have stopped to camp for

sometime like omitted ex.. explore

the night in the mountains, Toledo sometimes likes to go exploring by himself.

(Work out the total words, the correct words, the total self-corrections, and teacher interventions)

Table 2 Reading miscue checklist

1. Asks for help
2. Skips
3. Re-reads
4. Looks back
5. Adds word
6. Substitutes word
7. Omits word
8. Repeats word
9. Reverses word
10. Uses grapho-phonetic cues
11. Uses pictures/visual cues
12. Uses context
13. Uses background information
14. Recognizes miscues
15. Self-corrects miscues

Student performance can be coded thus: E = uses in a consistently Effective way, S = strategy used Sometimes effectively and sometimes ineffectively, I = uses strategy in an Ineffective way, N = Not observed to use the strategy. For more precision, the actual numbers of times students do each can be recorded.

shows a running record and **Table 2** shows a sample analysis. Teachers can write A if the child asked for the word, and T if they were told about it. A dot above a word indicates the child pointed. Interpretations from running records are, of course, subjective – the teacher constructs reasons for student error which may or may not be the actual ones. However, the running record can prove highly revelatory of the precise nature of each student's pattern of error – and every student's pattern will be different and have different implications for intervention.

Retellings

Retelling of a text is a powerful diagnostic tool for students whose abilities of comprehension are not good. The retelling can be audio-recorded for a more detailed analysis, including the ability to summarize, the level of retention of detail (main and supporting), the order of events, coherence, completeness, generalization beyond the text, and so on and so forth. Some students may be preoccupied with reporting facts – they should be encouraged to consider the wider message of the text. Other students may not be confident in their language abilities, and this can, erroneously, suggest poor comprehension. It is unwise to have the retelling in written form, since the act of writing interposes another variable to be considered.

Fluency

It has been said earlier (under Teacher Use of Tests) that the rate of reading is a somewhat doubtful indicator of competence in reading. So, how can we assess fluency, which is closely connected with the rate of reading? Fluency is an adaptive, context-dependent process that can

operate at a number of layers or levels. Even expert readers will show dysfluency when confronted with a text on an unfamiliar topic that provides a challenge which is beyond their independent reading level – however high that level might be. For silent reading, fluency can be defined as the extraction of maximum meaning at maximum speed in a relatively continuous flow, leaving spare simultaneous processing capacity for other higher-order processes. This assumes that the text is at an appropriate level of difficulty for the reader. For reading out loud, the task – and, therefore, the definition – is more demanding because, among the higher-order processes, the reader must have an awareness of audience needs and the capability to manage the prosodic demands for expressiveness (varying phrasing, stress, intonation, pitch, rhythm, loudness, pauses, etc.). Fluency can be considered as: (1) surface fluency (speed of accurate and automatic word recognition), (2) strategic fluency (control of speed of reading to yield comprehension and expression at the optimal level required for a specific purpose), and (3) deep fluency (control of speed of reading to maximize comprehension, expression, and deep reflection for specific purposes, enhancing explicit awareness and self-regulation of these processes). Deep fluency is tricky to assess. It can only be done in the context of systematic observation, RMIs, and self-assessment (above and below).

Portfolios

Portfolios give a longitudinal perspective on student performance and show a broader range of literacy achievements than other methods. Portfolios can also give greater depth on a particular piece of work or project. They, certainly, enable students to explore their reflections and feelings about what they are doing. They may include recordings of oral reading, student voluntary-reading logs, drawings pertaining to reading, book reports, journals, discourse during reading interviews, videotaping of book dramatizations, research projects, oral histories, parent questionnaires, and other material. Much of this can be stored electronically rather than in hard copy. Toward the end of a year, students can select and promote what they see their best work from their portfolio – this eases the assessment burden, since, otherwise, there is too much for a teacher to assess. Portfolios, often, have a section for out-of-school reading – otherwise easily overlooked. Portfolios can, of course, be taken home for parents to see and comment on longitudinal progress.

Computer-Aided Assessment of Real Books

At present, a number of software programs enable students to self-assess in relation to their comprehension of real books (e.g., Accelerated Reader, Reading Counts, and

Book Adventure). These require the student to read a real book of their choice independently, then – at the end – take a quiz on a computer to see if they understood the book. The basic quizzes are of literal comprehension in the interests of psychometric stability, but there are also quizzes requiring more inference which are less stable. The software gives feedback to the individual student (hopefully enhancing their motivation to read more and more carefully) and also to the teacher (who can see who is reading what, at what level and how effectively). Feedback is also available for parents. Students need to read a sufficient quantity of books at the right level and with the right degree of comprehension in order to make progress. Older students have been particularly found to read books at levels below their ability, resulting in a lack of challenge.

Self-Assessment

If assessment is to be formative as well as summative, the student's role in it must not be merely as a passive recipient. Students must be encouraged to actively engage in learning how to evaluate their own progress and do this better over time, since this reflects what will happen when they are adults. Self-assessment enables students to reflect and give feedback on whether teacher diagnosis is correct, as well as enhancing their ability to self-diagnose and become less dependent on teacher input. Increased student control of assessment is likely to enhance student self-confidence and self-regulation of their own learning processes – and these are important for reading beyond school – at present and in the future. Thus, self-assessment is actually a powerful intervention. Self-assessment should occur frequently, not be scheduled for occasional sessions. Students should choose what they want to evaluate (assuming they are not avoiding their problem areas). Teachers can help by giving information about what they think students are doing well. If students struggle, they should be provided with a short checklist of reading activities – as a basis for reflection. Which do they do better, which worse? **Table 3** gives a list of suggestions.

Peer Assessment

Of course, self-assessments may be highly subjective – more able students might overestimate their capabilities, while weaker students might underestimate theirs. Practice with teacher feedback is likely to improve accuracy. However, teachers have too much to do. Peer assessment has the advantage of offering some degree of external objectivity, albeit not nearly as competent as that of the teacher, but much more readily available. For peer assessment, pairs should be matched which are not too disparate

Table 3 What would help me read better?

Name _____

Date _____

Think about these items. Think about yourself as a reader.

Number the items in order from 1 to 10 to show which might be most helpful for you.

_____ Having someone listen to me reading and give me suggestions

_____ Working one to one with the teacher

_____ Having more time to practice reading in class

_____ Learning more strategies to help me understand what I read

_____ Helping tutor someone else become a better reader

_____ If I could get rewards for improvement

_____ Being able to work at my own speed

_____ Work should be more challenging (or less)

_____ More group work rather than doing everything individually

_____ Assignments should be more interesting

_____ I need to read more non-fiction (or fiction)

_____ I need to read more out of school

Please add anything you can think of that would help you become a better reader – do not worry if it is actually available just now.

in reading ability. Activities should be suggested, and simple checklists provided for students to assess the performance of other students. The pair might begin by discussing each other's self-assessment, to become familiar with the other's perceptions, and give them some sort of reality check. Beyond this, peers can discuss the selection of best work for portfolios, give feedback on retellings, render evaluative comments in group discussions, and discuss the motivational and affective aspects of their book readings. Maybe, they could even do more complex things.

Conclusion

Thus, a range of different measures, which teachers can use in the classroom, have been considered in this article. These are of various utilities and are also various in terms of the teacher time they consume. Consequently, teachers need to consider, very carefully, which measures they will use for which students. However, teachers also need to consider to what extent computers can be used to lighten their assessment load, or at least to place them as interpreters of richer information. Even more important is consideration of the extent to which students can be involved in their own self-assessment and peer assessment, since through these means, students can become more engaged in self-regulating their own performance and less dependent on teacher intervention. Certainly, these means for formative assessment of reading in the classroom lead more directly to intervention and are, thus, more important than external high-stakes or psychometric testing.

See also: Classroom Assessment Tasks and Tests; Formative Assessment; Impact of Assessments on Classroom Practice; Peer and Self-assessment; Portfolio Assessment.

Further Reading

- Bagnato, S. J. (2009). *Authentic Assessment for Early Childhood Intervention*. New York: Guilford Press.
- Bosco, D. and Buchner, J. (2004). *The Reading Portfolio*, 6th edn. Boston, MA: Houghton Mifflin.
- Bureau of Instructional Support and Community Services (2003). Florida Cool Tools Informal Reading Assessments. <http://reach.ucf.edu/~CENTRAL/Warehouse/Brochures-etc/Cool%20Tools%20-%20Elementary.pdf> (accessed August 2009).
- Caldwell, J. S. and Leslie, L. (2008). *Intervention Strategies to Follow Informal Reading Inventory Assessment: So What Do I Do Now?* 2nd edn. Boston, MA: Pearson Custom.
- Goodman, Y. M., Watson, D., and Burke, C. (2005). *Reading Miscue Inventory: From Evaluation to Instruction*, 2nd edn. Katonah, NY: Richard C. Owen.
- Greve, S. M. (2007). *Phonics for Dummies*. Hoboken: Wiley.
- Hebert, E. A. (2001). *The Power of Portfolios: What Children Can Teach Us about Learning and Assessment*. Hoboken: Jossey Bass.
- Jett-Simpson, M. and Leslie, L. (1999). *Authentic Literacy Assessment: An Ecological Approach*. Boston, MA: Allyn and Bacon.
- Keene, E. (2006). *Assessing Comprehension Thinking Strategies*. Huntington Beach, CA: Shell Educational Publishing.
- Mariotti, A. S. and Homan, S. P. (2005). *Linking Reading Assessment to Instruction*, 4th edn. London: Routledge.
- McAndrews, S. L. (2002). Enhancing reading strategies through teacher, peer, and self-assessment. *Illinois Reading Council Journal* 30(3), 32–41.
- McKenna, M. C. and Kear, D. J. (1990). Measuring attitude toward reading: A new tool for teachers. *The Reading Teacher* 43(8), 626–639.

- Rasinski, T. V. and Padak, N. P. (2008). *From Phonics to Fluency: Effective Teaching of Decoding and Reading Fluency in the Elementary School*, 2nd edn. Boston, MA: Allyn and Bacon.
- Topping, K. J. (1999). Formative assessment of reading comprehension by computer: Advantages and disadvantages of the Accelerated Reader software. *Reading OnLine*. (I.R.A.). <http://www.readingonline.org/critical/topping> (accessed August 2009).
- Topping, K. J., Samuels, J., and Paul, T. (2007). Does practice make perfect? Independent reading quantity, quality and student achievement. *Learning and Instruction* 17(3), 253–264.

Relevant Websites

- <http://en.wikipedia.org> – Authentic Assessment – Fluency, Wikipedia.
- <http://en.wikipedia.org> – Authentic Assessment, Wikipedia.
- <http://en.wikipedia.org> – Career Portfolios, Wikipedia.
- <http://www.familylearning.org.uk> – Family Learning: Pre-Reading Skills, Other Sub-Skills.
- <http://lrs.ed.uiuc.edu> – Informal Reading Inventories, Technology Studies in Education.
- <http://En.Wikipedia.Org> – Miscue Analysis, Wikipedia.
- <http://www.nclrc.org> – Peer and Self Assessment, National Capital Language Resource Center.
- <http://en.wikipedia.org> – Phonic Sub-Skills, Wikipedia.
- <http://en.wikipedia.org> – Reading, Wikipedia.
- <http://www.sedl.org> – Reading Assessment Database, SEDL Reading Resources.
- <http://www.nwrel.org> – Reading Miscue Inventories, Creating Eager Readers.
- <http://www.readingrecovery.org> – Reading Recovery: Systematic Observation.
- <http://www.readinga-z.com> – Retellings, Reading A–Z.
- <http://en.wikipedia.org> – Self Assessments, Wikipedia.

Impact of Assessment on Students' Learning Strategies and Implications for Judging Assessment Quality

M Segers, University of Leiden, Leiden, The Netherlands

F Dochy, University of Leuven, Leuven, Belgium

D Gijbels, University of Antwerp, Antwerpen, Belgium

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Authenticity – This criterion is meant to convey that assessment assignments need to include all relevant knowledge, skills and attitudes that are also needed to operate in realistic contexts.

Consequential validity – Refers to the effect of assessment on learning and instruction.

Fairness – This implies the absence of bias towards certain groups and the exclusion of irrelevant variance in the assessment process.

Generalizability – This describes the extent to which tasks and performances cover a broader 'universe' of similar tasks that measure the same content or task performance.

Post-assessment effect – Assessment influences student learning by looking back at the learning process and the learning product after the completion of the assessment task. It refers to the backwash effect of assessment on learning.

Pre-assessment effect – The influence of assessment on student learning *before* a student has taken the assessment. It refers to the pro-active behaviour of students, based on their expectations of what the assessment will look like and what will be required from them.

Pure assessment effect – Assessment itself becomes a rich learning experience for students. By working on assessment assignments, students develop their thinking and understanding of the subject, supported by the continuous loop of feedback and feed-forward of teacher and/or peers.

Systematic validity – 'Evolutions in the form and content of instruction and students' learning engendered by use of the test' (Frederiksen and Collins, 1989, p. 28).

Transparency – This criterion indicates that the assessment process itself needs to be clear and comprehensible to all participants.

Introduction

As far back as the early 1970s, researchers were engaged in studies of student learning at prestigious universities (e.g., Snyder, 1971). They found, unexpectedly, that what influenced students most was not the method of teaching, but the method of assessment. Students described all aspects of their studies – what classes they attended, how much work they did, and how they went about studying – as being completely dominated by the way in which they perceived the demands of the assessment. Snyder's work particularly gave birth to the notion of the hidden curriculum – the different and unwritten curriculum that students had to discover and pay attention to if they wanted to succeed.

During the past 20 years, the issue of the positive or negative impact of testing and assessment on learning, teaching, and achievement has come very much to the fore. Findings of a number of studies (e.g., Ramsden, 1992) have demonstrated that assessment that is perceived as threatening and that provokes anxiety may push students to employ surface learning approaches. This problem is not new. Already, in 1984, Frederiksen was claiming that educators have to concentrate on developing assessment instruments that improve the educational process instead of damaging it (Frederiksen, 1984). In this regard, Frederiksen and Collins (1989) introduced the term systematic validity, referring to "evolutions in the form and content of instruction and students' learning engendered by use of the test" (p.28). Since the 1990s, the concept of consequential validity has been commonly used to refer to the effect of assessment on learning and instruction.

Over the past decade, many classroom teachers have experimented with the implementation of so-called new modes of assessment, such as peer assessment, portfolio assessment, and case-based assessment. Educators expected that – compared to conventional testing – the effects on student learning would be more positive. More precisely, it was theorized that using modes of assessment that required students to use higher-order thinking skills – such as problem-solving and reflection – would stimulate students to approach their learning in a deep, instead of a surface, way. This would, in turn, lead to significant learning gains.

The studies of Scouller (1995, 1998), Scouller and Prosser (1994), and Tang (1994) were the first to compare traditional multiple-choice (MC) tests – asking students to reproduce learned information – with an essay test and portfolio assessment. The authors concluded that the essay test and portfolio assessment were perceived by the students as assessing on a deep level and were related to deep learning approaches; the reproduction-oriented MC tests were perceived as measuring on a surface level and were related to surface learning approaches. Since then, the effects of assessment – and, more specifically, of new modes of assessment – have been a subject of discussion. Although the discussion on the existence and proof of effects is interesting for educators, far more interesting is the question of what conditions are necessary for assessment to support student learning. To date, the evidence for any conclusive answer to this question is still very scarce. This article summarizes a few recent studies addressing this issue.

A distinction is made among three kinds of effects: pre-, post-, and pure assessment effects. Pre-assessment effects refer to the influence of assessment on student learning before a student has taken the assessment. It refers to the proactive behavior of students – based on their expectations of what the assessment will look like and what will be required from them. The post-assessment effect refers to the backwash effect of assessment on learning. How does having undergone the experience of assessment practices influence student learning behavior and products? Post-assessment effects can be formulated in terms of motivation (e.g., the influence of assessment on student self-efficacy), in terms of learning approach (e.g., how does having long-term experience with knowledge-reproduction tests influence student learning approaches?), as well as in terms of cognitive gains (e.g., to what extent does the feedback received following the assessment support or deepen understanding of the learning content?). Finally, it might be expected that, by working on assessment assignments, students develop their thinking and understanding of the subject – supported by the continuous loop of feedback and feedforward of teacher and/or peers. This is called the pure assessment effect. Project-based assessment is an example where students have ample opportunity to learn from working on the assessment assignment itself (the project).

Evidencing Pre-Assessment Effects

It has been argued that assessment can have an effect on different levels which depends on the function of the assessment (summative vs. formative) (Gielen *et al.*, 2003). The influence of summative assessment on learning behavior is

mainly proactive. The question “do we have to know this for the examination?” will be recognizable for nearly every teacher and illustrates that students tend to adjust their learning behavior to what they expect to be assessed. These effects can be described as pre-assessment effects, since the effects occur prior to the assessment. An example of an empirical study on the pre-assessment effect of new modes of assessment is the study by Segers *et al.* (2006). These researchers explored the conditions for assessment to steer learning – by investigating the impact of the implementation of a new assessment instrument (the OverAll Test) on students' learning strategies. The OverAll Test is a case-based assessment instrument which measures the extent to which students are able to use knowledge (models, theories, etc.) to define, analyze, and solve authentic problems. The OverAll Test is implemented as part of a course redesign and in alignment with the features of the redesigned learning environment. The main differences between the original and the redesigned learning environment are the format of learning tasks (study tasks in the original course and problem tasks in the redesigned course) and the mode of assessment (a knowledge-reproduction test in the original course and the OverAll Test in the redesigned course). When redesigning the course, it was expected that, by implementing the OverAll Test, the students would change their learning strategies. More specifically – given the features of the OverAll Test – it was expected they would adopt more deep learning strategies in the OverAll Test condition than in the original condition. In order to unravel the mechanism through which the OverAll Test influences students' learning, students' intended learning strategies as an indicator of their habitual learning strategies were addressed, as well as their perceptions of the assessment's demands. In addition, students' learning strategies were measured. The results indicated that – in contradiction with expectations – the students in the original course adopted more deep learning strategies and fewer surface learning strategies than the students in the redesigned course. Although the students were informed concerning the differences between both courses in the various course information resources, there were no significant differences between both groups of students in the learning strategies they intended to employ as well as in their perceptions of the assessment's demands.

In both conditions, the students – who intended to employ surface learning strategies – perceived the assessment's demands as surface and actually used surface approaches to learning. The results showed, clearly, that those students who expressed their intentions to employ a certain learning strategy perceived the assessment demands as such and actually employed a related learning strategy. In short, the change in assessment practice did

result in a change in learning strategy; however, this was toward more surface learning. Lack of familiarity with this mode of assessment might have stimulated them to go back to old habits, using the learning strategies they had experienced as being successful in the past. Students seem to be more stable than expected in the learning strategy they employ and perceived the assessment's demands accordingly.

Evidencing Post-Assessment Effects

Assessment influences student learning by looking back at the learning process and the learning product following the completion of the assessment task. This is referred to as post-assessment effects. When teachers give students information with regard to the quality of their performance and support students in reflecting on the learning outcomes and the learning processes they are based on, gains in terms of achievement and learning approaches can be expected. When students dispose the necessary metacognitive knowledge and skills, teacher feedback can be reduced. Students may become capable enough to draw conclusions themselves with regard to the quality of their learning behavior (self-generating feedback or internal feedback), following the completion of the assessment task. It is clear that, especially when assessment tasks are used for formative purposes, the feedback supports students in diagnosing their own learning and in taking the next steps in their development. However – in comparison with formative assessment – post-assessment effects of summative assessment on student learning are small. Moreover, it is evidenced that the effects of testing on the long term are negative for student motivation (Harlen and Crick, 2003). Regular exposure to conventional testing leads to significant motivational problems.

A recent study by Segers *et al.* (2008) illustrates the post-assessment effect of portfolio assessment, or more precisely how portfolio assessment influences student learning.

This study was conducted within a competency-based program on applied sciences, with portfolio assessment as its core mode of assessment. Based on an intake assessment, students developed a personal development plan (PDP) and a personal activity plan (PAP). Within the PDP, the student reflected on his/her personal competence development in relation to the levels and behavioral indicators described in the student-competencies guide. It formed the basis for coaching interviews with the study coach. The PAP consisted of the student's concrete working plan; the activities he/she would undertake to develop the competencies at the required level. The digital portfolio consisted of the PDP and PAP of the student, his/her study dossier (the work – such as project reports, project products, and performance assessments – that he/she was actually doing and had finished), and a private

dossier with documents the students wanted to keep for personal purposes. The development of the digital portfolio was supported by a study coach who had regular meetings with students (individually and in small groups). There was feedback on the work done and the progression shown, discussions were held about which competencies, at what level, and which specific behavioral indicators that a student will work on in the next step. It was expected that reflection and feedback – as part of the portfolio process – would result in students adopting a deep approach to learning. More specifically, the researchers were interested in unraveling the elements of the portfolio-assessment practice that were related to how students approached their learning. Therefore – at the end of the academic year – students' perceptions of these assessment practices (after experiencing the portfolio assessment) and how they approached their learning (deep and surface approaches to learning) were measured. The results indicated that students who adopted a deep approach to learning read the feedback carefully and used it to look critically at their portfolio, then tried to improve it based on the suggestions made. Additionally, it was found that students who adopted a deep learning approach perceived the portfolio assessment as stimulating their learning, being motivational, and requiring a deep understanding of the learning materials; they perceived the criteria for the portfolio to be clear and perceived the portfolio assessment as challenging.

In short, these findings suggest that how the students went about their learning was influenced by how they had experienced different features of the portfolio-assessment practice.

Evidencing Pure Assessment Effects

Nevo (1995) and Struyf *et al.* (2001) point to a third kind of learning effect from assessment. The main idea is that – when assessment is part of daily classroom pedagogy – there is an ongoing feedback and feedforward loop informing students concerning their growth and the next steps to take. This means that assessment itself becomes a rich learning experience for students. This means that – contrary to the old American saying “weighing the pig doesn't fatten it” – in some cases, high-quality learning is fostered by simply assessing the learning outcomes, or “weighing can fatten the pigs.” This can be called the pure assessment effect. Illustrative is the study of Gijbels *et al.* (2005). This study investigated the effect of integrating written assessment tasks into the learning environment on students' performance in final exams. The study can be seen as a small, but progressive, step toward the integration of assessment, learning, and instruction. In order to gain more insight into the effects, the researchers formulated two research questions. First, did students

who did the assessment tasks perform better in their final exams compared to students who did not? Secondly, what were the most important concerns in students' and teachers' perceptions of the assessment tasks? Students' final exam results were used to find out whether students who did the assessment tasks performed better than students who did not. Answers from questionnaires and semi-structured interviews were used to discover the most important concerns in students' and teachers' perceptions of the assessment tasks.

The research was carried out within the context of a compulsory second-year law course. A total of six assessment tasks were distributed over different topics and weeks in the course. Assessment on the course was two-fold. The final exam consisted of 40 MC questions and took place at the end of the course. During the course, students had the opportunity to complete the six assessment tasks on a voluntary basis, which could result in an extra bonus point – added to the score of the final exam. Both the assessment tasks and the MC questions set several cognitive problems to the students in line with the instructional goals of problem-based learning.

Students were stimulated to produce high-quality learning activities by giving the one extra bonus-point only if all six assessment tasks showed sufficient quality and effort. The feedback students got from their tutor and from the plenary discussion in the tutorial group could help them to do their next assessment task better and to get a better understanding of the learning materials to be studied in order to pass the final exam at the end of the course.

The results of the study indicated that working with assessment tasks had a significant positive influence on the students' performance – not only on assessment task-related topics, but also on nonrelated topics. These results indicated that the introduction of the assessment task helped students to address more appropriate student learning activities – going beyond the six assessment tasks and their content. This was confirmed by the results of the survey, indicating that students who took on the integrated assessment tasks worked in a different way – in terms of learning approach as well as time management. A crucial condition for the integrated assessment tasks to affect learning seemed to be the feedback students received – the clarity of goals, the criteria of the assessment tasks, and the implementation of the tasks.

Quality Indicators of Assessment

The aforementioned studies illustrate how classroom assessment practices influence how students go about learning. However, various quality aspects of the assessment practice might hinder or enhance the consequential validity of the assessment. For example, portfolio assessment might not lead to the expected reflection and

growth in competence development when the procedures are not transparent. Moreover, when portfolio assessment is mainly based on written artifacts and written reflections, English as a Second Language (ESL) learners might be disadvantaged. In the case of project-based assessment, the learning effects might be jeopardized by using unrealistic projects or projects that require students to only perform routine procedures and tasks to reach the project goals.

These examples illustrate the importance of reflecting on quality indicators for current classroom assessment practices. There has been an ongoing debate on the relevance of psychometric quality indicators. During the past decade, our thinking about assessment as well as classroom assessment practices has been changing. In general, these changes are:

- There is a stronger emphasis on the integration of assessment and instruction. Many assessment specialists take the position that appropriately used educational assessments can be seen as tools that enhance the instructional process as well as the student learning process.
- The position of the student is increasingly that of an active participant who shares responsibility for the process, practices self-assessment, reflection, and collaboration, and conducts a continuous dialog with the teacher. Students participate in the development of the criteria and the standards for evaluating their performance.
- Both the product and process are being assessed.
- The assessment takes many forms, all of which could generally be referred to as unstandardized assessments embedded in instruction. There is, often, no time pressure, and a variety of tools that are used in real life for performing similar tasks are permitted.
- The assessment tasks are more often interesting, meaningful, authentic, challenging, and engaging, involving investigations of various kinds.

The question is whether the traditional psychometric quality indicators align with the aforementioned features of assessment. Various authors have proposed replacing or extending the traditional quality criteria of reliability and validity. To illustrate the quality criteria that are argued to align with current assessment practices, the quality criteria as expressed in the literature on peer assessment are described as follows:

- *Authenticity*. This criterion is meant to convey that assessment assignments need to include all relevant knowledge, skills, and attitudes that are also needed to operate in realistic contexts. It refers to the requirement that assignments must measure how students learn actively in settings that resemble reality itself. In the literature referring to authenticity, four alternatives – or

for that matter underlying or subsumed criteria – are to be found. First, representativeness refers to the criterion or requirement that assessment tasks must do the reality domain at stake. Second, the criterion meaningfulness stresses the importance of assessing relevant and worthwhile contributions of the learner. Third, the criterion cognitive complexity involves inclusion of higher-order skills. A fourth criterion – content coverage – is meant to highlight the breadth of inclusion of the assessment domain.

- *Transparency*. This criterion indicates that the assessment process, itself, needs to be clear and comprehensible to all participants. Only then can it be expected that assessment has the potential to support student learning.
- *Fairness*. This implies the absence of bias toward certain groups and the exclusion of irrelevant variance in the assessment process. In essence, this criterion indicates that each student must receive the same chance to demonstrate their knowledge, skills, abilities, and competencies.
- *Generalizability*. This describes the extent to which tasks and performances cover a broader universe of similar tasks that measure the same content or task performance. Under generalizability, three more specified criteria are to be found: Comparability is defined as measurement under same conditions and criteria for everyone. Reproducibility is defined as the accuracy and consistency of decisions over time and assessors. Transferability refers to the extrapolation of measured performance to other situations and tasks.

Conclusions

During the past decade, the world of assessment in education has been dominated by a plea for the implementation of an assessment culture where learning in its various aspects, as well as the learner, should be central in the process. As part of this assessment culture, so-called ‘new modes of assessment’ have been widely implemented, in general, being characterized by measuring higher-order thinking skills and competencies. Examples of this are peer assessment, performance assessment, and portfolio assessment. Studies have been conducted to better understand the hypothesized positive effects of new modes of assessment on student learning: pre-assessment effects, post-assessment effects, and pure assessment effects.

Moreover, organizing assessment to positively influence student learning demands critical monitoring of the quality of our assessment practices. Several authors have argued that new forms of assessment require new quality criteria. When we want to make assessment practices supportive for learning, attention should be paid to quality

criteria, such as authenticity, transparency, fairness, and generalizability. The research agenda of the coming decade will be dominated by the search for empirical evidence, not only for the right conditions of assessment to support learning, but also for valid quality criteria to enhance the quality of these modes of assessment.

See also: Assessment Practice in Policy Context: Latin American Countries; Cultural Issues that Can Affect the Validity of Educational Evaluations; Evaluation Use; Impact of Assessment on Learner Groups (Boys/Girls); Impact of Assessment on Students’ Test Anxiety; Impact of Assessments on Classroom Practice.

Bibliography

- Frederiksen, J. R. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist* **39**(3), 193–202.
- Frederiksen, J. R. and Collins, A. (1989). A system approach to educational testing. *Educational Researcher* **18**(9), 27–32.
- Gielen, S., Dochy, F., and Dierick, S. (2003). Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post-, and true assessment effects. In Segers, M., Dochy, F., and Cascallar, E. (eds.) *Optimizing New Modes of Assessment: In Search of Qualities and Standards*, pp 37–54. Dordrecht: KluwerAcademic.
- Gijbels, D., van de Watering, G., and Dochy, F. (2005). Integrating assessment tasks in a problem-based learning environment. *Assessment and Evaluation in Higher Education* **30**(1), 73–86.
- Harlen, W. and Crick, R. D. (2003). Testing and motivation for learning. *Assessment in Education* **10**(2), 169–207.
- Nevo, D. (1995). *School-Based Evaluation: A Dialogue for School Improvement*. London: Pergamon.
- Ramsden, P. (1992). *Learning to Teach in Higher Education*. London: Routledge.
- Scouller, K. M. (1995). Different learning approaches of undergraduate students in two assessment contexts. *Electronic Disc 3, Paper 89, Research and Development in Higher Education* **17**.
- Scouller, K. M. (1998). The influence of assessment method on students’ learning approaches: Multiple choice question examination versus assignment essay. *Higher Education* **4**, 453–472.
- Scouller, K. M. and Prosser, M. (1994). Students’ experiences in studying for multiple choice question examinations. *Studies in Higher Education* **19**, 267–279.
- Segers, M., Gijbels, D., and Thurlings, M. (2008). The relationship between students’ perceptions of portfolio assessment practice and their approaches to learning. *Educational Studies* **34**(1), 35–44.
- Segers, M., Nijhuis, J., and Gijbels, W. (2006). Redesigning a learning and assessment environment: The influence on students’ perceptions of assessment demands and their learning strategies. *Studies in Educational Evaluation* **32**(3), 223–242.
- Snyder, B. R. (1971). *The Hidden Curriculum*. Cambridge, MA: MIT Press.
- Struyf, E., Vandenbergh, R., and Lens, W. (2001). The evaluation practice of teachers as a learning opportunity for students. *Studies in Educational Evaluation* **27**(3), 215–238.
- Tang, C. (1994). Effects of modes of assessment on students’ preparation strategies. In Gibbs, G. (ed.) *Improving Student Learning: Theory and Practice*, pp 151–170. Oxford: Oxford Centre for Staff Development <http://teaching.polyu.edu.hk/datafiles/R126.pdf> (accessed September 2009).

Further Reading

- Black, P. and William, D. (1998). Assessment and classroom learning. *Assessment in Education* **5**, 7–74.
- Crooks, T. (1988). The impact of classroom evaluations on students. *Review of Educational Research* **58**(4), 438–481.
- Segers, M., Dochy, F., and Cascallar, E. (eds.) (2003). *Optimising New Modes of Assessment: In Search of Qualities and Standards*. Dordrecht: Kluwer Academic Publishers.
- Shephard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher* **29**(7), 4–14.

Relevant Websites

- <http://www.assessment-reform-group.org> – Assessment Reform Group.
- <http://www.iml.uts.edu.au> – Key Assessment Elements, Institute for Interactive Media and Learning (IML), University of Technology, Sydney.

EDUCATIONAL ASSESSMENT – ASSESSMENT ACTIVITIES

Contents

Alternative Assessment
Classroom Assessment Tasks and Tests
Dynamic Assessment
Instructional System Provided Feedback
Peer and Self-assessment
Portfolio Assessment
Records of Achievement: Beyond Traditional Tests

Alternative Assessment

G Stobart, Institute of Education, University of London, London, UK
C Gipps, University of Wolverhampton, Wolverhampton, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Dynamic assessment – A form of diagnostic assessment which looks at what progress a learner can make when offered adult support in a one-to-one situation.

Educational assessment – An approach which focuses on an individual achievement relative to him or herself rather than to others, places little emphasis on standardization and seeks to help the individual progress. It is contrasted with psychometric testing.

Formative assessment – Assessment used to inform teaching and learning so that further learning takes place.

International Baccalaureate (IB) – A high status international diploma offered in three languages around the world.

Performance assessment – Judging what students can do in terms of the actual tasks and end performances that are the goals of instruction. (Used interchangeably with 'authentic assessments' and 'direct assessments'.)

Psychometric testing – Assessment which focuses on standardization of tasks and scoring, on norms and distributions.

Realistic simulation – Seeking to represent the intended real-world performance, for example, a flight simulator for trainee pilots.

Social constructivist – Cognitive and sociocultural theories of learning which emphasize the

importance of students actively making meaning in their learning.

Summative assessment – Assessment which 'sums up' where learners have got to in their learning.

Systemically valid testing – Assessment which leads to curriculum and instructional changes that foster the cognitive skills that the test measures.

Definitions

Alternative assessment is an elastic term which stretches from doing the same thing in a different format to describing radically different ways of approaching assessment. These include performance, authentic, formative, and diagnostic assessments which reflect alternative approaches to learning. A function of this contribution is to clarify, and question, the way the term is used.

As part of clarifying usage, we classify alternative assessments into three levels: alternative forms of assessment, which we will call alternative formats; alternative models of assessment; and alternative purposes of assessment, where purpose can drive form. While there are considerable overlaps among these, we believe that the distinctions are useful in discussing the claims that are made. Examples are considered from each of these levels.

Level 1: Alternative Formats

The alternative label is usually applied at this level to new or different delivery of an assessment – for example, transferring a pencil-and-paper test to a computer screen. The test content and demands remain much the same while the means of delivery changes. It may be the case that as the computer testing becomes more sophisticated, it begins to make new demands and might move to the next level – for example, using multimodal stimulus materials or forms of computer-adaptive testing.

Where a particular form of assessment is dominant – for example, multiple-choice testing in the USA – then, more open-ended tests requiring constructed answers have also been claimed to be alternative, in the sense that they are different. It may be that, once again, little else changes in the test demands and accompanying pedagogy. We would challenge whether this kind of claim should automatically be called alternative, since it is essentially business-as-usual with varied assessment formats. Little changes in what is assessed and in classroom teaching and learning.

Level 2: Alternative Models of Assessment

At this level, changes in assessment are underpinned by the intention to provide a different approach to assessment to reflect a different approach to learning and pedagogy. We illustrate this through developments in performance assessment and through approaches resulting from social constructivist views of learning. Performance assessment emphasizes direct, or as direct as possible, demonstration of the skill being assessed and minimizes the use of tests as a proxy for actual performance. This is a long-standing tradition in many occupational areas, and we consider moves to apply this within more general education. We use the term social constructivist loosely, to cover those cognitive and sociocultural theories of learning which emphasize the importance of students actively making meaning in their learning. These approaches have encouraged a more holistic approach to assessment than the knowledge-in-bits of most multiple-choice tests – underpinned for many years by more behaviorist views of learning.

Along with new conceptions of learning, the drive to develop alternative models of assessment received support from the fact that traditional types of test (multiple-choice standardized tests in the US and formal written examinations in the UK) have had unintended and negative effects on teaching and on the curriculum. “Under intense political pressure, test scores are likely to go up without a corresponding improvement in student

learning. In fact, distortions in what and how students are taught may actually decrease students’ conceptual understanding” (Shepard, 2000: 13).

Performance Assessment

Shepard and Bleim (1995) have argued for the equivalence of some of the key terms: “We use the terms performance assessments, authentic assessments and direct assessments interchangeably, the idea being to judge what students can do in terms of the actual tasks and end performances that are the goals of instruction” (p. 25). While this equivalence may not satisfy all, it provides us with a broad base from which to work. The spirit of this approach is to offer assessments which relate as closely as possible to the real-life knowledge and skills required. How close this match is may depend on the context. The philosophy behind this approach is straightforward: to be assessed as having a skill or competency we need to demonstrate it. This seeks to counter what Hanson has called “fabricating quality of tests” in which “the likelihood that someone will be able to do something, as determined by the tests, becomes more important than one’s actually doing it” (1994: 288).

We begin by briefly considering two well-established forms of performance assessment. Both have their roots in occupational and vocational training – where the emphasis is on practical demonstrations of competence. With pencil-and-paper testing so dominant in the education sector, these serve as a reminder of models of assessment which predate standardized testing.

Competency assessment in the workplace

The most direct correspondence would be competency assessments in the workplace; we would expect our trainee doctors, plumbers, and pilots to have demonstrated competency in real-life situations. We would require something similar of musicians and drama students; they will be assessed on the basis of actual performances. A key feature of these assessments is that they rely essentially on expert human judgment. This is a distinct difference from psychometric approaches which seek to minimize judgment and in which competency is often reduced to achieving the cut-score on multiple-choice tests. Work-based assessment has a distinguished history and predates our use of written assessments. It was embodied in the medieval guild system, in which persons would be apprenticed and would then – after learning their trade for a number of years – demonstrate the skills in their craft through their products. It was the quality of these products that determined their admittance to the guild.

Realistic simulation

This approach also seeks to align the assessment closely with the intended real-world performance, but may take

place in a context which relies on simulation – for example in a college. So medical training may involve direct simulation – for example, working with actors who mimic patients’ symptoms. Similarly, trainee pilots will use flight simulators and trainee mechanics work on workshop engines. Another form of realistic simulation may be working under highly controlled and supervised conditions – for example, special catering events for trainee chefs, or conducting scientific experiments in school science laboratories. We consider these alternative assessments because they represent a model of learning-by-doing which differs from the dominant educational model in which tests become a proxy for the required skills and practical knowledge.

While we categorized early computer-based testing as little more than changing the format of the tests (paper to screen), many of the more sophisticated applications are able to offer realistic simulations through multimedia stimulus materials. We also recognize that adaptive testing – which responds to the student’s choices and decisions – offers an alternative when, for example, it works to capture the student’s own understanding of a problem.

Alternative Educational Assessment

Here, we consider examples of school-based assessment which are based on models of learning and assessment which challenge some of the assumptions on which conventional standardized assessments are based. Some of these may be loosely described as performance assessments since they focus on what learners can demonstrate in less constrained settings than conventional standardized tests. These are often summative in purpose, though they contain many opportunities for formative assessment (see below).

The philosophy behind these alternative approaches distinguishes educational assessment from psychometric assessment. For present purposes, the key issues are that – in the psychometric model – the focus on reliability requires the standardization of administration and tasks as well as of scoring. The preoccupation of psychometrics with norms and norm-referenced testing stemmed from tests focused on aptitude, selection, and prediction. An individual’s performance was reported in relation to that of his/her peers, so that performance was seen in relative, rather than in absolute, terms.

In contrast, Wood (1986) has argued that educational assessment deals with the individual’s achievement relative to him or herself rather than to others; takes place in relatively uncontrolled conditions and so does not produce well-behaved data; looks for best – rather than typical – performances; is most effective when rules and regulations characteristic of standardized testing are relaxed; and embodies a constructive outlook on assessment, where the aim is to help rather than sentence the individual.

Shepard (2000) has developed a conceptual framework for this understanding of assessment based on social constructivist theories of learning. She summarizes this view of learning in the following principles: (1) intellectual abilities are socially and culturally developed; (2) learners construct knowledge and understandings within a social context; (3) new learning is shaped by prior knowledge and cultural perspectives; (4) intelligent thought involves metacognition or self-monitoring of learning and thinking; (5) deep understanding is principled and supports transfer; and (6) cognitive performance depends on dispositions and personal identity.

Educational assessment seeks to realize these principles through the kind of assessment tasks presented to students. These require a greater emphasis on broader-based performance tasks which encourage a deeper understanding. Frederiksen and Collins’ (1989) have approached this through their principles of systemically valid testing. At the heart of this approach is the principle that:

A systemically valid test is one that induces in the education system curricular and instructional changes that foster the development of the cognitive skills that the test is designed to measure. Evidence for systemic validity would be an improvement in those skills after the test has been in place within the education system for a period of time. (p. 27)

They proposed standards for assessment which included:

Directness. This involves assessing the cognitive skill of interest – the emphasis being on the authenticity of the extended tasks. They call for sets of tasks which are ecologically valid, which are representative of the way knowledge and skills are used in real-world contexts, and for which the primary traits have been identified.

Scope. This considers the range of skills needed to do well in the tasks.

Reliability. This seeks effective ways of assessing which, at the same time, fosters learning.

Transparency. This is the concern that those being assessed are clear about how they are being judged. This should enable learners to “assess themselves and others with almost the same reliability as the actual test evaluators achieve” (p.30).

In practice, this approach has been expressed through approaches such as portfolio-based assessment, extended assignments, and assessment by teachers which contribute to credentials. We briefly describe examples of each of these.

Portfolio assessment

At one level, portfolios may be regarded simply as collections of pieces of student work which are kept as a record. However, these collections can also be used as a basis for

assessment, and so portfolios are often described as a method of assessment.

A portfolio used for assessment and learning purposes involves documentation of achievements, self-evaluation, process artifacts, and analyses of learning experiences. It is, therefore, significantly more than a collection of assignments (Klenowski, 2000: 219).

Using collections of student work for assessment allows a wide range of approaches to marking – from traditional marks and grades assigned by teachers, through detailed written comments by teachers (with no grades), teacher–student discussion, and negotiation of the assessment, to student self-evaluation. As with other forms of alternative assessment, the proponents of portfolio assessment present a rationale based on the enhancement of learning. First, assessment that celebrates achievement on a self-referenced or criterion-referenced basis builds students' confidence and ultimately enhances their performance. These can be contrasted with the situation in which assessment draws attention to failure through the use of grades and marks, particularly when norm-referenced. Second, learning is enhanced if students understand clearly what they are trying to achieve, and are provided with personal and substantive feedback in trying to reach their target. Third, collecting, sorting, and annotating evidence to demonstrate achievement in the form of a portfolio of work is a valuable aid in developing the skills of self-evaluation, as well as for communicating standards in a relatively unambiguous and accessible way.

Advocates see portfolio assessment as the embodiment of good social constructivist principles. While the validity of using these for summative purposes at a local level is strong, there have been far more challenges when portfolios have been used in relation to wide-scale accountability testing. Portfolio experiments in the USA, in the states of Vermont and Kentucky, led them to be incorporated into accountability measures, which then raised questions of reliability. For high-stakes purposes, portfolio assessments remain vulnerable to reliability problems relating to sampling, standardization, and scoring. In high-stakes assessment contexts, this will often leave them exposed to criticisms of bias and unreliability, particularly where it is the student's own teacher who is making the assessment.

Extended assignments/projects

In standardized tests, even the inclusion of constructed (open-ended) answers offer the candidates little room to express their own understandings since they generally require convergent answers which are anticipated by the mark scheme. An alternative approach, which sits well with constructivist principles, is to provide the learner with opportunities to create a substantial piece of work in which they have choice over what they do. At classroom level, this may be projects which explore local issues – for

example, history projects, or which feature the students' own designs in technology. This may also incorporate group activities which would encourage the collaborative aspects of learning.

As with portfolios, it is possible to construct a strong case for the validity of such approaches when the summative purposes are restricted to local use. Their use in high-stakes assessments will again raise issues of consistency in their assessment and the problems of equating performance on very different tasks. In England, the use of such projects – assessed by teachers – has been a part of external qualifications such as the national examinations at ages 16 and 18. However, reliability concerns have led to the progressive erosion of the individual's choice of projects; so in subjects such as mathematics and science, the tasks were externally defined. This, in turn, led to such predictability about what was needed for a good grade that model solutions were easily available on the Internet. This in turn caused a further tightening of the specifications in the form of controlled assessments. These have to be done largely in the classroom under teacher supervision – giving students very little choice or autonomy in their work.

A more constructive example of how individual projects can be managed is provided by the extended essay in the International Baccalaureate (IB) – a high-status international diploma for 16–18-year-olds offered in three languages around the world. The extended essay is an independent, self-directed piece of research, culminating in a 4000-word paper which contributes to the diploma. It involves an in-depth study of a topic of interest within a chosen subject, with the emphasis placed on the research process: formulating an appropriate research question; engaging in a personal exploration of the topic; communicating ideas; and developing an argument. The assessment of this process is by the students' own teachers, who have received professional development in marking the essays. There is considerable external confidence in the dependability of IB assessments, especially among universities for whom the extended essay is a good preparation for similar work at degree level.

Assessment by teachers for external purposes

Teachers regularly assess their students' classroom learning for both formative and summative purposes. Our interest here is in those assessments which are used as part of external certification or for high-stakes selection purposes. Such assessments can be considered as alternative assessments when they both replace conventional standardized testing and involve the principles of educational assessment we have previously outlined. Thus, we would look for different approaches to teaching and learning that educational assessment encourages – Frederiksen and Collins' systemic validity.

Much classroom assessment may not meet these requirements since it mimics standardized tests and is little more than frequent summative testing which is used for reporting or management purposes rather than as formatively contributing to students' further learning. Continuous assessments such as the grade point average (GPA) in US classrooms can be seen as a means of keeping learning and effort on track – a form of classroom control, rather than encouraging the principles of educational assessment.

A more positive example of assessment by teachers is that found in Queensland, Australia, which abolished examinations over 30 years ago and replaced them with criteria-based subject-achievement assessments by teachers. The system involves extensive moderation by over 400 district review panels, composed mainly of teachers. These panels review samples of work across subjects and schools and provide advice to schools. The process is seen as an effective form of professional development in assessment. The only external assessment is the Queensland Core Skills Test (cognitive skills based on the common curriculum elements) which is taken by 17-year-olds. This – together with teacher assessment – provides the basis for entry to tertiary education. We believe this qualifies as alternative assessment by providing an alternative to selection examinations and encouraging a richer curriculum in which teachers and students are actively involved in the assessments.

Summary

This section has reviewed examples of alternative assessments which seek to directly measure the desired skills and competencies rather than through the proxy of written tests. This is an approach historically grounded in occupational and vocational approaches to learning represented by the guild system. In educational settings, we have considered alternative approaches to assessment based on different models of learning. Social constructivist approaches encourage students to make meaning and to understand the principles behind what they are doing. This leads to forms of assessment such as portfolios and extended project work. Classroom assessment by teachers will need to be interrogated to see if it is offering such opportunities. One way of ensuring this is through more effective formative assessment, which we now consider.

Level 3: Alternative Purposes of Assessment

The assessments we have considered so far have been primarily used for summative purposes, summing up where students have got in their learning. This section considers an alternative uses of assessment: the formative and the diagnostic. Formative assessment serves different

purposes since its main intention is to assist the learning process rather than to measure what has been learned. Diagnostic assessment can be seen as a subset of this in seeking to find out what is known in order to help with further learning. There are some considerable overlaps with the previous level: summative assessments can be used formatively if the results are used to guide further learning; performance and educational assessments will often involve the kind of self-regulation and feedback which characterize formative assessment.

What is the Purpose of Formative Assessment?

The purpose of formative assessment is to inform learning and teaching processes so that further learning takes place. This can be contrasted with summative assessment, in which the intention is to provide a summing up of where learners have got to in their learning. At the heart of formative assessment is the assumption that assessment, broadly understood as information gathering, can help identify what learners know – so that further teaching can build on this. It also assumes that learners will play an active part in the process and will increasingly be able to monitor and regulate their own learning. A key mechanism in this process is feedback, since this can help the learner bridge the gap between current and desired performance. We prefer to call this approach assessment for learning as this makes the purpose clear. A useful definition is:

The process of seeking and interpreting evidence for use by learners and their teachers, to identify where the learners are in their learning, where they need to go to and how best to get there. (Assessment Reform Group, 2002: 2–3)

This is essentially about day-to-day classroom assessment and depends on the quality of classroom interactions. It supports a pedagogy in which teachers seek to negotiate, and make explicit, learning intentions and success criteria ('where they need to go to'). It also involves finding out where learners are in their learning. This may be by diagnostic assessment – for example, miscue analysis in reading or investigating errors on tests. On a day-to-day basis, it occurs through classroom interaction – for example, questioning and dialog.

One small example of the way in assessment for learning can modify approaches to teaching and learning is the practice of wait-time. It is known that when teachers ask questions they expect an almost immediate reply. If effective learning is about the student-constructing meaning, what chance is there for this when an immediate answer is expected? Also, what kind of question only takes a second to process before answering? Wait-time, therefore, encourages the teacher to ask richer questions and allow students more time to process the answers (and there

maybe more than one correct answer). This approach also encourages students to discuss briefly the answer with each other before the teacher asks for responses. This may also lead to following up on unexpected or incorrect answers since these misconceptions may help in understanding where learners are in their learning.

Wait-time is just one small example of how formative assessment may change the quality of classroom interaction. This is no longer the teacher using surface-level questions to check quickly whether students know the correct answers, it involves a more deliberate attempt to gauge their understanding.

Feedback

One of the key components of assessment for learning is feedback. Its role is to help move learners from where they are toward the desired standard – the ‘and how to get there’ of our definition. What we know about feedback is that it does not always work as intended. In their meta-analysis of research on feedback, the psychologists Kluger and DeNisi (1996: 275, 277) concluded that:

In over one third of cases Feedback Interventions reduced performance . . . we believe that researchers and practitioners alike confuse their feelings that feedback is desirable with the question of whether Feedback Intervention benefits performance.

Whether feedback has positive learning effects depends on many interacting factors: motivation, the complexity of the task, the expertise of the learner, and the level and quality of the feedback. This makes it highly situational: the same feedback given to two learners could have opposite effects. Simply telling novice learners that they are wrong may set back learning, telling engaged experts the same may be enough to get them to increase their efforts and change their strategy.

Assessment for learning encourages feedback which focuses on the task rather than on the learner, and which has the goal of learners who become increasingly self-regulating. This involves being sufficiently aware of the desired standard that they can begin to give themselves, and their peers, feedback. This leads to alternative thinking about classroom practices, for example, questioning whether praise is an effective form of feedback (how does praise provide information to close the gap in learning?). It may also question routine marking and grading practices. For example, if the feedback to a piece of work is a mark (7/10) how does that provide information to move learning forward? Most teachers may put a comment alongside this – but research suggests that this is largely ignored, it is the mark that counts.

Where this leads is to classroom practice with an emphasis on developing self-regulated learners who receive informative feedback. This involves more focus

on the task, with feedback information intended to inform future performance. It will also lead to less emphasis on marks and grades, and more encouragement for comment-only marking.

Our claim is that formative assessment represents an alternative purpose for assessment: to contribute directly to the learning process. In terms of the approach to learning, it overlaps with the social constructivist views that we considered in the previous section and aligns well with approaches such as performance assessment and portfolios.

Diagnostic assessment

A further example of different purposes of assessment is its use for diagnostic purposes. Genuine diagnostic testing can be seen as a subset of formative assessment (where the learners are in their learning). We distinguish this from diagnostic testing which is used for managerial purposes – for example, to allocate students to groups. Diagnostic testing can involve informal analysis – for example, of a pupil’s reading skills and limitations – or the use of test results to identify what has not been understood. This test-and-remediate approach would often be called formative assessment in the US. In this, the formative element is in the remediation work following a test, occupying, for example, 1 week in a 6-week module. This approach to formative assessment differs from assessment for learning, which treats it as a continuous everyday classroom practice.

Dynamic assessment

An alternative approach to diagnostic assessment is provided by dynamic assessment. This has drawn on the sociocultural theories of Vygotsky (1978), who pointed to the importance of tools and aids in learning. What this diagnostic approach seeks to find out is what progress a learner can make when offered adult support in a one-to-one situation. So it is not so much what learners know, but how effectively they learn when assisted. In the traditional examination and test, the student is denied the use of external tools. As Lunt (1994) explains:

Dynamic assessment procedures . . . involve a dynamic interactional exploration of a learner’s learning and thinking process and aim to investigate a learner’s strategies for learning and ways in which these may be extended or enhanced. Since it offers individuals an opportunity to learn, dynamic assessment has the potential to show important information about individual strategies and processes of learning and, therefore, to offer potentially useful suggestions about teaching. (p. 152)

This approach – like formative assessment and educational assessment – represents a challenge to approaches to assessment found in the psychometric tradition. They

are alternative not because they do the same thing differently but because they attempt to achieve something different. This difference is rooted in current understandings of how we learn and in a search for ways of aligning assessment with approaches to teaching and learning which emphasize the importance of using assessment to support good-quality learning. As such, it requires a different relationship between assessor and learner, just as learning theory tells us the learner should be an active participant in his/her learning.

See also: Assessment and the Regulation of Learning; Assessment in Vocational Education; Dynamic Assessment; Formative Assessment; Moderation of Student Work by Teachers; Peer and Self-assessment.

Bibliography

- Assessment Reform Group (2002). *Assessment for Learning: 10 Principles*. University of Cambridge, UK: Assessment Reform Group.
- Fredericksen, J. R. and Collins, A. (1989). A systems approach to educational testing. *Educational Researcher* **18**(9), 27–32.
- Klenowski, V. (2000). Portfolios: Promoting teaching. *Assessment in Education: Principles, Policy and Practice* **7**, 215–234.
- Kluger, A. V. and DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* **119**, 252–284.
- Lunt, I. (1994). The practice of assessment. In Daniels, H. (ed.) *Charting the Agenda*, pp 145–170. New York: Routledge.
- Shepard, L. A. (2000). *The Role of Classroom Assessment in Teaching and Learning*. In Richardson, V. (ed.) *Handbook of Research on Teaching*. Washington, DC: American Educational Research Association.
- Shepard, L. A. and Bleim, C. L. (1995). Parents' thinking about standardized tests and performance assessments. *Educational Researcher* **24**(8), 25–32.

- Vygotsky, L. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.
- Wood, R. (1986). The agenda for educational measurement. In Nuttall, D. L. (ed.) *Assessing Educational Achievement*. London: Falmer.

Further Reading

- Bennett, R. (1998). *Reinventing Assessment*. Princeton, NJ: Educational Testing.
- Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice* **5**, 7–74.
- Gipps, C. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. London: Falmer.
- Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education* **24**, 357–394.
- Gipps, C. and Stobart, G. (2003). Alternative assessment. In Kellaghan, T. and Stufflebeam, D. L. (eds.) *International Handbook of Educational Evaluation*, pp 549–576. Norwell, MA: Kluwer.
- Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of Educational Research* **77**, 81–112.
- Koretz, D. (1998). Large-scale portfolio assessments in the US: Evidence pertaining to the quality of measurement. *Assessment in Education: Principles, Policy and Practice* **5**, 309–334.
- Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* **18**, 119–144.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher* **29**(7), 4–14.
- Stiggins, R. J. (2001). *Student-Involved Classroom Assessment*, 3rd edn. Upper Saddle River, NJ: Prentice-Hall.

Relevant Website

- <http://www.assessment-reform-group.org> – Assessment Reform Group.
- <http://www.cse.ucla.edu> – Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Classroom Assessment Tasks and Tests

J Phelan and J Phelan, University of California, Los Angeles, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Classroom assessment – The process of observing, collecting, analyzing, and interpreting evidence that can be used to draw inferences and inform decision making.

Formative assessment – The assessments used by teachers to guide them in adjusting instruction so as to improve learning (also called assessment for learning).

Performance assessment – The assessments designed for students to create an original response that showcases their competencies, which are then judged on predetermined criteria and observation.

Portfolio assessment – A systematically selected sample of student work intended to demonstrate growth and progression toward a specified learning goal.

Scoring rubric – A set of criteria against which student work is scored.

Summative assessment – The assessments used to evaluate how well a student has attained a certain learning goal or level of competency (also called assessment of learning).

Validity – Refers to how well what is being assessed actually corresponds to what was intended to be assessed and is intrinsically linked to the inferences one makes from an assessment.

Classrooms are filled with situations requiring teachers to make difficult decisions with important consequences: Should Sam and Jack be sitting next to each other? Is Charlie still struggling with his reading? Has the class mastered the content needed to pass the next benchmark assessment? What topics should be covered in math class tomorrow?

Some of the decisions that teachers have to make are about individual students, while others pertain to the class as a whole. Some decisions are related to what content to teach and how to teach it, some concern the classroom environment, and others involve evaluating how well students are learning the material and responding to the curricula. Many of these decisions must be made on a daily basis, while others may only happen once or twice a year. The common thread tying all of these decisions together, however, is that they should all be based on evidence gathered in the classroom, through the process

of classroom assessment. In this article, we describe some of the tasks, tests, and other routine classroom activities – or assessments – that provide a basis for observing, analyzing, interpreting, and judging students' learning and achievement in order to make these decisions.

What Is Assessment?

How are my students doing? Classroom assessment is the process of observing, collecting, analyzing, and interpreting evidence that can be used to draw inferences and inform decision making, helping you to answer this fundamental question. Moreover, careful consideration of the question helps to reveal that it is a question that can (and must) be answered in many different ways; classroom assessment is not a one-size-fits-all activity.

There is no single way to collect assessment evidence. Paper-and-pencil tests come to mind as the stereotypical classroom assessment, but many other options are available. These include the more traditional classroom assessments including quizzes, homework, written tests, final examinations, and so on, as well as many newer assessment methods such as portfolios, performance assessments, open-ended questions, online assessment, and computer simulations – sometimes considered alternative assessments. We should note that assessments delivered online or in an electronic format tend to be very similar to other assessments – however, with a different delivery and scoring method. Simulations are often outgrowths of more traditional-type assessments – designed to create a more realistic context in which learning and assessment can take place. For more on simulations and assessment, see the article by O'Neil elsewhere in this encyclopedia. Just as there are multiple types of classroom assessment, so too are there multiple reasons for teachers to use classroom assessment. These include: assigning a grade, quantifying the effectiveness of a new teaching method, gauging improvement toward a learning goal, and evaluating mastery of state standards.

Purposes of Assessment

Educational assessments are used for many purposes. On the one hand, teachers may use results of an assessment to form different reading groups, or assess the impact of various teaching strategies on the understanding of a

mathematical algorithm. On the other hand, assessment results may be used by organizations outside of the school walls to evaluate the effectiveness of a new district-wide reading program, or to evaluate the school as a whole. Thus, assessments can perform evaluative functions just as they can be used in a summative way to measure student achievement or in a more formative way to help refine student learning. Here we focus on the latter two, using assessment as a tool for measuring student achievement and for refining student learning within the classroom context.

Some assessments provide very detailed information about a particular student while others provide broader snapshots of achievement at the school or classroom level. The information useful to a teacher in figuring out how well a group of students understand chemical bonding differs from the large-scale assessment information designed to test mastery of the state chemistry standards. The specific purpose of an assessment, of course, plays a central role in how the assessment is designed. In the next section, we review the chief principles of assessment design. We also discuss examples of each, highlighting their particular value to teachers.

Assessment Design

Effective classroom assessment requires that teachers have a clear and complete understanding of the learning goals, have tasks that will allow them to see if these goals are being met, and finally, have the ability to interpret the evidence collected from these observations. Here, we discuss some of the important issues in satisfying each of these requirements.

Learning Goals

When designing an assessment, you must define clearly the learning goals. These learning goals encompass the knowledge or skills you wish to assess in your students. They must be articulated in such a way that students have a clear picture as to what the goals are and what they can do to achieve them. In many areas, models of cognition and learning have been developed and these can be used to guide design of both curricula and assessments. For example, in a middle-school math assessment project underway at the Center for Research on Evaluation, Standards and Student Testing (CRESST), development of formative assessments was preceded by a detailed analysis of the domains to be assessed and the creation of an algebra ontology. Experts were asked to consider which ideas were most important in their own thinking and problem solving. The expert panel first identified the big ideas that organized their thinking and work in the domain, then the subordinate or supporting ideas that elaborated and gave meaning to the organizing concepts.

An ontology showing the big ideas and relationships between them is shown in **Figure 1**. The ontology was then used to guide assessment-development efforts.

Learning goals also may be closely tied to standards set by the state, or may include the performance of a skill such as ballroom dancing, or painting. A model of knowledge to be learned (such as the one outlined above) can be used to make sure that what is being taught in the classroom is the important content. In any case, it is critical to identify the set of skills and body of knowledge that is important to know and measure. First and foremost, learning objectives must cover the most important concepts in a subject area, and assessments in turn should be closely aligned with those objectives. Assessment design must also focus on the cognitive demands of an assessment task – and not the superficial features.

Interpreting Evidence

Once the model of learning is in place, the types of inferences to be drawn from the assessments must be considered. What are the educational goals: memorization of multiplication tables or the application of acquired knowledge to a novel situation? If the former, then a complex performance task will be a poor choice, but if the latter, such an assessment would be valuable. The interpretation of evidence determines how information will be taken from the assessment task and used to produce results, ultimately allowing instructors to make inferences about whether or not their students have mastered a particular concept or acquired a particular skill.

Assessment Tasks

Assessment design – as well as the procedures for evaluating performance – should take into consideration both the learning goals and the way that evidence from the assessment will be interpreted. Of chief importance to designing a useful assessment is that it will provide you with information you can use to draw inferences. Assessments must be intrinsically connected to both the curriculum and classroom instruction and provide students the opportunity to show that they have attained a learning goal. Another key feature is construct validity – does the assessment task actually tap into the skills and content you intend it to? In other words, are you measuring what you want to be measuring? For a biology course with a learning objective focused on understanding Charles Darwin's theory of evolution by natural selection, an assessment task asking the name of the ship he sailed on around the world is poorly aligned with the objective. A task requiring students to describe the three conditions that must be present for natural selection to occur, on the other hand, will allow a useful measurement of what students know. Knowing whether a student responded

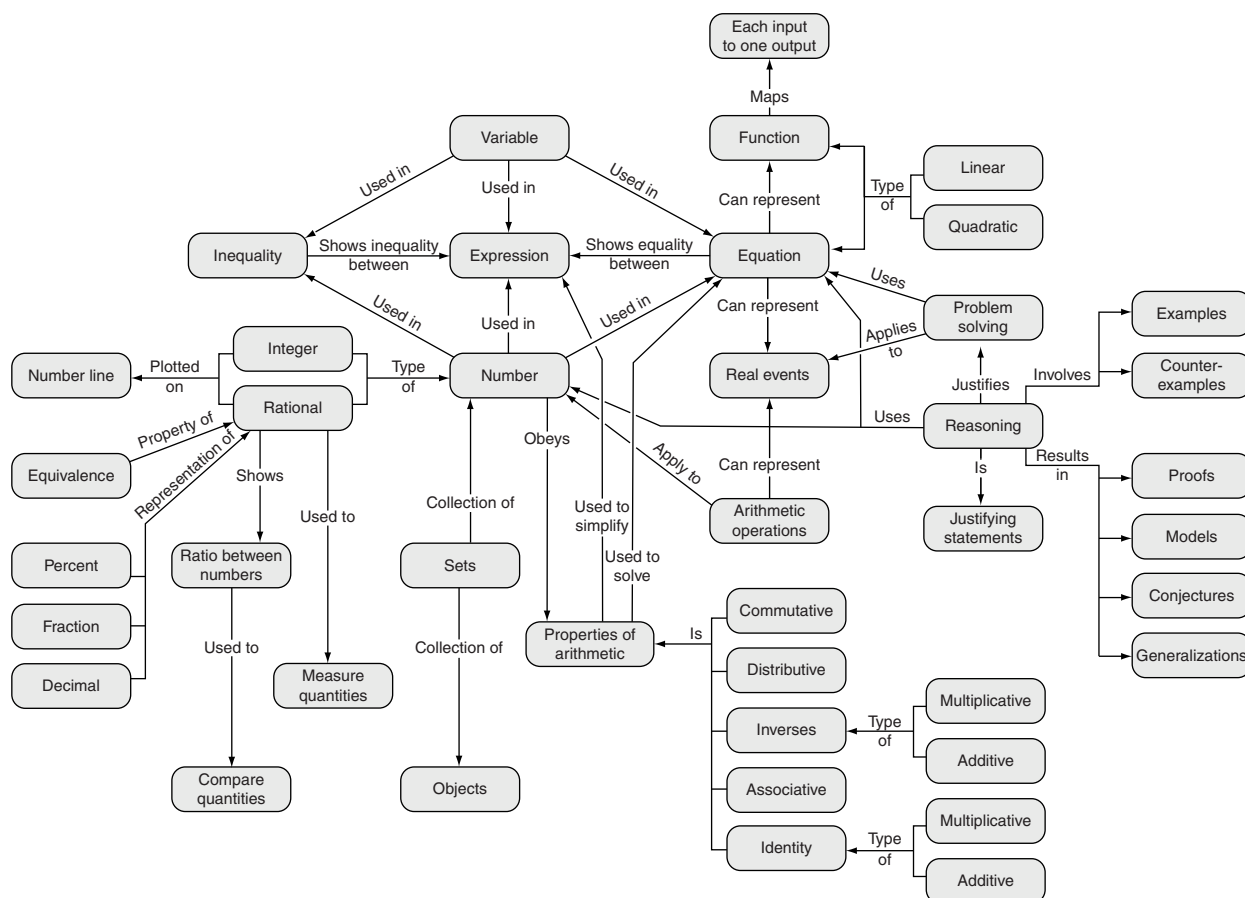


Figure 1 Map showing relationship of key big ideas.

correctly or incorrectly to the question about Darwin's ship, does not give you any information about whether they understand the concept of evolution by natural selection.

The method of generating assessment evidence – whatever it may be – varies. The common thread, however, is that an assessment should provide enough of the right information for teachers to be able to make appropriate decisions based on the results. There are many examples of classroom assessments, more than we have room to discuss here. We briefly describe the two most broad categories of assessment and then discuss some different types of classroom assessment, how they are designed, and the types of evidence and inferences we can expect to get from them.

Types of Assessment

Classroom assessments can be grouped into two main categories: summative assessments and formative assessments. This distinction rests primarily on how the information from these assessments is used. We will not focus too much here on the distinctions between the broad types of assessment used within the classroom. There is

more information on formative assessment and summative assessment elsewhere in this encyclopedia. It is, however, critical to understand that the two types of assessment described here do not necessarily align with different forms of assessment. What defines the assessment categories is how the results of the assessments are used – and not how they were collected. The key here is that formative assessment is used to make decisions that impact teaching and learning and summative assessments are used to evaluate how well a student has attained a certain learning goal or level of competency.

In the following sections, we describe some of the assessment types and techniques used within the classroom. Some of these assessment types are considered formative while some are more summative in nature and all with varying degrees of formality. What becomes clear as we survey the different assessment types is that formative and summative assessments are not necessarily mutually exclusive. An end-of-term test may be used to provide a grade for a course, but may also be used to gather information to be used in the next grade level. Teachers may not have the opportunity to use information from assessments to change their instruction mid-stream. They may, however, file the information away

until the next year when they revisit a topic with a new set of students. So we see that some assessments can be summative in the short term, but formative in a long-term way.

Among the many types of classroom assessments that we discuss below, some are more traditional and some fall into the category often called alternative assessments. For each of the assessment types discussed, we describe the features of the assessment and illustrate the potential benefits and possible pitfalls surrounding their use in the classroom.

Classroom Assessment Methods

Paper-and-Pencil Tests

For most grades, formal paper-and-pencil tests are the most commonly used form of gathering evidence about student learning. Tests can come from myriad sources, including textbooks, teacher created, online, or standardized tests designed by school districts. The selection of assessment items on such tests is critical. If students have gained a rich understanding of a particular historical period, for example, but are presented with confusing test questions, they may not be able to demonstrate their knowledge of the material. If students are not able to accurately demonstrate what they know, the opportunity for teachers to make effective instructional decisions based on the results is greatly diminished. Indeed, the use of a poor-quality assessment measure can cause teachers to make inappropriate decisions, perhaps concluding that a class has mastered a topic when that is not the case.

Paper-and-pencil tests comprise a set of questions or items. Typically, the questions on a test are independent of each other (although they may be on similar topics). Here we present some general principles to keep in mind when creating paper-and-pencil tests for use in the classroom. The overarching goal is to maximize the opportunities for students to demonstrate what they have learned, while minimizing confusion and ambiguity. There are two basic types of items: selected-response items and items where students construct their own responses. First, we consider selected-response items.

Selected response items

There are several different types of selected response items commonly used on classroom assessments, including multiple choice, true/false, and matching tasks. While there are some specific guidelines appropriate to the specific types, in general, items should focus on the learning goals previously determined, provide as much information as to how well students have mastered these goals, and be constructed in a valid way so that we know we are assessing what we want to be assessed.

Multiple-choice items

Multiple-choice items – including a question stem and a set of answer choices from which the student must pick the best one – are appropriate both for measuring factual information and for evaluating mastery of tasks with higher-level cognitive demands. Some of the limitations of multiple-choice items are that students can guess the answer, they only need to recognize the correct response (rather than generating it), and students do not have the opportunity to construct their own responses. Properly designed multiple-choice questions can, however, be a useful classroom-assessment tool and can measure a rich variety of student knowledge and skills.

Students may either choose an answer if they know it is correct. Five of the most frequently cited item-writing guidelines for multiple-choice items are:

1. The question stem should present a self-contained question or problem, for example, “How many chromosomes are there in a human gamete cell?”
2. Negatively stated stems should be used sparingly, for example, “Which of the following cities is NOT in Europe?”
3. The correct response should not consistently be markedly different than all the incorrect responses (e.g., much longer or shorter). This may give unintended clues to the students.
4. The correct answer’s position among the possible responses should be randomly assigned.
5. “All of the above” should be used sparingly as an answer choice. Students may either choose A (e.g.,) if they know it is correct and not read the rest of the answer choices, or they may only select “all of the above” if they know at least two of the answer choices are correct. “None of the above” can be used if it aligns with the goals of your assessment. Some have argued that it will cause students to be more careful about guessing one of the first four answer choices, without carefully considering each option. If students are not certain that the correct response is there, they will be less likely to pick an answer which is a close approximation to the correct one.

Contrary to common belief, a multiple-choice test is not limited to factual recall tasks. Multiple-choice items can also be used to assess complex skills, such as interpreting a graph or piece of data, or identifying a relationship between two phenomena, and so on. For example, **Figure 2** shows a multiple-choice item that requires students to analyze an experimental situation and determine how to manipulate the experimental setup in order to get a particular outcome.

Matching items

These items consist of two columns, or lists of words or phrases. The list for which students are looking for a match

Problem: In the following situation, two identical steel marbles M1 and M2 are to be launched horizontally off their respective tracks. They each leave the ends of their respective tracks at the same time, but M2 will leave its track traveling twice as fast as M1. The track for M1 can be set at any height in relation to M2.

(a) If we want the two marbles to collide, how will we need to arrange the horizontal launch tracks?

- A. The track for M1 should be much higher than the track for M2.
- B. The tracks for M1 and M2 should be at the same elevation.
- C. The track for M1 should be much lower than the track for M2.

Figure 2 A multiple-choice item constructed to tap complex cognitive processes. Adapted from Minstrell (1992).

is called the premises, and the list from which they will find the match is called the responses. The task for students is to match an entry on one list with an appropriate entry on the other list, see **Figure 3**. A matching item is a relatively easy way to assess factual material and one that takes up a small amount of space and can be easily scored.

When creating matching items, the lists should be homogenous, relatively short, have logically ordered responses, have more responses than there are premises, be all placed on the same page, and be preceded by a concise description of the match to be made. Instructions should also be clear on how many times a response can be used. It is important to have more responses than premises to reduce the likelihood that students can guess the correct answer once they have eliminated a few choices. A limitation of this type of item is that it can encourage memorization rather than deeper comprehension of concepts.

True-false items

The true-false format consists of a statement which the student must determine is correct or incorrect. There are two response categories, for example: yes/no, correct/incorrect, or true/false. The main feature is that there are two choices. A strength of this type of classroom assessment is that students can typically respond to many of these items in a short period of time and so a large amount of content can be covered in a relatively short assessment session. Limitations of these questions are that students have a 50% chance of guessing the correct answer, and in some cases, it may be hard to find statements that are unequivocally true or false. Typically, true-false items are quite short, but this should not necessarily lead to them being easy questions. Guidelines for creating true-false items include:

1. Care should be taken to not construct items that are very obviously true or false.
2. Avoid using negatives in true-false items; this may cause difficulties for students. For example, London is not the capital city of France. Students may know that London is not the capital of France but may not know if they should say true it is not, or that the statement is false.

On the line to the left of column A, write the letter of the person, from column B, with whom Tom Cruise co-starred in that movie.

Column A	Column B
___ <i>Top Gun</i>	A Dustin Hoffman
___ <i>The Color of Money</i>	B Nicole Kidman
___ <i>Far and Away</i>	C Thandie Newton
___ <i>The Firm</i>	D Paul Newman
___ <i>Rainman</i>	E Richard Gere
	F Gene Hackman
	G Val Kilmer

Figure 3 Example of a matching item.

3. The number of true responses should be approximately equal to the number of false responses.
4. The length of the statements should be similar for both the true and the false question stems. The correct statements should not necessarily be longer than the incorrect statements.

Constructed-response items

Constructed-response items require students to generate their own answer, rather than choosing a correct answer or simply indicating whether a statement is true or false. Here, we discuss two types of constructed response – the short-answer question and the essay. Although portfolio and performance assessments can be classified as constructed-response items, we discuss them separately in the next section. The major value for constructed-response tests is the ability to gain insight into what students are thinking and the depth of their knowledge.

Short answer

Short-answer items require students to answer a direct question such as, “Who is the Prime Minister of England?” Some may also write completion items which require students to complete a sentence such as, “The Prime Minister of England is _____”. Most guidelines for writing short-answer questions suggest using direct questions rather than the fill-in-the-blank option. The direct question is a more familiar format and leads to less confusion than the alternative. Typically, the answer required is very short – either one word, or a short phrase.

The strength of using a question like this is you decrease the likelihood that students will simply recognize the correct answer as they have to generate it themselves, rather than choosing it from a list of options. As with all constructed-response items, the main limitation of this type of question is that it is more difficult to score than a selected-response item. The longer the response, the more difficult it is to accurately score. The trade-off between getting richer information and reduced accuracy in scoring must be weighed.

Essay or extended response

Essay questions provide opportunities for students to create longer, more detailed responses, and to draw their own conclusions. Students have latitude in determining how they organize their response and what they include. Essay questions have the potential to provide teachers with a more detailed, richer response and can be especially useful in allowing teachers to evaluate students writing and composition skills. An example essay question is shown below:

What does scientific testing reveal about the accuracy of eyewitness testimony in courtrooms?

This question would allow you to see if students understood the limitations of eye-witness testimony and if they could adequately explain them using logical arguments.

Limitations of essay questions include that they are harder to score, rely heavily on students' writing ability, which can cause the score to reflect something other than what you are trying to measure, are time-consuming to answer, and allow for less breadth of coverage of the content taught. Guidelines for creating essay items include:

1. be clear as to how extensive the response to a question should be;
2. explicitly describe the task;
3. provide students with the point value for the question; and
4. create an example of a possible response prior to administering the assessment.

Scoring essay-assessment items is more time intensive than scoring other assessments. There may also be a greater degree of subjectivity than in multiple-choice or selected-response tests. One of the most important issues when scoring an essay is to decide if it will be scored analytically – with a more finely grained, point-by-point approach, or as a whole. When scoring an essay as a whole, teachers can use their general impression of the work, or have some criteria against which they will judge the student's response. If scoring a written composition, these criteria might include organization, clarity, word choice, spelling, consideration of the audience, and so on. If scoring an essay in a more analytic fashion, teachers

may create a scoring guide in which they allocate a specified number of points depending on the satisfaction of a particular criteria. For example, there may be between 0–3 points available for organization of the essay. An advantage of scoring an essay in a more analytic way is you can better identify areas where students need help. It may also be the case, however, that by combing through an essay looking to see if a student has satisfied particular specific criteria, a grader misses the overall feel of the essay as a whole. Both subgroups of question types – selected and constructed response – can allow teachers greater insight into higher-order thinking skills (e.g., analyzing, explaining, evaluating, and creating).

Performance Assessment

All assessments require some element of student performance – be it checking a box, selecting a response or writing an answer. Performance assessment, however, specifically refers to an assessment designed for students to create an original response that showcases their competencies that is then judged on predetermined criteria and observation. Performance assessment is an assessment approach designed to investigate how a student completes a particular task and demonstrates what he/she has learned. The components of a performance assessment are such that they build directly on learning and measure students' ability to apply their knowledge in new situations. Performance assessments have three basic features:

1. allow students to apply the skills and knowledge they have learned;
2. students complete tasks in the context of a real or simulated assessment; and
3. the assessment response or product is observed and rated according to specific predetermined criteria.

The latter point is crucial to the success of both instruction toward a goal and the performance assessment to see if students have achieved that goal. Just as you would not assign an essay without knowing what a good response would entail, so also a performance assessment must have clearly defined performance criteria against which students are compared. These criteria define the key aspects of the task and provide a guide as to how students should be both taught and assessed.

With performance assessments, it is possible to evaluate a student's ability to translate what they have learned into an action of some sort – be it a speech, speaking a foreign language, a response to art, or the art itself. These assessments may be in response to a written prompt or to directions given by an examiner. Consider students learning to speak French. After several months of study, they are called into a room and are asked to engage in conversation in French with the examiner. They have no idea ahead of time what the topic of conversation will be

and so they must draw on what they have learned and attempt to generalize it to the topic at hand. The examiner makes notes on the student's performance during the conversation based on predetermined criteria (which might include their accent, number of incorrect words, use of the incorrect tense, and so on). Following the conversation, a score is given to each of the students indicating how well each of them satisfied the various criteria.

Another example of a performance assessment might be dissection of an animal in biology class, an identification of an unknown chemical substance, or an evaluation of a historical document (see **Figure 4** for an example of a writing performance assessment).

Performance assessments can, alternatively, be more informal – perhaps based on observations made by a teacher as to how well a student can read aloud or even shoot a free throw in gym class. Performance assessments are particularly valuable when evaluating how well students can translate or transfer their mastery of a topic or task to solve a problem they have previously not had experience or practice with. Performance assessment is supposed to focus on tasks calling for complex thinking, deep understanding of subject matter, and open-ended responses. As school districts place more emphasis on higher-level thinking and problem solving, they increasingly turn to performance assessments to evaluate student learning.

Some issues associated with performance assessments, as compared with other types of assessment, are: (1) they take a long time to evaluate and require skilled judges and detailed rubrics; (2) some curricula and even teachers may not teach in a way that lends itself to performance assessment; (3) they can be difficult and expensive to develop; and (4) the amount of information gleaned from a 30 minute performance assessment may not measure up to that learned from a 30 minute multiple-choice test. The stakes become somewhat higher for students,

given the reduced breadth of material covered in a performance assessment. The student who succeeds at the task may have been lucky and the topic assessed could be the only thing about which he/she has any mastery. Given that most teachers will only have time for a small number of performance assessments, it is crucial that they are used to assess important content and content that will allow teachers to accurately generalize about their students' capabilities.

Performance assessments can be scored using rubrics, or checklists (as described for essay questions). Scoring may focus on specific skills (an analytical approach), or be very general (a holistic approach). Regardless of the approach, there must be specific criteria identified to judge the student responses against and there must be clear descriptions of the differences between score points. But if the assessment is to serve the purpose of highlighting specific areas of weakness in a student's performance, a rating scale, rubric, or separate score for a set of performance criterion may be more appropriate. Some rubrics may be a combination of a holistic score for overall quality or level of understanding followed by more specific criteria (see **Figure 5**).

Rubrics such as the above depend heavily on strong teacher knowledge. Teachers must be able to readily identify any misconceptions revealed by a student's performance and determine the degree of importance. Specific examples can be added to rubrics to help provide more structure, although doing so increases the risk of having teachers look for exact phrases or sentences, and also the idea that there is only one correct performance.

Portfolio Assessments

Portfolio assessments are a form of performance assessment created in part as a response to the need for greater authenticity of assessment tasks. In educational settings, a portfolio is a systematically selected sample of student

Writing assignment

Imagine you are taking a chemistry class with a teacher who has just given the demonstration of chemical analysis you read about earlier.

Since the start of the year, your class has been studying the principles and procedures used in chemical analysis. One of your friends has missed several weeks of class because of illness and is worried about a major exam in chemistry that will be given in two weeks. This friend asks you to explain everything that she will need to know for the exam.

Write an essay in which you explain the most important ideas and principles that your friend should understand. In your essay you should include general concepts and specific facts you know about chemistry, and especially what you know about chemical analysis or identifying unknown substances. You should also explain how the teacher's demonstration illustrates important principles of chemistry.

Be sure to show the relationships among the ideas, facts, and procedures you know.

Figure 4 Sample prompt for a writing-performance assessment in chemistry. Adapted from Baker, Aschbacher, Niemi, and Sato (1992).

Essay scoring guidelines

1. General impression of content quality (GICQ)

How well does the student know and understand this historical content?
(0–5 point global rating; 0 = no response, 5 = highest level of understanding)

2. Misconceptions (MIS)

This is measure of the amount of incorrect information, or the number of misconceptions or misinterpretations, in the essay.

Score point guidelines:

- 0 no response
- 1 one or more serious misconceptions central to the essay
- 2 at least one serious misconception
- 3 several minor errors and/or a moderate misconception
- 4 very minor misconception
- 5 no misconceptions.

Figure 5 Example of two essay scoring guidelines used in a historical-writing performance assessment. Adapted from Baker, Aschbacher, Niemi, and Sato (1992).

work intended to demonstrate growth and progression toward a specified learning goal. One of the most commonly used applications of educational portfolios are writing portfolios. These consist of collections of students writing – including outlines, rough drafts, finished essays, and various different forms of writing. These writing portfolios can be used by pupils, parents, and teachers to get a sense of development and improvement over time.

Effective portfolio assessment requires more than just the evaluation of a nonstructured collection of a student's work. As in performance and other assessments, there must be clear purposes for the work included, performance criteria on which students can be evaluated, and also rating or scoring criteria. Thus, the collection of the student work must be guided by a particular educational purpose – for example, to demonstrate improvement in writing persuasive essays. In terms of the writing portfolios, students may be working toward a particular learning goal and the portfolios can be a way of measuring whether they are adequately doing so. If the learning goal is to be able to identify conflict in a narrative passage and create one's own conflict in a written piece, the portfolio may showcase samples of student work along the road to this goal – a piece of criticism, drafts of works in progress, a response to literature, or an essay. Then when the portfolio is examined as a whole, it is possible to evaluate where the student began in terms of a particular goal and how he/she improved over a period of time.

Methods for scoring portfolio assessments depend on the purposes for the assessments. Like all of the assessments we have discussed, portfolios may serve summative or formative purposes. The particular function of the assessment should inform how it is evaluated. If the portfolios are meant to serve as a collection of student work to provide descriptive information to parents, or other teachers, the contents may not need to be scored at all. If the

purpose of the portfolio is to document growth or provide a grade for a student, some form of scoring is necessary. Most use checklists, rating scales, or rubrics to score portfolios as the amount of work contained within is typically large and scoring can be quite a cumbersome and time-consuming task.

Rating scales for portfolio assessments may include criteria such as level of understanding, quality of reasoning skills used, and use of voice or tone. When scoring a portfolio, the scoring must reflect a larger body of work than just one assignment, so different scoring criteria must be used than those used for scoring individual assignments. If the purpose of a portfolio is to determine how much a student has improved in literary-analysis skills, for example, attention must be paid to the change over time, from the early assignments to the later ones.

While portfolios can provide a unique insight into the development of skills and provide a view of a large body of work, there are challenges associated with their use. One challenge is that portfolios should be looked at as a form of performance assessment and the same steps taken in creating them. Without doing so, the portfolios will simply consist of a collection of student work with no clear way to get useful information or evidence. Therefore, there must be a clearly stated purpose of the portfolio, which will influence the type of work collected. If the purpose of a portfolio is to demonstrate improvement in writing, samples must be collected across the whole period being assessed. If only final drafts are included, there will be no way to determine if the students have improved. It must be clear from the start. Another potential pitfall of portfolios is that the support provided to students may be varied and the work in the portfolios may not reflect just the student's own work, but be a combination of student work with outside help.

Informal Classroom Observations

The classroom environment is filled with opportunities for teachers to gather valuable, but less-tangible information about students and their behavior. However, these characteristics of students (such as motivation, interest in school, or level of attention to a class discussion) are difficult to assess using formal, structured classroom assessments. Teachers frequently use informal observations to collect information about students that they are unable to collect in a more traditional way. They may want to gather information to make instructional decisions (should I speed up, or slow down this lesson?), in which case, the focus is more on the class as a whole. Or they may make an observation about an individual student, for example, “Andy does not talk much to the other students,” or “Elisabeth always puts her hand up first when a question is asked.” Unlike formal assessment, where there is an *a priori* decision made to collect information, with informal assessment, the information gathering has a spontaneous nature. Like formal assessment, however, informal assessment also requires interpreting and synthesizing information, determining what characteristic the information gathered is describing, and taking action based on evidence.

A teacher might ask – why does Andy not talk much to the other students? Then form a tentative explanation of why this might be. Next, a teacher might determine whether the current observation fits with previous observations of the student. If Andy was previously extroverted, the sudden shyness may be indicative of a change in another area. This is something the teacher must decide. There are many opportunities to gather this sort of information about a student – and this information may or may not fit with the teacher’s prior conception of that student. It is then up to the teacher to interpret the information and add it to that already collected. In this cyclical process, new information is constantly coming in and a teacher must be able to reassess a student if subsequent information contradicts the prior information. The cyclical nature of informal assessment is crucial.

Classroom observations have the benefit of providing more realistic information about student behavior than traditional formal assessments, but some limitations also exist:

1. Behavior must occur for it to be observed. So you can only measure behaviors that occur spontaneously.
2. Different characteristics for different students are also observed. So, one may get a more complete picture of one student than another.
3. The observations have a great degree of subjectivity. If a student is not paying attention and looks bored, they may not be understanding the material, they may be tired, or they may have already mastered the material and are wishing they could move on to something else.

There are many possible explanations for an observed behavior and the way the teacher responds to it relies heavily on their interpretation of the behavior.

4. Informal classroom observations are descriptive in nature. A student may be categorized as confident, lazy, motivated, bored, etc. Thus, students may be given descriptive labels that can be difficult to change even if future behavior is incongruous with the behavior that shaped the original impression.
5. Informal observations are rarely recorded. So the perceptions formed from the observations exist in memory only.

More systematic observations can improve the value of classroom observations. Simply recording observations, for example, can increase the value of such observations for helping to make educational decisions. This can be done in several ways:

1. Narrative, for example, audio/videotapes that can be coded or rated;
2. Checklist – set of predefined categories for tallying and classifying observed behavior of narrative data. These checklists can be used to gather information about how much a student understood in a lesson, or how much effort they put forth and so on; and
3. Rating scales – either graphic, numerical, or descriptive.

Validity

Teachers have at their disposal a great deal of assessment information from the many sources discussed above – some very formal, and some more informal. They use this assessment information to help make myriad decisions – about student understanding of a learning objective, about their own teaching, about the way their classroom is organized, and so on. Whether or not the assessment information available to teachers will help them make good decisions depends on the quality of that information. Classroom assessments must, therefore, allow teachers to arrive at valid inferences about their students. The accuracy of the decisions they make depends on the accuracy of the assessment.

Assessment validity is concerned with the accuracy of the inferences that teachers make about their students and may be the most fundamental consideration when developing assessments. Validity measures how well an assessment actually corresponds to what was intended to be assessed and is intrinsically linked to the inferences one makes from an assessment.

Teachers need to know how well students have mastered a particular learning goal – for example, determining equivalent rational numbers – in order for them to make effective decisions – such as how to proceed with instruction, or how well their students have mastered a concept. As it is difficult to create an assessment covering

everything that was taught in an instructional unit (or lesson, or part of a lesson), however, teachers must select a representative sample of content to be assessed.

Validation begins with the explicit statement of both the interpretations you would like to make (all students understand rational number equivalence, in the example above) and a good rationale for the relevance and importance of the material being tested (your original model of cognition). If a good model exists, choosing the appropriate content to assess becomes easier for the teacher. Not only must the content be clearly defined, but the assessment must also closely match that content. Including something extra may be just as problematic to validity evidence as leaving something out.

Consider a 12-item math test on rational number equivalence given to a group of students. The test is given after a course of study on rational number equivalence and the items are similar to those used during instruction. Validity is an overall evaluation of the degree to which we can conclude that a student scoring 100% on the 12-item test has actually achieved a full mastery of the rational-number-equivalence concepts. In other words, if the test was constructed well, and the domain or construct to be assessed was defined well, then the decision made based on the results is said to be valid. It is not the test that is valid, but the inferences made about the results of the test that are valid. If a test has a high degree of validity, one can be confident about the inferences made from a particular assessment.

In formative assessment, evidence that students have achieved particular learning goals is used by the teacher to determine the next instructional steps. Feedback is considered by the teacher and helps determine both the pace and content of subsequent instruction. So, when considering the validity of formative assessments, again there must be strong alignment between the specified learning goals and the criteria by which the evidence of achieving those goals are judged or compared. In the earlier example, we discussed a test focused on rational number equivalence. If we were to create a formative assessment in this domain, the lesson goal might be that students had to understand and be able to explain why any number multiplied by 1 is equal to that same number. So, when looking at the results of the formative assessment, the teacher would be looking to see if students understood this specific concept. If they did not, then additional instruction would be required. This assessment would be said to have a high level of validity if the intention of the lesson is well captured by the collected evidence.

Another consideration to be taken into account when designing assessments is reliability. The reliability of an assessment refers to the consistency with which the test measures what it is supposed to be measuring when repeatedly given to groups of students. An assessment is

not reliable if it yields significantly different results when administered by different individuals, or in different environments, or at different times.

In the context of classroom assessments, reliability is less of a concern than validity. Unlike in situations such as large-scale testing, teachers interact with the same students over an extended period and so there are multiple opportunities to compare assessments over time and adjust judgment accordingly. Particularly in the case of formative assessment, when there are no grades being assigned, assessment involves only the teachers and the student – no outside observers or other students to whom scores will be compared. In these cases, judgments based on the assessment are used in determining how best to proceed with instruction, and information is primarily used by the teacher to provide feedback to the student. Thus, reliability is not a strong concern. If however, test results were to be used by other people in order to compare students, reliability becomes more important. An example of this arises when applying to a new school, for example, middle school. The new schools require applicants to take a standardized test (one with consistent test materials, and administration and scoring procedures), results of which can be used to make comparisons across the whole applicant pool. This type of assessment has high reliability and is of great value relative to a comparison of students' classroom grades, which are likely to have been obtained in less-consistent ways (different schools, different teachers, different curricula, and so on). The consideration of the standardized tests scores allows students to be compared on a common measure with a higher degree of reliability.

Conclusion

There are a variety of diverse methods for assessing student learning. Some are summative, used to evaluate how well a student has attained a certain learning goal or level of competency, and others are formative, used by teachers to guide them in adjusting instruction so as to improve learning. In all types of assessment, however, it is essential that teachers have a clear and complete understanding of the learning goals, have tasks with high validity that allow them to see if these goals are being met, and have the ability to interpret the evidence collected from these observations.

We have presented here a description of some of the most useful types of classroom assessment, ranging from the more traditional paper-and-pencil tests to some alternative assessments including portfolios, performance assessments, and informal classroom observations. Each type of assessment has particular strengths and weaknesses, which we describe, that should be considered when formulating a comprehensive assessment strategy.

With proper design and implementation of classroom assessment, significant and valuable educational gains are possible.

See also: Alternative Assessment; Formative Assessment; Instructional System Provided Feedback; Portfolio Assessment; Summative Assessment by Teachers; The Multiple Purposes of Assessment.

Further Reading

- Airasian, P. (1994). *Classroom Assessment*, 2nd edn. New York: McGraw-Hill.
- American Educational Research Association (1999). *Standards for Educational and Psychological Testing*. Washington DC: American Educational Research Association.
- Anderson, L. W. (2003). *Classroom Assessment: Enhancing the Quality of Teacher Decision Making*. Mahwah, NJ: Erlbaum.
- Baker, E. L. (1998). *Model-Based Performance Assessment*. (Tech. rep. No 465). Los Angeles: University of California, Center for the Study of Evaluation.
- Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2002). *Working Inside the Black Box*. London: King's College, London.
- Black, P. J. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice* **5**(1), 7–74.
- Black, P. J. and Wiliam, D. (1998). *Inside the Black Box: Raising Standards through Classroom Assessment*. London: School of Education, King's College.
- Black, P. J. and Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan* **80**(2), 139–148.
- Downing, S. M. and Haladyna, T. M. (eds.) (2006). *Handbook of Test Development*. Mahwah, New Jersey: Erlbaum.
- Gardner, J. (ed.) (2006). *Assessment and Learning*. London: Sage.
- Gearhart, M., Herman, J. L., Baker, E. L., and Whittaker, A. K. (1993). *Whose Work is it? A Question for the Validity of Large-Scale Portfolio Assessment*. (Tech. Rep. No 363). Los Angeles: University of California, Center for the Study of Evaluation.
- Haladyna, T. M., Downing, S. M., and Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education* **15**(3), 309–334.
- Harlen, W. (2007). *Assessment of Learning*. London: Sage.
- McMillan, J. H. (1997). *Classroom Assessment*. Boston: Allyn and Bacon.
- McMillan, J. H. (ed.) (2007). *Formative Classroom Assessment: Theory into Practice*. New York: Teachers College Press.
- National Research Council (2001). Knowing what students know: The science and design of educational assessment. (Committee on the Foundations of Assessment. Pelligrino, J., Chudowsky, N., and Glaser, R. (eds.) Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education). Washington, DC: National Academy Press.
- Popham, J. W. (2008). *Classroom Assessment: What Teachers Need to Know*. Boston: Pearson.
- Shavelson, R. J. (2006). *On the integration of formative assessment in teaching and learning with implications for teacher education*. Paper prepared for the Stanford Educational Assessment Laboratory and the University of Hawaii Curriculum Research and Development Group. Retrieved October 19, 2006, from http://www.stanford.edu/dept/SUSE/SEAL/Reports_Papers/Paper.htm (accessed July 2009).
- Stiggins, R. (1987). Design and development of performance assessments. *Educational Measurement: Issues and Practice* **6**, 33–42.
- Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan* **86**, 22–27.
- Stiggins, R. J. (2007). Conquering the formative assessment frontier. In McMillan, J. H. (ed.) *Formative Classroom Assessment: Theory into Practice*, pp 8–28. New York: Teachers College Press.
- Stiggins, R. J. and Bridgeford, N. J. (1985). The ecology of classroom assessment. *Journal of Educational Measurement* **22**, 271–286.
- Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment* **11**, 283–289.

Relevant Website

<http://www.Diagnoser.com> – Instructional Tools for Science and Mathematics.

Dynamic Assessment

J G Elliott, Durham University, Durham, UK

E L Grigorenko, Yale University, New Haven, CT, USA

W C M Resing, Leiden University, Leiden, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Clinical DA – An approach to dynamic assessment in which the tester is free to determine the nature and form of feedback that should be provided to the testee.

Dynamic testing – An approach whereby contingent feedback is provided during the test session so that learning can take place and the testee's capacity for improvement can be gauged.

Learning potential/potential for learning – One's capacity to learn when optimal conditions are provided.

Standardized DA – A dynamic approach in which the nature of contingent feedback is not determined by the clinician but follows a prespecified path.

Static testing – The traditional approach to testing whereby the tester assumes a neutral stance and any feedback on performance during the testing session is proscribed.

Zone of proximal development – A term introduced by the Russian psychologist Lev Vygotsky to describe, "... the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance, or in collaboration with more capable peers."

The Need for a Dynamic Approach

When standardized intelligence tests are employed in educational settings, their primary purposes are usually for clinicians to be able to predict and describe the likely academic progress of the child, offer a diagnosis, and explain reasons for failure. Most tests used in schools have a static nature, even though, at the very beginning of the twentieth century, Binet the father of intelligence testing, defined intelligence in terms of the ability to learn. A static test, usually employing a set of standardized instructions at the outset, is characterized by its absence of contingent feedback to the child. Test scores are aggregated and global test or factor scores are interpreted in terms of the (cognitive)

abilities of the child. Scores on static instruments are indicative of the child's actual level of performance(s) reflecting learning that has taken place in the past.

Despite their continuing popularity, the weaknesses of traditional intelligence tests are widely recognized. It is often argued that such measures can underestimate the abilities and potential of children, in particular, those from minority groupings, disadvantaged socioeconomic contexts, or those with linguistic difficulties. Conventional intelligence tests are typically measures of achievement involving skills and knowledge that are generally acquired at home or in school. Children from economically disadvantaged contexts, or from non-Western cultures, are more likely to struggle with tasks and test procedures that are unfamiliar and unpractised and may become disconcerted by a formal testing situation. As a result, their performance in such contexts may be far from optimal.

A second criticism of such tests is that they typically have little utility for guiding classroom-based educational interventions. Static-intelligence test scores emphasize an end product (i.e., a set of scores of some kind) rather than an ongoing process (e.g., a description of problem-solving processes). Thus, test performance on these measures sheds little light upon how an individual learns, or fails to learn. For this reason, it is argued that rather than focusing exclusively on the end result of previous learning, cognitive testing should be concerned with underlying latent abilities, the child's potential. Testing while learning (Sternberg and Grigorenko, 2002) or learning while testing, usually called dynamic testing or dynamic assessment (DA), is hypothesized as yielding more predictive and diagnostic information about the level of the individual's intellectual functioning and the cognitive processes and strategies they employ. While dynamic approaches are employed with adults in a wide range of diagnostic activities, even for broader clinical purposes, the focus in the present article is upon the use of this approach with respect to children's learning.

A dynamic test procedure is primarily aimed at examining how training or individualized instruction during the test session(s) can lead to an improvement in a person's performance. While differing in structure, content, and degree of standardization, all such tests have one element in common; children are given hints or training to enable them to show individual differences in progress made during the process of solving a variety of cognitive tasks.

Drawing upon the Vygotskian concept of the zone of proximal development, the approach seeks to go beyond the assessment of existing capacities to examine the individual's learning potential. Rather than focusing upon where the child is now, given her previous educational experience, educationalists and psychologists should be concerned with where they can be tomorrow, given appropriate forms of education. It is considered that such potential is more likely to be revealed within a reciprocal testing situation in which assistance is offered as necessary and close examination is made of the individual's response to this (i.e., his/her modifiability) and, subsequently, his/her ability to transfer new learning to other items and problems. The aim is not to achieve permanent progression but, rather, to detect individual differences in progression during the test sessions. Clinicians typically seek to gain insights into the way in which these differences emerge, the variability or constancy in children's use of cognitive and metacognitive strategies, differences in the ways they verbalize their solving processes, their reactions to failure and success, and how they react to the provision of assistance.

Forms of Assistance

Dynamic testing or assessment can be seen as an umbrella term describing a heterogeneous group of approaches all linked by one essential element: instruction and feedback are provided as part of the assessment process (for an outline summary of major approaches, see Elliott, 2003: 29–32). Depending on the theory behind the specific model of dynamic testing or assessment, the form of feedback is either fixed, that is, it is the same for all children tested, or individualized, that is, it is applied in a tailored fashion contingent upon the child's ongoing performance in the test situation. In this latter situation, the nature and amount of assistance provided depend upon individual differences manifested within the assessment context. This instructional component within the testing situation represents a significant departure from most conventional testing procedures which usually prohibit any forms of assistance other than strictly delineated inputs geared to assist administration and maintain rapport.

Dynamic approaches can be differentiated on the basis of the timing of the provision of assistance during assessment. Sternberg and Grigorenko (2002) differentiate between the two most common forms of intervening – the sandwich and the cake formats. In the former, a pretest is administered without any assistance. After this, instruction is provided geared to helping the child maximize his or her performance. This intervention can be given in group or individual testing settings and may be standardized and scripted or left to the clinician's own judgment. Subsequently, a posttest is provided. While some testers seek to measure the differences between pre- and posttests as a measure of

learning potential, it has been noted that these gain scores pose a number of significant measurement difficulties (e.g., Sternberg and Grigorenko, 2002). For this reason, it is generally accepted that it is wiser to use posttest scores in isolation. The alternative approach – the cake format – dispenses with the pretest/posttest format altogether. Here, the testee receives assistance as soon as a significant difficulty is encountered, item by item during a single testing period. Again, intervention can be based upon a structured approach with a predetermined hierarchy of hints or individualized according to the particular understandings and perceptions of the tester. However, since there is no baseline measurement, it is impossible to predict the level of performance the child would have attained without such assistance.

Dynamic tests also differ in their content. While some have drawn upon academic areas such as mathematics, reading, and spelling, and others upon speech and language, including second-language learning, the majority of researchers in DA utilize items very similar to those used in intelligence tests. These are typically abstract in nature and are considered to tap cognitive processes deemed to underpin learning and problem solving. Many of these involve inductive reasoning tasks that require the generation of an underlying rule based upon systematic observation and comparison, in particular the examination of communalities and differences. This rule can then be used to solve future problems. Typical tasks often involve classification, series completion, or analogical reasoning using either verbal or spatial–figural modalities.

DA: Standardized or Clinical?

Yet another way to categorize dynamic approaches is to divide them into those which are predominantly clinical in their operation and others which seek to conform to rigors comparable to those of the traditional psychometric approach. To a significant extent, tester preferences reflect different agendas, purposes, and roles. Those who adhere to a traditional scientific approach tend to emphasize the importance of a strictly standardized approach in order to ensure test reliability and validity. Here, when errors are made, assistance typically takes the form of a series of scripted hints or prompts that offer progressively more guidance.

Those who emphasize a more clinical approach, usually adopting the cake format (e.g., Feuerstein *et al.*, 1979) contend that a heavily standardized procedure reduces the unique diagnostic contribution that dynamic approaches can offer. They argue that in order to maximize the power of the assistance offered, it is necessary to provide finely grained scaffolded or mediated instruction, aiming to maximize feelings of competence and efficacy

that cannot be detailed in advance as a set of preordained procedures. For this reason, the nature and extent of the assistance provided to the child is individually determined by the tester and is likely to vary from one clinician to another, and perhaps, from day to day with the same child and clinician. While this necessarily impacts upon test reliability (an element that is a *sine qua non* for most psychologists), advocates of Feuerstein's approach argue that psychometric concerns are less important than the quality of clinical insights or the magnitude of change in the child's performance.

Believing that the performance of many students on intelligence quotient (IQ) tests belied their true potential, particularly when they were from disadvantaged backgrounds, Feuerstein and his colleagues set out to devise measures that could assess the individual's modifiability. To aid in this process, a detailed list of cognitive functions (e.g., lack of, or deficient need for precision and accuracy in data gathering) were identified which were assessed by a suite of specially designed subtests that constituted the learning potential assessment device' (Feuerstein *et al.*, 1979). Although the procedure is very different, the content of many of these subtests is similar to that found in traditional IQ tests.

While the work of Feuerstein and his colleagues has had a seminal influence upon many psychologists and educationalists, his approach has failed to gain widespread acceptance by researchers or clinicians. Criticisms include the suggestion that his theory and techniques have changed little over time and thus fail to reflect more recent thinking and theorizing. Others, for example, Büchel and Scharnhorst (1993) and Sternberg and Grigorenko (2002) contend that a lack of conceptual clarity renders operationalization difficult and validation of the theory problematic. They point to the lack of a standardized approach (in test administration, analysis, and interpretation of results), weak test-retest reliability, and the poor quality of empirical studies examining the approach. Others have queried low rates of interrater reliability, which, perhaps, might be anticipated given this approach. In order to address these criticisms, some have tried to modify Feuerstein's approach by employing rather more standardized forms of assistance (e.g., Tzuriel, 2001; Lidz, 2000) although it should be noted that their use of these measures is primarily to inform intervention.

At the other, fully standardized end of the dynamic testing and assessment continuum are approaches that seek to assess learning potential that is defined in terms of the maximal degree of improvement in performance that can be achieved by training. Irrespective of whether these learning potential tests have a sandwich (pretest – training – posttest) or a cake (training-within-test) format, children are provided with the same kind and amount of instruction. This form of help appears to offer heightened predictive validity but the use of such high levels

of structured response limits its value for educational intervention.

In other tests, the nature and amount of standardized hints are more varied. Here, short-term dynamic tests employ only one test session comprising all stages – pretest, training, and posttest. Training consists of structured feedback indicating whether the response is correct or not, and, in some cases, is supplemented by very short explanations. This approach can be found in recent dynamic tests such as the Leipzig Learning Test (LLT), the *Evaluación del potencial de Aprendizaje*, and the Analogical Reasoning Learning Test (ARLT). A further measure, the Learning Potential Test for Ethnic Minorities (LEM), is a Dutch test (more recently extended to Switzerland) that is designed to assess the general cognitive abilities of young children, from ethnic minorities, who have limited knowledge of their host country's language. It provides standardized feedback that consists of repetition of items, and nonverbal feedback or demonstration appropriate to the task.

Other structured approaches provide a series of prompts or hints in response to errors or omissions. An early example of this is the graduated-prompts approach developed by Campione and Brown (1987). This approach involves the ongoing provision of assistance to the child until she is able to solve the test item concerned. Of particular interest, is the child's ability to transfer newly acquired learning to other items. The approach differs in emphasis from that typically used by other dynamic assessors in that, rather than focusing upon the ultimate level of performance that can be achieved with help, it is principally concerned with the amount of assistance that is necessary to achieve prespecified outcomes, and then transfer the learned rules and principles to novel situations. Thus the child's learning potential is not defined in terms of maximizing task performance but rather, is represented by the inverse of the minimal number of hints necessary to reach a specified amount of learning.

Resing's (1993, 2000) approach, using inductive reasoning tasks, similarly seeks to measure the minimum amount of help necessary for the child to succeed in independent problem solving. In order to draw inferences about the child's capacity for improvement (i.e., their zone of proximal development), hints are ordered hierarchically: from general, metacognitive prompts to concrete, cognitive, task-specific ones. After providing the correct answer, the child is asked to verbalize the reasons for her selection. The time taken, the number and type of hints necessary to reach a learning criterion, and posttest scores, are all seen as indicators of learning potential.

The shift toward more adaptive models of testing, in which test items presented to the child vary in difficulty according to their performance, adds to the complexity of the DA process. Computerized approaches such as the adaptive computerized intelligence learning test battery, are

one way of managing this. After correctly solving a number of target items, all operating at a similar level of complexity, the program moves to a higher level. If the child fails on the subsequent item, suggesting that the jump in complexity was too great, the next item presented is reduced in difficulty. After each response, the child receives feedback (correct/wrong), and where an error has been made, a series of prompts is provided until the right solution is found. The program can provide a detailed printout of the child's performance with information on the number of tasks completed, the number of prompts required to solve the questions, the latency times for each item (from presentation of the item until the child enters a response), and the total time taken to complete the test.

While most studies in DA have taken place in the Western world, it is possible that these approaches may have particular value in countries where children have very limited experience of traditional intelligence test measures. In a series of research studies in Africa, Sternberg, Grigorenko, and colleagues have shown how using various tests, normally employed in static mode, in a dynamic fashion can yield more meaningful information about children's intellectual potential.

What Are Dynamic Approaches Seeking to Achieve?

Despite the theoretical appeal of dynamic testing and assessment, its relatively long history – it is only slightly younger than the notion of intelligence testing itself – and growing interest in the topic, it is notable that the approach occupies relatively little space in the assessment literature and is rarely practised by clinicians. It is this relatively low profile that sees it erroneously as a novel assessment paradigm.

Elliott (2003) argues that the lack of recognition and impact of DA is, in large part, a consequence of the professional psychologist's long-standing emphasis upon classification. Some researchers and clinicians continue to see the approach as primarily representing a more sophisticated means of answering the questions for which IQ tests were originally designed. One of the key roles of the educational (school) psychologist has long been to obtain a measure of the child's intellectual ability in relation to others. Those scoring significantly below average (traditionally two standard deviations or more) were typically deemed to require some form of special education. While a seductively simple process, it is essential that data obtained from such measures do not misrepresent the individual's true capacities, especially where these are used to guide decision making about the child's educational future.

The early designers of DA (e.g., Feuerstein, Budoff) were driven by a desire to ensure that children's true potential was not underestimated because of his/her unfamiliarity

with the artificial test situation or the test content. Thus, their primary concern was to improve upon the level of prediction afforded by traditional IQ tests. Nevertheless, this is a contested area with some (e.g., Sternberg and Grigorenko, 2002) remaining unpersuaded, despite the claims of leading researchers in this field (Beckmann, 2006). To some extent, the debate is muddled by the use of differing external criteria: teacher grades, test scores, and problem-solving tasks. It is also likely that prediction will vary according to the groups that are under consideration. Beckmann (2006), for example, argues that traditional IQ test scores are likely to provide as sound an indicator of intellectual ability as dynamic measures in the case of those who have grown up under optimal conditions and who have had every opportunity to achieve a level of performance that reflects their true level of ability. In contrast, the greater predictive potential of DA is more likely to be evidenced when employed with minority or disadvantaged children.

A particular difficulty with respect to prediction concerns the role of the environment in maximizing the child's potential. Consider, for example, an underperforming child whose true potential is revealed by a dynamic measure. Should the child's environment not be subsequently modified or enriched, it is very likely that underperformance will continue and the child's potential will remain unrealized. Thus the predictive power of the measure is inherently intertwined with practitioner action.

The suggestion that notions of potential should be used to make decisions about special education (and, in a few cases, eligibility for gifted programs) has been shown to be oversimplistic as it makes sweeping and sometimes inappropriate assumptions about the nature of provision and resource levels in different types of classrooms, for example, that special-education settings will typically have lowered expectations of the children in their care (Elliott, 2003). However, given the widespread shift toward more inclusive approaches to education, particularly for those with learning difficulties, this debate is now less relevant.

In an inclusive education system, the classification function of the psychologist is less important and the value of psychological assessment would appear to lie increasingly in its capacity to optimize educational intervention, a task for which IQ tests are largely unhelpful. Thus, English educational psychologists' reports have been criticized (e.g., Wood, 1998) for being of little use to practitioners on the grounds that they tended merely to repeat what had been reported by others, and typically offered little insight or understanding about specific difficulties or guidance as to how to overcome these. As another commentator stated, education services appear to need more psychology, not necessarily more psychologists. Dynamic measures, with their emphasis upon children's response to intervention

appear to offer powerful ways by which the psychologist may address such criticisms.

Rendering observations from DAs in clinical settings into meaningful prescriptions for practice in busy classrooms is, however, rather more difficult than has often been acknowledged. While many leading proponents of DA continue to be heavily dependent upon Feuerstein's work, his theory and concepts are complex, and classroom implications are not always easy to provide or make accessible to teachers. Other clinicians appear to be drawing upon a broader and more eclectic theoretical base. Elliott (2000), for example, argues that psychologists can draw upon dynamic approaches to provide helpful accounts about the children's cognitive functioning, their dispositions, motivations and effect, and their various responses to differing forms of adult support and guidance. Such information can complement that which focuses upon academic mastery such as curriculum-based assessment (Jeltova *et al.*, 2007). In fact, the concept of response to instruction, currently being featured in the USA's system of eligibility for special education, is closely related to that of DA (Grigorenko, 2009).

However, the requirement that psychologists should be classroom-based problem solvers is tempered by their continuing role as gatekeepers to resources. Decision making based upon short and standardized procedures is seemingly more easily addressed by traditional psychometrics. The more complex, qualitative accounts that emerge from DAs may not be considered helpful by administrators whose primary concern involves resource allocation based upon clear and reliable diagnostic criteria. Furthermore, it should be recognized that the power and influence of educational (school) psychologists has long been rooted in their classificatory and diagnostic functions. Thus, greater challenge to the professional autonomy of psychologists, together with an increasingly litigious culture, may render conservative and defensive practices that draw upon longstanding approaches to assessment, more attractive.

Although it is widely accepted that dynamic measures should complement rather than replace conventional tests, it remains the view of most researchers and clinicians that this approach requires a strong psychometric foundation involving the use of standardized forms of assistance. While some accept Feuerstein's (1979) argument that the key issue for the purposes of validation is whether student functioning, informed by DA, can be shown to have improved after intervention, it is unlikely that anecdotal evidence will result in widespread acceptance. To date, we continue to have little strong supportive evidence beyond case-study exemplars, and calls for controlled studies that can provide strong support for the incremental benefit of interventions based upon dynamic, as opposed to static, assessments (Elliott, 2003) continue to go unheeded.

Future Possibilities?

Where does the future lie for DA? The development of reciprocal scaffolding techniques (Granott, 2005) and graduated-hint structures (Resing *et al.*, in press), in combination with advances in computerized testing (e.g., Beckmann, 2006), may provide new test instruments that can offer both adaptive and standardized means of investigating (variations in) item-solving processes and learning (Sternberg and Grigorenko, 2002; Resing, 2006). The difficult task ahead is to attune the hint structures of dynamic tests to classroom instructional processes and strategies. Reciprocally, teaching strategies may also be moderated in the light of the results of the dynamic test. As a result, teachers and other classroom-based practitioners would be offered a renewed view of how best to elicit optimal change in the child's learning. Such an eventuality would signal a successful paradigm shift for DA from its former (aspired) position as a superior intelligence test, to becoming a tool by which psychology, and psychologists, can offer teachers more meaningful insights for educational practice.

See *also*: Assessment and the Regulation of Learning; Formative Assessment.

Bibliography

- Beckmann, J. F. (2006). Superiority: Always and everywhere? On some misconceptions in the validation of dynamic testing. *Educational and Child Psychology* 23(3), 35–49.
- Büchel, F. P. and Schamhorst, U. (1993). The learning potential assessment device (LPAD); the discussion of theoretical and methodological problems. In Hamers, J. H. M., Sijtsma, K., and Rujssenaars, A. J. J. M. (eds.) *Learning Potential Assessment: Theoretical, Methodological and Practical Issues*, pp 83–111. Amsterdam: Swets and Zeitlinger.
- Campione, J. C. and Brown, A. L. (1987). Linking dynamic assessment with school achievement. In Lidz, C. S. (ed.) *Dynamic Assessment: An Interactional Approach to Evaluating Learning Potential*, pp 82–109. New York: Guilford Press.
- Elliott, J. G. (2000). The psychological assessment of children with learning difficulties. *British Journal of Special Education* 27, 59–66.
- Elliott, J. G. (2003). Dynamic assessment in educational settings: Realising potential. *Educational Review* 55, 15–32.
- Feuerstein, R., Rand, Y., and Hoffman, M. B. (1979). *The Dynamic Assessment of Retarded Performers: The Learning Potential Assessment Device, Theory, Instruments, and Techniques*. Baltimore, MD: University Park Press.
- Granott, N. (2005). Scaffolding dynamically towards change. Previous and new perspectives. *New Ideas in Psychology* 23, 140–151.
- Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities* 42(2), 111–132.
- Jeltova, I., Birney, D., Fredine, N., *et al.* (2007). Dynamic assessment as a process-oriented assessment in educational settings. *Advances in Speech-Language Pathology* 9, 1–13.
- Lidz, C. S. (2000). The application of cognitive functions scale (ACFS): An example of curriculum-based dynamic assessment. In Lidz, C. S. and Elliott, J. G. (eds.) *Dynamic Assessment: Prevailing Models and Applications*, pp 407–439. New York: Elsevier.

- Resing, W. C. M. (1993). Measuring inductive reasoning skills: The construction of a learning potential test. In Hamers, J. H. M., Sijtsma, K., and Ruijsseenaars, A. J. J. M. (eds.) *Learning Potential Assessment. Theoretical, Methodological and Practical Issues*, pp 219–241. Amsterdam: Swets and Zeitlinger.
- Resing, W. C. M. (2000). Assessing the learning potential for inductive reasoning in young children. In Lidz, C. S. and Elliott, J. G. (eds.) *Dynamic Assessment: Prevailing Models and Applications*, pp 229–262. Oxford: Elsevier.
- Resing, W. C. M. (2006). Using children's "ability to learn" in diagnosis and assessment. *Educational and Child Psychology* **23**, 6–11.
- Resing, W. C. M., Tunteler, E., De Jong, F. J., and Bosma, T. (in press). Dynamic testing: Individual differences and learning patterns in indigenous and ethnic minority children. *Learning and Individual Differences* **19**.
- Sternberg, R. J. and Grigorenko, E. L. (2002). *Dynamic Testing: The Nature and Measurement of Learning Potential*. New York: Cambridge University Press.
- Tzuriel, D. (2001). *Dynamic Assessment of Young Children*. New York: Kluwer.
- Wood, D. (1998). Okay, then: What do EPs do? *Special Children* **5**, 11–13.
- Haywood, H. C. and Lidz, C. S. (2007). *Dynamic Assessment in Practice: Clinical and Educational Applications*. Cambridge: Cambridge University Press.
- Hessels, M. G. P. (2000). The Learning Potential Test for Ethnic Minorities (LEM): A tool for standardized assessment of children in kindergarten and the first years of primary school. In Lidz, C. S. and Elliott, J. G. (eds.) *Dynamic Assessment: Prevailing Models and Applications*, pp 109–131. Oxford: Elsevier.
- Hessels, M. G. P. and Hessels-Schlatter, C. (2002). Learning potential in immigrant children in two countries: The Dutch and the Swiss-French version of the learning potential test for ethnic minorities. In Van der Aalsvoort, G. M., Resing, W. C. M., and Ruijsseenaars, A. J. J. M. (eds.) *Learning Potential Assessment and Cognitive Training*, pp 227–245. New York: Elsevier.
- Hessels-Schlatter, C. (2002). Moderate mental retardation and learning capacity: The analogical reasoning learning test. In Van der Aalsvoort, G. M., Resing, W. C. M., and Ruijsseenaars, A. J. J. M. (eds.) *Learning Potential Assessment and Cognitive Training*, pp 249–271. New York: Elsevier.
- Lidz, C. S. and Elliott, J. G. (eds.) (2000). *Dynamic Assessment: Prevailing Models and Applications*. Oxford: Elsevier.
- Lidz, C. S. and Elliott, J. G. (eds.) (2006). Use of dynamic assessment with gifted students. *Gifted Education International* **21**, 151–161.
- Van der Aalsvoort, G. M., Resing, W. C. M., and Ruijsseenaars, A. J. J. M. (eds.) (2002). *Learning Potential Assessment and Cognitive Training*. New York: Elsevier.

Further Reading

- Fernandez-Ballesteros, R. and Calero, M. D. (2000). The assessment of learning potential: The EPA instrument. In Lidz, C. S. and Elliott, J. G. (eds.) *Dynamic Assessment: Prevailing Models and Applications*, pp 293–323. Oxford: Elsevier.
- Feuerstein, R., Feuerstein, R. S., Falik, L. H., and Rand, Y. (2002). *The Dynamic Assessment of Cognitive Modifiability. The Learning Propensity Assessment Device: Theory, Instruments and Techniques*. Jerusalem: ICELP Press.
- Guthke, J. and Beckman, J. F. (2000). The learning concept: Application and practice. In Lidz, C. S. and Elliott, J. G. (eds.) *Dynamic Assessment: Prevailing Models and Applications*, pp 17–69. Oxford: Elsevier.

Relevant Websites

- <http://www.dynamicassessment.com> – Dynamic Assessment.
- <http://www.iacep.coged.org> – International Association for Cognitive Education and Psychology.
- <http://www.mindladder.com> – International Center for Mediated Learning.
- <http://www.icelp.org> – International Centre for the Enhancement of Learning Potential.

Instructional System Provided Feedback

H F O'Neil, University of Southern California, Los Angeles, CA, USA

S Chuang, National Taiwan Normal University, Taiwan, Republic of China

T Huang, Wenzao Ursuline College of Languages, Taiwan, Republic of China

© 2010 Elsevier Ltd. All rights reserved.

The use of feedback to inform learners of their state and progress in learning is an essential condition of learning. Feedback is an important element for learning, motivation, and self-regulation. With feedback, learners can restructure their knowledge, correct what they do wrong, and further support their metacognitive processes. Extensive research has been done in the area of feedback in the 1960s and 1970s, mostly in the behavioristic tradition. More recently, feedback research has declined and what is published reflects either a cognitive or situated learning focus.

The purpose of this article is to review the literature on feedback provided by an instructional system. Instructional systems, defined broadly, would include teacher-led classroom environments. However, our use of the term implies some technological aspect. The article is organized into several parts. First, the authors address the definition, different types of feedback in learning situations to include both system-assigned versus learner-controlled feedback, and timing of feedback. Subsequently, representation of feedback, feedback provided by technology-based system, and dynamic testing are discussed. Finally, the authors describe a current implementation of an instructional system that provides formative assessment feedback. The authors' intent is to describe the big ideas in instructional system feedback and provide a representative research study. Only seminal or recent critical references are provided.

Definition and Feedback Type and Timing

Historically, research on feedback started with feedback provided by teachers. Pintrich and Schunk (1996) summarized and distinguished four types of teacher feedback. Their distinction was made based on the type of information included in the feedback. The types were performance feedback, motivational feedback, attributional feedback, and strategy feedback. **Table 1** gives a definition and examples for each type of teacher feedback. The table has been adapted from Pintrich and Schunk (1996: 336).

Even though there were four types of teacher feedback identified by the researchers (Pintrich and Schunk, 1996), their overall implementation in the classroom is not equal. Research found that among teachers who were observed over a 10-week period, there was a significantly high use

of performance feedback, whereas significantly lower use of attributional feedback. The situations in the classroom and in the instructional system situation were the same. Most feedback in technology-based systems is performance or motivation oriented.

According to the amount of information included in the performance feedback, feedback can be categorized into verification, elaboration, and adaptation. Verification feedback includes knowledge of response feedback, knowledge of correct-response feedback, and answer-until-correct feedback.

Knowledge of response only informs the learner whether the answer is correct or incorrect. Knowledge of correct response not only points out whether the answer is correct or incorrect, but it also goes on to inform the learner of the correct answer. Answer-until-correct feedback only informs the learners when they come up with a correct answer. Elaboration feedback provides an explanation for why the answer is correct or incorrect, in addition to knowledge of correctness of the response. Adapted feedback was designed in computer-based instruction with attention to customizing feedback for the users' needs rather than just one fixed form of feedback. Task-specific adapted feedback, which provides more information to students, has the most effective influence on students' performance. In general, elaboration feedback is better than verification feedback. Adapted feedback is also more effective than elaboration feedback. Task-specific adapted feedback works the best (Chuang and O'Neil, 2006).

Feedback timing can be immediate or delayed. Kulik and Kulik (1988) defined immediate feedback as performance feedback provided to a learner or examinee as quickly as the computer's hardware and software allow after a test item or after a whole test. Delayed feedback was defined by them as performance feedback provided to a learner or examinee after a specified delay interval following a test. Furthermore, the idea of immediate feedback and delayed feedback is relative rather than absolute. Immediacy and delay only have meaning when they are in contrast with each other. Immediate feedback can result in a faster rate of task acquisition. Delayed feedback is more effective when used with feedback focusing on task processing (Hattie and Timperley, 2007). Immediate feedback produces a better result for automatic processing, while delayed feedback is favored in deliberative and effortful processing. When students

Table 1 Teacher feedback type

Type	Description	Examples
Performance	Performance feedback provides students with information on accuracy of work; it may include corrective information	"That's right." "That's correct. Do not forget to deduct the number."
Motivational	Provides information on progress and competence; may include social comparison and persuasion	"You are getting better." "Great job."
Attributional	Links student performance with one or more attributions	"You've been working hard and you are doing well."
Strategy	Links student performance with one or more strategies used	"The tutoring after school works, you are now very good at this."

have a low mastery level of the material or acquire the material in the beginning, giving them immediate feedback may be more appropriate. Providing delayed feedback results in a better outcome for long-term retention and students who have prior knowledge or a high mastery level.

Several interpretations were proposed to explain why delayed feedback was more helpful in test-acquisition conditions. A theoretical possibility is that delayed feedback in the task serves as a repetition that reinforces the item again for the learners. Therefore, when compared to immediate feedback, the delayed feedback group was provided two chances for processing and therefore had better performance. There is some empirical support for this theory provided by Clariana *et al.* (2000). Among items with differential difficulty, they found retention of initial learning responses was greater for delayed feedback compared to immediate feedback across all items; however, the result was more pronounced with difficult items.

King *et al.* (2000) found similar evidence for an interaction of feedback type (immediate or delayed) with the nature of the learning task (rote or requiring effortful processing), immediate feedback being more effective with verbal learning tasks and delayed feedback more effective with concept learning or problem-solving tasks. The presumed mechanism for the effect is that a delay allows more time for metacognitive activities, identifying and filling knowledge gaps, and restructuring knowledge, activities that make the delayed feedback more effective.

Representation of Feedback

Representation is an important aspect in feedback research as well. Basically, two categories of feedback representation were frequently compared. They were visual and verbal representations. A visual form of feedback may include feedback in a graphic format, while the verbal form usually contains linguistic information coded in texts or words. Before discussing the visual and verbal feedback effects, an introduction of cognitive load theory and dual-coding theory is indispensable.

Cognitive load theory has two assumptions (Sweller, 2006). The first is that the human cognition process has two parts: working memory and long-term memory. The working memory is limited in terms of its capacity and duration, while the long-term memory is unlimited. The second assumption is that schemata are cognitive structures that allow information temporarily stored in working memory to be transferred into long-term memory and thus reduce working memory load. Thus, it would be expected that when presenting more information in the feedback than the cognitive load can handle, the information is lost rather than being used by the learners. This situation is called cognitive overload. Cognitive load theory offers many helpful solutions to overcome this problem. For example, two kinds of feedback representations are compared and discussed: verbal mode and visual mode. In substance, visual-audio presentation or auditory presentation of text is superior to text-only visual form. When instructors use a text-auditory presentation of the to-be-learned text, text feedback should not appear on a computer screen but audio should be used (Mayer, 2006).

As another example, research in Sweller's lab and University of New South Wales, Australia, demonstrated that by appropriately mixing auditory and visual representation modes, cognitive load was reduced for mathematics learning. Their research also revealed that a visual-audio presentation mode promotes a deeper understanding of materials than a visual-visual presentation mode. They experimented with the effects of animated graphical feedback and textual feedback on undergraduates in a computer-based simulation program concerned with the laws of motion. They found that when given animated graphical feedback, subjects performed better, completed the game task in less time, and were less frustrated.

Likewise, O'Neil *et al.* (2000) examined training applications of a virtual environment simulation. They chose understanding an F-16 aircraft's fuel system as the learning task. They fixed most basic instructional design variables and only allowed the feedback representation to vary (e.g., audio and text feedback). They have shown, consistently with other Mayer research, that the audio group did better than the on-screen pop-up text group on three measures: the transfer test, the matching test, and the knowledge mapping test.

Another theoretical framework is the dual-coding theory proposed by Paivio (1991). Dual coding suggests that human cognition is divided into two processing systems: visual and verbal. The visual system deals with graphical information processing and the verbal system deals with linguistic processing. These two systems are separate and are activated by different information. Paivio also suggested that words and pictures are coded and decoded mentally in different ways. For example, pictures are most likely to be coded both visually and verbally, whereas words are usually coded verbally rather than visually. Following these assumptions, Paivio went on to propose that if the information to be processed is coded both visually and verbally, the acquisition chances for learners would be doubled because the information is presented physically as a whole. Cognitive load theory and dual-coding theory, together, have been widely used to explain feedback presentation research. In general, these theories have much empirical support.

Feedback Provided by Technology-Based System

Computerized instructional systems with interactive multimedia features are potentially optimal for feedback delivery as all types, delays, and representations are available. A critical issue is whether feedback is assigned by the system, or system control, or the learner chooses the type and characteristics of the feedback, or learner control. The research evidence (Clark, 2005; Niemiec *et al.*, 1996) is clear, system control works better than learner control in general. The exception is that for students of high aptitude, high prior knowledge, and high motivation, learner control is appropriate.

Traditionally, the majority of computer-based instruction programs that include feedback mostly utilize performance feedback. However, a new trend in educational research is to include strategy feedback into the instructional system. Strategy feedback has been used in several studies and was effective in helping students learn (Chuang and O'Neil, 2006). In these studies, strategy feedback in a collaborative computer-based instruction task was provided to students to help them complete a science concept mapping task on a computer networked system. The adapted feedback group did better on the concept mapping than the knowledge of response feedback group. When search information was added as task-specific feedback while keeping the adapted feedback, Chuang and O'Neil found that task-specific feedback was better than adapted feedback when it comes down to content understanding and problem-solving skills, such as general searching and Boolean search.

Based on these studies, the National Center for Research on Evaluation, Standards, and Student Testing

(CRESST) researchers sought to investigate the effect of after-action review feedback on team performance in the same collaborative problem-solving tasks, that is, a computer-based searching and knowledge-mapping task. The after-action review (AAR) is a method for providing feedback to learners commonly used in military team training, for example, following a simulated tactical exercise, the action. As it is not possible to interrupt the training exercise to provide feedback on specific responses, the AAR is necessarily delayed. The AAR reviews what was supposed to happen, identifies what actually happened during the execution, and stimulates team discussion on why it happened. During the discussion, team members learn from their mistakes and benefit from the lessons learned by other team members. The AAR becomes the bridge between the completed training event and the next training event, providing learning on how to improve that enables leaders to fix training weaknesses (Meliza and Goldberg, 2008).

The military considers AAR an effective form of feedback. The literature on delay of feedback would indicate that delay strengthens the effect of feedback when the learning task requires effortful cognitive processing, for example, problem solving.

Given the United States military's success with AAR in team training, and the suggestion that delayed feedback may be more effective with learning tasks requiring effortful problem solving, AAR feedback may be effective in distance learning for the training of teams on a collaborative task with high germane cognitive load. CRESST research indicated that after receiving the AAR feedback, the AAR group, compared to a control group, indicated significantly more content understanding, problem-solving strategies (e.g., searching), and an increase in team communication. CRESST researchers also manipulated the delivery mode of the AAR feedback. The feedback was presented to one group in traditional text format and to another with partial text plus partial audio. Their study indicated that the text-plus-audio feedback had significant effect on the content understanding as measured by their map scores compared to the text-only group. Moreover, compared to the control group the text-plus-audio-feedback group performed significantly better on all measures. These results are similar to those reported by Mayer (2006).

As Mayer (2006) pointed out, the cognitive processing differs for processing spoken words versus printed words. For the text-only condition, the feedback information often exceeded the capacity of the visual working memory. In contrast, for spoken words, a person hears it through his/her ears. There is a hypothesized aural working memory. For printed words, a person first sees it through his/her eyes. There is a hypothesized visual working memory. Thus, this textual spoken condition lessens the impact on working memory and improves performance.

Effects of Feedback in Dynamic Testing

Another type of feedback in assessment that has been investigated is dynamic testing. According to Grigorenko and Sternberg (1998), "Dynamic testing is a collection of testing designed to quantify not only the products or even the processes of learning but also the potential to learn" (p. 75). In order to fulfill the claims made for it, dynamic testing involves testing not only end products but also learning processes at the same time. This type of testing is quite different from traditional, static testing which only assesses the learned end products. Another difference between dynamic testing and static testing is the role of feedback. In traditional static testing, feedback about the performance is usually not given during the test. In dynamic testing, feedback is given during the test to help assess learning. In dynamic testing, an examiner presents a sequence of gradually more difficult tasks. After each performance from the student, the examiner gives the test-taker feedback and continues until the examinee either solves the problem or chooses to give up. The basic goal of dynamic testing is to see when feedback is given whether test-takers change and how they change. This is done through provision of feedback. However, there are no agreed-upon ideas about how much information should be included in the feedback. Currently, different approaches of dynamic testing vary in the amount of information contained in the feedback (Grigorenko and Sternberg, 1998). In general, dynamic testing increases learning.

Future Trends: The Formative Assessment Challenge

Formative assessment has been a characteristic of technology-based systems from their initial implementation in the 1950s. More recently, there has been renewed interest in formative assessment in teacher-led classroom environments. For example, the POWERSOURCE intervention (Baker, 2007) is a formative assessment classroom strategy that can be integrated with an ongoing mathematics curriculum to improve teachers' knowledge and practice and, in turn, student learning. In other words, POWERSOURCE includes both a system of learning-based assessments and an infrastructure to support teachers' use of those assessments to improve student learning. Its initial testing focuses on middle school mathematics, starting in grade 6, and on helping to ensure that students possess the required key understandings for success in algebra. This focus comes from most research findings showing the relationship between unsuccessful command of algebra and subsequent academic performance, including high school graduation and college entry.

Formative assessment with feedback information is extremely influential on learning (Black and Wiliam,

1998). For example, in Black and Wiliam's (1998) meta-analysis of 250 studies, effect sizes averaged between 0.4 and 0.7. In addition, the effect sizes are even more pronounced for low-achieving students, including students with learning disabilities.

Yet, the use in classrooms of formative assessment is not so encouraging. Assessments used in districts and schools are mostly summative rather than formative. The assessment is still considered by most, if not all, as the peripheral step in the curriculum rather than a core, essential element in teaching and learning. Moreover, teachers who should be using formative assessment often lack the ability and time to develop adequate assessment practices. Further, teachers' professional training on the subject matter brings another challenge. According to research, many US teachers have insufficient mathematics knowledge to teach and to assess effectively.

The result of implementing POWERSOURCE indicated an effect size of 0.83. This effect size is large for classroom interventions. Teachers become more proficient in their subject-matter knowledge and more skilled in their use of formative assessment. They also focus their instruction better on key ideas, and, as a result, are more effective in helping students to improve their understanding, as shown by measures of student learning. Ultimately, we expect the improvements in student understanding to drive better performance on No Child Left Behind (NCLB)-mandated state tests, transfer measures, and future coursework.

Therefore, in order for formative assessment to be used efficiently on a large scale in districts and schools, it requires significant changes at the district level, and more importantly, at the teacher and student level. The districts need to make sure that teachers have enough time and resources to support their assessment-related activities. Teachers need to deepen their level of mathematical concept understanding, familiarize themselves with key ideas and skills, and use their knowledge to plan strategically for increasing students' learning. For students, change will involve understanding the feedback information provided by the teachers from the assessment.

Although feedback in general has been demonstrated to have positive effects on learning in many studies, the specific influence of feedback on cognitive processes and achievement is involved with characteristics of students and tasks. Feedback needs to be properly designed to meet the needs of students. Inappropriate use of feedback can inhibit learning (Clark, 2005).

Acknowledgments

The work reported here was supported under the Office of Naval Research contract number N00014-08-1-0126 to the National Center for Research on Evaluation, Standards, and

Student Testing (CRESST), as administered by the Office of Naval Research, and under the Educational Research and Development Centers Program, as administered by the Institute of Education Sciences (IES), US Department of Education contract number R305A050004.

See also: Assessing Group Work; Assessment in the Workplace of Performance, Developing Expertise and Competence; Dynamic Assessment; Formative Assessment.

Bibliography

- Baker, E. L. (2007). The end(s) of testing (2007 AERA Presidential Address). *Educational Researcher* 36(6), 309–317.
- Black, P. J. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice* 5(1), 7–74.
- Chuang, S. H. and O'Neil, H. F. (2006). Role of task-specific adapted feedback on a computer-based collaborative problem-solving task. In O'Neil, H. F. and Perez, R. S. (eds.) *Web-Based Learning: Theory, Research, and Practice*, pp 239–254. Mahwah, NJ: Erlbaum.
- Clariana, R. B., Wagner, D. M., and Lucia, C. R. (2000). Applying a connectionist description of feedback timing. *Educational Technology Research and Development* 48(3), 5–21.
- Clark, R. (2005). Instructional strategies. In O'Neil, H. F. (ed.) *What Works in Distance Learning: Guidelines*, pp 25–40. Greenwich, CT: Information Age.
- Grigorenko, E. L. and Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin* 124(1), 75–111.
- Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of Educational Research* 77(1), 81–112.
- King, P. E., Young, M. J., and Behnke, R. R. (2000). Public speaking performance improvement as a function of information processing in immediate and delayed feedback interventions. *Communication Education* 49(4), 365–374.
- Kulik, J. A. and Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research* 58(1), 79–97.
- Mayer, R. E. (2006). Ten research-based principles of multimedia learning. In O'Neil, H. F. and Perez, R. (eds.) *Web-Based Learning: Theory, Research, and Practice*, pp 371–390. Mahwah, NJ: Erlbaum.
- Meliza, L. R. and Goldberg, S. L. (2008). Impact of after-action review on learning in simulation-based U.S. army training. In Baker, E. L., Dickieson, J., Wulfbeck, W., and O'Neil, H. F. (eds.) *Assessment of Problem Solving Using Simulations*, pp 255–272. Mahwah, NJ: Erlbaum.

- Niemiec, R. P., Sikorski, C., and Walberg, H. (1996). Learner-control effects: A review of reviews and a meta-analysis. *Journal of Educational Computing Research* 15(2), 157–174.
- O'Neil, H. F., Mayer, R. E., Herl, H. E., et al. (2000). Instructional strategies for virtual aviation training environments. In O'Neil, H. F. and Andrews, D. H. (eds.) *Aircrew Training and Assessment*, pp 105–130. Mahwah, NJ: Erlbaum.
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology* 45(3), 255–287.
- Pintrich, P. R. and Schunk, D. H. (1996). *Motivation in Education*. Englewood Cliffs, NJ: Prentice-Hall.
- Sweller, J. (2006). Why understanding instructional design requires an understanding of human cognitive evolution. In O'Neil, H. F. and Perez, R. S. (eds.) *Web-Based Learning: Theory, Research, and Practice*, pp 279–294. Mahwah, NJ: Erlbaum.

Further Reading

- Black, P. J. and Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan* 80(2), 139–148.
- Brackbill, Y., Bravos, A., and Starr, R. H. (1962). Delay improved retention of a difficult task. *Journal of Comparative Psychology* 62, 148–156.
- Clark, K. and Dwyer, F. M. (1998). Effects of different types of computer-assisted feedback strategies on achievement and response confidence. *International Journal of Instructional Media* 25(1), 55–63.
- Foot, C. J. (1999). Attribution feedback in the elementary classroom. *Journal of Research in Childhood Education* 13(2), 155–165.
- Hsieh, I. G. and O'Neil, H. F. (2002). Types of feedback in a computer-based collaborative problem-solving group task. *Computers in Human Behavior* 18(6), 699–715.
- Mayer, R. E. (2001). *Multimedia Learning*. New York: Cambridge University Press.
- Mayer, R. E. and Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology* 90(2), 312–320.
- Ngu, B. H., Low, R., and Sweller, J. (2002). Text editing in chemistry instruction. *Instructional Science* 30(5), 379–402.
- O'Neil, H. F. and Andrews, D. H. (eds.) (2000). *Aircrew Training and Assessment*. Mahwah, NJ: Erlbaum.
- O'Neil, H. F. and Perez, R. S. (eds.) (2006). *Web-Based Learning: Theory, Research, and Practice*. Mahwah, NJ: Erlbaum.
- Paas, F., Renkl, A., and Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist* 38(1), 1–4.
- Scheiter, K. and Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review* 19, 285–307.

Peer and Self-assessment

C Harrison, King's College, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This type of assessment involves students in the assessment process where they can learn to make judgments concerning their own work (self-assessment) and that of their fellow students (peer-assessment). Through self- and peer-assessment, students can become involved in an analysis and constructive criticism of their work. Taking part in peer- and self-assessment practices may also form a learning opportunity for students. This helps students develop a sense of what standards of work is expected of them by their teachers and can begin to acculturate them in the assessment process.

Classroom assessment plays a major role in how students learn and the way teachers teach. Peer- and self-assessment is not evident in all classrooms and, where it does occur, teachers have generally made pedagogic decisions with regard to its importance in learning because it requires time, preparation, and skill to implement successfully. This article attempts to pull together the research on peer- and self-assessment and is organized into three sections titled 'Peer and self-assessment in practice', 'Self-assessment is intrinsic to learning', and 'Self-assessment is part of lifelong learning'.

Peer and Self-Assessment in Practice

Moving to peer- or self-assessment is not a simple process as it requires training and experience before it can be successfully employed. A peer-evaluation experiment in a United Kingdom (UK) higher education institute found peer grading correlated positively with teacher grading. A survey of participants ($n = 70$) found that students believed their work had been marked fairly and the marks should count toward final grades (Conway *et al.*, 1993). The majority of studies compares the reliability of peer and teacher rating and indicates that peer-assessment is generally a useful, valid, and reliable method of assessment. An experiment in self-assessed learning – in which college students set their own goals weekly and prepare a self-assessment open to discussion by peers – encountered some problems in the implementation phase but was found to be the only method of achieving deep, as contrasted with surface, learning (Boyd and Cowan, 1985). This suggests that peer- and self-assessment are not simply replacements for teacher assessment but have merit in themselves.

Peer- and self-assessment can take many forms. There is a substantial amount of literature referring to peer-assessment of performance both in work settings such as hospitals and in theatrical and dance productions. Peer- and self-assessment may or may not involve discussions of quality before an assessment is undertaken. It may simply involve the use of rating instruments or checklists or mark schemes, which may have been designed – by others – before the peer-assessment exercise, and which peers simply apply to the work of their fellow students. Sometimes, self- and peer-assessments in classrooms are carried out on products or completed pieces of work, and here the purpose of making the assessment may be summative or formative, or applied during the process of learning when the main function is to utilize its formative function.

Peer-assessment is an organized event where learners of similar status consider and come to recognize the quality of work of one or more fellow peers. Sometimes, peers are asked to allocate marks or grades to a peer's work and, possibly, to provide some feedback on the judgment that they have reached. Peer-assessment can be formative or summative. Such feedback can be corrective, suggestive of alternatives, or simply confirmative. Peer-assessment is available in high volume, as there are generally more students than teachers in any class, and is generally more immediate than teacher-assessment (Topping, 2009). Peer-assessment is also perceived differently, by the learner, to feedback from their teacher. It is less authoritative and, thus, the student receiving feedback is less likely to simply accept the judgment – preferring to weigh up the options suggested by their peer and make decisions about improvements suggested. In this way, it can be an effective feedback mechanism for students as it can provide scaffolding for improvement by allowing the learner to review and renegotiate the quality of their work.

In many educational institutions, peer-assessment is introduced alongside, or as a precursor to, self-assessment. This is because teachers find it easier to model self-assessment practices through the vehicle of peer-assessment. Often, teachers begin training students in assessment by helping them peer-assess anonymous work so that they start to develop the ways and means of making judgments and of expressing these in an appropriate manner. Developing sufficient assessment-skill and literacy, to convey to others what our judgments are, takes time and experience to develop. By evolving these practices on anonymous work, supported by peers, students' self-esteem and motivation

is not affected. It is also easier for students to peer-assess rather than self-assess as they find it easier to stand back from the work and judge it for its strengths and weaknesses, rather than confounding such judgments with concern about the efforts they have put into the work or how they appear compared to their peers.

Most of the research studies that have focused on reliability of peer-assessment are set in higher education settings, with over 70% evaluated as adequate in assessment quality. Social processes that are indicative of any group of learners can mitigate against the reliability of peer-assessment, with friendship bonds, anxiety about sharing work with fellow students, or commenting on the work of others being likely inhibitors to full engagement with the process. Indeed, some studies noted a tendency for peer marks to bunch around the median. These studies also indicated that reliability was increased when supported by training and accompanied by checklists or exemplified and involved teachers in assisting or monitoring the process (Topping, 2009). While on the one hand, peer-assessment might save the teacher time in that she is not required to mark as much work, time and effort, on the other hand, are required to coach the students in peer-assessment practices.

Over the last decade, there has been a growing and sustained interest in formative assessment following the publication *Inside the Black Box*, by Black and Wiliam (1998b). Self-assessment has a key role to play in formative assessment. Black and Wiliam's (1998a) review of research on assessment and classroom learning showed a strong body of evidence that formative assessment practices raise standards in student learning. The review also indicated that such practices were only weakly developed in most classrooms. Since then, attempts to strengthen and develop formative assessment in classrooms in several countries have been underway.

The term assessment for learning was suggested by the Assessment Reform Group (1999) to try to help teachers differentiate the formative practices in classrooms from the summative ones, which tend to dominate in many classrooms. The Group stated that the research indicates that improving learning through assessment depends on five, deceptively simple, key factors:

- the provision of effective feedback to students;
- the active involvement of students in their own learning;
- adjusting teaching to take account of the results of assessment;
- a recognition of the profound influence assessment has on the motivation and self-esteem of students – both of which are crucial influences on learning; and
- the need for students to be able to assess themselves and understand how to improve.

Self-Assessment Is Intrinsic to Learning

Self-assessment is an essential component in the learning process because it helps students gauge suitable targets for their learning. This is because learning has to be done by learners, it cannot be done for them (Black *et al.*, 2003) and so helping them evolve a means of setting self-targets that fosters this process is useful. The assessment activity could be a test but, more often, it takes the form of a searching question or challenging activity, which encourages the learner to show what they know, what they partly know, and what they do not know (Black and Harrison, 2004). The assessment activity will, therefore, take place alongside or as part of the learning and is used as a tool to diagnose current understanding. Using self-assessment information requires control over one's cognitive activities or metacognition: understanding what strategies and skills are needed for a task and knowing how and when to use them (Brookhart, 2004).

This process helps them form frameworks for assessing their own work and also when applying assessment criteria to new pieces of work. When they can do this, learners are able to steer their own learning toward the learning goal because they have developed self-assessment techniques that help them in learning the ideas and developing the skills for that piece of work at the same time. By doing this, it can help students regulate their future learning and to keep on track and ensure that they are progressing. Through this ipsative process, learners realize what they have learnt and become more aware of what they need to do to improve, and – as a consequence – are often more motivated to learn. However, many studies of self-assessment do not involve students in the selection of criteria and simply ask them to rate themselves according to some preestablished scale. This inhibits students in acquiring a sense of quality that they can use to look at their own work – and that of others – in a variety of contexts. In other words, it prevents students developing self-regulatory skills.

For self-assessment to work, learners need to be able to gauge how much their current ways of working fit with expected practice. Teachers often model their expectations of students by referring to the work of previous students or by producing work themselves that would be typical of student work. This degree of anonymity allows students to engage with the work without conflicting concerns for their own or fellow students' self-esteem and, thus, discussions about work quality are entered into freely. In the UK, this approach has been popular over recent years and has been encouraged through professional development for teachers on strengthening questioning skills and collaborative group-work so that the teachers have the skills to support their students in these discussions (Black *et al.*, 2003).

When students are faced with new experiences, they need to make sense of them, and they do this by linking new ideas with old ones and then self-monitoring and generating new ways of thinking. Language is at the heart of this process. The learner uses talk to engage with the new knowledge and to try to understand it within their own personal frameworks through interactions with other learners and their teacher. They achieve a part of this through comparison with their previous thinking in that area, but the major part of this learning is in negotiating common meaning with others who are also engaged in the learning experience. Learners know when they do not understand the ideas arising within a learning situation. It is only through entering the dialog about shared ideas that the learner can begin to see other aspects of the ideas and so make judgments about where they are in their own sense-making. Nonengagement not only deprives the group of the learner's position but also prevents the learner from revealing their own sense-making to themselves. When the learner offers their emergent understanding, then they can both assess and modify their learning since movement forward – in terms of improvement or progress – rests on the reaction from others in the group. In this way, new knowledge is socially constructed (Vygotsky, 1978) and the immediacy of communication through dialog is essential in achieving this.

It is not simply that the learner hears several voices through dialog but that the ideas from individuals get challenged, molded, and reexamined through the collective voice of the group. Each individual student is able to check their own understanding against that emerging from the group dialog and, thus, self-assessment is at the centre of this process. Isaacs (1999) argues that this is not simply shared knowledge that arises from dialog but a sense of meaning that he terms “collective sensibility” that has evolved from the interactions and from which learners can capture their own sense of understanding. Because each learner brings their own knowledge, aspirations, and limitations to the interactive process, they will have a particular lens on the shared knowledge that arises through the dialog. Therefore, what they focus on, capture, and retrieve from the shared knowledge will depend on the lens they select as well as their capacity to engage with the shared knowledge as it arises. This process helps students regulate their learning. They can consider what they understand and misunderstand against the backdrop of what the group understands. This helps them locate where they are in their learning and where they need to focus to improve. So, for assessment for learning, classroom talk is an exploratory process that enables learners to match their current understanding against that of the groups and through which they might gain better understanding, since the cumulative knowledge of the group is likely to be more than that of any one individual within

the group. In engaging in the classroom talk, the learner is not only revealing to themselves their current understanding but is also providing their teacher with the evidence with regard to this and thus is also setting into action teacher planning for the next steps.

Deakin-Crick *et al.* (2007) review of the impact on students of self- and peer-assessment highlighted the need to include learners in decisions concerning how and what is being studied and insights into how the work is being assessed. In this way, learners become more informed regarding the learning process and play a more equal role with the teacher in making learning decisions. This has important messages for pedagogy since it is the ecology of learning that is called into question here and many teachers will need to change their classroom environments significantly to treat learners as co-constructors of the learning process.

A further factor to consider is self-efficacy, which involves the learner's estimate of their capabilities and their likely success in a particular task. Students with high expectations of their capability will tend to persist with tasks and put in greater effort than those with low self-perceptions. The former are, therefore, likely to be more successful and this feeds into a cycle of success for these learners (Schunk, 1995). The concern is that students with low self-efficacy believe that they cannot achieve and avoid tasks where they cannot quickly gain success, and so their self-assessment skills do not get a chance to develop very much.

Self-Assessment Is Part of Lifelong Learning

The fundamental role of school is to foster lifelong learning and to support children in learning how to learn. The dilemma for teachers is how to provide learning experiences that help children develop the skills they will need in their adulthood for a world that is in the future. Most university prospectuses, at present, include some statements relating to lifelong learning. In general, these focus on two aspects – active learning approaches and reflection. Universities increasingly expect their students to utilize active learning approaches where the student is encouraged to engage, research, and analyze ideas and to develop these collaboratively with others. Second, many courses encourage students to be more reflective by revisiting learning experiences, by making explicit how they approached the learning, and evaluating what they learnt about the task itself and their response to the task. In most institutions, this approach tends to emphasize:

- Learning-how-to-learn skills
- Disposition to learn and take learning opportunities
- Problem-formulation and problem-solving

- Ability to learn with and from others, teamwork
- Ability to identify and access appropriate resources for learning and assessment. (Boud and Falchikov, 2004)

Carr and Claxton (2002) suggest the way forward is to focus on developing learning power, which – they argue – consists of two related facets: skills for learning – which they call capabilities – and willingness and readiness to learn – which they term dispositions. The researchers believe that disposition requires high degrees of resourcefulness and reflectiveness and that these dispositions are inherent and evolvable in all children. Their approach to lifelong learning suggests that these dispositions can be trained and nurtured.

Developing these skills of self-regulation is not simple. Boud and Falchikov (2004) highlights that it is not simply a matter of adding self-assessment to the learning and assessment experiences, but of rethinking learning and assessment from a new point of view and examining the consequences for practice. The problem that arises is that, in learning-how-to-learn and becoming a lifelong learner, the capacity to become an assessor of learning is sometimes omitted, retained by the teacher, or not given sufficient focus. While self-assessment is implicit in learning how to learn, many approaches to develop lifelong learning skills fail to recognize and develop this aspect sufficiently to allow the learner to develop these skills and become more independent in their learning.

Sadler (1989) in his seminal paper on formative assessment highlights the importance of developing assessment skills in learners:

... providing guided but direct and authentic evaluative experience for students enables them to develop their evaluative knowledge, thereby bringing them within the guild of people who are able to determine quality using multiple criteria. It also enables transfer of some of the responsibility for making evaluative decisions from teacher to learner. (p. 119)

To implement this in the classroom, requires skill and time. Rolheiser (1996) suggests a four-stage plan for introducing self-assessment into the classroom that gradually skills the students in assessment practices, while gradually increasing their involvement in the assessment process.

Stage 1 – involve students in defining the criteria.

Stage 2 – explore with students how to apply the criteria to their own work.

Stage 3 – give students feedback on their self-evaluations.

Stage 4 – assist students in developing productive goals and action plans.

Rolheiser found that such approaches helped both teachers and students engage with self-assessment in the classroom and that these processes had a marked effect on intrinsic motivation.

Assessment practices can undermine students' capacity to judge their own work and this can work to constrain the lifelong learning agenda. It encourages students to look to their teachers to make judgments and inhibits them from developing their ability to assess their own learning outcomes. In other words, if students are always passing the responsibility for assessment to others, they may not acquire self-assessment skills.

... if students are to be encouraged to be lifelong learners, they must be weaned away from any tendency towards over-reliance on the opinions of others. Ultimately, in real world contexts, they must be able to judge or evaluate the adequacy, completeness or appropriateness of their own learning, so whatever assessment practices are used must be comprehensible to the learners so that they can be internalised as criteria for critical self-evaluation. (Candy *et al.*, 1994: 150)

Marshall and Wiliam (2006) introduce the notion of apprenticing pupils into the guild knowledge as the chief lever of progression and formative practice. Before learning can even commence, there is a need for learners to identify for themselves what they need to learn – taking into account a range of contextual factors – and to judge for themselves what counts as good work. The work by Orsmond and colleagues suggests that teachers need to carefully scaffold developing assessment ideas to enable students to begin to understand and utilize the criteria for quality (Orsmond *et al.*, 2002). Nevertheless, the goal must always be that students themselves can learn to judge for themselves what constitutes quality work. Students will not develop this if they do not have practice in determining appropriate standards for themselves.

There are substantial barriers to having students accept assessment as a key element of their continuing learning. The greatest of these are their prior experiences of being assessed and the ways in which these experiences influence their expectations and behavior in the present. While this has been little discussed in the literature, it appears to us to be an issue of such importance that it must be considered high on the agenda. Indeed, James (2000) suggests that grades and inadequate feedback to students on specific work can negatively impact on students' overall self-perception and confidence. If peer- and self-assessment is handled in such a way that it negatively affects student self-esteem, then this can demotivate learners and may lead to situations where students refuse to take part in future learning.

Summary

Peer- and self-assessment involve learners in the assessment process and this provides these individuals with an insight into how successful they have been in their performance

or with their product. This way of working can also have a formative function in that it can be utilized by the student formatively so that they become aware of their next steps for improvement. Teachers need to be aware that students require support and training if they are to acquire the necessary skills to assess themselves and others effectively. Involving students in peer- and self-assessment can lead to an understanding of the quality that teachers expect in pieces of work and this can lead to self-regulation of learning, which is the ultimate goal of lifelong learning.

Bibliography

- Assessment Reform Group (1999). *Assessment for Learning: Beyond the Black Box*. Cambridge: University of Cambridge School of Education.
- Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2003). *Assessment for Learning: Putting it into Practice*. London: OUP.
- Black, P. and Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice* 5(1), 7–73.
- Black, P. and Wiliam, D. (1998b). *Inside the Black Box*. London: GL Assessment.
- Boud, D. and Falchikov, N. (2004). Beyond formative and summative assessment: Developing a new agenda for assessment for lifelong learning. *Second Biannual Joint Northumbria/EARLI SIG Assessment Conference*, Norway: University of Bergen.
- Boyd, H. and Cowan, J. (1985). A case for self-assessment based on recent studies of student learning. *Assessment and Evaluation in Higher Education* 10(3), 225–235.
- Brookhart, S. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education* 8(2), 153–171.
- Candy, P., Crebert, G., and O'Leary, J. (1994). Developing lifelong learners through undergraduate education. *Commissioned Report No. 28 National Board of Employment, Education and Training*. Canberra: Australian Government Publishing Service.
- Carr, M. and Claxton, G. (2002). Tracking the development of learning dispositions. *Assessment in Education* 9(1), 9–37.
- Conway, R., Kember, D., Sivan, A., and Wu, M. (1993). Peer assessment of an individual's contribution to a group project. *Assessment and Evaluation in Higher Education* 18(1), 45–56.
- Isaacs, W. (1999). *Dialogue and the Art of Thinking Together*. New York: Broadway Business.
- James, D. (2000). Making the graduate: Perspectives on student experience of assessment in higher education. In Filer, A. (ed.) *Assessment: Social Practice and Social Product*, pp 151–168. London: RoutledgeFalmer.
- Marshall, B. and Wiliam, D. (2006). *English Inside the Black Box*. London: GL Assessment.
- Orsmond, P., Merry, S., and Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education* 27(4), 309–323.
- Rolheiser, C. (1996). *Self-Evaluation. Helping Students Get Better at It! A Teacher's Resource Book*. Toronto, ON: Visutronx.
- Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* 18, 119–144.
- Topping, K. (2009). Peer Assessment. *Theory into Practice* 48, 20–27.
- Vygotsky, L. (1978). *Mind in Society*. Cambridge: Harvard University Press.

Further Reading

- Black, P., Harrison, C., Lee, C., Marshall, B., and Dylan, W. (2002). *Working Inside the Black Box: Assessment for Learning in the Classroom*. London: GL Assessment.
- Coffield, F. (2002). Skills for the future: I've got a little list. *Assessment in Education* 9(1), 39–43.
- Cowie, B. (2000). *Formative Assessment in Science Classrooms*. Unpublished doctoral Thesis. New Zealand: University of Waikato.
- Knight, P. T. and Yorke, M. (2003). *Assessment, Learning and Employability*. Maidenhead: Society for Research into Education & Open University Press.
- Sebba, J., Crick, R. D., Yu, G., et al. (2008). Systematic review of research evidence of the impact on students in secondary schools of self and peer assessment. *Technical Report. Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Portfolio Assessment

V Klenowski, Queensland University of Technology, Brisbane, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

E-portfolios – A digital collection of diverse evidence of an individual's achievements over time involving selection, design, and reflection for a particular purpose and presentation to one or more audiences.

Portfolio – A purposeful collection of process artefacts and products that involves selection of evidence to demonstrate achievement over time and reflection on the process and value of the learning itself.

Introduction

At a time when global uncertainty is paramount and when a new form or re-form of curriculum is emerging – with content displaced by skills and knowledge acquisition by learning – assessment, too, begins to take on a new form or re-form. The focus for assessment has shifted to that which engages and promotes learning as a process rather than an assessment that focuses solely on measuring and reporting learning as product or score. The use of the portfolio for assessment offers the potential for the process and progress – integral to learning – to be included.

Portfolio use for assessment purposes – at the individual or systems level – to support different kinds of decision, process, or action “for which assessment results are used,” has expanded over recent years to include digital media. At present, portfolios are used in all phases of education from early childhood through to higher education and in diverse fields of study such as the arts, engineering, medicine, sport, and recreation. Both students and teachers make use of the portfolio for purposes ranging from improving learning, acquiring and demonstrating new knowledge and skills, to preparing professionally for employment.

This article commences with an analysis of the numerous definitions of the portfolio. An overview of the historical and contextual use of portfolios and the utility of emergent classificatory systems is offered. There has been limited conceptualization and theorizing with regard to the development and the assessment processes of the portfolio and these issues are examined. The article concludes with an examination of digital portfolios and implications for future use.

Definitions

By the early 1990s, the term portfolio had become a popular buzz word in educational contexts. It was first defined as a systematic collection of learned material (McLean, 1990). More comprehensive definitions followed (Paulson *et al.*, 1991; Linn and Baker, 1992; Herman *et al.*, 1992; Forster and Masters, 1996) with the essential features that remain constant in these definitions, including the notion of the portfolio as:

- a collection of student work;
- an opportunity for student selection of items for inclusion in the portfolio;
- active student engagement in the assessment process by demonstrating through evidence what he or she knows and can do;
- student self-assessment on progress and accomplishments; and
- reflection on the process and the value of the learning itself.

The portfolio has also been described as “expansionist” (Mabry, 1999) because of the possibility of submitting a large amount of information and a variety of modes and types of evidence for the assessment of student achievement of knowledge, skills, attributes, or competencies. In this way, the portfolio offers an alternative to more reductionist strategies such as multiple-choice-type testing where achievement is reduced to a score.

Overview

Adult Learning

In an adult learning context, portfolios have been used since the mid-1980s. In England, for example, the field of medical education recognized the value of portfolio-based learning in the professional development of general practitioners, in particular, and for the medical profession, in general. The English National Board for Nursing, Midwifery and Health Visiting developed a framework for post-registration, education, and practice which is used by members of the nursing profession to plan and implement continuing professional education. The portfolio requires record keeping and critical reflective practice and is used as a method of accreditation. The important shift in this context is toward self-directed learning as the

learner assumes responsibility for planning, managing, and evaluating the learning.

Serving teachers too have made use of portfolio assessment since the late 1980s. Teachers' competences – defined as their knowledge, skills, and attributes – are performance assessed using the assessment tool of the portfolio. It is possible to assess the level of attainment, the range of skills, and/or the progression achieved. The complexity of teaching, its multifaceted nature, and the qualities of effective teaching can be captured in the portfolio (Shulman, 1998), if the selection framework offers appropriate guidance for implementation of the portfolio and for judging the evidence included.

In the context of teacher education, the use of the portfolio for the assessment of achievements of pre-service teachers has thrived (Lyons, 1998; Chetcuti, 2007). Pre-service teachers use portfolios to develop their personal philosophy of teaching, to record their learning with regard to teaching, to engage in reflection (Zeichner and Wray, 2001, Reis and Villaume, 2002, Groom and Maunonen-Eskelinen, 2006), and to self-assess (Klenowski, 2002).

Primary and Secondary Education

Records of Achievement (RoA) – implemented in England – were an alternative to traditional forms of assessment in that profiles were used to engage students in dialog about their achievements and to record these in the form of a collection of certificates and/or work samples (Broadfoot, 1988). Arter and Spandel (1992) were the first to provide guidelines to assist in the development of portfolios for teaching and assessment purposes. These included:

- identifying the purpose of the portfolio;
- directions for selecting work samples;
- student participation;
- self-reflection in the purposeful selection of evidence;
- criteria to judge performance; and
- portfolio use to improve and to assess learning.

More specific guidelines for assessing the portfolio developed, in that, the number and nature of the selections of evidence to be included began to be specified to illustrate how the acquisition of certain abilities, understandings, and attitudes related to particular learning experiences could be demonstrated in the collection of evidence.

By the late 1990s, portfolios were beginning to be used for large-scale summative assessment purposes in Vermont (Koretz, 1998), Kentucky (Callahan, 1997), and England, Wales, and Northern Ireland in vocational qualifications programs (Wolf, 1998). Problems emerged from the lack of alignment of curriculum and pedagogy with the purposes and paradigms of educational assessment which included portfolio use. Teachers in their classroom assessment practice need to be aware of the

interrelationships of the structure of the body of knowledge, the curriculum design, and key concepts being taught to make effective use of assessment evidence to understand and promote learning.

Developments in the use of portfolios for large-scale purposes revealed a mismatch of policy contexts and assessment and learning paradigms. The tensions arising from conflicting paradigms meant that the impact of portfolio use for large-scale assessment purposes was limited. Teachers found the workload associated with classroom assessment demanding, particularly while simultaneously trying to meet the competing demands of accountability pressures of summative assessment at a systems level.

Purposes

As is evident from the range of contexts in which portfolios have been used, there is not one portfolio – there are many. They have been used for: learning, teaching, assessing, appraising, promotion, and professional development. These portfolio purposes have been described in summative terms of accountability, certification, selection, promotion, appraisal, and, formatively, to support teaching and learning (Klenowski, 2002). However, as Newton has expressed, attempts to classify and simplify can confuse rather than clarify because such classifications can suggest a level of similarity in the forms and function of the assessment used. It may be more useful to consider each purpose and context for which the portfolio is used as a category in, and of, itself.

Numerous classification systems of portfolio use have emerged in the context of teacher education (Hauge, 2006). One such classification includes: a “learning portfolio” that engages pre-service teachers in “authentic enquiry” (Harland, 2005) with regard to their teaching and records progress; a “credential portfolio” for standards-based assessment of pre-service teachers' readiness for certification and a “professional portfolio” of best work samples for interview and employment (Zeichner and Wray, 2001).

In the context of professional education programs, portfolios have been classified into four types: “a dossier portfolio” of stipulated work samples for selection or promotion to a profession or program; “a training portfolio” of collected evidence – reflective of achievements gained in the program; “a reflective portfolio” of best work to showcase the extent of achievements for promotion or further study; and a “personal development portfolio” a record of personal growth and personal evaluation (Smith and Tillema, 2003).

In higher education, four modes of portfolio implementation have been classified: upon admission, during the course, on entry to the profession, and for ongoing professional development (Meeus *et al.*, 2006). The portfolio types are categorized according to the profession-specific or

learning competencies they aim to develop. It is the learning competencies of: working independently, planning, reflecting, and modifying behavior that are claimed to add value for continuing learning. The purpose of the learning-competencies portfolio is to foster a self-directed learning process. However, these authors claim that if the learning processes involved in attaining the learning competencies are assessed and the same assessor assesses both the profession-specific competencies and the learning competencies, then tensions emerge. The implications are that, in high-stakes assessment, the purpose of the portfolio must be made clear and if, indeed, progress is to be illustrated and assessed then the assessment criteria must be inclusive of this quality. The ethics of allowing the demonstration of progress in a safe and secure manner needs to be respected.

These classifications strengthen Newton's argument that such schemes need to highlight the differences between and within purposes and supports the need to tailor assessment design to assessment purpose.

Portfolio-Development Processes

Most models of portfolio development include processes such as selection, collection, reflection, and connection. The development of the portfolio involves documenting achievements, judging work samples or artifacts for inclusion, self-evaluation of achievements, analyses of learning, experiences, learning strategies, and dispositions. The portfolio is, therefore, significantly more than just a collection of artifacts or assignments. The portfolio-development processes involved are equally important and need to be recognized. They include: critical self-evaluation, interactive learning conversations, reflection on practice and/or learning, and connection to learning experience, theory, and/or practice. These processes are integral to the development of a portfolio and are common for all purposes and audiences for which the particular portfolio can be constructed.

Selection and Collection

All portfolios include a collection of selected work samples or evidence in support of claims with regard to learning and achievements. The varying portfolio purposes of learning, assessment, or employment require different evidence to be collected and selected. A key property in selecting evidence is the relevance – the more relevant it is, the more useful it is for the assessment purpose (Forster and Masters, 1996).

In designing any assessment task, “fitness for purpose” and the positive impact of assessment on teaching and learning need to be considered (Gipps, 1994). These principles apply equally to the design of the selection framework for the development and collection of evidence for

the portfolio itself. Research into the design of a content-selection framework (Simon and Forgette-Giroux, 2000) for the use of the portfolio as an assessment instrument indicates that showcasing or selecting evidence of one's best work is inadequate. For example, in the holistic assessment of a particular competency, cognitive, affective, behavioral, metacognitive, and developmental dimensions need to be identified in the selection framework with explicit examples of entries suggested. This approach maximizes the potential for the assessment of progress and the processes of learning, yet, students require directions with regard to how to use the standards-referenced and selection frameworks when making judgments concerning their own work and selecting evidence for inclusion.

Reflection and Connection

Portfolio-based learning and assessment is characterized by valuing: one's past experience and learning; autonomy; reflection in the learning process; and the connectivity among experiences, learning opportunities, and role requirements. The importance of reflection in learning is well documented in directed and self-directed contexts (Boud *et al.*, 1985; Moon, 1999). The reflective statements or commentaries included in the portfolio are intended to illustrate how students connect their learning and/or their professional experience with the theoretical understanding gained from course work, assessment tasks, critical readings, writings, and/or research. However, such reflection has proven to be a problematic process. Students need guidelines – which specify the nature, scope, purpose, and audience when writing self-reflective statements.

Clarification is also needed regarding the value of this process in its intended positive contribution to learners in supporting them to develop confidence and competence, in the tensions created between academic and professional knowledge, and between learning and understanding of the process of becoming a professional. Reflecting on that process is to be valued and supported in the pursuit of critical, learner agency and metacognitive development. However, achieving the full value of portfolio-based learning and assessment requires teachers or mentors who have particular skills (Pietroni and Millard, 1997). They work with students or colleagues in portfolio development to facilitate the important intellectual work of moving beyond description – of what has been learnt and achieved – to analysis of what further learning and education needs to occur. The importance of critical reflection in the learning process and its role in new learning is realized through portfolio use.

Assessment and Teacher Judgment

Summative assessment of the portfolio is designed to provide quality information with regard to student learning in

a timely, manageable, and inexpensive manner without impacting negatively on teaching and learning. It is high stakes and occurs for certification and selection in a range of contexts. The selection or certification process can provide the individual with a statement of achievement for entry into a profession, to further education, or it can even lead to promotion. An adequate level of reliability is, therefore, required for comparability purposes. Consistency of standards, and consistent grading, must be implemented to ensure equity and fairness and to ensure quality in the overall assessment process and outcomes. Specified standards and contents frameworks aim to achieve a reasonable degree of reliability and to ensure a level of confidence in the results and comparability across institutions. The standards framework and attributes to be assessed are generally specified by awarding or professional bodies such as: teacher education (Lyons, 1998), higher education (Fry *et al.*, 1999), and doctoral studies (Shulman *et al.*, 2006). Professional development for continuing learning and development is also assessed using portfolios (Baume and Yorke, 2002; Orland-Barak, 2005; Jackson and Ward, 2004; Hay and Moss, 2005).

When assessing the portfolio summatively, the consistency of approach to the assessment tasks and consistency of teacher judgment in assessing the portfolio of work, using the standards framework, need to be monitored. Replicability and comparability are the key qualities (Gipps, 1994). Professional judgment in the use of criteria and standards for assessment is developed through moderation practice. Such professional collaboration helps to maintain consistent application of standards.

Holistic or analytic approaches can be adopted to assess the portfolio of work. In the latter approach, different aspects of the portfolio are assessed independently and judgments with regard to the quality of the parts are aggregated to obtain a total grade. Holistic approaches require a judgment with regard to the overall quality with attention to how the individual tasks or samples of work contribute to the whole. The multiple entries of the portfolio require the assessor to engage in iterative and cyclical processing sequences which differ to the assessment of a single work. Major threats to validity and reliability can occur when assessors omit the use of important criteria provided, “construct underrepresentation” or give particular weighting to criteria while not attending to the given criteria with equal evaluative attention, “construct-irrelevant variance” (Messick, 1995).

The use of portfolios for formative purposes to enhance learning and for professional development is well established. One of the major advantages of the portfolio for learning purposes is the opportunity it provides to monitor development and for teachers and instructors to provide feedback to the learner to fulfill a transformative function in the learning. To achieve this, there is a need for substantive conversation concerning

the qualities of the learning. This implies a facilitator role for mentor or teacher. Student self-reflection – promoted through conversation – requires interactive dialog that facilitates student recognition of strengths or weaknesses in their learning. Insights regarding how to improve are an intended consequence of this process. Thus, the portfolio provides the structure and process to facilitate understanding of one’s own learning or professional practice (Klenowski *et al.*, 2006) and for increasing one’s self-regulating capacity.

Research on how best to support learning in professional contexts and how to assess the rich, qualitative materials in portfolios concludes that a hermeneutic, interpretative approach is appropriate (Tigelaar *et al.*, 2005).

The resource implications of this approach are acknowledged any method for interpreting portfolio evidence in an equitable and responsible manner will require time and substantive conversation. Professional dialog in the context of teacher education, for instance, is fundamental for realizing the potential of this form of assessment and for engaging pre-service teachers in understanding deeply the meaning of effective teaching.

The problematic nature of the use of reflective statements in portfolios for learning and professional development has been researched and is more apparent when portfolios are assessed summatively (Chetcuti *et al.*, 2006; Orland-Barak, 2005). Students are reluctant to include authentic reflections that illustrate their areas of weakness or gaps in learning and can revert to “tactical writing” to convince the assessor of their achievements (Meeus, *et al.*, 2006). Students will not be inclined to reveal their failures (Smith and Tillema, 2003). Such reflections incorporated in the portfolio will not be reliable or genuine.

E-Portfolios

Electronic portfolios – also referred to as e-portfolios, digital portfolios, or webfolios – were developed in the early 1990s (Barrett and Wilkerson, 2004). The term e-portfolio has been adopted as an umbrella concept (Siemens, 2004) to include webfolios as in the concept of e-learning which includes web-based training. The characteristics that distinguish the e-portfolio from the paper-based portfolio are electronic access and digitization. Webfolios are static websites that make use of links to HTML-driven sites; however, e-portfolios are dynamic. They have been defined by the EDUCAUSE National Learning Infrastructure Initiative (NLII, 2003) as:

a digital collection of authentic and diverse evidence, drawn from a larger archive representing what a person or organisation has learned over time on which the person or organisation has reflected, and designed for presentation to one or more audiences for a particular rhetorical purpose.

This larger archive can also be considered an e-portfolio that contains multiple views of the contents constructed for a range of audiences. The multiple views relate to the purpose and the reader of the e-portfolio. For instance, the e-portfolio can be developed to demonstrate achievement, to meet formative or summative assessment requirements, or be used in searching for a job. The collection includes evidence of learning and performance, reflections or interpretations on the evidence, and representations of relationships between and among evidence, interpretations and evaluation criteria (NLII, 2003). An e-portfolio-management platform supports the learner in organizing the archive and enables connection to services that facilitate transactions and processes such as getting feedback from peers or cross-referencing of evidence with competency or professional standards.

Given the context of the knowledge society, the changing nature of the curriculum and of learning the e-portfolio offers features that address the changing needs of learners themselves (Siemens, 2004). E-portfolios are multipurpose, adaptable, interactive, transportable, and searchable (Cotterill *et al.*, 2004). The digital nature enables the adoption of a range of structures, facilitates cross-referencing, reduces administration, and allows secure access from a range of locations. Such features provide a rich learning tool for the learner to demonstrate their knowledge and understanding through evidence that can take the form of work samples, artifacts, or illustrations of development and learning including feedback from teachers, self-, and peer-assessments. The use of multiple media to facilitate, enhance, and demonstrate learning is epitomized in the e-portfolio.

At present, learning takes place in contexts that are no longer confined to the traditional, formal education settings. Learning on the job – in communities and through networks (Siemens, 2004) – is common and it is technology that has aided learning at a distance and in remote locations. With online teaching and learning becoming increasingly popular (Russell and Finger, 2005), there are important implications for teacher education and development. Teacher educators must teach pre-service teachers to be competent in the use of information and communication technologies (ICTs). It is, therefore, not surprising that the growth in the use of e-portfolios has occurred first in the contexts of teacher and higher education (Lane, 2007; Lind, 2007; Mason *et al.*, 2004).

Research has indicated that e-portfolios offer certain benefits to learners in that they provide valuable opportunities for reflective learning (Zubizarreta, 2004) and encourage students' active engagement with learning. The interactive nature of the e-portfolio assists learners to make connections between formal and informal learning experiences and among components within a subject, across subjects within a program, between field experiences, and components of programs. Young

learners, at present, who have a confidence with the use of multimedia are multi-processors and are more comfortable with a hyperlinked approach to thinking, as opposed to one that is linear (Brown, 2002). For such a learner, the advantages are many as they manage personal knowledge, the growth of their learning, planning, and goal setting, and take control over their learning history (Siemens, 2004). Meta-learning (Klenowski *et al.*, 2006) is facilitated when, in this context, the learner has the opportunity to review the learning in the e-portfolio and learns about their own learning – which can then be applied to future learning. It is the identification of new insights and understandings that brings about the changes.

The Web enables the user to be both receiver and sender of broadcast, which has been described as both push *and* pull (Brown, 2002). In this new learning context, it is possible for young people to engage in their preferred way of learning because multiple intelligences are honored in this medium. Students acquire skills in multiple literacies (text, screen, and image) and learn how to navigate through multiple-media genres and complex information spaces. Unlike the formal learning environments of the previous generation, young people today are engaging in discovery learning when browsing the digital “libraries” and the information available on the Web (Brown, 2002). The skills of judgment, discernment, and synthesis become crucial.

Research – in the context of teacher education – suggests that e-portfolios can help enhance pre-service teachers' meta-abilities. These include: higher-order cognitive skills, self-knowledge, personal resilience, and management of knowledge and time. Another finding suggests that a greater sense of interconnection seems to be fostered for pre-service teachers across and between disciplines and between theory and practice. There appears to be a greater acceptance for responsibility to plan, develop, and evaluate one's own learning outcomes and achievements (Broadbent, 2005).

Some of the principles of electronic portfolios as identified by (Cotterill, 2004) are as follows. E-portfolios have significant advantages over those that are paper based, they need to be considered in a social context of human processes, they must have clarity of purpose, they should be learner-focused, integral to the learning process, support lifelong learning, continued research and evaluation is essential, and one particular view will not suit all purposes.

Some of the limitations of the e-portfolio are that they preclude the inclusion of physical objects that may have been constructed by the student and, in terms of equity, students will have different levels of information technology (IT) experience and varied access to computers and Internet while the reliability of some IT systems could present another weakness.

Conclusion

While the potential value of portfolios remains significant in addressing the varied learning needs of students and professionals a number of uncertainties still remain. Further research to guide the use of portfolios for assessment needs to be conducted. There is also a need for further development of conceptual frameworks to guide the pedagogy and assessment of the portfolio that engage students in the negotiation of these. The emergence of the e-portfolio offers exciting opportunities for promoting learning through assessment that values reflection, participation, and narrative; however, research and development will remain a priority in this creative, digital, context.

See also: Challenges of Developing and Implementing Formative Assessment Practices in Schools; Formative Assessment; Records of Achievement: Beyond Traditional Tests; The Multiple Purposes of Assessment.

Bibliography

- Arter, J. A. and Spandel, V. (1992). Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice* **11**, 36–44.
- Barrett, H. and Wilkerson, J. (2004). Conflicting paradigms in electronic portfolio approaches. <http://electronicportfolios.com/systems/paradigms.html> (accessed September 2009).
- Baume, D. and Yorke, M. (2002). The reliability of assessment by portfolio on a course to develop and accredit teachers in higher education. *Studies in Higher Education* **27**, 7–25.
- Boud, D., Keogh, R., and Walker, D. (eds.) (1985). *Reflection: Turning Experience into Learning*. London: Kogan Page.
- Broadbent, C. (2005). Use of digital professional portfolios to enhance pre-service teachers' meta-abilities. In Cooper, M. (ed.) *Teacher Education: Local and Global*, proceedings of the Australian Teacher Education Association, pp 70–76.
- Broadfoot, P. (1988). Profiles and records of achievement: A real alternative. *Educational Psychology* **8**, 291–297.
- Brown, J. S. (2002). *Growing up Digital: How the Web Changes Work, Education and the Ways People Learn*. United States Distance Learning Association. http://www.usdla.org/html/journal/FEB02_Issue/article01.html (accessed September 2009).
- Callahan, S. (1997). Tests worth taking? Using portfolios for accountability in Kentucky. *Research in the Teaching of English* **31**(3), 295–336.
- Chetcuti, D. (2007). The use of portfolios as a reflective learning tool in initial teacher education: A Maltese case study *Reflective Practice* **8**, 137–149.
- Chetcuti, D., Murphy, P., and Grima, G. (2006). The formative and summative uses of a professional development portfolio: A Maltese case study. *Assessment in Education, Principles, Policy and Practice* **13**, 97–112.
- Cotterill, S. (2004). Electronic portfolios for formative and summative assessment, BeSST Conference. <http://www.eportfolios.ac.uk/references/?display=eportfolios> (accessed September 2009).
- Cotterill, S., McDonald, T., Drummond, P., and Hammond, G. (2004). Design, implementation and evaluation of a 'generic' eportfolio: The Newcastle experience. *Paper Presented at the ePortfolio 2004 Conference in La Rochelle*. <http://www.eportfolios.ac.uk> (accessed September 2009).
- Forster, M. and Masters, G. (1996). *Portfolios*. Melbourne, VIC: Australian Council for Educational Research.
- Fry, H., Ketteridge, S., and Marshall, S. (1999). *A Handbook for Teaching and Learning in Higher Education*. London: Kogan Page.
- Gipps, C. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. London: Falmer.
- Groom, B. and Maunonen-Eskelinen, I. (2006). The use of portfolios to develop reflective practice in teacher training: A comparative and collaborative approach between two teacher training providers in the UK and Finland. *Teaching in Higher Education* **11**, 291–300.
- Harland, T. (2005). Developing a portfolio to promote authentic enquiry in teacher education. *Teaching in Higher Education* **10**, 326–337.
- Hauge, T. E. (2006). Portfolios and ICT as means of professional learning in teacher education. *Studies in Educational Evaluation* **32**, 23–36.
- Hay, T. and Moss, J. (2005). *Portfolios, Performance and Authenticity*. Frenchs Forest: Pearson Education Australia.
- Herman, J. L., Aschbacher, P. R., and Winters, L. (1992). *A Practical Guide to Alternative Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Jackson, N. and Ward, R. (2004). A fresh perspective on progress files – a way of representing complex learning and achievement in higher education. *Assessment and Evaluation in Higher Education* **29**(4), 423–449.
- Klenowski, V. (2002). *Developing Portfolios for Learning and Assessment: Processes and Principles*. London: RoutledgeFalmer.
- Klenowski, V., Askew, S., and Carnell, E. (2006). Portfolios for learning, assessment and professional practice in higher education. *Assessment and Evaluation in Higher Education* **31**, 267–286.
- Koretz, D. (1998). Large-scale portfolio assessments in the US: Evidence pertaining to the quality of measurement. *Assessment in Education: Principles, Policy and Practice* **5**(3), 309–334.
- Lane, C. (2007). The power of 'e': Using ePortfolios to build online presentation skills. *Innovate* **3**(3). <http://electronicportfolio.pbworks.com/f/reading03presentationkills.pdf> (accessed September 2009).
- Lind, V. (2007). ePortfolios in music education. *Innovate* **3**(3). <http://www.innovateonline.info/index.php?view=article&id=351>.
- Linn, R. and Baker, E. (1992). Portfolios and accountability. *CRESST Line Fall*, 1–2.
- Lyons, N. (1998). *With Portfolio in Hand: Validating the New Teacher Professionalism*. New York: Teachers College Press.
- Mabry, L. (1999). *Portfolios Plus: A Critical Guide to Alternative Assessment*. Thousand Oaks, CA: Corwin.
- Mason, R., Pegler, C., and Weller, M. (2004). E-portfolios: An assessment tool for online courses, *British Journal of Educational Technology* **35**, 717–727.
- McLean, L. (1990). Time to replace the classroom test with authentic measurement. *Alberta Journal of Educational Research* **36**, 78–84.
- Meeus, W., Van Petegem, P., and Van Looy, L. (2006). Portfolio in higher education: Time for a classificatory framework. *International Journal of Teaching and Learning in Higher Education* **17**(12), 127–135.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific enquiry into score meaning. *American Psychologist* **50**, 741–749.
- Moon, J. A. (1999). *Reflection in Learning and Professional Development: Theory and Practice*. London: Kogan Page.
- NLLI (National Learning Infrastructure Initiative) (2003). <http://www.educase.edu> (accessed September 2009).
- Orland-Barak, L. (2005). Portfolios as evidence of reflective practice: What remains untold. *Educational Research* **47**, 25–44.
- Paulson, F. L., Paulson, P. R., and Meyer, C. (1991). What makes a portfolio a portfolio? *Educational Leadership* **48**, 60–63.
- Pietroni, R. and Millard, L. (1997). Portfolio-based learning. In Pendleton, D. and Halser, J. (eds.) *Professional Development in General Practice*, pp 81–93. Oxford: Oxford University Press.
- Reis, N. and Villaume, S. (2002). The benefits, tensions, and visions of portfolios as a wide-scale assessment for teacher education. *Action in Teacher Education* **23**, 10–17.
- Russell, G. and Finger, G. (2007a). ICTs and tomorrow's teachers: Informing and improving the ICT undergraduate experience. In Townsend, T. and Bates, R. (eds.) *Handbook of Teacher Education: Globalization, Standards and Professionalism in Times of Change*, pp 711–724. Dordrecht: Springer.
- Russell, G. and Finger, G. (2007b). Teacher education futures: Implications of teaching and learning in an online world.

- In Townsend, T. and Bates, R. (eds.) *Handbook of Teacher Education: Globalization, Standards and Professionalism in Times of Change*, pp 625–640. Netherlands: Springer.
- Shulman, L. (1998). Teacher portfolios: A theoretical activity. In Lyons, N. (ed.) *With Portfolio in Hand: Validating the New Teacher Professionalism*, pp 23–37. New York: Teachers College Press.
- Shulman, L. S., Golde, C. M., Bueschel, A. C., and Garabedian, K. J. (2006). 'Reclaiming education's doctorates: A critique and a proposal. *Educational Researcher* **35**, 25–32.
- Siemens, G. (2004). ePortfolios. <http://www.elearnspace.org/Articles/portfolios.htm> (accessed March 2009).
- Siemens, G. (2004). ePortfolios. <http://www.elearnspace.org/Articles/eportfolios.htm> (accessed September 2009).
- Simon, M. and Forgette-Giroux, R. (2000). Impact of a content selection framework on portfolio assessment at the classroom level. *Assessment in Education: Principles, Policy and Practice* **7**, 84–101.
- Smith, K. and Tillema, H. (2003). Clarifying different types of portfolio use. *Assessment and Evaluation in Higher Education* **28**, 625–648.
- Tigelaar, D., Dolmans, D., Wolfragen, I., and van der Vleuten, C. (2005). Quality issues in judging portfolios: Implications for organizing teaching portfolio assessment procedures. *Studies in Higher Education* **30**(5), 595–610.
- Wolf, A. (1998). Portfolio assessment as national policy: The National Council for Vocational Qualifications and its quest for a pedagogical revolution. *Assessment in Education: Principles, Policy and Practice* **5**(3), 413–445.
- Zeichner, K. and Wray, S. (2001). The teaching portfolio in US teacher education programs: What we know and what we need to know. *Teaching and Teacher Education* **17**, 613–621.
- Zubizarreta, J. (2004). *The Learning Portfolio: Reflective Practice for Improving Student Learning*. Bolton, MA: Anker.

Records of Achievement: Beyond Traditional Tests

P Broadfoot, University of Gloucestershire, Cheltenham, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Formative assessment – An assessment that is designed to support the learning process.

National record of achievement – The government-sponsored records for English school students.

One of the central purposes of educational assessment is the reporting of learners' achievements. Although educational assessment has a key role to play in providing feedback to learners that can guide, inform and motivate the learning process, it is the role of educational assessment in communicating the results of learning – in recording achievement – that is arguably its most visible manifestation. It is certainly the most powerful in determining the priorities of students and teachers. As records of achievement, in one form or another, are the basis for certification and selection, they act as a powerful gatekeeper to opportunities for further study and for prestigious jobs.

As is now widely accepted, over the last 150 years, educational assessment has gradually become the major means by which society selects and allocates individuals, first to different schools, colleges, and universities and afterwards, to employment. Given the widespread belief in the appropriateness of such procedures to provide for the necessary social processes of educational and occupational selection, we have come to recognize and accept the consequence that what is assessed, when, and how, exerts a major influence on curriculum priorities and on the way in which learners approach the task of learning. Teachers train students to hone the skills that will be required to obtain a good result in the assessment and students focus their study on what will be tested.

Traditionally, such assessments, in the context of formal education at least, have typically taken the form of paper-and-pencil tests which are easy to administer and police as well as relatively easy to mark. Whether in the form of multiple choice, the so-called objective tests, or in the extended essay type, such written tests and examinations have necessarily emphasized those achievements that can most readily be tested in a written form – the recall of knowledge and the demonstration of understanding. We accept that it is very difficult to design tests and examinations that go beyond such intellectual operations in their

scope, if the results are to be sufficiently comparable for the outcomes to be an acceptable basis for certification and selection. The concept of the reliability of assessment, that is, broadly speaking, the degree to which a given test or assessment is likely to produce a similar result on a different occasion or with a different test, has long been deemed a top priority for any assessment with so-called high-stakes consequences. Implicitly and explicitly, the principle that assessment results must be trustworthy and defensible to key stakeholders has long been regarded as of the utmost importance, because without such trust, the legitimacy of assessment as the central gatekeeper for life-chances is likely to crumble.

It is now increasingly widely recognized, however, that the prioritization of reliability in assessment design, has tended to squeeze out the equally important issue of validity – broadly, whether the assessment addresses the full range of learning goals in a given course of study. Messick's seminal work, exploring the different aspects of validity and the consequences of the tension between validity and reliability for the educational process, has powerfully demonstrated the high price currently paid within formal education systems as a result of the prioritization of reliability over validity (Messick, 1995). If, as suggested above, what is tested is what is taught, the emphasis on narrowly framed forms of assessment and testing must necessarily constrain curriculum priorities such that some of the most sought-after learning outcomes are ignored. In short, as is often said, there is a tendency to value what is measured rather than measuring what is valued.

This is the issue, and the dilemma, that lies at the heart of the recent trend toward the development of portfolio and records of achievement schemes. Such developments are a direct challenge to the history of educational assessment over the last 100 years and more, in which by far the greatest proportion of summative assessment at least, has focused on a very narrow range of skills and achievements. The development of portfolio and records of achievement schemes reflects a recognition that many of the other achievements and skills that students have are neither assessed nor recorded, despite being in many cases, both centrally important curriculum goals and also of great potential interest to selectors and recruiters. Qualities such as the ability to work in a team, or to solve problems, and dispositions such as creativity and resilience, are typical of the aspects of individual achievement that are of increasing interest to employers in particular but

which traditional forms of assessment cannot readily address. In short, the significant mismatch between many curriculum priorities and many of the skills needed for successful employment in today's world is now fueling the search for novel ways of promoting and recording learning and achievement through innovative forms of assessment.

Reasons for Change

In their 1998 international review of developments in the use of portfolios and records of achievement, Koretz and co-workers include examples of initiatives from Canada, the United States, and South Africa, as well as from the UK. They could have easily taken similar examples from Australia and New Zealand, Hong Kong, and many other countries since such initiatives have become so widespread in the search for forms of assessment and certification that are more appropriate to today's educational goals and student's needs.

At the heart of this search, as suggested above, are the following concerns:

- the need for a better match between the content and mode of assessment and curriculum goals;
- a recognition that examination and testing systems conceived when a minority of students were able to participate in senior secondary school and higher education are inappropriate for an era when the vast majority of the age cohort aspires to high school and college; and
- a recognition that if the new curriculum goals associated with preparation for work are to become central to the curriculum, they must be recognized in assessment procedures.

As Koretz and co-workers point out, such recognition can be expressed in terms of three key themes that are broadly characteristic of the contemporary portfolios and records of achievement movement. These are:

- An increased emphasis on validity in ensuring that the full range of learning goals is represented in the assessment;
- An increased emphasis on the authenticity of the task so that the predictive validity of the assessment is maximised; and
- The encouragement of skills related to '*action planning*' including the ability to identify one's own learning needs, to set an appropriate course through the diverse range of materials and provision and to evaluate one's own achievements and recognise where one has gone wrong. (Koretz *et al.*, 1998: 303).

The last of the three themes characterizing the recent development of records of achievement identified by

Koretz and co-workers – action planning – is perhaps the most novel highlighting as it does, an increasingly important formative dimension to such new developments. What was initially very largely an initiative focusing on summative assessment issues, has now become a substantially more generic concern for greater integration of learning and assessment goals throughout the educational process.

Thus, recent decades have seen a sustained international attempt to create not only new, more comprehensive, approaches to recognizing students' attainments in the form of summative records of achievement. They have also seen a growing concern to develop new formative assessment procedures that support the development of the capacity to learn itself, procedures that help students develop the skills they will need to be autonomous and independent, equipped to manage themselves as learners throughout life.

Not surprisingly, as Klenowski describes, the representations of these aspirations in practice are hugely varied depending on the age and stage of the students concerned, the principal purpose of the assessment and often, specific national assessment traditions and institutions. One of the countries in which records of achievement and portfolio assessment have arguably been most comprehensively developed is the United Kingdom. A brief review of developments in this national context can thus offer a useful case study of the impact of the international trends currently underpinning the evolution and use of new approaches to assessment and reporting.

The Evolution of Records of Achievement in England

The first stage of this development took place in the late 1970s and took the form of an attempt to offer the otherwise unqualified school leavers who lacked any external examination credentials, a broad-based profile in the form of a documentary record of their other achievements and experiences while at school drawn both from within the curriculum and from their life more generally. It was felt that many young people could provide evidence of significant achievements and personal qualities demonstrated in activities such as social service, sport, work experience, and family and community life. Moreover, it was argued, that many of the achievements and skills so demonstrated were more relevant to future working life and citizenship than conventional examination and test success.

The idea of recording and celebrating a much wider range of student achievement rapidly developed into a major educational movement that captured the enthusiasm of teachers across the country and convinced the government to adopt a policy of offering all students an accredited record of achievement. Thus, in 1990, the

government instituted the requirement that all school leavers in England should be issued with a National Record of Achievement (NRA).

The National Record of Achievement

The NRA could be produced in both hard copy and electronic forms. Compiled by each student individually, with support from teachers, the NRA was usually given more status as a credential by the involvement of an external accrediting body. The substantial portfolio provided a detailed account of a student's academic attainments, supplemented by information about other kinds of achievement such as sporting and artistic successes, social, cultural, and service activities. Where appropriate, evidence of these achievements, such as certificates, was also included. The NRA also included a personal statement written by the student which was intended to highlight their key achievements and experiences. Given its emphasis on positive achievements and inclusivity, the NRA was intended to give all students a sense of achievement and self-esteem. Many schools organized presentation ceremonies at which students would receive their NRA and so have their achievements publicly celebrated by friends and family.

The NRA was also intended to provide useful information about the individual for recruiters. It was felt that this kind of more comprehensive documentation of a student's achievements, experiences, and skills would provide potential employers with much more useful evidence about a student's suitability for the post in question. It was also anticipated that the possession of such documentary evidence of their achievements would make it easier for students to sell themselves at interview (Broadfoot, 1998).

In practice, the outcomes of the initiative were very mixed. Many students found the process of organizing and recording the evidence to go into their NRA, tiresome and tedious. They did not believe that employers and other recruiters would take such a record seriously even when, as in many cases, the NRA bore the stamp of an external validating body to testify to the authenticity of its evidence base. In the face of continuing and often increasingly intense pressures for more high-status traditional examinations, students and teachers were prone to seeing the whole exercise as an irrelevant chore.

Records of Achievement as a Formative Assessment Tool

In one of the many ironies of educational development, however, while the introduction of summative records of achievement in England largely failed to achieve its explicit intentions, it nevertheless resulted in some significant new insights concerning the potential use of such records for formative purposes. During the 1980s and 1990s, there was a growing recognition that young people

were growing up in a world where lifelong learning would be necessary if they, and the economies in which they worked, were to prosper. This underpinned a developing interest on the part of policymakers in how such dispositions could be developed and supported. Increasingly, both those involved in education and policymakers formed the view that young people were growing up in a society in which technological change would make multiple careers and hence, regular retraining, a necessary norm. In such a portfolio society, individuals would thus need to be equipped with the skills to manage their own learning; to recognize for themselves their own strengths and weaknesses and hence, the new knowledge and skills they would need to acquire in order to fulfill their personal life and career goals.

Personal Development Planning and Progress File

Thus, the concept of a record of achievement as a one-off, summative record, which was issued to the student on leaving school, gradually became transformed into the much more generic concept of a student's ongoing portfolio of recorded evidence and aspirations. Officially, the NRA was renamed Progress File, the new name reflecting the ambition of the new scheme to train and support individuals potentially throughout their whole life and certainly throughout formal education, in the skills needed to monitor their own learning, communicate achievements to others, and to plan next steps. Thus, at the present time, throughout their time in school, college, or university, students are to be helped to engage in personal development planning (PDP). Progress file will provide them with tools which they can use to record particular achievements both within and outside the institution; record examination and test results, reflect on their achievements, and to engage in curriculum and career planning. Increasingly available as an online set of tools, progress file is intended to be a dynamic system which institutions can adapt to suit their own learning priorities and concerns.

Thus, what started out in England, at least as simply an initiative to provide a more comprehensive form of school-leaving certificate, has become transformed into the concept of PDP. This transformation has heralded a much more profound educational change than simply a change of emphasis in the recording of achievement. It has been the engine for, and the visible manifestation of, a growing recognition that the curriculum must contain more than just predefined packages of knowledge and skill. Rather, changes in society and the world of work are fueling the recognition that individual learners must also have opportunities to articulate and navigate their personal learning journeys as part of their curriculum experience.

Clearly, this evolving emphasis on recognizing and equipping the individual as the driver of their own learning, represents a fundamental challenge to traditional models of formal education. Typically, these have been based on a designated curriculum which is to be taught to students in order that they may acquire the knowledge and skills deemed to be needed for adult life. Assessment then determines whether that process of inculcation has been successful. PDP, by contrast, requires provision within the timetable for students to engage in the necessary reflection and recording – a not insignificant challenge for the instructional and administrative systems of most educational institutions. It also requires teachers to engage with students in different ways. In the development of the NRA in England, for example, it was found that real educational benefit was demonstrated where teachers were able to engage with students individually to discuss their achievements, progress, and future objectives on a one-to-one basis. This finding was particularly significant given that few students reported that they had ever had the opportunity for such a planned, one-to-one conversation with a teacher before.

More General Lessons

The development of records of achievement in England, their subsequent translation into the progress file initiative, and, more recently, into more generic approaches to PDP, highlights an increasingly widespread recognition that students' learning is likely to be facilitated by a more personalized approach to curriculum and pedagogy, as well as to assessment. As such, this represents a significant broadening of understanding concerning the range of potential purposes for educational assessment. Alongside the undiminished international importance of summative educational assessment for certification and selection, there is now widespread international recognition of the potential role of assessment as a tool to support learning. In particular, it is increasingly being recognized that the processes that support the preparation of student portfolios are likely to play a central part in supporting and equipping learners to cope with a world of rapid change.

Thus, around the world, a host of contemporary assessment initiatives reflect this growing awareness and hence, share a similar focus on PDP. Their rationale typically includes:

- an increased use of evidence in the form of portfolios for formative purposes, to stimulate student reflection, to enhance summative communication and as a base for setting standards.
- An increased emphasis on comprehensive descriptive recording rather than the use of grades and marks;
- An increased emphasis on teacher assessment in recognition that only the teacher can access students'

performance on some of the most significant competencies and

- A growing involvement of students themselves in the preparation of portfolios in recognition of the power of assessment to inhibit or enhance student motivation. (Koretz *et al.*, 1998: 303)

One of the most powerful and generic manifestations of this new approach to assessment concerns its incorporation as part of many professional training programs. As such, it provides a second case-study illustration of the more general characteristics of records of achievements and the educational trends they reflect.

A Success Story: Professional Portfolios

The training associated with preparing for a professional role typically has several dimensions. In addition to the need to inculcate the would-be practitioner with the relevant specialist knowledge for their chosen profession, there is a need to engender a particular value system and a commitment to continuing professional development. Indeed, within programs of professional training such as those for doctors and accountants, lawyers, and teachers, it has long been recognized that such professionals need to be equipped with the capacity to reflect critically on their own work so that they can be self-managing and self-motivating, for such attributes lie at the heart of what it is to be a professional. It follows that, while traditional forms of summative assessment in the form of tests and examinations can provide for the attestation of achievement in relation to professional knowledge, the associated emphasis on personal development and the creation of a professional identity requires a rather different set of assessment tools.

Thus, in recent years, many such training programs have involved the preparation of a reflective portfolio of evidence in which the trainee professional regularly engages in critical reflection on their performance. The resultant dossier often forms part of the final summative assessment for the program as the basis for judging not only whether the student has demonstrated sufficient understanding of key professional issues and skills in their reflective diary, but also whether they have sufficiently developed the skill of reflection itself. (Klenowski, 2002).

Electronic Portfolios

The ready access to e-learning platforms in recent years has added a new dimension to such developments with the increasing use of electronic (e)-portfolios. The ability to incorporate within e-portfolios a much wider range of types of evidence, such as sound recordings and video

clips, has served to extend still further the potential authenticity of such portfolios as records of both achievement and the student's critical reflection. As such, these developments represent the latest stage in the search for the best possible match between espoused learning goals and assessment procedures. Moreover, it is not hard to see how the skills developed for such program-specific portfolio preparation may in due course complement the more generic skills that an individual will need to manage the kind of overarching progress-file type of portfolio related to their learning career as a whole.

Transcripts and Transferability

The developments referred to above hardly merit the term assessment in its traditional sense. In place of the familiar emphasis on the standardization of tasks and the comparability of outcomes in the form of grades and marks, is a highly individualistic, often narrative-based record that does not readily lend itself to either of the traditional approaches of norm or criterion referenced assessment. It is rather an ipsative assessment – a term that echoes the emphasis on a more personalized curriculum approach referred to above. The familiar distinction between formative and summative assessment also begins to break down in relation to portfolios since the latter is likely to be explicitly formative while in preparation, yet are also often used for summative accreditation purposes as well. The result is the relatively frequent juxtaposition of two different types of evidence in one overall record.

A similarly hybrid approach is developing within the European Union where the use of a transcript that documents a student's achievements while at university is now a requirement. The Diploma Supplement, as it is called, is intended to provide a common currency of assessment information in all the countries of the Union. Typically, these transcripts contain the conventional mixture of coursework and examination results. Increasingly, however, students are also incorporating a more broad-based portfolio of other accredited achievements and evidence of relevant experience. Currently the Higher Education Achievement Record (HEAR), being piloted in a number of English Universities, presents a good example in this respect.

The Future for Records of Achievement?

The coexistence of different kinds of assessment data in one record of achievement highlights the central issue that will define the subsequent evolution of such initiatives. On the one hand, there is now a substantial body of research evidence that documents the educational benefits of portfolio assessment as described elsewhere in the

encyclopedia. The use of portfolios, when students are appropriately supported and trained, appears to provide both a positive motivational effect and the capacity to enhance students' learning. This evidence is arguably in turn further reinforced by the much larger body of research evidence testifying to the role that formative assessment more generally can play in supporting learning and learners. (James *et al.*, 2007; Black and Wiliam, 1998; Crooks, 1988) There is now compelling evidence to suggest that the use of assessment to support learning is potentially just as significant as its role in measuring learning outcomes, and that new assessment tools, such as portfolios and records of achievement, have a significant part to play in such developments.

However, to the extent that the role of assessment is defined as one of supporting learning, rather than simply measuring it, there are significant implications for prevailing modes of educational organization and system delivery. These focus on problems of implementation and practical constraints on the one hand, and the continuing hegemony of apparently less-resource-hungry and well-understood existing mechanisms for selection, allocation, and accountability, on the other.

Practical Issues

It is implicit in the above account of the development of records of achievement in England, that one of the biggest hurdles to be overcome has been that of finding time in an already-crammed curriculum for students to engage in reflection and recording and even more difficult for time to be found for teachers to talk to students individually. Over and above these practical problems, the need for significant training for both teachers and students if they are to engage positively with these new approaches, and the scale of change management required, is readily evident. Portfolios and record of achievement schemes of all kinds are almost always resource hungry in terms of both time and skill. Thus, to be successful, they require the whole-hearted commitment not only of the students and teachers involved, but also of the institution as a whole and indeed, in the end, of the educational system itself.

These findings are reinforced by the findings of Anderson and Bachor (1998) reviewing the use of portfolios in Canada and by Stecher in reviewing developments in the United States. In both cases, the authors also highlight the fact that the impetus for portfolio use is their perceived instructional benefits in encouraging students to take responsibility for their own learning; that the use of portfolios is being driven by their perceived educational benefits for students rather than by external assessment agencies. Stecher (1998) provides further validation of the difficulty of using portfolio assessment in high-stakes accountability contexts. Moreover, Koretz *et al.* (1998) identifies a tension between the role of portfolios in

encouraging improved instruction and the difficulty of using portfolio assessment for large-scale external assessments.

Tensions and Contradictions

It is perhaps ironic that the development of the new approaches to assessment and recording described here has been paralleled in many countries by a significantly increased emphasis on the accountability of educational institutions. Such accountability arrangements typically involve some form of national testing and the use of the results of such testing to judge the quality of individual institutions and teachers, as well as the system as a whole. There is now compelling evidence that such testing negatively impacts the way in which individual students engage with the learning process and further compounds the washback effect of focusing energy on teaching the test. As Mabry and Stake point out: “in the past several years, the few have become the many, decrying and documenting the disastrous effects of large-scale, high-stakes standardised and criterion-referenced testing,” their “invalidity and negative educational effects,” and their “obvious failure . . . to improve education” (Mabry and Stake, 1994: 33). However, despite such criticisms, there is little indication that the more traditional roles of assessment associated with accountability and selection are reducing in significance.

Thus, in educational systems around the world, there is currently an increasing tension between the continued dominance of a testing discourse on the one hand, with its characteristic emphasis on reliability and comparability and, on the other, the emerging and powerful evidence in support of more personally empowering forms of assessment as the key to motivating and preparing learners to acquire the skills, attitudes, and dispositions they and their societies will need if they are to be successful in the twenty-first-century world.

This tension is likely to exert a strong influence on the future development of portfolios and record of achievement initiatives. High-stakes, summative tests exert an overwhelmingly powerful influence on students’ priorities. In this context, records of achievement and other more descriptive, personalized forms of assessment are likely to continue to suffer from not being valued by students posing

the question whether there is any real future for such developments in the mainstream of formal education.

See also: Formative Assessment; Portfolio Assessment; Validity of Achievement Gains on High-Stakes Tests.

Bibliography

- Anderson, J. O. and Bachor, D. G. (1998). A Canadian perspective on portfolio use in student assessment. *Assessment in Education: Principles, Policy and Practice* 5(1), 353. Oxford: Cafax Publishing.
- Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education* 5(1), 7–74.
- Broadfoot, P. (1998). Records of achievement and the learning society: A tale of two discourses. *Assessment in Education: Principals, Policy and Practice* 5(3), 447–477.
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Education Research* 58(4), 438–481.
- James, M., McCormick, R., Black, P., et al. (2007). *Improving Learning How to Learn, in Classrooms, Schools and Networks*, 250p. London: Routledge.
- Klenowski, V. (2002). *Developing Portfolios for Learning and Assessment*. London: RoutledgeFalmer.
- Koretz, D., Broadfoot, P., and Wolf, A. (eds.) (1998). *Special Issue: Portfolios and Records of Achievement. Assessment in Education* 5(3), 301–481.
- Mabry, L. and Stake, R. E. (1994). Aligning measurement with education (essay-review). *Educational Researcher* 23(2), 33–34.
- Messick, S. (1995). Validity of psychological assessment. *American Psychologist* 50(9), 741–749.
- Stecher, B. (1998). The local benefits and burdens of large-scale portfolio. *Assessment in Education: Principles, Policy and Practice* 5(1), 335. Oxford: Cafax Publishing.

Further Reading

- Broadfoot, P. (2007). *An Introduction to Assessment*. London: Continuum.
- Moss, P. (1994). Can there be validity without reliability? *Educational Researcher* 23, 5–12.
- Prensky, M. (2001). *On the Horizon*, vol. 9, no. 5. NCB University Press.
- Resnick, L. B. and Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In Gifford, B. R. and O’Connor, M. C. (eds.) *Changing Assessments: Alternative Views of Aptitude, Achievement and Instructions*, pp 37–75. London: Kluwer Academic Publishers.
- Sacks, P. (1999). *Standardized Minds: The High Price of America’s Testing Culture and What We Can Do to Change It*. Cambridge, MA: Perseus Books.
- Shepherd, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher* 29, 4–14.
- Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan* 86, 22–27.

EDUCATIONAL ASSESSMENT – ASSESSMENT IN DOMAINS

Contents

Assessment in Higher Education
Assessment in Schools – Affective Domain
Assessment in Schools – Creative Subjects
Assessment in Schools – Learning to Learn
Assessment in Schools – Literacy Writing (Extended)
Assessment in Schools – Mathematics
Assessment in Schools – Oracy
Assessment in Schools – Primary Science
Assessment in Schools – Secondary Science
Assessment in Schools – Technology Education and ICT
Assessment in the Early Years
Assessment in the History, Civics and Social Studies Domains
Assessment in the Workplace of Performance, Developing Expertise and Competence
Assessment in Vocational Education

Assessment in Higher Education

D R Sadler, Griffith University, Mt Gravatt, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Assessment principles and practices across all sectors of education have many characteristics in common. This article is mainly about those aspects that are distinctively related to higher education.

Internationally, interpretations of the scope of assessment in higher education differ widely. The broadest use is probably that in the USA. As an unadorned term, assessment includes the evaluation of all aspects of an institution, including teaching and course quality, research output, health services, academic support, physical facilities, public relations, and financial performance. The student side includes admissions, retention, credit transfer, and student learning. The narrow interpretation focuses specifically on the appraisal of student learning in courses that lead to academic degrees. The other matters listed above are then referred to as institutional research. The narrow interpretation is in widespread use outside the USA, and is the one adopted for this article.

Assessment inevitably shapes how students approach learning, including what they focus on and how they go about learning it. Information arising from an assessment event may be used for a variety of purposes, including:

- formal certification of achievements as grades which are recorded on student transcripts for future reference

or decision making in employment, professional licensing, or admission to advanced studies;

- provision of diagnostics and advice to students on how they could improve their work or performance, this feedback ideally being clear, informative, timely, and relevant;
- evaluation of the quality of teaching, the curriculum, and course structure, with a view to facilitating their improvement; and
- accountability to the institution, accrediting bodies, employers, and the wider community.

Of the various purposes for assessment, the first two are primary and have their own descriptors. The summary record of attainment is said to be summative, whereas assessment geared specifically toward improving student performance is formative.

The Assessment Process

The typical pattern runs as follows: faculty design (or adapt from elsewhere) specific tasks relevant to a course of study; students respond to those tasks; the teacher or a competent marker appraises the quality of the student responses; and the appraisal is coded in compact numerical or verbal form.

(For objective tests, scoring may be simplified by using a template or high-speed scanner.) The assessment process produces primary data from which an inference is made about the learning that has taken place by the end of the course. Assessment tasks may take a wide variety of forms, and responses may take a variety of formats. These include class quizzes, multiple-choice or other objective tests, field or project reports, take-home essays (term papers, assignments, etc.), seminar presentations, studio works, creative designs, specialized artifacts, clinical consultations, client interviews, professional performances or procedures, and traditional time-restricted examinations.

Summative performance in a course is usually reported as a grade on a fairly coarse scale. The anchor points on the scale may be denoted by words (distinction, merit, credit, etc.), numerals (7, 6, 5, etc.), or letters (A, B, C, etc.). Pass/fail grading has only two steps on the scale; expanded scales can have ten or more. The scales may be fixed by the higher education institution, or be required to comply with provincial or national policies. Where grade scales or systems differ across higher education institutions, equivalence tables allow conversion to facilitate student transfers with academic credit. For example, the European Credit Transfer System, by adopting a uniform scale across all signature countries as part of the pan-European Bologna Process, has student mobility across Europe as one of its major aims.

Grades may also be used for reporting appraisals of individual student responses to assessment tasks. For this purpose, the scale may be elaborated by the use of \pm symbols. Alternatively, numerical marks or scores on an arbitrary fine-grained scale such as 0–30, or 0–100 may be used for individual responses to assessment tasks. A course grade is then calculated by adding marks sourced from different assessment tasks. Nonoverlapping ranges of aggregate values correspond to the various grades. Because the same grade is assigned to all total scores within a given range, a course grade represents an achievement band, not an exact position.

A grade point average (GPA) is calculated as a weighted average of all the course grades a student receives over a specific period of time (such as one semester) or over a number of completed courses (such as an entire degree program). The weightings reflect the relative magnitudes of the contributing courses. Literal and verbal grades are given numerical equivalents for purposes of calculating GPAs. Institutional rules based on grades typically govern the conditions under which students may or may not progress to more advanced studies. Too low a GPA, or too many marginal or failing grades, may preclude a student from enrolling. Consistently poor performance may lead to a period of exclusion from the degree program or the institution.

The second primary purpose of assessment is formative. In ordinary discourse, formative activity is associated

with forming or shaping something during developmental stages in order to bring it closer to a desired end. Formative assessment is concerned with how the outcome of a competent appraisal of a student work can improve the quality of the learner's future performance. Feedback is the main formative tool, and consists of information about how successfully something has been done, with advice as to how it could be improved. The primary distinction between formative and summative relates to purpose and effect, not to timing. In a purely technical sense, any feedback that makes no contribution to the improvement of future performance is not formative, by definition. However, information arising from an appraisal may serve other (nonimprovement) purposes, such as explaining the grounds for a particular mark or grade that has been awarded. A frustrating aspect about feedback is that, despite the effort that generally goes into providing comprehensive guidance to students, little improvement often occurs. Turning this state of affairs around is an enduring challenge. The design of course assessment programs (including timing) is crucial if the likelihood of improvement through assessment is to be increased.

Although assessment tasks should closely reflect the purposes and aims of each course, the actual tasks specified form only a sample of all possible tasks that could be devised. Furthermore, students make real-time choices as to how they will respond on a particular occasion, and markers typically differ somewhat in their judgments about quality. The combined effect of these and other factors is to introduce a level of fuzziness into allocated marks and grades, and consequently into inferences about learning. Quite apart from the coarseness of typical grading scales, small differences in marks between students do not necessarily reflect differences in underlying achievement.

Because a mark or grade can tap into learning only indirectly through inference, it is different in nature from a physical measurement, such as the length of a table. The assessment procedures in all disciplines employ blunt instruments that invariably rely on subjective judgments at various stages in the process. As a consequence, marks can never be made exact, or equivalent to one another. Further discussion of this topic quickly leads into consideration of the basic measurement concepts of reliability and validity, and is not pursued here.

If different assessment programs were developed and implemented for the same students in the same course, and if it were assumed that students could respond to them as independent events, scores for individual students would rarely be identical. There is always some margin of error. Despite this, marks are typically regarded as strictly accurate measurements once they are allocated. Marks are also treated as if they were units of currency, with those from different sources being given equivalent worth. Numbers, in particular, have the appearance of objectifying judgments. Measurement error can be reduced (but

not eliminated) by paying strict attention to the design of assessment programs, quality control on the assessment task specifications, and the proper calibration of markers. However, these three aspects do not account for all potential sources of error.

Reference Frameworks

Universal scales with standardized units for measuring academic achievement do not exist in higher education. Assessing and grading student work always take place within a frame of reference which confers meaning on the marks or grades. Each reference framework consists of a set of conventions, rules, and procedures for deriving, interpreting, and using the grades. The most commonly employed are the norm-referenced and the criteria-based (criterion-referenced) frameworks. The norm-referencing framework has a long history. It is so deeply embedded in higher education practice that, for many teachers, at least elements of it seem to come naturally.

In principle, norm referencing uses the performance of a group of students as the comparative background for grading decisions, the group usually being defined as all of the students enrolled in a course at one time. Technically, it is the group performance that constitutes the norm. Marks or grades reflect not only how student performances are ranked but also their relative separation. If all course groups could be assumed to perform at roughly equivalent levels on the whole, a reasonable corollary would be that the top performers in each group deserve the same grade. The same would hold true for middling performers and for poor performers. Despite differences in course content, structures, and outcomes, and although the top performers in one course are not necessarily the same students as the top performers in another course, this presumption prevails. Norm referencing is simple to conceptualize as a principle and to operationalize.

Institutions that embrace norm referencing have procedures in place to ensure that the grades in different courses are apportioned according to the same basic rules (grading on the curve). For example, only the top 10% of students (within some acceptable tolerance) might be awarded the highest grade, the next 15% the second-highest grade, and so on. In this way, the proportions of the different grades awarded are subject to fairly strict control. The proportions are determined essentially arbitrarily but, once fixed, apply to all courses. The aim is to ensure that the relative worth of grades is comparable across the institution. This principle depends only on the apportionment of the various grades, so the shape of the frequency distribution of the underlying marks is irrelevant. It could be symmetrical and bell shaped; it could be skewed, or even rectangular. There is no intrinsic connection between norm referencing and the normal probability distribution in

statistics, although it often happens that the distributions of aggregate scores for large classes turn out to be roughly bell shaped.

Norm referencing provides perhaps the simplest type of brake on grade inflation, the latter referring to the award of undeservedly high grades. The basic principle is that limiting the supply of a desirable commodity (high grades) is a way of maintaining its relative value. This is the classic market approach to determining worth or value in contexts where there are no absolute reference points or independent standards. Regardless of how well students perform individually, or the caliber of the performance of the cohort as a whole, the apportionment of grades is always about the same.

Not surprisingly, this grading framework is the subject of concern for several reasons, some of which have ethical substance. Among them is that it is unfair to compare a particular student's performance with the performances of others in a group over whose membership neither the student nor the teacher has any control. Strict norm referencing avoids considerations of academic standards in any absolute sense. It is also structurally self-adjusting in that it is automatically blind to such factors as the quality of teaching and the design of the assessment program. This could be seen as either a liability (because it masks poor teaching and assessment) or a strength (because it compensates for them), depending on the observer's vantage point. Even where norm referencing is not embraced explicitly as the grading framework, norm-based practices are still common. Two examples are: reading through a set of essays to set a baseline of expectations before beginning to assign marks, and scaling marks (using the mean or standard deviation) before or after combining them.

Criteria-based assessment has a shorter history than does norm referencing. The theoretical basis is less familiar and not as well developed. The basic idea is to award marks and grades according to absolute levels of academic achievement. In this context, absolute means that the quality of work submitted is compared with fixed reference points that, once set, function independently of other students' performances and remain stable over successive offerings of the same course. Hypothetically at least, all students in a defined group could be awarded the highest grade if their work were of sufficient quality to meet or exceed the required threshold level. In practice, of course, this is unlikely to occur, although probability theory suggests that, other things being equal, it is more likely for very small groups. The criteria (and associated standards, if applicable) are communicated to students at the time the task is set. Students can then attend to them while shaping their work before submitting it. Apart from this proactive role, the stated or implied aims of using explicit criteria include: increasing the transparency of the assessment process by enabling students to understand the

grounds on which grades are determined; making more efficient and timely provision of specific formative feedback; and minimizing grade inflation by ensuring that grades directly reflect the quality of student work. The latter can be achieved only when appropriate standards are set for the various grade levels, which is a challenging task in its own right.

In practice, criteria-based grading has been open to a wider variety of interpretations than has norm referencing. Interpretations differ fundamentally on their definitions of what constitute the criteria. The first of three examples given here is the criterion as a threshold aggregate score. For example, on a 0–100 scale, an aggregate score in the range 90–100 might correspond to high distinction, 80–89 to distinction, and so on. With this interpretation, a score of 90 forms the (minimum) criterion for the highest grade. This approach to grading has long been common practice in higher education, predating the criterion-referencing label by decades. As it fails to address the issue of actual achievement standards and makes few, if any, assumptions about the basis on which the underlying marks are derived, it is not intrinsically criteria based.

The second interpretation of criteria is as specified qualities of performance set out as verbal descriptions. For example, the highest grade in a course may correspond to the following description: clear attainment of all course objectives, with complete and comprehensive understanding of course content; development of relevant skills, and intellectual initiative to an extremely high level. Although these provide some guidance, key terms such as clear, complete, and comprehensive are open to interpretation. In the third interpretation, the criteria are the characteristics or properties of responses to assessment tasks. For written works, relevant criteria could include: support for assertions, validity of arguments, logical organization, and quality of presentation. To date, the lack of a common understanding of what criteria based means and implies for practice has inhibited high-quality discourse, research and development among scholars and practitioners.

In recent decades, approaches to higher education teaching and learning have shifted away from being teacher centered (a retrospectively applied label) and toward being student centered. Assessment that is student centered not only incorporates a wider range of assessment techniques but also reflects a resolve to bring students into an appreciation of how extended or complex responses are graded. Criterion referencing forms part of this international movement. Particular tools used to facilitate these approaches include the use of rubrics, grading schemes, and scoring guides that typically set out the criteria which the markers will use in appraisal, along with other information such as standards on the criteria, the weightings of the criteria, and how levels on the individual criteria are combined into aggregate scores.

Besides the norm-referenced and criteria-based grading frameworks, other informal bases of reference can affect the actual assignment of marks for individual students at the point of appraisal. These tend to operate implicitly rather than as formal frameworks. They include awarding bonus marks to reward risk taking, dogged persistence, significant effort, or substantial improvement. (The last of these is a simple form of self-referencing.)

Approaches to determining academic achievement standards vary widely. Among the subissues is the authority of individual academics to determine standards. Internationally, perspectives on this extend right across the spectrum, the positions at both ends being strongly advocated and defended by their respective constituencies. At one end is the view that academics, departments, and institutions are accountable to students and society for the standards they employ, and that broad consistency across courses is both desirable and achievable. It is therefore entirely appropriate for institutions to formulate and implement policies and procedures aimed at establishing and maintaining the meaning and worth of grades. Although academics often retain substantial degrees of autonomy in the design of assessment programs and the award of grades, many universities insist on collaborative review procedures intended to assure the integrity of grades. These procedures include having all assessment task specifications reviewed by peers; requiring cross-marking (moderation) of completed student work in all courses; reviewing all distributions of grades to protect against grade inflation; and using networks of external examiners to scrutinize the structure and content of assessment tasks along with samples of actual student works (with the grades assigned). This list reflects aspects of both norm-referencing and criterion-referencing frameworks. Regardless of the grading framework adopted, they have in common an underlying concern with the maintenance of academic achievement standards. Ultimately at stake, of course, are institutional reputations.

At the other end of the spectrum is the view that, because academics are highly qualified professionals and presumed leaders in their respective disciplines and fields, they are in an authoritative position to judge the quality of students' work in the courses they teach, without interference. This position may be reinforced by contextual conditions. In particular, the sovereignty of academics to decide grades can be viewed as a fundamental aspect of academic freedom (essentially a professional norm), or the right to free speech (a national constitutional right), or both.

Quality of Assessment Programs

Attention to the design of all elements of the assessment program for a course is critical in obtaining high-quality data, regardless of the assessment approach used. All

assessment designs involve sampling from the possibilities. The initial step is to tap into course outlines, objectives, and desired outcomes to identify those aspects of the curriculum to be assessed. Especially important are the relative emphases on lower order outcomes (memorized knowledge, understanding, and application) versus higher order outcomes (analysis, critique, design, and synthesis). Certain assessment task types lend themselves more to particular classes of outcomes than to others, so getting a good match is crucial. However, the lines are not entirely fixed, so generalizations about the appropriateness of various types of response formats are inevitably broad brush. Multiple-choice items often default to testing factual knowledge. However, with time and expertise they can be constructed to test higher order intellectual outcomes. Similarly, essays are often suited to assessing complex cognitive outcomes. However, the literal interpretation of many essay task descriptions allows them to be fulfilled by regurgitating memorized knowledge. Ultimately, it comes down to two factors: how precisely the task specifications are formulated and, for nonobjective items, how strictly the marker holds student responses to the task specifications.

Administration of the assessment process also offers choices, especially among modes of communication between student and teacher. These involve the format and medium for task specifications and student responses. Although the award of formal grades takes place within the applicable institutional framework, prior decisions often depend on the standards that individual markers employ. The nature and extent of feedback and other information supplied depend on later steps in the learning path. Effectiveness and efficiency are key factors, because of the prevalence of semester-length courses, intensives, and online course design. The key to effectiveness is to ensure that important learning outcomes take precedence over lesser outcomes in assessment. The key to efficiency is designing assessment events that maximize information yield within the resource constraints, including time available for marking. Extensive guidelines are readily available on all of these aspects. They are listed here to emphasize that although innovations and advances in teaching and technology properly reflect the changing character of higher education, certain fundamental principles of good assessment practice are not negotiable.

Group Assessment

Group assessment in higher education is a troublesome area on which experience and opinion are mixed, and on which the literature is large and growing. Often associated with group work and peer assessment, a common form of implementation involves dividing the students into teams, each team working on a complex group project.

The single major piece of work produced is assessed by the teacher, or by members of other groups as peers, or by all involved. Various schemes have been proposed for awarding individual student marks that eventually contribute to course grades. Options include awarding the same mark to each group member, and creating a notional mark pool which is apportioned according to each individual's contribution to the finished work. If handled properly, quality, productivity, and student satisfaction with group assessment can be high. Poorly handled, group assessment gives rise to disputes about the influence of team composition and group dynamics on individual marks, and consequent fairness to individuals in marking and grading. However, group work that is directed toward creating highly productive collaborative sites for learning does not necessarily entail using group-assessment tasks as well. Individual assessments are often appropriate.

Technology and Assessment

Information and communication technologies are revolutionizing teaching and learning in higher education, including assessment. The rate of development and shifts in terminology are rapid; however, the following is indicative of directions to date and possibilities opening up. In general terms, online technologies provide access to, and high speeds of, synchronous, asynchronous, and isosynchronous communications. These enable interactions among students, teachers, course materials, and learning processes in radically new ways. Equally, they facilitate certain types of assessment, including testing modes for which the marking and the return of results are virtually instantaneous.

Commercial and open-source course and learning management systems generally include modules that facilitate record keeping of student results as these accumulate from periodic assessments. Scores are entered, weighted appropriately, and aggregated in a standardized grade book. Students may arrange to take tests at times and places that suit them, and may opt to take practice tests. The tests themselves may be adaptive in the sense that software selects graded items from an item bank, allowing acceleration through simpler items which students find easy. The testing density increases as the limit of the student's existing knowledge and understanding is approached. This minimizes the time and number of test items needed for a sound determination of proficiency yet still provides relevant feedback to students. The structure of items can also be dynamic, and schematically show the development of some phenomenon, in color and complete with labels. Sophisticated technological problems can be posed as simulations with which students engage interactively, and the path or solution strategy tracked automatically. In such cases, the strategy used may be as important as

the actual solution arrived at. Complex task structures, such as these, are time consuming and expensive to develop.

Another form of testing, labeled as certainty based or confidence based, requires students to respond to objective items not only with their primary responses, but also, by indications on a simple scale, their degree of confidence in the answers selected. The aggregation algorithm takes confidence into account, giving most weight to correct answers about which the students are confident.

Significant developments are also occurring for non-objective assessment. For assessment tasks that require extended written responses, model responses or student's actual responses may be shared quickly and securely within a cohort, subject to appropriate ethical clearances. This can provide students with authentic evaluative experience, which is necessary if students are to develop the ability to self-monitor the quality of their own works during construction. A group of students can also peer-share their work online in order to provide mutual feedback, including tentative grades, and can participate in processes that calibrate them as appraisers.

Technology can facilitate diagnostic quizzes during face-to-face classes to enable faculty to gauge how well students are grasping particular concepts on which the next segment of material builds. All students have access to a small keypad, which is either permanently wired into the teaching space or is wireless and hand held. The teacher poses questions about the material covered, students respond individually, and the class distribution of results is projected on a screen and discussed.

The requirement for students to use word processors when producing text-based assessment responses is standard in most developed countries. When coupled with widespread Internet access, word processing both simplifies and complicates production, handling, appraisal, and the provision of feedback. Electronic submission is common. Many markers read student text directly on the screen, annotating it as necessary. Software packages allow markers to tap into banks of stored, previously composed comments, which are then incorporated into the feedback using one or two keystrokes. The same capability is available through creating macros within regular word-processing applications.

On the negative side, digital communications in general, and Internet technology in particular, have facilitated student plagiarism. The dominant issue in discussions on academic integrity tends to be about plagiarism and its control. Issues, such as cheating and data fabrication, generally take second place, unless an institution already has sound measures to make plagiarism unattractive. One key way of dealing with it routinely is to screen all electronically submitted student works using plagiarism detection software. These applications compare a student's work with all works in an enormous database (which may include Internet sources, and term papers

available from commercial distributors or class peers) to identify overlap. A different approach is to make plagiarism-detection software available to students, who are then responsible for running their own works through before submission. This transfers the onus to the student, and can play a significant role in shifting the teaching and learning culture away from policing by the institution and toward acceptance by students of responsibility for the academic integrity of their submissions.

Automated essay grading is being promoted as a cost-effective way to mark essays, provide feedback, and improve student learning. Algorithms are generally based on research in artificial intelligence, natural language processing, text structures, and concept maps. Claims for these packages indicate that they may be best suited to assessing students' substantive knowledge about a topic, their ability to reason about the subject matter in specified ways, and their ability to comprehend specific written works and express it in their own words. In some systems, the software developer compiles a bank of assignment items from which the teacher chooses. In others, the teacher submits an essay topic together with a batch of pregraded student responses. This allows calibration for the purpose of grading future responses to the same assessment item.

Conclusion

Assessment is a high-stakes activity for students, and has a major impact on how they approach learning. Regardless of innovations in assessment techniques, developments in interpretive frameworks and increased adaptability made possible by new and forthcoming technologies, the core activities that cover the design and production of appropriate assessment tasks, the emphasis on higher order cognitive outcomes, the criteria for appraisal, the assignment and interpretation of marks and grades, and the overall maintenance of academic standards clearly remain ongoing responsibilities for the higher education enterprise as a whole.

Further Reading

- Crisp, G. (2007). *The e-Assessment Handbook*. New York: Continuum.
- Falchikov, N. (2005). *Improving Assessment Through Student Involvement*. London: RoutledgeFalmer.
- Freeman, R. and Lewis, R. (1998). *Planning and Implementing Assessment*. London: Kogan Page.
- Huba, M. E. and Freed, J. E. (2000). *Learner-Centered Assessment on College Campuses: Shifting the Focus from Teaching to Learning*. Needham Heights, MA: Allyn and Bacon.
- Morgan, C., Dunn, L., Parry, S., and O'Reilly, M. (2004). *The Student Assessment Handbook: New Directions in Traditional and Online Assessment*. London: RoutledgeFalmer.

- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* **18**, 119–144.
- Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment and Evaluation in Higher Education* **30**, 175–194.
- Schwartz, P. and Webb, G. (eds.) (2002). *Assessment: Case Studies, Experience and Practice from Higher Education*. London: Kogan Page.
- Stevens, D. D. and Levi, A. J. (2004). *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback and Promote Student Learning*. Sterling, VA: Stylus Publishing.
- Suskie, L. (2004). *Assessing Student Learning: A Common Sense Approach*. Boston, MA: Anker Publishing.
- Walvoord, B. E. and Anderson, V. J. (1998). *Effective Grading: A Tool for Learning and Assessment*. Etobicoke, ON: Wiley.

Assessment in Schools – Affective Domain

M Ventura and R D Roberts, Educational Testing Service, Princeton, NJ, USA
M Zeidner, University of Haifa Haifa, Israel

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Five-factor model (FFM) – A leading personality model that proposes there are five major universal personality dimensions: extraversion, neuroticism, agreeableness, conscientiousness, and openness.

Meta-analysis – A data analytic technique that uses study sample means and correlations to derive summative evaluations to a research question.

Situational-judgment test – A test that presents complex work or social situations in a written or video format, and requires the respondents to assess the desirability of various action alternatives.

Understanding human emotion has been one of the most enduring topics in philosophy and a principal driver in the development of the modern science of psychology. For example, both in the Rhetoric and Nichomachean Ethics, Aristotle provides one of the earliest known systematic accounts of what the Greeks called *pathe* (emotions), with much of these treatises focusing on the varieties of human emotion and speculation as to how they might ultimately influence human judgment and lead to a virtuous life. In modern psychology, one of the most important and influential classifications of human behavior involves the tripartite division of cognition, conation (motivation), and affect (emotion), originally suggested by the Platonic school. Not surprisingly then, the relationship between affect and academic outcomes has been an important topic within both education and psychology. Regrettably, the role of emotion in education has often been overlooked by researchers. Furthermore, despite over a century of research, the empirical relationships between affective factors and school outcomes remain relatively poorly understood, largely because the methods for measuring affect are often less precise than those assessing cognitive factors, in particular.

This article presents an overview of several affective factors deemed to have some importance in the school context. In any discussion of affective factors, it is essential to distinguish broad, relatively stable personality traits or dispositions to behave in predictable ways (e.g., trait anger) from relatively fluctuating and ephemeral emotional states (e.g., state anger), which change as a function of the individual's personality traits and parameters of the

social situation. Thus, the degree of anger a student will manifest in a classroom altercation will be a function of her personality trait and the degree to which she feels slighted or frustrated by her classmates. First, we describe constructs that might be labeled as affective traits and supporting literature for the impact of these traits in influencing valued educational outcomes. Next, we introduce various methods that have been used to create assessments of affect, including some that are newly emerging such as the situational-judgment test. Finally, we review assessments of affect and their exemplary application in specific school contexts.

Affective Traits

Although there is little consensus among scholars with respect to the term affect, it is often used to refer to the experience of feeling or emotion. Whereas a person may either be conscious or unconscious of the specific emotion being experienced, feeling refers to the person's conscious awareness of an emotion. There are two fundamental dimensions of affect, positive affect, referring to pleasant, excited emotion, and negative affect, bringing together anxiety, sadness, and anger. The role of personality in emotional responses assumes particular importance in the educational context. Personality traits, as we will show, are linked to negative and positive affect, and so may influence the educational experience. Personality factors also influence vulnerability to emotional disorders including anxiety, depression, and cognate personality dysfunctions. Increasingly, psychological research involving traits is structured around the five-factor model (FFM) of personality. Below we discuss several traits that are fundamentally affective constructs. These include neuroticism and extraversion from the FFM, and subconstructs that may be subsumed under these super-factors, as well as measures assessing the state–trait continuum of mood and emotional intelligence (EI).

Neuroticism

Neuroticism is characterized as displaying persistence states of fear, anxiety, anger, guilt, and depression. Meta-analyses have suggested modest negative relationships between neuroticism, assessed via trait-anxiety, as proxy

measure, and academic achievement measures (e.g., Seipp, 1991), plausibly mediated through cognitive interference or reduction in working-memory capacity. In general, high levels of neuroticism predict poorer academic performance among school-aged children. Many cross-sectional studies show that facets of neuroticism, aside from anxiety, such as hostility, impulsivity, and low self-esteem, are associated with various indices of aggression, delinquency, and illegal acts, in both children and adults. A meta-analysis has shown that the traits associated with antisocial behavior typically relate to dimensions of the FFM including neuroticism (Furnham and Heaven, 1999). Comparable findings are obtained in school settings, using criteria such as violence, vandalism, and theft. Others have found that angry hostility and impulsiveness, measured as facets of neuroticism, correlated with antisocial behavior, but other facets, such as anxiety and stress vulnerability, did not (for a review, see Matthews *et al.*, 2006).

Various longitudinal studies in the USA and Europe have also confirmed that childhood temperament measures related to neuroticism are predictive of misconduct later in school life. Studies have shown that children classified as being high on neuroticism and aggression are more likely to report antisocial behaviors, to have a criminal conviction, and to be diagnosed with antisocial personality disorder. Situational factors such as parenting style and deprivation also influence antisocial behavior; these factors may moderate the impact of temperamental factors (Matthews *et al.*, 2006). Bullying children also tend to be high on aspects of aggression, but their victims are marked by reactive aggression only. Studies have shown that while envy and resentment may be motives for bullying, there is little evidence to suggest that bullies suffer from any deficit in self-esteem. However, bullying can also be used as a tool to conceal shame or anxiety or to boost a person's self-esteem – by demeaning others, the abuser him/herself feels empowered.

Coping with Stress

Coping is another key factor in understanding the role of affect. Stress researchers have differentiated several strategies of coping, with most results converging around three dimensions (Folkman and Lazarus, 1988). Problem-focused coping involves determining effective strategies for reducing strain levels, establishing specific behavioral targets, and engaging in the behavior that will help solve problems. Researchers in the stress literature usually describe problem-focused coping in positive terms, with evidence that frequent use of this coping style reduces long-term strain levels. Emotion-focused coping involves direct efforts to reduce one's strain level without affecting the actual presence of stressors and includes activities like reappraising the situation, receiving reassurance from

friends, and focusing on one's strengths. Avoidance coping consists of not thinking about the problem, distracting oneself, drinking or using drugs, or removing oneself from situations that instigate the stress process. For chronic school-setting stressors, avoidance coping is considered maladaptive and may lead to considerably higher levels of strain in the long term. Research has shown that students with higher levels of neuroticism choose more maladaptive coping strategies. Other researchers note that individuals who are prone to negative emotions have less hope that they can solve problems effectively, and this lack of hope may reduce efforts to combat stressors. In addition, recent research has underscored the pivotal role of positive emotions in moderating the negative effects of stressful emotions on adaptive outcomes.

A widely used measure related to stress and coping in high school students is the ways-of-coping questionnaire (WCQ) (Folkman, and Lazarus, 1988). The WCQ considers coping as a process that is determined by the person–situation–transaction and, therefore, can only be adequately represented by the combined analysis of both the person and the stressor. In the WCQ, students report stressful events that have occurred in the last 2 months, then describe how well they coped with the stressful event. Coping strategies can be classified broadly along the three main coping strategies mentioned (problem, emotion, and avoidance coping) and more specifically under dimensions such as aggression, denial of guilt, avoidance, and relaxation.

Test Anxiety

The term test anxiety refers to the negative affect, worry, physiological arousal, and behavioral responses that accompany concern about failure or lack of competence on an exam or similar evaluative situation. Hundreds of studies have investigated the complex pattern of relations between anxiety and a wide array of conventional measures of school performance at elementary, high school, and college levels (Zeidner, 1998). Meta-analysis of available data reveals the moderate mean correlations between anxiety and achievement, somewhere around -0.20 , with academic deficits somewhat more strongly related to the worry component of test anxiety (e.g., Seipp, 1991). It is also acknowledged that effect sizes are larger for low-ability versus for high-ability students and for difficult tasks relative to those perceived as easy. Hence, detrimental effects of anxiety in the real world may represent more than just distraction from performance by the person's immediate worries about the test situation.

Behavioral avoidance also plays a key role in maintenance of evaluative anxiety and concomitant skill degradation. Procrastination, including failure to study or to complete homework, leads to failure acquiring requisite

knowledge. In turn, this lack of preparation leads to poor performance and anxiety under test conditions, increasing subsequent anxiety, and avoidance of study. However, the nature of the task plays an important moderating role. Generally, test anxiety is more detrimental to demanding tasks, and may even facilitate performance on easy tasks. In any case, it is important to diagnose and attempt to assess the specific etiological factors underlying a specific student's test anxiety (anxiety blockage, deficient study skills, self-handicapping, procrastination, and low ability) before attempting to intervene to alleviate the student's test anxiety (Zeidner, 1998).

Extraversion

A second affective trait under the FFM is extraversion. Extraversion is characterized as being social, outgoing, and assertive. Studies suggest that positive affect is a main component of extraversion and that the effect of extraversion on academic success appears age dependent. Whereas before the age of 11–12 years, extraverted children seem superior to introverted children, among adolescents and adults, introverts show higher achievement than extraverts. This change in the direction of the relationship has been attributed to the move from the sociable, less competitive atmosphere of primary school to the rather formal atmospheres of secondary school and higher education, in which introverted behaviors such as avoidance of intensive socializing become advantageous. Extraverts and introverts also differ in parameters of information processing such as speech production, attention, and reflective problem solving, with performance varying along meaningful dimensions. For example, extraverts have been shown to be better at oral contributions to seminars but poorer at essay writing than introverts (see Matthews *et al.*, 2006).

Traits of extraversion promote interaction with the environment, leading to greater opportunities to engage in social interactions in the classroom and elsewhere due to the positive affective states that are generated from such interaction. Thus, extraversion may influence emotional development both directly through rewarding affective states and indirectly through exposure to situations for practicing and learning skills for specific emotional challenges. By contrast, the qualities of the distress-prone child, such as hyper-awareness of threat and personal deficiencies, may lead to avoidance of feared social situations. This behavior pattern in turn affords fewer opportunities to develop emotion-recognition skills, leading to poorer understanding of what happens in emotional situations and hence poorer management of such situations. The resultant skills deficits lead to further avoidance and maladaptive self-beliefs that typically lead to further withdrawal.

State-Trait Measures of Mood (Positive and Negative Affect)

Both neuroticism and extraversion are traits that empower positive and negative emotions. An assessment for measuring affective states is the positive affect negative affect scale (PANAS; Watson *et al.*, 1988). There are ten descriptors for the positive affect (PA) scale (attentive, interested, alert, excited, enthusiastic, inspired, proud, determined, strong, and active), and ten descriptors for the negative affect (NA) scale (distressed, upset-distressed; hostile, irritable-angry; scared, afraid-fearful; ashamed, guilty; and nervous, and jittery). Participants are asked to rate how they felt for each of the emotions at various time periods (e.g., present moment, past few days, or past few months). High-NA is epitomized by subjective distress and unpleasurable engagement, and low NA by the absence of these feelings. By contrast, PA represents the extent to which an individual experiences pleasurable engagement with the environment. Thus, emotions such as enthusiasm and alertness are indicative of high PA, while lethargy and sadness characterize low PA. The PANAS has been useful in predicting important outcomes in school and at home. NA has been shown to be associated with more stress and coping problems, whereas PA is associated with increased engagement in school and healthy social relationships.

Emotional Intelligence

Recently, the construct of EI has emerged as an important affective component in school. Broadly defined, EI represents a set of core competencies for identifying, processing, and managing emotions (Mayer *et al.*, 2008). There are four branches to the model: (1) perception – perceiving and expressing emotions; (2) facilitation – integrating emotions into thought processes; (3) understanding – knowing the relations between emotions, between emotions and circumstances, and transitions among emotions; and (4) management – managing emotions to moderate negative emotions and enhance positive emotions. Studies have shown that even after correcting for statistical artifacts, there is a modest relationship between EI and school achievement (Mayer *et al.*, 2008).

Research has shown that EI predicts academic and social outcomes in youth. For example, a recent meta-analysis showed that emotional understanding is related to performance on cognitive tests, and particularly tests of the acculturated or learned knowledge that is valued in school curricula (Roberts *et al.*, 2008). In this analysis, emotion management is shown to be linked with pro-social behaviors, including the behavioral tendencies to be kind and trusting. In another study, Rivers *et al.* (2008) found that higher levels of emotional skills were related

to higher grade-point average and greater social outcomes (both positive and negative). Specifically, the following significant relationships were revealed for negatively valenced social outcomes: less emotionally skilled students displayed more negative attitudes toward school and teachers, higher self-reported clinical maladjustment, a higher amount of negative emotional symptoms, higher externalization and internalization of problems, more school problems, more behavioral symptoms, and more positive attitudes toward fighting, drinking alcohol, and smoking cigarettes. Significant relationships were also found for more positively valenced outcomes: that is, more emotionally skilled students had better interpersonal relations, greater self-reliance, more positive relationships with their parents, and better social, leadership, and study skills. At present, however, the intervening causal mechanisms through which EI impacts on academic outcomes is still unsettled.

Given the links between EI and both academic achievement and student well-being, one obvious question is how EI can be improved. Equally important is whether such improvement will result in concomitant gains in achievement and well-being. This is an emerging field of lively research where clearly affective assessment might play a vital role in evaluation.

Assessment Methods

Affective characteristics can be assessed in many different ways. In this section, we briefly describe both well-established and recently developed assessment tools, namely self-report, other-report, and situational-judgment tests. Some consideration is also given to more objective techniques that are also used in this domain.

Self-Report Assessments

Self-report assessments are the most widely used approaches to capturing traits such as neuroticism and extraversion. Most insights concerning the relationship between affect and educational outcomes stem from research conducted with questionnaires. Self-assessments usually ask individuals to describe themselves by answering a series of standardized questions. The answer format is often a Likert-type rating scale, but other methods may be used (such as Yes–No questions or constructed response). Typically, questions assessing the same construct are aggregated; this aggregated score serves as an indicator of the relevant affective domain.

At first glance, assessing affective qualities with self-assessments seems to be an easy, fairly cost effective, and efficient way of gathering information about the individual. However, many issues need to be taken into account

when developing a psychometrically sound questionnaire. For instance, the response-scale format has been proven to influence individual's responses. Whether one should use positively and negatively keyed questions (to avoid acquiescence) is still a controversial issue. Research has also shown that especially young male respondents are likely to use extreme answer categories and that individuals attribute different meanings to scale points along the response continuum (e.g., frequent vs. very frequent, or to some extent vs. to moderate extent).

Another problem with self-assessments is that respondents can fairly easily fake their responses to appear more attractive to a prospective institution. When self-assessments are faked, the validity of the assessment score can be compromised. Faking is a major obstacle in using self-assessments (say, of test anxiety) in high-stakes-testing situations such as taking the ACT or SAT. However, recent research on the issue of faking suggests that this confounding factor may not be as serious a threat to self-assessments as previously thought. Nevertheless, more work is needed to address how faking can be reduced in assessing affective components. The next section introduces one such methodology that addresses this issue.

Other-Report Assessments

Others' ratings are assessments in which raters (e.g., teachers, parents, or friends) judge applicants on various qualities. There is a long history of using other's ratings in schools to address the problem of faking and because it can be used at virtually any age level. The child behavior check list (CBCL; Achenbach, 1991) is one such instrument, and it involves parent- and teacher-ratings of behavioral and emotional problems and competencies for children aged 6–18 years. Constructs assessed by the CBCL include anxious/depressed, withdrawn/depressed, somatic complaints, rule-breaking behavior, and aggressive behavior. In addition, the 2001 revision to the CBCL features six new scales to be consistent with diagnostic categories of the American Psychiatric Association's fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM–IV). The DSM–IV-oriented scales are designated as affective problems, anxiety problems, somatic problems, attention-deficit/hyperactivity problems, oppositional defiant problems, and conduct problems.

CBCL items span diverse behavioral and emotional problems but are not assumed to include every possible problem that might be reported for every child in every society. CBCL findings for general population samples have been published from more than 25 societies. Across societies, girls scored significantly higher than boys on somatic complaints, anxious/depressed, and internalizing, whereas boys scored higher than girls on attention problems and delinquent behavior. At the time of writing, the test had been re-normed to include both classroom and

recess observations on a large sample of ethnically diverse 6–11-year-olds. Inter-rater reliability for the revised form across the scales was reported to range from 0.71 to 0.97.

Situational-Judgment Tests

Situational-judgment tests (SJTs) represent an assessment methodology that can be applied to measure many different constructs. Generally, SJTs present scenarios that describe incidents critical to effective behavior. A typical scenario could show several students involved in an activity that requires teamwork. One of the students might admit that he or she did not prepare for the task. After the description or display of scenarios, respondents are prompted to either specify the most appropriate behavior in such incidents or indicate their level of agreement with a given behavior. In short, SJTs can be regarded as fairly simple, economical simulations of understanding affect in social situations. For this reason, SJTs are commonly used in assessing EI, although in principle the technique could be used to construct tests of each of the affective constructs covered previously.

Virtually any critical incident given by a respondent can be used to create a scenario in a SJT, with the technique lending itself quite well to subtle manipulation (and hence assessment) of affective components. Numerous SJTs, ranging from print-based measures of emotional management to video-based measures of communication skills and social competence have been developed by industry. For the most part, however, SJTs remain mostly research devices in education. As this research shows that SJTs predict more nuanced and often neglected outcomes of the educational experience, such as retention, and appear less prone to adverse impact, the educational community is beginning to take the approach seriously. It is certainly a methodology that we suspect will grow in relevance in school settings, more especially given educational imperatives by influential groups such as the Conference Board to assess twenty-first-century skills, including affect-laden components such as teamwork, leadership, and communication skills. Of note, SJTs may also be repurposed as training tools.

Other Measures

Isard has pioneered the use of objective tests of emotion knowledge in children (see Mayer *et al.*, 2008). He has employed tasks such as identifying facial emotion from photographs, and asking children to say how characters in vignettes would feel. These emotional-knowledge tests have been shown to be useful predictors of social maladjustment and emotional pathology. Additional measures may include psychophysiological indicators (e.g., salivary cortisol to assess exam-related stress), unobtrusive indices

(e.g., assessing test anxiety through pressure students exert on pen or pencil, amount of nail biting), or unobtrusive observations, such as fidgeting.

Affective Assessments in School Settings

Many individuals have argued for a need for early identification of and intervention for children who are at risk for emotional and behavioral disorders (EBD). In this context, EBD could be defined as low levels of EI and above normal levels of neuroticism, for example. Current legislation has required school personnel to become proactive in their identification of children who display problem behavior. Specifically, the Individuals with Disabilities Education Act has mandated that children in the United States who are in need of specific early intervention or special-education services be identified through the statewide, comprehensive Child Find System. The increased emphasis on early identification and prevention encourages professionals to intervene with children before problem behaviors develop into EBD. As a result, many preschool and school personnel have incorporated the use of positive behavioral-support programs to address children's behavioral difficulties.

Behavior-checklist measures from parents or teachers are helpful in identifying children who might be at risk for EBD and in need of systematic intervention. The use of measures such as the CBCL has appeared to be a proactive step to identify children and to provide them with early intervention. A behavioral measure for use at a universal level is dependent on several features. The measure must be reliable, developmentally appropriate for preschool and kindergarten-age children, economical in terms of time and resources, and relatively easy to administer with large numbers of children.

The collaborative for academic, social, and emotional learning (see e.g., Zins and Elias, 2007) aims to reduce EBD through after-school program interventions. Such ideas and their practical implementation are collectively referred to as social and emotional learning (SEL), a broadly defined term encompassing understanding affect in school and social settings. Zins and Elias (2007) categorize SEL competencies as “the capacity to recognize and manage emotions, solve problems effectively, and establish positive relationships with others” (p. 2). SEL programs tend to be heterogeneous, broad ranging, and diffuse in their approaches to learning and to the goals of the learning itself. However, the goals of an SEL program are generally to increase children's self-awareness, social and emotional knowledge, ability to make responsible decisions, and level of competency in self- and relationship-management. Over the past decade, numerous published studies have demonstrated that SEL is a crucial factor to academic success as well as life outside of the school

environment. SEL plays an important role both in learning and in understanding issues related to school retention–attrition. For example, the National Center for Education Statistics (2002) reports that students cite social and emotional factors as the main reason for dropping out of school (e.g., 35% of students report that they dropped out of school because they do not get along with teachers or peers, and 23% of students said they dropped out because of feeling left out).

With the increasing awareness of SEL influence in school, the need for more affective-based assessments are needed to evaluate how these interventions impact students. Measures such as EI, mood, and stress and coping are key affective traits that should be affected by SEL interventions. With the growth in using affective-based assessments, counselors and school administrators can more easily understand how to diagnose and ameliorate social and emotional problems in school.

Beyond use as diagnostic and evaluation tools, affective assessments also play a role in educational policy both at the micro- and macro-level. Various accreditation agencies demand, for example, levels of accountability concerning student's level of stress and coping, and there is a move to replace anecdotal evidence with firmer metrics such as might be possible with some of the assessments currently reviewed. In addition, various large-scale assessments such as the National Assessment of Educational Progress (NAEP) and the Programme of International Student Assessment (PISA) have included assessments of affect such as anxiety and self-efficacy along with cognitive measures in every test cycle, with these ultimately influencing economic and political decisions at the local, state, national, and international levels.

Concluding Comments

In this article, we have surveyed the pervasive influence of affective traits on academic outcomes. Next, we reviewed a number of different assessment techniques that may be used to develop affective assessments. Finally, we showed how these affective assessments can be used helping policymakers, administrators, and teachers understand how to reduce many social and emotional problems in schools. The range and variety of educational issues raised by studies of affect in schools is vast, and we emphasize that our coverage here has been somewhat selective.

See also: Affect, Mood and Emotions; Anxiety; Assessment in Schools – Dispositions; Cognition and Emotion; Coping with Stressful Situations: An Important Aspect of Self-Regulation; Emotion in Educational Contexts; Impact of Assessment on Students' Test Anxiety; National Assessments; Noncognitive Measures for Higher Education Admissions; Peer and Self-assessment; Social Competence.

Bibliography

- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/ 4–18 and 1991 Profile*. Burlington: University of Vermont, Department of Psychiatry.
- Folkman, S. and Lazarus, R. S. (1988). *Ways of Coping Questionnaire*. Palo Alto, CA: Mind Garden.
- Furnham, A. and Heaven, P. C. L. (1999). *Personality and Social Behavior*. London: Arnold.
- Matthews, G., Zeidner, M., and Roberts, R. D. (2006). Personality, affect, and emotional development. In Alexander, P. A. and Winne, P. H. (eds.) *Handbook in Educational Psychology*, 2nd edn, pp 163–186. Mahwah, NJ: Erlbaum.
- Mayer, J. D., Roberts, R. D., and Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology* **59**, 507–536.
- National Center for Education Statistics (2002). *Dropout Rates in the United States 2000*. Washington, DC: U.S. Department of Education, Offices of Educational Research and Improvement.
- Rivers, S. E., Brackett, M. A., Reyes, M., and Savoy, P. (2008). Emotion skills in early adolescence: Relationships to academic and social functioning. *Paper Presented at the Meeting of the American Educational Research Association*, New York.
- Roberts, R. D., Schulze, R., and MacCann, C. (2008). The measurement of emotional intelligence: A decade of progress? In Boyle, G., Matthews, G., and Saklofske, D. (eds.) *The Sage Handbook of Personality Theory and Assessment*, pp 461–482. New York: Sage.
- Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research* **4**, 27–41.
- Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scale. *Journal of Personality and Social Psychology* **54**, 1063–1070.
- Zeidner, M. (1998). *Test Anxiety: The State of the Art*. New York: Plenum.
- Zins, J. E. and Elias, M. J. (2007). Social and emotional learning: Promoting the development of all students. *Journal of Educational and Psychological Consultation* **17**, 233–255.

Further Reading

- Chamorro-Premuzic, T. and Furnham, A. (2003). Personality traits and academic examination performance. *European Journal of Personality* **17**, 237–250.
- Collaborative for Academic, Social, and Emotional Learning (2003). *Safe and Sound: An Educational Leader's Guide to Evidence-Based Social and Emotional Learning Programs*. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.
- Costa, P. T. Jr. and McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences* **135**, 653–665.
- Heaven, P. C. L. (1996). Personality and self-reported delinquency: A longitudinal analysis. *Journal of Child Psychology and Psychiatry and Allied Disciplines* **37**, 747–751.
- Kyllonen, P. C. and Lee, S. (2005). Assessing problem solving in context. In Wilhelm, O. and Engle, R. W. (eds.) *Handbook of Understanding and Measuring Intelligence*, pp 11–25. Thousand Oaks, CA: Sage.
- Lopes, P. N., Salovey, P., and Straus, R. (2003). Emotional intelligence, personality, and the perceived quality of social relationships. *Personality and Individual Differences* **3**, 641–659.
- Mayer, J. D., Salovey, P., and Caruso, D. R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) User's Manual*. Toronto: MHS Publishers.
- Zeidner, M. and Matthews, G. (2000). Intelligence and personality. In Sternberg, R. J. (ed.) *Handbook of Intelligence*, pp 581–610. New York: Cambridge University Press.
- Zeidner, M., Roberts, R. D., and Matthews, G. (2002). Can emotional intelligence be schooled? A critical review. *Educational Psychologist* **37**, 215–231.

Relevant Website

<http://www.childfindidea.org>
– Idea Child Find, US Office of Special Education Programs.

Assessment in Schools – Creative Subjects

T E Costantino, University of Georgia, Athens, GA, USA

L Bresler, University of Illinois at Urbana–Champaign, Champaign, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Authentic assessment – Evaluative tools that reflect the nature of the learning activity that is being assessed.

Formative assessment – An evaluation of student learning in process.

Performance assessment – An evaluative method in which the learner demonstrates his or her knowledge and understanding of concepts, and skills and technical abilities by producing something.

Summative assessment – Final evaluation of student learning that comes at the end of a project, assignment, or course.

Introduction

Challenges and Common Practices

Assessment in fine arts education has been a challenging topic for educators, whether they are early childhood, elementary, or secondary school arts teachers, artists in residence or teaching artists, or educators in higher education studying arts education or preparing arts teachers. Traditionally arts educators have been uncomfortable with making an evaluative judgment by assigning a grade in subjects that are characterized by inherently creative processes and products. This attitude is changing, as arts educators understand that assessment is not necessarily focused on the outcome but can be used as a tool for helping teachers and students reflect on the learning process. Elliot Eisner makes the point that “without some form of assessment and evaluation, the teacher cannot know what the consequences of teaching have been. Not to know, or at least not to try to know, is professionally irresponsible” (2002: 179).

A complementary way of framing the importance of assessment is in terms of how cultural and societal values influence what are perceived as the most important kinds of knowledge. The famous question asked by Herbert Spencer, “what knowledge is of most worth?” reflects the role of values in education. The kinds of knowledge a society values will directly influence which content and skills are assessed. This is applicable in both the microcontext of a teacher’s perspective in the classroom

(as Elliot Eisner refers to above), and the macrocontext from a disciplinary and institutional perspective.

An oft-cited challenge by arts educators has been how to assess learning in a discipline whose processes and products are essentially nonverbal – in the case of visual art, music, and dance (drama is more verbal, whether oral or written, with the exception of pantomime) – and personal creative expressions of meaning and understanding – in the case of all the fine arts. Essentially, the arts go beyond the verbal and the numerical and what can be predicted *a priori*. Esthetic philosopher Morris Weitz has noted that the nature of art is to go beyond those *a priori* expectations. For this reason, the arts do not lend themselves to the prespecified measures that most other academic subjects accommodate. As Eisner (2002) has insightfully explained, the diversity of forms of representation in the arts implies that there is no easy translation between the modes of the discipline and the evaluation tool. Assessing the arts is further complicated by the unpredictable nature of the processes and products to be evaluated. In other words, traditional modes of assessing learning through objective, often quantifiable criteria applied to observable outcomes are typically not effective in assessing learning and development in the arts disciplines where ways of knowing and representation are inherently interpretive, metaphoric, and symbolic. The power of the arts is in their expressive and interpretive qualities, but expression and interpretations are harder to evaluate than the lower levels of Bloom’s taxonomy, such as knowledge and comprehension, which are more easily assessed through objective measures. In arts assessment, therefore, there is an important need for both analytic and holistic measures.

Considering this legitimate challenge, arts educators at all levels emphasize performance and authentic assessment strategies utilized frequently for formative as well as summative purposes. Performance and authentic assessment strategies reflect the process of revision that is inherent to arts practice as artists from all arts disciplines create, reflect, revise, and create again in an ongoing cyclical process. The critique, practice, rehearsal, and presentation components of the visual and performing arts are natural examples of authentic assessment. Authentic assessments refer to evaluative tools that reflect the nature of the learning activity that is being assessed; indeed, authentic assessments are often the actual learning activity. Dorn, Madeja, and Sabol provide this definition “Authentic

assessment does not focus on factual knowledge as an end in itself. Rather, it focuses on the ability to use relevant knowledge, skills, and processes for solving open-ended problems during meaningful tasks” (2004: 15). Performance assessment is an evaluative method in which the learner demonstrates his or her knowledge and understanding of concepts, and skills and technical abilities by producing something, whether it is, for example, a written essay (an example of performance assessment commonly used in nonarts disciplines), an abstract painting, an improvised eight-part dance, and so on. Often, a student’s performance assessments will be collected in a portfolio that demonstrates growth in the student’s learning over time. This portfolio will include authentic assessments. For example, in the visual arts, a portfolio might contain gesture drawings, compositional studies, a figural painting, and written statement from the artist. Examples of formative performance assessments include the in-process critique in a painting class or the dress rehearsal in a drama, dance, or music performance. Summative uses of performance assessments might be an exhibition of student artwork or a recital. The emphasis on formative performance assessments by arts educators reflects the nature of art practice, which is an ongoing iterative process of creation, reflection, and revision according to the perceived effectiveness of expressive products.

A third significant challenge in arts-education assessment, in addition to arts teachers’ traditional reluctance to formally assess, and the difficulty of assessing primarily nonverbal creative processes and products, occurs in a broader context in the form of standardized assessments at the national and state or regional levels. The current political context of high accountability and standardized curriculum, especially in the United States and the United Kingdom, puts pressure on the fine arts disciplines to consider standardized assessment in order to secure a place in the curriculum, heeding the warning that what gets tested gets taught. A recent example of research from the Center on Education Policy (McMurrer, 2008) regarding the influence of the No Child Left Behind Act on narrowing of the curriculum in the United States shows that subjects that are not currently tested through standardized assessment instruments receive reduced instructional time in elementary schools, including 36% of reporting school districts reducing instructional time on social studies by 76 min each week, 28% of reporting school districts indicating 75 min less per week in science, and 16% of reporting school districts reducing instructional time in art and music by 57 min per week. This reflects a decades-long pattern in the United States going back to the Sputnik-generated curriculum reform of the early 1960s prioritizing standardized curriculum and objective assessment.

There are actually two challenges within standardized assessment of fine arts learning: the first is how to assess arts knowledge and ability in a standardized way that

reflects the central bodies of knowledge characteristic of each arts discipline, rather than test what is easier to test and hence make what is assessed a central, highly valued body of knowledge. The second challenge is how standardized arts assessment might affect arts-education policy. In other words, there are benefits and disadvantages to national or regional standardized assessments of fine arts learning which are discussed briefly below.

Related to this, an additional challenge is how to assess the long-term impact of arts learning. Arts educators often emphasize the significance of arts instruction for developing habits of mind that support creative and critical thinking, such as persistence, comfort with ambiguity, and divergent thinking (e.g., Eisner, 2002; Hetland *et al.*, 2007). Standardized assessments consisting of multiple choice questions will not capture this kind of arts learning. A performance-based longitudinal standardized assessment is more appropriate for providing one perspective on this kind of learning to inform policy decisions, such as the 1997 and the 2008 fine arts national assessment of educational progress (NAEP) in the United States.

With the challenges and common practices of arts assessment outlined, the remainder of the article briefly discusses standardized testing of fine arts learning in order to acknowledge this context, but then focuses on classroom/individual student assessments in each fine arts discipline, as the most significant activity in arts assessment occurs at the classroom level.

Standardized Assessments in the Fine Arts

The Standards Movement

While the United Kingdom has had a national standardized curriculum including the arts since the Education Reform Act of 1988, standardized assessments in the fine arts became prominent in the United States in the late 1990s within the context of standardized curriculum brought on by the national reform movements resulting in Goals 2000 and the No Child Left Behind Act. Eager to be included in the national effort, the fine arts disciplines created national standards for visual art, music, dance, and drama, published in 1994. Individual states followed suit, developing their own state content standards in fine arts, in the mid- and late 1990s. As of 2006, according to the Arts-Education Partnership’s state arts-education-policy database, 49 states have developed state content standards for the fine arts, with most of these standards mandated as part of the curriculum (Nebraska has developed fine arts curricular guides for districts to use to develop local fine arts standards), while a few states have voluntary adoption by districts (e.g., North Dakota, South Carolina, and Wisconsin).

Along with the development of state fine arts standards in the late 1990s came the development of state

standardized assessments of fine arts learning. In some cases (e.g., Illinois), student achievement on the state fine arts standards was assessed in the social studies section of the annual state assessment, administered to specific grades. As of 2006, only Kentucky administered a statewide fine arts assessment (in grades 5, 8, and 11). Several states require fine arts assessment at the local school district level, and many states provide assessment resources for local districts that voluntarily assess their students' achievement on the fine arts standards.

The lack of statewide assessments for mandatory fine arts content standards reflects several realities about large-scale assessments. Importantly, standardized assessments are costly to develop and administer. A large-scale fine arts assessment would have to include performance-based test items to meaningfully assess student knowledge of and aptitude in the fine arts that reflect the content standards' emphases on creating and responding to works of art in all disciplines. Performance-based national assessments have been used successfully internationally, specifically the use of portfolios in the visual arts in the International Baccalaureate Program, the Dutch National Assessment System, the British GCSE art assessment, and in the twelfth-year state and national assessments in Australia and New Zealand (Boughton, 2004).

Considering the increase in standardized assessment in language arts and math under the No Child Left Behind Act in the United States, there are little to no funds available to states and local districts to develop fine arts standardized assessments. Performance-based assessments are also complicated and costly to score. The schedule for implementation of standardized assessments in the No Child Left Behind Act reflects societal priorities – language arts and math first (basic literacies), then science, and then social studies. With statewide-standardized arts assessments discontinued for the foreseeable future, the only other primary means for assessing students' knowledge of and ability in the arts disciplines on a large scale is the NAEP. There is a benefit in trying to assess arts knowledge and abilities on a large scale (as with the NAEP), as poor performance could galvanize support for increased instructional time in the arts and acceptable performance could safeguard the current level of instructional time.

National Assessment of Educational Progress

The US Congress mandated the NAEP in 1969. Called the nation's report card, the program's objective is to track US students' knowledge of and ability in academic disciplines over time (mathematics, reading, science, writing, the arts, civics, economics, geography, and US history). NAEP first evaluated visual arts achievements in 1974 and visual arts and music in 1978. The next NAEP arts assessment in 1997 included visual arts, music, and theater, and

consisted of items designed to measure eighth-graders' knowledge and skills in creating, performing, and responding to works of music, art, and theater (items compatible with the national arts standards and with current classroom practices). The 1997 arts assessment did not include dance because there was not a nationally representative sample of dance-education students. In other words, too few schools offered dance instruction at the eighth grade.

The 1997 NAEP assessment had a significant influence on how large-scale arts assessment is conceptualized. In addition to the expected multiple-choice and written-response items, it also included performance-based items. For example, music students were asked to create and perform a rock-and-roll improvisation on an MIDI keyboard. Theater students worked in groups to develop an improvisation about a camping trip. After analyzing a collage by Romare Bearden, visual arts students were asked to create a collage based on a memory of a childhood place.

Students' arts achievement on the 1997 assessment was not encouraging. On a 100% scale for creating and performing, the national average for music was at 34%, for theater 49%, and for visual arts 43%. On a 300-point scale for responding to works of art, the national average was 150 points in each discipline (music, theater, and visual arts) (Persky *et al.*, 1998). The NAEP arts assessment in music and visual arts was administered most recently from January to March of 2008 to eighth-grade students across the US. The results were comparable to the 1997 assessment.

Classroom-Based Assessment

On standards, Eisner (1995) writes that "In assessing works of art, standards are inappropriate; criteria are needed. Standards fix expectations; criteria are guidelines that enable one to search more efficiently for qualities that might matter in any individual work" (p. 763). Definitions of art in contemporary postmodern societies are open and contested. While this relativity can be applied to art made by children and youth (Bresler, 2002), there are criteria that can be assessed, such as specific parameters outlined in an assignment. Moreover, each arts discipline emphasizes different criteria for different genres within each discipline. For example, in classical music, accuracy and obedience to the score is all important as well as playing in the performing practice of the period. In contrast, jazz requires the ability to improvise. The ways in which artistic achievement is assessed also reflects philosophical (and often times political) assumptions regarding the goals and nature of teaching, learning, and the subject matter (arts discipline). This is elaborated upon within the discussions of each arts discipline below.

Assessing understanding about the arts (i.e., content knowledge) requires fundamentally different types of assessment from creating/performing in the arts. For example,

in discipline-based approaches to arts education, which incorporate history, criticism, and esthetics as content areas in addition to creating/performance, the first three fall more easily into the categories of other disciplines and can be assessed through objective measures, such as multiple-choice and written-response items. This is a particular issue in Continental Europe, where the understanding of the art of music is emphasized more than the making of music. In the US, there is more emphasis on music performance whereas in the United Kingdom and Australia music composition is integral to the curriculum. The visual arts in both the US and the UK emphasize creating. The emphasis on creation and performance requires more performance-based and authentic forms of assessment.

Visual Arts

With the prevailing approach to art education being creative self-expression in the first half of the twentieth century, assessment of students' artwork was considered detrimental to their learning and creative development. This is reflected in the prominent textbooks prior to the 1970s, including several editions of Lowenfeld & Brittain's *Creative & Mental Growth* and Wachowiak's *Emphasis Art*. Art-education curriculum began to change significantly at mid-century with the Russian launch of Sputnik, the cognitive revolution and the resulting emphasis on disciplinary knowledge in curriculum reform (under the leadership of Jerome Bruner). In the 1960s, influenced by Bruner's writings on the structure of knowledge in the disciplines, educators in higher education, most notably Manuel Barkan, Elliot Eisner, Harry Broudy, Ralph Smith, and Edmund Feldman began to apply these ideas to art-education curriculum. These concerns for art-education-curriculum reform were articulated at the Penn State Seminar in 1965 and subsequent book publications by Eisner & Ecker (*Readings in Art Education*) and Smith (*Aesthetics and Criticism in Art Education*) in 1966 and Broudy in 1972 (*Enlightened Cherishing*) emphasizing the disciplinary structure of art and the role of evaluation. This work was the precedent for the development of discipline-based art education (DBAE) in the 1980s, which emphasized a written sequential curriculum balanced with content from the four art disciplines of art production, esthetics, art criticism, and art history, and structured assessment of student learning. With the support of the Getty Center for Education, DBAE had a significant influence on art-education curriculum and the development of systematic assessment of student knowledge of and ability in art. The development of DBAE in the 1980s and 1990s closely follows the standards movement at the national level in the US. The national visual arts content standards and most states' visual arts standards reflect a DBAE approach to curriculum by

addressing art production and art history and responding to works of art (aesthetics and art criticism).

With a DBAE approach reflected in the standards, assessment of visual arts knowledge and ability could potentially be more concrete. A recent in-depth study of K-12 visual arts assessment in the US documents teachers' common practices (Dorn *et al.*, 2004). The most common tools, listed in order of priority, include: "work samples, professional judgment, teacher-developed tests, portfolios, discussions, critiques, sketchbooks, checklists, exhibits, reports, and research papers" (p. 16). The majority of these tools are performance-based, reflecting the emphasis on art making in US schools. This study also found that elementary-level art teachers use the fewest types of measures and secondary level, the most. Secondary-level teachers also use written assessments more often (reports, journals, research papers). With the emphasis on art production in US schools, the researchers in this study asked art teachers at all levels how they determined their evaluative criteria for art work. There was wide agreement on five criteria out of a participant-generated list of 23; these five included elements of art, the principles of design, composition (use of space), and creativity (p. 25).

Contemporary, twenty-first-century arts-education curriculum is becoming significantly influenced by social reconstructionist aims, embodied especially in the visual-culture approach to art education (VCAE) (Chalmers, 2005; Duncum, 2002). VCAE emphasizes developing students' visual literacy skills and abilities to critique visual culture represented especially in the popular media, with art production serving as a means of reflection and representation, often, of this critique. With this new curricular emphasis on critique in greater proportion to art making for creative self-expression, it remains to be seen to what extent and in what ways this might influence teachers' assessment practices.

Music

Assessment and evaluation have been a continued focus in the field of music. Historically, these centered on aptitude tests, used for nearly a century, reflecting the belief that when resources are limited, they should be expended on students who will benefit the most, in other words, the talented. Early pioneers are Carl Seashore's work on evaluating music talents, and later on, Ed Gordon's theory of development of musical aptitude, focusing on rhythmic and tonal pattern tasks. These assessment tools focus on reproduction rather than on the creative aspects of music.

The prevalence of these tests reflects the focus of school and out-of-school music on reproduction (as compared to the visual arts, dance, and creative drama focus on composition). A focus on listening is manifested in the other important measurements – the assessment of attitudes and preferences, for example, attitudes to music

classes, toward performing music, and toward listening to particular musical styles, among others (Cutietta, 1992).

The systematic study of creative thinking in music and its meaningful evaluation are relatively new, starting in the 1970s and 1980s. The work of Howard Gardner, and following him, Bennett Reimer on specific musical intelligences has been influential in expanding the scope of evaluation. Assessment of creativity in music has been shaped by psychometric research, cognition research, and environment research (Webster, 1992).

Dance

Unlike visual arts and music, which, for the most part, are part of the school curriculum, most dance education happens in magnet schools for the arts, private schools, and dance studios. Assessment tends to center on students' performances in dance class, based on criteria set by the instructors or the institutions. Where students take dance regularly over a period of years, the criteria for evaluating success and quality are usually specific, ongoing, and formative, primarily directed toward helping students and teachers monitor progress and improve learning (Oreck, 2007). In a globalized, technological world, many effective tools for assessing dance are accessible from Internet sites developed in England, Canada, New Zealand, and Australia. Video and computer technology can be powerful learning and evaluation tools. Students collect and analyze movement, add personal commentary, and examine and reflect on stages in the dance-making process (Parish, 2007).

Often, dance programs articulate broad educational and personal outcomes as primary goals of dance education, for example, in England's national dance curriculum, where evaluators are asked to consider how dance contributes to students' spiritual, moral, social, and cultural development (see the Office for Standards in Education website (Ofsted)). Other aspects of affective development are creativity, attitudes toward life, behavior toward others, self-management and self-esteem, and identity, which require long-term data collection.

Drama

School drama, when it is practiced, is often incorporated into other curricular disciplines (e.g., English/language arts). In general, it includes two distinct subdisciplines: creative drama, mostly for early childhood and elementary level, and drama as theater, typically in middle and high schools. These two are characterized by fundamentally different contents, forms of representation, and criteria (see Bolton, 2007; O'Toole and O'Mara, 2007). As in all the arts, the issue of assessing a product versus assessing a process is a key issue, but is manifested differently in these two subdisciplines.

Addressing the difficulties of assessment of process, Schonmann (2007) suggests drawing on a set of dialectical questions that create a profile of the student's achievement, including the student's journey to become an artist in the making and reflective understanding of drama. The issues of how we know whether students have undergone a change and what constitutes enough change are perennial questions. The deliberation process through a series of dialectic questions is formative, a motivational way of assessing students' achievements. It examines process versus goal orientation; discovery versus didacticism; rational versus intuitive methods; collaborative and cooperative versus solitary and competitive; breadth, versus specialized; and personal versus impersonal (the relationships between self and the content of drama that the student is learning).

Conclusion

The above discussion details the emphasis on formative performance assessments in classroom assessments of the fine arts within historical and contemporary contexts of both a push toward standardization and, most recently, a concern for students' social and emotional learning and development of habits of mind conducive to creative and critical thinking. As governments are concerned about remaining economically competitive, there is heightened discussion about the need to develop creative and innovative thinkers (e.g., the UK national curriculum emphasizes creativity), thereby providing an economic justification for increased arts education. It remains to be seen how the increasing globalization of curriculum and the influence of shifting economic powers and competition will influence assessment of the fine arts at local and national levels internationally.

See also: Curriculum in the Arts; Formative Assessment; Portfolio Assessment; Summative Assessment by Teachers.

Bibliography

- Bolton, G. (2007). A history of drama education: A search for substance. In Bresler, L. (ed.) *International Handbook of Research in Arts Education*, pp 45–62. Dordrecht: Springer.
- Boughton, D. (2004). Assessing art learning in changing contexts: High-stakes accountability, international standards and changing conceptions of artistic development. In Eisner, E. W. and Day, M. D. (eds.) *Handbook of Research and Policy in Art Education*, pp 585–605. Mahwah, NJ: Erlbaum/NAEA.
- Bresler, L. (2002). School art as a hybrid genre: Institutional contexts for art curriculum. In Bresler, L. and Thompson, C. (eds.) *The Arts in Children's Lives: Context, Culture, and Curriculum*, pp 169–183. Dordrecht: Kluwer.
- Chalmers, G. (2005). Visual culture education in the 1960s. *Art Education* 58(6), 6–11.

- Cutietta, R. (1992). The measurement of attitudes and preferences in music education. In Colwell, R. (ed.) *The Handbook on Research in Music Teaching and Learning*, pp 295–309. New York: Macmillan.
- Dorn, C., Madeja, S., and Sabol, F. R. (2004). *Assessing Expressive Learning: A Practical Guide for Teacher-Directed Authentic Assessment in K-12 Visual Arts Education*. Mahwah, NJ: Erlbaum.
- Duncum, P. (2002). Visual culture art education: Why, what and how. *Journal of Art and Design Education* 21(1), 14–23.
- Eisner, E. W. (1995). Standards for American schools: Help or hindrance? *Phi Delta Kappan* 76(10), 758–764.
- Eisner, E. W. (2002). *The Arts and the Creation of Mind*. New Haven, CT: Yale University Press.
- Hetland, L., Winner, E., Veenema, S., and Sheridan, K. (2007). *Studio Thinking: The Real Benefits of Visual Arts Education*. New York: Teachers College Press.
- McMurrer, J. (2008). *Instructional Time in Elementary Schools: A Closer Look at Changes for Specific Subjects*. Washington, DC: Center on Education Policy.
- Oreck, B. (2007). To see and to share: Evaluating the dance experience in education. In Bresler, L. (ed.) *International Handbook of Research in Arts Education*, pp 341–355. Dordrecht: Springer.
- O'Toole, J. and O'Mara, J. (2007). Proteus, The giant at the door: Drama and theater in the curriculum. In Bresler, L. (ed.) *International Handbook of Research in Arts Education*, pp 203–218. Dordrecht: Springer.
- Parish, M. (2007). Technology in dance education. In Bresler, L. (ed.) *International Handbook of Research in Arts Education*, pp 1381–1399. Dordrecht: Springer.
- Persky, H., Sandene, B. A., and Askew, J. M. (1998). *The NAEP 1997 Arts Report Card: Eighth Grade Findings From the National Assessment of Educational Progress*. NCES 1999-486.
- Schonmann, S. (2007). Wrestling with assessment in drama education. In Bresler, L. (ed.) *International Handbook of Research in Arts Education*, pp 409–421. Dordrecht: Springer.
- Webster, P. R. (1992). Research on creative thinking in music: The assessment literature. In Colwell, R. (ed.) *The Handbook on Research in Music Teaching and Learning*, pp 266–280. New York: Macmillan.
- Boyle, J. D. (1992). Evaluation of music ability. In Colwell, R. (ed.) *The Handbook on Research in Music Teaching and Learning*, pp 247–265. New York: Macmillan.
- Broudy, H. (1972). *Enlightened Cherishing: An Essay on Aesthetic Education*. Urbana, IL: University of Illinois Press.
- Clark, J. and Goode, T. (eds.) (1999). *Assessing Drama*. London: National Drama Publications.
- Colwell, R. (2003). The status of arts assessment: Examples from music. *Arts Education Policy Review* 105(2), 3–10.
- Gruber, D. D. and Hobbs, J. A. (2002). Historical analysis of assessment in art education. *Art Education* 55(6), 12–17.
- Murphy, R. and Espeland, M. (eds.) (2007). Assessment and evaluation (section 3). In Bresler, L. (ed.) *International Handbook of Research in Arts Education*, pp 337–444. Dordrecht: Springer.
- Ross, J. (1994). The right moves: Challenges of dance assessment. *Arts Education Policy Review* 96(1), 11–17.
- Salazar, L. G. (1996). Act IV: Theater teacher assessment and evaluation. *Arts Education Policy Review* 98(1), 27–31.
- Sefton-Green, J. and Sinker, R. (eds.) (2000). *Evaluating Creativity: Making and Learning by Young People*. London: Routledge.
- Spruce, G. (2001). Music assessment and the hegemony of musical heritage. In Philpott, C. and Plummeridge, C. (eds.) *Issues in Music Teaching*, pp 118–130. London: Routledge.
- Taylor, P. (2006). *Assessment in Arts Education*. Portsmouth, NH: Heinemann.
- Webster, P. R. (2002). Computer-based technology and music teaching and learning. In Colwell, R. and Richardson, C. (eds.) *The Handbook on Music Teaching and Learning*, 2nd edn., pp 416–439. New York: Macmillan.
- Winner, E. and Hetland, L. (2000). *Special Issue: The Arts and Academic Achievement: What the Evidence Show*. *Journal of Aesthetic Education* 34(3–4).

Relevant Websites

- <http://www.aep-arts.org> – Arts Education Partnership.
- <http://www.music.miami.edu> – Frost School of Music – University of Miami: Music Assessment Website.
- <http://www.nces.ed.gov> – National Center for Education Statistics (NCES): NAEP Arts.
- <http://www.ofsted.gov.uk> – UK Office for Standards in Education (Ofsted).

Further Reading

- Armstrong, C. (1994). *Designing Assessment in Art*. Reston, VA: National Art Education Association.
- Beattie, D. K. (1997). *Assessment in Art Education*. Worcester, MA: Davis.

Assessment in Schools – Learning to Learn

A Hautamäki, J Hautamäki and S Kupiainen, University of Helsinki, Helsinki, Finland

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Globalization and the development of information technology (IT) have set new challenges and options for education. Factors pertaining to lifelong learning are gaining momentum in educational discourses. The concept of learning to learn (L2), that has a central position amid the cross-curricular competences, has been adopted in response to these demands. Especially within the European Union, L2 has a place among the key competencies intended to be catered by educational policy considerations (Centre for Research on Education and Lifelong Learning (CRELL); Fredriksson and Hoskins, 2007).

L2 is used to refer to two diverse domains of discourses. On the one hand, L2 refers to the cognitive and affective-dispositional factors central to the application of existing skills and schemata to novel tasks and to new learning (commitments) in adult life and in the constitution of knowledge at work. On the other hand, L2 refers to potential new educational practices in basic or adult education – which would increase the appeal of education and learning to all pupils in order to enhance the acquisition of skills, abilities, and dispositions necessary in future society. Depending on the approach, there are also differing answers to the question of the criteria of L2 (Klauer, 1988). Different answers lead to different choices in assessments.

L2 is to be seen to comprise and adhere to several theoretical traditions within educational sciences. Even if, in psychology, L2 refers to repeated practice with tasks in which material varies, which enhances learning of new material, there is no general agreement in the educational literature on the definition of L2. On a general level, the search for an understanding of L2 can be seen to fall under Hirsch's (1996) Tool Conception of Education. Hirsch's idea comprises notions such as accessing skills, critical-thinking skills, higher-order skills, lifelong learning, metacognitive skills, and problem-solving skills. The L2 notion embraces this idea of general tools formed through good teaching, but acknowledges many structural constraints embedded in learning and development. Yet, L2 should be more than a set of study or survival skills (Smith, 1990). Study skills, strategies, and techniques with the related dispositions and beliefs may be the narrow(est) area of L2, but there are also attempts to broad(er) interpretations (Claxton, 2007; Hautamäki *et al.*, 2002).

L2, as an assessment issue, can be approached applying Snow's 1990-model for educational assessment. When L2 assessment is based on learning tasks that require

reflection and adaptation by the student, the tasks of the assessment exercise can be seen as both aptitude (input) and achievement (output) assessment and also to tap students' attitudes or beliefs. Snow (1994) shows – in his extended model – that, before being accepted as a task, an assignment or an exercise activates the orientation to task type, subject-matter characteristics (relevance to learner, novelty to learner, dominant symbol system, etc.), treatment dimensions (ambiguity, risk and evaluation, stress and importance of outcome, novelty, meaningfulness, complexity of information processing required, structuredness and completeness, adaptiveness to learner, etc.), and the instructional-social context. Sometimes, these orientative actions are considered to be the core of L2, together with components or elements which have been referred to as metacognition (after Flavell).

L2 leads to a state of mind which could be named: the prepared mind of Shayer and Adey, the epistemic mentality and resilience of Claxton, or the autodidactic learner of Klauer.

History of L2

Early History

Cronbach and Snow, in their treatise on aptitudes and instructional methods (ATI; 1977), point out that any study of aptitude–outcome relations ought to take learning-to-learn (Cronbach and Snow use the acronym LTL) into account: “A learner does considerably better after he has experience with many tasks of the same kind. The learning ability he displayed on the first few tasks of that series may not be the most significant indication of his ability to perform in an instructional situation or in any long-continued learning. This understanding leads to a definition where LTL is defined as a repeated practice with one kind of tasks but learning different material each time, which facilitates the ability to learn new material.”

The early start for this definition is in Harlow's (1949) studies on monkeys' learning. He demonstrated L2 – or learning set – by teaching monkeys discriminations between pairs of objects. Harlow used two experimental conditions: (1) the oddity problem, where the correct answer is determined not by the choice of one stimulus as correct but by the application of a concept: “3 stimuli, two alike, choice of the odd object is rewarded.” A subject who discovers operating rule is able to solve a new problem on the first trial. (2) The learning-set problem, where one of two objects is arbitrarily chosen as correct. Subject cannot

choose correctly on the first trial save by chance, but he can achieve perfect success on the second trial. The second type of study gives evidence of L2.

Later History

Later cognitive and developmental psychological traditions have provided the understanding of thinking and higher cognitive functions as malleable (theoretical ideas by Adey, Bruner, Flavell, Demetriou, Gardner, Klauer, Kuhn, Markman and Gentner, Nuthall, Piaget, Resnick, Shayer, and Sternberg). In this tradition, one can emphasize – when constructing a framework for assessments – the different modes of the tasks, automated skills in new contexts, identifying and using propositions, or mental tools. In the cognitive domain, critical thinking, problem solving, use of resources and rules, and formal operational reasoning are core skills. By combining differential-experimental and developmental traditions, Resnick (1987) showed the importance and possibility of thinking about thinking as something to be learned and accordingly teachable.

Flavell introduced metacognition, which has, since, made a major impact on learning studies. Sometimes, the term is used to refer assessments where the learning activity itself and acts of learning have vanished and metacognition is disembodied from cognition and reasoning as something independent. However, to plan, to monitor, and to evaluate learning activity is paramount in regulating learning. It is possible to test pupils for their knowledge of them (Meijer, 2007).

The sociocultural tradition (theoretical ideas by Bruner, Claxton, Cole, Galperin, Leontjev, Lompscher, and Vygotsky) can provide ideas relating to learning activity, to relations among motives, goals, and actions, to the roles of context factors and significant others, and to the general idea of an epistemic mentality. The contextual variables, which should be taken into account in analyzing the conditions and implications of L2 assessment, include not only the school, class, and teachers but also students' interpretation of the attitudes of their parents and peers toward school. These relate to students' self-efficacy and not only to students', but also to teachers', understanding of L2 as well.

The motivational and volitional tradition (theoretical ideas by Atkinson, Covington, Harter, Little, Niemivirta, Pintrich, Shavelson, and Skinner) attests to both positive and negative effects of beliefs in goal-oriented action. Even if the motivational, volitional, and dispositional factors can also be analyzed independently, it is important to remember that dispositions do not replace mental activity or the use of mental powers to solve tasks and to learn. All mental acts – based on comparing ideal goals and representations of the actual situations – produce emotions (happy when a goal is achieved, sad when there is a discrepancy). Therefore, cognitive development is tied to

the development of also mastering emotions. Emotions, interests, and other dispositional factors assist reasoning, and as teaching or an intervention increases children's abilities, their belief in those abilities also changes, slowly leading to a process of growth of self-concepts and self-awareness. Accordingly, only studying and assessing generalized emotions as beliefs, orientations, and motivations will not provide a complete description of the status of L2 in some school class or in school.

Current Trends in Assessment

Theoretically, the link between L2 competence as a cognitive process, and the will or attitudes and beliefs that steer its use, can be built through the construct of the autonomous personal control in the upkeep of reflective thinking and the self, learning, and personal development. Seen from this perspective, the concept of L2 is closely attached to those of voluntary learning (Vygotsky and Lompscher) and goal-directed action (Leontjev, Nuttin, and several other modern writers).

Students' goals at school are seen to be linked to learning tasks that the teacher gives and students are expected to accept them as their own. In this process, the outer social context is to be replaced – for a moment – by the inner context of the self. Then, the solution – be it positive or negative – will be evaluated by the student on two fields. Social comparison and achievement assessment tie it to the social system – to teachers, classmates, and parents – whereas inner evaluation ties it back to the goals set by the student, and the inner norm either gets reinforced or gets an incentive for change.

In the assessment of L2, students are given tasks that they are invited to accept as their own with all the motivational, goal related, and ability conditions (or beliefs thereof) attached, and the processes of L2 are set in motion in this acceptance. Irrespective of the knowledge or skill level of the student, the acceptance of the task (or the refusing of it) activates processes that either enhance or hinder flexible intellectual work necessary for learning a new thing. However, with the noncurricular nature of the L2 tasks as compared to regular curricular assessment, it is easier to see the task itself as instigating and summoning up the appropriate skills and abilities in the student, anticipating his/her willingness for lifelong learning.

As a summary, L2 could be defined as the learning set of the prepared mind to adapt to novel tasks in new and surprising circumstances and, often, within constraints which mean a high mental load.

Variation and Individual Differences

Cronbach and Snow (1977) address the issue of individual differences. They point out that the L2 generalization is

practically significant only if individual differences in learning are fairly consistent. If individual differences prove to be stable and predictable, one can capitalize on findings from the assessment in which learning is observed only for a short time, perhaps on just one task or topic. If individual differences are radically altered during learning-set formation, the short-term experiments on ATI will not give practically useful conclusion. That also implies that L2 studies would not give added value to other types of studies in learning and teaching.

Between-student differences in any cognitive and educational measure are large and significant for teaching. This is also true for L2 measures, both in cognitive and in belief factors. This outcome arouses some ideological discussions, even to the extent of creating debate of the validity of a measure which shows the L2-scores correlating with school achievement. Some of L2-discussants would like to have a measure that would show a compensatory outcome saving those pupils who do not do so well in formal education. The hoped-for L2 measures as saving virtues should show both zero or even negative correlations with school achievement and positive correlations with some later criterion measure, which also correlates positively with school marks, and, say, Organization of Economic Cooperation and Development (OECD)'s Programme for International Student Assessment (PISA) measures. However, without variations in scores, one cannot do meaningful empirical studies and describe education, and, without a good description, no modeling (theory formation) is possible when planning for a new school for all – where everybody would be better able to learn the modern contents.

When the L2 is understood from the psychological point-of-view – that is, to include studies and theories of learning – then it is useful to analyze the relation of L2 to learning ability, and, accordingly to general cognitive ability. This generalized L2 expresses itself, for example, in the developmentally most important change, when a child who has been learning-to-read moves into the mode of reading-to-learn.

Collectively Weaved Definition: Social Practice Creating Meaning

There is a relatively new phenomenon where international political institutions (United Nations (UN), United Nations Educational, Scientific, and Cultural Organization (UNESCO), OECD, and European Union (EU)) define educational terms. This process is an example of measures and activities, which create new social practices. These definitions are constructed as a response to attempts to try to achieve compromises, using methods such as the EU's open method of coordination. This process can be described as collective and institutional weaving – where participating countries try to defend their options for a leading position and, possibly, money and various experts

defend their academic fields. An example is the frameworks of OECD's PISA program. Similarly, the EU has worked on a framework for key or basic competencies, of which L2 is one. The following political or working definition is given in the EU documents.

'Learning to learn' is the ability to pursue and persist in learning, to organise one's own learning, including through effective management of time and information, both individually and in groups. This competence includes awareness of one's learning process and needs, identifying available opportunities, and the ability to overcome obstacles in order to learn successfully. This competence means gaining, processing and assimilating new knowledge and skill as well as seeking and making use of guidance. Learning to learn engages learners to build on prior learning and life experiences in order to use and apply knowledge and skills in a variety of contexts: at home, at work, in education and training. Motivation and confidence are crucial to an individual's competence. (paragraph 5, annex, Education Council, 2006)

This definition is a compromise of educational experts and policy advisors from EU countries, and is, therefore, creating new social practices when countries and schools are looking for resources and sources of innovation. In this sense, the EU definition is weaved collectively to provide a politically convenient definition upon which to also base decisions concerning money and joint actions.

Examples of L2 projects

L2 assessments in educational settings seem to be targeted to studies on how to learn in educational settings, learning how to learn, or to studies on the outcomes of processes intended to form L2 competencies, learning to learn or learned to learn. *The Curriculum Journal* has devoted a special issue for L2 (2007, volume 18, number 2).

Studies of Learning Low to Learn

Learning How to Learn – in classrooms, schools and networks is a part of Teaching and Learning Research Programme in UK (Black *et al.*, 2006; James *et al.*, 2006). The project worked with teachers, schools, and networks of teachers and schools in order to learn the actual practices of promoting how to learn. One of the frameworks has been Assessment for Learning (AfL). Studies were also conducted to understand conditions for school improvement and innovation. There are, in the literature, other good examples of careful studies of life in classrooms, and of practices expected to support the making of prepared minds.

In these studies, the researchers are taking part in life in classrooms observing, video recording, and interviewing teachers and students. In addition, questionnaires are used. Often, the programs produce www-newsletters and other support for discussion, debate, and innovations. The

approach could be called collaborative research and development. This kind of research can be placed also under the tradition of social and educational practices. There are also other L2 projects under the name learning to learn in schools in the United Kingdom (University of Durham, Steven Higgins), as well as in Italy (Christina Stringher) and in adult education in several European countries.

The Already Achieved Level of L2: EU L2-Pilot

An attempt to combine the cognitive and the affective approach to L2 is the EU prepilot (Frefriksson and Hoskins, 2007), where a combination of the Finnish definition of L2 as the commitment (ability and willingness) to adapt to novel tasks, activating one's mastery of thinking and the perspective of hope by means of maintaining one's cognitive and affective self-regulation in-and-of learning action, and British scales, Dutch scales, and Spanish metacognition scales has been prepared and tested in seven countries (Kupiainen *et al.*, 2008). It is expected that this attempt will lead to further studies and empirical reports.

Instruments are paper-and-pencil tests and questionnaires, which comprise three dimensions – the cognitive, the affective, and the metacognitive. The cognitive is divided into identifying a proposition, using rules, testing rules and propositions, and using mental tools. The affective dimension is divided into learning motivation, learning strategies and orientation toward change, academic self-concept and self-esteem, and learning environment. The metacognitive is divided into metacognitive monitoring, accuracy, and confidence. The test is given in two booklets and the testing time has been two lessons (45 min).

Transfer as an evidence for L2

Transfer is one kind of evidence for good learning – and is very relevant for L2 studies. Transfer refers to those processes which should be improved if one has learned to learn. A person is showing his/her L2 competencies when he/she faces unexpected chance encounters, can reflect upon his/her own situation, and can adapt him/herself to possibilities and options in the task situation in order to start to learn new requirements.

Speelman and Kirsner (2005) aim to generalize transfer and learning studies into a coherent framework. According to them, transfer reflects the way in which a person adapts to a task situation. In an efficient adaptation to a task environment, people will learn to cope with a task or lesson variation, and to do so efficiently they, say, will rely on the abstraction of features that are common to many items – that is, people will acquire skills that represent their adaptation to the environment in which they are performing at present.

Schooling can be conceptualized in many ways but, in connection to L2, it is possible to say that one of the major

objective tasks of schooling is to arrange teaching and training to adapt to student variation by organizing education in schools and, within schools, in classes. The curricula determine – in their nationally specific ways – the lessons, tasks, and textbooks to provide suitable variation of instructional tasks for ensuring as general a transfer as possible. However, in this process, schooling is determined and limited to take into account student variation. These variations can be analyzed in terms of plastic general abilities or in terms of mastery of L2 cognitive competencies. It is not a question of given, unchangeable, or nonmalleable intelligence, but more of the ways through which pupils adapt themselves in the learning–development transitions at school, and of how former pupils as adult persons have learnt to cope with unanticipated task and environment variation. This learning is also a process of the growth of self-awareness in the pedagogical and intentional relation between teachers and pupils.

So What?

L2 assessment attempts to measure nonspecific outcomes of learning using tasks which are both taught in principle (as required by the curricula) and new in some (surprising) sense. Irrespective of the practical assessment solution, the whole assessment is for learning (Black *et al.*, 2006), and should guide innovation of schooling.

L2 assessments in their two major forms – learning how to learn and learned to learn – are both needed in a comprehensive system for assessment for learning. The observations and real-time interviews provide information of the ongoing processes of the lessons in a school, and the measurement and description of variations and their different contextual factors provide information about the school as an institution; both of these require communication. It appears that this communicative act can take place in three instances. L2 assessment outcomes inform (1) teachers about their success in communicating during and through lessons how failures and errors are treated post-failure, reflectively – that is, it is how pupils learn that errors and failures can be of use in the process of learning; (2) how teachers have succeeded in telling the criteria of records and markings which are delivered to pupils at the ends of terminus – that is, how well the seriousness of schooling has been understood; and (3) how accurately the school as a whole has understood tasks and goals of education, as a mediator of precious knowledge of past generations to the next new generation.

Bibliography

- Black, P., McCormick, R., James, M., and Pedder, D. (2006). Learning how to learn and assessment for learning: A theoretical inquiry. *Research Papers in Education* 21(2), 119–132.

- Claxton, G. (2007). Expanding young people's capacity to learn. *British Journal of Educational Studies* 55(2), 115–134.
- Cronbach, L. J. and Snow, R. E. (1977). *Aptitudes and Instructional Methods. A Handbook for Research on Interactions*. New York: Irvington Publishers.
- Education Council (2006). *Recommendation of the European Parliament and the Council of 18 December 2006 on Key Competencies for Lifelong Learning*. Brussels: Official Journal of the European Union, 30.12.2006.
- Fredriksson, U. and Hoskins, B. (2007). The devevelopment of learning to learn in a European context. *Curriculum Journal* 18(2), 127–134.
- Hautamäki, J., Arinen, P., Eronen, S., et al. (2002). *Assessing Learning-to-Learn. A Framework. Evaluation 4*. Helsinki: National Board of Education.
- Hirsch, E. D. (1996). *The Schools We Need and Why We Don't Have Them*. New York: Doubleday.
- James, M., Black, P., Carmichael, P., et al. (2006). *Learning How to Learn. Tools for Schools*. London: Routledge.
- Klauser, K. J. (1988). Teaching for learning to learn: A critical appraisal with some proposal. *Instructional Science* 17, 351–367.
- Kupiainen, S., Hautamäki, J., and Rantanen, P. (2008). *European Union Pilot Test on Learning to Learn. A Report to EU*. Helsinki: Centre for Educational Assessment.
- Meijer, J. (2007). Cross-curricular skills testing in the Netherlands. *Curriculum Journal* 18(2), 155–174.
- Resnick, L. (1987). *Learning to Think*. Washington, DC: National Academy Press.
- Smith, R. M. (ed.) (1990). *Learning to Learn Across the Life Span*. San Francisco, CA: Jossey-Bass.
- Snow, R. (1994). Abilities in academic tasks. In Sternberg, R. J. and Wagner, R. K. (eds.) *Mind in Context. Interactionist Perspectives on Human Intelligence*. Cambridge: Cambridge University Press.
- Speelman, C. P. and Kirsner, K. (2005). *Beyond the Learning Curve. The Construction of Mind*. Oxford: Oxford University Press.
- Crick, R. D. (2007). Learning how to learn: The dynamic assessment of learning power. *Curriculum Journal* 18(2), 135–153.
- Demetriou, A. and Kazi, S. (2006). Self-awareness in g (with processing efficiency and reasoning). *Intelligence* 34, 297–317.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring. A new area of cognitive-developmental inquiry. *American Psychologist* 34(10), 906–911.
- Hautamäki, J., Kupiainen, S., and Arinen, P. (2006). Learning-to-learn assessment in Finland – versatile tools to monitor and improve effectiveness and equity of the educational system. In Niemi, H. and Jakku-Sihvonen, R. (eds.) *Research-Based Teacher Education in Finland. Research in Educational Sciences 25*. Turku: Finnish Educational Research Association.
- Hoskins, B. and Fredriksson, U. (2008). *Learning to Learn: What is It and Can It Be Measured? JRC Scientific and Technical Reports EUR 23432 EN 2008*. Ispra: Joint Research Centre of European Commission.
- Hounsell, D. (1979). Learning to learn: Research and development in student learning. *Higher Education* 8, 119–132.
- Jensen, A. R. (1998). *The g Factor. The Science of Mental Ability*. Westport, CT: Praeger.
- Moreno, A. and Martin, E. (2007). The development of learning to learn in Spain. *Curriculum Journal* 18(2), 175–193.
- Nuthall, G. (1999). Learning how to learn: The evolution of students' minds through the social processes and culture of classroom. *International Journal of Educational Research* 31, 129–257.
- Olson, D. (2003). *Psychological Theory and Educational Reform*. Cambridge: Cambridge University Press.
- Shayer, M. and Adey, P. S. (eds.) (2002). *Learning Intelligence: Cognitive Acceleration across the Curriculum From 5 to 15 Years*. Milton Keynes: Open University Press.
- Snow, R. (1990). New approaches to cognitive and conative assessment in education. *International Journal of Educational Research* 14, 455–473.
- Waeytens, K., Lens, W., and Vandenberghe, R. (2002). 'Learning to learn': Teachers' conceptions of their supporting role. *Learning and Instruction* 12, 305–322.

Further Reading

- Adey, P., Demetriou, A., Csapo, B., Hautamäki, J., and Shayer, M. (2007). Can we be intelligent about intelligence: Why education needs the concept of plastic general ability? *Educational Research Review* 2, 75–97.
- Bonnet, G., Svecnik, E., Hautamäki, J., et al. (2006). *Final Report of the Learning to Learn Expert Group to the European Commission, DG EAC A6*. Paris/Brussels/Ispra.

Relevant Websites

- <http://www.learn2learn.ac.uk> – Learning how to Learn: A Project of the ESRC Teaching and Learning Research Programme.
- <http://www.learning2learn.eu> – Learning to Learn (L2L).

Assessment in Schools – Literacy Writing (Extended)

J M Parr, University of Auckland, Auckland, New Zealand

L Earl, Aporia Consulting, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Writing research is a potentially dynamic field as the practices, uses, and modes of writing are constantly changing, particularly with the rapid growth of new technologies. Research in the field is eclectic, drawing on varied traditions, notably those of rhetoric, cognitive psychology, and linguistics. As a theoretical construct, writing is complex. Writing is a social and cultural act; it is problematic to specify what develops or progresses in writing and what it develops toward. Evaluation of writing is an area under-researched and under-theorized. (The term evaluation is used to refer to the process of making a judgment. If the use of the information is for an end-point type decision, a summative purpose, the word testing is employed. If the purpose is to inform the teaching and learning of writing, then the term assessment or formative assessment is used.) Links between theory in writing and evaluation of writing are not well articulated; equally, the links between evaluation and the teaching of writing are underdeveloped.

This article examines the current status of evaluating writing, particularly in relation to classrooms and teaching. This is particularly apt, given that key ideas and principles from cognitive, constructivist, and sociocultural theories have reconceptualized assessment and its role in teaching and learning. While the written form has long been used to examine knowledge, it is only relatively recently that evaluation of writing has been seen as something that is an integral part of informing teaching and learning. This article commences by briefly considering the substantial body of work on writing tests designed to make a specific judgment – to select, place, and grade – then moves to portfolios as a means of gathering data to evaluate writing. Finally, the relatively limited research related to assessing student writing in ways that potentially inform teaching and learning is examined. There is brief reference to response to writing in the form of conferences which have the potential as a site for formative assessment, and written response which is, sometimes, employed formatively on draft work although, more commonly, it accompanies a grade to justify it and/or provide feedback designed to improve subsequent performance. The use of assessment tools in writing designed to provide information for formative purposes concludes the discussion on assessment of writing.

Evaluating Performance: Tests of Writing

Writing is the means by which students in educational settings present and display what they know. As students move through the schooling system, their competence in using writing for this purpose assumes greater importance. Some tests may examine the ability to write directly or the ability to write may impact indirectly by content being tested in essay form. The widespread use of written tests to select, place, certify, or hold to account has been a shaping influence on writing scholarship. Writing as a means of testing literacy skills and intellectual abilities was seen in the Han dynasty (206 BC to AD 220) with its formalized system of written examinations for selection for imperial, governmental, or feudal service; in Britain, written tests became an efficient way to select candidates for the civil service. Universities employ written tests to select and place although, more commonly now, student writing performance is tested to determine the need for support to increase chances of success in tertiary study.

The resulting emphasis in writing evaluation research has been on judgments whose purpose (and timing) suggests they are of a summative nature. A measurement emphasis has led to issues in the evaluation of writing being framed technically. In fact, scholars have often characterized the history of testing of writing as one of balancing the requirements of reliability and validity with a large body of literature discussing how to obtain, maintain, and enhance agreement among markers and other literature dealing with the effects of testing variables such as prompts or time limits.

Criticism has centered around two major issues. The first concerns the fact that scholarship around testing of writing was based on outmoded notions of validity and seldom considered the implications of decisions – including how testing might connect to, and improve, learning. Teachers of writing tended to shun writing tests as something that either seemed unrelated to teaching or had a negative effect in that administering and scoring placement tests, for example, took time and energy away from teaching. There was little questioning of why some methods of evaluating writing are privileged and what the outcome of use of writing-test scores is both in terms of making the best educational decisions and in terms of unintended consequences.

There has been relatively limited attention to the consequences of testing writing, although unintended consequences have been a focus of some recent research. Tests impact on curricula and instruction – a complex phenomenon referred to as washback. At present, widespread accountability testing has highlighted views with regard to the negative effects of assessment on teaching. Hillocks – in *The Testing Trap: How State Writing Assessments Control Learning* – points out that assessments, not standards, have an influence on what happens in classrooms in terms of rhetorical stance, instructional mode, and writing process; they privilege curricula content that appears related to the assessment, such as particular forms of writing. They downplay the communicative interaction between writer and readers, turning writing into an act of formal display. Scoring rubrics can indicate what is perceived as important in writing and promote a narrow definition of writing by students.

The second focus of criticism centered around the idea that writing testing – unsupported by resources commonly used by writers and involving a single, often timed, response to a prompt – was out of step with current views of literacy and how it is acquired. As writing theory changed, its conceptions no longer matched the construct implied in – and the information provided by – traditional formats for evaluating writing.

Moving Beyond Traditional Test Formats: Writing Portfolios

The introduction of portfolios as a means of gathering writing for evaluation came at a time when composition studies moved from a focus on form and products to the processes involved in writing. The portfolio concept as it applied to writing has been discussed in a number of collections that incorporate diverse perspectives. Similar to the craft origin of the concept, portfolios are, essentially, multiple examples of writing which could be gathered over time to present richer, more complex information concerning a writer's efforts, processes, progress, or performance. Portfolios allow writers to demonstrate writing for a range of purposes, for different audiences, and in different contexts, thus seeming to encompass a view of writing more aligned with theories of writing as social practice to accomplish communicative goals (although social motives and functions of communication were not necessarily captured in the writing represented). They were also thought to be a way of collecting information to evaluate writing performance that might impact positively on instruction and student outcomes.

Although they originated in the classroom as part of a move to connect assessment to the process of learning, portfolios have morphed and the initial uses

for which they were designed have been extended and adapted. They have been widely employed for large-scale assessment – for comparisons across classrooms at different levels from school to state. These attempts to link classroom portfolio assessment to large-scale testing are seen as problematic. The limitations of portfolios – from pragmatic, resourcing issues to questions of scoring and standardization to notions of empirical validation – were widely addressed in scholarship of the 1990s.

Portfolios have the potential to be used for different purposes – to illustrate development or patterns in writing behavior over time or differential expertise across genres, purposes, or social contexts; to allow teachers insight into a writer's strengths and weaknesses and to help teachers to hone teaching to meet student needs. They provide a potential site for collaborative inquiry of student and teacher into student writing and an opportunity for students to learn to exercise judgment with regard to their own work to reflect and to self-assess. Ideally, portfolios would not be used to summarize performance but to serve a formative function linking instruction, performance, and ongoing assessment to guide the learning process.

Assessment of Writing for Teaching and Learning

While a major issue in testing writing has been aligning evaluation to theories of writing, a major concern in formative assessment is working out how to move toward the desired outcome whatever that may be. Whether – in writing – this outcome is a defined goal or a broad horizon is a dilemma discussed in the literature. The goal model of progression is like a skills or knowledge model which assumes that what is necessary to be good at writing is not only known but can be measured and then turned into some form of teaching program. The horizon model suggests less specified outcomes but also multiple pathways.

In order to understand progression in writing, it appears that teachers may rely less on a knowledge model than on value judgments that are formed through the extensive process of making those judgments. Teachers are thought to develop understandings of progression and make reliable judgments of written work through the process of construct referencing. Understanding of a construct is refined through experience and processes, such as moderation where collegial discussion is involved. The shared meaning that develops among those interpreting the evidence in the form of the writing has been referred to as guild knowledge.

Underpinning formative assessment is the promotion of student autonomy and self-regulation. To promote these, students are initiated into the guild knowledge that guides teachers' judgments with regard to student

writing. Knowledge is built for students through classroom practices closely associated with formative assessment such as sharing goals and what counts as success with the learner – through feedback, peer reviews, self-assessment, and the use of exemplars.

The research community has barely begun to consider formative assessment in writing. Formative assessment can be seen to be of two kinds: planned and interactive. The purpose for planned formative assessment is to obtain information from the whole class with regard to progress in learning through undertaking a specific activity. Interactive formative assessment, on the other hand, is that which takes place during student–teacher interactions arising out of a learning activity. In writing, the research literature is framed around the traditional, major ways in which teachers of writing would consider they gain information in order to use it to give feedback to writers, namely, by responding to the writer through conferencing and by providing written feedback. As potential sources of information to be used formatively, peer- and self-assessment have received some research attention with respect to writing.

Potential for Formative Assessment: Responding to Developing Writers

Teachers need to make an assessment of student writing in order to respond. Researchers have struggled to delineate the conditions for constructive response in writing. Writers need such response to be in a form that allows them to monitor their own progress toward a desired outcome and to help decide how to move forward; but they also need the response to be such that they discover their readers' needs. Teacher response is given in the context of an unequal power relationship; they have an ambiguous, often conflicting, role as facilitator, evaluator, and audience and the issues of text appropriation and ownership surface. Various vehicles for feedback on writing are possible, including conferencing, teacher-written feedback, and peer feedback. There is also feedback from self-assessment. While conferencing is more a feature of elementary writing classrooms, written feedback characterizes classrooms beyond this level. The sites for response only provide the opportunity for obtaining and using information for formative purposes.

Conferencing as a Site for Responding to Writing

Conferencing is viewed as fundamental to teaching and learning in process-oriented writing classrooms. In conferences, teachers learn about writers and developing writers learn to interact with – and craft meaning for – a reader.

Effective conferences provide writers with opportunities to develop the meta-cognitive awareness related to the writing process and the self-regulatory strategies required for reflecting on their texts, together with the personal responsibility necessary to becoming a writer.

Literature on conferencing is grounded in practice. Practice-oriented, classroom-based writers clearly locate the role of the teacher as a follower of a developing writer's lead, rather than that of an instructor. Young writers are to take responsibility for managing and reflecting on their own learning. The role of the teacher is couched in terms of listening for clues with regard to the kind of support writers need and providing that support at the appropriate moment. For teachers, even the most skilled conferencing is complex and challenging. The research illustrates numerous problematic issues surrounding conferencing, such as teacher control and a predominance of a narrow, low-level focus. There is little research that illustrates how the properties of the interactions are used to achieve targeted instructional purposes in a particular context. Conferences contain opportunities for evidence-informed, moment-by-moment decision making by teachers assessing where the writer currently is and the best way to move them forward. Effective writing conferences – where teachers understand and respond appropriately to student need – move learners' development forward in meaningful ways.

Written Comments as Response to Writing

Although feedback on writing products has been considered a common form of writing instruction at secondary and tertiary levels, scholarship on response to student writing has not been theoretically grounded and linked to assessment and improvement. Written feedback is intended to improve student learning in writing; however, most of the studies have been conducted outside of any context – pedagogical, theoretical, or communicative. Research has treated the texts that teacher-responders create as if they stand alone, ignoring the perspective that the meaning of text will be constructed differently depending on the discourses brought to bear on the text by the reader.

In general, students do not find written feedback on their writing either helpful in itself or as a catalyst for discussion; the traditional ways in which teachers make comments on students' drafts are seen as not effective in improving students' writing; perhaps, because the work is not often viewed as one in progress and feedback is corrective rather than designed to foster development.

A major stumbling block to instructive feedback is the nondialogic nature of much of the reported practice. In commenting on drafts, teachers – in their role as expert and ultimate evaluator – impact on what decisions students make with respect to revision. Student writers,

arguably, have a constrained level of choice with regard to whether or not to use the feedback. Students are seldom co-negotiators, co-evaluators. The lack of critical involvement by the student lessens the likelihood of the feedback becoming internalized and having effect beyond the current piece. Huot argues that comments should be transformative; they should be open-ended and force students back into their text.

Reflective and interactive logs are one way to do this. Dialoging is a means by which teachers help students to think about their writing and writing processes and to develop agency and control. Logs have taken the form of a learning journal where the teacher may write feedback and the student records thoughts and goals and examines and explores these to better understand their writing. Through the log, the student interacts with the teacher (or others) in an ongoing, formative, and dialogic fashion to enhance writing performance.

Self and Peer-Assessment

Teachers of writing describe methods by which they incorporate student self-assessment into their classroom through reflective writing including through contracting criteria that describe the work required to reach a particular level or grade, and through written or oral dialogs in which student and teacher evaluate writing together. The benefits of reflection and self-assessment are readily claimed and there is a developed rationale for linking self-assessment to achievement, but there is limited empirical support in terms of research into the forms and consequences of this, particularly in writing.

The dominant form of self-assessment in writing discussed in the literature involves the selection of work for portfolios or reflection on the completed portfolio and the recording of the evaluation often through a memo or covering letter. These acts are, generally, not for a formative purpose and are, arguably, too late to impact student learning. Inviting student input in an attempt to change the teacher–student dynamic of grading does not necessarily change the consequences or outcomes of the activity. The argument is made that the very nature of teaching makes it reasonable that, in self-assessing, students may simply reflect back teacher language and judgments; it is also noted that both selection and reflection are frequently teacher guided.

In the broader self-assessment literature, there has been an emphasis on explicit articulation of assessment criteria in order to promote understanding of what is needed. Describing standards in relation to writing is acknowledged as difficult and such is best transferred in joint evaluative activity through using annotated exemplars, together with the verbal descriptors. In writing, approaches to understanding assessment have utilized the principle of annotated samples which help students

gain a sense of perspective with regard to the quality of their written work and to both self- and peer-assess appropriately.

Peer response has been primarily utilized as a means to enable writers to receive feedback from an audience rather than as an explicit form of peer assessment. It has been instantiated in practices in elementary school, such as the author's chair. It is regarded as easier to evaluate and give feedback on a peer's work than on one's own. However, research suggests that peer-evaluation and assessment tasks that cede control to students are not a common occurrence in writing classrooms.

Supporting Formative Assessment Practices in the Classroom: Planned Assessment of Writing

The message in many writing-resource books for teachers presents a traditional view of classroom assessment as almost an afterthought phase in the planning cycle where what is taught is determined by curriculum objectives and writing-process specifications. A focus on instructional goals, however, does not help in decision making concerning student learning needs. Planned formative assessment can provide the teacher with insight into the particular learning needs of individual students and groups of students and form the basis for differentiated teaching of writing.

A major issue with respect to enabling formative assessment in writing relates to the notion of guild knowledge – the teacher's professional and content knowledge in the area of writing. Teachers have been shown to lack knowledge of how language works. This knowledge is needed in order to notice and recognize what is significant in terms of written language and then to respond. Arguably, teachers need considerable support in working out how to diagnose and teach to the needs of their developing writers. How to provide this support without an entirely prescriptive lock-step progression to a closely defined horizon is a problem that was addressed in the design of a planned formative assessment tool to assist teachers to gain information from student writing in their classrooms. (The tool is part of project as TTle (Assessment Tools for Teaching and Learning) which provides detailed assessment against curriculum objectives in reading, writing and mathematics for years 4 to 12. It is a CD Rom-based/web-based assessment suite which gives teachers choice in the design and timing of assessments and access to a range of reporting formats, including comparisons to norms.)

This tool is an example of a diagnostic assessment that aligns with current theory of writing as a social and cultural practice where the term genre refers to the processes involved in getting things done through language.

The theoretical framing views writing broadly as serving social purposes: writing was conceptualized as serving six major purposes – a core set of generic processes that encapsulate what the text is doing. For each of the major purposes that inform – or processes that form – texts, an analytic rubric was developed. The content of these rubrics drew on research which considers that a text can be seen from two perspectives, namely, as something that can be recorded, analyzed, and discussed, but also as a process that is the outcome of a socially produced occasion. Forms of text (i.e., genres) produced in and by specific social institutions (such as within schooling) will have some stability. Therefore, descriptions of features and text structures commonly associated with a generic social purpose were provided in the scoring rubrics.

The rubrics assess student performance using criterion statements relating to seven dimensions of writing. While, all dimensions of analysis of the text are seen as interdependent in terms of judging the effectiveness of the piece of writing, for purposes of assessment, the dimensions are considered and scored separately. Within the same framework and dimensions, the criteria at each of the levels articulate a developmental progression. This helps teachers to work out what the next level of development – the way forward – might look like for any given writer. They map a writer's development, allowing diagnosis of areas of difficulty and those of strength to be situated within the context of the communicative purpose of the task.

The diagnostic assessment tool is intended to be just one of those in the writing-teacher's armory, albeit a fairly powerful one. The classroom teacher who knows the students in a variety of contexts is able to view the results of a single assessment piece of writing in the light of other sources of information and make informed decisions concerning, likely, the most efficacious instruction. Further, the discussion surrounding moderating writing samples builds teachers' knowledge.

Conclusion

A consideration of the scope of the literature related to evaluation of writing suggests that scholarship has spent considerable time and effort focusing on performance testing for summative purposes. Technical concerns are an instance of a research literature which reveals a surprising disconnection to practice. Even in an area as fundamental to teacher practice in writing as response – while principles of sound written feedback or conferencing have been advocated – research has not established links between these and efficacious outcomes for developing writers. The field has been tardy in taking on board notions of validity that foreground the use and consequences of evaluation. There is a significant gap in the literature in terms

of research that considers the assessment of writing for formative, diagnostic purposes, to inform teaching, and to support learning in a forward-looking manner.

See also: Alternative Assessment; Assessment and the Regulation of Learning; Formative Assessment and Instructional Planning; Formative Assessment; Moderation of Student Work by Teachers.

Further Reading

- Ball, A. (1997). Expanding the dialogue on culture as a critical component when assessing writing. *Assessing Writing* 4, 169–202.
- Black, L., Daiger, D., Sommers, J., and Stygall, G. (eds.) (1994). *New Directions in Portfolio Assessment: Reflective Practice, Critical Theory, and Large-Scale Scoring*. Portsmouth, NH: Boynton.
- Brown, A. (2005). Self-assessment of writing in independent language learning programs: The value of annotated samples. *Assessing Writing* 10, 174–191.
- Calfee, R. C. and Perfumo, P. (eds.) (1996). *Writing Portfolios: Policy and Practice*. Hillsdale, NJ: Erlbaum.
- Elbow, P. (1993). Ranking, evaluating, and liking: Sorting out three forms of judgement. *College English* 55, 187–206.
- Faigley, L. (1989). Judging writing, judging selves. *College Composition and Communication* 40, 395–412.
- Fife, J. M. and O'Neill, P. (2001). Moving beyond the written comment: Narrowing the gap between response practice and research. *College Composition and Communication* 53, 300–321.
- Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing Writing* 8(1), 5–16.
- Haneda, M. (2004). The joint construction of meaning in writing conferences. *Applied Linguistics* 25, 178–219.
- Hillocks, G. (2002). *The Testing Trap: How State Writing Assessments Control Learning*. New York: Teachers College Press.
- Huot, B. (2002). *Re-Articulating Writing*. Logan, UT: Utah State University Press.
- Huot, B. and Neal, M. (2006). Writing assessment: A techno-history. In MacArthur, C. A., Graham, S., and Fitzgerald, J. (eds.) *Handbook of Writing Research*, pp 417–432. New York: Guilford.
- Hyland, F. (2000). ESL writers and feedback: Giving more autonomy to students. *Language Teaching Research* 4, 33–54.
- Knapp, P. and Watkins, M. (2005). *Genre, Text, Grammar: Technologies for Teaching and Assessing Writing*. Sydney, NSW: University of New South Wales Press.
- Kress, G. (1993). Genre as social process. In Cope, B. and Kalantzis, M. (eds.) *The Powers of Literacy – a Genre Approach to Teaching Literacy*, pp 1–21. London: Falmer Press.
- Luce-Kepler, R. and Klinger, D. (2005). Uneasy writing: The defining moments of high-stakes literacy testing. *Assessing Writing* 10, 157–173.
- Mabry, L. (1999). Writing to the rubric: Lingering effects of traditional standardized testing on direct writing assessment. *Phi Delta Kappan* 80, 673–686.
- Marshall, B. (2004). Goals or horizons – the conundrum of progression in English: Or a possible way of understanding formative assessment in English. *Curriculum Journal* 15, 101–113.
- Murphy, S. (2000). A sociocultural perspective on teacher response: Is there a student in the room? *Assessing Writing* 7, 79–90.
- O'Neill, P. (2003). Moving beyond holistic scoring through validity enquiry. *Journal of Writing Assessment* 1, 47–65.
- Parr, J. M., Glasswell, K., and Aikman, M. (2007). Supporting teacher learning and informed practice in writing through assessment tools for teaching and learning. *Asia-Pacific Journal of Teacher Education* 35, 69–87.
- Sperling, M. (2004). Constructing the perspective of teacher as reader: A framework for studying response to student writing. *Research in the Teaching of English* 28(2), 175–207.

- Sperling, M. and Freedman, S. (1987). A good girl writes like a good girl: Written response and clues to the teaching/learning process. *Written Communication* **4**, 343–369.
- Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.
- White, E. M., Lutz, W. D., and Kamuskiri, S. (eds.) (1996). *Assessment of Writing: Politics, Policies, Practices*. New York: The Modern Language Association.
- Williamson, M. and Huot, B. (1993). *Validating Holistic Scoring for Writing Assessment: Theoretical and Empirical Foundations*. Cresskill, NH: Hampton.
- Yancey, K. B. (1999). Historicizing writing assessment. *College Composition and Communication* **50**, 483–496.
- Yancey, K. B. and Weiser, I. (eds.) (1997). *Situating Portfolios: Four Perspectives*. Logan, UT: Utah State University Press.

Assessment in Schools – Mathematics

M Brown and J Hodgen, King's College London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Trends in Assessment in Mathematics: 1970–90

In the 1970s and 1980s, the main developments in assessment in mathematics were new and wider forms of assessment than the traditional written tests and examinations of technical performance, which had previously predominated in many countries. Whereas the earlier models had been mainly for purposes of teachers' and schools' monitoring of learning, end-of-course certification, or academic/vocational selection, the new forms of assessment were intended to improve the validity of the assessment of individual students against wider curricular objectives.

This broadening of the objectives of learning mathematics mainly reflected socioeconomic changes resulting from greater automation of employment and growing social equality. Previously, for the majority, it was thought that some basic calculation and measuring skills would suffice in everyday life and in jobs requiring manual and/or office-clerking skills. An elite minority were inducted into traditional mathematical knowledge in order to pass examinations allowing entry to careers, predominantly in mathematics teaching, engineering, and accountancy. A few of these would be needed to be original thinkers, but this was a talent perceived to arise rarely and naturally. Thus, assessment was mainly of proficiency in calculating or more technical procedures, and of knowledge of mathematical results.

With the advent of computers, calculators, and automation in the workplace, it was clear that numbers in purely manual jobs would decline and many more people would need to be able to interpret and make decisions based on the output of computers, calculators, and other machines, and to employ this technology to solve problems in both work and life. The pace of change and the degree of complexity of life were accelerating, alongside a greater respect for the rights of, and thus increasing autonomy for, all classes and races, and both genders.

Mathematics education adapted to become more utilitarian and less elitist, and more oriented to interpretation, reasoning, problem solving, and communication skills (e.g., National Council of Teachers of Mathematics, 1989). Tests of knowledge and procedures alone, whether set in abstract contexts or in artificial real-world contexts where there was just the right amount of information in easily manipulable form, were no longer perceived to provide valid ways to assess the wider set of skills and processes which formed the

new curriculum. Further, traditional tests were perceived as not allowing students who could, given time, use mathematics effectively to solve practical problems, to demonstrate their competencies; some such students had frequently scored few marks in abstract written tests.

Thus, mathematical assessment had developed in the 1970s and 1980s to incorporate more authentic assessment based in classrooms and allowing students to demonstrate their wide range of skills over longer time spans than traditional examinations. Evidence might be presented in the form of student portfolios, which allowed external validation of grades. Some countries introduced assessment of mathematics through practical tasks, using everyday equipment. Others used extended practical problems and/or mathematical investigations, either in the classroom or in controlled conditions. Some of these more open tasks allowed students to tackle the problem using whatever level of mathematics they knew, demonstrating assessment by outcome rather than assessment by task.

These assessments were designed not only to be more valid using the traditional sense of validity, but also in relation to the more contested notion of consequential validity (Messick, 1994). Thus, widening the assessment was designed to encourage broader forms of classroom teaching, demonstrating the power of assessment as a lever for curriculum change.

The purposes of assessment were no longer limited to monitoring of achievement, certification, and selection, but reflected both the greater technical background to assessment and the human rights developments in wanting to provide a broad description of the developing mathematical competencies and skills of individual students across the full attainment range, using the notions of criterion-referenced assessment which took into account the broad categorization of assessment objectives first outlined by Bloom *et al.* (1971). Rather than tests taken at particular times of year, ages, or grade levels, assessment could document progression in individual mathematical attainment – in England, this linked with the graded assessment movement.

With the increasing use of information and communication technology (ICT) in assessment, in the US criterion-referenced testing developed into adaptive testing, where the computer could adapt the later questions asked to the knowledge demonstrated in answers to earlier questions, thus providing a more efficient use of testing time and a more comprehensive description of individual attainment. However, the items used were restricted to multiple-choice

computer-marked format, which were better suited to knowledge, procedures, and understanding rather than analysis and synthesis. Adaptive tests required the design and use of an item bank related to specific objectives in the assessment domain. Item banks were also introduced during this period and used in the 1980s for a number of purposes, including (e.g., in Scotland) teacher-based national assessment.

Trends in Assessment in Mathematics: 1990–2008

What has changed in mathematical assessment since about 1990 has been less the modes and techniques of assessment and more the purposes. These include a desire for greater accountability for student progress not only at national or state level but also at school and teacher level. This requires more standardization, a process to which increasing deployment of ICT can contribute. However, the consequences of such assessments in terms of narrowing the mathematics curriculum require continuing searches for practical ways to broaden summative assessment in schools.

Finally, in opposition to these influences on summative assessment, there has been a growing emphasis on formative assessment in mathematics as a means to improve both teaching and student attainment.

These trends are described in more detail in the sections which follow.

Assessment of Mathematics as Part of School Accountability

Alongside mathematical assessment for national accountability, using the now well-known international Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA) surveys which provide league tables of countries, in some countries, there has been a significant growth in assessment for school accountability, holding schools accountable for their students' results and progress. Test results are also being used to evaluate and manage the performance of individual teachers.

The growth in accountability testing may also reflect a change in perceptions of the nature and distribution of mathematical ability. While external assessment of mathematics was once regarded as appropriate only for the minority endowed with mathematical talent, it later became a democratic necessity to recognize and improve all students' mathematical attainments, and now all students are regarded as capable of and entitled to attain a specified national standard, or at least a minimum degree of progress, with schools held responsible if they fail.

Such systemic use of assessment for evaluation and accountability has spread with the increased use of large

and complex databases, some at national level, which allow monitoring of the progress of individual children at intervals through their whole period of schooling. In England, the government has announced new plans for parents to be able to check using the Internet, at any time they wish, their child's detailed progress up the levels of the national curriculum. Partly because of its perceived hierarchical structure, mathematics has been a key subject in these developments, often alongside national languages.

Initially, school performance may be judged on the proportions of students who reach one or more specified attainment standards. This requires either strict moderation of teachers' assessment within and between schools, or national- or state-designed tests to guarantee comparability. Where teacher performance is also being measured, there is likely to be a call for external forms of assessment.

The more sophisticated databases, which include results of successive test administrations, such as those used in England and in separate states in the USA, allow schools to be judged not only on absolute standards but also, more equitably, on the progress their students make from their previous test results. They also allow school data to be reported in terms of ethnic, gender, and socio-economic groups. This systematic testing and data collection forms the basis of the No Child Left Behind Act introduced by the Bush administration in the USA and the national testing in England. One result of the detailed reporting of results for different groups and the raised expectations for all students is that the gender gap in mathematical achievement has been almost eliminated in most countries and differences between ethnic groups are at last gradually closing. However, differences in mathematics attainment between different socioeconomic groups, although generally less than those in languages, seem more intractable.

External testing provides a measure of school success that is open and apparently objective, enabling identification and remedial action for schools where mathematics teaching is poor. However, the downsides are now being discovered in England and in the USA, where the testing systems have been running for the longest.

In relation to mathematics, externally designed tests tend to not only lead to excessive time being spent on test practice but also to narrow the curriculum to include only those topics, and those types of artificial single-step question, which are included in the tests. There is little incentive for teachers to include in the curriculum work on longer and more realistic problems, for them to organize students to work in groups, or to spend time using calculators or computers, if this is not included in the external assessment. Teachers, who are under pressure to achieve high test results, have limited opportunities to concern themselves with wider aims of education, such as independent thought, creativity, positive attitudes to learning,

application to novel situations, depth of understanding, or ability to retain ideas and skills over long periods.

Children may become anxious about the assessments and those who know as the result of repeated practice that they are likely to do badly in mathematics may well develop negative attitudes. Data from TIMSS (Mullis *et al.*, 2008), for example, show that between 1995 and 2007, during the period when there has been increased emphasis in England on schools meeting test targets, student attitudes to mathematics have dropped significantly although their test results have risen.

Some administrations have sought to reduce these ill-effects by not publishing the results (e.g., Scandinavian countries), or by testing at the start rather than at the end of the academic year (e.g., France). In such countries, the results are less to evaluate school performance and more to provide teachers with comparative information about their students' performances on which they are expected to act.

Finally, the more complex the testing system the more likely it is to become logistically difficult to administer it. Such problems are one reason, but not the only reason, to lay more emphasis on ICT-based assessment which avoids the need to employ and train large numbers of markers.

The Use of ICT in Mathematics Assessment

As elsewhere in education, the use of ICT in the assessment of mathematics has increased, particularly in marking and administration. In mathematics, this use of technology in assessment lags behind its use in teaching and learning. This lag, in turn, has a limiting effect on technology use in classrooms as teachers prioritize preparation for high-stakes assessments, which remain predominantly pencil and paper based.

Calculators of all types remain distrusted by many. Hence, although students are commonly allowed to use them within high-stakes assessments, this use is still sometimes restricted.

It is likely that pressures to increase efficiency and to reduce costs will result in greater use of computer-based testing, including adaptive testing as already noted, although there remain significant difficulties that are yet to be overcome. Increasingly, in the developed world, students have ready access to computers in and out of schools and access is no longer as significant a problem as it once was. However, students taught in a pencil-and-paper mode may find it difficult to fluently handle mathematical symbolism on a computer in an assessment.

Computer-based tests are dominated by the multiple-choice format. However, there have been various innovations that allow students to use computers to do mathematics using constructed-response items. Some innovations have explored the potential for the use of more realistic contexts in school mathematics (Richardson *et al.*, 2002). There is some evidence that dynamic and colorful

on-screen presentation leads to greater student motivation, although this may be a novelty effect.

Computer-based assessment does have several potential advantages over pencil and paper, such as single-screen item presentation and interactivity. Interactivity can lead to less emphasis on visualization strategies. Candidates tend to use guess-and-check strategies more often and appear less willing to use pencil-and-paper jottings. Whether these different affordances lead to increased validity may depend on the interpretation of the concepts being assessed. Comparison of equivalent items presented in both formats reveals that for some items on-screen presentation increases validity, whereas for others validity is reduced (Threlfall *et al.*, 2007).

There has also been increasing use of technology for formative assessment purposes, for example, technologies that allow students to vote in real time thus providing immediate feedback to the teacher. There is little evidence as to whether such technologies are effective. Some research points to the efficacy of integrated learning packages in which computers provide feedback on learning, although such studies are generally small scale and related to specific interventions. There is increasing use of the Internet, particularly for students' individual use in preparing for high-stakes examinations, although again research in this area is limited.

Practical Ways of Broadening Summative Forms of Assessment

In the section titled 'Trends in assessment in mathematics: 1970–90' of this article, it was noted that by the 1980s experimental forms of assessment found richer and more practical tasks were being pioneered by a number of agencies in different countries, for example, the Freudenthal Institute in the Netherlands, the Shell Centre, the National Foundation for Educational Research, and King's College London in England, and the University of California at Berkeley in the United States.

To some extent, this was incorporated into international, national, or state-organized assessments, mostly using centrally designed tasks which were either centrally marked or assessed by trained teachers or local markers. For example, the TIMSS survey of 1995 included voluntary participation in the assessment of practical mathematics using problem-solving items administered and assessed by trained visiting staff. Mathematics examinations at age 16 in England included alongside external tests an element of rich activities set in the classroom and for homework over several weeks and marked by teachers according to clear instructions. These activities were voluntary from the late 1980s, but compulsory by the late 1990s. In other countries, teachers included students' performances in this type of activity into school-designed and/or portfolio-based summative assessment.

There is a major problem with the inclusion of these longer tasks in high-stakes summative assessments. Accountability measures have encouraged teachers to provide more guidance than was intended to increase the standards achieved by the teacher and the school, and have encouraged students to demand assistance from teachers, parents, and peers to enable them to submit perfect pieces of work. The availability of the Internet has also made it difficult to stop students accessing solutions posted by others inside or outside the school. Furthermore, if the problems are complex then marking equally becomes complex and it will be difficult to ensure reliable marking across students, teachers, and schools.

Although there is pressure in high-stakes assessment to avoid or remove rich, extended, problem-solving items (both pure and applied), investigation, discussion, and use of computers wherever appropriate, to encourage teachers not to narrow the curriculum it would seem to be necessary to incorporate some form of this activity as part of the summative assessment. There are several possible solutions to this problem.

One of the solutions is that adopted by the PISA surveys. PISA is administered for the Organisation for Economic Co-operation and Development (OECD) by the Australian Council for Educational Research, and in mathematics the aim (in contrast to those of TIMSS) is to assess the attainment only of mathematical literacy (the ability to apply mathematics to model and solve realistic problems). The assessment framework was developed internationally but the initial mathematics items were mainly developed by the Freudenthal Institute, reflecting a tradition of realistic mathematics education (Van Den Heuvel-Panhuizen and Becker, 2003).

Although some PISA items are short and can be computer marked, others are designed to be more complex, to require application of more than one idea or process and some degree of strategy/creativity/reasoning/communication skills. However, these still have to be restricted in scope to fit within a short timed test and, hence, none can take longer than about 20 min. They are externally marked by trained markers. Such short items are not ideal as an assessment of extended problem solving. Furthermore, in the absence of an explicit curriculum focus on the application of mathematics, the greater linguistic complexity and cultural assumptions associated with such items may bias the assessment against students from lower socioeconomic groups. Nevertheless, such items at least incorporate some of the features of extended problem solving while allowing assessment under controlled conditions.

An interesting consequence of the introduction of PISA is that countries as diverse as South Africa and Israel have adapted the PISA framework as a model for their own mathematics curriculum and assessment for students who are not planning to specialize in mathematics.

Of course, where assessment is less high stakes, more autonomy can more easily be given to teachers to assess students' abilities to solve holistic problems, including working in groups and using computers and the Internet. Such assessment increases validity but at the cost of reducing comparability between students, teachers, and schools.

Formative Assessment

While some countries have relied on national summative assessment to raise mathematical standards, others have attempted to reach the same ends by introducing more formative assessment, sometimes known as assessment for learning. Here, assessment is not being used so much a measure of attainment but to provide feedback to improve the quality of curriculum provision and teaching.

Although the recent interest in formative assessment can be linked to the influential review by Black and Wiliam (1998), the term itself has a long history in education. Building on advances in diagnostic assessment in the 1980s and early 1990s, formative assessment focuses on how assessment can be used by teachers to inform and adapt the teaching of mathematics to suit the learning needs of students. While formative assessment, in general, has been shown to be a particularly effective pedagogic strategy, it is poorly described in ways that can be implemented in practice. Hence, recent research addresses how to use assessment to inform learning in typical classrooms where the time a teacher has to spend with individuals is extremely limited.

Somewhat paradoxically, given that much of the evidence relating to the efficacy of formative assessment has been collected in the context of mathematics, the descriptions of formative assessment in practice are largely generic. Here, we discuss the specific mathematical issues relating to the key generic aspects of formative assessment:

- sharing learning intentions;
- eliciting students' ideas and understandings; and
- regulating learning through feedback.

Sharing learning intentions is remarkably difficult in mathematics. Recent policy initiatives in mathematics education in the UK, for example, have emphasized the importance of sharing lesson objectives with students. However, while the aims behind this are laudable, the tendency is to focus on short-term lesson objectives, whereas the majority of learning objectives in mathematics is medium or long term. Hence, lesson objectives are likely to be either trivial or meaningless to students. An alternative approach is to encourage students to engage in more discussion, yet this implies a considerably more dialogic approach to teaching than is currently the case in most mathematics classrooms.

Teachers have traditionally used classroom tests to elicit students' understandings and one formative approach is to develop better classroom tests. For example, the didactical assessment approach developed at the Freudenthal Institute in the Netherlands has shown how open-ended pencil-and-paper questions can elicit a variety of student thinking. Of course, for such assessment to be formative, it needs to be carried out before the relevant teaching. Another approach has been to examine ways in which teachers can use rich tasks and questioning to find out what students know in the course of normal classroom activity. Asking richer and more challenging questions is not in itself sufficient; responding appropriately to students' answers requires teachers to listen more interpretively in order to work out why students respond in particular ways rather than the more common evaluative practice of listening for correct answers (Davis, 1997).

Feedback can enable the teacher to adapt, or regulate, teaching to the learning needs of students. It can also inform students about their next steps in learning, thus enabling students to self-regulate their learning. A great deal of feedback to students in mathematics is in the form of marking. Typically, this marking identifies correct and incorrect answers and the number of correct answers is given as an overall mark. This is of limited use formatively, because it gives no advice to students on what to do to improve their mathematical understandings. In addition, such marking tends to de-motivate low attainers. Rich comments providing specific feedback to the student can be very effective. Yet, providing comments together with marks is just as ineffective because students focus on the marks not the comments (Butler, 1988). An effective alternative is providing comments without marks.

An important factor in the success of formative assessment has been the communication of this powerful pedagogic idea in simple terms. This has enabled a widespread take-up of the ideas, for example in Australia, New Zealand, Scotland, and England. Much of this take-up has involved a loose, weak, and, at times, procedural implementation of the core ideas. Inevitably, this has attracted a great deal of criticism. While much of this criticism is misplaced, formative assessment does nevertheless place considerable demands on teachers' pedagogic skills and their knowledge of school mathematics and mathematics learning. However, evidence suggests that engaging in professional development relating to formative assessment can improve teachers' pedagogic practice. It is not yet clear the extent to which improvements in knowledge can be achieved in the same manner. Further, the predominance of generic descriptions of formative assessment can lead to an emphasis on generic process skills, over and above mathematical processes and content (Watson, 2006). Another issue relates to misinterpretations by policymakers of formative assessment as

either continual or more personalized assessment and, hence, as simply a justification for increased summative assessment.

Looking Forward

The development of assessment in mathematics, as in other subject disciplines, has been shown to follow an ever-widening spread of purposes for assessment as well as reflecting a broader set of aims for learning mathematics and incorporating an expanding set of assessment techniques in response. In assessment, there is always a tension between validity and reliability/comparability. Consequential validity, in the form of washback on curriculum and pedagogy, is a very significant issue as it potentially affects the whole of a student's learning experiences.

We can expect greater computer power to deliver closer and more continuous monitoring of student progress, greater accountability of all partners in education, increased emphasis on formative assessment, and a closer connection between formative and summative assessment, where summative assessment is mainly to verify the ongoing record of student progress. All these depend on the existence of a sound and proven framework for mathematical progress with a bank of matching assessment tasks generated within and outside the school, and with a knowledgeable and well-trained teaching workforce – these are exacting requirements. However, the risks associated with high-stakes summative assessment, which is simplistic in form and purpose and narrow in scope, are likely to remain a constant political threat to a broad and effective mathematical education, internationally, nationally, and locally.

See also: Assessment and the Regulation of Learning; Challenges of Developing and Implementing Formative Assessment Practices in Schools; Classroom Assessment Tasks and Tests; Criterion-Referenced Measurement; Formative Assessment; IEA Studies in Mathematics and Science; Impact of Assessments on Classroom Practice; Mathematics Learning; Technology Supports for Assessment Design.

Bibliography

- Black, P. J. and William, D. (1998). Assessment and classroom learning. *Assessment in Education* 5(1), 7–73.
- Bloom, B. S., Hastings, J. T., and Madaus, G. F. (eds.) (1971). *Handbook of Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology* 58, 1–14.
- Davis, B. (1997). Listening for differences: An evolving conception of mathematics teaching. *Journal for Research in Mathematics Education* 28(3), 355–376.

- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher* **23**(2), 13–23.
- Mullis, I. V. S., Martin, M. O., Foy, P., et al. (2008). *TIMSS 2007 International Mathematics Report*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.
- National Council of Teachers of Mathematics (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Richardson, M., Baird, J. -A., Ridgway, J., et al. (2002). Challenging minds? Students' perceptions of computer-based world class tests of problem solving. *Computers in Human Behaviour* **18**, 633–649.
- Threlfall, J., Pool, P., Homer, M., and Swinnerton, B. (2007). Implicit aspects of pencil and paper mathematics assessment that come to light through the use of the computer. *Educational Studies in Mathematics* **66**, 335–348.
- Van Den Heuvel-Panhuizen, M. and Becker, J. P. (2003). Towards a didactical model for assessment design in mathematics education. In Bishop, A. J., Clements, K., Keitel, C., Kilpatrick, J., and Leung, F. K. S. (eds.) *Second International Handbook of Mathematics Education*, pp 689–716. Dordrecht: Kluwer.
- Watson, A. (2006). Some difficulties in informal assessment in mathematics. *Assessment in Education* **13**(3), 289–303.
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., and Williams, C. C. (2008). Gender similarities characterize math performance. *Science* **321**(5888), 494–495.
- Niss, M. (ed.) (1993). *Cases of Assessment in Mathematics Education and Investigations into Assessment in Mathematics Education (New ICMI study Series 1 and 2)*. Dordrecht: Kluwer Academic Publishers.
- Office for Standards in Education (2008). *Mathematics: Understanding the Score*. London: OFSTED.
- Pellegrino, J. W., Chudowsky, N., and Glaser, N. (eds.) (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academies Press.
- Shepard, L. A. (2008). Commentary on the national mathematics advisory panel recommendations on assessment. *Educational Researcher* **37**(9), 602–609.
- Van Den Heuvel-Panhuizen, M. (1996). *Assessment and Realistic Mathematics Education*. Utrecht: CD-β Press/Freudenthal Institute.
- William, D. (2007). Keeping learning on track: Formative assessment and the regulation of learning. In Lester, F. K. J. (ed.) *Second Handbook of Mathematics Teaching and Learning*, pp 1053–1098. Greenwich, CT: Information Age.
- Wilson, L. D. (2007). High stakes testing in mathematics. In Lester, F. K. J. (ed.) *Second Handbook of Mathematics Teaching and Learning*, pp 1099–1110. Greenwich, CT: Information Age.

Further Reading

- Black, P. J., Harrison, C., Lee, C., Marshall, B., and William, D. (2003). *Assessment for Learning: Putting it into Practice*. Buckingham: Open University Press.
- Brown, M. (ed.) (1992). *Graded Assessment in Mathematics (GAIM)*. Walton-on-Thames: Nelson.
- Cooper, B. and Dunne, M. (2000). *Assessing Children's Mathematical Knowledge: Social Class, Sex and Problem-Solving*. Buckingham: Open University Press.
- de Lange, J. (2007). Large-scale assessment and mathematics education. In Lester, F. K. J. (ed.) *Second Handbook of Mathematics Teaching and Learning*, pp 1111–1142. Greenwich, CT: Information Age.
- Hodgen, J. and William, D. (2006). *Mathematics Inside the Black Box*. London: NFER-Nelson.

Relevant Websites

- <http://www.fi.uu.nl> – Freudenthal Institute.
- <http://www.kcl.ac.uk> – King's College London Assessment Research Group.
- <http://www.nottingham.ac.uk> – Mathematics Assessment Resource Service (MARS).
- <http://www.ncetm.org.uk> – National Centre for Excellence in Teaching Mathematics (Graded Assessment in Mathematics materials).
- <http://www.nctm.org> – National Council of Teachers of Mathematics.
- <http://www.pisa.oecd.org> – OECD Programme for International Student Assessment (PISA).
- <http://timss.bc.edu> – Trends in International Mathematics and Science Study (TIMSS).
- <http://www.toolkitforchange.org> – Toolkit: Resources for Leaders in Mathematics Education.

Assessment in Schools – Oracy

A L Bailey, University of California, Los Angeles, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Discourse – The organization of language beyond the sentence/utterance level (e.g., conversation, narration, exposition).

Grammar – The rules/conventions governing word order and inflections.

Lexicon/Vocabulary – Receptive and expressive world knowledge.

Oracy – Oral language skills (i.e., listening and speaking) and proficiency.

Phonology – The sound system of a language.

Pragmatics – The socially appropriate uses of language.

Oral language, or oracy, and its assessment for educational purposes have often been overlooked by individual educators and school systems. The teaching and testing of student knowledge of print-based language skills are tantamount to much, if not all, the focus of language-arts skills that native-speaking students experience during their formal education. For example, review of publicly available documents published by the 50 US states and the District of Columbia, reveals that 41 of their education agencies do not include an oral-language component in their state-wide tests of English language arts (ELA), despite the fact that 48 of the agencies include oral-language skills in their mandated language-arts-content standards. (At the time of writing, six US state education agencies have an oral-language component to their annual assessment of students; one state only assesses children at the kindergarten through second grades; and documentation for three states is not clear on their practices. One state has standards for oral-language skills during the early grades only; one state has standards for oral-language skills from ninth grade and above only; and one state has no reported standards for oral language.) Assessment of ELA for accountability purposes is required under federal law (No Child Left Behind Act, 2001) in order to annually evaluate all students in second through eighth grades (7–14-year-olds).

The first section of this article provides the rationale for assessment of oracy in school. Next, construct definitions of oral language that are commonly given in the literature are discussed and major skill areas are described. The article then continues with key aspects of oral-language assessment, including testing for different

purposes, test development, and administration considerations, and the assessment needs of students who have language-learning difficulties or are second-language learners. It concludes by identifying several areas for future research and development.

Motivating the Need for Oral-Language Assessment in School

The demands placed on students by advancements in media technology as they become young adults in our society have increasingly called for a level of spoken language abilities that puts oracy sophistication on par with advanced literacy skills. True, the invention of text messaging, e-mail, online communities, and the use of blogging for personal and business purposes have all led to a heightened need for reading and writing abilities. However, also witness the use of digital video and audio for sharing ideas orally via personal blogs, e-journals, and YouTube, and even via more formal conduits such as business-related podcasts and academic webinars. Moreover, as information technology increasingly makes modern societies dependent on the service industries that require face-to-face or telephonic interactions rather than on the nonverbal skills often used in manufacturing, oral language emerges as critical for a competitive edge in the labor force of the twenty-first century.

Beyond these and other functional aspects of oral-language proficiency related to student needs once they have completed formal education, there are other important reasons for monitoring the progress of children's oral-language abilities in school. Assessment of oral language is critical for knowing how much of instruction across the entire academic curriculum a student is able to comprehend and make use of. Listening comprehension is a prerequisite for any learning that takes place through classroom interaction – from discussions with peers to listening to explanations and directions from the teacher. Subsequent display of this knowledge through students' own verbal responses and contributions to classroom discourse can alert a teacher to the level of learning, be that in mathematics, science, social studies, or the language arts themselves (Cazden, 2001; McKay, 2006).

Knowing how well a child is acquiring oral-language skills is important information for teachers as children learn to read and write. Oral-language skills are closely tied to children's development of literacy skills in terms of

the rate and ease by which they acquire reading and writing abilities, as well as their ultimate attainment in these domains. Certain forms of oral language, for example, narrative discourse which requires the child to decontextualize information for the listener, may even serve as early rehearsal for print-based skills encountered later on. Thus, the information yielded by assessment of these earlier acquired oral-language skills can be useful to educators even before students have documented signs of reading difficulties.

Knowing how well children are able to listen and speak the language that is spoken around them is particularly important with diverse populations of school children; specifically, those who speak a nonstandard variety of the language, those who have a history of language delay or impairment, and those who are acquiring the language of school as a second or additional language alongside their first language. When the language spoken at school is still being acquired or is a source of difficulty for a student, the student's opportunity to learn from instruction and, consequently, the validity of all content-area assessments, are called into question (e.g., Kopriva, 2000).

So why is there neglect of oral language in schools, especially for native speakers of a language and, most pertinently for this article, its assessment? Perhaps the most salient reason is that the assessment of oral-language skills is made challenging by both the amount of time and specialized arrangements it often necessitates. Unlike the assessment of literacy skills, oral-language assessment largely requires one-on-one administration. For administration of standardized assessments, individually testing students in a suitable quiet place takes time. Often administration of this kind requires test administrators to have training in order to obtain the required homogeneity in testing conditions (e.g., uniformity in prompts, types of allowable feedback or responses by the tester, etc., so that the test remains fair and its interpretation valid for a student). Understandably, administration of this type is costly. Nor is spoken language readily measured if scored in real time without audio recording for later verification. Oral language is transitory and must be processed immediately by the test-taker, unlike printed language that students can read and write and possibly reread and rewrite at their own pace. Moreover, oral-language assessment of both listening and speaking skills has to contend with performance characteristics (e.g., anxiety of responding aloud) not just the test-taker's accuracy on the language skill(s) being measured. As a result of these restrictions placed on traditional assessment formats, by its very nature oral language may have been ignored in large-scale, standardized testing.

Classroom-based assessments (both commercial and teacher created) offer more flexibility than large-scale assessments. For example, with small-group administration of speaking and listening assessments, a teacher can

take note of each student's ability to comprehend or respond. However, even these assessments may miss the communicative abilities of students interacting in authentic ways in the classroom rather than interacting in a testing context. Alternatively, formative assessment can allow a teacher to gauge and monitor each student's oral-language abilities within the classroom context, for example, through observation of spontaneously occurring student dialog. This type of classroom-based assessment is only just becoming part of most teachers' assessment repertoires in the US (Bailey and Heritage, 2008; Butler and Stevens, 1997; Wiliam and Thompson, 2007), but has already become a major focus in the UK. England, for instance, has adopted Assessment for Learning, an initiative that has recognized the importance of assessing speaking and listening skills and is replacing the national curriculum tests at the secondary level with a focus on individualized evidence of pupil progress through teacher observations of group work, class discussion, informal conversation, and role play. The Department for Children, Schools, and Families (DCSF, 2009) has developed standards and descriptors of oral-language progression, using exemplars in an attempt to establish a degree of consistency in Assessment for Learning.

Separate from the challenges of administering oral-language assessments in schools, a second reason oral language has likely suffered neglect in both instruction and assessment is because oral-language development among native speakers is traditionally thought to be complete by the time students arrive for formal schooling at about age 5. Yet, there are many important aspects of oral language that are only acquired once children start school. For example, the specialized knowledge of how to use language (pragmatics) for different academic tasks, and the ability to reflect on language as a object of study (metalinguistic awareness) are just two aspects of oral language that are expected to develop throughout the school years and consequently require monitoring and evaluation.

Defining the Oral-Language Construct for Assessment Purposes

There are many facets of language that can make up the oral-language construct that is purported to be measured by an assessment. **Table 1** catalogs the different characteristics that can be identified as making up a definition of oral language, including oral-language registers, oral-language domains, and oral-language skills. (The domains of language include listening, speaking, reading, and writing and are more commonly referred to as modalities by applied linguists.) **Table 1** also lists the major purposes of oral-language assessment, and provides examples of different types of assessment.

Table 1 Key characteristics of oral-language assessments in schools.

Contexts of language use	Oral-language domains	Oral-language skills	Assessment purposes	Assessment examples
Social	Listening (receptive)	Phonological	Instruction: a. Screening b. Formative c. Portfolio d. Benchmark	Language sampling a. Story (re)telling b. Personal narrative c. Picture description d. Dialogue
Academic: a. General b. Specialized (technical)	Speaking (expressive)	Lexical		
	Listening and speaking combined	Grammatical Discourse/pragmatics	Diagnosis Accountability: a. Summative b. Benchmark	Scoring Rubrics: SOLOM (observation) Rating scales Large-scale: a. Norm-referenced b. Criterion-referenced (standards-based)

Oral-Language Registers

There are many registers of oral language, such as casual speech to friends, technical language used by scientists to colleagues, and mothers' simplification of language as they talk to their young children. The types of language most pertinent for the school context are social language used in school and academic language. The social language that students encounter in school can be considered the everyday language of casual conversation between school friends during break time, lunch, or in the home-room, and between teachers and students when they are not engaged in formal instruction. For some students, this register can be predominant in their oral-language use and exposure even during formal instruction; not all teachers teach using a formal academic register all the time, nor might they always hold students accountable for academic language use.

Academic language itself has been characterized in different ways, with some questions of whether academic language is necessary for learning in school, whether it is limited to literate uses of language only, and even if it is meaningful to postulate such an overarching construct for the language of school given that different content areas may have very different linguistic characteristics. Bailey and Heritage (2008) further divide academic language into two subtypes: the language used for navigating a school environment that may still be more formal than social language (e.g., requesting and understanding teacher directions), and the language needed for teaching and learning the content of curricula. Within the latter, researchers have identified aspects of language that may make it challenging, for example, complex grammatical structures such as nominalizations and passive voice, language specialized to specific content areas, such as the technical vocabulary used in subfields of science, as well as the language that cuts across many different disciplines

that can be thought of as general academic language (e.g., Bailey and Butler, 2007). Specialized vocabulary can include words and phrases such as *circumference*, *base*, and *absolute deviation* used during algebra, for example, and general academic vocabulary can include words and phrases such as *evaluate*, *refer*, and *explanation* that tend to have the same meanings across content areas.

Domains of Oral Language

Oracy encompasses two language domains – listening and speaking. Listening is the receptive domain, while speaking is the expressive domain. Together they are needed to carry out human interaction in the here and now. Listening requires the comprehension of spoken language, whereas speaking requires the production of spoken language. In terms of assessment, educators need to test listening and speaking skills in isolation or they may fail to distinguish between abilities in the two domains. However, sometimes educators will want to test combined listening and speaking abilities because it is often at the intersection of the two that instruction occurs (McKay, 2006). For example, students first listen to a textbook passage read aloud by their teacher then answer questions orally to communicate their comprehension.

Oral-Language Skills

Regardless of which register and domain of language is being assessed, it is important to be clear what skill or skills of language are the target so that the assessment can be an effective measure of the skill(s). The following are brief descriptions of the different oral language-skill areas. While all skills are necessary for oral-language development, not all are readily tested (e.g., authentic conversation skills).

Phonological skills

Knowledge and facility with the sound system of a language are fundamental to oral-language development and proficiency. For younger students in school, this means the ability to distinguish units of sound at different levels of organization (e.g., phonemes at the single-sound level, syllables combining vowels and consonants). Later, students will need to use prosodic features (e.g., word stress, sentence intonation) to speak fluently and comprehend accurately (e.g., stress distinguishing nouns from verbs sharing the same form, such as ‘pro’duce (noun) and pro’duce’ (verb)).

Lexical skills

Word recognition, a receptive skill, and word use, an expressive skill, are key components of oral-language development and proficiency. A student’s lexicon, or store of known words can be measured in terms of its breadth and depth. Breadth of word knowledge is the number of different words known, whereas depth includes semantic connections between words. Lexical skills also include student knowledge of derivational morphology (e.g., the variety of affixes that can be added to known words to create additional words with different parts of speech, such as chemic’al’ (adjective) and chemist’ry’ (noun)).

Grammatical skills

Oral abilities include knowing the rules that govern the order of words in a sentence. In English, this syntactic knowledge includes the abilities to comprehend and use: (1) simple sentences (e.g., simple declaratives and imperatives), (2) complex utterances that contain embedded clauses (e.g., relative clauses), (3) compound sentences, and (4) certain phrase structures, such as nominalizations, long noun phrases, and long or multiple prepositional phrases.

Discourse and pragmatic skills

The skills necessary for the organization of language beyond the level of the single sentence include the knowledge of different genres of discourse, namely conversational, expository, and narrative. Each is characterized by its own set of rules for effective communication, with conversation requiring abilities in face-to-face interaction (e.g., taking account of audience knowledge of a topic), expository used for conveying factual information about people, objects, and processes, and narrative used for relaying sequenced past events, both personal and fictional. Discourse-level skills also include the repertoire of language functions or the communicative purposes for which we use language. These include a large range of functions but in the school setting will most likely require students to comprehend and produce explanations, descriptions, and comparisons, for example, and be able to integrate the vocabulary and syntactic features that predictably

attend these functions (e.g., comparative adjectives such as bigg’er’ and great’er’ in service of producing comparisons). Pragmatic rules, the rules that govern how we use language, must also be acquired; knowledge in this area is critical for knowing when it is culturally appropriate to use a particular register. These rules dictate use of vocabulary such as *locate* and *indicate* during geography instruction, whereas their everyday equivalents *find* and *show* are expected in noninstructional conversation with peers and teachers.

Oral-Language Assessment: Development and Use**Operationalization of Oral-Language Constructs**

All assessment users – not just those assessing oral-language abilities – face the issue of interpreting the meaning of test scores. In the case of oral-language assessment, test developers need to make a convincing argument that they are testing listening skills, for example, and not a student’s ability to write a response if a written-response format is being used for a comprehension test. To guard against such eventualities, a well-articulated construct definition at the beginning of all test specifications (i.e., the guiding document for ensuing test item-writing and construction of a finalized test) is critical in order to make the argument that a test is a valid test of oral language. This holds whether creating a large-scale commercial assessment of oral language or a teacher record of student vocabulary use during math problem solving.

Purposes and Uses of Oral-Language Assessment

Oral language is difficult to assess but oral language is at the core of early curriculum (McKay, 2006). Data generated by interpreting the scores obtained on oral-language assessments can be used in many different ways, but the three main purposes of oral-language assessment are to inform instruction, diagnosis, and accountability.

Instruction

Oral-language assessments used for instructional purposes are all designed to provide feedback for the teacher and in many cases for students as well. These can include screening tests, formative assessment, portfolio assessment, and benchmark assessment. Results from these assessments are meant to inform classroom-level decisions and are commonly referred to as classroom assessments of oral language. A screening test may be teacher created or a commercial test that provides sufficient information about a student’s potential oral-language abilities to decide whether certain kinds of instructional experiences will be suitable for a student or if further testing is needed.

Formative assessment is not one specific assessment but rather an approach to the continual monitoring of student progress so next-steps instructional decisions can be made. Formative assessment tools are commercially available, but formative assessment, as already mentioned, can be the observation of student oral interactions during class time with a teacher noticing which students will need further help with question formation or conditional tenses, for example. Providing students with information on desired learning goals and feedback about what they have observed are key components of teachers assessing formatively. Learning goals can be systematically reflected in a learning progression to which a teacher can target instructional activities and articulate success criteria for students meeting the goals (see, e.g., the Speaking and Listening Progression Chart for English 11–16, of the Secondary National Strategy, DCSF, 2009). Formative assessment can also focus on student self-assessment with students taught how to monitor their own learning, for instance, asking themselves key questions about the sequence of events in a story to be sure of comprehending the main idea.

Portfolio assessment is an accumulation of evidence of student learning taken from a variety of student work produced in different contexts. This evidence could take the form of oral reports from different content areas, such as a science report, a review of a novel, or a history-project presentation. Finally, benchmark assessments are typically created by external test developers and may be tied to a specific curriculum. They provide teachers with periodic information about how well a student is acquiring key content of a curriculum.

Diagnosis

The primary purpose of a diagnostic assessment of oral language is to provide teachers and education specialists with very specific information about a student's language abilities. Diagnostic assessments include sufficient items of a given skill area to provide detailed information about the skill so that a teacher or specialist can intervene as necessary. For example, assessments of language production in the areas of vocabulary and pragmatics are used to help determine an expressive language delay. In addition, diagnostic assessments often provide norming information usually by age and grade so that an individual student's performance can be compared with and understood in relation to the performance of typically developing students in a norming sample.

Accountability

Assessment can be used to hold educators accountable for student achievement. Often, summative assessments that come at the end of a school year or course of study (as well as benchmark assessments that come after a unit or a period

of study) are used to monitor student progress for the purposes of evaluating student performance and evaluating program efficacy (and implicitly, teacher and school performances). These assessments tend to be norm referenced or standards based (i.e., a form of criterion-based assessment discussed further below). The information these assessments yield can be used for reporting to governing or funding bodies (e.g., local and state education agencies, or central governments) to fulfill obligations or learning targets such as raised test scores or increased numbers of students reaching certain standards for oral language.

Approaches to Oral-Language Test Development and Administration

In terms of the content of an oral-language assessment, the test items should match the content of a curriculum, or match a set of standards for skills previously identified and valued by a set of stakeholders (e.g., national teacher organizations or state education agencies). The array of test items should ideally be tied to a theory of language use or development. For example, if a communicative theory of language is adopted, then tasks should capture the authentic language demands of classroom interactions between teachers and students and between students. Assessments that will be used to measure growth or annual gains in language development must take account of theories of acquisition (e.g., attention to the order and weight given to knowledge of complex/late-acquired grammatical structures).

Establishing test validity with oral-language assessments

Establishing the validity of an oral-language assessment involves constructing an argument that the test measures the language constructs it claims to measure. While this involves the explicit articulation of the construct and the content of the test in the test specifications, there are also qualitative and quantitative analyses that play a role. For example, small-scale pilot testing of item types and administration procedures using think-aloud protocols allows the test developer to examine how students respond to item difficulty, formatting, and administration at the very moment test-takers are answering test questions. Field testing with larger numbers of students allows the test-developer to use psychometric techniques to supplement the validity argument with statistical analyses of item difficulty, item discrimination (e.g., between known groups of low- and high-performing speakers), reliability, and differential item functioning for different groups of test-takers to investigate, for instance, potential gender or cultural biases. However, in building a convincing validity argument, these analyses cannot substitute for test specifications that clearly articulate the construct to be measured and the process of item content selection.

Threats to test validity

Perhaps in no other scholastic testing environment does face-to-face interaction play so critical a role than in the assessment of oral-language skills. The assessment of oral-language skills requires the student to interact with the tester, and different testers will have different styles of interaction and degrees of support to elicit responses from students. The heterogeneity in interaction between testers and test-takers during administration is a threat to the validity of oral-language assessments because it violates the assumption that each test administration is essentially carried out in the same manner. There are also age-related considerations that must be taken into account when developing, administering, and interpreting oral-language assessments. Performance can be impacted by the cognitive and social developments in young students that are construct-irrelevant for language skills. It is also important to keep demands on the memory of young children to a minimum, else memory load also inadvertently becomes part of what is measured by an oral-language assessment.

Standards setting for oral language

As mentioned, large-scale summative assessments are often used for accountability purposes and can provide a scale score and a percentile ranking of students if norm referenced (i.e., evaluation of a student's performance relative to other students taking the assessment). However, to be most useful to educators, assessments should also provide levels or bands that are accompanied by descriptors of what skills students at a particular level can be expected to perform. Thus large-scale assessments can be both norm referenced and criterion based. These assessments undergo standards setting in order to determine what levels or cut scores on the test are considered good, fair, or poor performances. Standards setting is typically conducted by a panel of educators familiar with the language expectations of the different grade levels and the skills associated with different levels of oral-language proficiency. The issue of which standard or standards should be used to gauge a student's fluency is a continuing challenge and must be established anew for each testing situation and student population. In recent years, there has been a shift to standards-based assessment in schools worldwide. However, while these tests attempt to be overtly aligned to a set of desired standards, often the standards themselves have not been validated (e.g., shown empirically to be skills necessary for functioning as a proficient speaker). Even the most proficient speakers are known to make dysfluencies and errors at times (Davidson, 1994). Moreover, there are few norms for the oral-language development of monolingual school-age children (Nippold, 1995) that test developers can turn to for guidance.

From a sociocultural perspective, the standards-setting procedure faces the issue of privileging certain oral-language characteristics over others. For example, correct pronunciation is a cultural artifact. Setting the target

pronunciation consistent with a particular regional or even national variety of a language is primarily a function of sociolinguistic preferences and even prejudices and not any inherent correctness of spoken English. (For example, Received Pronunciation in the UK and the Midwestern dialect in the US are used as the standard variety of the language in native speaker and English as a second language (ESL) contexts, and American-accented English may be chosen instead of British-accented English in the English as a foreign language (EFL) context.) Standards setting should at the very least be guided by which forms of the language will make the language learner intelligible to most speakers of the language, as well as take account of what level(s) of functioning society desires of the language learner.

Box 1 provides examples of classroom-based assessments by language domain that reflect points made in the prior discussion. Note that depending on the purpose of use, with little modification, these assessments can be

Box 1 Examples of classroom-based assessments by domain of oral language

Listening-only assessments

Novel natural speech input

Tying the topic of the stimuli to other content areas can maximize information gained by the teacher (both language and concept knowledge).

Responses can be nonverbal: drawing, picture matching, diagramming, model building, total physical response (head nod/shake; or pointing to objects) to show comprehension.

Grammaticality judgment task

Response can be simple check marks or true/false circling to indicate whether a spoken utterance is a legal sentence in the language.

Speaking-only assessments

Production of oral reports, story retelling, personal narratives, or picture descriptions

Again, the stimuli can be made maximally useful by choosing topics from other content areas.

Production will overtly require discourse skills (e.g., genre knowledge, organization of language); however, other language skills will be revealed (e.g., pronunciation, lexical knowledge and diversity, and grammatical accuracy).

SOLOM (student oral-language observation matrix)

Production is spontaneous speech during classroom activities scored using a multilevel rubric for a variety of oral-language skills.

Listening and speaking assessments combined

Grammatical imitation task

Response is the direct imitation of the stimulus utterances. Students typically cannot accurately imitate what they do not already have command over.

Authentic interaction

These include question-and-answer sessions with the teacher, interviewing peers, participating in plays, and partner/group discussions and debates.

Referential communication tasks

These can take the form of a barrier task, for example, describing a route or an object separated from a naive partner by a screen.

used formatively for guiding ongoing instruction, diagnostically to drill down to a specific oral-language challenge for a student, or summatively to judge student attainment after a period of instruction.

The Oral-Language Assessment of Diverse Learners

Special education and the education of linguistic-minority students are two areas that have tended to be exceptions to the neglect of oral-language assessment. Many oral-language assessments administered with monolingual students in school are used for diagnosing language delay or impairment. Speech-language pathologists working in schools have at their disposal norm-referenced assessments that can help diagnose whether a student has an issue with speech production or comprehension, including pronunciation difficulties, expressive language delays, specific language impairment that often affects grammatical abilities, and receptive language difficulties that suggest a language-comprehension disorder.

Due to the No Child Left Behind Act (2001), the US has seen much activity in language test development in the last decade. This activity is directed at new large-scale language assessments that capture the language demands that English-language learners encounter in school and will need in order to succeed academically. Performance on previously available assessments did not reflect whether a student was ready to perform in an academic setting because they tended to primarily measure everyday or social language. Educators now recognize the importance of academic language to academic-concept knowledge and the need to track its development in second-language learners (Bailey and Butler, 2007; Gottlieb, 2006).

Assessment of the language skills of second-language learners is recognized as highly complex and prone to biases that result from a lack of awareness about cultural practices for language usage. For example, students may have been exposed to a tradition of learning rich descriptors for objects, people, and concepts rather than the names of objects that are the typical content of vocabulary assessments used with school-age children (Peña *et al.*, 2001). Assessing second-language learners in culturally sensitive ways is critical for determining whether a language impairment or developmental disorder (e.g., autism) is at issue or whether a student is still in the process of learning his or her second language (Rollins *et al.*, 2000). The same holds for students who speak a nonstandard variety of the language such as African-American English (Craig and Washington, 2002).

The Future of Oral-Language Assessment

Technology initiatives and research of the oral-language construct are among the necessary future developments

in this area. Two areas in which technology can assist are (1) enabling test developers to overcome the challenges of individually administered assessment. For example, automated speech recognition (ASR) technology can record and process student responses so that they can be assessed by computer reducing the administration burden on teachers and standardizing procedures (e.g., Alwan *et al.*, 2007). This and other technology initiatives (e.g., telephonically administered assessment) will need further study to determine whether these approaches yield the same results and interpretations as traditional face-to-face assessments, and (2) innovations such as web-delivered videoed lessons, that can be used in teacher professional development for formative assessment; an area with acknowledged need for greater teacher learning and system-wide initiatives to increase the consistency with which formative-assessment strategies are applied (e.g., DCSF, 2009). (The Center for Applied Linguistics is also active in this area.)

Neglect of oral assessment in schools is mirrored by a lack of research of what language demands all students need to master in order to succeed academically. Future research can redress this dearth of information by focusing on at least three identifiable areas of research and development: (1) expanded study of first and second languages as they are acquired by school-age children, (2) creation and study of professional-development strategies that enable teachers to recognize discipline-specific oral language that is prerequisite for successful academic performance in mathematics, science, etc., and to integrate the evaluation of these language demands with their assessment of content-area knowledge, and (3) creation and validation of new standards and learning progressions for oral-language development to which new assessments (both formative and summative) can be aligned. Research of the oral-language construct in the different ways suggested will not only further our ability to develop fair, valid, and useful assessments but will also expand our fundamental understanding of how oral language is acquired and operates in school, thus fulfilling both the applied and the basic purposes of the research endeavor (e.g., Stokes, 1997).

See also: English Language Learners with Special Needs; Formative Assessment; Language Development in Children with Special Needs; Portfolio Assessment; Second Language Assessment; Test Development; Validity.

Bibliography

- Alwan, A., Bai, Y., Black, M., *et al.* (2007). A system for technology based assessment of language and literacy in young children: the role of multiple information source. In *Proceedings of IEEE International Workshop on Multimedia Signal Processing*, Greece, 26–30. http://diana.icsl.ucla.edu/Tball/publications/tball_mmsp07.pdf.
- Bailey, A. L. and Butler, F. A. (2007). A conceptual framework of academic English language for broad application to education. In

- Bailey, A. L. (ed.) *The Language Demands of School: Putting Academic English to the Test*, pp 68–102. New Haven, CT: Yale University Press.
- Bailey, A. L. and Heritage, M. (2008). *Formative Assessment for Literacy, Grades K-6: Building Reading and Academic Language Skills across the Curriculum*. Thousand Oaks, CA: Corwin/Sage.
- Butler, F. A. and Stevens, R. (1997). Oral language assessment in the classroom. *Theory into Practice* **36**(4), 214–219.
- Cazden, C. (2001). *Classroom Discourse: The Language of Teaching and Learning*, 2nd edn. Portsmouth, NH: Heinemann.
- Craig, H. K. and Washington, J. A. (2002). Oral language expectations for African American preschoolers and kindergartners. *American Journal of Speech-Language Pathology* **11**, 59–70.
- Davidson, F. (1994). Norms appropriacy of achievement tests: Spanish-speaking children and English children's norms. *Language Testing* **11**(1), 83–95.
- DCSF (Department for Children, Schools and Families) (2009). *The National Strategies: Assessment and Progression in Speaking and Listening*. http://nationalstrategies.standards.dcsf.gov.uk/node/2116?uc+force_uj (accessed August 2009).
- Gottlieb, M. (2006). *Assessing English Language Learners: Bridges from Language Proficiency to Academic Achievement*. Thousand Oaks, CA: Corwin.
- Kopriva, R. (2000). *Ensuring Accuracy in Testing for English Language Learners*. Washington, DC: Council of Chief State School Officers.
- McKay, P. (2006). *Assessing Young Language Learners*. Cambridge, UK: Cambridge University Press.
- Nippold, M. A. (1995). Language norms in school-age children and adolescents: An introduction. *Language, Speech, and Hearing Services in Schools* **26**, 307–308.
- No Child Left Behind Act (2001). Pub. L. No. 107–110, 115 Stat. 1425.
- Peña, E. D., Iglesia, A., and Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology* **10**, 138–154.
- Rollins, P. R., McCabe, A., and Bliss, L. (2000). Culturally sensitive assessment of narrative skills in children. *Seminars in Speech and Language* **21**(3), 223–234.
- Stokes, D. E. (1997). *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington, DC: Brookings Institution.
- William, D. and Thompson, M. (2007). Integrating assessment with learning: What will it take to make it work? In Dwyer, C. A. (ed.) *The Future of Assessment: Shaping, Teaching, and Learning*, pp 53–84. Mahwah, NJ: Erlbaum.
- Language and Education, *Language Testing and Assessment*, pp. 379–398. Berlin: Springer.
- Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.
- Davidson, F. and Lynch, B. K. (2002). *Testcraft: A Teacher's Guide to Writing and Using Language Test Specifications*. New Haven, CT: Yale Press.
- Garcia, G., McKoon, G., and August, D. (2006). Language and literacy assessment of language-minority students. In August, D. and Shanahan, T. (eds.) *Developing Literacy in Second-Language Learners*, pp 597–624. Mahwah, NJ: Erlbaum.
- Lazaraton, A. (2002). *A Qualitative Approach to the Validation of Oral Language Tests*. Cambridge: Cambridge University Press.
- Lumley, T. and Brown, A. (2004). Research methods in language testing. In Hinkel, E. (ed.) *Handbook of Research in Language Teaching*, pp 833–856. Mahwah, NJ: Erlbaum.
- Luoma, S. (2004). *Assessing Speaking*. Cambridge: Cambridge University Press.
- McKay, P. (2005). Research into the assessment of school-age language learners. *Annual Review of Applied Linguistics* **25**, 243–263.
- Nelson, N. W. (1989). Curriculum-based language assessment and intervention. *Language, Speech, and Hearing Services in Schools* **20**, 170–184.
- Oliver, R., Haig, Y., and Rochecouste, J. (2003). *Oral Language Assessment and the Communicative Competence of Adolescent Students*. Mount Lawley: Centre for Applied Language and Literacy Research, Edith Cowan University.
- Peña, E. D. and Quinn, R. (1997). Effects on the test performance of Puerto Rican and African American children. *Language, Speech, and Hearing Services in Schools* **28**(4), 323–332.
- Pray, L. (2005). How well do commonly used language instruments measure English oral-language proficiency? *Bilingual Research Journal* **29**(2), 387–409.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing* **18**, 429–462.

Further Reading

- Bailey, A. L. (2007). *The Language Demands of School: Putting Academic English to the Test*. New Haven, CT: Yale University Press.
- Bailey, A. L. (2008). Assessing the language of young learners. In Shohamy, E. and Hornberger, N. H. (eds.) *Encyclopedia of*
- Language and Education, Language Testing and Assessment*, pp. 379–398. Berlin: Springer.
- Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.
- Davidson, F. and Lynch, B. K. (2002). *Testcraft: A Teacher's Guide to Writing and Using Language Test Specifications*. New Haven, CT: Yale Press.
- Garcia, G., McKoon, G., and August, D. (2006). Language and literacy assessment of language-minority students. In August, D. and Shanahan, T. (eds.) *Developing Literacy in Second-Language Learners*, pp 597–624. Mahwah, NJ: Erlbaum.
- Lazaraton, A. (2002). *A Qualitative Approach to the Validation of Oral Language Tests*. Cambridge: Cambridge University Press.
- Lumley, T. and Brown, A. (2004). Research methods in language testing. In Hinkel, E. (ed.) *Handbook of Research in Language Teaching*, pp 833–856. Mahwah, NJ: Erlbaum.
- Luoma, S. (2004). *Assessing Speaking*. Cambridge: Cambridge University Press.
- McKay, P. (2005). Research into the assessment of school-age language learners. *Annual Review of Applied Linguistics* **25**, 243–263.
- Nelson, N. W. (1989). Curriculum-based language assessment and intervention. *Language, Speech, and Hearing Services in Schools* **20**, 170–184.
- Oliver, R., Haig, Y., and Rochecouste, J. (2003). *Oral Language Assessment and the Communicative Competence of Adolescent Students*. Mount Lawley: Centre for Applied Language and Literacy Research, Edith Cowan University.
- Peña, E. D. and Quinn, R. (1997). Effects on the test performance of Puerto Rican and African American children. *Language, Speech, and Hearing Services in Schools* **28**(4), 323–332.
- Pray, L. (2005). How well do commonly used language instruments measure English oral-language proficiency? *Bilingual Research Journal* **29**(2), 387–409.
- Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing* **18**, 429–462.

Relevant Websites

- <http://ericae.net> – Eric Clearinghouse on Assessment and Evaluation.
- <http://www.mlpp-msl.net> – Michigan Literacy Progress Profile.
- <http://www.cresst.org> – National Center for Research on Evaluation, Standards and Student Testing (CRESST).
- <http://www.oafccd.com> – Ontario Association for Families of Children with Communication Disorders.
- <http://www.cs.cmu.edu> – Project LISTEN.
- <http://nationalstrategies.standards.dcsf.gov.uk> – The National Strategies: The U.K. Department for Children, Schools and Families.

Assessment in Schools – Primary Science

W Harlen, University of Bristol, Bristol, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Assessment in primary science, as in other subjects, can serve a number of purposes:

- a formative purpose – when assessment is part of, and supports, teaching and learning;
- a summative purpose in relation to individual students – to record and report on progress to parents, other teachers, and the students themselves; and
- a summative purpose in relation to the achievement of groups of students – giving information that can be used in the context of evaluating school effectiveness.

(Primary here refers to students from the age of 5 to 12 or 13. In some countries schools for students of this age are described as elementary and at the upper end of the range students may be in middle or junior high schools.)

The great majority of primary science programs includes some guidance for teachers for summative assessment; very few directly support the formative use of assessment. The summative assessment material may take the form of tests taken prior to and following a module or unit of work and, sometimes, teachers are helped to assess students' written work within the unit. When a program is newly being introduced, this information is used to evaluate its impact as well as to report progress of individual students. However, deeper understanding of the goals of primary science – as now identified – brings with it the realization that to assess the understanding, inquiry skills, and attitudes involved requires a range of information beyond that provided by conventional tests.

This discussion of assessment, therefore, begins with a brief overview of the nature and goals of learning in science at the primary school level. Proceeding further, it considers how the challenges that these present for assessment can be met in relation to the three purposes above.

Learning in Science at the Primary School Level

Primary science programs that reflect modern views of learning share several key features:

- building on the ideas that students bring to the classroom at each stage;

- setting out a progression in small ideas toward a relatively small number of big ideas which students gradually grasp;
- providing activities in which students collect evidence from firsthand investigation of materials and phenomena around them and from other sources; and
- promoting collaborative learning through dialog, discussion, and argumentation among students and between students and the teacher.

These features are summed up in the concept of learning through inquiry – the active participation in the construction and reconstruction of their ideas by the students themselves. The concept of inquiry is not new, being implicit in Dewey's ideas with regard to active learning. Nor is it confined to science education, although it is often defined in that context, for example, by the National Research Council (NRC, 1996: 23). Inquiry is now a widely adopted approach for science education and there are over 30 developed and developing countries where it is in operation in at least some primary schools (Harlen and Allende, 2006). Its focus is both on the processes involved in inquiry and on the understanding of scientific ideas developed through the use of these processes (National Science Foundation (NSF), 1997: 7). Indeed the rationale for inquiry – and the reason for inquiry-based science programs being so strongly supported by scientists as well as by science educators throughout the world – is that it aims to promote scientific literacy.

Through engaging in scientific inquiry, students acquire a general understanding of the important ideas of science, of the nature of scientific investigation, and of how to evaluate and interpret evidence. The development of this scientific literacy has to start in the primary school grades – for the ideas and frameworks that students need to know are big ones that cannot be taught directly. They need to be built from the small ideas relevant to the objects and events familiar to the students. The role of primary school science is to help in building understanding by ensuring that the small ideas are consistent with evidence and are not just the students' preconceived ideas.

Active participation in learning is no longer taken to mean only hands on experiences – as it once was in the days of the first wave of primary science programs in the 1960s and 1970s. The importance of social interaction is seen as having a central role in the development of understanding.

Teachers are urged to encourage students to express their ideas, however unformed, to speculate about the meaning of evidence, to challenge each other's interpretation, and engage in what is described as exploratory talk (Barnes, 1978; Asoko and Scott, 2006; Mercer *et al.*, 2004).

Teaching for understanding, in this way, necessarily takes more time than the transmission of facts that has been the traditional way of teaching science in some countries. It requires a curriculum that is based on a few broad concepts rather than on a large number of narrower ones presented in what is often a fragmented way. Such a curriculum is described by Duschl *et al.* (2006) as being based on "learning progressions." A learning progression is a sequence of understandings about a topic that build on each other as students move from starting to completing their primary education – a span of 6–8 years.

It follows that the learning outcomes of primary science will include cognitive skills relating to obtaining and using evidence, conceptual understanding of science and how it works, and attitudes or dispositions.

Assessment to Help Learning in Primary Science

There is much in common between the activities of teachers and learners when engaged in inquiry and when assessment is being used to help learning (formative assessment). Inquiry-based science education requires teachers – in addition to providing opportunity for students to encounter materials and phenomena to explore and investigate – to do the following:

- Ask open, rich, questions to elicit present understandings and how students are explaining what they find.
- Listen to students' ideas and take them seriously.
- Engage students in suggesting how to test their ideas or answer their questions through investigation or in finding evidence from secondary sources.
- Set up opportunities for collaborative learning and dialogic talk.
- Scaffold alternative ideas that may explain the evidence from their investigation.
- Gather information – through observation, questioning, and interaction – with regard to students' developing skills and ideas.

The practice of formative assessment requires very similar conditions to these. This is not surprising since formative assessment aims to develop learning with understanding and to enable students to take responsibility for identifying what they need to do to achieve the goals of their activities. In science, assessment for learning is closely associated with inquiry (Black *et al.*, 2003). It could almost be said that inquiry requires formative assessment. The additional features of formative assessment – the provision of formative

feedback and the involvement of students in self- and peer-assessment – all support active engagement in learning and encourage students to take ownership of their learning and progress. However, as already noted, few current primary science programs include all the features that would support formative assessment.

Formative assessment is a continuing cyclic process which informs ongoing teaching and helps learners' active engagement in learning. The cycle involves the collection of evidence concerning learning, the interpretation of that evidence in terms of progress toward the goals of the work, the identification of appropriate next steps and decisions with regard to how to take them, so as to close the gap (Sadler, 1989) between what has been achieved and what is aimed for, or to move on. The cycle is repeated and the effects of decisions at one time are assessed at a later time as part of the ongoing process. The value of students' participation at all stages in this cycle is widely recognized (Hein and Price, 1994; Stiggins, 2001; Black *et al.*, 2002). It requires that the students as well as the teachers have a clear idea of the goals of the work. We now consider how the main elements of the formative assessment cycle are practiced in primary science, beginning with the clarification of goals.

Specific and General Goals

Clarity of goals is a pivotal requirement for all assessment; the difficulty that this often presents to primary teachers, in science, accounts for a good deal of poor practice in assessment in this area. Goals in science learning can be expressed with different degrees of specificity. The identification of standards, attainment targets, or other ways of specifying intended learning in national curricula has helped to communicate overall goals. However, these are in very general terms, such as ability to plan and conduct a scientific investigation or understanding the diversity and adaptation of organisms. Such statements are too general to be achieved in a single lesson or even a set of lessons on a particular topic. The goals of a specific lesson might include the understanding of how the structure of particular plants or animals is suited to the places where they are found. This will contribute to a broader goal of understanding how living organisms, in general, are suited to their habitats, but achieving this understanding will depend on looking at a variety of organisms, which will be the subject of other lessons with their own specific goals. Similarly, skills such as planning a scientific investigation are developed not in one lesson, but in different contexts in different lessons and topics.

Gathering Evidence

The specific goals indicate the evidence to be gathered to help learning. Without goals, there can be no way of

deciding what to pay attention to. This is particularly important in primary science because what students do and produce in their regular work is the main sources of evidence for formative assessment and while almost everything a student does can give useful evidence, not knowing what to look for will mean that relevant signs will be missed.

Observation of actions and review of written work may not be sufficient to indicate students' thinking and need to be supplemented by questioning and discussion. Not all questions are helpful in this context. The useful one are expressed as open and person-centered, for example, 'what do you think will happen if ...?' or 'what do you think could be the reasons for ...?' Stiggins *et al.* (1989) found that few teachers ask questions of this kind, focusing instead on asking for recall. Black *et al.* (2002) listed questioning as a particular feature of talk between teacher and students that requires change if formative assessment is to be implemented in classrooms. Further, questions that are carefully worded to provoke thought should be followed by time for thought. Row (1974) found that increasing wait time from the average of less than one second to eight or more seconds was found to increase the quality of the students' responses.

Interpreting Evidence

The evidence collected in these various ways as part of teaching is used in formative assessment to decide where students are in the progression toward achieving the goals and so to identify the next steps to take. This is based on a view of learning as producing a qualitative change both in development of processing skills and conceptual understanding. Skills become more elaborate, more consciously and widely used; small ideas – relating to particular phenomena – are linked together to form big ideas that are widely applicable. Although not enough is known to map out the course of development of some learning outcomes, research findings (SPACE research reports, 1990–1992; Black and Lucas, 1993; Harnqvist and Burgen, 1997) have been used to provide frameworks for describing progression. For instance, Masters and Forster (1996) have produced progress maps in several subjects and Wilson and Sloane (2000) describe progress variables as a key principle of the BEAR assessment system. **Table 1** displays suggested indicators of progression in development of ability to plan and conduct investigations:

For conceptual development, the statements in **Table 2** bring together common threads of development from small to big ideas in generic statements which can be translated into the context of particular activities.

Although the statements in **Tables 2** and **3** are numbered, this is for ease of reference only. There is no claim that there is an equal scale of progression, merely an ordinal sequence. The statements can be related to stages

Table 1 Indicators of progression in planning and conducting investigations

Students:

1. Suggest a useful approach to answering a question or testing a prediction by investigation, even if details are lacking or need further thought
2. Make suggestions about what might happen when certain changes are made
3. Identify the variable that has to be changed and the things which should be kept the same for a fair test
4. Identify what to look for or measure to obtain a result in an investigation
5. Select and use equipment and measuring devices suited to the task in hand
6. Succeed in planning a fair test using the support of a framework of questions or planning board
7. Spontaneously structure a plan so that variables are identified and steps taken to make results as accurate as possible.

Based on Harlen, W. (2006). *Teaching, Learning and Assessing Science 5–12*, 4th edn., p 107. London: Sage.

Table 2 Indicators of progression in ideas

Students:

1. Offer a description only with no attempt to explain a situation.
2. Use a preconceived nonscientific idea to explain a situation or make a prediction.
3. Name a relevant idea (e.g., friction, gravity) without explaining its relationship to the situation.
4. Use a relevant idea to explain a specific situation but not other situations where it also applies.
5. Use a relevant idea to explain several related experiences or situations.
6. Refer to a bigger idea that explains a number of linked phenomena.
7. Use a bigger idea to predict events not yet encountered.

Based on Harlen, W. (2006). *Teaching, Learning and Assessing Science 5–12*, 4th edn., p 148. London: Sage.

Table 3 The walled-garden task: An example of an embedded task

These materials take the form of a small class project about an imaginary walled garden. The students are introduced to various features of the walled garden – water (in a pond with a fountain), walls, wood, mini-beasts (i.e., small creatures), leaves, bark, and a sundial. For each of these seven features a poster is displayed in the classroom, giving suggestions for activities and posing some questions. Students work on one poster at a time, answering the questions in the answer booklet linked to the activities for that feature. They carry out the activities in any sequence and so a whole class can be working at the same time. The questions are designed for written answers, but the element of active exploration in arriving at the answers extends the range of skills tested. The questions focus on inquiry processes, the different content provided by the seven features creating opportunities for the skills to be used in different contexts and so reducing some of the task-sampling variation. Since the activities are well designed, attractively presented, and intriguing, students readily engage with them as if they were normal work.

or levels but, for the purpose of formative assessment, this is not necessary. What matters is to see what progress has been made and to identify what needs attention.

Feedback to Students

The decision with regard to next steps and how to take them is ideally shared by the teacher and students but there are also many occasions when the teacher provides feedback either through marking written work or spoken comment. The manner in which this feedback is given has a critical impact on motivation. The seminal research of Butler (1988) indicated that giving feedback in terms of comments had a positive effect on students' achievements and interest in undertaking further work compared with feedback in the form of grades, or grades plus comments. These results suggest that giving grades or marks diverts attention from any comments, however useful these may be. Good practice in feedback is identified as having a focus on the work, saying what is good about it, as well as how it can be improved, avoiding superficial comments (such as a good try), and reflecting the goals rather than other qualities such as tidiness or spelling, unless these were part of the task.

The Students' Role

Involving students in their own assessment means that they themselves gather the information that they need to make progress. It gives them direct feedback, without waiting to receive it from the teacher. Taking part in assessment of their own and their peer's work also means that students perceive assessment as something in which they have an active part rather than being the object of others' assessment. It helps them to be responsible for taking action and to be committed to the goals that they have participated in deciding. Developing these skills in students is not a simple matter and indicates that teachers have to take action to ensure that students understand the standards of quality that they should apply to it. While directly stating these criteria of quality is advocated sometimes (e.g., Clarke, 2001), it is more effective to involve students in thinking about what makes good work in a particular case. This can be done, for instance, by brainstorming on what makes a good investigation, or a good report, or discussing examples of satisfactory and unsatisfactory reports (Hein and Price, 1994). Students can then be encouraged to refer to the agreed criteria when they are working and particularly when reporting to each other and reflecting on their work.

Peer-assessment is a particularly useful form of interaction among students. As Sadler (1998) points out, students may accept and take criticism more seriously from their peers than from their teacher. They express it in language they would naturally use and so avoid misunderstanding. Arranging for students to talk to each other about their

work encourages them to review and explain it without the pressure that comes from the unequal relationship between learner (novice) and teacher (expert). It helps the development of a learning environment where learners recognize each others' strengths and are prepared to help each other. It also requires such an environment – where there is an ethos of cooperation rather than competition – which is a prerequisite for the practice of several features of formative assessment.

Reporting Individual Student Performance

Summative assessment provides information with regard to what students have achieved at certain times. As in the case of formative assessment, the assessment process involves gathering evidence, interpreting it, and reporting it. In some circumstances, the information is fed back into teaching, but this is not the primary purpose.

Evidence for summative assessment can be gathered in various ways:

- by special tests or tasks designed for students to show what they can do at a particular time;
- by summarizing evidence from regular workup to the time of reporting; and
- by combining evidence from ongoing work and special tasks of tests.

Special Tasks or Tests

Assessment for summative purposes needs to be highly reliable as it may be used for grouping or selection that can affect the future learning opportunities; for that reason, there is an attraction in using special tasks or tests. The items or tasks can be specified and controlled and presented in the same way to all students, and marked using the same rubrics or criteria. Thus, they appear to give students the same opportunities to show what they can do at a particular time and, therefore, are fair. However, giving the same tasks is not the same as giving the same opportunities; students vary in their reaction to – and, therefore, in their performance in – different tasks which appear to make the same demands. The reliability of a test is limited by the facts that only a small number of tests items or tasks can be included and a different selection from all possible ones could lead to a different result. Calculations based on the national tests taken by students in England at the end of the primary school indicate that this effect could result in at least a third of pupils being given the wrong level (William, 2001).

The effect of the selection of items or tasks is particularly serious in science since tasks that attempt to assess inquiry skills, problem solving, and the application of

concepts in real situations are necessarily time consuming and only a small number can be included in a test. Students' response to them is highly dependent on the choice of content. In a study, in the United States, by Pine *et al.* (2006), fifth grade students were assessed using several hands-on performance tasks, including one based on Paper Towels (in which they had to test which of three different kinds of paper towels would hold most water) and one called Spring (about the length of a spring when different weights were hung on it). The researchers found "essentially no correlation for an individual student's scores. Students with either a 9 or a 1 Spring score had Paper Towels scores ranging from 1 to 9" (Pine *et al.*, 2006: 480). Because this is a large task-sampling variation, obtaining a reliable score for an individual student would require the individual to tackle a totally unacceptable number of tasks.

This is an example of the interdependence of validity and reliability – where greater validity tends to mean lower reliability. The corollary is also true. Where greater reliability is paramount, the selection of items for a test gives preferences to those that can be marked unambiguously as correct or incorrect. These are usually ones requiring factual knowledge rather than the application of skills – lowering the overall validity of the test. Low validity is a serious defect of tests since what is tested is inevitably taken as an indication of what is important to teach. The effect is exacerbated by high-stakes use of test results, which not only demand high reliability in the items but put pressure on teachers to use methods that seem to be necessary for passing the tests (Harlen and Deakin Crick, 2003). This was shown clearly by Johnston and McClune (2000) in their work on the effect of the high-stakes tests taken by students at the end of primary school to decide the type secondary school to which they transfer. They found a high proportion of teaching through transmission of information and highly structured activities, with little time for direct investigation by students. When interviewed, the teachers indicated that they felt constrained to teach in this way on account of the nature of the tests.

Summarizing Evidence from Regular Activities

The problems of test validity are avoided by using the evidence that teachers collect during teaching. In theory, this will cover all goals of learning and will mean that evidence can be used both for formative assessment and for summative assessment. However, the important difference between these purposes has to be borne in mind – that, for summative assessment, the evidence has to be interpreted in terms of levels or grades, whereas, in formative assessment, interest is only in identifying next steps and how to take them. The indicators of development in **Tables 2** and **3** are at too detailed a level for summary reports. What is needed for reporting to parents or to other teachers is, at most, an

overall judgment concerning what has been achieved in terms of, for instance, knowledge and understanding of life processes and living things or scientific inquiry skills.

The process of making a judgment involves comparing the best evidence from the work during the time over which performance is being reported with the criteria defining grades or levels. This does not require the retention of vast amounts of work since the aim is not to arrive at an average level of performance, but one based on the most recent and best performance. In the case of written work, this can be accumulated in a folder in which earlier pieces of work are replaced by later ones that better reflect students' developing abilities. Students can also help in this selection and, in the process, acquire some understanding of the broader goals of their work and of the criteria by which its quality is judged.

While teachers' observations and the collection of students work can provide highly valid information, its reliability may well be low – unless appropriate action is taken to ensure consistency in using criteria. There are various ways in which teachers' judgments can be aligned, or moderated. These include teachers meeting to discuss specific examples of work and comparing their judgments, using exemplars of assessed work – either those published or created within a school, or using a brief test or set of tasks designed to indicate work at a certain level.

A Combination of Ongoing Work and Special Tasks

In some systems, both teachers' judgments and test scores are required in summative assessment – in recognition that tests cannot cover the full range of goals. However, difficulties arise when attempts are made to combine or compare these measures. For reasons mentioned earlier, greater weight is generally given to tests, and it is often forgotten that since tests and teachers' judgments assess different aspects, it is to be expected that their results differ. Thus, using tests to moderate teachers' judgments is a dubious practice. It is preferable for tests and special tests to be used to supplement the evidence teachers have in coming to their judgments. Used in this way, tests and tasks can fill the gaps where regular activities have, for one reason or another, not provided evidence of certain kinds. They are also particularly useful, as good examples, for new and inexperienced teachers. In order to avoid some of the disadvantages of tests, these tasks can be embedded in normal work. The Walled Garden tasks described by Schilling *et al.* (1990) provide an example of embedded tasks (see **Table 3**).

Reporting the Performance of Groups

It is the practice in several jurisdictions – including England, some states of the United States, and several

countries of Southeast Asia – to publish student performance results for schools and to use them as measures of the effectiveness of teachers, schools, and district administrations. When sanctions or rewards are attached to the results, attention is inevitably focused on maximizing the outcomes that are assessed, which then acquire high stakes. The consequence is to focus teaching and the curriculum on what is assessed. As noted earlier, test results are preferred for this purpose – whether or not this can be justified – and the importance of high reliability has the effect of restricting what is tested largely to factual knowledge and consequently narrowing students' learning opportunities. Teachers become concerned that to spend time on inquiry-based work would endanger coverage of the required subject matter. Although this by no means follows, since students develop knowledge and understanding while engaged in inquiry, the perception is a strong one and is reinforced by parents' views of science education.

For a more positive impact, accountability is best based on information about a range of student achievements and learning activities – judged by reference to the context and circumstances of the school and used positively to improve students' opportunities for learning. If the aim is to improve school effectiveness, then – as in the formative assessment of students – self-evaluation at the school level ought to have a role in systems of accountability.

Conclusion

Current views of the goals of primary science emphasize the use of inquiry skills by the students to develop their understanding of key broad ideas in science and of the nature of scientific activity. The assessment of these goals presents a considerable challenge, which conventional tests are far from being adequate to meet. Fortunately, there is an alternative since primary teachers – being generalists in contact with the same pupils for most of each day – are in a good position to gather the rich variety of evidence about their students that can be used both formatively and to summarize learning. Evidence collected by teachers – covering the full range of goals – can be used in the short term to adjust teaching and provide feedback to students to help them identify and take their next steps. The same evidence can be used to identify where students have reached at times when performance is to be reported, by judging the latest and best evidence against the criteria for grades or levels of achievement. Moderation of these judgments through discussion among teachers benefits understanding of goals and how to achieve them as well as the reliability of the assessment results.

Further changes in current practice are desirable to reduce, to the extent possible, the emotional impact which is associated with assessment (Assessment Reform Group

(ARG), 2002). A good deal of this impact derives from fear and anxiety of the unknown. This can be avoided by greater openness about the need for and purposes of assessment. Those more closely involved – parents and students as well as teachers – ought to be fully aware of how evidence is gathered and how judgments are made. Even the youngest students can be given some understanding of what evidence they and their teachers can use to judge whether they are making progress. The involvement of students is particularly important in using assessment to help learning but equally important for summative assessment so that there are no surprises (for students or parents) in the reports of the level reached at a particular time.

The practice of using the results of assessment to set targets for schools and teachers is adding to the stakes of assessment and creating obstacles to achieving the goals of inquiry teaching. The effectiveness of schools and teachers ought to be based on a range of indicators which can also support self-evaluation. Experience in a range of countries (see Harlen, 2007) indicates that disconnecting school accountability from assessment of individual students enables the latter to be used most effectively to improve learning.

Bibliography

- Asoko, H. and Scott, P. (2006). Talk in science classrooms. In Harlen, W. (ed.) *ASE Guide to Primary Science Education*, pp 158–166. Hatfield: the Association for Science Education.
- Assessment Reform Group (ARG) (2002). *Assessment for Learning: 10 Principles*. www.assessment-reform-group.org (accessed September 2009).
- Barnes, D. (1976). *From Communication to Curriculum*. Harmondsworth: Penguin.
- Black, P. J., Harrison, C., Lee, C., Marshall, B., and William, D. (2003). *Assessment for Learning: Putting it into Practice*. Maidenhead: Open University Press.
- Black, P. J., Harrison, C., Lee, C., Marshall, B., and William, D. (2002). *Working inside the Black Box*. King's College London: Department of Education and Professional Studies.
- Black, P. J. and Lucas, A. M. (1993). *Children's Informal Ideas in Science*. London: Routledge.
- Clarke, S. (2001). *Unlocking Formative Assessment*. London: Hodder and Stoughton.
- Duschl, R. R., Schweingruber, H. A., and Shouse, A. W. (eds.) (2006). *Taking Science to School: Learning and Teaching Science in Grades K-8*. Washington, DC: The National Academies Press.
- Harlen, W. (2007). *Assessment of Learning*. London: Sage.
- Harlen, W. and Allende, J. (2006). *Report of the Working Group on International Collaboration in the Evaluation of Inquiry-Based Science Education Programs*. Santiago, Chile: Fundación para Estudios Biomédicos Avanzados de la Facultad de Medicina.
- Hamqvist, K. and Burgen, A. (1997). *Growing up with Science, Developing Early Understanding in Science*. London: Jessica Kingsley.
- Hein, G. E. and Price, S. (1994). *Active Assessment for Active Science, a Guide for Elementary School Teachers*. Portsmouth, NH, Heinemann.
- Masters, G. and Forster, M. (1996). *Progress Maps*. Camberwell, VIC: Australian Council for Research in Education.

- Mercer, N., Dawes, L., Wegerif, R., and Sams, C. (2004). Reasoning as a scientist: Ways of helping children to use language to learning science. *British Educational Research Journal* **30**(3), 359–377.
- National Science Foundation (NSF) (1997). The challenge and promise of K-8 science education reform. *Foundations* 2, p. 7. Arlington, VA: NSF.
- NRC (National Research Council) (1966). *National Science Education Standards*, p. 32. Washington, DC: National Academy Press.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* **18**, 119–144.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education* **5**(1), 77–84.
- SPACE Research Reports (1990–1998). Titles include: *Evaporation and Condensation* (1990), *Light* (1990) *Growth* (1990) *Electricity* (1991), *Materials* (1991), *Processes of Life* (1992), *Rocks, Soil and Weather* (1993), *Earth in Space* (1996), *Forces* (1998). Liverpool: University of Liverpool Press.
- Stiggins, R. J. (2001). *Student-Involved Classroom Assessment*, 3rd edn. Upper Saddle River, NJ: Merrill Prentice Hall.
- Stiggins, R. J., Griswold, M. M., and Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement* **26**, 233–246.
- Wilson, M. and Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education* **13**(2), 181–208.

Further Reading

- Black, P. J. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education* **5**(1), 7–74.
- Black, P. J and Harrison, C. (2004). *Science inside the Black Box*. London: nferNelson.
- Harlen, W. (2006). *Teaching, Learning and Assessing Science 5–12*, 4th edn. London: Sage.
- Harlen, W. and Winter, J. (2004). The development of assessment for learning: Learning from the case of science and mathematics. *Language Testing* **21**(3), 390–408.
- Naylor, S., Keogh, B., and Goldsworthy, A. (2004). *Active Assessment: Thinking, Learning and Assessment in Science*. Sandbach, UK: Millgate House.
- White, R. and Gunstone, R. (1992). *Probing Understanding*. London: Falmer.

Relevant Website

- www.assessment-reform-group.org – Assessment Reform Group.

Assessment in Schools – Secondary Science

J Shemwell, A Fu, M Figueroa, R Davis and R Shavelson, Stanford University, Stanford, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Computer-adaptive testing – A method for administering tests that successively selects questions so as to maximize the precision of the test based on what is known about the examinee from previous questions.

Concept map – A network showing the relationships among named ideas (concepts); concepts are connected with labeled arrows that describe the nature and direction of the connection.

Conceptual inventory – A test of student conceptions about the natural world which has been developed from research on student difficulties with particular topics and has been refined, tested, and validated by detailed observations with a robust sample of students.

Conceptual item cluster – A focused set of conceptual questions embedded in a larger assessment. Item clusters are designed to probe student conceptions in detail; they are developed and validated in the same manner as questions for conceptual inventories.

Diagnostic testing – A form of formative assessment usually administered before a sequence of instruction that informs the subsequent planning process.

Formative assessment – An assessment that takes place in the midst of the learning process and is used to guide further instruction and student learning.

Performance assessment – An assessment that involves concrete, goal-oriented tasks in which students manipulate physical objects to try to solve a scientific problem or address a scientific question.

Psychometrics – The field of study concerned with the theory and technique of educational and psychological measurement, including measurement of knowledge, abilities, attitudes, and personality traits.

Question–demonstration assessment – A question and discussion assessment activity involving prediction followed by explanation of a demonstrated science phenomenon.

Reliability – The extent to which an assessment consistently obtains the same scores with the same group of students while they are in a steady state.

Science notebook – An ongoing written account of what students do and learn in their science class.

Scoring rubric – A set of criteria and standards linked to learning objectives that is used to score student performance on a variety of tasks.

Summative assessment – An assessment of learning after it has occurred that is primarily used for reporting the results of instruction to stakeholders in the educational process.

Validity – The extent to which an assessment measures what it is intended to measure.

Introduction

Science assessment in secondary schools ranges considerably in purpose and use, from formative assessment that is indistinguishable from instruction to strictly summative assessment such as standardized testing for accountability. Some assessment tasks lend themselves to both uses, but the differing goals and requirements of formative versus summative assessment result in different design logics and implementation strategies. When formative assessment is melded with instruction, to change assessment is to change teaching directly and vice versa. Yet, both kinds of assessment are inextricably linked to classroom practice. If the school environment, particularly its economy of scale, necessarily shapes science assessment, then assessment – either directly or indirectly – shapes the school environment. Whether formative or summative, the process of assessment – especially the form and content of assessment tasks – signals to students, teachers, and other stakeholders what science in school is supposed to be about.

To assess science learning is to find out what reasoning and actions a student will perform across a range of scientific knowledge domains and situations. In secondary schools, the enduring challenge is for one teacher or proctor to efficiently assess the learning of many students while doing justice to what it means to know and do science. Doing justice to science – the validity of assessment – necessarily involves value judgments and interpretations of what knowing and doing should look like in different situations and across disciplines. While particular content standards vary, most acknowledge that assessed performance depends upon students' knowledge (of theories, facts, concepts, procedures, and strategies), on the one hand, and their acts of reasoning, on the other.

Reasoning and knowledge are inseparable. For example, students who can reason effectively to control experimental variables in one science domain may perform differently in another. Knowledge and reasoning also differ with the social, cultural, historical, and environmental contexts that are reflected or embodied in assessment activities. For example, students' explanations relating force and motion for realistic situations can stand in direct opposition to those they provide when answering textbook questions. For these reasons, valid assessment of science learning requires a wide array of contexts and tasks, a diversity that mirrors the manifold and interconnected ways of knowing and doing that characterize science itself.

In what follows, we briefly describe six categories of science assessment, sampling widely from extant and developing formats and techniques. This sample, while by no means comprehensive, is nevertheless intended to convey some idea of the breadth of available assessment practices in secondary science. In the interest of brevity, we place less emphasis on some time-honored, yet effective, formats such as multiple-choice items and classroom questioning. Instead, we focus on certain advances in assessment that have helped science educators meet the challenge of achieving ever-better validity (doing justice to science) while working within the demands of the secondary school setting.

Question–Demonstration Assessments

Demonstrations of science phenomena, when structured as interactive, question-driven activities, can be powerful tools for probing students' developing knowledge and reasoning. While question–demonstration formats vary, the predict–observe–explain (POE) sequence popularized by Richard White and Richard Gunstone is typical and depicted in **Figure 1**. In this version, students are presented with the initial conditions of a situation with an uncertain outcome and asked to: (1) predict the outcome; (2) observe what happens; and (3) interpret and explain their observations, reconciling them with their initial predictions.

Question–demonstration sequences are especially useful for formative assessment, and when amplified and extended with discussion and group problem solving, they can form the core of the classroom learning process. They also lend themselves well to larger group settings in which an expert moderator or teacher can intensify engagement, dynamically assess student thinking, and promote and guide the productive exchange of emerging ideas.

Some form of prediction is the signature element of a question–demonstration assessment. While a demonstration's initial conditions will be designed to focus students' attention on specific facts, features, and relationships, the

act of making a prediction brings these ideas into heightened awareness. An important feature of prediction, in contrast to more general forms of questioning, is that students must decide what knowledge applies to the situation at hand. As a result, the act of predicting often elicits students' initial, naive conceptions, making them visible as part of the assessment process.

Observation of the phenomenon provides students with feedback on their predictions. Typically, a teacher or facilitator will perform a single demonstration for the group to observe. Students often observe different things, so teacher and peer mediation are involved in interpreting what happened. With careful moderation, the observation process can stimulate not only productive thinking about science principles but also awareness of the foundations of science in the sense that observed facts are fundamentally interpretations, not mere recordings, of physical reality.

The process by which students explain what they have observed, reconciling these explanations with predictions, is a knowledge building process. When used as formative assessment, a question–demonstration sequence will involve bringing explanations to light while simultaneously assisting students to advance them further. Typical procedures involve short writing prompts extended by discussion and lecture. Responding effectively to students' initial attempts at explanation can demand considerable skill and experience on the part of the teacher, a requirement that stands as a significant challenge to widespread use of this kind of formative assessment. Another limitation is the restricted role allocated to students in the highly scripted question–demonstration process. Although this process may provide for intense engagement and prompt deep reflection, it nevertheless denies students the chance to pose questions and manipulate materials for themselves.

Question–demonstration assessments can be cast in summative form. In one version, students are presented with an initial demonstration setup and asked simply to predict an outcome and justify their prediction; in another, they observe a phenomenon and explain it in terms of science principles and causal mechanisms. These kinds of items are sometimes incorporated into large-scale assessments to signal what is valued in science learning – not only answering questions but also reasoning about and making sense of the physical world.

Performance Assessments

Implementing hands-on tasks in the classroom and in large-scale assessments reflects the importance of doing science. This involves reasoning not just with knowledge in memory but also with an external environment that simulates the conditions and resources with which science is done. Thus, hands-on performance assessment involves concrete, goal-oriented tasks in which students manipulate physical

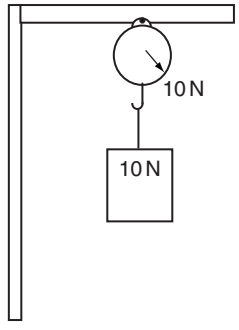
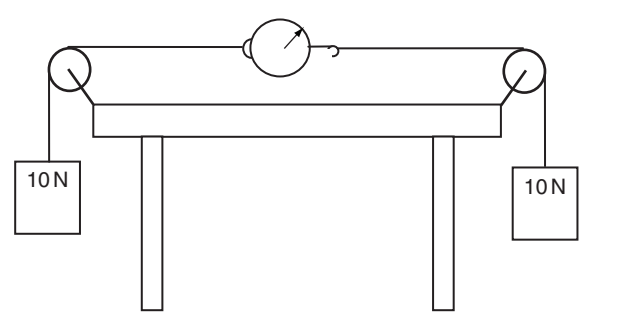
	
<p>Situation I: In this case, the spring balance reads 10 N</p>	<p>Situation II: What will the spring balance read in this case? (ignore the mass of the balance itself)</p>
<p>Part 1: Prediction</p> <p>Which of the choices below best predicts the reading on the balance in situation B?</p> <p>A. 20 N B. 10 N C. 5 N D. 0 N E. 15 N</p> <p>In the space provided, fully explain your reasoning.</p> <p>{A student might predict that the answer is choice A, 20 N, because in the second situation there are two forces, not just one, acting on the balance, providing twice as much force overall.}</p> <p>Part 2: Observation</p> <p>What do you observe? That is, what is the reading on the spring balance in situation II?</p> <p>{The balance in situation II reads 10 N as a result of two opposite 10 N forces. Does the balance in situation I also have 2 opposite 10 N forces? Yes it does!}</p> <p>Part 3: Explanation</p> <p>Why did that happen? Explain the physical principles involved that provide for the reading you observed.</p> <p>{In order for a spring balance to register a force of 10 N and remain at rest, two equal but opposite forces, each of 10 N, must be exerted. Any time there is a tensile force on a static object, an equal and opposite force must also be present.}</p>	

Figure 1 Question–demonstration item in predict–observe–explain format.

objects to try to solve a scientific problem or address a scientific question. The solution or answer is evaluated by a rater or teacher who takes into account not only the student's final result but also the method by which the result was achieved.

Technically, a performance assessment includes a challenge, a response, and a scoring system. The challenge requires students to work with concrete materials to solve a problem and does not specify the steps to be taken. For instance, a student could be given a wire, a bulb, and a battery and asked to light the bulb. The student's response includes his or her actions as well as artifacts produced in the process of tackling the challenge. Responses can be

registered in different formats extending from multiple-choice questions to science-notebook entries. The scoring system delineates the critical knowledge and reasoning expected of students and captures the full range of performance demanded by the task. This could include identification of the right answers, justifiability of procedures, appropriate use of evidence, and effectiveness of problem-solving approaches. **Figure 2** shows a physics performance assessment; and **Figure 3** shows its task instructions and scoring rubric.

Task demands for performance assessment may be thought of as ranging from knowledge-rich to knowledge-lean and process-open to process-constrained. Some tasks,

Students are asked to identify the contents of each of the six boxes (A–F) by using the batteries, bulbs, and wires they are given to complete a circuit. This task requires knowledge of series circuits but leaves problem-solving procedures up to the student.

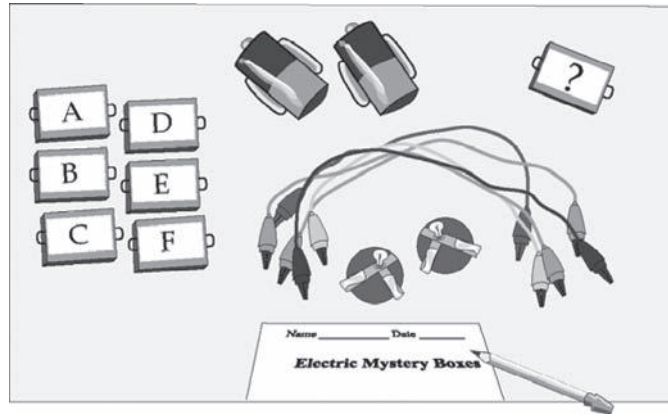


Figure 2 Electric mystery boxes performance assessment. From National Assessment Governing Board. (2006). *Science assessment and item specifications for the 2009 National Assessment of Educational Progress* (pre-publication edition).

The following is a brief description of two warm-up tasks:

1. Students are asked to connect one battery, one bulb, and wires so the bulb lights. They are then asked to draw a picture of this simple circuit.
2. Given mystery box '?,' students are asked to identify whether it contains a battery or a wire. They are told that they can determine the contents of the mystery box by connecting it in a circuit with a bulb.

The following is an excerpt from the main task instructions given to students:

Find out what is in the six mystery boxes A, B, C, D, E, and F. They have five different things inside, shown below. Two of the boxes will have the same thing. All of the others will have something different inside.

[The five options – two batteries, a wire, a bulb, a battery, and a bulb, nothing at all – are presented in words and drawings. Drawings are not provided here.]

For each box, connect it in a circuit to help you figure out what is inside. You can use your bulbs, batteries, and wires in any way you like.

When you find out what is in a box, fill in the spaces on the following pages.

The following is an example of the student response format:

Box A: Has _____ inside.

Draw a picture of the circuit that told you what was inside **Box A**:



The following is a brief description of the scoring system:

For each of the six boxes (A–F), students' responses are scored on two components: (1) identification of the contents of the box and (2) the circuit used to make the conclusion. For each box, if both components are correct, the student receives 1 point; if one or both components are incorrect, the student receives 0 points. Total maximum score is 6 points.

Figure 3 Instructions and scoring guide for electric mysteries task. From Shavelson, R.J., Baxter, G.P. and Pine, J. (1991). Performance assessment in science. R. Stiggins and B. Plake (Guest Eds.), *Applied Measurement in Education* [Special issue] 4(4), 347–362.

like the electric mysteries assessment shown in **Figure 2**, focus on domain-specific knowledge. Other tasks may focus on more general scientific skills. An assessment of this latter type might ask students to determine which of three paper towels soaks up more water and allow them to design an experiment to answer this question.

Hands-on performance assessments have a wide range of applications from small-scale formative assessment in classrooms to large-scale summative assessment such as the US National Assessment of Educational Progress (NAEP) and the Trends in International Mathematics and Science Study (TIMSS). Whether in the classroom or on large-scale assessments, the use of hands-on tasks signals the importance of doing science.

Performance assessment in science is challenging. Tasks are difficult to design and costly in terms of materials as well as the time and effort required for administration and scoring. In particular, large-scale assessments require meeting exacting standards for materials and scoring systems. Yet, these challenges are not insurmountable. For instance, the electric mysteries assessment can be used in ordinary classrooms at reasonable cost and with reasonable effort. The use of information technologies to simulate performance tasks can also reduce many of these difficulties.

Information Technologies and Assessment

By transforming the medium in which students carry out science-related tasks, information technologies have the potential to extend the reach of assessment to probe unique aspects of what students know and can do. For example, computers can support interactive models of hard-to-replicate phenomena such as predator–prey interactions. However, information technologies cannot replace other modes of assessment, such as hands-on performance assessment.

As one example of how information technologies can yield rich assessments, the Science Framework for the 2009 NAEP described several types of interactive computer tasks (ICTs), including information search and analysis, empirical investigations, and simulations. While these types are specified in the framework for summative assessment, they illustrate the techniques that potentially apply to both summative and formative purposes.

Information search and analysis tasks echo how scientists and science learners progress by working with the accumulated knowledge of a domain. These tasks provide students with an information database, pose questions, and ask students to find answers by querying the database. Students are assessed on their abilities to select, evaluate, and synthesize information.

Empirical investigation tasks move performance assessments to the computer platform. Doing so can bypass

certain challenges of the hands-on format. For example, computer simulations of experiments can eliminate the hazards of working with certain materials; facilitate manipulation of variables; collect data automatically; and alleviate the costs and logistical complexities associated with procuring, distributing, and storing the physical materials required for hands-on investigations.

Simulation tasks allow students to model, manipulate, and observe scientific phenomena in ways that are difficult with other formats. Some things are not easily seen in real time (e.g., erosion, planetary motion, and chemical reactions) or by the naked eye (e.g., atoms and bacteria), but they can be sped up, slowed down, or magnified in simulations. For example, students could use a simulation of erosion to analyze the effects of various farming practices on erosion rates. **Figure 4** shows a computer screenshot from an assessment that asks students to use a model to conduct experiments about population dynamics in a mountain lake ecosystem.

Information technologies can capture a range of student responses. Evidence of students' knowledge, reasoning, and skills can be gathered not only from their final answers but also from the actions that they take while working through an assessment task. On a computer-based task, certain keystrokes and actions can be automatically identified and recorded. Examples of relevant actions include the proportion of time spent on various websites in an information search and analysis, the number of trials performed in an empirical investigation, and the manipulation of parameters for a simulation. Scoring these captured sequences of actions yields an unusually direct assessment of students' strategies for approaching scientific tasks, an important but often elusive aspect of science learning.

Information technologies can also improve the efficiency of assessment. Computer-adaptive testing selects items from an item bank based on a student's responses to prior items. By choosing items that are targeted to a particular student, this technology provides an accurate estimate of individual capability with fewer items. Information technologies can scaffold the administration of complex tasks such as concept maps. Software can machine score students' constructed responses (e.g., essays and concept maps), which is faster and less expensive than human scorers while achieving roughly the same accuracy.

Improved efficiencies can support formative uses of assessment. Information technologies are able to efficiently analyze and summarize the vast amounts of performance data that may easily confound teachers' efforts to understand the state of students' learning. With more information on student performances gathered and summarized more quickly, teachers and students can receive feedback and change courses of action on shorter and potentially more effective timescales. Immediate feedback and interactivity built into information technologies can also directly guide student learning. For example, a

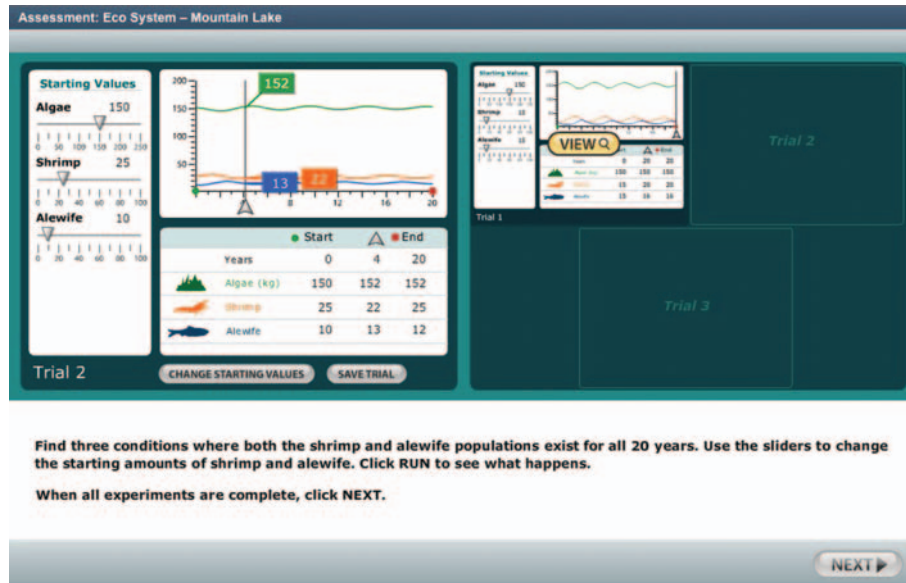


Figure 4 Screenshot from Calipers II predator-prey populations modeling tool. From Quellmalz, E. S., Timms, M. J., and Buckley, B. C. (2009). Using science simulations to support powerful formative assessments of complex science learning. Presented at the annual meeting of the American Educational Research Association, San Diego, CA.

computer-based assessment may immediately alert students if they answer or do something incorrectly, permitting an opportunity to reflect on their understandings and resubmit their answers.

The above examples highlight the promise of information technologies, but significant challenges remain. Unresolved issues include those concerning equity, privacy, and financial and other practical constraints. Further, the use of information technologies in assessment largely represents new psychometric territory. For example, it has not been firmly established whether computer tasks measure the same competencies as hands-on tasks and whether they measure these competencies with comparable reliability and validity. As the use of information technologies in science assessment progresses and expands, these and other challenges will require considerable attention and research.

Conceptual Inventories and Item Clusters

Conceptual inventories and item clusters are developed and used to probe students' knowledge and reasoning about specific science topics in depth. Unlike traditional tests with stand-alone items, these instruments use sets of conceptually related items that allow for deep and explicit investigations of students' particular explanations or mental models of the natural world. The degree of item relatedness ranges in scope from instruments covering broad topic areas to those targeting a single idea or conception. Conceptual inventories and item clusters may be used as tools for diagnostic testing that in turn inform

instructional planning, or researchers and curriculum developers may use them to evaluate students' performances in a randomized trial before and after an experimental intervention. Item clusters probing particular conceptions in depth can be used in large-scale standardized testing. Such clusters provide samples of more detailed information about student learning than typical science achievement items.

Conceptual inventories and item clusters are developed in conjunction with rigorous research into typical naive science conceptions. These instruments generally take the form of relatively short multiple-choice tests, but open-ended questions with detailed scoring systems may also be used. Multiple-choice items such as the one shown in **Figure 5** are carefully constructed and validated such that each distractor corresponds to a prevalent conception identified from the research literature. Thus, incorrect responses can reveal as much about students' knowledge and reasoning as correct responses.

Responses to questions such as the one in **Figure 5** may seem to suggest that students hold relatively stable and coherent conceptions of the natural world. However, such interpretations should be made with caution, as students' science conceptions are influenced by a range of contextual factors, including the unique sociocultural and affective features of a given situation. The confidence with which an individual's response to an item might be interpreted as reflecting stable knowledge and beliefs therefore depends upon the conditions under which the item was developed and tested.

The Force Concept Inventory is one of the best-known and most widely used conceptual inventories (see **Figure 6**).

This 30-item multiple-choice test probes conceptions of force and motion using everyday language and semi-realistic situations. Physics students taking this test can often exhibit profoundly nonscientific conceptions despite having mastered more traditional, problem-based assessments. Results like these demonstrate that being able to solve problems does not necessarily entail well-grounded conceptual understanding. This fact, together with continuing research on students' conceptions, has led to a proliferation of conceptual inventories in other science topics. **Figures 7 and 8** show sample items from two of these.

Many conceptual inventories and item clusters are not intended for formative assessment. An exception is a web-based instrument called Diagnoser (see **Figure 9**). This instrument provides students with immediate feedback on their responses to dynamically ordered multiple-choice and short-answer questions. Diagnoser takes more time and asks more questions than most conceptual inventories

with items tightly clustered around specific science ideas (e.g., the effect of pushes and pulls, explaining constant speed). The result is an intensive assessment of student conceptions that can guide both teachers and students in their choices regarding further instruction and study.

Conceptual inventories and item clusters define, through the questions they ask, what it means to have a strong grasp of fundamental principles in a domain. These instruments can impact science teachers' views about what should be learned in science and how this should be achieved. Generally, teachers voluntarily select and employ conceptual inventories and item clusters. Students' performances on these instruments confront teachers with the prevalence and sturdiness of students' naive conceptions. By exemplifying essential knowledge and reasoning, conceptual inventories and item clusters illustrate an important way in which science assessment can guide and contribute to teaching practice.

Concept Maps

Concepts maps are important tools for measuring the structure of students' conceptual knowledge in a science domain. The integration of ideas, including connections between key concepts, is an important aspect of expert knowledge and a key feature of scientific literacy. As students acquire expertise in a domain through learning, training, and experience, their representations of knowledge begin to more closely resemble the highly integrated knowledge structures that are characteristic of experts.

A concept map consists of nodes and labeled directed lines (see **Figure 10**). The nodes correspond to key terms representing concepts; the lines symbolize a relationship between a pair of concepts (nodes); and the label on the directed line indicates how the concepts are related. Two nodes and a labeled directed line combine to form a proposition, the essential unit of meaning in a concept

What causes day and night?

A. The earth spins on its axis. (0.66)

B. The earth moves around the sun. (0.26)

C. Clouds block out the sun's light. (0.00)

D. The earth moves into and out of the sun's shadow. (0.03)

E. The sun goes around the earth. (0.04)

Figure 5 Example item from the 47-item Project STAR Astronomy Concept Inventory. Some conceptual inventories include proportions of students typically selecting each response option. Here, 26% of students in the research sample responded that day and night are caused by the earth's movement around the sun. From Sadler, P. M. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. *Journal of Research in Science Teaching* 35(3), 265–296.

Imagine a head-on collision between a large truck and a small compact car. During the collision:

A. The truck exerts a greater amount of force on the car than the car exerts on the truck.

B. The car exerts a greater amount of force on the truck than the truck exerts on the car.

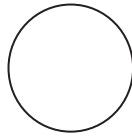
C. Neither exerts a force on the other; the car gets smashed simply because it gets in the way of the truck.

D. The truck exerts a force on the car but the car does not exert a force on the truck.

E. The truck exerts the same amount of force on the car as the car exerts on the truck.

Figure 6 Example item from the Force Concept Inventory. From Hestenes, D., Wells, M., and Swackhamer, G. (1992). Force concept inventory. *Physics Teacher* 30(3), 141–158.

The figure below shows a hollow conducting metal sphere which was given initially an evenly distributed positive (+) charge on its surface. Then a positive charge $+Q$ was brought up near the sphere as shown. What is the direction of the electric field at the center of the sphere after the positive charge $+Q$ is brought up near the sphere?



$+Q$

- (a) Left
- (b) Right
- (c) Up
- (d) Down
- (e) Zero field**

Figure 7 Example item from the Conceptual Survey of Electricity and Magnetism (CSEM). From Maloney, D. P., O’Kuma, T. L., Hieggelke, C. J., and Van Heuvelen, A. (2001). Surveying students’ conceptual knowledge of electricity and magnetism. *American Journal of Physics* **69**, S12–S23.

An elevator moves from the basement to the tenth floor of a building. The mass of the elevator is 1000 kg and it moves as shown in the velocity-time graph below. How far does it move during the first three seconds of motion?

- (A) 0.75 m
- (B) 1.33 m
- (C) 4.0 m
- (D) 6.0 m**
- (E) 12.0 m

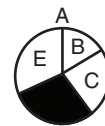
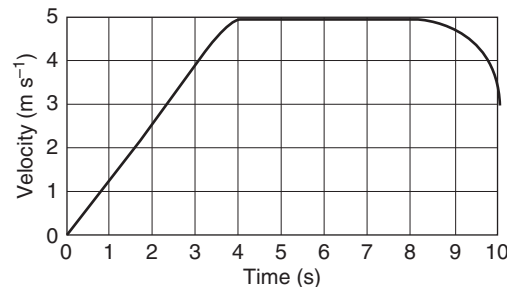


Figure 8 Example item from the Test of Understanding Graphs – Kinematics (TUG-K). The pie chart on the right shows the proportion of students choosing each response option. From Beichner, R. J. (1994). Testing student interpretation of kinematics graphs. *American Journal of Physics* **62**(8), 750.

John and his friends watch their radio-controlled car move along a straight path at their school track. John’s friends mark the position of the car as it travels down the track. Some of the data are shown in the table below. Determine the car’s average speed for the time interval shown in the table.

Position (m)	Time (s)
10	2
15	3
18	4
21	5
23	6
24	7
25	8

Type your answer in the box below.
Your answer must be a number.

m s^{-1}

Feedback to response of 3 m s^{-1} :

While the method you used might work in some situations, it will not give you the average speed for the motion unless the object is changing speed uniformly (at the same rate) throughout the motion.

Figure 9 Example item and feedback from Diagnoser. Diagnoser, unlike most conceptual inventories, provides immediate feedback. The correct answer is 2.5 m s^{-1} . From Minstrell, J. (2008). *Diagnoser project: Instructional tools for science and math*. Retrieved February 27, 2008, from <http://www.diagnoser.com>

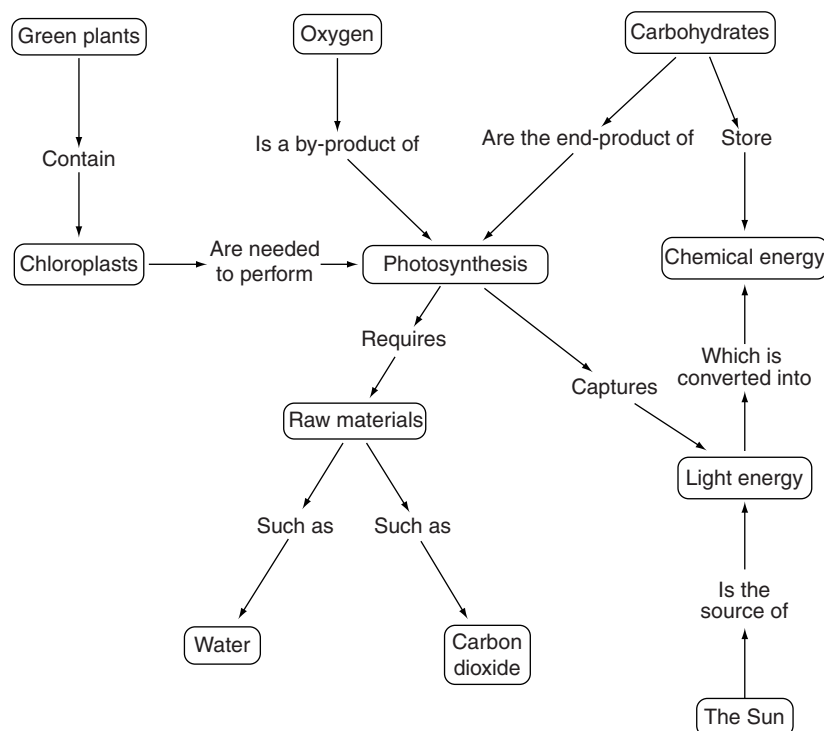


Figure 10 Photosynthesis concept map.

map and the smallest unit that can be used to evaluate the validity of the proposed relationship between two concepts. Concept-mapping tasks encompass a wide variety of techniques that can generally be grouped into two main categories: construct-a-map tasks and fill-in-map tasks. Construct-a-map tasks provide a set of concepts and require students to create all of the nodes and labeled lines in their maps, whereas fill-in-map tasks provide students some or all of the nodes or lines.

In a strict sense, a concept map used as an assessment tool should include a task that draws upon a student's knowledge structure in a domain, a format for the student's response, and a rubric for scoring maps accurately and consistently. Such structured concept-mapping assessments require advanced planning and provide unique challenges in scoring. Scoring concept maps involves evaluating each proposition on the map or comparing students' propositions (two nodes and a labeled directed line) to those on an expert or criterion concept map.

Scoring construct-a-map assessments is generally more challenging than scoring fill-in-map assessments. When students construct their own maps, choices about how to score missing propositions, incorrect propositions, partially correct propositions, and superfluous propositions must all be considered in defining the scoring system. However, while fill-in-map tasks are easier to score than construct-a-map tasks, research suggests that the two techniques do not tap identical aspects of students' understanding:

- Concept terms appear only once on the map
- The map can be organized any way you want
- Use only the concept terms that are provided
- Use only one labeled arrow between two concepts
- You can link a concept to more than one other concept, but you must use separate labeled arrows
- You can only draw arrows between concepts, not to another arrow

Figure 11 Rules for constructing a simple concept map (made available by the Stanford Education Assessment Laboratory); <http://www.stanford.edu/dept/SUSE/SEAL/>

construct-a-map tasks more accurately measure differences in students' knowledge structures.

Training students in the construction of concept maps is an additional challenge of concept map assessments. Concept maps also require students to follow a strict set of instructions (see **Figure 11**). Yet, once the process is understood, concept-map assessments can be easily administered to large numbers of students with minimal direction. Many computer programs are now available to assist students in constructing concept maps, and these programs help to minimize the learning curve.

Concept maps can be used effectively as both summative and formative assessments. They appear in science classrooms as homework assignments, small group work activities, full class collaborations, and individual formal and informal assessments. In formative use, teachers sorting through a set of maps can quickly develop an

understanding of prevalent student conceptions. Spontaneous construct-a-map assessments can be assigned in the middle of a lesson, and more structured fill-in-map assessments are often included on major summative assessments in the classroom as well as on large-scale standardized tests. The Science Framework for the 2009 NAEP specified construct-a-map items. It remains to be seen whether this becomes a standard for such items on other large-scale science assessments.

Science Notebooks

Science notebooks are a compilation of entries that provide a partial and time-bounded record of students' classroom experiences. As instructional artifacts generated alongside student activities, notebook entries are tightly linked to everyday classroom learning.

The characteristics of notebooks vary as a reflection of diverse classroom activities and routines. Entries may include defining concepts, identifying relationships, describing experimental procedures, recording observations, and discussing theoretical models. However, science notebooks go beyond writing; they incorporate drawings, data sets, diagrams, graphs, and tables (see **Figure 12**). These varied forms of representation are essential aspects of scientific inquiry and communication for both students and scientists. Indeed, an important reason for treating notebooks as assessments is to survey students' learning as they engage in a practice that is prevalent among professional scientists.

Scoring criteria for notebooks vary with type of entry. The scoring of an entry on experimental procedures may focus on replicability, while the scoring of recorded observations may focus on level of descriptive detail. When scoring a student's explanation, assessors may focus on the quality of the claim, the type of evidence provided, and the reasoning that links claim and evidence.

Scoring must be carefully aligned to the overall purpose of an assessment. For example, scoring for formative purposes may require looking for particular conceptual difficulties and learning needs. When teachers provide feedback and guidance as written comments, notebooks serve as a valuable record of the ongoing dialog between teacher and student. Teachers and external stakeholders can also score notebooks at the end of an instructional unit for summative purposes.

One challenge of assessing science notebooks is the considerable time and effort required to read and comment on them. This may be ameliorated to some degree by sampling among entries and using tightly focused scoring rubrics. Another difficulty is the requirement to carefully establish procedures and expectations for using notebooks as authentic scientific tools. Failing to do so can result in entries that misrepresent students' true processes of learning or underrepresent their knowledge and skills. These challenges notwithstanding, notebooks provide a unique source of insight into students' knowledge and reasoning in the context of day-to-day classroom activity.

Conclusion

Assessment of science learning in secondary schools is a challenging endeavor. It must encompass a large sphere of activities, many of which can be difficult to render faithfully in a testing environment. It must address complex ways of knowing that are distributed across individuals and specialized tools, and which can shift with changes in context. It must deliver useful information quickly, so that teachers, students, and other stakeholders can readily decide where they stand and what adjustments to make. It must do all of these things at limited cost and with reasonable effort on the part of teachers and proctors.

We have presented here a sample of methods and resources that can begin to address these challenges. In doing so, we have not tried to circumscribe the field but to

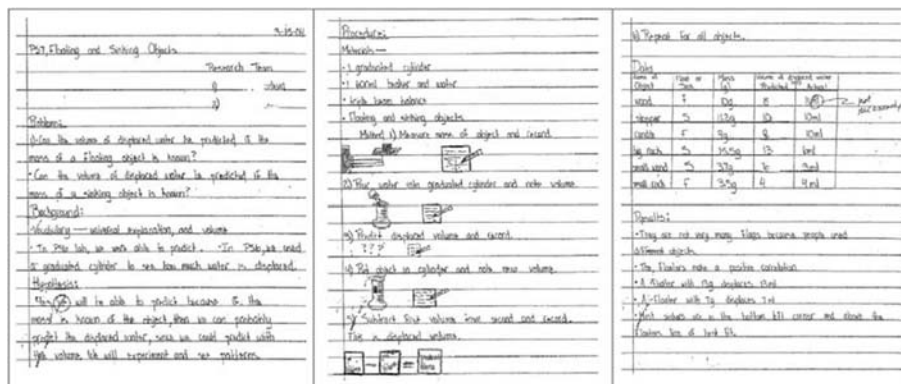


Figure 12 Example of a student's science notebook From Ruiz-Primo, M. A., Li, M., Tsai, S., and Schneider, J. (2007). Testing one premise of scientific inquiry in science classrooms: Examining students' scientific explanations and student learning. *Presented at the annual meeting of the American Educational Research Association, Chicago, IL.*

illuminate certain points of progress within it. Such progress is critical if science educators are to close the gap between what is truly valued in school science learning and what comes to be valued because it can be readily assessed.

See also: Assessment in Schools – Primary Science; Formative Assessment; Instructional System Provided Feedback; Portfolio Assessment; Summative Assessment by Teachers.

Further Reading

- Bennett, R. E., Persky, H., Weiss, A. R., and Jenkins, F. (2007). Problem solving in technology-rich environments: A report from the NAEP Technology-Based Assessment Project (NCES 2007-466). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Black, P. (2003). The importance of everyday assessment. In Atkin, J. M. and Coffey, J. E. (eds.) *Everyday Assessment in the Science Classroom*, pp 1–11. Arlington, VA: NSTA Press.
- Halloun, I. and Hestenes, D. (1985). Common sense concepts about motion. *American Journal of Physics* **53**(11), 1056–1065.
- Halloun, I. and Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics* **53**(11), 1043–1055.
- Hein, G. E. and Price, S. (1994). *Active Assessment for Active Science: A Guide for Elementary School Teachers*. Portsmouth, NH: Heinemann.
- Mintzes, J. J., Wandersee, J. H., and Novak, J. D. (2000). *Assessing Science Understanding: A Human Constructivist View*. San Diego, CA: Academic Press.
- National Assessment Governing Board (2006). *Science Assessment and Item Specifications for the 2009 National Assessment of Educational Progress* (pre-publication edition).
- National Research Council (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Committee on the Foundations of Assessment, Pellegrino, J., Chudowsky, N., and Glaser, R. (eds.). Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- Novak, J. D. and Gowin, D. B. (1984). *Learning How to Learn*. Cambridge: Cambridge University Press.
- Pine, J., Aschbacher, P., Roth, E., et al. (2006). Fifth graders' science inquiry abilities: A comparative study of students in hands-on and textbook curricula. *Journal of Research in Science Teaching* **43**(5), 467–484.
- Quellmaiz, E. S. and Pellegrino, J. W. (2006). Technology and testing. *Science* **323**, 75–79.
- Resnick, L. B. and Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In Gifford, B. R. and O'Connor, M. C. (eds.) *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction*, pp 37–75. Boston, MA: Kluwer.
- Rosenquist, A., Shavelson, R. J., and Ruiz-Primo, M. A. (2000). On the "Exchangeability" of Hands-on and Computer-simulated Science Performance Assessments. *Report*. Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education and Information Studies, University of California, Los Angeles: US Dept. of Education, Office of Educational Research and Improvement, Educational Resources Information Center.
- Ruiz-Primo, M. A., Li, M., Ayala, C. C., and Shavelson, R. J. (2004). Evaluating students' science notebooks as an assessment tool. *International Journal of Science Education* **26**(12), 1477–1506.
- Shavelson, R. J., Baxter, G. P., and Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement* **30**(3), 215–232.
- Shavelson, R. J., Baxter, G. P., and Pine, J. (1991). Performance assessment in science. *Special Issue: Performance Assessment*. Stiggins, R. and Plake, B. (guest eds.) *Applied Measurement in Education* **4**(4), 347–362.
- Solano-Flores, G. and Shavelson, R. J. (1997). Development of performance assessments in science: Conceptual, practical, and logistical issues. *Educational Measurement: Issues and Practice* **16**(3), 16–24.
- White, R. T. and Gunstone, R. F. (1992). *Probing Understanding*. London: Falmer.

Relevant Websites

- <http://cmap.ihmc.us> – CmapTools – Knowledge Modeling Kit.
- <http://ipat.sri.com> – Integrative Technology Performance Assessments.
- <http://www.ncsu.edu> – North Carolina State University, Assessment Instrument Information Page.
- <http://www.sciencenotebooks.org> – Science Notebooks in K12 Classrooms.
- <http://www.stanford.edu/dept/SUSE/SEAL> – Stanford Education Assessment Laboratory.
- <http://www.capsi.caltech.edu> – The Caltech Precollege Science Initiative.

Assessment in Schools – Technology Education and ICT

A Jones, B Cowie and J Moreland, University of Waikato, Hamilton, New Zealand

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In this article we address the issues related to school-based assessment of student learning in technology, including information and communication technology (ICT). We begin by explicating the nature of technology and the characteristics of technology education and then explore the relationship between curriculum, assessment and pedagogy. A sociocultural perspective on the nature of the discipline, curriculum, learning and assessment, and their associated interrelationships underpins our analysis.

Nature of Technology and Technology Education

People use technology to intervene in the world to expand their possibilities, applying both intellectual and practical resources. Technology encompasses a broad range of activities, including the transformation of energy, materials, and information. Technology is not only about artifacts, but also how and why those artifacts are developed and the impact they might have on people and our world. It encompasses more than ICTs, such as computers, the web, and e-mail; it is about products, systems, and environments. It can include electronics and control technology, materials technology, food technology, structures and mechanisms technology, and production and process technology.

Technology is included as a curriculum area in many countries in all continents and regions, including Europe, Asia, Pacific, Africa, and the Americas, where increasing the levels of technological literacy is seen as of intrinsic value for individual development, as a particular insight into culture and for the betterment of society. The most compelling personal reason for studying technology is that it is a major and, some would argue, a determining feature of the world we inhabit. As part of culture, young people need to be introduced to technological practice so that they can understand its nature and be able to participate in it at some level. If technology is indeed a determining feature of the world we inhabit, it follows that young people, as future citizens, need to understand how it shapes the world.

Technology Education and Assessment

Assessment models must accommodate the specific features of the subject and the characteristics of teaching and

learning in that subject. In the case of technology, there is a need to accommodate its complex multidimensional nature. When we look at the development of technology education many of its roots can be traced to traditional technical subjects that emphasized only skills teaching. More recently, it has been linked with and subsumed by science and also conflated with ICT. Consequently, technology education has struggled to establish itself as a school subject with its own identity. The primary focus of technology is to intervene in the made world to extend human capabilities as opposed to understanding the world. Technology encompasses both technological knowledge and technological practice and their interrelationships. Technological knowledge and practices are context dependent. They are associated with and structured by objects, artifacts, and tools in action embedded in social and cultural practices. It's characteristics as practices, as well as a body of knowledge that are crucial to technology education. The uniqueness of technological knowledge, processes, and skills has not always been recognized in general education. Technology education needs to be concerned with involving students with the identification, exploration, and solving of technological problems. Technological problems encompass multiple and interrelated conceptual, procedural, societal, and technical aspects. A comprehensive technology curriculum includes an understanding of the nature of technology, technological knowledge and understanding, technological practices, and the relationship between technology and society. This complexity poses a number of challenges to traditional approaches to assessment.

Assessment and Views of Learning in Technology and ICT

This article offers an analysis of assessment in technology and ICT that is grounded in a sociocultural view of learning. Individual and cognitive notions of learning and achievement have dominated the thinking about assessment. This has contributed to the predominance of individual testing as well as to a focus on written assessment tools in tightly controlled environments. Such approaches are inappropriate for technology and technology education. Social views of learning endorse the view that knowledge is socially constructed and context dependent. Human actions are situated within a historical, cultural, and institutional setting. Sociocultural theory proposes that knowledge emerges through social and cultural activity where tools,

artifacts, and systems mediate thinking, action, and interaction. From this viewpoint, technology assessment is a situated social and cultural activity that cannot be separated completely from the classroom or from ongoing student–teacher interaction.

Assessment of Technology Including ICT

Assessment brings with it connotations of testing but international research suggests that formative assessment, now often referred to as assessment for learning, is one of the keys to enhancing student learning and engagement. In formative assessment, teachers and students use assessment information to enhance student learning within the classroom and on a daily basis. In summative assessment, teachers and students use assessment information to sum up what has been learned. Traditionally, distinct summative assessment tasks have been used to generate data for this purpose. Formative assessment and summative assessment do not need to be seen as independent practices. Effective formative assessment practices can contribute to summative assessment. When formative and summative assessment practices mutually inform each other, teacher assessment practice can become seamless and optimally productive. Effective summative practice reviews learning and can contribute to decisions about further teaching and learning and thus can have a formative function for both teachers and students.

Role of the Teacher in Classroom Assessment

Teachers cannot design and evaluate valid assessment tasks or interact formatively unless they have a clear sense of the ideas of the subject. Teachers need to have an appreciation of the nature of the discipline, its organizing concepts, mediational tools, cultural values, and symbolic and language systems. For teachers to be effective technology assessors, they require clear understandings of the complex multifaceted nature of technology. Valid teacher assessment in technology needs to encompass:

1. conceptual (knowledge and understanding of relevant concepts and procedures);
2. procedural (knowing how to do something, what to do, and when to do it);
3. societal (aspects related to the interrelationship between technology and groups of people); and
4. technical (skills related to manual/practical techniques) aspects.

Teachers need a clear sense of the conceptual terrain and a pedagogical sense of the understandings the students will bring. Teachers cannot provide experiences and activities that guide and monitor student progress toward the understanding of ideas if they themselves do not know

what the ideas are. Successful facilitation of student–teacher formative interaction requires a flexible grasp of the subject matter being explored. With sound content and pedagogical knowledge, teachers can respond to student ideas both formatively and summatively.

Often in technology lessons, students design and create an artifact or a virtual solution in response to a scenario undertaken over several days. The long-term nature of such technology tasks poses particular issues for assessment. One way to help with building connections, continuity, and coherency is to think about these aspects when planning. Teachers can first define the macro-task, that is, the overall task. This macro-task needs to encompass the technological conceptual, procedural, societal, and technical learning outcomes. Then teachers can arrange a series of interrelated subtasks, meso-tasks, which are mutually important for achieving a solution to the macro-task. Micro-tasks may also be planned by teachers and are more localized tasks, embedded within meso-tasks. The macro-, meso-, and micro-tasks form a connected network that provides structure, support, and direction for students. In technology, it is essential that students work iteratively when designing, making, and testing. Often in technology classes, the design process is treated as a series of steps. These steps can become ritualized with lessons structured around each step so that students undertake the process in a stepwise fashion, giving rise to a veneer of accomplishment. When teachers include opportunities for students to distil out the essential criteria for the creation of a technological artifact, and these criteria are used as the foci, students are able to work iteratively across designing, making, and testing process.

Research indicates that when teachers plan for technological, conceptual, and procedural goals, they pay less attention to social and managerial aspects in the classroom, such as the need to take turns, work in groups, work independently, and finish on time. Technology goals can be expressed in terms of the knowledge, design processes, and technical skills. After planning appropriate technology goals, teachers are able to identify appropriate tasks that afford opportunities for students to accomplish the goals. Consideration of the demands and affordances of tasks is essential for planning for the incorporation of assessment for learning strategies, including the provision of feedback. By knowing the ideas and skills inherent in the tasks, teachers can be more clear about their focus for assessment. Teachers may anticipate students' possible actions and ideas and they may rehearse how they might respond. Rehearsal enables teachers to ascertain the technology knowledge and skills required, the suitability of the activities for their students, and to foreshadow potential problems. By prior testing the teachers are more aware of the demands, both conceptually and technically of the tasks, and can interact more confidently and effectively with students.

Effective assessment in technology, both formative and summative, focuses on the multifaceted and multimodal nature of technology. Effective assessment accommodates multiple modes such as drawing and modeling, not just talk, but to communicate and develop ideas. Technologists draw, make models and prototypes to develop and test ideas. In technology, teachers typically engage students in tasks that have a practical aspect. This provides students with access to multiple modes for developing, representing, and communicating their technological ideas. Drawing, modeling, and manipulating materials can contribute to, and are integral to, teachers and students exploring tasks and negotiating ideas together. ICTs provide a tool in technology for representing ideas. They can also engage learners in building understanding through the collaborative construction of an artifact or a shareable product in a virtual environment. They can construct a virtual reality, play with animated puppets, or build a three-dimensional (3D) model of solar system. It is important that students are allowed to use these multiple modes to represent their ideas in assessment settings.

Classroom Interactions and Assessment for Learning

Classroom interactions and feedback are central to teaching technology formatively. One of the strengths of technology lies in the way that ideas and concepts can be expressed in concrete, practical ways. Through exploration of technological ideas through talking and designing, students can begin to see for themselves what they know and can do, and how well they know and can do it. By listening to and interacting with students, a teacher can provide feedback that suggests ways in which students can improve their learning. Feedback, whether from teacher or students, is useful to both teachers and students in providing information that enables both teachers and students to modify the teaching and learning activities in which they are engaged. There are a variety of design and technology pedagogical approaches that can contribute to students learning through formative interactions and assessment for learning practices. These include: comparing and contrasting technologies, categorizing and grouping examples of technology and recognizing exceptions, and making predictions about technology activities. Examining existing technologies helps students distil the features of the technology. What is essential in these activities is that the teacher makes it clear to the learners that the activity is to explore what they think rather than for them to guess the answer the teacher has in their head. This process helps the teacher and the students to identify their current understandings and what they might need to know more about and where they might go next.

Categorizing and grouping examples of technology and recognizing exceptions and talking about them are

activities that challenge students to justify their classifications in terms of technological ideas. These types of activities enable learners to test their understandings about the nature of technology and about particular technologies. Incorporating prediction activities encourages students to apply their knowledge and understanding to future situations, through which they can then either test or realize their ideas. Problems may not be foreseen by the students, and therefore need to be dealt with on the spot by themselves, their peers, and/or the teacher. How teachers deal with the emerging problems students face as they engage with the task at hand, impacts on students' learning. Hence, how teachers' engage with students formatively strongly affects how students undertake technological processes and their learning in technology. One strategy that teachers have used successfully is to keep the key goals (learning intentions) upfront so that students can identify and clarify their own problems within the bounds of the learning goals. When students are engaged with drawing, designing, and making activities, it is important for the teacher to work alongside them so that they can interact with the students' emerging ideas. Leaving the students to work unassisted conveys the message that design and technology is a self-help activity. Teachers need to model techniques, processes, and procedures in front of students.

The process of formative assessment includes students finding out about their learning, recognizing, reacting, and evaluating their learning from their own viewpoint or others. When students monitor their own progress, evaluate their strengths and weaknesses, and devise strategies for improving their learning, their commitment to learning is strengthened. Technological specifications, intended learning, and success criteria align strongly with the technological process and self-assessment process. Key to the success of both peer- and self-assessment in technology is teachers and students discussing about technology together, as this provides students with the language to discuss technological concepts and processes. When teachers target conceptual and procedural aspects and talk about them with their students, more opportunities are afforded for students to learn about these technological aspects and include them in their work. When teachers concentrate on technological learning outcomes and make these clear to the students as outcomes for them to achieve, students are able to review their work and undertake self-assessment focused on these aspects. Technological concepts and procedures require embedding in learning activities through constant articulation and exploration, so that students begin to understand their strengths and weaknesses and how to deal with them. Engaging students in conferring, consultative conversations related to technological aspects, along with the provision of continual support, reinforcement, and overt cueing by teachers, leads to effective student understandings. In technology, teachers need to help

students focus on the conceptual aspects inherent in the processes and procedures involved in reaching solutions. Often a practical focus can overshadow conceptual aspects. When students are encouraged to review their accomplishments in technology in an ongoing manner they are able to do this throughout the technological process, rather than undertaking self-review only as an endpoint summative process. Discussion plays an important part in the classroom assessment. Through discussion students are able to help each other, clarify and develop ideas, communicate ideas and solutions, critique, and seek help. Class and peer discussion provides the teacher with opportunities to notice, recognize, and respond to students' ideas as they are brought to light. ICTs can allow for group collaboration over time and space.

Classroom Summative Assessment

Teachers are responsible for summing up and reporting on student learning. Teachers can summarize formative assessment information during and at the end of a teaching sequence. This information can be used to improve teaching and inform students on where to go next. Accumulated formative assessment information can assist in developing a richer picture of the sum of student learning and achievement, one that includes both the process and product of their learning. Summative assessment as a summing up of learning across time, across individuals, and across classes can support a formative function. Effective formative assessment practices can also contribute to summative assessment, and vice versa. This can be of benefit to teacher, students, and their parents.

Student performance in technology is often assessed by an evaluation of the end product. When this is the case, the product should first be assessed holistically against criteria pertaining to its form, function, and quality. This judgment needs to be followed by identifying strengths and weaknesses. This way the whole can be evaluated and also how the individual elements work together. Judgments need to be a balanced reflection of the whole of student learning, which are then backed with detail. Such summative assessment includes information that can be used to set the next direction for learning. At the same time as looking back at what students have accomplished, teachers also look forward to what they might accomplish next. This fits with a sociocultural perspective that is concerned with learning and achievement over time.

Assessment and ICT

There are three broad overlapping views of the role of ICT in schools, namely, ICT as:

1. a set of skills or competencies to be mastered in an information-rich society;
2. a vehicle for supporting and enhancing the teaching, learning, and assessment of other curriculum subjects; and
3. a subject in its own right as part of technology education.

The first focus is associated with current concerns that students leave school with the knowledge and skills to take an active part in society, that is, with a focus on information and ICT literacy. The second focus is associated with research and development that seeks to utilize the affordances of ICT to enhance student motivation, make teaching more relevant, incorporate multimodal items in lesson materials, represent and communicate ideas, and to generate, store, and analyze student assessment data for formative and summative purposes. The latter role is currently related to ICT as an area of study computing, software design, gaming, and development of ICT-based systems. ICT now lies at the heart of most activities that constitute social inclusion. ICT use is ubiquitous in what it means to be socially, economically, culturally, and politically involved in current society. In this section, we elaborate on the second view of ICT as the first is beyond the scope of this article and the third has been covered in the previous section on technology and technology education generally.

ICTs can enhance teaching and learning through a contribution to formative and summative assessment. They can be used for assessment purposes at the national, school, class, and individual level. National electronic (e)-assessment banks provide access to well-designed assessment items that allow teachers to make comparative judgments in real time to provide evidence of learning which can be used both formatively and summatively. Not only are assessment banks online but ICTs also are useful for developing electronic (e)-portfolios to document learning and achievement in a multimodal way. E-portfolios can contain written work, photographs, video, audio, and other digital media. These allow both teachers and students to document the process and products of learning as well as the teacher feedback and the student response to this. This type of e-portfolio can be used in national assessment and moderation processes.

Digital evidence is easy to share, easy to search, and easy to ask analytical questions of. This means that teachers and schools are able to ask probing questions about the nature of the learning of their students and to think about the implications of this for their teaching and learning programs. School and class e-portfolios allow for ease of sharing between teachers and provide a productive forum for discussion between teachers to compare and contrast student work to explore possibilities and potential for learning through questioning each other on how students have been supported to achieve to a high standard. Teachers are able to expand their teaching repertoire to enhance student learning. Students are also

able to see the potential in the learning tasks and what might be possible for them. For example, the use of photographs as a story line of images adds to student confidence and the sense that they are making progress. Social technologies are on the rise for classroom use. Students can immerse themselves in contexts that challenge and extend their understanding and use these social networkings as part of peer-assessment processes. These social networking ICT environments allow for multiple perspectives in assessment for learning and assessment of learning, including peers, teachers, students, and their families. As ICTs become more integrated into classrooms the process of assessment has the potential to become more dynamic, multimodal, and interactive.

Conclusion

Technology encompasses a very broad range of activities, including the transformation of energy, materials, and information. It encompasses products, systems, and environments. Technology is a multifaceted discipline with a strong multimodal aspect. Technology education requires students to develop and use knowledge, procedures, skills, and ethical/value/societal sensibilities to create technological solutions. A sociocultural view of learning accommodates the social, cultural, and historical dimensions of technology. Assessment models need to reflect this complexity. Too often assessment models do not reflect the characteristics of the subject under consideration. Classroom assessment for technology can have two main purposes: formative and summative. For teachers to assess in both a valid and reliable way, they need to have an understanding of the nature of technology as well as insights into its knowledge and practice. Technology is often taught and learned in long-term tasks requiring that teaching and assessment establish a sense of continuity and coherence. Effective planning not only addresses the multifaceted long-term nature of technology tasks but also accounts for how students and teachers might be actively involved in the assessment process. Clear learning goals and a shared understanding of the specifications of the technology solution enable teacher, student, and peer formative assessment. When teachers systematically collect formative evidence, this can serve a summative function and valid summative

assessment can be used formatively. ICTs have the potential to enhance classroom assessment practices through the provision of additional modes of representing, recording, and reviewing information on student learning process and products.

Assessment is often constructed as an external evaluation of learning. However, school-based classroom assessment has a greater influence on student learning and perceptions of a subject. Valid and reliable classroom assessments in technology need to represent and reflect the multidimensional and multimodal nature of technology. When this occurs the student's experience of teaching, learning, and assessment is enriched.

Further Reading

- Black, P. and William, D. (1998). Assessment and classroom learning. *Assessment in Education* 5(1), 7–74.
- Cowie, B. (2005). Pupil commentary on assessment for learning. In Black, P. (ed.) *Special Issue: Formative Assessment. Curriculum Journal* 16(2), 137–151.
- Gess-Newsome, J. and Lederman, N. (eds.) (1999). *Examining Pedagogical Content Knowledge*. Dordrecht: Kluwer Academic Publishers.
- Gipps, C. (1999). Sociocultural aspects to assessment. *Review of Educational Research* 24, 353–392.
- Jones, A. and Moreland, J. (2004). Enhancing practicing primary school teachers' pedagogical content knowledge in technology. *International Journal of Technology and Design Education* 14(2), 121–140.
- Kimbell, R. A., Stables, K., and Green, R. (1996). *Understanding Practice in Design and Technology*. Buckingham: Open University Press.
- Loveless, A. and Ellis, V. (eds.) (2001). *ICT, Pedagogy and the Curriculum: Subject to Change*. London: Routledge/Falmer.
- Means, B., Roschelle, J., Penuel, W., Sabelli, N., and Haertel, J. (2003). Technology's contribution to teaching and policy: Efficiency, standardisation, or transformation? *Review of Education Research* 27, 159–182.
- Moreland, J., Cowie, B., and Jones, A. (2007). Assessment for learning in primary technology classrooms. *Design and Technology Education: An International Journal* 12(2), 37–48.
- Roth, W. -M. (1995). From 'Wiggly structures' to 'Unshakey towers': Problem framing, solution finding and negotiation of course of actions during a civil engineering unit for elementary children. *Research in Science Education* 25(4), 365–382.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review* 57, 1–22.
- Staudenmaier, J. (1985). *Technology's Storytellers: Reweaving the Human Fabric*. Cambridge, MA: The Society for the History of Technology/MIT Press.
- Wells, G. and Claxton, G. (eds.) (2002). *Learning for Life in the 21st Century*. Oxford: Blackwell.

Assessment in the Early Years

M J Drummond, Formerly of the Faculty of Education, University of Cambridge, Cambridge, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

There are at least three reasons why assessment in the early years merits an entry of its own: first, because of the age of the learners, the youngest children in society, who have not yet learned to be pupils. More important for the discussion of assessment is the fact that these young children are not inexperienced learners: indeed, they have been learning from the day of their birth, if not before. As they enter our preschools and kindergartens, they bring a wealth of learning to the new experiences on offer. However, what they bring is not school learning, the result of the sticks or carrots of the education system; it is the spontaneous learning of children who are, in Wells' (1987: 120) words, "innately predisposed to make sense of their experience." Their learning is not confined by curriculum boundaries or specific subject areas; it is a robust foundation for learning throughout life. One of the challenges for early years educators is to respond to the undisclosed riches and appetite for learning that children bring with them into preschool.

Second, the settings that receive children of 3, 4, and 5 years of age are almost entirely nonstatutory, and, thus, likely to be free from imposed curriculum structures, policies affecting pedagogy, or prescribed assessment practices. There are, inevitably, some pressures and constraints on early years educators, but, on the whole, the early childhood sector is relatively unaffected by the pervasive discourse of levels, standards, and outcomes that characterizes primary/elementary education. Therefore, in looking at assessment in the early years, we may be able to see how learning might be fostered, indeed improved, through practices that are uncoupled from the very different project of documenting individual test scores, or compiling test scores into sets of results for classroom, school, and local systems. This is the central theme explored by Fleer (2002) in a paper speculating on the current Australian context, and the possibility of devising assessment practices in harmony with a sociocultural perspective on early learning.

We may note, in passing, that the alternative that Fleer is advocating, an education for early childhood that eschews levels and standards, is by no means a new invention. In England and Wales, for example, the 1933 Hadow Report on Infant and Nursery Schools took a very firm line on the undesirability of applying standards to the learning of young children:

In none of this should a uniform standard to be reached by all children be expected. The infant school [for children aged five to seven] has no business with uniform standards of attainment. (Board of Education 1933: para. 105)

Third, as a result of these first two factors, there are far more choices for early years educators to make in their assessment practices than for educators in other phases of education: choices about children, about learning, about the purposes of early education, and about the methods, purposes, and outcomes of assessment. These choices are examined in the remainder of this article, which takes as its starting point a definition of assessment as the process by which, in their daily practice, educators observe children's learning, try to understand it, and then put their understanding to good use (a definition used by Carr, 2001).

Choices About Children

A fascinating anthropological study, *Preschool in Three Cultures* (Tobin *et al.*, 1989), explores the very different beliefs about children held by educators in Japan, China, and the USA. The Japanese have found an elegant solution to the problem of defining their shared understanding of early childhood. When asked what sort of child they are trying to educate in their preschools, the Japanese educators reply *kodomorashii kodomo* – a child-like child. A fast and furious debate ensues among the educators from all three countries, revealing strongly contrasting versions of the salient qualities of a child-like child.

Turning to history, rather than geography, Dahlberg *et al.* (1999) offer an analysis of conflicting constructions of children and childhood, starting with the empirical philosopher John Locke (1632–1704), who conceptualized the child as an empty vessel or *tabula rasa*. In this model of childhood, the early years are spent in being prepared for school, being made ready to learn (a view that is by no means extinct, as we shall see). In contrast to Locke's child, Dahlberg *et al.* place "Rousseau's child . . . an innocent in the golden age of life" (p. 45) and "Piaget's child . . . the scientific child of biological stages" (p. 46). Dahlberg *et al.* argue cogently for their own preferred construction, "the child as co-constructor of knowledge, identity and culture," but their strongest argument is about the significance of the choices that have been made in the past and continue to be made today:

We have choices to make about who we think the child is and these choices have enormous significance since . . . they determine the institutions we provide for children and the pedagogical work that adults and children undertake in these institutions. (Dahlberg *et al.*: 43)

Moss and Petrie (2002) take up the theme of alternative categorizations and their determining effect upon provision, arguing that, in the UK, children's services are based on a construction of children as "poor, weak, and needy"; as a result, these services "are not provided as places for children to live their childhoods" (p. 63). The settings where poor, weak, and needy children spend their childhoods are places of rescue, where children can be protected, normalized, shaped, and developed. Moss and Petrie advocate a different way of thinking about children and public provision for them, a way that "addresses questions about the good life, including a good childhood, and starts with ethics and politics" (p. 79).

Moss and Petrie are not alone in seeing other possibilities: in recent years, the thinking of preschool educators in Reggio Emilia, Italy, has become increasingly well known, though it is not always sufficiently recognized that their construction of childhood has its roots in many earlier theorists, from Dewey to Piaget, and, significantly for English educators, the great Susan Isaacs (1885–1948), whose proud claim was to study children as they really are, not as we would like them to be. Her chief contribution to our understanding was to document, richly and analytically, the intense intellectual and social activity with which children spontaneously fill their days. Only in early childhood, she insisted, "before children have been taught to separate learning from playing, and knowledge from life, will you see the strength and spontaneity of the wish to know and understand" (Isaacs 1932: 113).

Malaguzzi, one of the founding fathers of the Reggio system of infant/toddler centers (for children from birth to 3) and preschools (for children from 3 to 6), takes a similar position on the intellectual lives of young children: "Children show us they know how to walk along the path to understanding . . . always and everywhere children take an active role in the construction and acquisition of learning" (Edwards *et al.*, 1998: 67). Rinaldi, who succeeded Malaguzzi as Director of Services to Young Children in Reggio insists, in words that have almost become a slogan for the Reggio approach, that

the cornerstone of our experience is the image of children as rich, strong and powerful . . . They have the desire to grow, curiosity, the ability to be amazed and the desire to relate to other people. (quoted in Edwards *et al.*, 1998: 114)

In a more recent text on the Reggio approach, Rinaldi (2006) elaborates on the same theme, claiming that "the image of the child" is "a cultural and therefore social

and political convention," which allows educators to give children the experiences and opportunities ("the space of childhood" and "a space of life") that value their qualities and their potential, rather than negating or constraining them: "What we believe about children thus becomes a determining factor in defining the education contexts available to them" (p. 83). In addition, of course, part of those contexts, in every early years setting, is the practice of assessment. Our constructions of children shape, for good or ill, the ways in which we assess their learning.

Brooker (2002) gives a telling example, in a disturbing study of 4-year-olds starting school in a single class in a run-down, inner urban neighborhood in England. She shows how assessment practices in that classroom create a "powerful hierarchy of achievement." Brooker describes one apparently child-friendly task, designed for assessing early mathematical learning, in which individual children are invited to sort miniature farm animals into fields; she points out that this task is, in fact, "unfriendly to those children who have never visited an English farm, played with farm animals or encountered small-world toys of any description" (p. 103). She observes one child, Jemma, who assembles in a single field a cat, a farmer, a chicken, and some sheep; when asked why she has placed them together, she replies "Because they friends." Brooker herself takes a credit, not a deficit, view of children, and so her response to Jemma's reasoning is to record her thinking as matching the descriptor "sorts objects using one criterion." However, the official math record card describes her as "unable to sort." Here a construction of children has prevailed in which their divergent, unpredictable responses have no value. Jemma is thus seen as a weak, incompetent child, rather than as "rich, strong and powerful," making the best of a pointless, dead-end task in a way that makes human sense to her.

Choices About Learning

In thinking about learning, there is an abundance of models and metaphors to choose from, some of which will serve children's interests better than others. In recent years, English educators of children under 5 have been offered the construct of "Early Learning Goals", now embedded in the statutory requirements for the Early Years Foundation Stage (for children from birth to 5). These numerous goals, organized into 13 scales, are specified in such bland terms that they are unlikely to do justice to what experienced educators see as the most significant qualities of young children's learning: its unpredictability, its inventiveness and creativity, and its purposefulness and emotional engagement. The metaphor of goals suggests a much more convergent, normative process, in which children proceed, in an orderly manner, from one observable

achievement to another: one by one, the goals are ticked off, and one of the nine levels for each scale is recorded. This process culminates in the completion of the Early Years Foundation Stage Profile, a statutory requirement, for every child, including those in independent, private, and voluntary preschools. The individual profiles, with spaces for 117 ticks, are to be put on record, and submitted, school by school, to the local authority, at the end of the year in which the children turn 5. It is as if learning were a ladder, up which, rung by rung, every child must go.

A different perspective on learning comes from Fromm (1976), psychologist and social theorist, whose books offer a series of critical perspectives on the ills of society. In *To Have or To Be*, he describes what he sees as a confusion between the two alternatives of his title, arguing that we have come to think of certain key concepts – such as love, happiness, and knowledge – as things to have, rather than as ways of being. He suggests that the Western obsession with material goods has spilled over into other categories of experience, so that we have come to treat happiness and love as if they were thing-like, as if we could accumulate them, and lay up ever-increasing stores of them.

We may apply these alternatives to early learning: is it a thing that young children have? Will some children have more and others less? Or is it more important for young children to be learners, fully engaged in being the rich, strong, and powerful learners that the Reggio educators describe?

If we choose to think of learning as a way of being, we will want to know more about the characteristic features of the beings who are immersed in their learning: we can do no better here than turn to the work of Carr and her colleagues in New Zealand. Rejecting levels, goals, targets, and other manifestations of the ladder model, Carr and her colleagues have chosen a narrative metaphor: they see each child's learning as a story. In early childhood settings in New Zealand, each child's learning is recorded, day by day, in individual books, with entries made spontaneously by the observant educators, by parents, and extended family members, and often by the children themselves.

The Learning Stories approach is holistic, not divided into areas of knowledge or specific skills. The stories recognize the unique developing individuality of every learner; they draw families in – parents find their children's stories irresistible. What is more, they document progression: over time, the stories get longer, broader, and deeper. The educators attentively observe the dispositions of the learners, the characteristics that dispose them to be ready, willing, and able to learn. They are especially interested in the five dispositions that reflect the five broad strands of their national early childhood curriculum:

Five broad strands

- belonging
- well-being
- exploration
- communication
- contribution

Five important dispositions

- taking an interest
- being involved
- persisting with difficulty
- communicating with others
- taking responsibility. (Carr, 2001)

In a more recent publication, Carr (2008) claims that resourcefulness and agency are the two most significant dispositional outcomes for early education. Linking her work to the Assessment for Learning projects in the UK, Carr argues that when assessment focuses on learning dispositions, then those dispositions are promoted; indeed, she suggests that when educators focus on resourcefulness and agency, children too begin to recognize and respond to opportunities for expressing just those qualities.

Learning dispositions are also central to the work of Laevers and his colleagues at the University of Leuven, in particular the learner's well-being and involvement, which are seen as essential factors in successful early learning. Laevers (2000) characterizes this as "deep level learning," opposing it to superficial learning, which does not affect the basic competences of the child, and does not transfer to real-life situations. Without high levels of well-being and involvement, only superficial learning is possible: when well-being and involvement are maximized, children function at the height of their capacities, with intense energy and concentration, experiencing the deep satisfaction of purposeful exploration. Laevers and his team have developed observation schedules for identifying levels of well-being and involvement, which are widely used by early years educators, in England as well as in the Netherlands.

Fleer's (2002) critique of current Australian practices in assessment, mentioned briefly above, concludes with an exposition of her preferred approach to learning, which is based on a sociocultural perspective. She echoes a number of themes that have already appeared in this discussion, especially the principle that learning is a process, not a test score, and should be seen as "the transformation of understanding and not some end-point" (p. 112). Here Fleer espouses a view of early learning that is practically synonymous with living. Carr describes an educator who has taken this view one step further, into her own professional life:

When I asked one early childhood staff team whether they had made any changes to the Learning Story process, one of them replied: "Well, we live a Learning Story here." (Carr, 2001: 187)

The choices educators make about learning, Carr reveals, affect their own lives, as well as those of the learners in whose interests they act.

Choices About the Purposes of Early Education

For Rinaldi (2006), the choices are already made: “the entire education system today . . . is involved in the process of evolution from a school of teaching to a school of learning.” She amplifies this position in terms of values, those that “inspire and orient daily life in the schools in Reggio.” One of these core values, she explains, is democracy; she describes how, in Reggio, the schools do more than merely transmit culture, they aspire to be places where

culture is constructed and democracy is lived. School and democracy, a theme that was dear to Dewey, is an important commitment for all of us: school as a place of democracy, where we can all live democracy. (pp. 141–142)

Dewey’s work is, indeed, threaded through with the grand theme of democracy, and equally with the metaphors of direction and growth – powerful alternatives to the ladder model. Indeed, his pithiest of summaries of the overall purpose of the entire educational enterprise centers on the notion of learning as growth: in a late work, *Democracy and Education*, he claims “Education has as its aim at every stage an added capacity for growth” (Dewey, 1916: 54).

However, there are other constructions of the purposes of early education. A striking example appears in an unexpected place: Zigler *et al.* (2006) describe *A Vision for Universal Preschool Education* (in the US, the term universal applies to programs open to all children, rather than programs that target children at risk or in poverty). Their vision is based on a highly successful school-based program known as the School of the 21st Century, a comprehensive program of family and child-care services, based in public (i.e., state-funded) schools, offering all-day, all-year provision for all 3- and 4-year olds. There is a very great deal to admire in this program, and, for English early years professionals, much to envy. However, there is a fatal flaw in the vision, announced in the introductory pages:

We all agree that the primary goal of pre-school education is school readiness. (p.xviii)

This “agreement” is reiterated throughout the book and appears one last time in the closing pages:

The purpose of pre-school, after all, is school readiness. (p. 245)

Malaguzzi is not the only one who would violently dissent from this view, but he speaks for many. He declares roundly:

If the school for young children has to be preparatory and provide continuity with the elementary school, then we as educators are already prisoners of a model that ends up as

a funnel. I think, moreover, that the funnel is a detestable object . . . Its purpose is to narrow down what is big into what is small. This choking device is against nature. (Edwards *et al.*, 1998: 88)

Malaguzzi chooses to see the purpose of early years education very differently: as a sequence of rich and provocative opportunities for children to become 2-, 3-, 4-, and 5-year-olds, on their own terms, in their own ways, nourished by a healthy curriculum diet of big ideas, supplemented with experiences and activities that matter to them. Far from it being the Reggio educators’ first concern to get children ready for school, the overriding priority is to ensure that when rich, strong, and powerful learners enter the statutory system (the elementary school in Italy, the primary school in England, or the K-12 system in the US) the schools are ready for them.

Zigler’s interpretation of the purpose of preschool has serious implications for assessment in the early years, its practices, and its purposes. In the chapter ‘Defining school readiness’, the persistent theme is that when children fail to succeed, or even to thrive, in school or kindergarten, the fault lies in themselves. Two examples will illustrate this position:

Citing results from a survey of more than 7000 kindergartens across the nation, Boyer (1991) reported that 35 per cent of all kindergarteners start school without the skills they need to succeed, and 42 per cent of teachers feel the problem is getting worse. (p. 22)

In a sample of childcare programs in Massachusetts, nearly 3 per cent of the preschoolers exhibited behavioural difficulties so great that they were permanently excluded from their program . . . their inability to negotiate the social and behavioural demands of their early care and education setting meant they could not take full advantage of the learning opportunities open to them . . . the best curriculum and literacy training in the world offer nothing to the child who is . . . disengaged from the lesson. (pp. 27–28)

Who is to blame for this shocking state of affairs? Why, the children! Who have not learned, the authors explain, the necessary school or kindergarten skills of self-control, paying attention to the teacher, following directions, working independently, managing the anxiety of parting from their parents, and ignoring noise and distractions: children, in short, who are simply “poorly prepared” (p. 99).

These are examples of a very strange form of assessment: a form that, as a default position, locates learning problems in the learners, a classic expression of the deficit approach. However, there are still choices to be made: there is another way of seeing. If schooling results in widespread failure for children, might there not be something wrong with the school, its curriculum, and pedagogy? Or with the

preschools and kindergartens that failed to prepare children properly? How good is the best curriculum and literacy training if it does not engage children? If kindergarten teachers can only educate properly prepared children, what sort of education are they offering?

The message from the choices that Zigler and his colleagues seem to have made is plain: if we are to do our best by young children, we need to rethink our attributions of blame, transferring them from the 3-, 4-, and 5-year-old learners to the settings that are failing them.

Choices About Assessment: Methods, Purposes, and Outcomes

This section turns to more practical issues, to consider how assessment is done in the early years, what it is for, and what it does for children. Everything that follows from this final set of choices is rooted in those outlined above, choices about children, about learning, and about early education as a whole.

In New Zealand, as we have seen, the use of Learning Stories as an assessment tool is based on a particular view of learning and learning dispositions. However, it has still deeper roots, in the four principles on which the national early childhood curriculum is founded, which are integral to their assessment practice. The principles are:

- empowerment – the early childhood curriculum empowers the child to learn and grow;
- holistic development – the early childhood curriculum reflects the holistic way children learn and grow;
- relationships – children learn through responsive and reciprocal relationships with people, places, and things;
- family and community – the wider world of family and community is an integral part of the early childhood curriculum.

These principles are enshrined in legislation: they have statutory force. The Learning Stories approach has been designed to realize the principles in practice; the approach is indeed empowering, it enhances children's sense of themselves as competent learners. It reflects the holistic way that children learn, and the reciprocal relationships between children, people, and the learning environment; parents, guardians, and the extended family are seriously involved.

For all these fine-sounding claims, the approach is also entirely practical and down to earth; in a series of booklets to be used as an early childhood assessment resource, educators are encouraged to tackle assessment as a sequence of five active verbs: noticing, recognizing, responding, recording, and revisiting. Carr (2008) reports that "Professional development with this resource indicates that as teachers record and revisit the learning, they notice, recognise and respond to children's activities with

more understanding" (p. 44). Such is the power of a simple, practical resource, rooted in both values and principles.

Where the New Zealand educators turn to narrative and story, the Reggio educators use the concept of pedagogical documentation. This term refers to a whole family of practices (video and audio recording, photographs, transcripts, observations, and notes of discussion and debate) through which they make learning visible. Rinaldi (2006) explains the consequences of using this process as the basis for assessment: "[it] gives us an extremely strong 'antibody' to a proliferation of assessment tools which are more and more anonymous, decontextualised and only apparently objective and democratic" (p. 62). Her definition of the process of assessment is breathtakingly simple "deciding what to give value to." The next step is plain; since assessment is, essentially, "a perspective that gives value," it is a perspective that allows educators "to make explicit, visible and shareable the elements of value" as they work together in the process of documentation (p. 70). It is striking how, in this approach, the responsibility of ascribing value to particular acts of learning is located firmly in the hands of the educators, who are face to face, in relationship, with the children. It is not the task of those constructing "anonymous and decontextualised" assessment tools.

However, not all early years educators are lucky enough to work in Reggio Emilia, or New Zealand, where, over the years, through collaborative enquiry and extremely well-supported continuing professional development for educators, these innovative forms of assessment have gradually taken shape. Fortunately, there are still choices to be made, in every early years setting, whatever the local or national frameworks within which educators work. There are still choices to be made about the curriculum, about pedagogy, about the educators' every interaction with their children, and so on – too many to discuss here. We will just note, in passing, how the provision of a "generous environment," to use an evocative phrase from Susan Isaacs, can have an enormous impact on the quality, depth, and intensity of children's learning.

For example, Worthington and Carruthers (2003) describe how young children's spontaneous mathematical explorations, discoveries, and acts of mark-making are significantly fostered by the availability of materials from the real world, rich in mathematical meaning (such as telephone directories, books of raffle tickets, clocks, diaries, real stamps, real money, and so on). In such a generous environment, Worthington and Carruthers demonstrate that there is more, richer, deeper mathematical learning for educators to assess.

In terms of the pedagogical relations established through talk between adults and children, educators can always choose between, on the one hand, traditional, empty, question-and-answer routines, and, on the other,

what the important Effective Provision of Preschool Education (EPPE) and Researching Effective Pedagogy in the Early Years (REPEY) studies have called “sustained shared thinking,” passages of meaningful, purposeful talk, in which both adults and children contribute ideas worth listening to and thinking about. The REPEY project has identified this style of interaction as a significant factor in settings of the highest quality; what is more, in the settings where this kind of talk is a frequent element of the children’s experience, educators have many more opportunities of seeing children’s minds actively at work, making meaning for themselves. They will be able to recognize the validity of the Reggio concept of “rich, strong and powerful children,” and act on their responsibilities to them.

Conclusion

This exploration of educators’ choices in assessment has shown that their practices can sometimes act against children’s best interests. However, there are always positive alternatives. In defining these, it is necessary, first, to distinguish between the purposes and outcomes of any set of early childhood assessment practices: purposes are nearly always benevolent and can certainly be made to sound benevolent, while the consequences of assessment, what it actually does to or for children, for parents, for educators, may be far from benevolent. The damage done by “ability labeling,” for example, and by “ability-focussed” practices, such as setting and streaming, has been fully documented by Hart *et al.* (2004), who also propose an alternative “anti-determinist” pedagogy. A detailed paper by Rogers (1989) reviews a number of studies on the effects of early labeling, showing how educators’ expectations of individuals, for good or bad, have long-lasting outcomes for young children, especially in the formation of their attitudes to themselves as learners. Carol Dweck and her colleagues, in a long series of ingenious problem-solving experiments, have identified two distinct patterns of learned behavior, which they term “mastery” and “helplessness,” and which have a significant effect on later learning (Dweck, 1999).

Assessment practices in early childhood are most likely to have benevolent outcomes when values (what we value in young children, in learning, in early education) and principles (the carefully considered ways in which we enact our values) are tightly aligned in the pursuit of a shared vision for children, a shared understanding of their powers (not least their spontaneous search for meaning and understanding), and a shared view of the kinds of learning that matter most.

When all these considerations are in place, when wise choices have been made at every level, early years educators are well placed to practice assessment in ways that

work for children. In effective assessment, the educators’ tasks of observing children’s learning, trying to understand it, and their responsibility to put their understanding to good use, are all at the service of young children.

Bibliography

- Board of Education (1933). *Report of the Consultative Committee on Nursery and Infant Schools*. The Hadow Report. London: HMSO.
- Boyer, E. L. (1991). *Ready to Learn: A Mandate for the Nation*. Lawrenceville, NJ: Princetown University Press.
- Brooker, L. (2002). *Starting School: Young Children Learning Cultures*. Maidenhead: Open University Press.
- Carr, M. (2001). *Assessment in Early Childhood Settings: Learning Stories*. London: Paul Chapman.
- Carr, M. (2008). Can assessment unlock the doors to resourcefulness and agency? In Swaffield, S. (ed.) *Unlocking Assessment*, pp 36–54. Abingdon: Routledge.
- Dahlberg, G., Moss, P., and Pence, A. (1999). *Beyond Quality in Early Childhood and Care: Postmodern Perspectives*. London: Falmer Press.
- Dewey, J. (1916). *Democracy and Education*. New York: Macmillan.
- Dweck, C. (1999). *Self Theories: Their Role in Motivation, Personality and Development*. Philadelphia, PA: Taylor and Francis.
- Edwards, C., Gandini, L., and Forman, G. (1998). *The Hundred Languages of Children: The Reggio Emilia Approach – Advanced Reflections*. Norwood, NJ: Ablex.
- Fleer, M. (2002). Sociocultural assessment in early years education – myth or reality? *International Journal of Early Years Education* 10(2), 105–120.
- Fromm, E. (1976). *To Have or to Be*. London: Jonathan Cape.
- Hart, S., Dixon, A., Drummond, M. J., and McIntyre, D. (2004). *Learning Without Limits*. Maidenhead: Open University Press.
- Isaacs, S. (1932). *The Children We Teach*. London: University of London Press.
- Laevers, F. (2000). Forward to basics! Deep-level learning and the experiential approach. *Early Years* 20(2), 20–29.
- Moss, P. and Petrie, P. (2002). *From Children’s Services to Children’s Spaces*. London: RoutledgeFalmer.
- Rinaldi, C. (2006). *In Dialogue with Reggio Emilia: Listening, Researching and Learning*. London: Routledge.
- Rogers, C. (1989). Early admission: Early labelling. In Desforges, C. (ed.) *Early Childhood Education*, pp 94–109. *The British Journal of Educational Psychology Monograph Series No. 4*. Edinburgh: Scottish Academic Press.
- Tobin, J. J., Wu, D., and Davidson, D. (1989). *Preschool in Three Cultures*. New Haven: Yale University Press.
- Wells, G. (1987). *The Meaning Makers: Children Learning Language and Using Language to Learn*. London: Heinemann.
- Worthington, M. and Carruthers, E. (2003). *Children’s Mathematics: Making Marks, Making Meaning*. London: Paul Chapman.
- Zigler, E., Walter, S., and Jones, S. (2006). *A Vision for Universal Preschool Education*. Cambridge: Cambridge University Press.

Further Reading

- Adams, S., Alexander, E., Drummond, M. J., and Moyles, J. (2004). *Inside the Foundation Stage: Recreating the Reception Year*. London: Association of Teachers and Lecturers.
- Drummond, M. J. (2003). *Assessing Children’s Learning*, 2nd edn. London: David Fulton.
- Drummond, M. J., Rouse, D., and Pugh, G. (1992). *Making Assessment Work: Values and Principles in Assessing Children’s Learning*.

- Nottingham: NES Arnold, in association with the National Children's Bureau.
- Isaacs, S. (1930). *Intellectual Growth in Young Children*. London: Routledge.
- Rich, D., Drummond, M. J., and Myer, C. (2008). *Learning: What Matters to Children*. Woodbridge: Rich Learning Opportunities.
- Siraj-Blatchford, I., Sylva, K., Muttock, S., Gilden, R., and Bell, D. (2002). *Researching Effective Practice in the Early Years. Report No. RR356*. London: DfES.
- Wiles, M. (1983). *Children into Pupils*. London: Routledge and Kegan Paul.

Assessment in the History, Civics and Social Studies Domains

P G Avery, University of Minnesota, Minneapolis, MN, USA

© 2010 Elsevier Ltd. All rights reserved.

Assessment in the domains of history, civics, and social studies is associated with a myriad of complex issues. Similar to other content area domains, the content, format, and purposes of assessment are the subject of intense debate, as is the impact of assessment on teaching and learning. However, there are also challenges and issues associated with the assessment of history, civics, and social studies that are unique. This article describes large-scale assessments in the domains, classroom-level assessments, and the issues that make assessment in these domains particularly problematic.

Large-Scale Assessments

Large-scale assessments such as the National Assessment for Educational Progress (NAEP) in civics and US history, the National Assessment Program (NAP) in civics and citizenship in Australia (MCEETYA, 2006), and the International Association for the Evaluation of Educational Achievement (IEA) CivEd study (Torney-Purta *et al.*, 2001) are not intended to provide information about an individual student's achievement, but rather to give a profile of student achievement at a particular point in time. Using complex sampling procedures, the results often serve as a baseline for trend studies; the NAEP assessments in civics and US history are conducted in grades 4, 8, and 12 in the United States approximately every 5–10 years; the NAP in Australia is scheduled to be conducted every 3 years in grades 6 and 10; and the IEA, which has conducted two international studies of adolescents' civic knowledge and understanding, is scheduled to conduct another study of civic and citizenship education in 2008–2009.

National and international frameworks provide insight into how the domains are conceptualized for the purpose of assessment. In the area of civics and citizenship, the frameworks for the 2006 NAEP in the United States, the 2004 NAP in Australia, and the 1999 (IEA) CivEd Study conducted in 28 countries, provide interesting comparisons. The NAEP framework includes civic knowledge, skills, and dispositions; the NAP divides the domain into two components: civics knowledge and understandings, and citizenship skills and dispositions. The CivEd study assessed students' knowledge, skills, understanding of concepts, attitudes, and actions. It is significant that even in large-scale assessments the subject is recognized as more than knowledge and skills. This broad conceptualization of civics and citizenship, which includes dispositions and

behaviors, is supported by civic leaders, professional organizations, and educational mission statements across many countries.

The NAEP and NAP items assessed students' dispositions; however, items were worded such that they assessed whether students could identify and describe the significance of particular civic dispositions in a democracy; importantly, an individual student's civic dispositions were not assessed. For example, students were asked to identify the significance of a sense of civic duty in a democracy as opposed to reporting their own sense of civic duty.

The IEA CivEd study included a test of knowledge and skills with right or wrong answers, and a survey of students' conceptions of democracy, citizenship attitudes, and their current and expected political behaviors, with no right or wrong answers. Whereas national assessments often focus on the structures, functions, and characteristics of government of a given country, the IEA CivEd assessment was framed around broad democratic concepts and principles not specific to a particular nation-state. The primary concepts underlying items in the CivEd study were democracy, citizenship, national identity, international relations, social cohesion, and diversity. The NAEP and NAP assessments included multiple choice and constructed response items; the CivEd Study did not include constructed response items (presumably because of the many languages of the students), but did include Likert-type response items in assessing student attitudes as well as current and expected citizenship participation.

The framework developed for the 2006 NAEP US history is organized around four major themes (e.g., change and continuity in American democracy), eight chronological periods in American history, and two cognitive domains. The two domains are described as "ways of knowing and thinking about U.S. history," and are (1) historical knowledge and perspective, and (2) historical analysis and interpretation. Approximately two-thirds of the items are focused on what the test developers identify as historical analysis and interpretation.

Although multiple choice items still dominate large-scale assessments, constructed response items now often comprise a significant portion. On the 2006 US Civics NAEP, for example, 20% of the items for grade 12 were constructed response. One item asked students to "explain three ways in which the power of the president can be checked by the legislative or the judicial branch." Responses were scored as 'Complete', 'Acceptable', 'Partial', or 'Unacceptable'. Note, however, that constructed

response items do not necessarily require higher order thinking skills. Although there are a range of possible correct responses to this item, it is still knowledge that is being assessed.

In the United States, the No Child Left Behind Act (NCLB) has led to an increased use of statewide tests. Grant and Salinas (2008), who have conducted the most comprehensive review of state assessments in the social studies to date, report that approximately half of the 50 states administer tests in social studies subjects, and about 10 states attach high-stakes consequences to them. They note that the exact number of states engaged in testing is a moving target, with two to three states opting into testing the same year that two to three states opt out.

There is wide variation in what is tested, how it is tested, the consequences attached to the results, and for whom the consequences apply (student, teacher, school, and/or school district). Even the name of the test is not necessarily an indicator of the content of the test. Social studies tests can be predominately tests of history knowledge, while history tests can include items that could easily be interpreted to be within the broader field of social studies. Passing scores are determined by the states, and vary widely. A student who passes a social studies test in one state could quite conceivably move to a neighboring state and be rendered a failure in social studies. The results are not comparable across states because of wide variation in content and the determination of proficiency levels.

As with national assessments, multiple-choice items dominate the state-level assessments because of the ease and efficiency with which they can be scored. Further, when high stakes are attached to the results, such as student retention or failure to graduate, the scoring of multiple-choice items with clear right or wrong answers is more defensible from the standpoint of the test administrators than is the scoring of open-ended, constructed response items.

The document-based question (DBQ), however, has been used in the New York Regents History exam since 2000. The format is to present students with parts of 5–8 historical documents, such as newspaper articles, political cartoons, letters, speeches, advertisements, and posters, each followed by short-answer questions, culminating with an essay question. In the 2008 New York Social Studies Regents Examination for US History and Government, for example, students were given eight documents related to reform movements in the 1800s and early 1900s (e.g., an anti-child labor poster). The directions for the essay were: “Discuss the social, political, *and/or* economic problems addressed by reformers in the 1800s and early 1900s. In your discussion, include the methods used by reformers to expose these problems.” Students were further instructed to:

- develop all aspects of the task;
- incorporate information from at least five documents;

- incorporate relevant outside information;
- support the theme with relevant facts, examples, and details; and
- use a logical and clear plan of organization, including an introduction and a conclusion that are beyond a restatement of the theme.

The task clearly requires more high-level thinking than the typical multiple-choice or short-answer question. However, is it more authentic with respect to the type of work historians do?

Grant *et al.* (2004) analyzed another DBQ from the New York State Regents Examination in Global History and Geography, and concluded that the task fell far short in terms of representing or reflecting the type of work historians do. They noted that historians always examine an entire document (the New York exam often edits primary source documents), engage in sourcing (evaluating the source of documents) and corroborating (comparing the facts presented in different documents), and present historical arguments based on their analyses. The researchers concluded that “if statewide tests claim to assess accurately students’ historical understanding, then those tests should resemble more closely the work that historians do” (p. 337).

The results and secondary analyses of large-scale assessments provide some insights into student achievement in subject areas. Some of the conclusions are predictable; for example, all studies show students of parents with higher incomes, education, and resources outperform their counterparts. Achievement gaps exist between dominant and nondominant groups, and between native-born students and immigrants. However, other findings are somewhat less expected, and potentially more valuable. For example, frequent (weekly) tests in social studies classrooms are associated with lower achievement levels on the US History NAEP. Minority and lower income students often receive less of the high-quality, active instruction that is associated with higher achievement scores (Smith and Niemi, 2001). Students who report participating in frequent, substantive discussions in their social studies classrooms tend to demonstrate higher achievement scores in history and civics, and are more likely to state that they expect to vote as adults (Torney-Purta *et al.*, 2001). But it is unclear if and how results such as these impact policymakers and classroom teachers.

The national climate of accountability and testing in the United States appears to affect most social studies teachers, regardless of whether their state mandates social studies tests or uses the tests to determine student promotion or graduation; however, the degree to which it changes individual teacher’s classroom practices varies widely. Grant (2006) surmised that teachers are most likely to change their selection of social studies content as a result of testing, less likely to change their assessment

procedures, and least likely to change their instruction. The next section describes what we know (and do not know) about classroom-level assessments.

Classroom-Level Assessments

The content of classroom assessments is presumably shaped by state/national standards, textbooks, and local curriculum guidelines. However, there has been very little research on the content, format, or impact of classroom assessments. On the IEA CivEd survey, teachers of civics-related subjects in 26 countries reported that written essays and oral participation were the most common type of assessments used in classrooms. Teachers from Eastern European countries were more likely to report the use of multiple-choice tests than were those from Western European countries.

In the United States (not included in the IEA teacher reports), the NAEP studies provide an indication of the types of assessments used most frequently in social studies classrooms. On the 2006 US History NAEP, teachers were asked to report how often they used five different types of assessments (see **Table 1**). Fill-in-the-blank assessments were used most often in grade 4, while written paragraphs were the most frequently reported type of assessment in grade 8. It is difficult to draw any conclusions about the use of projects without knowing the type and quality of projects, but unlike the other four assessments, projects are more likely to involve peers. The other four types of assessments are almost always conducted at the individual level. Not surprisingly, the emphasis on writing increases from grade 4 to grade 8; however, the preponderance of fill-in-the-blank assessments, which almost always require lower level recall, at both grade levels is striking.

As reported by the students responding to the 2006 US History NAEP, the emphasis on writing decreases at grade 12, perhaps because teachers at this level often have a much higher number of students than do teachers at grade 4 or 8 (see **Table 2**). One of the more interesting (or alarming) findings is that 66% of the grade 12 students report taking a test or quiz at least once a week. Students

who take a test or quiz 1–2 times a month, however, score significantly higher on the US History NAEP than do students who are tested more frequently.

There are no systematic studies that examine the use of more nontraditional assessments in the classroom. Similar to educators in other subject areas, however, social studies teachers' repertoire of assessment types was enhanced when notions of alternative assessments and authentic assessments received greater attention in the 1990s. Portfolios, performances, exhibitions, debates, etc. – all modes of assessment teachers had used previously – became more legitimized as valued means of assessing students' understandings. Rubrics, appropriately developed, prompt teachers to be more explicit about their goals, and the criteria associated with those goals. For students, well-designed rubrics can clarify tasks and grading criteria.

Fred Newmann's work on authentic assessment has had a special impact on social studies education, in part because his original work was with the social studies. He conceptualized assessment as integrally connected to instruction and student performance. In the early 1990s, Newmann and his colleagues at the University of Wisconsin's Center on Organization and Restructuring Schools (CORS) initiated a major research program focused on students' authentic intellectual achievement. They reasoned that meaningful teaching and learning should focus on the quality of work that students produce. The CORS group suggested that authentic intellectual achievement should involve: (1) students constructing knowledge through (2) disciplined inquiry that has (3) value beyond the classroom (Newmann *et al.*, 1995). Their conceptualization of authentic intellectual achievement reflects much of what we know about good teaching and learning: students are actively involved in producing and developing understandings, rather than reproducing isolated bits of information; they are using disciplinary methods, concepts, and generalizations; and they are recognizing a strong connection between the form and substance of the work they are doing in school and the intellectual work that takes place outside of school.

Newmann and coworkers suggest that authenticity may be reflected in three areas: assessment tasks, instruction, and student performance. They developed parallel,

Table 1 US social studies teachers' report of assessments used one or more times per week^a

Type of assessment	Grade 4 teachers	Grade 8 teachers
Extended essays	20%	36%
Fill in the blank	72%	74%
Projects	53%	62%
Written paragraphs	62%	81%
Multiple choice	13%	22%

^aData from the 2006 US History NAEP, public schools.

Table 2 US grade 12 students' report of assessments used one or more times per week^a

Type of Assessment	
Take a test or quiz	66%
Write short answers to history/social studies questions	59%
Write long answers to history/social studies questions	18%
Group project	18%
Give a report	11%
Write a report	11%

^aData from the 2006 US History NAEP, public schools.

Table 3 CORS standards for authentic assessment tasks^a*Construction of knowledge:**Standard 1. Organization of information*

The task asks students to organize, synthesize, interpret, explain, or evaluate complex information in addressing a concept, problem, or issue.

Standard 2. Consideration of alternatives

The task asks students to consider alternative solutions, strategies, perspectives, or points of view as they address a concept, problem, or issue.

*Disciplined inquiry:**Standard 3. Disciplinary content*

The task asks students to show understanding and/or use of ideas, theories, or perspectives considered central to an academic or professional discipline (e.g., democracy, social class, market economy, and theories of revolution).

Standard 4. Disciplinary process

The task asks students to use methods of inquiry, research, or communication characteristic of an academic or professional discipline.

Standard 5. Elaborated written communication

The task asks students to elaborate on their understanding, explanations, or conclusions through extended writing.

*Value beyond the classroom:**Standard 6. Problem connected to the world beyond the classroom*

The task asks students to address a concept, problem, or issue that is similar to the one that they have encountered, or are likely to encounter, in life beyond the classroom.

Standard 7. Audience beyond the school

The task asks students to communicate their knowledge, present a product or performance, or take some action for an audience beyond the teacher, classroom, and school building.

^aFrom Newmann, F. M., Secada, W. G., and Wehlage, G. G. (1995). *A Guide to Authentic Instruction and Assessment: Vision, Standards, and Scoring*, pp 81–85. Madison, WI: Wisconsin Center for Education Research.

though different, standards and scoring criteria for each area. The authenticity of assessment tasks was evaluated according to seven standards (Table 3).

The CORS researchers conducted an extensive study of authentic intellectual achievement in math and social studies classes in 24 restructured elementary, middle, and high schools. They found a high correlation between level of authentic pedagogy and student performance, that is, the higher the quality of instruction and assessment, the higher the quality of student performance. Their observations of classrooms and review of student tasks, however, suggested that in most classrooms, the authenticity of pedagogy (i.e., instruction and assessment) was fairly low. When teachers did implement more authentic pedagogy, all students benefited, regardless of ethnicity, socioeconomic status, gender, or achievement level. The study is one of the few that has looked systematically at the quality of classroom assessments used by social studies teachers.

Although the criteria for authentic intellectual achievement can be applied across subject areas, social studies educators have been particularly receptive to the framework.

The National Council for the Social Studies (NCSS) published sets of professional development materials at the elementary, middle, and high-school levels based on the CORS work. Social studies curriculum materials have been developed around the framework for authentic assessment. At least two states, Wisconsin and Michigan, have embedded parts of the framework into their state standards and assessments.

Frameworks or rubrics have been developed for assessing many of the rich experiences that take place in engaging social studies classrooms, such as mock trials, role plays, historical debates, and simulated legislative hearings. Discussions about controversial contemporary or historical issues in open and supportive classroom environments are considered by many social studies educators and researchers to be critical experiences for young people in preparing them for adult citizenship. Discussion formats that have received the most attention in social studies include Socratic seminars, structured academic controversies, or public issues discussions. An assessment for an issues discussion, such as the one developed by Harris (2002), includes substantive and procedural criteria. For example, students receive credit for stating and identifying issues, using disciplinary knowledge, and elaborating with reasons or evidence (substantive), or for inviting other students to contribute, recognizing another student's contribution, and providing summarizing statements (procedural). Credit is deducted for negative behaviors such as monopolizing the discussion.

Although professional organizations are very supportive of more authentic assessments, little research has been conducted to indicate how widely such assessments are used in social studies classrooms or how they impact student learning.

Issues and Challenges

There are at least five significant challenges and issues that are particularly salient to assessment in the domains of history, civics, and social studies. First, and perhaps most important, the purpose of each of these subjects in the schools is integrally tied to citizenship education. The social studies is defined by NCSS as "the integrated study of the social sciences and humanities to promote civic competence." The study of a nation's history is undertaken in part to give students a sense of their heritage, and a historical perspective that will enable them to make more informed decisions as citizens. Civics as a subject of study, of course, has as its focus the knowledge, skills, and values that enable participation in a democratic society. The citizenship purpose associated with these domains can be problematic in terms of assessment. Engaged and enlightened citizenship, the goal of most democratic societies, entails much more than knowledge

of the structures and functions of government or even an understanding of the principles of democracy. It also involves values and behaviors that are not only difficult to agree upon, but are also very difficult to assess. Many argue, for example, that service learning provides students with experiences that are likely to promote more concerned, community-oriented citizens. However, the understandings of people and communities that often develop through service learning do not lend themselves well to traditional assessments.

Although it is possible to assess citizenship attitudes and behaviors (as the CivEd study did), if the results are linked to individual students, how should the results be used? Also, at what point does the assessment of an individual student's citizenship values and behaviors become undemocratic? This is particularly troublesome in countries that are home to peoples with different ethnic, religious, and cultural backgrounds. However, if assessment is limited to knowledge and skills, it becomes what most social studies educators would consider a very shallow and minimalist vision of citizenship.

Second, the domains of history, civics, and social studies are more likely to provoke public debate than many of the other subject areas in schools. The controversial nature of these domains has implications for assessment. The question "Whose history should we teach?" is intertwined with "Whose history should we assess?" Do we assess students on their knowledge and understanding of the dark side of our countries' histories? Do we teach and assess conceptions of citizenship that are critical, participatory, community oriented, and/or law-abiding? Assessments often reflect the knowledge and skills about which there is the most agreement in a society and those that are least likely to provoke controversy; unfortunately, this sometimes reduces history to trivial pursuits and civics to a dull study of the structures and functions of government, all of which are unlikely to engage many students.

Third, there is less agreement within the profession as to the nature of history, civics, and social studies, particularly in comparison to disciplines such as math and science. In a study of five subject-area departments at 16 high schools, Grossman and Stodolsky (1995) found that English and social studies teachers viewed their subject areas as less defined than faculty in science, math, and foreign languages, and were less likely to develop common exams with their colleagues. The Ministerial Council in Australia reported that "the definitions associated with certain key [citizenship and civics] concepts were not generally agreed across jurisdictions, nor was their appearance in formal curriculum documents universal" (MCEETYA, 2006: 3). Similarly, in the IEA CivEd study, teachers from across 26 countries reported a lack of consensus as to what constitutes civic education. Moreover, the study conducted by the European Commission (2005) on citizenship education in Europe found different

conceptualizations of citizenship and citizenship education across the 30 countries surveyed. The ambiguity associated with the social studies field, and citizenship education in particular, suggests greater variation in the content of classroom-based assessments, and a difficulty achieving consensus among social studies educators about the content and nature of assessments.

Fourth, increased calls for accountability, combined with the lower status of civics and social studies in schools, present difficult dilemmas for social studies educators. The NCLB, passed in 2001 in the United States, identifies civics, government, history, economics, and geography as core subjects. However, unlike reading and math, states are not required to administer tests in these areas. When social studies tests are mandated at the state level, social studies tends to enjoy increased status and resources. The negative consequences of mandated state-level tests often include the narrowing of the curriculum, teaching to the test, and a focus on the recall of information. However, when social studies is not among the state-level mandated tests, it becomes marginalized. The intensive focus on reading and math has led to a decrease in social studies instruction, particularly at the elementary level in high-minority, low-income schools (Jennings and Rentner, 2006; von Zastrow and Janc, 2004). It may take 5–10 years before we can assess the effect of less social studies instruction on young people's social studies knowledge and understandings.

Concerns about the marginalization of social studies led the NCSS, in collaboration with national history, economics, geography, and civics education organizations, to issue a joint statement on NCLB in 2007 calling for the US government to include the social studies disciplines among those for which state-level assessments are mandated. At the same time, the NCSS (1994) supports the development of comprehensive assessment plans that:

1. are designed to inform pedagogical practices;
2. are aligned with local standards and curriculum;
3. include multiple types of assessments, such as portfolios, teacher observation, essays, student performance, and multiple-choice responses;
4. directly connect to the goals of citizenship education; and
5. assess students' knowledge, thinking skills, values, and social participation.

Most social studies educators are supportive of comprehensive assessment plans, and are concerned about the use of single, high-stakes tests that determine whether a student is promoted to the next grade level or graduates from school. They tend to view assessment more as a classroom-based tool integrated into curriculum and instruction for the purpose of providing ongoing support and feedback to teachers and students. They also express concern about the validity and reliability of all assessments, but particularly those with high stakes attached. Yet, the intense pressure for accountability, and concerns

about the marginalization social studies when it is not tested, suggest that state-level testing will continue and possibly increase.

Fifth, history education, in particular, is often ill-served by the way in which assessments have been conceptualized and implemented. That which can be readily tested in civics and history is often that which is least significant and meaningful. The broad survey nature of courses makes it difficult to assess students' in-depth knowledge and understandings. The type of historical skills that are most meaningful, that help students to understand historical significance, take historical perspectives, and develop and analyze historical interpretations, for example, are not those that can be assessed well in one setting at one time with one test or assessment task. Levstik and Barton (2005) noted that perspective recognition, an aspect of historical interpretation, "cannot be taught, practiced, and mastered during the course of a single, seven-step lesson; it requires sustained attention in a variety of contexts over the course of many lessons, many units, and many years" (p. 160). They advocate the type of comprehensive assessment plan that integrates curriculum, instruction, and assessment; is an ongoing part of the students' educational experiences; and involves students in reflecting on their own development. This vision stands in contrast with, say, the high percentage of students who report completing fill-in-the-blank exercises on a weekly basis in their US history classes.

In summary, the knowledge and skills that can be easily assessed in the history, civics, and social studies domains are often those which are least significant to academicians, and least meaningful to students. The movement toward more authentic assessments holds promise, but is frequently at odds with a culture of testing and accountability. Most striking, however, is the lack of systematic research on the ways in which assessments are implemented in history, civics, and social studies classrooms, and their impact on teaching and student learning.

See also: Alternative Assessment; Portfolio Assessment.

Bibliography

- European Commission (2005). *Citizenship Education at School in Europe*. Brussels: Eurydice European Unit.
- Grant, S. G. (2006). Research on history tests. In Grant, S. G. (ed.) *Measuring History: Cases of State-Level Testing across the United States*, pp 29–56. Greenwich, CT: Information Age.
- Grant, S. G., Gradwell, J. M., and Cimricz, S. K. (2004). A question of authenticity: The document-based question as an assessment of students' knowledge of history. *Journal of Curriculum and Supervision* 19(4), 309–337.
- Grant, S. G. and Salinas, C. (2008). Assessment and accountability in the social studies. In Levstik, L. S. and Tyson, C. A. (eds.) *Handbook*

- of Research on Social Studies Education*, pp 219–236. London: Taylor and Francis.
- Grossman, P. L. and Stodolsky, S. S. (1995). Content as context: The role of school subjects in secondary school teaching. *Educational Researcher* 24, 5–11; 23.
- Harris, D. E. (2002). Classroom assessment of civic discourse. In Parker, W. C. (ed.) *Education for Democracy: Contexts, Curricula, Assessments*, pp 211–232. Greenwich, CT: Information Age.
- Jennings, J. and Rentner, D. S. (2006). Ten big effects of the No Child Left behind Act on public schools. *Phi Delta Kappan* 88(2), 110–113.
- Levstik, L. S. and Barton, K. C. (2005). *Doing History: Investigating with Children in Elementary and Middle Schools*, 3rd edn. Mahwah, NJ: Erlbaum.
- Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) (2006). *National Assessment Program: Civics and Citizenship Years 6 and 10 Report*. Carlton South: MCEETYA.
- National Council for the Social Studies (NCSS) (1994). *Expectations for Excellence: Curriculum Standards for Social Studies*. Washington, DC: Author.
- Newmann, F. M. and associates (1996). *Authentic Achievement: Restructuring Schools for Intellectual Quality*. San Francisco, CA: Jossey-Bass.
- Newmann, F. M., Secada, W. G., and Wehlage, G. G. (1995). *A Guide to Authentic Instruction and Assessment: Vision, Standards, and Scoring*. Madison, WI: Wisconsin Center for Education Research.
- Smith, J. and Niemi, R. G. (2001). Learning history in school: The impact of course work and instructional practices on achievement. *Theory and Research in Social Education* 29(1), 18–42.
- Torney-Purta, J., Lehmann, R., Oswald, H., and Schulz, W. (2001). *Citizenship and Education in Twenty-Eight Countries: Civic Knowledge and Engagement at Age Fourteen*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Von Zastrow, C. and Janc, H. (2004). *Academic Atrophy: The Condition of the Liberal Arts in America's Public Schools*. Washington, DC: Council for Basic Education.

Further Reading

- Grant, S. G. (2003). *History Lessons: Teaching, Learning, and Testing in U.S. High School Classrooms*. Mahwah, NJ: Erlbaum.
- Grant, S. G. (ed.) (2006). *Measuring History: Cases of State-Level Testing Across the United States*. Greenwich, CT: Information Age.
- Harris, D. E. and Yocum, M. (2000). *Powerful and Authentic Social Studies (PASS): A Professional Development Program for Teachers*. Washington, DC: National Council for the Social Studies.
- Lee, J. and Weiss, A. R. (2007). *The Nation's Report Card: U.S. History 2006 (NCES 2007-474)*. U.S. Department of Education, National Center for Education Statistics. Washington, DC: US Government Printing Office.
- Lutkus, A. D. and Weiss, A. R. (2007). *The Nation's Report Card: Civics 2006 (NCES 2007-476)*. U.S. Department of Education, National Center for Education Statistics. Washington, DC: US Government Printing Office.
- Miller, B. and Singleton, L. (1997). *Preparing Citizens: Linking Authentic Assessment and Instruction in Civic/Law-Related Education*. Boulder, CO: Social Science Education Consortium.
- National Council for the Social Studies (NCSS) (1999). *Special Issue: Authentic Assessment in Social Studies*. *Social Education* 63(6).
- Niemi, R. G. and Sanders, M. S. (2004). Assessing student performance in civics: The NAEP 1998 civics assessment. *Theory and Research in Social Education* 32(3), 326–348.

Pike, M. A. (2007). The state and citizenship education in England: A curriculum for subjects or citizens? *Journal of Curriculum Studies* 39, 471–489.

US Department of Education (2006). *U. S. History Framework for the 2006 National Assessment of Educational Progress*. Washington, DC: Author.

Wineburg, S. (2004). Crazy for history. *Journal of American History* 90(4), 1401–1414.

Relevant Websites

<http://www.socialstudies.org> – National Council for the Social Studies.

<http://www.nysedregents.org> – State Assessment: Social Studies Regents Examinations.

<http://nces.ed.gov> – The Nation's Report Card: The National Assessment of Educational Progress (NAEP).

Assessment in the Workplace of Performance, Developing Expertise and Competence

A J Brown, University of Warwick, Coventry, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Competence-based assessment – An assessment process involving the collection of evidence about the ability of a person to perform a particular activity to the required standard.

Workplace-based assessment (WPBA) – The assessment of performance based on what a person (often a trainee) actually does in the workplace when engaged in work tasks.

Introduction

Assessment in the workplace has historically performed a number of functions that have included use in job allocation, appraisal, pay or promotion procedures; recognition of achievement; accreditation of competence; facilitating access to further education and training; and encouraging further personal learning and skill development. Assessment in the workplace in this article is considered in three contexts: assessment of performance linked to initial education and training (as in medical education); assessment of the development of expertise in knowledge-intensive work (as in aerospace companies); and the accreditation of competence, as represented in national systems of competence-based assessment. Assessment in the first two contexts is strongly situated and generally works well, whereas the success of national competence-based systems is much more mixed. There has been some success where the primary concern has been on supporting individuals getting their skills, knowledge, and understanding related to current work roles assessed and accredited. However, recent trends have seen growing interest in how more-formative workplace assessment can align individual development with organizational development. This interest has included an examination of the extent to which processes of formative assessment and critical reflection in the workplace can support employees' commitment to lifelong learning and knowledge-transformation processes at work. This article examines the success of situated approaches to performance assessment and expertise development in the workplace, some of the shortcomings of national competence-based systems, and the characteristics

of an approach to the assessment of workplace learning that is oriented toward the future, focusing upon how best to support employees' commitments to lifelong learning and active knowledge transformation.

Workplace-Based Assessment of Performance in the Workplace

Thomas Huxley famously argued in 1876 that something should be done to aid the medical student toward the acquirement of real knowledge by modifying the system of examination so as to have a greater focus upon practice and performance. Finally, in clinical education in most countries, there are now moves toward systems of which Huxley would have approved, where examinations of knowledge and clinical skills are retained but now they are part of an overarching assessment framework that has extensive workplace-based assessment (WPBA) of performance. The WPBA involves an evaluation of a doctor's progress in his/her performance over time, in hospitals or other clinical settings, in those areas of professional practice best tested in the workplace. WPBA in this context works well because it offers the opportunity to reconnect teaching, learning, and assessment and is authentic. It provides an opportunity for the assessment to get as close as possible to the real situations in which doctors work. WPBA has also led to a softening of the distinction between formative and summative assessment, with trainee doctors being provided with continuing feedback on their performance. WPBA therefore provides feedback to individuals on areas of strength and development needs; identifies trainees who require additional support; drives learning in important areas of developing expertise; and determines fitness of individuals to progress to the next stage of their careers. The UK Postgraduate Medical Education and Training Board (PMETB, 2004) has published the principles underpinning the approach and Schuwirth and van der Vleuten (2006) have highlighted the challenges that this provides for educationalists to increase the effectiveness of clinical education. In order to assemble evidence of performance, increasing use is made of portfolios, mini-clinical evaluation exercises, and 360° feedback. In a mini-clinical evaluation exercise, a consultation is observed and scored on a generic rating scale including items such as problem analysis, history

taking, and organization and efficiency. In 360° feedback, the candidate asks colleagues to complete a questionnaire on his or her performance that rates technical skills, interpersonal skills, team skills, education and research skills, etc. These instruments focus on observable behavior and can be used to provide feedback to trainees, but it is also clear that rather than using a single instrument to provide an assessment of a particular competence, it is better if the whole picture of someone's medical competence is portrayed using a variety of assessment instruments (Schuwirth and van der Vleuten, 2006>). Clinical education and training therefore provides a powerful example of the value of WPBA in the assessment and development of performance in initial education and training.

Assessment of the Development of Expertise in Knowledge-Intensive Workplaces

Another area where WPBA has become increasingly important is in the assessment of the development of expertise in knowledge-intensive workplace such as aerospace companies. In such contexts, WPBA faces two particular challenges. First, a focus on competence in the workplace in the sense of outlining what workers did in the recent past will be an insufficient basis for preparation for future performance. Second, much learning takes place while working, rather than in recognizable education or training settings. In knowledge-intensive work settings, such as aerospace, it is important to look beyond current competences as a basis for development, and instead take a developmental view of expertise. Such an approach requires that the development of expertise should itself be viewed as a continuing process. Thus, even if employees are able to produce competent performance in a range of more-or-less challenging work settings, there has to be a facility within teams or the workforce as a whole to go beyond this. From this perspective, it is interesting that some companies are explicitly using a developmental view of expertise that goes well beyond expecting technical proficiency and a commitment to continuing improvement. Thus, some companies, working in technologically advanced sectors, build up competence inventories of their staff in order to differentiate between:

- those who are technically able to perform a task but have very limited practical experience of actually doing so (e.g., the company could use them in an emergency or, if necessary, for a one-off activity);
- those who have successfully performed the task on a small number of occasions (e.g., the company could use them if the intention was to develop their expertise further; in a support role or if time is not necessarily a key criterion);

- those who have performed the task many times and under a variety of conditions (i.e., experienced worker standard – they are completely reliable);
- those who have substantial experience but are also able to support the learning of others (i.e., they can perform a coaching or mentoring role); and
- those who are world class, that is they are able to think through and, if necessary, bring about changes in the ways that tasks are tackled (e.g., could be chosen as a team leader for performance-improvement activities).

The interesting thing here is that this approach to professional development recognizes the importance of having a capacity to support the learning of others as well a capacity to change the way things are done. Now, WPBA has traditionally focused on the first three levels with a clear focus upon how workers perform the tasks being assessed. If the assessment is broadened to cover aspects of interaction with others, both on task and at other times, it might also be possible to pick up those who are able to support the learning of others, although in Germany such assessments have long been part of the *Meister* examinations for skilled workers seeking development and the attainment of higher-level qualifications. The highest level of performance can be recognized retrospectively (and the examples documented), but it is very difficult to predict through other forms of assessment. This difficulty is because this type of expertise is often partly built around recognition of the importance of the integration of different kinds of knowledge. Professionals and other highly skilled workers often find that the most important workplace tasks and problems require the integrated use of several different kinds of knowledge. Eraut (2004)> argues that this process typically involves five interrelated stages:

- the extraction of potentially relevant knowledge from the context(s) of its acquisition and previous use;
- understanding the new situation, a process that often depends on informal social learning;
- recognizing what knowledge and skills are relevant;
- transforming them to fit the new situation; and
- integrating them with other knowledge and skills in order to think/act/communicate in the new situation.

The whole process is much more complicated than just de-situating and re-situating particular pieces of knowledge. However, trying to predict in advance whether someone will be able to combine knowledge from education, training, work, and possibly other settings is very challenging, but WPBA in the form of recording of instances of such performances is authentic, fit for purpose, and signals that the overriding concern is with the holistic nature of performance in the workplace, if organizational performance is to be improved.

Competence-Based Assessment in the Workplace

Assessment in the workplace in the two contexts illustrated above is strongly situated and generally works well, but can these lessons be applied to the development of national competence-based systems where the picture is much more mixed. Assessment and learning in the workplace are not only related to each other, they are also linked to qualifications. Indeed national competence-based assessment regimes explicitly seek to link workplace assessments more toward qualifications than learning, in that they are expressly concerned with outcomes and downplay the significance of how things are learned. Such qualifications are assessment driven and the link to individual development is through achievement of qualifications, including progression to the next level in the qualification framework. These systems are by definition, national in outlook and do not necessarily engage with individual development in the workplace or improved organizational performance; nor do they necessarily encourage individuals to continue learning in the workplace. Indeed, Grugulis (2000) shows how compiling evidence of achievements at work against detailed performance criteria for competence-based qualifications, such as national vocational qualifications in England, can actually be antithetical to learning and development, because so much time was spent on the bureaucratic requirements of assembling evidence of existing competences.

Furthermore, competence-based work-related qualifications sometimes portray learning progression as an individual going through a single hierarchy of levels, whereas, in practice, an individual's learning requirements are likely to vary between domains and contexts and an expert in one area may be a novice in another, even if the fields are closely related. Moreover, individuals have to learn to make judgments in different spheres (e.g., in academic, cognitive, managerial, interpersonal, and experiential contexts) and again an individual may be at very different levels in the different spheres. Grugulis *et al.* (2004)> pointed out that the major failing of narrow views of competence was that they had an almost exclusive concern with measurable outcomes. Through qualifications, competences were seen as a proxy to measure the increase in skills of a population.

National competence-based work-related qualifications typically are drawn up according to similar templates that focus upon immediate achievements and measurable outcomes in the workplace. The qualifications supply information primarily about a person's skills and the knowledge necessary for carrying out the tasks and functions associated with particular jobs: that is, the assessment is concerned with the skills and knowledge necessary to underpin current performance. Such an approach says nothing at all about a person's potential or

his/her broader skills, knowledge, and understanding that may not be needed in his/her current work but that could be utilized in future. Assessment in the workplace, however, does not have to focus narrowly upon immediate achievements and outcomes, as it is possible for even competence-based assessment to reflect a developmental view of performance in the workplace that acknowledges the value of promoting continuing learning. For example, more sophisticated notions of competence have been developed that acknowledge that competence can be viewed as being held collectively by, for example, a workgroup and that there can be considerable value in competence development being contextualized for particular work environments (Mills *et al.*, 2000; Sandberg, 2001>).

The narrow approach to competence-based assessment has a focus upon measurable outcomes and an essentially binary conception of competence: you have either reached the appropriate standard (typically defined as experienced worker standard) or not. However, it is far more beneficial in inducing a more expansive frame of mind toward learning among employees to have a developmental view of competence as demonstrated by the example quoted earlier of companies working in knowledge-intensive industries. A developmental view of competence reflects attitudes toward learning and development that go well beyond achievement of simple technical or behavioral competence. That such orientations are not amenable to reductionist, analytical approaches to competence development is clear from the work of Sandberg (2001)> who shows that there are very different ways in which highly skilled professionals can approach their work: competence needs to focus upon higher levels of aggregation than just effective task completion. The increasing need for employees with proactive approaches to learning and knowledge transformation makes this requirement ever more pressing.

Formative Assessment and Critical Reflection in the Workplace Supporting Employees' Commitment to Lifelong Learning

One common method of professional development, used in a wide variety of occupational and organizational contexts such as healthcare, is to get employees to make an assessment or critical reflection on their own learning and development through, for example, the use of personal development portfolios or discussions with peers. The use of reflection upon learning at work can be a powerful formative assessment process, especially if it has a strong dialogical component, and can highlight not only what was learned but also the nature of the learning process. One outcome of a reflective formative assessment process might be the production of a portfolio that couples

hard achievements with development of soft processes that provide examples of experience and achievement. This reflective process can provide a spur to lifelong learning because it recognizes the emotional component to learning; it aligns with the idea of continuing personal development (as the examples can be updated); and it acknowledges the interrelatedness of changes in work practice and organization, personal development, and organizational performance. Williams (2001)> highlights how when learning is mainly coming through working rather than formal education and training, then the development of self-consciousness (reflection) and continual self-critique (critical reflection) are crucial to continued competence. Such workplace-based self-assessment, in this case in nursing, could be facilitated either through active repeated guided practice or could be an outcome of earlier education and training that was focused on problem-based learning.

This type of assessment does not have to remain formative. At any stage of the process, it should be possible to authenticate any particularly striking examples of experience and achievement through formalizing the process, such that it could form the basis of a later claim for credit against vocational or academic qualifications at a future date. This type of approach to assessment is primarily, rather than exclusively, formative, with the goal being encouraging learners to engage in further learning and skill development, and the use of mediating artifacts (portfolios, diaries, records, photographs, etc.) can encourage active reflection and review. The assessment process thereby adds value to the learning process in recognizing the achievements of the learner and the context in which the learning took place. One key aspect of this approach to assessment is that it increases the likelihood that learners would value the process and outcomes of assessment as supportive of their learning and further development.

Workplace-Based Assessment Supporting Processes of Knowledge Transformation

One of the key characteristics of a move toward a more knowledge-based society is that processes of knowledge development and transformation of different forms of knowledge become much more ubiquitous in where they are situated. In particular, organizations and workplaces are becoming important sites of knowledge generation and transformation, and some forms of WPBA have been designed to support such processes. The underlying idea is that there is considerable value in attempting to link processes of knowledge creation with approaches to tackling the core problems of manufacturing practice or service delivery as a means of engaging employees as

learners, including those employees in contexts such as small companies that have traditionally been difficult for formal education and training institutions to reach (Brown, 2005>). These ideas link to processes of continuing improvement in that attention is given to problems and dilemmas that are central to manufacturing practice or service delivery and performance improvement, although there is also express concern for work-related learning, development, and assessment. These problems and dilemmas have significance both for individual and organizational performance. The problems are likely to contain combinations of practical concerns, organizational issues, and sociocultural problems.

Processes of continuing development and process improvement require not only individuals to be reflective about their own learning but also for the networks involved in these development processes to be able to move knowledge and understanding between the individual and collective, the tacit and the explicit, and assessment in the workplace can play a role in these transformations. While a focus on the core problems of practice (and projected performance improvements in quality, cost and delivery) can act as a strong catalyst to galvanize the interest of companies and individuals, a vehicle is also needed to broaden the interest of companies and participants in both learning and organizational effectiveness (Brown *et al.*, 2004>). The approach to learning through networking could be seen as an example of an active model of learning whereby learners are engaged in processes of self and peer assessment and reflection leading to the creation of new contextualized knowledge, not recipients of a largely passive process of knowledge transmission (compare the processes of organizational knowledge creation outlined by Nonaka and Takeuchi, 1995>). This approach makes use of a social model of knowledge creation and transformation, where for genuine knowledge transformation to occur, knowledge has to move from the individual level into wider communities of interaction that cross organizational boundaries. This approach involves the spiraling of knowledge creation and transformation through continuing cycles of socialization, externalization, combination, and internalization. Processes of formative assessment are critical in order to allow ideas to be fed into the immediate change processes and to form the basis of continuing reflection on learning in the workplace over time that can form the basis of a step change in how organizations think about and support processes of learning, development, and knowledge transformation as a foundation for both individual learning and organizational development. This approach, when it works well, possesses the dynamism continually to create new knowledge, fueled by processes of assessment, reflection, and development.

The assessment processes used to power the knowledge development cycle include critical reflection, with

a focus on adaptability and forward thinking, and learning portfolios allowing individual and collaborative reflection on learning and knowledge-transformation processes. The portfolios could help employees pull their learning together; provide supporting evidence for use in company appraisal processes; help learning become more shareable, portable, and transferable; and act as a stimulus to innovation and the development of adaptability of the team as a whole, evidenced by the ability to perform effectively in a range of contexts (Brown *et al.*, 2004>).

Assessment of Workplace Learning Supporting Lifelong Learning and Active Knowledge Transformation

Exemplary forms of WPBA as outlined here are already supporting processes of lifelong learning and active knowledge transformation at an organizational level. At the national level, however, such approaches will need to be underpinned by a developmental view of competence. This change would imply that rather than the focus being on individuals being viewed as competent to perform current tasks at a particular level, it would recognize that people could still develop in a number of ways (at a range of levels) in order to improve their own performance, that of a team, or enhance the effectiveness of the organization. The forms of WPBA of engaging in peer assessment and self-reflection in order continue to develop a range of skills and to have a broad conception of expertise would also seem to offer, at a societal level, some possibility of moving toward a knowledge-based society. In such a society, there would be a need to focus more upon supporting the processes of formative assessment, learning and development, and to adopt a more expansive view of the nature of skills, knowledge, and competence development. This more expansive view could pay particular attention to the need to address issues of transfer of skills, knowledge, and experience between different settings; how to support individuals in developing a frame of mind whereby they continually look to improve their own performance through learning and development and to support the learning and development of others; and to recognize that in any organization a commitment to continuing growth and development of its members is strategically important.

The narrow view of competence-based assessment, by contrast, could be seen as the societal equivalent of what, at an organizational level, Argyris (2004)> called skilled incompetence, where the focus on doing current activities well can result in neglect of professional growth and longer-term development. A drift toward skilled incompetence could be challenged, however, by the use of processes of formative assessment and critical reflection within a more developmental view of how competence

and expertise can be fostered. This broader view could also help deal with a second problem: in many occupations, the types of knowledge developed through education and work differ, and either is often insufficient as it is the combination and integration of different types of knowledge that is often the major challenge. From this perspective, processes of formative assessment and critical reflection in the workplace could play a key role in the immediate post-qualifying period by recognizing that this is a time in which a great deal of learning takes place and support offered to individuals for their learning and development could have significance for establishing themselves in their career. People early in their careers learn a great deal from challenges at work, and provided that they receive support as required to facilitate processes of formative assessment and critical reflection, a virtuous circle of confidence, support, and challenge can be created (Eraut *et al.*, 2004>).

There is also value in building a stronger dialogical element into the assessment of work-related competences, especially where there is a focus of work-related learning upon the core problems of practice (Onstenk and Brown, 2002>). Learning from their own experience is important for the newly qualified, but so is learning from the experience of others. Newly qualified staff need opportunities to discuss and practice thinking about complex cases handled by their more experienced colleagues. The approach seeking to tackle complexity through processes of formative assessment and critical reflection puts interpretation and a shared search for understanding at the heart of “the discursive nature of professional practice” (Webb, 1996>: 111).

Assessment of workplace learning also needs to help meet the challenge of coping with the demands for flexibility, adaptability, and the ability to transfer skills, knowledge, and understanding between contexts in the workplace, particularly for those operating in highly skilled or professional contexts. The hallmark of successful occupational practice is the ability to draw on knowledge, abilities, skills, and attitudes used in an integrated, holistic way (Gonczi, 1994>). This approach to the performance of professional practice draws attention to three important features. First, complex professional duties can be performed in a variety of ways. Second, these duties can draw on different combinations of knowledge, skills, abilities, and attitudes in effective performance. Third, this approach implies that there is scope for professional judgment, not least in the ability to balance competing demands and the pressures of time. This means that individuals may come up with very different ways of responding to the demands for flexibility and transferability in the workplace. Indeed, one way forward for the newly qualified and experienced practitioners alike may be to use processes of formative assessment and critical reflection to review the different ways individual

practitioners seek to tackle their workload as a whole. By this means, it should be possible to discuss and share ideas about the most effective ways to tackle a range of problems in practice.

Summary

There is increasing evidence drawn from practice in a number of settings of how workplace assessment has affected employees' willingness to engage in lifelong learning and/or has supported processes of knowledge transformation in the workplace. From these particular examples, it is possible to put forward the design characteristics of an approach to the assessment of workplace learning that is oriented toward the future in that it focuses upon how best to support employees' commitments to lifelong learning and active knowledge transformation. It is possible to devise pedagogically driven assessment regimes that support employee development and improvements in organizational performance in ways that are meaningful to the learner and encourage further learning, thereby facilitating commitments to both lifelong learning and active knowledge transformation. The cornerstones of such an approach would acknowledge:

- the importance of formative assessment;
- focus upon individual performance in the enterprise or wider labor market rather than being almost exclusively concerned with seeking alignment with formal education and training programs and qualifications;
- build a stronger dialogical element into the assessment of work-related competences with the focus of work-related learning and assessment being upon the core problems of occupational practice;
- value in acknowledging the importance of soft qualifications in strengthening individual commitments to undertake continuing vocational training; and
- draw on the experience of organized programs of learning and development for organizational change agents that focus both upon organizational development and the value of continuing learning and development.

See also: Assessment in Vocational Education; Formative Assessment.

Bibliography

- Argyris, C. (2004). *Reasons and Rationalisations. The Limits of Organisational Knowledge*. Oxford: Oxford University Press.
- Brown, A. (ed.) (2005). *Learning While Working in Small Companies: Comparative Analysis of Experiences Drawn from England, Germany, Greece, Italy, Portugal and Spain*, SKOPE Monograph No 7., ESRC funded Centre on Skills, Knowledge and Organisational Performance, Oxford and Warwick Universities. <http://www.skope.ox.ac.uk/WorkingPapers/SKOPEMON07.PDF>.
- Brown, A., Rhodes, E., and Carter, R. (2004). Supporting learning in advanced supply systems in the automotive and aerospace industries. In Rainbird, H., Fuller, A., and Munro, A. (eds.) *Workplace Learning in Context*, pp 166–182. London: Routledge.
- Eraut, M. (2004). Transfer of Knowledge between Education and Workplace Settings. In Rainbird, H., Fuller, A., and Munro, A. (eds.) *Workplace Learning in Context*, pp 201–221. London: Routledge.
- Eraut, M., Maillardet, F., Miller, C., et al. (2004). *Learning in the Professional Workplace: Relationships between Learning Factors and Contextual Factors*. AERA Conference Paper, San Diego. http://www.tlrp.org/dspace/retrieve/93/AERA_2004.pdf (accessed July 2009).
- Gonczi, A. (1994). Competence-based assessment in the professions in Australia. *Assessment in Education* 1(1), 27–44.
- Grugulis, I. (2000). The management NVQ: A critique of the myth of relevance. *Journal of Vocational Education and Training* 52(1), 79–99.
- Grugulis, I., Warhurst, C., and Keep, E. (2004). Whatever happened to skill? In Warhurst, C., Grugulis, I., and Keep, E. (eds.) *The Skills that Matter*. Basingstoke: Palgrave MacMillan.
- Mills, T., Tyson, S., and Finn, R. (2000). The development of a generic team competency model. *Journal of Competency and Emotional Intelligence* 7(4), 37–41.
- Nonaka, I. and Takeuchi, H. (1995). *The Knowledge Creating Company. How Japanese Companies Create the Dynamics of Innovation*. Oxford: Oxford University Press.
- Onstenk, J. and Brown, A. (2002). A Dutch approach to promoting key qualifications: Reflections on 'core problems' as a support for curriculum development. In Kämäräinen, P., Attwell, G., and Brown, A. (eds.) *Transformation of Learning in Education and Training: Key Qualifications Revisited*. Cedefop Reference series 37. Luxembourg: Office for Official Publications of the European Communities.
- PMETB (2004). *Principles for an Assessment System for Postgraduate Medical Training*. London: Postgraduate Medical Education and Training Board.
- Sandberg, J. (2001). Understanding competence at work. *Harvard Business Review* 79(3), 24–28.
- Schuwirth, L. and van der Vleuten, C. (2006). Challenges for educationalists. *British Medical Journal* 333, 544–546.
- Williams, B. (2001). Developing critical reflection for professional practice through problem-based learning. *Journal of Advanced Nursing* 34(1), 27–34.
- Webb, G. (1996). *Understanding Staff Development*. Buckingham: Society for Research in Higher Education and the Open University Press.

Further Reading

- Ashton, D. and Sung, J. (2002). *Supporting Workplace Learning for High Performance Working*. Geneva: ILO.
- Billett, S. (2001). *Learning in the Workplace: Strategies for Effective Practice*. Sydney: Allen and Unwin.
- Boreham, N. and Moragn, C. (2004). A socio-cultural analysis of organisational learning. *Oxford Review of Education* 30(3), 307–325.
- Coffield, F. (ed.) (2000). *The Necessity of Informal Learning*. Bristol: Policy Press.
- Eraut, M. (1994). *Developing Professional Knowledge and Competence*. London: Falmer.
- Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology* 70, 113–136.
- Huxley, T. (1893). Address on university education (1876). *Science and Education: Collected Essays Volume 3*. reproduced as a Project Gutenberg eBook (2004). <http://www.gutenberg.org/dirs/etext04/8scd10h.htm> (accessed July 2009).
- Kämäräinen, P., Attwell, G., and Brown A. (eds.) *Transformation of learning in education and training: Key qualifications revisited*. Cedefop Reference series 37. Luxembourg: Office for Official Publications of the European Communities.

- Mott, V. (2000). The development of professional expertise in the workplace. *New Directions for Adult and Continuing Education* 86, 23–31.
- Oates, T. (2004). The role of outcomes-based national qualifications in the development of an effective vocational education and training system: The case of England and Wales. *Policy Futures in Education* 2(1), 53–71.
- Rainbird, H., Fuller, A., and Munro, A. (eds.) (2004). *Workplace Learning in Context*. London: Routledge.
- Schön, D. (1987). *Educating the Reflective Practitioner*. San Francisco, CA: Jossey-Bass.
- Stasz, C., Hayward, G., Oh, S., and Wright, S. (2004). *Outcomes and Processes in Vocational Learning: A Review of the Literature*.

London: Learning and Skills Research Centre. <http://www.lsda.org.uk/files/PDF/1689.pdf> (accessed July 2009).

Wolf, A. (1995). *Competence-Based Assessment*. Buckingham: Open University Press.

Relevant Websites

<http://www2.warwick.ac.uk/fac/soc/ier/glacier/learning> – Learning at Work.

<http://www.tlrp.org> – Teaching and Learning Research Programme.

Assessment in Vocational Education

K Ecclestone, Oxford Brookes University, Oxford, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The ways in which definitions of outcomes, standards, and curriculum content for vocational education relate to the content and criteria of assessment are affected by competing purposes for vocational education in relation to academic or general education. Some countries have particular difficulty in creating parity of esteem or status between vocational and academic education. Although the status of vocational education in relation to academic education varies between countries, questions about what differentiates general and vocational education, and which young people are suitable for both or either options affect all systems to a greater or lesser extent.

These underlying problems affect assessment approaches and methods used for summative purposes in vocational education which has to show credible standards of achievement in comparison to academic standards, while also opening up and widening opportunities for young people to participate in meaningful education and to progress either to workplace learning or further and higher education. Assessment in vocational education is therefore more complex than in workplace training or general education. This section cannot do justice to the complex theoretical and political debates about what standards of assessment are in academic and vocational education but simply note at this point that such debates affect assessment systems in different ways (see Goldstein and Heath, 2000).

While the particular problems of repeated reform and confusion about the relationship between academic and vocational education are more starkly evident in the UK than in other countries, many of its characteristics apply to assessment systems in other countries, particularly Australia and parts of Europe. This article draws on the features of vocational education in the British education system to summarize the main purposes for a general vocational education pathway, to describe typical assessment methods for formative and summative purposes, and to evaluate some of the links between goals for assessment, learning identities, and uncertainty about what vocational education comprises.

Although this section draws mainly on the UK's vocational education system, many of the features it describes and the questions it raises about goals, identity, and status of vocational education are pertinent to other countries too, albeit in different ways.

General Vocational Education

In the UK, general vocational education has evolved after 30 years of repeated attempts to create a coherent alternative to general academic education into a clear pathway. This runs alongside general academic education, workplace training, and vocational education and training courses that prepare people for different roles within a broad occupational area. General vocational education is now (since 2008) an option for young people at age 14 years before leaving compulsory education, an option after compulsory schooling at age 16 years as an alternative to general academic education or a work-based apprenticeship.

The roots of the general vocational education pathway are in earlier programs for prevocational education that enabled young people deemed to be unsuitable or unwilling to do general education to explore very broadly defined areas such as leisure and tourism, health and social care, or business studies, and to do practical assignments and projects rather than examinations as an alternative to general education.

At the ages of 16 and 18 years, a general vocational education qualification enables young people to progress to more focused vocational education and training courses in postcompulsory education (in Britain, further education colleges and universities, in Finland vocational colleges and polytechnics) that prepare them for occupational roles in a clearly defined area, such as media and performing arts, catering, nursery nursing, and engineering (among others).

One problem in creating stable assessment systems for vocational education that is well-understood by teachers, admission tutors in universities, and employers is confusion created by 30 years of reform and changes to the vocational routes available, and to their assessment systems and methods. In contrast, most Western European countries have managed to create a coherent, long-standing vocational education and training pathway that prepares young people for occupational roles at different levels, from leaving compulsory schooling at 16 years, to workplace apprenticeships and/or to higher level qualifications in polytechnics.

While parity of esteem between vocational and academic qualifications is often an underlying question in many countries, good levels of resourcing, including

highly qualified teachers, and continuity in policy have enabled countries like Finland to have clearly differentiated, well-respected pathways for vocational education and training courses that take place in a college and a local workplace, work, or general academic education.

Purposes

In a context of changing employment patterns, the decline of unskilled jobs (or home populations unwilling to do them), and social change in many countries which sees growing numbers of young people disaffected from compulsory schooling, vocational education and its assessment methods have to meet a growing number of sometimes competing aims:

- motivating learners who would otherwise not stay on in postcompulsory education or who are disaffected at school, by responding to, and rewarding, learners' expressed interests and notions of relevance;
- expanding routes into higher education while also making sure that expansion does not lead to oversubscription for limited places;
- preparing students for progression into work- and job-related training and assessment;
- encouraging young people to carry on gaining qualifications;
- convincing young people, teachers, admissions tutors, and employers that vocational education has parity of esteem with academic qualifications;
- ameliorating poor levels of achievement in numeracy and literacy through key/basic/functional skills;
- satisfying demands from different constituencies, such as employers, subject, and professional associations, etc., to include essential content and skills; and
- having credibility in compulsory schooling where there is less vocational expertise among teachers than there is in vocational/technical colleges in postcompulsory sectors.

Assessment Methods

Shifts from Norm-Referenced to Criterion-Referenced Assessment

The current system of assessment for vocational education in the UK has evolved over the past 50 years to offer diverse forms of certification and these derive from strong norm-referenced, graded systems, strong criterion-referenced systems, open-ended records of achievement, and non-graded competence-based assessment.

Since the late 1970s, the educational and social purposes of assessment have been to encourage wider access and participation and to motivate and engage those disaffected

from schooling and traditional academic assessment methods. A corresponding impetus is for assessment systems to recognize and certificate a much broader range of life and personal skills than in the past and to engage people with their learning in deep rather than instrumental ways.

Such goals have shifted strong norm-referenced systems developed in the 1950s to strong and weak forms of criterion referencing (see Baird *et al.* 2000). In theory, criterion-referenced systems can measure a wider range of real-life skills and attributes while enabling people to get the grade they deserve, providing they meet the publicly defined criteria. Such aims have dominated the strong criterion-referenced system of competence-based workplace qualifications. In criterion-referenced systems, standards come from raising levels of competence and performance in relation to well-defined specifications, thereby providing more valid, authentic measures of performance and coverage or mastery of performance that can be specified in the criteria. Formative assessment, feedback, setting, and reviewing targets not only enable more students to reach the required standard but are seen to be inherently democratic and motivating (see Jessup, 1991).

A Hybrid Assessment Model

Increased participation in both general academic and vocational pathways has eroded old beliefs that any cohort contains a limited pool of innate ability that can be measured reliably as the basis for selection through competitive examinations and norm-referenced grading. In this approach, failure is an inevitable adjunct to success. Although norm-referencing still operates in contexts where summative assessments must select people for licenses to practice or limited jobs and places for higher education, the expansion of work-based assessment and general vocational education has moved all the UK's assessment systems toward a hybrid model, with varying degrees of strong or weak criterion referencing. A powerful influence is the rise and dominance of competence-based assessment (see Jessup, 1991).

Assessment in general vocational education reflects the shift in many countries towards weak criterion-referencing, with the following features:

- the idea that specifications and criteria can be defined in more detail as an impetus for higher achievement because it enables better formative assessment, transparency of assessment demands for students, shared understanding of the criteria, and standardization of assessment decisions;
- strong emphasis on teacher assessment and locally designed assessments against predefined outcomes, with a focus on initial diagnostic assessment, individual target setting based on the criteria, recording achievement, and portfolio building;

- external assessment and moderation by an external authority such as an awarding body, or other national/state agency, and standardization of grading which combines criterion referencing and loose notions of rank-ordering; and
- a contrast to demotivating images of assessment that compare students with each other in general academic education through an upbeat focus on achievement and opportunities to succeed.

Methods of assessment in vocational education combine approaches used in academic education, practical subjects, and workplace training. There are broadly three types of assessment tasks in a vocational education course:

- summative tests such as an end of unit/module multiple-choice test on underpinning theory or knowledge or a written examination paper requiring short answers;
- compilation of a portfolio of evidence that enables students to accumulate different pieces of work (these might include testimonials from work experience, evidence of tasks carried out during work experience, photographs, audio and film recordings, and written work) to show competence or achievement against a set of predefined outcomes; and
- coursework summative assessment based on discrete modules or units, where students complete group and/or individual projects or assignments that have practical and written elements and come in a variety of sizes in terms of credit value. For example, students on a health and social care course might plan and run a health-awareness day for fellow students or for a school that involves research into different topics, presentations through posters, talks and written work, interviews with local health practitioners, and a written evaluation of how well they did in the project.

Many assignments can be locally devised and teacher assessed, provided they address the specifications of outcomes and criteria. There may also be nationally set assignments by the awarding or accrediting body, marked by teachers locally and standardized or moderated regionally or nationally.

Grading criteria and detailed descriptors are based on skills in learning to learn, teamwork, self-assessment, and communication as well as on evidence of subject understanding. A final summative qualification mark is usually derived from a grade-point average from the modules completed during the course, with some weighting given to a final project.

Formative and Summative Activities and Processes

In the UK's vocational education system, many tutors regard their main assessment role as a translator of official criteria and specifications which are usually detailed and

prescriptive. It is commonplace for tutors to break up strongly framed assignment briefs into sequential tasks to meet each criterion. However, at higher levels of the vocational pathway, this becomes less prescriptive and more open ended. Students prepare their assignments by working to the official criteria specified for grades in each unit. In many courses, students can submit a completed draft for feedback which tells them how they can improve their responses to different criteria. However, there is wide variation in formal arrangements for this: in some courses, drafting is done numerous times while in others, only one opportunity is offered. A large amount of lesson or contact time is used to introduce students to each assignment and also to talk through the outcomes of draft assignments, and to allow students to work on assignments.

Formative feedback often takes the form of written advice to plug and cover gaps in the criteria, cross-referenced to the assessment specifications. Vocational students have strong expectations that teachers would offer advice and guidance to improve their work.

Portfolios and assignments or projects provide opportunities for both formative and summative assessment. Reviews of progress and formative feedback can be based on a discussion of the portfolio of achievement, while unit/module-based assignments can be improved through feedback on a draft, or a number of drafts in order to pass or gain better summative marks.

Different countries place different emphases on formative or summative purposes. For example, the two purposes are blurred in the UK, with a strong emphasis from awarding bodies (regulatory examination bodies) on standardization of summative judgments between different centers. In Finland, vocational teachers working with workplace assessors are accustomed to using skills tests and other projects for formative purposes and less accustomed to national regulation or standardization. However, a project by the National Board of Education in Finland aims to use the outcomes of summative tests for national evaluation purposes (see Rakkolainen and Ecclestone, 2005).

In the UK's system, summative assessment, achievement and learning have become to a large extent synonymous, where assessment is not merely for learning or of learning: instead, it is learning:

The clearer the task of how to achieve a grade or award becomes, and the more detailed the assistance given by tutors, supervisors and assessors, the more likely the candidates are to succeed; but succeed at what? Transparency of objectives, coupled with extensive use of coaching and practice to help learners meet them, is in danger of removing the challenge of learning and reducing the quality and validity of outcomes achieved. . . . assessment procedures and practices come completely to dominate the learning experience, and 'criteria compliance' comes to replace 'learning' (Torrance *et al.* 2005: 46).

The Effects of Assessment on Attitudes to Learning in General Vocational Education

Discussion in this section draws on research studies on students' and teachers' attitudes and practices in relation to formative and summative assessment in the UK's system (see Ecclestone, 2002; Torrance *et al.* 2005; Ecclestone and Pryor, 2003; Ecclestone, 2004; Ecclestone, 2008). It summarizes some key features of the ways in which assessment affects students' and teachers' educational goals, attitudes to learning, and choices of pathway.

First, vocational teachers and students emphasize the importance of ipsative (self-referenced) achievement based on students' growing confidence and their ability to overcome previous fears and failures. They regard personal development as more important than the acquisition of skills or subject knowledge. In contrast, staff and students in general academic courses see achievement as predominantly about progression in subject-related knowledge and understanding and gaining suitable grades for progression to higher education.

Second, vocational students often have a strong sense of identity as second-chance learners, an image that their tutors empathize with from their own educational experience. Teachers' and students' attitudes to achievement and failure and their beliefs about students' preferences, capabilities, and attitudes both reflect and reinforce the ethos and culture of different qualification tracks. Attitudes about the purpose and ethos of a qualification and its assessment methods are reinforced by teachers' images of what students liked and wanted. Vocational tutors regard good assessment as practical, authentic, and relevant activities, work-experience, and field trips: there is often a strong view that these students do not want or like written assessment, that they are less secure, need more group affinity and be in a more protected, safe environment. In their beliefs about comfort zones and protecting students, vocational teachers see assessment as integral to a strong ethos of personal development that minimizes stress or pressure.

Third, assessment methods have had a strong impact on expectations for particular types of teaching and assessment activity, and for particular relationships between teachers and students. Many vocational education courses offer students a lively and relaxed atmosphere that enables group work, teacher input, and time to work on assignments individually or in small friendship-based groups. Vocational students also like the familiarity of continuing with teaching and assessment methods they had experienced in an intermediate vocational qualification in the final year of school.

Fourth, students have strong expectations of appropriate feedback, help, and support to enable them to get a good grade. A cultural understanding of assessment, used in the studies cited in this section, illuminates the ways in

which techniques such as feedback, explaining the criteria, encouraging student self-assessment, and using detailed grade descriptors are seen widely as good practice. Yet, in some assessment cultures they can encourage superficial compliance with tasks derived from the assessment criteria, bureaucratic forms of self-assessment against the criteria, and high expectations of coaching and support.

Finally, despite have a strong identity as vocational students, and despite a progression route from vocational courses to university, many young people following vocational courses, particularly at age 14 have erratic, uncertain, and vague aspirations and understanding of what this might mean for progression (see Davies and Biesta, 2006).

Uncertainty about Vocational Education

Thirty years of initiatives in the general vocational curriculum in the UK system have not clarified what vocational knowledge and skills really are. An emphasis on creating assessment methods to foster motivation, confidence, and personal development has led to growing uncertainty about what vocational means in the vocational education pathway. There is a strong view that assessment in vocational education should encompass relevant life skills and attributes associated with lifelong learning, personal development, basic and generic interpersonal and communication skills, and dispositions and attitudes for employability. Themes of relevance and real-life application are reinforced by concerns that assessment methods should engage and retain young people in formal education who are deemed to be demotivated and disengaged.

One effect of uncertainty about vocational knowledge and skills as opposed to generic ones, and particular images of types of young people and their attitudes to education and learning is a growing tendency to avoid burdening vocational students with too much written work or with methods that alienate them from formal education. A more recent phenomenon is to associate ideas that young people are disaffected or disengaged with images that they have fragile learning identities and low self-esteem (see Ecclestone and Hayes, 2008).

These images reflect and reinforce an enduring problem in the UK's education system, that vocational education and its assessment continue to be associated with practical activities and the need to motivate, rather than being a coherent curriculum in its own right. Assessment initiatives over the past 30 years have been a central factor in creating and maintaining these images (see also Hayward *et al.* 2005).

Finally, confusion is exacerbated by different constituencies and bodies being involved in deciding a vocational curriculum and assessment methods. In the UK, this is extraordinarily complicated with occupational

sector councils, employers' representatives, awarding bodies, the government's regulation agency, and educational providers all involved in processes to design vocational education qualifications. In contrast, countries such as Finland involve the key social democratic partners of unions, employers, and educationalists in a more stable and simple system.

Conclusions

General vocational education has become a crucial option for young people deemed to be unable or unwilling to undertake general academic education or who are not ready for workplace training. Assessment methods that are seen to motivate and engage young people, while also providing a credible basis for summative assessment that enables them to progress to further and higher education and to work, have evolved into a distinctive system. This system is criterion referenced and adapts methods from competence-based approaches developed for the workplace and approaches used in academic courses.

However, questions remain about the purposes of assessment in vocational education and the subject base and skills that are seen to comprise vocational'as opposed to practical or'relevant-to-life education. There are also questions about the status and identity of young people opting to follow a vocational education pathway and whether these reinforce particular attitudes to assessment and learning. Although these questions are particularly pertinent in the UK's system, readers need to relate them to the particular features of vocational education in other countries.

See also: Assessment in Schools – Dispositions; Assessment in Schools – Mathematics; Assessment in the Workplace of Performance, Developing Expertise and

Competence; Formative Assessment; Impact of Assessment on Students' Test Anxiety; Records of Achievement: Beyond Traditional Tests; Summative Assessment by Teachers.

Bibliography

- Baird, J., Cresswell, M., and Newton, P. (2000). Will the real gold standard please step forward? *Research Papers in Education* **15**, 213–229.
- Davies, J. and Biesta, G. (2006). Going to college or getting out of school? The experiences of vocational learning of 14–16 year olds in an FE college. *Research Papers in Education* **22**(1), 23–41.
- Ecclestone, K. (2002). *Learning Autonomy in Post-Compulsory Education: The Politics and Practice of Formative Assessment*. London: RoutledgeFalmer.
- Ecclestone, K. (2004). Learning in a comfort zone: Cultural and social capital in outcome-based assessment regimes. *Assessment in Education* **11**(1), 30–47.
- Ecclestone, K. (2008). *Transforming Formative Assessment in Post-Compulsory Education*. Buckingham: Open University Press.
- Ecclestone, K. and Hayes, D. (2008). *The Dangerous Rise of Therapeutic Education*. London: Routledge.
- Ecclestone, K. and Pryor, J. (2003). 'Learning careers' or 'assessment careers'? The impact of assessment systems on learning. *British Educational Research Journal* **29**, 471–488.
- Goldstein, H. and Heath, A. (eds.) (2000). *Educational Standards*. Oxford: Oxford University Press/British Academy.
- Hayward, G., Hodgson, A., Johnson, J., et al. (2005). *The Nuffield Review of 14–19 Education and Training: Annual Report 2004–2005*. Oxford: University of Oxford.
- Jessup, G. (1991). *Outcomes: NVQs and the Emerging Model of Education and Training*. London: Routledge.
- Torrance, H., Colley, H., Garratt, D., et al. (2005). *The Impact of Different Modes of Assessment on Achievement and Progress in the Learning and Skills Sector*. Learning and Skills Development Agency. <http://www.lsda.org.uk/cims/order.aspx?code=052284&src=XOWEB> (accessed June 2009).

Relevant Website

www.lsda.org.uk – Learning and Skills Development Agency, UK.

EDUCATIONAL ASSESSMENT – CONCEPTS AND ISSUES

Contents

Assessing Group Work

Assessment and the Regulation of Learning

Challenges of Developing and Implementing Formative Assessment Practices in Schools

Formative Assessment

Formative Assessment and Instructional Planning

Progression and Assessment: Developmental Assessment

Students' Rights in Assessment and Evaluation

Summative Assessment by Teachers

The Multiple Purposes of Assessment

The Relationship Between Assessment and the Organization and Practice of Teaching

Assessing Group Work

M Galton, University of Cambridge, Cambridge, UK

© 2010 Elsevier Ltd. All rights reserved.

The need to have pupils working in groups has become a universal panacea among policymakers for the problems of globalization in the workplace. In Singapore, despite its favorable positioning at the top of the league table of international studies, the Prime Minister's 2007 National Day speech urged schools to teach less so that children could learn more. In Hong Kong, since the millennium, the revised curriculum, learning to learn, has also stressed the need to move away from teacher-directed approaches toward greater pupil participation through the use of group work. The social constructivist viewpoint that talk drives learning has become an accepted part of current educational wisdom worldwide.

Yet, although studies in the West, where working in groups has a longer history, attest to the value of group working as an effective pedagogy, both in terms of improving pupil attainment and also attitudes, particularly racial ones, the fact remains that in many classrooms, group work is still a neglected art. While pupils often sit in groups, presumably for social reasons, they rarely work as a group. In the UK, research by Kutnick *et al.* (2007) has confirmed that teachers use groups relatively infrequently in subjects such as English, mathematics, and science. When group work was undertaken, it was often for practical reasons, as in the science laboratory where the determining factor was the availability of equipment. In other cases, the decision to move to group work was often arbitrary. For example, in one instance reported in Kutnick

et al. (2007) the rationale for changing from class to groups was "because children were becoming bored with the classwork and I thought they needed a change."

Assessment: One Possible Impediment to the Use of Group Work

When teachers were asked why they did not use group work more often, they presented a number of arguments. Pupils who were asked to work in groups usually wished to choose the composition. Since these choices tended to be based on friendship, groups were often unbalanced because their composition was strongly associated with the attainment level, gender, and ethnicity of the pupils. Teachers claimed that they were under too much pressure, because of the demands of the curriculum, to find the necessary time for setting up and carrying out group tasks (Kutnick *et al.*, 2007). Other practitioners had concerns about loss of control over the learning environment since it was difficult to be certain that the talk among pupils was productive, or even on-task. This contrasts sharply with the situation during whole-class discussion where pupils generally only speak when requested to do so by the teacher. Teachers also saw the use of mixed-ability groups as an additional source of difficulty, since it often meant placing a disruptive pupil among others who normally worked well together with likely consequences

of aggression or withdrawal. Many teachers said that they did not know what to do with the loners who withdrew from group activity completely (Kutnick *et al.*, 2007).

Another key reason for teachers' reluctance to use grouping to enhance academic performance concerns the requirement that to work effectively, pupils needed to be trained in communication skills (Mercer, 2000) and this took valuable time. A recent study of teachers' work-life balance in several countries has shown that workloads have increased, mainly because of the extra administrative burdens connected with the target-driven curriculum which requires that increased time be spent on assessment and therefore in marking (Galton and MacBeath, 2008). Many teachers thought that assessing the individual contributions to a group task was likely to be both problematic and time consuming (Kutnick *et al.*, 2007). For this reason alone, teachers were reluctant to promote group activity in their classrooms.

Johnson and Johnson (2004) argue that if teachers are to be persuaded to use group work more often for academic purposes, they need to be disabused of a number of misconceptions about the assessment process in general and its use for groups in particular. The first of these misconceptions, and perhaps the most powerful, is that an individual assessment always requires an individual performance since an assessment to be valid needs to measure unassisted learning. Hence, it follows that assessment data can only be collected once the pupil has completed the piece of work on which the marks are to be awarded. A further misconception is that it is only possible to have unbiased assessments if they are carried out by the teacher. From this viewpoint, therefore, asking pupils to assess each other, or assessing the contribution of individual pupils to a group product provides unreliable and invalid data. For teachers to accept that it is possible and feasible to assess group work, they need to be convinced that groups can have a powerful effect on the individual performance of its members. Thus group work can offer insights into the potential of some pupils that might not emerge when more conventional forms of testing are used.

What Aspects of Group Work Should Be Assessed?

There tends to be a division of opinion as to what aspects of group work should and should not be assessed. Most proponents of group work agree that it is important to distinguish between the assessment of group working skills and the assessment of what is learned during group work but not all agree with Dunne and Bennett (1994) that the pupils' cognitive achievements, both individual and joint constructions, should be included in any assessment. In designating what constitutes essential group work skills, Webb (1995)

puts the emphasis on interpersonal skills including the nature of the interactions with other group members (active listening, raising questions, making useful suggestions, explaining, etc.) and an individual's role in bringing about successful collaboration (acting as gatekeeper, as an organizer, as negotiator in attempts to arrive at a consensus view, etc.). Others emphasize various social processes such as degree of participation, the degree of empathy with other points of view, and sensitivity to the needs of others. Some writers argue that individual contributions should not be assessed as this weakens the extent to which certain pupils, particularly the more able ones, are willing to share their ideas, while others argue that if only the group's contribution as a whole is assessed, then social loafing will occur whereby some pupils will stand aside and let other group members do most of the work. For these reasons, a compromise position has been adopted by most advocates of group work whereby both the individual and the group contributions are assessed. Slavin (1995) reports that when both contributions are assessed, the improvement in pupils' performance produces medium-effect sizes of the order of 0.40 whereas when only one of the contributions is used, the effect sizes tend to be much smaller (typically 0.17).

Assessing Individual Contributions to Group Work

Perhaps the most straightforward method of assessing individual contributions to a group task involves various forms of team games tournament approaches as designed and used by Slavin (1995). The gist of the assessment procedure is that each individual takes a test before and after the group activity. The raw-score gains of each individual in the group are then added together to give a group score. Only by working together with the more-able pupil helping the slower learning ones can a group achieve a winning total. At the same time, the teacher can assess the improvements of individual team members by examining differences between pre- and posttest scores. Slavin argues that this method deals with the criticism that individual assessments tend to make more-able pupils less willing to share their knowledge with others. Slavin's approach has been modified by Race (2000) who suggests that the individual pupil's score should be weighted in some way by the group's score.

Race's recommendation is similar to that suggested by Johnson and Johnson (2004) who argue that an individual's grade should also reflect how well the group worked together. Johnson and Johnson suggest combining the individual score with the average group score or providing bonus points based on either the lowest individual score or the average improvement score. However, critics of the original Slavin approach, such as Jacobs *et al.*

2002 point to the fact that there is a need to distinguish between the products resulting from working in a group and the effectiveness of the group working skills which bring about these as a satisfactory joint outcome. These latter skills include understanding and implementing the task, making best use of available resources and the degree of appropriate help offered to others.

As a solution to this problem, Johnson and Johnson (2004) suggest the use of peer assessment whereby each group member rates the others on various group working skills such as listening carefully to other points of view, taking one's turn, offering reasons when making suggestions, and so forth. It is argued that this not only avoids potential teacher bias when basing their judgments on informal observations, but it can also be a learning process for the pupils since as the list of skills are added to over time, to include, for example, maintenance strategies, pupils gradually gain metacognitive understanding of what it is to be a good group member. A variation of this approach involves self-assessment with the pupil's own evaluation being compared with those of his or her peers. Other methods involve choosing a pupil to carry out assessments on the rest of the group using a simple category-observation system (Dunne and Bennett, 1994). These authors also offer a variety of alternative methods including post-task interviews, whole class debriefings, and video or audio recordings. Although the work involved can be considerable for teachers, Dunne and Bennett argue that it can provide them with important insights into pupil behavior. This, in turn, can have positive effects on their teaching and their expectations, since teachers may be surprised at the amount of work that some pupils can do unaided and the extent to which others turn out to be effective collaborators.

Assessing the Group's Contribution

When the group's contribution is to be assessed, a number of additional considerations need to be taken into account. In this connection, Webb (1995) points out that it is sometimes the case that the most often-cited reasons for using group work in the classroom do not match the choice of assessment method. This situation generally arises because only certain methods constitute a valid measure of a particular purpose of assessment, particularly when these purposes represent competing goals of learning from group work versus group productivity (Webb, 1995). It is therefore important in designing group assessments to appreciate that the manner in which groups are required to function will influence the outcomes of the assessment and these group processes need to be taken into consideration in the design of the assessment procedure.

Where the purpose of the assessment is to establish how much learning over and above the pupil's individual

competence has been constructed in collaboration with others, it is necessary to employ some pretest measure prior to the group activity. Webb (1995) cites an example in science, where pupils first had to list the various factors that influence yeast's activity in food. They then carried out an activity in groups which required them to design, carry out, interpret, and summarize the results of an experiment which investigated what happened when yeast was mixed with food. The summary was constructed by the group written up and then presented orally to the rest of the class. Then, in the third part of the exercise, pupils again worked individually to analyze and critique a summary written by another group. The difference in scores on the first task and the third task was then taken to represent the extent to which each pupil's learning was enhanced as the result of the group's collaborative activities. A group assessment can then be arrived at by adding the scores of individual members.

When, however, the purpose of the assessment is group productivity, then the effect on individual members is ignored and only the group outcome is considered. In this case, either the quantity or the quality of the product or a combination of both measures can be used in the assessment. However, it is important from the start that pupils understand the criteria that are being employed since an assessment based on the amount of the product (the number of ideas generated by the group) may result in different kinds of collaborative behavior than when the quality of these ideas is the main criterion. In the former case, where the quantity of the ideas matter, success does not depend on every group member understanding what is being proposed. Thus there may be fewer attempts by group members to justify their ideas or to give and receive elaborated explanation as an aid to understanding, whereas when quality is the criterion, then these kinds of interactions are an important part of the group process when attempting to arrive at a consensus about the most appropriate idea or the best solution to a problem.

A different kind of purpose of assessment concerns itself with the functioning of the group itself. Here social constructivist theories allied to that of Piaget's emphasis on promoting cognitive dissonance, as a way of enabling individuals to construct new schema and to reconstruct existing ones, provide a theoretical framework for the kinds of interactions that need to take place within the group if learning is to take place. These include higher-level cognitive behaviors such as making alternative suggestions, raising questions, offering explanations, and the social skills of being able to participate in the cut and thrust of debate by agreeing and disagreeing without provoking animosity on the part of other group members.

In all these different interpretations of what it means to undertake a group assessment, there is a need to ensure that the composition of the groups is such that, as far as it is feasible, no one group enjoys an advantage over another.

Johnson and Johnson (2004) summarize the research which shows that the ability, gender, ethnicity, and status of individual members influences the way in which the group performs. Lower-achieving pupils have been found to make greater progress when in heterogeneous groups, provided the more-able pupils have been trained to offer appropriate help, particularly in the matter of offering relevant explanations. The evidence concerning high achievers is less clear. Some studies show that it matters little whether the groups are homogeneous or heterogeneous, while others show that in the latter type of group, the more-able pupils tended to assume the role of the teacher and clarify their own ideas by having to explain them to others.

There is considerable evidence that boys tend to dominate the group conversations whether in a majority or a minority, although in the United States, this phenomenon is not so prevalent among African-American pupils. Social status is often associated with ethnicity in Western societies, where most of the research has been undertaken. Status can also be relative (Cohen, 1994) so that a pupil might participate actively in one group but not in another. It is generally recommended that teachers try to ensure, as far as possible, that the membership of the group is representative of the individual social and personal characteristics which define status in that particular classroom. In particular, there is a need to avoid placing single individuals of a particular status within a group such as a white male among a group of Afro-Caribbean pupils or of varying more than one status characteristic such as having a mix of ethnicity and gender. Thus, it is preferable to have two white males and two Afro-American females or vice versa rather than have one white and one Afro-American male (Miller and Harrington, 1993).

Procedures for Assessing Group Productivity

Race (2000) argues that assessing group tasks is one of the most difficult things to do well. He suggests that there is no ideal way to carry out this process. The main problem is that allocating a single group score can be perceived as unfair by those who do most of the work. For this reason, he suggests that all assessments should carry with it some form of additional differential. If, for example, a teacher gives a group of four pupils, a mark of 60%, then the group can be asked to partition this mark according to their estimate of individual contributions. These contributions can stand and be recorded separately or carry an added mark for particular contributions. While this approach may be perceived to be fairer, it requires a certain maturity on the part of the participants to obtain a consensus about who should get most marks. If there is some borderline level, such as an award of a credit for a mark above a specific score, then groups may distribute

the marks accordingly to gain the maximum advantage for its members. To avoid this possibility, Race (2000) recommends that the individual contributions should also be assessed by means of a short oral viva. Even here, there are problems since some pupils do not perform well under this form of assessment, either due to shyness or modesty, and there may be some advantage from being interviewed last since the questions asked become common knowledge and allow later participants to prepare for the interview. Moreover, in a class of 30 pupils, conducting the assessment adds enormously to the teacher's workload.

All commentators agreed, however, that pupils need to be informed, prior to the assessment, not only about what is to be assessed, but also the means by which the assessment is to be carried out. Where group processes are to be included, then the choice usually lies between the use of some simple observation checklist where the teacher follows different pupils from each group in a predetermined order or a self-evaluation sheet which pupils complete at the conclusion of the group activity. A refinement of this approach is to ask pupils to complete some form of learning diary or narrative log of group work which charts progress over time or to fill in trigger sheets at selected time intervals during the group activity. Most of the suggested observation schedules are watered-down versions of research instruments. They include categories such as whether the pupil was on task, whether he or she initiated or responded in a conversation with other group members, and whether the conversation was sustained. More detailed systems will also attempt to record the nature of the conversation; whether it involved making a suggestion, raising a question, offering an explanation, or expressing disagreement, etc. Other versions of this approach use videotape or audiotape to capture these pupil-pupil exchanges and then award a summative rating for each group on the various categories. As this form of assessment can be time consuming, it has been suggested that it is carried out by means of a one-off specially chosen assignment and that as far as possible, the task chosen as part of the group activity should be an authentic one which is designed to measure pupils' performance in real-life contexts (Crotty, 1994).

In creating criteria which allow pupils to evaluate their group's performance, it is necessary to cater for different levels of complexity which match the expected capabilities of the group. At elementary level, pupils might be required to report on whether members of the group took turns in talking and listening, helped other members to express their ideas, stayed on task, and so forth. In the middle and high school, the questions would more likely also to concern the planning process (whether deadlines were met), execution (the division of labor), degree of cooperation (all members contributed to decision making), participation (all took an equal part), and how the group responded to feedback from other groups and the teacher. Whatever the sophistication of the evaluation sheets,

however, it is essential that training in carrying out these assessments is provided in the provision of guidelines and exemplars. Not only does this ensure improved reliability, but it also conveys to the pupils what is required of them when working in groups. Through this self-evaluation process, pupils therefore gradually also gain a metacognitive understanding of what it is to be a good group person and this is likely to enhance future performance. Johnson and Johnson (2004) suggest that the assessment procedures work best when it includes both group products and processes. Furthermore, they report that when teacher and pupils share responsibility for the assessment, the improvement in performance was greater than when the processing was the sole responsibility of either the teacher or the pupils.

The growth in the use of information and communication technology in most schools has led to increased interest in the use of portfolio assessment described by Johnson and Johnson (2004) as an organized collection of group work samples accumulated over time, including individual work samples from each member, together with accounts of presentations and the subsequent critiques of teachers and peers. Electronic portfolios allow pupils to display evidence of both academic and social growth over time and when these are made publicly available, they can become a powerful tool for sharing learning with teachers and peers. To be effective, however, teachers need to instruct pupils in the best ways of planning and implementing their portfolios so that the criteria for good work are recognized and examples which articulate these qualities are selected. Johnson and Johnson (2004) also suggest that group assessment can be carried out through a series of conferences. At the first of these, the goal-setting conference, groups or the whole class set targets. At the second progress-assessment conference, groups assess the extent to which these targets are being met while at the final postevaluation conference, the groups discuss and decide their level of achievement. This, in turn, leads to the next goal-setting conference.

A Final Comment

The above review suggests that there are a number of principles which should guide the teacher or researcher when seeking to assess group work. First, a fair system should be used that rewards both individual effort and group collaboration. Second, the assessment of group productivity should involve both teachers and pupils in the undertaking. Third, in designing an assessment of group productivity, a distinction should be made between quantity and quality because this may well change the way in which group members approach the task. Fourth, a judgment also needs to be made between the purposes of the group work, whether the emphasis is to be on learning outcomes or on the group processes such as the kind's interactions that

theories and empirical evidence suggests enhance learning through cooperation. Fifth, the criteria to be used in the assessment must be made explicit to pupils beforehand, and training opportunities for practice and for review should be provided. Sixth, if the assessment is to be a one-off measure, akin to a summative test, it is better to design an activity which is as authentic as possible.

This is a formidable list of requirements so it is not surprising that much of the evidence which underpins these principles has been collected through compact, carefully designed research studies, including some that took place in the laboratory rather than in a classroom. So far, there are few attempts to collect evidence of the consequences of putting these principles into practice in naturalistic classroom settings, where teachers and pupils go about their business within the usual constraints which govern teaching and learning in a typical school.

It is partly because the task of assessing group work appears so complicated that teachers have, in the past, fought shy of using this teaching strategy, particularly in an era of targets and high-stakes testing where numbers are everything. It is for this reason that teachers who do favor the use of groups tend either to assess individual performance by written tests, or base their judgments on informal observations using a simple form of checklist, supplemented by a PowerPoint display to provide a summative evaluation of the group's productivity (Gillies, 2007). If the measures used to assess group work in naturalistic classroom situations are to become more reliable and result in increased validity, two things need to happen. First, teachers need to gain a better understanding of the impact of certain types of assessment on the kinds of behavior which pupils will exhibit when placed in groups. Second, researchers need to spend more time in classrooms working alongside teachers to make use of these various assessment procedures more suited to the ups and downs of a typical school day.

See also: Assessment Conversations: Reading and Writing Conferences with Students and with Parents; Peer and Self-assessment; Portfolio Assessment; Summative Assessment by Teachers.

Bibliography

- Cohen, E. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research* 64(1), 1–35.
- Crotty, E. (1994). The role of cooperative learning in an authentic performance assessment approach. *Social Science Record* 31(1), 38–42.
- Dunne, E. and Bennett, N. (1994). *Talking and Learning in Groups*. London: Routledge.
- Galton, M. and MacBeath, J. (2008). *Teachers under Pressure?* London: Sage.
- Gillies, R. (2007). *Cooperative Learning: Integrating Theory and Practice*. Thousand Oaks, CA: Sage.

- Jacobs, G., Power, M., and Loh, W. I. (2002). *The Teachers' Sourcebook for Cooperative Learning*. Thousand Oaks, CA: Corwin Press.
- Johnson, D. and Johnson, R. (2004). *Assessing Students in Groups: Promoting Group Responsibility and Individual Accountability*. Thousand Oaks, CA: Corwin Press.
- Kutnick, P., Hodgkinson, S., Sebba, J., et al. (2007). Pupil grouping strategies and practices at key stage 2 and 3. *Research Report 796*. Nottingham: DfES Publications.
- Mercer, N. (2000). *Words and Minds: How We Use Language to Think Together*. London: Routledge.
- Miller, N. and Harrington, J. (1993). Social categorization into groups: Principles for the design and development of cooperative learning teams. In Hertz-Lazarowitz, R. and Miller, N. (eds.) *Interaction in Cooperative Groups: The Theoretical Anatomy of Group Learning*, pp 203–227. New York: Cambridge University Press.
- Race, P. (2000). *Tips on Group Learning*. London: Kogan Page.
- Slavin, R. (1995). *Cooperative Learning: Theory, Research and Practice*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.
- Webb, N. (1995). Group collaboration in assessment: Multiple objectives processes and outcomes. *Educational Evaluation and Policy Analysis* 17, 239–261.
- Foster, M. (2005). Assessing group work. *Independent Education* 35(2), 13–14.
- Jacques, D. (2000). *Learning in Groups: A Handbook for Improving Group Work*. London: Kogan Paul.
- Jolliffe, W. (2007). *Cooperative Learning in Classrooms: Putting It into Practice*. London: Paul Chapman.
- Lowe, K. (2000). Graded self-assessment. *Literacy Learning: The Middle Years* 8(1), 54–60.
- North American Division Office of Education (2004–2010). *Assessing Cooperative Learning*. <http://www.journeytoexcellence.org/practice/assessment/cooperative> (accessed June 2009).
- Shulman, J., Lotan, R., and Whitcomb, A. (eds.) (1988). *Group Work in Diverse Classrooms: A Case Book for Educators*. New York: Teachers College Press.
- Torres-Skoumal, M. (2001). Alternative assessment models: Assessment through group work and the use of computer algebra systems as a self-assessment tool. *International Journal of Computer Algebra in Mathematics Education* 8(1), 61–83.
- Webb, G. Group Work (1981). An evaluation of techniques for analyzing small group work. *Programmed Learning and Educational Technology* 18(2), 64–66.
- Webb, N., Nemer, K., Chizhik, A., and Sugrue, B. (1998). Equity issues in collaborative group assessment: Group composition and performance. *American Educational Research Journal* 35(4), 607–651.

Further Reading

- Bain, D. (1988). School based assessment (part 4): The assessment of group work. *Mathematics in School* 17(1), 10–11.

Assessment and the Regulation of Learning

L Allal, University of Geneva, Geneva, Switzerland

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Affordances – Resources and constraints that are provided by the design of an instructional setting and can be incorporated into the learner's activities.

Formative assessment – Assessment that allows adaptation of teaching and learning activities in order to take into account students' knowledge, strategies, goals, needs, and interests.

Summative assessment – Assessment that sums up information in order to certify students' attainments at the end of a segment of an educational program.

Predictive assessment – Assessment that provides information for decisions about admission to an educational program or about placement in one of the sectors or tracks of a program.

Regulation of learning – Entails the processes of goal setting, monitoring progress towards the goal, interpretation of feedback from monitoring, adjustment of goal-directed actions and/or of the definition of the goal.

- Self-regulation: The processes of regulation are carried out by the learner
- Other-regulation: Another person (teacher, mentor, peer) intervenes in the processes of regulation

The assessment of student learning provides information that is analyzed, interpreted, and used for the regulation of educational activities. Two directions of regulation can be distinguished in association with the functions of student assessment defined in the well-known handbook edited by Bloom *et al.* (1971). The first direction concerns regulation of students' progression through the entrance, transition, and exit points of the educational system. It includes predictive or diagnostic assessments used for decisions about admission to an educational program or about placement in one of the sectors or tracks of a program. It also includes summative assessments that certify the quality of student achievement at the end of the segments of a program. These assessments, which determine students' access to educational resources and assure public recognition of their accomplishments, are based on the degree to which students' knowledge, competencies, or other attainments meet the standards fixed by the educational system. The second direction of regulation concerns the

adaptation of teaching and learning activities in order to take into account students' knowledge, strategies, goals, needs, and interests. This direction of regulation is closely linked to the formative function of assessment. In the initial conception of Bloom and co-workers, formative assessment intervenes in each unit of instruction and provides feedback that guides the choice of corrective measures aimed at the remediation of learning difficulties.

The concept of regulation has introduced, particularly in the French-language literature, a broader approach to formative assessment (Allal and Mottier Lopez, 2005). It includes remediation, which is a form of retroactive regulation that intervenes when students return to a task they have failed to master and, through the use of corrective materials or other forms of assistance, devote additional time and effort to attain the instructional objectives. Regulation of learning can, however, take on two other forms. Interactive regulation occurs when formative assessment is integrated into ongoing instructional activity; it results from the student's interaction with the components of the activity – that is, with the teacher, other students, and/or with material designed to encourage active reflection and self-assessment. Proactive regulation occurs when several sources of assessment information allow the preparation of new educational activities designed to take into account differences among students. It entails differentiation of instruction to ensure enrichment and consolidation of learning, according to the needs and interests of all students, rather than focusing on remediation of learning difficulties.

The remainder of this article examines two topics: first, the models of regulation that have been developed in research on learning and their implications for the design of assessment and, second, the contributions from research on assessment aimed at promoting student learning. The conclusion looks at how the continuity between formative and summative assessment can support the regulation of learning.

Models of Regulation of Learning

All theories of learning propose a mechanism of regulation that ensures adaptation of the learner's behavior and thought processes. Well-known mechanisms include reinforcement in behaviorist theory, equilibration in Piaget's constructivism, feedback devices in cognitive models, and social mediation in sociocultural and social constructivist approaches. Although there are important differences

among these theories, they all consider the processes of regulation to be a central feature of learning.

Regulation involves four main processes: goal setting, monitoring progress toward the goal, interpretation of feedback derived from monitoring, and adjustment of goal-directed actions and/or of the definition of the goal itself. A large number of empirical studies of these processes have been carried out in research on self-regulated learning, often abbreviated as SRL (Zimmerman and Schunk, 2001). This research aims at determining the ways students regulate their learning activity, the instructional, social, and contextual factors that influence self-regulation, and the effects on student achievement and well-being. Several models have been developed to describe self-regulation as an individual psychological process. One of the most comprehensive models is the three-layer model proposed by Boekaerts (1999). The inner core of the model consists in the regulation of cognitive processing, and particularly the choice of cognitive strategies. The next layer entails the regulation of learning and the use of metacognitive knowledge and skills to direct one's learning. The outermost layer corresponds to the regulation of the self and is linked to the formulation of goals and the allocation of resources (time, effort, etc.) to their attainment. Complex linkages among the layers allow self-regulated guidance of the individual's activities.

Self-regulation is often contrasted with external regulation or other-regulation. External regulation is associated with the structural features of the learning environment that stimulate and support learning, such as the types of tasks proposed and the sequencing of tasks, the tools and other available resources, the organization of time and space in the instructional setting, and the mechanisms of feedback and reinforcement that are present. Other-regulation refers to the interventions of other persons (e.g., teachers, mentors, and peers) who provide scaffolding and interactive guidance for learning. It has been observed that, in the classroom context, there is often a complex interplay between student regulation and teacher regulation of learning (Vermunt and Verloop, 1999). Perspectives derived from social constructivism, from sociocultural theory and from work on situated cognition treat learning and teaching as interdependent dimensions of educational activity. This has led to recognition of the reciprocal, or even dialectical, relations between student self-regulation and the sources of regulation situated in the learning environment (structure of the tasks, teacher interventions, peer interactions, and assessment tools). The regulation of learning in educational settings can thus be considered fundamentally as a process of co-regulation or of shared regulation (Hadwin and Oshige, 2007). This means that student self-regulation develops in interaction with multiple sources of regulation in the learning environment and, at the same time, contributes to the deployment and exploitation of these sources in the learning activities undertaken in class.

To take an example from ordinary classroom practice: the teacher's way of formulating questions provides a framework for students to learn to ask themselves questions; the questions they ask themselves frame, in turn, how they respond to the teacher and, more importantly, how they contribute to the evolving dialog.

The transition from models of individual self-regulation to models emphasizing the interplay between individual and social or contextual aspects of regulation has several implications for the design of assessment. First, it means that assessment is embodied as much in the social interactions taking place in the classroom as in the formal assessment tools and procedures that are used. Ways of structuring these interactions are therefore part of the design of assessment. A second implication is that the effects of assessment tools on learning can be amplified by their integration in social interactions. In order for an assessment tool, proposed in curriculum material or devised by the teacher, to have positive effects on student learning, it needs to become an object that students are able to appropriate. This means designing activities that allow students to discover the aims and properties of assessment tools and that encourage them to enter into discussions about the uses of assessment.

Contributions of Assessment to the Regulation of Learning

A review of the research by Black and Wiliam (1998) showing how assessment affects student learning, both positively and negatively, was the starting point for the development of the concept of assessment for learning, formulated by the Assessment Reform Group (1999). This concept includes several guiding principles for the conceptualization of assessment designed to promote learning. Assessment for learning is primarily concerned with the formative function of assessment, with how assessment is embedded in teaching and learning activities, with the quality of feedback provided by assessment, and with student involvement in assessment. Assessment for learning can, however, encompass forms of summative assessments that are devised to exert a positive influence on the way students approach learning. Shepard (2000) has described the transformations of the curriculum and the movement toward social constructivist learning theories that provide the foundations for a new vision of classroom assessment.

Integrating Assessment in Teaching and Learning

The initial conception of formative assessment, as described by Bloom and co-workers, divided instruction

into successive phases: teaching (or another form of delivery of instruction), formative testing, and remediation based on corrective material. Formative assessment can play a more pervasive role in the regulation of learning when it is integrated from the beginning in each teaching and learning activity. This implies looking for ways of embedding formative assessment in curricular materials, in teachers' ways of interacting with their classes, in the collaborative activities that students undertake in small groups, and in the individual tasks that students perform inside and outside of the classroom. Learning environments need to be designed so that they include resources and constraints (referred to as affordances in the literature on instructional design) that enhance the regulation of learning. This means the introduction of feedback at each stage of a learning activity, the use of questioning techniques that stimulate student reflection about alternative ways of carrying out a task, and the provisions of some degree of task differentiation to take into account learner interests and choices. The interactions that take place in class, between teacher and students, and among students, are the principal occasions for ongoing formative assessment.

The in-depth study carried out by Torrance and Pryor (1998) in English primary schools is an outstanding example of research on the formative assessment components present in teacher–student interactions. This study combines a social constructivist perspective on learning with an interest in the micro-sociology of classroom interaction. Although it does not frame its analyses and interpretations in terms of the concept of regulation, it provides a multitude of insights into the processes of regulation identified in the first section of this article. Torrance and Pryor's research is based on extensive audio and video recordings as well as on interviews with teachers and students. Through the presentation and interpretation of a large number of transcripts of assessment incidents, the authors reveal the patterns of interaction that embody formative assessment. They also describe the ways in which the meaning of assessment is negotiated in teacher–student dialogs and the influence exerted both on students' learning and on teachers' conception of their role. Their research is particularly important because it shows that teacher interventions in group settings can make a positive contribution to the regulation of learning for some children but not for others. It raises, in this respect, questions about the equity of interactive formative assessment. Several transcripts demonstrate, nevertheless, the power of this form of assessment when the teacher is able to appropriate individual children's words and actions and integrate them into a collective strategy for scaffolding and regulating learning. In a social constructivist interpretation, teacher–student interactions, and the processes of co-regulation they assure, are the contexts in which students construct and consolidate their individual strategies of self-regulation.

Providing Effective Feedback

In the literature on regulation of learning, feedback is information derived from monitoring the learner's progression toward a goal. It indicates how close the learner is to reaching the goal and may also provide indications on the types of obstacles to be overcome. It is the learner's or the teacher's interpretation of feedback that allows it to be used for the adjustment of goal-directed actions and/or of goal definitions. In this perspective, feedback is merely one aspect of regulation, albeit an essential one. In much of the literature on assessment, however, feedback is considered to be a system that includes several components: information about the student's present level of learning, a mechanism for comparing this level to a reference level (goal, objective, and standard to be attained), and a means for closing the gap between the two levels (Black and Wiliam, 1998). This broad conception of feedback encompasses the mechanisms of regulation that allow the student to progress.

In discussions of feedback, closing the gap generally means enabling students to adjust their behavior and understanding so as to reach the reference level of attainment defined by the educational system. There are circumstances, however, in which differentiation of the reference level may be needed in order to better take into account differences in students' needs and interests. A partial differentiation of goals, and not only of means for attaining them, does not mean a reduction of educational expectations for some categories of students but rather a more fine-tuned adaptation of goals so as to allow all children to acquire essential competencies while being able to express their individuality and cultural heritage. When feedback is integrated within a larger framework of co-regulation of learning, teachers and students discuss the goals to be attained, the criteria and standards of reference, and look for ways of personalizing goals, of making them meaningful in the learning trajectory of each individual.

A synthesis of meta-analyses on the effects of feedback has shown that it is one of the most powerful mediators available for fostering student learning (Hattie and Temperley, 2007). The average effect size of feedback is nearly 0.80, which is considerably higher than the effects of most other instructional factors. This synthesis also reveals that feedback can have powerful negative as well as positive effects on the regulation of learning. Learned helplessness and the fear of failure are as much the result of feedback as improved cognitive processing and feelings of self-efficacy. For a better understanding of the effects of feedback, Hattie and Temperley propose a model that distinguishes four types of feedback:

1. feedback about the student's level of performance or degree of understanding of a task,
2. feedback about processes (procedures, strategies, etc.) needed to understand and perform the task,

3. feedback that concerns the student's self-regulation (goal-setting, monitoring, and adaptation) with respect to the task, and
4. feedback about the self concerning the student's qualities as a learner.

Their review of the research evidence indicates that feedback about the processes involved in the task and about the student's self-regulation are more effective in promoting learning than feedback focused on task performance or on the learner as a person.

Since feedback about the learner as a person can have negative consequences, precautions need to be taken in the way information from assessment is communicated to students. Feedback that is useful for self-regulation of learning cannot, however, be entirely disconnected from feedback about the self. Learning involves the construction of one's identity, in addition to the acquisition of knowledge and skills. Moreover, all students, whatever their strengths and weaknesses, need to have the conviction that they are valued members of a learning community. This means that assessment procedures need to engage students in active reflection about what it means to be a learner and in participation with others in the construction of shared knowledge.

Involving Students in Assessment

Even in standard assessment situations based on teacher-made tests or external examinations, students not only demonstrate what they have learned but also learn new things about themselves. In this sense, self-assessment is ubiquitous; it intervenes implicitly in all assessments and continually affects students' learning and identity formation. It is possible, however, to introduce ways of conducting assessments that encourage active student involvement as well as a certain degree of student empowerment. Self-assessment tools can be embedded in curriculum materials or devised by the teacher. The impact of these tools on student self-regulation is likely to be enhanced when teachers conduct discussions that allow students to analyze the tools (their aims, their uses, and their possible misuses) and to confront the results of their ways of using them. Teachers can also go a step further and assist students in the construction of self-assessment instruments, such as a personal checklist of problems to look out for when writing texts, or a list of three questions to ask one's self when solving word problems.

Student involvement in assessment also includes various forms of peer assessment or of peer participation in the formative assessments conducted by teachers. Reciprocal peer assessment occurs when students assess the work produced by other students and then communicate their observations and suggestions for improvement. Joint peer assessment occurs when students examine a piece of

work together and share their ideas on how to improve it. Peer assessment can also be integrated in whole-class discussions orchestrated by the teacher in which criteria for an activity are constructed collectively and applied to various student productions.

In addition to the use of self-assessment and peer assessment tools, student involvement can go further in regulating learning when students actively participate in reflection about learning goals and about the meaning of different learning activities. This implies looking for ways in which at least some of the standard goals of education can be adapted or personalized so as to foster their appropriation by all students, given their learning histories and cultural backgrounds.

Research, demonstrating the effects of self-assessment and peer assessment on achievement, has been conducted primarily in the context of higher education (Boud, 1995). There are, however, research findings that offer support for the implementation of student-involved assessment during the entire course of schooling, and a wide variety of procedures, practices, and tools have been developed in this direction (see Stiggins *et al.*, 2004).

Continuity between Formative and Summative Assessment

Although formative and summative assessments have clearly different goals, the question can be raised as to their possible synergy in promoting learning (Harlen, 2005). The pressure of summative assessment, particularly when it is linked to frequent standardized testing, often leaves little space for the practice of formative assessment to develop. The consequences for the regulation of learning can be highly detrimental, especially for students who encounter failure repeatedly and cease trying to exert any control over their own learning. At the same time, summative assessment is necessary as a means of assuring social recognition of students' accomplishments both in school and outside. Students themselves inevitably ask: What knowledge and skills have I in fact acquired? Also, how do they measure up to expectations in society at large?

Continuity between formative and summative assessment can be developed in several ways. The first is through the alignment of both types of assessment with the curriculum goals underlying teaching and learning in the classroom. If this alignment is clearly perceived by students, the impact on their own goal setting can be very strong. A second point of continuity concerns the development of means of reporting the results of summative assessment so as to provide students with high-quality feedback about learning outcomes. When students receive a profile of test results, a graph comparing outcomes on different parts of a test, a set of rubrics describing the

qualities of a text, or teacher comments that accompany a grade, they can use this information to regulate their subsequent investment in learning. A third point of continuity has to do with student involvement in summative assessment. This form of assessment inevitably entails a judgment formulated by a professional (teacher, examiner, or other expert) about the quality of student learning. It is possible, nevertheless, to develop some degree of active student engagement in the way summative assessment is conducted. For example, in portfolio assessment used for summative purposes (grading and certification), students can participate in the selection of the work samples to include in the portfolio and be asked to write self-reflective commentaries that accompany and put into perspective their work. In professional education, summative assessment often takes place in conferences where the self-assessment expressed by the student is confronted with the assessment formulated by the teacher or supervisor. Students' knowledge of the conditions in which summative assessment will take place can have an important influence on the regulation of their investment in learning prior to being assessed. Teachers' knowledge of these conditions can have an equally important influence on how they organize learning activities and interact with students. To conclude: both formative and summative assessments provide explicit frames of references that guide the processes of co-regulation of student learning.

See also: Classroom Assessment Tasks and Tests; Formative Assessment; Peer and Self-assessment; The Relationship Between Assessment and the Organization and Practice of Teaching.

Bibliography

- Allal, L. and Mottier Lopez, L. (2005). Formative assessment of learning: A review of publications in French. In *Formative Assessment: Improving Learning in Secondary Classrooms*, pp 241–264. Paris: Organisation for Economic Co-operation and Development.
- Assessment Reform Group (1999). *Assessment for Learning: Beyond the Black Box*. Cambridge, UK: University of Cambridge School of Education.
- Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education* **5**, 7–74.
- Bloom, B. S., Hasting, J. T., and Madaus, G. F. (eds.) (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.
- Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research* **31**, 445–457.
- Boud, D. (1995). *Enhancing Learning through Self-Assessment*. London: Routledge.
- Hadwin, A. F. and Oshige, M. (2007). Self-regulation, co-regulation and socially shared regulation: Examining the many faces of social in models of SRL. *Paper Presented at the Conference of the European Association of Research on Learning and Instruction*. Budapest: Hungary.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning: Tensions and synergies. *Curriculum Journal* **16**, 207–223.
- Hattie, J. and Temperley, H. (2007). The power of feedback. *Review of Educational Research* **77**, 81–112.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher* **29**(7), 1–14.
- Stiggins, R. J., Arter, J. A., Chappuis, J., and Chappuis, S. (2004). *Classroom Assessment for Student Learning: Doing it Right – Using It Well*. Portland, OR: Educational Testing Service.
- Torrance, H. and Pryor, J. (1998). *Investigating Formative Assessment: Teaching, Learning and Assessment in the Classroom*. Buckingham: Open University Press.
- Vermunt, J. D. and Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and Instruction* **9**, 257–280.
- Zimmerman, B. J. and Schunk, D. H. (eds.) (2001). *Self-Regulated Learning and Academic Achievement*, 2nd edn. Mahwah, NJ: Erlbaum.
- Allal, L. and Pelgrims Ducrey, G. (2000). Assessment of–or in–the zone of proximal development. *Learning and Instruction* **10**, 137–152.
- Baumeister, R. F., Heatherton, T. F., and Tice, D. M. (1994). *Losing Control: How and Why People Fail at Self-Regulation*. San Diego, CA: Academic Press.
- Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2003). *Assessment for Learning: Putting It into Practice*. Berkshire: Open University Press.
- Crooks, T. (1988). The impact of classroom evaluation practices in students. *Review of Educational Research* **58**, 438–481.
- Davies, A. (2000). *Making Classroom Assessment Work*. Courtenay, BC: Connections Publishing.
- Dochy, F. and McDowell, D. (1997). Assessment as a tool for learning. *Studies in Educational Evaluation* **23**, 279–298.
- McCaslin, M. and Hickey, D. T. (2001). Self-regulation and academic achievement: A Vygotskian view. In Zimmerman, B. J. and Schunk, D. (eds.) *Self-Regulated Learning and Academic Achievement*, 2nd edn., pp 227–252. Mahwah, NJ: Erlbaum.
- Munns, G. and Woodward, H. (2006). Student engagement and student self-assessment: The REAL framework. *Assessment in Education* **13**, 193–213.
- Perrenoud, P. (1991). Toward a pragmatic approach to formative evaluation. In Weston, P. (ed.) *Assessment of Pupils' Achievement: Motivation and School Success*, pp 70–101. Amsterdam: Swets and Zeitlinger.
- Sadler, R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education* **5**, 77–84.
- Shepard, L. A. (2001). The role of classroom assessment in teaching and learning. In Richardson, V. (ed.) *Handbook of Research on Teaching*, 4th edn., pp 1066–1101. Washington, DC: American Educational Research Association.

Further Reading

Relevant Websites

- <http://www.assessment-reform-group.org> – Assessment Reform Group, UK.
- <http://www.assessmentinst.com> – ETS Assessment Training Institute, Portland, OR, USA.

Challenges of Developing and Implementing Formative Assessment Practices in Schools

R Sutton, Education Consultant, Cumbria, UK

© 2010 Elsevier Ltd. All rights reserved.

Case-Study Approach

Elsewhere in this encyclopedia, various authors have drawn our attention to the principles and purposes of assessment. The key purpose of this contribution is to explain and illustrate some of the practical implications and outcomes of implementing current assessment ideas in real schools and classrooms. For this purpose, two case studies are presented here. Ourtown Elementary School and Neartown High School do not exist: they represent a wide range of the issues surrounding the development and implementation of formative assessment in two stages of schooling. The content of these two case studies is supported by research but not entirely drawn from others' work. The author's own work over the past 20 years has been with schools and teachers, mainly in the UK, New Zealand, and Canada, supporting the introduction of assessment strategies that support learners and learning. Some, but not all, of this work has been published elsewhere.

The choice of a case-study approach is deliberate, to highlight the complex reality of every school, riddled with contextual variables that crucially influence the change process. Technical understanding of formative assessment is of little use without respect for the historical, logistical, and emotional context of the institution.

From Principles to Practice

In their synthesis of the international research on classroom and school-based assessment in 1998, Paul Black and Dylan William reported that researchers are often puzzled by the gap that repeatedly emerges between teachers' apparent understanding of the principle of assessment and their classroom practice. What follows is an attempt to focus on what schools do, not what they know, and to explain why this gap may exist and what could be done to close it. There is much to gain: formative assessment can have a marked positive impact on both the short-term achievement of students and their longer term success as learners. It can also affect positively the motivation and engagement of students, which will make the teacher's day easier.

Case Study 1: Ourtown Elementary School

Ourtown Elementary School could be anywhere. Typically, it would have one or two classes for each year or

grade level. The teachers will have been well-trained, but – at least until quite recently – many of them will not have been made deeply aware of the principles of formative assessment and how to apply them in their teaching. Over the past decade or so, the teachers and the school's leadership will have been challenged by the introduction of new, comprehensive, and specific curricula emanating from outside – the state, or the province. These new requirements come with a strong sense of accountability and pervasive systems of monitoring.

Summative assessments are undertaken to ascertain the measured success of learners relative to these mandated expectations. The results of these assessments will be reported to the parents. At Ourtown Elementary external tests are required, which have had the effect of making teachers anxious and prone to overprepare their pupils. Some of the teachers believe that external tests make the results more reliable, but others insist that there would be a more positive impact on teachers' professional practice if moderated teachers' judgments were used instead of a test result. In any case, they have no choice, and the school principal has had to steer a delicate path between discussion about the purpose and impact of the tests and the recognition that they are mandatory. Damage limitation has been her private motto since the testing regime started several years before, and in public she speaks regularly about the school's overriding interest in learning, motivation, and meta-cognition, all of which will deliver better test results and, more importantly, make children better prepared for their learning lives. The principal and the teachers have also found useful ways of using the school's test results formatively.

Making Formative Use of Summative Results

The tests used at Ourtown Elementary are developed externally and sent away to be marked. The results are sent back to the school several weeks later, too late to be useful for that group of children, but useful as a means of raising questions about the teachers' future practice. Encouraged by the leadership team, when the test results arrive, the teachers and the leadership take a really detailed look at them. The overall average is of no formative interest to them: they want to know which aspects of the tests children did well on, and which not. They need to analyze which groups of children did well, and less well.

They look for patterns and trends that indicate where possible improvements in teaching and learning can be made. These next steps are then built into the teachers' plans and classroom practice, and into the overall improvement plan for the school.

In the first year or two of these test results being available, the school was not sure how best to respond. Some teachers were inclined to be defensive, wanting to blame the circumstances of the school, and even the children or their parents, if the results were not as good as they had hoped for. Gradually, however, they have become less defensive, more curious and more willing to examine the evidence, looking for questions rather than answers.

Assessment Data and School Self-Review

Last year, it was clear that the pattern of boys achieving less well in writing in the later years of the school (ages 8–11) was more than a temporary blip and was becoming a worrying trend, worsening over time. The teachers needed to find out more about what was really going on, and how they might tackle it. Accordingly, they re-examined how they were teaching writing, spoke to a sample of boys about their approach and their response to writing, and what the school might do to make it better. Individual teachers were dispatched to other schools, and to various courses, to bring back ideas for improvement. They shared what they had learned and incorporated the most promising strategies into their future programs. The improvement in boys' writing began to show almost immediately: now the teachers will look forward to the test results as part of the evidence and they will investigate to check for the desired improvement. They will also look for evidence in the children's classroom work, and their responses to more interview questions. The school also decided to canvass parents on the issue, to discover the types of support available at home, and to find ideas to share about parents' contributions to their children's progress in writing.

Using Triangulated Evidence

Triangulating the evidence in this way, looking at three sources rather than just one, provides the school with a far richer picture of the issue and its remediation than they could have gleaned from the test results alone. The test results piqued their professional curiosity: the rest of the improvement strategy was led and fostered by the school's leadership team through building a climate of enquiry, distributing the responsibility for research, and involving all the people whose efforts could lead to improvement – the teachers, the children, and their parents. Summative assessment at Ourtown Elementary is used for evaluative purposes, intelligently, professionally, and effectively.

Developing Formative Assessment

The province and the school district have both been influenced by the wealth and depth of research evidence about the potential of assessment for learning to enhance children's learning. The teachers attended the training provided, but some felt that they were just being asked to do things they already did. "We know all this," some of them said, "So what are we expected now to do differently?" Other teachers understood and accepted the strategies being presented to them but felt that they simply would not have the time to make them work, or that their classes were too large. One or two argued that children in their early years of schooling are not yet ready to manage any responsibility for self-critique and self-correction.

Gradually, again with patient and continual encouragement from the school leaders, the teachers reflected more deeply about the implications of the principles of assessment for learning (as formative assessment was now relabeled) and their implications for the fundamentals of teaching. Here are some the insights they gained, through their reading and discussion, and – just as importantly – through experimentation in their classrooms:

- Children can successfully critique and correct their own and each others' work if they are clear about the criteria to be applied.
- This understanding of criteria can be found in very young children: in the nursery, for example, the teacher talked to her children about lining up. She and her children discussed this basic classroom routine, its purpose, and the criteria for success. Then through questions and specific feedback the teacher routinely involved the children in self and peer critique and correction. As the children absorbed responsibility for their actions, their lining up improved, and the teacher needed to remind them less frequently. Investing in self and peer critique saved the teacher both time and patience.
- Clear criteria will help, but the children still need to feel safe and able to examine their work critically, recognizing the difference between feedback and friendship.
- Teachers will need to understand the criteria for quality themselves, and to find ways of sharing these with the children.
- In order to achieve this, teachers will first need to be clear about what the children are going to learn, not just what they are going to do.
- Once the criteria are clear and shared, and the children have a safe climate to examine their work and provide effective feedback for each other, the teachers need to plan opportunities for the children's assessment skills to be developed and practiced, so that they can receive high-quality feedback – so necessary for improved learning – from each other as well as from the teacher.

- Parents will need to understand that involving the children in the assessment process is not an abdication of the teachers' responsibility but an intentional step toward both improved learning and also meta-cognitive development.

In response to these insights, the school took the following action:

1. *Planning.* They reiterated that the purpose of teachers' planning was learning, not coverage.
 - They looked at the level and type of involvement of children that they would expect to see year/grade level by level, so that everyone would be clear about this vertical articulation and the contribution being made by each year/grade level toward the overall meta-cognitive development of the children.
 - Each teacher plans time to involve children in the coconstruction of success criteria. This will focus on the learning to be achieved, not just the activities to be completed. Time is also included in their plans for regular brief reviews of progress relative to these criteria, as a whole class, a small group, a pair, or by individual children. These reflections are planned into the work as it progresses, so that children are encouraged to improve their work, rather than wait until it is done, when they might be less inclined to go back and fix it.
2. *Questioning.* Realizing that one of the purposes of assessment for learning is to encourage children to think, the teachers at Ourtown Elementary next looked at the way they pose questions to children. They realized the importance of the wording of questions, shared those that seemed to generate the most thoughtful responses, and wrote them into their plans. They practiced waiting longer in class to allow children time to think and possibly discuss their responses. They changed the hands-up rule in class so that children no longer raised their hands to answer a question and the teacher was then able to include more children in building effective answers, rather than looking for the right answer straight away and picking different children until she found it.
3. *Marking and feedback.*
 - The teachers agreed that feedback makes most sense if it connects with clear criteria understood by the learner. They investigated ways of co-constructing criteria with rather than for their students, tried out various methods and invited other teachers into their rooms to see how they did it. The principal helped by providing class cover for these short visits to take place, or by taking the opportunity to spend time in the classroom herself while the teacher was visiting another room.
 - They realized that parents expected some marks on the reports, especially in the later years of schooling,

and that they should look again at their reports to see how they could manage this need for marks without adversely affecting the self-efficacy of the children. The principal took responsibility for sharing with parents, in writing and through a meeting, the reasons why the school is interested in formative assessment and what changes parents might expect to see.

- It was not difficult for teachers to agree on what constituted effective feedback. One of the teachers shared what she had learned from her assigned task of checking the research in greater detail on this point. She convinced her colleagues that their praise must focus not on who the child is ("You are so smart"), but rather on what the child has done, "I really like the way you used colourful adjectives/describing words here." The feedback also needs to be descriptive rather than evaluative, and to help the child to understand the process by which they he/she achieved success, rather than just the outcomes he/she have achieved. Sharing these ideas, and ways in which they translated into daily teacher practice, helped to improve the overall quality of feedback right across the school.

Changing Teachers' Habits

Changing these basic classroom routines was much harder than the teachers had expected. For a while, they wondered if they could ever make it work. From their own previous experience of change, and from the experiences of other schools, the school's leadership understood that this is a long game, requiring clear goals, small steps, continuing feedback, practice, and perseverance. The principal herself undertook to articulate the principles that they were striving for, and to keep these regularly at the front of people's minds. She and her colleagues were determined to model these principles in action, both in their own teaching, and in their approach to the teachers as adult learners, providing for them the clear purpose, expectations, criteria, feedback, and opportunities for improvement that they wanted to see teachers offering to their pupils. Gradually, through teachers' trying out and sharing their formative-assessment strategies, the norms of teaching changed and formative assessment became embedded in the classrooms of Ourtown Elementary.

Case Study Two: Managing Assessment at Neartown High School

In the high school, summative assessment has been part of schooling for many generations. Teachers are so accustomed to the requirement to cover the given program and prepare students for regular testing that some of them find

it difficult to consider assessment in any other way. Marks, scores, and grades are what they, their students and their parents take for granted as the way the classroom does business. Some subjects were less likely to organize assessment in this way: the physical education (PE) department, for example, had always used feedback in terms of specific words and advice – as good coaching always does – and did not mark students' work in the same way as others. The art teachers too used group critique as a routine part of assessment for improvement in their rooms. The more academic subject teams might not know what happened in subject areas other than their own, and in any case would not regard these practices as necessarily relevant to them. Different subject areas have often been organized in quite different ways within the same school. In this environment, the challenge was to find some degree of consistency across the school, while still respecting the rich variety of experiences, ideas, and skills that form the beauty of secondary education.

There were technical questions also about the school's assessment habits. Prompted by presentations at conferences, and by their own reading, some teachers began to question how they handled late assignments, for example, or gave students zero under some circumstances. The calculation of final grades and the allocation of merit awards were critically examined for the first time in quite a while, and some existing practices were hard to explain or defend. The school began to divide on these issues, and the leadership worried about allowing this fragmentation to continue. The assessment habits of this school, like many others, were very hard-wired indeed: questioning them caused many teachers disquiet, and this rethinking had to be led rather than merely managed. As in the elementary school, teachers looked to the school's leaders rather than to outside experts for a clear sense of purpose and direction. The first essential, therefore, was for the leadership team to do their homework, find out what the technical issues were, and then decide how best to motivate the diverse range of adult learners among the teaching staff. They could, and were prepared to, impose a certain amount of change from the top, but they knew that real change in classroom practice and attitudes toward assessment practice would only truly happen by encouraging a gradual change in teachers' daily practice. Appealing to teachers' enlightened self-interest would also be an important part of the change strategy at the high school.

Introduction to the Principles of Assessment for Learning

Like their colleagues at the local elementary school, the secondary teachers were first introduced to the ideas of assessment for learning through a professional development event offered by the school district. It was clear that these ideas had been developed from extensive research,

but this was not sufficient for the teachers to truly grasp their implications. As the implications of these principles gradually became clearer, how did the high-school teachers react? These are some of the reactions:

- Some teachers were simply daunted by the workload implications of providing high-quality feedback for all the students they teach. They could not see how they would have the time to provide such detail for all the work the students did. They had to find a pragmatic way through this discomfort: the easiest choice for this group was to reject the ideas altogether and continue with what they had always done and knew to be achievable.
 - Other teachers felt that they were unavoidably constrained by the demands of others: parents demand marks and teachers must therefore supply them; entry to college makes students compete for grades, and competitive grading will therefore backwash through the whole school. "Our hands are tied," they said: adoption of formative assessment practice, however desirable, was pragmatically impossible. The principal acknowledged these constraints, but still argued that the learning needs of the students were the school's first priority: formative assessment was not a distraction from the focus on achievement but rather a vital step toward it.
 - Given a clear message from the school leadership team that no change was not an option, teachers then looked at ways of providing effective feedback through formative-assessment strategies, using the students themselves as one source of feedback rather than all of it coming from the teacher. They looked again at the learning activities they were planning and considered the appropriate type and source of feedback to be provided.
1. *Teacher feedback to individual students.* They agreed that work that was new to the students, and important as the basis for further learning, would need and deserve the best-possible feedback from the teacher herself. This would entail clarifying criteria with the students beforehand, providing students as soon as possible with as much specific feedback as they could absorb, and giving them the chance in class to focus on this feedback and correct their work before moving on.
 2. *Teacher feedback to whole group.* Some of the work set for the students did not need detailed individual feedback, but it did need to be looked at by the teacher. In this case, the teacher took a quick look at all the students' work, used what she learned from it to share with the group, and adjusted her plans to accommodate any necessary consolidation or remediation. She did not provide individual feedback for each student.
 3. *Peer and self-assessment.* Another type of work done by the students could be assessed by the students themselves, given the necessary training and practice. They would

learn how to apply criteria to a given piece of work, and then practice doing so and provide useful feedback to each other. As the teachers discussed how best to make this work, a couple of examples of existing practice emerged, that helped others to see what they might do.

This pragmatic approach to formative assessment was achieved by lowering the temperature of the debate, to release the practical creativity and ingenuity of the teachers involved. It was helpful for the principal to accept and state that he did not expect a complete change in all teachers' assessment practice immediately. He suggested that each teacher choose one teaching group with which to try new strategies. Teachers would then be given time first in their subject teams and then in cross-curricular groups to share and discuss what they did and what happened. They were not expected to get it all right first time, and were forewarned that both they and the students might find the unfamiliar practices more difficult to start with, as is the case whenever we change our habits. The schedule and purpose of subject-team meetings was changed slightly to accommodate the necessary sharing. The vice principal talked to all the subject-team leaders about the particular circumstances and approach to change in team, and how to support the team leader in effecting change.

Two months later, when subject teams had been trying out formative assessment strategies in their classes for a while, subject leaders came together, led by the same vice principal, to share their experience with other colleagues, to look for common themes, and identify the road blocks. At this meeting, a number of issues were raised:

1. Some teachers were concerned that the students could not separate feedback from friendship, and this affected the accuracy of feedback in peer assessment. They agreed that students would be given plenty of training with applying criteria to neutral examples of work before graduating to peer assessment when the teachers felt that they were skilled enough not to be deflected by their relationship with the person they were working with.
2. The teachers' experience bore out the research finding that students are distracted by marks, scores, and grades on their work, and it was agreed that they would try keeping marks separate from feedback, where possible, to encourage students to focus on improvement. This raised the whole question of the frequency of marks, which led in turn to further discussion about how students' final grades are determined. As this issue affected every subject team, a special task group was created to investigate the issue in greater depth and report back with a number of ways forward for the whole school to decide on.
3. The group discovered that when formative-assessment strategies had been running for a while the students themselves began to see the positive impact on their work and performance. It was decided that the school's planned explanations to parents should include input from the students themselves, with concrete examples of improved achievement.
4. With more time to focus on classroom strategies, the teachers reporting back to the vice principal began to highlight examples of the effect on students' motivation and levels of independence. It was decided that another meeting for parents and senior students should be addressed by a representative from the local college, talking about the students' need in postsecondary to be less dependent on their teachers and more able to accept responsibility for their own learning, both of which were goals of the school's changing assessment practice.

Roadblocks

During the school's implementation of their intentional and consistent approach, they ran into several roadblocks that needed to be handled carefully:

- Some subject areas appeared to lend themselves to these ideas more easily than others. In English, the teachers' normal way of working was through feedback and revision, as the students practiced a sets of skills repeatedly in different contexts and at growing levels of complexity. In math and science, on the other hand, the teachers felt they needed to be constantly moving on to different aspects of the subject that were not clearly related to what had gone before. They struggled to help students see the links, and the relative fragmentation of the subject, in view of some of the teachers, meant that they could not afford the time to correct skills that would not necessarily be used again. Other math and science teachers argued to the contrary, using problem solving, investigations, or lab-work as the continuing skills to be improved through clear criteria, feedback, and improvement. The best help for the math team came from detailed work with the math team from another school, where their special context allowed them to get down to more detailed and useful discussion, next steps, practice, and sharing their successes.
- A further roadblock arose from the high-stakes nature of the summative assessments undertaken at various stages in the high school. Teachers, students, and parents all took these results very seriously, and teachers were unwilling to experiment with new teaching strategies if they felt any uncertainty about the outcome. The principal's pragmatic suggestion in this case was to start changing the approach to criteria, marking, and feedback and the involvement of the students in lower stake years – not grade 12 in the first instance – so they could practice the skills involved and gather the evidence they needed of the positive impact on student performance.

- A further roadblock, encountered when this apparently new initiative was introduced, was that many Neartown teachers believed that this was separate from other requirements and initiatives that confronted them. Lacking the means to find connections with their daily work, and already annoyed by the increase in external interference and accountability, some of the teachers kept their heads down and hoped it would go away, while the skepticism of others affected their motivation to try the strategies with any real conviction. In this case, the excellent leadership of Neartown High School clarified and communicated the connection between assessment for learning and other initiatives. Instead of merely reacting to these external mandates, and treating them all separately, the leadership team developed its own vision of the school's goals for excellent learning and teaching, to which the external imperatives were then connected. Various initiatives were woven together, and the big picture was presented as clearly as possible, including in visual form. The daily behaviors of all the school's leaders were then influenced by their collective interest in formative-assessment strategies. As they observed teachers and teaching for the purpose of teacher evaluation, they would be applying agreed criteria for high-quality teaching that included formative-assessment strategies. As they walked through teaching areas they would pick up and comment on evidence of students' involvement in both learning and assessment. They would use open-ended questioning in their conversations with teachers to encourage reflection on their classroom practice.

School Change Is Influenced by System Change

All publicly funded schools will be influenced by the political and educational imperatives that accompany their funding. Our two case-study schools, which could be anywhere, will be affected by both these external factors and also, crucially, by the reactions of the school-based leadership. The monitoring of schools – by the state as in the UK and New Zealand, or by the school district – for example, can be handled by schools as either a curse or a blessing: the intelligent leadership team will use the event as a welcome opportunity to benchmark their pursuit of quality against external benchmarks, and not further undermine the motivation of their colleagues by raising anxiety about something beyond the school's control. In other forums, of

course, school leaders will try to use their collective influence to improve the inspection process, but in school they maintain a cheerful and encouraging stance.

Conclusion

The key question in assessment is about its purpose. Schools are expected to both help children learn and to check and report their achievement. Both formative and summative assessment practices are needed to achieve this, and schools have to manage them side by side, despite the continuing tension between the two, competing for space and attention in a space already crowded with expectations to cover fixed curricula and programs. Gradually, assessment for learning is affecting teachers' and schools' practice, but the progress is slow.

See also: Assessment: Pre-service and In-service Teacher Education; Formative Assessment and Instructional Planning; Formative Assessment; Impact of Assessments on Classroom Practice; Peer and Self-assessment; School Policies and Practices to Support Effective Classroom Assessment for Learning.

Further Reading

- Absolum, M. (2006). *Clarity in the Classroom*. Auckland: Hodder Education.
- Assessment Reform Group (1999). *Assessment for Learning: Beyond the Black Box*. Cambridge: University of Cambridge School of Education.
- Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2002). *Working Inside the Black Box*. London: King's College.
- Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education* 5(1), 7–74.
- Clarke, S. (2003). *Enriching Feedback in the Primary Classroom*. London: Hodder and Stoughton.
- Clarke, S. (2005). *Formative Assessment in the Secondary Classroom*. London: Hodder and Stoughton.
- Clarke, P., Owens, T., and Sutton, R. (2006). *Creating Independent Student Learners*. Winnipeg: Portage and Main Press.
- Davies, A. (2006). *Making Classroom Assessment Work*, 2nd edn. Courteney: Connections Publishing.
- Gardner, J. (ed.) (2006). *Assessment and Learning*. London: Sage.
- James, M., McCormick, R., Black, P., et al. (2007). *Improving Learning How to Learn*. London: Routledge.
- O'Connor, K. (2002). *How to Grade for Learning*. Toronto: Corwin.
- Stiggins, R. J., Arter, J., Chapuis, J., and Chappuis, S. (2005). *Classroom Assessment for Student Learning: Doing it Right – Using It Well*. Princeton, NJ: ETS.
- Sutton, R. (1995). *Assessment for Learning*. Salford: RS Publications.
- Wiggins, G. and McTighe, J. (2005). *Understanding by Design*. Virginia: ASCD.

Formative Assessment

P Black, King's College London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

The Early Development

The use of the term formative in education was introduced in the context of curriculum evaluation by Scriven and extended 2 years later to the improvement of teaching by Bloom (1969):

By formative evaluation we mean evaluation by brief tests used by teachers and students as aids in the learning process (p. 48).

However, as Black (1986) pointed out, the use of feedback from assessments to guide learning has always been a feature of education. His development of diagnostic assessment in Scotland was however exceptional in that the growth of standardized testing overshadowed formative developments for many years. Crooks (1988), in his survey of the literature on the impact of classroom evaluation practices on students, contrasted the many positive effects of good assessment with the evidence that classroom assessment often encouraged rote and superficial learning, with the grading function overemphasized; several authors described the difficulty of preventing the summative process from swamping formative work (Black, 1993).

This article presents a brief account of developments in formative assessment in the last 30 years, and goes on to discuss the principles and practices as they have emerged to date. A closing section reviews some of the obstacles to further development.

Policy Impact in the UK: The Task Group on Assessment and Testing Report

Fresh emphasis on formative assessment was developed in the UK with the report of a government-appointed Task Group on Assessment and Testing (TGAT), which recommended that a new national assessment system should:

- be criterion referenced,
- be formative,
- use moderated teachers' assessments, and
- be based on a framework of progression in learning.

In its closing summary, the groups' report emphasised that:

Our terms of reference stress that the assessment to be proposed must be "supportive of learning in schools". We reiterate that the four criteria set out in Section 1 are

essential if this support is to be secured. . . The **formative** aspect follows almost by definition. For the system to be formative it must produce a full and well-articulated picture of the individual child's current strengths and future needs. (DES, 1988 paragraph 221)

While the report's recommendations were accepted in principle, important reservations were expressed by the government. For the national tests taken by all at ages 7, 11, and 14, they rejected the TGAT proposal that these be based on moderated teacher assessment combined with scores from external tests, ruling instead that the test results and teachers' assessment results should be published separately. The predictable, and no doubt intended, result was that the test results became dominant (Black, 1997).

In and from these events, three of the problems that obstruct the implementation of formative assessment were emerging. These were:

- the lack of clear definition of formative assessment, which was seen only as the formative use of testing;
- the conflict between formative assessment and the accountability pressures of summative testing; and
- the failure to accept that formative assessment was not merely current good teaching, but rather a practice which would require from teachers a radical and difficult change in their role.

These three problems continue to inhibit formative work to the present day.

The New Perspective – 1990–2000

While several individuals made further study of formative assessment (e.g., Black, 1993), the main source for new thinking was the activity of the Assessment Policy Task Group of the British Educational Research Association (BERA), subsequently renamed as the Assessment Reform Group (ARG), which developed a critique of the development of national assessment policy in the UK, and were led to obtain funding for a comprehensive review of the literature on formative assessment. For this task, they commissioned the author and Dylan Wiliam, who produced a 55-page review (Black and Wiliam, 1998a) titled 'Assessment and classroom learning'. The first of its six main sections gave summary accounts of eight selected studies, the subjects of which ranged from 5-year-olds to university students, which described a variety of formative assessment practices, and all of which showed firm evidence

that these practices produced significant learning gains. There followed five sections which surveyed studies of:

- students and formative assessment – including discussions of goal orientation, self-perception, self- and peer assessment, and links to theories of learning;
- strategies and tactics for teachers – including choice of task, discourse, questions, the use of tests, and quality of feedback; and
- systems – including mastery learning, assessment-driven models of instruction, portfolios, and summative examination models.

These were followed by two more reflective and theoretical sections on:

- feedback – including discussions of the nature of feedback, meta-task processes, and motivation and learning processes;
- the prospects for theory and practice of formative assessment – commenting on the current theoretical basis, expectations and the social setting, further research, and potential implications for policy.

This review has been very widely cited – it was indeed long overdue. One reason for its influence was that it set out a broad agenda for further work on formative assessment. It gave particular impetus to such work by publicizing evidence, from many rigorous experimental studies, that formative assessment did indeed raise standards of performance in learning. These findings have been confirmed by re-analyses of the same and later evidence by other authors.

Widespread impact among teachers followed through a short booklet for teachers titled *Inside the Black Box* (Black and Wiliam, 1998b) in which the main conclusions, and their implications for both policy and practice, were summarized. This attracted wide publicity and to date over 50 000 copies have been sold, mainly in the UK.

Developing Practice

While the ARG followed the Black and Wiliam review in 1999 with a booklet developing its implications, what was lacking was evidence about how teachers might apply the many potential lessons from the review in their everyday practice. This challenge was taken up by a group at King's College London in a collaborative project with two local school district authorities (LAs) entitled the King's Medway Oxford Formative Assessment Project (KMOFAP).

The KMOFAP Project: Practices

Between 1999 and 2002 this project involved 48 teachers of mathematics, science, and English in six secondary

comprehensive schools. It was collaborative in that the King's group and the LAs emphasized that the lessons from research did not provide recipes for success, but were a source of suggestions which teachers might use as starting points to develop new practical knowledge about formative assessment. This strategy (described in Black and Wiliam, 2003) helped to encourage both the full engagement of, and practical innovations from, the teachers. Evidence collected from schools' test results confirmed that significant learning gains could be achieved within the normal working conditions of secondary schools, while the teachers involved also felt that the work had enhanced their own commitments to, and realization of, their professional values (Wiliam *et al.*, 2004).

The findings of the project led to the production of materials to guide schools and teachers in their own development of formative assessment methods. A short booklet, *Working inside the Black Box* (Black *et al.*, 2002), and a book (Black *et al.*, 2003), have both proved very popular. They set out the two main outcomes, which were to specify the principles on which formative assessment should be based, and to map out the main areas of formative practice. These two are discussed in the following two sections.

Basic Principles

A test as such may be either formative or summative or both – the distinction lies in the purpose for which information is interpreted and used, not in the means used to elicit that information (Wiliam and Black, 1996). The distinction is emphasized in the following definition:

Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting pupils' learning. It thus differs from assessment designed primarily to serve the purposes of accountability, or of ranking, or of certifying competence.

An assessment activity can help learning if it provides information to be used as feedback, by teachers, and by their pupils, in assessing themselves and each other, to modify the teaching and learning activities in which they are engaged. Such assessment becomes 'formative assessment' when the evidence is actually used to adapt the teaching to meet learning needs (Black *et al.*, 2002 – inside front cover).

Others have produced essentially similar statements (ARG, 2002a).

Given that evidence is collected to promote learning, its use must be consistent with established principles for effective learning, which may be summarized as:

- Start from a learner's existing understanding.
- Involve the learner actively in the learning process.

Develop the learner's understanding of the aims and criteria for effective learning.

Promote social learning, i.e. learning through discussion.

These are concerned with the cognitive dimension of learning. The affective and motivational dimensions are also important. The work of Butler (1988) showed that feedback given as rewards or grades enhances ego rather than task involvement and that task-involved students did better than the ego-involved. The more extensive work of Dweck (2000) likewise showed that for ego-involvement, with a mind-set focused on competitive performance, both high and low attainers are reluctant to take risks and react badly to new challenges, while failures simply damage self-esteem (see also ARG, 2002b), whereas with task-involvement, learners believe that they can improve by their own effort, are willing to take on new challenges and to learn from failure.

Areas of Practice

Oral and Written Feedback

One of the KMOFAP teachers summarized very effectively the changes she made in her use of questions in class:

Questioning

- My whole teaching style has become more interactive. Instead of showing how to find solutions, a question is asked and students given time to explore answers together. My Year 8 target class is now well-used to this way of working. I find myself using this method more and more with other groups.

No hands

- Unless specifically asked students know not to put their hands up if they know the answer to a question. All students are expected to be able to answer at any time even if it is an 'I don't know'.

Supportive climate

- Students are comfortable with giving a wrong answer. They know that these can be as useful as correct ones. They are happy for other students to help explore their wrong answers further (Black *et al.*, 2003: 40).

Research results on wait time impressed teachers with the need to give pupils time to think; many encouraged students to talk with one another before answers were called for, so that as many as possible would feel confident about expressing their ideas. The corollary of this approach was that each question had to be sufficiently open, from the learners' perspective, that it called for thought, and had to be sufficiently central to the learning aims that it justified the time spent on discussing the ideas that ensued. It followed that closed questions, checking only on knowledge, would be seen to make little contribution to students' learning. If students had to think, their

wrong or partly right answers would give teachers the information about the students' understanding, which would help them choose the optimum way for the work to proceed.

The key issue was that the formative purpose of a question is achieved only if it elicits from students some significant indicator of their understanding and then enables the teacher or other students to respond by trying to correct or develop that understanding, and perhaps through several responses lead the whole class to share a discussion of the issue. Thus, what is required is very different from the all-to-frequent, and still dominant, pattern in which the teacher may dominate by a sequence of triads of initiation, response, and judgment/correction. What is needed, if the classroom is to engage all students in thinking, is a radical shift toward dialog (Alexander, 2006) or, in Van Lier's (1996) terms, toward conversation. This latter view highlights the delicate task of teachers, for on the one hand tight control over the discussion can inhibit involvement of the learners, but on the other hand loose control can lead to digression so that the purpose of the learning is lost. Given that student participation is often unpredictable, the task of steering such dialog is a delicate one: this is an issue that requires further study.

Just as in questioning, in written work the purpose has to be to provide feedback that helps learning. The difficult change here was for teachers to stop giving marks or grades on written work. They were encouraged by research findings that comment-only marking led to improved learning (Butler, 1988), and came to justify the new practice as follows:

- students rarely read comments preferring to compare marks with peers as their first reaction on getting work back;
- teachers rarely give students time in class to read comments that are written on work and probably few, if any students, return to consider these at home;
- often the comments are brief and/or not specific, for example "Details?";
- the same written comments frequently recur in a student's book, implying that students do not take note of or act on the comments (Black *et al.*, 2003: 43).

In consequence, teachers' realized the need to spend more time on careful formulation of comments that would help students to understand their faults and to improve their work. Such comments as "Be more sensitive" and "You are mixing up the words 'solution' and 'mixture'" give the learner no useful guidance, but if the latter were expanded to read "Allan, this is generally fine but you are mixing up the words 'solution' and 'mixture'. Look up what we all wrote down about the difference and then check through this piece again." should help Allan to improve his understanding. The key point is that the written work is not

to be treated as a terminal test, for which the marks produced are the end of the matter, but as an opportunity, through formative feedback, for improving learning.

One feature that distinguishes written from oral feedback is that there is time to frame comments that encourage further learning, with more opportunity for differentiation through separate interaction with individuals. The effects on motivation of the type of feedback are also important: as one teacher expressed it:

The results were especially noticeable in lower attainers since grades can often have a de-motivating effect with such students which can be extremely destructive to their self-esteem (KMOFAP project, unpublished transcript).

For all feedback, any intervention by the teacher must be based on some assumptions about the nature of the thinking that produced that response. In the absence of any effective model of that thinking:

...part of the feedback given to pupils in a class is like so many bottles thrown out to sea. No one can be sure that the message they contain will one day find a receiver. (Perrenoud, 1998: 87).

This judgment highlights the difficulty of the problem. Authors concerned with the self-regulation involved in learning have explored relevant models (Greene and Azvedo, 2007), but whether, and how, such models might guide the day-to-day formative interventions by teachers has yet to be explored. Comparisons with dynamic assessment, which is centered on formative interpretation of feedback using a specific theory of cognitive development, might also be helpful (Leung, 2007).

Peer- and Self-Assessment

Self-assessment by students is essential for their development as independent and responsible learners. To meet this need, practice must involve them in assessing one another's work so that they might thereby develop the skills both of peer assessment and, by seeing their work through the eyes of their peers, of self-assessment. To do this, students have to understand the aims of the work involved, and the criteria by which this work should be assessed. In certain topics in some subjects, the criteria might be well defined (e.g., calculating the value of a force in a physics problem), in others there might be many different ways in which work could achieve excellence (e.g., writing a critical appreciation of a poem). Overall, the aim was to move students away from dependence on the teacher toward independence in the power to guide their own learning. As one teacher expressed it:

The kids are not skilled in what I am trying to get them to do. I think the process is more effective long term. If you invest time in it, it will pay off big dividends, this process

of getting the students to be more independent in the way that they learn and taking the responsibility themselves (Black *et al.*, 2003: 52).

In helping students to develop their self-assessment, teachers must clearly specify aims so that students can both steer their work toward attaining them and understand the criteria by which that work can be judged.

Peer assessment is another powerful tool: helping pupils collaborate effectively in groups is essential to such work. However, while there is strong evidence that collaborative groups can improve attainment, a survey of group work in a large sample of UK classrooms (Blatchford *et al.*, 2006) shows that the type of collaboration that can engage pupils in reasoned arguments about their own and one another's contributions is not often found. Intervention programs by Blatchford and colleagues and by Mercer *et al.* (2004) have demonstrated that students trained in such collaboration produce improved attainments, both in the quality of their arguments and in subsequent tests. The work of Dawes *et al.* (2004) is a good example of dissemination of the lessons learned about effective group work.

The formative use of summative tests extends this area of practice. Students learn by marking one another's test responses, particularly if they have to develop the marking schemes in the light of their understanding of the criteria. As for the new way of looking at written work generally, the point is that a test, while serving the summative function, might also be used, through formative feedback, as an opportunity for improving learning. One addition in this area was that teachers would strengthen preparation by asking students to compose questions which might well be suitable for the coming tests, for in so doing they would have to identify for themselves the main aims of the work. They would also learn about the need to frame the text of a question with care, and to exercise similar care in responding to questions framed by others.

Growing Points

The above accounts indicate areas where further work is needed to deepen understanding and improve the practical implementations of formative assessment. There is also a need to explore the differences in formative practice between different school subjects (Hodgen and Marshall, 2005). Insofar as formative interaction is at the heart of teaching and learning, it may seem inevitable that ideas about it may expand into a complete theory of pedagogy. It is necessary here to be more specific: the heart of the formative interaction is the active involvement of the learner who produces evidence that is used to guide further learning. The means to produce such evidence, and the aims to which the guidance is directed, are also essential aspects of pedagogy, but the specific

formative element is the contingent and unpredictable interactions which advance the learning by challenging and extending the learners' thinking.

Why Is It Difficult?

Attempts to implement formative assessment have been hindered by superficial interpretations. For example, teachers may ask open questions, but then correct the answers rather than use them to redirect their work, or they may ask pupils to self-assess their work but then fail to use this information to formulate formative feedback. Many have interpreted the related terms assessment for learning as incorporating all uses of assessment, and in this perspective focused on frequent summative testing, while Bloom's (1969) account in terms of brief tests showed a limited perspective on the practice. Such particular misinterpretations have been prevalent in the USA where, with a few notable exceptions, discussions often present the topic under the heading 'Classroom assessment' in terms which leave unclear whether it is the summative of the formative functions which are being considered. A recent text titled *Learner-Centred Classroom Practices and Assessments* does not even mention, let alone discuss, formative assessment (McCombs and Miller, 2007).

There are outstanding obstacles to further improvement of formative practices. The first is that implementing new formative practices is a very demanding task for any teacher. For many, a deep change in their beliefs about their role in the classroom is required, with practical implications which many find pretty scary. As one teacher put it:

It was clear that new ways of working together with new boundaries had to be established and to do this successfully took all my skills as a teacher. . . (Black *et al.*, 2003: 90).

Such difficulties pointed out in much earlier literature:

To incorporate formative assessment into their teaching would involve teachers in far more than acquisition of the necessary skills. . . The changes in their classroom practice might also involve profound changes of role, even for teachers regarded by themselves and others as already successful . . . devotion to formative assessment is risky, taking a great deal of time and energy. In particular, since many pupils may have acquired habits of doing just enough to get by, or have ceased to believe that they can be competent at the subject, the contract between teacher and pupil has to be reformulated. (Black, 1993: 79)

The initial disorientation was felt not only by teachers but also by students:

The first time I asked a Year 10 top set to work in groups to examine the errors they had made in a test and to help

each other to understand fully what was being asked of them, it was unsuccessful . . . That lesson I heard several times 'Why don't you just tell us?' . . . they have realised that they have more to do in my classroom than absorb the syllabus – they have to take responsibility for their learning (Black *et al.*, 2003: 87).

The changes needed cannot be produced by a short training session – they need the sustained support of collegial collaboration. Studies of individual teachers show that each changes in a different way and with different personal perceptions of the task (Black *et al.*, 2003: chapter 6; Lee and Wiliam, 2005).

A second obstacle is the dominance, in many national systems, of external summative testing with attendant high-stakes consequences. The perceived need to teach the superficial tactics which improve performance on superficial tests, limits teachers' freedom of maneuver in classroom work. The resolution of this problem will depend on trusting teachers to play a more responsible part in the summative assessment of their pupils (ARG, 2006). Then the problems of establishing synergy over the formative summative interface might be tackled.

A third obstacle is that to achieve such synergy, teachers need to have a basic expertise in such assessment issues, based on a broad understanding of the main principles and practices of assessment, including a thorough understanding of the criteria of validity and reliability. The lack of such expertise is a notable weakness because it means that the profession cannot take control of an essential component of its work, it is a partial profession. This view is borne out in work that has sought to develop teachers' summative assessment practices (Black *et al.*, 2007). Thus, developing understanding and skills in formative practices is only one component in building the assessment strengths of the profession – one that can only achieve limited success on its own.

See also: Alternative Assessment; Dynamic Assessment; Summative Assessment by Teachers.

Bibliography

- Alexander, R. (2006). *Towards Dialogic Teaching: Rethinking Classroom Talk*, 3rd edn. Cambridge, UK: Dialogos.
- ARG (2002a). *Assessment for Learning: 10 Principles*. Cambridge, UK: Assessment Reform Group, University of Cambridge Faculty of Education.
- ARG (2002b). *Testing, Motivation and Learning*. Cambridge, UK: Assessment Reform Group, University of Cambridge Faculty of Education.
- ARG (2006). *The Role of Teachers in the Assessment of Learning*. London: Institute of Education.
- Black, H. (1986). Assessment for learning. In Nuttall, D. L. (ed.) *Assessing Educational Achievement*, pp 7–18. London: Falmer.
- Black, P. J. (1993). Formative and summative assessment by teachers. *Studies in Science Education* 21, 49–97.

- Black, P. J. (1997). Whatever happened to TGAT? In Cullingford, C. (ed.) *Assessment vs. Evaluation*, pp 24–50. London: Cassell.
- Black, P., Harrison, C., Hodgen, J., Marshall, B., and Serret, N. (2007). Riding the interface: An exploration of the issues that beset teachers as they strive for assessment systems. *Paper Read at the Annual BERA Conference*, London September 2007. <http://www.kcl.ac.uk/schools/sspp/education/research/groups/assesspubs.html> (accessed June 2009).
- Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2002). *Working Inside the Black Box: Assessment for Learning in the Classroom*. London: nferNelson [Reprinted 2004, *Phi Delta Kappan* **86**, 8–21.].
- Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2003). *Assessment for Learning: Putting It into Practice*. Buckingham: Open University Press.
- Black, P. J. and Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education* **5**, 7–74.
- Black, P. J. and Wiliam, D. (1998b). *Inside the Black Box: Raising Standards through Classroom Assessment*. London: nferNelson [Reprinted 1998, *Phi Delta Kappan* **80**, 139–148.].
- Black, P. and Wiliam, D. (2003). In praise of educational research: Formative assessment. *British Educational Research Journal* **29**, 623–637.
- Blatchford, P., Baines, E., Rubie-Davies, C., Bassett, P., and Chowne, A. (2006). The effect of a new approach to group-work on pupil-pupil and teacher-pupil interaction. *Journal of Educational Psychology* **98**, 750–765.
- Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In Tyler, R. W. (ed.) *Educational Evaluation: New Roles, New Means: The 68th Yearbook of the National Society for the Study of Education*, part 2, pp 26–50. Chicago, IL: University of Chicago Press.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation; the effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology* **58**, 1–14.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research* **58**, 438–481.
- Dawes, L., Mercer, N., and Wegerif, R. (2004). *Thinking Together*, 2nd edn. York: Imaginative Minds.
- DES (2008). Task group on assessment and testing: A report. London: Department of Education and Science and the Welsh Office.
- DES (1988). Task group on assessment and testing: Three supplementary reports. London: Department of Education and Science and the Welsh Office.
- Dweck, C. S. (2000). *Self-Theories: Their Role in Motivation, Personality and Development*. London: Taylor and Francis.
- Greene, J. A. and Azvedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research* **77**, 334–372.
- Hodgen, J. and Marshall, B. (2005). Assessment for learning in English and mathematics: A comparison. *Curriculum Journal* **16**, 153–176.
- Lee, C. and Wiliam, D. (2005). Studying changes in the practice of two teachers developing assessment for learning. *Teacher Development* **9**, 265–283.
- Leung, C. (2007). Dynamic assessment: Assessment for and as teaching. *Language Assessment Quarterly* **4**, 257–278.
- McCombs, B. L. and Miller, L. (2007). *Learner-Centered Classroom Practices and Assessments: Maximizing Student Motivation, Learning and Achievement*. Thousand Oaks, CA: Corwin Press.
- Mercer, N., Dawes, L., Wegerif, R., and Sams, C. (2004). Reasoning as a scientist: Ways of helping children to use language to learn science. *British Educational Research Journal* **30**(3), 359–377.
- Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning processes: Towards a wider conceptual field. *Assessment in Education* **5**, 85–102.
- Van Lier, L. (1996). Contingency. In *Interaction in the Language Curriculum: Awareness, Autonomy and Authenticity*, ch. 7, pp 147–187. Longman: London.
- Wiliam, D. and Black, P. J. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment. *British Educational Research Journal* **22**, 537–548.
- Wiliam, D., Lee, C., Harrison, C., and Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education* **11**, 49–65.

Relevant Websites

- <http://www.assessment-reform-group.org> – Assessment Reform Group.
- <http://www.aia.org.uk> – Association for Achievement and Improvement through Assessment.
- <http://www.kcl.ac.uk> – King's College London Assessment Research Group.
- <http://www.learn2learn.ac.uk> – 'Learning to Learn' Project.
- <http://www.gtce.org.uk> – UK General Teaching Council.
- <http://www.standards.dfes.gov.uk> – UK Government Department for Children Schools and Families.
- <http://www.qca.org.uk> – UK Qualifications and Curriculum Authority.

Formative Assessment and Instructional Planning

B Marshall, Kings College London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The term assessment for learning (AfL), or formative assessment as it is sometimes referred to, has recently gained great currency in education, particularly in the UK. It has become a coverall term for improving learning of pupils and has been adopted into various frameworks including the inspection of schools. One aspect school inspectors consider is whether AfL is in evidence in lessons. In other words, they consider whether or not the teacher has used instructional planning to further the learning of the children in that class. In this way AfL and instructional planning are intimately entwined. AfL cannot take place unless it has been planned for. In this article, assessment and instructional planning are referred to almost, but not completely, synonymously because there are differences.

Two things are important in assessment and instructional planning – what the teacher thinks about what he/she is teaching and what he/she subsequently does in the lesson. What the teacher thinks is crucial for two reasons. The first is that AfL has at its heart, the desire to create an independent learner (James and Pedder, 2006). As part of a major research project, Learning How to Learn, it was found that unless teachers believed it was important that their lessons created autonomous learners, the spirit, as opposed to the letter, of formative assessment did not prevail (Marshall and Drummond, 2006). Black and others, have called it the Trojan horse of AfL (Black *et al.*, 2006).

The second is slightly tangential but nevertheless important. Dweck (2003) observed that there were two contrasting and complementary views of learning – task versus ego involvement and incremental versus entity theories of learning. Those who hold task and incremental views of learning tend to believe that learning makes them smarter, than what they learn from experience; whereas, ego and entity learners concentrate on their ability, which they think of as fixed, or the circumstances in which the task is undertaken. Those who believe in the importance of autonomous learning tend to share Dweck's profile of task/incremental views of learning (Marshall and Drummond, 2006).

High Organization Based on Ideas

However, what teachers actually do in lessons is also important. John Dewey, the originator of progressive education, claimed that lessons and so learning ought to be high organization based upon ideas (Dewey, 1966). In other

words, they should not be just a piecemeal series of activities, however entertaining, but should have some overall rationale or logic behind them. It is this principle that underpins the planning of formative-assessment lessons. John Dewey, in his defense of progressive teaching, wrote that it was a difficult task to work out the kinds of materials, of methods, and of social relationships that are appropriate when trying to discover and put into operation a principle of order and operation which follows from understanding what the educative experience signifies. He was debating the kind of teaching and planning that needs to be in place if one is to put into practice a high organization based upon ideas. In other words, Dewey, the originator of progressive teaching, was saying it is not easy to plan a lesson.

For Dewey, education was about the promotion of independent learners but crucially it was to come about in a social situated setting. The teacher's role was to construct learning in such a manner, through the activities that he/she organized, that pupil autonomy was fostered and so gained. Independent learning, therefore, came about through high organization based upon ideas practically worked out in the lesson. It was less about questioning, feedback, sharing the criteria with learners, and peer- and self-assessment and more about the realization of certain principles of teaching and learning.

Assessment and Instructional Planning in Practice

So what does it mean to plan for assessment to take place in a lesson? If we look at an English lesson, which is based on peer assessment, we can begin to see what Dewey meant when he said that the lesson must have high organization based upon ideas. Each of the activities builds on the one preceding it so that as the pupils come to each new activity, they can use what they have already learned through earlier tasks. There is a rationale to the lesson.

Year-8 lesson (12- and 13-year-olds) pre-twentieth century poem

- Class draw up list of criteria guided by teacher
 - Teacher and learning-support assistant perform poem
 - Pupils asked to critique performance
 - Pupils rehearse performance
 - Pupils peer assess poems based on criteria
 - Pupils perform poems based on criteria
-

Above is an outline of the main events of the lessons.

In the lesson the teacher, Angela, asked the pupils to consider a dramatic rendition of a nineteenth-century poem that they had begun looking at on the previous occasion. The lesson has the potential for pupils to engage with the question of what makes for quality in a piece of work – an issue which is difficult in English and hard for pupils to grasp. Significantly, Angela adopts the procedures of formative assessment identified – sharing the criteria with the learner, and using peer and self-assessment as a means to this end. For this, two activities – modeling and peer assessment – are linked. The modeling activity at the start of the lesson appears to be designed to help pupils know what to do when they peer assess.

Angela began the lesson by asking the pupils to draw up a list of criteria for performing a poem. All suggestions came from the pupils while she probed, challenged, and polished their contributions. For example:

Pupil: You could speed it up and slow it down

Angela: Yes – pace, that's very important in reading
[teacher then writes the word 'pace' on the board].

Interestingly, the Japanese have a useful term for describing such a process – *neriage*, which literally means polishing. In Japan, recapitulating the contributions made by pupils is an important part of teachers' classroom practice. It provides an opportunity for teachers to synthesize the contributions made by different pupils, to interject specific vocabulary, and also to refine or re-contextualize ideas.

The next activity built well on the previous task when Angela and the classroom assistant then performed the poem to the class and then invited pupils to critique their performance based on the criteria. A similar form of probing took place in these exchanges too:

Pupil: It [the performance] was boring.

Angela: What do you mean boring?

Pupil: There wasn't enough expression in your face when the poem was being read or in the reading.

Angela: So what could I have done to make it better?

Pupil: You could have looked and sounded more alarmed

Angela: Like this? [strikes a pose]

Pupil: Not quite

Angela: More like this? [strikes another pose]

Pupil: Yeah.

These three tasks in Angela's lesson the creation of the criteria, the performance of the poem, and the application of the criteria to Angela's and the learning-support assistant's performance all governed the pupils' thinking about what was needed when they acted out the poem themselves and the peer assessment of those performances. So we see again that each stage of the lesson built on the one that had preceded it. There is a logic to

it. Angela's questioning of the pupils provided further opportunity for both her assessment of their performance and the pupils ability to improve on it, as Dweck suggests. Angela takes a more task/incremental view of learning believing that something can always be done or changed to make the pupils in her class understand better what they can do to improve.

Two crucial but subtle elements emerge – the potential scope of the tasks, the way they interlock, and the opportunities these afforded for current and future pupil independence. Although it is hard to separate out the various aspects of the lessons, as they overlap, it is possible to use James and Pedder's (2006) analysis, which has certain headings, as a way of organizing the analysis – making learning explicit, promoting learner autonomy, and performance orientation.

If we start with making learning explicit we find that pupils in Angela's lesson engaged both in technical considerations, such as clarity and accuracy, as well as the higher-order, interpretive concepts of meaning and effect. In addition, the modeling of what was required in Angela's lesson ensured that pupils went beyond an imitation of that model because it challenged them to think about the variety of ways they might enact their interpretation of the poem.

The sequence of activities, each building on the other, guided the pupils in Angela's lesson toward being independent or autonomous learners (the second of our factor headings) because the tasks, including encouraging the pupils to create their own criteria, helped them to think for themselves about what might be needed to capture the meaning of the poem in performance.

Pupils in Angela's lesson, therefore, also began to engage in the more complex issues of any performance be it verbal or written. That is the pupils were asked to explore the relationship between the meaning of a product and the way in which that meaning is expressed – between form and content. This leads us to the final element of performance orientation. Crucially, Angela always described the tasks as opportunities for the pupils to improve their performance. In this way, the activities had an open, fluid feel which corresponded with the notion of promoting pupil autonomy; it reinforced a sense of limitless progress whereby assessment is always seen as a tool for future rather than past performance. This was mainly done by creating tasks, all of which worked effectively together, designed to enable children to enter the subject community or guild (Sadler, 1989). More importantly, Angela has prepared a lesson which is high organization built upon ideas.

The Regulation of Learning and Dialog

Another way of understanding Angela's lessons is to examine the way they were regulated both in the way

they were planned and carried out. Perrenoud (1998) describes the regulation of learning, as the key concept on which AfL is based. To illustrate this, he describes different types of classrooms. In some, those he calls traditional, learning is highly prescribed, the scope of the activities tightly defined. The outcomes of the learning are largely content driven and predetermined and pupils complete a series of narrow activities, which are designed to cover the prescribed learning objectives. There is little opportunity for the pupils to own their own learning and the only information it gives the teacher is a deficit model of what they cannot do according to the narrowly defined terms of reference.

The other type of lesson is much more like Angela's. Here, the tasks are more open ended, thus, the scope for pupils to govern their own thinking is greater, and the possibility for the teachers to provide feedback meaningfully enhanced. For in this type of classroom, "regulation does not include setting up activities suggested to, or imposed on, the pupils but their adjustment once they have been initiated" (Perrenoud, 1998: 88). If we consider Angela's classroom we can see, to an extent, what that means. In her dialog with the pupils, she adjusts the pattern of her teaching once they have been initiated.

And dialog is the final consideration we must take into account. Without effective dialog between pupils and between the teacher and pupil, little AfL can take place. Black and Wiliam observed that "the quality of the interaction [between pupil and teacher] ... is at the heart of pedagogy" (Black and Wiliam, 1998: 16). What counts is the relationship and the aspirations teachers have of their class – the very language of learning.

Vygotsky, the Russian psychologist's view of learning, like Dewey's, was socially constructed and his notion of the zone of proximal development (ZPD) was crucial in understanding the way in which he believed pupils progressed toward autonomy. In particular, Vygotsky saw the teacher's role in facilitating activities to encourage pupils to engage with the lesson as part of this process. This is redolent of Bruner's concept of scaffolding learning within Vygotsky's ZPD. Within the ZPD, Bruner suggested, pupils need to interact with those who are more capable. This interaction with peers and teachers provides a scaffolding for the learner, through which the learner develops the capabilities herself.

If we look at a history lesson, we can begin to see what this might look like. This lesson examined the slave-trade triangle between England, Africa, and the Caribbean. In it, the pupils were being asked to consider what type of evidence certain sources might proffer, how to use such sources, and also about the role of chronology. Toward the end of the lesson, they also began to consider the type of language and register necessary for writing about historical sources. What is significant again is how one activity naturally gives rise to the next – as with Dewey, all

activities are connected, and each relies on dialog either among the pupils themselves or between the pupil and teacher, very much like Vygotsky's ZPD.

The lesson was divided into three main activities – a recap on the previous lesson; in the second and main activity of the lesson, the pupils were given first five, and later four more, contemporaneous pictures that illustrated aspects of the slave trade and finally they were asked to write captions underneath the pictures. Each section had significant discussion but if we consider the dialog in the last of these sections, we see how important both types of conversation are in AfL. As with the English lessons, the teacher, David, modeled the type of writing required before they embarked on the task. Taking what was arguably the first picture of the sequence he asked for contributions. Again, like Angela's lesson, he developed and refined these. He began with:

T Why would you not write 'Slaves walking across a field'?

P Because they could be anywhere.

T So?

P Captured Africans taken to a ship.

T Can we see a ship? How else could we add to it?

P Captured Africans being marched to a ship on the West African coast.

T Have we left anything else out?

P Fellow African tribes capturing and taking their own people to captivity.

What is interesting about this exchange was the way that through questioning, the teacher pushed the pupil to develop the density and precision of the caption. As with Angela and the classroom assistant's performance, David set up an initial response, such as a pupil might make, to be critiqued and developed. Again, through the model he makes the criteria for a good caption clear. But this was done less in terms of what was right and wrong but by what would make for quality in a caption – in this case evidential and inferential density about the slave trade from a visual source. In this way, subject content and processes were blended in the articulation of a caption. Moreover, the use of a caption is, in itself, a stepping stone to enable pupils to use this type of language or discourse within an essay.

David then developed their language or discourse capacity as he went around the class listening to and reading their attempts, again selecting one caption and exchange, as a result of his sampling of pupil engagement, and getting the whole-class feedback to reinforce it. Moreover in doing so, he added another, what might be called moral, dimension to the exercise. As he went around, he had found two girls who had written of a picture, "slaves being sold at auction."

T How can you tell they are being auctioned?

P Because there's a sign that says 'Horses, negroes, cattle.

T What's so bizarre and shocking about how that's set up?

P They're being sold as if they are cattle.

In the whole-class feedback, using the same two girls, he rehearsed this exchange for the rest of the pupils, having focused their attention on the relevant picture.

T What does that sign tell you?

P That they think of people in the same way as they do cows and horses.

T What else might you add to the caption, 'slaves sold at auction'?

P Slaves who are sold at auction are treated like items not people.

The teacher used the girls as experts to enable the rest of the class to develop their own understanding. In this way, although he was orchestrating the commentary, by focusing attention on the girls' observations, he reinforced the sense of the classroom as a community where all learned from one another. At each stage, the teacher uses dialog to develop the pupils.

Conclusion

This article focused on two lessons to identify what assessment and instructional planning can mean. It has considered the overall plan that a teacher must have before embarking on a lesson, but it has also shown that impromptu dialog or conversations within a lesson, once it has started are also vital. Good AfL and the instructional planning that has to go with it depends on the teacher. As Angela commented, "If I've taught a lesson, then I'll go over it, reflect, think, what could I do better next time?" And again, "So I do a lesson with one and then I think, okay, how could I improve that for the next time?" That is the key. For Angela, nothing is fixed or beyond her control, everything is left to play for. If the children she has taught have not learned what she intended, she sees it as her job to revise what she has done and try again,

differently. The classroom is the place where she learns and in learning she considers the way she can improve her performance. It should always be so.

See also: Assessment and the Regulation of Learning; Formative Assessment; Summative Assessment by Teachers; The Relationship Between Assessment and the Organization and Practice of Teaching.

Bibliography

- Black, P., McCormick, R., James, M., and Pedder, D. (2006). Learning how to learn and assessment for learning: A theoretical inquiry. *Research Papers in Education* **21**(2), 119–132.
- Black, P. J. and William, D. (1998). *Inside the Black Box: Raising Standards through Classroom Assessment*. London: NFER.
- Dweck, C. (2003). *Self Theories: Their Role in Motivation, Personality and Development*. Philadelphia, PA: Psychology Press.
- James, M. and Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *Research Papers in Education* **17**(2), 109–138.
- Marshall, B. and Drummond, M.-J. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education* **21**(2), 133–154.
- Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning processes. Towards a wider conceptual field. *Assessment in Education: Principles, Policy and Practice* **5**(1), 85–102.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* **18**, 119–144.

Further Reading

- Alexander, R. J. (2006). *Towards Dialogic Teaching: Rethinking Classroom Talk*. York, UK: Dialogos.
- Black, P. J., Harrison, C., Lee, C., Marshall, B., and William, D. (2003). *Assessment for Learning: Putting It into Practice*. Buckingham: Open University Press.
- Black, P. J. and William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice* **5**(1), 7–73.
- Bruner, J. (1985). Vygotsky: A historical and conceptual perspective. In Wertsch, J. (ed.) *Culture, Communication and Cognition: Vygotskian Perspectives*. Cambridge: Cambridge University Press.
- Dewey, J. (1966). *Experience and Education*. London: Collier Books.
- Vygotsky, L. (1978). *Mind in Society*. Cambridge, MA: MIT Press.
- Vygotsky, L. S. (1985). *Thought and Language*. Cambridge, MA: MIT Press.

Progression and Assessment: Developmental Assessment

M Forster and G Masters, Australian Council for Educational Research, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Developmental assessment – The process of monitoring students' progress through an area of learning so that decisions can be made about the best ways to facilitate further learning. The unique feature of developmental assessment is the use of a progress map.

Progress map – A map (also sometimes referred to as a learning continuum, learning progression, developmental continuum, or progress variable) that describes the nature of development in an area of learning: the knowledge, skills, and understandings of a learning area in the order in which they typically develop.

Introduction

There is no concept more fundamental to teaching and learning than that of progress. This concept is invoked whenever teachers describe, for example, students becoming better readers, using more sophisticated language, becoming more tolerant of others, developing deeper understandings, acquiring higher order skills, solving more difficult problems, or mastering more advanced knowledge. Teachers talk about students developing and improving and they use such words such as better, deeper, and higher to describe the direction of student progress.

This article explores the conceptual and empirical underpinnings, and practice of developmental assessment – an approach to the assessment, reporting, and monitoring of learning that directly addresses the concept of progress through the construction and use of progress maps.

Progress maps describe typical progress in an area of learning: the developing knowledge, skills, and understandings against which students' learning can be assessed and monitored over time. They provide a framework that bridges the apparent conceptual gap between teachers' classroom assessments and the assessments undertaken for system accountability. They provide teachers with a shared understanding of the nature of development across the years of school, and a language for talking about growth with each other, with parents, and with students. They provide education systems with a coherent frame of reference for monitoring learning, for comparing the

achievements of subgroups of the student population, and for setting expectations of performance.

The construction and use of these explicit frameworks of progress can be understood in part in terms of advances in two fields: measurement theory and learning research. Advances in these two fields have converged around a few key ideas:

- learning is an ongoing, continuous process;
- at any given time, in any given domain of learning, every learner is at some point in their learning or development;
- while learners are at different stages and are progressing at different rates, every learner is capable of progressing beyond his or her current level of attainment;
- assessment is fundamentally about establishing where individuals are in their learning/development; and
- progress in most areas of school learning involves the development of deeper understandings resulting in increased abilities to organize, generalize, transfer, apply, and use knowledge.

Definitions and Conceptual Underpinnings

Developmental assessment is the process of monitoring students' progress through an area of learning so that decisions can be made about the best ways to facilitate further learning.

The unique feature of developmental assessment is the use of a progress map. A progress map is the backbone of developmental assessment – the framework against which evidence of student learning is collected, assessed, recorded, reported, interpreted, and monitored.

A progress map (also sometimes referred to as a learning continuum, learning progression, developmental continuum, or progress variable) describes the nature of development in an area of learning: the knowledge, skills, and understandings of a learning area in the order in which they typically develop; the things that teachers can watch for and collect evidence about. The terms developmental assessment and progress maps first appeared in the 1990s in materials published by the Australian Council for Educational Research.

A progress map is a construction that begins in the imagination as an attempt to deal with the complexity and multidimensionality of student development. A progress

map focuses attention on a particular area of learning and proposes that, within that area, individuals have levels of achievement which can never be known exactly, but can be estimated. These levels of achievement can be described and illustrated with the kinds of performances

and student work typically observed at particular levels of attainment from the lower more elementary levels of knowledge, skills, and understandings to more advanced levels. An example of a speaking progress map is provided in **Figure 1**.

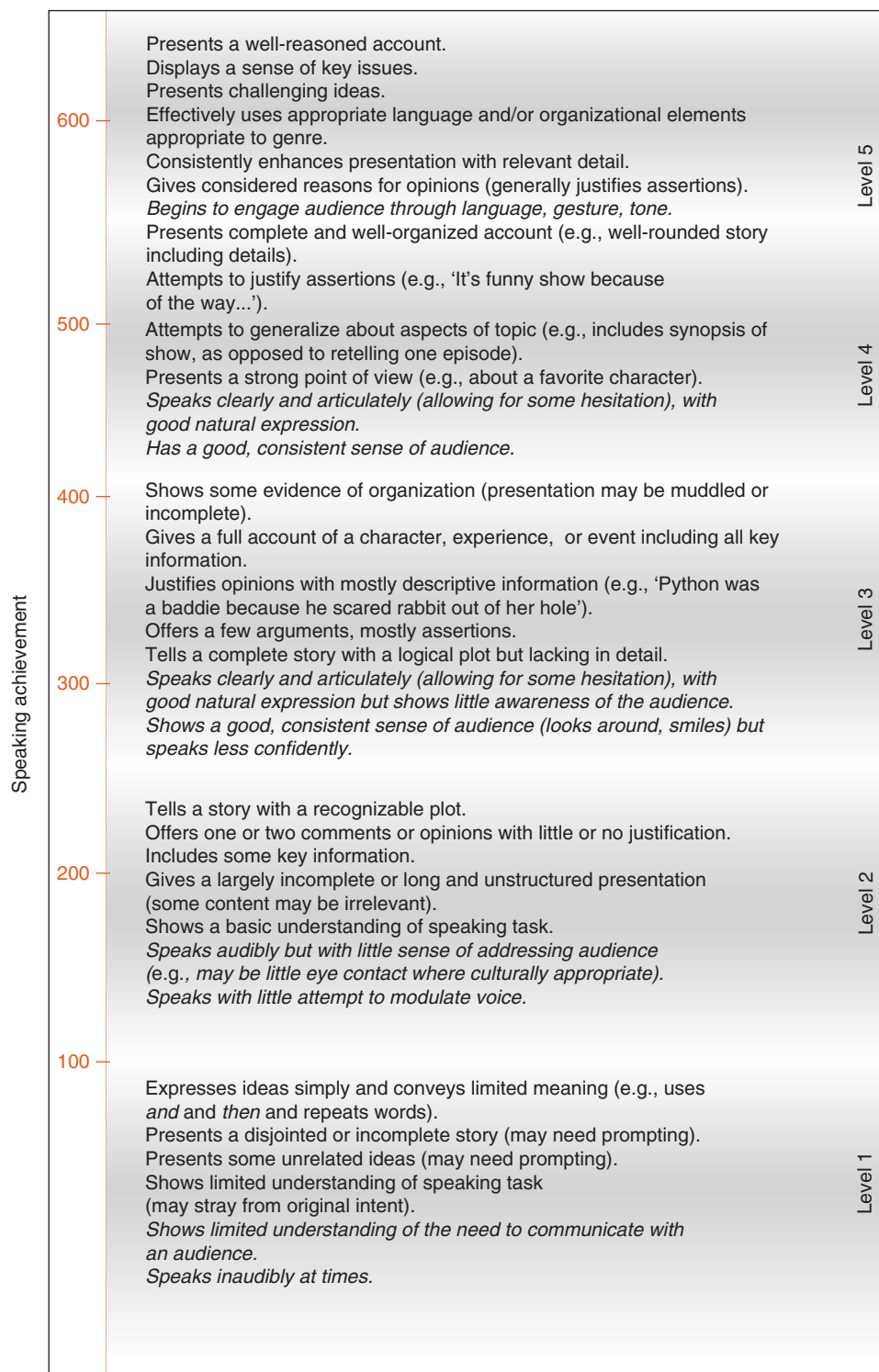


Figure 1 Speaking progress map. Adapted from Masters, G. and Forster, M. (1997). *ARK Progress Maps*. Camberwell: Australian Council for Educational Research.

The structure of many curriculum documents attempts to deal with the complexity and multidimensionality of student development in a similar way, focusing attention on a particular area of learning and describing achievement in that area. Examples include the Hong Kong Curriculum, the United Kingdom Attainment Levels, and the Australian State curriculum documents (e.g., the Victorian Curriculum and Standards Framework).

A progress map is built on the pragmatic observation that it is possible to identify and describe a path of typical progress through an area of learning. That is, although every learner has his/her own idiosyncratic path of development, it is still possible to see a common pathway (the wood as well as the trees). This path of typical progress is not based on a single theory of learning (e.g., Piagetian, constructivist, socioculturalist) or a natural sequence of development. Rather, it is the result of both natural sequences of student development and common conventions for the content and delivery of curricula.

In developmental assessment, progress is monitored against a progress map in much the same way as a child's physical growth is monitored against a framework. From time to time an estimate is made of a student's or a group's location on a progress map and changes in location provide measures of growth over time. The map provides a frame of reference for teachers (and education systems) to monitor an individual's or a group's development and for studying the idiosyncratic development patterns of individuals. It does not provide a prescription for a learning sequence or specify a sequence of learning activities.

If expectations of achievement are set along the progress map (e.g., the level of achievement expected of students by the end of the third grade), then development can be monitored against these expectations also.

Perspectives on School Learning

Developmental assessment represents a paradigm shift in the assessment, monitoring, and interpretation of student learning. Traditionally, assessment and monitoring focus on whether a student has learned what has been taught at grade level – at the end of a lesson, the semester or term, or the school year. For example, the end-of-year report marks the end of the school-year cycle for teachers, for students, and for parents who receive information about their child's achievement before the child moves on and up to the next year level. What have they learned during the year? Have they learned enough to pass? How are they doing relative to others in the class? How well will they cope next year? Do they need to try harder? The child then moves on . . . to the new grade, to the new yearly cycle, to the new teacher.

Developmental assessment takes a rather different perspective on the assessment, monitoring, and interpretation

of student learning. It begins with the understanding that learning is an ongoing process that transcends particular lessons, teachers, classrooms, grades, and even schools and jurisdictions. From this perspective, the main purpose of assessment and monitoring becomes one of tracking a student's development in an area of learning over time. In this context, the interpretations drawn about a student's achievements are drawn primarily in relation to their individual trajectory in an area of learning rather than age/grade expectations or the achievements of others. Where are the students in their learning now, compared with where they were before? What knowledge, skills, and understandings do they demonstrate now in relation to where we want them to travel next in their learning? What misconceptions that might block further learning do they still hold? What do they still need to know and what might be done to support their further development? The progress map provides the framework for teachers, parents, and students themselves to monitor and interpret learning across the years of schooling.

Advances in Modern Measurement

The use of a progress maps as the reference against which student learning is assessed and monitored also represents a paradigm shift of a different kind – an outcome in part of advances in measurement. Perhaps the single most important development in educational measurement in the twentieth century was the shift in focus from stand-alone instruments to underlying reporting frameworks. For much of the twentieth century, educational and psychological measurement was based on the construction of individual instruments (tests, questionnaires, examinations) and the development of norms for each instrument. Each instrument and its accompanying norm tables allowed students to be compared with each other.

The second-half of the twentieth century saw the development of methods for constructing and using reporting frameworks that are not tied to any particular instrument (in much the same way that measures of temperature or length have meanings that transcend the particulars of the instruments used to obtain them). One consequence of this development is that assessment instruments have moved from the foreground to the background: they have become mere vehicles for collecting evidence about students' levels of attainment within a domain of learning.

Advances in measurement theory have also made many of the old dichotomies less relevant. Modern educational assessment is fundamentally about inferring where individuals are in their learning. Students' performances on particular tasks are of interest only to the extent that they can be used to infer levels on attainment on the relevant underlying reporting framework. Inference is thus central to modern assessment, to developmental assessment, and to the use of progress maps.

An individual's level of attainment can be interpreted by reference to the underlying reporting framework (i.e., criterion or standards referencing) or by reference to the achievements of other students (i.e., norm referencing); but with the use of progress maps these different interpretations do not require different approaches to assessment. Similarly, an understanding of the stage a student has reached in his or her learning – including an identification of his or her knowledge gaps and misunderstandings – can be used either formatively to inform further learning or summatively to report that student's current level of attainment or to report progress since some earlier occasion.

In developmental assessment, the progress map provides the structure for this new way of assessing and monitoring learning. Regardless of the instrument used to collect evidence of learning, a student's level of achievement can be estimated (inferred) as a location along the map and can be reported in relation to the domain itself. That is, descriptively in terms of the attainments (knowledge, skills, understandings, attitudes, or values) typically demonstrated by students at that location or level of achievement in the particular domain. When the locations of a number of students are plotted on the same map, an individual's level of achievement also can be interpreted in relation to the achievements of others.

Progression

A progress map describes the knowledge, skills, and understandings of a learning area in the order in which they typically develop. These descriptions, sometimes called learning outcomes (descriptors, or indicators), are broadly described student behaviors, or sets of behaviors. For example, an outcome in physics might be “defines common forces using mathematical expressions and diagrams.” An outcome in writing might be “writes for a range of audiences and purposes.” Outcomes are accompanied by specific examples, sometimes called indicators or pointers, which illustrate and clarify the scope of each outcome. Locations along a progress map can be further illustrated with samples of student work: exemplars or work samples.

To facilitate the reporting and monitoring of student learning, a progress map is usually divided into stages, phases, or levels. The levels are sometimes numbered or given verbal labels that attempt to summarize in one or two words the kinds of observations typically made in each phase. For example, the levels of a reading progress map might be emergent, beginning, bridging, and expanding. The number of levels is always somewhat arbitrary.

Map Construction

The construction process begins with a decision about the number of maps to be defined within a domain.

For example, will a single progress map be developed for science, or will separate maps be developed for different aspects of science such as biological concepts and principles, and physics concepts and principles? The development of progress maps in some domains is more difficult than in others. For example, it is relatively straightforward to describe growth in factual and procedural knowledge when there are existing, well-defined grade expectations on which to draw. But for domains made up of key discipline concepts and principles, or for cross-curricular competencies, there may be less available knowledge about how learning progresses.

The next step is to develop a first draft of the progress map following one of two general approaches. Using the first approach, a top-down approach, teachers and subject-matter specialists draw on their professional knowledge together with relevant research findings about student learning to develop a picture of the sequence in which the knowledge, skills, and understandings of a learning area are typically developed. The draft that results from this approach is based on expert knowledge and opinion. Maps developed in this way need to be tested, refined, and enriched using actual student performances.

The second, bottom-up approach, uses only observations of student responses to develop a picture of progress. The draft that results from this approach is constructed from an analysis of recorded observations and judgments of actual student work. Because draft maps developed in this way are usually based on a small set of assigned tasks, they also need to be refined, enriched, and generalized by reference to a much larger bank of tasks and student work samples.

The step of constructing a first draft is followed by a refinement and enrichment stage. At this stage, two issues in particular are considered: the degree to which the described learning outcomes work together in practice to define a single learning domain against which students' levels of attainment can be estimated and monitored; and the degree to which the sequence of the described learning outcomes is supported by observations of student performance. The usefulness of the map as an assessment framework will depend on clarity about the nature of progression within each domain. If the learning outcomes do not work together to define a single learning domain, or if the map does not reflect progression of learning as observed in classrooms, then it will be difficult for teachers to use the framework to make decisions about attainment. If there is a very poor match – for example, if students typically demonstrate the knowledge, skills, and understandings described at level 4 on a map before they demonstrate those appearing at level 3, the draft map may be of little use as a frame of reference for monitoring learning.

Both issues are addressed through the collection and analysis of information about actual students' achievements. Common refinements include the reordering of misplaced

outcomes, the rewording of ambiguous or unclear statements, and the removal of outcomes that are not helpful in estimating a student's level of attainment (e.g., if the statement is true of all levels of attainment).

A further and ongoing task is to enrich the progress map. Each new assessment task, and its associated student work samples, has the potential to further elucidate the nature of development in an area of learning.

Progress maps developed from empirical evidence (bottom-up frameworks) are now used in international achievement surveys including the Programme for International Student Assessment (PISA), the Trends in International Mathematics and Science Study (TIMSS), and the Progress in International Reading Literacy Study (PIRLS); in national surveys such as the US National Assessment of Educational Progress (NAEP) and, in Australia, in the national literacy and numeracy tests (NAPLAN) and in resources for classroom use. **Figure 2** provides an example from PISA. In this figure, six levels of overall scientific literacy are described. These levels are constructed from an analysis of recorded observations and judgments of actual students' responses to multiple choice questions and constructed response tasks.

Assessment

The progress map provides the framework against which evidence of student learning is collected and assessed. In developmental assessment, records of observations of student achievement provide the evidence needed to estimate or infer a student's level of achievement on a progress map. Developmental assessment recognizes that different assessment methods provide relevant information about different kinds of learning. For this reason, a range of assessment methods will be needed to ensure that observations are relevant. For example, if the learning outcomes require evidence of students' abilities to write for a range of purposes and audiences, then a portfolio will be the most appropriate assessment method. One piece of writing will not provide the specified range of evidence. If the outcomes require evidence of students' abilities to collect, analyze, and report information, then a project may be the most appropriate assessment method.

Developmental assessment also recognizes that a wide variety of methods will be used to judge and record students' work. Some records will indicate whether tasks have been completed correctly; others will record partial

At level 6, students can consistently identify, explain, and apply scientific knowledge and knowledge about science in a variety of complex life situations. They can link different information sources and explanations and use evidence from those sources to justify decisions. They clearly and consistently demonstrate advanced scientific thinking and reasoning, and they are willing to use their scientific understanding in support of solutions to unfamiliar scientific and technological situations. Students at this level can use scientific knowledge and develop arguments in support of recommendations and decisions that center on personal, social, or global situations.

At level 5, students can identify the scientific components of many complex life situations, apply both scientific concepts and knowledge about science to these situations, and can compare, select, and evaluate appropriate scientific evidence for responding to life situations. Students at this level can use well-developed inquiry abilities, link knowledge appropriately, and bring critical insights to situations. They can construct explanations based on evidence and arguments based on their critical analysis.

At level 4, students can work effectively with situations and issues that may involve explicit phenomena requiring them to make inferences about the role of science or technology. They can select and integrate explanations from different disciplines of science or technology and link those explanations directly to aspects of life situations. Students at this level can reflect on their actions and they can communicate decisions using scientific knowledge and evidence.

At level 3, students can identify clearly described scientific issues in a range of contexts. They can select facts and knowledge to explain phenomena and apply simple models or inquiry strategies. Students at this level can interpret and use scientific concepts from different disciplines and can apply them directly. They can develop short statements using facts and make decisions based on scientific knowledge.

At level 2, students have adequate scientific knowledge to provide possible explanations in familiar contexts or draw conclusions based on simple investigations. They are capable of direct reasoning and making literal interpretations of the results of scientific inquiry or technological problem solving.

At level 1, students have such a limited scientific knowledge that it can only be applied to a few, familiar situations. They can present scientific explanations that are obvious and follow explicitly from given evidence.

Figure 2 Summary descriptions for six levels of overall scientific literacy in PISA. Adapted from Organization for Economic Co-operation and Development (OECD). (2006). *PISA 2006 Science Competencies for Tomorrow's World. Volume 1: Analysis*. Paris: Author.

understandings. Sometimes ratings of student work may be recorded either based on separate judgments of aspects of a piece of work (analytic ratings) or based on single overall (holistic) ratings.

As a progress map is a description of the path students typically follow as they progress through an area of learning, the records made for any particular student will only more or less resemble this path and an on-balance (or best-fit) assessment of the student's level of achievement will need to be made. For example, the levels of a progress map for the domain of writing may include descriptions of attainment that relate to the quality of ideas expressed in an extended piece of writing, the organization of those ideas, and the evidence of control over the conventions of punctuation, spelling, and grammar. The writing of an individual student may demonstrate strengths in all areas except spelling and punctuation, requiring an on-balance judgment of the student's level of achievement. In developmental assessment, there is no assumption that students will demonstrate all the skills, knowledge, and understandings below their estimated level of achievement.

Reporting, Interpreting, and Monitoring Learning

The progress map provides the structure against which the point-in-time achievements of an individual, a cohort, or subgroups are reported, and against which progress over time can be monitored. Where both individual and group information is reported concurrently, it is possible to compare the achievements of an individual with the achievement of others (norm referencing).

Figure 3 shows the point-in-time reading achievement of an individual in relation to the achievements of a cohort of students who were assessed at the same time. To the left of the figure is the scale against which achievement is reported; to the right is the alignment of the scale with the levels of the reading progress map. In the center of the figure is the distribution of the cohort of students' achievements. Students with the lowest levels of reading achievement are at the bottom of the picture; students with the highest levels are at the top. The level of an individual's achievement (student #201) is marked with an arrow.

Figure 4 shows the distribution of reading measures at years 3 and 5 for a subgroup of the population: low, medium, and high socioeconomic background.

Figure 5 shows the distribution of reading measures of the same cohort of students at year 3 and then at year 5.

While the progress map, by itself, does not answer questions about the adequacy of students' achievements in relation to expected achievement (absolute benchmarks) age/grade expectations are sometimes overlaid.

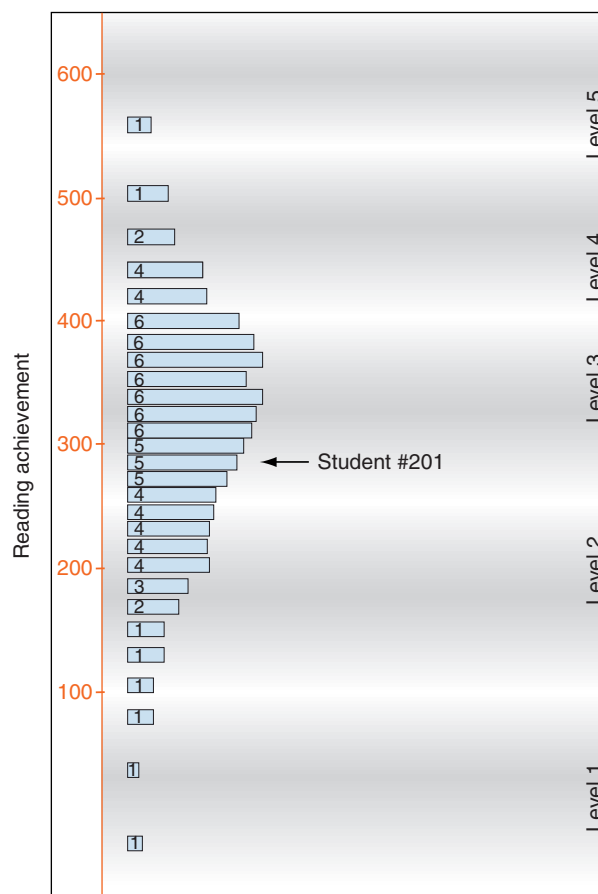


Figure 3 Individual student and cohort achievement in reading.

Figure 6 provides an example. Here, the level of expectation for year 3 students in reading is set at upper level 2 on the progress map (or 250 on the reading scale).

Whether the reported achievements relate to an individual or group, they are interpreted in relation to the map level descriptions – the kinds of knowledge, skills, and understandings typically demonstrated at each level of attainment. The level descriptions provide signposts – where a student has come from and where he/she is yet to advance in learning. What is the nature of the work that a student produces? What is the evidence of his/her learning? What does this evidence tell us about his/her ongoing learning? How much progress has he/she made over time, and what does further improvement look like?

The level descriptions send the message that all learners are on a path of development across the years of school and all learners are capable of making progress. They encourage teachers and parents to focus on the central issue of teaching and learning: what it means to get better (to grow, improve, or progress) in an area of learning. In addition, they provide a framework for students to monitor and reflect on their own progress and to develop metacognitive skills important in taking control of their learning.

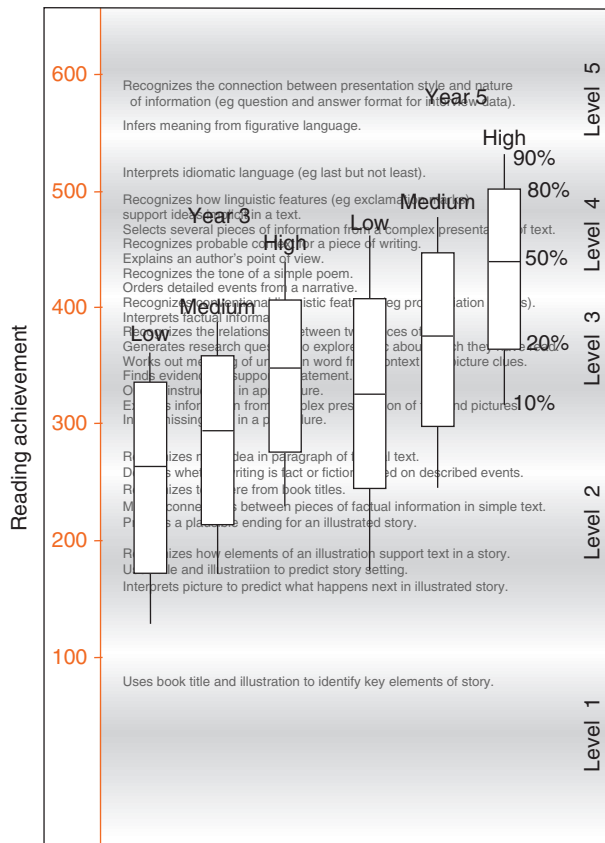


Figure 4 Subgroup achievement by socioeconomic background. Adapted from Masters, G. and Forster, M. (1997). *ARK Progress Maps*. Camberwell: Australian Council for Educational Research.

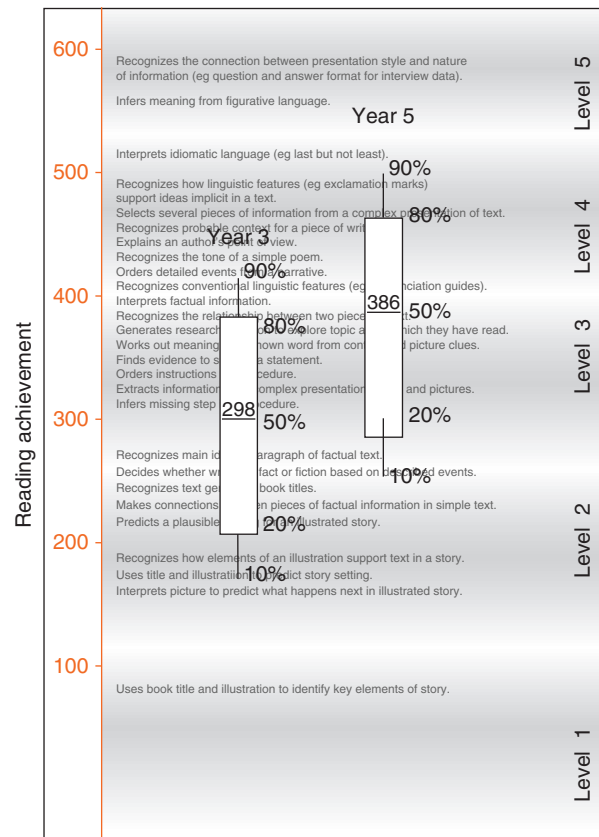


Figure 5 Single-cohort achievement over time in reading. Adapted from Masters, G. and Forster, M. (1997). *ARK Progress Maps*. Camberwell: Australian Council for Educational Research.

Advances in Learning Research

The use of progress maps as a central support for teachers and students is consistent with advances in learning research. Research shows that learning is enhanced when teachers:

- see learning as a continuous, lifelong process;
- adopt a learner-centered approach focused on understanding the learning progress of individuals;
- draw out and work with individuals' current knowledge and beliefs;
- use this knowledge as a starting point for instruction and further learning; and
- carefully monitor individuals' developing understandings.

To do these things effectively, teachers require a deep understanding of learning and how it typically occurs within particular domains: the typical paths students traverse in achieving understanding, the difficulties they commonly encounter, the misconceptions they commonly develop, and effective strategies for promoting further learning. These understandings are an important part of a teacher's pedagogical content knowledge.

Research into the nature of learning in a range of domains is also drawing attention to the kinds of learning that result in deeper understandings of concepts and their application. The development of conceptual understanding allows knowledge to be organized and relationships to be understood. Deep understandings of concepts make it easier for learners to retrieve knowledge, interpret new information, and transfer learning to new situations.

Finally, learning research is identifying the importance of supporting learners to monitor and take control of their own learning. Learning is enhanced when students are supported in identifying where they are in their own learning, in identifying gaps and failures to comprehend, and in monitoring their progress over time. These meta-cognitive skills are important in students' development as lifelong learners.

The implications of these research findings are that assessments must be based not on the testing of disconnected facts (which may actually undermine learning with understanding), but on the depth of students' understandings. New technologies are needed for assessing learning in ways that track the growth of understanding, not just the accumulation of facts. A well-constructed progress map organizes knowledge in a way that brings conceptual

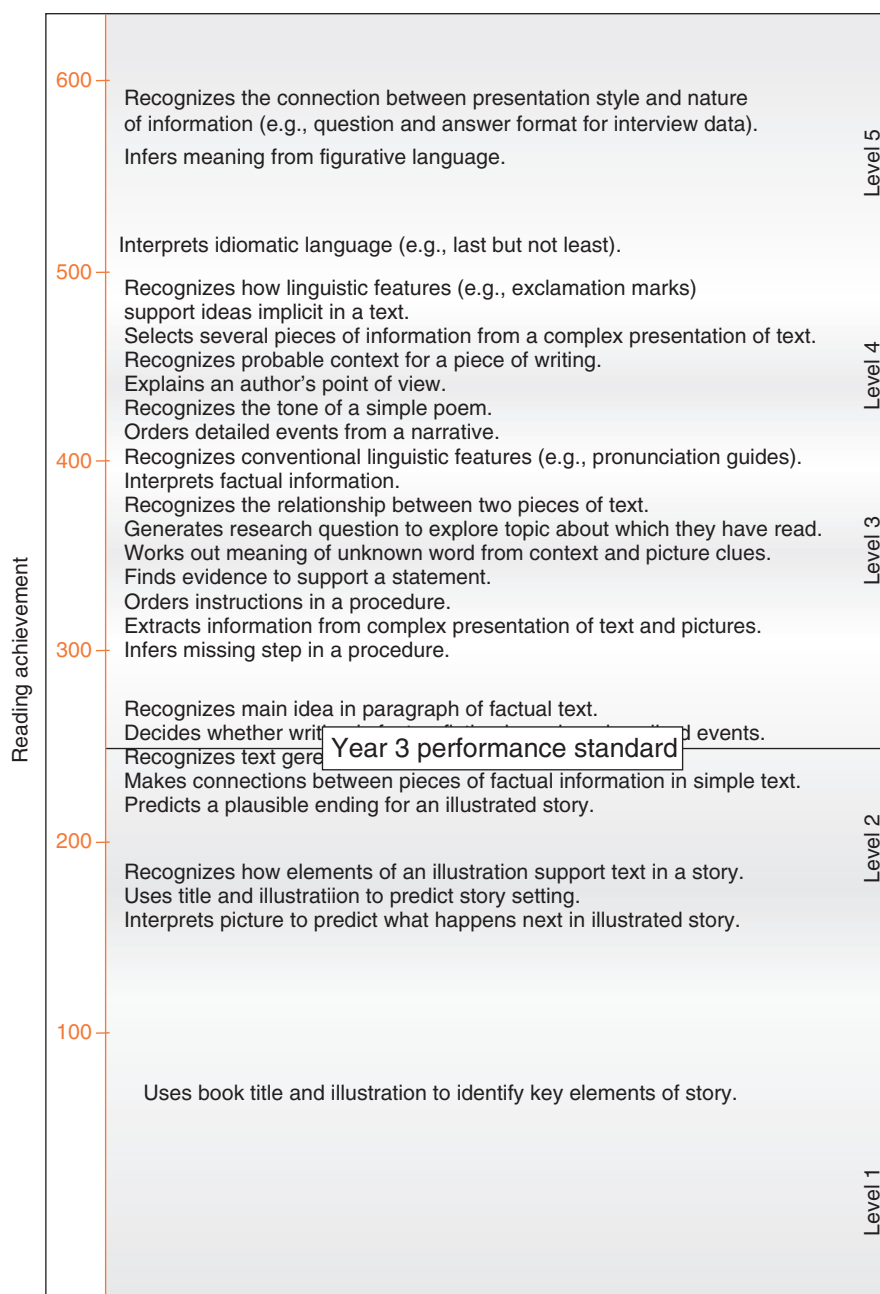


Figure 6 Reading-performance standard.

ideas to the fore, assists teachers to identify students' current beliefs, and provides a map of where students have come from and are going in their learning.

The use of progress maps is also consistent with formative-assessment research which suggests that in order to improve learning, teachers should discuss with pupils the purpose of their learning and provide feedback that will help the learning process; encourage students to judge their work by how much they have learned and the progress they have made; help students to understand where they are in relation to learning goals and how to

make further progress; and give feedback that enables students to know the next steps and how to succeed in taking them. Each of these strategies is facilitated by the use of a progress map.

Summary

Significant advances have occurred in recent decades in methodologies for assessing student learning. These advances can be understood in terms of advances in

measurement theory and in our understanding of learning itself. Central to modern approaches to assessment is the construction of research-based assessment and reporting frameworks in particular domains of learning. These frameworks capture and make explicit the nature of progress within a domain and provide frames of reference for monitoring learning across the years of school. They support teachers in identifying and working from individuals' current levels of knowledge and understanding. Consistent with current understandings of learning, these frameworks are most useful when they describe progress toward deeper conceptual understandings of subject matter. Empirically based reporting frameworks also play an important role in the communication of students' levels of achievement and progress and in supporting learners to monitor their own learning.

The key features of these frameworks (or progress maps) are that they:

- represent progress or learning as a continuous process (i.e., occurring on a continuum) across the years of school;
- describe the direction and nature of progress within a domain (i.e., what it means to learn or develop);
- provide a framework for monitoring growth and change;
- are research-based (i.e., are developed from empirical studies of how students learn);
- provide a frame of reference for monitoring and describing progress independently of the instruments or methods of assessment used; and
- provide a shared basis for describing and communicating achievement and progress.

See also: Challenges of Developing and Implementing Formative Assessment Practices in Schools; Educational Measurement: Overview; Portfolio Assessment.

Further Reading

Assessment Reform Group (2002). *Testing, Motivation and Learning*. Cambridge: University of Cambridge Faculty of Education.

Bodey, W., Darkin, L., Forster, M., and Masters, G. (1997). *DART English (Middle Primary). Developmental Assessment Resource for Teachers*. Camberwell: Australian Council for Educational Research.

Bransford, J. D., Brown, A. L., and Cocking, R. R. (eds.) (2000). *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. Washington, DC: Committee on Developments in the Science of Learning/National Academy Press.

Forster, M. (2000). A developmental assessment approach to assessing and reporting students' speaking skills. In Smith, P. (ed.) *Talking Classrooms: Shaping Children's Learning through Oral Language Instruction*, pp 141–156. Newark, DE: International Reading Association.

Forster, M. and Masters, G. (2002). Accepting the challenge: Assessing the moral and ethical outcomes of schooling. In Pascoe, S. (ed.) *Values in Education College Year Book 2002*, pp 140–152. Deakin West, ACT: The Australian College of Educators.

Forster, M. and Masters, G. (2004). Bridging the conceptual gap between classroom assessment and system accountability. In Wilson, M. (ed.) *Towards Coherence between Classroom Assessment and Accountability. 103rd Yearbook of the National Society for the Study of Education, Part II*, pp 51–73. Chicago, IL: University of Chicago Press.

Forster, M., Mendelovits, J., and Masters, G. (1994). *DART English (Upper Primary) Developmental Assessment Resource for Teachers*. Camberwell: Australian Council for Educational Research.

Griffin, P. (2003). Profiles and reporting in a developmental assessment framework. In Zbar, V. and Mackay, A. (eds.) *Leading the Education Debate*. Melbourne: IARTV.

Masters, G. and Forster, M. (1996). *ARK Developmental Assessment*. Camberwell: Australian Council for Educational Research.

Masters, G. and Forster, M. (1997). *ARK Progress Maps*. Camberwell: Australian Council for Educational Research.

Masters, G. N., Adams, R. J., and Wilson, M. R. (1990). Charting student progress. In Husine, T. and Postelthwaite, T. N. (eds.) *International Encyclopaedia of Education: Research and Studies*, vol. 2, pp 628–634. London: Pergamon Press.

Masters, G. N. and Forster, M. (1997). *Mapping Literacy Achievement: Results of the 1996 National School English Literacy Survey*. Canberra: Commonwealth of Australia: Department of Employment, Education, Training and Youth Affairs.

Organization for Economic Co-operation and Development (OECD). (2006). *PISA 2006 Science Competencies for Tomorrow's World. Volume 1: Analysis*. Paris: Author.

Pellegrino, J., Chudowsky, N., and Glaser, R. (eds.) (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: Committee on the Foundations of Assessment/ National Academy Press.

Thomson, S. and De Bortoli, L. (2008). *Exploring Scientific Literacy: How Australia Measures up. The PISA Survey of Students' Scientific, Reading and Mathematical Literacy Skills*. Camberwell: Australian Council for Educational Research.

Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Erlbaum.

Wilson, M. and Draney, K. (2004). Some links between large-scale and classroom assessments: The case of the bear assessment system. In Wilson, M. (ed.) *Towards Coherence between Classroom Assessment and Accountability. 103rd Yearbook of the National Society for the Study of Education, Part II*, pp 132–154. Chicago, IL: University of Chicago Press.

Students' Rights in Assessment and Evaluation

J Gardner, Queen's University Belfast, Belfast, Northern Ireland

B Holmes, Hamdan Bin Mohammed e-University, Dubai, United Arab Emirates

© 2010 Elsevier Ltd. All rights reserved.

If rights are interpreted strictly as legally supported entitlements, enacted through statute, then this article should probably end at this point. As far as we are able to determine, there are no specific rights for students in relation to the assessment or evaluation of their learning in any nation's legislation. (Arguably, the terms assessment and evaluation are nuanced in meaning but for the most part, assessment will be used in the article to relate to any process designed to appraise student learning.) However, there do exist educational principles, widely endorsed standards and codes of practice, and at least one legally binding set of international articles that can be used to underpin a discussion of students' rights in assessment.

Most jurisdictions have education-related rights based on statutory or common law, which may apply to assessment contexts depending on the circumstances. Some of these rights have considerable case law, most notably in the US and to a lesser extent the UK, Europe, and Australasia. They cover issues such as discrimination (numerous instances of disability, race, and gender actions over the past half century), parents' versus children's rights, free speech on campus, and youth culture. However, the most amenable measures, which afford at least the potential for legal protection of students in contexts involving assessment, are espoused under two internationally respected codes: the Standards for Educational and Psychological Testing (AERA *et al.*, 1999) and the Code of Fair Testing Practices in Education (JCTP, 2004). A third axis of potential protection derives from two articles of the United Nations Convention on the Rights of the Child (UNCRC, 1990), namely articles 3 and 12. These standards, code, and articles are discussed below, but first it is worth outlining why students should have the right to a greater say in the assessment of their learning.

Purposes of Assessment

Assessment can be argued to have many purposes, roughly positioned along a spectrum that ranges from the ad hoc question-and-answer sessions of teachers and students in classrooms to the formal examination style of assessment that tests students' levels of achievement after a period of learning activity. The former is generally located at what is called the formative end of the spectrum, where the purpose is primarily to provide support for students' learning, while the latter is generally located at the

summative end of the spectrum, where the purpose is primarily to identify a measure of the students' learning.

Assessment as Support for Learning

Over the past 10–15 years, there has been a growing awareness that the primary role of assessment should be to support learning. This realization is emerging as the result of a number of key developments, most notably the rise in importance of formative assessment, or assessment for learning as it has become known. A growing body of research-informed theory and knowledge has been accumulating in the field of formative assessment in relation to its practice and implications for policy (Gardner, 2006). In common with pedagogical approaches such as dynamic assessment and dialogic teaching, assessment used as support for learning promotes a variety of processes involving increased student participation and their owning of the learning activities. These include: sharing in the process of identifying learning objectives and success criteria between the teacher and the students; the development of peer and self-assessment; effective questioning, listening, and feedback techniques; and a conscious determination on the part of the teacher to recognize and affirm the full range of student achievement. Such processes promote the students' voice in their learning and assessment, while seeking to maximize engagement, not just with the lesson activities and content but with each other, with the teacher, and with the learning process itself.

The theory of formative assessment eschews the traditional didactic model of passive learners and directive teachers, instead giving the students their say in important aspects of their learning and its assessment. The argument draws on widespread evidence that formative assessment approaches can contribute to higher levels of student motivation, achievement, and participation.

Assessment of Learning

At the other end of the assessment spectrum, assessment of learning, which is generally summative in design and intention, largely continues to illustrate the phenomenon of doing to rather than with students. Such summative assessments may have several purposes but generally speaking they come at the end of a period of learning

(end of semester, end of year, etc.), are formal compared to classroom assessments, and purport to be objective and trustworthy. The underlying traditions of assessments for summative purposes include isolating students from sources of advice, support, or knowledge, and requiring them to address a set of tasks (questions, essays, problems, etc.) in time-bound circumstances, usually by means of pen-and-paper tests. Such conditions do not readily enable students to participate in the design or the content of the assessment. However, once the work is assessed, the students do enjoy some widely accepted entitlements, including specified circumstances in which they can seek to have grades or scores reviewed. In most national-education systems today, the main purposes of summative assessments may be summarized under two headings: accountability and credentialing.

Accountability Purposes

There is a variety of purposes to which assessments may be put and the most topical are arguably those that serve accountability. Aside from the traditional accountability of learners (how well they have performed), the last couple of decades has witnessed vigorous programs in the US and UK that incorporate the assessment of student learning to hold teachers, schools, and school authorities to account. One of these, the hugely important No Child Left Behind Act (NCLB) 2001 in the US, is silent on students' rights or perspectives on any aspects of its processes or impact, including assessments, though parents' rights are exhaustively detailed and secured. It has been in operation for 6 years without any noticeable attention to the students' perspectives on its processes or impact. The US NCLB and the UK Making Good Progress programs are typical of many recent educational policy developments around the world, inasmuch as they set targets for schools, based on the dubious proposition that testing and increased accountability raises standards. NCLB has attracted from educationalists who see it as damaging to its main disadvantaged target groups and to individual students (e.g., Houston, 2007; Nichols *et al.*, 2005).

It is only a small step from local accountability to making whole nations accountable for the success of their education provisions through, for example, international comparative programs such as the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA). Almost without exception in these cases, the students are the cannon fodder in a battle between policy-makers and the school communities they oversee. They have next to no voice (except perhaps through the ventriloquism of their advocates) and precious little in the way of rights. Houston (2007), for example, claims that one of the deadly sins in the US NCLB is the "subjecting

of children to days of examinations annually, with the time taken away from instruction" while Skidmore (2003: 34/35) in the UK cites the statistic that "... an average pupil [in England] will now take 70 exams by the time they finish school [at 16]." In none of these circumstances has the student any rights to constructive participation in, for example, contributing to the identification of assessment criteria or design of assessment items and tasks. They are merely the subject of the tests.

Doing all of this to students has considerable cost implications. For example, the UK Association of College and School Leaders claims that many schools in the UK are spending more on examination fees than on learning resources, and that costs for examinations have risen by 51% in the past 3 years.

PriceWaterhouseCoopers (PWC, 2004) has modeled the examinations system in England (ranging across a wide variety of testing programs) and reported that in the year 2003–2004, its direct cost was £370 million with an additional £240 million on time costs, but excluding teachers' time in their assessment-support activities. There were also huge environmental costs in, for example, the distribution to test centers of 4.2 million test papers, 80 000 packages, and 237 million A4 sheets (Durant, 2003). In the US, Wallis and Steptoe (2007) estimate that the costs of independent state testing programs in the US (separate to National Assessment of Educational Progress (NAEP) testing) is in the region of US\$600 million.

Although the costs of external assessments may give cause for concern, the justification for such a huge investment in testing appears to be its perceived reliability. What, then, if external tests are not as reliable as the confidence in them implies? Research shows that systemic errors arising from the difference between students' observed score on specific test-taking occasions and their true score, as would be determined by an infinite number of test-taking occasions, can have considerable impact on results. These standard errors of measurement can cause an astonishing degree of misclassification in high-stakes testing (Black and Wiliam, 2006). In 2006, some 65 000 UK General Certificate in Secondary Education (GCSE) results were appealed and over 14 305 were regraded. Though a small percentage of the huge number of individual examinations taken (6.5 million), the fact is that many more students were likely to have been misgraded but they either did not know this or were unaware or did not take up their right to a review.

The most important observation in this situation is that these students exercised their right to request a review of their grading. This right has been offered by the UK examination authorities to students for some time. However, to assume that such a right is granted primarily from a natural-justice perspective would probably be an overly generous analysis. Newton and Whetton's (2002) four objectives for student appeals illustrate that the right of

appeal has less of a rights perspective than a pragmatic purpose. They identify appeals as serving a measurement objective, which is designed to maximize the measurement accuracy of the test in question (i.e., more accurate grade attribution) and a political objective, which is intended to safeguard public and professional confidence (i.e., the “if we find a mistake, we fix it!” assurance). There is also an educational objective for student appeals, one of contributing to teachers’ professional development by sharing the outcomes and reasons for appeal decisions. Finally, there is a tip of the hat in the direction of students’ interests with the objective of aiming to minimize any negative psychological consequences for them.

Credentialing Purposes

On an individual level, the credentialing purpose of assessment, that is, the award of qualifications in academic or vocational settings, is perhaps the most pressing for many students. Test programs that lead to qualifications or credentials are generally taken by students because they need them to demonstrate that they have a certain set of skills or that they have the necessary qualifications to progress to the next step in their careers. For some years now, qualifications-oriented programs have been expanded to include key skills (sometimes known as basic or transferable skills) such as literacy, numeracy, information technology, and working together. The concept of twenty-first-century skills is also gaining ground, for example, Baker’s “... adaptive problem-solving, risk assessment, managing distraction, self-management and changeable roles” (Baker, 2007). Such skills are simply not being assessed properly at present, if at all, and much needs to be progressed before students can avail of their entitlement to valid and reliable assessment in important new areas of learning.

The pursuit of qualifications may also have the objective of seeking admission to a college or university course, to a selective school or to the world of employment. As such they are high-stakes tests or assessments and it is entirely reasonable for students to expect and indeed demand accurate and valid assessments of their performance. To this end, two codes of good practice, the Code of Fair Testing Practices in Education (JCTP, 2004) and the Standards for Educational and Psychological Testing (AERA *et al.*, 1999), have been widely endorsed as guides to good assessment practice.

Writ large in the American Educational Research Association (AERA) standards is a recognition of the perspectives of the different stakeholders in assessment contexts; from parents, teachers, and students through to the test developers and those who use them for assessment purposes. Most importantly, for our purposes, they also provide a comprehensive overview of the rights and

responsibilities of students (or test-takers as the standards more precisely define those who are assessed) whether in educational, legal, or employment contexts. For example, the first reliability standard, Standard 2.1 (S 2.1) states that:

... estimates of relevant reliabilities and standard errors of measurement or test information functions *should* be reported (our emphasis) (AERA *et al.*, 1999: 31).

It would seem reasonable that students should be able to determine the accuracy of any measurements of their performance, but in some significant instances, most notably the formal assessments of school students of the UK, the examination authorities are silent on the technical performance of the tests. Examination processes remain arcane with little or no information being made available in relation to reliability or validity, and no recognition of the rights of students to have access to this type of information. Yet, as we mention above, reliability is demonstrably a problem.

The large majority of the 264 standards are characterized in their language by the use of the word *should*. If this was the imperative *should*, and if there was a means by which the standard could be enforced (which there currently is not), then they would underpin legally protected student rights. However, there is little prospect of them being formally enshrined in law and it will likely remain an advisory *should*. The term *rights* is therefore something of a misnomer, and its substitution by a term such as *expectations* is arguably more accurate. Although not immediately obvious, the language of the standards is quite circumspect, carefully written to avoid any charge of being dogmatic or any potential of being used inappropriately, such as might be the case in attempts to override professional judgment. In situations in which assessments are being subjected to legal scrutiny or challenge, the standards are crystal clear about their advisory role: “... the intent of the Standards is to offer guidance for those judgments” (AERA *et al.*, 1999: 4).

The standards may stop short of speaking to legally enforceable rights but they are no less important in informing professional judgment and in their advocacy of fair and good practice (see, e.g., Camara and Lane, 2006). Among the instances of good practice in seeking to protect student interests and rights, are those standards that specifically address issues of information, consent, privacy, and appeals. For example, the standards articulate the view that if any information about an assessment is given to any candidate, it should be given to all of them (S 8.1). The information should cover all important aspects of the assessments including the intended use of the results, scoring criteria, confidentiality, re-sits, and the release of results (S 8.2). The consent of the student to be assessed is seen as a major principle of procedural propriety, though several potential exclusions are offered: when

the testing is mandated by government; when the testing is a routine part of school activities; and when consent is obviously implied. For example, consent may be assumed in an employment-selection context when the act of applying is deemed to include consent to undertake any assessment that is a part of the process (S 8.4). There should be an entitlement to have the privacy of individuals' results maintained (S 8.5) and secure arrangements for storage of the results (S 8.6). In decision situations, the student should be entitled to receive details of the results and scores, and of their interpretation (S 8.9). Such information should also inform students of any recourse to appeals, challenges, and reviews (S 8.10).

While one whole chapter (chapter 8) of the standards is given over to the rights and responsibilities of test-takers, the other chapters also contain standards that speak to students' rights. For example, if the integrity of a test score is challenged, the student should be informed of any rights to appeal and to have legal representation (S 11.11). Standard 13.14 also states that information on the extent of possible misclassification and errors in measurement should be released.

Much of the intention behind developing the standards was to ensure reliable, fair, and valid testing of students and others, particularly in high-stakes settings, but they are nonetheless daunting for the average student, parent, or person who is not *au fait* with assessment terminology. As a means of making the underpinning principles of the standards more accessible, the Joint Committee on Testing Practices (JCTP) issued their 1988 Code of Fair Testing Practices in Education. This was updated in 2004 and "...applies broadly to testing in education (admissions, educational assessment, educational diagnosis, and student placement)" (JCTP, 2004: 2).

In essence, the JCTP code addresses the importance of students (test-takers) having the fullest information possible, including what rights they may be granted in terms of appeals, having their scripts returned, or being able to take re-sits. As with the standards from which they are derived, they are also advisory. However, in combination with the standards they present a formidable array of widely accepted good practices in assessment, to which most courts and legally constituted tribunals will have due regard.

United Nations Convention on the Rights of the Child

As mentioned above, the United Nations Convention on the Rights of the Child is the third axis of possible protection for students' rights in assessment, or at the very least those students who are under 18 years of age, which is the definition of the child under the Convention. At first sight, this might appear an unlikely bastion for

students' rights in assessment, focusing as it does on issues such as children's fundamental rights to life, freedom from abuse, and from recruitment to armies. Ratified by 192 countries, with the remaining two as signatories (the US and Somalia), courts have less leeway to reject rights actions that arise under the relevant articles of the Convention. Aside from the education-specific articles such as 28 and 29, which respectively assert the right to primary education and to the development of personality and talents, the key articles for our purposes are:

1. *Article 3.* In all actions concerning children, whether undertaken by public or private social-welfare institutions, courts of law, administrative authorities, or legislative bodies, the best interests of the child shall be a primary consideration.
2. *Article 12.* States/Parties shall assure to the child who is capable of forming his or her own views the right to express those views freely in all matters affecting the child, the views of the child being given due weight in accordance with the age and maturity of the child.
3. *Article 3.* This has the potential to call into doubt the legality of the testing regimes currently holding sway in the US and England (not, it should be noted, in the three other UK home countries, Scotland, Northern Ireland, and Wales. In these nations, much of the London-led accountability testing has been abandoned as counter-educational by their devolved governments). Many commentators have called the preponderance and nature of the testing into serious question and some researchers are beginning to bring a United Nations Conventions of the Rights of the Child (UNCRC) lens to bear on the issue (e.g., Yates, 2006). Clearly there is an argument that state testing programs in the US, and the national-curriculum tests of England, are not in the best interests of the young students concerned. Probably more to the point, however, is the tenuous validity (and ethical basis) of using such tests of students' learning for a purpose that is some way removed from their individual learning and progression, that is, school and state accountability.

Article 12 brings the matter of assessment propriety into particular focus. High-stakes tests are so called because they have the potential for considerable impact, whether positive or negative, on students' educational or career choices. More often than not, however, the students concerned have little say either in formulating or responding to any aspect of the major decisions that directly affect them and which arise from such tests. The extent to which re-sits are offered, actual annotated test scripts are returned for scrutiny, information on scoring is shared, or feedback on performance is offered, is generally in the gift of the examiners – and not an established right of the students. The AERA Standards and JCTP Code set out clearly what should happen and it is possible that the UNCRC will

provide a legal platform from which to challenge the more arcane of assessment practices today. However, one major obstacle stands in the way – the special place given to academic judgment.

Academic Judgment

Unless there are procedural or technical grounds for challenging a grade or score, it is most likely that students appealing the assessment of their performance in a test will face the dictum, especially in higher education institutions, “We cannot accept challenges to academic judgment.” For example, the UK’s Quality Assurance Agency (with a remit for the quality assurance of teaching and learning in UK universities and colleges) say simply that “[Since] most institutions do not allow appeals against the exercise of academic judgment . . . it would therefore be helpful to define what constitutes an academic judgment” (QAA, 2008: 7). It defines an appeal as “. . . a request for a review of a decision of an academic body charged with decisions on student progression, assessment and awards” (QAA, 2008: 4). Another body, the Office of the Independent Adjudicator for Higher Education (England and Wales) similarly states: “We cannot look at a complaint if . . . it relates to a matter of academic judgment” (OIAHE, 2008) yet it does allow what it terms academic appeals. The adjudicator can recommend compensation and in the examples of the 32 completed cases provided on the website, compensation payments of up to £9000 for students are recorded. Although these are clearly not court-of-law judgments, they are beginning to form a body of case law in the context of university students’ rights in the assessment of their performance.

Disbarring appeals against academic judgment is not a simple expedient on the part of examination authorities, whether university boards of examiners or major testing agencies. Judgments of student performance, other than the purely correct/incorrect dichotomies of aspects of some disciplines such as mathematics, are generally supported by processes such as those that promote inter-judge (inter-rater) reliability. An example of this process is known in the UK as moderation, in which examiners’ grades and scores are checked and calibrated by other examiners. Agreement trials are processes in which a group of examiners take samples of student scripts and grade them together, aiming to develop a connoisseurship among the examiners on the standards to be applied. There are also systems of double marking (where two examiners independently grade the work) and anonymous marking (to ensure that knowledge of the student does not influence the examiner’s grading). The purpose of all such procedural devices is to ensure that the assessments are as reliable as possible and executed fairly across the

test-taking group. Notwithstanding these efforts, however, much of the process remains largely hidden and unknown to the student. Disappointment with their grading will often prompt students to appeal, to be told that only technical or procedural grounds are allowed.

Courts for their part are also notoriously wary of becoming involved in challenging the internal judgments of schools, universities, or examination agencies. Unless there is *prima facie* evidence of malpractice, which is amenable to statutory or common law redress (e.g., some form of discriminatory practice), they will often not proceed. Even if there is an assessment appeal purporting to be a contravention of UNCRC Article 3 or 12, the prospect of a case coming down to the consideration of the reliability or validity of an academic judgment may well jeopardize any potential finding for the student concerned, specifically because it is considered so difficult to mount an effective challenge to academic judgment.

Summary

Students’ rights in relation to the assessment of their learning or performance do not specifically exist in legislation in most jurisdictions. However, the long-standing and highly respected AERA and JCTP guidelines enshrine many expectations for fair treatment. Arguably, these have the intention of prescribing entitlements or rights and the rationale seems eminently reasonable for expectations such as receiving academic information on tests (e.g., the curriculum covered and assessment criteria), receiving technical information on tests (e.g., reliability and validity), having annotated scripts returned, having recourse to appeals, and having opportunities to re-sit a test. Two main factors seem to be in play: a keen sense of natural justice among educators and an increasing academic and professional understanding of the extent to which uncertainty in scores undermines the pursuit of absolute measures of performance.

As time goes on, the hope must be that increased compliance with the AERA Standards and JCTP Code will reduce the potential of unfairness in treatment or outcome in assessment of students. But when needed, there is the increasing prospect of legal strengthening of students’ rights, especially in high-stakes testing. This will most likely occur through bringing to bear national instruments such as the Bill of Rights or international treaties such as the UNCRC, both of which have seen considerable impact in contexts of discrimination in education over recent years.

See also: Assessment and the Evaluation of Institutional Effectiveness; Formative Assessment; Summative Assessment by Teachers.

Bibliography

- AERA, APA, and NCME (1999). *Standards for Educational and Psychological Testing*. American Educational Research Association, American Psychological Association and National Council on Measurement in Education. Washington, DC: American Psychological Association.
- Baker, E. (2007). *The End(s) of Assessment*. Presidential Speech, American Educational Research Association Annual Conference, 11 April, Chicago. <http://www.cmcgc.com/Media/Synch/270409/40/default.htm> (accessed August 2009).
- Black, P. and Wiliam, D. (2006). The reliability of assessments. In Gardner, J. (ed.) *Assessment and Learning*, pp 119–131. Thousand Oaks, CA: Sage.
- Camara, W. J. and Lane, S. (2006). A historical perspective and current views on the standards for educational and psychological testing. *Educational Measurement, Issues and Practice* 25(3), 35–41.
- Durant, D. (2003). A comparative analysis of key stage tests and teacher assessments. *Paper Presented at BERA Annual Conference*, Heriot-Watt University. www.leeds.ac.uk/educol/documents/00003153.htm (accessed August 2009).
- Gardner, J. (ed.) (2006). *Assessment and Learning*. Thousand Oaks, CA: Sage.
- Houston, P. D. (2007). The seven deadly sins of No Child Left Behind. *Phi Delta Kappan* 88(10), 744–749.
- JCTP (2004). *Code of Fair Testing Practices in Education*. Washington, DC: Joint Committee on Testing Practices. <http://www.apa.org/science/FinalCode.pdf> (accessed August 2009).
- Newton, P. and Whetton, C. (2002). Technical and practical implications of the right to appeal test results. National Foundation for Educational Research. Slough: NFER. <http://www.nfer.ac.uk/publications/other-publications/conference-papers/technical-and-practical-implications-of-the-right-to-appeal-test-results.cfm> (accessed August 2009).
- Nichols, S. L., Glass, G. V., and Berliner, D. C. (2005). High stakes testing and student achievement: Problems for the No Child Left Behind Act. Arizona State University: Educational Policy Studies Laboratory. <http://epicpolicy.org/files/EPSSL-0509-105-EPRU.pdf> (accessed August 2009).
- OIAHE (2008). Resolving student complaints, Office of the Independent Adjudicator for Higher Education. <http://www.oiahe.org.uk/summary.asp> (accessed August 2009).
- PWC (2004). *Financial Modelling of the English Exams System 2003–04*. PriceWaterhouseCoopers. http://www.qca.org.uk/qca_9940.aspx (accessed August 2009).
- QAA (2008). Code of practice for the assurance of academic quality and standards in higher education, section 5: Academic appeals and student complaints on academic matters – October 2007. <http://www.qaa.ac.uk/academicinfrastructure/codeofpractice/section5/default.asp> (accessed August 2009).
- Skidmore, P. (2003). *Beyond Measure: Why Educational Assessment is Failing the Test*. London: DEMOS. <http://demos.co.uk/files/beyondmeasure.pdf> (accessed August 2009).
- United Nations Convention on the Rights of the Child (UNCRC) (1990). Office of the High Commissioner for Human Rights. United Nations: New York. <http://www.unhchr.ch/html/menu3/b/k2crc.htm> (accessed August 2009).
- Wallis, C. and Steptoe, S. (2007). How to fix No Child Left Behind. *Time* 169(23), 34.
- Yates, T. L. (2006). A criticism of the No Child Left Behind Act: How the convention on the rights of the child can offer promising reform of education legislation in America, 5. *Whittier Journal of Child and Family Advocacy* 399.

Further Reading

- Cumming, J. J. and Dickson, E. A. (2007). Equity in assessment: Discrimination and disability issues from an Australian legal perspective. *Education and the Law* 19(3/4), 201–220.
- Cumming, J. J. and Mawdsley, R. D. (2005–2006). Student rights and parent rights in education in Australia. *Australia and New Zealand Journal of Law and Education* 10(2);11(1), 37–54.
- Gardner, J. and Cowan, P. (2005). The fallibility of high stakes 11-plus tests in Northern Ireland. *Assessment in Education* 12(2), 145–165.
- Kornhaber, M. L. (2004). Appropriate and inappropriate forms of testing, assessment, and accountability. *Education Policy* 18(1), 45–70.
- Making Good Progress (2007). Department for children, schools and families, London. <http://www.teachernet.gov.uk/mgppilot/> (accessed August 2009).
- Mansell, W. (2007). *Education by Numbers: The Tyranny of Testing*. London: Politico's.
- No Child Left Behind Act 2001 (2002). US Congress Public Law 107–110. <http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf> (accessed August 2009).
- Pitoniak, M. J. and Royer, J. M. (2001). Testing accommodations for examinees with disabilities: A review of psychometric, legal, and social policy issues. *Review of Educational Research* 71(1), 53–104.
- Rogosa, D. (2003). How accurate are the STAR national percentile rank scores for individual students? – an interpretive guide: Version 2.0, CAT/6 Survey. <http://www-stat.stanford.edu/~rag/acccuide/> (accessed August 2009).
- The White House (2007). The No Child Left Behind Act: Five years of results for America's children, fact sheet, 8 January 2007. Washington, DC: White House. <http://www.whitehouse.gov/news/releases/2007/01/20070108-6.html> (accessed August 2009).

Relevant Websites

- <http://www.pisa.oecd.org> – Programme for International Student Assessment Paris: OECD.
- <http://timss.bc.edu> – Trends in International Mathematics and Science Study.

Summative Assessment by Teachers

R Daugherty, Cardiff University, Cardiff, Wales, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Teachers routinely sum up what their students have learned. They cannot expect to be able to move students on in their learning, either individually or as a group, if they have not established where those students have reached. The skill of summarizing is therefore integral to instruction and to informing instructional practice.

Teachers sum up in a multiplicity of ways, from an informal comment on a particular piece of work to a formal grade generalizing about overall performance. What marks out such summing up as summative assessment is not the extent of the formality or of the generalization but the purpose or purposes for which it is undertaken. Where the teacher is making a judgment that is part of a dialog between teacher and student to help the student learn more effectively that summing up is part of the process of formative assessment. If, however, the teacher's judgments are directed toward a record of student performance that is to be communicated to others, within or beyond the school, we are into summative assessment – evaluating learning in ways that are determined by the needs of others for evidence of student performance.

Assessments designed to meet summative purposes may sometimes have formative spin-offs; teachers and students can adapt subsequent learning to take account of what such assessments have revealed. But the form those assessments take and the approach of both teacher and student to them will reflect the needs of the user(s) for the performance data that emerge from them.

There are numerous possible reasons why information about the performance of students might be needed by someone other than the teacher and the student. These purposes will be referred to here in three broad categories:

- reporting on the individual student to other teachers in the same school, to another school, or to a parent/guardian;
- reporting on individual performance as part of a process leading to the student being awarded an external qualification; and
- reporting on the performance of groups of students to supply aggregate data which managers of the school system will use to monitor the performance of teachers and schools.

Each category places particular demands on the way the evidence of performance is generated, evaluated, and reported. The school and its teachers have to be able to

respond to needs as diverse as a parent wanting to know how his/her child progressing, an agency managing a national qualification that is awarded to many thousands of students, and a school district monitoring the performance of the schools for which it is responsible.

One definition of summative assessment by teachers, precluding the possibility of teachers assessing the performance of students other than those they have taught, is:

The process by which teachers gather evidence in a planned and systematic way in order to draw inferences about their students' learning, based on their professional judgment, and to report at a particular time on their students' achievements (Assessment Reform Group, 2006: 4).

In the article, discussion of the teacher's role in summative assessment extends beyond systems of assessment in which teachers are making judgments about their own students. The above definition is, however, helpful in highlighting four crucial features of summative assessment by teachers:

- it is systematic, occurring within a system;
- it requires teachers to draw inferences from the evidence they have of student learning;
- it calls for teachers to exercise their judgment; and
- it involves some form of report on student achievement.

It should also be noted that the systems, inferences, judgments, and forms of report can be expected to be purpose specific; a teacher will be responding concurrently to several different sets of requirements for data about student performance. What is needed by the student's parent will make demands on the teacher's record-keeping, judgment, and reporting that are different from, for example, the requirements of the school's monitoring procedures or of the agency managing a national qualification. For many teachers it is the within-school procedures that make the most frequent demands on them to sum up student performance. However, it will be on the other two categories – the award of external qualifications and the supplying of data for system monitoring – that this article focuses.

Why a Role for Teachers?

The case for teacher judgment contributing to the way student performance is evaluated, whatever the use that is to be made of the evidence thus generated, has been well

rehearsed. In principle, it rests on the argument that teachers can sample the range of a student's work more fully than can any assessment instruments devised by an agency external to the school, thus enhancing both the reliability (more evidence) and the validity (a wider range of evidence) of the assessment. Various interpretations in different cultural contexts, this argument has fueled the case for performance assessments in the USA (Resnick and Resnick, 1992), for assessment by teachers in high school qualifications in the UK (Wilmot, 2004), and for reliance on teacher judgment in school-based assessment at the end of secondary school in Queensland, Australia.

Broadfoot (1995) interpreted the increased use of performance assessments as part of a broader international trend with five main elements:

- greater reliance on formative, learning-integrated assessment;
- a commitment to improving teacher expertise in assessment;
- increasing emphasis on validity in relation to the full range of curriculum objectives;
- increasing emphasis on describing learning outcomes in terms of specified standards; and
- increasing emphasis on using assessed learning outcomes as an indicator of the quality of educational provision.

It is worth noting here that the advocacy of teachers having a role in summative assessments is based on the teacher being able to draw on a wider range of evidence available (rather than a small, time-limited sample) and being better placed to draw inferences from the evidence (through familiarity with what students' work reveals about their learning). It is not in essence a matter of the assessment instruments that are deployed. For example, both the external agency and the teacher may well draw inferences from test-based assessment instruments though, for practical reasons, the external agency will necessarily be more dependent on such instruments. To discuss summative assessment by teachers in contradistinction to external tests/examinations is to conflate the who of judgment with the how of assessment procedures.

If the arguments in favor of the teacher's role in summative assessments are so persuasive, then that way of assembling evidence and drawing inferences might be expected to have a prominent place in systems of assessment across the globe. Some of the reasons why this has not occurred will become apparent from a review of practice in a range of contexts.

Summative Assessment by Teachers in Practice

There is enough experience from different countries of summative assessment of student work by teachers within

large-scale systems of assessment for some conclusions to be drawn about its use in practice. Three case studies are briefly summarized here as a prelude to the discussion being broadened and to identifying recurring themes.

Portfolio Assessment in a State-Wide Assessment Program: Vermont, USA

Disillusionment with models of state-wide testing that relied on multiple-choice test instruments led a minority of US states to experiment in the 1980s and 1990s with various forms of performance assessment. One such form is the assessment of a portfolio of student work, sampling a student's performance more fully than either multiple-choice or constructed response tests presented under standard conditions. In this form of assessment, the work to be judged is accumulated during a course of study, allowing for greater diversity both in the individual student's response to the tasks set and in the conditions under which it is completed. Maintaining a degree of standardness that should, in principle, make consistent judgment possible across a large student cohort is achieved partly through a task specification that is common to all and partly through rules for scoring the portfolios.

The original plans for state-wide assessment in the state of Vermont gave portfolio assessment in the two assessed subjects (mathematics and writing in grades 4 and 8) equal status with the test-based assessments of those subjects. Students and teachers selected what they judged to be the best pieces of work for each student while also taking account of requirements for specified types of work. For state-level reporting, after early experiments with teachers scoring their own students writing portfolios, the portfolios were scored by teachers other than the students' own.

There were marked disparities in the scoring of portfolios in the early years of the Vermont program. Subsequent refinements in the rubrics and in the training of teachers achieved relatively high correlation between scorer ratings in mathematics, with improved but lower levels of consistency in the scoring of writing. Rater inconsistency was not, however, the largest source of unreliability in the scoring of portfolios. From the evaluation of the early implementation of the Vermont program, it was apparent that the sampling of tasks contributed substantially to the unreliability of scoring of the mathematics portfolios; different selections of tasks would have resulted in substantially different ratings. Another variable in the compiling of portfolios was the extent to which teachers provided assistance to students, including whether students were encouraged to revise the pieces of work that would be selected for inclusion.

The conclusions drawn by Koretz (1998) from his review of the evidence from four large-scale programs using portfolio assessment in the USA are not encouraging:

Portfolio assessment has attributes that make it particularly appealing to those who wish to use assessment to encourage richer instruction – for example, the ‘authentic’ nature of some tasks, the reliance on large tasks, the lack of standardization, and the close integration of assessment with instruction. But some of these attributes may undermine the ability of the assessments to provide performance data of comparable meaning across large numbers of schools (Koretz, 1998: 332).

Coursework Assessment in a National Qualification: The General Certificate of Secondary Education, England

The General Certificate of Secondary Education (GCSE) was established in the mid-1980s as a subject-based national qualification for students at age 16 in England, Wales and Northern Ireland. Its main rationale was that the high-status qualification that it replaced, O level, had relied too much on time-limited written examinations which prioritized a narrow range of academic attainments. The new qualification, it was claimed, would recognize a wider range of attainments; it would assess not only student performance in written examinations but also what students know, understand and can do. In short, it would be more valid.

Coursework assessment was to be the means of achieving this goal, building on the experience of in-course assessment that had been gained through the lower-status qualifications that were, along with the O level, replaced by GCSE subject certification. The arguments for teachers contributing to summative assessments of their students were accepted and signed up to by all the main interest groups, including the Conservative government responsible for implementing the new qualification. However, it remained unclear what was actually meant by coursework assessment and how an acceptable level of consistency would be achieved in the summative assessments for which teachers would be responsible.

In the event, each of the several agencies responsible for awarding the new qualification put in place its own arrangements, lightly regulated by the national organization overseeing the examination. The activities encompassed under the umbrella term of coursework were extraordinarily diverse, from portfolios in art and English to laboratory tasks in science but also including a proliferation of activities in minority subjects that were also scored by the students’ own teachers. In the overall GCSE grade, the notional weighting accorded to coursework ranged from as low as 10% to 100% in one high-profile scheme in English. The only system-level attempt to enhance consistency of teachers’ practice and of teacher judgment was a national (but not compulsory) scheme of

training for one teacher from each school in the ten most popular subjects, supported by publication of guidance materials.

Within a very short time, media reports of students plagiarizing the work of other students, parents completing their children’s work for them, teacher bias, and sundry other indefensible practices undermined public confidence in the new qualification. The GCSE was always going to find it difficult to match the respect that had been accorded to the O level, an examination tailored to what were assumed to be the academic needs of the minority of students who attended selective secondary grammar schools. However, those responsible for its implementation had failed to think through, plan for, and regulate effectively a qualification that would depend substantially on tens of thousands of teachers in over 100 subjects managing and scoring work undertaken by their own students.

An initial phase of reduction in the reliance on coursework for the award of GCSE grades, coupled with tighter regulation at national level, quickly followed in the early 1990s. A second phase, abandoning the term coursework and further limiting the role of teachers in contributing their judgment to the award of the qualification, was implemented from 2009. The new controlled assessments would be designed to ensure an acceptable level of consistency in the setting, taking and scoring of in-course tasks.

Continuing concerns about the reliability of teachers’ assessments of individual students were not the only reason for coursework assessment being abandoned. The GCSE had been introduced as a qualification for individuals to aim at. Twenty years on, with governments in England as elsewhere in the world using data from a range of sources to monitor system performance, GCSE grades aggregated across cohorts of students had become the main indicators of the performance of high schools in England. High school teachers were not only being expected to contribute to the grading of their students, they were also participating in supplying the data on which their own, and their school’s, performance would be judged.

National Monitoring of Educational Achievement: New Zealand

Some national education systems rely on a sample of performance data collected for that purpose for the monitoring of system performance. Other systems analyze and interpret data obtained from every student in the selected cohorts. For example, England abandoned its sampling system, the assessment of performance unit (APU), in the late 1980s in favor of using data on every student in the cohort gathered either from national-curriculum assessments at ages 7, 11, and 14 or from the grades achieved by students in national qualifications at ages 16 and 18.

A national system of sampling student performance in which teachers have a central role is New Zealand's National Education Monitoring Project (NEMP). Each year, about a quarter of the areas within the New Zealand Curriculum Framework for children in years 4 and 8 are targeted by a range of tasks designed to elicit students' attainments through approaches that interest and stimulate them. These tasks are administered to a sample of about 2.5% of the two cohorts by teacher administrators seconded to work in pairs assessing children in several schools over a 5-week period.

The students' responses are forwarded to be marked centrally. Teachers are also recruited at that stage in the NEMP to mark the tasks that require higher levels of professional judgment. Evidence of performance is not confined to the students' written responses to tasks. By 2006, two-thirds of the evidence of student attainment was in the form of videotapes of the interaction between student and teacher administrator, allowing for a rich analysis of both process and task achievement.

Several features of the role of teachers in the NEMP are worth highlighting in the context of this review of summative assessment by teachers:

- Only a selected group of teachers is engaged either in administering or in marking students' work. By giving priority in each year's selection as teacher administrators to teachers who have not previously been engaged as teacher about 5% of primary teachers had undertaken that role over the first 12 years of the NEMP.
- There is at least as much focus on how the tasks are administered as on how they are marked, ensuring close scrutiny of how the evidence of attainment has been obtained.
- The teachers who manage the task administration and those who mark the students' responses are trained; teacher administrators undergo a week of specialist training.
- Substantial gains in professional learning are reported (Gillmore, 2002) for those teachers.
- Teachers are engaged in managing and judging the performance of students on curriculum-related tasks but, crucially, not the work of the students they teach or of other students in the schools in which they are employed.

Overview

Reference to only three of the many examples that could have been quoted runs the risk of distorting what emerges about the role of teachers in large-scale systems of assessment. For example, more than 30 years experience in the Australian state of Queensland of school-based assessment at the end of secondary school paints a more positive picture of teacher judgment having a central role in the award of a qualification. On the other hand, 20 years

experience of a national-curriculum assessment system in England has shown that, when the stakes are high and no more than token recognition is given to the judgment of teachers, the role of teachers as assessors in national systems can become peripheral (Whetton, 2009).

A systematic review of the evidence of reliability and validity in assessment by teachers for summative purposes (Harlen, 2004) brought together evidence from a range of academic studies of the practice of summative assessment. Several of the findings from that review point to ways of enhancing the contribution of teachers to the judgments made about large cohorts of students:

- The low reliability of portfolio assessment where tasks are not closely specified, as was the case in Vermont, argues for closer task specification.
- Clear specification of the criteria by which performance is judged can also improve reliability.
- The clearer teachers are about the goals of students' work, the more consistently they apply the criteria used to judge performance.
- Where teachers are required to assess their own students, steps need to be put in place, through training, to minimize bias in teacher judgment and, through monitoring outcomes, to detect malpractice.

As is inherent in any review of relevant research, the systems of summative assessment by teachers included in Harlen's review varied greatly in many respects: teacher expertise; the training available; whether or not teachers are judging their own students; and the way teacher judgments are moderated. The contexts in which teachers were making judgments of student performance were also diverse: different purposes (including qualifications for individual students, system-wide monitoring and holding teachers and schools to account); different types of demand for performance data; and the degree of trust accorded to teachers' professionalism. As Madaus and Kellaghan (1993: 459) noted in their observations on experiments with authentic testing in England, the arguments for teachers to have a role in large-scale systems of summative assessment must be evaluated in relation to two realities:

First, there are many practical, technical and infrastructural issues that must be resolved before such techniques can safely be deployed as policy instruments on a large scale in schools. Second, claims about the positive effects of such techniques have to be examined in light of how the results will be used.

Summative Assessment by Teachers: A Process

The teacher drawing inferences from evidence presented by a student is only one stage in a process that has several

key components. To understand summative assessment by teachers, we need to understand how each stage plays its part in the overall process.

Task Specification

A task must have been completed if the student is to present evidence on the basis of which a performance judgment can be made. This is equally true when the student is responding to a question in a test but that test question is notionally the same for every student in the cohort. However, to define closely the nature of the task(s) to be undertaken by each student in a system of summative assessment by teachers would be to compromise its potential to be more valid than a system that is wholly dependent on evidence from the student's performance in a test.

The tasks should, in principle, reflect the educational outcomes of the course of study. If speaking a language is an integral part of learning it, then an assessable task will call for oral proficiency to be demonstrated. If practical skills in the laboratory are valued outcomes from a course of study in science, then laboratory skills should be assessed. Systems of summative assessment by teachers open up a wider range of task possibilities but do not preclude the use of test-based instruments.

Amidst that diversity there needs to be some commonality of tasks across the whole student cohort. For the designers of systems of summative assessment, there is a delicate balance to be achieved along a continuum from precise task specification to loose guidelines. It is possible within a framework of specified tasks, as the New Zealand monitoring project illustrates, to offer students considerable scope to show what they are capable of.

Task Conditions

For a student to produce her/his best possible response to a given task, there will need to be some flexibility allowed in the conditions under which the task is completed – time to plan, to revise drafts, or to access relevant information. Two main threats to the validity and reliability of the outcomes can arise from that flexibility. One is that the conditions under which the students across the cohort have completed the task are too variable for their performances to be judged fairly. The other is that the openness of conditions under which the work was completed will have resulted in much or all of it not being the work of the student in whose name it is submitted.

Once again there is a delicate balance to be achieved by system designers. Students may be required to complete tasks under supervision and with minimal support but where would be the potential in those conditions for the student to produce a best performance across a range of curriculum goals? On the other hand, as with the nature

of the tasks themselves, if the conditions under which those tasks are completed are too variable, the outcomes will be, and will be seen to be, unreliable.

Which Teachers' Judgments?

Discussion of the merits of summative assessment by teachers has sometimes been clouded by a lack of clarity as to which teachers are called upon to judge student performance. There are three possible answers to that question:

- a teacher who has also been responsible for setting and managing the work that is to be judged (the GCSE example above);
- a teacher working closely with the teacher who has set and managed the work, possibly in the same school (the Vermont portfolios example above); and
- a teacher whose workplace is equivalent to, but entirely separate from, the school in which the work they are judging was undertaken (the New Zealand NEMP example above).

For teachers in the first of these categories, their knowledge of the students concerned could enable them to make more appropriate inferences from the evidence than would be possible for a teacher who did not know those students. On the other hand, there is also evidence from research (Harlen, 2004) of bias in judgments by a student's own teachers, much of it unintentional. In the third category, the rationale for teachers judging student performance is that they are drawing upon the professional expertise they bring from their own classroom experiences.

Criteria and Performance Standards

Any judgment of student performance will involve the application of criteria to match the available evidence against expected standards of performance. The criteria identify the critical features of the student's work, the standards identify the level(s) of performance against which that evidence is to be judged. As Sadler (2009: 160) explains it:

Divergent works are typically complex, in the sense that their quality can be explained only by reference to multiple criteria, possibly including some that are abstract in nature . . . A *criterion* refers to a property, quality, characteristic or attribute of a student response. It is distinct from a *standard*, which refers to a particular degree or level of quality.

In a straightforward marking rubric for a test in which the person scoring the test is not expected to exercise her/his own judgment, the criteria may remain implicit and the performance standards may be expressed only in terms of an overall cut-off score.

But for a more sophisticated judgment of less straightforward evidence, such as that in a set of portfolios, it is essential to make the criteria as explicit as possible and to show how the performance standards can be operationalized. Experience from systems that have relied, at least in part, on teacher judgment has shown that the involvement of teachers in the generation of criteria and in shared interpretations of standards can, over a period of time, enhance both the validity and reliability of outcomes.

What Inferences

Once the tasks specifications, task conditions, criteria, and performance standards are in place, it may sound deceptively easy to say that teachers will then be able to draw inferences from the available evidence of student performance. But what kind of inferences are they expected to draw and to what purpose? It is one thing for a classroom teacher to grade the performance of each student for internal monitoring within a school but quite another for a teacher to draw inferences that will be used to determine whether a student is awarded a national qualification.

The translation from evidence via inference to the reporting of performance can take many different forms. It can result in a numerical score, a grade, or a profile of the attributes that the work has displayed. Even within the same system, the form of the inferences teachers are expected to draw from the evidence can be radically changed. The teacher assessment element in national-curriculum assessment in England moved on from compartmentalized judgment in terms of statements of attainment and attainment targets to an expectation that teachers would make holistic best-fit judgments, matching any available evidence of student performance to multifaceted levels of attainment. There is evidence from a number of contexts of significant variability in the way teachers draw inferences and make judgments, suggesting a need for research into the deep structures of teacher judgments and the underlying conceptualizations.

A Systematic Approach

The definition quoted at the beginning of the article drew attention to the fact that, as with any large-scale arrangements for assessing the performance of students, a system must be in place within which those stages in the process of assessment can be operationalized. The structures and procedures within that system will need to be fit for purpose. Experience of summative assessment systems in which teachers have an active role suggest that there are several aspects of system design that are critical to the system being fit for purpose.

1. *Explicit procedures for each stage in the process.* It might seem obvious from the previous section of this article that the answers to the system design questions referred to there would be decided and communicated to the teachers involved in advance of implementation of a new assessment system. That there are many instances of this not being done suggests that the subtleties and complexities of enabling teachers to contribute their judgments to large-scale systems of assessment have often not been recognized. For example, initial planning at the national level for the teacher-assessment element of national-curriculum assessment in England consisted of little more than pamphlets that were long on generalization and short on practical suggestions.
2. *Teacher expertise.* Systems of assessment where the teacher has a role in setting and managing the tasks, as well as in drawing inferences from, and making judgments about, students' responses, place considerable demands on teachers' expertise. The teachers involved will bring to that work what they have learned from their training and their experience; to varying degrees, each teacher will have a generic guild knowledge of assessment. If the system depends on the assessment expertise of every teacher across a whole student cohort (about 600 000 students for national-curriculum assessment in England), that is a lot of variability. The problem is reduced, as in the case of the NEMP in New Zealand, where fewer teachers are involved and where they are selected on the basis of the level of expertise they possess.
3. *Teacher training and support.* Whatever expertise teachers bring to their role(s) there will be a need for system-specific training and support. That training may take the form of induction into the system as a whole and ongoing support as the system evolves or it may be targeted at perceived problems such as inconsistency or bias in teacher judgments. For a system of summative assessment by teachers to operate efficiently, all of those forms of training – initial and ongoing, generic and targeted – need to be in place. It is also the case, as was noted earlier, that teachers are likely to fulfill the roles allotted to them more effectively if they are active participants in, rather than passive recipients of, training and support.
4. *Quality assurance and quality control.* The quality of the outcomes from any system of assessment can in principle be enhanced by quality-assurance arrangements to establish confidence in the procedures and judgments. It will also need to be reinforced by quality-control arrangements to check that those procedures and judgments are resulting in assessed outcomes that are fit for purpose. Both quality and assurance and quality control depend on moderation procedures designed to produce consistency across assessors in qualitative judgments of student performance.

A range of approaches to moderation, in particular social moderation through interaction among participants in an assessment system, is discussed more fully elsewhere in the encyclopedia. Maxwell helpfully characterizes the distinction between assurance and control as moderation for improvement (through professional development) versus moderation for accountability (quality control). This aspect of a system of summative assessment by teachers is crucial to the system's effectiveness and Maxwell refers to no fewer than 12 factors to consider in designing moderation processes.

Conclusion

These examples of summative assessment by teachers in practice show that the social and cultural context is a crucial factor in ensuring successful implementation of a system of assessment. In the USA, Shepard (2000: 4), after reviewing how conceptualizations of assessment evolved during the twentieth century, concluded that there had been:

... consistent theoretical frameworks in which conceptions of 'scientific measurement' were closely aligned with traditional curricula and beliefs about learning.

In the decade following Koretz's (1998) review of portfolio assessments, the role of scientific measurement has been reinforced by the US federal government's No Child Left Behind program.

In contrast, the government of New Zealand has since 1994 entrusted the monitoring of student performance in its schools to teachers seconded into their roles as administrators and markers. There are other examples too of countries, for example, Scotland (Hutchinson and Hayward, 2005) and Wales (Daugherty, 2009), where teachers are trusted to judge the performance of students, including their own students.

Within any one context, there are also technical dimensions to systems of assessment by teachers for summative purposes. Some of the examples referred to here have had serious, sometimes fatal, design deficiencies. The five stages in the process and the four aspects of a system of summative assessment by teachers that were discussed in the two preceding sections of this article offer a checklist for system designers. In their discussion of threats to the valid use of assessments, Crooks *et al.* (1996) proposed a model of a chain with eight links, from administration of the task through to impact on the student and others. Validation in any system of assessment, not only one in which teachers are responsible for some of the links, requires careful consideration of the strength of each link (Crooks *et al.*, 1996: 269).

Commentators on the conceptualization and practice of student assessment (e.g., Resnick and Resnick, 1992; Broadfoot, 1995; Shepard, 2000; Harlen, 2007) have

typically interpreted recent international trends as being in the direction of greater reliance on the judgment of teachers. That interpretation may underestimate the contradictory pressures that are associated with demands for measures of system accountability, both within and across national-education systems. Even in Australia, a country where all six state-education systems have developed standards-referenced, teacher-based procedures, there are tensions between those systems and federal government requirements for test data to enable comparisons of performance against benchmark standards (Cumming and Maxwell, 2004). In Japan, a perceived crisis in school standards, fueled by the country's ranking in international comparative studies of achievement (Tsuneyoshi, 2004) and the inadequacy of the data available on student performance at the level of schools and districts, led to the reintroduction of national tests in 2007. The conflict that is inherent when teacher judgment has a role in summarizing student performance and the performance data then used for system accountability purposes was highlighted in a review of large-scale assessment systems in seven countries:

... while many systems rely on teacher judgment for assessments that are high stakes for students, there are ... no systems that rely on teacher judgment for assessments that are high stakes for teachers (Black and Wiliam, 2007: 11).

Summative assessment by teachers can and does have a role in large-scale systems of assessment. In a social and cultural context where teachers are trusted to exercise their judgment of student performance, there are two essential features if such systems are to be designed and implemented effectively. First, the role that teachers have in a system must be compatible with the use(s) that will be made of the student-performance data produced by that system. Second, a sophisticated infrastructure is required if large-scale systems are to realize the potential benefits, in terms of both validity and reliability, of extending the involvement of teachers beyond the use of their judgment for internal reporting purposes.

See *also*: Classical Test Theory Reliability; Classroom Assessment in Policy Context (Australia); Formative Assessment; Moderation of Student Work by Teachers; Portfolio Assessment; Test Development; The Multiple Purposes of Assessment; Validity.

Bibliography

- Assessment Reform Group (2006). *The Role of Teachers in the Assessment of Learning*. London: University Institute of Education.
- Black, P. and Wiliam, D. (2007). Large-scale assessment systems: Design principles drawn from international comparisons. *Measurement* 5(1), 1–53.
- Broadfoot, P. (1995). Performance assessment in perspective: International trends and current English experience. In Torrance, H.

- (ed.) *Evaluating Authentic Assessment*, pp 9–43. Buckingham: Open University Press.
- Crooks, T., Kane, M., and Cohen, A. (1996). Threats to the valid use of assessments. *Assessment in Education* **3**(3), 265–285.
- Cumming, J. and Maxwell, G. (2004). Assessment in Australian schools: Current practice and trends. *Assessment in Education* **11**(1), 89–108.
- Daugherty, R. (2009). National curriculum assessment in Wales: Adaptations and divergence. *Educational Research* **51**(2), 247–250.
- Gilmore, A. (2002). Large-scale assessment and teachers' assessment capacity: Learning opportunities for teachers in the national education monitoring project in New Zealand. *Assessment in Education* **9**(3), 343–361.
- Harlen, W. (2004). A systematic review of the reliability and validity of assessment by teachers used for summative purposes. In *Research Evidence in Education, Issue 1*. London: Institute of Education EPPI Centre.
- Harlen, W. (2007). *Assessment of Learning*. London: Sage.
- Hutchinson, C. and Hayward, L. (2005). The journey so far: Assessment for learning in Scotland. *Curriculum Journal* **16**(2), 225–248.
- Koretz, D. (1998). Large-scale portfolio assessments in the US: Evidence pertaining to the quality of measurement. *Assessment in Education* **5**(3), 309–334.
- Madaus, G. and Kellaghan, T. (1993). The British experience of 'authentic' testing. *Phi Delta Kappan* **74**, 458–469.
- Resnick, L. and Resnick, D. (1992). Assessing the thinking curriculum: New tools for educational reform. In Gifford, B. R. and O'Connor, M. C. (eds.) *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction*, pp 39–75. Boston, MA: Kluwer.
- Sadler, R. (2009). Indeterminacy in the use of pre-set criteria for assessment and grading. *Assessment and Evaluation in Higher Education* **34**(2), 159–179.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher* **29**(7), 4–14.
- Tsuneyoshi, R. (2004). The new Japanese educational reforms and the achievement 'crisis' debate. *Educational Policy* **18**(2), 364–394.
- Whetton, C. (2009). A brief history of a testing time: National curriculum assessment in England 1989–2008. *Educational Research* **51**(2), 137–159.
- Wilmut, J. (2004). *Experience of Summative Teacher Assessment in the UK: A Review*. London: Qualifications and Curriculum Authority.

The Multiple Purposes of Assessment

P E Newton, Cambridge Assessment, Cambridge, UK

© 2010 Elsevier Ltd. All rights reserved.

Anyone planning to invest in a new vehicle ought first to have a good idea of what they are going to use it for. If the foremost desire is to cruise in style along winding roads, then a Harley Davidson motorcycle might fit the bill. If, instead, the major concern is to transport colleagues, tools, and the odd animal across country fields, then a four-wheel-drive utility vehicle might be much more appropriate. In just the same way, those who design vehicles need to start by considering transportation purposes, as do those who review them.

Designers, users, and evaluators of educational assessments also need to start by considering purposes. Just as it would be daft to try to use a bicycle to cross a desert, it would be daft to try to use certain forms of assessment for certain kinds of purpose. Strangely, though, when it comes to educational assessment, people often fail to appreciate this point.

What Is an Assessment Purpose?

Admittedly, it can be confusing to discuss the multiple purposes of educational assessment, since there are multiple levels at which these can be envisaged. At one level, the purpose of assessment might be to generate a particular kind of judgment, result, or outcome (e.g., a pass or a fail, a grade linked to a performance descriptor, or a percentile rank). At another level, the purpose might be to bring about a particular kind of social or educational impact or consequence (e.g., to motivate learners to work hard, or to cause teachers to align their teaching with a specific curriculum). At yet another level, the purpose might be to support a particular kind of decision, process, or action (e.g., to select the best candidate for a job, to guide subsequent teaching and learning, or to monitor educational standards over time).

All of these three levels – result, impact, and decision – need to be borne in mind by users, designers, and evaluators alike. However, in keeping with the emphasis in the literature, and with its centrality to theory and practice, the remainder of this article focuses upon assessment purposes at the decision level: the decisions, processes, or actions for which assessment results are used.

What Does It Mean to Be Fit-For-Purpose?

Fit-for-purpose is a phrase that is often used when discussing whether or not something can be expected to

function effectively. In the world of educational assessment, fitness-for-purpose is closely related to the concept of validity. This validity is broadly interpreted to embrace any form of evidence which bears upon the adequacy and appropriateness of inferences from results in relation to a specified use (e.g., Messick, 1989; Kane, 2006).

Inferences from exactly the same set of assessment results might be fit (i.e., valid) for certain purposes, but unfit (i.e., invalid) for others. This might seem counterintuitive, but it is fairly straightforward to explain. It depends upon factors such as the following:

- the kind of inference drawn from results (e.g., results from a single test might be interpreted either in terms of attainment in relation to a specific curriculum, or in terms of potential to succeed in a related course at a higher level, and the adequacy and appropriateness of inferences drawn will differ accordingly);
- the form of the inference drawn from results (e.g., results in the form of an overall grade might be appropriate for placing students in suitable educational programs, but inappropriate for identifying individual teaching and learning needs); and
- the accuracy of the inference drawn from results (e.g., results from a single test might be quite accurate in identifying those who have failed to achieve basic proficiency, but quite inaccurate in distinguishing excellent students from very good ones).

In short, decisions made when designing an assessment instrument, process, or system will secure the validity of inferences in relation to certain purposes, but will threaten the validity of inferences in relation to others (see AERA/APA/NCME, 1999, Standard 13.2 comment).

How Many Assessment Purposes Are There?

There are very many different kinds of decisions, processes, or actions that result from educational assessments might support. They might, for instance, be used:

1. to decide whether students are making sufficient progress in attainment in relation to expectations or targets; and, potentially, to allocate rewards or sanctions (student monitoring);
2. to identify students' proximal learning needs, guiding subsequent teaching (formative);

3. to judge the social or personal value of students' achievements (social evaluation);
4. to clarify the type and extent of students' learning difficulties in light of well-established criteria (diagnosis);
5. to determine whether students meet eligibility criteria for special educational provision (provision eligibility);
6. to identify students who differ significantly from their peers, for further assessment (screening);
7. to segregate students into homogeneous groups, on the basis of aptitudes or attainments, to make the instructional process more straightforward (segregation);
8. to identify the most suitable courses, or vocations for students to pursue, given their aptitudes (guidance);
9. to identify the general educational needs of students who transfer to new schools (transfer);
10. to locate students with respect to their position in a specified learning sequence, to identify the level of course which most closely reflects it (placement);
11. to decide whether students are sufficiently qualified for a job, course, or role in life – that is, whether they are equipped to succeed in it – and whether to enrol them or to appoint them to it (qualification);
12. to predict which students – all of whom might, in principle, be sufficiently qualified – will be the most successful in a job, course, or role in life, and to select between them (selection);
13. to provide legal evidence – the licence – of minimum competence to practice a specialist activity, to warrant stakeholder trust in the practitioner (licensing);
14. to provide evidence – the certificate – of higher competence to practice a specialist activity, or subset thereof, to warrant stakeholder trust in the practitioner (certification);
15. to identify the most desirable school for a child to attend (school choice);
16. to decide whether institutional performance – relating to individual teachers, classes or schools – is rising or falling in relation to expectations or targets; and, potentially, to allocate rewards or sanctions (institution monitoring);
17. to identify institutional needs and, consequently, to allocate resources (resource allocation);
18. to identify institutional failure and, consequently, to justify intervention (organizational intervention);
19. to evaluate the success of educational programs or initiatives, nationally or locally (program evaluation);
20. to decide whether system performance – relating to individual regions or the nation – is rising or falling in relation to expectations or targets; and, potentially, to allocate rewards or sanctions (system monitoring);
21. to guide decisions on comparability of examination standards for later assessments on the basis of cohort performance in earlier ones (comparability); and
22. to quality adjust education output indicators (national accounting).

This is not an exhaustive list, but it helps provide a sense of the number and range of different kinds of decisions, processes, or actions that can be identified. Furthermore, the distinctions drawn above are illustrative rather than definitive, using labels which have been informed by the literature, but not necessarily circumscribed by it. This is for a range of reasons, including the following:

- a tendency for different names to be used for the same purpose; for example, Chauncey and Frederiksen (1951: 109) used articulation instead of the more common placement;
- a tendency for the same name to be used for different purposes; for example, Nitko (1989: 455) described as diagnostic what many would describe as formative;
- a lack of precision in applying the terminology; for example, certification is sometimes used vaguely, as though to imply the process of summarizing end-of-course attainment (e.g., Sadler, 1989), but sometimes far more precisely to refer to a specific professional purpose (e.g., AERA/APA/NCME, 1999); and
- the identification of novel purposes; for example, comparability and national accounting (see Newton, 2007, for further details).

Is It Useful to Classify Assessment Purposes?

It is fairly easy to develop schemes for classifying assessment purposes. For instance, those up to purpose 14 (above) relate primarily to individuals and are based upon individual results, while those from 15 to 22 relate primarily to larger entities (such as schools, examination cohorts, or nations) and are based upon aggregated results. Equally, though, the same 22 purposes could be classified in an entirely different manner, as illustrated by the following scheme:

- personal decisions, made by students and parents (e.g., social evaluation, guidance, school choice);
- educational decisions, made by teachers and psychologists (e.g., formative, diagnosis, screening, placement, transfer, and segregation);
- administrative decisions, made by gatekeepers (e.g., provision eligibility, qualification, and selection);
- managerial decisions, made by those supervising and directing performance (e.g., student monitoring, institution monitoring, system monitoring, resource allocation, and organizational intervention);
- endorsed decisions, made by consumers (e.g., licensing, certification);
- technical decisions, made by technicians and analysts (e.g., comparability, national accounting); and
- evaluative decisions, made by researchers (e.g., program evaluation).

As it happens, most writers who have discussed assessment purposes in any depth have provided their own version of a classification scheme (e.g., Cook, 1951; Findley, 1963; Bloom *et al.*, 1971; Millman and Greene, 1989; Nitko, 1989; Hopkins *et al.*, 1990; Mehrens and Lehmann, 1991; Black, 1998; AERA/APA/NCME, 1999; Pellegrino *et al.*, 2001; Schmeiser and Welch, 2006; Gronlund, 2006). This, in itself, seems to cast some doubt over the utility of the exercise. In an ideal world, a classification scheme might help the reader to understand how certain kinds of purposes are associated with certain forms of assessment. But the reality is simply not that straightforward and the many classification schemes in existence seem to provide little more than relatively arbitrary and contestable groupings which lack obvious utility.

While both Millman and Greene (1989) and Schmeiser and Welch (2006) have noted the lack of a universally accepted scheme, Newton (2007) went a step further by claiming that classification schemes can sometimes do more harm than good. Having noted an increasingly common distinction between formative, summative, and evaluative purposes, he suggested that the distinction between formative and summative is spurious and that the distinction between summative and evaluative is trivial. To explain this first claim, the term ‘formative’ describes a decision-level purpose (and might be compared with others like ‘student monitoring’ and ‘diagnosis’); while the term ‘summative’ describes a judgment-level purpose (and might be contrasted with others like ‘descriptive’). To explain the second claim, the only difference between summative and evaluative, as the distinction is typically drawn, is that the former concerns results for individual students while the latter concerns aggregated results, which seems to be a fairly superficial difference. Most importantly, he suggested that making such broad-brush classifications can mislead policymakers, users, and stakeholders into inappropriate inferences, by implying that there might be deeper functional similarities between similarly classified purposes. For instance, it can mislead them into thinking that similarly classified purposes can be supported by the same assessment instrument.

Instead of seeking simplifying schemes, it may be more helpful to think of each purpose as a category in itself. Emphasizing differences between and even within purposes helps to foreground the necessity of tailoring assessment design to assessment purpose.

Tailoring Assessment Design to Assessment Purpose

Even within the same category of purpose, there are often subpurposes which necessitate very different assessment-design decisions. Program evaluation is a good case in point. As explained many decades ago: “If you wish only

to know how well a curriculum is achieving its objectives, you fit the test to the curriculum; but if you wish to know how well the curriculum is serving the national interest, you measure all the outcomes that might be worth striving for” (Cronbach, 1964: 243).

Similar within-purpose distinctions could be made for many other uses. This explains why the national monitoring system which operates in New Zealand is quite different from that which has been operating for many decades in the USA. In New Zealand, the National Education Monitoring Project was intended to be of particular use to teachers and curriculum designers; and, consequently, it was designed to report at the level of individual tasks and small clusters of tasks. In the USA, although originally reporting at the level of individual tasks, the long-term trend National Assessment of Educational Progress was specifically redesigned to be of particular use to policymakers and stakeholders, subsequently reporting only at the overall subject level. Within-purpose distinctions also explain why Nitko (1989: 455) was able to identify five different categories of diagnostic test and their correspondingly divergent technical characteristics.

In exactly the same way, between-purpose distinctions highlight different assessment-design implications. Thus, Millman and Greene (1989: 338) illustrated how assessments designed for program evaluation might differ from those designed for other purposes, such as student monitoring, by:

- targeting learning objectives beyond those specific to the course;
- employing item formats unfamiliar to those following the course;
- allocating different items to different students, to sample learning objectives thoroughly; and
- scoring by learning objective, rather than overall.

This kind of assessment would not even report at the level of individual students, let alone allow a user to monitor student progress in relation to a specific course of instruction.

To illustrate the point further, an assessment intended to support student monitoring ought to be designed according to criterion-referenced principles, rather than to norm-referenced ones; which would not be necessary if only a selective purpose needed to be served (e.g., Thorndike, 1971). Similarly, when results are to be used for purposes with high stakes – such as qualification, selection, or licensing – all aspects of the assessment system need to be designed with test security in mind; from identity checking at the test center to the disposal of completed test scripts (e.g., Millman and Greene, 1989; Schmeiser and Welch, 2006). Such considerations simply do not apply when results are used only for low-stakes purposes, where cheating does not present a significant threat to validity.

Can a Single Assessment Support Multiple Purposes?

Although it might seem possible to use outcomes from a particular assessment instrument, process, or system for a range of disparate assessment purposes, not all of these uses may prove to be defensible. The problem – a classic problem of assessment – lies in determining the appropriate cut-off point between defensible and indefensible.

To illustrate the point, it might be considered indefensible to use an assessment designed principally for system monitoring, and based upon a matrix-sampling approach, for student monitoring. This is because system monitoring tends to emphasize overall content validity at the expense of reliability for individual students. To design an assessment system which thoroughly sampled both curriculum-learning objectives and assessment formats, it would be necessary to allocate different task blocks to different students and to link the results statistically. In theory, if the task blocks were long enough, reliability for individual students would not need to be sacrificed and it might be defensible to report results at multiple levels. In practice, though, if the task blocks were constructed to be long enough, the complexity, expense, and goodwill required to administer the assessment would probably render it unmanageable. So pragmatic compromises tend to be made and results from assessments based on matrix-sampling are generally not reported back to students or to their schools.

It will often be possible to identify a variety of reasons why results from a particular assessment might be considered indefensible for certain uses. One such reason might be that the inferences drawn are of an inappropriate kind. Linn (1989: 7) cited the case of *Larry P. versus Riles*, which concerned the use of intelligence tests to place (predominantly black) children in special classes for the educable mentally retarded. This practice might be considered defensible if it were possible to infer, from performance on such tests, the potential to benefit educationally from segregation. That this inference could not be validated supported the legal decision which prohibited the practice. Although the possibility of drawing valid inferences concerning intelligence was not in question, the possibility of drawing more-distant inferences was.

Perhaps the most common reason for limiting the use of results to specific purposes concerns accuracy – when results which are sufficiently accurate for one purpose are insufficiently accurate for another. This is often attributable to compromises during the assessment-design stage. As noted above, system monitoring assessments typically sacrifice reliability of results for individual students, for reasons of manageability and cost effectiveness. If a system-monitoring assessment were to be designed also to report results for individual students, then it would need, in effect, to be over-engineered for the purpose of system monitoring.

This reasoning is relevant to the debate over whether a single assessment system can effectively support both formative purposes and other purposes too. On the one hand, it does seem theoretically possible to use the same assessment evidence to generate low-level micro-inferences for formative purposes and high-level macro-inferences for other purposes – just as long as the original evidence base is sufficiently comprehensive, structured, and detailed (e.g., Findley, 1963; Harlen, 2006). On the other hand, a single assessment system designed to this specification would probably turn out to be both over-engineered for formative purposes (in terms of comprehensiveness and structure) and over-engineered for other purposes too (in terms of detail). In circumstances like this, it can actually turn out to be more manageable and cost effective to develop separate assessment systems, each tailored to their intended uses of results and each with their own compromises and trade-offs.

Maybe, though, we should look optimistically to the technology of the future for genuinely multifunctional, cost effective, and manageable assessment solutions (e.g., Pellegrino *et al.*, 2001: 284).

Dealing with Multiple Assessment Purposes

The USA *Standards* document (AERA/APA/NCME, 1999) provides some sensible practical guidance on how to deal with multiple assessment purposes, for example:

- A rationale should be presented for each recommended interpretation and use of test scores, together with a comprehensive summary of the evidence and theory bearing on the intended use or interpretation (Standard 1.1).
- The test developer should set forth clearly how test scores are intended to be interpreted and used (Standard 1.2).
- If validity for some common or likely interpretation has not been investigated, or if the interpretation is inconsistent with available evidence, that fact should be made clear, and potential users should be cautioned about making unsupported interpretations (Standard 1.3).
- If a test is used in a way that has not been validated, it is incumbent on the user to justify the new use, collecting new evidence if necessary (Standard 1.4).

Unfortunately, recommendations such as these have implications that are not always palatable to measurement professionals (e.g., Green, 1998). For a start, the responsibilities assigned to test developers, publishers, and users are contestable. Given that any particular test might potentially be used to generate inferences in support of a range of purposes, should test developers and publishers really be expected to provide validity evidence in relation to all of them? If not, then how exactly ought the list to be restricted – to one principal purpose? And, while it may, from a theoretical perspective, be incumbent upon users

to justify new uses, how practical a recommendation is this? Particularly, when the user group extends beyond those professionals with formal responsibility for assessment to stakeholders like parents and many employers.

The situation becomes more complicated still when test developers are required to produce tests for sponsoring agencies, like regional or national governments. Often, policymakers responsible for commissioning the development process will have limited understanding of the challenge of multiple purposes and may need to be disabused of the false belief that a single assessment system can support a multitude of disparate purposes. This may not be easy to do, especially if the policymaker has already earmarked a limited budget for developing the new system.

Hopefully, the responsible test developer will begin by helping policymakers to identify the purposes they consider most important to support, and those which it will be less possible to support. Ideally, the developer would encourage the policymaker to prioritize between desired purposes, and orient the development process to the principal purpose or purposes. However, ensuring that these principal purposes are made explicit for all users and stakeholders to see may prove less straightforward. Moreover, either during the development process or as years go by, ensuring that there is no unannounced drift or accumulation of assessment purposes may prove harder still to prevent. Finally, there is a clear tension for policymakers who are both responsible for publicizing the limitations of their assessment systems in relation to specified assessment purposes and accountable for the success of those systems as mechanisms for organizing and informing a fair society (see Newton, 2005).

The challenge of multiple assessment purposes is not simply technical, but social and political. It requires full and frank debate between developers, policymakers, users, and other stakeholders. All parties need to recognize that a single set of assessment results may not be capable of supporting a range of disparate uses of results. And all parties need to recognize that, when an assessment which has been designed for one specific purpose is used for another, the consequence will be a larger number of incorrect assessments and bad decisions than might otherwise have been the case.

See also: Assessment and the Evaluation of Institutional Effectiveness; Formative Assessment; Summative Assessment by Teachers; Test Development; The Purpose of Educational Evaluation; Validity.

Bibliography

AERA/APA/NCME (American Educational Research Association, American Psychological Association and National Council on Measurement in Education) (1999). *Standards for Educational and*

- Psychological Testing*. Washington, DC: American Educational Research Association.
- Black, P. J. (1998). *Testing: Friend or Foe? The Theory and Practice of Assessment and Testing*. London: Falmer Press.
- Bloom, B. S., Hastings, J. T., and Madaus, G. F. (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.
- Chauncey, H. and Frederiksen, N. (1951). The functions of measurement in educational placement. In Lindquist, E. F. (ed.) *Educational Measurement*, pp 85–116. Washington, DC: American Council on Education.
- Cook, W. W. (1951). The functions of measurement in the facilitation of learning. In Lindquist, E. F. (ed.) *Educational Measurement*, pp 3–46. Washington, DC: American Council on Education.
- Cronbach, L. J. (1964). Evaluation for course improvement. In Heath, R. W. (ed.) *New Curricula*, pp 231–248. New York: Harper and Row.
- Findley, W. G. (1963). Purposes of school testing programs and their efficient development. In Findley, W. G. (ed.) *The Impact and Improvement of School Testing Programs. The Sixty-Second Yearbook of the National Society for the Study of Education. Part II*, pp 1–37. Chicago, IL: National Society for the Study of Education, University of Chicago Press.
- Green, D. R. (1998). Consequential aspects of the validity of achievement tests: A publisher's point of view. *Educational Measurement: Issues and Practice* 17(2), 16–19.
- Gronlund, N. E. (2006). *Assessment of Student Achievement*, 8th edn. Boston, MA: Pearson/Allyn and Bacon.
- Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. In Gardner, J. (ed.) *Assessment and Learning*, pp 103–117. London: Sage.
- Hopkins, K. D., Stanley, J. C., and Hopkins, B. R. (1990). *Educational and Psychological Measurement and Evaluation*, 7th edn. Needham Heights, MA: Allyn and Bacon.
- Kane, M. T. (2006). Validation. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 17–64. Westport, CT: American Council on Education/Praeger Publishers.
- Linn, R. L. (1989). Current perspectives and future directions. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn., pp 1–10. New York: American Council on Education/Macmillan.
- Mehrens, W. A. and Lehmann, I. J. (1991). *Measurement and Evaluation in Education and Psychology*, 4th edn. Fort Worth, TX: Holt, Rinehart and Winston.
- Messick, S. (1989). Validity. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn., pp 13–103. New York: American Council on Education/Macmillan.
- Millman, J. and Greene, J. (1989). The specification and development of tests of achievement and ability. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn., pp 335–366. New York: American Council on Education/Macmillan.
- Newton, P. E. (2005). The public understanding of measurement inaccuracy. *British Educational Research Journal* 31(4), 419–442.
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principle, Policy and Practice* 14(2), 149–170.
- Nitko, A. J. (1989). Designing tests that are integrated with instruction. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn., pp 444–474. New York: American Council on Education/Macmillan.
- Pellegrino, J. W., Chudowsky, N., and Glaser, R. (eds.) (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* 18, 119–144.
- Schmeiser, C. B. and Welch, C. J. (2006). Test development. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 307–353. Westport, CT: American Council on Education/Praeger Publishers.
- Thorndike, R. L. (1971). Educational measurement for the seventies. In Thorndike, R. L. (ed.) *Educational Measurement*, 2nd edn, pp 3–14. Washington, DC: American Council on Education.

The Relationship between Assessment and the Organization and Practice of Teaching

L B Resnick, D William, R Apodaca, and E S Rangel, University of Pittsburgh, Pittsburgh, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Associationist epistemology – The theory of knowledge and learning in which knowledge is defined as a collection of established facts and generalizations.

Embedded assessment – The teacher judgments made in the course of classroom interactions with students.

External assessment – The standardized tests or exams are administered under controlled conditions.

Formative assessment – An embedded, diagnostic assessment.

Interpretive epistemology – The theory of knowing and learning in which learners work to establish meaning, and competing interpretations are considered by a community of learners.

Tracking (also called grouping, setting, or streaming) – The policies in which students of different abilities are placed in different track groups, sets, or streams which receive instruction that proceeds at different rates and may involve different content.

embody an interpretive theory of knowing and learning in which learners work to establish meaning and competing interpretations are considered by a community of learners. These two dimensions of assessment are far from perfectly correlated with one another – interpretive epistemologies can be used in external examinations, for example; and associationist epistemologies can produce engaging classroom discussions. This leads to considerable confusion as educators and psychometricians discuss the role of assessment in adaptive teaching. We examine the two-dimensional assessment space defined by the external-embedded and the associationist-interpretive distinction, showing how goals of reliable measurement and agile response to individual students can sometimes conflict.

Diagnostic Assessment: From Tracking and Sorting to Classroom-Based Adaptation

Early attempts to adapt classroom instruction to individual differences generally led to grouping, tracking, and streaming policies in which students of different abilities were placed in different tracks or streams. These students received instruction that proceeded at different rates and that often involved radically different curriculum offerings. Tests and assessments were intimately involved in this process of instructional sorting – from Alfred Binet's individually administered intelligence quotient (IQ) test to later group-administered aptitude and achievement tests. Tests were routinely used by schools, either as admission screens or as a means of placing students in the appropriate classes or courses within an institution. Group tests found widespread applications beyond classroom walls as well, such as the US Army's use of IQ tests to channel recruits into different specialties.

These macro-strategies of sorting and streaming placed assessment external to the classroom and became the domain of a separate profession of psychometrics. They were based on widely held beliefs among scholars and education professionals that different scores on placement and sorting tests reflected fixed differences in ability, and that sensible adaptation to these differences in ability would probably require not just slower rates of instruction, but different ways of teaching, and perhaps even different content.

Introduction

With the advent of compulsory public education, for most modern societies came efforts to adapt instruction to students. Strategies for adapting have ranged from external assessments of abilities and knowledge in which standardized tests or exams are administered under controlled conditions to deeply embedded teacher judgments made in the course of classroom interactions with students. External assessments provide reliability and replicability, but standardization along with delays in the cycle of testing and feedback can limit their pedagogical utility. Embedded assessments are tuned to the instructional moment, but usually require highly skilled teachers and may be subject to biased judgments. Assessments also differ in their epistemological assumptions. Some are based in an associationist theory of knowledge and learning in which knowledge is defined as a collection of established facts and generalizations. Others attempt to

Starting in the mid-1960s, a number of psychologists began to suggest that intelligence and aptitude were far more malleable than earlier theories had assumed (Glaser and Resnick, 1972). It soon became clear that tracking and streaming were *de facto* denying many students the opportunity to learn more cognitively demanding curricula that they might, with appropriate adaptations and perhaps some extended learning time, be able to learn successfully (Oakes, 1985). Evidence also grew that students of color and language minorities were being disproportionately assigned to special-education categories that isolated them from cognitively demanding instruction.

In response, educators began searching for ways to adapt instruction within heterogeneous classrooms. To guide such adaptive teaching, many scholars and theorists proposed embedding brief evaluative or stock-taking observations within instruction. Not surprisingly, the psychometrics profession sought to respond by developing practices of assessment that would support the ongoing adaptation of instruction. They distinguished between formative (i.e., embedded diagnostic) assessment and summative (i.e., external, evaluative) assessment.

Glaser (1963) proposed that tests should be designed so as to produce information not just on how an individual student ranked within a class but also what, specifically, had been learned. Bloom (1969) advocated the use of formative evaluation:

to provide feedback and correctives at each stage in the teaching-learning process. By formative evaluation we mean evaluation by brief tests used by teachers and students as aids in the learning process. While such tests may be graded and used as part of the judging and classificatory function of evaluation, we see much more effective use of formative evaluation if it is separated from the grading process and used primarily as an aid to teaching. (p. 48)

Toward the end of the 1980s, two substantial articles reviewed the impact of classroom-assessment practices on students and their learning. One (Natriello, 1987) used a model of the assessment cycle, involving purposes, setting of tasks, criteria, and standards, evaluating performance and providing feedback. The author concluded that much of the research reviewed was largely irrelevant because of weak theorization, resulting in the conflation of key distinctions (e.g., the quality and quantity of feedback).

The second review (Crooks, 1988) had a narrower focus – the impact of assessment practices on students. This author suggested that the summative function of assessment predominated, and that little attention was paid, either in research, or in the practice of educators, to the potential of classroom assessments to assist learning.

The view of assessment as a separate process from teaching continues to dominate approaches to the use of assessment in the support of learning. A review of research by Bangert-Drowns *et al.* (1991) focused on the

effects of feedback in what they called test-like events (evaluation questions in programmed learning materials, review tests at the end of a block of teaching, etc). Throughout this strand of research, the assumption is that assessment is a related, but separate process from teaching; and that assessments need to be psychometrically sound in order to allow valid inferences about students to be made; but that we should explore the potential for these assessments to do double-duty by also being instructionally supportive. It should be noted that this characterization extends also to the experiments with authentic or performance assessment in the 1980s and 1990s (see, e.g., Koretz, 1998), and to more recent work on student-involved classroom assessment (Stiggins, 2001). The emphasis has been on trying to make assessments that are primarily designed to support a summative function to function formatively as well. Most recently, the Commission on Instructionally Supportive Assessment in the US explored the possibilities for tests used primarily for purposes of accountability also to provide information useful to teachers in their teaching (Popham, *et al.*, 2005).

Over the last 20 years, there has been a strand of work on formative assessment that seeks to examine the real-time classroom-evaluation processes used by teachers not only as aspects of teaching, but also as particular kinds of assessment (Sadler, 1989; Black and Wiliam, 1998). Rather than seeking to make existing assessments support learning, this work has explored how an assessment perspective on certain aspects of teaching might illuminate, and improve, learning (Wiliam and Thompson, 2007).

Accountability Assessment: From End Results to Data-Based Management

Even as some psychometricians have been turning attention to formative assessment, large-scale summative testing has come to exert substantial pressure on teachers and classrooms. Test-based international comparisons of academic achievement are eagerly reported in many countries, and these motivate lively and often quarrelsome debates about how to increase school achievement results. In some countries, schools are held accountable for achievement gains, with varying positive and negative incentives attached to schools' test results. The end-of-course external assessment results are increasingly becoming requisites for promotion and graduation from one level of schooling to the next, thus attaching important consequences to test performance.

The comparative and accountability tests now in wide use were not designed to guide adaptive instruction although tests used in these evaluations can have a distinct influence on the nature of instruction – despite rhetoric in some countries that claim that the tests are curriculum neutral. Teachers work to prepare students to do well on

school-leaving exams, and even on yearly tests that carry important consequences – as in the current American accountability system. In addition, psychometric criteria of reliability and generalizability, along with cost, often drive out assessment of interpretation, problem solving, and reasoning.

The line between embedded and external assessment is blurred when tests designed as external measurement instruments are used as interim assessments to check on progress through an externally mandated sequence of learning topics. This use is current now in several countries with test-based accountability policies. Evidence suggests that unless the interim tests closely match the intended curriculum, tests over time come to trump curriculum as schools and teachers focus increasing attention on teaching to the test.

Truly diagnostic tests, aimed at adaptive instruction, should examine the units or modules of instruction in the classroom. The information such tests yield is intended for a very local audience – the teacher and even students themselves. Assessment results are meant to be used to guide the very next bit of instruction. Thus, if an error in judgment is made, it can be corrected quickly and with no important long-term impact on student-learning opportunities. This means that psychometric criteria for large-scale external assessments need not drive out attention to the details of content and performance that are essential in diagnostic assessment. To be useful to teachers and students, embedded assessments should be tightly linked to the specifics of curriculum – what is taught. Teachers – rather than administrators or accountability monitors – are the people who need to see the data quickly. Indeed, to an outside observer, embedded diagnostic assessments may not appear to be assessments at all but rather part of the teaching process.

Epistemologies of Learning and the Nature of Assessments

Especially strong tensions arise when assessments intended to assist teachers in adapting instruction to their students do not share the same epistemological assumptions as the curricula teachers are attempting to implement. Most of the evaluation and accountability tests in the US today are based on epistemological assumptions of the early twentieth century. Major developers of educational psychology, such as E. L. Thorndike, who is arguably the father of standardized testing, worked from within the associationist theory of learning. To know something, for Thorndike and his colleagues, was to have a mental collection of many separate associations and the ability to call up these associations quickly and accurately. Thorndike's influential book on *The Psychology of Arithmetic*, published in 1922, listed hundreds of individual

associations that people who knew arithmetic would need to hold in their heads. The best form of teaching, it followed, was well-designed practice of all the many associations that constituted a given body of knowledge.

The appropriate form of testing, given the associationist definition of knowledge, were short items that examined which associations a student had mastered and which needed further study. Associationist testing initially began as collections of short-answer (often single-word) constructed responses, but with the advent of machine scoring, multiple-choice testing took over. With this type of testing as the dominant form of summative evaluation, it was a logical step to try to create formative tests that were essentially subsets of the larger summative evaluation.

From the 1960s onward, however, psychology and cognitive science were giving us a new view of knowledge and skill with profound implications for what to teach and how to assess. These new interpretive theories of learning and knowing range from Piagetian constructivism, to cognitive science's focus on mental models and problem solving, all the way to sociocognitive theories that place learning in the nexus of social interaction as pairs or groups of people work to align their understanding. In all of these theories, knowledge is the result of sense making by learners. To know something is to have constructed a web of understanding that links individual associative bits into structured schemas that are used in reasoning and in acquiring new knowledge. The learner is understood to be an active constructor of knowledge, bringing to bear existing, often schematized, knowledge on new problems. Learning, even of simple material, is thus an interpretive process. Readers, for example, do not just extract a mental copy of what they read; they build a mental model of the situation described by the text (Kintsch, 1986); they question the author's intent (Beck and McKeown, 2001) in order to understand the text's meaning; they construct self-explanations that tell them how elements are linked and why things work the way they do.

With the cognitive revolution in psychologists' understanding of knowledge, which brought a focus on instruction for understanding, meaning making, and the creation of explanations, new forms of teaching have been developed. Students work on tasks and problems that are designed to engage their meaning-making capacities. These often involve both collaborative work with other students, and structured, teacher-led talk as integral parts of the learning process. In classrooms that are engaging in this kind of learning activity, the use of tests matched to the coming end-of-year summative assessment may disrupt teaching rather than support it. Recognizing this, many assessment specialists, along with teachers, have called for types of formative assessment that would better match the new forms of pedagogy they were trying to implement.

Embedded Assessment in Two Epistemologies: Classroom Talk as Diagnostic Testing

Each of the epistemologies we have discussed – the associationist and the interpretive – gives rise to a distinctive set of classroom practices. We might call them signature pedagogies – pedagogies that can be recognized by a small set of classroom moves that signal quite different expectations of how students learn. The differences can be seen to some extent in the tasks students are assigned; but they are most clearly evident in the nature of the conversations about the content embedded in the tasks. These conversations are simultaneously diagnostic and evaluative.

From Catechism to IRE: Associationist Teaching and Assessment

Much of associationist classroom instruction, at least in Western societies, takes its origin in early forms of religious education in the Christian church. Protestants first, and then Catholics, developed a form of basic instruction that proceeded in oral question-and-answer form – known as catechism. Children (and neophyte adults) were asked a set series of questions and were expected to provide standardized answers.

As mass schooling developed, the catechism form moved into the secular classroom.

The content changed (now to include basic arithmetic, a broader range of texts for reading, attention to spelling and correct grammar as the public definition of literacy enlarged to include written composition, and nationalist history), but the form of classroom interaction remained remarkably constant. In reading, history and geography and science, students were assigned a brief text to read – perhaps during class, perhaps overnight as homework – and were then quizzed by the teacher with a series of questions that checked out whether each student had read and remembered what the text said. Arithmetic was similar: a set of calculation problems to solve and a classroom check on whether answers were correct. The classroom check might be in the form of a written quiz, a review of homework, or an oral question and answer: questions by the teacher, answers by the student. Teaching was a matter of setting a task and evaluating how students performed the task. Sometimes a teacher was called upon to explain or to tell students about concepts or events that textbooks had not yet caught up with (with more extended teacher talk as students grew older – right up to the college lecture hall). But evaluation of bite-sized bits of information was the most common form of teaching in schools.

The catechism form is still with us in what we now know as standard recitation teaching. In 1979, Hugh Mehan, aiming to provide a structured account of the

first-grade teaching that he had observed in elementary-school classrooms, described a standardized sequence that came to be known as the I–R–E sequence (Mehan, 1979). The teacher initiated (I) a three-step exchange. The student, who might be a volunteer or someone picked by the teacher, responded (R). The teacher then evaluated (E) the student's response, either accepting it as correct or indicating that a different answer was needed.

Characteristic of IRE, as of catechism, is the assumption that what students are to learn is a prespecified set of facts or explanations. It is understood that the teacher already knows the answers – in fact, it is the preestablished set of questions and answers that defines what is to be learned. In IRE conversations, the students' job is to provide the answers expected by the teacher. The teacher's job is to find a student who can provide the expected answer in order to keep the instructional conversation on track and on time. Assessment is undetachable from the act of teaching, which in turn is composed of a string of rehearsals of the proper questions and proper answers.

In their book on the teaching gap, Stigler and Hiebert (1999) describe standard American teaching of mathematics as very close to this catechism approach. In a prototypical lesson, the teacher provides a set of arithmetic, or algebra, problems to be worked. The students are expected to solve the problems at home or during in-class work periods. During the subsequent class period, a review of homework is conducted (a form of micro-assessment). This review, conducted largely in the IRE form, identifies right and wrong answers to the problems, but there is little discussion of why the answers are right, or of what alternative solutions might have been considered. This recitation form of teaching is standard practice in other subject matters as well. For example, students might be expected to read a chapter in a history book and be prepared to respond from memory to a series of questions posed by the teacher the next day.

Multiple efforts underway now use various forms of technology to expand the power of IRE teaching to diagnose where each individual is in the process of learning the expected set of knowledge components and procedural skills. For example, students can use wireless handheld computers (personal digital assistants or PDAs) to respond to teachers' multiple-choice or short-answer questions and the teacher can receive almost instantaneously a summary of responses. The teacher can then adapt the next piece of a lecture or other presentation to the average knowledge displayed by the class; and can also pick out specific students who may need individual tutoring in order to keep up. Going even further into a technology-based future, today, there are already available sophisticated computer tutors that use a highly interactive combination of problem presentation and solution elicitation to tailor a substantially personalized form of

instruction in which students' steps of reasoning and explanation are immediately evaluated by the computer. Depending on the outcome of the evaluation, the tutor may silently wait for further reasoning, provide corrective feedback, prompt for an alternative strategy, or provide a new problem to address a knowledge growth need (e.g., Koedinger and Aleven, 2007). Interactive computer tutors in use in schools today monitor student reasoning and explanation to provide deeper assessment and learning support beyond simple practice of facts and associations. We are also coming closer to the day when computer-student interactions can be used to provide formative assessment of student-learning strategies. As this happens, computer-based instruction may also be able to support interpretive learning.

Communities of Learners: Interpretive Teaching and Assessment

The classroom conversation takes a very different form in classrooms based on interpretive epistemologies. Over the past two decades, research has accumulated on how discussion methods are used in classrooms and why such discussion may support learning of important school subject matter as well as the process of reasoned participation (Michaels *et al.*, 2007). This research – blending sociology, linguistics, and psychological strands of educational research – has repeatedly demonstrated the role of certain kinds of structured talk in learning with understanding. The discourse focuses on understanding and explanation, and admits multiple student solutions to the classroom discussion.

We can turn once again to the teaching gap book for an interpretive pedagogical image, this one based on the typical Japanese mathematics classroom. Variants of this approach have appeared in many countries in the past 15–20 years, and can now be said to constitute an emergent signature pedagogy for interpretive teaching and learning.

As in associationist teaching, the process begins with the teacher assigning a task for students to work on. The task for an interpretive mathematics lesson typically describes a situation that needs to be mathematized by the students. That is, variables have to be identified and their relationships understood even before specific quantities come into the discussion. Typical tasks for this form of instruction have more than one route to solution, although there is no compromising on the need for finding mathematically correct answers through different routes.

Once the task has been set out, there follows a period of individual and small group work in which students develop candidate solutions to the problem. It is typical at this time for teachers to move around among students, keeping everyone focused, and sometime suggesting an approach to a group that has reached a dead end in their thinking. This is also a time when the teacher can make

mental notes of what the thinking of groups and individuals is. This work period is thus a privileged moment for embedded assessment.

Next, there is a teacher-led whole-class session in which solutions are presented to the whole class by students. This is followed by class discussion of the validity of solutions and of interrelations among suggested solutions. During this time, the teacher uses a number of talk moves in which individuals are asked to elaborate their own ideas or those of other students. Several of the most common moves are shown in **Table 1**.

In this form of carefully guided discussion students' ideas may be challenged – by the teacher or by other students – but a developed set of norms for classroom discussion insure that talk remains accountable to the community, to appropriate forms of reasoning, and to the specific knowledge that is under discussion.

This form of interpretive discussion represents a large change from recitation or catechism – even as it contains a large measure of teacher direction, and hence occasion for embedded assessment of student knowledge. But what counts as knowledge is expanded, the chunks to be learned and talked about are larger; there are more occasions for the teacher to be surprised by students' understanding (and misunderstanding), and it is clear that there is more at stake than finding the single answer that the teacher is looking for. The teacher is not simply searching for a succession of right answers to simple problems or steps within a larger problem. The opportunities for assessment are embedded in the learning activities, and it is difficult to imagine separating them out as individual test items.

Table 1 Characteristic talk moves

<i>Talk-move action</i>	<i>Prototypical form</i>
Revoicing	"So let me see if I've got your thinking right. You're saying XXX?" (With time for students to accept or reject the teacher's formulation)
Asking students to restate someone else's reasoning	"Can you repeat what he just said in your own words?"
Asking students to apply their own reasoning to someone else's reasoning	"Do you agree or disagree and why?"
Prompting students for further participation	"Would someone like to add on?"
Asking students to explicate their reasoning	"Why do you think that?" or "How did you arrive at that answer?" or "Say more about that."
Challenge or counter example	"Is this always true?" "Can you think of any examples that would not work?"

Adapted from O'Connor, R. and Michaels (in press). How (well structured) talk builds the minds. In Sternberg, R. and Preiss, D. (eds.), *Pathbreaking Discoveries in the Learning Sciences*.

Conclusion

In this article, we have suggested that strategies for adapting instruction to the needs of students range from formal external, standardized assessments to informal teacher judgments deeply embedded in day-to-day classroom practice. These strategies also differ in the epistemological assumptions underlying the assessment, with some grounded firmly in an associationist theory of knowledge, while others embody an interpretive theory of knowing and learning. We have argued that distinguishing clearly between these two aspects of assessment defines a two-dimensional assessment space (external-embedded and associationist-interpretive) that shows how conflicts can arise between the goals of reliable measurement on the one hand and agile response to individual students on the other.

Bibliography

- Beck, I. L. and McKeown, M. G. (2001). Inviting students into the pursuit of meaning. *Educational Psychology Review* 13(3) (Special Issue edited by A. Renninger and S. Wade), 225–241.
- Black, P. J. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice* 5(1), 7–73.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: some questions. *American Psychologist* 18, 519–521.
- Glaser, R. and Resnick, L. B. (1972). Instructional psychology. *Annual Review of Psychology* 23, 207–276.
- Kintsch, W. (1986). Learning from text. *Cognition and Instruction* 3(2), 87–108.
- Koedinger, K. R. and Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review* 19(3), 239–264.
- Koretz, D. M. (1998). Large-scale portfolio assessments in the US: Evidence pertaining to the quality of measurement. *Assessment in Education: Principles, Policy and Practice* 5(3), 309–334.
- Michaels, S., O'Connor, C., and Resnick, L. B. (2007). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education* 24, 283–297.
- Oakes, J. (1985). *Keeping Track*. New Haven, CT: Yale University Press.
- Popham, W. J., Keller, T., Moulding, B., Pellegrino, J. W., and Sandifer, P. (2005). Instructionally supportive accountability tests in science: A viable assessment option? *Measurement: Interdisciplinary Research and Perspectives* 3(3), 121–179.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* 18, 119–144.
- Stiggins, R. J. (2001). *Student-Involved Classroom Assessment*, 3rd edn. Upper Saddle River, NJ: Prentice-Hall.
- Stigler, J. and Hiebert, J. (1999). *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom*. New York: The Free Press.

Wiliam, D. and Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In Dwyer, C. A. (ed.) *The Future of Assessment: Shaping Teaching and Learning*, pp 53–82. Mahwah, NJ: Erlbaum.

Further Reading

- Andrade, H. L. and Cizek, G. J. (eds.) (2009). *Handbook of Formative Assessment*. New York: Taylor and Francis.
- Binet, A. and Simon, T. (1916). *The Development of Intelligence in Children*, Kite, E. S. (trans.). Baltimore, MD: Williams and Wilkins.
- Bloom, B. S., Hastings, J. T., and Madaus, G. F. (eds.) (1971). *Handbook on the Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record* 64(8), 723–733.
- Gardner, J. (ed.) (2005). *Assessment and Learning*. London: Sage.
- Looney, J. (ed.) (2005). *Formative Assessment: Improving Learning in Secondary Classrooms*. Paris: Organisation for Economic Cooperation and Development.
- Mehan, H. (1979). *Learning Lessons*. Cambridge, MA: Harvard University Press.
- Newell, A. and Simon, H. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice.
- Oakes, J. (1990). *Multiplying Inequalities: The Effects of Race, Social Class, and Tracking on Opportunities to Learn Mathematics and Science*. Santa Monica, CA: RAND.
- Piaget, J. (1960). *Psychology of Intelligence*. Paterson, NJ: Littlefield, Adams and Company.
- Resnick, L. B., Stein, M. K., and Coon, S. (2008). Standards-based reform: A powerful idea unmoored. In Kahlenberg, R. D. (ed.) *Improving on No Child Left Behind: Getting Education Reform Back on Track*. New York: The Century Foundation Press.
- Sternberg, R. J. (1985). *Beyond IQ: A Triarchic Theory of Intelligence*. Cambridge: Cambridge University Press.
- Terman, L. M. (1916). *The Measurement of Intelligence. An Explanation of and a Complete Guide for the Use of the Standard Revision and Extension of The Binet–Simon Intelligence Scale*. New York: Houghton Mifflin.
- Thorndike, E. L. (1922). *The Psychology of Arithmetic*. New York: Macmillan.
- Wiliam, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In Lester, F. K. Jr. (ed.) *Second Handbook of Mathematics Teaching and Learning*, pp 1053–1098. Greenwich, CT: Information Age.

Relevant Websites

- <http://www.aaia.org.uk> – Association for Achievement and Improvement through Assessment.
- <http://www.assessment-reform-group.org> – Assessment Reform Group.
- <http://www.instituteforlearning.org> – University of Pittsburgh Institute for Learning.

EDUCATIONAL ASSESSMENT – ENHANCING QUALITY AND USE

Contents

Assessment Conversations: Reading and Writing Conferences with Students and with Parents
Assessment: Pre-service and In-service Teacher Education
Classroom Assessment in Policy Context (Australia)
Classroom Assessment in Policy Context (England and France)
Classroom Assessment in Policy Context (French Sources)
Classroom Assessment in Policy Context (Hong Kong)
Classroom Assessment in Policy Context (New Zealand)
Classroom Assessment in Policy Context (Sub-Saharan Africa)
Moderation of Student Work by Teachers
School Policies and Practices to Support Effective Classroom Assessment for Learning
Student Assessment: Policy and Practice in Eastern Europe

Assessment Conversations: Reading and Writing Conferences with Students and with Parents

M Barrs, Centre for Literacy in Primary Education, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Defining the Territory

In exploring this aspect of assessment it has to be recognized that we quite soon come up against problems of terminology: these assessment events are variously referred to as conferences, consultations, discussions, or interviews. For the purposes of this article we will refer to them all as assessment conversations, a term that has already been used in another context by Johnston (2003).

All of these terms suggest some kind of formalization of classroom talk, the making of a special occasion for teachers to discuss students' progress in reading/writing or literacy with them – either in groups or, more commonly, with individuals. Sometimes assessment conversations may be held with students' parents, or sometimes conversations with students may include their parents.

Assessment conversations are often recorded in some way and the record becomes part of a teacher's ongoing assessment, or of a final overall assessment.

Purposes and Principles

In the field of literacy, there can be several purposes to this quite well-known but underdocumented practice.

In relation to conversations with students these may include:

- enabling the teacher to find out more about students' views of literacy and attitudes to literacy, both in general and in relation to their own work;
- enabling teacher and student to discuss the student's home literacy experiences and linguistic range;
- enabling the teacher to find out more about a student's literacy work by observing their approach to a literacy task (e.g., reading a text and discussing it);
- reviewing a student's reading or writing with them and evaluating their progress as a reader/writer; and
- encouraging students in reflection, making them more conscious of their own progress and of their strengths and needs

The purposes of assessment conversations with parents may include:

- enabling the teacher to find out more about students' home literacy experiences and attitudes to literacy;
- promoting discussion between teacher and parent of different aspects of students' home/school literacy; and
- reviewing students' reading or writing with their parents and evaluating their progress as readers/writers.

History of this Practice

There are a number of antecedents that may have contributed to the development of assessment conversations in literacy:

1. *Case studies in literacy research.* The use of case study in qualitative research in literacy has demonstrated how much can be learned about students' literacy development by observing them and talking to them about their reading and writing. Important studies such as Fry's (1985) *Students Talk about Books: Seeing Themselves as Readers*, *Inquiry into Meaning* by Chittenden *et al.* (2001), and Hugh and Crago's (1983) *Prelude to Literacy* have documented students' individual paths of literacy development and revealed their personal constructs of literacy.
2. *The analysis of reading strategies.* Miscue analysis was developed by Goodman (1973) in the late 1960s as a way of looking inside students' errors or miscues in the oral reading of a text. The Running Record, developed by Marie Clay in 1972, is essentially another type of miscue analysis, but is more suitable for younger children beginning to read. Both of these assessments involve teachers in listening closely to students' reading and recording/analyzing their errors as a guide to their characteristic behavior as readers; they are sometimes followed by conversations with the student about the text or about their reading in general.
3. *Classroom observation or kid watching.* Assessment conversations form part of a general approach to assessment in the classroom that includes the informal observation of children's reading and writing behavior. Teachers in the early years have been at the forefront of using child observation as a means of understanding and documenting students' learning. In the 1980s, teachers of older students began to observe students more systematically, recording their observations and reflections on what they were finding out about students' patterns of learning. Yetta Goodman coined the term kid watching to describe this kind of informal observation in 1978, and has written extensively about it since (Goodman and Owocki, 2002).
4. *Writing conferences.* A writing conference is a face-to-face discussion of a text between the student author of the text and their teacher. The term first came into general educational use through Graves' (1982) work in the late 1970s on students' writing processes, detailed in his book *Writing: Teachers and Students at Work*. Graves and his associate Lucy Calkins (1994), author of *The Art of Teaching Writing*, developed the use of writing conferences as a way of moving students on in their writing. This kind of interaction can be viewed as formative assessment; it allows teachers both to evaluate students' progress and to help them to develop their writing.

Conferences and Conferencing

Assessment conversations in literacy are sometimes known as conferences. Although this term – perhaps borrowed from parent–teacher conferences – came into use mainly in relation to writing, it is also used in relation to reading. Reading and writing conferences are usually informal, one-to-one conversations between teachers and students to discuss students' work. In this context, the word conference is sometimes used as a verb as well – to conference can mean to conduct writing or reading conferences with students.

Reading and writing conferences can range along a continuum from informality to formality. At the informal end of the continuum, teachers can drop by while children are writing and ask "How's it going?" (Anderson, 2000). This can lead to an informal conference in which teacher and pupil discuss the progress of a particular piece of writing. At this end of the continuum, the conference is as much a part of the everyday business of teaching as an informal assessment. Lucy Calkins, the American educationalist who has developed the workshop approach to teaching writing most fully, always includes mini-lessons in her approach to conferences.

However, informal conferences do not have to involve direct teaching; they can be opportunities for teachers to observe a child's literate behavior in some depth, in order to assess her/his needs and plan future teaching and learning activities. Therefore, a reading conference can take a form as simple as a child reading and discussing a picture book with the teacher, while the teacher notices and records all that she/he can observe about the child's approach to the reading and the text.

At the more formal end of the continuum, teachers can organize times for individual conferences with students to discuss their literacy progress, perhaps with special reference to their writing folders or reading journals. Conferences of this kind can provide valuable information about pupils' attitudes to their own literacy learning and can also involve discussions of their home literacy experiences, the range of those experiences, their personal tastes, and the social nature of their literacy – the involvement of family members in their learning. A record of language/literacy conference with a child is as follows:

'What Is A Ghost Going To Do' is D's favourite book. 'I like to be able to read so I can read stories to my sister, I sometimes make up stories to tell her. I'm not that good a reader 'cos sometimes I forget the words, I don't know them. I ask my sister and she gives me a clue. I read at home every night.' Given the choice D would watch television rather than read. 'I make up ghost stories and copy from books – my mum helps with spellings. My best bit of writing was when I wrote about the tele programme Dungeons and Dragons. Sometimes I'm a bit lazy with

handwriting and I chat too much but sometimes it's OK' (Barrs *et al.*, 1988).

A well-conducted individual literacy conference is an opportunity for a child to engage in self-assessment and to reflect on her/his own progress. But the way that children assess themselves as literacy learners will reflect the discourse of their classrooms. Johnston (1992), in discussing evaluation through interviews, observes that "It is probably worth noting that the kind of interaction you set up in interviews about books and about writing will form the basis of those interviews that take place between students." Conferencing helps to establish a classroom discourse in which listening to others and reflecting with others is normal.

Assessment Conversations as Part of Informal Assessment Systems

Assessment conversations form part of a range of techniques and practices that teachers now have at their disposal for evaluating students' literacy progress. These ways of assessing are sometimes collectively called informal assessment or authentic assessment. What they have in common is the fact that they are carried out in the classroom, in the course of normal classroom activities, and recorded by the teacher. These assessments are formative; they help the teacher to assess students' strengths and needs, and they support further planning.

Such informal assessments include teachers' diaries and anecdotal records, running records and miscues, reading records, reading and writing journals, the analysis of writing samples, and conferences with students about their reading/writing. Some of these practices are more highly structured, others less so. At best, they work together to give a full and rounded picture of students' progress, based on their daily behavior in normal classroom contexts rather than in test situations.

Effective teachers often include these informal assessments in their normal classroom practice. The survey conducted by Medwell *et al.* (1998) into the effective teaching of literacy in the UK suggested that a group of 228 teachers whom they had identified as effective practitioners were less likely than their validation (control) group of 71 teachers to rely on tests for their assessments in literacy. Instead, they were much more likely to use assessment techniques, such as marking, error analysis, and observation. When given a list of approaches to assessment in literacy and invited to add more, the effective teachers added a further 55 approaches, which included reading conferences – mentioned by 20% of those responding (but only by two of the validation teachers).

There have been relatively few attempts to fully incorporate informal assessment techniques into an official

system of assessment. Although many national and state systems (e.g., Florida, Kentucky, Ontario, and Tasmania) have recommended the use of a range of informal assessments as a part of a wider system of assessment, not many have attempted to blend these practices into a coherent whole.

Fewer administrations still have made assessment conversations part of such a system. This is not surprising: assessment conversations are a relatively unstructured form of assessment and depend heavily on a teacher's sensitivity and responsiveness in conducting a discussion with a student. Because these conversations are one-to-one events, they also take up substantial teacher time. The hesitancy around the use of assessment conversations in official systems means that there are few guidelines about how to conduct assessment conversations, either with students or with parents.

Often, early-years practitioners lead the way in making conversations with children and with parents a key part of an assessment system (Smidt, 2005). The New Zealand *Learning Stories and Teaching Stories* research report (Carr *et al.*, 2000) emphasized the necessary complexity of assessment in the early years and the importance of taking into account:

The funds of knowledge children bring to school, and the advantage to which they might be turned for literacy development, the languages the child brings and his or her ability to transfer one to the other, the permeability of classroom interactions around print to the language(s) and patterns of interaction in the child's home, the child's understanding of literate activity... what he or she can know and do under which circumstances....

In order to capture the complex nature of this learning, the project placed narrative assessment, including teachers' observations, transcripts, photos, children's work, and children's and parents' comments at the heart of its approach to assessment. Learning stories document children's engagement in learning experiences and are presented in children's portfolios for children, families, and teachers to read and discuss.

In the UK, the Primary Language Record (PLR; 1988) included conversations with both students and parents at the beginning and at the end of each school year. This practice provides teachers with information about children's home literacy and linguistic experiences early in the year, and encourages the review of children's progress over the course of the year. Assessments that derive from or are influenced by the PLR generally include both student and parent conferences. Following is a record of discussion between child's parents and the class teacher:

M. thoroughly enjoys reading, reading every evening. She selects books from school as well as the Public Library. She reads to her mother at the weekend. M. writes a lot

too, writing about special events and outings. She enjoys drawing cards and makes her own books. Loves to paint and draw.

M. is busy all the time helping her mother in the house. She enjoys some television, mainly Indian films. M. speaks Panjabi at home and is a fluent speaker. She enjoys maths at home and using text books from the library. (Ellis, 1993).

Conversations with students and with parents give teachers more insights into students' linguistic and cultural experiences and the implications for their literacy learning. They help teachers to see how they could plan their teaching better, and how students could be better supported in the classroom. In addition, assessment conversations can establish a dialog between teachers, students, and parents, which makes for better working relationships and better learning.

Links with Self-Assessment and Peer Assessment

One of the purposes of holding assessment conversations with students is to develop their confidence, their ability to reflect on their own learning, and their capacity for self-assessment. Such conferences require students to review their work and consider what they have already learned – they encourage self-evaluation. They make students more conscious of the learning strategies that they are using, and help them to become aware of their use of unhelpful or negative strategies.

The development of students' ability to discuss their learning and their understandings is an essential component of self-assessment. The UK-based Assessment Reform Group (2002) has for some years been promoting assessment for learning, as distinct from the assessment of learning, and has articulated ten research-based principles which characterize assessment for learning. These include:

Principle 9: Independent learners have the ability to seek out and gain new skills, new knowledge and new understandings. They are able to engage in self-reflection and to identify the next steps in their learning. Teachers should equip learners with the desire and the capacity to do this for themselves through developing the skills of self-assessment (Assessment Reform Group, 2002).

Assessment conversations provide an ideal opportunity to help learners to develop and practice these skills.

Links with Portfolio Assessment

Assessment conversations are sometimes structured by being based on a student's portfolio, and then they form part of portfolio assessment. Discussions between teacher

and student are generally seen as being a vital component of portfolio assessment.

In portfolio assessment, students make a collection of samples of their work over an extended period – the principle of compiling a portfolio is sometimes expressed as: collect, select, and reflect. Student-kept records of reading and other evidence of learning can be included in a portfolio. Teacher observations and notes on conferences can also form part of the portfolio. Every aspect of keeping a portfolio, from the initial selection of work to the review of the final collection, can be discussed with teachers and/or with other students. In the process, students' awareness of the development of their work – its range, variety, and its relation to key criteria – is enhanced. Several authorities on portfolio assessment, including Roberta Camp, Ellen Krogh, and Val Klenowski, have commented on the basis for dialog that portfolios offer and how they make learning more visible (Camp, 1998; Klenowski, 2002).

Portfolios can be used to support peer assessment, where students review their work together. Portfolio assessment can also involve parents; portfolios can be shared with parents at regular parents' evenings or parent-teacher meetings. More ambitiously, parents have sometimes been involved as partners and stakeholders in portfolio assessment, and assessment conversations around the portfolio have been conducted as three-cornered discussions between parent, student, and teacher.

Assessment Conversations with Parents

Assessment conversations with parents can go further than the discussion of students' portfolios. They can contribute to the establishment of links with homes that will support students through their school careers. Schools can also benefit hugely from parents' knowledge of their children's home literacy experiences, their interests, and their approaches to learning. This knowledge can complement teachers' own observations of students' school learning, and help them to arrive at more-informed judgments of their literacy progress.

However, some parents feel disenfranchised from school settings, and some schools do not do enough to reach out to parents who did not have good school experiences themselves. Gaitan (2001) has documented the difficulties that parents can experience in participating fully in their children's education and has explored the particular issues that affect bilingual Latino children's literacy achievement. She argues that interactions between Latino parents, and between parents and school staff, are more than usually important to children's progress; they enable parents to understand the system and begin to make their viewpoints heard, and can affect practice in schools.

Teachers are not always sufficiently well prepared by their training to conduct effective parent-teacher

conferences. A national study of teacher-education programs in the USA found that in most universities parental involvement was not a required course for trainee teachers. Although classroom teachers assert that working with families is important to children's positive school outcomes, they receive little formal training and preparation to work with parents.

In her book *The Essential Conversation*, the distinguished African American educationalist Sara Lawrence-Lightfoot (2003) analyses the tensions that can exist when parents and teachers come together in the parent-teacher conference, but argues that effective and authentic dialogs in this setting are both possible and necessary for children's achievement in school. She views the student as the missing link in the parent-teacher conference and advocates the inclusion of children as full participants in the conference.

Some pioneering classroom teachers arrived independently at the belief that parent-teacher conferences would have more meaning and be more effective if the child/student were involved. In a self-study of her professional practice, an Alaskan elementary teacher, Terri Austin, has described how she moved to student-led conferences with parents:

I created student-led conferences as a result of seeing students negated in the assessment process. They virtually had no voice in any step of the traditional reporting. Now at the end of each quarter, my students prepare a portfolio of work that best represents them as a learner. The student and I review the portfolio and complete the report card together. On one evening, all the students and their families come for our conference night. The student leads their own conference, showing their work and answering parent questions (Austin, 2000).

In recent years, student-led conferences have begun to be part of more general practice. Schools find that preparing for the conferences helps students to be more reflective and more effective learners. They do not always supplant traditional conferences, but can supplement them (Hiatt-Michael, 2001).

Whether or not students are included in parent-teacher conferences, the student is always the center of the conversation. It is particularly important that such important conferences should be true dialogs in which all parties can share their perceptions, stories, and their concerns. If the conference is open and interactive, issues may surface that will help both parents and teachers to support children better.

Recent Developments in Assessment Conversations

Conversations between students and teachers are the essential stuff of teaching and learning. Applebee (1996),

in his book *Curriculum as Conversation*, has argued that we should design curriculum around "domains for conversation" and then help students enter these conversations "seeking not consensus but understanding difference, the willingness to listen, and the ability to disagree."

This article has argued for assessment as conversation – if students' perspectives on their own learning are to be taken fully into account, assessment also must be designed to include conversations and conferences between students, parents, and teachers. This last section looks at two recent developments in assessment that give full weight to conversations between students and teacher-evaluators.

The Learning Record

The Learning Record is a technology-based development from the print-based California Learning Record, which was in turn adapted from the UK PLR (1988). The international collaboration between the UK originators of the PLR and the developers of the Learning Record in California and in the University of Texas at Austin dates back to the late 1980s.

Regular conversations between teachers and children, and between teachers and parents, were an integral part of the original PLR and of the Learning Record in California. This emphasis on the involvement of children and their parents contributed to the reputation of the Learning Record as an assessment responsive to bilingual children and to cultural difference. The Learning Record lends itself readily to use with bilingual children and has been translated into Spanish. A paper on its use with deaf and bilingual children (Allen, 2005) can be found on the website of the University of California's Linguistic Minority Research Institute.

Between 1988 and 1994, the Learning Record became the only alternative assessment approved by the California State Department of Education as an alternative to standardized testing. It also became the first initiative in the history of the Bureau of Indian Affairs (BIA) that was supported both by BIA officials and tribal elders. In a series of countrywide moderations attended by BIA and California teachers (1999–2004), it achieved an inter-rater reliability rate of 90–92% – a consequence of thorough professional development and clear, shared descriptions of progress (Hallam, 2001). The Learning Record has been strongly endorsed by the US organization Fairtest.

A later version of the Learning Record was originally designed as an online portfolio and this form was used by hundreds of teachers and students at 14 institutions of higher education in the USA. However, it has now been redesigned – as part of a drive to radical simplicity – and has become a stand-alone Word version of the record. In this version, it can be accessed and transferred from the Learning Record website, administered by Dr. Margaret

Syverson at the Department of Rhetoric and Writing, University of Texas at Austin.

In its current form, the Learning Record is an essentially dialogic record. It is an evidence-based assessment in which information is collected by the student and organized, using a simple Word or Rich Text Format (RTF) document linked to a selection of their work. The student's commentary on this evidence of their learning is an integral part of the portfolio.

The assessment of the record takes place as students discuss their progress with their teacher – either in a face-to-face conversation or an online one. The eventual negotiation of a grade, based on the evidence provided and on clear, shared criteria, is part of this dialog. Students are paired to read and discuss their learning records in a peer moderation process. Examples of students' records and their commentaries, of teachers' responses, and of the moderation discussion process can all be found at the Learning Record website.

Assessment Conversations in the New Zealand Education Monitoring Project

The New Zealand Education Monitoring Project (NZEMP) is one of the most developed forms of national monitoring, and one of its interesting features is the fact that it employs specially trained teacher administrators to conduct assessments. Among the task approaches used is a one-to-one interview with a teacher, with the whole session recorded on videotape, and scored centrally by a team of assessors.

A serious effort is made throughout the NZEMP to engage students, both through the content and design of the actual tasks and through the involvement of the teacher administrator. One of the key aims of the NZEMP is to gain information about students' attitudes to learning as well as their factual knowledge and concepts. A remarkable feature of the project is the way in which, following the assessments, students are asked to evaluate the tests, saying which they most enjoyed doing (Crooks, 2007).

The teacher administrator aims to create a genuinely communicative situation in which the student responds to her directly and personally. For instance, in the *Whaka-boaboa* (personal introduction) oral presentation task, reported in Reading and Speaking Assessment Results 2004, the teacher first invites the student to introduce him/herself. After this, the teacher introduces her/himself and then encourages the student to say more: "Is there anything else you would like to tell me about yourself?" This task is videotaped and then scored centrally, with assessors considering the content of the student's personal introduction and its presentation (organization, clarity, fluency, and overall effectiveness).

The one-to-one reading assessments include a retelling, with the teacher as a receptive audience. Another reading assessment, *Favourite Book*, presents students with pictures of children reading and simply asks the student to think of a book that they really enjoyed and tell the teacher administrator as much as they can about it. Although the task is a standard one, the benefits of having an encouraging listener in a one-to-one assessment situation like this are obvious.

The NZEMP also includes a range of group assessments and Terry Crooks, the co-director, describes why these are an important part of the assessment system:

From the outset, we were aware that much classroom learning takes place in groups, and that interpersonal attitudes, communication skills, and group problem solving processes would appropriately be assessed through the use of group tasks. Groups provide an appropriate context for some listening, speaking, drama, and musical performance activities, as well as for areas where debate and discussion are appropriate (such as social studies). A further strength associated with the use of group tasks is that collectively a group can undertake substantial tasks that would be beyond the capabilities of some individual group members. . . NZEMP team tasks are administered to groups of four children, who work on each task for 10 to 20 minutes while their work is facilitated by a teacher and recorded on videotape.

To reduce the reading load for students, many of the tasks are presented through audio or video clips.

The NZEMP has taken the idea of assessment conversations, both between teachers and students and between groups of students, a long way. It recognizes that the social context of assessment strongly influences what students are able to do, and sets out to provide supportive interpersonal contexts in which to assess children's learning. One of its most innovative features is that children are asked to evaluate the tasks they have undertaken, and to say whether they enjoyed them – these responses are taken into account in the design of future tasks.

The Validity of Assessment Conferences

Johnston and Yanoff (2002), in an article on early childhood assessment and family literacy, draw attention to the concept of consequential validity in assessment. Assessments which feed back positively into classroom practice are greatly to be preferred, they say, to narrow assessments that diminish the curriculum:

We have to think about the learning community – how assessment practices affect the teachers, parents, children, and families involved. (Johnston and Yanoff, 2002)

The consequential validity of literacy conferences with students and with children is high. As teachers take into

account learners' perspectives and self-evaluations, as they involve parents in dialog about their children's progress, they are recognizing the complex and personal nature of literacy learning and making strong links with children's families and cultures. In addition, they are discovering more about the sense that their students are making of school literacy; discoveries which can help them to reshape their curriculum and their teaching.

See also: Assessment in Schools – Literacy Writing (Extended); Assessment in Schools Related To Literacy: Reading; Formative Assessment; Portfolio Assessment.

Bibliography

- Anderson, C. (2000). *How's it Going? A Practical Guide to Confering with Student Writers*. NH: Heinemann: Portsmouth.
- Applebee, A. N. (1996). *Curriculum as Conversation: Transforming Traditions of Teaching and Learning*. Chicago, IL: University of Chicago Press.
- Assessment Reform Group (2002). *Assessment for Learning, 10 Principles: Research-Based Principles to Guide Classroom Practice*. London: ARG.
- Austin, T. L. (2000). Treasures in the snow: What do I know and how do I know it through my educational inquiry into my practice of community. <http://people.bath.ac.uk/edsajw/austin.shtml>. (accessed June 2009).
- Barrs, M., Ellis, S., Hester, H., and Thomas, A. (1988). *The Primary Language Record: Handbook for Teachers*. London: ILEA/CLPE.
- Calkins, L. (1994). *The Art of Teaching Writing*. Heinemann: Portsmouth, NH.
- Camp, R. (1998). Portfolio reflection: The basis for dialogue in ERIC Clearing House. c.72 no.1 p.10–12, Sept–Oct 1998.
- Carr, M., May, H., and Podmore, V. N. (2000). Learning and teaching stories: Action research on evaluation in early childhood education: Final report to the Ministry of Education. New Zealand Council for Educational Research.
- Chittenden, E., Salinger, T., and Bussis, A. M. (2001). *Inquiry into Meaning: An Investigation of Learning to Read*. New York: Teachers College Press.
- Crago, M. and Crago, H. (1983). *Prelude to Literacy: A Preschool Child's Encounter with Picture and Story*. Carbondale, IL: Southern Illinois University Press.
- Crooks, T. (2007). Philosophical choices and technical consequences in the design, execution, and reporting of results in New Zealand's National Education Monitoring Project (NEMP). *Annual Meeting of the American Educational Research Association*.
- Ellis, S. (1993). Changing the pattern. In Barrs, M. and Pidgeon, S. (eds.) *Reading the Difference: Gender and Reading in the Primary School*. London: CLPE.
- Fry, D. (1985). *Children Talk about Books: Seeing Themselves as Readers*. Milton Keynes: Open University Press.
- Gaitan, C. D. (2001). *The Power of Community*. Lanham, MD: Rowman and Littlefield.
- Goodman, K. (ed.) (1973). *Miscue Analysis: Applications to Reading Instruction*. ERIC/NCTE: Urbana, IL.
- Goodman, Y. and Owocki, G. (2002). *Kidwatching: Documenting Children's Literacy Development*. Heinemann: Portsmouth, NH.
- Graves, D. (1982). *Writing: Teachers and Students at Work*. Portsmouth, NH: Heinemann.
- Hallam, P. J. (2001).
- Hiatt-Michael, D. B. (ed.) (2001). *Promising Practices for Family Involvement in Schools*. Greenwich, CT: Information Age.
- Johnston, P. (2003). Assessment conversations. *Reading Teacher* 57(1), 90–92.
- Johnston, P. (1992). *Constructive Evaluation of Literate Activity*. White Plains, NY: Longman.
- Johnston, P. and Yanoff, L. (2004). *Early Childhood Assessment and Family Literacy in Handbook of Family Literacy Research*. Mahwah, NJ: Erlbaum.
- Klenowski, V. (2002). *Developing Portfolios for Learning and Assessment*. Oxford, Routledge.
- Lawrence-Lightfoot, S. (2003). *The Essential Conversation: What Parents and Teachers Can Learn from Each other*. New York: Random House.
- Medwell, J., Wray, D., Poulson, L., and Fox, R. (1998). *The Effective Teachers of Literacy Project*. Exeter: University of Exeter.
- Smidt, S. (2005). *Observing, Assessing and Planning for Children in the Early Years*. Oxford: Routledge.

Further Reading

- Barrs, M., Craig, D., Syverson, M., and Fiset, D. (1999). *Assessing Literacy with the Learning Record: A Handbook for Teachers, Grades K-6*. Portsmouth, NH: Heinemann.
- Clay, M. (1985). *The Early Detection of Reading Difficulties*. Heinemann: Portsmouth, NH.
- Fryar, R. (2003). Students assess their own reporting and presentation skills. In Fehring, H. (ed.) *Literacy Assessment: A Collection of Articles from the Australian Literacy Educators Association*. Newark, DE: International Reading Association.

Relevant Websites

- <http://www.fairtest.org> – Assessment of ELL Students under NCLB: Problems and Solutions.
- <http://www.cwrl.utexas.edu> – Computer Writing and Research Lab, the University of Texas at Austin.
- <http://www.inca.org.uk> – International Review of Curriculum and Assessment Frameworks Internet Archive.
- <http://nemp.otago.ac.nz> – New Zealand National Education Monitoring Project.

Assessment: Pre-service and In-service Teacher Education

A Davies, Classroom Connections International, Courtenay, BC, Canada

© 2010 Elsevier Ltd. All rights reserved.

Elizabeth and her high school colleagues had a problem. Numerous ninth-grade students were not academically ready to learn what needed to be learned. They teach at an urban high school of 2500 students that serves students from a large area including four housing projects. This means working with students who may enter ninth grade with behavior plans, identified learning deficits or identified as students with special needs, including English Language Learners.

While working with these students over time, a variety of interventions were tried teacher by teacher and student by student. While there was some success, there wasn't enough. Elizabeth and her colleagues expanded their informal conversations to include the school leadership team. They talked about what they wanted for these students. They came to agreement on what success would look like. Together, as they considered what had worked and what hadn't, a possible solution emerged.

Now a larger intervention is being tried by more teachers under the umbrella of a school-wide effort to help every student be successful – “whatever it takes.” New ninth-grade students identified as being at risk by their eighth-grade teacher are put into a special team for 1 year. The teacher to student ratio has been reduced because classes are taught by a team – one regular education teacher and one teacher qualified in special education methods. These students need the skills to learn and are given extended time to learn them. As part of this intervention teachers are deliberately using assessment for learning to support success for all students. This larger intervention isn't perfect but thus far it is working better than the previous patchwork of small interventions tried by a few teachers. Informally day-by-day and formally during the year and at the end of the year, student successes and challenges are reviewed. Then, based on the qualitative and quantitative data, the intervention is adjusted and renewed.

With a growing recognition of the complexity of teaching and learning, the term ‘professional learning’ has emerged to replace ‘teacher training’ to describe the ongoing education of teachers and other educators. The example above describes a continuous assessment for learning cycle that engages teachers in professional learning that leads to increased student learning and achievement. It is learning that arises from the needs of students under the district and state commitment to success for all learners. Results are documented by informal and formal qualitative and quantitative data (feedback)

collected continuously. It is supported through ongoing reflection and collaboration with colleagues. Over time, the feedback feeds forward to more learning as teachers reflect and review student evidence and help students set new goals that lead to further action and learning. This is professional learning supported by assessment for learning that is embedded in the work of schools and the professionals who work within them.

The focus of this article is on supporting educators to learn about classroom assessment through engagement in assessment for adult learning. It is important for educators to learn about and be able to use effective classroom assessment to both support learning and evaluate student progress and achievement. Second, it is important that educators be able to apply these practices for student learning. Third, educators need to experience and be supported to use assessment for learning during professional learning opportunities. This is illustrated through a series of vignettes, a description of the acquisition of essential skills and understandings about classroom assessment and a consideration of professional learning from an assessment for learning perspective. This learning needs to occur during preservice education and throughout an educator's career given continuous curricula renewal, changing student needs, and evolving community expectations. This perspective complements the work of Darling-Hammond *et al.* (2009) in a review of research related to professional learning and Timperley and Parr's (2004) research showing the need for evidence-based professional learning. Further, it is supported by research that demonstrates the importance of initial and ongoing teacher training in the area of classroom assessment (James *et al.*, 2007).

Classroom Assessment

Preservice and in-service educators need to learn about classroom assessment: research, theory, and practices. Research in the area of classroom assessment is comprehensive and shows that the way to support more learning is through increased formative assessment that involves learners in the assessment process (Crooks, 1988; Black and Wiliam, 1998; Hattie and Timperley, 2007). There are four areas of research that support work in this area: motivation, classroom assessment, feedback, and summative evaluation. Each area is linked to classroom assessment practices that support student learning.

Research: Motivation

Educators report, and research confirms, that when learners are involved in the assessment process, they:

- engage in learning,
- develop a sense of ownership and commitment to their learning, and
- make choices about what to focus on next in their learning.

When the focus is on learning and performance (rather than evaluation and judgment), learners are more likely to be intrinsically motivated and encouraged to take risks that challenge and expand their learning. When feedback is specific and descriptive, learners are more likely to remain focused on learning and make informed choices about what to do next. When learners make choices about their learning, achievement increases; when choice is absent, learning decreases. The use of rewards of any kind – marks, grades, and evaluation – can result in learners being de-motivated and less willing to take the risks needed to learn. While at times summative evaluation may be required, it is important to try to keep the feedback as specific and descriptive as possible (rather than evaluative) in order to support learning (Covington, 1992; Dweck, 2000; Hattie and Timperley, 2007).

Research: Classroom Assessment

Research shows that when learners are involved in the assessment process, they learn and achieve more. When learners are involved in their own assessment – co-constructing criteria, self-assessing in relation to criteria, giving themselves information to feed forward their learning, collecting, and presenting evidence of their learning, and reflecting on the strengths and needs that are revealed – motivation and achievement increase. Learners are able to set informed and appropriate goals when they apply feedback to feed forward into learning. When learners are involved in their own assessment, mistakes become information or feedback to use to adjust what they are doing. When learners pay attention to mistakes in their learning and identify ways to improve, they learn. This is self-assessment. Self-assessment can provide feedback information to inform and further the learning. This kind of feedback/feedforward information is essential for learning (Black and Wiliam, 1998; Crooks, 1988; Hattie and Timperley, 2007).

When learners are involved in the assessment of their work, they consider their own strengths and the areas they need to improve. By doing so they have the opportunity to make the most of their strengths and correct or compensate for their weaknesses. This leads to greater learning success. Black and Wiliam (1998) put it this

way, “The research reported here shows conclusively that formative assessment does improve learning. The gains in achievement appear to be quite considerable, and as noted earlier, amongst the largest ever reported for educational interventions” (p. 61).

Research: Quality Feedback for Learning

There are two general categories of feedback: specific, descriptive feedback and evaluative feedback. Each has a different impact on learning. Specific, descriptive feedback comes during as well as after the learning. It is formative. It says what is working and what needs to be adjusted or changed. It provides information about what needs to be learned next or done differently – it feeds forward to more learning. This kind of feedback is easily understood by learners and relates directly to their learning. It is specific so performance can improve. It is often given in comparison to models, exemplars, samples, or descriptions.

Evaluative feedback often comes at the end of the learning. It is summative. It tells the learner how she or he has performed compared to others (norm-referenced assessment) or in relation to what was to be learned (criterion-referenced assessment). Evaluative feedback is communicated using letters, numbers, checks, or other symbols; it is encoded. Learners who receive evaluative feedback usually understand whether or not they need to improve. However, unless learners can decode the evaluative feedback, they may not have enough information to understand what they need to do differently in order to improve. Researchers report that letter grades and other symbols that communicate evaluative feedback can have a negative effect on learning. The negative effects are most pronounced with low-achieving students. Learners with poor marks are more likely to see themselves as failures. Learners who see themselves as failures are less likely to succeed as learners (Butler, 1988; Hattie and Timperley, 2007).

Research: Summative Assessment

Summative assessment – evaluation – comes at the end of learning, while formative assessment provides information and support during the learning. Summative assessment aims to evaluate what students know, can do, and can articulate at a given point in time. This evaluation is reported verbally or in writing to others. Summative assessment is more reliable and valid when evidence of learning is collected from multiple sources over time and when the evidence is examined in light of quality expectations or achievement indicators. Summative

assessment, at the classroom level, is based on evidence collected both during the learning by students and teachers as well as evidence collected at the end of learning. Evidence of learning may include observations of students engaged in the process of learning, products students create, test results, and student articulations of their understandings as evidenced through sources such as teacher notes, student self-assessments, or recordings of discussions. Anything students do, create, or articulate is potentially evidence of learning. It is important that evidence of student learning be in relation to clear learning targets, be of high quality and free of bias (James *et al.*, 2007; Stiggins, 2007). In recent years, there has been growing concerns regarding the quality of teacher-made tests and the appropriate use of external examination results in summative assessments (ARG, 2006). It is important that educators learn how to use tests and the information they provide appropriately.

Summative evaluation requires sufficient evidence that the intended learning has been achieved. In the past, the assumption has been that such evaluation was best done externally – with tests and other forms of evaluation created and monitored by outside sources. What research has revealed, however, is that when teachers are involved in becoming assessment literate and engaged in a conscious development and application of consistent criteria for summative evaluation, valid and reliable summative evaluation of the learning are more likely to result. Further, when clearly specified criteria that describe progressive levels of competence and procedures are developed and used to judge student work for evaluation purposes, teachers are more able to reliably assess a greater range of classroom work. Looking at a greater range of student work as they apply shared criteria increases the validity of professional judgments (ARG, 2006; Sadler, 1989). It is important that classroom teachers understand the role of external evaluations in terms of:

1. how the information is used to impact classrooms and education practice;
2. informing teachers' understanding of the effectiveness of classroom programs and instructional techniques;
3. providing information concerning trends and patterns with regard to indicators such as student learning, student achievement, evolving needs of learners, and changing emphasis of curricula; and
4. informing system-level decision-making so appropriate supports and resources can be provided.

In summary, when preservice and in-service educators learn about classroom assessment research, theory, and practices, they are better able to support student learning through assessment. Educators can then thoughtfully employ classroom assessment practices such as:

- setting clear learning targets;
- using samples to show quality and possible pathways to success;
- co-constructing criteria about important products, processes or other evidence of learning;
- engaging students in reflection, self-assessment, and peer assessment using a common language;
- ensuring that they give themselves and receive from others specific descriptive feedback;
- collecting ongoing evidence of their learning;
- preparing collections of evidence to show proof of learning; and
- involving students in communicating proof of learning to an audience.

When students are involved in creating and collecting evidence of learning in relation to clear learning goals, they have a greater opportunity to show proof of learning and use the language of assessment. When they communicate proof of learning to others using the language of assessment they inform others, receive feedback, and can consolidate plans for next steps. Lastly, teachers themselves also collect evidence of learning from multiple sources over time in relation to clear learning goals. When it is time to report, teachers engage in a process of summative assessment – evaluation – that involves professional judgment.

Professional Adult Learning

Classroom assessment research, theory, and practices, while important foci for professional learning, have opened new possibilities for helping adults engage in learning themselves. In this section, research regarding effective professional learning is described, and then assessment – as a vehicle for adult learning – is detailed.

Ongoing research related to adult learning demonstrates that they need to be engaged in their learning in a variety of ways, both alone and with others. Efficiencies cannot be gained by doing the same thing, at the same time, and in the same way (Darling-Hammond *et al.*, 2009). Research has also shown that adult learning experiences that result in lasting changes to beliefs and practices that make a difference for student learning are multifaceted and occur over time (Timperley and Parr, 2004). Professional learning may focus on classroom teachers, school or district leaders, or any certificated adult who supports students or the adults who work with them so in this article the term adult learning (not training) is used. It is essential that professional learning in the area of assessment not be limited to teachers. Teachers who are most successful in implementing assessment for learning are supported in their learning by school leaders (James *et al.*, 2007). Further, when school and district leaders use

assessment for learning strategies in their work in schools, implementation success is enhanced.

Research has shown that in different educational contexts adult learning is supported when it:

- is purpose-driven and related to overall vision;
- engages learners so they learn by doing and provides for learning to emerge through interaction with others;
- involves adult learners in learning alongside colleagues in response to personally identified needs, concerns, or issues;
- is supported by reflection and focused conversation with the larger community;
- results in sustained commitment and re-commitment to improving their teaching habits since change – learning – necessarily occurs at the individual level;
- invites us to learn from others whose knowledge, expertise, and wisdom we respect, with learning being embedded in apprenticeship-like experiences that acknowledge the complexity of the work; and
- is continually informed by evidence (quantitative and qualitative) about student learning from students, from self, others, and the system that feeds forward to new learning and new actions (reflective feedback).

Professional Learning

Assessment for learning needs to be an essential part of any professional learning initiative because it helps adult learners understand what success is, have a common language to use with one another, as well as monitor and make adjustments in response to ongoing assessment information. When assessment is used to guide adult learning in the same way that classroom teachers use assessment for learning to guide and support student learning, effective research and experience-based practices are modeled while ongoing learning is supported. Assessment for learning supports the learning of educators as they:

- are involved in co-constructing criteria which lead to clear learning goals, shared understanding of success, and a common language;
- give and receive specific, descriptive feedback about the learning – during the learning – for themselves and for others through increased self- and peer assessment in relation to the criteria;
- collect, select, reflect, and communicate evidence of their learning;
- set realistic and manageable goals for next steps; and
- engage in providing evidence of their own learning in preparation for performance reviews.

In the following section various aspects of assessment for learning practices are described and their application to adult learning articulated: co-constructing criteria, specific feedback, evidence of learning, and summative assessment.

Co-Constructing Criteria

Ms. C. meets with teachers and co-constructs criteria around what's important when setting criteria with students. The teachers then reflect on their own practice, self-assess, and set goals for involving their own students in setting criteria. Teachers then set criteria with their students and later share what they did with one another, explaining what worked and what didn't work. They then set goals and return to fine-tune the process of co-constructing criteria.

Criteria define and describe evidence or proof that demonstrates learning and achievement. When adult learners are involved in co-constructing criteria about key processes or products related to classroom assessment, important ideas are made explicit to all learners in language that they share and understand. They begin to gain the knowledge they need to make decisions to help close the gap between where they are in understanding how to support students and where they need to be. The process of co-constructing criteria provides adult learners with opportunities to confirm, consolidate, integrate new knowledge, and make better choices for next steps. Adult learners also have opportunities to share their ideas and check their personal theories. They not only witness a key process being modeled, they experience it and, as a result, are better prepared to engage their own students in similar ways.

Professional development leaders sometimes choose to use samples and models to inform the criteria setting process. Adult learners are more likely to understand what success is when it comes to classroom assessment when they can examine samples of work, watch video footage that illustrates key concepts, listen to teachers and other educators describe their practices, or visit classrooms to watch classroom assessment in action. Samples are useful to help adult learners understand:

- the attributes of quality classroom assessment;
- development from novice to experienced to expert; and
- different ways to adapt what is being learned.

When adult learners are engaged in discussion with professional development leaders about the learning expected before an activity or task and then co-construct criteria, the conversation clarifies options, highlights possible pathways to learning, and shares learning information across the group. Knowing what they are learning and what must be present in their work gives adult learners the information they need to keep themselves on track.

Giving and Receiving Specific, Descriptive Feedback

A whole school faculty co-constructed criteria around quality classroom assessment prior to district staff conducting a

walk-through to observe students and teachers at work. When the guests arrived to look for evidence, classroom doors had signs highlighting what part of the classroom assessment process could be witnessed. The signs said, “As you walk through my classroom, please notice that we are...”. Teachers also used the criteria to assist their own self-assessment and provide feedback to each other as they work to improve classroom assessment.

Specific feedback is important for learning. Each time adult learners reflect on their learning using co-constructed criteria, they are better able to articulate their learning and give themselves and others specific, descriptive feedback that will feed forward the learning. Quality feedback gives adult learners specific information they need to adjust or change what they are doing so they can learn and do better. When adult learners are assessed during the learning and evaluated at the end of the learning, they have time to practice and extend their learning; and, they learn more.

Multiplying good-quality feedback leads to more learning for adults. The learning brain needs continuous feedback – much like a fish needs water. The amount of specific, descriptive feedback that feeds forward to more learning is increased when sources of feedback are increased. Consider the co-constructed criteria described earlier. Adult learners can use the language of criteria to give themselves and others specific information to improve. They can observe their peers and give specific feedback, analyze professional practice, and select two elements that are powerful and something specific that could be done to improve their own practice.

Collecting, Selecting, Reflecting, and Presenting Evidence of Learning to Others

School leadership teams meet as part of a long-term ongoing commitment to improving quality classroom assessment. They commit to working with their colleagues and collecting evidence of learning. At the beginning of each session and at the end of each year members bring evidence of learning forward to share with others. They also set goals for the next learning period.

It is important to collect evidence of the impact (or not) of a professional learning initiative. Everything a learner says, does, or creates is potential evidence of learning. Professional development leaders involve adult learners in collecting proof of their learning. Involving adult learners in collecting, selecting, reflecting, and presenting evidence of learning is important because:

- Adult learners have an opportunity to affirm their understandings as they select evidence of learning to

show others. This helps them reflect on what has been learned.

- Adult learners have an opportunity to articulate what they have learned and what they still need to work on. The process of articulating what has been learned and considering how best to communicate it to others helps the learner learn more.
- When adult learners show evidence of learning to an audience of peers, they receive feedback and recognition for their learning and everyone has an opportunity to learn more.
- When evidence of adult learning is collected from a variety of sources over time and examined in terms of patterns and trends, the reliability and validity of the findings is increased.

Adult learners vary in terms of their background experiences, knowledge, and skills (inherent and learned). They require different kinds of input and levels of support in order to learn. Often adult learners need different learning paths and different evidence is created along the way. As a result, the collection of evidence of learning becomes increasingly complex because learners produce different kinds of evidence to show that they have met the criteria.

Once adult learners have collected evidence of their learning, a next step shows their evidence to others since the presence of others influences what and how we present. Performance reviews that involve learners in communicating evidence of their learning assist adult learners to step back and reflect – to assess what they are doing and reexamine their current practices. Some of the ways adult learners present evidence of learning include sharing student work samples, professional portfolios, or postings online. No matter what the format, it is important that adult learners identify what has been learned, what the evidence shows, and ask for specific feedback that can help them frame their own next learning steps.

Selecting Next Steps to Feed Forward Learning (Goal Setting)

Ms. G. uses protocols to support adult learning by providing feedback for their work and to their peers. Some of the teachers with whom she has worked have further adapted the protocols for work with their students to improve feedback/feedforward. Students insist on receiving good-quality feedback that helps their learning improve.

When adult learners self-assess using co-constructed criteria, they come to understand what they have learned. As they reflect and present evidence of learning to others, they identify possible next steps needed for their

continued learning. This is feedback that feeds forward into further learning for adults. Goal setting is simply the process of deciding what feedback to take forward as next steps. Next steps are realistic, manageable goals in relation to co-constructed criteria. Research and theory indicates that closing in on a goal triggers a part of the brain linked to emotions, motivation, and learning (Csikszentmihalyi, 1997). When professional development leaders ask adult learners to reflect and set realistic, manageable goals for learning in this way – to look at what they are doing well and what they need to improve in relation to co-constructed criteria – they help them prepare to learn more.

Summative Assessment: Assessment of Learning

In order to bring alignment to a system-wide initiative at the secondary school level, district leaders revamped the performance review system so as to mirror the classroom assessment process. Educators were asked to set professional learning goals, collect evidence of their learning from multiple sources over time, and meet with their supervisors to consider their strengths, areas of learning needs, and to set goals for next steps.

As professionals learn more about summative assessment, it is important that they engage in processes to inform their professional judgment since professional judgment is a key aspect of summative assessment. Here are a few of many possible examples of educators engaged in processes of summative assessment:

- Engaging in the process of moderation of student work in relation to co-constructed criteria is a key element.
- Using structured learning protocols to give oneself and one's colleagues specific feedback.
- Using protocols to engage groups of educators in examining student work to enhance common understandings of the quality required to meet achievement indicators.
- Involving adult learners in showing proof of their own learning through collecting evidence of learning from multiple sources overtime in relation to shared criteria and using that evidence in the performance review process.

Complex learning systems collect evidence – feedback information – from adult learners as well as consider a wide range of other data from a variety of sources over time. Significant learning – whether individuals or systems – takes time. Data collected from yearly achievement tests, for example, will not provide immediate evidence. Evidence has to be based on more sensitive data. Consider the marketplace. Economists use leading and

lagging indicators to help explain how change occurs. Education research shows that increased student learning and achievement (lagging indicator) follows powerful professional learning (leading indicator). External examinations are an example of a lagging indicator. Classroom assessment evidence of learning provides leading indicator information.

While it is critical that professional learning result in increased student learning and achievement, the demand for immediate results is unrealistic as it does not give adult learners time to adapt ideas for their students and their context nor systems time to align their policies, regulations, and practices. In order for adult learning to be successful, it needs to be ongoing, sustained, and connected to other aspects of change in the school community.

Professional learning that supports and sustains life long, independent, self-directed learners draws upon educators' strengths, communicates self-worth and capability, and views participants' individual differences as a value-added benefit to the group. Professional learning that works is purposeful, grounded in work with students, relevant to the adult learner, and uses assessment for learning to initiate and sustain learning that makes a difference for adults and then students over time.

See *also*: Alternative Assessment; Assessment and the Regulation of Learning; Challenges of Developing and Implementing Formative Assessment Practices in Schools; Formative Assessment and Instructional Planning; Formative Assessment; Impact of Assessments on Classroom Practice; Peer and Self-assessment; School Policies and Practices to Support Effective Classroom Assessment for Learning; Summative Assessment by Teachers; The Relationship Between Assessment and the Organization and Practice of Teaching.

Bibliography

- ARG (Assessment Reform Group) (2006). The role of teachers in assessment of learning. *Pamphlet Produced for Assessment Systems for the Future Project (ASF)*, UK. <http://www.assessment-reform-group.org/ASF.html> (accessed October 2009).
- Black, P. and William, D. (1998). Assessment and classroom learning. *Assessment in Education* 5(1), 7–75.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology* 58, 1–14.
- Covington, M. V. (1992). *Making the Grade: A Self-Worth Perspective on Motivation and School Reform*. New York: Cambridge University Press.
- Crooks, T. (1988). The impact of classroom evaluation on students. *Review of Educational Research* 58(4), 438–481.
- Csikszentmihalyi, M. (1997). *Finding Flow: The Psychology of Engagement with Everyday Life*. New York: Basic Books.
- Darling-Hammond, L., Chung Wei, R., Andree, A., Richardson, N., and Orphanos, S. (2009). *Professional Learning in the Learning Profession: A Status Report on Teacher Development in the United States and Abroad*. Stanford, CA: National Staff Development Council.

- Dweck, C. (2000). *Self-Theories: Their Role in Motivation, Personality, and Development*. Philadelphia, PA: Psychology Press.
- Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of Educational Research* **77**(1), 81–112.
- James, M., McCormick, R., Black, P., et al. (2007). *Improving Learning How to Learn – Classrooms, Schools and Networks*. Oxon: Routledge.
- Stiggins, R. (2007). *Introduction to Student-Involved Classroom Assessment*, 5th edn. Columbus, OH: Pearson Prentice Hall.
- Timperley, H. and Parr, J. (2004). *Using Evidence in Teaching Practice: Implications for Professional Learning*. Auckland: Hodder Moa Beckett Publishers.
- Davies, A., Herbst, S., and Reynolds, B. P. (2008). *Transforming Barriers to Assessment for Learning*. Courtenay, BC: Connections Publishing.
- Deci, E. and Ryan, R. M. (2002). *Handbook of Self-Determination Research*. New York: University of Rochester Press.
- Glaude, C. (2005). *Protocols for Professional Learning Conversations*. Courtenay, BC: Connections Publishing.
- Harlen, W. and Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education* **10**(2), 169–208.
- Popham, W. J. (2009). *Instruction That Measures Up: Successful Teaching in the Age of Accountability*. Alexandria, VA: ASCD.
- Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* **18**, 119–144.

Further Reading

- Davies, A. and Busick, K. (eds.) (in press). *Assessment in Professional Learning*. Courtenay BC: Connections Publishing.
- Davies, A. and Busick, K. (eds.) (2008). *Classroom Assessment: What's Working in High Schools?* Courtenay BC: Connections Publishing.
- Davies, A., Herbst, S., and Reynolds, B. P. (2008). *Leading the Way to Making Classroom Assessment Work*. Courtenay, BC: Connections Publishing.

Relevant Websites

- www.assessment-reform-group.org – Assessment Reform Group.
- www.annedavies.com – assessment for learning.
- <http://www.essentialschools.org> – Coalition of Essential Schools.
- <http://www.tlrp.org> – Learning How To Learn.
- <http://www.lasw.org> – Looking at Student Work.
- <http://www.tki.org.nz> – New Zealand.

Classroom Assessment in Policy Context (Australia)

J J Cumming, Griffith University, Brisbane, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

The one certainty about education policy is constant change. Charting educational practice within policy is fraught with difficulties. Certainty of practice at a point in time can change through legislative action, public pressures, and financial restrictions. This article describes the commonalities of Australian classroom assessment policy and practices that have emerged from the vagaries of cultural practice to be recognized as the standards of practice expected across the nation. Necessarily included in the discussion are the policy directions emerging in the twenty-first century that may or may not reach maturity, and may or may not affect the basic principles underlying classroom assessment in Australia.

Purposes of Classroom Assessment in Australia

Classroom assessment in Australia serves many purposes, both to chart student progress in learning and to provide the basis for reporting and accountability. A review of future schooling states:

[t]he primary purpose of assessment is for use in providing teachers with the information they need to improve each student's results and parents with information on their child's progress (Federalist Paper 2, Council for the Australian Federation Steering Committee, 2007).

Classroom assessment is valued in Australia for the scope it provides to teachers and students to engage in rich and complex curriculum and assessment activities, increasing the likelihood of knowledge transfer to settings beyond school. As one curriculum authority notes, "assessment tasks such as independent research projects and major works are likely to be the most challenging learning" students will undertake (Board of Studies (NSW), undated).

The international focus on assessment for learning (Assessment Reform Group, 1999) is clearly compatible with the Australian philosophy that assessment is to improve student learning. In most Australian classroom assessment contexts, distinctions between formative and summative assessments are not a major focus. Assessment of learning is ongoing, with summative assessment a point-in-time summary of the student's achievement for reporting, drawing on both formal and informal assessment data (Assessment Resource Centre, NSW Board of Studies).

Classroom assessment in Australia, then, is firmly focused on enhancing student learning. Moreover, Australian education policy endorses the principle of inclusivity – schools are a microcosm of society. Classrooms include children with differing levels of academic achievement, abilities and disabilities, and sociocultural and language backgrounds.

Policy Influences in Australian Education and Classroom Assessment

Inclusivity and Diversity of the Classroom

Australian education has long catered to students with special needs, originally in special schools. However, Australian policy during the last decades of the twentieth century and early twenty-first century endorsed inclusion of all students, to the extent possible, within regular classrooms, with legislated equitable practice established by the Disability Standards for Education 2005 (Cth), including curriculum and assessment design to allow all students to demonstrate their knowledge and achievement (s 6.3). Students generally attend specialist schools by parental (and student) choice. Otherwise, regular schools across all levels of schooling are expected to provide appropriate support for students with intellectual disability and/or physical impairments.

Australian classrooms also reflect our multicultural society. Students from migrant and refugee backgrounds enter school at all ages, often with limited or no English-language skills. In New South Wales, more than one in every four students is from a non-English-speaking background, that is, a language other than English is spoken in the home.

Not only do students within a classroom represent a broad range of abilities and linguistic and cultural backgrounds, but Australian schools are also socioculturally and geographically diverse. While most Australians live in coastal regions, many children live in rural or isolated areas, attending very small schools.

The diverse nature of classrooms has impact for classroom assessment. Assessment guidelines published by the Australasian Curriculum, Assessment and Certification Authorities (ACACA, 1995), while primarily focused on assessment for high school certification, provide principles for good practice across all levels of education. Schools are expected to reflect and respect student diversity, and to value different knowledge and experience (ACACA, 1995: 1).

As knowledge, particularly early knowledge, develops within cultural contexts, Australian schools are encouraged to modify their curriculum, learning and assessment practices, even in the final years of schooling (QSA, undated), to reflect and integrate with their communities, in order to promote effective learning and engagement.

Consideration of diverse ways and means of learning, as well as identifying and supporting communities to integrate their own cultural content into the curriculum, must be taken into account. The NTCF (Northern Territory Curriculum Framework) provides scope for learners to demonstrate outcomes within a range of contexts, including localised contexts. For example, Two-Way learning emphasises the importance of locally developed learning programs that uphold the values and learnings of particular communities (DEET(NT), 2007: 1).

Assessment is also expected to incorporate a range of instruments, under a range of conditions, in a variety of modes, to meet the needs of different learners, and facilitate student demonstration of achievement rather than creation of barriers due to the form of the assessment (ACACA, 1995: 3). This expectation further facilitates the design of different forms of assessments to suit students with special needs or diverse cultural backgrounds.

Within this framework of diverse classrooms and communities, the expectation is that Australian teachers will be able to identify students' knowledge backgrounds and needs, cultural and linguistic diversity, and plan appropriate learning activities and a variety of assessments to chart learning. Australian teachers are given considerable responsibility to plan student learning and assess students' progress, including assessment for end-of-schooling certification. A fundamental principle underpinning this responsibility is that education systems and communities in Australia respect teachers' professional capacity (Cumming and Maxwell, 2004). However, regard for, and expectation of, teacher professionalism occur in a highly structured educational environment managed through a range of national and state and territory policies.

Professional Standards for All Teachers

The second major policy context for Australian classroom assessment is the expected standard of teacher professionalism. All states and territories in Australia, except one, require school teachers to be registered or accredited through a state or territory authority. At the time of writing, the remaining territory (the Australian Capital Territory (ACT)) was finalizing registration processes.

Registration requirements generally specify personal suitability to teach (a fit and proper person) and 4 years of formal tertiary (university) training in an accredited teacher program, in conjunction with studies in discipline

areas in one or two areas for high school teachers. Teachers may also become specialists in special needs, or areas such as music, languages, or physical education.

Teacher registration authorities establish accreditation guidelines for teacher education program content and structure, including specified periods of practical experience with school mentors. Authorities elaborate professional teacher standards across skill levels, from graduate to advanced, and may require ongoing professional development for registration retention.

Classroom assessment knowledge is identified by the authorities as an integral component of quality teaching and hence required in teacher-education programs. For example, in Tasmania, a commencing teacher is expected to know how to develop a teaching sequence and "assess, plan and teach for the learning needs of a range of students" (Teachers Registration Board, 2007). In Queensland, new graduates should be able to:

- identify and apply learning experiences that incorporate flexible individual and group learning, teaching, assessment and 'have a sound fundamental knowledge of ...'
- effective ... assessment strategies and resources where ICT is embedded (and)
- authentic literacy and numeracy assessment strategies for gathering information and making judgements about students' language, literacy and numeracy development (Queensland College of Teachers, 2007: 10–11).

The expected assessment skill requirements are clearly demanding. Criticism has occurred regarding the extent to which such assessment skills are realized for new teachers (House of Representatives Standing Committee on Education and Vocational Training, 2007: 8). However, such concerns may be more indicative of high expectations of beginning teachers by the Australian public, than lack of appropriate novice skill level.

Control of Schools and Prescribed Curriculum

The third and fourth policy contexts influencing Australian classroom assessment are state and territory syllabi or curriculum frameworks, and regulation of schools, respectively. State and territory government authorities (often tripartite representing government, independent, and Catholic-sector interests) develop syllabi in numerous discipline areas, providing general but official guidelines on topics for instruction.

Two-thirds of Australian students attend government schools with the remaining one-third attending nongovernment schools, both secular and nonsecular. Schools and instructional programs are regulated in two ways in Australia. First, all schools receive federal funding, either directly or through funding to the state or territory government. Federal funding provision legislation specifies a

range of school requirements, using the power of the purse (Schools Assistance Act 2004). While nongovernment schools were less monitored by such legislation until the 1990s, legislative control gradually increased. By the twenty-first century, legislative requirements paralleled those for government schools.

The second mechanism for controlling schools occurs at the state or territory level. To operate, schools must meet accreditation and registration criteria under relevant state or territory legislation, including teacher qualifications, resources, capacity to maintain the health and safety of students, and, most importantly, “curriculum (including the framework of the curriculum and the principles on which the curriculum is based) (that) meets the curriculum requirements for students attending government schools” (Education Act 2004 (Tas), s88(6)(c)). Thus, all schools, both government and nongovernment, must use curriculum developed or recognized by their relevant curriculum authority.

Breadth of syllabus, dimensions and progressions, and classroom assessment

In Australia, syllabus frameworks in general separate into two levels. The first framework, from kindergarten/preparatory year to year 10 (K-10), focuses on key learning areas identified in a national agreement, the Adelaide Declaration (MCEETYA, 1999). Eight key learning areas were endorsed for the compulsory years (generally 6–15 years inclusive) – the arts, English, health and physical education, languages other than English, mathematics, science, studies of society, and environment technology. The second framework of syllabi addresses the senior secondary years leading to high school graduation and certification. Senior students select study areas, typically studying five or six subjects.

K-10 syllabi and classroom assessment

Using the K-10 syllabi, schools are expected to plan and integrate assessment across classrooms, years of schooling, and curriculum. In Western Australia, each school is required to have a whole-school assessment policy, consistent with the ACACA guidelines. Other states and territories incorporate the ACACA guidelines in advice to schools about good assessment practices.

Syllabi developed across Australia emphasize the interconnectivity of teaching, learning, and assessment. Therefore, it is difficult to talk about classroom assessment independent of curriculum contexts. During the last decade of the twentieth century, building from initial attempts to develop a national curriculum, states and territories endorsed outcomes-based frameworks for K-10 syllabi, outlining curriculum topics and charting development across levels of proficiency. Outcomes-based syllabi describe performance criteria and knowledge bases to be achieved by a student. The principle is that a student progresses when they have mastered a level, independent

of their year level in school, or the progress of their classroom peers.

Teachers are expected to use outcomes-based syllabi and descriptors of achievement to design learning activities and assessments that enable students to develop and demonstrate the target outcomes. Best practice addresses both multidisciplinary outcomes and multiple levels. Students work alongside one another to attain different learning goals extending their previous achievements. Reports to parents describe the level of achievement attained by their child and the type of skill the child has demonstrated.

The concept of outcomes-based syllabi and assessment is familiar. The same principle underpins grading in many sports, such as the martial arts and gymnastics. Students attain a skill level, learn a new repertoire, are assessed in performance, and, if successful, attain a new level of achievement. The motivation is the advancement, and self-improvement against specified standards, with the learner assessed when ready, and the repeated failure experienced by many in traditional school examinations avoided.

Although outcomes-based syllabi had yet to be fully realized in Australian classroom practice, by the turn of the twenty-first century they started to fall from favor. Reasons included media barrage for a return to the basics, and simpler reporting to parents. The outcomes-based syllabi did have two, possibly fatal, flaws. A syllabus should provide general guidelines for topics and sequences of instruction. However, outcomes-based syllabi became very detailed. While such detail could have been helpful for teachers for planning assessment, the volume of information across eight subjects and multiple levels overwhelmed many. Teachers were inadequately prepared to develop holistic, integrated learning and assessment activities, with the default adoption for many a checklist approach and questions such as how many times a child had to demonstrate achievement of a skill – 1 time? 3 times? 10 times? – to be assessed as having met a level (Cumming *et al.*, 2006). More professional development and support materials may have addressed these issues.

The second flaw was that the detailed reporting to parents was often couched in educatese, that is, the language used by educational professionals to describe a learning outcome to each other, rather than language parents could understand.

Senior schooling curriculum and classroom assessment

Australia is noted internationally for its use of classroom assessment for high-stakes assessment. Again, centralized syllabi guide classroom instruction and assessment. Across Australia, school graduation certification and achievement reports are based on a mix of classroom assessment and external-curriculum-based examinations. For example, in

Western Australia, classroom assessments in some subjects are weighted equally with external examinations for tertiary entrance purposes. However, with the raising of compulsory education years in that state to 17 years of age, not all students will be undertaking programs of study for university entrance, and many may never be required to take an external examination. In Victoria, the weighting of school assessments versus examinations varies according to the nature of the subject, with areas with high practical components, such as art, having higher school-based weighting. For most subjects, the weighting is 50–50. Queensland and the ACT differ from the other states, as final certification of achievement is based solely on classroom-based assessment and achievement.

Classroom-based components of assessment for high-stakes certification operate within a system of central governance. Relevant curriculum authorities monitor classroom assessment programs with subjects studied for university entrance monitored most extensively. In Queensland, samples of student achievement are monitored against the descriptors of levels of achievement in the syllabus (moderation) to provide comparability across schools.

Across Australia, classroom assessments in the senior years are expected to follow the ACACA guidelines (ACACA, 1995: 3) that good assessment practice requires “explicit, clear, unambiguous criteria declared in advance”: the provision of information to students outlining the curriculum framework in a subject; the assessment activities that will be completed and evidence that an appropriate range of assessment modes will be used; how these assessments will be weighted toward an overall grade; and guidelines for the criteria and standards to be used for assessment. Australian classroom assessment generally endorses criteria-referenced assessment (Sadler, 1987), with assessments against descriptors of performance rather than relative to other students.

It should be noted that in addition to university prerequisite and general subjects, students in high school in Australia may study vocational training subjects, either alone or as part of recognized certification. While partnership with training providers may occur, school teachers with appropriate qualifications may both instruct and assess students, using a competency-based approach, against competencies outlined in national training packages. This area of classroom assessment is beyond the scope of this article.

Ensuring student work is their own

When classroom assessment is used for summative purposes and high-stakes assessment, concerns are raised regarding assurance that work is the student’s own, often argued as a reason against using classroom assessment for such purposes. A further issue is the balance between acceptable teacher guidance and formative assessment during the development of student’s work, and its

summative assessment. Curriculum authorities and schools in Australia have developed a range of strategies to address these concerns. For example, the Senior Secondary Assessment Board of South Australia (SSABSA) Supervision and Verification Policy suggests close monitoring of and discussions with students during the production of the work, undertaking some components of assessment tasks in class (particularly for larger activities), team work, and developing new assignment activities regularly. In New South Wales, a program for teachers and students, All My Own Work, takes an ethical education stance about the issue, to raise student awareness about scholarly activity and plagiarism.

Initiatives to assist professional practice in classroom assessment

Current education and policy agendas have highlighted the need to provide ongoing professional development support for teachers. While external accountability regimes can have negative impact on the breadth of classroom curriculum and assessment practice, they do draw attention to the need for teachers to collect evidence on student progress. Teachers’ intuition or internalized knowledge of a student is no longer sufficient – reportable information is necessary, whether formal or informal such as classroom observation notes, practiced very successfully in early years of schooling.

Work to enhance teacher assessment knowledge and practice is ongoing, including seminars and workshops around Australia. Exemplary assessment activities, often cross curricular with technological foundations, are available through websites for use and modification. These activities identify syllabus components, often providing criteria and standards or rubrics for assessing student work and student work to exemplify different standards. A major goal of such work is not just to offer professional development, but also to provide teachers with greater assurance regarding the reliability of their judgment of students’ achievement and performance relative to other students in the state or territory – especially valuable to teachers in remote areas. However, the weakness of the assessment-sharing activities to date is the focus on summative assessment and grading of student performance but with limited scope for providing feedback to teachers to help to improve student learning, and the imposition of tasks that may not be embedded within the curriculum and context of the classroom.

Increasing Federalization of Curriculum and Classroom Assessment

A major (fifth) policy direction influencing classroom assessment in Australia for the twenty-first century is increasing federalization of both curriculum and classroom assessment. The Commonwealth of Australia was

created as a federation of states in 1901, under a Constitution that promoted both the common good of the nation, the Commonwealth, while ensuring the rights of states to manage their own affairs. While areas were identified in the original Constitution as the responsibility or reserve of the states, education was not mentioned. Unlike constitutions of some other nations, the Australian Constitution is not interpreted as giving reserve powers to states for matters that are not explicitly allocated in the Constitution, although the Australian government may not legislate to interfere in the policymaking powers of a state (Engineers case).

While there are ongoing contestations about increasing federalization of education, the ministers of education for each state and territory have been in agreement that national goals of education and curriculum are desirable since the Hobart Declaration of 1989, revisited in the Adelaide Declaration that identified the eight key learning areas.

Concerns about the crowded curriculum and reemphasis of basic areas of learning such as literacy and numeracy, especially in the early years of schooling, led to reconsideration of these key learning areas by the state and territory ministers. New discipline areas under consideration to replace the key learning areas include English, maths and science (including physics, chemistry, and biology), languages, humanities and social sciences (including history, geography, and economics), the arts (performing and visual), health and physical education, and cross-disciplinary learning areas (technology including information and communication technology (ICT) and design, civics and citizenship, and business). Such new policy directions do not apparently reduce curriculum but the weight accorded to different discipline areas would change from year to year.

Further national agreement has occurred through national statements in English, mathematics, science, civics and citizenship, and ICT (MCEETYA, 2007a) that states and territories have incorporated in syllabi, referred to as essential learnings but with some interpretative differences. However, the national statements are linked to mandated national testing (through the funding provisions), and the latter more than the former may have most impact on classroom teaching and assessment.

While outcomes-based syllabi were being challenged in public forums, the Australian Government intervened for the first time in classroom assessment practice, mandating that all schools must report student progress twice yearly to parents, in “plain language,” and “relative to the performance of the child’s peer group at the school” (Schools Assistance Act 2004, ss13, 22). Regulations further stipulated that such reports must include “an assessment against achievement levels or bands defined by the education authority or school,” in bands that “must be labelled as A, B, C, D, E (or an equivalent),” “clearly

defined against specific learning standards,” and must include, for each subject, “the child’s achievement relative to the achievement of the child’s peer group at the school by at least quartile bands” (Schools Assistance Regulations 2005, ss 2.3, 3.3).

As noted, school funding is contingent upon legislative compliance. Although the Tasmanian State Government indicated initially that it would not comply with the reporting regulations, report formats deemed to meet the guidelines were developed and a compromise reached. School sectors with strong commitment to nurturing the individual student in a positive manner, such as the Catholic sector, were also critical of the new peer-comparison requirement. However, school systems can meet the requirements using the outcomes-based syllabus assessment framework. Some schools address the quartile regulation by providing information on the percentage of students achieving each of the A–E levels (Catholic Education Archdiocese of Brisbane (undated)). Reporting on quartiles against peer group poses more challenges for the many small rural schools in Australia, where total enrolment of children across all primary school years may number fewer than 25 students, and each year level may be represented by only one or two children.

Classroom Assessment and External Assessment in the K-10 Years

The final (sixth) policy context affecting classroom assessment in Australia is the balance and impact of external versus classroom assessments of student achievement across the years of schooling. From the mid-1960s to late twentieth century, with the exception of a year 10 certificate in New South Wales, student assessment and reporting in the compulsory years have been classroom based.

In the first moves toward national collaboration, the Adelaide Declaration, the states and territories agreed to assess and report literacy and numeracy achievement against nationally developed literacy and numeracy benchmark standards for years 3, 5, and 7, initially on a sample basis but becoming cohort testing. By 2007, further agreement was reached that common national assessments would be used. Such tests apparently make little intrusion into classroom assessment practices. States and territories have in general avoided school league tables, pointing to the diversity of classrooms and school contexts. However, new federal legislative requirements for public reporting have exerted pressure on schools and internal pedagogy and assessment practices. Evidence is emerging of the negative effects, noted in other nations, of reduction in the breadth of classroom instruction and assessment practices, and teaching to the test (Cumming *et al.*, 2006).

External monitoring of student performance is increasing. In conjunction with the literacy and numeracy testing

at years 3, 5, 7 and 9, national testing will be implemented in ICT literacy at years 6 and 10 and science literacy in year 6 (MCEETYA, 2007b). Over time, these programs of external monitoring must impact on classroom assessment practices in Australia.

Future Challenges

Classroom assessment in Australia provides a quality of practice and sophistication that can be taken for granted. However, even within this context of practice, four areas offer new or increased challenges.

Increasing Student Diversity in the Senior Years of Schooling

As noted earlier, an area of challenge for the future for senior schooling is the goal of increased school retention. Approximately two-thirds of the potential year level (67%) complete the final year of schooling, year 12, a retention rate that has been consistent for a decade. A further 15% complete year 12 equivalence through other education (MCEETYA, 2006). Increasing the age of compulsory education (or training or employment) will require an increase in study options to address the needs of all students. Increasing diversity in senior schooling will require considerable change in the assessment knowledge and practices of teachers who have previously worked within narrower discipline studies and with cohorts of students academically motivated by the goal of university entrance.

Diagnosis of Special Learning Needs

An area of classroom assessment needing attention is diagnostic assessment, the ascertainment of specific areas of need or learning difficulties, as opposed to formative assessment to shape student progress. The corollary of inclusive practices is that classroom teachers encounter students with diverse learning needs, many of whom may have learning difficulties or disabilities that have not yet been identified. If a teacher considers the student has a special need, the student is referred to a specialist for assessment. The critical issue is that classroom teachers need to identify when such a referral should occur. Evidence of lack of knowledge to do so is emerging through the law courts. Legal challenges are already occurring in Australia and the United Kingdom by students whose disabilities were not appropriately diagnosed in time to provide intervention.

The extent to which teachers are provided with support and guidance in this area of classroom assessment is still limited. Special instructional programs in early reading and

mathematics that often incorporate ongoing diagnostic and formative assessment activities are in place in schools. In Queensland, the Year 2 Diagnostic Net provides a structured interview and resources for teachers to individually assess students' strengths and needs in literacy and numeracy. Initially, all students were assessed. However, the Net is now used for students believed to be at risk, negating its effectiveness in identifying students who might otherwise escape teacher notice and the Net. Victoria is undertaking work in this area with an online interview in mathematics to assist teachers to identify students' strengths and needs in this area (DEECD, undated).

The Role of Technology in Classroom Assessment

The expectations that teachers maintain comprehensive records of student assessments against curriculum outcomes or essential learnings, the use of portfolios to record students' work across the years of schooling, and comparative reporting of student achievement to parents have placed considerable demand on teachers to maintain records and materials. States and territories, and government and nongovernment sectors, are using technological resources to assist teachers. Electronic records, the provision of exemplars previously discussed, and online recording of student outcomes are some of the technological solutions in progress, and which can be expected to expand exponentially in the twenty-first century.

Assessing All Goals of Learning

State and territory syllabi fulsomely articulate learning goals that are not just content based but also directed at what has previously been called the affective domain. All Australian education authorities aim to develop lifelong learners with skills in learning how to learn, effective Australian and global citizens, critical thinkers and problem solvers, and persons with resilience and a range of other personal characteristics – incorporating what have been referred to as conative knowledge (Snow, 1989), personal learning (Cumming, 2001), or learning power (Broadfoot, 2007). While regarded as an essential component of the K-10 syllabi, this area has yet to be developed across all states in classroom assessments and reporting. The Victorian Department of Education has developed frameworks and assessment tools for personal learning that give the fullest guidance to teachers. Given endorsement in the goals of education for the twenty-first century of the principles espoused in conative and personal learning, assessment in this area is a challenge to come. The classroom assessment practices of Australia provide the most opportunities for such an assessment reform to occur.

Conclusion

Classroom assessment in Australia is regarded as a fundamental dimension of education: to help students, identify needs, promote further learning, and to report achievement. Teachers are trained professionals who make judgments about student performance within a structured environment of syllabi and with increased central governance, as assessments become more high stakes. The benefit of classroom assessment approaches in Australian schooling has been the breadth of learning and performance facilitated by classroom assessment. Differences among the states and territories do occur. These can be seen either as detrimental to Australian national unity and cohesion, or as laboratories that enable experimentation with curriculum and assessment. The latter view allows best practices to be identified and used in other states. However, when state and territory assessment practices are examined, strong commonalities emerge. A watching brief for classroom assessment in Australia during the twenty-first century will be the impact of increasing federalization of curriculum, and testing, on what have been to date highly student focused and enriched classroom assessment practices with successful learning outcomes.

See also: Moderation of Student Work by Teachers; Summative Assessment by Teachers; The Multiple Purposes of Assessment.

Bibliography

- Assessment Reform Group (ARG) (1999). Assessment for learning: Beyond the Black Box. <http://www.assessment-reform-group.org> (accessed June 2009).
- Australasian Curriculum, Assessment and Certification Authorities (ACACA) (1995). *Guidelines for Assessment Quality and Equity*. Brisbane: Queensland Board of Senior Secondary School Studies.
- Board of Studies (NSW) (undated). *HSC Assessments and Submitted Works*. http://www.boardofstudies.nsw.edu.au/manuals/advice_parents.html (accessed June 2009).
- Broadfoot, P. (2007). *An Introduction to Assessment*. Norfolk: Continuum.
- Catholic Education Archdiocese of Brisbane (undated). *School Reporting FAQs*. <http://www.bne.catholic.edu.au/asp/index.asp?pgid=10778> (accessed June 2009).
- Council for the Australian Federation Steering Committee (2007). *Federalist Paper 2. The Future of Schooling in Australia*. Melbourne: Department of Premier and Cabinet.
- Cumming, J. J. (2001). Valuing learning and learners: The forgotten dimensions of educational assessment worth remembering in the 21st century. *Paper Presented at the 30th International Conference of the International Association for Educational Research*. May 2001, Rio de Janeiro, Brazil.
- Cumming, J. J. and Maxwell, G. S. (2004). Assessment in Australian schools: Current practices and trends. *Assessment in Education: Principles, Policy and Practice* **11**, 89–108.
- Cumming, J. J., Wyatt-Smith, C. M., Elkins, J., and Neville, M. (2006). *Teacher Judgment: Building an Evidentiary Base for Quality Literacy and Numeracy Education*. Brisbane: Queensland Studies Authority/Griffith University.
- Department of Education and Early Childhood Development (Vic) (DEECD) (undated). Resources for teachers – prep to year 10 resources. *Mathematics Online Interview*. <http://www.education.vic.gov.au/studentlearning/teachingresources/year10to12.htm> (accessed June 2009).
- Department of Employment, Education and Training (Northern Territory) (DEET(NT)) (2007). *Northern Territory Curriculum Framework (NTCF)*. http://www.det.nt.gov.au/education/teaching_and_learning/curriculum/ntcf (accessed June 2009).
- Disability Standards for Education (Cth) 2005*.
- House of Representatives Standing Committee on Education and Vocational Training (2007). *Top of the Class. Report on the Inquiry into Teacher Education*. Canberra: Commonwealth of Australia.
- Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) (1999). *Adelaide Declaration on National Goals for Schooling in the Twenty-First Century (Adelaide Declaration)*. http://www.mceetya.edu.au/mceetya/adelaide_declaration,11576.html (accessed June 2009).
- Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) (2006). *National Report on Schooling in Australia*. http://cms.curriculum.edu.au/anr2006/pdfs/2006_appendix1_stats.pdf. Appendix 1 Statistical Annex. (accessed June 2009).
- Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) (2007a). *The Statements of Learning*. http://www.curriculum.edu.au/mceetya/statements_of_learning,22835.html. Appendix 1 Statistical Annex (accessed June 2009).
- Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) (2007b). *2008 National and International Assessment Program*. [http://www.mceetya.edu.au/mceetya/national_assessment_program_\(nap\),16358.html](http://www.mceetya.edu.au/mceetya/national_assessment_program_(nap),16358.html). Appendix 1 Statistical Annex (accessed June 2009).
- Queensland College of Teachers (2007). *Program Approval Guidelines for Preservice Teacher Education*. Brisbane: Queensland College of Teachers.
- Queensland Studies Authority (QSA) (undated). *Benefits of Moderated School-Based Assessment*. Brisbane: QSA. <http://www.qsa.qld.edu.au> (accessed June 2009).
- Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education* **13**, 191–209.
- Schools Assistance Act (2004). *Schools Assistance (Learning Together — Achievement Through Choice And Opportunity) Act 2004*.
- Schools Assistance Regulations (2005). *Schools Assistance (Learning Together — Achievement Through Choice and Opportunity) Regulations 2005*.
- Snow, R. E. (1989). Toward assessment of cognitive and conative structures in learning. *Educational Researcher* **18**(9), 8–14.
- Teachers Registration Board (Tasmania) (2007). Tasmanian professional teaching standards framework graduate standards. <http://www.trb.tas.gov.au> (accessed June 2009).

Further Reading

- Amalgamated Society of Engineers v Adelaide Steamship Co Ltd (Engineers case)* [1920] HCA 54.
- Education (Accreditation of Non-State Schools) Act 2001* (Qld).
- Education Act 1972* (SA).
- Education Act 1990* (NSW).
- Education Act 1994* (Tas).
- Education Act 2004* (ACT).
- Education Act* (undated) (NT).
- Education and Training Reform Act 2006* (Vic).
- Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) (2005). *National Report on Schooling in Australia, ch. 5. Participation and attainment*. http://cms.curriculum.edu.au/anr2005/ch5_particip.htm (accessed June 2009).
- Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) (2005). *National Report on Schooling in Australia, 2002–2005, Appendix 1: Statistical annex, Table 21 (2001 data)*. http://cms.curriculum.edu.au/anr2005/pdfs/2005_stats.pdf (accessed June 2009).
- School Education Act 1999* (WA).

Relevant Websites

- <http://www.austlii.edu.au>– Australian Commonwealth, State and Territory Legislation.
- <http://www.boardofstudies.nsw.edu.au>– Board of Studies NSW.
- <http://www.curriculum.edu.au>– Curriculum Corporation (for links to all national, state and territory educational organizations, both government and nongovernment).
- <http://www.curriculum.wa.edu.au>– Curriculum Council.
- <http://www.decs.sa.gov.au>– Department of Education and Children's Services (DECS), Government of South Australian.
- <http://www.deet.nt.gov.au>– Department of Education and Training – Northern Territory Board of Studies.
- <http://www.education.tas.gov.au>– Department of Education, Tasmania.
- <http://www.mceetya.edu.au>– Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA), for joint papers and discussion by Ministers of Education across Australia and the *National Reports on Schooling in Australia*.
- <http://www.qsa.qld.edu.au>– Queensland Studies Authority.
- <http://www.vcaa.vic.edu.au>– Victorian Curriculum and Assessment Authority (VCAA).
- <http://www.vit.vic.edu.au>– Victorian Institute of Teaching (for links to Teacher Accreditation bodies in Australia).

Classroom Assessment in Policy Context (England and France)

M Osborn, University of Bristol, Bristol, UK

M Raveaud, Université du Maine, Le Mans, France

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Assessment is a social practice and social product that entails a conception of learning and of the learner. It is value laden and socially constructed (Broadfoot *et al.*, 2001). In this article, we compare assessment policy and practice in two countries with very distinctive educational priorities and traditions, England and France, to illustrate the extent to which classroom assessment is culturally situated and closely linked with the key values of an educational system (Planel *et al.*, 2000).

The article contrasts the policy context in England and France, and draws upon comparative empirical research on assessment practice in both countries. It points not only to a shared experience of a gap between policy and practice, but also to the cultural embeddedness of assessment practice.

The Policy Context for Assessment in England and France

The article assumes that the English policy context will be more familiar to readers (Gipps *et al.*, 1995; Scott, 2007). This is therefore outlined briefly while more space is given to describing the context in France.

England

Currently, in England, all pupils of years 2 (approximately 7 years old) and 6 (approximately 11 years old) in primary school are assessed through statutory assessment involving teacher assessment (TA) and a combination of standard tasks and/or tests.

In addition to statutory assessment, ongoing assessment by teachers has become a more overt feature of classroom life in England. The PACE (Primary assessment, curriculum, and experience) study which studied teachers, schools, and classrooms over 6 years of reform from 1989–95, found that the majority of teachers believed that assessment was the aspect of their practice and skills that had changed most significantly following the continuous policy changes of the last 20 years (Osborn *et al.*, 2000; Pollard *et al.*, 2000). The introduction of more rigorous assessment of subject knowledge as specified by the national curriculum, introduced by the Education Reform Act (ERA), 1988, represented a move

in the direction of more overt and explicit assessment. Many teachers hoped to be able to develop their required TA as formative assessment and to build it as an integral part of their classroom practice. However in the early days of the reforms, the scale and complexity of the record keeping and evidence gathering required, threatened to overwhelm teachers and Gipps *et al.* (1995) reported that TAs remained intuitive rather than being evidence based. In addition, the requirement for teachers in years 2 and 6 to aggregate their assessment for each attainment target, for each child, into a national-curriculum level, which was to be reported alongside each child's scholastic aptitude test (SAT) results, was seen as an attack on teacher professionalism (Osborn *et al.*, 2000).

Virtually all schools in England have explicit assessment policies and teachers are encouraged to collect the following for each child in their class: individual reading record; individual maths record; child's grouping for maths/English and other subjects; an individual portfolio containing samples of a child's work; TAs plus task or test results; and national-curriculum summary record, showing work covered and individual progress. While much of this evidence would have been collected prior to the ERA, the records became more detailed and more directly tied to national curriculum attainment targets as the PACE study progressed. With the increasing age of pupils and pressure of national testing, there was also a decrease in the use of more formative approaches to assessment, such as listening and observing, and an increase in activities, such as marking work and recording, and regular classroom tests. The PACE study (Pollard *et al.*, 2000; Osborn *et al.*, 2000) and the research of Gipps *et al.* (1995) suggest that there is much evidence that teachers have developed increased skills and confidence in their assessment practices. However, the PACE data also showed that although the national tests may be accommodated by teachers, the effect on classroom practice is profound. Besides the more obvious sessions of revision and teaching to the test which were observed in year-6 classes, there was also increasing evidence of a wash-back effect which reached into year groups that were not immediately involved with the tests. This appeared to be diminishing the opportunities for teachers to work in a way that enabled them to develop the whole child and address the social concerns of parents and the wider society.

Paradoxically, while British government policy was moving in the direction of more overt, explicit, and

summative assessment, similar to a traditional French model, French policy discourse appeared to be moving toward a more traditional English emphasis although changes in practice did not necessarily follow.

France

As in England, policy discourse in France has adopted research-based categories distinguishing summative and formative (in which case the French official terminology is diagnostic) assessment.

The French policy context has been characterized by a top-down drive toward diagnostic assessment as part of a wider trend to use assessment – not only of students but also of systems, of management procedures, of institutions – as a form of governance in an increasingly decentralized education system. Assessment, particularly summative assessment, has traditionally had a very high status in French education. However, the two major education acts passed recently (*Lois d'orientation* of 1989 and 2005) have attempted to introduce a culture shift from summative to diagnostic/formative assessment. While they stem from opposed political views (left-wing government in 1989, right-wing in 2005) with diverging views about the missions and management of the education system, both resort to diagnostic assessment as a tool to identify student needs, and thus ultimately improve the education system and increase school effectiveness.

Pedagogically, these reforms have contributed to making differentiated teaching possible and to providing (or imposing) tools for diagnostic assessment. For instance, curricular aims, traditionally set out as annual goals, have been restructured into triennial cycles to encourage a degree of differentiation in teaching instead of making pupils repeat a year. Formative assessment tools have been supplied by the Ministry, complete with keys for

analyzing their results, to encourage teachers to plan their teaching according to students' needs. A pool of such resources is available online for all levels, and national compulsory evaluations for 8-year-olds and 11-year-olds are carried out at the beginning of a new cycle. **Table 1** summarizes major reforms in this attempt at a culture change and their expected impact on classroom practice.

So far, these top-down initiatives appear to have had little impact on classroom practice as described in academic research and inspection reports (IGEN-IGAEN, 2005). Traditional practice remains largely unchanged in France despite research pointing to the limits of summative assessment as practiced in the French context. For instance, the pedagogical benefit of repeating a year has been called into question. Similarly, pseudoscientific annual results expressed as averages out of 20 remain the basis for determining whether a pupil should be held back a year despite evidence that grading varies depending on factors such as the teacher carrying out the assessment, the order in which papers are marked, pupil characteristics such as gender or sociocultural background, school and class variables. Classroom-assessment practice in France typically continues to focus on subject-specific aims in 97% of cases rather than cross-curricular aims (31%), with assessment of knowledge (84%) taking precedence over that of skills and attitudes (29%), and it remains under the control of the class teacher with little collaborative planning, coordination, or sharing of resources with colleagues, and an equally infrequent use of resources such as the Ministry's pool of tests (*banque d'outils*) (Ministère de l'Éducation Nationale, 2005a).

Teaching practice in France has proved largely resistant to policy-led changes, due to a culture of summative assessment rooted in a strong consensus regarding the educational and cultural values underlying the aims and practice of classroom assessment (Broadfoot *et al.*, 2001;

Table 1 Policy reform in France and intended impact on classroom practice

Year	Reform	Intended impact
1989	<i>Cycles</i> : restructuring the national curriculum into cycles (key stages) instead of setting annual targets	Allows for a degree of pedagogical differentiation, maintaining common goals to be reached by all pupils, but within 3-year cycles
1989	National compulsory evaluations with formative aims at ages 8 (CE2) and 11 (6 ^e), carried out at the beginning of a new cycle	Top-down introduction of compulsory formative-assessment tools, complete with feedback mechanisms for teachers to plan their teaching accordingly
2005	<i>Socle commun</i> : lists the knowledge, skills, and attitudes that each child must be equipped with, by school-leaving age	Change in relevant assessment: replacing averages with a binary notion (acquired or not) that excludes compensation between knowledge of one notion and ignorance of another
2005	<i>Banque d'outils d'aide à l'évaluation</i> : a pool of formative assessment resources set up by the Ministry for all levels of compulsory schooling	Resources and guidance including tests and analysis of student responses so as to identify the areas of knowledge and skills that have not been acquired
2005	<i>Programme personnalisé de réussite éducative</i> (PPRE): students are entitled to personalized support programs if they appear likely to fail to meet end-of-cycle targets and to acquire the <i>socle commun</i>	Cross-disciplinary remedial or preventive approaches based on assessment to identify target students, set up their PPRE and monitor progress; partnership between various education, social, and healthcare actors

Planel *et al.*, 2000). While policy stresses formative assessment to plan teaching using pedagogical differentiation to remedy individual students' identified weaknesses, dominant teaching practice relies heavily on summative assessment expressed as grades that are perceived as an objective measurement of student performance against standard curricular aims that are identical for the whole class and based on the national curriculum rather than previous student attainment (Raveaud, 2004). In other words, teachers tend to make little use of new statutory possibilities for classroom autonomy and remain attached to the tradition of universalism and national uniformity in teaching, perceived as a safeguard against arbitrary treatment of students. French teachers resist longitudinal monitoring of student progress and the transmission between classes or cycles of student profiles, as such mechanisms are considered as giving students a criminal record (IGEN-IGAEN, 2005: 54) that could lead to perpetuating prejudice.

Comparing National Assessment and Classroom-Assessment Practice in England and France

This section considers similarities and differences in both the national tests and in classroom-assessment practice in England and France. In both countries, there is evidence of discrepancies between the aims, content, and ideology of national policy discourse and of national tests and actual classroom practice. For example, while in France the emphasis of policy discourse is on the use of national tests for diagnostic evaluation for teaching purposes, in England the aim of national testing is to monitor and evaluate output and to raise standards through market competition between schools. As we saw earlier, in France,

national assessment policy has had little impact on classroom practice, whereas in England, most teachers believe that they have changed their practice in terms of assessment more than in any other area. In the PACE and STEP (Systems, teachers and education policy) studies, 84% of teachers said that they had given assessment skills more importance following reforms as against 69% of French teachers. Almost one-third of French teachers felt that their practice had not changed at all, following reforms (Broadfoot *et al.*, 2001). In England, there is also considerable concern about washback effects of national testing on classroom practice and on teacher and learner perceptions (Scott, 2007).

Table 2 summarizes the key areas of difference in the English and French national tests at the end of primary and beginning of secondary schooling.

So although there are similarities between the aims of the two countries in terms of their prioritizing of educational outcomes and the monitoring of the progress of the reforms they have introduced, there are also key differences in their educational ideologies. The introduction of national testing caused much tension between politicians and professionals in England, whereas in France there appeared to be greater consensus between policymakers and professionals about testing, reflecting the traditional values of universalism. The 1988 ERA in England, which introduced national testing, reflected a fundamental change toward the marketization of education where schools would compete for clients. The *Lois d'Orientation sur l'éducation* of 1989 and of 2005 were less politically charged and did not represent such a major shift (Broadfoot *et al.*, 2001). In both theory and practice, French national testing is more directly supportive of professional educators and the French national tests were designed to guide teaching and learning while the English tests functioned more for purposes of accountability and influencing the curriculum and teaching at a time when the

Table 2 The key areas of difference between English and French national tests at the end of primary/beginning of secondary schooling

	England	France
National test aims and uses	Monitor and evaluate output, indirectly raise standards through competition between schools	Diagnostic evidence for teaching purposes over time, consideration of trends over time, awareness of socioeconomic variables
Relationship between curriculum and test	Matching width, approach, and differentiation	Matching a narrower but deeper curriculum; matching uniform level and subject compartmentalization
Test structure; administration and marking	Differentiated, individualized approach, more subjective marking	Uniform structure, more objective marking
Relationship between test and learning	Matching open and more experimental approach; matching emphasis on ability	Matching more closed, formulaic, and abstract approach; matching emphasis on knowledge and skills learnt

From Planel, P., Broadfoot, P., Osborn, M., Sharpe, K., and Ward, B. (2000). National assessments: Underlying cultural values revealed by comparing English and French national tests. *European Journal of Education* 35(3), 363.

government was seeking to carry through its educational reforms.

These comparisons and those that follow were carried out as part of a series of research projects conducted in both France and England since the 1980s. These include the Bristaix project that ran from 1984 to 1987, the PACE project that ran from 1989 to 1997, the STEP project (1992–93), the Quality of education systems transnationally (QUEST) project (1996–98), and the Education and national culture: A comparative study of attitudes to secondary schooling (ENCOMPASS) project that ran from 1998 to 2000 (Broadfoot *et al.*, 2000; Osborn *et al.*, 2003; Raveaud, 2006). These various studies embraced both teachers' and pupils' perspectives, policymaking, and classroom practice in England and France. They were linked by the use of a comparative approach to illuminate key features of the teaching and learning process and focused particularly on the impact of the educational policy change in both countries and the challenge these posed for teachers to change, in more or less fundamental ways, their professional perspectives and ways of working. This article draws upon all of these to illuminate cultural influences on the meaning and practice of assessment as perceived by both the teachers and the pupils in the study with the aim of demonstrating the relativity of this aspect of classroom experience. In particular, the QUEST study used specifically devised literacy and math assessments as well as interviews, questionnaires, and observation to establish the ways in which pupils' performance differed from one country to another and to explore the pedagogical and cultural reasons for those differences. The study aimed to analyze the significance of national differences in teaching and learning in England and France and to compare pupil perceptions of learning and assessment.

The study found that the underlying national attitudes and traditions were, to a certain extent, resistant to change. Despite reforms that brought more control to the center in England and attempts to create more diversity in France, the English system continued to be characterized by individuality and diversity, while the French system remained a centralized, bureaucratic structure.

Thus, although the *loi d'orientation* (1989) in France was introduced to promote a more child-centered approach to education and aimed to increase school autonomy in the organization of learning, the classroom practice we observed had not fundamentally changed. The focus of lessons in France was still on the acquisition of skills to standardized national levels, a transmission style of teaching, and moving the class forward as a single unit. The French classrooms we observed were work centered and relatively formal. Teacher expectations of performance were explicit and children knew precisely how well they were doing on the basis of the regular, public allocation of marks. Poor performance in class could mean public disapproval and even humiliation by the teacher. Pupil

satisfaction was often linked to evidence of academic achievement and pupils were motivated to try harder. Pupils were able to discuss their progress or lack of it in an articulate way and were aware of reasons for their success and failure. Academic success was seen by both teachers and pupils as closely linked with effort, which may also account for the continuing practice of holding students back a year rather than differentiating teaching to deal with the range of attainment in the classroom.

In contrast, English primary education was informed by a continued concern for the whole child and for the school as a community. There was an emphasis on trying to make learning fun and on a broadly based curriculum but this very focus on the whole person led to a tendency both for teachers to judge pupils in terms of their personal and social characteristics and an innate level of ability, and for pupils to develop an identity as a learner on those terms (Broadfoot *et al.*, 2000, 2001). Although English primary teachers emphasized a positive and pupil-centered evaluation of performance and tried to minimize the effects of overt, categorical assessment, many English primary pupils were less motivated and positive about school than their French counterparts.

Overall, the QUEST study found that, despite some recent convergence between the values underpinning the French and English systems, English pupils are still expected to reach widely varying levels of attainment which government policy discourse sees as reflecting their innate ability. In contrast, French pupils are expected to strive to achieve a common level of attainment and this achievement is seen as reflecting the amount of effort they are prepared to exert. This emphasis on common rather than differentiated learning outcomes has an impact on assessment criteria. When we compared French and English teachers' approaches to assessing children's writing (Osborn and Planel, 1999), we found that French teachers expected correctness and conformity to a predetermined plan, while teachers in England looked for creativity and divergence and emphasized process rather than product. French teachers tended to make greater use of negative language and sanctions, using finite marks where English teachers made open-ended comments.

These findings were echoed in the research carried out by Raveaud (2006) with younger children in England and France. Observations were carried out in 15 classrooms of pupils aged 4–6, between 1998 and 2001, and continued in 2007–08. Assessment practice in these classes confirmed the culturally situated nature of key educational concepts, and acted as a lens revealing different priorities in the English and French education systems (Raveaud, 2004). One example shows the extent to which the notion of a mistake is socially constructed, for instance, when children are learning to write. In the classes observed in England, in parallel with an emphasis on spelling, much writing was linked to communication and

expression, whatever the pupils' degree of competence in spelling. Many emergent writing strategies were encountered, where young children wrote the sounds they heard, leaving out the parts of difficult words they could not identify. In this way, a high-attaining 4-year-old wrote "The b- is in a b- h-as n in l- to go on a l-" (The butterfly is in a butterfly house and likes to go on a leaf.). This method was compatible with the national curriculum which, despite a concern with spelling, states that pupils should be taught to write familiar words and attempt unfamiliar ones (DfEE/QCA, 1999: 48).

By contrast, French pupils were given fewer opportunities to write their own texts: they are intended to spell words correctly as soon as they start writing so that they do not memorize mistakes. Opportunities for creative writing are limited in French children's early schooldays, until they have mastered enough spelling and grammar to enable them to produce correct texts. Until then, much writing consists in copying from the blackboard or from a book, in line with the French curriculum: "until basic skills (forming letters, mastering the main rules of spelling) have been acquired and become automatic, it is difficult for the child to concentrate fully on more delicate tasks such as ordering information, structuring a text or composing their own sentences" (Ministère de l'Éducation Nationale, 2002: 87). As a result, what would be considered as a mistake in French classes was tolerated, if not encouraged, in English infant classes where pupils gradually built up their own understanding of sound-symbol relationships. The emphasis shifted from the outcome of a task to the act of learning, highlighting the product or the process of learning.

The form and scope of assessment also differed. While teachers in both countries gave oral and written feedback to pupils, French teachers were more likely to limit the written component of assessment to a grade, while English teachers favored comments not only on cognitive outcome but also on efforts, progress, and neatness. These factors were also taken into account in French classes, but only orally. The mark was conceived as a measurement of work against an absolute norm that could not be relative to an individual child's efforts or prior attainment, whereas English schools sought to reflect the stage reached in the development of pupils' learning. The French notion of an external, criterion-referenced standard of attainment gives way to a more diffuse evaluation akin to Bernstein's invisible pedagogy where a pupil's behavior and dispositions are taken into account, thus laying open more of the child to the teacher's assessing eye.

Some of these dimensions of difference found in these studies are summarized in **Table 3**.

It is striking that in some of the dimensions identified here, the policy discourse in each country appeared to be shifting each education system closer to the other,

Table 3 Dimensions of assessment

	<i>England</i>	<i>France</i>
Emphasis on	Process of learning	Product of learning
Assessment reference	Relative to child's attainment	Absolute, criterion-referenced
Classroom-assessment tasks	Frequently differentiated by attainment	Identical for all
Form of written assessment	Comment	Mark, grade
Dimensions taken into account	Prior attainment, effort, neatness	Cognitive outcome
Institutional stakes for students	Grouping, occasionally setting	Repeating a year (redoubler)

Adapted from Broadfoot, P., Osborn, M., Planel, C., and Sharpe, K. (2000). *Promoting Quality in Learning: Does England Have the Answer?* London: Cassell.

although the classroom practice observed held closer to the original dimensions and was more resistant to change.

Teachers' response to assessment in the two countries has also been very different. As Alexander (2001) argues, while teachers in many countries talk about assessment, it is above all in England that teachers see what they call the burden of assessment as oppressive and alienating. As he puts it, following the reforms of the late 1980s and 1990s, teachers in England found assessment burdensome "because it was outside their control, because it had been imposed on them in a form and at a time that they were not in a position to question" (p. 372) and because it had become heavily politicized.

Although there is evidence that teachers in England now feel more at ease with the current assessment requirements, it is still the case that there has been concern among educators and parents with what is seen as the overassessment of children and a possible link with anxiety, unhappiness, and a loss of motivation in learning. Certainly, the PACE study (Pollard *et al.*, 2000) found that children's feelings about testing became increasingly negative as they progressed through school with a peak in year 6 at the key stage-2 tests. Currently, schools and teachers are under pressure to demonstrate constant improvement in performance as a result of target setting and due to parental pressure, Office for Standards in Education (Ofsted) inspections, and publication of results (Osborn *et al.*, 2000; Scott, 2007). As James (2000) notes, the pressure on schools and teachers is far from accidental but is a form of intentional washback based on the expectation that the publication of league tables of performance would lead to open competition for students and push schools to improve. Teachers in England feel that their work

is evaluated more by the test results than by their TA. As one of the teachers in Scott's (2007) study argued:

I think most people look at the test results. I don't think they give two splodges of marmalade for what I put. . . . at the end of the day it's like 74% of people passed their test and got at least a Level 2 rather than me saying well, I think he can do (such and such) (Scott, 2007: 37).

In contrast, teachers in France generally feel positively about assessment and share similar goals. Ninety-five percent of secondary school teachers agreed that the purpose of assessment is to measure students' attainment, with 20% only using it as a decision-making tool (Ministère de l'Éducation Nationale, 2005b). It is also the case, as the following quote from the QUEST study illustrates that there has been little compulsion to change their assessment practice in the classroom. A typical comment from a French teacher was

Les ministres passent, les enseignants restent et évoluent à leur rythme. (Ministers (of education) come and go: teachers stay and change in their own time.)

Or:

Il y a les belles idées; et il y a ce qui se passe en classe. (There are beautiful ideas and there is what happens in class.)

Two factors, however, have led to a growing diversity in assessment practice over the past decade. The first is related to student characteristics: in underprivileged schools where academic achievement tends to be weaker, there is evidence of teachers adapting their academic goals, methods, and assessment practice, for instance, by lowering academic standards so as to avoid failure on a massive scale, or by taking into account relational dimensions and not only cognitive outcomes. Second, changing teacher characteristics have led to renewed ethical references shifting away from egalitarian principles and a strict republican ethos toward more student-centered concerns. What change has been observed in classroom practice in France may be more strongly related to these two factors than to top-down policy reforms.

Conclusion

The purpose of comparing these two assessment cultures has been to demonstrate the cultural situatedness of assessment beliefs and practices. As Broadfoot argues, in a sense the French system has had to move toward the mode of control traditional in English education, and central to the efficacy of these changes has been a very substantially increased role for assessment of students, of

schools, and of the system as a whole. Assessment is an inevitable feature of both classroom life and of institutions and educational systems as a whole. It is one of the most powerful forces shaping learning but we should not lose sight of the extent to which it is a social, culturally embedded process and will thus vary considerably from one educational culture to another. The article has emphasized the differences as well as the similarities of assessment approaches taken by the two education systems with differing values and priorities and the impact of this on teaching, learning, and pupil identity. The English education system, even after the 1988 ERA, takes a wider view of the pupil than does the French system, informed by the republican model of social justice that requires assessment to be conceived, as far as possible, as an objective, universal measurement. Such cultural differences in educational values are conveyed to pupils through varying assessment discourses and practices, and lead to differing pupil expectations of the learning process. The "effective harnessing of assessment to promote learning" (Broadfoot *et al.*, 2001: 59) depends on our understanding of the way in which assessment conveys these cultural messages and of the effect this has on learner identity.

See also: Classroom Assessment in Policy Context (French Sources); Impact of Assessments on Classroom Practice; Social Practices in School Assessment and their Impact on Learner Identities; Summative Assessment by Teachers; The Multiple Purposes of Assessment.

Bibliography

- Alexander, R. (2001). *Culture and Pedagogy: International Comparisons in Primary Education*. Oxford: Blackwell.
- Broadfoot, P., Osborn, M., Planel, C., and Sharpe, K. (2000). Promoting Quality in Learning: Does England Have the Answer? London: Cassell.
- Broadfoot, P., Osborn, M., Sharpe, K., and Planel, C. (2001). Pupil assessment and classroom culture: A comparative study of the language of assessment in England and France. In Scott, D. (ed.) *Curriculum and Assessment, International Perspectives on Curriculum Studies*, vol. 1, pp 41–62. Westport, CT: Ablex Publishing.
- DfEE/QCA (1999). *The National Curriculum Handbook for Primary Teachers in England*. London: QCA.
- Gipps, C. V., McCallum, B., and McAlister, S. (1995). *Intuition or Evidence? Teachers and National Assessment of Seven Year Olds*. Buckingham: Open University Press.
- IGEN-IGAEN (2005). *Les Acquis des élèves, pierre de touche de la valeur de l'école? Report 2005-079*. Paris: Ministère de l'Éducation nationale.
- James, M. (2000). Measured lives: The rise of assessment as the engine of change in English schools. *Curriculum Journal* 11(3), 343–364.
- Ministère de l'Éducation Nationale (2002). *Qu'apprend-on à l'école élémentaire?* Paris: CNDP/XO.
- Ministère de l'Éducation Nationale (2005). *Les pratiques d'évaluation des enseignants au collège. Les Dossiers* 160.
- Ministère de l'Éducation Nationale (2005). *Évaluer autrement*. Paris: MEN-Desco.

- Osborn, M., McNess, E., Broadfoot, P., Pollard, A., and Triggs, P. (2000). *What Teachers Do? Changing Policy and Practice in Primary Education*. London: Continuum.
- Osborn, M., Broadfoot, P., McNess, E., et al. (2003). *A World of Difference: Comparing Learners across Europe*. Maidenhead: Open University Press.
- Osborn, M. and Planel, C. (1999). Comparing children's learning, attitude and performance in French and English primary schools. In Alexander, R., Broadfoot, P., and Phillips, D. (eds.) *Learning from Comparing: New Directions in Comparative Educational Research. Vol I: Contexts, Classrooms and Outcomes*, pp 259–281. Symposium Books: Oxford.
- Planel, P., Broadfoot, P., Osborn, M., Sharpe, K., and Ward, B. (2000). National assessments: Underlying cultural values revealed by comparing English and French national tests. *European Journal of Education* **35**(3), 361–374.
- Pollard, A., Triggs, P., Broadfoot, P., McNess, E., and Osborn, M. (2000). *What Pupils Say: Changing Policy and Practice in Primary Education*. London: Continuum.
- Raveaud, M. (2004). Assessment in French and English infant schools: Assessing the work, the child or the culture? *Assessment in Education* **11**(2), 193–211.
- Raveaud, M. (2006). *De l'enfant au citoyen*. Paris: PUF.
- Scott, C. (2007). Stakeholder perceptions of test impact. *Assessment in Education* **14**(1), 27–49.

Further Reading

- Anderson-Levitt, K. (2002). *Teaching Cultures: Knowledge for Teaching First Grade in France and the United States*. Cresskill, NJ: Hampton Press.
- Aubégnay, J. and Clavier, L. (2006). *L'Evaluation, entre permanence et changement*. Paris: L'Harmattan.
- Broadfoot, P. (1996). *Education, Assessment and Society*. Buckingham: Open University Press.
- Gipps, C. V. (1994). National curriculum assessment: A research agenda. *British Educational Research Journal* **18**(3), 277–286.
- Gipps, C. V. and Cumming, J. (2005). Assessing literacies. In Bascia, N., Cumming, A., Datnow, A., Leithwood, K., and Livingstone, D. (eds.) *International Handbook of Educational Policy 13*, pp 695–714. Dordrecht: Springer.
- Goldstein, H. (2001). Using pupil performance data for judging schools and teachers: Scope and limitations. *British Educational Research Journal* **27**(4), 433–442.
- Osborn, M. and Broadfoot, P. (1992). A lesson in progress? Primary classrooms observed in England and France. *Oxford Review of Education* **18**(1), 3–15.
- Sharpe, K. (1992). Catechistic teaching style in French primary education: Analysis of a grammar lesson with seven year olds. *Comparative Education* **28**(3), 249–268.

Classroom Assessment in Policy Context (French Sources)

D Laveault, University of Ottawa, Ottawa, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

The increased recognition of the impact of better and more efficient classroom-assessment practices on student learning has led government and local administrations to develop policies on classroom assessment to capitalize on its benefits. Such policies are varied and depend on national and local context. French sources are used in this article to report on how policy decisions have influenced classroom assessment to enhance its quality and use. French sources used here are limited to four countries which share French as an official language of instruction and which have contributed several scientific publications on the subject: Belgium, Canada, France, and Switzerland. All four countries are part of the same international organization, *l'Association pour le développement des méthodologies en éducation* (ADMEÉ) and share the same scholarly journal *Mesure et évaluation en éducation* (MÉEÉ). Both have contributed to the scientific exchange of information among the four countries and to the development of some distinct perspectives on classroom assessment.

Four main issues seem to have shaped the way educational policies influence classroom-assessment practices and have emerged as crucial in French sources in the context of school reform and in the consequent renewal of classroom-assessment practices:

1. the impact of competency-based programs on classroom assessment;
2. the priority on the formative function of classroom assessment;
3. the communication of classroom-assessment results;
4. the influence of standards and external testing on classroom assessment.

Before addressing each of these issues, some clarification about the meaning of French sources and classroom assessment is necessary. This will help us understand the terminology and the scientific context and background associated with classroom assessment in French-language publications.

French Sources

Each country's governing jurisdictions have addressed classroom-assessment issues through their own system of rules and laws. International comparisons are thus dependent on each country-specific context and should be interpreted with great care. On the one hand, Belgium, Canada, and Switzerland have more than one official

language and are highly decentralized federations. In these countries, education is under the state or other local jurisdictions. On the other hand, France is a republic with a single official language and a lot of decisions on education are centralized at the national level. Each country or jurisdiction posts its policies and the approved classroom-assessment material on the Web (see list of websites).

A few comparative studies among those countries have started to emerge within the last 10 years, especially in the context of the ongoing school reforms. Such reforms have presented classroom assessment with important challenges and have contributed to the publication of an important documentation. Several of these publications include references in both French and English. A study of the references cited in papers published in *Mesure et évaluation en éducation* from 1999 to 2005 (Laveault, 2006) revealed that 41% of the references came from English sources as compared to 46% from French sources. The most frequent sources reported were scholarly journals in English (26%) followed by books published in French by European editors (23%). References to scholarly journals in French accounted for no more than 12% of total references.

The previous statistics tend to show that French sources refer to English and French sources in almost equal proportions. Among European authors, books are the most important vehicle of publications in French. Forty percent of the references from European authors originate from books. Such statistics are also indicative of the scarcity of French scholarly journals in the field of educational assessment.

Classroom Assessment: French and English Meanings

Classroom assessment is best translated in French as *évaluation en salle de classe*. The French word *évaluation* has two possible translations in English: assessment and evaluation. This should be taken into account when reading directly from French sources. Scallon (2004) imported the word assessment into French. Such a practice is not widespread, but is symbolic of the need to differentiate the timely operation of professional judgment (evaluation) from the more global task of collecting and organizing a large volume of information and appreciations on student achievement (assessment). This conceptual

distinction is in line with English definitions of the term assessment and has taken on an increased significance with the renewal of classroom-assessment practices which have occurred in the context of school reform. While assessment and evaluation may be differentiated conceptually, they may not be or need not be in actual practice. Most representative of this issue is the growth of competency-based classroom assessment where assessment can hardly be separated from instruction and the complex situations which are developed for student learning.

The Impact of Competency-Based Programs

Competency-based programs of studies have come to replace learning objectives defined in terms of measurable and observable outcomes in Belgium, France, Canada (Québec), and Switzerland. Although the definition of what is meant by competency may not be exactly the same in each country, they all involve the capacity to act effectively by mobilizing an integrated set of resources (knowledge, skills, and attitudes) to work out a family of situations. Such curriculum developments have been highly influenced by the French socioconstructivist theories which emphasize the central role of the student and social interactions in the construction of knowledge. Such changes in program descriptions have had a strong impact, from the standpoint of both the curriculum developer as well as the teachers' and students' classroom-assessment practices. They have resulted in program descriptions of disciplinary and cross-sectional competencies which provide teachers with more holistic and situated descriptions of learning acquisitions.

From the perspective of curriculum development, the description of competencies has required innovative ways to organize a program of studies while trying to maintain a certain level of alignment with teacher's classroom assessment (Jonnaert and Masciotra, 2007). Important distinctions have been made by Le Boterf (1994) to define more precisely what a competency is and what it is not. Though the notion of competency involves some form of mobilization, integration, and transfer of knowledge and skills, it is not limited to any one of these conditions. It is not to be confounded with terminal objectives of a program or with objectives of knowledge integration. Finally, competencies are inferred not only from the resources to be mobilized but in the sequence of actions necessary to mobilize such resources in the appropriate situations.

From the teachers' perspective, the interpretation of a competency-based program requires that they develop a clear understanding of the way competencies are developed and assessed throughout the curriculum. The notion of family of situations introduced by Roegiers (2004) is central to this goal. It requires teachers to think of a set of

problems, complex assessment tasks, or projects that will have students mobilize the same pool of resources in a variety of meaningful situations. Such complex tasks will involve one or more competencies and will present students with situations they have not yet encountered. To demonstrate their competencies, students may use different procedures or exhibit diverse productions, performances, or answers.

The focus of competency-based programs on complex tasks or situations has a profound impact on the way teachers conduct assessment. Competencies can hardly be treated separately from the learning context or from the situations which will be used to infer them. Consequently, complex assessment tasks or situations require some form of harmonization between instruction and assessments procedures used to support learning and certify competencies. They will also involve students to a greater extent. Here is a short description of the changes in assessment practices which occur at each step of the evaluation process of competencies.

1. *Planning.* Planning assessment of competencies requires the appropriate selection of a series of complex learning situations that will provide an adequate coverage of the program along a continuum of progression. The sampling and sequencing of situations must be suitable for gathering relevant information to allow teachers to make a decision on how to support or recognize a competency. Descriptive scales are used to inform teachers of what are the major stages of a student progression on a continuum. They also provide teachers with benchmarks on what are the key features and final outcomes which are the targets of a cycle of training (Ministère de l'éducation, 2006).
2. *Information gathering and interpretation.* To support and render account of competency development, teachers observe students' performance in a variety of situations, on more than one occasion and use diverse assessment instruments. Grids are often used to help teachers collect and code observations on students' performances or productions. They also assist teachers in interpreting their observations when such grids refer to evaluation criteria which are formulated in the curriculum. Such tools are provided to reduce teachers' subjectivity or bias in the case of summative evaluation or to regulate students' progress through appropriate feedback and intervention in the case of formative assessment. When such grids are adapted for students' use, they help teachers communicate the evaluation criteria. Students may thus be involved in multiple forms of assessment that assist them at all critical steps of their work and ultimately will contribute to set individual targets for personal improvement. Occasionally, gathering data through quizzes or tests will help find out whether students' difficulties results from a lack of skill or knowledge or from the incapacity to

mobilize such resources appropriately in a specific context. The same result may be accomplished by repeating observations in different contexts where the level of guidance is varied. The three-step assessment design developed by Rey *et al.* (2003) is a good illustration of an assessment practice where teachers may determine whether a student's difficulty derives from problems in (1) the mastery of basic procedures; (2) the capacity to select the appropriate one; or (3) the capacity to combine several of them in the appropriate context.

3. *Judgment and decision.* Teachers' professional judgment plays an important role in supporting students' learning and in making decisions concerning students' progressive acquisition of a competence. It is based on the information gathered on students' achievements and on various guidelines regarding what should be recorded and how it should be organized and reported. Scallon (2004) has shown that depending on the type of continuum used to collect data on students' achievement, teacher's judgment may take different forms. It could have a bearing on the student's ability to achieve increasingly complex tasks or on the capacity of the student to perform to standards with varying degrees of autonomy. Teachers' professional judgment is also required to weight the relative importance of each different competency in rating or marking students' work. Competencies may be considered of equal importance or not and compensation among competencies may be permitted or not. Policy directives assist teachers on such matters through the criteria and outcomes specified in the curriculum or through general guidelines indicating what values and principles should be taken into account (Ministère de l'éducation, 2003). To reduce personal subjectivity in teachers' interpretations, criteria and performance standards may be exemplified in a variety of government publications such as scales of competency levels, exemplars of students' performance, or generic evaluation grids. Mandatory or optional diagnostic assessment tools may also provide teachers with indications as to whether their own appreciation of students' work is properly aligned with the curriculum performance standards (IGEN, 2007; Rey *et al.*, 2003).

On certain aspects, classroom assessment of competencies still raises important questions on the notion of learning progression or developmental trajectory (Tardif, 2006). The implementation of cognitive models into developmental scales of progression requires deep knowledge of how a competency is learned. Other important questions also concern how teachers use their professional judgment to certify students through retrospective judgment or decision algorithms. Consequently, the way teachers use their professional judgment is the object of increased study.

The Formative Function of Classroom Assessment

The new programs of studies developed in the French countries all emphasize the major role of formative assessment in one form or another. Such an emphasis is not new since formative assessment practices have long been associated with values of differentiation of instruction and equal opportunities for learning for all students. The introduction of competency-based programs, however, has required a redesign of the instruments of formative assessment and the inclusion of additional methodologies to gather information on student learning. Such a renewal had already been prepared for years as several papers and books had already been published in French on formative assessment.

Following an extensive review of the French-language publications, Allal and Mottier Lopez (2005) have concluded that it has contributed to an enlargement of the conception of formative assessment. In such an enlarged perspective, highly influenced by constructivist theories of learning such as Piaget's and Vygotsky's, a much greater emphasis is placed on the regulation of cognitive and motivational processes occurring in learning. Still another direction of enlargement occurs when formative assessment is used to inform teacher planning of future instructional activities. It thus goes beyond remediation that directly benefits the students and includes regulation of teachers' future instructional activities to the point where decisions on instructional strategies and assessment are closely related.

The concepts of regulation and self-regulation have largely contributed to redefine formative assessment as an object of study in the French-language publications. External regulation (by the teacher, by the test, by remedial materials) is "redefined as scaffolding that assists students' development of self-regulation" (Allal and Mottier Lopez, 2005). Classroom assessment is thus considered as an instrument in developing students' autonomy in learning through procedures of self-assessment, reciprocal peer-assessment, and joint teacher–student assessment. New methods of student involvement in assessment are emphasized to such an extent as to include motivational regulations in addition to cognitive and metacognitive regulations.

The work of French researchers has led to an increased diversity in the ways of carrying out formative assessment. A lot has been accomplished at the theoretical level, but very little has been published on the verification of the impact of formative assessment on student learning. In the light of the new curricula based on competencies, the French-language work on formative assessment is in need of considerably more empirical research.

Classroom Assessment as Communication

Competency-based programs of studies have brought in several changes in classroom assessment that have

required innovative ways of reporting students' results. The performance on a complex task may require knowledge and skills derived from several subject matters such as mathematics, sciences, and writing as well as from more global competencies which are not linked to a single subject matter such as cross-curricular competencies in Quebec or the common base of knowledge and skills in France. This makes reporting students' progress through subject-matter content areas most difficult and somewhat artificial. Furthermore, students' learning progression involves both qualitative and quantitative changes which can hardly be summed up on a single scale.

Through a comparative study of different report-card systems in France, Quebec, and the French regions of Belgium and Switzerland, IGEN (2007) has listed a series of conditions for the effective use of report cards and communication of students' results and has made recommendations which extend well beyond the context prevailing in France:

1. Report cards should serve a clear purpose: either report on the best way to follow-up on a student's learning progression or to summarize student's achievement at certain period in time. Different report cards may have to be developed to serve distinct purposes.
2. Report cards do not need nor can they report exhaustively on the very fine details of a learning progression. To assess competencies require that a delicate balance be found between expected competencies and the whole set of resources associated with them. Decisions should be made as to what competencies and resources should be reported as priority and how they should be structured in the report card.
3. The same report card system cannot be used for both internal and external communication. The way information on students' achievement is conveyed depends on the intended recipient of the communication: parents or tutors, students, or other teachers or professionals.

Recent experiences with various forms of report cards have shown the difficulty of transmitting useful information on both qualitative and quantitative changes occurring in competency development to different stakeholders and for multiple purposes such as guidance, certification, and regulation of learning. Cross-sectional competencies have been found especially difficult to report on. They raise questions as to who should assess them and how, especially when such competencies may be acquired in disciplines taught by different teachers and validated in contexts which are not necessarily identical to the situations used to learn them. More development is required to build efficient report-card systems which articulate and coordinate several sources of information, at different times and for different purposes.

The Influence of Standards and External Testing

French countries have echoed the global world trend for a better monitoring and accountability of school systems. They have all developed some form of external testing to inform decision making at different levels of the school system. Such initiatives, however, are not translated in the same terminology as in the English publications. For instance, French countries which have tried to import the concept and word standards have been confronted with its multiple English definitions and interpretations. As Mons and Pons (2006) have shown, standards of evaluation policies are often shaped locally to take into account and weight differently, notions such as school efficiency, student learning progress, accountability, responsibility on one side, and the structural characteristics of the national systems such as the level of centralization, the multicultural nature of the country, the relationship between teachers' union and school administration, to name just a few.

A review of French sources by Mons and Pons (2006) shows that external tests may have both desirable and undesirable effects on classroom instruction and on classroom-assessment practices. To really have a positive impact, such external evaluations must be part of a *culture d'évaluation* (culture of evaluation) where decision making and management at all levels of the school system value the collection of data and its use to plan school improvement. Such a culture of evaluation provides teachers with formative opportunities to work collaboratively, reflect on the efficiency of their instructional practices, on their shared understanding of the curriculum, and on the best ways to use it to assess students' learning acquisitions (Demailly, 2001).

National Education Inspection–National Education and Research Administration Inspection (IGEN–IGAENR) (2005) have also observed that the introduction of new standardized tests in France have had an almost immediate impact on what is taught and assessed by teachers. The introduction of geometry as part of the grade 6 diagnostic test was sufficient to reactivate the teaching of this strand of mathematics which had been neglected for some time. The authors conclude that external forms of testing do not need to be related to high stakes to have an impact. They warn, however, that the introduction of new tests may have both, a dynamic effect, such as in the previous case, or a reductive effect when such tests validate a limited number of competencies mentioned in the program of studies.

Both the assessment of students' individual learning and teacher's instructional methods will benefit if assessment comes from a variety of sources and uses a diversity of strategies. External evaluations, however, should not be limited to multiple-choice tests or easily measurable learning acquisitions. There is a general apprehension as to whether such standardized tests can measure students' competencies as efficiently as with classroom-authentic

and complex assessment tasks. The development of large-scale standardized tests to assess competencies has presented test builders with important challenges such as duration of such tests, the authentic nature of the complex task or situation, and the capacity of students to mobilize extracurricular and out-of-school learning acquisitions.

Conclusion

The notion of competency has captured and crystallized a long-term interest relayed in French-language publications on the cognitive and developmental processes in student regulation of learning and on the role of formative assessment to involve both teachers and students in these processes. Such interests are inherited from the French constructivist models of learning and teaching and the notion of competency fits well within these models.

French-language publications on classroom assessment have contributed to show how it is closely interrelated with teaching and learning to the point of it being difficult to differentiate when assessment involves complex learning situations. This high degree of interrelationship is also evidenced by the domino effect which the introduction of competencies in the curriculum has had on all aspects of classroom-assessment practices, from planning assessment to the report of results.

Due to the high degree of interconnectedness of classroom assessment with the multiple facets of student learning, each country's policies have tried to maintain a certain coherence and alignment between the expectations of the program of studies and classroom-assessment practices. This has been translated with various degrees of success into materials linked to competency-based programs of studies. Such materials ranged from ready-to-use complex learning situations to diagnostic instruments of competency development. Descriptive scales and exemplars of student performances are also used to provide teachers with a better understanding of students' learning progression, a central concept in the assessment of competency development.

In actual practice, classroom assessment is much less coherent than what it is intended to be. A separate look at each country would reveal different kinds of problems in the implementation of competency-based assessment in the classroom depending on national priorities, the political and social situation of each country, and the mere resistance to change. External assessments have been introduced to monitor and regulate the effect of policies on the school system. The consequential validity and, especially, the side effects on classroom assessment are still being debated upon and are a matter of concern.

The difficulties in implementing appropriate assessment practices also result from a shortage of empirical research in the field. A few comparative studies among

French countries have started to emerge and it is expected that their results will increasingly contribute to inform policy decisions. In the meantime, French sources will continue to refer to English publications to complement the knowledge base on classroom-assessment practices. In the past, such combinations of sources have been very stimulating and led to original syntheses. They, however, require being appropriately adapted to each national context. Moreover, they make the case for an increased attention to empirical validation even more compelling.

See also: Alternative Assessment; Assessment and the Regulation of Learning; Classroom Assessment Tasks and Tests; Formative Assessment and Instructional Planning; Formative Assessment; Impact of Assessments on Classroom Practice; Peer and Self-assessment; Progression and Assessment: Developmental Assessment; Records of Achievement: Beyond Traditional Tests.

Bibliography

- Allal, L. and Mottier Lopez, L. (2005). Formative assessment of learning: A review of publications in French. *Formative Assessment: Improving Learning in Secondary Classrooms*, pp 241–264. Paris: Organisation for Economic Co-operation and Development.
- Demailly, L. (2001). Enjeux de l'évaluation et régulation des systèmes scolaires. In Demailly, L. (ed.) *Évaluer les politiques éducatives*, pp 13–30. Brussels: De Boeck.
- IGEN (Inspection générale de l'éducation nationale) (2007). Les livrets de compétences: nouveaux outils pour l'évaluation des acquis. *Rapport No. 2007-048*. Paris: IGEN.
- IGEN-IGAENR (Inspection générale de l'éducation nationale) (Inspection générale de l'administration de l'éducation nationale et de la recherche) (2005). *Les acquis des élèves, pierre de touche de la valeur de l'école?* Paris: IGEN-IGAENR.
- Jonnaert, P. and Masciotra, D. (2007). Socioconstructivisme et logique de compétences pour les programmes d'études. In Lafortune, L., Moussadak, E., and Jonnaert, P. (eds.) *Observer les réformes en éducation*, pp 53–75. Québec: Presses de l'Université du Québec.
- Laveault, D. (2006). Les échanges scientifiques de part et d'autre de l'Atlantique. Une étude des articles parus dans «Mesure et évaluation en éducation», la revue de l'ADMÉE. In Figari, G. and Mottier Lopez, L. (eds.) *Recherche sur l'évaluation en éducation*, pp 213–221. Paris: L'Harmattan.
- Le Boterf, G. (1994). *De la compétence: essai sur un attracteur étrange*. Paris: Éditions d'Organisation.
- Ministère de l'éducation (2003). *Politique d'évaluation des apprentissages*. Québec: Gouvernement du Québec.
- Ministère de l'éducation (2006). *L'évaluation des apprentissages au secondaire. Cadre de référence (version préliminaire)*. Québec: Gouvernement du Québec.
- Mons, N. and Pons, X. (2006). *Les standards en éducation dans le monde francophone. Une analyse comparative*. Neuchâtel: Institut de recherche et de documentation pédagogique.
- Rey, B., Carette, V., Defrance, A., and Kahn, S. (2003). *Les compétences à l'école*. Brussels: De Boeck.
- Roegiers, X. (2004). *L'école et l'évaluation. Des situations pour évaluer les compétences des élèves*. Brussels: De Boeck.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Montréal: Éditions du Renouveau Pédagogique.
- Tardif, J. (2006). *L'évaluation des compétences. Documenter le parcours de développement*. Montréal: Chenelière Éducation.

Further Reading

- Bonniol, J. -J. and Vial, M. (1997). *Les modèles de l'évaluation*. Brussels: De Boeck.
- Bressoux, P. and Pansu, P. (2003). *Quand les enseignants jugent leurs élèves*. Paris: Presses universitaires de France.
- Cardinet, J. and Laveault, D. (1996). Dix années de travaux européens et nord-américains sur l'évaluation : Quelles lignes de force? *Mesure et Évaluation en Éducation* **18**(3), 1–24.
- De Peretti, A. (1998). Encyclopédie de l'évaluation en formation et en éducation. Issy-les-Moulineaux: ESF.
- Figari, G. and Mottier-Lopez, L. (eds.) (2006). *Recherche sur l'évaluation en éducation*. Paris: L'Harmattan.
- Lafortune, L. and Allal, L. (2008). *Jugement professionnel en évaluation. Pratiques enseignantes à Québec et à Genève*. Québec: Presses de l'Université du Québec.
- Legendre, R. (2005). *Dictionnaire actuel de l'éducation*, 3rd edn. Montréal: Guérin.
- Perrenoud, P. (1998). *L'évaluation des élèves. De la fabrication de l'excellence à la régulation des apprentissages. Entre deux logiques*. Brussels: De Boeck.
- Roegiers, X. and De Ketele, J. -M. (2001). *Une pédagogie de l'intégration. Compétences et intégration des acquis dans l'enseignement*, 2nd edn. Brussels: De Boeck.
- Scallon, G. (2000). *L'évaluation formative*. Montréal: Éditions du Renouveau Pédagogique.
- Vial, M. (2001). *Se former pour évaluer. Se donner une problématique et élaborer des concepts*. Brussels: De Boeck.

Relevant Websites

- <http://eduscol.education.fr> – ÉduSCOL. Le site pédagogique du ministère de l'éducation nationale.
- <http://www.ides.ch> – Information and documentation center.
- <http://www.irdp.ch> – Institut de recherche et de documentation pédagogique.
- <http://iate.europa.eu> – InterActive Terminology for Europe (Translation of terms in education).
- <http://www.admee.ca> – L'Association pour le développement des méthodologies d'évaluation en éducation-Canada.
- <http://www.irdp.ch> – L'Association pour le développement des méthodologies d'évaluation en éducation-Europe.
- <http://enseignement.be> – L'enseignement en Communauté française de Belgique.
- <http://www.education.gouv.fr> – Ministère de l'éducation nationale.
- <http://educ-eval.education.fr> – Ministère de l'éducation nationale. Portail de l'évaluation.
- <http://www.mels.gouv.qc.ca> – Ministère de l'éducation du loisir et du sport.
- <http://www.restode.cfwb.be/index2.asp> – Serveur pédagogique de l'enseignement organisé par la Communauté française.
- <http://www.educa.ch> – The Swiss education server.

Classroom Assessment in Policy Context (Hong Kong)

D Carless, University of Hong Kong, Hong Kong, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Confucian-heritage culture – Those societies which subscribe to the tenets of Confucian thought.

School-based assessment – Assessments taken by students within the school, marked by their teachers and counting as part of the external examination score for the subject.

Washback – The phenomenon whereby due to the influence of a test, teachers, and/or students do things that they would not normally do.

The potential of classroom assessment to enhance student learning has, in recent years, attracted considerable attention in the main Anglophone countries. Learning-oriented assessment has been less established in jurisdictions that share a Confucian-heritage culture (China, Hong Kong, Japan, Korea, Singapore, and Taiwan) and are usually considered to be highly test-driven. Such examination-oriented systems may be due in part to high expectations of parents for the academic success of their offspring and the residual influence of the Chinese imperial examination system. The intense pressure and potentially negative social impacts of highly competitive examination-oriented systems in East Asia have been much discussed. Pong and Chow's (2002) portrayal of the pressures exerted on Hong Kong youngsters contains a distressing depiction of the risks of heaping examination stress on vulnerable teenagers. More positive images arise from the seminal work of Biggs (1996) who elaborates the paradox of the Chinese learner in that large class sizes, transmissive teaching, and test-dominated education can still lead to successful deep learning outcomes. A further positive performance indicator is that Hong Kong students generally perform well in international tests of achievement, such as TIMSS (Trends in International Mathematics and Science Study) and PISA (Programme for International Student Assessment).

Within such examination-oriented systems, the prospects for classroom assessment in East Asia might seem unpromising at first sight. A more positive outlook arises from Hong Kong's historical position as a combination of Eastern and Western influences and so there might be greater potential for formative approaches to assessment than in neighboring economic success stories, such as Japan or South Korea. There are grounds for believing that there is scope for classroom assessment to embed

itself in the Hong Kong context and a number of relevant policy initiatives have manifested themselves in recent years. The aim of this article is to review the prospects for classroom assessment in Hong Kong with developments divided into those in the primary sector (students aged 6–11 years old in years 1–6) and those in the secondary sector (students aged 12–18 in years 7–13). Each section includes both policy and implementation perspectives, mindful that there are tensions and contradictions between policy intentions and actions, as teachers re-interpret, subvert, or ignore policy prescriptions. This can particularly be the case when policy borrowing from overseas fails to account adequately for local contextual factors.

Assessment Developments in Primary Schools

Primary schooling in Hong Kong has traditionally been textbook dominated and test driven with school communities attaching great importance to indicators of academic achievement, such as textbook completion, homework, test scores, and report cards. Testing from an early stage is emphasized as preparation for future competition and selection through high-stakes examinations. From the age of 6 years, primary school students experience multiple tests and examinations in each school year. Within this test-dominated setting, two significant policy initiatives related to classroom assessment in the primary sector are: the target-oriented curriculum (TOC) in the period 1993–99 and the promotion of reforms related to learning to learn; and assessment for learning, from around 2000 until the time of writing. These two initiatives are discussed below.

Assessment Reform in the TOC

TOC was a radical and wide-ranging reform that sought to transform curriculum, pedagogy, and assessment through multiple components, including targets, tasks, and criterion-based assessment. It is probably fair to say that although there were pockets of encouraging developments, an overall judgment would indicate that TOC was overambitious in scope and failed to achieve the fundamental changes its adherents sought. Here the focus is only on the assessment aspects of the TOC reform. The TOC framework document, developed by Clark *et al.* (1994), sets out what they saw as the main problems with assessment in Hong Kong:

1. too much assessment involving gap-filling and multiple-choice questions;
2. a focus on the ranking and sorting aspects of assessment;
3. lack of integration of assessment with the curriculum; and
4. Failure to assess what is important in learning rather than what is easily assessable.

The proposed solutions tended to be more at the level of exhortations rather than concrete practical strategies and included pleas for: greater use of criterion-referenced assessment; more attention to formative assessment and student improvement over time; and the development of records of student achievement.

Evaluation reports indicated that the assessment aspects of TOC were the most difficult to reform and proved most problematic for teachers. Morris *et al.* (2000) reported resistance to TOC-assessment processes in that teachers held a set of interlinked beliefs, which involved a preference for objective, reliable, formal assessments, and doubts about the feasibility of formative assessment. Teachers did not see assessment as something which would involve their professional judgement and had a reluctance to assess through any means which might be regarded as nonobjective. Formative assessment in TOC became associated with the recording of data about learners, with teachers not having the time, skills, or support necessary to feed back this data into the classroom. Teachers reported doing extra work in collecting assessment data and became frustrated as they realized this was not contributing to enhanced pupil learning. The combination of increased paperwork through record-keeping and a system which was both misunderstood and not congruent with teacher beliefs resulted in formative assessment being considered by teachers as the most problematic and unpopular aspect of the TOC reform.

TOC classroom assessment was further hindered in view of the failure to reform the complex secondary school-selection processes at the end of primary schooling, whereby internal school tests in years 5 and 6 are scaled against a standardized test. It became apparent that neither the system, nor educational stakeholders were ready for the kind of assessment system envisaged by TOC. As a central plank of government educational policy, TOC was beginning its decline by the time of the retrocession to mainland China in 1997 and began fading out as the incoming postcolonial government sought to make its mark. Challenges in the implementation of formative classroom assessment were one of the factors in the eventual abandonment of the TOC reform.

Learning to Learn, Assessment for Learning and Basic Competency Assessments

It was acknowledged that TOC had done little to modify the prevailing examination-oriented culture. A number of TOC principles or concepts were however, repackaged as a new curriculum reform was launched at the turn of the

millennium under the theme of learning to learn. Following contemporary terminology in the UK and elsewhere, the nomenclature formative assessment was changed to assessment for learning. The Education Bureau made a number of recommendations for the development of assessment for learning, including the following:

1. the development of more diversified modes of assessment and a reduction in tests and examinations;
2. opportunities to do assessment collaboratively with students or encourage students to carry out peer or self-assessment;
3. sharing with students the goals of learning, so that they can recognize the standards they are aiming for; and
4. the use of assessments which probe higher-order thinking skills, creativity, and understanding rather than rote memorization of facts.

Within this policy context, Carless (2005) discusses two case studies of progressive teachers implementing assessment for learning, mindful that scaling up of these experiences is likely to be difficult. He proposes a multi-level model of factors impacting on the implementation of assessment for learning in Hong Kong, comprising three components: teacher factors, including attitudes toward and understanding of assessment for learning; microlevel school factors, including internal resources, support from management, and collaboration with colleagues; and macrolevel societal factors, including the wider reform climate, educational culture, impact of relevant governmental agencies, and washback from high-stakes testing.

A further subcomponent of the learning-to-learn reforms relates to the introduction of basic competency assessments for the three main subjects of mathematics, Chinese, and English. Basic competencies are the essential knowledge and skills or basic standards required by students in relation to the learning targets and objectives set out in the curriculum for each key stage. Progress toward the basic competencies is monitored through a territory-wide system assessment (TSA) administered by the government at primary three (first conducted in mid-2004) and primary six (first conducted mid-2005), comprising paper-and-pen mode with an oral-assessment component for the two languages. Its stated purposes are to provide feedback to schools about their standards in the three key subjects so that schools can draw up plans to increase effectiveness in learning and teaching.

The government claims that the TSA is low stakes in nature and this is true for students because no individual student grades are made available. For schools and teachers however, it is generally interpreted as high stakes because it indicates how well schools are performing. In a context with a low birth rate, competition for students and associated threats of school closures, the TSA exerts pressure on teachers, which is then transmitted to students.

Table 1 Percentage of students achieving basic competency levels

<i>Subject and year level</i>		<i>Percent achieving basic competency</i>			
		<i>2004</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>
Chinese language (listening, reading, and writing)	Yr 3	82.7	84.7	85.2	84.9
	Yr 6	–	75.8	76.5	76.7
	Yr 9	–	–	75.6	76.2
English language (listening, reading, and writing)	Yr 3	75.9	78.8	79.4	79.5
	Yr 6	–	70.5	71.3	71.3
	Yr 9	–	–	68.6	69.2
Mathematics	Yr 3	84.9	86.8	86.9	86.9
	Yr 6	–	83.0	83.8	83.8
	Yr 9	–	–	78.4	79.9

Adapted from government press release: <http://www.gov.hk>.

The government documents tend to reinforce this by stating that monitoring of the extent to which individual schools attain the basic standards in key learning areas is premised on the need for accountability. The TSA seeks to incorporate both summative (overall assessment data about school performance) and formative (information which can feed back to teaching) functions. The reality, however, is a focus on statistics when results are published. **Table 1**, for example, illustrates the percentage of students who have reached the basic competency standard. It indicates an initial improvement, perhaps as teachers and students become familiar with the test format, and then evidence of tapering off in 2007.

In terms of the formative function of TSA, schools are provided with data which summarize the performance of their students on different test items. Teachers find it difficult to channel this raw feedback into classroom practice. A consequence of the emphasis on statistics and the difficulty of making formative use of the data is to reinforce the prevalent notion in Hong Kong that almost all tests are perceived by stakeholders as high-stakes and formative assessment tends to be drowned by the power of summative assessment. The use of standardized assessment data to inform teaching and learning still needs further development if its formative potential is to be realized.

Assessment Developments in Secondary Schools

Secondary schools in Hong Kong tend to be even more examination oriented than primary schools, as students work their way toward the key selective public examinations at the end of years 11 and 13. Years 10 and 11 are often characterized by drilling, rote learning, examination tips, and repeated study of past papers or examination-practice

exercises. A narrowing of the curriculum results from the tendency to disregard anything that is not directly examination related. Given the importance of the major public examinations, TSA (at secondary 3, year 9) seems less influential than in primary schools.

Unlike in the primary sector, where initiatives such as TOC or assessment for learning can be adopted, tinkered with, or ignored, changes to high-stakes examinations in the secondary sector have an immediate impact on classroom practice, particularly as teacher effectiveness is principally judged by examination results. In line with the understanding that one of the principal influences on what goes on in the classroom is the nature of high-stakes tests, most examination changes in Hong Kong during the last 20 years have had the aim of promoting positive washback on classroom pedagogy. Washback can be defined as the phenomenon whereby because of a test, teachers and students do things that they would not normally do. Washback studies have been a particular feature of assessment research in Hong Kong, particularly with respect to English language, one of the key subjects in the school curriculum. Cheng (2005), for example, studied the impact of changes to the 1996 English language Hong Kong Certificate of Education Examination (HKCEE) taken at the end of year 11. These changes were premised on the aim of improving students' English communication skills and encouraging a more communicative orientation to language teaching. Cheng argues that washback works quickly and efficiently in bringing about changes in teaching materials, which is due largely to the highly adaptable and commercial nature of Hong Kong society, and slowly and with difficulties in terms of the methods teachers employ.

Two current developments should also be noted. First, the Hong Kong Examinations and Assessment Authority HKEAA (2006) with effect from 2007 is moving from norm-referenced to standards-referenced reporting with grades of A–F being replaced by levels 1–5, with the top-scoring students within level 5 being awarded a 5*. Under the standards-referenced approach, the standards are held constant and the percentage of students being awarded a given level varies according to the proportion of students meeting the defined standard. Key aims are to allow students to gain recognition for what they know and can do, and reduce the number of students who are currently awarded unclassified grades and hence no recognition of their learning. It is also hoped that transparent standards with associated criteria and exemplars can be useful in illustrating to students targets and areas for improvement. This aspect is congruent with the role of learning intentions and explicit criteria in classroom assessment.

Second, Hong Kong is moving from 7 to 6 years of secondary schooling with 4-year undergraduate degrees to align itself with mainland China and North America. As part of a new senior secondary (NSS) curriculum,

instead of two high-stakes external public examinations, – HKCEE at the end of year 11 and Hong Kong advanced level examination at the end of year 13 – from 2012 onward there will be one qualification, a new Hong Kong diploma of secondary education (HKDSE) at the end of year 12. It is hoped that moving from two examinations to one will reduce the amount of time preparing for and taking examinations and thereby increase the time available for student learning. The overall aim of the assessment change is to assess a wide range of outcomes relevant to the aptitudes, needs, and abilities of the entire cohort of students eligible for the HKDSE. Each student would also have a learning profile to record their main achievements during the process of senior secondary school. One school of thought is that the HKDSE may reduce pressure on students as there is only one instead of two public examinations for secondary school students. A contrary viewpoint is that having only one high-stakes examination process actually intensifies the pressure on students. A further feature of the HKDSE is that a proportion of marks is being awarded for school-based assessment in all subjects, thereby bringing classroom assessment a higher profile. This development is discussed below.

School-Based Assessment

The most significant and longstanding classroom assessment initiative in the secondary section is the promotion and development of school-based assessment (SBA). SBA in the Hong Kong context refers to in-school assessments graded by teachers and contributing around 15–25% of the marks to the HKCEE or HKDSE. The earliest manifestations of SBA date from a teacher assessment scheme (TAS) for science subjects, initiated from the late 1970s onward. Motivations for SBA include: the potential for assessing a wider range of achievements than through examinations; involving teachers and students more actively in the processes of external examinations; facilitating student improvement over time; and not determining students' academic standard by a single examination. As indicated above, this and other assessment innovations have been generally motivated by the need to temper the negative impact of examinations on classroom practice.

Benny Yung Hin-Wai's research into practical work in science for the TAS has developed particularly useful insights into the challenges of combining both summative and formative functions of assessment. In TAS/SBA, teachers need to award a grade for certification purpose and also provide feedback to students for ongoing learning purposes. This represents a tension which Yung (2001) analyses using the metaphors of policeman and companion to illustrate Hong Kong teachers' conceptions of assessment. The former saw the teacher adopting a picky and fault-finding attitude as represented in the following quotation, "I had to behave like a policeman who had to grasp every chance to give out the assigned

quota of illegal parking tickets in order not to be scolded by his superiors" (Yung, 2001: 254). The latter was represented by a teacher who upheld his beliefs that the teacher's role was to assist students' learning and that the process of awarding an examination grade should be of subsidiary importance to that goal. As Yung acknowledges, the teachers in his study did experience challenges and confusion concerning formative and summative uses of the same set of evidence; illustrative of challenges also found elsewhere in reconciling formative and summative functions of assessment.

Yung also discusses the crucial issue of fairness and describes three views of fairness exhibited by the case-study teachers in his research: fair in the sense of assessing students reliably; fair in terms of not jeopardizing students' chances of learning the subject matter; and fair in the sense of not depriving students of opportunities to receive a well-rounded education. Yung concludes by suggesting that the educational benefits derived from the TAS are at a cost to reliability. The evocative subtitle of Yung's (2006) recent book *Fairness and Fear* reinforces these issues and the pressures of assessment for both teachers and students.

SBA is currently being expanded to all subjects of the curriculum within various timeframes with statistical moderation being used to enhance the reliability of teacher grading. The subject of English has been a particularly high-profile example of SBA. Early progress indicates some positive impacts on students in terms of engaging with assessment criteria and having the explicit opportunity to improve over time, balanced by concerns about increased workload in preparation for SBA. Teachers' responses vary and concerns are mainly focused on the pressures of being responsible for part of the public-examination score; the issue of fairness and interrater reliability; and the heavy workload accruing from SBA preparation, scoring, moderation, and attendance at training programs.

The extent to which SBA can operate fruitfully in a Confucian-heritage culture setting is still open to question. There remain a number of societal barriers to its successful implementation: a preoccupation with reliability and fairness at the expense of validity; lack of understanding, and few well-publicized local models of formative assessment; and a societal perception that teachers cannot be trusted to grade their own students fairly. The following comment of Pong and Chow may still carry some validity, "the highly selective nature of examinations has forced examiners to put fairness and objectivity of marking above all other concerns" (Pong and Chow, 2002: 143).

Summary

This article has focused on assessment developments in Hong Kong primary and secondary schools. Hong Kong

represents an interesting case in which global trends (assessment for learning; learning to learn; school-based assessment) meet a traditional competitive test-driven culture dating back to the Han dynasty. In primary schools, despite some pockets of good practice, classroom assessment is still struggling to become firmly embedded. In secondary schools, SBA has been and will remain to be a crucial topic and one which is likely to generate much discussion or controversies. The new HKDSE and standards-referenced reporting are also issues which are likely to be extensively debated in Hong Kong in the years to come. The high demands for education from parents, the public, and employers, allied with perennial concerns about perceived falling standards continue to put pressure on government, schools, teachers, and students.

The prospects for classroom assessment to become a positive force in Hong Kong remain unclear. In an examination-oriented system, a contextually suitable version of classroom assessment would need to acknowledge the power and dominance of summative assessment. Developing formative learning power from in-school and external tests might be a useful way of developing productive synergies between summative and formative functions of assessment. SBA also has the potential to develop a stronger interface between examinations and productive student learning but still faces the barriers of societal values and teacher resistance. For these potentials to be realized, there remains a pressing need for the enhancement of assessment literacy through teacher development and an associated program of collaborative classroom-based research.

See also: Challenges of Developing and Implementing Formative Assessment Practices in Schools; Summative Assessment by Teachers.

Bibliography

- Biggs, J. (1996). Western misperceptions of the Confucian-heritage learning culture. In Watkins, D. and Biggs, J. (eds.) *The Chinese Learner: Cultural, Psychological and Contextual Influences*, pp 45–67. Hong Kong: Comparative Education Research Centre and Australian Council for Educational Research.
- Carless, D. (2005). Prospects for the implementation of assessment for learning. *Assessment in Education* **12**(1), 39–54.
- Cheng, L. Y. (2005). *Changing Language Teaching through Language Testing: A Washback Study*. Cambridge: Cambridge University Press.
- Clark, J., Scarino, A., and Brownell, J. (1994). *Improving the Quality of Learning: A Framework for Target-Oriented Curriculum Renewal*. Hong Kong: Institute of Language in Education.
- HKCAA (2006). School-based assessment: A brief introduction and a response to major concerns. <http://hkeaa.edu.hk> (accessed June 2009).
- Morris, P., Lo, M. L., and Adamson, B. (2000). Improving schools in Hong Kong: Lessons from the past. In Adamson, B., Kwan, T., and Chan, K. K. (eds.) *Changing the Curriculum: The Impact of Reform on Hong Kong's Primary Schools*, pp 245–262. Hong Kong: Hong Kong University Press.
- Pong, W. Y. and Chow, J. C. S. (2002). On the pedagogy of examinations in Hong Kong. *Teaching and Teacher Education* **18**(2), 139–149.
- Yung, B. H. W. (2001). Examiner, policeman or students' companion: Teachers' perceptions of their role in an assessment reform. *Educational Review* **53**(3), 251–260.
- Yung, B. H. W. (2006). *Assessment Reform in Science: Fairness and Fear*. Dordrecht: Springer.
- Adamson, B., Kwan, T., and Chan, K. K. (eds.) (2000). *Changing the Curriculum: The Impact of Reform on Hong Kong's Primary Schools*. Hong Kong: Hong Kong University Press.
- Biggs, J. (ed.) (1996). *Testing: To Educate or Select? Education in Hong Kong at the Crossroads*. Hong Kong: Hong Kong Educational.
- Carless, D., Joughin, G., and Liu, N. F., and Associates (2006). *How Assessment Supports Learning: Learning-Oriented Assessment in Action*. Hong Kong: Hong Kong University Press.
- Curriculum Development Council (2001). *Learning to Learn: The Way Forward in Curriculum Development*. Hong Kong: Government Printer.
- Curriculum Development Council (2006). Basic Competency Assessments. <http://www.edb.gov.hk/> (accessed June 2009).
- Davison, C. (2007). Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly* **4**(1), 37–68.
- Kennedy, K. (2005). *Changing Schools for Changing Times: New Directions for the School Curriculum in Hong Kong*. Hong Kong: Chinese University Press.
- Morris, P. (1995). *The Hong Kong School Curriculum*. Hong Kong: Hong Kong University Press.
- Morris, P. (2002). Promoting curriculum reforms in the context of political transition: An analysis of Hong Kong's experience. *Journal of Education Policy* **17**(1), 13–28.
- Stimpson, P., Morris, P., Fung, Y., and Carr, R. (eds.) (2003). *Curriculum, Learning and Assessment: The Hong Kong Experience*. Hong Kong: Open University of Hong Kong Press.
- Watkins, D. and Biggs, J. (eds.) (1996). *The Chinese Learner: Cultural, Psychological and Contextual Influences*. Hong Kong: Comparative Education Research Centre and Australian Council for Educational Research.
- Watkins, D. and Biggs, J. (eds.) (2001). *Teaching the Chinese Learner: Psychological and Pedagogical Perspectives*. Hong Kong: Comparative Education Research Centre and Australian Council for Educational Research.

Further Reading

Relevant Websites

- <http://www.edb.gov.hk> – Education Bureau, The Government of the Hong Kong Special Administrative Region, people's Republic Of China.
- <http://hkeaa.edu.hk> – Hong Kong Examinations Authority.

Classroom Assessment in Policy Context (New Zealand)

T J Crooks, University of Otago, Dunedin, New Zealand

© 2010 Elsevier Ltd. All rights reserved.

New Zealand has an area similar to that of Japan or the British Isles, but a population of about 4.3 million people. About 70% are of European origin, predominantly from the British Isles. The indigenous Maori population makes up about 15% of the balance, with substantial percentages also from other Pacific Islands and Asian countries.

School attendance is compulsory in New Zealand from age 6 to 16. However, it is normal to enrol children in early childhood care and education well before their fifth birthday, to enter them into the school system on or soon after their fifth birthday, and to continue their formal education past their 16th birthday. A majority proceed into some form of tertiary education.

Primary schooling (years 1–8) usually starts at age 5 and lasts 7.5–8.5 years. Secondary schooling normally lasts for up to 5 additional years (years 9–13). More than 80% of students remain until year 12, and almost 60% remain until year 13.

The most common patterns for schooling, in decreasing order of student numbers, are:

- approximately 6 years in primary school, 2 years in intermediate school, and up to 5 years in secondary school;
- approximately 8 years in primary school and up to 5 years in secondary school; and
- approximately 6 years in primary school and up to 7 years in a secondary school that includes years 7 and 8, as well as years 9–13.

The focus here is on classroom assessment in New Zealand schools. Crooks (2002) has written an overlapping but more wide-ranging review of assessment in the New Zealand school system, while Carr (2001) has given an extensive description of recent developments in assessment practices for early childhood education in New Zealand.

School Curriculum Policy

Prior to 1990, school curricula evolved slowly, with leadership from experienced specialists in the national Department of Education and extensive involvement of teachers in developing successive drafts and trying them with students. The syllabi that resulted from this process were not highly prescriptive, but commonly were accompanied by custom-developed resources that teachers could use. In 1991, a new government and minister of education introduced sweeping changes modeled on the

curriculum and assessment changes made in the late 1980s in England and Wales. These changes have substantially influenced classroom assessment practice.

Working groups of curriculum experts and teachers identified several strands for each curriculum, and within each strand a substantial number of achievement objectives. The objectives were then placed into eight levels, representing the planned progress of students over their years of primary and secondary education. The first five levels were spaced approximately 2 years apart, with the last three levels each intended to represent 1 year of normal progress. The details for each curriculum area were described in separate volumes: one each for mathematics, science, English, technology, social studies, health and physical, education, and the arts. Special versions of each of these were developed for students learning in Maori-language immersion programs, and additional volumes were prepared for the teaching of languages other than English.

These curriculum changes saw a considerable tightening of curriculum and assessment expectations for schools and teachers. Pressure to implement the fairly detailed structures of the new curricula – with their strands, levels, and achievement objectives – largely has come through the reviews of schools conducted by the Education Review Office (a school inspection agency), supplemented by nationally funded but regionally implemented professional development contracts focusing on each new curriculum document. Many primary school teachers were quite overwhelmed trying to cover the full range of achievement objectives, and, toward the end of the 1990s, changes to the national education guidelines for schools offered greater flexibility in curriculum implementation, particularly for the first 4 years of school, where renewed priority was given to literacy and numeracy.

There has been significant debate about the merits of these curriculum changes. On the one hand, they were seen as helping teachers by offering them more structured guidance for their teaching and increasing continuity of learning programs for students moving between schools. On the other hand, many teachers felt that they and their students were on a treadmill of curriculum requirements, that the large list of achievement objectives encouraged shallow rather than deep learning, and that the loss of the trust and creative freedom that teachers previously enjoyed undermined their sense of professionalism.

Late in 2007, after a 4-year gestation, a revised, integrated curriculum document was released. This maintained the

general structure of the previous curriculum, with its strands, levels, and achievement objectives, but considerably streamlined it into a single volume with fewer and broader objectives. Schools are encouraged to develop their own curriculum plans within the broad national framework.

School Assessment Policy

In New Zealand, assessment of individual students in schools predominantly has been viewed as a responsibility of their teachers. Where standardized tests have been used in schools, it usually has been in a low-stakes manner, offering teachers additional information on which to make their judgments.

The main exception to this has been assessment for national qualifications in the final 3 years of secondary education. These use a mixture of national examinations and teacher assessments, with inter-teacher moderation procedures for the latter that are intended to achieve greater national consistency of standards.

Until the late 1980s, formal assessment and reporting in New Zealand schools were predominantly norm referenced, emphasizing the relative performance of different students. There is now widespread acceptance of the arguments for criterion-referencing or standards-based assessment (as it is commonly called in New Zealand), which is now integral to the national curriculum and qualifications. However, many teachers who have found the arguments for standards-based assessment convincing have struggled to develop practices consistent with the goal.

The basic structure for assessment of students in New Zealand schools from 1993 to 2007 was set out in *The New Zealand Curriculum Framework* (Ministry of Education, 1993). It stated that:

The primary purpose of school-based assessment is to improve students' learning and the quality of learning programmes. Other purposes of assessment include providing feedback to parents and students, awarding qualifications at the senior secondary school level, and monitoring overall national educational standards. Assessment also identifies learning needs so that resources can be effectively targeted (p. 24).

The emphasis on the use of assessment to improve learning and teaching has been maintained and strengthened in the 2007 version of *The New Zealand Curriculum* (Ministry of Education, 2007):

The primary purpose of assessment is to improve students' learning and teachers' teaching as both student and teacher respond to the information that it provides. With this in mind, schools need to consider how they will

gather, analyse and use assessment information so that it is effective in meeting this purpose.

Assessment for the purpose of improving student learning is best understood as an ongoing process that arises out of the interaction between teaching and learning. It involves the focused and timely gathering, analysis, interpretation and use of information that can provide evidence of student progress. Much of this evidence is "of the moment". Analysis and interpretation often take place in the mind of the teacher, who then uses the insights gained to shape their actions as they continue to work with their students. (p. 39).

In describing the key features of good assessment, the new document states that effective assessment benefits students, involves students, supports teaching and learning goals, is planned and communicated, is suited to the purpose, and is valid and fair (p. 40).

Classroom Assessment of Students in Primary Schools

For the purposes of this section, the term primary schools includes schools specializing in the education of early adolescents (intermediate and middle schools).

Most of the assessment in New Zealand primary schools can be described as assessment for learning (formative assessment). This is focused on enhancing student development, and often involves relatively unstructured interaction between student and student or teacher and student rather than planned formal assessment events. A few well-timed words, an insight shared, or shared enjoyment of notable achievement may have a lasting beneficial impact. Much of the assessment in New Zealand primary schools is of this nature, because teachers spend a lot of their teaching hours walking around their classes and monitoring and guiding the work of their students. More formal assessment of work done in class or for homework also focuses mainly on providing feedback rather than on accumulating marks or grades.

Assessment for learning is thus an integral part of teaching and learning, and appears to have more affinity to good teaching practice than to the usual prescriptions for good practice in assessment of learning (summative assessment). The goal of the latter is to provide a trustworthy, clear, and up-to-date picture of a student's current capabilities or attitudes, progress over time, or further growth needs and potential. Summative assessment in New Zealand primary education is evident mainly in reports to parents about their children's progress and in school-wide monitoring of student achievement by principals and boards of trustees.

When the curriculum documents described earlier were released during the 1990s, each listing a large number of

achievement objectives classified into levels, the Ministry of Education and the Education Review Office encouraging the use of these objectives and levels in teachers' assessments of students. A common early response from teachers and schools was to develop detailed lists of the achievement objectives, tabulate these alongside the names of all children in a class, and record in the resulting table whether or not each child appeared to have met the objectives.

Several things soon became apparent to teachers and researchers:

- Because of the large number of objectives, a teacher's judgment for one objective might be made on the basis of just one task, yet many tasks could be developed for each objective and students could be expected to perform differently on different tasks.
- Children who could do a particular task on one day often could not do that task or a very similar one the next day.
- Trying to complete this process for all of the achievement objectives in the primary school curriculum for a particular class was overwhelmingly time consuming, and threatened the quality of teaching by fragmenting the teacher's educational focus.
- There were major difficulties in summarizing student performance by aggregating across achievement objectives in a curriculum strand or whole curriculum area, because student performance varied markedly across objectives and strands.
- Different teachers applied quite different standards in judging whether or not an objective had been met or a level achieved.
- The gap between adjacent levels (2 years of normal progress) was too large to give a satisfying sense of progress.

There have been varied responses to these dilemmas from teachers, schools, and professional development consultants. Some tried to make assessment decisions more straightforward by subdividing each objective into a number of more specific outcomes, making their assessment easier but exacerbating the other difficulties by, in effect, increasing the number of objectives. Others subdivided each level into finer divisions, making evidence of progress more attainable but not resolving the other issues. Still others concluded that while the sequences of achievement objectives offered useful guidance for program planning, the levels should not play a significant role in assessment and reporting. The reduced number of objectives in the 2007 version of the curriculum and the greater degree of choice that it offers to schools appear likely to reduce the emphasis on curriculum levels in teacher assessments and reporting.

New Zealand has not used achievement tests nationally for whole cohorts of primary school students. Reports

to parents have primarily focused on teacher judgments of what students do well, how much progress they are making, and what they most need to work on. This recognizes the extensive knowledge of students that primary teachers develop from more than 700 h of observations during a school year. However, there has been concern from some parents, education officials, and politicians that reports to parents on their children's progress have tended to accentuate positive aspects (areas of good performance and good progress) and minimize comment on less positive aspects. This has been used as an argument for introducing national testing, with proponents suggesting that national test results would give parents a true picture of their children's progress within their national age cohort. To date, however, national testing has been rejected on the grounds that it could be expected to do more harm than good. While teacher judgments remain central in reporting student progress, there is increasing support for using professionally developed assessment resources (such as nationally standardized tests) to further inform their judgments.

Compared to many other countries, New Zealand has not made heavy use of standardized tests in its primary schools. Where they have been used, the results often have not been passed on directly to parents, but used as additional evidence for teacher decision making.

The most pervasive standardized technique has not been a test, but rather a means of monitoring the progress of student's oral reading skills: the running record (Clay, 1993). This technique is very widely used in the early primary school years, and to a lesser extent in the later primary years. At about age 6, a year after most children start school, many students have been assessed for their reading development through the use of the Observational Survey (Clay, 1993). In many schools, this has helped to identify the weakest readers for personalized reading recovery tuition.

At more senior levels (mainly year 4 upwards), many schools since the early 1970s have made annual use of the Progressive Achievement Tests (covering listening comprehension, reading vocabulary, reading comprehension, mathematics, and study skills). Although this testing is a well-established practice, schools have varied widely in how much use has been made of the results. In many cases, the scores have had little impact on teacher decision making and reporting. New, more sophisticated versions of these tests are now being released. These use item response modeling and provide a range of reports, at individual student and class level, designed to give teachers and principals more insight into the performance and learning needs of students.

Since 2000, the government has been funding the development of Assessment Tools for Teaching and Learning (asTTle), designed for assessing reading, writing, and numeracy in both English and Maori, predominantly with

students between years 4 and 11. These are currently available in CD-ROM format, but an Internet-based version is currently being developed. They are based on banks of nationally normed items, calibrated through item response modeling, together with software to assemble tests and print reports in several easily understood graphic formats. The reports are intended to identify strengths and learning needs for individuals and groups of students, and to compare the performances of individuals and groups to selected norms. Through task-selection criteria, tests can be tailored to the preferences of particular schools or teachers, for instance, by selecting particular curriculum strands to focus on and emphasizing easier or harder items. Results can be viewed by curriculum level and compared against a variety of norms, such as students attending similar schools. Ideas for the next instructional steps based on the patterns of responses are available online.

An ongoing project that commenced in 1993 has involved the development of Assessment Resource Banks. These now contain about 3600 assessment resources (tasks) for mathematics, science, and English, predominantly suited to students in years 4–10. Many of the tasks use multiple-choice or short-answer formats, with smaller numbers involving extended written answers or practical science activities. Marking guides are included. The resources have been normed on nationally representative samples and are coded by curriculum strand, levels, response format, and other features such as task difficulty. They can be accessed and searched through the Internet.

Between 2000 and 2004, the Ministry of Education funded the development of exemplars of student performance in many areas of the national curriculum, from early childhood education through to lower secondary school. These are nationally moderated, annotated examples of student work that are designed to help answer the question “What do we mean by quality work?” They aim to assist teachers to choose appropriate criteria for evaluating student work, understand which aspects are of particular importance, and, therefore, focus ongoing teaching and guidance more effectively.

In addition to funding the development of these assessment resources to help classroom teachers make better-informed judgments about student achievement, the Ministry of Education has also devoted substantial resources to assessment-focused professional development programs for teachers. Since 1995, school-based programs on ‘assessment for better learning’ and, more recently, ‘assess to learn’ have been offered throughout the country. The main focus of these has been to further develop the formative assessment practices of individual teachers, but more recently there has also been emphasis on the use of assessment data collaboratively by clusters of teachers or by all staff at a school. The latter approach is supported by some research evidence, suggesting that

individual teachers can best be helped to deal with teaching and learning issues in their individual classrooms by engaging the collective skills, ideas, and insights of a group of teachers to review the available assessment information and discuss desirable next steps.

In summary, assessment in New Zealand primary schools is predominantly low-stakes assessment focused on monitoring students’ learning and improving learning through direct feedback to students or adjustments to teaching programs. Written or oral reports to parents can be seen as complementing the formative role by giving guidance to parents and students, while also having a comparatively low-stakes summative role. A substantial range of standardized tests and resources is now available, but New Zealand does not have whole-cohort testing in its primary schools.

Classroom Assessment in New Zealand Secondary Schools

While some secondary schools begin at year 7, a majority have students in years 9–13. Most of the students in years 11–13 are seeking to earn credits toward national qualifications, and the high focus on the assessments for these qualifications has a major influence on assessment practices throughout secondary schools. In this section, the general character of assessment in years 9 and 10 is described briefly and subsequently assessment for qualifications in years 11–13 is discussed more fully.

The most notable differences between assessment in primary schools and assessment in secondary schools involve the timing and forms of feedback and the use of marks or grades. In primary schools, and particularly the earlier years, the majority of teacher feedback is given while work is in progress and work is not generally returned with marks or grades on it. Teachers keep records of their observations and judgments about some of the work done by the student, but these are not usually numerical, and they are not mathematically aggregated to produce end-of-term or end-of-year marks or grades. By years 9 and 10 in secondary school, students usually have different teachers for each subject. Much of the feedback that students receive comes on completed work, and it usually includes formal marks or grades. Significant pieces of work are marked and these marks accumulated in a grade book, so that they can be aggregated to produce a cumulative mark or grade.

For much of the last century, secondary schools produced written reports to parents that give, for each subject, a cumulative percentage mark, a grade and/or place in class, and a fairly brief written comment. About 20 years ago, national efforts were made to introduce a form of standards-based assessment into secondary

schools. Several (usually 5–7) important learning goals were identified for each subject and year level, and for each of those goals a descriptive scale with about five levels was developed to describe levels of development or performance. For instance, in year 12 physical education, one of the goals was knowledge of exercise physiology. The performance levels associated with this goal were that the student:

1. can identify the physiological body systems;
2. has a basic knowledge of physiological body systems;
3. has a basic knowledge of physiological changes in body systems in response to exercise and knows how to apply this;
4. has a broad knowledge of physiological changes in body systems in response to exercise and applies this to familiar activities; and
5. demonstrates a clear understanding of the physiological changes in body systems in response to exercise and applies this in new and different activities.

Many secondary schools have adopted variants of this approach for use in years 9 and 10. Within this approach, each assessment task involves one or more of the learning goals, and students receive feedback in terms of the levels of those goals. Many teachers find that this approach helps them to be precise in setting tasks for students and allows students to become more skilled at assessing their own work against the goals and performance levels. At the end of each school term or year, parents receive reports in which a level has been assigned to each goal, a generic description of that level is provided, and often a more specific or personal comment added.

For more than 50 years, students in years 11–13 (the final 3 years of secondary school) have attempted national qualifications through national end-of-year exams, moderated school-based assessments, or a combination of both. The precise form of these assessments has changed several times over the 50 years, while the percentage of students remaining in school until years 12 and 13 has risen dramatically in the last 20 years. A brief account of those changes is given in Crooks (2002).

The current form of qualifications for years 11–13 is a series of National Certificates of Educational Achievement (NCEAs). Students can attain these certificates by gaining credit on Achievement Standards or Unit Standards. Most established secondary school subjects are subdivided into several (usually five to eight) achievement standards, each of which has three passing levels (achieved, merit, and excellence) and a specified number of credits toward an NCEA for those who pass the standard. In addition, there are unit standards, which are assessed on a pass/fail basis, each with a specified credit value. Unit standards are most commonly used for more vocational subjects, but are also available as alternatives to

achievement standards in many secondary school subjects. Schools vary in the extent to which they allow or encourage use of unit standards rather than achievement standards. A recent government decision that NCEAs can be awarded with overall merit or excellence if 50 or more of the required 80 credits are gained at the specified level appears likely to encourage more use of achievement standards rather than unit standards in subjects where both types are available.

About half of the achievement standards are assessed through national examinations held late in the school year (external assessment). The remaining achievement standards and all unit standards are awarded through teacher assessments conducted within schools (internal assessment). These two alternatives have clear implications for classroom assessment practice. For externally assessed standards, classroom assessment is largely assessment for learning, with teachers trying to help their students prepare well for the national examinations. For internally assessed standards, classroom assessment is a complex mixture of assessment for learning and high-stakes summative assessment. Teachers need to walk a delicate line between their mentor and assessor roles. In most cases, this line is partly established by having quite formal assessment events for awarding the standards, so that the remainder of class time can be focused on helping the students to prepare for the assessments. The more tricky aspect is that the teachers usually know the details of the assessments tasks that will be used, and have to decide how specifically to prepare the students for those tasks. To help ensure national consistency in the assessment of internally assessed standards, there are nationally supervised processes for moderation of the internal assessment tasks and their marking, but the extent and effectiveness of this moderation remain a matter for substantial debate. At its best, moderation can be a valuable source of professional development for teachers.

See also: Formative Assessment; Moderation of Student Work by Teachers; Summative Assessment by Teachers.

Bibliography

- Carr, M. (2001). *Assessment in Early Childhood Settings: Learning Stories*. London: Sage.
- Clay, M. M. (1993). *An Observational Survey of Early Literacy Achievement*. Portsmouth, NH: Heinemann.
- Crooks, T. J. (2002). Educational assessment in New Zealand schools. *Assessment in Education: Principles, Policy and Practice* 9(2), 237–253.
- Ministry of Education (1993). *The New Zealand Curriculum Framework*. Wellington: Learning Media.
- Ministry of Education (2007). *The New Zealand Curriculum*. Wellington: Learning Media.

Relevant Websites

<http://www.minedu.govt.nz> – Ministry of Education.

<http://www.nzcer.org.nz> – New Zealand Council for Educational Research.

<http://www.nzqa.govt.nz> – New Zealand Qualifications Authority.

<http://www.tki.org.nz> – Te Kete Ipurangi: The Online Learning Centre; Assessment.

<http://www.tki.org.nz> – Te Kete Ipurangi: The Online Learning Centre; asTTle.

<http://www.tki.org.nz> – The New Zealand Curriculum Exemplars.

Classroom Assessment in Policy Context (Sub-Saharan Africa)

S J Howie, L Zimmerman and K Draper, University of Pretoria, Pretoria, South Africa

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This article focuses on the policy and practice pertaining to assessment in schools from the sub-Saharan region. In particular, the case of classroom assessment in South Africa is presented in some detail, while information and examples are also provided from other countries such as Botswana, Ghana, and Tanzania. As a starting point, it is first necessary to contextualize classroom assessment by providing an overview of issues that continue to play a role in education in sub-Saharan Africa (SSA).

SSA's population of more than 694 million people is one of the poorest (41.1% of people lived on less than US\$1 or US\$2 a day in 2004 (World Bank, 2008a)) and most underresourced regions of the world with some of the highest illiteracy rates, highest exclusion of females from education, and the lowest access to education. The average life expectancy is 46 years, compared to 78 years in developed countries (EFA, 2007: 228). The gross national product (GNP) per capita is the lowest of all the world's regions at US\$601 (compared to US\$1416, the next lowest) (EFA, 2007: 229). The adult literacy rate stands at 61% compared to 77% of developing countries and 99% of developed countries (EFA, 2007: 236). Poverty and substandard living conditions prevail throughout the region, exacerbated by violent civil and tribal conflicts. As a result, access to and the quality of education offered to children, vary greatly within and between countries in the region. Vast inequities characterize the region with children from the richest (20%) of the households having on average, more than 11 times the chance of reaching grade 9 than those from the poorest (40%) of households (Lewin, 2004 in World Bank, 2008a).

Access to education in SSA was strongly emphasized after the 1990 World Conference on Education for All (EFA) in Jomtien, Thailand, particularly with the move to provide universal primary education. Nonetheless, the gross enrolment rates for this region are the lowest in the world. The primary education gross enrolment ratio for SSA is 79% with the next lowest being the Pacific region at 94% (EFA, 2007: 268). In contrast, the gross enrolment ratio for secondary education in SSA is 36% compared to the next lowest region at 64% (EFA, 2007: 292). Across Africa, gross enrolment rates for secondary students vary below 20% and above 80%, with Botswana having one of the highest transitions from primary education to secondary (almost 100%) (Bregman and Simmonet, 2004; in World Bank, 2008b: 5).

In SSA, an estimated 86 million children are of secondary school age; however, approximately 63 million are estimated to be excluded from school (Bregman and Simmonet, 2004 in World Bank, 2008b: 5).

The curriculum structure also varies across SSA as seen in Table 1. Primary education varies between 6 and 7 years for the five countries listed, while secondary education ranges from 5 to 6 years.

Types of Assessments Undertaken in SSA

There is a paucity of literature and research emanating from the SSA region on assessment practices in general and classroom-assessment practices in particular. Countries conduct and participate in a range of assessment activities across the regions. These include national and international assessments, external examinations, continuous assessment, and traditional and alternative forms of assessment in the classroom. In this section, the first three are briefly described.

National and International Assessments

While national examinations have long been prevalent in African education systems, national assessments are a relatively new occurrence. National assessments are designed to describe achievement levels of an entire education system or a clearly defined part of it via collection of school-level data. Four major categories of national assessments have been identified in Africa. Three have involved similar activities in several countries. These are the monitoring learning achievement (MLA) project; the Southern Africa Consortium for Monitoring Educational Quality (SACMEQ), which has been implemented twice thus far and is about to be undertaken for a third time; and the *Programme d'Analyse des Systèmes Educatifs des Pays de la CONFEMEN* (PASEC). In MLA I, in which 40 African countries participated, grade-4 learners were assessed in literacy, numeracy, and life skills. In MLA II, in which 11 African countries took part, grade-8 learners' performance in mathematics and science were assessed (Kellaghan and Greaney, 2005).

At least 20 countries in Africa have participated in large-scale assessments through major international studies organized under the auspices of the International Association for the Evaluation of Educational Achievement

Table 1 Curriculum structure in five countries of sub-Saharan Africa

<i>South Africa</i>	<i>Botswana</i>	<i>Ghana</i>	<i>Tanzania</i>	<i>Senegal</i>
Basic education 9 years	Basic education 10 years	Free compulsory universal basic education	–	–
Foundation phase 3 years	Primary lower level 4 years	Primary education 6 years	Primary education 7 years	Primary education 6 years
Intermediate phase 3 years	Primary upper level 3 years	–	–	–
Senior phase 3 years	Junior secondary 3 years	Junior secondary 3 years	Junior secondary 4 years	Junior secondary 4 years
Further education and training 3 years	Senior secondary 2 years	Senior secondary 3 years (to become 4 years)	Senior secondary 2 years	Senior secondary 2 years

Adapted from World Bank (2008b). Curricula examinations and assessment in secondary education in Sub-Saharan Africa. World bank working paper No.128. Washington: World Bank, p 16.

and the SACMEQ (see Howie, 2008). The last decade has seen a significant increase in the number of studies and in particular, an increase in the number of African countries participating in international studies:

- South Africa was the only African country to participate in Trends in International Mathematics and Science Study (TIMSS) 1995 whereas in 1999, they were joined by Morocco and Tunisia and then in 2003, Botswana, Egypt, and Ghana joined South Africa, Morocco, and Tunisia.
- A voluntary grouping of 15 ministries of education in Southern and Eastern Africa, together with the International Institute of Educational Planning (IIEP) in Paris, cooperate to carry out educational policy research (SACMEQ). Between 1995 and 1998, eight of these ministries participated in SACMEQ I by collecting information on educational inputs, human and material resource allocation, and the literacy levels of grade-6 learners. Teachers were also assessed except in South Africa and Mauritius. From 1999 to 2002, Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, mainland Tanzania and Zanzibar, Uganda, Zambia, and Zimbabwe participated in SACMEQ II (Kellaghan and Greaney, 2005: 279–280).
- South Africa and Morocco participated in the second study of the Progress in International Reading Literacy studies (PIRLS), 2006. Morocco previously participated as the only African country in the PIRLS, 2001.

Public (External) Examinations in SSA

Public or external examinations have played a major role in modern education in African countries. These examinations may be administered at the end of primary schooling when students are assessed in the major subjects of the curriculum; after 2–3 years of secondary school usually in a wider range of subjects; and at the end of secondary school. In secondary education, these examinations have played a key role in determining the implemented

curriculum in schools and also acting as gatekeepers to schools and providing an evaluation of students at school exit which is likely to have important consequences for their further education and even life chances (Kellaghan, 2004).

In South Africa, as in many other SSA countries, the final year of schooling is dominated by the high-stakes National Senior Certificate (NSC) exam. As with the previous so-called matric examination, the examination is perceived to have various roles. For instance, it is seen by many as a gatekeeper to employment and higher education and is held in high esteem by members of the public as it represents an indication of achievement. In SSA, examination results need to be interrogated. In Tanzania, the pass rate is high but the University of Dar es Salaam has developed its own examination tests because the quality of school-leaving students is not considered at par with the examination results (World Bank, 2008: 60).

In addition, inadequacies in the quality of public examinations in Africa have been identified (Kellaghan and Greaney, 2005). Most examinations take the form of the paper-and-pencil variety ignoring areas of knowledge and skill, which although often specified by curricula, cannot be measured by these tests. A high emphasis is placed on cognitive skill in these tests with little on practical skills. In some instances, these examinations also focus on low taxonomic skills such as recall and recognition of factual content, rather than more advanced higher-order skills such as synthesis of information or application of knowledge. These examinations may also contain little reference to everyday life outside of school and the quality of the examination questions is often poor. Another key factor in less than optimal performance in these examinations for many African learners, is a lack of proficiency in the language used for testing (Kellaghan and Greaney, 2006).

Continuous Assessment in SSA

Examinations only allow for a limited measurement of certain skills and knowledge. Countries in SSA acknowledge the need for a wider range of assessment practices,

using a variety of tools and strategies to assess learning progress. This model of assessment is generally referred to as continuous assessment or CASS. CASS is criterion-referenced and allows teachers to continuously monitor the progress of their learners over time through cumulative judgments.

In **Table 2**, the emphasis in assessment was found to be on CASS in the early phase of language education in South Africa, Botswana, Ghana, and Tanzania. No formal assessments such as tests are used, except in Ghana. The general approach is informal assessment with teachers observing learners' language performance in speech and writing. In Tanzania and in Ghana, CASS is practiced as the continuous evaluation of subject content covered (norm-referenced tests) with its main purpose being to promote students through periodic tests to the next educational level or class. Tanzania has now included a 50% CASS in the final examination (HSRC, 2008). The role of CASS in South Africa will be considered within the realms of the overall policy on assessment.

Assessment Policy and Practice in South Africa

The end of apartheid in the early 1990s and the emergence of a democratic South Africa after the first democratic elections in 1994 resulted in significant changes in education policy. The fundamental framework of the Department of Education was set out in the first White Paper in 1995: Education and Training in a Democratic South Africa after which a number of policies were promulgated for implementation in school. Thus developments in education and assessment policies in South Africa can be discussed in two distinct phases since the end of apartheid in 1994 (Lubisi and Murphy, 2002).

Table 2 Assessment types found in grades 1–3 for language assessment

	<i>Assessment</i>		
	<i>Continuous/formative</i>	<i>Portfolio/assignments</i>	<i>Summative/tests</i>
South Africa	Yes	Yes	No
Botswana	Yes	No	No
Ghana	Yes	Yes	Yes (tests and CASS)
Tanzania	yes	No	Yes

From Human Sciences Research Council (2008). A review of the curriculum documents for English second language learning in Grades 1–3 in twelve African countries. *Report Prepared for Maskew Miller Longman*. Pretoria: Human Sciences Research Council, p 7.

Curriculum 2005

The systemic transformation in the schooling system started in 1996 with the introduction of Curriculum 2005 (C2005), introduced to teachers in 1998. It was so named because after its introduction in 1998, a planned phasing-in of the curriculum was expected to result in all school grades up to grade 12 implementing the new curriculum by 2005, a timeframe that later proved to be unmanageable. The new curriculum model drew upon a number of ideas and trends in the international arena, shaping them to fit local conditions (Botha, 2002: 362). This curriculum was guided by principles of outcomes-based education (OBE) (Spady, 1994) which the new government had undertaken as a central feature of its educational-reform plan.

OBE is a learner-centered approach that emphasizes what the learner should know, understand, demonstrate, and become. Teachers and learners focus on predetermined results or outcomes which must be achieved by the end of each learning process. The idea of critical cross-field outcomes is fundamental to this model, and these are subdivided into seven critical and five developmental outcomes, which give prominence to the cultivation of cognitive capacity (Lombard and Grosser, 2008).

As part of this educational-reform process, the school system was restructured. **Table 3** illustrates the old and new structures since the introduction of C2005.

The old structure was based on a primary–secondary division while the new structure is based on a division between the general education and training (GET) phase from grades R (reception year) to 9 and the further education and training (FET) phase from grades 10 to 12. The FET band was designed to diversify curricula beyond traditional concerns for tertiary education. Learners in this phase of education can choose options that may prepare them for higher education, vocational

Table 3 Comparison of old and new structure of education in South Africa pre- and post-1994

<i>Old structure (pre-1994)</i>	<i>New structure (post-1994)</i>
Senior Certificate Exam (matric)	Further Education and Training (FET) Certificate
Std 6 (grade 8)–std 10 (grade 12)	Grades 10–12
	General Education and Training (GET) Certificate
Std 3 (grade5)–std 5 (grade7)	Grades 7–9
	Senior phase GET
Std 1 (grade 3)–std 2 (grade 4)	Grades 4–6
Class 1 (grade 1)–class 2 (grade 2)	Intermediate phase GET
	Grades 1–3
	Foundation phase GET
	Grade R (Reception)

From Lubisi, R. C. and Murphy, R. J. L. (2002). Assessment in South African Schools. *Assessment in Education* 9(2), p 258.

Table 4 Exit points, continuation, and exit skills in South Africa, compared to Botswana, Tanzania, and Ghana

	<i>When?</i>	<i>Continuation into</i>	<i>Exit skills</i>
Exit point 1 After primary education	Botswana – after 7 years of primary education, pupils write the primary-leaving school examination (PLSE) Tanzania – pupils leave after sitting for the PSLE	World of work (unskilled labor) and society Junior secondary education (JSE)	Reading and writing Basic numeracy Language proficiency in the instructional language for continuation into JSE Basic problem solving
Exit point 2 After junior secondary	South Africa – pupils exit with GET certificate, 9 years of basic education. Botswana – pupils leave with a junior secondary (JC) certificate, 10 years of basic education Tanzania – pupils leave with general certificate of education (GCE O-level), 11 years schooling Ghana –	Secondary education, general streams Secondary education, technical streams Vocational education. World of work (low-skilled labor) and society	Self and social responsibilities Basics of learning to learn Problem solving English (French) Mathematics (including geometry) Academic knowledge for continuation into general secondary education
Exit point 3 After senior secondary education	South Africa – pupils leave with FET certificate, 12 years of schooling Botswana – pupils leave with Botswana general certificate of secondary education, 12 years of schooling Tanzania – pupils leave with GCE A-level, 12 years schooling Ghana –	General tertiary education. Technical tertiary education (polytechnics) World of work and society	Advanced learning to learn Problem solving Specific subject and discipline knowledge

Adapted from World Bank (2008b). Curricula examinations and assessment in secondary education in Sub-Saharan Africa. World bank working paper No.128. Washington: World Bank, p 18.

education, careers, and self-employment (Lubisi and Murphy, 2002). **Table 4**, illustrating the exit points for South Africa compared to three other SSA countries, is shown below.

The rapid and dramatic change to the curriculum and so many parts of the education system highlighted the weakness in C2005 and in 2000, a review committee was appointed (Chisholm, 2000). Chisholm (2007) observes that C2005's "... original formulations were clothed in a complex framework of outcomes that provoked a range of criticisms drawing attention to the behaviourist underpinnings, excessive assessment requirements, and difficulty of implementation in under-resourced contexts with poorly-trained teachers," the review led to the conclusion that it made little difference to what actually happened in classrooms. Well-resourced schools were thought to be better able to adopt learner-centered approaches and new assessment methods than poorly resourced schools. There were complaints about the language used, excessive paperwork related to new forms of CASS and expectations that were too complex. A revision was therefore proposed and accepted (Chisholm, 2007: 298).

The Revised National Curriculum Statement

In 2002, a new curriculum, the revised national curriculum statement (RNCS) was issued. The fundamental principles of C2005, that is OBE, remained the same. It covered education for grade R (reception year) to grade 9 together with a draft version of the new suggested curriculum documents for grades 10–12. The RNCS introduced a more streamlined and less-technical curriculum. The RNCS was implemented in schools in 2004 for the foundation phase (grades R–3), 2005 for the intermediate phase (grades 4–6), and then 2006/7/8 for grades 7, 8, and 9. The new national curriculum statement (NCS) was introduced in the FET phase in 2006 with the first cohort of -grade-12 pupils writing the NSC in 2008 (Lubisi and Murphy, 2002).

Current National Assessment Policy in South Africa

The prominence of assessment in South Africa emerged particularly in the late 1990s as teachers grappled with the

new (and alternative) vision on assessment as stated in the new curriculum documents. According to the RNCS, assessment is still the continuous planned process of gathering information on learner achievement as it was in C2005, based on the principles of OBE. Outcomes-based assessment (OBA) (authentic assessment) is practiced (Combrinck, 2003). This approach is aimed at clarifying what learners are expected to achieve and assists the process of learner assessment by placing assessment standards at the center of the assessment process in every grade. The assessment standards describe the level at which learners in each grade should demonstrate their achievement of the learning outcomes as well as the ways of demonstrating their achievement. OBA is a move from a summative norm-referenced approach to a formative criterion-referenced approach, hence the shift to CASS which is a formative continuous model based on the principle of authentic assessment that encompasses group or individual projects, portfolios, and performance-based assessment (Combrinck, 2003).

CASS in South Africa was intended to be used for formative and diagnostic purposes within the classroom. South African CASS should be based on tests, classroom interaction, project work, investigations, and assignments. The use of learner portfolios is part of the strategy of CASS. Evidence of learners' work is filed and used as a means to determine progress or promotion to the next grade. The portfolio is used to track progress, growth, and learner achievement with regard to expected outcomes, create an opportunity for the learner to reflect on his/her growth and set goal for self-development, and inform teachers' planning as well as intervention strategies. The South African basic education examination consists of 75% CASS, and 25% so-called common-task assessment (CTA), a standardized external examination written by grade 9 learners in all government-funded schools. The inclusion of CASS is often problematic when it comes to grading. CASS marks are usually far above the final examination marks of students indicating that there is a problem. Most often, the CASS marks are statistically moderated by the national quality assurance body for the national examinations body, to bring them in line with the final examination results. However, students passing examinations are widely reported as lacking in basic reading, writing, and mathematical skills.

In South Africa, assessment in the NCS is an integral part of teaching and learning and should be part of every lesson to complement learning activities. In addition, teachers should plan a formal year-long program of assessment. Together, the informal daily assessment and the formal program of assessment should be used to monitor learner progress through the school year. South Africa has adopted a policy of CASS in the form of informal daily assessment together with the formal program of assessment. These assessment approaches should be used in

conjunction to develop learners' knowledge, skills, and values; assess learners' strengths and weaknesses; provide additional support to learners; revisit or revise certain sections of the curriculum; and motivate and encourage learners.

Classroom Assessment Under the RNCS (GET) and NCS (FET)

Before the introduction of C2005, assessment in South African schools was considered a straightforward process of testing pupils' ability to reproduce information that was presented to them through a content-based curriculum. The first years of the change to an OBE approach required teachers to use forms of criterion-referenced assessment to monitor pupil progress which stated learning outcomes which pupils had achieved. Before teachers had become competent at using this assessment system, the RNCS guidelines shifted to a standards-referenced approach which required of teachers to make qualitative decisions on frames of reference on different levels of achievement, an activity foreign to many South African teachers (Vandeyar and Killen, 2006).

Assessment within an OBE approach, now focuses on the achievement of defined outcomes, referenced against pre-defined criteria. A variety of tools and strategies are used with the aim of encouraging lifelong learning skills. Classroom assessment is based on a program of daily assessment tasks as well as a year-long formal program of assessment (see **Table 5**). The daily assessment tasks take place in the classroom with learner progress being monitored during learning activities. This informal daily monitoring of progress can be done through question-and-answer sessions; short assessment tasks completed during the lesson by individuals, pairs, or groups; or homework exercises. These tasks can be marked by the learners themselves and this allows learners to learn from and reflect on their own performance. The results of the informal daily assessment tasks do not have to be formally recorded.

South Africa's school assessment policy is guided by the *National Protocol on Assessment for Schools* in both GET and FET and training band (grades R–12) published 21 October 2005 by the Department of Education.

Table 5 Weighting assigned to various forms of assessment in grades 10–12 in South Africa

	<i>Daily assessment tasks</i>	<i>Formal program of assessment</i>	<i>External formal examinations</i>
Grade 10 and 11	25%	75%	none
Grade 12	none	25% (moderated)	75% (examination)

All schools are expected to implement this national protocol on assessment from January 2006. It was written within the frameworks of the outcomes-based national curriculum statements (grades R–9 and grades 10–12) and provides schools with a regulatory framework for the management of school assessment records and basic requirements for learner profiles, teacher and learner portfolios, and report cards. The protocol gives guidelines for recording and reporting assessments (see **Table 6**).

Assessment is viewed as an integral part of teaching and learning to be incorporated into all levels of planning. Assessment standards stated in the RNCS (2002) define the minimum requirements for achieving the stated learning outcomes for a grade and for a learning program or learning area. These assessment standards “describe the level at which learners should demonstrate their achievement of the learning outcome(s) and the ways (depth and breadth) of demonstrating their achievement.” The assessment standards are also designed to show how conceptual progression will occur in a learning area. The goal of the assessment standards is to help teachers know when learners have achieved a learning outcome; to show the minimum levels which learners should achieve in a specific grade; and to show learners’ level of achievement and progress in a specific grade (DoE, ND, p. 8).

A Final Word

It is apparent that the same problems with classroom assessment that are reported in the international literature (Kellaghan and Greaney, 2005) have also been observed in the African context which means that in Africa, the quality of teachers’ assessment practices is seen to be deficient in many ways. Prevalent assessment procedures in African classroom contexts, including South Africa, have been attributed to the nature of the teaching and learning situation in which teachers are often dominant and learners are passive. Other explanations for inadequate existing assessment practices include large class sizes, poorly qualified teachers, poor facilities, and a shortage of learning materials and storage. The use of drill exercises, recitation, and mock examinations in preparation for public examinations have also been cited as a potential cause of poor assessment procedures (Kellaghan and Greaney, 2005). Teachers too, tend to use poorly focused questions, questions requiring short answers drawing on factual knowledge and as a result, questions may lead to repetition and recall instead of reflection and application of knowledge and skills. Another difficulty may be the inability of teachers to use strategies and procedures designed to develop students’ higher-order cognitive skills (Kellaghan and Greaney, 2005).

In South Africa, despite changes to the curriculum policy for assessment, research reveals that teachers may be either unwilling or unable to change their assessment

practices (Vandejar and Killen, 2007). Although government educational-policy changes were aimed at redressing past inequalities in educational provision, they have not necessarily resulted in major changes at classroom level with some teachers still applying the same pedagogical practices they used a decade ago. It appears as if entrenched assessment practices may be hampering the government’s transformation plans (Vandejar and Killen, 2007). If it is accepted that all pedagogical acts are influenced by teachers’ conceptions about teaching and learning and the processes and purposes of assessment, then these conceptions may also be barriers to change in practices. One of the strongest influences on teachers’ conceptions of assessment is their understanding of their subject(s), which is an ongoing difficulty in South Africa because many non-White teachers were trained during the apartheid era under a teacher-education system which left them poorly trained and underqualified. The apartheid education system emphasized content, conformity, and high-stakes summative assessment and, as it is difficult to change teachers’ conceptions of practice, it is not surprising that ideas of assessment, founded on learner and school accountability, still abound (Vandejar and Killen, 2007). In a small exploratory study, Vandejar and Killen, (2007) found that teachers may be struggling with OBA and may still hold conceptions of assessment resulting from policies and practices that are no longer relevant, that is, practices that were used prior to 1994. In South Africa, concerns have been raised about teachers’ ineffective teaching methods, weak subject knowledge, and misunderstandings of the demands of the curriculum (Fleisch, 2007) which no doubt incorporate difficulties with assessment practices too. Indeed, teachers’ difficulties with classroom assessment and reported assessment practices are indicated as issues of continued concern (Vandejar and Killen, 2007; Moloi and Strauss, 2005).

While education policy in Botswana (Revised National Policy on education, 1994) recommends the use of criterion-referenced assessment, teachers still commonly use norm-referenced testing which largely test knowledge recall. End-of-term school-based examinations are used for ranking pupils according to performance. Senior secondary results are typically used to stream students into science or humanity streams, depending on performance in the examinations. In 2006, only terminal examinations were used in secondary education but practical work, project work, CASS with criterion-referenced testing are expected to complement the terminal examinations over time. However, in Botswana, many teachers lack sufficient skills in assessment to develop their own tests and examinations and often resort to copying other tests and examinations. This means that teachers in Botswana as well, have still to adopt and implement real-CASS practices (SEIA, 2008).

If the implementation of a new curriculum demands new forms of assessment but the implementation of the assessment strategy lags behind, there is little change of

Table 6 Formal recorded assessment tasks (formal program of assessment)

<i>Learning program</i>		<i>Term 1</i>	<i>Term 2</i>	<i>Term 3</i>	<i>Term 4</i>	<i>Total</i>
<i>Grades R–3</i>						
Literacy (languages)		4	4	4	4	16
Numeracy (mathematics)		3	3	3	3	12
Life skills (life orientation)		1	1	1	1	4
<i>Grades 4–6</i>						
Language 1		3	3	3	3	12
Language 2		2	2	2	2	8
Language 3 (optional)		1	1	1	1	4
Mathematics		3	3	3	3	12
Natural sciences		2	2	2	2	8
Social sciences		2	2	2	2	8
Technology		1	1	1	1	4
Economic and management sciences (EMS)		1	1	1	1	4
Life orientation		1	1	1	1	4
Arts and culture		1	1	1	1	4
<i>Grades 7–9</i>						
Language 1		4	4	4	4	16
Language 2		3	3	3	3	12
Language 3 (optional)		3	3	3	3	12
Mathematics		3	3	3	3	12
Natural sciences		2	2	2	2	8
Social sciences		2	2	2	2	8
Technology		1	1	1	1	4
Economic and management sciences (EMS)		1	1	1	1	4
Life Orientation		1	1	1	1	4
Arts and culture		1	1	1	1	4
<i>Grades 10 and 11</i>						
Home language (HL)		5	5 ^a	5	4 ^a	19
Language 2 choice of HL or first additional language (FAL)	HL	5	5 ^a	5	4 ^a	19
	FAL	4	4 ^a	4	3 ^a	15
Life orientation		1	1	1	2	5
Mathematics or maths literacy		2	2*	2	2 ^a	8
Choice subject 1		2	2*	2	1 ^a	7
Choice subject 2		2	2 ^a	2	1 ^a	7
Choice subject 3		2	2 ^a	2	1 ^a	7
<i>Grade 12</i>						
Home language (HL)		6	6	5		17
Language 2 choice of hl or first additional language (FAL)	HL	6	6	5		17
	FAL	5	5 ^a	4 ^a		14
Life orientation		1	2	2		5
Mathematics or maths literacy		3	2 ^a	2 ^a		7
Choice subject 1		2	2 ^a	(2)3 ^a		6–7 ^b
Choice subject 2		2	2 ^a	(2)3 ^a		6–7 ^b
Choice subject 3		2	2 ^a	(2)3 ^a		6–7 ^b

^aOne of these tasks must be an examination.

^bThe number of internal tasks per subject differs from 6 to 7 as specified in section 3 of the subject-assessment guidelines.

curricula changes making their way into the classroom. The example of South Africa reveals this as a serious challenge but this is a problem across SSA where there is a lack of alignment between the intentions of the curriculum and the quality of assessment.

Combrinck (2003) asserts that in implementing OBA, certain factors need to be in place for the new assessment approach to be successful:

1. an adequate in-service training program;
2. a support system for schools and teachers in implementing the new approach; and
3. a clear school assessment policy that addresses all aspects of assessment including a policy on cultural diversity.

There are a number of logistical and practical problems, such as the lack of clear assessment instruments

and marking schemes in the hands of teachers and these are persistent problems in assessment reform. Coupled with the practical problem is that regardless of the quality of assessment protocols and tools, the high incidences of learners being assessed in languages other than their vernaculars across Africa, languages in which in many cases they are not educationally proficient, will continue to impact negatively on learner outcomes and the quality of educational output across Africa.

Bibliography

- Botha, R. J. (2002). Outcomes-based education and educational reform in South Africa. *International Journal of Leadership in Education* 5(4), 361–371.
- Bregman, J. and Simmonet (2004). What's next? Coping with Successful Primary Education For All. In Govender, P. and Grudz, S. (eds.) *Back to the Blackboard; Looking Beyond Universal Primary Education in Africa*. Johannesburg: The South African Institute for International Affairs.
- Chisholm, L. (2000). A South African curriculum for the twenty first century report of the review committee on curriculum 2005. *Report of the Review Committee on Curriculum 2005*. Pretoria: Ministry of Education.
- Chisholm, L. (2007). Diffusion of the national qualifications framework and outcomes-based education in southern and eastern Africa. *Comparative Education* 43(2), 295–309.
- Combrinck, M. (2003). An international comparative perspective on outcomes-based assessment: Implications for South Africa. *Perspectives in Education* 21(1), 51–66.
- Department of Education (2005). *National Protocol on Assessment for Schools in both the General and Further Education and Training Band* (Grades R-12).
- Department of Education (ND). Assessment guidelines for Foundation Phase Grades R-3.
- Fleisch, B. (2007). *Primary Education in Crisis: Why South African Schoolchildren Underachieve in Reading and Mathematics*. Cape Town: Juta and Company.
- Howie, S. J., Barnes, H., Herman, C., van Loggerenberg, A., and Mapile, S. (2006). Secondary education in Africa. Science, mathematics and ICT Study. South Africa country profile. Africa region human development Working Paper Series. Amsterdam: Vrije Universiteit Amsterdam.
- Human Sciences Research Council (2008). A review of the curriculum documents for English second language learning in Grades 1–3 in twelve African countries. *Report Prepared for Maskew Miller Longman*. Pretoria: Human Sciences Research Council.
- Kellaghan, T. (2004). Public examination, national and international assessments and educational policy. http://siteresources.worldbank.org/INTAFRREGTOPSEIA/Resources/paper_Kellaghan.pdf (accessed July 2009).
- Kellaghan, T. and Greaney, V. (2005). Chapter 11. Monitoring performance: Assessment and examinations. In Verspoor, A. D. (ed.) *The Challenge of Learning: Improving the Quality of Basic Education in Sub-Saharan Africa*, pp 271–292. Paris: Association for the Development of Education in Africa (ADEA).
- Lewin, K. (2004). Beyond primary education for all: Planning and financing secondary education in Africa (SEIA). *Paper Presented at the SEIA Donor Conference* Amsterdam, October.
- Lombard, K. and Grosser, M. (2008). Critical thinking: Are the ideals of OBE failing us or are we failing the ideals of OBE? *South African Journal of Education* 28, 561–579.
- Lubisi, R. C. and Murphy, R. J. L. (2002). Assessment in South African Schools. *Assessment in Education* 9(2), 255–268.
- Mafumiko, F. and Leliveld, M. (2006). Secondary education in Africa. Science, mathematics and ICT Study. Tanzania country profile. Africa region human development working paper series. Amsterdam: Vrije Universiteit Amsterdam.
- Moloi, M. and Strauss, J. (2005). The SACMEQ II project in South Africa: A study of the conditions of schooling and the quality of education. *South Africa Working Report*. Harare, Zimbabwe: SACMEQ; Ministry of Education, South Africa.
- Ramorogo, G. J., Koosimile, A. T., Tawana, L., and Prophet, R. B. (2006). Secondary education in Africa. Science, mathematics and ICT study. Botswana country profile. *Africa Region Human Development Working Paper Series*. Amsterdam: Vrije Universiteit Amsterdam.
- Spady, W. G. (1994). *Outcomes-Based Education: Critical Issues and Answers*. Arlington, VA: American Association of School Administrators.
- Vandeyar, S. and Killen, R. (2006). Beliefs and attitudes about assessment of a sample of student teachers in South Africa. *Africa Education Review* 3(1+2), 30–47.
- Vandeyar, S. and Killen, R. (2007). Educators' conceptions and practice of classroom assessment in post-apartheid South Africa. *South African Journal of Education* 27(1), 101–115.
- World Bank (2008a). Global Monitoring Report: MDGs and the Environment, Agenda for Inclusive and Sustainable Development. Washington: World Bank.
- World Bank (2008b). Curricula, examinations and assessment in secondary education in Sub-Saharan Africa. World bank working paper No.128. Washington: World Bank.

Further Reading

- Ampiah, J. G. and Akyeampong, K. (2006). Secondary education in Africa. Science, mathematics and ICT Study. Ghana country profile. *Africa Region Human Development Working Paper Series*. Amsterdam: Vrije Universiteit Amsterdam.
- Howie, S. J. (2008). Measuring the health of the education system: Lessons from international studies for South Africa. *Keynote Address Presented at the Association for the Study of Evaluation and Assessment in Southern Africa conference*, (ASEASA), Pretoria, South Africa, 9–11.
- Howie, S. J., Venter, E., Van Staden, S., et al. (2007). Progress in international reading literacy study 2006. *Summary Report. South African Children's Reading Literacy Achievement*. Centre for Evaluation and Assessment: University of Pretoria.
- Pryor, J. and Lubisi, C. (2002). Reconceptualising educational assessment in South Africa: Testing times for teachers. *International Journal for Educational Development* 22, 673–686.
- UNESCO (2007). *Education for All: Strong Foundations, Early Childhood Care and Childhood*. Paris: UNESCO.
- Vandeyar, S. (2005). Conflicting demands: Assessment practices in three South African primary schools undergoing desegregation. *Curriculum Enquiry* 35(4), 462–481.
- World Bank (2007). Developing science, mathematics and ICT education in Sub-Saharan Africa. World bank working paper No. 101. Washington: World Bank.

Relevant Websites

- <http://WWW.ADEAnet.org> – Association for the Development of Education in Africa (ADEA).
- <http://www.sacmeq.org> – Southern and Eastern Africa Consortium for Monitoring Educational Quality.

Moderation of Student Work by Teachers

G S Maxwell, Brisbane, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

External assessment – An assessment designed and managed by an agency outside schools, as with public examinations and standardized tests.

Internal assessment – An assessment designed and conducted by teachers within a school.

Moderation – A process for producing consistency across assessors in qualitative judgments of student performance or achievement.

Standards – The representations of the quality of achievement or performance that needs to be demonstrated for awarding particular grades, ratings, or score ranges.

Definition

Moderation is a process for producing consistency across assessors in qualitative judgments of student performance or achievement. It is especially relevant to teacher assessments (teacher-managed or school-based assessments) where results are expected to be comparable across teachers and across schools. However, it can be relevant to any situation involving multiple assessors (perhaps across multiple sites or observations) where judgment is required concerning performance quality.

Moderation contributes to quality assurance of assessment procedures and outcomes. Quality assurance, in its broadest sense, refers to processes for establishing confidence in the quality of procedures and judgments. Confidence is always a matter of degree. More rigorous processes are required where the stakes are higher and more confidence is needed. Where less confidence can be tolerated, processes can be less rigorous. This means that moderation procedures need to be tailored to the context.

Quality assurance sometimes refers more narrowly to processes for encouraging or approving practices that are assumed to produce desired consequences, without necessarily checking that they do. Checking on the outcomes can be characterized as quality control, which means reviewing, and where necessary adjusting or correcting, the outcomes (Harlen, 1994). That is, moderation as quality control involves checking assessment procedures and outcomes to ensure that there is consistency in the judgments (student results) before they are used or reported.

Typically, quality control is embedded within broader processes of quality assurance – that is, supported by other aspects of quality assurance (maybe all seen as part of the moderation process) such as assessor training, program approval, and implementation monitoring (Daugherty, 1994).

This distinction may also be characterized as moderation for improvement (through professional development) versus moderation for accountability (quality control) (Maxwell, 2002). The former develops the capability of teachers to undertake appropriate assessments and make consistent and comparable judgments. The latter provides official confirmation of results that are reported or used publicly. Clearly, each can complement the other: quality control can contribute to improvement in future assessments through the feedback it provides to assessors; professional development can provide the competent assessors needed for producing consistency.

How much emphasis to place on quality control depends on how the assessments are used, their public visibility and their status. High-stakes assessment, that is, assessment with external consequences – such as selection decisions (individual consequences), professional certification (public consequences), or resource allocation (institutional consequences) – demands more consistency and therefore greater quality control. Where the consequences are less public and more diffuse – such as classroom assessments that do not contribute to public reporting or student certification – the aim may be to raise assessment quality and consistency in the longer term through professional development processes rather than expect substantial consistency in the short term. In many situations, this is more realistic, affordable, and effective, especially where the emphasis is on formative assessment.

Situations Where Moderation Is Applicable

Moderation is relevant to all levels and types of education and training. It is also potentially relevant to other judgment situations such as workplace assessments, case-work assessments, and quality audits – wherever application of consistent standards across multiple assessors and/or multiple sites is important for fairness and equity or the promotion of consistent high-quality professional judgments.

Some illustrative situations are school-based end-of-secondary-school certification, other school-based assessments, vocational training assessments, and university course assessments, which are discussed below.

School-Based Certification (High Stakes)

Moderation has been keenly identified with school-based (that is, teacher-directed) assessments contributing to state-awarded (provincial or national) certificates in secondary school (high school). The aim of school-based assessment is improved validity through use of an expanded range of assessments, authentic performance assessments, multiple assessments over time (continuous or progressive assessment), and achievement portfolios. For public certification, equity considerations require consistency and comparability across teachers and schools. Otherwise confidence in the certificate is compromised. This is especially important where selection decisions, such as for higher education, are based on the certificate (ACACA, 1999).

The Australian state of Queensland has since 1972 implemented and refined a system of moderation for high-stakes school-based assessments at the end of secondary school (Maxwell, 2007; QSA, 2006). Some other Australian states also have moderation processes for school-based assessments at the end of secondary school but only Queensland and the Australian Capital Territory (ACT) are entirely school based with no external examinations (apart from a core skills test – see later). High levels of comparability have been achieved in Queensland (Masters and McBryde, 1994; QSA, 2000–2007) and the system enjoys public confidence.

Moderation in such situations is directed at comparability of grades reported against a particular subject label (such as Physics or French). In Queensland, such subjects have a syllabus (content standards) and grading standards (performance standards). Schools submit their teaching and assessment programs for approval by a central agency (Queensland Studies Authority) that manages between-schools moderation processes and issues the Queensland Certificate of Education. This system allows flexibility for schools to design their own assessment programs within an overall framework of curriculum expectations and achievement standards. Other systems may allow less flexibility, for example, by imposing common assessment tasks. In all cases, successful moderation requires some way of establishing common expectations and standards against which student performance (achievement) can be judged.

Other School-Based Assessments (Low Stakes)

At all levels of education, there are situations where assessments are set and/or conducted and/or marked by

teachers and where consistency in the reported results is desirable. For example, student progress reports should reflect consistent application of standards among teachers within the school. School districts might want also to see consistency in reporting across schools within the district. Moderation processes can contribute to raising the quality of teacher assessments and developing both within-school and between-school consistency in teacher judgments (Daugherty, 1997).

Some systems impose common assessment tasks at particular year levels. Other systems involve centrally determined assessment specifications and performance standards but not prespecified assessment tasks. Assessments may also be marked in the schools by the teachers or marked centrally by specially selected marking teams. Central markers can be closely monitored. On the other hand, local marking can enhance teacher professional growth and development. The specifics of moderation, including whether assessment is centrally set or open and whether marking is central or local, have to be worked out in the political, social, and institutional context and depend on the purposes, uses, and consequences of the assessments as well as the amount of consistency expected. Recent enactments of moderation-supported teacher assessments in schools can be found, for example, in England, Scotland, Wales, and Australia.

Vocational Training Assessments

National vocational training systems, such as those in England, Scotland, Australia, New Zealand, and South Africa, have developed vocational-qualification frameworks based on workplace skills. Performance expectations for different levels of expertise, typically expressed as competency requirements, are specified and assessed through *in-situ* workplace observation. Quality assurance tends to focus on front-end-delivery requirements for each training organization, such as proper training of assessors. In some countries, there is support for consistency of assessment judgments through verification or validation processes. These processes involve mandatory checks on the appropriateness of the assessment processes and judgments – for example, through site visits by external verifiers appointed by the qualifications agency or an industry-training body. Alternatively, there may be encouragement of voluntary participation in assessor networks that support the development of common understanding and application of the competency requirements. These networks allow assessors to share and compare their assessment processes and judgments with other assessors. External verification and assessor networks are instances of assessment moderation even if not so labeled. Other possibilities deserve more deliberate attention (Maxwell, 2001).

University Course Assessments

Universities worldwide are increasingly concerned about improving the quality and consistency of their assessments as part of their overall teaching and learning strategies. In the UK and Denmark, universities engage external examiners as part of their quality-assurance processes for assessment; the aim is to maintain national consistency of standards among universities although there is ongoing debate about their purpose and function and the need for a broader participatory process. External examiners are essentially external moderators. Internal moderation is beginning to emerge in the practices of UK universities, encouraged by the *Quality Assurance Agency Code of Practice* (QAA, 2006) which references moderation.

Australian universities use external examiners only for higher-degree theses, but most have specific policies on moderation for course assessment; these policies typically include a role for moderators (external to the course but not necessarily the university) to review and approve course assessments. More generally, most Australian universities now encourage internal moderation processes for developing consistency in assessment of students, especially for large classes with several markers but also for coherence of assessments across courses within the same degree program. Such practices are recent, uneven, and evolving (DEST, 2002: paragraphs 146–150).

Social Moderation versus Statistical Moderation

Linn (1996) makes a distinction between social moderation and statistical moderation. The former involves personal comparison and alignment of assessor judgments (review and coordination of teachers' ratings of student products, p. 102) whereas the latter involves statistical scaling of the results of each group of students against that group's performance on an external anchor test or examination. These have different purposes and effects and it may be confusing to refer to both as moderation. Social moderation (including the consensus moderation discussed by Linn) is the focus of discussion here. Social moderation is concerned with the direct apprehension and comparison of standards (standards referencing) whereas statistical moderation is a norm-referencing strategy. That is, in social moderation, the codes for ratings of performance (often expressed alphabetically rather than numerically) are tied to performance levels or standards. In statistical moderation, the ratings are assumed to have only relative meaning (ordinal and perhaps equal-interval scaling) and hence can be arbitrarily adjusted to match another distribution without losing their meaning.

Statistical moderation may be better called scaling or linking. Linn (1996) notes the two forms it typically takes:

scaling of internal (teacher assessed) scores against external (public examination) scores through distribution equating; and scaling of scores in different subjects against a measure of general academic achievement (in Queensland a core-skills test). The latter is done to produce an appropriate aggregation of results across subjects taken in different combinations by different students. Both depend on an appropriate match between the external measure and the internal measures and this requires careful delineation of the relevant construct (Maxwell, 1996; Newton, 2005). The focus of the discussion here is social moderation.

Moderation as a Participative Process

Social moderation is a participative process that treats assessors professionally. Moderation can be framed as a normal and positive consultative process that values consensus and builds confidence rather than as a corrective activity that compensates for presumed inadequacies or mistakes of the assessors. Arbitrarily overriding and adjusting the grades, levels or scores awarded by an assessor can undermine professional responsibility and confidence. Professional engagement is strengthened by dialog and negotiation to arrive at consensus. Even so, in high-stakes assessment, external moderators or panels may have greater power in any negotiation. How far a certifying agency chooses to go in forcing the issue depends on how high the stakes are and how much responsibility it wishes to accept for the assessment judgments. In school-based assessments, there is shared responsibility between certifying agency and schools and the certifying agency cannot usually verify directly all of the assessments. Moderation typically provides a feedback loop to advise the assessor. In the moderation processes for the Queensland Certificate of Education, checks are made on whether advice is heeded and further negotiation may be necessary before the results are approved (Maxwell, 2007).

Moderation as a Proactive Process

The aim of moderation is typically represented as delivering comparability across all the assessment results (grades, ratings, levels, and scores). Sadler (1986) argues that comparability has two meanings:

- the results must be compare-able, that is, able to be compared, and
- achievements reported by the same code (grade, level, or score) must be equivalent in terms of the standard they represent.

Assessments are compare-able only if they assess common qualities, characteristics, or criteria. Equivalence requires consistency in the application of common standards

so that all performances awarded the same grade are in fact at the same standard.

Delivering comparability implies action before judgments are final, not simply checking whether they were satisfactory afterward. This means that moderation needs to be a planned proactive process covering all phases of assessment: before (at planning phase), during (while assessments are administered and marked), and after (before the results are reported, while there is still time to amend them). In addition, there could be a follow-up evaluation of the whole process to confirm that comparability was achieved (to a satisfactory extent) and provide feedback into the next round of assessments (to strengthen and improve comparability in future) as with *post-hoc* random sampling in Queensland (QSA, 2000–2007).

Achieving Consistency in the Application of Standards

Consistency in the application of common standards is more likely if those standards are explicitly stated and exemplified. Clarification of the standards is the key activity in moderation. However, no matter how apt the standards statements and exemplars, assessors can and will interpret them differently. Moderation is directed at aligning assessors' different understandings and judgments of the relevant standards to produce shared interpretations.

In the before phase of moderation, the focus is on assessor training. This includes explanation and discussion of the performance standards (and in some cases the collaborative design of those standards). It also beneficially includes discussion focused on exemplars of the standards, either deliberately designed or drawn from previous assessments undertaken by similar students in similar circumstances. Where the assessors design the assessments themselves, as in school-based assessment designed by teachers to fit local circumstances, training is also necessary in assessment design. In some cases this training might lead to formal certification of the assessors as being qualified to assess, as is required in some vocational training systems. In school-based assessments designed by teachers, training might be followed by submission and approval of an assessment plan, including example assessments and supporting documentation, before the assessment plan is enacted. In addition, external moderators or moderation panels need to be trained both in the standards and in the role they are expected to adopt.

In the during phase of moderation, the focus is on fine-tuning assessor judgments, trouble-shooting any concerns and problems and making sure that assessors stay on target, that is, setting up conditions that support consistency among the assessors. This could involve joint

marking of a small sample of assessments early in the marking process to provide initial orientation, or periodic meetings to consider borderline or unusual cases, or more formal procedures at particular stages of the marking. The longer the delay in checking for consistency, the more difficult it is to rectify any misapplication of the standards. In large-scale school-based assessment, complex management systems may be necessary to develop consistency at three levels: within schools; between schools within a district or region; and between districts or regions.

In the after phase of moderation, the focus is on final validation or verification of student results so that they become official. This is only relevant when there is official reporting on an authorized certificate or transcript of results. Final validation or verification may involve further negotiation between an external moderator or moderation panel and an assessor over whatever adjustments are deemed necessary to establish comparability. Sufficient time has to be allowed for this, but the process might begin while assessments are still being marked, allowing early identification of problematic assessments. Successful negotiation also requires goodwill on both sides. Any breakdown in negotiation can be costly in time and resources, especially in large systems. There is usually no time to re-mark large numbers of assessments.

Factors Affecting Moderation Design

Moderation processes are tailored to the circumstances. Factors to consider in designing moderation processes include the following:

- locus of authority: whether local (e.g., school), central authority (e.g., state), or shared;
- Layers: how many layers must be managed (e.g., school, district, and state)?
- Stakes: higher stakes (e.g., assessments used for selection or placement) demand stronger moderation (tighter comparability)?
- Risks: what are the consequences of lack of comparability (personally for students or publicly for education authorities)?
- Size: how many students, but more especially how many assessors?
- Time and timing: are there set timelines to manage?
- Dispersal: how geographically or institutionally dispersed are assessors (making communication, meetings, and/or visitation more difficult)?
- Scope: single task or multiple tasks; standardized tasks (e.g., written examinations) versus situated tasks (e.g., work based); common task (centrally set) versus tailored or individualized tasks (locally set and therefore different across locations and students); projects, presentations, or portfolios?

- Tolerance: latitude in public acceptance – put most effort where comparability is most needed;
- Logistics: larger systems require more management (e.g., training moderators, assembling panels, transporting portfolios, and developing IT communication);
- Cost benefit: is it affordable; does the benefit warrant the cost?
- Professional capacity: how skilled are the assessors; can they be up-skilled; do the moderation processes contribute to long-term skills development?

Some Specific Moderation Processes

The following discussion is divided into weak quality control and strong quality control. The former is typically considered sufficient when the stakes are low but the latter necessary when the stakes are high. However, weak and strong (just as low and high) are relative concepts. Therefore, weak and strong quality control are best seen as two ends of a spectrum of possibilities allowing potential mixing of approaches.

Weak Quality Control (Low Stakes)

Assessor meetings (consortium moderation)

These can be group meetings based on communities of interest (course, school, and district). There might be a nominal leader to chair the meetings and ensure balanced participation. The purpose is to share understandings of the relevant performance standards and their application to examples of student performance brought by the participants and put on the table for consideration. The focus is on learning from each other and convergence toward common understandings of standards. This can improve comparability within the group. Assembling different groups on different occasions can create more widespread convergence over time.

Assessor partnerships and conversations (moderation exchanges and networks)

Partnerships can be small groups, perhaps just pairs, based on proximity or similarity. Relationships can be formal (requiring reporting) or informal (emphasizing mutual support). The aim is exchange of assessment practices and convergence of viewpoints on standards, again in relation to specific examples of student performance. This is sometimes extended to voluntary assessor networks and communities of practice that encourage wider sharing and dissemination of assessment ideas and practices.

Self-audit and self-moderation

Many quality-assurance processes involve self-audit. This is always a necessary first step. For example, in student

assessment, there is no point in moderating between schools if schools are internally incoherent in their practices. External moderation assumes internal consistency, that is, assumes that there is consistency across all assessors within the school. Similarly, moderation between assessors assumes that each assessor is internally consistent, that is, applies standards consistently across all the assessed performances. This is especially relevant where there are many scripts or portfolios that can take several sessions to mark. Later ratings should be checked for consistency with earlier ratings; performances given the same rating should be checked for their comparability. Assessor consistency is not only important for fairness and equity among those students; it is a prerequisite for consistency with other assessors.

Strong Quality Control (High Stakes)

Group (consensus) moderation

Where the group of assessors is small and self-contained (e.g., assessors for the same subject or course within a school or university), the group may take collective responsibility for quality control among themselves, perhaps under the direction of a coordinator. The aim is comparability across the group achieved through consensus. Where assessors have latitude in setting the assessment tasks, task equivalence has to be designed and agreed upon before assessment begins and clear understandings reached on the common performance standards to be applied. Where external approval or intergroup comparison is required, group moderation may be accompanied or followed by external expert or panel moderation.

Expert moderation

Expert moderation can be both internal and external to a school or institution. It can supplement or replace group consensus. The role of an expert moderator can vary from offering an independent viewpoint (advisory) to requiring compliance (regulatory). In large systems (e.g., in state-wide between-school moderation) teams of moderators may be needed. Moderators may conduct site visitations for observation and discussion with assessors and/or may review samples of performance evidence. Moderators provide feedback to the assessors on all relevant aspects of the assessments, but especially on ratings against the defined performance standards. Action may be requested and negotiated (e.g., adjustments to ratings). Success depends on the competence, skills, and authority of the moderator(s). The moderation team itself will need moderation (perhaps group moderation).

Panel moderation

Panel moderation can replace external moderators and fulfill the same function. The constitution, management, and operation of panels are matters for thoughtful design.

Typically, chairs and members of panels are selected from the most skilled and experienced assessors, making this a form of peer moderation. Scripts or portfolios may be formed into batches for allocation to subgroups of panelists (perhaps only two members per subgroup in a mixed design). Small differences of judgment within a chosen level of tolerance may be averaged. Larger differences require discussion and resolution within the subgroup, with the panel chair arbitrating difficult cases. In large multidistrict systems, district panels themselves need to be moderated.

Issues and Alternatives

The issues related to moderation and the alternatives offered can be summed up as below:

- A certain level of trust is necessary for successful moderation. Assessors are assumed to be professionally competent and ethical. In some situations, persistently incompetent and unethical assessors can be replaced. In other situations, for example, where assessment is continuous or *in situ*, assessors are integral. In general, provided assessor judgments are respected, moderation can provide powerful professional development and therefore strengthen trust. A balance must be struck between trust and surveillance.
- Moderation consists essentially of a series of feedback loops to improve overall quality and consistency of assessment processes and judgments. Feedback is usually based on a sample of assessments. Most assessments are not reviewed. Moderation involves negotiation and agreement on standards in relation to the sample. The assumption is that assessors will then apply those agreed standards to all the other assessments. Sometimes assessors are asked to demonstrate that they have done this but options for final validation are usually limited.
- Sampling can be random or deliberate and chosen internally or externally. Being deliberate is more efficient since it can cover the range of performance and include mid-grade and borderline cases. Atypical or unusual cases are sometimes instructive and assessors may want to seek guidance on them, while noting their atypical or unusual characteristics. In general, typical cases may be more helpful for exemplifying general standards.
- With external moderation (expert or panel) a choice must be made between blind re-marking and confirmatory review. The latter is usually more practical; keeping scripts free of assessor marks and comments is often unrealistic. The question is then whether the moderator agrees with the assessor's judgment. The moderator looks for the evidence to justify that judgment. This lays a better basis for negotiation through professional respect rather than a clash of judgments.
- Moderation emphasizes consensus, that is, averaging judgments is a weak alternative that leaves different interpretations and applications of standards unchallenged and unresolved; besides, negotiation allows standards to be discussed, clarified, and aligned, generating greater transparency and consistency into the future.
- Reaching for agreement encourages and strengthens professional engagement. A choice sometimes must be made between accepting assessor judgments at face value versus substituting the moderator's judgment for the assessor's. The latter may sacrifice long-term for short-term benefits and is best limited to exceptional cases such as recalcitrant assessors with whom agreement is unattainable.
- Once agreement is reached, some form of follow-up may be necessary. This might involve assessor sign-off that agreed actions have been taken. Alternatively, there might be further monitoring of those actions, such as through additional moderation. Clearly, the extent of follow-up depends on time and resources.
- What constitutes successful moderation will vary. Long-term refinement and improvement depends on asking *post hoc* whether the implemented moderation processes served their purpose and had desirable consequences. Such a review can be informal (e.g., by discussion) or formal (e.g., by systematic research).
- Little research has been conducted on moderation even though moderation practices are now widespread and becoming accepted as part of normal educational practice. More research is needed to establish a more extensive and systematic body of knowledge on the varieties of moderation practice, the rationales and circumstances of their adoption, and their effects and effectiveness.

See also: Equating and Scaling; Impact of Assessment on Classroom Practice; Portfolio Assessment; Summative Assessment by Teachers.

Bibliography

- ACACA (Australasian Curriculum, Assessment and Certification Authorities) (1999). *Principles for the Integrity, Quality and Long-Term Credibility of Certificates of Achievement*. Brisbane: ACACA.
- Daugherty, R. (1994). Quality assurance, teacher assessments and public examinations. In Harlen, W. (ed.) *Enhancing Quality in Assessment*, pp 100–115. London: Paul Chapman.
- Daugherty, R. (1997). Consistency in teachers' assessments: Defining the problem, finding the answers. *British Journal of Curriculum and Assessment* 8(1), 32–38.
- DEST (Department of Education Science and Training) (2002). *Striving for Quality: Learning, Teaching and Scholarship*. Canberra: Commonwealth of Australia.

- Harlen, W. (1994). Issues and approaches to quality assurance and quality control in assessment. In Harlen, W. (ed.) *Enhancing Quality in Assessment*, pp 11–25. London: Paul Chapman.
- Linn, R. L. (1996). Linking assessments. In Kane, M. B. and Mitchell, R. (eds.) *Implementing Performance Assessment: Promises, Problems, and Challenges*, pp 91–106. Mahwah, NJ: Erlbaum.
- Masters, G. N. and McBryde, B. (1994). *An Investigation of the Comparability of Teachers' Assessments of Student Folios*. Brisbane: Tertiary Entrance Procedures Authority (now Queensland Studies Authority).
- Maxwell, G. S. (1996). Technical issues in comparing and aggregating achievement measures for tertiary selection in Queensland. In *A Systems Analysis of Selection for Tertiary Education: Queensland as a Case Study*, PhD Thesis, pp 139–230. Brisbane: The University of Queensland.
- Maxwell, G. S. (2001). *Moderation of Assessments in Vocational Education and Training*. Brisbane: Queensland Department of Employment and Training.
- Maxwell, G. S. (2002). *Moderation of Teacher Judgments in Student Assessment*. Brisbane: Queensland School Curriculum Council (now Queensland Studies Authority).
- Maxwell, G. S. (2007). *Implications for Moderation of Proposed Changes to Senior Secondary School Syllabuses*. Brisbane: Queensland Studies Authority.
- Newton, P. (2005). Examination standards and the limits of linking. *Assessment in Education: Principles, Policy and Practice* **12**, 105–123.
- QAA (Quality Assurance Agency) (2006). *Code of Practice for the Assurance of Academic Quality and Standards in Higher Education (Section 6: Assessment of Students)*. London: QAA.
- QSA (Queensland Studies Authority) (2000–2007). *Random Sampling Reports*. Brisbane: QSA.
- QSA (Queensland Studies Authority) (2006). *Moderation Processes for Senior Certification*. Brisbane: QSA.
- Sadler, D. R. (1986). *School-Based Assessment and School Autonomy (Assessment Unit Discussion Paper 11)*. Brisbane: Queensland Board of Secondary School Studies (now Queensland Studies Authority).
- Bloxham, S. and Boyd, P. (2007). *Developing Effective Assessment in Higher Education: A Practical Guide* (Chapter 8: Moderation). Maidenhead: McGraw Hill/Open University Press.
- Booth, R., Clayton, B., House, R., and Roy, S. (2002). *Maximising Confidence in Assessment Decision-Making: Resource Kit for Assessors*. Adelaide: National Centre for Vocational Education Research.
- Hall, K. and Harding, A. (2002). Level description and teacher assessment in England: Towards a community of assessment practice. *Educational Research* **44**(10), 1–15.
- Harlen, W. (ed.) (1994). *Enhancing Quality in Assessment*. London: Paul Chapman.
- New South Wales Department of Education and Training (NSWDET) (no date). *Consistent Teacher Judgement: Moderation*. Sydney: NSWDET. www.curriculumsupport.education.nsw.gov.au.
- Orr, S. (2007). Assessment moderation: Constructing the marks and constructing the students. *Assessment and Evaluation in Higher Education* **32**(6), 645–656.
- University of Southern Queensland (USQ) (2007). *Moderation of Assessments*. Toowoomba: USQ Learning and Teaching Support Unit.
- William, D. (1992). Some technical issues in assessment: A user's guide. *British Journal of Curriculum and Assessment* **2**(3), 11–20.
- Wilson, M. (2004). Assessment, accountability and the classroom: A community of judgment. In Wilson, M. (ed.) *Towards Coherence between Classroom Assessment and Accountability: 103rd Yearbook of the National Society for the Study of Education (Part III)*, ch. 1, pp 1–19. Chicago, IL: NSSE.

Relevant Websites

- <http://www.acaca.org.au> – Australasian Curriculum, Assessment and Certification Authorities (ACACA).
- <http://www.curriculum.org> – Curriculum Services Canada (collaborative assessment and moderation).
- <http://www.nzqa.govt.nz> – New Zealand Qualifications Agency. National Qualifications Framework Moderation.
- <http://www.trainingconsultants.co.za> – QED, South Africa (moderator training program).
- <http://education.qld.gov.au> – Queensland Department of Education, Training and the Arts (Year 2 Diagnostic Net Moderation).
- <http://www.qsa.qld.edu.au> – Queensland Studies Authority (moderation processes for school-based assessment).
- <http://www.sqa.org.uk> – Scottish Qualifications Agency (assessment and quality assurance).
- <http://www.qca.org.uk> – United Kingdom Qualifications and Curriculum Authority (moderation arrangements).

Further Reading

- Adult Community and Further Education Board (1994). Moderation in the certificates of general education for adults: Guidelines for providers. *ERIC EDRS ED387700*, Melbourne.
- Banks, T. and McNeil, M. (2006). *Assessment of Learning: Local Moderation/Sharing the Standard*. Glasgow: Assessment is for Learning.

School Policies and Practices to Support Effective Classroom Assessment for Learning

D Pedder, University of Cambridge, Cambridge, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

At root, what makes assessment educational is the use of assessment information for supporting improvements in learning by teachers and their students. The classroom assessment strategies that teachers develop to support learning are commonly known as formative assessment or assessment for learning (AfL). Five broad classroom strategies commonly associated with formative assessment or assessment for learning are sharing; learning intentions and success criteria with learners so they understand criteria of quality; developing classroom talk, questioning and learning tasks that elicit evidence of learning; giving feedback that identifies strengths and weaknesses and helps learners know what to try next to move forward and improve; and providing opportunities for peer- and self-assessment so that learners can eventually develop and apply the four other strategies in this list and become autonomous learners and sources of learning support for one another. Indeed, a central underlying core principle reflected in assessment for learning is its value for promoting learning autonomy (James *et al.*, 2007). More specific practices and techniques have been developed for realizing the five broad AfL strategies listed above in different classroom contexts, and these are discussed further in Wiliam (2007).

In understanding how teachers develop and use different classroom assessment strategies and practices in different contexts, it is always important to ask how they manage to do so in ways that reflect the underlying core principles of using classroom assessment to improve the quality and autonomy of students' learning. It then becomes useful to ask whether teachers' use of the five broad AfL strategies in classroom lessons brings them to a genuinely closer appreciation and fuller understanding of their students' learning and orientations to learning to (1) support it better and (2) provide more scope for their students to exercise and develop increasing autonomy and responsibility in their learning. For example, do teachers observe their students' learning as it happens in classroom lessons? Do they focus on what their students are learning and also on how they are learning? Do they provide opportunities for their students to assess their own and one another's learning? Do they help their students to plan appropriate ways to take their learning forward? And then, do teachers widen the scope for their students to take on more responsibility in deciding or negotiating learning objectives and in assessing their own and one another's learning?

Related to these questions on how broad classroom assessment strategies can be realized by teachers in specific classroom contexts and in ways that remain true to the original precepts and values of assessment for learning, there are also questions about the scalability of effective classroom assessment. These questions relate to how AfL can be embedded beyond the practices of small pockets of committed advocates. For example, how can knowledge of effective classroom assessment be developed by a teacher for use with students in all the different classes he or she teaches? How can transfer of expertise and know-how be spread across all teachers within a school and across schools throughout entire school systems? Also, what school systems and practices support the development and transfer of such expertise? In short, on what scale can effective classroom assessment that supports improvements in learning be incorporated into classroom practice within schools and between schools throughout entire school systems?

Earlier small-scale experiments indicated powerful effects of a number of classroom assessment interventions, but the problem of scaling up these innovations across all schools and teachers remained. The Learning How to Learn (LHTL) project in the United Kingdom (see James *et al.*, 2007) is the main large-scale longitudinal study that was designed from the outset to understand and support development of the organizational conditions that would enable such scaling up of effective classroom assessment practices. There is, therefore, extensive reference to this project throughout this article.

In addressing these questions, it is important to recognize sources of complexity and challenge involved for teachers and students when asked by researchers and policy makers to incorporate assessment into their everyday classroom practices. This is the focus of the next section. In the light of these complexities and challenges, later sections consider what school practices and policies are needed to support development and embedding of effective classroom assessment practices to scale.

Sources of Complexity and Challenge in Enhancing the Use of Effective Classroom Assessment

The Classroom Environment

One source of complexity and challenge that shapes any kind of classroom practice, including classroom assessment,

is the nature of the classroom itself as a place in which teaching, learning, and assessment take place. McIntyre (2000: 98) suggests that the sustained high-quality oral and/or written interactions involved in effective formative assessment so adds to the complexity of classroom teaching as to make it a quite impracticable option for teachers. He refers to six features of this complexity: the multidimensionality, simultaneity, immediacy, unpredictability, publicness, and historical embeddedness of the demands made on teachers in classroom lessons. McIntyre (2000: 99) goes on to ask how far the advice and guidance to teachers from researchers such as Black and Wiliam (1998) and Torrance and Pryor (1998) about using formative assessment in classrooms takes sufficient account of the complexity of classroom life and of the sophisticated ways in which expert teachers have learned to work effectively in classrooms through rigorous prioritization, simplification, and intuitive decision making. Fairly low uptake of assessment for learning practices has been reported by James *et al.* (2007). Decisions by teachers not to build regular high-quality formative interactions with individuals or small groups into their classroom lessons might well be due to constraints of the classroom as an environment in which teaching and learning take place.

However, over the past 10 years, researchers have been designing research and development projects on the belief that, despite such complexities, teachers can build assessment for learning into their everyday classroom practices so long as they are supported appropriately in doing so. These researchers believe that integration of AfL or formative assessment within normal everyday classroom practice is possible because all aspects of student activities during classroom lessons carry some potential for informing teachers and students about current levels of understanding. When AfL strategies are successfully integrated into routine classroom practice, the boundaries between teaching, learning, and assessment become blurred. As James (1998: 172) argues, the aspiration is that assessment should become fully integrated with teaching and learning, and therefore part of the educational process rather than a bolt-on activity. Since then, research and development projects such as those undertaken by Black *et al.* (2003), James *et al.* (2007), Thompson and Wiliam (2008), and Wiliam (2007) have actively supported groups of teachers and schools in using and improving assessment for learning as part of their everyday practice in classrooms.

Changes in Teachers' and Students' Conceptions of their Classroom Roles

A second and related source of complexity and professional challenge in enhancing the use of effective classroom assessment relates to changes in teachers' and students' conceptions of their classroom roles. An important thread

that runs through the LHTL project data (James *et al.*, 2007) is that important shifts in the roles of teachers and students, and their conceptions of those roles, are needed if the promotion of learning autonomy through the development of AfL is to become a reality in classroom lessons. A central feature of any AfL approach is that students are actively supported by their teachers in taking increased responsibility for their own and one another's learning. This increased responsibility is encouraged when learning itself becomes established as an explicit and critical focus of classroom talk through, for example, discussing with students what to try next to improve, negotiating learning objectives with students, or supporting students to clarify their understandings of success criteria. As learning becomes more explicit and visible, learners are helped to become more conscious of the how and the why of learning as an invaluable support for developing more effective strategies and increasing learning autonomy. To make this a reality, new understandings and perspectives need to be developed among teachers and students about each other and, therefore, about the nature of learning and teaching. Furthermore, new attitudes to and practices of learning and teaching, shaped by explicit and critically reflective modes of participation, need to be acquired and developed.

This shift in emphasis in the roles and relationships of teachers and students is not trivial and further complicates teachers' work in an already complex classroom environment. Teachers are challenged to change the way they see their role: from prescribing what tasks students perform to a kind of orchestration of the learning itself (James *et al.*, 2007: 217). Similarly, there is a perspective shift involved for students as they are challenged to take on an expanded role in classroom lessons. Students as well as teachers have an explicit role in instigating classroom teaching and learning. They are not merely the objects of their teacher's behavior, but animators or co-constructors of their own effective teaching and learning processes. As such, AfL challenges teachers and students to develop conditions of trust and norms of participation that are conducive to the kinds of power-sharing that support pupils to move much closer to decision making that influences the conditions of their classroom learning.

In the next section, we move beyond the individual classroom and ask what is involved in encouraging and supporting scaling up of AfL practices within and between schools.

Fidelity and Flexibility

As more teachers adopt and adapt AfL approaches, a third source of complexity and challenge lies in striking an appropriate balance between what Thompson and Wiliam (2008) refer to as fidelity and flexibility. How, they ask, can teachers adapt AfL practices to local opportunities in different classroom contexts while at the same time

remaining true to the core principles and theory of action that underpin AfL? In their various classroom contexts, teachers need to be flexible in their development and adaptation of AfL strategies (as with any strategy they decide to work into their practice) to recognize, take advantage of, and then optimize opportunities for promoting high-quality learning.

This tension between flexibility and fidelity was evident in the reported experiences of teachers in the LHTL project (James *et al.*, 2007). From analysis of large-scale survey data from this project, three groupings of teachers' classroom assessment practices were identified and these are summarized in **Table 1**.

Measures of the value teachers placed on different classroom assessment practices were compared with their reported levels of practices. The LHTL project data show that teachers tend to place highest value on making learning explicit, and these high values were in line with similarly high levels of reported practice. However, the majority of teachers in the project (about 80%) struggled to bring practice into line with values with regard to promoting learning autonomy and performance orientation. Levels of reported practice for promoting learning autonomy were significantly behind their values, whereas levels of practice reported for performance orientation were significantly ahead of their values.

These survey findings are reflected in the interview and classroom observation evidence from the same project reported by Marshall and Drummond (2006). They report genuine difficulties faced by teachers as they attempted to transform classroom cultures through the use of AfL strategies aimed at promoting pupil autonomy. They found only about a fifth of the lessons they observed captured the spirit of AfL – assessment practices with an explicit orientation to the promotion of learning autonomy. Many teachers felt constrained by a policy context that encouraged rushed curriculum coverage, teaching to the test, and a tick-box culture.

What is clear from these different data sets is that in attempting to develop classroom assessment practices, teachers struggle to realize their values in practice with

regard to both the promotion of learning autonomy and performance orientation. At the heart of implementation is this struggle to resolve the tension, discussed by Thompson and Wiliam (2008), between fidelity to core principles and flexibility in contexts of constraint. In developing classroom assessment practices, teachers are constantly balancing and aiming for synergy between learning and performance orientations. These tensions are an important source of complexity in scaling up genuine AfL approaches to classroom assessment in classrooms, schools, and networks of schools throughout systems.

The main argument so far is that if teachers are to build genuine AfL approaches to their classroom assessment practices, changes are needed in their assessment values and their conceptions of classroom roles. Changes are also needed in the role conceptions of their students. This, together with the complexity of the classroom itself as a place in which to teach, learn, and assess, represents a set of significant professional challenges and cultural shifts for both teachers and students. Therefore, if teachers are to sustain support for high-quality learning through their adaptation of AfL in multiple school and classroom contexts, then they need to continue learning and be encouraged and supported to do so by schools with a culture and commitment to teachers' professional learning (Pedder *et al.*, 2005; Pedder and MacBeath, 2008).

The Importance of Teachers' Professional Learning for the Development of Effective Classroom Assessment

What kinds of professional learning then are most conducive to the development of assessment for learning practices in classrooms? This is a question posed by James and Pedder (2006) with reference to the LHTL project. Four main groupings of teachers' professional learning practices were identified through analysis of the LHTL survey data (Pedder *et al.*, 2005) and are presented in **Table 2**.

The most striking finding from analysis of the LHTL survey data was the importance of the inquiry group of

Table 1 Teachers' classroom assessment practices: Factors and definitions

Promoting learning autonomy	Widening scope for pupils to take on greater independence over their learning objectives and the assessment of their own and one another's learning
Making learning explicit	Eliciting, clarifying, and responding to evidence of learning; working with pupils to develop a positive learning orientation
Performance orientation	A concern to help pupils comply with performance goals prescribed by the curriculum through closed questioning and measured by marks and grades

Table 2 Teachers' professional learning practices: Factors and definitions

Inquiry	Using and responding to different sources of evidence; carrying out joint research and evaluation with colleagues
Building social capital	Learning, working, supporting, and talking with one another
Critical and responsive learning	Reflection, self-evaluation, experimentation, and responding to feedback
Valuing learning	Valuing their own and their students' learning

teacher learning practices. This group of practices reflects a range of research-informed, classroom-based approaches to collaborative teacher learning, and it is these that were most directly and powerfully associated with promoting learning autonomy and making learning explicit dimensions of classroom assessment. This association suggests that teachers' uses of and responses to different sources of evidence (from more formal research and their own enquiries) together with their collaboration with colleagues in joint research and evaluation activity are important for the development of assessment practices that lead to autonomous, independent, and active learning among their students.

Classroom-based teacher research and inquiry is not only an important strand of teachers' continuing learning and sustained critical reflection but also an important factor in helping teachers to enhance the ways they support students develop independence and autonomy in their learning. In fact, as James and Pedder (2006) argue, teachers who are prepared to engage in the risky business of problematizing their own practice, seeking evidence to evaluate to judge where change is needed and then to act on their decisions, are thus engaging in assessment for learning with respect to their own professional learning. Helping students to do the same with respect to their learning becomes less challenging because teachers are familiar with the principles and processes through inquiry into their own practices. In other words, they are well on the way to conceptualizing, developing, and valuing expanded roles for themselves and their students in teaching and learning.

Collaborative classroom-focused inquiry can take many forms. Teachers can visit one another's lessons and have focused conversations about what took place and how practice can be further refined. Teachers can consult their pupils about what they do that is helpful or not for supporting their students' learning (e.g., Rudduck and McIntyre, 2007). Research Study Lessons or Lesson Study represents a more formal approach to collaborative professional learning pioneered in Japan and developed in England by Dudley. Thompson and Wiliam (2008) report the centrality of teacher learning communities as opportunities for knowledge creation and transfer of AfL expertise and for sustained reflective practice.

A collaborative classroom inquiry orientation helps teachers develop subject content knowledge and find ways of applying it appropriately as they realize broad AfL strategies as specific and increasingly refined practices in particular subject and classroom contexts. Thompson and Wiliam (2008) argue that well-developed content knowledge is a necessary precondition if teachers are to ask good questions, interpret responses of their students, provide richly formative feedback that focuses on what students can do to improve, and adapt their teaching to developments as they unfold in lessons, based on the information they are gathering about their students' understanding of particular content.

Similar findings related to the importance of classroom-based professional learning in specific subject and classroom contexts have been replicated in a number of other research and development studies such as the King's-Medway-Oxfordshire Formative Assessment project in the United Kingdom (e.g., Black *et al.*, 2003); Classroom Assessment as a basis for Teacher Change project in the Netherlands and the United States (e.g., Dekker and Feijs, 2005); the Assessment is for Learning project in Scotland (e.g., Hutchinson and Hayward, 2005); the Keeping Learning on Track programme in United States (e.g., Thompson and Wiliam, 2008); and a number of case studies involving Canada, Denmark, England, Finland, Italy, New Zealand, Australia, and Scotland as part of the OECD study of formative assessment (e.g., Sebba, 2006).

Given the importance of collaborative, classroom-based professional learning and inquiry for fostering effective classroom assessment, it is vital that teachers sustain such learning at the center of their practice. In the next section, we turn to the question of what school policies and practices support and encourage teachers to continue learning together in these ways as a basis for advancing classroom assessment.

School Policies and Practices to Support Promotion of Effective Classroom Assessment

A guiding assumption of the LHTL project (see Pedder and MacBeath, 2008) was that schools that get to know themselves better, that approach pedagogy with a more informed grasp and critical understanding, and with an enhanced disposition to discern the integral connections between learning in classrooms, schools, and networks are best placed to support teachers and students in developing effective classroom assessment. At the heart of school policies and practices that support promotion of effective classroom assessment is school self-evaluation (e.g., MacBeath, 2006) and organizational learning. Schools that develop double-loop approaches to organizational learning adopt a critical stance to prevailing school routines and raise questions such as what are we noticing about the process of our learning as a school? Also, how can we learn more about ourselves as a learning community? Core values and assumptions are brought to the surface, and defensive routines and barriers to learning that are resistant to change are acknowledged in a climate of openness and willingness to critically reflect and respond at all levels of the school organization. The importance of organizational learning processes to promoting effective classroom assessment, and the teachers' learning that is a necessary condition for it is also recognized in Scotland's Assessment is for Learning project (e.g., Hutchinson and Hayward, 2005). However, the

organizational processes and cultures of support tend not to have been emphasized in the design of this and other research projects. The LHTL project is an exception; therefore, consideration of school policies and practices to support promotion of effective classroom assessment rely on insights from that project.

We have seen that teachers' collaborative classroom-based professional learning and inquiry embody assessment for learning processes of sustained critical reflection and self and peer assessment as a basis for reexamining and taking forward professional practices and values. Similar processes can be realized organizationally. Schools that are committed to self-evaluation through deep critical inquiry and continuous double-loop organizational learning enact assessment for learning processes and practices at the level of the institution. Schools that have well-developed systems and practices for learning critically in these ways lay the basis for development of the kinds of culture and institutional environment in which teachers become more collaborative and critically reflective in their learning and practice, while encouraging their students to do the same: to become more reflective and critical with regard to their own and one another's learning through, for example, effective classroom assessment practices.

Analysis of the LHTL survey data leads to the identification of four groupings of school leadership systems and practices as given in **Table 3** (Pedder and MacBeath, 2008).

Analysis of relationships between school practices and systems, teachers' professional learning, and classroom assessment practices through the LHTL survey data provided the basis for developing a model of the organizational conditions that foster successful promotion of assessment for learning in classrooms (see **Figure 1** below). This model shows empirical links between classroom assessment practices and the organizational conditions that support wider uptake of those practices within and across classrooms and schools (see Pedder, 2006).

A striking feature of this pattern of relationships is that organizational strategies such as developing a sense of where we are going, supporting professional development, and auditing expertise and supporting networking have an indirect influence on promoting learning autonomy and making learning explicit. In other words, school leadership practices and systems do not have a direct impact on classroom assessment practices; they are mediated by the inquiry group of professional learning practices. This points to the importance of school leaders focusing on school strategies that help teachers develop collaborative classroom-based, research-informed approaches to professional learning. The model also suggests that an emphasis on performance-oriented classroom assessment is disconnected from both professional and organizational learning.

Principal and senior leader interview accounts (as part of the LHTL database) focused on resolving tensions

Table 3 School leadership systems and practices: Factors and definitions

Deciding and acting together	Involving staff in decision making and using their professional know-how in the formulation and critical evaluation of school policy
Developing a sense of where we are going	Clear communication by senior management of a clear vision and the fostering of staff commitment to the whole school, based on good working knowledge among staff of school development priorities, which they view as relevant and useful for learning and teaching
Supporting professional development	Providing formal and informal training opportunities so that teachers, for example, can develop skills to assess their pupils' work in ways that move them on in their learning and to observe learning as it happens in the classroom
Auditing expertise and supporting networking	Information is collected on practices that staff themselves think they do effectively and on informal teacher networking in which they play an active role; teachers are supported in sharing practice with other schools through networking

between a performance-oriented policy environment and the support provided for the development of learning-oriented classroom assessment practices. Principals varied in the degree to which they embraced or critiqued this policy environment, differing in their perspectives along a continuum from compliance to subversion (see MacBeath, 2008). It was clear from school leaders' accounts how deeply embedded in thinking and practice the performance levels and attainment targets were. The prevalence of language related to targets and levels, monitoring, standardization, and consensus contrasted sharply with the absence of terms such as dialog, dissent, disagreement, or conflict. It seemed as if these forms of interaction were viewed as negatives to be avoided, and not part of a discourse among professional colleagues through which decisions about a sense of where we are going or acting together could be made.

Schools differ in the leadership approaches adopted for embedding effective classroom assessment. Swaffield and MacBeath (2006) distinguish between a structural and a cultural emphasis in the leadership strategies developed at different schools. Although all LHTL schools used both structural and cultural approaches to embedding effective classroom assessment, the conception of change described by some principals could clearly be located within a structural approach. These principals emphasized clearly

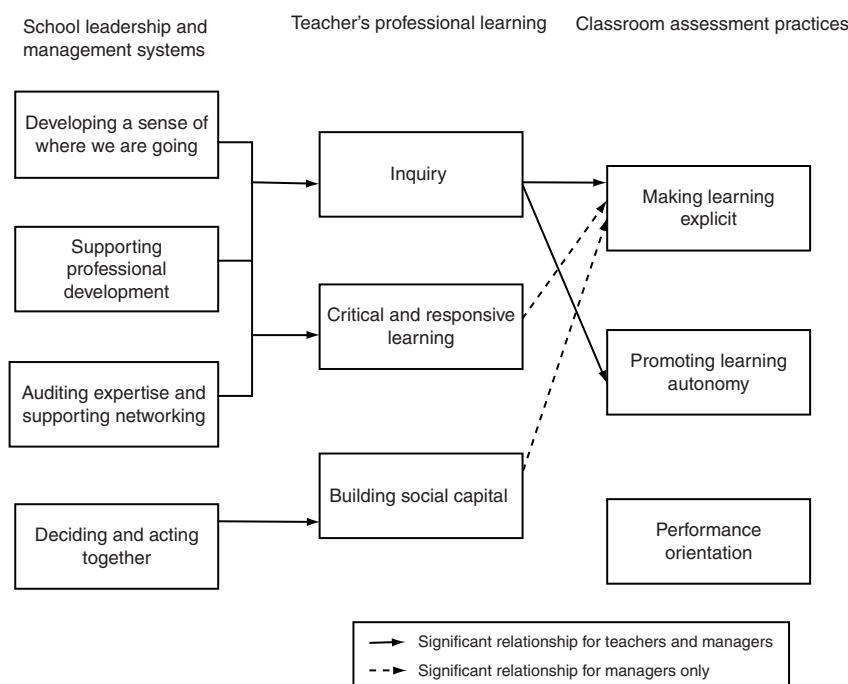


Figure 1 Organizational conditions that foster successful promotion of classroom assessment practices.

defined organizational structure with identified roles, linking mechanisms between roles and clear communication patterns. Specific tasks were identified with explicit lines of authority leading, within a traditional school hierarchy of responsibility, to a downward flow of decision-making from senior leadership to classroom teachers and classroom assistants. Other heads envisaged embedding more in terms of a cultural shift in which values and systems of belief change over time, through persuasion and support. Changing cultures for these school leaders required an extensive time frame; the use of a horticultural metaphor was common; change was portrayed as an organic growth and nurturing process. The term bedding in was used to describe a settling down of ideas and practices into new soil where they take root over time and with watchful tending. A culinary metaphor was another common evocation of processes of cultural change. Senior leaders spoke of the bubbles in the oil, assessment for learning was constantly bubbling away. Embedding was viewed as a slow simmering – a constant informal presence in the background of school life.

In practice, however, there was a great deal of common ground among schools in the adoption of strategies for embedding assessment for learning throughout the school. Committed leading teachers were used to move classroom assessment forward. Senior leaders made constant reference to the thinkers, the innovators, the change agents, and the champions as key advocates involved in fostering a climate of change. Commonly, these key personnel occupied a middle leadership role. Other schools, however, set up cross-departmental teaching and learning groups or

working parties to develop and test practical classroom assessment ideas, devise policies and feedback to senior leadership teams, and lead professional development. See MacBeath, 2008; James *et al.*, 2007 for further discussion of these aspects of leadership and organizational approaches to enhance the use of classroom assessment. What is clearly suggested from the LHTL project is that there is no simple formula and no single leadership model for promoting and embedding effective classroom assessment for learning in schools.

Conclusion

Schools that are serious about supporting teachers as they build authentic assessment for learning practices and values into their classroom routines are, at the same time, serious about promoting learning at all levels of the school organization. This is a key argument running through the research considered in this article. Teachers and school leaders as well as students go to school to learn. These schools are likewise serious about the need to support teachers and students in transforming their relationships and roles – a necessary condition if the development of classroom assessment is genuinely to be a collective endeavor and to genuinely support advances in the quality of both teachers' and students' learning. This much is clear from the LHTL project. But fundamental changes are needed, and not just in the classroom. To remain at the cutting edge of innovation in classroom assessment, and to ensure all school members have

genuine opportunities to contribute to its development and promotion, a school is invited to seek ways of broadening its organizational perspective by first asking the question, what a school needs to become to transform relationships and the learning of its members. And this question involves schools and everyone associated with them in continuously reflecting on what a school can be.

The rationale for schooling, that schools and classrooms exist to serve the learning needs of their pupils, can carry assumptions of dependent relations, which is to say that in the school community there are those who do things for others and those who have things done for them by others. And this is a common trend in performance-oriented assessment cultures. For example, it might be assumed that there is a division between teachers – those who teach pupils – and pupils – those who learn from their teachers. However, what might be involved in a richer, more expansive understanding of school relationships that goes beyond this strict division of labor between teachers (those who teach) and pupils (those who learn)? A new definition of schooling is needed in which schools recognize in each member of the school community something of value for the learning and social development of all members. In this sense, the school certainly exists for the sake of all members of the school community but not on the basis of a dependent model of service that dictates that there are professionally qualified teachers who teach and students who learn from their teachers.

New frames of school that are more in line with the transformational and collective potential of assessment for learning go beyond potential and actual divisions between agents and recipients of service – those who do and those who are done to. An alternative vision can be articulated and developed in a school on the basis of autonomy, interdependence, and human capacity for independent learning and transformation. Such a vision is highly inclusive: an all-members-as-subject vision. Research discussed in this article provides powerful evidence to suggest that assessment for learning practices and values, with their focus on the promotion of learning autonomy and making learning explicit, offer one powerful set of professional orientations and practical starting points for realizing such a vision.

However, the persistence of performance-oriented models of assessment continue to risk shrinking human agency and capacity to a narrow range of achievement, mode of development, and participation, and it can entrench teachers and pupils in traditional roles of doers and done-to which ill prepares them for active citizenship and lifelong learning. A preoccupation with league tables and test results has all too frequently led schools to narrow their organizational perspective in ways that limit scope for nurturing agency and a genuine responsiveness to the capacities of all members. Hence, the realization of the all-members-as-subject vision is frustrated.

What is it then that an assessment for learning perspective brings which helps a school move closer to this inclusive and participatory vision? At the heart of such a vision is a model of learning and achievement that belongs both individually and collectively to teachers and pupils as agents, connected to choices they make and directions they decide in their own and one another's learning. Learning is not defined for students or for teachers by the power and external constraints of someone else or by school or by some other body or authority. Predominantly students and teachers are accountable to themselves and those in their milieu for their learning and the relationships which underscore that learning. They, as subjects, take responsibility for their actions and choices with others. This is both a collective and an individual responsibility because learning practices are embedded in the lives of others and as such constitute a corporate practice in the common life of the school. But ultimately an individual's learning is something that fulfills the life of an individual agent and thus it is easy to forget that learners are not only accountable to one another but also to themselves.

Such learning and fulfillment can never be reduced to a curriculum and testing package provided or delivered to a learner by another. Nor can it be defined by school for learners. It is shaped through negotiation, dialog, conversation and consultation. There is co-construction and joint discovery by agents with other agents. Freedoms and agency flow together in joint enterprises of learning, assessment, and inquiry. In this sense, it is not the school's job to provide learning and fulfillment. It is not about someone else's responsibility for my learning. But it is the school's job to release teachers and pupils from any restrictive organizational constraints or barriers that prevent a flourishing, a shaping, and a discovery of what can be learnt and developed with others in the school community. And this is the challenge of developing school conditions conducive to the development of classroom assessment that supports genuine improvements in learning.

A school that develops such conditions then becomes a place where members are not relieved of burdens of choice, decision making, responsibility for their own and one another's learning, and the struggle to become an agent. A school does not exist to make people passive. Therefore, the school as an institution serving the needs of pupils becomes an inadequate institutional arrangement for the educational flourishing that can take place in schools. A better characterization of school centers on it as a place that fosters learning in terms of students' and teachers' agency being released – in consultation and negotiation, conflict and harmony, joint decision making and inquiry. Assessment practices that emphasize formative processes have been found to engender such a release of agency and organizational renewal in a small number of schools. Embedding and scaling up such practices organization- and system-wide remains the long-term goal. Consideration of

the research suggests that finding appropriate strategies involves collaborative experimentation among teachers and students and a great deal of adaptation to context – the kind of challenge to which the distinctive professional expertise of teachers is so well suited.

See also: Alternative Assessment; Assessment and the Regulation of Learning; Assessment in Schools – Learning to Learn; Assessment: Pre-service and In-service Teacher Education; Challenges of Developing and Implementing Formative Assessment Practices in Schools; Formative Assessment and Instructional Planning; Formative Assessment; Impact of Assessments on Classroom Practice; Peer and Self-assessment; Summative Assessment by Teachers; The Relationship Between Assessment and the Organization and Practice of Teaching.

Bibliography

- Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2003). *Assessment for Learning. Putting it into Practice*. Maidenhead: Open University Press.
- Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice* **5**, 7–75.
- Dekker, T. and Feijls, E. (2005). Scaling up strategies for change: Change in formative assessment practices. *Assessment in Education: Principles, Policy and Practice* **12**, 237–254.
- Hutchinson, C. and Hayward, L. (2005). The journey so far: Assessment for learning in Scotland *Curriculum Journal* **16**, 225–248.
- James, M., Black, P., Carmichael, P., et al. (2007). *Improving Learning How to Learn in Classrooms, Schools and Networks*. London: Routledge.
- James, M. (1998). *Using Assessment for School Improvement*. Oxford: Heinemann.
- James, M. and Pedder, D. (2006). Professional learning as a condition for assessment for learning. In Gardner, J. (ed.) *Assessment for Learning: Theory, Policy and Practice*, pp 27–43. London: Sage.
- MacBeath, J. (2006). *School Inspection and Self-Evaluation: Working with the New Relationship*. London: Routledge.
- MacBeath, J. (2008). Stories of compliance and subversion in a prescriptive policy environment. *Educational Management, Administration and Leadership* **36**, 123–148.
- Marshall, B. and Drummond, M.-J. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education* **21**, 133–149.
- McIntyre, D. (2000). Has classroom teaching served its day? In Moon, B., Ben-Peretz, M., and Brown, S. (eds.) *The Routledge International Companion to Education*, pp 83–108. London: Routledge.
- Pedder, D. (2006). Organisational conditions that foster successful classroom promotion of learning how to learn. *Research Papers in Education* **21**, 171–200.
- Pedder, D., James, M., and MacBeath, J. (2005). How teachers value and practise professional learning. *Research Papers in Education* **20**, 209–243.
- Pedder, D. and MacBeath, J. (2008). Organisational learning approaches to school leadership and management: Teachers' values and perceptions of practice. *School Effectiveness and School Improvement* **19**, 207–224.
- Rudduck, J. and McIntyre, D. (2007). *Improving Learning Through Consulting Pupils*. London: Routledge.
- Sebba, J. (2006). Policy and practice in assessment for learning: The experience of selected OECD countries. In Gardner, J. (ed.) *Assessment for Learning: Theory, Policy and Practice*, pp 185–196. London: Sage.
- Swaffield, S. and MacBeath, J. (2006). Embedding learning how to learn in school policy: The challenge for leadership. *Research Papers in Education* **21**, 201–216.
- Thompson, M. and Wiliam, D. (2008). Tight but loose: A conceptual framework for scaling up school reforms. In Wylie, E. C. (ed.) *Tight but Loose: Scaling up Teacher Professional Development in Diverse Contexts*, pp 1–44. Princeton, NJ: Educational Testing Service.
- Torrance, H. and Pryor, J. (1998). *Investigating Formative Assessment: Teaching, Learning and Assessment in the Classroom*. Buckingham: Open University Press.
- Wiliam, D. (2007). Keeping learning on track: Classroom Assessment and the Regulation of Learning. In Lester, F. K. (ed.) *Second Handbook of Research on Mathematics Teaching and Learning*, pp 1051–1098. Greenwich, CT: Information Age.

Further Reading

- ARG (2002). *Assessment for Learning: 10 Principles*. Cambridge: University of Cambridge Faculty of Education (Assessment Reform Group).
- Gardner, J., Harlen, W., Hayward, L., and Stobart, G. (2008). *Changing Assessment Practice: Process, Principles and Standards*. London: Assessment Reform Group.
- McIntyre, D., Pedder, D., and Rudduck, J. (2005). Pupil voice: Comfortable and uncomfortable learnings for teachers. *Research Papers in Education* **20**, 149–168.

Relevant Websites

- <http://www.learntolearn.ac.uk> – Learning how to Learn.
- <http://www.ncsl.org.uk> – National College for School Leadership.
- <http://www.tlrp.org> – Teaching and Learning Research Programme.

Student Assessment: Policy and Practice in Eastern Europe

G Bethell, Anglia Assessment Ltd., Stowmarket, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Backwash effect – The effect, intentional or unintentional, that testing and other assessment practices have on the behavior of teachers and learners. Positive backwash has a beneficial effect, while negative backwash is detrimental.

Formative assessment – An assessment that takes place during the teaching/learning process for the specific purpose of providing developmental feedback so that teachers and learners can adjust their plans for future learning and, hence, improve outcomes. An alternative term is assessment for learning.

High-stakes assessment – A high-stakes assessment is where individual candidates stand to gain or lose much according to their results. For example, in Armenia, young men who do well in the university selection exams win financial support for their further studies, while those who fail have to enter military service. Here, the stakes for candidates are extremely high.

National assessment – An assessment exercise carried out to determine average national educational standards in a particular curricular domain at a particular stage of education. Usually this involves testing a relatively small, but representative sample of students from the cohort.

Summative assessment – An assessment that takes place at the end of a learning activity or program for the specific purpose of judging and reporting a student's overall achievement.

Value-added modeling (VAM) – This predicts the likely educational outcome for an individual student by applying a statistical model to current or past test scores taking into account a range of background characteristics for the student concerned. By comparing actual outcomes, as measured by tests, with predicted outcomes the effects of individual schools or teachers can be estimated. Where the actual outcome exceeds the predicted outcome by a significant amount, we conclude that the school or teacher has added value.

dramatic shift. Within 10 years of the fall of the Berlin Wall in 1989, Germany was reunited, the Soviet Union and the Yugoslav Republic had collapsed, and other aligned nations in the region had, to a greater or lesser extent, broken free of restrictive ideologies and authoritarian rule. Throughout Central, Eastern, and Southern Europe, newly independent republics were free to rebuild their national identities and formerly closed states were suddenly open to the Western influences from which they had long been isolated (**Table 1**). In these countries, political parties, special-interest groups, and the wider public debated two key issues: how to generate economic growth in a competitive global market and how best to bring about desired social and political reforms. These internal voices were joined from outside by international development banks, multinational agencies, and numerous nongovernmental organizations (NGOs) – each keen to promote its own economic model or recipe for social change.

From the outset it was apparent that the reform of education would have a major role to play in both areas. If a country is to succeed in a fiercely competitive open market, the knowledge and skill levels of all young people, not just those of a small elite, must be raised and made more relevant to an increasingly technological world. At the same time, an open, civil society has far more chance of being sustained where students have the competencies and confidence necessary to question long-established orthodoxies and where they are given the opportunity to propose alternatives freely. As a consequence, policymakers across Central and Eastern Europe (CEE), the Commonwealth of Independent States (CIS), and the Baltic States sought to transform not only the structure and management of their education systems, but also the content of the curriculum and its delivery in schools. Many of the models considered were those operating in the West where, over a period of 25 years, the crucial role of assessment in measuring, monitoring, and raising educational standards had been firmly established and subsequently incorporated into sophisticated national assessment and qualification frameworks and rigorous international studies of learners' achievements.

Against this background, three main themes related to educational assessment emerged:

1. the reform of examination and qualification systems – especially at the secondary/tertiary interface;
2. the development of new national standards for educational outcomes and mechanisms for monitoring them and comparing them with international norms; and
3. the introduction of new approaches to classroom assessment in order to enhance learning.

Context

In the late 1980s and early 1990s, the European order which had prevailed since the end of World War II underwent a

Table 1 Regional groupings in the post-socialist countries of Central, Eastern, and Southern Europe

<i>Regional groupings of independent states</i>	
Commonwealth of Independent states	Armenia, Azerbaijan, Belarus, Georgia, Kazakhstan, Kyrgyzstan, Moldova, Russian Federation, Turkmenistan, Ukraine, and Uzbekistan
Central and Eastern Europe	Albania, Bosnia and Herzegovina, Bulgaria, Croatia, Montenegro, Romania, Serbia, Slovenia, Macedonia, and Turkey
Baltic states	Estonia, Latvia, and Lithuania

Reform of School-Leaving and University Selection Examinations

Assessment of students in the CEE/CIS region was, and to a great extent remains, a task delegated to schools. Teachers are required to use a variety of methods to assess students regularly throughout the academic year, but they are given much freedom in the choice of assessment tasks and the way in which they set standards. Approved methods include teacher-made tests and control works – a term from the Soviet period for classroom tasks specifically set for assessing student achievement. In addition, and in contrast to Western practices, formal oral examinations remain common throughout basic and secondary education. For each assessment, student performance has traditionally been reported on a simple numerical scale of either five or ten points. For example, on the five-point scale, 1 is considered a failing score, 2 as satisfactory, and 5 as excellent. On the ten-point scale, 4 is the minimum passing score and 10 is considered a perfect score. At the end of each year, numerical scores from all assessments in a particular subject are combined to give an overall mark. At the end, for example, of secondary education these school marks are noted on the student's diploma where results from any formal, centralized examinations may also be recorded.

Prior to 1990, any centralized examinations used in the region bore little relation to the highly sophisticated, external examinations common throughout Western Europe. Ministries of Education allocated few human, physical, or financial resources to the production and delivery of examinations. For example, in Romania, a single ministry official worked largely in isolation, preparing question papers for the baccalaureate examinations taken at the end of secondary education. The question papers were, as a rule, constructed using items from previous tests and approved textbooks. Master copies of tests were duplicated for distribution to schools where, at the appointed time, teachers would copy the questions on to blackboards for the students to answer. Teachers then evaluated (marked) the responses of their own students; however, since it was assumed that

all teachers would be familiar with the correct answers, detailed marking criteria were not provided.

An alternative model, still used in Serbia, involved the publication of all possible examination tasks in advance. This open item bank was widely available at very low cost so that students could prepare, or be prepared by teachers and tutors, in advance. On the day of the examination, the numbers of the tasks selected for testing were announced on national radio. Students wrote their answers to the chosen tasks and these were then marked and graded by their teachers.

Such examinations were extremely cheap and generally secure, but they encouraged rote learning and did not promote the development of higher skills in the classroom. In addition, because conditions for taking the test were not uniform and there was no mechanism for standardizing marking, the results were too unreliable to be used for comparing standards or selecting students for further education. For example, it was found that the results awarded by teachers to students in disadvantaged schools in remote rural areas were not dissimilar to those of students attending the most prestigious urban schools. In addition, disproportionately large numbers of students were awarded perfect scores, making it impossible for higher education institutions (HEIs) to select the most able students (Bethell and Zabulionis, 2000). As a result, university faculties set, and in some countries continue to set, their own selection examinations. Such bifurcated systems, where school leaving examinations have no formal link with university selection procedures, lead to several problems. First, the powerful backwash effect of university admissions tests tends to undermine the central position of the national curriculum in schools. Students approaching the end of secondary education direct all their efforts toward preparing for the entrance examinations of their preferred faculties. In some cases, schools simply suspend teaching programs toward the end of the year to enable students to practice for the selection tests. Where competition for university places is particularly fierce and where the stakes associated with success and failure are high, private coaching for university entrance tests becomes commonplace – and a problem for ministries of education, parents, and students. In some countries, the perception, and perhaps the reality, is that unless a student has paid for private coaching from a university tutor, their chances of success in the entrance examination is minimal. Where such tutors have the opportunity to influence selection decisions directly either as markers of written answers or as members of evaluation panels for oral examinations, prevention of corruption becomes extremely difficult and public confidence is undermined.

In order to eliminate the pressure on students (and their families) of having to sit multiple exams in the same subject at different locations, and to reduce the perceived

levels of corruption in the allocation of university places, politicians and educators sought to introduce new examination systems at the school/university interface. Two alternative approaches have been used. The first, adopted by several CEE and CIS countries, was to develop a centralized examination system based on the European baccalaureate model. In this case, the majority of school leavers take formal examinations in compulsory subjects (usually the national language and mathematics) and a number of optional or elective subjects. The results are used for two purposes: certifying achievement in a school-leaving diploma and selecting candidates for places in institutions of higher education. The second approach was to introduce a single, centrally controlled university entrance examination system to replace the multiplicity of entrance tests set by the various university faculties. Here, the general model is more closely related to the system of standardized, objective tests used in the United States and in Turkey.

The first transitional country to introduce radical examination reforms was the former Yugoslav Republic of Slovenia where, in 1989 the Ministry of Education recommended that a centrally controlled Maturity (*Matura*) examination system should be introduced to serve both as a school-leaving qualification and as a passport to university (Gabršček, 2001). To achieve this, a National Matura Commission was established to determine the principles, rules, and procedures for the new examinations. After much deliberation, the model chosen was one based on the International Baccalaureate and much influenced by the A-level examinations of England. The compulsory subjects were mother tongue, mathematics, and a foreign language. In addition, students were to choose two elective subjects from a prescribed list. In order to pass *Matura*, and hence be eligible to compete for a place in an institution of higher education, candidates had to achieve at least a minimum pass in each of the five subjects taken.

Administering high-stakes examinations on a national scale requires far more resources than operating a decentralized system where much responsibility is delegated to schools. In Slovenia, this challenge was met by establishing an independent National Examinations Centre. This is responsible for setting question papers, printing and distributing examination materials to schools, organizing the marking of student scripts, and processing the results. The costs associated with accommodating, equipping, and staffing such an organization are considerable. In most CEE/CIS countries, these costs are met from the state budget for education.

Slovenian reforms were generally considered a great success and they influenced policymakers throughout the region. The World Bank, the European Union, and other international agencies have been keen to provide financial support for similar reforms in CEE, the CIS, and the Baltic states. Typical projects have included the establishment

of new national assessment agencies, the training of assessment specialists and the development and piloting of new assessment procedures. Countries that have successfully adopted this model to date include Armenia, Estonia, Latvia, Lithuania, and Romania.

New Matura and Baccalaureate examination systems successfully address many of the educational challenges facing transition states. However, not only are they costly to operate, but it is also difficult to keep them secure because of their large scale and dispersed administration. Indeed, leakage, that is, the loss of a question paper in advance of the examination date, can be both expensive and politically damaging. In countries where competition for places in institutions of higher education is fierce and where the perception of corruption is high, maintaining the security of the examination is the first priority. In these cases, policymakers have favored the introduction of standardized university selection examinations based on the use of objective questions and automated scoring methods.

An early and extreme example of this approach can be found in Azerbaijan which was the first former Soviet republic to introduce a centralized examination built around automated systems and new technologies for admission to all institutions of higher education. The admissions testing system is conducted by the State Students Admission Commission (SSAC) which was established in 1992 under the control, not of the Ministry of Education, but of the president's office. Currently, about 90 000 students apply annually for approximately 25 000 places in HEIs. All students applying for each of the five specialties (i.e., medicine, mathematics, economics, humanities, and sociology) take their examinations on a given day. For example, in 2004 approximately 23 000 applicants for mathematics took their examinations on the same day. This required 20 variants of the question paper. For this group, results were issued within 7 days. (Compare this with Matura-type examinations which typically include constructed-response and essay questions. In this case, marking and the processing of results takes more than a month from the date of the examination.)

To achieve a high level of security both before and during the examination session, all variants of the test are produced from the large item bank that SSAC has built since its inception. The items in the bank are categorized by level of difficulty and the content/skill being tested. When a test variant is required, the information system selects items at random until the test's specification is met. No further human intervention is made to ensure secrecy. Answer sheets are marked by optical mark readers to remove all elements of subjectivity.

Azerbaijan's automated testing system meets its key objectives in that the admission system is standardized for all candidates and that corruption in processing tests is eliminated. However, critics argue that the use of objective

tests built from open item banks obviates the need for school students to develop higher skills and promotes rote learning and the use of private tuition. Indeed, it is estimated that over 80% of university applicants in Azerbaijan pay for private tutors or preparation classes at an average cost of about USD 200 (ESP, 2006).

Establishing and Monitoring Educational Standards

Formerly, the term standards was used in Eastern Europe to refer to the document setting out the content, duration, and objectives for a particular teaching program. The intended outcomes were not, in general, differentiated and so served as targets for all students. This was in contrast to the concept of educational standards prevailing in the West which, in recognition of the fact that there is a broad spectrum of student ability, described discrete levels of achievement or, in the case of England, a complete ladder of achievement targets against which the progress of each child could be judged. In the Netherlands, basic and higher levels of achievement were based not only on the judgment of experts but also on quantitative evidence gained through testing representative samples of students. Encouraged by the World Bank and other funding agencies, many transitional countries built the concept of graduated standards into their new curriculum frameworks and, most importantly, developed national assessment systems to validate and monitor the new standards – particularly in mathematics and native/state languages.

In all cases, a sample-based model was used. This was, to all intents and purposes, the model used both in international comparative studies of learner achievement and in well-established national systems such as the National Assessment of Educational Progress (NAEP) used in the United States since 1969. Conducting sample-based national assessments requires a high level of expertise in test construction, sampling, and statistical analysis. In particular, it requires the use of item response theory in calibrating items, linking tests, and developing achievement scales. Capacity in these fields was developed through technical assistance from international assessment agencies and by participating in international studies (see below).

The result of much investment is that many CEE/CIS countries now have assessment units with experience of running sophisticated national assessments and reporting the outcomes. However, the impact of these surveys is generally disappointing. Assessment specialists are frustrated that decision makers do not pay due attention to their findings and that educational policy is not influenced by the objective data that national assessments yield. One reason for this is that ministers of education

throughout the region are keen to use the results of testing to increase accountability. In particular, they would like to identify failing schools and teachers so that sanctions could be imposed. However, it is not possible to use sample-based results to produce a meaningful league table of schools or to detect inadequate teachers. These objectives can only be met by regularly testing every child and by calculating the value added by schools and teachers between stages of education as in the English system. While no CEE/CIS country has yet invested in developing such a system, there is increasing interest in using test results for accountability and some steps toward this are being taken. In the Krakow region of Poland, for example, the regional examination board is investigating the use of value-added models based on national examination results.

International Comparative Studies

On entering a more open Europe where it was intended that there should be greater freedom of movement for employment and study, many CEE/CIS countries were keen to compare their national educational standards with those of others by participating in international surveys of learner achievement. Early adopters included Bulgaria, Latvia, Lithuania, Romania, the Russian Federation, and Slovenia – all of which participated in the Third International Mathematics and Science Study (TIMSS) conducted by the International Association for the Evaluation of Educational Achievement (IEA) in 1995. This policy was strongly encouraged by international agencies and, in particular, by the World Bank which gave financial support through grants or through loans to countries embarking on education reform programs with assessment as a component. Such support allowed, for example, Moldova and Macedonia to join the next round of the renamed Trends in Mathematics and Science Study (TIMSS) in 1999.

In 2001, IEA conducted the first Progress in International Reading Literacy Study (PIRLS). This aimed to measure how well 9- and 10-year-olds read and to investigate the factors influencing reading achievement. This was seen as an important development and the CEE/CIS and Baltic countries that had taken part in TIMSS in 1999 also participated in PIRLS (Mullis *et al.*, 2003).

In a parallel development, the Organization for Economic Co-operation and Development (OECD) established its Programme for International Students Assessment (PISA) and conducted the first study in 2000. PISA looks at students nearing the end of compulsory education (15-year-olds) but, unlike TIMSS, focuses on the knowledge and skills needed in adult life rather than simple mastery of the school curriculum. Its assessment framework is based on four domains: reading literacy, mathematical literacy, scientific literacy, and the cross-curricular competence of problem solving. With its emphasis on competencies and

everyday contexts, PISA was seen by many transitional countries as being particularly relevant. As a result, Albania, Bulgaria, Latvia, Macedonia, the Russian Federation, and Romania were among the non-OECD countries that took part in a re-run of the study in 2001 (PISA+).

Prior to participation in such international studies, judgments as to national educational standards in Eastern Europe were often based on the performance of students in prestigious subject Olympiads. In these competitions, extremely gifted students demonstrated their advanced skills in, for example, mathematics and physics. The number of gold medals that a nation's students won was taken as a proxy for more general standards – unfortunately, the achievements of the few often flattered to deceive. For the first time, TIMSS and PISA allowed countries to measure with precision the average standards achieved by the complete student population and, on the basis of highly reliable data, to compare typical achievement with that of Western counterparts and regional neighbors. For most CEE/CIS countries, the results, particularly from PISA were, to say the least, disappointing. In PISA 2000, the Russian Federation and Latvia fell significantly below the OECD average, while Bulgaria, Albania, and Macedonia fell in the group of poorest performers with scores comparable to those of low-income countries in South America (OECD, 2003). In PISA 2003, Serbia's average mathematics score (432) placed it 33rd out of 40 participating countries (OECD, 2004).

Reactions to the results varied. Some countries were defensive, arguing that the unfamiliar test formats and the contexts in which the tasks were set placed their students at such a disadvantage that the results were misleading. Reactions in Albania and Serbia were particularly strong with the international results being decried and national reports being suppressed. However, in other CEE/CIS countries it has been accepted that the results of such surveys give a true picture of relative standards and hence indicate where new policies are required. Latvia, for example, has invested in major projects to raise national achievement – especially in mathematics and science.

Classroom Assessment Practices

Reform of high-stakes examinations, particularly those used for selecting candidates for universities, inevitably has the highest profile as far as politicians and the public are concerned. CEE/CIS countries have made unprecedented investments in implementing new assessment systems – the outcomes of which have often attracted media attention and resulted in controversy. This has, in many ways, overshadowed important developments in the day-to-day assessment of students in classrooms.

Prior to 1990, the classrooms of Eastern Europe were generally places where the teacher's role was to deliver the

approved curriculum in the prescribed manner to children who were expected to be passive learners. As countries became independent, old curricula were rejected and, in the interregnum, a space appeared where enthusiastic educators pressed for new approaches to teaching and learning. NGOs promoted their own programs and, in some cases, supported local initiatives for pedagogical reform. Two major themes emerged. First, there was a move toward active learning especially in kindergartens and early-years classrooms. Projects such as the Step-by-Step program, originally supported by the Soros Foundation and now maintained by a network of legacy foundations, promoted child-centered teaching with teachers and children as partners in the learning process. This and similar initiatives percolated upwards into the basic and secondary phases of education. Second, there was a reaction against the traditional emphasis on the memorization of facts and the ability to reproduce standard solutions to familiar problems. Curriculum developers and pedagogs argued for a curriculum and methodology that would encourage the development of higher-level cognitive skills, including independent critical thinking and the ability to solve problems in unfamiliar contexts – something where the PISA results of 2000 and 2003 had revealed significant deficiencies across the region (OECD, 2003).

As new outcome standards have been developed and more active approaches to teaching encouraged, complementary approaches to assessment have been proposed. These mirror recent and ongoing developments in Western education systems, and have the dual role of ensuring that teachers constantly monitor achievement through regular testing (i.e., assessment for reporting) and improving student achievement (i.e., assessment for learning).

At the primary level, there is much evidence of increased emphasis on judging performance against attainment standards defined in behavioral terms and helping young learners through increased use of diagnostic testing and remedial support. In Romania, for example, one of the first tasks of the new National Assessment and Examinations Service (established 1998) was to develop band descriptors, which described expected student performance at four levels in each subject of the curriculum. The intention was that these should serve as criteria against which teachers would judge and report the performance of each child. This was a radical departure from traditional practice in which teachers placed students on a numerical, 1–10 scale – a system which inevitably led to norm referencing with schools producing similar distributions of scores, regardless of absolute student achievement (Bethell and Mihail, 2005). In order to overcome the inevitable resistance to change, the Ministry of Education instructed all teachers of lower primary classes to report student achievement using comments, that is, statements related to the band descriptors, and to abandon the use of the 1–10 scale. This development may seem superficial to

some observers; however, in fact, it exemplifies a significant shift in approach to teaching, learning, and assessment, evidence of which can be found across the CEE/CIS region. First, it depends on a new interpretation of standards. Second, it focuses attention on the absolute rather than the relative performance of each student. Finally, assessment is seen as an integral tool in raising standards. In essence, it indicated a first, albeit small, step toward formative assessment to enhance learning.

Beyond the primary phase, as in the West, resistance to change is greater with secondary school teachers tending to be more conservative than their primary counterparts. In addition, high-stakes examinations at the end of secondary education exert a powerful backwash effect encouraging schools to focus on the use of summative tests to discriminate across the ability range in order to produce a rank order of students. Notwithstanding the above, there are initiatives to introduce new assessment procedures in many CEE/CIS countries.

One emerging issue is the perceived need of ministries to control teaching standards by requiring teachers to set regular and frequent written tests, oral examinations, assessed homework tasks, and other assessment exercises. In a throwback to earlier practices, student work on each assessment task is graded using traditional numerical scales and, at the end of the term or academic year, all results are combined using a prescribed formula. Such systems are based on repeated summative measurements (i.e., continuous assessment) and, as such, are a long way from the formative assessment they purport to promote. That having been said, many educators in the region are vocal advocates of true formative assessment as described by Black and Wiliam (1998) in *Inside the Black Box: Raising Standards through Classroom Assessment*. In particular, there is increasing emphasis on the use of tests for diagnostic purposes, especially for students making the transition from one grade to the next. In Armenia, for example, teachers have been provided with model diagnostic tests have been trained in setting similar assessment tasks and interpreting results. Progress here and elsewhere in the region is slow and limited, but this is not unexpected. Even in the relatively well-resourced classrooms of the West, high-quality formative assessment remains a distant goal and the number of teachers who, in Black's terms, have become pioneers remains limited.

Summary

Traditional approaches to student assessment, such as the value placed on oral examinations, remain strong in many parts of Central, Southern, and Eastern Europe. However, the reforms that transitional countries have put in place to make their educational systems more relevant to the modern world have incorporated significant developments in

the field of educational assessment. Perhaps the most visible signs of change are the numerous national agencies that have been established to administer high-stakes examinations. In particular, where new *Matura*, baccalaureate, or university selection examinations have been implemented, they have had a major impact on schools, teachers, students, and their families. Some of the reforms remain controversial but, at least in principle, they have increased equity by standardizing assessment procedures and setting uniform standards for all candidates.

Parallel developments, including the conduct of sample-based national assessments and participation in major international comparative studies, have yielded important objective data about the absolute and relative standards achieved by learners. At times, the results have been difficult for countries to accept, but many are now building on the lessons learned with the objective of closing the gap between their own results and those of, for example, their OECD counterparts.

Through these initiatives, educators throughout the region have been exposed to international best practice and the research upon which it is built. In particular, a much larger cadre of assessment professionals has been established – many of whom have benefited from technical assistance from some of the world's leading assessment agencies. As a result, the countries of the CEE, the CIS, and the Baltic States are now fully integrated into the international assessment community and, increasingly, are making their own contributions to the ways in which students are assessed.

See also: Assessment and the Evaluation of Institutional Effectiveness; Formative Assessment; Impact of Assessments on Classroom Practice; Summative Assessment by Teachers; The Multiple Purposes of Assessment.

Bibliography

- Bethell, G. and Mihail, R. (2005). Assessment and examinations in Romania. *Assessment in Education: Principles, Policy and Practice* 12(1), 79–96.
- Bethell, G. and Zabulionis, A. (2000). *Examination Reform in Lithuania*. Vilnius: National Examination Centre.
- Black, P. and Wiliam, D. (1998). *Inside the Black Box: Raising Standards through Classroom Assessment*. London: School of Education, King's College.
- ESP (2006). *Education in a Hidden Marketplace: Monitoring of Private Tutoring*. Budapest: Education Support Program (ESP) of the Open Society Institute.
- Gabršček, S. (2001). Access to higher education in Slovenia: Assessment reform and the re-creation of a national examination. *Assessment in Education: Principles, Policy and Practice* 8(3), 369–389.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., and Kennedy, A. M. (2003). *PIRLS 2001 International Report*. Boston, MA: International Study Center, Lynch School of Education.
- OECD (2003). *Literacy Skills for the World of Tomorrow – Further Results from PISA 2000*. Paris: Organisation for Economic Co-operation and Development.

OECD (2004). *Learning for Tomorrow's World – First Results from PISA 2003*. Paris: Organisation for Economic Co-operation and Development.

Greaney, V. and Kellaghan, T. (2008). *Assessing National Achievement Levels in Education*. Washington, DC: The World Bank.

OECD/CERI (2005). *Formative Assessment – Improving Learning in Secondary Classrooms*. Paris: Organisation for Economic Co-operation and Development/Centre for Educational Research and Innovation.

Further Reading

Braun, H. I. (2005). *Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models*. Princeton, NJ: Educational Testing Service.

Crighton, J. and West, R. (1999). Examination reform in Central and Eastern Europe: Issues and trends. *Assessment in Education: Principles, Policy and Practice* 6(2), 271–289.

Relevant Websites

<http://www.pisa.oecd.org> – OECD Programme for International Student Assessment (PISA).

<http://timss.bc.edu> – TIMSS and PIRLS International Study Centre, Lynch School of Education, Boston College, USA.

EDUCATIONAL ASSESSMENT – KEY RELATIONSHIPS

Contents

Assessment Practice in Policy Context: Latin American Countries

Impact of Assessment on Learner Groups (Boys/Girls)

Impact of Assessment on Learner Groups (Disabilities)

Impact of Assessment on Students' Test Anxiety

Impact of Assessments on Classroom Practice

Social Practices in School Assessment and their Impact on Learner Identities

Assessment Practice in Policy Context: Latin American Countries

F M Rizo, Universidad Autónoma de Aguascalientes, Aguascalientes, Mexico

© 2010 Elsevier Ltd. All rights reserved.

Background and First Steps

Latin American educational systems arose in the nineteenth century, but the region's unstable situation delayed their development, in most countries, until the mid-twentieth century. By the 1970s and 1980s, education was obligatory only through primary (elementary) school. During this time, mandated education began to extend to the lower secondary (middle school) years. Even today, however, poverty in sectors of several Latin American countries prevents children from completing their primary schooling.

Two or three decades ago, the chief concern of educational policies was access to primary and lower secondary schooling. In some countries, this continues to be the priority, but student learning levels are gaining in importance, especially since the mid-1990s. In part, this is due to the low results among Latin American countries participating in international student-achievement assessments. It is also due to the economic globalization which demands attention to the preparation of future workers and citizens needed for a competitive society and effective democracy.

The development of standardized tests of school achievement began in the early twentieth century. Across the first-half of this century, this movement built momentum to the extent that teacher evaluations began to get integrated into standardized testing. The influence of these developments was felt around the mid-century in Latin America when some teachers were asked to prepare multiple-choice tests. They were asked to standardize their students' scores – known as grading on a curve – but since no special training was provided, the quality of teacher-made tests was poor. By 1960, standardized tests were used as entrance exams to high-demand programs in higher education in some schools at Mexico's National University, and the Technological

Institute of Monterrey began to use a Spanish version of the Scholastic Aptitude Test (SAT) – prepared by the Puerto Rican College Board. In 1968, the Colombian Institute for Higher Education Promotion (ICFES) was established to develop college-entrance exams. For primary and lower secondary education, the first large-scale assessments were conducted in Chile, Mexico, and Costa Rica.

Assessment Development since the 1990s

Previous efforts to develop student-assessment systems in Latin America were characterized by poor technical quality and low continuity. Thus, prior to the 1990s, accurate assessment systems did not exist. Efforts to this end were sporadic and incomplete. It was not until the 1990s that large-scale assessments were given increased importance. At the end of the first decade of the twenty-first century, more than half of the countries in the region have stronger assessment systems. Technical and organizational improvements have been made which translate into greater continuity of test administration and higher-quality data.

Argentina

In 1993, the Ministry of Education established the National Assessment System of Educational Quality to make National Assessment Operatives (ONEs). Initially, these were administered to students in the final year of primary and lower secondary schooling. Tests evaluated math and language. Later, they were extended to other grades and subject areas. The primary school assessments were

administered until 1999. In 1997, they began universal testing of the lower secondary level, which were repeated in 1998, 1999, and 2000. In 2000, the creation of the Institute of Educational Quality was approved. The serious economic crisis that struck this country prevented plans from advancing, and the ONEs intended for 2001 were not carried out.

Following the 2000–2001 period, crisis assessment was conducted in 2002 and 2003, and it was decided to administer tests – in the future – every other year. The exams were given in 2005, 2007, and 2009, under the direction of the National Division of Information and Evaluation of Educational Quality (DINIECE). Many of the nation's 24 jurisdictions administer their own assessments. In 2003, equating techniques were introduced and, in 2005, achievement levels were defined in relation with the Common Curricular Content. In 2009, all the students in the last year of secondary education were assessed (census-based), as well as a sample of ninth grade. The country is planning to switch to a 3-year cycle, with assessments of third and sixth grade students in a year, ninth and 12th grade in another, and a third year for analysis and dissemination of results.

Bolivia

In 1996, a System of Quality Measurement was established. In 1997, language and math achievement tests were administered to children in the third and sixth grades. In 2000, tests were administered to children in the first, third, sixth, and eighth grades, as well as in the 12th grade. With these results, a baseline was established as a reference for subsequent assessments. Tests in Quechua and Aymara – two native languages spoken by a large portion of the population, were developed. Following 2005, the unit in charge of assessments was dismantled and no testing has taken place since then.

Brazil

From 1990 to 1995, student samples attending the first, third, fifth, and seventh grades were assessed. During this time, the Basic Education Assessment System (SAEB) was initiated. In 1997, this institution was encompassed by the National Institute of Educational Studies and Research (INEP), which administers the National Secondary Education Exam and the National Course Exam – a college entrance exam known as *Prova*. Until 2001, samples of the fourth-, eighth-, tenth-, and eleventh-grade students were assessed every other year. The SAEB exams were redesigned in 2001. Revised editions included achievement levels and test equating to compare results over time. Other census-based tests have also emerged since 2005. The *Prova Brasil* is administered every other year to students in the fourth and eighth grades to assess language and math. Including more than 5 million children – all

students attending public schools in urban areas – results are available by school, municipality, and state. The *Provinha Brasil* seeks to measure emergent literacy at the end of the first grade. It is administered on a voluntary basis to children in public schools, where the results are processed. Several of the 27 states, such as Sao Paulo, Minas Gerais, and Parana, have developed their own assessments.

Chile

Chile had developed a true national assessment system prior to any other Latin American country. The first tests were administered to eighth grade students in 1968. A renewed interest in these tests emerged in 1978 with the redistricting movement and a step toward privatizing several public schools. The Ministry of Education entrusted the Catholic University with the system's development. In 1982, they began to administer the Tests of Performance Evaluation (PER) to students in the fourth and eighth grades. In 1988, the Educational Quality Measurement System (SIMCE) was established, which was absorbed by the Ministry of Education in 1995. Since the 1988 high stakes, universal tests have been administered every other year, alternating the grades. All the students are assessed in language and math. Tests in natural science, history, and geography are administered to student samples. In 1993, they administered tests to students in the 11th grade, and again in 1994, 1998, 2001, and 2003.

Since 1998 the item response theory (IRT) has been used for test equating between assessment years. In 2003, a committee of experts conducted a general evaluation of the SIMCE. Based on this, the ways in which assessment results are used were adjusted, but the assessment design remained the same. Achievement levels were instated in 2005. The tests are aligned with the Core Objectives and Minimum Obligatory Contents from the National Curriculum Standards, which define Progress Maps. From 2006, the tests are administered each year to fourth-, eighth-, and tenth-grade students. The SIMCE continues to have a strong impact because results are provided by school. Attempts are made to leverage the tests' potential and to use the results to develop policies that lead to the improvement of school quality.

Colombia

In 1968, the Colombian Institute for Higher Education Promotion (ICFES) was created to develop college entrance exams. In the early 1990s important educational reforms were undertaken and an interest in large-scale assessments emerged. At this point, ICFES was entrusted with test development for students in basic education. In 1991, language and math tests began to be administered to samples of students attending the third, fifth, seventh, and ninth grades. Social and natural science assessments were then developed.

In 1997, the Quality Division of the Ministry of Education and the ICFES launched the National System to Assess Educational Quality (SABER) – which sought to evaluate a longitudinal cohort of students with annual exams. This idea was later abandoned.

In 2002–2003, the first universal tests in reading, math, natural sciences, and civic competence were administered. The second round was during the 2005–2006 school year. Since 1998, IRT and test-equating techniques have been used. Bogota has their own assessments, and there are similar efforts in other regions. In 2008, the ICFES initiated an important effort to redesign its tests. The intention is to improve the technical quality and measurement precision of different curriculum areas – combining universal and stratified student samples in a multiyear cycle.

Costa Rica

The Institute for Research and the Improvement of Costa Rican Education (IIMEC), at the University of Costa Rica, began to administer large-scale tests in 1986 – with the endorsement of the Ministry of Education. This led to the creation of the National Center of Educational Evaluation in 1997. However, because assessment efforts could not be consolidated, national assessments are now administered by the Ministry. Universal assessments, in 1986, were administered to students in the third, sixth, and ninth grades in language and math. In 1987, 1990, 1996, and 1997, some of these grades were assessed, and in other subject areas as well. Currently, tests are administered to several grades, covering an array of subject areas. All children completing the ninth grade and secondary education (high school) are administered graduate exams which include all subject areas.

Cuba

The Evaluation System of Educational Quality began its development in 1996. The Central Institute of Pedagogical Sciences (ICCP), of the Ministry of Education, oversees the Quality Assessment Administrations. There are also provincial and local assessment units. From 1996 to 1997, language and math tests were administered to children attending the third grade in 100 schools. In 1999, sixth-grade students were added to the sample. Later, grades and the sample size have expanded, and the areas of natural science and history were added.

In recent years, some grades of technical/professional schooling as well as teacher training programs are being assessed in Cuba. These subsample assessments include reviews of math and language notebooks. In 2008, it was also planned to conduct assessments of special education. Achievement levels are used to report the results and produce briefs for several teachers, school administrators, and provincial supervisors. In addition to student

achievement, a socioeconomic indicator and other indices specify system efficiency, teacher performance, and school inspection.

Dominican Republic

Assessments began in 1991 with tests for fourth-grade students. In 1992, exams were administered to eighth-grade students, and to individuals attending a third term of adult education. In 1993, the last year of secondary education was included. From their inception, the tests are universal. Since 1994, they are administered annually to all students completing their final year of basic (eighth grade) and secondary (12th grade) education. The results of each student are combined with school scores to decide if the student will be permitted to continue to the following educational level. In 2009, the country is planning to create a new institution to develop a more comprehensive assessment system.

Ecuador

From 1996 to 2000, basic education assessment was carried out using tests developed by the National System for Academic Achievement Measurement, known as APRENDO. This was conducted by an external group of evaluators funded by the World Bank. They administered language and math tests to student samples attending third, seventh, and tenth grades. From 2001, the responsibility of testing was passed to the Ministry of Education, who continued to administer the same tests to the same grades. The APRENDO tests used the Classic Theory and equating methods of results over time. These tests were made public following each administration. In 2006, it was decided to substitute these tests with the National Assessment System – headed by the Planning Division within the Ministry of Education. Every 3 years, universal reading and math assessments will be administered in the fourth, seventh, tenth, and 12th grade. Natural and social science tests will be administered to samples in the seventh and tenth grade using a matrix design and IRT procedures.

El Salvador

The first tests – known as SABE – were administered from 1993 to 1996 as part of a project designed to strengthen achievement in basic education. Each year, language, math, natural science, and social science tests were administered to a national sample of students attending third through sixth grades and ninth grade. In 2001, the National Monitoring and Evaluation Division launched the National Learning Assessment System (SINEA) – which administers several achievement tests in various primary and lower secondary grades. In 2001, 2003, and 2006, the tests were

administered to student samples, and, in 2005 and 2007, they were applied universally. These tests contain defined achievement levels. The SINEA assumed the administration of the Learning and Aptitude Tests for Secondary School Graduates which, since 1997, were administered by the Simeon Canes University, and required to graduate from secondary schooling all over the country. Test scores contributed 20% of secondary graduates' final grades in 2005. In 2007, their weight increased to 25%.

Guatemala

From 1992 to 1996, the National Testing Center (CENPRE) of the Ministry of Education – with support of the US Agency for International Development – carried out their first assessments. In 1997, the University of the Valley of Guatemala assumed the tasks of the CENPRE and created the National System of Academic Achievement Measurement – later renamed the National Assessment Program of School Performance (PRONERE). From 1992 to 1996, it assessed children attending the third grade, and, from 1997 to 2001, children in the sixth grade. In 2004 – using a test called *Graduandos* – the PRONERE undertook new assessments of other primary grades, as well as of students finishing secondary schooling. In 2004, the University of the Valley administered *Graduandos*, for the first time, to students finishing secondary school. While results do not influence students' final grades, all students are required to take the test. In 2005, the Ministry of Education established the National Assessment and Educational Research System, within the General Division of Evaluation, Research and Educational Standards. The most recent tests include achievement levels, and seek to assess high-level skills; they are developed in Spanish and four native languages; results are available for all the schools – ranked by results considering students' socioeconomic status.

Honduras

From 1990 to 1994, the first large-scale assessments took place. This was part of the Program for Educational Quality by the National Pedagogical University, with support from United States Agency for International Development (USAID). In 1995, the Measurement Unit of Educational Quality was created by the National Pedagogical University. This office assumed assessment activities as assigned by the Ministry of Education. In 1997, the first test was given in Spanish and math to third- and sixth-grade students. In 1998, the second iteration of tests in the same subject areas was carried out, extended to second-, third-, fourth-, and sixth-grade students. Grades two through five were assessed in 1999 and grades three and six in 2000, 2002, and 2004.

Mexico

The Secretariat of Education launched initiatives, in the 1970s, which led to the creation of the Assessment Division. The first achievement tests were administered in 1972 as entrance exams to secondary schooling. By the end of the decade, tests were administered to the first samples of primary school students. Qualitative advances in testing administration did not occur until the early 1990s. In relation to an economic incentive program for teachers, called *Carrera Magisterial*, large-scale assessments were administered to students attending the last four grades of primary school, and the 3 years of lower secondary. The first administration was in 1993 to more than 4 million students. In 2005, the number approached 8 million. In 1991 – with support of the World Bank – the Program to Combat Educational Setback (PARE) was developed in four poor states, with student assessments. Similar programs continued, always with a student-assessment component. In 1996, an initiative to define curriculum standards was undertaken. Assessment instruments were developed, called the National Standards Tests. These were first administered in 1998. Until 2004, they were administered annually to national samples of primary school students. Since 2000, they were given to lower secondary school students as well.

In 2002, the National Institute for Educational Evaluation (INEE) was created, to develop independent assessments of the country's educational system. These tests have high technical quality and a wide distribution of the results – which contrasts with the previous practice of refraining to share test results publicly. As of 2004, the INEE developed Tests of Quality and Educational Achievement (EXCALE) using a matrix design. These cover the core curricular areas. EXCALE tests use multiple-choice and open-response items, and are administered to representative samples of students from all states within a cycle of four school years – successively assessing students in the ninth, third, and sixth grades, and the final preschool year (kindergarten). In 2006, the Secretariat of Education began to administer additional census-based exams (National Tests of Educational Achievement in School Centers (ENLACE)) during the four final grades of primary school and the last grade of lower secondary. In total, every year, more than 10 million students are tested. In 2008, students attending their final year of secondary school were tested for the first time.

Nicaragua

Nicaragua participated in a project by the Organization of Iberoamerican States (OEI) and the Central American Cultural and Educational Coordination (CECC) to establish standards for primary education – with an assessment component. In 1996, the Assessment Division of the

Ministry of Education began to administer tests from the APRENDE project. In 1999, the National Assessment System for secondary and basic education was created with support of USAID and United Nations Educational, Scientific, and Cultural Organization (UNESCO). Language and math tests are administered to samples of third- and sixth-grade students. The last administrations reported are from 2002 and 2006.

Panamá

Large-scale assessments can be traced to 1985. At the time, tests were administered to a national sample in a project led by the Organization of American States (OEA). In 1988 and 1992, the Ministry of Education carried out additional testing. In 1996, a National Assessment System for Educational Quality (SINECE) was established; samples of students of the third, sixth, and ninth grades were tested, and again in 1998, 2000, and 2001. In 2002, the National Direction for Educational Assessment was created and the National Assessment System was redefined as one for Learning Quality (SINECA). The methodological quality has been improved; in 2005, samples of students of third, sixth, and ninth grades were tested on Spanish, math, and natural and social sciences, and students at 12th grade on Spanish, math, and English. The first results were released on July 2006; later that year, a full report has been published – with analyses of the answers by content area and their association with contextual factors. A Plan for the Improvement of Learning has been defined.

Paraguay

In 1995, the Ministry of Education created the National Assessment System of Educational – process which administers exams to student samples. The first tests were given, in 1996, in language and math to sixth-grade students. In 1997, they were administered to third- and ninth-grade students, and, in 1998, to sixth- and 12th-grade students. From 1999 to 2006, students of one school grade were assessed each year, including grades third, sixth, ninth, and twelfth. Social science, natural science, and writing have been added. In 2001, an assessment of schools at risk was conducted.

Peru

The Quality Measurement Unit (UMC) was created within the Ministry of Education in 1996. Since then, six assessments have been conducted. The first four – from 1996 to 2004 – used student samples, while the final two were universal assessments. In 2006, universal assessments began with second-grade students. However, only 55% of schools and 44% of students were covered. In 2007, 91%

of schools and 80% of students participated in the testing, which included fourth grade students as well as those attending Quechua- and Aymara-speaking Intercultural Bilingual Education programs. In 2007, a longitudinal project began to follow a cohort beginning primary school until they finish the sixth grade in 2012.

Uruguay

The National Administration of Public Education (ANEP) established the Measurement Unit of Educational Results (UMRE) in 1995. In 1996, assessments were conducted for the first time. Universal administrations in language and math were given to sixth-grade students. Teachers administered these exams. Similar universal assessments were administered to sixth-grade students in 1999, 2001, and 2005. In 1998, third-grade students were also assessed, and, in 1999, ninth-grade students were tested. In 2001, linguistic competence and math were assessed at the start of primary school – using tests administered by teachers in kindergarten, and the first and second grades. This sort of assessment was repeated in 2007 among students attending the first grade. A universal, voluntary assessment was also administered to sixth-grade students.

In 2008, natural science content was added to previous assessments. Universal tests continue to be administered by teachers in each school. A control sample is used for test equating and research. The focus that distinguishes the work of the ‘Measurement unit’ remains in force; that is, efforts are made so that test results benefit teachers in all schools of the country. They are used in important training efforts. Evaluation results and test materials are counted as resources to improve teaching practice. The tests include multiple-choice and open-response items. A project was debated in 2008 to transform the Measurement Unit into an Evaluation Institute with greater autonomy.

Venezuela

The National Learning Assessment System was created by the Ministry of Education at the end of the 1990s. In 1998, a language and math exam was administered to a national sample of sixth-grade students. The instruments were developed with support from the Central University of Venezuela. This experience did not have continuity and, until 2008, no new assessments have been conducted.

Regional Participation in International Assessments

The First Study of the Latin American Assessment Laboratory of Educational Quality (LLECE) was promoted by UNESCO’s Latin American and Caribbean Regional Office (OREALC) in 1997. Argentina, Bolivia, Brazil,

Chile, Colombia, Cuba, Dominican Republic, Honduras, Mexico, Paraguay, and Venezuela participated in this project. Argentina, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Mexico, Nicaragua, Panama, Paraguay, Peru, and Uruguay participated in the Second Study in 2006.

Increased interest in assessment is reflected in the growing participation in other international projects as well, particularly the Programme for International Student Assessment (PISA) studies by the Organization for Economic Cooperation and Development (OECD). In its first testing cycle, only Mexico – which was already a member of the OECD – and Brazil participated. A follow-up administration (PISA Plus) added Argentina, Chile, and Peru. These three countries did not participate in 2003, but Uruguay did – along with Mexico and Brazil. Argentina and Chile were integrated again in 2006, as was Colombia. In 2009, several more Latin American countries plan to participate.

Among Latin American nations, Chile has participated most in assessments conducted by the International Association for the Assessment of Educational Achievement: the Six Subjects Study in 1971, Institute of Advanced Legal Studies (IALS) in 1998, Trends in International Mathematics and Science Study (TIMSS) in 1998 and 2003, and Progress in International Reading Literacy Study (PIRLS) and the Civic Education Study (CivEd) in 2001. Colombia participated in TIMSS in 1995 and 2000, as well as in PIRLS and CivEd in 2001. Brazil participated in TIMSS in 2003. The IEA conducted a special administration of the civic education study for six Latin American countries in 2008. Regional cooperation has developed in relation to student assessment. Responsible parties for the administration of LLECE's Second Study met regularly from 2004 to 2008. They are now preparing the Third Study. The six countries participating in PISA 2006 have met regularly – together with Spain and Portugal – to prepare a joint report, published in 2009. This group supports countries participating in PISA for the first time.

Conclusion: Promises and Perils

Despite a late start and a shaky psychometric tradition, assessment systems in Latin America are consolidating and learning from international experience. Systems in Chile, Mexico, Brazil, and Uruguay already provide their educational authorities with information valuable to policymaking. Technical advances have been made in Argentina, Colombia, Ecuador, El Salvador, and Guatemala. Cuba's system – although very different from the others – demonstrated surprising efficiency according to results from the two regional assessment projects. Growing regional cooperation can contribute to the strengthening of the weakest systems.

On the other hand, two considerable perils are detected on the assessment horizon. The first – which is long standing – is the lack of continuity brought by political instability; its destructive effects are verified in decades passed. The second – a new danger to the region – is diminished educational quality as a by-product of inadequate understanding of the reaches and limitations of large-scale assessment. Indeed, the recent trend to conduct census-based assessments can have positive effects, but also negatives. There are no empirical studies on the subject, but qualified observers point out that the wrong practices of teaching to the test and poor instruction are increasingly present in several Latin American countries, following the implementation of census-based assessments with public results in the form of rankings of schools.

It was only a decade ago that few assessments existed in Latin America and, when administered, they produced results that were rarely distributed to the public. In sharp contrast, at present, testing has proliferated to the extent that it begins to tire the schools. Mass media fervently debates results – often with little knowledge of assessment complexity – and education authorities tend to take decisions based more on fads and political pressures than on solid knowledge.

In order for the promise of large-scale assessment to be realized, an objective vision of its real potential must prevail over superficial notions. It is essential to build teachers' capacity to do good classroom formative assessment, and to avoid the risk of poor replication of large-scale tests. Well-balanced assessment systems are needed.

See also: Classroom Assessment in Policy Context (French Sources); Formative Assessment; Impact of Assessment on Classroom Practice; National Assessments; The Relationship Between Assessment and the Organization and Practice of Teaching.

Further Reading

- Carnoy, M., Gove, A. K., and Marshall, J. H. (2007). *Cuba's Academic Advantage. Why Students in Cuba do Better in School*. Stanford, CA: Stanford University Press.
- Corvalán, J. and McMeekin, R. W. (2006). *Accountability educacional: Posibilidades y desafíos para América Latina a partir de la experiencia internacional*. Santiago: PREAL-CIDE.
- Ferrer, G. (2006). *Educational Assessment Systems in Latin America: Current Practice and Future Challenges*. Washington, DC: Programa de Promoción de la Reforma Educativa en América Latina y el Caribe, PREAL.
- González Jiménez, D. A., et al. (2008). *Evaluaciones externas en América Latina. Caracterización de los procesos logísticos utilizados para la aplicación de pruebas censales y muestrales en Cuba, Chile, Argentina, México, Uruguay y Colombia*, 6 vols. Cali-Bogotá: Universidad de San Buenaventura-ICFES.
- Grupo de Trabajo sobre Estándares y Evaluación (GTEE) (2007–2009). *Observatorio Regional de Políticas de Evaluación Educativa. Boletines 1 a 10*. Lima: PREAL.
- Grupo Iberoamericano de PISA (2009). *Iberoamérica en PISA 2006. Informe regional*. Madrid: Santillana.

- Horn, R., Wolff, L., and Vélez, E. (1991). *Developing Educational Assessment Systems in Latin America: A Review of Issues and Recent Experience*. Washington, DC: World Bank.
- Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) (1998). *Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados, para alumnos del tercer y cuarto grado de la educación básica*. Santiago: UNESCO-OREALC.
- Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) (2008). *Los aprendizajes de los estudiantes de América Latina y el Caribe. Primer Reporte. Segundo Estudio Regional Comparativo y Explicativo*. Santiago: UNESCO-OREALC.
- McMeekin, R. W. (1998). *Estadísticas educativas en América Latina y el Caribe. Informe técnico*. Washington, DC: Interamerican Development Bank.
- Popham, W. J. (2003). *Test Better, Teach Better. The Instructional Role of Assessment*. Washington, DC: Association for Supervision and Curriculum Development.
- Shepard, L. A. (2006). Classroom assessment. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 623–646. Westport, CT: Praeger.
- Stiggins, R. (2008). *Assessment Manifesto: A Call for the Development of Balanced Assessment Systems*. Portland: ETS Assessment Training Institute.
- Tiana Ferrer, A. (2000). *Cooperación internacional en evaluación de la educación en América Latina y el Caribe. Análisis de la situación y propuestas de actuación*. Washington, DC: Interamerican Development Bank.
- Torres Fernández, P. (2008). Evaluación educativa: los estudios internacionales y su influencia en Cuba. In *IV Jornadas de cooperación educativa iberoamericana sobre evaluación*. Antigua, Guatemala: Agencia Española de Cooperación Internacional.
- Valeirón, J. and Esquivel, J. M. (2006). *Situación de la Evaluación y la Investigación Educativa en los Países Centroamericanos y la República Dominicana*. San José: Coordinación Educativa y Cultural Centroamericana y Agencia Española de Cooperación Internacional.
- Wolff, L. (2004). Educational assessments in Latin America: The state of the art. *Applied Psychology: An International Review* **53**(2), 192–214.
- Wolff, L. (2006). Las evaluaciones educacionales en América Latina: Avance actual y futuros desafíos. In Arregui, P. (ed.) *Sobre estándares y evaluaciones en América Latina*. Santiago: PREAL.

Relevant Website

<http://www.preal.cl> – Programa de Promoción de la Reforma Educativa de América Latina y el Caribe, PREAL.

Impact of Assessment on Learner Groups (Boys/Girls)

J Elwood, Queen's University Belfast, Belfast, Northern Ireland, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Assessment – An umbrella term that covers any practice that is used to make a judgment or evaluation of a student's learning. This judgment may be formative or summative and be made with the aid of tests, orals, practicals, portfolios, tasks, examinations, coursework, etc.

Bias – A term which indicates test items (or whole tests) that are discrepant and that may affect one group of test takers more adversely than another.

Differential performance – Different groups of test takers get different score levels on the same assessment either because of bias in the test or because of real differences in performance that may be the result of differing access to learning, resources, and/or educational opportunities.

Formative assessment – The assessment that is continuous throughout a course of study and through feedback supports the learning process.

Summative assessment – The assessment at the end of a course or program of study which, usually through a single grade or mark, describes the achievement obtained.

Validity – The extent to which an assessment dependably reflects the achievement being measured and the use to which the result is put.

Introduction

Research that is interested in the impact of assessment on learner groups of boys and girls considers differing views that promote various theoretical positions that help frame explanations and account for different performances by boys and girls on assessment tasks and tests.

At the core of these differing positions is how learning and the learner are viewed, how best to assess learning, and how assessment items and tasks interact with what the learner knows and understands. Moreover, how the learner interprets what is required from the task itself is also key, as well as those aspects of assessment tasks that may influence boys' and girls' understandings of what are necessary for a successful response. Furthermore, research exploring the social impact of assessment on learner groups is fundamentally concerned with the

role that gender plays in contributing to any differences in performance observed.

This article presents an overview of key research from this field that has a bearing on how we understand the complex range of influences and impacts that assessment has on the learning and achievements of boys and girls.

Assessment and Learning – Identifying Links

Current research in the area of assessment concerns itself with links that exist between our understanding of how children learn and assessment practice, with a focus on how the mode of assessment chosen aligns to a particular model of the learner and how the practice of assessment itself can be used as a tool to improve learning (see CERI, 2005; Wilson, 2004). In earlier investigative approaches, assessment researchers were less clear in their articulation of the model of learning that underpinned the assessment technique chosen. However, more recent studies have shown that one cannot consider systems of assessment and testing without acknowledging their relationship to the model of the learner they promote or wish to promote. Furthermore, there is a connection between learning and assessment within which the notion of gender in all its complexities needs to be considered. This is because how gender mediates learning has a profound impact on how students and teachers experience assessment practice and its influence on educational success.

Views of Learning and Links to Assessment

Within the literature on how children learn, there are many theories presented that consider what learning is and how it occurs (Murphy, 1998). These theories are varied, change in prominence over time, and are often contested. In reviewing these many theories of learning, writers tend to categorize certain types of theories and/or approaches to learning into larger groupings or families to help understand key themes and influences. For example, some categorizations bring together theories of behaviorism (where notions of learning are fixed and learning happens through transmission), theories of constructivism or social constructivism (where learning is seen as a social activity and learning happens through social interaction), and theories of socioculturalism (where learning is seen as something that happens within the social, cultural, and

historical contexts of students' lives) (Elwood, 2006; Murphy, 1998). In relation to work on assessment, different categories of learning have been associated with different approaches to assessment and how students' learning is evaluated. For example, assessment located within a traditional psychometric model is seen as something done to an individual to measure that individual's learning (the behaviorist influence), with underlying assumptions about the presence of psychological attributes that can be observed and evaluated through responses to test items.

More recent considerations of assessment practice see it as a social activity and promote assessment as something done with the learner, acknowledging that assessment is measuring learning that is acquired through the social mediation of teacher and students (the constructivist and social constructivist influence). Assessment within this tradition still attempts to check up on what the individual has learned even though they have acquired that learning through social interaction. Researchers who place emphasis on the complex interactions of gender and assessment, especially within classroom-based assessment tasks and practices, prioritize the cultural and social contexts within which assessment takes place and consider the gendered relationships between teacher and student, and student with student, as key in comprehending the impact and interaction of all this on girls' and boys' assessment outcomes (the sociocultural influence). Thus, how we understand how students learn has major implications for how we choose to assess them, how we understand their responses to assessment tasks, and how we consider gender in relation to the outcomes observed.

Assessment and Boys and Girls: Sex Group Differences and Fixed Notions of Learning

Assessment within a traditional psychometric model considers learning as an individual activity and something that the student acquires through instruction; learning gets stored within the student, and the act of assessment is to measure this stored learning and to check up on what a student knows. These views of the learner and how to assess them is aligned to a static view of gender, that is, as sex group (male or female). Thus, gender is seen as a fixed, dichotomous variable against which assessment scores can be analyzed and reported.

From the 1970s, much more educational research began to focus on, and take account of, sex differences in educational achievements. Within the field of assessment and testing, much of the early research into differences between boys' and girls' achievements on tests and assessments considered sex group of the candidate as a key variable in helping to understand differential achievement

and discrimination between learner groups in education more generally (see Gipps and Murphy, 1994; Willingham and Cole, 1997). The analysis of test data from a sex-group perspective has been helpful in the field of research into boys' and girls' achievements. The publication of national and international assessment results by sex group, alongside more detailed analysis of large-scale international assessment programs (OECD, 2004) and research evidence from other studies focusing on sex differences in performance on tests (Gipps and Murphy, 1994; Willingham and Cole, 1997) have shown distinct patterns of performance between males and females which are seen across different subjects, different ages, and different testing situations. To illustrate this point, a brief overview of male and female performance across large-scale assessment programs is given below.

Sex Differences on International Assessment Programs

In international tests of achievement, distinct patterns of performance for males and females across the subjects of language (native), math, and science have been identified. For example, in the assessment of English language across a number of studies, females tend to perform better than males in all main aspects of the subject, especially in reading and writing (Gipps and Murphy, 1994; OECD, 2004). The gaps in performance between males and females tend to show themselves in the beginning of primary school and continue to grow until females perform better than males to a significant degree by the end of compulsory schooling. In math, many international and cross-national surveys (OECD, 2004) show that on average males and females in the earlier stages of schooling perform similarly in math, but as age increases, males generally tend to outperform females, and by age 15/16, males achieve better performances in virtually all aspects of mathematics tested. In science, evidence from large-scale assessment programs at both cross-national and international level show that males perform better than females in science, but that the gaps in science are the smallest across the three subject areas (OECD, 2004).

Published data of male and female achievement on large-scale assessments has improved over the last three decades to the extent that we now have extensive datasets that provide information on how males and females achieve within subjects across a range of test types. Research in the field of gender and assessment has greatly benefited from this data, especially as much of it is disaggregated at a number of levels – the test as a whole, the test paper, the test item, and mode of response. Thus, we are better informed as to how males and females perform on various types of tests and assessments. These data have certainly enhanced our knowledge of differential performance as it allows us to monitor the results of different

groups across a range of tests and assessments and enables us to pursue questions of fairness and equity at a macrolevel.

Explanations for Differences

Reasons put forward to explain why such differences occur are often situated within a technical discourse of test development (e.g., test bias) and/or within a psychological cognitive discourse which positions psychological traits or characteristics of the male or female test taker as explanatory variables for the differences observed. In relation to test bias, the explanations presented are not unproblematic. This is mainly due to contradictions in definition, usage, and understanding of the term in relation to test and/or item characteristics as opposed to genuine group differences in test performance (see Goldstein (1996) for a review). While statistical procedures can be used to identify discrepant items in tests that can then be discarded in the aim of creating fair tests, such procedures tell us only about the difficulty levels of the items for different groups and tell us little about the construct being assessed (Gipps and Murphy, 1994). Items tend to have different meanings for different groups and items that do not fit a common pattern of response may simply be assessing a different attribute.

In relation to cognitive factors, many studies have addressed a variety of psychological characteristics as explanations for sex-group differences in educational achievement (Hyde, 2005). For example, the area of learning or cognitive styles has had varying popularity over the years as a way of explaining sex-related differences in performance. Research within this area has often positioned learning styles as fixed characteristics of learners that are innate (i.e., that boys or girls have particular learning styles) and that learning styles are responsible for how boys and girls respond to different assessment items and/or tasks and that curriculum and/or instruction should be adjusted accordingly to reflect these learning styles. However, Head (1996) shows quite clearly that learning styles are not fixed attributes and that individuals may show some flexibility in alternating between styles, adjusting them to cope with the situation in which the learning is being evaluated. What seems to have become more important over time is not so much that boys and girls have particular learning styles, but that they choose different learning styles and adopt different coping behaviors in learning situations which affect their opportunity to learn and the ability to show this learning to good effect.

Limitations of the Sex-Group Approach

While there are benefits to the types of data and research that have developed around treating gender as a sex-group variable there are also limitations to such approaches. One

such limitation is that the publication of sex-group differences in achievement on tests and assessments seems to attract little disagreement about the variable's dichotomous nature or that it is telling us something important about the test taker as well as the test itself. However, such data can only tell us what is happening in relation to male and female differences in performance, it can offer no explanation as to why such differences in performance occur. Thus, it becomes difficult to extract whether the differences observed are an artifact of the test item or whether there is fundamentally something more important going on in the educational experiences of boys and girls more generally.

A further limitation is that such an approach often accentuates differences that are small or of little significance. An alternative view is that there are far more similarities between boys' and girls' performances than differences. Hyde (2005: 581) has positioned this argument as the gender-similarities hypothesis where the notion of difference presents a negative and self-fulfilling model, whereas the notion of similarities positions males and females as performing similarly across most, but not all, psychological attributes, with any differences (of sizeable magnitude) being small in number. This hypothesis fits with a wide range of data for male and female performances in educational tests which show much more overlap between male and female performance with bigger differences (and variations in performance) occurring within female and within male groups. This also leads to consideration of variables other than (or as well as) sex of the test taker (such as ethnic group and/or social background) as having as much impact on learner groups as does their sex.

The traditional view of learning and assessment and sex group considered above is no longer helpful in understanding the complexities around gender and performance on tests and assessments. If we advocate that learning does not take place in isolation and that assessments are not socially neutral then we are forced to look beyond such fixed notions of these concepts to better understand gender differences in performance.

Assessment and Boys and Girls: Social Considerations of Gender, Learning, and Assessment Tasks

There is a considerable body of research in the area of gender and assessment that has taken seriously the limitations of the gender as sex-group-variable approach outlined above and has developed an alternative framework within which to consider the impact of assessments on learner groups (Elwood, 2006). In rejecting the notion of learning and assessment as fixed concepts, research within this alternative arena positions assessment and learning as

social activities that are socially generated and mediated. Furthermore, gender is seen less as a dichotomous variable against which results can be reported and more as a socially constructed category that has a much more complicated relationship to, and with, the practices of assessment; the consequences of which are manifested in differences in educational outcomes observed between boys and girls.

Influential movements within the assessment field over the last decade have advocated a reconsideration of the use and purpose of summative assessment, promoted the use of formative (or classroom-based) assessment to enhance students' learning, and championed a view of the learner that fits within a social constructivist theoretical position (James, 2006). Here, assessment is positioned as something that is done with the student to illicit clarification of students' current knowledge and understanding so they know where to go next in their learning. New forms of assessment (performance assessments, portfolios, coursework, and extended response items) have tended to reflect these shifting views of learning and achievement and have aimed to bring tests and classroom assessments into alignment. A consequence of this is that research has been able to focus on how the techniques and structures of assessment systems impact on learner groups and how different assessment and learning activities and processes contribute to differences in achievement.

Assessment Technique and Gender Differences in Performance

Research has identified several factors in the design, structure, and use of tests and assessments that contribute significantly to differences in performance between boys and girls (Elwood, 2005). Factors identified include: type and mode of item response; teacher-assessed components of qualifications (known as coursework or school-based assessment); different levels of assessment with restricted grade ranges within the same qualification (known as tiering); the use of real-life contexts in assessment items and tasks; and the sampling of subject content on test papers (see Gipps and Murphy, 1994; Willingham and Cole, 1997 for a detailed review).

For example, if we take the factor of teacher-assessed components of qualifications (i.e., coursework) there is a widespread perception (especially in the UK) that while such components generally benefit pupils, it is girls who gain from them most. Research (see Elwood, 2005) looked at whether teacher-assessed coursework accounts for the advantage that girls demonstrate over boys in end-of-school qualifications. In relation to differences in mean marks obtained on coursework components, it was found that in all cases, girls' mean coursework marks were higher than those of boys, and that these differences were statistically significant. However, teacher-assessed

coursework marks for girls tended to be slightly more bunched than those awarded to boys; there was also less variation in the marks awarded to girls than those awarded to boys. This bunching effect of coursework marks provided an interesting paradox; it indicated that coursework actually contributed more to the final outcomes of boys than it did for girls. If coursework marks are bunched and examination component marks spread more widely, then it is the examination component marks which determine the students' rank order and hence the final overall grade. Thus, the research showed that coursework had less of an influence on the final grades awarded to girls than to those awarded to boys.

Hence, different assessment components do not operate in practice in the same way for boys as they do for girls. This is an important point given the different attention paid by girls and boys to different assessment tasks, and teachers' perceptions of how different types of assessment contribute to the final success of boys and girls. The validity of different assessment modes is thus compromised by them operating in different ways from that intended. The social consequences of this are backwash effects into the curricula experiences of students; the negative consequence of which provide misrepresentations of boys' and girls' success. Therefore, the interaction of gender and assessment technique suggests that not all boys and girls are affected equally and fairly by similar types of assessment activity.

Limitations of Assessment-Technique Approach

There are limitations to the approach of focusing on the assessment systems or techniques used and their contribution to gender differences in performance. These limitations are connected to the ways in which focusing on the assessment technique and students' experiences of the technique does not allow for a consideration of the interaction between the two, nor does it allow for a more subtle understanding of how gender mediates students' experiences of subject knowledge(s), how these are encountered in classrooms, and how they are ultimately assessed. To move the work forward in a significant way, we need to consider how the social nature of assessment shapes the experiences of students and teachers, their perceptions of the subjects they teach and learn, and the ways in which knowledge(s) and understanding(s) are assessed.

The research outlined above does help in giving us a comprehensive picture of how different aspects of assessment activities (i.e., the test, the item, the task, etc.) influences the outcomes observed and how different groups of students perform across different subjects and contexts of learning. Thus, they do provide, at one level, important information about boys and girls as learner groups but tend to continue to see any differences in

performance as belonging to, and located within, the test taker when all technical aspects of the test have been accounted for. The research in this area has concerned itself with large-scale summative assessment (even that which is teacher assessed). Any such attention to gender and its interaction with formative assessment tasks and practices are less common as are any perspectives that consider problematic issues in classroom assessment. It is to these aspects of assessment and its impact on learner groups that we turn to next.

Future Directions

New interests in the field of gender and assessment are interested in the links between assessment and a view of learning and the learner that are situated within sociocultural theoretical understandings and the implications of such approaches for the understanding of how gender is integral to any consideration of assessment outcomes (Elwood, 2006; Moss, 2004; Murphy and Ivinson, 2004). The work within this field positions learning as a cultural activity that cannot be considered in isolation from the social and historical contexts within which it occurs and defines gender as a fluid, social construct that represents a set of cultural ideas, conventions, and norms within which boys' and girls' social beings are created and played out. Assessment within this view becomes a cultural artifact that only has meaning in relationship with and between the teacher and student and/or peer and the context in which it occurs; the outcomes are relational and cannot be abstracted away from the situation in which they were created.

Thus, to understand the achievements of boys and girls we need to consider these more essentially as relational to, and with, the interaction of learning with assessment practice. The relationships between teachers, students, their learning, and how this learning is assessed are entangled as are the gendered lives of students and teachers and their understandings of subject knowledge(s). All these factors come into play in comprehending how boys and girls learn and how this is represented in outcomes on assessment tasks and tests. From this perspective, to understand boys' and girls' learning through assessment, we need to see gender as located within the outcomes observed.

For example, research by Murphy and Ivinson (2004) considered how teachers present gendered messages to students, either consciously or not, about what it means to do well in a subject, and thus what is considered to be valued subject knowledge and equated with success. They acknowledge that boys and girls interact with these messages in different ways, that boys and girls themselves have different notions of what is relevant or what is of value to them, as well as different ideas about what are gender-acceptable things to be seen to know within the

contexts of socially constructed classrooms. They take the subject of English and an assessment activity of creative writing to illustrate their argument. Murphy and Ivinson (2004) outline an assessment task where boys and girls (aged 14/15) were asked to draft three openings to a novel using a range of genres, were given the opportunity to redraft pieces with the support of teachers and peers, and in some cases were asked to read aloud their paragraphs and receive oral feedback from teachers. It was through this latter part of the activity that different messages were given to students as to what constituted a good piece of writing – that is, the pieces that teachers choose to hear read aloud were considered of value and portrayed to students as the best type of writing. However, boys and girls both chose different genres within which to present their work and also tended to stick with genres that were seen (by them) as gender-appropriate (especially if reading aloud were part of the assessment). Thus, teachers' representations of good pieces of English writing (i.e., their social representation of good subject knowledge), which were those that were chosen to be read aloud, did not always fit with those genres chosen by boys (mainly) or girls in relation to the full nature of the task and the social contexts in which it was played out.

Thus, how teachers view success in the community of the subject, its conventions, forms, practices, and cultural settings can significantly influence their judgments of boys' and girls' abilities. These definitions of success are entangled with teachers' and students' experiences of learning; it is only by looking into these experiences of learning and the assessment of this learning we can obtain a more humble understanding of the impact of assessment on the learning of girls and boys.

Conclusions

This article explored the various research approaches and positions within the field of gender and assessment that attempt to help us understand the impact of assessment activities and practice on learner groups of boys and girls. Learning and gender are fundamental and interrelated concepts that interact with assessment practice in profound ways. As this overview has shown, how these concepts are viewed and defined will dictate how important they are seen to be in relation to assessment practice and the understanding of students' experiences of schooling more generally. If the move toward more formative assessment systems within national assessment programs continues, then consideration must be given to the socio-cultural contexts of classrooms and the relations between teachers and students where this learning takes place and is assessed. The gendered nature of students' lives and experiences do, and will, play significant roles in determining success in assessment outcomes.

Bibliography

- Centre for Educational Research and Innovation (CERI). (2005). *Formative Assessment: Improving Learning in Secondary Classrooms*. Paris: OECD.
- Elwood, J. (2005). Gender and achievement: What have exams got to do with it? *Oxford Review of Education* **31**, 373–393.
- Elwood, J. (2006). Gender issues in testing and assessment. In Skelton, C., Francis, B., and Smulyan, L. (eds.) *The SAGE Handbook of Gender and Education*, pp 262–278. London: Sage.
- Gipps, C. and Murphy, P. (1994). *A Fair Test? Assessment, Achievement and Equity*. Buckingham: Open University Press.
- Goldstein, H. (1996). Group differences and bias in assessment. In Goldstein, H. and Lewis, T. (eds.) *Assessment: Problems, Developments and Statistical Issues*, pp 85–93. Chichester: Wiley.
- Head, J. (1996). Gender identity and cognitive style. In Murphy, P. and Gipps, C. (eds.) *Equity in the Classroom: Towards Effective Pedagogy for Girls and Boys*, pp 59–70. London/Paris: Falmer Press/UNESCO.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist* **60**, 581–592.
- James, M. (2006). Assessment, teaching and theories of learning. In Gardner, J. (ed.) *Assessment and Learning*, pp 47–60. London: Sage.
- Moss, P. A. (2004). The risks of coherence. In Wilson, M. (ed.) *Towards Coherence between Classroom Assessment and Accountability. The 103rd Year Book of the National Society for the Study of Education, Part 2*, pp 217–238. Chicago, IL: Chicago University Press.
- Murphy, P. (1998). *Learners, Learning and Assessment*. London: Paul Chapman.
- Murphy, P. and Ivanson, G. (2004). Gender differences in educational achievement: A socio-cultural analysis. In Olssen, M. (ed.) *Culture and Learning: Access and Opportunity in the Curriculum*, pp 365–386. Greenwich, CT: Information Age Publishing.
- Organization for Economic Co-operation and Development (OECD). (2004). *Learning for Tomorrow's World – First Results from PISA 2003*. Paris: OECD.
- Willingham, W. W. and Cole, N. S. (1997). *Gender and Fair Assessment*. Mahwah, NJ: Erlbaum.
- Wilson, M. (2004). *Towards Coherence between Classroom Assessment and Accountability the 103rd Year Book of the National Society for the Study of Education, Part 2*. Chicago, IL: Chicago University Press.

Further Reading

- Alperstein, J. F. (2005). Commentary on girls and boys, test scores and more. *Teachers College Record*. <http://www.tcrecord.org> (accessed June 2009).
- Black, P. (1998). Assessment, learning theories and testing systems. In Murphy, P. (ed.) *Learners, Learning and Assessment*, pp 118–134. London: Paul Chapman.
- Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education* **5**, 7–71.
- Broadfoot, P. (2007). *An Introduction to Assessment*. London: Continuum.
- Brunner, J. (2006). *In Search of Pedagogy (Volume II); The Selected Works of Jerome S. Bruner*. London: Routledge.
- Elwood, J. (1995). Undermining gender stereotypes: Examination and coursework performance in the UK at 16. *Assessment in Education* **2**, 283–303.
- Epstein, D., Elwood, J., Hey, V., and Maw, J. (eds.) (1998). *Failing Boys? Issues in Gender and Achievement*. Buckingham: Open University Press.
- Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education* **24**, 355–392.
- Hargreaves, A., Earl, L., and Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal* **39**, 69–95.
- Murphy, P. and Whitelegg, E. (2006). *Girls in the Physics Classroom*. London: Institute of Physics Publications.
- National Assessment of Educational Progress (NEAP) (2000). *Trends in Academic Progress: Three Decades of Student Performance*. Washington, DC: National Centre for Educational Statistics.
- Rogoff, B. (2003). *The Cultural Nature of Human Development*. Oxford: Oxford University Press.
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher* **29**, 4–14.
- Skelton, C., Francis, B., and Smulyan, L. (eds.) (2006). *Handbook of Gender and Education*. London: Sage.
- Skelton, C., Francis, B., and Valkanova, Y. (2007). *Breaking down the Stereotypes: Gender and Achievement in Schools*. Manchester: Equal Opportunities Commission.
- Wertsch, J. (1991). *Voices of the Mind: A Sociocultural Approach to Mediated Action*. Cambridge, MA: Harvard University Press.

Impact of Assessment on Learner Groups (Disabilities)

M Rouse, University of Aberdeen, Aberdeen, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Few areas of education have seen changes as profound as those seen in the field of assessment and these changes are having a major impact on the education of children with disabilities and special educational needs. Many countries, especially in the English-speaking world, have enacted policies in which assessment plays a crucial role in attempts to improve standards and to ensure that schools are accountable for the attainment of all learners. These reform agendas are intended to raise attainment overall and to close the achievement gap between the highest- and lowest-performing students through the introduction of challenging curriculum content and achievement standards, systems of assessment that reflect the curriculum and accountability for the performance of all learners. In spite of claims that the reforms should be relevant for all learners, there was little mention of students with disabilities and special educational needs in the first wave of educational reforms. However, more recently, several countries have introduced reforms specifically designed to include such students and to raise standards for all children.

Although children with disabilities and special educational needs were not of central concern to policymakers when standards-based reforms were being formulated, the reforms have had a fundamental influence on their education and assessment. If the aim of these reforms is to raise the educational attainment of all learners, then some serious challenges must be faced, not least that children with disabilities, who may have been excluded from mainstream education and assessment in the past, will also be expected to achieve pre-specified academic standards. In the face of such a challenge, special needs education is undergoing a shift from a preoccupation with identification, resource allocation, and procedural safeguards, to a focus on outcomes and accountability. For example, in the United States, the President's Commission on Excellence in Special Education (US Department for Education, 2002) stressed that the special education must be judged by the outcomes achieved by each child. According to Nagle and Thurlow (2008), this recommendation fits well with the standards-based reforms already underway, specifically those specified under the No Child Left Behind Act (NCLB). In other countries too, questions are being asked about who should, and should not, be included in national systems of assessment and the extent to which national assessments might need to be adapted and modified if all students are to participate.

While the policy responses in various countries have been different, the tensions and implications associated with moving special needs education from a largely separate and highly specialized endeavor into a more inclusive system will be similar (Rouse and McLaughlin, 2007). Furthermore, many countries are looking at what is happening elsewhere, in part because of the impact of international comparisons of attainment and cross-national reviews of special needs education provision (OECD, 2005). Therefore, while this article is largely informed by developments in the English-speaking world, there is also relevance to other countries.

Standards-based reforms raise questions about the consequences of developing national/state systems of assessment that aspire to include all learners and whether new approaches and understandings about inclusive educational assessment might be challenging old assumptions and approaches to the assessment and classification of children with disabilities. In attempting to answer these questions, this article considers the ways in which changes in assessment policies and practices, together with other developments in classification systems outside education, have influenced basic assumptions and approaches to the classification of disability in education. It also examines the extent to which universal standards and common educational outcomes are possible, given the deeply embedded assumptions about human differences that pervade school structures and approaches to assessment.

Disability Classification in Education: Purposes

Before considering the impact of educational assessment on learners with disabilities, it is necessary to ask why some learners have been classified as disabled in the first place.

Historically, there were two main purposes for disability classification in education. The first purpose was to identify the children who might struggle in mainstream schools: those who would need to be placed in alternative forms of provision, such as special schools, or who might need additional help or support within a mainstream setting. It seeks to answer 'the who' questions.

The second purpose involves a more in-depth assessment of the nature and extent of the difficulty or disability. It seeks to answer questions about why some children have difficulties and what might be done about it. This process collects evidence using assessment tools and techniques,

by particular professionals such as school psychologists, in order to allocate a child to a particular category of handicap or disability. The intention here is to provide information that might lead to specific help tailored to the needs of such children based on a diagnosis of their difficulties.

Therefore, the identification and assessment of children's disabilities and educational needs have not only been an important element of selecting some children for alternative forms of provision, both inside and outside of mainstream schooling, but they also continue to be an important component of decisions about resource allocation and of informing pedagogical approaches.

Disability Classification in Education: A Historical Overview

Since schooling became compulsory, it was soon clear that some children did not easily fit into the schools that had been created and a system of sorting out these children was required. More than a century ago Alfred Binet was invited to devise the first tests to identify abnormal children who would need extra help or who would be too difficult to educate in mainstream schools. Subsequently, the Binet–Simon scale was developed for measuring intelligence of children. Since this time, the technologies of testing and systems of classification have changed, as has the language used to describe the children. For example, children are no longer described as imbeciles or subnormal, but many of the fundamental assumptions apparent in Binet's time still inform policies and practices. Underpinning these assumptions is a deep-seated belief that some children are normal while others are different. The distinction between normal and different (not normal) is based on two overlapping sets of assumptions, the first is quantitative, the second qualitative.

Quantitative assumptions are based on beliefs that abilities are distributed normally throughout the population according to the statistical laws of the bell-shaped curve. Thus, 2% or so of the population are located at more than 2 standard deviations below the mean (those typically scoring less than 70 on a standardized test of ability or attainment where the mean is 100 and standard deviation is 15). Such people were not only statistically unusual, but they were seen as not normal or indeed subnormal in the terminology used in the UK until the 1970s. Interestingly, about 2% have historically been placed in special schools in England and Wales.

Differences between the normal and different may also be understood as being qualitative. Some children's development or behavior may not be judged as meeting a set of pre-specified criteria thought to be normal within a particular cultural or educational context. Qualitative assumptions of normality are socially constructed and lie at the heart of many developmental and behavioral checklists. Using

such approaches with children who are not from the dominant cultural group is likely to lead to overrepresentation of minorities in special forms of provision and negative consequences resulting from low expectations.

Special education was built on the foundation of assessment practices designed to sort out children who were considered to be quantitative and qualitatively different. Nevertheless, the educational assessment of children and young people with disabilities has become a contested arena in recent years as long-held assumptions and practices about the nature and purpose of this task have been subjected to scrutiny and critique. For example, in the United Kingdom, there were attempts to move away from assigning children to categories of handicap following the Warnock Report (Department for Education and Science, 1978) to a broader concept of special educational needs. The identification of children's needs was to be made: "not in terms of a particular disability which a child may be judged to have, but in relation to . . . all the factors which have a bearing on educational progress" (Department for Education and Science, 1978: 3.6).

The process of assessment moved from being predominantly about allocating children to categories of disability to an individualized assessment of the child's strengths and difficulties based on an ecological perspective that recognizes disability and special educational needs result from an interaction between an individual's strengths and weaknesses and the environmental context.

Meanwhile, in the United States, the Individuals with Disabilities Education Act (IDEA) continued to use a framework for assessing entitlement to special education services based on the allocation of children to one of 13 disability categories. However, the President's Commission on Excellence in Special Education could find no firm practical or scientific reasons for supporting the existing classification of disabilities in IDEA (US Department for Education, 2002: 22) and suggested new ways of approaching this task. In other parts of the world too, assessment and classification have been changing in response to new understandings about why children have difficulties in learning. Yet, while the movement toward inclusive education has reduced the role of assessment in informing decisions about the placement of children in separate special schools and classes, old approaches to assessment and classifying children as disabled or having special education needs are still alive and well and are being used to justify why some children are not successful learners in schools.

Disability Classification: A Broader Perspective

The classification of disability involves a process of sorting people and collecting evidence in order to make

decisions about people's needs. This does not only occur in education, but it is also a process widely used in other human services, including employment, health, and social services. Disability classification systems have evolved across time within specific social, economic, and political contexts and crucially they exist within legal frameworks. They are used to inform decisions on a range of matters, including eligibility for services and benefits, and are often seen as playing important roles in equal opportunities and antidiscrimination policies (Pullin, 2008).

In education, this process is justified by arguments that it is in the best interests of these children (and the rest of the system) to identify them so that targeted extra help can be provided. It is argued that children have to be classified to protect their rights as part of the struggle for access and equity in education. However, a central dilemma is that attempts to help and protect children may have the opposite effect. By focusing on children's deficits as well as by classifying children as different, there may be unintended negative consequences. The philosopher Martha Minow has explored this dilemma when referring to the dilemma of difference, which results when people are classified as different so that they can be provided with legal protection and services designed to meet their needs, while at the same time recognizing that this very process involves imposing labels and potential stigma (Minow, 1990: 20). Such classification may also result in relocation to a different place where help may be provided but at the price of restricting opportunities to learn and lowering expectations. Thus, there is inevitably a trade-off between the benefits and the negative consequences of classification which manifests itself differently in different contexts, changing over time as conceptions of the common good and notions of difference shift (Pullin, 2008).

Although disability classifications systems are ubiquitous and are resistant to change, there have been challenges to old assumptions about human differences upon which they were constructed. More than 30 years ago, Nicholas Hobbs and colleagues began a major review of the classification and labeling of children in education. It demonstrated the complexity and ambiguity in classifying children, because while it may provide benefits, there are also risks. At that time, Cromwell *et al.* (1975) pointed out the need for regular review of the use of classification and labels to determine what purposes they are serving and injustices they might be creating.

In spite of such challenges, there are many questions that remain unanswered about the utility of educational classification systems as well as concerns about the negative consequences for those who are identified (Florian *et al.*, 2006). In a highly critical report, the President's Commission on Excellence in Special Education pointed out that current methods of identifying children with disabilities lack validity and that thousands of children are misidentified every year (US Department for Education, 2002: 8).

Significantly, assessment has been the predominant means through which additional resources and funding have been allocated to support some children's education. As more children were identified, there was an expansion of services and provision designed to meet special needs, with some countries seeing as many as a third of all children receiving additional support (OECD, 2005). The classification process was thought to be out of control. It was not transparent; it took too long and was consuming a disproportionate amount of professional time and energy (US Department of Education, 2002). The allocation of additional resources depended on a child's failure and schools were reluctant to demonstrate success for fear that any additional resources would be removed. Thus, assessment was at the heart of a system of perverse incentives that was not always working in the best interests of children.

At the same time, there has been movement from the assessment of deficits in individuals, an approach influenced by long-established medical approaches of the *Diagnostic and Statistical Manual of Mental Disorders-Text Revision* (DSM-TR) currently in its fourth version and the International Classification of Disease (ICD). The ICD developed from attempts to agree on a list of causes of death in the early part of the twentieth century and has evolved through ten versions to the ICD-10 (Classification of Mental and Behavioral Disorders), which contains clinical descriptions and diagnostic guidelines. Subsequently, the International Classification of Disability and Handicap (ICIDH) moved the focus from the causes to the consequences of disease and disability (Hollenweger, 2008). However, the ICIDH did not go far enough in acknowledging social and ecological models of disability. In response, there has been a shift to the assessment of a person's functioning within the environment in which they live and learn as can be seen in the International Classification of Functioning, Disability and Health (ICF; World Health Organization, 2001), which sees disability as the result of the interaction between the person and the environment (Hollenweger, 2008). The ICF is built upon an ecological perspective of disability and it has already begun to inform thinking about the assessment in educational settings because it can provide evidence linked to schooling, the curriculum, and teaching and is consistent with approaches based on dynamic and authentic assessment.

Nevertheless, the adoption of new approaches to classification is likely to be slow because even when the reasons for the establishment of old systems may no longer be relevant, complex systems are slow to change. Professional beliefs, commercial self-interest, and institutional inertia mean that approaches to assessment of disability are remarkably resistant to change, even when the original underpinning assumptions or purposes may no longer be relevant.

Assessment for Pedagogical Reasons

In one form or another, most countries have ongoing, formative assessment that is used to inform decisions about teaching and learning. Many of these approaches involve pupils and classroom teachers, are linked to the curriculum, and have been informed by the principles of assessment for learning. The benefits from using these approaches with children who have difficulties in learning are apparent and significant progress has been made in recent years in this regard in several countries (Watkins, 2007). However, in the field of special needs education, it is thought that more specific information about children's difficulties is required so that teaching can be matched to children's needs. In spite of the common-sense power of this argument, the usefulness of such approaches has been challenged. For example, in an extensive review covering 25 years of work, Ysseldyke (2001) has shown that there is insufficient evidence to support a diagnostic-prescriptive approach to teaching.

Assessing and Classifying Special Educational Needs: A Summary

For many years, the development of tests and diagnostic instruments was seen as a conceptual and technical challenge and the field of psychometrics developed to meet the demand for increasingly more sophisticated ways of sorting people out and diagnosing their needs. This grew into a multimillion-dollar industry and the products of this endeavor have had a profound influence on the development, not only of special education and disability classification, but also more broadly on schooling and systems of assessment throughout the twentieth century. Crucially, it has left a deep-rooted set of beliefs about human (dis)ability that continue to be used to explain and justify the educational failure of some children.

Including All Learners in National Systems of Assessment: The Impact on Students with Disabilities

Serious questions remain about how children with disabilities and special educational needs might be included in national systems of assessment and the extent to which state and national assessments might need to be adapted and modified if all students are to participate. In many countries, it has been recognized that some children may require assessment accommodation, such as additional time, a reader, large print, or alternative forms of response. The term accommodation is used to mean those changes that are not considered to alter the construct measured. Nevertheless, questions remain about who should decide which children are allowed such accommodations and the

extent to which such accommodations affect validity and reliability of test scores.

In England, flexibility within the national curriculum and its assessment systems have permitted children to be assessed outside of key-stage level and there are well-developed alternatives for children with severe cognitive impairments to work at levels below the starting point of the national curriculum on what are known as 'p' levels.

In the USA, the legislation requires states to annually assess the performance of all students in language, mathematics, and science against three levels: proficient, advanced, and basic. States have some flexibility; however, assessment must reflect the content standards, must measure the achievement of all children, and the results must be used to hold schools and school systems accountable for the achievement of all children. In part, this is carried out through setting annual adequate yearly progress (AYP) performance targets for individual schools and school systems designed to ensure that all children reach grade-level proficiency in reading and mathematics and science by 2014. AYP requires that not less than 95% of children with disabilities should be assessed. School districts and schools that fail to make AYP toward statewide achievement are subject to mandatory and increasingly severe sanctions aimed at getting them to meet state standards.

Despite the rhetoric that all children should take part in statewide assessments, the US Department of Education recognizes that it is not appropriate to measure all children with disabilities against grade-level achievement standards and that not all can participate in normal statewide assessment. States have been granted the flexibility to measure the achievement of children with the most significant cognitive disabilities against alternate achievement standards and to develop assessments based on alternate or modified achievement standards.

Interestingly, some unintended consequences are becoming apparent. Nagle and Thurlow (2008) suggest that children with disabilities are increasingly being described by the type of assessment they take; by the achievement standards against which they are measured and by the proximity of their score to the level deemed as proficient and that the established disability categories are less relevant than previously. The current regulations defining how students with disabilities will participate in state assessments may have unintentionally created a new categorization system: those children with disabilities, who are held to grade-level achievement standards, take the regular assessment, with or without accommodations, and those children who are held to alternate or modified achievement standards and take an alternate form of assessment. The practice of labeling children by their performance level encourages schools to engage in what Gillborn and Youdell (2000) in England and Booher-Jennings (2005) in the US refer to as educational triage, a process in which schools make judgments about where to target the

extra help. The evidence from Gillborn and Youdell (2000) is that extra help is provided for those students who are within reach of the passing grade to ensure that the school reaches its specified targets. Interventions are not provided to students who are either clearly proficient or unlikely to pass even with extra help.

There is considerable evidence that schools are focusing their resources on children who are just below the proficient level. Schools concentrate on the almost-passing children because they have a chance at moving them over the proficiency bar and improving their chances of making AYP (Booher-Jennings, 2005). Research conducted by the Educational Policy Reform Research Institute (Nagle and Thurlow, 2008) supports this contention. Practitioners involved in their research indicated that they concentrated their resources on these children, regardless of whether the child has or does not have a disability. Teachers and administrators indicated that they had a range of initiatives in place to improve the achievement of these children as the children who are unlikely to pass or are well within the basic range, may be perceived as lost causes and marginalized. It seems as though teachers do not believe that they can raise their performance and that there was no point in helping them because such students have already reached their potential. Thus, current systems of high-stakes accountability may be leading to new systems of classification based on whether children are deemed to be worth helping.

In the both England and the USA, national systems of assessment have created new dilemmas in which individuals and schools may be using disability classification to gain advantage. For example, individuals and their parents may have incentives to acquire a disability or special needs status if the benefits (such as additional time on tests) are thought to outweigh the potential costs of any stigma associated with the label. One of the most problematic aspects of this search for relative advantage can be seen in the huge increase in requests for modified assessment arrangements to compensate for dyslexia and other disabilities. In addition to bringing advantage to individuals, it may be to the advantage of schools if children are classified as disabled or having special educational needs if it means that accountability standards are lowered, or the conditions under which the assessments are taken are more favorable. This trend raises complex questions about fairness, equity, and equal opportunity (Pullin, 2008) as well as threatening the validity of the assessment process.

Conclusion

This is a time of great change and uncertainty for state/national systems of assessment and for disability

classification in education. A number of challenges face systems of assessment as they struggle to include students who are currently excluded from participation. If assessment systems are primarily for accountability purposes, then they will distort the ways in which schools view learners who may find it difficult to achieve the pre-specified standards. Equally, if an important purpose of assessment is to sort children out, then many will continue to be excluded. However, if national systems of assessment can be developed in which all learners can demonstrate their achievements, then we will be not only on the way to having an inclusive framework for assessment but also one of the fundamental purposes of disability classification will be redundant. However, while deep-seated assumptions about the so-called normal distribution of human abilities remain, it is unlikely that children with disabilities and special educational needs will be able to participate in national assessment without reinforcing notions of failure for children, teachers, and schools.

Nevertheless, the nature and purpose of disability classification in education is changing from a focus on deficits in individuals to one that acknowledges functioning and recognizes the importance of the broader learning environment. Such approaches do not depend on norm-referenced distributions, nor are they focused on sorting people out so that alternative forms of provision can be made. Perhaps some of the lessons learned from such approaches can be incorporated into the national systems of assessment. This is important because as Kirp (1982) reminds us: a way of knowing about a phenomenon can become the way of knowing about it. For example, in special education, systems of assessment and classification are deeply embedded at the heart of the enterprise and assessment decisions have particular consequences for where children are educated, what kind of curriculum they will receive, who they are educated alongside, which professionals they encounter, and what life courses are available for them. Traditional assessment practices were not only a way of knowing, they became the way of knowing; they had become deeply institutionalized. Therefore, shifting assessment practices from those based on norm-referenced assumptions and a focus on deficits in individuals, to one that focuses on functioning and achievement could bring benefits for all, not only for children with disabilities. It would also create the conditions for fairer and more relevant national systems of assessment that are capable of recognizing the achievements of all learners.

See also: Alternative Assessment; Dynamic Assessment; Formative Assessment; Students' Rights in Assessment and Evaluation.

Bibliography

- Booher-Jennings, J. (2005). Below the bubble: 'Educational triage' and the Texas accountability system. *American Educational Research Journal* **42**(2), 231–268.
- Cromwell, R. L., Blasfield, R. K., and Strauss, R. S. (1975). Criteria for classification systems. In Hobbs, N. (ed.) *The Futures of Children: Categories, Labels and Their Consequences*, vol. 1, pp 4–25. San Francisco, CA: Jossey-Bass.
- Department for Education and Science (1978). Special educational needs, *The Warnock Report*. London: HMSO.
- Florian, L., Hollenweger, J., Simeonsson, R. J., et al. (2006). Cross-cultural perspectives on the classification of children with disabilities: Part 1: Issues in the classification of children with disabilities. *Journal of Special Education* **40**(1), 36–45.
- Gillborn, D. and Youdell, D. (2000). *Rationing Education: Policy, Practice, Reform, and Equity*. Buckingham: Open University Press.
- Hollenweger, J. (2008). Cross-national comparisons of special education classification systems. In Florian, L. and McLaughlin, M. (eds.) *Disability Classification in Education: Issues and Perspectives*, pp 11–30. Thousand Oaks, CA: Corwin.
- Kirp, D. L. (1982). Professionalization as a policy choice: British special education in comparative perspective. *World Politics* **34**(2), 137–174.
- Minow, M. (1990). *Making All the Difference: Inclusion, Exclusion and American Law*. Ithaca, NY: Cornell University Press.
- Nagle, K. and Thurlow, M. (2008). Classification of children with disabilities in the context of performance-based education reform – an unintended classification system. In Florian, L. and McLaughlin, M. (eds.) *Disability Classification in Education: Issues and Perspectives*, pp 109–204. Thousand Oaks, CA: Corwin.
- OECD (2005). *Student with Disabilities, Learning Difficulties and Disadvantages. Statistics and Indicators*. Paris: OECD.
- Pullin, D. (2008). Implications for human and civil rights entitlements: Stigma, stereotypes and civil rights in disability classification systems. In Florian, L. and McLaughlin, M. (eds.) *Disability Classification in Education: Issues and Perspectives*, pp 94–108. Thousand Oaks, CA: Corwin.
- Rouse, M. and McLaughlin, M. (2007). Changing perspectives of special education in the evolving context of educational reform. In Florian, L. (ed.) *The Sage Handbook of Special Education*, pp 85–106. London: Sage.
- US Department of Education (2002). *President's Commission on Excellence in Special Education. A New Era: Revitalising Special Education for Children and Their Families*. Washington, DC: US Department of Education.
- Watkins, A. (ed.) (2007). *Assessment in Inclusive Settings: Key Issues for Policy and Practice*. Odense, Denmark: European Agency for Development in Special Educational Needs.
- World Health Organisation (2001). *International Classification of Functioning, Disability and Health (ICF)*. Geneva: WHO.
- Ysseldyke, J. E. (2001). Reflections on a research career: Generalizations from 25 years of research on assessment and instructional decision making. *Exceptional Children* **67**(3), 295–309.

Further Reading

- Cline, T. (1992). *The Assessment of Special Educational Needs: International Perspectives*. London: Routledge.
- Florian, L. and McLaughlin, M. (eds.) (2008). *Disability Classification in Education: Issues and Perspectives*. Thousand Oaks, CA: Corwin.
- Gardner, J. (ed.) (2006). *Assessment and Learning*. London: Sage.
- Gould, S. J. (1996). *The Mismeasure of Man*, 2nd edn. Harmondsworth: Penguin Books.
- Harry, B. and Klingner, J. (2006). *Why Are So Many Minority Students in Special Education? Understanding Race and Disability in Schools*. New York: Teachers College Press.
- Hart, S., Dixon, A., Drummond, M. J., and McIntyre, D. (2004). *Learning Without Limits*. Maidenhead: Open University Press.
- Hobbs, N. (ed.) (1975a). *The Futures of Children: Categories, Labels and Their Consequences*. San Francisco, CA: Jossey-Bass.
- Hobbs, N. (ed.) (1975b). *Issues in the Classification of Children: Categories, Labels and Their Consequences*. San Francisco, CA: Jossey-Bass.
- McLaughlin, M., Dyson, A., Nagle, K., et al. (2006). Cross-cultural perspectives on the classification of children with disabilities part II. Implementing classification systems in schools. *Journal of Special Education* **40**(1), 46–58.
- Salvia, J., Ysseldyke, J., and Bolt, S. (2010). *Assessment in Special and Inclusive Education*. Wadsworth.
- Sattler, J. (2008). *Assessment of Children: Cognitive Foundations*, 5th edn. La Mesa, CA: Jerome M Sattler Publisher.

Impact of Assessment on Students' Test Anxiety

P Buchwald and C Schwarzer, Heinrich-Heine University Dusseldorf, Dusseldorf, Germany

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The assessment of student learning can be seen as an essential component of evaluating the effectiveness of learning in an organization. Identifying factors influencing student learning and performance is a major goal for teachers and educational researchers. Test anxiety is a major predictor of academic performance and various studies have demonstrated that it has a detrimental effect (Zeidner, 1998). Test anxiety can be defined as a student's reaction to testing or evaluating situations. However, testing and assessment does not only have an impact on test anxiety, but also influences motivation to learn and further performance. The potential impact seems to differ from how and by whom the assessment is done, in which way it affects the educational career, and also how it is perceived by the person assessed.

The effects of test anxiety on performance have first been explained by the Yerkes–Dodson law. Other explanations follow the interference and learning-deficit models. The interference model focuses on the mediating role of negative thoughts on the anxiety–performance linkage. According to the learning-deficit model, the student with high test anxiety tends to use inadequate learning skills while in the preparation stage of test taking. Due to their theoretical underpinnings, however, all these models look at anxiety in an individualistic manner. In order to get a more comprehensive understanding of the impact of assessment on student's test anxiety we introduce the conservation of resources (COR) theory (Hobfoll, 1998) which does not only stress individuals' perceptions but also takes environmental contingencies into account. COR theory makes it possible to look at both, teachers and students simultaneously and combines the aspects of psychosocial well-being of teachers and students with different categories of assessment.

Foremost, the effects of test anxiety on performance have been depicted as a curve (see **Figure 1**) by Yerkes and Dodson (1908). The Yerkes–Dodson curve shows that up to a threshold anxiety is related to increased performance. Conversely, performance efficiency decreases if anxiety is still at an ever-increasing rate. Up to the summit of the curve, anxiety can be adaptive because it motivates students, helps to prepare for a test, and improve their functioning. Beyond the optimal level, the peak, anxiety is considered maladaptive as it causes distress and impairs functioning. In general, high test anxiety is more closely associated with

lowered performance in low-ability students than in their high-ability counterparts (Hembree, 1988).

Two key components of test anxiety are cognition and emotion. Liebert and Morris (1967) suggested that test anxiety was composed of worry and emotionality. Worry was conceptually identified as cognitive expression of concern about one's own performance, while emotionality referred to autonomic reactions which tend to occur under examination stress as stated by Liebert and Morris. Negative effects of test anxiety on academic performance have been explained by two models: the interference model (Sarason, 1986) and the learning-deficit model (Naveh-Benjamin *et al.*, 1987). According to the interference model, anxious students are distracted due to worrying and task-irrelevant cognitions and the learning-deficit model proposes that it is students' ineffective study habits during preparation for a test that leads to anxiety and less performance on the test.

These stated aspects demand for answers and possible resultant actions. Assessment models and evaluative situations have to be modified allowing for successful learning, coping strategies, and better performance. In general, education has to rethink assessment procedures and integrate alternative assessment approaches such as portfolios, project assessment, or other classroom-assessment techniques which focus on students' resources rather than on their deficits in order to diminish feelings of test anxiety. Therefore, broader resource-oriented theories are required to get a more comprehensive understanding of the impact of assessment on student's test anxiety. First, we would like to introduce the COR theory as such an inclusive framework, and second, provide an overview about alternative assessment methods.

Stress, Anxiety, and Student Learning in the Framework of the Conservation of Resources Model

Students facing the challenge of coping with upcoming tests and assessment situations is a common phenomenon. Students seem to be more vulnerable at examination times, and adaptive coping with assessment is important for their achievement and psychological well-being (Zeidner, 1998). The role of stress, anxiety, and coping in an assessment situation has been investigated by a number of researchers. They mainly focused on the cognitively based

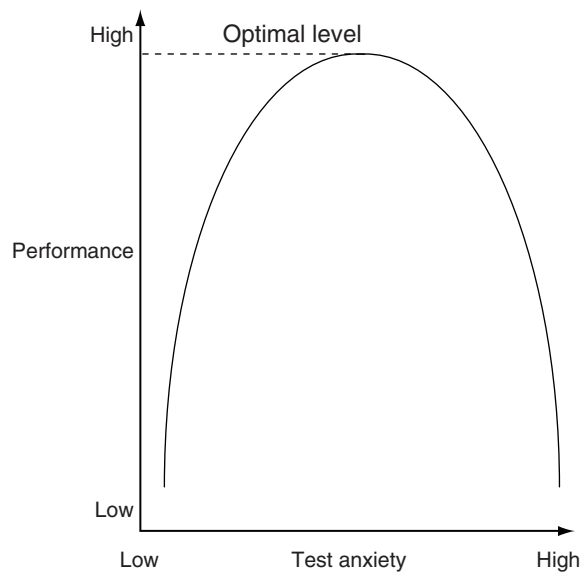


Figure 1 Effects of test anxiety on performance.

transactional stress concept of Folkman and Lazarus (1985) viewing emotions as arising from the individuals' perceptions of imbalance between individual resources and environmental demands. By highlighting perceptions when defining stress and coping capacity, primarily, the importance of individual difference factors are stressed and only secondary emphasis is placed on environmental contingencies. Hobfoll *et al.* (1998) developed the COR theory in response to the need to incorporate more fully both the objective and perceived environment into the process of coping with stress.

Theoretical Concepts of Conservation of Resources Theory

COR theory (Hobfoll, 1998) offers a comprehensive framework for understanding the impact of assessment on emotions and performance by focusing on the resources of individuals and groups (Buchwald, 2003). First, we outline the theoretical principles of COR theory and give more detailed examples of the application of COR theory to educational assessment settings on the individual and group level. Finally, we discuss future directions for integrating COR theory into best-practice models to enhance students' learning development.

According to COR theory, human beings' primary motivation is to build, protect, and foster their resource pools in order to protect the self and the social bonds that support the self. The theory provides a model for preventing resource loss, maintaining existing resources, and gaining resources necessary for engaging in appropriate behaviors. COR theory argues that resources are the key components

to determining individuals' appraisals of events as stressful and resources define how individuals are able to cope with stressful situations. As a result of the strong association to broader life conditions, COR theory (see Figure 2) can augment our understanding of stress and coping, particularly in complex learning situations in which students have differential access to resources.

Further, COR theory proposes that those already lacking in resources will be more vulnerable to the experience of loss spirals and those with plenty of resources will have more opportunity for resource gain. Loss spirals, as Hobfoll (1998) explains, occur when resources are expended, and are therefore not available to cope with future-loss threats, thus potentially leading to further loss. Initial loss leaves individuals, groups, and communities more vulnerable to the negative impact of ongoing resource challenges. Those endowed with greater resources will be more resilient, but ongoing resource loss will challenge even richly resource-endowed individuals or groups. Thus, loss spirals are a powerful force that is evident in individuals and communities already lacking resources.

As Figure 2 shows, the processes of resource conservation are a product of both broader life circumstances as well as resource-loss events. Conditions of resource loss tend to lead to further resource loss, sometimes engendering a cyclical process. When losses occur, individuals apply resource-conservation strategies, whereby they utilize resources available to them in order to adapt as successfully as possible. Successful adaptation generates new resources which, in turn, replenish people's resource pools and offset the conditions that produce acute and chronic resource loss. Those with less coping capabilities will have to employ riskier resource protection and gain strategies, which are less likely to yield the results hoped for (Hobfoll, 1998). Unsuccessful strategy employment results in both psychological distress as well as material loss, such as the diminishment of the resources invested. Such unsuccessful loss-prevention strategies further generates secondary resource loss, resulting in loss spirals.

Conservation of Resources Theory: Application to Learning, Assessment, and Test Anxiety

COR theory can be applied to education and assessment by conceptualizing test anxiety as a loss of resources (loss of self worth, motivation, and of productive cognitive processing) affecting students' learning and performance. Within the framework of COR theory, test anxiety can be understood as a severe stressor. Resources can be outlined commonly used to cope with resources loss in school and other learning settings. When individuals generally strive to obtain, retain, and protect what they value, testing situations increase the probability for test anxiety and a loss of the associated valued resources. Stress in testing situations can be so excessive that it hinders a person's

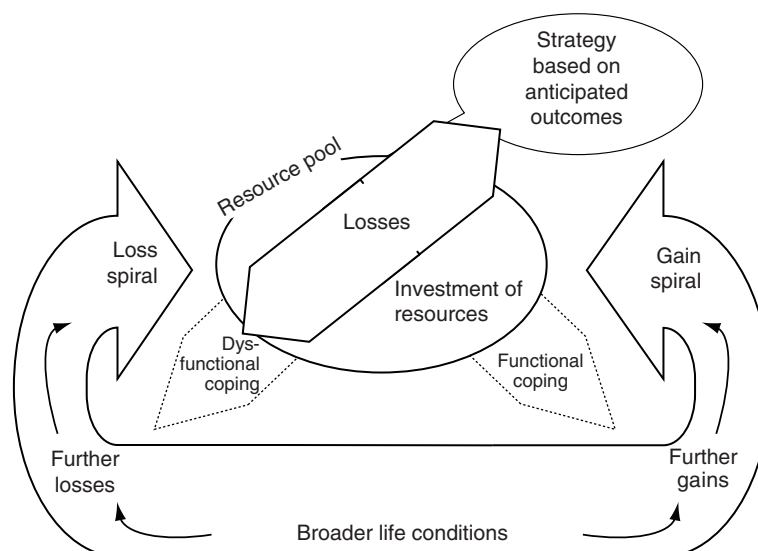


Figure 2 Processes of resource conservation.

ability to prepare properly and test effectively. Test anxiety is associated with poor performance, attacking self-worth and prospects for personal growth.

COR theory outlines four general categories of resources: objects, conditions, energies, and personal resources. Each of these resource categories aids coping efforts. For students facing testing situations we begin by considering object resources important to this group, such as a room of one's own, teaching books, and a computer. These resources provide the basis for coping: a computer provides access to knowledge and support; and a workroom provides a calm atmosphere for uninterrupted learning. In line with this is the finding that task interruption affects judgment and decision making. Condition resources facilitate acquisition or protection of valued resources. Critical conditions for students are: the kind of school, curriculum, stereotypes, assessment methods, and reference norm. For example, it is not uncommon for students to face social-referenced tests. The resulting phenomenon is also known as the big-fish-little-pond effect (BFLPE) and states that it is better for academic self-concept to be a big fish in a little pond (gifted student in regular reference group) than to be a small fish in a big pond (gifted student in gifted reference group). Empirical support for the BFLPE comes from numerous studies based on a variety of different experimental approaches. Zeidner and Schleyer (1999) examined this effect with respect to academic self-concept, test anxiety, and school grades in a sample of 1020 gifted Israeli children. The authors could confirm that academically talented students enrolled in special gifted classes, perceived their academic ability and chances for academic success less favorably compared to students in regular mixed-ability classes. These negative self-perceptions, in turn, served to deflate students' academic self-concept, elevated their levels of test anxiety,

and resulted in depressed school grades. Further, academic self-concept and test anxiety were observed to mediate the effects of reference group on school grades.

Other resources used to respond to stressful test situations are personal resources, including personal characteristics and skills. Individual characteristics most frequently studied in student samples include variables such as motivation, sense of control, and learning strategies. Anxiety has been shown to lower levels of motivation in highly evaluative learning settings and to impact learning strategies. It could be shown that anxiety was significantly negatively correlated with learning strategies involving rehearsal, active reflection, written help seeking, practical application, emotional control, motivational control, and comprehension monitoring. Self-efficacy, the individual's belief in their ability to execute behaviors necessary to achieve a certain goal, can be viewed as a potent resource that aids in coping with assessment and resultant stress. A series of recent studies has provided consistent evidence for the association between self-efficacy and test anxiety (see Zeidner, 1998 for a review). It can impact the goal of performing as well as enhance the internal motivation to manage the testing situation. Overall, results indicate that self-efficacy is a salient and powerful predictor of test anxiety and is negatively correlated with test anxiety.

Finally, energy is the fourth resource category in COR theory. Energy resources include money, time, and knowledge, and it allows access to other resources. With regard to test anxiety, examination-related knowledge is a significant energy resource. Empirical evidence was found for the fact that students with high test anxiety developed and maintained less complete conceptual representations of the course content. Their inadequate time management and procrastination of study tasks lead to various study problems. Thus, timely preparation and knowledge

gaining is a fundamental way to minimize test anxiety. A confident knowledge of course material is the first step in reducing test anxiety.

Resource Loss and Gain Spirals

COR theory states that those who lack resources are more vulnerable to resource loss, and initial loss leads to further loss. By contrast, those who possess resources are more capable of gain, and initial resources create further gain (Hobfoll, 1998). Many empirical findings in studies of learning, assessment, and test anxiety are consistent with such gain and loss cycles (see **Figure 3**). Young people with text anxiety are usually ambitious students who work hard and have high expectations of themselves. The situation may start with inadequate performance on a particular test, which then creates a general fear of the testing situation that hinders future performance, producing a vicious cycle of anxiety, learning, and poor performance associated with several losses in many ways and at many points in the learning process. Anxiety affects motivation during the learning process, the assessment of learning, and the learning strategies that are employed by students. The cumulative effect of anxiety can ultimately be detrimental to a student's overall learning. When considering the nature of the feedback process of emotion, cognition, learning, and assessment, test anxiety can continue to negatively impact a student's academic career in terms of an educational loss

spiral. Still, COR theory with its emphasis on conservation and protection of resources also has implications for understanding the potential positive impact of stressful events as described by Yerkes and Dodson (see **Figure 1**). Students experiencing stress and anxiety facing a testing situation seek both to repair the damage and to mobilize resources for further resource protection (Hobfoll, 1998). A little nervousness can actually help motivate students and provoke them to learn and seek for support. A limited amount of stress is natural and helps to keep learners mentally and physically alert. However, too much may cause physical distress, emotional upset, and concentration difficulties.

Different Assessment Approaches: A Comparison

Students' performance can be assessed in several ways in order to gain information that tells directly or indirectly about the learning process or results. Direct evidence of the process is gained if students show their mastery of learning goals through the demonstration of knowledge, skills, and other gained resources. Indicators for that mastery are resource-oriented ratings of students' competences in research projects, presentations, and performances. Indirect evidence is provided by grades of class tests or examinations that allow making inferences about learning, but does not demonstrate actual learning.

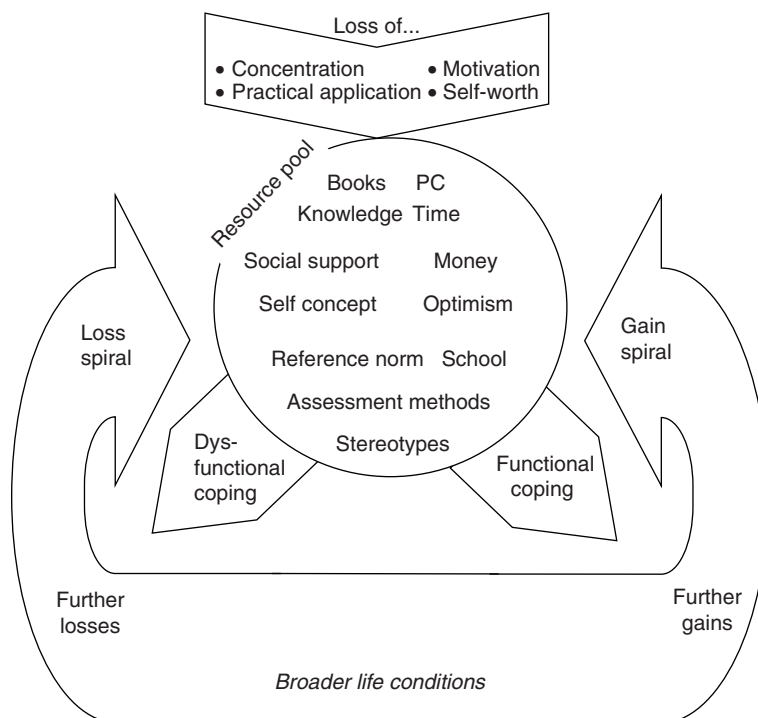


Figure 3 Processes of resource conservation of students facing an assessment. Note that words in upper case refer to object resources; words underlined refer to energy resources, words in lower case refer to condition resources ; and words in bold refer to personal resources.

(Standardized) Written Tests

A test or a written examination is a technique to assess students' knowledge, skills, or abilities. Tests are usually divided into several parts, each covering a different area of the field to be tested. It is generally recognized that these commonly used tests focus on deficits and can provide a teacher with only a limited view of students' achieved knowledge during a semester course or class (Slater, 1997). A standardized version of a test is one that measures ability of every individual subject against a norm or a standard that may be established independently, or by statistical analysis of a large number of subjects. Standardized tests, using multiple choices true–false or fill-in-the-blanks questions, also provide only a limited basis for understanding and evaluating student performance. They deal primarily with factual information. These deficiencies and others have been thoroughly described and documented. However, in the US and other countries, multiple-choice tests have come to be used for assessments of great importance (e.g., high-stakes tests). Testing becomes more and more omnipresent, and the consequences associated with test performances are ever increasing. Consequently, examination stress and test anxiety have become pervasive problems in modern society and the level of test anxiety in both students and teachers has increased as well.

Oral Examinations

An oral examination provides a meaningful procedure in terms of testing for extended problem-solving ability. It normally consists of 30–120-min sessions and is spent with an oral examiner or conducted by a panel. A typical oral exam involves one or two students interacting only with the instructors. Examinees have to demonstrate that they possess the knowledge, skills, and abilities required by giving complete answers to the questions. If the examinees do not know the answer to a question, anxiety can arise, hindering appropriate thinking. However, examiners may be assessing examinees' ability to deal with this difficult or complex situation. Besides exam-related knowledge, other skills are assessed, such as nonverbal language and communication skills. Considering the great variety of social tasks in an oral examination, it is evident that interpersonal resources are of potential relevance and are broadly recognized as important in this context. Buchwald (2003) focused on the examiner–examinee relationship with respect to test anxiety and communal coping during an oral examination. She explored how coping strategies of examiners were related to nonverbal stress behavior and performance of examinees. The study revealed that in examiner–examinee dyads, where male examiners utilize more dominant coping patterns (e.g., aggressive–antisocial action), female examinees were more likely to utilize passive–submissive patterns reflected by avoidance and showed poorer performance.

Online Assessment

Online and computer-based assessment, particularly for universities, are natural consequences of the rapid growth of using the Internet and computer technologies to enhance learning. As more students seek flexibility in their courses, there will be growing expectations for flexible assessment as well. The major educational question is whether online assessment, including multiple-choice questions, is having any influence on the learning benefits and on test anxiety. Although there is some evidence that online assessment can encourage students to focus on lower-level cognitive skills, there are possible positive effects as well on test anxiety. On the one hand online assessment concentrates primarily on true/false or multiple-choice responses, and can have direct negative effects on student approaches to learning by encouraging narrow reproduction rather than the development of higher-order cognition. On the other hand, students taking tests online reported lower levels of perceived test threat (Cassady and Gridley, 2005). This might be due to the fact that online tests enable students to schedule tests following their own workload and allow the students to reduce the level of contextual stress by strategically placing their testing times in convenient time slots.

Portfolios

Portfolios provide a forum for extended and complex learning activities and observations (Klenowski, 2002). Generally speaking, a portfolio is a systematic collection of a variety of teacher observations and student products, collected over time. It reflects a student's developmental status and progress made in a specific field. In such a procedure that focuses on learners' resources, much of the responsibility of both learning and assessment is transferred to the student. Slater (1997) explored the effectiveness of portfolio-assessment strategies and could show that students assessed by portfolios scored just as well on a traditional multiple-choice final examination as their traditionally assessed counterparts. In addition, the students reported reduced levels of test anxiety and enjoyed class discussion more because of the atmosphere promoted by the assessment strategies employed.

Observation of Cooperative Learning Activities

What are needed in addition to traditional assessment methods are those methods for understanding students' learning skills and learning styles (Panitz, 2001). Cooperative learning activity offers teachers an excellent opportunity to observe students interacting, explaining, arguing, helping their peers, and being helped. These observations can provide significant insights into a student's ability and performance level (Panitz, 2001). One important point is that it brings about informal conversations between

individual students, cooperating groups, and the teacher which help to create a more personal class environment. Students get to know the teacher and vice versa, which, in turn, might enhance sympathy and reduce uncertainty and test anxiety. Cooperative learning can also foster cooperation and connections with their peers, enhance team building and teamwork skills, and eliminate cheating. Further and most importantly, cooperating creates a more humane learning environment, is associated with higher levels of student satisfaction, and lowers test anxiety (Muir and Tracy, 1999). Observation of cooperative learning activities could also help identify students who are shy and anxious in order to encourage their participation in nonthreatening ways. Due to the permanent observation process, teachers know which students perform well and which need individual promotion. Teachers can help students understand when they have gained resources by mastering course material, which can reduce anxiety, raise students' self esteem, put them in control of their own destiny, and emphasize that they are responsible for their own learning. The results they obtain are based upon their efforts, not the teacher's. Only very few studies have examined the relationship between collaboration in an evaluative context and test anxiety and even fewer have attempted an empirical assessment of this association.

Product/Project Assessment

A project can be a task given to an individual student or a group and results in a product. The product as well as the processes used during the project can be assessed. Projects are primarily learning experiences, and secondly, an assessment task. There are several advantages of project assessment tasks; for example, they allow teachers to assign projects at different levels of difficulty to account

for individual learning styles and ability levels. They can be motivating and can provide an opportunity for positive interaction and collaboration among peers, and increase students' self-esteem as they would not feel tested as in tests or traditional writing assignments.

Overview of Alternative Assessment Approaches

We outlined the pros and cons of several assessments techniques and their relatedness to test anxiety. **Table 1** presents a wide range of assessment ideas classified into four categories: tests, project assessments, performance assessments, and process skills assessments.

Implications and Future Directions

We have introduced test anxiety as a major predictor of performance and have presented the basic framework of COR theory and offered it as a possible explanatory model for the study of assessment of student learning, stress, and anxiety. COR theory may be applicable to stress during testing situations as resources are often lost and challenged under such circumstances. Testing and assessment might produce a loss of resources and the consequences of such circumstances are often long term. For students, increasing pressure to perform well in tests, combined with cognitive interference, a lack of self-efficacy and motivation, as well as negative life conditions (e.g., inadequate opportunities to learn), can lead to test anxiety and feelings of hopelessness. This, in turn, can cause students to stop trying to improve their performance. Consistent evidence of poor performance repeatedly

Table 1 Alternative assessment approaches

<i>Tests</i>	<i>Products</i>	<i>Performances</i>	<i>Process skills</i>
Essay	Art exhibit	Activities	Checklist observations for processes
Fill in the blank	Audio cassettes	Character sketches	Experiences checklists
Matching	Books	Cooperative learning group activities	Interactional analyses
Multiple choice	Collages	Dances	Interviews
True-false	Computer creations	Discussions	Metaphor analyses
	Drawings	Demonstrations	Observations
	Essay	Experiments	Oral questioning
	Journal	Laboratory experiences	Question production
	Lab reports	Reports	Retelling in own words
	Poem	Simulations	Describing how something was done and justifying the approach used
	Portfolios	Exercise routines	
	Posters		
	Projects		
	Results of surveys		

reported to families, peers, and to the public can result in a long-lasting loss spiral.

Focusing on resources helps to appreciate that evaluative stress affects students, teachers, and schools, both on a personal level as well as on an environmental level. COR theory highlights that these different resource levels are interrelated and produce a web of interactions that must be understood if we are to explain stress reactions and test anxiety. That is, the potential impact of stress in evaluative settings seems to differ from how and by whom the assessment is done (object and condition resources), in which way it affects the educational career (condition resources), and also how it is perceived by the person assessed (personal and energy resources). Evaluative stress threatens and challenges resources and can result in the loss of the most precious resources. How well students respond depends, in part, on the degree of resources that are lost and the extent to which they can preserve resources in order to initiate the recovery and rebuilding of resources.

Using COR theory as a framework to understand the loss and resilience of students, we can work to make schools and other educational institutions more responsive to their unique needs. Despite large-scale assessment or written class tests, those assessments are needed that enable teachers gather information about student learning on a day-to-day basis and enable them to use it to benefit students. Individual classroom assessment might increase performance and decrease the loss of self-worth and self-efficacy. Black and William (1998) could confirm in their comprehensive meta-analysis, of more than 40 controlled studies, the impact of improved classroom assessment on subsequent student success. Underlying the various approaches to improving classroom assessment are assumptions about what is responsible for effective learning is in particular that students have to be actively involved in the assessment process. This includes a resource-oriented, student-involved feedback process. Moreover, student-involved record keeping brings students into the process of monitoring improvements in their performance through repeated self-assessments over time. One way to accomplish this is by having students build portfolios of evidence of their success over time and requiring periodic student self-reflections about the gains of resources.

COR theory also provides guidance for future research questions. Although much is known about stress, anxiety, and coping in evaluative settings on an individually perceived and deficit-focused level, a resource-oriented perspective leads to new questions. How can teachers help their students want to learn? How do they encourage them to feel self-efficacious and take responsibility for maximizing their own success? How can assessment be used not as a source of stress and anxiety but as a source of confidence? How can we provoke a gain spiral?

Using COR theory to better understand the impact of evaluating settings on the processes of resource loss and

gain might greatly expand our understanding of how students experience loss as well as how they become able to cope with these losses.

Summary

The topic of this article is in a way a tricky one because it stresses two facets of the role of teachers which are not really compatible: on the one hand stands the obligation to create a student-centered pedagogy, including, for example, problem solving in teams by using communal coping strategies, and on the other hand stands the requirement to evaluate these students. This dilemma which comes to its peak with respect to high-stakes testing can lead to negative feelings such as stress, anxiety, or bad mood in both, teachers as well as students. With respect to teachers, a wide range of literature refers to burnout as an outcome of this conflict, while the detrimental effects of assessment on students worldwide is documented roughly in this article. The research outcomes for both groups are depicted mostly in different bodies of literatures which might be due to the theoretical underpinnings of looking at the emotion of anxiety in an individualistic manner. The introduction of the COR theory focuses on the different aspects of resources and the possibility of experiencing loss and win spirals makes it possible to look at both, teachers and students simultaneously as persons who strive to build, protect, and foster their resource pools. This viewpoint gives way to new research combining the aspects of psychosocial well-being of teachers and students with different categories of assessment.

See also: Alternative Assessment; Formative Assessment.

Bibliography

- Black, P. and William, D. (1998). Assessment and classroom learning. *Assessment in Education* **5**, 7–74.
- Buchwald, P. (2003). The relationship of individual versus communal state-trait coping and trust, empathy, and responsibility. *Anxiety, Stress, and Coping* **16**, 307–320.
- Cassady, J. C. and Gridley, B. E. (2005). The effects of online formative and summative assessment on test anxiety and performance. *Journal of Technology, Learning and Assessment* **4**, 5–31.
- Folkman, S. and Lazarus, R. S. (1985). If it changes it must be a process: Study of emotion and coping during three stages of a college examination. *Journal of Personality and Social Psychology* **48**, 150–170.
- Hembree, R. (1988). Correlates, causes, and treatment of test anxiety. *Review of Educational Research* **58**, 47–77.
- Hobfoll, S. E. (1998). *Stress, Culture and Community*. New York: Plenum.
- Klenowski, V. (2002). *Developing Portfolios for Learning and Assessment. Processes and Principles*. London: RoutledgeFalmer.
- Liebert, R. M. and Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports* **20**, 975–978.

- Muir, S. and Tracy, D. (1999). Collaborative essay testing. *College Teaching* **47**, 33–36.
- Naveh-Benjamin, M., McKeachie, W., and Lin, Y. (1987). Two types of test-anxious students: Support for an information processing model. *Journal of Educational Psychology* **79**, 131–136.
- Panitz, T. (2001). *Learning Together: Keeping Teachers and Students Actively Involved by Writing across the Curriculum. A Source Book of Ideas and Writing Assignments*. Stillwater, OK: New Forums Press.
- Sarason, I. (1986). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology* **46**, 929–938.
- Slater, T. F. (1997). The effectiveness of portfolio assessments in science. *Journal of College Science Teaching* **26**(5), 315–318.
- Yerkes, R. M. and Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology* **18**, 459–482.
- Zeidner, M. (1998). *Test Anxiety*. New York: Plenum.
- Zeidner, M. and Schleyer, E. J. (1999). The big-fish-little-pond effect for academic self-concept, test anxiety, and school grades in gifted children. *Contemporary Educational Psychology* **4**, 305–329.
- Johnson, D., Johnson, R., and Stanne, M. (2000). Cooperative learning methods: A meta-analysis. <http://www.clcrc.com/pages/cl-methods.html> (accessed June 2009).
- McCombs, B. (2007). *Learner-Centered Classroom Practices and Assessments. Maximizing Student Motivation, Learning, and Achievement*. Thousand Oaks, CA: Sage.
- Pekrun, R., Goetz, T., and Titz, W. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist* **37**, 91–105.
- Schwarzer, C. and Buchwald, P. (2003). *Special Issue: Examination Stress. Anxiety, Stress, and Coping* **16**.

Further Reading

- Cizek, G. J. and Burg, S. S. (2006). *Addressing Test Anxiety in a High-Stakes Environment: Strategies for Classrooms and Schools*. Thousand Oaks, CA: Sage.

Relevant Websites

- <http://www.ncsu.edu>– NC State University, Silence Junction – Professional Development. Assessment Vocabulary and Differentiating Performance Assessment.

Impact of Assessments on Classroom Practice

J L Herman, University of California, Los Angeles, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Since its inception, assessment has been advanced as a tool for improving educational practice. Initially, assessment was thought to exert such influence through the information it provided to educators and policymakers and thus presumably to improve decisions. Over time, however, recognition of assessment's sociocultural and broader influences on teaching and learning has grown, and recently researchers have differentiated the effects of assessments of learning – the historical focus of assessment policy, practice, and research – and those of assessments for learning – the use of assessment during the course of instruction to directly inform and improve the learning process. These two types of assessment, respectively, summative and formative, operate from distinct theoretical bases and have unique theories of action, which are explored below. The article starts by providing perspective on the purpose and uses of assessment of learning and how these may impact classroom practice before considering the research base on assessment's actual effects on practice and the factors that influence such impact. Finally, the article addresses theory and research on how assessment for learning, so-called formative assessment, influences classroom teaching and learning practice and draws implications for future research.

A note about use of terms: this article uses the terms test and assessment interchangeably as denoting measures of student learning, although the former tends to denote traditional, multiple-choice-dominated standardized tests and the latter a broader array of methods. Similarly, a number of terms have been used in the literature to capture the concept of the impact of assessment, including assessment effects, consequences, washback, and backwash. These terms too are considered interchangeable.

Historical Perspective: Assessments for Selection and Improvement Purposes

The imperial examinations of the Han Dynasty (220–206 BC) are credited as the first use of standardized testing and the idea that gathering standard information on individuals' capabilities could lead to better and more equitable advancement decisions, in this case, decisions about coveted civil-service appointments. One can imagine that faced with the exam, elite, would-be candidates would prepare for the test and engage tutors to do so, a first instance of the impact of assessment on educational practices. Used for purposes of

selection, the Han Dynasty assessments also are a first example of how an assessment result can influence the practices in which individuals are subsequently placed.

From this very early precedent, educational assessment's use in selection thrived in the last century and continues today worldwide. Following World War I in the United States, for example, many school districts used standardized, intelligence tests for placing students into homogenous classrooms, reinforcing tracking schemes that provided lower-track students with less-rigorous instruction than higher-track students. Similarly, Great Britain's tradition of using assessment to certify, select, and track students for subsequent education dates back to the nineteenth century. Although more essay based than those in the United States, these set the stage for tests that motivate performance and with it the substantial influence of testing on educational practice.

Side by side with traditions of testing for selection and placement developed that of using assessment to improve educational effectiveness. The father of modern educational measurement, Edward Thorndike, advocated the value of testing for guiding teaching, and Ralph Tyler's 8-year study (in the 1930s) created assessments designed to support progressive education goals. With a new objectives-based framework as the foundation for development, Tyler articulated the ways in which assessment could support teaching and learning, for example, monitoring student learning to inform instructional planning; affording school- and program-level accountability and improvement; and in loosening the confining influence of traditional scholarship tests. That is, predating current interest in assessment consequences, Tyler recognized that traditional tests constrained instructional practice, and that new forms of assessment were needed for permitting new, progressive approaches to education to take hold. Similarly, like hope-for-performance assessment in the United States nearly 50 years later, Tyler's assessments were intended to educate the public about important goals of schooling and help teachers' clarify their instructional goals and practices, serving also to sow the seeds for changes in practice.

Assessment Impact in the Context of Overlapping Purposes

These fundamental, formal functions or purposes of measurement – to sort and select students, to certify accomplishment, to provide data for purposes of improving the quality of education – have coexisted since the

beginning of educational testing. Further, while measurement experts advise that tests be designed to serve a specific purpose, the reality is that current educational policy tends to intertwine multiple purposes, and thus large-scale assessments of learning can impact instructional practice through multiple pathways. Whether the context be the test-based accountability systems in the United States, or the operation of A- and O-level-type examination systems in many countries across the world, testing serves not only to provide technical evidence that informs multiple categories of decisions and policy functions (e.g., accountability, admissions/selection, placement, diagnosis, progress monitoring, certification, accreditation, and evaluation), but also attention to and use of the technical data carrying important sociopolitical meanings that can translate into important consequences for educational practice.

Consider, for example, high school exit exams common in the United States that students need to pass for a high school diploma; Key Stage, General Certificate of Secondary Education (GCSE), or A-level-type examinations in the United Kingdom; and higher education admissions tests that operate in countries across the globe. The ostensible purpose of the assessment is to certify that students have the knowledge and skills that the state has determined are needed for subsequent success.

As doing well on the assessment is important to students' futures, the existence of the test is expected to motivate students to learn, to prepare for the test, and to acquire the knowledge and skills that are assessed, as well as to motivate parents to pay attention and support their children's education. Moreover, even in the absence of tangible rewards and sanctions, educators will be motivated to help their students to do well, both because they care about their students' futures and because they see public reports of their students' test performance as evidence of their own professional efficacy. Accordingly, schools are motivated to do what they can to assure that their students will succeed and to do so may modify the content of their classes, may develop or purchase special materials to prepare students for the test, identify and establish new programs to help students succeed, etc. Instructional materials providers may respond by adapting existing materials or creating new products to meet the needs created by the test. In the case of high school exit exams, educators also may use test results of students who do not pass the test to place them in special classes, provide extra tutoring options, and if the test provides subscale data on different aspects of performance, they may use the results to diagnose individual students' learning needs.

Those same results aggregated at the class, school, or district levels may be used to analyze and respond to curriculum strengths and weaknesses, to suggest specific topic areas on which new or supplementary materials may

be needed and in which teachers may benefit from professional development, or to infer priorities for budget allocations. Actors at the district, regional, state, and/or federal levels may similarly analyze and respond to the same data at different levels of aggregation, identifying schools and districts that are doing particularly well or that are in most need of help, including deriving implications for system priorities, needing changes, and/or new programs. Thus, data from a single test may fuel analysis and decisions at multiple levels, which in turn have cascading effects on instructional practice.

As another example, take a state-mandated teacher-competency assessment that teachers must pass to gain tenure. By mandating such an assessment, state policymakers communicate their commitment to educational quality, their recognition of the importance of students having access to good teachers, and their responsibility for assuring the quality of the teaching force. As the assessment is developed, it makes concrete and communicates valued knowledge and competencies for beginning teachers. Just as secondary schools may respond to the requirements of high school exit or college-readiness exams, so too can teacher preparation institutions be expected to adapt their curriculum and instructional pedagogy to the expectations of the teacher assessment and to use their students' results to strengthen their program's success in enabling students to pass. To the extent that sizable proportions are not passing, an institution may change its admission procedures, create special test-preparation courses or tutoring opportunities, and/or try to pressure pre-preparation institutions to prepare potential candidates better. To the extent that the assessment delves deeply into meaningful aspects of teachers' content and pedagogical knowledge and requires demonstration of quality practice, preparing for the test may well help new teachers become more effective, and responding to the assessment requirements may strengthen and deepen existing preparation programs. In any event, unless it is a trivial assessment, the assessment may function to homogenize or standardize preparation programs and the content and pedagogical knowledge that teachers gain from them. As they can influence the what and how teachers learn in teacher-preparation programs, and who is allowed to enter the profession, such teacher assessments may substantially influence subsequent instructional practice.

These examples are not meant to be exhaustive, but rather to illustrate the confounding of test purposes, the merging or accountability/summative and improvement/formative functions, and of the technical, sociopolitical, and symbolic roles that high-stakes or high-visibility assessments play in influencing practice. A single test can launch a whole network of interrelated motivations and cascading actions, only some of which directly involve the technical use of the test.

These same overlapping functions also may accrue in classroom assessments of learning, particularly those that are used for grading. They too function to motivate and focus students' effort, communicate what is important to learn and how to do it and, for learning-oriented teachers and students, to stimulate reflection on what has been learned, how to fill the gap between that and intended-learning goals, and how to improve the teaching and learning sequence. (See also the section on formative assessment.)

Research on the Effects on High-Visibility External Tests

What of the actual impact of high-visibility tests? Researchers around the globe have studied this question, drawing largely on interview and survey methods, sometimes supplemented with observations of practice.

Consistent Results on Potential Benefits

The rigor of design and methodology vary substantially from study to study, but results have been quite consistent regardless of locale, for example, studies of state accountability tests in more than a dozen states in the United States, of A- or GCSE, or Key Stage Exams in England, and language and higher education admissions-testing programs in countries such as Australia, Central and Eastern Europe, China, Hong Kong, Israel, Japan, New Zealand, and Sri Lanka:

- Testing signals priorities for curriculum and instruction: high visibility tests serve to focus the content of instruction. School administrators and teachers pay attention to what is tested, analyze test results, and adapt curriculum and teaching accordingly.
- Teachers tend to model the pedagogical approach reflected on high-visibility test. When a high-visibility assessment is composed of multiple-choice tests, teachers tend to rely heavily on multiple-choice worksheets in their classroom instruction. However, when the assessments use open-ended items and/or extended writing and rubrics to judge the quality of student work, teachers prepare students for the test by incorporating the same types of activities in their classroom practice.
- Curriculum developers, particularly commercial interests, respond to important tests by modifying existing textbooks and other instructional materials and/or developing and marketing new ones to address test content and format. These products in turn may become primary resources that influence practice and also influence teachers' understanding of test expectations.
- Test results can make visible and promote responsiveness to the needs of students who previously have been

underserved. In the United States, for example, where all students are held to the same standards and schools must reach test-based accountability targets for all subgroups, research shows schools and teachers paying more attention to traditionally low-performing students, including second-language learners and students with disabilities. Test results are used to identify and provide special help for faltering students, for example, through the creation of new courses, purchase of new materials, and provision of extra school opportunities (e.g., before or after school tutoring, summer school).

Consistent Results on Problematic Effects

While the above points demonstrate some ways that assessment, depending on the nature of the test, can leverage productive changes in instructional practice, research shows unintended, negative consequences:

- Schools and teachers may focus on the test rather than underlying standards or learning goals. With sanctions and incentives riding on test performance, educators give primary attention to what is tested and how it is tested, rather than to the standards or learning goals the test is intended to measure. For example, by how much teachers may shift classroom-instruction time, accorded to particular topics or curriculum subjects, will depend on the content emphasis of the test.
- What is not tested can become invisible. As a corollary to focusing on the test rather than the standards, that which is not tested tends to get less attention or may be ignored all together. Both the broader domain of the tested disciplines and important subjects that are not tested may get short shrift. In the United States, for example, many state tests tend to give relatively little attention to complex thinking and problem solving and tend to focus on lower levels of learning, which can lead to similar emphases in classroom practice.
- Focusing on the test, rather than underlying learning, may encourage performance orientation and transmission-type teaching. When passing the test, rather than learning, becomes the goal, schools may unwittingly promote a performance orientation in students, which can work against students' engagement and persistence in learning, metacognition, and self-regulation. Especially for high-visibility tests that are predominantly multiple choice, teachers may concentrate on direct instruction to help students acquire specific content, rather than use pedagogy that helps students build conceptual understanding and problem-solving capability.
- Instructional time is diverted to specific test-preparation activities. Schools provide students with practice on the specific types of tasks and formats that are expected on the test, through commercial test-preparation packages,

special classes, and homework. Such activities aim specifically to help students do well on the test, rather than promote students' learning, and depending on the school and the pressure to improve test scores, can divert weeks or more of instructional time.

- Assessment-driven instruction can spur more and more testing. In the United States, more and more districts are mandating interim assessments during the school year, largely mimicking the expected content and format of their annual state tests, to provide quarterly or so feedback on how students are doing, and to encourage teachers to keep their eye on student progress on the knowledge and skills that will be tested. Moreover, many districts have become more prescriptive about how and what teachers are supposed to teach and have created pacing guides detailing what is to be covered when.
- Rather than motivating teachers to improve teaching and learning, high-visibility testing can cause teachers to feel pressured and demoralized, particularly when they view testing targets as unrealistic.
- Testing schedules can distort teaching and learning plans. Teachers even may adjust the amount of time they devote to basic subjects (e.g., reading, math) depending on when each subject is assessed. For example, if math is tested after grade 4 and reading in grade 5, then math may receive relatively more attention in grade 4 and reading in grade 5. Similarly, the best teachers may be moved to the particular grades at which important assessments are given. The desire to boost test scores thus can trump goals for teaching and learning in curriculum and personnel decision making.

Factors That Influence Assessment Impact

While research shows various potential effects of assessment on instructional practice, it also makes clear that such effects are not automatic; neither do they tend to be uniformly positive or negative. A first issue in shaping potential impact is the nature of the test itself. In fact, worry in the 1990s that traditional, standardized multiple-choice testing in the United States was producing a narrow, drill-and-skill curriculum led advocates to advance large-scale performance assessments to be, in the words of Lauren Resnick, "tests worth teaching to." By requiring students to engage in more authentic tasks that required complex thinking and problem solving and extended responses, performance assessment was intended to drive instruction in a similar constructivist direction, an early case of testing policy being explicitly used to leverage change. Similarly, more than 20 years ago, J. Charles Alderson identified the potentially powerful washback of language testing and argued that changes in testing could support desired changes in language curriculum and

teaching. The 1993 Hong Kong Certificate of Education in English, as an example, was launched to encourage secondary teachers of English to adopt a more communicative and purposeful approach, rather than a structural approach, to language development. The new assessment included more authentic and interactive tasks that required students to integrate listening, writing, reading, and/or speaking and to apply their skills in context. Or more recently, Great Britain's key-stage tests ostensibly are designed to encourage teachers to address a broad curriculum within each assessed subject, to enable students to respond to and create a variety of forms of writing and to engage students in applying mathematical and scientific knowledge and skills.

Yet, the path from potential or intended impact to actual changes in practice is neither simple nor clear cut. The move from policy mandate to operational test involves a process of implementation, from public dialog and public discourse about the meaning of the policy, to how and by whom the test is specified and developed, and through a complex chain of visible and invisible decisions and actions that may redefine initial intentions. From operational test to impact on practice similarly involves a circuitous and unpredictable process, as the concept of impact necessarily involves some change in practice.

Succinctly summarizing the many factors that influence it, theorists have noted that change requires both the will and the capacity to change and have suggested that policymakers hoping to use assessment to leverage changes in practice must consider the resources and strategies needed to support such change. From a motivational perspective, for example, in response to any new testing mandate, stakeholders may choose to accept and use it, ignore it, or fight it. Only those in the first category have immediate potential for intended effects and if change is to occur, action must be taken to convert those in the other two categories.

Yet, even with willingness, intended changes in practice will occur only when educators understand what is expected and have the prerequisite knowledge, skills, and understandings, as well as relevant resources, to make the change. As a new test rolls out, for instance, educators may differ in their understandings of what the test actually assesses and thus in its implications for their practice. Teachers will vary as well in their capability to engage in new, expected practices, depending on how complex intended changes are and how far from existing practice. Thus what teaching to the test, a common response to high-visibility tests, actually means depends on the beliefs and capacities of the educators who are responding and the resources and support they have available to them. In response to the same test, some schools and/or teachers may make meaningful and productive changes in teaching and learning, while others may react more rote with test-preparation activities simply mimicking the test.

Research on the consequences of assessment world-wide, in fact, indicates that while teachers may easily modify the content of their teaching, pedagogy is more resistant to change. Teachers' initial pedagogical responses tend to be superficial, for example, simply incorporating samples of the new assessments as an add-on to existing curriculum. Over time and only with sufficient support do new practices become more integrated in ongoing instruction and new standards and quality criteria more deeply understood and acted upon. Effects on practice, in short, are not the product of a test, but the result of a process that is influenced by stakeholders' attitudes and efficacy, the meanings they make of what is expected, the school context, power, or other relationships in which the change is embedded, the stakes riding on test results, the nature of the intended change, and the resources and strategies used to help teachers incorporate new practices, to name just a few. Assessments may be mandated top-down with the intention of accomplishing various impacts, but even bolstered by policy incentives and sanctions, actual impact is neither easily controlled nor easily predicted.

The Effects of Assessment for Learning: Classroom Formative Assessment

In contrast to assessments of learning that are mandated top-down, assessments for learning, so-called formative assessments, occur bottom-up, within the actual context of classroom teaching and learning. Similar to standardized testing of learning, however, the roots of formative assessment also can be found in ancient history, that of Socrates in fifth century BC, Greece. Socrates educated his students through questioning and dialog, breaking down important ideas into a series of questions, whose answers, with probing, gradually helped students acquire new insights. Questions and answers were not judged right or wrong, but were the means to building knowledge and understanding. Similarly in formative assessment, assessment is used to elicit students' understanding in order to form subsequent teaching and learning.

The last decade has seen increasing, worldwide interest in formative assessment, fueled in large part by Black and William's landmark meta-analysis showing the strong effects of formative assessment on student learning, particularly for low-ability students. Policymakers and educators increasingly have recognized that while external tests of learning may help communicate priority goals, motivate improvement, and identify who has succeeded or not – summative functions – the results are too little and too late to directly improve practice. Rather, the most powerful use of assessment occurs hand-in-hand with classroom teaching and learning.

The use of data is key to the idea: to be considered formative, assessment evidence must be acted upon during

the course of classroom instruction. Rather than focusing backward on what has been learned, formative assessment help chart the learning road forward, by identifying and providing information to fill any gaps between the learners' current status and goals for learning. Moreover, more than solely a source of evidence that informs subsequent teaching and learning, formative assessment may enhance student learning directly. For example, in asking students to make public their thinking, formative probes can provide scaffolding that helps students confront their misconceptions, refine and deepen their understandings, and move to more sophisticated levels of expertise. By making learning goals explicit and involving students in self-assessment, formative assessment also can promote students as agents in their own learning, increasing student motivation, autonomy, metacognition, as well as learning.

Formative assessment itself involves a change in instructional practice. Rather than imparting knowledge in a transmission-oriented process, in formative assessment teachers guide students toward significant learning goals and actively engage students as assessors of themselves and their peers. Formative assessment occurs when teachers make their learning goals and success criteria explicit for students, gather evidence of how student learning is progressing, partner with students in a process of reciprocal feedback, and engage the classroom as a community to improve students' learning. The social context of learning is fundamental to the process as is the need for classroom culture and norms that support active learning communities – for example, shared language and understanding of expected performance; relationships of trust and respect; and shared responsibility for and power in the learning process. Theorists observe that enacting a meaningful process of formative assessment influences what students perceive as valued knowledge, who can learn and who controls, and is valued in the learning process. However, as with the effects of assessments of learning on practice, those of formative assessment do not happen automatically but involve a complex process of change. Moreover, at this point, the anticipated effects of formative assessment on instructional practice are based on theoretical argument and are yet to be fully confirmed through empirical or other research.

Conclusion

Available theory and research suggest the ways in which assessments of learning and those for learning may exert influence on instructional practice. In both cases, such effects are not automatic but rather dependent on a complex process of change that is influenced by many, difficult to control, factors. Marshalling the power of assessment to improve practice requires better understanding of how such change occurs and the nature of the assessments

and support processes that can best achieve positive effects while minimizing unintended negative ones. Continuing research needs to explore whether high stakes and formative assessments are effective in meeting their intended aims, how effective each is relative to other strategies for influencing instructional practice, and whether and how assessment may undermine or support other teaching and learning goals, including effects on educational equity. Most studies thus far have examined short-term effects. Additionally needed are longer-term studies that examine implementation and effects over time.

See also: Classroom Assessment Tasks and Tests; Formative Assessment; Student Test Results in School Accountability; Summative Assessment by Teachers.

Further Reading

- Alderson, J. C. and Wall, D. (1993). Does washback exist? *Applied Linguistics* **14**(2), 115–129.
- Black, P., McCormick, R., James, M., and Pedder, D. (2006). Learning how to learn and assessment for learning: A theoretical inquiry. *Research Papers in Education* **21**(2), 119–132.
- Black, P. and William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan* **80**, 139–148.
- Cheng, L. and Curtis, A. (2004). Washback or backwash: A review of impact. In Cheng, L., Watanabe, Y., and Curtis, A. (eds.) *Washback in Language Testing: Research Contexts and Methods*, pp 3–18. Mahwah, NJ: Erlbaum.
- Cheng, L., Watanabe, Y., and Curtis, A. (eds.) (2004). *Washback in Language Testing: Research Contexts and Methods*. Mahwah, NJ: Erlbaum.
- Daugherty, R. and Ecclestone, K. (2006). Constructing assessment for learning in the UK policy environment. In Gardner, J. (ed.) *Assessment and Learning*, pp 149–168. London: Sage.
- Dwyer, C. A. (ed.) (2007). *The Future of Assessment: Shaping Teaching and Learning*. Mahwah, NJ: Erlbaum.
- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In Cheng, L., Watanabe, Y., and Curtis, A. (eds.) *Washback in Language Testing: Research Contexts and Methods*, pp 191–210. Mahwah, NJ: Erlbaum.
- Fuhrman, S. and Elmore, R. (eds.) (2004). *Redesigning Accountability*. New York: Teachers College Press.
- Gardner, J. (ed.) (2006). *Assessment and Learning*. London: Sage.
- Haertel, E. H. and Herman, J. L. (2005). A historical perspective on validity arguments for accountability testing. In Herman, J. L. and Haertel, E. L. (eds.) *Uses and Misuses of Data for Educational Accountability and Improvement*. 104th Yearbook of the National Society for the Study of Education (NSSE), 104(2), pp 1–34. Malden, MA: Blackwell.
- Harlen, W. (2006). The role of assessment in developing motivation for learning. In Gardner, J. (ed.) *Assessment and Learning*, pp 61–80. London: Sage.
- Herman, J. L. and Baker, E. L. (2009). Assessment policy: Making sense of the babel. In Sykes, G., Schneider, B., and Plank, D. N. (eds.) *AERA Handbook on Education Policy Research*, pp 176–190. New York: Routledge.
- Herman, J. and Haertel, E. (eds.) (2005). *Uses and Misuses of Data for Educational Accountability and Improvement*. 104th Yearbook of the National Society for the Study of Education (NSSE), Part 2. Malden, MA: Blackwell.
- Kirkup, C., Sizmur, J., Sturman, L., and Lewis, K. (2005). Schools use of data in teaching and learning. *DfES Research Report #671*. London: DfES. <http://www.nfer.ac.uk/research-areas/pims-data/summaries/schools-use-of-data-in-teaching-and-learning.cfm> (accessed July 2009).
- QCA (2007). *National curriculum tests: Building effective assessment*. Paper 2. www.qca.org.uk/libraryAssets/media/National_curriculum_tests_building_effective_assessment.pdf (accessed July 2009).
- Qi, L. (2004). Has a high stakes test produced the intended changes? In Cheng, L., Watanabe, Y., and Curtis, A. (eds.) *Washback in Language Testing: Research Contexts and Methods*, pp 171–190. Mahwah, NJ: Erlbaum.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* **18**, 119–144.
- Saville, N. and Hawkey, R. (2004). The IELTS impact study: Investigating washback on teaching materials. In Cheng, L., Watanabe, Y., and Curtis, A. (eds.) *Washback in Language Testing: Research Contexts and Methods*, pp 73–96. Mahwah, NJ: Erlbaum.
- Shepard, L. A. (2006). *Classroom Assessment*. In Brennan, R. L. (ed.) *Educational Measurement*, vol. 4, pp 623–646. Washington, DC: National Council on Measurement in Education and American Council on Education/Praeger.
- Stecher, B. and Chun, T. (2001). School and classroom practices in Washington state during two years of education reform. *CRESST Report #550*. Los Angeles, CA: CRESST.
- Wall, D. (2005). *The Impact of High-Stakes Examinations on Classroom Teaching*. *Studies in Language Testing* 22. Cambridge: Cambridge University Press.
- Watanabe, Y. (2004). Teacher factors mediating washback. In Cheng, L., Watanabe, Y., and Curtis, A. (eds.) *Washback in Language Testing: Research Contexts and Methods*, pp 129–146. Mahwah, NJ: Erlbaum.
- Willis, J. (2007). Assessment for learning: Why the theory needs the practice. *International Journal of Pedagogies and Learning* **3**(2), 52–59.

Relevant Websites

- <http://research.acer.edu.au> – Australian Council for Educational Research.
- www.CRESST.org – National Center for Research on Evaluation, Standards, and Student Testing.
- www.nfer.ac.uk – National Foundation for Educational Research.
- www.qca.org.uk – Qualifications and Curriculum Authority.

Social Practices in School Assessment and their Impact on Learner Identities

A Pollard, University of London, London, UK

A Filer, University of Cambridge, Cambridge, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In a review of sociocultural aspects of assessment, Gipps (1999) calls for more interpretive studies of the social world of the classroom. This contribution reviews work of this sort and offers a conceptualization of the factors in play. But why is this approach of consequence?

Accounts of the history of assessment practices show its roots in the selection and certification processes which were developed as national requirements for personnel standards emerged. Indeed, the first examinations are believed to date from the Han dynasty in China (206 BC to AD 220) and were used to select staff for government service. In seventeenth-century Europe, the Jesuits introduced competitive examinations within their schools and these became widely established in the late eighteenth and early nineteenth centuries as nation-states industrialized. The need for commonly understood and dependable standards of skill and knowledge in workplaces and the professions grew. Mass testing for service in the armed forces was thoroughly established by the two world wars of the first-half of the twentieth century, while the second-half saw school-outcome exams regulating entry with increasing precision to employment and more advanced study. In the twenty-first century, we now have many international standards and the market for specialized personnel is global.

Codified procedures and technologies for testing and assessment developed alongside these historic needs and applications. Fueled by early educational psychology, for instance, by studies of mental capacities and of intelligence quotient (IQ), these offered a scientific legitimation. Testing could be objective, valid, and reliable. People and their attributes and capabilities could be measured with accuracy. Examinations of various sorts thus provided an appropriate basis for decision making in the public interest. Such taken-for-granted confidence gave rise to a technical discourse of assessment in which required ends (e.g., practices of grading, comparing, and selecting individuals or schools on the basis of test scores) were not generally in dispute. Rather, the focus of concern became the means, whereby given ends could be achieved as fairly and objectively as possible. Within such a discourse, technicalities concerned with test validity and reliability, criterion and norm referencing, and so on tend to prevail. Such a discourse is thus about maintaining and improving

confidence in systems of assessment and results; thus, ultimately, legitimizing the uses to which they are put.

A contemporary educational critique of technicist assessment reflects concern with its consequences for learning itself. In particular, as this encyclopedia demonstrates, the significance of formative feedback as an integral part of teaching–learning processes can hardly be underestimated. Less constructively, the same is true of the capacity of high-stakes assessment to distort the curriculum and classroom practices. In any event, assessment as an integral part of teaching and learning processes, rather than simply as a source of input and output data, is now generating an important new discourse and body of knowledge. This directs attention to the social processes through which assessment is manifest.

Sociologists, such as Broadfoot (1996), have argued that assessment has a primary rationale in controlling access to positions in society. It is thus seen as a mechanism through which power and control is exercised and a means of social reproduction. In this analysis, testing and assessment are far from neutral or objective. Rather, they systematically enable and disable, enhance and retard. The sociological discourse of assessment also draws on social–psychological, anthropological, and some post-modern studies. It thus analyzes how, as well as having educational purposes, assessment fulfills a range of political and social functions within modern societies. These wider functions are concerned with social differentiation and reproduction, social control and the legitimizing of particular forms of knowledge and culture of socially powerful groups. In this, the discourse has a critical role in examining some of the myths and assumptions embedded in the activity of educational testing. It critiques the science of testing and offers insights into the fact that assessment, from the most formal to the most informal, takes place within social contexts. Assessment is conducted on, by, and for inherently social actors (William, 1997: 396) and so the social and cultural values, perceptions, interpretations, and power relations of assessors and assessed carry important implications for processes and outcomes. Sociological analysis is therefore particularly concerned with the social impact of assessment and the perpetuation of educational and social disparity, and its cumulative affects in shaping ways in which individuals and groups in society come to be seen, and to see themselves (Filer, 1995).

We have briefly reviewed assessment in terms of its historical evolution and its social consequence. We have also considered the technical, educational, and sociological discourses associated with it. However, in this contribution, our concern is mainly with the latter, sociological discourse, and with some of the ways in which it serves to illuminate aspects of those technical and educational discourses. In particular, we focus on a somewhat neglected dimension concerning the impact on individuals of assessment policies, procedures, and practices, and some of their hidden, as well as their taken-for-granted, roles in everyday schooling.

Assessment as a Social Process and Product

Assessment is not limited to the formal testing, grading, and classification of students. More expansively, the term also concerns ongoing, formative, and diagnostic judgments, day-to-day marking and recording, and numerous informal, often implicit, evaluations made of students' work, progress, and potential. It is also important to remember that many assessments explicitly or implicitly embody a number of social, emotional, and physical characteristics of students. Thus an assortment of behavioral, attitudinal, socioeconomic, cultural, and family characteristics often constitutes a social diagnosis in accounting for students' failure to make satisfactory progress or fulfill their potential.

But how does this occur? What social processes have this effect? Are assessment results really a social product? How can we make sense of these issues? Exploration of assessment as a lived experience is particularly important to this new strand of work. Thus, we are concerned with the sociocultural interpretations which pupils and teachers bring to their interaction, and to the differentiating consequences of classroom processes. There then follows the issue of how assessment outcomes are interpreted, contested, or otherwise mediated by learners and significant others and how the consequences of such outcomes are incorporated into future lives.

In recent years, we have begun to explore how a sociology of assessment could tackle such issues. A major source is Filer's (2000) edited collection, *Assessment: Social Practice and Social Product*, which identifies key themes within a sociological discourse of educational assessment and offers expert contributions to illuminate them. Another source is our collaborative theoretical development within *The Social World of Pupil Assessment* (Filer and Pollard, 2000) on which we draw below.

We will take each element of this model in turn:

- *When and where is the assessment taking place?* As we have seen, assessment policies, requirements, and prescriptions are culturally, structurally, and politically embedded in particular societies at particular times.

Comparative and historical studies have demonstrated significant variations over time and place, but it is also the case that modern, technologically and competitively driven societies are showing remarkable convergence in deploying the results of educational assessments as an evaluative and accountability device for the modern centralist state. We thus locate our analysis of social influences on assessment within specific sociocultural contexts of state policy, region, and community. That is to say, we ask the question, "Where and when is the assessment taking place?" In the case-study schools in which we carried out ethnographic empirical research over a 7-year period in the early 1990s, the contexts in question were those of two southern English primary schools serving what we would contrast as white, middle-class, and skilled working-class families. In these case studies, we were particularly concerned with the introduction of teacher assessment (TA) and standard assessment tasks and tests (SATs) into primary schools in England. Over the period, tests, administration, marking, and moderation procedures were introduced and standardized. Results were scrutinized from year to year to ensure that consistent standards were being maintained. Through such processes, assessment practices were legitimated and assumptions of impartial objectivity were established.

- *Who is being assessed?* This question focuses on the identity of the learner being assessed. We see self-perceptions held by individuals and judgments made about individuals as being inextricably linked to the social relationships through which they live their lives. Indeed, we argue that we can only make sense of individuals and their learning within the context of that dynamic relationship with the wider world. Of course, there are factors that are internal to the individual in terms of capacities and potentials. However, the realization of a pupil's capacity and potential is, to a very significant extent, a product of external social relationships and cultural circumstances, among which we can count the processes and outcomes of assessment, in all its various forms. In our case studies, we observed, year on year, formalized and informal representations of children as social beings, as learners, and as pupils. The succession of objective, highly personalized, and invariably confident accounts were, of course, very powerful in shaping pupils' emerging and ongoing classroom and learning identities, as well as their relationships with parents and peers. Equally inevitably, however, those assessments were partial and contingent, obscuring the ways in which individual behaviors, relationships, attitudes to work, and perceived intelligences were, in part, products of the social, emotional, and organizational aspects of classroom life that teachers themselves created. Despite the confidence that most primary teachers held regarding their

all-round knowledge of individual pupils, we found it is extremely problematic for a teacher to know the whole child for assessment purposes – as we illustrate later in this article through the story of a pupil, Elizabeth, and her primary-school career.

- *Who is assessing?* Having argued that pupil identity can only be understood in context, we clearly need to focus on teachers, since they are undoubtedly the most powerful classroom participants with whom pupils must interact. In particular, we need a sociological conception of pedagogy and its link to each teacher's own sense of personal identity. For this, we deployed the concept of coping strategy (Woods, 1977; Pollard, 1982) and traced how satisfying role expectations and the constant pressures of teaching must be balanced with maintaining a sense of personal integrity and fulfillment. In the immediacy of classroom dynamics, this can be seen as teachers juggle with the immediate pressures they face. At the level of the school, it is played out through negotiation between different interest groups and the formation of taken-for-granted institutional assumptions. We traced how teachers in the case-study schools coped with the national curriculum and assessment requirements which progressively challenged their autonomy and traditional practices through the 1990s. A case study of one teacher, Marie Tucker, and her classroom practice demonstrated the detailed application of this analysis. In particular, it shows how her coping strategies, classroom organization, and associated pedagogies produced particular contexts with which pupils, in turn, had to cope. In particular, Mrs Tucker prided herself on her reputation for maintaining classroom order and routine, and for being good with problem children. Indeed, she routinely perceived and assessed pupils in terms of their adaptation to her personal strategic criteria. Inevitably, requirement for new classroom and assessment practices constituted a very real threat to Mrs Tucker's ability to maintain her established order and routine. At the same time, she began to perceive her many problem children, less as a source of professional pride, and more in terms of their threat to her grip on classroom control, and to the reputation she was proud of – something rather different from assessment based on objective performance.
- *What is being assessed?* An official answer to such a question might point to the subject content of a test, or to listed criteria of judgment, and would draw conclusions in terms of the attainment of pupils. More colloquially, inferences about the particular abilities of children may be legitimated by faith in the objectivity and categoric techniques of standardized assessment. In our analysis, we argued that such confident conclusions are misplaced because pupil knowledge, skills, and understandings are embedded in particular, and often conflicting, sociocultural understandings and are further conditioned by factors such as gender, ethnicity, and social class. We particularly focused on the influence of peer-group relationships and the ways in which peer culture and the sociocultural identity of each pupil can condition performance. Thus, while pupils' subject knowledge, skill, or understanding may seem to be objectively revealed by the neutral, standardization technique of a test or by a classroom task or teacher questioning, tests and tasks also reveal the facilitation or constraint of sociocultural influences and forms of understanding. We illustrated this proposition through analysis of a year-3 news session at Albert Park Primary School. This showed how classroom meanings were created through interaction of circumstances, strategies, and identities. In particular, it showed how language was used to satisfy pupil agendas for entertainment, novelty value, and peer status in ways that were often in conflict with teacher-led instruction or approval. Assessment, the analysis suggests, can thus never tap pure knowledge or capability – any result will also always reflect the wider sociocultural circumstances of its production. Assessment is an interpretive process. For this reason, beyond academic subject matter, we must ask what else is being assessed?
- *How does assessment function in classrooms?* Our analysis foregrounds the links between assessment and other sociologically important influences on classroom life – ideology, language, and culture. As a whole, these factors are played out through particular power relations between teachers and pupils and have significant consequences for social differentiation. We explored these ideas by drawing on some of Bernstein's (2000) work. This highlights the ways in which classroom language is conditioned by patterns and forms of control, which are embedded in teachers' routine and everyday practices. The consequence is that it is not possible for teachers to be neutral either in their impact on pupil performance or in their assessment of pupil performance. We show how, irrespective of intentions, each teacher generates a particular set of circumstances in which interaction with each child takes place, and which will affect each child differentially. The scope for variability in the overall effect is enormous.
- *How are assessments interpreted and mediated?* Assessments are produced and delivered within networks of social relationships. We thus need to consider the audiences of assessment and, in school contexts, this requires particular reference to families and to peers. These are the major significant others in children's lives and we therefore traced their influence throughout the 7 years of schooling in the case-study primary schools. More specifically, we considered how families interpret, mediate, and give meaning to assessment outcomes, so that their impact on their child is shaped

and filtered. Once again, we found that the outcomes of assessment cannot be seen as categoric and direct in their consequence. Rather, their meaning is malleable and is likely to be drawn into existing frames of reference, relationships, and patterns of social interaction. For each learner, this is an extremely important process in the development of further phases of their personal narrative and in the construction of identity.

Overall, in relation to each of the major questions set out in the simple, cyclical model in **Figure 1**, we emphasize the influence of social factors on assessment. Learner, assessor, focus, process, and interpretation are all embedded in their sociocultural context and caught up in webs of social relationships. In such circumstances, we believe that the technical objectivity of assessment is a myth too far.

The Case of Elizabeth

We can illustrate the social processes and products of assessment through the case study of a young learner, Elizabeth – the only child of Eleanor and John Barnes, living in an immaculately kept, small, modern terraced house in the area of skilled, working-class families served by Albert Park Primary School. How did school assessment influence her life?

We described Elizabeth's school career at some length in *The Social World of Pupil Assessment* and recorded, year on year, her emerging identity and self-esteem as a pupil and as a girl. We saw her evolving classroom status and relationships among teachers, peers, and within the family, and her evolving approaches to academic tasks from

age 5 to age 11. Elizabeth was a lively child with a strong will. However, while her classroom playfulness and boldness allowed her to develop an attractive identity among her peers, her parents and many of teachers repeatedly expressed their disapproval of these aspects of her performance as a pupil and as a girl. These social and behavioral outcomes developed in a dynamic relationship with assessment outcomes concerning attainment and progress levels, etc. These in turn, of course, culminated in her SAT and teacher-assessment scores at age 11, and informed her progression to secondary education.

Social and academic outcomes were thus interrelated and, in addition to its ostensible role in supporting learning and monitoring outcomes, classroom assessment made a significant contribution to socializing Elizabeth into school life. As in the work of Torrance and Pryor (1998), we recorded this early socializing process in Elizabeth's career when, as well as being concerned with her academic progress, TAs (Teacher Assessments) and parental responses to assessment, focused on monitoring and supporting her adaptation to school expectations. While Elizabeth's intellectual and physical skills and progress across the curriculum were assessed as good, both formal and verbal reports to her parents were dominated by a range of negative assessments of her as noisy, quarrelsome, and disruptive with peers, especially boys. Such assessments led to friction, punishments, and warnings at home. As we have described above, in relation to Mrs Tucker, such socializing processes are linked to issues of control and the maintenance of a classroom environment in which effective teaching and learning can develop. Thus, the socializing and control functions of assessment in Elizabeth's case were related to her teachers' interests and

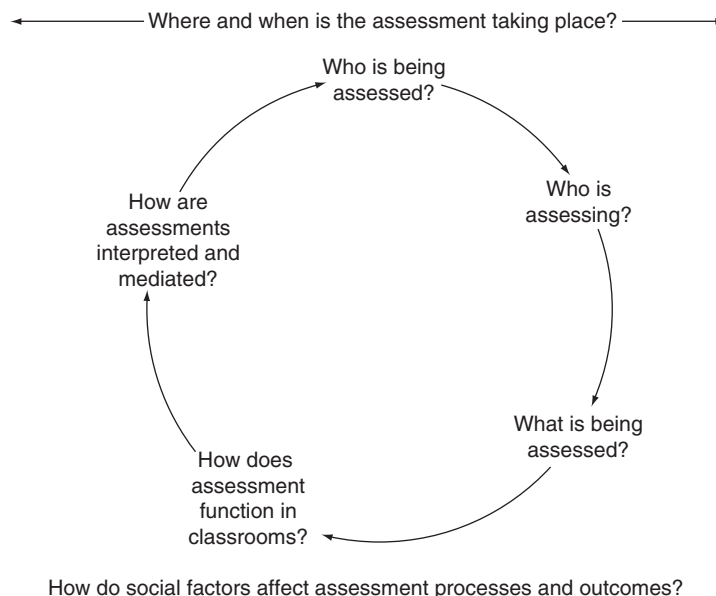


Figure 1 Social influences on assessment: A simple model. From Filer, A. and Pollard, A. (2000). *The Social World of Pupil Assessment*. London: Continuum.

to other children's learning as much as with Elizabeth's development as a pupil. However, assessment of Elizabeth's behavior and relationships went beyond straightforward conceptions of her as a pupil. The problem of her physical liveliness and association with boys, especially naughty boys, was a recurring theme in early TAs of Elizabeth, and a cause for further conflict and unhappiness at home. While accepting her daughter's identity as a girl – "that *is* Elizabeth, that's *her*" – her mother nevertheless strived to support teacher attempts to curb her daughter's boisterous association with boys, wishing, she admitted, that Elizabeth was more of a dainty child, and expressing the notion that her daughter, with all her enquiring energy was more like a boy. However, in the context of peer relationships, year-on-year sociometric questioning repeatedly revealed Elizabeth to be one of the most popular children in the class, for her playfulness, confidence, and boldness, for her artistic skills and willingness to help, academically. Further, she achieved this status on account of her appeal to boys, as well as girls. Thus, while continual attempts were made by her teachers, and on behalf of her teachers by her mother, to socialize Elizabeth into appropriate pupil behavior, and especially to socialize her as a female pupil, these efforts were in direct conflict with socializing process within the peer culture.

Thus, Elizabeth's case makes explicit some of the gendered norms and values that can remain implicit in assessments of pupils' behavior unless a pupil persistently contravenes them, as Elizabeth did. One of the functions of assessment in Elizabeth's life, therefore, was in providing a form of communication through which gender expectations were shared and communicated among her teachers, her parents, and her peers. While implicit norms relating to gender were made explicit here, other norms and values relating to culture and ethnicity of pupils similarly remain implicit until challenged. For example, Adams (1997), writing within the US context, presents the case study of a white adolescent girl, Sharon, who contravened gender and racial norms. Sharon's toughness and powerful identification with black gang culture gave rise to sustained criticism, in approaches to her mother and before her peers, as teachers tried to re-socialize her into an appropriate pupil identity.

Reviewing again the question of who is being assessed, we can note that Elizabeth's resources and potential with regard to her physical and healthy attractiveness and liveliness and her linguistic, physical, and artistic skills, were highly valued within the peer group. However, many of her teachers felt ambivalence regarding her social identity. For instance, her year-2 teacher made overt and strenuous attempt to curb Elizabeth's deviant pupil behavior, while recognizing that those same qualities made her attractive to her peers, exciting, and funny. Indeed, she recognized that Elizabeth's liveliness, lack of conformity, and well-developed sense of humor made for

an extremely vital and attractive social identity, among adults, as well as children. However, for many of her teachers and her mother, they also stood in the way of an acceptable pupil identity and, in some respects, in the way of an appropriate identity for a pupil who was also a girl.

Figure 2 is an elaboration of the simplified model previously introduced. Some elements of it have been referred to in Elizabeth's story, but others would repay further study within the original sources.

Conclusion

The conceptual model which has been introduced is an attempt to represent important, recursive cycles which underpin the social consequences of schooling. In particular, it highlights the ways in which identity, teacher practices, peer and family cultures, and the interpretations and mediations of teachers, family, and peers feed into, reinforce, and condition each other. Importantly too, of course, we have suggested that the microprocesses of pupil and teacher identities, practices, and interpretations are embedded in macro-sociohistorical and political contexts which shape the policies of schools and the perceptions and expectations of teachers and families.

This type of model was initially developed through two related studies, *The Social World of Children's Learning* (Pollard and Filer, 1997) and *The Social World of Pupil Career* (Pollard and Filer, 1999) and the recursive cycles represent the analytic insights generated through longitudinal tracking of pupils' identity, learning, and school careers.

The significance of this approach to assessment is that power, process, and consequence are foregrounded. Detailed study of assessment practices reveal patterns of social influence which question the aspiration to objectivity. Research on classroom interaction, identity, and relationships suggest that assessment outcomes are tainted and should be, at least in part, seen as social products as well as products reflecting purer forms of capability or achievement. Holistic enquiry into the social mediation of assessment outcomes by significant others shows the enormous variability of consequence for pupil identities and self-perceptions as learners.

A social analysis of assessment practices and products thus generates concerns in respect of the categoric, summative certification of performance. Indeed, we could go further and argue that established assessment practices often yield patterns and systematic effects which risk being fundamentally divisive. As policymakers configure education systems to meet the demands of international competition, they may also unwittingly reinforce social divisions and widen the life-chance gaps which some groups of children face.

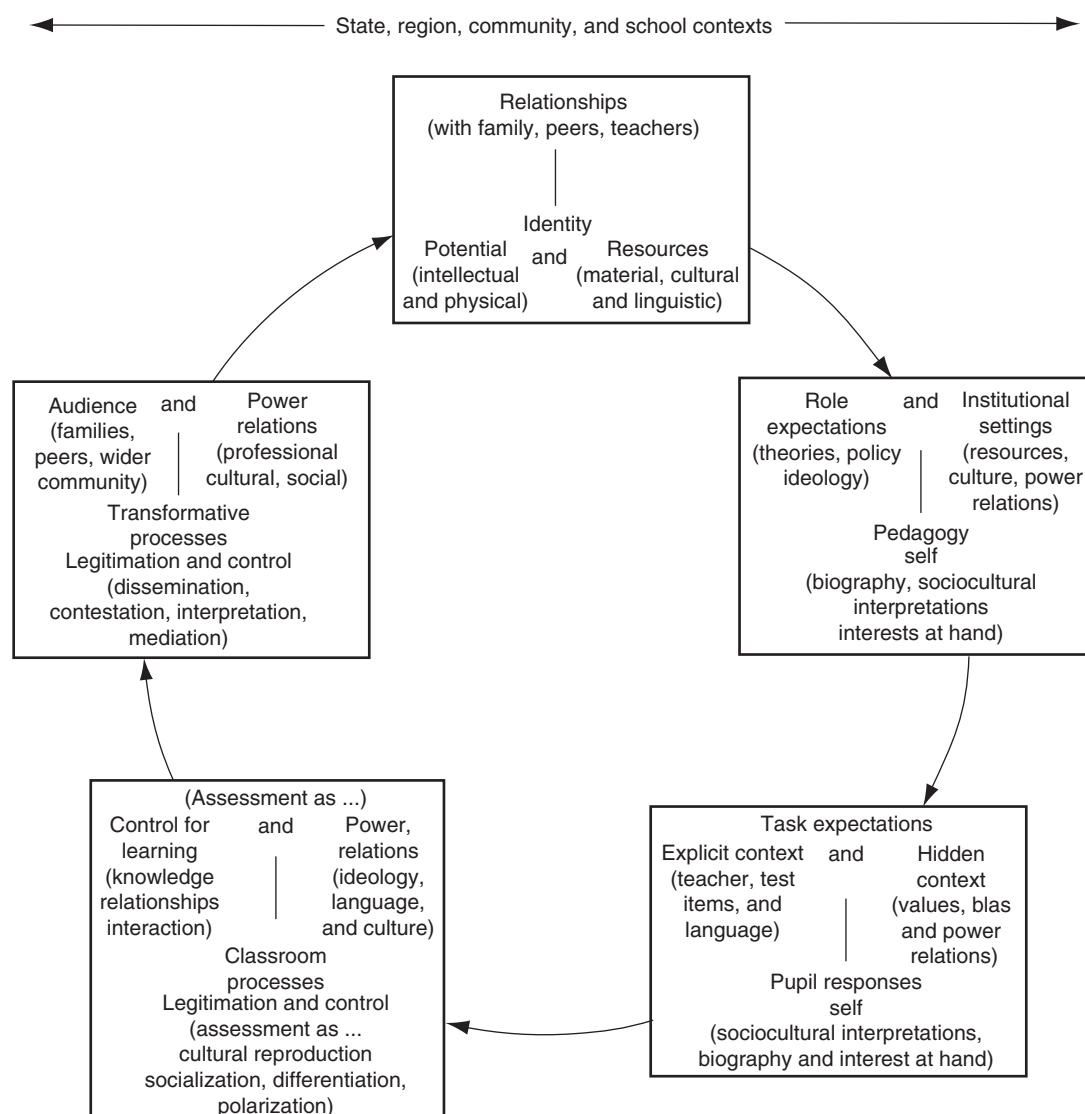


Figure 2 Social influences on assessment: A developed model. From Filer, A. and Pollard, A. (2000). *The Social World of Pupil Assessment*. London: Continuum.

Bibliography

- Adams, N. (1997). Towards a curriculum of resiliency: Gender, race, adolescence and schooling. In Marshall, C. (ed.) *Feminist Critical Policy Analysis*. London: Falmer.
- Bernstein, B. (2000). *Pedagogy, Symbolic Control and Identity*. London: Routledge.
- Broadfoot, P. (1996). *Education, Assessment and Society*. Milton Keynes: Open University Press.
- Filer, A. (1995). Teacher assessment: Social process and social product. *Assessment in Education* 2(1), 23–38.
- Filer, A. (ed.) (2000). *Assessment: Social Practice and Social Product*. London: Routledge.
- Filer, A. and Pollard, A. (2000). *The Social World of Pupil Assessment*. London: Continuum.
- Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education* 24, 353–392.
- Pollard, A. (1982). A model of classroom coping strategies. *British Journal of Sociology of Education* 3(1), 19–37.
- Pollard, A. and Filer, A. (1997). *The Social World of Children's Learning*. London: Continuum.

Pollard, A. and Filer, A. (1999). *The Social World of Pupil Careers*. London: Continuum.

Torrance, H. and Pryor, J. (1998). *Investigating Formative Assessment: Teaching, Learning and Assessment in the Classroom*. Buckingham: Open University Press.

William, D. (1997). Review of Broadfoot (1996). *British Educational Research Journal* 23(3), 396–397.

Woods, P. (1977). Teaching for survival. In Woods, P. and Hammersley, M. (eds.) *School Experience*. London: Croom Helm.

Further Reading

Torrance, H. (ed.) (1995). *Evaluating Authentic Assessment*. Buckingham: Open University Press.

Relevant Websites

<http://www.assessment-reform-group.org> – Assessment-Reform-Group.
<http://www.tltp.org> – Teaching and Learning Principles into Practice.

EDUCATIONAL EVALUATION – CONCEPTS, PRACTICE, AND FUTURE DIRECTIONS

Educational Evaluation: Concepts, Practice, and Future Directions

Y J Ni, The Chinese University of Hong Kong, Hong Kong SAR, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Functions of educational evaluation –

Educational evaluation in general can serve the function of describing what happens and how and why it happens, judging whether or not what happens meets the expectations, and then recommending what actions to take. However, evaluation practioners and evaluation researchers may view differently on the matter whether or not a particular evaluation should or can serve the three functions. Evaluation practioners tend to think an evaluation needs to fulfill these functions in order to meet the needs of the targeted evaluation users, whereas evaluation researchers tend to think that to provide an accurate description is most important for an evaluation. And judgment and recommendation will be constructive only when such descriptions are verified and collaborated with other and relevant replications.

Generalization of evaluation studies – The extent to which that a conclusion drawn from an evaluation study or a set of evaluation studies can be generalized to a different situation. There is disagreement with regarding whether or not evaluation study needs to concern about the generalization as the debate continues whether evaluation should be focused on particularization or on generalization. However, the author of this chapter thinks that choosing between the two is false.

Outcome-based accountability – Individuals (e.g., teachers) or institutions (e.g., schools) are asked to be accountable not only for what they do, but more importantly for whether or not what they do produce student learning outcomes, mainly achievement test scores. The outcome-based accountability is controversial especially when it is narrowly conceived as merely indicated by student test scores.

No child left behind act – The US legislation passed in 2001 requires the testing by the states of all children in third through eighth grades. Schools will be judged to be successful or in need of improvement based on student performance on the states' administered standardized assessments. This Act enacted the theories of standard-based education reform, which has raised prospects for public education, as well as problems and controversies.

Professional accountability – Regulations by a professional body to inform and to make the individual professionals to conduct the profession in a socially and ethically responsible way.

Program evaluation standards – The Program Evaluation Standards (1994, second edition) was published by the Joint Committee on Standards for Educational Evaluation of USA and Canada. The publication explicates four categories of standards concerning utility, feasibility, propriety, and accuracy of evaluation for evaluators and evaluation users to respect and to reflect in conducting educational evaluations and using evaluation results.

Regulatory accountability – Government regulations of schools, school personnel, and school practice to evaluate how well the schools and school personnel perform their institutional duties as specified by legislations and as expected by the society. School inspection by government-appointed assessors is an example.

Role-based accountability – Schools, school principals, and teachers are assigned institutional roles in the education of young people. Role-based evaluation establishes evaluation criteria based on job descriptions for a principal or for a teacher and uses the criteria to judge how well the principal or the teacher performs the institutionally expected roles.

The office for standards in education, children's services and skills (OFSTED) – OFSTED was

established in 2007 from a merger between the old Office for Standards in Education and the Adult Learning Inspectorate under the Education and Inspection Act 2006 of British. As a non-ministerial government department reporting directly to the Parliament, OFSTED is responsible for regulating and inspecting schools and other education and childcare services including adult learning and training, children's social care services, and initial teacher training in England.

Theory-driven evaluation – The evaluation approaches that emphasize program theory of why a program would produce expected changes and program implementation of how the program would achieve changes, planned or not.

An item of recent news from China caught the attention of the author while she was working on this article, that she was compelled by its significance to the present topic. On 30 June 2009, the Mainland Chinese government announced a delay in implementing a mandated installation of website-blocking software Green Dam Youth Escort into all computers sold in the country. This initiative – planned to commence from the first of July – was to block pornography as well as information on the Internet that was officially designated as undesirable. Implementation was delayed because of protest by hundreds and thousands of Internet users throughout the country. An incident such as this, forcing the government to stop what had been decided (also announced) under pressure from ordinary people – Internet users – was unprecedented in modern Chinese history. There are significant ramifications for this situation as it demonstrates the rise of civil society in Mainland China. It is also an indication of the extent to which information technology has helped to change modern China and the world. In relating this turn of events to the topic of the present article, a parallel could be drawn. As was the case with information technology, the development of educational evaluation practices and theories has been nurtured and has contributed to the rising and maturing of civil societies around the world.

Educational evaluation has become an integral part of civilized society because – as a social institution – its role is to judge the merit, worth, value, and impact of educational enterprise. Any educational evaluation engages consistent inquiry into what is good and the reason for its elevated status. The value feature distinguishes evaluation from other types of inquiry. Evaluation is also an empirical endeavor committed to finding evidence and the establishment of its credibility via a systematic procedure. The process must define and defend the judgment on the

merit of an educational enterprise. Educational evaluation has developed on these two fronts. In the last 15 years – as an indicator – the publication of the *Program Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 1994) demonstrates the maturity of the field of educational evaluation as a field and a profession. Under its guidance, there have been changes that are both significant and numerous. However, these have been achieved without compromising the ultimate purpose of educational evaluation – namely evaluation for accountability and for learning improvement. It could be said that educational evaluation is multifaceted, hence the need to determine, precisely, what is meant by effective accountability and effective learning improvement – which has driven value and methodological inquiries in the field and in the process – shaped its current landscape. The section 'Educational evaluation' is intended to reflect on the major developments in the evolving field of expertise that has taken place over the past one and half decades. This article aims to provide an overview of this section. Initially, the article presents an outline of the content and structure of the section. Subsequently, details are provided that illustrate the evolution. Finally, some sustained issues and challenges in the field are identified with suggestions for future development in educational evaluation.

Content and Organization of the Section

By utilizing Schwab's conceptualization of knowledge organization (Schwab, 1978), the section 'Educational Evaluation' is organized into five areas: purposes and models of educational evaluation, contexts of educational evaluation, evaluation methods, evaluation domains, and contemporary themes in evaluation. The first area consists of five articles that contribute to the objects and concepts of educational evaluation and theoretical approaches to educational evaluations. The second area – relating to contexts – includes seven articles and addresses influences of the circumstances, such as social, political, cultural, and ethnical aspects – factors that can affect educational evaluation. The third section – methods of evaluation – consists of four articles that describe various research tools with which to conduct educational evaluation. The fourth section deals with evaluation domains and contains 14 articles that discuss the theoretical constructs/models and the tools to attend to the major domains for educational evaluation. These include (but are not limited to), curriculum evaluation, program evaluation, school evaluation, educational personnel evaluation, educational policy evaluation, national evaluation programs, and international evaluation. The final area is one that deals with contemporary themes and includes seven articles on the emerging educational practices and their

evaluation, such as formative evaluation of teacher and teacher education, evaluation of educational technology and e-learning, and evaluation of integrated school health programs.

In terms of Schwab's conceptualization of knowledge organization, the first two parts correspond to the dimension of boundaries (the objects and the concepts) of the field; the third part addresses the dimension of methodology in relation to the objects and concepts of evaluation; and the last two parts relate to the domain or emerging applications of the objects, the concepts and the methodology in the field. The section contains a total of 46 entries, with 38 entries being full papers and eight using cross references of the articles from other sections of the International Encyclopedia of Education (IEE). The intention is for the content and its organization to provide the reader a relatively comprehensive perspective of the contemporary educational evaluation as a field.

Evolution of Educational Evaluation in the Last 10–15 Years

There is, undoubtedly, consensus among authors contributing to this body of research that educational evaluation has changed significantly over the last 10–15 years, but they may disagree with how to characterize these changes. Based on reading each of the contributions and others, descriptors have been identified and will be used in this article to capture the collective reflections of the contributions on these changes to educational evaluation. The descriptors are as follows: outcome-based accountability, stakeholder involvement, globalization of educational evaluation, diversity in value and method repertoires, and theory-driven evaluation. In the elaboration of each of the characterizations below, every effort was made to identify specific meanings and implications relevant to the field.

Outcome-Based Accountability

Educational evaluation is always about accountability since it obtained the legitimate status when the US congress passed the 1965 Elementary and Secondary Education Act (ESEA) – a document that includes an evaluation clause requiring the evaluation of federally funded programs. Evaluation for accountability is one key reason for evaluation to stay. That being said – in terms of concept and practice – educational evaluation has been transformed in a significant way as public education faces challenges from a number of political, social, and cultural developments. This impression has been reinforced by reading the articles of the section, particularly those such as 'School inspections' (Chapman and Earley), 'School-based evaluation' (MacBeath), 'Principal evaluation' (Glassman and Glasman), 'Evaluation of teacher

quality and practice' (Robinson and Campbell), 'Evaluation and accountability' (Sato and Rabinowitz), 'Evaluating education in three policy eras (Peters, Weber, and Britez)', and 'Purposes of educational evaluation' (Mathison). The transformation has taken place because evaluation for outcome-based accountability has become statutory and, in reality, had material consequences. This process began in the 1980s and was facilitated by increasing public dissatisfaction with public education. There was also a trend toward global neoliberalism that conceptualized education less as a process and more as a product or outcome. One of the key terms – outcome-based accountability – is in contrast to earlier education policy when accountability was more about what schools and teachers do than how student-learning outcomes are influenced by what they do. Another key concept is regulatory accountability – mandatory by governments for individual schools, classrooms, and teachers and in contrast to professional accountability which is regulated and highly regarded by the teaching profession – and, consequentially, in contrast to what was considered inconsequential. These changes have obtained prominence and become an integral part of educational systems in the last two decades. Two outstanding examples of this include the No Child Left Behind Act of 2001 in the US and the Education and Inspections Act in England which, in 2007, established the Office for Standards in Education, Children's Services and Skills (OFSTED). This important legislation was responsible for the systematic inspection of schools (and other providers of education and childcare) in the country.

Many articles found in this section provide lively and thoughtful accounts of this transformation in various aspects of education. Most of the authors see the potential benefits of this transformation while pointing to the critical issues brought about by the changes. Among the desirable effects brought about by the change, a school's operation and performance have become more explicit and transparent – not only to the parents and the public but also to schools, teachers, and students. Making school practice and performance explicit, discussable, and subject to a valid form of assessment and measurement has encouraged schools to recognize their own strengths and weaknesses. The emphasis on evidence of student-learning outcomes promotes and strengthens both internal and external accountability of the stakeholders in school education. For example, it has been suggested that the quality-assurance-evaluation system has greater validity for evaluating university teaching's impact on student learning and development than the common practice – one that relies on anonymous student evaluations. These have been found to reflect tutors' popularity, personality, and the ease of higher-grade achievement more than teaching quality. A teacher's promotion and career development can be equitably supported when it is based on evidence of the teacher's performance and its impact on student learning rather than on the

informal – or even nepotistic – arrangements for promoting teachers. The concern for evidence of student-learning outcomes also prompts the research and development designed to capture the connections between teaching and learning. (Under this section see articles by Kyriakides, Creemers, Teddlie, and Muijs; MacBeath.)

Along with the potential benefits, there are also questions and problems to consider that are associated with the outcome-based, regulative, and consequential accountability. First, it is necessary to identify what outcomes require accountability. Examples of typical educational outcomes are academic achievement, school attendance, school completion, graduates' employability, civic participation, social development, and so on. These outcomes require no justification. However, educational outcomes are unlike products of business organizations that are clearly defined and measured. Educational outcomes (or standards) vary considerably. They can be described as short or long term, intended or unintended, and singular, multiple, or hierarchical. For example, there is multiplicity of outcomes resulted from integrated school health programs, (see Franke; this section); complexity in identifying educational outcomes in cost analysis in evaluation (see Fletcher, this section). Consequently, it is, often, difficult to establish direct links between outcome and cause in educational settings (Scriven, 2005; Schwab, 1978). It remains very problematic in research to disaggregate the value added to student learning by individual teachers from that added by school factors (Reynolds *et al.*, 2003). With the difficulties in establishing direct links between student learning outcomes and actions of individual schools and individual teachers for rewards or punishments (e.g., performance-related pay arrangement for teachers, school closure because of insufficient student enrollment), connecting that with the outcome-based accountability is inevitably controversial. In addition, the regulative accountability has placed constraints on professional autonomy. This is a particular problem where working conditions in schools are more likely to be imposed from outside the profession, usually by government. This causes lower morale and learned incompetence in teaching professions, which, in turn, defeats the very purpose of outcome-based accountability intended to improve school effectiveness and student learning.

To address these concerns, three key relationships have been proposed by several contributors. These need to be addressed in practice so that evaluations for outcome-based accountability have a chance to succeed. The first of these is the relationship between professional and regulatory accountability, followed by between role-based and outcome-based accountability, and between bottom-up process and top-down formulas of accountability policy and implementation. With regard to the first relationship, Mathison (this section) discusses the necessity and limitations both of regulatory and professional accountability. Therefore,

the proposal is for pluralistic democratic accountability to balance the two. The pluralistic democratic approach to accountability requires educational stakeholders, including students, parents, government, schools, principals, teachers, to enter into a contract of mutual, collective responsibility for education and schooling. This places greater emphasis on local context and internal accountability – that is, mutual responsibility within contexts and communities.

The second key factor relates to a balance between role-based and outcome-based accountability. Schools, school principals, and teachers assume a significant institutional role in the education of young people – empowering them to become self-sufficient individuals and responsible citizens. Therefore, the focus of criteria should be whether or not teachers or principals fulfill their institutional roles for teacher or principal evaluation. In the article 'Principal evaluation', Glassman and Glasman explicate the complexity in role-based or outcome-based evaluation of school principals. They argue that role-based evaluation (assessing the qualities of the activities) needs to be the primary component of principal evaluation although outcome-based evaluation (assessing the results of the activities) presents an increasing pressure. Role-based evaluation establishes evaluation criteria as based on job descriptions for a principal or a teacher – that is, the roles of teachers or the roles of principals and how well they perform the institutionally expected roles (e.g., judging the principal's knowledge and skills with regard to managing the school's finances effectively and efficiently). Outcome-based evaluation sets evaluation criteria in terms of student-learning outcomes. Role-based and outcome-based evaluation should not be exclusive. Performing the institutional roles by principals or by teachers is believed to bring about expected educational outcomes in students. School principals and teachers may resign if they believe their efforts have nothing to do with student-learning outcomes. Role-based evaluation enhances the sense of agency among principals and teachers as professionals. Outcome-based criteria – for example, percentage of students that score at, or above, expected levels – when appropriately conceived by principals and teachers, or by a school as a whole, will provide a sense of purpose and encouragement for principals and teachers to perform their expected roles.

The third factor has as its focus the balance of the bottom-up and top-down process in formulating and implementing policies of evaluation for accountability (Chapman and Earley; MacBeath; Nevo, this section). School-based internal evaluation or teacher self-evaluation has been considered a good exemplar of pluralistic democratic accountability (Mathison, this section). This process should be initiated and governed by the school or by the teachers. However, schools and teachers are, often, perceived by society to be self-protective and conservative. Therefore, the internal or self-evaluations by schools and

teachers are under considerable influence of governments. MacBeath (this section) observes that the school-based internal evaluation in Britain is changing. It is seen as being very much influenced by top-down formulas from the government. Significant ways to make school-based evaluation or teacher self-evaluation, genuinely, professionally regulated and accountable by schools, principals, teachers, as well as governments, will be always a practical challenge for continual reflection and improvement.

The likelihood is high that the direction of evaluation for outcome-based accountability will not be reversed. The tensions inherent in the three sets of relationship, probably, are always present and no definitive course exists to achieve a resolution. Continuing dialog between the stakeholders is imperative in order to reach a consensus and provide new understanding to make the outcome-based accountability systems productive.

Stakeholder Involvement

Education has always been political because it is never without social and economic consequences for individuals (Bruner, 1996; Cronbach, 1982). Education is, therefore, too important for social constituencies to leave it to the educational profession. The same can be said of educational evaluation. Contemporary evidence indicates that nations have established outcome-based accountability systems. Consequently, the scope and depth of stakeholder involvement is ever increasing as there are greater demands being made for education and educational evaluation to be more responsive, transparent, and valid. Stakeholders (e.g., students, parents, principals, teachers, business community, and government agency) who have a vested interest in what is being evaluated are, at present, not merely a source of data but equally important as contributors, shaping the evaluation efforts, including formulating evaluation questions, identifying data sources, interpreting data, and assigning values to evaluation results. This has been documented in *The Program Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 1994). Many entries reflect this trend and discuss critical issues associated with it in educational evaluations, including 'Program evaluation' (Christie and Fierro), 'Curriculum evaluation' (Levin), 'Internal and external evaluation' (Nevo), 'Evaluation, governance and planning' (Picciotto), 'Evaluation reporting and communicating' (Torres), 'Evaluation and accountability' (Sato and Rabinowitz), and 'Moral and ethic issues in evaluation' (Morris). In particular, the article 'The roles of stakeholders in educational evaluation' (Taut and Alkin) provides a focused discussion on the constructs underpinning the notion of stakeholder involvement and factors affecting the stakeholder involvement and its results.

The evaluation literature indicates two theoretical orientations that underscore the notion of stakeholder

participation. One is the line of inquiry into evaluation utilization. It argues that stakeholder involvement increases and improves evaluation use. The other is from the democratic participation perspective – making evaluation more inclusive and interactive. It has been shown that social justice is not only concerned with distributing social benefits but also with distributing power and participation from different groups of people – especially by minority groups whose social participation is often marginalized. These two orientations are considered to result in deep involvement in contrast to broad involvement of stakeholders in its implementation (Taut and Alkin, this section). The utilization orientation focuses on securing the participation of a few key stakeholders – intended users of evaluation to enhance the extent to use evaluation results (e.g., Sato and Rabinowitz, this section). The emphasis of the democratic orientation is on securing the participation of a diverse range of stakeholders – especially those with least power in the context to achieve broader participation and social justice. Whether to implement a deep involvement or a broad involvement of stakeholders for an implementation, certainly, depends on difference in theoretical orientation. However, there is the factor of resource constraint to consider, such as time and money. Taut and Alkin's review of the recent literature concludes that the implementation of evaluations prioritizing deep over broad involvement seems less likely to fail. Furthermore, actively involving stakeholders to attain evaluation use seems to have been better documented in the literature. The conclusion indicates that a broad involvement implementation is under more constraints – such as time, money, and with more diverse stakeholders (also see, Levin, this section).

Stakeholder involvement is now a key principle in the implementation of educational evaluation. It is believed to make the evaluation in question more credible and more accessible to those who are involved in – and affected by – the educational evaluation. Evaluation and evaluative judgment that is anchored in a shared reality by participating stakeholders is expected to make evaluation, itself, a powerful learning system – allowing for organizational learning and improvement. Applying Bruner's analysis of education as aculturing, the learning can take place at two levels. At the macro level, participants will learn to see the culture (or a school) as a system of values, rights, exchanges, obligations, opportunities, and power. On the micro level, participants will learn to examine how the demands of a cultural system (or a school system) affect those who must operate within it. In that latter spirit, participants will learn how "individual human beings construct realities and meanings that adapt them to the system, at what personal cost and with what expected outcomes" (Bruner, 1996: 12).

As anticipated, stakeholder involvement brings challenges for educational evaluation. Many articles, such as 'The roles of stakeholders in educational evaluation' (Taut and Alkin), 'Evaluation and accountability'

(Sato and Rabinowitz), 'Program evaluation' (Christie and Fierro), 'Curriculum evaluation' (Levin), 'Informal education and evaluation' (Jeffs, this section) elaborate on the complications in various educational evaluations. For example, there is considerable stakeholder disagreement and apathy – given the fact that stakeholders differ in authority and power. The standard 'Political viability', therefore, warns against possible pressures from stakeholders and asks evaluators to prevent their involvement from biasing the evaluation and its findings. In addition, factors – such as inadequate individual expertise and experience, imbalanced representation of knowledge and perspectives, inadequate training and orientation, and lack of sufficient time – all pose threats to the validity and utility of evaluation and accountability efforts.

Stakeholder involvement is a necessity – as well as a challenge – for the contemporary educational evaluation. The ultimate goal is to improve evaluation use and to prompt democratic participation and organizational learning. There are constraints to the realization of the full potential with influence from many factors, such as those mentioned above. Affirming stakeholder involvement, dialog along with empirical and practical work are required to identify and create the conditions to maximize the positive effects and minimizing negative ones on educational evaluation.

Globalization in Educational Evaluation

The third feature of the contemporary educational evaluation appears to be a global trend toward outcome-driven or performance-driven educational evaluation. As a result international evaluations are obtaining prominence by providing an internationally comparative perspective to view school effectiveness across nations. Two of the more notable systems are the Organization of Economic Cooperation and Development (OECD)'s Programme for International Student Assessment (PISA) and the International Association for the Evaluation of Educational Achievement (IEA)'s international evaluations. This started, in the late 1950s, when the IEA was established. The scale and influence of the international evaluation has accelerated in the recent two decades. When the IEA conducted the first international mathematics study in 1964, 12 countries participated. Twenty countries took part in the second international mathematics study in 1980–81. The Third International Mathematics and Science Study (TIMSS) was completed in 1995, involving 20 countries. The subsequent TIMSS (now known as Trends in Mathematics and Science Study) took place in 1999, 2003, and 2007. More than 60 educational systems plan to participate in the fifth TIMSS evaluation to take place in 2011. In addition to mathematics, IEA conducts the cycle evaluation of student achievements in reading literacy (Progress in Reading Literacy Study (PIRLS)) and civic and citizenship education (International Civic and Citizenship Education Study

(ICCS)). The target student populations of the IEA evaluation are fourth-grade and/or eighth-grade students. PISA surveys every 3 years on the academic achievements of 15-year olds who complete the compulsory education in reading, mathematics, and science. PISA was conducted in 43 countries in the first assessment, in 2000; in 41 countries in the second assessment, in 2003; and in 57 countries in third assessment, in 2006. Sixty-two countries have signed up to participate in the fourth assessment, in 2009.

The original purpose of international evaluations by IEA was research motivated. More recently, however, the increasing prominence of PISA and TIMSS illustrates that the internationally comparative student-learning outcomes have become the benchmark for nations' public policy in education. One of the forces that determine this is the institutional and financial support of the international evaluations by the supranational bodies such as the OECD, United Nations Educational, Scientific, and Cultural Organization (UNESCO), the European Commission, and the World Bank. Another significant factor is the need for self-knowledge of student achievement by governments because of the accountability motive through which nations are compared – based on the performance of their students. These international evaluations help shape the current performance-driven educational systems around the world. The article 'International evaluations' by Andreas Schleicher (this section) provides an assessment of the developments and the impacts of the international evaluation.

The international evaluations contribute to the field of educational evaluation in several significant ways. The international evaluations inform nations, set national educational targets, and contribute to greater understanding of the relationship among input, processes, and educational outcomes. For example, the Netherlands is using the PISA performance levels to benchmark its school performance against the other OECD countries (Thijs *et al.*, 2009). The PISA assessment results allowed Germany to see the problem of educational equity in comparison to the OECD countries achieving better educational equity and influencing national policies to address the problem (Schleicher, this section). The TIMSS research indicates that the students obtained higher scores and displayed smaller achievement gap when taught by teachers who majored in mathematics or mathematics education compared to students whose teachers did not major in these fields. The findings influenced the Honk Kong Special Administrative Region (HKSAR) government to adopt the policy to require that primary school mathematics teachers have qualifications of either mathematics major or mathematics education major (the requirement exists already for secondary school teachers). In examining the history and development of the National Assessment of Educational Progress (NAEP) in the USA, Schneider (this section) observes that state governors of that country now want to compare their students to those in Korea or

Germany rather than Kansas or Georgia. PISA has certainly gained international attention. Indeed, many countries have considered administering PISA to a state-wide representative sample to benchmark their student performance against students from countries around the world. This also occurs in Shanghai – the only city in the Mainland China to take part in the PISA assessments. The impact of international evaluations on nation's public policy in education is unprecedented.

These international evaluations support the data-based decision making by building an infrastructure of data development, sharing, and access for educational research and development. The infrastructure makes it possible for investigators to work with the same set of variables, theoretical constructs, and to understand how variables and relationships between variables change over time (Committee on Scientific Principles for Education Research, 2002). The international evaluations also advance the techniques to implement cross-country/cross-culture international assessment programs. Andreas Schleicher (this section) provides an informative explanation of the challenges and the solutions in designing and implementing the international evaluations. Leonidas Kyriakides and his colleagues (this section) describe and explain the purpose of as well as the procedure for the development of the international system for teacher observation and feedback for teacher evaluation.

The strong presence of international evaluations is expected to continue given their authority in providing comparative evaluation of student-learning outcomes and, thus, inferring effectiveness of national educational systems. Moreover, the time may come for local evaluation systems to request the use of international data to evaluate program effectiveness or school effectiveness. For example, Hong Kong's PISA provides student-performance information for individual participating schools. Individual schools can use the information to infer how well their students perform compared to those of the other participating schools. The self-knowledge could be a useful tool for the school, students, and staff. However, any international evaluation is perspectival. Caution needs to be taken to prevent the international evaluations from becoming a dominating force – dictating local educational evaluations.

Diversity

The significance of diversity in modern society seems too obvious to be reiterated here. However, the introduction article would not be complete without reiterating the influence of diversity as a feature of this section. Diversity is a fact of life and, certainly, a fact of education. For example, stakeholder participation and the globalization of educational evaluation increase diversity in value, method, and practice for educational evaluation. Education

faces a range of diversity such as ethnic diversity, cultural diversity, language diversity, family diversity, gender diversity, ability diversity, and more. Diversity encompasses differences, conflicts, and vitality, and, consequently, there is a need for choice, equity, and inclusion in order to address these issues. All these facts, issues, and proposals are legitimate objects for educational evaluation. Diversity is also a DNA that is part of the birth and growth of the field of educational evaluation. From the very beginning, the theories and methods were heavily influenced by those of social sciences. Baker and Niemi (1996) attributed to four of such sources that have contributed to the field of educational evaluation. These include experimentation and measurement from psychology, system analysis from research on schooling and administration, and interpretive approaches from humanities critical of empiricism. The different orientations determine the types of evaluation questions asked, the evaluation goals to be realized, the type of interpretations considered valid or useful; the value system chosen and meaning ascribed to quality; the context of the evaluation process; and the evaluator identity and roles (Dahler-Larsen; Levin, this section).

The diversity in epistemological paradigms, methodological tools, and professional training continues to permeate today's evaluation field. The diversity brings viability as well as tension in the field and is clearly reflected in this section. The article 'Understanding approaches to evaluation' (Love, this section) takes on the daunting task in an attempt to delineate numerous evaluation approaches and their classifications from the instrumental chronological framework to many more value-rich frameworks. Each of the approaches has different assumptions, a different form, and different factors affecting it. In a similar vein, the article 'Evaluation methodology' (Chatterji, this section) – in describing various implementations of quantitative, qualitative, and mixed-method designs – addresses the complexity of interdependence of methods, uses, and values in making methodological decisions for doing educational evaluations. The article 'Curriculum evaluation' portrays a paradigm shift from instrumental evaluation to adaptive and situated evaluation and to emergent and emancipatory evaluation. In the article 'Cultural issues that can affect the validity of educational evaluations', Durán (this section) argues for the importance of understanding and taking into consideration the cultural ways of persons targeted by educational interventions in making educational evaluation more valid. In the *Zeitgeist*, the article 'Defining quality in evaluation' (Dahler-Larsen, this section) recounts the complication and controversy in defining quality in evaluation.

These articles reflect ways that the field of educational evaluation has developed more diverse perspectives and more varied methodological repertoire in response to the contemporary educational realities that are multiple, divergent, and interrelated. Diversity is an asset.

It reduces hierarchical and structural control from both inside and outside the field – providing the field with intellectual viability. Diversity also causes anxiety and incoherence (Mathison, this section). Incoherence is not necessarily a problem; rather, it can be the drive for pursuing new dialog, consensus, and understanding. However, this is only possible when diversity itself is prevented from becoming dogma and causing fissures between different thoughts and practices – that is, an evaluation approach, a practice, or even diversity itself can always be questioned and examined, similar to any other entity.

Theory-Driven Evaluation

In general, educational evaluation includes outcome evaluation and process evaluation. However, educational evaluation has focused on outcome evaluation in order to establish the merit of an educational intervention. It concerns outcomes – not the program itself – in program evaluation. This situation is changing. The author has learned this through personal experience when she submitted a manuscript last year to *American Education Research Journal* (AERJ). The paper reports a longitudinal study to compare learning outcomes of students receiving alternative mathematics curricular. The editor returned the submission to the author, explaining that the AERJ does not accept a manuscript on program evaluation if it does not provide evaluation details of the program implementation linked to the program outcomes. This was surely not an isolated incident. Rather, it indicated the field of evaluation research appeared to have developed – moving beyond a black box, input–output models focusing on outcomes to the approaches that emphasize program theory of why a program would produce expected changes and program implementation of how the program would achieve changes, planned or not – thus called theory-driven evaluation (Weiss, 1998).

In this section ‘Educational Evaluation,’ the move for theory-driven evaluation is evident in two aspects – one involves conceptual inquiry of critical concepts/issues, the other prompts systematic approaches to evaluate program theories, that is, the way in which produced educational outcomes are related to educational processes. The changing conceptualization of the construct evaluation use is an exemplar of the conceptual investigation. Thomas (Thomas, V., this section) traces this development of how the present multifaceted concept of evaluation use (e.g., process use, conceptual use) has been informed by practical needs and also informed practice. The enriched conceptualization of evaluation use not only helped clarify the theoretical question of what constitutes evaluation use but also stimulates new questions and new practices (Weiss, 1998). For example, the concept of process use inspired the innovative evaluation practices that has prompted organizational learning and capacity building – such as school-based evaluation (MacBeath, this section), self-assessment for

teacher professional development (Tillema, this section), self-assessment for adult education (Pauline, this section), and evaluating schools as learning communities (Thomas, S., this section).

The concept of process use refers to the evaluation process itself being utilized as a learning process to improve and to empower for participating individuals and organizations. Process evaluation is concerned with understanding reasons why an educational intervention works to bring about outcomes – expected or not. An emphasis on process evaluation that asks the questions of what happens, and how and why it happens, highlights the function of educational evaluation for knowledge development similar to any educational research. The article ‘Program evaluation’ (Christie and Fierro, this section) explains the importance of process evaluation for program understanding and program improvement. This is, clearly, echoed in the notion of multiple representations of curriculum – namely, intended, planned, implemented, and achieved curriculum, and their interrelationship for evaluation (Levin, this section). The thinking of multiple representations of curriculum or program and their evaluation have marked the present practice in educational evaluation. As an example, there is evaluation of mathematics-education programs (Cai, this section), evaluation of science-education programs (Berberoğlu, this section), and evaluation of social studies and civic education programs (Grant, this section).

Process evaluation prompts and is also consistent with a system-analysis approach to evaluate school effectiveness at four different – but interrelated – levels: societal or macro (system, nation, or state), institutional or meso (school), instructional (classroom) or micro (classroom), and individual or nano (Levin, this section). For example, PISA operates within the frameworks that require data from the different levels of the educational system. These include the educational system as a whole – that is, the educational institutions and providers of educational services; the classrooms or instructional setting; and the learners themselves. Each level is being investigated for education and learning outputs and outcomes, policy levers, and contexts shaping educational outcomes and antecedents or constraints that contextualize policy (Schleicher, this section). The NAEP High School Transcript Study (HSTS) gathers not only test scores but also information about school practices and how these practices might affect performance outcomes (Schneider, this section). The system-analysis approach integrating outcome evaluation and process evaluation has contributed greatly to interpret observed variation in learning outcomes between students, classrooms, schools, and educational systems. In educational evaluation – to understand what is happening at one level – it is, often, necessary to understand other levels. For example, student-learning outcomes in a given subject matter may be influenced by the teacher’s approach to instruction, the value a

principal places on the subject within the curriculum, the curriculum adopted by the school district, and different familial and community factors (Committee on Scientific Principles for Education Research, 2002; Ni, Chiu, and Cheng, in press).

Theory-driven evaluation that emphasizes process evaluation also makes it possible and more meaningful for meta-evaluation (Wingate, this section) and meta-analysis of educational evaluations (Weiss, 1998). Theory-driven evaluation is imperative for the field to contribute to understanding of the multiplicity of intervening factors operating in the educational systems that effect educational outcomes.

Sustaining Challenges for Future Directions

The five perspectives elaborated above are intended to provide a rough framework showing the new developments in educational evaluation in the last 10–15 years. These new developments – collectively reflected by their contributions – are taking place to meet the need for educational innovations that address social, cultural, and economical changes. Consequently, each of the new developments – such as outcome-based accountability, stakeholder participations, diverse approaches and practices, globalization of educational evaluation, and theory-driven evaluation – is raising a set of new questions and challenges. Some of the substantive issues associated with each of the developments have been identified in the previous parts of the article. In the present segment, the focus is on two basic issues that have serious ramifications for the field of educational evaluation. The first is the problem of generalization which is concerned with the extent to which a conclusion – drawn from specific evaluations – can be generalized to a different situation. The other issue relates to the specific functions of educational evaluation. It has been assumed that description, judgment, and recommendations provide legitimacy to educational evaluation. However, it can be argued that a concise account of relevant facts is the most important function of educational evaluation.

The problem of generalization is particularly acute for educational evaluation. The reason for this is that evaluation is deeply embedded in social and political activity (Cronbach, 1982). Evaluation and research are considered different in terms of the origin of study questions and the purpose for which study information is gathered. Evaluation questions are, often, elicited from program stakeholders whereas research questions are from researcher(s). Evaluation studies are, often, decision oriented, designed to generate information for program improvement and for informing the decisions based on the program under evaluation. Research is, often, not decision oriented but aims to understand the phenomena under investigation and contributes to knowledge development in an area

(see Christie and Fierro, this section). Therefore, some researchers argue that evaluation should be focused on particularization rather than generalization. Differentiating evaluation from research encourages evaluators to be more sensitive to local contexts that may be a constraint to an educational program under evaluation. It also encourages more orientation toward producing evaluation findings that are beneficial for the users. Acknowledging the distinction, however, evaluation or educational evaluation should not be exempt from the requirement of generalization. Conversely, it makes generalization more challenging and having a more critical role in educational evaluation. Overemphasizing the distinction may reduce evaluation or educational evaluation to merely a commodity to be contracted or purchased. Evaluation or educational evaluation has to be concerned with regard to generalization. Without this assurance of integrity, evaluation or educational evaluation will have no legitimacy. It is of the utmost important for knowledge development to inform policy and practice in a substantial and sustaining way. Raz (1999) uses an example of understanding Japanese humor to illustrate the relationship between social dependence of value and universality of value (see quote below):

How can the points about social dependence or creation of value be reconciled with the requirement of intelligibility which leads to the irresistible attraction of universality of value? Granted that we cannot have a thick understanding of the concept “Japanese humor” without familiarity with their society and with instances of their humor, still we cannot recognize their humor for what it is unless we can subsume it under a more general concept, and recognize theirs as an instance of humor. (p. 208)

The example provides a good illustration of the relationship between a general principle and its particulars. On one hand, a particular humor takes its meaning from the situations to which it refers; on the other hand, the particular humor is only illuminated with the general principles governing what makes a humor. The same point can be made with regard to the relationship between particularization and generalization in educational evaluation.

The need for improving generalization of evaluation and evaluation studies is being addressed within the field. For example, many contributors have raised this concern. Robinson and Campbell (this section) suggest that the present teacher-evaluation practice is derived from specific programs that have been *ad hoc*, lacking a robust evidence-based justification, and a robust knowledge basis. Glassman and Glasman (this section) observe that the practices of principal evaluation have been affected by stakeholder involvement. Unfortunately, there is no systematic research on how stakeholders have shaped the nature of principal evaluation and its results. The matter of whether and by what means school inspections contribute to school

improvement has been studied systematically in very few countries (Chapman and Earley, this section). Stufflebeam and Shinkfield (2007) have expressed dissatisfaction that the evaluation theories that prevail at present are mostly prescriptions for practice and have not been systematically examined in empirical studies. King (1998) made the same assessment, stating that the field of evaluation has been based more on theorists' ideas than on systematic research evidence.

The present literature has indicated that there are two groups of evaluation studies, among others, that provide exemplars for improving generalization of educational evaluations – making them more accountable for informing educational policy and practice. The first of these consists of a constellation of evaluation programs that assess effectiveness of similar and ongoing educational programs. There are concerted efforts that allow for the examination of robustness of similar education programs in different circumstances and with different methods. A good example of this is the evaluation of the federally funded early-childhood programs in the United States (Heckman *et al.*, 2009; Reynolds, 2000). On one hand, the implementations of the early-childhood education varied in different states and communities. This made evaluation-based generalizations with regard to its effectiveness difficult to obtain. On the other hand, the assemblage of evaluations on the programs that were carried out in diverse situations provided an excellent opportunity to examine whether or not a given finding about the effects of the programs could be observed over diverse circumstances. Converging evidence was obtained that indicated cognitive advantages for the children participating in the programs tend to disappear approximately 3 years after leaving the programs. However, those children who participated did benefit in terms of increased likelihood of retention in grade school, high school graduation, college education, and employment. The conclusions converging from numerous evaluations of the early-childhood programs have contributed to well-informed educational policy and practice for early-childhood education.

As the institutionalization of data systems is a reality (Baker and Niemi, 1996) at the multiple levels – such as state-assessment systems, national assessment systems (e.g., NEAP), and international assessment systems (e.g., PISA) – educational evaluations can be conducted in a less fragmented but more concerting way to achieve better generalization of evaluation studies. With the data systems as well as the advances in statistical tools (Chatterji; Ma, this section), multiple levels or systematic analysis using classroom and school as the relevant unit are starting to be prevail (Akiba *et al.*, 2007; Tarr *et al.*, 2008). With the powerful data systems and the advanced analysis methods there, whether one has some program theory becomes ever significant to identify new questions and to find

regular and meaningful patterns in social relations (e.g., school, classroom, teacher, family, and community) that produce educational outcomes.

The second group consists of experimental studies into effectiveness of educational programs. Generally, there is understanding that experimental studies present great challenges for conducting educational evaluations (Scriven, 2005). For example, the article 'Methods to evaluate technology' (Bewley, this section) discusses the difficulties with experimentally evaluating effects of technology use on student-learning processes and outcomes. The difficulties are apparent and so are the demands. Understandably, there is pressure for educational evaluations to provide an evidentiary base to inform educational programs or policies for their development, implementation, and refinement. Educational evaluations can be a *post hoc* process – understanding and evaluating an educational intervention afterward. It can also be a planned experiment – testing anticipated outcomes from educational activities. The fact that cooperative learning becomes a standard and widespread teaching procedure in classroom should be attributed to hundreds of experimental studies that verified and established the effectiveness of the educational procedure (Johnson and Johnson, 2009).

The issue of generalization is related to the second issue on the functions of educational evaluation. As the early-childhood education and the cooperative learning examples indicated, external validity is difficult to demonstrate with a single evaluation and has to be accumulated over many replications. Therefore, educational evaluation, in general, can serve the functions of describing what occurs as well as how and why it happens. Other important aspects are judging whether or not what happens meets the expectations and recommendations as to what action should be taken. However – for a single evaluation – it is neither productive nor feasible to include all three. In reality, the most important contribution of an educational evaluation is to provide a description that is as accurate as possible. These descriptions will have to be verified and collaborated with other similar and relevant replications. Then, it will be possible for judgment and for recommendations that are constructive.

Recently, the author and her colleagues were conducting an evaluation on the effects of a reformed mathematics curriculum on teaching and learning in Mainland China (Ni *et al.*, 2009). The results showed that – after accounting for the socioeconomic status (SES) variable – the reform group performed better than the nonreform group on open-ended tasks requiring mathematical explanation. However, the reform group did not perform as well as the nonreform group on measures of mathematical calculations and routine mathematics problem solving. Another finding showed that there was a decrease in the achievement gap in calculations and routine mathematics problem solving for both groups. However, the decrease

was more significant for the nonreform classrooms. The trend was not obvious for both groups on the complex problem-solving tasks. It was concluded that the differentiated patterns of mathematics achievement between the two groups appeared to reflect the relative emphases of the respective curricular. In this sense, the new curriculum was effective in achieving the intended outcome to improve students' competence to solve open-ended tasks. The researchers were then asked to judge whether or not the observed tradeoff was worthwhile, comparing the proficiency and the automaticity in computation skills and the competence in solving nonroutine mathematical problems. However, it was not possible to make a judgment about the matter according to these evaluation results. To understand the tradeoff in theory and to make the judgment about the merits of the tradeoff, among others, three kinds of empirical evidence need to be obtained. The first is concerned with relationship between the different dimensions of mathematics competence. The second requires longitudinal tracking of the development of the different mathematics competences in students receiving the new curriculum. The third kind of evidence needs to look into comparative results involving Chinese students using international assessments. The example suggests that it is feasible for an educational evaluation to make a specific judgment, such as whether or not a specific educational program produces expected results. However, it will be challenging for an evaluation to make a broader assessment – such as whether or not it is worthwhile for a program to gain in some areas at the expense of those in other areas. That would extend beyond the capacity of an educational evaluation to make recommendations for action. Scriven (1995) made an excellent analysis of this dilemma. According to these findings, educational evaluation needs to focus on expertise in the field to describe any given situation and factors that influence the outcome. Based on the careful descriptions of many evaluations, collective efforts can be exercised to compare different evaluation studies and dialogue to identify regular and meaningful patterns of relationships for the service of informing educational policy and practice in a substantial and sustaining way.

In sum, educational evaluation must be an integral part of civil society. The establishment of the importance of evaluation has seen consistent growth and is one that depends on dialog with practitioners in the field as well as between the field and other disciplines and the public. It is the purpose of this section 'Educational Evaluation' to facilitate wider access to the new developments, and also to provide greater understanding for educational professionals and the public – individuals who have a vested interest in education and educational evaluation and thus to help make the field more self-enhancing and socially empowering.

See also: Cost Analysis in Evaluation Studies; Cultural Issues that Can Affect the Validity of Educational Evaluations; Curriculum Evaluation: Approaches and Methodologies; Curriculum Evaluation; Defining Quality in Evaluation; Evaluating Education in Three Policy Eras; Evaluating E-Learning; Evaluating Schools as Learning Communities; Evaluation and Accountability; Evaluation Governance and Planning; Evaluation Methodology; Evaluation of Adult Education and Training Programs; Evaluation of Integrated Health Programs in School; Evaluation of Mathematics Education Programs; Evaluation of Science Education Programs; Evaluation of Social Studies and Civic Education Programs; Evaluation of Teacher Quality and Practice; Evaluation Reporting and Communicating; Evaluation Use; Formative Assessment in Teacher Education and Teacher Professional Development; Impact of Assessment on Students' Learning Strategies and Implications for Judging Assessment Quality; Informal Education and Evaluation; Internal and External Evaluation; International Evaluations; Longitudinal Evaluation Designs; Meta-evaluation: Purpose, Prescription, and Practice; Methods to Evaluate Technology; Moral and Ethical Issues in Evaluation; National Assessment Programs: The Example of the U.S. National Assessment of Educational Progress; Needs Assessment in Education; Principal Evaluation: Concepts, Needs and Realities; Program Evaluation; School Based Evaluation: Purposes, Protocols and Processes; School Inspection/External School Evaluation; The International system for Teacher Observation and Feedback: A Theoretical Framework for Developing International Instruments; The Purpose of Educational Evaluation; The Role of Stakeholders in Educational Evaluation; Understanding Approaches to Evaluation.

Bibliography

- Akiba, M., LeTendre, G. K., and Scribner, J. P. (2007). Teacher quality, opportunity gap, and national achievement in 46 countries. *Educational Researcher* 36(7), 369–387.
- Baker, E. and Niemi, M. D. (1996). School and program evaluation. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 926–944. New York: Macmillan.
- Bruner, J. (1996). *The Culture of Education*. Cambridge, MA: Harvard University Press.
- Committee on Scientific Principles for Education Research (2002). *Scientific Research in Education*. Washington DC: National Academy Press.
- Cronbach, L. L. (1982). *Designing Evaluations of Educational and Social Programs*. San Francisco, CA: Jossey-Bass.
- Heckman, J., Doyle, O., Harmon, C., and Tremblay, R. (2009). Investing in early human development: Timing and economic efficiency. *Economics and Human Biology* 7, 1–6.
- Johnson, D. W. and Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher* 38, 365–379.
- Joint Committee on Standards for Education Evaluation (1994). *The Program Evaluation Standards*, 2nd edn. Thousand Oaks, CA: Sage.
- King, J. A. (1998). Making sense of participatory evaluation practice. *New Directions for Evaluation* 80, 5–23.
- Ni, Y. J., Chiu, M. M., and Cheng, Z. J. (in press). Chinese children learning mathematics: from home to school. In M. H. Bond (ed.)

- The Oxford Handbook of Chinese Psychology*. UK: Oxford University Press.
- Ni, Y. J., Li, L., Cai, J., and Hau, K. T. (2009). *Has Curriculum Reform Made a Difference? Looking for Evidence of Change in Classroom Practice*. Hong Kong: The Chinese University of Hong Kong.
- Raz, J. (1999). *Engaging Reason*. New York: Oxford University Press.
- Reynolds, A. J. (2000). *Success in Early Intervention: The Chicago Child-Parent Centers*. Lincoln: University of Nebraska Press.
- Reynolds, A. J., Wang, M. C., and Walberg, H. J. (eds.) (2003). *Early Childhood Programs for A New Century*. Washington DC: Child Welfare League of America.
- Schwab, J. J. (1978). Education and the structure of the disciplines. In Westbury, I. and Wilkof, N. J. (eds.) *Science, Curriculum, and Liberal Education. Selected Essays Joseph J. Schwab*, pp 229–274. Chicago: University of Chicago Press.
- Scriven, M. (1995). The logic of evaluation and evaluation practice. *New Directions for Evaluation* **68**, 49–70. Thousand Oaks, CA: Sage.
- Scriven, M. (2005). Causation. In Mathison, S. (ed.) *Encyclopedia of Evaluation*, pp 43–47. Thousand Oaks, CA: Sage.
- Stufflebeam, D. L. and Shinkfield, A. J. (2007). *Evaluation Theory, Models, & Applications*. San Francisco, CA: Jossey-Bass.
- Tarr, J. E., Reys, R. E., Reys, J. B., et al. (2008). The impact of middle-grades mathematics curricular and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education* **39**, 247–280.
- Thijs, A., van Leeuwen, B., and Zandbergen, M. (2009). *Inclusive education in the Netherlands*. The Hague: National Institute for Curriculum Development.
- Weiss, C. H. (1998). *Evaluation: Methods for Studying Programs and Policies*, 2nd edn. Upper Saddle, NJ: Prentice-Hall.

EDUCATIONAL EVALUATION – CONTEMPORARY THEMES IN EVALUATION

Contents

Evaluating E-Learning

Evaluating Schools as Learning Communities

Evaluation of Adult Education and Training Programs

Evaluation of Integrated Health Programs in School

Formative Assessment in Teacher Education and Teacher Professional Development

Informal Education and Evaluation

Methods to Evaluate Technology

Evaluating E-Learning

D D Williams and C R Graham, Brigham Young University, Provo, UT, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Blog – A website on which the author keeps a record or log of whatever he/she wants.

Course-management system – An Internet-based system for organizing and operating a learning and teaching course.

Blended courses – The combinations of face-to-face and online or distributed instruction.

E-learning – Learning utilizing any form of electronic technology or media.

Evaluand – An object of evaluation.

Evaluation – Assessing the merit or worth of an evaluand.

Metaevaluation – Evaluation of evaluations.

Personal-learning environments – The combinations of digital learning resources learners assemble to bring information to them and through which they can communicate with others to improve their own learning.

Stakeholder – A person with interests in the evaluation of an object.

Web analytics – An analysis of how viewers are using a website.

Wiki – A website that allows multiple authors to jointly create the contents.

teaching people of all ages (Wesch, 2007). Multiple participants' innovations with technology-mediated learning require dynamic evaluative inquiry (Allen and Seaman, 2007; WestEd, 2008). Evaluators identify evolving evaluation objects (evaluands) in idiosyncratic combinations, involve stakeholders in prioritizing their values to select from thousands of possible evaluation questions, gather, analyze, and report answers to those questions, and make recommendations for action, while simultaneously meta-evaluating and improving their evaluation activities.

This article employs a flexible evaluation framework (Williams, 2006), exemplified in most articles of this volume and in evaluation literature, to illustrate a way to conduct e-learning evaluations (see **Figure 1**). Consideration of challenges and opportunities facing e-learning participants and evaluators, as educational uses of technologies proliferate, complete the article.

As **Figure 1** illustrates, evaluation involves several activities for reaching judgments by clarifying what should be. This usually entails understanding the evaluation context by consulting with stakeholders to identify evaluands to evaluate, clarify criteria that stakeholders value, and compose questions about the evaluands' nature and performance. Then, evaluators collaborate with stakeholders to determine what is by clarifying methods to effectively answer the questions, and collecting and analyzing associated data. Finally, evaluators and stakeholders juxtapose what should be with what is to judge how well what is meets what should be, make recommendations, and meta-evaluate the entire process and results.

Evaluating E-Learning

Hopeful visions for using new technologies are transforming the nature of learning and possibilities for

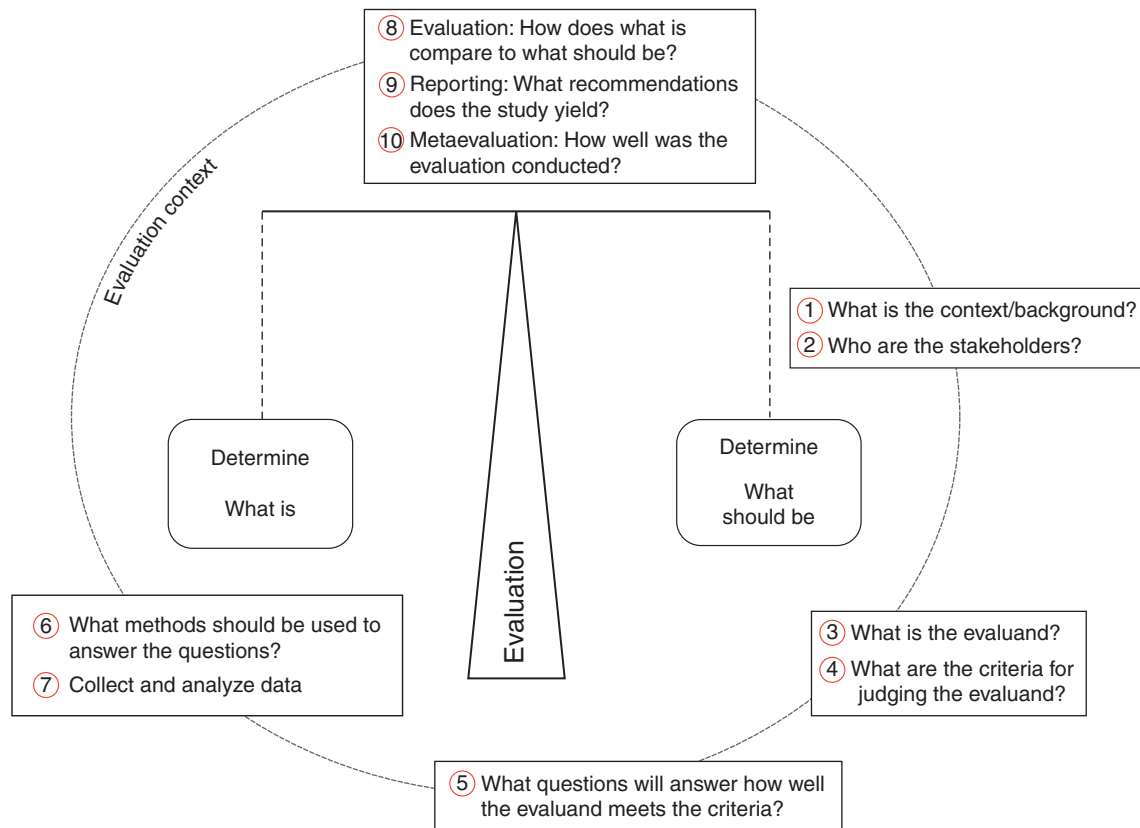


Figure 1 Key components in the evaluation process.

Table 1 Scope, stakeholders, and possible evaluands in an e-learning context

Scope of context	Primary stakeholders	Sample evaluands	Example criteria
Institutional level	Administrators	E-learning initiative Online course offerings E-learning policies	Cost effectiveness Enrolments Completion rates/satisfaction
Program level	Administrators	Distance-learning program	Enrolments Student satisfaction
Course level	Instructor/learners	Online course	Student satisfaction, learning, engagement Student access Resource/tech requirements
Activity level	Instructor/learners	Course activity	Student satisfaction, learning, engagement Participant time involvement

Elements of E-Learning Evaluation

After illustrating how activities summarized in **Figure 1** facilitate evaluation of e-learning experiences and results, an extended case is presented to demonstrate processes for clarifying what should be, describing what is, and comparing them.

Clarifying What Should Be

Clarifying what should be involves identifying context, stakeholders, evaluands, and criteria to generate evaluation questions.

Context

Beginning an e-learning evaluation involves understanding the context and setting the stage for deciding who the stakeholders will be, what evaluands will be studied, and what the criteria for judging the evaluands will be. For example, **Table 1** shows four scope levels that are often considered in e-learning contexts (Graham, 2006). At the institutional and program levels, administrators are often the primary stakeholders, while at the course and activity levels, faculty and learners are often the primary stakeholders.

Context also includes other variables summarized in **Table 2**: The setting for e-learning, the evaluation purpose,

Table 2 Context variables and examples associated with e-learning

Context variables	Examples
Setting	<ul style="list-style-type: none"> • Distance • Face to face • Blend of both
Purpose	<ul style="list-style-type: none"> • Formative (for improvement) • Summative (for judgment)
Who is evaluating?	<ul style="list-style-type: none"> • Learners (internal) • Teachers (internal) • Instructional designers (internal and external) • Administrators (internal and external) • Professional evaluators (internal and external)
Who is responsible for learning?	<ul style="list-style-type: none"> • Learners through personal-learning environments • Learners through institutional learning-management systems under instructor direction
Formality of evaluation	<ul style="list-style-type: none"> • Informal • Formal
Evaluand development stages during which evaluation is being conducted	<ul style="list-style-type: none"> • Before development and implementation • During implementation • After implementation
Evaluability of the evaluand	<ul style="list-style-type: none"> • Stability of the evaluand • Availability of evaluation resources • Availability of guidance from literature • Interest and attention of stakeholders

who is evaluating, who is learning, evaluation formality, the evaluand's stage of development, and evaluability of the evaluand. Diversity of contexts complicates the evaluation of e-learning; however, clarifying contexts involved as part of the evaluation process helps focus selection of stakeholders, possible evaluands or their characteristics, and criteria for judging the evaluands, and simplifies data gathering and interpretation.

The diversity of possible e-learning evaluation contexts is apparent in a review of article titles in the *International Journal on E-Learning*. The October 2008 issue includes articles on online communities, blended learning, students' perceptions of value of online learning, self-regulation by learners, designing and implementing courseware and systems, and many others.

Stakeholders

To clarify what should be, the contextual issues and others discussed below are clarified, usually through

negotiation among stakeholders (those internal to the learning process such as students, teachers, and instructional designers, as well as those who have a more external role, such as some administrators and independent evaluators), as led by both internal and external evaluators. Who are the stakeholders? Who should have a say regarding the quality of e-learning associated with various combinations of technologies, pedagogies, learners, and teachers?

Logically, the most important stakeholders for e-learning are e-learners, ranging from preschool children to university students, to people of all ages participating in social networks and/or learning on their own, to adults participating in corporate-training events. Learners are the ultimate evaluators of the knowledge and learning environments they receive; so e-learning evaluations should take their views, needs, and values into account along with the views of teachers, parents, administrators, educational company owners, government leaders, instructional designers, and others who care about the learners.

In recent years, evaluation theorists have promoted various participatory evaluation approaches (Cousins and Whitmore, 1998) to identify and include values of all people with interests in evaluands, particularly recipients of services, such as learners. Learner participation is especially important in e-learning evaluation because increasingly, there is no designated teacher in many e-learning situations. Learners' values and needs are becoming more essential in evaluations of their learning experiences and outcomes as growing uses of personal-learning environments and Web 2.0 tools suggest should happen. An Internet search on personal-learning environments provides videos and documents exploring how learners are using emerging tools to personalize e-learning.

A growing literature explores the nature of learners as key stakeholders in evaluating e-learning. For example, Clifford (2006) and Ito *et al.* (2008) conducted studies of learners' uses of social media and other technology tools to identify values, habits, learning styles, and characteristics of e-learning stakeholders. Evaluations of e-learning expand this growing literature and invite stakeholders to clarify their idiosyncratic values and evaluands.

Evaluands

What evaluands do stakeholders care most about and how do they define them and the issues associated with them? This is one of the more complicated questions to be answered in clarifying what should be in e-learning. Why? This is because stakeholders often struggle to agree about what the evaluand is in e-learning. **Table 3** lists a range of categories of evaluands often considered in e-learning contexts. The growing literature on the interface between technology, pedagogy, and content knowledge (TPCK) suggests most evaluands are complex combinations

Table 3 Examples of a few evaluands and criteria in diverse e-learning contexts

<i>Categories of evaluands</i>	<i>Example evaluands</i>	<i>Example criteria</i>
Content	Module Materials Course	Quality Accuracy Cost Access
Pedagogy	Pedagogical assumptions Course activities	Student engagement Student satisfaction Student learning
Technology	Course-management systems Production and delivery tools Labs/specialized equipment	Access to technology Student/instructor technology skills Cost of infrastructure Scalability
Instructor/learners	Online teaching/learning knowledge and skills Instructor/students dispositions toward e-learning	Instructor training Instructor/student comfort in online setting
Business model	Independent study model Teacher-led e-learning model Open-learning models Others	Cost effectiveness Enrolments Completion rates/satisfaction Scalability Balance between human and materials interaction

of these and other variables. To evaluate isolated effects of technological or pedagogical interventions does not make sense (Koehler and Mishra, 2008), as the literature on no significant differences due to technology indicates. Technology never operates alone in learners' personal-learning environments or in curriculum-supporting learners (Oblinger and Hawkins, 2006).

Criteria

What are stakeholders' values, criteria, and standards for judging evaluands as they operate in particular contexts? Given the variety of e-learning stakeholders and their definitions of evaluands they want to evaluate, there are many criteria to consider in evaluating e-learning (see **Table 3** for examples). Most criteria may be organized as learning quality (knowledge, skills, and dispositions as measured through direct and indirect performance indicators or learning outcomes), efficiency (lower costs for equal or higher performance), and improved access to learning opportunities (Graham, 2006) – or in terms of inputs, processes, and outcomes for each of five pillars of

e-learning (Lorenzo and Moore, 2002): cost effectiveness and institutional commitment, learner access, learning effectiveness, student satisfaction, and instructor satisfaction.

In addition, industry standards for judging subsets of a learning experience might be considered important by some stakeholders and could be taken into account. For example, there are standards for web-based e-learning called Sharable Content Object Reference Model or SCORM (Advanced Distributive Learning, 2004) and for content areas such as math (National Council of Teachers of Mathematics, 2000) and reading (International Reading Association, 2004).

Questions

Exploring the four components of the evaluation framework yields what should be or criteria and standards that evaluators translate into questions stakeholders want answered so they can evaluate what is against those criteria. For example, in a university economics department, faculty and students as key stakeholders might agree to ask if the computer programs they anticipate buying to train students in economic statistics meet standards such as minimal cost, easy Internet access to minimize lab use, large data-set capacity, reduced need for faculty mentoring, and high student-performance scores when using them, while also meeting SCORM and other technical standards. Once criteria are identified, articulated evaluation questions focus on assessing how well competing computer programs meet these standards.

Clarifying What Is

Once what should be and associated questions are established, e-learning evaluators select methods, and gather and interpret data to answer the question of what is.

Selecting methods to fit questions

Funding pressure has grown to evaluate with randomized control group trial studies. However, professional evaluators (AEA, 2003) and other experts (AERA, 2003) agree that answering questions about what is compared to what should be involves quantitative, qualitative, or mixed data, collected through many designs including experiments, quasi-experiments, surveys, and case studies.

Likewise, a review of methods and results for evaluating e-learning for K–12 children (WestEd, 2008) suggests that the most useful evaluation studies triangulate multiple stages and methods to replicate and confirm findings. In this review of seven online learning-program evaluations, the US Department of Education guide identifies challenges and recommends methods for meeting needs of multiple stakeholders, building on existing literature, recognizing evaluands as multifaceted, doing comparison studies whenever possible, gathering valid data, and

moving stakeholders to action with evaluation results. Many resources for facilitating K–12 and higher education e-learning evaluations are noted.

Data gathering and analysis

Depending on the questions asked, data gathering may involve observing, interviewing, artifact reviewing, testing, surveying, and measuring unobtrusively (e.g., web analytics). Analyses include statistical description and hypotheses testing to make inferences about the power of findings as well as thick qualitative descriptions and identification of patterns and themes to facilitate interpretation of results.

The seven studies mentioned earlier (WestEd, 2008) illustrate these multimethods. The evaluators used comparison studies when possible but remained flexible in order to address emerging interests of multiple stakeholders through multistage studies that addressed formative and summative questions. They used surveys, observations, interviews, tests, and computer-usage data to examine the implementation and impact of distance-education programs, teacher-training programs, self-paced online classes, blended courses, online mentors, digital libraries, online learning activities and diagnostics, video-conferences, podcasts, blogs, games, and other digital tools to enhance e-learning.

Completing the Evaluation

Evaluation: Comparing what is to what should be

The heart of evaluation involves comparing the current activities and/or outcomes associated with an evaluand to the criteria that stakeholders care about and deciding how worthy or meritorious that evaluand is. E-learner could make this comparison regarding their own personal-learning environment and their learning process and outcome goals, asking if they are learning what they want and need from sources they select. Parents could compare what their child is learning while participating in a computer lab to what they want them to learn or what the lab sponsors claim they should be learning. A business could compare how efficiently they are able to train their employees using an Internet-based simulation and how well employees can perform before and after training. In each case, evaluation compares described reality to the stakeholders' ideals and invites them to make formative or summative decisions about the evaluand.

Carliner and Shank (2008) combine many evaluations made by experts in e-learning based on their experiences and evaluations of efforts to use innovative technologies to enhance learning. They conclude that e-learning has been over-sold and efforts to evaluate and enhance e-learning should be more critical and rigorous. Several authors in their edited book note that e-learning has not grown as fast as projected and some stakeholders treat e-learning as another educational fad – demanding the latest new developments without demanding evidence that the results

are effective or available cost efficiently. A few meta-analyses have been done of e-learning (see several in the section titled 'Further reading') but conclusive evidence regarding the effects of technology, pedagogy, teacher content knowledge, and the many other variables discussed in this article continues to elude researchers. Ongoing evaluation of e-learning over time may estimate how well educators using technology to enhance learning are succeeding and how e-learners judge the quality of their experiences in those programs as well as their own personal-learning environments.

Reporting and recommendations

Reports may be automatically generated and shared efficiently by posting user ratings, as exemplified on Amazon.com and other websites. Khribi *et al.* (2008) encourage such web-usage mining to generate recommendations from internal consumers' perspectives. Usually, findings are presented in written and/or oral reports, for review by stakeholders, with encouragement to collaborate with evaluators to generate and implement recommendations.

Recommendations may urge stakeholders to buy an e-learning tool after comparing it to standards, to revise an e-learning curriculum to better meet standards, or to reject an e-learning data-management system from further consideration. Stakeholders who will need to implement them are in the best position to decide how realistic and sensible recommendations are, even though they usually benefit from collaboration with external third-party evaluators who are less invested in the outcomes.

Metaevaluation

Professional evaluators have collaborated with stakeholder groups to create standards for judging evaluations (Sanders, 1994) and principles to guide sound and ethical evaluation practice (AEA, 2004). Stakeholders of e-learning evaluations can use these resources to ensure that the evaluations proposed and implemented are trustworthy, valuable, and useful. However, very few examples of e-learning stakeholders using evaluation standards and principles are found in the literature (Williams and Campbell, 2001).

Extended Case Vignette

To holistically illustrate principles and practices identified above and demonstrate how e-learning can be evaluated, a sample study is analyzed below. In this case, an online e-learning activity was integrated into a face-to-face university class. Such blended learning has been called the single greatest unrecognized trend in higher education (Young, 2002: A33) and was also identified by the American Society for Training and Development as one of the top ten trends in the knowledge-delivery industry (Rooney, 2003). Examples from informal social network-based

learning, K–12 schooling, adult education, and corporate training could also be used to illustrate the use of e-learning evaluation.

Clarifying What Should Be

The framework activities presented in **Figure 1** for clarifying what should be are illustrated in this example.

Context

A university instructor decided to create a class wiki to encourage class members to build a learning community. She invited students to join her in contributing weekly to the wiki by summarizing readings, issues, and accomplishments associated with their own research projects, questions they wanted to answer, and responses to others' entries. Throughout this innovative experimentation, the instructor invited the students to internally evaluate the experience with her.

The scope of the evaluation study was focused on the wiki activity within a blended setting. The instructor wanted to formatively refine use of wikis or some form of virtual collaboration for use throughout the semester and beyond. Many of the students were summatively judging whether they would ever participate in a class using wikis in the future. The instructor and the students together were internally evaluating their own use of the wiki and associated learning. Although this process was initiated by the instructor as a way to enhance the students' learning through an institutional-management system, the students were free to interpret the use of the wiki in various ways and could make it part of their own personal-learning environments to whatever degree they preferred.

This evaluation was formal in that the instructor planned to gather data systematically from the students by observing their use of the wiki and by interviewing them. The evaluation continued through all stages of the wiki's development as the instructor interviewed students about her plan and invited them to join her in evaluating their use of and learning from it during and after implementation of the wiki. The wiki was evaluable because the instructor introduced and monitored it in a stable way, provided sufficient evaluation resources by making evaluation of the wiki a part of the course, and built upon literature she had reviewed regarding the use of class wikis. The instructor and students, as co-stakeholders, were willing to participate in making the experience evaluable.

Stakeholders' evaluands, criteria, and questions

The stakeholders were the instructor and students using a class wiki. The instructor was tempted several times to dominate the study because the wiki was her idea and she would be teaching this class after the students finished as well. However, she wanted to understand the students' experiences and judgments of the wiki activity too. So, she invited them to not only evaluate it but to tell her at

the earliest stages possible what their expectations and worries were concerning use of the wiki. They had the opportunity to collaborate with her in creating the wiki and in evaluating this learning activity and their own learning while using it.

The most apparent evaluand was the class wiki. That was what the instructor explicitly invited the students to jointly evaluate. But she also invited them to help create it. So, the wiki emerged during its evaluation and changed throughout the semester by accretion from the entries participants made. Each student chose to contribute to the wiki in slightly different ways, and thereby made it their own to different degrees. Some students gravitated to some parts of the wiki more than others and their views of the value of the wiki modulated, as did their learning associated with what they put into and took from the wiki activity. Thus, the evaluand became a combination of the Internet and computers which made creation and use of the wiki possible, the pedagogical purposes guiding the instructor's identification of this learning community-building activity, the wiki product the collaborators generated together, the differential energy and interest given to this activity by the students, and the related learning activities and outcomes in which they participated.

The instructor's main criteria for the wiki activity were that the students would participate in using it to record their experiences and understandings associated with their readings and projects and they would learn better by participating in a collaborative learning community. The students' criteria varied from being able to earn a good grade by meeting the instructor's expectations sufficiently, to learning from the wiki activity, to expanding their personal-learning environments. They wanted to meet not only their class-learning goals but also their own personal-learning goals.

Based on the context, stakeholders, evaluands, and criteria, the main evaluation questions for this emerging evaluation became:

1. Do the students participate to the instructor's satisfaction in using the wiki?
2. Do the students learn better by participating in the wiki activity than they would have without participating?
3. Do the students meet additional personal goals to their satisfaction?

Clarifying What Is

The instructor selected formal methods for gathering and interpreting data regarding the e-learning associated with the students' use of the wiki. However, the students played key roles in that data collection and analysis and therefore modified it from the instructor's original plan.

Selecting methods to fit questions, data gathering, and analysis

As this was a class-level study, the instructor could have randomly assigned half the students to participate with the wiki and the other half to form a control group. However, she knew they would be referring to the students' use of the wiki during class discussions; so she decided she would compare the students' behaviors, performance, and attitudes to data she had collected from students in previous semesters and to the students' own judgments of their experiences and views before, during, and after the class.

In addition to her analysis of the wiki itself and the students' participation in writing entries in it, the instructor assessed students' performance on the course-learning outcomes and interviewed them about their learning experiences, opinions regarding their growth, and how well they achieved their personal goals associated with the wiki activity, if they had any. She also invited the students to interview one another about their experiences and individual goals they had set for themselves. The analyses were both quantitative and qualitative. The instructor counted the number of entries per student and the length of their entries as estimates of their level of participation. But she focused mainly on the quality of the students' entries, descriptions of their experiences and opinions about their growth, and understanding what their personal goals had been, how they evolved, and how well they were achieved.

Completing the Evaluation

The teacher invited the students to collaborate with her in comparing their experiences to their goals and to identify recommendations for continually evaluating and improving the use of technology to enhance learning in future semesters.

Comparing, reporting, recommending, and metaevaluating

In response to the evaluation questions, which reflected the instructor's and students' criteria, they found that 85% of the students participated very actively (with more than 50 words per entry and made entries in the wiki at least 90% of the 30 times they could have contributed). This was more than satisfactory to the instructor. In addition, the instructor judged that the students learned better by participating in the wiki activity than they would have without it, based on her experiences with prior classes and the students' ratings of their learning before, during, and after the semester. Finally, the majority of the students noted that they had developed and accomplished several additional personal goals to their satisfaction.

As this evaluation was completely internal and formative, the instructor did not write up a report but shared the

results with the students and with her rank-advancement committee to certify that she was seeking feedback on her teaching initiatives and would continue using the wiki as an e-learning tool because most students said it was helpful. She noted that a few students did not participate in using the wiki and they did not learn as well or have as positive views of the content as did those who participated. She visited with them about their experiences and decided she should explore ways to engage future students who do not participate in the wiki activity. The students who participated were so positive about their additional learning and achieving of personal goals that she decided to continue the wiki activity the coming semester with just a few modifications.

Students who participated most in the wiki activity also reported they expanded their personal-learning environments to include blogs, wikis associated with other classes, and tools associated with other social media they believe will enhance their formal education, personal, and informal learning experiences.

The instructor did not metaevaluate her evaluation of the wiki activity before or during its implementation, which she learned later is the recommended process. But she was able to go to a website where a checklist (Stufflebeam, 1999) guided her through several key questions regarding the evaluation in terms of metaevaluation standards. She plans to use what she learned from this self-evaluation to conduct better studies of her teaching and her students' e-learning in subsequent semesters.

Challenges and Opportunities

Challenges

Major challenges with evaluating e-learning include understanding how to engage stakeholders in clarifying their evaluands of interest and criteria for judging them, and helping stakeholders systematically gather and analyze data to compare against those ideals. Metaevaluations are needed to raise the quality of e-learning evaluations.

Why do these challenges persist? There is a tendency to champion new learning technologies without collecting associated evidence. There are also difficulties evaluating phenomena that are constantly changing and as complicated as those found in e-learning situations. As the WestEd (2008) studies show, evaluators of e-learning must anticipate difficulties and adjust designs to skillfully complete helpful studies. Multiple sources of information and replications of studies are needed to reach dependable conclusions.

Opportunities

In spite of challenges, there are hopeful opportunities associated with evaluation of e-learning. Government efforts to encourage evaluation of various forms of e-learning

instruction are growing. Many stakeholders in the private sector also acknowledge that data-based decisions can make their efforts more profitable, ethical, effective, and efficient. There have been significant advances in evaluation theory and methodology, which can enhance e-learning evaluations.

Perhaps most hopeful is the growth in social networks and other forms of communication that encourage learners to take responsibility for their own learning, whether in formal education settings or on their own. Development of personal-learning environments implies development of personal-evaluation skills through which learners decide what resources to use to meet their learning goals, test themselves through reality-based performance opportunities, and share their findings and self-evaluations with others through various peer-based reviews.

Research into e-learning itself will continue but inquiry into the evaluation of e-learning must also proceed. This field is dynamic and involves people of all ages and in all walks of life who are shaping their criteria and definitions of what constitutes e-learning success faster than evaluation can evolve by itself.

See also: An Evaluation Framework for E-Learning Effectiveness in the Arab World; An Overview of Technology and Learning; Assessment and the Regulation of Learning; Assessment in Schools – Technology Education and ICT; Challenges of Developing and Implementing Formative Assessment Practices in Schools; Classroom uses of Technology to Manage Instruction; Conceptions of Technology Literacy and Fluency; Cost Analysis in Evaluation Studies; Cultural Issues that Can Affect the Validity of Educational Evaluations; Curriculum, Digital Resources and Delivery; Curriculum Evaluation; Defining Quality in Evaluation; Effective Use of Technology in Teaching and Learning in HE; Evaluation and Accountability; Evaluation Methodology; Evaluation Reporting and Communicating; Evaluation Use; Examining the Effects of Educational Technology Programs: Challenges and Strategies; Formative Assessment and Instructional Planning; Formative Assessment; Informal Education and Evaluation; Instructional System Provided Feedback; Internal and External Evaluation; International Evaluations; Internet-based Education; Leadership and Technology; Learning as Inquiry; Learning from Multiple Information Sources; Learning Outside of School; Longitudinal Evaluation Designs; Media Use and School Achievement; Meta-evaluation: Purpose, Prescription, and Practice; Methods to Evaluate Technology; Moral and Ethical Issues in Evaluation; Needs Assessment in Education; New Media, Learning from; Peer and Self-assessment; Peer Interaction and Learning; Program Evaluation; Progress Monitoring; Summative Assessment by Teachers; Technology Supports for Acquiring Inquiry Skills; Technology Supports for Lifelong Learning; The Purpose of Educational Evaluation; The Role of Stakeholders in Educational Evaluation; Understanding Approaches to Evaluation.

Bibliography

- Advanced Distributive Learning (2004). *SCORM 2004*, 3rd edn. <http://www.adlnet.gov/scorm/index.aspx> (accessed August 2009).
- AEA (American Evaluation Association) (2003). *AEA Response to US Department of Education on Scientifically Based Evaluation Methods*. <http://www.eval.org/doestatement.htm> (accessed August 2009).
- AEA (American Evaluation Association) (2004). *American Evaluation Association Guiding Principles for Evaluators*. <http://www.eval.org/GPTraining/GPTrainingOverview.asp> (accessed August 2009).
- AERA (American Educational Research Association) (2003). *Letter to US Department of Education on Scientifically Based Evaluation Methods*. <http://www.eval.org/doeaera.htm> (accessed August 2009).
- Allen, I. E. and Seaman, J. (2007). *Online Nation: Five Years of Growth in Online Learning*. Needham, MA: Sloan-C.
- Carliner, S. and Shank, P. (eds.) (2008). *The E-Learning Handbook*. San Francisco, CA: Pfeiffer.
- Clifford, L. (2006). *Learner Experiences of E-Learning*. <http://www.jisc.ac.uk/whatwedo/programmes/elearningpedagogy/learnerxp.aspx> (accessed August 2009).
- Cousins, J. G. and Whitmore, E. (1998). Framing participatory evaluation. In Whitmore, E. (ed.) *Understanding and Practicing Participatory Evaluation. New Directions for Evaluation*, No. 80, pp 5–23. San Francisco, CA: Jossey-Bass.
- Graham, C. R. (2006). Blended learning systems: Definition, current trends, and future directions. In Bonk, C. J. and Graham, C. R. (eds.) *Handbook of Blended Learning: Global Perspectives, Local Designs*, pp 3–21. San Francisco, CA: Pfeiffer.
- International Reading Association (2004). *Standards for Reading Professionals*. <http://www.reading.org/downloads/resources/545standards2003/index.html>.
- Ito, M., Horst, H., Bittanti, M., et al. (2008). *Living and Learning with New Media: Summary of Findings from the Digital Youth Project*. Chicago, IL: MacArthur Foundation. <http://digitalyouth.ischool.berkeley.edu/report> (accessed August 2009).
- Khribi, M. K., Jemni, M., and Nasraoui, O. (2008). Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval. In *Proceedings of the 2008 Eighth IEEE International Conference on Advanced Learning Technologies*, pp 241–245. IEEE Computer Society. <http://portal.acm.org/citation.cfm?id=1381543&jmp=cit&coll=&dl=> (accessed August 2009).
- Koehler, M. J. and Mishra, P. (2008). Introducing TPCK. In AACTE Committee on Innovation and Technology (ed.) *Handbook of Technological Pedagogical Content Knowledge (TPCK) for Educators*, pp 3–29. New York: Routledge.
- Lorenzo, G. and Moore, J. (2002). *The Sloan Consortium Report to the Nation: Five Pillars of Quality Online Education*. <http://www.aln.org/effective/pillarreport1.pdf> (accessed August 2009).
- National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. <http://standards.nctm.org> (accessed August 2009).
- Oblinger, D. G. and Hawkins, B. L. (2006). The myth about no significant difference “using technology produces no significant difference.” *EDUCAUSE Review* 41(6). <http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume41/TheMythaboutNoSignificantDiffe/158103> (accessed August 2009).
- Rooney, J. E. (2003). Blending learning opportunities to enhance educational programming and meetings. *Association Management* 55(5), 26–32.
- Sanders, J. R. (1994). *The Program Evaluation Standards: How to Assess Evaluations of Educational Programs*. Thousand Oaks, CA: Sage.
- Stufflebeam, D. L. (1999). Program evaluations metaevaluation checklist. http://www.wmich.edu/evalctr/checklists/program_metaeval_10point.htm (accessed August 2009).
- Wesch, M. (2007). Human futures for technology and education. *Presentation at the Sixth Annual ECAR/HP Summer Symposium for Higher Education IT Executives*, 11–13 June 2007, Boulder, CO. <http://www.educause.edu/ECAR/HumanFuturesforTechnologyandEd/161880> (accessed August 2009).

- WestEd (2008). *Evaluating Online Learning: Challenges and Strategies for Success*. <http://www.wested.org/cs/we/view/rs/890> (accessed August 2009).
- Williams, D. D. (2006). Measurement and assessment supporting evaluation in online settings. In Williams, D. D., Hricko, M., and Howell, S. (eds.) *Online Assessment, Measurement and Evaluation: Emerging Practices*, pp 1–9. Hershey, PA: Idea Group.
- Williams, D. D. and Campbell, J. O. (2001). Effectiveness of e-learning instruction and of evaluations of e-learning. *Paper Presented at the Annual Meeting of the American Evaluation Association*, 7–10 Nov, St. Louis, MI (accessed August 2009).
- Young, J. R. (2002). 'Hybrid' teaching seeks to end the divide between traditional and online instruction. *Chronicle of Higher Education*, A33. <http://chronicle.com/weekly/v48/i28/28a03301.htm> (accessed August 2009).
- ## Further Reading
- Ardito, C., Costabile, M. F., Marsico, M. D., et al. (2006). An approach to usability evaluation of e-learning applications. *Universal Access in the Information Society* 4(3), 270–283.
- Arthur, W., Bennett, W., Edens, P. S., and Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology* 88(2), 234–245.
- Attwell, G. (ed.) (2006). *Evaluating E-Learning: A Guide to the Evaluation of E-Learning*. <http://www.pontydysgu.org/2007/11/evaluating-e-learning> (accessed August 2009).
- Bernard, R. M., Abrami, P. C., Lou, Y., et al. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research* 74(3), 379–439.
- Boettcher, J. V., Justice, L., Schenk, K., Rogers, P. L., and Berg, G. A. (2005). *Encyclopedia of Distance Learning*. Hershey, PA: IGI Global.
- Bonk, C. J. and Graham, C. R. (eds.) (2006). *Handbook of Blended Learning: Global Perspectives, Local Designs*. San Francisco, CA: Pfeiffer.
- Britain, S. and Liber, O. (2004). A framework for the pedagogical evaluation of elearning environments. http://www.jisc.ac.uk/index.cfm?name=project_elearn_ped_learning_design_tools (accessed August 2009).
- Chen, C. L. (2006). Development. In Berman, P. (ed.) *E-Learning Concepts and Techniques*. http://iit.bloomu.edu/Spring2006_eBook_files/index.htm (accessed August 2009).
- Clark, D. (2002). Psychological myths in e-learning. *Medical Teacher* 24(6), 598–604.
- Cochrane, T. (2006). Learning with wireless mobile devices and social software. *Proceedings of the 23rd Annual Ascilite Conference: Who's Learning? Whose Technology?* 143. http://64.233.179.104/scholar?hl=en&lr=&q=cache:2Fz5qswalhCJ:ascilite.org.au/conferences/sydney06/proceeding/pdf_papers/p50.pdf+social+media+learning+evaluation (accessed August 2009).
- Corte, E. D., Verschaffel, L., Entwistle, N., and Merrienboer, J. V. (2003). *Powerful Learning Environments*. Oxford, UK: Elsevier.
- Farmer, B., Yue, A., and Brooks, C. (2008). Using blogging for higher order learning in large cohort university teaching: A case study. *Australasian Journal of Educational Technology* 24(2), 123–136. <http://www.ascilite.org.au/ajet/ajet24/farmer.html> (accessed August 2009).
- Joint Committee on Standards for Educational Evaluation (2008). *Various Evaluation Standards*. <http://www.wmich.edu/evalctr/jc> (accessed August 2009).
- Lam, P. and McNaught, C. (2005). Management of an e-learning evaluation project: e3learning. In *Proceedings of the 17th Annual World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pp 2261–2268.
- McNaught, C. and Lam, P. (2005). Building an evaluation culture and evidence base for e-learning in three Hong Kong universities. *British Journal of Educational Technology* 36(4), 599–614.
- Moore, M. G. and Anderson, W. G. (eds.) (2003). *Handbook of Distance Education*. Mahwah, NJ: Erlbaum.
- Reeves, T. C. (2002). Keys to successful e-learning: Outcomes, assessment, and evaluation. *Educational Technology* 42(6), 23–29.
- Ruiz, J. G., Mintzer, M. J., and Leipzig, R. M. (2006). The impact of e-learning in medical education. *Academic Medicine: Journal of the Association of American Medical Colleges* 81(3), 207–212.
- Sitzmann, T., Kraiger, K., Stewart, D., and Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. *Personnel Psychology* 59(3), 623–664.
- Waight, C. L., Willging, P. A., and Wentling, T. L. (2002). Recurrent themes in e-learning: A meta-analysis of major e-learning reports. http://learning.ncsa.uiuc.edu/papers/AHRD2002_waight-willging-wentling.pdf (accessed August 2009).
- Welsh, E. T., Wanberg, C. R., Brown, K. G., and Simmering, M. J. (2003). E-learning: Emerging uses, empirical results and future directions. *International Journal of Training and Development* 7(4), 245–258.
- Wiecha, J. M., Gramling, R., Joachim, P., and Vanderschmidt, H. (2003). Collaborative e-learning using streaming video and asynchronous discussion boards to teach the cognitive foundation of medical interviewing: A case study. *Journal of Medical Internet Research* 5(2), e13.
- Williams, D. D., Hricko, M., and Howell, S. (eds.) (2006). *Online Assessment, Measurement and Evaluation: Emerging Practices*. Hershey, PA: Idea Group.
- ## Relevant Websites
- <http://blogs.aace.org> – Association for the Advancement of Computing in Education (AACE), with information on e-learning and related topics.
- <http://www.aace.org> – Association for the Advancement of Computing in Education, International Journal on E-Learning.
- <http://link.brightcove.com> – Brightcove, Wired Campus TV.
- <http://chronicle.com/blog/Wired-Campus/5> – Chronicle's Wired Campus Newsletter; Chronicle of Higher Education Information Technology.
- <http://online-journals.org/i-jet> – Online Journals, International Journal of Emerging Technologies in Learning (iJET).
- <http://www.shambles.net> – Shambles, Training Evaluation.
- <http://www.sreb.org> – Southern Regional Educational Board, Educational Technology Cooperative.
- http://video.google.com/videosearch?q=personal+learning+environments&ie=utf-8&oe=utf-8&rls=org.mozilla:en-US:official&client=firefox-a&um=1&sa=X&oi=video_result_group&resnum=4&ct=title# – Videos on Personal Learning Environments.
- <http://www.inacol.org/research/nationalstandards/>

Evaluating Schools as Learning Communities

S M Thomas, University of Bristol, Bristol, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

CPD – Continuing Professional Development.
DCSF – UK's Department for Children, Schools and Families.
DFES – UK's Department for Education and Science (replaced by DCSF in 2007).
GCSE – General Certificate of Secondary Education.
GNVQ – General National Vocational Qualifications.
GTC – UK's General Teaching Council.
KS2 – National Curriculum Key Stage Two (England).
KS4 – National Curriculum Key Stage Four (England).
NCSL – UK's National College for School Leadership.
PLC – Professional Learning Community.
PLCs – Professional Learning Communities.
VA – Value-Added measure.

Introduction

Recent emphasis on the fundamental – but not particularly new – idea that teachers working and learning together can enhance the quality of learning and teaching in schools has led to renewed interest in the potential of professional learning and professional learning communities (PLCs) to advance school improvement. Indeed, it is now central to England's National College for School Leadership's (NCSL's) revised National Standards for Head teachers and the Department for Children, Schools and Families (DCSF) Core Principles for raising standards in teaching and learning. Research studies in a range of international contexts, particularly in the USA, have similarly emphasized the value and role of professional learning communities to enhance professional practice (Stoll and Louise, 2007, Hipp *et al.*, 2008). Effective PLCs are seen as more likely to generate and support sustainable improvements in student and other outcomes because they build the necessary professional skill, capacity, and shared good practice. However, can the effectiveness and quality of a PLC be measured and does it have a quantifiable positive impact on pupil outcomes? Although relatively few research studies address these specific questions, this article reviews the available evidence and reports in detail select findings from a large-scale research project funded by the UK's Department for Education and Science (DFES),

General Teaching Council (GTC), and National College for School Leadership (NCSL) 'Creating and Sustaining Effective Professional Learning Communities' (Bolam *et al.*, 2005). In particular, this article illustrates the quantitative research tools used to define and measure specific aspects of PLCs and examines the relationship between these measures and pupils' academic outcomes. At this point, it is important to point out that the term professional learning community, used in this article, overlaps with – but is, nevertheless, more focused on – professional staff learning, than the term school learning community which is, generally, seen as broader and refers to all learners in the school community.

Approaches to Evaluate Schools as Learning Communities

One common approach to evaluating schools as learning communities is to examine the school characteristics and processes that have been identified as key features of a PLC. International research suggests that PLCs are characterized by five aspects: shared values and vision; collective responsibility for pupils' learning; reflective professional inquiry; collaboration; and the promotion of group, as well as individual learning (Stoll *et al.*, 2003, 2006a). However, more recent research has identified three new characteristics and four processes relevant to the creation and effective functioning of PLCs (Bolam *et al.*, 2005). Therefore, although PLCs do not look exactly the same in all schools, the research has found that they seem to display eight characteristics which – irrespective of the educational phase or context – they exhibit to a greater or lesser degree:

1. shared values and vision;
2. collective responsibility for pupils' learning;
3. collaboration focused on learning;
4. individual, group, and collective professional learning;
5. reflective professional enquiry;
6. openness, networks, and partnerships;
7. inclusive membership; and
8. mutual trust, respect, and support.

In addition, four processes are required to create and sustain a PLC:

1. optimizing resources and structures to promote the PLC;

2. promoting individual and collective professional learning;
3. evaluating and sustaining the PLC over time; and
4. leading and managing to promote the PLC.

These findings are illustrated in greater detail below by presenting select results from a questionnaire survey used to test out the relevance of different characteristics and processes of PLCs in different kinds of schools in England (Bolam *et al.*, 2005).

A second approach to evaluating schools as learning communities is to use statistical techniques to identify and examine a finite set of key indicators or factors related to the processes of continuing professional development (CDP) and, more broadly, to the development of PLCs. This approach has been used in a study in the Netherlands by Visscher and Witziers (2004) who identified six PLC concepts (policy and evaluation; consultation and cooperation; consensus; decision-making, school leadership; and departmental leadership) and measured these via a teacher questionnaire administered to mathematics teachers in 39 schools. Pedder *et al.* (2005) have also explored the nature of professional learning via a factor analysis of teachers' responses to a questionnaire and identified four factors related to teacher learning. The factors were:

1. *Inquiry*. Using and responding to different sources of evidence; carrying out joint research and evaluation with colleagues.
2. *Building social capital*. Learning, working, supporting, and talking with each other.
3. *Critical and responsive learning*. Through reflection, self-evaluation, experimentation, and by responding to feedback.
4. *Valuing learning*.

One purpose of this kind of approach is to establish the existence of one or more factors that could be useful as indicators to measure and evaluate particular aspects of schools' capacity and success in developing as a PLC. Moreover, indicators of this type could provide a helpful feedback to inform school self-evaluation and improvement processes (see, e.g., Scheerens *et al.*, 2003).

A third – and, arguably, the most crucial – approach to evaluating schools as learning communities is to examine the relationship between the indicators of PLCs and a range of pupil outcome data in order to establish whether or not the extent of school characteristics and processes that are associated with PLCs are also linked to improved student outcomes. There is very limited evidence with regard to the impact of continuing professional development – and, more generally, of PLCs – on student outcomes and progress (see, e.g., Lee and Smith, 1996; Louis and Marks, 1998; Wiley, 2001). However, findings in this area are extended by a more recent study (Bolam *et al.*, 2005), which provides a rare

opportunity – via the collection of school-questionnaire-survey data alongside access to the national pupil-assessment databases – to produce some relatively hard data. The nature of this data is, of course, correlational – not causal – but, nevertheless, this evidence is arguably more robust than previous studies and is employed in this context to illustrate, in detail, the findings and methods used to evaluate the effectiveness of PLCs in terms of promoting student outcomes.

The three approaches mentioned above are discussed, in detail, in the following sections by drawing on findings mainly from Bolam *et al.* (2005) work.

Characteristics and Processes of Schools as PLCs

As noted above, previous research has identified eight key characteristics and four processes relevant to the creation and sustainability of schools as PLCs. The study conducted by Bolam and colleagues employed both case-study and questionnaire-survey methods to draw its final conclusions. However, the detailed findings from the questionnaire administered to almost 400 English schools provide a useful example of the most and least common PLC features reported by schools.

The questionnaire was developed from a review of the literature on PLCs and was designed to be completed by the headteacher or CDP coordinator of each school participating in the study – in consultation with other staff. The questionnaire comprises three parts:

- *Part 1*. Items designed to gather opinions regarding aspects of PLC in the school; these questions are framed in terms of the proportion of staff in the school engaged in particular activities or having particular views. These items reflect different aspects of CDP and teachers, professional life: teacher support for pupil learning; professional learning; support for professional learning; collaboration and culture; research and professional enquiry; external links and networking; and participation of teaching assistants and nonteaching support staff in PLC.
- *Part 2*. Items exploring the perceptions of a definition of a PLC and the factors that the respondent felt facilitated, or were barriers, to the school becoming a PLC.
- *Part 3*. Items related to factual information with regard to the range and extent of professional development and school self-evaluation activities in the school.

The questionnaire findings indicated which PLC characteristics and processes the respondents were most and least likely to report, on average – in terms of the proportion of staff engaged in particular PLC activities, as well as which activities were the most and least variable across schools – thereby suggesting agreement among schools on the importance of some key PLC features (see **Table 1**).

Table 1 Perceived aspects of PLC by primary schools and secondary schools: three items (indicated with the original questionnaire numbers) with the highest/lowest means and standard deviations

	<i>Mean</i>	<i>Standard deviation</i>	<i>n</i>
<i>Highest mean – primary schools</i>			
Q1: collective responsibility for pupil learning	3.77	0.56	224
Q3: create conditions for pupils to feel the confidence to learn	3.74	0.52	225
Q4: learn together with colleagues	3.73	0.51	223
<i>Highest mean – secondary schools</i>			
Q1: collective responsibility for pupil learning	3.64	0.54	164
Q3: create conditions for pupils to feel the confidence to learn	3.40	0.61	164
Q24: are members of at least one professional team	3.80	0.55	160
<i>Lowest mean – primary schools</i>			
Q9: carry out classroom-based research	1.81	1.00	209
Q36: experience job rotation	1.83	1.07	214
Q38: have opportunities for work shadowing	1.75	1.03	215
<i>Lowest mean – secondary schools</i>			
Q36: experience job rotation	1.13	0.34	160
Q38: have opportunities for work shadowing	1.36	0.71	157
Q44: receive financial support from the school for award-bearing courses	1.42	0.85	153
	<i>Mean</i>	<i>Standard Deviation</i>	<i>n</i>
<i>Highest standard deviation – primary schools</i>			
Q13: have dedicated time for classroom observation	2.48	1.25	222
Q44: receive financial support from the school for award-bearing courses	1.95	1.26	207
Q45: have some protected time for joint planning and development	2.57	1.37	218
<i>Highest standard deviation – secondary schools</i>			
Q13: have dedicated time for classroom observation	2.28	1.13	162
Q43: use professional development profiles/portfolios	2.41	1.14	158
Q45: have some protected time for joint planning and development	2.36	1.18	163
<i>Lowest standard deviation – primary schools</i>			
Q1: collective responsibility for pupil learning	3.77	0.56	224
Q4: learn together with colleagues	3.73	0.51	223
Q18: share a common core of educational values	3.71	0.55	225
<i>Lowest standard deviation – secondary schools</i>			
Q1: collective responsibility for pupil learning	3.64	0.54	164
Q24: are members of at least one professional team	3.80	0.55	160
Q36: experience job rotation	1.13	0.34	160

Only includes items from the PLC survey part 1.

The four-item response categories are treated as a numerical scale from one to four as follows: nearly all staff (>80%) = 4, most staff (50–80%) = 3, some staff (20–49%) = 2, a few staff (<20%) = 1.

Highest/lowest mean in bold; highest lowest standard deviation in bold (3).

Adapted from Bolam, *et al.* (2005). Creating and sustaining effective professional learning communities. *DfES Research Report RR637*. University of Bristol.

The findings show that two question items, in particular, are crucial in terms of primary and secondary schools reporting the highest mean level of teacher involvement. These are: collective responsibility for pupil learning and creating conditions for pupils to feel the confidence to learn. Responses to these two items also demonstrate a high level of consensus among schools, given the low item-score variability in comparison to other items. These findings suggest, not surprisingly, that pupil learning is the foremost concern of teachers and that this is a PLC characteristic that is very common in all schools.

In contrast, the results for two different items (experience job rotation and have opportunities for work shadowing) indicate that both primary and secondary schools,

typically, reported a low level of teacher involvement in these activities. The finding suggests that these particular aspects of CDP related to learning and developing staff roles, were not a common characteristic in most schools.

Finally, the results for two further items (have dedicated time for classroom observation and have some protected time for joint planning and development) suggest that, although a medium level of staff involvement is reported by primary and secondary schools, there is a notable lack of consensus (i.e., high item-score variability) concerning these two activities across schools. This finding suggests that some aspects of CDP – concerned with improving classroom practice – are very variable across schools, with some schools reporting a high level of

Table 2 Percentage of schools reporting different stages of PLC development

Stage	Nursery and primary schools (n = 227)	Secondary schools (n = 166)
Mature/ established	25	16
Developer	57	67
Starter	14	15
Working to reestablish PLC	2	1
No response	2	1

The primary and the secondary school samples included special schools, respectively.

Adapted from Bolam, *et al.* (2005). Creating and sustaining effective professional learning communities. *DfES Research Report RR637*. University of Bristol.

staff involvement and others reporting a low level of staff involvement. The issues of teachers' time management and professional culture may have an impact here.

Schools in the sample were found to be adequately representative of the national picture in terms of context factors, such as the percentage of pupils in the school entitled to free school meals. However, it is important to note that schools were asked specifically, in part 2 of the survey, to assess their current stage of development as a PLC. The provisional working definition of a PLC quoted in the survey was as follows: "An effective professional learning community has the capacity to promote and sustain the learning of all professionals and other staff in the school community with the collective purpose of enhancing pupil learning." A majority of respondents – 57% from nursery and primary schools and 67% from secondary schools – reported that their school was a developing PLC (see **Table 2**). Therefore, the survey findings reported above reflect the features of this self-reported group of schools more than any other. Moreover, it is important to note that the survey findings are a summary of data collected at one point in time from one respondent – either the head or CDP coordinator – in each school responding to the survey. Of course, the situation may have changed since the data were collected and, in addition, other members of the school may have different opinions concerning the questions. Clearly, these issues should be considered when interpreting the survey findings, although interviews in case-study schools – also conducted as part of the study – generally confirmed survey respondents' perceptions of their school's current stage of PLC development.

Summary Indicators of PLCs

In addition to the questionnaire-item results outlined above, the study also provides a useful example of findings

Table 3 Questionnaire items included on the factor of professional and pupil learning ethos

Factor 1	Professional and pupil learning ethos
1:	Collective responsibility for pupil learning
2:	Base their approach to change on the use of good evidence
3:	Create conditions for pupils to feel the confidence to learn
4:	Learn together with colleagues
5:	Ensure pupils receive constructive feedback about their work
6:	Set learning targets for individual pupils
7:	Routinely collect, analyze, and use data and evidence to inform their practice
8:	Have low expectations of children
9:	Actively seek and use feedback from pupils
10:	Regularly monitor the learning and progress of individual pupils
11:	Share a common core of educational values
12:	Use the staff room at break times for professional links
13:	Are involved in seeking solutions to problems facing the school
14:	Regularly discuss teaching methods
15:	Share their professional experiences and successes
16:	Experiment and innovate in their work
17:	See the school as stimulating and professionally challenging
18:	Routinely share information with parents and the community
19:	Learn from each other
20:	Take responsibility for their own professional learning
21:	Give priority to learning more about pupils learning

that employ quantitative factor-analysis techniques to establish the existence of one or more summary indicators or factors that could be used to measure and evaluate schools as professional learning communities. Four independent factors were identified from the questionnaire results which relate to:

1. professional and pupil learning ethos;
2. within-school policy, management, and support for professional learning;
3. enquiry orientation (external and internal); and
4. participation of nonteaching support staff in PLC.

As an illustration of the results, the questionnaire items loading on factor 1 (professional and pupil learning ethos) are shown in **Table 3** (for further details, see Atkins and Thomas, 2004). Correlations between the factor scores showed that, although there is some overlap between factors (i.e., all correlations are positive), there are, nevertheless, observable differences between these PLC summary indicators. Therefore, the evidence indicates that at least four dimensions reflecting different aspects of PLCs exist and individual schools can have a very different profile of results across the four factors. The findings related to professional and pupil learning and enquiry orientation are also reflected in a comparable questionnaire study which focused specifically on professional learning and was conducted on the basis of a teacher survey – over 1000 individual teacher responses from 32 schools (Pedder *et al.*, 2005). As previously noted, in

the study carried out by Pedder and his colleagues, four factors were identified using standard factor-analysis techniques which comprised aspects of inquiry and using different sources of evidence, building social capital, critical and responsive learning, and valuing learning. Broad similarities could also be found with the PLC concept factors identified by Visscher and Witziers (2004).

The validity of the PLC concept at the primary and secondary phases of education was also examined by Bolam and colleagues in terms of whether schools from different educational phases report similar levels of staff involvement as measured by the PLC factors. **Table 4** depicts the mean and standard deviation of the factor scores from primary and secondary schools. Overall, the findings indicate that primary schools, generally, report higher levels of staff involvement in PLC processes than secondary schools. Interestingly, the highest level of primary-phase staff involvement suggested by the factor scores is in terms of pupil and professional learning ethos (factor 1). In contrast, for the secondary phase, the highest level of staff involvement is in terms of enquiry orientation (external and internal; factor 3).

In the survey, respondents were also asked to assess their current stage of development as a PLC in three developmental stages: mature/established, developing, and starting the journey to become a PLC. The validity of conceptualizing PLCs in three stages was examined by looking at the relationship between schools' four-factor scores and their self-reported stage as a mature, developer, or starter. In some cases, the findings suggest an overlap between mature and developer PLCs – with starter PLCs falling behind on some factors. In addition – not surprisingly – in all cases, the primary mean factor score for a particular PLC stage is higher than the equivalent secondary mean factor score, again reflecting the overall tendency for primary schools to have reported higher teacher PLC involvement than secondary schools. Given these findings, it is not clear whether it should necessarily be expected that primary and secondary phases would report a similar level of staff involvement

in the PLC. This is because secondary schools are, usually, much larger and it is possible that smaller PLCs exist within the schools surveyed but were not identified because the focus was at the whole-school level. For example, one or more secondary school subject departments may operate as a PLC and display the characteristics and features of a mature PLC, but the proportion of total school staff involved may be low.

Nevertheless, overall – from these findings – it appears that whole-school PLCs within secondary schools may be less developed than those within primary and nursery schools, because the results indicate that a smaller percentage of staff are perceived to be engaging in the measured PLC behaviors. Of course, such a response pattern is not uncommon in questionnaires of this type (e.g., McCall *et al.*, 2001), and achieving secondary school improvement is notoriously more difficult than that in primary schools (Louis and Miles, 1990).

Overall, the results from the PLC factors do appear to be broadly in line with schools' self-reported stage of development – in all cases, schools reported as mature PLCs have higher average factor scores than starter PLCs. Despite this, it is also important to point out that by combining the results from different questionnaire items into summary indicators (i.e., factor scores), a greater degree of differentiation can be employed in measuring each factor than the three crude PLC developmental stages of mature, developers, and starter. This evidence does seem to suggest that the conceptualization of a PLC needs to shift from the idea of a single dimension in three developmental stages to multiple dimensions – each being on a continuum – although the two approaches could also be seen as broadly complementary.

Relationship Between Pupil Outcomes and PLC Processes and Contexts

The research conducted by Bolam and colleagues also provides evidence of the relationship between the

Table 4 Means and standard deviations of the four PLC factor scores by educational phase

	<i>Educational phase</i>			
	<i>Nursery and primary schools n = 227</i>		<i>Secondary schools n = 165</i>	
	<i>Mean</i>	<i>Standard deviation</i>	<i>Mean</i>	<i>Standard deviation</i>
<i>Learning ethos (factor 1)</i>	0.29	0.94	–0.41	0.95
<i>Support for professional development (factor 2)</i>	0.28	1.03	–0.38	0.82
<i>Enquiry orientation (factor 3)</i>	0.22	1.06	–0.32	0.81
<i>Participation of nonteaching staff (factor 4)</i>	0.26	0.82	–0.36	1.12

Factor scores were standardized across the whole sample with mean = 0 and standard deviation = 1.

The primary and the secondary school samples included special schools, respectively.

Adapted from Bolam, *et al.* (2005). Creating and sustaining effective professional learning communities. *DfES Research Report RR637*. University of Bristol.

characteristics of PLCs and pupil outcomes and these findings address the crucial question of whether there is a statistically significant relationship between school PLC indicators (e.g., factor scores) and school academic performance measures. As outlined in greater detail below, only tentative evidence of a positive relationship has been found. Nevertheless, the findings support previous research by several researchers. Lee and Smith (1996: 103) found that “Achievement gains are significantly higher in schools where teachers take collective responsibility for students’ academic success or failure rather than blaming students for their own failure. Achievement gains were also higher in schools with more cooperation among staff.” Louis and Marks (1998) also found that the promotion of a professional community for teachers has a positive relationship with the organization of classrooms for learning and the academic performance of students. Similarly, Wiley (2001) found a positive effect for professional community on student achievement – but only in situations where teachers’ experience of (transformational) leadership was also positive.

Relatively sophisticated analyzes (multilevel modeling) have also been conducted to establish whether or not there is a statistically significant link between schools’ value added in terms of pupils’ relative progress and their capacity as a PLC. Crucially, sophisticated value-added (VA) measures of relative pupil progress provide a more valid measure of educational and school effectiveness than the raw results (Thomas and Mortimore, 1996). However, for comparison purposes, raw (unadjusted) assessment and examination results have also been employed – as published in the annual league tables in the UK.

Calculating VA and raw school performance

A variety of different multilevel models were tested prior to deciding on the final ones to use for calculating VA school performance (see Thomas (2001) for a similar approach). This approach was seen as preferable to assess the long-term impact of a school on a pupil’s progress in learning and attainment as the models utilize outcome and baseline data that span, by and large, the period spent in that school phase by pupils. However, both the VA and raw (unadjusted) models have been used to analyze the end of phase-schooling outcomes. The Outcomes for primary schools comprise the mean National Curriculum Key Stage 2 (KS2) assessment score across the core subjects of English, mathematics, and science for 11-year-olds. Outcomes for secondary schools comprise the Total National Curriculum Key Stage 4 (KS4) assessment score, including grade scores for all subjects taken by 16-year-olds in the General Certificate for Secondary Education and in General National Vocational Qualifications (GCSE/GNVQ).

The final multilevel VA models take account of the variability in pupil outcomes due to the characteristics of the

pupil intake for a school. Thus, pupils’ prior attainment, gender, socioeconomic background, and other variables are controlled for in the statistical analysis (see Smith and Thomas, 2004). The VA score (or residual) for each school is determined from the model results and used to make more robust (reliable) comparisons about how pupils in a school perform compared with pupils in other schools. The raw model is used to calculate school performance in terms of raw pupil outcomes when no adjustment for intake and other factors has been made.

Correlations between school performance and PLC indicators

The four PLC factors were correlated with primary and secondary schools’ VA and raw performance. However, only a minority of these comparisons were found to be statistically significant at the 0.05 level and the findings are summarized below (see Bolam *et al.*, 2005, for further details of statistical results). At the primary level, positive and statistically significant correlations were found between schools’ factor 1 score professional and pupil learning ethos and their 2002 KS2 performance – both raw and VA. At the secondary level, positive and statistically significant correlations were also found for both factor 1 professional and pupil learning ethos and factor 2 support for professional learning – but only in terms of schools’ VA GCSE/GNVQ performance. Of course, it should be noted that the relationships are fairly weak (no correlations reported are greater than 0.3). However, it could be argued that this would be expected, given that any statistical relationship is likely to be tenuous between process measures collected as a snapshot at one point in time and pupil performance and progress over a relatively long time period (4–5 years).

The relationship between pupil outcomes and the individual items from part 1 of the survey, were also examined. As expected, the findings support those from the PLC factor results reported above. Several positive and statistically significant correlations (at 0.05 level) were found, but only two were fairly strong ($r \geq 0.3$) and these relate only to the primary phase (teachers have low expectations of pupils and teachers share a common core of educational values – note that both of these items are included on factor 1).

There are only three items that are statistically significant (at 0.05 level) across both the primary and secondary phases and the evidence tentatively indicates that these aspects of a PLC are key in terms of enhancing pupil progress. Positive relationships were found between schools’ VA residual outcomes and their responses to the following items teachers create the conditions for pupils to feel the confidence to learn, teachers share a common core of educational values, and teachers see the school as stimulating and professionally challenging. A positive

relationship (at the 0.05 level) was also found between primary VA residuals and teachers' experiments and innovation in their work.

Correlations between school performance and PLC factual characteristics

The relationship between pupil outcomes and the PLC factual characteristics, from part 3 of the survey, were also examined. Both positive and negative statistically significant correlations (at 0.05 level) were found – six indicating a fairly strong relationship ($r \geq 0.3$). These six strongest correlations were found between pupil outcomes and items related to the status of the school – that is, being in an Education Action Zone or part of an Excellence in Cities initiative (negative relationships) or being a Beacon School, having other working links or Specialist status (positive relationships). Clearly, a school's involvement in these kinds of initiatives is often related to school performance and effectiveness. (A school being in an Education Action Zone or Excellence in Cities initiative indicates that it receives additional funding and support due to being located in a particularly economically disadvantaged area. Beacon School or specialist status indicates that a school has met the criteria for excellence and/or additional resources in particular aspects of the curriculum.)

In addition, interestingly at the primary level, weak but nevertheless statistically significant positive associations were found between school VA outcomes and the degree to which staff were monitoring pupil progress. This finding supports previous research that underlines the need for systematic monitoring of pupil progress in schools to improve school effectiveness (Thomas *et al.*, 2007). At the primary level, the total number of supply teaching days for 2001/02 was significantly negatively correlated with raw pupil outcomes suggesting that frequent teacher absences have a detrimental effect on pupil outcomes.

Finally, in addition to the correlational analyzes described above, further more sophisticated multilevel analyzes were also carried out to determine whether the previously identified PLC factors and other school characteristics have a combined (as well as individual) impact on pupil progress in the VA models (i.e., given pupil intake factors have already been accounted for in the analysis) and whether they further reduced the remaining unexplained school and total variance. In line with the correlational analyzes, it was found that, in the primary school analyzes, the addition of factor 1 professional and pupil learning ethos to the VA model was statistically significant (at the 0.05 level); however, this resulted in a very small reduction in school and total variability in KS2 mean score outcomes. Similarly, in the secondary school analyzes, the addition of both factor 1 professional and pupil learning ethos and factor 2 internal support for professional

learning individually into the VA model, was statistically significant. However, overall, the findings indicated that any positive link between school PLC factors and pupil outcomes is fairly weak and may be difficult to distinguish when other (possibly confounding) school factors and context – such as percentage of pupils entitled to free school meals – is taken into account.

In summary, these findings are interesting as they do suggest that a weak, but positive, link exists between individual PLC factors and pupil outcomes – particularly, VA performance, which measures the relative progress of pupils in a school in comparison to pupils in other schools (arguably one of the most valid measures of school effectiveness). Findings of this kind are important because it may be argued that particular features of PLCs can only be validly evaluated as effective if a positive impact on student outcomes can be demonstrated. Of course, it may be that quantitative measures used in the research conducted so far are not sufficiently sensitive to fully reflect the complex relationship among PLCs and pupil performance and progress. Therefore, clearly, further research is necessary to contribute stronger evidence on this issue which has only rarely been addressed using sophisticated quantitative techniques.

Conclusion

In conclusion, a limited number of relevant studies have been reviewed and the findings indicate that some key PLC characteristics are more (or less) commonly reported in schools (e.g., opportunities for work shadowing and experience job rotation were, generally, less common). In addition, some key PLC characteristics are more-or-less variable among some schools than others (e.g., dedicated time for classroom observation and protected time for joint planning and development were more variable). These findings reveal an important message for policymakers – that there are some key areas of teachers' professional development that either happen very infrequently in schools or are highly variable across schools.

The Bolam *et al.* (2005) and other similar studies found that the summary indicators of different aspects of a PLC can be identified using quantitative factor-analysis techniques. However, it is important to recognize that the different factors identified are not, necessarily, comprehensive for the purpose of evaluating the nature and quality of PLCs. The qualitative case-study data, also reported by Bolam *et al.* (2005), tapped into considerable detail and depth concerning the nature of PLCs and, subsequently, revealed important new aspects of PLC processes and characteristics such as trust and distributed leadership, which were not directly included on their questionnaire survey.

Moreover, in Bolam and colleague's final conclusions, it was suggested that PLCs should be evaluated not only in terms of pupil outcomes – indeed, overall, the quantitative evidence linking student outcomes and PLC functioning, seems to be fairly weak – but also via impact on staff morale and practice, and potential for developing leadership capacity, as well as by establishing that PLC characteristics are in place and processes are operating smoothly. Further research is, of course, needed to explore these issues in more detail, given that future theory building will require stronger evidence of this kind to clearly differentiate between the effects of the wide varieties of learning, development, and improvement activities that teachers, schools, and PLCs engage in.

In conclusion, this article indicates fairly strongly that a school, as a PLC, is multidimensional and complex in nature. Evidence also, tentatively, suggests that some PLC characteristics and processes may be enhanced or hindered by the advantages or disadvantages of particular contexts (e.g., smaller size may allow greater opportunities for collaboration, higher percentage of student disadvantage indicated by their eligibility for free school meals may indicate higher funding levels for professional development activities). The idea of three developmental stages (mature, developer, and starter) may be helpful as a starting point in understanding the PLC concept but, in the end, may be too crude to measure the complexity of PLC characteristics and processes. This article demonstrates that a more detailed approach is needed to help teachers and other professionals working in schools to understand the conditions and contexts that underpin systematic and effective professional learning as well as to enhance PLC development in schools. Further evidence is also needed to better understand the links between professional learning and the learning of other members of the school community – such as students and parents – in order to explore whether the processes and outcomes of a PLC mirror those of the school learning community as a whole.

Overall, the review findings indicate that quantitative and qualitative methods can be used to evaluate different aspects of the effectiveness and quality of a PLC and demonstrates the potential for creating valid measures of professional learning within schools. These kinds of measures could provide a useful method of enhancing information available to schools and, if used alongside other types of evidence, could meaningfully inform school improvement and self-evaluation activities (see Stoll *et al.* (2006b) for examples of PLC evaluation materials produced for use in schools).

See also: Assessment and the Evaluation of Institutional Effectiveness; Professional Learning Community.

Bibliography

- Atkins, A. and Thomas, S. (2004). *EPLC Working Paper No. 3: Factor Analysis of Part One of EPLC Questionnaire: The Findings about Professional Learning in the School*. Bristol: University of Bristol Graduate School of Education.
- Bolam, R., McMahon, A., Stoll, L., *et al.* (2005). Creating and sustaining effective professional learning communities. *DfES Research Report RR637*. University of Bristol. <http://www.dfes.gov.uk/research/data/uploadfiles/RR637.pdf> (accessed September 2009).
- Hipp, K., Huffman, J., Pankake, A., and Olivier, D. (2008). Sustaining professional learning communities. *Journal of Educational Change* 9, 173–195.
- Lee, V. E. and Smith, J. B. (1996). Collective responsibility for learning and its effects on gains in achievement for early secondary students. *American Journal of Education* 104, 103–147.
- Louis, K. S. and Marks, H. (1998). Does professional community affect the classroom? Teachers' work and student experience in restructured schools. *American Journal of Education* 106(4), 532–575.
- Louis, K. S. and Miles, M. B. (1990). *Improving the Urban High School: What Works and Why*. New York: Teachers College Press.
- McCall, J., Smith, I., Stoll, L., *et al.* (2001). Views of pupils, parents and teachers: Vital indicators of effectiveness and for improvement. In MacBeath, J. and Mortimore, P. (eds.) *Improving School Effectiveness*. Buckingham: Open University Press.
- Pedder, D., James, M., and MacBeath, J. (2005). How teachers value and practise professional learning. *Research Papers in Education* 20(3), 209–243.
- Scheerens, J., Glas, C., and Thomas, S. (2003). *Educational Evaluation, Assessment and Monitoring: A Systemic Approach*. Tokyo: Swets & Zeitlinger.
- Smith, M. and Thomas, S. (2004). *EPLC Working Paper No. 5: Multilevel modelling of pupil outcomes*. Bristol: University of Bristol Graduate School of Education.
- Stoll, L., Bolam, R., McMahon, A., Wallace, M., and Thomas, S. (2006a). Professional learning communities: A review of the literature. *Journal of Educational Change* 7(4), 221–258.
- Stoll, L., Bolam, R., McMahon, A., *et al.* (2006b). *Professional Learning Communities: Source Materials for School Leaders and Other Leaders of Professional Learning*. London: Innovation Unit, DfES, NCSL and GTC.
- Stoll, L. and Louise, K. S. (eds.) (2007). Professional Learning Communities: Elaborating new approaches. In *Professional Learning Communities*, pp 1–14. Maidenhead: Open University Press.
- Stoll, L., McMahon, A., Bolam, R., *et al.* (2003). *Creating and Sustaining Effective Professional Learning Communities – Questions Arising from the Literature and How They are Being Addressed in the Project*. DfES research Brief no. RBX 12-03.
- Thomas, S. M. (2001). Dimensions of secondary school effectiveness: Comparative analyses across regions. *School Effectiveness and School Improvement* 12(4), 285–322.
- Thomas, S. and Mortimore, P. (1996). Comparison of value added models for secondary school effectiveness. *Research Papers in Education* 11(1), 5–33.
- Thomas, S. M., Peng, W. -J., and Gray, J. (2007). Value added trends in English secondary school performance over ten years. *Oxford Review of Education* 33(3), 261–295.
- Visscher, A. J. and Witziers, B. (2004). Subject departments as professional communities? *British Educational Research Journal* 30(6), 785–800.
- Wiley, S. (2001). Contextual effects of student achievement: School leadership and professional community. *Journal of Educational Change* 2(1), 1–33.

Further Reading

- Stoll, L., McMahon, A., and Thomas, S. (2006). Identifying and leading effective professional learning communities. *Special Issue:*

Leadership and Student Learning in Professional Learning Communities. Journal of School Leadership **16**(5), 611–623.
Thomas, S., Smees, R., MacBeath, J., and Robertson, P. (2000).
Valuing pupils views in Scottish schools. *Educational Research and Evaluation* **6**(4), 281–316.

<http://www.dcsf.gov.uk> – Department for Children, Schools and Families.
<http://www.gtce.org.uk> – General Teaching Council.
<http://www.ncsl.org.uk> – National College for School Leadership.
<http://www.innovation-unit.co.uk/about-us/publications/professional-learning-communities.html> – UK Innovation Unit school source materials and activities to support staff development see Stoll *et al.* (2006b).

Relevant Websites

<http://www.epic.info> – Creating and Sustaining Effective Professional Learning Communities.

Evaluation of Adult Education and Training Programs

G McNamara, P Joyce, and J O'Hara, Dublin City University, Dublin, Republic of Ireland

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Competences – The integration of knowledge, skills, and attitudes concerned primarily with developing the ability to acquire further knowledge, rather than simply the possession of a particular amount of current knowledge.

Evaluation – The systemic assessment of an object's worth, probity, feasibility, safety, significance, and/or equity.

Formative evaluation – Evaluation conducted during the development of a program.

Program evaluation – The use of social research procedures to systematically investigate the effectiveness of social intervention programs such as education and training.

Quality – The ongoing and continuous analysis of the provision and outcomes of programs.

Summative evaluation – Retrospective assessments of completed or established programs.

Introduction

Evaluation occupies an increasingly major place in public and private decision-making processes, serving many political functions such as analysis of spending, allocation of funds, and provision of accountability. On the other hand evaluation is also increasingly concerned with learning, with empowering program participants and enabling social transformation.

Program evaluation is the use of social research procedures to systematically investigate the effectiveness of social intervention programs such as education and training (Stufflebeam and Shinkfield, 2007). At one level the demand for an appropriately skilled workforce in an evolving global economy has made the evaluation of adult education and training programs a high priority but one which seemed relatively straightforward. However ideological, curricular, and political rethinking has fundamentally undermined this seeming simplicity. The conceptual framework underpinning theories of adult education and training has become deeply contested. In terms of curricular structure the traditional focus on behavioral outcomes has been replaced by the broader concept of competences.

Equally, increased demands for value for money in public services have resulted in concerns about accountability and quality moving to the fore. Evaluation has also progressed dramatically and is now no longer simply concerned primarily with measurable outcomes but with process, stakeholder roles, values, and quality.

In this article, the contextual issues to be considered in the evaluation of adult education and training programs are outlined. Key questions which need to be answered whichever approach is undertaken are presented. Evaluation models and approaches are summarized and two particular models are highlighted as popular but contrasting approaches to evaluating adult education and training programs. Finally a greater emphasis on self-evaluation is suggested as an approach worth considering for the twenty-first century in the field of adult education and training.

The Changing Face of Evaluation in Adult Education and Training

Evaluation theory and practice in adult education and training has, in recent times, undergone significant development. This development reflects not only evolving concepts of evaluation but equally dramatic changes in the philosophy and curriculum of adult education and training. For example, education and training in nursing and medicine have undergone significant reform in the past 10–15 years. Continuing with a strong practice-based focus it now also has an explicit focus on problem-based learning, change management, and policy development. The focus has also changed from one of instruction, for example, an apprenticeship model, to a learning paradigm, where the emphasis is on student-centered learning. The role of the adult learner is one of active participant who sees a relevance to learning which can be applied to practice.

The purposes and ideology of much adult education and training have become deeply contested. Traditionally, adult education and training had been seen as primarily concerned with issues such as skills, labor market productivity, and economic development. More recently, post-modernist thinking has moved the focus to issues of social transformation concerned primarily with personal and societal development as key purposes in adult education. Particularly influential have been constructivist theories

reminding us that knowledge, while individually constructed, requires collective and collaborative interaction as a key element of learning. Thus, while for certain purposes objectives-based training in the form of skills development may well be defensible and appropriate, the concept of adult education and training has expanded significantly in line with postmodernist views of epistemology and ontology.

In the light of this changing conceptual focus, the nature and structure of the curriculum in adult education and training has also been reformed. In response to changes in society, and in particular the emergence of the knowledge economy, adult education and training is increasingly concerned with the development of broad competences. Competences integrate knowledge, skills, and attitudes and are concerned primarily with developing the ability to acquire further knowledge rather than being simply about the possession of a particular amount of current knowledge. In consequence, the structure of adult education and training programs has undergone radical change, with an ever greater emphasis on issues of process and less concern with the acquisition of a particular skill set.

Somewhat paradoxically, in tandem with this change in the philosophy and curriculum structure of adult education and training programs, there has been also a parallel increase in policies designed to ensure quality and value for money in the provision of such programs. This is in line with the international neoliberal movement often referred to as new public management which has resulted in greater oversight of programs and initiatives in the public sector and determination to introduce the alleged efficiencies of the free market. Quality is understood to be concerned with the ongoing and continuous analysis of the provision and outcomes of programs (Kells, 1992). It is a broad concept which includes both external requirements for accountability and accreditation such as national standards and benchmarks and internal control of quality within educational institutions and programs.

Evaluation has become dramatically altered as a result of these developments. From its traditional role as a once-off measurement of program outcomes, evaluation is now perceived as an integral part of a continuous cycle of quality assurance which includes program philosophy, curriculum development, definition of quality standards, assessment, strategic planning, and internal and external evaluation. Moreover, since in most cases evaluation, as now constituted, encompasses both accountability and improvement focuses it must be multilevel, capable of responding to different needs and expectations from a variety of audiences.

The above trends have paralleled and are closely linked with changing concepts and practices in the evaluation of adult education and training. Consistent with

developments in other fields there is an increased focus in evaluation on methods which include community or stakeholder input from the beginning of program definition and design. Such methods are designed to help program participants to evaluate themselves and their programs, still with the goal of improving outcomes but also of fostering autonomy and decentralized decision making.

Before the days of competences and quality assurance, the evaluation of adult education and training was primarily concerned with the measurement of traditional behavioral objectives. More recently, evaluation theory and practice has become increasingly defined by a more sophisticated analysis of programs involving the inclusion of stakeholder perceptions and multiple forms of evidence, data, and indicators. There has been a tendency to break away from the classical, objectivist, outcome-based, and performance-orientated concept of evaluation toward a multiplicity of new models. Among these models are responsive evaluation as illumination, ethnographic evaluation, naturalistic evaluation, utilization-focused evaluation, the integrated information development and evaluation model, fourth-generation evaluation, empowerment evaluation, participative evaluation, self-evaluation, and others. In the more recent past, there has been a move from debates between positivists and post-positivists to a dialog between the paradigms leading to a greater emphasis on multimethod approaches.

Therefore, the design of evaluation of adult education and training programs is now more complex than in the past. It has to take into account the changing priorities of the curriculum in such programs, emphasizing key competences at many levels. Evaluation must also be reconceptualized to fit within structures designed not as once-off appraisals but rather continuous cyclical quests for improvement. In design it must reflect these imperatives, and in consequences, educational evaluation theory and practice has moved from simplistic notions of measuring outcomes to more complex concerns with stakeholders' roles and the process of learning.

Designing Evaluation for Adult Education and Training

From the aforementioned discussion it should be clear that the design and conduct of evaluation in the field of adult education and training presents difficulties peculiar to that field. For example, a curriculum based on the complex notions of competences requires standards, indicators, criteria, and appropriate assessment procedures if it is to be coherent. All these features must be evaluated. Since by their nature many competences are in the expressive domain and resistant to traditional notions of measurement, a variety of largely qualitative methodologies will be

applied alongside more traditional modes of testing. If, as likely in the bulk of cases, the evaluation is concerned with both accountability and improvement, and internal and external audiences, the evaluation design must be able to meet all these needs. This may be achievable in theory but recent work in the field by McNamara and O'Hara (2004) suggests that evaluation for accountability and for improvement are not complementary but competing, and that formative goals focused on learning can easily be derailed by the demands for summative judgment. Finally, in the modern discourse of evaluation, the emphasis on negotiation and collaboration and on iterative research processes and methods are often at odds with contractual requirements and realistic resources and timescales.

The increasing complexity of educational evaluation design is illustrated by the definition by Stufflebeam and Shinkfield (2007), which describes evaluation as the systematic assessment of an object's merit, worth, probity, feasibility, safety, significance, and/or equity. This more values-oriented definition is an extension of that put forward by the Joint Committee on Standards for Educational Evaluation (1994) in the USA which defined evaluation as the systematic assessment of the worth or merit of an object. The merit of a program can be judged by examining if it does well what it is supposed to do. Worth refers to a program's combination of excellence and service in an area of clear need within a specified context. In evaluating probity, assessments are made of the program's honesty, integrity, and ethical behavior (Stufflebeam and Shinkfield, 2007). A good evaluation should provide direction for efficient use of time and resources and be politically viable. Many programs require evaluations that examine safety of facilities, influence, importance, and visibility. Evaluators need to consider the possibility that a program can have far-reaching implications outside the local arena and evaluate its significance over time and in other settings. The criterion of equity argues for equal opportunities for all people and emphasizes freedom for all. It is also concerned with the roles and rights of all legitimate stakeholders including politics, ideology, and imbalances in power relationships. Taking this array of concerns into account in evaluation design is clearly a formidable task.

In the following section we look at two contrasting models of adult education and training evaluation. Kirkpatrick's four-level model is very much in the Tylerian behavioral objectives measurement tradition and is still widely utilized in the evaluation of skills-training programs (Thackwray, 1998). In contrast Jacobs' (2000) approach is representative of models concerned with stakeholder empowerment, constructivist concepts of knowledge, and social transformation. We begin by explaining the categorization of evaluation into formative and summative and propose key questions which need to be answered whichever approach is undertaken. Finally, the increasing emphasis on self-evaluation is mentioned and it is suggested

that for both conceptual and practical reasons it will become increasingly influential in the field of adult education and training.

Evaluation Models

Designing an evaluation of adult education and training programs involves a number of key steps. Such a structure is important so as not to omit any vital parts of the process. In addition the timing of evaluation is very important and this should be decided in advance of setting up any adult education and training programs.

Formative and Summative Evaluations

The timing of evaluations has been generally differentiated into summative evaluations (retrospective assessments of completed or established programs) and formative evaluations (conducted during the development of a program). Formative evaluations are used to modify and improve a program and this is frequently used to provide feedback to staff while the program is in operation. These evaluations assess and assist with the formation of goals and priorities, provide direction for planning, and guide program management. Information from formative evaluations is directed to improving operations and serves quality assurance purposes. In contrast, summative evaluations are used to prove something, satisfy accountability, or make a judgment about the overall quality of the program. They draw together previously collected information, for example, from formative evaluations. Both formative and summative evaluations are needed in the development of a program (Rossi *et al.*, 1999). In general, formative evaluation will be dominant in the early stages of a program and summative will take over as the program concludes.

Key Questions

Regardless of model or approach chosen, a number of key questions need to be answered when organizing a program evaluation (Thackwray, 1998):

- How is evaluation defined?
- What are the functions of the evaluation?
- What are the objects of the evaluation?
- What kinds of information should be collected regarding each object?
- What criteria should be used to judge the merit of an evaluation object?
- Whose interests should be served by the evaluation?
- What methods of enquiry should be used?
- Who should do the evaluation?
- By what standards should the evaluation be judged?
- How and when should the results be presented?

The process of considering these issues and integrating them into evaluative research design is very well exemplified in two evaluation models which have become hugely influential, namely case study evaluation as developed by Robert Stake and utilization-focused evaluation created by Michael Quinn Patton. An interesting range of examples of evaluation practice in complex learning organizations which respond to the above questions is provided in a recent publication of the American Evaluation Association (Braverman *et al.*, 2008).

Kirkpatrick's Model

Introduced over four decades ago, this model is still used by many for evaluating training and development programs and can be categorized as a results or goal-based evaluation (Figure 1). Initially a four-level model, it was later adapted to include a fifth level to measure return on investment. Each level measures different but complementary aspects of training and development. In essence, Kirkpatrick sought to stimulate those with responsibility for the management of training and development to increase their efforts in evaluating training and development actions. Critics of Kirkpatrick assert that his evaluation process may not always produce genuinely meaningful, long-term results. This model implies that evaluation is a standardized, prepackaged process, which is clearly not always the case.

Reaction or level one is easy to measure and quantify and is interpreted as determining how people feel about the program. Criticisms of this level state that it does not measure any learning which takes place so that moving to level two is important. Recommendations for level two include the use of a before-and-after approach so that learning can be related to the program. Evaluation of behavior or level three of this model is more difficult. It may be possible to appraise performance before and after the program or to have a posttraining appraisal 3 months or more after the program so that participants have an opportunity to put into practice what they learned. The fourth or results level is the most difficult area to evaluate effectively. This level defines results to include an institution's capacity to learn, change, and develop in line

with its stated agreed objectives. MacDonald *et al.* (2000) used Kirkpatrick's model to examine the impact of applying adult education principles to training in advanced technology companies. The evaluation explored levels one to four and data was collected by observation notes, post module surveys, post assignment surveys, and the participants' summative evaluation of the program. The authors suggest that the assignment data and summative evaluations confirm a level four on Kirkpatrick's model was reached on this evaluation.

Jacobs' Model

However, for many evaluators today, the process as well as the outcomes of modern evaluation has to be meaningful and compelling for the stakeholders if they are to have any useful impact on daily activities in organizations (Figure 2). The epistemic changes in both adult education and training and evaluation theory and practice outlined earlier in this article demand an operational model that permits built-in evaluations with continuous streams of information for both descriptive and evaluative uses. These changes also demand a methodology of evaluation that covers both the technical and social aspects and is comprehensive enough to include the historical, the political, the analytical, and the naturalistic aspects. Such an evaluation will be supported by statistics and stories, quantities and qualities, careful description, and impressions and anecdotes. The current literature on evaluation has many examples of such models. The following based on the work of Jacobs is representative.

Jacobs outlines a ten-stage evaluation model. The opening three stages are concerned with collaborative and negotiated understandings around the context and policy framework of the intervention, the goals of the evaluation, and the identification of and consultation with the principal stakeholders and relevant constituencies. Stages four through seven again involve iterative negotiated agreement on evaluation purposes and criteria, the interests to be served by the evaluation, the aspects of the program to be evaluated, the sources of information to be used, and the evaluation methods most appropriate. Stage eight involves the collection of data from the agreed sources and stage nine involves the analysis and interpretation of the data in the context of the agreed, negotiated criteria. Stage ten involves the dissemination of findings to the stakeholders and constituencies identified at the beginning of the process.

Jacobs stresses that evaluators should employ this negotiated and iterative process, regardless of where the call and resources for the evaluation have come from. She argues that it is central to the appropriate role of the evaluator to recognize the interests and rights of all legitimate stakeholders and to take into account power differentials between groups. Jacobs also advises that

Level 1 Reaction:	What the participants felt about the project or program – the happy sheet?
Level 2 Learning:	Internal validation – where the objectives met?
Level 3 Behavior:	External validation – has training transfer taken place?
Level 4 Results:	Has the project/program made a difference? That is, what has been the impact on the institution?

Figure 1 Kirkpatrick's four-level model of evaluation.

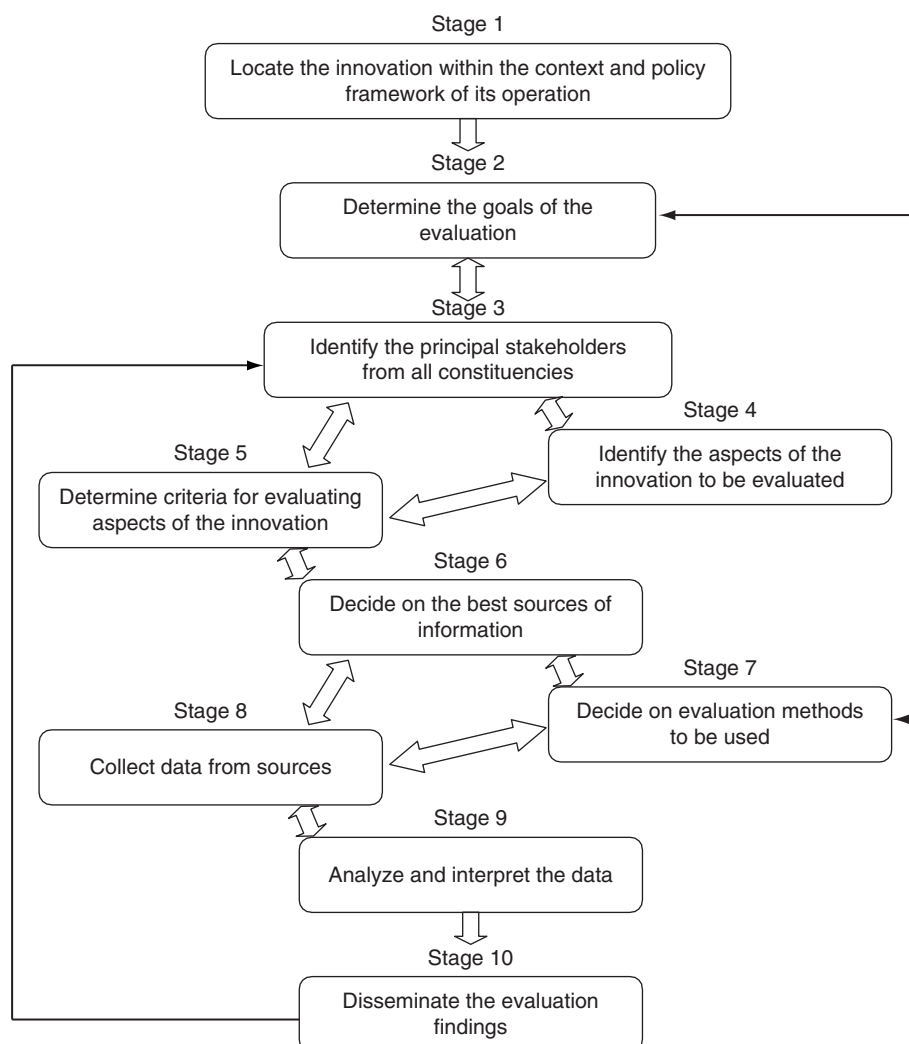


Figure 2 Stages in the process of evaluating educational innovation. Adapted from Jacobs, C. (2000). The evaluation of educational innovation. *Evaluation* 6(3), 261–280, with permission from Sage.

evaluators should give due weight to formative, summative, and illuminative goals and draw from the widest possible range of evidence.

Process-centered models of evaluation such as Jacobs' are widely used in areas of adult and vocational education where the emphasis is on personal and professional development rather than skills acquisition. A good example of collaborative and iterative evaluation in this vein can be found in the work of Bhola (1998).

Self-Evaluation

Despite the developments in evaluation theory and practice outlined so far, many of these models, including more progressive recent ones such as Jacobs', represent a form of evaluation which involves judgments made through the eyes of the external evaluator and the connotation persists of evaluation as an external monitoring of professional practice. The professionalization of evaluation,

the dominance of the contract and terms of reference, and the increasing use of consultants with little knowledge of the field in which they are trying to apply generic research methods are all likely to contribute to the legitimization of market-driven innovations which deskill and disenfranchise practitioners. Educational evaluation is not therefore an objective, external, value-free process, but rather is deeply influential in shaping educational philosophy and policy. The conceptual and ethical stance it adopts is influencing the educational debates to a significant degree.

One particular approach the present authors suggest is to move the focus away from external evaluation to one focused on empowering practitioners to self-evaluate. In the case of adult education and training programs the focus of judgment can move from the evaluators to the practitioners and the former can find a new role in supporting the professional development of the latter. Inviting educators to become the key evaluators of educational innovation as opposed to measuring the outcomes in

some external or objective way is of course controversial. For example, it can be argued that educators cannot be objective evaluators as at one level it is their work and effectiveness that is being evaluated. On this account evaluation must be primarily external. However, herein lies the dilemma not just of evaluation but also of perceptions of teaching and learning. On the one hand, there is increasing pressure to reduce teaching to merely implementing a proven or tested program of instruction. On the other hand, the literature of educational improvement has come increasingly to emphasize that the quality of student learning has to be seen in relation to the quality of teachers' learning.

This view of educators' professional learning emphasizes that the quality of teaching is closely bound up with the capacity of teachers to make professional research-based judgments on their own practice and on the programs and methodologies they are being required to implement. This approach is coming to fruition in initiatives such as peer review and peer observation of teaching, best-practice research scholarships, action research, reflective practice, practitioner-led research, and institution-wide development planning. However, for this to truly happen, the values and methodologies of self-evaluation must be inculcated into specific education and training programs.

Conclusion

Learning from and about evaluation often requires us to change our mental models – to rethink our assumptions and beliefs and to develop new understandings about our programs and evaluation processes. This logically should lead on to an organizational learning approach to evaluation. Such an approach to evaluation would be context-sensitive, ongoing, support dialog, reflection, and decision making at department and organization-wide levels, and contain strong commitments to self-evaluation and practitioner empowerment.

Effective evaluation is a significant contributor to quality but does not necessarily guarantee that those in authority will heed the outcomes of evaluation and take necessary corrective action. According to Kells (1992) institutions and programs can be strengthened substantially through effective evaluation and the basis for choices about the future can be soundly established by a combination of internal self-assessment and unbiased, informed peer review. If adult education and training programs are to gain from evaluation procedures and processes then the proper conditions for securing these improvements must be established. A vision of evaluation for the twenty-first century may be one which is made honest, accurate, and useful by engaging in a partnership with practitioners, people, and programs. Evaluators will be held to a higher standard and will be expected to do good through evaluation. In fact evaluation

must move from just generating findings about specific programs to generating knowledge.

To an increasing extent the evaluation model chosen is influenced by the evaluators' own philosophy about evaluation, although other factors such as time, resources, expertise, and availability of staff also strongly influence procedures used. Most program evaluation experts agree that there is no one best model. It is necessary, therefore, for the program evaluator to select a model which matches the requirements of a situation to produce evaluation findings which are most likely to accurately appraise a program's merits, worth, probity, feasibility, safety, significance, and equity.

See also: Cost Analysis in Evaluation Studies; Curriculum Evaluation; Evaluating Education in Three Policy Eras; Program Evaluation.

Bibliography

- Bhola, H. S. (1998). Constructivist capacity building and systemic evaluation of adult basic education and training in South Africa: One discourse, two scripts. *Evaluation* 4(3), 329–350.
- Braverman, M., Engle, G., Arnold, M. E., and Rennekamp, R. A. (eds.) (2008). *Program Evaluation in a Complex Organisational System: Lessons from Cooperative Extension*. San Francisco, CA: Jossey-Bass/American Evaluation Association.
- Jacobs, C. (2000). The evaluation of educational innovation. *Evaluation* 6(3), 261–280.
- Joint Committee on Standards for Educational Evaluation (1994). *The Program Evaluation Standards*. Thousand Oaks, CA: Corwin Press.
- Kells, H. R. (1992). *Self-Regulation in Higher Education. A Multi-National Perspective on Collaborative Systems of Quality Assurance and Control*. London: Jessica Kingsley.
- MacDonald, C. J., Gabriel, M. A., and Cousins, J. B. (2000). Factors influencing adult learning in technology based firms. *Journal of Management Development* 19(3), 220–240.
- McNamara, G. and O'Hara, J. (2004). Trusting the teacher: Evaluating educational innovation. *Evaluation* 10(4), 467–478.
- Rossi, P. H., Freeman, H. E., and Lipsey, M. W. (1999). *Evaluation: A Systematic Approach*, 6th edn. Thousand Oaks, CA: Sage.
- Stufflebeam, D. L. and Shinkfield, A. J. (2007). *Evaluation Theory, Models, and Application*. San Francisco, CA: Wiley.
- Thackwray, B. (1998). *Effective Evaluation of Training and Development in Higher Education*. London: Kogan Page.

Further Reading

- Bezzi, C. (2006). Evaluation pragmatics. *Evaluation* 12(1), 56–76.
- Fetterman, D. M. (2001). The transformation of evaluation into a collaboration: A vision of evaluation in the 21st century. *American Journal of Evaluation* 22(3), 381–385.
- Greenwood, D. J. and Levin, M. (2007). *Introduction to Action Research: Social Research for Social Change*, 2nd edn. London: Sage.
- Hansen, H. F. (2005). Choosing evaluation models: A discussion on evaluation design. *Evaluation* 11(4), 447–462.
- Haug, P. and Schwandt, T. A. (eds.) (2003). *Evaluating Educational Reforms. Scandinavian Perspectives*. Greenwich, CT: Information Age/Corwin Press.
- House, E. R. and Howe, K. R. (2003). Deliberative democratic evaluation. In Kellaghan, T. and Stufflebeam, D. L. (eds.) *International Handbook of Educational Evaluation*, pp 79–102. Norwell, MA: Kluwer.

- Kushner, S. (2000). *Personalising Evaluation*. London: Sage.
- Kushner, S. and Norris, N. (2007). *Advances in Program Evaluation: Volume 10. Dilemmas of Engagement: Evaluation and the New Public Management*. London: Elsevier Science.
- Marton, F. (1988). Phenomenography: A research approach to investigating different understandings of reality. In Sherman, R. and Webb, R. (eds.) *Qualitative Research in Education: Focus and Methods*, pp 143–161. New York: Falmer.
- McNamara, G. and O'Hara, J. (2008). *Trusting Schools and Teachers: Developing Educational Professionalism through Self-Evaluation*. New York: Peter Lang.
- Micari, M., Light, G., Calkins, S., and Streitwieser, B. (2007). Assessment beyond. *Evaluation* **28**(4), 458–476.
- Patton, M. Q. (1997). *Utilization-Focused Evaluation: The New Century Text*. Thousand Oaks, CA: Sage.
- Patton, M. Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *American Journal of Evaluation* **22**(3), 329–336.
- Preskill, H. (2008). Evaluation's second act. A spotlight on learning. *American Journal of Evaluation* **29**(2), 127–138.
- Shackman, G. (2008). What is program evaluation? A beginners guide. <http://gsociology.icaap.org/methods/evaluationbeginnersguide.pdf> (accessed August 2009).
- Simons, H. (2004). Utilizing evaluation evidence to enhance professional practice. *Evaluation* **10**(4), 410–429.

- Stake, R. E. (2004). *Standards Based and Responsive Evaluation*. Thousand Oaks, CA: Sage.
- Torres, R. T. and Preskill, H. (2001). Evaluation and organizational learning: Past, present, and future. *American Journal of Evaluation* **22**(3), 387–395.

Relevant Websites

- <http://www.cgu.edu> – Claremont Graduate University, Institute of Organizational and Program Evaluation Research.
- <http://www.eaea.org> – European Association for Education of Adults.
- <http://www.europeanevaluation.org> – European Evaluation Society.
- <http://www.iveta.org> – International Vocational Education and Training Association.
- <http://www.cse.ucla.edu> – National Centre for Research on Evaluation, Standards and Student Testing.
- <http://www.nzqa.govt.nz> – New Zealand Qualifications Authority (NZQA), National Qualifications Framework.
- <http://reflective.bazaar.org> – The Bazaar, Reflective Evaluation.
- <http://www.edpubs.ed.gov> – US Department of Education Publications.
- <http://www.voced.edu.au> – VOCED, an Australian Government Funded Operation.

Evaluation of Integrated Health Programs in School

T M Franke, University of California at Los Angeles, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

School-based health clinics provide an essential resource for children and families around the world. From a historical perspective they have clearly evolved in form and function over the past 100 years. One needs, only looking at the current outbreak of swine flu in the United States and throughout the world, to see an example of the role that school-based health clinics can play in identification, referral, and the dissemination of current information. This article focuses primarily on the development of school-based health clinics in the United States and the role evaluation has played in demonstrating the efficacy of integrated health programs in schools. In examining the literature around the evaluation of school-based health centers, several prominent themes appear. These themes include the evaluation of school-based health centers from the standpoint of utilization, mental health access/services, reproductive health, and clinic effectiveness/costs.

The History of School-Based Health Clinics in the United States

The beginnings of the American school-based health system can be traced back to the Progressive Era. From the mid-1850s through the early 1900s, industrialization, urbanization, and immigration were permanently changing city living. As more people moved to the cities, living and working conditions became increasingly worse, creating perfect conditions for the spread of disease and epidemics. These conditions were instrumental in bringing some level of healthcare to schools, where epidemics could easily be spread. Funding was provided to prevent such an occurrence. As was evident in the 1872 smallpox epidemic in Elmira, New York, which began the vaccination of the school children still practiced today (Dryfoos, 1994; Lear, 1996).

Disease would continue to open doors for school healthcare. For the next 20 years, epidemics swept through Boston and its schools. Boston's response was to start city-wide inspections, especially screening children at schools, for disease that spread quickly to other cities. By 1911, 300 cities required medical inspection; nine states had mandatory school health inspection laws while ten states permitted local agencies to hire school health inspectors (Kort 1984; Dryfoos, 1994).

The end of the Progressive Era (1890s–1920s) brought critical changes for school-based healthcare. The first was

a distinction between the medical treatment and preventive health services. Preventive health services would offer screening and health education, while medical treatment is the actual practice, diagnosis, and treatment of the disease. Free medical treatment in schools was perceived as taking away from the private practice of medicine. Some labeled it as a form of socialism and saw it as infringing on the rights of the family and home. One of the leading voices against healthcare in schools came from the American Medical Association that used its political power to discourage the medical community and society from supporting any service that could threaten the private fee-for-service medicine (Dryfoos, 1994; Lear, 1996). This effectively put a stop to the fledgling movement of school-based healthcare for many years.

School health policy would make drastic shifts between the 1920s and 1970s. By the end of the 1920s, the idea that the school was not meant to maintain health but rather just educate had emerged. Schools simply became referral points for private practice physicians. By the 1930s, educational psychology began to take over school healthcare. The focus in school-based healthcare shifted from treatment, case finding, and medical care to prevention, universal intervention, and health education (Dryfoos, 1994) and remained the primary focus of school-based health centers until the 1960s. With President's Johnson's War on Poverty, policy began to shift the emphasis back to case finding, and medical treatment as a result of new services such as (1) Medicaid, (2) Medicare, (3) Head Start, (4) Early Periodic Screening, Diagnosis, and Treatment (EPSDT), and (5) the Elementary and Secondary Education Act (ESEA). These led to later policies such as (6) the Education for the Handicapped Children Act of 1975 (subsequently renamed the Individuals with Disabilities Education Act).

1. *Medicaid* – Title XIX of the Social Security Act is a Federal/State entitlement program that pays for medical assistance for certain individuals and families with low incomes and resources. This program became law in 1965 as a cooperative venture jointly funded by the Federal and State governments to assist States in furnishing medical assistance to eligible needy persons.
2. *Medicare* – health insurance program administered by the United States government, providing health insurance coverage to people aged 65 or older, people under age 65 with certain disabilities, and people of all ages with End-Stage Renal Disease (permanent kidney failure requiring dialysis or a kidney transplant).

3. *Head Start* – a national program that promotes school readiness by enhancing the social and cognitive development of children through the provision of educational, health, nutritional, social and other services to enrolled children and families.
4. *Early Periodic Screening, Diagnosis, and Treatment (EPSDT)* – EPSDT is the child health component of Medicaid. It is required in every state and is designed to improve the health of low-income children, by financing appropriate and necessary pediatric services.
5. The Elementary and Secondary School Act (ESEA) – It is a United States federal statute enacted on 11 April 1965, which funds primary and secondary education such as professional development, instructional materials, resources to support educational programs and parental involvement programs. Recent reauthorizations of the Act include: Education Consolidation and Improvement Act of 1981, Improving America's Schools Act of 1994, No Child Left Behind Act of 2001.
6. The Education for All Handicapped Children Act (sometimes referred to using the acronyms EAHCA or EHA, or Public Law (PL) 94-142) was enacted by the United States Congress in 1975. This act required that all public schools accepting federal funds to provide equal access to education for children with physical and mental disabilities. Public schools were required to evaluate handicapped students and create an educational plan with parent input that would emulate as closely as possible the educational experience of non-disabled students.

Shortly after, new models for school-based clinics and services began to emerge. These school-based health clinics (SBHCs) were programs created under medical supervision that typically operated out of a health department hospital or community health center. The first three models were prototypes for the school-based health clinics to come. In Cambridge, Massachusetts, the Cambridge program featured primary care medical clinics within elementary schools staffed by pediatric nurse practitioners. The Hartford project made dental and primary care services available to all students in one urban and one suburban school in Hartford, Connecticut. In Galveston, Texas, pediatrician Philip Nader linked Galveston School District health services to those offered at University of Texas Medical Branch (Lear, 1996). The Posen-Robbins program pioneered setting up medical clinics in the very disadvantaged neighborhood of Cook County, Illinois with a team of two full-time school nurse practitioners, four health aides, a lab technician, and outreach workers along with medical backup offered by two neighboring hospital physicians. SBHCs would learn from these first schools and start to proliferate quickly (Dryfoos, 1994).

As many new SBHCs appeared, research was needed to evaluate the efforts; however, this proved to be a difficult

endeavor. The main problem was that the children and families who benefited from the programs were moving targets. A significant proportion of children and families did not stay in the same schools for very long – a challenge that still exists. (The United States has the highest rate of residential mobility among industrialized countries. According to the US Census Bureau (2000), every year nearly 1 in 6 persons (17%) in the US relocate or change their place of residence.)

Although school mobility has the potential to affect school performance, social and emotional performance, and peer relationships (Mehana and Reynolds, 2004; Nelson *et al.*, 1996), it is especially problematic when attempting to perform evaluation of school-based health services. The cost of evaluation could end up equaling or exceeding the cost the service provided (Dryfoos, 1994). Despite this, by the early 1990s, many programs had appeared throughout the USA and substantial evaluation of school-based health clinics was being conducted.

Finances have emerged as another core issue for SBHCs. Government healthcare programs operate through managed care delivery systems, which makes receiving reimbursements for out-of-network service very difficult. Nevertheless, school-based health clinics have adapted and evolved. The 2004–05 National Assembly on School-Based Health Care census identified 1235 school-based health centers, 805 offering mental health services and 153 offering dental health services (NASBHC, 2005).

Evaluations of School-Based Health Centers

Overall, more discourse exists on the positive findings associated with SBHCs. The focus of the evaluations ranges from utilization to program and cost effectiveness and typically addresses health broadly. The following sections will examine several prominent themes in the evaluation of school-based health centers.

Utilization

Numerous evaluations have been implemented to compare students' use of medical, mental health, and substance abuse services located in school-based centers with traditional locations. The School Health Demonstration Program (SHDP) in New York provided low-income families with children in preschool through 12th grade with primary care. This program was created with the help of the New York State Department of Health, Education, and Social Services and comprised a team consisting of a nurse practitioner, health aide, and a pediatrician or family practice physician acting as supervisor and was linked to a local health facility that provided medical, dental, and mental health counseling services.

It allowed parents to choose level of care on the enrolment forms and provided health education curricula, which focused on family life education, dental health, summer safety, etc.

The evaluation examined data from nine states in which the SHDP was first implemented. Approximately 36 000 children from 55 public, parochial, and private schools were invited to participate in the program; with 22 689 finally enrolled. The evaluation found that parents chose levels of care greater than the regular school health services, with 53% selecting comprehensive physicals for complete care (NYSDoH, 1983). Overall, the SHDP had 83 602 encounters with children during the school term, including 38 514 screenings, 22 373 primary care services, 15 055 first aid visits and responses to minor complaints, 9121 follow-up visits, 7212 comprehensive physical exams, and 2119 immunizations. In addition to an increase in healthcare utilization, the NYSDoH (1983) evaluation found that principals and teachers reported support for the program, appreciated having it as a quick-response resource for disease and unintentional injuries and felt the program had positive impacts on attendance and learning.

Another study by Anglin *et al.* (1996) focused on 3818 adolescents and three high school-based student health centers using 4 years of archived data to find the frequencies of student use of the SBHC. Adolescents who used the SBHC had higher rates of visit frequency than those that used traditional medical care. The evaluation found that 94% of the students used the clinic for medical providers; 25% for mental counseling; and 8% for substance abuse counseling. The visit frequency increased significantly for students who used two and three categories of service; going up from four to five visits for a single service type to 13–15 for two categories of providers and 32 total visits if all three services were used. Overall, SBHCs increased adolescents' access to medical, mental health, and substance abuse services.

Kaplan *et al.* (1999) looked at the effects of SBHC access and utilization of physical and mental health services by children between 4 and 13 years old. The study compared a school with an SBHC and a comparison school in Denver, CO both having 93% Hispanic students with over 85% being on free-lunch programs. The study was a retrospective cohort analysis of parent surveys from both schools with return rates being 570/728 at the intervention school and 440/571 at the comparison school (Kaplan *et al.*, 1999). The results showed that families utilizing the SBHC reported less difficulty receiving physical healthcare for their children compared to the comparison group; access to the SBHC was independently and significantly related to lower emergency department use and greater likelihood of a yearly physician's visit and annual dental examination. Respondents using the SBHC as their primary health service reported significantly more

satisfaction with the services offered than respondents who mainly used hospital and community clinics. Traditionally underserved minority children, with access to SBHCs, receive significantly better healthcare than those without SBHC access, independent of insurance status and confounding variables (Kaplan *et al.*, 1999).

Mental Health Access/Services

By the mid-1990s, there was mounting evidence that traditional mental health services were not available and not adequate to meet the needs of school age children (Weist *et al.*, 1996). Pastore *et al.* (1998) examined students with mental health problems who had access to SBHCs. A questionnaire was administered on health center use and mental health concerns in an urban high school with a 2-year-old SBHC in New York, NY; 630 responses were received from both users and nonusers of the SBHC (Pastore *et al.*, 1998). Three significant findings emerged. First, average users, frequent users, and nonusers did not differ in mental health problems measured. Second, students who used the SBHC's mental health services were strongly satisfied with the care they received (92%). Third, the study found that respondents had a variety of reasons for nonuse: 60% already had a physician, 50% said they did not need it, and 45% preferred continuing previous healthcare. While the study has some limitations, it demonstrated satisfaction and utilization of mental health services available through an SBHC.

Research continued to look at mental health programs and the ability to bridge the gap between service need and service utilization at the school level. Armbruster *et al.* (1997) looked at the central clinic collaboration with the New Haven Public Schools to establish health services for thirteen inner city schools in 1992. The purpose was to compare the characteristics, both clinical and socio-demographic, of the children in the school-based clinics to those in the central clinic. A similar follow-up study by Armbruster and Lichtman (1999) looked at efforts to bridge the gap between service need and service utilization at 36 inner city schools with a sample of 220 clinic cases and 256 school cases of children and adolescents 5–18 years olds. The study found that children from both groups had similar Children's Global Assessment Scale (C-GAS) and Global Assessment of Functioning Scale (GAF) improvements even though the school case children were seen for a shorter period of time (5 months vs. 8 months for the clinic case children), with both samples having an equally frequent level of service of three sessions per month in each setting. The study also notes that the school-based mental health program was also more cost-effective for three reasons: (1) no attrition as children were seen in school; (2) due to program effectiveness, grant funding from public and private sources has increased; (3) school-based services

were part of the child's psychiatric outpatient clinic and were easier to reimburse.

Reproductive Health

One of the very early programs was the Health Start program; a nonprofit organization based in St. Paul, Minnesota, which beginning in 1973 offered comprehensive, multidisciplinary healthcare to adolescents, which included laboratory tests, including pregnancy tests, onsite. The main purpose of the program was to prevent unwanted pregnancies and provide prenatal care to female students in seventh through twelfth grades (Edwards *et al.*, 1980). From its inception through the 1978–79 school year, the clinic serviced 403 students, 85 of which were given prenatal care and subsequently delivered at the St. Paul-Ramsey Medical Center.

Data on all 403 students who received family planning services and/or prenatal care at the three Health Start school clinics were reviewed. The researchers also calculated contraceptive usage and the fertility rates for the female population for each year the clinic was open. The evaluation had limited findings; however, it did show that by the 1977–78 school year, 25% of the schools' female students were receiving family planning services through the Health Start clinics, and 87% of those receiving family planning services were still using contraception three years later (Edwards *et al.*, 1980). The Health Start clinics later evolved into general adolescent health clinics in five St. Paul high schools.

Another early program, the Self Center Program, was a project that dealt with pregnancy and associated issues for seventh through 12th grade students. It tried to raise students' knowledge on issues such as reproductive biology, pregnancy, and sexual activity. The goals were to postpone the onset of intercourse, increase clinic attendance and contraception use, while lowering the risk of pregnancy. The program was a collaboration between The John Hopkins University School of Medicine and the Baltimore City Departments of Education and Health and included both mental health services and clinical services along with the health education. The program provided classroom presentations, education and counseling services within the school, and education, counseling and medical services provided after school in a storefront clinic. The staff consisted of a registrar, nurse's aide or licensed practical nurse, a physician (on some days) and two teams of a social worker and a pediatric nurse practitioner or nurse-midwife.

Extensive evaluation conducted from 1981 to 1984 as part of a study was focusing on African-American students in both junior and senior high schools living in public housing in inner-city Baltimore (Zabin *et al.*, 1986). The program groups consisted of 667 males and 1033 females from one junior and one senior high school, and the control

group included 944 male and 1002 female students from two other junior and senior high schools. The program and control groups completed pre- and post-program questionnaires. The program groups completed follow-up questionnaires for three subsequent years. The researchers found that after 16 months of exposure to the program, the pregnancies rose by 13% among program females, while rising 50% among controls; after 20 months, pregnancy rates fell 22.5% among program females, while rising 39.5 among controls. The trend continued in the same direction, as after 30 months, pregnancy rates fell 30.1% among program females, while increasing 57.6% among controls. The post-program analyses revealed that less than 20% of the program females had participated in unprotected intercourse versus the 44–49% of the control group.

Kirby *et al.* (1991) focused on the reproductive health programs of six school-based clinics that were only part of a comprehensive health program and differed in emphases on reproductive health, sexuality education, and family planning. However, each clinic employed at least one physician and nurse practitioner within the school. The main purpose of the program was to influence sexual behavior and contraceptive usage among ninth through twelfth graders. The schools represented various socioeconomic populations, both rural and urban; however, the majority of the students were low-income families with large proportions of African Americans and other minorities with limited access to healthcare.

The Kirby *et al.* (1991) study had somewhat limited findings overall. In one of the three sites that provided contraceptives in the clinic, students were significantly more likely than comparison school students to use birth control during last intercourse. At the same time, Kirby *et al.* (1991) found that one clinic school with a strong AIDS education program observed a sharp increase in condom use; one clinic school with a strong pregnancy prevention component observed a higher use of condoms and birth control pills relative to its comparison school. However, the study could not show a difference in pregnancy rates between schools with clinics and schools without clinics. The study also found that students in schools with clinics did not engage in sexual activity earlier than students at schools without clinics.

Research throughout the 1990s and 2000s continued to look at SBHCs impact on fertility rates. Ricketts and Guernsey (2006) examined the decline in fertility rates among African-American adolescents in Denver between 1991 and 1997 and its possible association with the Denver SBHCs. The team compared the changing fertility rates among areas with and without SBHCs by using geocoded birth certificates and school enrollment data. The African-American fertility rate declined significantly in areas with SBHCs, enough to suggest that SBHCs were effective in meeting the needs of students at risk of pregnancy (Ricketts and Guernsey, 2006). The African-American fertility rate

in SBHC areas dropped from 165/1000 in 1992 to 38/1000 in 1997. The team found that the rate of decline in areas with SBHCs was significantly greater than areas without SBHCs, 77% and 56%, respectively. The decline is credited to SBHCs being able to properly identify, intervene, and follow-up on high-risk students (Ricketts and Guernsey, 2006).

Clinic Effectiveness/Costs

One of the central concerns with school-based health clinics in the United States is whether or not there is a real benefit to the children, and at what financial cost. Research has shown that SBHCs offer both educational benefit and medical benefit. The Gillespie Student Health Project in Greensboro, North Carolina, is an example of a program that showed educational benefits to students from its evaluation. The program involved the Guilford County Department of Public Health, the Moses Cone Hospital, Greensboro Public Schools and the Robert Wood Johnson Foundation in providing an Alternative Education Program (AEP) which taught health education, health promotion, and screening. The clinic served low-income, at-risk students, over half living in single-parent female-headed households (Klein and McCord, 1992).

The researchers sought to determine associations between clinic use and positive scholastic outcomes, lowered rates of absences, suspensions, dropping out and increased rates of promotion and graduation; data were taken from archival student records, school attendance records, and clinic use records on all AEP students ($n = 322$) during the 1990–91 school year. Students who used the clinic had a lower rate of absence than students who registered but never used the clinic and those who did not register. They also found that students who were registered for the clinic were more likely to stay in school as opposed to the students who did not register (44% vs. 29%), and that registered students were also more likely to graduate and/or more likely to be promoted than the nonregistered students (Klein and McCord, 1992).

The School-Based Adolescent Health Care Program (one of the bigger programs) offered evidence of medical benefits to the students in the SBHCs. Started in 1986 by the Robert Wood Johnson Foundation, the program awarded grants of up to \$600 000 to public and private institutions to set up adolescent health centers (Brodeur, 2000). Twenty-four sites were established across the nation with the goal to deliver primary healthcare services to high school adolescents (Lear *et al.*, 1991). The program offered clinical, health education, counseling and mental health services as well as community coordination.

The evaluation focused mainly on students from 23 clinics in ethnically diverse high schools in 14 cities and 11 states (Lear *et al.*, 1991). The grantees sent quarterly

reports to the Robert Wood Johnson Foundation for analysis, which looked for trends, characteristics, and other important findings over the 3 years of reports. Lear *et al.* (1991) found that of the 34 106 students that had been enrolled in the 23 clinic schools, 46% of the students were clinic users; each of the clinics had, on average, 685 users. Acute illness or injury made up 26% of the total visits; it was the leading service provided by the clinic. Mental health-related care was second in frequency, making up 21% of the clinics visits while reproductive health made up only 12% of the total visits. A trend was also found in terms of parental consent to school-based healthcare, with a steady rise from 34% in 1987 to 71% in 1989–90.

Kisker and Brown (1996) independently evaluated the School-Based Adolescent Health Care Program for whether or not SBHCs offered improvements in access to health care, health status, and risk-taking behavior. The study compared a cohort of students attending 19 participating schools with a national sample of urban youths, using logit models to control for observed differences in the groups. The team compared health center utilization, use of other healthcare providers, knowledge of key health facts, substance use, contraceptive use, sexual activity, pregnancy and birth rates, and health status. They found small, nonstatistically significant impacts on health status and risky behavior although health centers did increase students' access to healthcare and improved their knowledge about health. The team recommended more intensive or completely different services in order to produce significant reduction in risk-taking behaviors.

Other programs, such as the Middletown Adolescent Health Project (MAHP), demonstrate that SBHCs have financial outcomes that are, in fact, cost effective. The project was launched with the cooperation of the local school district, with clinic staff from the Delaware Division of Public Health and intended for ninth through twelfth grade students in rural Middletown, Delaware. A regular school nurse served as a liaison, referring students to the clinic. The main purpose was to provide school-based, comprehensive health services such as treatment of minor and acute illnesses, sports and routine physical examination, laboratory tests, screenings, medical social services, and a number of other services.

The evaluation of MAHP was conducted from 1985 to 1986, was rather simple, consisting of assessing and comparing project cost of the program with the costs of private health available and surveying users' feelings about the project (Siegel and Kriebel, 1987). All the students in the project received the survey, which was distributed and collected over a 3-month period by the evaluation team. Evaluation concluded that the school clinic was able to provide services to many students by being geographically accessible, whereas 68% of the students had a regular doctor outside of their rural community (Siegel and Kriebel, 1987). The clinic was also able to help those who

were economically disadvantaged, from families with one or no wage earners (58% of enrollees). In addition, the clinic also had fewer than 6% of scheduled appointments cancelled due to student no shows. The clinic also proved to be cost-effective to the parents; the treatment costs of MAHP were lower than the costs of the same treatment or services from private physicians. By the end of the first school year of operation, half of the 658 students had used the services.

The effectiveness of SBHCs has also been studied with regard to sexually-transmitted disease (STD) screening and treatment. A study performed by Cohen *et al.* (1999) attempted to determine whether SBHCs' repeated screenings and treatment for gonorrhea and chlamydia would have effects on the infection rate. The team used three high schools with SBHCs and offered over 2000 students at each school an opportunity to get tested for three consecutive years. The team's comparison group was 5063 students at five comparable schools that were used as wait-list controls. Annually, 52–65% of all enrolled students participated, while those who had been in the school for at least 2 years had participation rates of 83.4% for at least one test (Cohen *et al.*, 1999). Intervention and treatment included education, counseling, and treatment with oral single-dose antibiotic therapy.

Research has shown that SBHC may end up saving the states' Medicaid program money. Adams and Johnson (2000) assessed the effects of the Whiteford Elementary School-Based Health Clinic (WESBHC) in Atlanta, Georgia for the years of 1994 through 1996. Their target population of children 4–12 on Medicaid compared to those from another district with no SBHC. Adams and Johnson (2000) identified the sporadic users as well as the frequent users of the WESBHC and analyzed Medicaid claims and expenses per each child enrolled for both the SBHC group and the control group. In 1994, no change in expense was found; in 1995, WESBHC students had significantly lower emergency department expenses; in 1996, WESBHC had lower inpatient, emergency department, nonemergency department transportation and drug Medicaid expenses (Adams and Johnson, 2000). Multivariable analysis confirmed the effect of the WESBHC on lowering the Medicaid emergency department expenses.

To justify SBHCs, some studies looked at the impact those centers made on students, specifically with medical conditions. The Guo *et al.* (2005) study looked at the SBHC impact on children with asthma in terms of hospitalization and emergency department visits. The team conducted a longitudinal quasi-experimental time-series repeated measures design from 1997 to 2003 at four SBHC intervention school districts and two comparable non-SBHC districts in Cincinnati, Ohio. Only children with at least 2 years of continuous enrolment for asthma claims and medications were selected for the sample of 273 children, 196 in SBHC schools, and 77 in non-SBHC schools. The team used student enrolment data and Ohio

Medicaid claims data as well as generalized estimating equation and covariance analysis.

The SBHC programs saw a lowered risk of hospitalization and emergency department visits for children with asthma (Guo *et al.*, 2005). After the SBHC opened, the relative risks of hospitalization decreased 2.4-fold, while emergency department visits decreased by 33.5%. The emergency department visit costs for children in the SBHC were significantly lower than children in non-SBHC schools, and the cost of hospitalization was significantly decreased for children in the SBHC with potential cost-savings estimated at \$970 per child.

One of the more recent evaluations was the Newark Case Study (Silberberg and Cantor, 2008), which looked at three elementary and two high schools with a low-income, largely minority, high-needs population that had SBHCs in Newark, New Jersey. These clinics were created through a partnership with a local hospital and the public school system and had a full-time pediatric nurse practitioner, a social worker, and an administrative assistant working there. They offered referrals for dental, primary and preventive care services, health education and held health fairs.

The study to establish whether or not new models of care were needed in Newark, assess whether clinics already in place were having an effect on health, and shed light on the extent to which student care might be covered by Medicaid and the State Children's Health Insurance Program. A total of 478, predominantly African American, parents were interviewed, 323 from clinic schools, and 155 from nonclinic schools (Silberberg and Cantor, 2008). The findings demonstrated that SBHCs made a difference in health service utilization; however, after the first visit, the study found no association between clinic availability and health-services utilization, which led to a high level of unmet health needs. However, parents rated clinic care significantly and substantially higher than HMO care and gave clinics significantly higher marks for cleanliness, physical condition and translation services when needed (*ibid*). Silberberg and Cantor (2008) found insufficient support for the integration of clinics into public insurance programs. Specifically they found: (1) no general impact on emergency-department use and hospitalization; (2) magnitude of the difference of using healthcare only eight percent with medical, dental, and mental health; (3) after the first-visit effect, there was no impact on health-service utilization (visits to clinics substituted for visits to the community-based provider); (4) no difference between clinics and nonclinic schools on common measures of access to care or health status.

Lessons Learned and Next Steps

Driven by a variety of mechanisms including statutes, policy, and principles, those committed to supporting and enhancing the efforts of SBHC continue to strive to

identify and implement effective models of practice, service provision, and appropriate research/evaluation designs. Despite these efforts, the body of evidence remains somewhat limited and rigorous research/evaluation efforts are needed to continue to build the current evidence base. Over the past 15–20 years, the performance expectations placed on SBHC receiving funds has steadily increased, as it has in most other health, education, and social service settings (e.g., Government Performance and Results Act). The Government Performance and Results Act (GPRA) is a United States law enacted in 1993. It is one of a series of laws designed to improve government project management. The GPRA requires agencies to engage in project management tasks such as setting goals, measuring results, and reporting their progress. In order to comply with GPRA, agencies produce strategic plans, performance plans, and conduct gap analysis of projects). The complexity of the problems, settings, issues, and populations being addressed by school-based health centers presents a number of challenges to randomized designs (e.g., RCT) similar to a few of those discussed previously. Other distinct and equally rigorous frameworks and methods exist that may present distinct advantages in the situations typically found in investigating school-based health centers and the services they provide. These include the use of regression discontinuity designs or propensity score matching procedures. Regression discontinuity (RD) designs are used for evaluating the causal effects of intervention when random assignment is not possible. Assignment to treatment is determined by a cut-off value on the pre-treatment measure. Methodologically, inferences that are drawn from a well-implemented RD design are comparable in addressing threats to internal validity to conclusions from randomized experiments (see Shadish *et al.*, 2002). Another way of addressing the issue of selection bias in nonrandomized experiments is the use of propensity score matching. In its most straightforward form, propensity score matching employs a predicted probability of group membership (e.g., treatment versus control) based on observed predictors, usually obtained from logistic regression to create a counterfactual group (see Luellen *et al.*, 2005).

Trends toward the use of evidence-based practice (e.g., What Works Clearinghouse) are likely to be strong and enduring. There are guidelines and benchmarks being developed both in the United States and internationally across numerous disciplines. Researchers involved in SBHC evaluation need to be involved in their development and application. The increased attention on program outcomes also leads to a challenge that, so far, evaluations of school-based health centers have often navigated successfully – linking program measures to program objectives.

One of the primary challenges in measurement is identifying outcome measures that accurately capture

the objectives of the program. This will become increasingly more important, both in the USA and internationally, as the pressure to be evidenced based is combined with the scarcity of funds and an increase in the number of entities competing for those funds. These combined pressures have a tendency to influence programs into drifting away from actual program objectives and aligned measures, what it was designed to accomplish (e.g., reduce absence), toward a more global level, yet on the surface equally important objectives (e.g., improvement in grades). The idea of drift or distance from the core program objectives can be thought of as a continuum ranging from more proximal to more distal outcomes; outcomes closer to program objectives are referred to as being more proximal while those that are more removed are referred to as distal. For example, measuring the effect of service provision alone, assuming it is effective, will be increasingly more difficult as one uses measures that are further from the actual program objectives, more distal along the continuum (e.g., reducing school-wide absence to increasing school-wide academic performance). To put it more simply, the measurable impact of service provision decreases as the distance between real program objectives and measured program outcomes increases. Evaluators need to be constantly aware of this potential drift. School-based health centers work and the body of evidence supporting this is growing.

See also: Cost Analysis in Evaluation Studies; Evaluation Methodology; Evaluation Use; Program Evaluation; The Role of Stakeholders in Educational Evaluation.

Bibliography

- Adams, K. E. and Johnson, V. (2000). An elementary school-based health center: Can it reduce medicaid costs? *Pediatrics* **105**(3), 780–788.
- Anglin, T. M., Naylor, K. E., and Kaplan, D. W. (1996). Comprehensive school-based health care: High school students' use of medical, mental health, and substance abuse services. *Pediatrics* **97**(3), 318–329.
- Armbruster, P., Gerstein, S. H., and Fallan, T. (1997). Bridging the gap between service need and service utilization: A school-based mental health program. *Community Mental Health Journal* **33**(3), 199–211.
- Armbruster, P. and Lichtman, J. (1999). Are school based mental health services effective? Evidence from 36 inner city schools. *Community Mental Health Journal* **35**, 493–504.
- Brodeur, P. (1998). School-based health clinics. In Isaacs, S. L. and Knickman, J. R. (eds.) *To Improve Health and Health Care, 1998–1999, The Robert Wood Johnson Foundation Anthology*. San Francisco: Jossey-Bass.
- Cohen, D. A., Nsuami, M., Martin, D. H., and Farley, T. A. (1999). Repeated school-based screening for sexually transmitted diseases: a feasible strategy for reaching adolescents. *Pediatrics* **104**(6), 1281–1285.
- Dryfoos, J. G. (1994). *Full-Service Schools: A Revolution in Health and Social Services for Children, Youth, and Families*. San Francisco: Jossey-Bass.
- Edwards, L. E., Steinman, M. E., Arnold, K. A., and Hakanson, E. Y. (1980). Adolescent pregnancy prevention services in high school clinics. *Family Planning Perspectives* **12**(1), 6–14.

- Guo, J. J., Jang, R., Keller, K. N., *et al.* (2005). Impact of school-based health centers on children with asthma. *Journal of Adolescent Health* **37**, 266–274.
- Kaplan, D. W., Brindis, C. D., Phibbs, S. L., *et al.* (1999). A comparison study of an elementary school-based health center: Effects on health care access and use. *Archives of Pediatrics and Adolescent Medicine* **153**, 235–243.
- Kirby, D., Waszak, C., and Ziegler, J. (1991). Six school-based clinics: Their reproductive health services and impact on sexual behavior. *Family Planning Perspectives* **23**(1), 6–16.
- Kisker, E. E. and Brown, R. S. (1996). Do school-based health centers improve adolescents' access to health care, health status, and risk-taking behavior? *Journal of Adolescent Health* **18**, 335–343.
- Klein, J. D. and McCord, M. T. (1992). Evaluation of the Gillespie Student Health Project, Greensboro, North Carolina, Evaluation and Technical Assistance Report to the Robert Wood Johnson School-Based Adolescent Health Care Program. April 4, 1992.
- Kort, M. (1984). The delivery of primary health care in American public schools, 1890–1980. *Journal of School Health* **54**, 453.
- Lear, J. G. (1996). School-based services and adolescent health: Past, present, and future. *Adolescent Medicine* **7**(2), 163–180.
- Lear, J. G., Gleicher, H. B. St. Germain, A., and Porter, P. J. (1991). Recognizing health care for adolescents: The experience of the school-based adolescent health care program. *Journal of Adolescent Health* **12**, 450–458.
- Luellen, J., Shadish, W., and Clark, M. (2005). Propensity scores: An introduction and experimental test. *Evaluation Review* **29**(6), 530–558.
- Mehana, M. and Reynolds, A. J. (2004). School mobility and achievement: A meta-analysis. *Children and Youth Services Review* **26**, 93–119.
- National Assembly on School-Based Health Care. (2005). National Census School Year 2004–2005. <http://www.nasbhc.org/atf/cf/%7BCD9949F2-2761-42FB-BC7A-CEE165C701D9%7D/Census2005.pdf> (accessed October 2008).
- Nelson, P. S., Simoni, J. M., and Adelman, H. S. (1996). Mobility and school functioning in the early grades. *The Journal of Educational Research* **89**, 365–369.
- New York State Department of Health. (1983). *Evaluation of The New York State School Health Demonstration Program, September 1982–June 1983*. EDRS Document Resume ED 246 055, TM 840 172.
- Pastore, D. R., Juszczak, L., Fisher, M. M., and Friedman, S. B. (1998). School-based health center utilization: A survey of users and nonusers. *Archives of Pediatrics and Adolescent Medicine* **152**, 763–767.
- Ricketts, S. and Guernsey, B. (2006). School-based health centers and the decline in black teen fertility during the 1990s in Denver, Colorado. *American Journal of Public Health* **96**(9), 267–275.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: MA: Houghton Mifflin Co.
- Siegel, L. P. and Kriebel, T. A. (1987). Evaluation of school-based, high school health services. *Journal of School Health* **57**(2), 323–325.
- Silberberg, M. and Cantor, J. C. (2008). Making the case for school-based health: Where do we stand? *Journal of Health Politics Policy and Law* **33**, 3–37.
- Weist, P. W., *et al.* (1996). Treatment outcomes of school-based mental health services for urban teenagers. *Community Mental Health Journal* **32**, 149–157.
- Zabin, L. S., Hirsch, M. B., Smith, E. A., Streett, R., and Hardy, J. B. (1986). Evaluation of a pregnancy prevention program for urban teenagers. *Family Planning Perspectives* **18**(3), 119–126.

Relevant Websites

- <http://www.nasbhc.org> – National Assembly on School-based Health Care.
- <http://www.healthinschools.org/> – The Center for Health and Health Care in Schools.
- <http://ies.ed.gov/> – What Works Clearinghouse.

Formative Assessment in Teacher Education and Teacher Professional Development

H Tillema, Leiden University, Leiden, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In teacher education and the professional development of teachers, assessment is increasingly being recognized as a valuable tool to promote development and growth in the profession, that is, not only as a means to certification and qualification (Cochran-Smith, 2005) but also as a tool for learning. First, the nature of formative assessment is circumscribed; it essentially means providing informative feedback to the learner and its prime function is to deliver insight into improvements in performance. Second, formative assessment in teaching is delineated as a process of compiling performance evidence, appraising the collected information, and, subsequently, providing feedback in such a way that teaching practice(s) can be improved. It is this (in)formative nature of assessment that makes it valuable to professional growth. Third, it is clarified how different assessment tools or interventions may help the (student)teachers as learners to benefit from assessment which depends to a large extent on the (1) nature of feedback given and (2) the assessment arrangement (i.e., tools and approaches). Finally, reviewing the research on effective formative assessment arrangements, it is shown in what ways feedback can concentrate on (bridging the discrepancies between) where the learner is and wants to go, that is, develop (Sadler, 1998). This implies that formative assessment needs to be integrated and aligned with instruction (Segers *et al.*, 2003) and occurs alongside with mentoring for development (Wang and Odell, 2004).

The Nature of Formative Assessment

Formative assessment is about feedback. In teacher formative evaluation, assessment instruments are means to provide feedback on action, and it is through this informative feedback that one learns (Hattie and Timperley, 2007). Supporting teachers as learners requires moving beyond providing mere knowledge of results to a type of functional (i.e., performance-related) feedback that can articulate developments in practice, anticipate professional learning needs, and monitor the learner's progress during a course of teacher action. Assessment instruments, thus viewed as learning tools, are to enhance insight into both current and aspired levels of performance and, based on this, are to provide possible or alternative learning routes (Zeichner

and Wray, 2003). We need to distinguish this learning-oriented, (in)formative assessment from a summative, often externally initiated or mandated, assessment that documents and appraises work performance in relation to mandate evaluation requirements (De Landshere and Arens, 2003). Assessment in this latter sense is interested in the certification of marked achievements which are being appraised and judged against a set of external standards (often labeled as standard testing or competence testing; Sachs, 2003). Summative assessment for achievement evaluation usually occurs after learning has taken place; and as such, it has its own legitimized function in teacher education (i.e., serving an accountability warrant). Formative assessment, however, occurs in alignment with instruction and learning (i.e., to inform the learner about redirecting and regulating development in competence growth) and adds value to recognition, interpretation, and utilization of teaching experiences in line with the (often personal) goals for achievement (Zimmermann, 2000). Collecting such targeted feedback information by assessors, often together with their assesseees, is meant to scaffold a meaningful appraisal and subsequent decision making on the (alternative) routes to be taken for further professional action (Loughran, 2003). Typically, formative assessment tools document and elucidate a cyclical or recurrent process of professional activity in order to provide a better understanding of one's practice experiences in teaching (Smith and Tillema, 2003). This is accomplished through a continuous monitoring of performance over extended periods of time. As such, it is usually a highly authored process of personal data collection (Wang and Odell, 2003). The formative assessment process, therefore, evolves alongside professional practice to scaffold and integrates professional development goals and learning needs (Havner and McDowell, 2007).

Providing Informative Feedback

Reviews on the impact of formative assessment (OECD, 2005; Lievens, 2001; Darling-Hammond and Bransford, 2004) show that learners neither value feedback *per se* nor accept its deliverance wholeheartedly. Indeed, they are sometimes reluctant to use feedback information that is provided to them. Therefore, simply providing corrective feedback can be counterproductive and may elicit resistance,

Collection	(Reflection)	Appraisal	Recommendation	(Further learning plans)
Compilation and documentation phase	Conceptualization and elucidation of collected data	Provision of feedback and dialog phase	Acceptance of feedback and following recommendations	Intended actions in practice contributing to further professional development

Figure 1 Stages in the deployment of formative assessment for teaching. The steps in brackets refer to an extended interpretation of the formative assessment process, with special relevance to teacher development.

especially when issued too frequently (Falchikov, 2005). Furthermore, studies show clearly that acceptance of feedback is related to the way it is provided, that is, tool dependent (Hattie and Timperley, 2007), which essentially means that particular assessment tools may serve specific learning goals. Moreover, it needs to be noted that beliefs on personal efficacy filter the feedback information (Waldman and Atwater, 1998). In sum, it has been found that factors in feedback delivery (i.e., the assessment tools used) result in divergent acceptance of feedback (Shute, 2008). Formative assessment, therefore, is a complex process, involving distinct steps or stages in its deployment. **Figure 1** depicts the interplay between them with special reference to teacher learning and development.

The collected and to-be-appraised assessment information must fulfill certain requirements to be followed for action, that is, be of relevance to the improvement of teacher practice. Above all, it must be cognitively acceptable in that it correlates with a learner's prevalent goals or perceptions about aspired performance levels (i.e., both in expectations and actual goal setting – Zimmerman, 2000). Crucial to acceptance of feedback is the learners' resolution whether following recommendations is of any advantage to their future action (Topping, 1998). Positive recommendations tend to be accepted and followed more easily than negative or corrective feedback (Waldman and Atwater, 1998). Recommendations, however, given along different performance dimensions (Thornow, 1993) are not found to contribute to a targeted learning. Clear and univocal recommendations that may guide future action are to be preferred over breadth and completeness of advice. Typically, assessment, in which information is collected and appraised in reference to personal standards and aspirations, proves to be a helpful feedback instrument (McDowell, 1995). Furthermore, studies indicate that the amount of feedback information given during appraisal may influence its acceptance (Wade and Yarbrough, 1996); a focused provision and delivery helps to integrate feedback information and subsequent action. In overview, several features in the arrangement of formative assessment have impact on following recommendations. Particularly, assessment as an intervention is considered positively when it is characterized by the features listed in **Table 1**.

Table 1 Principles for providing feedback in assessment for learning

- Being task and goal oriented
- Consistent with pre defined criteria
- Give positive praise and constructive comments
- Focused on particular qualities of the person involved
- Identification of strength and weaknesses in performance
- Guidance about what to improve in performance
- Held in climate of trust
- Occurs strategically throughout learning
- Has a shared understanding between assessor and assessee about purpose of feedback

These principles of assessment interventions and feedback delivery are extracted from proposals of the Assessment Reform Group in Great Britain: *Assessment for learning: 10 principles* (2002) www.assessment-reform-group.org.uk and the OECD report on *Formative Assessment: Improving Learning in Secondary Schools* (2005).

Assessment as Intervention

The way formative assessments are being arranged (i.e., as a deliberate intervention in support of learning) strongly influences what will be learned (i.e., what will be followed as recommendations for action). The strength of an assessment intervention (which is more than the tool itself) is essentially determined by the way it exhibits, appraises and interprets the collected evidence on performance. An assessment arrangement starts with the collected record of performance evidence and ends with the modification of that performance (see **Figure 1**). Typically, in assessment for teacher development, it is the collector, that is, (student) teacher, who compiles instances of practice-related, authentic (Wiggins, 1998) performance and in fact determines what is offered for appraisal. The assessor (i.e., teacher educator, or mentor), in turn, typically is put in the role of feedback provider (Loughran, 2003); and both, that is, (student) teacher and assessor/mentor, are involved in valuing the gathered information (Delandshere and Arens, 2003). Although this process of assessment as intervention in teaching may seem straightforward, it requires a set of framing factors that leads to different assessment arrangements. The arrangement of an assessment is at least governed by the following framing or design factors:

1. the way information is collected, that is, how the collected information is gathered as evidence;
2. the content to be provided in the assessment (the domains and topics in the collected information), that is, what is being collected;
3. the nature of evidence collected (the type of information that will be regarded as relevant), that is, what counts as evidence;
4. the involvement of different types of raters to appraise the collected materials, that is, who is being rated by whom;
5. the purpose of further development (degree of autonomy in support and mentoring), that is, at what stage in development is the information utilized;
6. that is, the criteria for appraisal (i.e., the selection of levels in achievement), which means by what standards it is reviewed;
7. the (time and contextual) frame in which information is collected (setting and duration of the collected information), that is, when and where it is applied.

Table 2 explicates the design decisions that are possible in the arrangement of a formative assessment intervention.

How an assessment for teacher development actually is arranged (i.e., as highly structured with strict standards and pre-specified conditions or not) will determine the nature and impact of feedback. That is, the assessment as an intervention articulates the functionality of feedback.

Different Tools of Assessment for Learning in Teaching

There are several assessment tools in use in teacher education and teacher development to support professional competence development.

The Portfolio

The portfolio (or its digital version: the e-portfolio) is one of the most widely used formative assessment tools in teacher preparation and development; in essence, it is a purposeful collection of work over a period of time. The core elements of the portfolio as a product are:

- *An index.* An indication of what performance evidence is to be included in the portfolio; often in reference to

the teacher-education program's intended learning outcomes.

- *The description of the portfolio entry.* For each entry, clarifying information indicates what actually has led to or would scaffold, and helps to interpret the collected evidence.
- *The evidence.* The core of the portfolio compilation consists of several folders in which the learner has gathered materials that demonstrate certain performance outcomes for each identified goal. This may take several forms, such as analysis of pupil observation, critical incidents in classroom management, work/timetables, notes from project meetings, mentor evaluations, or other suitable materials.

The following elements of the portfolio can be optional:

- *Learning experiences.* An indication of the learning experiences encountered with respect to each particular learning goal/entry.
- *Proposed further actions.* A self-determined indication of the future steps needed to be taken for further development.

The information provided in the portfolio product constitutes the input for the subsequent feedback meeting with the assigned teacher mentor or professional development coach as the feedback provider. The utilization of framing factors in the design of the portfolio (see **Table 2**) leads to different types of teaching portfolios:

1. *Dossier portfolio.* This type of portfolio specifies products of work-related performance included by the learner as best evidence of accomplishments in teaching. This dossier portfolio is a vehicle to collect practice materials to be evaluated in appraisal meetings with the (student) teacher's mentor, coach, or critical friend (Day, 1999). The provided feedback results in concrete assignments and redirections of future work.
2. *The (professional) development portfolio.* During a course (in teacher education or in a professional development program), a detailed description of evidence is compiled to show the attainment of program objectives by the learner. This type of portfolio highlights accomplishments regarding specific targets and favors a display of successive steps in the attainment of goals. The following additional steps to the portfolio's core elements are included (see **Table 2**):

Table 2 Design features of an assessment arrangement

Collection of information	Content to be provided	Nature of evidence	Involvement of raters	Purpose of support	Criteria in use	Context of assessment
Guided	Bounded, restricted	Best evidence	Single assessors – single assessee	Reflective oriented	Prespecified	Constrained
Open	Unspecified	Sampled collection	Multiple assessors – multiple assessees	Performance oriented	Negotiable	Open

- a. *Defined competences/targets to be attained.* Often in the form of a contract, the learner together with the teacher educator/coach specifies which competences are to be given attention during a certain period of time.
 - b. *Self-assessment.* As the contract represents the personal learning goals for the learner, the goals themselves are often framed as discrepancies between self-perceptions about competence levels and the external standards set by the program.
 - c. *Portfolio compilation.* For each specific learning goal, performance evidence is collected, showing visible and tangible proof of competence attainments in the actual work environment, relating it to the standards of competences stipulated in the contract.
 - d. *Presentation of the portfolio.* At agreed-upon moments during the course/training/development program, the learner offers his or her portfolio for identifying the accomplishments that merit further effort in development.
3. *Reflective or personal portfolio.* This type of portfolio is best described as a learner report or account on the professional growth of being a teacher, attained over an extended period of time. The reflective portfolio records notable achievements in daily professional practice (not only best evidence as is the case in the dossier). The choice of records is defined, and selected beforehand by the learner in the introductory, reflective part of the portfolio-index. The learner is the prime owner of both the portfolio construction process as well as the subsequent appraisal process. The reflective portfolio adds the following elements to the core portfolio:
- a. *Focus and scope.* The learner makes a deliberate choice regarding the relevant professional targets to pursue during the period of time and selects various learning goals to be worked out at greater depth.
 - b. *Evidence.* The nature of evidence can be highly diverse and may consist of narrative, comment, or biographical notes for each selected entry/goal, thus helping the learner to embark on a reflective learning dialog to be shared with a critical friend as an assessor.
 - c. *Reflections.* The learners examine their own learning, their efforts, and their achievements as a means of monitoring and critiquing their own work.
 - d. *Future action.* The appraisal is used to initiate either a process of collecting new evidence or establishing new professional learning goals.

Peer Assessment

Peer assessment is a process by which learners rate their peers, and is, as such, of great relevance to teacher development. It signifies the joint collaboration by those

involved in the learning process in the appraisal of their own learning. In a peer-assessment arrangement, the learners consider “the amount, level, value, worth, quality or success of learning of peers of similar status” (Topping, 1998). Peer assessment is not only a direct appraisal of what has been learned (outcomes) but also of the how of learning (process). The supposed beneficial effects of peer assessment are not only diverse, but also inconclusive (Falchikov, 2005). Peer assessment (and its related format: co-assessment, that is, mentor/assessee) is said to help learners develop meta-cognitive skills, for example, communication skills, self-evaluation skills, observation skills, and self-criticism (Havner and McDowell, 2007); and this may lead more readily to acceptance of feedback. However, supposed effects of peer assessment for learning vary considerably. The findings range from better attendance, learning gains, impact on the ability to self-assess, developing critical thinking, to no effects at all (Topping, 1998). Peer assessment in essence is a social appraisal process where feedback is given to and received by others, aimed at enhancing the performance of the learner. Therefore, interpersonal and interactional processes play an important role, such as psychological safety, value diversity between peers, interdependence in social relations, and trust in the other as an assessor. Framing features in the arrangement of peer assessment might condition how peers step into the process of appraising each others’ learning results. A first set of framing features has to do with specifying the contextual arrangement of the assessment, that is:

1. the why, that is, reasons for utilizing peer assessment;
2. the what, that is, objectives, teaching areas, and products/outcomes;
3. when, that is, time;
4. where, that is, place; and
5. how, that is, is it supplementary to grading or required; compulsory or voluntary?

A second set of framing features considers the interaction among peers in the appraisal; because of the interpersonal factors mentioned, the assessment might vary with respect to who assesses whom. This directionality in peer assessment can be one way (from assessor to assessed), reciprocal (peers assess each other, e.g., in pairs), and mutual (all peers assess all peers). In addition, peer assessment may differ in level of privacy (anonymous, confidential, and public) and nature of contact between assessor and assessee (at a distance or face to face).

A third set of framing features refers to the composition of the peer group that provides feedback – it can differ in ability or not; its constellation can vary or not.

In teacher education, student teachers often work and practice together during practice teaching/mentoring. Also in professional development programs, teachers work together as colleagues and share learning experiences.

This setting provides a platform for peer assessment in which the 'learners' appraise each other as critical friends (Edwards *et al.*, 2002).

Self-Assessment

Self-assessment is the relatively autonomous and deliberate engagement in reviewing and critiquing one's work in an appraisal of progress made over a period of time. Often self-assessment is closely aligned with (self)-monitoring and reflection (on action). It is meant to increase the learner's self-responsibility and self-regulation in learning (Zimmermann, 2000). In this sense, self-study, action research, or inquiry and analysis of one's work are the main vehicles of self assessment. Learners engaged in self-assessment are found to be more interested in their work and more able to interpret why and what they are doing. Also by analyzing one's own work self-assessment builds ownership and high expectations on the improvement of one's work, mainly through the internalization of criteria and standards of performance. Deployment of self-assessment therefore gains insights that can be used for further learning. As an arrangement, self-assessment involves procedures as to making judgments about accuracy, worth, and appropriateness within a learning context. In essence, self-assessment is a reflective activity and develops reflective as well as meta-cognitive skills. There are three major formats of self-assessment, mainly differing in whether agreed upon judgments are made on the level of performance or whether monitoring progress in work is involved, thus varying in the level of control by the learner:

- Self-evaluation is a process of critical evaluation of one's performance, in which explicit criteria are being used to scrutinize and appraise work against a set of agreed upon standards. Self-collected performance results are offered for deliberate evaluation in conferencing or supervision meeting with an assessor. This appraisal builds on the application of explicit criteria to the improvement of performance.
- Self-monitoring is a way of continuously looking at practice and meant to scaffold a better understanding of one's task performance to gain insight in its strength and weaknesses. This is often considered a more informal self-evaluation due to its openness and fluidity of goals and criteria. It merges often with activity scheduled within a formal course or curriculum.
- Self-reflection exceeds almost its assessment function as a generic vehicle to scrutinize one's accomplishments. As reflection on action it has found a strong foundation in teacher education and, scaffolded by self-regulation, is meant to direct action and clarify performance.

However it needs to be noted that self-assessment in any context needs to be conducted with a clear reference to

explicit (i.e., shared) standards and deliberate judgments by (external) assessors, who may be operating as either a critical friend or a supervisor.

Research-Based Principles of Assessment for Learning

Strongest benefit of assessment for learning is the insight it provides in performance and the subsequent support it gives for its improvement. Formative assessment's outstanding meaning is to spell out acquired competence in a concrete, visible, and tangible way (Smith and Tillema, 1998). Assessment formats may differ, however. Self-assessment preferably should be carried out in light of (1) feedback from significant others, (2) with a clear focus on content or performance, and (3) in comparison with external sources of evidence. Peer assessment requires favorable conditions regarding interpersonal exchange, fairness, and trust, while openness and reflexivity to look at one's accomplishments appear to be crucial in the portfolio. The manner in which different formative assessment arrangements orchestrate feedback ultimately must correspond, that is, needs to be aligned, with performance goals (i.e., what is actually intended to be achieved). In determining such a precise match between assessment/feedback, it is important to look at how feedback (as a method) can provide for acceptance and following recommendations of the assessment information. Nonetheless, no single feature of feedback delivery serves each type of assessment best. However, what seems to be essential in all of them is a performance-based focus in the delivery of feedback. In reference to teacher development, the link with practical action and work assignments seems to be an additional crucial feature in all of the assessment methods mentioned. Looking at differences between assessment tools will indicate that reflection is a differential but highly effective element (**Figure 1**) in bringing about performance change. Reflection enables a cognizant and meta-cognitive focus on one's self-determined levels of performance, both in the collection of evidence phase as well as later on, in the appraisal and action phase.

The results from studies on assessment interventions in teacher education support a view of formative assessment as a process in which feedback is closely associated with the expressed intentions of both the assessee and the assessor and connected to explicated criteria that are considered relevant by both in the appraisal of the compiled information. A valuable attribute of formative assessment, then, would be the increase of learner control over the content to be assessed as well as over the criteria by which the teaching performance will be scrutinized. In this way, deployment of tools can become both continuous and nonthreatening; or put otherwise, assessment can become a bridging tool between learning needs and actual

Table 3 Principles of formative assessment for teaching, teacher education, and quality assurance

<i>Formative assessment of teachers: guiding principles for assessment practices</i>	<i>Principles for educating teachers in formative assessment</i>	<i>Quality measures to develop formative assessment</i>
Assessments should:	(Student) teachers, practicing teachers and teacher educators, at all levels of teacher education (pre-service, in-service, professional development activities for teachers and teacher educators), need to be given ample opportunity, time and guidance in order to:	Assessment quality, at the local school, or institutional level, entails:
1. Focus on how (student) teachers learn	1. Master the principles of formative assessment and learn how to integrate formative assessment in the learning/teaching process	1. A vision of the relevance and the power of formative assessment for improving learning
2. Be included in the planning of teaching and learning	2. Experience the power of formative assessment as learners in teacher education programs	2. Develop legislation recognizing the role of teachers' professional judgement in the practice of formative assessment
3. Be an integral part of classroom practice	3. Implement various forms of formative assessment and develop personal conceptions of assessment practices that can empower learning	3. Links between formative and summative assessments (e.g., use of formative assessment to prepare summative assessment; use of data from summative assessment for formative purposes at the formative and school levels)
4. Be aimed at significant developmental goals and objectives	4. Undertake diversified activities that allow critical reflection about formative assessment (analysis of examples of (student) teachers' work, self-observation through video-recording, reciprocal observation with peers and colleagues)	4. Guidelines regarding the integration formative assessment practices in curriculum and other teaching materials
5. Recognize the full range of all (student) teachers' achievements	5. Discuss beliefs and attempts to implement formative assessment, with the support of experienced and knowledgeable mentors	5. Tools and teaching resources to support formative assessment and share experiences with formative assessment
6. Promote shared understanding of learning goals and assessment criteria	6. Carry out informed, evidence-based professional decisions about formative assessment having lasting positive effects on (student) teachers learning and achievements	6. Further research and pilot projects to develop strategies and techniques for effective formative assessment
7. Be constructive and sensitive to (student) teachers' feelings	7. Foster the common goal of improving learning for each and every learner across the different levels of the educational system (preschool, primary, secondary, vocational)	7. Schools and districts to be responsible for continuous monitoring of the learning progress of their (student) teachers
8. Take into account (student) teacher diversity	8. Explain the power of formative assessment to stakeholders in the educational system	8. Measures to inform all stakeholders (school management, professional association, teachers, parents, (student) teachers) about the benefits of formative assessment to improve learning

9. Foster (student) teacher motivation for learning
 10. Encourage positive relations and productive exchanges among (student) teachers and between teachers and (student) teachers
 11. Entail forms of regulation (appropriate feedback, differentiated guidance, adaptive instructional activities) that promote (student) teachers learning
 12. Encourage active (student) teachers involvement in assessment (self-assessment, peer and teacher and (student) teachers co-assessment)
 13. Be based on varied sources of information (written/oral, individual/group, quantitative/qualitative)
 14. Situate (student) teachers learning outcomes with respect to educational objectives
 15. Be communicated in a transparent and coherent way to concerned parties ((student) teachers, other teachers)
 16. Enhance (student) teachers' capacity to undertake both independent learning and shared learning in group settings
9. Conditions of professional development that allow teachers to feel confident that their job, status and professional recognition is strengthened when implementing formative assessment to promote learning
 10. Conditions of external assessment (for accountability and quality control) that allow formative assessment to develop fully its role in supporting learning

The principles presented here evolved out of exchanges at the International Conference on Classroom Assessment, held in Portland, Oregon, in September 2005. This conference examined classroom assessment experiences in Australia, Canada, Great Britain, New Zealand, United States, as well as in several countries on the European continent. This overview is authored by: *Linda Allal (University of Geneva), Janet Looney (OECD, Paris), Kari Smith (University of Bergen), Harm Tillema (Leiden University), and Joke Voogt (University of Twente)* and was modified for the purpose of this publication. A copy of the original leaflet can be downloaded from www.uib.no/iuh/forskning/pca.pdf.

competence levels. Admittedly, student teachers and their teacher educators often have great difficulty with such a self-reliant approach to assessment (i.e., being more accustomed to viewing assessments as summative, externally controlled, and with objective procedures) (Wiggins, 1993). Actively collecting and negotiating assessment information, however, lies at the heart of teacher development. Therefore, mastering a self-directed and reflective use of assessment instruments is an effective way of performance monitoring and a valuable tool for redirecting one's learning.

Still, it needs to be recognized that no single intervention produces all relevant feedback. The feedback process is complex (Darling-Hammond and Snyder, 2000). Functional feedback essentially means setting the goals for learning and reflection first (Sadler, 1998) and then focusing on the tools that offer and scaffold careful diagnosis and monitoring of performance (Redman, 1994). A formative assessment process undoubtedly offers self-initiated opportunities for reflection in which the individual learner appraises set goals. However, on the other hand, this self-initiated process may fall short of scaffolding an illuminative evaluation that can promote clear vision and better conceptual understanding of advancements in learning. Without any form of mentoring or supervision, the feedback intervention may leave the learners stuck within their own realm of thinking (Zeichner and Wray, 2000). A collaborative or joint approach in feedback delivery is more conducive to the pursuit of goals in teacher development. Formative assessment, irrespective of who initiates the evaluation, needs to be framed and structured in a way that enables appraisal of collected experiences for the purpose of further development. Principles that may guide this process are listed and summarized in **Table 3** to put formative assessment into perspective for teacher development.

See also: A Pedagogy of Teacher Education; Accreditation and Standards in Teacher Education; Assessment in Higher Education; Assessment in the Workplace of Performance, Developing Expertise and Competence; Conceptions of Teacher Education; Contemporary Approaches to Teacher Professional Development; Evaluation of Adult Education and Training Programs; Evaluation of Teacher Quality and Practice; Impact of Assessment on Students' Learning Strategies and Implications for Judging Assessment Quality; International Evaluations; Learning as Inquiry; Mentoring in Teacher Education; Narratives and Biography in Teacher Education; Needs Assessment in Education; Peer Learning in the Classroom; Professional Development of Teacher Educators; Taking Prospective Teachers' Beliefs into Account in Teacher Education; Teacher Education and the Educational Foundations Knowledge Base; Teacher Education for Elementary Education; Teacher Induction; Teacher Learning with Lesson Study; The Competency Model.

Bibliography

- Cochran-Smith, M. (2005). The new teacher education, for better or for worse. *Presidential Address to the American Educational Research Annual Meeting*, Montreal, 11–15 April 2005.
- Darling-Hammond, L. and Bransford, J. (2004). *Preparing Teachers for a Changing World, What Teachers Should Learn and Be Able To Do*. San Francisco, CA: Jossey-Bass/Wiley.
- Darling-Hammond, L. and Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education* **16**, 523–545.
- Day, C. (1999). *Professional Development of Teachers*. Buckingham: Open University Press.
- Delandshere, G. and Arens, S. A. (2003). Examining the quality of the evidence in pre-service teacher portfolios. *Journal of Teacher Education* **54**(1), 57–73.
- Edwards, A., Gilroy, P., and Hartley, D. (2002). *Rethinking Teacher Education, Collaborative Responses to Uncertainty*. London: RoutledgeFalmer.
- Falchikov, N. (2005). *Improving Assessment through Student Involvement. Practical Solutions for Aiding Learning in Higher and Further Education*, 151 p. London: RoutledgeFalmer.
- Hattie, H. and Timperley, J. (2007). The power of feedback. *Review of Educational Research* **77**(1), 81–112.
- Lievens, F. (2001). Assessor training strategies and their effects on accuracy, interrater reliability, and discriminant validity. *Journal of Applied Psychology* **86**, 255–264.
- McDowell, L. (1995). The impact of innovative assessment on student learning. *Innovation in Education and Training International* **32**, 302–313.
- OECD (2005). *Formative Assessment: Improving Learning in Secondary Education*. Paris: OECD.
- Sachs, J. (2003). Teacher professional standards, controlling or developing teaching. *Teachers and Teaching* **9**(2), 175–187.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education* **5**(1), 77–85.
- Segers, M., Dochy, F., and Cascallar, E. (2003). *Optimising New Modes of Assessment: In search for Qualities and Standards*. Dordrecht: Kluwer Academic.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research* **78**(1), 153–189.
- Smith, K. and Tillema, H. H. (2003). Clarifying different types of portfolio use. *Assessment and Evaluation in Higher Education* **28**(6), 621–644.
- Thornow, W. W. (1993). Perception or reality? Is multiperspective measurement a means or an end. *Special Issue: 360 Degree Feedback. Human Resource Management* **32**, 221–230.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research* **66**, 249–276.
- Wade, R. C. and Yarbrough, D. B. (1996). Portfolios: A tool for reflective thinking in teacher education. *Teaching and Teacher Education* **12**(1), 63–79.
- Waldman, D. A. and Atwater, L. A. (1998). *The Power of 360-Degree Feedback: How to Leverage Performance Evaluations for Top Productivity*. Houston, TX: Gulf.
- Wiggins, G. (1998). *Educative Assessment, Designing Assessment to Inform and Improve Student Performance*. San Francisco, CA: Jossey-Bass.
- Zeichner, K. and Wray, S. (2000). The teaching portfolio in US teacher education programs: What we know and what we need to know. *Teaching and Teacher Education* **17**, 613–621.
- Zimmerman, B. J. (2000). Attaining self-regulation, a social cognitive perspective. In Boekaerts, M., Pintrich, P. R., and Zeidner, M. (eds.) *Handbook of Selfregulation*, pp 13–39. San Diego, CA: Academic Press.

Further Reading

- Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education* **5**, 7–74.

- Cochran-Smith, M. and Fries, M. K. (2002). The discourse of reform in teacher education: Extending the dialogue. *Educational Researcher* **31**(6), 26–28.
- Havnes, A. and McDowell, L. (eds.) (2007). *Balancing Dilemmas in Assessment and Learning in Contemporary Education*, pp 183–197. London: Routledge.
- Loughran, J. (2007). Researching teacher education practices responding to the challenges, demands, and expectations of self-study. *Journal of Teacher Education* **58**(1), 12–20.
- MacLelland, E. (2004). How convincing is alternative assessment for use in higher education? *Assessment and Evaluation in Higher Education* **29**, 311–321.
- Shephard, L. (2000). The role of assessment in a learning culture. *Educational Researcher* **29**(7), 4–15.
- Wang, J. and Odell, S. J. (2002). Mentored learning to teach according to standards based reform, a critical review. *Review of Educational Research* **72**(3), 481–546.

Informal Education and Evaluation

T Jeffs, Durham University, Durham, UK and University of Bedfordshire, Luton, UK

© 2010 Elsevier Ltd. All rights reserved.

In Disraeli's novel *Sybil*, published in 1845, the life course of the hero Egremont is changed by an accidental conversation with a stranger whom our hero never encounters again or learns the identity of. During the course of that conversation Egremont for the first time reflects on the nature of community and the two nations that comprised Victorian England – the rich and the poor. Over time he revisits that conversation in ways that lead him to reassess both his political beliefs and the direction of his life until eventually he feels obligated to devote himself to seeking ways to remove the social and political chasm separating rich and poor. This example is no mere literary device. Biographies and conversations with older people are replete with similar examples of men and women who at formative moments in their lives experienced similar encounters or experiences: unplanned, unintended encounters that, for good or ill, altered the direction of their lives by demanding that they question and reflect upon who they are, what they believe in, and what they are doing – exchanges that made the familiar unfamiliar, that cultivated learning through reflection and analysis.

Informal educators seek to play a role similar to that stranger in relation to the lives of those they work with. Through the medium of conversation and dialog, they attempt to encourage others to reflect upon what they believe and what they do. They deliberately seek out opportunities to promote educational encounters that foster learning through creative dialog. In doing so they rarely operate randomly, rather they specifically aim to work with and alongside individuals and groups in particular settings. Historically informal educators, paid and unpaid, have predominately, but not exclusively, worked with two groups. One has been young people as they undergo the transition for childhood to adulthood. The intention being to help them clarify their hopes and aspirations; to assess where they are and what they wish to become; and help them negotiate the tensions of adolescence. The second has been those living in troubled and disadvantaged communities.

Ordinarily informal education takes place beyond the classroom – on the street, in youth clubs and settlement houses, in residential facilities and faith settings, as well as in schools and colleges (Jeffs and Smith, 1990). Essentially, it can occur wherever individuals and groups congregate. With regard to schools and colleges, this will be in corridors, dining halls, playgrounds, and public spaces during breaks and before and after formal lessons. Although times and venues are not fixed, this should not be taken to imply that informal education happens purely

by chance. Rather, as Hazler (1998) explains, it usually happens because practitioners behave in ways that encourage others to engage with them and because they pay careful attention to the timbre of the environment in order to maximize the opportunities for conversation. In other words, informal educators make themselves accessible to others and act in ways that foster trust and engagement. Although some practitioners do refer to themselves as informal educators, many do not. They might call themselves, for example, youth workers or youth ministers, community or settlement workers, and adult or outreach educators. In addition, they may at other times be designated, for example, a counselor, school teacher, social worker, or minister of religion. Irrespective of the title what they have in common is that they adopt for at least part of their working day a form of educational practice that consciously seeks to promote learning and impart skills via the medium of conversation and dialog. Also, whatever the setting the relationship is invariably a voluntary one, because those who work alongside retain the freedom to refuse to engage in conversation with the informal educator or to curtail it, and the connection, whenever they choose. Given the basis for the relationship, and the mode of teaching, the content of the conversation and trajectory of the dialog will always be open to negotiation. As Gadamer explains, it is “a characteristic of every true conversation that each opens himself to the other person, truly accepts his point of view as worthy of consideration and gets inside the other to such an extent that he understands not a particular individual, but what he says. The thing that has to be grasped is the objective rightness or otherwise of his opinion, so that they can agree with each other on a subject” (Gadamer, 1979: 347). As a consequence of informal education being based on dialog and conversation, those who engage upon it must teach without recourse to a syllabus or a curriculum.

Nevertheless, although conversation may be unpredictable, informal educators must operate in methodical and professional ways. This entails paying careful attention to:

- the presentation of self to ensure they are both approachable and accessible;
- the sustenance of relationships;
- their knowledge base to enable them to respond in an informed way when required to do so;
- their understanding of the social and cultural norms prevalent within the community in which they work so as to ensure that they behave in appropriate ways;

- the development of a repertoire of techniques relating in particular to the arts of conversation and group work so that opportunities for learning are maximized;
- their behavior and public persona. Besides teaching via conversation and dialog, informal educators do so through the modeling of behavior. Those around them are likely to scrutinize their demeanor for informal educators are engaged in a moral craft that demands constant examination of their motives and conduct.

With respect to all the above, and more besides, informal educators must interrogate and reflect upon their practice to ensure that they help others learn and flourish. They must seek to become at all times what Schon (1983) terms a reflective practitioner some one who accumulates a catalog of examples, images, insights, understandings, and procedures – a repertoire that helps them to make sense of their practice and the situations they encounter and evaluate their actions *in situ* and strengthen their practice via reflection.

Emergence of an Evaluation Culture

Josephine Macalister Brew's (1946) *Informal Education: Adventures and Reflections* (see also Smith, 2001) was the first text to explicitly discuss this genre and employ the term. Although the label may be of recent origin, this educational format, characterized by spontaneity and built upon the interplay inherent in dialog and conversation, has an unbroken history that can be traced back to Athenian society and the ideal of the Socratic dialogue (Jeffs, 2001). Neither Brew nor any of the subsequent literature to emerge during the next quarter of a century broached the topic of evaluation. However, they consistently argued that it was essential for practitioners to reflect upon their work and continually engage in what amounted to a form of evaluation-in-practice, in other words undertake formative evaluation – usually via formal supervision (Tash, 2000) or the interrogation of practice with colleagues in time set aside, often after the completion of a session or piece of work. A helpful example of the incorporation of this approach is to be found in Marie Paneth's *Branch Street* (1944) an account of an action-research project undertaken with young people in a deprived bomb-damaged neighborhood during wartime. Paneth repeatedly interrogates and reflects upon her work and relationships with colleagues, the community, and the young people involved in the project. Evaluation, though she does not use the term, takes the form of dialog and reflective discussion with those involved and is directed toward moving the project forward and discovering, and then responding to, the needs of the young people.

Historically, informal educators conveyed a relaxed stance with regard to the importance of measureable or

identifiable outcomes. According to Brew they “must try not to be too concerned about results, and at all costs not to be over-anxious. It is always difficult to measure any good influence in the total lives of human beings” (Brew, 1957: 183). However sporadic attempts were made during this early period to evaluate effectiveness via the measurement of outcomes. For example, the study undertaken by Brown (1956) of the Red Shield Boys' Club in Louisville claimed to demonstrate that it achieved a 52% decrease in juvenile delinquency in the 11 years it operated from 1944 to 1955. In comparison, delinquency rates tripled in one comparable neighborhood and increased 33% in another. Such empirical evaluative research was uncommon for a number of reasons. First, because informal education, in the guise of youth work, adult education, and community work, was viewed as being of self-evident value. An essential component of civil society sustained by an army of volunteers who by undertaking this work contributed to the vibrancy of democratic structures and the accumulation of what Putman (2000) subsequently labeled social capital. Therefore, for key writers on community work such as Follett (1929), adult education (Lindeman, 1926), and youth work (Brew, 1957), the value resided not only in the measurable outcomes of daily practice but also in the ways informal education enriched the social networks that fostered a sense of community and a democratic way of life. Citizenship, equity, and service lay at the heart of informal educational practice both as guiding principles and outcomes. As Paneth (1944: 46) explained “my goal in educational work was to make good, independent citizens in a good community.” Second, because evaluation was not imposed on practitioners or required by those who helped support the work via charitable donations. In part, this was because the work was overwhelmingly undertaken by volunteers who saw a transparent value in what they freely chose to do and fund. Third, both young people and community groups who engaged with the informal educators in clubs and settlements were not, at least in theory, viewed as clients or users, but as members or, especially the young people, as citizens-in-the-making. Therefore, democracy and self-government must be given practical expression which meant informal educators sought ways to make themselves answerable to the members and management committees elected in part or whole by those same members. Consequently evaluation, as we might understand the term, was not a priority for members who had direct control or at least an influence over issues relating to priorities, outcomes, and where applicable, the performance of professional staff. The quality of the program and the worth of the project was continuously monitored by members who were able to seek change through democratic mechanisms or could withdraw their financial support by not paying membership subscriptions. The staff, salaried or not, were unremittingly evaluated by

those they served. Judgments regarding the quality and value of their contribution were inevitably incorporated into the daily round of relationships for they were the servants of those they served.

Changing Focus of Evaluation

State involvement in the funding of informal education was, prior to the early 1960s, minimal, or in many localities, nonexistent. In the United States, the type of agencies currently employing informal educators rarely secured significant federal and state funding until the late 1960s and 1970s and it was from that period onward that evaluation gradually became an integral part of service delivery. Although local government funding was more common in the United Kingdom and Europe from an earlier period, formal evaluation again did not begin to play a significant role until approximately the same period when central government started to usurp local government as the main source of subsidies. The growth of state funding reflected, and indeed may have contributed to a decline in the vibrancy of the public domain and the weakening of civil society (Marquand, 2004). For whatever reason, informal socializing, volunteering, and affiliation to membership-based organization has rapidly declined during the last three decades (see Putman, 2000; Smith, 2000). Those organizations that historically generated and sponsored the bulk of informal education waned, and the space they vacated was partially filled by professionals paid directly or indirectly by the state and, to a lesser extent, corporate charity.

At a macro-level, this shift created within all Western industrialized nations what Neave describes as an evaluative state, within which governments employ evaluation as a means of enforcing strategic change and use evaluation techniques to “elicit how far goals have been met, not by setting the prior conditions but by ascertaining the extent to which overall targets have been reached through the evaluation of ‘product’” (Neave, 1988: 9). Within this environment, evaluation *apropos* informal education has shifted from a historic emphasis upon process, the practitioners reflecting on their interaction with learners and the community, toward a focus upon outcomes or product. This has had a profound impact upon practice. In particular, it has meant the informal educator and the citizen are now rarely viewed as partners in dialog. Rather, the latter is perceived as a learner, client, or customer deficient with regard to specific skills or social attributes, while the former is judged in relation to the extent to which the learner and customer acquire new skills and patterns of behavior. As a consequence in recent years, the focus has shifted from formative to summative evaluation.

This change has meant that evaluation in relation to informal education practice has, according to France,

predominately “been imposed, designed and planned by ‘experts’ or ‘managers’” (France, 2001: 277). The function and purpose being:

- to ensure what is being done for, and to, young people, communities, and others has a positive outcome;
- to clarify aims and purposes, set unambiguous objectives concerning the boundaries of the work; and clearly state what the work is, is not, about; and
- to support funding claims and justify external investment.

The result has been that evaluation, which in its original Latin context meant ‘to strengthen,’ has come to mean something else in relation to informal education. It has become a tool of funders rather than being an integral part of educating. The focus has shifted toward counting and comparing rather than an ongoing evaluation that is intrinsic to dialog and conversation.

Three Approaches to Evaluation

Whatever the prime focus, three broad approaches to evaluation are commonly encountered in relation to informal education practice.

Directed Evaluation

In this approach, external agents, usually funders, set criteria. The focus is predominately upon measurable outcomes and outputs. Attention is paid to the number of contacts or conversations, measurable changes in the behavior of those with whom the informal educator works. Comparisons are made between agencies and workers. The format is designed to measure efficiency, effectiveness, and value for money. The result is often the creation of conformity as agencies and workers endeavor to deliver the required outputs and outcomes. In addition, it generates competition, the fabrication of results, and rejection of unprofitable customers in order to sustain funding and maximize income. Some of the most visible signs of the use of this mode of evaluation are the translation of its findings into data for league tables, funding report forms, and inspection reports. Within the last few decades, it has encouraged the imposition of programmatic work directed toward assessed or measured outcomes (Hirsch, 2005) so that those involved in community work, adult education, and youth work have had less space to engage in informal education.

Negotiated Evaluation

Here, the judgments regarding practice are made according to criteria agreed between the different parties involved. Management committees, funders, workers, and participants may all make a contribution toward

deciding what is to be evaluated and why. There is debate and discussion, although not all may be involved in everything. The criteria and the method may be imposed by outside bodies and can also vary from situation to situation. However, they are set out in advance. The methods can also be contiguous to those employed in directed evaluation. Apart from collecting data (about, for example, the numbers using a facility or contacts made on the street), this approach can involve the widespread use of evaluation forms and questionnaires. This model is advocated, for example, by Feuerstein (1988) and Kuriel (1991), both of whom argue that it offers the potential for partnership and engagement with the least powerful members of a group or community.

Dialogical Evaluation

This approach places the prime responsibility for evaluation on educators and participants. Its purpose is to enrich performance and it is integral to practice. Here informal educators seek conversations that focus on issues concerning the value of people's experience and learning. This entails engaging with people to describe experiences, explore meanings, confront issues, and reconstruct practice (Smyth, 1993). As part of the daily round of working, the informal educator encourages people to look at what they have learnt and their experiences of learning. They listen to what others are saying (and not saying) and reflect on their own experiences in order to evaluate what is taking place. The educator will ask questions or make suggestions so that others may develop their capacity to evaluate their own experiences. At the same time, they might invite them to join in making judgments about the work undertaken. For much of the time, this will not be planned and will not make use of formal tools. It will interweave with and develop through conversation and dialog. This means it may take some unraveling – hence the importance of recording on the part of the informal educator and the need for nonmanagerial supervision. More formal activities may be required to stimulate reflection. These can range from the production of annual reports, creation of focus groups, and holding of business meetings. The keynotes are dialog and the exercise of democratic power. This approach involves the making of decisions and an expectation that those who make the evaluative judgments will live with the consequences of the actions that they lead to.

Dialogical evaluation requires not merely clarity regarding how it might be undertaken and the methodology, but also a degree of precision concerning what questions need to be asked. In particular, attention needs to be paid to the following:

- *Interactions.* What are the characteristics of these? What purpose did they serve? What or who initiated them?

To what extent, if any, were they educative? Are they sustained? Did they reflect the sort of values the informal educator is seeking to encourage?

- *Focus.* What issues and topics form the focus for conversation and dialog? Which of these are initiated by the informal educator and which by others? What are the most common subjects or concerns that arise?
- *Setting.* Where is the work being undertaken? What physical settings best stimulate conversation and dialog? What is the impact of setting upon subject matter, the nature of those worked with, and the quality of interaction?
- *Aims.* What was the informal educator aiming to achieve? What were the aims of others involved? What actual or potential conflicts might arise between the two?
- *Strategies.* How did we, as educators, plan to achieve our aims? Who set these, and were they agreed or imposed? How, if at all, were they altered during the course of the interaction and who influenced this? What strategies did others have? How did they change?
- *Outcomes.* Were outcomes set, and if so by whom? What appeared to be the outcomes for different participants? What did we learn from our engagement? Are there questions and issues that now need to be addressed? Who needs to know about these?

By considering these – and how they relate to each other – it becomes possible to begin making judgments relating to the value and experience of informal educational practice interventions, to evaluate the quality and impact of practice and make decisions about future modes of intervention and strategies.

Problems in Evaluating Informal Education

Irrespective of the methodology adopted in relation to the evaluation of informal educational practice, a number of problems emerge. First, the different things that influence the way people behave cannot be easily broken down. For example, an informal educator working with a project to reduce teen crime in two localities might notice that the one with a youth center open every weekday evening has less crime than the area without such provision. But what will this variation, if it even exists, prove given most teen crime remains unreported? It could equally be explained by differences in the ethos of local schools, policing practices, housing quality, unemployment rates or the opportunities that exist for offending in a given neighborhood.

Second those who may be influenced by the work of informal educators are often not easily identified. It may be possible to list those who have been worked with directly

over a period of time. However, much contact is sporadic and may, like Egremont's meeting with the stranger, take the form of a single accidental encounter. The indirect impact is just about impossible to quantify. The efforts of the informal educator may result in significant changes in the lives of people we do not work with. This can happen as those we work with develop. Consider, for example, how we reflect on conversations that others recount to us, or ideas that we acquire second or third hand. Good informal education seeks to achieve a ripple effect to stimulate and contribute to the full and free communication that Dewey (1927: 11) argues is essential for the creation of "the great community." Informal education aspires to encourage learning through conversation, dialog, and example. Indeed, the more successful it is in reaching this goal, the more limited will the awareness of what has been achieved, more irrelevant the attempts to measure that success.

Third, change can rarely be monitored even on an individual basis. For example, informal educators who focus on say alcohol abuse within a particular group can face an insurmountable problem if challenged to provide evidence of success. They will not be able to measure levels of use prior to intervention, during contact, or subsequent to the completion of their work. In the end, all the informal educator will be able to offer, at best, is vague evidence relating to contact and anecdotal material.

Fourth, there is an issue with timescales. Change of the sort with which an informal educator is concerned does not happen overnight. Changes in values and the ways people come to appreciate themselves and others are notoriously hard to identify – especially as they are happening. What may seem ordinary at the time can, with hindsight, be recognized later as special and transformative.

Lastly, we encounter the thorny topic of candor within evaluation. The growth of the audit culture with its concern that everything public servants do must be measured and assessed to ensure value for money has created a cascade of forms and returns the bric-a-brac of a system rooted in a distrust of the professional. These accumulate much information regarding how informal educators spend their time, who they meet, and what they do. It consumes a great deal of time and reduces the proportion of the working day spent on face-to-face work. The result is that much of the data collected is likely to be of dubious value, for it is gathered and collated for others rather than to strengthen the practice of those who provide it and who consequently have no incentive to ensure its authenticity. As Squires and Measor (2005) note, it also frequently fails to offer the level of analysis needed to strengthen practice for such "evaluation does not have the free pursuit of knowledge as its central focus and, by virtue of the fact that it is tied to a particular funder, some of its independence and space to develop criticism is diminished" (2005: 29).

Conclusion

Informal education like all forms of educational practice, is future oriented. It is about stimulating development and growth within individuals and groups. Informal education, like so much education that is relational, is essentially based upon faith and hope (Halpin, 2003). For this reason, those who practice it are as eager as colleagues operating within classrooms and other formal settings to evaluate their practice, to know what forms of intervention are most likely to be effective, and to know what factors are most likely to generate productive and worthwhile educational encounters. Having said that, as Brookfield and Preskill (1999: 170) conclude, it is probably the case that there really are no "standardized protocols or universal measures we can apply to assessing" the competence of those required to foster discussion, debate, and dialog. Informal educators may wish to secure better ways of evaluating their practice but that does not mean that they will be able to secure tools and methods that possess similar accuracy to those available to educators operating in other settings. For Paneth, her prime aim was to offer the young people a "store of happy memories" (Paneth, 1944: 86), something that by definition cannot be effectively measured; likewise Brent (2004), who in an account of working with a group of bereaved young people, identifies one measure of success as being able to see them smile more freely. For Paneth and Brent, like so many informal educators, the more challenging the task, the more pressing the need, the more laudable the aspiration, the more elusive it seems as a mechanism for evaluating success and failure.

See also: Community Based Adult Education; Informal Learning: A Contested Concept; Internal and External Evaluation; Lifelong Learning; Popular Adult Education.

Bibliography

- Brent, J. (2004). Communicating what youth work achieves: The smile and the arch. *Youth and Policy* 69, 69–73.
- Brew, J. M. (1946). *Informal Education: Adventures and Reflections*. London: Faber and Faber.
- Brew, J. M. (1957). *Youth and Youth Groups*. London: Faber and Faber.
- Brookfield, A. D. and Preskill, S. (1999). *Discussion As a Way of Teaching: Tools and Techniques for University Teachers*. Buckingham: Open University Press.
- Brown, R. C. (1956). *A Boys' Club and Delinquency: A Study of the Statistical Incidence of Juvenile in Three Areas in Louisville, Kentucky*. New York: New York University.
- Dewey, J. (1927). *The Public and Its Problems*. New York: Holt.
- Feuerstein, M. (1986). *Partners in Evaluation*. London: Macmillan.
- Follett, M. (1929). *The New State*. New York: Longmans Green.
- France, A. (2001). Evaluation in informal education. In Deer, R. L. and Wolfe, M. (eds.) *Principles and Practice of Informal Education: Learning through Life*, pp 270–288. London: Routledge.
- Gadamer, H.-G. (1979). *Truth and Method*, 2nd edn. London: Sheen and Ward.
- Halpin, D. (2003). *Hope and Education: The Role of the Utopian Imagination*. London: RoutledgeFalmer.

- Hazler, R. J. (1998). *Helping in the Hallways: Advanced Strategies for Enhancing School Relationships*. Thousand Oaks, CA: Corwin Press.
- Hirsch, B. J. (2005). *A Place to Call Home: After School Programs for Urban Youth*. New York: Teachers College Press.
- Jeffs, T. (2001). First lessons: Historical perspectives on informal education. In Deer, R. L. and Wolfe, M. (eds.) *Principles and Practice of Informal Education: Learning through Life*, pp 34–50. London: Routledge.
- Jeffs, T. and Smith, M. K. (1990). *Using Informal Education: An Alternative to Casework, Teaching and Control?* Buckingham: Open University Press.
- Kuriel, C. P. (1991). Participatory evaluation – methodological guidelines based on Indian experience. *Journal of Rural Development* **10**, 225–230.
- Lindeman, E. (1926). *The Meaning of Adult Education*. New York: New Republic.
- Marquand, D. (2004). *Decline of the Public*. Cambridge: Polity Books.
- Neave, G. (1988). On the cultivation of quality, efficiency and enterprise: An overview of recent trends in higher education in Western Europe 1986–88. *European Journal of Education* **23**(1–2), 7–23.
- Paneth, M. (1944). *Branch Street*. London: George Allen and Unwin.
- Putman, R. (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.
- Schon, D. (1983). *The Reflective Practitioner. Towards a New Design for Teaching and Learning in the Professions*. San Francisco, CA: Jossey-Bass.
- Smith, D. H. (2000). *Grassroot Associations*. Thousand Oaks, CA: Sage.
- Smith, M. K. (2001). Josephine Macalister Brew: Youth work and informal education. In Gilchrist, R., Jeffs, T., and Spence, J. (eds.) *Essays in the History of Community and Youth Work*. Leicester: Youth Work Press.
- Smyth, J. (1993). *Teachers as Collaborative Learners*. Buckingham: Open University Press.
- Squires, P. and Measor, L. (2005). Below the decks on the youth justice flagship: The politics of evaluation. In Taylor, D. and Balloch, S. (eds.) *The Politics of Evaluation: Participation and Policy Implementation*, pp 21–40. Bristol: Polity Press.
- Tash, M. J. (2000). *Supervision in Youth Work*. London: YMCA George Williams College.

Further Reading

- Disraeli, B. (1845). *Sybil or the Two Nations*. Oxford: Oxford University Press [Reprinted 1981, Oxford: Oxford University Press].
- Jeffs, T. and Smith, M. K. (2005). *Informal Education: Conversation, Democracy and Learning*. Nottingham: Educational Heretics Press.

Methods to Evaluate Technology

W L Bewley, University of California at Los Angeles, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Formative evaluation – This is conducted during program development to collect information which can be used to improve the program, including the technology, the intervention, and its implementation.

Performance assessment – This requires students to perform a task rather than select a response from a set of provided options. A performance assessment task is designed to be as similar as possible to a task delivered in the intervention, usually including a context and a complex problem to solve. Performance assessments are also known as alternative or authentic assessments.

Portfolios – The alternatives to performance assessments as measures of higher-level knowledge and skills, which include examples of student work, usually with student self-reflections and teacher comments. Scoring is usually done using rubrics.

Qualitative methods – The evaluation methods used to describe and understand the program by contact with the program and participants using observation, interviews, and surveys.

Quantitative methods – The evaluation methods involving measurement of prespecified outcomes or process to show a relationship between program activities and outcomes in order to provide information on program effectiveness.

Quasi-experiments – These have many of the features of experiments except random assignment to experimental and control groups and full control of all variables, for example, the timing of exposure to the experimental treatment.

Random-assignment experiment – This experiment requires random assignment of the unit of treatment application, for example, students, classrooms, or schools, to experimental and control groups in order to achieve equivalent groups in terms of variables not explicitly controlled by the experimenter.

Summative evaluation – This is conducted following implementation to provide information on the program's effectiveness.

Technology evaluation – An evaluation determining the effectiveness of technology for use in a particular context for a particular purpose. Based on systematic collection of information about how a

program operates and/or about the effects it may be having in order to support decisions about the technology application – how to improve it, whether to expand it, or whether to discontinue it.

Theory of change – A description of the causal processes through which change is believed to happen as a result of an intervention's strategies and action.

The goal of technology evaluation is to determine the effectiveness of technology for use in a particular context for a particular purpose. In educational settings, contexts include such things as teacher and student demographics, access to technology resources, local support, and time and budget constraints. Purposes of technology use include instructional applications intended to improve student learning and productivity applications intended to reduce student, teacher, and administrator time and effort. Productivity applications are not unique to education, and there is a large literature on fairly standardized approaches to their evaluation. In this article, we focus on technology evaluations unique to education – evaluation of applications of technology for instructional purposes, which also has an extensive literature but with less agreement on methodologies.

Technology evaluation is important because evaluation results are used to guide or justify decisions that lead to investments of scarce resources on technology-based educational interventions that may have positive effects with varying degrees of cost effectiveness, negative effects, or no effects at all. The results of high-quality evaluations can help make the impacts positive and cost effective. The methods used to evaluate technology are important because they affect the quality of the evaluation. Method selection and design are not easy tasks because technology evaluation is very difficult, for all the reasons educational research is difficult, and there are additional difficulties that come with the use of technology. The effectiveness of an educational intervention is due to a combination of causes, not one, and the causes may interact in complex ways. The instructional experience depends on many variables, including teacher background, teaching philosophy, training, and experience; the support of school management; characteristics of the students; school district; and community context. When technology is part of the experience, there are additional variables, including availability of hardware, software, and technical support;

professional development in technology integration strategies; student prior experience with and expertise in using technology; teacher expertise in technology; home access to technology; technology use in classes other than those in which the intervention is conducted; and teacher skill in implementing the intervention.

This article discusses the current state of technology evaluation methodology. We begin with an overview of three major methodological approaches, the random-assignment experiment, quasi-experiments, and alternatives based on qualitative methods; then we discuss issues which must be addressed by any method. We conclude with a discussion of heuristics for matching methods to situations.

Methodological Approaches

Random-Assignment Experiments

A random-assignment experiment requires random assignment of the unit of treatment application, for example, students, classrooms, or schools, to experimental and control groups. The unit of treatment application is the unit of analysis, and it defines the sample size. Random assignment is required in order to achieve equivalent groups in terms of variables not explicitly controlled by the experimenter. Variables explicitly controlled by the experimenter are the treatment – the introduction of the technology-based intervention – and all measures and procedures that may affect the results. Treatments must be applied within same time period following a written experimental protocol.

The use of random-assignment experiments for educational research has been dominant from as far back as the Thorndike era in the 1920s. More recently, motivated by sponsors of large-scale curriculum-development projects who wanted summative information to guide funding decisions, and by US government policymakers who wanted to ensure accountability, technology evaluations of the 1960s and 1970s emphasized experimental methods using random assignment of students, classrooms, or schools to experimental groups receiving the treatment (technology intervention) and control groups not receiving the treatment. The approach is also specified by the No Child Left Behind Act in the US, which calls for long-term evaluation of the impact of educational technology using scientifically based research and control conditions.

The argument for the use of random-assignment experiments is that they provide better evidence for causal inferences than any other method. This is true, assuming that the conditions required for experiments are met. The difficulty of meeting these conditions has led to strong objections to experiments in education research, including technology evaluations. The key problem is the requirement for random assignment. Schools do not typically assign students to classrooms and teachers randomly, and

students and teachers are not randomly assigned to schools. It is also difficult to meet the requirement for a control group not receiving the treatment. Technology use may not be a legal entitlement, but parents and school boards do not readily accept withholding the use of technology for the sake of an experiment. It may also be the case that technology use in other classes and at home is so widespread that it is difficult or impossible to have a control group with no experience that might be relevant. Many argue that the goal of educational technology is to provide experiences not possible without the technology, which means that it is impossible to have a control group receiving the same experience but without the technology intervention. In addition, the fidelity of the implementation over time is a problem and differential attrition from experimental or control groups can introduce bias.

Another criticism of the experimental approach is that although it provides better evidence for causal inferences, it does not provide information on why the intervention had its effects. The argument is that the experiment is a black box that provides evidence of connections between causes and effects, but does not provide information on the processes inside the box that explain why the intervention caused the effects, many of which are based on the context of the intervention.

Finally, there are the practical problems of cost and time. Experiments are expensive and time consuming. They may require all the funds available for evaluation and take so long to complete that decisions are made before results are available. Whether this is unique to random-assignment experiments is arguable, but it is a common criticism nonetheless.

Quasi-Experiments

Quasi-experiments have many of the features of experiments except random assignment to experimental and control groups and full control of all variables, for example, the timing of exposure to the experimental treatment. One example is the time-series experiment, in which periodic measurements are taken over time and an experimental change is inserted at some point in the time series of measurements. Changes after insertion not only may indicate an effect caused by the experimental change, but may also be caused by other events occurring during the time series because there is no control over events other than the introduction of the experimental change. Another example is the nonequivalent control group design, one of the more common designs in educational research. There are two groups – an experimental group and a control group. Both are given a pretest and a posttest, but only the experimental group receives the experimental treatment between the two tests. This is similar to an experimental design, but students are not randomly assigned to each group. Causation can be inferred if there is an experimental

versus control group difference in the posttest score. As the two groups are naturally assembled, for example, students in two different classrooms, not randomly assigned, they cannot be considered equivalent, and it is possible that some difference affecting the groups other than the experimental treatment could be the cause. Although this may seem unlikely, it is possible. The point is that the evidence from quasi-experiments is not as strong as the evidence from random-assignment experiments, but it is also true that quasi-experiments are usually more feasible and practical in an education setting.

Qualitative Methods

Qualitative methods do not attempt to compare experimental and control groups at all, or to control variables. They investigate the intervention through observation, review of artifacts, and interviews, studying cases in their natural setting to consider variables as they appear in all the complexity of the context. These methods are very popular in education research, including technology evaluation, due in part to the difficulties in doing experimental research in educational settings, and in part to the desire to obtain information on why the intervention had its effects – the processes and mechanisms that lead from specifics of the intervention to effects – and the contextual conditions under which the intervention is more or less effective. The focus is on the context of the technology intervention, for example, local engagement, collaboration, and feedback, and looking at the why of the result. Understanding the why of the result involves developing a theory of change, a description of the processes through which the effects are produced. Qualitative methods are weak on causal inference, but the contextualization makes them very useful to decision makers by providing models (theories of change) describing how and why the intervention works or does not work in the existing system, information needed to decide whether, how, and when to use the technology-based intervention.

Qualitative methods are especially useful for studying a broad range of naturally occurring practices found in many different parts of the education system, not from a particular intervention using a particular technology, which would usually be evaluated with an experiment. Such studies are often descriptive, interested in the frequency of various technology uses and practices, not their effects. Some correlate descriptive data with student outcomes to attempt to identify relationships, if not causes. Concluding anything about causation from correlations is, of course, problematic.

Combined Methods

As is usually the case when there is a debate over the merits of radically different points of view, the practical truth lies somewhere in between. There is no one right

way to do technology evaluation. The approach depends on the purpose of the evaluation, the nature of the intervention, and the context in which it is situated. Some will require quantitative methods, some will require qualitative methods, and usually the evaluation will benefit from a combination providing both quantitative and qualitative data on student learning and attitude outcomes, context, and the implementation of the intervention. A process for deciding when to use what methods and combinations is described in the section titled ‘Matching methods to situations’ at the end of this article.

Issues

There are several important issues which must be addressed by any methodology. These include defining the implementation, maintaining the treatment, and defining the measures.

The Implementation

Designing and developing the technology are only part of the problem. At least as important, if not more important, is the implementation of the technology. Implementing the same technology in a different way is a different intervention. There are two aspects to consider. First, for any intervention, certain conditions must be met for it to be effective. These include things such as the availability of the required technology, teacher and student competency in using the technology, access to the technology, and sufficient time to learn and use the technology. Second, even when conditions for a successful implementation are met, the intervention may not be implemented as planned. How the intervention is implemented depends on the teacher, students, the school facility and technical support, and many other contextual variables. Differences may include how the intervention is introduced, how well it is integrated with the curriculum, how much time is spent on it, and the teacher’s attitude. If implementation differences cannot be avoided, which is usually the case, they should be minimized, and at least recorded to define the actual implementation.

Maintaining the Treatment

For random-assignment experiments and quasi-experiments, the difference between treatment and control groups must be maintained over time. This is difficult due to the problems mentioned earlier, for example, the control group could have access to the technology in other classes or at home, and the school may be reluctant to maintain the treatment-control difference because it could be viewed as depriving some students of equal access. In addition to differences in how the intervention is implemented, there

are typically differences in implementation over time. A teacher may, for example, decide that there is a better way to implement the intervention, eliminating elements that do not work and adding elements that may work better. Teachers want to improve their instruction, but these changes change the intervention.

Defining the Measures

Measures should be considered in five general categories: cost, context, participants, implementation, and outcomes. In addition to information about the effectiveness of a technology-based intervention, decision makers may want information about resources required by the intervention and its cost effectiveness. Cost information includes direct costs of staff time, materials, equipment, and facilities. It may also be useful to measure opportunity cost, the difference between benefits of the evaluated intervention and the benefits that could have been received from an alternative course of action. Context includes variables that may influence the effectiveness of the intervention, such as local support, leadership, communication, and time and budget constraints. Participant characteristics may also influence the effectiveness of the intervention. Characteristics of interest include age, sex, language, background, and school record. As noted above, how the intervention is implemented is extremely important in influencing outcomes because implementation is part of the definition of the intervention.

Outcome measures must be aligned with the intended outcomes of the intervention as implemented. Using an outcome measure not aligned with the intervention will not detect the effects of the intervention. This seems obvious, but many technology evaluations use standardized multiple-choice tests of basic skills as outcome measures for interventions targeting higher-level knowledge and skills such as problem solving. Standardized tests are readily available, are relatively easy to use and inexpensive to administer and score, and are understood and valued by consumers of the evaluation results. They also have known and fairly good technical quality, for example, reliability, which makes them seem a good choice. Without alignment to the intervention, of course, they are not valid, but this is often overlooked when validity is not understood. Outcome measures must tap the entire range of knowledge and skills at the same level of complexity addressed by the intervention, and they must be validated for the purposes and situations to which they are applied. Validity is the key indicator of technical quality, but it is not a general property that applies to all uses for all time. Rather, it depends on the context and inferences to be drawn from the results.

Another popular outcome measure is the opinion survey. If teachers and students like the intervention and think that learning has occurred, then the intervention is a success. Opinion surveys are relatively easy to develop

and administer, and they are understood by information consumers. Again, however, there is a validity problem.

To evaluate technology-based interventions targeting higher-level knowledge and skills, we need measures of the targeted knowledge and skills. For example, if we are interested in conceptual understanding, an essay or knowledge map may be appropriate measures, but for interventions targeting higher-level cognitive processing, measures derived from performance assessments or portfolios may be required. A performance assessment is a task designed to be as similar as possible to a task delivered in the intervention. These tasks usually include a context and a complex problem to solve, for example, designing an experiment to measure the sources of groundwater contamination, and they may be performed in groups, requiring measures of collaboration as well as planning and design. Problem-solving strategies and behaviors in performance assessments can be measured using scoring rubrics.

Another option is using measures embedded in the technology itself. Many technology-based interventions targeting higher-level knowledge and skills are based on simulations. Measures embedded in a simulation enable the unobtrusive measurement of task performance as the task is delivered in the simulation. This can provide valuable information on the problem-solving strategies used by the student as well as measures focused on the outcome of the process, such as a rating of overall success, the number of errors, and time to complete. In tasks performed by manipulating objects on a computer screen, it is possible to record the clickstream – the responses of the examinee in performing the task, usually mouse clicks, with the associated location, time, and task context of the clicks as appropriate. Using the technology to deliver instruction as well as measure the outcomes has a certain elegance, providing both alignment and efficiency of measurement as it does, but there is a danger. Embedded assessments are necessarily specific to the implementation of the intervention, which limits the generality of conclusions and eliminates the possibility of an equivalent control group. A student who has not used the technology-based intervention would probably not have been exposed to the same content, may not have access to the assessments, and if the assessment is accessible, may not know how to use the technology to perform the assessment task.

However they are delivered, because performance assessments are complex and take longer to complete, a limited number can be administered in the time available for collection of data. This may limit the generalizability of the results because, unlike selected response tests that provide equivalent forms, the problem of designing equivalent performance assessment tasks has not been solved. If time is available for only one assessment task, there is uncertainty as to whether performance on a different task thought to require the same knowledge and skills would provide the same results.

Portfolios are alternatives to performance assessments as measures of higher-level knowledge and skills. These include examples of student work, usually with student self-reflections and teacher comments. As with problem-solving strategies and behaviors in performance assessments, scoring rubrics can be used for portfolio measures, but like performance assessments, these measures have typically not had good technical quality. Technical quality is, of course, critical when scores are to be used to make important decisions.

Whatever the measure, studies should try to measure student performance longitudinally, not at a single point in time. There should also be an attempt to define and measure variables that may be alternative causes of the outcomes, and to estimate effects for students with different characteristics and at different levels of the education system, for example, classroom, school, and district.

Matching Methods to Situations

As already noted, there is no one right way to do technology evaluation. What is required is selection from a range of methods, from random-assignment experiments to quasi-experiments to qualitative methods, such as in-depth case studies and ethnographic analysis, the choice depending on the purpose of the evaluation, the nature of the intervention, the context in which it is situated, and practical constraints including site cooperation and available time, finance, equipment, and support resources. The choice need not be limited to a single design. Depending on the purpose, intervention, context, and practical constraints, the appropriate technology evaluation may and usually should consist of a combination of designs.

Figure 1 summarizes a heuristic process for matching evaluation methods to situations and requirements. The major distinction is between investigations of specific projects, evaluations concerned with the impact of a specific technology-based intervention (project-linked), and studies of promising practices, in which the goal is to study existing technology-based practices or access to technology rather than investigate a particular project. Studies of naturally occurring practices obviously require the use of qualitative methods, but when the goal is to identify successful practices, the evaluation should start with a quantitative study to identify successful sites based on some measure, then qualitative methods should be used to understand the differences between successful and unsuccessful sites and the practices related to success.

If the investigation is project-linked, there is a distinction based on the purpose of the evaluation: a formative evaluation to improve the intervention, or a summative evaluation to determine the effectiveness of the intervention. Formative evaluations are used to improve early-stage projects by

collecting information that can be used to guide the development and implementation of the intervention. This requires the use of qualitative methods to provide information on how the technology, the intervention, and the implementation work in the intended setting, used by the intended users. The evaluator will be interested in how features of the environment interact with features of the intervention, and how features of the intervention will influence behavior. To understand the contexts in which the intervention is and is not likely to be effective, the evaluation should look at it in several contexts, for example, classrooms with students from a range of economic backgrounds and students for whom English is not the native language, and schools with minimal resources.

If the purpose of the evaluation is to provide summative information on the intervention's effectiveness, there is another distinction, this one is based on whether causal information is desired. If not, qualitative methods are appropriate. If causal information is desired, quantitative methods are indicated. Random-assignment experiments are best for determining causation and should be considered first, but before selecting an experiment, the evaluator must determine whether the key requirement can be met: it must be possible for students, classrooms, or schools to be randomly assigned to conditions. If the answer is yes, a random-assignment experiment may be possible; if no, a quasi-experiment should be considered.

Before selecting either, however, the feasibility of conducting a random-assignment experiment or quasi-experiment must be determined. Major feasibility criteria are summarized in **Figure 1**. For either experiment type to be feasible, the intervention must be different from standard practice in order to achieve a meaningful comparison, the intervention must be maintainable, participation must not deny students access to an entitlement, human subjects' protection requirements must be met, participants and the site must be willing to cooperate, and funding, equipment, and support resources must be available. These criteria are usually easier to meet when the intervention deals with a relatively small instructional unit and is supported by a particular technology. Feasibility is less likely when the intervention is wide in scope, open ended, and long term, for example, using word processing to support instruction in writing. Different teachers would probably implement such an intervention differently, especially over time. With a long-term intervention, it will be difficult to maintain the composition of experimental and control groups, as students change classes and use the technology for similar purposes in different classes.

If random assignment is possible and feasibility requirements can be met, the evaluator should consider whether there is a requirement for information on conditions of applicability or the process producing the outcomes. If the answer is yes, and this is usually the case, a random-assignment experiment combined

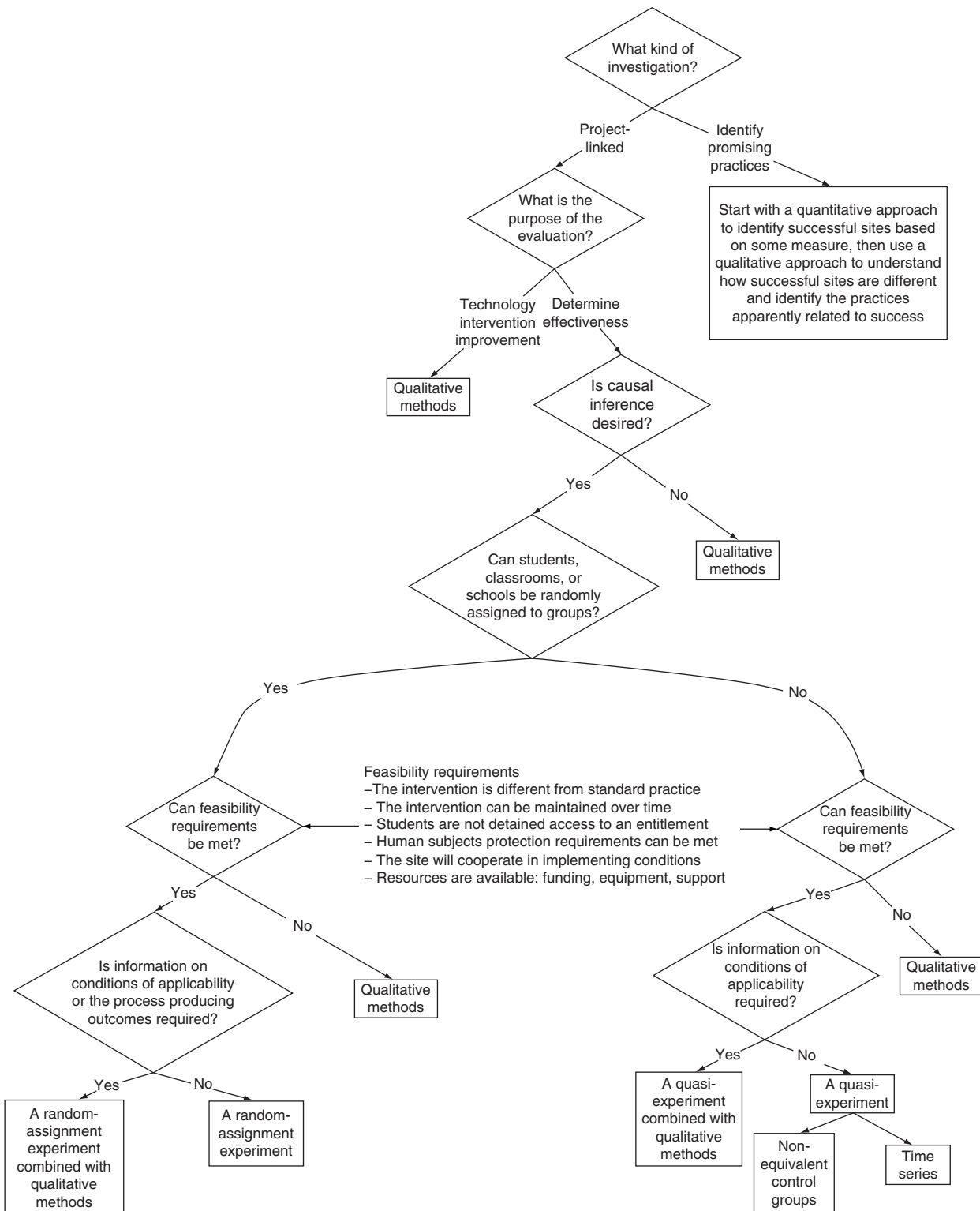


Figure 1 The evaluation method selection process.

with qualitative methods is appropriate. If the answer is no, the random-assignment experiment is sufficient. If feasibility requirements cannot be met, qualitative methods are indicated.

If random assignment is not possible, but feasibility requirements can be met, and there is a requirement for information on conditions of applicability or the process producing the outcomes, a quasi-experiment combined

with qualitative methods would be appropriate. If there is no requirement for conditions of applicability or process, which is not usually the case, a quasi-experiment is appropriate. As with the random-assignment experiment branch of the method-selection process, if the quasi-experiment or quasi-experiment/qualitative method combination is not appropriate, qualitative methods are the choice.

See also: An Overview of Technology and Learning; Classroom uses of Technology to Manage Instruction; Cost Analysis in Evaluation Studies; Curriculum Evaluation: Approaches and Methodologies; Educational Measurement: Overview; Effective Use of Technology in Teaching and Learning in HE; Evaluating E-Learning; Evaluation Methodology; Examining the Effects of Educational Technology Programs: Challenges and Strategies; Field Experimentation in Education; Focus Groups; Intelligent Systems; Internal Validity; Internet-based Education; Methods for Approximating Random Assignment: Regression Discontinuity and Propensity Scores; Mixed Methods; Multitrait–Multimethod Designs; Multivariate Longitudinal Data Analysis; Needs Assessment in Education; Sampling; Survey Research Methods; Technology and Formative Assessment; Technology for Large-Scale Assessment; Technology Supports for Assessment Design; The Quality of Evidence in Qualitative Research; Validity: Mapping Diverse Perspectives.

Further Reading

- Baker, E. L. (1999). *Technology: How Do We Know It Works?* Center for Applied Research in Educational Technology. caret.iste.org/index.cfm/index.cfm?fuseaction=studySummary&StudyID=395&resultSet (accessed August 2009).
- Baker, E. L. and O'Neil, H. F. Jr. (eds.) (1994). *Technology Assessment in Education and Training*. Hillsdale, NJ: Erlbaum.

- Clark, R. E. (ed.) (2001). *Learning from Media: Arguments, Analysis and Evidence*. Greenwich, CT: Information Age.
- Cuban, L. (1986). *Teachers and Machines: The Classroom Use of Technology since the 1920s*. Cambridge, MA: Harvard University Press.
- Cuban, L. (2001). *Oversold and Underused: Computers in the Classroom*. Cambridge, MA: Harvard University Press.
- Haertel, G. D. and Means, B. (eds.) (2003). *Evaluating Educational Technology: Effective Research Designs for Improving Learning*. New York: Teachers College Press.
- Heinecke, W. F. and Biasi, L. (2001). *Methods of Evaluating Educational Technology*. Greenwich, CT: Information Age.
- Herman, J. L., Morris, L. L., and Fitz-Gibbon, C. T. (1987). *Evaluator's Handbook*. Newbury Park, CA: Sage.
- Kozma, R. (ed.) (2003). *Technology, Innovation, and Educational Change: A Global Perspective*. Eugene, OR: International Society for Technology in Education.
- Means, G. and Haertel, G. D. (eds.) (2004). *Using Technology Evaluation to Enhance Student Learning*. New York: Teachers College Press.
- O'Neil, H. F. Jr. and Baker, E. L. (eds.) (1994). *Technology Assessment in Software Applications*. Hillsdale, NJ: Erlbaum.
- Oppenheimer, T. (2003). *The Flickering Mind: The False Promise of Technology in the Classroom and How Learning Can Be Saved*. New York: Random House.
- Roblyer, M. D., Castine, W., and King, F. J. (1988). *Assessing the Impact of Computer-Based Instruction: A Review of Recent Research*. New York: Haworth Press.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton-Mifflin.
- Stoll, C. (1999). *High Tech Heretic: Why Computers Don't Belong in the Classroom and Other Reflections by a Computer Contrarian*. New York: Doubleday.

Relevant Websites

- <http://ctl.sri.com> – SRI International, Center for Technology in Learning.
- <http://www.cse.ucla.edu> – The Center for Research on Evaluation, Standards, and Student Testing.
- <http://www.iste.org> – The International Society for Technology in Education, Research and Evaluation.

EDUCATIONAL EVALUATION – CONTEXTS OF EDUCATIONAL EVALUATION

Contents

Cost Analysis in Evaluation Studies

Cultural Issues that Can Affect the Validity of Educational Evaluations

Evaluation and Accountability

Evaluation Governance and Planning

Evaluation Use

Moral and Ethical Issues in Evaluation

The Role of Stakeholders in Educational Evaluation

Cost Analysis in Evaluation Studies

J D Fletcher, Institute for Defense Analyses, Alexandria, VA, USA

© 2010 Elsevier Ltd. All rights reserved.

A core function of evaluation is to inform decisions intended to improve the effectiveness and/or productivity of an enterprise, policy, or program. The hallmark of such decision-making is not just identifying promising opportunities but determining what must be given up to take advantage of them. What must be given up may be defined as cost. Cost may be measured in monetary terms or not, but the most accessible measures of cost are usually monetary (e.g., dollars, pounds, and euros). Monetary cost, therefore, is frequently a key factor in education decision-making, determining whether or not to proceed with some course of action. Cost analysis provides a way to include costs in assessing the net value of a course of action. This article discusses cost analysis as a vital component of education evaluation.

Cost analysis may be called, among other things, efficiency assessment, cost-benefit analysis, return on investment, economic analysis, cost-effectiveness analysis, and cost-utility analysis. There are differences among these analyses, but cost analysis serves as well as anything else as an overarching, inclusive term for them.

The basic idea behind cost analysis is the well-established and straightforward notion that there is no free lunch. The cost analyst looks for opportunities and resources that must be sacrificed to take any course of action. In this way, the cost analyst pursues a marginalist or incrementalist approach. The analyst must determine how much is gained and how much is given up by taking, or continuing to take, some action.

Cost analyses are as subject to controversy as other assessments. Differences in data, data definitions, analysis techniques, models, and assumptions are as likely to be

found among cost analyses as elsewhere. Education decision-makers would be well served by the adoption of generally accepted standard practices for cost-element definitions, data models, and analysis techniques. A variety of commentators have provided a basis for such standards. However, more needs to be done to standardize the practice of cost analysis in education. The best that can be done at present in performing education cost analysis is to be sufficiently explicit to allow decision-makers to determine for themselves and their objectives how well and to what extent any such analysis informs their decisions.

Interest in cost analysis, which began in the manufacturing sector and has now worked its way through the service, healthcare, and public sectors, has arrived as an assessment factor in industrial training (Phillips, 2003) and military training (Orlansky and String, 1983; Fletcher and Chatham, in press); however, it has yet to catch on as an essential component in education evaluation studies (Hummel-Rossi and Ashdown, 2002; Levin, 2001; Rice, 1997; Ross *et al.*, 2007). This need seems particularly urgent, as expenditures for education compete with other public needs that now defend their budgets with cost analyses. Rice described this neglect of cost in education research as a paradox, wondering “why such a seemingly relevant form of analysis has been so underutilized in the field of education” (p. 309). Ross *et al.* (2007) argue convincingly that the neglect continues.

The goals of education, which largely focus on the learning, development, and socialization of young members of society, seem particularly vulnerable in competition with hard-headed analyses of programs in areas such as food safety, transportation, health, and national defense. It is notable that the first myth listed by Thompson (1990)

in his discussion of program evaluation is that many program effects are inherently unmeasurable and unquantifiable. He declares these myths to be false. Both history and more recent practice support his position.

For instance, as early as 1667, public health officials in London defended their expenditures by calculating a benefits-to-cost ratio of 84:1 for investments in combating the plague (Thompson, 1980). More recently, the 1973 Principles and Standards for Planning Water and Related Land Resources, required analyses of trade-offs between national economic development, regional economic development, environmental quality, and social well-being (Sassone and Schaffer, 1978). Accountability for Great Society programs in the 1960s was required by the Planning, Programming, Budgeting System, which was specifically implemented for these programs and now remains a core requirement for defending today's public expenditures in many areas. The continuing emphasis on accountability and on determining quantitative relationships between public investments and their returns suggests the importance of cost assessment in education.

In education, the investment component is more amenable to cost analysis than the return. We can, for instance, determine the costs of lengthening the school day, installing computers and education software, or providing after-school professional tutoring more easily than we can determine the cost value returned to students and their society by improvements in their learning, but it is by no means impossible. For example, an extensive survey by Ross *et al.* (2007) identified 31 cost studies in education that provided an appropriate design for assessing effects, data on costs and benefits, and defensible analyses relating costs and benefits. The studies in their analysis produced credible assessments of the effects of education expenditures on student achievement, private and social educational attainments, the cost-effectiveness of single programs, and the cost-effectiveness of competing program alternatives.

Investment costs generally fall into four major categories: research and development; initial investment; operations and maintenance; and disposal and salvage (Mishan and Quah, 1988). Research and development costs consist of the materials, people, and facilities needed to create, test, and evaluate a capability in which people might invest. Initial investment costs cover the one-time costs of procuring and deploying the resources needed to implement for operational use what research and development has created. Operations and maintenance costs cover those needed to manage, operate, support, and maintain what has been created after it has been implemented. Disposal and salvage costs are the one-time costs of removing what has been implemented from operational use. Cost analyses intended to inform decisions on courses of action usually omit research and development costs, and disposal and salvage costs, leaving as their basis initial investment and operations and maintenance costs.

As difficult as assessment of investment costs may be, returns from improved efficiency and productivity in education seem far more multivariate and difficult to quantify using any metrics, let alone monetary values. Assessments of return from enhanced capacities for innovation, entrepreneurial vigor, a better-informed and more-literate citizenry, economic health, national morale, and so on and so forth turn out to be more difficult, controversial, and resistant to quantifiable measurement than the assessments of investment costs. This difference may account for the preference in education on cost-effectiveness analyses over benefit-cost analyses, which differ in ways discussed below.

Issues in Cost Analysis

Generally, there are two primary issues to consider in cost analysis: the cost model to use and the choice between benefit-cost analysis and cost-effectiveness approaches.

Cost Model

The cost model is the foundation for cost analysis. It identifies, lists, and defines the cost elements that will be included in any analysis that might be performed. As suggested above, specificity and explicitness are critically important for cost analyses. It is, obviously, important for decision-makers and the analysts themselves to both know and be able to articulate what they are talking about. What is not included in a cost element should be as clear as what is included.

Early on, Levin (1983) suggested five classes of elements, or ingredients, to be considered in a cost model. These are: personnel, facilities, equipment and materials, other program inputs, and client inputs. Personnel costs include all the resources required for the human resources needed by the approach. Levin recommended that personnel be classified according to their roles (instructional, administration, clerical, etc.), qualifications (training, experience, specialized skill, etc.), and time commitments (full time, part time, etc.). The costs of facilities include all resources required to provide physical space for the approach. Equipment and materials include furnishings, instructional equipment, and supplies.

Other inputs in Levin's scheme include components that do not fit elsewhere, for instance, instructor training and insurance costs. Client inputs include resources that must be contributed by students and/or their employers. These inputs are especially relevant in military and industrial training where student pay and allowances may be provided by the client, thereby increasing interest in the speed with which students finish the instruction. Much of the rationale for applying technology in industrial and military training is, therefore, keyed to its capacities for self-pacing and the earlier release of students for duty

	Personnel	Facilities	Equipment	Materials	Indirect costs
Analysis					
Design					
Development					
Implementation					
Evaluation					

Figure 1 A cost-model framework for educational evaluations. Reproduced from Kearsley, G. (1982). *Costs, Benefits and Productivity in Training Systems*. Reading, MA: Addison-Wesley.

(Fletcher, 2004). Both other and client inputs are typically treated together in other costing models as indirect costs, and they are shown that way in **Figure 1**.

About the same time, Kearsley (1982) developed a similar model with one dimension very much like Levin's but with an added dimension keyed to the stages of instruction system development: analysis, design, development, implantation, and evaluation. Although Kearsley's approach is intended for use in training, this second dimension with its instructional elements is readily adapted for education. Its elements should be familiar to most educators. Analysis determines educational objectives – what should be taught to what students so that they attain specified levels of knowledge and skill. Design concerns the selection and application of educational approaches that will most reliably and efficiently accomplish these educational objectives. Development covers the production of the educational materials to design specifications. Evaluation concerns the ability of the materials to produce students who can demonstrate that they have achieved the targeted objectives. Integrated, these two models yield the cost framework shown in **Figure 1**.

The components for each dimension seem appropriate for use as a framework in an educational cost model, but in neither case are the definitions presented here sufficient for the full development of an adequate cost model. Readers are encouraged to turn to Fitzpatrick *et al.* (2003), Kearsley (1982), Levin and McEwan (2007), and/or McDavid and Hawthorn (2005) for a more comprehensive discussion.

Benefit–Cost Analysis

Benefit–cost analysis is used to determine if the benefits returned by some course of action outweigh the costs of investing in it. The calculation of benefit–cost ratios is fairly simple. It reduces all costs of an action to a single unit. It does the same for all benefits. The ratio of benefits to costs is then calculated.

We can calculate a benefits–cost ratio using whatever metrics we choose, but the terms for input and output must be commensurable – both must be assessed using the same

units of measure. Monetary units tend to be those most readily translated from whatever investment resources are required and whatever returns are produced. For that reason, these ratios are almost always expressed in terms such as dollars, pounds, euros, or whatever communicates most easily and usefully to likely decision-makers.

As discussed by Phillips (2003), among others, a benefit–cost ratio is calculated as:

$$\frac{\text{Value of the result}}{\text{Cost of the investment}}$$

It tells us, in quantified terms, how many units of value we get for every unit of cost.

Return on investment is closely related to benefit–cost ratios. It, too, is a ratio, and calculating it is as straightforward as its name suggests. Again as discussed by Phillips (2003) among others, it is:

$$\frac{\text{Value of the result} - \text{Cost of the investment}}{\text{Cost of the investment}}$$

This ratio may be multiplied by 100, so it can be expressed as a percent rather than a proportion. Left as a proportion, any return on investment >1.0 indicates a positive net return. Return on investment must be calculated for some period of time, such as a year. As with monetary units, the length of time covered should be determined by analysts in consultation with decision-makers who are likely to use the results of the analysis.

As mentioned above, both analyses require value and cost to be commensurable. Of the two, return on investment may be preferred because it indicates how many units of net benefits are returned, after investment costs have been subtracted, for each unit invested. There are, of course, spikes, dips, and diminishing returns to be considered with differently timed units of investment, so averaging and curve smoothing may be required.

Neither ratio provides a one-size-fits-all analysis, and neither is likely to satisfy the needs of all decision-makers who turn to it for assistance. Again, what is done for the analysis can, and should, be made as explicit as possible based on a foundation of well-formed and explicated cost models, so that decision-makers receive sufficient

information to determine for themselves the degree to which the analysis is relevant, credible, and able to inform whatever decision they must make.

The issues that arise with the investment side of return on investment analyses usually concern what cost model should be used, what cost elements from it should be included, how specifically to define them, and what values should be assigned to parameters such as discount, interest, depreciation, inflation, and amortization rates. These are not simple issues, but they are easier to resolve and assign costs to than finding data to quantify returns arising from the units of enhanced human performance. For these reasons, we may turn to the cost-effectiveness analysis.

Cost-Effectiveness Analysis

When commensurability is difficult, cost-effectiveness analysis is used. The costs of investment can usually be expressed in monetary units, but the full return, the benefits, of education may be impossible to assess fully in monetary units, despite heroic attempts by economists to do so. Cost-effectiveness analysis allows benefits, such as information retention, job knowledge, and motivation of workers, supervisor ratings, and productivity and effectiveness of the client organization, to be measured in their own units and incorporated in the analysis. It allows the analysis to cover a more complete range of educational outcomes.

Cost-effectiveness is expressed in a way that is directly analogous to benefit-cost ratios. Effectiveness may be measured in units other than those used for cost, so net effectiveness is not measured as it is in assessing the return on investment. Instead, cost-effectiveness is calculated as the ratio of effectiveness divided by the cost of investment – as a direct ratio of benefits to cost. For these reasons, it is common practice in determining cost-effectiveness to hold either costs or effectiveness constant across all the alternatives being considered and observe variations in effectiveness or costs, respectively. Sometimes, either costs or effectiveness is simply assumed to be constant across the alternatives. One could argue that cost is implicitly assumed to be constant by its absence from many educational evaluations. Either assumption may be reasonable, but analysts should present data or information to support it so that decision-makers can decide for themselves if it is warranted.

Cost-effectiveness is a relative term, and relevant decision alternatives must be specified in assessing it. Despite common usage, we cannot properly say that an investment, all by itself, is or is not cost-effective, although there is no harm in calculating a cost-effectiveness ratio for it. An example of cost-effectiveness analysis was provided by Fletcher *et al.* (1990) who calculated the cost-effectiveness of several different education interventions (decreasing class size, lengthening the school day, providing professional tutors, providing peer tutors, and using

computer-assisted instruction) in raising mathematics scores one standard deviation on a standard test of mathematics. Additional examples and more extensive discussions are presented by Ross *et al.* (2007).

The addition of an alternative for achieving the objective(s) after a cost-effectiveness analysis is completed may change its conclusions and recommendations. If a single course of action is to be selected based on cost-effectiveness analysis, the set of alternatives must be as well defined and comprehensive as possible.

One form of cost-effectiveness analysis is cost-utility analysis where the return is assessed in terms of utility or value received by the beneficiaries of the investment. Cost-utility analysis is frequently urged but rarely used in sectors other than health services, where there is significant attention paid to quality-adjusted life years (Drummond and McGuire, 2001). In health services, these analyses are often used to trade-off expected years of additional life for a patient with the quality of his/her life during these years. A measure of expected quality of life is multiplied by the number of additional years of life expected under two or more treatment regimens. Ratios of these measures for different regimens help patients and their medical advisors assess the net benefit or the utility of different regimens for the patient.

Other examples of cost-utility analysis in healthcare are provided by Franke's thorough review of health programs in schools. He cites and discusses many studies, a number of which include costs, that assessed utility-benefits as access to, frequency of, and satisfaction with the mental and physical health services (including substance abuse and sexual activity advice) students received from school-based programs.

An analogous approach might be taken with regard to utility-related benefits of schooling and education. Judgments of quality-of-life improvements obtained from education might be balanced against the opportunity and actual costs of the years of education required to achieve them. Students and their advisors might assess the utility of different programs of education by calculating the ratios of these measures. Levin, starting in 1983, while cautioning against the subject nature of cost-utility analysis, has long advocated it as a component of education evaluation, particularly as a way to assess a wide range of alternatives in a short period of time. Nonetheless, it remains rarely used (Ross *et al.*, 2007).

Considerations in Performing Cost Analysis

A number of discussions suggest steps to use in performing cost analysis (e.g., Barnett, 1993; Fitzpatrick *et al.*, 2003; Hummel-Rossi and Ashdown, 2002; McDavid

and Hawthorn, 2005). Integrating their comments produces the following comments on performing the analysis.

Identifying the Objectives

An analysis must be performed to identify and define what objectives the course or courses of action are supposed to achieve. These objectives may be expressed in terms of what students know or can do once they finish a course of learning, but other objectives and considerations may be in play. Some objectives of significant concern to decision-makers may be implicit. Their inclusion in a cost analysis is a matter for discussion and consultation between analysts and potential decision-makers. Once identified and defined, the objectives then provide the basis for whatever analysis will be performed. As usual, they must be clearly identified and understood by all concerned.

Identifying Requirements for Scale

On a macro level, the education alternative must be able to produce the required number of graduates, given the constraints of time and budget, all of which can affect the outcomes of analysis. Different decision-makers will have different scales of application in mind. The scale of application assumed by the analysis in terms of time, budget, and productivity must be defined and explicated.

Identifying the Cost Model to Be Used

The definition of the cost model and its elements is, as emphasized above, a critical element in any cost analysis. These definitions should be as explicit and unambiguous as possible. Moreover, there may be important differences between planned budget expenditures and the actual costs. The actual costs should be used wherever possible.

For a Cost-Effectiveness Analysis, Identifying the Alternatives

Although a benefit-cost analysis can stand alone, cost-effectiveness analyses must include a realistic and comprehensive set of alternatives. Once the objectives along with the scale of application and the cost model have all been identified for a cost-effectiveness analysis, the alternative approaches that may reasonably be considered for satisfying them should be identified and defined. Generating these alternatives is a critical activity that may require imagination and creativity. It is, of course, unreasonable to include every conceivable alternative in the analysis. As with most cost-analysis issues, the set of alternatives to be considered is a matter for discussion and consultation between analysts and the decision-makers who may consult the findings of the analysis.

Designing the Analysis

Cost analyses may involve an experiment specifically designed to collect data for the assessment (e.g., Barnett, 1985; Fletcher *et al.*, 1990; Kreuger, 2003), or an analysis that takes advantage of already-collected empirical data (e.g., Levin *et al.*, 1985; Borman and Hewes, 2002). Of course, as Hummel-Rossi and Ashdown (2002) advise, the analysis should incorporate “rigorous experimental or quasi-experimental design with attention to . . . hidden and/or qualitative outcomes, and positive as well as negative outcomes” (p. 20).

A test of the adequacy of measures selected in a cost-effectiveness analysis is to determine whether an instructional alternative could excel in most of the measures and still not be best on some intuitive level. This result might occur, for instance, if cost-utility measures are absent. Cost analysis of different approaches to health and sex education might clearly point to programs that would be unacceptable in school districts with religiously conservative parents. In these cases, the assessments may need to be augmented with additional measures.

For a cost-effectiveness analysis, a fixed cost- or fixed-effectiveness approach should be selected. The distinction between education and training may affect the choice. In training, which is a means to an end, the instructional objectives are primarily concerned with producing given and specifiable levels of proficiency, and it may be wiser to choose fixed effectiveness and seek an alternative that achieves that level with minimum cost. In education, which is an end in its own right, decision-makers may be more interested in maximizing achievement. In addition, time and budgets are more likely to be fixed at the beginning of a program of education, so that a fixed-cost approach may be more appropriate for education.

Finally, Kazanowski's (1968) classic list of pitfalls still seems relevant. Four of these, the ratio fallacy, the quantification fallacy, the interrelationship fallacy, and the neglect of spillover effects, are illustrative and seem particularly relevant in education evaluation.

The ratio fallacy involves scaling of the course of action being considered. If the analysis deals with costs that are small in magnitude but the decision-maker is contemplating a sizable implementation, then the magnitude of costs needed to support that implementation must be taken into account. Benefits, effectiveness, and ratios aside, some investments are simply too large to be realistic. The quantification fallacy reminds both the analyst and the decision-maker that it is never safe to assume that everything relevant to the analysis can be quantified. The interrelationship fallacy involves the assumption that the relationship between variables is monotonic and linear. Particularly for educational variables, these relationships may be neither. The spillover effects concern interactions

among systems. A course of action for one system may well, if not inevitably, affect the operation of other systems that are either presently in place or likely to be implemented in the future. Such interactions may be dominant in decision-making and should be included both in designing the analysis and reporting its results.

Performing the Analysis

The analysis should employ data drawn from sufficient sample sizes, designs appropriate for an experimental, quasi-experimental, or analytical approach, and assess both short-term and long-term effects. It should be performed in a manner appropriate for a benefit–cost or cost-effectiveness study. In short, it should be as methodologically correct as time, budget, and circumstances allow.

Performing a Sensitivity Analysis

In any analysis, assumptions must be made about the values assigned to some of its components. Will a piece of equipment last 5 or 10 years? What will the rate of inflation be over the period covered by the analysis? Will 50, 500, or 5000 students use an instructional program or programs under consideration? How frequently will the instructional materials have to be updated? And so on. The outcome of a cost-effectiveness evaluation will be more sensitive to some of the assumptions on which it is based than others. The sensitivity analysis is intended to identify these differences so that decision-makers can see for themselves how robust the analysis is with regard to variations from its underlying assumptions. The analysis may best be designed as a model into which different assumptions or parameters can be substituted. In addition, what components are included in a cost analysis is always an issue. A solution is to assume some base set of elements and then to use sensitivity analysis to help decide what must be added and what may be excluded.

Reporting the Results

When the data are gathered, they should be tabulated in a form suitable for comparison. It is critical for decision-makers to know the strengths and limitations of an analysis. It is not a trivial task to identify the assumptions underlying an analysis, but these assumptions should be identified and described as explicitly as possible. The underlying models of cost and effectiveness should be documented so that decision-makers can see what has been excluded and, therefore, assumed either irrelevant, less important, or treated as equivalent across alternatives in cost-effectiveness analyses. All the cautions concerning clear, unambiguous, and explicit definitions and descriptions of the data model and its elements continue, of course, apply.

If cost data are drawn from different periods of time, they should be adjusted so that they are equivalent. In these cases, cost data reporting should take account of inflation and be reported in year-constant units. If the cost data results are extrapolated into the future, proper discounting should be used so that costs can be reported in whatever year-constant units are selected. Capital equipment (e.g., computers, desks, and texts) should be realistically amortized to account for their different life cycles of use. Fitzpatrick *et al.* (2003), Kearsley (1982), Levin and McEwan (2001), McDavid and Hawthorn (2005), Mishan and Quah (2007), and Phillips (2003), among others, may be consulted for more specific detail and advice about these matters.

Decision-makers may view a tabulated array of findings more as raw data than useful information. As a rule, decision-makers need information first, then the data to support it. Discussion of the findings, the strengths and weaknesses of the data-collection procedures, limitations of the assessment, and presentation of the alternatives in light of the evaluation, should be in order and should be provided.

Conclusion

Phillips (2003) points out that cost analyses allow behavioral scientists to join a transformation begun in the manufacturing sector but that is now arriving at military and industrial training. Education may not be far behind. This progression will transform education evaluation from reactive efforts starting only after some course of action has been taken or some program has been fielded, to a more proactive, integrated activity that begins much earlier in the decision, development, and implementation cycle. Such a transformation will more closely link program effects to specific requirements and objectives, more strongly emphasize not just doing things right but doing the right things, more effectively communicate choices, costs, and effects to decision-makers, help set priorities, and, overall, focus program reporting on the output and results.

See also: Evaluation of Integrated Health Programs in School.

Bibliography

- Barnett, S. W. (1985). Benefit–cost analysis of the Perry Preschool Program and its policy implications. *Future of Children* 3, 93–112.
- Barnett, S. W. (1993). The economic evaluation of home visiting programs. *Future of Children Home Visiting* 3, 93–112.
- Borman, G. D. and Hewes, G. M. (2002). The long-term effects and cost-effectiveness of success for all. *Educational Evaluation and Policy Analysis* 24, 243–266.

- Drummond, M. and McGuire, A. (2001). *Economic Evaluation in Health Care: Merging Theory with Practice*. Oxford, UK: Oxford University Press.
- Fitzpatrick, J. L., Sanders, J. R., and Worthen, B. R. (2003). *Program Evaluation: Alternative Approaches and Practical Guidelines*, 3rd edn. New York: Longman.
- Fletcher, J. D. (2004). Technology, the Columbus effect, and the third revolution in learning. In Rabinowitz, M., Blumberg, F. C., and Everson, H. (eds.) *The Design of Instruction and Evaluation: Affordances of Using Media and Technology*, pp 139–157. Mahwah, NJ: Erlbaum.
- Fletcher, J. D. and Chatham, R. E. (in press). Measuring return on investment in training and human performance. In Cohn, J. V. and O'Connor, P. E. (eds.) *Performance Enhancement in High Risk Environments*. Westport, CT: Praeger.
- Fletcher, J. D., Hawley, D. E., and Piele, P. K. (1990). Costs, effects, and utility of microcomputer assisted instruction in the classroom. *American Educational Research Journal* **27**, 783–806.
- Hummel-Rossi, B. and Ashdown, J. (2002). The state of cost-benefit and cost effectiveness analyses in education. *Review of Educational Research* **72**, 1–30.
- Kazanowski, A. D. (1968). Cost effectiveness fallacies and misconceptions revisited. In English, J. M. (ed.) *Cost-Effectiveness: The Economic Evaluation of Engineered Systems*, pp 151–165. New York: Wiley.
- Kearsley, G. (1982). *Costs, Benefits, & Productivity in Training Systems*. Reading, MA: Addison-Wesley.
- Kreuger, A. B. (2003). Economic considerations and class size. *Economic Journal* **113**, 35–53.
- Levin, H. A. (1983). *Cost-Effectiveness: A Primer*. Beverly Hills, CA: Sage.
- Levin, H. M. (2001). Waiting for Godot: Cost-effectiveness analysis in education. *New Direction for Program Evaluation* **90**, 55–68.
- Levin, H. M. and McEwan, P. J. (2001). *Cost-Effectiveness Analysis*. Thousand Oaks, CA: Sage.
- McDavid, J. C. and Hawthorn, L. R. L. (2005). *Program Evaluation: Separating the Forest from the Trees*. Thousand Oaks, CA: Sage.
- Mishan, E. J. and Quah, E. (2007). *Cost Benefit Analysis*. London: Routledge.
- Phillips, J. J. (2003). *Return on Investment in Training and Performance Improvement Programs*, 2nd edn. Oxford, UK: Butterworth-Heinemann.
- Rice, J. K. (1997). Cost analysis in education: Paradox and possibility. *Educational Evaluation and Policy Analysis* **19**, 309–317.
- Ross, J. A., Barkaoui, K., and Scott, G. (2007). Evaluations that consider the cost of educational programs: The contribution of high quality studies. *American Journal of Evaluation* **28**, 477–492.
- Sassone, P. G. and Schaffer, W. A. (1978). *Cost-Benefit Analysis: A Handbook*. New York: Academic Press.
- Thompson, M. S. (1990). *Benefit–Cost Analysis for Program Evaluation*. Thousand Oaks, CA: Sage.
- Harris, D. N. (2008). Toward policy-relevant benchmarks for interpreting effect sizes: Combining effects with costs. *Educational Evaluation and Policy Analysis* **31**, 3–29.
- Kettner, P. M., Moroney, R. M., and Martin, L. L. (1990). *Designing and Managing Programs: An Effectiveness-Based Approach*. Thousand Oaks, CA: Sage.
- Selameab, T. and Yeh, S. S. (2008). Evaluating intangible outcomes: Using multiattribute utility analysis to compare the benefits and costs of social programs. *American Journal of Evaluation* **29**, 301–316.
- Tang, M. C. (1997). Cost analysis for improved educational policymaking and evaluation. *Educational Evaluation and Policy Analysis* **19**, 318–324.
- Weimer, D. L. and Vining, A. R. (1992). *Policy Analysis: Concepts and Practice*. Englewood Cliffs, NJ: Prentice Hall.

Further Reading

Cultural Issues that Can Affect the Validity of Educational Evaluations

R P Durán, University of California Santa Barbara, Santa Barbara, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Activity system – The term drawn from cultural psychology and cultural historical activity theory refers to the organization of human action in everyday settings to accomplish goals associated with the settings. Attention is given in activity system analysis to the ways that participants in activity settings identify the “who,” “where,” “when,” “what,” “how,” and “why” tied to their mutual pursuit and accomplishments of goals.

Critical pedagogy – Critical pedagogy developed from the work of the Brazilian educator Paulo Freire. This approach focuses on the social empowerment of communities that are underserved by society through reliance on the cultural, linguistic, and social assets of community members.

Culture – Culture used in the present article is not a background trait of individuals or a trait inherited from their national or ethnic communities of origin. While drawing on these background influences, culture instead is viewed as a continuously constructed and negotiated way of seeing and acting on the world to in the moment-to-moment pursuit of activity.

Cultural psychology – Sometimes also referred to as ‘cultural-historical activity theory’ by some investigators, addresses the study of human action and its development in specific or local social, institutional, and historical circumstances.

Culturally responsive evaluation – Concerns the importance of making evaluation of educational interventions and strategies (and evaluation more generally) sensitive to the cultural values and practices of community participants from diverse cultural, linguistic, and social backgrounds in the context of everyday life.

Ecological systems theory – Ecological systems theory draws from the work of the developmental psychologist Uri Bronfenbrenner. This theory examines complex ways in which different ecologies surrounding human action such as family, school, community, geographical, policy interact and influence each other and the psychosocial development of humans.

Ecology – A general term which in this article refers to the social, institutional, community, and policy environments or settings surrounding human action.

Generative learning – A theory of human learning introduced by Merlin Wittrock. The theory asserts that humans learn most effectively by actively generating conceptual connections between new knowledge and prior knowledge. Asking learners to communicate new knowledge to others is an effective way to support generative learning.

Kamehameha Early Education Program (KEEP) – The name of an innovative language arts program developed in the 1970s in to serve low income Hawaiian origin children . The program was designed to draw on the home and community cultural, social, and linguistic knowledge of participants as a resource to support reading and comprehension skills development in classroom settings.

Participatory action research – An approach to research on social and educational problems that frames research in terms of questions, issues, evidence, and perspectives of community members most affected by the outcomes of research.

Photovoice – A research technique employed by community members to document and analyze problems in their communities by use of photography, narrative, and expository text for the purpose of undertaking grassroots social action.

Subtractive schooling – A concept introduced by Angela Valenzuela to refer to analyze negative schooling conditions faced by immigrant and US born youths from minority backgrounds who perform poorly in classrooms because of lack of teachers’, school staff’s and community members’ respect for their cultural, linguistic, and social knowledge and lack of understanding of how to apply these resources to improve schooling outcomes.

Youth Empowerment Strategies (YES!) Program – An intervention program implementing the Photovoice technique implemented with inner city youths to undertake identification and solution of local community social, health, and other problems.

Introduction

Educational interventions are often set in the context of a need to solve learning and achievement challenges faced by youths in schools or in informal learning settings relevant to development. The specification of desired educational outcomes that are not being met results from a public policy and educational stakeholder concern that something is not right regarding the educational achievement and progress of students. A prominent pattern of this sort found across nations in the context of formal schooling is that students' achievement test scores are lower than desired for a group of students of identifiable backgrounds. Such backgrounds include, for example, coming from an immigrant background where parents show low schooling attainment and where immigrant families' adaptation to a new host country involves acquiring a new language and new ways of behaving and acting that contrast with an upbringing in a different country with its own privileged languages, customs, and institutional structures related to education and social opportunities. Other student background characteristics often associated with lower-than-desired achievement and attainment also pertain to nonimmigrant students. For example, in many countries around the world there may be failure of certain social groups or ethnic or racial minority groups to attain publicly and policy-specified educational achievement and attainment levels. These patterns can be of a long-standing historical nature associated with differences in the extent to which ethnic-racial subgroups have been integrated into a national society as a whole, and into the wealth-generating and wealth-accumulating institutions that generate economic and social well-being.

Given such contexts, the designers, implementers, and evaluators of educational interventions typically set forth educational remedies intended to change the educational achievement and attainment results of target immigrant and racial-ethnic subgroups, so that these interventions alleviate gaps in achievement and attainment relative to other groups in a nation and setting that show greater educational achievement and attainment on the indicators identified for this purpose by education policymakers. This article explores some of the important challenges faced by educational stakeholders in evaluating the effectiveness of interventions addressing these concerns with attention to the possibilities that achievement and attainment objects and their received indicators themselves may be questioned, given their suitability for members of culturally diverse communities. The matters to be addressed are quite complex in scope and so the intent is to forward important vision elements that should underlie attempts to develop evaluation philosophies and methods attuned to culture as a central issue. Here, vision elements mean deeper conceptual issues that should undergird attempts

to frame the purpose and outcomes of an intervention that has educational aims and the range of evidence that such outcomes have been attained.

Ecology, Culture, Educational Interventions, and Evaluation

There are many ways to approach the development of evaluation philosophies and methods that can lead to educational intervention evaluations sensitive to the cultural lives of members of diverse communities. Here, attention is called to an ecocultural approach building on the work of theorists and researchers such as Bronfenbrenner (1979), Cole (1996), Freire (1974), and van Lier (2004). The basic idea is that the context for educational interventions needs to be broken down into interacting formal and informal institutional systems that compose the realities surrounding the daily lives of community members. An educational intervention such as homework tutoring, mentoring, or specialized reading skill curriculum implementation occurs in a specific setting such as a classroom, community center, or some other location. Any such setting is surrounded and connected to other settings and institutions which can influence what occurs in the actual setting for an intervention.

Classrooms and schools as educational intervention settings are influenced in various direct and indirect ways by governance units for schools, teacher unions, and educational policy setting bodies such as local, regional, and national governance bodies and city school boards, legislatures, and executive bodies of government, such as city councils and state agencies. Business institutions and public policy advocacy groups, such as legal rights organizations, religious organizations, and community organizations, represent other examples of institutions that can influence what is possible and what can occur in an educational intervention setting. Informal organizations influencing what can occur in an educational intervention setting are exemplified by nuclear and extended families, friendship and informal activity peer networks among community members, and information sharing and knowledge dissemination networks implemented on the Internet or through other forms of media. It is important to remember that not all educational interventions occur in schools as part of regular school-day instruction. They may occur after school or on weekends in a variety of community settings such as community youth recreation centers or civic engagement clubs including, but going beyond, the physical setting of schools. The aforementioned formal and informal nonintervention policy setting and governance bodies and other institutions nonetheless may also bear influences – for example, a teacher union might not allow a credentialed teacher to engage in volunteer teaching service after school in a community setting without a

formal contractual arrangement with the institution sponsoring the activity at hand.

Regardless of the setting for an educational intervention, be it in or out of regular school-based instruction, various social organizations may have explicit or implicit stakes regarding the success of different community members and groups being served by an intervention. Different social organizations may show differences in their attentiveness for the success of an intervention based on the social and cultural characteristics of learners served by an intervention. For example, a state educational agency may provide financial support for literacy programs promoting students' acquisition of English-language literacy, but not provide support for maintenance of literacy in a first language. The ways that organizations may become involved in designing, supporting, and sustaining educational interventions is a dynamic outcome subject to local political and participatory motivation regarding ways to improve community well-being. Bronfenbrenner's (1979) formulation of ecological systems theory is helpful in highlighting ways that relationships between and among such organizations and settings evolve and are mutually influential.

Bronfenbrenner's (1979) ecological system formulation can be used by evaluators and educational stakeholders to inform our understanding of ways in which organizations, and, in particular, dynamic relationships among organizations, can affect the implementation and history of educational interventions. According to Bronfenbrenner, microsystem environmental systems include the immediate settings and face-to-face relationships among people in which individual lives occur. Such systems include, for example, community members' family, peer networks, school and neighborhood settings, and the activities that occur within the functioning of these settings. These relationships establish the social identities of individuals and pathways for personal and social development in these worlds or settings. Bronfenbrenner characterizes meso-system environmental systems as those relationships that arise between persons' immediate environmental settings, such as between a family and school setting, because of explicit and expected accountabilities of parents for the care of children, including schooling and schools' expectations of parents. Exo-systems are characterized by Bronfenbrenner as informal relationships that arise between social and institutional settings that affect individuals' performance and development in these settings. For example, parents' experience of racial or ethnic prejudice in the workplace might spill over into home family life, leading parents to express anger to children about social unfairness and reduce the motivation of children to be engaged in schooling activities requiring cooperation across mixed ethnic-racial subgroups. Macro-systems were characterized by Bronfenbrenner to refer to the broader relationships between persons' cultural and social heritage that set the stage for everyday behavior.

For example, immigrant students may have socialized patterns for showing respect for elders and teachers that are appropriate in a native country, but which then contrast with expected or permissible ways to interact with elders and teachers in the social contexts of a new host country. Finally, Bronfenbrenner characterized chronosystems as relationships among other ecological systems that evolve over the life-course of individuals and that thereby influence developmental trajectories over time. Schooling interventions aimed at increasing the educational achievement and attainment of ethnic-racial subgroup members exemplify such systems, albeit they can illustrate both beneficial and disruptive affects on children's lives – for example, see Valenzuela's (1999) work on subtractive schooling which examines how efforts to improve second language and immigrant background schooling achievement might actually have a negative impact on schooling outcomes by disconnecting students from values and knowledge that is part of their natal cultural background.

Bronfenbrenner's notions of ecological systems can prove useful to evaluators by suggesting relationships among different systems that might bear on students' responsiveness to an educational intervention and issues that might arise regarding the appropriateness of an intervention. The field of cultural psychology (Cole, 1996) contributes additional and complementary insights. One key insight is the notion of activity system. Humans in social and institutional settings engage in practices tied to the characteristics and social functions realized in settings. In common sense terms, every setting can be seen to involve persons assuming social roles appropriate to exercising agency in the setting, given its community location and expected functions. Every activity setting involves participants' active and constant negotiation and social interpretation of "who," "where," "when," "what," "why," and "how" – see Wertsch (1998) for an elaboration of these notions and their interdependent nature. An educational intervention when seen as an activity system realized in an ecological setting can be seen as purpose-driven with the end of having the students acquire the learning proficiencies and skills targeted by the intervention. The conduct of the intervention requires that the participants actively negotiate and understand the "wh's" and "how" of the setting and how learners are expected to respond to the guidance of the setting coordinators/instructors.

That stated, there are complications from a cultural perspective. The ways of acting and interacting shown by participants in an intervention setting may not always involve participants interpreting the same meaning for the "wh's" and "how." Classic research studies of classroom interaction involving members of indigenous minority groups in classrooms organized around nonindigenous classroom participation norms have found, for example,

that the failure of students to volunteer answers to teacher questions can arise not because students do not know the answer to questions, but because a student's public display of knowledge conflicts with local community cultural norms regarding polite ways of showing knowledge (Philips, 1993). More generally, students from cultural backgrounds that are different from those of a majority culture may not perceive the nature and demands of academic learning tasks in the same ways as students brought up in majority cultural settings (Adamson, 1993). Ability to communicate in the language used in an intervention setting is particularly crucial. When communicative competence in the language of an intervention is not considered, it is possible that students will not be able to participate as intended in the intervention and thereby not benefit.

From a cultural psychology perspective, the foregoing discussion is undergirded by the point that all humans acquire mediational resources or cultural capital that enable them to participate in everyday activities and settings. These forms of capital apply to every domain of human functioning and knowledge about how to communicate and act in activity settings. When the cultural capital that students possess through their socialization differs from the cultural capital that is valued in schools in a dominant culture intervention setting, there can be problems in that students may not show the responsiveness expected of them in an intervention. This was shown in the previously cited work of Philips (1993). Specifically, she found that American-Indian students from the Warm Springs Reservation showed reluctance to answer a teacher's questions publicly in classroom settings, because such public displays of knowledge violated politeness conventions for display of knowledge learned as part of socialization practices for community members.

Researchers from a critical pedagogy perspective (Cummins, 2001; Freire, 1974; Freire and Macedo, 1987; Valenzuela, 1999; Wink, 2005) carry these ideas further in a way that complements an ecological perspective. They forward the idea that failure of an educational system (and interventions) to build on the cultural assets of immigrant and nondominant ethnic-racial social groups can actually disempower members of these groups from benefiting from education. They also forward the view that when educators and interventions build on the cultural capital or funds of knowledge of ill-served groups, educational outcomes can be raised (González *et al.*, 2004).

Valenzuela's (1999) work on subtractive schooling is illustrative of how such disempowerment occurs. Her ethnographic study of Latino students in Texas classrooms traces how teachers' inability to appreciate the natal cultural values and interaction style of students leads teachers to ascribe low learning motivation and learning ability to Latino students, which in turn leads to a teacher's display of animosity toward students, who then, in turn, display disinterest in classroom

learning and limited evidence of benefiting from instruction. A similar disempowerment was found in the research of Moll *et al.* (1992). These researchers found that when English as a Second Language (ESL) reading teachers believed that elementary school Latino English-language learners were incapable of comprehending the meaning of book stories written in English, they concentrated on teaching only basic English vocabulary skills to children and thereby did not permit children to exercise deeper comprehension and communications skills. In contrast, the researchers found that ESL teachers who required that similar background students write book reports on what they learned from book stories showed evidence of significant understanding of stories.

Van Lier (2004) working from an ecological perspective adds important additional insights. He effectively theorizes the many ways in which members of social and cultural groups construct meaning, social identity, and agency by continuously interpreting their environment and interactions with others. He calls attention to the important phenomena of prolepsis – the capacity of persons individually and collectively to project their identities and agency into the future based on current experiences and the affordances for action and participation in current events.

An example of prolepsis arises when youths describe to others their aspirations of who they will become as adult members of societies. While prolepsis is embodied in projections of future roles in the economy in terms of professions and jobs, it is a much broader notion and includes the ways in which youths see the range of identities as members of society and ways that they will contribute to society – this will be especially evident in the discussion of the out-of-school Youth Empowerment Strategies (YES!) educational intervention program that follows in a later section. From the perspective of prolepsis, if an educational intervention is to work, it must be capable of bridging the existing cultural assets of individual with new tools for thinking, problem solving, and language use that extend the identity and agency of individuals into future circumstances and opportunities for development.

Some Implications of an Ecological Perspective to Evaluation Philosophies and Methods

While there is no singular way to categorize educational evaluation philosophies, one can nonetheless contrast two very different evaluation approaches that illustrate how cultural issues may affect what is learned from an evaluation.

For example, it is possible to distinguish objectivist logic model approaches from transformative participatory evaluation approaches (Mertens, 2009). Objectivist approaches,

in their most simplistic and narrow implementation, for example, might use experimental or quasi-experimental methods to determine whether implementation of an educational treatment intervention is associated with a statistically significant gain by a treatment group over a comparison group on outcome measures such as tests of constructs targeted for improvement as a result of exposure to a treatment, versus nonexposure to the treatment. Such evaluations might take a formative or summative form. If a formative approach is used, the association of changes in outcome measures can be examined as they emerge over time as an intervention is administered, and it is possible that the results of such formative outcomes can be used by educational intervention program staff to improve design and delivery of the intervention with the aim of improving outcomes of a next phase formative assessment. A summative evaluation, in its simplest embodiment, only compares growth in outcome measures at the conclusion of an intervention.

It is common practice for objectivist evaluation approaches to apply an input–output logic model to specify conceptually the hypothesized relationship between exposure versus nonexposure to a treatment and targeted outcome measures. A logic model may also specify measures of ancillary input and process variables that may mediate such relationships. Ancillary input variables may include measures of relevance to the language, and socio-cultural status of participants (e.g., race, ethnicity, immigrant experience gender, and socioeconomic level) in the evaluation, and statistical modeling procedures may be used to assess direct or indirect effects of these variables on process measures and the outcomes produced by exposure to the intervention. In more complex designs, hierarchical linear modeling or structural equation modeling might be used to examine the effects of variables, such as school resources, community location, racial–ethnic school, or community characteristics, as variables independently contributing to changes in outcomes associated with exposure versus nonexposure to an educational intervention treatment.

One of the most basic characteristics of the objectivist approach to evaluation is that the evaluation team is desired to be a separate independent entity from the education program implementing the educational intervention. The reason for this separation is to ensure that the outcome of the evaluation cannot be influenced by inappropriate collection or analytic manipulation of data, leading to invalid specification of education intervention results.

Culturally responsive and transformative participatory evaluation approaches to educational evaluation can take a very different stance, one that is sensitive to an ecological perspective of human action and development. This perspective, in general, would view educational intervention implementers, recipients of an intervention and their families–communities, and evaluators all as mutual stakeholders in the design, development, implementation, and

research evaluation of an intervention. Such approaches can build on an account of how the cultural assets of target community members are the foundation for implementation of a successful intervention. An important quality of this approach is that evaluators can take a more active and accountable role in making educational interventions happen effectively than is the case with an objectivist orientation. The intervention itself can take the features of a clinical design experiment or formative intervention – these going beyond the goals of a traditional objectivist formative intervention to include an active, creative search and exploration for how to go about designing what the intervention is and how it can draw on cultural resources of communities and their formal and informal institutions, given findings collected during implementation of the intervention (Engeström, 2008).

Two educational interventions are mentioned in the following section that illustrate the richness of how cultural issues may affect the design and enactment of evaluations. The Kamehameha Early Education Program (KEEP) interventions and its evaluation illustrate how evidence of the effectiveness of a classroom literacy intervention serving at-risk Hawaiian youths, coupled with a study of Hawaiian community cultural communication practices led to improvements in the delivery of the classroom literacy intervention. The example also led to the discovery that effective delivery of the intervention in another cultural community required understanding the cultural norms for communication in that community and fine-tuning the literacy intervention accordingly.

The second example of an intervention, the YES! after-school program, illustrates how cultural perspectives might inform evaluations concerns, given the goal of having youths from diverse cultural backgrounds learn how to conduct research on how to make their communities better. It is fair to state that the YES! program has yet to evolve a definitive evaluation strategy because of its participatory action–research orientation. This latter orientation has as its goal having youths discover what they want to investigate about their communities and the outcomes of this discovery process and how youths present the outcomes to other community members are unpredictable in advance. The example also illustrates how culture needs to be treated as a continuous construction of meaning of what is going on in everyday life from an ecological perspective in contrast to being considered a background trait of individual community members and groups.

The Evaluation of KEEP as an Example of a Culturally Responsive Evaluation

The implementation and evaluation of the KEEP reading program intervention, serving elementary school youths from low-income Hawaiian backgrounds in the late 1970s

to early 1980s, is illustrative of a culturally responsive evaluation approach (Tharp and Gallimore, 1979, 1988). The KEEP reading comprehension program brought together teachers and teacher trainers, and an interdisciplinary team of researchers, including applied linguists, anthropologists, and educational psychologists. The goal of the program was to improve the reading performance of students as indexed by reading comprehension test scores and other indicators of reading competence. It took over 10 years for the program to arrive at a sustainable, stable configuration that was able to manifest statistically significant growth in reading test scores as compared to reading test scores of children from similar backgrounds not exposed to the curriculum. The early stages of implementation of the program focused on helping teachers acquire proficiency in the delivery of direct reading instruction targeted largely at discrete reading decoding and correct answering of questions about story texts. As the program implementers explored the improvement of reading instruction, the research intervention team developed awareness through studies of classroom videotapes that children showed preferences for how they interacted with each other during instruction that were not part of the direct instruction model. Specifically, they discovered that children working in small groups would sometimes spontaneously look around the classroom for students who appeared to be having problems completing their reading assignments. Students would on occasion then leave their group spontaneously to go and help a student in another group. A second finding was that during direct instruction reading aloud in small groups, students who were not the ones to be reading aloud to other group members had a proclivity to speak up and add vocal commentary regarding the interpretation and meaning of what was being read. It is important to underscore that both of the two behaviors arose spontaneously and not as a result of following the script for intended direct instruction.

The reading implementation and research team came to the discovery that these spontaneous behaviors that manifested students' engagement with learning activity were not accidental, but resulted from cultural practices that were common in the natal community outside of school. Anthropologist team members who were using ethnographic methods to document social and communication practices in the homes of children and in the community found this connection. Their home ethnographies, for example, found that children often helped each other perform chores at home and also that community storytelling practices (known as talk story) involved having story audience members break in to the oral delivery of a story with comments and annotations.

The foregoing discoveries led the KEEP reading intervention team to revise the design of the reading program so as to add in both types of behaviors as encouraged or permissible practices, built into the implementation of the

program. This shift in program implementation coincided with evidence of statistically significant improvement in the reading test scores of students exposed to the curriculum.

The implementers of the KEEP reading intervention went on to attempt to implement the program in the northern Arizona Navajo community of Rough Rock (Vogt *et al.*, 1993) in the elementary grades. The team discovered that the program could not be implemented successfully in this community without adjusting its reliance on small group activity design and the program implementers' expectations that students would naturally be inclined to help each other in the classroom. In particular, they discovered that members of small reading groups would not be engaged in their work unless they were separated by gender. They also discovered that, contrary to their expectations, students were not inclined to help each other spontaneously, as was the case with Hawaiian children. They also discovered that Rough Rock children were best managed by ignoring misconduct during learning activities and reminding the children of honorable behavior compared to the effectiveness of discouraging inappropriate behavior of Hawaiian children with gentle, but stern scolding.

Participatory Action Research and Evaluation: The YES! Program as an Example

The examples of culturally responsive evaluation cited above clearly build on understanding culturally valued ways of acting and communicating as a foundation for implementation of educational interventions, and for use of evaluation, coupled with research, as a way to arrive at effective interventions that can lead to measurable results. A participatory action research and evaluation perspective is a related but more proactive methodology that draws on an ecological account of human action (Fals Borda, 2006; Ulrich and Wenzel, 2009). The YES! Project is an example of such an approach (Wilson *et al.*, 2007, 2008).

The project in question utilizes a critical pedagogy orientation to empower urban, upper-elementary-school-age and adolescent youths to conduct projects they design to enact social action efforts that address local community problems. Youths use a Photovoice technique that employs photography to convey youths' perceptions of problems (Wang and Burris, 1997). Youths author essays that accompany the photos that discuss and analyze the community problem context captured in photos. The fuller curriculum of YES! involves six sequential steps: group formation, photography, Photovoice, community-organizing strategies, engaging the group in social action, and engaging the community in social action. Evaluating the effectiveness of an implementation of the YES! curriculum is a very challenging and creative endeavor. As noted

by the implementers, one of the most important outcomes of an implementation is the unsurfacing of community needs themselves.

The outcomes of YES! are not restricted to outcomes for individual youth participants. They include communications to the surrounding community and appropriate community agencies issues and problems, as seen from the perspectives of youth with the anticipation that both youths and community members can contribute to the resolution of problems.

The community problem identification outcome of the YES! curriculum is an example of a generative learning outcome (Wittrock, 1989) that contrasts with a traditional school curriculum where the target skills and knowledge to be learned by youths is well known and specified by the boundaries and expectations of grade-level learning standards in a subject-matter content area. In contrast, in the YES! curriculum, akin to the critical pedagogy process of problem identification coupled with problem analysis regarding solutions, youth must actively generate their learning regarding what is a significant community problem by exploring their community ecology and analyzing conditions that need to be addressed. The problem identified and partially represented in photographs needs to be explained in an accompanying essay or a creative performance such as a play, and these problem representations and enactments must then be brought to community members for action – proposed solutions, in a convincing manner that could lead to action by community members. Traditional school curricula seldom foster the integrative thinking and civic engagement goals targeted by a curriculum such as YES!, although there are calls by educational leaders for connecting students' school-based experiences with the development of students' capacities as community problem solvers, applying and blending what they learn in classrooms with their knowledge of the ecology of their communities (Association for Curriculum Development, 2009). The implementers of the YES! program describe the proximal outcome goals for YES! for participating youths as increased skill in collaborative decision making, conflict resolution, and increased political participation. Interestingly, these goals resonate well with the twenty-first-century skill priorities for development of a resourceful citizenry (see, e.g., Gardner 2006). However, the implementers of the YES! program do not restrict their vision to proximal goals associated with such skills. They also wish to measure effectiveness and success of the program in terms of youths' learning how to take care of themselves in terms of their broader well-being, as a result of becoming more effective problem solvers regarding community well-being. In this light, long-term outcome goals are described as positive changes in health and wellness awareness (Center for Research on Adolescent Health and Development, 2002–2006).

The breadth of federal, local, and community institutions involved in implementing YES! program illustrates how Bronfenbrenner's ecological perspectives can inform a deeper understanding of educational interventions. The funders of the YES! program include the federal Centers for Disease Control and Prevention via a grant to the Center for Research on Adolescent Health and Development, of the Public Health Institute based in Oakland California, and a local elementary school. The participants in the YES! program are community youths residing in Oakland attending the school in question. The federal–city agency–school–community youth links represent a chain of resources and health education and community improvement capacity building, cutting across very different but interdependent institutions. Other cross-institutional ecologies would also seem worthwhile to investigate as part of an evaluation of the YES! program. For example, youths' investigation of community problems, such as community sanitation needs, would involve their investigating the role of city agencies to supply needed sanitation services. On the whole, a deep evaluation of the outcomes of the YES! program would benefit from addressing how the program's limits and successes build on these relationships to transform the well-being of youths and community members. The YES! program also illustrates the importance of understanding how the everyday sense-making of youth participants coupled with the practices and sense-making of institutions interfacing with the program constitute the construction of culture itself. Culture is made to happen in the practices and meaning-making of community members and institutions as they negotiate their interpretation of reality and the consequences of these interpretations. This is a very different notion of culture, resonating with a cultural psychology and ecological perspective as compared to a perspective that culture is a characteristic or background variable describing intervention participants and stakeholders.

Clearly, much work remains to be done in operationalizing assessment and other forms of evidence of such outcomes. There is very little concrete evidence presented as yet regarding YES! program effects, although the provocative nature and sensitivity of this program in light of cultural and ecological issues augers for such evidence. Thus far the implementers report that youths show some positive qualitative evidence of meeting the ideal goals of the YES! curriculum, although the results also show limitations in youths' capacities to implement the fuller range of analytic thinking skills and explanatory communicative demands required to convey the importance of the problems they identify in a convincing manner (Wilson *et al.*, 2007, 2008). Similarly, a fuller, ecological sensitive evaluation of the YES! program would need to develop methods and lines of evidence to characterize and measure how the implementation and impact of the program is shown by the development of relationships between the program

and other institutions such as the funders, host school and school district, civic agencies, and community groups. It would seem useful to suggest a narrative historical account of how the program developed and the relationships developed as part of such an evaluation strategy.

Conclusion

There are many ways to approach the evaluation of educational interventions that can improve the validity of evaluations in light of cultural and ecological issues. This article has explored only a limited range of such possibilities, although it calls attention to the potential importance of insights drawn from ecological systems theory, cultural psychology, and critical pedagogy that connect well with interventions developed to be responsive to the cultural background and real-world community perceptions of youths targeted by educational interventions. Much work needs to be done to create and strengthen evaluation strategies for such interventions, given the challenges and even impossibilities of adequately capturing and measuring valued outcomes solely by statistical methods.

See also: Curriculum Evaluation; Evaluation and Accountability; Moral and Ethical Issues in Evaluation; Program Evaluation.

Bibliography

- Adamson, H. D. (1993). *Academic Competence: Theory and Classroom Practice—Preparing ESL Students for Content Courses*. White Plains, NY: Longman.
- Association for Curriculum Development (2009). Teaching social responsibility. *Educational Leadership* 66(8).
- Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by Nature and Design*. Cambridge, MA: Harvard University Press.
- Center for Research on Adolescent Health and Development (2002–2006). Youth Empowerment Strategies (YES!). <http://crahd.phu.org/projects/YES.html> (accessed October 2009).
- Cole, M. (1996). *Cultural Psychology: A Once and Future Discipline*. Cambridge, MA: Harvard University Press.
- Cummins, J. (2001). *Negotiating Identities: Education for Empowerment in a Diverse Society*. Ontario, CA: California Association for Bilingual Education.
- Engeström, Y. (2008). From design experiments to formative evaluations. *Keynote Presentation at the International Conference on Learning Sciences*, Utrecht.
- Fals Borda, O. (2006). Participatory (action) research in social theory. In Reason, P. and Bradbury, H. (eds.) *Handbook of Action Research*, pp 27–37. London: Sage.
- Freire, P. (1974). *Pedagogy of the Oppressed*. New York: Seabury Press.
- Freire, P. and Macedo, D. (1987). *Literacy: Reading the Word and the World*. South Hadley, MA: Bergin & Garvey.
- Gardner, J. (2006). *Five Minds for the Future*. Boston, MA: Harvard Business School Press.
- González, N., Moll, L., and Amanti, C. (eds.) (2004). *Funds of Knowledge: Theorizing Practices in Households, Communities, and Classrooms*. Mahwah, NJ: Erlbaum.
- Mertens, D. M. (2009). *Transformative Research and Evaluation*. New York: Guilford Press.
- Moll, L., Diaz, S., Estrada, E., and Lopez, L. (1992). Making contexts: The social construction of lessons in two languages. In Saravia-Shore, M. and Arvizu, S. (eds.) *Cross-Cultural Literacy: Ethnographies of Communication in Multiethnic Classrooms*, pp 339–366. New York: Garland.
- Phillips, S. (1993). *The Invisible Culture: Communication in Classroom and Community on the Warm Springs Indian Reservation*. Long Grove, IL: Waveland Press.
- Tharp, R. G. and Gallimore, R. (1979). The ecology of program research and development: A model of evaluation succession. In Sechrest, L. B. (ed.) *Evaluation Studies Review Annual*, vol. 4. Beverly Hills: Sage Publications.
- Tharp, R. and Gallimore, R. (1988). *Rousing Minds to Life. Teaching, Learning, and Schooling in Social Context*. New York: Cambridge University Press.
- Ulrich, S. and Wenzel, F. (2009). Participatory evaluation. A perspective for human rights education. <http://www.bpb.de/files/J44PTB.pdf> (accessed October 2009).
- Valenzuela, A. (1999). *Subtractive Schooling: U.S.–Mexican youth and the Politics of Caring*. New York: State University of New York Press.
- van Lier, L. (2004). *The Ecology and Semiotics of Language Learning*. Boston, MA: Kluwer Academic Publishers.
- Vogt, L., Jordan, C., and Tharp, R. (1993). Explaining school failure, producing school success. In Jacob, E. and Jordan, C. (eds.) *Minority Education: Anthropological Perspectives*, pp 53–65. Norwood, NJ: Ablex Publishing Corporation.
- Wang, C. and Burris, M. A. (1997). Photovoice: Concept, methodology, and use for participatory needs assessment. *Health Education & Behavior* 24(3), 369–387.
- Wertsch, J. (1998). *Mind as Action*. New York: Oxford University Press.
- Wilson, N., Martin, A., Wallerstein, N., Wang, C., and Minkler, M. (2007). Engaging young adolescents in social action through photovoice. *Journal of Early Adolescence* 27(2), 241–261.
- Wilson, N., Minkler, M., Dasho, S., Wallerstein, N., and Martin, A. (2008). Getting to social action: The youth empowerment strategies (YES!) project. *Health Promotion Practice* 9(4), 395–403.
- Wink, J. (2005). *Critical Pedagogy. Notes from the Real World*. Boston, MA: Allyn and Bacon.
- Witrock, M. (1989). Generative processes of comprehension. *Educational Psychologist* 24, 345–376.

Evaluation and Accountability

E Sato and S N Rabinowitz, WestEd, San Francisco, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Accountability – The condition or state of being held responsible for one's actions.

Evaluation – The careful appraisal of something in order to judge its significance or quality.

Framework – A guideline that includes criteria for systematically approaching the development and implementation of something.

Qualitative data – Data that cannot be quantified or measured and that are instead represented with words.

Quantitative data – Data that can be quantified or measured.

Stakeholders – The individuals or groups involved in or affected by a course of action or outcome.

Stakes – The level of risk or impact associated with a course of action or outcome.

Triangulation – The analysis and use of information from multiple measures or sources of data in the examination of something. These data may be convergent, inconsistent, or contradictory, but must be rendered sensible by the researcher or evaluator in order to minimize the limitations and intrinsic biases that may be present in a particular measure or data source.

Utility – The degree to which the inferences made on the basis of certain information are clear and easy to understand by key stakeholders, and the facility with which consequent actions can be consistently and accurately implemented as intended.

Validity – The degree to which the inferences made and actions taken on the basis of certain information are trustworthy and appropriate.

Introduction: Evaluation and Accountability

Evaluation is the careful appraisal of something (e.g., product, service, program, and policy) in order to judge its significance or quality (American Evaluation Association, 2006; Madaus and Stufflebeam, 1984; Merriam-Webster, 2003). Accountability is the condition or state of being held responsible for one's actions (American Evaluation

Association, 2006; Linn, 2001; Madaus and Stufflebeam, 1984; Merriam-Webster, 2003). Both evaluation and accountability have applications in many disciplines, and numerous models exist for both. Although a careful discussion of the relative merits of various models of evaluation and accountability across disciplines is of importance, it is not the focus of this article. Instead, the authors describe evaluation and accountability within the educational context and adopt a neutral position and do not endorse or advocate for any particular model. The authors focus on the common function both evaluation and accountability serve of improving learning, teaching, and educational administration, and discuss evaluation and accountability within a framework of validity and utility. Such a framework is important to consider because, although not identical in purpose and methods, both evaluation and accountability require judicious design and implementation to ensure appropriate, meaningful, and practicable outcomes that effectively support educational improvement. For general discussions of the background and models of educational accountability, see, for example, Ananda and Rabinowitz (2001), Baker *et al.*, (2002), Gong (2002), and Linn (2001). For general discussions of the background and models of educational evaluation see, for example, Alkin (2004), Christie (2003), Fuhrman and Elmore (2004), Madaus and Stufflebeam (1984), and Patton (2008).

Framework of Validity and Utility

Overview

The validity and utility framework that this article uses to describe educational evaluation and accountability extends the strong base of relevant research and practice that exists for validity and utility (e.g., Brennan, 2006; Kane, 1992; Messick, 1989). Assuming that the information gathered for evaluation or accountability purposes is accurate and reliable, validity refers to the degree to which the inferences made and actions taken on the basis of the information are trustworthy and appropriate (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999; Merriam-Webster, 2003; Messick, 1989). Utility refers to the degree to which the inferences made on the basis of the information are clear and easy to understand by a range of key stakeholders, and the facility with which consequent actions can be consistently and accurately implemented as

intended (Assessment and Accountability Comprehensive Center, 2009; Merriam-Webster, 2003).

In general, valid and useful educational evaluation or accountability has:

- a clearly defined purpose and scope,
- appropriate involvement of vital individuals or groups,
- information from multiple measures and data sources,
- context-sensitive outcomes,
- reasonable and defensible interpretations of information that lead to adequate and appropriate actions, and
- reporting that is available and accessible to appropriate stakeholders.

This article elaborates on each of these aspects of the framework and describes implications for educational evaluation and accountability. Additionally, the following sections present critical links between evaluation and accountability that strengthen their contributions to educational improvement, as relevant to each aspect of the framework.

Clearly Defined Purpose and Scope

The purpose – the intended outcomes and uses – of evaluation or accountability efforts should be explicit and clear in order to effectively guide their design and implementation. Additionally, the scope – the scale or extent of the effort, including any limitations or restrictions – should be explicit and clear in order to help determine what resources are needed for evaluation or accountability as well as their design and implementation.

The purposes and scope of evaluation and accountability typically differ. Accountability systems tend to focus on end results relative to an established standard or level of achievement. Evaluation efforts also can produce summative results related to a standard or level of achievement (e.g., whether an educational program or intervention accomplished its stated goals; a program or intervention's level of impact on changing behaviors or practices), but they also can be designed to yield formative information (e.g., specific ways in which or conditions under which an educational program or intervention is successful, or not, as it is being implemented), and data that inform the understanding of relationships among key variables relevant to the target of the evaluation (e.g., product, program, service, and policy).

Common challenges to crafting an effective statement of purpose and scope include having an appropriate focus (not too broad or too narrow), requiring an appropriate level of effort (e.g., relative to time and phases of the effort, level of associated risks or stakes, amount of resources, number and complexity of related tasks), and conforming to requisite limitations or restrictions (e.g., legislation). If these challenges are not addressed, they pose threats to

the validity and utility of evaluation and accountability efforts. Therefore, it is essential that key information for purpose and scope are considered upfront, and that the information appropriately addresses the critical, substantive questions or issues of focus for the evaluation or accountability effort.

More specifically, in the stated purpose and scope, information should be evident to:

- identify who needs to be involved (e.g., students, teachers, administrators, parents/guardians, policymakers, and other members of the community);
- define what the intended outcomes are (e.g., level of achievement, impact or effectiveness of a program, cost/benefit analysis, and quality assurance);
- determine what tasks need to be done (e.g., data collection activities, analyses);
- determine what is needed to accomplish specified tasks (e.g., data, finances, systems, and existing vs. additional resources);
- define how findings are intended to be used (e.g., formative, summative, for rewards and sanctions);
- determine how possible unintended consequences will be identified and minimized (e.g., indicators, procedures); and
- determine what measures will indicate when the intended outcomes have (or have not) been achieved.

Upfront consideration and explicit articulation of key information will more likely lead to the optimization of related design and implementation strategies and will more likely ensure that the outcomes of evaluation or accountability efforts will effectively provide direction and guidance for educational improvement.

Link Between Evaluation and Accountability

Related evaluation and accountability efforts can be conceived concurrently. A critical link that can be established between evaluation and accountability efforts is in the crafting of purposes and scopes that cohere, and such coherence has implications for the validity and utility of the efforts and their outcomes. The focus of accountability can provide information regarding the kind of evaluation that would be most useful in facilitating understanding of accountability findings and their implications, and data from evaluations can provide information regarding the design of accountability systems. That is, given the expected targets (end results) of an accountability system, evaluations can be conceived to provide specific information related to (1) the successes or failures of relevant people, processes, and conditions or contexts, (2) the appropriateness, including the technical adequacy of measures and

information used to determine the degree to which accountability targets are achieved, and (3) the appropriateness of the outcomes and their uses and impact. Therefore, upfront conceptualization of evaluation and accountability efforts that cohere can yield information that is not limited to whether something has achieved a standard or level of expectation (accountability); rather, coherent evaluation and accountability efforts can yield information that informs an understanding of the possible reasons for and the conditions under which something has been achieved (evaluation). Consequently, the information can provide direction for educational improvement.

Appropriate Involvement of Vital Individuals or Groups

The appropriate involvement of vital individuals or groups is essential to ensuring the validity and utility of outcomes. Those who should be involved in evaluation or accountability activities are individuals or groups who:

- have knowledge and experience that will provide information regarding the design of the evaluation or accountability effort;
- are necessary for the successful implementation of the effort;
- are critical to ensure valid interpretation of findings and usefulness of subsequent decisions and actions taken; and
- generally are those whose participation would provide the greatest assurance that the effort ultimately will lead to educational improvement.

The selection of individuals and composition of groups involved during different key phases of an evaluation or aspects of accountability may need to differ, depending on the particular needs or intended outcomes of a phase or aspect of an effort. For example, policymakers may be more heavily involved in early phases of evaluation or accountability efforts determining purposes and scope, while teachers may be better suited to inform implementation phases of evaluation or accountability efforts. No matter who is involved, to best ensure the validity and utility of evaluation or accountability efforts, individuals, at a minimum, should have appropriate expertise and experience, and there should be adequate balance of representation and perspectives across individuals involved and stakeholder groups. Appropriate inclusiveness is essential for effective evaluation, and broad participation and shared responsibility are necessary for successful accountability (American Evaluation Association, 2006; Linn, 2001).

Appropriate involvement, regardless of the phase or aspect of evaluation or accountability, also requires that individuals have:

- adequate training and orientation to the task to which they are attending;

- an understanding of the intended outcomes of the effort and how these outcomes relate to the broader evaluation or accountability objectives; and
- sufficient time for their involvement.

Inadequate individual expertise and experience, imbalanced representation of knowledge and perspectives, inadequate training and orientation, and lack of sufficient time pose threats to the validity and utility of evaluation and accountability efforts. Therefore, it is important that, in the initial stages of planning evaluation or accountability efforts, a general timeline is determined, key phases of the evaluation or aspects of accountability are identified, the particular needs and intended outcomes of each are defined, and the expected level of effort is determined in order to appropriately address such threats. Such upfront planning should provide information regarding the qualifications and range of individuals necessary for an effective evaluation or accountability effort.

Link Between Evaluation and Accountability

A key outcome of accountability is the proper attribution of responsibility for the success or failure of something (Fuhrman and Elmore, 2004; Linn, 2001). Therefore, to the degree that such an attribution could be assigned to an individual or group (e.g., teachers and administrators), those who could be held accountable should be included in accountability and related evaluation activities. Inclusion is not meant to create real or perceived conflicts of interest where those who may be accountable for results hold persuasion over the specific indices and standards for judgment. Rather, proper inclusion is intended to provide information regarding the mechanics of the accountability design and help ensure the meaningfulness of accountability efforts.

Similarly, evaluation efforts should be designed with input from vital individuals or groups in the:

- determination of appropriate indicators for the accountability and evaluation efforts;
- selection of appropriate measures to collect relevant data; and
- implementation of measures in order to ensure they are properly administered and that data are appropriately collected, analyzed, and reported (Alkin, 2004; Patton, 2008).

Additionally, evaluation efforts should gather information from those involved in the efforts about the relative value versus burden of the measures used and the information yielded in order to better understand circumstances related to fidelity of the efforts, interpret findings, and provide information regarding decisions and actions.

Multiple Measures and Data Sources

Multiple measures and data sources (both qualitative and quantitative) are critical to supporting valid and defensible interpretations of information that lead to appropriate actions. There may be limitations to any single measure or data source, and multiplicity provides a means for triangulating information. That is, there may be convergent, inconsistent, or contradictory information across measures or data sources that must be rendered sensible by the researcher or evaluator in order to minimize the limitations and intrinsic biases that may be present in a particular measure or data source (Mathison, 1988). Additionally, using multiple measures and data sources can yield a more robust depiction of the subject being examined, thereby facilitating more valid interpretations of information, minimizing uncertainty, and increasing the utility of the findings. Therefore, whenever possible and appropriate, multiple measures should be used, and both qualitative and quantitative data should be considered.

Common challenges that pose threats to the validity and utility of evaluation or accountability efforts include the quality of measures and data and the possibility of omitting a class of important, relevant information. For example, more standardized instruments (e.g., multiple-choice assessments, highly scripted observation protocols) may have high reliability but less relevance and validity than more open-ended methodologies (e.g., performance assessments, cognitive interviews). An understanding of the technical quality of the measures used, the degree to which something is able to be directly and accurately measured or observed, and possible inherent source(s) of bias of a measure or data source are essential to overcoming such threats. Such understanding will help establish appropriate parameters within which inferences about findings can be made (e.g., limitations, conditions, contingencies), thereby supporting the validity of interpretations and the utility of consequent decisions and actions.

Link Between Evaluation and Accountability

Given the high stakes often associated with accountability outcomes, it is vital that the information used to make accountability judgments is accurate and reliable and supports valid interpretations. Quality information is dependent on the quality of the measures used to gather the information. Therefore, evaluations should be designed to support accountability outcomes by:

- examining the technical quality of the measures used to gather information on which accountability judgments are based (e.g., validity, reliability, freedom from bias and sensitivity issues);

- determining whether the information collected is appropriately central or supplemental to the particular question or issue requiring examination;
- scrutinizing the degree to which and appropriateness with which results of individual indicators generalize to other related indicators;
- appraising the body of information collected and its sufficiency to support various accountability outcomes; and
- examining the efficacy of the indicators of the accountability system's success or failure – in terms of both intended and unintended outcomes.

Coherent evaluation and accountability efforts that include multiple measures and data sources are necessary for gathering a robust body of information, which can lend itself to the validity and utility of outcomes that appropriately guide educational improvements (American Evaluation Association, 2006; Baker *et al.*, 2002).

Context-Sensitive Outcomes

There are a number of contextual influences that can impact the validity and utility of evaluation or accountability outcomes (American Evaluation Association, 2006; Ananda and Rabinowitz, 2001; Linn, 2001; Patton, 2008). Such influences include:

- politics;
- conflicts among interests, values, and priorities;
- existing systems and structures;
- availability of resources;
- the cultural context; and
- the general readiness to act by those who are involved in the evaluation or accountability efforts and by those who will be affected by the outcome(s).

There are obvious ways in which aspects of context can negatively affect evaluation or accountability efforts, including placing restrictions on the focus of the evaluation and the consequences associated with evaluation findings or accountability outcomes, undermining efforts because the efforts are not consistent with or do not necessarily support political or fiscal objectives, manipulating availability or collection of information to serve specific interests, yielding unreliable or invalid information, and yielding biased interpretations that serve a particular agenda or interest group. Understanding the context within which the evaluation or accountability efforts are conducted and their outcomes enacted is essential to avoiding or minimizing threats to validity and utility. Such understanding can be gained through the purposeful inclusion of key stakeholders (e.g., policy-makers, technical experts). Key stakeholders should be identified upfront and opportunities should be planned for appropriate communication and collaboration across these stakeholders. Doing so will best ensure that various (possibly conflicting) perspectives and interests are

appropriately accounted for and balanced upfront, and a shared understanding of the goals and parameters of the evaluation or accountability efforts is established.

Link Between Evaluation and Accountability

Both evaluation and accountability occur within dynamic contexts that typically are laden with politics, practical issues, and values. Evaluation can aid the understanding of aspects of context that may impact the timing and scope of the evaluation itself (e.g., degree to which mid-implementation information and possible revisions are feasible and desirable) and how its findings might inform accountability decisions. Additionally, context-sensitive evaluation can better ensure the accuracy of information gathered, the efficacy of resulting recommendations, and the meaningfulness and appropriateness of the accountability outcomes. Often when accountability models are conceptualized and developed, there can be tension between the broad, overarching goals of policymakers (e.g., the broadly stated objective that all students will achieve at high levels) and the concerns of technicians that the available information is not refined or precise enough for the level of scrutiny and attribution needed to support broad accountability goals. For example, there often are debates in the research and policy arenas about the relative merits of status versus growth goals as the basis to determine school success (e.g., Betebenner, 2008; Goldschmidt *et al.*, 2005). This debate has led to the development and implementation of various value-added methodologies to evaluate teacher success (Sanders *et al.*, 1997). If a growth model is desired with value-added consequences for classroom teachers, it is important to determine the degree to which the available assessments are reliable and will yield data to support valid interpretations related to fine distinctions among different levels of student performance. Evaluation is critical for examining the technical adequacy of the measures used, the accuracy and reliability of the data collected, the validity of the interpretations, and the overall efficacy of the accountability model.

Reasonable and Defensible Interpretations and Adequate and Appropriate Actions

All aspects of the aforementioned validity and utility framework, if appropriately addressed, help to ensure effective evaluation and accountability efforts. The effectiveness of evaluation or accountability efforts in terms of educational improvement is dependent on reasonable and defensible interpretations of information and the adequacy and appropriateness of resultant actions. In order to best ensure the validity of interpretations of information and the utility of outcomes, consideration must be given to the following:

- Diversity of the data – If sources of data are too diverse, with little consistency or connection among the various types of information gathered, then such diversity, when aggregated, may limit the interpretability, generalizability, and utility of the information.
- Weighting of information – Given the critical questions and issues examined by the evaluation or accountability effort, *a priori* consideration must be given to the importance of information gathered from different sources and the influence each source of data and type of data (qualitative, quantitative) should have on interpretations and subsequent decisions and actions.
- Extent and trustworthiness of interpretations and extrapolations made from the data – The degree of inference involved in the interpretation of data, including the degree to which findings can generalize, should be considered.
- Limitations of the data (see related discussion on multiple measures and data sources).
- Amount and type of impact – Both positive and negative consequences should be considered, including the degree to which and manner by which understanding of such consequences can be substantiated with data.

Link Between Evaluation and Accountability

Accountability serves the function of improving learning, teaching, and educational administration, by assigning responsibility for the success or failure of a particular educational endeavor. Accountability efforts can benefit from evaluation because evaluation adds a level and type of scrutiny that can hone understanding of the who, what, and why delineated through accountability efforts:

- Who – To better understand who is responsible, evaluation can focus on group and subgroup characteristics at a greater level of detail than accountability efforts and can provide a richer depiction of the groups and subgroups affected (i.e., determined through the analysis of relevant aggregated and disaggregated quantitative data along with data from qualitative measures, such as surveys and observations).
- What – This aspect of accountability can be better understood through evaluation efforts that closely examine characteristics of the conditions that are affecting groups and subgroups. Such conditions include those aspects of context discussed above. As evaluation often uses group and subgroup characteristics as blocking variables or predictors in complex regression models, the possible differential effects of aspects of context on various subgroups can be better understood.
- Why – Evaluation can be designed to obtain information from representative samples on a number of indicators in a manner that should enable extrapolation of

findings to a broader population and/or context with similar characteristics as the sample. Therefore, evaluation can provide information regarding accountability efforts by providing more robust information related to the positive and negative effects of something – both in terms of what is intended and what may not be intended. Evaluation also can help to appraise the sufficient size and nature of an effect to determine something as a success versus determining something as a failure.

The level and type of detailed information provided by evaluation lends itself to more specific, actionable plans for educational improvement. Thus, evaluation and accountability efforts can be coordinated to increase both the validity and utility of the information.

Available and Accessible Reporting

Transparency of educational evaluation or accountability efforts is important to ensure that the efforts themselves are accountable for the appropriateness of procedures and the adequacy of findings, their interpretation, and subsequent decisions and actions. Transparency can be achieved by making reports available and accessible to stakeholders. Reports should cover key phases or critical findings of an evaluation or accountability effort. In order to make reports appropriately available and accessible, consideration should be given to:

- various audiences or stakeholders, their different information needs about the effort, as well as their need in terms of developing relevant understanding and capacity,
- different formats for communicating and reporting (e.g., paper-based, electronic), and
- establishing a clear way for audiences or stakeholders to ask questions about information presented in a report and to possibly refute the appropriateness and adequacy of procedures or outcomes.

Link Between Evaluation and Accountability

A distinction between evaluation and accountability reports is that evaluation reports typically tend to target professional or technical audiences; that is, individuals who may be charged with revising and implementing the next generation of programs or judging the quality of the evaluation itself. As such, they need to have adequate technical information to demonstrate their levels of internal and external validity. Accountability reports typically are less technical because their primary audiences include many nontechnical groups and others not part of the education establishment. Such constituencies include legislators, parents, business groups, and classroom teachers.

The ways in which information from an evaluation or accountability effort is shared can significantly affect the ways in which audiences perceive, understand, and learn from the outcomes of the efforts. Therefore, upfront consideration of the set of reports that would best serve both the needs of key constituencies and the ultimate goal of educational improvement is critical. Such a set of reports could consist of nontechnical tables with some summary language or a range of reports geared to the interests of the various constituencies. A technical report also could be developed, indicating the reliability and validity of the judgments that are made relative to the findings of the evaluation or accountability effort. An executive summary targeted to policymakers could be developed to provide a summary of findings, conclusions, recommended next steps, and resource needs for the implementation of the next steps. Throughout the set of documents, a description of how results are to improve education (i.e., learning, teaching, and/or educational administration) should be included.

Summary and Conclusion

The value of educational evaluation or accountability is ultimately determined by the degree to which it contributes to the improvement of learning, teaching, and educational administration. For the purpose of this article, discussion focused on this common contribution of both evaluation and accountability and presented relevant information within a framework of validity and utility. As discussed in this article, threats to the validity and utility of educational evaluation or accountability efforts can be minimized through careful attention to a clearly defined purpose and scope, appropriate involvement of vital individuals or groups, consideration of information from multiple measures and data sources, providing context-sensitive outcomes, making reasonable and defensible interpretations of information that lead to adequate and appropriate actions, and providing reports that are available and accessible to appropriate stakeholders to help minimize such threats.

Establishing links between evaluation and accountability efforts can strengthen their relative contributions to educational improvement. The focus of accountability systems can drive determinations of what kinds of evaluations are likely to be particularly useful in enhancing understanding of accountability system findings and their implications for future policy and practice. Data from evaluations can inform the design of accountability systems as well as the understanding of relationships among key components of the systems and possible reasons for and conditions under which the desired standard has been achieved. These linkages suggest that careful communication and coordination among key stakeholders and purposeful coordination of

efforts are essential. Such communication and coordination between those designing and implementing evaluation and accountability systems can better ensure the validity and utility of information to guide improvements. Linking evaluation and accountability in this manner is more likely to increase the overall quality of our education system and the likelihood of achievement and equity for all our students.

See also: Educational Evaluation: Concepts, Practice, and Future Directions; Evaluation Governance and Planning; Evaluation Reporting and Communicating; Evaluation Use; Internal and External Evaluation; The Purpose of Educational Evaluation; The Role of Stakeholders in Educational Evaluation.

Bibliography

- Alkin, M. C. (2004). *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks, CA: Sage.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: Author.
- American Evaluation Association (2006). *Public Statement: Educational Accountability*. Fairhaven, MA: Author.
- Ananda, S. and Rabinowitz, S. (2001). *Building A Workable Accountability System*. San Francisco, CA: WestEd.
- Assessment and Accountability Comprehensive Center (2009). *Framework for High-Quality English Language Proficiency Standards and Assessments*. San Francisco, CA: WestEd.
- Baker, E. L., Linn, R. L., Herman, J. L., and Koretz, D. (2002). *Standards for Educational Accountability Systems (CRESST Policy Brief 5)*. Los Angeles, CA: Center for the Study of Evaluation, Standards, and Student Testing (CRESST).
- Betebenner, D. (2008). *Norm- and Criterion-Referenced Student Growth*. Dover, NH: National Center for the Improvement of Educational Assessment.
- Brennan, R. L. (ed.) (2006). *Educational Measurement*, 4th edn. Westport, CT: American Council on Education and Praeger Publishers.
- Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *Practice–Theory Relationship in Evaluation. New Directions for Evaluation* **97**, 7–35.
- Fuhrman, S. H. and Elmore, R. F. (2004). *Redesigning Accountability Systems for Education*. New York: Teachers College Press.
- Goldschmidt, P., Roschewski, P., Choi, K., et al. (2005). *Policymakers' Guide to Growth Models for School Accountability: How do Accountability Models Differ?* Washington, DC: Council of Chief State School Officers.
- Gong, B. (2002). *Designing School Accountability Systems: Towards a Framework and Process*. Washington, DC: Council of Chief State School Officers.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin* **112**, 527–535.
- Linn, R. L. (2001). *The Design and Evaluation of Educational Accountability Systems (CSE Technical Report 539)*. Los Angeles, CA: Center for the Study of Evaluation, Standards, and Student Testing (CRESST).
- Madaus, G. F. and Stufflebeam, D. L. (1984). Educational evaluation and accountability. *American Behavior Scientist* **29**(5), 649–672.
- Mathison, S. (1988). Why triangulate? *Educational Researcher* **17**(2), 13–17.
- Merriam-Webster (2003). *Merriam-Webster's Collegiate Dictionary*. Springfield, MA: Merriam-Webster.
- Messick, S. (1989). Validity. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn, pp 13–103. New York: Macmillan.
- Patton, M. Q. (2008). *Utilization-Focused Evaluation*. Thousand Oaks, CA: Sage Publications.
- Sanders, W. L., Saxton, A. M., and Horn, S. P. (1997). The Tennessee value-added accountability system: A quantitative, outcomes-based approach to educational assessment. In Millman, J. (ed.) *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* pp 137–162. Thousand Oaks, CA: Corwin Press.

Further Reading

- Haertel, E. and Herman, J. (2005). *A Historical Perspective on Validity Arguments for Accountability Testing (CSE Technical Report 654)*. Los Angeles, CA: Center for the Study of Evaluation, Standards, and Student Testing (CRESST).
- Pellegrino, J. W., Chudowsky, N., and Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.
- Scriven, M. (1999). The fine line between evaluation and explanation. *American Journal of Evaluation* **15**(1), 75–77.

Relevant Websites

- <http://www.eval.org> – American Evaluation Association.
- <http://www.aacompcenter.org> – Assessment and Accountability Comprehensive Center.
- <http://www.nciea.org> – Center for Assessment.
- <http://www.cpre.org> – Center for Policy Research in Education.
- <http://www.cse.ucla.edu> – National Center for Research on Evaluation, Standards, and Student Testing.

Evaluation Governance and Planning

R Picciotto, King's College, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This article outlines the prerequisites for effective evaluation governance and planning in the education sector. Evaluation governance consists of the policies, rules, and practices that initiate and control evaluations, while planning identifies the methods, tasks, and resources needed to carry them out. Together, the two functions contribute to improved performance in the education sector.

The two concepts are complementary and interdependent. First, governance structures shape evaluation planning by setting the rules of the game according to which the evaluation mandate is discharged. Second, evaluation planning influences governance by focusing on the institutional areas that need adjustment in order to achieve improved educational outcomes.

Conditions for Good Evaluation Governance

Good governance aims at legitimacy, accountability, responsibility, fairness, transparency, efficiency, and probity (OECD, 2004). Evaluation contributes to these objectives. It provides external guidance to the supreme authorities of education bodies. In addition, it enhances accountability and probity and makes authority responsible – a role akin to that played by accounting and auditing in the private sector. On the other hand, evaluation only comes into its own within a governance system that is legitimate, guarantees the independence and quality of the function, and reflects intentionality to make use of lessons of experience.

Policy

Evaluation informs not only managerial self-governance (rectors, presidents, deans, etc.) and academic self-governance (e.g., teaching and research staff) but also government ministries, national quality assurance bodies, national-level advisory or consultative bodies, international organizations (e.g., the European Commission), or educational associations (e.g., the Network of Universities from the Capitals of Europe).

Reporting lines are bound to vary since educational structures differ in how they allocate responsibilities and decision-making powers. Hence, evaluation policy must be adapted to the governance architecture in place. However, at the national level, good evaluation practice requires an

official mandate, that is, a policy statement that specifies the basic purposes and features of the evaluation function and its role. International professional associations may also be called upon to draw up relevant evaluation procedures to ensure that they meet commonly agreed standards.

Normally, an explicit evaluation policy is embedded in legislation endorsed by the supreme governing body of the state or region that lays down the rule of operation of the education sector. In the United States, the No Child Left Behind Act has set explicit proficiency goals for all students by 2014 that are tracked at the state level and also at the federal level by an agency that operates under the aegis of the US Department of Education.

Similarly, national standards for evaluation of academic staff may be imposed to induce trust in the quality of education and training. Thus, it is common practice for evaluations of academic staff to form an integral part of compulsory quality assurance procedures while, in order to attest to the overall effectiveness of educational institutions, formal evaluation arrangements may be made by the governing bodies that carry ultimate responsibility for all the affairs of the institution.

Additional safeguards may be provided through independent evaluations contracted by the official bodies that provide educational funding. In Estonia, for example, accreditation requirements established by government state that higher educational institutions need to use criteria and methods that take all fields of activity into account (teaching, students' instruction, research, special tasks, etc). Similarly, evaluation arrangements for the overall effectiveness of education institutions may be prescribed by legislation.

In England and Wales, higher education institutions are legally independent corporate institutions and their councils or board of governors are formally tasked with contracting for quality audits that review the mechanisms and structures designed to monitor and ensure academic quality and standards. In addition, official funding bodies have a statutory duty to ensure the quality of instruction in individual subject areas through quality assessments.

It is a good practice for the legal framework that governs educational institutions to specify evaluation independence, stakeholder involvement, transparency of evaluation processes (including public disclosure of evaluation reports), and the relationships between the independent evaluation functions and the quality-assurance procedures that are the responsibility of the educational establishment.

Rigorous evaluation of academic staff is a characteristic of sound educational governance. In Latvia, evaluation of academic staff occurs before hiring and reevaluation of incumbents of academic posts occurs every 6 years based on the criteria set by the Cabinet of Ministers. In Finland, the performance-based salary scheme is based on the evaluation of two components – evolving job requirements and individual performance. In Romania, the methodological standards for the annual evaluation of higher education staff are set according to performance indicators established by the Ministry of Education and Research.

Independence

The objectivity and professional skepticism that should characterize evaluations cannot be secured unless evaluators are shielded from external threats and influences regarding the collection of information and the assessment process. Sound governance safeguards the independence and impartiality of evaluators. It shields them from the demands and biases of program managers and policy-makers. It provides them with mandatory access to relevant information and full autonomy in carrying out their investigations and reporting their findings. It protects them from intimidation, organizational pressures, financial leverage, and political influence.

Equally, evaluation independence requires distinctive personal characteristics, attitudes, and behaviors among individual evaluators and a frame of mind characterized by curiosity, skepticism, and hunger for evidence. However, evaluators should not be so detached from the activity being evaluated as to shirk interaction with operational managers, staff, or beneficiaries of the activity: independence is not isolation. Good evaluation overcomes information asymmetries, while protecting the integrity and fairness of the process from the influence of vested interests.

Four standards characterize evaluation independence (World Bank, 2007):

1. *Organizational independence.* It ensures that the evaluation unit and its staff are not under the control or influence of decision makers who have the responsibility for the activities being evaluated and that they have full access to the information they need to fulfill their mandate.
2. *Behavioral independence.* It measures the extent to which the evaluation unit is able and willing to set its work program, produce high-quality and uncompromising reports and to disclose its findings to the supreme governance authority without management-imposed restrictions.
3. *Protection from outside interference.* It keeps the evaluation function free to set its priorities, design its processes and products, reach its judgments, and administer its

human and budget resources without intrusion by management.

4. *Conflict of interest safeguards.* They guarantee that current, immediate future, or prior professional and personal relationships and considerations are not allowed to influence evaluators' judgments or create the appearance of a lack of objectivity.

All four criteria are important and none implies trade-offs with operational relevance or influence. In particular, behavioral independence is a privileged dimension of evaluation excellence since it is integrally connected to evaluation quality and can be ascertained not only by assessing governance structures, processes, and practices but also by examining whether, through effective evaluation planning and management, the independent evaluation unit produces quality evaluations.

Quality

Areas typically covered by educational evaluations include the institution's arrangements for designing, approving, and reviewing programs of study, grading protocols and degree classification, as well as internal feedback and monitoring arrangements. To carry out such evaluations, appropriate skills, sound methods, adequate resources, and transparency are required.

Intellectual engagement with the suppliers and beneficiaries of educational services are equally critical to evaluation effectiveness since accurate and fair evaluations combine detachment with empathy and deep understanding. Hence, the importance of a governance system that provides connectivity to stakeholders in order to secure their trust without jeopardizing the integrity and objectivity in the evaluation process.

The implications for organizational design are straightforward. In order to enhance effective use of educational resources, accountability for results, and learning from experience, all stakeholders – parents, teachers, children, school administrators, etc. – as well as political authorities and the general public have an interest in the quality of educational evaluations. In all European countries, the law requires that students be involved in evaluating the quality of teaching staff; in Italy, it is the only method currently used to this end.

With regard to the evaluation of educational programs, whether the emphasis is placed on formative or summative evaluations, stakeholders look to evaluation studies to generate independent and rigorous assessments of educational policies, programs, or projects, their design, their execution, and their outputs, outcomes, and impacts. (Formative evaluation aims at improved program design and performance, whereas summative evaluation helps to determine the extent to which planned outcomes and impacts have been achieved. The former kind of evaluations often takes

place during implementation so as to make use of lessons learned for improved outcomes, whereas the latter types are usually conducted at the end of a program phase to assess performance and assign responsibility for outcomes.) In the United States, evaluations are expected to measure educational impacts in a rigorous way through experimental or quasi-experimental methods.

Compliance with good professional practice means impartiality and absence of bias. Evaluators reach balanced judgments by relying on relevant and accurate evidence about program strengths and weaknesses. To this end, the views of program participants, partners, and beneficiaries should be sought. Where stakeholders wish to challenge the analysis contained in an evaluation, they should be given an opportunity to add dissenting footnotes or statements.

Monitoring

A responsibility of educational managers, monitoring, consists in the continuous assessment of progress made during program implementation and in the design of performance-related salary schemes. Monitoring relies on baseline data and on regular information about the use of inputs for major activities, their outputs, their expected outcomes, and, to the extent practicable, their impact. By providing timely information on program performance and by tracking staff compliance with standards, monitoring provides relevant feedback to education managers, staff, and partners and pinpoints the actions needed to improve performance toward agreed goals.

Thus, monitoring promotes organizational learning in real time. For effective dissemination, reporting is clear, concise, and easily accessible. Good monitoring systems use specific, measurable, attainable, relevant, and time-bound (SMART) indicators that reflect the values and policies embedded in the agreed education program. Toward this end, data collection is carried out professionally and at an appropriate frequency so as to provide timely feedback and facilitate proactive management.

Participation, Transparency, and Ethics

Embedding participation within the evaluation process is a prerequisite of fairness, objectivity, and utilization of evaluation findings. Principled consultation with judiciously selected stakeholders at critical phases of the evaluation process (from initial design to completion follow-up) should be an integral part of the evaluation plan as well as a standard feature of monitoring systems.

Thus, the partners that contribute to the governance, management, and funding of education programs should be consulted, and so should the beneficiaries and communities affected by the education program being evaluated. Special efforts should be made to include organizations

and individuals that represent disadvantaged and neglected groups. Such participation enhances the credibility of monitoring and evaluation arrangements and facilitates buy-in of evaluation lessons.

The reach and impact of evaluation are increased through explicit provision for interaction with stakeholders through advisory bodies, learning events, focus groups, formal surveys, and structured consultations at key stages of the evaluation process. Through such means, stakeholders have an opportunity to comment on the design, methods, findings, and conclusions of the evaluation. It is also desirable for the salient contributions of stakeholders to be transparently displayed in the final report.

Participation is fortified by transparency that enhances ownership of evaluation findings and helps to ensure that lessons drawn from evaluation become lessons learned. Through evaluation transparency, stakeholders and the general public are informed about the evaluation objectives, criteria, and planned use of findings, and information is made available about the confidentiality rules that will govern the evaluation for different categories of stakeholders and at various phases of the process.

The basic presumption should be that stakeholders and the general public have a legitimate claim to accessing evaluation results. Equally, administrators and decision makers should be tasked with informing their constituencies about the decisions taken in light of evaluation findings, for example, with respect to future program design and implementation.

This said, evaluators must strike a balance between the demands placed on them and the rights and sensitivities of those who contribute to the evaluation – or are potentially affected by its results. For evaluative decisions to be made in a way that respects the interests of students, parents, and the public as well as those of educational bodies and their staff, full transparency about the purpose of the evaluation is important.

As in social research, ethical considerations should guide the evaluation process. The rights to privacy of evaluation respondents should be protected and informed consent sought from those who deliver sensitive information about an education program, neither should evaluations contribute to a chilling effect on the openness of educational processes. Confidentiality rules should be ensured, for example, if commercial considerations intervene or access to data is restricted for security reasons, potential retaliation against respondents, or possible embarrassment of participants.

Anonymity of information sources should be respected in the processing of surveys, and evaluators should seek permission before accessing an educational establishment and agreements reached regarding protocols of engagement with pupils, teachers, and administrative staff. Special safeguards are needed in field work that involves children.

Final evaluation reports should be disclosed to the public and a reasoned judgment about the reliability and independence of the evaluation should be included in the final reports.

Such reports should include adequate information about the evaluation commissioning process, the evaluation framework, the financial and human resources employed, their provenance, the data collection arrangements, the assessment methods selected, and their limitations. The responses of the administration to evaluation recommendations should also be made public, and actual implementation of agreed recommendations should be tracked and disclosed.

Beyond the immediate stakeholders, the new knowledge generated by evaluations should be disseminated to the policy community. To this end, documentation resulting from educational evaluations (including abstracts and syntheses of performance data and evaluation lessons) should be clearly presented, be user friendly, and be tailored to diverse audiences. Evaluation accessibility is enhanced by the new information technologies (including websites) as well as the judicious use of workshops, conferences, and media events.

Conditions for Sound Evaluation Planning

Educational institutions must ensure that they remain responsive to the needs and demands of society. Strategic plans are the instruments through which educational priorities and activities are set. Educational strategies state the vision and specify the policy directions of education programs. They also outline educational objectives, means of implementation, and review processes. This is where evaluation fits in. By feeding into the strategic planning process, it bridges educational governance and strategy.

At the levels of both the evaluation program and individual studies, planning is an iterative process. It depends not only on the issues to be addressed but also on the feasibility of addressing them. Allocation of scarce evaluation resources is a major responsibility of evaluation managers. It calls for expert familiarity with evaluation methods and constraints as well as for a sound understanding of the policy trends and the organizational context within which educational programs are designed and implemented.

Evaluation Programming

At the level of the evaluation program, resource allocation within an evaluation unit concerns the choice of evaluation studies and activities and the deployment of staff and funds to implement them within a given time span. This requires evaluation managers to make assumptions

about the evolution of the policy and operating context. A far sighted approach is needed since the evaluation program will generate a result stretching well beyond the programming period, and continued relevance of the evaluation program calls for skills and methods that are tailor-made to the evolving programmatic challenges.

It takes time to nurture the human assets and the relationships needed for high-quality evaluations. Hence there is a need for evaluation strategies. They help evaluation managers to identify the evaluation skills and the knowledge partnerships needed to comply with the evaluation mandate. Evaluation strategies are also used to secure sustained support from the governing authorities regarding medium-term evaluation budget plans. They reflect the evaluation managers' assessments of the strengths and weaknesses of the institutional environment.

In setting priorities for the evaluation program, evaluation managers are also guided by their policy mandate, by the lessons of evaluation experience, and by the results of their regular consultations with decision makers and other stakeholders. As organizational and policy contexts evolve, evaluation strategies need periodic adjustment while evaluation plans address shorter time periods. Where operating environments are characterized by volatility and uncertainty, flexible rolling programs may need annual updating to help redeploy scarce evaluation resources in a timely fashion (Mathison, 2007).

Evaluation programming should meet four standards of quality:

1. criticality – a portfolio of evaluations that together assess educational performance in an objective and transparent fashion;
2. additionality – a distinctive contribution to operational knowledge creation or dissemination within the education sector;
3. timeliness – the delivery of educational evaluation findings and lessons early enough to inform decision making; and
4. materiality – a deliberate focus on topics and issues that have substantial relevance to educational effectiveness.

Planning of Evaluations

Once the diverse audiences and goals of the evaluation program have been ascertained and specific choices have been made as to which evaluations should have priority, the focus of evaluation planning turns to the design of individual evaluation studies. At this level, the nature of the questions posed and their precise formulation may have to be adjusted to take account of practical considerations. As for the overall program, this implies a cyclical process designed to reconcile the scope, content, and timing of the evaluation with the constraints imposed by available funds, skills, data, and methods (Weiss, 1998).

Evaluability assessments can help to define the appropriate direction, timing, and content of the study by examining the program to be evaluated. Typically, such assessments involve an overview of the program, a cursory examination of available data, a preliminary survey of stakeholders' perceptions, and a tentative judgment as to the likelihood of program success. This approach is a good way of involving managers and stakeholders early in the evaluation process. It also helps to delineate the precise timing of the evaluation and the roles it is expected to play in improving program outcomes.

Thus, evaluability assessments may be an economical way of generating improved program design recommendations and/or enhanced monitoring arrangements. In some cases, they may preclude, preempt, or postpone the need for the immediate or comprehensive evaluation studies that were originally envisaged. At a minimum, they can help clarify the key assumptions that underlie program design, consider their plausibility and, as a result, help to identify relevant evaluation questions and to select appropriate evaluation methods.

Evaluation quality calls for an evaluation framework that identifies the most relevant stakeholders and defines an appropriate role for them in the evaluation process. The framework should also include clear and consistent evaluation objectives, agreed evaluation criteria, relevant baseline and performance indicators, judicious selection of evaluation methods, data collection and interpretation arrangements, reporting formats, and dissemination plans. Next, the skills and resources needed to bring the evaluation process to fruition should be assembled. These are the challenges that evaluation planning is expected to meet.

Evaluation coverage and terms of reference depend on the criteria to be used and the nature of questions to be addressed. These may concern one or more of the following: rationale for program scope, content and/or timing, program goals and/or objectives, choice of program instruments, relevance, effectiveness and/or efficiency of processes, compliance with regulations, performance of participants, role of contextual factors, selection of success indicators, etc. Since stakeholders usually want to know not only what happened but also how and why, desired attribution of outcomes to particular program features or to the behavior of key program participants is critical to the choice of evaluation methods and the scope of data collection.

Timing and Scope

The optimum timing of an evaluation is not a straightforward matter. The longer the time used to ascertain results the more secure the assessment of program impacts. However, accountability considerations call for early organizational feedback so as to create wholesome incentives for program participants and provide the

timely knowledge needed to adjust program designs and fine-tune the program processes. Furthermore, the role of unanticipated events and the emergence of unintended consequences tend to increase as time horizons expand.

Similarly, the scope of the evaluation must strike a balance between what is desirable and what is realistic and achievable. Evaluation techniques should be chosen to ensure as complete, fair, and rigorous answers to the questions posed as practicable. But the larger the scope of the evaluation, the harder the methodological and data challenges, and hence, the greater the risks to external and internal validity of evaluation findings. Threats to external validity occur when evaluation conclusions are extrapolated beyond their specific contexts while threats to internal validity (the correct attribution of results to particular program features) result from changes in operating contexts, assignment bias, or unintended impacts on participants' or control group behaviors.

In deciding on the direction and scope of an evaluation, evaluation managers may establish peer review mechanisms or set up advisory committees or reference groups composed of sector specialists and expert evaluators. Such mechanisms can be used for quality assurance and for advice on methods, data sources, and interpretation of results. They answer the perennial question of who evaluates the evaluators. It is desirable to provide them with adequate resources and to disclose to the public their judgments about the quality of the final evaluation report.

Methods

No standard blueprint exists for selecting the right evaluation methods: the context of the evaluation and its purposes should determine its methodology. For example, if the effects of a specific intervention have to be demonstrated with rigor, its observed results should be compared with those that would have materialized without the intervention.

Where feasible, the counterfactual should be derived through randomized control trials that measure the net impact of an educational intervention by the differences observed between learning outcomes within an experimental group subject to the intervention and those secured by a group (designated as the control group) deprived of the intervention, endowed with the same characteristics and chosen through a random selection process.

A variant method is regression discontinuity. It selects the experimental and control groups on the basis of a pre-tested characteristic measured before and after the intervention: the shift between the regression lines constitutes the net effect of the education program. Interrupted time series assess the impact of various features of an educational intervention by looking at changes in learning variables.

When random allocation is not feasible, quasi-experimental designs are used by selecting subjects that closely match the experimental group (Scriven, 1991). Matched comparisons ascertain the desired characteristics of a control group excluded from the intervention against those of program participants. Propensity scores that reflect the likely importance of contributory factors to outcomes (based on reviews of the research literature) can be used to select the experimental and control groups. Such experimental methods involve ethical risks and tend to be costly and skill intensive.

They can be useful for standard educational treatments in stable environments, but in the real world, unless rigorously handled, they can yield misleading results because of selection bias (unrepresentative treatment group), substitution bias (access to other treatments in the control group), and unintended behavioral consequences within the treatment group (Hawthorne effect) or within the control group (John Henry effect).

As they do not invariably deliver what stakeholders are interested to know – for example, the what and how of program design and the distinctive contributions of educational actors and participants – experimental methods cannot be described as a gold standard for evaluation. Depending on the evaluation objectives, other methods may be more useful. In general, the limitations of any single method may be mitigated through triangulation of data collection instruments and multi-method approaches.

Longitudinal evaluations based on information collected over a period of time (from the same group of educational respondents) may be combined with cross-sectional evaluations that assess the results obtained by different respondent groups (at the same time). Alternatively, or in addition, trend studies may focus attention on changes in key factors that are deemed to impact on educational outcomes.

Case studies are often used to evaluate the effectiveness of an educational policy (Cohen *et al.*, 2007). Well-selected and -designed case studies have the capacity of illuminating program performance by providing rich narratives of policy applications in diverse operating contexts, analyzing the linkages between policy actions and results, and identifying the contribution of various actors to observed outcomes.

Participatory approaches favor a combination of interviews, questionnaires, focus groups, tests, and simulations. For example, in evaluating the behavior of a partner engaged in an educational endeavor, it is appropriate to study it from different perspectives making use of stakeholder perception surveys as well as of qualitative information and quantitative data that test performance according to agreed benchmarks.

Similarly, in evaluating the rationale of an educational technique, theory-based approaches have considerable merit: they model the theory or theories of change

implied by the intervention, examine critically their explicit or tacit assumptions, and seek to verify the extent to which they are valid.

For example, the theory-driven evaluation of a program aimed at reducing problem behavior in the middle school of a US southern state demonstrated conclusively that improvements in social bonding, social competency skills, and school success did not materialize due to a patent lack of academic and administrative support for the changes proposed in instruction and mentoring (Rossi *et al.*, 1999).

In a globalized world, public policy benchmarks are increasingly defined with reference to international comparators. But defining such benchmarks in ways that are comparable across countries without undermining their validity at the national and institutional level poses major methodological challenges. Hence, without a detailed understanding of the specific contexts within which educational services are delivered, lessons are not transferable from one country to another.

Contracting and Managing an Evaluation Team

Evaluation relies on all the social sciences. As a trans-discipline, it connects all the disciplines relevant to the assessment of merit, worth, and value of a collective endeavor. It is also a discipline in its own right since it has developed its own concepts and practices. Thus, the skills needed to carry out evaluations are diverse and no single individual can be expected to master the full repertoire of social research findings and evaluation methods. Nor can formal evaluation training, critical as it is, substitute for learning by doing evaluations.

Evaluators' competence largely determines evaluation quality. Hence, the assembly of appropriate skills and experience to carry out an evaluation assignment and deliver a quality study is a key evaluation management challenge (Rossi *et al.*, 1999). Team leaders and members should have the competencies (e.g., educational background, evaluation qualifications, professional skills, and work experience) needed to implement the evaluation according to agreed quality standards. Trust and mutual respect should characterize the relationship between commissioners of evaluations and the evaluators that they employ.

A multidisciplinary team composed of external subject matter specialists and evaluators may have to be constructed. Where an external group or a consulting firm is employed, transparent (preferably competitive) selection procedures should be used. These procedures should avoid conflicts of interest and provide a level playing field for potential bidders. The external parties who undertake to carry out an evaluation should endorse a written agreement

to fulfill the terms of reference of the evaluation and meet the conditions set forth in a legally enforceable contract that addresses matters such as confidentiality, privacy, reporting, as well as mediation procedures, in case disagreements arise or misconduct is alleged.

Summary

Good governance of evaluation in education is embedded in policies and standards that protect the independence of the function and promote impartial and rigorous assessment of educational programs. It involves stakeholders in the evaluation process through open and transparent processes. It ensures regular tracking of implementation progress toward agreed goals through effective monitoring and it encourages utilization of evaluation lessons by program managers. Finally, it complies with ethical standards that protect the rights of stakeholders that are asked to contribute to the fulfillment of the evaluation mandate. While sound governance is critical for evaluation effectiveness, it is not a sufficient condition of evaluation excellence. For evaluation resources to be allocated judiciously, sound evaluation planning feeds into the education strategy cycle so that lessons of experience inform education program design and implementation. Tailor-made approaches adapt the scope, content, and timing of evaluations to stakeholders' needs and educational priorities. Sound evaluation management also calls for

triangulation of evaluation methods and the assembly of adequate evaluation competencies.

See also: International Evaluations; National Assessment Programs: The Example of the U.S. National Assessment of Educational Progress.

Bibliography

- Cohen, L., Manion, L., and Morrison, K. (2007). *Research Methods in Education*, 6th edn, pp 253–263. Oxon: Routledge.
- OECD (2004). *OECD Principles of Corporate Governance*. Paris: Organization for Economic Cooperation and Development.
- Rossi, P. H., Freeman, H. E., and Lipsey, M. W. (1999). *Evaluation: A Systematic Approach*, 6th edn, p 106. Thousand Oaks, CA: Sage.
- Scriven, M. (1991). *Evaluation Thesaurus*, 4th edn, pp 102–103. Newbury Park: Sage.
- Weiss, C. (1998). *Evaluation*, 2nd edn, pp 73–91. Upper Saddle River, NJ: Prentice Hall.
- World Bank, Independent Evaluation Group (2007). *Sourcebook for Evaluating Global and Regional Partnership Programs: Indicative Principles and Standards*, pp 5–14. Washington, DC: IEG-World Bank.

Further Reading

- Eurydice (2008). *Higher Education Governance in Europe*. Brussels: European Commission.
- Mathison, S. (ed.) (2005). *Encyclopedia of Evaluation*, pp 316–317. Thousand Oaks, CA: Sage.

Evaluation Use

V G Thomas, Howard University, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Conceptual use – Using evaluation to better understand a program, policy, or related issues; issues; use to influence how key people think about a program or policy and better understand it in some significant way; no action or decision flows from the findings.

Evaluand – The thing that the evaluator is evaluating.

Evaluation misuse – Calculated and intentional suppression, misrepresented, or unbalanced use of evaluation findings to influence opinions or decisions.

Evaluation use – Application of evaluation information or evaluation process to achieve intended, desired ends.

Instrumental use – Using evaluations to directly inform a decision, improve a program or policy, develop new directions, or contribute to solving a problem.

Process use – Learning or changes resulting from participating in the evaluation process itself.

Strategic use – Using evaluation to persuade others or to gain particular outcomes.

Symbolic use – Token support for the evaluation process or findings with no real intent to take either the process or the findings seriously.

Utility standards – One of the four areas around which the 30 program evaluation standards are organized; the utility standards, consisting of seven standards, are intended to provide guidance that ensure that an evaluation will serve the information needs of intended users.

Use, or utilization, has long been recognized as a central issue in the evaluation field either as a goal of the evaluation or as a fundamental construct for discussion in theories, approaches, and research on evaluation. (In this article, as is generally the case in the evaluation literature, the terms use and utilization are used interchangeably and applied in a synonymous manner.) As an applied field, evaluation must be deeply concerned about its impact and the discipline rightfully should promote practices that advance utilization among intended audiences. During the late 1960s and 1970s, the use concept was generally accepted as a concise way of conveying

how evaluation findings affected programs and policies in ways that were clear, direct, obvious, and reasonable. Since then, discourse on evaluation use has evolved considerably. A reconceptualization of evaluation use emerged as the landscape of the field became more diversified and as increasing numbers of evaluators recognized that evaluations did not necessarily translate into immediate or direct decisions to continue, expand, eliminate, or revise programs and policies. The limited use of evaluation results, especially for improving programs, has been characterized as the Achilles heel of the profession. To address this issue, in the late 1970s and 1980s numerous studies were conducted that examined utilization with this period referred to as the golden age of research on evaluation use (Henry and Mark, 2003).

With increased theoretical emphasis and empirical study on evaluation use, by the end of the 1980s, the concept had expanded to include a more complex, multifaceted perspective beyond emphasis on use for overt decision making to also encompass other types of utilization such as conceptual, strategic, political, symbolic, and organizational learning. There was more attention given to effective utilization of evaluation rather than simply more utilization (Weiss, 1988). Further, issues related to evaluation misuse were infused into the dialog (Alkin and Coyle, 1988; Patton, 1998; Shadish *et al.*, 1991). Some of this discussion focused on the relationship between use and misuse and inherent challenges of differentiating misuse from use, particularly in view of conflicting values and disagreements about means and ends of social programs (Shulha and Cousins, 1997).

Despite changing perspectives on evaluation use, it has maintained a significant presence in the field and is listed first as one of four attributes around which the thirty program evaluation standards (Joint Committee on Standards for Educational Evaluation, 1994) are organized. These standards, currently in their second published edition with ongoing field testing of the third edition, initially emerged as the result of a 1975 pioneer project whose goal was the development of standards to help ensure useful, feasible, ethical, and sound evaluations of educational projects, programs, and materials. The standards were designed to serve as a basis for program evaluation design and implementation in North America; however, they have been translated into Spanish, German, French, and other languages for use around the world. Taken as a set, the program evaluation standards provide a working philosophy for evaluations in fields beyond education and

Table 1 Utility standards of the program evaluation standards

<i>Utility standards</i>	<i>Standard description</i>
U1 Stakeholder identification	Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed
U2 Evaluator credibility	The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance
U3 Information scope and selection	Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders
U4 Values identification	The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear
U5 Report clarity	Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood
U6 Report Timeliness and dissemination	Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion
U7 Evaluation impact	Evaluations should be planned, conducted, and reported in ways that encourage followthrough by stakeholders, so that the likelihood that the evaluation will be used is increased

From Joint Committee on Standards for Educational Evaluations (1994). *The Program Evaluation Standards*, 2nd edn. Thousand Oaks, CA: Sage.

they identify principles that, when addressed, are expected to result in improved evaluation practice. The seven utility standards, in particular, provide guidance for examining programs, projects, and materials in a manner that ensures that the evaluation serves the information needs of intended users. These standards are summarized in **Table 1**.

In addition to the prominence of utility in the program evaluation standards, it is also one of the key components highlighted in the mission, vision, and values statements of evaluation professional organizations, including those articulated by the American Evaluation Association (AEA), African Evaluation Association (AfrEA), Australasian Evaluation Society (AES), European Evaluation Society (EES), UK Evaluation Society (UKES), and the Canadian Evaluation Society (CES). Some of these organizations have special sections or groups that explicitly address issues related to use. For example, the AEA has a Topical Interest Group (TIG) on Evaluation Use, which is the conceptual strand of the organization focusing specifically on evaluation theorists, researchers, and practitioners interested in the nature, causes, and consequences of evaluation utilization.

In the United States, education and evaluation have long been closely connected as numerous educational interventions were initiated, especially beginning in the 1960s, with the goal of buffering the negative effects of poverty and improving educational outcomes of children from improvised backgrounds. Now into the twenty-first century, increasingly, legislators have been calling for more useful educational evaluations by urging that these studies meet high standards of scientific quality with the US Congress writing these standards into national legislation (Ginsburg and Rhett, 2003). With the push for evidence-based interventions in education, both within

and outside the US, decision makers are seeking to use evaluation results as an accountability tool that impacts the expenditure of funds related to the creation, expansion, and elimination of programs. Two examples of this increasing emphasis on educational accountability include the No Child Left Behind Act of 2001 in the US and the Education and Inspections Act in the UK which, in April 2007, established the Office for Standards in Education, Children's Services and Skills (OFSTED).

The following sections examine evaluation use, considering how it was originally conceptualized to examination of more recent perspectives encompassing a multidimensional, multifaceted view which has arguably led to an expanded understanding of the concept. Two fundamental, yet overlapping, issues that will be considered in greater detail include the centrality of use in defining evaluation success and the appropriateness of evaluation use terminology. Subsequently, a discussion of evaluation use in the field of education is provided. The article concludes with brief attention to particular issues related to evaluation use for improving the educational outcomes of students from diverse and disadvantaged settings.

Understanding Evaluation Use: Past to Present

Over the past four decades, an impressive body of literature has been published on evaluation use. Carol Weiss published the first paper on use of findings and a decade later she (Weiss, 1988) expanded her discussion of use (and sometimes nonuse) focusing on three important questions: (1) What do we mean by use? (2) What is it that is used? and (3) Who are the users? Weiss expressed concerns about underutilization and called for more empirical study of

the topic. Since then, there have been numerous developments in evaluation theory, research, and practice related to use. Notable among these include: (1) reconsideration of the use terminology, (2) expanding its scope to include use that is noninstrumental in nature, (3) recognition of contexts as critical to understanding and explaining use, (4) expansion of conceptions of use from the individual to the organizational level, (5) diversification of the role of the evaluator as facilitator, planner, and educator/trainer, and (6) consideration of the centrality of use in defining evaluation success (Kirkhart, 2000; Henry and Mark, 2003; Patton, 1998; Shulha and Cousins, 1997).

By the 1980s, there existed a coherent body of published literature on evaluation use, including several reviews and syntheses of literature shaping evaluation utilization as a field of inquiry. Studies on evaluation use identified predictors that facilitated use by intended uses. The major predictors identified include: (1) relevance, (2) credibility, (3) user involvement, (4) communication effectiveness, (5) potential for information processing, (6) clients' need for information, (7) anticipated degree of program change, (8) perceived value of evaluation as a management tool, (9) quality of evaluation implementation, and (10) contextual characteristics of the decision or policy setting (Shulha and Cousins, 1997).

Work by Cousins and Leithwood (1986) sought to go beyond simply describing predictors to assessing the relative weight of various factors in their ability to predict evaluation use. Utilizing a meta-analytic approach that considered empirical research conducted over a 15-year period on use of evaluation results, the investigators examined 65 studies in the fields of education, mental health, and social services. Based upon the results of the meta-analysis, Cousins and Leithwood developed a conceptual framework incorporating 12 factors influencing evaluation use. Six of these identified factors were associated with implementation (i.e., variables associated with designing, conducting, and reporting the evaluation) and six factors were related to decision or policy setting (i.e., variables addressing the environment in which the evaluation occurred). The six factors identified as influencing evaluation implementation include: (1) evaluation quality or characteristics of the evaluation process such as methods, rigor, and evaluation model; (2) credibility of the evaluator and process, defined in terms of objectivity and believability; (3) relevance of the evaluation to the information needs of the decision makers and the location of the evaluator; (4) communication quality, including clarity of reporting to evaluation audiences in terms of style, evaluator advocacy, and breadth of dissemination; (5) findings, that is, whether results aligned with evaluation audience expectations and their value to decision making; and (6) timeliness of the dissemination of results to decision makers. The six factors related to the decision or policy setting include: (1) information needs of the

evaluation audience(s), including the number of audiences and the perceived need for evaluation; (2) decision characteristics involving the type of decision to be made and the significance of the decision or evaluation problem; (3) political climate encompassing the political orientation of the evaluation sponsors, power struggles, and fiscal issues; (4) competing information from sources beyond the evaluation; (5) personal characteristics relative to the decision makers' organizational experience and social characteristics; and (6) commitment or receptiveness to evaluation as evidenced by the attitudes of decision makers to evaluation, organizational resistance, and open mindedness. Ranking the importance of the influencing factors on overall use as well as by various types of use, Cousins and Leithwood found that evaluation quality and decision characteristics were the most prevalent factors influencing evaluation use.

While it is beyond the scope of this article to provide a comprehensive discussion of the various issues and debates that emerged from the literature on evaluation use, two areas are examined in more detail in this article: (1) consideration of the centrality of use in judging evaluation success and (2) re-conceptualization of the use terminology.

Centrality of Use in Judging Evaluation Success

Use of results is generally one of the key assumptions underlying the planning, implementation, and dissemination of an evaluation. Traditionally, it has been assumed that a good evaluation will yield information that will be used by decision makers to take some future action. An area of debate in the field, however, relates to the centrality of use in determining the success of an evaluation. One perspective is that use is the criterion by which the success and worth of an evaluation should be defined and evaluation studies are only worthwhile if they are used.

Patton (1988) argues that evaluations should be judged by their utility and actual use. Evaluation use presents, maintains Patton, both a challenge and mandate. The underlying premise of his utilization-focused evaluation approach is that evaluators should facilitate the evaluation process and design the evaluation with careful consideration of how everything that is done, from beginning to end, will affect use. Here, use involves how real people in the real world apply evaluation findings and experience the evaluation process. Utilization-focused evaluations stress the need to move from general and abstract (e.g., from potential audiences and potential uses) to the real and specific (e.g., actual primary intended users and their explicit commitments to concrete, specific uses). Patton's emphasis was on intended use, both instrumental and conceptual, by intended users. Intended use can be classified as either instrumental or conceptual and it covers

three potential uses, including: (1) using evaluations for making overall judgments about the value, merit, and worth of program which often fulfills accountability mandates; (2) using evaluations results for program improvements; and (3) using evaluation to generate knowledge or influence thinking about the issues in a more general way.

While Weiss (1988, 1998), like Patton, acknowledged the importance of evaluation use and the need to increase use by intended users, particularly decision makers, she took a different position arguing that the success of an evaluation or the evaluator need not be tied to direct use. Evaluators, Weiss argues, should not be held accountable for the nonuse of results. She took this position acknowledging that various obstacles within a given context, having nothing to do with the evaluation or evaluator, could prevent use. These obstacles include things such as conflicting beliefs within the program's organization, disagreement over use of resources, rigidity of organizational rules and standard operating procedures, and shifts in external conditions, including budget cuts or changes in the political climate. Intense focus on the use of evaluation findings is not always viewed in a positive manner. In fact, Scriven (1991) points out that overemphasis on use can create a strong conflict for evaluators because it places undue pressure on them to adjust findings to what decision makers are willing to do rather than what they should do based on the evaluation evidence.

Others also question the legitimacy of use as a defining goal of the success of an evaluation. Henry (2000) maintains that the field does not suffer from a lack of relevance, but from heavy focus on emphasizing the short-term use. Social betterment, not use, argues Henry and colleagues (Henry, 2000; Henry and Mark, 2003; Mark and Henry, 2004), is the ultimate goal of an evaluation. Here, useful evaluations are thought to be ones that provide persuasive information that informs policies and programs aimed at alleviating human suffering and improving social conditions. Evaluations contribute to the pursuit of social betterment, according to Henry and Mark, by providing evidence and insights in determining the common good (which may be raising a social problem), selecting a course of action (which may be creating a policy or program), and adapting the course of action (which may be improving a policy, program, or organization).

Moving Beyond Use Terminology

Criticism has been lodged against the use terminology. The concept has been viewed as overgrown in the sense that the multiple forms of use (instrumental, conceptual, symbolic, and process) overlap and, in some instances, use is treated in a descriptive manner (referring to whether some consequence of evaluation did happen) and in other instances use is treated as normative concept use for

guiding the purpose of evaluation (Mark and Henry, 2004). Kirkhart (2000) maintains that the terms use and utilization have a conceptual and linguistic base that is too narrow because they promote an inappropriate imagery of instrumental use that focuses on a unidirectional, episodic application of evaluation findings. As such, she views this terminology an imprecise fit when considering nonresults-based applications and argues that work reflecting this terminology fails to adequately consider unintended results and the gradual incremental impact over time.

It has been recommended that the field abandon the use terminology replacing it with influence as the preferred concept. Influence, described by Kirkhart, is the capacity or power of persons or things to produce effects on others by intangible or indirect means. This is viewed as a broader concept that helps to create a framework for examining evaluation effects that are multidirectional, incremental, unintentional, and noninstrumental, as well as those that are unidirectional, episodic, intended, and instrumental. Taking this position, Kirkhart (2000) conceptualized an integrated theory of influence as a paradigm for rethinking the impact of evaluation by not simply holding results-based influence at the center, but instead, having multiple vantage points (instrumental, symbolic, and conceptual) for examining evaluation's influence. Kirkhart's theory of influence incorporates three overlapping dimensions: (1) the source or the active agent of change or starting point of a generative process of change; (2) the intention or the extent to which there is purposeful direction to exert a particular kind of influence through evaluation process or findings; and (3) the timing of influence, that is, if the influence is immediate, end of cycle, or long term. A theory of influence, argues Kirkhart, is good for the evaluation profession by making various types of influence visible and facilitating greater understanding of long-term evaluation impact which invariably builds credibility for the profession and generates support for evaluation.

There have been other criticisms lodged against the use concept and further expansion of frameworks emphasizing influence as an appropriate starting point of discussion. Henry and Mark (2003) view use as an inadequate motivation for evaluation and they maintain that emphasis on maximizing use is an inappropriate goal. They argue that a focus on use fails to provide a moral compass and it falls short of recognizing and promoting the ultimate purpose of evaluation – social betterment. Drawing concepts from the social science literature, Henry and Mark (2003) and Mark and Henry (2004) developed a theory of change to provide a more detailed and complex framework of influence than the perspective offered by Kirkhart. While Henry and Mark supported many of the arguments put forth by Kirkhart, they noted that her framework and other work in the field ignored the underlying mechanisms through which evaluation may achieve its influence. Attention to underlying

mechanisms, according to Henry and Mark, is an important factor that can guide evaluation practice to better influence individuals' attitudes and actions as well as aid in judging and improving policies and programs. Henry and Mark classified these underlying mechanisms into four types – general influence processes, cognitive and affective (or attitudinal) processes, motivational processes, and behavioral processes – and examined them at three levels of analysis (i.e., individual, interpersonal, and collective). Evaluation influence at the individual level involves changes in thought or actions of one or more individuals. Influence at the interpersonal level includes change arising from interactions between/among individuals. At the collective level, change is the influence of the evaluation on decisions and practices of the organization which may include agenda setting, policy-oriented learning, policy change, or diffusion. It is worth noting that shifting the ultimate purpose of an evaluation from use to social betterment has met with some resistance in the field. In particular, Cousins and Shulha (2006) maintain that viewing social betterment as the ultimate goal for evaluation is problematic because it excludes the many and valued types of evaluations that take place outside of the social policy domain.

Alkin (2003) argues that a push to replace the use concept with influence is an interesting, yet misguided suggestion and that the difference between the two concepts should be preserved. The important distinctions that Alkin poses include the use terminology as a way of capturing the evaluative setting and time frame, as opposed to influence terminology that includes other places and time. The evaluative time frame is seen as commencing when the evaluation is contemplated and concluding well beyond the completion of the report as long as the information is still thought of as stemming from the evaluation. The further removed from place and time of the evaluative encounter, Alkin maintains, the more one may view the evaluative results or processes as knowledge, and the influence might be considered knowledge use (as opposed to evaluation use, *per se*). Subsequently, King (2006) proposed a set of definitions to distinguish between evaluation use and influence. She defined evaluation use as the application of evaluation processes, products, or findings to directly produce an effect. On the other hand, evaluation influence was defined as the capacity of evaluation processes, products, or findings to indirectly produce a change in understanding or knowledge. While increased clarity in definition is certainly useful, it does not resolve the varying viewpoints in theory and research on evaluation use and influence.

Evaluation Use in Education

Programs and policies in the field of education have long been a primary target for evaluations. Evaluation in the

US educational arena can be traced back to the nineteenth century when, in 1897, Joseph Rice conducted a comparative study to evaluate the effects of spelling drills on the performance of 33 000 students. However, modern-day evaluation in education is often associated with the Tyler and Smith (1942) landmark longitudinal evaluation study on the achievement of high school students. The evaluation innovation of Tyler is credited as making the connection between educational outcomes and educators' intentions and with using evaluation as an integral part of the educational process. While educational evaluation shares commonalities with evaluation in other fields, such as criminal justice, public health, and social services, there are some distinguishable differences. Nevo (2006) documented three unique features: (1) the roots of evaluation in education are found in student testing and assessment, which has been conducted in schools around the world for centuries; (2) public involvement in the practice and use of educational evaluation is characteristically strong, since education is understood as a public good relevant to most members of society and far beyond the relevance of social services, health, and criminal justice; and (3) the role of teachers and their experience as evaluators (although mainly evaluators of students) cannot be ignored in designing and implementing educational evaluations that position teachers as evaluators, evaluation objects, and/or evaluation stakeholders.

In the United States, intense interest in utilizing social science research to evaluate educational and other social interventions was directly linked to expansion of government programs in the 1930s (described as the New Deal) and programming initiated by President John F. Kennedy and later expanded under President Lyndon B. Johnson's Great Society legislation during the 1960s. This legislation spurred a large investment of resources in education. Probably, the major stimulus for large-scale educational evaluations arose out of the inclusion of an evaluation clause in the Title I (compensatory education) section of the 1965 Elementary and Secondary Education Act (ESEA) which provided federal grants to local schools, state and regional agencies, and universities. This clause required the local education agency to submit an evaluation plan and the state agency to provide a summary report that described the effectiveness of the program. This legislation paved the way for evaluation requirements becoming an integral part of subsequent federal grants. Hence, educational evaluations expanded beyond assessment and student testing to examination of all aspects of the educational system, including instructional materials, classroom climate, school quality, teacher credentials and competency, and educational programs and activities. The increasing demand for evaluations in the field of education was coupled with discussions about how to improve educational evaluations and how such evaluations were used.

By the end of the 1960s, it was disappointing, yet evident, that evaluations of Great Society programs in education and other fields were largely ignored or politicized. Nonuse or underutilization was identified as one of the foremost problems in evaluation research at that time. Subsequently, in the 1980s and early 1990s, there was a decline in government funding for evaluation amidst a growing perception that evaluation procedures and practices were inadequate, not attuned to the realities of school situations, and generally fail to meet the informational needs of intended users.

More recently, funding for evaluation has increased and evaluation of education programs has emerged as a very deliberate research endeavor which can yield information of use and influence across multiple purposes. Thomas and McKie (2006), like other scholars, delineated major potential uses of educational evaluations, including providing information that could: (1) help shape the design and implementation of new and existing educational programs, (2) improve the management of existing educational programs and strategies, and (3) influence decision making regarding the development of educational policies and practices. The demand for instrumental use of evaluation findings for direct decision making has certainly escalated within the past decade with calls for greater educational accountability and efforts to raise the learning outcomes of all students, but particularly those placed at risk for underachievement and academic failure. More frequently, evaluations are utilized in a less direct, more subtle manner for reducing uncertainty, confirming existing points of view, or aiding in the day-to-day maintenance of a school, a program, or a system (Hofstetter and Alkin, 2003).

Over the past decade, consideration of useful evaluations in the field of education has been discussed within the context of the recent accountability movement. Utilizing a policy perspective, Ginsburg and Rhett (2003) characterized a useful evaluation as one that adds to the body of timely, relevant evidence to increase the likelihood that policy decisions improve program performance. They maintained that the value of any single evaluation is the additional contribution that it makes to the existing body of knowledge in the field. Ginsburg and Rhett noted that educational policymakers and program administrators will use findings from high-quality evaluations which are characterized by use of experimental designs, disaggregated data analyses, and new standards for literature reviews and case studies.

While evaluation scholars generally agree that scientific rigor is critically important, there is disagreement regarding the extent to which experimental designs are essential for an evaluation to be useful. Grob (2003) argues that many less-than-perfect evaluations can, in fact, be useful. He, like many others, contends that experimental designs are not always possible to achieve, can take years to

perform, and consume significant resources. Here, the usefulness of the evaluation is judged in terms of not simply the scientific rigor of the evaluation study, but also the combination of the evaluator's skilled engagement in the policy decision processes and the delivery of a sound, reliable, professional evaluation report which includes sufficient contextual information that allows readers to put things in perspective.

Evaluations in Diverse Education Settings

There is considerable evidence pointing to the failure of public schools in providing equitable and effective educational opportunities for children of color and those from low-income communities. Given increasing student diversity in US schools and the numerous reports of the achievement gap between white, middle-class students and students of color, there has been more pressure for utilizing evaluations to aid in efforts to elevate the achievement of low-performing students of color. Evaluations and evaluation use in diverse communities of color are very much social enterprises that can be best understood by taking into consideration the social, cultural, economic, and political contexts surrounding the program under consideration.

Educational evaluators working with programs serving students of color, in particular, can work in the public interest and on behalf of these communities by engaging their practice in a manner that produces high-quality data that will: (1) generate a more profound understanding of the context of education for marginalized students; (2) help educators and relevant decision makers strengthen marginalized students and the schools that serve them; and (3) enlighten and empower parents and communities from improvised areas to advocate for students. Thomas and McKie (2004) maintain that in order for evaluations to influence programs and policies seeking to enhance the achievement of students of color, evaluators studying interventions with this population must begin with a framework of designing and implementing evaluations to yield data that promote the creation of more equitable and socially just school reform practices. Further, evaluation data obtained from such a framework could provide information that moves beyond blaming the victim and by yield data and insights for improving achievement outcomes for struggling students. Some educational evaluators maintain that evaluations taking place in settings serving large numbers of students of color and other disadvantaged group should be educative inasmuch as they provide thoughtful, data-based insights into salient dimensions of schooling and contribute constructively to improvement and to a stronger and more contextually meaningful educational reform initiative (Greene *et al.*, 2006). A growing literature in the US is emerging on evaluation and evaluation use in educational settings serving students of color and those from disadvantaged backgrounds (e.g., Frierson *et al.*, 2000;

Madison, 1992; Thomas and McKie, 2006; Thomas and Stevens, 2004). Generally, this literature falls within a transformative paradigm (Mertens, 1999) that emphasizes the evaluator's role in facilitating utilization in ways that hold promise for ensuring that the inequities that created achievement gaps between white, middle-class students and students of color will be eliminated.

Conclusions

Evaluation use is a critical issue in the field. This is evident from the extensive literature on the topic ranging from Weiss's work beginning in the late 1960s and early 1970s to more recent publications in two volumes of *New Directions in Evaluation* (McLaughlin, 1988; Caracelli and Preskill, 2000), a special section of the *American Journal of Evaluation* (Sridharan, 2003), and numerous journal articles, book chapters, and conference presentations. Over the past 20 years, the emergence of more comprehensive frameworks on evaluation use has resulted in expanded conceptualization of the multiple domains of use, increased attention to the underlying motivation influencing users of evaluation, identification of specific avenues by which use takes place, and consideration of contextual factors influencing use and potential users. Such discourse has certainly advanced our conceptual, theoretical, and empirical understanding of evaluation use in multiple contexts. Given the purpose of evaluation, a conscious awareness of use-oriented effects, whether intentional (e.g., instrumental support for decision making) or not (e.g., skills development as a consequence of proximity to the evaluation or process use), must remain an important and continuing concern in the field (Alkin and Taut, 2003). Evaluators should engage their practice, from beginning to end, in a manner that promotes use and influence, both unidimensional and multidimensional, short-term and long-term and at multiple levels. Further, evaluators must maintain a delicate balance of paying sufficient attention to use without such overemphasis that it ultimately exerts influence on what findings are disclosed.

Frameworks for promoting evaluation use in the field of education, and other areas, have emerged and factors that increase the likelihood of use have been identified. Despite this reality, some critics maintain that there is still remarkably little empirical study to support many of the claims concerning evaluation use (Ginsburg and Rhett, 2003; Leviton, 2003). Leviton argues that the standard of evidence that many evaluators would not consider applying to the conduct of evaluations too often predominates in the study of evaluation use. While this suggests the need for more rigorous study of evaluation use, there is compelling evidence from case studies and meta-analytic investigations that evaluation use depends upon, among other factors, the quality of the study and evaluator

credibility. However, in large part, use, particularly in the policy arena, depends upon the situation and contextual environment impacting upon the evaluand. The evaluation is only one part of the policymaking process and often not even an important part, especially as political organization, or other changes in the environment may render the evaluation no longer relevant (Grasso, 2003). Therefore, an examination of evaluation use must incorporate a multifaceted perspective in seeking to better understand this concept and its place in evaluation theory and practice.

See also: Defining Quality in Evaluation; Evaluation and Accountability; Evaluation Methodology; Formative Assessment in Teacher Education and Teacher Professional Development; Internal and External Evaluation; School Based Evaluation: Purposes, Protocols and Processes.

Bibliography

- Alkin, M. C. (2003). Introduction (to Section on Evaluation Use). In Kelligan, T. and Stufflebeam, D. L. (eds.) *International Handbook of Educational Evaluation*, pp 189–196. Dordrecht: Kluwer Academic Press.
- Alkin, M. C. and Coyle, K. (1988). Thoughts on evaluation misutilization. *Studies in Educational Evaluation* **14**, 331–340.
- Alkin, M. C. and Taut, S. (2003). Unbundling evaluation use. *Studies in Educational Evaluation* **29**, 1–12.
- Caracelli, V. J. and Preskill, H. (eds.) (2000). *The Expanding Scope of Evaluation Use*. New Directions for Evaluation Use (No. 88). San Francisco, CA: Jossey-Bass.
- Cousins, B. J. and Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research* **56**(3), 331–364.
- Cousins, B. J. and Shulha, L. M. (2006). A comparative analysis of evaluation utilization and its cognate fields of inquiry: Current issues and trends. In Shaw, I. F., Greene, J. C., and Mark, M. M. (eds.) *The Handbook of Evaluation*, pp 266–291. Thousand Oaks, CA: Sage.
- Frierson, H. T., Hood, S., and Hughes, G. (2002). A guide to conducting culturally responsive evaluations. *The 2002 User-Friendly Handbook for Project Evaluation*, pp 63–73. Arlington, VA: National Science Foundation.
- Ginsburg, A. and Rhett, N. (2003). Building a better body of evidence: New opportunities to strengthen evaluation utilization. *American Journal of Evaluation* **24**(4), 489–498.
- Grasso, P. G. (2003). What makes an evaluation useful? Reflections from experience in large organizations. *American Journal of Evaluation* **24**(4), 507–514.
- Greene, J. C., DeStefano, L., Burgon, H., and Hall, J. (2006). An educative, values-engaged approach to evaluating STEM educational programs. In Huffman, D. and Lawrenz, F. (eds.) *Critical Issues in STEM Evaluation*. pp 53–71. New Directions for Evaluation (No. 109). San Francisco, CA: Jossey-Bass.
- Grob, G. (2003). A truly useful bat is one found in the hands of a slugger. *American Journal of Evaluation* **24**(4), 499–505.
- Henry, G. (2000). Why not use? In Caracelli, V. J. and Preskill, H. (eds.) *The Expanding Scope of Evaluation Use (New Directions for Evaluation)*, pp 85–98. San Francisco, CA: Jossey-Bass.
- Henry, G. T. and Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation* **24**(3), 293–314.
- Hofstetter, C. H. and Alkin, M. C. (2003). Evaluation use revisited. In Kellaghan, T., Stufflebeam, D. L., and Wingate, L. A. (eds.)

- International Handbook of Educational Evaluation*, pp 197–222. Boston, MA: Kluwer Academic Press.
- Joint Committee on Standards for Education Evaluation (1994) *The Program Evaluation Standards*, 2nd edn. Thousand Oaks, CA: Sage.
- King, J. (2006). Persistent challenges in studying evaluation use. *Paper Presented at the Annual Conference of the American Evaluation Association*. Portland, OR.
- Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. In Caracelli, V. J. and Preskill, H. (eds.) *The Expanding Scope of Evaluation Use (New Directions for Evaluation, 88)*, pp 5–23. Thousand Oaks, CA: Sage.
- Leviton, L. C. (2003). Evaluation use: Advances, challenges and applications. *American Journal of Evaluation* **24**(4), 525–535.
- Madison, A. M. (1992). Minority issues in program evaluation. *New Directions for Program Evaluation*, No. 53.
- Mark, M. M. and Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation* **10**(1), 35–57.
- McLaughlin, J. A. (1988). *Evaluation Utilization. New Directions for Program Evaluators*, vol. 39. San Francisco, CA: Jossey-Bass.
- Mertens, D. M. (1999). Inclusive evaluation: Implications of transformative theory for evaluation. *American Journal of Evaluation* **20**(1), 1–14.
- Nevo, D. (2006). Evaluation in education. In Shaw, I. F., Greene, J. C., and Mark, M. M. (eds.) *The Handbook of Evaluation*, pp 441–460. Thousand Oaks, CA: Sage.
- Patton, M. Q. (1998). Discovering process use. *Evaluation* **4**(2), 225–233.
- Scriven, M. (1991). *Evaluation Thesaurus*. Thousand Oaks, CA: Sage.
- Shadish, W. R., Cook, T. D., and Leviton, L. C. (1991). *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, CA: Sage.
- Shulha, L. M. and Cousins, B. (1997). Evaluation use: Theory, research and practice since 1986. *Evaluation Practice* **18**(3), 195–208.
- Sridharan, S. (2003). Introduction to special section on “What is a useful evaluation?” *American Journal of Evaluation* **24**(4), 483–487.
- Thomas, V. G. and McKie, B. (2006). Collecting and utilizing evaluation research for public good and on behalf of African American children. *Journal of Negro Education* **75**(3), 341–352.
- Thomas, V. G. and Stevens, F. I. (eds.) (2004). *Co-Constructing a Contextually Responsive Evaluation Framework: The Talent Development Model of School Reform. New Directions for Evaluation*, No. 101. San Francisco, CA: Jossey-Bass.
- Tyler, R. W. and Smith, E. (1942). *Appraising and Recording Student Progress*, Vol. 3: *Adventure in American Education*. New York: Harper.
- Weiss, C. H. (1981). Measuring the use in evaluation. In Ciarlo, J. A. (ed.) *Utilizing Evaluation: Concepts and Measurement Techniques*, pp 17–33. Thousand Oaks, CA: Sage.
- Weiss, C. H. (1988). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice* **9**(1), 5–19.
- Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation* **19**(1), 21–33.

Further Reading

- Cousins, J. B. (2004). Minimizing evaluation misuse as principled practice. *American Journal of Evaluation* **25**(3), 391–397.
- Henry, G. (2003). Influential evaluations. *American Journal of Evaluation* **24**(4), 515–524.
- Kellaghan, T., Stufflebeam, D. L., and Wingate, L. A. (eds.) (2003). *International Handbook of Educational Evaluation*, pp 197–222. Boston, MA: Kluwer Academic Press.
- Leviton, L. C. (2003). Evaluation use: Advances, challenges, and applications. *American Journal of Evaluation* **24**(4), 525–535.
- Mark, M. M., Henry, G. T., and Julnes, G. (2000). *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs*. San Francisco, CA: Jossey-Bass.
- Nutley, S., Walter, I., and Davies, H. T. O. (2003). From knowing to doing. *Evaluation* **9**(2), 125–148.
- Preskill, H. and Caracelli, V. (1997). Current developments and conceptions of use: Evaluation use TIG survey results. *Evaluation Practice* **18**(3), 209–225.

Moral and Ethical Issues in Evaluation

M Morris, University of New Haven, West Haven, CT, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Conflict of interest – A situation where an evaluator's personal or financial interests could influence, or be influenced by, the way in which the evaluation is designed, carried out, or reported.

Stakeholder – An individual or group with a vested interest in what is being evaluated.

Virtually all forms of research are subject to moral and ethical hazards, and this is certainly true of program evaluation. Within this context, the terms moral and ethical are essentially synonymous with the latter term being more frequently used in evaluation. An ethical issue is one that deals with what is good and bad and with moral duty and obligation. In other words, it deals with what individuals should or should not do. When evaluators behave ethically, they are conforming to accepted standards for professional conduct. Consistent with prevailing practice, the terms ethics and ethical are used throughout this article.

Most evaluators report that they have encountered ethical challenges in their work. It is likely that many of those who indicate otherwise have, in fact, faced problematic situations that other evaluators would have conceptualized in ethical terms. Individuals vary in the extent to which they view events as having ethical significance, and the existence of explicit professional standards pertaining to ethical issues can be counted on to narrow this variation but not eliminate it. What one evaluator defines as an ethical issue, another might see as a political, philosophical, or methodological one. Consider, for example, the question of how involved key stakeholders should be in various phases of the evaluation. Many evaluators would regard extensive engagement as an ethical imperative. Others might see it as primarily a political matter, in terms of making choices that have implications for managing dynamics of power and influence within the evaluation. For a third subgroup, the issue may be viewed as mainly one of personal preference and values; that is, some evaluators may be ideologically committed to in-depth stakeholder involvement, while others might not be. Finally, the methodological significance of stakeholder involvement might be an evaluator's chief concern. The focus here would be the impact of stakeholder participation on the quality of the evaluation design, the adequacy of data collection and analysis, and other factors affecting

the technical soundness of the evaluation. For example, in what ways could stakeholder involvement enhance, or compromise, the objectivity needed for a credible evaluation? As we can see, answering the question 'What is an ethical issue in evaluation?' is not always as straightforward a process as it might first appear.

Professional Standards and Guidelines

One of the hallmarks of an emerging profession is the development of standards and guidelines to promote ethical practice within the field. Accordingly, the past three decades have witnessed a growing number of efforts, both within and across national boundaries, to establish standards for evaluation. For example, specific sets of standards have been crafted for use in the United States, Canada, Switzerland, Germany, France, the United Kingdom, Eastern Europe, Western Europe, Australasia, and Africa. Two of the most prominent efforts in this regard are the Guiding Principles for Evaluators and the Program Evaluation Standards, both originating in the United States.

Guiding Principles for Evaluators

The Guiding Principles, developed by the American Evaluation Association in 1995 and revised in 2004, offer guidance that is more general than specific, and are intended to be relevant to all types of evaluation. The principles target five domains where evaluators face distinctive ethical obligations:

1. Systematic inquiry – evaluators conduct systematic, data-based inquiries.
2. Competence – evaluators provide competent performance to stakeholders.
3. Integrity/honesty – evaluators display honesty and integrity in their own behavior, and attempt to ensure the honesty and integrity of the entire evaluation process.
4. Respect for people – evaluators respect the security, dignity, and self-worth of respondents, program participants, clients, and other evaluation stakeholders.
5. Responsibilities for general and public welfare – evaluators articulate and take into account the diversity of general and public interests and values that may be related to the evaluation.

Each core statement is accompanied by an elaboration focusing on the meaning and application of the principle.

These elaborations clarify the nature of the evaluator's ethical responsibilities within a given domain. Thus, the Guiding Principles can be seen as providing evaluators with a foundation for building an ethically sound practice. The systematic inquiry principle, for example, requires evaluators to conduct research that is methodologically appropriate to the questions being asked and the setting in which the research is taking place, and to clearly communicate with stakeholders concerning the strengths/limitations of the approach employed.

The competence principle is closely linked to systematic inquiry, in that one's competence is a major determinant of one's ability to carry out a methodologically credible evaluation. Ethical evaluators only take on projects for which they possess the relevant skills, background, and expertise. This is a particularly important issue in evaluation, due to the general absence of regulatory bodies that accredit training programs and license practitioners in the field. When third parties are not available to officially endorse an evaluator's training and competence, stakeholders are in a disadvantaged position with respect to assessing the suitability of an evaluator for a specific project. Thus, evaluators must guard against exploiting this disadvantage to engage in evaluations for which they are not adequately prepared.

The significance of cultural competence is also highlighted by the competence principle. Evaluators increasingly work in settings, and with populations, whose norms, values, and cultures differ from their own. When evaluations are designed and implemented in ways that are sensitive to these differences, their validity and value are enhanced. Evaluators have an ethical responsibility not only to be mindful of their own value-based assumptions, and how these assumptions interact with those of other stakeholders, but also to use this knowledge to design responsive, effective evaluations.

The issue of transparency is at the heart of the integrity/honesty principle. Throughout the evaluation, from the planning phase to the communication and utilization of findings, ethical evaluators are proactive in their efforts to ensure that all aspects of the evaluation are accurately represented to stakeholders and understood by them. Among the myriad issues that can arise here, three of the most important are conflicts of interest, the presentation of findings, and the limitations of the evaluation. Opportunities for conflicts of interest abound in evaluation. For example, an evaluator may have a personal relationship with a key stakeholder in the evaluation, or the evaluator may be asked to help design a program that he or she will also be requested to evaluate. When conflicts of interest occur, it is essential that they are acknowledged and steps taken to address them.

An evaluation is rendered worthless, or worse, when its findings are not honestly communicated. Evaluators are often pressured by stakeholders to engage in misrepresentation

(at least in the opinion of evaluators) and this pressure must be resisted, lest the credibility of the entire endeavor is undermined. Evaluators have an ethical obligation to explore with stakeholders any significant disagreements over the reporting or interpretation of results, but such exploration should never lead to evaluators taking ownership of findings or interpretations that they believe are unjustified by the data.

This issue is closely related to acknowledging an evaluation's limitations. Acting consistently with the integrity/honesty principle means that evaluators are realistic in describing the extent to which a study can answer the questions it addresses. They must not succumb to the temptation to oversell the knowledge and conclusions that an evaluation can produce or has produced, even when stakeholders appear eager to accept such exaggerations.

The fourth principle, respect for people, encompasses the domains traditionally associated with the ethical treatment of human subjects. Thus, evaluators are expected to obtain informed consent, protect the confidentiality (or anonymity) of participants, minimize risks to participants while maximizing benefits, respect participants' dignity, and, in general, strive to do no harm when carrying out the evaluation.

Informed consent is the process by which individuals exercise their right to participate, or not participate, in research after having been provided with information concerning the study, its risks, and its benefits. Indeed, under informed consent, individuals are free to withdraw from a study after they have agreed to participate in it. This provision can pose a particular challenge to evaluators as they attempt to maintain the integrity of research design and data collection strategies throughout an investigation. Circumstances such as these can tempt evaluators to misrepresent a study's risk/benefit ratio when communicating with potential and current participants, a temptation that must be resisted if ethical practice is to be preserved.

Addressing confidentiality and anonymity can also be problematic for evaluators. Confidentiality refers to the protection of an individual's privacy when reporting findings (where the individual's identity is known to the evaluator), while anonymity involves gathering data in a way that ensures individual respondents cannot be linked by the evaluator to the specific information they provide. Both of these protections can be difficult to achieve, often because of the relatively small number of individuals in certain stakeholder groups (e.g., program staff) providing data for the evaluation.

At a more general level, upholding the do-no-harm principle is frequently a complex undertaking in evaluation. Evaluations can contribute to program termination/reduction, the loss of services for clients, and the release of staff from the program and related settings. Although

these actions may, in the long run, increase the availability of resources for more effective interventions, such an outcome is by no means guaranteed, and the short-term damage to the welfare of particular stakeholder groups can be significant. What is clear is that evaluators engage, and collect data from, not just program clients, but program staff, administrators, funders, community-based constituencies, clients' families, and other stakeholders. Safeguarding the rights and responding to the legitimate needs of one group can sometimes conflict with being sensitive to the needs and rights of others, with the potential for conflict growing as the number of relevant stakeholders increases. The generality of the Guiding Principles makes it difficult to resolve these issues by simply referring to the principles themselves. Indeed, as is the case with all ethical challenges, the details of the case at hand must be thoroughly examined before reaching a decision.

Finally, the principle, responsibilities for general and public welfare, is probably the most ambitious of the five Guiding Principles. In taking the position that "evaluators will usually have to go beyond particular stakeholder interests and consider the welfare of society as a whole" when conducting evaluations, this principle assumes that the welfare of society as a whole can in fact be discerned by the evaluator. Concepts such as the public interest, public welfare, and public good are referred to in the responsibilities principle but not specifically defined. Of course, visions of what constitutes the good society can greatly vary among individuals, groups, and cultures. Thus, the responsibilities principle serves as a vehicle for evaluators to explore the relationship of their and other stakeholders' notions of social justice to evaluation practice. Indeed, a number of evaluation approaches have been developed over the years – for example, empowerment evaluation, deliberative democratic evaluation, participatory evaluation, and collaborative evaluation – that appear to be explicitly grounded in such broader visions of the collective good.

The Program Evaluation Standards

The most detailed set of professional guidelines for ensuring the ethical practice of evaluation are the Program Evaluation Standards, developed by the Joint Committee on Standards for Educational Evaluation. The Standards were first published in 1981, followed by a second edition in 1994. The third edition is likely to be officially released in 2010. The second edition consists of 30 standards organized into four groups, with each group representing a key desired characteristic of an evaluation:

1. *Utility*. These standards emphasize the importance of evaluations being responsive to the information needs of intended users. In short, a good evaluation is a useful evaluation.

2. *Feasibility*. The focus of these standards is on ensuring that evaluations are realistic, prudent, diplomatic, and frugal. Evaluations high in feasibility are ones that take into account the practical constraints, political dynamics, and resource limitations operating in the settings in which they are conducted.
3. *Propriety*. These standards examine, in more specific terms, many of the issues addressed by the Guiding Principles of respect for people and integrity/honesty: rights of human subjects, informed consent, confidentiality, conflicts of interest, fair representation of findings, etc. Propriety standards tend to be the ones invoked most frequently in discussions of the ethics of evaluation.
4. *Accuracy*. Like the systematic inquiry principle, these standards stress the importance of the evaluation's methodological soundness. For example, was the program comprehensively examined? Was valid and reliable information competently collected, analyzed, and communicated to stakeholders?

The issues addressed in the second edition of the Standards are likely to receive substantial attention in the third edition as well. The evolution of professional standards is typically characterized by continuity and incremental change rather than dramatic shifts. However, the need for evaluators to be culturally and contextually sensitive will almost certainly be accorded greater prominence in the third edition, a development that reflects, at least in part, the globalization of the evaluation profession and major demographic changes in the United States and elsewhere.

Research on Ethical Challenges

Although anecdotal reports of ethical challenges encountered by evaluators are numerous, systematic studies of the experiences of representative samples of evaluators are rare. Viewed together, these two sources of information produce a relatively coherent picture of the ethical terrain that evaluators negotiate in their work. It is clear that every stage of the evaluation process can pose ethical problems for the evaluator, with some stages appearing to be more ethically hazardous than others. The following sections provide an overview of the ethical conflicts that evaluators claim they face most frequently.

Entry/Contracting Stage

In this phase, the evaluator and key stakeholders discuss the overall focus of the evaluation and the resources needed to conduct it. This is an exploratory stage, where the evaluator is forming judgments about the wisdom of taking on the project. Ethical challenges occurring during this phase include the following:

- the evaluator believes that a core stakeholder has already concluded what the evaluation results should be

or plans to make unethical use of the results (e.g., to punish a staff member);

- a perceived conflict of interest exists (e.g., a future evaluation contract depends on positive findings emerging from the proposed evaluation);
- issues and questions that the evaluator believes should be addressed in the evaluation are deemed off-limits by a stakeholder;
- in the opinion of the evaluator, legitimate stakeholders have been left out of the planning process;
- there is conflict among stakeholders surrounding key features of the proposed evaluation; and
- the evaluator encounters problems in identifying stakeholders.

These challenges, if not successfully addressed, can fatally damage the credibility of the ensuing evaluation. For example, if a stakeholder is seeking an evaluation to support a decision that he or she has already made, the evaluator is likely to be subjected to strong, and inappropriate, pressure from that stakeholder during the communication-of-results phase. Similarly, if serious disagreements among stakeholders emerge during the entry/contracting stage, whose views should be accorded priority by the evaluator if consensus cannot be achieved? The evaluator's client? The most knowledgeable stakeholder? The most powerful stakeholder? Or the most disenfranchised stakeholder? It is here where evaluators' judgments of the nature of the public interest and collective good can play a crucial role.

Designing the Evaluation

Methodological concerns are central to this phase; the major objective is to generate a research strategy that can provide sound answers to the core questions driving the evaluation. From an ethical perspective, the struggle most frequently reported by evaluators during this stage is the failure to achieve acceptance of the overall design by stakeholders. Evaluators can feel pressured to employ a design that they believe is ill-equipped to answer the questions the study is supposed to address. Is it ethically defensible to simply share with stakeholders one's views of the flawed nature of an evaluation design while proceeding to implement it in a particular project due to stakeholder preferences? Here we can see one of the consequences of not effectively engaging stakeholders during the entry/contracting phase.

Data Collection

As one might expect, assurances of confidentiality and anonymity to research participants are often threatened during the data-collection phase, a problem that is likely to increase in frequency and magnitude as the use of large, computerized databases in evaluation grows. Although

evaluators' self-reports do not indicate that violations of informed consent are a common occurrence, observational studies of informed-consent interactions suggest that such violations are probably more frequent than evaluators realize.

It is also the case that during data collection evaluators can become aware of stakeholder behaviors that are unethical, illegal, or dangerous (e.g., misappropriation of funds or sexual harassment). What to do with this knowledge is often a vexing question, especially in view of the fact that evaluators' legal responsibilities in these circumstances are not necessarily clear. Even when they are clear, an evaluator may be strongly committed to an ethical principle (e.g., honoring confidentiality) that conflicts with the relevant legal obligation.

Data Analysis and Interpretation

This stage encompasses the technical procedures evaluators use to process the data quantitatively and qualitatively, as well as the conceptual and value frameworks evaluators employ in reaching conclusions about what the results mean in substantive terms. It is in this latter domain where evaluators report the greatest ethical difficulties. Specifically, they sometimes fear that their interpretations and conclusions may be biased by unacknowledged values and sentiments (e.g., admiration for or hostility toward program staff) that could cause them to view the study's findings, and the programs the findings describe, more positively or harshly than the data warrant. In this fashion, the evaluator's role as an honest broker is seriously compromised.

Communication of Results

Being pressured by a stakeholder to misrepresent a study's findings is, by far, the ethical challenge that evaluators claim they encounter most frequently in this or any other stage of an evaluation. Although the pressure usually focuses on making the program look better than the evaluator believes is justified by the data, occasionally the influence attempt is in a negative direction. Pressure can range from subtle encouragement to soften a finding to forceful declarations that certain results must be deleted from a final report. There can also be instances when the dispute concerns not how positive or negative the findings are, but the issue of what the findings mean substantively. The common denominator of all these episodes is that the evaluator sees the recommended changes as inappropriate, and that agreeing to them would represent a violation of ethical responsibilities involving honesty, integrity, and accuracy. Interestingly, research on the views of other stakeholders (e.g., evaluation funders) suggests that evaluators themselves can provide the impetus for misrepresentation, attempting to do whatever they believe is necessary to please the evaluation client.

Confidentiality issues also arise during the communication-of-results phase. For example, evaluators indicate that stakeholders sometimes pressure them to reveal the identities of individuals who provided particular interview responses. Threats to confidentiality, however, are not always due to stakeholder pressure. In certain instances, evaluators are concerned that the simple act of reporting findings can undermine confidentiality. This is especially likely when sample sizes are small and the breaking down of results according to multiple background characteristics (e.g., gender, race/ethnicity, and age) is being considered. Tables in final reports containing cells with few respondents can reveal more than the evaluator intends about who said what.

Utilization of Findings

In theory, evaluations are undertaken to influence decision making in the short and/or long term. This purpose generates many opportunities for ethical problems, and it is not surprising that evaluators report frequent encounters with a number of them. These include the following:

- a stakeholder suppresses or ignores evaluation findings;
- disagreements take place over who owns the raw data and/or the right to distribute/publish the evaluation's findings;
- a stakeholder uses the findings to punish another stakeholder (e.g., termination of employment) or the evaluator (e.g., harshly and unfairly criticizing the evaluation's quality);
- a stakeholder intentionally modifies the results in a way that misrepresents the substance of what was found;
- a stakeholder, operating in good faith, misinterprets the results; and
- a stakeholder (usually the evaluator's supervisor) engages in plagiarism or misrepresentation of authorship regarding the evaluation.

These utilization challenges vary on at least two dimensions. One is the degree to which the evaluator feels obligated to take action to address the situation. For example, perceptions of deliberate misrepresentation are likely to provoke a stronger response than perceptions of a stakeholder simply not using the evaluation's results. Indeed, some evaluators might view the latter circumstance as one where it is the stakeholder's prerogative to use the results or not. The second dimension involves the feasibility of responding to the situation. Correcting a benign misinterpretation is usually a much less challenging endeavor than intervening in an unjust termination. This issue highlights an important reality: of all the phases in an evaluation, the utilization-of-results stage is the one where the evaluator typically has the least amount of direct control over the course of events, ethical or otherwise.

Preventing and Responding to Ethical Challenges

The most effective strategy for dealing with ethical conflicts is to prevent them from occurring. However, problems can arise in an evaluation that even the most foresighted evaluator could not have anticipated. What follows is an overview of the major steps that evaluators can take to ensure that their practice is ethically grounded.

Actively Manage the Entry/Contracting Stage

This is the phase of the evaluation where preventive efforts are likely to bear the most fruit. Evaluators can do two things during this stage to enhance the ethical quality of the evaluation:

1. Discuss with stakeholders the implications of the professional guidelines most relevant to the proposed evaluation. Specifically, evaluators should attempt to identify potential problem areas involving ethics, and explore with stakeholders the steps that can be taken to address these issues proactively. In this regard, it is important that evaluators clearly communicate the professional standards they must uphold in conducting the evaluation.
2. Encourage stakeholders to share any ethical concerns they have about the upcoming project. Given that stakeholders often lack the experience and knowledge to feel confident in raising such issues, evaluators may need to take the lead in providing examples.

These entry/contracting activities have both content and process objectives. The former is to achieve consensus on how to address specific ethical issues in the evaluation. The latter is to generate a climate where evaluators and stakeholders are willing to bring ethical matters to the table as the evaluation unfolds. Ethical problems left unaddressed are unlikely to resolve themselves.

Apply Relevant Professional Guidelines

When ethical challenges arise, evaluators should consult the professional principles and standards they deem most relevant. For many evaluators in the United States, these will be the Guiding Principles and/or the Program Evaluation Standards. As previously indicated, a number of other countries/regions have developed their own evaluation standards, often based, at least in part, on those used in the United States. As principles and standards are inherently general and abstract, they are constrained in their ability to offer detailed guidance on how to respond in specific circumstances. Rather, their main purpose is to highlight overall imperatives that evaluators should be mindful of as they deal with ethical challenges. Reinforcing this point is

the fact that situations can arise where upholding one principle or standard can compromise the ability to adhere to another principle or standard from the same set of guidelines. Professional guidelines in evaluation typically do not prioritize individual principles or standards, leaving to the evaluator the task of choosing between them when conflicts occur, based on careful consideration of the case at hand. Although these limitations of professional guidelines are significant, they do not erode the more fundamental value that guidelines offer to the practicing evaluator.

Consult with Colleagues

Experienced colleagues in evaluation can be an invaluable resource when grappling with ethical challenges. Most importantly, they can broaden and deepen one's perspective, thereby helping one appreciate aspects of the problem that may have been overlooked. They can also expand the range of options the evaluator considers when deciding how to respond to the conflict. Seeking input from colleagues has become much easier in recent years, as Internet listservs, such as EVALTALK, have grown in the evaluation community.

Appreciate Culture and Context

As previously noted, the settings in which evaluations take place increasingly embody cultures, values, and norms that differ significantly from those represented by the evaluator. These differences can have substantive implications for how ethical issues are conceptualized and addressed (e.g., views of confidentiality and privacy in individualistic vs. collective cultures). Prudent researchers learn as much as possible about these features of the host environment prior to embarking upon an evaluation. Doing so reduces the likelihood that the evaluator will inadvertently violate the ethical sensibilities of key stakeholders, and increases the chances that the professional standards subscribed to by the evaluator can be successfully adapted to the setting.

Examine One's Own Values

Professional standards and principles represent just one of the value frameworks upon which ethical decisions can be based in an evaluation. Evaluators inevitably bring their personal values to their work, and these can play a crucial role in determining how one responds to ethical conflicts, especially when professional guidelines provide unclear or conflicting advice.

With this background in mind, it is essential that evaluators possess a highly articulated set of personal values, and understand how these values interact with professional standards and the settings in which they conduct

their work. These settings may embrace value orientations that are fundamentally at odds with the evaluator's core beliefs concerning social justice and the public interest. In such instances, evaluators must decide whether it is appropriate for them to take on a particular project. This decision can be a particularly wrenching one if the material self-interest of the evaluator will be affected in a major way by the choice that is made.

Underlying this recommendation is an essential truth: ethical issues in evaluation, and the professional standards applied to them, are inextricably linked to more general ethical questions and challenges faced by people in all walks of life. Individuals with a well-grounded sense of who they are in this broader sense are in a much stronger position to address the ethical challenges awaiting them in the field of evaluation.

See also: Defining Quality in Evaluation; Evaluation Reporting and Communicating; Evaluation Use; Internal and External Evaluation; Meta-evaluation: Purpose, Prescription, and Practice; Program Evaluation; The Role of Stakeholders in Educational Evaluation.

Further Reading

- African Evaluation Association (2002). *The African Evaluation Guidelines 2000*. <http://www.afrea.org/content/index.cfm?navID=5&itemID=204> (accessed August 2009).
- American Evaluation Association (2004). *Guiding Principles for Evaluators* (revised). <http://www.eval.org/Publications/GuidingPrinciples.asp> (accessed August 2009).
- American Evaluation Association (2007). *American Evaluation Association Guiding Principles Training Package*. <http://www.eval.org/GPTraining/GPTrainingOverview.asp> (accessed August 2009).
- Australasian Evaluation Society (1997). *Australian Evaluation Society Guidelines for the Ethical Conduct of Evaluations*. <http://www.aes.asn.au> (accessed August 2009).
- Blustein, J. (2005). Toward a more public discussion of the ethics of federal social program evaluation. *Journal of Policy Analysis and Management* **24**, 824–846.
- Cooksy, L. J. (2005). The complexity of the IRB process: Some of the things you wanted to know about IRBs but were afraid to ask. *American Journal of Evaluation* **26**, 353–361.
- Fitzpatrick, J. L. and Morris, M. (eds.) (1999). *Current and Emerging Ethical Challenges in Evaluation. New Directions for Evaluation*, No. 82. San Francisco, CA: Jossey-Bass.
- House, E. R. and Howe, K. R. (1999). *Values in Evaluation and Social Research*. Thousand Oaks, CA: Sage.
- Joint Committee on Standards for Educational Evaluation (1994). *The Program Evaluation Standards: How to Assess Evaluations of Educational Programs*, 2nd edn. Thousand Oaks, CA: Sage.
- Mabry, L. (1999). Circumstantial ethics. *American Journal of Evaluation* **20**, 199–212.
- Mertens, D. M. and Ginsberg, P. E. (eds.) (2009). *The Handbook of Social Research Ethics*. Thousand Oaks, CA: Sage.
- Morris, M. (1998). *Ethical Challenges*. Section of the *American Journal of Evaluation*, beginning with Volume **20**.
- Morris, M. (2007). Foundation officers, evaluation, and ethical problems: A pilot investigation. *Evaluation and Program Planning* **30**, 410–415.
- Morris, M. (ed.) (2008). *Evaluation Ethics for Best Practice: Cases and Commentaries*. New York: Guilford.
- Morris, M. and Cohn, R. (1993). Program evaluators and ethical challenges: A national survey. *Evaluation Review* **17**, 621–642.

- Morris, M. and Jacobs, L. (2000). You got a problem with that? Exploring evaluators' disagreements about ethics. *Evaluation Review* **24**, 384–406.
- Newman, D. L. and Brown, R. D. (1996). *Applied Ethics for Program Evaluation*. Thousand Oaks, CA: Sage.
- Penslar, R. L. (1993). *Institutional Review Board Guidebook*. http://www.hhs.gov/ohrp/irb/irb_guidebook.htm (accessed August 2009).
- Pocciotto, R. (2005). The value of evaluation standards: A comparative assessment. *Journal of MultiDisciplinary Evaluation* **3**, 30–59. http://survey.ate.wmich.edu/jmde/index.php/jmde_1/index (accessed August 2009).
- Rossi, P. H. (ed.) (1982). *Standards for Evaluation Practice. New Directions for Program Evaluation*, No. 15. San Francisco, CA: Jossey-Bass.
- Russon, C. and Russon, G. (eds.) (2004). *International Perspectives on Evaluation Standards. New Directions for Evaluation*, No. 104. San Francisco, CA: Jossey-Bass.
- Schwandt, T. A. (2007). On the importance of revisiting the study of ethics in evaluation. In Kushner, S. and Norris, N. (eds.) *Advances in Program Evaluation: Dilemmas of Engagement: Evaluation and the New Public Management*, vol. 10, pp 117–127. New York: Elsevier.
- Shadish, W. R., Newman, D. L., Scheirer, M. A., and Wye, C. (eds.) (1995). *Guiding Principles for Evaluators. New Directions for Program Evaluation*, No. 66. San Francisco, CA: Jossey-Bass.
- Simons, H. (2006). Ethics in evaluation. In Shaw, I. F., Greene, J. C., and Mark, M. M. (eds.) *The SAGE Handbook of Evaluation*, pp 243–265. Thousand Oaks, CA: Sage.
- Strike, K. A., Anderson, M. S., Curren, R., et al. (2002). *Ethical Standards of the American Educational Research Association: Cases and Commentary*. Washington, DC: American Educational Research Association.
- Thompson-Robinson, M., Hopson, R., and SenGupta, S. (eds.) (2004). *In Search of Cultural Competence in Evaluation: Toward Principles and Practices. New Directions for Evaluation*, No. 102. San Francisco, CA: Jossey-Bass.
- Walker, R., Hoggart, L., and Hamilton, G. (2008). Random assignment and informed consent: A case study of multiple perspectives. *American Journal of Evaluation* **29**, 156–174.
- Widmer, T., Landert, C., and Bachmann, N. (2000). *Evaluation Standards of the Swiss Evaluation Society*. http://www.seval.ch/en/documents/SEVAL_Standards_2000_en.pdf (accessed August 2009).

Relevant Websites

- <http://www.evaluationcanada.ca> – Canadian Evaluation Society Guidelines for Ethical Conduct.
- <http://www.degeval.de> – DeGEval, Evaluation Standards of the German Evaluation Society.
- <http://www.coventrypartnership.com/upload/documents/webpage/document242.pdf> – UK Evaluation Society Guidelines for Good Practice in Evaluation.

The Role of Stakeholders in Educational Evaluation

S Taut, Pontificia Universidad Católica de Chile, Santiago, Chile

M C Alkin, University of California, Los Angeles, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In this article, we discuss the way in which stakeholders are defined in the evaluation literature and what is said in evaluation writings about the role of stakeholders in educational program evaluation. We dichotomize stakeholder involvement into two categories that we refer to as deep involvement and broad involvement. Focusing on deep involvement, which is the framework for most of the research literature, we discuss why stakeholder involvement is important and offer a summary of key factors that lead to such involvement.

Who Are Stakeholders?

Stakeholders are all those individuals who have a vested interest in what is being evaluated. This includes evaluation clients and all those who will use the evaluation and its results for various purposes. It also includes those whose work is being studied and who will be affected by the evaluation.

How have Stakeholders Typically been Involved in Evaluations?

The extent to which different stakeholders are engaged in the process of an evaluation depends on various factors not limited to the evaluator's theoretical positioning, funding, and time. Nonetheless, stakeholder involvement can roughly be thought of as being either deep or broad. On the one hand, stakeholders are deeply engaged in an evaluation when they are closely engaged in various evaluative activities. Such activities range from determining the evaluation questions and designing the evaluation to identifying potential data sources and collecting and analyzing data. In some cases, stakeholders may even participate in writing the evaluation report. In addition, when one of the evaluation's goals is to engage stakeholders on a more significant level, the number of stakeholder groups involved is typically limited to individuals who are in positions to make decisions about what is being evaluated. On the other hand, broad involvement entails the involvement of a wide range of stakeholders in a smaller portion of the evaluation process. Unless funds are unlimited, stakeholders tend to play a more limited role. This may be only during the beginning stages of the evaluation by contributing to the development of the program theory.

Stakeholders in Educational Program Evaluation

What Do the Program Evaluation Standards Say About the Role of Stakeholders?

The *Program Evaluation Standards* developed by the Joint Committee on Standards for Educational Evaluation (1994) is a guide for what constitutes high-quality educational program evaluation practice in the United States. The second edition's 30 standards are organized around four attributes of an evaluation – utility (U1–U7), feasibility (F1–F3), propriety (P1–P8), and accuracy (A1–A12) – and have been approved as American National Standards. The *Program Evaluation Standards* reflect what the field of evaluation in this country thinks is responsible conduct and use of evaluation while staying largely independent of the theoretical and methodological preferences of evaluators. Therefore, the *Program Evaluation Standards* is a prominent source for finding out what the field of evaluation as a professional entity thinks should be the role of stakeholders in evaluation.

There are three program evaluation standards in the second edition that directly address the role of stakeholders. These are U1 stakeholder identification, U4 values identification, and F2 political Viability. The first is central to the topic and asserts that “it is crucial to identify potentially important stakeholders and, within resource and time limitations, to involve them in the planning and conducting of the evaluation” (1994: 25). This standard statement even goes one step further when it observes that “in many evaluations, special efforts may be necessary to promote the appropriate inclusion of less powerful groups or individuals as stakeholders, such as racial, cultural or language minority groups” (1994: 25). The standard statement on values identification posits that evaluators, clients, and multiple stakeholders should be involved in determining how values are assigned in the evaluation. Finally, the statement related to political viability warns against possible pressures from stakeholders and asks evaluators to prevent their involvement from biasing the evaluation and its findings.

In summary, the *Program Evaluation Standards* assigns stakeholders an important role in the conduct of an evaluation. It implies that stakeholder involvement helps increase the likelihood that evaluations are used and that evaluation also helps ensure the inclusion of less powerful groups, giving them a voice in evaluation. The *Program*

Evaluation Standards also cautions that evaluators may need to somewhat restrict the number of stakeholders to be included because of the difficulty of dealing with too many conflicting views.

Program Evaluation Theoretical Positions and the Role of Stakeholders

The *Program Evaluation Standards* attempts to represent a range of evaluation theoretical perspectives on a broader level. However, the book is designed as a guide that should be appropriately adapted to individual evaluation contexts. As such, it does not reflect the universe of theoretical positions that is reflected in the evaluation literature. To address this issue, in this section we discuss categories of evaluation theoretical positions and the role that they afford stakeholders more specifically.

The theoretical positions referred to here are use-, values-, and methods-oriented approaches, which represent the three main branches of an evaluation theory tree developed by Marvin Alkin. Use- and values-oriented approaches include those that explicitly address stakeholder involvement as their central tenet. For example, approaches that are mainly concerned with evaluation use include Michael Patton's utilization-focused evaluation and Brad Cousins' participatory evaluation. Approaches that are value laden may stress social justice concerns. Examples of these theoretical positions include deliberative democratic evaluation by Ernest House and Kenneth Howe, and fourth-generation evaluation by Egon Guba and Yvonna Lincoln. As mentioned earlier, each family of approaches involves stakeholders differently. Specifically, those that are mainly use oriented tend to emphasize deep involvement with a limited number of primary users as stakeholders throughout the evaluation process. Theoretical positions that are values oriented, those whose goal is to serve social justice, for instance, tend to involve a wide range of stakeholders on a broad level. Finally, methods-oriented approaches prioritize appropriate methods of inquiry when conducting an evaluation and typically involve a limited number of stakeholders for limited engagement – primarily at the early stages of the evaluation. Thus, they are neither deep nor broad. Huey-Tsyh Chen's theory-driven evaluation is one example of this theoretical position (also see Stewart Donaldson and Carol Weiss for others).

Christina Christie studied the extent to which current evaluation theorists say they involved stakeholders in their evaluations. None of the eight theorists included in the study saw stakeholders as mere audiences. Rather, all found stakeholders to have some kind of role, even if this role was limited to helping define evaluation issues (e.g., Ernest House). Others reported extensive participation of stakeholders in all aspects of the evaluation process (e.g., David Fetterman). It is interesting that Christie, in studying a group of evaluation practitioners, found that

their practice was more diverse in terms of stakeholder involvement, including practitioners who considered stakeholders to have no role in the evaluation except as its audience.

Although the majority of the evaluation field agrees that stakeholders should be involved at least to some degree in an evaluation, a more detailed analysis of the role of stakeholders is warranted. Following Jennifer Greene we summarize the issues to be discussed as follows: (1) How should the evaluator decide which stakeholders to involve, that is, whether to involve a few of them in depth, or a broad range less actively? (2) If the evaluator decides to involve a few stakeholders in depth, how should he or she do so, and what are the effects of such practice? Also (3) if the evaluator decides to involve a broad range of stakeholders, how should he or she do so, and what are the effects? Finally, in turning to empirical research we ask under what conditions positive effects of stakeholder involvement are maximized, and negative effects minimized, irrespective of the theoretical and practical approach implemented. We would like to elaborate on these questions in the following sections.

Deciding which stakeholders to involve and how

Some evaluation scholars have posited that both the election of stakeholders and the prioritizing of breadth or depth in their involvement are decisions the evaluator must make based on the values he or she holds relative to the main purpose or role of evaluation, for example, in furthering social justice versus promoting evaluation use for organizational learning. Two examples illustrate these different values. Ernest House, an advocate of the social justice perspective states, "I worry that the interests, values, and perspectives of the poorest, the weakest, and the less vocal will be excluded from the study [...] I worry that the evaluation study will reinforce the interests of the powerful, even if inadvertently" (House, 2003: 54). In contrast, Brad Cousins and Lorna Earl who are interested in evaluation use for the promotion of organizational learning, "see participatory evaluation as a powerful learning system designed to foster local applied research and thereby enhance social discourse about relevant organizational issues. [For them,] the requirement of direct involvement in the research process [...] will heighten opportunities for staff to discuss process and outcome data, to rethink their conceptions and challenge basic assumptions [...]" (Cousins and Earl, 1992: 401).

In the first case, social justice is clearly the evaluator's main concern, and this implies different rules for the involvement of stakeholder groups than in the second case, which is mainly concerned with the utilization of the evaluation process and results. Melvin Mark and Lance Shotland offer an excellent discussion of the role of values in stakeholder-based evaluation, differentiating two criteria for stakeholder selection: power and legitimacy.

They argue that an evaluator concerned with utilization for decision making would tend to work with high-power stakeholders, while an evaluator with a desire for social justice through evaluation would favor a large group of stakeholders, including low-power stakeholders, thus giving them voice.

Even if an evaluator intended to serve both purposes, few evaluations have the resources necessary to involve the full range of stakeholders in a just and equitable as well as deep and active manner. In other words, feasibility usually demands a trade-off between depth and breadth of stakeholder participation. Although many evaluation scholars see the decision of how to shape this trade-off as value based and therefore quite invariant over time, one could imagine an evaluator deciding situationally which value or purpose to prioritize, based on perceived needs as well as on practical and strategic considerations.

Deep involvement for the purpose of maximizing evaluation use

Participatory, empowerment, and utilization-focused evaluation are some of the approaches most inclined to involve few stakeholders actively in order to maximize positive effects such as empowerment and evaluation utilization. We cannot offer a comprehensive review of this literature here, but will include a short discussion of Michael Patton's utilization-focused evaluation to illustrate the possible actions and rationale of the utilization-focused evaluator for involving primary intended users in the evaluation.

The evaluator identifies primary intended users of the evaluation early, upon being contacted to do the evaluation. Primary intended users are people (specific people by name) who have a vested interest in the evaluation, not only as an activity to fulfill external requirements, but also to take decisions based on what they have learned from the evaluation in terms of the main purposes and uses which the evaluation serves. Importantly, it is the primary intended users' questions that the evaluator sets out to answer in the evaluation. Likewise, the evaluator plans the evaluation, defines its methods, collects and analyzes data, and interprets results in close collaboration with the primary intended users to ensure their ownership of the evaluation and commitment to evaluation use. Ownership and use are more likely to occur if intended users have been actively involved in the evaluation. The extent of involvement is subject to negotiation with primary users and is contingent on the evaluation's purpose and other situational factors, which the evaluator should be aware of and remain sensitive to throughout the evaluation. Patton refers to this cognizance and sensitivity as the "active-reactive-adaptive-interactive" approach: The evaluator's role will depend on the context and purposes of the evaluation as negotiated with primary intended users" (Patton, 2008: 122). Similar to utilization-focused evaluation, practical participatory evaluation is said to be most effective when participation

in the evaluation is limited to primary users or those with a vital interest in the program, and when these primary users exert control over the evaluation process and are actively involved in all its phases.

Broad involvement for the purpose of promoting social justice

Ernest House and Kenneth Howe, in describing their deliberative democratic evaluation approach and its three requirements – inclusion, dialog, and deliberation – acknowledge that "the procedures and techniques for meeting [these] requirements are more raw, untested, and uncertain than the technical data collection and analysis procedures developed over the past many decades. Much work needs to be done [...]" (House and Howe 2000: 11).

Jennifer Greene describes a concrete set of processes and methods to promote stakeholder involvement in a democratic deliberative evaluation. Examples include use of an evaluation committee as a forum for inclusive conversations, public forums planned as open reports on the progress of the evaluation and as open invitations for discussion and dialog, a mix of methods, sampling strategies directed toward diversity, and criteria to judge program quality that are multiplistic and broad. In the case she describes, however, these practices met with little success.

Turning to evaluation textbooks in search of more concrete advice, we find that Jody Fitzpatrick *et al.*, for example, promote the involvement of a broad set of stakeholders in the planning (divergent phase) of the evaluation and, most importantly, in the process of identifying evaluation questions. These authors propose a matrix for ranking the evaluation questions that emerge during this phase to help select those that are of most interest to key audiences, reduce uncertainty the most, and are critical to the study's scope and comprehensiveness, which they call the convergent phase.

Another example of an evaluation approach that calls for broad stakeholder involvement is stakeholder-based evaluation, as developed by the National Institute of Education (NIE). The intention of the NIE was to encourage broad involvement of stakeholders, at least in the planning phase of the evaluation. Empirical evidence regarding this approach can be found in the next section.

Empirical Research on the Effects of Stakeholder Involvement in Program Evaluation

The benefits of the participatory and utilization-focused approaches have been studied in some detail, so we limit our discussion to a few seminal studies. For example, Jennifer Greene published a set of articles on her work about stakeholder participation and evaluation utilization (Greene, 1987, 1988). She discusses three factors necessary to successfully link stakeholder participation and increased utilization. First, the evaluator should involve

appropriate stakeholders, that is, people who are legitimate program stakeholders possessing enough program knowledge (but not necessarily evaluation knowledge); who have a high, self-defined stake in the program; and who come from “organizations that are receptive to evaluation and perhaps that are democratic, decentralized, collegial, and participatory themselves” (Greene, 1987: 390). Second, the evaluator needs to use an appropriate process for involvement, that is, one characterized by frequent group discussions providing multiple, cumulative opportunities for participation and including tangible evidence that the participants contributed to evaluation-related decisions. Third, Greene (1988) highlights evaluator skills such as responsiveness, listening skills, acceptance of diverse stakeholder views, ability to invoke trust and rapport, as well as technical skills and status as an impartial, credible outsider. Greene’s research showed that stakeholders benefited in two ways from their participation. First, they learned more about their work, the program, and the organization. Second, they learned more about evaluation and developed more favorable attitudes toward evaluation. Greene links these conceptual (process) uses of evaluation to the cognitive, and, to a lesser degree, the affective dimensions of the participatory process.

Another case study discussing the positive outcomes of evaluations that deeply involve a small group of stakeholders comes from Torres *et al.* (2000). These authors summarize the benefits of their collaborative evaluation as “time for reflection, learning on numerous levels, efficiencies in data collection and program planning, deepened professional and personal relationships, and the conviction to take action” (p. 35). The authors go on to discuss the conditions necessary for these positive outcomes to become possible: “arguably the single most important requisite for collaborative, participatory evaluation was present from the outset: administrative and leadership commitment to evaluation for the purposes of learning and improvement.” In their case, this meant providing the motivation and resources, specifically time scheduled to engage in fruitful dialogs and beneficial reflections. Other studies have also provided evidence that active stakeholder participation increases evaluation use (e.g., O’Sullivan and D’Agostino, 2002; Turnbull, 1999), and more specifically, process use (e.g., Forss *et al.*, 2002; Preskill *et al.*, 2003; Taut, 2007). In addition, there is some indication that the validity of evaluation conclusions can be improved when the evaluator successfully collaborates with those stakeholders who are knowledgeable about the evaluated program; taps into their program expertise by applying appropriate methods for stakeholder participation; and ensures that no group’s expertise is ignored (Brandon, 1998).

Cousins and Earl (1992) offer a summary of a number of organizational and evaluator characteristics that provide positive conditions for participatory evaluation.

According to their review of the literature, evaluator characteristics include motivation to work directly with participants, readiness to take on the role of a teacher of evaluation, and tolerance toward problems and imperfections. Organizational requirements include an evaluation culture, time, and resources to undertake the participatory evaluation activities, and participants motivated to get involved in the evaluation. Greene summarized her experiences regarding the influence of the evaluation and program context on successful implementation of participatory evaluation as follows: “Participatory evaluation is readily accepted in contexts that share its value commitments [...]. Finding acceptance and understanding of participatory evaluation is much more difficult in contexts where (1) traditional evaluation is known and familiar and other approaches to evaluation are unknown and thereby suspect, (2) value commitments are not consonant, (3) the values of participatory evaluation are actually threatening to some stakeholders, usually those with extant decision making authority and power, and (4) there is considerable stakeholder disengagement and apathy” (see Ryan *et al.*, 1998: 106).

Despite the positive outcomes associated with the approaches described, the literature also discusses their limitations. Cousins (2003) points to the key challenges of participatory evaluation: the need for technical sophistication and external credibility of the evaluation on the one hand, and the responsiveness and extent of participation of stakeholders on the other. Regarding the issue of credibility, another challenge of active stakeholder participation is to ensure that the evaluation does not get biased by giving the advantage of direct and active involvement only to a small (and thereby privileged) group, or because the evaluator is perceived to lose independence by giving up control of the process and closely collaborating with program staff.

As for evaluations involving a broad range of stakeholders, but less actively, early evidence on the effects of the stakeholder-based evaluation approach comes from Bryk (1983). Bryk’s volume presented two case studies meant to test stakeholder-based evaluation as it had been developed by the NIE. However, the evidence shows that the two case studies did not fully implement the approach as it had been intended. Stake (1983) describes one of the evaluations, of the Cities-in-Schools program, and gives the idea that few stakeholders were actually involved (mainly program staff), and that their participation in terms of actual influence on the evaluation process was rather limited. According to Stake, “nothing was noted to suggest that use of the stakeholder approach made the evaluation study more fair, more accurate, or more efficient. [...] In summary, the evaluation team did not take full advantage of the stakeholder concept, limiting its use to information enhancement” (p. 29).

Likewise, Farrar and House’s (1983) discussion of the evaluation concerning the Push/Excel program revealed

that, “all in all, the stakeholders contributed little to the evaluation design or to the plan for disseminating findings” (p. 45). Gold (1983), on the other hand, who represented the NIE, concluded that stakeholder involvement was active (although with only marginal impact on the evaluation design itself), that the evaluation was in fact used for formative purposes, and that the approach assured access and cooperation for the evaluation. Similarly, the principal evaluator of the two programs, Charles Murray, came to the conclusion that, on the positive side, “the stakeholder approach is a useful device for getting the leading players to cooperate, for understanding a program intimately, for attracting attention to interim evaluation findings, and perhaps even for getting decision makers to take evaluation findings into account [...]” (Gold, 1983: 59).

Mercier (1997) offers a description of a stakeholder-based evaluation that neither increased the fairness of the evaluation’s conduct, nor enhanced the use of its findings by the represented individuals and agencies. The author points to a number of factors responsible for this failure. These include distrust regarding the evaluation by some groups, inequalities among the groups regarding organizational culture, status, and what was at stake for them in the evaluation, and a restrictive time frame of the evaluation. The stakeholder groups used the evaluation competitively to defend political and financial interests, and the strategy of appointing a steering committee did not help resolve these issues. This could have also been due to a lack of evaluator skills, particularly regarding group facilitation and conflict resolution. Other authors have also supported the latter point by noting that the more heterogeneous the stakeholder group, the greater were the facilitation skills required by the evaluator (Torres *et al.*, 2000; Torres and Preskill, 1999).

The literature describes other potential limitations or difficulties in the context of applying the stakeholder-based evaluation approach. For example, stakeholders would not be able to specify their information needs clearly because they lacked knowledge about what evaluation can offer; the multiple information demands of various stakeholder groups exceeded the research capacities of an evaluation; technical compromises diminished the quality of the evaluation; contamination of the evaluator’s judgments occurred; and conflicts between evaluation purposes and the interests of certain stakeholder groups could bias the evaluation (Mark and Shotland, 1985; Murray, 1983; Gold, 1983). Similarly, O’Brecht (1992) examined the use of evaluation steering committees in a government agency and concluded that broad stakeholder involvement increased evaluation costs and duration, reduced flexibility in evaluation planning, and could diffuse responsibilities and dilute commitment to the evaluation. Limitations are also due to a lack of effective evaluator strategies for involving a broad range of divergent stakeholder groups in an evaluation

(Gold, 1983). An indirect result of this observation is House’s (2003) call upon the profession to “develop methods for controlling [...] stakeholder biases, just as we have methods for controlling other biases” (p. 54).

In summary, the empirical evidence is based primarily on case studies conducted in the context of participatory evaluation (deep involvement) and stakeholder-based evaluation (broad involvement). The studies we have presented use systematic data collection in order to derive credible conclusions related to the evaluation process and evaluator and contextual characteristics that enable or hinder desired evaluation outcomes for stakeholders. The studies we reviewed show that, first of all, the implementation of evaluations prioritizing deep over broad involvement seems less prone to failure. Furthermore, actively involving stakeholders to attain evaluation use seems to have been better documented in the literature. However, approaches that tend to cursorily engage stakeholders to obtain desired outcomes, such as giving all stakeholder groups a voice, appear less so. Therefore, in the section titled ‘Synthesis and conclusions’, we focus on deep involvement approaches and summarize the factors that link deep involvement to increased evaluation use.

Synthesis and Conclusions

Conclusions on the Importance of Stakeholder Involvement

There are some who diminish the importance of stakeholder involvement. Their argument centers on concern about the introduction of political interests and competing points of view into the process, potentially compromising the evaluation’s credibility. Indeed, we would maintain that political interests and conflicting points of view are ever present in any evaluation and that the involvement of stakeholders makes explicit what would otherwise be implicit. A similar argument and rebuttal can be made with respect to the introduction of potential stakeholder bias. These arguments are also reflected in the *Program Evaluation Standards* that relate to the stakeholder topic. Further, we acknowledge the skills that are necessary to work successfully with stakeholders and recognize that the act of engaging stakeholders in the evaluation process could increase the costs of conducting the evaluation.

However, we believe that the importance of stakeholder involvement far exceeds any potential detriments. Much depends upon the extent to which stakeholder participation is fully and appropriately implemented, which in turn depends on evaluator and stakeholder characteristics and contextual prerequisites.

There are multiple ways in which stakeholder involvement is important. The first of these is that it provides stakeholder participants an opportunity to learn more about their work, about their program, about their organization and about evaluation. Second, when stakeholder

participation is real and means some degree of control over the evaluation by way of evaluation-related decision making, this results in a sense of ownership of the evaluation. Moreover, a positive evaluation experience generally leads to a more favorable attitude toward evaluation.

We maintain, further, that stakeholder involvement leads to improved validity of the evaluation conclusions. By tapping into stakeholder program knowledge and expertise the evaluator, provides a more relevant and appropriate evaluation.

Finally, increased understanding about the program, ownership of the evaluation, more favorable attitudes toward evaluation, and more relevant and valid evaluative conclusions increase the conviction to take action. Action may be on the part of stakeholders themselves or it may involve their role in getting decision makers to pay heed to evaluation findings.

Summary of Factors Linking Deep Stakeholder Involvement to Increased Evaluation Use

Some of the factors that can impact evaluation use are as follows:

1. *Whom to involve.* Not only evaluation theoretical writings, but also much of the empirical evidence presented above call for the involvement of fewer stakeholders, particularly if the intended focus of the evaluation is to foster evaluation use. When a small group of key intended users is involved, dangers such as lack of commitment, political abuse, and inflexible evaluation planning are less likely to occur. This does not negate obtaining limited input from a wide group of stakeholders prior to the active engagement phase.
2. *Participatory evaluation process characteristics.* The reviewed studies show that successful evaluation involving stakeholders must be characterized by multiple, cumulative opportunities for participation, enabling participants to effectively contribute to evaluation-related decisions. Evaluators should strive to make evaluation more efficient and effective by weaving evaluation activities into ongoing work routines.
3. *Evaluator skills for promoting effective participation.* Many authors highlight the importance of an evaluator's group facilitation and conflict-resolution skills. Jennifer Greene specifically calls for evaluators to invoke trust, show listening skills, responsiveness, and acceptance of diverse stakeholder views. Along these lines, Michael Patton promotes an active–adaptive–reactive–interactive evaluator role and stresses the importance of an evaluator's interpersonal skills.
4. *Adequate resources for participatory processes.* This is a key factor in facilitating the successful involvement of stakeholders. Evaluators often struggle with resource constraints in the sense that stakeholder suggestions for

evaluation cannot be implemented, thus giving them the feeling that they have no control over the process despite the fact that the evaluation is formally labeled participatory.

5. *Leadership support for involving stakeholders in evaluation.* Some studies mention leadership support as a major enabling factor in their work.
6. *Stakeholder knowledge about evaluation.* Brad Cousins describes quality concerns regarding stakeholder contributions to evaluation as one difficulty in conducting participatory evaluation. However, other authors do not share these concerns based on their experiences.
7. *Trust in participatory evaluation.* Some authors highlight the topic of trust between stakeholders and evaluator as a main factor for successful implementation of their participatory evaluation; others mention it as a major factor in evaluations failing to produce the desired effects. Diverse stakeholder values, especially those diverging from a participatory philosophy, are closely connected to issues of trust.

In summary, the key factors we have highlighted above have to do with evaluation process characteristics (who gets involved and how), evaluator skills, stakeholder predispositions, and contextual prerequisites. We hope to have given an overview of those factors that, according to the literature, seem crucial when trying to involve stakeholders in order to increase evaluation utilization.

See also: Defining Quality in Evaluation; Educational Evaluation: Concepts, Practice, and Future Directions; Evaluation and Accountability; Evaluation of Teacher Quality and Practice; Evaluation Use; Moral and Ethical Issues in Evaluation; Principal Evaluation: Concepts, Needs and Realities; School Based Evaluation: Purposes, Protocols and Processes; School Inspection/ External School Evaluation; The Purpose of Educational Evaluation; Understanding Approaches to Evaluation.

Bibliography

- Brandon, P. (1998). Stakeholder participation for the purpose of helping ensure evaluation validity: Bridging the gap between collaborative and non-collaborative evaluations. *American Journal of Evaluation* 19(3), 325–337.
- Bryk, A. (ed.) (1983). *Stakeholder-Based Evaluation. New Directions for Program Evaluation*, No. 17. San Francisco, CA: Jossey-Bass.
- Cousins, B. and Earl, L. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis* 14(4), 397–418.
- Cousins, J. B. (2003). Utilization effects of participatory evaluation. In Kellaghan, T., Stufflebeam, D., and Wingate, L. (eds.) *International Handbook of Educational Evaluation*, pp 245–266. Boston, MA: Kluwer.
- Farrar, E. and House, E. (1983). The evaluation of push/excel: A case study. In Bryk, A. (ed.) *Stakeholder-Based Evaluation. New Directions for Program Evaluation*, No. 17, pp 31–57. San Francisco, CA: Jossey-Bass.

- Forss, K., Rebién, C., and Carlsson, J. (2002). Process use of evaluation: Types of use that precede lessons learned and feedback. *Evaluation* 8(1), 29–45.
- Gold, N. (1983). Stakeholder and program evaluation: Characterizations and reflections. In Bryk, A. (ed.) *Stakeholder-Based Evaluation. New Directions for Program Evaluation*, No. 17, pp 63–72. San Francisco, CA: Jossey-Bass.
- Greene, J. (1987). Stakeholder participation in evaluation design: Is it worth the effort? *Evaluation and Program Planning* 10, 379–394.
- Greene, J. (1988). Stakeholder participation and utilization in program evaluation. *Evaluation Review* 12(2), 91–116.
- House, E. (2003). Stakeholder bias. In Christie, C. (ed.) *The Practice–Theory Relationship in Evaluation, New Directions for Evaluation*, vol. 97, pp 53–56. San Francisco, CA: Jossey-Bass.
- House, E. and Howe, K. (2000). Deliberative democratic evaluation. In Ryan, K. and DeStefano, L. (eds.) *Evaluation as a Democratic Process: Promoting Inclusion, Dialogue, and Deliberation. New Directions for Program Evaluation*, No. 85, pp 3–12. San Francisco, CA: Jossey-Bass.
- Joint Committee on Standards for Educational Evaluation (1994). *The Program Evaluation Standards*, 2nd edn. Thousand Oaks, CA: Sage.
- Mark, M. and Shotland, L. (1985). Stakeholder-based evaluation and value judgements. *Evaluation Review* 9(5), 605–626.
- Mercier, C. (1997). Participation in stakeholder-based evaluation: A case study. *Evaluation and Program Planning* 20(4), 467–475.
- Murray, C. (1983). Stakeholders as deck chairs. In Bryk, A. (ed.) *Stakeholder-Based Evaluation. New Directions for Program Evaluation*, No. 17, pp 59–61. San Francisco, CA: Jossey-Bass.
- O'Brecht, M. (1992). Stakeholder pressures and organizational structure for program evaluation. *Canadian Journal of Program Evaluation* 7(2), 139–147.
- O'Sullivan, R. and D'Agostino, A. (2002). Promoting evaluation through collaboration. *Evaluation* 8(3), 372–387.
- Patton, M. Q. (2008). *Utilization-Focused Evaluation*, 4th edn. Thousand Oaks, CA: Sage.
- Preskill, H., Zuckerman, B., and Matthews, B. (2003). An exploratory study of process use: Findings and implications for future research. *American Journal of Evaluation* 24(4), 423–442.
- Ryan, K., Greene, J., Lincoln, Y., Mathison, S., and Mertens, D. (1998). Advantages and challenges of using inclusive evaluation approaches in evaluation practice. *American Journal of Evaluation* 19(1), 101–122.
- Stake, R. (1983). Stakeholder influence in the evaluation of cities-in-schools. In Bryk, A. (ed.) *Stakeholder-Based Evaluation. New Directions for Program Evaluation*, vol. 17, pp 15–30. San Francisco, CA: Jossey-Bass.
- Taut, S. (2007). Studying self-evaluation capacity building in a large international development organization. *American Journal of Evaluation* 28(1), 45–59.
- Torres, R., Stone, S. P., Butkus, D., et al. (2000). Dialogue and reflection in a collaborative evaluation: Stakeholder and evaluator voices. In Ryan, K. and DeStefano, L. (eds.) *Evaluation as a Democratic Process: Promoting Inclusion, Dialogue, and Deliberation. New Directions for Evaluation*, vol. 85, pp 27–38. San Francisco, CA: Jossey-Bass.
- Torres, R. and Preskill, H. (1999). Ethical dimensions of stakeholder participation and evaluation use. In Fitzpatrick, J. and Morris, M. (eds.) *Current and Emerging Ethical Challenges in Evaluation. New Directions for Evaluation*, vol. 82, pp 57–66. San Francisco, CA: Jossey-Bass.
- Turnbull, B. (1999). The mediating effect of participatory efficacy on evaluation use. *Evaluation and Program Planning* 22, 131–140.

Further Reading

- Alkin, M. (ed.) (2004). *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks, CA: Sage.
- Alkin, M., Hofstetter, C., and Ai, X. (1998). Stakeholder concepts. In Reynolds, A. and Walberg, H. (eds.) *Advances in Educational Productivity*, vol. 7, pp 87–113. Greenwich, CN: JAI Press.
- Chen, H. (1990). *Theory-Driven Evaluation*. Newbury Park, NJ: Sage.
- Christie, C. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. In Christie, C. (ed.) *The Practice–Theory Relationship in Evaluation, New Directions for Evaluation*, vol. 97, pp 7–35. San Francisco, CA: Jossey-Bass.
- Donaldson, S. (2007). *Program Theory-Driven Evaluation Science: Strategies and Applications*. Oxford: Taylor and Francis.
- Fetterman, D., Kaftarian, S. J., and Wandersman, A. (eds.) (1996). *Empowerment Evaluation*. Thousand Oaks, CA: Sage.
- Fitzpatrick, J., Sanders, J., and Worthen, B. (2003). *Program Evaluation: Alternative Approaches and Practical Guidelines*, 3rd edn. Englewood Cliffs, NJ: Allyn and Bacon.
- Greene, J. (2000). Challenges in practicing deliberative democratic evaluation. In Ryan, K. and DeStefano, L. (eds.) *Evaluation as a Democratic Process: Promoting Inclusion, Dialogue, and Deliberation. New Directions for Program Evaluation*, vol. 85, pp 13–26. San Francisco, CA: Jossey-Bass.
- Guba, E. and Lincoln, Y. (1989). *Fourth Generation Evaluation*. Newbury Park, CA: Sage.
- House, E. (1976). Justice in evaluation. In Glass, G. V. (ed.) *Evaluation Studies Annual Review*, vol. 1, pp 75–100. Beverly Hills, CA: Sage.
- King, J. and Ehler, J. (2008). What we learned from three evaluations that involved stakeholders. *Studies in Educational Evaluation* 34(4), 194–200.
- Mathie, A. and Greene, J. (1997). Stakeholder participation in evaluation: How important is diversity? *Evaluation and Program Planning* 20(3), 279–285.
- Morris, D. (2002). The inclusion of stakeholders in evaluation: Benefits and drawbacks. *Canadian Journal of Program Evaluation* 17(2), 49–58.
- Patton, M., Grimes, P., Guthrie, K., et al. (1977). In search of impact: An analysis of the utilization of the federal health evaluation research. In Weiss, C. (ed.) *Using Social Research in Public Policy Making*, pp 141–164. Lexington, MA: Lexington Books.
- Poth, C.-A. and Shulha, L. (2008). Encouraging stakeholder engagement: A case study of evaluator behavior. *Studies in Educational Evaluation* 34(4), 218–223.
- Weaver, L. and Cousins, B. (2004). Unpacking participatory evaluation. *Journal of MultiDisciplinary Evaluation* 1(1), 19–40.
- Weiss, C. and Birckmayer, J. (2000). Theory-based evaluation in practice: What do we learn? *Evaluation Review* 24, 407–431.

EDUCATIONAL EVALUATION – EVALUATION DOMAINS

Contents

Curriculum Evaluation

Evaluating Education in Three Policy Eras

Evaluation of Mathematics Education Programs

Evaluation of Science Education Programs

Evaluation of Social Studies and Civic Education Programs

Evaluation of Teacher Quality and Practice

International Evaluations

National Assessment Programs: The Example of the U.S. National Assessment of Educational Progress

Principal Evaluation: Concepts, Needs and Realities

Program Evaluation

School Based Evaluation: Purposes, Protocols and Processes

School Inspection/External School Evaluation

The International system for Teacher Observation and Feedback: A Theoretical Framework for Developing International Instruments

Curriculum Evaluation

T Levin, Tel Aviv University, Tel Aviv, Israel

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Adaptive and situated evaluation – An evaluation orientation which conceives curriculum evaluation as an adaptive, productive knowledge construction process driven by in-school, situated, specific considerations and needs but directed and conditioned by mandated and prescribed evaluation policy.

Emergent and emancipatory evaluation – An evaluation orientation aiming to influence and understand the curriculum evolution process and conceives curriculum evaluation as a continuous knowledge co-construction process in individual contexts and through social collaboration, dialogical-negotiation, and authentic experiences regarding curriculum meanings, actions, and values.

Instrumental evaluation – An evaluation orientation which conceives curriculum evaluation as a technical-rational productive and dissemination process, with an instrumental rationale, aiming and valuing control, order, and standardization.

Although linked to the general field of educational evaluation, nowadays curriculum evaluation is considered an area within curriculum development and innovation. It is consequently conceived as an intervention that can affect what is taught and learned, and which aims to describe, judge, qualify, understand, and/or enhance new knowledge concerning school curricula. The processes and outcomes of curriculum evaluation are based on the ontological, epistemological, and methodological assumptions that are reflected in the types of evaluation questions asked, and on the evaluation goals expected to be realized. They are also based on the evidence used, the type of interpretations considered valid or useful; the value system chosen and meaning ascribed to quality; the curricular context of the evaluation process and the evaluator identity and roles.

Curriculum Evaluation in Transition

Curriculum evaluation draws on two distinct and complex fields – curriculum and evaluation – both of which encompass dozens of different definitions, approaches, and methods. Thus, curriculum evaluation can neither be simple nor standardized. There are 30 different evaluation

approaches listed in the literature, each of which has a different meaning, a different form, and different factors affecting it (Patton, 1986). This shows that there is no ideal, all-purpose approach to evaluation suitable for every occasion. The multiplicity of evaluation typologies, models, concepts, and methodologies suggest a development from a monolithic conceptualization to a pluralist one containing multiple methods, measures, criteria, perspectives, audiences, and even interests. Evaluation has lost its exclusive reliance on technical and analytical procedures favoring negotiation instead. In methodological terms, this has marked a shift away from the early quantitative emphasis toward an emphasis that is less rigid and finds a variety of qualitative research methods and measures and mixed methodologies acceptable (Guba and Lincoln, 1989; House, 1993; Stake, 1995). These changes essentially reflect a departure from the traditional logical positivistic approach of evaluation as the work of a lone individual, to a constructively orientated, collaborative inquiry approach, and interpretivist philosophy in which the entire community participates to construct new knowledge.

Likewise, we are still very far from consensus on the questions of defining what the curriculum is, what it should be, how to plan and implement it, and who should make curricular decisions. Views of the meaning of the curriculum range from the narrower view, which sees the curriculum as a given set of knowledge, skills, and activities to be delivered to students, to an expanded view which takes full account of the cultures, languages, and lived and emergent experiences of the teachers and students. Curricular approaches also differ in their assumptions, values, in the epistemological, pedagogical, and organizational perspectives inherent in curriculum development decisions and practices, and in the scope of curriculum development technique (Schubert, 1996). These changes reflect a development away from an approach that is mechanistic, linear, authoritative, controlled, pre-planned and rooted in the positivist orientation, and toward a nonlinear, system-oriented, flexible, tentative, and emergent approach marked by open-ended planning and freedom to accept the challenges of spontaneity and situativity, which is grounded in constructivist and complexity theories.

However, despite the conceptual changes in both the curriculum and evaluation fields, a review of the relevant literature shows that neither curriculum evaluation practices nor curriculum evaluation methodology have changed much over the last 20 years and that curriculum evaluation still uses obsolete models and methods (Jasparro, 1998). According to Patton (1998), evaluators unconsciously tend to fall back on old familiar patterns even when facing new situations. There is thus a gap separating curriculum evaluation theory and practice. This is particularly true for federally funded and prescribed curricular programs.

Paradigmatic Orientations to Curriculum Evaluation

Based on these paradigmatic changes and generalizing from the multitude models of evaluation and the diverse meanings of curriculum, different models of curriculum evaluation have been defined. These reflect the conceptualizations of curriculum evaluators regarding the meanings, functions, methodologies, values, and uses of curriculum evaluation (Aoki, 1978; Lather, 1991; Leathwood and Phillips, 2000; Levin, 2002; MacDonald, 1974; Melrose, 1998; Norris, 1998). Each researcher reflects his/her own body of knowledge and professional experience, defining specific schemes, and using his/her own unique language to describe developments in the twin fields of evaluation and curriculum. Each differentiates between various approaches, assuming them legitimate and effective under different curricular circumstances or given different ideological assumptions.

The various approaches to curriculum evaluation can be grouped into three different orientations based on the epistemological views, educational and methodological beliefs, and axiological perspectives that they espouse. Each orientation imputes different meaning to the evaluation process, reflecting the various mindsets regarding the meaning of evaluation, evaluation goals, functions, and design structure, evaluators' identities and roles, the value system involved and the evaluation's uses. The following three orientations are characterized as follows: (1) instrumental evaluation; (2) adaptive and situated evaluation; and (3) emergent and emancipatory evaluation.

Instrumental Evaluation

The first approach to curriculum evaluation, conceived as a technical-rational, productive, and dissemination process, has an instrumental rationale. Its goal is largely one of control. Its main purpose is to provide information for instrumental reasons, dictated by an authority. This approach to curriculum evaluation is considered technical since it regards curricular decisions or actions as standardized, systematic, and objectively reached mainly by evaluation experts, based on a means-ends rationale. It is productive in that it conceives curriculum evaluation as a process whose main purpose is rendering unequivocal judgment or an absolute curricular decision. It is disseminative in that it regards evaluation as a top-down model of search for a proven model of a quality curriculum to be generalized and assimilated everywhere. Its key value is that of order and congruence, and evaluators working within this orientation often evaluate alone as an external expert or unbiased outsider. In this top-down approach, curriculum evaluation is an instrument for improving and encouraging commitment to the curriculum. It reflects an audit mentality and compliance monitoring, and as such,

its discourse contains concepts like expectations, discrepancies, fidelity, congruence, precision, satisfaction, feedback and correction, clients, and products.

This curriculum evaluation orientation assumes there is a concrete truth to be discovered regarding the worth of the curriculum and that evaluation can provide either a correct answer to a curriculum problem or show whether a curriculum should maintain its present form (Melrose, 1998). It presumes that scenarios can be predicted and contrived and that order, stability, content, and meaning are imposed on the curriculum. This orientation represents a culture that puts a predefined ideology of a centralized authoritative figure into practice and reflects a belief in power and control, the universality and stability of knowledge, and the existence of well-defined and authoritative sources of knowledge. This orientation has three main characteristics: (1) the evaluators' input is limited to technical decision making; (2) the impact on the curriculum is limited leading to little or no understanding of the specific circumstances of the evaluation; and (3) there is a monologic relationship between curriculum developers, users, and evaluators, where the set of evaluation standards authoritatively dictates expectations of a quality curriculum. In sum, this orientation applies the one-size-fits-all approach to curriculum evaluation, ignoring its danger of overlooking the specific needs of local communities and subgroups.

The basic rules guiding this orientation are standards and the search for the best practices. It employs predefined, externally mandated standards, and standardization practices. Curriculum quality is thus determined through judgment and is mainly interested in discrepancies between the planned and the achieved curriculum; whether a curriculum succeeds in producing student knowledge or skills as expected, and whether it is more efficient than a different competitive curriculum. The most important and most desired quality measures are of the curriculum's effectiveness and efficiency relative to the mandated standards. In other words, instrumental evaluation assesses the merit of a curriculum in terms of the number and quality of its outputs and the efficiency obtained in transforming inputs into outputs. Instrumental evaluation studies are therefore analytical and comparative and reflect the ideology of customer accountability.

This approach seems best suited to curricula that are defined as a program of planned activities or intended learning outcomes, or as cultural reproductions. Influenced by the behavioral objectives or goal-attainment models of Tyler (1949) and Popham (1988), this orientation corresponds with Aoki's (1978) ends-means orientation; MacDonald's bureaucratic evaluation (MacDonald, 1974); Melrose's (1998) functional evaluation paradigm; with what Lather (1991) refers to as the realist tale, and with Leathword and Phillips' (2000) quality assurance orientation.

Situated and Adaptive Evaluation

The second evaluation orientation conceives curriculum evaluation as an adaptive, productive knowledge construction process driven by in-school, situated, specific considerations and needs but directed and conditioned by mandated and prescribed evaluation policy. It thus conceives curriculum evaluation as occurring within the unique context of a school with its teachers, and students while considering their individual and curricular needs, and expectations, beliefs, capabilities, and actions, though still framed and controlled by a central authority. It combines top-down and bottom-up evaluation processes. This means that by limiting hierarchical structure and control, curriculum evaluation shares both power and control between local and centralized agents. Whereas the instrumental evaluation orientation focuses on prediction, alignment, accuracy, and control, the situated and adaptive evaluation orientation stresses understanding and seeks to improve a practice by applying situated-local wisdom. It is a flexible process whose discourse employs such concepts as circumstances, uniqueness, local needs, situatedness, responsiveness, deliberation, stakeholders, consensus, dialogue, conditionality-autonomy, and products.

This orientation mainly exists in educational contexts that conceive schools as social organisms, which struggle to adapt to curricular demands both from within and beyond their boundaries. In this sense, while broad goals and curriculum evaluation frameworks may be – and usually are – nationally or state determined, the evaluation process has the defined responsibility of transforming and/or adapting these generalities to form a specific, local-situated evaluation. Thus, the considerations driving the curriculum evaluation processes and responsibilities of this orientation cannot be determined either by relating to the individual school or with reference simply to the centralized authority's decisions regarding what schools generally are required to do. Rather a sophisticated blend of the two is needed in which schools and teachers have a wide and complex set of curriculum evaluation roles to perform. This orientation has three characteristics: (1) the evaluation experience involves questioning curricular goals and processes with a view to improving the curriculum, its relevancy, and its accomplishments for specific circumstances; (2) the evaluation process is deliberative and aims to achieve knowledge growth by responding to local perspectives and needs, theoretical considerations, and mandated guidelines; and (3) the relationship is dialogic: evaluators, authority representatives, and local stakeholders, mainly teachers, enter into dialogue aimed at understanding and influencing the evaluation processes, and arriving to a consensus concerning dimensions of curriculum quality.

Through its sensitivity to both the general and locally situated environment, this orientation conceives reality as

multiple, divergent, and interrelated rather than as the singular, convergent, and fragmented reality conceptualized by the instrumental orientation to evaluation. It conceives knowledge as an end product, but nevertheless reflects a belief that all knowledge is socially, culturally, and geographically situated. It also sees practical knowledge as constructed during a process of negotiation and recognizes a consensus among the stakeholders regarding the justification of the knowledge. In particular, it calls for consensual decision making among stakeholders based on criteria relating to contents, processes, and outcomes, and a quest for situated and adjusted evaluation criteria concerning what and how curricular components function in a given school context. Standards-setting processes are often considered participatory, with input sought from teachers and lower-level school/local administrators. Yet, this is still a top-down approach, since standards are first set at the national or state level after which they trickle down to districts and schools, which are compelled, often through incentives such as standardized testing, to adopt the official dominant standards.

Adaptive and situated evaluation seeks optimum practices which are aligned to the circumstances in which the curriculum is implemented or examined, and assumes interdependence between circumstances, goals, processes, and outcomes. It also seeks a more general knowledge that pinpoints differences and similarities between situations, aiming to contribute to a better understanding of a specific curriculum or curricular policy. Therefore, differences between required curricular goals, processes, or products and implemented and achieved ones are examined in order to understand and learn from the discrepancies rather than seeing the discrepancies as problems to be solved. In this way, curriculum quality is defined by how far the goals and expectations accepted by all stakeholders are fulfilled, taking into account the characteristics of the specific situation relative to the more general mandatory circumstances. Here, curriculum quality is based on curriculum description and curricular judgments, which are both conceived as products of equal significance with its own unique role (Stake, 1975). Typically, the main evaluative focus of this orientation is whether a curriculum succeeds in producing students' knowledge or skills as expected and whether it plans and uses locally meaningful curriculum structure and processes. It also focuses on describing and understanding how, and regarding which dimensions, the curriculum differs from, or alternatively complies with, the more general quality characteristics of the national curriculum.

Thus, for example, curriculum evaluation can engage both teachers and school-based curricular experts, who, along with an external evaluator or external evaluation team share responsibility for the evaluation. Besides the standardized evaluation goals, tools, methodology, and evaluation standards, the dialogue between the external

and internal evaluation agents should consider the specific and situated (local) evaluation processes including locally set curricular goals, curriculum sequence or structure, evaluation methodology and curriculum tools, timing, and criteria. Examining situated evaluation data in conjunction with standardized nationally determined data allows authentic and conditional curriculum evaluation knowledge to emerge. The shared evaluation wisdom orchestrated by the evaluation team (external and internal combined) reaches toward consensual evaluation decisions and is expected to provide evaluation insights that explain the value of the situated curriculum versus the normative curriculum and its appropriateness to a given classroom, school, or district. This process can enhance our understanding of the curriculum's value and flexibility and its potential for adapting different curriculum dimensions (goals, structure, instructional strategies, assessment procedures, etc.) to different situations. It can also, for example, incorporate other curriculum goals besides those implemented on a nationwide or statewide scale.

Such evaluation which integrates top-down with bottom-up guidelines is able to generate knowledge for practice, in other words guidelines for sharing, transforming, and constructing knowledge of a general curricular value that can also be of value in a variety of different situations. Naturally, though, this sort of evaluation process is not at all simple due both to its dependence on consensual decision making involving diverse stakeholders, internal and external, and its reliance on multiple evaluation modes, some of which may also conflict with the process's philosophical and practical principles.

Influenced by Stake's (1975) responsive evaluation model, Eisner's (1998) connoisseurship model, and Parlett and Hamilton's (1972) illuminative model, this orientation corresponds, though not identically, with Aoki's situational evaluation orientation; MacDonald (1974) autocratic approach; Melrose's (1998) transactional paradigm of evaluation, and with what Lather (1991) refers to as the critical tale. These models are all based on conceptions involving liberal humanism, subjectivist ethics, and communal activities: the use of mutual negotiation, communication, and partnership, aimed at and achieving consensual interpretation.

Emergent and Emancipatory Evaluation

The third orientation to curriculum evaluation entails a continuous knowledge co-construction process in individual contexts and through social collaboration, dialogical-negotiated, and authentic experience regarding meanings, actions, and values. It is also a reflective meaning-making process that is situated, systemic, and interwoven with curriculum evolution and contributes to ongoing learning and change. The approach is considered emergent in that it is not pre-planned but evolves together with curriculum

evolution through dialogue, collaboration, and praxis. It therefore represents a culture, which turns practice into ideology and knowledge, while emphasizing authenticity and emergentism. It is considered emancipatory in its combination of inquiry and action with critical insight into the social construction of human society. It expresses a concern for the moral and ethical dimensions underlying human action by asking what sorts of experiences can help lead the people involved toward a life influenced by equity, caring, and compassion. It therefore also demonstrates sensitivity to factors such as gender, culture, ability, and sociogeographic context. It identifies and exposes that which is oppressive and authoritarian by sensitively eliminating false consciousness and unjust values. It accomplishes this by considering historical, political, and social dimensions and factors relating to goals, contents, and classroom processes in an effort to change the curricular context, and by monitoring the results of the attempted changes.

Based on the social and radical constructivist and participatory paradigms, and grounded in the epistemological belief that all knowledge is socially, culturally, and geographically situated, and that truth and knowledge are created (Schwandt, 1997), this orientation holds that in order to understand the world, human beings use socially constructed dialectic strategies, which are shared through systems of language and symbols and adapted to meet the needs and intents of human activity. This orientation also believes that a system can only be understood when viewed within its wider contexts and that because each curriculum is unique due to the context of its use, it is important to know and evaluate the specific and individual biography of each curriculum. This assumption reflects Stenhouse's (1975) conditionality criteria, which emphasizes the influence of the context on the curriculum. This orientation also holds that there is plenty of room for every stakeholder: students, teachers, parents, superintendents, evaluators, and others, and therefore appreciates every voice. It expresses openness to learning at any time and from any source, and appreciates the naturalness of chaos, implying that everything is relative and therefore not always expected.

This mindset conceives curriculum evaluation as a dynamic progression that stimulates action and evolves over a long period hand-in-hand with evolutionary curriculum planning, based on the negotiated needs, goals, beliefs, and knowledge of all those involved. It proceeds through mindful and sensitive negotiation between stakeholders. In theory, the process can continue indefinitely, uniting external and internal, individual and social, process and product, not as separate entities, but as elements of mutually constituted social-educational-cultural and virtuous activities. Thus, emergent and emancipatory curriculum evaluation focuses on the whole process and the whole persons involved, students and teachers alike.

It is therefore complex, dynamic, and holistic, as well as open, flexible, context-contingent, time-dependent, and reality-based.

Emergent and emancipatory curriculum evaluation is a highly participatory mode of continuous inquiry that not only fosters a dynamic mode of curriculum planning, but also helps those more directly involved in planning the curriculum, for example, teachers and students, to achieve their individual and common goals. So, evaluators are part of a design team, functioning as an action research or learning group and helping to monitor classroom practices and outcomes, all within an evolving and rapidly changing environment of constant feedback, reflection, and change. The evaluator functions as guide, monitor, facilitator, and a critical learner. Teachers, students, and stakeholders play key roles in implementing and regulating the process (learning and evaluation). All are both subjects of evaluation and evaluators. The cardinal concepts of this evaluation approach are therefore: collaboration, inter-subjectivity, context-specificity, diversity, meaning, construction, reciprocity, pluralism, unpredictability, and participants/stakeholders. This contrasts markedly with the nature of the discourse surrounding the orientations above, which reflect modern values like absolutism, stability, objectivity, certainty, prediction, and authoritative or consensual decision making.

This orientation in curriculum evaluation does not value the traditional mores of excellence such as input standardization, consistency of intervention or implementation processes, outcome uniformity, and clarity or linearity of causal linkages. Instead it assumes a world of situated and multiple needs and causes, diversity of outcomes, inconsistency of interventions, and interactive effects at every level. It examines the individual student, teacher, classroom, school, district, and nation, and is therefore the goal of a pluralistic society. This orientation therefore offers a diversified conception of quality in which participants suggest the evaluation standards and criteria depending on the educational contexts, settings, views, and needs of their environment. Through dialogue among diverse stakeholders curriculum quality is defined as the extent to which the goals of all stakeholders are fulfilled or will be fulfilled in future, taking into account that not all goals can be simultaneously fulfilled or agreed upon. Setting evaluation standards and criteria is therefore a bottom-up adaptive process that considers not only past and present contextual factors, but also future needs and consequences. Unlike the adaptive and situated evaluation orientation, no stakeholders' consensus is required and divergent voices are appreciated. There is enough space for every voice, openness to learning at any time and from any source, and an appreciation for the naturalness of chaos. Everything is relative and therefore not always predictable, making this type of evaluation not easy to implement and often complicated to interpret.

This is due to the lack of absolute or contextual standards, to the fact that different stakeholders' views must be considered simultaneously, and to the demand for sensitivity to multiple factors such as gender, culture, ability, and sociogeographic context, which must be considered all together without the availability of compatible implementation processes.

The evaluation process is actually an internal, context-dependent, development and evaluation process. Thus, for example, curriculum evaluation might engage teachers, students, school-based curricular experts, parents, and any relevant stakeholder that the team co-opts. It progresses hand-in-hand with curriculum planning and provides a means of acquiring evaluation knowledge-in-action, of transforming or taking responsibility for this knowledge, and of producing and reproducing the curriculum. In one context students or teachers may decide on a particular curriculum path, discussing it with and convincing meaningful others who might be simply the students' teacher or teachers' colleagues that this is a worthwhile route to take. As the evaluation process unfurls additional stakeholders may follow up the development and learning process and engage in short- as well as long-term reflections on the process, its learning development, and outcomes. Each representative of the stakeholders considers the evaluation process from his/her own unique perspective. It is a process, which describes what can be done in the learning situation, what is chosen to be done, and what conditions enable it to be done. Finally, the process is likely to produce insights into the specific value of the chosen curricular path within the larger context of the evolving curriculum. The evaluation ideas, issues, and suggestions offered by the stakeholders at different points in time are thus subject to constant reconsideration and reexamination in a dialogic process embracing all relevant parties, viewpoints, and interests. In this scenario, the professional evaluator functions as a guide, monitor, facilitator, and learner, and not as a powerful decision maker. His or her standpoint carries no more weight than any of the other stakeholders.

This evaluation orientation is probably best suited to a curriculum which is considered an agenda for social reconstruction or an emergent curriculum. In terms of curriculum evaluation, the emergent and emancipatory orientation corresponds to Aoki's situational, interpretive, and critical approach; Melrose's (1998) critical paradigm of evaluation; with what Lather (1991) refers to as the deconstructionist tale, with MacDonald's (1974) democratic evaluation and with Kemmis and Hughes' (1979) conception of evaluation as self-reflection in a critical community. Thus, emergent and emancipatory curriculum evaluation is a communicative, dynamic, and holistic paradigm, and its vision is one of evaluation as both a guiding and learning process.

In sum, these evaluation orientations apply to curriculum in all subjects, as well as to interdisciplinary and multidisciplinary curricula, computer-based curricula, and to all grade levels. While the three evaluation mindsets are not entirely mutually exclusive, since some characteristics of each paradigm are present at least partially in one of the other two, they do differ in their primary purposes and uses, methodologies, and in what each considers an evaluation priority.

Purposes and Uses of Curriculum Evaluation

There are three different purposes or uses of curriculum evaluation: (1) curriculum improvement, commonly linked to formative evaluation; (2) judgment purposes, which is commonly associated with summative evaluations; and (3) knowledge generation, which involves curriculum-oriented knowledge building (Patton, 1998), commonly related to developmental evaluation, and sometimes referred to as conceptual use (Shadish *et al.*, 1991). Each has its own focus: formative evaluation emphasizes on discovering how to improve the curriculum, whereas summative evaluation focuses on appraising outcomes and deciding whether to continue or discontinue a curriculum. Developmental evaluation, on the other hand, seeks to influence thinking, deepen understanding, and generates learning by increasing knowledge of what constitutes a meaningful curriculum, based on evaluation wisdom. Thus, while formative evaluation carries a bias about making the curriculum better rather than just making it different, developmental evaluation conceives of changes in the curriculum not necessarily as an indication of progress or improvement but rather as an adaptation.

These three uses of curriculum evaluation have different meanings in the context of the different evaluation orientations. For example, the differentiation between formative and summative evaluation that holds true in the instrumental orientation does not apply in the emergent and emancipatory orientation. In the former, formative evaluation is a phase in the evaluation process, which is carried out during the early stages of curriculum development seeking to invite improvement (Lewy, 1990). Summative evaluation, for its part, focuses on determining whether a curriculum is efficient and answers identified needs. However, in the context of emerging and emancipatory evaluation this distinction between formative and summative is no longer relevant since the curriculum evaluation constantly evolves synergistically with the curriculum planning process (Levin, 2002) and therefore the more appropriate evaluation use is that of developmental evaluation, a process in which the evaluator becomes part of the development team helping to monitor

what is happening, both processes and outcomes, in an evolving, rapidly changing environment of constant feedback and change (Patton, 1994).

Multiple Objects and Multiple Representations in Curriculum Evaluation

Any discussion of curriculum evaluation must specify the objects of evaluation and the evaluative information needed. Due to the systemic nature of a curriculum, we expect curriculum evaluation to address both the trees and the forest: namely, its components, their interrelationships, and the holistic structure. In the past, mostly under the influence of Scriven's (1994) belief that in order to establish the merit of all or part of a curriculum there is no need to know how it works or why it fails to work, curriculum evaluation mainly focused on results and outcomes. Later, influenced by other evaluation models, it was argued that curriculum evaluation should reflect the fullness, complexity, and importance of a curriculum (Stake, 1980, 2004; Alkin, 1994; Stufflebeam, 1983). That is, curriculum evaluation should focus on antecedent conditions and contextual factors relevant to the curriculum act, including curricular policy and goals; classroom transactions; goals (intended and unintended), and outcomes, planned or not. It is also accepted that the objects of evaluation would include the processes that lead to the decisions taken by the participants. Furthermore, the more we see the curriculum as a complex system, rather than a linear and well-defined sequence of tasks, the more we need not only to evaluate the different evaluation objects but also to explore their interrelationships. Although this is true for all three orientations, the evaluation orientation will determine the boundaries and specifications of the evaluation components, and the depth and breadth of the evaluative process.

When we evaluate, what curriculum are we evaluating? Is it the intended curriculum, the planned curriculum, the acted curriculum, or the achieved curriculum (Goodlad *et al.*, 1979). Every layer of the curriculum has its own evaluative value for describing, judging, understanding, and explaining a curriculum's evolution, practices, and outcomes. In particular, the relationships between the various curricular representations are worthwhile exploring. For example, the relationships between the acted and achieved curriculum can offer some explanations of the results achieved, in both cases, whether the results are satisfactory or not. The relationships between the planned and acted curriculum can also provide insights into the potential possibilities for implementing a particular curriculum. However, the meaning attached to these relationships depends on the evaluation orientation. For example, evaluators employing the instrumental orientation seek

alignment and regard discrepancies between the intended and the acted curriculum as a failure in interpretation or implementation. In contrast, one expects in situated and adaptive evaluation to reveal discrepancies between the different representations and be beneficial for enhancing understanding of a situated curriculum and its alignment with a set of specific curricular needs.

Influential Levels and Methods of Curriculum Evaluation

Curriculum evaluation is influential on four different levels: societal or macro (system, nation, or state), institutional or meso (school), instructional (classroom) or micro (classroom), and individual or nano. In each case, the evaluation questions differ, as do the purposes and methodologies used. Typically, on the macro and meso levels, evaluation questions are oriented toward curricular policy, accountability and judgmental issues, aiming to construct general knowledge in the form of best practices and prescriptive models. On the micro and nano levels in contrast, curriculum evaluation often concentrates mainly on the unique characteristics of the curriculum in specific contexts, and relates to curricular needs, processes, and outcomes, as well as the webs of its contextual relationships in order to generate contextual knowledge in the form of lessons learned and principles.

Methodologically, when embedded in a positivist worldview, mostly within the instrumental orientation, curriculum evaluation applies methods and techniques to obtain objective and generalizable facts. Mandating a rigorous separation of facts and values and aiming to generalize and predict, the inquiry process adopts a clinical distance, which we usually find in the more traditional research approaches, often applying the same methods as in controlled life science experiments. The main procedures are asking questions, developing a study design, defining variables, developing instruments, collecting data as objectively as possible, analyzing data, and arriving at conclusions or recommendations through a process of synthesis. In these processes, there is substantial emphasis on achieving precision, reliability, and validity. The typical inquiry methods focus mainly on quantifiable indicators of success and failure. Within the adaptive situated orientation, some methods also incorporate qualitative measures, particularly in the descriptive part of the evaluation, and case analysis is also likely to be performed (Stake, 1995).

In terms of the emerging and emancipatory orientation, however, since the evaluation process is designed to learn from events as they occur and to help promote and accelerate change, the methodological mindset accepts the fact that prior to an investigation one cannot know all it will entail or uncover. The methods therefore

become part of a continuous interplay with new emerging data as the data collection and analysis progresses. The key processes of the evaluative inquiry in this context are: asking questions; identifying and challenging values, assumptions, and beliefs; reflection; dialogue; collecting, analyzing, and interpreting data; action planning and implementation (Preskill and Preskill, 1997). Qualitative and quantitative measurement can both be applied depending on the specific issues investigated (Stake, 2004). The data is not objective but understood in terms that invoke the interpreters' values. Informal logic thus guides the evaluation methodology within the emerging and emancipatory culture, countering formal logic within the inductive–deductive framework of the instrumental orientation to curriculum evaluation (Scriven, 1994).

Summary

As previously held beliefs and certainties blur into post-modern complexities, many of the tried and tested assumptions regarding curriculum evaluation seem unsatisfactory. Modern conceptions of curriculum evaluation encourage us to value stability, regularity, certainty, and universality. They prompt us to apply guidelines and standards and to assume that curriculum practices are predictable and carefully delineated. Nowadays, influenced by constructivist and complexity theories, and challenged by the globalization demands of new capabilities and proficiencies in cultures marked by uncertainty and pluralism, curriculum evaluation is viewed as a perpetual, dynamic, context-specific, collaborative, meaning-making developmental process. It is flexible, yet methodical, open, yet directive, sensitive to ethical dimensions underlying human action, and respectful of the diverse, complex curricular visions, needs, and constraints encountered in society, schools, and classrooms. It embodies new categories of relationships between people/stakeholders all grounded in an on-going dialogue that is open to changing events. It also establishes different meanings of quality and accountability and shifts the emphasis away from an external value system of fitting to standards, to refocus it on respecting the idea that a curriculum's intrinsic value is its meaning, and that accountability should relate to values and commitments rather than to external authoritative control. Thus, curriculum evaluation takes on different forms and different trajectories of meaning in different contexts and situations. There is no evaluation model to follow or implement, no standardized procedures or explicit guidelines. Instead, we find ideas and principles that provoke and challenge our thinking, beliefs, and routines in the dual fields – curriculum and evaluation.

See also: Curriculum Evaluation; Formative Assessment in Teacher Education and Teacher Professional Development.

Bibliography

- Alkin, M. C. (1994). Curriculum evaluation models. In Husen, T. and Postlethwaite, T. N (eds.) *The International Encyclopedia of Education*, 2nd edn., vol. 3, pp 1279–1283. Oxford: Pergamon Press.
- Aoki, T. T. (1978). *Toward New Conceptions for Evaluation in Social Studies. Curriculum Evaluation in a New Key (Monograph Series, No. 1)*. Vancouver: University of British Columbia.
- Eisner, E. W. (1998). *The Enlightened Eye. Qualitative Inquiry and the Enhancement of Educational Practice*. Upper Saddle River, NJ: Prentice Hall.
- Goodlad, J. I., Klein, M. E., and Tye, K. (1979). The domains of curriculum and their study. In Goodlad, J. I. and Associates (eds.) *Curriculum Inquiry: The Study of Curriculum Practice*, pp 43–76. New York: McGraw-Hill.
- Guba, E. G. and Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Newbury Park, CA: Sage.
- House, E. R. (1993). *Professional Evaluation*. Newbury Park, CA: Sage.
- Jasparro, R. J. (1998). Applying systems thinking to curriculum evaluation. *National Association of Secondary School Principals. NASSP Bulletin* 82(598), 80–86.
- Kemmis, S. and Hughes, C. (1979). Curriculum evaluation in higher education: Self-reflection in a critical community. *Paper in the Proceedings of the Fifth Annual Conference of the Higher Education Research and Development Society of Australasia (HERDSA)*. Brisbane, Australia: HERDSA.
- Lather, P. (1991). *Getting Smart: Feminist Research and Pedagogy with/in the Postmodern*. New York, NY: Routledge.
- Leathwood, C. and Phillips, D. (2000). Developing curriculum evaluation research in higher education: Process, politics and practicalities. *Higher Education* 40, 313–330.
- Levin, T. (2002). Stability and change in curriculum evaluation. *Studies in Educational Evaluation* 28, 1–33.
- Lewy, A. (1990). Formative and summative evaluation. In Walberg, H. J. and Hurdle, G. D. (eds.) *The International Encyclopedia of Educational Evaluation*, pp 26–28. Oxford: Pergamon.
- MacDonald, B. (1974). Evaluation and the control of education. Reprinted in: Murphy, R. and Torrance, H. (eds.) (1987). *Evaluating Education: Issues and Methods*. London: Harper and Row.
- Melrose, M. (1998). Exploring paradigms of curriculum evaluation and concepts of quality. *Quality in Higher Education* 4(1), 37–43.
- Norris, N. (1998). Curriculum evaluation revisited. *Cambridge Journal of Education* 28(2), 207–219.
- Parlett, M. R. and Hamilton, D. (1972). *Evaluation as Illumination, a New Approach to the Study of Curriculum Projects*. Occasional Paper, No. 9, Center of Research in Educational Sciences, Edinburgh University.
- Parlett, M. and Hamilton, D. (1976). Evaluation as illumination: A new approach to the study of innovative programs. In Glass, G. V. (ed.) *Evaluation Studies Review Annual*, vol. 1, pp 140–157. Beverly Hills, CA: Sage.
- Patton, M. Q. (1986). *Utilization-Focused Evaluation*. London: Sage.
- Patton, M. Q. (1994). Developmental evaluation. *Evaluation Practice* 15, 311–319.
- Patton, M. Q. (1998). The challenges of diversity in evaluation: Narrow versus expansive perspectives. *Science Communication* 20(1), 148–164.
- Popham, W. J. (1988). *Educational Evaluation*. Englewood Cliffs, NJ: Prentice Hall.
- Preskill, S. L. and Preskill, H. (1997). Meeting the postmodern challenge: Pragmatism and evaluative inquiry for organizational learning. In Mabry, L. (ed.) *Advances in Postmodern Evaluation: Evaluation and the Postmodern Dilemma*, vol. 3, pp 155–169. London, UK: JAI.
- Schubert, W. H. (1996). Perspectives on four curriculum traditions. *Educational Horizons* 74(4), 169–176.
- Schwandt, T. A. (1997). Reading the 'problem of evaluation' in social inquiry. *Qualitative Inquiry* 3(1), 4–25.
- Scriven, M. (1994). The final synthesis. *Evaluation Practice* 15, 367–382.
- Shadish, W. R., Cook, T. D., and Leviton, L. C. (1991). *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, CA: Sage.

- Stake, R. E. (1975). *Evaluating the Arts in Education: A Responsive Approach*. Columbus, OH: Merrill.
- Stake, R. E. (1980). Program evaluation, particularly responsive evaluation. In Dockrell, W. B. and Hamilton, D. (eds.) *Rethinking Educational Research*, pp 72–87. London: Hodder and Stoughton.
- Stake, R. E. (1995). *The Art of Case Study Research*. London: Sage.
- Stake, R. E. (2004). *Standard-Based and Responsive Evaluation*. London: Sage.
- Stenhouse, L. (1975). *An Introduction to Curriculum Research and Development*. London: Heineman Educational.
- Stufflebeam, D. L. (1983). The CIPP model for program evaluation. In Madaus, G. F., Scriven, M., and Stufflebeam, D. L. (eds.) *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, pp 117–141. Boston, Massachusetts: Kluwer-Nijhoff.
- Tyler, R. W. (1949). *Basic Principles of Curriculum and Instruction*. Chicago, IL: University of Chicago Press.
- Research Review. <http://www.nfer.ac.uk/publications/pdfs/downloadable/NCAe-report.pdf> (accessed August 2009).
- Mabry, L. (1997). Implicit and explicit advocacy in postmodern evaluation. In Mabry, L. (ed.) *Advances in Postmodern Evaluation: Evaluation and the Postmodern Dilemma*, vol. 3, pp 191–203. London: JAI.
- Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods*. Sage: London.
- Scriven, M. (1980). *The Logic of Evaluation*. Inverness, CA: Edge Press.
- Scriven, M. (1983). Evaluation ideologies. In Madaus, G. F., Scriven, M., and Stufflebeam, D. L. (eds.) *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, pp 229–260. Boston, MA: Kluwer-Nijhoff.
- Scriven, M. (1999). The fine line between evaluation and explanation. *Research on Social Work Practice* 9(4), 521–524.
- Stake, R. E. (1998). Hoax. In Davis, R. (ed.) *Proceedings of the Stake Symposium on Educational Evaluation*, pp 363–374. Champaign, IL: University of Illinois.
- Stevens, F., Frances, F., and Sharp, L. (1997). *User-Friendly Handbook for Project Evaluation in Science, Mathematics, Engineering and Technology*. New York: NSF.
- Worthen, B. R. and Sanders, J. R. (1987). *Educational Evaluation: Alternative Approaches and Practical Guidelines*. New York: Longman.
- Abma, T. A. (1997). Sharing power: Facing ambiguity. In Mabry, L. (ed.) *Evaluation and the Postmodern Dilemma. Advances in Program Evaluation*, vol. 3, pp 105–119. Greenwich, CT: JAI.
- Chelimsky, E. (1997). Thoughts for a new evaluation society. *Evaluation* 3(1), 97–118.
- Fetterman, D. (1994). Empowerment evaluation. American Evaluation Association Presidential Address. *Evaluation Practice* 15(1), 1–15.
- Greene, J. C. (2005). The generative potential of mixed methods inquiry. *International Journal of Research and Method in Education* 28(2), 207–211.
- Greene, J. C. and Caracelli, V. J. (1997). *Advances in Mixed-Method Evaluation: The Challenges and Benefits of Integrating Diverse Paradigms. New Directions for Evaluation*, vol. 74. San Francisco, CA: Jossey-Bass.
- Hartman, R. S. (1967). *The Structure of Value: Foundations of scientific Axiology*. Carbondale, IL: Southern Illinois University Press.
- Iannone, R. (1995). Chaos theory and its implications for curriculum and teaching. *Education* 115(4), 541–547.
- Kirkpatrick, D. L. (1967). Evaluation of training. In Craig, R. L. and Bittel, L. R. (eds.) *Training and Development Handbook*, pp 87–112. New York: McGraw-Hill.
- Lord, P. and Jones, M. (2006). Pupils' experiences and perspectives of the national curriculum and assessment. *Final Report for the*

Further Reading

Relevant Websites

- <http://www.project2061.org> – AAAS Conference on Developing Textbooks That Promote Science Literacy, Project 2061, American Association for the Advancement of Science.
- <http://www.thirteen.org> – Assessment, Evaluation, and Curriculum Redesign, Thirteen.
- <http://www.sde.ct.gov> – CT Curriculum Development Guide, Connecticut State Department of Education.
- <http://www.med.ufl.edu> – Curriculum Evaluation Plan (Exhibit 6), College of Medicine – University of Florida.
- <http://www.cast.org> – Curriculum-Based Evaluations, CAST: Center for Applied Special Technology.
- <http://www.uwex.edu> – Evaluation Curriculum Overview, UW-Extension.
- <http://www.aecom.yu.edu> – Evaluation Policies and Procedures, Einstein: Albert Einstein College of Medicine.
- <http://www.vtvsba.org> – Vermont School Boards Association.

Evaluating Education in Three Policy Eras

M A Peters and R Britez, University of Illinois at Urbana-Champaign, Champaign, IL, USA
S M Weber, Marburg University, Marburg, Germany

© 2010 Elsevier Ltd. All rights reserved.

Introduction: Education Policy and Policy Evaluation

In this article, we demonstrate that the question of how to analyze the relationship between educational policy and policy evaluation leads directly into the heart of theoretical positions in policy science and into historical reconstructions of knowledge and rationalities underlying educational policy and evaluation. We show that evaluation is not a neutral knowledge, just a methodological and technical knowledge for betterment in general, “designed and conducted to assist some audience to judge and improve the worth of some educational object” (Stufflebeam and Webster, 1980: 6). The word value directly raises the question of whose value? and immediately steps into the political dimension. Carol Weiss (1988a, 1988b) an early voice in the field of evaluation argued for critical perspectives, taking the context conditions and the political dimension into account.

Over the past 50 years, criteria and the question of interests, multiperspectivity, and the political dimension has increasingly come in our view (House, 1990: 24). This article shows that according to different rationalities of governing the state and the present, the knowledge base, methodology, conceptualizations and use of evaluation differ. We demonstrate this, analyzing three relationship patterns between concepts of governing the state and evaluation knowledge. Three political eras are differentiated: the welfare-state model, which is oriented toward equality and security of its citizens; the neoliberalism model of the state, where the notion of freedom implies a loss of state-based securing systems and rising levels of individual risks; and knowledge capitalism, where new mixes and practices of knowledge are to be identified. Seen from a Foucauldian perspective (Foucault, 1992, 2004), evaluation has to be analyzed as power knowledge (Bröckling, 2004; Höhne, 2006), creating specific knowledge orders, knowledge architectures, and structures of attention. In a governmentality perspective (Peters *et al.*, 2009), knowledge is analyzed as discursive, social, and political practice – taking the activity of the state and the rationality of steering and governing into account. Following this perspective, we analyze evaluation as power/knowledge within three specific policy eras of the welfare state, neoliberalism, and knowledge capitalism. It becomes clear that within three political eras, different relationships to truth, to power, to practice, and

to experience are observed in the way in which evaluation is being seen and used (see **Table 1** for summary).

The first pattern shows the relationship between education policy and policy evaluation as spheres of autonomy. Evaluation is, mostly, regarded as neutral scientific knowledge, supplying needed information and findings to society in general. In a delivery relationship, policy is donor to evaluation and customer of findings. In this relationship pattern, evaluation transfers scientific expert knowledge to society, supplying policy with relevant information. The concept of neutral science supplying society with relevant data for means of legitimation is inherent in the guiding image of the welfare state.

In a second pattern of relationship between educational policy and policy evaluation, conditions and interdependence become crucial: In this perspective, policy is to be seen as a condition and framework for evaluation activities. The function of evaluation activity changes into an active measurement function. The relationship between evaluation and policy analysis becomes closer and more intense; evaluation knowledge shifts from legitimation to control and measurement. Relevant information and findings are not primarily oriented toward the past (legitimation), but directed toward a steering function.

The third relationship pattern between educational policy and policy evaluation is to be drawn as overlapping rationality: policy rationalities shape the modeling of evaluation approaches and empirical research and in a reflexive view become visible as an epistemological relationship of power/knowledge. Evaluation is becoming a mix of visibility and learning, of quantitative ranking, and of creative and future-oriented knowledge creation.

Table 1 indicates the rationality shifts of the three eras in which evaluation practices are not necessarily empirically replacing each other totally, but overlap and coexist, and change their position, importance, and acceptance within predominant patterns of evaluation regimes.

Era of Keynesian Welfare State

The beginning of the political era of the welfare state differs according to the specific historical national backgrounds, but in a very broad perspective can be dated to the end of the nineteenth and the beginning

Table 1 The evaluation of education in three policy eras

	<i>Welfare-state era of policy evaluation</i>	<i>Neoliberal era of policy evaluation</i>	<i>Knowledge-economy era of policy evaluation</i>
Main levels of action of evaluation programs	Local National	National International	Local National Regional Global
Predominant epistemology of evaluation programs	Objectivist technocratic reproduction theory	Objectivist marginalized social-constructivist approach	Social-constructivist approach
Predominant methodology of evaluation programs	Description and measurement of input and effects	Accountability and large-scale assessment	Hermeneutic inquiry and collective truth production by inquiry and negotiation of truth and results
Predominant orientation of evaluators	Research/knowledge production	Performance measurement	Learning/collaborative knowledge production
Main objectives of policy evaluation	Legitimation Research	Legitimation Control Allocation of resources	Legitimation Policy learning Policy change

and throughout the twentieth century. The political model of the welfare state is oriented toward prevention or solution of poverty, social security, and aims at diminishing social inequalities by money transfers as well as social services, health services, labor politics, legal regulations, and restrictions. This approach goes along with the concept of a state-steered economy, developed by John Maynard Keynes. Esping-Anderson (1990) differentiates three types or regimes of welfare states: liberal regimes (USA, Canada, Switzerland, and Japan), conservative regimes (France, Italy, Germany, and Austria), and social democrat regimes (Sweden, Denmark, Norway, and Finland). In this classical categorization, the relationships between state and market differ with regard to the transfers of social services, the mode and quality of services, and the effects of social policy on social stratification and power in society. The welfare state in its ideal type should offer social security to every citizen. Welfare states are securing societies establishing complex strategies to care for their citizens, formal education of citizens being part of those strategies.

In a Foucauldian notion we can see that the steering rationality of the welfare state has been based on hegemonial disciplinary knowledge (Peters *et al.*, 2009), like, for example, psychological knowledge of testing and measurement. This disciplinary knowledge produces specific relationships of the subject of the diagnosis and diagnostics. Survey rationality became the prominent movement of the 1920s and 1930s. It grew along with the rise of psychology as a discipline and it was based within a specific concept of science as natural science, where experimental and causal logic rationality is promoted. Linear thinking, test theoretical approaches, and the notion of measurement were institutionalized in education – for example, in the Cincinnati public schools as early

as 1929 (Felix, 1979). Evaluation as measurement was proposed by Tyler (1950), viewing educational objectives as changes in behavior and following the concepts of behaviorism.

The rise of the welfare state grew along with the rationality of planning and measurement, and in the discourse of human capital, education was regarded as a matter of investment and predictable prosperity. Education was seen as both a capital and a consumer good (Kogan, 1979: 20), as functional investment for social equality and economic growth. In this phase of optimism and rationalism, educational policymaking followed an expansionist idea (Kogan, 1975). For example, US investment in education grew around 20 times and investment in social politics like housing in 1980 was 129 times higher than in 1950, while the United States population increased only by half. Given the extraordinary input in the phase of 1950–1980, Murray (1994: 14) calls it a generous revolution. This phase of massive welfare states' financial investment was true for other Organization for economic Cooperation and Development (OECD) areas as well. OECD, World Bank, and the United Nations Educational, Scientific and Cultural Organization (UNESCO) supported the world expansion of education. In OECD reviews, remarkable parallelisms of national development are observed, although countries had adopted various policy/planning approaches and operational strategies (Kogan, 1979:16).

In general, policy at that time was driven by a state approach, where political strategies were defined and driven by the state. Political strategies were input- and top-down oriented. At that point in time, strategies were not developed with the participation of citizens in a bottom-up mode. Peoples' acceptance of top-down strategies and their lack of participation in education were not

questioned. In this political climate of steering optimism, science and scientific methods for program planning and evaluation arose. Within the so-called war on poverty, large state programs were set up. In order to bring about large-scale effects, social action programs in social security, health and education (Rogers-Dillon, 2004: 8) like Elementary Secondary Education Act (ESEA), Headstart, or Follow Through were established. At that point in time, few OECD background reports linked their statements of goals to an evaluation of how they were implemented (Kogan, 1979: 27). At the national level in the US, legitimation for welfare and education programs was a need. Political strategies therefore needed evaluation. Following the rationalist model of planning, predominant evaluation concepts directly referred to the – already established – scientific-measurement movement with its attendant objectivity and reliability (Jemelka and Borich, 1979: 264). Assuming measurability of learning and change, until the 1970s, evaluation rationality followed a mechanist and objectivist input-output model.

Under the shock of economic recession, the political and economical climate shifted into uncertainty, pessimism (Kogan, 1979: 19), and into a second generation of evaluation (Ravitch, 1983), and bottom-up multiperspectivity and qualitative approaches became more acceptable. Legitimation-oriented evaluation ceased to be seen as producing the information required to alter either program or policy. As Ginsburg *et al.* (1992: 24) show, criticisms of large program evaluation are related to four points:

1. Evaluations showed a preoccupation with measuring overall program impacts, particularly test score changes.
2. Evaluations focused almost entirely on federally funded services and failed to recognize that federal programs are part of a larger service-delivery system.
3. Evaluation studies were frequently funded as single large-scale studies. Study findings were not well integrated into decision making. A common complaint from congressional staff was that studies were completed out of phase with the legislative cycle.
4. Evaluations were criticized as designed and operated without any involvement from staff in the program office. Therefore, in the 1970s, programs were asked to open up more to multiperspectivity. Comprehensive change started to be regarded as necessary (Ravitch, 1983: 258).

Disciplinary dominance of psychology and professional dominance of academia's program evaluators (Concliff-Lagemann, 1997) began to be increasingly questioned. Scientists lost their status of being assumed as value-neutral experts. Between quantitative and qualitative advocates, battles were hard fought (Boruch and Riecken, 1975). In the expanding and professionalizing field

of evaluation, a discipline and profession of evaluation emerged. Although the evaluation market was still retained by powerful stakeholders, evaluation became publicly politicized, and the voice of qualitative approaches became louder, criticizing the troubles of this troubled crusade (Ravitch, 1983). Major differences in the field of evaluation were seen between qualitative versus quantitative methodology, accountability versus policy orientation, and client participation versus nonparticipation (House, 1990: 26). Even when representing stakeholders' views in evaluation became more common, the critical political question still remained – whose interests did the evaluation serve (Guba and Lincoln, 1982)? The political question of evaluation policy now became a topic to address (Braybrooke and Lindblom, 1970).

Due to the cuts in the social programs, 1980s was a quiet decade. Neutrality and objectivity became political priorities again. The formulated purpose of centralized accountability was renewed by hiring international evaluation staff and establishing standardized achievement testing. Evaluation knowledge shifted from legitimation of the past to controlling of the present. Testing and measurement once more showed to be instruments of discipline than of diagnosis (House, 1990: 24). International comparison studies like the Program for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS) contributed to establish a European space of modularized knowledge. Within the political climate of the activating state, transnationalization in educational policy accelerated. Within a transnational framework, pre-given political parameters and standards applied pressure on national states. The OECD and the World Bank emphasized the policy of neoliberal modernization from above and evaluation became an effect of globalization. Following House (House 1990: 25), "what had begun as an era of social consensus dissolved into an age of conflict and diversity." It was seen that policy no longer orientated toward inclusion of marginalized groups. The function of evaluation changed toward a medium of exclusion and selection (Höhne, 2006).

Neoliberal Paradigm

As we have mentioned in the introduction, we explore the three eras of policy evaluation welfare, neoliberal, and knowledge economy. In this article, we follow Pollit's example of not ascribing hard edges to any of those broad periodizations. We can see that the origins of policy evaluation as a systematic and formalized practice of government activity can be traced to the early 1960s. Fischer (1995) indicates that the development of policy evaluation today is associated with the expansion of many of the governmental programs of this period. During this time,

the evaluation of policies became a formalized practice of public agencies, as policy-analysis studies increasingly influenced and informed the decision of policymakers and public-policy debates. The relevance that social research and policy analysis studies acquire in the public debate about education can be observed in the publication of the 1966 report *Equality of educational opportunity*, also known as the Coleman report (see Coleman *et al.*, 1966). In the United States, Lyndon Johnson's Great Society and War-on-Poverty programs, as well as the expansion of other government bureaucracies, were driven by the common belief of policymakers that the "decision-making process could be effectively rationalized," through rigorous analysis enabling access to information capable of improving the decisions of public organizations (Fischer, 1995: 4).

The history of policy evaluation as a formalized government practice, informed by different philosophies of governance, has followed different trajectories across countries and agencies in different periods. However, the rationale for the implementation of public-policy-evaluation practices follows the common assumption that rigorous rational-policy analysis could improve the effectiveness of public organizations. In those terms, policy analysts then followed a technocratic approach to evaluation.

The advantages of the development and use of policy-analysis capabilities soon became clear, and policy evaluation rapidly became a common mandate for public agencies. For instance, in the 1970s, the US Congress legislated legally mandate program evaluations (Fischer, 1995: 5) for public agencies. It is during this period of expansion that we observe the transition toward the neoliberal era in policy evaluation.

The economic crisis of the 1970s not only signaled a profound crisis in the model of the welfare state and the ascension of its critics, but also triggered a criticism of the rationale of types of policy analysis associated with the implementation of welfare-state programs. Associated with this critique was a desire by opponents of the welfare-state model to create alternative uses for the tools of policy analysis. In the United States and United Kingdom, critics of the welfare-state government programs such as the Great Society began to develop and fund policy research and institutes suited to their own political needs (Fischer, 1995: 5) (e.g., in the US, the Heritage Foundation, American Enterprise Institute (Fischer, 1995) and in the UK, the Centre for Policy Studies and Adam Smith Foundation (Pollitt, 1993)). In short, the same policy-analysis tools used for implementing and coordinating the 1960s' government programs were used for eliminating them in the 1980s.

The victories of Margaret Thatcher in 1979 and a year later the ascension of Ronald Reagan to the US presidency signaled 1980s as the opening of the neoliberal era of

policy evaluation. This was an era of evaluation dominated by the following philosophy of governance:

Government had taken on more than it could handle . . . decentralized decisions of the market-place should whenever possible replace the inevitably inadequate plans of central or local government (Pollitt, 1993: 12).

Following this philosophy, policy evaluation was to be concerned with measuring the effectiveness, in economic terms, of policy programs. In other words, the interest of the preceding era to identify effectiveness and impact was replaced with a focus on evaluating the efficiency and economy (Pollitt, 1993: 13) of policies. We would like to mention three main features of this evaluation era:

- First, there was a new concern with the management of the allocation of public resources, which in the case of education, serves to justify cuts in the direct financing of public programs, as well as to link educational outcomes to the requirements of the economy. Furthermore, this concern for the control of the allocation of resources affects not only the methods but also the purposes of evaluation. Evaluators increasingly became auditors rather than analysts, with a methodological emphasis on the measurement of the economic performance of the policies evaluated.
- Second, the new public management (NPM) school emerged as one of the most relevant evaluation approaches. Neoliberal government reforms looked for the emulation of business ideas and purposes in public sectors that required the modes of evaluation supported by NPM. In other words, the goal of the government reform was "transplanting business management ideas and practices in the public sector" (Saint-Martin, 1998: 324), thus allowing for the subsequent rationalization and privatization of state efforts. In that sense, it is not surprising that governments looked for the kind of skills and expertise that professionals educated in the NPM provided. It was an expertise that mirrored the expertise of managers, professionals who in the past were employed by private business consulting firms.
- Finally, the rapid globalization of this perspective on evaluation and governance in education can be associated with the increasing influence on international organizations (IOs) and the drive to implement international comparative studies. For instance, Kellaghan and Greaney (2001) describe the globalization of assessments of performance results across national educational systems in the last decades. Such assessments have become frequently associated with the provision of baseline data for educational reforms (p. 90), and their implementation has been actively supported by intergovernmental organizations (IGOs) such as the World Bank, UNESCO, and the OECD.

Those IOs were among the first institutions to shift toward neoliberal positions. For instance, Klaus Armingeon points out that the OECD "...from the mid-1970s until the end of the 1990s, ... exerted a unidirectional effect on national welfare states, supporting the idea of welfare-state retrenchment and an increased bonus on individuals and families to shoulder greater personal responsibilities for their security in times of need" (Armingeon, 2004: 227).

At the same time, in educational policy, the information supplied by IOs influenced national policymakers in specific directions. The common denominator was a shared instrumental view of education as serving the national economic development of countries, while linked to individual success in the global economy. Comparative assessment provided comparative quantitative indicators of the knowledge skills deemed economically competitive. An example of this type of assessment is the PISA.

The neoliberal era of policy evaluation is preeminently positivistic and economic, supporting what Fischer (1995) calls a technocratic world view that claims the value neutrality of the evaluation and the authority of the manager as policy expert. As an instrument of analysis, the objectives of the evaluations were limited to those outcomes considered of interest to the neoliberal paradigm – a paradigm that states that education is a private economic good. It was only in the mid-1990s that one started to observe a process of transition toward what appeared to be a new era in policy evaluation.

Era of Knowledge Economy

The focus of the OECD's (1996) influential view based on the early work of Machlup (1962), Porat (1977), and new growth theory (Romer, 1994) emphasized the importance of question of knowledge codification (know what, know why, know how, and know who), the dimension of tacit knowledge, the need for continuous learning, and the importance of knowledge networks with strong policy implications for employment policy. As the OECD report suggests: the economy becomes a hierarchy of networks, driven by the acceleration in the rate of change and the rate of learning. Accordingly, government policy and its evaluation should aim at (1) enhancing knowledge diffusion, (2) upgrading human capital, and, (3) promoting organizational change to increase flexibility, particularly relating to work arrangements, networking, multiskilling of the labor force, and decentralization. In this context, evaluation policies have focused on performance management and related research-performance-monitoring of staff in higher education together with increased global benchmarking and active management of knowledge-assets management including the careful audit of intellectual property. Much of the human-capital thrust has led to

the design of competency criteria both for teachers, teacher education, and for students. Student assessment increasingly emphasizes what is called competence-based assessment, focusing on the assessment of the competences to learn and to create new knowledge in the learning community, such as the ability to apply and create knowledge to solve problems, the ability to communicate domain knowledge to various audiences, and the ability to work with others of diverse backgrounds.

The OECD report is divided into three sections, focusing on trends and implications of the knowledge-based economy, the role of the science system in the knowledge-based economy, and indicators, essentially a section dealing with the question of measurement. In the summary, the OECD report discusses knowledge distribution (as well as knowledge investments) through formal and informal networks as being essential to economic performance, and hypothesizes the increasing codification of knowledge in the emerging information society. In the knowledge-based economy, innovation is driven by the interaction of producers and users in the exchange of both codified and tacit knowledge. The report points to an interactive model of innovation (replacing the old linear model) which consists of knowledge flows and relationships among industry, government, and academia in the development of science and technology. With increasing demand for more highly skilled knowledge workers, governments will need to enhance the capacity to learn and the knowledge-distribution power of the economy through collaborative networks and the diffusion of technology.

The science system – public research laboratories and institutions of higher education – is seen as one of the key components of the knowledge economy, and the report identifies the major challenge as one of reconciling traditional functions of knowledge production and training of scientists with its newer role of collaborating with industry in the transfer of knowledge and technology. Economies are more strongly dependent on knowledge production, distribution, and use than ever before and that knowledge-intensive service sectors (especially education, communications, and information) are the fastest-growing parts of Western economies, which, in turn, are attracting high levels of public and private investment.

New growth theory, in particular, demonstrates that investment in knowledge is characterized by increasing rather than decreasing returns, a finding which modifies the neoclassical production function which argues that returns diminish as more capital is added to the economy. Knowledge also has spillover functions from one industry or firm to another; yet types of knowledge vary: some kinds can be easily reproduced and distributed at low cost, while others cannot be easily transferred from one organization to another or between individuals. Thus, knowledge (as a much broader concept than information) can be considered in terms of know what and know why,

broadly as what philosophers call propositional knowledge (knowledge that), embracing both factual knowledge and scientific knowledge, both of which come closest to being market commodities or economic resources that can be fitted into production functions. Other types of knowledge, what the OECD identifies as know how and know who, are forms of tacit knowledge (after Polanyi (1967); see also Polanyi (1958)), which are more difficult to codify and measure. The OECD report indicates that “Tacit knowledge in the form of skills needed to handle codified knowledge is more important than ever in labour markets” (p. 13) and reasons that, “Education will be the centre of the knowledge-based economy, and learning the tool of individual and organisational advancement” (p. 14), where learning by doing is paramount.

It is argued that the knowledge economy is different from the traditional industrial economy because knowledge is fundamentally different from other commodities, and that these differences, consequently, have fundamental implications both for public policy and for the mode of organization of a knowledge economy. Following the New Keynesian, Joseph Stiglitz (1999), we can analyze the knowledge economy in terms of the scarcity-defying characteristics of ideas. Stiglitz argues that knowledge is a public good because it is nonrivalrous, that is, knowledge, once discovered and made public, operates expansively to defy the normal law of scarcity that governs most commodity markets. Knowledge in its immaterial or conceptual forms – ideas, information, concepts, functions, and abstract objects of thought – is purely nonrivalrous, that is, there is essentially zero marginal costs to adding more users. While nonrivalrous, knowledge can be excluded – the other property of a pure public good – from certain users through various forms of legal protection. Yet, even though knowledge is not a pure public good, there are extensive externalities (spillovers) associated with innovations, which do not necessarily accrue to the innovators.

Higher education, and in particular, the universities, are at the heart of the new policy developments surrounding the concept of the knowledge economy. There is a move toward a system rationalization at all levels, with an accent on enhancing and rewarding the quality of research. The transformation of higher education in many Western countries from a universal welfare entitlement, first, into a private investment in human capital and, second, to a fully consumer-driven system, has followed a now-familiar pattern: a transparent alignment of the university system to reflect the needs of an emerging postindustrial economy, with increasing demands for highly trained, multiskilled, tertiary-educated workers; new forms of corporate managerialisms have been introduced with the emulation of private-sector management styles; and, the introduction of user charges, student loans, has led to the creeping privatization of the system as a whole.

The massification of higher education has been based on new funding mechanisms involving an everincreasing proportion of income from student fees and contestable research funding. Even with the diversification of funding sources, universities have struggled to cope financially. Increasingly, institutions have been forced not only to compete with each other in the market for student places but also to absorb the cost of providing extra, unfunded student places, at declining levels of state funding. With the emergence of the policy concept of the knowledge economy, universities are struggling to take competitive advantage in a new, complex environment that no longer privileges national or regional sites.

Neoliberal market fundamentalism which holds that markets are self-correcting has been discredited and it is important to revisit the goals of knowledge-economy policies and their evaluation in the Obama era. The move to state-centric policies and to forms of federal regulation in the United States and elsewhere now seem almost inevitable. Government intervention is now suddenly back in fashion and on the books at the International Monetary Fund (IMF) and the World Bank. The move to federal regulation and a reform of the financial system seems to chime with the development of state capitalism elsewhere, especially in East Asia, and other forms of state centrism seen as necessary for job creation and national reinvestment in infrastructure (Peters, 2008a). The Obama administration is aware that investment in education and America’s social infrastructure is an important part of the successful recipe for long-term growth and recovery. Obama’s policy advisors are also aware that education policies and their successful evaluation now depend upon the recognition of alternative modes of social production and the developing international networks of collaboration that are important for the recovery and continued success of American science and technology.

Openness has emerged as an alternative mode of social production based on the growing and overlapping complexities of open source, open access and open archiving, and open publishing. It has become a leading source of innovation in the world global-digital economy increasingly adopted by world governments, international agencies, and multinationals as well as leading educational institutions. It is clear that the free software and open-source movements constitute a radical nonproprietary alternative to traditional methods of text production and distribution. This alternative nonproprietary method of cultural exchange threatens traditional models and the legal and institutional means used to restrict creativity, innovation, and the free exchange of ideas. In terms of a model of communication, there has been a gradual shift from content to code in the openness, access, use, reuse, and modification reflecting a radical personalization that has made these open characteristics and principles increasingly the basis of the cultural sphere (Peters, 2008b).

Increasingly, in this context, the evaluation of education policy must take into account the huge reversal of neoliberalism and the shift toward new Keynesianism and especially the connection between investment in education and its relation to high-skill job creation in the vital sectors of energy and new technologies. Evaluation of education policy therefore must acknowledge the changing conceptions of education in relation to dominant economic theories and take into account shifts in ideology. In the era of knowledge economy, policies must be designed with an understanding of the logic of networks and evaluation itself must be an indigenous part of the question of design.

Conclusion

In this article, we have described how evaluation of education policy differs according to policy era and in response to the main underlying policy values. We have analyzed the three main policy eras in education policy – the welfare state, the neoliberal era, and the era of the knowledge economy. In each of these policy eras, education has been seen as fulfilling different functions. Evaluation in each case differs accordingly. One of the metatheoretical questions for policy evaluators is the extent to which their theories and methodologies recognized macro policy changes and match the era and the policy evaluated.

Acknowledgments

The authors acknowledge the helpful and constructive comments of an anonymous reviewer on an earlier version of this paper.

See also: Ability Testing; An Overview of Research in Curriculum Inquiry; Educational Measurement: Overview; Equity and Educational Effectiveness; Evaluation Methodology; Philosophy and Educational Research; Studies of School Improvement in Developing Countries; Theoretical Concepts in the Economics of Education; Understanding Approaches to Evaluation.

Bibliography

- Armington, K. (2004). OECD and national welfare state development. In Armington, K. and Beyeler, M. (eds.) *The OECD and European Welfare States*, pp 226–242. City: Edward Elgar Publishing.
- Boruch, R. F. and Riecken, W. H. (eds.) (1975). *Experimental Testing of Public Policy*. Boulder, Co: Westview Press.
- Braybrooke, D. and Lindblom, C. E. (1970). *A Strategy of Decision. Policy Evaluation as a Social Process*. Free Press/Macmillan: New York.
- Bröckling, U. (2004). Evaluation. In Bröckling, U., Krasmann, S., and Lemke, T. (eds.) *Glossar der Gegenwart*, pp 76–81. Frankfurt a.M.: Suhrkamp.

- Coleman, J. S., Campbell, E. Q., Hobson, C. J., et al. (1966). *Equality of Educational Opportunity*. Washington, DC: US Government Printing Office.
- Concliff-Lagemann, E. (1997). Contested terrain. A history of education research in the United States. *Educational Researcher* 26(9), 5–17.
- Esping-Andersen, G. (1990). *The Three Worlds of Welfare Capitalism*. Cambridge: Cambridge University Press.
- Felix, J. L. (1979). Research and evaluation to improve instruction: The Cincinnati Strategy. In *Educational Analysis and Policy Analysis*.
- Fischer, F. (1995). *Evaluating Public Policy*. Chicago: Nelson-Hall Publishers.
- Foucault, M. (1992). *Archäologie des Wissens*. (Archeology of Knowledge). Frankfurt a.M. 5. Auflage. Suhrkamp.
- Foucault, M. (2004). *Geschichte der Gouvernementalität Band II. Die Geburt der Biopolitik (History of Governmentality II. The Birth of Biopolitics)*. Frankfurt a.M.: Suhrkamp.
- Ginsburg, A., Mc Laughlin, M., Plisko, V., and Takai, R. (1992). Research news and comments: Reinvigorating program evaluation in the US Department of Education. *Educational Researcher* 21(3), 24–27.
- Guba, E. G. and Lincoln, U. S. (1982). *Effective Evaluation*. San Francisco, CA: Josey-Bass.
- House, E. (1990). Research news and comment: Trends in evaluation. *Educational Researcher* 19, 24–28.
- Höhne, T. (2006). *Evaluation als Medium der Exklusion. Eine Kritik an disziplinärer Standardisierung im Neoliberalismus* (Evaluation as medium of exclusion). In Weber, S. M. and Maurer, S. (eds.) *Gouvernementalität und Erziehungswissenschaft. Wissen – Macht – Transformatio*, pp 197–218. Wiesbaden. S.
- Jemelka, R. P. and Borich, G. D. (1979). Traditional and emerging definitions of educational evaluation. *Evaluation Review* 3(2), 263–276.
- Kellaghan, T. and Greaney, V. (2001). The globalisation of assessment in the 20th century. *Assessment in Education* 8(1), 87–102.
- Kogan, M. (1975). *Educational Policy Making. A Study of Interest Groups and Parliament*. London: George Allen & Unwin Ltd.
- Kogan, M. (1979). *Reviews of National Policies for Education. Education Policies in Perspective. An Appraisal*. OECD.
- Machlup, F. (1962). *The Production and Distribution of Knowledge in the United States*. Princeton, NJ: Princeton University Press.
- Murray, C. (1984). *Losing Ground. American Social Policy 1950 – 1980*. New York: Basic Books.
- Organisation for Economic Cooperation and Development (OECD) (1996). *The Knowledge-Based Economy*. Paris: OECD.
- Peters, M. A. (2008a). The global failure of neoliberalism: Privatize profits; socialize losses. *Global-e, A Global Studies Journal*. Available at: <http://global-ejournal.org/2008/11/06/the-global-failure-of-neoliberalism-privatize-profits-socialize-losses/>.
- Peters, M. A. (2008b). Globalization and the virtues of openness in higher education. *Global-e, A Global Studies Journal*. Available at: <http://global-ejournal.org/2008/08/29/globalization-and-the-virtues-of-openness-in-higher-education/>.
- Peters, M. A., Besley, A. C., Olssen, M., Maurer, S., and Webers, S. (eds.) (2009). *Governmentality Studies and Education*. Rotterdam: Sense.
- Polanyi, M. (1958). *Personal Knowledge: Towards A Post-Critical Philosophy*. London: Routledge.
- Polanyi, M. (1967). *The Tacit Dimension*. London: Routledge & Kegan Paul.
- Pollitt, C. (1993). Occasional excursions: A brief history of policy evaluation in the UK. *Parliamentary Affairs* 46(3), 353(10). Academic OneFile. Gale. University of Illinois Urbana-Champaign.
- Porat, M. (1977). *The Information Economy: Definition and Measurement*. Washington DC: US Government Printing Office.
- Rogers-Dillon, R. (2004). *The Welfare Experiments* Stanford: Stanford Law and Politics.
- Ravitch, D. (1983). *The Troubled Crusade*. New York: Basic Books.
- Romer, P. (1994). The origins of endogenous growth. *Journal of Economic Perspectives* 8.
- Saint-Martin, D. (1998). The new managerialism and the policy influence of consultants in government: An historical-institutionalist analysis of Britain, Canada and France. *Governance* 11(3), 319–356. Retrieved January 17, 2009, from Academic Search Premier database.
- Stiglitz, J. (1999). *Public Policy for a Knowledge Economy*. Remarks at the Department for Trade and Industry and Center for Economic

- Policy Research, London, Jan. 27, available at: <http://www.worldbank.org/html/extdr/extme/jssp012799a.htm>.
- Stufflebeam, D. L. and Webster, W. J. (1980). An analysis of alternative approaches to evaluation. *Educational Evaluation and Policy Analysis* 2, 5–20.
- Tyler, R. W. (1950). *Basic Principles of Curriculum and Instruction*. Chicago: University of Chicago Press.
- Weiss, C. H. (1988a). If program decisions hinged only on information: A response to Patton. *American Journal of Evaluation* 9, 15–28.
- Weiss, C. H. (1988b). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice* 9(1), 5–19.

Further Reading

- Peters, M. A. (2001). National education policy constructions of the knowledge economy: Towards a critique. *Journal of Educational Enquiry* 2(1), 1–22.
- Peters, M. A. (2002). The university in the knowledge economy. In Cooper, S., Hinkson, J., and Sharp, G. (eds.) *Scholars and Entrepreneurs: The Universities in Crisis*, pp 137–152. Melbourne: Arena.
- Popham, W. S. (1975). *Educational Evaluation*. Englewood Cliffs, NJ: Prentice Hall.
- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy* 98(5), 71–102.

Relevant Websites

- <http://www.bc.edu/research> – Centre for the Study of Testing, Evaluation, and Educational Policy (CSTEED).
- <http://www.cemcentre.org> – Curriculum, Evaluation and Management (CEM) Centre. The website of the Curriculum, Evaluation and Management Centre is based in Durham University in North-East England.
- <http://www.intute.ac.uk> – Further Education Research Association (FERA).
- It aims to help disseminate advances in good practice, research findings, and policy evaluation relevant to the further education and training sector.
- <http://ies.ed.gov> – Institute of Education Sciences. The Education Sciences Reform Act of 2002 established a new organization within the US Department of Education, the Institute of Education Sciences.
- <http://www.iea.nl> – International Association for the Evaluation of Educational Achievement (IEA).
- <http://www.nationalschool.gov.uk> – Policy Hub, Magenta Book is a website, developed by the Government Social Research Unit, which aims to improve the way public policy is shaped and delivered in UK.
- <http://www.spear.govt.nz> – Social Policy Evaluation and Research Committee.
- A cross-agency group established by the New Zealand Government in 2001 to oversee the government's investment in social policy research and evaluation.

Evaluation of Mathematics Education Programs

J Cai, University of Delaware, Newark, DE, USA

© 2010 Elsevier Ltd. All rights reserved.

Mathematics Education as a Discipline

While mathematics has been a subject of study in school for many centuries, mathematics education is a relatively new field of study. Early in the twentieth century, it was established as an independent field of research. The most significant milestone was establishing The International Commission on Mathematical Instruction (ICMI) in 1908. In 1969, the first International Congress on Mathematical Education (ICME) was held in Lyon. Since then, the ICME is held every four years under the auspices of the ICMI. The aim of the Congress is to present the current states and trends in mathematics education research and the practice of mathematics teaching at all levels. In 2008, the 11th ICME was held in Monterrey, Mexico.

In its first century, the field of mathematics education has prospered, particularly in the past four decades. The publications of research journals and training doctoral students are two indications for the prosperity of mathematics education as a field. So far, the field has launched nearly 20 research journals, such as *Journal for Research in Mathematics Education*, *Educational Studies in Mathematics*, *ZDM-International Journal on Mathematics Education*, *For the Learning of Mathematics*, *Mathematics Education Research Journal*, *Mathematical Thinking and Learning*, *Journal of Mathematical Behavior*, *School Science and Mathematics*, *Journal of Women and Minorities in Science and Engineering*, *Journal of Mathematics Teacher Education*, *Statistics Education Research Journal*, *Journal of Computers in Mathematics and Science Teaching*, and the *International Journal of Mathematical Education in Science and Technology*. Many institutions of higher education around the world have doctoral programs in mathematics education.

Mathematics education, as a field of study, is multidisciplinary as well interdisciplinary and deals with theories, practices, policies, curriculum, and issues about the teaching and learning of mathematics. It is multidisciplinary, because the field adopts methods and perspectives in other disciplines, such as psychology and education, to study issues in mathematics education. In recent years, there is a major shift to interdisciplinary collaboration. For example, a group of cognitive scientists, educational researchers, and mathematics educators bring their respective expertise together to explore a particular issue in mathematics education.

Research in mathematics education is quite varied, ranging from studies of young children to adults, from

large-scale experimental designs to single-subject case studies, from studies in a local context to studies involving multiple nations, and from studies by single authors to studies involving international collaboration.

Formative and Summative Evaluations in Mathematics Education

The evaluation of mathematics education programs is an integral part of mathematics education. Evaluation is the process of judging the value or worth of something. Like in any evaluation, the critical words in the evaluation of mathematics education programs are value or worth. When we evaluate a mathematics education program, we engage in a process designed to provide information to help make a judgment about the program. In evaluating such programs, the worth or value is usually related to students' learning.

Sometimes in mathematics education, evaluation has been used interchangeably with assessment and testing, without a clear distinction (National Council of Teachers of Mathematics, 1989). In fact, assessment is a broad term defined as a process for obtaining information that is used to make decisions about students, curricular programs, and policy (National Council of Teachers of Mathematics, 1995). Thus, assessment can provide information to evaluate a mathematics education program, but is not always involved in the judgment of worth. A test is defined as an instrument or systematic procedure for observing and describing one or more characteristics of a student. In other words, a test is a special form of assessment using a formal instrument, but an assessment does not have to involve a test because we can assess a student's learning through informal observation. Test scores are usually used to judge how well students perform or how effectively teachers teach. Please note that testing itself is neutral and does not involve any value judgment. As a result, a test is a powerful tool for evaluating mathematics education programs.

Generally speaking, every evaluation process involves an object to be evaluated, a scale of value, and a way of collecting information so that the object can be placed on the scale of value for judgment (Eisner, 1994; Fitzpatrick *et al.*, 2004). The ultimate goal of program evaluation is to improve students' learning of mathematics. To reach this goal, both formative and summative evaluations are needed.

An evaluation is defined as formative if the primary goal is to provide information for program improvement. For example, if the goal of evaluating teaching of mathematics is to improve teaching, then it is formative evaluation. Formative evaluation determines if a mathematics education program is delivered appropriately and then determines ways for improvements if they are needed. In the formative evaluation of teaching, for example, the evaluation provides a way to understand teachers' current way of teaching and then creates a plan to enhance their professional growth (Fitzpatrick *et al.*, 2004; NCTM, 1991).

An evaluation is defined as summative if the primary goal is to determine the effectiveness of a mathematics education program. It looks at the results. It is common to speak of short-term and long-term outcomes for a mathematics education program. For example, we can evaluate if a particular instructional intervention is effective within one year (short-term outcome) or in four or more years (long-term outcome).

Curriculum and Mathematics Education Programs

In mathematics education, even though the term program is used in a rather broad sense, it usually refers to a curricular program. Curriculum plays a significant role in mathematics education because it influences what students learn and when and how well they learn it. Advocates of mathematics education reform often attempt to change classroom practice, and thereby students' learning, by changing the curriculum. Historically, curriculum conveys what students should learn and at the same time also serves as an agent for instructional improvement (Howson *et al.*, 1981; Senk and Thompson, 2003).

Since the International Association for the Evaluation of Educational Achievement (IEA)'s First International Mathematics Study (Husen, 1967) used three levels of curriculum (intended, implemented, and attained), these distinctions of the curriculum have been widely accepted in mathematics education. The categorization in the first IEA international study highlights the differences in what a society would like to have taught, what is actually taught, and what students have actually learned (Akker *et al.*, 2003; NRC, 2004; Pinar, 2003; Senk and Thompson, 2003).

The intended curriculum refers to the formally written documents that set system-level expectations for learning of mathematics. It usually includes goals and expectations set in the educational system with textbooks, official syllabi or curriculum standards, and course. The intended curriculum deals at the system level.

The implemented curriculum refers to school and classroom processes for teaching and learning of mathematics as interpreted and implemented by teachers,

according to their experiences and beliefs for particular classes. It deals at the classroom level. The classroom is central to students' learning since students acquire most of their knowledge and form their attitudes from classroom instruction (Robitaille and Garden, 1989). Regardless of how well a curriculum is designed, it has little value unless it is implemented in the classroom.

The attained curriculum refers to what is learned by students and is manifested in their achievements and attitudes. It is at the student level. The attained curriculum deals with the aspects of the intended curriculum that are taught by teachers and actually learned by students.

The conceptualization of the three levels of curriculum is widely used in evaluating curricular programs in mathematics education (NRC, 2004). In the following sections, evaluating the curriculum at each level is discussed separately. However, it should be remembered that all three levels are related to one other and each one supports the other in the evaluation process, as shown in Figure 1.

Evaluation of Intended Curriculum

The intended curriculum specifies goals, topics, sequences, instructional activities, and assessment methods and instruments. The evaluation of the intended curriculum should be guided by the goals of mathematics education. Recently, a set of new goals for mathematics education has been discussed, with an emphasis on the importance of thinking, understanding, reasoning, problem-solving, connections, applications, and communication (e.g., National Council of Teachers of Mathematics, 1989, 1991). These goals stand in contrast to a more conventional view of mathematics education, which involves memorizing and reciting decontextualized facts, rules, and procedures, with an emphasis on the application of well-rehearsed procedures to solve routine problems. For example, in the United States, the National Council of Teachers of Mathematics (1989) proposed five new goals for students to learn in school mathematics:

1. they learn to value mathematics,
2. they become confident in their ability to do mathematics,
3. they become mathematical problem-solvers,
4. they learn to communicate mathematically, and
5. they learn to reason mathematically.

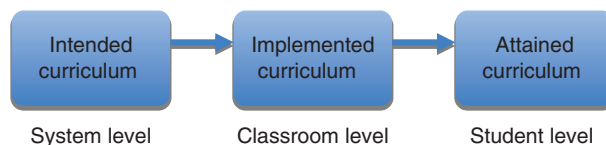


Figure 1 The conceptualization of the three levels of curriculum.

In other countries too, there are similar goals for school mathematics (e.g., Basic Education Curriculum Material Development Center, National Ministry of Education, 2001; Cockroft, 1982; Treffers, 1987). The new goals of mathematics education represent a major shift from traditional practice in mathematics instruction (NCTM, 1989), including shifts:

- toward classrooms as mathematical communities – away from classrooms as simply a collection of individuals;
- toward logic and mathematical evidence as verification – away from the teacher as the sole authority for right answers;
- toward mathematical reasoning – away from merely memorizing procedures;
- toward conjecturing, inventing, and problem-solving – away from an emphasis on mechanistic answer-finding; and
- toward connecting mathematics, its ideas, and its applications – away from treating math as a body of isolated concepts and procedures.

In measuring the effectiveness of a curricular program at the intended level, we must determine if all the relevant contents are covered and if they are sequenced in a coherent manner so that the curriculum meets the goals of mathematics education. For example, in determining if the curriculum includes the goals of developing students' problem-solving, reasoning, and communication abilities, we need to evaluate how these processes are integrated across different mathematical topics.

The evaluation of intended curriculum must take into account the quality of activities, their use in instruction, and their frequency of use. If an intended curriculum claims to be problem-based, we should expect to see a large proportion of cognitively demanding tasks. When evaluating the effectiveness of a standards-based mathematics curriculum in the United States, for example, Cai *et al.* (in press) analyzed instructional tasks in both standards-based mathematics curriculum (called Connected Mathematics Program or CMP in short) and a more traditional mathematics curriculum (called nonCMP). The intent of CMP curriculum is to build students' understanding of major ideas in numbers, algebra, geometry, measurement, data analysis, and probability through explorations of real-world situations and problems. We classified the instructional tasks from CMP and nonCMP curricula into four increasingly demanding categories of cognition: memorization, procedures without connections, procedures with connections, and doing mathematics (Stein and Lane, 1996). As **Figure 2** illustrates, the CMP curriculum had significantly more higher-level tasks (procedures with connections and doing mathematics) ($\chi^2(3, N = 3311) = 759.52, p < 0.0001$) than the nonCMP. Such an analysis of the intended

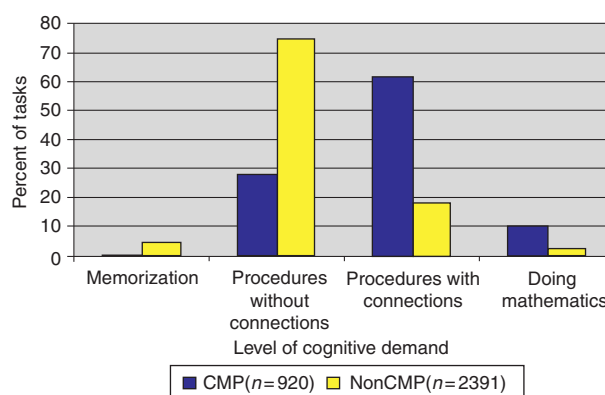


Figure 2 Percentages of various types of tasks in CMP and nonCMP curricula.

curriculum provides insightful information to distinguish different curricula.

The most common method of evaluating an intended curriculum is content analysis which involves judging the quality of contents and their presentation in a curriculum. The National Research Council proposed a list of factors to consider when conducting content analysis to evaluate the intended curriculum (**Table 1**). When doing this, we may focus on one or more factors, depending on the specific purpose of the evaluation.

Because of the conceptual and procedural nature of mathematics, it is important to achieve an appropriate balance between concept development and computational proficiency in a curriculum (NCTM, 1989). An inordinate emphasis on skill proficiency at the expense of a strong conceptual framework to support such skills renders curricular materials inappropriate. On the other hand, it is equally inappropriate to overemphasize conceptual understanding at the expense of basic mathematical skills.

Besides the factors mentioned for evaluating an intended curriculum, we may also consider the qualifications of both, the authors of the curriculum and the qualifications of evaluators. It is ideal to create a team of evaluators with diverse areas of expertise, including mathematicians, mathematics educators, school mathematics teachers, curriculum developers, and school administrators, in the evaluation of an intended curriculum.

Evaluation of Implemented Curriculum

The implemented curriculum concerns both, the mathematics that is taught and how it is taught. One of the ideas about the implemented curriculum acknowledges that what teachers teach may or may not be consistent with what is intended in the curriculum. The implemented curriculum then can be evaluated in terms of what teachers actually do in the classroom with respect to the curriculum.

Table 1 Factors to consider in content analysis of mathematics curriculum materials

Listing of topics
Sequence of topics
Clarity, accuracy, and appropriateness of topic presentation
Frequency, duration, pace, depth, and emphasis of topics
Grade level of introduction
Overall structure: integrated, interdisciplinary, or sequential
Types of tasks and activities, purposes, and level of engagement
Use of prior knowledge, attention to (mis)conceptions, and student strategies
Reading level
Focus on conceptual ideas and algorithmic fluency
Emphasis on analytic/symbolic, visual, or numeric approaches
Types and levels of reasoning, communication, and reflection
Type and use of explanation
Form of practice
Approach to formalization
Use of contextual problems and/or elements of quantitative literacy
Use of technology or manipulatives
Ways to respond to individual differences and grouping practices
Formats of materials
Types of assessment and relation to classroom practice

Adapted from NRC (National Research Council) (2004). *On Evaluating Curricular Effectiveness: Judging the Quality of K-12 Mathematics Evaluations*, p. 42. Washington, DC: National Academy Press.

Evaluating the implemented curriculum includes examining the congruity between the instruction to students and the goals (also called fidelity of implementation). It is possible that teachers may vary widely in their commitment to the intended curriculum. Therefore, it is important to determine whether, how, and to what extent the intended curriculum influences the instruction of the teachers who use them in evaluating the implemented curriculum. The factors listed in **Table 1** can also be considered in examining the fidelity of implementation. For example, we can determine an implementation level (high, medium, or low) according to the extent of coverage of the curricular materials, the consistency between the intended and implemented curricular goals, and the congruence between instructional sequence and curricular sequence.

In classroom instruction, it is possible that teachers may supplement the intended curriculum extensively. Thus, to judge the extent to which a curriculum is being implemented as intended by the program, it is necessary to document in detail the degree to which the curriculum has been used in the classroom and to see if teachers supplement the curriculum, and if they do, what supplementary materials they use.

A faithfully implemented curricular program may not be effective for students' learning. Therefore, in evaluating the implemented curriculum, it is important to examine how mathematics is taught by analyzing

the experiences students and teachers undergo in the classrooms as they use the intended curriculum. Research has shown that high quality mathematics instruction should not only provide students with the opportunity to learn important mathematics and participate actively in the processes of constructing knowledge, but should also provide a setting for students to explain and justify their thinking and challenge the explanations of their peers and teachers (NCTM, 1991; NRC, 2001). Students' ability to reason, solve problems, and use mathematics to communicate their ideas will develop only if they actively and frequently engage in these processes. Whether students come to view mathematics as an integrated whole instead of a fragmented collection of arbitrary topics and whether they ultimately come to value mathematics will depend largely on how the subject is taught (NCTM, 1991). Therefore, evaluating classroom instruction also requires a careful analysis of the classroom events that instructionally guide students' learning of important mathematics.

In mathematics education, different paradigms and methods have been used to identify important features of classroom instruction. One of the important constructs to study classroom teaching is the instructional task presented and implemented in the classroom. The term instructional tasks has been referred to by other researchers as academic tasks or mathematical tasks (Doyle, 1983). Mathematical tasks can be defined broadly as projects, questions, problems, constructions, applications, or exercises in which students engage. Mathematical tasks provide intellectual environments within which students can learn and develop mathematical thinking. Tasks help regulate not only students' attention to particular aspects of content but also their ways of processing information. However, only worthwhile problems give students the chance to solidify and extend what they know and stimulate mathematical learning (NCTM, 1991). In the classroom, actual opportunities for students to learn depend on the type of mathematical tasks presented and implemented. Regardless of the context, for a task to be worthwhile, it should be intriguing and should provide a level of challenge that invites speculation and hard work. Most importantly, worthwhile mathematical tasks should direct students toward explicit learning goals by encouraging them to investigate important mathematical ideas and ways of thinking. NCTM (1991) has recommended that students should be exposed to truly problematic tasks in classrooms so that they can practice mathematical sense-making. Doyle (1983) argues that tasks with different cognitive demands are likely to induce different kinds of learning. Mathematical tasks that are truly problematic have the potential to provide intellectual contexts for students' rich mathematical development. Such tasks can promote their conceptual understanding, foster their ability to reason and communicate mathematically, and capture their interests and curiosity (NCTM, 1991).

For example, in evaluating the effectiveness of a standards-based curriculum in the United States, we analyzed instructional tasks implemented in the classrooms using both CMP and nonCMP (mentioned earlier). Using the same scheme developed by Stein and Lane (1996), we classified the instructional tasks from classrooms, using CMP and nonCMP curricula, into four increasingly demanding categories of cognition: memorization, procedures without connections, procedures with connections, and doing mathematics (Cai *et al.*, in press). **Figure 3** illustrates the percentage distributions of the cognitive demand of the instructional tasks implemented in CMP and nonCMP classrooms. A chi-square test shows that the CMP and nonCMP percentage distributions are significantly different ($\chi^2(3, N=1390)=209.42, p<0.0001$). The difference is due to the fact that the CMP classrooms had implemented a larger percentage of high cognitive demand tasks (procedures with connection or doing mathematics) than the nonCMP ones ($z=13.79, p<0.001$). On the other hand, the nonCMP classrooms had implemented a larger percentage of low cognitive demand tasks (procedures without connection or memorization) than the CMP ones. This analysis of instructional tasks is quite informative to examine the curricular programs at the classroom level.

In evaluating an implemented curriculum, it is important to collect various types of information, such as sustained classroom observations, interviews with teachers, written self-reports by teachers, questionnaires, analysis of lesson plans, and summaries of peer observations. Because teachers are at the heart of using curricular materials and making instructional decisions for teaching, they should be active participants in evaluating the implemented curriculum. In particular, the teachers' own reflections of their teaching could be a valuable source in the evaluation process, especially as it relates to formative evaluation.

Evaluation of Attained Curriculum

The ultimate goal of educational research, curriculum development, and instructional improvement is to

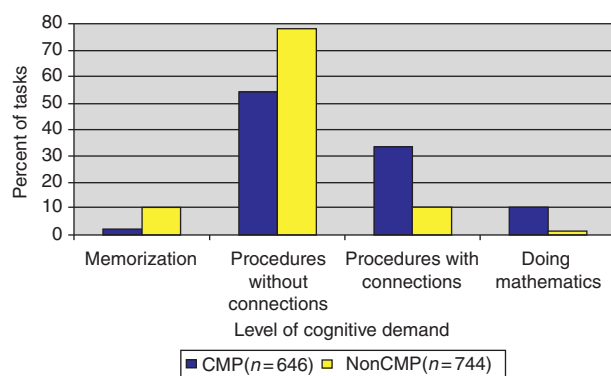


Figure 3 Instructional tasks implemented in CMP and nonCMP classrooms.

enhance student learning. Thus, the evaluation of a mathematics education program at the student level – evaluation of attained curriculum – is of critical importance. To evaluate the attained curriculum, we must address the different facets of mathematical thinking (Sternberg and Ben-Zeev, 1996). Mixed methods, such as observing students doing mathematics, performing tasks, and taking tests, can be used to collect information to evaluate the attained curriculum. In particular, evaluators must pay special attention to the selection of assessment tasks and employment of analysis methods in the evaluation of mathematics education programs.

Assessment tasks

Even though various methods can be used to measure students' learning, the heart of measuring mathematical performance is the set of tasks on which students' learning is to be evaluated (NRC, 2001). It is desirable to use various types of assessment tasks, thereby measuring different facets of mathematical thinking. For example, different formats of assessment tasks (such as multiple-choice and open-ended tasks) may be used to measure students' learning. Multiple-choice tasks have many advantages. For example, more items can be administered within a given time period, and scoring responses can be done quickly and reliably. However, it is difficult to infer students' cognitive processes from their responses to multiple-choice items. Thus, in addition to multiple-choice tasks, open-ended tasks may be used. For open-ended tasks, students are asked to produce answers as well as show their solution processes and provide justifications for their answers. In this way, the open-ended tasks provide a better window into the thinking and reasoning processes involved in their learning of mathematics. Of course, a disadvantage of open-ended tasks is that only a small number of these tasks can be administered within a given period of time. Also, grading students' responses is labor-intensive. To help overcome the disadvantages of using open-ended tasks, a matrix sampling design of administering open-ended tasks to students is recommended. In this way, we can reduce both testing time and grading time but still obtain a good overall estimate of students' learning of mathematics.

Analyses of mathematical performance

It is useful to know the learning outcomes of students in terms of mean scores on various types of tasks. However, comparing their performances from different mathematical programs in terms of correctness of individual tasks is not particularly revealing unless the reviewers explore the thinking and methods that led students to their correct answers. For example, two students may receive the same mean score but use very different solution strategies. Also, two students may receive the same mean score but may make very different errors. Therefore, in evaluating their

learning outcomes, it is important to examine the cognitive aspects of their problem-solving, such as solution strategies, mathematical misconceptions/errors, mathematical justifications, and representations. In fact, examining solution strategies can reveal the qualitative aspects of the mathematical thinking and reasoning of the students, such as how they go about formulating goals and purposes in their learning and mathematical problem-solving (Sternberg, 1991). Similarly, the examination of solution justifications and representations reveals the ways that students process a problem and express their mathematical ideas and thinking processes.

In addition, the unit of analysis should be appropriately selected in analyzing students' performance. In evaluating mathematics education programs at the student level, a reasonable number of students from different classes with different teachers in different schools, or even different school districts, is needed. To examine the impact of a program, pre- and post-tests are usually used and statistical analyses conducted. In traditional analysis using *t*-test or analysis of covariance (ANCOVA), each student in the sample is treated as the same. This implies that students from different classes with different teachers in different schools are considered to have similar experiences with the program. However, the reality is that students from different classes with different teachers in different schools are likely to have different experiences with the program. Thus, in evaluating mathematics education programs, researchers must select an appropriate unit of analysis, take into account the degree to which a program is implemented as well as the relationship between the degree of implementation and students' achievement. Advanced analysis techniques, such as hierarchical linear modeling, can be used to examine the impact of a curriculum on student-learning at the intended (system level) and implemented level (classroom level) simultaneously (Raudenbush and Bryk, 2002; Tarr *et al.*, 2008).

Conclusion

This article discussed the evaluation of mathematics education programs at three levels (intended, implemented, and attained) for both formative and summative purposes. While evaluating mathematics education programs can be discussed and conducted at three levels, the evaluation process at each level should reflect the ultimate goal: to improve and maximize students' learning. Put another way, learning of important mathematics by students should be an important source of information when evaluating a mathematics education program.

It should be indicated that since teaching is a cultural activity, students with different ethnic and cultural backgrounds may respond differently to a new program in

mathematics education. Globalization makes new mathematics education programs more accessible for students around the world. A critical component of any mathematics education program is that all students have an equal opportunity to take advantage of the full benefits of the program (NCTM, 1989). So, the question arises: to what extent does implementing a new mathematics education program improve the learning of mathematics for all students, regardless of the gender, ethnicity, cultural background, language use, and nationality? Clearly, program evaluations in mathematics education should include indicators that specify if the program is meeting these essential criteria. In addition, the evaluation of mathematics education programs should provide information to foster equal access to the programs.

Acknowledgments

During the preparation of this article, the author was partially supported by a grant from the National Science Foundation (NSF) (ESI-0454739). As always, the support from NSF is greatly appreciated. Any opinions expressed herein are those of the author and do not necessarily represent the views of the National Science Foundation.

See also: Assessment in Schools – Mathematics; Cultural Issues that Can Affect the Validity of Educational Evaluations; Curriculum Evaluation; Evaluation Use; Mathematics Learning; Program Evaluation.

Bibliography

- Akker, J. J. H., van den Kuiper, W., and Hameyer, U. (eds.) (2003). *Curriculum Landscape and Trends*. Dordrecht: Kluwer Academic Publishers.
- Basic Education Curriculum Material Development Center, National Ministry of Education (2001). *National Mathematics Curriculum Standards at the Compulsory Education Level (Draft for Consultation)* (in Chinese). Beijing: Beijing Normal University.
- Cai, J., Moyer, J. C., Wang, N., and Nie, B. (in press). The development of middle school students' algebraic thinking: A longitudinal study. In Cai, J. and Knuth, E. (eds.) *Early Algebraization*. New York: Springer.
- Cockroft, W. H. (1982). *Mathematics Counts: Report of the Committee of Inquiry into the Teaching of Mathematics in School*. London: Her Majesty's Stationery Office (HMSO).
- Doyle, W. (1983). Academic work. *Review of Educational Research* **53**, 159–199.
- Eisner, E. W. (1994). *The Educational Imagination: On Design and Evaluation of School Programs*, 3rd edn. New York: Macmillan.
- Fitzpatrick, J. L., Sanders, J. R., and Worthen, B. R. (2004). *Educational Evaluation: Alternative Approaches and Practical Guidelines*. Boston, MA: Pearson.
- Howson, G., Keitel, C., and Kilpatrick, J. (1981). *Curriculum Development in Mathematics*. Cambridge: Cambridge University Press.
- Husen, T. (1967). *International Study of Achievement in Mathematics: A Comparison of Twelve Countries*, vol. I and II. New York: Wiley.

- NCTM (National Council of Teachers of Mathematics) (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: Author.
- NCTM (National Council of Teachers of Mathematics) (1991). *Professional Standards for Teaching Mathematics*. Reston, VA: Author.
- NCTM (National Council of Teachers of Mathematics) (1995). *Assessment Standards for School Mathematics*. Reston, VA: Author.
- NRC (National Research Council) (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.
- NRC (National Research Council) (2004). *On Evaluating Curricular Effectiveness: Judging the Quality of K-12 Mathematics Evaluations*. Washington, DC: National Academy Press.
- Pinar, W. (ed.) (2003). *International Handbook of Curriculum Research*. Mahwah, NJ: Erlbaum.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edn. Thousands Oaks, CA: Sage.
- Robitaille, D. F. and Garden, R. A. (1989). *The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics*. New York: Pergamon Press.
- Senk, S. L. and Thompson, D. R. (eds.) (2003). *Standards-Based School Mathematics Curricula: What Are They? What Do Students Learn?*. Mahwah, NJ: Erlbaum.
- Stein, M. K. and Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation* 2(1), 50–80.
- Sternberg, R. J. (1991). Cognitive theory and psychometrics. In Hambleton, R. K. and Zaal, J. N. (eds.) *Advances in Educational and Psychological Testing: Theory and Applications*, pp 367–393. Boston, MA: Kluwer Academic Publishers.
- Sternberg, R. J. and Ben-Zeev, T. (eds.) (1996). *The Nature of Mathematical Thinking*. Hillsdale, NJ: Erlbaum.
- Tarr, J. E., Reys, R. E., Reys, B. J., et al. (2008). The impact of middle-grades mathematics curriculum and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education* 39(3), 247–280.
- Treffers, A. (1987). *Three Dimensions: A Model of Goal and Theory Description in Mathematics Instruction – The Wiskobas Project*. Dordrecht: Reidel.

Evaluation of Science Education Programs

G Berberoğlu, Middle East Technical University, Ankara, Turkey

© 2010 Elsevier Ltd. All rights reserved.

Evaluation of Science-Education Programs

Science Education and Basis for Evaluation

At present, science education is one of the primary requirements of a literate society where students should be strengthened with basic scientific skills. They should develop curiosity and excitement as well as positive attitudes toward science in order to understand the natural world around them, use scientific processes in dealing with their personal decisions, cope with the scientific and technological aspects of today's world, and, eventually, act as scientifically literate persons to contribute to productivity in society (Buxto and Provenzo, 2007; Carin and Bass, 2001; Gregory and Hammerman, 2008). Therefore, a science program should stimulate the engagement in science by emphasizing the processes in the making of science, organizing scientific ideas, and developing a scientific temper through inquiry rather than imparting mere scientific knowledge (Bloom, 2006; Freeman and Taylor, 2006). Within this context, evaluation becomes a process dealing with all the possible indicators of a successful implementation through a systematic procedure which consists of a formal appraisal of the quality of educational phenomena (Popham, 1993). This is the program-accountability issue (Scheerens *et al.*, 2007). Thus, program accountability is established through appraising the quality in line with the program's overall planned conceptual framework in terms of its scope and good choice of objectives, its implementation practices in line with planned activities, and the level of attainments of the objectives (Scheerens *et al.*, 2007). For program accountability, it is also important to consider the satisfaction of various stakeholders, education policymakers, experts, and any other agency dealing with the program outcomes, since every program should serve the intellectual and social development of a society (Marsh and Willis, 2007). Thus, evaluating a science program is a dynamic process in which any information about the quality will provide continuous feedback to its conceptual framework as well as the implementation practices.

Steps for evaluating a science program

In general, a program evaluation attempt needs to address the quality issue – in terms of input, process, and output (Posner, 1995). **Figure 1** represents the model for evaluating a science program.

As it is seen in **Figure 1**, program evaluation in science consists of three general phases, such as a contextual

evaluation of the program's overall framework with its goals and objectives, evaluation of the implementation process, and evaluation of the program's output. For creating a base in program accountability, the evaluation attempts should be (1) comprehensive in collecting information from all the related elements and stakeholders, (2) combinations of different program evaluation approaches, and (3) using both qualitative and quantitative data collection and analysis procedures.

For ensuring comprehensiveness of the information collected, program accountability should also rely upon judgmental evaluation of experts, teachers, school administrators, education policymakers, stakeholders, and customers, besides student-related learning outcomes (Gredler, 1996; Morris *et al.*, 1987; Posner, 1995; Stecher and Davis, 1987; Worthen *et al.*, 1997).

Even though evaluation approaches vary in terms of research methodologies, criteria to be focused on, nature of the data collected, and standards to be addressed for quality indicators, all the approaches try to collect as much evidence as possible to prove that the program under investigation operates as intended (Herman *et al.*, 1987). Thus, a unified approach with the combination of various evaluation methods would provide a wider range of information with regard to the quality of a science program.

The use of both quantitative and qualitative approaches should be favored since any important finding in quantitative analyses could be probed via qualitative techniques.

On the other hand, as seen in **Figure 1**, the successful implementation of a science program also relates to student and teacher background characteristics, as well as the infrastructure of the school, as intervening variables.

Evaluation is a decision-making process through which a judgmental conclusion is drawn, based on information collected at different phases of the program-evaluation model, as depicted in **Figure 1**. The most critical issue in this process is selecting appropriate criteria to decide about the quality of the program. When the steps of the science evaluation model is considered, coverage of the standards defined in the conceptual framework of the program – with respect to basic skills and knowledge universally accepted in science education, valid and congruent implementation of instructional modes, materials, laboratories, etc., in the classroom – and attainment of skills and attitudes by students, could be the major focus of the evaluation process.

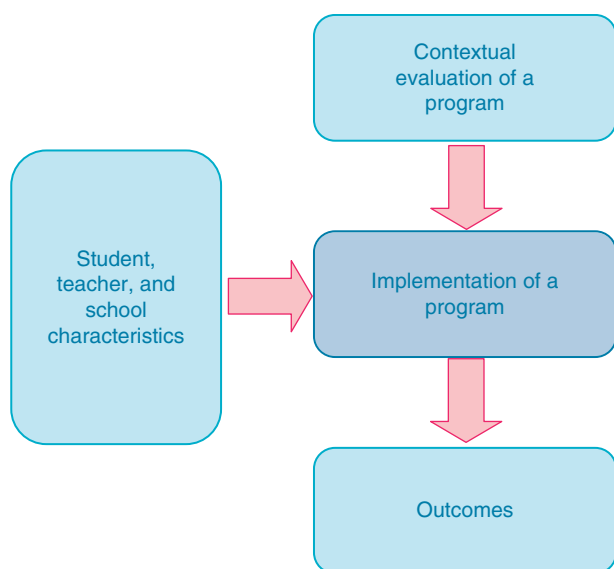


Figure 1 Steps for evaluating a science program.

A program evaluator should consider the following for each phase of the program evaluation:

1. Determining the tasks to be evaluated.
2. Planning the process to collect information about the tasks.
3. Determining criterion against which the quality will be addressed.
4. Integrating all the information collected.
5. Judgmental decision with regard to the overall success of the program.

The following sections provide examples for the tasks to be evaluated as well as criteria to be considered in each step of the evaluation model given in **Figure 1**.

Contextual evaluation

This phase, basically, focuses on the definition of the program. Successful implementation of the program requires a well-defined overall goal, instructional objectives, and planned activities. There are always differences between the program's overall planned structure and what is being implemented in practice (Marsh and Willis, 2007). Thus, before evaluating the program implementation and outcomes, the contextual evaluation of a science program constitutes a base for the program implementation.

In the definition of a science program, there should be at least two dimensions to consider: subject matter and cognitive processes.

The selection of appropriate subject matter and their sequence and organization are the primary tasks to be considered at this very first stage. Selection of subject matter is a function of age and grade levels, but it is linked to cognitive processes as well. In general, subject matter in science consists of facts, concepts, and principles. For

meaningful learning to take place, the subject matter should be in congruence with the developmental level of the students. Any concept which is too abstract for the learners will hinder meaningful learning. This issue is rather the validity of the subject matter domain and its significance with reference to the needs of the students and society (Oliva, 2001). A program evaluator should take the developmental stages of the students into consideration as well as the needs of the stakeholders in assessing the appropriateness and sequential organization of the subject matter.

Cognitive processes, on the other hand, are the higher-order thinking processes which are inquiry skills in science that are required through all the age and grade levels. After clarifying the overall goal of a science program, the primary concern of the program developer is to link the cognitive processes with the subject matter, through the program's objectives, as learning outcomes.

For an evaluator, the major questions to ask at this phase are: How well was the program planned and designed in line with the subject matter and cognitive process? Were the objectives correctly selected and defined? In a science program, the objectives could be planned and designed with respect to various approaches, such as behaviorist or cognitive (Klopfer, 1971; Marzano *et al.*, 1988; Royer *et al.*, 1993). With the increasing impact of constructivism on program-development attempts, science programs could be designed in spiral curriculum format – through which the same cognitive skills can be repeatedly introduced to students through all the grade levels (Henson, 2006). Since the evaluation of this domain constitutes the appropriateness of the program's overall structure, no matter which approach is being used, the primary concern should be the content of the cognitive processes program covers. No matter which program-development approach is used, besides basic conceptual understanding, the science program should emphasize inquiry skills in stating its objectives.

Table 1 exemplifies the higher-order and inquiry skills and their definitions.

There are, actually, two criteria in deciding the quality of the content of the program in this phase. The first criterion is external in nature in which the curriculum or syllabus of another traditional or alternative program constitutes the main source. Similarly, there could be criterion reflecting societal needs with regard to the outcomes of a science program which are described at the national level and that is congruent with the internationally accepted criteria. The National Standards in Science Education (National Research Council, 1996) could be a good example and source for this kind of assessment. In this task, the evaluator needs to conduct a content-wise comparison to justify the congruence of the content of the science program with reference to globally accepted standards.

The second criterion is internal, through which consistency among different components of the program is

Table 1 Examples of science process skills

Process skills	Explanation
Transfer	Converting one form of information into another.
Prediction	Based on a criterion anticipating a future condition.
Relating	Link at least two body of information with each other.
Providing examples	With reference to a given situation proposing a novel example.
Classification	With reference to at least two groups of information, locating or naming a given case.
Hypothesizing	Based on an observation result proposing a statement which is testable.
Observing	Based on five common senses describing attributions or properties of a given event or object.
Conducting experiments	Based on a hypothesis, designing an experimental apparatus.
Data collection	Collecting or analyzing data with reference to a given hypothesis.
Drawing conclusion	Providing a statement that may support an experimental hypothesis or observations.

assessed. The tasks to be actualized in this phase are whether: (1) the program emphasizes development of inquiry-based thinking skills in the overall goal, (2) the program provides well-defined objectives in line with these skills, (3) there is congruence between the overall framework and the objectives, (4) there is congruence between the objectives and the subject matter selected, and finally (5) there is congruence between the objectives and the activities planned to achieve these objectives. Well-defined science objectives are those that are attainable at the respective age level and should be stated in a way that reflects higher-order cognitive and inquiry skills acquisition of which can help students perform to a minimum acceptable level (Henson, 2006).

The cognitive processes are not the only concern at this step. A science program should also cover objectives and tasks to develop students' skills in conducting experiments with safety precaution. Development of attitudes, values, and beliefs is also considered for a successful science program. In this context, conceptions about the nature of science, positive attitudes, and interest toward science are the basics to be covered in order to develop scientific curiosity in students. It is expected that a science program with inquiry emphasis would also develop these psychomotor and affective components of the program. All the procedures suggested above for cognitive skills are also valid for these outcomes.

At this particular step, the appropriateness of the program's content should be assessed by the panel of experts consisting of related stakeholders, such as experts in science education, education policymakers, school

administrators, teachers, and parents. This phase of the program evaluation requires a qualitative description of the content of the program and content-wise comparison with the existing internal or external criteria. A detailed report for the tasks given above will help in deciding the quality of the program in terms of its relevance (Scheerens *et al.*, 2007). Within this context, opinions of the related experts, policymakers, and teachers could be used as supportive information for making a contextual analysis.

Evaluation results in this phase will be useful input for the following phases of the program-evaluation process.

Evaluation of program implementation

The implementation, primarily, focuses on the process with actual settings. In this respect, an evaluator should deal with the possible sources of barriers and limitations which could hinder the attainment of program goals (Rossi and Freeman, 1985). The program implementation, basically, uses formative evaluation techniques through which the evaluator should obtain an answer to the question "How well is the program being implemented?" Thus, answering this question requires close interaction of the evaluator with the school administrator, teachers, and students. Interaction includes all sorts of activities being carried out in the classroom environment as well. Adequacy and effectiveness of materials, such as textbooks, worksheets, and laboratory equipment, have to be assessed in this phase of evaluation. As a result of this interaction, the evaluator needs to prepare a detailed description of the program operations (Patton, 1987). Thus, questionnaires, open-ended in-depth interviews, direct observation of the process, and contextual analyses of written documents and anecdotal records are used to describe the program operation. On the other hand, the evaluator could also collect quantitative information through structured questionnaires. Since the major concern is to understand the strengths and weaknesses of program implementation, any problem singled out as a result of quantitative analyses could be further analyzed by qualitative methods, in order to understand the possible sources of the problems that hinder successful implementation. For instance, if teachers and students report that they do not spend much time on student-based experiments in science classes, further in-depth interviews with teachers may provide information on the reasons for this particular outcome found in the quantitative questionnaire. Thus, in this phase, qualitative analyses provide further evidence to support the general findings reported by quantitative data.

Table 2 indicates examples of the possible formative evaluation tools and how they could be designed to collect information about the implementation of science programs.

The techniques proposed to use in process evaluation – such as questionnaires, interviews, observations, and analyses

Table 2 Examples of formative evaluation tools for program implementation

<i>Decisions</i>	<i>Source</i>	<i>Instrument</i>	<i>Rating</i>
<i>Congruence of activities with respect to program objectives</i>	Student	Questionnaire	Frequency/likelihood of activities carried out in the classroom/laboratory
	Teacher	Questionnaire	Frequency of activities
	External evaluator	Observation record	Frequency and quality of laboratory or classroom activities
<i>Quality of the instructional materials and laboratory equipment</i>	Teacher	Interview	How teachers value textbooks and other instructional materials being used
	Student	Interview	How students perceive the instructional process and the materials they use
<i>Adequacy of materials, laboratory equipment</i>	Teachers and school principals	Interview	How they value the adequacy of the materials and laboratory equipments and their effective use in the classroom
<i>Impact of program implementation on student learning</i>	Student	Unobtrusive measurement	Student response on questions embedded in the instructional processes
	Teacher	Interview	How teachers appreciate the impact of ongoing programs on student development and emphasize on the strengths and weaknesses of the program implementation

of written documents – have their own advantages and disadvantages with respect to information provided and the technical characteristics such as reliability and validity of the results (King *et al.*, 1987). Since results obtained in these techniques are relatively less reliable and valid than assessment of cognitive skills, the key feature in this particular step should be comparing the results on similar issues across the groups, such as students, teachers, and school administrators. For instance, information on the frequency of student-based inquiry experiments in the classroom could be obtained from a students and teachers, and a comparison of their responses provides evidence for the validity of the information collected. Similarly, triangulation is necessary for the qualitative data for increasing the strength and rigor of the information collected (Patton, 2002).

In this phase of the evaluation, equity and economic use of resources are two important features that have to be considered by the evaluator. As indicated in **Figure 1**, this is one of the primary concerns of the evaluation process since the variations in the infrastructure among schools will reflect the quality of the program implementation. The success of the program will partly be dependent upon equity in sharing resources among schools. Thus, this particular variable might be controlled in anticipating the quality of the program's output.

Evaluation of outcomes

The major focus in this phase is “How much of the objectives are attained by the students?” Answering this question, as the major task, includes the summative evaluation of objectives related to cognitive and affective domains.

The summative assessment of students' attainment of cognitive objectives could be performed in norm-referenced and criterion-referenced approaches. The norm-reference approach ranks the students, and, in general, overall mean of the group or the percentile scores of the students are used as the criteria to make a decision regarding the success of the program. Outcome evaluation is, generally, carried out by an experimental design in which a new program is compared with the traditional one in terms of students' summative test scores. However, this approach seems quite misleading since, in general, summative tests used for the comparison is biased against the traditional group since, in general, tests used for this particular comparison are designed in line with the objectives of the new curriculum. On the other hand, the mean difference between traditional and experimental groups does not indicate for which cognitive skills the students' attainment is different across two groups being compared. Similarly, surveying the achievement of students across a new science program and the traditional one without any experimental control may be misleading as well. The mean could be different in favor of the new science program but it may not display a satisfactory level of attainment of the objectives. Mere ranking of students with respect to their achievement levels does not give a clear picture concerning the quality of new science programs with reference to attainment of the cognitive skills. Thus, a more elaborative assessment is needed to interpret the skills possessed by the students who were exposed to new science programs.

Elaborate assessment could be carried out by criterion-referenced interpretations of the test scores. The criterion-referenced approach would be very informative

with regard to the success of a program if the cognitive skills corresponding to different score levels are clearly defined (Hambleton and Jurgensen, 1990). Definition of score intervals – in line with the cognitive skills achieved by the students – is possible through item-mapping procedures (Ercikan, 2006; Karantonis and Sireci, 2006). Using this approach indicates not only the achievement difference, but the level of attainment of the program objectives at different score levels as well. This kind of information can be gathered through the use of item-response-theory models – in which probabilistic interpretations of the correct response rate at different score intervals are available (Hambleton *et al.*, 1991). Moreover, having a score level and cognitive process link in any assessment program is also helpful in describing the minimum standards students should achieve in science education throughout the country.

In this summative evaluation process, students' individual or group products – such as laboratory reports, project papers, workbooks, concept maps, portfolios, etc. – can also be used as informative documentation with regard to the success of the program. This is a rather content-wise evaluation of the quality of the products produced by students, which may provide clues regarding the attainment of program objectives. For improving the reliability and validity of product evaluation, a panel of experts and teachers could develop the standards for assessment, and at least two independent groups of scorers could evaluate the content of student products.

In terms of affective domain outcomes, attitudes toward science and science instruction, interest, academic concept in science, student willingness to engage in science in their careers, nature of scientific concept, value given to science etc., have to be assessed in this phase. The assessment of affective variables is, generally, carried out with Likert-type scales. The comparisons between the new science program and traditional or alternative programs, with respect to these affective variables, may also provide information about the success of the program output. As in the case of cognitive measurement, the criterion-referenced interpretations of the mean scores could also be used in evaluating the results of affective measurements. The scale mean on a five-point Likert-type scale could be used as the success or failure of the program in terms of attainment of affective objectives.

Considering Background Characteristics in Evaluation

The success of a science program depends on some external factors – such as the school's infrastructure, socioeconomic level of the parents, students' needs and preachievement in science, and qualifications of science teachers. The school's infrastructure includes variables such as science-teacher-student ratio, facilities in science laboratories, availability of

instructional materials, etc. The information concerning these characteristics could be assessed through the use of written documents, or via interview or questionnaire data obtained from school principals. The socioeconomic characteristic of a student is an important factor related to science achievement. The educational level of parents and facilities at home, such as the number of books and the availability of separate study desks are the primary indicators of this particular variable (Kohr *et al.*, 1988). This could be assessed through the student questionnaire used in the second phase of the program evaluation. Students' needs and expectations in science education, and their preachievement levels as assessed by the teacher grades or standardized tests could be important inputs for a successful implementation. Finally, teachers' experience in teaching science, their efficacy in science, their understanding of the content and aim of the new science program, their understanding with regard to the nature of scientific concepts, etc., are the variables that might reflect the quality of program implementation that an evaluator needs to assess and control in the program-evaluation design.

These external variables are either controlled experimentally – if the program evaluation is based on experimental design – or statistically – if the evaluation design is primarily based on a survey. The covariance analysis could be used to infer success of the program when school-, teacher-, and student-related factors are statistically removed from the output variables. Alternatively, these variables could be the major concern of the evaluator in assessing the success of the science program under different conditions, such as experienced versus nonexperienced teachers and high versus low socioeconomic status. The gender variable, on the other hand, should be specifically taken into consideration for a science program since gender differences in science achievement are, generally, reported across all countries (Martin *et al.*, 2008). A successful science program should be effective for both gender groups.

Data Analysis and Reporting the Results

As stated previously, program-evaluation data are both qualitative and quantitative in nature. Thus, two broad evaluation reports need to be developed. However, these two data sources should be aggregated to compensate each other. For the rigor of the evaluation attempt, there are two concerns for the evaluator: reliability and validity of the data collected and generalizability of the results.

Reliability and validity of the measurement results could be enhanced by triangulation of qualitative data and by using reliable and valid standardized tests or questionnaires in quantitative analyses. For quantitative data, outcomes of a program could also be partly evaluated with respect to available data sources at the national and international level such as National

Assessment of Educational Progress (NAEP), The Third International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA).

For generalizability of the results, the evaluator should consider representativeness of the sample. Based on the design of the study, samples should include all the strata in the population.

There needs to be an analysis design for the qualitative and quantitative data sets. The evaluator needs to combine these two sources of information in a way to support the validity of the conclusions derived. The qualitative reports should be providing supportive evidence to the findings reported on quantitative data. On the other hand, the quantitative data could be analyzed via descriptive statistical techniques, as well as inferentially. Descriptive analyses can document all the descriptive statistics across different strata of the sample – such as schools, urban and rural areas, and gender groups. In the inferential analyses, the linear structural model or the hierarchical linear model could be used to aggregate all the variables and data sources collected through tests and questionnaires (Jöreskog and Sörbom, 1993; Raudenbush and Bryk, 2002). In the modeling process, the major problem is the criterion through which one can generalize with regard to the effectiveness of the new science program. The models comparing traditional and new programs could help to infer the effectiveness of the new science program. If there is no alternative or traditional program to compare, the evaluator might constitute a model based on pretest and posttest design. Considering the difficulty of the experimental approach, a criterion-referenced interpretation of the test and questionnaire scores could help to decide on the quality of the science program. Since the effectiveness of the program could be observed in time, a time-series statistical design to follow-up student development in cognitive and affective areas could constitute a valid basis for evaluation.

A program has a dynamic nature. Any feedback from an evaluation attempt could be used for revising the program structure, as well as the implementation process. Thus, program evaluation in science should be considered as a continuous process to enhance the quality of program structure and implementation.

See also: Cultural Issues that Can Affect the Validity of Educational Evaluations; Curriculum Evaluation; Evaluation Use; Impact of Assessment on Students' Learning Strategies and Implications for Judging Assessment Quality; Program Evaluation; The Role of Stakeholders in Educational Evaluation.

Bibliography

Bloom, J. W. (2006). *Creating a Classroom Community of Young Scientists*, 2nd edn. New York: Routledge Taylor and Francis Group.

- Buxto, C. A. and Provenzo, E. F., Jr. (2007). *Teaching Science in Elementary and Middle School: A Cognitive and Cultural Approach*. Los Angeles, CA: Sage.
- Carin, A. A. and Bass, J. E. (2001). *Teaching Science as Inquiry*, 9th edn. New Jersey: Merrill Prentice Hall.
- Ercikan, K. (2006). Examining guidelines for developing accurate proficiency level scores. *Canadian Journal of Education* 29, 823–838.
- Freeman, G. and Taylor, V. (2006). *Integrating Science and Literacy Instruction: A Framework for Bridging the Gap*. Lanham, MD: Rowman and Littlefield Education.
- Gregory, G. H. and Hammerman, E. (2008). *Differentiated Instructional Strategies for Science, Grades K-8*. Thousand Oaks, CA: Corwin Press/Sage.
- Hambleton, R. K. and Jurgensen, C. (1990). Criterion-referenced assessment of school achievement. In Reynolds, C. R. and Kamphaus, R. W. (eds.) *Handbook of Psychological and Educational Assessment of Children: Intelligence and Achievement*, pp 456–476. New York: Guilford Press.
- Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Henson, K. T. (2006). *Curriculum Planning: Integrating Multiculturalism, Constructivism, and Educational Reform*, 3rd edn. Long Grove, IL: Waveland Press.
- Jöreskog, K. and Sörbom, D. (1993). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Chicago, IL: Scientific Software International.
- Karantonis, A. and Sireci, S. G. (2006). The bookmark standard setting method: A literature review. *Educational Measurement: Issues and Practice* 25(1), 4–12.
- Kohr, R. L., Coldiron, R., Skiffington, E. W., Masters, J. R., and Blust, R. S. (1988). The influence of race, class, and gender on gender on self-esteem for fifth, eighth, and eleventh grade students in Pennsylvania Schools. *The Journal of Negro Education* 57(4), 467–481.
- Marsh, C. J. and Willis, G. (2007). *Curriculum: Alternative Approaches, Ongoing Issues*, 4th edn. Columbus, OH: Pearson, Merrill Prentice Hall.
- Martin, M. O., Mullis, I. V. S., and Foy, P. (eds.) (2008). *TIMSS 2007 International Science Report*. Boston, MA: TIMSS and PIRLS International Study Center.
- Marzano, R. J., Brandt, R. S., Huges, C. S., et al. (1988). *Dimensions of Thinking: A Framework for Curriculum and Instruction*. Alexandria: The Association for Supervision and Curriculum Development.
- Morris, L. L., Fitz-Gibbon, C. T., and Freeman, M. E. (1987). *How to Communicate Evaluation Findings*. Newbury Park: Sage.
- National Research Council (1996). *National Science Education Standards*. Washington, DC: National Academy Press.
- Oliva, P. F. (2001). *Developing the Curriculum*, 5th edn. New York: Longman.
- Patton, M. Q. (1982). *Practical Evaluation*. Newbury Park, CA: Sage.
- Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods*, 3rd edn. Thousand Oaks, CA: Sage.
- Posner, G. J. (1995). *Analyzing the Curriculum*. New York: McGraw-Hill.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edn. Thousand Oaks, CA: Sage.
- Rossi, P. H. and Freeman, H. E. (1985). *Evaluation: A Systematic Approach*, 3rd edn. Beverly Hills, CA: Sage.
- Royer, J. M., Cisero, C. A., and Carlo, M. S. (1993). Techniques and procedures for assessing cognitive skills. *Review of Educational Research* 63(2), 201–243.

Further Reading

- Gredler, M. E. (1996). *Program Evaluation*. Upper Saddle River, NJ: Prentice Hall.
- Herman, J. L., Morris, L. L., and Fitz-Gibbon, C. T. (1987). *Evaluator's Handbook*. Newbury Park, CA: Sage.
- King, J. A., Morris, L. L., and Fitz-Gibbon, C. T. (1987). *How to Assess Program Implementation*. Newbury Park, CA: Sage.

- Klopfer, L. E. (1971). Evaluation of learning in science. In Bloom, B. S., Hastings, J. T., and Madaus, G. F. (eds.) *Handbook on Formative and Summative Evaluation of Student Learning*, pp 559–641. New York: McGraw-Hill.
- Morris, L. L., Fitz-Gibbon, C. T., and Freeman, M. E. (1987). *How to Communicate Evaluation Findings*. Newbury Park, CA: Sage.
- Patton, M. Q. (1987). *How to Use Qualitative Methods in Evaluation*. Newbury Park, CA: Sage.
- Popham, W. J. (1993). *Educational Evaluation*, 3rd edn. Boston, MA: Allyn and Bacon.
- Scheerens, J., Glas, C., and Thomas, S. M. (2007). *Educational Evaluation, Assessment, and Monitoring*. London: Taylor and Francis.
- Stecher, B. M. and Davis, W. A. (1987). *How to Focus an Evaluation*. Newbury Park, CA: Sage.
- Worthen, B. R., Sanders, J. R., and Fitzpatrick, J. L. (1997). *Program Evaluation: Alternative Approaches and Practical Guidelines*, 2nd edn. New York: Longman.

Evaluation of Social Studies and Civic Education Programs

S G Grant, Binghamton University, State University of New York, Binghamton, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Center for civic education – A California-based organization that promotes the civic knowledge and awareness.

CityWorks – A curriculum package developed by Constitutional Rights Foundation to promote civic knowledge and awareness.

Civic education – A field of academic endeavor intended to promote students' knowledge and understanding of government, the political process, and citizenship rights and responsibilities.

Committee on the social studies – A group of US academics enlisted in 1916 by the National Education Association and empowered to draft a set of curriculum guidelines for a new school subject called "social studies."

Constitutional rights foundations – An organization based in California that develops curriculum and instructional materials to promote student understanding of the US Constitution and Bill of Rights.

Documented-based questions – A form of test task that asks students to use primary and secondary documents in the construction of their responses to an essay prompt.

Man: A course of study – A social studies curriculum package developed in the 1960s.

National center for history in the schools – A curriculum development center located at the University of California, Los Angeles that develops and promotes curriculum and instructional resources around the teaching of history.

National council for the social studies – The leading professional organization for educators interested in social studies education.

No child left behind – US national legislation created in 2002 featuring a standardized test-based approach to educational reform.

Social studies – The broad label for a series of school coursework that features attention to history and the social sciences.

We the people – A curriculum package designed to promote students' civic knowledge and dispositions developed by the Center for Civic Education.

The evaluation of social studies and civic-education programs in the United States is both challenging and rare. The first challenge is curricular: Social studies and civic education curricula have long been contentious and hide-bound fields. Many would argue that the lines within and between the two fields are murky at best. A second set of challenges surrounds social studies and civics assessments. There, the key issues are the validity and reliability of differing assessment systems as well as a set of larger concerns related to interpreting what children know and understand. Finally, uncertainty abounds at the classroom level. In that realm, we find that teachers are paying some attention to the state curriculum guides and to state-level tests. But in each case, there is as much variability in their responses as there is consistency. Each of these challenges undercuts the potential for robust curriculum evaluation. Moreover, although there have been evaluations of some foundation-sponsored curriculum projects (e.g., We the People and Kids Voting USA), these projects are almost exclusively civics oriented, small scale, and unrelated to the recent drive to create state curriculum.

A Contentious Curriculum

Not only are the current manifestations of social studies and civic education contentious (so contentious in fact that neither is among the areas to be tested under the No Child Left Behind legislation of 2002), but there is considerable and continuing disagreement about the legitimacy of social studies as a distinct school subject. Civic education has longer history, but even advocates write about it in ways that blur the distinction with social studies.

Although they trace its antecedents back to the late 1890s, most curricular historians (Barr *et al.*, 1977; Hertzberg, 1989; Jenness, 1990; Wattras, 2002) date the birth of social studies as a school subject to 1916 and the Committee of Social Studies (sponsored by the National Education Association) report whose aims included the "development of an appreciation of the nature and laws of social life, a sense of the responsibility of the individual as a member of social groups, and the intelligence and the will to participate effectively in the promotion of the social well-being" (National Education Association, 1916). Born of twin impulses – a need to enliven the pedantic teaching of history and a desire to help students engage more directly with their

social environment – the founders of social studies (e.g., Thomas Jesse Jones and Arthur Dunn initially, George Counts and Harold Rugg later) hoped to create a school curriculum that served both progressive and patriotic ends (Wattras, 2002). Among the curricular changes offered was the high school problems-of-democracy course which supporters envisioned as a way to engage topics of “immediate interest to the class and [of] vital importance to society” (Hertzberg, 1989: 87).

What the Committee on Social Studies and subsequent reformers have failed to do, however, is install a single, coherent program of studies. Instead, social studies tends to be a blanket term for a wide range of disciplinary elements (e.g., history, geography, political science, and economics) and for a wide range of courses. There is something of a *de facto* curriculum (see Table 1), but variety has been more the norm than consistency.

The history of curriculum-reform efforts demonstrates a stubborn resistance to real change (Dow, 1991; Jenness, 1990). In fact, recent efforts have only added to the curricular incoherence. The most comprehensive of those efforts, the National Council for the Social Studies (National Council for Social Studies, 1994) *Expectations of Excellence* and the National Center for History (National Center for History in the Schools, 1996). *National History Standards* illustrate the scatter plot that exemplifies social studies. The National Center for History *Standards* offers a K-12 sequence of history-based topics (e.g., Living and working together in families and communities, Now and long ago for grades K–4 and Contemporary United States for grades 5–12). The authors of the National Council for the Social Studies standards, by contrast, forswear a particular sequence of course topics in favor of a list of 10 standards (e.g., time, continuity, and change and science, technology, and society).

The variety evident in the national standards documents is echoed in the various state social studies curricula (Grant, 2006). Many feature elements of the *de facto*

curriculum listed in Table 1, but considerable variation exists. For example, depending on which state one is looking at, the seventh-grade curriculum could be world history and geography (California), US history and geography (New York), or Asia, the Middle East, and Africa (Georgia). Civics, which is generally included in state social studies curriculum sequences, may or may not be identified as such. For example, twelfth graders in California take a course titled Principles of American Democracy while the course their peers in New York take is called Participation in Government. By contrast, states like Florida and Georgia list courses of study like Civics, American Government, and American Government/Civics, but they do so as topics to be taught anywhere from grades 9–12.

The conceptual elasticity of a catch-all term like social studies can be viewed as an asset: certainly, few of the social situations we encounter are political or economic or geographical alone, so the capacity to think and act across disciplinary boundaries makes sense. The extant social studies curriculum sequences (which typically include civics), however, offer little in the way of rich, sustained study. The crazy-quilt quality evident in most state social studies curriculum meanders in and around history and civics at the secondary level (with some attention to geography) after a vague expanding-horizons curriculum at the elementary level (Jenness, 1990).

The expansive social studies/civics curriculum scene, then, poses problems for curriculum evaluators in that one is seldom sure what curricular area to bring into focus. For example, the concept of voting could be considered a civics idea, a history idea, or a generic social studies idea. Further complicating the evaluation of curricular efforts, however, is the rise of state-level, standardized testing, which in many instances eclipses the relevance to state-level curriculum efforts.

A Problematic Testing Program

For most of its history, social studies and civics have been untested areas. True, they are often part of general achievement tests and some states, like New York, have a long history of state-level exams in history and geography. But while many states have long-standing programs of literacy and mathematics testing, only recently has this been the case with social studies/civics. Given this relatively recent attention, it should surprise few readers to learn that there are a range of psychometric and interpretive challenges associated with social studies exams (Horn, 2006). These challenges undercut the potential for useful evaluation of social studies and civics curriculum programs.

The testing challenges stem largely from the circumstance that the basic psychometric principles of validity and reliability are so elusive. True, no test of social behavior can be perfectly valid nor can any set of results be

Table 1 Typical sequence of social studies courses K–12 in the US

Grade	Course topic(s)
Kindergarten	Myself and my family
Grade 1	My family and my neighborhood
Grade 2	My local community
Grade 3	Communities around the world
Grade 4	My state
Grade 5	The United States, Canada, and Mexico
Grade 6	Area studies (e.g., Europe, Asia)
Grade 7	Area studies, civics, law-related education
Grade 8	US history
Grade 9	World history, social studies electives
Grade 10	World history
Grade 11	US history
Grade 12	Participation in government, civics, electives

perfectly reliable. Those conditions notwithstanding, it is important to understand how problematic the social studies/civics testing landscape really is.

On the validity side, questions arise primarily about the validity of test items. Although it seems reasonable that propositional knowledge can be measured with objective-style questions, a number of observers argue that the inferential quality of some dimensions of some curricular areas (e.g., history) and the dispositional quality of some others (e.g., civics) cannot. Since few observers would argue that social knowledge ought to be reduced to a parade of easily tested facts, social studies/civics test designers immediately run into validity roadblocks.

Although those validity roadblocks can come in any form (i.e., content, criterion, and consequential), it is construct validity that proves most challenging. Horn (2006) notes that, “in the case of history tests, assessing construct validity determines the extent to which a test accurately and adequately captures how much a student knows or can do related to the specific area of interest” (p. 64). In order to determine the construct validity of an exam, the National Research Council (National Research Council, 1999) proposes four key questions: Are the right things being measured in the right balance? Is the scoring system consistent with the structure of the domains about which inferences or predictions are being made? Are the scores reliable and consistent across the different contexts for which they are used, as well as across different population groups? What are the short-term and long-term consequences of score interpretation and use?

Horn (2006) offers a set of more pointed questions that begin with: how is a competent student of social studies/civics defined and how does the test define what a competent student will demonstrate during the test to show that she or he is indeed competent? These would be daunting questions even in the most coherent field of study. Given the struggles social studies and civics educators have faced in defining their fields, the answers to these key questions remain elusive. As a result, although students’ social studies and civics knowledge can be assessed through standardized measures, few observers have confidence that those assessments measure something more than surface-level knowledge (Burroughs *et al.*, 2005).

A different challenge emerges on the reliability front. There, the issue is one of commensurate assessments. Not only have social studies and civics not been regularly tested subjects, but the assessments given typically have not remained the same long enough to provide reliable data over time or from state to state. On the national scene, students’ knowledge and understanding of history and civics have been tested on the National Assessment of Educational Progress (NAEP) exams on multiple occasions. However, the structure of the exams has changed with each administration, so that it is impossible to track

the progress of students over time. State-level tests vary on two dimensions – within and across states. For example, the New York state high school history and geography exam administered at grades 10 and 11 underwent a dramatic change in the late 1990s through the addition of a document-based question (DBQ) and a series of constructed response questions tied to the DBQ documents. This structural change resulted in changes to the scoring rubrics such that comparisons with students who took the old exam could no longer be made. Even less reliable is any comparison of one state’s test scores with those of another state (Burroughs *et al.*, 2005). In effect, the exams vary to such a degree that achieving any kind of equivalency across state scores is out of the question. As an example, consider the differences between the eighth-grade exams in New York and Wisconsin. First, the New York test focuses only on US history and geography whereas the Wisconsin exam covers a broad range of social studies content. Second, the New York exam offers a range of question formats: multiple-choice questions, a thematic essay, constructed-response questions, and a DBQ. The Wisconsin exam, by contrast, consists entirely of multiple-choice questions. Finally, the eighth-grade New York students take an exam that closely parallels the one they will take in eleventh grade. Eighth graders in Wisconsin, however, take a general battery of social studies questions that are unrelated in focus to the tenth-grade exam. The net result is that observers interested in comparing student performance from state to state will find no points of coherence.

The validity and reliability problems present an interpretive challenge: How are we to understand what students know and understand about social studies and civics when the available instruments are both flawed and incompatible? Yet, vexing as these issues are, they pale when compared with the larger issues associated with assessing students in the areas of social studies and civics.

One concern is that, in some cases, students appear to know and understand different things, even over a relatively short period of time (Grant, 2007; VanSledright *et al.*, 2006). In other words, the knowledge that children demonstrate over time seems to change – ideas that once appeared to be solidly held may be forgotten or may mutate into something unrecognizable (VanSledright, 1995). A second concern revolves around the phenomenon that, when asked different kinds of questions about the same topic, children’s responses suggest that they know different things (Grant, 2007; Nuthall and Alton-Lee, 1995; Rogers and Stevenson, 1988). The idea here is that how one probes children’s ideas can influence what they appear to know. One last concern centers on the notion that what children think they know can be a powerful factor in determining what they need to know (VanSledright and Brophy, 1992). Prior knowledge is potent force; what we know shapes what we learn. But

prior knowledge can be faulty and, unless it is challenged effectively, the inaccurate ideas we bring to bear can override more useful new learning. Findings like these ought to give pause to anyone who claims to know what children know about social studies and civics: Not only are our instruments flawed, but the nature of knowing is a fluid construct at best.

Given these challenges – some general, some specific to social studies/civics – it is not surprising that state-level social studies/civics curriculum have been largely ignored by curriculum evaluators.

Uncertainty at the Classroom Level

As noted above, the current tendency to pair curricular revisions with new, and often more assertive assessments, complicates an evaluation of the effect of those revisions on teachers' classroom practices. Teaching is a complicated act: minimally, teachers plan, enact, and assess lessons and units, but each of these actions is only a label for a much more complex situation. As teachers respond to new curriculum and new testing programs, then, it should surprise few observers to learn that these changes produce no single, coherent, or consistent effect. In brief, curricular and testing changes influence teachers' content, instructional, and assessment decisions differently. As a result, the particular effect of any new curriculum program may be difficult to discern.

New curricular and testing programs seem to have the largest effect on teachers' content decisions; they have the least effect on teachers' classroom assessments. Most surprising is the relatively modest effect that the new curriculum and assessments seem to have on teachers' instructional planning and delivery. The notion that testing (and curriculum) drives teaching may make a convenient sound bite, but it neither makes good educational policy nor does it make sense as a way to capture the essence of good teaching.

Curricular and Testing Influences on Teachers' Content Decisions

In terms of pedagogical effects, it appears that new social studies and/or civics curriculum and tests have the biggest impact on teachers' content decisions. The subject matter changes they report range from small (i.e., adding or eliminating a historical person, group, or event) to large (i.e., adding or eliminating whole eras). As state curriculum and state tests send clearer messages about what to teach rather than how to teach it, this finding should surprise few observers.

Although novice teachers struggle with the content choices they make that will benefit their test-taking students most (Segall, 2006; van Hover, 2006), so too do veteran

teachers (Fickel, 2006; Salinas, 2006; Vogler, 2006). Despite the evidence of curriculum narrowing (McNeil, 2000) associated with the new testing programs, many teachers withstand the temptation to tailor their curriculum to either state curriculum or state exams. These teachers plan units on topics covered on state exams, but they also plan units on topics barely mentioned (Bolgatz, 2006; Grant, 2003, 2006). Some researchers find that many teachers assert content control over the elective courses they teach (Fickel, 2006; Gerwin and Visone, 2006), others have documented the work of teachers who resolve the content dilemmas they face by choosing to teach more ambitiously in classes that enrol students of all levels (Gradwell, 2006, 2003).

Curricular and Testing Influences on Teachers' Assessment Practices

The clear, but modest impact that new social studies and civics curriculum and tests are having on teachers' content decisions is not mirrored in terms of teachers' assessment plans. Teachers are making adjustments in the assessments they give, especially in response to state-based tests, but their general approaches to assessing their students seem largely consistent.

On the whole, few teachers appear to be making wholesale changes in their student assessments (Grant, 2003, 2006). This finding is less remarkable than it might seem on the surface: Social studies and civics teachers typically employ test questions that mirror those of state exams – multiple-choice questions, short-answer tasks, and essays. However, teachers are not designing test-based exams exclusively; instead, they use state-like exams as part of their larger assessment plans (van Hover, 2006).

Although they are not making radical changes in their assessment strategies, teachers are not sanguine about the state tests they administer: They dislike the pressure on their practices and on their students, the ways that scores are used, the kinds of test items employed, and the mixed messages that tests send about what is important. Yet, few teachers dismiss outright the idea of a state-level test. Many protest one or more features of state-test construction or the ways in which scores are interpreted; few protest against the very existence of a test. Many reasons support this conclusion – coercion by school and district administrators, pressure from parents and the public, and uncertainty about what seems like an inevitable trend in American education. If these explanations account for the lack of teacher resistance to the concept of testing, so does one other: Most Americans accept the validity of tests as a means of judging student performance. Like the public at large, teachers seem to accept the premise that tests are useful and that multiple-choice questions and essay prompts represent reasonable ways of judging what students know and understand.

Testing now may be an inextricable part of the US school landscape but, if it is, so too is the idea that tests have limited utility. Tests can provide an efficient screening mechanism, presumably separating those who know from those who do not. At the same time, they are a screen with a wide mesh: most Americans can recite a litany of students whose accomplishments were inaccurately predicted by their standardized test scores. In the end, an odd schizophrenia surfaces – faith in tests coexisting with doubts about their relevance.

Curricular and Testing Influences on Teachers' Instructional Practices

In many ways, the most notable finding in the research literature is the uncertain influence state curriculum and testing programs seem to have on teachers' instructional decisions (Au, 2007; Grant, 2003, 2006). Again, new tests seem to be cited as far more influential than new curricula but, in any event, the effects are far from predictable.

More than social studies and civics curriculum, standardized tests are a factor in teachers' practices: Some teachers are engaged in more test preparation than they would prefer, some are discontinuing activities that they have done in the past, and some are using more pedantic approaches than they think wise (Burroughs *et al.*, 2005; McNeil, 2000; Vogler, 2006). The consternation that a Virginia teacher expresses when he observes that “it's facts—names, dates, places. I used to be a good teacher—now I'm cramming this stuff down their throats” (Smith, 2006) echoes throughout a survey of Mississippi teachers' test-influenced instructional practices (Vogler, 2006).

As fact laden as social studies and civics curriculum and tests are, it would surprise few observers to see teachers take the path of least resistance by centering their teaching around rote recall. There are teachers taking and justifying this approach.

Schools are complicated places, however, and simple cause–effect assumptions rarely surface: within the research that demonstrates considerable test-based influence on teachers' instruction are signs of practices that buck this trend. In fact, many teachers continue to teach in ways they think are most appropriate for students to learn powerful ideas (Bolgatz, 2006; Gradwell, 2006; Grant, 2006). Existing alongside lectures, rote recitation, and other seeming concessions to state-level testing are debates, projects, and class discussions (Fickel, 2006; Smith, 2006). It is not the case that tests are inconsequential; it is the case that there is little evidence demonstrating wholesale instructional change toward test-centric instructional practices (Au, 2007).

Although this finding may cause critics of test-based reforms to cheer and the proponents to wince, both groups would be wise to proceed cautiously. Critics of standardized testing (and the accompanying curriculum)

suggest that teachers would routinely plan and enact substantive and engaging lessons were it not for state-level tests that inhibit this creativity (Ohanian, 1999). As sensible as this presumption is, it fails to explain the fact that researchers described much of instruction in social studies and civics classrooms as pedantic at best (Cusick, 1983; Goodlad, 1984). Test-based reforms may not inspire ambitious teaching, but neither do they destroy it (Gradwell, 2006; Grant, 2003, 2006; Libresco, 2005). Instead, new social studies and civics curriculum and tests are no more likely to induce large-scale instructional change as any other innovation (Tyack and Cuban, 1995). Critics ought not to rejoice, however, for although weak teachers may be no more likely to invoke test-influenced practices than their more ambitious peers, that does not mean that their students are any better served.

The overall effect of new state-level curriculum and testing programs on teachers' instructional decisions, then, seems uncertain. Although research evidence exists to support the conclusion that reforms, especially test-based reforms, are provoking considerable change and in less ambitious directions, equally persuasive evidence points in the opposite direction (Au, 2007).

Evaluations of Social Studies and Civics Education Programs

As noted above, the number, range, and scope of evaluations of social studies and civics curriculum are limited. To date, there have been no large-scale, empirical studies of the various state-level social studies/civics curriculum efforts. There have been studies conducted on a small number of civics-specific curriculum projects sponsored by national foundations, but the only government-developed curriculum project to be evaluated in social studies is the Man: A Course of Study program developed in the 1960s.

Civics education is a more narrowly defined area than social studies, so groups like the Center for Civic Education and the Constitutional Rights Foundation have developed and introduced curriculum projects into schools and then commissioned evaluations of the effect. For example, Brody (1994) studied the We the People curriculum created by the Center for Civic Education and found that the students exhibited greater political tolerance than average American citizens. Similarly, Kahne *et al.* (2006) found that students who participated in the Constitutional Rights Foundation CityWorks curriculum exhibited stronger interest in civic participation than their peers.

Broad-based social studies curriculum proposals rarely have been evaluated. Perhaps the best example is the Man: A Course of Study program funded by the National Science Foundation. Although well-funded and promoted widely, the program soon failed. The common view is that social conservatives killed the effort based on their

perception that it promoted cultural relativity (Dow, 1991). Others, however, cite the major problem as the seeming attempt to create a teacher-proof curriculum, one in which teachers have very little control over what is taught (Cort and Peskowitz, 1977).

Recommendations

Anyone making recommendations for more ambitious social studies and civics education programs faces a daunting challenge: as these areas are left out of the No Child Left Behind testing apparatus, school and state-level leaders have little incentive to elevate these elements of the curriculum. Most states have curriculum and standardized tests for social studies, if not specifically for civics, but because the test scores do not figure as prominently as reading and mathematics scores do, many educators focus their attention elsewhere.

That said, two recommendations are offered for improving the quality of social studies and civics-education programs. The first is to appropriate more time and money to teacher professional development. Smith and Niemi (2001) make a compelling case for the relationship between teachers' instructional practices and student performance on standardized tests – that is, ambitious teaching practices are strongly correlated with high student test scores. It follows, then, that investments in teachers' subject matter and pedagogical expertise may offer large academic returns.

The second recommendation relates to the structure of standardized testing. Such tests may not be terribly authentic, but they fail as an assessment only when they are the only assessment. With no serious challenge to state-level testing on the horizon, the issue of how to mitigate the worst effects of testing looms large. Imagine, then, a situation in which sets of more authentic assessments (e.g., a position paper on a local or state-level issue) were used in conjunction with the extant state exams. Imagine also that students had to present a portfolio of their work on these assessments to local committees of teachers, parents, and community members and that these committees held the power to decide when students were ready to graduate. Making decisions about students' success based only on state-mandated exams disenfranchises the people who have the greatest stake in their schools – students, teachers, parents, and community members. A social studies and civics education that expands both the kinds of assessments used and the authority over those assessments would offer powerful incentives for teachers and students to engage in rich and ambitious teaching and learning.

Conclusion

Although social studies and civics are largely ignored in the national No Child Left Behind legislation, these are

subjects that have generated considerable activity in terms of state-level curriculum and testing. These efforts, however, have not been coherently or consistently evaluated. Most of the recent curriculum work has occurred at the state level; yet, most of the curriculum evaluation has been focused on narrowly defined, foundation-sponsored projects. These evaluations demonstrate some effect, but the challenges of assessing the large state-level curriculum efforts have thwarted any real sense of the impact.

See also: Assessment in the History, Civics and Social Studies Domains; Classroom Assessment in Policy Context (New Zealand); National Assessments; Testing in History.

Bibliography

- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher* 36(5), 258–267.
- Barr, R., Barth, J., and Shermis, S. (1977). *Defining the Social Studies*. Arlington, VA: National Council for the Social Studies.
- Bolatz, J. (2006). Using primary documents with fourth-grade students: Talking about racism while preparing for state-level tests. In Grant, S. G. (ed.) *Measuring History: Cases of State-Level Testing Across the United States*, pp 133–156. Greenwich, CT: Information Age.
- Brody, R. (1994). *Secondary Education and Political Attitudes: Examining the Effects on Political Tolerance of the "We the People..." curriculum*. Calabasas, CA: Center for Civic Education.
- Burroughs, S., Groce, E., and Webeck, M. L. (2005). Social studies education in the age of testing and accountability. *Educational Measurement: Issues and Practice* 24(3), 13–20.
- Cort, H. R. and Peskowitz, N. (1977). *A Longitudinal Study of Man: A Course of Study. Summary Report*. Washington, DC: National Science Foundation.
- Cusick, P. A. (1983). *The Egalitarian Ideal and the American High School*. New York: Longman.
- Dow, P. (1991). *Schoolhouse Politics: Lessons from the Sputnik Era*. Cambridge, MA: Harvard University Press.
- Fickel, L. (2006). Paradox of practice: Expanding and contracting curriculum in a high-stakes climate. In Grant, S. G. (ed.) *Measuring History: Cases of High-Stakes Testing Across the US*, pp 75–104. Greenwich, CT: Information Age.
- Gerwin, D. and Visone, F. (2006). The freedom to teach: Contrasting history teaching in elective and state-tested courses. *Theory and Research in Social Education* 34(2), 259–282.
- Goodlad, J. (1984). *A Place Called School*. New York: McGraw-Hill.
- Gradwell, J. M. (2006). Teaching in spite of, rather than because of, the test: A case of ambitious history teaching in New York state. In Grant, S. G. (ed.) *Measuring History: Cases of High-Stakes Testing Across the US*, pp 157–176. Greenwich, CT: Information Age.
- Grant, S. G. (2003). *History Lessons: Teaching, Learning, and Testing in U. S. High School Classrooms*. Mahwah, NJ: Erlbaum.
- Grant, S. G. (ed.) (2006). *Measuring History: Cases of High-Stakes Testing Across the U. S.* Greenwich, CT: Information Age.
- Grant, S. G. (2007). Understanding what children know about history: Exploring the representation and testing dilemmas. *Social Studies Research and Practice* 2(2), 196–208.
- Hertzberg, H. (1989). History and progressivism: A century of reform proposals. In Gagnon, P. (ed.) *Historical Literacy: The Case for History in American Education*, pp 69–99. New York: Macmillan.
- Horn, C. (2006). The technical realities of measuring history. In Grant, S. G. (ed.) *Measuring History: Cases of High-Stakes Testing Across the U.S.*, pp 57–74. Greenwich, CT: Information Age.
- Jenness, D. (1990). *Making Sense of Social Studies*. New York: Macmillan.

- Kahne, J., Chi, B., and Middaugh, E. (2006). Building social capital for civic and political engagement: The potential of high-school civics courses. *Canadian Journal of Education* **29**(2), 387–409.
- Libresco, A. (2005). How she stopped worrying and learned to love the test... sort of. In Yeager, E. and Davis, O. L. Jr. (eds.) *Wise Social Studies Teaching in an Age of High-Stakes Testing*, pp 33–49. Greenwich, CT: Information Age.
- McNeil, L. (2000). *Contradictions of School Reform: Educational Cost of Standardized Testing*. New York: Routledge.
- National Center for History in the Schools (1996). *National Standards for History, Basic Edition*. Los Angeles, CA: Author.
- National Council for Social Studies (1994). *Expectations of Excellence*. Washington, DC: Author.
- National Education Association (1916). *The Social Studies in Secondary Education*. Washington, DC: Author.
- National Research Council (1999). *High-Stakes Testing for Tracking, Promotion, and Graduation*. Washington, DC: National Academy Press.
- Nuthall, G. and Alton-Lee, A. (1995). Assessing classroom learning: How students use their knowledge and experience to answer classroom achievement test questions in science and social studies. *American Educational Research Journal* **32**(1), 185–223.
- Ohanian, S. (1999). *One Size Fits Few: The Folly of Educational Standards*. Portsmouth, NH: Heinemann.
- Rogers, V. and Stevenson, C. (1988). How do we know what kids are learning in school? *Educational Leadership* **45**, 68–75.
- Salinas, C. (2006). Teaching in a high-stakes testing setting: What becomes of teacher knowledge? In Grant, S. G. (ed.) *Measuring History: Cases of High-Stakes Testing Across the US*, pp 177–194. Greenwich, CT: Information Age.
- Segall, A. (2006). Teaching in the age of accountability: Measuring history or measuring up to it? In Grant, S. G. (ed.) *Measuring History: Cases of High-Stakes Testing Across the US*, pp 105–132. Greenwich, CT: Information Age.
- Smith, A. M. (2006). Negotiating control and protecting the private: History teachers and the Virginia standards of learning. In Grant, S. G. (ed.) *Measuring History: Cases of High-Stakes Testing Across the US*, pp 221–248. Greenwich, CT: Information Age.
- Smith, J. and Niemi, R. (2001). Learning history in school: The impact of course work and instructional practices on achievement. *Theory and Research in Social Education* **29**(1), 18–42.
- Tyack, D. and Cuban, L. (1995). *Tinkering Toward Utopia*. Cambridge: Harvard University Press.
- van Hover, S. (2006). Teaching history in the old dominion: The impact of Virginia's accountability reform on seven secondary beginning history teachers. In Grant, S. G. (ed.) *Measuring History: Cases of High-Stakes Testing Across the US*, pp 195–220. Greenwich, CT: Information Age.
- VanSledright, B. (1995). "I don't remember – the ideas are all jumbled in my head" eighth graders' reconstructions of colonial American history. *Journal of Curriculum and Supervision* **10**(4), 317–345.
- VanSledright, B. and Brophy, J. (1992). Storytelling, imagination, and fanciful elaboration in children's historical reconstructions. *American Educational Research Journal* **29**(4), 837–859.
- VanSledright, B., Kelly, T., and Meuwissen, K. (2006). Oh, the trouble we have seen: Researching historical thinking and understanding. In Barton, K. (ed.) *Research Methods in Social Studies Education*, pp 207–233. Greenwich, CT: Information Age.
- Vogler, K. (2006). The impact of a high school graduation examination on Mississippi social studies teachers' instructional practices. In

- Grant, S. G. (ed.) *Measuring History: Cases of High-Stakes Testing Across the US*, pp 273–302. Greenwich, CT: Information Age.
- Wattras, J. (2002). Debating the curriculum: Social studies or history, 1892–1937. *Social Studies* **93**(6), 245–250.

Further Reading

- Abrams, L., Pedulla, J., and Madaus, G. (2003). Views from the classroom: Teachers' opinions of statewide testing programs. *Theory into Practice* **42**(1), 18–29.
- Amrein, A. and Berliner, D. (2002). High-stakes testing, uncertainty, and student learning. *Educational Policy Analysis Archives* **10**(18).
- Burroughs, S., Groce, E., and Webeck, M. L. (2005). Social studies education in the age of testing and accountability. *Educational Measurement: Issues and Practice* **24**(3), 13–20.
- Firestone, W., Monfils, L., Camilli, G., et al. (2002). The ambiguity of test preparation: A multimethod analysis in one state. *Teachers College Record* **104**(7), 1485–1523.
- Gaudelli, W. (2006). The future of high-stakes history assessment: Possible scenarios, potential outcomes. In Grant, S. G. (ed.) *Measuring History: Cases of High-Stakes Testing Across the US*, pp 321–334. Greenwich, CT: Information Age.
- Grant, S. G. (ed.) (2006). *Measuring History: Cases of High-Stakes Testing Across the US*. Greenwich, CT: Information Age.
- Hillocks, G. (2002). *The Testing Trap: How State Writing Assessments Control Learning*. New York: Teachers College Press.
- Horn, C. (2003). High-stakes testing and students: Stopping or perpetuating a cycle of failure. *Theory into Practice* **42**(1), 30–41.
- Linn, R. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher* **32**(7), 3–13.
- Newmann, F., Bryk, A., and Nagaoka, J. (2001). *Authentic Intellectual Work and Standardized Tests: Conflict or Coexistence?* Chicago, IL: Consortium on Chicago School Research.
- Yeager, E. A. and Davis, O. L. Jr. (eds.) (2005). *Wise Social Studies Teaching in an Age of High-Stakes Testing*. Greenwich, CT: Information Age.

Relevant Websites

- <http://www.cde.ca.gov> – California Department of Education.
- <http://www.crf-usa.org> – Constitutional Rights Foundation.
- <http://www.civiced.org> – Center for Civic Education.
- <http://www.doe.k12.ga.us> – Georgia Department of Education.
- <http://nces.ed.gov> – The Nation's Report Card, National Assessment of Educational Progress.
- <http://www.nchs.ucla.edu> – National Center for History in the Schools.
- <http://www.socialstudies.org> – National Council for the Social Studies.
- <http://www.nationalacademies.org> – National Research Council, The National Academies.
- <http://www.nysed.gov> – New York State Education Department.
- <http://www.ed.gov> – No Child Left Behind, US Department of Education.

Evaluation of Teacher Quality and Practice

W Robinson, University of Exeter, Exeter, UK

J Campbell, University of Warwick, Coventry, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Context specificity – In evaluating teachers, this means to give high importance to the particular circumstances (organizational structures, culture, characteristics of students, etc.), within which teachers work.

Credentialism – The use of the level of qualifications held by teachers as a major variable in evaluation.

Multilevel modeling – A set of complex statistical techniques developed to enable researchers to attribute the contribution made to variance in student achievement to factors at different levels in hierarchical structures, such as in schools, leadership, classroom, and in individual student.

Proletarianization – A neologism used to mean the reduction of teaching from full professional status, in which working conditions are controlled by the professional bodies, to a status where conditions are imposed from outside the profession, usually by government.

Value-added – The gain in student attainment from an agreed baseline attributable to the individual teacher or the school, or a combination of both.

In the literature on teacher evaluation, there has been a distinction between teacher quality and teacher practice. Teacher practice is seen as an empirical or descriptive concept referring to all the work of teachers, both inside and outside the classroom. A large number of studies have attempted to delineate the constituent elements of teachers' work, identifying the time given to work overall, and to different activities, such as teaching, planning, assessing, pastoral care, and participating in meetings, within it. These studies have shown that teaching, meaning face-to-face contact with students in classrooms, occupies only about a third of teachers' work, the rest of the time being taken up with planning, marking administration, and other activities. Some studies have attempted to use the data gathered to demonstrate the causes of teacher burn-out, stress, and overconscientiousness, while others have shown that some of teachers' practice is spent on low-level administration and clerical activities that could be carried out by support assistants. Related studies used this

kind of data to analyze teachers' working conditions. (International Labour Office, 1991). Other studies of teacher practice have attempted to describe and analyze the nature of teachers' behavior in classrooms, constructing from observational data different styles of teaching, for example, democratic, child-centered, and progressive, class enquirers, individual monitors.

Teacher quality, by contrast, is a normative concept, concerned with the role of the teacher in enhancing or hindering students' learning, motivation, self-esteem, and aspirations, thereby raising educational attainment. The focus has been almost exclusively upon how the teachers achieve, or fail to achieve, these goals through their interactions with students in classrooms. In these terms, the concept is twofold: what teachers should know (subject knowledge) and what they should be able to do (pedagogical skills). It has therefore, by definition, been concerned with what is thought, by evaluators, to be good practice rather than the routine day-to-day business of teaching, even though there is some overlap between the two concepts.

Although both teacher practice and teacher quality are in principle able to be evaluated, most approaches to teacher evaluation use teacher quality as their organizing focus.

The Purposes of Teacher Evaluation

The last three decades have seen the growth of interest worldwide in formal arrangements for teacher evaluation. Although the forms of such evaluation may vary by education system, and by levels within systems, there are three commonly articulated purposes for it.

The first is that it serves a performativity ideology, whereby public-sector services, including schooling, are required to develop greater accountability to their stakeholders, including parents, students, tax payers, and policy-makers. The defining characteristics of this accountability are primarily economic; teachers are expected to deliver increased productivity, enhanced value for money, and improved educational attainment by their students. This constructs teacher evaluation as part of a wider evaluation of the education system as a whole, attempting to measure and increase the system's efficiency, especially in relation to its ability to provide a skilled workforce. The principal critique of this view is that it has led, under modernizing capitalism, to the proletarianization of teachers and the

intensification of teaching, leading to reduced quality of teaching and learning, as teachers are pressed to cut corners to achieve efficiency (Apple, 1986).

A second purpose is to provide robust evidence upon which teachers' promotion and career development can be equitably supported. Although many education systems retain elements of informal, or even nepotistic, arrangements for promoting teachers, drawing mainly on the inexplicit judgments of inspectors or principals, without observation of teaching, modernizing systems have attempted to develop more apparently rational models; they use formally established arrangements with explicit criteria, appraisal interviews and direct observation of teaching, leading to an official record indicating strengths and weaknesses. This constructs teacher evaluation as a process focused on rewarding or penalizing the individual teacher. In those systems where performance-related pay has been implemented, this aspect has taken increased significance. Where teacher self-evaluation is part of this process, it is designed to improve teachers' skills, expertise, and overall professional development.

However, the rational planning models underlying evaluation for promotion and career development have been criticized as being divorced from the political and economic realities of teacher supply and demand. When economies are booming, the supply of teachers is restricted because of competition from other more lucrative professions, yet schools need to fill teaching positions even if the teachers' quality has been judged to be poor. When economies decline, recruitment of highly qualified graduates into teaching improves. Even in these economies, however, teachers of subjects with high market value, such as mathematics and physics and information technology, are in short supply in education. Thus, the rational models of teacher evaluation, though promising greater equity, have less effectiveness than they imply. Moreover, where there are strong and effective teacher unions, the forms, nature, and consequences of teacher appraisal have to be negotiated with them rather than be imposed on their members. Such negotiations have tended to lead to conservative evaluation procedures that carry few, if any, career consequences.

Third, teacher evaluation serves as an important element in assessing school effectiveness. This purpose envisages the teacher in the classroom as part of the wider school context. Teacher evaluation thereby is constructed as, potentially, a major mechanism for school improvement, assuming it is used formatively.

A main problem with this approach is empirical; it arises from the history of separate development of teacher effectiveness research, school effectiveness research, and the school improvement movement. All three have until recently developed as separate and largely unconnected fields of endeavor, relatively uninfluenced by each other.

There is now a strong movement to link teacher and school effectiveness research, in order to examine the interaction of school and teacher effectiveness variables, and to tease out the contribution made by individual teachers to student learning, drawing on methodological advances in research, such as multilevel modeling. In theory, this would provide a much stronger evidence-based approach to teacher evaluation, although there is further research needed in this respect. In current approaches to school improvement, there remains a strong emphasis on school leadership development as the major mechanism, with the consequence that in almost all countries, the movement has not drawn sufficiently upon the strong evidence base of the educational effectiveness research. This has more than academic interest, since the competing models have substantive consequences for teacher evaluation; the educational effectiveness research identifies the teacher operating at the classroom level as contributing most to student outcome, and therefore prioritizes student learning as the major focus for teacher evaluation.

Forms of Teacher Evaluation

It is common to draw a distinction between formative and summative purposes for evaluation, the former designed to serve an improvement function, the latter a comparative one. However, as can be seen from the preceding paragraph, teacher evaluation arrangements can take both forms; they are complementary rather than oppositional.

A second distinction is between self-evaluation and external evaluation. Self-evaluation is developed within the individual school, often with informal protocols and instruments, and tends to rely on qualitative methodologies and outcomes. Typically, it has drawn upon the teacher's own statement of objectives for the lesson as a basis for observation, or even the teacher's preferred focus for the observation (e.g., quality of the teacher's questioning techniques, or the teacher's use of information and communications technologies) and has concluded with a confidential oral feedback by the observer, who is often a peer rather than a superior. Its principal purpose is the improvement of the quality of the individual teacher's teaching skills, and the evaluation outcomes are confidential to those involved, in order to prevent the process becoming high stakes.

High-stakes evaluation is a term taken to imply that the evaluation will lead to serious and potentially damaging consequences for those teachers whose outcomes are negative. These consequences may include harm to professional reputation, compulsory retraining, reduced promotion prospects, and, in extreme cases, redundancy. The obverse side of the high-stakes coin is sometimes

forgotten; those teachers receiving positive ratings may gain both material and reputational benefits and improved career prospects.

External evaluation is conducted by national, regional, or local inspection agencies operating with official or quasi-official status and formally agreed instruments and protocols. It tends to rely primarily on quantitative methodologies, typically using rating scales of teaching quality, following classroom observation, attainment test data, and estimates of the value added to students' attainment by the quality of teaching. In some systems, the inspection process has statutory force, and the principal purpose is to provide data that can be used for comparing the performance of schools or individual teachers. This makes it high stakes.

A third distinction is between generic and differentiated models of evaluation, with the former being a one-size-fits-all model for evaluation, and the latter more tailored to the different contexts and needs of the individual teachers. The generic model, empirically based on teacher effectiveness research in the USA and UK, has been the most common in practice, for two reasons: it enables the development of fairly simple instruments and carries the appearance of equity, with all teachers being evaluated by identical instruments and procedures; and it has been developed out of the most robust evidence base.

More recently, as teacher evaluation has begun to spread through the whole of schooling, the differentiated model has attracted attention (Campbell *et al.*, 2004). This proposes that five different aspects of teaching may require differences in evaluation. These are the backgrounds of the students taught, the nature of the subject taught, the range of role demands, students' characteristics including age, and the cultural and organizational context within which teaching occurs. For example, the differentiated model proposes that the skills and knowledge required to teach a university doctoral class and a kindergarten class are different. Likewise, in terms of the subject being taught, the methods appropriate for teaching mathematics differ from those needed for teaching drama. There have also been studies suggesting that different teaching methods may be needed with students of different socioeconomic and ethnic backgrounds. It follows that teachers should be evaluated in ways that take account of such differences. A particular interest here has been whether teachers' values and purposes for teaching and learning can be taken into account through a differentiated model. Although there is considerable interest among researchers and scholars in a differentiated model, it has yet to gain widespread adoption in school systems. This is for two reasons: it has the appearance of being too complex to be implemented in practice; and the evidence base is too limited as yet for full development.

Policy Framework

Until recently, teacher evaluation procedures have been derived from specific programs and therefore have been ad hoc, lacking not only a robust evidence-based justification, but also a robust theoretical basis. This remains the state of play in many systems, including the USA, which is thought to be at the forefront of innovation. A wide-ranging analysis of teacher evaluation by a Washington-based think tank, reported that it was still common to use credentialism, that is, the qualifications of teachers, sometimes with a fleeting classroom visit from the principal, as the basis for evaluation in many states (Toch and Rothman, 2008). They point to the need for multiple methods of evaluation, linked to student learning.

However, an important study using data on teacher evaluation procedures in five countries, has attempted to remedy this defect, by proposing the adoption of formal policy linkages between variables found in teacher effectiveness research, teacher evaluation, and school improvement research (Teddle *et al.*, 2003). This demonstrated the relatively loose links between these three in most countries, with the best-established links having been developed in the USA, both because of the longer history of education research there and because researchers in school effectiveness and in school improvement collaborated early, and were able to embed research perspectives into some of the policies driving school reform movements there.

In most other systems, according to Teddle and his colleagues, the links are merely between two elements in the framework, school effectiveness research and teacher evaluation instruments, for example, or are nonexistent. Developing approaches to teacher evaluation which incorporate the key variables from school and teacher effectiveness research, and linking it to teacher and school improvement, could give teacher evaluation the empirical and theoretical justification currently lacking. It might also, through standardizing models and instruments, provide a methodological improvement needed to conduct robust comparative studies.

This latter aim, to develop rigorous and appropriate instruments for teacher evaluation, is nonetheless extremely ambitious. There are myriad instruments and procedures, many of them developed in the 1970s and 1980s as part of state-mandated evaluation systems, and most focusing on observable teacher behavior in classrooms. After criticisms that these instruments tended to focus on teacher behavior rather than student learning, a new generation of instruments is being developed, primarily in the USA, designed to capture the connections between teaching and learning. Although these promise significant improvement and conceptual sophistication over earlier approaches, they remain system specific, and it is an open question whether they could be transferred to other systems, without cultural distortion.

Evaluation and Professional Autonomy

The research shows that the main variation across systems is the extent to which teacher evaluation is controlled by the state and imposed in a standard way upon all teachers or is subject to a degree of professional autonomy at the level of the individual teacher or school. The origin of these differences lies in the political culture of countries. The clearest example in the literature is a study of teacher evaluation in China and Hong Kong after the latter became a Special Administrative Region (SAR) of China in 1997 (Lee *et al.*, 2003). This showed that in Hong Kong SAR each school was expected to design its own staff appraisal system, whereas in the People's Republic of China each teacher was evaluated according to a national model, comprising four domains with agreed maximum scores: morality (10), diligence (5), abilities (45), and student performance (40).

The highest degree of school autonomy in teacher evaluation is reported to operate in the Netherlands, where there is a strong political commitment to decentralization of schooling. Teacher evaluation is usually devolved to the school principals, whose responsibility is to develop and implement an appraisal system according to their own priorities. Although annual appraisal discussions between principals and individual teachers occur in many schools, not all are based on classroom observations. Therefore, there is little standardization in practice and Dutch analysts report that teacher evaluation is mainly formative, is not linked to salary or promotion, and contributes little to systematic teacher improvement. Although there is an inspectorate in the Netherlands, its prime focus is on the quality of schools rather than individual teachers (Reezigt *et al.*, 2003). It is ironic that, alongside this *laissez faire* approach to evaluation, academics in the Netherlands have made one of the most distinguished contributions to research into the measurement of educational effectiveness.

The most-centralized, comprehensive, and penetrative model of teacher evaluation was developed in England through the government's Office for Standards in Education (OFSTED) set up in 1992. The model had statutory force. All publicly maintained schools were inspected every 4 years, on a protocol which required each teacher to be observed teaching up to four lessons by trained inspectors, and each lesson to be graded on a seven-point scale, with the last two grades being considered as unsatisfactory. The evaluation process also involved interviews with students and analysis of students' assignments. The inspectors' grades were made available to the teacher, the governing body of the school, and were entered onto a government-controlled database. The OFSTED methodology gave very strong emphasis to direct observation of teaching and learning in classrooms, with observation schedules that resonate with the teacher effectiveness research in the USA in the 1980s. The report on the quality of teaching

and learning in the school by the inspectors was made public and schools were ranked on league tables, to provide comparative performance data for parents and others to use for school choice. Originally restricted to maintained primary and secondary schools, the inspections have been extended to further-education colleges, teacher-training providers, local authorities, nursery and other preschool settings, residential care homes, private schools, and specialist education agencies.

As the system is high stakes, it was criticized for causing stress and lowered morale among teachers, leading to substantial numbers of teachers, including those rated highly, leaving the profession. However, among most parents the system was widely supported, and as experience with this evaluation system grew, there was a shift to more school self-evaluation, especially for those teachers and schools which the series of external inspections had shown to be consistently good.

At the same time, the quality of teaching in universities was subject to evaluation by a national agency, the Quality Assurance Agency (QAA), with quality ratings being awarded out of a maximum of 24. Although universities whose departments scored highly under this system were pleased to publicize the fact, the system has been criticized as being overly bureaucratic, with the QAA spending undue effort on reading documentation, and interviewing staff, but relatively little time observing teaching. It was also seen as undermining the long-standing tradition in universities, enshrined in their regulations, of academic autonomy. Despite this, the QAA system had greater validity for evaluating university teaching than the common practice of relying on anonymous student evaluations, which have been found to reflect tutors' popularity, personality, and the ease of high-grade achievement, rather than teaching quality.

Professional Standards

The development of formally articulated professional standards has provided teachers and their evaluators with published criteria by which teachers can be assessed. In most modernizing societies, these take the form of hierarchical levels, each level representing improved professionalism. A widespread concern has been that these kinds of standard are technicist reflecting a lack of interest in professional values, judgment, and experimentation, thereby promoting a view of teaching which is, at best, semiprofessional and, at worst, robotic. There is also the argument that such approaches do not reflect the complexity of modernized professions (Scriven, 1988). The dissonance in relation to teacher evaluation implied in technicist models, and others stressing degrees of teacher autonomy, is substantial.

In the USA, it has been argued that historically the focus of competency standards has changed from articulating the moral qualities of teachers, through licensing people to teach, through classroom teaching skills, to, recently, students' learning. An example of the emphasis on learning is reflected in the development of the National Board for Professional Teaching Standards, which in 2009 was basing its standards on five core beliefs, rather than a set of classroom behaviors. These were: commitment to student learning; knowledge of their subjects and how to teach them; responsibility for managing and monitoring student learning; systematic thinking about their practice and learning from experience; and membership of learning communities.

In England, nationally applicable standards were developed by a government agency, the Teacher Training Agency, later the Training Development Agency for Schools (TDA). By 2008, these were designed to assess competency at five career stages: initial training, core standards after the induction year, post-threshold standards, excellent teachers, and advanced skills teachers. Each set of standards comprised three elements: professional attributes; professional knowledge and understanding; and professional skills. The intention was to evaluate and support the progression of teachers holistically through their career, using increasingly complex criteria building on achievement in preceding stages. The emphasis is, like in the USA, upon evidence about teachers' impact on their students' learning and their own professional development, and has the advantage of specifying in considerable detail what a teacher needs to do to obtain a successful evaluation.

The developments in both countries resonate with emerging concerns in Australia, where it has been argued that the practice of evaluating teaching by econometric methods, which make connections between teacher qualifications, or teachers' classroom behavior and student test outcomes is simplistic and methodologically flawed. What is proposed as improvement was an evaluation approach with two elements: building teacher capacity and the articulation of professional standards emphasizing learning (Ingvarson and Rowe, 2008).

A problematic aspect of these new-generation evaluation schemes, despite their significant advances on preceding attempts, is that it remains very difficult, technically, to disaggregate the value added to student learning by individual teachers from that added by school factors (Reynolds *et al.*, 2003).

Toward a Typology of Teacher Evaluation

Overall from this wide range of research, it is possible to generate a tentative threefold typology of teacher evaluation, based primarily on differing conceptions of teaching quality. Although the classifications are not pure, it can be

hypothesized that they reflect historical changes, with the first type being found in systems where there has been only limited change toward modernization, and the last being found in systems that have begun to implement modernized education systems, reflecting the complexity of teaching in highly differentiated contexts.

Teaching as a craft, in which evaluation is carried out by connoisseurship, with inspectors or principals exercising their judgment based on professional experience. The accountability is top down and runs the risk of not being open to challenge. It is exclusively external evaluation.

Teaching as a science, in which evaluation is conducted on the basis of scientifically established teaching behaviors and their contribution to standardized test scores. The accountability is technicist, and uniform, and open to challenge only if the procedures adopted can be shown to be erroneous. It is largely external evaluation.

Teaching as a science of an art, in which evaluation is conducted on the basis of its contribution to student learning and linked to teachers' professional improvement, as judged by the teacher. The accountability is open to individual interpretation, personalized, and seen as focused on the student. It is primarily self-evaluation.

Problematic Issues

Despite the progress made, there remain seven unresolved and problematic issues in teacher evaluation. These are:

1. to improve validity and reliability, especially in high-stakes models, where the evaluations carry career implications, and also to highlight the contested nature of teaching quality;
2. to take account of context specificity through developing further a differentiated model of teacher evaluation without undue complexity;
3. to broaden the range of performance indicators so as to incorporate other educational objectives in addition to narrow cognitive ones, such as character development, problem solving, attitudes to learning, meta-cognitive attributes, and self-esteem;
4. to demonstrate that external evaluation can lead to sustainable teacher improvement;
5. to overcome the democratic deficit implicit in almost all evaluation procedures to incorporate recognition of different educational values and purposes held by individual teachers, as against the dominance of organizational purposes;
6. to gain recognition from policymakers and researchers of the absolute impossibility of devising appropriate evaluation schemes to take into account, and value, the personal power of individual teachers, that is,

“the personal flair, instinct or intuition. . . those qualities often referred to as gifts” (Robinson, 2004); and

7. to resolve the tensions between personal/professional development and high performativity. The dominant drivers for much current teacher evaluation tend to be associated with high performativity and run counter to models of reflective self-evaluation and professional development.

Priorities for Future Research

There are three categories of research where further investigation is particularly required. The three are different in character and substance:

1. Conceptual and philosophical investigations, including the educational values reflected in, or absent from, current models of teacher evaluation, and the basis for a differentiated model.
2. Sociological and political investigations exploring the social functions, intended and unintended, of teacher evaluation on professional practice, and the power relations between government, state, regional bodies, and teachers and their associations in developing, implementing, and revising teacher evaluation procedures (Stronge and Tucker, 2000), and the assumptions made about accountability.
3. Technical investigations, attempting to develop improved reliability and validity in instruments and procedures (including self-evaluation methods), to identify the scope and limitations of evaluation data for judging teachers and for predicting future performance (Goldstein, 2001), and to contribute to the development of comparative studies within and across education systems.

See also: Accreditation and Standards in Teacher Education; Research for Leaders: The National College for School Leadership; School Inspection/External School Evaluation; The International system for Teacher Observation and Feedback: A Theoretical Framework for Developing International Instruments.

Bibliography

- Apple, M. (1986). *Teachers and Text*. New York: Routledge and Kegan Paul.
- Campbell, R. J., Kyriakides, L., Muijs, R. D., and Robinson, W. (2004). *Assessing Teacher Effectiveness: Developing a Differentiated Model*. London: RoutledgeFalmer.
- Goldstein, H. (2001). Using pupil performance data for judging schools and teachers: scope and limitations. *British Journal of Educational Research* 27(4), 433–442.

- Ingvarson, L. C. and Rowe, K. (2008). Conceptualising and measuring teacher quality: Substantive and methodological issues. *Australian Journal of Education* 52(1), 5–34.
- International Labour Office (ILO) (1991). *Teachers: Challenges of the 1990s: Second Joint Meeting of Conditions of Work of Teachers* (1991). Geneva: International Labour Office.
- Lee, J. C. K., Lam, W. P., and Li, Y. Y. (2003). Teacher evaluation and effectiveness in Hong Kong: Issues and challenges. *Journal of Personnel Evaluation in Education* 17(1), 41–66.
- Reezigt, G. J., Creemers, B. P. M., and de Jong, R. (2003). Teacher evaluation in the Netherlands and its relationship to educational effectiveness research. *Journal of Personnel Evaluation in Education* 17(1), 67–82.
- Reynolds, D., Muijs, D., and Treharne, D. (2003). Teacher evaluation and teacher effectiveness in the United Kingdom. *Journal of Personnel Evaluation in Education* 17(1), 83–100.
- Robinson, W. (2004). *Power to Teach: Learning through Practice*. London: RoutledgeFalmer.
- Scriven, M. (1988). Duty-based teacher evaluation. *Journal of Personnel Evaluation in Education* 1(4), 319–334.
- Stronge, J. H. and Tucker, P. D. (2000). The politics of teacher evaluation: A case study of new system design and implementation. *Journal of Personnel Evaluation in Education* 13(4), 339–359.
- Teddlie, C., Stringfield, S., and Burdett, J. (2003). Comparisons of the relationships among educational effectiveness, evaluation and improvement variables: An overview. *Journal of Personnel Evaluation in Education* 17(1), 5–20.
- Toch, T. and Rothman, R. (2008). *Rush to Judgement: Teacher Evaluation in Public Education*. Washington, DC: Education Sector.

Further Reading

- Campbell, R. J. and Neill, S. R., St. J. (1994). *Primary Teachers at Work*. London: Routledge.
- Danielson, C. and McGreal, T. (2000). *Teacher Evaluation to Enhance Professional Practice*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Etzioni, A. (1969). *The Semi-Professions and Their Organisation*. New York: Free Press.
- Gray, J., Goldstein, H., and Thomas, S. (2001). Predicting the future: The role of past performance in determining trends in institutional effectiveness at A Level. *British Journal of Educational Research* 27(4), 391–406.
- MacBeath, J. (1999). *Schools Must Speak for Themselves: The Case for School Self-Evaluation*. London: Routledge.
- MacBeath, J., Schratz, M., Jakobsen, L., and Meuret, D. (2000). *Self-Evaluation in European Schools*. London: RoutledgeFalmer.
- Millman, J. and Darling-Hammond, L. (eds.) (1990). *The New Handbook of Teacher Evaluation*. Newbury Park, CA: Sage.
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D., and Ecob, R. (1988). *School Matters*. Wells: Open Books.
- OECD (2005). *Attracting, Developing and Retaining Effective Teachers – Final Report: Teachers Matter*. Paris: OECD/CERI.
- Peterson, K. D. (2000). *Teacher Evaluation: A Comprehensive Guide to New Directions or Practices*, 2nd edn. Thousand Oaks, CA: Corwin Press.
- Robinson, W. (2003). *Pupil Teachers and Their Professional Training in Pupil-Teacher Centres in England and Wales, 1870–1914*. New York: Edwin Mellen Press.
- Robinson, W. (2004). *Power to Teach: Past and Present Perspectives on Learning through Practice*. London: RoutledgeFalmer.
- Robinson, W. and Campbell, R. J. (2009). *School Ethos and Effective Teaching: Case Studies in the Education of Gifted and Talented Students*. London and New York: RoutledgeFalmer.
- Stronge, J. H. (ed.) (1997). *Evaluating Teaching: A Guide to Current Thinking and Practice*. Thousand Oaks, CA: Corwin Press.
- Wittrock, M. C. (ed.) (1986). *Handbook of Research on Teaching*. New York: Macmillan.

Relevant Websites

<http://www.nbpts.org> – National Board for Professional Teaching Standards.
<http://www.ncsl.org.gov> – National College for School Leadership.

<http://www.teacherevaluation.net> – Teacher Evaluation: New Directions and Practices hosted by Kenneth D. Peterson.
<http://www.tda.gov.uk> – Training Development Agency for Schools.
<http://www.oecd.org> – OECD.
<http://www.ofsted.gov.uk> – OFSTED.
<http://www.qaa.gov.uk> – Quality Assurance Agency.

International Evaluations

A Schleicher, OECD, Paris, France

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Parents, students, and those who teach and run education systems seek good information on how well their education systems prepare students for life. Many countries now monitor students' learning in order to provide answers to this question: for example, among 36 countries with comparable data – including the 30 members of the Organization for Economic Cooperation and Development (OECD) – 22 countries undertake student examinations and/or assessments and 17 require schools to be evaluated (either self-evaluations and/or inspections by an external body) at regular intervals. For student-performance measures, student assessments (evaluations without direct consequences for an individual student's further educational career) are used in 17 countries, whereas national examinations are used in ten OECD countries (OECD, 2008).

Comparative international evaluations can extend and enrich the national picture by providing a larger context within which to interpret national performance. They have gained prominence – over recent years – since the benchmarks for public policy in education are no longer solely national goals or standards, but, increasingly, the performance of the most successful education systems internationally. International evaluations can provide countries with information that allows them to identify areas of relative strengths and weaknesses, and monitor the pace of progress of their educational system. They can also stimulate countries to raise aspirations by showing what is possible in education, in terms of the quality, equity, and efficiency of educational services provided in other countries or regions, and they can foster better understanding of how different education systems address similar policy challenges.

Following a brief introduction to the history of international evaluations, this article sets out the potential that international evaluations offer for educational policy and practice as well as some of the methodological challenges they face in providing valid, comparable, and reliable evidence.

History of International Evaluations

While efforts to compare education systems internationally can be traced back to the early nineteenth century (e.g., Jullien, 1817), the discourse on international comparisons of learning outcomes started to emerge during the 1950s

and 1960s. In 1958, an expert group led by William Douglas Wall and including prominent researchers, such as Benjamin Bloom, Robert Thorndike, Arthur Wellesley Forshay, Arnold Anderson, Gaston Mialaret, and Torsten Husen, met under the auspices of United Nations Scientific, Educational, and Cultural Organization (UNESCO)'s International Institute of Education, in Hamburg, to launch a feasibility study to compare student performance internationally. The feasibility study involved 12 000 13-year-olds in 12 countries and its results were published in 1962 (Forshay *et al.*, 1962). The International Association for the Evaluation of Educational Achievement (IEA) emerged out of this collaboration which then conducted a series of international evaluations (see Siniscalco, 2007). The most prominent regular surveys conducted by the IEA are now the 4-yearly Trends in Mathematics and Science Study (TIMSS) and the 5-yearly Progress in Reading Literacy Study (PIRLS).

In 1988, the Education Testing Service in the United States conducted the International Assessment of Educational Progress (IAEP) (Lapointe *et al.*, 1989) and a follow-up study in 1991 (Lapointe *et al.*, 1992).

The latest generation of international evaluations has been developed by the OECD as part of the Program for International Student Assessment (PISA). PISA surveys have been conducted every 3 years, since 2000, in key content areas – such as reading, mathematics, and science – but they also cover cross-curricular domains, such as problem solving, and digital reading, as well as a range of non-cognitive outcomes. The latest PISA survey, in 2006, tested over 400 000 students who represented more than 20 million 15-year-olds in 60 countries that made up close to 90% of the world economy.

Research Frameworks of International Evaluations

The international evaluations of the IEA and OECD contextualize measures of student-learning outcomes with background information collected from students, school principals, and, sometimes, teachers and parents in order to interpret the observed variation in learning outcomes between students, classrooms, schools, and education systems. To facilitate this, they operate with research frameworks that provide data at up to four levels of the education system, namely: (1) the education system as a whole; (2) the educational institutions and providers of educational services; (3) the classrooms or instructional setting; and (4) the learners themselves (see **Table 1**).

Table 1 Research frameworks for international assessments

	<i>Education and learning outputs and outcomes</i>	<i>Policy levers and contexts shaping educational outcomes</i>	<i>Antecedents or constraints that contextualize policy</i>
Individual participants in education and learning	The quality and distribution of individual educational outcomes	Individual attitudes, engagement, and behavior	Background characteristics of individual learners
Instructional settings	The quality of instructional delivery	Curriculum, pedagogy, and learning practices and classroom climate	Student learning conditions and teacher working conditions
Providers of educational services	The output of educational institutions and institutional performance	School environment and organization	Characteristics of service providers and their communities
The educational system as a whole	The overall performance of the educational system	System-wide institutional settings, resource allocations, and policies	The national educational, social, economic, and demographic contexts

The research frameworks of the international evaluations typically address different classes of research issues: (1) A first class relates to simple comparisons of learning outcomes at each of the four levels. (2) A second class provides information on the policy levers or circumstances which shape the outputs and outcomes at each level of the education system. This includes, for example, measures of attitudes and behaviors at the level of students, measures of student learning and teacher working conditions and human and material resources at the level of instructional settings and institutions, and measures of structures and resource allocation policies and practices at the level of the educational system. These policy levers and contexts, typically, have antecedents – factors that define or constrain policy, and which are represented in a third class of research areas. (3) This third class which, for example, provides information on the socioeconomic context of students, schools, or systems, is particularly important in an international comparative context, as it allows to ensure like-with-like comparisons, that is, comparisons of schools that have a similar socioeconomic intake or countries that operate under similar socioeconomic conditions. Each of the cells resulting from cross-classifying the above two dimensions can then be used to address a variety of research issues from a comparative perspective relating, for example, to the quality of educational outcomes and educational provision; to issues of equality of educational outcomes and equity in educational opportunities; or to the adequacy, effectiveness, and efficiency of resource management.

The Potential of International Evaluations for Policy and Practice

The design and conduct of international evaluations was, originally, motivated by research objectives. More recently, governments too have begun to attribute growing importance to international evaluations and have invested

considerable resources into their development and implementation. This interest derives from several considerations:

- By revealing what is possible in education in terms of the performance levels demonstrated in the countries with the highest performance levels, international evaluations can enhance the quality of existing policies but also create a debate about the paradigms and beliefs underlying policies.
- While international evaluations alone cannot identify cause-and-effect relationships among inputs, processes, and educational outcomes, they can shed light on key features in which educational systems show similarities and differences, and make those key features visible for educators, policymakers, and the general public.
- In some countries, international evaluations are also used to set policy targets in terms of measurable goals achieved by other systems, and seek to identify policy levers and establish trajectories as well as delivery chains for reform.
- International evaluations can assist with gauging the pace of educational progress, through assessing to what extent achievement gains observed nationally are in line with the achievement gains observed elsewhere.
- Last, but not the least, international evaluations can support the political economy of educational reform, which is a major issue in education where any payoff to reform almost inevitably accrues to successive governments, if not generations.

These issues are examined more closely in the remainder of this section.

Revealing What Is Possible in Education and Identifying Factors that Contribute to Educational Success

The level of public awareness raised by comparisons has, in some countries, created an important political momentum and engaged educational stakeholders – including

teacher or employer organizations – in support of policy reform (Hopkins, 2008). International evaluations have, sometimes, had a significant impact in countries that found themselves confronted with results that differed from how educational performance was generally perceived. In Germany, for example, equity in learning opportunities across schools had often been taken for granted, as significant efforts were devoted to ensuring that schools are adequately and equitably resourced. The results from the PISA 2000 evaluation, however, revealed large socioeconomic disparities in educational outcomes between schools. Further analyses that separated equity-related issues between those that relate to the socioeconomic heterogeneity within schools and those that relate to socioeconomic segregation through the school system suggested that German students from more privileged social backgrounds are directed into the more prestigious academic schools which yield superior educational outcomes whereas students from less privileged social backgrounds are directed into less prestigious vocational schools which yield poorer educational outcomes, even where their performance on the PISA evaluation was similar. This raised the specter that the German educational system was reinforcing – rather than moderating – socioeconomic background factors. Such results – and the ensuing vivid public debate – inspired a wide range of equity-related reform efforts in Germany, some of which have been transformational in nature. These include giving early-childhood education, that had hitherto been considered largely an aspect of social welfare, an educational orientation and better institutionalizing early-childhood provision; reducing the degree of tracking in some of the German states; establishing national educational standards for schools in a country where regional and local autonomy had long been the overriding paradigm; introducing full-day schooling in a system where half-day schooling had been the norm for centuries; or enhancing the support for disadvantaged students, such as students with a migration background.

For many educators and experts in Germany, the socioeconomic disparities that PISA revealed were unsurprising. However, it was, often, taken for granted and outside the scope of public policy that disadvantaged children would fare less well in school. The fact that PISA revealed that the impact which the socioeconomic background has on students and school performance varied considerably across countries – and that other countries appeared to moderate socioeconomic disparities much more effectively – showed that improvement was possible and provided the momentum for policy change.

Showing that strong educational performance and improvement is possible seems to be one of the most important aspects of international evaluations. Whether in Asia (e.g., Japan, Korea, and Singapore), in Europe (e.g., Finland, and the Netherlands), or in North America (e.g., Canada),

many countries display strong overall performance in the international evaluations conducted by the IEA and OECD and, equally important, some of these countries also show that poor performance in school does not automatically follow from a disadvantaged socioeconomic background. Last, but not the least, some countries show that success can become a consistent and predictable educational outcome: In Finland, for example – the country with the strongest overall results in PISA – the performance variation between schools amounted, in 2006, to only 5% of students' overall performance variation. So, parents can rely on high and consistent performance standards in whatever school they choose to enroll their children. Considerable research has been invested into the features of these educational systems. In some countries, governments have used knowledge provided by PISA as a starting point for a peer review to study policies and practices in countries operating under similar contexts that achieve better results (Döbert *et al.*, 2004). Such peer reviews – each resulting in a set of specific policy recommendations for educational improvement – are also being carried out by the OECD, the results of which have been published so far for Denmark and Scotland (OECD, 2005, 2007).

As a result, the benchmarks for public policy in education are no longer national goals or standards alone, but, increasingly, the performance and achievement gains of the most successful education systems internationally.

Putting National Targets into a Broader Perspective

International evaluations can also play an important role in putting national performance targets into perspective. Often, educators are faced with a dilemma: if, at the national level, the percentage of students achieving good exam scores in school increases, some will claim that the school system has improved; others will claim that standards must have been lowered, and behind the suspicion that better results reflect lowered standards is, often, a belief that overall performance in education cannot be raised. International evaluations allow those perceptions to be related to a wider reference framework, by allowing schools and educational systems to compare themselves with schools and educational systems in other countries. Some countries have actively embraced this perspective and systematically related national performance to international evaluations, for example, by embedding the components of the PISA or TIMSS evaluations into their national evaluations.

Assessing the Pace of Change in Educational Improvement

A third important aspect is that international comparisons provide a frame of reference to assess the pace of change

in educational development. While a national framework allows progress to be assessed in absolute terms, an internationally comparative perspective allows an assessment of whether that progress matches the pace of change observed elsewhere. Indeed, while all educational systems in the OECD area have seen quantitative growth over past decades, international comparisons reveal that the pace of change in educational output has varied markedly.

A Tool for the Political Economy of Reform

International evaluations can also support the political economy of reform. For example, in the 2007 Mexican national survey of parents, 77% of parents interviewed reported that the quality of educational services provided by their children's school was good or very good. However, in OECD's PISA 2006 evaluation, roughly half of the Mexican 15-year-olds who are enrolled in school performed at or below the lowest level of proficiency established by PISA (IFIE-ALDUCIN, 2007; OECD, 2007). There may be many reasons for such a discrepancy between perceived educational quality and performance on international evaluations. For example, in part, this may be due to the fact that the educational services which Mexican children receive are significantly enhanced over the quality of schooling that their parents experienced. However, the point here is that justifying the investment of public resources into areas for which there appears to be no public demand poses difficult challenges for the political economy of reform. One recent response by the Mexican presidential office has been to include a PISA performance target into the new Mexican reform plan. This performance target that is based on the outcome of an international evaluation and that is to be achieved by 2012, will serve to highlight the gap between national performance and international standards, and monitor how educational improvement feeds into closing this gap. It is associated with a reform trajectory and delivery chain of support systems, incentive structures, as well as with improved access to professional development to assist school leaders and teachers in meeting the target. Such reforms draw on the experience of other countries. Brazil has taken a similar route – providing each secondary school with information on the level of progress that is required to perform at the OECD average performance level on PISA, in 2021.

Japan is one of the best-performing educational systems on the various international evaluations. However, PISA revealed that while students tended to do very well on tasks that require reproducing subject-matter content, they did much less well on open-ended constructed tasks requiring them to demonstrate their capacity to extrapolate from what they know and apply their knowledge in novel settings. Conveying that to parents and a general public who are used to certain types of tests, poses a challenge for the

political economy of reform too. One policy response in Japan has been to incorporate PISA-type open-constructed tasks into the national evaluation, with the aim that skills that are considered important become valued in the educational system. Similarly, Korea has recently incorporated advanced PISA-type literacy tasks in its university entrance examinations, in order to enhance excellence in the capacity of its students to access, manage, integrate, and evaluate written material. In both countries, these changes represent transformational change that would have been much harder to imagine without the challenges revealed by PISA.

Design of International Evaluations

The design of international evaluations of learning outcomes needs to fulfill different, and, sometimes, competing demands:

- First, international evaluations need to ensure that their outcomes are valid across cultural, national, and linguistic boundaries and that the target populations from which the samples in the participating countries are drawn, are comparable.
- Second, they need to offer added value to what can be accomplished through national evaluation and analysis.
- Third, while results from international evaluations need to be as comparable as possible, they also need to be country specific so as to adequately capture historical, systemic, and cultural variation among countries.

Some of the design issues involved in meeting and balancing these various demands are laid out in the remainder of this section.

Cross-Country Validity and Comparability in the Evaluation Instruments

International evaluations necessarily are limited in their scope. This is because:

- There is no overarching agreement on what fundamental competencies students in a particular grade or at a particular age should possess.
- Any evaluation can only capture a selection of competencies.
- Various methodological constraints limit the nature of competencies that are currently amenable to large-scale evaluation.

International evaluations have made considerable progress toward assessing knowledge and skills in content areas such as mathematics, reading, science, and problem solving. However, they are still limited in the coverage of important cognitive outcomes – in particular, the

assessment of creative competencies. Similarly, the tension between achieving high degrees of objectivity in the evaluations – which tend to favor multiple-choice tasks that can be scored without human judgment, on the one hand – tends to detract from the evaluation of the higher-order competencies and the production of knowledge – which require open-ended evaluation tasks, on the other. At times, international evaluations have also sacrificed validity gains over efficiency gains by giving undue weight to assessment tasks that can be easily administered and scored, to make the evaluations affordable. Even less progress has been made to assess interpersonal dimensions of competencies which are, often, recognized as of increasing importance, such as the capacity of students to relate well to others, to manage and resolve conflicts, or to respect and appreciate different values, beliefs, or cultures. Last but not least, international evaluations provide only very crude self-reported measures of intrapersonal dimensions of competencies.

Even in established content areas, comparative evaluation poses major challenges. Countries vary widely in their intended, implemented, and achieved curricula. Inevitably, international evaluations need to strike a balance between narrowing the focus to what is common across the different curricula of school systems, on the one hand, and capturing a wide enough range of competencies to reflect the content domains to be assessed adequately, on the other. Leaning toward the former – as has been the tendency for the international evaluations of the IEA – ensures that what is being tested internationally reflects what is being taught nationally. At the same time, such an approach drives assessments toward the lowest common denominator of national curricula – at the expense of important aspects of curricula that are not taught in all countries. Leaning toward the latter – as is the case for the evaluations of the OECD with their focus on the capacity of students not merely to reproduce what they have learned but to extrapolate from what they have learned and apply their knowledge and skills in novel settings – enhances content validity but risks that students are being confronted with evaluation material they may not have been taught.

In whatever way the various international evaluations have struck these balances, they have tried to build them through a carefully designed interactive process between the agencies developing the assessment instruments, various international expert groups working under the auspices of the respective organizations, and national experts charged with the development and implementation of the surveys in their countries. Often, a panel of international experts led – in close consultation with participating countries – the identification of the range of knowledge and skills in the respective assessment domains that were considered to be crucial for student's capacity to fully participate in and contribute to a successful modern

society. A description of the assessment domains – the assessment framework – was then used by participating countries, and other test development professionals, as they contributed assessment materials, which typically involved:

- The development of a working definition for the assessment area and description of the assumptions that underlay that definition.
- An evaluation of how to organize the set of tasks constructed in order to report to policymakers and researchers on performance in each assessment area among 15-year-old students in participating countries.
- The identification of a set of key characteristics to be taken into account when assessment tasks were constructed for international use.
- The operationalization of the set of key characteristics to be used in test construction, with definitions based on existing literature and the experience of other large-scale evaluations.
- The validation of the variables, and assessment of the contribution which each made to the understanding of task difficulty in participating countries.
- The preparation of an interpretative scheme for the results.

Once the assessment framework is established and agreed – which tends to be the most challenging aspect of an international evaluation – assessment items are developed to reflect the intentions of the frameworks and they need to be carefully piloted prior to the establishment of final assessment instruments. To some extent, the question to what extent the tasks in international evaluations are comparable across countries can be answered empirically. Analyses to this end were first undertaken for the IEA TIMSS (Beaton *et al.*, 1996). The authors compared the percentage of correct answers in each country according to the international evaluation, as a whole, with the percentage correct in each country on the items reported by the country to address its curriculum in mathematics. Singapore, for example, had 144 out of 162 items that were said to be covered by the Singaporean curriculum. The percentage of items correct on the whole test and on the items covered in the curriculum was 79 in both cases. Singapore also scored between 79% and 81% correct on the items that other countries considered as covered in their own curricula. These ranged from 76 items in Greece to 162 items in the United States. For most countries, the results were similarly consistent – suggesting that the composition of the tests had no major impact on the relative standing of countries in the international comparisons. Such analyses have also been conducted for PISA, and yielded similar results.

International evaluations pay close attention to reflecting the national, cultural, and linguistic variety among participating countries. OECD's PISA evaluations employ

the most sophisticated and rigorous process to this end. The agency charged with the development of the instruments uses professional test-item-development teams in several different countries. In addition to the items developed by these teams, assessment material is contributed by participating countries that is carefully evaluated and matched against the framework. Furthermore, each item included in the assessment pool is rated by each country: (1) for potential cultural, gender, or other bias; (2) for relevance to the students to be assessed in school and nonschool contexts; and (3) for familiarity and level of interest.

Another important aspect concerns the nature and form of the evaluation – as reflected in the task and item types. While, as noted before, multiple-choice tasks are the most cost-effective way to assess knowledge and skills, and have, therefore, dominated earlier international evaluations, they have important limitations in assessing more complex skills – particularly, ones that require students not just to recall but to produce knowledge. Moreover, since the nature of assessment tasks – and, in particular, student familiarity with multiple-choice tasks – varies considerably across countries, heavy reliance on any single item type, such as multiple-choice tasks, can be an important source of response bias. The PISA evaluations have tried to address this through employing a broad range of assessment tasks, with about 40% of the questions requiring students to construct their own responses. These tasks require students to either provide a brief answer (short-response questions) or to construct a longer response (open-constructed-response questions) – allowing for the possibility of divergent individual responses and an assessment of students' justification of their viewpoints. Partial credit can be given for partly correct or less complex answers, with questions assessed by trained specialists using detailed scoring guides which gave direction on the codes to assign to various responses. Open-ended assessment tasks, however, raise other challenges – in particular, the need to ensure inter-rater reliability in the results. For PISA, subsamples of the assessment booklets are coded independently by four coders and examined by the international contractor. In order to examine the consistency of this coding process in more detail within each country and to estimate the magnitude of the variance components associated with the use of coders, an intercoder reliability study on the sub-sample of assessment booklets is being applied and homogeneity analysis is applied to the national sets of multiple coding. Similarly – at the between-country level – an international coding review is implemented to check on the consistency of application of response coding standards across all participating countries, with the objective to estimate potential bias (either leniency or harshness) in the coding standards applied in participating countries.

In order to cover the intended broad range of content while meeting the limits of individual assessment time, assessments, such as PISA or TIMSS, are now using multiple forms which are administered to students in a matrix design, with sufficient overlap between the forms to relate their results psychometrically.

Ensuring that international evaluations are comparable across countries is one thing, but the more important challenges actually relate to their external validity – which involves verifying that the assessments measure what they set out to measure. An important question is whether the knowledge and skills that are being assessed are predictive for the future success of students. In the case of PISA, the Canadian Youth in Transition Survey (YITS) – a longitudinal survey which investigates patterns of and influences on major educational, training, and work transitions in young people's lives – provided a way to examine this empirically. In 2000, 29 330 15-year-old students in Canada participated both in YITS and PISA. Four years later, the educational outcomes of the same students – then aged 19 – were assessed and the association of these outcomes with PISA reading performance at age 15 was investigated. A strong relationship between performance at age 15 and subsequent success in access to higher education and successful labor-market transitions, even after accounting for the socioeconomic background (Knighton and Bussière, 2006), was revealed.

Comparability of the Target Populations

Even if the assessment instruments are valid and reliable, meaningful comparisons can only be made if the target populations being assessed are also comparable. International evaluations, therefore, need to use great care when: (1) defining comparable target populations; (2) ensuring that they are exhaustively covered with minimal and well-defined population exclusions; and (3) ensuring that the sampled students do participate in the evaluation.

With regard to defining target populations, important tradeoffs need to be made between international comparability, on the one hand, and relating the target populations to national institutional structures, on the other. Differences between countries in the nature and extent of pre-primary education and care, the age of entry to formal schooling, and the institutional structure of educational systems, do not allow the establishment of internationally comparable grade levels of schooling. Consequently, international comparisons of educational performance, typically, define their populations with reference to a target age group. International evaluations of the IEA have defined these target groups on the basis of the grade level that provides maximum coverage of a particular age cohort (such as the grade in which most 13-year-olds are enrolled). This gives an advantage that a grade level can be easily interpreted within the national institutional structure and provides a cost-effective way toward assessment – with

Table 2 International assessments conducted by the IEA

	<i>Study</i>	<i>Data collection</i>	<i>Target population</i>	<i>Coverage</i>
60	Pilot Study	1960	13 years	Mathematics, science, reading, geography, nonverbal abilities
	First International Mathematics Study, FIMS	1964	13 years and last year of upper secondary education	Mathematics
70	The Six Subject Survey	1970–71	10 years, 14 years, and last year of upper secondary education	Reading comprehension Literature First International Science Study English as a Foreign Language French as a Foreign Language Civic Education
80	Second International Mathematics Study – SIMS	1980–82	13 years and last year of upper secondary education	
	Classroom Environment Study	1982–83	9 and 15 years	Instructional methods in mathematics, science, and history
	Second International Science Study – SISS	1983–84	10 years, 14 years, and last year of upper secondary education	Science
	Written Composition	1985	10 years, 14–16 years, and last year of upper secondary education	Writing
	Computers in Education – COMPED	1989 e 1992	10 and 13 years	Availability and use of computers and technology
	Preprimary project	1986–94 1989–2003 1993–2003	Longitudinal study following children from 4 to 7 years	Quality of early childhood provision
90	Reading Literacy Study – RLS	1990–91	9 and 14 years	Reading comprehension
	Third International Mathematics and Science Study – TIMSS 1995	1994–95	9 and 13 years and last year of upper secondary education	Mathematics and science
	Civic Education Study – CIVED	1996–97 1999–2000	14 and 16–18 years	Civic education
	Second Information Technology in Education Study Module 1 – SITES-M1	1998–99		Availability and use of technology
	Third International Mathematics and Science Study Repeat – TIMSS-R 1999	1998–99	9 and 13 years and last year of upper secondary education	Mathematics and science
2000	Progress in International Reading Literacy Study 2001 – PIRLS	2001	9 years	Reading comprehension
	Trends in Mathematics and Science Study 2003 – TIMSS 2003	2002–03	9 and 14 years	Mathematics and science
	(in progress) Progress in International Reading Literacy Study 2006 – PIRLS 2006	2005–06	9 years	Reading comprehension
	(in progress) Second Information on Technology in Education Study 2006 – SITES 2006	2006		Availability and use of technology
	(in progress) Teacher Education and Development Study – Mathematics TEDS-M 2008	2007–08		Teacher training
	(in progress) Trends in International Mathematics and Science Study 2007 – TIMSS 2007	2006–07	9 and 13 years	Mathematics and science
	(in progress) TIMSS Advanced 2008	2007–08	Last year of upper secondary education	Mathematics and physics
	(in progress) International Civic and Citizenship Education Study – ICCS 2009	2008–09	13 years	Civic education

From Siniscalco (2007).

minimal disruption of the school day when administering the assessment to students of the same grade. However, a disadvantage is that slight variations in the age distribution of students across grade levels, often, lead to the selection of different target grades in different countries – or between educational systems within countries – raising serious questions about the comparability of results across – and, at times, within – countries. In addition – because not all students of the desired age are usually represented in grade-based samples – there may be a more serious potential bias in the results if the unrepresented students are typically enrolled in the next higher grade in some countries and the next lower grade in others. This excludes students with potentially higher levels of performance in the former countries and students with potentially lower levels of performance in the latter. To address these problems, the evaluations of the OECD use an age-based definition for their target populations, that is, a definition that is not tied to the institutional structures of national educational systems. For example, PISA assesses students who were in the age range of 15 years and 3 (complete) months and 16 years and 2 (complete) months at the beginning of the assessment period and who were enrolled in an educational institution, irrespective of the grade levels or type of institution in which they were enrolled, and also irrespective of whether they were in full-time or part-time education. The disadvantage of this age-based approach is that selecting students of an age cohort from different grades is more costly and disruptive, and it is more difficult to relate the results of an age cohort to the classrooms from which these students were drawn.

The accuracy of any survey results also depends on the quality of the information on which national samples are based as well as on the sampling procedures. For the latest international evaluations, advanced quality standards, procedures, instruments, and verification mechanisms have been developed that ensure that national samples yielded comparable data and that the results could be compared with confidence.

Last, but not the least, even the best international samples will only translate into comparable results if the sampled schools are willing to take part in the evaluation. At times, schools do not perceive to receive sufficient benefit from an evaluation that only yields national or subnational outcomes. Some countries have started to link international evaluations more closely to participating schools – either through providing them with school-level outcomes from the evaluation or the related questionnaires, or through the provision of better information on the objectives and nature of these evaluations. Incentives or feedback that have been deployed include: (1) better explanation of the context and usefulness of international evaluations at the start of the process to help engage teachers and schools; (2) preparing a briefing pack to prepare teachers, schools, pupils, and parents to overcome

initial anxieties of pupils and stimulate better communication within schools; (3) setting up an international buddies scheme with schools doing the international evaluations in other countries – in particular, for sharing ideas for using the results to improve; (4) giving out student certificates on the day; and (5) preparing electronic versions of feedback.

Comparability in Survey Implementation

Last but not least, well-designed international evaluations require to be well implemented to yield reliable results. The process begins with ensuring consistent quality and linguistic equivalence of the assessment instruments across countries. Precise translation and adaptation guidelines are supplied, also including instructions for the selection and training of translators. For each country, the translation and format of the assessment instruments (including test materials, marking guides, questionnaires, and manuals) are verified by expert translators appointed by an agency charged with the development of the assessment instruments (whose mother tongue was the language of instruction in the country concerned and who were knowledgeable about education systems) prior to their use.

The evaluations then need to be implemented through standardized procedures. Comprehensive manuals, typically, explain the implementation of the survey, including precise instructions for the work of school coordinators and scripts for test administrators for use during the assessment sessions. In the case of PISA, specially designated quality monitors visited all national centers to review data-collection procedures and school quality. Monitors from the international agency visited a sample of 15 schools from each participating country or region during the assessment. Marking procedures are designed to ensure consistent and accurate application of the internationally agreed marking guides.

Conclusions

In a globalized world, the benchmarks for public policy in education are no longer national goals or standards alone, but, increasingly, are the performance of the most successful educational systems internationally. International evaluations can be powerful instruments for educational research, policy, and practice by allowing educational systems to look at themselves in the light of intended, implemented, and achieved policies elsewhere. They can show what is possible in education – in terms of quality, equity, and efficiency in educational services – and they can foster better understanding of how different educational systems address similar problems. Most importantly – by providing an opportunity for policymakers and practitioners to look beyond the experiences evident in their own systems

and, thus, to reflect on some of the paradigms and beliefs underlying these – they hold out the promise to facilitate educational improvement. As this article shows, designing and implementing valid and reliable international evaluations pose major challenges – including defining the criteria for success in ways that are both comparable across countries while remaining meaningful at national levels, establishing comparable target populations, and carrying out the surveys under strictly standardized conditions. However, more recently, international evaluations have made significant strides toward this end.

See also: Sampling.

Bibliography

- Beaton, A. E., Mullis, I. V. S., Martin, M. O., et al. (1996). *Mathematics Achievement in the Middle School Years. Center for the Study of Testing, Evaluation, and Educational Policy*. Chestnut Hill, MA: Boston College.
- Döbert, H., Klieme, E., and Sroka, W. (2004). *Vertiefender Vergleich der Schulsysteme ausgewählter PISA-Teilnehmerstaaten*. Frankfurt a. M.: Deutsches Institut für pädagogische Forschung.
- Foshay, A. W., Thorndike, R. L., Hotyat, F., Pidgeon, D. A., and Walker, D. A. (1962). *Educational Achievement of Thirteen-Year-Olds in Twelve Countries*. Hamburg: UNESCO Institute for Education.
- Hopkins, D., Pennock, D., and Ritzén, J. (2008). *Evaluation of the Policy Impact of PISA*. Paris: OECD.
- IFIE-ALDUCIN (2007). *Mexican National Survey to Parents Regarding the Quality of Basic Education*. Mexico City: IFIE.
- Jullien, M. A. (1817). *Esquisse et vues préliminaires d'un ouvrage sur l'éducation comparée*. Paris: L. Colas.
- Knighton, T. and Bussière, P. (2006). Educational outcomes at age 19 associated with reading ability at age 15 (research paper), Statistics Canada, Ottawa.
- LaPointe, A. E., Mead, N. A., and Askew, J. M. (1992). *The International Assessment of Educational Progress Report*. Princeton, NJ: Educational Testing Service.
- LaPointe, A. E., Mead, N. A., and Phillips, G. W. (1989). *A World of Differences: An International Assessment of Mathematics and Science*. Princeton, NJ: Educational Testing Service.
- OECD (2004). Reviews of national policies for education – Denmark: Lessons from PISA 2000. Paris, OECD.
- OECD (2007). Reviews of national policies for education: Quality and equity of schooling in Scotland. Paris: OECD.
- OECD (2008). *Education at a Glance – OECD Indicators 2007*. Paris: OECD.
- Siniscalco, M. T. (2007). PISA e le valutazioni internazionali dei risultati della scuola. In Siniscalco, M. T., Bolletta, R., Mayer, M., and Pozio, S. (eds.) *Le valutazioni internazionali e la scuola italiana*. Bologna: Zanichelli.

Further Reading

- OECD and Statistics Canada (2000). *Literacy Skills for the Information Age*. Ottawa/Paris, OECD/Statistics Canada.

National Assessment Programs: The Example of the U.S. National Assessment of Educational Progress

M Schneider, American Institutes for Research, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Achievement levels, scale scores – Ideally, scale scores measure what students know and can do, achievement levels (often presented as descriptive terms, such as basic, proficient, or by level) are performance standards showing what students should know and be able to do.

National Assessment Governing Board (NAGB) – An independent bipartisan board appointed by the secretary of education, sets policy for NAEP and is responsible for developing the framework and test specifications.

National Assessment of Educational Progress (NAEP) – Also known as the nation's report card, it is a large nationally representative assessment of what America's students know and can do.

National Center for Education Statistics (NCES) – A federal statistical agency, located in the Institute for Education Sciences, the research arm of the US Department of Education.

PISA, the Programme for International Student Assessment – An assessment of the literacy of 15-year-olds administered by the Organization for Economic Cooperation and Development.

TIMSS, Trends in International Math and Science Study – An international assessment of science and math skills of fourth and eighth graders administered by the International Association for the Evaluation of Educational Achievement.

National testing programs of students are fairly common around the world. They can have different purposes, ranging from deciding which students go on to which secondary or postsecondary school, to being a core component of accountability systems, designed to find out which schools are doing a better job than others, to acting as a more general indicator of how well a province, a region, or a country is doing over time.

In the United States, the National Assessment of Educational Progress (NAEP), also known as the nation's report card, is a large nationally representative assessment of what America's students know and can do. NAEP is administered by the National Center for Education

Statistics (NCES), a federal statistical agency, located in the Institute for Education Sciences, the research arm of the US Department of Education.

Tests differ in the consequences they have for students, schools, or governments. For students, in particular, tests can vary widely in their impact. Some tests have very high stakes – for example, a growing number of states are putting into place high school exit tests that a student must pass in order to get a diploma. In contrast, for students, NAEP is a no-stakes test – there are no rewards or punishments for performance – indeed, there are no individual student-level scores.

For individual American states, the stakes are somewhat higher – when NAEP releases state results, there is an inevitable series of horse-race stories about who is winning and who is losing. While there are no direct consequences for doing well or badly on NAEP, governors and heads of state education departments can be embarrassed by declining scores or poor performance.

NAEP has also become a highly visible indicator of the nation's progress toward meeting the 100% proficiency goal set by the national No Child Left Behind (NCLB) Act for all students by 2014.

The Commissioner of Education Statistics, who heads the NCES, is responsible for carrying out NAEP. However, the National Assessment Governing Board (NAGB), an independent bipartisan board appointed by the secretary of education, sets policy for NAEP and is responsible for developing the framework and test specifications. By law, NAGB is designed to balance various education interests and its members must include governors, state legislators, local and state school officials, researchers, educators, business representatives, and members of the general public. There is an inevitable tension between NAGB and NCES, with NAGB given the major policy role, but NCES and its commissioner charged with maintaining the technical quality and validity of NAEP. Moreover, NCES contracts for and supervises operations that turn policy guidance into actual assessments. In addition, NCES is ultimately responsible for the fiscal integrity of NAEP and making sure that the sometimes overly ambitious goals of NAGB can be met within existing resources.

There are two main strands of NAEP: long-term trends (LTT) and main NAEP (which itself has national, state, and urban school district components).

Long-Term Trend NAEP

The NAEP LTT assessments have been administered nationally every 4 years (more or less) since the early 1970s. While LTT has assessed other subject areas, it now focuses on student performance in mathematics and reading at ages 9, 13, and 17. (An LTT science study was conducted from 1969 through 1999, but the framework was viewed as too outdated to continue. LTT writing was also administered six times between 1984 and 1996, but was dropped for technical reasons.) NAEP first assessed reading in 1971 and mathematics in 1973. The next LTT assessment results (for 2008) will be released in 2009.

The main advantage of LTT is the length of the time series and the use of the same framework since its inception. Preserving trend is the very reason for LTT and keeping the framework in place is the result of a specific policy decision to present a consistent long-term view of the progress of American students in two key subjects. As we note below, balancing the desire to update frameworks and the need to preserve trend data are perennial issues in main NAEP.

Both NAEP's and LTT give us a picture of the performance of American students and suggest some of the challenges of doing low-stakes assessments. This is evident in **Figures 1** and **2**, which are examples of recent LTT findings. In LTT (and in many other NAEP tests), gains in performance are much more noticeable in younger cohorts (defined by age or by grade) and they are more noticeable in math than in reading. For example,

in reading, since LTT was first administered in 1971, reading scores have increased significantly among 9- and 13-year-old students (with a greater gain among the youngest students) and stayed flat among 17-year-olds. These patterns raise the issue of motivation: since NAEP is a low-stakes test, as students get older, they have been found to be less engaged with NAEP tests, which may account for poorer performance in older students. NCES has conducted several studies about how to increase student motivation (e.g., through payments for correct answers), but how well these methods could scale to a full assessment is a serious question.

For math, as in reading, the size of the gains among younger students is larger than the gains among older ones. For many, the increase in math scores has been linked to improved math instruction but also to more time on task. The fact that math scores have moved up faster than reading is also often interpreted as showing that teaching math is easier than teaching reading, in part because math is a more school-based skill while reading draws on a wide range of cognitive skills that young children need to absorb in a wider range of venues than just the classroom.

These patterns, which are reinforced by the main NAEP results, have become part and parcel of policy debates around the US curriculum and the effects of NCLB. Both critics and supporters of NCLB have looked at the same trend lines and have found evidence that NCLB is working or that it is not. This highlights one of the fundamental problems with NAEP – it is best viewed

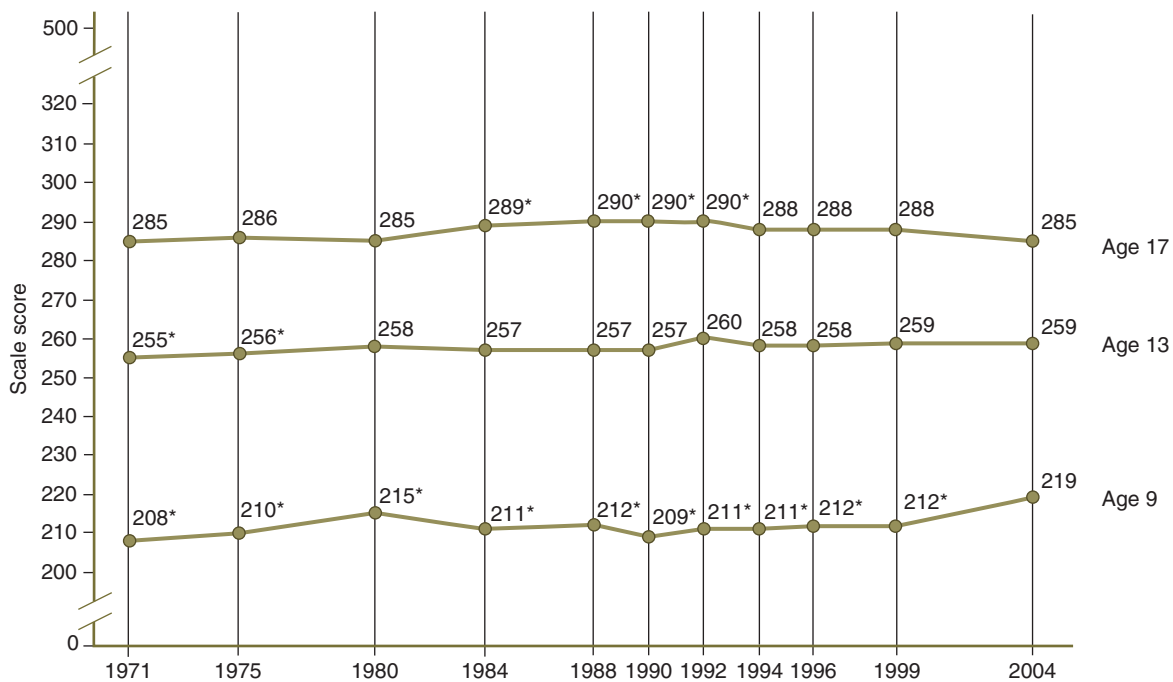


Figure 1 An example of NAEP long-term trends (LTT) in average reading-scale scores for students ages 9, 13, and 17: 1971–2004.

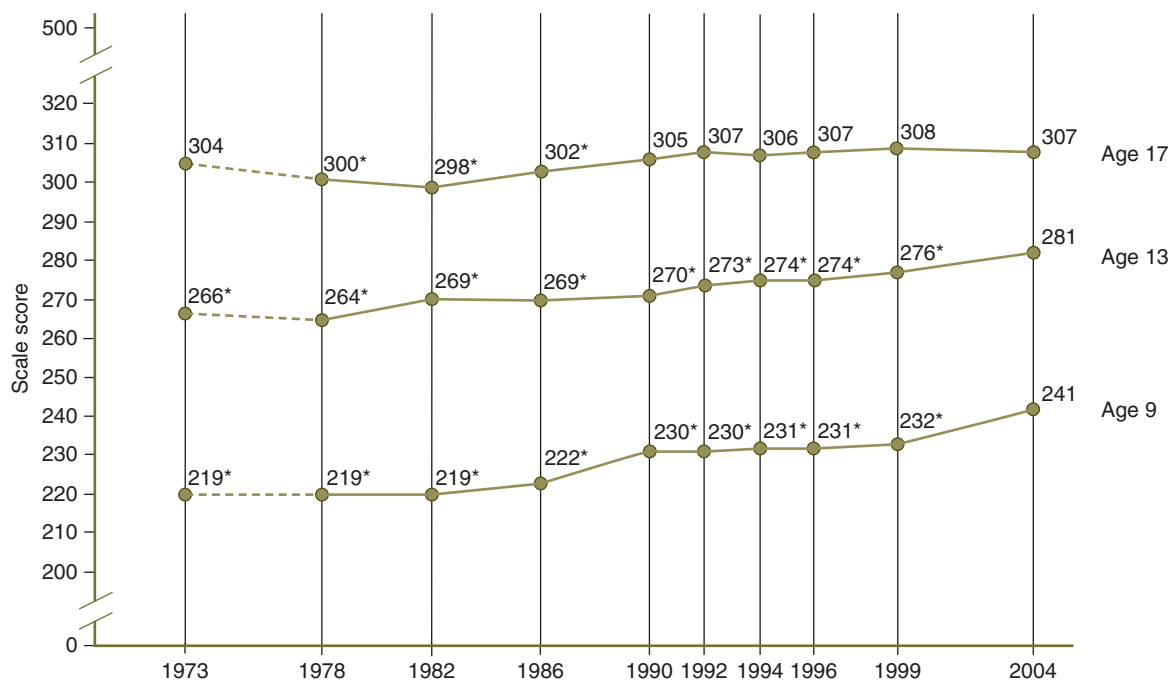


Figure 2 An example of NAEP LTT in average mathematics-scale scores for students ages 9, 13, and 17: 1973–2004.

as a benchmarking exercise and it is really too blunt an instrument to answer many important policy questions (a point returned to below).

Scale scores versus achievement levels

Both main NAEP and LTT NAEP report scale scores and achievement levels.

Scale-score results are reported either on a 0–500- or a 0–300-point scale, overall and for each of the subscales that may be included in an assessment. (The 0–500 scale is usually used for cross-grade assessments, such as reading and math in grades 4, 8, and 12 in main NAEP and at ages 9, 13, and 17 in LTT, while the 0–300 scale is used for within-grade assessments in main NAEP.) In the most recent reading assessment, these subscales were defined as contexts: reading for literary experience and reading for information. For mathematics, there were five content areas, including number properties and operations, measurement, geometry, data analysis and probability, and algebra. The balance of items across these contexts or content areas differs from grade to grade, as determined by the assessment framework.

As NAEP scales are developed independently for each subject and for each content area within a subject, the scores should not be compared across subjects or across content areas within the same subject. This is a constant problem when NAEP scores are released because people see similar numbers and want to make the kind of comparisons just presented above (i.e., where it is stated that the gains in math were greater than the gains in reading – to

make sense of the comparisons, these scores should be turned into standardized effect sizes. However, only on rare occasions does NAEP report these standardized difference measures). (While effect sizes would improve one's ability to make comparisons across content areas, it would not eliminate incompatibilities across content areas that have items of differing difficulty and depth.)

To help the public and policymakers have a better understanding of the meaning of main NAEP results, the NAGB sets specific achievement levels for each subject area and grade. While scale scores ideally measure what students know and can do, achievement levels are performance standards showing what students should know and be able to do.

In LTT, for each of the subject-area scales, performance levels are set at 50-point increments from 150 through 350. The five performance levels – 150, 200, 250, 300, and 350 – are described in terms of the knowledge and skills likely to be demonstrated by students who reached each level. As an example, these are the achievement levels for math:

- Level 150: Simple arithmetic facts.
- Level 200: Beginning skills and understandings.
- Level 250: Numerical operations and beginning problem solving.
- Level 300: Moderately complex procedures and reasoning.
- Level 350: Multistep problem solving and algebra.

In main NAEP, results are reported as percentages of students performing at or above the basic and proficient

levels and at the advanced level. Below are the core ideas that define main NAEP's achievement levels:

- Basic denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at a given grade.
- Proficient represents solid academic performance. Students reaching this level have demonstrated competency over challenging subject matter.
- Advanced represents superior performance.

The simple user-friendly names and description of achievement levels in main NAEP make them far more compelling than the levels as defined by LTT (or by the numbered levels used by the best-known international assessments – the Programme for International Student Assessment (PISA) or the Trends in International Math and Science Study (TIMSS)). However, there has been controversy over these achievement levels. The current advisory that appears in NAEP reports reads: “As provided by law, NCES, upon review of congressionally mandated evaluations of NAEP, has determined that achievement levels are to be used on a trial basis and should be interpreted with caution. The NAEP achievement levels have been widely used by national and state officials.” As with many aspects of the high-visibility NAEP program, technical cautions are often ignored in the face of pressing needs of policymakers and the public for usable and understandable information.

Main NAEP

NAEP has been testing a variety (some would say too wide a variety) of subjects with differing degrees of frequency and consistency since 1969. Some subjects have come and gone (career/occupational development in 1973; an index of basic skills in 1974); some have run into trouble (foreign language); others appear with an irregular frequency, depending on money as well as political interest and pressure (civics, geography, economics, and the arts). The core subjects of NAEP, not surprisingly, are reading and math, a focus reinforced by the provisions of the NCLB, but writing and science are, if not by statute, then by practice, at the core of NAEP's testing program.

With the passage of NCLB, NAEP's budget increased dramatically and is now in the vicinity of US\$100 million per year. This budget growth allowed the growth of NAEP, including an expansion in the range of subjects assessed, the creation of state representative samples for several of NAEP's assessments, and the assessment of performance in a number of large urban districts. We describe state NAEP and the trial urban district assessments (TUDA) in more detail below, but first we need to describe some of the characteristics of how NAEP is administered.

First, while NAEP's LTT is based on age-defined samples, main NAEP results are based on representative samples of students at grades 4, 8, and/or 12. NAEP collects data on subject-matter achievement, instructional experiences, and school environment for populations of students (e.g., all fourth-graders) and by defined policy-relevant subgroups (most notably, gender, race/ethnicity, and income). (Income is measured by participation in the federal free and reduced-price lunch program. This has been an increasingly weak indicator of income, and NCES has undertaken studies trying to identify better measures.)

NAEP does not provide scores for individual students or schools. This is fundamental to the design of NAEP, which, because of its large number of test questions, uses matrix sampling and a focused balanced incomplete block (BIB) or partially balanced incomplete block (pBIB) design in which blocks or groups of cognitive items are assigned to a set of student booklets.

Given this design, NAEP can sample enough students to obtain precise results for each test question within an average test time of about an hour and a half per student. As a result, while the full framework is covered, there are no individual student-level test scores. This design was a major measurement technical breakthrough in the early days of NAEP and allowed the testing of a wide range of skills without forcing students to sit through hours upon hours of testing. However, the cost is the absence of individual-level scores. (This also solved a political objection: since no individual student scores were calculated, privacy concerns were sidestepped, removing one of the bones of contention that were raised (and continue to be raised) about NAEP.) Without individual scores, NAEP does not allow for longitudinal analysis, which is increasingly the way in which researchers prefer to measure progress. NAEP shares this research design (and its shortcomings) with the major international assessments, such as PISA and TIMSS.

Despite this limit (and it is a severe one), the most powerful aspect of NAEP is that it is the only nationwide test administered uniformly to all participating students using the same test items and procedures. As a result, NAEP serves as a common yardstick against which state assessments can be compared. This allows NAEP to serve as a comparative tool to gauge the rigor of individual state-assessment standards. This role was central to the NCES report *Mapping 2005 State Proficiency Standards onto the NAEP Scale* (National Center for Education Statistics, 2007).

Under the NCLB, states are required to report the percentages of students achieving proficiency in reading and mathematics for grades 3 through 8 (and at least once during grades 10 through 12). But they do so using their own assessment instruments. For each subject and grade combination, the percentages of students reaching proficiency (or any other achievement level) vary widely

across states. However, because NAEP is a uniform test, at grades 4 and 8, state percentages can be compared to the percentages of students achieving proficiency on NAEP, providing a gauge of the rigor of state standards.

In the 2007 report, large discrepancies in state versus NAEP proficiency standards were reported for both grades 4 and 8 in math and reading. For each of the four subject and grade combinations, the NAEP score equivalents to the states' proficiency standards varied widely, spanning a range of 60–80 NAEP score points (roughly two standard deviations in the NAEP test). But far more importantly, most of the NAEP score equivalents fall below the cut-point corresponding to the NAEP proficient standard, and many fall below the cut-point corresponding to the NAEP basic standard. **Figure 3** shows how NAEP can be used as a yardstick to identify differences in state proficiency standards, using eighth-grade reading as an example.

State NAEP

State NAEP is based on large state-wide representative samples of students taking NAEP. This allows state scores to be calculated and changes in state performance measured over time (remember, the P in NAEP stands for progress). The assessments given in the states are the same as those

given nationally, allowing states to compare themselves to the nation and to one another.

Until 1990, NAEP reported only the academic achievement of the nation as a whole and student groups within the nation. In 1988, Congress authorized a voluntary trial state assessment, in which state-representative samples of students were selected from each state or jurisdiction that agreed to participate in state NAEP. These trial state assessments were conducted in 1990, 1992, and 1994, after which time the state component was no longer labeled a trial. Between 1990 and 2002, state NAEP was a mix of testing at grades 4 and 8 in reading, math, science, and writing. For example, in 1990, only math at grade 8 was tested. In 1992, 1994, and 1996, math was assessed in volunteer states at grades 4 and 8, but math was not assessed at all in 1994 and 2002. State assessments in science were conducted at grades 4 and 8 in 1996 and 2000.

In short, the state testing program was erratic. The NCLB regularized state assessments. Beginning in 2003 and continuing through the present, states that accept federal title 1 monies must participate in statewide assessments in reading and math in grades 4 and 8 every other year. (Title 1 of the Elementary Secondary Education Act of 1965, as amended most recently by the NCLB, is the primary vehicle by which the US federal government

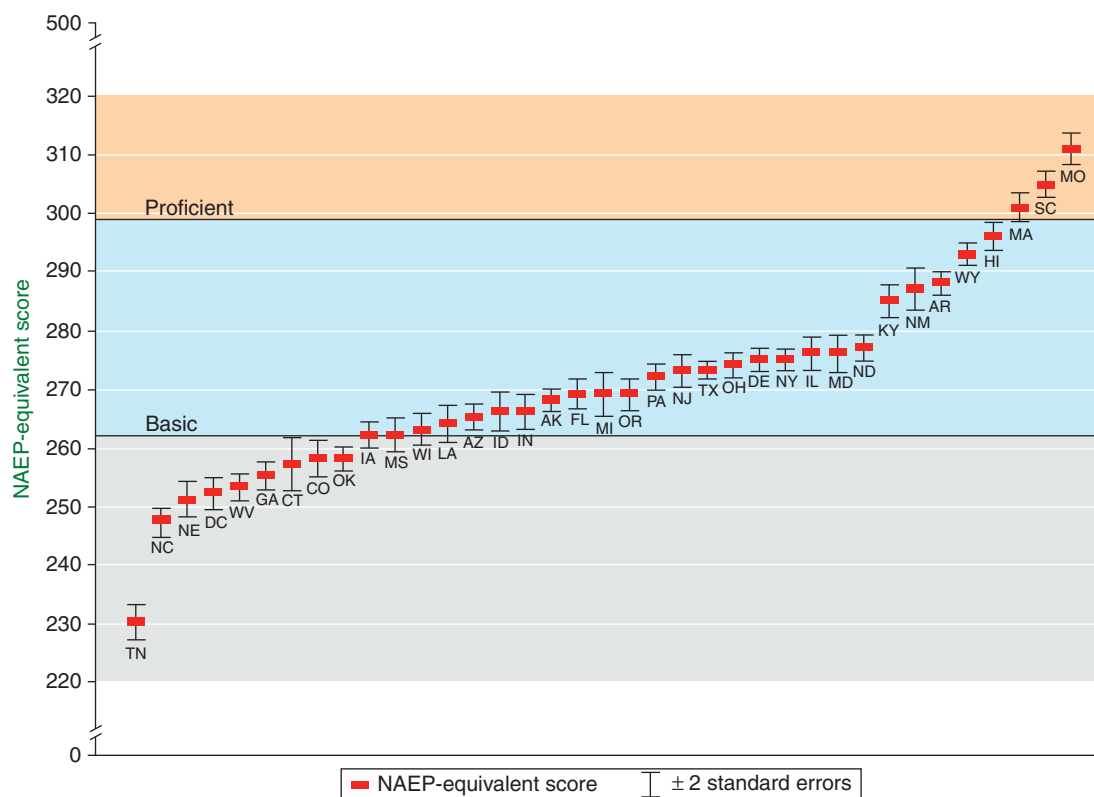


Figure 3 Mapping state proficiency standards using NAEP as a common yardstick. Grade 8 reading. From Lee, J., Grigg, W., and Dion, G. (2007). *The Nation's Report Card: Mathematics 2007 (NCES 2007-494)*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education. <http://nces.ed.gov/nationsreportcard/pdf/main2007/2007494.pdf>.

supports local schools. The program provides financial assistance to school districts and schools with high numbers or high percentages of poor children. Title 1 represents about 1/3 of total federal expenditures on elementary and secondary education, an annual budget allocation of well over US\$10 billion.) In addition to these mandatory state tests, since passing the NCLB, NAEP has offered (and paid for) state assessments in science at fourth and eighth grades in 2005 and writing at grade 8 in 2007.

As noted, since 1990, state NAEP has assessed students at grades 4 and 8, but not at grade 12. However, a pilot program has been funded by the US Congress, so that 11 states will participate at the twelfth-grade level in 2009 and get reading and math scores.

Even though NAEP produces state scores, like all NAEP assessments, it does not produce individual scores for the students or schools assessed.

NAEP TUDA program

Just as in the 1990s when a trial state testing program was launched, NAEP now has a TUDA program that is designed to explore the use of NAEP to assess the performance of public school students at the district level.

TUDA began with five urban districts in 2002 – Atlanta, Chicago, Houston, Los Angeles, and New York City – and has since added Austin, Boston, Charlotte-Mecklenburg, Cleveland, and San Diego. The District of Columbia is included in many TUDA reports.

Since the TUDA assessment is administered as an expanded sample of students being tested for state NAEP, the subjects tested in the urban districts coincide with state assessments. In 2002, six urban districts participated in reading and writing at grades 4 and 8. In 2003, nine districts participated in reading and math. In 2005 and 2007, ten districts reported math and reading scores. In addition, the District of Columbia, which appears in state NAEP also appears in these TUDA assessments.

In 2005, district science scores were released and in 2007, district grade 8 writing scores. (The District of Columbia is too small to support science and writing oversamples and was not included in those TUDA reports.) TUDA has been used to track changes in performance in areas of the country where many of the challenges of education are the most severe. In 2009, Congress authorized sufficient funds to add several more urban districts.

NAEP special studies

In addition to the assessments, NAEP coordinates a number of special educational studies. Perhaps the most important is the NAEP High School Transcript Study (HSTS) which goes beyond just scores, gathering information about school practices and how these practices might affect performance outcomes (Shettle *et al.*, 2007). As the

HSTS was conducted in 1982, 1987, 1990, 1992, 1994, 1998, 2000, and 2005, NAEP HSTS provides a wealth of information about changing course-taking patterns over time. The 2005 report presented data from transcripts collected from a nationally representative sample of 26 000 high school graduates.

The report presented information about the types of courses 2005 graduates took during high school, how many credits they earned, and the grades they received. As the 2005 results were the most recent in a long line of such transcript studies, the results were compared over time and differences among graduates by race/ethnicity, gender, and parent education were examined.

Among the most notable findings, the report showed that students were taking more academic courses and more rigorous ones (at least judging by their titles) than ever before. As an example, the study found that 2005 graduates took about 360 additional hours of instruction during their high school careers than their 1990 counterparts. Increased percentages of White, Black, Hispanic, and Asian/Pacific Islander graduates completed at least a mid-level curriculum in 2005 compared with 1990. However, differences by race/ethnicity persisted: in 2005, both Black and Hispanic graduates were less likely than White graduates to have completed calculus or advanced science courses.

Conclusions

The US NAEP serves as the nation's report card. Since the passage of the NCLB, NAEP has become a central tool in measuring educational progress in the United States as a whole, in individual states, and in a growing number of large urban districts. Moreover, while NCLB mandates testing in reading and math at grades 4 and 8 at the state and national levels, state-level testing in science and writing has been and will likely continue to be high priority. NAEP also tests a range of other academic subjects on a less regular basis, including civics, history, and economics. NAEP unsuccessfully tried to assess foreign languages in 2004, and in 2008, NAEP assessed the arts, although a previous arts assessment in 1997 was very expensive. The budget for NAEP increased more than threefold with the passage of NCLB and, as it has established itself as the nation's report card, Congress has appropriated further funds to support its testing program.

While on many levels this signals a successful national assessment program, there are challenges facing NAEP.

Any large-scale assessment is shaped by its framework, and every framework has to be both forward-looking (what skills should students have tomorrow) and also reflective of what skills they need today. This is complicated by the fact that, as noted, the P in NAEP stands for progress – and measuring progress implies maintaining

the framework and the skills tested over time. However, at some point, today's skills become yesterday's, and the framework and the assessment need to change. The challenge is to make these changes while somehow continuing to report on trends. While LTT is specifically designed for this purpose, and has, thus, kept changes to its framework to a minimum, it has been eclipsed by the far more visible and frequent main NAEP, which is not as committed to keeping a stable framework. Introducing new frameworks, while making the assessment more relevant, has costs.

As an example of how jarring the effects of a new framework can be, consider the 2005 report on 12th-grade reading and math. While that year's reading test has a depth of trend analysis, the math report notes that "Because of changes in assessment content and administration, the results for 2005 could not be directly compared to those from previous years" (see Grigg *et al.*, 2007). In turn, the report released relatively little information, such as presented in **Figure 4** below.

NAGB made the policy decision that the old 12th-grade math framework was too far out of line with modern math curricula and skills and that breaking trend was a cost worth absorbing. Breaking trend for the mandated NCLB testing areas (reading and math, grades 4 and 8) was considered by NAGB, but the political costs of doing that were far higher than for the 12th-grade (nonNCLB mandated) assessment. In turn, NCES has spent considerable time and resources trying to work out mechanisms to introduce changes into these assessments while still preserving trend.

While pedagogical changes continually challenge NAEP frameworks, technological challenges also loom. Despite some advances, especially in the upcoming NAEP science assessment, which has a large interactive computer-based component, NAEP is still a paper-and-pencil test. Moving NAEP to computer-based assessments poses a new set of challenges.

Some of the challenges are logistical: the computer infrastructure of the 100 000 or so schools in the US varies

widely. NCES discovered that assessing computer capacity in schools was like an archeological dig. Not only was the world divided into Apples versus PCs, but in the PC world there were layers upon layers of Windows operating systems (although NCES did not find any DOS-based machines, there were machines running Windows 95, 98, 2000, VISTA, and all the permutations of each). There were also differences in the support staff available. In one school the NAEP team visited to assess the possibility of computerized assessment, the IT person excused himself at 2:30, promising to return at 4:30 – because he had to drive a school bus to take students home!

There is also a question of how moving to a computer-based assessment might affect performance. If some groups are more familiar with a technology than other groups, then construct-irrelevant variance can be introduced – that is, some set of skills extraneous to the purpose of the assessment might bias results.

Even more important is the changing nature of assessments themselves. Recall NAEP is based on matrix sampling and a BIB design. This design allows a whole framework to be tested, while not forcing any student to sit through the hours-long exam that would be required to accurately assess their skills across the whole framework. But this also means that there are no individual-level scores.

For many subjects, especially math, there are more elegant solutions. In the last few years, computer-based adaptive testing has made great strides and NAEP should be moving toward that approach to better assess student skills. In addition, longitudinal data on student performance has become the new gold standard. But NAEP is a cross-sectional measurement without individual student scores. This means that NAEP is not amenable to some of the most interesting work using growth models to measure progress and to identify what works and for whom.

Another looming issue is how NAEP will fit into an increasingly global environment. The concrete manifestation of this is the role of NAEP versus the three big international assessments: the Program for International Reading Literacy (PIRLS), TIMSS, and the PISA. PIRLS is a fourth-grade literacy assessment and TIMSS tests students in grades 4 and 8 in science and math. While on different cycles than NAEP, both are testing in the same grades and there is considerable overlap in their purposes, frameworks, and actual items. The bigger question is the growing importance of PISA, an assessment administered through the Organization for Economic Cooperation and Development (OECD), which includes all 30 OECD countries plus about an equal number of nonOECD jurisdictions.

PISA is a yield study of reading, math, and science competencies. The students tested and the basic ideas motivating PISA do not align as closely with NAEP as do TIMSS and PIRLS. Indeed, while NAEP is designed

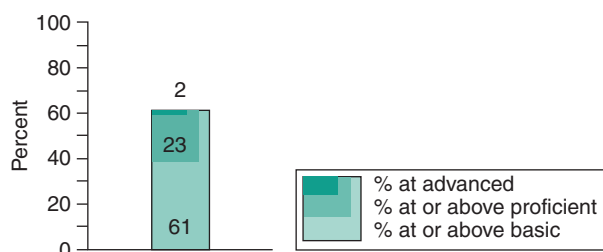


Figure 4 An example of how changing a NAEP framework resulted in limited information. 2005 12th-grade math. From US Department of Education, Institute of Education Services, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), and 2005 Mathematics Assessment.

to reflect (more rather than less) the curricula goals of American schools, PISA is much less concerned with assessing school-based knowledge. Rather, as a yield study, PISA is designed to measure the cumulative product of all educational activities in a given country in developing student competencies to apply knowledge to real-world situations.

As state governors now want to compare their students to those in Korea or Germany rather than Kansas or Georgia, PISA has gained national attention. Indeed, many states have considered administering PISA to a statewide representative sample to benchmark their student performance against students from countries around the world. (To date, no states have actually signed up for state-level PISA, although a number of states have considered it and the National Governors Association has been supportive of the idea. Note that in the 2007 TIMSS, two states (Massachusetts and Minnesota) did have state representative samples and were able to compare their states' students to the USA as a whole and to other nations which administered TIMSS. See the 2008 report *Benchmarking for Success: Ensuring U.S. Students Receive a World-class Education* as an example of the tension between NAEP and these international assessments (National Governors Association *et al.*, 2008).) NAEP is designed for state benchmarking; if attention continues to be focused on international comparisons, one of NAEP's key functions may lose its prominence.

Finally, there is the perennial issue of the role NAEP can play as a research database. As noted above, NAEP is the nation's prime method for assessing national educational progress and for benchmarking performance over time and across jurisdictions. Despite the challenges it faces, NAEP still performs that benchmarking function admirably. However, given the size of the database, its quality, and the fact that it is repeated on a regular basis, NAEP tempts many to use it for research and policy analysis. The problem is that it is far too easy to ignore the limits on NAEP and use it for purposes for which it is not suited. Most notably, state-assessment systems that track individual students over time and are suitable for things such as growth and value-added modeling, are more likely to yield reliable policy-relevant results than NAEP, despite its size, its cost, and its prominence.

See *also*: Evaluating Education in Three Policy Eras; Evaluation Governance and Planning; Evaluation Methodology; Evaluation Use; International Evaluations.

Bibliography

- Grigg, W., Donahue, P., and Dion, G. (2007). *The Nation's Report Card: 12th-Grade Reading and Mathematics 2005* (NCES 2007-468). Washington, DC: US Department of Education, National Center for Education Statistics/US Government Printing Office. http://nationsreportcard.gov/reading_math_grade12_2005/s0301.asp (accessed August 2009).
- National Center for Education Statistics (2007). *Mapping 2005 State Proficiency Standards Onto the NAEP Scales* (NCES 2007-482). Washington, DC: National Center for Education Statistics, US Department of Education. <http://nces.ed.gov/nationsreportcard/pubs/studies/2007482.asp> (accessed August 2009).
- National Governors Association, Councils of Chief State School Officers, and Achieve (2008). *Benchmarking for Success: Ensuring U.S. Students Receive a World-Class Education*. <http://www.achieve.org/BenchmarkingforSuccess> (accessed August 2009).
- Shettle, C., Roey, S., Mordica, J., *et al.* (2007). *The Nation's Report Card: America's High School Graduates* (NCES 2007-467). Washington, DC: US Department of Education, National Center for Education Statistics, US Government Printing Office. <http://nces.ed.gov/nationsreportcard/hsts> (accessed August 2009).

Further Reading

- Lee, J., Grigg, W., and Dion, G. (2007). *The Nation's Report Card: Mathematics 2007* (NCES 2007-494). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education. <http://nces.ed.gov/nationsreportcard/pdf/main2007/2007494.pdf> (accessed August 2009).
- National Center for Educational Statistics (2006). Comparing Mathematics Content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 Assessments. *Technical Report*. <http://nces.ed.gov/pubs2006/2006029.pdf> (accessed August 2009).
- Stephens, M. and Coleman, M. (2007). Comparing PIRLS and PISA with NAEP in Reading, Mathematics, and Science. *Working Paper*. Washington, DC: National Center for Education Statistics, US Department of Education. <http://nces.ed.gov/Surveys/PISA/pdf/compaper12082004.pdf> (accessed August 2009).

Relevant Websites

- <http://nces.ed.gov> – National Center for Educational Statistics, NAEP Data Explorer; NAEP Questions Tools; NAEP Restricted-Use Data Files and Variables in Each File.
- <http://nces.ed.gov> – National Center for Educational Statistics, National Assessment of Educational Progress Technical Documentation; Overview.

Principal Evaluation: Concepts, Needs and Realities

N S Glasman and L D Glasman, University of California, Berkeley, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Evaluation Terms

At the risk of oversimplification but for the purpose of clarification, we present here the fundamental components of evaluation and corresponding examples.

1. *An evaluation problem.* This relates to situations containing issues that interfere or seem to interfere with meeting expectations. A decision has been made to examine these issues in a focused way. Some examples are:
 - a. Despite the fact that additional resources are allocated to certain schools, why are student test scores not improving?
 - b. In districts with a high first-year teacher turn over what can be done to reduce it?
 - c. Can closer collaboration between a high school's principal and counselor improve the quality of the counselor's contributions to students' well-being?
2. *An evaluation purpose.* This refers to learning about something (or someone) and judging the level of success obtained by something (someone) in order to make a decision about it. Some examples are:
 - a. to determine the most effective and efficient ways of utilizing added resources for improving student test scores;
 - b. to plan stronger support systems for first-year teachers; and
 - c. to develop ways for the principal and counselor to learn more about each other's training, role, and educational philosophy in order to improve the quality of the contributions of the counselor to the students' well-being.
3. *An evaluation object.* This refers to the specific thing or person about which information is gathered. Some examples are:
 - a. ways in which additionally allocated funds are used to improve student test scores;
 - b. nature of first-year teacher support system and rate of first-year teacher turnover; and
 - c. nature of the interactions between the principal and the counselor and the counselors' perceptions about the effect of these interactions on their contributions to the well-being of students.
4. *Evaluation information.* This refers to the data describing the objects that is being evaluated. Some examples are:
 - a. documents depicting fund allocation and student test scores;
 - b. first-year teachers' perceptions about the support system that the principal establishes for them and percentages of pre- and post- (support system establishment) rates of turnover; and
 - c. perceptions of counselors and principals about their interactions and the counselors' perceptions about the pre- and post- (interaction activities) quality of their student counseling work.
5. *Evaluation criteria.* These refer to the standards by which something is judged. Some examples are:
 - a. the manner and the extent to which using additional allocation of resources helps improve student test scores;
 - b. the manner and the extent to which stronger support systems of first-year teachers helps reduce their turnover; and
 - c. the manner and the extent to which the principal and counselor learning about each other's training, role, and philosophy help improve the quality of the counselor's contributions to students' well-being.
6. *An evaluation-based decision.* A formal judgment and a conclusion made with regard to what action(s) has (have) to take place as a result of an evaluation. Some examples are:
 - a. Whether to continue or modify the use of the additional funds, or the amount of allocated additional funds?
 - b. Whether to continue or modify the first-year teacher support system?
 - c. Whether to continue or modify the nature of the principal/counselor interactions, or the nature of the student counseling services?

Selected Publications on Principal Evaluation

Two literature-based and detailed chapters on principal evaluation have been published in recent years (Glasman and Heck, 1996, 2003). We first offer current reflections on the data, conclusions, and suggestions presented in each of these publications.

The 1996 Publication: Role-Based Evaluation of Principals – Developing an Appraisal System

Compared to earlier years, increased attention to principal evaluation between the 1970s and the early 1990s

came about because of the general increased interest in evaluation in the field of education. In fact, politically mandated evaluation was part of a great number of categorical federal and state funding allocations to school districts.

Research work on educational evaluation theories, models, and principles intensified during this period. It was mostly psychologists, researchers in the area of learning, and newly trained professional evaluators who designed methods for evaluating educational programs and personnel. Emphasized in these designs were the selection of evaluation purposes, objects to be evaluated, and standards to which evaluators must adhere.

In the late 1980s and early 1990s, specific attention focused on teacher evaluation and its relation to principal evaluation. Teachers and principals were identified as being pivotal in relation to the curricula, programs, student behavior, and student learning. Therefore, the need to evaluate their performance seemed obvious. At the same time, some structural changes in the governance of education were implemented. The state acquired additional curricular and budgetary authorities in relation to educating the increasing number of new immigrants. In addition, the federal government increased its funding of support programs for these children. Increased government funding required increased auditing too. Principals' responsibilities expanded and the need to monitor their activities became apparent.

Role-based evaluation of the principal became an important concept used in guiding research. For example, some writing suggested that it is unfair to evaluate principals on the basis of their performance of responsibilities if they possess insufficient authority to carry out these responsibilities. Evaluation of principals' performance, it was argued, should be limited only to tasks over which principals have or can have full control. This led to research on the evaluation of primarily inherent personal traits rather than on actual performance. However, the usefulness of such evaluations was questioned. Not many school districts are known to have engaged in this kind of principal evaluation and not many resulting decisions are known to have occurred.

The 2003 Publication: Principal Evaluation in the United States

It is generally agreed that until the 1980s principal evaluations were mostly role based. Items to be evaluated were typically selected from the school-district-prepared formal job description. Various scales were selected as evaluation criteria. For example, a common evaluation item might be: guides teachers and provides support to them. This item would require a rating on a scale of points ranging from poor to excellent.

In the late 1980s, contextual forces emerged that brought about changes in the nature of the practice of role-based principal evaluation. Parents, legislators, and other policymakers made demands that schools become accountable for how well students learn.

This implied that the evaluation of teachers (and later, also, principals) does not remain as just role based (assessing the quality of the activities), but also as outcome-based (assessing the results of the activities). Accountability demands implied that student outcomes would be included as items to be assessed in the evaluation of school personnel.

Teachers strongly resisted these demands. They considered themselves to be professionals who must be evaluated on what they do in the classroom and how well students do in relation to what teachers teach. Principals were in a difficult position. They could not respond to the new demands without the cooperation of the teachers. On the other hand, the demand for student-learning-based accountability became a mandated policy.

Within a short time, teachers began to insist that their principals agree with their arguments and even serve as the teachers' advocate to external stakeholders. Evaluation policymakers did not accept this argument but agreed to hold the school as a whole accountable for student learning. This notion implied that teachers alone would not be accountable for student learning. Teachers' resistance was reduced but not eliminated. However, in the context of school-wide and outcome-based accountability, the evaluation of the person who carries the overall responsibility for the school received increased attention. Thus was born the notion of principal accountability. It gave rise to studies focusing on the intersection of evaluation and the school principal, a new academic inquiry.

These contextual forces impacted principal evaluation in new ways. No longer were the principals serving only as educational officers whose evaluation was a perfunctory activity, if it occurred at all. From now on their role was also conceived of as leaders who need to have authority and exercise influence. The differentiation between the principals' leadership and managerial tasks was first conceived of in relation to principal evaluation: Which of the two is more important to evaluate?

The next step in the design of principal evaluation included a more detailed examination of:

1. the criteria by which information is selected about the object to be evaluated;
2. the merit of the information judged for use in the evaluation; and
3. the ways by which the raw findings, such as scores and statements, are utilized in the evaluation process.

Thus, educational evaluation utilization emerged as an important new field of investigation.

Research at that time increasingly focused on the formal positions of those who conducted the evaluation, including their authority to conduct it, their qualifications to evaluate, and the evaluation consequences.

A Summary of Suggestions Made in the Two Publications

The publication titled 'Role-based evaluation of principals' proposed one possible sequence of principal evaluation activities that focuses on principal performance (Glasman and Heck, 1996: 387–388). The sequence included six steps, with some detail given for each. The steps were: setting the agenda, delimiting the scope of the evaluation, developing evaluation standards, specifying criteria for selecting information, measuring and collecting data, and making an evaluation decision.

The publication 'Principal evaluation in the United States' offered a conceptual framework for examining the principal's stage of employment (Glasman and Heck, 2003: 650, 651). The first stage is the period of his/her hiring and assignment. Evaluation activities include setting expected tasks, selecting types of data to be collected, and choosing indicators to measure the extent to which expectations will be met. The second stage is the improvement stage. Formative evaluative activities include data gathering, comparing them to expectations, judging the value (if needed) to adjustments that the principal makes, and (if needed) fine-tuning the evaluation activities themselves. The third stage involves summative evaluation of the principal's performance and the use of the results of this evaluation as inputs to making specific decisions pertaining to the principal's future career.

Stakeholders Define the Meaning of Principal Evaluation

Multiple meanings have been attributed to the concept of principal evaluation. Generally speaking, a definition is a function of how the vested interest of a specific stakeholder shapes it. For example, the vested interest of the Council of Chief State School Officers' (2005) has been to expand the number of states that examine school administrators for licensure purposes according to agreed-upon indicators. The Council developed six professional standards that constitute a very detailed definition. Each of the standards includes dozens of kinds of knowledge, personal attributes, and behaviors.

Another example shaping definitions of principal evaluation is the influence of a very large number of stakeholders, namely the public, and particularly, parents of school children. Most of what these people want is that student learning is at least satisfactory. Low student test

scores disturb this group. Through their political representatives, these stakeholders have called for increased school accountability. As student test scores did not significantly improve and as the clarity of how to implement school accountability policies was lacking, the evaluation component of these policies soon took centerstage and focused on principals' responsibilities (Fowler, 2000 ch. 11; Glasman and Glasman, 2007: 3–4; Wong *et al.*, 2007: 33–35; 160–169). In this case, principal evaluation has essentially been defined as ascertaining the degree to which principals are responsible for student learning. With stronger emphasis on outcome-based accountability, some principals have faced the challenges of the results of high-stakes testing of students (DeMoss, 2002), which has indicated using student scores on norm-referenced tests as the central input to judging and deciding about the success of principals, students, and schools. As a result, principals' professionalism has been seen as being damaged (Firestone and Riehl, 2005: 117–118).

Multiple other stakeholders have also shaped the nature of principal evaluation activities. Among them are the principals' superiors in the district office, the principals themselves, the staff members who work in their schools, various principals' professional organizations, members of their district boards of education, local business and industry employers, as well as other lay people who have had a general interest in the relation between school and society. However, systematic research has not been carried out on the processes by which these stakeholders have shaped the nature of principal evaluation.

Professional evaluators have usually been trained in economics, psychology, statistics, or administration. They engage in evaluation activities such as education costs and programs, teachers and student achievement, but only very little in principals' performance (Popham, 2006). School principals as evaluators rather than those who are being evaluated is another matter and should not be confused with principal evaluation as it is considered in this article.

Most publications in the area of educational administration have reported the existence of almost total agreement among stakeholders about the pivotal role that the principal plays in education. One would assume that this fact alone should have accelerated the progress in studying and in improving the practice of principal evaluation; however, progress has been slow. There is little consensus among stakeholders as to what principal evaluation should mean. Moreover, the nature of the principal's role itself has been changing.

In the professional educational evaluation literature only two central contributions have been made to principal evaluation: (1) standards were developed by which evaluation of educational personnel should be conducted and (2) also a range of possible purposes for which personnel evaluation could be conducted was specified

(The Joint Committee on Standards for Educational Evaluation and Stufflebeam, 1988).

Four groups of standards have been proposed:

Propriety standards imply suitability and honesty of the evaluator. Utility standards include usefulness of the evaluation results as input to decisions that are to be made on the basis of the evaluation. Feasibility standards refer to the availability of resources to conduct the evaluation and the probability of attaining the desired information. Accuracy standards consider the validity and reliability of the findings and the analyses of the data. As to the possible purposes for which personnel evaluation could be conducted, this includes making decisions about entry to training, certification, role defining, selection, tenure, staff development, professional feedback, accountability, merit awards, promotion, and termination (p.17).

Twenty years after these standards and purposes were published it is still not generally known who has made use of these contributions other than evaluation professionals.

A General Working Definition of Principal Evaluation

To the best of these writers' knowledge, there is no specific definition of principal evaluation to which all stakeholders would subscribe. However, there might be an acceptable one if it is sufficiently general in scope. This kind of a definition would allow for multiple specific interpretations so as to suit multiple stakeholders as they shape their own specific terms. One such possible definition is as follows:

Principal evaluation involves gathering information in the educational context about a principal and judging her/his performance on the basis of this information for the purpose of making a decision about his/her future.

An acceptance of such a definition, these writers feel, might allow most stakeholders of principal evaluation to comfortably subscribe to the following guiding assumptions:

1. A specific and agreed-upon administrative purpose of the principal evaluation needs to be articulated by those who are involved in performing it (e.g., improvement, reassignment, and release from responsibilities).
2. Specific and agreed-upon objects to be evaluated that describe and relate to the principal need to be articulated (e.g., personal characteristics, knowledge, skills, visions, behaviors, attitudes, decisions, actions, task accomplishments, performance, general direct and/or indirect impact, general performance, and specific objectively measured effects).
3. Evaluation information and criteria of performance (that are directly relevant to the specific evaluation purpose) need to be selected for assessing the performance (e.g., degree of satisfaction of parents regarding

the quality of education their children are receiving and the extent to which the principal enhances the quality of education as the parents perceive it; degree to which the principal reports success in resolving conflicts and the perceptions of the parties to the conflict about that resolution; teachers' perceptions about the degree to which the principal is effective in achieving school goals given the resources that are available to him/her to accomplish these goals; and percentage of students that score at or above expected levels and the principal's perceptions about helping teachers understand the scores for improvement purposes).

4. Maximum amount of suitable, possible, useful, and accurate information should be collected.

Objects to be Evaluated for Principal Evaluation

Depending on the purpose of the principal evaluation, a number of selections may be made as to what the objects to be evaluated will be. Two general rules can be followed. One rule applies to a situation where the purpose of the evaluation is quite specific. In such situations, specific objects to be evaluated could be selected in correspondence to the purpose. Two examples of specific purposes are: (1) assessing the principal's capacity to make the school accountable to parents and (2) judging the principal's knowledge and skills regarding managing the school's finances effectively and efficiently.

In the first example, specific objects that could be evaluated include (a) parental satisfaction with how well the principal works with teachers on students' problems and (b) parental reports about how well the principal has responded to requests that they have made of her/him. In the second example, where the specific evaluation purpose is determining the principal's knowledge and skills dealing with finances, the specific objects that could be evaluated include (a) how effectively (extent to which a goal has been achieved with a given amount of money) does the principal spend available funds and (b) how efficiently (the most benefit for the least cost) does the principal spend available funds.

The other situation is where the evaluation purpose is insufficiently specific. An example of such a situation might be learning about the role of the principal in improving the school climate. In such a situation, some goal-free, unstructured and open-ended exploration needs to be done. The needs are to become familiar with (1) the dimensions of the climate that are characteristic to the particular school whose principal is to be evaluated (e.g., a tradition of formal interpersonal communication; a high principal turnover) and (2) the principal's connections with any or all of these elements. Only after acquiring a closer familiarity with the situation, can more specific objects be

selected. Correspondingly, these could include (a) the extent to which each dimension of the school climate is central to the school overall climate and (b) what the principal does or does not do in relation to the dimension described in (a).

In many school districts no specific purpose for principal evaluation is established. Evaluation activities are mandatory, routinely conducted, and nonconsequential. Typical evaluation objects are articulated in broad terms, similar to the way items are listed in the principals' job descriptions. In such cases, a yes or a no answer to each item on the evaluation instrument is considered sufficient. At times evaluation is done by grading on a scale of ordinal numbers that have no verbal descriptions corresponding to them.

Over the years, scholarly attention has been paid to some of these job-description-related evaluation objects. Particularly close attention has been given to objects about which the principals themselves have reported and those that have included district-expected outcomes. These evaluation objects included such activities as carrying out responsibilities of instructional leadership, administration of pupil services, financial dealings, legal matters, and ethical principles. (Rebore and Walmsley, 2007). Others objects include making decisions, solving problems (Zey, 1992; Glasman, 1994), communicating, interpersonal relations, as well as behaviors and attitudes (Hoy and Miskel, 2008). In the past 30 years, in addition to the knowledge principals possess in these administrative areas, principals' corresponding skills in using the knowledge have also been considered as appropriate objects to evaluate (Glasman, 1979, 1985, 1986, 1994; Glasman and Glasman, 2007; Glasman and Nevo, 1988).

In addition, over the years, the general title (as it is recognized in textbooks and conversations among professionals) of the professional work in which principals engage has gradually changed. It used to be called administration. Now, it is referred to as leadership and management. Perhaps changing times and expectations were the general determinants of this change in wording. What has now become clear, however, is that governing schools today includes both leadership and management, where the two complement each other: guiding and handling, providing direction, and controlling the operation.

A Special Case of the Principal's Leadership as the Object to be Evaluated

Studying successful leaders and defining leadership as what leaders actually do, Bennis and Nanus (1985) suggest that the potential for effective leadership is more widespread than is generally believed. Nanus (1989) followed this up by asserting that institutional leaders need to possess mega skills if they are to shape institutions effectively. These mega skills include:

- *Farsightedness*. Future creating, operating in the future, always searching for possible opportunities and threats.
- *Mastery of change*. Skilled in regulating the direction and rhythm of the institution so that its changes match the external pace of events.
- *Ability to design the organization*. Determining the human context, establishing the policy guidelines, fixing the setting, designing the structure and support system, and implementing and making adjustments as needed.
- *Anticipatory learning*. Of what is important, how the world works, how to increase the clarities of sources of uncertainty, new options, how much lead time is available for making decisions.
- *Initiating skills*. Being curious, inquisitive, and genuinely interested in what clients say so as to feel their intentions, frustrations, and needs.
- *Mastery of independence*. Skills associated with inspiring people, learning to trust them, developing commitments, communicating, and seeking collaboration.
- *Possessing high standards of integrity*. Undertaking a sacred trust and communicating ethical and even ennobling visions and action.

For the purpose of evaluating school principals as institutional leaders, some adjustments need to be made to the list of mega skills that Nanus has offered. Nanus derives his suggestions from studying executives and skills that they need in order to succeed. The executives he studies typically lead organizations that are hierarchical in nature and whose product is clearly defined and accurately measured (financial profit). Schools are nonhierarchical organizations. There is neither agreement on what their central output is nor on the criteria with which to measure it. Power is highly distributed in schools. Besides, principals do not teach students in the classrooms. Principals might possess different degrees of the corresponding two mega-skills: (1) designing institutions and (2) being masters of change. However, principals' limited authority and their only indirect influence on student outputs imply that their evaluators must adjust accordingly the information that they decide to collect about them (e.g., Firestone and Riehl, 2005; Glasman and Glasman 2007).

In this connection, Firestone and Riehl (2005) focused on skills related to supporting teachers and incorporating communities as contexts of student learning. Glasman and Glasman (2007) focused on skills related to meeting eight more specific challenges that include a total of 23 areas that could describe the skills required by principals. The first major challenge is setting clearly defined and achievable annual goals. The second goal is enforcing safety. This could be viewed as a managerial challenge, but a closer look at the specific relevant tasks identifies leadership behaviors, such as becoming comfortable with acting from authority and helping teachers to feel the same way.

The third challenge is helping students to maximize the benefits they derive from school such as knowledge and social skills. The fourth challenge is helping to solve problems, primarily by reflecting on possible decisions and how to communicate them, learning how to market the school and learning to make decisions more from the heart (than what is customarily done). The fifth challenge involves taking fiduciary responsibility. The sixth challenge is providing comfort when needed. The seventh is adjudicating conflicts. The eighth challenge is making beneficial uses of student achievement data.

These and similar suggestions about objects (or principals' behaviors, attitudes, and activities) that might be evaluated have brought about controversies. Some disagreements focus on the possible lack of sufficient appropriateness of the object as it relates to the evaluation purpose. Other disagreements involve the questionable fairness about including a particular item over which principals have little or no control. In regard to the fairness issue, the most controversial expectation has been that, since some studies have shown that principals have positive effects on students, principals should be evaluated on the basis of these effects (e.g., Gonzales *et al.*, 2002: 274–277). Those who believe that such expectations are unfair point to numerous other studies that show that principals do not have positive effects on students or that positive effects were not found. Furthermore, it should be of note is that of the 60 studies reported in the Gonzales *et al.*, (2002) publication, 12 were theoretical arguments and 39 focused on only one case study. Only nine studies were literature reviews. In very few of the studies mentioned above did the data lend themselves to developing generalities. Those who do not accept such reviews as strong evidence that principals have positive effects on student achievement feel, therefore, that it is unjustified to evaluate principals on the basis of such effects.

Collecting Information about the Object and Performance Criteria

There are an infinite number of information items that might describe the object to be evaluated. The items that are selected must correspond to the evaluation purpose and describe a characteristic of the principal. The capacity to become aware of teachers' needs for the purpose of guiding and supporting them is an example of one of many items that relate to evaluating the instructional leadership of the principal. The possession and use of qualities that match those that are needed in a newly constructed school is an example of a set of items that relates to evaluating the suitability of a principal being appointed to head a new school. Information about the object to be evaluated could be provided by several sources, including evaluation instruments.

There are probably thousands of ways in which school districts might collect principal evaluation information. No recently published data exist that delineate specific ways of collecting evaluation data about principals. Such an effort would require developing a useful conceptual framework for guiding the collection of such data and a very large funding base and full cooperation by school districts to collect the data. Specific ideas about purpose-related principal characteristics to be evaluated have recently been offered in the context of outcome-based principal accountability (Gonzales *et al.*, 2002; Firestone and Riehl, 2005; Glasman and Glasman, 2007).

Evaluation of each object requires information that is specific to it. However, triangulations are possible and at times desirable. For example, consider the object to be evaluated interpreting conflicting demands to improve student learning. This constitutes a highly complex object. One type of information might include answers to questions such as what, how, and to whom the principal presents interpretations. Another type of information may be generated from instances in which the principal exchanges ideas about these accountabilities (say political and professional) with the staff or with superiors. But it would still be necessary to conceptualize meritorious information about interpretations and exchanging ideas that include direct links to the improvement of specific student-learning component of the object to be evaluated.

To be placed in a situation of conflicting demands to improve students, forces the principal to make difficult choices about what he/she considers a justified object on which to be evaluated. In this situation, one object on which to be evaluated includes responses to political accountability demands. Another object to be evaluated includes responses to teachers' professional accountability demands whose hallmark claim is that no one should evaluate them as professional educators on the basis of student end of the year test scores. Several other factors (besides what teachers do with students), such as those related to students' families, significantly contribute to the level of student achievement. In this situation of conflicting demands, an ethical principal cannot tell politicians what the school, the teachers, and her/himself really cannot do about certain aspects of student achievement.

Of note also is that there are numerous validity and reliability issues associated with testing student improvement itself. This raises another set of issues regarding the use of student improvement as the object on the basis of which principals are evaluated (e.g., Amrein and Berliner, 2002).

Another example of an object to be evaluated might be making beneficial uses of student achievement data. Here, too, it is not difficult to conceive of, gather data about, and judge data's merit requiring what principals can do as a result of looking at the data and discussing them with teachers, parents, district office personnel, and community

leaders. However, the problem is that many principals have had no training in these areas and, therefore, either do not engage in such activities or engage in them in a perfunctory level. It would be unfair to use these kinds of activities as an object on the basis of which principals would be evaluated. These are specialties and one needs training in order to practice them. Assessing student test results and making inferences from them is one of these important set of skills (Popham, 2006).

Delimitations, Implications, and Suggestions

This article does not include an analysis of any of the probably hundreds of existing formal or informal principal evaluation procedures. Neither does it cover the probably hundreds of ideas that conceptualize and design specific principal evaluations.

In order to make significant progress in thinking about and performing useful principal evaluations, we need to interconnect evaluation knowledge and skills with those of school principals. Here is a suggested way to do it. Begin with evaluation, narrow it down to evaluation in education, and further narrow it down to evaluation of educational personnel. Simultaneously, start with education, narrow it down to educational leadership and management, and further narrow it down to school principals. At the end of these two imaginary lines, they intersect each other. This intersection is crucial to the understanding that although the cultures of those two theories and practices have been different, they must merge if the dual theory/practice of principal evaluation is to expand and improve.

We believe that the following might serve as a useful first step:

1. School principals as leaders/managers must become experts in certain specializations (such as analyzing student test scores, dealing with financial matters, and handling violence) while maintaining their generalist roles.
2. Evaluation of school principals for educational accountability involves primarily the assessment of what they say and do in attempting to enhance student outcomes.
3. Because of local contextual variables, principal accountability-related evaluations vary.
4. Nonconsequential accountability demands to evaluate principals might constitute a simple procedure involving asking an evaluator or the principal her/himself to respond to a few questions with a yes or a no answer.
5. Parental requests to evaluate principals for information and discussion purposes might involve more detailed and also partly open-ended evaluation questions.
6. School outcome-based principal evaluations might include information about some school parameters that

are independent of the principals (student backgrounds, teacher availability, and salary schedule) and information about what principals say and do to enhance the outcomes in light of those parameters.

7. Finally, the evaluation of principals on their performance to enhance the climate of their schools and on the growth of all students in their schools might include a principal-initiated strategic plan (written with her/his faculty and other service providers in the school) for initiating, maintaining, and evaluating a multiyear set of relevant activities. This would require a special financial grant.

See also: Internal and External Evaluation; School Based Evaluation: Purposes, Protocols and Processes; School Inspection/External School Evaluation; The Role of Stakeholders in Educational Evaluation.

Bibliography

- Amrein, A. L. and Berliner, D. C. (2002). High-stakes testing, uncertainty and student learning. *Educational Policy Analysis Archives* **10**(18), 2–71.
- Bennis, W. and Nanus, B. (1985). *Leaders: Their Strategies for Taking Charge*. New York: Harper and Row.
- Council of Chief State School Officers (2005). *Interstate School Leader Consortium Standards for School Leaders*. Washington, DC.
- DeMoss, K. (2002). Leadership styles and high-stakes testing: Principals make a difference. *Education and Urban Society* **35**(1), 111–132.
- Firestone, W. A. and Riehl, C. (eds.) (2005). *A New Agenda for Educational Research*. New York: Teachers College.
- Fowler, F. C. (2000). *Policy Studies for Educational Leaders*. New York: Prentice Hall.
- Glasman, N. S. (1979). A perspective of evaluation as an administrative function in education. *Education Evaluation and Policy Analysis* **1**(5), 39–44.
- Glasman, N. S. (ed.) (1985). *Special Issue: Evaluation as a Management Tool*. *Studies in Educational Evaluation* **11**(2), 119–238.
- Glasman, N. S. (1986). *Evaluation Based Leadership*. Albany, NY: State University of New York Press.
- Glasman, N. S. (1994). *Making Better Decisions about School Problems*. Newbury Park, CA: Sage.
- Glasman, N. S. and Glasman, L. D. (2007). *The Expert School Principal*. Lanham, MD: Rowman and Littlefield Education.
- Glasman, N. S. and Heck, R. (1996). Role-based evaluation of principals: Developing an appraisal system. In Leithwood, K., Chapman, J., Corson, P., Hallinger, P., and Hart, A. (eds.) *International Handbook of Educational Leadership and Administration*, pp 369–394. Boston, MA: Kluwer Academic Publishers.
- Glasman, N. S. and Heck, R. (2002). Principal evaluation in the United States. In Stufflebeam, D. and Kellaghan, T. (eds.) *International Handbook on Educational Evaluation*, pp 630–656. Boston, MA: Kluwer Academic Publishers.
- Glasman, N. S. and Nevo, D. (1988). *Evaluation in Decision Making: The Case of School Administration*. Boston, MA: Kluwer.
- Gonzales, M., Glasman, N. S., and Glasman, L. D. (2002). Daring to link principal preparation programs to student achievement in schools. *Leadership and Policy in Schools* **1**(3), 265–283.
- Hoy, W. K. and Miskel, C. G. (2008). *Educational Administration: Theory, Practice and Research*. Boston, MA: McGraw Hill.

- Nanus, B. (1989). *The Leader's Edge*. Chicago, IL: Contemporary Books.
- Popham, W. J. (2006). *Assessment for Educational Leaders*. Boston, MA: Pearson.
- Rebore, R. W. and Walmsley, A. L. E. (2007). *An Evidence-Based Approach to the Practice of Educational Leadership*. Boston, MA: Pearson.
- The Joint Committee on Standards for Educational Evaluation and Stufflebeam, D. L. (1988). *The Personnel Evaluation Standards*. Newbury Park, CA: Sage.

- Wong, K. K., Nicoretta, A., and Guthrie, J. W. (2007). *Successful Schools and Educational Accountability*. Boston, MA: Pearson.
- Zey, M. (ed.) (1992). *Decision Making*. Newbury Park, CA: Sage.

Further Reading

- Glasman, N. S. and DelVecchio, V. (2004). *Leadership for the Rest of Us*. Morrow Bay, CA: Reference Desk Books.

Program Evaluation

C A Christie, University of California, Los Angeles, California, USA
L A Fierro, Claremont Graduate University, Claremont, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Evaluation is the process of characterizing and appraising something of interest. In everyday life, we evaluate naturally. For example, we make evaluative judgments about the type of milk we buy in relationship to a set of standards – fat content, calories, taste, etc. This type of everyday evaluation is distinct from formal evaluation, which involves the systematic study of some evaluand (the object of the evaluation) to address questions related to the merit, worth, value, and impact of the evaluand. While in education we evaluate personnel, products, and programs, most of the theoretical, conceptual, and practical evaluation work that has been done over the years is grounded in evaluation of programs. Thus, the field that is now known as evaluation has evolved primarily from program evaluation.

In education, program evaluation is mostly concerned with the study of curriculum, programs, and policies. These programs can be examined in relation to their process, outcomes, and impact. The primary purpose of a process evaluation is to examine the development and implementation of the program. This information helps to explain what program components worked and why. In addition, this information is useful for monitoring program implementation, and more generally for program improvement. A process evaluation focuses on what services were provided to whom and how, and describes how the program was implemented, who was involved and what problems were experienced.

Gomby and Larson describe the evaluation of a complex school-linked service collaboration among a number of child-serving agencies and suggest several roles for a process evaluation. For example, a process evaluation might examine how the agencies interacted during implementation and document changes in relationships among social and health-service providers, educators, and client families. A process evaluation might also document system change as evidenced by new intake procedures, new forms, memoranda of understanding, or interagency linkage agreements. In some cases, changes in relationships and systems may be specifically planned goals of the program. In other cases, such changes may only be intermediate outcomes accomplished *en route* to the outcome of most interest to the program – changes in the students.

Outcome evaluation examines the program's benefits to recipients – the direct effects of the program on participants. For example, an outcome evaluation would examine whether participants in a 3-month program designed to teach adolescents to practice safer sex can demonstrate

the skills successfully. This type of evaluation is not unlike what happens when a teacher administers a test before and after a unit, to make sure the students have learned the material. However, the scope of an outcome evaluation can extend beyond knowledge or attitudes to examine the immediate behavioral effects of programs.

Impact evaluations look beyond the immediate results of the program to identify possible longer-term effects. An impact evaluation typically examines whether a program's immediate positive effects on behavior were sustained over time. Many school districts limit their evaluations to process and outcome studies. However, there are other agencies that are interested in and have the resources to examine whether program activities affect participants over time. For example, in the USA, there has been a desire to understand the longer-term impact of early childhood education programs such as Head Start on students' academic skills over time. Thus, academic outcomes of students who attended Head Start have been studied through high school and beyond.

A significant proportion of educational program evaluation takes place in traditional school settings (e.g., K-12 schools and higher education). In the USA, federal, state, and local departments of education support a great number of studies investigating the impact of educational programs so that policy and other decision makers can modify or enhance existing programs and initiate new education initiatives. Information generated from educational program evaluations can be used formatively, meaning that the information generated is used to improve programs on an ongoing basis. Summative evaluations may also be conducted, generating information that enables schools to determine, for example, which reading curriculum has the greatest impact on student literacy scores. Hence, summative evaluation conducted in the school setting is an endpoint evaluation of a particular program that has specific and expected outcomes.

Ralph Tyler conducted one of the first large-scale educational evaluations in the USA. Known as The Eight Year Study (1939–41), this longitudinal study of progressive education through the 4 years of secondary school and the 4 years of college is still considered to be one of the most significant evaluations of the twentieth century. Tyler's work on this study has been described as one of the best available examples of how evaluators can work cooperatively with teachers to clarify instructional objectives and develop indicators of students' continuous progress on a range of learning outcomes.

While other important studies of school-based programs were conducted during the 1940s and 1950s, it was not until the Great Society legislation of the 1960s that a broad and coordinated effort to evaluate federal programs was supported. This legislation introduced new programs into US classrooms and appropriated funds specifically for evaluating their impact. The same stipulation was placed upon other large social initiatives designed to support people and the communities in which they lived (e.g., Aid to Families and Dependent Children (AFDC)). This new demand for evaluation served as the foundation for the professionalization of evaluation, which continued to grow and mature during the 1970s. Reductions in federal spending during the 1980s translated into less funding for program evaluation and, as a result, a decline was experienced in the number of program evaluations conducted. This changed however in the 1990s, during which the US federal and state government spending for evaluation increased. This increase in available funds for program evaluation resulted in renewed energy for studying social and educational programs. This energy continues today.

Defining Program Evaluation: What It Is and How It Is Different from Research and Assessment

The evaluation literature includes a diversity of opinions about the ultimate purpose of program evaluation. While these differences are subtle, they are meaningful and become clearer when reviewing how some leading academicians describe program evaluation in their texts. For example, Michael Patton defines evaluation as the systematic collection of information about the characteristics, activities, and outcomes of programs to improve program effectiveness, make judgments about the program, or inform decisions about future programming. Peter Rossi, Mark Lipsey, and Howard Freeman not only highlight the importance of using information generated from evaluations for decision making, but also emphasize the role of evaluation in promoting social betterment. Others focus on the use of evaluation to identify causal outcomes of programs, that is, the systematic assessment of the extent to which a program caused the observed results. Amid these differences, one striking similarity emerges – the systematic nature of evaluative inquiry. To this end, evaluators employ the same methodological tools used by researchers such as study designs, data collection, and analysis techniques. Yet, research and evaluation differ in important ways: the origin of study question(s) and the purpose for which study information is gathered.

With respect to the origin of study questions, evaluation questions are typically elicited from program stakeholders, whereas research questions originate from

a researcher or research team. That is, researchers hypothesize about phenomena, and then develop research studies to answer their question. Regarding study findings, evaluation generates information for program improvement and decision making; that is, evaluation is decision oriented and is intended for use by individuals seeking the information. Evaluation studies, then, are designed to offer decision makers, such as program directors and policy-makers, systematic information about a program so that some actions can be taken, such as whether to continue a program. Alternatively, research is conclusion oriented. That is, the information generated from research studies is intended to enhance our general understanding of an area. The study of learning and cognition, for example, includes consideration of basic psychological theory, educational issues, and application. By and large, researchers generate hypotheses about phenomena with the purpose of producing information that can be generalized across context, people, and time, whereas evaluators typically serve an identified group of stakeholders (e.g., program staff) with the goal of providing information that will be useful for making decisions about specific programs within a relatively defined timeframe.

Evaluation is also different from assessment. While the terms assessment and evaluation are often used interchangeably in our everyday vernacular, in the context of education, assessment and evaluation are two distinct (although related) activities. Assessments are tests that are developed and administered to measure student performance. They are often used to measure specific student outcomes in the context of a broader educational program evaluation. An assessment system consists of some combination of norm-referenced, criterion-referenced, alternative, and classroom assessments reported to inform decisions about students, schools, districts, or states. What distinguishes an assessment system from the more general program evaluation is the focus on testing; assessment systems use tests as the basis for decision making. In evaluation, information about program processes and outcomes is generated for decision making using any combination of quantitative and qualitative data-collection techniques, which may or may not include testing.

Models and Frameworks for Program Evaluation

In 1967, Edward Suchman, a sociologist at the University of Pittsburgh, wrote a founding book on evaluation research citing Campbell and Stanley's 1963 manuscript, *Experimental and Quasi-Experimental Designs for Research* as the appropriate guide for developing evaluation designs. This citation brought experimental and quasi-experimental designs to the center of evaluation. Given that most of the evaluations at the time were being conducted by

university-based social scientists, these designs were valued as scientific and thus, desirable. Rather quickly, however, many individuals conducting evaluations began to recognize that, for a variety of reasons, experimental designs were very difficult to implement in educational settings. The school context, and the districts in which they operated, were not necessarily conducive to the Campbell and Stanley designs, particularly the randomized experiment. Evaluators acknowledged that in many contexts, experiments are not achievable, ethical, or desirable. Even the alternatives to the true experiment, quasi-experimental designs, that were developed to deal with the messy world of field research, were not always practical or popular. Alternative approaches for conducting evaluation were developed in response to these critiques and observations. Today, a diverse set of ideas and beliefs exist about how to best conduct evaluation in the educational setting.

The vast majority of evaluation approaches (referred to as models, frameworks, or theories) offer principles, rationales, and organizational structures for the procedural choices made by evaluators thereby orienting them to the issues and problems with which they must deal. By and large, they are qualitative models, points of view, persuasions, and approaches to the process of evaluation, which are intended to guide practice rather than explain phenomena. The US Centers for Disease Control and Prevention (CDC) developed a framework for program evaluation with input from hundreds of professionals and practitioners working in evaluation and research in universities as well as local communities. This framework describes the general steps involved in designing and conducting evaluations of social and educational programs. Although developed by a public health agency, the steps are generic and easily adapted for use in a variety of program contexts, particularly educational settings.

The framework consists of six-steps:

1. engage stakeholders,
2. describe the program,
3. focus the evaluation,
4. gather credible evidence,
5. justify conclusions, and
6. ensure use and share lessons learned (see **Figure 1**).

During each of these steps, evaluators are encouraged to apply the program-evaluation standards of utility, feasibility, propriety, and accuracy as described by the Joint Committee on Standards for Educational Evaluation. These standards assist evaluators in making difficult decisions encountered in evaluation practice; therefore, it is anticipated that use of these standards will result in high-quality evaluations. The linear arrangement of these steps is intended to clarify how program evaluations are structured and to emphasize the importance of fully addressing each step prior to making final determinations in subsequent steps. However, it is recognized that some of

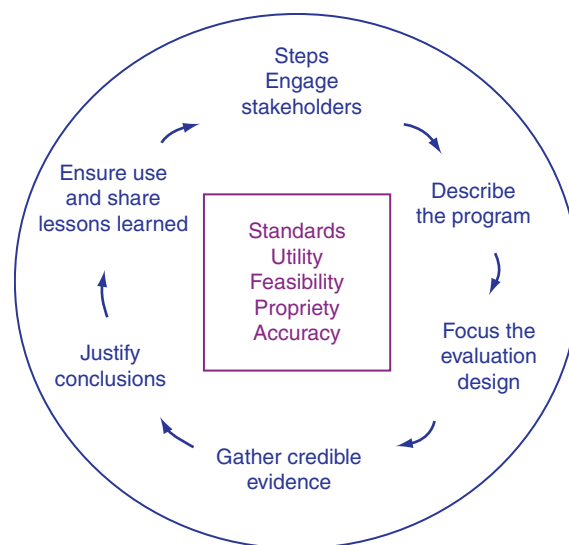


Figure 1 Centers for Disease Control and Prevention (CDC) evaluation framework.

the steps residing at the end of the framework may be addressed earlier in the process (e.g., how the evaluative findings will be used is often considered when engaging stakeholders).

Engaging Stakeholders

Stakeholders are people who have a vested interest in the evaluation findings. However, the level of interest stakeholders have in the program or its evaluation can differ. For this reason, stakeholders are sometimes categorized based on their relationship to the evaluation: primary stakeholders are in a position to use findings from an evaluation to alter a program's course (e.g., program designers and program staff); secondary stakeholders are likely to be affected by programmatic changes (e.g., teachers, parents, and students), and tertiary stakeholders may be interested in the findings from the evaluation but are not directly impacted by its results (e.g., superintendents of other school districts interested in adopting the program).

Two key considerations associated with stakeholder engagement include when to involve stakeholders and which stakeholders to involve. Program evaluators vary in the breadth with which they engage stakeholders – some evaluators focus solely on including primary stakeholders whereas others make extensive efforts to engage secondary and tertiary stakeholders. When making determinations about which stakeholders to involve, it is helpful to consider how nontraditional partners may contribute to the evaluation process. For example, it may be tempting to limit stakeholder involvement to individuals who directly influence the program (e.g., program managers), yet, their experiences may differ greatly from those who receive program services. It may also seem counterintuitive to

involve individuals who are not supportive of the program. However, engaging these nontraditional evaluation stakeholders may result in a richer program description, identification of important evaluation questions that would have been overlooked, and lend greater credibility to evaluation findings to a broader audience. Despite the potential value that each stakeholder perspective can bring to the table, evaluators need to carefully balance the value of adding more stakeholders with the need to make progress in the evaluation.

Program evaluators also vary with respect to the depth with which they engage stakeholders. Some evaluators involve stakeholders in every step of an evaluation, including data collection, analysis, and interpretation whereas others limit the scope of stakeholder involvement to the early stages of an evaluation (e.g., defining the program and determining evaluation questions). This variation in practice is a result of evaluator characteristics (e.g., the evaluator's beliefs about the ultimate purpose of evaluation or the evaluator's training) and logistical considerations associated with the evaluation (e.g., deadlines, funding levels, or access to stakeholders).

Describing the Program

Studies of evaluation practice show that it is common for program evaluators to engage stakeholders in developing

a description of the program. A common technique used to do this is logic modeling, in which a pictorial description is developed with stakeholders to clearly articulate how program activities will lead to fulfillment of programmatic goals. A general depiction of a logic model is provided in **Figure 2**.

When developing a logic model, evaluators may review existing program documents, conduct observations, and meet multiple times with stakeholders to obtain information from various perspectives about what constitutes a successful program. Logic models vary in their composition and structure; however, common components include inputs, activities, outputs, as well as short-term, intermediate, and long-term outcomes. Logic models sometimes also include a description of programmatic assumptions and denote the external forces that have the potential to impede or enhance progress toward reaching the stated program outcomes.

While developing the logic model, it may be discovered that the theory underlying the program is underdeveloped, unclear, or is unlikely to produce the anticipated results based upon findings from previous research studies. In these instances, the program may be modified prior to embarking upon an evaluation. In the event that the evaluation continues, as is typically the case, the presumed causal pathways expressed in the logic model can be helpful in generating appropriate and useful evaluation

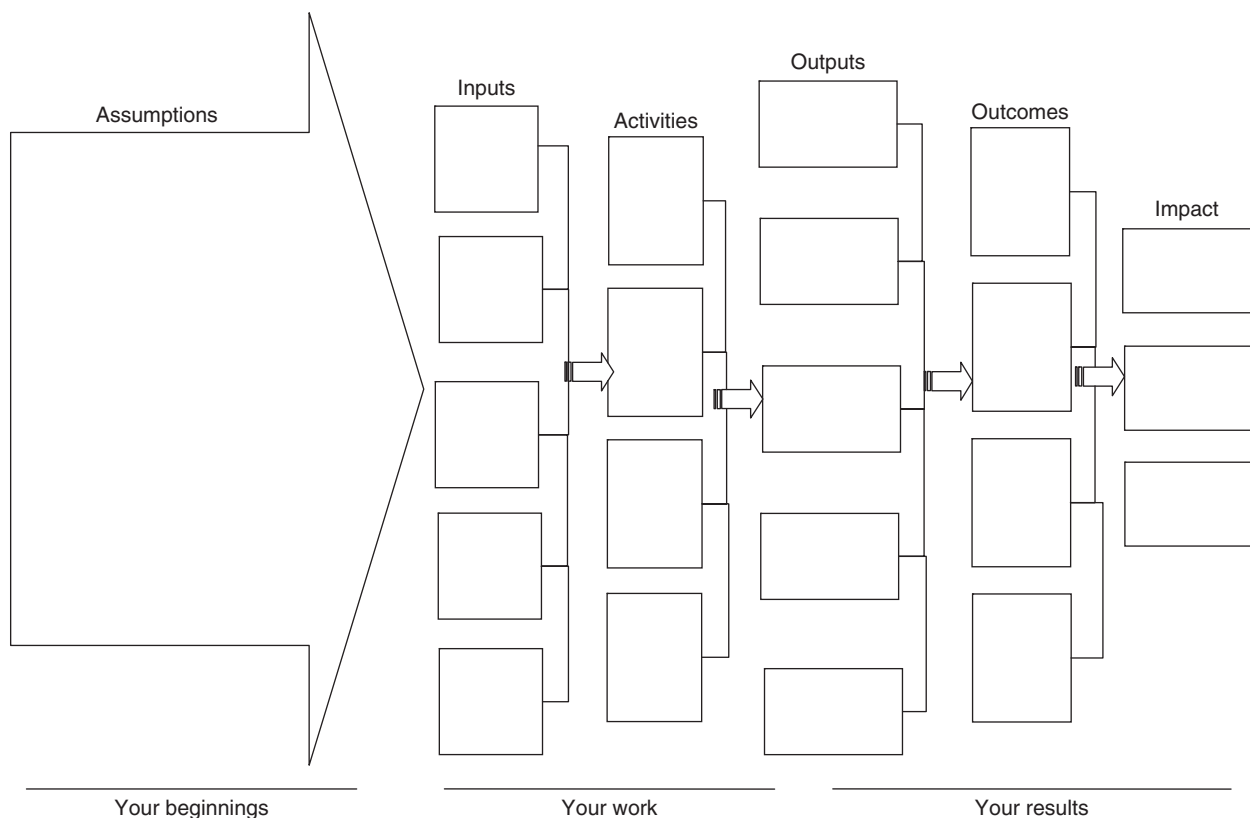


Figure 2 Logic model template.

questions, selecting an evaluation design, identifying appropriate time periods to collect data on the variables in the model, and ultimately in communicating findings. Logic modeling is only one approach to describing a program; related techniques include but are not limited to program theories, action models, and impact models, all of which are discussed as part of an approach to evaluation known as theory-driven evaluation and described in detail by evaluation scholars such as Peter Rossi, Huey-Tsyh Chen, and Stewart Donaldson.

In addition to developing a logic model, or a similar programmatic representation, evaluators may also work with stakeholders to elucidate and document additional factors that are important to fully describing the program. Specifically, it is important to place the description of the program itself within a broader context. Clearly articulating the need(s) addressed by the program, the context in which it resides (e.g., budgetary considerations or political considerations), the length of time the program has been in operation (i.e., whether the program is in a planning or implementation phase), and the degree of interconnectedness that exists between this program and related endeavors can help the evaluator and stakeholders in designing, implementing, and interpreting the findings and implications of the program evaluation.

Focusing the Evaluation

When focusing the evaluation, evaluators work with stakeholders to clarify the intended purpose of the evaluation, generate a list of clearly articulated evaluation questions that align with the intended purpose, prioritize the questions on this list, and select an appropriate evaluation design. This process requires a thorough understanding of why the evaluation is being commissioned and how stakeholders who are in a position to use evaluation findings actually intend to use the findings. During this step, the evaluator works with stakeholders to carefully balance considerations based upon the program evaluation standards of utility, feasibility, propriety, and accuracy.

As previously mentioned, the logic model developed with stakeholders can be used to generate a list of potential evaluation questions. As part of this process, it is likely that a number of evaluation questions will be raised; however, practical considerations, such as the length of time the program has been in operation and the resources available to conduct the evaluation (i.e., staff and money), make it unlikely (and in some cases inappropriate) to address all of the questions proposed. After a limited number of meaningful and useful evaluation questions have been prioritized with the stakeholders, an evaluation design that is capable of addressing these questions is selected. Evaluation designs include those that are experimental (e.g., randomized controlled trials),

quasi-experimental (e.g., pretest–posttest control group), or nonexperimental (e.g., ethnography) in nature.

Gathering Credible Evidence

During this stage of the program evaluation, consideration is given to the types of data that will be collected as part of the study as well as the specific data-collection methodologies that will be employed. In the previous step, the evaluator worked with the stakeholders to articulate the purpose of the evaluation and to generate and select a set of clearly worded evaluation questions. The data collected should contribute directly to answering the evaluation questions selected and should produce credible evidence to answer them.

Data-collection strategies used by the evaluator may include the collection of new data (i.e., primary data) through such methods as surveys, interviews, observations, or focus groups. Evaluators may also choose to use data that is already available (i.e., secondary data), which might include archival information from existing electronic or paper records, or analyzing data that is used to monitor, for example, student progress (e.g., student-assessment data). The credibility of data gathered through these methods can be affected by the procedures used to train individuals in data collection, the validity of the instruments used to gather data, the reliability of the measures used, and the quality-control procedures used to monitor data integrity and accuracy. Additionally, it is important for evaluators to adhere to ethical standards associated with the protection of data, and to maintain confidentiality and anonymity in instances where these protections have been assured.

The credibility of evaluation evidence is, at least in part, in the eye of the beholder. To develop an evaluation that is regarded as credible to the stakeholders, many evaluators will work with stakeholders to gain a clear understanding as to what constitutes credible evidence from their perspectives. Stakeholders may have a preference for the type of design selected in the previous step (e.g., view randomized controlled trials as the gold standard) as well as the type of data that are collected for the evaluation (e.g., qualitative, quantitative, or both). Engaging stakeholders in discussions to better understand the expectations they hold about what types of evaluative data will be collected is an important step in an evaluation, and can help to increase the likelihood that evaluative findings will be used.

Justifying Conclusions

After data have been collected the next step is to analyze, synthesize, and interpret the findings from the evaluation. To prepare for this step, evaluators will often work with stakeholders during the design phase to determine what level of achievement constitutes programmatic success.

These shared value statements form the standards that can be used to judge the evidence produced through the evaluation. When interpreting the findings, the evaluator and/or stakeholders may embark on a process in which they critically examine the proposed reasons for why the specific evaluative findings emerged and attempt to find alternative rationales (e.g., plausible rival alternative hypotheses). Additionally, they may choose to see how their results measure up to similar indicators of program performance – such as their performance on the same indicator in a previous time period or an aggregate measure of performance at a state or national level.

Irrespective of the method of interpretation used, it is important to consider what the practical implications of the findings are for the program and how the findings can be used to make changes or inform decisions about the future of the program. Any next steps or recommendations made during discussions with stakeholders should be firmly grounded in the analysis and synthesis performed on the credible evidence gathered in the previous step. Moreover, findings from a program evaluation should be applied in other settings with caution. Generalizability of the findings should be done in accordance with the limits imposed by the evaluation design and statistical sampling techniques employed (if any).

Ensuring Use and Sharing Lessons Learned

Although ensuring use and sharing lessons learned reside within the final step of the framework, considerations about how to facilitate the use of evaluative findings should begin early in the process of developing the evaluation and continue throughout the evaluation process. For example, when focusing the evaluation, it should become clear as to how the stakeholders ultimately anticipate using findings from the evaluation as well as their information needs. Thus, the evaluator should work early in the evaluation process to determine the types of decisions that might result from the evaluation. For example, will the information from the study be used to determine whether the program will be continued or if it may be implemented in other geographic regions and if so what changes might need to be made to the program so that implementation elsewhere is successful. This interaction aids the evaluator in tailoring the evaluation in a manner that increases the likelihood that findings will be used when the evaluation is complete.

Other steps can be taken throughout the evaluation process to prepare stakeholders for using the evaluation results. For example, interim findings or information about the progress that is being made in the evaluation may be shared with the appropriate stakeholder group(s). These interactions can help to prevent surprising the stakeholders with unanticipated findings, which may lead to stakeholders rejecting the final results. Additionally, the

evaluator can engage stakeholders in mock scenarios while the evaluation is underway. During these scenarios, the evaluator presents fallacious, but realistic, evaluative findings to the stakeholders and works with them to think through how they would interpret these findings and take action.

Another technique used to facilitate the use of evaluation findings is the development of a communication and reporting plan that is tailored to stakeholder needs. A communication and reporting plan clearly describes when, how, and with whom communications about the evaluation will be shared and identifies the specific methods that will be used to communicate with stakeholders. Potential communication methods include but are not limited to hard copy reports, presentations, e-mails, working sessions, and newsletters. As is demonstrated through these examples, communications range in their level of formality (i.e., informal emails to formal presentations) as well as in their level of interaction (i.e., highly interactive working sessions to static newsletters).

Summary/Conclusion

The evaluation of social and educational programs play an important role in understanding which programs are effective and why. A range of approaches has been developed to help guide evaluators in the conduct of program studies, which are intended to generate reliable, valid, and relevant information for decision makers to use when determining which programs to revise, expand, or terminate. To this end, it is critical that evaluation information is timely, meaningful, credible, and accessible to all that have a stake in both the program and its evaluation.

See also: Cultural Issues that Can Affect the Validity of Educational Evaluations; Evaluation Methodology; Evaluation Reporting and Communicating; Evaluation Use; Moral and Ethical Issues in Evaluation; The Role of Stakeholders in Educational Evaluation.

Further Reading

- Centers for Disease Control and Prevention (1999). Framework for program evaluation in public health. *Morbidity and Mortality Weekly Report* 48(No. RR-11), 1–40.
- Chen, H. (1990). *Theory-Driven Evaluations*. Newbury, CA: Sage.
- Chen, H. (2005). *Practical Program Evaluation: Assessing and Improving Planning, Implementation, and Effectiveness*. Thousand Oaks: Sage.
- Donaldson, S. I. (2007). *Program Theory-Driven Evaluation Science: Strategies and Approaches*. New York: Erlbaum.
- Donaldson, S. I., Christie, C. A., and Mark, M. (eds.) (2008). *What Counts as Credible Evidence in Evaluation and Evidence-Based Practice?* Thousand Oaks, CA: Sage.
- Fitzpatrick, J., Christie, C. A., and Mark, M. (eds.) (2008). *Evaluation in Action: Interviews with Expert Evaluators*. Thousand Oaks, CA: Sage.

- Fitzpatrick, J. L., Sanders, J. R., and Worthen, B. R. (2004). *Program Evaluation: Alternative Approaches and Practical Guidelines*, 3rd edn. Boston, MA: Pearson Education.
- Gomby, D. S. and Larson, C. S. (1992). Evaluation of school-linked services. *Future of Children* 2, 68–84.
- Joint Committee on Standards for Educational Evaluation (1994). *Program Evaluation Standards: How to Assess Evaluations of Educational Programs*, 2nd edn. Thousand Oaks: Sage.
- Madaus, G. and Stufflebeam, D. (eds.) (1989). *Educational Evaluation: Classic Works of Ralph W. Tyler*. Dordrecht: Kluwer Academic Publishers.
- Patton, M. Q. (1997). *Utilization Focused Evaluation: The New Century Text*, 3rd edn. Thousand Oaks: Sage.
- Patton, M. Q. (2008). *Utilizational Focused Evaluation*, 4th edn. Thousand Oaks, CA: Sage.
- Rossi, P., Lipsey, M., and Freeman, H. (2004a). *Evaluation: A Systematic Approach*. Thousand Oaks, CA: Sage.
- Rossi, P. H., Lipsey, M. W., and Freeman, H. E. (2004b). *Evaluation: A Systematic Approach*, 7th edn. Thousand Oaks, CA: Sage.
- Russ-Eft, D. and Preskill, H. (2001). *Evaluation in Organizations: A Systematic Approach to Enhancing Learning, Performance, and Change*. New York: Basic Books.
- Shadish, W. R. Jr., Cook, T. D., and Leviton, L. C. (1991). *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, CA: Sage.
- Torres, R. T., Preskill, H., and Piontek, M. E. (2005). *Evaluating Strategies for Communicating and Reporting: Enhancing Learning in Organizations*, 2nd edn. Thousand Oaks, CA: Sage.
- US Department of Health and Human Services. Centers for Disease Control and Prevention. Office of the Director, Office of Strategy and Innovation (2005). *Introduction to Program Evaluation for Public Health Programs: A Self-Study Guide*. Atlanta, GA: Centers for Disease Control and Prevention.
- Wholey, J. S., Hatry, H. P., and Newcomer, K. E. (eds.) (2004). *Handbook of Practical Program Evaluation*, 2nd edn. San Francisco, CA: Jossey-Bass.
- W. K. Kellogg Foundation (1998). *Evaluation Handbook*. <http://www.wkkf.org/Pubs/Tools/Evaluation/Pub770.pdf> (accessed September 2009).

Relevant Websites

- <http://www.eval.org> – American Evaluation Association.
- <http://www.cdc.gov> – Center for Disease Control and Prevention's Evaluation Working Group.
- <http://www.wmich.edu> – Western Michigan University: The Evaluation Center.

School Based Evaluation: Purposes, Protocols and Processes

J MacBeath, University of Cambridge, Cambridge, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Over a century ago, Arnold Tomkins, a New York administrator wrote the following:

The organisation of the school must be kept mobile to its inner life. To one who is accustomed to wind up the machine and trust it to run for fixed periods, this constantly shifting shape of things will seem unsafe and troublesome. And troublesome it is, for no fixed plan can be followed; no two schools are alike; and the same school is shifting, requiring constant attention and nimble judgement on the part of the school leader. (Tomkins, 1895: 4)

This is an apt description of school-based evaluation, the purpose of which is to keep a school mobile to its inner life, a continuing process of reflection and search for improvement. A school's values and the priorities that count are implicit in the way people (teachers, students, and administrators) think and talk about their work, but evaluation is what they do to make their practice explicit, discussable, and subject to a valid form of measurement. The latter is a term that has to be approached with some caution as it has connotations with a mechanistic or numerical accounting, whereas exploring the inner life of schools and classrooms is a much more complex and messy business. This is not to gainsay the need for rigor; indeed, the greater the recognition given to the subtleties and complexities of measurement, the more rigorous the approach is likely to be.

Evidence from countries in many different parts of the world leads to the same conclusion – schools that know themselves, that adopt a systematic and critical approach to evidence are schools that are able to take charge of change rather than be controlled by it. Self-evaluating schools are likely to be more effective and to improve more rapidly than ones that rely on external sources to validate their quality (Stoll and Myers, 1997; Ouston and Davies, 1998; Rosenthal, 2001).

What constitutes school-based evaluation is, however, problematic. It has been defined in a number of different ways and enjoys a confusing lexicon of terms – self-assessment, self-review, audit and quality assurance, for example. It is a useful point of departure however, to make some distinctions among this varied terminology:

1. *Audit*. The origins of audit lie in financial practice designed to ensure that the organization's books are kept rigorously and ethically. It has also connotations of

stock taking and in a school context, implies an overview of the educational stock or available resource. It is essentially summative in nature.

2. *Quality assurance*. This is a form of audit. The term denotes a systematic examination of quality, usually by an external body and, like audit, is essentially geared to accountability in return for the investment (financial and moral) in the school. It is primarily summative although it may be followed by recommendations for improvement.
3. *Self-review*. This tends to be used synonymous with self-evaluation although a distinction may be made between one-off review or snapshot and self-evaluation as an ongoing process. The term tends to be used summatively but may also be formative.
4. *Self-assessment*. This is a term also used interchangeably with self-evaluation although in many regimes, important distinctions are made between assessment and evaluation, the former referring to judgments of student performance whereas the latter refers to broader judgments of organizational effectiveness. Both terms are used summatively and formatively.
5. *Inquiry*. This is a term with a North American resonance denoting an open-ended process, not necessarily tied down to prespecified criteria. A form of this, known as appreciative inquiry, sets out to understand the strengths of a school within its own frame of reference. It is essentially formative rather than summative.
6. *Self-evaluation*. This is a term with growing currency in European and Asia-Pacific countries. Interpretations of its form and purpose vary widely but the self in question applies to the school as an institution and is primarily a formative process. A simple singular definition is:

A process of reflection on practice, made systematic and transparent, with the aim of improving student, professional and organisational learning and rendering an account to key stakeholders.

Three Driving Motives for Self-Evaluation

There are essentially three driving motives for advocating or adopting self-evaluation. These are the improvement motive, the accountability motive, and the economic motive.

The improvement motive rests on a belief that, to build the school's capacity to respond to and manage change

cannot be achieved without a commitment to self-evaluation. Lack of self-knowledge and self-delusion work to the detriment of students and staff and mislead parents. Social capital is built through a continuing commitment to knowledge creation, drawing on the hidden capital within a school enabling it to become more intelligent than its individual members. Soo Hoo described one source in this way:

Somehow educators have forgotten the important connection between teachers and students. We listen to outside experts to inform us, and, consequently overlook the treasure in our very own backyards, the students. (Soo Hoo, 1993: 389)

The origins of the accountability motive, born out of school effectiveness research and the increasing availability of performance data, have acquired a sharper edge in an international climate in which nations are compared according to the performance of their students. As politicians and policymakers do not wish to be taken by surprise by external evaluations undertaken by the Organization for Economic Cooperation and Development (OECD) or by other influential agencies, systematic self-knowledge and reporting of achievements assumes a high priority.

There is a compelling economic rationale for school-based evaluation. It is much easier on the public purse for schools to provide performance evidence than for this to be the job of external policing bodies such as inspection teams and external review agencies. In England, the Gershon Report of 2004 recommended a saving of over 50 million pounds by, among other cost-cutting devices, shifting the responsibility from an inspectorate to schools, holding them responsible for their own self-evaluation (Gershon, 2004).

While the rationales for each of these three motives may ultimately serve the same end, there are nonetheless tensions among them. What a school deems to be an improvement priority may not accord with the accountability agenda or meet what are seen as the pressing professional-development needs of staff. Writing in a Canadian context, Ben Jaafar (2006) describes the tensions between economic bureaucratic accountability and ethical professional accountability. These can, she argues, be addressed by inquiry-based accountability. In this model evaluation at classroom, school and external levels are used as entry points for professional discussions about learning experiences, opportunities, and outcomes and the priorities for the young people that schools are expected to serve. Self-evaluation then functions as the handmaid of professional development so that staff become more self-aware, more reflective, and more self-critical by virtue of the tools they use to monitor their own professional growth and the accounts they render to their colleagues, communities, and paymasters.

What these varying approaches have in common comes down to some key ideas – that:

- schools themselves require some form of internal review of their own quality and effectiveness;
- internal processes of review/evaluation take account of evidence from a range of data sources and stakeholders;
- evaluation strives for consensus and common understanding while welcoming differences in perspective as growth promoting;
- it is forward looking and essentially formative in nature;
- its findings provide the base for development and improvement planning; and
- its concern for evidence promotes and strengthens both internal and external accountability.

These are, however, demanding principles and for many schools present a challenge that cannot easily be met without guidance and support from external sources.

The Need for Critical Friends

The notion of the self-evaluating, self-improving school may be more of an aspiration than a reality in a context where the weight of policy and the pace of change conspire to keep teachers preoccupied and make reflective and collegial dialog something of a luxury. In the United States, Apple (2006) argued that as teachers work in increasingly intensified conditions and, those who consult or collaborate with them have to act as story tellers and secretaries, intermediaries who help teachers and school leaders help give shape to their stories and enable their voices to be heard.

Such external support is crucial, conclude Baker *et al.* (1991), who compared schools in the United States which drew on external support and those that did not. Their finding received confirmation from research undertaken in Scotland between 1997 and 2000 in which researchers and critical friends worked with 80 Scottish schools, collecting and feeding back data to support schools in their own self-evaluation and improvement planning (MacBeath and Mortimore, 2001). Half of the schools had the support of a critical friend who advised on collection and use of data, feeding back findings and helping headteachers and staff to make sense of that information. The importance of informed, supportive, and challenging critical friendship was one of the main findings of that study.

One of the key foci common to these various studies was the schools' capacity for what Argyris (1978) has termed double-loop learning, that is, the willingness and ability to stand back from the processes of seeking evidence and data gathering to reflect more deeply on the evaluation experience itself. The double loop is one of the distinguishing marks of ongoing self-evaluation as against the more routinized process of self-inspection. While the

latter adopts a single-loop process of data collection – interpretation – target setting and improvement planning – the double loop offers an escape route from that cycle, keeping its own evaluation activity under continuing and critical review. It is in this respect that the visitor's eye view as the Icelanders call it, is significant.

The argument that schools need friends who both support and challenge now seems beyond dispute. Such friendship does not, however, guarantee an accountability to a government agenda, as critical friends may be drawn from universities, private agencies or colleagues in other schools, bringing with them no necessary allegiance to government targets or time scales. As a consequence, most administrations have seen the need for some form of marriage between internal, school-based evaluation and external review or inspection.

While a judicious combination of internal and external evaluation is widely seen as the way forward, aligning improvement, accountability and economics, no country can claim to have found the ideal relationship. There is, however, a strong emerging consensus that external review works best when there is robust self-evaluation in place and that effective self-evaluation is bolstered by well-designed external support. A 2004 study conducted by the Standing International Conference on Inspection (SICI) (SICI 2004) in Europe found that:

The school visits conducted as part of the project have shown that self-evaluation is most effective in countries that have the strongest external support to the process and thus have created a culture and climate for effective school self-evaluation.

Models of Self-Evaluation and External Review

The Norwegian academic, Trond Alvik identified three models which describe the relationship between internal and external evaluation, to which may be added a fourth, the supporting model.

Parallel systems are those in which the two systems run side by side, each with their own criteria and protocols. While schools pursue their own evaluation criteria, inspection or review carries out its own system, perhaps entirely independently of the school's own internal efforts. This is true of systems in which inspection is fairly new or in which self-evaluation is not yet widely developed as, for example, in Poland, Slovenia, or some Swiss cantons. In such systems, self-evaluation is entered into voluntarily and with a focus on improvement rather than accountability, while the balance of power rests entirely with the inspectorate who set the criteria, the nature, and timing of the visit and reporting of outcomes. France is archetypical in this. While there are schools with

their own approaches to whole-school evaluation, inspectors evaluate and grade classroom teachers' performance, a grade which carries high stakes for the teacher because it affects the speed at which his or her career progresses.

Parallel systems may, however, be the seedbed for the growth, as they mature developing a more coherent joined-up approach. This is dependant on the extent to which inspection regimes are willing to investigate and learn from school-led innovation while schools are able to adapt and learn from external protocols.

Supporting systems are those in which the role of the administration (district, state, or national body) is to provide support for schools in carrying out self-evaluation more effectively. These forms of support tend to be characterized by a power-sharing relationship and schools are involved in a more or less voluntary basis. As there is a greater sense of ownership on the part of the school, the stakes are low and decisions about public reporting tend to be negotiated. In Italy, for example, there are myriad bodies and clusters of schools pursuing their own forms of self-evaluation with support from universities and provincial advisers, with an explicit focus on improvement rather than accountability. Similar approaches may be found in Denmark (Larsen and Højdalthe, undated) the Czech Republic (SICI, 2001), and some German Lander (Stern and Dobrich, 1999), although in these states and countries there are moves toward a more sequential system. Accountability tends not to be emphasized, although internal, or professional, accountability may be seen as integral to the process of self-evaluation.

Sweden's Skolverket is a rare example of a supporting system at national level. The focus of external review is on classroom practice but in a collegial relationship, where the power dimension is minimized and inspectors neither criticize nor grade teachers. The inspection process is primarily one of professional discussion rather than summative judgment. Inspectors cannot tell teachers what to do, but they can direct them to best practices elsewhere. Their role is, through collaborative conversations with teachers, to stimulate school-wide improvement. This may be seen as a soft option but these conversations are challenging, drawing on sources of evidence, using results, statistics, and observation to engage the discussion with teaching staff.

Sequential systems are those in which external bodies follow on from a school's own evaluation and use that as the focus of their quality-assurance system. The excellence model in Singapore, adapted from a business excellence model, exemplifies one of the most comprehensive approaches to self-evaluation within a sequential paradigm. Schools are provided with extensive guidelines, indicators, and tools for self-appraisal which they use to evaluate themselves, their scorecard comprising a source of evidence for external review. Schools complete their own scorecard (SEMS), also available in an electronic

version, with a given set of criteria and allocation of marks under each of six key performance domains – cognitive, physical, esthetic, social and moral, student leadership, and student morale.

The sequential model is the model favored by the SICI, evidence for which can be found in Portugal, Ireland, England, and Scotland. There are also examples in a number of Australian states where inspection was firmly rejected by academics and teacher professional bodies in favor of a more supportive form of external review such as the continually evolving process of independent external verification in Victoria (Cairns *et al.*, 2009).

The model adopted in Hong Kong as SSE/ESR (school self-evaluation and external school review) is also exemplary of this model, evolving in response to independent research findings (MacBeath, 2007). After a systematic evaluation of the system over a 4-year period, the education bureau has moved to a system which is responsive to the school's own context and development priorities rather than working to a prescribed common template. In these various sequential systems, SSE precedes external review but, as in the case of Israel, for example, the sequence may be reversed with inspection providing the impetus to self-evaluation.

Cooperative systems are those in which external agencies cooperate with schools to develop a common approach to evaluation. It is in the last of the four types of evaluation above that the power balance is most equitable. As a cooperative system implies, the power issues are not covert but explicit from the outset so that there are no hidden agendas, no nasty surprises, or resentment. These systems are characterized by reciprocity, what the Dutch academic Van Leeuw calls "me-too-you-too" principle because the arrangement is voluntary and negotiated on a collegial footing. While the Swedish system has some of the hallmarks of a cooperative system, it is not cooperative in the sense of constituting a peer relationship in which the nature and conduct of the review is entirely open to negotiation and mutual consent.

In the USA, examples of this can be found at state level or district level. In Rhode Island, the School Accountability for Learning and Teaching (SALT) scheme takes as its focus, teachers' practice evaluated by teachers from another school. The review team is composed of practicing Rhode Island teachers together with a parent, an administrator, and a member of university staff. The team spends 4–5 days in the school, writes a report which is then negotiated with school staff, a process of which can be lengthy but is highly valued in teasing out evidence and the basis for judgments made. The team then draws up a compact for learning, the purposes of which are to ensure that school staff have the capacity to implement improvement. Strategic Inquiry in Trenton, New Jersey, is a cooperative process in which the external team act as critical friends, spending an intensive week in the school,

suspending judgment, getting to know the school by shadowing students, taking lunch with them and the staff, observing in classrooms, and participating in professional-development sessions. They report back to the school not with a set of categorical judgments but with a set of questions to stimulate further dialog and planning for improvement.

Another example of a cooperative model is KEYS, a self-evaluation program and one which schools can buy into. Developed at the University of Washington, it is a process involving a wide variety of stakeholders and vested interests – teachers unions, the school district, the community, as well as school leadership and taking as its starting point the individual interests of staff members. KEYS aims for honest appraisal through self-reflective renewal activity using a framework which holds up a mirror to the school, opens up issues of preexisting beliefs and patterned behavior, and invites staff to reconsider the assumptions on which they rest. One of the strengths of this approach is its factoring in of the changing context and dynamic of school, leadership turnover, community conflict, previously engaged projects, and trying to create conditions that pave the way for new ways of working.

Reciprocity and the Balance of Power

As can be seen from these examples, systems do not fall discretely into one of these four models (Alvik's three plus one). There is always a mix of purposes and matters of emphasis. The twin purposes of self-evaluation and inspection have provided a constant source of tension in different countries and official documents place the emphasis on different facets. However well chosen and rational the model devised, every regime is subject to political pressure whether from change of government, by virtue of the particular individuals who are in power, and in response to pressure groups such as professional bodies and teacher unions. There are also the highly influential supranational bodies such as the OECD, the United Nations Educational Scientific and Cultural Organization (UNESCO), and the European Commission, whose comparative data on school performance are treated with extreme seriousness by politicians and policy advisers. Factored into this mix are the media, watched increasingly closely by political parties for whom reelection depends on the nature of the stories told in the press and on television. Educators are aware that there is always a distance between espoused theory and theory in practice, between what is desirable and what is pragmatically possible and politically acceptable.

The balance of power between the inspection teams and school teams and the weight given to the external report and the internal narrative differ widely, although virtually by definition, a government body is likely to have

the last word. The last word of course, carries quite different significance in countries where the outcomes of review are publicly available, are competitive, and high stakes as against places where the data are seen purely for the school itself. It is the inequality of the relationship in many regimes that prompted the Dutch academic Leeuw (2001) to argue for a reciprocity of relationship, what he calls the “me-too-you-too” principle. In other words, if external evaluators are to make judgments about school or classroom practice, professional principles require that judgments be mutual, negotiated, and shared on an equal basis.

Questions of Power, Provenance, and Processes

The questions we may ask of these systems are, for example:

- The voluntary or statutory nature of the process – Are schools required to self-evaluate and to prove evidence of that or is that a matter of the school’s choosing?
- The balance of power between the school and the review body – Who has the final say? How much credence is given to the school’s own judgment?
- The low- or high-stakes nature of the process and outcomes – Do sanctions or tangible rewards follow from external review?
- The nature of central direction as against flexibility for the school self-determination – Who decides on the focus, agenda, and timing of the school’s report and the team’s visit?
- The constitution of the inspection/review team – Is it led by a government body, an independent agency, trained inspectors, lay people, or peers?
- The framework and structure of protocols, indicators, and success criteria to be used by the school – Who decides them?
- The frequency, length, and intensity of the external visit – How often, for how long, and with what advance notice?
- The focus of external review – Whole school, departments, or individual teachers?
- The transparency, public recognition, or confidentiality of the report – Is the report publicly accessible or restricted to particular stakeholders?
- The balance of essential purpose – Where is the explicit emphasis – accountability or improvement? What is the impact and feeling of those on the receiving end?

As self-evaluation becomes more embedded in school practice and increasingly mandated by governments, its form tends to move from grounded bottom-up processes to more tightly defined top-down formulae. In longstanding inspection regimes, such as in the UK countries, governments keen on standardization, devolve their own

frameworks to schools, providing the goals, criteria, and protocols with the consequence that schools begin to engage in a form of self-inspection. Self-inspection and self-evaluation may be characterized (**Figure 1**) as arising from quite different assumptions about the nature of improvement.

The paradox is that the more the governments provide the frameworks, indicators, and tools, the less inventive and spontaneous the process at school and classroom level. Self-evaluation becomes a ritual event, a form of audit in which senior leaders assume the role of an internal inspectorate applying a set of common criteria.

The previous chief inspector in England now permanent undersecretary at the Department for Children, Schools and Families (DCSF) said this:

I have always been cautious in saying that inspection causes improvement because, frankly, we don’t. But it has to be an important part of our thinking about inspection . . . To say inspection causes improvement is fundamentally unprovable. I think there are examples of where you have greater evidence of improvement being brought about by inspection, but again it’s still not quite the same as saying it causes it. For example, our monitoring of schools with special measures is not causing improvement but most headteachers say to us that the process of professional debate and discussion with HMI brings some real bite to the improvement process. I think it’s a bit too simplistic to say that either OFSTED does cause improvement or OFSTED doesn’t cause improvement. (MacBeath, 2006: 30)

The keywords in this statement are the process of professional debate and discussion.

This is where school-based evaluation comes into its own and where internal accountability is nurtured. As Elmore (2003: 17) has argued, without a strong sense of internal accountability, schools and teachers will always be subject to external pressures and remain reactive to externally driven change. They are more able to counteract the local, national, and international forces at work when there is shared understanding of the difference

Self-inspection	Self-evaluation
<p><i>Top down</i> A one-off event</p> <p>Provides a snapshot at a given time Is time-consuming Is more about accountability than improvement Applies a rigid framework Uses a set of predetermined criteria Creates resistance Can detract from learning and teaching Encourages playing safe</p>	<p><i>Bottom up</i> Is continuous and embedded in teachers’ work Is a moving and evolving picture Is time saving Is more about improvement than accountability Is flexible and spontaneous Uses, adapts, and creates relevant criteria Engages and involves people Improves learning and teaching Takes risks</p>

Figure 1 Self-inspection and self-evaluation.

between what they can and cannot do, but continuously push at the boundaries of the possible. In the best schools, change forces arise from the inside, from a deeply rooted commitment to what is important and of lasting value.

Bibliography

- Apple, M. (2006). *Markets, Standards, and Inequality – Keynote Address to International Congress on School Effectiveness and Improvement*, Fort Lauderdale, Florida, January 6th.
- Baker, P., Curtis, D., and Berenson, W. (1991). *Collaborative Opportunities to Build Better Schools*. Bloomington, IL: Illinois Association for Supervision and Curriculum Development.
- Ben Jaafar, S. (2006). From performance-based to inquiry-based accountability. *Brock Education* 16(2), 62–77.
- Cairns, L., Corrigan, D., and Forsyth, A. (2009). Accounting for school accountability and improvement in Victoria, Australia. *Paper delivered at the International Congress on School Effectiveness and Improvement*, Vancouver, 5–8 January.
- Gershon, P. (2004). *Releasing Resources to the Front Line*. London: HM Treasury.
- Larsen, J. O. and Højdalthe, R. Denmark: *First National Report, Quality School Evaluation in Denmark*. http://209.85.229.132/search?q=cache:HZgTfTsIC_kJ:reflective.bazaar.org/wp-content/stall_reflective_uploads/2006/07/RE-WP1dk.pdf+schools+self+evaluation+in+denmark&hl=en&ct=clnk&cd=2 (accessed August 2009).
- Leeuw, F. (2001). Reciprocity and the evaluation of educational quality: Assumptions and reality checks. *Keynote Paper for the European Union Congress*, Karlstad, Sweden, 2–4 April.
- MacBeath, J. (2006). *School Inspection and Self Evaluation: Working with the New Relationship*. London: Routledge.
- MacBeath, J. (2007). *The Impact Study on the Effectiveness of External School Review in Enhancing School Improvement through School Self Evaluation in Hong Kong: Final Report*. Hong Kong: Education Bureau.
- MacBeath, J. and Mortimore, P. (2000). *Improving School Effectiveness*. Buckingham: Open University Press.
- Ouston, J. and Davies, J. (1998). OfSTED and afterwards: Schools responses to inspection. In Earley, P. (ed.) *School Improvement After OfSTED Inspection: School and LEA Responses*, pp 126–139. London: Sage Publications.
- Rosenthal, L. (2001). *The Cost of Regulation in Education: Do School Inspections Improve School Quality?* Keele: Department of Economics, University of Keele.
- SICI (Standing International Conference of Central and General Inspectorates of Education) (2001). *School Self Evaluation in Thirteen European Countries*. Brussels: European Commission.
- SICI (Standing International Conference of Central and General Inspectorates of Education) (2004). *Evaluation of schools providing Compulsory Education in Europe*. Brussels: SICI.
- Soo Hoo, S. (1993). Students as partners in research and restructuring schools. *Educational Forum* 57, 386–393.
- Stern, C. and Dobrich, P. (1999). *Wie Gut ist Unsereschule? Network of Innovative School Systems*. Gutersloh: Bertlesmann Stiftung.
- Stoll, L. and Myers, K. (1997). *No Quick Fixes: Perspectives on Schools in Difficulty*. London: Falmer Press.
- Tomkins, A. (1895). *School Administrator*. New York quoted in Riley, K. (1895). *School Leadership and School Culture*. Washington: World Bank.

Further Reading

- Alvik, T. (1996). *Self Evaluation: What, Why, How, by Whom, for Whom, Collaborative Project Self-evaluation in School Development*. Dundee: CIDREE.
- Argyris, C. (1993). *Knowledge for Action: A Guide to Overcoming Barriers to Organizational Change*. San Francisco, CA: Jossey-Bass.
- Argyris, C. and Schön, D. (1978). *Organizational Learning: A Theory of Action Perspective*. Reading, MA: Addison Wesley.
- Cullingford, C. and Daniels, S. (1998). *The Effects of Ofsted Inspection on School Performance*. Huddersfield: University of Huddersfield.
- Department for Education and Skills (2004). *A New Relationship with Schools: School Improvement Partners' Brief*. London: DfES.
- Elmore, R. (2005). *Agency, Reciprocity, and Accountability in Democratic Education*. Boston, MA: Consortium for Policy Research in Education.
- Learmonth, J. (2000). *Inspection: What's in it for Schools?* London: Routledge.
- Love, A. J. (1998). *Internal Evaluation: Building Organizations from Within*. Newbury Park, CA: Sage.
- MacBeath, J. and Sugimine, H. (eds.) (2001). *Self Evaluation in the Global Classroom*. London: Routledge.
- National College of School Leadership (2005). *Self Evaluation: A Guide for School Leaders*. Nottingham: NCSL.
- Nevo, D. (ed.) (2002). *School Based Evaluation: An International Perspective*. Boston, MA: Elsevier.
- Pang, N. (2004). Developing schools through self-evaluation. In Sun-keung Pang, N., MacBeath, J., and McGlynn, A. (eds.) *Self-evaluation and School Development*. Hong Kong: The Chinese University of Hong Kong.
- Rogers, E. M. (1962). *Diffusion of Innovations*. New York: Free Press.

School Inspection/External School Evaluation

C Chapman, University of Manchester, Manchester, UK

P Earley, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Education quality – The quality of education received by students within an education system.

Specific factors influencing quality include the effectiveness of teaching and learning and leadership arrangements within a school.

External evaluation – The assessment of the quality of structures and processes operating within a school as judged by an outside agency or perspective.

Internal review/self-review – A review conducted by members of the organization to assess the effectiveness of structures and processes in place.

International education – Of relevance to more than one educational system. Drawing on examples from a range of systems.

Ofsted – The Office for Standards in Education is the inspection agency responsible for the systematic inspection of schools (and other providers of education and child care) in England.

School effectiveness – The extent to which a school adds value to students. An effective school is a school where students achieve more than might be expected after a range of background factors have been taken into account.

School evaluation – The evaluation of the effectiveness of the structures and processes in place within a school.

School improvement – The processes and strategies associated with improving student outcomes and creating the capacity for enhanced effectiveness.

School inspection – The process of examining a school's structures, processes, culture, and capacity for improvement.

Self-inspection – The process of examining or reviewing a schools' strengths and weaknesses conducted by members of the organization using an existing inspection framework rather than one devised by the school.

outcomes – but there is still considerable discussion about the best ways of achieving this. Among educationists and policymakers, the school improvement debate often centers around the relative importance of the processes of external review compared with those of internal school evaluation. Those advocating external evaluation systems associated with school inspection, whom we might refer to as hawks, see self-evaluation as soft-centered; they argue that without the hard-edged external probing that inspection brings, difficult questions and judgments will always be shirked. On the other hand, the doves remark that unless schools are in charge of any process of evaluation, then change will remain superficial as there is the need for ownership of the school improvement process. We argue that there is a need for a combination of both external inspection and self-evaluation and that promoting a culture of restless self-evaluation, a feature of improving schools, should be a key feature of any national inspection system.

MacBeath (1999) has utilized Alvik's three categories of external and internal school evaluation in a wide-ranging account of practice worldwide:

- *Parallel*. Both school and the external review body conduct their own evaluations. They may afterwards share and compare findings.
- *Sequential*. The school conducts its own evaluation and then the external body uses it as a basis for its review. It can also work in the opposite direction with the external body giving feedback to the school or offering its main findings for the school to work on.
- *Cooperative*. The two parties discuss and negotiate the process and different interests and viewpoints are taken into account simultaneously.

Learmonth (2000) uses these three categories to discuss arrangements in Scotland, USA, the Channel Islands, the Netherlands, Victoria in Australia, and a London borough where there have been attempts to get the balance right between the external and internal perspectives. However, systems of external inspection or review, the focus of this article, are established for a number of reasons; school improvement is only one of the many functions an inspection system can perform. School inspection or external review may emphasize one or more of several functions (see below) and any national system of school inspection may give greater prominence to one or another at different points in its history.

Introduction

Education systems throughout the world are being asked to improve – to raise standards and increase pupil

The Purposes of Inspection

The purposes of inspection are given below:

- Information giving – to parents and the local community; for regional and national government.
- Monitoring – assessing the performance and improvement of schools.
- Accountability – for quality, standards, and use of resources; the effective use of public funds.
- Improvement – by identifying strengths and weaknesses and the action to improve.

The functions outlined above can be used to illustrate how inspection varies across different education systems. In other words, how different systems view the main function of inspection. For example, is inspection for school improvement or to hold schools to account? Inspection systems might also differ according to the degree to which they are delivered or administered by central or local government agencies, the extent to which they are staffed by private contractors or civil servants, their levels of public exposure, and the importance they are given within the system. Differences also exist between inspection systems with regard to the degree to which they encourage schools to improve by giving advice, offering recommendations, and identifying key issues for action. In some cases, national or local inspectorates act as in-service trainers for teachers. The level of monitoring for compliance with legal regulations and the emphasis given to accountability may also vary from one inspection system to another. To what extent, for example, is inspection high stakes and part of an accountability framework?

A number of educational systems around the world have adopted external evaluation systems or approaches based on high levels of accountability. For example, in the United States, high-stakes testing and the No Child Left Behind policy tend to dominate a culture of performativity and accountability. A number of systems in developing countries have also adopted such approaches. For example, in African countries, such as Malawi, systematic inspection has been introduced as a lever for school improvement and several European countries have developed various accountability systems which are underpinned by the elements of inspection combined with use of performance data and target setting.

In an overview of the characteristics of school inspectorates in Europe, Standaert (2000) shows how in Portugal, Denmark, and France the range of school inspections varies from full inspections of both schools and individual teachers, to inspections of specific subjects only. However, all schools in England, Scotland, Flanders, the Czech Republic, and Northern Ireland receive systematic, full inspections. As noted above, some inspection systems see themselves as being complementary to the processes of internal self-evaluation, while others give greater

emphasis to the monitoring and accountability functions than they do to institutional improvement. Many claim to do both, although the extent to which school inspection leads to school improvement is hotly contested and the mechanisms used by national inspectorates to realize the goal of improvement vary from country to country.

Case Study: School Inspection in England

The history of school inspection in England can be traced back to the 1800s but the current regime emerged from the 1992 Education (Schools) Act which provided the legal framework for the launch of the Office for Standards in Education (Ofsted). This policy was implemented by the Conservative government and the underpinning rationale was that if schools had to attract students, then standards would automatically rise in the face of local competition (Learmonth, 2000). Ofsted's early years were not without controversy. Some commentators claimed the Conservative government had legitimately pitted public interests against the self-interest of the educational establishment in an attempt to demystify the closed world of schools and classrooms. Others from within the educational establishment argued that the process was demoralizing and de-professionalizing (Gray and Wilcox, 1995). A number of studies challenged the value of the Ofsted system of inspection, but Ofsted has developed its own research and publications arm to support the claim of improvement through inspection (e.g., Ofsted, 1994, 2000).

In 1996, the Conservative government outlined a new framework, but for the most part implementation would be left to their successors, New Labour, who were elected in May 1997. The new framework marked a number of changes to the system including: the stretching of the cycle from 4 to 6 years for most schools, a reduction in the notice of inspection to 6–10 weeks and an entitlement of individual feedback to teachers on their lesson performance. The new framework moved Ofsted further into the terrain of school improvement by promoting “school improvement by identifying priorities for action” (Ofsted, 1996: 2), as well as assessing the schools capacity to manage change and review its internal systems (Earley *et al.*, 1996). However, in many cases, the key elements of trust and mutual respect between the inspecting and the inspected remained limited and restricted Ofsted's contribution to school improvement, therefore the improvements made because of inspection continued to be challenged (Cullingford, 1999).

Despite calls for change from within the educational establishment, the election of the New Labour government in 1997 heralded more of the same. Intervention in inverse proportion to success continued as an overt policy with schools being identified as failing or at risk of failing, being subjected to compressed inspection cycles and

public exposure through extravagant local and occasionally national media headlines. However, many teachers and schools were becoming tired of the treadmill of inspection cycles and the process seemed to be losing some of its potency. Teachers began to realize for the most part inspections came and went, policies and plans were prepared, and reports published but unless there was a crisis not much changed.

A government response to this situation was to create a cycle for continuous improvement through a new relationship with schools. The Minister for School Standards outlined the three key aspects of the relationship as being rigorous, ongoing, self-evaluation combined with focused external inspection all linked into the improvement cycle of the school. It was argued that this simplified improvement process underpinned by a single conversation with a school improvement partner to discuss and negotiate appropriate targets, priorities for development, and strategies for support provided the high challenge and high support environment necessary to deliver the progress and performance necessary to become a world class education system (MacBeath, 2006). The new relationship with schools led to significant changes in the inspection system including an increased focus on self-evaluation.

Self-evaluation is currently at the heart of the new inspection arrangements in England because Ofsted believes that schools are best placed to recognize their own strengths and weaknesses, and have designed an online self-evaluation form (SEF) for them to do so. This is not statutory, but all schools seem to use it which is not surprising as their inspection is largely based around the SEF, which includes the school's performance data. The form asks schools to evaluate themselves under seven headings:

1. characteristics of the school;
2. views of learners, parents/carers, and other stakeholders;
3. achievement and standards;
4. personal development and well-being;
5. the quality of provision;
6. leadership and management; and
7. overall effectiveness and efficiency.

Schools can edit, save, and submit their online interactive SEFs as many times as they wish through the secure password system. They are asked to grade themselves on a four-point scale (grade 1: outstanding; grade 2: good; grade 3: satisfactory; and grade 4: inadequate) on each of these aspects:

- learners' achievement and standards in their work;
- learners' personal development and well-being;
- quality of teaching and learning;
- quality of the curriculum and other activities;
- quality of care, guidance, and support for learners;
- effectiveness and efficiency of leadership and management;

- overall effectiveness;
- capacity to make further improvement;
- improvement since the last inspection; and
- quality and standards in foundation stage (Ofsted, 2004).

A study recently conducted by the National Foundation for Educational Research highlighted that the SEF should be "something that should be carried out collaboratively, in an ongoing, continuous way, reflecting developments across the school" (McCrone *et al.*, 2006: 8). Just over half of the NFER survey respondents reported that they had attended local authority training on how to complete the SEF. In just under one in five schools an independent consultant was brought in to help. This study found that in one in ten schools the headteacher had completed the SEF alone, but that staff were consulted about the SEF in most schools and made an input in over three-quarters of schools. There were similar levels of governor involvement, with over 90% being consulted and two-thirds making an input.

The SEF has become the driver of a streamlined inspection that now lasts no more than 2 days, focuses on the core systems and senior managers, and is informed by statistical data and lesson observations, rather than being led by them (Bubb *et al.*, 2007). The size of the inspection team is smaller than in the past, and likely to be led by an experienced Her Majesty's Inspector. Notice of an inspection was also reduced to 2–5 days and schools can expect to be inspected once every 3 years. Inspection reports have also been streamlined from over 30 pages in length to five or six pages and they are presented to the governing body (in at least draft form) the same week of the inspection. Rather than preparing a response to the key issues raised, schools now incorporate their response into the school development plan.

In 2008, Ofsted published a consultation document when further changes to the inspection framework were proposed. The consultation sought views to proposed changes which included inspecting good and outstanding schools every 6 years, the possibility of unannounced inspections, national surveys which capture the views of children, school staff and parents, school leaders playing a greater part in inspection by shadowing the inspectors, giving more attention on the achievement of different groups of pupils, taking more account of the capacity of the school to improve, continuing to use contextual value added (CVA) as a measure of schools' progress, and the impact of partnerships on outcomes for pupils. (Contextual value added is the difference between the predicated and actual academic attainment made by an individual after taking into account the individual's prior attainment, characteristics and background, and the school's prior attainment.) In addition, the consultation asked should the inspectors' recommendations focus more precisely on the action the school should take to become good or better?

The ever-changing nature of Ofsted inspections has made it difficult to research its effects. However, researchers have attempted to explore the relationship between inspection and institutional development. To reflect on this fundamental issue, we draw on evidence from two national systems – England and the Netherlands.

Inspection and School Improvement

The matter of how or whether school inspections contribute to school improvement has been studied systematically in very few countries. The only country with any research tradition in this area is England and to a lesser extent the Netherlands. However, the findings of these research studies are not consistent and there is a dearth of research, as opposed to hyperbole, into the effects of school inspection and how it impacts upon school improvement.

England

Since its introduction in the early 1990s Ofsted has undergone a number of developments, reflecting the changing demands of the political and educational landscapes; in terms of experience and tone current, future school inspections bear little resemblance to those conducted in the early 1990s. However, despite structural and cultural shifts and the evolution of a more streamlined process, the budget needed to sustain inspection in England remains considerable, totaling £238 million in 2007–2008 of which £90 million has been allocated to education (schools, teacher training, and joint area reviews of children's services). Essentially, inspections of English schools are intended to identify strengths and weaknesses so that schools may improve the quality of education they provide and raise the educational standards achieved by their pupils.

During the early years of Ofsted, there was little independent research investigating inspection's contribution to school improvement (Earley, 1996). However, the agency itself produced, and continues to produce, a plethora of literature to support its claim of improvement through inspection (Matthews and Sammons, 2004; Ofsted, 2000), the extent to which (or not) Ofsted inspection contributes to school improvement remains contested at a number of levels. One key issue is that of reliability of judgment. Fitz-Gibbon (1998) argues that inspectors make inaccurate guesses about progress and the effectiveness of schools. Richards (2001: 9) observes that inspection "involves the interpretation, not just the reporting, of activities". He states that "crucially it's about making judgments as to the worthwhileness of what is observed, collected and reported" and notes that "to be valid interpretations such judgments need to be informed by an understanding of the aims and values of the activity or organization being inspected and of how these relate to the aims and values of the educational system

itself" (Richards, 2009: 9). If it is the case that judgments are not reliable, then national educational policy may have been based on inaccurate data (Fitz-Gibbon, 1998) obtained through potentially unreliable methodologies (Gray and Wilcox, 1995; Richards, 2001). In turn, this may lead to inappropriate priorities for improvement being identified and, subsequently, important opportunities for improvement overlooked.

Some researchers have considered the impact of inspection prior to the inspection period, arguing notice of inspection can trigger improvements during the run-up period. Improvement efforts include smartening up school buildings and the preparation of new interactive displays of the pupils' work mounted on walls, high quality of lesson preparation, and marking by teachers but that these improvements are likely to be short lived with "normality returning when the inspection is over" (Gray and Wilcox, 1995: 82). It would seem the reduction in notice time is likely to have further eroded any potential for improvement during the build-up period. However, the publication of a 3-year cycle allows school leaders to predict when they should be inspected and to prepare accordingly.

Other researchers have chosen to examine post-inspection impact. Lowe (1998) focuses on the implementation of recommendations 1-year after inspection. This research highlights a wide variation in how schools respond to inspection and the extent to which recommendations related to teaching and learning are tackled. Other research takes an alternative perspective on classroom change. Brimblecombe *et al.* (1996) and Chapman (2001) report approximately one-third of teachers intend to change their practice as a result of inspection and of those, most report they intend to do so as a result of the direct feedback they receive from inspectors (Chapman, 2001). However, despite an entitlement to feedback, considerable variations in quality and in quantity to feedback persist (Ferguson *et al.*, 2000). It would seem that lesson observation and quality feedback are potentially important levers for improvement but these have progressively been lessened as the length of inspections has shortened and the number of observations per inspection reduced.

One area where we may expect to see inspection making considerable contributions to school improvement is in the weakest or badly underperforming schools. Here, inspection and performance data suggest that schools identified as failing within the system are more likely to sustain the improvement they make after inspection than those that are relatively more effective but still causing concern (Matthews and Sammons (2004)). However, we also see a number of schools being removed from special measures, only to find themselves being placed back in the category after subsequent inspections (Gray, 2000) and in England there are a number of cases where despite significant intervention including compressed inspection

cycles and high levels of additional Her Majesty's Inspectorate (HMI) monitoring visits schools have not managed to establish the momentum to bring about significant improvement. Perhaps, for these schools the odds were so highly stacked against them that any intervention was doomed to failure? Alternatively, maybe the inspection process lacked the potency or sensitivity to pull the appropriate levers for change in their particular contexts? For many of these schools within the English education system the reality becomes a structural solution whereby the school is closed, and reopened as an academy, or joined with another school in a federation or collaborative restart.

The Netherlands

In the Netherlands, schools are inspected annually and an inspection lasts between half a day and 1 day. Inspectors have responsibility for about 100 primary schools or 40 secondary schools and each school will know its inspectors, thus ensuring that the inspection is not carried out by a stranger, as it is under the English system of privatized or contracted inspection teams. There is therefore a continuing relationship between the school and its inspector.

The Dutch Supervision Act of 2002 states that "through inspection the government guarantees that schools will deliver a satisfactory level of educational quality for all citizens," and that as a result of inspection "the government stimulates schools to develop their own quality assurance systems, which will lead to improvement in the quality of education." The inspection framework for assessing school performance includes aspects of the teaching-learning process, the school's results and the school organizational conditions (e.g., policies and processes). Three core aspects of school inspection which, in the Supervision Act, are assumed to promote school improvement are: quality assessment, proportionate inspection, and the publication of inspection findings.

When a school is deemed to be failing or underperforming, the Dutch inspectorate can only take action if the school does not comply with the legal regulations. Schools in the Netherlands have considerable freedom and autonomy; central government and the inspectorate have a reserved role compared to their English counterparts. Whereas action plans are obligatory in England, and follow-up school visits in cases of school underperformance will take place regularly, this is not always the case in the Netherlands. Dutch schools may also be requested to describe the implementation of their school improvement action plans which should be monitored by the school's inspector. In both countries, school inspection reports are published on the Internet.

Research shows that school inspectors vary in the ways they inspect schools and stimulate improvement. As Ehren and Visscher (2008) note:

Some inspectors assess schools more thoroughly than others, some also explain more than others about the methods they use, and some focus more on improvement suggestions, some especially put effort into agreements and improvement plans, while yet others invest more in monitoring the school following the inspection visit. These differences were interpreted as variations in inspection style which were labelled as 'directive', or 'reserved'. (p. 211)

The researchers conclude that:

School inspections probably lead to improvement processes if school inspectors give feedback about those aspects on which the school underperforms, if school inspectors assess these aspects as unsatisfactory, and if they make appointments with the school about improving on aspects of underperformance. (p. 236)

The authors stress the limitations of their research but suggest that the combination of techniques in some sort of integrated approach to school inspection is effective in promoting school improvement and their research is a useful addition to the existing knowledge base regarding the effects of school inspection. They also found that the amount of inspector feedback provided "does not contribute significantly to school improvement, but the way it is given does: in a directive sense, together with unsatisfactory assessments in the school report, and agreements about improvement of the weak aspects of functioning within a certain period" (p. 236).

Conclusion

National systems of inspection can perform a number of functions but school improvement and the raising of standards is often an important one. However, assuming that external inspection alone will automatically improve schools is naive, and few people now would adopt such a simplistic stance to school development. Contingencies are significant: a major study in England found that features of school inspections, which influenced school improvement, included such things as the inspector's attitude, the perceived inspection quality, the manner in which the inspector provided feedback, the extent to which the inspection report is perceived by the school as relevant and correct, and the congruence between the inspector's recommendations and the school culture. Interestingly, this study by Ferguson *et al.* (2000) used the term school self-inspection in preference to self-evaluation. Similarly, in the Netherlands, Standaert (2000) notes that the impact of inspections depends, among other things, on the school staff's attitude and capacity to change, whereas Matthews and Sammons (2004), writing about Ofsted inspections, point to the importance of strategic leadership and effective

action planning including the identification of necessary resources.

The inability of some schools to secure improvements despite frequent inspection and high levels of associated intervention may suggest some contexts are so challenging, the odds stacked so highly against the school that no amount of external intervention can secure improvement through inspection. However, what is particularly depressing for school improvers is that inspection does not operate in isolation. In these schools, additional resources including continuing professional development focusing on teaching and learning, target setting, and use of data are all common features of standard school improvement initiatives. Therefore, the lessons would seem to suggest that however one quantifies inspection's contribution to improvement, inspection alone is an insufficient ingredient for success and, in this case, the framework for continuous improvement drawing on inspection combined with a raft of additional resources and school improvement initiatives were also insufficiently powerful to overcome the challenging circumstances faced by some schools and communities. This can only support the arguments made in the literature (Muijs *et al.*, 2004) for linking school and community interventions.

School improvers, researchers, and policymakers are still discussing the right combination of external and internal review processes. The doves and the hawks are still to be found but now self-evaluation (or is it self-inspection?) plays a significant part in virtually all inspection arrangements. Matthews and Sammons (2004) offer a cautionary note in their recent evaluation of Ofsted's contribution to school improvement:

Ofsted will need to be vigilant, however, as the contribution of self-evaluation to the inspection process becomes more prominent, to guard against any 'halo' effect, in which self-evaluation grades have an undue influence on the judgements of inspectors. This is particularly important given the evidence that institutions tend to make more favourable judgements of their own practices. For these reasons, validated school self-evaluation should not be regarded as a substitute for inspection. (Matthews and Sammons, 2004:93)

In the international race to raise achievement and to improve standards, we need to consider carefully the role of external inspection and how it can best be related to internal processes of school self-evaluation and review. Without doubt, we need internal review and external inspection – they complement each other – but the extent or prominence of one or the other in any education system will depend on a number of factors, including the system's maturity, the individual school context, and the competence of schools to self-evaluate effectively. The era of school self-evaluation requires schools to

invest in the professional development of their staff in reflection and self-evaluation skills.

In sum, an important general conclusion can be drawn from this article about external school inspection. It does encourage school improvement but under specific conditions and in certain circumstances. We would argue that there should always be a role for the external eye or outsider perspective. The expertise and authority of school inspectors may place them in an excellent position to promote school improvement as well as perform the other functions that systems of external inspection perform.

Bibliography

- Brimblecombe, N., Ormston, M., and Shaw, M. (1996). Teachers' perceptions of inspections. In Ouston, J., Fidler, B., and Earley, P. (eds.) *Ofsted Inspections: The Early Experience*, pp 126–134. London: David Fulton Publishers.
- Bubb, S., Earley, P., Ahtaridou, E., Jones, J., and Taylor, C. (2007). The self-evaluation form: How are schools handling the SEF? *Management in Education* 21(3), 32–37.
- Chapman, C. (2001). Changing classrooms through inspection. *School Leadership and Management* 21(1), 59–73.
- Cullingford, C. (ed.) (1999). *An Inspector Calls*. London: Kogan Page.
- Earley, P. (1996). School improvement and Ofsted inspections: The research evidence. In Earley, P., Fidler, B., and Ouston, J. (eds.) (1996). *Improvement through Inspection? Complementary Approaches to School Development*, pp 11–22. London: David Fulton Publishers.
- Ehren, M. and Visscher, A. (2008). The relationship between school inspections, school characteristics and school improvement. *British Journal of Educational Studies* 56(2), 205–227.
- Ferguson, N., Earley, P., Fidler, B., and Ouston, J. (2000). *Improving Schools and Inspection: The Self Inspecting School*. London: Sage/PCP.
- Fitz-Gibbon, C. (1998). Ofsted: Time to go? *Managing Schools Today* 7(6), 22–25.
- Gray, J. (2000). *Causing Concern but Improving: A Review of Schools' Experience*. London: DfEE.
- Gray, J. and Wilcox, B. (1995). *Good School, Bad School: Evaluating Performance and Encouraging Improvement*. Buckingham: Open University Press.
- Learmonth, J. (2000). *School Inspection: What's in it for Schools?* London: Routledge/Falmer.
- Lowe, G. (1998). Inspection and change in the classroom: Rhetoric and reality? In Earley, P. (ed.) (1998). *School Improvement after Inspection? School and LEA Responses*, pp 110–125. London: PCP.
- MacBeath, J. (1999). *Schools Must Speak for Themselves: The Case for School Self-Evaluation*. London: Routledge.
- MacBeath, J. (2006). *School Inspection and Self-Evaluation: Working the New Relationship*. London: Routledge.
- Matthews, P. and Sammons, P. (2004). *Improvement through Inspection: An Evaluation of the Impact of Ofsted's Work*. London: Ofsted.
- McCrone, T., Rudd, P., Blenkinsop, S., and Wade, P. (2006). *Impact of Section 5 Inspections: Maintained Schools in England*. Slough: NFER.
- Muijs, D., Harris, A., Chapman, C., Stoll, L., and Russ, J. (2004). Improving schools in socially disadvantaged areas: A review of research evidence. *School Effectiveness and School Improvement* 15(2), 149–151.
- Office for Standards in Education (2004). *Self Evaluation and the Self Evaluation Form*. London: Ofsted.
- Ofsted (1994). *A Focus on Quality*. London: HMSO.
- Ofsted (1996). *Inspection Quality 1994/1995*. London: HMSO.
- Ofsted (2000). *Improving City Schools*. London: HMSO.

- Richards, C. (2001). *School Inspection: A Reappraisal*. Oxford: Philosophy of Education Society of Great Britain.
- Standaert, R. (2000). *Inspectorates of Education in Europe: A Critical Analysis*. Flanders: Ministry of Education.

Further Reading

- Chapman, C. (2006). *Improving Schools through External Intervention*. London: Continuum.
- Earley, P. (ed.) (1998). *School Improvement After Inspection? School and LEA Responses*. London: PCP.
- Ferguson, N., Earley, P., Fidler, B., and Ouston, J. (2000). *Improving Schools and Inspection: The Self Inspecting School*. London: Sage/PCP.
- Learmonth, J. (2000). *School Inspection: What's in it for Schools?* London: Routledge/Falmer.
- MacBeath, J. (2008). Leading learning in the self-evaluating school. *School Leadership and Management* 28(4), 385–399.
- Plowright, D. (2007). Self evaluation and Ofsted inspection: Developing an integrative model of school improvement. *Educational Management, Administration and Leadership* 35(3), 373–393.
- Richards, C. (2001). *School Inspection: A Reappraisal*. Oxford: Philosophy of Education Society of Great Britain.
- Rothstein, R., Jacobsen, R., and Wilder, T. (2008). *Grading Education: Getting Accountability Right*. New York: Economic Policy Institute and Teachers College Press.
- Standaert, R. (2000). *Inspectorates of Education in Europe: A Critical Analysis*. Flanders: Ministry of Education.

Relevant Websites

- <http://www.dcsf.gov.uk> – Department for Children, Schools and Families.
- <http://www.ofsted.gov.uk> – Ofsted.
- <http://publications.teachernet.gov.uk> – Online Publications.

The International System for Teacher Observation and Feedback: A Theoretical Framework for Developing International Instruments

L Kyriakides, University of Cyprus, Nicosia, Cyprus

B P M Creemers, University of Groningen, Groningen, The Netherlands

C Teddlie, Louisiana State University, Baton Rouge, LA, USA

D Muijs, University of Manchester, Manchester, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The globalization of the economy and the relation between economic growth and education has increased interest in education and the results of educational research, in particular comparisons between countries with respect to educational outcomes. International studies undertaken by the International Association for the Evaluation of Educational Achievement (IEA) and the Organization for Economic Cooperation and Development (OECD), such as the Third International Mathematics and Science Study (TIMSS), the Progress in International Reading Literacy Study (PIRLS), and the Program for International Student Assessment (PISA) study revealed that there are huge variations in student outcomes from country to country. This raises questions regarding the identification of factors associated with variation in student outcomes. However, secondary analyses of comparative studies reveal that most of the variations situated at the country level remain unexplained. This can be attributed to a number of methodological weaknesses of these studies (Kyriakides and Charalambous, 2005; Lassibille and Gomez, 2000) and especially to the fact that more attention should be paid to investigating the dynamic interaction between children and educational systems by understanding education as experienced by children at the point of delivery (Reynolds, 2006). This argument is also supported by the findings of studies conducted in different countries which revealed that the classroom effect is more important than the school effect in explaining variation on student achievement in both cognitive and affective outcomes. Effectiveness studies have also revealed that quality of teaching is the most important factor at the classroom level (Brophy and Good, 1986; Kyriakides, 2005a).

Unless data on the quality of teaching are collected, researchers and policymakers cannot easily interpret the results of these comparative studies and may attribute variation in student outcomes to variables that show differences between countries, but are not necessarily related to student outcomes. For example, when the results of PISA 2000 study were announced, policymakers in different countries attempted to emulate parts of the educational policy from Finland in their own countries.

However, results achieved in one country cannot be replicated in another country by a simplistic transplant of educational factors to another context, especially if these factors were not found to explain variation in achievement from country to country. Therefore, effort should be put into developing valid instruments which are able to measure quality of teaching.

Although national-observation instruments have been developed and validated in some countries, there is no internationally valid teacher-observation instrument available. Moreover, in an international study on educational effectiveness (Reynolds *et al.*, 2002), it was discovered that one of the most commonly used instruments in USA (i.e., the Virgilio Teacher Behavior Inventory (Teddlie *et al.*, 1990)) did not produce valid data on quality of teaching in several of the participating countries. This implies that in the construction of an international teacher-observation instrument, different views on quality of teaching should be taken into account for identifying important aspects of teacher effectiveness shared in different countries, as well as aspects that are more important for specific countries.

In this context, the main aim of the International System for Teacher Observation and Feedback (ISTOF) was to develop a valid international instrument for assessing teacher effectiveness that can be used for conducting comparative studies which may help us to understand the reasons why some countries are more effective in education than others. Since national studies reveal that teacher factors are important predictors of educational outcomes, a comparative study measuring teacher effectiveness may also reveal which aspects of the quality of teaching can explain variation on student outcomes across countries and which aspects are context specific (i.e., they demonstrate relationships with student outcomes in specific countries only). Thus, similarities in characteristics of effective teaching will be identified. At the same time, differences between countries in what constitutes effective teaching will be identified and they will help us understand the conditions where specific reform policies could contribute to learning.

Therefore, the main aim of this article is to inform readers about the background and the rationale for the ISTOF project and to present its major outcomes. We also

draw implications of the findings of this project for developing international instruments that can measure effectiveness factors operating at different levels, such as the school and the system. In this way, valid instruments may be developed for use in comparative studies seeking to identify factors that explain variation in student outcomes across countries. In the next section, we describe the rationale upon which the design of ISTOF is based by referring to four techniques used in developing ISTOF.

The Rationale for Designing ISTOF

Technique 1: Merging Inductive and Deductive Procedures

We decided to make use of both inductive and deductive procedures for the development of ISTOF. The inductive procedure involved the collection of input from different groups of participants in as many countries as possible with regard to the components of effective teaching, their indicators, and the items that can be used to collect data for each indicator. On the other hand, the use of deductive procedure implies that the knowledge base on teacher effectiveness, which emerged through syntheses of research findings, should guide the development of the conceptual map, the list of indicators, and the items of the instrument. By using both approaches, we tried to take into consideration not only research findings concerned with teacher effectiveness, but also the viewpoints of different groups of experts in different countries (i.e., researchers, professionals, and policymakers) on what makes an effective teacher and how this can be measured. The reasons for using this approach in developing ISTOF are discussed below.

First, in many countries, studies looking at teacher effectiveness have been conducted only recently, if at all. By using only the deductive approach, it is possible to overemphasize views stemming from specific countries where there is a long tradition in teacher-effectiveness research (TER) such as the USA where TER originated in the 1960s and 1970s (Gage, 1977; Brophy and Good, 1986). By using deductive methods based only on the existing TER literature, we may not take into consideration contextual differences that can affect the way research is framed.

Second, by using the inductive approach, you can take into account the experience and views of different stakeholders in different countries who are currently involved in the training, evaluation, and counseling of teachers. In this way, we can combine the knowledge and experience of professionals and policymakers within and between countries and identify whether consensus can be achieved. Therefore, this approach helps us establish an instrument capable of generating meaningful and useful data for different countries and different groups of stakeholders.

Third, the use of both the inductive and deductive approaches facilitates the search for elements of effective teaching that are considered important by stakeholders in different countries and are also in line with the knowledge base of educational effectiveness research. This enables us to generate a part of the instrument (i.e., the common part) that can be used across different countries for collecting valid and meaningful data with respect to teacher effectiveness. Another goal of ISTOF was to identify elements that are more country specific, meaning that these characteristics of effective teachers are important in only some countries. The country-specific part of the instrument may generate data that can explain the conditions under which the common aspects of effective teaching operate in improving learning.

Finally, the use of inductive and deductive approaches for developing ISTOF gives us the opportunity to examine the internal validity of the procedure used to define components of effective teaching, and generate their indicators and items per indicator. Moreover, it demonstrates the value of using both quantitative and qualitative approaches in collecting and analyzing data. If similar results emerge from both approaches, one can assume that it is likely that both have generated valid data.

Technique 2: Using the Mixed-Methods Approach

We also decided to conduct a mixed-methods study to develop ISTOF. The reasons for using such an approach are related to our intention to use both inductive (qualitative) and deductive (quantitative) techniques in developing the instrument. Specifically, in order to find out how each expert participating in the country panels understands effective teaching, appropriate qualitative methods such as individual and group interviews were used and the data were analyzed using qualitative strategies. In this way, different views of effective teaching emerged and were taken into account in generating the instrument. But since our intention was to also develop a common part of the instrument, it was important to use appropriate quantitative methods (i.e., administering questionnaires and analyzing numeric data). The quantitative approach emphasizes the identification of those aspects of effective teaching that are considered important across the entire group of country-based expert panels.

Technique 3: Using a Modified Delphi Technique

Activities for developing ISTOF centered on an iterative, multiple-step, Internet-based, modified Delphi technique (Teddle *et al.*, 2008). Modified means that the ISTOF queries asked experts their opinions about what constitutes effective teaching, whereas the original Delphi technique

asked experts to forecast events in the future (Gordon and Helmer, 1964). There are three main reasons for using this technique. First, it enables us to determine whether consensus can be reached among members of an expert panel. In the case of ISTOF, we used the Delphi technique to generate the components, indicators, and items related to teacher effectiveness. The items that constitute the common part of ISTOF are those for which consensus about their perceived importance could be reached within and between country-based expert panels. Second, the Delphi technique has been applied in many settings, such as educational policymaking and even in previous studies on what constitutes effective teaching and it was found to be useful for reaching consensus among groups of participants. Finally, using this technique, feedback is given to members of the country panels consisting of the results of analyzing data that emerged from each stage of developing ISTOF, combined with lessons drawn from analyzing the literature. Then, by applying the Delphi technique, consensus could be reached about the final product of the activity and a new round based on the results of the previous one could be initiated.

Technique 4: Using Generalizability Theory to Analyze Data

Since the findings of the ISTOF project are based on the responses of the panel of experts in each country, it is important to evaluate the dependability (reliability) of the behavior of each member and/or interest group in the panel of each country. At the same time, the issue of reaching consensus among the panel of experts is very important for generating the common part of the ISTOF. Therefore, the analysis of quantitative data must identify the extent to which such consensus was established in each step of the process. Generalizability theory provides answers to these two issues (dependability and degree of consensus) and it is for this reason that it was used for developing ISTOF.

The conceptual framework underlying generalizability theory involves an investigator asking about the precision or reliability of a measure because she/he wishes to generalize from the observation in hand to some class of observations to which it belongs (Shavelson *et al.*, 1989). In the case of ISTOF project, it was considered important to determine the extent to which the responses of each member of the panel of experts to the questionnaire items depend on his/her professional status (e.g., researcher, teacher, teacher-educator, policymaker, or evaluator) and/or on his/her origin from a particular country. Specifically, this analysis helped us to identify the extent to which experts from different countries agreed among themselves about the appropriateness of: (1) the components of effective teacher, (2) their indicators, and (3) the

items designed to measure the set of indicators of teacher effectiveness. Therefore, the use of generalizability theory gave answers to questions concerned with the extent to which the conceptual map of the ISTOF is in line with the opinions of experts in different countries who belong to different groups of stakeholders.

Taking into consideration the rationale given above, the following procedure was used to develop ISTOF by first generating components of effective teaching, and then producing indicators for each component and finally creating items associated with each indicator.

Description of Six Steps/Queries

At the beginning, in each participating country, panels of experts were established. There was a general procedure and requirements for selecting group members and the team leader that each country followed. Details about this procedure are contained in Teddlie *et al.* (2006). Countries participated in the project are as follows: Argentina, Belarus, Belgium (Flanders), Brazil, China, Cyprus, Denmark, Finland, India, Ireland, Japan, Malaysia, Netherlands, South Africa, Turkey, United Kingdom, and the United States of America. This sample of countries is very diverse but due to contextual differences between countries, there were no strict guidelines with respect to the composition of expert panels, but country coordinators were advised to include representatives of the three main categories: researchers, practitioners (i.e., teacher trainers, teacher evaluators, and advisors), and policymakers. Then six Internet-based queries were formulated to guide the development of ISTOF.

Query 1 asked the country-based expert panels to generate broad components of effective teaching in their country and give clear definitions of them. Seventeen country teams sent in lists of 103 components with definitions in response to this query. Using the qualitative data analysis program ATLAS.ti, two different teams of analysts conducted content analyses of these responses by following the constant comparative method (Denzin and Lincoln, 1998). The two lists generated by the teams were reconciled and resulted in a final list of 11 components with definitions (see **Appendix 1** for the list of components and their definitions).

Query 2 asked each country team and its members to assess each of the components by (1) rating how important each component was on a five-point Likert scale and (2) rank ordering the 11 components. Responses to query 2 were received from all participating countries and 257 individual participants. The main aim of the analysis was to identify those components of effective teaching which were regarded as the most important by all participating countries. Using both parametric and nonparametric tests for each part of the questionnaire respectively, it was found out that all components were

rated highly by each country team and by each category of experts. Moreover, it was not possible to classify these components into different groups according to their perceived importance (see Kyriakides, 2005b). These components were then subjected to a generalizability analysis (Shavelson, *et al.* 1989). Specifically, the country and the professional status of each participant were the two facets, which were taken into account. Using the GENOVA software (Crick and Brennan, 1983), a random effects component X (expert:countries), analysis of variance was initially performed. Thus, this analysis makes explicit the country facet acknowledging that experts share some characteristics and help us identify the extent to which the responses of experts depend on their origin from specific countries. Although it was found out that failing to include the country facet led to an overly optimistic estimate of the precision of the measurement procedure during query 2, the component of effective teaching by country and the component by expert (nested in countries) effects were very small. These findings justify our attempt to analyze the data of query 2 by looking at differences across countries. It also reveals only a small effect of the country of origin of each expert upon his/her responses during the query 2. By introducing the facet of professional status, we also attempted to identify the extent to which the response of each member of the panel of experts depends on his/her professional status (i.e., researcher, teacher, teacher-educator, policymaker, or evaluator). The results of this study revealed that failing to include the professional-status facet in the analysis has minimal impact on the estimated error variance.

Comparison of the components with the research literature on teacher effectiveness resulted in minor amendments with respect to the definition of some components and the grouping of them (see Teddlie *et al.*, 2006). Decisions also occurred regarding how each component could be measured especially since four of the components could not be measured through observation. Other ways of measuring each component, such as student ratings, documentary analysis, and self-reports were recommended.

Query 3 asked the country team members to generate up to five indicators for each of the 11 refined components. Altogether, the team members generated about 65 indicators per component, which resulted in almost 750 indicators for the entire set of components. Following the same procedure as in query 1, content analysis of these indicators resulted in a reduced set of 43 indicators (the number of indicators per component ranged from two to six).

Query 4 asked the country team members to assess the importance of each of the 43 indicators generated through query 3 on a five-point Likert scale. Altogether there were 213 individual responses from the participating countries. The procedure for analyzing the data was similar to the one that was used with respect to query 2. It was

determined that 21 indicators, which were seen as highly appropriate from each country, were sufficient for collecting data with respect to the seven observable components of teaching (i.e., clarity of instruction; instructional skills; promoting active learning and developing metacognitive skills; assessment; differentiation and inclusion; classroom climate; and classroom management). Moreover, eight indicators were seen as appropriate by each country panel for collecting data with respect to the nonobservable components (i.e., planning of single lessons; long-term planning; teacher knowledge; and teacher professionalism and reflectivity). However, these components were not utilized in conducting queries 5 and 6, which deal with the development of the observation instrument.

Queries 5 and 6 were concerned with generating items per indicator for the observation instrument. The country coordinators and experts from different countries were split into four groups, which then drafted an average of about three items for each of the seven observable indicators. These items were exchanged for comments and critiques. A final set of 45 items emerged through consensus reached within and between the four groups. Input from experts in the construction of the teacher-observation instruments was also provided and contributed to the development of the final version of the observation instrument.

Main Results

ISTOF resulted in different products which can be used independently from each other. First, as a result of queries 1 and 2, a conceptual map of components of effective teaching that are considered highly important from experts in 17 countries was produced (see **Table 1**). This conceptual map refers to 11 components of effective teaching that are grouped into five overarching factors. It is important to note that the 11 components do not refer exclusively to teacher behavior in the classroom. For example, the fifth overarching factor refers to the professional development of teachers, which is seen as a characteristic of effective teachers by stakeholders in different countries, but only some aspects of teacher knowledge and teacher professionalism and reflectivity could be measured through observation of teacher behavior in classroom. This indicates that beyond demonstrating specific behaviors in the classroom, effective teachers are also expected to develop their knowledge further with respect to teaching (i.e., subject, pedagogy, and pedagogical content knowledge) and to develop themselves further as professionals through reflection and other practices (see components 10 and 11, respectively, in **Table 1**). This view of the country panels seems to be in line with the continuous learning model (Cheng and Tsui, 1999) which assumes that teachers are effective if they can adapt to external and internal changes, cope with different challenges, meet

Table 1 Conceptual map of ISTOF components, with overarching factors of effective teaching

Components from Deductive Committee	Ways of measuring the component				Overarching factors
	Observation	Student ratings	Documentary analysis	Self-reports	
1 Clarity of instruction	√	√			Quality of teaching
2 Instructional skills	√	√			
3 Promoting active learning and developing metacognitive skills	√	√			
4 Assessment and evaluation	√	√	√	√	Adaptive teaching
5 Differentiation and inclusion	√	√			Classroom environment
6 Classroom climate	√	√			
7 Classroom management	√	√			Planning
8 Planning of single lessons		√	√	√	
9 Long-term planning			√	√	
10 Teacher knowledge (subject, pedagogy, and pedagogical content knowledge)					Teacher as a professional
11 Teacher professionalism and reflectivity					

diverse expectations, and develop themselves through continuous learning.

The other two components which are not directly related to teacher behavior in the classroom refer to short- and long-term planning activities (see components 8 and 9 in **Table 1**). This implies that stakeholders in different countries expect their teachers to be competent in planning their activities and in working according to those plans. These two components may also be related to the components concerned with teacher professional development especially since a cycle of planning, acting, and evaluating through reflection is assumed to contribute to the learning of teachers and ultimately to improve the quality of teaching.

The first seven components are more directly associated with the teaching process and they are in line with the working-process model, which assumes that the effective teaching and functional learning processes help teachers to achieve the predetermined objectives measured through students' cognitive and affective learning outcomes (Brophy and Good, 1986). Consequently, teachers are judged as effective when they can ensure the quality of their teaching. Based on the results from the queries, it appears that stakeholders in different countries agreed regarding these first seven components of effective teacher and their definitions. These components were also found in the reviews of the TER literature that accompanied this research project.

Specifically, looking at the three components of the first overarching factor, we observe that they are not only in line with the tenets of structured or direct teaching (Joyce *et al.*, 2000), but also with approaches associated with constructivism (Schoenfeld, 1998). This implies that

stakeholders believe that an integrated approach to teaching should be adapted rather than focusing exclusively on a specific approach. This view of stakeholders is supported by recent findings of effectiveness studies which have shown that by adapting an integrated approach to teaching, teachers can become more effective (Kyriakides and Creemers, 2008). Moreover, stakeholders across a large number of countries agreed that formative assessment is an important component of effective teaching (see component 4 in **Table 1**). Again, empirical evidence from a number of studies provided strong support for this component (Kyriakides, 2005a).

Furthermore, stakeholders in different countries underlined the importance of differentiation in teaching, thereby agreeing with published research and theoretical models that treat differentiation and inclusion as important elements of effective teaching (see component 5 in **Table 1**). Finally, the classroom-environment overarching factor and its two components indicate that stakeholders believe that effective teachers should organize and manage the classroom environment as an efficient learning environment and thereby maximize engagement rates (Creemers and Reezigt, 1996; Doyle, 1986). With respect to this overarching factor, the two components (i.e., classroom climate and classroom management) are seen as important for measuring teacher effectiveness by both the stakeholders and the research literature, indicating that the classroom environment should not only be businesslike but also needs to be supportive for the students (Walberg, 1986). Therefore, effective teachers are expected not only to manage their classroom in order to keep students on task but also to create a positive, learning-centered environment with an atmosphere of mutual respect between

students and between students and the teacher (Creemers and Kyriakides, 2008).

Results from the first two queries point to the fact that consensus about what constitutes effective teaching among stakeholders from different professional groups was easily achieved in a large number of countries, and all of the components which were generated were seen as very important by all the country panels. It was, therefore, not possible to treat any of them as country specific. Since only a handful of effectiveness studies had been conducted in most participating countries, most of the members of the country-based expert panels were not aware of the knowledge base of effectiveness research. Despite this, these country-based panels conceived of teacher effectiveness in the same way as more experienced members and researchers. Therefore, reaching consensus easily among the participants in the ISTOF project can be interpreted as a strong argument for the existence of a common understanding of what constitutes effective teachers.

These results are also in line with the findings of a recent collaborative work by inspectorates of education in Europe, which reveal that the quality of teaching in four countries could be compared in a reliable and valid way with regard to constructs such as a safe and stimulating environment, clear instruction, adaptation of teaching, teaching-learning strategies, and classroom management (Van de Grift, 2004). The constructs that emerged from that European study are similar to those identified through ISTOF where a larger number of countries from different parts of the world participated. These findings imply that there are factors which can describe the process of teaching and learning that are common in different educational settings irrespective of their specific social and cultural context (see also Creemers and Kyriakides, 2008).

A list of 43 indicators for these 11 components of effective teaching emerged from queries 3 and 4. Again, consensus within and between the country panels was reached, but only 29 out of 43 were seen as very important from all the participating countries and the various groups of stakeholders. We, therefore, decided to retain these 29 indicators in the common part of ISTOF and treat the others as country specific. This overall pattern of results indicates that although there is a common understanding among stakeholders in the participating countries on what constitutes effective teaching, some indicators are seen as appropriate in specific countries only due to social and cultural reasons. For example, only the Chinese team treated the time that each teacher spends on his/her career (e.g., hours spent at students' home-visits or for assisting students in his/her spare time) as an indicator of his/her professionalism.

On the other hand, indicators which were rated as very important by each country panel can be used not only for establishing the common part of ISTOF but also for

helping policymakers collect evaluation data. For example, policymakers and practitioners in different countries may make use of them to establish their own instruments based on a more specified frame that is applicable across countries. Moreover, they can create a basis for comparative studies concerned with the effectiveness of teaching practices in their countries.

Three indicators associated with two different observable components (i.e., Components 1 and 3), which were used for the construction of the common part of ISTOF, are presented below as an illustration of the common indicators which were developed:

Component 1: Assessment and evaluation

- *Indicator 1.1.* The teacher gives explicit, detailed, and constructive feedback.
- *Indicator 1.2.* Assessment is aligned with goals and objectives.

Component 3: Clarity of instruction

- *Indicator 3.1.* Lessons are well structured.

The final product of ISTOF was a high-inference observation instrument consisting of 45 items associated with the 21 indicators of the observable components. Each indicator is represented by two or three items which are rated on a five-point Likert scale. Together with the instrument, guidelines for training observers to use the instrument were produced. We recommend in the guidelines that observers should be familiar with the 45 items on the observation instrument before conducting any observations. The guidelines also stressed that during the training, each indicator and each item of the observation instrument should be discussed by the trainees before using the instrument. In addition, we have advocated that trainees should conduct observations using the instrument and the training should be continued until interrater reliability (computed using Cohen's kappa) at 0.70 level is achieved. Two items used for measuring two indicators are given below to illustrate the content and form of the observation instrument:

- *Indicator 3.3.* Lessons are well structured.
- *Item 13.* The teacher presents the lesson with a logical flow that moves from simple to more complex concepts.
- *Indicator 4.2.* The teacher possesses good questioning skills.
- *Item 18.* The length of the pause following questions varies according to the difficulty level of questions (e.g., a question calling for application of abstract principles requires a longer pause than a factual question).

ISTOF was conceived as a protocol that would employ mainly observational items. While this is the case, there are components of effective teaching that were generated through queries 1 up to 4 that are unobservable. The country teams indicated that planning, teacher knowledge, and teacher professionalism and reflectivity should be included on the ISTOF protocol even if it meant

developing alternative data sources. For this purpose, alternative methods of measuring these components have been developed and a student questionnaire is currently being validated in different countries.

Conclusions

The final section of this article draws implications from the ISTOF project for the process of developing valid international instruments that can be used in comparative studies searching for multilevel factors that may explain variation on student outcomes in different countries. We believe that the use of both inductive and deductive approaches in developing ISTOF made important contributions in producing both the common and the country-specific parts of the instrument. Specifically, the inductive process, facilitated by the use of modified Delphi technique, made it possible to identify and classify the views of different stakeholders and relate them to each other. That process served the purpose of identifying those views that are shared by the great majority of stakeholders in different countries, as opposed to those that are supported either by individuals or a few stakeholders in only a small number of countries. Then by comparing the results of the inductive approach with the results of the deductive approach, which is concerned with establishing consensus from all the available empirical evidence including that from the literature, the common part of the instrument was produced. This common part of the instrument, therefore, consists of constructs/indicators/items that are supported both by the general research evidence and the experts/stakeholders from ISTOF. At the same time, this comparison revealed the importance of generating country-specific constructs/indicators/items. However, it should be acknowledged that using both inductive and deductive approaches in developing the framework of an international instrument can be problematic both in terms of collecting data from a large number of stakeholders in different countries and in terms of establishing consensus among them. Nevertheless, our experience with ISTOF indicates that the Internet-based approach that characterizes much of the project contributed greatly to its success.

The use of both deductive and inductive approaches in developing the ISTOF resulted in an effective teaching framework that is shared by a large number of countries in different parts of the world and with different experience in educational effectiveness research. Countries where there is not much of a research tradition in educational effectiveness can use this framework as a point for departure for launching their own research. In fact this research is not only relevant for their countries, but can also contribute to the knowledge base of the field. Therefore, using inductive and deductive approaches to developing

international instruments concerned with effectiveness factors operating at different levels (e.g., school climate, school self-evaluation, and parental involvement) may result not only in establishing valid instruments that are meaningful for different countries, but may also contribute to the development of a program for future research in parts of the world where it did not exist before.

Finally, it is important to acknowledge the dilemmas that may arise when using this approach. Given that there is huge variation in experiences and knowledge in the participating countries, the danger for more experienced researchers to impose their agenda on other participants is an issue that has to be considered. Ultimately, bringing together stakeholders and researchers from different countries should result in establishing a learning environment where participants can develop knowledge and experience further by cooperating with colleagues from different countries. At the same time by advocating this international cooperation, we do not imply that quality standards for educational research should be ignored or disregarded. The lack of experience in some countries may be compensated by undertaking some tasks centrally (e.g., using advanced qualitative and quantitative techniques to analyze data). In such cases, the presentation of the outcomes of these tasks should be done in a way that is meaningful for all members and that allows participation of all country panels in the discussion about the meaning of these results, or even the need for further analysis of data and further development of the project. Thus, this article demonstrates that although projects aiming to develop international valid instruments are complicated, such projects can also contribute significantly in the further development of policy, practice, and theory of educational effectiveness.

Appendix 1: Definition of ISTOF Components of Effective Teaching

Clarity of Instruction

The teacher is able to communicate clearly and directly with students and is able to adapt her/his communication to student's levels of understanding and their individual differences. The teacher clearly explains the purposes of class activities. The teacher has good listening skills and is a good facilitator who is able to draw out student potential.

Instructional Skills

The teacher uses a variety of methods and strategies effectively. S/he has also basic instructional skills including orientation, structuring the lesson, presentation, questioning, listening, balancing teacher input and student activity, and making use of appropriate teaching aids.

Promoting Active Learning and Developing Metacognitive Skills

The teacher is able to connect the learning contents to the students' prior knowledge about the world. The teacher makes use of students' own views on their learning as well as their self-conceptions as learners in different learning domains and promotes active involvement of students in her teaching and instructional practices. S/he also fosters metacognition, critical thinking, creativity, self-regulation, and reflection. S/he also helps students develop their own strategies for solving different types of problems.

Assessment and Evaluation

The teacher uses various techniques to assess and evaluate student learning during and after teaching sequences, in order to check whether learning goals have been reached, and to adapt teaching to student needs and progress. Based on information gathered from student assessment, the teacher also evaluates herself/himself and makes decisions on how to improve her/his teaching skills and practice. Assessment procedures include immediate and formative feedback, aimed at enhancing student self-evaluation and metacognitive skills.

Differentiation and Inclusion

The teacher is aware of, and adapts instruction to, the individual needs of her/his students. S/he focuses on their individual development, while making sure all are included and empowered in the learning community.

Classroom Climate

The teacher creates a positive, learning-centered environment with an atmosphere of mutual respect between students and between students and the teacher. The environment is both safe and caring. High expectations exist for all students and are clearly communicated to students by the teacher. The teacher expresses genuine affection for children, and all students feel valued and experience enhanced self-concepts.

Classroom Management

The teacher manages the classroom effectively. S/he has good time management, establishes a purposeful lesson flow, and allows all students learning opportunities through effective behavior management, which includes clear and accountable procedures and rules while allowing students to exercise autonomy.

Planning of Single Lessons

The teacher effectively plans instruction and the creation of a challenging learning environment to foster student learning. S/he bases her/his teaching on clear and explicit goals that are communicated to students. The teacher is well prepared with materials and resources ready at the outset.

Long-Term Planning

Long-term planning is the organization of the curriculum and especially of the content to be covered over time within grade level and over grades. It refers to the time needed to cover the curriculum and the emphasis given on topics and objectives. It asks from the teacher not only to look forward but also to see what she has done and to change her teaching plan if necessary.

Teacher Knowledge (Subject, Pedagogy, and Pedagogical Content Knowledge)

The teacher has a good knowledge of relevant subject content, pedagogy, theories of learning, motivation and human behavior, and is able to use this knowledge to shape her/his practice. S/he applies these knowledge bases in the classroom within an organized curriculum that facilitates student learning.

Teacher Professionalism and Reflectivity

The teacher reflects critically on her/his own and others' practice, engages in dialog with peers, is committed to lifelong learning, and regularly updates her/his knowledge and skills. S/he is able to work in a fast-changing environment and adapt to policy initiatives and is characterized by proactivity toward innovation.

See also: International Evaluations; School Effectiveness in Developed Societies; Validity of Educational Indicators; Whole School Designs for Enhancing Student Achievement.

Bibliography

- Brophy, J. and Good, T. L. (1986). Teacher behavior and student achievement. In Wittrock, M. C. (ed.) *Handbook of Research on Teaching*, 3rd edn., pp 328–375. New York: MacMillan.
- Cheng, Y. C. and Tsui, K. T. (1999). Multimodels of teacher effectiveness: Implications for research. *Journal of Educational Research* 92(3), 141–150.
- Creemers, B. P. M. and Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice and Theory in Contemporary Schools*. London: Routledge.
- Creemers, B. P. M. and Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement* 7, 197–228.

- Crick, J. E. and Brennan, R. L. (1983). *Manual for GENOVA: A Generalized Analysis of Variance System*. Iowa City, IA: ACT.
- Denzin, N. K. and Lincoln, Y. S. (eds.) (1998). *Collecting and Interpreting Qualitative Materials*. Thousand Oaks, CA: Sage.
- Doyle, W. (1986). Classroom organization and management. In Wittrock, M. C. (ed.) *Handbook of Research on Teaching*, 3rd edn., pp 392–431. New York: Macmillan.
- Gage, N. L. (1977). *The Scientific Basis of the Art of Teaching*. New York: Teachers College Press.
- Gordon, T. J. and Helmer, O. (1964). *Report on a Long Range Forecasting Study*. Rand Paper P-2982. Santa Monica: Rand Corporation.
- Joyce, B., Weil, M., and Calhoun, E. (2000). *Models of Teaching*. Boston: Allyn & Bacon.
- Kyriakides, L. (2005a). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement* **16**(2), 103–152.
- Kyriakides, L. (2005b). Report of preliminary results emerging from ISTOF Query #2. *Paper presented at the Annual Meeting of the International Congress for School Effectiveness and Improvement*, Barcelona, Spain.
- Kyriakides, L. and Charalambous, C. (2005). Using educational effectiveness research to design international comparative studies: Turning limitations into new perspectives. *Research Papers in Education* **20**(4), 391–412.
- Kyriakides, L. and Creemers, B. P. M. (2008). Using a multidimensional approach to measure the impact of classroom level factors upon student achievement: A study testing the validity of the dynamic model. *School Effectiveness and School Improvement* **19**(2), 183–205.
- Lassibille, G. and Gomez, L. N. (2000). Organization and efficiency of educational systems: Some empirical findings. *Comparative Education* **36**(1), 7–19.
- Reynolds, D. (2006). World class schools: Some methodological and substantive findings and implications of the international school effectiveness research project (ISERP). *Educational Research and Evaluation* **12**(6), 535–560.
- Reynolds, D., Creemers, B., Stringfield, S., Teddlie, C., and Schaffer, G. (eds.) (2002). *World Class Schools*. London: RoutledgeFalmer.
- Schoenfeld, A. H. (1998). Toward a theory of teaching in context. *Issues in Education* **4**(1), 1–94.
- Shavelson, R. J., Webb, N. M., and Rowley, G. L. (1989). Generalizability theory. *American Psychologist* **44**(6), 922–932.
- Teddlie, C., Creemers, B. P. M., Kyriakides, L., Muijs, D., and Fen, Y. (2006). The international system for teacher observation and feedback: Evolution of an international study of teacher effectiveness constructs. *Educational Research and Evaluation* **12**(6), 561–582.
- Teddlie, C., Tashakkori, A., and Johnson, B. (2008). Emergent techniques in the gathering and analysis of mixed methods data. In Hesse-Biber, S. and Leavy, P. (eds.) *Handbook of Emergent Methods in Social Research*, pp 389–413. New York: Guilford.
- Teddlie, C., Virgilio, I., and Oescher, J. (1990). Development and validation of the Virgilio Teacher Behavior Inventory. *Educational and Psychological Measurement* **50**(2), 421–430.
- Van de Grift, W. (2004). *Quality of Teaching in England, Belgium Flanders, Lower Saxony (Germany) and the Netherlands*. Utrecht, The Netherlands: The Netherlands Inspectorate of Education.
- Walberg, H. J. (1986). Syntheses of research on teaching. In Wittrock, M. C. (ed.) *Handbook of Research on Teaching*, 3rd edn., pp 214–229. New York: Macmillan.

Further Reading

- Campbell, R. J., Kyriakides, L., Muijs, R. D., and Robinson, W. (2004). *Assessing Teacher Effectiveness: A Differentiated Model*. London: Routledge/Falmer.
- Creemers, B. P. M. (1994). *The Effective Classroom*. London: Cassell.
- Fraser, B. J., Walberg, H. J., Welch, W. W., and Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research* **11**, 145–252.
- Kyriakides, L., Creemers, B. P. M., and Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education* **25**(1), 12–23.
- Muijs, D. and Reynolds, D. (2001). *Effective Teaching: Evidence and Practice*. London: Sage.
- Scheerens, J. and Bosker, R. J. (1997). *The Foundations of Educational Effectiveness*. Oxford: Pergamon.
- Teddlie, C. and Reynolds, D. (2000). *The International Handbook of School Effectiveness Research*. London: Falmer.

EDUCATIONAL EVALUATION – EVALUATION METHODS

Contents

Evaluation Methodology

Evaluation Reporting and Communicating

Longitudinal Evaluation Designs

Meta-evaluation: Purpose, Prescription, and Practice

Evaluation Methodology

M Chatterji, Columbia University, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Causal inferences – A conclusion about whether a program directly leads to some outcomes in a target population.

Cost-efficiency studies – Studies aimed at evaluating the long-term benefits, effects, feasibility or perceived utility of a program or policy intervention in monetary terms. Examples include cost–benefit, cost–effectiveness, cost utility and cost–feasibility studies.

Ecological validity – The extent to which evaluation results are pertinent in real settings where the program is likely to operate.

Educational evaluation – Systematic inquiry processes used for monitoring, improving, and/or appraising the quality of educational entities in stakeholder contexts.

Evaluability assessments – A pre-evaluation study, conducted to examine the extent to which a program is ready for a formal evaluation.

Evaluand – The object of inquiry in an evaluation.

Evaluation research – The systematic use of a variety of social research methods and tools of inquiry for answering evaluation questions, guided by stakeholder needs for information about programs or policies.

Evaluation researchers – See the definition of the glossary term ‘evaluators’.

Evaluation utility, evaluation use – The extent to which the results of an evaluation are meaningful and useful to stakeholders and help improve prevailing educational or social conditions.

Evaluators – Professionals, usually with graduate-level training in social research methods, who

provide evaluation research services to clients and stakeholders.

Evidence hierarchies – Related to policies in evidence-based practices, this term refers to a research design hierarchy linked to the quality of evidence that particular research designs yield. In a widely-used evidence hierarchy, randomized experiments are considered best for generating impact evidence on programs.

Evidence-based education (EBE) – A policy that requires educators to use practices, programs, and interventions that are supported with empirical evidence obtained via research and evaluation studies. The policy is also called evidence-based practice (EBP).

Formative evaluations – Evaluations in which results are used for refining and improving an entity.

Generalizability – The extent to which evaluation results can be extrapolated from smaller samples of observations to the larger population.

High-stakes testing – A policy or practice where results from test administrations are used by authorities to impose sanctions or grant rewards to students, teachers, or leaders in an education system.

Impact evaluations – Studies aimed at monitoring effects and evaluating the expected outcomes of a program or policy intervention.

Large-scale assessments – School-wide or system-wide testing of student achievement outcomes in order to document trends or assess present status, usually driven by public accountability requirements.

Logic models – Causal path diagrams revealing assumptions about how an entity is expected to work

in actual contexts (see the definition of the glossary term ‘program theory-driven evaluation’).

Needs assessments – Evaluation studies intended to determine the scope and dimensions of a social or educational problem, where results are typically used to plan or inform the design of a new program or policy intervention.

Performance assessments – Systematic procedures designed to observe and rate the quality of performances (e.g., classroom behaviors of students and teachers).

Process evaluations – Evaluation studies aimed at monitoring the implementation processes of a program or policy intervention.

Product assessments – Systematic procedures designed to rate or classify the quality of an educational product (e.g., teacher-developed instructional materials).

Program theory-driven evaluation – An approach to evaluation design that uses causal path models to depict and test the underlying assumptions of how a program is expected to work, identifying key variables and expected relationships (see also the definition of the glossary term ‘logic models’).

Replications – Repeated studies conducted with different samples to verify results.

Social research methods – Formal tools of inquiry used in the social sciences, including those employed in quantitative, qualitative, and mixed-method research designs.

Summative evaluations – Evaluations in which findings are used to make a final judgment of the merit or worth of an entity.

Systemic evaluation – A comprehensive evaluation design that examines a program’s context, inputs, processes, or outcomes within the organizational or community system.

Unit of analysis – The units that are formally selected, observed, described, analyzed, or appraised in an evaluation.

Educational evaluation, also called evaluation research in education, involves inquiry processes that rely on a range of social research methods. Evaluative inquiry is aimed at answering questions relevant to making decisions in educational practice and policy contexts. The primary goal of educational evaluation is to monitor and gauge the effectiveness of various educational programs and to improve their functioning. Educational evaluations are also intended to appraise the extent to which students learn and attain educational goals, both in and out of formal school settings.

While there are long-standing debates about what types of methods are most useful for particular evaluation

purposes and problems, there is widespread agreement that some form of systematic and formal methodology is necessary to support most kinds of evaluation research (Cronbach *et al.*, 1980; Guba, 1969; Stake, 1984; Campbell and Stanley, 1963; Eisner, 1994; Cook, 2002). A variety of research tools, procedures, and approaches have, thus, emerged over time to serve the needs of educational evaluation.

Purpose

This article provides a broad overview of methodological approaches available today for the conduct of educational evaluations and discusses their applications, advantages, and limitations. To set the stage, it first discusses the unique characteristics of evaluative inquiry, stressing on issues surrounding the object of inquiry, units of analysis, and question-method alignment. The article aligns different classes of research methods with guiding evaluation questions and decision-making contexts where the evidence is likely to be used. It concludes by touching on the interdependence of evaluation methods, use and values; the role of stakeholders; and the methodological issues relevant to current movements of evidence-based education (EBE), comparative international assessment, and the globalization of the education enterprise. Salient terms are defined in the text or glossary.

To examine how the use of various methods fits within the larger scheme of conceptualizing, designing, and conducting evaluations, consider **Table 1**. The table shows a range of possible educational objects that could be the focus of evaluations, linked to corresponding questions and decision-making contexts. Although not an exhaustive list, the table serves as a starting point for discussing key evaluation methodology issues.

The Nature of Evaluative Inquiry

Understanding the Object of Inquiry

A first concern in conducting an evaluation is the object of inquiry (see **Table 1**). Michael Scriven coined the term ‘evaluand’ to refer to what we evaluate in an evaluation (Scriven, 1991). A precondition to appropriate choice and application of evaluation methods is to have a sound grasp of the evaluand and a clear specification of the purposes for doing an evaluation.

In most educational evaluations, the evaluand is a program implemented in a school setting (**Table 1**, example 1). As shown in **Table 1**, however, the focus of an evaluation can include a variety of other entities (see **Table 1**, examples 2–7). Recently, there is growing recognition that many field-based educational interventions have socially complex and multivariate configurations. Further, they

Table 1 Educational evaluation: Evaluand types, evaluation questions, and decision-making contexts

<i>What is evaluated: Evaluand type</i>	<i>Specific example</i>	<i>Evaluation question(s)</i>	<i>Possible decision-making contexts</i>
1. A curriculum, program, or educational service	A new science curriculum for K-12 grades	Is the new program more effective in improving student achievement than the current one?	Actions on program adoption or continuation
2. A system-wide policy	Class size-reduction policy in elementary schools	How consistently is the policy being implemented in schools? To what extent does the policy help improve students' academic achievement when consistently implemented?	Decisions on improving policy implementation levels
3. An institution or school	A college undergoing national accreditation	Do services, operations, and outcomes of the institution meet standards of quality set by external accreditation boards? What are the areas in which improvements or changes are needed?	Decisions on institutional accreditation Strategic planning and goal-setting in programs and in the institution
4. Personnel	Teachers and staff in a school	Does staff performance meet criteria valued in the institution?	Conducting annual performance reviews of personnel Giving formative feedback to individual staff
5. Individual Students	Elementary-level students	Have the students met expectations set by the teachers and school authorities to move to the next level? What are the students' strengths and weaknesses in different domains?	Decisions on student promotion Individual educational planning
6. An education system, group of institutions or programs within institutions	Secondary schools in a system	What percent of students, in different subgroups, meet the achievement standards set by local or public authorities?	Decisions related to public accountability
7. An educational product	An Internet-based science curriculum	Does this product meet quality criteria set by relevant experts?	Actions on curriculum adoption
8. A joint policy and program package	School-level class size-reduction policy combined with classroom-level curriculum delivery strategies	What types of curriculum delivery strategies work best under different class size conditions?	Resource-allocation decisions in a system to reduce class sizes and utilize effective curricular strategies

frequently operate in open organizational systems with multiple levels.

A complex intervention example from **Table 1** would be a class-size reduction policy implemented at an upper organizational level, operating in unison with a teacher-delivered, classroom curriculum (example 8). Both these entities jointly target a common outcome, improving student achievement. Curriculum resources, delivery processes, teacher characteristics, and infrastructure supports, are all potential factors influencing outcomes. Some factors could interact with one or both intervention elements. These multiple and dynamic variables together define the program and its context.

Not all entities evaluated are as complex as example 8. Regardless, a thorough understanding of the evaluand is necessary to ensure that it is appropriately specified in the

conceptual framework underlying an evaluation design. A limited or erroneous specification of the object of inquiry poses barriers to formulating relevant evaluation questions, and leads to inadequate data-gathering tools and analytic approaches. Findings might miss the mark in such circumstances and lose their utility in decision-making contexts. A common error in specifying complex programs is in the form of a binary variable (denoting its presence or absence) that oversimplifies an entity's multi-factor composition.

Program theory-driven evaluations (also 'logic models' in glossary) are useful heuristic mechanisms for improving evaluation designs of complex programs like example 8. Logic modeling techniques appeared in the methodological literature in the 1990 (Chen, 1990; Donaldson, 2007; Rossi *et al.*, 2003; Weiss, 1997).

Basically, logic models are flow charts. They help identify major variables that define a program and its context, and show how these variables logically lead to outcomes in specified populations. In doing so, logic models reveal the implicit and explicit assumptions underlying how particular programs are expected to work. Logic models make it easier for researchers to focus on relevant variables, to formulate better questions, and to test hypotheses about various interrelationships and effects. A program's logic may be based on some underlying social science theory or on beliefs of the developers and users.

Units of Analysis in an Evaluation

A related but not always overlapping factor with the object of inquiry is the unit of analysis in an evaluation. It is the units of analysis that are formally selected, observed, described, analyzed, or appraised in an evaluation. The examples in **Table 1** illustrate differences in the units of analysis in different evaluation problems.

In an evaluation examining effects of an intervention on school populations, the units of analysis are typically the students on whom outcomes are measured (see examples 1, 2, and 6 in **Table 1**). In a study of a school undergoing accreditation, on the other hand, the institution as a whole would serve as both the unit of analysis and object of inquiry (example 3, **Table 1**). When a product is being evaluated, such as a packaged curriculum, the unit as a whole is again the object of inquiry (example 7, **Table 1**). When measuring outcomes at different levels of an organization, units of analyses can be taken from multiple levels of the system (example 8, **Table 1**). We could sample students as units from classrooms. We could also sample classes as units from schools or larger education systems.

Identifying appropriate units of analysis is an important concern from both a conceptual and methodological standpoint in designing evaluations. Inferences and overall conclusions ultimately pertain to these units. As evident, the units relate directly to the object of inquiry even when they are not one and the same.

Aligning Evaluation Purposes and Questions with Methods

Along with clarity on the evaluand and units of analysis, the evaluation purposes and specific evaluation questions must be clear to both evaluation researchers and users of the information. The motivating purposes of an evaluation directly affect the choice of methods. Well-framed questions instantly point to the methods that are most appropriate.

Traditionally, four main categories of evaluations have been identified in the literature, based on their guiding purposes: needs assessments; process evaluations and

implementation monitoring studies; impact evaluations; and (cost) efficiency studies. Yet another type of study, labeled as an evaluability assessment, is driven by a need to assess the readiness of a program to be evaluated. Evaluability assessments occur before an entity is subjected to a formal evaluation (Rossi *et al.*, 2003). More comprehensive evaluation design frameworks have also been proposed, and these frameworks combine the above purposes to different degrees and in different ways.

A more general classification of different evaluations, given by Scriven (1991), distinguishes between formative and summative evaluations. This dichotomy is based on the decisions and actions taken with results. Formative evaluations utilize results for developing, modifying, or improving the evaluand (see the second question in example 3 of **Table 1**), while summative evaluations use findings for making some definitive or final judgments of worth (see example 1 in **Table 1**).

A possible range of evaluation questions associated with different types of evaluands is shown on the right-hand side of **Table 1**. To elaborate on the principle of question-method alignment, let us review the examples closely.

In **Table 1**, the first example “Is the new program more effective than the current one?” is an example of an impact evaluation question. The interest here is in a program's effects in terms of outcomes on some target population. The framing of the question clearly suggests a need for comparative experimental tools to assess a new program against an existing one.

The second question in **Table 1**: “How consistently is (a) policy being implemented in schools?”, on the other hand, reflects a process evaluation problem. Here, useful methodological tools would help describe the implementation processes of a policy in schools against the expected policy theory.

The third question, showing an accreditation example, reflects a need for an in-depth examination of an institution as a self-contained unit. Here, case study methodology may be the best match.

Product and performance assessments are common tools for curriculum and staff evaluations. Product or performance reviews call for valid and reliable assessment instruments that tap into relevant constructs, generating data that can be interpreted meaningfully and reliably.

In contrast, altogether different methodological approaches would help answer needs assessment questions. Needs assessment problems are not represented in **Table 1** call for evaluations of the scope and dimensions of a community-wide problem, such as, and how widespread is the high school dropout problem in New York? The information can be fed into the planning and development of a suitable intervention to address the issue. Large-scale survey methods, archival data analyses, and focus group interviews are common approaches for conducting needs assessments.

Integrated designs that are more ‘systemic’ have been recommended for educational program evaluations, both in the past and, more recently, in the context of evidence-based education. A well-known systemic framework is the context, input, process, product (CIPP) model (Stufflebeam, 2004). The CIPP approach is a developmentally oriented and organization-focused approach to evaluation, with an aim to improve rather than simply prove whether an entity and its parts work (p. 247).

Systemic designs are particularly necessary when the evaluand has a complex definition and operates in a complex environment. Because of their breadth, systemic evaluation questions demand the integration of multiple research methods, including necessary tools to gauge program effectiveness (Chatterji, 2005). Such evaluations may call for a layered set of questions aimed at capturing how, why, and when complex entities work. For a school-based program, guiding questions might look like this: What types of delivery practices produce the desired effect? Under what conditions, and with what type of student are the effects realized?

Evaluation Design Options and Methodological Tools

We now turn to three broad classes of research designs that offer methodological options to evaluation researchers to address a range of questions: quantitative, qualitative, and mixed-method designs. The three design categories are characterized based on the research purposes they best serve, underlying assumptions, and their strengths and limitations in applied contexts.

Quantitative Design Options

In the quantitative tradition of research, data on all variables are collected, coded, and analyzed in numeric terms. This tradition is founded on logical-positivistic and post-positivistic philosophies of science. The purpose of science, from this perspective, is the pursuit of universal laws that explain cause and effect relationships and natural covariations. A basic belief in this tradition is that there is an objective reality outside the researcher that can be investigated (Robson, 2002; Best and Kahn, 2003).

There are three main methodological options in this category for evaluation researchers: randomized experiments, quasi-experimental designs, and nonexperimental or observational designs. The degree of certainty with which each allows causal inferences depends on how well their assumptions are met and how rigorously the methods are executed in field settings.

Manipulation, comparison, and control are the hallmarks of a sound experiment aiming to detect causal effects of programs. When a clean, causal link can be established between a manipulated experimental variable

(say, a mathematics program) and a desired outcome (mathematics achievement in students), the study is said to have ‘internal validity’. An isolated effect is a ‘net’ effect with extraneous or confounding factors eliminated or controlled; otherwise, effects are described as ‘gross’ (Rossi *et al.*, 2003). When the effect generalizes to other units in the population or to different settings, the findings are said to have ‘external validity’. These two validity considerations have to be balanced in designing and executing field-based impact evaluations (Campbell and Stanley, 1963).

Randomized field experiments

In a randomized experiment, participants are placed in a treatment (usually, a new program) or a control group using procedures that ensure each person an equal chance of belonging in either condition. Case assignment is independent. Each participant in the experimental group is then exposed to the program, while with the control group is withheld from the same.

When randomization is well executed and there are enough participants, the procedure results in groups that are equivalent on pre-existing factors. Without any treatment, average outcomes are expected to be the same in both groups. If outcome means differ at the end of the experiment and favor the experimental group, the method permits a direct causal inference that the program caused that effect.

The main assumptions of experimental designs are:

- presence of a well-defined and manipulated treatment;
- stable and authentic delivery of the treatment;
- conditions that control or hold all other factors constant; and
- use of defensible outcome measures.

In theory, the experimental method may be applied in any impact evaluation seeking to make causal inferences linking a program to outcomes. The guiding questions for such studies are typically framed as follows: Does the entity have a statistically significant effect on measured outcomes, on average, as compared to a control program? What is the size and direction of that effect?

Strengths and limitations

Randomized experiments originated as a laboratory method. When assumptions are met, they are considered the most powerful designs for confirming cause–effect relationships (Schneider *et al.*, 2007). However, assumptions are difficult, often impossible, to meet under field conditions (Cook, 2002). Common threats to internal and external validity include poor execution of the randomization procedure; differential levels of subject attrition resulting in nonequivalent comparison groups; poor specification of evaluand; poor fidelity in program

implementation; treatment noncompliance (e.g., due to student absenteeism); nested, hierarchical, and open systems unaccounted for in designs that violate independence assumptions; overlooked or mis-measured environmental factors in analytic models; inaccurately interpreted effects of treatment; compromised external validity due to very tight controls; and inadequate sampling of relevant units, conditions, and time points (Chatterji, 2007; Shadish *et al.*, 2002; West *et al.*, 2000).

Quasi-experimental designs

A range of quasi-experimental designs offers different ways to address some of the preceding challenges in conducting field experiments under real-world conditions. Quasi-experiments are also the best design option when randomization is precluded due to ethical, logistical, or other barriers. Three designs are considered the best alternatives to randomized experiments for making causal inferences in impact evaluations (Shadish *et al.*, 2002; Schneider *et al.*, 2007; West *et al.*, 2000). They are:

1. matched, nonequivalent groups designs;
2. regression discontinuity designs; and
3. interrupted time series designs.

In the matched, nonequivalent groups design, two or more groups are identified and they are exposed to different conditions, one of which is the treatment under investigation. Randomization is not used for group assignment, as procedures for subject placement are often not in the researcher's control. The groups are matched on a few selected variables to establish equivalence, so that on average, they are similar (say, within a range of half a standard deviation unit). An example of an off-cited study compared the achievements of students attending public and private schools, and was done by Coleman, Hoffer, and Kilgore in 1982.

Both groups are usually measured before and after the treatment. Internal validity is instituted through statistical controls or matching methods using preexisting characteristics known to influence the outcome variable. A traditionally used matching variable is a pretreatment measure, such as prior achievement scores. Mean outcome measures are compared at the end of the experiment.

Several statistical adjustment techniques are available to correct for selection biases in applying quasi-experimental approaches. These include the analysis of covariance (ANCOVA), ANCOVA with correction for unreliable measurement; gain score analysis; and a number of regression-based econometric methods, such as propensity score matching. Unlike traditional matching described above, propensity score matching involves matching on multiple covariates that are probabilistically predicted to be most pertinent for establishing group equivalence. The use of instrumental variables is another econometric method that is now a part of an education evaluator's toolbox.

The regression discontinuity design is applicable when placements are deliberately made to nonequivalent treatment groups based on a prior quantitative measure, usually a test to determine merit or a special need for treatment. For example, students may be placed in English as a Second Language (ESL) classes due to a diagnosed language deficiency, based on a cut-score on a placement test. Suppose, the outcome of interest is reading achievement administered at the end of the term. In applying this design, the reading outcome measure would be regressed on the placement measure when the experiment ends. If the ESL intervention works, a break would be evident in the regression line at the cut-point separating the treatment and control groups (see West *et al.*, 2000 for details).

In the last design option in this category, the interrupted time-series design, measurements on the outcome variable are collected at equally spaced intervals in a selected group. The intervention program is then introduced at particular points in time, breaking the series. In a graphic representation of the series of data points, spikes would be observed during each intervention spell, if the latter was effective.

Time-series analysis would be a useful method to examine changes in levels of achievement of children originally attending a public school, who are subsequently rezoned to a charter school. Time-series analyses can be enhanced with the use of comparison groups, so that changes in outcomes can be attributed more definitively to the intervention.

Strengths and limitations

Quasi-experiments need judicious application, as specific designs apply to particular types of evaluation problems. The designs discussed are considered to approximate the effects of randomized trials (Schneider *et al.*, 2007; West *et al.*, 2000). A main assumption is that variables defining the evaluand, outcomes, covariates, and matching variables are all well specified and properly measured. A further assumption is that all alternative hypotheses explaining outcomes are systematically ruled out. Covariate-based regression procedures assume that there is no interaction between the treatment and covariates – an assumption that needs testing.

There are several threats to quasi-experiments that mirror those listed for randomized experiments. Some common internal and external validity issues include the following: persistent group selection biases; poorly measured or selected covariates, matching variables, or outcomes; subject attrition resulting in nonequivalent samples; reduced statistical power due to loss of subjects or limited sample sizes; external factors influencing the outcomes during the course of the study (e.g., environmental, history or maturation of sample); lack of attention to nonlinear relationships; regression toward the

mean; weakened generalizability due to poor sampling of relevant units; and lack of attention to hierarchical, nested structures where applicable.

Observational (nonexperimental) designs

In this design category, we consider three overlapping approaches: survey research, secondary analysis of large databases, and retrospective case control designs. At a basic level, observational methods are useful in quantitatively describing the status or trends in a population on selected characteristics. When comparative experimentation is impossible, these approaches help establish associations (or covariations) among independent and dependent variables under less controlled and more natural conditions. The methods are so called because there is no direct manipulation of an experimental variable. Data from such studies permit descriptive, correlational, predictive, or explanatory analyses.

Large-scale education evaluations using survey methods, both cross-sectional and longitudinal, are widely used today. For example, the National Center for Educational Statistics in the United States (U.S.) publishes periodic reports called *The Condition of Education*, based on sample survey research – a nonexperimental approach. In the developing world, India's National Sample Survey Organization conducts similar nationwide surveys to monitor population trends vis-à-vis national policy changes. Bangladesh's *Education Watch* reports published by the Campaign for Popular Education since 2002, are another international evaluation relying on nonexperimental methodology.

Most large-scale surveys employ probability sampling procedures, with mail-out, web-based, telephone, or interview-based surveys of subjects conducted by trained field researchers. Survey databases can be used to generate a series of subsequent and ongoing studies. A distinction, thus, is made between a primary analysis of an original survey, and a secondary analysis of the same database by other researchers.

With the development of newer statistical methods, some degree of causal inferences are possible with these data sets when supported with appropriate and theoretically grounded causal models, statistical controls, and variable measures (Schneider *et al.*, 2007). Although there is debate, some researchers recommend confirmatory evaluations using covariance structure models, incorporating causal inferences.

Survey research involves the design of self-report questionnaires that tap variables and domains of interest, such as public opinions, perceptions, and attitudes towards teachers, and schools. Researchers should gather the necessary evidentiary support to claim validity and reliability of survey-based indices. Target populations for education surveys may include students, teachers, parents, or administrators.

To see how a retrospective, case control design might be designed, let us consider a study investigating the

correlates of high school dropout status. To begin, cases are selected that are known to belong in two conditions (e.g., school dropouts vs. completers). Next, researchers go backwards in time to collect data on potential predictors and background characteristics of the groups. Predictor selection may be based on existing knowledge or social science theory. Cases in the two groups may be matched or statistically controlled on selected background characteristics, and an analysis ensues to identify variables maximally predicting the propensity of students to drop out or stay in school.

Strengths and limitations

Many evaluative questions are best answered with observational designs. Needs assessments (e.g., identifying incidence of smoking in schools), process evaluations (e.g., monitoring instructional delivery practices), trend studies (e.g., comparing achievement trends in different groups in education systems) call for some combination of descriptive, longitudinal, or correlational approaches that rely on surveys or large database analyses.

Some variables relevant for particular policy evaluations, such as gender, simply cannot be experimentally manipulated. For example, monitoring male versus female outcomes is necessary for in global education evaluations examining gender equity of initiatives such as the United Nations' Millennium Development Goals and Education for All. With carefully selected probability samples, surveys make it possible to extrapolate from sample statistics to population parameters with estimates of error. Further, survey methods are relatively easy and practical to administer to school populations; statistical procedures are, likewise, well developed and tested.

Simultaneously, survey questionnaires and data retrieved from existing databases place a measurement burden on researchers. Common sources of error include low response rates, missing data, low validity and reliability of survey indices due to poorly constructed items or data collection instruments, acquiescence bias and faking by respondents, and variabilities caused by irregular survey administration or data collection procedures. In addition, errors can be introduced through other breakdowns in the sampling procedures and fieldwork, threatening representativeness and external validity of findings. Causal-comparative inferences with observational data are, at best, tentative without addressing all the statistical issues that apply to experimental methods, such as covariate selection, control of subject selection biases, and appropriate modeling of moderators and mediators.

Qualitative Design Options

Qualitative research methods arise from a phenomenological philosophy of inquiry (Creswell, 2003). Here, the interest is in understanding and interpreting the essence

of a phenomenon through direct lived experience. Several discipline-based traditions to qualitative inquiry exist, including ethnography (from anthropology), phenomenology (from philosophy), hermeneutics (from theology, philosophy, and literary criticism), heuristics (humanistic psychology), and ethnomethodology (from sociology).

Common themes in these approaches are immersion of the researcher in the details of occurrences through direct induction; unobtrusive, unmanipulated, naturalistic study of real-world phenomena; personal contact, narration, and subjective insights; attention to processes and process dynamics; sensitivity to context and context dynamics; holistic approaches to interpreting and understanding phenomena; unique case orientation; and emergent rather than fixed, predetermined designs (Ryan and Bernard, 2000).

Qualitative data sources include running text, video or audio data from interviews, field notes of external observers, participant observations, and documents and artifacts. Focus group interviewing is another popular method used in evaluations. Although some methods of data reduction and categorization might yield counts and frequencies (such as classical content analysis rooted in linguistics), rich, thick description is a primary characteristic of qualitative methodology. Themes are extracted from data gathered from natural environments and organized to tell a story. By design, the researcher becomes an instrument in the processes of data collection, analysis, and interpretation.

Qualitative methods are useful in answering ‘why’ and ‘how’ things happen as they do; the approach emphasizes interpretation and use of insights and personal experiences of both the researchers and participants to deepen understandings. The purpose of one kind of qualitative inquiry is description, narration, and interpretation (e.g., content analysis and historiography). Researchers collect data through direct participation and observation; then, they record, code, and construct meanings from themes and patterns that emerge.

The aim of another kind of qualitative inquiry is in rendering expert, subjective criticism or value-based insights (e.g., literary criticism and critical review).

Specific aims vary with individual studies. For example, a case study may be more descriptive in purpose, employed to examine bounded social units and how they function. An ethnography may be critically interpretive, aiming to both describe and compare cultures from a value-based position.

There are techniques that add rigor to qualitative inquiry. These include triangulation, cross-validation, internal and external criticism, use of formal arguments, and justification of findings with constant comparison methods to cross-corroborate findings.

Strengths and limitations

Qualitative methods provide techniques and tools to document processes and contextual factors holistically and under natural conditions. They give evaluation

researchers insights into questions on how and why particular programs work, and the meaning and nature of effects found in authentic settings. They offer useful tools for conducting evaluability assessments, context evaluations, input and process evaluations, as well as in implementing systemic evaluation designs.

Qualitative methods, at present, are not as formally developed and well tested as the quantitative methodologies discussed in earlier sections. There are, thus, high levels of variability in both researcher training and applications of these methods in education and the larger field of evaluation.

The subjectivity of interpretations and confounding of researchers’ values and personal factors with the data may be viewed as a limitation in evaluation contexts. The more open-ended and flexible format of inquiry (as opposed to fixed, hypothesis-driven research) can lead to lack of closure and large volumes of hard-to-decipher and unwieldy data. Researchers must thus be careful in keeping their focus and utilizing defensible methods of data reduction and coding that are consistent with the goals of evaluation research.

Mixed-Method Design Options

The last category of evaluation methods involves mixed-method designs (see Caracelli and Greene, 1993). Although typologies have been proposed, there is no single way to do mixed-methods research. Depending on research purposes, it is possible for researchers to mix quantitative designs (e.g., surveys with comparative experiments), to combine qualitative methods (e.g., ethnography with focus group interviews), or align qualitative–quantitative designs (field observations, surveys, with comparative experiments). The preceding sections suggest that such approaches may be useful for either exploratory or confirmatory studies. An example of a mixed-method evaluation design intended for making generalizable, causal inferences, follows.

As the fundamental assumptions of comparative experimentation are violated or severely tested in field settings where most programs operate, broader design options are often necessary. Drawing on a body of historical work, Chatterji (2005; 2007; 2009) proposed the extended-term mixed-method design (ETMM). The ETMM approach offers several strategies to enhance the grade of evidence on the effects and workings of complex education programs in their natural settings. The design strategies include the use of:

1. expanded and layered questions, guided by logic models;
2. multiphase designs incorporating exploratory (formative) and confirmatory (summative) phases;
3. comprehensive evidence-gathering approaches with multiple research methods;

4. delayed use of comparative experiments, when applicable, to gauge effects; and
5. contextualized inference-making on program effects using both qualitative and quantitative evidence (Chatterji, 2005: 19–21).

Several ETMM design elements are found in the global research framework for the Strategic Impact Inquiry (SII) studies of CARE international's women's education and empowerment program, which has a highly complex configuration. CARE's SII studies focus on international agency collaborations that offer multiplicity of services to women in developing nations, including programs in basic literacy, health and nutrition education. To gauge the impact, the SII studies employ multiple methods and tools, including secondary analysis of quantitative databases, site-based interviews, surveys, and qualitative participant-oriented methods. The multiphase studies are conducted over a few years, using a learning-process approach that mimics the formative–summative model in ETMM. Researchers could not experimentally manipulate CARE's program due to logistical barriers; however, they applied a more systemic and contextually based mixed-method design to shed light on causal connections and make generalizations about the impact of a socially complex program.

Strengths and limitations

Mixed-method tools allow for more flexible evaluation design options, with a possibility of combining evidence in various ways, as dictated by the purposes of an investigation and object of inquiry. The advantages and disadvantages of mixed-method designs stem from the properties and limitations of particular quantitative or qualitative approaches. Trade-offs must be weighed. For example, well-executed sampling techniques focusing on persons, treatments, time points, and settings would give researchers several ways to increase generalizability and ecological validity of findings. On the other hand, qualitative tools will facilitate a more detailed study of context, changing processes, with attention to particular and unique cases. Causal inferences would be best made with experimental designs, scaffolded with descriptive and qualitative methods. However, when experiments are impossible to mount, observational designs can also be effectively combined with other methods to permit some degree of causal inferences. With multilevel interventions, analytic models accounting for nested-ness of data and longitudinal change designs would allow for better testing of sustained effects.

Other Issues in Evaluation Research

Interdependence of Methods, Use, and Values

When conducting evaluations, social research methods must be balanced against the social and public utility of

evaluation results, as well as the values of stakeholder groups and, in some cases, the values of evaluators themselves. The use of research methods generates data. However, the methods are mute with respect to whether the results are good or bad for society. Values are a necessary part of the evaluation research process and influence how the findings are viewed, prioritized, and used by stakeholders. Likewise, methods alone cannot solve the problem of evaluation utility. As evaluations aspire toward social betterment, the research is wasted if stakeholders do not use the findings (an unfortunate but common occurrence). There are, thus, three interdependent components to sound evaluative inquiry that run parallel to the roots of the evaluation enterprise: research methods, use, and values (Alkin, 2004).

To maximize evaluation use, formulation of evaluation questions should begin by identifying the information needs of relevant stakeholders. The evaluation issues examined should be given priority based on the identified stakeholder needs. How methods are selected, deployed, and resourced in evaluations, are affected by priorities set (see American Evaluation Association's *Guiding Principles for Evaluators*). Evaluation approaches and standards are inevitably value based, and can involve evaluator judgments to different degrees. In the connoisseurship model (Eisner, 1994), the evaluator serves as the expert, outside judge. In other approaches, selected stakeholder groups make judgments about results through the application of some systematic procedures, often facilitated by trained evaluators. Standard-setting methods for large-scale testing and educational assessment programs, for example, follow the latter procedures.

Evaluator–Stakeholder Roles and Relationships

Methodological decisions in an evaluation study are influenced by roles, responsibilities, and relationships among evaluators, participants, and other stakeholders. The role played by the evaluation researcher, and relationships developed among evaluators, sponsors, and stakeholders can vary widely depending on the type of evaluation study and the timing of the evaluation with respect to the evaluand (see Fitzpatrick *et al.*, (2003) for a classification of models attending to this criterion). Here again, professional standards and guidelines on ethics and propriety can be of great assistance to evaluation researchers (The Joint Committee On Standards for Educational Evaluation, 1994).

Some evaluations involve external evaluators, who maintain a clinical distance from both the entity and local stakeholders, controlling the design and implementation of a study. Other investigations call for more participatory approaches; here, evaluators may be internal or external to the community or organization, with program participants involved in the evaluative exercise to different degrees.

Methods appropriate for one form of evaluation may not be appropriate for another. Different types and levels of training, experience, and rigor may be required for a thorough implementation of different approaches. Typically, it is the evaluation researchers who contribute to the methodological expertise.

The Evidence Debate and Methodological Hierarchies

Internationally, the advent of policies pushing for evidence-based education (EBE) has given educational evaluation a new and different focus (see, e.g., the Evidence for Policy and Practice Information and Co-ordinating Centre in the United Kingdom; the Iterative Best Evidence Synthesis Programme in New Zealand; and the What Works Clearinghouse in the US). EBE is based on the principle that interventions, tools, and materials used to educate or provide services should be scientifically tested and proven to work. Borrowing from the Cochrane Collaboration's advocacy for evidence-based medicine in the US, these organizations have established standards for synthesizing evidence on what works in education from published research and evaluation studies. Schools in the US are now asked to use research-based practices.

A widely published position among policymakers worldwide calling for EBE, is that some research designs are better than others in yielding impact evidence on education programs. This policy position, called the evidence hierarchy, places randomized controlled experiments that are replicated on a large scale at the topmost rung of the methodological ladder for generating the best evidence on the causal effects of education interventions. Quasi-experiments occupy the next rung in the evidence hierarchy. The hierarchy rejects all other research methods as useful in generating sound impact evidence.

Evidence hierarchies have resulted in heated debates in the educational research and evaluation community. As is obvious from the discussion in this article, no single method is foolproof, particular methods are best suited for particular purposes, and all of them can be implemented poorly or well. Evidence-based policies place a responsibility on the research and evaluation community for using appropriate methods with rigor and integrity.

Globalization and Large-Scale Assessments

Nations now compete in a global marketplace. Governments are, thus, paying attention to the performance of their educational systems and outcomes of education. Global policy circumstances have serious implications for the proper use of tests, assessments, and evaluation methodology. They also influence the roles and responsibilities borne by assessment and evaluation specialists.

See also: International Evaluations; National Assessment Programs: The Example of the U.S. National Assessment of Educational Progress; Needs Assessment in Education; The Role of Stakeholders in Educational Evaluation.

Bibliography

- Alkin, M. C. (ed.) (2004). *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks, CA: Sage.
- Best, J. W. and Kahn, J. V. (2003). *Research in Education*. Boston, MA: Pearson.
- Campbell, D. T. and Stanley, J. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally.
- Caracelli, V. J. and Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis* 15(2), 195–207.
- Chatterji, M. (2005). Evidence of "What Works": An argument for extended-term mixed-method (ETMM) evaluation designs. *Educational Researcher* 34(6), 13–24.
- Chatterji, M. (2007). Grades of evidence: Variability in quality of findings from complex field interventions. *American Journal of Evaluation* 28(3), 3–17.
- Chatterji, M. (2009). Enhancing impact evidence on the how global educational initiatives work: Theory, epistemological foundations, and guidelines for applying multiphase, mixed method designs. In Ryan, K. B. and Cousins, J. B. (eds.) *The SAGE International Handbook of Educational Evaluation*. Thousand Oaks, CA: Sage Publications.
- Chen, H. T. (1990). *Theory-Driven Evaluations*. Newbury Park, CA: Sage.
- Cook, T. D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Educational Evaluation and Policy Analysis* 24(3), 175–199.
- Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative, and Mixed-Method Approaches*, 2nd edn. Thousand Oaks, CA: Sage.
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., et al. (1980). *Toward Reform in Program Evaluation*. San Francisco, CA: Jossey-Bass.
- Donaldson, S. I. (2007). *Program Theory-Driven Evaluation Science: Strategies and Applications*. New York: Erlbaum.
- Eisner, E. W. (1994). *The Educational Imagination*, 3rd edn. Upper Saddle River, NJ: Prentice-Hall.
- Fitzpatrick, J. L., Sanders, J. L., and Worthen, B. R. (2003). *Program Evaluation: Alternative Approaches and Practical Guidelines*, 3rd edn. Boston, MA: Pearson.
- Guba, E. (1969). The failure of educational evaluation. *Educational Technology* 9(5), 29–38.
- Robson, C. (2002). *Real World Research*, 2nd edn. Malden, MA: Blackwell Publishing.
- Rossi, P. H., Freeman, H. E., and Lipsey, M. W. (2003). *Evaluation: A Systematic Approach*, 6th edn. Thousand Oaks, CA: Sage.
- Ryan, G. W. and Bernard, H. R. (2000). Data management and analysis methods. In Denzin, N. K. and Lincoln, Y. S. (eds.) *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage.
- Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., and Shavelson, R. J. (2007). *Estimating Causal Effects: Using Experimental and Observational Designs*. Washington, DC: AERA.
- Scriven, M. (1991). *Evaluation Thesaurus*, 4th edn. Newbury Park, CA: Sage.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton-Mifflin.
- Stake, R. E. (1984). Program evaluation, particularly responsive evaluation. In Madaus, G. F., Scriven, M., and Stufflebeam, D. L. (eds.) *Evaluation models*, pp 287–310. Boston, MA: Kluwer-Nijhoff.
- Stufflebeam, D. L. (2004). The 21st century CIPP model. Origins, development and use. In Alkin, M. C. (ed.) *Evaluation Roots: Tracing Theorists' Views and Influences*, pp 245–266. Thousand Oaks, CA: Sage.
- The Joint Committee on Standards for Educational Evaluation and Sanders, J. R. (1994). *The Program Evaluation Standards: How to Assess Evaluations of Educational Programs*, 2nd edn. Thousand Oaks, CA: Sage.
- Weiss, C. H. (1997). How can theory-based evaluation make a greater headway? *Evaluation Review* 21, 501–524.

West, S. G., Biesanz, J. C., and Pitts, S. C. (2000). Causal inference and generalization in field settings. In Reis, H. T. and Judd, C. M. (eds.) *Handbook of Research Methods in Social and Personality Psychology*, pp 40–84. Cambridge, UK: Cambridge University Press.

Further Reading

Ryan, K. B. and Cousins, J. B. (eds.) (2009). *The SAGE International Handbook of Educational Evaluation*. Thousand Oaks, CA: Sage.

Relevant Websites

<http://www.eval.org> – American Evaluation Association.
<http://www.care.org> – CARE: Defending Dignity, Fighting Poverty.
<http://www.csps.emory.edu> – CSPS: CARE Strategic Impact Inquiry on Women's Empowerment.
<http://www.minedu.govt.nz> – New Zealand Ministry of Education.
<http://eppi.ioe.ac.uk> – The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre).
<http://www.w-w-c.org> – What Works Clearinghouse.

Evaluation Reporting and Communicating

R T Torres, Torres Consulting Group, New Orleans, LA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Audiences – Those who receive information about the evaluation and its findings. Audiences include, but are not limited to stakeholders – for example, staff from other programs who would benefit from information about a particular program.

Evaluand – The object of an evaluation; the program, project, process, organization, product, or issue that is the subject of an evaluation.

Formative evaluation – Evaluation that provides information about how a program or organization operates and how to improve it. Typical audiences for formative evaluation are program staff and managers.

Stakeholders – Those individuals, groups, or organizations who may be affected by the planning, activities, and/or findings of an evaluation. For instance, program staff are stakeholders. Stakeholders might also include groups such as parents of students in a particular school program. They do not directly participate in a program, but do have a vested interest in it.

Summative evaluation – Evaluation that provides information about the overall effectiveness, impact, and/or outcomes of a program. Summative evaluation serves accountability purposes. Typical audiences for summative evaluation are funders, sponsors, and/or organizational leaders.

Introduction

Evaluation tends to emphasize one of two purposes and sometimes both: accountability or summative evaluation; improvement or formative evaluation. The following definitions focus on summative evaluation, the process of determining the evaluand's merit or worth:

Evaluation: The systematic investigation of the worth of an object. (Joint Committee on Standards for Educational Evaluation, 1994: 3)

Program evaluation is... a process dedicated to making, generating, or feeding judgments about the worth or significance of a program. (Mathison, 2004: 334)

Evaluation refers to the process of determining the merit, worth, or value of something, or the product of that process. Terms used to refer to this process or part of it include appraise, analyze, assess, critique, examine, grade, inspect, judge, rate, rank, review, study, test. (Scriven, 1991: 139)

Typically, merit or worth is assessed in relationship to some criteria or standards.

Evaluation is defined as the “periodic assessment of the relevance, performance, efficiency and impact (both expected and unexpected) of the project in relation to stated objectives.” (Fort *et al.*, 2001: 1)

Evaluation is the comparison of the condition or performance of something to one or more standards. (Stake, 2004: 4)

The evaluation process normally involves some identification of relevant standards of merit, worth, or value; some investigation of the performance of evaluands on these standards; and some integration or synthesis of the results to achieve an overall evaluation or set of associated evaluations. (Scriven, 1991: 139)

Other definitions emphasize evaluation in service of program improvement:

Program evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming. (Patton, 1997: 28)

Program evaluation is the use of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social action to improve social conditions. (Rossi *et al.*, 2004: 16)

Program evaluation examines programs to determine their worth and to make recommendations for programmatic refinement and success. (Spaulding, 2008: 5)

Most evaluators agree that evaluation communicating and reporting is the central activity through which evaluation serves its intended purpose, regardless of the specific focus of that purpose. Almost 30 years ago, Lee J. Cronbach positioned evaluation as fundamentally a learning process:

The proper function of evaluation is to speed up the learning process by communicating what might otherwise

be overlooked or wrongly perceived Success is to be judged by . . . success in communication Payoff comes from the insight that the evaluator's work generates in others. (Cronbach, 1982: 8)

The value of an evaluation is no greater than the effectiveness of its reporting in, meaningfully conveying the evaluation's findings. In the author's view, this perspective rightfully burdens the reporting and communicating process with maximizing the possibility that the evaluation audience will assimilate and learn from the evaluation findings. It contrasts sharply with historically traditional views which held that the evaluation's job was done once a comprehensive, valid, and credible evaluation report had been written and delivered to the evaluation's primary stakeholders (i.e., typically those who funded and/or requested the evaluation).

This view has shifted gradually but solidly toward a more integrative, participatory, and learning approach to evaluation (see, e.g., Patton, 2008; Preskill and Torres, 1999; Russ-Eft and Preskill, 2009), which engages stakeholders in all phases of the evaluation in an effort to maximize the credibility of the evaluation process and the use of its findings. The nature of the reporting effort and indeed the communication that takes place throughout the evaluation is crucial to meeting this goal. While a final evaluation report is the most commonly used means of conveying findings, the degree to which it is accepted, seen as credible, and ultimately used in most cases is largely mediated by the communication that takes place during the evaluation between the evaluator(s) and audiences for the report. This communication includes not only written interim reports that may be delivered to stakeholders, but also the verbal and written interaction that occurs during the evaluation. In a participatory approach to evaluation, the evaluator(s) and the intended evaluation users (see Patton, 1997, 2008) collaborate throughout all phases of the evaluation, from its design to follow-up activities specifically supporting the use of the findings. The more meaningful this engagement with stakeholders is, the more likely they are to use the final evaluation report.

The remainder of this article overviews the purposes of communicating and reporting, and then focuses on the final evaluation report and ways to increase its use. Further, additional strategies for increasing the effectiveness of communications and reports are covered: (1) adding visual appeal to evaluation products, (2) using highly interactive formats, (3) maintaining communication, (4) engaging audiences with creative formats, and (5) introducing cross-cultural considerations. The article concludes with the development of a communicating and reporting plan, designed to address all considerations of the endeavor: audiences, purposes, formats/strategies, timing, and resources.

Purposes of Communicating and Reporting

In general, the purposes of evaluation communicating and reporting are to convey information, build understanding and create meaning, and/or support decision making. Specifically, during the evaluation communicating and reporting serves to (1) include stakeholders in decision making about evaluation design and activities; and (2) inform them about (i) specific upcoming evaluation activities, (ii) overall progress of the evaluation, and (iii) interim findings. After the evaluation, it serves to (1) inform about a program and its evaluation to create awareness/support, (2) communicate final findings to support change and improvement (formative evaluation), and/or (3) communicate final findings to show results and demonstrate accountability (summative evaluation). These purposes can best be met by making information in the communications and reports accessible and easy to assimilate, and by facilitating interaction wherever possible and appropriate.

The Final Evaluation Report

As mentioned above, the final evaluation report is the most commonly expected and used format for communicating evaluation processes and results. Yet, evaluators continually experience frustration with hours spent on writing reports that are seldom read, or shared with others. This frustration is especially apparent when learning and evaluation use is an expected goal of the evaluation. Typically, the contents of a comprehensive, final evaluation report are modeled after a social science research report (see **Table 1**). This academic focus has its origins in the notion that sufficient information about methodology, findings, and conclusions so that (1) readers can understand how claims for an evaluation's validity are justified, and/or (2) another researcher/evaluator could replicate the study.

There are ways, however, in which these researcher-focused purposes can be served, and evaluation reports also can be more readable, accessible, and useful for audiences who are primarily concerned with the substance of the evaluation findings (especially in cases of participatory evaluation when stakeholders have been involved in decision making about the evaluation methods; see **Table 2**).

Adding Visual Appeal to Evaluation Products

Making evaluation products visually appealing is a fundamental means of increasing their accessibility to users, specifically aiding in the assimilation of text-based documents. Text-based formats can be made more readable, visually appealing, and instructive, with features of layout and design and the inclusion of photographs. Visual appeal can be maximized with the use of video to convey evaluation findings.

Given the wide variety of visually appealing material all readers and audiences are exposed to in everyday life, there is an increasing expectation for easy access and readability in professional documents and products.

Table 1 Typical sections and contents of comprehensive written evaluation reports

Introduction

Purpose of the evaluation, including evaluation approach
Brief description of the program
Evaluation stakeholders/audiences
Relationship between/among organizations involved and those serving in evaluator roles
Overview of contents of the report

Program description

Program history, background, development
Program goals/objectives
Program participants and activities

Evaluation design and methods

Evaluation questions
Data collection methods (including participants and schedule) used to address each question
Analysis methods for each type of data collected

Findings/results

Description of how the findings are organized (i.e., by evaluation questions, logic model components, themes/issues, etc.)
Results of analyses of quantitative and/or qualitative data collected (usually represented in tables, charts, graphs, illustrations, and text)

Conclusions and recommendations

Conclusions drawn about the evaluation results
Recommendations for action based on these conclusions
Suggestions for further study, if applicable

Design, graphics, and layout

With the advances in technology, the sophistication of today's print media is ever growing, increasing expectations of lay and professional audience for readability and easy access to information. Therefore, it goes with evaluation. Text-based evaluation reports and other communications (e.g., executive summaries, newsletters, or brochures) can be made more readable, visually appealing, and instructive through the use of:

- variation in headings with different typefaces;
- bulleted lists;
- boxed text, charts, graphs, tables, figures, and graphic images; and
- variations in color from shades of gray to the full color spectrum.

The technology for producing evaluation products using these features is readily available in software applications accessible to most computer users. Many evaluation organizations are opting for professional design and layout of their reports, producing reports similar to those of annual reports from foundations and government agencies.

Photography

Still photographs can be added to evaluation reports and other text-based products to:

- illustrate the context and the realities of program participants;
- convey various perspectives; and
- show program activities.

Table 2 Strategies for increasing the relevance and utility of final evaluation reports

1. Maximize the accessibility of findings, conclusions and recommendations:
 - Describe the methodology briefly in an introductory section, and present it fully within an appendix
 - Organize the report by the evaluation questions (or other questions that guide the reader or main topic areas). Immediately after each question, describe the data that addresses the question, and then present a summary of the findings and recommendations. Detailed presentations of methods and findings can be included in appendices
 - Organize the report by the main body of the recommendations. For each recommendation, present conclusions and evidence (evaluation findings) to support it
 - In the case of numerous tables or graphs, put them in an appendix and reference the specific appendix pages throughout the earlier sections in the report that describes these findings
 - Provide a reader's guide or description of the report structure in the front inside cover of the report.
2. Select an organizing framework for the presentation of findings that is most salient for the primary report audiences:
 - The evaluation questions
 - The program's logic model or theory of action
 - Decision-making processes with the life cycle of the program
 - Emergent themes or categories that tell the program's story.
3. Get feedback from primary audiences about the style and focus of the final report.
4. Develop shorter versions of the final report in language and design/layout appropriate for different audiences.
5. Maximize dissemination of the final report as well as shorter renditions by:
 - Posting them on websites accessible to various audiences
 - Using e-mails, newsletters, staff meetings, and/or latest news sections of websites accessed by stakeholders to announce the availability of evaluation findings and where different versions of the report can be found.
6. Design and facilitate interactive working sessions with primary stakeholders (and others as appropriate):
 - To present findings and provide an opportunity for stakeholder input on the interpretation of findings and development of recommendations
 - To answer questions and carry out action planning based on findings and recommendations developed by the evaluator(s).

Photography is particularly useful when language or cultural barriers may inhibit participants' ability to easily assimilate information in written reports. It can also be used as a data-collection device to count, measure, compare, qualify, or track artifacts or information that can be captured visually. Further, when evaluation findings are presented through photographs, they can stimulate audiences' participation in interpreting important events and experiences, and engage them in dialog about the meaning and implications of findings.

Video

Video is an enhancement to still photography with motion and sound. In evaluation, it is particularly useful for:

- conveying information about new or innovative programs;
- disseminating findings to broader audiences than those directly involved with a program;
- providing findings to groups with limited time, and/or to those not accustomed to reading reports;
- projecting multisite evaluations to depict events and activities at different sites; and
- documenting evaluation processes.

Evaluation videos have the advantage of easy dissemination (depending on length) as electronic files: attached to e-mails, included within electronic versions of text-based reports, or PowerPoint presentations, and/or posted on websites. Longer videos can be disseminated on digital versatile discs (DVDs) that are mailed or delivered to evaluation audiences.

Using Highly Interactive Formats

Participatory evaluation is based on significant degrees of interaction and opportunities for learning that occur between the evaluator(s) and the evaluation user(s). Providing social contexts for decision making about evaluation activities and the interpretation of findings is one example of how group learning can support evaluation use. This section explores two highly interactive formats, working sessions and web conferences.

Working sessions

Evaluation working sessions typically engage stakeholders in some aspect of the evaluation. They are similar to meetings, but are usually planned and facilitated by the evaluator(s) to realize specific objectives. That is, they can be used throughout the evaluation, for:

- identifying concerns about the study;
- establishing buy-in;
- making decisions regarding the evaluation's implementation and/or use of the findings;

Table 3 Steps for carrying out evaluation working session

1. Assess and refine your skills for facilitating working sessions.
2. Build support for participation in working sessions by explaining the purposes and benefits to stakeholders.
3. Specifically tailor each working session for your participants and carefully plan it to achieve clear objectives.
4. Do your best to get key stakeholders at the table.
5. Determine what information or documents participants should receive ahead of time.
6. Consider conducting working sessions as part of other regularly scheduled meetings.
7. Provide sufficient background information at the beginning, and close session with a clear outline of what is to follow.
8. Reach common understandings at the outset about basic issues/information regarding the evaluand.
9. Use worksheets to organize activities.
10. Develop ground rules about confidentiality.
11. Follow up with stakeholders who may not have been able to attend.

- obtaining input about evaluation design, data-collection methods, and procedures; and
- engaging stakeholders in interpreting findings and developing recommendations and action plans.

Evaluators must draw on group process skills to successfully carry out working sessions with stakeholders (see **Table 3**). In addition, they require careful planning with consideration of purpose and specific objectives, participants, group dynamics, need for follow-up sessions, processes, and activities to engage participants, length of session, and materials preparation.

Web conferences

Web conferences allow participants to exchange messages in real time without being in each other's physical presence. They are meetings of two or more individuals conducted over the Internet with features of audio and visual communication, simultaneous viewing of electronic files including PowerPoint presentations. In evaluation, they are generally set up as working sessions (described above) and are used when face-to-face meetings are impractical or impossible. However, Web conferences:

- can lead to more timely dissemination of findings and subsequent decision making;
- can provide opportunities to include stakeholders who might not otherwise be able or available to travel to a meeting; and
- can be used for any phase of the evaluation work, but are best after some face-to-face meetings have occurred.

Computer programs for conducting Web conferences are available as part of many software applications and/or operating systems, and through Web-based service providers. Web conferences should be planned with the same

care as face-to-face working sessions, with the additional burden of making sure that all participants can successfully access the necessary computer and Internet technology.

Maintaining Communication

In addition to regular meetings that evaluators typically have with those stakeholders directly involved in the evaluation, communication can be maintained with e-mail and postcards and online collaborative tools, as well as newsletters and brochures.

E-mail and postcards

E-mail is the mainstay of most communication among those involved in professional activities, including evaluation. It provides for easy reading and assimilation of information at the beginning of the evaluation (e.g., regarding evaluation purpose and activities), throughout the evaluation (e.g., regarding data-collection activities, feedback from stakeholders, interim findings), and at the end of the evaluation (e.g., regarding final findings and next steps).

Postcards are an alternative to e-mail for short communications. They can be used to maintain contact and build interest in the evaluation, invite stakeholders to and/or remind them of data-collection events, and solicit reactions to preliminary findings. In this age of e-mail overload, evaluators can use postcards with color and graphic illustrations that accompany text to gain attention, pique interest, and solicit engagement of stakeholders.

Online collaborative tools

Many online collaborative tools are in use today that provide ongoing communication and/or project management functions for teams of individuals working on the same project. Programs available at the time of writing this article include Base Camp, Central Desktop, and Google Groups. Their features useful for evaluators are summarized in **Table 4**. Some orientation and training is likely necessary for fully utilizing the different features. These programs are probably the most useful for stakeholders involved in carrying out the evaluation.

Newsletters and brochures

Newsletters and brochures that contain evaluation information are typically created and disseminated by organizations that house the evaluand, house the evaluator(s), fund the evaluand or the evaluation, support the evaluand, and/or support the development of the evaluation profession (including capacity building). They provide another means of maintaining contact with evaluation audiences, and:

- can be disseminated to publicize program information, upcoming evaluation activities, and/or evaluation findings;

Table 4 Features and evaluation uses of online collaborative tools

<i>Feature</i>	<i>Evaluation use</i>
Email/ message boards	Share evaluation updates, ask questions, have others respond; keep message exchanges in one place
Milestones/ to-do lists	Establish milestones and create lists of tasks for specific upcoming evaluation activities
Calendar	Make aware of scheduling of activities for carrying out the evaluation
Wikis/whiteboards	Collaborate to write documents
Blogs	Provide access to continuous stream of information about evaluation events, findings, etc.
Document posting and retrieval	Share evaluation documents in progress as well as final reports

- can be distributed after evaluation activities to publicize evaluation findings or actions taken based on the findings;
- provide opportunities to reach broad groups of people;
- can be mailed and/or posted on the web;
- allow for presentation of text and graphics;
- are typically 1–4 pages in length; and
- typically provide contact information from where further information about the evaluation can be obtained.

Newsletters take many shapes and forms. A newsletter may be devoted solely to communicating evaluation findings from one study; or, articles about the evaluation activities and/or findings may be included in existing internal or external newsletters, which have a broader purpose. A newsletter can be a single page containing key findings, or it can be several pages created in a newspaper format describing particular elements of the evaluation and its findings.

Brochures typically are intended to generate interest and some follow-up on the part of readers. This might be to find out more about a program, an organization, or an evaluation. Brochures are sometimes used to disseminate and provide easy access to evaluation findings. Today's inexpensive desktop publishing software has dramatically increased the ease with which newsletters and brochures can be created. Word-processing software often contains design templates. Alternatively, evaluators may choose to engage professional services for design, layout, content development, and publishing.

Engaging Audiences with Creative Formats

The nontraditional, but creative formats of cartoons, poetry, and drama are gaining use among evaluators, particularly those who want to facilitate personalized representation and interpretation of evaluation findings.

Poetry, and drama, in particular, can “honor participants’ complex lives and highlight the emotion, drama, and subtleties of [their] stories” (Goodyear, 2007: 38). These formats are particularly good at capturing multiple voices including those of the evaluator whose perspectives are represented through their data analyses and presentations of the findings (often developed in collaboration with stakeholders). Each person who then views the cartoons, reads the poems, or participates as an audience member of a dramatic performance creates his/her own meaning and interpretations. For evaluation communicating and reporting purposes these creative formats share a number of commonalities; they:

- invoke creatively inspired individual interpretations and tend to engage audiences emotionally;
- increase the accessibility of evaluation information for particular audiences (see **Table 5**);
- may be more culturally or audience-appropriate in some cases than in others;
- are most typically used for conveying findings based on qualitative data-collection methods (e.g., interviews, observations, document analysis, etc.);
- are particularly powerful when used as a basis for interaction with program participants and stakeholders to interpret and discuss their meaning;
- warrant special efforts to communicate the methods used for their creation and to obtain stakeholder feedback prior to dissemination;

Table 5 Some suggested audiences for creative formats in evaluation communicating and reporting

Cartoons

- Program administrators and staff involved in the program’s design and implementation
- Youth
- Parents, guardians, caregivers, and teachers of participants
- Stakeholders with low-level reading ability
- Funders who have in depth experience with the context of the program and its participants

Poetry

- Program participants, administrators, and staff, especially those who have participated in case studies or provided qualitative data such as interviews
- Stakeholders of programs related to education, social services, arts and culture
- Audiences whose cultural/social traditions value verbal communications

Drama

- Program administrators and staff involved in the program’s design and implementation
- Other stakeholders who have experiences related to the program’s focus/purposes
- Youth
- Stakeholders of evaluands related to education, social services, arts and culture
- Stakeholders and other audiences who are available to attend a performance

- add costs in terms of time and budget to a typical evaluation endeavor; and
- are best used in conjunction with traditional formats (e.g., written reports and summaries).

Cartoons

In addition to the strategies described earlier in this article, cartoons are another way of adding visual appeal to evaluation reports. They:

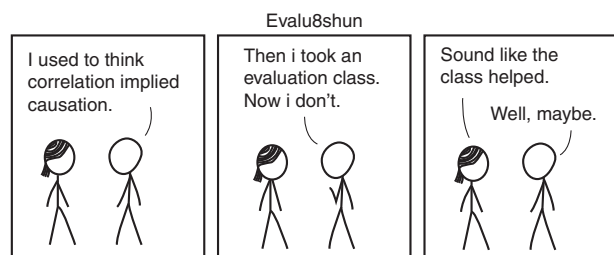
- can illustrate simple or symbolically complex issues in an accessible and concise format;
- may be used to convey information to children and those with low reading levels;
- can provide audiences with an informal, visually oriented understanding of program impact; and
- may be useful when language or cultural barriers inhibit participants’ ability to assimilate information from written reports.

One method for using cartoons in an evaluation (Chin, 2002) involved asking teachers, parents, and students exposed to an elementary reading program what amusing sidelights or funny incidents they had experienced in the program. Their responses reflected benefits, critical issues, and challenges associated with the program and also evidenced in other data sources. The evaluator created four cartoons from the interview findings and classroom observations to reflect major themes, and included them in the annual report. Feedback from stakeholders revealed that the cartoons (1) provided a visual picture of the program, (2) helped them remember the findings, (3) were helpful for those who did not read the entire report, and (4) were amusing and allowed the reader to feel the attitudes of participants.

Cartoons are often used to teach about evaluation to other evaluators and/or evaluation clients and stakeholders. The purpose is typically to engage the audience through humor (see **Figure 1**).

Poetry

As a means of communicating and reporting, poetry can illuminate the emotions, contradictions, and complexities



Adapted by kistler kreatives with permission from xkcd.com

Figure 1 Cartoon created to engage evaluation audiences with humor.

of evaluation contexts, and can be used to integrate the experiences and perspectives of multiple stakeholders (program staff, participants, funders, evaluators, etc.) into a collective voice. It is especially suited for communicating the tacit and the implicit, and making evaluation results more accessible by using participants' language and avoiding evaluation or other academic jargon. It can be developed from evaluation findings using key words, excerpts, and phrases that appear particularly poignant, insightful, and compelling.

For the evaluation of a mental health peer support center, MacNeil (2000) created the poem shown in **Figure 2**. It addressed the evaluation question, "What are the unique qualities a peer professional has to offer?" and was based upon responses from a staff focus group. The poem was created from a transcript of the focus group, reproduced on a large poster, and hung in the center's main meeting room. It served to report evaluation findings, remind staff of the evaluation question, and sustain attention on the evaluation. Occasionally, it has been used in orientations to the center, showing viewers that the center is a place that promotes a different way of thinking, and talking about mental health and peer support.

Insider knowledge and heart gladdening:
What peers have to offer

Art
Music
Poetry
Writing
Journaling
Inspiration
Rejuvenation
Survivor skills
Bringing light
A lot of courage
Capacity seeking
Important purpose
Listening to stories
Antidote to urgency
Releases from isolation
Knack for interviewing
Honoring of spirituality
Appreciation of animals
Retriever of better times
Getting off to a new start
Willingness to love oneself
A campground atmosphere
Finding that something special
Bringing people in from the cold
A common bond beyond mental illness
A different way of being in conversation
Understanding what people give value to
Rescuing the sparkling tellings from the told
Responding that honors ones lived experience
Helping people stand back from traumatic memories

Figure 2 Poem created to convey evaluation findings. From MacNeil, C. (2000). The prose and cons of poetic representation in evaluation reporting. *American Journal of Evaluation* 21, 359–367.

Drama

Drama is a more complex form of communicating and reporting. It can be used to recreate the lived experiences of program participants, combining realism, fiction, and poetic genres. Like poetry, it can help create a balanced representation of the perspective of multiple stakeholders and is particularly powerful when used as a basis for interaction with program participants and stakeholders. **Table 6** shows three formats for dramatic presentations of evaluation findings based on the work of the University of Michigan's Center for Research on Learning and Teaching.

Evaluators can develop options for creating the script of a dramatic performance based on evaluation findings through consideration of the following questions:

- should the performance be a verbatim conversation from interview or focus group transcripts or a dialog created by arranging excerpted quotes?
- how many of the participants in the data collection/evaluation should be represented in the performance?
- what criteria will I use to choose whom to represent in the performance? By stakeholder group? By level of involvement in the program? To show a variety of impacts of the program? To represent positive and negative aspects of the program?

Table 6 Summary of theatrical formats for communicating and reporting evaluation findings

Traditional sketch

- Used to evoke assumptions, motivations, feelings, experiences, themes or metaphors related to evaluation findings
- Developed from original evaluation data (especially focus groups and interviews)
- Performers can be characters or represent abstract themes/metaphors
- Facilitator guides discussion following performance

Interactive sketch

- Uses provocative vignettes from findings to engage audience members
- Typically portrays complexities, challenges of an issue/theme
- Following performance, audience dialogs with the actors who stay in character
- Facilitator guides discussion
- Actors repeat performance incorporating outcomes of discussion

Forum theater workshop

- Used to promote and extend discussions (e.g., action planning) related to findings
- Facilitator presents findings/topic
- Participants serve as both actors and audience members
- Participants create mini-scenes of situations, contexts, scenarios; exchange roles as actors/audience to explore/expand the topic
- Facilitator elicits discussion and reflections about performances

Adapted from <http://www.crlt.umich.edu/theatre/performances-workshops/kinds.php>.

- if the script will be a dialog created from excerpted quotes, should those quotes be attributed only to the individual who said them, or should I create composite characters?
- should the performance highlight the most significant or dramatic aspects of participants' experiences, or should it cover the range of experiences during the entire time frame of the data collection?
- should I develop the sketch as a typical conversation that took place during the interviews or focus groups, or should I include information gathered through observations, surveys, and document review (including quantitative statistics and demographic information)?
- should I include myself as a character in the performance? If so, should I fictionalize my dialog or take it verbatim from transcripts?

Cross-Cultural Considerations

From the standpoint of communicating and reporting, evaluators must design their communications and reports to maximize stakeholders' participation as contributors to

evaluation activities and users of evaluation findings. Participation and use should not be limited by communications and reports that stakeholders do not understand (due to diversity in age, language, educational levels, etc.), or reject because the communications and reports are insensitive to their cultural orientations. Evaluators can increase the cultural sensitivity of their communications and reports by:

- using a fully participatory approach with diverse evaluation stakeholders;
- forming teams of evaluators to include those from multiple cultural contexts;
- providing all communications and reports in the dominant language(s) of evaluation stakeholders;
- using language appropriate for diverse participants and reflective of their voices;
- avoiding primarily relying on written communications; and
- selecting photographs and video footage that are representative of stakeholder's race/ethnicity, socio-economic status, age, and gender.

Table 7 Importance of various evaluation audience characteristics to communicating and reporting

<i>Audience characteristic</i>	<i>Description</i>	<i>Example implication for communicating and reporting</i>
How accessible is this audience, and through what venues?	Audiences may be more or less accessible based on a number of factors: physical/geographical distance, access to use of the Internet, availability in terms of time, etc.	For less-accessible audiences, additional time and effort will be needed to strategize access and select appropriate formats, e.g., for busy executives
How familiar with the evaluand is this audience?	Different audiences are likely to vary in their familiarity with an evaluand; e.g., most funders will know less about the details of the program than program managers	More basic information about the evaluand is necessary for audiences who do not have regular contact with it
What attitude toward or interest level in the evaluand does this audience have?	Audiences may vary in their attitude toward (positive, negative, and indifferent) or interest level (low, medium, and high) with any given evaluand	Additional efforts may be necessary to get participation in presentations of evaluation findings to audiences who are disinterested in a particular program
What role in decision making about the evaluand or the evaluation does this audience have?	Role in decision making is often closely aligned with interest level and can vary from no role at all with respect to a particular evaluand to a significant role regarding a very similar evaluand	The greater the role in decision making, the more detailed that information about the findings the audience is likely to want
How generally familiar with research and evaluation is this audience?	Different audiences for any one evaluation likely vary on this dimension based on their prior experiences with research and evaluation	Those unfamiliar with research evaluation will need explanations of statistical terms and/or other aspects of methodology
What attitude toward or interest level in the evaluation does this audience have?	As audiences vary in their interest level or attitude in the evaluand, so may they vary with respect to the evaluation	Audiences hostile toward the evaluation may require more interaction and personal contact, with the evaluator adopting an educative role about the benefits of the evaluation
What experience using evaluation findings does this audience have?	Some audiences have never had any occasion to read about or use evaluation findings	Audiences with little prior experience using evaluation findings might need an explanation of the purpose and benefits of participating in a working session to interpret findings, for example
What is this audience's reading level and primary language?	Not all audiences have the same reading level and primary language as the evaluators who typically write evaluation reports	Audiences at lower reading levels might be successfully engaged with visually appealing and/or more creative formats such as cartoons or drama

Indeed, as addressed in the next section, evaluators have long been encouraged to match their communications and reports to audience characteristics and needs.

Creating a Communicating and Reporting Plan

Choices among the strategies and formats presented in this article must be based on a comprehensive set of considerations: audiences and their characteristics, purposes and timing, and resources. After addressing each of these, this article concludes with suggested formats for creating a communicating and reporting plan.

Audiences

Evaluation audiences are typically identified with the following questions. Which individuals or groups:

- were mentioned when the rationale for the evaluation was established?
- will be asked to provide data for the evaluation?
- should be involved in planning or carrying out the evaluation?
- might use the findings for program development or improvement?

- might use the findings for making funding or resource decisions?
- might be interested in the findings but are not in a decision-making position relative to the program?
- within the public at large should know about the evaluation's findings?

Audiences are typically categorized or prioritized into three groups. Primary audiences requesting the evaluation are the major decision makers for the evaluand, and include program staff, supervisors, senior managers, and/or funders. Secondary audiences are usually involved with the evaluand, but may have little or no daily contact with them; they include program participants, their supervisors or managers, and others impacted by the evaluand. Tertiary audiences usually are more distant, but may possibly be interested in the evaluation findings; they can include future program participants, general public, and members of same/related professions. It is important to note that use of these categories is arbitrary, relative to each other, and specific to any given evaluation. That is, a type of group that is considered a primary audience in one evaluation, might be considered a secondary audience in another. The most effective communications and reports are designed to

Step 1: List a single audience below: (individual or group) ①	② Audience characteristics							
	How accessible?	Reading level?	Familiarity with program or evaluand?	Attitude toward/ interest level in program?	Role in decision making about program or evaluation?	Familiarity with R&E in general?	Attitude toward/ interest level in this evaluation?	Experience using evaluation findings
Step 2: For each characteristic to the right, circle the response that best describes this audience.	Easily With some effort With substantial effort Don't know	High level Low level Non-reader Don't know	Very familiar Somewhat Familiar Not familiar Don't know	Positive/high +/– Negative/low Don't know	Crucial important minor no role don't know	Very Familiar Somewhat Familiar Not familiar Don't know	Positive/high +/– Negative/low Don't know	Substantial Some None Don't know
Step 3: Check the purposes for communicating with this audience.	③ Communicating/reporting purpose		④ Priority	⑤ Ideas about content	⑥ Format(s) to use	⑦ Dates	⑧ Resources needed	
	<input type="checkbox"/> Include in decision making about evaluation design & implementation		HI MED LO					
Step 4: Considering the audience and the purpose, prioritize each communicating and reporting task.	<input type="checkbox"/> Inform about specific upcoming evaluation activities		HI MED LO					
	<input type="checkbox"/> Keep informed about overall progress of the evaluation		HI MED LO					
	<input type="checkbox"/> Communicate Interim findings		HI MED LO					
Step 5: Note the implications that the characteristics of this audience may have for contents of communication/report.	<input type="checkbox"/> Inform about program and evaluation to build awareness, and/or support		HI MED LO					
	<input type="checkbox"/> Communicate final findings to support change and improvement		HI MED LO					
	<input type="checkbox"/> Communicate final findings to show results, demonstrate accountability		HI MED LO					
Step 6: Indicate appropriate formats to use.								
Step 7: Indicate date for communication/ report.								
Step 8: Indicate resources needed.	<input type="checkbox"/> Other:		HI MED LO					

Figure 3 Detailed evaluation communicating and reporting plan for each audience.

match the audience characteristics described in **Table 7**. Although the answers to these questions about each audience may be obvious, it is important to systematically consider each question for each audience because addressing their communicating and reporting needs may take additional planning and resources.

Purposes and timing

As explained in the section above on “Purposes of Communicating and Reporting,” timing is a key factor in considering what purposes are to be served by the evaluation’s communications and reports. To review, during the evaluation communications and reports typically are for the purpose of (1) including stakeholders in decision making about evaluation design and activities, and/or (2) informing them about specific upcoming evaluation activities, the overall progress of the evaluation, and/or interim findings. After the evaluation, communications and reports typically inform audiences about the program and its evaluation to create awareness/support, or communicate final findings to support change and improvement and/or to show results, to demonstrate accountability.

Planning for communications and reports that takes place during the evaluation is particularly important at its outset, especially for participatory evaluations designed to substantively engage stakeholders and facilitate the ownership of evaluation products that are produced once data collection and analysis is complete. Planning for post-evaluation communications and reports is especially important because they tend to be the more resource-intensive (see below) and require significant lead time to complete.

Strategies/formats and resources

At this point in the planning process for communicating and reporting, it is appropriate to consider the following questions for each audience:

- what strategies and formats are the best to use, given the audience and their characteristics, purpose, and timing of the communication?
- what resources (including time) will be needed for desired strategies/formats?
- what alternatives might be used if sufficient resources are not available?

① List audiences below:	② Describe any implications based on audience characteristics:*	③ Check the communicating and reporting purpose(s) for each audience:	④ List appropriate format(s) for each purpose**	⑤ Indicate date needed for each format:	⑥ Resources needed for each format:
		<input type="checkbox"/> Change, improvement			
		<input type="checkbox"/> Show results, accountability			
		<input type="checkbox"/> Build awareness/support			
		<input type="checkbox"/> Change, improvement			
		<input type="checkbox"/> Show results, accountability			
		<input type="checkbox"/> Build awareness/support			
		<input type="checkbox"/> Change, improvement			
		<input type="checkbox"/> Show results, accountability			
		<input type="checkbox"/> Build awareness/support			
		<input type="checkbox"/> Change, improvement			
		<input type="checkbox"/> Show results, accountability			
		<input type="checkbox"/> Build awareness/support			
		<input type="checkbox"/> Change, improvement			
		<input type="checkbox"/> Show results, accountability			
		<input type="checkbox"/> Build awareness/support			

*Relevant audience characteristics include accessibility; familiarity with evaluand, with research and evaluation; attitude toward/interest level in the evaluand, in the evaluation; role in decision making about the evaluand, about the evaluation; experience using evaluation findings, reading level.

**Note that a single format may serve more than one purpose.

Figure 4 Plan for post-evaluation communicating and reporting.

The communicating and reporting plan

A plan for evaluation communicating and reporting can vary in its detail and complexity. **Figure 3** shows a detailed plan that can be completed for each differently identified audience. It accounts for audience characteristics, the communicating and reporting purposes and their priorities, implications for content based on the audience characteristics, formats/strategies, dates, and resources needed.

An alternative to this plan focuses on post-evaluation reporting only (see **Figure 4**); it accounts for more than one audience on the same page and accounts for audience characteristics, communicating and reporting purposes, formats/strategies, dates, and resources.

See also: Evaluation and Accountability; Evaluation Governance and Planning; Evaluation Methodology; Evaluation Use; Moral and Ethical Issues in Evaluation; The Purpose of Educational Evaluation; The Role of Stakeholders in Educational Evaluation.

Bibliography

- Chin, M. C. (2002). Evaluation of the Flare Elementary School SRA Reading Program. *Report for the 2001–2002 School Year*.
- Cronbach, L. J. (1982). The case for participatory evaluation. *Educational Evaluation and Policy Analysis* **14**(4), 397–418.
- Fort, L., Martinez, B. L., and Mukhopadhyay, M. (2001). *Integrating a Gender Dimension into Monitoring and Evaluation of Rural Development Projects*. Washington, DC: World Bank.
- Goodyear, L. K. (2007). Poetry, performance, and pathos in evaluation reporting. In Kushner, S. and Norris, N. (eds.) *Dilemmas of Engagement: Evaluation and the New Public Management* pp 37–52. Bingley, UK: Emerald Group.
- Joint Committee on Standards for Educational Evaluation (1994). *The Program Evaluation Standards*, 2nd edn. Thousand Oaks, CA: Sage.
- Mathison, S. (ed.) (2004). *Encyclopedia of Evaluation*. Thousand Oaks, CA: Sage.
- MacNeil, C. (2000). The prose and cons of poetic representation in evaluation reporting. *American Journal of Evaluation* **21**, 359–367.
- Patton, M. Q. (1997). *Utilization-Focused Evaluation*, 3rd edn. Thousand Oaks, CA: Sage.
- Patton, M. Q. (2008). *Utilization-Focused Evaluation*, 4th edn. Thousand Oaks, CA: Sage.
- Preskill, H. and Torres, R. T. (1999). *Evaluative Inquiry for Learning in Organizations*. Thousand Oaks, CA: Sage.
- Rossi, P. H., Lipsey, M. W., and Freeman, H. E. (2004). *Evaluation: A Systematic Approach*, 7th edn. Thousand Oaks, CA: Sage.
- Russ-Eft, D. F. and Preskill, H. S. (2009). *Evaluation in Organizations: A Systematic Approach to Enhancing Learning, Performance, and Change*, 2nd edn. Boston, MA: Perseus.
- Scriven, M. (1991). *Evaluation Thesaurus*. Thousand Oaks, CA: Sage.

- Spaulding, D. T. (2008). *Program Evaluation in Practice: Core Concepts and Examples for Discussion and Analysis*. San Francisco, CA: Jossey-Bass.
- Stake, R. E. (2004). *Standards-Based & Responsive Evaluation*. Thousand Oaks, CA: Sage.

Further Reading

- Gullickson, A. and Keiser, N. (2002). Dissemination of evaluation findings and recommendations: Emerging approaches. *Paper Presented at the Annual Meeting of the American Evaluation Association*, Washington, DC, <http://www.wmich.edu/evalctr/ate/presentations/aea2002disspaper.pdf> (accessed October 2009).
- Hood, S., Hopson, R., and Frierson, H. (eds.) (2005). *The Role of Culture and Cultural Context in Evaluation: A Mandate for Inclusion, the Discovery of Truth, and Understanding in Evaluative Theory and Practice*. Greenwich, CT: Information Age.
- Hopson, R. (2009). Reclaiming knowledge at the margins: Culturally responsive evaluation in the current evaluation moment. In Ryan, K. and Cousins, B. (eds.) *International Handbook on Evaluation*. Thousand Oaks, CA: Sage.
- Mertens, D. M. (2004). *Research and Evaluation in Education and Psychology: Integrating Diversity with Quantitative, Qualitative, and Mixed Methods*, 2nd edn. Thousand Oaks, CA: Sage.
- Preskill, H. S. and Russ-Eft, D. (2005). *Building Evaluation Capacity*. Thousand Oaks, CA: Sage.
- Rosenstein, B. (2002). Video use in social science research and program evaluation. *International Journal of Qualitative Methods* **1**(3) Article 2.
- Russ-Eft, D. (2007). Communication in evaluation: A systems approach. In Hannum, K. M., Martineau, J. W., and Reinelt, C. (eds.) *Handbook of Leadership Development Evaluation*, pp 487–510. San Francisco, CA: Jossey Bass.
- Torres, R. T., Preskill, H. S., and Piontek, M. E. (1997). Communicating and reporting: Practices and concerns of internal and external evaluators. *Evaluation Practice* **18**(2), 105–125.
- Torres, R. T., Preskill, H. S., and Piontek, M. E. (2005). *Evaluation Strategies for Communicating and Reporting: Enhancing Learning in Organizations*, 2nd edn. Thousand Oaks, CA: Sage.

Relevant Websites

- <http://www.gatesfoundation.org> – Bill & Melinda Gates Foundation.
- <http://www.cdc.gov> – Centers for Disease Control and Prevention (CDC).
- <http://www.fordfound.org> – Ford Foundation.
- <http://www.macfound.org> – MacArthur Foundation.
- <http://www.cse.ucla.edu> – National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- <http://www.ohioline.osu.edu> – Ohioline.
- <http://www.extension.psu.edu> – Program Evaluation, Pennsylvania State University.
- <http://www.packard.org> – The David and Lucile Packard Foundation.
- <http://www.wmich.edu> – The Evaluation Center, Western Michigan University.
- <http://www.rwjf.org> – The Robert Wood Johnson Foundation: Health and Health Care Improvement – RWJF.
- <http://www.worldbank.org> – The World Bank.
- <http://www.nsf.gov> – US National Science Foundation (NSF).
- <http://www.wkkf.org> – W.K. Kellogg Foundation.

Longitudinal Evaluation Designs

X Ma, University of Kentucky, Lexington, KY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cohort design – A type of research design in which data are collected repeatedly from a group of individuals who possess a common trait, share a common condition, or receive a common intervention, often accompanied by a different group of normal individuals.

Cross-lagged model – A statistical model to determine cause–effect relationships in data that are obtained from nonexperimental, longitudinal research designs. In such a model, each variable is predicted (as effect) by all variables that precede it in time (as causes).

Cross-sectional design – A type of research design in which data are collected from a number of distinct age groups of individuals at a single point in time.

Multilevel growth model – A multilevel model to analyze data that are obtained from longitudinal research designs. Such a model typically estimates the trajectory of change over a period of time and identifies variables at different levels that contribute to variation in rate of change.

Multilevel model – An extension of the standard multiple regression technique from single level to multilevel. With simultaneous multiple regressions at different levels, such a model accounts for correlated responses at levels where dependencies of observations (or clustering effects) occur.

Panel design – A type of research design in which a group of individuals is surveyed or measured repeatedly over a period of time.

Structural equation model – A statistical technique to test and estimate cause–effect relationships based on both quantitative data and qualitative (causal) assumptions.

Time-invariant variables – Variables in longitudinal research designs whose values are constant over a period of time (e.g., race-ethnicity).

Time series design – A type of research design in which data are collected from a group of individuals on a small number of variables across a large number of points in time.

Time-varying variables – Variables in longitudinal research designs that take on different values at different points in time (e.g., achievement in mathematics).

Trend design – A type of research design in which a different sample of individuals is drawn from a clearly defined population at more than one point in time.

In a longitudinal evaluation, one or more clearly defined outcome measures are assessed repeatedly on a sample of subjects (e.g., students) over a period of time. Longitudinal data can be collected in the form of survey and experiment and can be analyzed for the examination of change and causality. In particular, longitudinal design is essential to study causality in nonexperimental conditions. In educational research and evaluation, longitudinal design has been gaining great momentum for a couple of decades because “the very notion of learning implies growth and change” (Willett, 1988: 346). Research questions under longitudinal design can be reasonably classified into two categories: (1) whether change has occurred and (2) when change has occurred. For example, an evaluation examining mathematics achievement over the entire secondary school (i.e., grades 7–12) is about whether change (i.e., growth) in mathematics achievement has occurred over time (see Ma, 1999), whereas an evaluation examining student avoidance of (i.e., dropout from) mathematics courses during the entire secondary school is about when change (i.e., dropping out of mathematics courses) has occurred over time (see Ma, 1997). With these general longitudinal research questions in mind, this article serves as a latest survey on methods to collect and analyze longitudinal data in educational evaluation.

Research Designs for Collecting Longitudinal Data

Researchers have employed a variety of research designs to collect longitudinal data with different degrees of threats to the validity of the resulting statistical inferences. A research design is truly longitudinal if, and only if, the same sample of subjects (e.g., students) is followed and evaluated repeatedly over time. However, researchers often settle into some sorts of quasi longitudinal design because a true longitudinal design requires considerable commitment of time, labor, and resource, especially when multiple waves of data are planned. This section reviews common strategies to collect longitudinal data.

Cross-Sectional Design

In a cross-sectional design, researchers work with a number of distinct age groups of students at a single point in time. For example, elementary students from grades 1 to 6 are surveyed at the end of a school year, resulting in data describing six groups of elementary students. Obviously, this is not a true longitudinal design. By comparing students from different grade levels or age groups, however, a sense of development emerges. Some researchers distinguish between standard cross-sectional design and simultaneous cross-sectional design. These two designs differ only in terms of how age is intended for data analysis. The former uses age as a control variable (i.e., to remove the impact of age or maturity), whereas the latter uses age as a chief variable (i.e., to study change due to age or maturity).

A cross-sectional design is easy to plan and implement, requiring the least amount of time, labor, and resource. However, threats to the validity of statistical inferences are serious. Obviously, this type of design lacks the ability to control different historical events that students of different ages have experienced, assuming that historical conditions do not attribute to differences across the grade levels. Even with a survey questionnaire that collects historical conditions unique to students at different grade levels as statistical adjustments, it is still impossible to overcome the fundamental difficulty that age effects are confounded with change effects because both come from the same variable – age.

Trend Design

In a trend design, a sample is drawn from a clearly defined population at more than one point in time. In most cases, a different sample is obtained at a different point in time (from the same population), and samples do not need to have the same size. For example, a sample of teachers is drawn from the population of teachers in a certain state to survey their support for the state-run testing system. One year later, another sample of teachers is drawn from the same population regarding the same issue. A comparison of data indicates change among teachers in terms of their support for the testing system. Obviously, this is not a true longitudinal design, unless the same sample is used at different points in time.

A trend design does not require a large investment of time, labor, and resource. It is a good choice for researchers who are concerned about change at the group level rather than at the individual level. A trend design opens the doors for a secondary data analysis of the state, national, or international database to evaluate change over time at the state, national, or international level. As different samples are used across time, it is impossible to identify how many individuals have experienced change. Comparability of data from different samples is always the

key issue. A trend design requires highly valid instruments to ensure that instruments do not discriminate against certain samples. If an instrument has not been psychometrically tested, there is a risk that the same instrument may function differently for different samples. When this happens, data from a trend design become incomparable.

Cohort Design

In a cohort design, cohorts are formed. A cohort is a group of persons who possess a common trait, share a common condition, or receive a common intervention, often accompanied by a different cohort of normal persons. Data are collected repeatedly from cohorts over a period of time and compared to examine whether change comes from age or maturity (which can often be observed across cohorts) or whether cohorts demonstrate differential outcomes of interest. For example, age at entry (into schooling) is often used as a cohort variable (i.e., cohorts of students who enter schooling at an age earlier or later than normal age) to examine whether early (late) entry promotes optimal learning.

A cohort design can be applied to data from both survey and experiment to separate age effects from change effects. It is also a good alternative to randomized experimental design. In fact, there are situations where a cohort design is the only appropriate research and evaluation approach. For example, to examine the impact of school violence (e.g., school bullying) on schooling outcomes of students, researchers cannot apply randomization to subject a group of students to school violence. Instead, a cohort of victims of school violence is followed, accompanied by a cohort of students who have not experienced school violence. Unfortunately, the main threat to validity, as inherited in a cohort design, also comes from the lack of randomization. Without randomization, confounding factors may come into play so that cohort differences may not be due to trait, condition, or intervention but some factors over which researchers do not have control.

Panel Design

In a panel design, a group of persons is surveyed or measured repeatedly (e.g., by means of a common questionnaire) over a period of time. Some panel designs collect data on a regular basis; others on an on-demand basis. For example, the Longitudinal Study of American Youth (LSAY) followed a national probability sample of some 3000 students from grades 6 to 12, collecting achievement and questionnaire data on a yearly basis (see Miller *et al.*, 2000).

As the same sample is used at different points in time, a panel design greatly advances over a trend design, allowing analysis of change at both individual and group levels. With an appropriate number of points in time, a panel

design is highly effective in investigating the pattern of change (i.e., characteristics of trajectory) over time. As a result, it is able to make a credible prediction of long-term change and the cumulative effects of change. A panel design is particularly popular in survey studies because it represents the best opportunity to collect a large quantity of detailed longitudinal data which allow not only the examination of change but also the identification of agents of change. Furthermore, a panel design is perhaps the best (nonlaboratory) approach to examine causality when a random sample is obtained and multiple waves of data are collected.

For a panel design, the main threat to validity stems from the continuous use of a common instrument at different points in time. The repeated use of a common test, for example, creates carryover effects (i.e., students become familiar with the test or even memorize certain items on the test). When parallel (equated) tests are hard to develop, carryover effects pose a serious threat to validity in that improvement in performance is contaminated by an increasing familiarity with the test. Like all other field-oriented (as opposed to laboratory-oriented) research and evaluation designs, a panel design is incapable of control over confounding factors. For example, in survey studies, such as the LSAY, if some clusters of students receive special performance-related treatments at a certain grade level as, say, a state education reform program, an artificial peak in the growth trajectory may occur among those students. Obviously, it is improper to relate this change (caused by an uncollected variable) in any causal manner with collected variables.

Time-Series Design

A panel design, as discussed above, often tends to focus on a large number of variables (measures) across a small number of points in time. In this sense, the power of a panel design resides in its ability to collect information on a huge number of variables, hopefully from which salient agents of change can be identified. The opposite of a panel design is a time-series design that tends to focus on a small number of variables across a large number of points in time. Because of this, a time-series design is particularly effective to examine the developmental characteristics of intraperson change. The large number of points in time produces data that are very sensitive to local change at the person level and allows the intraperson change curve to take on various mathematical functions (e.g., a cubic function) for optimal mathematical representation of change.

The following two examples help differentiate a panel design and a time-series design. A researcher who wants to identify the key causes dropping out of school, is likely to adopt a panel design to focus on the senior high school years (i.e., grades 10–12) and include a large number of student-level and school-level variables as potential candidates

for key characteristics of school dropouts. On the other hand, a researcher who wants to explore how the intention to drop out of school develops into action, is likely to adopt a time-series design that may begin to follow students right after elementary school (i.e., grades 6–12). Furthermore, instead of collecting data once or twice each year as likely in the panel design above, this researcher may interact with students every 2 or 3 months each year. With such a frequent interaction with students, this researcher will strive to minimize the disruption of learning on sampled students by asking them only a few questions (i.e., collecting data on a very small number of variables).

Often, a time-series design involves different age groups of persons so that intraperson change can be compared across age or maturity. Given that a time-series design and a panel design belong to the same family of research design, a time-series design also inherits threats to validity associated with its counterpart. When multiple age groups are involved in a time-series design, the potential issue of differential instrument functioning for different age groups also emerges (e.g., there is a need for a test to be psychometrically equated for use by persons of different ages so that cross-age comparisons can be made).

Statistical Techniques for Analyzing Longitudinal Data

To a large extent, longitudinal design is a modern concept, representing a considerable progress over the conventional conception that views individual change as an increment from before to after (e.g., the typical pre- and post-design is a good reflection of this conception). Willett *et al.* (1998) made a classical argument against this conception, stating that

Individual change takes place continuously over time, and comparison of each person's "before" and "after" status is not the most subtle, nor the most effective, way to reveal the features of that trajectory. To measure individual change well, a truly longitudinal perspective must be adopted – a sample of people must be followed over time allowing the research to collect multiple waves of data at sensibly spaced intervals. (p. 397)

This argument can be viewed as the very research condition that allows a meaningful statistical analysis of longitudinal data to address the two research questions posed at the beginning of this article. In this section, three general analytical frameworks are introduced: (1) multilevel growth model (to analyze whether change has occurred), (2) survival model (to analyze when change has occurred), and (3) cross-lagged model (via structural equation modeling, to specifically address the issue of causality in longitudinal research and evaluation).

The survival model and the cross-lagged model are more restricted than the multilevel growth model on data requirement in that both models require a true longitudinal design that follows one or more groups of persons repeatedly over a period of time. As a result, the survival and the cross-lagged models are appropriate analytical techniques for cohort design, panel design, and time-series design. The multilevel growth model, on the other hand, is quite flexible, and with an appropriate specification of the model, it can be applied to the cross-sectional design and trend design (apart from cohort design, panel design, and time-series design) (see Ma and Ma, 2007). This section focuses on the analysis of data that are obtained from true longitudinal designs (i.e., cohort design, panel design, and time-series design).

Multilevel Growth Model

Many research and evaluation studies in education have data that possess a multilevel or hierarchical structure (e.g., students nested within classes, classes nested within schools, and schools nested within districts). Because of these clustering effects, students are no longer independent in data analyzes (i.e., their responses are correlated, resulting in the loss of independence among observations) (e.g., Hox, 2002; Raudenbush and Bryk, 2002). Consequently, instead of traditional statistical techniques, multilevel modeling or hierarchical linear modeling (HLM) needs to be used to analyze education data (see Rasbash *et al.*, 2000; Raudenbush and Bryk, 2002). Multilevel modeling is an extension of the standard multiple regression technique. With simultaneous multiple regressions at different levels, multilevel analysis specifically accounts for correlated responses at levels where dependencies of observations (or clustering effects) occur; hence, it overcomes the difficulty in analyzing hierarchical data, as the assumption of independence of observations is relaxed. When the multilevel nature of educational data is coupled with the longitudinal perspective, the multilevel growth model is the most effective statistical technique.

Ma (1999) represents a typical multilevel growth model in the form of a three-level HLM (see Raudenbush and Bryk, 2002). Using longitudinal data from LSAY discussed earlier, Ma (1999) examined (among other issues) whether there was a statistically significant growth in mathematics achievement among American students when they progressed through secondary schooling. The first level of the model is a set of separate linear regression models, one for each student. These regression equations model students' outcome scores (dependent variable) on their grade levels (time variable):

$$Y_{ijt} = \pi_{0ij} + \pi_{1ij}(\text{grade})_{ijt} + R_{ijt}$$

where Y_{ijt} is the mathematics achievement score for student i in school j is in at occasion t , $(\text{grade})_{ijt}$ is the grade

that student i in school j is in at the testing time t , and R_{ijt} is an error term, assumed to be independent and normally distributed with a common variance of σ^2 . The parameters π_{0ij} and π_{1ij} represent, respectively, the initial status (referring to grade 7 in this case) and the rate of growth in mathematics achievement for student i in school j .

The second level of the model contains two between-student regression equations modeling the initial status and rate of growth based on student background covariates (student-level independent variables), including age, socioeconomic status (SES), the number of parents (numpar), and the number of siblings (numsib):

$$\begin{aligned}\pi_{0ij} &= \beta_{00j} + \beta_{10j}(\text{age})_{ij} + \beta_{20j}(\text{SES})_{ij} + \beta_{30j}(\text{numpar})_{ij} \\ &\quad + \beta_{40j}(\text{numsib})_{ij} + u_{0ij} \\ \pi_{1ij} &= \beta_{01j} + \beta_{11j}(\text{age})_{ij} + \beta_{21j}(\text{SES})_{ij} + \beta_{31j}(\text{numpar})_{ij} \\ &\quad + \beta_{41j}(\text{numsib})_{ij} + u_{1ij}\end{aligned}$$

where β 's are parameters for student-level covariates, representing the relationships of initial status and rate of growth to student-level covariates, and u_{0ij} and u_{1ij} are student-level error terms. If student-level covariates are centered around their grand means, then β_{00j} and β_{01j} represent the initial status and the rate of growth for what is often referred to as a typical student (with average age, SES, number of siblings, and who attends a school that has the same proportion of single parents as in the sample of students).

The third level of the model has two between-school equations, which regress the average initial status and the average rate of growth on school background covariates (school-level independent variables), including school size, school location represented by suburban (vs. urban) schools and rural (vs. urban) schools, and school mean (average) SES:

$$\begin{aligned}\beta_{00j} &= \phi_{000} + \phi_{001}(\text{size})_j + \phi_{002}(\text{suburban})_j + \phi_{003}(\text{rural})_j \\ &\quad + \phi_{004}(\text{meanSES})_j + v_{00j} \\ \beta_{01j} &= \phi_{010} + \phi_{011}(\text{size})_j + \phi_{012}(\text{suburban})_j + \phi_{013}(\text{rural})_j \\ &\quad + \phi_{014}(\text{meanSES})_j + v_{01j}\end{aligned}$$

In these equations, the average initial status and the average rate of growth are represented as an average value (ϕ_{000} or ϕ_{010} , i.e., grand mean), an error term (v_{00j} or v_{01j}), and the contribution of each school-level covariate.

Statistical analysis is often completed in two steps. First, a simple multilevel growth model is estimated without between-student and between-school covariates. In this simple model, the initial status and the rate of growth are described as an average value (fixed effect) plus a variation (random effect). This provides an opportunity to examine not only the average values of the initial status and the rate of growth but also their variances and covariances.

The second step of analysis introduces between-student and between-school covariates, establishing a complex multilevel growth model. The purpose is to use those covariates to explain variation between students within schools and between schools regarding both the initial status and the rate of growth. Adopted and modified from Ma (1999), **Table 1** presents the analytical results from the above (complex) multilevel growth model that estimates the rate of growth in mathematics achievement across secondary schooling and identifies variables that predict such a growth.

Survival Model

Unlike the multilevel growth model that typically deals with the trajectory of change, another category of longitudinal research and evaluation typically deals with the occurrence of an event of interest. Traditionally, researchers tend to ask whether an event occurs before the end of a certain period. Willett and Singer (1991) argued that, rather than asking whether, researchers should ask when an event occurs during a certain period:

Although logically intertwined, these two types of questions are also conceptually distinct, the “When?” question being far more general than the “Whether?” In fact, . . . by asking when events occur, a researcher learns not only

whether these events occur by each of several points in time but much much more. (p. 408)

To some extent, researchers’ asking ‘whether’ instead of ‘when’ is due to the fact that traditional educational statistics are inappropriate to address the ‘when’ question. Willett and Singer (1991) stated that traditional statistical techniques, such as regression analysis or analysis of variance, are not capable of handling censored event times:

No matter when data collection begins and no matter how long any subsequent follow-up, some study participants do not experience the target event while the researcher watches – some students do not drop out; some children do not leave day care; some teachers do not quit. These people have censored event times. What value of the outcome should they be assigned? Will they experience the event soon after the end of data collection, or will some of them never undergo the transition of interest? (p. 408)

Researchers cannot answer these questions on the basis of traditional statistical techniques. Survival analysis is the most effective statistical approach to work with censored data (Willett and Singer, 1991).

Unlike traditional statistical techniques for longitudinal data analysis, the survival model is capable of incorporating time-varying (independent) variables (i.e., variables that take on different values at different points in time). Willett and Singer (1991) argued that “traditional analytic methods offer few mechanisms for including predictors whose values vary over time or for permitting the effects of predictors to fluctuate over time” (p. 426). As a result, researchers tend to use “values corresponding to a single point in time, the average of values over time, or the rate of change in values over time” in their analyses (p. 426). In doing so, “traditional methods force researchers into building static models of dynamic processes; survival methods [on the other hand] allow researchers to model dynamic process dynamically” (p. 427).

Using data from LSAY, Ma and Willms (1999) developed a survival model to examine participation of American students in mathematics coursework across secondary schooling (i.e., dropping out of mathematics courses). As events, such as dropout, are dichotomous (e.g., 1 = drop and 0 = stay), the nature of data analysis is often logistic regression. There is special treatment for time-varying variables – “for each unit of time that each individual is known to be at risk, a separate observational record is created” (Allison, 1984: 18). Hazard probability (the proportion of persons available at the beginning of a certain period who experience the event during the period) is an important part of any survival model. It is often assumed that hazard probability varies across points in time (e.g., different across the 5 years in the study of dropping out of mathematics courses). To assume varying hazard probability, a set of dummy variables is created, one for each of the later points in

Table 1 Results of multilevel growth model explaining variation in mathematics achievement between students and schools in grade 7 status and rate of growth

	<i>Females</i>		<i>Males</i>	
	<i>Effect</i>	<i>(SE)</i>	<i>Effect</i>	<i>(SE)</i>
<i>Average within-student model</i>				
Grade 7 status	59.79*	(1.35)	59.89*	(1.39)
Rate of growth	3.75*	(0.49)	4.19*	(0.64)
<i>Effects of between-student covariates on grade 7 status</i>				
Age	−3.30*	(0.56)	−4.07*	(0.48)
Socioeconomic status (SES)	2.62*	(0.36)	3.06*	(0.35)
Number of parents	1.74*	(0.71)	2.29*	(0.78)
Number of siblings	0.09	(0.27)	0.37	(0.26)
<i>Effects of between-student covariates on rate of growth</i>				
Age	−0.64*	(0.21)	−0.80*	(0.20)
SES	0.56*	(0.13)	0.45*	(0.13)
Number of parents	0.50	(0.26)	0.45	(0.31)
Number of siblings	0.13	(0.10)	0.19	(0.10)
<i>Effects of between-school covariates on grade 7 status</i>				
School size	−0.01	(0.00)	−0.00	(0.00)
School mean SES	4.75*	(1.72)	6.56*	(1.65)
Suburban	1.38	(1.28)	−1.43	(1.32)
Rural	1.09	(1.30)	−0.68	(1.32)
<i>Effects of between-school covariates on rate of growth</i>				
School size	−0.00	(0.00)	−0.00	(0.00)
School mean SES	−0.33	(0.65)	0.42	(0.76)
Suburban	1.21*	(0.47)	0.34	(0.61)
Rural	1.17*	(0.47)	0.15	(0.61)

* $p < 0.05$.

time, with the first point in time as the reference (see Allison, 1984). In Ma and Willms (1999), one dummy variable was created for each of the last 5 years (grades 8–12), with grade 7 as the base for comparison. The probability of dropping out of mathematics courses is measured as:

$$P(u = 1) = \frac{1}{1 + e^{-Z}}$$

where $P(u = 1)$ denotes the probability that event u (dropping out of mathematics courses) occurs and Z is a linear composite of independent variables. In Ma and Willms (1999), the probability of dropping out of mathematics courses was predicted by students' gender, SES, prior achievement in mathematics (PAIM), and prior attitude toward mathematics (PATM); that is:

$$Z = \sum_{i=1}^5 \beta_i(G_j) + \beta_6(\text{gender}) + \beta_7(\text{SES}) + \beta_8(\text{PAIM})(t) + \beta_9(\text{PATM})(t)$$

where $j = 8, 9, 10, 11, 12$, and $\beta_i (i = 1, 2, 3, 4, 5)$ represents the time-varying hazard probability. For example, $\beta_1(G_8)$ represents the hazard probability in grade 8. Time-varying variables denoted by t may have different effects on the probability of dropping out of mathematics courses at different points in time, which can be modeled by interaction terms between grade-level dummies (G_8 to G_{12}) and independent (time-constant or time-varying) variables. For example, gender differences in dropping out of mathematics courses is expected to occur mainly in the last year of high school (grade 12).

$$Z = \sum_{i=1}^5 \beta_i(G_j) + \beta_6(\text{gender}) + \beta_7(\text{SES}) + \beta_8(\text{PAIM})(t) + \beta_9(\text{PATM})(t) + \beta_{10}(\text{gender} \times G_{12})$$

Adopted and modified from Ma and Willms (1999), **Table 2** presents the analytical results from the above survival model that estimates time-varying hazard probabilities and identifies student background variables that have constant effects (see SES and PAIM) or time-varying effects (see gender and PATM) on dropping out of mathematics courses.

Cross-Lagged Model

Cross-lagged panel design is an effective way to determine the cause–effect relationship between two variables of interest (see Marsh, 1990). Using data from LSAY, Ma and Xu (2004) developed a cross-lagged model via structural equation model (SEM) with latent variables to analyze longitudinal data on attitude toward mathematics (ATM) and achievement in mathematics (AIM) for the purpose of determining the causal ordering between ATM and AIM. The simplest cross-lagged model contains two waves of data, even though Willett *et al.* (1998) have argued against designs producing only before and after data. The cross-lagged model in Ma and Xu (2004) contains six waves of data (grades 7–12). The cross-lagged model via SEM (with latent variables) has a huge advantage of integrating a structural model with a measurement model. This is desirable in that structural relationships can be directly adjusted for measurement qualities. The development of the cross-lagged model is usually achieved in two stages.

In Ma and Xu (2004), the first stage includes specification of measurement models. There are three attitude items:

1. mathematics is useful in everyday problems,
2. mathematics helps a person think logically, and
3. I will use mathematics in many ways as an adult.

Table 2 Results of survival model explaining variation in mathematics participation from grade 8 to 12

Time effect	Effect	SE	Proportion
Grade 8 (D)	3.83*	(0.07)	0.98
Grade 9 (D)	2.55*	(0.05)	0.93
Grade 10 (D)	2.57*	(0.03)	0.93
Grade 11 (D)	2.15*	(0.10)	0.90
Grade 12 (D)	0.71*	(0.11)	0.67
Variable effect	Effect	SE	Exp
Female (D)	0.09*	(0.04)	1.09
Socioeconomic status (SES)	0.14*	(0.02)	1.15
Prior mathematics achievement (T)	0.27*	(0.03)	1.31
Prior attitude toward mathematics (T)	0.11*	(0.02)	1.12
Female \times grade 12	−0.45*	(0.13)	[0.70]
Prior attitude toward mathematics \times grade 11	0.27*	(0.06)	[1.46]
Prior attitude toward mathematics \times grade 12	0.32*	(0.06)	[1.54]

* $p < 0.05$.

Effects in parentheses are calculated by e raised to the power of the sum of interaction effect and base-line effect. For example, the effect of prior attitude toward mathematics in grade 11 is $\text{Exp}(0.27 + 0.11) = 1.46$. D denotes dummy variables; T denotes time-varying variables; SE denotes standard errors; Exp denotes the regression results in terms of e raised to the power of each effect; grades 8–12 (D) represent hazard probabilities in grades 8–12.

They are specified as indicators of the latent variable, ATM, in a measurement model (see **Figure 1**). Specified also in this measurement model are the stability effects of ATM from one year to the next across grades 7–12 (see unidirectional paths in **Figure 1**). Obviously, such a model specification yields estimates of the stability effects that are free from random measurement errors and the effects of a person's unique responses specific to each item over time.

AIM is measured in four mathematics subjects: (1) basic skills, (2) algebra, (3) geometry, and (4) quantitative literacy. The measurement model for AIM is developed in the same way as shown in **Figure 2**. The relationship between two non-neighboring latent variables is completely mediated by latent variable(s) in-between them in these models (**Figures 1 and 2**). These measurement models include sound common statistical practices when analyzing longitudinal data: (1) allowing measurement errors (not shown in the figure for simplicity) to correlate across different points in time for each observed indicator and (2) allowing measurement errors among various indicators to covary at each point in time.

Structural models can then be developed to connect with measurement models by adding cross-lagged effects. In Ma and Xu (2004), cross-lagged effects between ATM and

AIM are specified (see **Figure 3**). Such a cross-lagged model with all possible stability effects and cross-lagged effects, is an ideal approach to investigate the causal relationship between ATM and AIM. With all specifications about measurement errors as discussed in the measurement models maintaining the same, the latent variables of ATM and AIM can now be specified to correlate at each point in time. All of this specification process until now establishes a complete cross-lagged model. There are variations of this cross-lagged model for analytical purposes. A cross-lagged model without paths being specified as equal is unconstrained, often serving as the baseline model. This unconstrained model can then be compared with models that specify stationarity (invariance of factor loadings across measurement waves) and tau equivalence (invariance of factor loadings within measurement waves) to modify the unconstrained model for the causal relationship between ATM and AIM (see Pitts *et al.*, 1996). Model comparisons for goodness of fit can be achieved with common indices such as the chi-square index, the criterion fit index (CFI), the Tucker–Lewis index (TLI), and the standardized root mean square residual (SRMR). **Figure 3** summarizes the causal analysis reported in Ma and Xu (2004) that AIM demonstrates causal predominance over ATM across secondary schooling.

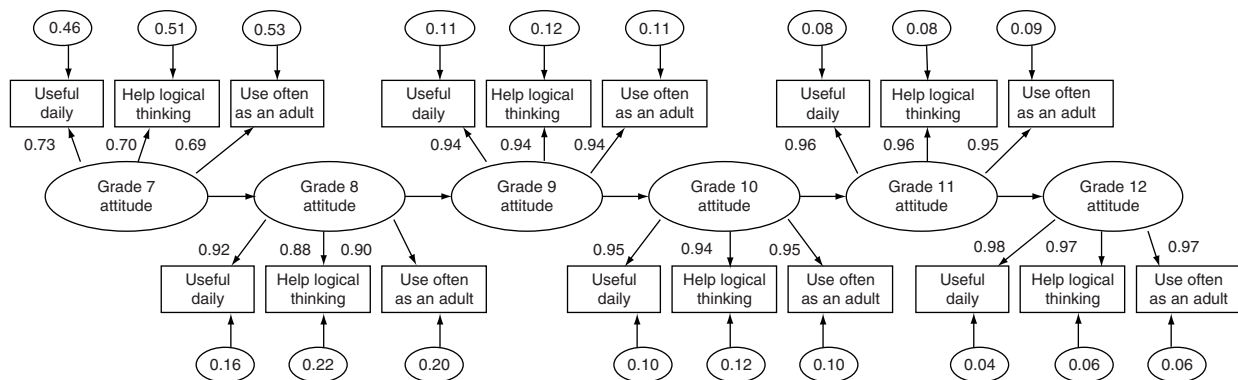


Figure 1 Measurement models of attitude toward mathematics across grades 7–12.

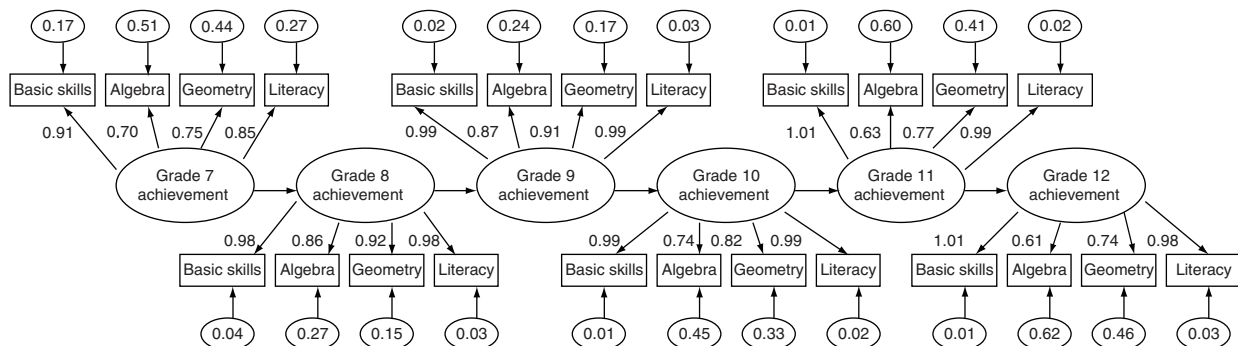


Figure 2 Measurement models of achievement in mathematics across grades 7–12.

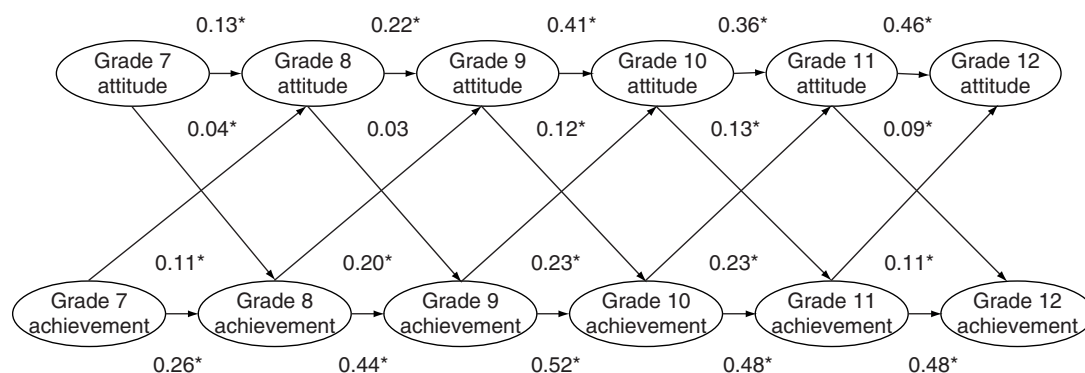


Figure 3 Results of cross-lagged model estimating the causal relationship between attitude toward mathematics and achievement in mathematics across grades 7–12. * $p < 0.05$. Large ovals represent latent factors and unidirectional arrows represent causal links. All parameter estimates for unidirectional paths are standardized.

See also: Evaluation Research; Generalized Linear Mixed Models; Growth Modeling; Hierarchical Linear Models; Latent Class Models; Mixed Methods; Survival Data Analysis.

Bibliography

- Allison, P. D. (1984). *Even History Analysis: Regression for Longitudinal Event Data*. Beverly Hills, CA: Sage.
- Hox, J. J. (2002). *Multilevel Analysis: Techniques and Applications*. Mahwah, NJ: Erlbaum.
- Ma, X. (1997). *A National Assessment of Mathematics Participation in the United States: A Survival Analysis Model for Describing Students' Academic Careers*. Lewiston, NY: Edwin Mellen.
- Ma, X. (1999). Gender differences on growth in mathematical skills during secondary grades: A growth model analysis. *Alberta Journal of Educational Research* **45**, 448–466.
- Ma, X. and Ma, L. (2007). The challenge of separating project effects on student achievement: The case of AMSP and ARSI. *Paper Presented at the Annual Meeting of the American Educational Research Association*, Chicago, IL.
- Ma, X. and Willms, J. D. (1999). Dropping out of advanced mathematics: How much do students and schools contribute to the problem? *Educational Evaluation and Policy Analysis* **21**, 365–383.
- Ma, X. and Xu, J. (2004). Determining the causal ordering between attitude toward mathematics and achievement in mathematics. *American Journal of Education* **110**, 256–280.
- Marsh, H. W. (1990). Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology* **82**, 646–656.
- Miller, J. D., Kimmel, L., Hoffer, T. B., and Nelson, C. (2000). *Longitudinal Study of American Youth: Users Manual*. Chicago, IL: International Center for the Advancement of Scientific Literacy, Northwestern University.
- Pitts, S. C., West, S. G., and Tein, J. (1996). Longitudinal measurement models in evaluation research: Examining stability and change. *Evaluation and Program Planning* **19**, 333–350.
- Rasbash, J., Browne, W., Goldstein, H., et al. (2000). *A User's Guide to MLwiN*. London: University of London.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models*, 2nd edn. Newbury Park, CA: Sage.
- Willett, J. B. (1988). Questions and answers in the measurement of change. In Rothkopf, E. Z. (ed.) *Review of Research in Education* pp 345–422. Washington, DC: American Educational Research Association.
- Willett, J. B. and Singer, J. D. (1991). From whether to when: New methods for studying student dropout and teacher attrition. *Review of Educational Research* **61**, 407–450.
- Willett, J. B., Singer, J. D., and Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychology in context: Statistical models and methodological recommendations. *Development and Psychology* **10**, 395–426.

Further Reading

- Bijleveld, C. C. J. H. and van der Kamp, L. J. (eds.) (1998). *Longitudinal Data Analysis: Designs, Models and Methods*. London: Sage.
- Singer, J. D. and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.
- Taris, T. W. (2000). *A Primer in Longitudinal Data Analysis*. London: Sage.

Relevant Websites

- <http://www.longitudinal.stir.ac.uk> – Longitudinal Data Analysis for Social Science Researchers.
- <http://tigger.uic.edu> – Longitudinal Data Analysis, University of Illinois at Chicago.

Meta-evaluation: Purpose, Prescription, and Practice

L A Wingate, Western Michigan University, Kalamazoo, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Evaluation client – The individual or group that requested and/or paid for an evaluation.

Evaluation consumer – Anyone who uses information produced by an evaluation.

Evaluation practitioner – An individual who provides evaluation services.

Evaluation stakeholder – Anyone who is affected by or has a vested interest in an evaluation.

Formative evaluation – The systematic assessment of an object to determine ways in which it can be improved as it is being developed or implemented.

Meta-analysis – The aggregation and analysis of the quantitative results from multiple studies that investigated the same phenomena.

Meta-evaluation – The formative or summative assessment of the merit and worth of an evaluation, group of evaluations, or evaluation system.

Summative evaluation – The systematic determination of the merit and worth of an object (e.g., program, policy).

The term metaevaluation was introduced by Michael Scriven in 1969. In defining it, he explained that metaevaluation has both a theoretical function and a practical one. The former, he said, is “the methodological assessment of the role of evaluation.” The latter “is concerned with the evaluation of specific evaluative performances” (Scriven, 1969: 36). A few years later, Scriven (1975) noted that metaevaluation may seem like an opaque neologism to some (p. 16). It may well have been an opaque neologism in 1975, but it is no longer. Metaevaluation has taken a firm hold in evaluation theory since then. But it is the latter function – that of evaluating particular evaluations – that has taken root, and metaevaluation is commonly understood as the evaluation of an evaluation. Metaevaluation may focus on a single evaluation, a set of evaluations, or an evaluation system. It is applicable to any type of evaluation, such as program, personnel, student, product, and policy. This article focuses on metaevaluation as applied to program evaluations.

Many books, presentations, and websites that provide a general orientation to evaluation start with the issue of why programs should be evaluated. Given that program stakeholders often need convincing about why they should embrace evaluation and expend resources on it,

it is likely that the value of metaevaluation is even less obvious to many. It is for that reason, then, that this article begins with an overview of the various purposes for which metaevaluation may be conducted and the benefits that may be reaped from it. Next, we discuss what has been prescribed by the evaluation profession’s strongest metaevaluation advocates with regard to how metaevaluation should be performed. Then, departing from theoretical and methodological considerations, we turn to issues regarding real-world metaevaluation practice.

Purpose

All the reasons that may be invoked for why programs should be evaluated also apply to why evaluations should be evaluated, such as to provide guidance for improvement, ensure that resources have been expended wisely, ensure that stakeholders have been properly engaged and respected, verify that outcomes were worthwhile and met needs, and learn about what works and does not in particular evaluation contexts.

The range of purposes that metaevaluation serves may be grouped into four categories: formative evaluation, summative evaluation, synthesis, and research. These purposes are not mutually exclusive, and a single metaevaluation may serve multiple purposes. We discuss them independently not because there are unique approaches or criteria associated with them, but to better illuminate the various purposes for and uses of metaevaluation. This article focuses on formal metaevaluation performed by another evaluator, whether internal or external to the primary evaluation. However, evaluations are frequently subject to scrutiny and assessment through means that are not officially labeled as metaevaluation. Such means include review and commentary by project advisory panels, clients, school boards, and organizational boards of directors. Evaluators may also self-metaevaluate by reflecting on their practice in terms of how well it meets the standards and guidelines set forth by the profession. These activities serve similar functions, albeit in a less overt and systematic way.

Formative Evaluation

Formative metaevaluation takes place while an evaluation is underway in order to provide guidance for improvement. Draft evaluation work products, such as evaluation plans, instruments, protocols, and reports, may be subject to review

and feedback by a metaevaluator prior their finalization, implementation, or dissemination. Less tangible aspects of an evaluation may also be assessed in formative metaevaluation, such as data collection, verification, and analysis procedures; competence of the evaluation personnel; and responsiveness to stakeholders' information needs.

The same arguments that are made for formative evaluation of programs can be made for formative metaevaluation. For example, it helps ensure that wise choices are made with regard to how to focus the study, expend resources, perform analyses, and so on. Another important reason that formative metaevaluations are conducted is to reduce bias in evaluation (Fitzpatrick *et al.*, 2004; Scriven, 1975). The mere presence of a metaevaluator has the potential to reduce bias and increase rigor as an evaluation is being implemented (Scriven, 1975), simply because the evaluator knows his or her performance is being scrutinized, not just by the evaluation client, but by someone else with evaluation expertise.

The greatest benefit of formative metaevaluation is that it can reveal deficiencies in the primary evaluation at a point in time when they can still be addressed, thus preventing the determination and dissemination of invalid conclusions and increasing the primary evaluation's utility and cost-effectiveness (Fitzpatrick *et al.*, 2004; Joint Committee on Standards for Educational Evaluation, 1994; Stufflebeam and Shinkfield, 2007).

Summative Evaluation

A summative metaevaluation assesses the quality of a completed evaluation – the appropriateness of its processes and validity of its conclusions. A summative metaevaluation may examine, for example, the extent to which evaluators delivered what was promised, used resources efficiently, and acted ethically. In addition to informing such accountability issues, a summative metaevaluation may serve a verification purpose, that is to confirm (or disconfirm) that an evaluation's conclusions (and recommendations, if present) were sound and justified. Summative metaevaluation that checks the veracity of findings and appropriateness of the methods is especially important for evaluations where stakes are high and important decisions hinge on the results (Patton, 2008; Stufflebeam and Shinkfield, 2007).

When a summative metaevaluation validates a primary evaluation, it adds credibility to it and enhances users' confidence in using the evaluation findings to inform decisions to expand, modify, or cancel programs. When a summative metaevaluation finds serious flaws in a primary evaluation, it can prevent decision makers from taking actions based on faulty information. Evaluation consumers, clients, and practitioners can use summative metaevaluations in a formative manner by taking what was learned in the summative metaevaluation and applying those lessons in their subsequent evaluation experiences.

Synthesis

Metaevaluation may also be performed as a precursor to synthesizing findings across evaluations of a particular program or set of programs (Cooksy and Caracelli, 2005; Scott-Little *et al.*, 2002; Stufflebeam and Shinkfield, 2007). Such syntheses are undertaken to achieve a deeper understanding of the effectiveness of a certain type of intervention or program. In this case, metaevaluation is a means to an end in that it is a tool for determining which evaluations are of sufficient quality to be included in a synthesis of findings to address research questions that are about program – not evaluation – quality or effectiveness. Metaevaluation is a critical step in such syntheses, whether qualitative or quantitative (i.e., meta-analytic), so that methodologically weak evaluations (Cooksy and Caracelli, 2005) or those with incorrect conclusions may be appropriately excluded.

Research

Metaevaluation is a useful vehicle for conducting research on evaluation (Fitzpatrick *et al.*, 2004; Stufflebeam and Shinkfield, 2007). Here, the focus is on addressing questions about the practice and/or use of evaluation. Stufflebeam and Shinkfield (2007) advocate for more empirical research on evaluation that will facilitate the development of “research-based principles for conducting program evaluation” (p. 73). The evaluation theories that prevail today are mostly prescriptions for practice and have not been systematically examined in empirical studies. Stufflebeam and Shinkfield identify numerous hypotheses about evaluation that have been proffered in the literature and could be subject to empirical investigation, and several of these – such as the ones listed here – could be addressed through metaevaluation:

- Appropriate application of evaluation and standards and principles enhances an evaluation's quality and resolution of ethical problems;
- Stakeholder involvement enhances use of evaluation findings;
- Participatory and collaborative approaches used for capacity building enhance program effectiveness and increase evaluation use; and
- Effectiveness of employment of a variety of reporting strategies beyond the written report, applied differentially to audiences, increases their use of findings (pp. 71–73).

In order to test these hypotheses, researchers must be able to discern evaluation quality, effectiveness, usefulness, ethicality, cost-effectiveness, etc. Making judgments about these aspects of evaluations is the domain of metaevaluation.

Prescription

Evaluation theories are essentially prescriptive theories – recommendations about how to go about the business of evaluation – that are informed by both scholarship and practice. Such theories address the focus and function of evaluation, role of evaluation stakeholders, evaluation design and implementation, sources of criteria for determining quality, and use of evaluation results. Although almost any theoretical orientation to or model for evaluation – for example, CIPP, utilization-focused, deliberative-democratic, empowerment – could be applied in metaevaluation, those who advocate for evaluation at the meta level typically do not prescribe the use of a particular model. (Metaevaluation is sometimes confused with meta-analysis, but they are very different endeavors, with meta-analysis being the aggregation and analysis of the quantitative results from multiple studies that investigated the same phenomena. However, a legitimate focus of a metaevaluation may be the extent to which an evaluation adhered to a particular evaluation model, in which case issues related to evaluation models would come into play, but not so much with regard to how the metaevaluation itself is implemented.) Rather, they tend to focus on the central issues of questions and criteria, which illuminate the issues on which metaevaluation should focus, and methodological approaches, that is, how metaevaluations should be implemented.

Questions and Criteria

By nature, metaevaluation focuses on questions about an evaluation's merit and worth. Stufflebeam and Shinkfield (2007) articulate the fundamental question of merit as "how well [the evaluation] meets the requirements of a sound evaluation" and the fundamental question of worth as "the extent to which it meets the audiences' needs for evaluative information" (p. 667). Patton (2008) suggests some basic metaevaluation questions that focus on evaluation processes: "To what extent was the [evaluation] implemented as designed? What adaptations occurred during implementation and what are the implications of those adaptations for intended and unintended outcomes?" (p. 570). The Joint Committee on Standards for Educational Evaluation (1994) identifies the following questions among those that might be addressed in a metaevaluation:

- Were the findings for the program evaluation clear?
- Were the design and implementation of the design appropriate for the study?
- Were the conclusions justified by the data?
- Were the results adequately disseminated? Were they used? (Joint Committee on Standards for Educational Evaluation, 1994: 186)

Questions define the scope of inquiry in a metaevaluation and criteria provide the basis for arriving at judgments in relation to those questions. As in any type of evaluation, the criteria and questions that drive a metaevaluation should be negotiated in advance with the evaluation client and other stakeholders to ensure their information needs are met.

However, metaevaluators should strive to ground their assessments of evaluations in commonly accepted criteria or standards for evaluation. This is important not only for individual metaevaluations, but also for advancing the relatively young evaluation profession by invoking and promoting a common understanding of what constitutes sound evaluation practice – not just among evaluation practitioners and scholars, but also evaluation participants, clients, consumers, and managers. Unlike well-established professions such as accounting, medicine, law, psychology, and the construction trades (to name just a few), there are no regulated certification or licensure processes to verify that individuals are qualified to practice evaluation or call themselves evaluators (although the Canadian Evaluation Society's Professional Designations Project, which started in 2007, is working toward that end –). In the absence of such regulations, it is incumbent on evaluation practitioners, clients, managers, and consumers to educate themselves about and use established criteria and standards to facilitate consistency and professionalization in evaluation practice. Some of evaluation's most distinguished scholars and practitioners, including Patton (2008), Scriven (2007), and Stufflebeam (2001) identify *The Program Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 1994) as one of the most important sources of criteria for assessing evaluation quality, if not the most. Grounding metaevaluation in the Standards promotes consistency in practice and diminishes the potential for evaluations to be judged on idiosyncratic criteria that do little to advance the profession or educate evaluation consumers about what constitutes sound evaluation. Judgments about the quality of evaluations should not hinge on methodological or theoretical preferences, personal opinion, or arbitrary criteria, but on recognized requirements for sound and credible evaluation, such as *The Program Evaluation Standards*, which are summarized in **Table 1**.

The subtitle of *The Program Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 1994) book, *How to Assess Evaluations of Educational Programs*, reveals why these standards figure so prominently in discussions of metaevaluation. But they were developed specifically for application to educational evaluations in North America. As someone entrenched in educational program evaluation in the American context, it follows that these are the standards that dominate the author's own evaluation practice and scholarship. However, other standards may be more appropriate – either in

Table 1 The Program Evaluation Standards (Joint Committee 1994)

<i>Utility standards : The utility standards are intended to ensure that an evaluation will serve the information needs of intended users.</i>	
<i>U1 stakeholder identification</i>	Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed.
<i>U2 evaluator credibility</i>	The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance.
<i>U3 information scope and selection</i>	Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders.
<i>U4 values identification</i>	The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.
<i>U5 report clarity</i>	Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.
<i>U6 report timeliness and dissemination</i>	Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.
<i>U7 evaluation impact</i>	Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.
<i>Feasibility standards: The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.</i>	
<i>F1 practical procedures</i>	The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained.
<i>F2 political viability</i>	The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations, or to bias or misapply the results can be averted or counteracted.
<i>F3 cost-effectiveness</i>	The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified.
<i>Propriety standards: The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.</i>	
<i>P1 service orientation</i>	Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.
<i>P2 formal agreements</i>	Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.
<i>P3 rights of human subjects</i>	Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects.
<i>P4 human interactions</i>	Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed.
<i>P5 complete and fair assessment</i>	The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed.
<i>P6 disclosure of findings</i>	The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation and any others with expressed legal rights to receive the results.
<i>P7 conflict of interest</i>	Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.
<i>P8 fiscal responsibility</i>	The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate.
<i>Accuracy standards: The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.</i>	
<i>A1 program documentation</i>	The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified.
<i>A2 context analysis</i>	The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified.
<i>A3 described purposes and procedures</i>	The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.
<i>A4 defensible information sources</i>	The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.
<i>A5 valid information</i>	The information-gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.
<i>A6 reliable information</i>	The information-gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use.

Continued

Table 1 Continued

<i>A7 systematic information</i>	The information collected, processed, and reported in an evaluation should be systematically reviewed, and any errors found should be corrected.
<i>A8 analysis of quantitative information</i>	Quantitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.
<i>A9 analysis of qualitative information</i>	Qualitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.
<i>A10 justified conclusions</i>	The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.
<i>A11 impartial reporting</i>	Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.
<i>A12 metaevaluation</i>	The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses.

Note: In *The Program Evaluation Standards* book, each standard is supported by an overview, guidelines, common errors, illustrative cases, and references to supporting documentation.
Reprinted from the Joint Committee on Standards for Educational Evaluation.

conjunction with or in lieu of *The Program Evaluation Standards* – depending on the subject matter and context. In addition to standards for program evaluations, the Joint Committee has issued standards for educational personnel evaluation (Joint Committee on Standards for Educational Evaluation, 2008) and student evaluation (Joint Committee on Standards for Educational Evaluation, 2002). The United States Government Accountability Office's (2007) *Government Auditing Standards* include standards for evaluations (referred to as performance audits in that document) conducted by and for U.S. federal agencies. The American Evaluation Association's *Guiding Principles for Evaluators* provide general guidance for how professional evaluators should conduct themselves. Some organizations have developed their own criteria for assessing evaluations. For example, ALNAP, the Active Learning Network for Accountability and Performance in Humanitarian Action, has developed a tool for rating evaluations of humanitarian interventions, which is available from its website. In this volume's article on moral and ethical issues in evaluation, Michael Morris directs readers to standards and principles for use in an array of international contexts and provides the website addresses for many of these. Evaluators, metaevaluators, and evaluation clients, consumers, and other stakeholders should make themselves aware of the standards that are most pertinent for their contexts and apply them accordingly.

The increasing attention to the development and use of standards for professional evaluation in these diverse contexts is an indicator of the maturation of the evaluation profession. Use of such established and widely respected standards for metaevaluation purposes further promotes indoctrination of practitioners and scholars from diverse backgrounds into the global community of professional evaluators.

Individuals who request a metaevaluation may have special interests or concerns with regard to a particular evaluation, and their information needs should be respected

and addressed as the criteria and questions for a metaevaluation are negotiated.

Methodological Approaches

Stufflebeam has prescribed a practical, generic approach to metaevaluation that involves eleven tasks, which parallel the tasks in primary evaluation (Stufflebeam, 2001; Stufflebeam and Shinkfield, 2007):

1. Staff the metaevaluation team with one or more qualified evaluators.
2. Identify and arrange to interact with the metaevaluation's stakeholders.
3. Define the metaevaluation questions.
4. Agree on standards, principles, or criteria to judge the evaluation system or particular evaluation.
5. Issue a memo of understanding or negotiate a formal metaevaluation contract.
6. Collect and review pertinent available information.
7. Collect new information as needed.
8. Analyze and synthesize the findings.
9. Judge the evaluation's adherence to appropriate standards, principles, and/or criteria.
10. Convey the metaevaluation findings through reports, correspondence, oral presentations, and other ways.
11. As appropriate, help the client and other stakeholders interpret and apply the findings. (Stufflebeam and Shinkfield, 2007: 663)

The options for data collection methods and analytic techniques for metaevaluation include all of those that are used in primary evaluation. Metaevaluations, whether formative or summative, should almost always examine evaluation documents, such as plans, meeting minutes, instruments, protocols, and reports – that is, pertinent available information, noted in Task 6 above. Such documents, however, cannot tell the full story about an

evaluation, which is why Task 7, which calls for the acquisition of additional information, is often warranted. Additional information may be collected through a variety of means, such as interviews, focus groups, surveys, observations, and advocate–adversary techniques. Methodological choices should be keyed to the metaevaluation questions and criteria and, just as importantly, must consider the scope of the metaevaluation.

Cook and Gruder's (1978) typology of metaevaluation approaches provides a different way to conceptualize how to approach the task of metaevaluation. For them, key considerations are whether the metaevaluation takes place after the primary evaluation or concurrently, whether data collected for the primary evaluation are manipulated, and whether there are one or multiple data sets. They organize the possible options into seven basic approaches for metaevaluation:

1. Essay review of an evaluation report;
2. Review of the literature about a specific program;
3. Empirical reevaluation of an evaluation or program;
4. Empirical reevaluation of multiple data sets about the same program;
5. Consultant metaevaluation;
6. Simultaneous secondary analysis of raw data; and
7. Multiple independent replications (p. 17).

Cook and Gruder's (1978) metaevaluation typology was an influential contribution to early conceptualizations of metaevaluation and a must-read for anyone interested in the topic, but is somewhat dated in light of developments in evaluation that have occurred over the past 30 years. For example, since it was published prior to the issuance of the first edition of *The Program Evaluation Standards* in 1981, it lacks references to standards for judging evaluations. Although the formative/summative functions of metaevaluation are implicit in the dichotomy of whether metaevaluations are conducted subsequent to or simultaneous with primary evaluations, the formative/summative terminology, which has become so entrenched in how evaluators think and talk about evaluation, is absent. Nonetheless, the range of options suggested by Cook and Gruder is broader than what is offered by more contemporary writings on metaevaluation, and persons interested or engaged in metaevaluation would be well served by revisiting this seminal work.

Scriven (2005, 2007) suggests three options that overlap Cook and Gruder's (1978) typology substantially:

- Replicate evaluation using same methodology and compare results;
- Evaluate same evaluand using different methodology and compare results; and
- Apply Scriven's (2007) Key Evaluation Checklist, *The Program Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 1994; Stufflebeam,

1999), or United States Government Accountability Office (2007) standards to the evaluation.

The first two approaches above and their correlates in Cook and Gruder's typology are applicable to summative metaevaluations that check on the accuracy of a primary evaluation's findings. Scriven (1975) recommends a procedure for reaching convergence when independent evaluations are conducted of the same evaluation, whereby the involved evaluators critique each other's reports, then modifies their own based on the feedback. This process is critical for achieving metaevaluative objectives – otherwise the client just has two evaluations of the same program, which may or not agree. The application of standards or criteria, indicated by the last option above, can address accuracy, but also facilitates a focus on evaluation processes.

Guidance is available for how to apply the criteria embodied in *The Program Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 1994) to reach judgments about evaluation quality. The Standards book includes a chapter on applying the standards, which recommends that metaevaluators rate an evaluation on each standard in terms of whether the standard was addressed, partially addressed, not addressed, or not applicable. Fitzpatrick *et al.*, (2004) illustrate how a checklist could be developed for evaluating specific aspects of an evaluation, with specific subcriteria to support each standard, and they provide an example of a checklist designed to assess evaluation reports and designs. Stufflebeam (1999) designed an elaborate checklist for evaluating evaluations based on *The Program Evaluation Standards*. His checklist includes subcriteria for each standard and a scoring system to facilitate interpretations of quality. Use of this checklist is also suggested by other prominent advocates of metaevaluation (e.g., Fitzpatrick *et al.*, 2004; Patton, 2008; Scriven, 2007).

Regardless of the criteria, approach, or tools one uses in metaevaluation, it is almost a necessity that there be a connoisseurial (Eisner, 1983) element to it, or what Datta (1999) calls the “I know it when I see it” (p. 346) approach to metaevaluation. Rigorous and extensive data collection is not typically practical in the metaevaluation context, and in the absence of such data, metaevaluators must bring to bear their expert knowledge in evaluation, in combination with what they learn about the particular evaluation through additional means. Primary evaluators can facilitate metaevaluation by keeping thorough records of evaluation processes and decisions, thereby minimizing, albeit not eliminating, the need for additional data collection.

Practice

It is difficult to determine the frequency with which metaevaluation is actually practiced. An evaluation that

the author has been involved with at The Evaluation Center at Western Michigan University has been subject to five separate metaevaluations over the course of 9 years. But this may be an exception and not the rule in evaluation practice, due largely to the lead evaluator's deep commitment to the tenets of The Program Evaluation Standards.

In a recent search of the peer-reviewed literature, the author located 18 examples of metaevaluations. Concurrently and independently of my effort, Cooksy and Caracelli (2009) conducted a search for examples of metaevaluation and also located 18. Cooksy and Caracelli did not limit their search to the peer-reviewed literature, but did restrict it to metaevaluations of single evaluations for formative or summative purposes (excluding metaevaluations of multiple evaluations and those conducted for theory development purposes). Collectively, our two searches yielded just 34 examples of formal metaevaluation (two of these metaevaluations were identified by both searches), suggesting that only a tiny proportion of the evaluations that are conducted are subject to systematic metaevaluation. Although this conclusion may seem to have been reached by rather unsystematic means, it is common perception (Fitzpatrick *et al.*, 2004; Henry and Mark, 2003). It is also noteworthy that 17 of the 18 metaevaluations located by Cooksy and Caracelli were produced by just 3 entities – the Center for Instructional Research and Curriculum Evaluation at the University of Illinois, The Evaluation Center at Western Michigan University, and Cornell University's graduate program in evaluation. There are tentative indications, however, that the practice is may be growing in frequency – of the 34 metaevaluation examples located by both searches, 15 of them were issued in 2000 or later; 10 were published in the 1990s, and the remaining 9 were put forth in the 1980s. Neither search specified any parameters regarding years of publication.

Although formal metaevaluation may be practiced on a relatively limited basis, informal metaevaluation happens all the time. For example, evaluations are often subject to review and feedback by project advisory panels, clients, and organizational boards of directors. Potential clients and competitors may also be paying attention to quality of an evaluator's performance. Any evaluator who uses The Program Evaluation Standards or other evaluation guidelines or principles to keep his or her work on track and/or to inform decisions about how to conduct an evaluation is engaged in metaevaluation.

Examples of metaevaluation in the literature are relatively rare, but Stufflebeam and Shinkfield (2007) and Cook and Gruder (1978) describe several cases of metaevaluation, with clear descriptions of what various approaches to metaevaluation look like in practice and insights that only such seasoned evaluators and metaevaluators can provide. Based on the author's own experience as an evaluator who has conducted internal and external

metaevaluations and who has had her own work subject to metaevaluation, she has identified some evaluation practice issues that are specific to metaevaluation. The author cautions readers that these observations are based on her own experience, although many are substantiated in the literature on metaevaluation, and likely do not represent the full range of issues that are uniquely metaevaluative.

The Program Evaluation Standards were previously discussed in terms of their use as criteria in assessing evaluation quality. However, they are also a guide to evaluation practice. In thinking about the Standards as a guide to practice, it is the author's opinion that some of them manifest slightly differently in metaevaluation than in primary evaluation. These standards and their special considerations in the context of metaevaluation are highlighted as follows: (refer to **Table 1** for the standards' summary statements).

Evaluator Credibility

This utility standard is concerned with the evaluators' trustworthiness and competence. In any evaluation context, it is crucial that stakeholders have confidence in evaluators so that they will trust and use the evaluation findings. As metaevaluation is an unfamiliar concept outside of the evaluation world, it is possible that clients and other primary evaluation stakeholders may think that the presence of a third party checking on the work of the primary evaluator means they should be concerned about his or her competence. Metaevaluators and those who engage them should take care to protect the credibility of primary evaluators. As evaluation consumers and practitioners alike become more knowledgeable of the guidelines for professional evaluation practice, they should come to understand that metaevaluation is part and parcel of sound evaluation practice and is not intended to detract from or discredit a primary evaluation, but rather to enhance its utility and/or credibility.

Values Identification

This utility standard states that evaluators should be upfront about the values they use to interpret findings so that bases for judgments are clear. As noted repeatedly throughout this article, judgments about evaluation quality should be grounded in recognized and widely accepted criteria for sound and credible evaluation. In addition to contributing to the professionalization of evaluation, this focus on common standards also aids in educating evaluation consumers about what they can and should expect from evaluation, providing the metaevaluation results are disseminated (see section titled 'Disclosure of findings' below).

Practical Procedures

This feasibility standard advises evaluators to keep disruption to a minimum (Joint Committee on Standards for Educational Evaluation, 1994: 65) during data collection. Metaevaluators need to be cognizant that primary evaluation stakeholders may have already been called upon to give time or other resources to the primary evaluation. Calling upon them again to provide information for the metaevaluation may contribute to an unfavorable opinion of the evaluation enterprise, especially if they did not value or obtain benefit from their contribution to the primary evaluation. An overzealous metaevaluation effort could also threaten the primary evaluation by burdening the primary evaluators with time-consuming requests (Cook and Gruder, 1978). Cook and Gruder (1978) advise metaevaluators to “be very sensitive to the trade-offs between the needs of busy evaluation staffs for helpful and supportive feedback and the... requirements to monitor quality and discover problems before they become salient” (p. 37).

In addition, metaevaluation budgets are typically fairly small. Stufflebeam and Shinkfield (2007) have observed that is common for metaevaluation budgets to be less than 2% of the primary evaluation’s budget, which means it is especially important to ensure that procedures are realistic.

Political Viability

Evaluation is often a politically charged endeavor. The feasibility standard on political viability calls on evaluators to anticipate the different and possibly competing interests of evaluation stakeholders to avoid derailment of the evaluation process, bias, or misuse of results. Metaevaluation efforts that require the acquisition of anything other than publicly available documents must achieve buy-in from the affected/involved parties. For example, when the focus of a metaevaluation is on evaluations across multiple projects, metaevaluators are dependent on the cooperation of project directors, who may be reticent to relinquish evaluation reports, especially if they were intended as internal project documents, to a third party for review. Metaevaluators should expect to spend a fair amount of time explaining the nature of the work and ensuring the confidentiality of the information provided.

In addition, as much as evaluators may believe in practicing what they preach, they are not accustomed to having their work evaluated and may be reluctant to cooperate with external metaevaluators. As Fitzpatrick *et al.*, (2004) have astutely observed, “Evaluators are human and are no more ecstatic about having their work evaluated than are professionals in other areas of endeavor” (p. 455).

Cost Effectiveness

The feasibility standard on cost-effectiveness states that evaluators should use resources efficiently and produce

valuable information so that evaluation costs are justified. To primary evaluation clients, metaevaluation may seem redundant with the initial evaluation and therefore not worth the expense – either to them or to their evaluators. Care must be taken to explain the value of metaevaluation to primary evaluation clients. Moreover, metaevaluators should be frugal in their efforts to ensure that their work brings value to the evaluation endeavor. This has implications for the selection of methods to perform the metaevaluation. Also important to consider are the stakes involved. High-profile, controversial evaluations warrant greater expense and rigor than internal metaevaluations conducted primarily for improvement purposes (Patton, 2008; Stufflebeam and Shinkfield, 2007).

Human Interaction

The propriety standard on human interaction is about respecting the human dignity and worth (Joint Committee on Standards for Educational Evaluation, 1994: 99) of the persons involved in an evaluation. “A *sin qua non* of metaevaluation,” said Cook and Gruder (1978: 36), “is tact.” In their discussion of consultant metaevaluation (what is essentially now considered formative evaluation), they emphasize the importance of a collegial relationship between the metaevaluator and primary evaluator; without a collaborative and cooperative working relationship, the metaevaluator will find it difficult to do his or her job. Of course, one has to balance such collaborative arrangements with potential threats to credibility and objectivity. Indeed, nonformative metaevaluations would not require any interaction at all between the primary evaluator and metaevaluator.

Disclosure of Findings

This propriety standard advises evaluators to ensure that evaluation findings are disseminated to the individuals impacted by the evaluation. It is especially important for evaluators whose work has been evaluated to share metaevaluation results to demonstrate they practice what they preach, that they are willing to expose their work to criticism and are not shy about sharing them with others. As suggested throughout this article, the practice of metaevaluation can contribute to the advancement of the evaluation discipline, but for this to happen, metaevaluation the findings need to be made available to others who can benefit and learn from them. Broader dissemination of metaevaluations serves to educate both evaluation practitioners and consumers about what should be emulated and what should be avoided in evaluation practice.

Conflict of Interest

The point of this propriety standard is not necessarily to create conditions in which there are no conflicts of

interest present in an evaluation, but instead to deal with such conflicts openly and honestly so that the evaluation's validity and credibility are not jeopardized. Since the community of professional evaluators is fairly small, evaluators and their clients must take care to engage critical friends instead of friendly critics as metaevaluators. Thinking about this issue in opposite terms, it is possible that unscrupulous evaluators could use metaevaluation as an opportunity to discredit the primary evaluator in order to obtain future evaluation contracts for themselves. In short, evaluators evaluating other evaluators create situations in which different types of conflicts of interest exists than in typical program evaluation contexts.

Metaevaluation. The final standard in The Program Evaluation Standards is the Accuracy standard on metaevaluation. It calls on evaluators to subject their work to systematic formative and summative evaluation. This is the one standard that does not transfer well to metaevaluation practice. That is, metaevaluations usually are not subject to metaevaluation – one must consider the point of diminishing returns. That said, Scriven (2009) recommends that metaevaluators have their work independently critiqued by another evaluator. His suggestion that the compensation for this work would be a good dinner implies that meta-metaevaluation would not involve much more than a careful review of metaevaluation plans or reports and/or an in-depth conversation about the process.

Conclusion

Metaevaluation serves several important functions for a wide range of evaluation stakeholders. Formative metaevaluation helps evaluators ensure their services are relevant and technically sound. Summative metaevaluation adds credibility to primary evaluations and helps evaluation audiences gauge how much confidence they should place in the findings. Metaevaluation performed for synthesis purposes ensures that only evaluations that are methodologically sound will be used to extend knowledge about the effectiveness of a given type of intervention, which has implications for policymakers. Metaevaluation as a vehicle for research on evaluation aids in the evaluation discipline's self-knowledge and in the development of more sophisticated models and tools for evaluation. Independently, each type of metaevaluation represents what Scriven (1969) referred to as the practical role of metaevaluation – to evaluate “specific evaluative performances” (p. 36). Collectively, these metaevaluation efforts contribute to our knowledge about what constitutes sound evaluation practice, where evaluation discipline falls short of what it strives to accomplish, evaluation's role in contributing to the betterment of society, and how to advance the discipline. This is what Scriven described as the theoretical function of metaevaluation, which goes beyond

looking at particular evaluations to “the evaluation of the very function and practice of evaluation itself” (Fitzpatrick *et al.*, 2004: 443).

The third edition of The Program Evaluation Standards, forthcoming in 2010, will give metaevaluation a higher profile than in previous editions. A fifth domain of standards called evaluation accountability will join utility, feasibility, propriety, and accuracy, and encompass matters pertaining to metaevaluation. The increased prominence of metaevaluation in The Program Evaluation Standards is consistent with the growing interest and attention to metaevaluation within the profession.

See also: Defining Quality in Evaluation; The Purpose of Educational Evaluation.

Bibliography

- Cook, T. D. and Gruder, C. L. (1978). Metaevaluation research. *Evaluation Review* 2(1), 5–51.
- Cooksy, L. J. and Caracelli, V. J. (2005). Quality, context, and use: Issues in achieving the goals of metaevaluation. *American Journal of Evaluation* 26(1), 31–42.
- Cooksy, L. J. and Caracelli, V. J. (2009). Metaevaluation practice. *Journal of MultiDisciplinary Evaluation* 6(11), 1–15.
- Datta, L. (1999). CIRCE's demonstration of a close-to-ideal evaluation in a less-than-ideal world. *American Journal of Evaluation* 20(2), 345–354.
- Eisner, E. W. (1983). Educational connoisseurship and criticism: Their form and functions in educational evaluation. In Madaus, G. F., Scriven, M., and Stufflebeam, D. L. (eds.) *Evaluation Models*, pp 87–102. Boston, MA: Kluwer-Nijhoff.
- Fitzpatrick, J. L., Sanders, J. R., and Worthen, B. R. (2004). *Program Evaluation: Alternative Approaches and Practical Guidelines*, 3rd edn. Boston, MA: Pearson.
- Henry, G. T. and Mark, M. M. (2003). Toward an agenda for research on evaluation. In Christie, C. (ed.) *The Practice–Theory Relationship in Evaluation*, pp 69–80. (New Directions for Evaluation, 97). San Francisco, CA: Jossey-Bass.
- Joint Committee on Standards for Educational Evaluation (1994). *The Program Evaluation Standards*, 2nd edn. Thousand Oaks, CA: Sage.
- Joint Committee on Standards for Educational Evaluation (2002). *The Student Evaluation Standards*. Thousand Oaks, CA: Sage.
- Joint Committee on Standards for Educational Evaluation (2008). *The Personnel Evaluation Standards*, 2nd edn. Thousand Oaks, CA: Sage.
- Patton, M. Q. (2008). *Utilization-Focused Evaluation*, 4th edn. Los Angeles, CA: Sage.
- Scott-Little, C., Hamann, M. S., and Jurs, S. G. (2002). Evaluations of after-school of programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation* 23(4), 387–419.
- Scriven, M. (1969). An introduction to meta-evaluation. *Educational Products Report* 2, 36–38.
- Scriven, M. (1975). Evaluation bias and its control. *Occasional Paper Series 4*. Kalamazoo: The Evaluation Center, Western Michigan University. <http://www.wmich.edu/evalctr/pubs/ops/ops04.pdf> (accessed August 2009).
- Scriven, M. (2006). Metaevaluation. In Mathison, S. (ed.) *Encyclopedia of Evaluation*, pp 249–250. Thousand Oaks, CA: Sage.
- Scriven, M. (2007). *Key Evaluation Checklist*. <http://www.wmich.edu/evalctr/checklists> (accessed August 2009).
- Scriven, M. (2009). Meta-evaluation revisited. *Journal of MultiDisciplinary Evaluation* 6(11), iii–viii.
- Stufflebeam, D. L. (1999). *The Program Evaluations Metaevaluation Checklist*. <http://www.wmich.edu/evalctr/checklists> (accessed August 2009).

- Stufflebeam, D. L. (2001). The metaevaluation imperative. *American Journal of Evaluation* **22**(2), 183–209.
- Stufflebeam, D. L. and Shinkfield, A. J. (2007). *Evaluation Theory, Models, and Applications*. San Francisco, CA: Jossey-Bass.
- United States Government Accountability Office (2007). *The Government Auditing Standards*. . <http://www.gao.gov/govaud/ybk01.htm> (accessed August 2009).

Further Reading

- Chatterji, M. (2005). Applying the Joint Committee's 1994 standards in international contexts: A case study of education evaluations in Bangladesh. *Teachers College Record* **107**(1), 2372–2400.
- Dowell, K., Haley, J., and Doyno-Ingersoll, J. A. (2006). Evaluating the evaluator: Development, field testing, and implications of a client-based method for assessing evaluator performance. *New Directions for Evaluation* **111**, 95–108.
- Lynch, D. C., Greer, A. G., Larson, L. C., et al. (2003). Descriptive metaevaluation: Case study of an interdisciplinary curriculum. *Evaluation and the Health Professions* **26**, 447–461.
- Maxwell, G. S. (1984). A rating scale for assessing the quality of responsive/illuminative evaluations. *Educational Evaluation and Policy Analysis* **6**(2), 131–138.
- Nilsson, N. and Hogben, D. (1983). Metaevaluation. In House, E. R. (ed.) *Philosophy of Evaluation (New Directions in Program Evaluation 19)*, pp 83–97. San Francisco, CA: Jossey-Bass.

- Russon, C. and Russon, G. (eds.) (2004). *International Perspectives on Evaluation Standards. New Directions for Evaluation 104*. Hoboken, NJ: Wiley.
- St. Pierre, R. G. (1982). Follow through: A case study in meta-evaluation research. *Educational Evaluation and Policy Analysis* **4**(1), 47–55.
- Stufflebeam, D. L. (1978). Meta evaluation: An overview. *Evaluation and the Health Professions* **1**(1), 17–43.
- Thomas, R. M. (1984). Mapping meta-territory. *Educational Researcher* **13**, 161–18.

Relevant Websites

- <http://www.eval.org> – American Evaluation Association Guiding Principles for Evaluators, American Evaluation Association.
- <http://www.scu.edu.au> – Metaevaluation Bibliography, Southern Cross University, Australia.
- <http://www.alnap.org> – The Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP) Metaevaluations.
- <http://www.gao.gov> – United States Government Auditing Standards (The Yellow Book), Government Accountability Office.
- <http://www.wmich.edu> – Western Michigan University – The Evaluation Center: Evaluation Checklists.
- <http://www.wmich.edu> – Western Michigan University – The Evaluation Center: Joint Committee on Standards for Educational Evaluation.

EDUCATIONAL EVALUATION – PURPOSES AND MODELS OF EDUCATIONAL EVALUATION

Contents

Defining Quality in Evaluation
Internal and External Evaluation
Needs Assessment in Education
The Purpose of Educational Evaluation
Understanding Approaches to Evaluation

Defining Quality in Evaluation

P Dahler-Larsen, University of Southern Denmark, Odense, Denmark

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Anti-criteriological approaches – Views of quality based on the belief that no list of quality criteria exhaust what quality might mean under specific situational circumstances.

COM-list – A list of Criteria of Merit that can be used as evaluation criteria.

Criteriological approaches – Views of quality based on the belief that quality can be defined on the basis of a definitive list of quality criteria.

Evaluation – A systematic, methodological, value-based assessment of some activity, program, or product, having in mind some practical decision-making or learning related to that activity, program, or product or to other activities, programs, or products.

Evaluation theory – Fairly rational and general assumptions about how an understanding of values, knowledge, evaluands, and the use of evaluation, can guide better evaluation practice.

Maximum standards – Ideal standards from which one may deviate if there are good reasons to do so.

Minimum standards – Standards which must be met under all circumstances.

The issue of quality in evaluation is of continuing and increasing relevance for at least three reasons. First, controversies about any particular evaluation often take a starting point in a discussion of whether that evaluation is good or good enough. Low-quality evaluations may hurt human beings in unethical ways, and may be discarded as bases for decision making.

Second, the reputation of evaluation in society is likely to depend to some extent on its perceived quality. To the extent that management tools and political steering mechanisms such as documentation, benchmarking, and quality assurance are perceived as forms of evaluation, their legitimacy may be in jeopardy if their quality is perceived as low.

Third, if evaluation as a field is to consolidate itself and perhaps move toward increased professionalization of evaluators, a vision of what quality of evaluation might mean for the field should be related to how evaluators define the mission of evaluation for society. Low quality is thus a problem for users of evaluation, for evaluators, and for society.

However, it is not easy to define quality. Except that quality is almost exclusively positive, the meaning of quality is ambiguous, and the term relates to esthetic, epistemological, ethical, organizational, managerial, and statistical domains of meaning. Therefore, it is difficult to keep a fixed idea of quality.

While some argue that quality in evaluation can be defined according to particular criteria (although several sets of criteria may be possible), others argue that quality should be appreciated and enhanced through ongoing dialog, education, and reflexivity instead of fixed criteria. Evaluation takes place in many different situations. These situations call for different approaches and models of evaluation. While some quality criteria are specific to each of these models, not all quality criteria may be generally applied to all evaluations.

The first section of this article gives a definition of evaluation, which helps clarify fundamental dimensions in all evaluations. The next short section looks at special issues in educational evaluation. Third, different sets of

criteria for quality of evaluation are discussed. Four such sets of criteria will be considered under the headlines of product-oriented, process-oriented, and user-oriented, and utility-oriented approaches. Fourth, anti-criteriological approaches to quality are considered. In the fifth and final section, various tools and initiatives to enhance quality in evaluation are discussed.

Definition of Evaluation

Evaluation is a systematic, methodological value-based assessment of some activity, program, or product, having in mind some practical decision making or learning related to that activity, program, or product or to other activities, programs, or products. The definition includes many activities or aspects of activities which are usually not called evaluation in daily life, such as benchmarking, quality assurance, audit, accreditation, assessment, etc. Discussions of the quality of evaluation therefore also apply to these other activities and processes.

Based on the above definition, evaluation includes four fundamental dimensions: one has to do with method and systematic knowledge; a second has to do with values; a third has to do with an understanding of the evaluand (object of evaluation); and the fourth has to do with the intent to use the evaluation. A good theory of evaluation describes how evaluators would design a good evaluation as a result of sound reasoning in all four dimensions. Most evaluators agree, however, that evaluation in practice is complicated, and evaluations are not exhaustively designed through logical algorithms. There is more than one good theory in each dimension of evaluation.

Particularly, the value dimension is controversial. Although, say, ethical theory and democratic theory may help clarify value choices, there continues to be an element of choice in selecting values on which evaluation criteria depend. Furthermore, in a given situation, not all factors in all four dimensions point in the same direction. For example, a concern for immediate use may be in conflict with a concern for in-depth analysis. In addition, most evaluators hold explicit or implicit views about their favorite values, methods, evaluands, and types of intended use. Two experienced evaluators do not approach the same situation in exactly the same way. A good evaluation is therefore a more or less thought-through interpretation of a situation rather than a mechanical response to factors in that situation.

Issues in Educational Evaluation

Although all evaluation may be controversial, educational evaluation involves especially complicated issues in all four of the above dimensions. In the value dimension, different

pedagogical views represent different value systems, and the ultimate purposes of education seem to hinge on different interpretations of what constitutes a good life, not only for the pupil or the student, but for the adult as a member of our civilization. In the knowledge-and-method dimension, a pressing question is how closely educational evaluation should attempt to identify causal relationships between teaching and its outcomes. Some critics of the causal view argue that educational situations are existential meetings which cannot be understood instrumentally, or that educational activities have value in themselves. Others argue that situational and human factors make causal links in various educational situations too complicated to map, and that methods oriented toward causal analysis (such as randomized experiments) are unethical or reductionist. At the same time, there is continuous public pressure on modern education systems to demonstrate the outcome of their activities. In the evaluand dimension, people hold fundamentally different views about the nature of teaching, learning, schooling, and education. In the use dimension, a central conflict has to do with the relation between two major purposes of evaluation, learning, and accountability. Some argue that the two are mutually exclusive. If evaluation is to enhance continuous learning among students and teachers, daily evaluation in schools should be protected against external accountability. According to another view, learning and accountability, or, if you will, the formative and the summative function of evaluation, are both possible, even in the same evaluation. The actual use of an evaluation is only determined by the actual users in the actual moment of use, so pupils, teachers, and school principals can take different actions based on the same evaluation. Finally, some argue that in the broader social arena, learning and accountability are both relevant and necessary functions of evaluation. The practical implication of this view is that even though schools work hard to learn internally from evaluation, the societal need for documentation of performance of schools is not likely to disappear.

Criteriological Approaches to Quality

Common for criteriological approaches is their assumption that quality can be exhaustively defined by means of a list of quality criteria. Some of the approaches are discussed below.

Product-Oriented Approaches

These fundamentally view the written product (such as an evaluation report) as the crux of evaluation. Thus, they define quality in evaluation in terms of how well an evaluation report (or a similar product) scores on various questions such as the following: Is the main question clearly stated? Is the report conceptually clear and theoretically

informed? Is the evaluation model well chosen and well explained? Is the method for data collection transparent and well chosen based on the question which needs to be answered? Are the data convincing? Is the analysis systematic? Are recommendations logically flowing from the analysis?

Since these dimensions of quality are methodological, conceptual, and analytical, they are often held in high regard by readers of evaluation reports who have an academic inclination and who see more similarities than differences between evaluation and research. However, product-oriented criteria for quality in evaluation may also include credibility, relevance, timeliness, communication quality (whether the report is well written and easy to read), and documentation quality (whether it provides proper documentation about the evaluator, the evaluation, and the costs of evaluation).

An obvious problem with product-oriented approaches is that they exaggerate the parallel between a tangible product and evaluation as a service. The latter is an interactive process. If the evaluator has played an important role as consultative, facilitative, enabling, or change agent, it is reductionist to gauge the quality of evaluation based solely on an evaluation report. Some evaluation reports are merely a summary of the most important learning points from the whole evaluation process. In some evaluations, innovative and creative communication formats are developed which render a classical evaluation report nonessential or redundant.

Process-Oriented Approaches

These, instead, define quality in terms of key events throughout the evaluation as a social process. Quality is thus something which happens as this process unfolds. Quality has to do with the management of the evaluation process which includes the following phases: focusing the evaluation, designing the evaluation, collecting information, analyzing and interpreting information, reporting information, managing the evaluation, and evaluating the evaluation.

In a narrow interpretation of process-oriented quality, the evaluation process conforms to a flow-chart-like algorithm. Evaluation is seen as sets of activities in each of the above phases, each of which include quality check points, and potential corrective measures if necessary. Critics of this view argue that evaluation services are not so well defined, controllable, and repetitive. In a broader interpretation, process-oriented quality is defined by good behavioral principles for evaluators, and by the interaction and communication between the evaluator, the commissioner, and other relevant stakeholders. More specifically, how well do the commissioner and the evaluator understand each other? How clear is their communication? Are all relevant stakeholders involved appropriately in the

evaluation process? Is the purpose of the evaluation clear to all? Does the evaluator behave ethically? Does the evaluator select the right method and implement it competently? Is the evaluator culturally competent and does he or she respect the values and views of diverse stakeholders? Are evaluation results communicated to all relevant stakeholders? Does the evaluator balance the concerns of the commissioner with the broader concerns for general welfare in society?

Although a process-oriented view of quality of evaluation has many advocates, it does not provide any guarantee for good evaluation without an authority which establishes standards for good evaluation practice and monitors how well evaluators comply with these standards. Furthermore, since the evaluation process includes interaction between evaluators, commissioners, and other stakeholders, principles for the behavior of evaluators alone cannot guarantee the quality of the overall process.

User-Oriented Approaches

These define quality in evaluation from a client's or commissioner's perspective. Criteria for good evaluation are deduced from how well the evaluation responds to their needs and wants. In other words, if the client is satisfied, it is a good evaluation. Although this view sounds evident, it implies a number of problems. One is that clients may be inconsistent. They may change their views several times during an evaluation. They may regret an otherwise good evaluation when they see the results. Second, they may not have the expertise that allows them to specify the type of evaluation that serves their needs. They may therefore demand an evaluation which is methodologically inferior, and be dissatisfied unless it is delivered. Third, client satisfaction may hinge on superficial characteristics of the evaluation or the evaluator. Among the most problematic of such characteristics is that the evaluation is harmless. Faced with managerial or political obstacles or with ideological tunnel vision, clients may sometimes be satisfied with evaluations which support status quo or which have low utility. User-oriented approaches are therefore not always promoting use of evaluation, and neither do they guarantee anything more than one subjective view of the quality of evaluation. For these reasons, clients' views on good evaluation are difficult to synthesize into a philosophically coherent perspective on evaluation quality, but this does not mean that their views or their role in the utilization of evaluations are generally irrelevant for initiatives to improve evaluation quality.

Utility-Oriented Approaches

These define quality in evaluation on the basis of the actual use of evaluations or at least utility (potential for use). Use or utility may in the long run be much broader

than what immediate clients or commissioners may have in mind.

Over the years, the field of evaluation has identified lack of use of evaluation as a major problem. Evaluators have developed different strategies to overcome the problem, and new categories of use have been identified.

Some evaluators develop strategies to maximize the likelihood of only intended use by intended users. These strategies include maintaining a fruitful dialog throughout the evaluation process, understanding the needs of clients, building personal relationships, sharing information about the pros and cons of all choices in the evaluation process, delivering a timely and well-communicated evaluation product, and not leaving the scene until the evaluation has been utilized.

Critics of this view argue that if evaluators focus too much on immediate use, evaluation will tend to focus on minor problems which are easy to repair. Pragmatic, managerial concerns will reign over larger policy decisions, which are more difficult to change. Evaluators may adopt a clientilistic view, where client satisfaction is enhanced at the expense of broader societal concerns.

Others have developed alternative categories of use, such as enlightenment. In this type of use, evaluations may make little difference here and now. However, evaluation can be useful even when it contradicts the views of clients and other decision makers. Evaluations may help challenge conventional practices, develop new ideas, inspire critical thinking, and facilitate policy change in the long run. If evaluations are carried out rigorously, and become publicly known, their overall broad use may be much more beneficial for society than the sum of the intended uses for intended users of a given number of evaluations. What Weiss called enlightenment through evaluation may transgress or contradict what clients or commissioners expect from the evaluation.

Acknowledging that evaluation takes place in political contexts, there may also be strategic or tactical uses of evaluation. Strategic use refers to the use of evaluation results as political ammunition. Tactical use refers to the use of the evaluation process to divert attention from an issue or delay a decision.

In recent years, a number of additional categories of use have been identified. These include process use, which is a reflexive type of thinking that flows not from evaluation results, but from participation in the evaluation process. Symbolic use has to do with the status flowing from evaluation as a fashionable thing to do. Constitutive effects of evaluation capture how human beings respond to having their activities measured, whether or not these responses enhance the quality of services. In education, teaching to the test is one such effect.

Since the categories of use are multiple, and since it is not easy to distinguish clearly between use and misuse of evaluation (one man's use may be another man's misuse),

it is complicated to gauge the quality of evaluation solely on the basis of the use. Use or utility is decided by a number of factors in the social and political context around evaluation. Due to political interests or ideological barriers, a good evaluation can be ignored and a bad one used. It would therefore be unfair to regard use or utility as dimensions in the quality of the evaluation itself.

Situational and Anti-Criteriological Approaches

According to an anti-criteriological view, no fixed set of criteria can be used to determine the quality of a complex situational and relational practice. The choice of the term practice is significant. While evaluation looks at complex, situationally embedded practices such as teaching and social work, it is also not merely a methodological tool, but itself a practice, too. Referring to Aristotle's view of *phronesis*, or practical wisdom, good practice takes into account factors too numerous to count. Practice is embedded in tradition and in social institutions. Practice is situational, and takes into account circumstantial and even personal factors. Practice is also relational, and often carried out in a community of practice. More often than not, complex and sometimes conflicting moral concerns are tightly woven into practical actions so that the instrumental and the moral are inseparable. A good evaluator exercising practical wisdom will begin not by measuring a practice according to standards, but by seeking to understand the practice at hand. The evaluator will then seek to find out what quality might mean in that particular context. Based on responsiveness and participation, the evaluator will then engage in a dialog with people in the evaluated field. The evaluator plays the role of a critical friend, a teacher, or partner rather than a methodological technician. A central issue for the evaluator is the conflicting moral concerns related to explicit or implicit evaluation criteria both from an external perspective and from an internal perspective in the practice under evaluation. Another critical issue is the degree to which people in the field can tolerate critique of their practice. This depends partly on the relation which the evaluator has built with these people. Finally, the evaluator would understand the evaluated practice in the light of broader institutional and political issues in society. Based on this view, what might constitute quality in evaluation is as contextual as in any other practice, although of course the evaluator would be prudent in consulting methodological expertise, ethical guidelines, other respectable evaluators, and his or her own sense of good practice.

Therefore, contextual quality does not imply a relativistic or nihilistic view. Quality in a particular evaluation can be understood by people in the field as well as by educated evaluators who engage in a serious attempt to

understand how the evaluation has responded to the particular situational challenges at hand. Views from methodological specialists, managers, patients, pupils, or other stakeholders, etc. are also relevant, although no fixed set of criteria would exhaust the total meaning of quality in evaluation.

Initiatives to Enhance Quality in Evaluation

The field of evaluation is dynamic, and its borders are loosely defined. People go in and out of evaluation either in the same job function or over a whole career. Many actors in society buy, sell, or commission evaluation in fairly unregulated ways. The demand for evaluation seems to be increasing, and so are attempts to do it quickly. No authority has a high degree of control of the field.

For these reasons, quality in evaluation is more difficult to control than say, medical or legal practice. Under these circumstances, attempts to enhance quality in evaluation have included the following.

Checklists

An evaluation checklist is a list of factors, tasks, dimensions, or criteria, etc. which need to be taken into consideration when performing a particular evaluation. Checklists are of various types, ranging from simple to-do lists to elaborate criteria of merit (COM)-lists which include all relevant criteria to be used in an evaluation. A checklist is usually expressed in less-sophisticated ways than a formal methodology or theory. Yet, a checklist may be useful if it summarizes earlier experiences with a particular form of evaluation or evaluand. To illustrate the usefulness of a checklist, one might think of the pilot's checklist before take off, which hopefully expresses the total sum of relevant systematized knowledge from earlier events that went wrong with that type of airplane. Some experienced evaluators would argue that evaluation is a more situational, relational, and contingent practice than a mechanical checklist would suggest. Checklists at best support the exercise of practical wisdom in evaluation, but do not replace it. Checklists are often specific to particular evaluation models or approaches. A number of checklists for various evaluation approaches can be found below.

Standards and Guidelines

Many evaluation associations have voluntarily adopted a set of guidelines or standards which describe good evaluation practice. When developing guidelines, evaluators discuss questions such as: Should guidelines be minimum (mandatory) standards or maximum ideals (which should not be met in all situations)? If guidelines cannot be met in

a particular situation, may it still not be better to do evaluation than not do it? How should evaluator's guidelines relate to professional guidelines such as those of psychologists, etc.? How should guidelines in one country relate to potential international guidelines (a pressing question in e.g., Europe)? Should the behavior of commissioners, too, be described in the guidelines? Is it possible to develop guidelines which do not favor one evaluation methodology over another? How and how often should guidelines be revised? What is the meaning of guidelines, if sanctions cannot be imposed on those who break them? Finally, if guidelines are to cover all evaluation situations, and they should be backed by consensus, will they not just be broad, self-evident principles?

Among the more notorious ones are American Evaluation Association's guiding principles, which include systematic inquiry, competence, integrity/honesty, respect for people, and responsibilities for general and public welfare. These principles are not prioritized. Each principle, as well as the history of the guiding principles, is explained at the AEA website, see below. Although most evaluators find these guiding principles fairly evident at first glance, interesting debates often occur when the meaning of a specific principle is unfolded in a particular case, or when a balance between several principles must be struck. Respected evaluators may disagree in those situations about how to best respect the guiding principles. Guiding principles may thus be useful not because they directly control the behavior of evaluators, but because they help raise the awareness among evaluators about ethical issues and quality in evaluation. Evaluators meet the guiding principles regularly in teaching, at conferences, on websites, in folders, and in collegial discussions.

Evaluation Culture, Evaluation Capacity, and Evaluation Policy

Over the years, the field of evaluation has learned that evaluations done on an ad-hoc basis run the risk of having little organizational impact. Therefore an interest has developed in a more holistic and integrated system for the use of evaluative information. Evaluation capacity building is an integrated set of activities which aims at increasing the organizational potential for sustaining continuous and effective evaluation. Evaluation capacity building includes structural, cognitive, and communicative features which support evaluation (offices, staff, training, managerial support, organizational communication, etc.). Evaluation culture is a term which describes the norms, values, and practices related to evaluation in an organization. Some believe that evaluation is only effective if an evaluation culture is developed to support it. Evaluation policy is a term used to describe the deliberate and official strategy toward evaluation which is adopted by management. Some believe that without an evaluation

policy, evaluation activities tend to remain sporadic, fragmented, and unsystematic. A common idea in evaluation culture, capacity building, and evaluation policy is thus the ambition to plan how continuous evaluation fits into a larger organizational picture. The principles for evaluation embedded in these attempts of course affect the quality of evaluation carried out. Although many believe in the advantages of a systematization of evaluation, some fear that evaluation will be reduced to a management information system, and that truly innovative and provocative findings will rarely come from managerially approved evaluation systems.

Education

Some believe that evaluation is too complex for checklists, guidelines, and managerial policies to be of more than limited value. Instead, evaluators should be so thoroughly educated that they have both sufficient competences, a strong sense of quality in evaluation, and an ability to reflect upon their own practice. Some universities therefore offer programs in evaluation. However, many evaluators only do evaluation as a part of their total job function, and some discover the need for training in evaluation only after they already have taken on a full-time job. Most evaluators and commissioners of evaluation have no formal education in evaluation, and some have only a limited amount of short-term training.

Professionalization

Some believe that the quality of evaluation can only be enhanced substantially through professionalization. This includes some sort of education or training plus official control mechanisms regarding who call themselves evaluators. Credentialing includes a set of courses one must go through. Certification includes the assessment of an external body of a person's skills, competencies, and credentials. Licensing involves the approval by some authority of that person's ability to actually carry out the practice. Accreditation is a mechanism whereby the program or institution offering courses is examined and approved by an external body. All these mechanisms involve decisions about what constitutes a good or acceptable evaluator, and they involve decisions about who has the authority to make those decisions. None of these decisions are evident. Few countries have gone very far in the direction of professionalization of evaluators.

See also: Evaluation and Accountability; Moral and Ethical Issues in Evaluation; The Purpose of Educational Evaluation; The Role of Stakeholders in Educational Evaluation; Understanding Approaches to Evaluation.

Further Reading

- Altschuld, J. W. (2005). Certification, credentialing, licensure, competencies, and the like: Issues confronting the field of evaluation. *Canadian Journal of Program Evaluation* 20(2), 157–168.
- Brinkerhoff, R. O., Brethower, D. M., Hluchyi, T., and Nowakowski, J. (1983). *Program Evaluation*. Boston, MA: Kluwer-Nijhoff.
- Cousins, B. J. and Earl, L. M. (1995). *Participatory Evaluation in Education – Studies in Evaluation Use and Organizational Learning*. London: RoutledgeFalmer.
- Cousins, J. B. and Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research* 56(3), 331–364.
- Hansen, H. F. (2005). Kvalitet i evalueringspraksis. *Nordisk Administrativ Tidsskrift* 1(86), 27–43.
- Hodgkinson, G. P. and Starbuck, W. H. (2008). *The Oxford Handbook of Organizational Decision Making*. Oxford, NY: Oxford University Press.
- House, E. R. (1993). *Professional Evaluation*. Newsbury Park, CA: Sage.
- Mathison, S. (2005). *Encyclopedia of Evaluation*. Thousand Oaks, CA: Sage.
- Morris, M. (ed.) (2008). *Evaluation Ethics for Best Practice*. New York: Guilford Press.
- Patton, M. (1997). *Utilization-Focused Evaluation. The New Century Text*. Thousand Oaks, CA: Sage.
- Schwandt, T. A. (1990). Defining 'Quality' in evaluation. *Evaluation and Program Planning* 13, 177–188.
- Schwandt, T. A. (2002). *Evaluation Practice Reconsidered*. New York: Peter Lang.
- Schwartz, R. and Mayne, J. (2005). *Quality Matters – Seeking Confidence in Evaluating, Auditing, and Performance Reporting*. New Brunswick, NJ: Transaction Publishers.
- Scriven, M. (2000). The logic and methodology of checklists. http://www.wmich.edu/evalctr/checklists/papers/logic&methodology_dec07.pdf (accessed August 2009).
- Shadish, W. R. Jr., Cook, T. D., and Leviton, L. C. (1991). *Foundations of Program Evaluation – Theories of Practice*. Newbury Park, CA: Sage.
- Shaw, I. F., Greene, J. C., and Mark, M. M. (2006). *The SAGE Handbook of Evaluation*. London: Sage.
- Stake, R. E. (1997). The fleeting discernment of quality. *Advances in Program Evaluation* 3, 41–60.
- Taylor, D. and Balloch, S. (2005). *The Politics of Evaluation – Participation and Policy Implementation*. Bristol, UK: Policy Press.
- Weiss, C. H. (1977). Research for policy's sake: The enlightenment function of social research. *Policy Analysis* 3, 531–545.

Relevant Websites

- <http://www.eval.org> – American Evaluation Association: Guiding Principles for Evaluators.
- <http://www.wmich.edu> – Western Michigan University, The Evaluation Center, Evaluation Checklist Project.

Internal and External Evaluation

D Nevo, Tel Aviv University, Tel Aviv, Israel

© 2010 Elsevier Ltd. All rights reserved.

Internal evaluation and external evaluation have long been under discussion in the evaluation literature (Scriven, 1967, 1991; Stufflebeam *et al.*, 1971; Love, 1991; Mathison, 1991; Sonnichsen, 2000). Scriven (1991) gave the following definitions for internal and external evaluation in the context of program evaluation:

Internal evaluators (or evaluations) are those done by project staff, even if they are special evaluation staff – that is, even if they are external to the production/writing/service part of the project. (p. 197)

An external evaluator is someone who is at least not on the project or program's staff, or someone – in the case of personnel evaluation – other than the individual being evaluated, or their family or staff. . . . It is best to regard externality as a continuum along which one tries to score as high as possible. . . . (p. 159)

At the school level, internal evaluation can be performed by a teacher or a group of teachers, by other members of the school's professional personnel, by the principal or other school administrators, or by a special staff member designated by the school to serve as a school evaluator. An external evaluation of the school can be performed by the school district, the state department of education, or a ministry of education using professional evaluators, regional inspectors, or a district/state/national evaluation department. An external evaluation of the school could also be conducted by an independent evaluation consultant or evaluation firm, commissioned by the school itself or its governing board.

External Evaluation

For many years and in many countries, school evaluation was tantamount to external evaluation (Nevo, 2006). In some countries, it was done mainly by inspectors (e.g., in the UK or other European countries), and, in some others (e.g., in the USA), it was done by means of state or district assessment programs. Many educational systems combined both student assessment programs and overall school reviews, some of them conducted in a very systematic way by central units in the educational system (e.g., OFSTED in the UK or OER in New Zealand). The idea of accountability, which has been around for over a quarter century, and relatively newer ideas of setting standards and benchmarks as major means for school improvement,

can all be traced back to this long tradition of controlling schools by means of external evaluation.

Even before the term accountability was used, there was a clear demand by politicians, administrators, parents, and the public at large that schools be evaluated externally to find out if they were fulfilling their duties. There was also a hope that such external evaluations would motivate teachers and school principals to work harder to improve their schools. This was true for democratic and nondemocratic societies, as well as for centralized and decentralized educational systems alike. The demand never ceased, even when external evaluation was highly criticized by innovative educators and when internal evaluation was encouraged by way of an alternative.

Internal Evaluation

Parallel to the almost-universal phenomenon of external school evaluation, many countries have more recently tended to apply newly developed evaluation methods at the school level in the form of internal evaluation or self-evaluation. Participatory evaluation (Cousins and Whitmore, 1998), empowerment evaluation (Fetterman *et al.*, 1996), total quality management (TQM), and Action Research are some attempts made to apply internal evaluation methods at the school level. These seem to be in line with other prevailing ideas such as reflection (Schon, 1983) and professionalization of teachers (Darling-Hamond, 1992) and educational administrators/leaders. Reflection is, in a way, one kind of self-evaluation, although its proponents do not sufficiently stress the important role of systematic data collection as a basis for reflection. A professional perception of teaching regards teaching as a complex undertaking, suggesting that teachers should identify needs, analyze goals, choose instructional strategies, and plan and monitor their work. Evaluation becomes an integral part of the teaching profession, relevant to various aspects of teachers' responsibilities, and no longer narrowly limited to evaluating student achievements.

Internal evaluation is also a very important component of schools in decentralized or decentralizing educational systems, which are following models of school-based management or autonomous schools (Nevo, 1997). Autonomous schools are urged to define their own educational aims, be in charge of the educational process, and to evaluate their actions. With the expansion of school authority, schools are also expected to take greater responsibility

and to be accountable for their deeds. These demands have emphasized the importance of internal self-evaluation not only to support improvement but also to respond to the school accountability requirement.

Internal evaluation becomes a major tool for school-based management, serving decision making at various administrative levels. The improvement of decision-making processes is especially important within the broader context of ongoing decentralization and school empowerment, mentioned above. Schools are granted significant authority and are, in return, expected to make decisions autonomously. Internal self-evaluation is highly salient in such situations: it can provide schools with means to improve decision-making processes and make them more effective. An internal evaluator is usually better acquainted with the local context of the evaluation and less threatening to those being evaluated. He or she knows the local problems, communicates better with the local people, and remains on site to facilitate the implementation of the evaluation recommendations. Developing an internal evaluation mechanism in a school is also an investment in an enduring resource for serving the information needs of the school by means of data pools and school portfolios.

Internal evaluation is also an expression of school empowerment and transfer of authority from the center to the periphery, that is, from the central government to the local community. Participation in the evaluation process may contribute to the empowerment of an organization as a whole, as it acquires the ability to monitor itself in a more systematic way and by gaining greater confidence in its educational direction. Participation in the evaluation process can furthermore empower individuals in the organization by providing them with evaluation skills, which they can later apply in various contexts. The principals and the teachers participating in evaluation activities can apply the newly acquired evaluation skills and knowledge in other areas of their work. For example, teachers can use their knowledge of research methodology to teach pupils to perform investigative tasks as part of their schoolwork.

Another benefit of participation in evaluation activities is that it may increase teachers' involvement in decision-making processes outside the classroom; it may foster collegiality and collaboration among the teachers and serve as a means of promoting reflection. All these are central to development of teacher professionalization (Darling-Hammond, 1992).

However, pointing out the importance of internal evaluation or self-evaluation and its relevancy to professional teachers and autonomous schools should not lead us to ignore external evaluation. Control and accountability requirements and the right of the public to know in democratic societies have also to be kept in mind.

Although internal evaluation, by providing information to parents and the community at large, can also enhance accountability, the credibility of its findings might be limited without external evaluation.

Mutual Benefits

If internal and external evaluation are both important, we need both of them and they might even benefit from each other.

Internal evaluation can benefit from external evaluation in at least three ways. External evaluation can stimulate internal evaluation, expand its scope, and it can legitimize its validity. A more detailed discussion of these three ways is given below.

Stimulating Internal Evaluation

Evaluation is a demanding undertaking whose benefits are yet to be proved. It requires a significant amount of resources, such as funds, time, and skilled personnel. We evaluators believe in its potential usefulness, but our research on evaluation utilization has so far yielded only meager support to our beliefs, and mainly in relation to conceptual use of evaluation (Cousins and Leithwood, 1986; Cousins and Shulha, 2006). The common wisdom is that evaluation is a required activity and hopefully a useful one.

It would be difficult to argue that doing self-evaluation comes naturally to teachers and schools, even less so – being evaluated by others. Schools will surrender with or without protest to inevitable external evaluation requirements, imposed on them by the educational system. In some cases, they might opt for internal self-evaluation if they believe that this will release them from external evaluation. Experience shows that schools might be motivated to engage in self-evaluation if faced with an external evaluation requirement, even when internal evaluation is not suggested as an alternative to external evaluation but only as a prior condition and counterpart. Some schools will tend to do internal evaluation to help them confront the external evaluation. Others believe, as mentioned earlier, that they might be released from the burden of external evaluation if they do internal evaluation or just establish an internal evaluation team. Sometimes it is difficult to avoid the notion that the most important function of external evaluation is to motivate people and organizations to do internal evaluation.

Expanding the Scope of Internal Evaluation

Although internal evaluation usually has the advantage of being more sensitive to the local context of schools and

their unique characteristics, it might suffer from a narrow perspective on their overall qualities. External evaluation can add commonalities to the uniqueness of the school as well as provide a basis to judge its qualities. School inspectors can provide useful observations from their inspection region. Other external evaluators, such as professional state evaluation teams, may be able to supply the school with relevant information available at the central office.

External evaluation can also provide the school with information on national standards or benchmarks or comparative data from other schools (e.g., student achievements) that might help the school interpret its own data and assess its quality. Local standards might be narrow in their scope or not available at all. For example, a national evaluation that has focused on school violence prevention projects, can provide schools engaged in internal evaluations of school climate not only with national comparative data on mean rates of violence, but also with a more comprehensive definition of school violence and its measurement.

Legitimizing the Validity of the Internal Evaluation

Although internal evaluation should exist in its own right, it is always suspected of being biased and subjective. External evaluation can be as biased too, but the accusation of subjectiveness tends to be directed more toward internal evaluation. Being blamed for lack of objectivity is not only a treat to the credibility of internal evaluation, but it also jeopardizes its existence. An evaluation that is not trustworthy might have difficulties in obtaining the resources necessary for its existence. Why should the public waste money on such activity? Why should teachers spend their precious time on an internal evaluation which nobody trusts?

External evaluation can help legitimize internal evaluation – not by providing a formal stamp of approval but by respecting its existence as an important ingredient of school evaluation. External evaluation can engage in a dialog with internal evaluation on equal grounds and with mutual respect, sharing the common cause of gaining a better understanding of the school and its problems. External evaluators do not have to agree with all the findings of the internal evaluation, but they have to perceive them as good enough to argue with.

External evaluation can benefit from internal evaluation in many ways and not less than how it can contribute to internal evaluation. Internal evaluation can deepen the scope of external evaluation by increasing its awareness of local issues; it can improve the interpretation of external findings by making them more sensitive to local needs; and it can also improve the utilization of external evaluation

by diminishing resistance to it and increasing evaluation mindedness (Glasman and Nevo, 1988).

Expanding the Scope of External Evaluation

External evaluation is too often criticized for its narrow scope and its tendency to focus on commonalities rather than the uniqueness. In its attempt to seek comparability and generalization, external evaluation is forced in many cases to address the lowest common denominator, which might be either trivial or insensitive (or both) to local needs and priorities. Our own experience in evaluating the Thirty Townships Project (Nevo and Friedman, 1999) is a case in point. This project was initiated and implemented by the Israeli Ministry of Education, with the aim of improving local education systems of 30 townships suffering from low levels of student achievement, high dropout rates, and low community self-image. Although the main goal of the project was the same for all townships, various intervention agencies implemented the project, together with each township, developing unique intervention programs to meet local needs and in line with the educational perspectives of the intervention agency. The result was 30 different intervention programs. The Ministry of Education commissioned an external evaluation to focus on project outcomes and to be used mainly for accountability. The external evaluation was expected to supply information about the overall impact of the project and thus used an overall set of indicators, which affected its capability of supplying in-depth portrayals of specific interventions. Internal evaluation could, in this case, extend the scope of the external evaluation by pointing out the relevancy of additional data reflecting the unique character of a particular school and local educational system.

Improving Interpretation of Findings

The external evaluation usually has the advantage of allowing to interpret findings regarding the quality of a specific school by comparing them to other schools or national standards and benchmarks. While such interpretations are usually important and legitimate, they might also overlook the local perspective, reflecting the special needs and opportunities of the school. It is the internal evaluation that can add on that local perspective of school and community. The above-mentioned Thirty Townships Project, is a good example of the limited ability of external evaluation to interpret specific findings regarding local perspectives, needs, and constraints.

Moreover, as societies become increasingly diverse and multicultural, it is not only important that external evaluation be complemented by locally specific data, but also important that local perspectives and values be considered in interpreting evaluation findings.

Increasing Evaluation Utilization

Schools that have internal evaluation teams, for whom evaluation is part of their pedagogical and administrative life, have a better chance of understanding the meaning of evaluation and its significance in education. Such schools also have a better chance of appreciating the potential usefulness of external evaluation to the lives of their schools rather than opposing it defensively.

Schools that practice evaluation, by establishing internal evaluation teams or other evaluation mechanisms, increase their institutional evaluation literacy. They have a better grasp of the significance of educational evaluation and can speak its language. They gain technical skills to understand external evaluation reports and argue with their findings, if necessary, rather than opposing them.

A school that engages in self-evaluation, and collects information on its pedagogical and administrative activities for that purpose, might increase its self-confidence and be less defensive when confronted with negative findings from an external evaluation; thus trying to make use of evaluation rather than opposing it.

A school engaged in internal evaluation will tend to develop a commitment to evaluation, which has been found in various evaluation utilization studies (Cousins and Leithwood, 1986; Cousins and Shulha, 2006) as an important factor affecting the use of evaluation.

Terms of Coexistence

If internal and external evaluations are both important and can also benefit from each other, then the remaining question is: How can they coexist? Nevo suggests that they can coexist in a constructive way if an appropriate ground can be created for a dialog between the two (Nevo, 1995, 2001).

We all love the concept of dialog and not many people would dare to argue against it. However, dialog cannot flourish on all grounds and in all climates. Some terms of existence have to be provided to create an appropriate habitat for constructive dialog. These terms are not self-evident and sometimes quite difficult – although not impossible – to achieve. They relate to three aspects of dialog evaluation: conception, methodology, and communication.

On the conceptual level, evaluation has to be perceived as a means for understanding rather than judgment, providing quality profiles rather than composite scores.

On the methodological level, dialog evaluation should be practiced as a process and not as a one-shot activity. The interaction between internal and external evaluators should be based on a two-way flow of information in a process of mutual learning, and evaluation should focus on relevant issues and pertinent data.

On the communication level, there must be mutual respect and trust between the parties. Both have to believe that each has a genuine interest in understanding what is

at stake and can make a significant contribution to such an understanding; evaluation has to be fair to both parties involved in the dialog and both parties should take some responsibility for handling the consequences of the evaluation.

Conclusion

In many educational systems, everybody seems to hate external evaluation while nobody trusts internal evaluation. This article contends that both types of evaluation are needed as they both have important roles in the life of schools, teachers, and educational systems. The article also suggested some conditions that have to be met to establish a constructive dialog between internal and external evaluation as a basis for their coexistence.

Thus, those of us who are proponents of external evaluation should find ways to empower schools and teachers to participate as equal partners in the evaluation process and make use of it; and those of us who believe in internal evaluation as a means for school autonomy and teacher professionalization must admit the legitimacy of accountability and the right of the public to know. They, in their turn, should seek external evaluation as a partner for dialog rather than an object for rejection.

The leaders in many countries and the public at large usually believe in external evaluation and are ready to spend a lot of money on implementing grandiose national school evaluation systems, but have a hard time to waste money on internal evaluation, which nobody trusts anyhow. They must realize that if they can overcome the resistance of teacher and other educators to external evaluation and if they think that it is also useful and being used, they should not waste any money on internal evaluation. However, deep in their hearts they know that this is not the case.

On the other side, school teachers and other professional educators have to agree with external evaluation requirements and waste time on its implementation. They must realize that if they will not respect the authority and responsibility of the ministry of education and the right of parents to know about the schools their children go to, they cannot expect them to respect their right for autonomy and reflection and trust their judgment as professional teachers.

See also: School Based Evaluation: Purposes, Protocols and Processes; School Inspection/External School Evaluation.

Bibliography

- Cousins, J. B. and Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research* 56(3), 331–364.

- Cousins, J. B. and Shulha, L. M. (2006). A comparative analysis of evaluation utilization and its cognate fields of inquiry: Current issues and trends. In Shaw, I. F., Greene, J. C., and Mark, M. M. (eds.) *The SAGE Handbook of Evaluation*. London: Sage.
- Cousins, J. B. and Whitmore, E. (1998). Framing participatory evaluation. In Whitmore, E. (ed.) *Understanding and Practicing Participatory Evaluation: New Directions for Evaluation*, 80, pp 2–23.
- Darling-Hammond, L. (1992). Teacher professionalism. In Alkin, M. C. (ed.) *Encyclopedia of Educational Research*, 6th edn, pp 1359–1366. New York: MacMillan.
- Fetterman, D., Kaftarian, S. J., and Wandersman, A. (eds.) (1996). *Empowerment Evaluation*. Thousand Oaks, CA: Sage.
- Glasman, N. S. and Nevo, D. (1988). *Evaluation in Decision Making*. The Case of School Administration. Boston: Kluwer.
- Love, A. J. (1991). *Internal Evaluation: Building Organizations from Within*. Newbury Park, CA: Sage.
- Mathison, S. (1991). What do we know about internal evaluation? *Evaluation and Program Planning* 14, 159–165.
- Nevo, D. (1995). *School-Based Evaluation: A Dialogue for School Improvement*. Oxford: Pergamon.
- Nevo, D. (1997). The function of evaluation in school autonomy. In Shapira, R. and Cookson, P. W. (eds.) *Autonomy and Choice in Context: International Perspectives*. Oxford: Pergamon.
- Nevo, D. (2006). Evaluation in education. In Shaw, I. F., Greene, J. C., and Mark, M. M. (eds.) *The SAGE Handbook of Evaluation*. London: Sage.
- Nevo, D. and Friedman, E. (1999). The common and the diverse: Evaluation of multi-dimensional projects. *Paper Presented at the Annual Meeting of the European Conference for Research on Learning and Instruction*. Gothenburg, Sweden, August, 24–28.
- Schon, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.
- Scriven, M. (1967). The methodology of evaluation. In Stake, R. E. (ed.) *AERA Monograph Series on Curriculum Evaluation No. 1*. Chicago, IL: Rand McNally.
- Scriven, M. (1991). *Evaluation Thesaurus*, 4th edn. Thousand Oaks, CA: Sage.
- Sonnichsen, R. C. (2000). *High Impact Internal Evaluation*. Thousand Oaks, CA: Sage.
- Stufflebeam, D. L., Foley, W. J., Gephart, W. J., et al. (1971). *Educational Evaluation and Decision-Making*. Itasca, IL: Peacock.

Further Reading

- Patton, M. Q. (1997). *Utilization-Focused Evaluation*, 3rd edn. Thousand Oaks, CA: Sage.
- Ryan, K. E. and Cousins, J. B. (eds.) (2009). *The SAGE International Handbook of Educational Evaluation*. Thousand Oaks, CA: Sage.
- Stake, R. E. (2004). *Standards-Based and Responsive Evaluation*. Thousand Oaks, CA: Sage.

Needs Assessment in Education

J W Altschuld and T L Lepicki, The Ohio State University, Columbus, OH, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Causal analysis – Examining needs in terms of their potential causes particularly in regard to identifying those that can be changed by the school (district) and those that are not under its aegis.

Need – The measured discrepancy between the current (or present) situation and the desired (ought to be) one.

Needs assessment (NA) – A formal process to obtain information on the two states (current vs. desired); identify gaps between them; and arrive at needs-based priorities for actions.

Needs Assessment Committee (NAC) – A group that oversees the conduct of all aspects of an NA, including the development of plans for resolving needs. (The NAC may also be involved in the implementation of NA procedures and resultant solution strategies.)

Introduction

We probably have a sense of what a need is and may have participated in or led a needs assessment (NA). However, after thinking about things a few more questions arise:

- What really are these things – needs and NA?
- How do we organize for and conduct assessments?
- Why should we do an NA now – what is motivating it?
- Who or what groups are pushing for it?
- What kinds of data might be collected?
- Would any individuals or groups not want to provide input into the NA?
- How will we analyze what we find and what good will come from it?
- If identified needs lead to new directions, are teachers, school staff, administrators, and communities ready and motivated to act?
- Who is going to work on resolving needs?
- What sacred cows in school systems might pose obstacles for the assessment or be encountered when implementing activities to resolve needs?

A seemingly straightforward idea is not so simple after all. In this respect, it has been noted that if assessments are viewed just as methods (planning, asking, analyzing, etc.),

the likelihood of success is sharply reduced (Altschuld, 2004). Methods are important, but an NA should be a blueprint for organizational improvement, learning, and change – a perception essential for maximum effect.

To explain what NA is, definitions followed by descriptions of a generic process for making assessments and methods often employed for looking at needs are given here. Suggestions are then offered as to typical uses of NA and when it might be done as shown by brief examples from the literature.

Defining Key Terms

What Is a Need?

A need is the measured discrepancy between the current (or present) situation and the desired (ought to be) one. Without data about both states and the ability to contrast them, it is not possible to have a defined need. Surprisingly, until recently, many NAs did not investigate the two conditions. The focus was more on solutions to problems rather than on probing into and understanding them. Needs indicate the size and direction of gaps, and when compared and prioritized they serve as input for actions to alleviate underlying concerns. While the definition implies quantification ($x - y = z$), most assessors promote the use of mixed methods (including qualitative ones) for full comprehension of needs.

Examples of needs (several pertaining to education) are given in **Table 1** with many subtle features embedded in them. One is that the desired condition may be stated in different ways with implications arising from how they are framed. For example, it could be worded as what is minimally expected (as in minimal competency standards) or what is ideally expected. Such wordings indicate that there are multiple perspectives about needs. Another feature is that needs might be narrowly or globally assessed (see the last entry in **Table 1**) or have more of a future orientation, whereas others are in the near term. Finally, note that the row about wealth illustrates a want not a need. It is a discrepancy, but it would be considered by most to be a want (not a need).

Are There Levels of Needs?

Yes, there are. Witkin and Altschuld (1995) wrote about the three levels of needs:

- Level 1 – receivers of goods and services (students in schools, patients in hospitals, consumers or purchasers of products, those in need of guidance counseling, etc.).

Table 1 Examples of the “What is,” “What ideally should be,” and “What is likely” states

Area	What is	What ideally should be	What is likely
Health	30% of US is overweight	100% at or near a reasonable weight for age, height, gender, and body build	75% will reach the standard within a 5-year period
Mathematics	62.8% of district students achieve the state standard for the fourth-grade mathematics test	100% reach the standard or 75% reach the standard to remove the district from possible state sanctions	65% or more achieve the standard by this time next year
Reading	75% of eighth-grade students read and understand the instructions on an aspirin bottle or on a bottle of patent medicine	100% should be able to do the task	85% are able to do the task 2 years from now after exposure to improved reading instruction
Youth recreation	A community does not have a recreation center and adequate recreation activities for youth	A recreation center will be built and open 5 years from now Within 1 year a recreation program will be started in the community	The recreation center will be a reality 10 years from now A small recreation program will start in 2 years and gradually expand
Immunization	The inoculation rate for preschool children in a particular state is currently at approximately 70–75%	A rate of 90–95% will be achieved, thus greatly reducing the likelihood of increasing the incidence (spread of certain diseases)	Rate of inoculation will gradually increase to 80% over a 5-year period Rate will remain the same without understanding the causes of the problem
Wealth	An individual is currently worth \$1 000 000	With inflation and worries about job stability, the individual would prefer to be at \$2 000 000 to feel more secure	\$1 500 000 would be likely in light of the general growth of investments within a 10-year period
Overall education system	Current state standards for courses and areas required for a high school degree	Given possible changes in knowledge and the world of work, what standards should we develop for children now entering the educational system and who graduate in 13 years	What are reasonable expectations for change in complex multidimensional systems like education

Adapted from Altschuld, J. W. and Kumar, D. D. (2009). *Needs Assessment: An Overview (Book 1)*. Thousand Oaks, CA: Sage.

- Level 2 – providers of services and goods for level 1 (teachers, health care workers, company trainers who keep staff up to date with the latest technologies, etc.).
- Level 3 – the management, administration, and support necessary to enable work by making available finances and other similar entities. It is the structure or system that makes it possible for level 2 to serve level 1.

What frequently happens in NAs is that by identifying the needs of level 1 those of levels 2 and 3 become obvious. On the other hand, misguided NAs place too much stress on the training of level 2 (teachers) or looking at organizational problems at the expense of level 1, thus subverting the systematic approach required for the best NAs. After all, the function of education is to make certain it attends to level 1 (students) and that what levels 2 and 3 do should positively impact on the needs of the ultimate recipients of goods and services. That is the *raison d'être*.

What Is NA?

As a guide to decisions and actions, NA is a formal process to obtain information on the two states (current vs. desired); identify gaps between them; and arrive at needs-based

priorities for actions. Causal analyses of needs (why is the need occurring) are often part of the endeavor. NAs begin when a school or educational group feels that it has problems that must be explored in depth. This perception leads to a study that deduces the priorities and plans for change and/or new directions.

Going back to **Table 1** it is clear that the wording of the ‘what should be’ condition has a major impact on what the organization does for identified needs. Resolution of a realistic or a short-term slight gap is a far cry from resolution of one that is long term and major in nature. It is much more difficult to muster support and enthusiasm for the latter.

Models for Assessing Needs

Numerous models, frameworks, and approaches from a variety of fields are available for assessing needs. While an extensive examination of these is beyond the scope of this article, the reader is recommended to the overview developed by Watkins and the website on which he lists many of them.

Here, we describe two well-known and widely used (complementary) models. One is a model of Kaufman and the other

a version of a three-phase model recently updated by Altschuld and Kumar (2009). In the Kaufman approach (Watkins and Wedman, 2003), needs are discrepancies between desired (what should be) accomplishments and current (what is) achievements. Gaps in such results can occur at any point along a continuum that goes from the mega- to the micro-levels, which are linked together in a systems type of way:

- Mega-level – results (and needs therein) for which the primary client and beneficiary is society as a whole.
- Macro-level – results (and needs therein) for which the primary client and beneficiary is the organization.
- Micro-level – results (and needs therein) for which the primary client and beneficiary are individuals and groups within the organizations.

Kaufman's thinking and usage of the model are prevalent in training in business and industry settings. The system dimension in his writings aligns closely with the three phases of NA underlying the second framework where the emphasis is on the process and procedures for conducting the NA. Its main steps are given below.

1. *Phase 1.* Pre-assessment focuses on the NA and deals almost entirely with existing data and information relative to the needs area, to:
 - a. develop the focus for the assessment;
 - b. establish a Needs Assessment Committee (NAC) responsible for the remainder of the activities within this phase;
 - c. sort out the values driving the assessment;
 - d. identify and locate existing resources and information for the need area; and
 - e. arrive at a consensus regarding recommendations for the organization and its decision makers.
2. *Phase 2.* Assessment is based on the premise that if phase 1 does not produce enough information, new data must be collected; that is:
 - a. identify discrepancies at levels 1, 2, and 3, which can be done by a variety of data-collection methods (see **Table 2**) with multiple methods usually being part of the landscape;
 - b. prioritize discrepancies;
 - c. conduct causal analysis of needs; and
 - d. identify preliminary solution criteria and possible solution strategies.
3. *Phase 3.* Post-assessment is when there is compelling evidence for high-priority needs and we prepare action plans to tackle the problems and the causes underlying them, such as:
 - a. make final decisions to resolve needs and select solution strategies;
 - b. develop action plans for solutions, communicate, and build support for them;

- c. implement and monitor plans; and
- d. evaluate the overall NA endeavor (document with an eye to revisit and reuse).

A number of things are important to highlight. In the early activities of the two approaches (especially in phase 1), stress is placed on using existing sources of data. To the extent this occurs, the costs of conducting an assessment are substantially reduced.

For the health status of children, there are reams of easily accessible data in the National Health and Nutrition Examination Survey (NHANES). From this database, current health can be looked at in terms of health standards to ascertain discrepancies. In education, local and state educational systems have volumes of relevant data and information. Every state has massive archives of information from accountability systems and tests conducted on a yearly basis. Results can be contrasted to norms for tests. Since the data are collected over time, it is possible to study trends and look at emerging needs and then to establish policies and programs for rectifying problems that are uncovered. Sources such as the two just given should be studied before collecting any new data and incurring attendant expenses.

In conducting NAs it is recommended that an NAC be formed (the second step in phase 1) to guide the process. In general, such groups consist of 8–12 individuals chosen on criteria such as being: respected in the school or organization for their opinions and accomplishments; open to new ideas and ways of considering what the organization is doing and what it might be doing in the future; willing to express their views honestly and frankly even if they differ from those of others; respectful of the perceptions of others with an appreciation for how committees work and get things done; and committed to the endeavor by pitching in to get the necessary work completed.

The committee is responsible for looking at the general questions, focusing them down, establishing what should be explored, and locating and analyzing pertinent information. Eventually, the NAC comes to one of three decisions:

1. there is not much of a need and it is not fruitful to go further;
2. more information is desired, and we must move into phase 2; and
3. enough has been learned about the problem and it is clear, hence we should move on to phase 3 and action planning.

NACs are mostly facilitated by an external person with NA experience and knowledge of what is involved in a successful effort. As the assessment progresses into the third phase of an NA, it may be desirable to have the leadership become more internal in nature. This is because the translation of needs into action plans are to be undertaken by staff and organizational personnel. Having a well-chosen

Table 2 An Overview of needs assessment methods

<i>Data type</i>	<i>Comments/description</i>	<i>Information generated</i>
<i>Archival</i>		
Records/logs	Data exist usually in databases or records	Mostly quantitative data about the current (what is) status of target groups
Social indicators	In some instances, it may be possible to initiate new record keeping	Sources may reveal ideas about causes contributing to needs
Census data	Existing data may not exactly match the intent of the needs in question	Some databases/records include comments necessitating qualitative analyses and interpretation
Epidemiological studies		
Test data		
Information derived from databases		
Other existing sources		
<i>Communicative–noninteractive</i>		
Written questionnaires	Methods rely primarily on structured instruments or forms	While some of the data may be quantitative in nature, they come from values, judgments, and opinions of those providing responses and perspectives
Critical incident technique		
Mailed Delphi surveys	Surveys mostly have scaled questions with perhaps a few open-ended ones	
Web-based surveys		
Observations	Observations may follow detailed protocols or permit more freedom in describing phenomena	
<i>Communicative–interactive</i>		
Public hearings	Procedures involve the use of small or large groups with varying degrees of interaction	Highly qualitative data that are summarized into themes and recurring concepts
Community group forums	Group leadership is especially critical to the success of the procedures and the results produced	Data will be about perceptions, opinions, judgments, and values
Nominal group techniques (NGTs)		Information might be about consensus on goals, courses of action, causes, priorities, and the like
Focus group interviews (FGIs)		
Cyber or virtual FGIs		
Key informant interviews		
DACUM process		
Scenario discussions		
<i>Analytic</i>		
Fishbone diagrams	Processes that examine solution strategies and causes or risks associated with needs and/or ways to resolve them	Problems that might lead to the failure of a solution strategy
Cause and consequence analysis		Guidance in choosing a strategy with a high likelihood of resolving a need (with other information from NA, makes for a more comprehensive understanding of needs)
Quality function deployment (QFD)		
Fault tree analysis (FTA)	Results might be shown in graphs or diagrams emanating from the analytic process	
Success mapping		
Task analysis		
Risk assessment		
Trend analysis		
Cross-impact analysis		
Force-field analysis		

Adapted from Altschuld, J. W. and Kumar, D. D. (2009). *Needs Assessment: An Overview (Book 1)*. Thousand Oaks, CA: Sage.

NAC can be very helpful for identifying what is available and for obtaining existing data sources.

Phase 2 requires the creation of instruments and data-collection techniques. Interviews and surveys have to be developed and then implemented. Systematic observations are undertaken. Various group procedures (hearings, community group forums, etc.) are planned and conducted to provide more information and insights into needs. All such activities take additional money and time, so only enter the phase if absolutely necessary. Factor in that as more data are obtained from multiple and different sources the complexity of data analysis and integration escalates dramatically. This is not to say that multiple sources are not useful and should not be pursued but that the situation is a bit more complicated when they are part of the equation of NA.

Some of the methods in phase 3 are different in that they are analytical and employed on the assumption that

needs have been delineated and prioritized. The focus is on probing into what might be causing the needs and if those variables can be determined, then it becomes somewhat easier to think about solution strategies. They also may push us into thinking about what might happen if nothing was done about needs (could they lessen or abate over time) and what might be the risks in attending to or not attending to them. New data are not, for the most part, necessary here; rather, what has been found is subjected to procedures like cause and consequence analysis, fault tree analysis, and fishbone diagrams to see what may be influential in producing problems.

Once the causes are determined, we should be better able to devise solution strategies and to see how effective they would be in resolving the problem. This leads to an action plan for rectification, and since the problem is understood in greater depth, we should have a better grasp of key variables to evaluate as programs are implemented to reduce

needs. If problems are not resolved, the entire process may be thought of as unsuccessful. Therefore, the third phase entails evaluation of not only the solution strategy but also the NA itself. Evaluation is part of the other phases as well but becomes particularly prominent in this phase.

The methods of NA are quite diverse with some more appropriate for one part of the process than another (see **Table 2**). Detailed explanation of the methods in **Table 2** can be found in the work of Witkin and Altschuld (1995), and Altschuld and Kumar (2009). As noted, experienced needs assessors recommend using a combination of quantitative and qualitative methods to understand needs that might be quite complex (such as what is the underlying nature of the problem in the first row of **Table 1**). Quantitative outcomes indicate the size and direction of discrepancies between the two conditions but cannot get into what they mean for the individuals in need. For example,

- Can numbers fully portray a person's feelings about being homeless, foreclosed out of a home, or losing a job?
- What are the feelings of students who are in jeopardy of not passing tests mandatory for graduation from high school?
- How might raising or lowering the passing scores on tests be perceived by students, teachers, and the general public?
- Even if we know that schools in the US are below other countries in mathematics and the sciences, can we find out why from just the test scores or does the situation require other kinds of studies?

The value of qualitative data is apparent for digging beneath numeric information. The juxtaposition of the two creates a more powerful sense of meaning and sensitivity to issues. The practice of NA suggests that the combination of methods is the way to go.

Across all phases, results and progress should routinely be communicated to the organization, its key personnel and decision makers, and stakeholding groups. There is no set schedule; judgment of the NAC and its facilitator should determine when it seems best. Why is this critical?

NA is oriented toward improvement of an organization and is a frank and realistic way of delving into problems with which it must deal. Action plans might change habitual ways of doing work and/or might require shifting resources or emphasis from one department to another. If not handled well and tactfully, this can lead to acrimony and loss of morale and trust. Communication is essential for establishing the right kind of social climate, and the process generally will not work without attention to it.

Now that the foundation has been built, how does the NA get off the ground and when does it occur? What follows is an approximation of how NAs start and trigger events that typically set the process in motion. Every situation is unique and other concerns and foci may be the active ingredients elsewhere.

For What Situations and When Might NA Be Used in Education

NA can be used for an array of educational concerns (lower-than-desired achievement in mathematics and science, school dropout rates, violence in and around schools, problems related to the education of children from broken homes who are in foster care, the guidance and counseling needs of elementary through high school students, problems arising from alcohol and substance abuse, and so on). It goes without saying that almost any area is a possibility for an assessment, but at the same time that does not mean that all needs will be investigated.

Therefore, when would you do an assessment and how do you narrow the focus so that it is meaningful and the probability of findings getting used are increased? NAs are often thought about at the point administrators or involved stakeholders (teachers, parents, or community groups) become concerned about problems. Examples might be: "Our students aren't doing as well academically as we would like," "We feel that there is an emerging drug and alcohol problem in this area," and "We are seeing less minority student enrollment and retention in math and science disciplines." People sense that their concerns merit a closer look and seek a mechanism for making this happen.

NA occurs early in the process of change as opposed to evaluation which is about how well a program is implemented and the degree to which it is resulting in the outcomes and impacts it was intended to achieve. NA and evaluation overlap, but they are done at different times and have unique functions.

General Applications in Education

One interesting application is the Trends in International Mathematics and Science Study (TIMSS), which consists of standardized tests to measure and compare the progress of students in relevant areas. The US did not perform as well as expected in relation to many other countries. Follow-up investigations as to what might be causing this situation indicated that the curriculum might be a factor contributing to lower achievement. In the US, it was characterized as being a mile wide and an inch deep (featuring many topics but not going into them in much depth) in comparison to what is done elsewhere. This observation along with others has led to much introspection about mathematics and science education in the US – its structure and the instructional strategies used (Schmidt and Wang, 2002).

In essence, testing programs and educational accountability systems have all the elements of NA. The test helps get at the 'what is' condition or state, and then the scores are looked at either for individuals or groups in terms of

standards or expected condition status – what ought to be. When the discrepancy between them is too great (outside of a reasonable range of tolerance), school systems will consider what actions they could take to reduce the deficits. Testing programs and school accountability are not ordinarily posed as NA, but indeed they are.

Hunt *et al.* (2002) employed the technique in a school system to explore the nature of violence and ways to resolve and/or reduce such incidents. They studied the problem for over a year through mixed methods to obtain the perspectives of students, teachers, staff, parents, and community members. By means of the prolonged involvement on site with stakeholders, the needs assessors felt that they were able to get more buy-in to solutions to the problem.

An NA was incorporated into an evaluation of a state-wide university consortium dealing with the retention of minority students in the fields of science, technology, engineering, and mathematics (STEM). Students and faculty/administrators were surveyed with similar but slightly different versions of scaled questions. In general, the two groups were in agreement but when they were not, a second survey was sent to them asking them in open-ended format to explain why the groups differed. By this means, the subtle and underlying complexity of quantitative needs was uncovered (Lee *et al.*, 2007a, b).

Chauvin and Anderson (2003) looked at the potential training needs of public health workers. They developed a sophisticated questionnaire asking various levels of stakeholders about the frequency of key tasks, the competency of workers to perform them, and the motivation to have further training for them. They then contrasted different groups of respondents on the three distinct questions and pinpointed where it might be most beneficial to offer training and where it would tend to be well received. The three areas of the questions were rated on a five-point scale. While the scaling technique was somewhat intricate, it was offset by the amount of information generated, the in-depth understanding of needs, and by the possibility of implementing better and more meaningful training for the public health workforce.

Observations from the four NAs described above underscore key, important aspects of NA. It was apparent across all cases that the process helped to illuminate subtle, under-the-surface dimensions of problems. Another aspect was the influence, in some instances, on how actions plans were developed and/or implemented. Third and, perhaps the most notable, was that the data changed the nature of the discourse related to thinking about policies. Even if there was not much in the way of tangible change, decision makers became more thoughtful and attentive to implications of policies and possible directions to be pursued. This is to us the most valuable outcome and emphasises our bottom line that NA is well worth the effort.

Acknowledgment

This article is partially based on the work of Altschuld and Lepicki (2009).

See also: Evaluation Methodology; Evaluation of Adult Education and Training Programs; Evaluation of Integrated Health Programs in School; The Role of Stakeholders in Educational Evaluation.

Bibliography

- Altschuld, J. W. (2004). Emerging dimensions of needs assessment. *Performance Improvement* 43(1), 10–15.
- Altschuld, J. W. and Kumar, D. D. (2009). *Needs Assessment: An Overview (Book 1)*. Thousand Oaks, CA: Sage.
- Chauvin, S. W. and Anderson, A. C. (2003). Training needs assessment questionnaire for public health competencies in emergency preparedness and bioterrorism. *South Center for Public Health Preparedness*. New Orleans, LA: Tulane University.
- Hunt, M. H., Meyers, J., Davies, O., *et al.* (2002). A comprehensive needs assessment to facilitate prevention of school drop-out and violence. *Psychology in the Schools* 39(4), 399–416.
- Lee, Y. -F., Altschuld, J. W., and White, J. L. (2007a). Effects of the participation of multiple stakeholders in identifying and interpreting perceived needs. *Evaluation and Program Planning* 30(1), 1–9.
- Lee, Y. -F., Altschuld, J. W., and White, J. L. (2007b). Problems in needs assessment data: Discrepancy analysis. *Evaluation and Program Planning* 30(3), 258–266.
- Schmidt, W. H. and Wang, H. A. (2002). What role should TIMSS play in the evaluation of U. S. science education? In Altschuld, J. W. and Kumar, D. D. (eds.) *Evaluation of Science and Technology Education at the Dawn of a New Millennium*, pp 23–48. New York: Kluwer Academic/Plenum Publishers.
- Watkins, R. and Wedman, J. (2003). A process for aligning performance improvement resources and strategies. *Performance Improvement* 42(9), 9–17.
- Witkin, B. R. and Altschuld, J. W. (1995). *Planning and Conducting Needs Assessments: A Practical Guide*. Thousand Oaks, CA: Sage.

Further Reading

- Altschuld, J. W. and Lepicki, T. L. (2009). Needs assessment. In Watkins, R. and Leigh, D. (eds.) *Handbook of Improving Performance in the Workplace, Vol. 2, Selecting and Implementing Performance Interventions*. New York: Wiley.
- Altschuld, J. W. and Witkin, B. R. (2000). *From Needs Assessment to Action: Translating Needs into Solution Strategies*. Thousand Oaks, CA: Sage.
- Chauvin, S. W., Anderson, A. C., and Bowdish, B. E. (2001). Assessing the professional development needs of public health professionals. *Journal of Public Health Management Practice* 7(4), 23–37.
- Gupta, K., Sleezer, C. M., and Russ-Eft, D. F. (2007). *A Practical Guide to Needs Assessment*. San Francisco, CA: Pfeiffer.
- US Department of Education (2003). *TIMSS (Trends in International Mathematics and Science Study)*. Washington, DC: National Center for Education Statistics.
- US Department of Health and Human Services (2006). *National Health and Nutrition Examination Survey (NHANES)*. Hyattsville, MD: National Center for Health Statistics.
- Wedman, J. (2007). *The Performance Pyramid: Needs Assessment Made Easy*. Columbia, MO: University of Missouri.
- Witkin, R., Leigh, D., Platt, W., and Kaufman, R. (1998). Needs assessment – a digest review, and comparison of needs assessment literature. <http://home.gwu.edu/~rwatkins/articles/NAdigest2.pdf>.

The Purpose of Educational Evaluation

S Mathison, University of British Columbia, Vancouver, BC, Canada

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Accountability – Demonstrating that some predetermined level of performance has been achieved with the understanding that meeting expectations results in rewards while not meeting expectations results in sanctions.

Amelioration – Making something better.

Formative – Evaluation done for the purpose of improvement focusing on implementation and process, and which is conducted while the evaluand is ongoing or in the development stage.

Summative – Evaluation done for the purpose of accountability focusing on outcomes and effects, and which is conducted when the evaluand is completed or in its final form.

- States/governments are accountable for providing adequate and equitably distributed resources for education, and certification and other processes to insure educators are properly prepared.
- School districts/units are accountable for their policies, the fair distribution of resources, and for being responsive to their constituents.
- Schools or universities are accountable for a fair internal distribution of resources, promoting continual improvement, and being responsive to their constituents.
- Teachers are accountable for identifying and meeting the needs of students.

The nature of evaluation of accountability can be understood by looking at three contexts: learning, interventions, and organizations.

Two Main Purposes of Educational Evaluation

In education, evaluation serves two primary purposes: accountability and amelioration; the latter purpose is sometimes divided into development and knowledge generation. Both of these purposes are relevant at multiple levels in education, from individual learning to programs to complete educational systems.

Accountability

While accountability can mean more than evaluation, evaluation is always necessary for accountability. In education, individuals, learners, teachers, administrators, specific interventions, organizations, and states may be held to account by someone for demonstrating that something has been accomplished in an explicitly specified way. Evaluation as accountability typically implies there are rewards or sanctions depending on whether the level of performance has met preset standards, although there is considerable variation in whether rewards and sanctions are direct and severe (withdrawal of funding) or indirect and modest (public embarrassment). Accountability is most often associated with summative evaluation, or evaluation that focuses on outcomes and effects, and which is conducted when the evaluand is completed or in its final form.

At different levels the foci for evaluation change and are often expressed as follows:

Learning

Often referred to as summative assessment or assessment of learning, this is evaluation that occurs at the end of a period of instruction and often results in a grade that summarizes what a learner knows. The period of instruction may be a unit of study, a project or paper, a complete course, or even a larger timeframe like fourth grade. Evaluation of learning in this context might take the form of quizzes, chapter tests, end of term examinations, or government mandated standardized tests. Evaluations of learning done for accountability purposes do not play a role in learning or teaching as they occur after the completion of educational interventions and are meant to label the level of performance or achievement.

Performance measurement is the term that typically captures a more general accountability for learning, the sort of broad assessments of educational attainment such as international tests of achievement like the Program for International Student Assessment (PISA) or the US National Assessment of Educational Progress (NAEP). These evaluations are meant to provide information on student outcomes to make judgments about the quality of education at local, national, and international levels. In addition, these evaluations often draw attention to differential effects, such as the achievement gap between social classes or racial groups.

While evaluations of learning that are done for accountability purposes are often equated with summative evaluation, this is accurate only for the specific learning context and participants. Such evaluations can be used for formative

purposes, but only in the sense of the improvement of similar, but future teaching and learning events (Figure 1).

Interventions

Evaluations of interventions may focus on accountability of different kinds – accountability for implementation, process, and long-term outcomes.

Educational interventions (such as curriculum, pedagogical strategies, or processes) are often developed with the expectation that an educational problem will be resolved with a prescribe set of activities implemented under certain conditions. These prescribed activities and conditions are based on a theoretical or logical understanding of needs and effective means for reaching outcomes. But for those outcomes to occur, there is a presumption that the educational intervention must be implemented with fidelity, that is, as it is prescribed. Evaluation that focuses on the degree of implementation is one form of accountability. This is sometimes referred to as quality assurance.

Accountability evaluation may also focus on process, what is often referred to as the black box of interventions, the ways in which resources come together with teaching and learning activities and create an educational process. A process evaluation focuses on more than implementation fidelity and includes a description of how the intervention was implemented, who was involved and received educational services, and examines internal and external factors that have an impact on the intervention. A process evaluation often focuses on untangling differential implementation and effects, such as what components of an intervention work effectively for subgroups of students.

Interventions often have long-term intended outcomes that evaluation may focus on discerning. An outcome evaluation is intended to determine the extent to which

outcomes occur and to establish that the changes are significantly attributable to the intervention. Outcomes are typically attitudes, knowledge, behaviors, or skills that programs aim to positively influence. Examples of typical educational outcomes are academic achievement, school completion, conflict resolution, employment, civic leadership, social development, and so on. An outcome evaluation of a peer mediation program in elementary schools, for example, might focus on determining the extent to which elementary age students are able to resolve conflict for themselves and that it was participation in the peer mediation program that contributed significantly to those conflict resolution skills. Another example is an evaluation of the extent to which a simulation-driven first-year college biology class leads to greater student control of learning and tolerance of uncertainty. While many interventions have such intentions, the multiplicity of intervening factors that may effect outcomes make this kind of accountability focused evaluation challenging (Table 1).

Organizations

Educational evaluation may focus on accountability at an organizational level, that is, a school, school district, college, or any organization with an educational mission. Such accountability may take the individual as the level of data collection, but focuses on the aggregate. Currently in the United States the requirements of the No Child Left Behind legislation require the testing of all children in third through eighth grades, and accountability occurs at a number of levels, but ultimately it is schools that are judged to be successful or in need of improvement. In both K-12 and postsecondary education, accreditation is an example of accountability evaluation at the organizational level. Educational institutions and individual programs are required to demonstrate they have met standards set by recognized external professional bodies. Successful accreditation may directly effect getting operational funding and attracting students, while being unsuccessful might mean loss of funding, inability to recruit students, and/or an inability to attract research grants. This sort of accountability evaluation requires a demonstration of and justification of performance to external audiences, and indeed is often required by external authorities such as inspectorates and professional accrediting agencies. It is evaluation that can be described as top-down and is often motivated by political agendas, at the local, national, and international levels (Table 2).

Amelioration

Evaluation can also play a significant role in planning and development and is manifest in evaluation that focuses on needs assessment, improvement, learning and appreciation, building self-evaluation capabilities, determining fidelity of

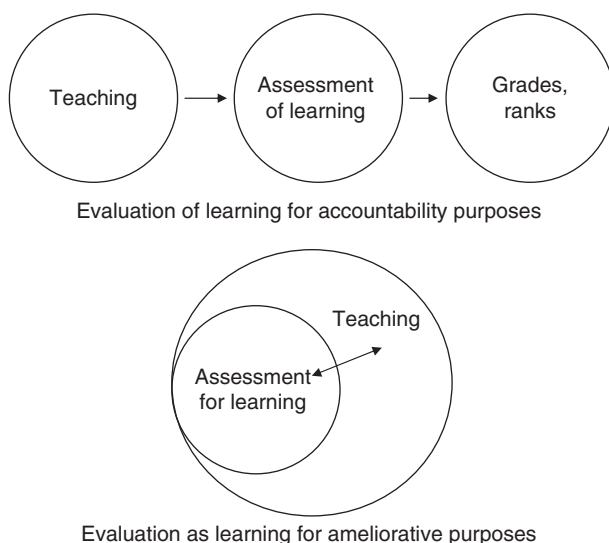


Figure 1 Relationship between evaluating learning for accountability and ameliorative purposes.

Table 1 Purposes of evaluation in education by level of evaluand

	<i>Amelioration</i>	<i>Accountability</i>
Learning	Occur during the teaching learning process Focus on identifying the need for remediation Example Teacher analysis of error patterns in student work to identify topics for reteaching	Occur at the end of a teaching–learning event Focus on identifying knowledge, skills, or attitudes acquired Example End of course grades
Intervention	Support planning and implementation of specific educational interventions or components Focus on implementation, process, and short-term outcomes Example Improve mathematics achievement by creating relevant learning activities	Support resource allocation or selection decisions about educational interventions Focus on outcomes and effects Example Full-scale adoption of a reading program that has been pilot tested
Organization	Support organizational improvement Focus on strategy planning, organizational learning, process, short-term outcomes Example Identify strategies for improving human resource development in the use of technology	Demonstrate financial responsibility and mission accomplishment Focus on outcomes and effects Example Accreditation of a university

Table 2 Characteristics of evaluation for amelioration and accountability

	<i>Amelioration</i>	<i>Accountability</i>
Purpose	To improve, to learn, to generate insight about educational issues, to promote ongoing evaluation, to strength programs and organizations	To provide judgments about outcomes, effectiveness, value, to determine costs, to assess efficiency
Character of evaluation	Internal, formative, diagnostic	External, summative
Audience	Teachers, learners, curriculum developers, parents, policy shaping community	Policy makers, educational administrators, funders, public
Role of evaluator	Often internal to organizations; collaborative, evaluator as critical friend	Often external, distant, independent
Reporting	Informal	Formal

implementation, making informed decisions, and contributing to enlightened discussions. Most ameliorative evaluation efforts give precedence to finding solutions and making improvements in specific contexts and in the short term. Ameliorative evaluation speaks most often to internal audiences such as administrators, educational planners, and teachers or other education professionals.

Often, evaluation that is ameliorative is equated with formative evaluation, that is, evaluation done for the purpose of improvement focusing on implementation and

process, and which is conducted while the evaluand is ongoing or in the development stage. Some evaluation theorists, for example, Eleanor Chelimsky and William Shadish, separate this ameliorative purpose into development and knowledge generation.

Needs assessment

Needs assessment is a key strategy in evaluation that aims to set priorities for improvement and allocation of resources. It is a strategy to identify issues and to develop plans for change by systematically analyzing what is and considering what should be. Evaluation is used to identify gaps between what is desirable or expected and what is actually happening, prioritizing which gaps to be bridged through a determination of where inevitably limited resources should be expended. In educational contexts, a school district might do a needs assessment to determine what is expected of schools by its largely immigrant, English as a second language community, or a statewide university system might use needs assessment to identify community goals and make resource allocations in order to meet those goals.

Improvement and learning

In the context of learning, this type of evaluation is sometimes referred to as assessment for learning or formative assessment. Evaluations of learning that are meant to ameliorate occur during the learning process and provide information that cycles back into the teaching–learning connection so that both the quality of teaching and learning can be enhanced. The idea is that assessment for learning will increase learners' motivation by genuinely engaging them in making judgments about their knowledge and skills. Typical evaluation strategies include questioning,

feedback by comments, peer assessment, self-assessment, and analysis of errors. **Figure 1** illustrates that evaluation in this ameliorative sense is embedded within the teaching and learning activity rather than being separate from and following the teaching/learning activity.

Evaluations are meant to support improvement and learning at programmatic and organizational levels. During the development, planning and piloting of programs and interventions evaluation is a means for providing information about implementation, inputs, processes, and short-term outcomes that feeds back into a planning and development cycle. Such evaluation may be formal or informal, and evaluators and program developers work closely to monitor what is happening and make adjustments as the program is ongoing. Evaluation for this purpose is often a bottom-up enterprise in which educational programs, organizations, and systems are motivated by a desire to identify areas in need of improvement and the strategies for improvement.

The school self-evaluation movement in Europe is a good example of ameliorative evaluation that focuses on improvement and learning at an organizational level. While school self-evaluation focuses on the same elements as other evaluations (outcomes, process, and context), the definition of what the evaluation will focus on and the use of the results is local and specifically targeted to judging how that school is doing with an eye to improving something in that school. Responsibility for doing evaluation is vested in school evaluation teams, which may include students, parents, teachers, school administrators, and other community members. School self-evaluation is a particular case of evaluation capacity building.

Formative evaluations focused on improvement and learning may, however, also serve accountability purposes in evaluation, such as in the case of accreditation where the self-evaluation is largely prescribed by an external agency and will ultimately be reviewed by external evaluators or inspectors.

Evaluation capacity building

Evaluation capacity building (ECB) intends not just to do evaluation for others but instead to build a system of knowledge, process, and practice within an organization so that quality evaluation becomes ordinary and ongoing. Emphasizing ECB means evaluators support the development and sustaining mechanisms for those within an organization to learn about and do evaluation for themselves. For example, ECB within a college or university might focus on the implementation of changes to a student rating of teaching survey, distribution of information about evaluation-related professional development opportunities for academic department heads, reviewing academic department self-reviews, and so on.

Contributing to deliberation about educational problems

Evaluation also serves a more general ameliorative purpose of creating and contributing to public discourse about education, what Lee Cronbach called the policy-shaping community. Cronbach highlighted the contribution evaluation makes by enriching discussions within the polity about important social and educational problems. Evaluations help in clarifying what the issues are, in analyzing programmatic assumptions, and in revealing the political, social, and organizational contexts that influence both problem definition and solution. In this way, evaluations contribute to a broad deliberative discourse about how to solve problems and improve educational and social outcomes.

A useful illustration of this idea is the cumulative evaluative effort of the widely adopted Drug Abuse Resistance Education (DARE) program. Because the DARE program is so ubiquitous, in virtually all schools in the United States and Canada as well as many other countries, many evaluations of the program have been conducted. The evaluation results are mixed – early evaluations of DARE concluded it was effective but more recent evaluations suggest the program does little to decrease the incidence of drug use. Over time and with the investment of considerable programmatic and evaluative resources, evaluations of DARE have resulted in any significant changes in the program. Evaluations have contributed in a broad sense to the public discourse about why and how to discourage youth from drug use.

Evaluation Within Socio-Political Contexts

Accountability and amelioration co-occur as reasons for doing evaluation in education, but often the contemporary socio-political context gives precedence to one or the other. Since the early 1980s, educational evaluation has been preoccupied with the accountability purpose, but in a particular way. Before the current era, which emphasizes regulatory accountability, professional accountability was common and considered trustworthy. However, accountability in education, much like in other domains of public and private life, became too important to leave to educators, and local, state, and sometimes multi-state authorities have assumed a more significant role.

Public confidence in institutions to meet individual and collective needs has been eroded. While once there was confidence that professionals (e.g., teachers, doctors, and accountants) and leaders (e.g., politicians and ministers) had deeply held and ethical commitments to doing the right thing, too many transgressions have led to public skepticism. This skepticism has ushered in what has been called the era of new public management.

Until the late 1970s and early 1980s, accountability was more an expectation than a process. In general, what is referred to as professional accountability prevailed – the expectation that professional judgment and action was informed by good reasons and an acceptance of the authority of those providing the reasons. Professional accountability is self-regulation by a group of professionals. In the past, educators experienced more autonomy in determining how much learning had occurred, what quality teaching was, if schools were doing a good job, and whether organizational missions were appropriately being met.

While scandals are often cited as the reason for a shift away from professional accountability, in fact, there are examples of what are now commonly and widely held senses of inadequate or inappropriate professional knowledge. Historically, both disagreements about whether the public good has been served by professional accountability and incidences of professional misconduct, including but not limited to malfeasance, have led to a disintegration of professional accountability and the rise of a regulatory accountability. Regulations begin to take the place of educators' professional judgment.

Regulatory accountability is a shift to determinations about what professionals should do that is external to the institutions within which they work, rather than being decided by the professionals working within those institutions. This form of accountability vests authority in governments, but not by simply putting the government in charge. Rather, governments are agents that support free markets by creating the conditions that allow markets to operate to maximize effectiveness, profits, and efficiency. Government regulations are the primary means for creating these conditions. Regulatory accountability is facilitated by a global neoliberalism that conceptualized education less as process and more as products or outcomes, a shift in emphasis on competing purposes of education. Vocationalism and democratic citizenship have long competed as the main purpose of education. Taylorism came to US schools early in the twentieth century, an approach that emphasized efficiency of production but developed alongside progressivism's focus on the effectiveness of schools to promote democratic principles. Education has always been conceived as an institution that serves the public interest by preparing young people for work and citizenship, promoting a common culture (especially in nations of immigrants), and reducing race, ethnic, and class inequalities. What is different about these purposes is whether they are conceived in the interest of individuals or the collective, public interest. The current emphasis is on the private and economic benefits (vocational purpose or schooling for the market that serves individual and corporate economic interests), rather than the public benefits (schooling for democratic citizenship with attention to mediating special interests for the common good).

Neoliberal values currently capture the public attention regarding the purpose of schooling and, consequently, how education is evaluated. As politicians, corporate CEOs, and free marketers continue to dominate the public rhetoric about quality in education, so too will the strategies for educational evaluation reflect those values.

The Future Purposes of Evaluation

The fundamental purposes of evaluation will not change; educational evaluation will continue to provide the means for accountability and amelioration through development and knowledge generation. The current era of regulatory accountability emphasizes the accountability purpose in a particular way, but educational evaluators are developing alternate views of accountability that combine both purposes of evaluation. Broadly, these approaches to educational evaluation might be called pluralistic democratic accountability.

Pluralistic democratic accountability is an approach that asks educational stakeholders to enter a compact of mutual, collective responsibility for education and schooling. Such an approach to accountability places far greater emphasis on local context and internal accountability, that is, mutual responsibility within contexts and communities where social actors can and do know each other, and in which there is real participation in education. It is an approach that focuses on learning and improvement, rather than blaming and punishing.

The accountability and ameliorative purposes of evaluation may be perceived as antithetical to each other, but there is no logical reason for this. Indeed, evaluation that is done to improve can be used for accountability and accountability-oriented evaluation can be used to make plans for improvement. Pluralist democratic accountability makes the complementarity of these purposes explicit by promoting evaluation that guides improvement of education (through involvement of multiple stakeholders, school self-reviews, and local analysis of student achievement data) and demonstrates accountability to various publics (through organizational report cards and reporting to a broad range of education's constituencies).

See also: Educational Evaluation: Concepts, Practice, and Future Directions; Evaluating Education in Three Policy Eras; Evaluating Schools as Learning Communities; Evaluation and Accountability; Evaluation of Teacher Quality and Practice; Formative Assessment in Teacher Education and Teacher Professional Development; Principal Evaluation: Concepts, Needs and Realities; School Based Evaluation: Purposes, Protocols and Processes; School Inspection/External School Evaluation; The Role of Stakeholders in Educational Evaluation.

Further Reading

- Chelimsky, E. and Shadish, W. R. (1997). *Evaluation for the 21st Century*, 2nd edn. Thousand Oaks, CA: Sage.
- Compton, D. W., Baizerman, M., and Stockdill, S. H. (2002). The art, craft, and science of evaluation capacity building. *New Directions for Evaluation* **93**.
- Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record* **64**, 672–683.
- Cronbach, L. J., Ambron, S. R., Donrbusch, S. M., et al. (1980). *Toward Reform of Program Evaluation*. San Francisco, CA: Jossey Bass.
- MacBeath, J. and McGlynn, A. (2002). *Self-Evaluation: What's in It for Schools?* London: Routledge.
- Mathison, S. (2008). Educational assessment's role in the neo-liberal, neo-conservative agenda. In Gabbard, D. (ed.) *Knowledge and Power in the Global Economy: The Effects of School Reform in a Neoliberal/Neoconservative Age*, 2nd edn, pp 161–169. Mahwah, NJ: Erlbaum.
- Mathison, S. (2009a). Public good and private interest: A history of educational evaluation. In Ayers, W. C., Quinn, T., and Omatoso Stovall, D. (eds.) *The Handbook of Social Justice in Education*, pp 5–14. London: Routledge.
- Mathison, S. (2009b). Educational evaluation and the social good. In Cousins, C. B. and Ryan, K. (eds.) *The SAGE International Handbook of Educational Evaluation*, pp 525–538. Thousand Oaks, CA: Sage.
- Mathison, S. and Ross, E. W. (eds.) (2008). *The Limits of Standards Based Reform and Testing*. New York: Teachers College Press.
- Nevo, D. (1983). The conceptualization of educational evaluation: An analytical review of the literature. *Review of Educational Research* **53**(1), 117–128.
- William, D. and Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal* **22**(5), 537–548.
- Witkin, B. R. and Altschuld, J. W. (1995). *Planning and Conducting Needs Assessments: A Practical Guide*. Thousand Oaks, CA: Sage.

Understanding Approaches to Evaluation

A J Love, Independent Evaluation Consultant, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Introduction

During the last two decades policymakers, grant-making bodies, educators, and the public around the globe increasingly have turned to educational evaluation to achieve higher levels of school and teacher effectiveness, guide the reform of education systems, and foster greater accountability. To achieve these aims, a seemingly endless stream of evaluations is taking place at all levels in the education sector. There are extensive in-depth evaluations of national programs and materials, such as the Student Mentoring Program and the Even Start Family Literacy Programs in the United States, and Early Reading First initiatives in England. There are evaluations of student achievement at the state and provincial level, for example, the Illinois Standards Achievement Test and the work of the Education Quality and Accountability Office in Ontario, Canada. There are evaluations of student performance at the national level, such as the US National Assessment of Educational Progress (NAEP) and Australia's National Assessment Program, and at the international level, for example, the Organization for Economic Cooperation and Development (OECD)'s Program for International Student Assessment (PISA). There are also comparative evaluations of the effectiveness of school systems across several regions, such as those conducted across Europe and in the Asia-Pacific region.

In addition, school administrators and teachers worldwide are using internal evaluations (self-evaluations) to drive school improvements and create innovative learning environments. Spurred by greater awareness of the importance of factors outside of the classroom in achieving student and school success, grant-making organizations are demanding evaluations of the relative contributions of the diverse support programs they fund, such as the Healthy Minds School Breakfast Program and Homework Clubs Programs.

These examples reveal a wide variety purposes for educational evaluation. As a partial list, the purposes for educational evaluation include assessing the quality and effectiveness of educational curricula, programs, and materials; accrediting schools and ensuring that they meet minimum standards; assisting parents and students in selecting schools; improving accountability by monitoring student performance at local, state/provincial, and national levels; assessing the effectiveness of school systems across countries and regions; guiding school-reform efforts; and empowering school administrators and teachers to learn

from evaluation as a way of improving their teaching effectiveness and school performance.

If educational evaluation had only a single purpose, then perhaps a single approach to evaluation would be possible. But this is not the situation. Educational evaluation is an applied activity that takes place in the turbulent environments of real-world settings. Evaluators commonly encounter multiple and competing purposes for an evaluation, numerous stakeholders with contradictory information needs, limited time and budget for evaluation, and conflicting views about evaluation methodology. The following section presents several conceptual frameworks that evaluators use to manage the complexity of the evaluation field, select relevant evaluation approaches, and successfully conduct evaluations in challenging practice settings.

Selected Frameworks for Understanding Approaches to Evaluation

In the evaluation field, the term model is used broadly to denote a conceptual perspective or structure, rather than more narrowly in the sense of a scientific or theoretical model. Many evaluators prefer the term approach, however, because it denotes a way of viewing evaluation, its values and purposes, the roles of evaluators, and the recommended steps for designing and implementing the evaluations. In turn, most approaches to evaluation identify relevant methods of data collection and analysis, such as needs assessment or cost-effectiveness analysis. In general, an evaluation approach imparts coherence to the evaluation process and serves as a guide for evaluation training and practice. Given the diversity of purposes for educational evaluation and its varied social and political contexts, literally dozens – if not hundreds – of approaches to evaluation exist today that reflect the views of thoughtful evaluators about the theory and practice of evaluation. To navigate through this labyrinth and manage its complexity, this article concentrates on four major frameworks for organizing and understanding the multiplicity of approaches in the evaluation field.

These four frameworks meet the following criteria: (1) the developers of each framework are highly experienced educational evaluators, even though their frameworks apply beyond the confines of the education field; (2) the developers refined and tested their frameworks over a period of decades and their frameworks have stood

the test of both time and peer scrutiny; (3) the frameworks relate to the entire evaluation field rather than one single approach or methodology; (4) there are widely available books, articles, and reference materials available for each framework that can supply the interested reader with the detailed information not possible within the limitations of this article; and (5) universities and training programs use these four frameworks, alone or in combination, to educate professional program evaluators, graduate students, and teachers about program evaluation and give reliable direction to seasoned evaluators striving to improve their evaluation practice. The four frameworks presented here are:

- chronological framework;
- alternative approaches framework;
- evaluation forms and approaches framework; and
- framework based on a meta-evaluation of evaluation approaches.

In terms of the organizing principles of each framework and ease of comprehension by novice evaluators and those new to the evaluation field, these four frameworks follow a general progression from the simpler to the more complex.

Chronological Framework

The chronological framework is arguably one of the most commonly used frameworks for training novice evaluators in the fundamentals of program evaluation and in selecting appropriate evaluation approaches and methods. The chronological framework is based on the notion that the sequence of events (i.e., chronology) during the developmental life cycle of a program strongly influences the primary purposes for an evaluation, the focus of the specific evaluation questions addressed by the evaluation, and the choice of evaluation approaches and methods required for answering those evaluation questions. The chronological framework also reinforces the idea that program evaluation should be used throughout the program life cycle and not just at the end.

Although elements of the chronological framework appeared in the late 1970s, it received closer attention after its description in the evaluation primer, *Thinking about Program Evaluation* (Berk and Rossi, 1998), and its integration into successive editions of the widely adopted evaluation text, *Evaluation: A Systematic Approach* (Rossi et al., 2004). In addition to the university courses that use *Evaluation: A Systematic Approach*, the chronological framework has made an important contribution to the structure of several national training programs in evaluation, such as those supplied by the Canadian Evaluation Society, the Japan Evaluation Society (Nagao et al., 2005), and the basic course in program evaluation offered by The Evaluators' Institute. The chronological framework is also congruent

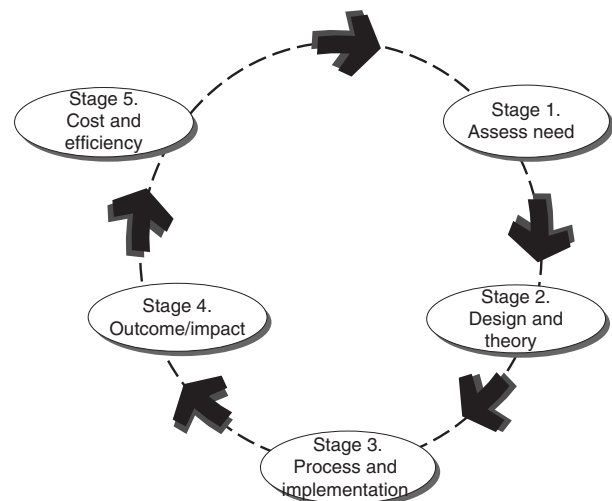


Figure 1 Sequence of events in program life cycle.

with the standards developed by the Joint Committee on Standards for Educational Evaluation (1994).

For a better understanding of the chronological framework, **Figure 1** portrays the major sequence of events in the program life cycle, arranged in the fashion of a clock. **Figure 1** shows that stage 1 in the program life cycle is assessing the need for the program; stage 2 is planning the program design and clarifying the program theory; stage 3 is establishing the processes of program delivery and implementing the program; stage 4 is the generation of outcomes by a fully implemented program; and stage 5 is weighing the cost and benefits of the program.

As most programs follow a similar developmental sequence, the chronological framework is intuitive and readily understood by managers (administrators) and staff (teachers and other staff). With assistance from an evaluator, managers and staff can quickly pinpoint their program's location in the program life cycle. Knowing the location in the program life cycle helps clarify the purposes for evaluation and identifies the relevant evaluation questions. For example, a typical purpose for evaluating a program at stage 3 (implementation stage) is to assess the strengths and weaknesses of the program with an eye to improving its operations. Relevant questions at stage 3 include those about the process of program delivery, whereas a program at stage 4 usually raises questions about program outcomes. In short, applying the chronological framework allows an evaluator to rapidly focus the evaluation by using this shortcut to confirm its purpose and identify the key evaluation questions.

The chronological framework points not only to specific evaluation questions but also suggests appropriate evaluation approaches for each stage. For example, programs at stages 1 through 3 in the program life cycle frequently require a formative approach aimed at adjusting the program theory to the local context and fine

Table 1 Stage of program life cycle matched to major evaluation methods

Stage of program life cycle	Evaluation methods
Stage 1. Need for program	Need identification Needs assessment
Stage 2. Design and theory	Evaluability assessment Logic modeling Describing and assessing program theory
Stage 3. Process and implementation	Program process monitoring (service utilization and organizational functions) Program process evaluation
Stage 4. Outcome/Impact	Identifying intended and unintended outcomes Measuring/monitoring program outcomes Impact assessment Meta-analysis
Stage 5. Cost and efficiency	Efficiency analysis Cost-benefit analysis Cost-effectiveness analysis

From Rossi, P., Freeman, H., and Lipsey, M. (2004). *Evaluation: A Systematic Approach*, 7th edn. Thousand Oaks, CA: Sage.

tuning the program implementation, whereas programs at stages 4 and 5 call for a summative approach that assesses the outcomes and cost effectiveness of the program. The chronological framework also suggests relevant evaluation methods. For example, a program at stage 2 (design and planning stage) usually benefits from specific methods such as evaluability assessment or clarification of the program logic. **Table 1** illustrates the relationship between program stage and choice of evaluation methods.

In short, the chronological framework has clear advantages: it portrays evaluation holistically throughout the stages of program development; introduces novice evaluators to important terms and concepts; and the model provides clear direction about major evaluation questions, appropriate evaluation approaches, and relevant data-collection methods. On the other hand, a central limitation of the chronological framework is that it tends to oversimplify the process of program development and evaluation. The framework implies a systematic or step-by-step process that may not reflect the realities of real-world settings. As evaluators gain experience and gain the capacity to deal with complex and nonlinear evaluations, they may find that one of the other evaluation frameworks described in the following sections more adequately meets their needs.

Alternative-Approaches Framework

The alternative-approaches framework, as described in *Program Evaluation: Alternative Approaches and Practical*

Guidelines (Worthen *et al.*, 2004), imparts order to complex evaluation situations by clustering a selection of the more popular and influential evaluation approaches into five alternative approaches to evaluation:

- Objectives-oriented approaches;
- Management-oriented approaches;
- Expertise-oriented approaches;
- Consumer-oriented approaches; and
- Naturalistic and participant-oriented approaches.

The alternative-approaches framework takes the position that there are many approaches to evaluation and each can be useful depending on the purpose of the evaluation and other factors, such as the organizational context. In developing the five categories of alternative evaluation approaches, this framework considers several important dimensions of the alternative-evaluation approaches, such as epistemological differences (objectivism vs. subjectivism), value differences (utilitarian vs. intuitionist-pluralist), and methodological differences (quantitative vs. qualitative).

The resulting five categories of evaluation approaches are loosely ranged across these dimensions. The categories reflect both the major questions addressed by the approach and its key features. For example, an objectives-oriented approach to school achievement is likely to address questions on the attainment of learning objectives and assess attainment using standardized tests. In contrast, a naturalistic and participant-oriented approach is more prone to evaluate school achievement in terms of questions identified jointly by teachers and students and assess the achievement using qualitative methods, such as portfolios. The following sections describe each alternative approach briefly; and **Table 2** summarizes the major purposes and chief characteristics of the five alternative approaches.

Objectives-oriented approaches

The essence of the objectives-oriented approaches is specifying the goals and objectives of a program and then assessing how well program participants attain them. One of the earliest approaches to educational evaluation assessed student behaviors against the intended educational objectives by using impartial and systematic measures. The differences between the intended and the actual achievement of objectives were used to improve program performance. There are several objectives-based approaches, for example, discrepancy evaluation that measures the discrepancy or difference between program standards and actual program performance. Critics of the objectives-based approaches contend that they promote an inflexible approach to evaluation and lack clear criteria to judge the worth of a program. One reaction to the limitations of objectives-based evaluation is goal-free evaluation, that is, evaluation that focuses on the actual outcomes of a program and unanticipated results, rather than the intended goals/objectives alone.

Table 2 Purposes and characteristics of alternative approaches

<i>Approaches</i>	<i>Major purposes</i>	<i>Major characteristics</i>
Objectives-oriented	Measure extent goals and objectives are achieved	Measurable goals and objectives; assess gaps; apply scientific tools
Management-oriented	Supply useful information for decision making	Support data-based decision making; employ systems analysis; evaluate all stages of program development
Consumer-oriented	Aid decisions about products or purposes	Product testing; apply checklists; render judgments using established criteria
Expertise-oriented	Render professional judgments of quality and achievement of standards	Professional expertise and judgment; apply consensus standards; use of formal and informal review panels
Participant-oriented	Respond to stakeholder information needs; describe program activities and processes	Involve evaluation users; apply naturalistic and mixed methods; portray multiple perspectives of program realities

From Worthen, B., Sanders, J., and Fitzpatrick, J. (2004). *Program Evaluation: Alternative Approaches and Practical Guidelines*, 3rd edn. New York: Pearson Education.

Management-oriented approaches

Rather than assessing goals and objectives, management-oriented approaches concentrate on identifying and meeting the evaluation-information needs of program managers or administrators. The context, input, process, and product (CIPP) approach is an example of an important management-oriented approach to program evaluation in the education field. In these approaches, the evaluator works closely with program managers to identify key decisions and then selects the appropriate evaluation methods to inform those decisions. The management-oriented approaches place primacy on the role of program managers and the importance of a systems approach that values both program context (e.g., classroom environment) and processes (e.g., following lesson plans) in generating program outcomes. Management-oriented approaches also view program managers and program evaluators as having complementary roles and working cooperatively to overcome obstacles to program success.

Some strengths of management-oriented approaches include a clear focus on evaluation questions relevant to managers, application of evaluation throughout all phases of program planning and implementation, and increased use of evaluation information for both learning and accountability. On the other hand, management-oriented approaches experience limitations unless there is strong management commitment to the evaluation process, acceptance of the evaluation interests of managers over the evaluation needs and concerns of other stakeholders, clear formulation of decision alternatives and relevant evaluation questions during the evaluation planning process, and specific mechanisms for ensuring that evaluation findings inform decision making and action planning.

Consumer-oriented approaches

Every year billions of dollars are spent worldwide on educational products and services – all without adequate information about their effectiveness. Consumer-oriented

approaches to evaluation focus on providing consumers (e.g., governments, school authorities, and the public) with dependable evaluation information about competing products and services offered in the education marketplace. Some examples include independent evaluation of learning programs (e.g., tutoring, summer programs abroad), educational products (e.g., literacy instructional materials, technology equipment, and software), and educational services (e.g., in-service training, workshops).

The strengths of the consumer-oriented approach include well-established criteria and checklists for evaluating educational products and services. For nearly half a century, the use of these tools has greatly improved the knowledge of consumers and spurred greater accountability for the effectiveness of educational goods and services. Potential disadvantages of the consumer-oriented approaches include the cost and time to collect credible evaluation information, as well as discourage local solutions because of the effort involved to assess their performance.

Expertise-oriented approaches

These approaches place emphasis on the judgment of experts in rendering an evaluative judgment about the program. Evaluators with the requisite professional expertise collect information about a program through a variety of methods, such as reviews of documents, curricula, student projects, and classroom observation. Then they apply their expertise to make an evaluative judgment about the performance of the program.

Expertise-oriented approaches may be formal or informal. Formal expertise-oriented approaches include professional review systems, such as the national and regional accreditation reviews of school systems and teacher training programs (e.g., National Council for the Accreditation of Teacher Education) that have published standards agreed upon by the profession, clear criteria for rating each standard, site review by experts to validate the attainment of the standards, review of the site-review

report by a distinguished accreditation committee, an appeal procedure, and a final accreditation decision by the accrediting organization. Informal expertise-oriented approaches include *ad hoc* panel reviews, funding-agency-review panels, and blue-ribbon panels (e.g., Rosenberg Commission in Maryland and the Wessell Commission in New York) with members selected for their experience, expertise, and credibility.

The strength of the expertise-oriented approaches is that they draw upon the years of training, wisdom, and insight of professionals to examine a program with the depth and discernment not available through other approaches. On the other hand, this is also seen as the weakness of these approaches – the subjective and possibly narrow judgment of experts is not a substitute for objective criteria and the application of systematic analysis. Furthermore, without a strong mandate and formal structure, critics of these approaches contend that is relatively easy to ignore the findings of expertise-oriented evaluations.

Participant-oriented approaches

In general, participant-oriented approaches invite the involvement of a broad range of stakeholders to identify evaluation questions, select data sources and data-collection tools, and in some approaches, assist in the interpretation of the findings. These approaches view the program managers, staff, and participants as important stakeholders and seek to meet their evaluation needs, as well as the needs of stakeholders external to the program (e.g., government departments or community partners). The primary task for the evaluator is to balance the values and perspectives of the various stakeholders and to protect the pluralism of views when gathering data and reporting evaluation findings. In practice, this means that participant-oriented evaluations are more likely to draw data from multiple sources, employ both qualitative and quantitative data-collection methods, and report evaluation findings in ways that give voice to multiple viewpoints, such as oral briefings, multimedia presentations, and public meetings.

A wide variety of participant-oriented approaches have been influential in educational evaluation, such as the stakeholder-based and responsive-evaluation approaches that urge program evaluators to respond to the participants' concerns and need for evaluation information; the utilization-focused evaluation approach that stresses that an evaluation must be useful for its audiences; participatory evaluation approaches that involve program managers and staff directly in the evaluation process; the empowerment evaluation approaches that give program participants the tools to evaluate their programs and learn from their experiences; and the deliberative democratic evaluation approaches that support transparency and democratic discussion about programs and their outcomes.

Critics of participant-oriented approaches argue that involving multiple stakeholders in the process of designing, implementing, and interpreting the evaluation findings tends to make the evaluation too complex and can present a bewildering array of contradictory viewpoints. The net result is a murky description of the program rather than a clear evaluative judgment about its worth and merit.

In summary, the alternative-approaches framework acknowledges the diversity of approaches to evaluation and the importance of considering differences in values, philosophy, and methodology when selecting an evaluation approach within a specific organizational context. While the chronological framework adopts a sequential viewpoint that matches evaluation approaches and methods to specific stages in the program life cycle, the alternative-approaches framework arranges evaluation approaches along a continuum that reflects their similarities and differences. Although this strategy reflects more accurately the range of evaluation approaches, to apply the alternative-approaches framework well in different settings requires a firm grounding in the theories of evaluation, assessing the evaluation setting, and an understanding of a broad range of evaluation approaches.

Forms-and-Approaches Framework

John M. Owen, who served as director of the Centre for Program Evaluation at the University of Melbourne's Faculty of Education, developed the forms-and-approaches framework. The latest edition of *Program Evaluation: Forms and Approaches* (Owen, 2007), presents an up to date and comprehensive framework for designing program evaluations. The forms-and-approaches framework depends on an analysis of the evaluation situation, rather than arranging evaluation approaches sequentially according to program life cycle, as portrayed in the chronological framework, or clustering approaches according to common characteristics, as found in the alternative-approaches framework. The analysis of the evaluation situation allows the evaluator to apply a framework based on five evaluation forms that provide the conceptual mechanism for organizing different approaches to program evaluation.

Evaluation forms

The five evaluation forms are the proactive form, clarificative form, interactive form, monitoring form, and impact form. Each evaluation form relates to a specific evaluation purpose and focuses on a common set of issues. Here is a brief description of the five evaluation forms:

- the proactive form provides input to decisions about how to design and develop a program before the program enters the planning stage;

- the clarificative form clarifies the theory or logic of the program to assess program feasibility, identify design problems, and strengthen its design;
- the interactive form supplies information about delivery or implementation of a new or rapidly changing program to strengthen the program;
- the monitoring form offers regular feedback about the ongoing performance of well-established programs, often using performance indicators; and
- the impact form determines the extent and level of long-term attainment of outcomes and results.

Evaluation approaches linked to evaluation forms

As noted earlier, each evaluation form links to a specific set of evaluation approaches. In turn, the evaluation approaches for a given evaluation form link to a specific set of evaluation methods. In this way, identifying the underlying evaluation form helps the evaluator select appropriate evaluation approaches and methods.

As the use of the concept of evaluation forms is new for many evaluators, the forms-and-approaches framework

presents the details of each evaluation form in depth. It presents the rationale and focus of each evaluation form, defines the major evaluation approaches related to that evaluation form, and outlines the major methods for data collection that are linked to those approaches. Many detailed examples give life and substance to these discussions and the vast majority of the examples come from the education field. **Table 3** summarizes the major evaluation approaches and the evaluation methods that are linked to each evaluation form.

For example, the interactive form of evaluation suits situations that raise questions about program delivery and the desire to improve decision making by using evaluation to learn more about program implementation. To continue this example, according to the forms-and-approaches framework, the major approaches linked to the interactive form of evaluation include responsive evaluation, action research, quality reviews, developmental evaluation, and empowerment evaluation. In turn, these approaches use specific methods to obtain the desired evaluative information and translate it into action. These methods are the core methods shared in common among the evaluation

Table 3 Major approaches and methods linked to each evaluation form

<i>Evaluation form</i>	<i>Major evaluation approaches</i>	<i>Major evaluation methods</i>
Proactive form	Needs assessment and needs analysis Research review and evidence-based synthesis Review of best practices Benchmarking	Review documents and databases Site visits Interviews Focus groups Nominal groups Delphi method
Clarificative form	Logic modeling Evaluability assessment Accreditation	Document analysis Interviews Site visits Direct observation
Interactive form	Responsive evaluation Action research Quality review (institutional self-study) Developmental evaluation Empowerment evaluation	Intensive onsite studies Direct observation Depth interviews Self-studies
Monitoring form	Component analysis Performance measurement Systems analysis	Program Records Client and management Information systems Performance indicators Client feedback via interviews, focus groups, and/or surveys Key informant interviews
Impact form	Goal-based evaluation and Objectives-based evaluation Process-outcomes studies Needs-based evaluation Goal-free evaluation Performance audit of effectiveness and efficiency with financial and nonfinancial measures Realistic evaluation	Experimental or quasi-experimental designs Standardized tests Direct observation Qualitative exploratory and investigative inquiry

From Owen, J. (2007). *Program Evaluation: Forms and Approaches*, 3rd edn. New York: Guilford Press.

approaches included in the interactive form. The methods for the interactive form are intensive onsite studies, direct observation, depth interviews, and self-studies. The specific mix of core methods and the addition of supplementary methods (e.g., databases and focus groups) vary according to each evaluation situation.

The forms-and-approaches framework has several advantages. It provides a flexible framework that covers a wide range of purposes for educational evaluation and links them to contemporary evaluation approaches (e.g., developmental evaluation, and realistic evaluation) and methods (e.g., performance indicators, qualitative exploratory and investigative inquiry), as well as traditional ones, in a balanced and coherent manner. The forms-and-approaches framework also provides a structure for conversations about the application of evaluation approaches in hotly contested and high-stakes areas of the educational evaluation field, such as outcomes indicators, standardized testing, evidence-based practice, and the use of randomized control studies as the gold standard for impact analysis.

On the other hand, the forms-and-approaches framework requires an accurate appraisal of the evaluation situation before the evaluator can select the appropriate evaluation form that, in turn, guides the choice of evaluation approaches and methods. Assessing the situational complexities of an evaluation can be challenging for novice evaluators. For this reason, the forms-and-approaches framework is arguably more suitable for evaluators, school administrators, and teachers who already have a good working knowledge of evaluation fundamental and also grounding in basic approaches to evaluation. In such cases, the forms-and-approaches framework provides a pathway to develop a deeper understanding of the evaluation situation and to select evaluation approaches and methods that generate useful evaluative information.

Framework Based on a Meta-Evaluation of Evaluation Approaches

The final framework follows a different way of categorizing evaluation approaches than found in the other evaluation frameworks in this article. To develop their framework, Stufflebeam and Shrinkfield (2007) used a meta-evaluation (evaluation-of-evaluations) process for identifying and rigorously assessing the most frequently used approaches for program evaluation before presenting their findings in the book *Evaluation Theory, Models, and Applications*. The credibility of a framework based on a meta-evaluation depends largely on the expertise of the developers and their ability to rigorously apply extensive criteria in their analyses. Both developers are well qualified for the task: Stufflebeam led the work of the Joint Committee for Educational Evaluation Standards in developing the Program Evaluation Standards and for

nearly half a century, he served as director of the Evaluation Centers at Ohio State University and then at Western Michigan University; and Shrinkfield was both headmaster of a college and an evaluator for independent schools and universities in Australia.

The framework that emerged from this exhaustive meta-evaluation identified 26 approaches to evaluation. These were sorted into three major categories according to a set of descriptors: pseudoevaluations, quasi-evaluations, and true evaluations. In turn, quasi-evaluations are divided into two subcategories: questions-oriented and methods-oriented quasi-evaluations; and true evaluations are separated into three sub-categories: improvement/accountability true evaluations, social agenda/advocacy true evaluations, and eclectic true evaluations. **Table 4** outlines these categories and identifies the evaluation approaches associated with each one.

Pseudoevaluations

The authors identify pseudoevaluations as the first category. Pseudoevaluations are studies that masquerade as valid evaluations, but conceal, distort, or falsify the evaluation findings. The motivation behind pseudoevaluations is usually the desire to promote a program or political aims. The five types of pseudoevaluations include public relations studies that present program strengths but conceal weaknesses; politically controlled studies that deceive and mislead through selective and biased evaluation, pandering evaluations, evaluation-by-pretext studies, and empowerment under the guise of evaluation, that is, studies fully controlled by a program but done under the name and status of an evaluator to gain credibility. The discussion of pseudoevaluations makes a unique contribution by clearly flagging those approaches an evaluator should avoid and by demonstrating the importance of ethics as a central consideration throughout the evaluation process.

Quasi-evaluations

The second category includes approaches that legitimately evaluate programs, but have some deficiencies. Stufflebeam and Shrinkfield classify 14 approaches as quasi-evaluations because they narrow the scope of the evaluation and thereby do not provide a complete assessment of a program's merit or worth. The first two quasi-evaluation approaches are objectives-based studies and accountability studies, especially payment-by-results studies. The authors call them questions-oriented approaches because they tend to focus on a limited set of questions related to a program's objectives or the accountability requirements of a funding body.

Table 4 shows that the other 12 quasi-evaluation approaches include the success case method, objective testing programs, outcome evaluation as value-added assessment, performance testing, experimental studies, management information systems, benefit-cost analysis, clarification hearings, case study evaluations, criticism

Table 4 Approaches by categories as identified by meta-evaluation

Categories	Approaches
Pseudoevaluations	1. Public relations-inspired studies 2. Politically controlled studies 3. Pandering evaluations 4. Evaluation by pretext 5. Empowerment under guise of evaluation 6. Objectives-based studies
Quasi-evaluations: Questions-oriented evaluation approaches	7. Accountability studies, particularly payment-by-results studies 8. Success case method
Quasi-evaluations: Methods-oriented evaluation approaches	9. Objective testing programs 10. Outcome evaluation as value-added assessment 11. Performance testing 12. Experimental studies 13. Management information systems 14. Benefit–cost analysis 15. Clarification hearing 16. Case-study evaluations 17. Criticism and connoisseurship 18. Program-theory-based evaluation 19. Mixed-methods studies
True evaluations: Improvement/accountability-oriented evaluation approaches	20. Decision/accountability-oriented studies 21. Consumer-oriented studies 22. Accreditation/certification approach
True evaluations: Social agenda/advocacy evaluation approaches	23. Client-centered studies or responsive evaluation 24. Constructivist evaluation 25. Deliberative democratic evaluation
True evaluations: Eclectic evaluation approaches	26. Utilization-focused evaluation

From Stufflebeam, D. L. and Shrinkfield, A. J. (2007). *Evaluation Theory, Models, and Applications*. San Francisco, CA: Jossey-Bass.

and connoisseurship, program theory-based studies, and mixed-methods approaches. Stufflebeam and Shrinkfield call these methods-oriented approaches because they prescribe the application of a specific method (e.g., literacy test, success case method) or set of methods (e.g., use of both quantitative and qualitative methods, whether appropriate or not). Since many of the shibboleths of the evaluation field fall into the quasi-evaluations category, this categorization is both controversial and thought provoking.

True evaluations

The authors reserve the term true evaluations for those approaches that assess a program's merit and worth in a comprehensive manner. True evaluations also foster program improvement by supplying information to program decision makers and assist consumers with relevant information about program options and competing programs.

The authors cluster true-evaluation approaches into three groups. The first group is the improvement and accountability oriented evaluation approaches that include decision and accountability oriented studies, consumer-oriented approaches, and accreditation and certification approaches. Decision and accountability oriented

approaches encourage evaluators to interact with stakeholders and supply useful information that contributes to program improvement, rather than rendering a judgment about outcomes of a program alone. Both consumer-oriented and accreditation and certification approaches strive to assist and protect consumers and the public.

The second group is the social agenda and advocacy evaluation approaches that support equity and social justice. These approaches include client-centered studies or responsive evaluation, constructivist evaluation, and deliberative democratic evaluation approaches. A key aspect of these approaches is the inclusion of safeguards that eliminate the biases and distortions that characterize pseudoevaluations.

The third group is called eclectic evaluation approaches but it contains only one approach – the utilization-focused evaluation approach developed by Patton (1997). Utilization-focused evaluation focuses on designing evaluations so they are useful and developing evaluation questions that provide important information to a range of stakeholders. This approach views evaluators working collaboratively with program managers, staff, and clients throughout the process of program development. As the program develops, it also evolves; therefore, the evaluation questions and the specific evaluation methods change as

well. Stufflebeam and Shrinkfield acknowledged that it was difficult to classify utilization-focused evaluation because although it shares common elements with other approaches, such as involvement with stakeholders and a concern for program improvement, it also has characteristics that set it apart.

It is worth noting that checklists for a number of the highest-rated approaches are available to the public for download from the website of The Evaluation Center at Western Michigan University. The checklists were developed by experts in particular evaluation approaches to capture lessons learned from practice and to remind evaluators what they are supposed to check, observe, and do during an evaluation.

A major contribution of this framework based on meta-evaluation is the application of a clear set of criteria to assess the comparative strengths and weaknesses of a wide variety of approaches to evaluation. The recognition of the pitfalls leading to pseudoevaluations is a particularly valuable feature of this framework, especially those new to the evaluation field. Another contribution is the acknowledgment that all evaluation approaches have weaknesses as well as strengths. Knowing the strengths and weaknesses of the evaluation approaches during the design process allows evaluators to reduce potential biases and overcome practical obstacles in the application of different approaches.

One of the limitations of this framework is that the meta-evaluation is based on the judgments of only two developers, even though these evaluation experts applied detailed checklists and drew on their many decades of experience in the evaluation field to make those judgments. Another consideration is that the range and amount of detail contained in the descriptions of the individual evaluation approaches can be overwhelming even for knowledgeable and experienced evaluators. Novice evaluators will likely require extensive period of time to study, master, and apply this framework in real-world situations. An additional factor is that complex educational-evaluation environments may require the application of more than one approach or a hybrid approach during an evaluation, but Stufflebeam and Shrinkfield's framework tends to emphasize the use of one approach alone.

Summary

Educational evaluation has grown rapidly during the last two decades. As the boundaries of educational evaluation continue to widen, the purposes for evaluation are expanding also. Evaluation is being used to measure student progress, reform education systems, and enhance accountability for outcomes. School administrators and teachers alike are conducting evaluations of their own to improve school performance and foster creative spaces for learning. This

proliferation of the purposes for evaluation has brought with it a greater appreciation of the demand for evaluators to understand the evaluation needs of stakeholders, recognize the importance of contextual factors (e.g., normative, political, ethical, and organizational) in the evaluation enterprise, and select an evaluation approach that is responsive and supplies useful information to specific audiences.

The contextualizing of evaluation has fundamentally changed the way evaluations are designed. Instead of consistently applying one approach or set of methods to all evaluation situations, evaluators are aware that the choice of evaluation approaches and methods for data collection and analyses must be matched to the unique purposes of an evaluation within a unique evaluation setting. In response to these new realities, literally dozens of approaches to evaluation have emerged. Essentially, the approaches to evaluation define the content of the evaluation field.

The challenge in this article was finding a way to make sense of this complexity while acknowledging the enormous range and diversity of evaluation approaches. The overall strategy made use of four major frameworks developed by highly experienced educational evaluators. These frameworks have proved their value as lenses for viewing more clearly the realities of contemporary evaluation and as a guide to evaluation practice. The use of four frameworks gives insight into alternative characterizations of the evaluation field and portrays the evaluation process in a less linear fashion than by examining one framework alone.

The presentation of these four evaluation frameworks is intended as a point of departure, because the number and range of evaluation approaches are likely to increase in the years ahead. Although at one time educational evaluation was influenced primarily by the social and behavioral sciences, contemporary evaluation is considered by many to be a trans-discipline (Scriven, 2003) that has its roots in many distinct disciplines. As evaluation continues to be influenced by new and emerging disciplines, our frameworks and approaches to program evaluation are likely to evolve regarding their basic assumptions, purposes, processes and tools, roles of evaluators, and modes of disseminating and using evaluation information (Love, 2001).

See also: Defining Quality in Evaluation; Educational Evaluation: Concepts, Practice, and Future Directions; Evaluating Education in Three Policy Eras; Evaluation Use; Internal and External Evaluation; Meta-evaluation: Purpose, Prescription, and Practice; The Purpose of Educational Evaluation.

Bibliography

- Berk, R. and Rossi, P. (1998). *Thinking about Program Evaluation*, 2nd edn. Thousand Oaks, CA: Sage.
- Love, A. J. (2001). The future of evaluation: Catching rocks with cauldrons. *American Journal of Evaluation* 22, 437–444.

- Nagao, M., Kuji-Shikatani, K., and Love, A. J. (2005). Preparing school evaluators: Hiroshima pilot test of the Japan Evaluation Society's accreditation project. *Canadian Journal of Program Evaluation* 20(2), 125–155.
- Owen, J. (2007). *Program Evaluation: Forms and Approaches*, 3rd edn. New York: Guilford Press.
- Patton, M. Q. (1997). *Utilization-Focused Evaluation – The New Century Text*, 3rd edn. Thousand Oaks, CA: Sage.
- Rossi, P., Freeman, H., and Lipsey, M. (2004). *Evaluation: A Systematic Approach*, 7th edn. Thousand Oaks, CA: Sage.
- Scriven, M. (2003). Evaluation in the new millennium: The transdisciplinary view. In Donaldson, S. I. and Scriven, M. (eds.) *Evaluating Social Programs and Problems: Vision for the New Millennium*, pp 19–42. Mahwah, NJ: Erlbaum.
- Stufflebeam, D. L. and Shrinkfield, A. J. (2007). *Evaluation Theory, Models, and Applications*. San Francisco, CA: Jossey-Bass.
- The Joint Committee on Standards for Educational Evaluation (1994). *The Program Evaluation Standards: How to Assess Evaluations of Educational Programs*, 2nd edn. Thousand Oaks, CA: Sage.
- Worthen, B., Sanders, J., and Fitzpatrick, J. (2004). *Program Evaluation: Alternative Approaches and Practical Guidelines*, 3rd edn. New York: Pearson Education.
- Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Evaluation* 97, 7–35.
- Christie, C. A. and Alkin, M. C. (2008). Evaluation theory tree re-examined. *Studies in Educational Evaluation* 34(3), 131–135.
- Hansen, H. F. (2005). Choosing evaluation models: A discussion on evaluation design. *Evaluation* 11(4), 447–462.
- Schwandt, T. A. (2005). A diagnostic reading of scientifically-based research for education. *Educational Theory* 55, 285–305.
- Scriven, M. (1991). *Evaluation Thesaurus*, 4th edn. Thousand Oaks, CA: Sage.
- Stufflebeam, D. L. (2000). Foundational models for 21st century program evaluation. In Stufflebeam, D. L., Madaus, G. F., and Kellaghan, T. (eds.) *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, pp 33–84. Boston, MA: Kluwer Academic Publishers.
- Stufflebeam, D. (2001). Evaluation models. *New Directions for Evaluation* 89, 7–106.
- Stufflebeam, D. L., Madaus, G. F., and Kellaghan, T. (eds.) (2001). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, 2nd edn. New York: Springer.

Further Reading

- Alkin, M. C. (ed.) (2004). *Evaluation Roots: Tracing Theorists' Views and Influences*. Thousand Oaks, CA: Sage.
- Alkin, M. C. and Christie, C. A. (2004). An evaluation theory tree. In Alkin, M. C. (ed.) *Evaluation Roots: Tracing Theorists' Views and Influences*, pp 12–66. Thousand Oaks, CA: Sage.

Relevant Websites

- <http://www.wmich.edu> – Evaluation Checklists may be downloaded from the Evaluation Center at Western Michigan University.
- <http://www.wmich.edu> – The Joint Committee on Standards for Educational Evaluation's website offers The Program Evaluation Standards, The Student Evaluation Standards, and The Personnel Evaluation Standards.

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

This page intentionally left blank

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

EDITORS-IN-CHIEF

PENELOPE PETERSON

School of Education and Social Policy
Northwestern University
Evanston, IL
USA

EVA BAKER

UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

BARRY MCGAW

McGraw Group Pty Ltd
Mont Albert Nth, VIC
Australia



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
The Boulevard Langford Lane, Kidlington, Oxford OX5 1GB, UK

Copyright © 2010 Elsevier Ltd. All rights reserved

LATVIA
Copyright © 2010 A Joma

STUDENT DATA SYSTEMS AND THEIR USE FOR EDUCATIONAL IMPROVEMENT
Copyright © 2010 J C Wayman, V Cho, and M P Richards

QUALIFICATIONS FRAMEWORKS AND THEIR ROLE IN THE REFORM OF EDUCATION AND TRAINING
Copyright © 2010 M Young

LANGUAGE OF VOCATIONAL EDUCATION AND TRAINING
Copyright © 2010 J Misko, J MacKenzie, and R-A Polvere

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44)(0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at (<http://elsevier.com/locate/permissions>), and selecting *Obtaining permissions to use Elsevier material*

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Catalog Number
A catalogue record for this book is available from the Library of Congress

ISBN: 978-0-08-044893-0

For information on all Elsevier publications
visit our website at books.elsevier.com

PRINTED AND BOUND IN THE UNITED KINGDOM
09 10 11 12 13 10 9 8 7 6 5 4 3 2 1

PREFACE

A preface usually provides a brief introduction to a work, intended to set the stage, provide some background insight, and whet the appetite of the reader. In our case, however, this preface has to address a fundamental question, one that was in our minds at the time we were recruited as Editors-in-Chief for the International Encyclopedia of Education. The question was “Why do we need an Encyclopedia? Its subtext was inspired by the ever-growing, ever-popular Internet. We believe that *this* Encyclopedia is desperately needed and will become a valued resource in education and associated social sciences and arts. The reasons are intellectual and procedural. Anyone with a modicum of knowledge knows that finding and trusting information gleaned from the Internet are two separate actions. The reliance on browsers to help discover references and comments result in resources based on popularity not quality. Pithy titles catch the eye and references rise in the ranks of browser searchers. Related to this is the “editing” in the Internet realm of populist efforts at encyclopedia, references, and other compilations. Once again, after removing offensive material, the accuracy, completeness, lack of bias, and other provenance for entries simply do not exist. Experienced researchers in education can sort through and make intelligent choices. Novices and many journeyman, or practitioners, parents, and policy makers cannot. Contrast how this Encyclopedia was built. Key domains of educational research were identified, and a tentative list of sub-domains or useful applied areas was posited. Then the Editors-in-Chief (apologies for the awkwardness of the term) identified the leading researcher in a particular domain, and with surprisingly little effort, recruited them to participate. They in turn identified the two best researchers in a sub-domain, such as formative assessment or the training of pre-school teachers. The authors of the sections of the Encyclopedia do not represent a collective group of friends and acquaintances, although friendships have been made. Rather they embody a deep and broad scholarly community. The difference from compiled Internet resources is the built-expertise and intellectual engagement of the authors. The summary of the developments and futures in their personal areas of scholarship have been filtered through their years of experience, both as scholars and communicators. Quality, then, is endemic to each piece, developed through this top-down identification of expertise, and made indelible by the bottom-up application of high standards from people leading the sub-domains – the authors, and the domains themselves, the section editors.

On a procedural level, the publishers early committed to the notion that this Encyclopedia would also be an online resource, and access would be available through print, for those with strong bookcases and the persisting love of turning real pages. The Internet version will allow multiple prisms through which the reader may access articles and provide, as it were, an emulation of the Internet in our field, albeit bounded by expertise and high quality.

What must be underscored in the assessment of this effort are the Editors-in-Chief and the publishers’ commitment to find excellence worldwide. We tried very hard to persuade notable scholars from all parts of the world to make contributions. Less than to fulfill the title of “International,” we were on the hunt for perspectives that would enrich the scope and depth of the sections. Our section editors put in enormous time attempting to find the best in the field, wherever they resided. Yet, not everyone is in the volume. Some were overcommitted. Many were not fully confident of their English, and the automated translation software has not yet met standards for technical writing. We believe that such writing and editing tools will make the outreach to an even broader International group of scholars possible in future revisions, or online updates. Furthermore, the birth of the World Educational Research Association (in 2009) will provide a better set of interlocking networks to find and evaluate scholarship from any place on the globe.

Finally, the scope of the effort must be acknowledged: 28 section editors, 926 articles were commissioned, drafted, reviewed, redrafted, edited, and put together in the space of four years. The publishers underwent some internal changes, and alterations in management. We as Editors-in-Chief, changed roles, moved, and also had to keep our own research and development enterprises afloat. Deadlines wobbled; authors dropped from view and had to be replaced.

Yet, at times frustrating as all development is, we find the final product exhilarating. We are enthusiastic not simply because it came into being at all, but because the collective light of the minds that wrote have left a bright resource for the future, one that will impact the way our colleagues understand and experience the educational knowledge, improvement, and impact in the future.

Eva Baker
Penelope Peterson

EDITORS

EDITORS-IN-CHIEF

Penelope Peterson

School of Education and Social Policy
Northwestern University
Evanston, IL
USA

Eva Baker

UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

Barry McGaw

McGaw Group Pty Ltd
Mont Albert Nth
Melbourne, VIC
Australia

SECTION EDITORS

Walter Allen

HEOC Division
University of California, Los Angeles
Los Angeles, CA
USA

Vibeke Grøver Aukrust

Institute for Educational Research
University of Oslo
Oslo
Norway

Dominic J Brewer

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

Kai-Ming Cheng

Faculty of Education
The University of Hong Kong
Hong Kong

Gregory Chung

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Penny Enslin

Department of Educational Studies
University of Glasgow
Glasgow
UK

Noelle Griffin

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Dennis Hocevar

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

David Hopkins

Institute of Education
University of London
London
UK

Mary James

Faculty of Education
University of Cambridge
Cambridge
UK

Sanna Järvelä

Department of Educational Sciences and
Teacher Education
University of Oulu
Oulu
Finland

Thomas Karmel

National Centre for Vocational Education Research
Adelaide, SA
Australia

Harold G Levine

School of Education
University of California, Davis
Davis, CA
USA

Robert L Linn

University of Colorado at Boulder
Ouray, CO
USA

Allan Luke

Faculty of Education
Queensland University of Technology
Brisbane, QLD
Australia

Ann M Mastergeorge

Human Development and Family Studies
University of California, Davis
Davis, CA
USA

Patrick J McEwan

Department of Economics
Wellesley College
Wellesley, MA
USA

Barbara Means

Center for Technology in Learning
SRI International
Menlo Park, CA
USA

Ingrid Moses

University of Canberra
Canberra, ACT
Australia

Bill Mulford

Faculty of Education
University of Tasmania
Hobart, TAS
Australia

Yu Jing Ni

Faculty of Education
The Chinese University of Hong Kong
Hong Kong

Festus Obiakor

Department of Exceptional Education
University of Wisconsin-Milwaukee
Milwaukee, WI
USA

P David Pearson

Graduate School of Education
University of California, Berkeley
Berkeley, CA
USA

Colin Power

Graduate School of Education
University of Queensland
Brisbane, QLD
Australia

Kjell Rubenson

Faculty of Education
The University of British Columbia
Vancouver, BC
Canada

Sandip Sinharay

Education Testing Services
Princeton, NJ
USA

Collette Tayler

Queensland University of Technology
Brisbane, QLD
Australia

Nico Verloop

ICLON, Graduate School of Education
Leiden University
Leiden
Netherlands

HOW TO USE THE ENCYCLOPEDIA

The International Encyclopedia of Education is intended for use by students, research professionals, and interested others. Articles have been chosen to reflect major disciplines in the study of education and common topics of research by academics in this domain. Each article serves as a comprehensive overview of a given area, providing both breadth of coverage for students, and depth of coverage for research professionals. We have designed the encyclopedia with the following features for maximum accessibility for all readers.

The contents of the encyclopedia are arranged alphabetically by section, and within sections, alphabetically by article. The Subject Index is located in Volume 8. Some topics are covered in a multitude of articles from differing perspectives, while other topics may have only one entry. We encourage use of the index for access to a subject area, rather than use of the Contents list alone, so that a reader has a full notion of the coverage of that topic.

The articles include cross-references to other related encyclopedia articles, suggested further readings where applicable, and many contain relevant websites for additional information. We encourage readers to use the cross-references to locate other encyclopedia articles that will provide more detailed information about a subject.

The Further Reading sections include recent secondary sources to aid the reader in locating more detailed or technical information. Review articles and research articles that are considered of primary importance to the understanding of a given subject area are also listed. These suggested further readings are not intended to provide a full reference listing of all material covered in the context of a given article, but are provided as next steps for a reader looking for additional information.

This page intentionally left blank

CONTENTS

Preface	v–vi
Editors	vii–viii
How to use the Encyclopedia	ix

VOLUME 1

A

Adult Education

Adult Education Overview	<i>K Rubenson</i>	1
--------------------------	-------------------	---

Adult Education – Adult Learning, Instruction and Program Planning

Adult Learning	<i>S B Merriam</i>	12
Adult Learning and Instruction: Transformative Learning Perspectives	<i>P Cranton</i>	18
Adult Learning in a Biographic Perspective	<i>L West</i>	25
Adult Learning, Instruction and Programme Planning: Insights from Freire	<i>P Mayo</i>	31
Characteristics of Adult Learning	<i>K Illeris</i>	36
Informal Learning: A Contested Concept	<i>P Hodgkinson</i>	42
Organizational Learning	<i>P-E Ellström</i>	47
Program Planning	<i>A L Wilson and R M Cervero</i>	53
Workplace Learning Frameworks	<i>S Billett</i>	58

Adult Education – as a Field of Study

Class Analysis in Adult Education	<i>T Nesbit</i>	64
Gender Analysis	<i>J Stalker</i>	70
Race and Ethnicity in the Field of Adult Education	<i>J Johnson-Bailey and D Drake-Clark</i>	77
Trends in Workplace Learning Research	<i>T Fenwick</i>	83

Adult Education – Concepts

Lifelong Learning	<i>J Field</i>	89
Rewriting the History of Adult Education: The Search for Narrative Structures	<i>B J Hake</i>	96

Adult Education – Domains and Provision

Adult Basic Education: A Challenge for Vocational Based Learning	<i>R Höghelm</i>	102
Adult Literacy Education	<i>L Tett and R St.Clair</i>	107
Citizenship and Immigrant Education	<i>D Schugurensky</i>	113
Community Based Adult Education	<i>B Connolly</i>	120
Continuing Professional Education: Multiple Stakeholders and Agendas	<i>R M Cervero and B J Daley</i>	127
Financing of Adult and Lifelong Learning	<i>Å Sohlman</i>	133
Labor Education	<i>K Forrester</i>	139
Learning Cities and Regions	<i>C Duke</i>	144
Museums as Sites of Adult Learning	<i>R S Grenier</i>	150
Overview of Lifelong Learning Policies and Systems	<i>D-B Kwon and C H O Daeyeon</i>	156
Popular Adult Education	<i>S Tøsse</i>	162
Provision of Prior Learning Assessment	<i>P Andersson</i>	169
The Age of Learning: Seniors Learning	<i>P Jarvis</i>	175
University Adult Continuing Education: The Extra-Mural Tradition Revisited	<i>M Slowey</i>	183

Adult Education – Economy and Society

Adult Education and Civil Society	<i>J Crowther and I Martin</i>	191
Adult Education and Nation-Building	<i>S Walters</i>	198
Economic Outcomes of Adult Education and Training	<i>A Ferrer and W C Riddell</i>	204
Health and Adult Learning	<i>P Bélanger and M Robitaille</i>	211
Modernization Processes and the Changing Function of Adult Learning	<i>H S Olesen</i>	217
The Political Economy of Adult Education	<i>G Rees</i>	223
Wider Benefits of Adult Education	<i>R Desjardins and T Schuller</i>	229

Adult Education – Participation

Barriers to Participation in Adult Education	<i>K Rubenson</i>	234
Participation in Adult Learning	<i>R Desjardins</i>	240

C**Comparative Education**

Comparative and International Education: English-Language Scholarly Publications	<i>D Post and M Fedri</i>	251
Comparative Education: Philosophical Issues and Concepts	<i>M Mason and C W Evers</i>	257
Comparative Education: Societies and Associations	<i>M Bray</i>	266

Curriculum Development

Curriculum and Religion	<i>C Kapitzke and P Graham</i>	272
Gender and Curriculum	<i>J Marsh</i>	277
Local Teacher-Based Curriculum Development	<i>M D Osborne and H B Barker</i>	283

Curriculum Development – Contexts

Curriculum and Human Rights	<i>C Lohrenscheit</i>	287
Curriculum and the Education of Cultural and Linguistic Minorities	<i>S May</i>	293
Curriculum in Postcolonial Contexts	<i>A Hickling-Hudson</i>	299
Curriculum, Economic and Cultural Development	<i>A Rampal</i>	307
Globalization and Curriculum	<i>R Dale</i>	312
Popular Culture	<i>K Schultz and R Throop</i>	318

Curriculum Development – Evaluation and Research

An Overview of Research in Curriculum Inquiry	<i>F M Connelly and S J Xu</i>	324
Curriculum Evaluation: Approaches and Methodologies	<i>V Klenowski</i>	335
Curriculum Studies, Discourse Analysis, and the Construction of Historical Time	<i>J Laspina</i>	342
Curriculum Use in the Classroom	<i>M Ben-Peretz and B Eilam</i>	348
The Roots and Routes of Teacher-Based Action Research and Curriculum Inquiry: An Historical Perspective	<i>L G Putney and J L Green</i>	355

Curriculum Development – Planning and Development

Curriculum and Syllabus Design	<i>A Woods, A Luke, and K Weir</i>	362
Curriculum and the Publishing Industry	<i>J Larson, A-R Allen, and D Osborn</i>	368
Curriculum and Teacher Change	<i>J G Ladwig</i>	374
Curriculum Governance and Planning	<i>B Levin</i>	379
Curriculum Planning and Systems Change	<i>Z Deng</i>	384
Curriculum, Digital Resources and Delivery	<i>Y Zhao, G Zhang, and C Lai</i>	390
Textbook Development and Selection	<i>P Shannon</i>	397

Curriculum Development – Subject Matter

Curriculum and Globalization: Higher Education	<i>K H Mok</i>	403
Curriculum Development as Subject Matter: Social Studies	<i>J Brophy and J Alleman</i>	411
Curriculum Development in the Area of Reading	<i>C L Madda and T E Raphael</i>	415
Curriculum in the Arts	<i>C M Thompson, L Bresler, and T E Costantino</i>	423
Curriculum: Foreign Language Learning	<i>Angel M Y Lin</i>	428
First Language Learning	<i>T Rogers and L Moffatt</i>	435
Learning a Second Language in First Language Environments	<i>E Bernhardt and D J Tedick</i>	441
Literature	<i>S Hynds</i>	447
Mathematics	<i>B Kaur and C P Vistro-Yu</i>	453
Physical Education and Sports	<i>D Kirk</i>	459
Primary and Elementary/Middle Grades Reading	<i>D E Alvermann and C A Mallozzi</i>	464
Science	<i>Y Kali and M Linn</i>	468

Curriculum Development – Theory

Bilingual Learning (Learning L1 and L2 in an L1 and L2 Environment)	<i>D Schwinge</i>	475
Curriculum and Complex Systems Theory	<i>W-M Roth and J Thom</i>	481

Curriculum and Constructivism	<i>B Davis and D Sumara</i>	488
Curriculum and Critical Theory	<i>L Yates</i>	494
Curriculum and Poststructuralist Theory	<i>J L Miller</i>	499
Curriculum and Structuralist Sociology: The Theory of Codes and Knowledge Structures	<i>J Muller and J Gamble</i>	505
Curriculum Reform	<i>R Gilbert</i>	510
Early Childhood Curriculum and Developmental Theory	<i>S M-Y Lim and C Genishi</i>	514

D

Demography and Social Change

An Evaluation Framework for E-Learning Effectiveness in the Arab World	<i>W AbuSneineh and M Zairi</i>	521
Globalization and Social Justice in Higher Education	<i>E Kimura-Walsh</i>	536
Racial Inequality and the Economic Payoffs of Higher Education	<i>S A Lachtman</i>	542
School Health Education	<i>J S Yang</i>	547

Demography and Social Change – Demography

A Paradigm of Contradictions: Racism and Science Education	<i>B A Brown and J W Mutegi</i>	554
Africana Studies: Past, Present And Future	<i>B R Hare, A Young, J B Stewart, and D P Aldridge</i>	565
Beyond Black Nationalism to Black Internationalism: The Compelling Case for Examining Black Educational Challenges Globally	<i>K Freeman</i>	573
Beyond Racial, Ethnic, and Gender Bias in Education Statistics	<i>J T Lott and C Bennett</i>	581
Caste and Education in India	<i>G G Wankhede</i>	589
Demography, Community, and the Education of Asian American and Pacific Islander Students	<i>R T Teranishi</i>	595
Educational Attainment and Mortality Differentials	<i>H D Horton, L Hixson, and S M Monnat</i>	602
Educational Opportunity and Latino/Chicano College Choice: New Findings, and Theoretical Perspectives	<i>M A Ceja and P A Pérez</i>	610
Ethnic Minority Identity and Educational Outcomes in a Rising China	<i>G A Postiglione</i>	616
Ethnic Studies and Women's Studies: From the Past to the Present, Providing a Space in the Academy for Scholarship Focused on the Histories and Experiences of Marginalized People	<i>C A Brown</i>	623
Gender Equity in Higher Education: Challenges and Celebrations	<i>L Morley</i>	629
Locating Space and Place in the College Access Debate: New Tools for Mapping and Understanding Educational Inequity and Stratification	<i>B Cantwell and J F Milem</i>	636
Public Policy and Inequality in Postsecondary Opportunity: Educational Statistics and the Failure of Education Reform	<i>E P St. John, K Williams, and K Moronski</i>	649
The Education of Indigenous Students	<i>J W Tippeconnic and S C Faircloth</i>	661
The Health Advantages of Educational Attainment	<i>G Y Dinwiddie</i>	667
To Go or Not to Go: Access and Barriers to Ethnic Minority Participation in Higher Education within the United Kingdom	<i>P Brug and T L Oliver</i>	673
Universal Secondary Education and the Two-Tiered School System in the English-Speaking Caribbean	<i>P Anderson and J Devonish</i>	679
Women of Color in Science, Technology, Engineering and Mathematics (STEM): Refining the Concepts, Reframing the Issues	<i>C B Leggon</i>	686

Demography and Social Change – Social Change

Educational Diversity in U.S. Law Schools	<i>C E Daye, A T Panter, W R Allen, M E Deo, and L F Wightman</i>	691
Educational Reform	<i>F Lopez-Segrera</i>	698
Emergence of For-profit Higher Education	<i>X Li</i>	705
Engaged Education: Experiential Learning, Intensive Field Experiences, and Social Change	<i>T R Rone</i>	711
Higher Education and Social Change in Mexico	<i>A Mungaray, G Estrella, P Moctezuma, and M Ramírez-Urquidy</i>	719
The Changing Role of the State in Higher Education	<i>K Gürüz</i>	728
The Role of the State in School Reform: Responding to Increasing Ethnic Diversity in Britain	<i>P Ratcliffe</i>	737
Transforming Mindsets Through Education for Sustainable Development	<i>V M Segovia</i>	746

Demography and Social Change – Sociology

Achievement Motivation in Ethnic Minority Youth	<i>S Graham and E Bryant</i>	753
Affirmative Action and Higher Education in Brazil	<i>J T dos Santos and D M Queiroz</i>	760
Black in White: Black Students at White and Black Colleges	<i>L Evans and J R Feagin</i>	767
Bologna Process: On the way to a Common European Higher Education Area	<i>A V Rezaev</i>	772
Critical Race Theory in Education Research	<i>R Vue and C B Newman</i>	779
Cultural Diversification and Japanese Education: Social Constructions of the New Diversity	<i>R Tsuneyoshi</i>	787
Education in War Countries: A Disempowerment of Society	<i>J L Nkata</i>	793
Educational Attainments of U.S. Black Males and Females: 1971 to 2003	<i>F B Bonner, M-C Jipguep-Akhtar, and R J Harrison</i>	799
Educational Preparation: Fostering the Self-Efficacy and Resilience of Urban Adolescent Youth	<i>M B Spencer and B Tinsley</i>	806
HIV Education for Low-Literate People: Transforming Students and Communities through Paulo Freire's Praxis and the Pedagogy of Action	<i>N Z Haniff</i>	814
How Are Discussions of Interdisciplinary Studies Linked to Diversity Discourse?	<i>C N Spence</i>	823
Identifying Patterns of Doctoral Attrition Across Academic Fields	<i>A James, K J Fosnacht, E B Stolzenberg, and D Thirumalai</i>	827
Male Underachievement in Education Across the Globe: A Shift in Paradigm for Gender Disparities Regarding Academic Achievement	<i>J F L Jackson, J L Moore III, and R A Leon</i>	838
Noncognitive Measures for Higher Education Admissions	<i>W E Sedlacek</i>	845
Peace Education	<i>J Page</i>	850
Remote Control: Africa's Development and Outflow to Developed Nations of Technical Talent	<i>F Haniff</i>	855
Scholar-Baller®: Student Athlete Socialization, Motivation and Academic Performance in American Society	<i>C K Harrison, S Bukstein, J Mottley, E Comeaux, J Boyd, C Parks, and D Heikkinen</i>	860
Service-Learning and Diversity as Mediums to Foster Civic and Community Engagement in Undergraduates	<i>L Oseguera</i>	866
Success and Failure in Education and Criminal Justice: Identifying Common Mechanisms	<i>S Ramirez and A Harris</i>	874
The History of Education: Race and Education	<i>J O Jewell</i>	881

VOLUME 2**E****Early Childhood Care and Education Theory**

- Early Childhood Care and Education: The Family, The Market, and The State *S B Kamerman and M J Neuman* 1
- Gender Issues in Early Childhood Education and Care *B Pocock* 11

Early Childhood Education and Care Governance

- Early Childhood Risk, Protection and Abuse Prevention *E Khoo* 17
- Evaluating Early Childhood Education and Care Programs *M Janus and S Brinkman* 25

Early Childhood Education and Care Investment

- Investing in Early Childhood Education and Care: The Health and Wellbeing Case *T Moore and F Oberklaid* 32
- Cost-effective Early Childhood Programs from Preschool to Third Grade *A J Reynolds, J A Temple, and B White* 38
- Early Childhood in Post-Modern Cultures: Thoughts and Some Concerns *P Gammage* 49
- Investing in Early Childhood Education and Care: Some Policy Implications *J Bennett* 55
- Investing in Early Childhood Education and Care: The Economic Case *G Cleveland and M Krashinsky* 63

Early Childhood Education and Care Programs

- Early Care and Education Programs for Infants and Toddlers *B Nyland* 69
- Literacies in Early Childhood: The Preschool Period *M Fler and B Raban* 75
- Participation in Early Childhood Education and Care Programs: Equity, Diversity and Educational Disadvantage *M Vandenbroeck* 81
- Teaching in Early Childhood Centers Instructional Methods and Child Outcomes *I Siraj-Blatchford* 86

Early Childhood Education and Care Provision

- Early Childhood Education and Care in the People's Republic of China *Zeng Xiaodong* 93
- Home-Based and Institutional Early-Childhood Education and Care Services *R Fukkink and H Blok* 98
- Integrated Early Childhood Education and Care Services – Care, Upbringing, Education and Health *S Colley* 104
- Issues of Access and Program Quality *J T Hustedt and W S Barnett* 110
- Kindergarten Transitions and Linkages to Primary School – Readiness Re-conceptualized *A Petriwskyj* 120
- Parent Support in Early Childhood – Approaches and Outcomes *K S Howard, K A Beckmann, and J Brooks-Gunn* 126

Early Childhood Education and Care Research

- Development and Implementation of Early Learning Standards in the United States *C Scott-Little* 132
- Global Status of Early Learning and Development Standards *P R Britto and S L Kagan* 138
- Rethinking Early Childhood Education and Care: Implications for Research and Evaluation *K Thorpe, D Cloney, and C Tayler* 144

Early Childhood Education and Care Theory

- Child Rearing and Early Education: Parents and Professionals: Theoretical and Historical Influences *H Penn* 151

Early Childhood Education and Care Workforce

- Early Childhood Education and Care Professional Profiles in the United Kingdom: Multiagency Teamwork and Transdisciplinary Challenges *A Anning and P Calder* 157
- Early Childhood Education and Care Staff Support: Mentoring and Professional Development in Finland *P Lindberg* 162

Economics of Education

- Education and Health *P Muennig* 169
- Data *M Perez and M Socias* 178
- Empirical Research Methods in the Economics of Education *P J McEwan* 187
- Theoretical Concepts in the Economics of Education *D J Brewer, G C Hentschke, and E R Eide* 193

Economics of Education – Costs and Financing of Schooling

- Cost–Benefit Analysis and Cost–Effectiveness Analysis *C Belfield and H M Levin* 199
- Economic Approaches to Adequacy *J Imazeki* 204
- Economic Approaches to School Efficiency *L L Taylor* 210
- School Finance Reform *T Downes* 216
- School Finance: An Overview *J K Rice, D H Monk, and J Zhang* 222
- The Economics of Tuition and Fees in Higher Education *R G Ehrenberg* 229

Economics of Education – External Benefits of Schooling

- Education and Civic Engagement *T S Dee* 235
- Education and Crime *L Lochner* 239
- Education and Economic Growth *E A Hanushek and L Wößmann* 245
- Parental Socioeconomic Status, Child Health, and Human Capital *J Currie and J Goodman* 253
- The External Benefits of Education *W W McMahon* 260

Economics of Education – Labor Markets, Education and Earnings

- Education and Inequality *J Blanden and S Machin* 272
- Human Capital *E R Eide and M H Showalter* 282
- Race Earnings Differentials *M Carnoy* 288
- Returns to Education in Developed Countries *M Gunderson and P Oreopoulos* 298
- Returns to Education in Developing Countries *H A Patrinos and G Psacharopoulos* 305
- School Quality and Earnings *J R Betts* 313
- Signaling in the Labor Market *M E Page* 321
- The Economics of High School Dropouts *R W Rumberger* 325

Economics of Education – Markets, Incentives and School Choice

- Competition and Student Performance *A A Payne* 331
- Educational Privatization *C Belfield and H M Levin* 337

Student Incentives	<i>E R Eide and M H Showalter</i>	342
The Economic Role of the State in Education	<i>D N Plank and T E Davis</i>	347
The Economics of Catholic Schools	<i>W Sander and D Cohen-Zada</i>	354
The Economics of Charter Schools	<i>R Zimmer and R Buddin</i>	360
The Economics of Parental Choice	<i>R Chakrabarti and J Roy</i>	367
The Economics of School Accountability	<i>D N Figlio and H F Ladd</i>	374
The Efficacy of Educational Vouchers	<i>R Zimmer and E P Bettinger</i>	380
Tiebout Sorting and Competition	<i>T J Nechyba</i>	388

Economics of Education – Production of Schooling

Desegregation, Academic Achievement and Earnings	<i>S Rivkin</i>	394
Education Production Functions: Concepts	<i>D N Harris</i>	402
Education Production Functions: Developed Country Evidence	<i>E A Hanushek</i>	407
Education Production Functions: Evidence from Developing Countries	<i>P Glewwe and S Lambert</i>	412
Family Environment in the Production of Schooling	<i>R Rothstein</i>	423
Neighborhoods and Peers in the Production of Schooling	<i>J Vigdor and J Ludwig</i>	431
Teacher Quality in Education Production	<i>I Pelayo and D J Brewer</i>	438
The Economics of Class Size	<i>D W Schanzenbach</i>	443
The Economics of Early Childhood Interventions	<i>M Nores</i>	450

Economics of Education – Teaching Labor Markets

An Overview of Teacher Labor Markets	<i>H Lankford and J Wyckoff</i>	456
Compensating Differentials in Teacher Labor Markets	<i>J Chambers</i>	465
Economic Approaches to Teacher Recruitment and Retention	<i>S Loeb and J Myung</i>	473
Teacher Incentives	<i>L Santibañez</i>	481
Teacher Supply	<i>P Dolton</i>	489
Teacher Training and Preparation in the United States	<i>M Podgursky, M G Springer, and R Hutton</i>	498
Teachers in Developing Countries	<i>F H Rogers and E Vegas</i>	504
The Economics of Teachers' Unions in the United States	<i>K Strunk</i>	512

Education of Children with Special Needs

Adapted Physical Activity for Students with Special Needs	<i>S R Hodge</i>	519
Assessment Accommodations for Children with Special Needs	<i>R J Kettler and S N Elliott</i>	530
Assistive Technology and Educational Practice	<i>H P Parette and G R Peterson-Karlan</i>	537
Augmentative–Alternative Communication	<i>C Mason and K M Chinn</i>	544
Community Integration and Employment of Youth with Special Needs	<i>B A Ford and A W Daviso III</i>	550
Cooperative Learning for Children with Special Needs	<i>E A Dooley and K R Kossar</i>	555
Co-Teaching	<i>E D Cramer</i>	560
Cultural and Linguistic Diversity and Children with Special Needs	<i>J M Patton, N Day-Vines, and A McSwain</i>	565
Curriculum-Based Assessment and Students with Special Needs	<i>D J O'Shea and J D Katsafanas</i>	571

Diagnosis and Classification	<i>M Sapp</i>	577
Direct Instruction and the Education of Children with Special Needs	<i>T Mehring</i>	584
Disabilities and Gifted Education	<i>V I Daniels and M J McCollin</i>	588
Early Childhood Special Education: Birth to Eight Years Old	<i>T S Lovelace, R Gardner III, J B Axe, and D M Sainato</i>	594
Early Intervention for Students with Special Needs	<i>P A Grant</i>	608
Early Transition of Children with Special Needs	<i>S A Fowler</i>	613
Educating Children Who are Deaf-Blind	<i>J Wood and K M Chinn</i>	621
Educating Children with Deafness and Hearing Impairments	<i>G M Mukuria and C J Eleweke</i>	628
Educating Children with Other Health Impairments (ADHD)	<i>G M Mukuria and F E Obiakor</i>	634
Educating Children with Physical Disabilities	<i>Z P Nkabinde, F E Obiakor, M T Ofor, and D J Smith</i>	640
Educating Children with Specific Learning Disabilities	<i>P A Grant and P B Grant</i>	646
Educating Children with Speech and Language Disorders	<i>S L Green and C D Qualls</i>	654
Educating Students with Autism and Related Disorders	<i>T T Dyches</i>	661
Educating Students with Cognitive Disabilities	<i>J P Bakken</i>	669
Educating Students with Emotional and Behavioral Disorders	<i>S W Smith and G G Taylor</i>	678
Educating Students with Special Needs: An Overview	<i>F E Obiakor, M K Harris, P Harris-Obiakor, and G M Mukuria</i>	688
Educating Students with Traumatic Brain Injury	<i>S Burkhardt and A Rotatori</i>	695
English Language Learners with Special Needs	<i>R Rueda and G Ragusa</i>	701
Functional Behavioral Assessment	<i>J P Bakken</i>	708
History of Special Education	<i>L M Garrick Duhaney and S J Salend</i>	714
Inclusion of Students with Special Needs in General Education Classrooms	<i>M A Prater</i>	721
Instructional Accommodations for Children with Special Needs in Inclusive Settings	<i>A L Yurick, S Musti-Rao, and G Cartledge</i>	727
Interagency Agreements	<i>L A Owens</i>	733
Interdisciplinary Collaboration in Early Childhood	<i>N File</i>	739
Language Development in Children with Special Needs	<i>M S L Stewart</i>	745
Large Scale Assessment and Accountability and Students with Special Needs	<i>M L Thurlow</i>	752
Literacy Instruction for Students with Special Needs	<i>L H Mason</i>	759
Math Instruction for Children with Special Needs	<i>B A Bottge</i>	767
Overrepresentation of Culturally and Linguistically Diverse Students in Special Education	<i>S C Trent</i>	774
Paraprofessionals in Special Education	<i>B Y Ashbaker, S C Dunn, and J Morgan</i>	780
Parent and Family Involvement in the Education of Children with Special Needs	<i>L P Prater</i>	788
Peer Relations and Socialization of Children and Adolescents with Special Needs	<i>L M Bullock, M Wong-Lo, and R A Gable</i>	794
Peer-Tutoring	<i>L J O'Shea and D J O'Shea</i>	802
Postsecondary Participation of Students with Special Needs	<i>R A Stodden and K D Roberts</i>	808
Prereferral Strategies	<i>C G Simpson and J P Bakken</i>	815

Preventing Antisocial Behavior and Delinquency: A Comprehensive School-wide Approach <i>P R Hallam, K R Young, P Caldarella, D G Wall, and L Christensen</i>	820
Programs and Instructional Strategies for Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	829
Related Services for Children with Special Needs <i>S O Obi</i>	834
Risk and Resilience Frameworks in Understanding Special Education <i>W-Y Hsieh and M L Donahue</i>	840
School Dropout Prevention <i>F E Obiakor</i>	845
School-Based Services for Children with Special Needs <i>S O Obi</i>	850
Self-Determination <i>M L Wehmeyer</i>	857
Social Competence <i>L A Owens and S Johnston-Rodriguez</i>	865
The Identification of Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	870
Transition from School to Adult Life <i>M J McCollin and F E Obiakor</i>	876
Visual Impairment <i>W Sapp</i>	880
Workforce Issues in Special Education <i>E Carlson and B Billingsley</i>	886

VOLUME 3

Education of Professionals

An Overview of Accreditation, Certification, and Licensure Processes <i>J J Norcini and M van Zanten</i>	1
Educating America's Military Officers <i>M H Adamshick</i>	7
Education for Architecture in the United States and Canada <i>S Schreiber</i>	13
Education for Medicine <i>M B Anderson</i>	19
Education for the Legal Profession <i>J W Wegner</i>	25
Education of Clergy <i>D O Aleshire</i>	31
Framing Professions Education <i>P S O'Sullivan</i>	37
Issues in Accounting/Business Education <i>W H Gijsselaers and R G Milter</i>	39
Professions – Engineering <i>J Heywood</i>	45
Professions – Health-Related <i>G M Jensen and C B Royeen</i>	51
Professions – Pharmacy <i>D E Beck</i>	58
Teaching for the Complex Practice of Nursing <i>P Benner and M Sutphen</i>	65
Veterinary Education <i>K M Edmondson</i>	71

Education Research Methodology: Quantitative Methods and Research

Abductive Research Methods <i>B D Haig</i>	77
Field Experimentation in Education <i>M A Constat</i>	83
Internal Validity <i>M H Clark and S C Middleton</i>	90
Methods for Approximating Random Assignment <i>B Schneider and S-K McDonald</i>	97
Methods for Approximating Random Assignment: Regression Discontinuity and Propensity Scores <i>L M Scheier</i>	104

Multitrait–Multimethod Designs	<i>M Bong and S Hong</i>	111
Multivariate Longitudinal Data Analysis	<i>H Liu and L Chang</i>	120
Quasi-Experimentation: Two Group Design	<i>E Hong</i>	128
Regression Discontinuity Designs	<i>A Kelava, S Rohrmann, and V Hodapp</i>	134
Sampling	<i>R S Brown</i>	142
Statistical Conclusion Validity	<i>P L Busk</i>	147
Survey Research Methods	<i>A Fink</i>	152

Educational Assessment

An Overview of Educational Assessment	<i>M James</i>	161
Assessment and the Evaluation of Institutional Effectiveness	<i>S M Thomas</i>	172
Assessment in Schools – Dispositions	<i>R D Crick</i>	181
Assessment in Schools Related To Literacy: Reading	<i>K J Topping</i>	189
Impact of Assessment on Students' Learning Strategies and Implications for Judging Assessment Quality	<i>M Segers, F Dochy, and D Gijbels</i>	196

Educational Assessment – Assessment Activities

Alternative Assessment	<i>G Stobart and C Gipps</i>	202
Classroom Assessment Tasks and Tests	<i>J Phelan and J Phelan</i>	209
Dynamic Assessment	<i>J G Elliott, E L Grigorenko, and W C M Resing</i>	220
Instructional System Provided Feedback	<i>H F O'Neil, S Chuang, and T Huang</i>	226
Peer and Self-assessment	<i>C Harrison</i>	231
Portfolio Assessment	<i>V Klenowski</i>	236
Records of Achievement: Beyond Traditional Tests	<i>P Broadfoot</i>	243

Educational Assessment – Assessment in Domains

Assessment in Higher Education	<i>D R Sadler</i>	249
Assessment in Schools – Affective Domain	<i>M Ventura, R D Roberts, and M Zeidner</i>	256
Assessment in Schools – Creative Subjects	<i>T E Costantino and L Bresler</i>	262
Assessment in Schools – Learning to Learn	<i>A Hautamäki, J Hautamäki, and S Kupiainen</i>	268
Assessment in Schools – Literacy Writing (Extended)	<i>J M Parr and L Earl</i>	273
Assessment in Schools – Mathematics	<i>M Brown and J Hodgen</i>	279
Assessment in Schools – Oracy	<i>A L Bailey</i>	285
Assessment in Schools – Primary Science	<i>W Harlen</i>	293
Assessment in Schools – Secondary Science	<i>J Shemwell, A Fu, M Figueroa, R Davis, and R Shavelson</i>	300
Assessment in Schools – Technology Education and ICT	<i>A Jones, B Cowie, and J Moreland</i>	311
Assessment in the Early Years	<i>M J Drummond</i>	316
Assessment in the History, Civics and Social Studies Domains	<i>P G Avery</i>	323
Assessment in the Workplace of Performance, Developing Expertise and Competence	<i>A J Brown</i>	330
Assessment in Vocational Education	<i>K Ecclestone</i>	337

Educational Assessment – Concepts and Issues

Assessing Group Work	<i>M Galton</i>	342
Assessment and the Regulation of Learning	<i>L Allal</i>	348
Challenges of Developing and Implementing Formative Assessment Practices in Schools	<i>R Sutton</i>	353
Formative Assessment	<i>P Black</i>	359
Formative Assessment and Instructional Planning	<i>B Marshall</i>	365
Progression and Assessment: Developmental Assessment	<i>M Forster and G Masters</i>	369
Students' Rights in Assessment and Evaluation	<i>J Gardner and B Holmes</i>	378
Summative Assessment by Teachers	<i>R Daugherty</i>	384
The Multiple Purposes of Assessment	<i>P E Newton</i>	392
The Relationship between Assessment and the Organization and Practice of Teaching	<i>L B Resnick, D William, R Apodaca, and E S Rangel</i>	397

Educational Assessment – Enhancing Quality and Use

Assessment Conversations: Reading and Writing Conferences with Students and with Parents	<i>M Barrs</i>	403
Assessment: Pre-service and In-service Teacher Education	<i>A Davies</i>	410
Classroom Assessment in Policy Context (Australia)	<i>J J Cumming</i>	417
Classroom Assessment in Policy Context (England and France)	<i>M Osborn and M Raveaud</i>	425
Classroom Assessment in Policy Context (French Sources)	<i>D Laveault</i>	432
Classroom Assessment in Policy Context (Hong Kong)	<i>D Carless</i>	438
Classroom Assessment in Policy Context (New Zealand)	<i>T J Crooks</i>	443
Classroom Assessment in Policy Context (Sub-Saharan Africa)	<i>S J Howie, L Zimmerman, and K Draper</i>	449
Moderation of Student Work by Teachers	<i>G S Maxwell</i>	457
School Policies and Practices to Support Effective Classroom Assessment for Learning	<i>D Pedder</i>	464
Student Assessment: Policy and Practice in Eastern Europe	<i>G Bethell</i>	472

Educational Assessment – Key Relationships

Assessment Practice in Policy Context: Latin American Countries	<i>F M Rizo</i>	479
Impact of Assessment on Learner Groups (Boys/Girls)	<i>J Elwood</i>	486
Impact of Assessment on Learner Groups (Disabilities)	<i>M Rouse</i>	492
Impact of Assessment on Students' Test Anxiety	<i>P Buchwald and C Schwarzer</i>	498
Impact of Assessments on Classroom Practice	<i>J L Herman</i>	506
Social Practices in School Assessment and their Impact on Learner Identities	<i>A Pollard and A Filer</i>	512

Educational Evaluation – Concepts, Practice, and Future Directions

Educational Evaluation: Concepts, Practice, and Future Directions	<i>Y J Ni</i>	518
---	---------------	-----

Educational Evaluation – Contemporary Themes in Evaluation

Evaluating E-Learning	<i>D D Williams and C R Graham</i>	530
Evaluating Schools as Learning Communities	<i>S M Thomas</i>	539
Evaluation of Adult Education and Training Programs	<i>G McNamara, P Joyce, and J O'Hara</i>	548
Evaluation of Integrated Health Programs in School	<i>T M Franke</i>	555

Formative Assessment in Teacher Education and Teacher Professional Development	<i>H Tillema</i>	563
Informal Education and Evaluation	<i>T Jeffs</i>	572
Methods to Evaluate Technology	<i>W L Bewley</i>	578

Educational Evaluation – Contexts of Educational Evaluation

Cost Analysis in Evaluation Studies	<i>J D Fletcher</i>	585
Cultural Issues that Can Affect the Validity of Educational Evaluations	<i>R P Durán</i>	592
Evaluation and Accountability	<i>E Sato and S N Rabinowitz</i>	600
Evaluation Governance and Planning	<i>R Picciotto</i>	607
Evaluation Use	<i>V G Thomas</i>	614
Moral and Ethical Issues in Evaluation	<i>M Morris</i>	622
The Role of Stakeholders in Educational Evaluation	<i>S Taut and M C Alkin</i>	629

Educational Evaluation – Evaluation Domains

Curriculum Evaluation	<i>T Levin</i>	636
Evaluating Education in Three Policy Eras	<i>M A Peters, R Britez, and S M Weber</i>	645
Evaluation of Mathematics Education Programs	<i>J Cai</i>	653
Evaluation of Science Education Programs	<i>G Berberoğlu</i>	660
Evaluation of Social Studies and Civic Education Programs	<i>S G Grant</i>	667
Evaluation of Teacher Quality and Practice	<i>W Robinson and J Campbell</i>	674
International Evaluations	<i>A Schleicher</i>	681
National Assessment Programs: The Example of the U.S. National Assessment of Educational Progress	<i>M Schneider</i>	690
Principal Evaluation: Concepts, Needs and Realities	<i>N S Glasman and L D Glasman</i>	698
Program Evaluation	<i>C A Christie and L A Fierro</i>	706
School Based Evaluation: Purposes, Protocols and Processes	<i>J MacBeath</i>	713
School Inspection/External School Evaluation	<i>C Chapman and P Earley</i>	719
The International System for Teacher Observation and Feedback: A Theoretical Framework for Developing International Instruments	<i>L Kyriakides, B P M Creemers, C Teddlie, and D Muijs</i>	726

Educational Evaluation – Evaluation Methods

Evaluation Methodology	<i>M Chatterji</i>	735
Evaluation Reporting and Communicating	<i>R T Torres</i>	746
Longitudinal Evaluation Designs	<i>X Ma</i>	757
Meta-evaluation: Purpose, Prescription, and Practice	<i>L A Wingate</i>	765

Educational Evaluation – Purposes and Models of Educational Evaluation

Defining Quality in Evaluation	<i>P Dahler-Larsen</i>	775
Internal and External Evaluation	<i>D Nevo</i>	781
Needs Assessment in Education	<i>J W Altschuld and T L Lepicki</i>	786
The Purpose of Educational Evaluation	<i>S Mathison</i>	792
Understanding Approaches to Evaluation	<i>A J Love</i>	798

VOLUME 4**Educational Measurement**

Ability Testing	<i>P B Tymms</i>	1
Admissions Testing	<i>R Zwick</i>	7
Appropriateness Measurement: Person Fit	<i>R R Meijer and A M Weekers</i>	15
Automated Essay Scoring: Writing Assessment and Instruction	<i>M D Shermis, J Burstein, D Higgins, and K Zechner</i>	20
Classical Test Theory Reliability	<i>M D Miller</i>	27
Criterion-Referenced Measurement	<i>S A Shrock and W C Coscarelli</i>	31
Differential Item Functioning	<i>M D Hidalgo and J Gómez-Benito</i>	36
Educational Measurement: Overview	<i>R L Linn</i>	45
Equating and Scaling	<i>A A von Davier</i>	50
Fairness	<i>C Gipps and G Stobart</i>	56
Generalizability Theory	<i>R L Brennan</i>	61
Impact of Assessment on Classroom Practice	<i>J L Herman</i>	69
Intelligent Systems	<i>V J Shute and D Zapata-Rivera</i>	75
Item Response Theory	<i>W J van der Linden</i>	81
National Assessments	<i>T J Crooks</i>	89
Norm-Referenced Measurement	<i>H D Hoover</i>	95
Progress Monitoring	<i>L S Fuchs and D Fuchs</i>	102
Rasch Models	<i>D Andrich</i>	111
Scoring Rubrics	<i>J A Arter</i>	123
Second Language Assessment	<i>L F Bachman</i>	140
Self Adaptive Testing	<i>J Revuelta</i>	148
Student Test Results in School Accountability	<i>L S Hamilton</i>	153
Test Development	<i>S M Downing</i>	159
Test Translation and Adaptation	<i>A Allalouf and P Hanani</i>	166
Testing Creativity	<i>M A Runco</i>	170
Testing in History	<i>S Wineburg and B Fogo</i>	175
Validity	<i>R L Linn</i>	181
Validity of Achievement Gains on High-Stakes Tests	<i>D Koretz</i>	186
Validity of Educational Indicators	<i>B P M Creemers and L Kyriakides</i>	193

G**Globalization of Universities and Institutions**

Globalization, Universities and Medium of Instruction	<i>J Lo Bianco</i>	201
The Washington Accord on Engineering Curriculum	<i>E Ko</i>	209

H

Higher Education

Higher Education: An Overview	<i>P Scott</i>	217
-------------------------------	----------------	-----

Higher Education – Management, Leadership and Governance

Academic Managers – Heads of Departments/Schools	<i>S Marshall</i>	229
Leading and Managing the University – Presidents and their Senior Management Team	<i>R Middlehurst, T Kennie, and S Woodfield</i>	238
Management of and in Higher Education Institutions	<i>B Sporn</i>	245
Managerialism and Collegialism in Higher Education Institutions	<i>M Shattock</i>	251
Strategic Planning in Higher Education	<i>L Zechlin</i>	256
The Changing Role of University Presidents, Vice-Chancellors and Rectors	<i>M F Green and P D Eckel</i>	264
The Changing Roles of University Governing Boards and Councils	<i>J Fielden</i>	273

Higher Education – Research and Scholarship

Funding of University Research	<i>G Harman</i>	279
Mobility of PhD Students and Scientists	<i>S Avveduto</i>	286
National Science Policy and Universities	<i>M Henkel and U Kogan</i>	294
Research Commercialization	<i>S Slaughter</i>	300
Research Quality Assessment	<i>I McNay</i>	307
The Management of University Research	<i>J Kirkland</i>	316
The Training of New Scientists	<i>R Neumann</i>	322

Higher Education – Society

Academic Freedom	<i>S F Macintyre</i>	328
Access and Equity in Higher Education	<i>G Baldwin and R James</i>	334
Boundaries of Institutional Autonomy and their Impact on Higher Education	<i>V L Meek</i>	341
Diversity of Higher Education	<i>U Teichler</i>	347
Employability of University Graduates and Graduate Outcomes	<i>J Brennan</i>	354
Higher Education and the Knowledge Society	<i>J Välimaa</i>	360
Higher Education and the Labor Market	<i>J Enders</i>	365
Higher Education and the Transformation of Society	<i>P Scott</i>	370
Quality Assurance in Higher Education – Practices and Issues	<i>D Dill</i>	377
The Role of Accreditation of Higher Education Institutions	<i>J S Eaton</i>	384
Universities and Regional Development	<i>J Goddard and J Puukka</i>	390
Universities' Engagement with Society	<i>D Watson</i>	398

Higher Education – Teaching and Learning in Higher Education

Distance Education and Open Universities	<i>A Kanwar and J Daniel</i>	404
Education Development in Higher Education	<i>Y Ryan and K Fraser</i>	411
Effective Use of Technology in Teaching and Learning in HE	<i>D Laurillard</i>	419

Facilitating Students' Success – Student Support and Learning Resources in Universities	<i>E Woodberry and C Barnes</i>	427
Faculty Values and Expectations and Their Implications for Teaching, Learning and the Organization of Higher Education Institutions	<i>K Harman</i>	433
Flexible Learning in Higher Education	<i>S Alexander</i>	441
Innovation in Teaching and Curriculum Design	<i>J M Dearn</i>	448
Productivity of University Faculty Staff	<i>S Kyvik</i>	455
Promoting Effective Student Learning in Higher Education	<i>K Trigwell</i>	461
Student Attrition from Higher Education Institutions	<i>M Hayden</i>	467
University Students' Experiences of Higher Education	<i>C McInnis</i>	473
What Do University Students Learn?	<i>G D Kuh and B K Bridges</i>	479

Higher Education – Tertiary Education in a Globalized World

Associations of Universities	<i>G F Moodie</i>	485
Collective Bargaining in Higher Education	<i>C Allport, K McAlpine, and S Roberts</i>	491
Financing of Higher Education	<i>B Chapman and P Tulip</i>	499
Higher Education Crossing Borders	<i>J Knight</i>	507
Higher Education in Post-Conflict Societies	<i>M Daxner</i>	514
Higher Education in the New Economies	<i>J Fielden</i>	520
Human Resource Issues in Higher Education	<i>E El-Khawas</i>	527
International Accreditation in Higher Education	<i>A Stella and D Woodhouse</i>	533
Internationalization of Higher Education	<i>M van der Wende</i>	540
National and International Rankings of Higher Education	<i>S Marginson</i>	546
Steering of Higher Education Systems – The Role of the State	<i>K Yokoyama and V L Meek</i>	554
Student and Faculty Transnational Mobility in Higher Education	<i>U Lanzendorf and B M Kehm</i>	559
The Bologna Process in European Higher Education	<i>D Rich</i>	566
The Economics and Finance of Higher Education	<i>A Wolf</i>	573
The Role of the OECD in the Development of Higher Education in a Globalized World	<i>R Yelland</i>	584
Transforming Higher Education in Developing Countries: The Role of the World Bank	<i>J Salmi and R M Bassett</i>	590
UNESCO's Role in the Development of Higher Education in a Globalized World	<i>S Uvalić-Trumbić</i>	597

Higher Education – The Organization and Financing of Higher Education

Higher Education in Federal Systems	<i>M A M Cortés and U Teichler</i>	603
Pathways and Articulation into Higher Education	<i>G F Moodie</i>	609
Unitary, Binary and other National Systems of Higher Education	<i>L Goedegebuure</i>	616
Viewing Private Higher Education: How Much, Where, Why, and What?	<i>D C Levy</i>	622

I

Immigration

Children of Migrant Populations	<i>M M Suárez-Orozco and C Suárez-Orozco</i>	629
Third Culture Kids	<i>R E Van Reken</i>	636

International Comparisons

IEA Studies in ICT	<i>N Law</i>	644
IEA Studies in Mathematics and Science	<i>F K S Leung</i>	650
IEA Study in Civic Education	<i>J Torney-Purta, J-A Amadeo, and J Schwillie</i>	656
IEA: Globalization and Assessment	<i>H Wagemaker</i>	663

International Organizations in Education

Education and the European Union	<i>S Hegarty</i>	669
National University on Educational Planning and Administration, New Delhi	<i>J B G Tilak</i>	676
On School Management – International Council on School Effectiveness and Improvement, Commonwealth Council on Educational Administration and Management	<i>A Walker and T Townsend</i>	681
UNESCO's International Institute for Educational Planning	<i>S J Hite and M Bray</i>	688

L**Leadership and Management**

Leadership and Management Overview	<i>B Mulford</i>	695
------------------------------------	------------------	-----

Leadership and Management – as a Field of Study

Contexts for Leadership at the Beginning of the 21st Century	<i>M Bottery</i>	704
Educational Leadership: Philosophical Issues	<i>C W Evers</i>	710
Ethical/Moral Issues in Educational Leadership	<i>R J Starratt, L Langlois, and P Duignan</i>	717
History of Educational Leadership/Management	<i>R Bates</i>	724
Educational Leadership as a Field of Study: Professional Journals 1994–2006	<i>A R Thomas</i>	731
Research on Educational Leadership – Approaches/Promising Directions	<i>W A Firestone and V M J Robinson</i>	740

Leadership and Management – Leader Preparation

Leaders for Productive Schools	<i>J Murphy, S N Elliott, E B Goldring, and A C Porter</i>	746
New Approaches in Preparing School Leaders	<i>S G Huber</i>	752
Research for Leaders: The National College for School Leadership	<i>G Southworth</i>	762
School Leadership Standards	<i>L Ingvarson</i>	768

Leadership and Management – Leadership Issues

Effective Leadership in Challenging Schools	<i>A Harris</i>	777
Esthetics, Educational Leadership and Management	<i>R Bates</i>	783
Leader Mistreatment of Teachers	<i>J Blase and J Blase</i>	790
Leadership and Gender	<i>J Blackmore</i>	797
Leadership and Student Voice	<i>C Lanskey and J Rudduck</i>	803
Leadership and Urban Education	<i>K A Riley</i>	809
Leadership for Learning	<i>J MacBeath</i>	817
Leadership in Diverse Cultures	<i>A Walker and P K Yu-kwong</i>	824

Leading Diverse Schools	<i>J Ryan</i>	832
Supervision and Personnel Administration	<i>P D Tucker and D G Pounder</i>	839
The Big Challenge: Leadership for Inclusion	<i>M Ainscow and A Sandill</i>	846
The Educational Leader's New Work	<i>P Gronn</i>	852

VOLUME 5

Leadership and Management – Leadership Types

A Distributed Perspective on School Leadership and Management	<i>J P Spillane</i>	1
Leadership: Authentic	<i>P T Begley</i>	7
Leadership: Democratic	<i>J Møller</i>	12
Leadership: Instructional	<i>P Hallinger, K Leithwood, and R H Heck</i>	18
Leadership: Transformative	<i>C M Shields</i>	26
Strategic Leadership	<i>B Davies and B J Davies</i>	34
Teacher Leadership and Organizational Development	<i>A Harris</i>	40
Understanding How Leadership Influences Student Learning	<i>K Leithwood and B Levin</i>	45

Leadership and Management – Politics and Governance

Equity and Educational Effectiveness	<i>P Sammons</i>	51
Leading/Managing Schools in Communities Made Poor	<i>P Thomson</i>	58
Principals Role in Restructuring Schools	<i>G M Crow and K D Peterson</i>	66
School Reform and Restructuring: Self Managing School	<i>B J Caldwell</i>	72
System Leadership	<i>D Hopkins and R Higham</i>	78

Leadership and Management – School and Community

Educational Leadership and Social Capital	<i>J West-Burnham</i>	85
Networked Learning Communities School-to-School Collaboration as an Essential Component of a System Reform Strategy	<i>D Jackson</i>	92
Networks and Communities of Knowledge	<i>T Schuller and H Theisens</i>	100
School Community Relations	<i>R L Crowson and E B Goldring</i>	106
Social Capital, Educational Institutions and Leadership	<i>S Kilpatrick, S Johns, and B Mulford</i>	113

Leadership and Management – School Effectiveness and Improvement

Leadership and Technology	<i>C F Webber</i>	120
Leadership in the Implementation of Innovations	<i>I E Bogotch, T Townsend, and M Acker-Hocevar</i>	128
Leadership: School Improvement	<i>R H Heck and P Hallinger</i>	135
Organizational Learning in Schools	<i>B Mulford and H Silins</i>	143
Professional Learning Community	<i>L Stoll</i>	151
Transformational School Leadership	<i>K Leithwood</i>	158

Learning and Cognition – Issues and Concepts

Attention in Cognition and Early Learning	<i>M Heimann, T Tjus, and K Strid</i>	165
Concept Learning	<i>S J Unsworth and D L Medin</i>	172
Cooperative Learning	<i>R E Slavin</i>	177
Intelligence	<i>R J Sternberg</i>	184
Learning from Multiple Information Sources	<i>H I Strømsø and I Bråten</i>	191
Memory	<i>S Magnussen and T Brennen</i>	197
Metacognition	<i>L Baker</i>	204
Personal Epistemology in Education: Concepts, Issues, and Implications	<i>I Bråten</i>	211
Piaget: Recent Work	<i>L Smith</i>	218
The Adult Development of Cognition and Learning	<i>E A L Stine-Morrow and J M Parisi</i>	225
Vygotsky and Recent Developments	<i>J V Wertsch</i>	231

Learning and Cognition – Issues and Concepts – Focus on Cognition

Cognition and Emotion	<i>F Pons, M de Rosnay, and F Cuisinier</i>	237
Cognition: Overview and Recent Trends	<i>J Wiley and B D Jee</i>	245
Development of Creativity	<i>M A Runco</i>	251
Knowledge Domains and Domain Learning	<i>L Maggioni and P A Alexander</i>	255
Problem Solving and Human Expertise	<i>T J Nokes, C D Schunn, and M T H Chi</i>	265
Problem Solving and Reasoning	<i>R E Mayer</i>	273
Reasoning, Moral and Social	<i>N S Jampol, C B Richardson, and M Killen</i>	279
Remembering as Social Activity	<i>D Middleton</i>	285

Learning and Cognition – Issues, Concepts, Types

Computer-Supported Collaborative Learning: Basic Concepts, Multiple Perspectives, and Emerging Trends	<i>S R Ludvigsen and A I Mørch</i>	290
Learning as Inquiry	<i>S R Goldman, J Radinsky, S Tozer, and D Wink</i>	297
Learning Through Narrative Socialization	<i>A Fasulo</i>	303
Neuroscience Bases of Learning	<i>M H Immordino-Yang and K W Fischer</i>	310
The Neuroscience of Reading	<i>U Goswami</i>	317

Learning and Cognition – Issues, Concepts, Types – Focus on Learning

Learning Strategies	<i>C E Weinstein, J Jung, and T W Acee</i>	323
Learning Through Play	<i>K Aronsson</i>	330
Situated View of Learning	<i>A Collins and J G Greeno</i>	335
The Neuroscience of Aging and Cognition	<i>W J Hoyer and B Bessette-Symons</i>	340

Learning and Cognition – Language, Literacy and Subject-Related

An Overview of Language and Literacy in Educational Settings	<i>V G Aukrust</i>	345
Bilingualism and Learning	<i>L S Verplaetse and E Schmitt</i>	355
Classroom Discourse and Student Learning	<i>L Hemphill</i>	361
Early Writing	<i>B E Hagtvet</i>	367

First Language Acquisition	<i>N B Ratner</i>	375
Foreign Language Learning	<i>C Kinginger</i>	382
Learning Science	<i>J Leach and P Scott</i>	387
Learning to Read	<i>M S Burns and J K Kidd</i>	394
Mathematics Learning	<i>L Verschaffel, B Greer, and E De Corte</i>	401
Music Learning	<i>S Hallam</i>	407
Reading Comprehension: Reading for Learning	<i>C E Snow</i>	413
Second Language Learning	<i>J S Arnfast, J N Jørgensen, and A Holmen</i>	419
Writing: Advanced	<i>R Andrews</i>	426

Learning and Cognition – Other Issues

Gender and Learning	<i>L Yates and J McLeod</i>	432
Parent–Child Relationships in Early Learning	<i>C P Edwards, S M Sheridan, and L L Knoche</i>	438
Peer Interaction and Learning	<i>S Blum-Kulka and S Dvir-Gvirsman</i>	444

Learning and Cognition – Places of Learning

Learning in a Cross-Cultural Perspective	<i>L-F Zhang and R J Sternberg</i>	450
Learning Outside of School	<i>T Nunes</i>	457
Reasoning and Explanatory Talk: Learning in Everyday Settings	<i>D E Beals</i>	464

Learning and Cognition – Theoretical Perspectives – Learning

Apprenticeship Approach to Learning	<i>K Nielsen</i>	469
Connectionism and Learning	<i>T R Shultz</i>	476
Constructivism and Learning	<i>S Sjøberg</i>	485
Cultural–Historical Activity Theory	<i>A Havnes</i>	491
Learning in a Sociocultural Perspective	<i>R Säljö</i>	498
Self-Regulated Learning and Socio-Cognitive Theory	<i>P H Winne and A F Hadwin</i>	503

N

National Systems of Education

Australia	<i>C N Power</i>	510
Austria	<i>M J Gaertner</i>	517
Bangladesh	<i>R Z Dewan</i>	524
Barbados	<i>R Rudder</i>	534
Bulgaria	<i>A Chavdarova and D Vesselinov</i>	540
England	<i>C Whatford</i>	550
Estonia	<i>M Keränen and J Kõrgesaar</i>	555
Fiji	<i>I Senibulu</i>	563
Finland	<i>J E Tuovinen</i>	569
France	<i>M Sellier</i>	576

Germany	<i>A Hüfner and K Hüfner</i>	582
Ghana	<i>R Palmer</i>	590
Greece	<i>K Toura and A Karagiannis</i>	599
Iceland	<i>M Hardardottir and M Gunnlaugsdóttir</i>	605
Indonesia	<i>A C Alwasilah and Furqon</i>	611
Israel	<i>A Volansky</i>	618
Jamaica	<i>E Hannam</i>	627
Latvia	<i>A Joma</i>	630
Lebanon	<i>N A Mneimneh</i>	636
Lithuania	<i>J Navickaitė, L Paurienė, R Zablackė, A Tijūnelytė, and V Kožemiakina</i>	656
Malawi	<i>F R Mkandawire and D Mulera</i>	666
Malta	<i>C Farrugia</i>	673
Mongolia	<i>B Yembuu</i>	681
Namibia	<i>J Ellis</i>	687
New Zealand	<i>R Baker and C Wylie</i>	692
Niger	<i>A A I Daouda</i>	699
Norway	<i>H Farstad</i>	709
Oman	<i>S A Issan</i>	718
Pakistan	<i>N Mahmood and R Malik</i>	727
Poland	<i>A Janowski</i>	737
Qatar	<i>S R Salman</i>	744
Russia	<i>V M Filippov and G F Tkach</i>	750
Rwanda	<i>E Bahizi</i>	758
Scotland	<i>D Raffe</i>	770
Singapore	<i>P T Ng</i>	776
Slovenia	<i>T Plevnik and A B Lakota</i>	782
Solomon Islands	<i>D P Malasa</i>	790
South Africa	<i>P Kgobe and J Pampallis</i>	799
Spain	<i>A Tiana</i>	806
Sri Lanka	<i>B M J Balasooriya, M M Wehella, and S U Wijeratne</i>	813
Suriname	<i>H R Ori</i>	826
Taiwan	<i>C-M Lee</i>	834
Tanzania	<i>R Wedgwood</i>	841
Thailand	<i>C Rukspollmuang</i>	847
The Gambia	<i>Y A-M Jobe</i>	858
The Philippines	<i>E A P Valenzuela</i>	865
Uruguay	<i>B Guinovart and C Troncoso</i>	873
USA	<i>B Pusser and M E R Lloyd</i>	888
Zanzibar	<i>I M Hija</i>	896

VOLUME 6**P****Philosophy of Education**

Philosophy of Education: Overview	<i>P Enslin</i>	1
-----------------------------------	-----------------	---

Philosophy of Education – Contemporary Issues

Neoliberalism, the Market and Performativity	<i>M A Peters</i>	11
Parents and Children's Rights	<i>M Clayton</i>	17
Peace Education	<i>I Gur-Ze'ev</i>	22
Philosophy and Educational Research	<i>D Bridges</i>	29
Race, Critical Race Theory and Whiteness	<i>B Applebaum</i>	36
School Choice and the Common School	<i>G Haydon</i>	44

Philosophy of Education – Philosophical Perspective

African Philosophy of Education	<i>K Horsthemke</i>	50
Feminism	<i>L Stone</i>	56
Hermeneutics	<i>C Bingham</i>	63
Islamic Education	<i>Y Waghid</i>	69
Liberalism and Education	<i>W Feinberg</i>	74
Phenomenology and Existentialism	<i>B Levering</i>	80
Pragmatism	<i>B Thayer-Bacon</i>	86
Psychoanalysis and Education	<i>P Smeyers</i>	92
Social and Cultural Capital in Education	<i>M Print</i>	97
The Aims of Education	<i>R Marples</i>	103
The Analytical Tradition	<i>D Carr</i>	108

Philosophy of Education – Philosophical Themes

Adult Learning: Philosophical Issues	<i>D Beckett</i>	114
Autonomy	<i>C A Wringe</i>	121
Bildung	<i>J Masschelein and N Ricken</i>	127
Critical Theory and Pedagogy	<i>M Mason</i>	133
Critical Thinking	<i>H Siegel</i>	141
Environmental Education	<i>M Bonnett</i>	146
Identity	<i>M S Merry</i>	152
Inclusion	<i>R Cigman</i>	158
Justice and Care	<i>M Tjiattas</i>	164
Lifelong Learning	<i>J Halliday</i>	170
Post-Structuralism and Education	<i>M Mason and M Clarke</i>	175
Reason and Rationality	<i>P Hogan</i>	183

Religious and Spiritual Education	<i>J C Conroy and S J McKinney</i>	189
The Capabilities Approach	<i>L Terzi</i>	194

Primary and Secondary Education – Learning and Teaching in School Age Education

Class Size – Arguments and Evidence	<i>P Blatchford and K C Lai</i>	200
Continuing Professional Development of Teachers	<i>P Earley</i>	207
Homeschooling	<i>P Farenga</i>	214
Innovations and Early Intervention in the Teaching of Literacy	<i>R E Slavin</i>	221
Personalized Learning in School Age Education	<i>D Hopkins</i>	227
Perspectives on Schooling in the Middle Years	<i>V J Russell</i>	233
Reform of the School Workforce	<i>H M Gunter and G Forrester</i>	242
Secondary Education in Developing Countries	<i>P Hughes and A Walker</i>	249

Primary and Secondary Education – Social Justice Issues in School Age Education

Community Focused Schools	<i>A Dyson</i>	255
---------------------------	----------------	-----

Primary and Secondary Education – The Effectiveness of School Systems

Reforming the American School System	<i>M Barber</i>	261
School Effectiveness in Developed Societies	<i>D Reynolds</i>	269
School to School Collaboration: Innovation and Improvement	<i>R Higham</i>	274
Studies of School Improvement in Developing Countries	<i>S E Anderson</i>	281
Sustainable Educational Reform	<i>A Hargreaves</i>	289
The Role of the District in Tri Level Reform	<i>M Fullan</i>	295
Whole School Designs for Enhancing Student Achievement	<i>S Stringfield and J Nunnery</i>	303

Q

Qualitative Research

Action Research in Education	<i>E Stringer</i>	311
Classroom Ethnography	<i>F Erickson</i>	320
Computer Assisted Qualitative Data Analysis	<i>C Silver and A Lewins</i>	326
Conversational Analysis	<i>H G Levine</i>	335
Critical Ethnography	<i>D Foley</i>	337
Critical Race Theory	<i>D Gillborn and G Ladson-Billings</i>	341
Critical Theory	<i>J J Ryoo and P McLaren</i>	348
Cultural Studies	<i>R Gaztambide-Fernández and K Cairns</i>	354
Cultural–Historical Activity Theory	<i>Laboratory for Comparative Human Cognition</i>	360
Design Experiments	<i>A Collins</i>	367
Discourse Analysis	<i>P Nikander and S Vehviläinen</i>	373
Document Analysis	<i>T Rapley and K N Jenkins</i>	380
Ethnography	<i>M Hammersley</i>	386

Ethnomethodology in Education Research	<i>D Macbeth</i>	392
Focus Groups	<i>D L Morgan and K Hoffman</i>	401
Grounded Theory	<i>K Charmaz and A Bryant</i>	406
Hermeneutics	<i>J Dyer</i>	413
Interpretive Research	<i>E Rauscher and M E Graue</i>	419
Interviews and Interviewing	<i>A Marvasti</i>	424
Life History	<i>J A Hatch and S K Newsom</i>	430
Narrative Inquiry	<i>D J Clandinin and J Huber</i>	436
Participant Observation	<i>J Schostak</i>	442
Phenomenology	<i>M van Manen and C A Adams</i>	449
Qualitative Case Studies	<i>S B Merriam</i>	456
Qualitative Data Management	<i>C Grbich</i>	463
Semiotics	<i>M Danesi</i>	475
Symbolic Interactionism, Naturalistic Inquiry, and Education	<i>J A Forte</i>	481
The Quality of Evidence in Qualitative Research	<i>S K Biklen</i>	488
Visual Data in Education Research	<i>J Wagner</i>	498

Qualitative Research – Standards of Proof

Validity: Mapping Diverse Perspectives	<i>G B Rossman, S F Rallis, and A M Kuntz</i>	505
--	---	-----

Quantitative and Qualitative Approaches to Research – Integration

Mixed Methods	<i>A Tashakkori and I Newman</i>	514
---------------	----------------------------------	-----

Quantitative Methods and Research Design

External Validity	<i>G E Matt, A Brewer, and M Sklar</i>	521
-------------------	--	-----

R

Role of NGOs in Globalization

NGOs and Globalization of Education	<i>M Ahmed</i>	529
-------------------------------------	----------------	-----

S

Social and Emotional Aspects of Learning

Achievement Goal Theory: Definitions, Correlates, and Unresolved Questions	<i>T Urdan</i>	537
Affect, Mood and Emotions	<i>S Schnall</i>	544
Anxiety	<i>M Zeidner</i>	549
Attribution Theory	<i>B Weiner</i>	558
Children's Friendship	<i>S R Asher and K L McDonald</i>	564
Coping with Stressful Situations: An Important Aspect of Self-Regulation	<i>M Boekaerts</i>	570
Culture in Motivation Research: A Challenging and Enriching Contribution	<i>K Kimmel and S Volet</i>	576
Early Social Development and Schooling	<i>G M van der Aalsvoort</i>	585

Emotion in Educational Contexts	<i>P A Schutz, P D Quijada, S de Vries, and M Lynde</i>	591
Empathy	<i>R Arnold</i>	597
Flow in Education	<i>J A Schmidt</i>	605
Interest	<i>M Ainley</i>	612
Intrinsic and Extrinsic Motivation	<i>C Levesque, K J Copeland, M D Pattie, and E L Deci</i>	618
Motivating Students in Classrooms	<i>J Brophy</i>	624
Motivation Regulation	<i>C A Wolters and S A Mueller</i>	631
Peer Learning in the Classroom	<i>N M Webb</i>	636
Poverty, Effects of on Social and Emotional Development	<i>N Philipsen Hetzner, A D Johnson, and J Brooks-Gunn</i>	643
Seeking Help as an Adaptive Response to Learning Difficulties: Person, Situation, and Developmental Influences	<i>S A Karabenick and R S Newman</i>	653
Self-Concept in Learning: Reciprocal effects model between academic self-concept and academic achievement	<i>H W Marsh and L F Scalas</i>	660
Self-Efficacy Beliefs	<i>D H Schunk and F Pajares</i>	668
Social and Emotional Outcomes of Learning	<i>J Vorhaus</i>	673
Social Aspects of Collaborative Learning	<i>M Arvaja and P Häkkinen</i>	685
Social Development and Schooling	<i>G W Ladd</i>	691
Social Interaction and Learning	<i>K Littleton</i>	698
Social Networks and the Education of Children and Youth	<i>B Schneider, T G Ford, and L Perez-Felkner</i>	705
Sociocultural Issues in Motivation	<i>R A Walker</i>	712
Socioemotional Aspects of Technology-Supported Learning	<i>S Järvelä</i>	718
Volitional Control of Learning	<i>J Husman and L Corno</i>	724
Wellbeing	<i>T Hascher</i>	732

VOLUME 7

Statistics

An Overview of Statistics in Education	<i>S Sinharay</i>	2
Analysis and Interpretation of Multivariate Data	<i>D J Bartholomew</i>	12
Analysis of Covariance	<i>N T Longford</i>	18
Analysis of Extreme Values in Education	<i>J Beirlant, G Dierckx, K Luwel, and P Onghena</i>	25
Analysis of Variance	<i>B M King</i>	32
Bayesian Statistical Analysis	<i>G Kokolakis</i>	37
Bootstrap Method	<i>K Singh and M Xie</i>	46
Canonical Correlation	<i>A Basu and A Mandal</i>	52
Categorical Data Analysis	<i>S E Fienberg and B W Junker</i>	58
Causal Inference	<i>D B Rubin</i>	66
Cluster Analysis: Overview	<i>S Landau and I Chis Ster</i>	72

Cognitive Psychology and Educational Statistics	<i>H S Kim and S Embretson</i>	84
Computational Statistics	<i>J E Gentle</i>	93
Continuous Probability Distributions	<i>S Sinharay</i>	98
Correspondence Analysis	<i>M Greenacre</i>	103
Data Mining	<i>R S J d Baker</i>	112
Decision Theory	<i>M Mendoza and E Gutiérrez-Peña</i>	119
Design of Experiments	<i>C A Albers and T R Kratochwill</i>	125
Discrete Probability Distributions	<i>S Sinharay</i>	132
Discrimination and Classification	<i>R C Neath and M S Johnson</i>	135
Empirical Bayes Methods	<i>E Ip and G Molenberghs</i>	142
Evaluation Research	<i>T Kellaghan</i>	150
Exploratory Data Analysis	<i>A Unwin</i>	156
Factor Analysis: An Overview and Some Contemporary Advances	<i>F A N Alhija</i>	162
Generalized Linear Mixed Models	<i>S Rabe-Hesketh and A Skrondal</i>	171
Generalized Linear Models	<i>J J Faraway</i>	178
Generating Random Numbers	<i>P J Pashley and A Amodeo</i>	184
Goodness-of-Fit Testing	<i>A Maydeu-Olivares and C García-Forero</i>	190
Graphical Models	<i>R G Almond</i>	197
Growth Modeling	<i>I Plewis</i>	203
Hierarchical Linear Models	<i>D Rindskopf</i>	210
Hypothesis Testing and Confidence Intervals	<i>T R Konold and X Fan</i>	216
Instrumental Variables	<i>E P Bettinger</i>	223
Jackknife Methods	<i>S Sinharay</i>	229
Large-sample Statistical Methods	<i>J M Singer and D F Andrade</i>	232
Latent Class Models	<i>J K Vermunt</i>	238
Markov Chain Monte Carlo	<i>A M Johansen</i>	245
Matrix Algebra	<i>H Yanai and H Ishii</i>	253
Measure of Association	<i>G-H Huang</i>	260
Measures of Central Tendency	<i>D Clark-Carter</i>	264
Measures of Dispersion, Skewness and Kurtosis	<i>D Rindskopf and M Shiyko</i>	267
Meta Analysis	<i>J Sánchez-Meca and F Marín-Martínez</i>	274
Missing Data	<i>C A W Glas</i>	283
Model Selection	<i>S J Haberman</i>	289
Monte Carlo Methods	<i>A M Johansen</i>	296
Multidimensional Scaling	<i>Z Zhang and Y Takane</i>	304
Multiple Comparisons	<i>C Lewis</i>	312
Multivariate Analysis of Variance	<i>N T Longford</i>	319
Multivariate Linear Regression	<i>S Ganesh</i>	324
Multivariate Normal Distribution	<i>J Tacq</i>	332

Nonlinear Regression Analysis	<i>H-H Huang, C K Hsiao, and S-Y Huang</i>	339
Nonparametric Statistical Methods	<i>K Sijtsma and W H M Emons</i>	347
Observational Studies	<i>P R Rosenbaum</i>	354
Order Statistics	<i>I Bairamov</i>	360
Point Estimation Methods with Applications to Item Response Theory Models	<i>F Bartolucci and L Scrucca</i>	366
Principal Components Analysis	<i>D J Bartholomew</i>	374
Probability Theory	<i>T Rudas</i>	378
Recursive Partitioning	<i>S Landau and S Barthel</i>	383
Sampling	<i>J Qian</i>	390
Sequential Probability Ratio Test	<i>T J H M Eggen</i>	396
Sequential Testing	<i>H J Vos</i>	401
Signal Detection Theory	<i>H Abdi</i>	407
Small Area Estimation	<i>A Noble</i>	411
Statistical Analysis of Functional Data	<i>F Ferraty, P Sarda, and P Vieu</i>	417
Statistical Inequalities	<i>H M Mahmoud and J L Gastwirth</i>	426
Statistical Paradoxes	<i>Y-K Tu</i>	429
Statistical Power Analysis	<i>L V Hedges and C Rhoads</i>	436
Statistical Significance Versus Effect Size	<i>X Fan and T R Konold</i>	444
Stochastic Processes	<i>F Baudoin</i>	451
Structural Equation Models	<i>S-Y Lee and X-Y Song</i>	453
Survival Data Analysis	<i>A F De Champlain</i>	459
The Normal Distribution and its Applications	<i>J Tacq</i>	467
Time Series Analysis	<i>P J Brockwell</i>	474
Univariate Linear Regression	<i>M L Hazelton</i>	482
Value-Added Models	<i>D F McCaffrey</i>	489

T

Teacher Education – Evaluation in the Teacher Education Program

Accreditation and Standards in Teacher Education	<i>R J Yinger and K L Daniel</i>	495
Determining Long Term Effects of Teacher Education	<i>C N Brouwer</i>	503

Teacher Education – Evaluation of the Prospective Teacher

The Evaluation of Prospective Teachers in Teacher Education	<i>D E H Tigelaar and J van Tartwijk</i>	511
---	--	-----

Teacher Education – Macro Factors

Diversity in Teacher Education	<i>T Wubbels</i>	518
Social and Cultural Influences on Teacher Education	<i>M Ben-Peretz and R Lotan</i>	525
The Role of Governance in Teacher Education	<i>G E Karlsen</i>	532

Teacher Education – Post-Initial Professional Development

Action Research as a Tool for Teachers' Professional Development	<i>P Ponte</i>	540
Contemporary Approaches to Teacher Professional Development	<i>H Borko, J Jacobs, and K Koellner</i>	548
School Development for Teacher Learning and Change	<i>P Sleegers and K Leithwood</i>	557
Teacher Induction	<i>D Beijaard, J Buitink, and C Kessels</i>	563
Teacher Learning as Workplace Learning	<i>J Imants and K van Veen</i>	569
Teacher Learning with Lesson Study	<i>A Murata</i>	575
Teachers Career Stages and Professional Development	<i>R Fessler and E Rice</i>	582

Teacher Education – Preservice Teacher Education

A Pedagogy of Teacher Education	<i>J Loughran</i>	587
Case Methods in Teacher Education	<i>M Sato and C Rogers</i>	592
Conceptions of Teacher Education	<i>J A Whitcomb</i>	598
Inquiry-Oriented Teacher Education	<i>R Schulz</i>	604
Narratives and Biography in Teacher Education	<i>G Kelchtermans</i>	610
Supervision in Teacher Education	<i>J McCarthy and L F Quinn</i>	615
Taking Prospective Teachers' Beliefs into Account in Teacher Education	<i>M Valcke, G Sang, I Rots, and R Hermans</i>	622
Teacher Education and Models of Teacher Reflection	<i>V K LaBoskey</i>	629
Teacher Education as Teaching for Understanding with New Technologies	<i>M S Wiske and D E Spicer</i>	635

Teacher Education – Preservice: The Knowledge Base

Experienced Teachers' Craft Knowledge	<i>P C Meijer</i>	642
Moral Values in Teacher Education	<i>W Veugelers</i>	650
Pedagogical Content Knowledge	<i>J H Van Driel and A Berry</i>	656
Teacher Education and the Educational Foundations Knowledge Base	<i>F B Murray</i>	662
The Relationship Between Theory and Practice in Teacher Education	<i>F A J Korthagen</i>	669

Teacher Education – Teacher Educators

Characteristics, Scholarship and Research of Teacher Educators	<i>M Lunenberg</i>	676
Professional Development of Teacher Educators	<i>P K Smith</i>	681
Self-Study by Teacher Educators	<i>T Russell</i>	689

Teacher Education – Teaching Specific Domains

Arts Teacher Education	<i>L Beudert</i>	695
Mathematics Teacher Education	<i>K Krainer and S Llinares</i>	702
Music Teacher Education	<i>M W-y Wong</i>	706
Science Teacher Education	<i>J H Van Driel and S K Abell</i>	712
Social Studies Teacher Education	<i>A-L Halvorsen and S M Wilson</i>	719
Teacher Education for Elementary Education	<i>K D Brown and G N Carnes</i>	725

Teacher Education for Teaching the Gifted	<i>A Y Baldwin</i>	732
Teacher Preparation in Mother Tongue Education	<i>M Kooy and H Hulshof</i>	738
Vocational Teacher Education	<i>S Johansson</i>	744

Teacher Education – The Cooperating School

Mentoring in Teacher Education	<i>P D Tomlinson, A J Hobson, and A Malderez</i>	749
Partnerships Between Schools and Higher Education	<i>C Day and L Smethem</i>	757

VOLUME 8

Technology and Learning

An Overview of Technology and Learning	<i>B Means and J Roschelle</i>	1
Educational Data Modeling	<i>S Iksal</i>	11
Student Data Systems and Their Use for Educational Improvement	<i>J C Wayman, V Cho, and M P Richards</i>	14
Technology and Learning: Access in Schools Around the World	<i>R Anderson and J Ainley</i>	21

Technology and Learning – as a Subject of Instruction

Conceptions of Technology Literacy and Fluency	<i>O Erstad</i>	34
--	-----------------	----

Technology and Learning – Assessment

Technology and Formative Assessment	<i>J W Pellegrino</i>	42
Technology for Large-Scale Assessment	<i>R E Bennett</i>	48
Technology Supports for Assessment Design	<i>R J Mislevy, F I Winters, I I Bejar, R E Bennett, and G D Haertel</i>	56

Technology and Learning – Education Reform and Economic Development

National Strategies to Build a Technology Workforce	<i>H-M Cheah</i>	66
Public–Private Partnerships for Educational Reform	<i>V Quah, C P Lim, and C Brook</i>	73
Relating Technology, Education Reform and Economic Development	<i>R B Kozma</i>	81

Technology and Learning – Evaluating Technology

Examining the Effects of Educational Technology Programs: Challenges and Strategies	<i>M Russell</i>	88
---	------------------	----

Technology and Learning – Global Trends

Equity in Technology Access and Opportunities	<i>T Matuchniak and M Warschauer</i>	95
Internet-Based Education	<i>M Bakia</i>	102

Technology and Learning – Issues

Media Use and School Achievement	<i>K Roe</i>	109
Technology and Physical and Social Health	<i>K Subrahmanyam</i>	112

Technology and Learning – Meeting Special Students’ Needs

Universal Design for Learning *D H Rose and J W Gravel* 119

Technology and Learning – School Leadership for Technology Integration

Technology as a Support for School–Community Connections *K M Culp* 125

Technology and Learning – Supports for Managing Instruction

Classroom Uses of Technology to Manage Instruction *W R Penuel* 133

Technology and Learning – Supports for Skill Learning

New Media, Learning from *W Schnotz and H Horz* 140

Reading and Technology *N DeFrance, D Khasnabis, and A S Palincsar* 150

Technological Supports for Acquiring Twenty-First-Century Skills *C Dede* 158

Technology Supports for Acquiring Inquiry Skills *T de Jong* 167

Technology Supports for Acquiring Mathematics *M J Nathan* 172

Technology Supports for Lifelong Learning *A Collins and R Halverson* 184

Technology and Learning – Supports for Student Collaboration

Knowledge Communities in the Classroom *J D Slotta and H Najafi* 189

Technology and Learning – Supports for Subject Matter Learning

Technology Supports for Science Learning *C Delgado and J Krajcik* 197

Technology Supports for Second Language Learning *D Nunan* 204

Technology and Learning – Teacher Learning

Teacher Skills and Knowledge for Technology Integration *N Law* 211

Technology in Preservice Teacher Education *N Davis* 217

Technology Resources for Teacher Learning *J E Hinojosa, C Labbé, and L López* 222

V

Vocational Education and Training

An Overview of Vocational Education and Training *T Karmel* 229

Informal and Non-Formal Learning in Vocational Education and Training *F Beddie and S Halliday-Wynes* 240

School-Based Vocational Education and Training *B Knight and R Sweet* 247

The Economics of Vocational Education and Training *P Carneiro, L Dearden, and A Vignoles* 255

Vocational Education and Training – Individuals and Students

Age Equality in Education and Training *P Taylor* 262

Career Guidance in Vocational Education and Training and in the Workplace *P Plant* 269

Displaced Workers, Unemployed and Vocational Education and Training *B Kriechel* 275

Immigrant Investment in Host Country Vocational Education and Training *I Kogan* 280

Indigenous People and Vocational Education and Training	<i>R G Schwab</i>	289
Persons with a Disability and Vocational Education and Training	<i>F Ferrier and C S Smith</i>	294
Prisoner Education and Training	<i>S Dawe, S H Wynes, and J Misko</i>	300
Vocational Education and Training and the School-to-Work Transition	<i>G Hayward</i>	306

Vocational Education and Training – Industry and Employers

Apprenticeships	<i>E Smith</i>	312
Skills Shortages: Concepts, Measurement and Policy Responses	<i>C Shah and G Burke</i>	320
Training and Learning in the Workplace	<i>R L Jacobs</i>	328
Training in Small Businesses	<i>H Matlay</i>	337

Vocational Education and Training – Research on VET

Language of Vocational Education and Training	<i>J Misko, J MacKenzie, and R-A Polvere</i>	343
Research on Vocational Education and Training	<i>J Robson and B Bailey</i>	352
Vocational Education and Training Information Resources	<i>L-A Harris, R Naidu, and M Saunders</i>	358

Vocational Education and Training – Teaching and Learning

Employability Skills	<i>E Smith</i>	368
Learning Styles in Vocational Education and Training	<i>J Vorhaus</i>	376
Modularization in Vocational Education and Training	<i>H Ertl and G Hayward</i>	383
Personalized Learning and Vocational Education and Training	<i>I Duckett</i>	391
Recognition of Prior Learning and Vocational Education and Training	<i>J Keating</i>	397
The Competency Model	<i>A Gonczi and P Hager</i>	403

Vocational Education and Training – VET in Context

Globalization and Vocational Education and Training	<i>P Loveder</i>	411
Learning to Work for the Future	<i>J Fien, R Maclean, and D N Wilson</i>	416
The Relationship between Technical and Vocational Education and Training and Innovation	<i>K Ruth and L Deitmer</i>	423
The Status of Vocational Education and Training	<i>J Lasonen</i>	429
Vocational Education and Training in Post-Conflict Countries	<i>E Lyby</i>	441

Vocational Education and Training – VET System

Dual System	<i>T Deissinger</i>	448
Industry Involvement in the Vocational Education and Training System	<i>M Hughes and C Hughes</i>	455
Non-Completion in Vocational Education and Training: Issues, Patterns and Possibilities	<i>S Lamb and A Walstab</i>	463
Planning and Policy Development for Technical Vocational Education and Training Systems	<i>R Maclean and M Pavlova</i>	469
Qualifications Frameworks and their Role in the Reform of Education and Training	<i>M Young</i>	476
Quality Assurance in Vocational Education and Training	<i>D Gunning</i>	482

Training Finance	<i>A Ziderman</i>	489
Training Markets	<i>P Ryan</i>	496
Vocational Education and Training Teacher Training	<i>S Nielsen</i>	503
Vocational Education and Training Workforce	<i>D L Wallin</i>	513
Subject Index		519
Contributors		747

EDUCATIONAL MEASUREMENT

Contents

- Ability Testing
- Admissions Testing
- Appropriateness Measurement: Person Fit
- Automated Essay Scoring: Writing Assessment and Instruction
- Classical Test Theory Reliability
- Criterion-Referenced Measurement
- Differential Item Functioning
- Educational Measurement: Overview
- Equating and Scaling
- Fairness
- Generalizability Theory
- Impact of Assessment on Classroom Practice
- Intelligent Systems
- Item Response Theory
- National Assessments
- Norm-Referenced Measurement
- Progress Monitoring
- Rasch Models
- Scoring Rubrics
- Second Language Assessment
- Self Adaptive Testing
- Student Test Results in School Accountability
- Test Development
- Test Translation and Adaptation
- Testing Creativity
- Testing in History
- Validity
- Validity of Achievement Gains on High-Stakes Tests
- Validity of Educational Indicators

Ability Testing

P B Tymms, Durham University, Durham, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bell curve – Also known as the Gaussian distribution and the normal distribution. It refers to a frequency distribution commonly found in social science and specifically in relation to ability tests.

COMBSE – Confidential Measurement Based Self Evaluation. A monitoring project from schools which emphasises the formative use of performance indicators. It is now called ALIS (A-level information system).

Correlation – A numerical indicator of the degree of association between two variables. It can vary from -1 to 1 where 1 indicates a perfect positive association and 0 no association.

Ecological fallacy – An inferential error in which associations found in aggregate statistics are assumed to apply at the individual level.

Factor analysis – A statistical technique designed to simplify a dataset by building on patterns in the data. It involves extracting a small number of factors from information on a large number of items.

Heritability – Variation in observed traits amongst individuals can be due to environmental and genetic factors. Heritability is the proportion of variance attributable to genetic variance in a population.

IQ – Intelligence quotient. The term was first used in its German version by William Stern in 1912.

Item Response Theory – A modern assessment approach which puts testees and items on a single equal interval scale. It can accommodate items of varying difficulty, varying discrimination, and varying susceptibility to guessing.

Rasch – A unique modern approach to assessment which may be seen as a subset of Item Response Theory.

Introduction

Ability has two meanings. One refers to established competence such as the ability to run or do arithmetic and the other involves a prediction that somebody will be able to do something. In common parlance, ability can be taken to be either of these two things, but in the educational/psychological context, ability is taken to mean an indication of the propensity to be able to do something, an underlying trait which shows that someone, given the right opportunities, will demonstrate competence in an area. Some would call this intelligence but, as will be noted later, the word intelligence and its measurement to produce an intelligence quotient (IQ), have acquired negative connotations. Perhaps, as a result of this the word, ability is much more widespread and accepted in educational circles.

Historical Perspective

Much research has gone into the measurement of ability and its different types. It has a long history within education and much can be traced back to Binet's work in the latter part of the nineteenth and the early twentieth centuries in France where he was involved in an effort to try to find those pupils who were educable. Officials wanted a method which would bypass teachers' judgments which, it was felt, might be unduly influenced by the trouble that individuals caused. This led to the development of the Binet–Simon scale which was revised and extended at Stanford to produce the Stanford–Binet intelligence scale. This proved to be enormously popular particularly in the United States and elsewhere as well. It went through several revisions and was widely used by clinical and educational psychologists.

Since then, the interplay between researchers' understanding of the psychological constructs, their meaning,

and etiology have run in parallel with efforts within schools and classrooms to try to use the concept and measures of ability. Some of this research has produced theoretical perspectives such as the distinction proposed by Raymond Cattell between fluid and crystallized intelligence. Fluid ability is the capacity to reason and adapt to new situations whereas crystallized intelligence employs stores of information which have already been acquired and which can be applied or extended. Others have tried to look at component parts of ability and so at the structure of ability. Key workers in this area include Charles Spearman, Philip Vernon, J. P. Guilford, and John Carroll. They all subscribe to a general factor running across all abilities – an idea which originated with Spearman and which has more recently been linked to brain structure. It has been termed 'g' (general ability) and its existence is widely although not universally accepted. Carroll's work is the most extensive in this area. Much of this work has involved a statistical technique known as factor analysis which is not without criticism.

Howard Gardner's idea of multiple intelligences which are proposed to consist of linguistic, musical, logical–mathematical, spatial, bodily kinesthetic, and personal intelligences can be seen as an extension of the more psychometric approaches of Carroll and others but based on hypotheses (ideas which are yet to receive strong empirical backing). An alternative perspective comes from a systems approach taken by Robert Sternberg. It is known as the triarchic theory of successful intelligence. Essentially, the theory brings together the interplay between three systems, the internal working of the mind, the interface between the individual and his or her world (experience) and the external world.

Much of the work noted above comes from differential psychology. Others have concentrated on developmental psychology – the most famous worker in that field being Jean Piaget. A third related area is known as cognitive psychology which is concerned with the flow and processing of information. To a large extent, these three traditions have developed independently but recent serious work by Andreas Demetriou has sought to provide a theoretical basis for the manifestations of abilities that we see and to integrate work from the three areas.

Testing and Measurement

A good deal of research in differential psychology has been devoted to exploring the underlying constructs of ability using assessments (also known as tests) and there has been an interplay between the development of tests and theory as the one informs and influences the other. For the majority of research it has been common practice in measurement to use what is known as classical test theory, in which the results from a sample for testees are used as the norm reference to produce scales with known

means and standard deviations (standard scores). A scale with a mean as 100 and a standard deviation of 15 has been so commonly used in ability testing that it has come to be known as the IQ scale. However, it has always been appreciated that this is not a perfect system and indeed Gene Glass has called the process "... pseudoquantification, a meaningless application of numbers to a question not prepared for quantitative analysis." Various workers including L. L. Thurston, Raymond B. Cattell, and Georg Rasch have tried to put psychometrics onto a more satisfactory basis. This has led to equal interval scales of measurement which require the researcher to focus on a single construct. The testee and the items making up the test are both located on the same unidimensional scale. The simplest approach which has a wide following is known as Rasch measurement but it should be noted that some see Rasch as a subset of Item Response Theory. Purists see Rasch measurement as being a major advance for the social sciences moving it forward by engaging in fundamental measurement. Only with such measurement, it is argued, can social science be put on to a serious scientific base.

These theoretical debates and associated advances in psychometrics have provided the necessary basis for the creation of computer-adaptive assessments rather than traditional pencil-and-paper assessments. Such tests employ algorithms which ensure that the test-taker is efficiently assessed. Responses to items are used to select new items and, as a result less time is spent obtaining the same information than in a test where everyone is required to answer all the questions. A downside of computer-adaptive testing is that it takes enormous time and energy to develop.

Controversy

It is to be regretted that some of the research surrounding ability has resulted in the whole field becoming tainted with accusations of racism. This has largely followed the work on differences between racial groups on measures of IQ by high-profile pieces of research by Arthur Jensen and Hans Eysenck. While reports of the research have often failed to recognize the nuanced views of the authors, an impression has been created. Various studies of heritability have indicated that inheritance has a part to play in explaining measures of IQ. The studies involve looking at IQ scores within and between families, at the scores of siblings and twins (fraternal and identical), at twins reared apart and together, and at adoption studies. There is argument over the precise figures and it is accepted that the figure must be different for different settings or cultures, but estimates vary from 30% to 70%. This is a complex area and not one that should be approached lightly or with little knowledge. Nevertheless, workers such as Jensen and Eysenck and more recently Herrnstein

and Murray have argued that the heritability figures can be used to explain the difference between different racial groups. These studies have created a furore and much condemnation from others within the scientific community. Of particular relevance to the debate have been empirical findings such as the changes in IQ over the years. The IQs of African Americans in the United States now exceed those of Whites when the work identified the difference in the first place. There has also been a decrease in the gap between the two groups.

A general rise in IQ scores, the so-called Flynn effect, has been observed across the Western world. It has been particularly seen in those tests which are apparently the purest measures of *g*, the most famous of which is Ravens progressive matrices. This has led to advances in theoretical explanations of how ability develops through its interaction with the environment. A particularly important theoretical perspective by William Dickens has sought to explain how differences can appear between races because of the way in which society is organized, and yet can still have substantial heritability. It also shows how *g* can appear as a result of personal selection, societal structure, and interaction with the environment.

Developed Ability

The baggage associated with ability testing and IQ scores indicates that there is a problem with the use of ability tests. It produces, in the minds of some, thoughts of innate differences and inherited IQ that cannot be changed and which are linked to race. In reality, ability is the propensity, at a certain point in time, to acquire certain skills, resulting from an individual's decisions, interaction with the environment, opportunities that have presented themselves, and genetic make-up. This developed ability is something which can be altered, albeit slowly, which schooling improves but only over a considerable period of time, and which can be suppressed by a culture and ways in which society operates.

Ability Testing by Psychologists and Teachers

It should be taken as read that the aims of ability testing within education must ultimately be to improve the education of the individuals who are being assessed and that high standards should be maintained. It is important to distinguish between different types of testing. It is rare to find that teachers develop their own measures of ability. They might make judgments of individuals but they rarely create ability tests in the way that they might create tests of geography or mathematics. These implicit judgments are important and probably have long-term consequences

for individuals but the actual tests themselves are left to specialists. Further, there are two types of tests known as open and closed tests. Closed tests are only available to psychologists and open tests, as the phrase implies can be bought and used more readily in schools.

The most famous of closed tests is the Weschler Intelligence Scale for Children (WISC), which has gone through a series of revisions and is now available in its fourth edition strongly influenced by Carroll's work. It is associated with systematic use by educational psychologists either as part of their duties within the schools or in clinical practice. The second stream of ability testing is in schools and districts by teachers and administrators. Sometimes ability testing has been used for the selection of pupils. In England, following the 1944 Education Act, and the creation of a national selective system of secondary schools, those who passed the ability test, the so-called 11+, went to grammar schools and those who did less well went to technical schools, and those lesser than that into secondary modern schools. The psychologist Cyril Burt was influential in devising the policy. Although many state secondary schools in England are now not selective, grammar schools in England continue to exist and continue to select on the basis of ability tests as do some private schools.

The use of IQ tests was widespread in schools in the United States but concern over the tests' validity with subgroups has led to their restriction. This restriction in use has been encouraged by legal action. For example, in 1984 the 9th Circuit Court of appeals ruled that IQ tests should not be used to place students into classes for the mentally retarded. Arguments continue in the courts and outside.

Dyslexia

It may seem natural to assume that students with high ability who find difficulty reading fluently and accurately should be treated differently from students with low ability who have similar problems. Such thinking led to a definition of dyslexia which involved a discrepancy between IQ and reading scores. It also led to policies in which different approaches were used for poor readers with high and low IQ scores. But a series of investigations has found that irrespective of IQ, poor readers are best helped by diagnosing their specific problems and tackling those directly. It is also generally accepted that dyslexia should be defined without reference to ability.

Monitoring

A quite different use of ability tests in schools is linked to monitoring. In England in 1982, the governor of a prestigious school came to the local university and sought out

an assessment expert, Carol Taylor Fitz-Gibbon, and asked her to look at some grades that had been given for preuniversity math tests, which are known in England as Advanced Level exams. The school were expecting Grade As, but they were largely Ds and Es or even unclassified. Fitz-Gibbon looked at the results, and perhaps to the disappointment of the governors, said that on the basis of those grades she could say nothing because she had not seen any indication of the ability of the math students, and she wondered what those students who had chosen mathematics as one of three subjects, thought about mathematics now. Did they enjoy their time doing it? How were they taught? A whole range of questions flooded out, none of which had ready answers. As a result, she set up a project with a group of schools and colleges in the North East of England, which would share the data anonymously, together with information about attitudes and about processes. A key to the project was the development of statistical controls so that like could be compared with like as one school looked at its data in comparison with others. This project was known as Confidential Measurement Based Self Evaluation (COMBSE) project. Since those ideas and since that original development, that project has grown every year until now in 2007, it covers well over half of all A-level provision.

One of the issues that Fitz-Gibbon had to face was what the controls should be. Of course, there were ready variables that could be put into the regression equations such as sex, eligibility to free school meals, and there were even prior achievement measures taken by the students 2 years before their A-level exams. These were leaving exams taken in England. They correlated at the 0.5–0.7 level and acted as good statistical controls for looking at the A-level results, but there was a problem with them – they were dependent on the teaching of the schools which then went on to teach the A levels. So, were a school to be seen as particularly good at teaching for the leaving certificate, they would then be potentially penalized when it came to look at their A-level grades because the controls would show that their progress had been less than expected. She therefore looked for an ability measure to act as a control, the idea being that this would be a cognitive measure which the school had not specifically taught. From Education Testing Services, she obtained the International Test of Developed Abilities (ITDA). It had been created for colleges in the United States in order to select students who had been in different countries and had varied backgrounds. Using that, a school would be able to look at its A-level grades and the gains made against the ITDA, also against the gains made against the national leaving certificate test, or against the gains made simply knowing the home background of the children. This allowed schools to look at the data in very different ways. It did not give an answer which was right and for which people could be held accountable, but rather it provided information for teachers to understand their

own situation and to integrate professional knowledge with quantitative data. Since the success of the COMBSE Project, a whole series of projects for different ages and different countries have been created in its image.

In using these data in schools, it became clear that the communication of the idea of cognitive measures, ability measures, which the schools had not deliberately enhanced was understood and largely well accepted. It was emphasized that ability is not something which is fixed, nor is it something which is entirely dependent on genes, but it is something which develops over time.

Some have tried to monitor school performance by using controls for home background alone. This is not adequate. Measures of the home typically correlate with schools success at about the 0.3 level in the United Kingdom explaining about 10% of the outcome measure. On the other hand, cognitive measures, such as an ability test, correlate at about the 0.7 level explaining about 50% of the variance. The correlations with home background are much increased if the unit of analysis is taken to be the school rather than the students and can get as high as 0.8. But, any suggestion that such correlations can be extrapolated to indicate that home background is a good predictor of school success is to fall foul of the ecological fallacy.

School Effectiveness

A project like COMBSE generates much data which can be examined to develop hypothesis about effective schooling and much research has been carried out into this area. Some of it, particularly in the Netherlands and the United Kingdom, has involved ability testing and an effort to look for associations between gain scores and other school-based factors.

Consequences

What have been the consequences of ability testing in education? Has it helped students? Over the last 100 years, it is clear that mistakes have been made. Many students have been wrongly classified by intelligence tests because of inaccurate measurement but, more worryingly, because of bias in the assessments and sometimes because of unjustified theoretical stances. On the other hand, society and schooling is such that judgments are inevitable. Some schools offer high-level math classes for which pupils must be chosen; some pupils have difficulties and need help with limited resources. Whenever human judgment is required, mistakes abound. We are notoriously biased and persistent in our mistaken homespun theories. If a badly behaved child from a poor background is to be selected then he (*sic*) may have more chance passing an

ability test than being chosen by his teacher, taking us back to the issue presented to Binet at the turn of the nineteenth century. On balance, have more selection mistakes been created or avoided as a result of ability testing? It is impossible to say. What cannot be doubted is that the potential for better decisions to be taken exists if knowledgeable educators have high-quality ability data available to them.

Have the professional monitoring systems with the use of ability tests enhanced education in the way that was originally hoped and envisaged? Is better information through feedback loops actually having an impact? There have been experiments in the area and the evidence has been synthesized by Robert Coe. On balance, feedback has a small positive effect although the evidence is not specifically linked to ability testing. However, the professional monitoring movement has proved to be important in enabling schools to defend themselves against unwarranted attacks on the basis of judgments made on the basis of inadequate controls or inspectors' judgments. Giving teachers the capacity to defend themselves has been very important. It has also provided very good data to look at the relative difficulties of subjects and thus to comment on the comparability of examinations.

Conclusion

Ability testing has been in use for a little over 100 years and it has been used so widely and investigated so thoroughly that more is probably known about the subject than any other area in education. Through that time, ability tests have been used to spot talent and to help to remove bias from judgment using objective measures. Other uses and perspectives have changed dramatically. The simplistic links made to race are now no longer tenable and concerns about bias within the tests themselves are much more clearly understood and widespread. Indeed, the knowledge base generated through research into abilities can now be used to counter racist perceptions. Further, our understanding of how intelligence develops is much stronger than it ever was as is our knowledge about the architecture of the mind and how we think.

There can be little doubt that ability testing has been used in harmful ways – condemning individuals to inappropriate situations. But there can also be little doubt that if used appropriately, ability testing can be used in a positive way to help education.

See *also*: Educational Measurement: Overview; Item Response Theory; Norm-Referenced Measurement; Piaget: Recent Work; Rasch Models; School Effectiveness in Developed Societies; Self Adaptive Testing; Testing Creativity; Testing in History; Validity of Educational Indicators.

Further Reading

- American Psychological Association, American Educational Research Association and National Council on Measurement in Education (1999). *The Standards for Educational and Psychological Testing*. American Educational Research Association.
- Bouchard, T. J., Lykken, D. T., McGue, M., Segal, N. L., and Tellegen, A. (1990). Sources of human psychological differences: The Minnesota study of twins reared apart. *Science* **250**(4978), 223–228.
- Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor Analytic Studies*. Cambridge: Cambridge University Press.
- Carroll, J. B. (1994). Cognitive abilities: Constructing a theory from data. In Derman, D. K. (ed.) *Current Topics in Human Intelligence: Vol. 4 Theories of Intelligence*, pp 43–63. Norwood, NJ: Ablex.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology* **54**, 1–22.
- Cronbach, L. J. (1990). *Essentials of Psychological Testing*, 5th edn. Longman.
- Daniels, M., Devlin, M., and Roeder, K. (1997). Of genes and IQ. In Devlin, B., Fienberg, S. E., Resnick, D. P., and Roeder, K. (eds.) *Intelligence, Genes and Success: Scientists Respond to the Bell Curve*, pp 46–70. New York: Springer.
- Demetriou, A. (2005). Mind intelligence and development: A cognitive, differential, and developmental theory of intelligence. In Demetriou, A. and Raftopoulos, A. (eds.) *Cognitive Developmental Change: Models, Methods, and Measurement*, pp 21–73. Cambridge: Cambridge University Press.
- Devlin, B., Fienberg, S. E., Resnick, D. P., and Roeder, K. (eds.) *Intelligence, Genes and Success: Scientists Respond to the Bell Curve*. New York: Springer.
- Dickens, W. T. (2007). What is g? *Paper Presented at the American Education Research Association*, Chicago.
- Dickens, W. T. and Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review* **108**(2), 346–369.
- Dickens, W. T. and Flynn, J. R. (2006). Black Americans reduce the racial IQ gap: Evidence from standardization samples. *Psychological Science* **17**(10), 913–920.
- Fitz-Gibbon, C. T. (1996). *The Design and Use of Monitoring Education: Indicators, Quality and Effectiveness*. London: Cassell.
- Flanagan, D. P. and Harrison, P. L. (2005). *Contemporary Intellectual Assessment: Theories Tests and Issues*. New York: Guilford Press.
- Flynn, J. R. (1987). Massive gains in 14 nations: What IQ tests really measure. *Psychological Bulletin* **101**, 171–191.
- Gardner, H. (1993). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.
- Glymour, C. (1997). Social statistics and genuine inquiry: Reflections on the bell curve. In Devlin, B., Fienberg, S. E., Resnick, D. P., and Roeder, K. (eds.) *Intelligence, Genes and Success: Scientists Respond to the Bell Curve*, pp 257–280. New York: Copernicus.
- Gould, S. J. (1996). *The Mismeasure of Man*. London: Penguin Books.
- Guiford, J. P. (1967). *The Nature of Human Intelligence*. New York: McGraw-Hill.
- Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., and Alkire, M. T. (2004). Structural brain variation and general intelligence. *Neuroimage* **4**(25), 1–9.
- Herrnstein, R. J. and Murray, C. A. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Simon & Schuster.
- Paiget, J. (1970). Paiget's theory. In Mussen, P. H. (ed.) *Carmichael's Handbook of Child Development*, pp 703–732. New York: Wiley.
- Sternberg, R. J. (1988). *The Triarchic Mind: A New Theory of Human Intelligence*. New York: Viking Press.
- Stuebing, K. K., Fletcher, J. M., LeDoux, J. M., et al. (2002). Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal* **39**(2), 469–518.
- Tymms, P. (1999). *Baseline Assessment and Monitoring in Primary Schools: Achievements, Attitudes and Value-Added Indicators*. London: David Fulton.
- Weschler, D. (2003). *Weschler Intelligence Scale for Children*, 4th edn. San Antonio, TX: Psychological Corporation.
- Winship, C. and Korenman, S. (1997). Does staying in school make you smarter? In Devlin, B., Fienberg, S. E., Resnick, D. P., and Roeder, K. (eds.) *Intelligence, Genes and Success: Scientists Respond to the Bell Curve*, pp 215–234. New York: Copernicus.

Admissions Testing*

R Zwick, University of California, Santa Barbara, Santa Barbara, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Correlation coefficient (multiple) – A measure of association, ranging from 0 to 1, between a set of (at least two) variables and a single variable. For example, a multiple correlation coefficient could be used to assess the degree to which test scores and high school grades, considered jointly, are associated with college grade-point average.

Correlation coefficient (simple) – A measure of linear association, ranging from – 1 to 1, between two variables (e.g., an admissions test score and a college grade-point average).

Differential item functioning (DIF) –

A phenomenon in which two (or more) groups of test takers (e.g., men and women) tend to perform differently on a test question despite having equal skill in the area that is being tested. One possible cause of DIF is the inclusion in the test question of material that is irrelevant to the proficiency that is intended to be assessed.

Linear regression analysis – A statistical technique used to assess the relationship between a set of predictors (consisting of one or more variables) and a criterion variable. In the admissions context, for example, test scores and high school grades are often used to predict college grade-point average.

Predictive validity – The degree to which a test or other measuring instrument can predict a criterion of interest, typically academic or job performance. In the case of admissions tests, the criterion is typically a grade-point average in a subsequent academic program.

Range restriction – A reduction in the range (distance between maximum and minimum values) of a variable that results from a selection process. For example, the SAT scores of students admitted to college typically have a smaller range than the SAT scores of applicants.

Background

This article provides a brief historical perspective and then describes several aspects of higher education admissions testing in the USA: the main tests used, the role of tests in the admissions process, and the evidence on predictive validity. Following this is a short discussion of test fairness issues. Admissions testing in several other countries is then described.

Accounts of testing history agree that the administration of tests for regulating admissions to universities began in Europe, most likely in the eighteenth century (Webber, 1989: 37), although some reports give an earlier date (e.g., Office of Technology Assessment, 1992; Stewart, 1998). Most descriptions agree that admissions testing had been instituted in Germany and England by the mid-1800s.

Standardized admissions testing first took root in the USA during the early part of the twentieth century, when the leaders of 12 northeastern universities formed the College Entrance Examination Board in 1900. The College Board created a common set of examinations that were administered by member institutions and then shipped back to the Board for scoring. Initially, the Board developed essay tests in several subject areas; it later developed a new exam that contained mostly multiple-choice questions – the SAT. This precursor to today's SAT was first administered in 1926 to about 8000 candidates.

The first SAT consisted of questions similar to those included in the Army Alpha tests, which had been developed for use in selecting and assigning military recruits in World War I. These Army tests, in turn, were directly descended from intelligence quotient (IQ) tests, which had made their first USA appearance in the early 1900s. The SAT gained popularity over the next two decades and, in 1947, Educational Testing Service (ETS), currently the largest USA testing organization, was founded. ETS has had a role in most of the standardized admissions tests used in the USA today.

Higher Education Admissions Testing in the USA

Main Admissions Tests Used in the USA Today

Admissions processes in the USA are largely under the control of individual institutions and instructional programs. These entities decide whether and how to use tests in making admissions decisions. The main admissions tests

* Portions of this article are adapted from *Fair Game? The Use of Standardized Admissions Tests in Higher Education*, RoutledgeFalmer, 2002. Copyright © 2002 by Rebecca Zwick.

used in the USA today are the SAT and ACT, used for college admission; the Graduate Record Examinations (GRE), used for graduate school admission; the Medical College Admission Test (MCAT); the Law School Admission Test (LSAT); and the Graduate Management Admission Test (GMAT), used for admission to business school. As described below, these tests are developed and produced by private organizations. (Tests that are not discussed here include the Test of English as a Foreign Language, which is required of foreign students by some USA colleges or graduate schools and the Miller Analogies Test, which is accepted by a small percent of graduate programs as an alternative to the GRE.)

Today's SAT

The SAT testing program is sponsored by the College Board; the tests are administered by ETS under a contract with the Board. According to the College Board website, the SAT Reasoning Test “measures critical reading, mathematical reasoning, and writing skills that students have developed over time and that they need to be successful in college.” (“SAT” is no longer considered to be an acronym, but the actual name of the test. Originally, it stood for Scholastic Aptitude Test, which was later changed to Scholastic Assessment Test. More recently, the test was known as the SAT I: Reasoning Test.)

The SAT has undergone many changes in content in its eight decades of existence; a history of these changes is provided by Lawrence *et al.* (2004). The most recent revisions occurred in 2005, following a nationwide controversy about the utility and fairness of the SAT. Historically, the SAT had provided math and verbal scores, but the new version of the test consists of three separately scored sections: math, critical reading, and writing. Analogy items were eliminated, and more advanced math content was added. All the critical reading questions and most of the math questions are multiple choice. Each SAT also includes some math questions that require student-produced answers – there are no response choices. The newly added writing section, which is mandatory, includes separately scored multiple-choice and essay sections. An overall writing score is also provided.

In addition to the SAT Reasoning Test, the current SAT program also includes 20 SAT Subject Tests (formerly called the College Board Achievement Tests and later, the SAT II: Subject Tests), which assess the candidates' knowledge in particular areas.

The ACT

In 1959, the SAT program acquired a competitor in the college admissions test market – the American College Testing Program. (Today, the company is ACT, Inc. and the test is simply the ACT. Like ‘SAT’, ‘ACT’ is no longer considered an acronym.) In its early years, the ACT was administered primarily in Midwestern states, but it is now considered interchangeable with the SAT by most institutions.

From the beginning, the ACT was somewhat different from the SAT in terms of underlying philosophy: While the SAT consisted only of verbal and mathematical sections, the ACT was more closely tied to instructional objectives. Today, the content of the test is informed by regular surveys of secondary school teachers and curriculum experts. The ACT consists of four multiple-choice sections: English, mathematics, reading, and science. Students receive a score in each subject area, as well as a composite score. Seven subscores are also reported – two in English, three in mathematics, and two in reading.

In 2002, after the College Board announced that a writing component would be added to the SAT, ACT, Inc. announced that it would add a 30-min essay test to the ACT. Unlike the SAT writing section, however, the ACT writing test, first administered in 2005, is optional. Students who elect to take it along with the ACT receive two additional scores, a writing test score, and a combined English/writing score, as well as comments about their essays.

The GRE

The GRE program of the Carnegie Foundation for the Advancement of Teaching administered its first exam in 1937. In 1948, the GRE program was transferred to the newly formed ETS. Today, the GRE, which is used to evaluate candidates for admission to graduate school, is developed and administered by ETS under the direction of the GRE Board, an independent 34-member organization that is affiliated with the Association of Graduate Schools and the Council of Graduate Schools. The GRE General Test was the first major admissions test to be administered as a computerized adaptive test, beginning in 1993. In a test of this kind, candidates are administered items that are intended to be tailored to their skill levels. The exam is available in test centers around the world.

The GRE has undergone several content changes since its inception. Among the more recent modifications was the incorporation of a writing assessment into the General Test in 2002, replacing the analytical reasoning section. According to the ETS website, today's GRE General Test “measures verbal reasoning, quantitative reasoning, and critical thinking and analytical writing skills that have been acquired over a long period of time and that are not related to any specific field of study.” The verbal and quantitative reasoning items are primarily multiple-choice, although some quantitative items requiring student-produced answers were added in 2007. The analytical writing section consists of two essay tasks, one asking the candidate to take a perspective on an issue, and one requiring the candidate to analyze an argument. Test-takers receive separate scores in quantitative reasoning, verbal reasoning, and writing. In addition to the General Test, eight GRE Subject Tests are available.

The MCAT

The Medical College Admission Test, first given in 1946, is sponsored by the Association of American Medical Colleges, which currently represents about 125 USA medical schools, 17 Canadian medical schools, 400 teaching hospitals, and various academic societies and medical professionals. Since 2007, the MCAT has been administered as a (nonadaptive) computer-based test by Thomson Prometric, a technology-based testing company.

As described in an MCAT bulletin, the test assesses “the skills and knowledge that medical educators and physicians have identified as key prerequisites for success in medical school and the practice of medicine,” including “mastery of basic concepts in biology, general and organic chemistry, and physics, . . . capacity for problem solving and critical thinking as well as general writing skills” (Association of American Medical Colleges, 2007). The test consists of three multiple-choice sections – verbal reasoning, physical sciences, and biological sciences – as well as a writing sample composed of two essay questions. A separate score is reported for each section of the MCAT.

The LSAT

The LSAT was conceived at a 1947 meeting between College Board staff and representatives of an association of nine law schools, the precursor of the Law School Admission Council (LSAC). In 1948, the first LSAT was administered by ETS. Since 1979, the LSAT has been developed by the LSAC itself, which today has more than 200 member institutions in the USA and Canada.

The LSAT is “designed to measure skills that are considered essential for success in law school: the reading and comprehension of complex texts with accuracy and insight; the organization and management of information and the ability to draw reasonable inferences from it; the ability to reason critically; and the analysis and evaluation of the reasoning and argument of others” (Law School Admission Council, 1999). Four sections serve as the basis for a single score: one reading comprehension, one analytical reasoning, and two logical reasoning sections. All LSAT questions are multiple-choice except a writing sample, added in 1982, which is not graded, but is sent to the law schools to which the candidate applies.

The GMAT

In the late 1940s and early 1950s, the GRE was used to screen business school applicants. In 1953, however, representatives of nine graduate schools of business agreed that they needed an admissions test of their own and, just a year later, the precursor to the GMAT was administered (Schmottter, 1993). For five decades, the test was developed and administered by ETS for the Graduate Management Admission Council (GMAC), which includes about 160 governing schools of business and management.

Since 2006, the GMAC has contracted with ACT, Inc. to develop the GMAT and with Pearson VUE, a company specializing in computer-based testing, to administer it.

Today’s GMAT, according to its information bulletin, “measures general verbal, mathematical, and analytical writing skills that are developed over a long period of time . . . The GMAT does not presuppose any specific knowledge of business . . . [and] does not measure achievement in any particular subject areas” (Graduate Management Admission Council, 2000: 5). Since 1997, the GMAT has been a computerized adaptive test, like the GRE.

The test contains verbal and quantitative sections, both of which are multiple choice, and an analytical writing section, which was added in 1994. The writing section requires the test-taker to produce essay responses to two questions, with half an hour allowed for each. Separate scores are reported for each of the three sections of the GMAT; a total score that reflects performance on the verbal and quantitative sections (but is not the total of the verbal and quantitative scores) is also reported.

The Use of Higher Education Admissions Tests in the USA

Use of admissions tests at undergraduate institutions

In the USA, use of admissions tests varies widely across institutions. At open-door colleges, tests play no role in the admissions process: all that is required is to complete an application and, in some cases, show proof of high school graduation. Eight percent of the 957 4-year institutions that responded to a survey conducted in 2000 by ACT, Inc., the Association for Institutional Research, the College Board, Educational Testing Service, and the National Association for College Admission Counseling (referred to hereafter as ‘the joint survey’), came under the open-door category; 80% of the 663 2-year institutions were open-door (see Breland *et al.*, 2002: 15).

According to the joint survey, the percentage of 4-year colleges requiring either the SAT or ACT held steady at slightly over 90% between 1979 and 2000 (Breland *et al.*, 2002). Two major sources of information, the joint survey and the National Association for College Admission Counseling (NACAC) Admission Trends Survey (Hawkins and Lautz, 2005) indicate that test scores are viewed as the second-most important factor, after high school grades, and that the importance attributed to test scores has increased in recent years (Breland *et al.*, 2002: 67; Hawkins and Lautz, 2005: 39). (The watchdog organization FairTest lists 740 schools that “deemphasize the use of standardized tests by making admissions decisions about substantial numbers of applicants who recently graduated from U.S. high schools without using the SAT or ACT.”)

Use of admissions tests at graduate and professional schools

Graduate and professional school programs tend to be smaller, and admissions policies less formalized and less public than at undergraduate institutions. Decisions are typically in the hands of faculty, rather than admissions officers and admission rates tend to be much lower.

Admissions policies and rates vary widely over the hundreds of fields of doctoral study available in the USA. Decisions are typically made at the department level, using very flexible procedures. As a College Board report noted, “published statements about [doctoral] admissions provide ample latitude for almost any decision” (Rigol and Kimmel, 1997: 13). Surveys have shown that standardized admissions tests are widely used in graduate admissions, in combination with undergraduate grades and other factors (Kuncel *et al.*, 2001; Rigol and Kimmel, 1997). The vast majority of USA graduate programs require the GRE General Test (Norcross *et al.*, 1996); some require certain GRE Subject Tests as well. The Miller Analogies Test, developed by The Psychological Corporation, is accepted by a small percentage of graduate programs as an alternative to the GRE.

Use of admissions tests is nearly universal at professional schools. The MCAT is required by nearly all American medical schools. All American and Canadian law schools that belong to the LSAC require the LSAT, and most other law schools do as well. According to its website, the GMAT is used by “thousands of graduate management programs around the world.” Undergraduate grades are also considered very important in professional school admissions, and medical schools place a high value on interviews as well (Johnson and Edwards, 1991).

Predictive Validity of Admissions Tests

Although they differ somewhat in the skills they are alleged to measure, all college, graduate school, and professional school admissions tests share one particular claim – that they are useful for predicting the grades students will receive in the educational programs they enter. Although a comprehensive evaluation of a test’s validity must involve a consideration of the test’s design, development, content, administration, and use, the validity of admissions tests as a selection tool for higher education institutions is, in practice, judged largely by the degree to which test scores can predict students’ grade-point averages in college, graduate school, or professional school (GPAs). Most often, first-year grades are used.

Conducting a predictive validity study requires that the GPAs for the cohort of interest be available so that the predicted GPAs can be compared to the GPAs actually earned by the admitted students. Predictive validity studies are usually conducted within a single institution, although results may later be averaged across institutions.

Linear regression analysis is typically applied to estimate an equation for predicting college, graduate school, or professional school GPA. The predictors are typically admissions test scores and high school grades (in the case of college admissions) or undergraduate grades (in the case of graduate or professional school admissions). The resulting multiple correlation, which can range from 0 to 1, provides an index of the effectiveness of the prediction equation. The regression analysis can then be repeated using prior grades alone as a predictor. Comparing the predictive effectiveness of the two equations gives an estimate of the value added by using admissions test scores. The simple correlations between test scores and subsequent GPAs are often examined as well.

A factor that complicates the interpretation of these correlations, or validity coefficients, is restriction of range: students whose test scores are too low to allow admission to a higher education institution will not have GPAs. Because of this restriction of range (of test scores, and, as a result, of other predictors and of GPAs as well), validity coefficients tend to be smaller for the admitted students than they would be for the entire population of applicants. Statistical corrections are often applied in an attempt to estimate how big the association would have been if the range had not been restricted (see Gulliksen, 1987: 165–166). Except where noted, the correlations reported in this article have not been corrected for range restriction. Typically, the corrected correlations are larger by 0.15 – 0.20.

An examination of large-scale studies (focusing on multi-institution studies and reviews published in 2000 or later) reveals some consistent patterns in the findings on the predictive validity of admissions tests in college, graduate school, and professional school. (See Kuncel and Hezlett (2007) for an excellent overview of graduate and professional school results.) The multiple correlation of the ACT or the SAT with college GPA is typically about 0.3 – 0.4 (ACT, Inc., 2007; Burton and Ramist, 2001; Camara and Echternacht, 2000; Kobrin *et al.*, 2008). In the case of the ACT, these results are based on analyses that use scores on the English, mathematics, reading, and science sections, considered together. In the case of the SAT, results are based on analyses that use math and verbal (or, after 2005, critical reading) scores considered together. The correlation between the admissions test (ACT or SAT) and college GPA is usually found to be slightly lower than the correlation between high school grades and GPA. Considering ACT or SAT scores as predictors along with high school grades yields correlations with GPA that average about 0.5.

A study of the SAT writing test, first administered in 2005, showed that scores were correlated about 0.5 with college GPA (Kobrin *et al.*, 2008). These results supported the predictive validity of the writing test, but also showed that it added only a small amount (0.02) to the validity

coefficient obtained using critical reading and mathematics scores alone. These results paralleled those obtained from an earlier study of a prototype version of the SAT writing test (Norris *et al.*, 2005; see also Mattern *et al.*, 2007). Similarly, research on the ACT writing test (ACT, Inc., 2009) showed that scores were predictive of grades in 'writing-intensive' college courses, but that inclusion of the writing test in the prediction equation increased the validity coefficient by only 0.02 over the value that is obtained using the ACT English test alone.

Most GRE validity research predates the substitution of the writing assessment for the analytical reasoning component in the main part of the test. Results from the GRE Validity Study Service collected between 1986 and 1990, which are based on more than 1000 departments and more than 12 000 test-takers (Educational Testing Service, 2003), as well as on other recent multi-institution studies (Burton and Wang, 2005; Kuncel *et al.*, 2001), show that the predictive validity of the GRE (as formerly constituted) is quite similar to that of the SAT and ACT. GRE scores (verbal, quantitative, and analytical reasoning considered together) typically have a validity coefficient of 0.3 – 0.4, and this is usually similar to the correlation between undergraduate GPA and first-year graduate school GPA. When undergraduate GPA and GRE scores are considered in combination, their correlation with graduate school GPA is 0.4 – 0.5. Including scores on the GRE Subject Tests as predictors usually boosts the correlation to 0.5 – 0.6.

Separate validity analyses of the GRE writing assessment (before it was added to the main assessment) were conducted based on approximately 2000 college juniors, college seniors, and first-year graduate students. Analyses showed that the writing assessment (the combined score on two essays) had correlations of about 0.3 with a GPA based on courses that 'required considerable writing'. Correlations with overall GPA were about 0.2, and correlations with GPA in the students' major field were smaller (Powers *et al.*, 1999: 33).

The association between the GMAT (verbal, quantitative, and analytical writing scores considered together) and business school GPA (evaluated halfway through the business school program) was investigated by Talento-Miller and Rudner (2005), based on results from 273 schools. They found a median multiple correlation of 0.5, substantially larger than the median correlation between undergraduate grades and business school GPA (0.3). (These correlations were adjusted for restriction of range.) According to a recent summary of results from 166 schools, the LSAT typically does a slightly better job of predicting first-year law school GPA than do undergraduate grades, yielding correlations averaging over 0.3 (Stilwell *et al.*, 2007). A study based on two student cohorts at each of 14 medical schools reported that the median correlation between MCAT scores and the average of

first- and second-year medical school GPA exceeded 0.4, higher than the median correlation between undergraduate grades and medical school GPA (Julian, 2005). In the case of the GMAT, LSAT, and MCAT, using test scores and undergraduate grades in combination to predict subsequent GPA was slightly more effective than using test scores alone.

In summary, college admissions test scores tend to be slightly weaker than high school grades as predictors of college GPA. Including test scores in addition to past grades in the prediction equation tends to increase the validity coefficient by about 0.1. In predicting graduate school grades, GRE scores and undergraduate grades tend to be equally effective. Test scores are frequently found to be more effective than undergraduate grades in predicting professional school grades.

Although first-year GPA is the most common criterion variable in admission test validity studies, other criteria have been studied as well. For example, there is considerable evidence that admissions test scores are helpful in predicting grades beyond the first year (Burton and Ramist, 2001; Kuncel and Hezlett, 2007; Zwick, 2006, 2007). The evidence on prediction of degree attainment is mixed (Zwick, 2006, 2007). The SAT appears to be useful for this purpose (Burton and Ramist, 2001; Carnevale and Rose, 2003), while graduate and professional school admission tests are less so (Kuncel and Hezlett, 2007). Some studies have examined prediction of other criteria, such as performance on licensing exams and measures of career success, with inconsistent results (see Zwick (2006) for a summary).

Use of personal characteristics to predict success in higher education

An area of ongoing debate is whether including formal measures of personal characteristics could improve the prediction of success in higher education. In a landmark 1985 study, Warren Willingham identified a characteristic he called 'productive follow-through,' defined as 'persistent and successful extracurricular accomplishment' in high school. Productive follow-through proved to be a poor substitute for high school grades and admissions test scores in predicting which students would be identified as most successful by their colleges, but taking this quality into account along with traditional academic qualifications substantially improved the prediction of success (Willingham, 1985: 184).

A measure of personal characteristics that has often been used for admissions purposes in recent years is the Noncognitive Questionnaire (NCQ) developed by William E. Sedlacek. The NCQ is intended to assess such factors as positive self-concept and preference for long-term goals (Sedlacek, 1998: 11). The NCQ is claimed to be useful for predicting college GPA and college persistence (e.g., Tracey and Sedlacek, 1984). However, a recent

meta-analysis based on over 9000 cases from 47 samples found NCQ scores to be “largely unrelated to college performance as measured by GPA, college persistence, and credits earned” (Thomas *et al.*, 2007: 648).

Although there is widespread agreement that factors such as perseverance, determination, and ability to delay gratification are relevant to success in higher education, measuring these characteristics accurately and without bias is difficult. One problem in using self-reports about these attributes in the admissions context is that the desired answers are often obvious to the respondent. Thus, the use of information from respondents other than the candidates themselves (possibly in the context of a standardized letter of recommendation) may yield more useful data.

Fairness of Admissions Tests to People of Color, Women, and Other Special Populations

Standardized admissions test results often reveal substantial average score differences across ethnic, gender, and socioeconomic groups. Among the general public, these differences are often regarded as sufficient evidence that these tests are biased. From a psychometric perspective, a test’s fairness is inextricably tied to its validity. According to Cole and Moss (1989), test bias occurs (i.e., fairness is violated) “when a test score has meanings or implications for a relevant, definable subgroup of test takers that are different from the meanings or implications” for other test-takers. “. . . [B]ias is differential validity of a given interpretation of a test score . . .” (p. 205).

Psychometric assessment of the fairness of admissions tests typically comprises two broad types of investigations. One type consists of analyses of differential prediction of a criterion, usually a GPA. Two distinct questions are typically addressed: first, are the test scores equally predictive of later grades for all groups? This question is typically investigated by obtaining separate prediction equations for each group and then comparing correlation or regression coefficients across groups. Second, if we obtain a single prediction equation for students as a whole, does this equation produce predicted GPAs that are systematically too high or too low for some groups? (See Zwick, 2002: 117–124, 147–150.)

Another common component of fairness assessment is an examination of item content. Before an item is approved for inclusion in an admissions test, it undergoes a ‘sensitivity review’ to make sure its content is not disturbing to certain student groups or offensive in some other way. Later, after the test is administered, a differential item functioning (DIF) screening is performed to determine whether equally skilled members of different groups (e.g., men and women) have statistically different rates of correct response on some items. The purpose of DIF analysis is to identify test items with content that may

be problematic for some student group. For example, the item may include content that is irrelevant to the construct being assessed, and is more familiar to some student groups than others, such as sports content in a mathematics test. Items found to have DIF are either discarded immediately or flagged for further study.

Zwick (2002, 2006, 2007) summarizes findings on differential prediction of GPAs for key student groups, including people of color, women, people with disabilities, and students with limited English skills (see also Young, 2004), and provides a discussion of DIF findings and fairness issues associated with test coaching.

Higher Education Admissions Testing Around the World

Higher education admissions processes vary internationally in terms of the degree of government oversight, the degree of centralization, and the role, if any, of admissions tests. Some examples follow.

In Europe, the minimum requirement for admission to higher education institutions is a certificate of completion of secondary school or its equivalent. (The certificate itself typically requires passage of exams based on the secondary school curriculum.) Most countries also have additional requirements for university entrance, such as admissions tests or interviews. In the UK, Ireland, and Norway, candidates for university admission apply through central offices, which process applications for multiple institutions (Eurydice, 2007). Turkey also relies on a centralized application system, which it uses in combination with a national university entrance exam, the Student Selection Examination (ÖSS). Scores on the ÖSS, which is intended to measure primarily verbal and quantitative reasoning, are used in combination with candidates’ high school grades (and in some cases, the results of a foreign-language examination) to determine eligibility for higher education programs (Student Selection and Placement Center (ÖSYM), 2006).

In other European countries, such as Sweden, the admissions process is fairly similar to that in the USA. Most Swedish universities require an entrance test, the Swedish Scholastic Aptitude Test (SweSAT), which consists of subtests in vocabulary, data sufficiency, Swedish reading comprehension, interpretation of diagrams, tables, and maps, and English reading comprehension (ERC). The test falls under the authority of the Agency for Higher Education but is developed and administered by the Department of Educational Measurement at Umeå University, with the exception of the ERC subtest, which is constructed at Göteborg University (Stage and Ögren, 2004). Originally developed for members of the work force who were at least 25 years of age and wanted to return to school (Wedman, 1995), the SweSAT is now used widely as a selection tool. Some institutions,

however, use only high school grades in making admissions decisions.

In Israel, admissions procedures closely parallel those in the USA. Universities generally require the Psychometric Entrance Test (PET), which consists of verbal reasoning, quantitative reasoning, and English sections. The PET is developed and administered by the National Institute for Testing and Evaluation, an independent body governed by the leaders of Israel's seven research universities (Beller, 2001). Universities develop their own admissions requirements and typically use a composite of PET scores and secondary school record in admissions decisions. The validity of the PET and of the composite are very similar to their American counterparts, as are the national concerns about the fairness of the PET to various groups and the possible advantages of test coaching.

At least two other countries have taken steps toward American-style college admissions testing: The British government is conducting a 5-year trial of the SAT, sponsored in part by the College Board (Labi, 2005), and Russia began nationwide use of a standardized admissions test similar to the SAT in 2009 (MacWilliams, 2007; Nemtsova, 2009).

See also: Ability Testing; Access and Equity in Higher Education; Assessment in Higher Education; Noncognitive Measures for Higher Education Admissions; Pathways and Articulation into Higher Education; Validity.

Bibliography

- ACT, Inc. (2007). *The ACT Technical Manual*. Iowa City, IA: ACT, Inc.
- ACT, Inc. (2009). *The ACT Writing Test Technical Report*. Iowa City, IA: ACT, Inc.
- Association of American Medical Colleges (2007). *2007 MCAT Essentials*. <http://www.aamc.org> (accessed August 2009).
- Beller, M. (2001). Admission to higher education in Israel and the role of the Psychometric Entrance Test: Educational and political dilemmas. *Assessment in Education* 8(3), 316–337.
- Breland, H., Maxey, J., Gernand, R., Cumming, T., and Trapani, C. (2002). Trends in college admission 2000: *A Report of a Survey of Undergraduate Admissions Policies, Practices, and Procedures*. (Sponsored by ACT, Inc., Association for Institutional Research, The College Board, Educational Testing Service, and the National Association for College Admission Counseling.) <http://www.airweb.org> (accessed October 2003).
- Burton, N. W. and Ramist, L. (2001). Predicting success in college: SAT studies of classes graduating since 1980. *Research Report 2001–2*. New York: College Entrance Examination Board.
- Burton, N. W. and Wang, M. (2005). Predicting long-term success in graduate school: A collaborative validity study. (Research Report 05–03). Princeton: Educational Testing Service.
- Camara, W. J. and Echternacht, G. (2000). The SAT and high school grades: Utility in predicting success in college. College Board Research Note RN-10. New York: College Entrance Examination Board.
- Camevale, A. P. and Rose, S. J. (2003). Socioeconomic status, race/ethnicity, and selective college admissions. *A Century Foundation Paper*. New York: The Century Foundation.
- Cole, N. S. and Moss, P. A. (1989). Bias in test use. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn., pp 201–219. New York: American Council on Education/Macmillan.
- Educational Testing Service (2003). *GRE 2003–2004 Guide to the Use of Scores*. Princeton, NJ: Educational Testing Service.
- Eurydice (The information network on education in Europe) (2007). Key data on education in Europe 2005. <http://www.eurydice.org> (accessed July 2007).
- Graduate Management Admission Council (2000). *GMAT Information Bulletin*. McLean, VA: Graduate Management Admission Council.
- Gulliksen, H. (1987). *Theory of Mental Tests*. Hillsdale, NJ: Erlbaum.
- Hawkins, D. A. and Lautz, J. (2005). *State of College Admission*. Alexandria, VA: National Association for College Admission Counseling. <http://www.nacacnet.org> (accessed January 2006).
- Johnson, E. K. and Edwards, J. C. (1991). Current practices in admission interviews at U. S. medical schools. *Academic Medicine* 66(7), 408–412.
- Julian, E. R. (2005). Validity of the Medical College Admission Test for predicting medical school performance. *Academic Medicine* 80(10), 910–917.
- Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., and Barbuti, S. M. (2008). Validity of the SAT for predicting first-year college grade point average. *College Board Research Report No. 2008–5*. New York: The College Board.
- Kuncel, N. R. and Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science* 315, 1080–1081.
- Kuncel, N. R., Hezlett, S. A., and Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin* 127, 162–181.
- Labi, A. (2005). England gives the SAT a try. *Chronicle of Higher Education* 52(6), A45.
- Lawrence, I., Rigol, G., Van Essen, T., and Jackson, C. (2004). A historical perspective on the content of the SAT. In Zwick, R. (ed.) *Rethinking the SAT: The Future of Standardized Testing in University Admissions*, pp 57–74. New York: RoutledgeFalmer.
- Law School Admission Council (1999). *LSAT and LSDAS Registration Information Book*. Newton, PA: Law School Admission Council.
- MacWilliams, B. (2007). Russian parliament approves standardized university admissions test. *Chronicle of Higher Education* 53(24), 1.
- Mattern, K., Camara, W., and Kobrin, J. L. (2007). SAT writing: An overview of research and psychometrics to date, Research Note 32. New York: College Board.
- Nemtsova, A. (2009, October 1). Russia cracks down on fraud in college entrance exam. *The Chronicle of Higher Education*. <http://chronicle.com/article/Russia-Cracks-Down-on-Fraud-in/48662/> (accessed October 2009).
- Norcross, J. C., Hanych, J. M., and Terranova, R. D. (1996). Graduate study in psychology: 1992–1993. *American Psychologist* 51, 631–643.
- Norris, D., Oppler, S., Kuang, D., Day, R., and Adams, K. (2005). *The College Board SAT Writing Validation Study: An Assessment of the Predictive and Incremental Validity*. Washington, DC: American Institutes for Research.
- Office of Technology Assessment (1992). *Testing in American Schools: Asking the Right Questions*. Washington, DC: U.S. Government Printing Office.
- Powers, D. E., Fowles, M. E., and Welsh, C. K. (1999). Further validation of a writing assessment for graduate admissions. *GRE Board Research Report No. 96-13R, ETS Research Report 99-18*. Princeton, NJ: Educational Testing Service.
- Rigol, G. W. and Kimmel, E. W. (1997). *A Picture of Admissions in the United States*. New York: The College Board and Educational Testing Service.
- Schmotter, J. W. (1993). The Graduate Management Admission Council: A brief history 1953–1992. *Selections* 9(2), 1–11.
- Sedlacek, W. E. (1998). Multiple choices for standardized tests. *Priorities* 10, 1–15.
- Stage, C. and Ögren, G. (2004). *The Swedish Scholastic Assessment Test (SweSAT): Development, Results and Experiences*. (EM No, 49). <http://www.umu.se> (accessed July 2007).
- Stewart, D. M. (1998). *Why Hispanic Students Need to Take the SAT*. The College Board. <http://www.collegeboard.org> (accessed April 1999).
- Stilwell, L. A., Dalessandro, S. P., and Reese, L. M. (2007). *Predictive Validity of the LSAT: A National Summary of the 2005–2006 Correlation Studies*. Newtown, PA: Law School Admission Council.

- Student Selection and Placement Center (ÖSYM) (2006). *Selection and Placement of Students in Higher Education Institutions in Turkey*. <http://www.osym.gov.tr> (accessed July 2007).
- Talento-Miller, E. and Rudner, L. M. (2005). GMAT validity study summary report for 1997 to 2004. *GMAC Research Report 05–06*. McLean, VA: Graduate Management Admission Council.
- Thomas, L. L., Kuncel, N. R., and Credé, M. (2007). Noncognitive variables in college admissions: The case of the Non-Cognitive Questionnaire. *Educational and Psychological Measurement* **67**(4), 635–657.
- Tracey, T. J. and Sedlacek, W. E. (1984). Noncognitive variables in predicting academic success by race. *Measurement and Evaluation in Guidance* **16**(4), 171–178.
- Webber, C. (1989). The mandarin mentality: University admissions testing in Europe and Asia. In Gifford, B. R. (ed.) *Test Policy and the Politics of Opportunity Allocation: The Workplace and the Law*, pp 33–57. Boston, MA: Kluwer.
- Wedman, I. (1995). Selection to higher education in Sweden. In Kellaghan, T. (ed.) *Admission to Higher Education: Issues and Practice*, pp 290–298. Dublin: Educational Research Centre.
- Willingham, W. W. (1985). *Success in College: The Role of Personal Qualities and Academic Ability*. New York: College Entrance Examination Board.
- Young, J. W. (2004). Differential validity and prediction: Race and sex differences in college admissions testing. In Zwick, R. (ed.) *Rethinking the SAT: The Future of Standardized Testing in University Admissions*, pp 289–301. New York: RoutledgeFalmer.
- Zwick, R. (2002). *Fair Game? The Use of Standardized Admissions Tests in Higher Education*. New York: RoutledgeFalmer.
- Zwick, R. (2006). Higher education admissions testing. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 647–679. Westport, CT: American Council on Education/Praeger.
- Zwick, R. (2007). College admissions testing. (Paper Commissioned by the National Association for College Admission Counseling). <http://www.nacacnet.org> (accessed July 2007).

Relevant Website

<http://www.fairtest.org> – The National Center for Fair and Open Testing (FairTest).

Appropriateness Measurement: Person Fit

R R Meijer, University of Groningen, Groningen, The Netherlands

A M Weekers, University of Twente, Enschede, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Person-fit analysis – A technique for determining if a person's results on educational or psychological test are valid. The purpose of a person-fit analysis is to detect item-score vectors that are unlikely given a hypothesized test theory model or unlikely compared with the majority of item-score vectors in the sample.

In educational and psychological standardized testing, measurement inaccuracy has received widespread attention. Examples are reliability theory, methods for estimating reliability, and statistics comparing groups with respect to the probability of correctly answering an item (differential item functioning). In this article, we discuss methods for determining the fit of an individual item-score pattern to a test model, that is, we discuss person-fit methods that can be used to investigate the validity of a test score. Validity of test scores may be threatened when factors other than the trait being measured, such as cheating, guessing, or plodding, influence an individual's test score.

In large-scale cognitive assessment, item response theory (IRT) modeling is often used for test construction, analysis, and scoring. Although empirical similarities of tests and questionnaires constructed according to classical test theory and IRT do exist, IRT offers better solutions to many psychometric problems than classical test theory (e.g., Embretson and Reise, 2000). The detection of invalid test scores, often called person-fit or appropriateness measurement, is an interesting and practically useful example (Meijer, 2003). Before we discuss different techniques used in person-fit research and its applications, we explain the basic principles of IRT and the idea of studying individual item-score patterns.

Item Response Theory

In most IRT models, test responses are assumed to be influenced by a single latent trait, denoted by the Greek letter θ . Both IRT models for dichotomous and polytomous item scores have been proposed. For dichotomous (correct/incorrect or true/false) data, the goal of fitting an IRT model is to identify an item response function (IRF) that describes the relation between θ and the probability

of correctly answering an item. In most IRT models, it is assumed that the probability of a correct response should increase as the trait level increases; thus, IRFs are monotonically increasing functions.

More formally, the IRF, denoted $P_g(\theta)$, gives the probability of a correct answer to item g ($g = 1, \dots, k$) as a function of θ . It is the probability of a correct response ($X_g = 1$) among persons with the latent trait value θ . For dichotomous items, $P_g(\theta)$ often is specified using the 1-, 2-, or 3-parameter logistic model (1-, 2-, 3PLM, see Embretson and Reise, 2000). The 3PLM is given by

$$P_g(\theta) = \gamma_g + \frac{(1 - \gamma_g) \exp[\alpha_g(\theta - \delta_g)]}{1 + \exp[\alpha_g(\theta - \delta_g)]}$$

where γ_g is the lower asymptote (γ_g is the probability of a correct score for persons with low θ 's, that is, $\theta \rightarrow -\infty$), α_g is the slope (item discrimination) parameter, and δ_g is the item location parameter.

The 2PLM can be obtained by fixing $\gamma_g = 0$ for all items, and the 1PLM or Rasch model by additionally fixing $\alpha_g = 1$ for all items.

As an alternative to these parametric IRT models, non-parametric IRT models have been proposed that only put order restrictions on the IRFs and do not assume a specific parametric function (e.g., Sijtsma and Meijer, 2007).

For polytomous item scores (more than two answer categories), polytomous IRT models have been proposed. In these models, category response functions specify the relation between endorsing a particular category and the latent trait. In this article, we mainly focus on person-fit statistics for dichotomous IRT models and refer to person-fit statistics for polytomous data and some important work in this area.

Studying Individual Item-Score Patterns

By means of an IRT model it is possible to predict a person's answering behavior when confronted with a particular set of questionnaire items. Let us illustrate this by means of **Figure 1**. For the sake of simplicity, we depicted five IRFs that do not intersect across the latent trait range. These IRFs comply with the Rasch model. Assume that we have an estimate of someone's trait level to be $\theta = 0$, then the probability of answering item 1 correctly equals 0.9 (this is the easiest item) and the probability of answering item 5 correctly equals 0.1 (this is the most difficult

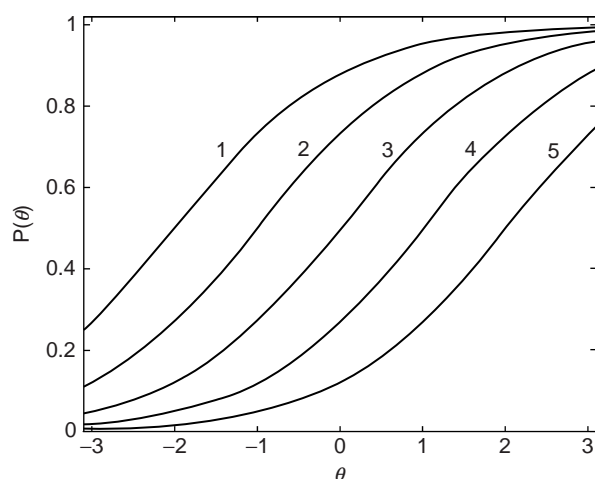


Figure 1 Item response functions that comply to the Rasch model. Meijer RR, Egberink IJL, Emons WHM, et al., (2008). *Journal of Personality Assessment* 90(3), 227–238.

item). Suppose now that the items are ordered to increasing item difficulty and that a person answers three items correctly, then the item-score pattern that has the highest probability of occurrence is 11100 and the item-score pattern with the lowest probability of occurrence is 00111. This second pattern thus is unexpected (misfitting or aberrant) and it may be questioned whether the total score of 3 has the same meaning for both patterns.

Several person-fit statistics and statistical tests have been proposed to identify unexpected item-score patterns (Meijer and Sijtsma, 2001). Most statistics have been proposed for parametric IRT models. We first, however, discuss a statistic that can be used in the context of nonparametric IRT modeling or using classical test theory.

Nonparametric IRT-Based and Group-Based Person-Fit Statistics

Nonparametric IRT-based or group-based person-fit statistics are statistics that compare the score of an individual against score patterns of other persons in the group. An individual's item-score pattern is evaluated given the nonparametric IRT model that can be applied to the group's data.

An often-used and simple person-fit statistic is the number of Guttman (1950) errors found in an individual's item-score pattern. Given that the items are ordered according to increasing difficulty, for dichotomous item scores the number of Guttman errors is simply defined by the number of 0 scores to the left of each 1 score. Thus, for example, the pattern (1110101) contains three Guttman errors (three (0,1) item pairs) (see Emons *et al.* (2005) for an application of this statistic).

For polytomous (e.g., Likert scale) items, the popularity of the item steps can be determined and the item steps

can then be ordered according to decreasing popularity. An item step is the imaginary threshold between adjacent ordered response categories. A Guttman error consists of endorsing a so-called less-popular item step in one item, while not endorsing a more popular item step in another item. To illustrate this, consider a scale that consists of six items with four response alternatives (coded 1 through 4). This implies that there are three item steps per item (from 1–2, 2–3, and 3–4). Thus, there are $6 \times 3 = 18$ item steps for each person. As an example, we consider two score patterns. One observed score pattern is (323322). From this pattern we obtained the item step score pattern 11111001101000000 with item steps ordered to decreasing popularity. These popularities can be obtained from the observed proportion of respondents who endorse a particular response category. This pattern contains seven Guttman errors (seven pairs of 0 before 1 scores). Consider now another pattern (144414) with the corresponding item step pattern 011100111010101111 which contains 45 Guttman errors. This latter pattern can thus be considered more unexpected than the first pattern. The question, however, is when do we classify a pattern as unexpected or misfitting?

In general, let t be the observed value of a person-fit statistic T . Then, the significance probability or probability of exceedance is defined as the probability under the sampling distribution that the value of the test statistic is smaller than the observed value: $p^* = P(T \leq t)$, or larger than the observed value: $p^* = P(T \geq t)$, depending on whether low or high values of the statistic indicate aberrant item-score patterns.

Nonparametric or group-based person-fit statistics, like the number of Guttman errors, are mainly used in a descriptive way and use some rule of thumb to classify a score pattern as misfitting. Sampling distributions that can be used to classify a pattern as fitting or misfitting are unknown. Although it may be argued that this is not a serious problem as long as one is only interested in the use of a person-fit statistic as a descriptive measure, a more serious problem is that the distribution of the numerical values of most group-based statistics is dependent on the total score (e.g., Drasgow *et al.*, 1987). This dependence implies that when one critical value is used across total scores, the probability of classifying a score pattern as aberrant is a function of the total score, which is undesirable.

Parametric IRT-Based Person-Fit Statistics

In parametric IRT-based person-fit, an observed item-score pattern is compared to the predicted item-score pattern under certain restrictions derived from a parametric IRT model. Most person-fit statistics are sensitive to unlikely response behavior and do not test against

specific model violations. For example, Wright and Stone (1979) proposed mean squared residual-based statistics such as U , which is given by

$$U = \sum_{g=1}^k \frac{[X_g - P_g(\theta)]^2}{kP_g(\theta)[1 - P_g(\theta)]}.$$

The nominator contains the difference between observed and expected item scores; the denominator contains the conditional variances of the individual item scores. U can be interpreted as the mean of the squared standardized residuals based on k items.

Perhaps the most popular person-fit statistic in the literature is the log-likelihood function:

$$l_0 = \sum_{g=1}^k \{X_g \ln P_g(\theta) + (1 - X_g) \ln [1 - P_g(\theta)]\}$$

This statistic was first used by Levine and Rubin (1979) and has been further developed and applied by, for example, Drasgow *et al.* (1985). l_0 is determined as the logarithm of the likelihood function evaluated at the maximum likelihood estimate of θ . Two problems occur when using l_0 as a fit statistic. (1) l_0 is not standardized implying that the classification of an item score pattern as model-fitting or misfitting depends on θ . (2) For classifying an item-score pattern as misfitting, a distribution under the null hypothesis of fitting scores is needed. The null distribution is unknown for l_0 . To overcome these problems, Drasgow *et al.* (1985) proposed a standardized version of l_0 that is less confounded with θ and purported to be standard normally distributed.

The standardized version of l_0 is given by

$$l_z = \frac{l_0 - E(l_0)}{[Var(l_0)]^{1/2}},$$

where $E(l_0)$ and $Var(l_0)$ are the expectation and variance of l_0 , respectively:

$$E(l_0) = \sum_{g=1}^k \{P_g(\theta) \ln [P_g(\theta)] + [1 - P_g(\theta)] \ln [1 - P_g(\theta)]\},$$

and

$$Var(l_0) = \sum_{g=1}^k P_g(\theta) [1 - P_g(\theta)] \left[\ln \frac{P_g(\theta)}{1 - P_g(\theta)} \right]^2,$$

Snijders (2001) proposed further refinements accounting for the fact that in practice $\hat{\theta}$ was used instead of θ and Glas and Dagohoy (in press) discussed distributions for polytomous item scores.

In contrast to the person-fit statistics discussed above, Levine and Drasgow (1988) proposed a statistically optimal method for the identification of misfitting item-score patterns against a particular alternative model. No other method can achieve a higher rate of detection at the same type I error rate. A likelihood ratio statistic was proposed

that provides the most powerful test for the null hypothesis that an item-score pattern is model fitting. A model for model-fitting behavior (e.g., 1-, 2-, or 3PLM) and a model for a particular type of misfitting behavior (e.g., a model with violations of local independence) are specified in advance. Other examples for testing against specific model assumptions can be found in Klauer (1995).

An alternative approach was proposed by Trabin and Weiss (1983). They used the person response function (PRF) to identify aberrant item-score patterns. At a fixed θ value, the PRF specifies the probability of a correct response as a function of the item location δ . In IRT, the IRF often is assumed to be a nondecreasing function of θ , whereas the PRF is assumed to be a nonincreasing function of δ . To construct an observed PRF, Trabin and Weiss (1983) ordered items to increasing δ values and then formed subtests of items by grouping items according to $\hat{\delta}$ values. For fixed $\hat{\theta}$, the observed PRF was constructed by determining, in each subtest, the mean probability of a correct response. The expected PRF was constructed by estimating according to the 3PLM, in each subtest, the mean probability of a correct response. A large difference between the expected and observed PRFs was interpreted as an indication of misfitting responses for that examinee. For developments with respect to the PRF, see Sijsma and Meijer (2001) and Ferrando (2007).

Depending on the type of data and the problems envisaged, a researcher may choose a particular statistic; however, not all statistics have equally favorable properties in a statistical sense. For example, for short tests and tests of moderate length (say, 10–60 items) and using the standard normal distribution, due to the use of $\hat{\theta}$ rather than θ for most statistics the nominal type I error rate is not in agreement with the empirical type I error rate. Recently, statistical theory for correcting the bias caused by using the maximum-likelihood estimate $\hat{\theta}$ rather than θ was developed (e.g., Snijders, 2001; Glas and Dagahoy, in press). Sound statistical methods have been derived for the Rasch model, but because this model is rather restrictive to empirical data, the use of these statistics is also restricted.

In general, it may be wise to first investigate possible threats to the fit of individual item-score patterns before using a particular person-fit statistic. As another example, if violations against unidimensionality are expected, one of the methods proposed by Klauer (1995) may be used instead of a general statistic such as the l_z statistic (see also Molenaar and Hoijtink, 1990). Not only are tests against a specific alternative more powerful than general statistics, but also the type of deviance is easier to interpret. Statistics such as l_z statistic are helpful in situations when the researcher has no idea which threats are most realistic. Statistics such as U or the PRF can be used as diagnostic tools to test whether item-score patterns on *a priori* specified subtests fit the IRT model.

Furthermore, research (see, e.g., Meijer and Sijtsma, 2001) has shown that detection rates are highly dependent on: (1) type of misfitting behavior (more unexpected item scores result in more power), (2) θ value (in general, at the extremes, power is lesser than in the middle of the distribution), and (3) test length (longer tests result in more power).

Applications

The literature on person fit is mainly technical in the sense that there are many studies devoted to the psychometric characteristics of the statistics and tests (such as the correct sampling distribution) but there are very few studies that illustrate the usefulness of these statistics in practice (e.g., Meijer and Sijtsma, 2001). There is a gap between the often very sophisticated articles devoted to the psychometric characteristics of several statistical tests and measures, on the one hand, and the articles that describe the practical usefulness of these measures, on the other. Two types of applications can be distinguished: (1) studies on the usefulness of person-fit statistics and their link to the respondents' characteristics and (2) the usefulness of the statistics in new applications such as computer adaptive testing (CAT).

Misfitting Patterns and Person's Characteristics

Several researchers applied person-fit statistics in practice. For example, Lamprianou and Boyle (2004) analyzed data of the National Curriculum tests in mathematics in England and found that pupils having English as an additional language and pupils belonging to ethnic minorities are significantly more likely to generate aberrant response patterns. Furthermore, they concluded that pupils having English as an additional language and pupils belonging to ethnic minorities produce significantly more misfitting patterns than the remainder of pupils by taking the mathematics National Curriculum tests.

Reise and Waller (1993) analyzed 11 unidimensional personality scales that had well-defined content clusters and were therefore able to link unsalability to aberrant responding for specific types of item content. In general, however, one should be careful when interpreting misfitting item-score patterns. Reise and Waller (1993) discussed that interpreting misfitting item-score patterns is difficult. Possible causes may be response faultiness, misreading, or random responding. Thus, person-fit statistics allow the identification of a deviant item-score pattern but not the recovery of the mechanism that created a deviant item-score pattern.

As a researcher usually does not know the cause of an atypical answering behavior, background information

about individual persons needs to be incorporated into the diagnostic process. Depending on the application, such information may come from previous psychological-ability and achievement testing, school performance (tests and teacher's accounts), clinical and health sources (e.g., about dyslexia, and learning and memory problems), or social-economic indicators (e.g., related to language problems at home). Although in many studies it has been suggested that quantitative and qualitative information should be combined, there are few studies where this has been done.

Recently, Meijer *et al.* (in press) integrated psychometric analysis with information from qualitative sources to make judgments about the validity of an individual's test score. More specifically, they (1) explored the usefulness of person-fit statistics to identify invalid test scores using real data from self-concept theory in a sample of children and (2) validated information obtained from IRT using personality theory and qualitative data obtained from observation and interviews. They showed that invalid test scores were the result of a less-developed self-concept in young children. We hope that future research will concentrate more on how person-fit statistics can be used in practice.

Computer Adaptive Testing

A new field of application for person-fit research is CAT (e.g., van der Linden and Glas, 2000). In CAT, the difficulty of the test is adapted to the θ level of the candidate. With the increasing use of CAT, the use of person-fit statistics may be helpful to detect item memorization or to detect examinees who, as a result of continuous test administration from the same item bank, are familiar with some of the items. Research showed that, in CAT, the distributional characteristics of existing person-fit statistics are not in agreement with the expected theoretical distributions (e.g., van Krimpen-Stoop and Meijer, 2001). This may be explained by the modest spread in the item difficulties, which results in an under-dispersion of the assumed null distribution of the person-fit statistics. Consequently, empirical type I errors are too small compared to nominal type I errors. Person-fit statistics that are especially designed for a CAT may be more powerful than conventional person-fit statistics, and the statistical properties of the former statistics should be less susceptible to the characteristics of a CAT. Both Bradlow *et al.* (1998) and van Krimpen-Stoop and Meijer (2001) proposed person-fit statistics that make use of the property of CAT that a fitting item-score pattern will consist of an alternation of correct and incorrect responses, especially at the end of the test when $\hat{\theta}$ comes closer to θ . Item-score patterns that consist of strings of 1 scores or 0 scores are unexpected. Another recent development is the link between response times administered in a CAT and the detection of misfitting response behavior (e.g., Van der

Linden and van Krimpen-Stoop, 2003). Very short response times to difficult items may indicate cheating.

Summary

We discussed methods that can be used to investigate the aberrance of individual item-score patterns under particular IRT models. The methods ranged from statistics for testing whether an item-score pattern is in agreement with the other patterns in the sample (group-based statistics) to methods for investigating whether there are specific violations against model assumptions (parametric IRT-based statistics).

A drawback of some person-fit statistics is that only deviations against the model are tested. This may result in interpretation problems. For example, item-score patterns that do not fit the Rasch model may be described more appropriately by means of the 3PLM. If the Rasch model does not fit the data, other explanations are possible. Because in practice it is often difficult, if not impossible, to substantially distinguish different types of item-score patterns and/or to obtain additional information using background variables, a more fruitful strategy may be to test against specific alternatives.

See also: Item Response Theory.

Bibliography

- Bradlow, E. T., Weiss, R. E., and Cho, M. (1998). Bayesian identification of outliers in computerized adaptive testing. *Journal of the American Statistical Association* **93**, 910–919.
- Drasgow, F., Levine, M. V., and McLaughlin, M. E. (1987). Detecting inappropriate test scores with optimal and practical appropriateness indices. *Applied Psychological Measurement* **11**, 59–79.
- Drasgow, F., Levine, M. V., and Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology* **38**, 67–86.
- Embretson, S. E. and Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum.
- Emons, W. H. M., Sijtsma, K., and Meijer, R. R. (2005). Global, local, and graphical person-fit analysis using person-response functions. *Psychological Methods* **10**, 101–119.
- Ferrando, P. J. (2007). A Pearson-type-VII item response model for assessing person fluctuation. *Psychometrika* **72**, 25–41.
- Glas, C. A. W. and Dagohoy, A. V. (in press). Lagrange multiplier person-fit tests for polytomous IRT models. *Psychometrika*.
- Guttman, L. (1950). The basis for scalogram analysis. In Stouffer, S. A., Guttman, L., Suchman, E. A., et al. (eds.) *Measurement and Prediction*, pp 60–90. Princeton, NJ: Princeton University Press.
- Klauser, K. C. (1995). The assessment of person-fit. In Fischer, G. H. and Molenaar, I. W. (eds.) *Rasch Models. Foundations, Recent Developments, and Applications*, pp 97–110. New York: Springer.
- Lamprianou, I. and Boyle, B. (2004). Accuracy of measurement in the context of mathematics national curriculum tests in England for ethnic minority pupils and pupils who spend English as an additional language. *Journal of Educational Measurement* **41**, 239–259.
- Levine, M. V. and Drasgow, F. (1988). Optimal appropriateness measurement. *Psychometrika* **53**, 161–176.
- Levine, M. V. and Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational Statistics* **4**, 269–290.
- Meijer, R. R. (2003). Diagnosing item score patterns on a test using item response theory-based person-fit statistics. *Psychological Methods* **8**, 72–87.
- Meijer, R. R., Egberink, I. J. L., Emons, W. H. M., and Sijtsma, K. (in press). Detection and validation of unscalable item score patterns using item response theory: An illustration with Harter's self-perception profile for children. *Journal of Personality Assessment*.
- Meijer, R. R. and Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement* **25**, 107–135.
- Molenaar, I. W. and Hoijtink, H. (1990). The many null distributions of person fit indices. *Psychometrika* **55**, 75–106.
- Reise, S. P. and Waller, N. G. (1993). Traitendness and the assessment of response pattern scalability. *Journal of Personality and Social Psychology* **65**, 143–151.
- Sijtsma, K. and Meijer, R. R. (2001). The person response function as a tool in person-fit research. *Psychometrika* **66**, 191–207.
- Sijtsma, K. and Meijer, R. R. (2007). Nonparametric item response theory and special topics. In Rao, C. R. and Sinharay, S. (eds.) *Handbook of Statistics*, vol. 26, pp 719–746. Amsterdam: Elsevier B.V.
- Snijders, T. A. B. (2001). Asymptotic distribution of person-fit statistics with estimated person parameter. *Psychometrika* **66**, 331–342.
- Trabin, T. E. and Weiss, D. J. (1983). The person response curve: Fit of individuals to item response theory models. In Weiss, D. J. (ed.) *New Horizons in Testing*. New York: Academic Press.
- Van Krimpen-Stoop, E. M. L. A. and Meijer, R. R. (2001). Detecting person-misfit in adaptive testing using statistical process control techniques. In van der Linden, W. J. and Glas, C. A. W. (eds.) *Computerized Adaptive Testing: Theory and Practice*. Boston: Kluwer.
- Van der Linden, W. J. and Glas, C. A. W. (2000). *Computerized Adaptive Testing: Theory and Practice*. Boston: Kluwer.
- Van der Linden, W. J. and van Krimpen-Stoop, E. M. L. A. (2003). Using response times to detect aberrant responses in computerized adaptive testing. *Psychometrika* **68**, 251–265.
- Wright, B. D. and Stone, M. H. (1979). *Best Test Design. Rasch Measurement*. Chicago, IL: Mesa Press.

Further Reading

- Emons, W. H. M., Glas, C. A., W, Meijer, R. R., and Sijtsma, K. (2003). Person-fit in ordered-restricted latent-class models. *Applied Psychological Measurement* **27**, 459–478.
- Meijer, R. R. (1994). The number of Guttman errors as a simple and powerful person-fit statistic. *Applied Psychological Measurement* **18**, 311–314.
- Smith, R. M. (1986). Person fit in the Rasch model. *Educational and Psychological Measurement* **46**, 359–372.
- Tatsuoka, K. K. (1984). Caution indices based on item response theory. *Psychometrika* **49**, 95–110.

Automated Essay Scoring: Writing Assessment and Instruction

M D Shermis, The University of Akron, Akron, OH, USA

J Burstein, D Higgins, and K Zechner, Educational Testing Service, Princeton, NJ, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Automated essay scoring – A measurement technology in which the computer evaluates human written responses.

Introduction

This article documents the advent and rise of automated essay scoring (AES) as a means of both assessment and instruction. The first section discusses what AES is, how it works, and who the major purveyors of the technology are. The second section describes outgrowths of the technology as it applies to ongoing projects in measurement and education.

In 1973, the late Ellis Page and colleagues at the University of Connecticut programmed the first successful AES engine, Project Essay Grade (PEG) (1973). The technology was foretold some 6 years earlier in a landmark *Phi Delta Kappan* article titled, The imminence of grading essays by computer (Page, 1966). At the time the article was provocative and a bit outrageous, though in hindsight, it can only be deemed prophetic. As a former high school English teacher, Page was convinced that students would benefit greatly by having access to technology that would provide quick feedback on their writing. He also realized that the greatest hindrance to having secondary students write more was the requirement that, ultimately, a teacher had to review stacks of papers. While PEG produced impressive results, the technology of the time was too primitive to make it a practical application. Text had to be typed on IBM 80-column punched cards and read into a mainframe computer before it could be evaluated. As a consequence, the technology sat dormant until the early 1990s and was revitalized with the confluence of two technological developments: microcomputers and the Internet. Microcomputers permitted the generation of electronic text from a regular keyboard and the Internet provided a universal platform to submit text for review (Shermis *et al.*, 2001).

AES is a measurement technology in which computers evaluate written work (Shermis and Burstein, 2003). Most of the initial applications have been in English, but past work has been applied to Japanese (Kawate-Mierzejewska, 2003, March), Hebrew (Vantage Learning, 2001), and

Bahasa Malay (Vantage Learning, 2002). Computers do not understand the written text being evaluated. So, for example, the computer would not get the following joke.

Q-When is a door not a door?

A-When it is ajar.

Unlike humans, a computer cannot interpret the play on words, and infer that the predicate in the answer (i.e., ajar) is being cleverly used as a noun (i.e., a jar).

What the computer does in an AES context is to analyze the written text into its observable components. Different AES systems evaluate different numbers of these components. Page and Petersen (1995) referred to these elements as proxies or approximations for underlying traits (i.e., intrinsic characteristics) of writing. It is the observable components that AES engines, identify, computationally, and subsequently use to compute essay scores. AES statistical models are developed by weighting the various observable components as they relate to intrinsic characteristics of writing. For example, a model in the PEG system, might be formed by taking five intrinsic characteristics of writing (content, creativity, style, mechanics, and organization) and linking proxies. An example of a simple proxy is essay length. The empirical evidence suggests that the longer an essay is, the more highly valued it is by a rater. This could be because the writer provides additional details which improves the essay's standing in the eyes of the grader. However, this relationship is not linear, but logarithmic. More specifically, it appears as if essay length is important up to a point, but beyond a certain threshold it carries little additional weight. Where that point is becomes a function of the average essay length. If most essays have 150 words and a candidate essay has 200 words, then the writer has probably already taken advantage of any contribution that essay length would influence in a rater's decision-making.

Again, in the PEG framework, an example of a complex proxy might be a count of the number of times because is used in an essay. Though this counting method is admittedly superficial, it is an indicator of sentence complexity, and is also tied conceptually to the intrinsic characteristic of style. It is useful to note that this one proxy is joined with a number of others to estimate the trait, and would not normally be used as a single indicator for the trait.

How are models formed? There are three general approaches, two of which involve the collection of empirical data, and one that is currently more of a theoretical option.

In the first approach, two samples of essays are collected – one for model building and the other for model evaluation/confirmation – each of which has been rated by multiple raters. For the purpose of creating a model, one might utilize the field distribution of ratings or select a fixed sample size (e.g., 300) with approximately an equal number of essays at each cut point. Using the latter technique if there were six points on the rating rubric that raters used, the model would be most discriminating if there were 50 essays for each point of the scale ($300/6 = 50$). As rating points at the extremes of the scale are more difficult to come by, one may need a pool of more than 500 essays from which to draw the data if using the field distribution of ratings. As alluded to previously, the prox (or prox cluster) variables are regressed against the essay ratings (or in the case where the model is not formulated on an a priori basis, to select the variables and optimize the weights). The validation sample (e.g., 200 cases) is used to evaluate the results from the first set of estimates. Most AES developers use multiple regression to create their models, but one developer uses multiple statistical techniques, and then selects the one that explains the most variance.

In the second approach, the evaluation of content may be accomplished through the specification of vocabulary (i.e., the evaluation of the other aspects of writing is performed as described above). Latent semantic analysis and its variants are employed by some developers to provide estimates as to how close the vocabulary in the candidate answer is to a targeted vocabulary set (Landauer *et al.*, 1998).

The third approach is to develop models that are based on a gold standard formulated by experts. To date, this mechanism for developing models is more theoretical than applied. If normative models can be created for the relevant dimensions of writing (i.e., age and writing genre), then other variables could be tailored on an a priori basis to generate a statistical blueprint for the ideal response. The blueprint may or may not be aligned with human ratings. For example, the guidelines for a few high-stakes writing programs explicitly direct raters to ignore expressions of nonstandard English. However, raters find this a difficult challenge, even when exposed to comprehensive training programs. When presented with an expression of nonstandard English, a typical rater will inevitably undervalue the essay even though an answer may be functionally equivalent to a response given in standard English. AES would have the capacity to overcome this human limitation if the relevant affected variables associated with nonstandard English can be isolated and adjusted.

AES Programs

Presently, there are three major developers of AES. The Educational Testing Service (ETS) has e-rater which is a

component of Criterion, a comprehensive electronic portfolio administration system. E-rater is also used as a scoring application for high- and low-stakes assessments, as well as a number of test-practice applications. Vantage Learning has developed Intellimetric which is also part of an electronic portfolio administration system called My-Access. Finally, Pearson Knowledge Technologies supports the Intelligent Essay Assessor which is used by a variety of proprietary electronic portfolio systems. All of the products have the capacity to receive text via a webpage and return feedback to both a student user and comprehensive database that may be accessed by teachers. In the paragraphs below, a short description is given that illustrate the kinds of factors/dimensions/variables used in building AES scoring models. References are provided for a more comprehensive description of the process.

The construction of e-rater v. 2.0 models is given in detail in Attali and Burstein (2006). (E-rater v.2.0 is the infrastructure for all subsequent version of e-rater. Currently, e-rater v.9.1 has been released. Some of the mathematical transformations implemented in the earlier e-rater v. 2.0 have been modified in more current versions.) It is composed of up to 12 features used by e-rater v.2.0 to score essays. These 12 features are associated with six areas of analysis: errors in grammar, usage, and mechanics (Leacock and Chodorow, 2003); style (Burstein, 2003); identification of organizational segments, such as thesis statement (Burstein *et al.*, 2003); and vocabulary content (Attali and Burstein, 2006).

Eleven of the individual features reflect essential characteristics in essay writing and are aligned with human-scoring criteria. These features included related to: (1) proportion of errors in grammar, (2) proportion of word-usage errors, (3) proportion of mechanical errors, (4) proportion of style comments, (5) number of required discourse elements, (6) average length of discourse elements, (7) score assigned to essays with similar vocabulary, (8) similarity of vocabulary to essays with score 6, (9) number word types divided by number of word tokens, (10) log frequency of least common words, (11) average length of words, and (12) total number of words (Attali and Burstein, 2006). (This feature is not used in the current version of e-rater.) E-rater uses a sample of human-scored essay data for model building purposes. E-rater identifies features and feature weights are assigned using a multiple regression procedure. E-rater models can be built at the topic level, in which case a model is built for a specific essay prompt. However, more often, e-rater models are built at the grade level. So, for instance, a model is built for sixth-grade writers in Criterion. Writers can respond to topics selected by the teacher from the set of Criterion prompts, or the teacher can assign his or her own topic, and the sixth-grade model will be used to score these teacher–topic responses.

A comprehensive specification of the Intellimetric model is given in Elliot (2003). The model selects from 500

component features (i.e., proxies) (and clusters the selected elements into at least five consolidated sets). These sets include content, word variety, grammar, text complexity, and sentence variety. Other dimensions of writing may be used, but these five are common to Intellimetric models. Intellimetric uses word nets based on Latent semantic dimension which is similar in nature to LSA (i.e., it determines how close the candidate response is, in terms of content, to a modeled set of vocabulary). Word variety refers to word complexity or word uniqueness. The grammar composite that evaluates things like subject-verb agreement, and text complexity is similar in nature to ascertaining the reading level of the text. The information gleaned is then used by a series of independent mathematical judges, or mathematical models, to predict the expert human scores and then optimized to produce a final predicted score. Typically, the judge that explains the largest proportion of rater variance will be employed in model development.

The technical details of the Intelligent Essay Assessor (IEA) are highlighted in Landauer *et al.* (2003). IEA is modeled using a two-pronged approach. The content of the essay is assessed by using a combination of external databases and LSA. For example, if the writing prompt had to do with the differentiation among Freud's concepts of superego, ego, and id, the reference database might include the electronic version of an introductory text in psychology. From that database, LSA would determine the likelihood of encountering certain words (e.g., the term conscience as a synonym for superego) given the constellation of vocabulary in the reference text. A candidate essay with more relevant vocabulary will be awarded a higher score. In setting up the models, IEA incorporates a validation procedure to check that LSA scores are aligned with those that might be given by human raters.

In contrast to e-rater and Intellimetric, the noncontent features (e.g., mechanics, style, and organization) of IEA are not fixed, but rather are constructed as a function of the domains assessed in the rating rubric. The weights for prox variables associated with these domains are predicted based on human ratings, and then are combined with the score calculated for content.

Reliability and Validity

As AES models often formed by using more than two raters, studies that have evaluated interrater agreement have usually showed that the agreement coefficients between the computer and human raters is at least as high or higher than among human raters themselves (Elliot, 2003; Landauer *et al.*, 2003; Page and Petersen, 1995). All AES engines have obtained exact agreements with humans as high as the mid-1980s and adjacent agreements in the mid-high 1990s – slightly higher than the

agreement coefficients for trained human raters. Several validity studies have suggested that AES engines tap the same construct as that being evaluated by human raters. Page *et al.* (1995) examined the construct validity of AES, Keith (2003) summarized several discriminant and true-score-validity studies of the technology, and Attali and Burstein (2006) demonstrated the relationship between AES and instructional activities associated with writing.

AES is not without its detractors. Ericsson and Haswell (2006) performed a comprehensive critique of the technology from the perspective of those who teach postsecondary writing. Objections to the technology ranged from a concern about the ethics of using computers rather than humans to teach writing to the lack of synchronicity between how human graders approach the rating task and the process by which AES evaluates a writing sample to failed implementations of AES in university placement-testing programs. Clearly, there are certain types of stylized text writing that AES may never be able to evaluate. Nevertheless, AES is now used as a scoring process for high-stakes tests (e.g., GMAT) and is provided as a common instructional intervention for writing.

AES was a technology trigger that has spawned several related, and new innovative education technologies. In the next section we provide descriptions of emerging technology that, based on AES, has migrated to other measurement domains.

Transformations into New Applications

The success of AES and short-answer scoring (Leacock and Chodorow, 2003) has set the stage for a number of new capabilities developed for text-based analysis for enhanced feedback related to technical and organizational writing quality to help both native and non-native English speakers, and applications that incorporate text-analysis capabilities to provide reading-comprehension support for English-language learners (ELLs). Until now, the majority of AES and related capabilities have focused on text. However, speech-based capabilities are also making their way into commercial applications. In light of this, the second half of this section is focused a discussion of a speech-based, instructional capability currently used for scoring the speech of ELLs.

Criterion Online Essay Evaluation Service

As mentioned above, automated essay-scoring engines are typically combined with electronic portfolio systems to provide a full-spectrum set of services for those involved with writing instruction and assessment. In this section, a description of the Criterion online essay evaluation service is provided. The application is designed to help

teachers in K-12 classrooms, in community college, and university classrooms who typically have a large number of writing assignments to grade. This limits the number of writing assignments that teachers can offer to students. In an effort to offer additional writing practice to students, researchers have sought to develop applications not only for AES, but also for offering more descriptive essay feedback similar to teacher feedback of student writing: indications of grammar, usage, and mechanics errors, stylistic, and organization and development issues. Pioneering work in automated feedback of this kind was initiated in the 1980s with the Writer's Workbench (MacDonald *et al.*, 1982), and continues in applications, including the Criterion online essay evaluation service (Burstein *et al.*, 2004), the AOL Writing Wizard, and Vantage Learning MY Access!®.

The Criterion online essay evaluation service combines e-rater AES, advisories indicating anomalies, such as off-topicality (Higgins *et al.*, 2006), and descriptive feedback. The descriptive feedback is comprised of a suite of programs that evaluate and, subsequently, flag essays for errors in grammar, usage, and mechanics; identify an essay's discourse structure; and recognize undesirable stylistic features. As the population of ELLs grows, researchers are working on enhancements to the grammatical error-detection component to accommodate the kinds of mistakes more common in the ELL population. This kind of feedback includes determiner and preposition errors (Han *et al.*, 2004), and collocation errors (e.g., strong computer instead of powerful computer) (Futagi *et al.*, 2008; Pantel and Lin, 2000; Shei and Pain, 2000).

Criterion offers a prewriting (planning) utility. This emphasis on planning was a logical outgrowth of the process-writing approach that Criterion embodies. Both earlier literature (Elbow, 1973) and more recent research have suggested that making plans can help students produce better-quality writing, just as revising drafts can (Chai, 2004; Goldstein and Carr, 1996). In light of this research, it is advisable to incorporate formal planning activities into writing-instruction applications. The ability to collect student-planning data through formal planning applications provides a new and authentic data source that can be used in writing research. Other computer-based instructional systems also offer a planning tool, including Inspiration Software, Inc., which offers elaborate graphic organizers for writing and research projects, while online writing-instruction applications such as the AOL Writing Wizard, CompassLearning Odyssey Writer, and Vantage Learning MY Access provide onscreen planning tools to aid in the process of composition.

Generally speaking, researchers continue to develop capabilities for online writing instruction programs that are aligned with different populations of students and their respective needs with regard to their becoming more proficient writers.

Text Adaptor: Technology to Support ELLs

Authentic texts for the classroom that are grade-level appropriate and accessible to ELLs are often unavailable, especially in middle school and high school. As a result, the time-consuming practice of manual text adaptation has become a required task for both ESL and content-area teachers. Research suggests that certain kinds of text modifications, specifically vocabulary expansion and elaboration (i.e., providing synonyms and native-language cognates) can facilitate students' comprehension of content in a text (Perez, 1981; Carlo *et al.*, 2004; Fitzgerald, 1995; Hancin-Bhatt and Nagy, 1994; Ihnot, 1997; James and Klein, 1994; Jimenez *et al.*, 1996; Nagy *et al.*, 1993; Oh, 2001; Yano *et al.*, 1994).

Text Adaptor, a web-based tool, was designed as an authoring tool to support K-12 teachers in the text-adaptation practice. While we continue to develop the tool, it currently incorporates several natural language processing (NLP) capabilities to support automated generation suggested text modifications for classroom texts. Tool suggestions are similar to the kinds of adaptations that teachers might implement for ELLs in their content-area classes.

Text Adaptor allows users to import a text or webpage, and subsequently, to generate the following types of adaptations of the imported text: English and Spanish text summaries, vocabulary support, including synonym (Lin, 1998), antonym, and Spanish/English cognate identification (WordNet lexical database). Text Adaptor also identifies complex phrasal and sentence structures, and academic vocabulary, fixed phrases (e.g., phrasal verbs and collocations), and cultural references. Teachers can then modify the text accordingly, given the learning needs of the ELL students in their classrooms. NLP capabilities used to generate these adaptations include, automatic summarization (Marcu, 2000), machine translation (Language Weaver's English-to-Spanish machine-translation system), and synonym and antonym identification. The adaptation capabilities include strategies used by teachers to manually create adaptations, such as summaries and varied vocabulary support, as well as translate a text into another language. Teachers can use Text Adaptor to author any kind of classroom text, including reading texts, activities, and assessments.

As part of this research, a 2008 pilot study was conducted in two online teacher professional development (TPD) for ELL teachers in the United States: one at a large, private university on the west coast, and another at a large, private university on the east coast (Burstein, 2009; Shore *et al.*, 2009). A central purpose of the pilot was to gauge whether Text Adaptor increased teachers' linguistic awareness, resulting in improved text adaptations for ELLs. A pre-post test design was implemented with approximately 70 teachers enrolled in the TPD courses. The pilot activities were integrated into the respective

TPD courses. All participants completed online background surveys about their educational and teaching experiences, and postsurveys that asked the control group about their experience adapting texts in the pilot study, and asked the treatment group about their experiences using Text Adaptor. All participants completed manual (pre-) adaptations to gauge baseline adaptation knowledge and ability. Both completed a mid- and post-adaptation activity. Control teachers completed these activities manually, while treatment teachers were trained to use Text Adaptor and complete their adaptations using the tool. An important outcome indicated that all teachers who participated in this pilot gained knowledge about linguistic features as a result of the TPD training. Teachers were better able to articulate how to modify content to make it accessible to all students. In comparative analyses of the pre- and postadaptations, we found that teachers who used Text Adaptor modified features in texts that were closely associated with best practices in modifying materials for non-native English speakers, and modified the language and content of the text more comprehensively than teachers who did not have access to Text Adaptor. Positive outcomes suggesting teachers' increased knowledge and linguistic awareness around text adaptations when they used the tool has inspired additional research toward developing additional tool features to support authoring of content-area texts for ELLs.

Automated Scoring of Spoken Responses

Automated scoring of speech follows a paradigm similar to that of AES. First, language-related features are extracted, and in a second step a scoring model is used to compute a score based on a combination of these features. The main difference between text and speech is that the word identities are unknown in speech, and additional programming is needed for a speech recognizer to generate word hypotheses from the digitized candidate's speech response to an item prompt. (Another difference from an assessment perspective is that speech testing is generally done for non-native speakers.)

Ordinate, a subsidiary of Harcourt, has been developing language tests since the 1990s where basic language abilities such as reading or repeating are tested (Bernstein, 1999). This is another way of avoiding the high error rate in open-ended speech recognition for spontaneous speech. They showed correlations around 0.80 between their tests and other widely used language tests such as ETS's Test of English as a Foreign Language (TOEFL) (Bernstein *et al.*, 2000).

Cucchiari *et al.* (1997a, 1997b) developed a speech-recognition-based automatic pronunciation scoring system for Dutch by using features such as log likelihood Hidden Markov Model scores, various duration scores, and information on pauses, word stress, syllable structure,

and intonation. They also found good agreement (correlations above 0.70) between machine scores and human ratings of pronunciation.

Stanford Research Institute (SRI) International, similarly, has been developing an automatic pronunciation scoring system, EduSpeak, which measures phone accuracy, speech rate, and duration distributions for non-native speakers who read English texts (Franco *et al.*, 2000). Unlike in Ordinate's test, the texts being read need not be known to the system for prior training.

At ETS, research in automated speech scoring has been conducted since 2002. In 2006, a first speech-scoring system, SpeechRater^M, was successfully deployed to score the speaking section of TOEFL Practice Online (TPO). This is an environment that helps students prepare for the TOEFL. Unlike for the aforementioned predecessors, the goal for developing SpeechRater is to provide scoring for assessments that cover a wide range of speaking proficiency (i.e., not only pronunciation) and to elicit spontaneous and natural speech from the test candidates as opposed to mere reading or repetition (Xi *et al.*, 2006, 2008; Zechner and Bejar, 2006; Zechner *et al.*, 2009).

The tasks scored by SpeechRater are modeled on those used in the speaking section of the TOEFL iBT (Internet-based test). These tasks ask the examinee to provide information or opinions on familiar topics based on their personal experience or background knowledge, as well as to respond to read or audio stimuli related to campus life and academic situations, such as lectures. The speaking time per item is about a minute. They are scored on a scale of 1–4, with a score of zero assigned to responses which do not address the task.

The design of SpeechRater is similar to that of e-rater, which underscores both the influence which the history of work in essay scoring has had on the development of speech-scoring systems, and the fundamental similarities in the two domains. Both systems proceed by first extracting a vector of features to represent a response, and then using a machine-learning system to predict the appropriate score based on those features.

In fact, there is a preliminary step in the case of SpeechRater: the response is first processed by a speech recognizer, the output of which provides a more pliable basis for the construction of scoring features than the raw speech stream. This speech recognizer is adapted to the speech of non-native English speakers from a wide variety of first-language backgrounds, but still manages a word accuracy rate of only around 50%. While this means that, on average, every other word is recognized incorrectly, this is the best that can currently be achieved by state-of-the-art wide-coverage speech-recognition systems on data from nonnative speakers with multiple-language backgrounds and proficiency levels, under variable recording conditions. (Since TPO is web-based,

examinees may record their responses in their own homes, using their own microphones.)

This level of recognizer performance means that the features extracted by SpeechRater must not be highly dependent on recognition accuracy. By the same token, this means that SpeechRater cannot presently be expected to account for the full range of elements mentioned in the TOEFL rubric. The rubric specifies three dimensions of attributes which contribute to the score of a response:

- delivery (low-level technical features of speech production, such as pronunciation and fluency),
- language use (formal cues of linguistic competence, such as grammar and diction), and
- topic development (higher-level semantic, pragmatic, and organizational aspects of the response).

Delivery features can most easily be extracted from the state of the speech-recognition system, while language use is more difficult to address, given the constraints of recognition accuracy. Of course, the appropriate development of the topic is even more challenging to assess without an accurate transcript of the response.

Most of the features actually used in SpeechRater address the delivery aspect of the TOEFL speaking construct in one way or another. A subset of the features do, however, relate to the language-use dimension of the construct as well. The statistical model SpeechRater uses to predict the score on the basis of these features is a multiple linear regression, although promising experiments have been performed using decision trees as well.

Currently, SpeechRater is in operational use to score the TOEFL Practice Online speaking section only. It provides the examinee with a predicted score for the entire section, comprised of six speaking items, and with a range within which their score is expected to fall with a certain probability. In the future, the application's use may be expanded to other testing programs, and work is being conducted to expand the construct coverage of the model, to bring it into closer alignment with the scoring of the operational TOEFL.

Conclusion

As amazing as the invention of television was in the late 1940s, it was clear that the kinescopic pictures and mono-channel sound that reflected the technology of the times was an inadequate substitute for re-creating the real thing. However, over time, improvements were made to the broadcasting enterprise – tape for the kinescope, color, multi-channel sound, and high-definition clarity. In addition, new uses were made for television beyond entertainment (e.g., instruction, security). Do these developments make the experience any more authentic? Maybe. How often have you heard someone say, “I’d rather watch it on TV”?

In a similar vein, AES might still be characterized as an emerging technology. Though it has been demonstrated to replicate human judgments in the grading of essays, over time it will be enhanced to do so with even more proficiency and accuracy. Moreover, it has branched out to perform other roles (instruction) and is now used as a conceptual platform for other applications (language-proficiency ratings). Finally, it has engendered a discussion about what constitutes good writing and how it is best achieved.

Bibliography

- Attali, Y. and Burstein, J. (2006). Automated essay scoring with e-rater V.2. *Journal of Technology, Learning, and Assessment* 4(3). <http://www.jtla.org> (accessed August 2009).
- Bernstein, J. (1999). *PhonePass Testing: Structure and Construct*. Menlo Park, CA: Ordinate Corporation.
- Bernstein, J., DeJong, J., Pisoni, D., and Townshend, B. (2000). Two experiments in automatic scoring of spoken language proficiency. *Paper Presented at the InSTIL-2000*, Dundee, Scotland.
- Burstein, J. (2003). The E-rater scoring engine: Automated essay scoring with natural language processing. In Shermis, M. D. and Burstein, J. (eds.) *Automated Essay Scoring: A Cross-Disciplinary Perspective*, pp 113–122. Mahwah, NJ: Erlbaum.
- Burstein, J. (2009). Opportunities for natural language processing in education. In Gebulkuh, A. (ed.) *Springer Lecture Notes in Computer Science*, vol. 5449, pp 6–27. New York: Springer.
- Burstein, J., Chodorow, M., and Leacock, C. (2003). Criterion: Online essay evaluation: An application for automated evaluation of test-taker essays. *Paper Presented at the Fifteenth Annual Conference on Innovative Applications of Artificial Intelligence*, Acapulco, Mexico.
- Burstein, J., Chodorow, M., and Leacock, C. (2004). Automated essay evaluation: The Criterion Online writing service. *AI Magazine* 25(3), 27–36.
- Carlo, M. S., August, D., McLaughlin, B., et al. (2004). Closing the gap: Addressing the vocabulary needs of English language learners in bilingual and mainstream classrooms. *Reading Research Quarterly* 39(2), 188–215.
- Chai, C. O. L. (2004). *Development and Validation of Procedures to Assess Writing Plans*. Vancouver: University of British Columbia.
- Cucchiarini, C., Strik, H., and Boves, L. (1997a). Automatic evaluation of Dutch pronunciation by using speech recognition technology. *Paper Presented at the IEEE Automatic Speech Recognition and Understanding Workshop*, Santa Barbara, CA.
- Cucchiarini, C., Strik, H., and Boves, L. (1997b). Using speech recognition technology to assess foreign speakers' pronunciation of Dutch. *Paper Presented at the Third International Symposium on the Acquisition of Second Language Speech: NEW SOUNDS 97*, Klagenfurt, Austria.
- Elbow, P. (1973). *Writing without Teachers*. New York: Oxford University Press.
- Elliot, S. (2003). Intellimetric: From here to validity. In Shermis, M. D. and Burstein, J. (eds.) *Automated Essay Scoring: A Cross-Disciplinary Perspective*, pp 71–86. Mahwah, NJ: Erlbaum.
- Fitzgerald, J. (1995). English-as-a-second-language learners' cognitive reading processes: A review of research in the United States. *Review of Educational Research* 65(2), 145–190.
- Franco, H., Abrash, V., Precoda, K., et al. (2000). The SRI EduSpeak system: Recognition and pronunciation scoring for language learning. *Paper Presented at the InSTIL-2000*, Dundee, Scotland.
- Futagi, Y., Deane, P., Chodorow, M., and Tetreault, J. (2008). A computational approach to detecting collocation errors in the writing of non-native speakers of English. *Computer Assisted Language Learning* 21, 353–367.

- Goldstein, A. A. and Carr, P. G. (1996). *Can Students Benefit from Process Writing?* Washington, DC: National Center for Educational Statistics.
- Han, N. -R., Chodorow, M., and Leacock, C. (2004). Detecting errors in English article usage with a maximum entropy classifier trained on a large, diverse corpus. *Paper Presented at the 4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal.
- Hancin-Bhatt, B. and Nagy, W. E. (1994). Lexical transfer and second language morphological development. *Applied Psycholinguistics* **15** (3), 289–310.
- Higgins, D., Burstein, J., and Attali, Y. (2006). Identifying off-topic student essays without topic-specific training data. *Natural Language Engineering* **12**(2), 145–159.
- Ihnot, C. (1997). *Read Naturally*. St. Paul, MN: Read Naturally.
- James, C. and Klein, K. (1994). Foreign language learners' spelling and proof reading strategies. *Papers and Studies in Contrastive Linguistics* **19**, 31–46.
- Jimenez, R. T., Garcia, G. E., and Pearson, D. P. (1996). The reading strategies of bilingual Latina/o who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly* **31**(1), 90–112.
- Kawate-Mierzejewska, M. (2003). E-rater software. *Paper presented at the Japanese Association for Language Teaching*, Tokyo, Japan.
- Keith, T. Z. (2003). Validity and automated essay scoring systems. In Shermis, M. D. and Burstein, J. (eds.) *Automated Essay Scoring: A Cross-Disciplinary Perspective*, pp 147–168. Mahwah, NJ: Erlbaum.
- Landauer, T. K., Foltz, P. W., and Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes* **25**, 259–284.
- Landauer, T. K., Laham, D., and Foltz, P. W. (2003). Automated scoring and annotation of essays with the Intelligent Essay Assessor. In Shermis, M. D. and Burstein, J. (eds.) *Automated Essay Scoring: A Cross-Disciplinary Perspective*, pp 87–112. Mahwah, NJ: Erlbaum.
- Leacock, C. and Chodorow, M. (2003). C-rater: Scoring of short-answer questions. *Computers and the Humanities* **37**(4), 389–405.
- Lin, D. (1998). Automatic retrieval and clustering of similar words. *Paper Presented at the 35th Annual Meeting of the Association for Computational Linguistics*, Montreal, Canada.
- MacDonald, N. H., Frase, L., Gingrich, P., and Keenan, S. A. (1982). The writer's workbench: Computer aids for text analysis. *IEEE Transactions on Communications* **30**(1), 105–110.
- Marcu, D. (2000). *The Theory and Practice of Discourse Parsing and Summarization*. Cambridge, MA: MIT Press.
- Nagy, W. E., Garcia, G. E., Durgunoglu, A. Y., and Hancin-Bhatt, B. (1993). Spanish–English bilingual students' use of cognates in English reading. *Journal of Reading Behavior* **25**(3), 241–259.
- Oh, S. Y. (2001). Two types of input modification and EFL reading comprehension: Simplification versus elaboration. *TESOL Quarterly* **35**(1), 69–96.
- Page, E. B. (1966). The imminence of grading essays by computer. *Phi Delta Kappan* **48**, 238–243.
- Page, E. B., Keith, T., and Lavoie, M. J. (1995). Construct validity in the computer grading of essays. *Paper Presented at the Annual Meeting of the American Psychological Association*, New York, NY.
- Page, E. B. and Petersen, N. S. (1995). The computer moves into essay grading: Updating the ancient test. *Phi Delta Kappan* **76**(7), 561–565.
- Pantel, P. and Lin, D. (2000). Word-for-word glossing with contextually similar words. *Paper Presented at the 1st Annual Meeting of the North American Chapter of Association for Computational Linguistics (NAACL2000)*, Seattle, WA.
- Perez, E. (1981). Oral language competence improves reading skills of Mexican-American third graders. *Reading Teacher* **35**(1), 24–27.
- Shei, C. and Pain, H. (2000). An ESL writer's collocational aid. *Computer Assisted Language Learning* **13**(2), 167–182.
- Shermis, M. D. and Burstein, J. (2003). Introduction. In Shermis, M. D. and Burstein, J. (eds.) *Automated Essay Scoring: A Cross-Disciplinary Perspective*, pp xiii–xvi. Mahwah, NJ: Erlbaum.
- Shermis, M. D., Mzumara, H. R., Olson, J., and Harrington, S. (2001). On-line grading of student essays: PEG goes on the web at IUPUI. *Assessment and Evaluation in Higher Education* **26**(3), 247–259.
- Shore, J., Burstein, J., and Sabatini, J. (2009). Web-based technology that supports ELL reading instruction. *Paper Presented at the Interactive Symposium for the Special Interest Group, Computers and Internet for Educational Applications, held in conjunction with the American Educational Research Association Annual Meeting*, San Diego. <http://ciaesig.ning.com/> (accessed August 2009).
- Vantage Learning (2001). *A Preliminary Study of the Efficacy of IntelliMetric™ for Use in Scoring Hebrew Assessments*. Newtown, PA: Vantage Learning.
- Vantage Learning (2002). *A Study of Intellimetric™ for Responses in Scoring Bahasa Malay*. Newtown, PA: Vantage Learning.
- Xi, X., Higgins, D., Zechner, K., and Williamson, D. M. (2008). Automated Scoring of Spontaneous Speech Using SpeechRaterSM v1.0. Educational Testing Service. *Research Report RR-08-62*, November.
- Xi, X., Zechner, K., and Bejar, I. (2006). Extracting meaningful speech features to support diagnostic feedback: An ECD approach to automated scoring. *Paper Presented at the National Council on Measurement in Education*, San Francisco, CA.
- Yano, Y., Long, M., and Ross, S. (1994). The effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning* **44**, 189–219.
- Zechner, K. and Bejar, I. (2006). Towards automatic scoring of non-native spontaneous speech. *Proceedings of the Human Language Technology Conference of the NAACL*, New York, NY.
- Zechner, K., Higgins, D., Xi, X., and Williamson, D. M. (2009). Automatic scoring of non-native spontaneous speech in tests of spoken English. *Speech Communication* **51**(10), 883–895.

Further Reading

- Ajay, H. B., Tillett, P. I., and Page, E. B. (1973). *Analysis of Essays by Computer (AEC-II)* (No. 8-0102). Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education, National Center for Educational Research and Development.
- Ericsson, P. F. and Haswell, R. (eds.) (2006). *Machine Scoring of Student Essays: Truth and Consequences*. Logan, UT: Utah State University Press.

Relevant WebSites

- <http://www.ets.org/criterion> – Educational Testing Service: Criterion.
- <http://www.languageweaver.com> – Language Weaver.
- <http://www.pearsonkt.com> – Pearson's Knowledge Technologies.
- <http://www.vantagelearning.com> – Vantage Learning.

Classical Test Theory Reliability

M D Miller, University of Florida, Gainesville, FL, USA

© 2010 Elsevier Ltd. All rights reserved.

Reliability refers to the degree to which test scores are free from errors of measurement. It is not a global property of an assessment. Instead, different types of errors of measurement might be considered for different test uses. For example, a test might provide reliable scoring without being stable over time. Reliability is also a matter of degree. Tests are not generally perfectly reliable or completely unreliable. Instead, a test is reliable to some degree.

Estimates of reliability quantify the amount of measurement error for a particular test use or interpretation for a specified population. Although reliability can be defined broadly in terms of consistency or generalizability, specific statistical indices of reliability will vary depending on the statistical model and the sources of error. The statistical model may be based on classical test theory, generalizability theory, or item response theory. For classical test theory, there are multiple indices of reliability based on different sources of measurement error, including item heterogeneity, equivalence of test forms, stability over time, or consistency of subjective ratings. Different sources of error would be of concern in different contexts. For example, the test score from a student writing an essay will be affected by errors in scoring, while the test score from a student taking a multiple-choice test would be affected by the heterogeneity of the items selected to measure the construct. In addition, a test score can be affected by multiple sources of error simultaneously. A student taking an admissions test would be affected by the heterogeneity of the items, the form of the test, and the subjectivity of the scoring for any open-ended written portion of the test. Thus, there are many types of reliability that vary depending on the sources of error being considered. These varying definitions should be selected based on the particular test use or score interpretation being made and one type of reliability should not be considered interchangeable for another.

Classical Test Theory

Classical test theory is defined such that any observed test score, X , is the sum of a true score, T , and a random error, E . That is,

$$X = T + E$$

This model assumes that the expected value of the random errors are zero and that the random errors are uncorrelated

with the true score, with errors on alternate forms of the test, and with other measures such as another test or grades in school. Validity focuses on the definition (i.e., use and interpretation) of the true score, while reliability focuses on the random error. The error is the sum of all random effects, while the true score is the sum of all consistent effects. Thus, the error is undifferentiated with respect to different sources of randomness unlike in generalizability theory. Broadly, two indices of reliability are commonly reported: the reliability coefficient and the standard error of measurement (SEM).

The reliability coefficient ($\rho_{XX'}$) is defined as the ratio of the true score variance (σ_T^2) to the observed score variance (σ_X^2), or the ratio of the true score variance to the sum of the true score variance and the error variance (σ_E^2). That is,

$$\rho_{XX'} = \sigma_T^2 / \sigma_X^2 = \sigma_T^2 / (\sigma_T^2 + \sigma_E^2)$$

Hence, the value of the reliability coefficient is the proportion of variation in test scores that can be attributed to consistent measurement (i.e., the true score). The reliability coefficient ranges from 0.0 to 1.0 with higher values being preferred. At $\rho_{XX'} = 0.0$, there is no consistency in the measurement procedure and the observed score is equal to the random error ($X = E$). At $\rho_{XX'} = 1.0$, the observed score has no error and the observed score is equal to the true score ($X = T$). In practice, the reliability coefficient will be somewhere between the two extremes.

The SEM is defined as the standard deviation of the errors of measurement, σ_E . The SEM ranges from 0.0 to the standard deviation of the observed scores, σ_X . When the SEM = σ_X , there is no consistency in the measurement procedures, the reliability coefficient is equal to 0.0, and the observed score is equal to the random error. When the SEM = 0.0, there is perfect consistency in the test scores, the observed score is equal to the true score, and the reliability coefficient is equal to 1.0. In practice, the SEM will fall somewhere between the two extremes.

The reliability coefficient is an easily interpreted index of the consistency of the test scores since it is in a standard range for all tests. While the SEM is more difficult to interpret initially, it is in the metric of the test scores which allows for the interpretation of the individual test scores via confidence intervals. Another advantage of the SEM is that it is not based on the true scores and, consequently, its estimation is not influenced by sampling errors. The reliability coefficient will be underestimated when

the sample range of scores is restricted, whereas the SEM will be largely uninfluenced by sampling fluctuations.

Types of Reliability

Within the framework of classical test theory there are several types of reliability coefficients based on the source of the random errors. The types of reliability discussed below are test–retest, alternate form, alternate form test–retest, interrater, split half, and internal consistency.

Test–retest reliability is used to examine the stability of the trait being measured over time. The reliability coefficient is the correlation between test scores for a sample taking the same test on two occasions. Generally, test–retest reliability is higher when the time span between test administrations is shorter. However, the test–retest reliability should be estimated with a time interval that mirrors the actual use of the test rather than trying to maximize the value of the coefficient.

Alternate form reliability is used to measure the equivalence of test scores across two (parallel) forms of a test. The reliability coefficient is the correlation between test scores on the two forms of the test taken by the same sample without a substantial time lag. Usually, half of the sample receives one form first (e.g., form A), while the other half of the sample receives the other form first (e.g., form B) so that there is no order effect. Then, examinees take the form they had not taken yet. Alternate form reliability is higher when care is taken to make sure that the two forms are equivalent in content and statistical properties (i.e., mean, standard deviation, and distribution shape).

Alternate form test–retest reliability follows the same procedure as with the alternate form reliability except that there is a time lag between test administrations. In this case, the errors of measurement include stability over time and equivalence of the forms. In general, this type of reliability will be lower than alternate form or test–retest reliability which targets only one type of random error.

Interrater reliability is used to measure the consistency of ratings from subjective scoring. The reliability coefficient is the correlation between the ratings from two raters on the same sample of writings/essays/performances. Interrater reliability is higher when standardized procedures are used by the raters to score the writings. At a minimum, the standardized procedures should include training of the raters and clearly defined rubrics. Large-scale assessments further standardize the procedures to include the use of benchmarks, monitoring the process, intervening when ratings are discrepant, and other procedures to check the rating process.

Split-half reliability is used to measure the consistency within a single administration of a test by examining

the relationship between two halves of the same test. The procedure for split-half reliability is to administer a single form of the test to a sample. The reliability estimate is then based on the correlation between two halves of the test adjusted for test length. That is, the test is divided into two equivalent halves based on test content and item statistics (often this can be accomplished by using odd- and even-numbered items to form the halves) and the halves are correlated. However, the reliability will be less for half of a test than it is for the full-length test. Consequently, the correlation between the halves is adjusted upward using the Spearman–Brown prophecy formula. Split-half reliability will be higher when the equivalence of the two forms is higher in terms of content and item statistics. However, the matching of the two halves should not be completed on the basis of the sample statistics since random sampling fluctuations could inflate the value of the reliability. Instead, careful matching should be completed based on content and item statistics from a prior data collection. The Spearman–Brown prophecy formula is:

$$\rho_{XX'} = 2\rho_{HH'}/(1 + \rho_{HH'})$$

where $\rho_{XX'}$ is the full-length reliability and $\rho_{HH'}$ the half-length correlation.

Internal consistency is used to measure the consistency of items within a single test form. The procedure for internal consistency is to administer a single form of the test to a sample and estimate the internal consistency using item and test statistics with an internal consistency formula. The formula for internal consistency has many equivalent forms in the literature, including the Kuder–Richardson 20 (KR20) formula for dichotomously scored items and Cronbach's alpha. Internal consistency is also easy to compute with most standard statistical software (e.g., SPSS or SAS). Internal consistency is higher when the items are more homogeneous. The formula for Cronbach's alpha coefficient is

$$\alpha = \frac{n}{n-1} \left(\sigma_X^2 - \sum_{i=1}^n \sigma_i^2 \right) / \sigma_X^2$$

where n is the number of items, σ_X^2 is the total test score variance, and σ_i^2 is the item variance.

Table 1 summarizes the reliability coefficients and their major sources of error that are reported in classical test theory.

Each of the reliability coefficients shown in **Table 1** differs in its data collection procedure, computation, and the major source of error. The appropriate reliability coefficient should match the intended use or interpretation of the test. For example, when subjective measurements are part of the assessment procedure, interrater reliability is needed. When multiple items are being used (which should be the case), internal consistency or

Table 1 The reliability coefficients and their major sources of error

<i>Reliability type</i>	<i>Source of error</i>
Test–retest	Stability over time
Alternate form	Equivalence across forms
Alternate form test–retest	Stability over time and equivalence across forms
Interrater	Consistency of ratings
Split half	Equivalence across halves
Internal consistency	Equivalence and item homogeneity

split-half reliability should be used. In short, the reliability estimate(s) should include all sources of error that will be part of the test use or interpretation. One type of reliability should not substitute for another.

Standard Error of Measurement

The reliability coefficient is used to quantify the precision of an assessment for a particular use or interpretation. The index is simple to interpret since it is always based on the same scale (0.0–1.0). However, the reliability coefficient fails to show the amount of error that might be expected in an individual's test score. The SEM is the standard deviation of the errors of measurement and can be used to create confidence intervals for examinee scores. Assuming a normal distribution, 68% of the observed scores will be within one SEM of their true score; and 95% if the observed scores will be within 1.96 SEMs of their true score. For example, if $SEM = 2.00$ and an examinee's true score was 25, upon repeated measurements, 68% of the scores for that examinee would be between 23 and 27. Note that the confidence interval is around the true score and not the observed score, which leads to the interpretation that 68% of the time that a confidence interval based on one SEM is constructed, it will contain the true score.

As with the reliability coefficient, an SEM can be created for different types of measurement error. In fact, the SEM is calculated using the appropriate reliability coefficient so that the appropriate source of error is being used. Thus, **Table 1** can be used for the SEM or the reliability coefficient so that the SEM can be created with each of the different sources of error.

Magnitude of Reliability

The literature does not provide definitive guidance on acceptable levels of reliability. However, what constitutes an acceptable level of reliability is determined by the use of the test. Uses of the test with higher stakes require higher levels of reliability. For example, reliability for a test being used in theoretical research may not require

the same level of consistency as would be required for high stakes uses of tests such as high-school graduation, certification, or licensure.

How to Increase Reliability

It is important to be able to increase reliability when developing instruments. In general, there are two ways that should always be considered in increasing reliability: greater standardization and increasing the number of items. Test administration and test development procedures should be standardized so that no random errors are introduced. The effect of the standardization will not only globally affect each type of reliability but will also have specific effects on certain types of reliability. Standardization includes methods to create equivalent forms of a test (alternate form), and methods to create homogeneous pools of items (internal consistency), or equivalent halves of tests (split half), as well as it includes methods to create consistency in scoring through the development of rubrics and standardized scoring procedures (interrater).

Another key element to increasing reliability is increasing the length of the test. The Spearman–Brown formula is based on the principle that longer tests are more reliable. Assuming that the conditions of testing do not change with increased length (i.e., fatigue, boredom, or item quality), increasing the number of items always leads to more reliable tests. Thus, short forms of a test or subscores are generally less reliable than the full form.

Reliability and Aggregation

Increasing the number of items will increase the reliability of test scores due to averaging over more data. Similarly, increasing the number of examinees and averaging across their scores will reduce the SEM and increase reliability. That is, the reliability of the mean will generally be higher than the reliability of an examinee. This applies to estimates for the full sample as well as to aggregates such as classrooms, schools, or states. Whether averaging across items or examinees, the estimate becomes more stable. Indeed, the SEM will almost always be lower for group means than it is for individuals. (Note: under some conditions the reliability coefficient could be lower for the group means when the true score variance in the groups is restricted in range. However, even under these conditions, the SEM will typically be lower and the group means will be more stable.)

Reliability and Growth Scores

The reliability of growth or difference scores, defined as posttest minus pretest, has received considerable attention

in the literature. Some have argued that the growth score is unreliable and that growth is negatively correlated with the pretest. That is, growth will be higher for examinees with low pretests. However, other researchers have pointed out that low reliability for the difference scores does not necessarily result in less power for comparisons among groups and the difference score may be the construct of interest. At any rate, caution should be used in interpreting growth scores at the examinee level.

Relationship of Reliability and Validity

Classical test theory assumes that an examinee's test score is the composite of a true score and random error. Validity addresses the true score by examining its uses and interpretations. Thus, any systematic error or bias is part of the true score, while only random errors are addressed in the reliability analysis. However, reliability is a necessary, but not sufficient, condition for validity. This means that there needs to be a true score (with some level of reliability) to examine validity; however, that the existence of a true score does not guarantee the validity of it. That is, the true score may still be a biased estimate of the construct of interest as a result of systematic errors. In addition to this relationship of validity and reliability, there may be a tension between the two psychometric properties of the test. Higher reliability can be attained by standardizing

the testing procedure which has the potential to reduce the breadth of the construct being measured and, thus, to decrease the validity. For example, higher internal consistency is attained by increasing item homogeneity. To the extent that the construct requires heterogeneity of the items, this will create a tension between reliability and validity. As a second example, interrater reliability is increased by standardizing the scoring procedure (e.g., clearly defined rubrics and training). This standardization can limit the definition of good writing by not including some types of writing in the rubric and, as a consequence, limit the breadth of the construct. As a result, it is important to consider the impact of any standardization on the validity of the test as well as the reliability so that the construct is still clearly being measured.

See also: Generalizability Theory; Item Response Theory.

Further Reading

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Haertel, E. H. (2006). Reliability. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 65–110. Westport, CT: Praeger.
- Lord, F. M. and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

Criterion-Referenced Measurement

S A Shrock and W C Coscarelli, Southern Illinois University, Carbondale, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Conjectural technique – A method of setting the passing score on a criterion-referenced or mastery test that relies on expert estimation of item or assessment task performance at the mastery borderline.

Contrasting groups – A method of setting the passing score on a criterion-referenced or mastery test by examining the performances of test-takers of known competency level.

Criterion-referenced measurement – The measurement of human performance in relation to a well-defined, accepted standard.

Domain-referenced measurement – The measurement of human performance by random sampling from a well-defined population of items or assessment tasks such that the percent correct on the measurement tool implies capability to perform the same percent of the population of items/tasks.

Domain specification – A set of descriptive criteria used to create the items or assessment tasks for a criterion-referenced or domain-referenced measurement tool.

Informed judgment – A method of setting the passing score on a criterion-referenced or mastery test by considering the preferred cutoff scores of all major stakeholders in the assessment.

Mastery testing – The measurement of human performance of specific competencies for the purpose of classifying the performer as a master or nonmaster of the competencies assessed.

Norm-referenced measurement – The measurement of human performance in relation to the performances of others.

Introduction

Criterion-referenced measurement typically refers to measurement of human performance in relation to an accepted, described standard. Performance, as used here, is evidence of a human competency – a demonstration of a skill or mastery of a knowledge base. Criterion-referenced measurement is frequently defined in contrast with norm-referenced measurement – the measurement of human performance in relation to the performances of

others. The two forms of measurement can be thought of usefully as two different ways to interpret test results; however, it is the role of test construction to produce results that allow the alternative interpretations. In other words, one cannot typically discriminate a criterion-referenced test from a norm-referenced one by simply examining the questions. However, the norm-referenced test will have been created to separate the test-takers' performances from one another, that is, to create test score variance, while the criterion-referenced test will have been composed of questions carefully matched to stated objectives or competencies. Norm-referenced measurements are ideal for making comparative, selection decisions among test-takers. Criterion-referenced measurement is useful for determining whether or not individual test-takers have specific competencies.

Criterion-referenced measurement is relatively new compared to norm-referenced measurement. Classical test theory that posits the assumptions and principles that undergird norm-referencing and its statistical indicators of reliability, validity, and error was developed and refined over the past century. Criterion-referencing was conceptualized as an alternative in the 1960s; its initial theory, principles and implementation procedures developed rapidly during the 1970s and 1980s, slowing considerably during the more recent past. On the other hand, the term is used more frequently now in the world of applied measurement than ever before, in school and university contexts as well as in those of government and business. To what extent current usage reflects adherence to the original tenets of the paradigm is a current issue in the field.

Distinctions Within Criterion-Referenced Measurement

The term criterion-referenced seems to be used interchangeably with domain-referenced, competency-based, objectives-based, and mastery testing, in education literature, if not among psychometric professionals. Actually, the terms imply important differences in test construction, as well as in test score interpretation.

Items included on a domain-referenced test are selected through a random or stratified random sampling from a pool of items generated to match a content domain specification. Because the items in the pool constitute essentially a population of tasks, the test-taker's percentage score on the domain-referenced test is taken to represent the percentage

of the total pool of tasks that the test-taker would perform correctly. The technology of domain-referenced testing was pursued aggressively during the 1970s. It presented some major challenges, most notably the need for tightly defined content domains that would ensure the generation of large numbers of comparable test items whose membership in the same population could be defended. This domain specification proved more feasible for some content areas (math and grammar) and cognitive levels (knowledge, comprehension, and application) than others (history, and science at the analysis level); the more complex the required cognition and the more divergent the universe of correct performances became, the harder the underlying premises of the domain-referenced model were to meet.

The terms objectives-based measurement and competency-based measurement suggest that test items are based only on descriptions of desired behaviors, rather than a comprehensive, precise content domain specification. These terms are largely nonspecific regarding test score interpretation.

Mastery testing introduces an alternative means of test score interpretation, that is, a dichotomous decision – master or nonmaster. In doing so, mastery test results open the analysis of test quality – especially reliability and validity indicators – to modifications of the statistical procedures used with norm-referenced tests. The major modification is reliance on the consistency and accuracy of mastery classifications, instead of score correlations, to calculate reliability and validity coefficients.

Criterion-referenced measurement could refer to any of the assessments described above. In this respect it has become an imprecise term and one is well advised to seek clarification of any usage with serious implications or consequences. The term now most often refers to mastery testing, although certainly not to the entire mastery learning model, and only rarely to domain-referenced measurement as it was originally proposed.

Criterion-Referenced Test Development

It is not true that criterion-referenced assessments are limited to tests composed of multiple-choice or other closed-ended questions, although these are the most common referents. Even though the procedures described below are phrased in terms of question-based test construction, a criterion-referenced measurement instrument could take the form of a performance test or other so-called authentic assessment composed of what might be more descriptively called assessment tasks.

Domain Specification

A criterion-referenced test begins with a thorough analysis of the competencies that the test must measure.

The result is a definitive set of behavioral descriptions that focus item creation, content validation, and score interpretation – a content domain specification. Failure at this stage of the process will doom the validity and, therefore, the usefulness of the test that results regardless of the efforts devoted to subsequent steps.

Popham (1978) provides guidelines for constructing content domain specifications. Descriptions of incorrect performance or responses are included in the elaborated objectives that constitute the domain specification.

Item/Assessment Task Writing

Using the guidance provided by the domain specifications and typically the assistance of a test design specialist, item writers draft a pool of items whose correct completion requires the criterion behaviors the test must measure. Instruction in the technical aspects of sound item writing at this draft stage pays great dividends in reduced revision of the items later. Novice item writers can be expected to have greater difficulty writing items that require cognition above the knowledge or memory level. The item-writing activity typically raises questions about the domain specifications; the two activities are actually reciprocal in practice.

The question of how many items should be created is not a simple one. More items should be written than will appear on the test – some would recommend creating twice that number – because during content validation and pilot testing, many flawed items will be discarded. The number also depends upon whether and how many parallel forms of the test are planned; because parallel forms of a test measure the same competencies using different items, additional items are required. Similarly, large numbers of items are needed if the goal is to build a bank from which items will be randomly chosen to form multiple parallel forms of the test. The length of the finished exam can be a function of many factors, perhaps the most important of which is the criticality of the decisions to be made based on the test results; test reliability will be a function partly of test length. Other factors include the breadth of the domain competencies assessed by the test, the number of different domains covered, and pragmatic yet critical factors such as available testing time and budget size.

Content Validation

The content validity of a criterion-referenced test is established by having content experts examine items and verify that they match (or elicit) the criterion behaviors described in the domain specification. If a criterion-referenced test cannot be said to have content validity, it probably does not matter what other kind of validity evidence might be presented in its defense. The resolution of litigation

surrounding these tests has tended to hinge on the content validity question.

At this stage in the test development process, the items should be examined for technical flaws such as cueing, improper reading level, and confusing language or visual images. In addition, the items should be screened for bias against any segment of the test-taker population, including stereotyping that could heighten test anxiety or other negative attitude in test-takers. If content experts are not skilled in detecting these flaws, then measurement specialists should be sought to provide this review. Flawed items must be revised or discarded. Substantial revision of items may mean that the content validation process must be repeated.

Item Analysis and Selection

At this point the test(s) are given to groups representative of the test-takers for whom the test was designed. The major goal of this pilot is the collection of item analysis data. Item statistics serve two major purposes: to detect flawed items for revision or elimination from the item pool and to provide a basis for matching items on difficulty and discrimination power to create parallel forms of the test. Suspected flawed items are those that appear substantially easier or more difficult than other items from the same domain specification, or any item with a negative discrimination coefficient. Numbers of test-takers included in the pilot sample should be larger as the stakes of the implemented test and the diversity of the intended test-takers increase. If the pilot sample is adequate, possible adverse impact of the test on subgroups of test-takers could be detected at this point.

It should be noted here that advocates of domain-referenced testing have argued that item analysis data should not play a role in the selection of items for domain-referenced tests (Millman, 1974). Domain-referenced testing theory posits that all items created to reflect the content domain are eligible for inclusion on the test; selection of the items is necessarily random, and this random selection guarantees the parallelism of all test versions and their fidelity to the criterion behaviors the test is designed to measure.

Standard Setting

After test revision is complete and the test assembled, the fundamental issue of standard setting – establishing the score that separates mastery from nonmastery – must be resolved. (Domain-referenced tests do not require a standard or cutoff score. The score interpretation is the percent of items correct.) Historically, psychometricians have relied on three classes of techniques: informed judgment, conjectural techniques, and contrasting groups. The informed judgment technique relies on a systematic review of the test by stakeholders and their consensus

decision on the standard for mastery. The conjectural approaches use subject-matter experts for review of test items and development of the cut-score. Contrasting groups approaches rely on gathering data of known masters and nonmasters and determining the cut-score based on the distributions of the two groups and the subjective determination of the risk associated with false positives or false negatives in selecting the standard. (An alternative standard setting process is still in use, although it provides no logical basis for mastery determination: “Standards by directive or authority” (Browning *et al.*, 2006) are established by legislative or managerial fiat.)

Current practice in standard setting has seen the ascendancy of conjectural techniques. Perhaps because the informed judgment process is too political and compromises test security, and contrasting groups procedures demand large sample sizes and are unwieldy for rapidly changing content areas, the conjectural approaches represent the mini-max strategy between feasibility and valid judgment. The Angoff technique was one of the first approaches proposed for standard setting; its applicability and ease of use with both closed-ended and open-ended testing approaches likely contribute to its continued popularity. A growing number of conjectural approaches are also being developed and implemented primarily for school accountability assessment, for example, Jaeger-Mills, Bookmark, and Body-of-Work procedures.

In the end, there is no technique in the psychometrician’s armamentarium that will provide a singularly statistical calculation of the correct standard; there is no substitute for professional judgment on the part of the standard setter.

Test Reliability

Establishing reliability for criterion-referenced tests requires a different set of assumptions than those for norm-referenced tests. As norm-referenced tests maximize variance to spread test-takers along a continuum of scores, criterion-referenced tests often result in a lack of variance in scores; for example, when all test-takers master the performance measured by the test. Further, since criterion-referenced tests often measure a series of discrete and independent performances, the assumption of a single unifying ability or construct is not met. Therefore, classic approaches to reliability must be reinterpreted in light of these alternative assumptions.

There are three general approaches to criterion-referenced reliability: decision consistency, single-test administration, and interrater reliability (Shrock and Coscarelli, 2007).

Decision consistency

In this approach, the test-taker is given two tests over a short period of time and the consistency of classification

of mastery is measured. This design can be either a test-retest model using the same form in both administrations or an equivalence model where two parallel forms are used to assess the same competencies. Reliability can be determined using the agreement coefficient (p_o), Cohen's kappa (κ), or phi (ϕ). The p_o is the easiest to conceptualize but provides an estimate inflated by chance; κ and ϕ will often provide similar indices, although κ is not a correlation coefficient and provides different information – the realized proportion of possible improvement in classification above chance level. A reliability estimate of 0.85 or higher is desirable.

Single-test administration

Mainly because of logistical, political, and economic issues, test designers must often accept a reliability estimate based on one test administration. Squared-error loss approaches use the squared distance between a person's score and the cutoff score to establish reliability. Livingston's coefficient kappa (κ^2) which draws on common norm-referenced internal consistency measures is one approach; the Brennan and Kane indexes of dependability $\Phi(\lambda)$ and Φ also provide similar information, although they seem to offer increased precision through a correction for chance. Subkoviak's Z also provides p_o and κ , although it requires support of an adjunct set of tables. Brennan's BB-CLASS software system provides a sophisticated assessment of decision consistency and accuracy. Generalizability (G) theory based on analysis of variance (ANOVA) models and item response theory (IRT) approaches also exist that provide reliability information (Cizek and Bunch, 2007). Livingston's κ^2 is probably more widely used due to its ease of computation and its use of norm-referenced statistics.

Interrater reliability

When mastery judgments of a performance are being made by observers, their reliability must also be determined. Simple percentages of agreement between or among judges should not be viewed as adequate measures of reliability. Both ϕ and κ , as used in the decision consistency approach, can be used to determine consistency among raters of a performance.

Concurrent Validation

Concurrent validation of a criterion-referenced test demonstrates the ability of the test to correctly classify masters and nonmasters. Only one administration of the test is needed for each test-taker and the estimate of validation is typically expressed through a ϕ correlation coefficient.

The concurrent validation process begins with the identification of known masters and nonmasters – a sample of test-takers who are representative of those between

whom the test is designed to distinguish. The potential range of errors includes:

1. introducing nonmasters who are far less competent than those nonmasters who will take the test,
2. choosing superior masters whose skills reside far above the cut-score,
3. choosing a sample of convenience rather than a representative sample,
4. choosing a sample that is too small for meaningful statistical purposes (fewer than 60), and
5. selecting judges who are not competent to make the selection of masters and nonmasters.

Once the test-takers are identified, the test is administered to each person. (Testing need not be done in a single administration time and location for all test-takers.) Phi is then calculated between the two dichotomous variables found in this design: the known status as determined by the judges (master/nonmaster) and the test's classification of each person (master/nonmaster).

Reporting Test Results

The most basic interpretation of criterion-referenced test results is a single decision: master/nonmaster. Currently, there is distinct pressure from test-takers and policy-makers alike to provide test results that reflect degrees of competence. Some seek to provide three outcomes, for example, basic, proficient, and advanced, while others advocate as many as five classifications. These categories are often developed not from a content analysis perspective but from statistical adjustments based, essentially, on the concept of standard error of measurement. Sophisticated approaches to create these psychometrically based classifications often rely on IRT to estimate standard error and provide a basis for creating categories.

At the moment, the confluence of political and personal stakeholders in test reporting, combined with the developing but unsettled issues in standard setting (Cizek and Bunch, 2007) should leave the test designer wary of making refined judgments. Organizations with access to large numbers of test-takers, stable content, and significant computing resources may be able to demonstrate relatively accurate classifications of competence from a single test. Those without such resources would be advised to adhere to the most basic form of criterion-referenced test development.

Application of Criterion-Referenced Measurement

Criterion-referenced tests are currently very widely used to test both adults and children of all ages. The early adopters of criterion-referenced tests seemed to be the

military and professional credentialing bodies. This application is perhaps not surprising for three reasons:

1. the required criterion behavior descriptions were more easily generated in the world of work than in the world of abstract school subjects;
2. the military and the professions had greater urgency to identify incompetent practitioners; and
3. resources to devote to the labor-intensive process of creating criterion-referenced tests were more available in defense departments and through a fee structure to professional credentialing groups than in school budgets.

More recently, businesses, especially large corporations, have sought to create criterion-referenced tests for numerous reasons. In this environment, the term certification testing is often used to describe criterion-referenced determination of masters and nonmasters.

On the other hand, major school accountability legislation passed by the United States federal government has demanded that individual states assess the academic progress of virtually every child in every school. This legislation has created an enormous demand for so-called criterion-referenced tests, aggravated by allowing each of the 50 states to develop unique tests. Whether the tests created in response are truly criterion-referenced is less clear.

Emerging Issues

Future issues in criterion-referenced measurement may well be: the applicability of rapidly developing test creation computer software; the influence of America's huge investment in school accountability testing on accepted practice and psychometric research; and the possible

impact of expanded corporate and school testing on test litigation and legislation. More research-oriented issues include application of IRT and computerized adaptive testing to criterion-referenced measurement; a consensus on selection of cut-score techniques; and Monte Carlo simulations to inform choice of an approach to reliability estimation.

See also: Impact of Assessment on Classroom Practice; Norm-Referenced Measurement; Scoring Rubrics; Student Test Results in School Accountability; Test Development.

Bibliography

- Browning, A. H., Bugbee, Jr., A. C., and Mullins, M. A. (eds.) (2006). *Certification: A NOCA Handbook*. [CD ROM]. Washington, DC: National Organization for Competency Assurance.
- Cizek, G. J. and Bunch, M. B. (2007). *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*. Thousand Oaks, CA: Sage.
- Millman, J. (1974). Criterion-referenced measurement. In Popham, W. J. (ed.) *Evaluation in Education*, pp 311–397. Berkeley, CA: McCutchan.
- Popham W. J. (1978). *Criterion-Referenced Measurement*. Englewood Cliffs, NJ: Prentice-Hall.
- Shrock, S. A. and Coscarelli, W. C. (2007). *Criterion-Referenced Test Development*, 3rd edn. San Francisco, CA: Pfeiffer.

Further Reading

- Downing, S. M. and Haladyna, T. M. (eds.) (2006) *Handbook of Test Development*. Mahwah, NJ: Earlbaum.
- Hambleton, R. K. (1999). Criterion-referenced testing principles, technical advances, and evaluation guidelines. In Gutkin, T. and Reynolds, C. (eds.) *The Handbook of School Psychology*, 3rd edn., pp 409–433. New York: Wiley.

Differential Item Functioning

M D Hidalgo, University of Murcia, Murcia, Spain

J Gómez-Benito, University of Barcelona, Barcelona, Spain

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bias – A test or item favors one group to the detriment of another, in other words, there is a disparity in its behavior according to one or more group variables (e.g., race, sex, socio-economic level, cultural background, and linguistic ability) which are, however, not relevant to the trait or variable that the test actually aims to measure.

Dichotomous versus polytomous – A dichotomous item is a categorical variable with two possible response values (Yes/No, Agree/Disagree, Success/Fail). A polytomous item is a categorical variable ordinal or nominal with more than two possible values (e.g. strongly disagree, disagree, agree, strongly agree). Most of the proposed methods have focused on detecting DIF in dichotomous items, more recently, and due to the development of new evaluation instruments in the educational context (portfolio-based assessment, authentic assessment, etc.) techniques for detecting DIF have also been adapted to polytomous items.

Differential item functioning (DIF)/differential test functioning (DTF) – Different probabilities of successfully responding to a given item/test for subjects with the same level of the measured characteristic, depending on the group to which they belong.

DIF amplification vs DIF cancellation – DIF amplification is produced when a subtest of items only shows DIF in one of the groups (e.g., focal group). DIF cancellation occurs when one subset of test items shows DIF in one of the compared groups, while another subset of items cancels out the effect of the second.

Focal and reference group – In DIF studies there are normally two groups, known as focal (in which DIF is assumed to exist) and reference (the majority group). However, some DIF techniques allow multiple focal groups.

Impact – Differences observed in item performance for two different population groups that are real differences in the underlying ability of interest measured by the test.

Non-uniform DIF – The difference between the two groups in terms of the probability of responding

correctly to the item is not the same at all levels of the trait, in other words, there is an interaction effect.

Matching criteria – For the groups to be comparable, they have to be equal with regard to a certain criterion, generally the total test score or an estimate of ability. The fact that the performance on the item is conditional on the ability level is fundamental to distinguish between differences in the functioning of the item and differences in the ability level of the groups.

Purification procedures – The DIF is re-evaluated in a second step (two-stage procedures) or by successive steps (iterative procedures), using as matching criteria only those items which do not present DIF in the previous analysis, the aim of which is to purge the items with DIF so that they do not form part of the matching criterion.

Uniform DIF – The probability of responding correctly to an item is greater for one group than another across all levels of the trait.

Introduction

Standardized tests are nowadays frequently used in the field of education to measure aptitudes, skills, interests, attitudes, and performance. The results they provide serve to guide decision making in areas such as personnel selection, promotion, and the evaluation of ability, and they also offer a framework for conducting cross-cultural studies. Given their extensive use in both educational decision-making processes and transcultural comparisons, the possibility of bias in test items has become a key concern when it comes to evaluating test validity, as is clear from the latest edition of the *Standards for Educational and Psychological Testing* (American Psychological Association *et al.*, 1999).

The existence of bias in educational measurement instruments can undermine the validity of a test, as some of its items may be benefiting certain population groups at the expense of others who are at the same level of the measured characteristic. Thus, conducting evaluations that seek to ensure the equity and validity of the interpretations and decisions made on the basis of a test has become a key feature of good professional practice in the field of

educational measurement. Indeed, it is essential to analyze the degree to which the measurement instruments used are free of bias and able to provide valid and fair assessments for groups of people from different demographic, social, cultural, and linguistic backgrounds. In this context, studies of differential item functioning (DIF) have drawn the attention of researchers and professionals at both the methodological and applied levels.

DIF, Bias, and Impact

Bias in a test or item means that the test or item in question favors one group to the detriment of another; in other words, there is a disparity in its behavior according to one or more group variables (e.g., race, sex, socioeconomic level, cultural background, and linguistic ability) which are, however, not relevant to the trait or variable that the test actually aims to measure. However, the term bias assumes that the reason why a given item behaves differently according to certain variables is known or under investigation, when in fact all that can generally be inferred from subjects' responses is that there are differences in the results obtained by different but equally able subjects.

Therefore, a more technical term to refer to this problem is DIF or differential test functioning (DTF). The precise definition of DIF varies across methods, according to whether binary or polytomous items are being examined. In general, DIF exists if an item's statistical properties vary according to the groups under study when these are matched on the attribute measured by the item. There are normally two groups: the focal group (generally the minority group), in which DIF is assumed to exist, and the reference group (generally the majority group). In dichotomous items, DIF is said to exist when the odds of a correct response to the item vary according to the grouping variable. In the framework of item response theory (IRT), DIF is defined as a difference in the conditional probabilities of item response in two or more groups.

When seeking to detect DIF it is important to distinguish between impact and differential functioning. Thus, item impact is present when the differences observed in item performance for two different population groups are real differences in the characteristic measured by the test; in contrast, DIF exists when subjects with the same level of the measured characteristic have different probabilities of successfully responding to a given item, depending on the group to which they belong.

Types of DIF

In general, two types of DIF can be distinguished: uniform and nonuniform. Previous research has identified both

DIF types in real data sets (Ellis, 1989; Hambleton and Rogers, 1989; Mellenbergh, 1982). DIF is said to be uniform when the probability of responding correctly to an item is greater for one group than another across all levels of the trait. In contrast, DIF is nonuniform when the difference between the two groups in terms of the probability of responding correctly to the item is not the same at all levels of the trait; in other words, there is an interaction effect.

DIF Techniques: A Classification

Although more than 40 years have passed since Cleary and Hilton (1968) first proposed using analysis of variance (ANOVA) to detect items that were biased by race or gender, the identification of such items in educational, psychological, and/or social measurement instruments is an issue that continues to arouse considerable interest. Indeed, many studies, whether from a methodological or an applied perspective, have addressed this controversial topic. Within the methodological field, various statistics have been proposed to detect DIF and the relative efficacy of each one has been analyzed under different applied conditions using simulation studies.

Given the wide range of statistical procedures available for detecting items with DIF, several authors have sought to classify them (Camilli and Shepard, 1994; Hidalgo and Gómez-Benito, 1999; Millsap and Everson, 1993; Potenza and Dorans, 1995). Camilli and Shepard (1994) group the procedures into three categories: (1) classical test theory (CTT) and analysis of variance methods; (2) methods based on IRT; and (3) contingency table methods. **Table 1** shows a classification of DIF detection methods according to the following criteria:

1. type of model: parametric or nonparametric;
2. type of matching variable: observed or latent; and
3. type of item: binary or polytomous.

As can be seen in **Table 1**, all these procedures take the matching variable as continuous and DIF is detected on the basis of the observed score or latent variable. Those methods which use an estimation of the latent variable first specify a measurement model and then check to see if the parameters of this model remain invariable for the different groups (Millsap and Everson, 1993). Although it is not very common to find continuous items, it is possible to detect DIF for them by using, for example, linear regression analysis or confirmatory factor analysis, or the more general framework of structural equation models.

Nonparametric DIF Techniques

The statistical techniques which fall into this category are generally based on contingency tables. Three approaches can be distinguished:

Table 1 Classification of DIF techniques

<i>Item format</i>	<i>Matching variable</i>	<i>Model</i>	<i>DIF technique</i>
Binary	Observed score	Nonparametric	Mantel–Haenszel (MH) binary statistic
		Parametric	Standardization Logistic regression analysis Log-linear models Latent class models
	Latent trait	Nonparametric	SIBTEST CATSIB
		Parametric	IRT log-likelihood ratio Lord chi-squared statistic Area measures IRT-Differential Item Functioning Test (DIFT) Confirmatory factor analysis (CFA) Multiple-indicator multiple-cause (MIMIC)
Polytomous	Observed score	Nonparametric	MH-generalized MH-ordinal Standardization
		Parametric	Multinomial logistic regression Discriminant logistic regression Log-linear models Latent class models
	Latent trait	Nonparametric	POLYSIBTEST
		Parametric	IRT log-likelihood ratio Lord chi-squared statistic Area measures IRT-Differential Item Functioning Test (DIFT) Confirmatory factor analysis (CFA) Multiple-indicator multiple-cause (MIMIC)

1. those which involve testing the hypothesis of equal proportions;
2. those which use a categorical model; and
3. the simultaneous item bias test (SIBTEST) procedure.

The first group includes χ^2 (Scheuneman, 1979), the standardization procedure (Dorans and Kulick, 1986), and the Mantel–Haenszel (MH) statistic (Holland and Thayer, 1988); the second would comprise logit and log-linear models (Mellenbergh, 1982), latent class models (Cohen and Bolt, 2002; De Ayala, *et al.*, 2002), and logistic regression (LR) analysis (Swaminathan and Rogers, 1990); and finally, the third group includes the SIBTEST framework (Shealy and Stout, 1993a, 1993b). Among these techniques, the most commonly used are MH and LR.

The MH procedure (Mantel and Haenszel, 1959) was first used to analyze DIF by Holland and Thayer (1988). It is now the gold standard for detecting items with differential functioning and is used by the Educational Testing Service. The MH procedure compares the item performance of the reference and focal groups, which were previously matched on the ability that the test measures. The observed total test score is normally used as the matching criterion (Holland and Thayer, 1988).

In the standard MH procedure, an item shows uniform DIF if the odds of correctly answering the item are different for the two groups at a given level j of the matching

variable. The MH method provides an estimation of the effect size based on the common odds ratio (α). Holland and Thayer (1988) proposed a logarithmic transformation of α for interpretive purposes, with the aim of obtaining a symmetrical scale in which a zero value indicates an absence of DIF, a negative value indicates that the item favors the reference group over the focal group, and a positive value indicates that the item is easier for the focal group than for the reference group. This transformation is expressed as $\Delta_{\alpha\text{MH}} = -2.35 \ln(\alpha)$. Zwick and Ercikan (1989) proposed the following interpretative rules to evaluate the DIF effect size:

1. type A items/negligible DIF: these are items in which $\Delta_{\alpha\text{MH}} < |1|$;
2. type B items/moderate DIF: items in which $|1| \leq \Delta_{\alpha\text{MH}} \leq |1.5|$ and where MH proved to be statistically significant; and
3. type C items/large DIF: items in which $\Delta_{\alpha\text{MH}} > |1.5|$ and MH proved to be statistically significant.

Zwick and Ercikan (1989) point out that type B items can be used in the test if there are no others to replace them, whereas type C items will only be selected if they are necessary to achieve test specifications.

There are several advantages to this method, the main one being its high power rate of DIF detection (Holland

and Thayer, 1988). Furthermore, the statistic is easy to calculate, includes a measurement of effect size, and can be readily interpreted by professionals without expertise in psychometrics. In addition, its application does not require excessively large sample sizes; for example, Mazor *et al.* (1992) found high power and good control of the type-1 error rate in samples of 200 subjects per group (focal and reference). Finally, it should be noted that the MH statistic can be calculated using easily accessible statistical software, whether that for general use (SPSS, SYSTAT, and SAS) or more specific packages (MHDIF: Fidalgo, 1994; EZDIF: Waller, 1998a; DIFAS: Penfield, 2005). However, there are a number of limitations:

1. it does not test for nonuniform DIF (Rogers and Swaminathan, 1993; Swaminathan and Rogers, 1990; Uttaro and Millsap, 1994), although Clauser *et al.* (1994) did propose a variation that is useful for its detection;
2. it works with different levels of a matching variable, which is somewhat arbitrary and may affect statistical decisions regarding DIF;
3. poor performance in terms of its power in detecting DIF with sample sizes of 100 or fewer per group (Fidalgo *et al.*, 2004; Mazor *et al.*, 1992), although exact nonparametric methods have been proposed for both binary and polytomous items (Meyer *et al.*, 2004; Parshall and Miller, 1995);
4. it works with an observed score metric; and
5. it may need to purify matching scores with a small number of items or in a test with a high percentage of items with differential functioning.

LR analysis (Swaminathan and Rogers, 1990) is a technique for detecting different types of DIF (uniform and nonuniform). It is a more sophisticated and complex procedure than others such as the MH statistic, but has greater power in terms of detecting nonuniform DIF (Clauser *et al.*, 1996; Hidalgo and López, 2004; Narayanan and Swaminathan, 1996; Rogers and Swaminathan, 1993). Moreover, LR analysis provides a general and flexible framework for analyzing and evaluating DIF, for example: different item formats (dichotomous or polytomous), different matching criteria or a combination of other criteria, and grouping variables (gender, race, culture, and language) with two or more levels (Zumbo, 1999). Other advantages of this method are:

1. it offers a statistical test and an effect size statistic (French and Maller, 2007; Gómez-Benito *et al.*, 2009; Hidalgo and López, 2004; Hidalgo and Gómez-Benito, 2006b; Jodoin and Gierl, 2001; Zumbo, 1999);
2. it is useful with relatively small sample sizes (Hidalgo and Gómez-Benito, 2006b, 2009); and
3. it can be calculated using readily available software, whether general statistical packages (SPSS, Systat, and

SAS) or more specific applications (LRDIF: Gómez-Benito *et al.*, 2005; EZDIF: Waller, 1998a).

Its limitations include the fact that its sample size requirements depend on the number of response categories and on the degree of skew, such that for items without skew, lower sample sizes (e.g., 100) may be adequate and both uniform and nonuniform DIF can be detected. Two more limitations are that it works with an observed score metric and it may need to purify matching scores with a small number of items.

The SIBTEST (Shealy and Stout, 1993a, 1993b) is based theoretically on the multidimensional model of IRT. This procedure tests for DIF by using differences in the expected scores that are conditional on ability across groups. The conditional expected scores in each group are estimated by conditioning on a valid subtest free of DIF. The results obtained by this method are similar to those of the MH procedure in terms of the direction and amount of DIF estimated (Nandakumar, 1993; Narayanan and Swaminathan, 1994; Roussos and Stout, 1996; Shealy and Stout, 1993b).

The most important advantages of this method include the fact that it is able to detect both DIF and DTF, and therefore it can explore the effects of DIF amplification and cancellation. Amplification is produced when a subset of items only shows DIF in one of the groups. In contrast, cancellation occurs when one subset of test items shows DIF in one of the compared groups, while another subset does so in the other group; this may mean that the effect of the first subset of items cancels out the effect of the second. The method has been extended in order to estimate DIF in polytomous items (Poly-SIBTEST: Chang, *et al.*, 1996) and to assess nonuniform DIF (Crossing-SIBTEST: Li and Stout, 1996). In addition, Nandakumar and Roussos (2004) have proposed another extension of the procedure (CATSIB), in which the observed total score is replaced by IRT-based estimates.

The main disadvantage of this method is that determining the criteria which will be used to select the valid subtest which does not contain items with DIF is a complex procedure. Furthermore, it may not be powerful with smaller sample sizes (Bolt, 2002). Finally, the detection of nonuniform DIF in polytomous items using Poly-SIBTEST has proved somewhat problematic.

Parametric DIF Techniques

IRT includes a wide range of DIF statistics, for example, parameter comparisons such as the Lord statistic (Cohen *et al.*, 1993; Lord, 1980), model comparisons using the likelihood ratio (Thissen, 2001; Thissen *et al.*, 1988), area measures (Cohen *et al.*, 1993; Kim and Cohen, 1991; Raju, 1988, 1990), and methods based on the differential functioning

of items and tests (DFIT) framework (Flowers *et al.*, 1999; Oshima *et al.*, 1997, 2006; Raju *et al.*, 1995).

The Lord statistic and the likelihood ratio test are based on the comparison of the item parameters obtained in different groups: in the first case, the parameters estimated separately for each group (reference and focal) are compared directly, while the second approach involves comparing models in which it is assumed that the parameters of one or more test items vary between the focal and reference groups. Area measures are also obtained by estimating separately the item parameters of the groups to be compared, but the method calculates the area between the item characteristic curve (ICC) of the reference group and the ICC of the focal group.

DFIT procedures can be used with both dichotomous and polytomous items, as well as with multidimensional tests (Flowers *et al.*, 1999; Oshima *et al.*, 1995, 1997). This procedure estimates the size of the difference between the true score of a subject as a member of the reference group and the true score of the same subject as a member of the focal group. The greater the difference between these scores the greater the DTF.

The main advantage of these methods is that they are based on a well-developed theoretical model, since IRT can examine both uniform and nonuniform DIF. When polytomous items are used, IRT methods can measure the impact of DIF on the total score using a test response function. However, some limitations of this procedure are:

1. its sample size requirements in order to ensure an appropriate estimate of the item parameters;
2. the model must fit the data;
3. the model assumptions must be met;
4. the DFIT method, Lord statistic, and area measures all require specific computer programs;
5. the IRT/log-likelihood method requires intensive computer calculation; and
6. the software is not very user friendly, although Thissen (2001) has developed IRTLRDIF, a software implementation of the model comparison approach, and Waller (1998b) has created LINKDIF, a package which implements area measures.

Structural equation models (SEMs) include several procedures to detect DIF, such as restricted factor analysis (RFA) (Gómez-Benito and Navas-Ara, 2000; Navas-Ara and Gómez-Benito, 2002; Oort, 1992, 1998), multi-group confirmatory factor analysis with latent mean and covariance structure (González-Romá *et al.*, 2006), or the multiple-indicator multiple-cause (MIMIC) model (MacIntosh and Hashim, 2003; Muthén, 1989). The advantage of RFA is that it can detect DIF with respect to various bias factors simultaneously, although it is not able to distinguish between uniform and nonuniform DIF. The multigroup CFA approach is useful for assessing the

presence of uniform and nonuniform DIF, but its main limitation is the need for large sample sizes, especially if more than two groups are to be examined; moreover, it does not enable direct modeling of the relationship between group membership and the latent trait (Finch, 2005). In contrast, the MIMIC approach involves the estimation of both direct and indirect effects for a grouping variable. Thus, this model introduces grouping variables, which are assumed to influence the latent factors and the DIF model is incorporated by adding direct effects from the grouping variables to the observed items unmediated by the latent factors. In addition, it is possible to check for the presence of DIF for more than two groups, and with multiple grouping variables, including both categorical and continuous types (Muthén, 1989). The main disadvantage of this approach is the inflated type-1 error rate for short tests of fewer than 50 items (Finch, 2005). In general, these SEM-based approaches are able to detect DIF in multidimensional models and provide various fit indices at the level of both overall and local models (e.g., in the detection of certain noninvariant items). It should also be noted that these procedures not only enable the assessment of DIF but are also very useful in the evaluation of construct equivalence and the analysis of cross-cultural measurement invariance.

Purification of Matching Variables

The methods currently used to detect DIF use as their matching criterion the total test score or an estimate of ability. However, matching on an internal criterion that involves some DIF items can lead to inaccurate ability estimates and, therefore, inaccurate DIF identification (Clauser *et al.*, 1993; Kim and Cohen, 1992).

To mitigate this problem, what are known as purification procedures have been proposed: in an initial stage, the DIF is evaluated in a standard way, but it is then evaluated again using only those items which do not present DIF in the initial analysis. An estimate of the competence level of subjects is then calculated.

Purification procedures are classified into two types:

- *Iterative.* The items with DIF are reevaluated by successive iterations, the aim of which is to gradually purge these items so that they do not form part of the criterion.
- *Two stage.* First, the DIF is evaluated and then, in a second step, those items which in the first stage were shown to be biased are eliminated from the total test score, before once again calculating the DIF.

The contributions of Marco (1977) and Lord (1980), within the framework of IRT, represent the first steps toward an estimate of the competence level of subjects which is not distorted by DIF items; other proposals can

be found in Park (1988) and Park and Lautenschlager (1990). Different purification procedures have also been used outside the framework of IRT, for example, the iterative logit method (Van der Flier *et al.*, 1984), the two-stage version of the MH method (Holland and Thayer, 1988), the two-stage estimation of multinomial LR (Hidalgo and Gómez-Benito, 2003), and iterative RFA (Navas-Ara and Gómez-Benito, 2002).

All the simulation studies carried out with the different DIF detection methods agree that purification procedures reduce false-positive error rates and/or increase power, and also that iterative procedures always present better results than when DIF is evaluated through just one analysis (Clauser *et al.*, 1993; French and Maller, 2007; Hidalgo and Gómez-Benito, 2003; Hidalgo and López, 2002; Kim and Cohen, 1992; Miller and Oshima, 1992; Navas-Ara and Gómez-Benito, 2002; Park and Lautenschlager, 1990).

Guidelines to Detect DIF

A number of practical statistical guidelines for the correct detection of DIF need to be taken into account. First, it is necessary to select an acceptable DIF detection method on the basis of what has been discussed previously. Second, in an empirical study more than one method for evaluating DIF should be used; for example, Millsap and Everson (1993) recommend that when the Lord statistic is significant it should be combined with an area measure. Third, it is important to use a purification procedure for the matching criteria. Finally, in DIF studies, it is advisable to support statistical results with an effect size statistic that allows a more accurate evaluation to be carried out and, subsequently, enables decisions to be made regarding the presence or absence of DIF in the item. Furthermore, the combination of a statistical test with an effect size measure will help to reduce false identification rates.

However, in DIF analysis, other nonstatistical recommendations also need to be taken into account. First, DIF analysis should be followed by content analysis, and, if it is feasible to identify some possible causes of the DIF, we can consider tentative bias factors as evidence of a threat to test validity. Second, the effect that would occur if the DIF item were removed from the test needs to be explored; for example, some findings indicate that item-level DIF does not necessarily make a difference when the scale is considered as a whole. Third, what would the consequence be of decisions made as a result of the assessment process, that is, consideration must be given to consequential validity.

In general, there has been a poor interrelationship between substantive and statistical analyses and attempts to identify the reasons for this have not evolved at the same pace as statistical procedures; thus, there is a deficit

in this area (Bolt and Stout, 1996; Englehard *et al.*, 1990; Gierl, 2005; Roussos and Stout, 1996; American Psychological Association *et al.*, 1999; Sudweeks and Tolman, 1993). Moreover, Gierl and Khaliq (2001) point out that substantive reviewers have not been successful in predicting which items will contain DIF. To strengthen the interrelationship between the two approaches Roussos and Stout (1996) recommended a two-step procedure. The first step consists in a substantive analysis in which DIF hypotheses are generated, and it is decided whether or not DIF is present as a secondary factor. In this stage, substantive reviewers use different sources for the review process, such as previously published DIF analyses, reviews of contexts present in other similar data, or substantive content considerations. In the second stage, statistical analyses are conducted in order to test the DIF hypothesis. Teresi (2004) presented a set of steps that could be performed in examining DIF:

1. convene focus groups to examine how members of different groups perceive the meaning of the studied construct;
2. conduct in-depth interviews to supplement information gathered in the focus groups;
3. convene a panel of experts to review closed-ended items; and
4. perform cognitive interviews, presenting individuals with closed-ended items followed by open-ended probes.

In sum, nowadays there is clearly a growing trend toward explaining DIF, rather than merely identifying it, and thus its study forms part of the validity framework.

Summary

As was pointed out by Gómez-Benito *et al.* (2005) in their historical review of the field over the last quarter of the twentieth century, more than 80% of publications on DIF appeared during the 1990s, and of these, three-quarters were published in the last 5 years of this decade. Moreover, one of the areas showing the greatest interest in the subject of DIF is education. This growing interest is not surprising if one considers the social and political dimensions of the problem, in which there is increasing concern to ensure equality of opportunity among individuals and social groups. Furthermore, possible bias must also be taken into account in tests used for clinical/educational diagnosis, due to the biopsychosocial and educational consequences of how the results are used.

Although enormous progress has clearly been made in terms of developing statistical methods to identify items that show DIF, whether these are dichotomous or polytomous, a key question which remains is that regarding the choice of a particular method. This will depend on several factors including sample size of focal and reference

groups, type of DIF, item scoring, matching variable, and availability of statistical software. In this regard, it should be noted that greater emphasis is now placed on studying the behavior of these statistical procedures in situations that are common in practice, for example, the identification of DIF in small samples (Fidalgo *et al.*, 2004; Hidalgo and Gómez-Benito, 2006b, 2007; Muñoz *et al.*, 2001) and the study of the advantages and limitations of polytomous procedures (Bolt, 2002; Hidalgo and Gómez-Benito, 2006a). Also noteworthy is the attempt to include effect size measures that complement the information provided by the test of significance, as well as the development of flagging rules and cutoff values in order to make decisions about items with DIF (Lai *et al.*, 2005; Hidalgo and López, 2004; Jodoin and Gierl, 2001; Zumbo and Thomas, 1997). Finally, mention should be made of the development of specific software programs that serve a dual purpose, that is, they offer a more user-friendly application for use by nonexpert professionals and also incorporate the latest developments in the field; examples of such software are DIF/DBF (Stout and Roussos, 1999), IRTLRFID (Thissen, 2001), TESTGRAPH (Ramsay, unpublished), and LRIDF (Gómez-Benito, *et al.*, 2005).

To summarize, although it is true that in recent decades DIF has been the focus of many studies carried out in the field of educational measurement, much work remains to be done. This is the case not only in terms of methodology but also with respect to the applied field, where DIF studies are increasingly being used as a routine step in the process of developing new measurement instruments and evaluating existing ones; furthermore, there is a growing awareness of the need to analyze the potential causes of the differential behavior shown by the items of a given instrument. This problem of bias analysis as a fundamental aspect of measurement instrument validity is a field characterized by much debate but one which, due to its intrinsic difficulty, has yet to be explored in depth. It is clear, therefore, that future efforts must seek to address this issue.

See also: Statistical Significance Versus Effect Size; Test Translation and Adaptation; Validity.

Bibliography

- American Psychological Association, American Educational Research Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.
- Bolt, D. M. (2002). A Monte-Carlo comparison of parametric and nonparametric polytomous DIF detection methods. *Applied Psychological Measurement* **15**, 113–141.
- Bolt, D. and Stout, W. (1996). Differential item functioning. Its multidimensional model and resulting SIBTEST detection procedure. *Behaviormetrika* **23**, 67–95.
- Camilli, G. and Shepard, L. A. (1994). *Methods for Identifying Biased Test Items*. Newbury Park, CA: Sage.
- Clauser, B. E., Mazor, K. M., and Hambleton, R. K. (1993). The effects of purification of the matching criterion on the identification of DIF using the Mantel-Haenszel procedure. *Applied Measurement in Education* **6**, 269–279.
- Clauser, B. E., Mazor, K. M., and Hambleton, R. K. (1994). The effects of score group width on the Mantel-Haenszel procedure. *Journal of Educational Measurement* **31**, 67–78.
- Clauser, B. E., Nungester, R. J., Mazor, K., and Ripkey, D. (1996). A comparison of alternative matching strategies for DIF detection in tests that are multidimensional. *Journal of Educational Measurement* **33**, 202–214.
- Chang, H., Mazzeo, J., and Roussos, L. A. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *Journal of Educational Measurement* **33**, 333–353.
- Cleary, T. A. and Hilton, T. L. (1968). An investigation of item bias. *Educational and Psychological Measurement* **28**, 61–75.
- Cohen, A. S. and Bolt, D. M. (2002). A mixture model analysis of differential item functioning. *Paper Presented at the Annual Meeting of the American Educational Research Association*. New Orleans.
- Cohen, A. S., Kim, S. H., and Baker, F. B. (1993). Detection of differential item functioning in the graded response model. *Applied Psychological Measurement* **17**, 335–350.
- De Ayala, R. J., Kim, S.-H., Stapleton, L. M., and Dayton, C. M. (2002). Differential item functioning: A mixture distribution conceptualization. *International Journal of Testing* **2**, 243–276.
- Dorans, N. J. and Kulick, E. M. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement* **23**, 355–368.
- Ellis, B. (1989). Differential item functioning: Implications for test translations. *Journal of Applied Psychology* **74**, 912–921.
- Englehard, G., Hansche, L., and Rutledge, K. E. (1990). Accuracy of bias review judges in identifying differential item functioning on teacher certification tests. *Applied Measurement in Education* **3**, 347–360.
- Fidalgo, A. M. (1994). MHDIF: A computer program for detecting uniform and nonuniform differential item functioning with the Mantel-Haenszel procedure. *Applied Psychological Measurement* **18**, 300.
- Fidalgo, A. M., Ferreres, D., and Muñoz, J. (2004). Utility of the Mantel-Haenszel procedure for detecting differential item functioning in small samples. *Educational and Psychological Measurement* **64**, 925–936.
- Finch, H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel-Haenszel, SIBTEST, and the IRT Likelihood ratio. *Applied Psychological Measurement* **29**, 278–295.
- Flowers, C. P., Oshima, T. C., and Raju, N. S. (1999). A description and demonstration of the polytomous-DFIT framework. *Applied Psychological Measurement* **23**, 309–326.
- French, B. F. and Maller, S. J. (2007). Iterative purification and effect size use with logistic regression for differential item functioning detection. *Educational and Psychological Measurement* **67**, 373–393.
- Gierl, M. J. (2005). Using dimensionality-based DIF analyses to identify and interpret constructs that elicit group differences. *Educational Measurement: Issues and Practice* **24**, 3–14.
- Gierl, M. J. and Khaliq, S. N. (2001). Identifying sources of differential item and bundle functioning on translated achievement tests. *Journal of Educational Measurement* **38**, 164–187.
- Gómez-Benito, J., Hidalgo, M. D., Guilera, G., and Moreno, M. (2005). A bibliometric study of differential item functioning. *Scientometrics* **64**, 3–16.
- Gómez-Benito, J., Hidalgo, M. D., and Padilla, J. L. (2009). Efficacy of measures of the effect size in logistic regression: An application in the detection of DIF. *Methodology* **5**, 18–25.
- Gómez-Benito, J., Hidalgo, M. D., Padilla, J. L., and González, A. (2005). Desarrollo informático para la utilización de la regresión logística como técnica de detección del DIF ('Software for the use of logistic regression as DIF detection procedure'). *Paper Presented at the IX Congreso de Metodología de las Ciencias Sociales y de la Salud*, Granada, Spain.
- Gómez-Benito, J. and Navas-Ara, M. J. (2000). A comparison of chi-square, RFA and IRT based procedures in the detection of DIF. *Quality and Quantity* **34**, 17–31.
- González-Romá, V., Hernández, A., and Gómez-Benito, J. (2006). Power and type I error of the mean and covariance structure analysis

- model for detecting differential item functioning in graded response items. *Multivariate Behavioral Research* **41**, 29–53.
- Hambleton, R. K. and Rogers, H. J. (1989). Detecting potentially biased test items: Comparison of IRT area and Mantel-Haenszel methods. *Applied Measurement in Education* **2**, 313–334.
- Hidalgo, M. D. and Gómez-Benito, J. (1999). Técnicas de Detección de Funcionamiento Diferencial en Items Politémicos. ('Differential item functioning detection techniques in polytomous items'). *Metodología de las Ciencias del Comportamiento* **1**, 39–60.
- Hidalgo, M. D. and Gómez-Benito, J. (2003). Test purification and the evaluation of differential item functioning with multinomial logistic regression. *European Journal of Psychological Assessment* **19**, 1–11.
- Hidalgo, M. D. and Gómez-Benito, J. (2006a). Nonuniform DIF detection using discriminant logistic analysis and multinomial logistic regression: A comparison for polytomous items. *Quality and Quantity* **40**, 805–823.
- Hidalgo, M. D. and Gómez-Benito, J. (2006b). DIF detection in small samples: Logistic regression analysis and effect size measure. *Paper Presented at the 5th Conference of the International Test Commission*. Brussels, Belgium.
- Hidalgo, M. D. and Gómez, J. (2009). DIF detection in small samples using logistic regression analysis with effect size measure (submitted).
- Hidalgo, M. D. and Lopez, J. A. (2002). Two-stage equating in differential item functioning detection under the Graded Response Model with the Raju area measures and the Lord statistic. *Educational and Psychological Measurement* **62**, 32–44.
- Hidalgo, M. D. and Lopez, J. A. (2004). DIF detection and effect size: A comparison between logistic regression and Mantel-Haenszel variation. *Educational and Psychological Measurement* **64**, 903–915.
- Holland, P. W. and Thayer, D. T. (1988). Differential item performance and Mantel-Haenszel procedure. In Wainer, H. and Braun, H. I. (eds.) *Test Validity*, pp 129–145. Hillsdale, NJ: Erlbaum.
- Jodoin, M. G. and Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with Logistic Regression procedure for DIF detection. *Applied Measurement in Education* **14**, 329–349.
- Kim, S. H. and Cohen, A. S. (1991). A comparison of two area measures for detecting differential item functioning. *Applied Psychological Measurement* **15**, 269–278.
- Kim, S. H. and Cohen, A. S. (1992). Effects of linking methods on detection of DIF. *Journal of Educational Measurement* **29**, 51–66.
- Lai, J. S., Teresi, J., and Gershon, R. (2005). Procedures for the analysis of differential item functioning (DIF) for small samples sizes. *Evaluation and the Health Professions* **28**, 283–294.
- Li, H. -H. and Stout, W. (1996). A new procedure for detection of crossing DIF. *Psychometrika* **61**, 647–677.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum.
- MacIntosh, R. and Hashim, S. (2003). Variances estimation for converting MIMIC model parameters to IRT parameters in DIF analysis. *Applied Psychological Measurement* **27**, 372–379.
- Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies disease. *Journal of the National Cancer Institute* **22**, 719–748.
- Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement* **14**, 139–160.
- Mazor, K. M., Clauser, B. E., and Hambleton, R. K. (1992). The effect of sample size on the functioning of the Mantel-Haenszel statistic. *Educational and Psychological Measurement* **52**, 443–451.
- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics* **7**, 105–118.
- Meyer, J. P., Huynh, H., and Seaman, M. A. (2004). Exact small-sample differential item functioning methods for polytomous items with illustration based on an attitude survey. *Journal of Educational Measurement* **41**, 331–344.
- Miller, M. D. and Oshima, T. C. (1992). Effect of sample size, number of biased items and magnitude of bias on a two-stage item bias estimation method. *Applied Psychological Measurement* **16**, 381–388.
- Millsap, R. E. and Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement* **17**, 297–334.
- Muñiz, J., Hambleton, R. K., and Xing, D. (2001). Small sample studies to detect flaws in item translations. *International Journal of Testing* **1**, 115–135.
- Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika* **54**, 557–585.
- Nandakumar, R. (1993). Simultaneous DIF amplification and cancellation: Shealy-Stout's test for DIF. *Journal of Educational Measurement* **16**, 159–176.
- Nandakumar, R. and Roussos, L. (2004). Evaluation of the CATSIB procedure in a pretest setting. *Journal of Educational and Behavioral Statistics* **29**, 177–200.
- Narayanan, P. and Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential item functioning. *Applied Psychological Measurement* **18**, 315–338.
- Narayanan, P. and Swaminathan, H. (1996). Identification of item that show nonuniform DIF. *Applied Psychological Measurement* **20**, 257–274.
- Navas-Ara, M. J. and Gómez-Benito, J. (2002). Effects of ability scale purification on the identification of DIF. *European Journal of Psychological Assessment* **18**, 9–15.
- Oort, F. J. (1992). Using restricted factor analysis to detect item bias. *Methodika* **VI**, 150–166.
- Oort, F. J. (1998). Simulation study of item bias detection with restricted factor analysis. *Structural Equation Modeling* **5**, 107–124.
- Oshima, T. C., Raju, N. S., and Flowers, C. (1997). Development and demonstration of multidimensional IRT-based internal measures of differential item functioning of items and tests. *Journal of Educational Measurement* **34**, 253–272.
- Oshima, T. C., Raju, N. S., Flowers, C., and Monaco, M. (1995). A Monte Carlo assessment of DFIT with dichotomously scored multidimensional tests. *Paper Presented at the Annual Meeting of American Educational Research Association*, San Francisco.
- Oshima, T. C., Raju, N. S., and Nanda, A. O. (2006). A new method for assessing the statistical significance in the differential functioning of item and tests (DFIT) framework. *Journal of Educational Measurement* **43**, 1–17.
- Osterlind, S. J. and Everson, H. T. (2009). *Differential item functioning*, 2nd edn. Thousand Oaks, California: Sage Publications, Inc.
- Park, D. G. (1988). *Investigations of Item Response Theory Item Bias Detection*. Unpublished doctoral dissertation, University of Georgia, Department of Psychology.
- Park, D. G. and Lautenschlager, G. J. (1990). Improving IRT item bias detection with iterative linking and ability scale purification. *Applied Psychological Measurement* **14**, 163–173.
- Parshall, C. G. and Miller, T. R. (1995). Exact versus asymptotic Mantel-Haenszel DIF statistics: A comparison of performance under small-sample conditions. *Journal of Educational Measurement* **32**, 302–316.
- Penfield, R. D. (2005). DIFAS: Differential item functioning analysis. *Applied Psychological Measurement* **29**, 150–151.
- Potenza, M. T. and Dorans, N. J. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. *Applied Psychological Measurement* **19**, 23–37.
- Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika* **53**, 284–291.
- Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement* **14**, 197–207.
- Raju, N. S., van der Linden, W. J., and Fleer, P. F. (1995). IRT-based internal measures of differential item functioning of items and tests. *Applied Psychological Measurement* **19**, 353–368.
- Ramsay, J. O. (unpublished). *TestGraph: A Program for the Graphical Analysis of Multiple Choice and Test Questionnaire*.
- Rogers, H. J. and Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement* **17**, 105–116.
- Roussos, L. A. and Stout, W. F. (1996). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel type I error performance. *Journal of Educational Measurement* **33**, 215–230.
- Scheuneman, J. D. (1979). A new method for assessing bias in test items. *Journal of Educational Measurement* **16**, 143–152.

- Shealy, R. T. and Stout, W. F. (1993a). An item response theory model for test bias and differential test functioning. In Holland, P. W. and Wainer, H. (eds.) *Differential Item Functioning*, pp 197–239. Hillsdale, NJ: LEA.
- Shealy, R. T. and Stout, W. F. (1993b). A model-based standardization approach that separates true bias/DIF from group differences and detects test bias/DIF as well as item bias/DIF. *Psychometrika* **58**, 159–194.
- Stout, W. and Roussos, L. (1999). *Dimensionality-Based DIF/DBF Package*. [Computer Program] William Stout Institute for Measurement. University of Illinois.
- Sudweeks, R. R. and Tolman, R. R. (1993). Empirical versus subjective procedures for identifying gender differences in science test items. *Journal of Research in Science Teaching* **30**, 3–19.
- Swaminathan, H. and Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement* **27**, 361–370.
- Teresi, J. (2004). Differential item functioning and health assessment. *Paper Presented at the Conference Advances in Health Outcomes Measurement: Exploring the Current State and the Future of Item Response Theory, Item Banks, and Computer-Adaptive Testing*. Bethesda, MD, June 23–25.
- Thissen, D. (2001). *IRTLDIF v2.0b. Software for the computation of the statistics involved in Item Response Theory Likelihood-Ratio Test for Differential Item Functioning*. Available on Dave Thissen's web page.
- Thissen, D., Steinberg, L., and Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In Wainer, H. and Braun, H. I. (eds.) *Test Validity*, pp 147–169. Hillsdale, NJ: Erlbaum.
- Uttaro, T. and Millsap, R. E. (1994). Factors influencing the Mantel–Haenszel procedure in the detection of differential item functioning. *Applied Psychological Measurement* **18**, 15–25.
- Van der Flier, H., Mellenbergh, G. J., Ader, H. J., and Wijn, M. (1984). An iterative item bias detection method. *Journal of Educational Measurement* **21**, 131–145.
- Waller, N. G. (1998a). EZDIF: Detection of uniform and nonuniform differential item functioning with the Mantel–Haenszel and logistic regression procedures. *Applied Psychological Measurement* **22**, 391.
- Waller, N. G. (1998b). LINKDIF: Linking item parameters and calculating IRT measures of differential item functioning of items and tests. *Applied Psychological Measurement* **22**, 392.
- Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.
- Zumbo, B. D. (2007). Three generations of DIF analysis: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly* **4**, 223–233.
- Zumbo, B. D. and Thomas, D. R. (1997). A measure of effect size for a model-based approach for studying DIF. *Working Paper of the Edgeworth Laboratory for Quantitative Behavioral Science*. University of Northern British Columbia: Prince George, BC.
- Zwick, R. and Ercikan, K. (1989). Analysis of differential item functioning in the NAEP history assessment. *Journal of Educational Measurement* **26**, 55–66.

Further Reading

- Clauser, B. E. and Mazor, K. M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice* **17**, 31–44.
- Holland, P. W. and Wainer, H. (eds.) (1993). *Differential Item Functioning*. Hillsdale, NJ: LEA.
- Mazor, K. M., Clauser, B. E., and Hambleton, R. K. (1994). Identification of nonuniform differential item functioning using a variation of the Mantel–Haenszel procedure. *Educational and Psychological Measurement* **54**, 284–291.
- Penfield, R. D. and Lam, T. C. M. (2000). Assessing differential item functioning in performance assessment: Review and recommendations. *Educational Measurement: Issues and Practice* **19**, 5–15.
- Roussos, L. A. and Stout, W. (2004). Differential item functioning analysis. In Kaplan, D. (ed.) *The Sage Handbook of Quantitative Methodology for the Social Sciences*, pp 107–116. Thousand Oaks, CA: Sage.
- Welch, C. and Hoover, H. D. (1993). Procedures for extending item bias detection techniques to polytomously scored items. *Applied Measurement in Education* **6**, 1–19.

Educational Measurement: Overview

R L Linn, University of Colorado at Boulder, Ouray, CO, USA

© 2010 Elsevier Ltd. All rights reserved.

The purpose of this article is to provide an overview of educational measurement. A more detailed discussion of educational measurement can be found elsewhere in the encyclopedia. We begin with a description of four major functions of educational measurement and then move to a consideration of some characteristics of modern educational measurement.

The most recent edition of *Standards for Educational and Psychological Testing* (American Educational Research Association (AERA), American Psychological Association (APA), and the National Council for Measurement in Education (NCME); AERA, APA, and NCME, 1999) began with high praise for the contributions of testing. "Educational and psychological testing and assessment are among the most important contributions of behavioral science to our society" (AERA, APA, and NCME, 1999: 1). This praise must be balanced against a variety of criticisms on the uses of educational tests. It is also important to recognize that the value of tests depends heavily on the purposes of testing and the uses that are made of test results. Hence, it seems appropriate to begin with a consideration of some of the major functions of educational measurement.

Major Functions of Educational Measurement

Educational tests have long been a part of the educational landscape. In the past 10 or 20 years, however, educational measurement has come to play an increasingly prominent role in education. Educational tests and assessments are used for a wide variety of functions. However, four separate functions are particularly noteworthy. First, educational tests are used to make a variety of decisions about individual students (e.g., placement, grade-to-grade promotion, high school graduation, and admission to college). Second, tests are used to monitor student achievement and progress. Third, tests are used to leverage educational reform by holding schools and teachers accountable for student achievement. Fourth, classroom assessments are used for formative purposes and as an integral component of teaching and learning.

The second and third functions sometimes overlap and so do the school accountability and individual decision-making functions. Although some would like to see greater overlap between the accountability and formative functions, the

characteristics of assessments required for those two functions are substantially different.

Decisions about Students

Tests play an important role in many decisions that are made regarding the education of individual students. Test results are used to make decisions about course placement and they may determine whether or not a student is required to attend summer school or be retained in the same grade for another year. Passing tests may also be a requirement for receiving a high school diploma. Although college admission decisions may depend heavily on the courses students have taken and the grades they have received in high school, standardized tests also have a major influence on admission decisions in many colleges as well as graduate and professional schools. Test results earned in high school may also be used to award college credit and tests are often used to determine placement in remedial courses or more advanced college courses.

Many questions regarding technical adequacy arise when tests are used to make decisions about individual students. Justifying the use requires evidence of validity (see Linn) and reliability (see Brennan; Miller). Is the content coverage of the test adequate for the particular use and interpretations of the test scores? Do scores provide adequate prediction of future performance of students? Measurement error is an unavoidable characteristic of any test, but the question is whether the magnitude of the errors is acceptably small for the use that is made of the test scores. For tests used for college admissions or high school graduation decisions there are often questions about the equivalence of scores that are obtained in different administrations of the test either within or between years. Some of these and other measurement issues are discussed after brief discussions of the other three major functions of educational measurement.

Monitoring Function

There are examples of tests being used to obtain a reading on student achievement that date back to the beginning of the twentieth century. Rice (1897), for example, used achievement test results to compare schools. The results were also used to make judgments about the adequacy

of education. Until the last half of the twentieth century (see Crooks), however, major efforts to monitor achievement at the national level or through international studies did not get underway. The International Association for the Evaluation of Educational Achievement (IEA) was officially organized in 1967, although its roots can be traced back to 1958 (IEA, Mission Statement). In the United States the National Assessment of Educational Progress (NAEP) was first administered in 1969 after nearly 6 years of preparation (see, e.g., Jones, 2004).

In the past 40 years, IEA has sponsored several international assessments, including the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS). A substantial number of countries have participated in those assessments. Reports of results from TIMSS and PIRLS are available online from the TIMSS and PIRLS International Study Center. NAEP has expanded to include state as well as regional results. NAEP results have become increasingly prominent in discussions of student achievement in the United States and several other countries have introduced national assessments.

More recently, a series of international studies of student achievement have also been undertaken under the auspices of the Organization for Economic Cooperation and Development (OECD). OECD developed and arranged for the administration of the Program for International Student Assessment (PISA) in cooperating countries. PISA assesses 15-year-old students in reading, mathematics, and science. The assessments are conducted every 3 years with a focus on one of the three subject areas. For example, the focus was reading literacy in 2000, mathematics literacy in 2003, and science literacy in 2006.

For the monitoring function, there is no need to test every student. Good estimates of the achievement of age or grade cohorts of students can be achieved by assessing samples of students and different subsets of items may be administered to different students. In principle, sampling of students and the administration of different subsets of items to different students might be used for a system designed to hold schools accountable. In practice, however, census testing of all students with the same items is generally used for the accountability function.

Accountability Function

Policymakers have looked to educational test results as key indicators of educational quality and progress (see Hamilton). Lackluster test results have been used to advocate educational reforms. Educational tests and assessments have also been used as a critical educational reform tool. Educational policymakers can implement reforms in some countries by introducing changes in the mandated curriculum. In many countries, however, curricular changes alone

are relatively ineffective without being reinforced by other change mechanisms. As McDonnell (2004) has noted, “elected officials and policymakers have few tools at their disposal [to reform classroom instruction]. Standardized tests are one of the most effective levers they have for influencing what happens in local schools and classrooms” (p. 9).

The large-scale, externally mandated educational tests are used not only to monitor student achievement, but to leverage educational reform as well. The tests used for both the monitoring and the accountability functions are built upon an impressive technological foundation. The ability of these tests to yield reliable scores that are comparable from year to year, and have acceptable levels of validity, depends on the sophisticated use of psychometric techniques and statistics.

Formative Function

The formative function of educational measurement differs from the student decision making, monitoring, and accountability functions in several respects. Tests used for these other three functions are generally administered at a specified time during the school year. Formative assessments, on the other hand, do not occur at a fixed time during the school year. Rather, formative assessments are administered as an integral part of instruction at the moment when the information is needed. Tests serving the student decision making, monitoring, and accountability functions are standardized and considerable attention is given to the technical quality of the tests. In contrast, formative assessments are tailored to the student and the information from the assessment is used to foster student learning. As Shepard (2006) has noted, “For teachers to be effective in supporting student learning, they must constantly be checking for student understanding” (p. 627).

Unlike tests used for the other functions, formative assessments are under the control of classroom teachers. They may not even appear to be a test in the usual sense, but instead may consist of a few oral questions or informal observations. Formative assessment has been defined in several different ways; however, the following definition seems to capture the essence of formative assessment in a way that distinguishes it from tests used for other functions. “Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students’ achievement of defined instructional outcomes” (Council of Chief State School Officers, no date). This definition makes it clear that formative assessment is not just a test, but it is a process, one that may use an array of different procedures, both formal and informal, that are selected by the teacher to assess individual students.

Effective formative assessment not only requires feedback, but also needs to help students identify strategies for learning and be tied to achievement standards or learning outcomes (Sadler, 1989; Shepard, 2006). Black and Wiliam's (1998) review provided evidence that well-conceived and well-implemented formative assessment practices can yield large improvements in student learning. Few instructional interventions have effects that are as large as those associated with effective formative assessment practices.

Discussions of ways to improve formative assessments are better integrated into discussions of instruction rather than discussions of measurement. This is so because effective formative assessment practices cannot be divorced from effective instructional practices. Thus, formative assessment is excluded from the discussion below not because it is unimportant, but because it has more to do with good instructional practices than with technical measurement characteristics.

Educational Measurement Theory and Practice

Measurement theory, also known as psychometric theory, provides the foundation for evaluating educational tests and their uses and interpretations. Validity and reliability are the most fundamental measurement theory concepts. Together, validity and reliability provide the primary basis for judging the technical quality and the appropriateness of the uses and interpretations of educational tests results. There is widespread agreement among measurement experts that while reliability is important, validity is the most important consideration in evaluating the uses of and inferences that are made from test results. Other considerations, such as fairness and the comparability of test results for different test takers or from one occasion to another, are also important, but such considerations can be readily subsumed under the more fundamental topics of validity and reliability.

Validity

The Standards for Educational and Psychological Testing (AERA, APA, and NCME, 1999), which provide the most authoritative statements of professional consensus about topics such as validity and reliability, define validity as follows: "Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed in the uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests" (AERA, APA, and NCME, 1999: 9) (see Linn).

It is not unusual to hear a test referred to as valid or invalid. Such references are in keeping with outdated

conceptions of validity, but not with contemporary conceptualizations. It is the uses and interpretations of test results that are validated, rather than the test itself. A particular use or interpretation may have high validity, while another use or interpretation of the same test has unacceptably low validity. Validity is also a matter of degree rather than an all or non phenomenon, therefore the unqualified claim that a test is valid for a particular use is misleading.

Many types of evidence may be relevant to judging the validity of a particular use or interpretation of test scores. Considerations of content are often of central importance to judging the degree of validity of a claim that a test provides an adequate measure of the achievement of students in a given content domain. Relationships to other variables, such as college grades, may also be relevant for judging other uses of test scores. There is not a single type of evidence that is relevant to every use or interpretation. Rather, the priority that different types of evidence should have depends upon logical reasoning that incorporates the evidence into a justification for a specific use or interpretation of test scores.

Building on the conceptualizations of earlier validity theorists (e.g., Cronbach, 1988; Messick, 1989), Kane (2006) has described validity as an argument. According to this formulation, validation requires the evaluation of the rationale for a proposed use or interpretation of test scores. Proposed interpretations and uses must be clearly specified. The intended interpretations and uses provide the starting point for developing two kinds of argument: an interpretive argument and a validity argument. The interpretive argument needs to include "a clear and fairly complete statement of the claims included in the interpretation and the goals of any proposed test uses" (Kane, 2006: 22). Kane refers to the evaluation of the interpretive argument as the "validity argument."

The evidence that needs to be marshaled in for a validity argument depends on the claims that are made in the interpretive argument. For example, if it is claimed that persons with high test scores will be more likely to obtain high grades in college than persons with low scores, then evidence is needed that test scores are reasonable predictors of future college grades. Conversely, if it is claimed that schools where students have high test scores are more effective than schools with lower test scores, then the validity argument needs to include evidence of school effectiveness that goes beyond the test scores on which the claim is based (e.g., gains in test scores from one year to the next, and independent observations of the schools).

Kane (2006: 24–25) uses the example of placement testing to illustrate interpretive and validity arguments. One of the identified claims in the interpretive argument for placement testing is that people with scores below some identified minimum are unlikely to succeed in an identified course. The validity argument needs to evaluate

this claim, possibly with evidence that students with scores below the identified minimum have a lower likelihood of successful completion of the course than students with scores above the minimum.

Reliability

All test scores are subject to measurement error (see Miller). Because of measurement error scores from alternate forms of a test, from one set of items to another, from one occasion to another, or, in cases where scores depend on human raters, from one rater to another will not be perfectly consistent. In other words, scores on tests lack perfect precision due to the inevitable errors of measurement.

As Feldt and Brennan (1989) have noted, “Historically, the reliability of a measuring instrument of process has been quantified via two indexes: the standard error of measurement and the reliability coefficient” (p. 105). In theory, the standard error of measurement is the standard deviation of a hypothetical set of replications of the measurement procedure for a single individual. According to Feldt and Brennan (1989), “The reliability coefficient quantifies reliability by summarizing the consistency (or inconsistency) among several error prone measurements” (p. 105).

A variety of statistical models and techniques are available for use in quantifying the reliability or consistency of scores or the converse, score inconsistency due to errors of measurement. As Haertel (2006) has noted, these include approaches of classical test theory and generalizability theory as well as estimates of measurement error that are based on item response theory (IRT). (IRT approach is briefly discussed as a separate topic below.) As is true of validity, reliability is a matter of degree rather than an all or none phenomenon and the level of reliability depends on the circumstances of test administration. The level of reliability that is achieved under standardized and carefully controlled test administration conditions can be easily eroded if the test is administered under haphazard, nonstandardized conditions.

In classical test theory, an observed score is assumed to be made up of two independent components: a true score, which is a score that would be obtained if the measurement process could be repeated and infinite number of times, and an error of measurement, which is assumed to be random and unrelated to other variables, such as the true score or scores obtained from other measures. Under the classical test theory assumptions, reliability is simply the ratio of the true score variance to the observed score variance and the standard error of measurement is simply the standard deviation of the errors of measurement.

Generalizability theory (see Brennan) can be viewed as an extension of the classical test theory that allows for the differentiation of multiple sources of measurement error

(Feldt and Brennan, 1989; Haertel, 2006). Generalizability theory provides a structure for identifying different components of measurement error (e.g., error due to item sampling, to raters, or to two-way or three-way interactions among persons, items, and raters). Although generalizability coefficients are used to summarize the overall outcome of a generalizability analysis, the outcomes of primary interest are the components of variance due to various sources (e.g., persons, items, raters, and interactions).

Item Response Theory

IRT (see van der Linden) includes a variety of statistical models of person responses to test items (Yen and Fitzpatrick, 2006). Although it did not come into widespread use in educational testing until the 1980s, when computer programs made it feasible to apply IRT models to large-scale testing programs, IRT has become the statistical approach of choice in many situations. In recent years, IRT models have been employed in tests used in making decisions about individuals, in national and international assessments, and in large-scale accountability testing programs. The popularity of IRT stems in large part from its utility for addressing many measurement problems, such as test design, equating, and estimation of measurement error that is conditional on the proficiency of a test taker. IRT models also have appeal because they provide a means of solving some problems that are intractable under the confines of classical test theory (e.g., the selection of items and scoring of item responses for a computerized adaptive test).

Two IRT models, the Rasch model and the three-parameter logistic model, are in widespread use for dichotomous test items. Under the Rasch model, the probability that a person will answer an item correctly is modeled with a single item parameter (the item’s location or difficulty) and a single person parameter (the person’s ability or proficiency). As the name suggests, the probability of a correct response is modeled with three item parameters (item location of difficulty, item discrimination, and the lower asymptote). In addition, there is a person parameter corresponding to the person’s ability or level of proficiency. Under either the Rasch or three-parameter logistic model the person and item difficulty parameters are located on the same scale.

There are also several IRT models for items with multiple response categories. Partial credit models that have a single person parameter and as many item parameters as there are unique response categories. Thus, the partial credit model is a natural generalization of the Rasch model. More generalized models that treat categories as nominal or as graded and allow for item categories to differ in discrimination as well as location are also available (see, e.g., Yen and Fitzpatrick, 2006).

Testing Technology

Testing and technological developments have been closely linked at least since the development of the first electronic scoring machines. The optical scanners made it possible to score large numbers of answer sheets with great efficiency and made multiple-choice test items the dominant approach to educational testing in the United States and some other countries. However, the current ties between technology and testing run much deeper than simply test scoring. Today, technology plays a role in an ever-expanding number of aspects testing. Drasgow *et al.* (2006) paint a vivid picture of the ways technology can be integrated into test design, item generation, test assembly, test delivery, and scoring of both selected-response and constructed-response test items.

Computer-based testing (CBT) makes it possible to report scores immediately after a test is completed; it also makes it possible to administer the test adaptively so the selection of the next item of set of items to present to a test taker depends on the responses that are made to previously administered items or subsets of items. Computerized adaptive tests make it possible to reduce the length of a test while maintaining a fixed level of measurement precision or it makes it possible to terminate testing when a desired level of precision is achieved for a given decision. Moreover, CBT provides the opportunity to introduce simulations and other types of test problems that are not feasible with paper-and-pencil tests. Automated scoring of essays reduces the need for the assembly and training of human scorers and speeds up the scoring process so that results can be returned sooner than is possible with reliance on human scorers.

Scheduling flexibility of test administration on the Internet and the immediate return of results make CBT appealing as a means of getting interim assessment results that provide early warning for end-of-the-year accountability testing. These features, together with the ability to add items to an electronic item bank and select items for administration from that bank for administration at the time deemed appropriate by a teacher, also have the promise of facilitating formative assessment.

Conclusion

Educational measurement is a critical component of education. Educational tests serve a variety of functions, including their use in making educational decisions about students, monitoring student achievement, schools and system accountability, and formative decision making by classroom teachers. Test theory provides a solid foundation for

the development and evaluation of tests and the uses of test results. Technological advances offer the opportunity to improve the quality and efficiency of educational testing. It remains to be seen, however, whether the improved technology will lead to more effective use of educational test results.

Bibliography

- AERA, APA, and NCME (American Educational Research Association, American Psychological Association, and the National Council for Measurement in Education) (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice* 5(1), 7–74.
- Council of Chief State School Officers (no date). *Formative Assessment and CCSSO: A Special Initiative – a Special Opportunity*. Washington, DC: CCSSO.
- Cronbach, L. J. (1988). Five perspectives on validation argument. In Wainer, H. and Braun, H. (eds.) *Test Validity*, pp 3–17. Hillsdale, NJ: Erlbaum.
- Drasgow, F., Luecht, R. M., and Bennett, R. E. (2006). Technology and testing. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 471–515. Westport, CT: ACE/Praeger Series on Higher Education.
- Feldt, L. S. and Brennan, R. L. (1989). Reliability. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn., pp 1035–146. New York: Macmillan.
- Haertel, E. H. (2006). Reliability. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 65–110. Westport, CT: ACE/Praeger Series on Higher Education.
- Jones, L. (2004). Chronology, 1963–2003. In Jones, L. V. and Olkin, I. (eds.) *The Nation's Report Card: Evolution and Perspectives*, pp 11–21. Bloomington, IN: Phi Delta Kappa Education Foundation.
- Kane, M. T. (2006). Validation. In Brennan, R. L. (ed.) *Educational measurement, Fourth Edition*, pp 17–64. Westport, CT: ACE/Praeger Series on Higher Education.
- McDonnell, L. M. (2004). *Politics, Persuasion, and Educational Testing*. Cambridge, MA: Harvard University Press.
- Messick, S. (1989). Validity. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn., pp 13–103. New York: Macmillan.
- Rice, J. M. (1897). The futility of the spelling grid – I. *Forum* 23, 163–172.
- Sadler, R. (1989). Formative assessment and the design of instructional assessments. *Instructional science* 18, 119–144.
- Shepard, L. A. (2006). Classroom assessment. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 623–646. Westport, CT: ACE/Praeger Series on Higher Education.
- Yen, W. M. and Fitzpatrick, A. R. (2006). Item response theory. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 111–153. Westport, CT: ACE/Praeger Series on Higher Education.

Relevant Websites

- <http://www.iea.nl> – IEA: Mission Statement.
- <http://nces.ed.gov> – Program for International Student Assessment (PISA): Frequently Asked Questions, National Center for Education Statistics (NCES).
- <http://www.timss.bc.edu> – The TIMSS and PIRLS International Study Center.

Equating and Scaling

A A von Davier, Educational Testing Service, Princeton, NJ, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The term linking has slightly different meanings in the field of educational assessments and it is used as (1) a general term for denoting a relationship between test forms (at the total score level, at the item parameter level, etc.); (2) as a weaker form of equating; and (3) as a synonym to the process of placing item response theory (IRT) item parameters on the same scale, which sometimes is also called IRT calibration. ("Calibration refers to the process of determining the relation between the output (or response) of a measuring instrument and the value of the input quantity or attribute, a measurement standard. In non-specialized use, calibration is often regarded as including the process of adjusting the output or indication on a measurement instrument to agree with value of the applied standard, within a specified accuracy."— www.wikipedia.org). In this article, we refer to equating as a subclass of linking methods (Holland and Dorans, 2006) and we refer to scaling as a common subsequent procedural step to equating (and to some other classes of linking, such as, for example, vertical linking), where scaling denotes the mapping of the already equated scores onto a reporting scale to aid in communications with test takers and test users. The framework and definitions we use here are common in the literature and aid the focus of the article (see Holland and Dorans (2006) and Kolen and Brennan (2004) for an extensive view of categories of linking methods and for other uses of the word scaling).

Here, we give an overview of the existing methodologies for comparing skills and establishing a common scale in educational assessment programs when the scores are reported at the individual level. Next, equating and scaling methodologies are briefly introduced. In the following section, observed-score and IRT equating methods are presented in detail.

An equating process has basically two components: a data collection design and a statistical method for achieving the desired relationship between two or more of the test forms.

In all equating applications, the outcome variable (i.e., the ability measured by the test forms) must be comparable across time points (i.e., each scale point of the measure must retain an identical meaning over time). The outcome variable must also remain construct-valid for at least the entire period of observation and maybe beyond, if prediction of future success is part of the study. These two requirements speak to the necessity of establishing a common scale for the test forms we use (see Harris *et al.*, 2004; Patz *et al.*, 2003).

Equating

The need for test equating arises when there are two or more test forms that measure the same construct and that can yield different scores for the same examinee. The most common example involves multiple forms of a test within a testing program, as opposed to a single testing instrument. In a testing program, different test forms that are similar in content and format typically contain completely different test questions. As the tests may contain different questions, the tests can vary in difficulty depending on the degree of control available in the test development process. Examinees tested with the more difficult test form will get lower scores than they would get had they been tested with the easier form. As testing programs often require comparability of the scores produced on these different forms, test-equating techniques were developed to adjust for these differences in test difficulty across test forms. The goal of test equating is to allow the scores on different forms of the same test to be used and interpreted interchangeably. Test equating requires some type of control for differential examinee ability, or proficiency, in the assessment of, and adjustment for, differential test difficulty; the differences in abilities are controlled by employing an appropriate data collection design (see next section for more details).

Not all assessments need an equating plan. This decision depends on the use of the tests. Some testing programs give one test a year, and students are ranked on that test for some purpose such as college admissions. If the test is used only to find the top performing students on it for that year, and the scores from one year are not ever compared to those from previous years, then equating the forms given annually may be a unnecessary (and expensive). However, if several test forms are used throughout the testing year for a common purpose and it is important that differences between the relative difficulties of the different test forms not affect the assessment of students taking the different forms, then test equating is necessary.

Equating is the most stringent form of score linking because it claims score interchangeability, not merely comparability, as do concordances and predictions (see Holland and Dorans, 2006, for more details and definitions of types of score linking). Other types of score linking may use the same computations as test equating but do not result in scores that are interchangeable. A linking typically does not qualify as an equating if the test forms are not constructed to the same specifications

or if the test forms measure different constructs. Test equating places several stringent requirements on the content and statistical properties of the test forms and on the samples of test takers involved, and is vulnerable to deviations from these requirements. These deviations may result in scores that are not interchangeable. In these circumstances, the intended test equating simply becomes weak, a form of test linking and the lack of interchangeability of scores can lead to unintended unfairness to some test takers.

So far, no systematic theory of test equating has been laid out. Over the years, methods have been developed in response to the need to create comparable test scores in practical circumstances. In order to evaluate these methods, Dorans and Holland (2000), Holland and Dorans (2006), Kolen and Brennan (2004), and Lord (1950, 1980) have laid out a framework that defines a good equating procedure. This framework is based on the following five requirements on the test forms and on the equating functions: the test forms should measure the same construct, and have equal reliability; and the function used for transforming the scores should be symmetric (i.e., the function for equating the new form, X , to the old form, Y , should be the inverse of the function equating Y to X), population invariant (i.e., the use of different subpopulations should yield the same equating function), and should lead to equity, that is, it should be a matter of indifference to an examinee to be tested by either one of two test forms that have been equated (Holland and Dorans, 2006). Many of these requirements are vague or arguable. In most situations, a failure of any of these requirements is hard to detect using available data. The combination of the lack of a theory and difficulties in detecting whether the equating requirements have been violated in practical settings creates a challenging situation for the practitioner.

Some problems in the equating process also appear when the data requirements for an equating method are not fulfilled; for example, when the samples available for equating are too small, and when large ability differences exist in the groups that take the two test forms to be equated, or when differences in subgroups lead to a linking function that is subgroup dependent (see Dorans and Holland, 2000). In an attempt to address these issues, researchers have developed new strategies to cope with design and data difficulties (equating with small samples, new approaches to anchor test construction, and new equating models) (von Davier, in press).

Scaling

Depending on how a test is scored – number correct, formula-scored, IRT scored – a raw score of an examinee on a test will look and be different. In order to aid the interpretability of the scores when provided to test users and test takers, the raw scores are transformed to scale scores. The scale scores of a test are what are reported to the test users, and therefore, they are the most visible, and

important part of an assessment. The reported scores provide the basis for decisions about whether one passes or fails a test, for comparisons across individuals, across institutions, or for policy decisions. Petersen *et al.* (1989) said that the “usefulness of a primary score scale depends in its fulfilling two important goals: facilitating meaningful inferences and minimizing misinterpretation and unwarranted inferences.” For more details about the scaling process see Kolen and Brennan (2004) or Dorans (2004).

Typically, scaling is established by mapping raw scores from a single test form (administered to a sample that is representative of the population of test takers for which the test is intended) to scale scores. Establishing the scale for reporting scores is a process that is both statistically and policy based and it should support the purpose of the assessment. The reporting scale should (1) have an established mean and variance; (2) allow for a good representation of easier or more difficult subsequent test forms; and (3) avoid (misleading) comparisons with different and already-established assessments, and it might incorporate score precision (such as reflecting a special relationship of the standard error of measurement across the score points, or deciding about the number of score scale points). This mapping, called scaling, can be linear or curvilinear. It is common to talk about equating and scaling as a two-step process. In practice, the scaling of the scores from a new test form is accomplished as follows: raw scores on the new test form are equated back to the raw scores of the previous (old) form for which the scaling has already been established.

Other approaches to establishing reporting scale or scale scores incorporate information about the content of the test, expert knowledge about the claims made about the items and test, and the intended message to the test users. Some of the most familiar approaches that consider the content of the test are: standard setting, scale anchoring, and item mapping, which are described in detail in the literature by Livingston and Zieky (1982), Allen *et al.* (1999), and Beaton and Allen (1992).

The rest of the article focuses on the observed-score equating methods and on the IRT calibration and equating methods. In the class of observed-score equating and linking methods, the traditional methods – equipercentile and linear – and a newly developed method – kernel equating (KE) – are discussed. The observed-score and IRT-based equating methods are the two major classes of statistical methods used for equating and linking tests in educational assessments. Each of them is followed by the scaling process, that is, placing the linked raw-scores onto a reporting scale.

Observed-Score Equating Methods

In this section, we introduce the notation and lay out a framework for a discussion of observed-score equating (see also Kolen and Brennan, 2004 for a related discussion).

There are two test forms to be equated, X and Y , and a target population, T , on which this is to be done. The data are collected in such a way that the differences in the difficulty of the test forms and the differences in the ability of the test takers that take the two forms are not confounded. There are two classes of data collection designs for equating: (1) designs that allow for common people (equivalent groups, single group, and counterbalanced designs) from a single target population of examinees T (see Livingston, 2004, for a slightly different view and definition of a target population); and (2) designs that allow for common items (the nonequivalent groups with an anchor test design or NEAT design, also referred to as the common item or anchor test design) where the tests, X and Y , are given to two samples from two test populations (administrations), P and Q , respectively, and a set of common items, the anchor test, is given to samples from both these populations. As the name implies, in a NEAT design, the samples from P and Q are not assumed to be of equivalent ability. The target population, T for the NEAT design, is assumed to be a weighted average of P and Q where P and Q are given weights that sum to 1. This is denoted by $T = wP + (1 - w)Q$.

Many observed-score equating methods are based on the equipercentile equating function. It is defined on the target population, T , as $e_{Y:T}(x) = G_T^{-1}(F_T(x))$, where $F_T(x)$ and $G_T(y)$ are the cumulative distribution functions (cdfs), of X and Y , respectively, on T . In order for this definition to make sense, and to insure that the inverse equating function exists, it is also assumed that $F_T(x)$ and $G_T(y)$ are strictly increasing and have been made continuous or continuized.

Several important observed-score equating methods may be viewed as only differing in the way the continuization is achieved. The traditional equipercentile equating method (percentile rank method) uses linear interpolation of the discrete distribution to make it piecewise linear and therefore continuous. The KE (von Davier *et al.*, 2004) method uses Gaussian kernel smoothing to approximate the discrete histogram by a continuous density function.

Equipercentile equating leads to linear equating if one assumes that $F_T(x)$ and $G_T(y)$ are continuous and have the same shape while differing in mean and variance. The linear equating function is defined by $\text{Lin}_{Y:T}(x) = \mu_{YT} + \sigma_{YT}((x - \mu_{XT}) / \sigma_{XT})$, where μ_{XT} , μ_{YT} , σ_{XT} , and σ_{YT} are the means and standard deviations of X and Y on T , respectively.

In von Davier *et al.* (2004), it is shown that any equipercentile equating function can be decomposed into the corresponding linear equating function and a nonlinear part.

The Kernel Method of Test Equating

The KE method (von Davier *et al.*, 2004; Holland and Thayer, 1989) is an observed-score equating procedure that unifies several observed-score procedures of test

equating into a single method while, at the same time, providing new statistical information that can be used in the practice of test equating. KE brings the steps of an equating process into an organized whole rather than treating them as disparate parts. KE exploits data pre-smoothing techniques by fitting loglinear models to score data and incorporates the error introduced by the pre-smoothing procedure into the standard errors. KE provides new tools for comparing two or more equating functions and for rationally choosing between them based on newly introduced indices.

KE is an equipercentile equating procedure in which the score distributions to be equated are converted from discrete distributions to continuous distributions by using a normal (Gaussian) kernel as opposed to using linear interpolation, as is done in the traditional equipercentile equating method. By varying the bandwidth values of the Gaussian kernel (see von Davier *et al.*, 2004), KE can approximate the traditional equipercentile and linear equating methods. The process of choosing the optimal bandwidth is fully automatic and involves the minimization of a penalty function. When the bandwidths used are 100 times the standard deviation of the scores or larger (i.e., large bandwidths), the continuized distributions will be nearly normal, in which case the KE functions can be regarded as approximately linear. Thus, linear equating can be regarded as special case of equipercentile equating in the framework of KE. The KE framework also introduces the percent relative error (PRE) that aids the diagnosis of the equating function and introduces the standard error of the difference between two equating functions (the SEED). The SEED can help in making a linear or nonlinear decision.

IRT Linking and Equating

If IRT is used in the equating of the test scores, it is necessary to first use some sort of IRT linking or calibration method to place the IRT parameter estimates on the same scale (see Kolen and Brennan, 2004; von Davier and von Davier, 2004). All IRT models have an intrinsic lack of identifiability of the item and person parameters that is usually addressed by imposing some restrictions on the item or person parameters (there are various ways to implement restrictions, and consequently different software use different approaches). This leads to a need for placing the item parameters on the same scale even in the absence of equating. In other words, even if the same test instrument would have been given to two equivalent groups and if the estimation of the IRT parameters was done separately in the two groups then, the item parameters still need to be placed on the same scale by some sort of linear transformation. This process is usually called

IRT linking, but it is also called IRT scaling or IRT calibration in the literature.

Hence, for equating purposes in a NEAT design and assuming that the IRT model used fits the data well, the two separate parameter estimates of the item parameters from anchor items in the two groups need first to be placed on the same scale.

Once this step is accomplished, then additional methods, such as IRT true- or observed-score equating (Kolen and Brennan, 2004), can be undertaken to establish a relationship at the (raw) score level. Then a third step is employed – scaling – which refers to placing the raw scores onto some reporting scale, as was described in the previous section. However, in many settings the second step is skipped altogether – see also the brief discussion in next subsection.

In this section, we describe the IRT linking or calibration procedures (i.e., procedures for placing the item parameters estimates on the same scale) used for data collection designs that involve common items (NEAT design) since this is the most common data collection design. Then, the IRT true score equating is briefly described.

IRT Linking Methods for the NEAT Design

Here, we describe the usual IRT linking or calibration methods: mean–sigma and mean–mean, concurrent calibration, fixed parameters calibration, the Stocking and Lord characteristic curves approach, and the Haebara characteristic curves approach (see Kolen and Brennan 2004: ch. 6, for a detailed description of these methods).

IRT models express the probability of a response z_{ni} of a given person, n ($n = 1, \dots, N$), to a given item, i ($i = 1, \dots, I$), as a function of the person's competency or ability (which is a latent variable), denoted θ_n and a possibly vector valued item parameter, β_i , that is,

$$P_{ni} = P(X = z_{ni}) = f(z_{ni} | \theta_n, \beta_i)$$

In the case of the well-known three-parameter logistic model (3PL) model (Lord and Novick, 1968) that is used to fit data from dichotomous items, the item parameter vector is three dimensional. Its dimensions are the slope or discrimination that is usually denoted by a , the difficulty b , and the guessing parameter c , respectively, that is, $\beta_i' = (a_i, b_i, c_i)$. However, most results presented here do not depend on the specific choice of the model and apply to models for both dichotomous and polytomous data.

In the NEAT design, X is not observed in population Q , and Y is not observed in population P . To overcome this feature of the NEAT design, all equating methods developed for the NEAT design (both observed-score and IRT methods) must make additional assumptions of a type that does not arise in the other designs. IRT calibration makes the following assumptions.

Estimation Methods

The tests to be equated, X and Y , and the anchor, V , are all unidimensional (i.e., all items measure the same unidimensional construct), carefully constructed tests, in which the local independence assumption holds at the item level (Hambleton *et al.*, 1991); and the chosen IRT model fits the data well.

The methods available for the estimation and calibration purposes are mentioned below. Operationally, the estimates of the item parameters are usually obtained using the marginal maximum likelihood (MML) estimation method for IRT models (Bock and Aitkin, 1981). In the past, the joint maximum likelihood (JML) method has been used as well. Conditional maximum likelihood (CML) methods are used for 1PL IRT models (1PL, Rasch model; see also von Davier, 2001) or for 1PL models with fixed slopes, such as the one-parameter logistic model (OPLM; Verhelst and Glas, 1995), and in some recent developments, Bayesian estimations methods have been used in the framework of Markov chain Monte Carlo (MCMC) estimation (see Patz and Junker, 1999).

Concurrent calibration

The item parameters from X , V (in both populations), and Y can be estimated jointly, in a single estimation run, assuming that the items in V are the same for both populations and coding the items that an examinee did not take as not administered, since these outcomes were unobserved and are missing by design.

Fixed parameters calibration

In this method, common items whose parameters are known (e.g., from a previous administration calibration or a separate calibration) are anchored or fixed to their known estimates during calibration of a particular form of the test. By treating the common item parameters as known and, therefore, not reestimating them, the remaining item parameters from the not common set of items in the two test forms are forced onto the same scale as the items with fixed parameters. This calibration procedure is more restrictive than concurrent calibration.

Mean–mean and mean–var IRT calibrations

If an IRT model fits the data, any linear transformation (with slope A and intercept B) of the theta scale also fits these data, provided that the item parameters are also transformed in the same way (see, e.g., Kolen and Brennan, 2004: ch. 6). The most straightforward way to transform scales when the parameters were estimated separately is to use the means and standard deviations of the item parameter estimates of the common items for computing the slope and the intercept of the linear transformation. Loyd and Hoover (1980) described the mean–mean method and Marco (1977) described the mean–var IRT

calibration (Marco, 1977). Both methods are seldom used in practice and tend to be inadequate if the groups taking the test forms differ in ability.

Stocking–Lord and Haebara linkings

Characteristic curve transformation methods were proposed (Haebara, 1980; Stocking and Lord, 1983) to avoid some issues related to the mean–mean and mean–var approaches, such as the fact that various combinations of the item parameter estimates produce almost identical item characteristic curves over the range of ability at which most examinees score.

The Stocking and Lord IRT method finds parameters for the linear transformation of the item parameters estimates from the anchor set in one population (say Q) that matches the estimated test characteristic function of the anchor set in the reference population (say P). Haebara (1980) expressed the differences between the estimated characteristic curves as the sum of the squared differences between the item characteristic functions for each item over the common items for examinees of a particular ability θ_r . The Haebara method is more restrictive than the Stocking and Lord method because the restrictions take place at the item level (i.e., for each item from the set of common items), while the Stocking and Lord approach poses a global restriction at the anchor test level.

von Davier and von Davier (2004) proposed a new perspective on IRT linking by viewing any linking function or calibration method as a restriction function on the joint log-likelihood function based on all the data. Rewriting any linking as a restriction function and estimating the model parameters under this restriction implies a greater degree of flexibility in the IRT calibration process.

Equating of Scores in the IRT Framework

If equating of scores is desired, different programs and testing companies use different techniques after the calibration process. In some programs, an equating step is performed, where the so-called IRT true-score or observed-score equating methods are used to establish a relationship between the raw scores on the two forms. After that, a third step, scaling, is employed, which refers to placing the raw scores onto some reporting scale, as discussed previously. In other programs, the second step is skipped altogether, and the ability scale is linked directly to the reporting scale (Yen, 1984). This method is used in scoring tests of state assessments and published tests, such as TerraNova (CTB/McGraw-Hill, 2000). Next the IRT true-score equating method is briefly described.

IRT true score equating

A common IRT-based equating process can be described as follows: in a NEAT design with tests with multiple-choice items, an IRT model (e.g., the 3PL model) is

separately fitted to the data from the two populations P and Q . Then, the characteristic curve method (Stocking and Lord, 1983) is used to place the separately estimated item parameters onto a common scale. After all the item and person parameters are on the same scale, they are used to estimate the true-score equating function. The number-correct true score for a given θ is obtained by summing the (conditional) probabilities over the number of items in the test. Then the true-score equating method is used to obtain equivalent scores on X and Y (see Kolen and Brennan, 2004; Petersen *et al.*, 1989; von Davier and Wilson, 2005). According to this method, for a given true score of the new test form X , one finds (via iterative procedures, such as the Newton–Raphson method) the value of competency/ability θ and then computes the true score on the old form Y .

The IRT true score equating requires that many assumptions are met and relies on some *ad hoc* decisions: it requires that the tests are number-right scored, which involves an implicit assumption that there are no omits (see Kolen and Brennan, 2004); if the tests are formula-scored, then some sort of transformation is necessary and usually this transformation will treat the omits as wrong. The IRT true score equating introduces one more assumption: The relationship between the true scores holds also for the observed scores. This assumption has not been theoretically proven, but was confirmed in research studies (Lord and Wingersky, 1984). In IRT true score equating that uses the common 3PL IRT model, the lowest possible true score is the sum of the c_j , the so-called guessing parameters, and not 0. In this case, there is some arbitrariness of the results of the conversion of the observed scores that are outside the range of possible true scores on the new form X (see Kolen and Brennan, 2004, for details).

Discussion

This article reviews the existing methodologies for equating tests that measure the same construct, where interchangeability of the scores is desired.

Nowadays, when more and more standardized testing is used nationally and internationally, we are discovering new challenges in ensuring that the testing process and its results are fair and accurate. For example, some of the challenges to test equating are the construction of the anchor sets, the choice of the reporting scale, and the characteristics of the samples used for establishing the scale, which ideally should be representative for the population of test takers. In addition, psychometricians worry about the maintenance of the reporting scale: how to introduce new forms, how to monitor the scale over time, and how to adjust to changes in the administration mode.

See also: Educational Measurement: Overview; Item Response Theory.

Bibliography

- Allen, N. L., Carlson, J. E., and Zelenak, C. A. (1999). *The NAEP 1996 Technical Report*. Washington, DC: National Center for Education Statistics.
- Beaton, A. E. and Allen, N. L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics* **17**(2), 191–204.
- Bock, R. D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **46**, 443–445.
- CTB/McGraw-Hill (2000). *TerraNova CAT*, 2nd edn. (Computer software). http://www.ctb.com/mktg/terranova/tn_intro.jsp (accessed August 2009).
- Dorans, N. J. (ed.) (2004). Using subpopulation invariance to assess test score equity. Special Issue: Assessing the Population Sensitivity of Equating Functions. *Journal of Educational Measurement* **41**(1), 1–68.
- Dorans, N. J. and Holland, P. W. (2000). Population invariance and the equitability of tests: Basic theory and the linear case. *Journal of Educational Measurement* **37**, 281–306.
- Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research* **22**, 144–149.
- Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Harris, J. D., Hendrickson, A. B., Tong, Y., Shin, S. -H., and Shyu, C. -Y. (2004). Vertical scales and the measurement of growth. *Paper Presented at the Annual Meeting of the National Council on Measurement in Education*. San Diego, CA.
- Holland, P. W. and Dorans, N. J. (2006). Linking and equating. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 189–220. Westport, CT: Praeger.
- Holland, P. W. and Thayer, D. T. (1989). *The Kernel Method of Equating Score Distributions*. (ETS RR-89-07) Princeton, NJ: ETS.
- Kolen, M. J. and Brennan, R. J. (2004). *Test Equating: Methods and Practices*, 2nd edn. New York: Springer.
- Livingston, S. A. (2004). *Introduction to Test Equating (without IRT)*. Princeton, NJ: ETS.
- Livingston, S. A. and Zieky, M. J. (1982). *Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*. Princeton, NJ: Educational Testing Service.
- Lord, F. M. (1950). *Notes on Comparable Scales for Test Scores* (Research Bulletin 5048). Princeton, NJ: Educational Testing Service.
- Lord, F. M. (1980). *Application of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum.
- Lord, F. M. and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- Lord, F. M. and Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score “equating.” *Applied Psychological Measurement* **8**, 452–461.
- Loyd, B. H. and Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement* **17**, 179–193.
- Marco, G. L. (1977). Item characteristic curves solutions to three intractable testing problems. *Journal of Educational Measurement* **14**, 139–160.
- Patz, R. and Junker, B. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of Educational and Behavioral Statistics* **2**, 146–178.
- Patz, R., Yao, L., Chia, M., Lewis, D., and Hoskens, M. (2003). Hierarchical and multidimensional models for vertical scaling. *Paper Presented at the Annual Meeting of the National Council on Measurement in Education*, Chicago, IL, USA.
- Petersen, N. S., Kolen, M. J., and Hoover, H. D. (1989). Scaling, norming, and equating. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn., pp 221–262. New York: Macmillan.
- Stocking, M. L. and Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement* **7**, 201–210.
- Verhelst, N. D. and Glas, C. A. W. (1995). The one parameter logistic model. In Fischer, G. H. and Molenaar, I. W. (eds.) *Rasch Models: Foundations, Recent Developments, and Application*, pp 215–238. New York: Springer.
- von Davier, A. A. (ed.) (in press). *Statistical Models for Equating, Scaling and Linking*. New York: Springer.
- von Davier, A. A., Holland, P. W., and Thayer, D. T. (2004). *The Kernel Method of Test Equating*. New York: Springer.
- von Davier, A. A. and Wilson, C. (2005). *A Didactic Approach to the Use of IRT True-Score Equating* (ETS RR-05-26). Princeton, NJ: ETS.
- von Davier, M. (2001). WinMira 32 Pro: A program system for analyses with a variety of discrete mixture distribution models (Computer software). Assessment System Corporation. <http://www.assess.com/Software/WINMIRA.htm> (accessed August 2009).
- von Davier, M. and von Davier, A. A. (2004). *A Unified Approach to IRT Scale Inking and Scale Transformations* (ETS RR-04-09). Princeton, NJ: ETS.
- Yen, W. M. (1984). Obtaining maximum likelihood trait estimates from number-correct scores for the three-parameter logistic model. *Journal of Educational Measurement* **21**(2), 93–111.

Fairness

C Gipps, University of Wolverhampton, Wolverhampton, UK

G Stobart, Institute of Education, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bias – A technical term used in test construction for when individuals with equal ability, in the subject being tested, but from different groups, do not have the same probability of success.

DIF (Differential Item Functioning) – DIF occurs when people in different groups perform in substantially different ways on a test question even though they may have very similar scores on a test.

Equality – An essentially quantitative approach to differences between groups.

Equity – The ‘spirit of justice’ which looks at the fairness of a given state of affairs. It is essentially a qualitative judgment.

Introduction

Fairness is a concept for which definitions are important, since it is often interpreted in narrow and technical terms. We set fairness in a social context and look at what this means in relation to different groups and cultures. Similarly, we are using educational assessment in a more inclusive way than is often the case; we include tests, examinations, teachers’ informal judgments, or evaluation (assessment in United Kingdom) of student performance. We then explore bias in measurement as well as the broader concept of equity – which goes beyond the more quantitative concerns of equality. Practical approaches to ensuring fairness are discussed in the subsequent section.

Fairness

How would we tell whether a test is fair for different groups (male/female; socially advantaged/disadvantaged; ethnic groupings)? The dilemma is that different groups will have different qualities and experiences, so fairness in assessment cannot assume equal scores or outcomes. Indeed, equal outcomes could represent bias in favor of a lower-achieving group.

Differences in performance may be due to differing access to learning, or there may appear to be differences as the test is biased in favor of one group. Wood (1987)

describes these different factors as the opportunity to acquire talent (access issues) and the opportunity to show talent to good effect (fairness in the assessment).

In our view, fairness in assessment cannot be considered in isolation from access issues in the curriculum, and the educational opportunities offered to the students: fairness in access opportunities and in what the curriculum offers provide the level playing field which must precede a genuinely fair assessment situation.

The approach we take includes these broader issues and, therefore, owes more to sociocultural theory than to measurement theory. Assessment is, after all, a socially embedded activity which can only be understood by taking account of the cultural, economic, and political contexts within which it operates (Sutherland, 1996; Gipps, 1999).

There is a historical legacy here. Examinations were introduced in order to encourage advancement through talent rather than patronage, an example being the Civil Service examinations in England in the 1850s (although the examination was only open to men, and furthermore only to those men who had had a particular type of schooling).

Of course, one can argue that public tests are important as a means of equalizing opportunities and as a necessary corrective to patronage, while at the same time understanding that tests may be biased in favor of one particular gender, social or ethnic group.

Educational measurement has traditionally presented itself as scientific and objective – providing a true picture of an individual’s performance (unless he or she cheats). It is, in this frame, seen essentially as a fair activity. If access to higher levels of education or to certain professions is determined by the open competition of examination, then this allows society to subscribe to promotion through merit, and unsuccessful individuals to accept their lot, since all have been allowed to compete on an equal basis, in that they took the same examination under similar conditions, to demonstrate their competence.

More recent analyses, however, show a rather different reality. Evidence in the 1950s began to emerge that non-verbal intelligence quotient (IQ) tests, which were viewed as culture free, were not independent of culture. A picture was built up of tests being biased toward the culture of the people who designed them. In the United States and the United Kingdom this meant, at the time, white, male, and middle class. Such tests are therefore unfair for people from other backgrounds, as these individuals do not have

the same chance to do well on the tasks. This was particularly important in relation to IQ tests because of the very powerful role of these tests in determining life chances. In the post-1965 Civil Rights legislation era in the United States, critics of advancement through testing pointed out that opportunities to acquire talent, through access to good schooling, were not equally distributed (Wood, 1987; Orfield and Kornhaber, 2001). In other words, these tests were biased in favor of the dominant social group, partly from the design of the tests, and partly from access issues.

In the case of school examinations and tests, the critique has essentially similar underpinnings. Social class is a significant determinant of school performance. Children from lower social groups are not less intelligent or less academically capable, but children from middle-class homes are better able to do well at school because of the correspondence of cultural factors between home and school.

We are now well aware that the form of assessment can differentially affect results for different groups. In England there has been far more analysis of this in relation to gender than to ethnicity. We know that during compulsory schooling (up to 16 years) girls are likely to outperform boys on tasks which involve open-ended writing, particularly when this involves personal response. Even within multiple-choice tests, traditionally seen as favoring boys, there are differential response patterns. Carlton (2000) has shown that in such tests females perform better than males, matched for ability, on questions in which the content is a narrative or is in a humanities field and when the content deals with human relationships. As the context of an item grows longer, the relative performance of females also improves. Males outperform females on questions relating to science, technical matters, sports, war, or diplomacy. We also know that where examinations have a coursework (or essay) element, the performance of girls is likely to be more consistent, although the effect this has on final grades in English-school-leaving examinations has often been overstated (Elwood, 1995).

We know less about other aspects of the form of assessment, particularly in relation to ethnicity. For example, oral assessment plays little part in the examination system in England outside examining languages. Does the emphasis on written response disadvantage groups who place more emphasis on oral communication in their culture?

Bias

Bias is a technical term used in test construction. In layman's terms a test is biased if "two individuals with equal ability (in the subject being tested) but from different groups do not have the same probability of success" (Shepard *et al.*, 1981).

A main cause of bias in a test would be that it was designed by one cultural group to reflect their own experience,

and thus disadvantage test-takers from other cultural groups, as in the example of IQ tests given above. Thus, bias may be due to the content matter in a test. However, it may also be due to lack of clarity in instructions which leads to differential responses from different groups, or to scoring systems that do not credit appropriate or correct responses that are more typical of one group than another.

The existence of group differences in performances on tests is often taken to imply that the tests are biased, the assumption being that one group is not inherently less able than the other. However, as we have argued, the two groups may well have been subject to different environmental experiences or unequal access to the curriculum. This difference will be reflected in group test scores, but the test is not strictly speaking biased.

There are technical approaches to evaluating bias, which we discuss below.

Equity

Equity is defined in the dictionary as moral justice. Equity does not imply equality of outcome and does not presume identical experiences for all – both of these are seen to be unrealistic, but it asserts that assessment practice and interpretation of results need to be fair and just for all groups.

For example, it is possible to have similar outcomes for two groups and yet to see this as unfair to one of them, which may have been disadvantaged in terms of access to the curriculum. Conversely, it is possible to have unequal group outcomes that may be seen as fair. An example would be where there are group differences in the application to learning and preparation, where each had similar resources and opportunities.

Equity is also a quasi-legal term. The legal meaning of equity is the spirit of justice and building on the work of Secada (1989), we see it as a qualitative concern for what is just. "Equity attempts to look at the justice of a given state of affairs, a justice that goes beyond acting in agreed upon ways and seeks to look at the justice of the arrangements leading up to and resulting from those actions" (p. 81).

The implication is that equity is not the same as equality. Equity represents the judgment about whether equality, be it in the form of opportunity and/or of outcomes, achieves just (fair) results.

Looking for equality requires essentially a quantitative approach to differences between groups, while equity goes beyond this and looks at the justice of the arrangements prior to the testing/examination.

Legal Challenges

Sitting behind these sociocultural and technical approaches to fairness are legal challenges to assessment programs and results. Such legal challenges are more common in the

United States and Australia than in the United Kingdom. According to Cumming and Dickson:

Sources of legal action around the world have different bases. For example, in the USA many education cases have been fought and won on constitutional grounds and individual rights. In countries such as Australia, where individual rights do not have a constitutional basis, legal challenges in this area are usually brought through tort, or negligence, law, or on a statutory basis, under anti-discrimination laws. (Cumming and Dickson, 2006: 4)

Simple examples illustrate how discrimination might arise in an assessment context. Direct discrimination in the administering of a test might happen, for example, if a marker marks a student of a particular race, sex, religion 'harder' than students not of that race, sex or religion. Indirect discrimination might arise in the administering of a test if there were, for example, a requirement that students complete the test in a set time. Students with a disability may not be able to comply with this requirement – they may not be able to write quickly, or they may have a processing disorder. (Cumming and Dickson, 2006: 5)

Legal challenges have been made on the grounds of: opportunity to learn – not having been taught the syllabus being assessed; certification decisions made on the basis of a single instrument (which violates the fairness principle of using a variety of assessment approaches); lack of funding – leading to low performance on national performance measures; incorrect scoring of tests; and incorrect (or absent) assessment of students with learning difficulties resulting in inappropriate placement and/or teaching.

A review of legal cases in the United States indicates the high standard required to establish lack of equity in assessment. These authors similarly identify lack of opportunity to learn and lack of opportunity to demonstrate it to good effect – for various reasons, including insufficient preparation time for a new examination or test – as major themes in legal challenge, and conclude:

In most cases, in hindsight, the matters challenged are preventable. Equitable access to the curriculum to be

assessed, provision of adequate resources and implementation of appropriate assessment regimes are surely public expectations of education.

Approaches to Ensuring Fairness

What we have outlined so far is a complex and contested picture, difficult to engage and move forward with. Yet, much work has been done to address this, which is both sociocultural and technical.

An Equity Approach

Apple (1989) argued that attention in the equity and education debate must be refocused on important curricular questions, to which Gipps and Murphy (1994) linked assessment questions (see **Table 1**). Based on accounts of the difficulties of providing effective education to all in countries such as South Africa (Meier, 2000) and Kenya (Mwachihi and Mbithi, 2000), Stobart has added access and resource questions (Stobart, 2005).

Some aspects of this approach are now reflected in the guidelines for test developers.

Test development approach

The most fully articulated approach comes from test developers. The Standards for Educational and Psychological Testing (1999), which has guided later documents, was published jointly by AERA/APA/NCME (1999) to guide test development, scoring, and administration. It has a whole section on fairness in testing and test use, dealing with many of the issues in this article. Educational Testing Service (ETS), the major American test development agency, has its own standards for quality and fairness (ETS, 2000) to guide test developers. There are eight standards for fairness aimed at ensuring that “construct – irrelevant (i.e. to the skill or concepts being addressed) personal characteristics of test takers have no appreciable effect on test results or their interpretation” (p. 17).

Table 1 Access, curriculum, and assessment questions in relation to equity

<i>Access questions</i>	<i>Curricular questions</i>	<i>Assessment questions</i>
Who gets taught and by whom?	Whose knowledge is taught?	What knowledge is assessed and equated with achievement?
Are there differences in the resources available for different groups?	Why is it taught in a particular way to this particular group?	Are the form, content, and mode of assessment appropriate for different groups and individuals?
What is incorporated from the cultures of those attending? (Stobart, 2005)	How do we enable the histories and cultures of people of color, and of women, to be taught in responsible and responsive ways? (Apple, 1989)	Is this range of cultural knowledge reflected in definitions of achievement? How does cultural knowledge mediate individuals' responses to assessment in ways which alter the construct being assessed? (Gipps and Murphy, 1994)

Table 2 The ETS International Principles for Fairness Review of Assessments (2004)

• Principle 1:	Treat people with respect in test materials
• Principle 2:	Minimize the effects of construct-irrelevant knowledge or skills
• Principle 3:	Avoid material that is unnecessarily controversial, inflammatory, offensive, or upsetting
• Principle 4:	Use appropriate terminology to refer to people
• Principle 5:	Avoid stereotypes
• Principle 6:	Represent diversity in depictions of people

From ETS (2004). *ETS International Principles for Fairness Review of Assessments*. Princeton, NJ: Educational Testing Service.

These have been extended to the ETS International Principles for Fairness Review of Assessments (ETS, 2004). The six principles are presented in **Table 2**.

The requirement is to select assessment content that accurately reflects the construct, even if it produces gender/ethnic group differences, and to avoid content that is not relevant to the construct and could affect such differences. The involvement of those with a minority background is crucial here.

The document outlines the procedures for fairness review, and other fairness actions, one of which is the use of a statistical measure:

A statistical measure related to fairness should be used, whenever sample sizes permit, as an empirical check on the fairness of questions. Statistical measures based on the way matched people in different groups perform on each test question, called differential item functioning or DIF, are preferred. DIF occurs when people in different groups perform in substantially different ways on a test question, even though they have very similar scores on the test. If DIF data are available, tests should be assembled following rules that keep DIF low. (ETS, 2004: 11)

Fifteen years ago, the emphasis was solely on a statistical approach to eliminating bias in test design. The shift in emphasis from statistical manipulation as a major approach to it being just one element of test review is notable, and we would argue, to be welcomed.

An important approach to offering fairness, it is now argued, is to use, within any assessment program, a range of assessment tasks involving a variety of contexts – a range of modes within the assessment and a range of response format and style. This broadening of approach is most likely to offer pupils alternative opportunities to demonstrate achievement if they are disadvantaged by any one particular assessment in the program (Linn, 1992):

Also, if we wish pupils to do well in tests/exams we need to think about assessment which elicits an individual's best performance (after Nuttall, 1987). This involves

tasks that are concrete and within the experience or the pupil (an equal access issue) presented clearly (the pupil must understand what is required of her if she is to perform well) relevant to the current concerns of the pupil (to engender motivation and engagement) and in conditions that are not threatening (to reduce stress and enhance performance). (Gipps, 1994)

Large-Scale Assessment Programs

There are four key areas within large-scale testing/examination systems in which to raise issues of fairness, particularly in relation to multicultural societies. These are:

- The nature and requirements of the assessment system itself, for example, how are cultural and linguistic diversity approached? (For example, in which languages are examinations conducted?)
- How do the assessment methods meet the cultural diversity of the candidates?
- How does the content of the assessment reflect the experiences of different groups? (For example, whose version of history is tested?)
- How effectively is the performance of different groups monitored and how is this fed back into the system? (Stobart, 2005)

Teachers' Informal Evaluations/Assessments

Fairness in assessment in the informal setting of the classroom can be both more difficult – because there are many complex issues for the teacher to consider – and more possible – since a range of assessment approaches is possible.

It is more feasible for the teacher to offer, in the informal assessment setting, a range of assessment tasks and modes, and an approach which supports fairness as we argued above. It is also more feasible to provide the situation that can elicit an individual's best performance, since it is under the teacher's control.

However, teachers' informal assessment is, to a certain extent justifiably, perceived as being unreliable and biased (Harlen, 2004a EPPI-Centre). This is usually to do with lack of clarity, and variability, in standards or criteria. It is possible to improve the consistency of teachers' assessments through: providing clear criteria, training teachers to assess against these, and supporting the process with moderation of judgments via discussion (ARG, 2006).

In relation to the curriculum offered and opportunity to learn, there is another inconvenient fact: teacher expectation can affect the curriculum and learning experiences offered to children. There is clear evidence that teachers offer a different curriculum to children for whom they

hold low and high expectations (Tizard *et al.*, 1988; Troman, 1988; Harlen, 2004b). This is pertinent to the equal access issue.

Conclusion

Fairness is both essential and elusive. It is the appeal to fairness that has made educational measurement a pivotal part of most cultures. We have argued that different groups being allowed to sit, and be judged by, the same test is a simplistic view. Equality of opportunity includes access to similar resources and curricular opportunities. The more familiar, and narrower, discussion of bias is only a small part of this.

We will never achieve fair assessment, but we can make it fairer: The best defense against inequitable assessment is openness. Openness about design, constructs, and scoring will bring out into the open the values and biases of the test design process, offer an opportunity for debate about cultural and social influences, and open up the relationship between assessor and learner. These developments are possible, but they do require political will.

Bibliography

- AERA/APA/NCME (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Apple, M. W. (1989). How equality has been redefined in the conservative restoration. In Secada, W. (ed.) *Equity and Education*, pp 7–35. New York: Falmer.
- ARG (2006). The role of teachers in the assessment of learning. Assessment reform group pamphlet. www.assessment-reform-group.org (accessed August 2009).
- Carlton, S. T. (2000). Contextual factors in group differences in assessment. *Paper Presented at 26th IAEA Conference*, Jerusalem, 14–19 May.
- Cumming, J. and Dickson, E. A. (2006). Equity in Australia: The legal perspective, *Paper presented at 32nd IAEA Conference*, Singapore, 21–25 May.
- Elwood, J. (1995). Undermining gender stereotypes: Examination and coursework performance in the UK at 16. *Assessment in Education* 2(3), 283–303.
- ETS (2000). *ETS Standards for Quality and Fairness*. Princeton, NJ: Educational Testing Service.
- ETS (2004). *ETS International Principles for Fairness Review of Assessments*. Princeton, NJ: Educational Testing Service.
- Gipps, C. (1994). *Beyond Testing, Towards a Theory of Educational Assessment*. London: Falmer.
- Gipps, C. (1999). Sociocultural aspects of assessment. *Review of Research in Education* 24, 357–392.
- Gipps, C. and Murphy, P. (1994). *A Fair Test? Assessment, Achievement and Equity*. Buckingham: Open University Press.
- Harlen, W. (2004a). A systematic review of the evidence of reliability and validity of assessment teachers use for summative purposes (EPPI-Centre Review). In *Research Evidence in Education Library Issue 3*, London: EPPI-Centre, Social Science Research Unit, Institute of Education. http://eppi.ioe.ac.uk/EPPIWeb/home.aspx?page=/reel/review_groups/assessment/review_three.htm (accessed August 2009).
- Harlen, W. (2004b). A systematic review of the evidence of the impact on students, teachers and the curriculum of the process of using assessment by teachers for summative purposes (EPPI-Centre Review). In *Research Evidence in Education Library Issue 4*, London: EPPI Centre, Social Science Research Unit, Institute of Education Available on the website at: http://eppi.ioe.ac.uk/EPPIWeb/home.aspx?page=/reel/review_groups/assessment/review_four.htm (accessed August 2009).
- Linn, M. C. (1992). Gender differences in educational achievement. In Pfeleiderer, J. (ed.) *Sex Equity in Educational Opportunity, Achievement and Testing*, pp 11–50. Princeton, NJ: Educational Testing Service.
- Meier, C. (2000). The influence of educational opportunities on assessment results in a multicultural South Africa. *Paper Presented at the 26th IAEA Conference*, Jerusalem, 14–19 May.
- Mwachihi, J. M. and Mbithi, M. J. (2000). Assessment and equity assurance in the Kenyan multicultural background. *Paper Presented at 26th IAEA Conference*, Jerusalem, 14–19 May.
- Nuttall, D. (1987). The validity of assessments. *European Journal of Psychology of Education* 11(2), 109–118.
- Orfield, G. and Kornhaber, M. L. (2001). *Raising Standards or Raising Barriers?* New York: Century Foundation.
- Secada, W. G. (1989). Educational equity versus equality of education: An alternative conception. In Secada, W. G. (ed.) *Equity and Education*, pp 68–88. New York: Falmer.
- Shephard, L., Camilli, G., and Averill, M. (1981). Comparison of procedures for detecting test item bias with both internal and external ability criteria. *Journal of Educational Statistics* 6, 317–375.
- Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in Education* 12(3), 275–287.
- Sutherland, G. (1996). Assessment: Some historical perspectives. In Goldstein, H. and Lewis, T. (eds.) *Assessment: Problems, Developments and Statistical Issues*, pp 9–20. Chichester: Wiley.
- Tizard, B., Blatchford, P., Burke, J., Farquhar, C., and Plewis, I. (1988). *Young Children at School in the Inner City*. Hove: Erlbaum.
- Troman, G. (1988). Getting it right: Selection and setting in a 9–13 middle school. *British Journal of Sociology of Education* 9(4), 1988.
- Wood, R. (1987). *Measurement and Assessment in Education and Psychology*. London: Falmer.

Relevant Website

www.assessment-reform-group.org – The Assessment Reform Group (ARG): The role of teachers in the assessment of learning.

Generalizability Theory

R L Brennan, University of Iowa, Iowa City, IA, USA

© 2010 Elsevier Ltd. All rights reserved.

Generalizability theory offers an extensive conceptual framework and a powerful set of statistical procedures for addressing numerous measurement issues. To an extent, the theory can be viewed as an extension of classical test theory (see, e.g., Feldt and Brennan, 1989, and Haertel, 2006) through an application of certain analysis of variance (ANOVA) procedures to measurement issues. Classical theory postulates that an observed score can be decomposed into a true score, T , and a single undifferentiated random error term, E . As such, any single application of the classical test theory model cannot clearly differentiate among multiple sources of error. Generalizability theory liberalizes classical theory by employing ANOVA methods that allow an investigator to disentangle the multiple sources of error that contribute to the undifferentiated E in classical theory.

The defining treatment of generalizability theory is a monograph by Cronbach *et al.* (1972) entitled *The Dependability of Behavioral Measurements*. A history of the theory is provided by Brennan (1997). Cronbach (1991) described the theory as “a tapestry that interweaves ideas from at least two dozen authors” (p. 394). In particular, Burt (1936), Hoyt (1941), Ebel (1951), and Lindquist (1953, chapter 16) discussed ANOVA approaches to reliability. Indeed, the work by Burt and Lindquist appears to have anticipated the development of generalizability theory.

The essential features of univariate generalizability theory were largely completed with technical reports in 1960–61. These were revised into three journal articles, each with a different first author (Cronbach *et al.* 1963; Rajaratnam *et al.* 1965; Gleser *et al.* 1965). In the mid-1960s, the Cronbach team began their development of multivariate generalizability theory, which is incorporated in their 1972 monograph.

Since the Cronbach *et al.* (1972) monograph, a number of publications have explicated and, in some cases, extended the theory at various levels of detail and complexity. Brennan (1983, 1992a) provides a monograph on generalizability theory that is quite extensive but still less detailed than Cronbach *et al.* (1972). Shavelson and Webb (1991) provide a primer on generalizability theory. Shorter treatments are given by Brennan (1992b, 2001). Allal and Cardinet (1994), Crocker and Algina (1986), Feldt and Brennan (1989), Haertel (2006), Shavelson and Webb (1992), and Webb *et al.* (2007). Cardinet *et al.* (1981) propose some extensions to the theory. Various misconceptions are discussed by Brennan (2000). Brennan (2001a) provides the most extensive current treatment of

generalizability theory. Cardinet and Tourneur (1985) and Yang and Chang (2002) provide books on generalizability theory in French and Chinese, respectively.

Generalizability theory has broad applicability. It has been employed in a vast array of educational testing programs and to a wide range of other types of tests, as well. In addition, generalizability theory has been employed in numerous areas of research in education, business, and medicine.

Framework of Generalizability Theory

Classical test theory and ANOVA can be viewed as the parents of generalizability theory, but the child is both more and less than the simple conjunction of its parents. For example, although generalizability theory liberalizes classical test theory, not all aspects of classical theory are incorporated in generalizability theory. In addition, the ANOVA issues emphasized in generalizability theory are different from those that predominate in many experimental design and ANOVA texts. In particular, generalizability theory concentrates on variance components and their estimation.

Perhaps the most important aspect and unique feature of generalizability theory is its conceptual framework. Among the concepts are universes of admissible observations and G (Generalizability) studies, as well as universes of generalization and D (Decision) studies. The concepts and the methods of univariate generalizability theory are considered next in the context of a specific example. Subsequently, multivariate generalizability theory is introduced using an expanded version of the same example.

Univariate Generalizability Theory

Brennan *et al.* (1995) describe a Listening and Writing assessment that was administered to 50 examinees. (Strictly speaking, there were three forms, each of which was administered to three different groups of 50 examinees. Here, we will simply use results averaged over the three forms.) Each examinee (p) listened to 12 tape-recorded messages (t). Examinees were told to take notes while each message was played. At the conclusion of each message, examinees were told to use their notes to construct a written message. The written messages were scored by three trained raters (r) on a five-point holistic

scale for listening skills (i.e., information accuracy); a different three raters evaluated responses for writing skills. In the discussion here of univariate generalizability theory, we consider the Writing scores, only. Later, in the discussion of multivariate generalizability theory, we consider both the Listening and Writing scores.

Universe of Admissible Observations and G Studies

Assume that the $n_t = 12$ prompts can be viewed as a sample from an indefinitely large (i.e., approaches infinity) universe of potential prompts. Similarly, assume that the $n_r = 3$ raters can be viewed as a sample from an indefinitely large universe of potential raters. We say that prompts and raters are facets in the universe of admissible observations. Assume, as well, that in principle any one of the prompts in the universe of admissible observations could be evaluated by any one of the raters in the universe. If so, we say that the two facets are crossed in the universe of admissible observations ($t \times r$). (Although the example considered here involves only two facets, generalizability theory can accommodate any number of facets.)

Note that in describing the universe of admissible observations, no explicit reference has been made to persons who respond to the prompts. In generalizability theory the word universe is reserved for conditions of measurement (prompts and raters, in this example), while the word population is used for the objects of measurement (persons, in this example). If any person in the population might respond to any prompt in the universe evaluated by any rater in the universe, we say that the population and universe of admissible observations are crossed, which is represented $p \times t \times r$. For this situation, any observable score for a single prompt evaluated by a single rater can be represented as:

$$X_{ptr} = \mu + v_p + v_t + v_r + v_{pt} + v_{pr} + v_{tr} + v_{ptr} \quad [1]$$

where μ is the grand mean in the population and universe and v_α designates an uncorrelated effect for the component α . (Actually, the effect ptr is a residual effect involving the triple interaction and all other sources of error, not explicitly represented in the universe of admissible observations.)

The variance of the scores given by eqn [1], over the population of persons and the conditions in the universe of admissible observations, is

$$\sigma^2(X_{ptr}) = \sigma^2(p) + \sigma^2(t) + \sigma^2(r) + \sigma^2(pt) + \sigma^2(pr) + \sigma^2(tr) + \sigma^2(ptr) \quad [2]$$

That is, the total observed score variance can be decomposed into seven independent parts called variance components. When it is assumed that the population and both

facets in the universe of admissible observations are indefinitely large, the variance components in eqn [2] are called random effects variance components. It is important to note that these variance components are for single person–prompt–rater scores, as opposed to average scores over prompts and/or raters, which fall in the realm of D study considerations.

The above discussion of the population and universe of admissible observations is motivated by the Writing assessment introduced previously, but nothing stated so far actually relates to data from the Writing assessment. A particular type of analysis of the actual data constitutes a G study. The purpose of a G study is to estimate the variance components for the population and universe of admissible observations. This is usually accomplished using mean squares (see, e.g., Brennan, 2001a, especially section 3.4 and appendix B). The resulting estimates are for a random effects model, since we are assuming here that both facets and the objects-of-measurement facet are indefinitely large.

For the Writing assessment, the estimated variance components are:

$$\left. \begin{aligned} \sigma^2(p) &= 0.691, & \sigma^2(t) &= 0.025, & \sigma^2(r) &= 0.010 \\ \sigma^2(pt) &= 0.159, & \sigma^2(pr) &= 0.047, & \sigma^2(tr) &= 0.008 \\ \text{and } \sigma^2(ptr) &= 0.218. \end{aligned} \right\} \quad [3]$$

These are estimates of the actual variances (parameters) in eqn [2]. For example, $\hat{\sigma}^2(p)$ is an estimate of the variance component $\sigma^2(p)$, which can be interpreted roughly in the following manner. Suppose that, for each person in the population, the person's mean score (technically, expected score) was obtained over all essay prompts and all raters in the universe of admissible observations. The variance of these mean scores (over the population of persons) is $\sigma^2(p)$. The other main effect variance components for the prompt and rater facets can be interpreted in a similar manner. Note that for the Brennan *et al.* (1995) universe of admissible observations, the estimated variance attributable to essay prompts, $\hat{\sigma}^2(t) = 0.025$, is over twice as large as the estimated variance attributable to raters, $\hat{\sigma}^2(r) = 0.010$. This suggests that prompts differ much more in average difficulty than raters differ in average stringency.

Interaction variance components are more difficult to describe verbally, but approximate statements can be made. For example, $\hat{\sigma}^2(pt)$ estimates the extent to which the relative ordering of persons differs by prompt, and $\hat{\sigma}^2(pr)$ estimates the extent to which persons are rank-ordered differently by different raters. It is especially important to note that $\hat{\sigma}^2(pt) = 0.159$ is over three times larger than $\hat{\sigma}^2(pr) = 0.047$. This fact, combined with the previous observation that $\hat{\sigma}^2(t)$ is over twice as large as $\hat{\sigma}^2(r)$, suggests that prompts are a considerably greater source of variability in persons' scores than

are raters. The implication and importance of these facts will become evident in what follows.

Infinite Universe of Generalization and D Studies

The G study estimated variance components can be used to design efficient measurement procedures for operational use and to provide information for making substantive decisions about objects of measurement (i.e., persons in this example) in various D (Decision) studies. Broadly speaking, D studies emphasize the estimation, use, and interpretation of variance components for decision-making with well-specified measurement procedures.

Perhaps the most important D study consideration is the specification of a universe of generalization, which is the universe to which a decision maker wants to generalize based on the results of a particular measurement procedure. For the Writing assessment, suppose the universe of generalization contains all the prompts and raters in the universe of admissible observations. Since both facets are assumed to be infinite, the universe of generalization would be called infinite as well. This implies that the investigator wants to generalize persons' observed scores based on the specific prompts and raters in the measurement procedure to their scores for a universe of generalization that involves all of the infinite number of prompts and raters. In analysis of variance terminology, such a model is described as random.

The universe of generalization is closely related to potential replications of the measurement procedure. Suppose that the measurement procedure has each person responding to n'_t prompts, with each response to every prompt evaluated by the same n'_r raters. Furthermore, suppose that decisions about a person are based on his or her mean score over the $n'_t n'_r$ observations associated with the person. This is a verbal description of a D study $p \times T \times R$ design. It is much like the $p \times t \times r$ design, but there are two important differences.

First, the sample sizes for the D study (n'_t and n'_r) need not be the same as the sample sizes for the G study (n_t and n_r). This distinction is highlighted by the use of primes with D study sample sizes. Second, for the D study, interest focuses on mean scores for persons, rather than single person–prompt–rater scores that are the focus of G study estimated variance components. This emphasis on mean scores in a D study is highlighted by the use of uppercase letters for the facets in the D study $p \times T \times R$ design.

Under the conceptualization presented above a replication of the measurement procedure would involve a different sample of n'_t essay prompts and a different sample of n'_r raters. Such measurement procedures are described as randomly parallel. These randomly parallel replications span the entire universe of generalization, in the sense that the replications exhaust all conditions in the universe.

Universe scores

In principle, for any person, one can conceive of obtaining the person's mean score for every replication of the measurement procedure in the universe of generalization. For any such person, the expected value of these mean scores is defined as the person's universe score. The variance of universe scores over all persons in the population is called universe score variance. It has conceptual similarities with true score variance in classical test theory.

D study variance components

For the D study $p \times T \times R$ design, the linear model for an observable mean score over n'_t essay prompts and n'_r raters can be represented as

$$X_{pTR} = \mu + v_p + v_T + v_R + v_{pT} + v_{pR} + v_{TR} + v_{pTR} \quad [4]$$

The variances of the score effects in eqn [4] are called D study variance components. When it is assumed that the population and all facets in the universe of generalization are infinite, these variance components are random effects variance components. They can be estimated using the G study estimated variance components in eqns [3].

For example, consider using sample sizes $n'_t = 6$ and $n'_r = 2$ for the Writing assessment. In this case, the estimated D study random effects variance components are

$$\left. \begin{aligned} \hat{\sigma}^2(p) &= 0.691, & \hat{\sigma}^2(T) &= 0.004, & \hat{\sigma}^2(R) &= 0.005 \\ \hat{\sigma}^2(pT) &= 0.027, & \hat{\sigma}^2(pR) &= 0.024, & \hat{\sigma}^2(TR) &= 0.001 \\ & \text{and } \hat{\sigma}^2(pTR) &= 0.018. \end{aligned} \right\} [5]$$

These estimated variance components are for person mean scores over $n'_t = 6$ prompts and $n'_r = 2$ raters. Obtaining these results is simple, as indicated by the formulas in the first column in Table 1. These formulas are based on one of the fundamental results in statistics: the variance of a mean is the variance of the individual elements divided by the sample size. For example,

$$\hat{\sigma}^2(pT) = \hat{\sigma}^2(pt)/n'_t = 0.159/6 = 0.027$$

Table 1 Random effects variance components that enter $\sigma^2(\tau)$, $\sigma^2(\delta)$, and $\sigma^2(\Delta)$ for the $p \times T \times R$ design and different universes of generalization

	D studies ^a	
	T, R random	T fixed
$\sigma^2(p)$	τ	τ
$\sigma^2(T) = \sigma^2(t)/n'_t$	Δ	
$\sigma^2(R) = \sigma^2(r)/n'_r$	Δ	Δ
$\sigma^2(pT) = \sigma^2(pt)/n'_t$	Δ, δ	τ
$\sigma^2(pR) = \sigma^2(pr)/n'_r$	Δ, δ	Δ, δ
$\sigma^2(TR) = \sigma^2(tr)/n'_t n'_r$	Δ	Δ
$\sigma^2(pTR) = \sigma^2(ptr)/n'_t n'_r$	Δ, δ	Δ, δ

^a τ is universe score.

Note that the division-by-sample-size rule does not apply to the objects of measurement facet p . In particular, for this random effects example, $\hat{\sigma}^2(p) = 0.25$ is unchanged from its G study estimate, and it is the estimated universe score variance for the Writing assessment. An examinee's universe score is simply the expected value of that examinee's observed scores for all potential replications of the measurement procedure involving the same universe of generalization, the same D study design, and the same D study sample sizes. Universe score variance is denoted generically $\sigma^2(\tau)$, which is simply $\sigma^2(p)$, here.

Error variances

Under the random model, variance components other than $\sigma^2(p)$ contribute to one or more different types of error variance. Considered next are absolute and relative error variances.

Absolute error variance, $\sigma^2(\Delta)$. Absolute error is simply the difference between a person's observed and universe scores:

$$\Delta_p \equiv X_{pTR} - \mu_p \quad [6]$$

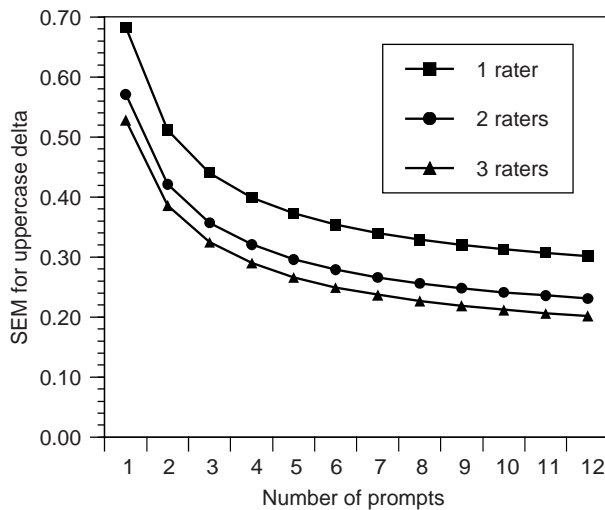
For the Writing assessment design,

$$\Delta_p = v_T + v_R + v_{pT} + v_{pR} + v_{TR} + v_{pTR} \quad [7]$$

Consequently, the variance of the absolute errors, $\sigma^2(\Delta)$, is the sum of all the variance components except $\sigma^2(p)$. This result is also provided in Table 1 under the column headed 'T, R random'.

Given the estimated D study variance components in eqns [5], the estimate of $\sigma^2(\Delta)$ for three prompts and two raters is

$$\begin{aligned} \hat{\sigma}^2(\Delta) &= 0.004 + 0.005 + 0.027 + 0.024 + 0.001 \\ &\quad + 0.018 = 0.079 \end{aligned}$$



and its square root is $\hat{\sigma}(\Delta) = 0.28$, which is interpretable as an estimate of the absolute standard error of measurement (SEM).

The sample sizes of $n'_t = 6$ and/or $n'_r = 2$ were chosen here for illustrative purposes, only. The same procedures and equations could be applied to any pair of sample sizes. For example, the left-hand panel of Figure 1 illustrates results for n'_t ranging from 1 to 12 and for n'_r ranging from 1 to 3. It is evident from Figure 1 that increasing n'_t and/or n'_r leads to a decrease in $\hat{\sigma}(\Delta)$. This result is sensible, since averaging over more conditions of measurement should reduce error. Figure 1 also suggests that using more than two raters leads to only a very slight reduction in $\hat{\sigma}(\Delta)$. Consequently, it would probably be unnecessary to use more than two raters for an actual measurement procedure. Further, Figure 1 indicates that using additional prompts decreases $\hat{\sigma}(\Delta)$ quicker than using additional raters. This is a direct result of the fact that $\hat{\sigma}^2(t) = 0.025$ is larger than $\hat{\sigma}^2(r) = 0.010$, and $\hat{\sigma}^2(pr) = 0.159$ is larger than $\hat{\sigma}^2(pr) = 0.047$. Finally, it does not appear that having more than about 10 prompts makes an appreciable difference in the magnitude of $\hat{\sigma}(\Delta)$; an investigator might even conclude that about six prompts is enough.

Relative error variance, $\sigma^2(\delta)$. Relative error is defined as the difference between a person's observed deviation score and his or her universe deviation score:

$$\delta_p \equiv (X_{pTR} - \mu_{TR}) - (\mu_p - \mu) \quad [8]$$

where μ_{TR} is the expected value over persons of the observed scores, X_{pTR} , for the $p \times T \times R$ design and an infinite universe of generalization. For the random model $p \times T \times R$ design, it can be shown that

$$\delta_p = v_{pT} + v_{pR} + v_{pTR} \quad [9]$$

and the variance of these relative errors is the sum of the variance components for the three effects in eqn [9].

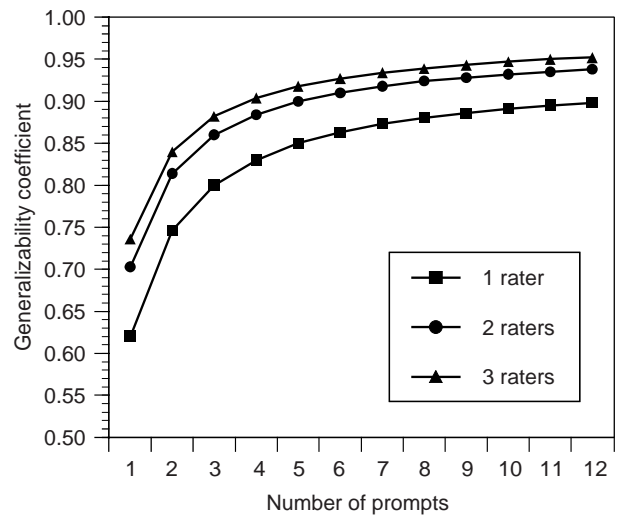


Figure 1 $\hat{\sigma}(\Delta)$ and $E\hat{\rho}^2$ for the $p \times T \times R$ design and an infinite universe of generalization.

This result is also given in **Table 1**, under the column headed '*T, R* random'. Relative error variance is similar to error variance in classical theory.

For the Writing assessment with $n'_t = 6$ and $n'_r = 2$,

$$\hat{\sigma}^2(\delta) = 0.027 + 0.024 + 0.018 = 0.069$$

and its square root is $\hat{\sigma}(\delta) = 0.26$, which is interpretable as an estimate of the relative SEM. Note that this value of $\hat{\sigma}(\delta)$ is smaller than $\hat{\sigma}(\Delta) = 0.28$ for the same pair of sample sizes. In general, $\hat{\sigma}(\delta)$ is less than $\hat{\sigma}(\Delta)$ because, as indicated in **Table 1**, $\hat{\sigma}^2(\delta)$ involves fewer variance components than $\hat{\sigma}^2(\Delta)$. In short, relative interpretations about persons' scores are less error prone than absolute interpretations.

Coefficients

Two types of reliability-like coefficients are widely used in generalizability theory. One coefficient is called a generalizability coefficient and denoted $E\rho^2$. The other coefficient is an index of dependability and is denoted Φ .

Generalizability coefficient, $E\rho^2$. The generalizability coefficient is the ratio of the universe score variance to itself plus relative error variance:

$$E\rho^2 = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\delta)} \quad [10]$$

It is the analog of a reliability coefficient in classical theory. For the Writing assessment with $n'_t = 6$ and $n'_r = 2$,

$$E\hat{\rho}^2 = \frac{0.691}{0.691 + 0.069} = 0.91$$

The right-hand panel of **Figure 1** provides a graph of $E\hat{\rho}^2$ for n'_t ranging from 1 to 12 and for n'_r ranging from 1 to 3. As observed in the discussion of SEMs, little is gained by having more than two raters or more than 10 prompts.

Index of dependability, Φ . An index of dependability is the ratio of universe score variance to itself plus absolute error variance:

$$\Phi = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\Delta)} \quad [11]$$

Φ differs from $E\rho^2$ in that Φ involves $\sigma^2(\Delta)$, whereas $E\rho^2$ involves $\sigma^2(\delta)$. Since $\sigma^2(\Delta)$ is generally larger than $\sigma^2(\delta)$, it follows that Φ is generally smaller than $E\rho^2$. The index Φ is appropriate when scores are given absolute interpretations, as in domain-referenced or criterion-referenced situations. For the Writing assessment with $n'_t = 6$ and $n'_r = 2$,

$$\hat{\Phi} = \frac{0.691}{0.691 + 0.079} = 0.90$$

Restricted Universes of Generalization and D Studies

The previous section assumed that the D study employed a $p \times T \times R$ design and the universe of generalization was infinite, consisting of two random facets, *T* and *R*. Recall that

the G study also employed a fully crossed design ($p \times t \times r$) for an infinite universe of admissible observations. In short, up to this point, it has been assumed that both designs are fully crossed and the size or extent of both universes is essentially the same. This need not be the case, however. For example, the universe of generalization may be narrower than the universe of admissible observations.

Suppose the Writing assessment investigator is not interested in generalizing over prompts. Rather, if the investigator were to replicate the measurement procedure, a different set of two raters would be used each time with the same six prompts. (Strictly speaking, it might be argued that the G study estimated variance components should be recomputed using only the actual six prompts under consideration, but we overlook this matter here.) If so, we would say that the investigator's universe of generalization is restricted in that it contains a *fixed* facet, *T*. Consequently, the universe of generalization under these circumstances is narrower than the universe of generalization for the random model. In ANOVA terminology, we are discussing a mixed model.

Suppose the investigator decides to use the same D study design structure as before – namely, the $p \times T \times R$ design. Under these circumstances, the last column of **Table 1** indicates which of the random effects D study variance components need to be summed to obtain estimates of universe score variance, $\sigma^2(\tau)$, as well as $\sigma^2(\Delta)$ and $\sigma^2(\delta)$. With $n'_t = 6$ and $n'_r = 2$.

$$\hat{\sigma}^2(\tau) = \hat{\sigma}^2(p) + \hat{\sigma}^2(pT) = 0.691 + 0.159 = 0.850$$

$$\begin{aligned} \hat{\sigma}^2(\Delta) &= \hat{\sigma}^2(R) + \hat{\sigma}^2(pR) + \hat{\sigma}^2(TR) + \hat{\sigma}^2(pTR) \\ &= 0.005 + 0.024 + 0.001 + 0.018 = 0.048 \end{aligned}$$

and

$$\hat{\sigma}^2(\delta) = \hat{\sigma}^2(pR) + \hat{\sigma}^2(pTR) = 0.024 + 0.018 = 0.042$$

It is particularly important to note that, with prompts fixed, $\sigma^2(pT)$ contributes to universe score variance, not error variance. Consequently, for a restricted universe of generalization with *T* fixed, universe score variance is larger than it is for a universe of generalization in which both *T* and *R* are random.

Given these results, it follows from eqn [10] that

$$E\hat{\rho}^2 = \frac{0.850}{0.850 + 0.042} = 0.95$$

Recall that, for these D study sample sizes ($n'_t = 6$ and $n'_r = 2$), when prompts were considered random, $E\hat{\rho}^2 = 0.91$. The estimated generalizability coefficient is larger when prompts are considered fixed because a universe of generalization with a fixed facet is narrower than a universe of generalization with both facets random. That is, generalizations to narrow universes are less error prone than generalizations to broader universes. It is important to note, however, that this does not necessarily mean that

narrow universes are to be preferred, because restricting a universe also restricts the extent to which an investigator can generalize. For example, when prompts are considered fixed, an investigator cannot logically draw inferences about what would happen if different prompts were used. Kane (1982) exploits and extends these notions in his discussion of a sampling model for validity (see also Kane, 2006).

Multivariate Generalizability Theory

Here, a brief introduction to some of the features of multivariate generalizability theory, which Cronbach *et al.* (1972) considered their most novel contribution to generalizability theory, is provided. (See Brennan (2001a, chapters 9–12) for a detailed, integrated treatment of multivariate generalizability theory).

For the details of the Listening and Writing assessment (Brennan *et al.*, 1995), which we will abbreviate *LW*, the reader is referred to the beginning of the section entitled ‘Univariate generalizability theory’. When we focus on Listening *or* Writing scores, only, the analysis is univariate. When we focus on both sets of scores simultaneously, the analysis is multivariate.

In characterizing the multivariate *LW* assessment, it is important to note that: (1) all tasks contributed to *both* Listening and Writing scores and (2) the three raters who evaluated responses for listening skills were different from the three raters who evaluated responses for writing skills. Under these circumstances, the multivariate G study design is symbolized $p^{\bullet} \times t^{\bullet} \times r^{\circ}$ with $k = 2$ dependent variables (Listening scores and Writing scores). The closed

circle after p indicates that Listening and Writing scores are obtained for all persons. The closed circle after t indicates that all tasks contribute to both Listening and Writing scores. The open circle after r indicates that different sets of raters are associated with Listening and Writing scores. A closed or open circle is sometimes said to refer to a linked or independent facet, respectively.

At its simplest level the variance components, $\sigma^2(\alpha)$, in univariate generalizability theory are replaced by variance–covariance matrices, Σ_{α} , in multivariate generalizability theory. The diagonal cells of these matrices always contain the variance components from the k univariate analyses, and some of these matrices contain covariance components in their off-diagonal cells. The second column in **Table 2** contains matrices of the estimates of the variance and covariance components for the *LW* assessment. Matrices contain covariance components when α contains only linked facets, that is, those with closed circles. For example, Σ_t contains covariance components for tasks because the same tasks contribute to both dependent variables. By contrast, Σ_r is diagonal because different raters are associated with Listening and Writing.

It is evident that $\hat{\sigma}_L^2(p)$ and $\hat{\sigma}_W^2(p)$ are large relative to the other estimated variance components, which suggests that examinees differ considerably with respect to their levels of proficiency in Listening and Writing. Furthermore, since $\hat{\sigma}_W^2(p) = 0.691$ is much larger than $\hat{\sigma}_L^2(p) = 0.324$, examinees appear to be much more variable in Writing proficiency than Listening proficiency. The estimated variance components $\hat{\sigma}_L^2(pr)$ and $\hat{\sigma}_W^2(pr)$ are also notably large, which suggests that the rank ordering of examinees differs by task for both Listening and Writing. By contrast, $\hat{\sigma}_L^2(r)$, $\hat{\sigma}_W^2(r)$, $\hat{\sigma}_L^2(pr)$, and $\hat{\sigma}_W^2(pr)$

Table 2 Multivariate results for the *LW* assessment with two raters

G study $p^{\bullet} \times t^{\bullet} \times r^{\circ}$ design		D study $p^{\bullet} \times T^{\bullet} \times R^{\circ}$ design	
		$n_t = 6$	$n_t = 12$
$\hat{\Sigma}_p$	$\begin{bmatrix} 0.324 & 0.356 \\ 0.356 & 0.691 \end{bmatrix}$	$\hat{\Sigma}_p$	$\begin{bmatrix} 0.324 & 0.356 \\ 0.356 & 0.691 \end{bmatrix}$
$\hat{\Sigma}_t$	$\begin{bmatrix} 0.127 & 0.039 \\ 0.039 & 0.025 \end{bmatrix}$	$\hat{\Sigma}_T$	$\begin{bmatrix} 0.021 & 0.006 \\ 0.006 & 0.004 \end{bmatrix}$
$\hat{\Sigma}_r$	$\begin{bmatrix} 0.012 & \\ & 0.010 \end{bmatrix}$	$\hat{\Sigma}_R$	$\begin{bmatrix} 0.006 & \\ & 0.005 \end{bmatrix}$
$\hat{\Sigma}_{pt}$	$\begin{bmatrix} 0.393 & 0.030 \\ 0.030 & 0.159 \end{bmatrix}$	$\hat{\Sigma}_{pT}$	$\begin{bmatrix} 0.066 & 0.005 \\ 0.005 & 0.026 \end{bmatrix}$
$\hat{\Sigma}_{pr}$	$\begin{bmatrix} 0.014 & \\ & 0.047 \end{bmatrix}$	$\hat{\Sigma}_{pR}$	$\begin{bmatrix} 0.007 & \\ & 0.024 \end{bmatrix}$
$\hat{\Sigma}_{tr}$	$\begin{bmatrix} 0.022 & \\ & 0.008 \end{bmatrix}$	$\hat{\Sigma}_{TR}$	$\begin{bmatrix} 0.002 & \\ & 0.001 \end{bmatrix}$
$\hat{\Sigma}_{ptr}$	$\begin{bmatrix} 0.317 & \\ & 0.218 \end{bmatrix}$	$\hat{\Sigma}_{pTR}$	$\begin{bmatrix} 0.026 & \\ & 0.018 \end{bmatrix}$
		$\hat{\Sigma}_{\Delta}$	$\begin{bmatrix} 0.128 & 0.011 \\ 0.011 & 0.078 \end{bmatrix}$
			$\begin{bmatrix} 0.070 & 0.006 \\ 0.006 & 0.053 \end{bmatrix}$

are all quite small, suggesting that raters are not nearly as large a contributor to total variance as are tasks.

There are positive estimated covariance components for p , t , and pt . The estimated covariance component for persons is particularly large (0.356) relative to the estimated variance components (0.324 and 0.691) suggesting that universe scores for Listening (μ_{pL}) and Writing (μ_{pW}) are highly correlated. The estimate of this disattenuated correlation is

$$\hat{\rho}_{LW}(p) = \frac{\hat{\sigma}_{LW}(p)}{\sqrt{\hat{\sigma}_L^2(p)\hat{\sigma}_W^2(p)}} = \frac{0.356}{\sqrt{0.324 \times 0.691}} = 0.75$$

Also, by the same line of reasoning, universe scores for tasks on Listening and Writing are highly correlated [$\hat{\rho}_{LW}(t) = 0.69$] suggesting that the rank ordering of the tasks in terms of difficulty is quite similar for Listening and Writing. By contrast, the pt interaction effects are only slightly correlated [$\hat{\rho}_{LW}(pt) = 0.12$].

The right-hand side of **Table 2** provides D study variance-covariance matrices for the $p^\bullet \times T^\bullet \times R^\circ$ design with two raters and two sample sizes for tasks. Just below these matrices are the associated $\hat{\Sigma}_\Delta$ matrices, which are obtained by summing all of the D study matrices except $\hat{\Sigma}_p$. The fact that the $\hat{\Sigma}_\Delta$ matrices have covariance components is a direct indication of the correlated error induced by using the same tasks to obtain both Listening and Writing scores. For example, when $n'_t = 6$, the correlated Δ -type error is $0.011/\sqrt{0.128 \times 0.078} = 0.11$. Multivariate generalizability theory is one of the few psychometric models that permits (indeed encourages) an explicit consideration of correlated error without much additional effort.

Multivariate generalizability theory does not require that a composite of the k scores be considered, but often one or more composites is/are of interest. In general, the composite universe score is

$$\mu_{pC} = w_1\mu_1 + w_2\mu_2 + \cdots + w_k\mu_k$$

where the w_j are investigator-specified weights. A simple example is difference scores, for which $w_1 = 1$ and $w_2 = -1$.

For the LW assessment and the Listening-minus-Writing difference-score composite, estimated universe score variance is

$$\begin{aligned}\hat{\sigma}_C^2(p) &= \hat{\sigma}_L^2(p) + \hat{\sigma}_W^2(p) - 2\hat{\sigma}_{LW}(p) = \\ &0.324 + 0.691 - 2(0.356) = 0.304\end{aligned}$$

For $n'_t = 6$ and $n'_r = 2$, the difference-score estimated absolute error variance is

$$\begin{aligned}\hat{\sigma}_C^2(\Delta) &= \hat{\sigma}_L^2(\Delta) + \hat{\sigma}_W^2(\Delta) - 2\hat{\sigma}_{LW}(\Delta) = \\ &0.128 + 0.078 - 2(0.011) = 0.183\end{aligned}$$

Therefore, $\hat{\Phi} = 0.304/(0.304 + 0.183) = 0.62$. Note that positively correlated error decreases $\hat{\sigma}_C^2(\Delta)$ when the composite is a difference score, which increases $\hat{\Phi}$.

Other Issues

One of the distinct advantages of multivariate generalizability theory is that it obviates otherwise complex estimation problems for designs that are unbalanced with respect to fixed facets, only. By contrast, unbalanced designs create complexities for estimating random effects variance components, as discussed extensively by Brennan (2001a, chapters 7 and 8). The essential problem is that there are numerous estimation procedures that give different estimates, with no obvious criterion for choosing among them.

Estimated variance components and functions of them (e.g., error variances and coefficients) are subject to sampling variability, as discussed extensively by Brennan (2001a, chapter 6). If it can be assumed that effects are normally distributed, often relatively straightforward procedures exist to estimate standard errors. When such assumptions are unreasonable (as they often are in generalizability analyses), nonparametric jackknife or bootstrap procedures can be considered. Recent research on the jackknife is provided by Feng (2002). Recent research on the bootstrap is provided by Brennan (2007), Tong and Brennan (2007), and Wiley (2000).

In most applications of generalizability theory, examinees or persons are the objects of measurement. Occasionally, however, some other collection of conditions plays the role of objects of measurement (e.g., class, district, or school means). It is relatively straightforward to apply generalizability theory in such cases (see, e.g., Brennan, 1995; Brennan *et al.*, 2003; Kane and Brennan, 1977).

Some general purpose computer packages for statistics (particularly SAS Institute Inc., 1996, Varcomp) can be used to estimate random effects G study variance components; other results for generalizability analyses usually are not provided by such packages, however. A few computer programs have been specifically designed for generalizability analyses. For any univariate, complete, balanced design GENOVA (Crick and Brennan, 1983) provides both G and D study results. urGENOVA (Brennan, 2001c) provides estimated G study random effects variance components for unbalanced designs. mGENOVA (Brennan, 2001b) provides multivariate G and D study results for a selected set of designs. These three programs are coordinated with Brennan (2001a).

In a sense generalizability theory is a macro-measurement model in that it pays particular attention to test scores, whereas item response theory focuses much more on item scores. Brennan (2006) argues that one of the most important challenges to psychometrics for the twenty-first century is to integrate these two models, if possible.

See also: Ability Testing; Admissions Testing; Analysis of Variance; Automated Essay Scoring; Writing Assessment and Instruction; Bootstrap Method; Classical Test Theory

Reliability; Criterion-Referenced Measurement; Educational Measurement: Overview; Equating and Scaling; Jackknife Methods; Multivariate Analysis of Variance; National Assessments; Norm-Referenced Measurement; Portfolio Assessment; Second Language Assessment; Validity.

Bibliography

- Allal, L. and Cardinet, J. (1994). Generalizability theory. In Husén, T. and Postlethwaite, T. N. (eds.) *The International Encyclopedia of Education*, vol. 5, pp 2458–2462. New York: Pergamon.
- Brennan, R. L. (1983). *Elements of Generalizability Theory*. Iowa City, IA: ACT.
- Brennan, R. L. (1992a). *Elements of Generalizability Theory*, revised edn. Iowa City, IA: ACT.
- Brennan, R. L. (1992b). Generalizability theory. *Educational Measurement: Issues and Practice* **11**(4), 27–34.
- Brennan, R. L. (1995). The conventional wisdom about group mean scores. *Journal of Educational Measurement* **32**, 385–396.
- Brennan, R. L. (1997). A perspective on the history of generalizability theory. *Educational Measurement: Issues and Practice* **16**(4), 14–20.
- Brennan, R. L. (2000). (Mis)conceptions about generalizability theory. *Educational Measurement: Issues and Practice* **19**(1), 5–10.
- Brennan, R. L. (2001a). *Generalizability Theory*. New York: Springer.
- Brennan, R. L. (2001b). *Manual for mGENOVA*. Iowa City, IA: Iowa Testing Programs, University of Iowa.
- Brennan, R. L. (2001c). *Manual for urGENOVA*. Iowa City, IA: Iowa Testing Programs, University of Iowa.
- Brennan, R. L. (2004). Generalizability theory. Lewis-Beck, M. S., Bryman, A., and Liao, T. F. (eds.) *The Sage Encyclopedia of Social Science Research Methods*, vol. 2, pp 418–420. Thousand Oaks, CA: Sage.
- Brennan, R. L. (2006). Perspectives on the evolution and future of educational measurement. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn, pp 1–16. Westport, CT: American Council on Education/Praeger.
- Brennan, R. L. (2007). Unbiased estimates of variance components with bootstrap procedures. *Educational and Psychological Measurement* **67**, 784–803.
- Brennan, R. L., Gao, X., and Colton, D. A. (1995). Generalizability analyses of work keys listening and writing tests. *Educational and Psychological Measurement* **55**, 157–176.
- Brennan, R. L., Yin, P., and Kane, T. M. (2003). Methodology for examining the reliability of group mean difference scores. *Journal of Educational Measurement* **40**, 207–230.
- Burt, C. (1936). The analysis of examination marks. In Hartog, P. and Rhodes, E. C. (eds.) *The Marks of Examiners*, pp 245–314. London: Macmillan.
- Cardinet, J. and Tourneur, Y. (1985). *Assurer la Mesure*. New York: Peter Lang.
- Cardinet, J., Tourneur, Y., and Allal, L. (1981). Extension of generalizability theory and its applications in educational measurement. *Journal of Educational Measurement* **18**, 183–204.
- Crick, J. E. and Brennan, R. L. (1983). *Manual for GENOVA: A Generalized Analysis of Variance System*. American College Testing Technical Bulletin No. 43. Iowa City, IA: ACT, Inc.
- Crocker, L. and Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt.
- Cronbach, L. J. (1991). Methodological studies – a personal retrospective. In Snow, R. E. and Wiley, D. E. (eds.) *Improving Inquiry in Social Science: A Volume in Honor of Lee J. Cronbach*, pp 385–400. Hillsdale, NJ: Erlbaum.
- Cronbach, L. J., Gleser, G. C., Nanda, H., and Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: Wiley.
- Cronbach, L. J., Rajaratnam, N., and Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology* **16**, 137–163.
- Ebel, R. L. (1951). Estimation of the reliability of ratings. *Psychometrika* **16**, 407–424.
- Feldt, L. S. and Brennan, R. L. (1989). Reliability. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn, pp 105–146. New York: American Council on Education and Macmillan.
- Feng, W. C. (2002). *Applicability of the Jackknife Procedure for Estimating Standard Errors of Variance Component Estimates in Selected Random Effects G Study Designs*. Unpublished Doctoral Dissertation. University of Iowa.
- Gleser, G. C., Cronbach, L. J., and Rajaratnam, N. (1965). Generalizability of scores influenced by multiple sources of variance. *Psychometrika* **30**, 395–418.
- Haertel, E. H. (2006). Reliability. In Brennan, R. L. (ed.) *Educational measurement*, 4th edn, pp 65–110. Westport, CT: American Council on Education/Praeger.
- Hoyt, C. J. (1941). Test reliability estimated by analysis of variance. *Psychometrika* **6**, 153–160.
- Kane, M. T. (1982). A sampling model for validity. *Applied Psychological Measurement* **6**, 125–160.
- Kane, M. T. (2006). Validation. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn, pp 17–64. Westport, CT: American Council on Education/Praeger.
- Kane, M. T. and Brennan, R. L. (1977). The generalizability of class means. *Review of Educational Research* **47**, 267–292.
- Lindquist, E. F. (1953). *Design and Analysis of Experiments in Psychology and Education*. Boston, MA: Houghton Mifflin.
- Rajaratnam, N., Cronbach, L. J., and Gleser, G. C. (1965). Generalizability of stratified-parallel tests. *Psychometrika* **30**, 39–56.
- SAS Institute Inc (1996). *The SAS System for Windows Release 6.12*. Cary, NC: SAS Institute.
- Shavelson, R. J. and Webb, N. M. (1991). *Generalizability Theory: A Primer*. Newbury Park, CA: Sage.
- Shavelson, R. J. and Webb, N. M. (1992). Generalizability theory. In Alkin, M. C. (ed.) *Encyclopedia of Educational Research*, vol. 2, pp 538–543. New York: Macmillan.
- Tong, Y. and Brennan, R. L. (2007). Bootstrap estimates of standard errors in generalizability theory. *Educational and Psychological Measurement* **67**, 804–817.
- Webb, N. M., Shavelson, R. J., and Haertel, E. H. (2007). Reliability coefficients and generalizability theory. In Rao, C. R. and Sinharay, S. (eds.) *Handbook of Statistics*, vol. 26, pp 81–124. Amsterdam: Elsevier.
- Wiley, E. W. (2000). *Bootstrap Strategies for Variance Component Estimation: Theoretical and Empirical Results*. Unpublished Doctoral Dissertation, Stanford.
- Yang, Z. and Chang, L. (2002). *Generalizability Theory and Its Applications*. Beijing: Educational Science Publishing House.

Relevant Website

<http://www.education.uiowa.edu> – Center for Advanced Studies in Measurement and Assessment, College of Education, University of Iowa.

Impact of Assessment on Classroom Practice

J L Herman, University of California, Los Angeles, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Since its inception, assessment has been advanced as a tool for improving educational practice. Initially, assessment was thought to exert such influence through the information it provided to educators and policymakers and thus presumably to improve decisions. Over time, however, recognition has grown of assessment's sociocultural and broader influences on teaching and learning, and recently researchers have differentiated between the effects of assessments of learning, the historical focus of assessment policy, practice, and research, and those of assessments for learning, the use of assessment during the course of instruction to directly inform and improve the learning process. These two types of assessment, summative and formative respectively, operate from distinct theoretical bases and have unique theories of action, which are explored below. The article begins by providing perspective on the purpose and uses of assessment of learning and how these may impact classroom practice before considering the research base on assessment's actual effects on practice and the factors that influence such impact. Finally, the article addresses theory and research on how assessment for learning, so-called formative assessment, influences classroom teaching and learning practice, and draws implications for future research.

This article uses the terms test and assessment interchangeably as denoting measures of student learning, although the former tends to denote traditional, multiple-choice-dominated standardized tests and the latter a broader array of methods. Similarly, a number of terms have been used in the literature to capture the concept of the impact of assessment, including assessment effects, consequences, washback, and backwash. These terms also are to be considered interchangeable.

Historical Perspective: Assessments for Selection and Improvement Purposes

The imperial examinations of the Han Dynasty (220–206 BC) are credited as the first use of standardized testing and the idea that gathering standard information on individuals' capabilities could lead to better and more equitable advancement decisions, in this case decisions about coveted civil service appointments. One can imagine that faced with the examination, elite, would-be candidates would prepare for the test and engage tutors to do so, a first instance of the impact of assessment on educational

practices. Used for purposes of selection, the Han Dynasty assessments also are a first example of how an assessment result can influence the practices in which individuals are subsequently placed.

From this very early precedent, educational assessment's use in selection thrived in the last century and continues today worldwide. Following World War I in the United States, for example, many school districts used standardized, intelligence tests for placing students into homogeneous classrooms, reinforcing tracking schemes that provided lower-track students with less rigorous instruction than higher-track students. Similarly, Great Britain's tradition of using assessment to certify, select, and track students for subsequent education dates back to the nineteenth century. Although more essay-based than those in the United States, these set the stage for tests that motivate performance and with it the substantial influence of testing on educational practice.

Side by side with traditions of testing for selection and placement developed that of using assessment to improve educational effectiveness. The father of modern educational measurement, Edward Thorndike, advocated the value of testing for guiding teaching, and Ralph Tyler's 8-year study (in the 1930s) created assessments designed to support progressive education goals. With a new objectives-based framework as the foundation for development, Tyler articulated the ways in which assessment could support teaching and learning, for example, monitoring student learning to inform instructional planning; affording school- and program-level accountability and improvement; and in loosening the confining influence of traditional scholarship tests. That is, predating current interest in assessment consequences, Tyler recognized that traditional tests constrained instructional practice, and that new forms of assessment were needed to permit new, progressive approaches to education to take hold. Similarly, like Hope for performance assessment in the United States nearly 50 years later, Tyler's assessments were intended to educate the public about important goals of schooling and to help teachers' clarify their instructional goals and practices, serving also to sow the seeds for changes in practice.

Assessment Impact in the Context of Overlapping Purposes

These fundamental, formal functions or purposes of measurement – to sort and select students, to certify

accomplishment, and to provide data for purposes of improving the quality of education – have coexisted since the beginning of educational testing. Further, while measurement experts advise that tests be designed to serve a specific purpose, the reality is that current educational policy tends to intertwine multiple purposes, and thus large-scale assessments of learning can impact instructional practice through multiple pathways. Whether the context is the test-based accountability systems in the United States, or the operation of A- and O-level type examination systems in many countries across the world, testing serves not only to provide technical evidence that informs multiple categories of decisions and policy functions (e.g., accountability, admissions/selection, placement, diagnosis, progress monitoring, certification, accreditation, and evaluation), but attention to and use of the technical data also carries important sociopolitical meanings that can translate into important consequences for educational practice.

Consider, for example, high school exit examinations common in the United States that students need to pass for a high school diploma; Key Stage, the General Certificate of Secondary Education (GCSE), or A-level type examinations in the United Kingdom; and higher education admissions tests that operate in countries across the globe. The ostensible purpose of the assessment is to certify that students have the knowledge and skills that the state has determined as needed for subsequent success.

As doing well on the assessment is important to students' futures, the existence of the test is expected to motivate students to learn, to prepare for the test, and to acquire the knowledge and skills that are assessed, as well as to motivate parents to pay attention and support their children's education. Moreover, even in the absence of tangible rewards and sanctions, educators will be motivated to help their students to do well, both because they care about their students' futures and because they see public reports of their students' test performance as evidence of their own professional efficacy. Accordingly, schools are motivated to do what they can to assure that their students will succeed and to do so may modify the content of their classes, may develop or purchase special materials to prepare students for the test, identify and establish new programs to help students succeed, etc. Instructional materials providers may respond by adapting existing materials or creating new products to meet the needs created by the test. In the case of high school exit examinations, educators also may use test results of students who do not pass the test to place them in special classes, provide extra tutoring options, and if the test provides subscale data on different aspects of performance, they may use the results to diagnose individual students' learning needs.

Those same results aggregated at the class, school, or district levels may be used to analyze and respond to curriculum strengths and weaknesses, to suggest specific

topic areas on which new or supplementary materials may be needed and in which teachers may benefit from professional development, or to infer priorities for budget allocations. Actors at the district, regional, state, and/or federal levels may be similarly analyzing and responding to the same data at different levels of aggregation, identifying schools and districts that are doing particularly well or that are in most need of help, including deriving implications for system priorities, needed changes, and/or new programs. Thus, data from a single test may fuel analysis and decisions at multiple levels, which in turn have cascading effects on instructional practice.

As another example, take a state-mandated teacher competency assessment that teachers must pass to gain tenure. By mandating such an assessment, state policymakers communicate their commitment to educational quality, their recognition of the importance of students having access to good teachers, and their responsibility for assuring the quality of the teaching force. As the assessment is developed, it makes concrete and communicates valued knowledge and competencies for fresh teachers. Just as secondary schools may respond to the requirements of high school exit or college-readiness examinations, so also can teacher preparation institutions be expected to adapt their curriculum and instructional pedagogy to the expectations of the teacher assessment and to use their students' results to strengthen their program's success in enabling students to pass. To the extent that sizable proportions are not passing, an institution may change its admission procedures, create special test preparation courses or tutoring opportunities, and/or try to pressure pre-preparation institutions to better prepare potential candidates. To the extent that the assessment delves deeply into meaningful aspects of teachers' content and pedagogical knowledge and requires demonstration of quality practice, preparing for the test may well help new teachers become more effective, and responding to the assessment requirements may strengthen and deepen existing preparation programs. In any event, unless it is a trivial assessment, the assessment may function to homogenize or standardize preparation programs and the content and pedagogical knowledge that teachers gain from them. As they can influence the what and how teachers learn in teacher-preparation programs, and who is allowed to enter the profession, such teacher assessments may substantially influence subsequent instructional practice.

These examples are not meant to be exhaustive, but rather to illustrate the confounding of test purposes, the merging of accountability/summative and improvement/formative functions, and of the technical, sociopolitical, and symbolic roles that high-stakes or high-visibility assessments play in influencing practice. A single test can launch a whole network of interrelated motivations and cascading actions, only some of which directly involve the technical use of the test.

These same overlapping functions also may accrue in classroom assessments of learning, particularly those that are used for grading. They too function to motivate and focus students' effort, communicate what is important to learn and how to do it, and, for learning-oriented teachers and students, to stimulate reflection on what has been learned, how to fill the gap between that and intended learning goals, and how to improve the teaching and learning sequence (see also the section on formative assessment).

Research on the Effects on High-Visibility External Tests

What of the actual impact of high-visibility tests? Researchers around the globe have studied this question, drawing largely on interview and survey methods, sometimes supplemented with observations of practice.

Consistent Results on Potential Benefits

The rigor of design and methodology vary substantially from study to study, but results have been quite consistent regardless of locale, for example, studies of state accountability tests in more than a dozen states in the United States, of A- or GCSE or Key Stage examinations in England, and language and higher education admissions testing programs in countries such as Australia, Central and Eastern Europe, China, Hong Kong, Israel, Japan, New Zealand, and Sri Lanka:

- *Testing signals priorities for curriculum and instruction: high-visibility tests serve to focus the content of instruction.* School administrators and teachers pay attention to what is tested, analyze test results, and adapt curriculum and teaching accordingly.
- *Teachers tend to model the pedagogical approach reflected on high-visibility test.* When a high-visibility assessment is composed of multiple-choice tests, teachers tend to rely heavily on multiple-choice worksheets in their classroom instruction. However, when the assessments use open-ended items and/or extended writing and rubrics to judge the quality of student work, teachers prepare students for the test by incorporating these same types of activities in their classroom practice.
- Curriculum developers, particularly commercial interests, respond to important tests by modifying existing textbooks and other instructional materials and/or developing and marketing new ones to address test content and format. These products in turn may become primary resources that influence not only practice but also teachers' understanding of test expectations.
- *Test results can make visible and promote responsiveness to the needs of students who previously have been underserved.* In the United States, for example, where all students are held

to the same standards and schools must reach test-based accountability targets for all subgroups, research shows schools and teachers paying more attention to traditionally low-performing students, including second-language learners and students with disabilities. Test results are used to identify and provide special help for faltering students, for example, through the creation of new courses, purchase of new materials, and provision of extra school opportunities (e.g., before or after school tutoring and summer school).

Consistent Results on Problematic Effects

While the above points demonstrate some ways that assessment, depending on the nature of the test, can leverage productive changes in instructional practice, research shows unintended, negative consequences:

- *Schools and teachers may focus on the test rather than underlying standards or learning goals.* With sanctions and incentives riding on test performance, educators give primary attention to what is tested and how it is tested, rather than to the standards or learning goals the test is intended to measure. For example, teachers may shift how much classroom instruction time is accorded to particular topics or curriculum subjects depending on the content emphasis of the test.
- *What is not tested can become invisible.* As a corollary to focusing on the test rather than the standards, that which is not tested tends to get less attention or may be ignored altogether. Both the broader domain of the tested disciplines and important subjects that are not tested may get short shrift. In the United States, for example, many state tests tend to give relatively little attention to complex thinking and problem solving and tend to focus on lower levels of learning, which can lead to similar emphases in classroom practice.
- *Focusing on the test, rather than underlying learning, may encourage performance orientation and transmission-type teaching.* When passing the test, rather than learning, becomes the goal, schools may unwittingly promote a performance orientation in students, which can work against students' engagement and persistence in learning, metacognition, and self-regulation. Particularly for high-visibility tests that are predominantly multiple choice, teachers may concentrate on direct instruction to help students acquire specific content, rather than using pedagogy that helps students build conceptual understanding and problem-solving capability.
- *Instructional time is diverted to specific test-preparation activities.* Schools provide students with practice on the specific types of tasks and formats that are expected on the test, through commercial test-preparation packages, special classes, and homework. Such activities aim specifically to help students do well on the test,

rather than promoting students' learning, and depending on the school and the pressure to improve test scores, can divert weeks or more of instructional time.

- *Assessment-driven instruction can spur more and more testing.* In the United States, more and more districts are mandating interim assessments during the school year, largely mimicking the expected content and format of their annual state tests, to provide quarterly or so feedback on how students are doing, and to encourage teachers to keep their eye on student progress on the knowledge and skills that will be tested. Moreover, many districts have become more prescriptive about how and what teachers are supposed to teach and have created pacing guides detailing what is to be covered when.
- Rather than motivating teachers to improve teaching and learning, high-visibility testing can cause teachers to feel pressured and demoralized, particularly when they view testing targets as unrealistic.
- *Testing schedules can distort teaching and learning plans.* Teachers may even adjust the amount of time they devote to basic subjects (e.g., reading and math) depending on when each subject is assessed. For example, if math is tested after grade 4 and reading in grade 5, then math may receive relatively more attention in grade 4 and reading in grade 5. Similarly, the best teachers may be moved to the particular grades at which important assessments are given. The desire to boost test scores thus can trump goals for teaching and learning in curriculum and personnel decision making.

Factors that Influence Assessment Impact

While research shows various potential effects of assessment on instructional practice, it also makes clear that such effects are not automatic – neither do they tend to be uniformly positive or negative. A first issue in shaping potential impact is the nature of the test itself. In fact, worry in the 1990s that traditional, standardized multiple-choice testing in the United States was producing a narrow, drill-and-skill curriculum led advocates to advance large-scale performance assessments to be, in the words of Lauren Resnick, “tests worth teaching to.” By requiring students to engage in more authentic tasks that required complex thinking and problem solving and extended responses, performance assessment was intended to drive instruction in a similar constructivist direction, an early case of testing policy being explicitly used to leverage change. Similarly, more than 20 years ago, J. Charles Alderson identified the potentially powerful washback of language testing and argued that changes in testing could support desired changes in language curriculum and teaching. The 1993 Hong Kong Certificate of Education in English, as an example, was launched to encourage secondary teachers of English to adopt a more communicative and purposeful approach, rather than a structural approach, to

language development. The new assessment included more authentic and interactive tasks that required students to integrate listening, writing, reading, and/or speaking and to apply their skills in context; or more recently, Great Britain's key stage tests ostensibly are designed to encourage teachers to address a broad curriculum within each assessed subject, to enable students to respond to and create a variety of forms of writing and to engage students in applying mathematical and scientific knowledge and skills.

Yet, the path from potential or intended impact to actual changes in practice is neither simple nor clear-cut. The move from policy mandate to operational test involves a process of implementation, from public dialog and public discourse about the meaning of the policy, to how and by whom the test is specified and developed, and through a complex chain of visible and invisible decisions and actions that may redefine initial intentions. From operational test to impact on practice similarly involves a circuitous and unpredictable process, as the concept of impact necessarily involves some change in practice.

Succinctly summarizing the many factors that influence it, theorists have noted that change requires both the will and the capacity to change and have suggested that policymakers hoping to use assessment to leverage changes in practice must consider the resources and strategies needed to support such change. From a motivational perspective, for example, in response to any new testing mandate, stakeholders may choose to accept and use it, ignore it, or fight it. Only those in the first category have immediate potential for intended effects and if change is to occur, action must be taken to convert those in the other two categories.

Yet even with willingness, intended changes in practice will occur only when educators understand what is expected and have the prerequisite knowledge, skills, and understandings, as well as relevant resources, to make the change. As a new test rolls out, for instance, educators may differ in their understandings of what the test actually assesses and thus in its implications for their practice. Teachers will vary as well in their capability to engage in new, expected practices, depending on how complex intended changes are and how far from existing practice. Thus, what “teaching to the test,” a common response to high-visibility tests, actually means depends on the beliefs and capacities of the educators who are responding and the resources and support they have available to them. In response to the same test, some schools and/or teachers may make meaningful and productive changes in teaching and learning, while others may react more mechanically with test-preparation activities simply mimicking the test.

Research on the consequences of assessment worldwide, in fact, indicates that while teachers may easily modify the content of their teaching, pedagogy is more resistant to change. Teachers' initial pedagogical responses tend to be superficial, for example, simply incorporating samples of

the new assessments as an add-on to existing curriculum. Over time and only with sufficient support do new practices become more integrated with the ongoing instruction and new standards and quality criteria are more deeply understood and acted upon. Effects on practice, in short, are not the product of a test, but the result of a process that is influenced by stakeholders' attitudes and efficacy, the meanings they make of what is expected, the school context, power or other relationships in which the change is embedded, the stakes riding on test results, the nature of the intended change, and the resources and strategies used to help teachers incorporate new practices, to name just a few. Assessments may be mandated top down with the intention of accomplishing various impacts, and even bolstered by policy incentives and sanctions but the actual impact is neither easily controlled nor easily predicted.

The Effects of Assessment for Learning: Classroom Formative Assessment

In contrast to assessments of learning that are mandated top down, assessments for learning, so-called formative assessments, occur bottom up, within the actual context of classroom teaching and learning. Similar to standardized testing of learning, however, the roots of formative assessment also can be found in ancient history, that of Socrates in fifth-century BC Greece. Socrates educated his students through questioning and dialog, breaking down important ideas into a series of questions, whose answers, with probing, gradually helped students to acquire new insights. Questions and answers were not judged right or wrong, but were the means to building knowledge and understanding. Similarly in formative assessment, assessment is used to elicit students' understanding in order to form subsequent teaching and learning.

The last decade has seen increasing, worldwide interest in formative assessment, fueled in large part by Black and Wiliam's landmark meta-analysis showing the strong effects of formative assessment on student learning, particularly for low-ability students. Policymakers and educators increasingly have recognized that while external tests of learning may help to communicate priority goals, motivate improvement, and identify who has succeeded or not-summative functions, results are too little and too late to directly improve practice. Rather the most powerful use of assessment occurs hand in hand with classroom teaching and learning.

The use of data is key to the idea: to be considered formative, assessment evidence must be acted upon during the course of classroom instruction. Rather than focusing backward on what has been learned, formative assessment helps to chart the learning road forward, by identifying and providing information to fill any gaps between the learners' current status and goals for learning.

Moreover, more than solely a source of evidence that informs subsequent teaching and learning, formative assessment may enhance student learning directly. For example, in asking students to make public their thinking, formative probes can provide scaffolding that helps students to confront their misconceptions, refine and deepen their understandings, and move to more sophisticated levels of expertise. By making learning goals explicit and involving students in self-assessment, formative assessment can also promote students as agents in their own learning, increasing student motivation, autonomy, meta-cognition, as well as learning.

Formative assessment itself involves a change in instructional practice. Rather than imparting knowledge in a transmission-oriented process, in formative assessment teachers guide students toward significant learning goals and actively engage students as assessors of themselves and their peers. Formative assessment occurs when teachers make their learning goals and success criteria explicit for students, gather evidence of how student learning is progressing, partner with students in a process of reciprocal feedback, and engage the classroom as a community to improve students' learning. The social context of learning is fundamental to the process as is the need for classroom culture and norms that support active learning communities—for example, shared language and understanding of expected performance; relationships of trust and respect; and shared responsibility for and power in the learning process. Theorists observe that enacting a meaningful process of formative assessment influences what students perceive as valued knowledge, who can learn, and who controls and is valued in the learning process. However, as with the effects of assessments of learning on practice, those of formative assessment do not happen automatically but involve a complex process of change. Moreover, at this point, the anticipated effects of formative assessment on instructional practice are based on theoretical argument and are yet to be fully confirmed through empirical or other research.

Conclusion

Available theory and research suggest the ways in which assessments of learning and those for learning may exert influence on instructional practice. In both cases, such effects are not automatic but rather dependent on a complex process of change that is influenced by many, difficult-to-control, factors. Marshaling the power of assessment to improve practice requires better understanding of how such change occurs and the nature of the assessments and support processes that can best achieve positive effects while minimizing unintended negative ones. Continuing research needs to explore whether high stakes and formative assessments are effective in meeting their intended aims, how effective each is relative to other

strategies for influencing instructional practice, and whether and how assessment may undermine or support other teaching and learning goals, including effects on educational equity. Most studies thus far have examined short-term effects. Also needed are longer-term studies that examine implementation and effects over time.

See also: Classroom Assessment Tasks and Tests; Formative Assessment; Summative Assessment by Teachers.

Further Reading

- Alderson, J. C. and Wall, D. (1993). Does washback exist? *Applied Linguistics* **14**(2), 115–129.
- Black, P., McCormick, R., James, M., and Pedder, D. (2006). Learning how to learn and assessment for learning: A theoretical inquiry. *Research Papers in Education* **21**(2), 119–132.
- Black, P. and William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan* **80**, 139–148.
- Cheng, L. and Curtis, A. (2004). Washback or backwash: A review of impact. In Cheng, L., Watanabe, Y., and Curtis, A. (eds.) *Washback in Language Testing: Research Contexts and Methods*, pp 3–18. Mahwah, NJ: Erlbaum.
- Cheng, L., Watanabe, Y., and Curtis, A. (eds.) *Washback in Language Testing: Research Contexts and Methods*. Mahwah, NJ: Erlbaum.
- Daugherty, R. and Ecclestone, K. (2006). Constructing assessment for learning in the UK policy environment. In Gardner, J. (ed.) *Assessment and Learning*, pp 149–168. London: Sage.
- Dwyer, C. A. (ed.) (2007). *The Future of Assessment: Shaping Teaching and Learning*. Mahwah, NJ: Erlbaum.
- Ferman, I. (2004). The washback of an EFL national oral matriculation test to teaching and learning. In Cheng, L., Watanabe, Y., and Curtis, A. (eds.) *Washback in Language Testing: Research Contexts and Methods*, pp 191–210. Mahwah, NJ: Erlbaum.
- Fuhrman, S. and Elmore, R. (eds.) *Redesigning Accountability*. New York: Teachers College Press.
- Gardner, J. (ed.) *Assessment and Learning*. London: Sage.
- Haertel, E. H. and Herman, J. L. (2005). A historical perspective on validity arguments for accountability testing. In Herman, J. L. and Haertel, E. L. (eds.) *Uses and Misuses of Data for Educational Accountability and Improvement, 104th Yearbook of the National Society for the Study of Education (NSSE)*, Part 2, pp 1–34. Malden, MA: Blackwell.
- Harlen, W. (2006). The role of assessment in developing motivation for learning. In Gardner, J. (ed.) *Assessment and Learning*, pp 61–80. London: Sage.
- Herman, J. L. and Baker, E. L. (in press). Assessment policy: Making sense of the babel. In Plank, D., Sykes, G., and Schneider, B. (eds.) *AERA Handbook on Education Policy*. Newbury Park, CA: Sage.
- Herman, J. and Haertel, E. (eds.) *Uses and Misuses of Data for Educational Accountability and Improvement, 104th Yearbook of the National Society for the Study of Education (NSSE)*, Part 2. Malden, MA: Blackwell.
- Kirkup, C., Sizmur, J., Sturman, L., and Lewis, K. (2005). Schools use of data in teaching and learning. *DfES Research Report #671*. London: DfES. <http://www.nfer.ac.uk/research-areas/pims-data/summaries/schools-use-of-data-in-teaching-and-learning.cfm> (accessed September 2009).
- QCA (2007). National curriculum tests: Building effective assessment. Paper 2. www.qca.org.uk/libraryAssets/media/National_curriculum_tests_building_effective_assessment.pdf.
- Qi, L. (2004). Has a high stakes test produced the intended changes? In Cheng, L., Watanabe, Y., and Curtis, A. (eds.) *Washback in Language Testing: Research Contexts and Methods*, pp 171–190. Mahwah, NJ: Erlbaum.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* **18**, 119–144.
- Saville, N. and Hawkey, R. (2004). The IELTS impact study: Investigating washback on teaching materials. In Cheng, L., Watanabe, Y., and Curtis, A. (eds.) *Washback in Language Testing: Research Contexts and Methods*, pp 73–96. Mahwah, NJ: Erlbaum.
- Shepard, L. A. (2006). Classroom assessment. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 623–646. Washington, DC: National Council on Measurement in Education and American Council on Education/Praeger.
- Stecher, B. and Chun, T. (2001). School and classroom practices in Washington state during two years of education reform. *CRESST Report #550*. Los Angeles, CA: CRESST.
- Wall, D. (2005). *The Impact of High-Stakes Examinations on Classroom Teaching. Studies in Language Testing, Vol 22*. Cambridge: Cambridge University Press.
- Watanabe, Y. (2004). Teacher factors mediating washback. In Cheng, L., Watanabe, Y., and Curtis, A. (eds.) *Washback in Language Testing: Research Contexts and Methods*, pp 129–146. Mahwah, NJ: Erlbaum.
- Willis, J. (2007). Assessment for learning: Why the theory needs the practice. *International Journal of Pedagogies and Learning* **3**(2), 52–59.

Relevant Websites

- <http://research.acer.edu.au> – Australian Council for Educational Research.
- <http://www.CRESST.org> – National Center for Research on Evaluation, Standards, and Student Testing.
- <http://www.nfer.ac.uk> – National Foundation for Educational Research.
- <http://www.qca.org.uk> – Qualifications and Curriculum Authority.

Intelligent Systems

V J Shute, Florida State University, Tallahassee, FL, USA

D Zapata-Rivera, Educational Testing Service, Princeton, NJ, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Different models of educational measurement engender different instructional practices in the classroom or within intelligent instructional systems, and thus have different effects on student learning. Historically, the main aim of measuring students' educational progress was to identify differences among students in order to rank order them by achievement. This type of measurement model makes heavy use of summative assessment, which is useful for accountability purposes but only marginally useful for guiding day-to-day instruction and supporting student learning. In contrast, student-centered measurement models rely mostly on formative assessment, which can be very useful in guiding instruction and supporting student learning, but provide an inadequate basis for accountability purposes. Now, instead of considering these as disjunctive measurement models, it may be possible to effectively combine them within new, enhanced intelligent systems that use both kinds of assessments – summative and formative – and which leverage computer technology, educational measurement, and cognitive science to address problems that face education today.

A unified approach to educational measurement may be justified based on the following assumptions: (1) individual differences among students have powerful effects on learning, (2) these effects can be quantified and predicted, and (3) technology can capitalize on these effects to the benefit of teachers and students (as well as others, such as administrators and parents). The goal is to figure out how to effectively and efficiently integrate appropriate assessment and instruction to improve student learning and education in general. This is substantially easier said than done, although significant advances have been made in both educational-measurement technologies and intelligent systems.

We begin this article by broadly defining educational measurement, specifically in terms of the critical role assessment plays in education. Summative and formative assessment both have their place in education and we examine how they may be more tightly and strategically coupled. Next, we define intelligent systems, focusing on the usage of assessment data by computers to make inferences about students' cognitive and other attributes. This provides the foundation for rendering decisions about what content to present next to the student. Measuring what students know and do not know (and to what degree)

is the first step in remediating students or advancing them to higher levels. This requires the development and use of valid and reliable assessment tools and results to make accurate diagnoses and guide learning. We conclude with an outline of an approach to incorporating evidence-based assessments into intelligent systems to improve learning.

Definitions

Educational Measurement

Educational measurement may be broadly defined as the application of a standard scale or measuring tool to determine the degree to which educationally valuable knowledge, skills, and abilities have been acquired. According to the website of the National Council on Measurement in Education, this includes theory, techniques, and instrumentation available for measurement of educationally relevant human, institutional, and social characteristics. We measure to obtain information, and such information may or may not be useful, depending on the accuracy of the instruments and the skillful manner with which they are used.

Assessment is a general term that includes testing. Progress toward educational goals is typically assessed through testing. Assessment is both an instrument and a process by which information is obtained relative to a known objective or goal. Since inferences are made about what a person knows on the basis of his/her responses to a limited number of assessment tasks or items, there is always some uncertainty in inferences made on the basis of assessments. The goal in educational measurement is to minimize uncertainty or error; thus, key aspects of assessment quality are validity and reliability. Reliability refers to the consistency of assessment results – the degree to which they rank order students in the same way. Validity refers to the extent to which the assessment accurately measures what it is supposed to measure, and the accuracy of the inferences made from task or test results.

Types of Assessment

We consider here two main types of assessment: summative and formative. Summative assessment reflects the traditional approach used to assess educational outcomes. This involves using assessment information for high

stakes, cumulative purposes, such as promotion, certification, and so on. It is usually administered after some major event, like the end of the school year or marking period or before a big event, like college. Benefits of this approach include the following: (1) it allows for comparing student performances across diverse populations on clearly defined educational objectives and standards; (2) it provides reliable data (e.g., scores) that can be used for accountability purposes at various levels (e.g., classroom, school, district, state, and national) and for various stakeholders (e.g., students, teachers, and administrators); and (3) it can inform educational policy (e.g., curriculum or funding decisions).

Formative assessment reflects a more progressive approach in education. This involves using assessments to support teaching and learning. Formative assessment is tied directly into the fabric of the classroom and uses results from students' activities as the basis on which to adjust instruction to promote learning in a timely manner. This type of assessment is administered much more frequently than summative assessment, and has shown great potential for harnessing the power of assessments to support learning in different content areas and for diverse audiences. In addition to providing teachers with evidence about how their students are learning so that they can revise instruction appropriately, formative assessment may directly involve students in the learning process, such as by providing feedback that will help students gain insight about how to improve, and by suggesting (or implementing) instructional adjustments based on assessment results.

We now turn our attention to intelligent computer-based systems which have been around for several decades, but have yet to be fully embraced by education. Their primary goal is to enhance student learning; so assessment should, in theory, play a key role in these systems.

Intelligent Systems

Intelligent systems (also known as intelligent tutoring systems) refer to educational software containing an artificial-intelligence component. The software tracks students' work, adjusting feedback and providing hints along the way. By collecting information on a particular student's performance as well as other cognitive and noncognitive variables, the software can make inferences about strengths and weaknesses, and can suggest additional work.

A summary of requirements for intelligent systems was presented by Hartley and Sleeman in the early 1970s. They argued that these systems must possess: (1) knowledge of the learner (student model), (2) knowledge of the domain (expert model), and (3) knowledge of teaching strategies (pedagogical model). It is interesting to note

that this simple list has not changed in more than three decades; however, advances have been made in each of the three areas. All of the computer-resident knowledge marks a radical shift from earlier knowledge-free, computer-assisted instructional programs. Furthermore, the ability to diagnose students' errors and adapt instruction based on the diagnosis represents a key difference between intelligent versus other computer-based systems, such as simulations. Intelligent systems are also aligned with the features and goals of formative assessment. The three main components of intelligent systems – student, expert, and pedagogical models – are now briefly described.

A student learns from an intelligent system primarily by solving problems – ones that are appropriately selected or tailor-made, and that serve as learning experiences for that student. The system may start by assessing what the student already knows. Information about the student is maintained within what is called the student model, which is updated during the course of learning. The system then must consider what the student needs to know. This information is embodied in the domain-expert model. Finally, the system must decide what unit of content (e.g., assessment task or instructional element) ought to be presented next, and how it should be presented. This is achieved by the pedagogical model (or tutor). From all of these considerations, the system selects or generates a problem, then either works out a solution to the problem (via the domain-expert model), or retrieves a prepared solution. The intelligent system compares its solution to the one the student has prepared and performs a diagnosis based on differences between the two as well as other information available in the student model. Feedback is offered by the system based on considerations such as how long it has been since feedback was last provided, whether the student already received some particular advice, and so on. After this, the program updates the student model, and the entire cycle is repeated, starting with selecting or generating a new problem.

Despite the great promises of intelligent systems, they are currently not widely used in classrooms, partly because of their cost, and also because of measurement limitations. We now focus on the latter in more detail, describing how assessments differ between traditional intelligent systems and newer, enhanced intelligent systems. This is intended to provide the foundation on which to consider a new view of educational measurement within intelligent systems.

Assessments' Role in Intelligent Systems

For the most part, traditional intelligent systems use a formative assessment model, where different student actions invoke different instructional decisions or paths. This comprises the basis for adaptive instruction. New

enhanced intelligent systems extend the assessment capabilities of traditional systems. Some of these enhancements include the use of: evidence-based assessment data, explicit links to state curriculum standards, formative and summative sources of assessment information, new measurement techniques from educational psychology and cognitive science, and an explicit and strong role for teachers. Both types of intelligent systems are now discussed in turn.

Traditional Intelligent Systems

As noted earlier, formative assessment is explicitly intended to support student learning, defining the role of the student as an active, creative, and reflective participant in the learning process. Learning environments that make use of formative assessment typically include individualized instruction, along with hands-on, authentic learning activities. Assessments are used primarily to inform teaching and improve student learning.

One major downside of this model is that formative assessment is often implemented in a nonstandardized and hence less-rigorous manner than summative assessment. This can hamper the validity and reliability of the assessment tools and data. As the validity and reliability of the assessment data affect the accuracy of the student diagnosis, and the diagnosis informs instructional support, if the first part of the chain is weak, the rest (i.e., diagnostic accuracy and effective instructional support) would consequently be compromised. In other words, the effectiveness of an intelligent system in achieving its goal hinges on the goodness of information in the student model (i.e., the inferences about what the student knows and can do).

Traditional intelligent systems that employ formative assessment utilize a rich source of student data from which to draw inferences. For example, evidence is captured from all past and current student–system interactions, and may differ in type and grain size. Thus, in addition to the nonstandardization of methods for implementing formative assessment in traditional intelligent systems, there are also problems with accurately modeling student knowledge within such multifaceted environments. This poses a number of psychometric challenges (e.g., modeling of multiple abilities, capabilities, and other learner characteristics) regardless of the measurement model employed.

We now take a closer look at new intelligent systems that are starting to integrate formative and summative sources of assessment information, and are employed within real classroom settings.

Enhanced Intelligent Systems

Most current intelligent systems reside primarily in the laboratory. This isolation from real classrooms explains why their designs have not been overly concerned with

summative types of assessment, and also explains to some extent why they have not been widely adopted. That is, learning systems deployed within laboratory-based environments do not have to comply with the same high standards (e.g., accountability requirements), as in real-classroom environments. However, as these systems move out of the laboratory and into the classroom, the need for accountability (e.g., standards and norm-referenced assessments) increases.

Summative assessments are explicitly designed for accountability purposes. They represent a source of valid and reliable evidence of student knowledge. Due to national and international accountability requirements and interests, summative assessments are widely used in schools. For example, in the US, summative assessments have received increased attention after the US Congress passed the No Child Left Behind (NCLB) initiative in 2001. And on the international front, the Programme for International Student Assessment (PISA) is being used to compare student achievement in countries all over the world. The measurement community has made important advances in the development of psychometric models (e.g., Rasch, item-response theory) that provide reliable and valid assessment information, typically presented as a single measure of ability at a particular point in time for any given student. These data, however, have limited use for formative purposes. And one often-cited downside of this emphasis on accountability is that teachers tend to view testing as time taken away from valuable instruction and learning.

Over a decade ago, Snow and Mandinach called for the development of principles for creating valid and useful instructional-assessment systems. We are, only now, beginning to see intelligent systems entering classrooms that integrate sound assessment and instruction. These systems are characterized by (1) a strong presence of teachers in all phases of the project, (2) a cognitive model that is used to drive instructional and assessment interactions, and (3) explicit connections to state standards and standardized state tests.

An example of a successfully deployed intelligent system can be seen in the web-based cognitive tutors designed and developed by Anderson, Koedinger, and colleagues at Carnegie Mellon University. A derivation of their cognitive-tutor approach is called *assistsments* – the merging together of robust assessment with instructional assistance into one system. *Assistsments* use real (i.e., released) items from the Massachusetts comprehensive assessment system (MCAS) state exams within the system for both assessment and instructional purposes.

Table 1 summarizes the main features that separate traditional from enhanced intelligent systems with regard to the role assessments play.

While assessments provide a good example of joining formative and summative models within an intelligent

system, this important blending is still uncommon despite calls for their union. A few other systems demonstrating similar capabilities include: Sistema de Evaluacion Inteligente mediante Tests (SIETTE) by Conejo and colleagues, adaptive content for evidence-based diagnosis (ACED) by Shute and colleagues, and English assessment-based learning environments (ABLE) by Zapata-Rivera and colleagues. The next section presents an evidence-based approach designed to create valid assessments for summative or formative purposes, and which may be implemented as part of an intelligent system.

Evidence-Centered Design and Intelligent Systems

An intelligent system that includes valid assessments, for formative and/or summative purposes, must elicit behavior from the student that bears evidence about key skills and knowledge. In addition, the system must provide principled interpretations of that evidence in terms that

suit the purpose of the assessment. **Figure 1** sketches the basic structures of an evidence-centered approach to assessment design, formulated by Mislevy and colleagues at Educational Testing Service (ETS).

Working out these variables and models and their interrelationships is a way to answer a series of questions posed by Messick in the early 1990s that get at the very heart of assessment design:

- What complex of knowledge, skills, or other attributes should be assessed? A given assessment – formative or summative – is meant to support inferences for some purpose, such as a licensing decision, provision of diagnostic feedback, guidance for further instruction, or some combination. Variables in the competency model describe the knowledge, skills, and abilities on which the inferences are to be based. The term student model is often used to denote a student-instantiated version of the competency model. That is, values in the student model express the assessor's current belief about a student's level on variables within the competency model.
- What behaviors or performances should reveal those constructs? An evidence model expresses how the student's interactions with, and responses to, a given problem constitute evidence about student-model variables. Observable variables summarize aspects of specific task performances, and may come from either formative or summative sources. Then depending on the type and origin of the source(s) of evidence, different parameters are used to update the student model.
- What tasks or situations should elicit those behaviors? Task-model variables describe features of tasks or situations that will be used to elicit performance. A task model provides a framework for characterizing and constructing situations with which a student will interact to provide evidence about targeted aspects of knowledge. The task models will vary in line with the purpose of the assessment and its administration.

Within intelligent systems employing such evidence-based assessment, the student model would accumulate and represent belief about the targeted aspects of skill. These beliefs are often expressed as probability distributions

Table 1 Assessments' role in traditional vs. enhanced intelligent systems

Issue	Traditional Systems	Enhanced systems
Design methods based on evidentiary argument	Mostly absent	Present
Assessment focus	Mostly formative assessment	Formative and summative assessment
Links to standards	Mostly absent	Present
Measurement models	Largely <i>ad hoc</i>	Variegated, informed by advances in educational measurement and cognitive science
Evaluations	Mostly laboratory based	Classroom based
Role of teacher	Very limited or absent	Strong

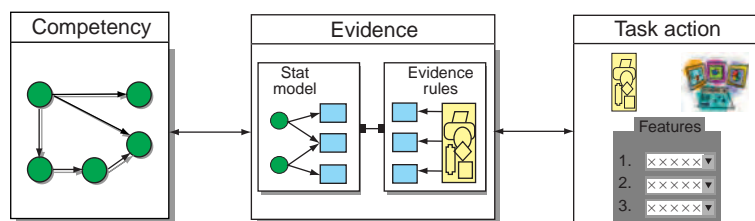


Figure 1 The three central models of an evidence-centered assessment design. Reproduced from Mislevy, R. J., Steinberg, L. S., and Almond, R. G. (2003). On the structure of educational assessment. *Measurement: Interdisciplinary Research and Perspective*, 1(1), 3–62.

for student-model variables. The way that this works in practice is that the evidence model extracts observables (e.g., scores) from student work (i.e., what the student says or does), and provides a way to aggregate scores that are then used to update the student model. In other words, the evidence model describes how evidence about a set of skills is connected to the competency-model variables using a psychometric model. Task models express situations that can evoke required evidence.

Based on the information in the student model, a viable approach to select and deliver content to the learner is needed – one that fits his or her needs at the time. This would provide context and coherence for delivering adaptive instruction, one of the main goals of intelligent systems. Following is an example of a model to support the select-and-deliver goal. It has been extended from the simpler two-process model that resides at the core of intelligent systems – diagnosis and prescription – and from a process model to support assessment, by Mislevy, Almond, and colleagues at ETS.

Four-Process Adaptive Cycle

The success of any intelligent system to promote learning requires accurate diagnosis of student characteristics (e.g., algebra knowledge, troubleshooting skill, and engagement). The collection of student information then can be used formatively for the prescription of optimal content, such as hints, explanations, hypertext links, practice problems, encouragement, metacognitive support, and so forth. Student information can also be used in a summative manner, such as providing reports on student achievement. The framework, described in this section, involves a four-process cycle connecting the student to appropriate educational materials and other resources (e.g., learning objects, peers, applications, and pedagogical agents) through the use of a student model, shown as the small human icon at the top of **Figure 2**.

The main components of this four-process cycle are (1) capture, (2) analyze, (3) select, and (4) present. The solid arrows in the figure show a normal, complete loop used in many intelligent systems, while the dashed arrows show variations of the cycle that have been used in other kinds of systems. For example, the dashed line that goes upward from the student (i.e., the large human icon in at the bottom of **Figure 2**) to the student model depicts an intelligent system where the student is allowed to interact directly with the student model. The nature of this interaction and the effects on the student model can vary, such as negotiating the value of a particular variable with the system or the teacher. The four processes are now briefly defined.

- **Capture.** The capture process entails gathering personal information about the student as he or she interacts with the environment. Relevant information, obtained

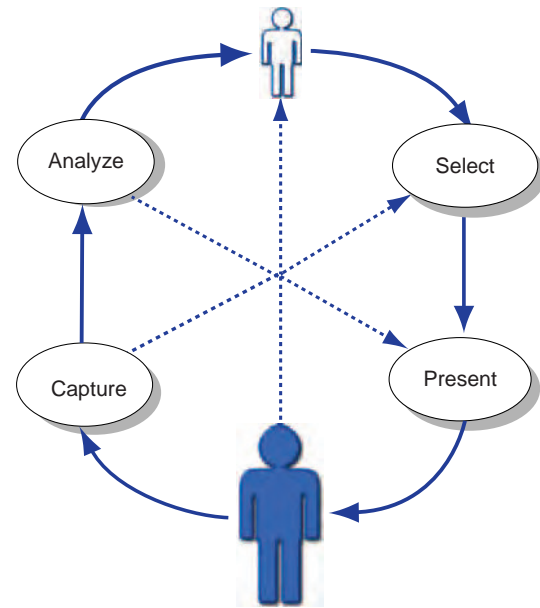


Figure 2 Four-process adaptive cycle. Reproduced from Shute, V. J. and Zapata-Rivera, D. (2008). Adaptive technologies. In Spector, J. M., Merrill, D., van Merriënboer, J., and Driscoll, M. (eds.) *Handbook of Research on Educational Communications and Technology*, 3rd edn, pp 277–294. New York: Lawrence Erlbaum Associates, Taylor & Francis Group.

through formative or summative assessment, can include cognitive as well as noncognitive aspects of the learner. This information is used to update the internal student model maintained by the system.

- **Analyze.** The analyze process requires the creation and maintenance of a student model by properly integrating evidence sources from student performance in the environment. This usually involves representing information in the student model through inference mechanisms in relation to students' proficiency states based on specific performance data.
- **Select.** Information (i.e., content in the broadest sense) is selected according to the model of the student maintained by the system and the goals of the system (e.g., next learning object or test item). This process is often required to determine how and when to intervene.
- **Present.** Based on results from the select process, specific content is presented to the learner. This entails appropriate use of different media, devices, and technologies to effectively and efficiently convey information to the learner.

Discussion

Educational measurement involves obtaining observations from students' responses in order to make inferences about their knowledge and skills. Reliable instruments are critical in this effort. The way in which information from assessments is captured, analyzed, and ultimately used, strongly influences the student model. First, the

granularity of the assessment information influences the kinds of assessment claims that the system can make at any point in time (e.g., general ability vs. component-skill level). Second, the type of evidence that is available from assessment tasks influences the reliability of the claims that the system can make about the student (e.g., data from calibrated test items are more reliable than data from, say, homework assignments). Third, the type of evidence available from assessment tasks influences the validity of the claims that the system can make about the student (e.g., the student is fully proficient at a particular skill vs. more evidence is needed to support the claim). Finally, the complexity of the assessment information – or the manner in which assessment information is interpreted – can range from very simple methods to quite sophisticated ones, such as using probability-based models. To effectively handle the diverse types of assessment information, from either formative or summative sources, an evidence-based assessment framework (e.g., evidence-centered design) is recommended.

Merging instruction with valid formative and summative assessment information in the form of intelligent systems opens up new possibilities and challenges that, if surmounted, could result in improved student learning and educational outcomes. New tools and methodologies are needed to get a more complete profile of the learner. These tools should focus on functions such as: (1) helping create student models that go beyond skills (e.g., conceptual understanding, social aspects of learning, emotional states, and other noncognitive attributes); (2) creating a framework to analyze incoming student work and update student model(s) properly (e.g., evidence-centered design (ECD), along with the four-process adaptive cycle); (3) informing the teacher, in easy-to-understand language, what to do next with the class and/or student and how to understand the data that are derived from the system; and (4) encouraging students to become more active and accountable for their own learning.

Early intelligent system design and development was typically accomplished without much input from psychometricians and/or assessment specialists. This may be partially explained by the chasm between traditional and emergent models of educational measurement. Current advances in intelligent systems, cognitive science, and educational measurement provide the opportunity for integrating assessment and instruction into powerful new intelligent systems. These new systems have the potential to improve student learning and also standardized test scores. Additional benefits of integrating formative and summative assessments with instruction include: (1) re-using rigorous assessment tasks and items from summative tests directly in the service of learning, and (2) increasing the chances of successful adoption of intelligent systems into mainstream education.

Further Reading

- Black, P. and Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice* 5(1), 7–74.
- Black, P. and Wiliam, D. (1998b). *Inside the Black Box: Raising Standards through Classroom Assessment*. London: School of Education, King's College.
- Conejo, R., Guzman, E., Millán, E., et al. (2004). SIETTE: A web-based tool for adaptive testing. *International Journal of Artificial Intelligence in Education* 14(1), 29–61.
- Linn, R. L., Baker, E. L., and Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher* 20(8), 15–21.
- Martin, J. D. and VanLehn, K. (1995). A Bayesian approach to cognitive assessment. In Nichols, P., Chipman, S., and Brennan, R. (eds.) *Cognitively Diagnostic Assessment*, pp 141–165. Hillsdale, NJ: Erlbaum.
- Messick, S. (1992). The interplay of evidence and consequences in the validation of performance assessments. *ETS Research Report, RR-92-39*, pp 1–42. Princeton, NJ: Educational Testing Service.
- Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 257–305. Westport, CT: Greenwood.
- Mislevy, R. J. and Gitomer, D. H. (1996). The role of probability-based inference in an intelligent tutoring system. *User-Modeling and User-Adapted Interaction* 5, 253–282.
- Mislevy, R. J., Steinberg, L. S., and Almond, R. G. (2003). On the structure of educational assessment. *Measurement: Interdisciplinary Research and Perspective* 1(1), 3–62.
- National Research Council (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*, Committee on the Foundations of Assessment, Pellegrino, J., Glaser, R., and Chudowsky, N. (eds.). Washington, DC: National Academy Press.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Kaufmann.
- Razzaq, L., Feng, M., Heffernan, N. T., et al. (2007). Blending assessment and instructional assistance. In Nedjah, N., deMacedo Mourelle, L., Borges, M. N., and Almeida, N. N. (eds.) *Intelligent Educational Machines within the Intelligent Systems Engineering Book Series*, pp 23–49. Berlin: Springer.
- Shute, V. J. (1995). SMART: Student modeling approach for responsive tutoring. *User-Modeling and User-Adaptive Interaction* 5(1), 1–44.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research* 78(1), 153–189.
- Shute, V. J. (2007). Tensions, trends, tools, and technologies: Time for an educational sea change. In Dwyer, C. A. (ed.) *The Future of Assessment: Shaping Teaching and Learning*, pp 139–187. New York: Erlbaum/Taylor and Francis.
- Shute, V. J., Hansen, E. G., and Almond, R. G. (2007). An assessment for learning system called ACED: Designing for learning effectiveness and accessibility. *ETS Research Report*. Princeton, NJ: Educational Testing Service.
- Shute, V. J. and Psotka, J. (1996). Intelligent tutoring systems: Past, present, and future. In Jonassen, D. (ed.) *Handbook of Research for Educational Communications and Technology*, pp 570–600. New York: Macmillan.
- Shute, V. J. and Zapata-Rivera, D. (2008). Adaptive technologies. In Spector, J. M., Merrill, D., van Merriënboer, J., and Driscoll, M. (eds.) *Handbook of Research on Educational Communications and Technology*, 3rd edn, pp 277–294. Mahwah, NJ: Erlbaum.
- Snow, R. E. and Mandinach, E. B. (1991). *Integrating Assessment and Instruction: A Research and Development Agenda*. Princeton, NJ: Educational Testing Services.
- VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education* 16(3), 227–265.
- Zapata-Rivera, D., VanWinkle, W., Shute, V. J., Underwood, J. S., and Bauer, M. I. (2007). English ABLE. *Proceedings of the 13th International Conference of Artificial Intelligence in Education*, 8p.

Item Response Theory

W J van der Linden, CTB/McGraw-Hill, Monterey, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Ability parameter – Parameter in a response model that represents the person's ability, skill, or proficiency measured by the test items.

Adaptive testing – Mode of testing with real-time ability estimation in which each item is selected to be optimal at the last estimate.

Cognitive-component models – Response models with parameters for the individual cognitive operations required to solve the test item.

Dichotomous response models – Response models for items with dichotomously scored responses, such as correct–incorrect, true–false, and agree–disagree items.

Hierarchical response models – Response models with random item and/or person parameters for which a distribution is specified.

Information function – Fisher's measure for the information in an item or test score as a function of the ability parameter.

Item parameter – Parameter in a response model that represents a property of the item, such as its difficulty and discriminating power.

Multidimensional response models – Response models with a multidimensional ability parameter.

Optimal test design – Use of techniques from mathematical programming to design tests that are optimal with respect to an objective function and meet the set of constraints representing its content specifications.

Polytomous response models – Response models for items with polytomously scored responses.

Response function – Probability of a specific response on an item as a function of the ability parameter.

Item response theory (IRT) has many roots, and it would be wrong to single out anyone as the most important. But the contributions by Louis Thurstone (1925) have been decisive in many respects. These contributions built on Binet's intelligence test developed to separate students in Paris schools that needed special education from regular underachievers (Binet and Simon, 1905) as well as earlier work in psychophysics by Fechner.

Although Binet's test has been hailed as the first to be a fully standardized test in the history of educational and

psychological measurement, his most significant contribution was his idea to scale the items before using them in the test. Of course, there is no natural intelligence scale, and neither were there any points of view to look at the matter in Binet's days. Binet's solution to the problem, however, was equally simple as ingenious: he chose chronological age as the scale on which he mapped his items and students. Using an empirical pretest, he defined the scale value of an item as the age group of which 75% of its members had solved it correctly. These scale values were then used to estimate the age group for which the student's achievements on the test were representative. The age of this group was the mental age of the student.

The chronological age scale used by Binet allows for direct measurement. Thurstone's innovation was to put Binet's items on an intelligence scale that cannot be measured directly. He did so because he recognized that intelligence is an example of what is now known as a latent variable; that is, an unmeasured hypothetical construct. The scale of this variable was defined by postulating a normal distribution for each age group and inferring the scale values of the items from the response data. Thurstone also showed how to check this distributional assumption. The assumption of a normal cumulative distribution function as a response function was not new but borrowed from the earlier work on psychophysics by Fechner, who used them to describe how psychological sensations vary with the strength of experimentally manipulated physical stimuli. But Thurstone's idea to separate intelligence from age and define it as a latent variable with a scale defined by such response functions was entirely new.

The idea of response functions on a latent variable was picked up again by authors like Ferguson, Lawley, and Mosier in the 1940s (and led to much confusion between the use of the normal ogive as a definition of a population distribution and a response function on a latent variable). But we had to wait until the seminal work by Lord (1952) and Rasch (1960) until the developments really began. From a more statistical point of view, later contributions by Birnbaum (1968) were important. He replaced the normal ogive by the logistic function, introduced additional item parameters to account for guessing on items (which is typical of most educational measurements), derived maximum-likelihood estimators for the model, and showed how to assemble tests from a bank of calibrated items to meet optimal statistical specifications for their application.

The next two decades showed much research on IRT models for test items with other response formats than simple dichotomous scores as well as on newer procedures for parameter estimating and model evaluation. Especially, the development of Bayesian procedures for parameter estimation and model validation was prominent. When computers became more powerful and cheaper in the 1980s, the routines for maintaining common scales and test assembly used in most large-scale educational testing programs became IRT based. The first programs to exploit IRT to score test takers in real time and deliver computerized adaptive tests were launched in the 1990s. Nowadays, IRT models and procedures are no longer the main instruments only in the educational testing industry but are becoming increasingly popular in psychological testing as well. In addition, they have hit areas such as medical diagnosis and marketing research.

Review of Response Models

Unidimensional Logistic Models for Dichotomous Items

Due to Thurstone's pioneering work, the assumption of a normal-ogive function as response function for test items remained popular for a long time. Lord's (1952) treatment of IRT for dichotomously scored items was also based on this assumption. Let U_i denote the response to item i , where $U_i = 1$ if the response on item i is correct and $U_i = 0$ if it is incorrect. The normal-ogive model describes the probability of a correct response as a function of the test taker's ability θ :

$$p_i(\theta) = \Pr\{U_i = 1|\theta\} = \int_{-\infty}^{a_i(\theta - b_i)} \frac{1}{\sqrt{2\pi}} \exp^{-z^2/2} dz. \quad [1]$$

where $\theta \in (-\infty, \infty)$ and $b_i \in (-\infty, \infty)$ and $a_i \in (0, \infty)$ are parameters for the difficulty and discriminating power of item i .

Observe that the model has its parameter structure in the upper limit of the integral. For this reason, it was rather intractable at the time, and hardly used in routine applications in testing. The model also seemed to ignore the possibility of guessing on test items. These two points were addressed in Birnbaum's (1968) model, which is now generally known as the three-parameter logistic (3PL):

$$p_i(\theta) = \Pr\{U_i = 1|\theta\} = c_i + (1 - c_i) \frac{\exp(a_i(\theta - b_i))}{1 + \exp(a_i(\theta - b_i))} \quad [2]$$

where additional parameter $c_i \in [0, 1]$ represents the height of a lower asymptote for the probability of a correct response. The asymptote is approached for $\theta \rightarrow -\infty$. This limit is assumed to represent random guessing without any knowledge.

The two models have no fixed unit and origin for the scale of θ . In practice, we therefore fix the scale following a practical convention. All later models in this article need comparable constraints to become identifiable. For an appropriate choice of scale unit, the models in eqns [1] and [2] predict success probabilities for identical sets of parameter values that are virtually indistinguishable from each other.

A graphical example of the logistic response function for a set of item parameter values is given in **Figure 1**. Difficulty parameter $b_i = 1$ controls the location of response function along the θ scale. A more difficult item has its location to the right of $\theta = 1$ and requires a higher ability to give the same probability of success. Discrimination parameter $a_i = 1.4$ controls the slope of the response function. A more discriminating item has a response function with a steeper slope than in **Figure 1**. Finally, a value of 0.23 for discrimination parameter c_i indicates the height of the lower asymptote for the response function. Typically, for multiple-choice items, the estimated values for this guessing parameter are close to the reciprocal of their number of response alternatives.

The two-parameter logistic (2PL) model is obtained if we put $c_i = 0$ in eqn [2]. The result is the logistic analog of eqn [1]. If we also assume equal values for the discrimination parameter (that is, $a_i = a$ for all i), the Rasch (1960) model is obtained. This model belongs to the exponential family of distributions in statistics and is statistically quite tractable.

Due to its parameter structure, the 3PL model is flexible and has been shown to fit large pools of items written for the same content domain in educational and psychological testing. In fact, it has become the standard of the testing industry for tests with dichotomous items

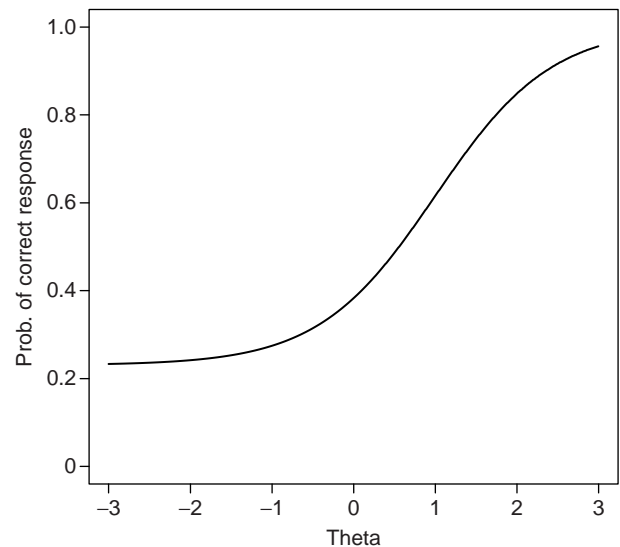


Figure 1 Example of a response function for the 3-parameter logistic model in eqn [2].

that measure a unidimensional ability. For more narrowly defined domains of educational items, or psychological tests of well-defined constructs, the Rasch model becomes an attractive alternative.

Models for Polytomous Items

Test items have a polytomous format when the responses are scored in more than two categories. Such models have a response function for each different category. **Figure 2** shows a typical set of response functions for an item with five different categories for one of the polytomous response models below.

Although the response categories of dichotomous items typically have a fixed order (e.g., correct–incorrect; true–false), this does not necessarily hold for polytomous items. Polytomous items have a nominal response format if their response categories can be classified but an *a priori* order between them does not exist. The nominal response model below is appropriate for items with this format. If an *a priori* ordering of the categories does exist, a graded-response model or partial-credit model should be chosen. The two models differ in the problem-solving process that is assumed to lead to the responses.

Models for a graded-response format are more generally known as cumulative models in ordinal categorical data analysis. The partial-credit models below are known as adjacent-category models. These two options do not exhaust all possibilities, a more comprehensive review of different polytomous response formats and IRT models is given in Mellenbergh (1994).

Nominal-response model

Response variable U_i is now assumed to have possible values $b = 1, \dots, m_i > 2$. Observe that different items in

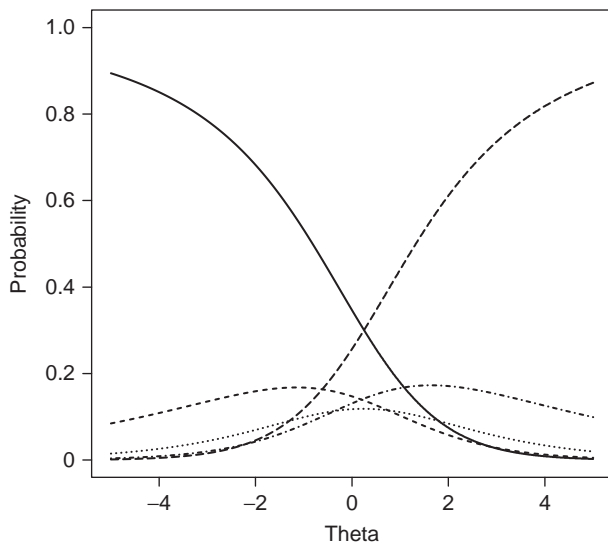


Figure 2 Example of response functions for a polytomous response model with five categories.

the same test may have different numbers of possible responses. According to the nominal-response model (Bock, 1972), the response function for category b is

$$p_{ib}(\theta) = \Pr\{U_i = b|\theta\} \quad [3]$$

$$= \frac{\exp(a_{ib}(\theta - b_{ib}))}{\sum_{b=1}^{m_i} \exp(a_{ib}(\theta - b_{ib}))}. \quad [4]$$

Parameters a_{ib} and b_{ib} maintain their interpretation as a discrimination and difficulty parameter. **Figure 2** illustrates that b_{ib} represents the location of the response function for category b along the ability scale, where the location is defined as the value of θ at which the function shows its largest change in curvature. The value of the discrimination parameters a_{ib} is proportional to this change.

In spite of the nominal response format, the values for the parameters b_{ib} , $b = 1, \dots, m_i$, do imply an order for the response categories. However, the actual order is only known when these parameters are estimated; it is thus empirical, not *a priori*. This feature makes the model less suitable for educational measurement, where performances on test items can always be ordered from worse to better in advance. Another reason why the model in eqn [4] may be less appropriate for this application is that it does not allow explicitly for guessing. A version of the nominal response model that does allow for guessing is given in Thissen and Steinberg (1997).

Graded-response model

Suppose index b reflects an *a priori* ordering of the response categories. The graded-response model (Samejima, 1969) addresses the probabilities of the compound events $U_i \geq b$,

$$P_{ib}(\theta) = \begin{cases} 1 & \text{for } b = 1 \\ \Pr\{U_i \geq b|\theta\} & \text{for } b = 2, \dots, m_i \\ 0 & \text{for } b > m_i \end{cases} \quad [5]$$

as a function of ability parameter θ . The more interesting probabilities are those for $b = 2, \dots, m_i$. They increase monotonically with θ because the probability of responding in any of these categories or higher goes to one if $\theta \rightarrow \infty$.

A typical choice for $P_{ib}(\theta)$ for $b = 2, \dots, m$ in eqn [5] is from the logistic functions in eqn [2] for $c_i = 0$. If the parameters a_i are free, the result is known as the nonhomogeneous case of the model. If we set $a_i = 1$ for all i , a more stringent version of the graded-response model is obtained, which is known as its homogenous case.

The response functions for the individual categories $b = 2, \dots, m_i$ can be derived from eqn [5] as

$$p_{ib}(\theta) = P_{ib}(\theta) - P_{i(b+1)}(\theta). \quad [6]$$

The shape of these response functions may not differ much from those for the nominal-response model.

The most distinctive feature of these functions, however, is that they reflect an *a priori* order, which is specified by the test specialist when assigning the labels $b = 1, \dots, m_i$ to the categories. These labels determine the way in which the differences in eqn [6] are calculated.

Partial-credit models

These models derive their name from an increasing credit given to the responses $b = 1, \dots, m_i$. All models are based on a sequential response process in which the test taker is assumed to compare two adjacent categories at a time and decide to prefer one category over the other with a certain probability.

The first version of the partial-credit model (Masters, 1982) was based on the assumption that the probability of preferring response b relative to $b - 1$ can be represented by the Rasch model. Upon some probability calculus, this assumption can be shown to lead to the following response functions for $b = 1, \dots, m_i$:

$$p_{ib}(\theta) = \frac{\exp\left(\sum_{k=1}^b (\theta - b_{ik})\right)}{\sum_{b=1}^{m_i} \exp\left(\sum_{k=1}^b (\theta - b_{ik})\right)}. \quad [7]$$

A generalized version of the model exists in which the probabilities of adjacent responses are based on the 2PL model with free discrimination parameters a_i (Muraki, 1992). The model therefore has parameter structures $(\theta - b_{ib})$ extended to $a_i(\theta - b_{ib})$.

Further, if we adopt a common number of categories m for all items, the generalized partial-credit model can be specialized to a model for a set of rating scales of the Likert type. Likert scales are well known in attitude measurement, where they are used to ask subjects to evaluate a set of attitude statements using scales with common categories such as strongly agree, agree, neutral, disagree, and strongly disagree. The steps necessary to obtain the rating scale model steps are: (1) decomposing the parameters b_{ib} additively as $b_i + d_k$, with b_i a location parameter for the entire item and d_k a threshold parameter for k th category on the Likert scale, and (2) constraining the discrimination parameters to special known constants. This rating scale model was introduced by Andersen (1977) and Andrich (1978).

Multidimensional Models

Test items may measure more than one ability, for example, a verbal ability in addition to the ability in another domain of achievement that is tested more explicitly. If this happens, the previous models have to be extended by more ability parameter. For the 3PL model in eqn [2], the extension with a second ability parameter may lead to response functions

$$p_i(\theta_1, \theta_2) = c_i + (1 - c_i) \frac{e^{a_{i1}\theta_1 + a_{i2}\theta_2 - b_i}}{1 + e^{a_{i1}\theta_1 + a_{i2}\theta_2 - b_i}}. \quad [8]$$

This model defines the probability of a correct response as a function of (θ_1, θ_2) . It has two discrimination parameters, a_{i1} and a_{i2} , which control the slope of the response surface along θ_1 and θ_2 , respectively. But it has only one (generalized) difficulty parameter b_i ; a similar model with two parameters $b_{i1} - b_{i2}$ would be nonidentifiable version of it. If $\theta_1, \theta_2 \rightarrow -\infty$, the probability of a correct response goes to c_i . The role of these parameters becomes clear if we view the graphical example of the response surface in Figure 3 as a two-dimensional generalization of the response function in Figure 1.

Cognitive-Component Models

IRT models in this class specify the probability of success on a test item as a function of the components of the cognitive task involved in solving the item. Two main types of models have been used. Both types agree in that they decompose the difficulty parameter as a set of more basic parameters for the relevant components. They differ, however, in how to impose a structure on these basic parameters.

Let $U_{ik} = 0, 1$ denote whether or not component $k = 1, \dots, K$ of the task involved in solving item i has been executed successfully. Suppose a correct response to item i requires a successful execution of each of these components. The probability of a correct response for this conjunctive structure is then given by the product of the probabilities for each of the components,

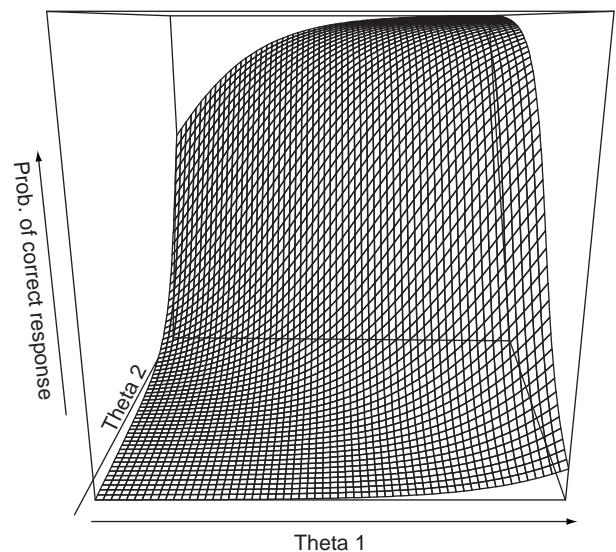


Figure 3 Example of a two-dimensional response surface for the logistic model in eqn [8].

$$\prod_{k=1}^K \Pr\{U_{ik}\}. \quad [9]$$

Models of this type were introduced by Embretson (1997). Her choice of success probabilities for the component tasks was the Rasch model with free ability parameters, θ_k , and difficulty parameters, β_{ik} , for $k = 1, \dots, K$. She also introduced a guessing parameter for the case a test taker misses a component and has to guess to the answer on the item.

A different approach is offered in the linear logistic model introduced by Fischer (1983). In this model, the structure is not imposed on the probabilities of successful completion of the component tasks, such as for the conjunctive structure in eqn [9], but directly on their difficulty parameters. Assuming the same set of component tasks $k = 1, \dots, K$ to hold for item i , the item-difficulty parameter is taken to be the following linear function of the component parameters:

$$b_i = \sum_{k=1}^K w_{ik}\beta_{ik} + c$$

where w_{ik} is the weight set by a content specialist to represent the impact of component difficulty β_{ik} on the difficulty of item i .

The idea of modeling underlying processes rather than the responses themselves has been a fruitful area of research. Its spirit is well captured by the name explanatory item response modeling given to the area by De Boeck and Wilson (2005).

Hierarchical Models

Hierarchical or multilevel modeling of responses is appropriate when the test takers can be assumed to be sampled from a population and the interest is in some of the features of its ability distribution. This application arises, for example, in large-scale educational assessments, where the development in the educational achievements of certain populations should be compared or followed over time. Alternatively, the interest may exist in test items that can be considered as randomly sampled, for example, from families of items generated from a set of parent items, as in the use of techniques of item cloning (Glas and van der Linden, 2003). In this case, to allow for the random sampling of items from different families, the distributions of their parameters in these families have to be modeled. Of course, it is possible to have applications in which the distributions of both the person and item parameters have to be specified.

As a first-level model, typically one of the earlier models is chosen, for example, the 3PL model for a correct response in eqn [2]. A second-level model for the ability distribution is the normal

$$\theta \sim N(\mu, \sigma^2) \quad [10]$$

with mean μ and variance σ^2 .

If the items are sampled from families $f = 1, \dots, F$, the second-level models for their distributions can be taken to be the multivariate normals

$$(a_{if}, b_{if}, c_{if}) \sim MVN(\mu_f, \Sigma_f) \quad [11]$$

where μ_f and Σ_f are the mean and covariance matrix for the item parameters of family f .

These models become more powerful if we are able to explain the second-level distributions of the abilities or item parameters by background variables or covariates, for example, group memberships of the test takers or structural features of the item families. This can be done by introducing linear regression structure for the means in eqn [10] or eqn [11]. Following the spirit of hierarchical linear modeling, the regression parameters can also be defined to be random at a higher level of modeling (Fox and Glas, 2001). In large-scale educational assessments, the introduction of such regression structures helps us to pinpoint differences in achievement between specific populations.

Other Models

The previous response models belong to the most important categories in the literature as well as the practice of educational measurement. But they do certainly not exhaustive the possibilities. To date, some 30–40 different response models have been proposed that are statistically tractable. The collection includes models for nonmonotone items. An item is called monotone if the function for its correct response is monotonically increasing with the ability measured by the items. This feature is typical of achievement test items but cannot be expected to hold for items in attitude scales or preference measurement, where respondents may endorse the same alternative for opposite reasons. The current review, moreover, does not include nonparametric approaches to IRT modeling or models for response times on test items. Nonparametric approaches avoid the assumption of a parametric family of response functions but try to use order assumptions with respect to probabilities of success, persons, and/or items only, which may vary in their degree of stringency. In addition, polytomous versions of nonparametric models exist (for a review, see Sijtsma and Molenaar, 2002). Response-time models have become important because of the increasing popularity of computerized delivery of educational tests with its automatic recording of the response times on the individual items. In order to use these times as an additional source of information on the persons or items, a response model with a typical IRT parameterization is necessary (van der Linden, 2006, 2009).

For more extensive introductions to larger collections of response models, the reader should consult Fischer and Molenaar (1995) and van der Linden and Hambleton (1997).

Applications in Educational Measurement

Item Calibration and Measurement

The use of IRT models in educational measurement typically consists of the stages of item calibration and measurement. During item calibration, response data are collected for a representative sample of test takers whereupon the item parameters are estimated from the data and the validity of the model is evaluated. The process of model evaluation may involve the checking of the model against the data for features such as the match between the dimensionality of the data and the ability parameters in the model, the assumption of local independence required to estimate the parameters, possible systematic differences in item parameters between relevant subpopulations (differential item functioning), and the general shape of the response functions. Additionally, it should be checked if the test takers tend to behave according to the response model. All checks take the form of statistical hypothesis testing, where the alternative hypothesis is the specific model violation against which the model is tested. For a more comprehensive treatment of these tests, see Glas and Meijer (2003) and Glas and Suarez Falcon (2003).

If the model fits and the item parameters are estimated with enough precision, the model can be used as measurement model. The test is then administered as part of operational testing and the person parameter is estimated for each test taker equating the item parameters to their estimates. The estimate is reported as the test taker's ability score. (Following a Bayesian approach, it would be more appropriate not to treat the item parameters as completely known but to account for their estimation error when estimating the person parameter. But this practice is rather unusual in educational measurement.)

Person parameters are typically estimated using the method of maximum likelihood (ML) or a Bayesian method. For the dichotomous model in eqn [2], these estimates are obtained as follows. Let (u_1, \dots, u_n) be the vector with the responses for a person on the items $i = 1, \dots, n$ in the test that is used to measure the person's ability, θ . In ML estimation, the estimate of θ is calculated from the probability of the response vector under the model taken as a function of θ . More specifically, this function is known as the likelihood function of θ associated with the responses,

$$L(\theta|u_1, \dots, u_n) = \prod_{i=1}^n p_i(\theta)^{u_i} [1 - p_i(\theta)]^{1-u_i}, \quad [12]$$

and the ML estimate is the value of θ for which this likelihood function reaches its maximum.

In a Bayesian approach, θ is estimated by its posterior distribution; that is, its probability distribution given the responses. The distribution is obtained by assuming a prior distribution of θ and combining this with its likelihood as:

$$f(\theta|u_1, \dots, u_n) = cL(\theta|u_1, \dots, u_n)f(\theta), \quad [13]$$

where $f(\theta)$ is the density of the prior distribution of θ , $f(\theta | u_1, \dots, u_n)$ is the density of its posterior distribution, and c is a normalizing constant. If a point estimate is needed, the mean of the posterior distribution can be taken (expected *a posteriori* or EAP estimation).

The power of IRT for educational measurement lies in the presence of the item parameters in the expression in eqns [12] and [13] from which the ability estimate is calculated. Due to this, the estimation automatically accounts for the properties of the items that were selected in the test. For example, if the items would have been more difficult, the values of the item difficulty parameters would have been greater, and the ability estimate for the response vector would automatically have been increased to account for this. This feature has been called item-free measurement – an expression that, when taken naively, seems to capture this feature nicely but is somewhat misleading because the statistical features of the estimate (e.g., its accuracy) still depend on the chosen items.

Another way of touting the power of IRT is by pointing at the fact that it is able to produce valid item calibration or educational measurement from response data collected using research designs with missing data. The applications in the next sections capitalize on this feature. Of course, the validity of parameter estimates from incomplete designs is only guaranteed if the fact whether or not the responses are missing does not contain any direct information on their correctness. It would be wrong to leave out a portion of the responses, for instance, because they were incorrect.

Specific Applications with Missing Data

Most of the large-scale educational testing programs now use item banking. In item banking, new items are written, pretested, and calibrated continuously. If an item passes all quality checks, it is added to the item bank. At the same time, new tests are assembled from the items in the bank. This practice differs dramatically from traditional test construction, which goes through a complete cycle of item writing, pretesting, and test construction for one test at a time. Obvious advantages of item banking are more constant use of the resources, permanent access to a large stock of pretested, high-quality items for test assembly, and less vulnerability to premature leaking of a test form. Its main advantage is, however, stable score scales defined by larger pools of items. Such scales permit testing programs to

produce scores that are comparable over time even when earlier stocks of items in the bank have been retired and replenished.

Item banking is possible because when using IRT, it is no longer necessary for two persons to answer identical collections of test items to have comparable scores. Likewise, in order to compare the parameter estimates of different items, it is no longer necessary for them to be taken by the same sample of test takers. It is therefore possible to build up banks with much less restricted versions of data-collection designs than in traditional pre-testing of test items. Likewise, once the items have been calibrated, we can deliberately select a subset of items for a specific measurement goal.

The idea of optimal test assembly capitalizes on the latter opportunity. One of the first to point at it was Birnbaum (1968). His idea was to translate the measurement goal into a target function for the accuracy of the test along the ability scale. The test was then assembled to match the target as closely as possible. For example, the test could be required to have a uniform target over a certain ability range when it is used for diagnostic purposes with respect to students in it, or a peaked function at a cutoff score when the goal is accurate admission decisions.

For this approach to be practical, it is necessary to have a measure of the accuracy of the test as a function of the ability θ measured by the items. A useful measure is Fisher's information in the item responses in the test, which for the 3PL model in eqn [2] can be shown to be equal to

$$I_i(\theta) = \frac{a_i^2 [1 - p_i(\theta)] \left[\frac{p_i(\theta) - c_i}{1 - c_i} \right]^2}{p_i(\theta)} \quad [14]$$

for item i . Taken as a function of θ , the measure is generally referred to as the information function of item i . The use of this measure for optimal test assembly is motivated by two different facts: First, information functions are additive. That is, if the items $i = 1, \dots, n$ are selected for a test, test information function $I(\theta)$ can be shown to be equal to the sum of the item information functions,

$$I(\theta) = \sum_{i=1}^n I_i(\theta). \quad [15]$$

It is thus straightforward to evaluate the effects of adding or removing an item to a test. Second, the sampling variance of the ML estimator of θ for a given test is asymptotically equal to the inverse of the information function in eqn [15]. This feature gives the use of the information function its statistical foundation. **Figure 4** illustrates the additivity of the information functions for a test of six items.

Although optimal test assembly can be illustrated graphically rather easily for a short test, the actual assembly of a test in a real-world application is no simple affair.

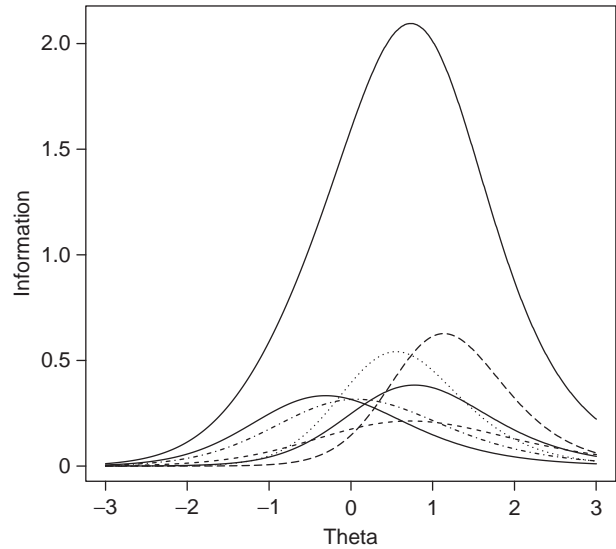


Figure 4 Example of six item information functions and their test information function.

In many applications no single tests but sets of parallel tests are to be selected. More importantly, tests always have to be assembled to sets of constraints that control their content specifications. In fact, it is not unlikely for real-world tests to be assembled to several hundreds of such constraints. Fortunately, such problems can be modeled to be a linear integer programming problem, which can easily be solved using standard commercial software. For solutions to a large variation of optimal test assembly problems with various objective functions and types of constraint, see van der Linden (2005).

In computerized adaptive testing (CAT), the items in the test are not selected simultaneously for a group of test takers but sequentially from the item pool for an individual test taker. The responses on the items are recorded and used to update the person's ability estimate in real time. Each next item in the test is selected to be optimal at the last estimate. Due to this adaptation, the ability estimate converges much faster to the person's true ability than for a traditional fixed test, even when it has been assembled to be optimal for a group of test takers.

In order to implement adaptive testing, a criterion for item selection is required. An obvious criterion is to use the information function $I_i(\theta)$ in eqn [14] and select the next item to have a maximum value for it at the last ability estimate among the items in the pool. This popular criterion of item selection is known as the maximum-information criterion. Alternatively, Bayesian item-selection criteria can be used, for example, a criterion that minimizes the variance of the posterior distribution of the test taker's ability in eqn [13].

Other practical issues that have to be addressed when implementing adaptive testing are how to impose a fixed set of content constraints on the test for different test

takers in real-time item selection, how to prevent different time pressure between test takers that get different selections of items, and how to maintain the integrity of the item pool against test takers that cheat and try to memorize and share test items. IRT-based solutions to these problems are reviewed in van der Linden (2005) and van der Linden and Glas (2010).

Bibliography

- Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika* **42**, 69–81.
- Andrich, D. (1978). A rating formulation for ordered categories. *Psychometrika* **43**, 561–573.
- Binet, A. and Simon, T. A. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologie* **11**, 191–336.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord, F. M. and Novick, M. R. (eds.) *Statistical Theories of Mental Test Scores*, pp 397–479. Reading, MA: Addison-Wesley.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* **37**, 29–51.
- De Boeck, P. and Wilson, M. R. (2005). *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York: Springer.
- Embretson, S. E. (1997). Multicomponent response models. In van der Linden, W. J. and Hambleton, R. K. (eds.) *Handbook of Modern Item Response Theory*, pp 305–321. New York: Springer.
- Fischer, G. H. (1983). Logistic latent trait models with linear constraints. *Psychometrika* **48**, 3–26.
- Fischer, G. H. and Molenaar, I. W. (eds.) (1995). *Rasch Models: Foundations, Recent Developments and Applications*. New York: Springer.
- Fox, G. J. A. and Glas, C. A. W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika* **66**, 271–288.
- Glas, C. A. W. and Meijer, R. R. (2003). A Bayesian approach to person-fit analysis in item response theory models. *Applied Psychological Measurement* **27**, 217–233.
- Glas, C. A. W. and Suarez Falcon, J. C. (2003). A comparison of item-fit statistics for the three-parameter logistic model. *Applied Psychological Measurement* **27**, 87–106.
- Glas, C. A. W. and van der Linden, W. J. (2003). Computerized adaptive testing with item clones. *Applied Psychological Measurement* **27**, 247–261.
- Lord, F. M. (1952). A theory of test scores. *Psychometrika Monographs* No. 7.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika* **47**, 149–174.
- Mellenbergh, G. J. (1994). Generalized linear item response theory. *Psychological Bulletin* **115**, 300–307.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement* **16**, 159–176.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research.
- Samejima, F. (1969). Estimation of ability using a pattern of graded responses. *Psychometrika Monograph* **17**, 1–100.
- Sijsma, K. and Molenaar, I. W. (2002). *Introduction to Nonparametric Item Response Theory*. Thousand Oaks, CA: Sage.
- Thissen, D. and Steinberg, L. (1997). A response model for multiple-choice items. In van der Linden, W. J. and Hambleton, R. K. (eds.) *Handbook of Modern Item Response Theory*, pp 51–65. New York: Springer.
- van der Linden, W. J. (2005). *Linear Models for Optimal Test Design*. New York: Springer.
- van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics* **31**, 181–204.
- van der Linden, W. J. (2009). Conceptual issues in response-time modeling. *Journal of Educational and Measurement* **46**, 247–272.
- van der Linden, W. J. and Glas, C. A. W. (eds.) (2010). *Elements of Adaptive Testing*. New York: Springer.
- van der Linden, W. J. and Hambleton, R. K. (1997). *Handbook of Modern Test Theory*. New York: Springer.

Further Reading

- Bock, R. D. (1997). A brief history of items response theory. *Educational Measurement: Issues and Practice* **16**, 21–33.
- Hambleton, R. K. and Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Boston, MA: Kluwer.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum.

National Assessments

T J Crooks, University of Otago, Dunedin, New Zealand

© 2010 Elsevier Ltd. All rights reserved.

National assessment involves gathering and presenting information on national patterns of educational outcomes. The focus in this article is on assessment within primary and secondary education. Most commonly, national assessments concentrate on student achievement outcomes, such as knowledge and skills, but some national assessment programs focus on affective outcomes such as motivation, attitudes and dispositions as well.

National assessment programs have their core focus on educational outcomes for the whole school population at selected year levels. However, most also present information about outcomes for subgroups within that population, such as boys and girls, students with different cultural or ethnic affiliations, students attending schools with different characteristics, and students from different regions. This more differentiated information can be useful in identifying policy needs associated with particular groups or the progress being made towards achieving equity of educational outcomes.

Functions

National assessment programs may have control, accountability and diagnostic functions. The balance among these for a particular program is determined largely by the design of the program and the prominence that it achieves.

National assessment programs can exercise control over participants in the education system through the emphasis given to particular educational outcomes. If particular subsets of curriculum objectives are selected as the focal areas for national assessment, and participants feel strong pressure to demonstrate good performance in those areas, the designers of the national assessment program can influence, quite strongly, the curriculum coverage and emphases of administrators, teachers, and students, and even the teaching and learning styles adopted. In federal systems, national assessments have the potential to alter the balance of control between national and state or regional education authorities.

Accountability involves taking or assigning responsibility within an enterprise for achieving particular outcomes. In education systems, national assessment provides information that can be used to decide if desired outcome targets are being met, and can therefore play a significant role in the accountability of agencies or individuals with responsibilities in those areas.

A third function of national assessment is diagnosis: identifying strengths and weaknesses in educational outcomes and possible reasons for the patterns identified. Good diagnoses suggest the next actions to be taken to achieve the most desired outcomes.

These three functions are not independent, or even necessarily competing. For instance, accountability mechanisms are usually central to the effectiveness of control strategies, and diagnostic information can be very useful to help individuals or agencies to meet accountability requirements.

Models for National Assessment

Participation in International or Regional Assessment Programs

One way that many countries gain national assessment information is by participating in major international or regional multicountry assessment programs. For instance, at least 40 countries participated in each of the most recent international assessments of the Progress in International Reading Literacy Study (PIRLS), the Trends in International Mathematics and Science Study (TIMSS), and the Programme for International Student Assessment (PISA).

PIRLS, TIMSS, and PISA are open to countries throughout the world, but there are also multinational assessment programs that are restricted to countries within a given region. For instance, many countries have participated in assessments organized by the Southern African Consortium for Monitoring Educational Quality (SACMEQ), while countries further north in Africa have made use of the francophone assessments of the *Programme d'analyse des systèmes de la CONFEMEN* (PASEC).

For national assessment purposes, participation in such programs has significant advantages. The consortia involved have access to strong technical expertise that is not available in many single nations, and the processes are highly developed and systematic. The results give comparative information between nations as well as substantial information for individual national reports.

There are, however, significant disadvantages also. Because of the different curricula of different countries, the assessment items included in an international assessment represent a compromise between the curricula and wishes of the different countries involved, so that it may not fit really well with the curriculum of a particular country. Moreover, the security and technical requirements

associated with these assessments often place restrictions on the number of items that can be revealed in national reports, and therefore on the specificity and usefulness of the reports to educators.

National Testing of Student Cohorts

A second way that countries gain national assessment information is by testing entire national cohorts of students. In this approach, the testing usually also serves other purposes, such as providing systematic information on the achievement of individual students and the collective achievement of students in different classes or schools.

This approach appears to be attractive if all students in a given grade or year cohort are going to be tested anyway. All that is required is to aggregate the results from individual students and schools nationally and analyze them in the ways required for national assessment. This can be completed quite economically.

There are two main dangers for national assessment with this approach. The first is that the large number of students to be assessed requires that the cost per student is kept quite low. This limits the range of educational outcomes that can be assessed. For instance, the testing might be administered by the students' usual class teachers and consist entirely of multiple-choice items that can be marked by scanner and computer. This would preclude assessment of many outcomes that involve open-ended responses or performances. The second danger is that the reporting of results for individual students and schools can make these assessments somewhat high-stakes ones, especially if there is public reporting of the comparative performance of different schools or school districts and rewards and sanctions for high or low performance. As a result of the pressure this places on teachers and administrators, the national assessment results are likely to be distorted by schools placing special emphasis on preparing the students for the tests, or perhaps even manipulating to some degree who takes the tests.

Assessment of National Samples of Students

A third way that countries gain national assessment information is by organizing assessments with national samples of students specifically for the purpose of national assessment. Because these assessments usually involve no more than a few percent of the schools and students nationally, they are typically not used to provide information about educational outcomes of individual schools, classes, or students and are therefore are low-stakes assessments for the schools, teachers, and students.

These survey-type assessments have significant advantages. As the number of students assessed is small compared to the national population of students, fairly complex and

expensive assessment approaches can be used without making the overall costs too high. This allows assessment of curriculum areas and aspects thereof that would normally be excluded from national testing programs for reasons of complexity and expense. The low stakes associated with the assessments reduce the risk of schools declining to participate or trying to exclude students who are expected to perform poorly. These factors increase the extent to which the sample of students assessed is representative of the national school population at the year or grade level for the assessments.

However, the low stakes of these assessments can also be a disadvantage. As there is no possibility of sanctions for individual students if they perform poorly, or school recognition or reward if they do well, they may not give the assessment tasks their full attention or effort, in which case the national assessment results may under-represent student capabilities. This problem can be greater with item types that require greater student effort (such as those that ask for extended constructed responses), in curriculum areas that are relatively unpopular, or if students are distracted by other interests or priorities (as has been observed for the National Assessment of Educational Progress (NAEP) in the final semester of high school in the United States, where students often seem to be more interested in social, cultural and sporting activities rather than serious academic work).

Assessment by School Inspection

Another option used by some countries for monitoring the processes and outcomes of education is inspection of schools by teams of experienced educators. By looking at school documentation, teaching and learning processes, examples of students work, and class achievement data, the inspectors judge educational standards in each school. These judgments can then be collated nationally, often from a representative sample of schools, to form the basis of a report on education standards countrywide.

As the evidence sought and obtained about educational outcomes is likely to be quite variable from class to class and school to school, inspection is not a very strong basis for drawing conclusions about national educational outcomes. It is more apt as a comment on the educational processes used in schools, considering that it is often the primary source of evidence nationally. The focus in this article is on national assessment of educational outcomes, so school inspection will not be discussed further.

Key Design Considerations

Choice of Outcomes to Assess

A key decision in the development of national assessment is the range of outcomes to be assessed. This involves the

identification of the curriculum areas that will be included, the aspects of those curriculum areas that will be assessed or not assessed, and how closely aligned the assessment of those aspects will be to the specific emphases of the curricula.

Only a few national assessment systems attempt to assess all curriculum areas, or almost all areas. They are far outnumbered by national assessment systems that focus their effort on areas that are perceived to be of prime importance and suitable for the more commonly used assessment approaches. This selective approach reduces the cost and complexity of the national assessment, but raises the risk that performance in areas that are not assessed may decline because educators place increasing emphasis on the areas that are assessed. Such shifts in curriculum coverage and emphasis are particularly likely if national assessment is based on whole cohort testing because of the high stakes to students, teachers, and administrators associated with the assessment outcomes.

The most common areas assessed in national assessment systems are literacy (reading and sometimes writing in the main national language or languages), mathematics, and science. Social sciences and additional languages are less commonly assessed, and the arts and physical education outcomes are included in few national assessment systems.

Within the areas that are assessed, knowledge, understanding, and skills that are more easily assessed by paper-and-pencil tests often take priority over those that are not. In reading, oral reading skills such as decoding and self-correction are less likely to be included than reading passages and answering comprehension questions. Similarly, in science, students are much more likely to be assessed on knowledge and its application than on practical and research skills.

Another issue to be addressed is whether the assessments will be tightly linked to school curriculum objectives, or sampled from a broader domain that reflects the diversity of possible learning outcomes in a curriculum area and the range of ways in which the learning is achieved, inside and outside of schools. For instance, in the early design of NAEP in the United States, a decision was taken that:

Although subjects were defined by the structure of school curricula, NAEP would assess knowledge and skills that could be gained from any source, not just from school learning. What citizens know can be measured; to discover how they acquired their knowledge would be far more resistant to discovery (Jones, 1996: 15).

If a nation does not have a nationally defined curriculum, it is almost inevitable that the assessments will be more generally focused because they cannot be tightly linked to one curriculum document. The same is, of course, true of the international assessments that must find common ground across the varied curricula of 40 or more nations.

Age Levels

National assessment is usually located in certain grade levels (or, less commonly, age levels). It is rare for a nation to assess at more than three levels. The most common level is at the end of primary education, usually around the age of 13. The lowest level is usually decided by a requirement that most participants should be able to read and write reasonably competently, and typically falls in the middle years of primary schooling. The highest level is usually located at or just below the point at which substantial proportions of the population leave school.

Inclusion Criteria and Participation

National assessment results are difficult to interpret if the population to which they apply is not clearly specified. It is important that who is to be included in or excluded from the assessments is fully described and the inclusion criteria are consistently followed. For instance, it is necessary to be clear whether both public and private schools, students whose main or first language is not the language of assessment, and students with particular levels or types of disability are to be included in the assessment. Special accommodations allowed for identified categories of students need to be described.

Another factor that can undermine clear interpretation of the results is non-participation by selected schools or students. The greater the level of non-participation, the more doubt arises about how representative the results are for the national population that is the intended focus of the assessment. Good national assessment programs therefore take considerable care to define inclusion criteria clearly and maximize participation.

Data Analysis and Reporting

There are three main options for presenting the overall results of national assessments: statistics on the performance of students on individual assessment tasks (items), aggregated statistics for clusters of tasks that all relate to a particular aspect or strand of the curriculum area assessed, and aggregated statistics for the entire curriculum area assessed.

The results for individual tasks may be presented as percentages of students who succeeded or scored at different levels on the task as a whole, or on components of the task. Alternatively, if item response modeling is used, results for tasks and their components may be presented using item characteristic curves.

Aggregated results across tasks can be created by simple addition of scores on the individual tasks, or by converting scores on each task to a standardized score (such as a z-score) and adding these scores together (perhaps with weights that reflect the perceived relative importance of

the knowledge and skills assessed by the different tasks). The third option is to create scale scores using item response modeling, allowing the modeling process to assign weights to the different tasks based on the statistical criteria of the modeling process.

All forms of aggregation involve explicit or implicit theories about the relative importance of what is assessed by the different tasks, and how they relate to each other. For instance, most item response theory (IRT) modeling assumes that the relative importance of tasks should be determined statistically by the relationships between performance on each task and the unidimensional scale that best fits the patterns of results for all of the tasks involved in the aggregation. Tasks that are outliers, having small correlations with most of the other tasks, are assigned low effective weights, and are at risk of being deleted from the assessment. This can narrow the range of outcomes assessed. This contrasts strongly with what happens when the scores on all tasks are standardized and the standardized scores added together. This implies that the knowledge and skills assessed by all tasks are to be treated as equally important in creating the aggregated score. In this case, low correlations among the tasks can result in difficulty in ascribing meaning to the aggregated score, and may suggest that the scale should be subdivided into more specific subscales that are more unidimensional.

Most of the international and national assessment programs that use samples of students (rather than whole cohorts) assign different clusters of assessment items to different students, using a matrix sampling approach. This reduces the testing load on students yet allows many assessment tasks to be used, potentially increasing the validity and comprehensiveness of the assessment. As a consequence, however, the students-by-tasks matrix has many empty cells. For the purpose of estimating what the performance statistics would have looked like if all students had attempted all tasks, statisticians have developed procedures that estimate those statistics. Mazzeo *et al.* (2006) have written a good account of the nature and purpose of these marginal estimation procedures.

Comparisons across Time

While national assessment programs aim to give an accurate picture of national patterns of student outcomes when the assessment occurred, they usually also want to show how those patterns have changed between successive assessments of the same curriculum area or, more ambitiously, across multiple assessments. The latter goal can be approached in two ways.

The first is to retain substantial proportions of tasks from one assessment to the next and not release details of these tasks until they will no longer be used. This allows comparisons across time to be based directly on

performance on these common tasks. Alternatively, the common tasks can be used to statistically link the entire assessments on consecutive occasions so that statistical comparisons across the two or more occasions can be based on the complete sets of tasks. The latter approach is usually taken if item response modeling is used. The trustworthiness of these procedures increases as the proportion of tasks retained increases, but retaining a high proportion of items limits the extent to which the results of the earlier assessment can be reported meaningfully. Also, care has to be taken not to base comparisons across time on tasks that have changed in appropriateness because of curriculum or societal changes.

The second way of providing comparisons across time uses no common tasks in consecutive assessments. New tasks are pre-tested alongside the old tasks so that performances on the two occasions can be statistically linked and then compared. This approach is most common where national assessment is based on whole cohort testing and the tests from each occasion must be publicly released. Finding a suitable sample of students for the pre-testing can be problematic, especially if the assessments carry high stakes.

Maintaining performance-trend information across multiple assessments is especially challenging. If large numbers of tasks continue to be used throughout all of the assessments, their relevance or appropriateness may decline over the series of assessments. On the other hand, if tasks are not used for more than two consecutive assessments, the long-term trend has been based on statistical linking that involves increasingly dubious assumptions as the number of assessments linked increases.

Examples of National Assessment Programs

In this final section of the article, three national systems are briefly described, to illustrate the range of approaches currently being used in national assessment.

England

From the mid-1970s to the 1980s, England had sample-based procedures for national assessment, using substantial proportions of complex performance assessments, overseen by the Assessment Performance Unit of the Department of Education and Science. These were abandoned in 1987 when the government implemented a national curriculum in England for the first time. Since then, national assessment has been based on whole cohort assessment (Daugherty, 1995; James, 2000). Originally, assessment was to occur at ages 7, 11, 14, and 16, at the end of stages described as key stages 1, 2, 3, and 4. The assessment results were to be used for several purposes:

- to provide students, their teachers, and their parents with information about their achievement;
- to provide teachers, schools, parents and the public with aggregated information about the comparative achievement of the students attending different classes, schools, and local education authorities (LEAs); and
- to provide education authorities, the government, and the nation with aggregated information about the achievement of students nationally.

The original intention was to include nationally designed standard assessment tasks (SATs) that were administered to whole classes simultaneously (under standard testing conditions) or to individual students or small groups by classroom teachers during class instruction. Teachers were also to report on whether or not students had reached the numerous attainment targets set out in the curriculum documents. Before long, however, because of concerns about teacher workload and consistency of administration, and also political preference for external testing, classroom based SATs and tests were retained only for key stage 1, and replaced by national tests for key stages 2 and 3. The teacher assessment against attainment targets was modified to focus on profiles of student performance rather than the individual targets.

In 2008, the national tests for key stage 3 were eliminated, with assessment and reporting for that stage subsequently to be based solely on moderated teacher assessments. The nature of the key stage 2 assessments was also to be reviewed.

The nationally reported results for the end of key stages 1, 2, and 3 currently focus on English, mathematics, and science. At the end of key stage 1, they are based on moderated teacher assessments taking account of the results of nationally designed SATs and tests. At the end of key stage 2 there are moderated teacher assessment results, but national test results as well in the three areas. At the end of key stage 3, there are moderated teacher assessments in the three areas. The assessments in a wide range of subjects for the General Certificate of Secondary Education (GCSE) serve as the assessments for Key Stage 4.

These assessment results are reported to students and parents. They are also aggregated for each class, school, and LEA, with the school and LEA achievement statistics made available to the general public for key stages 2, 3, and 4. These features make key stage 2, 3, and 4 assessments very-high-stakes assessments for teachers and schools.

In this system, the national results are obtained by aggregating the assessment results for all students assessed throughout England. The advantages and disadvantages of this whole cohort testing approach were discussed earlier in this article.

New Zealand

Prior to 1995, New Zealand had no systematic process for monitoring achievement of students in the first 10 years of

schooling. National examinations gave some relevant information in years 11 to 13. In 1995, the National Education Monitoring Project (NEMP) commenced. It assesses national representative samples of students in years 4 and 8 (the middle and end of primary education). Matrix sampling distributes three sets of tasks among 1440 students at each year level.

NEMP operates on a 4-year cycle, covering about one quarter of the national curriculum areas for primary and intermediate schools each year. The areas covered are literacy (reading, writing, listening, speaking, and viewing), mathematics, science, technology, social studies, art, music, health, physical education, information skills, and the use of graphs, tables, and maps.

Assessment tasks are designed to interest students and stimulate them to do their best. Laptop computers are used to present video and audio material, and in some cases to record student responses. Many tasks involve use of equipment and hands-on resources. Task components vary widely in difficulty, so that all students can enjoy some success and some challenge.

Each student has four assessment sessions, lasting about an hour each, supervised by specially trained teachers. These four sessions are:

- *One-to-one interview.* A student works individually with a teacher, attempting 15–20 tasks, with the whole session recorded on videotape.
- *Team.* Four students work collaboratively on several tasks, recorded on videotape.
- *Stations.* Four students move around a series of stations where tasks have been set up.
- *Independent.* Four students work individually on tasks such as paper-and-pencil tests, surveys, making works of art, or demonstrating physical skills (videotaped).

Experienced teachers mark tasks that require high levels of professional judgment, while other tasks are marked by senior teacher education students. Task administration and marking offer substantial, highly valued professional experience to about 260 teachers and 40 teacher education students each year.

About 45% of the tasks are kept constant from one cycle to the next. This allows trends across a 4-year interval to be observed and reported. The remaining tasks are released, making them available for teacher use and allowing detailed reporting of students' responses. The results are reported alongside the task details, showing the percentages of year 4 and/or year 8 students succeeding with, or performing at different levels on, each task component. Total scores for each task are graphed for boys, girls and three different ethnic groups. Effect sizes and statistical significance tests are used to compare the performance of various population subgroups. Trends across 4 years are displayed task by task and are also averaged across tasks. Multiple copies of the reports are

sent to all schools, accompanied by an initial response for all teachers from a national forum of educators, highlighting areas of good news, concerns, and suggestions.

United States of America

The NAEP began in 1969, with design and goals quite similar to New Zealand's NEMP (Jones, 1996, 2001). The initial intention was to assess national samples at ages 9, 13, 17, and young adult (late 20s). Ten areas were to be assessed, and results were to be reported task by task.

Over its first 10 years, cost and other considerations led to the exclusion of assessment of young adults, and of children at the other ages not attending school. Reporting shifted from individual tasks to clusters of related tasks. In the early 1980s, the Educational Testing Service proposed a new design and became the main contractor. More complex sampling arrangements were introduced, item response modeling was used to present results as scale scores for entire subjects (and in some cases major aspects of subjects), and most assessments occurred at grades 4, 8, and 12 (age sampling and audio-assisted task administration were retained for special long-term trend assessments). Overall control and oversight was given to a newly established National Assessment Governing Board, which established national standards with target achievement levels for advanced, proficient, and basic levels of competence (Jones, 1996).

Another major change to NAEP was the addition of state level reporting of results, and more recently reporting of results for some large school districts. These changes have required substantial expansion of the total sample size, in order to have adequate samples in each area reported, and have somewhat increased the stakes associated with NAEP. The larger samples have constrained the use of performance assessment approaches, because of their higher cost per student. For this reason, and because of their political importance, assessments of literacy, mathematics, and science have become much more frequent than assessments of other curriculum areas. Almost 30 years of changes in NAEP have made it

more effective for national policy and political use, but less useful for the educational goals of the original design.

Mazzeo *et al.* (2006) give a quite detailed account of the current NAEP design and the technical procedures used in data analysis and reporting, which are too complex to report here.

See also: Item Response Theory.

Bibliography

- Daugherty, R. A. (1995). *National Curriculum Assessment: A Review of Policy, 1987–1994*. London: Routledge.
- James, M. (2000). Measured lives: The rise of assessment as the engine of change in English schools. *Curriculum Journal* 11(3), 343–364.
- Jones, L. V. (1996). A history of the national assessment of educational progress and some questions about its future. *Educational Researcher* 25(7), 15–22.
- Jones, L. V. (2001). Assessing achievement versus high-stakes testing: A crucial contrast. *Educational Assessment* 7(1), 21–28.
- Mazzeo, J., Lazer, S., and Zieky, M. J. (2006). Monitoring educational progress with group-score assessments. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 681–700. Westport, CT: Praeger.

Further Reading

- Kellaghan, T. and Greaney, V. (2001). *Using Assessment to Improve the Quality of Education*. Paris: UNESCO: International Institute for Educational Planning.

Relevant Websites

- <http://www.iea.nl> – International Association for the Evaluation of Educational Achievement, PIRLS and TIMSS.
- <http://nemp.otago.ac.nz> – New Zealand's National Education Monitoring Project (NEMP).
- <http://www.nagb.org> – National Assessment of Educational Progress (NAEP).
- <http://www.pisa.oecd.org> – OECD Programme for International Student Assessment (PISA).
- <http://www.confemen.org> – Programme d'analyse des systèmes de la CONFEMEN (PASEC).
- <http://www.naa.org.uk> – QCDA, England National Assessments.
- <http://www.sacmeq.org> – The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEC).

Norm-Referenced Measurement

H D Hoover, University of Iowa, Iowa City, IA, USA

© 2010 Elsevier Ltd. All rights reserved.

A norm-referenced interpretation of test scores involves comparing the scores obtained by individuals or groups of individuals. For such comparisons to be meaningful, the tests must be administered under uniform, that is standardized, conditions. The first-known standardized tests were Chinese civil service exams introduced in 2357 BC. These exams, administered every 3 years and covering topics such as civil law, music, and Confucian classics, were designed to ensure greater fairness in selection decisions. They employed a number of sophisticated refinements such as multiple scorers to improve score reliability and the use of copyists to minimize the effect of handwriting skill on an examinee's score. Testing in European universities from the thirteenth through the seventeenth centuries consisted primarily of oral examinations, with little attention paid to standardization. However, in 1599, members of the Jesuit order, who had extensively used written tests for both placement and student evaluation, published a set of rules intended to standardize the administration of examinations.

When Horace Mann carried out the first American school survey in the grammar and writing schools of Boston in 1845, he argued that these exams should be written rather than oral, that the same questions should be asked of all students, and that the exams should be administered under uniform conditions. He also stressed that an examination is a limited sampling of a student's knowledge and that there is chance error associated with success or failure.

Another important step in the development of modern norm-referenced achievement tests was the series of surveys of educational attainment in spelling, arithmetic, and language undertaken in 1894 by Joseph Rice, a retired physician. The same exams were administered to large samples of school children. Mean scores to be expected at different grades were determined. Rice saw his exams as important in evaluating the effectiveness of different teaching strategies.

The first individual intelligence tests, put forth by Alfred Binet in 1905, greatly influenced the future of standardized tests. Binet's age-equivalent scale was the forerunner of modern developmental scales and made possible explicit normative comparisons among individuals. The relationship between mental age and chronological age was the basis for the ratio intelligent quotient (IQ) scale, which was eventually replaced by deviation IQs. These scales became the model for many of the single-group status scales discussed later. Group intelligence tests soon followed, notably the

Army Alpha, in 1917. Standardized batteries of tests for measuring the academic progress of school children, such as the Stanford Achievement Test, California Achievement Test, and Iowa Tests of Basic Skills, followed in the 1920s and 1930s. Achievement test batteries such as these are often considered synonymous with norm-referenced testing.

The discussion that follows will center on the scaling, norming, and interpretation of scores from standardized achievement test batteries. Many types of tests, such as college entrance examinations, educational placement tests, aptitude tests, certification and licensure exams, and classroom examinations, use normative information to enhance test-score interpretation. However, standardized achievement test batteries for elementary- and secondary-school students can be used to illustrate nearly all the issues surrounding the use of normative information in interpreting test scores.

Normative Scales

An examinee's raw score, or the number of test questions answered correctly, by itself seldom has much meaning. Knowing the number of items on the test, the difficulty of those items, and the conditions under which the test was administered begins to offer a context for the interpretation of the raw score. But a raw score becomes much more meaningful when it is seen in a relative sense, that is, when it is referenced to a distribution of scores obtained from a single group or several groups of examinees who took the same test under very similar conditions.

In general, there are two kinds of reference-group, or normative, comparisons (Lindquist and Hieronymus, 1964). The first describes an examinee's score in relation to the distribution of scores of the examinees in a single reference group. Score scales determined in this way are status scales. The most familiar type of status scale is probably the percentile rank. By way of example, imagine the raw score of a fifth-grader on a test being referenced to the scores of a national sample of fifth-graders who took the same test. If the student's score was higher than 72% of the scores achieved by members of the reference group, the student's score would have a percentile rank of 72. Status scales can also be used to determine an examinee's relative strengths and weaknesses in achievement across different content areas.

The second kind of normative comparison makes use of a series of reference groups. The average score for each

group is computed. A student's raw score is then compared to these group averages, and the group whose average score is most nearly like the student's raw score is identified. Score scales based on the average test scores of a series of reference groups are referred to as developmental scales and are useful in measuring growth. Two commonly used developmental scales are age equivalents and grade equivalents. If a fourth-grade student receives a grade-equivalent score of 6.0, for example, it means that the student's performance on the test is most like that of a typical sixth-grade student at the beginning of the school year.

Status Scales

Status scales have a long history of use in the social sciences. Thorndike (1904) in his seminal text *Mental and Social Measurements* stated that "Measurement by relative position in a series gives as true, and may give as exact, a means of measurement as that by units of amount." He went on to note that "Measures by amount of some unit have been the subject of great development in the hands of physical science, while measures of relative position have been comparatively neglected, though for the mental sciences they have been of utmost importance." However, he also pointed out that "Measures by position have one grave defect. Ordinary arithmetic does not apply to them. . . . We can not equate different positions in the series with each other as we can different amounts of the same thing."

The belief that ordinary arithmetic cannot be used with relative-position scales such as percentile ranks led to the development of a large number of alternatives. In general, these alternatives are simple nonlinear transformations of percentile ranks, most often those that normalize the scores. That is, these transformations convert the original scores into a normal distribution of scores. Raw scores on intelligence tests, for example, are usually transformed into a set of normally distributed scores with a mean of 100 and a standard deviation of 15. These transformed scores are commonly referred to as deviation IQs. To interpret an examinee's IQ score of 115, one needs to know that in a normal distribution, a score that is one standard deviation above the mean exceeds 84% of the scores. That is, the examinee has a percentile rank of 84. In addition to deviation IQs, there are a number of other widely used normalized status scales, such as stanines and normal curve equivalents.

The widespread use of normalized scores has helped fuel a common misconception about norm-referenced tests – that test developers assume that most mental traits of individuals are normally distributed. In fact, the term norm is sometimes assumed to mean normally distributed, when it actually refers to the scores of the reference group that make possible the interpretation of test scores by means of normative comparisons. The rationale and

history behind normalizing scores were described in detail by Angoff (1984).

Developmental Scales

Developmental scales are most commonly used in tests or test batteries such as the Stanford Achievement Test and The Iowa Tests of Basic Skills that are intended for measuring ability or educational achievement across a series of age or grade groups. The derivation of developmental scales is quite complex and is thoroughly discussed in Kolen (2006).

Developmental scales are the basis for constructing growth models, which show how students at all achievement levels increase in knowledge as they progress through the educational system. A growth model is constructed in the following way. First, within each content area assessed, a common scale is obtained linking the different grade levels of the test. Then, within-grade percentile ranks are attached to the scale. This results in a series of distributions of scores representing the overlap in achievement from grade to grade. The development of such a scale was first explicated (and the name growth model attached to it and the associated within-grade distributions of percentile ranks) by Hieronymus and Lindquist (1974). For a detailed explanation of such models, see Petersen *et al.*, (1989). Carefully constructed growth models are especially useful in longitudinal comparisons of achievement.

In summarizing the necessity of normative information in interpreting test scores, Lindquist (1953) stated at the 1952 ETS Invitational Conference on Testing Problems, "Any meaning that a scaled score has, in addition to that contained in the raw score, it has because of the normative data incorporated in the score, and that meaning applies strictly only to the particular reference population involved in the scaling process. In other words, no scaled score has any fundamental meaning attributed to the scale itself. Whatever meaning it has . . . , it has because of the normative data incorporated in the score."

Types of Norms

Normative data are crucial for most test-score interpretation. Norms provide the basis for comparing the achievement of an examinee with that of a relevant group of examinees or of a series of such groups. They can be used to assess the relative strength of an examinee's performance among a number of different content areas or to assess how that performance compares to that of series of age or grade groups in a single content area. It should be clear that the usefulness of every type of norm is dependent on the appropriateness and significance of the reference population. Interpretations based on poorly defined, vague, or unusually restricted populations have

little merit. It is also important that the norms sample be drawn from these populations in such a manner as to be truly representative. Different reference populations are used to establish different types of norms.

National Norms

Norms that are based on a nationally representative sample of those individuals for whom the test was designed (e.g., examinees of a certain age or at a certain grade level) are referred to as national norms. For instance, the scores of a nationally representative sample of third-grade students might be used as the reference group to which the scores of other third-grade examinees are compared. National normative data are also sometimes gathered from specific subgroups of examinees, such as gender-based groups, ethnic groups, public-school or private-school students, students in large-city schools, or students who attend schools in specific regions of the country. In order for subgroup norms to provide an accurate basis for comparison of examinees scores, they must be based on a large, carefully stratified reference group.

Local Norms

Norms based on the scores of a reference group of examinees in a particular educational or geographical unit, such as those in a local school district, are referred to as local norms. Local norms can sometimes be more useful than national norms. For instance, for choosing students for courses in which enrolment must be limited (e.g., an algebra class offered only to relatively high-achieving middle-school math students), local norms may be a more relevant basis for selection decisions.

User Norms

Norms based solely on the scores of examinees who have taken a test at a given time are referred to as user norms. In a number of testing programs, these are the only norms available.

The norms for the two primary college-entrance examinations in the United States, the ACT assessment and the SAT, are based on the scores of the examinees who choose to take these tests during a given time period. These norms are neither nationally representative nor comparable between the two test batteries. One reason for this lack of comparability is that the SAT is more widely used on the east and west coasts and the ACT is administered more often in the middle of the country. The racial and ethnic makeup and the educational experiences of these self-selected populations also change over time. This lack of representativeness can be especially problematic when results from these tests are used to evaluate longitudinal changes in student performance.

In some cases, user norms can be quite useful, for instance, when the examinees who take a test can be viewed as the population of interest for a given decision. This is frequently the case for licensure and certification exams, where all the examinees attempting to be certified at a particular time period are tested and included in the norms.

Norms by Age and Grade

For K-12 achievement or aptitude batteries, percentile ranks within age or grade are usually reported, as well as age- or grade-equivalent scores or scores on some other developmental scale. Status scales and developmental scales present very different indications of growth from year to year. A student who remains at the tenth percentile over the years usually falls further and further behind on a developmental scale such as grade equivalents, while the above-average student moves further and further ahead.

Group-Level Norms

When the average performance of groups is compared, group-level norms are sometimes calculated. For example, group-level norms are used when US states are ranked based on their average national assessment of educational progress (NAEP) scores. Another widely used example of a group-level norm is the school average, which can be helpful in evaluating relative strengths and weaknesses across content areas within a school building.

Item- and Skill-Level Norms

For school achievement batteries, norms based on performance on specific items or sets of items measuring a specific skill are sometimes provided. These norms may be used in a limited way in assessing strengths and weaknesses of individual students and can be extremely helpful in evaluating either classroom or building performance in specific content areas.

Types of Score Interpretation

Standardized educational achievement batteries are designed to provide primarily norm-referenced interpretations of the scores that students obtain on them. For this reason, they are commonly called norm-referenced tests. However, it is possible to make criterion-referenced interpretations of the scores obtained on such tests.

A criterion-referenced interpretation of a test score involves comparing examinees' scores with a subjective standard of performance rather than with the performance of a norm group. For example, deciding whether examinees have mastered a skill or demonstrated certain

levels of proficiency constitute criterion-referenced interpretations of their performance. In order to make criterion-referenced interpretations of scores from standardized, norm-referenced achievement tests, someone, usually a teacher or a group of content experts, must make subjective decisions on the cut scores that will be defined as separating mastery from nonmastery or various levels of proficiency from nonproficiency.

The school reform movements of recent years have placed new emphases on the scores of groups of students, including subgroups identified by such characteristics as race, ethnicity, and language proficiency. What percent of fourth-grade students are reading well enough? Has there been sufficient improvement in grade 8 math performance among Hispanic students? These are questions that require criterion-referenced interpretation of scores, and they require the establishment of standards of goodness in reading and math before they can be answered. The establishment of such performance standards is a complex endeavor and is beyond the scope of this discussion. (For a detailed description of methods of standard-setting, see Cizek and Bunch, 2006.)

Labels such as advanced, proficient, and less than proficient are commonly used to describe achievement based on performance standards. Although the terminology has expanded since the term criterion referenced first began to be used, the nature of the score interpretation really has not changed. That is, performance standards are criterion levels, just like those that have always been used in making criterion-referenced score interpretations. Some test users have been surprised to learn that commonly used terms describing performance standards do not carry an established, generally accepted meaning across states, test batteries, test publishers, or even users of the same test battery.

This lack of comparability can be illustrated using data from the national standardization programs of three widely used K-12 achievement batteries and from the criterion-referenced achievement levels in reading and mathematics established by the publishers of these batteries. **Table 1** lists the labels that the three test batteries, as well as the NAEP, use for various achievement levels. Since virtually identical labels are used, it would seem reasonable to assume that the achievement levels would be comparable across the batteries and NAEP. **Table 2** reports

the percent of students labeled proficient or advanced in reading and math by the standardized test batteries and by NAEP. The achievement levels were developed by each publisher independently and apparently with a great deal of care and effort involving a large number of qualified standard setters and significant expenditures of time and money. The numbers reported were taken directly from published materials and represent the percent of students in each battery's representative national standardization sample scoring at the proficient or advanced level. These results indicate that the lack of comparability of the standards across grades and between different test areas should be of considerable concern in interpreting results reported by school districts and states on the percent of their proficient students.

On Test X and NAEP, for example, note the difference at grade 12 between the percent of students proficient in reading and the percent proficient in mathematics. For Test Y, the difference between the percent of proficient students at the lower grades and those at the upper grades is significant, especially in mathematics. Note also the differences in percent proficient at grades 2 and 3 for Test Z in both reading and mathematics. For each of these comparisons, reports made to the public regarding percents of proficient students would likely be difficult to explain.

These data do not say that there are problems with the national norms of these test batteries but rather suggest that there may be a lack of comparability of the criterion-referenced achievement levels. It is the norms that make the differences in standards apparent. Compared to these differences, the sampling error in grade-to-grade differences in percentile ranks that are associated with group averages in national standardization samples is minuscule. Similarly, the lack of comparability of state standards makes it very difficult to compare, among states, the proportion of students who have achieved proficiency (Olson, 2002) or even to compare, within a state, the proportion of proficient students across grades or within grades across test areas. If standards-based reporting systems are to be used to compare longitudinal growth in achievement among school districts or states, it is likely that any model for growth used for such purposes will need to be developed using normative information. Such a growth model would seem to require an underlying framework based on the normative developmental scales discussed earlier.

Many educators object to the use of norm-referenced tests because they think it is inappropriate to compare children developmentally or because comparing children with an average does not seem meaningful. Statements such as "I don't care how a student compares to others; I want to be told what the student knows and is able to do" or "Norm-referenced tests create winners and losers" are common. However, comparisons are embedded in most statements we make about the developmental level and

Table 1 Labels used by three achievement batteries and NAEP

<i>Test X</i>	<i>Test Y</i>	<i>Test Z</i>	<i>NAEP</i>
Advanced	Advanced	Advanced	Advanced
Proficient	Proficient	Proficient	Proficient
Basic	Basic	Nearing proficiency	Basic
		Progressing	
Below basic	Below basic	Step 1	Below basic

Table 2 Percent of students in the nation labeled proficient or advanced by three test batteries and NAEP

Reading comprehension					Mathematics			
Grade	Test X	Test Y	Test Z	NAEP 1998	Test X	Test Y	Test Z	NAEP 2000
1	64	43	14		49	41	12	
2	58	28	41		49	34	40	
3	56	37	14		46	32	5	
4	55	40	24	31	44	34	15	26
5	49	32	33		42	28	30	
6	47	31	18		41	24	12	
7	47	39	25		39	19	21	
8	47	39	33	33	37	23	33	27
9	49	39	11		37	17	9	
10	50	34	15		35	11	12	
11	53	23	25		33	7	24	
12	54	22	29	40	30	5	27	17

Fom CTB/McGraw-Hill (1997). *Terra Nova* (Technical bulletin). Monterey, CA: McGraw-Hill; Harcourt Educational Measurement (1997). *Stanford Achievement Test Series*, 9th edn. (technical data report). San Antonio, TX: Harcourt Educational Measurement; National Center for Education Statistics (2008a). NAEP data. <http://nces.ed.gov/nationsreportcard/naepdata> (accessed May 2009); and Riverside Publishing Company (1998). *The Iowa Tests: Special Report on Riverside's National Performance Standards*. Itasca, IL: Riverside Publishing Company.

physical characteristics of any child. To say, for example, that a young child lacks self-control suggests that there is a notion of how much self-control children of this age usually exhibit. When we recognize that a student's language skills have developed early or that those of another are slow to develop, there is a norm (or standard of expectation) behind the judgment expressed. Infants are measured for height and weight, and the pediatrician reports the percentile ranks to the eager parents. Telling the parents of a 3-month-old son that their child is 64 cm (25.2 in) tall and weighs 6 kg (13.2 pounds) means little. However, telling them that he is taller than 75% and heavier than 50% of infant boys of his age clearly imparts important information.

Height and weight are different things observed on different metrics, and their comparison requires normative information. Comparing reading achievement and math achievement is no different. When the same reference group is used to establish score scales in more than one test-content area, these norms make it possible to identify relative strengths and weaknesses. This is the characteristic of norms that makes their use critical in evaluating the achievements of students and schools.

By way of example, consider trying to determine what a student knows and is able to do from both a standards-based (criterion-referenced) and a norm-referenced perspective. **Table 3** shows NAEP's grade-4 criterion-referenced achievement levels in reading. What kind of diagnostic information is available to a parent or teacher who receives a report on a student who has scored at the basic level, a level considered failing by current US educational policy? What intervention strategies are suggested?

By contrast, consider the norm-referenced report of a reading profile for a first-grade student shown in **Figure 1**.

Table 3 The NAEP reading achievement levels, grade 4

Basic	Fourth-grade students performing at the basic level should demonstrate an understanding of the overall meaning of what they read. When reading text appropriate for fourth-graders, they should be able to make relatively obvious connections between the text and their own experiences and extend the ideas in the text by making simple inferences
Proficient	Fourth-grade students performing at the proficient level should be able to demonstrate an overall understanding of the text, providing inferential as well as literal information. When reading text appropriate to fourth grade, they should be able to extend the ideas in the text by making inferences, drawing conclusions, and making connections to their own experiences. The connection between the text and what the student infers should be clear
Advanced	Fourth-grade students performing at the advanced level should be able to generalize about topics in the reading selection and demonstrate an awareness of how authors compose and use literary devices. When reading text appropriate to fourth grade, they should be able to judge text critically and, in general, give thorough answers that indicate careful thought

From National Center for Education Statistics (2008b). The NAEP reading achievement levels. <http://nces.ed.gov/nationsreportcard/naepdata> (accessed May 2009).

This report summarizes information from five of the eleven subtests on a level of the Iowa Tests of Basic Skills (Hoover *et al.*, 2003) typically taken by first-graders. At the top of the report, percentile ranks surrounded by error bands show the student's relative strengths and weaknesses in vocabulary, word analysis, spelling, listening, and

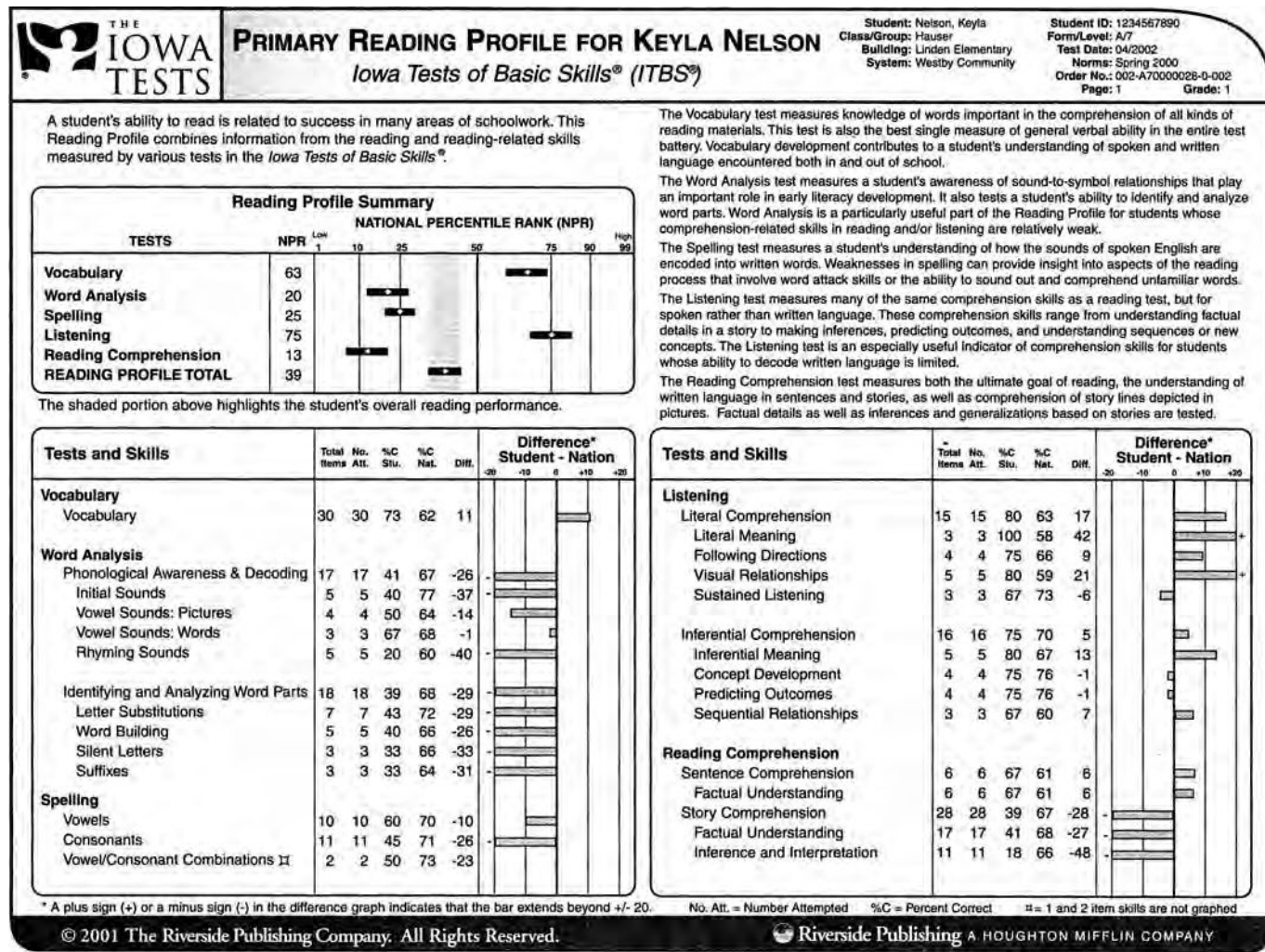


Figure 1 Student reading profile. With kind permission from Riverside Publishing.

reading comprehension. The text to the right of the profile explains to teachers and parents how each of the tests in the profile is helpful in understanding a student's reading comprehension skills. The poor reader with this profile, who shows a strong vocabulary and good listening comprehension skills, will need different instruction from that required by another poor reader who has strengths in word analysis and spelling. Comparisons like these are possible only when using norms. The bottom of the report gives a more detailed analysis of the student's performance on subskills within each test area. While criterion levels can be established for subskill scores, just as for total test scores, the low reliability of these levels is even more problematic than those for total test scores illustrated by the data in **Table 2**.

Summary

The number of questions an examinee answers on a test can seldom be meaningfully interpreted in an absolute sense. However, by referring a test score to the distribution of scores of some well-defined group of examinees, a relative judgment of the student's performance is easily obtained. The characterization of tests as either norm-referenced or criterion-referenced is unfortunate. It is score interpretations that are either norm- or criterion-referenced. Although many state and federal testing programs are viewed as criterion-referenced, it is inevitable that the test score information of most interest to the public and to policymakers is normative. For example, when results from an administration of NAEP are released, interest usually centers around the rank ordering of the states and whether there has been an increase or decrease in the proportion of students labeled proficient (a normative judgment) on the criterion-referenced achievement levels. Good score interpretation usually includes both norm-referenced and criterion-referenced components. In all cases, knowledgeable and well-reasoned interpretation of test scores should be the goal.

Bibliography

Angoff, W. H. (1984). Scales, norms, and equivalent scores. Princeton, NJ: Educational Testing Service. (Reprint of chapter in Thorndike, R. L. (ed.) *Educational Measurement*, 2nd edn. Washington, DC: American Council on Education, 1971).

- Cizek, G. J. and Bunch, M. B. (2006). *Standard Setting: A Guide to Establishing and Evaluating Performance Standards for Tests*. Thousand Oaks, CA: Sage Publications.
- CTB/McGraw-Hill (1997). *Terra Nova*. Monterey, CA: McGraw-Hill.
- Harcourt Educational Measurement (1997). *Stanford Achievement Test Series*, 9th edn. (technical data report). San Antonio, TX: Harcourt Educational Measurement.
- Hieronymus, A. N. and Lindquist, E. F. (1974). *Manual for Administrators, Supervisors, and Counselors. Levels Edition. Forms 5 and 6. Iowa Tests of Basic Skills*. Boston, MA: Houghton Mifflin.
- Hoover, H. D., Dunbar, S. B., and Frisbie, D. A. (2003). *Iowa Tests of Basic Skills: Interpretive Guide for Teachers and Counselors, Forms A and B, Levels 5–8*. Itasca, IL: Riverside.
- Kolen, M. J. (2006). Scaling and Norming. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn, pp 155–186. Westport, CT: American Council on Education and Praeger Publishers.
- Lindquist, E. F. (1953). Selecting appropriate score scales for tests. *Proceedings of the 1952 Invitational Conference on Testing Problems*, pp 34–38. Princeton, NJ: Educational Testing Service.
- Lindquist, E. F. and Hieronymus, A. N. (1964). *Manual for Administrators, Supervisors, and Counselors, Iowa Tests of Basic Skills*. Boston, MA: Houghton Mifflin Co.
- National Center for Education Statistics (2008a). NAEP data. <http://nces.ed.gov/nationsreportcard/naepdata>.
- National Center for Education Statistics (2008b). The NAEP reading achievement levels. <http://nces.ed.gov/nationsreportcard/naepdata>.
- Olson, L. (2002). A "proficient" score depends on geography. *Education Week* **21**(23), 1–14.
- Petersen, N. S., Kolen, M. J., and Hoover, H. D. (1989). Scaling, norming, and equating. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn, pp 221–262. New York: Macmillan.
- Riverside Publishing Company (1998). *The Iowa Tests: Special Report on Riverside's National Performance Standards*. Itasca, IL: Riverside Publishing Company.
- Thorndike, E. L. (1904). *Mental and Social Measurements*. New York: The Science Press.

Further Reading

- Brennan, R. L. (ed.) (2006). *Educational Measurement*, 4th edn. Westport, CT: American Council on Education and Praeger Publishers.
- Linden, K. W. and Linden, J. D. (1968). *Modern Mental Measurements: A Historical Perspective*. Boston, MA: Houghton Mifflin.
- Lindquist, E. F. (ed.) (1951). *Educational Measurement*. Washington, DC: American Council on Education.
- Linn, R. L. (ed.) (1989). *Educational Measurement*, 3rd edn. New York: American Council on Education/Macmillan.
- Linn, R. L. and Miller, M. D. (2005). *Measurement and Assessment in Teaching*, 9th edn. Upper Saddle River, NJ: Pearson Education Inc.
- Nitko, A. J. and Brookhart, S. M. (2007). *Educational Assessment of Students*, 5th edn. Upper Saddle River, NJ: Pearson Education Inc.
- Thorndike, R. L. (ed.) (1971). *Educational Measurement*, 2nd edn. Washington, DC: American Council on Education.

Progress Monitoring

L S Fuchs and D Fuchs, Vanderbilt University, Nashville, TN, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Curriculum-based measurement – The assessment of a students' overall competence in the curriculum.

Mastery measurement – The assessment of students' mastery through a hierarchy of skills that constitute the year's curriculum.

Progress monitoring – The assessment of students' rate of improvement in the annual curriculum for the purposes of determining whether a student's academic development over an academic year is proceeding well and, if not, to design an individualized program that promotes better academic growth.

Responsiveness to intervention – A multi-tier prevention system used to prevent academic failure and to identify students with learning disabilities.

Screening – Assessment on a relatively brief, inexpensive measure to identify students who may be at risk for academic failure.

Slope of improvement – A student's weekly rate of increase on curriculum-based measurement, used to quantify learning.

As a form of classroom assessment, teachers use progress monitoring for two purposes. First, they determine whether a student's academic development over an academic year is proceeding well. Second, when a student is not progressing adequately, teachers use progress monitoring to design an individualized program that promotes better academic growth. Most forms of classroom-based progress monitoring represent mastery measurement, a form of progress monitoring that has been predominant for many years. Beginning in the 1970s, a research program to develop a more technically adequate form of progress monitoring was initiated at the University of Minnesota. This more technically adequate form of progress monitoring has come to be known as curriculum-based measurement. In this article, we provide an overview of mastery measurement and explain some of its limitations. We then explain how mastery measurement and curriculum-based measurement differ. We focus the remainder of this article on curriculum-based measurement. We provide an overview for how teachers use curriculum-based measurement to make instructional decisions and then explain how curriculum-based

measurement is an important component in the responsiveness to intervention education reform movement. We next describe the most common forms of curriculum-based measurement in mathematics and in reading. We conclude by discussing future directions.

Mastery Measurement

For many years, mastery measurement was the dominant approach to progress monitoring. Using mastery measurement, teachers specify a hierarchy of skills that constitute the year's curriculum. For each skill in the sequence, teachers devise a test to assess student mastery. When a student achieves mastery, the teacher shifts instruction as well as assessment to the next skill in the hierarchy. In this way, learning is conceptualized as a series of short-term accomplishments, which are thought to accumulate into broader competence. Mastery measurement was popularized years ago with the Wisconsin Instructional Design System and Precision Teaching. Even today, most core curricula incorporate mastery measurement as their approach to progress monitoring.

In the 1970s, Stanley Deno at the University of Minnesota conceptualized and oversaw a systematic program of research on the technical features, logistical challenges, and instructional effectiveness of progress monitoring. In the initial phases of that research program, the focus was on mastery measurement. Several technical difficulties associated with mastery measurement, however, quickly emerged. For example, to assess mastery of a skill, each item on a mastery measurement test addresses the same skill, which limits the generalizability of the conclusions that can be drawn on the basis of the testing. For example, let us assume that a first-grade teacher's initial skill in her skills hierarchy is consonant–vowel–consonant words (e.g., bat, log, and sun), and her mastery test presents students with a list of 50 consonant–vowel–consonant words. Many low achievers can read consonant–vowel–consonant words if they know that all the words on the test are in this same phonics pattern but cannot read consonant–vowel–consonant words in naturally occurring text, as is the case in the real world and on high-stakes tests. In the same way, some low achievers can solve addition with regrouping problems if they know they must add and regroup for every problem on the test. However, when a test mixes math problems of different types, as is the case in the real world and on high-stakes

tests, these same students no longer perform well on addition with regrouping. In this way, mastery measurement can be misleading. This also raises questions about the assumption that mastery of a series of short-term accomplishments accumulates into broad competence. It also means that students who have mastered many objectives during the school year may not be able to use their knowledge in flexible ways or score well on the end-of-year state test. Mastery measurement also suffers from other technical problems. For example, for most criterion-referenced measures that are embedded within mastery measurement systems, reliability and validity are undocumented, and mastery measurement provides an inadequate basis for determining maintenance of previously mastered skills or generalization to untaught skills.

Curriculum-Based Measurement: An Alternative Framework for Progress Monitoring

How Curriculum-Based Measurement Differs from Mastery Measurement

In the 1970s, Deno (1985) began to address this and other problems associated with mastery measurement. He conceptualized the approach to progress monitoring that has come to be known as curriculum-based measurement. With curriculum-based measurement, an alternate form of a test is administered each week. Each alternate form represents the many skills and strategies the teacher expects students to be competent with at the end of the year. In contrast to mastery measurement, therefore, curriculum-based measurement requires students to demonstrate and integrate the various skills required for competent year-end performance on every weekly test. As the student learns the components of the annual curriculum, his/her curriculum-based measurement score gradually increases. Importantly, because each weekly test is comparable in difficulty and construction, it is possible to compute an average weekly increase in score, called slope. Teachers use slope to quantify a student's responsiveness to the instructional program and to determine when a student's instructional program needs to be adjusted. Therefore, curriculum-based measurement differs from mastery measurement in these ways (Fuchs and Deno, 1991). First, each weekly test is equivalent weekly tests and spans the curriculum reflected in the school year. Second, curriculum-based measurement is standardized so that the reliability and validity of the testing procedures can be assessed, and adequate reliability and validity have been demonstrated for a variety of curriculum-based measures.

The following example shows how curriculum-based measurement is used. Mrs. Jones, a third-grade teacher,

sets a reading goal for year-end performance as competent third-grade performance. Using established methods, she identifies enough passages of equivalent, third-grade difficulty to provide weekly assessments across the school year. Each week, she asks each student in her class to read aloud from one passage for 1 min and she scores this performance as the number of words read correctly. Each week's simple, brief assessment produces an indicator of overall reading expertise. Reading text aloud requires, for example, skill at automatically translating letters into coherent sound representations, unitizing those sound components into recognizable wholes and automatically accessing lexical representations, processing meaningful connections within and between sentences, relating text meaning to prior information, and making inferences to supply missing information. As competent readers translate text into spoken language, they coordinate these skills in a manner that appears effortless (Fuchs *et al.*, 2001). This approach to curriculum-based measurement is known as the overall indicator approach.

A second approach to curriculum-based measurement is to systematically sample the various skills that constitute the year's curriculum. With curricular sampling, the weekly curriculum-based measurement test presents students with items that represent the variety of skills the teacher will address over the academic year. With either approach, each progress-monitoring test collected across the school year is of equivalent difficulty; and with either approach, each week's assessment reflects the performance that can be used to characterize overall competence in the annual curriculum for an academic domain.

Curriculum-based measurement can also be used to track its development through the primary grades. That is, scores can be graphed and compared to each other. Moreover, as mentioned, a slope (i.e., average increase per week) can be calculated on the series of scores to quantify the rate of improvement. This method for characterizing growth has been shown to be more sensitive to individual differences than those offered by other classroom assessments. Moreover, curriculum-based measurement is sensitive to academic development made under a variety of instructional procedures. In a related way, research shows that teachers' instructional plans developed in response to curriculum-based measurement incorporate a wide range of instructional methods. For example, in reading, instructional plans developed in response to curriculum-based measurement incorporate decoding instruction, repeated readings, vocabulary instruction, story grammar exercises, and semantic-mapping activities. Therefore, curriculum-based measurement is not linked to a particular instructional approach. Most importantly, studies show that curriculum-based measurement progress monitoring enhances teachers' capacity to plan programs for students with serious learning problems and effect superior achievement for those students.

To inform instructional planning, teachers rely on curriculum-based measurement's scores in graphed form. If a student's rate of improvement, as shown on the graph, is inadequate, the teacher revises the instructional program. Research (see Fuchs and Fuchs, 1998 for summary) shows that with curriculum-based measurement decision rules, teachers' instructional plans are more varied and are more responsive to student needs. Besides, their plans incorporate more ambitious goals for students and result in stronger end-of-year scores on standardized tests, including high-stakes state tests. In addition, when curriculum-based measurement systematically samples the skills in the year's curriculum, performance on the individual skills represented on each assessment can be analyzed. This kind of skills analysis has been shown to enhance the quality of instructional programming.

Curriculum-Based Measurement and Instructional Decision Making

Regardless of whether one adopts an overall indicator or a curriculum sampling approach to curriculum-based measurement, teachers use validated curriculum-based measurement decision rules to determine whether a student's academic development is adequate and, if not, to design an individualized instructional program that prompts better growth.

In **Figure 1**, we show a sample curriculum-based measurement report for one class. The first page shows a composite class graph. The bottom path on the graph depicts the progress of students in the bottom 25% of the class. The middle path shows the progress of students in the middle 50% of the class. The top path characterizes the progress of students in the top 25% of the class. Under the class graph, the names of 'students to watch' and 'most improved' in the past month are displayed. Next, the class report identified skills on which the class has made good progress in the past month and skills on which most students in the class should profit from instruction. At the bottom of page 1, students are identified for small-group instruction. On the second page of the report, each student's mastery status on each skill embedded in the annual curriculum is coded; the darker the box, the higher the mastery status. The teacher can look across a row to obtain the profile of skills strengths and weaknesses for a given student. The teacher can look down a row to determine how her class is performing on a specific skill. Page 3 shows students' slopes of improvement, and on page 4, names are listed of students who are noticeably behind their classroom peers in terms of the level of their current curriculum-based measurement score and their slope of improvement over time.

To determine whether a student's academic development is adequate, the teacher considers the student's slope

of improvement against the rate of improvement demonstrated by other students in the class or against national norms for expected rates of improvement. When a student is chronically identified as not making good progress, as identified on page 4 of the class report in **Figure 1**, curriculum-based measurement is then used to design an individualized instructional program. The goal is to promote better academic growth.

This is usually accomplished by the special educator or the reading or mathematics specialist, using curriculum-based measurement decision rules. These decisions rules are applied to the slope of the graphed scores, after the year-end goal has been set and placed onto the student's individual graph. A straight line connects the student's initial scores with that goal. This is called the goal line, which is compared to the student's actual rate of improvement or slope, referred to as the trend line. When a student's trend line is flatter than the goal line, the teacher revises the instructional program in some important way in an attempt to effect a better slope of improvement. After several weeks of implementing this instructional revision, the teacher analyses the student's graph in the same way, comparing the goal line against the student's new trend line. If progress looks strong enough to realize the end-of-year goal, the program revision is considered successful (although data analysis continues to ensure strong improvement across the school year). If, however, progress looks inadequate to ensure reaching the end-of-year goal, then the teacher experiments with a different method for revising the instructional program. Research (see Fuchs and Fuchs, 1998) shows that this kind of data-based decision making, based on curriculum-based measurement, results in more varied instructional programs that are more responsive to individual needs; that result in more ambitious student goals; and that produce stronger end-of-year scores on commercial, standardized achievement tests including high-stakes tests.

Progress Monitoring: An Essential Component for Responsiveness to Intervention

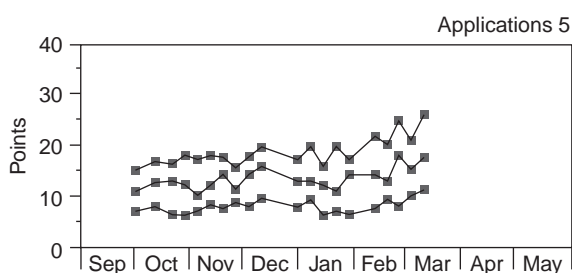
No instructional method, even those validated using randomized control studies, are universally effective. Therefore, when schools implement validated instructional programs, the effects of those programs on children's academic development must be assessed. That way, children who do not respond adequately can be identified promptly for different or more intensive intervention. For students who fail to respond to a second, more intensive level of programming, a third instructional tier, with greater individualization, is implemented. Throughout this process, the assessment of response continues. This assessment of response can be accomplished through curriculum-based measurement.

This iterative process, with which interventions of increasing intensity and individualization are conducted and their effects are assessed, describes the education reform movement known as responsiveness to intervention, which is conducted as part of a multi-tier prevention system. The first tier is primary prevention. The intensity of instruction reflects general education practice, with or without the use of individual accommodations and adaptations that fit within the general education program and can be managed by the general education teacher. The second tier is secondary prevention. Secondary prevention incorporates a greater level of intensity than

can be accomplished within general education. Typically, instruction is delivered in small groups of three to five students by well-trained, although not necessary, certified tutors, who use a prescriptive, validated intervention for 10–20 weeks. In some models, multiple doses of small-group, research-based tutoring are used to strengthen effects. If secondary prevention tutoring proves ineffective, then an even more intense level of intervention is conducted. This most intense level of instruction, referred to as tertiary prevention, is individually tailored. Instruction is delivered one to one or in pairs. It is often conducted under the auspices of special education resources, using

Class summary

Teacher: Mr. Simmons
Report through 3/15



Students to watch

David Torres
Amanda Smith
Dustin Montgomery
Kasha Michaels

Most improved

Dustin Montgomery
Kasha Michaels
Justin Price
Brittany Page
Yasmine Andrade

Areas of improvement: applications

WP Word Problems

CG Charts and Graphs

Whole class instruction: applications

Mn Money

57% of your students are either COLD or COOL on this skill

Small group instruction: applications

Me Measurement

Amanda Smith	Erica Johnson
Brittany Page	Kasha Michaels
David Torres	
Dustin Montgomery	

Figure 1 (continued)

special-education certified teachers who are skilled in formulating individually tailored instruction.

Within a responsiveness-to-intervention system, therefore, progress monitoring plays three important roles. First, it is used to identify who should be targeted for special attention. Second, for those targeted for special attention, assessment is used to quantify responsiveness to intervention. Third, for the most unresponsive subset of children, assessment is used to tailor individualized programs. Curriculum-based measurement can be used for all three purposes.

Identifying Who Should Be Targeted for Attention

Within a responsiveness-to-intervention prevention system, the first assessment function is identifying a group of students who suspected to be at risk for poor learning outcomes. These students become the focus of the responsiveness-to-intervention system. Such screening is typically accomplished by administering a brief measure to all students at a given grade level. A cut-score is applied to these score. The cut-score indicates the score that is

Class skills profile - applications

Teacher: Mr. Simmons

Report through 3/15

Name	Nu	Mn	Me	Ge	CG	FF	De	AC	WP
Amanda Smith									
Brandon Davis									
Brittany Alvarez									
Brittany Page									
David Torres									
Dustin Montgomery									
Erica Johnson									
Jackie Alexcander									
Justin Price									
Kasha Michaels									
LaKisha Thomas									
Stephen Miranda									
Victor Nobrega									
Yasmine Andrade									
COLD. Not tried	0	4	0	0	0	0	0	1	1
COOL. Trying these.	4	4	3	13	7	8	7	7	3
WARM. Starting to get it.	8	1	10	1	6	5	6	6	7
VERY WARM. Almost have it.	0	3	1	0	0	1	1	0	1
HOT. You've got it!	2	2	0	0	1	0	0	0	2

Ranked scores - applications

Teacher: Mr. Simmons

Report through 3/15

Name	Score	Growth
Brandon Davis	28	+0.36
Justin Price	25	+0.41
Stephen Miranda	23	+0.34
Jackie Alexcander	22	+0.29
Brittany Alvarez	18	+0.23
Yasmine Andrade	17	+0.30
Victor Nobrega	17	+0.07
Erica Johnson	15	+0.50
Brittany Page	14	+0.12
LaKisha Thomas	13	+0.22
David Torres	13	+0.19
Amanda Smith	12	+0.09
Dustin Montgomery	9	+0.15
Kasha Michaels	8	+0.01

Figure 1 (continued)

Class statistics: applications

Teacher: Mr. Simmons

Report through 3/15

Score

Average score 16.7

Standard deviation 5.9

Discrepancy criterion 10.8

Slope

Average slope +0.24

Standard deviation 0.14

Discrepancy criterion +0.10

Students identified with dual discrepancy criterion

	Score	Slope
Kasha Michaels	8	+0.01

Figure 1 A curriculum-based measurement class report for the half-month period starting on March 15. Page 1 shows (a) a class graph with the bottom path charting scores for the lowest quartile of the class, the middle path charting scores for the middle half of the class, and the top path charting scores for the highest quartile of the class; (b) a list of students to watch because their scores fall in the lowest quartile; (c) a list of students whose scores have improved in the most in the past half-month; (d) skills on which the class as a whole has improved in the past half-month; (e) skills on which the class as a whole can benefit from instruction in the upcoming half-month; and (f) skills on which small groups of students can benefit from instruction in the upcoming half-month. Page 2 shows (a) a skills profile in which the skills incorporated with each weekly assessment are summarized as mastered (black box), probably mastered (black box with a dot), attempted and partially mastered (plaid box), attempted but not mastered (striped box), and not attempted (open box) for each student in the class; (b) a frequency count of students in each mastery status for each skill; and (c) a rank ordering of students in terms of their most recent curriculum-based measurement score, with the student's slope of improvement through scores since the beginning of the academic year. Page 3 summarizes the class performance with (a) the average curriculum-based measurement score on March 15, the standard deviation around that average score, and the discrepancy criterion (i.e., 1 standard deviation below the mean, below which a student's curriculum-based measurement score may indicate difficulty); (b) the average curriculum-based measurement slope of improvement as of March 15, the standard deviation around that average slope, and the discrepancy criterion (i.e., 1 standard deviation below the mean, below which a student's curriculum-based measurement slope of improvement may indicate difficulty); and (c) a list of student(s) who meet the discrepancy criterion on both the average curriculum-based measurement score and the average curriculum-based measurement slope and therefore warrant attention to avoid poor long-term outcomes.

associated with inadequate performance on an important outcome measure, such as a high-stakes test, at a later time. All students scoring below this cutoff are designated as at-risk for poor outcome. Technically, screening, which involves a one-time test administration, is not a form of progress monitoring, which requires more frequent (typically, at least monthly) assessment. However, prominent screening measures have been borrowed from curriculum-based measurement tools due to the strong reliability and validity of curriculum-based measurement and because of its efficiency.

It is important to note that 1-time screening carries a significant danger of identifying false positives (i.e., designating students as at risk and in need of tutoring when, in fact, those students would develop strong academic skills without tutoring). In a first-grade responsiveness-to-intervention experiment (Fuchs *et al.*, 2006), for example,

50% of the control group, who had been designated at risk according to 1-time screening but who did not actually receive tutoring, made good progress by the end of the first semester of first grade. Such false-positive errors are expensive for responsiveness-to-intervention prevention systems because they waste costly resources on students who do not require them.

Because screening, especially at kindergarten and first grade, typically overidentifies students for secondary prevention tutoring, we recommend that 1-time screening constitutes only the first step in designating risk. Therefore, students who are suspected to be at risk, based on brief screening, are followed with 5–8 weeks of progress monitoring while Tier 1 general education is implemented. This short-term progress monitoring is used to gauge a student's actual response to Tier 1 general education and thereby confirm that the suspected risk, based on

screening, constitutes actual risk for reading or mathematics difficulties. Such short-term progress monitoring has been shown to increase the precision of designating who requires secondary prevention.

Quantifying Responsiveness to Intervention

Within a multi-tier prevention system, a second purpose for progress monitoring occurs with secondary prevention. With secondary prevention, tutoring is based on standard treatment protocol. With a standard treatment protocol, validated tutoring is conducted in small groups. The assumption, based on the validation research, is that most students should respond well to this tutoring. If a child responds poorly to instruction that benefits most students, then the responsiveness-to-instruction assessment process eliminates instructional quality as a viable explanation for poor academic growth and, instead, provides evidence of a disability. The purpose of progress monitoring at secondary prevention is, in part, to determine whether students do or do not respond to validated small-group tutoring and to identify students who do not benefit for a more intensive level of instruction. The students who fare well (and respond) are returned to general education, where progress monitoring continues to assess whether any additional tutoring is required.

Tailoring Individualized Instructional Programs

When, however, the secondary prevention progress-monitoring data indicate that the student has not responded adequately to secondary prevention tutoring, the student enters tertiary intervention, typically with special education resources and personnel. At the tertiary prevention level, instruction differs from secondary prevention because it is more intensive, involving longer sessions conducted in smaller groups or individually and because it is individualized, rather than relying on a validated treatment protocol. At tertiary prevention, progress monitoring is essential for two purposes: to inductively formulate instructional programs that are optimal for that student (as already described) and to determine when the student's response to tertiary prevention is adequate to return to the primary prevention general education program, possibly with accommodations or modifications, or to return to secondary prevention's small-group tutoring using a standard treatment protocol. In either case, progress monitoring continues so that tertiary intervention can be reinstated as needed.

Curriculum-Based Measurement Measures in Mathematics

At kindergarten in math, research has focused primarily on the overall indicator approach to curriculum-based

measurement. One promising measure is quantity discrimination. With quantity discrimination, students are shown a page with a series of pairs of numerals. The student has 1 min to circle the larger quantity for each pair. Another promising measure is number identification, with which students see a page of numerals and have 1 min to name the numbers. A third promising measure is missing number. With missing number, students are presented with strings of numerals, with the last numeral missing, and the student has 1 min to fill in the missing numerals. Some forms of validity have been shown to be strong (e.g., Baker *et al.*, 2002; Lembke and Foegen, 2006). Yet, some work suggests that when these measures are used for progress monitoring, students may learn these specific tasks and ceiling out of the test quickly, within 1 or 2 months (Lembke and Foegen, 2005). To address this problem, Seethaler and Fuchs (2006) developed and assessed a curriculum-sampling curriculum-based measurement approach at kindergarten. With this curriculum-sampling approach to curriculum-based measurement, a variety of computation and concepts/applications is systematically sampled on each alternate form. This study shows that the curriculum-sampling curriculum-based measurement systems, when used at the beginning of kindergarten, forecast students' development of math competence at the end of first grade with similar accuracy as do the overall indicator measures. Additional work is needed to determine which curriculum-based measure is sound for monitoring math progress over time at kindergarten.

At first grade in mathematics, most progress-monitoring research parallels the work conducted at kindergarten, using quantity discrimination, number identification, and missing number. As with kindergarten, although some forms of validity look strong, the threat of students reaching the top score on the test quickly looms larger even at first grade than at kindergarten. An alternative approach, using a curriculum-sampling approach to curriculum-based measurement, has also been used at first grade. It has been shown to index progress effectively and to forecast the development of serious math difficulties at the end of second grade (Compton *et al.*, 2006; Fuchs *et al.*, 2005). In a similar way, research at grades 2–6 (Deno *et al.*, 2006) demonstrates the validity of a curriculum-sampling curriculum-based measurement approach, and other research (see Fuchs and Fuchs, 1998) illustrates how teachers can use the curriculum-sampling curriculum-based measurement approach in math to design more effective instructional programs.

Curriculum-Based Measurement Measures in Reading

As already stated, curriculum-based measurement can take one of two forms. It can rely on a single behavior

that functions as an overall indicator of competence in an academic area or it can systematically sample the annual curriculum. In reading, most well-researched curriculum-based measurement systems take the overall indicator approach.

At kindergarten in reading, there are three major alternatives for curriculum-based measurement: phoneme segmentation fluency, rapid letter naming, and letter-sound fluency. With phoneme segmentation fluency, the tester says a word, and the child says the sounds that constitute the word (i.e., the tester says, *rat*; the child says, /r/ /a/ /t/). The tester administers as many items in 1 min as the rate of the child's response permits. With rapid letter naming, the tester presents a page showing lower- and uppercase letters randomly ordered, and the child says as many letter names as possible in 1 min. With letter-sound fluency, the tester also presents a page showing lower- and uppercase letters randomly ordered. This time, the child says sounds for 1 min. Compared to phoneme segmentation fluency, rapid letter naming and letter-sound fluency are easier for teachers to learn to administer, and reliability tends to be stronger. Yet, compared to rapid letter naming, phoneme segmentation fluency and letter-sound fluency are better instructional targets. They relate more transparently to what children need to learn for reading. Therefore, phoneme segmentation fluency and letter-sound fluency may guide the kindergarten teacher's instruction more effectively, although such studies have not been conducted.

For first-grade reading, two curriculum-based measurement measures have been the focus of study. With one approach, students begin the year on the nonsense-word fluency task and move to passage-reading fluency in January. With nonsense-word fluency, students are shown a page with consonant-vowel-consonant pseudo-words (such as *gav* or *jum*). They have 1 min to decode as many as they can. With passage-reading fluency, students are presented with first-grade text, with each weekly alternate form of roughly equivalent difficulty. The student reads aloud for 1 min.

As an alternative, schools use word-identification fluency across all of first grade. Students are presented with a page showing 50 high-frequency words. Each alternate form systematically samples high-frequency words in random order, and students read as many words as possible in 1 min. Nonsense-word fluency offers the advantage of mapping onto beginning decoding instruction and may therefore provide teachers with input to form instructional plans. The disadvantage of combining nonsense-word fluency with passage-reading fluency across the school year is that changing the measure at mid-year precludes calculating slope or modeling student development over first grade. Word-identification fluency, by contrast, can be used with strong reliability, validity, and instructional utility across the entire first-grade year, also making it possible to model the development of

reading skill across the entire time frame. In a study that contrasted nonsense-word fluency to word-identification fluency (Fuchs *et al.*, 2004), the technical features of word-identification fluency exceeded those of nonsense-word fluency.

In reading at grades 2 and 3, the curriculum-based measurement task of choice is passage-reading fluency. It is the most-studied curriculum-based measure, with strong technical features and utility for instructional planning. Each week, one passage is administered. The student reads aloud for 1 min. Then, the tester counts the number of words read correctly. The accuracy (reliability), meaningfulness (validity), and instructional usefulness of this simple measure have been demonstrated repeatedly (see Fuchs and Fuchs (1998) for summary).

It is important to note, however, at grades 4–6 – as students move from learning to read to reading to learn – studies suggest that the validity of the curriculum-based measurement passage-reading fluency task decreases (Espin, 2006). Therefore, beginning at grade 4 or 5, the use of a different measure, which more directly taps comprehension, may be indicated. One option at the higher grades, which is efficient for teachers to use on a regular basis, is curriculum-based measurement maze fluency. With maze fluency, students work on a passage from which every seventh word has been deleted and replaced with three choices. Only one option makes sense in the blank. The student has 2 or 3 min to read and restore meaning to the passage by replacing blank with words. The score is the number of correct replacements. Espin (2006) provided evidence that maze fluency demonstrates strong accuracy (reliability) and meaningfulness (validity) and models reading development well beginning at fourth grade, continuing through eighth grade.

Future Directions

An essential component of classroom assessment is systematic progress monitoring. Teachers use progress monitoring for two purposes: to determine whether a student's academic development during an academic year is proceeding well and, when a student is not progressing adequately, to design an individualized instructional program that promotes better academic growth. Curriculum-based measurement is the form of progress monitoring with the strongest scientific evidentiary base. A large literature shows that curriculum-based measurement produces accurate descriptions of student development in reading and math. In addition, randomized control trials in real-school settings, in which teachers are randomly assigned to plan instruction with and without curriculum-based measurement, document that when teachers use curriculum-based measurement to inform their instructional decisions,

students achieve better. Research on the use of curriculum-based measurement for the purpose of identifying students with learning disabilities, within a responsiveness-to-intervention framework, represents a new avenue of inquiry. Moreover, there is ongoing research exploring methods to expand curriculum-based measurement to other grade levels and content.

Bibliography

- Baker, S., Gersten, R., Flojo, J., *et al.* (2002). Preventing mathematics difficulties in young children: Focus on effective screening of early number sense delays. *Technical Report No. 0305*. Eugene, OR: Pacific Institutes for Research.
- Compton, D. L., Fuchs, L. S., and Fuchs, D. (2006). The course of reading and mathematics disability in first grade: Identifying latent class trajectories and early predictors.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children* **52**, 219–232.
- Deno, S. L., Jiban, C., and Spanjers, D. (2006). Monitoring math progress: Grades 1–6. *Paper Presented at the Annual Meeting of the Pacific Coast Research Conference*.
- Espin, C. (2006). The technical features of reading measures. *Paper Presented at the Annual Meeting of the Pacific Coast Research Conference*.
- Fuchs, D., Compton, D. L., Fuchs, L. S., and Bryant, J. D. (2006). Prevention and identifying reading disability. *Paper Presented at the Annual Meeting of the Pacific Coast Research Conference*.
- Fuchs, L. S., Compton, D. L., Fuchs, D., *et al.* (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology* **97**, 493–513.
- Fuchs, L. S. and Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children* **57**, 488–501.
- Fuchs, L. S. and Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice* **13**, 204–219.
- Fuchs, L. S., Fuchs, D., and Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children* **71**, 7–21.

- Fuchs, L. S., Fuchs, D., Hosp, M., and Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading* **5**, 239–256.
- Lembke, E. and Foegen, A. (2006). Monitoring student progress in early math. *Paper Presented at the Annual Meeting of the Pacific Coast Research Conference*.
- Seethaler, P. M. and Fuchs, L. S. (2006). Reliability, concurrent validity, and predictive validity of a curricular sampling approach to curriculum-based measurement in mathematics at kindergarten. Research in progress.

Further Reading

- Compton, D. L., Fuchs, D., Fuchs, L. S., and Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology* **98**, 394–409.
- Fuchs, D. and Fuchs, L. S. (2005). Responsiveness-to-intervention: A blueprint for practitioners, policymakers, and parents. *Teaching Exceptional Children* **38**, 57–61.
- Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review* **33**, 188–192.
- Fuchs, L. S., Fuchs, D., and Courey, S. J. (2005). Curriculum-based measurement of mathematics competence: From computation to concepts and applications to real-life problem solving. *Assessment for Effective Instruction* **30**(2), 33–46.
- Stecker, P. M., Fuchs, L. S., and Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools* **42**, 795–820.

Relevant Websites

- www.wids.org – Instructional Design System, Wisconsin, USA.
- www.interventioncentral.org – Intervention Central.
- www.jimwrightonline.com – Intervention Central.
- www.rti4success.org – National Center on Response to Intervention.
- www.celeration.org – The Standard Celeration Society, a special interest group of the Association of Behavior Analysis.

Rasch Models

D Andrich, The University of Western Australia, Crawley, WA, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Location – The value of a parameter of a person, or a threshold of an item, on a continuum.

Parameter – A variable whose values govern the responses in a probabilistic model.

Polytomous response format – A format in which there is a response in one of more than two mutually exclusive categories.

Sufficient statistic – A function of the responses with respect to a parameter such that, conditional on this statistic, the resulting distribution is independent of the parameter.

Threshold – The location where the probability of a response in either of two adjacent categories is identical.

Introduction

This article explains the unidimensional Rasch model (RM) for two or more ordered categories. It considers the rationale for the RM and principles of parameter estimation and tests of fit, but is not concerned with details of these procedures. There are a number of software packages that implement the RM at an advanced level. Detailed studies of the theory and applications of the RM can be found in Fischer and Molenaar (1995), Engelhard and Wilson (1996), Van der Linden and Hambleton (1997), and Smith and Smith (2004). The article begins with the simplest of these models for the simplest response format, that of two ordered categories, with just two items. It then proceeds to generalize the model for more items and more than two ordered categories in standard formats. The article comments on the distinctive paradigm of the data–model relationship in the application of the RM and shows some relationships between the RM and traditional test theory (TTT).

Standard formats involve one response in two or more categories deemed *a priori* to be ordered on a hypothesized continuum, often termed a latent trait. They are used in quantifying achievement, performance, attitude, and status in the social sciences. A common dichotomous response format in assessing achievement is multiple choice where only one of the given options is correct. Another is that of scoring an extended response to a

question as pass or fail. Table 1 shows typical formats for four ordered categories which are extensions of the two ordered categories, and which may be extended further. Such formats are termed polytomous. Figure 1 shows the format on a continuum that is envisaged when models such as the RM is applied to the response data. The case of two categories, for example, fail or pass, is simply a special case of more than two categories, for example, fail, pass, credit, or distinction as in Figure 1.

The Case for the RM

To increase the validity and precision beyond that which can be achieved with just one item, tests generally have many items, the number constrained by the time available due to administrative and fatigue factors. In assessing whether the items are working consistently with each other to separate persons on a single continuum, the RM was derived from the following requirements of invariant comparisons:

- The comparison between two stimuli should be independent of which particular individuals were instrumental for the comparison.
- Symmetrically, a comparison between two individuals should be independent of which particular stimuli within the class considered were instrumental for comparison (Rasch, 1961: 332).

Equivalent requirements for invariance were articulated by Thurstone (1928) and Guttman (1950), but Rasch was the first to formalize it in the form of a probabilistic mathematical model (Andrich, 2005). Thus, the model has the inherent requirements of the invariance of comparisons. Then to the degree that the data, too, conform to the model, to that degree the data have the property of invariance. In addition, mathematical derivations that rest on this requirement can lead to insights for measurement and test design that go beyond those that could be obtained descriptively. An illustration is provided at the end of the article.

The Item Characteristic Curve

The dichotomous RM takes the form

$$\Pr\{X_{ni} = x\} = (\exp x(\beta_n - \delta_i)) / \gamma_{ni} \quad [1]$$

where $X_{ni} = x \in \{0, 1\}$ is an integer random variable of the responses; β_n and δ_i are, respectively, the parameters

of person n and item i characterizing their locations on the same continuum; and $\gamma_{ni} = 1 + \exp(\beta_n - \delta_i)$ is a normalizing factor that ensures that the two probabilities in eqn [1] sum to 1.

Figure 2 shows the probability of the positive response $X_{ni} = 1$ as a function of β for three items with location parameter estimates δ_i equal to -1.01 , -0.07 , and 0.94 , respectively. These functions are known as item characteristic curves (ICCs). The probability of the positive response increases as a function of the location of the persons. Notice that the ICCs are parallel (Wright, 1997). This is a distinctive property of the RM and reflects the property of the invariance of comparisons. Further, the location of the item is identified as the position on the continuum where a person has equal probability of responding 0 or 1. From psychophysics, these locations are known as thresholds. In educational measurement, they are referred to as item difficulties and the locations of persons are referred to as abilities. This article also uses the generic term locations for these parameters.

Table 1 Standard response formats for more than two ordered categories

Fail	Pass	Credit	Distinction
Never	Sometimes	Often	Always
Strongly disagree	Disagree	Agree	Strongly agree

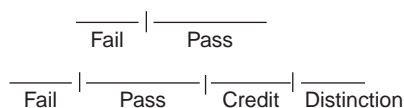


Figure 1 Graphical representation of contiguous intervals for ordered categories.

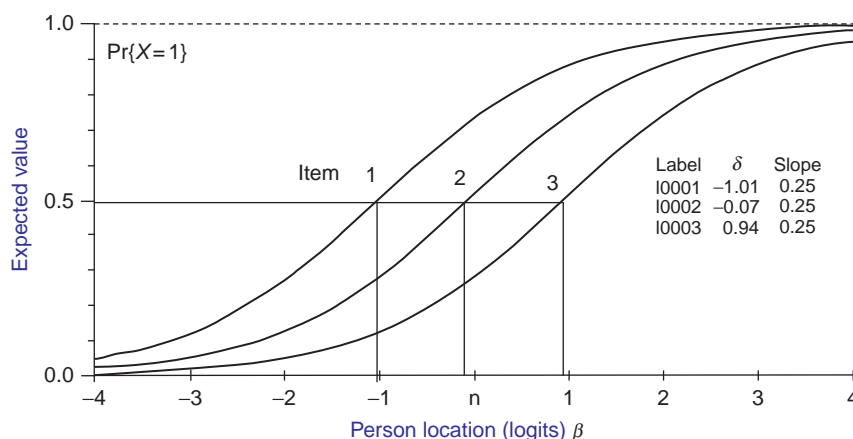


Figure 2 Item characteristic curves for three items.

In any analysis, the common slope of the ICC curves is a kind of average of the discrimination of the items. This rate of change is often termed the discrimination.

Implications of Dimensionality and Response Independence

Equation [1] shows that, because there is only one person parameter, the model is unidimensional, that is, that persons can be distinguished on just one continuum. It also implies that the responses of a person to more than one item are statistically independent in the sense that

$$\Pr\{(X_{ni}) = (x_{ni})\} = \prod_{i=1}^I \exp x_{ni}(\beta_n - \delta_i) / \gamma_{ni} \quad [2]$$

where (x_{ni}) denotes the vector of responses of person n to items $i = 1, 2, \dots, I$.

Thus, only the person's and item's locations govern each response. Clearly, the greater the ability of the person, the more likely is a positive response to all items; and the greater the difficulty of an item, the more likely is a negative response from all persons. This implies that the responses among items for a sample of persons with different abilities will be correlated. Nevertheless, statistical independence holds according to eqn [2]. This independence is often termed local independence.

Sufficiency of the Total Score and Invariance

The formalization of invariance of comparisons rests on a sufficient statistic for a parameter (Rasch, 1961). A sufficient statistic implies that, conditional on that statistic, the resultant distribution of responses of a person to the items is independent of the parameter. The sufficient statistic for the person parameter β_n is simply the total score

$r_n = \sum_{i=1}^I x_{ni}$. For a vector of responses (x_{ni}) of person n to I items,

$$\Pr\{(X_{ni}) = (x_{ni}) | r_n\} = \frac{1}{\gamma_r} \exp\left(\sum_{i=1}^I x_{ni}(-\delta_i)\right) \quad [3]$$

where $\gamma_r = \sum_{((x_{ni})|r_n)} \exp\left(\sum_{i=1}^I x_{ni}(-\delta_i)\right)$ is the summation over all possible vectors of responses of person n given a total score of r_n . Equation [3] is clearly independent of the person parameters β_n , $n = 1, 2, \dots, N$.

Estimates of Item Parameters

As it is central to appreciating the model, the implications of eqn [3] are illustrated in **Table 2** with 50 persons responding to just two dichotomous items. There are two sets of data. The responses are taken as either correct or incorrect.

Equation [3], which is a probability statement conditional on the total score r_n , has only one response pattern (0,0) for $r_n = 0$, and only (1,1) for $r_n = 2$. Therefore, $\Pr\{(0,0)|r_n = 0\} = 1$ and $\Pr\{(1,1)|r_n = 2\} = 1$, making both patterns uninformative regarding the relative difficulties of the items. For $r_n = 1$, however, there are two patterns, (1,0) and (0,1). Then eqn [3] becomes

$$\begin{aligned} \Pr\{(1,0)|r_n = 1\} &= (\exp(-\delta_1))/\gamma_1; \\ \Pr\{(0,1)|r_n = 1\} &= \exp(-\delta_2)/\gamma_1 \end{aligned} \quad [4]$$

where $\gamma_1 = \exp(-\delta_1) + \exp(-\delta_2)$.

The responses in which $r_n = 1$ are themselves dichotomous – the person is correct on only one or other of the items. Thus, with N_{12} persons with $r_n = 1$ and s_1 correct responses on item 1, the resultant distribution is binomial:

$$\Pr\{s_1 | r_n = 1\} = \binom{N_{12}}{s_1} \pi_{12}^{s_1} (1 - \pi_{12})^{N_{12} - s_1} \quad [5]$$

where $\pi_{12} = (\exp(-\delta_1))/(\gamma_1)$; $1 - \pi_{12} = \pi_{21} = (\exp(-\delta_2))/\gamma_1$.

With just two items, to estimate (δ_1, δ_2) equate the proportions in **Table 2** to the ratio of the corresponding probabilities of eqn [5]. In the case of set 1, this gives

$$\frac{\Pr\{(1,0)|r_n = 1\}}{\Pr\{(0,1)|r_n = 1\}} = \frac{\pi_{12}}{\pi_{21}} = \frac{\exp(-\delta_1)}{\exp(-\delta_2)} = \frac{10/30}{20/30} = 0.5 \quad [6]$$

Taking logarithms of both sides of eqn [6] gives $-\hat{\delta}_1 + \hat{\delta}_2 = \ln(0.5) = -0.69$. That is, $\hat{\delta}_1 - \hat{\delta}_2 = 0.69$ indicating that item 1 is more difficult than item 2. This is evident because of the 30 persons who answered both items correctly, only 10 answered item 1 correctly while twice as many, 20, answered item 2 correctly.

Notice that only the difference $\hat{\delta}_1 - \hat{\delta}_2$ is estimated, not a value for each of $\hat{\delta}_1$ and $\hat{\delta}_2$ individually. The estimates in the RM are comparisons, and not absolute values. To obtain a single value for each item, the additional constraint $\hat{\delta}_1 + \hat{\delta}_2 = 0$ is generally imposed. Then the respective estimates are $\hat{\delta}_1 = 0.345$; $\hat{\delta}_2 = -0.345$. In the case of I items,

the constraint is generalized to $\sum_{i=1}^I \hat{\delta}_i = 0$. This constraint sets the origin of the scale arbitrarily at 0.0.

Relative Item Difficulties and Conditional Estimates

Notice that the relative difficulties of the items are not given by taking the ratio of the total number correct obtained on each of the items. The numbers correct on items 1 and 2, respectively, are 15 and 25; this ratio gives $15/25 = 0.60$ rather than 0.5. To further consolidate the idea as to how the estimate of the relative difficulties of the items is independent of the abilities of the persons, consider data set 2 of **Table 2**. The ratio of eqn [3], conditional on the total score of 1, is the same as for set 1. However, the total number correct for both items is greater, 18 compared to 15 for item 1, and 28 compared to 25 for item 2. Thus, overall the persons in set 2 are more able than those in set 1. In addition, the ratio of the total number correct changes from $15/25 = 0.60$ to $18/28 = 0.64$. However, the difference in difficulties between the items remains the same as in set 1. This is the sense in which the comparison of the items is invariant relative to the locations of the persons, where the latter may show greater or lesser values overall.

The conditional maximum likelihood solution equation based on eqn [5], of which eqn [5] is a special case, is

$$s_i - \sum_{r=1}^{R-1} n_r \Pr\{x_{ri} = 1 | r\} = 0 \quad [7]$$

where s_i is the total number of correct responses to item i , n_r is the number of persons with a score of r , and

Table 2 Responses of 50 persons to two items: sets 1 and 2

Set 1					Set 2				
Item1	Item2				Item1	Item2			
	1	0	15			1	0	18	
	5	10				5	10		
Item1	0	20	15	35	Item1	0	20	15	32
	25		25	50		28		22	50

$\Pr(x_{ri} = 1|r)$ is the probability that a person with total score r answers item i correctly. This equation gives consistent estimates in the sense that as the number of persons increases, the estimates converge to the correct value, always with the condition that the data fit the model. However, in many educational data collection designs, for example, those concerned with equating and linking which are considered in the section titled 'Applications in test equating and linking,' not all persons respond to all items. Other methods of estimation have been employed to overcome some of the complexities that ensue with this situation. One method that gives consistent estimates and can handle missing data is that of conditional pairwise estimation. This takes the responses of persons to every pair of items, and generalizes eqn [5] directly (Andrich and Luo, 2003) to give the solution equation

$$s_i - \sum_{j, j \neq i} N_{ij} \Pr\{x_{nij} = 1 | r_{nij} = 1\} = \sum_{j, j \neq i} N_{ij} \left\{ \exp(-\delta_i) / (\gamma_{ij}) \right\} \quad [8]$$

where $x_{nij} = 1, r_{nij} = 1$ are, respectively, a correct response of person n to item i and the condition that the total score of the person on items i and j is 1, N_{ij} is the number of persons who have a score of 1 on item i when paired with any other item j , s_i is the number of correct responses to item i among these N_{ij} persons and summed over all items $j \neq i$, and $\gamma_{ij} = \exp(-\delta_i) + \exp(-\delta_j)$. The equation essentially handles missing responses by not summing into either N_{ij} or s_i a person who does not respond to one or both items i and j .

Estimate of the Person Parameter

Comparisons of persons could be estimated by eliminating the item parameters in the way shown above for the comparisons of items. However, because there are generally many more persons than items, this is not feasible. Therefore, estimates of persons are obtained using the item-location estimates as given. This approach also lends itself to conceptualizing two stages in educational measurement: first, the calibration of the items and the check that they are operating as required, and, second, the measurement of the persons. It, too, exploits the sufficient statistic for the person parameter.

From eqn [2] the likelihood L of the vector of responses x_{ni} for a person gives

$$\begin{aligned} L = \Pr\{(X_{ni}) = (x_{ni})\} &= \sum_{i=1}^I \exp(x_{ni}(\beta_n - \delta_i)) / \prod_{i=1}^I \gamma_{ni} \\ &= \exp(r_n \beta_n - \sum_{i=1}^I x_{ni}(\delta_i)) / \prod_{i=1}^I \gamma_{ni} \quad [9] \end{aligned}$$

where the coefficient r_n of the parameter β_n is simply the person's total score. Differentiating eqn [9] with respect to β and setting the derivative to 0, gives the solution equation

$$r_n = \sum_{i=1}^I \exp(\beta_n - \delta_i) / \gamma_{ni} = \sum_{i=1}^I \Pr\{X_{ni} = 1\} \quad [10]$$

Equation [10] needs to be solved iteratively. Notice that for fixed values of (δ_i) , and irrespective of the pattern of responses, the ability estimate will be the same. However, two other points need to be noted simultaneously. First, the required and expected pattern of responses is that the person tends to answer the easier items correctly and the harder items incorrectly. Thus, for ten items ordered in difficulty and a total score of 5, the vector (1,1,1,1,0,1,0,0,0,0) is consistent with the model, while the pattern (1,0,0,1,0,1,0,0,1,1) is not. Fit to the model is considered further under the section titled 'Test of fit of the data to the model.' Second, the person's estimate is based on the items that the person has completed. Therefore, if different persons completed different items from the same set of items, then for the same total score they will have different ability estimates – for the same score, a person who answered more difficult items will have a higher ability estimate than a person who answered easier items. This is an important property of the RM in linking and equating tests which is considered further in the section titled 'Applications in test equating and linking.'

Equation [10] gives the direct maximum likelihood estimates which are slightly biased for a fixed number of items and methods for reducing the bias have been devised.

The locations of the items and persons are said to be in logits.

Standard Errors

With every estimate, there is a standard error. Based again on the theory of maximum likelihood, good approximations are given for the item and person parameters, respectively, by

$$\hat{\sigma}_i = 1 / \sqrt{\sum_{n=1}^N (\Pr\{X_{ni} = 1\})(1 - \Pr\{X_{ni} = 1\})} \quad [11]$$

and

$$\hat{\sigma}_n = 1 / \sqrt{\sum_{i=1}^I (\Pr\{X_{ni} = 1\})(1 - \Pr\{X_{ni} = 1\})} \quad [12]$$

Applications in Test Equating and Linking

The Rasch class of models was developed from a design in which students with reading problems were to be monitored for reading improvement. A scale was to be developed which could measure a range of abilities so that as students improved in their reading ability they could be measured on the same scale. However, as their abilities

increased, they had to be given more difficult texts to read. To construct the scale, students in different grades were given different tests with difficulties similar to the students' abilities (Rasch, 1960).

A modern example of such a situation is placing students in different year groups onto the same achievement scale. Suppose students in years 5 and 6 are to be assessed for their reading ability on the same scale and there was time to give 30 items to each class. To construct the scale, a typical design would have easier items for year 5, more difficult for year 6, with some items in common. The common items, called link items, would typically be more difficult than the majority of items for year 5 and easier than the majority of the items for year 6. For example, suppose 250 students from years 5 and 6 each are assessed on 30 dichotomously scored items and that there are ten linking items. The total number of different items is therefore 50. The person by items data matrix would have structurally missing responses – year 5 students would not have completed items 31–50, and year 6 students would not have completed items 1–20, while both year groups would have completed link items 21–30. Nevertheless, because of the properties of the model and of the solution algorithm for the estimates, the location parameters for all items and all persons could be estimated simultaneously on the same scale.

Table 3 shows summary estimates of such a design. Notice that the mean of the items for year 5 is less than that for year 6, and that the mean of the year 5 students is less than that of the year 6 students.

Figure 3 shows the distribution of student estimates in the two year groups relative to the distribution of item estimates. It is evident that the students' distribution is bimodal reflecting the abilities of the two year groups.

From eqn [10] it is possible to present readily the transformation of every total score to its corresponding location estimate. **Figure 4** shows the transformation functions for the linking design data of **Table 3**. For example, as shown in the **Figure 4**, a score of 25 on the year 5 (easier test) is equivalent to a score of 12.80 on the year 6 (more difficult test). Of course, no student could score 12.80; therefore, 12.80 is interpreted as the expected value (theoretical mean) of all persons who score 25 on the year-5 test.

Test of Fit of the Data to the Model

As illustrated above, invariance of the locations of the items relative to the locations of the persons is a property of the model. Whether or not it is a property of the data is to be tested empirically. In principle, because it is a

Table 3 Estimates of item and person locations

	Item locations			Person locations	
	Year 5 items 1–30	Year 6 items 21–50	Link items 1–10	Year 5	Year 6
Mean	–1.094	1.091	–0.008	–0.203	0.223
St. dev.	1.021	1.019	0.658	1.353	1.266
Mean SE	0.149	0.148	0.106	0.498	0.488
Number	30	30	10	250	250

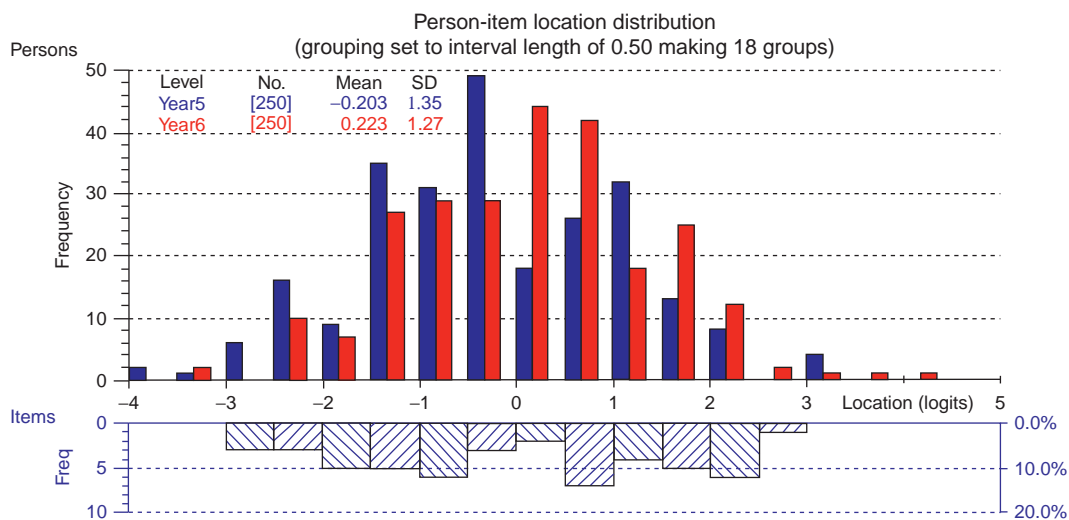


Figure 3 Distribution of student and item locations in the linking design.

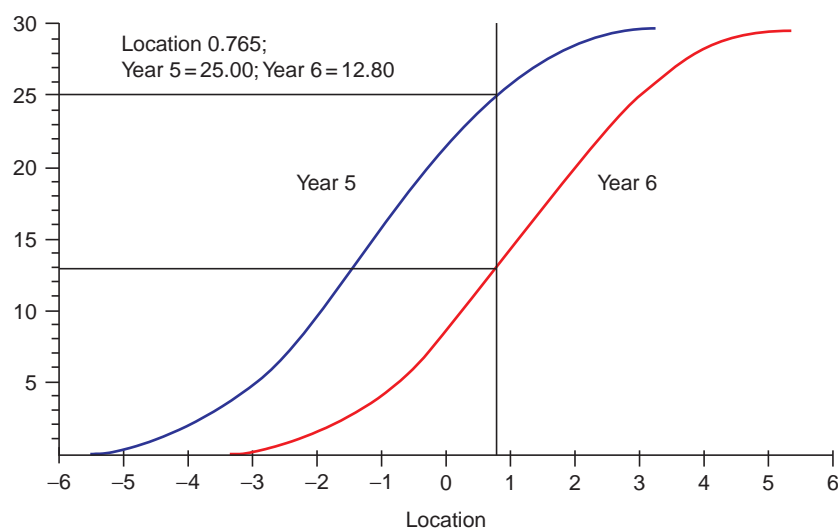


Figure 4 Equating of two tests of different difficulties.

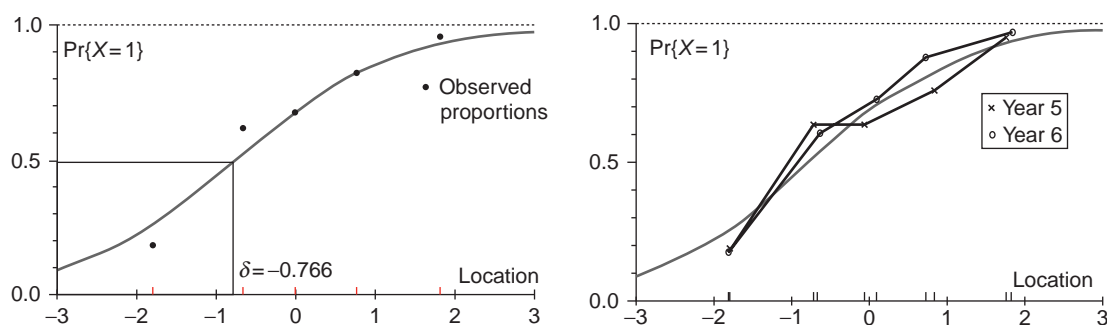


Figure 5 ICC for item 22 of the linking design: observed proportions in five class intervals (left) for two year groups (right).

property of the model, to check if it is a property of the data, the focus is on checking if the data fit the model.

No test of fit is necessary and sufficient and different tests of fit have different power in detecting misfit. The power of the tests of fit depends on the relative locations of persons to items, on the sample size, and how specific the particular hypothesis of misfit is.

The greater the distance between the person and item locations, the greater is the power to detect misfit. However, items close to a person's location provide the smallest standard errors of measurement. Thus, there is a trade-off between precision and the power of the test of fit. One general and one specific test of fit are described below. There are different kinds of test of fit. In this article, only those that can be generalized immediately to polytomous responses are summarized.

A Unified Approach to Differential Item Functioning and General Fit

A general test of fit, which capitalizes on the ICC, is to divide the persons into class intervals according to the

estimates of their locations, and then to compare the proportion of persons in these class intervals who score 1 on the item with the theoretical probability of scoring 1 given by the ICC. The left panel of **Figure 5** shows such information graphically for item 22 in the linking design.

A specific test of fit considers the responses of persons defined in two or more different groups, for example, students in years 5 and 6 in the linking design. In this case, it is necessary to establish that the link items are functioning equivalently across the groups, that is, that they show no differential item functioning (DIF). Again the ICC can be exploited to check this invariance. The requirement is that, irrespective of the year group, the same total score should imply the same probability of success. The right panel of **Figure 5** shows the same information as the left panel for item 22 (a link item), but now divided by the year group to which the students belong. For the first two class intervals, there is little difference between these proportions, but for the last three, the year 6 answers the items correctly at a higher rate.

This graphical test of fit can be complemented statistically. For each response, the standardized residual can be formed.

$$z_{ni} = \frac{x_{ni} - E[X_{ni}]}{\sqrt{V[X_{ni}]}} = \frac{x_{ni} - \Pr\{X_{ni} = 1\}}{\sqrt{\Pr\{X_{ni} = 1\}(1 - \Pr\{X_{ni} = 1\})}} \quad [13]$$

If the data fit the model, then these residuals should have a mean of 0 and a standard deviation close to 1.0. The resulting matrix can be partitioned by class intervals and groups giving a two-way structure for the residuals. This structure can be analyzed by analysis of variance (ANOVA) to indicate whether there is any effect (1) across the class intervals irrespective of groups, (2) across the groups irrespective of class intervals, and (3) whether there is an interaction between the two. The first of these checks the fit across the class intervals and therefore the continuum irrespective of DIF, the second checks for possible DIF, called uniform DIF, irrespective of fit across the continuum, and the third checks for an interaction between the two, commonly called nonuniform DIF.

The number of class intervals is arbitrary, but ideally there should be more than 20 persons for each cell of the two-way matrix of residuals. The ANOVA is robust against violations of its assumptions and takes into account the sample size. If different numbers of class intervals makes a difference in the decision that would be made regarding a significant effect, that suggests that the significance is marginal, and should be interpreted as such.

This ANOVA can be summarized readily in a single table to provide a unified test of fit. **Table 4** shows a summary of this analysis for the ten common items in the linking design used above. There is only evidence of very marginal misfit at the item level in **Table 4**. Item 23 shows marginal misfit across the class intervals, and item 22, which is displayed graphically in **Figure 5** and shows the worst uniform DIF, has $P < 0.08$. No item shows a significant nonuniform DIF. There are 30 tests of fit in **Table 4**, and therefore at the 5% level, up to two could be significant just by chance. Thus, these data fit excellently according the ANOVA analysis and for the given sample size. Excellent fit is required of linking items in such designs.

The same residuals can be studied in other standard ways to detect various violations of the data to the model. For example, they may be correlated, and large correlations between pairs of items would indicate some kind of dependence between the items, for example, either they assess a second dimension or they violate statistical independence.

Relationship to TTT

The RM is seen as part of the general area of item response theory (IRT). In this construction, it can be seen as in opposition to TTT (Lord and Novick, 1968). Conversely, it can be seen as formalizing the principles of TTT, in so doing, resolving two of its paradoxes.

The key identity between the RM and TTT is that the

total score, $r_n = \sum_{i=1}^I x_{ni}$, for a person is the key statistic that

indicates the relative status of the person. It is this by definition in TTT, while it arises from the requirement of invariance in the RM. Tests of fit to the RM can be seen as providing a check on whether the total score can be used meaningfully. Then the logit estimates from transformed raw scores in the RM, which are additive in the person and item parameters, satisfy the conditions of interval measurement for purposes of calculating means, variances, and correlations, better than do the raw scores (Embretson, 1996). In this sense, the RM improves upon TTT.

Reliability

In TTT, the key statistic for the assessment of the operation of a test is its reliability, often denoted as r_{tt} . This statistic is anchored in the sample (or population of the sample) to which the test is administered and defined generically as the proportion of true variance relative to

Table 4 Statistical analysis of item fit for the link items

Class interval (CI)					Year (uniform DIF)				CI by year (nonuniform DIF)			
Item	MS	F	DF	P<	MS	F	DF	P<	MS	F	DF	P<
21	0.70	0.85	7	0.55	2.14	2.57	1	0.11	0.49	0.59	7	0.77
22	1.62	1.86	7	0.07	2.64	3.03	1	0.08	0.33	0.38	7	0.91
23	2.32	2.50	7	0.02	0.03	0.03	1	0.86	0.82	0.88	7	0.52
24	1.07	1.12	7	0.35	1.19	1.25	1	0.26	0.58	0.61	7	0.75
25	0.71	0.70	7	0.67	0.84	0.83	1	0.36	0.66	0.65	7	0.72
26	0.65	0.68	7	0.69	1.78	1.86	1	0.17	0.79	0.83	7	0.56
27	0.95	1.00	7	0.43	0.12	0.13	1	0.72	0.88	0.92	7	0.49
28	1.56	1.53	7	0.16	1.90	1.86	1	0.17	0.95	0.93	7	0.48
29	1.30	1.56	7	0.15	0.68	0.81	1	0.37	0.55	0.67	7	0.70
21	1.15	1.21	7	0.30	0.06	0.07	1	0.80	0.70	0.73	7	0.64

the total variance (which includes the error of measurement). It takes the form

$$r_{tt} = \frac{\sigma_{\beta}^2}{\sigma_{\beta}^2 + \sigma_e^2} \quad [14]$$

where, using the notation of the RM above, σ_{β}^2 is the (true) variance among the persons in the relevant population and σ_e^2 is the error variance. This relationship is based on the fundamental equation of TTT:

$$r_{tt} = \beta_{tt} + \varepsilon_{tt} \quad [15]$$

where ε_{tt} is the error between the observed score r_{tt} and the ability β_{tt} .

There are different ways of estimating r_{tt} , a common one being Cronbach's α which can be calculated from one administration of the test and which is therefore an index of the internal consistency of the items relative to the sample of persons. An index α_R which is comparable to α , both in concept and in value, can be calculated from an RM analysis constructed as

$$\alpha_R = \frac{\hat{\sigma}_{\beta}^2}{\hat{\sigma}_{\beta}^2 + \hat{\sigma}_e^2} = \frac{\hat{\sigma}_{\beta}^2 - \hat{\sigma}_e^2}{\hat{\sigma}_{\beta}^2} \quad [16]$$

where $\hat{\sigma}_{\beta}^2$ is the variance of the estimates of the person parameters obtained from eqn [10], and $\hat{\sigma}_e^2$ is the mean error variance obtained as the mean of the squares of the standard errors obtained from eqn [12].

This index, as the traditional reliability, is a function of the internal consistency of the items and, in the same way as the traditional reliability, a function of the true variance of the sample. It is a useful index to check the degree to which the test separates the persons in the particular sample. Depending on the circumstances, this may or may not be required to be high.

In the RM, this is also an index of the power of the test of fit of the data to the model. Thus, if the value of α_R is low, for example, 0.5, then it indicates that the persons have not been separated very well by the test and the power to detect misfit is not high. Therefore, before reporting good fit to the model, this index should be checked. Therefore, also, when a test is being developed, it is necessary to have a relatively high value, of the order of 0.9, in order to check the functioning of the items in the test.

As the RM handles missing data, this index can be readily calculated and interpreted in the presence of missing data.

Resolution of the Attenuation and Standard Error Paradoxes

In TTT, the goal is to maximize the reliability of a test. One paradox of TTT is that as the reliability becomes higher beyond a certain point, then the validity becomes lower. This can be seen from the condition of maximum

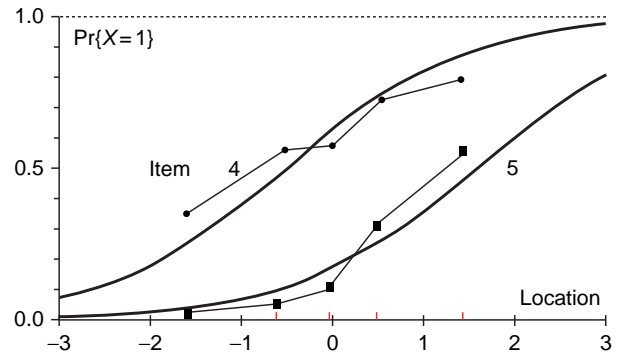


Figure 6 ICCs of two items and observed proportions in five class intervals.

reliability which arises when (1) all items have the same difficulty (2) their difficulty is the mean of the person abilities, and (3) the correlations among the items is 1. In this case, every person has the easiest half of the items correct and the most difficult half of the items incorrect. However, now the validity of the test is that of just one item with all other items redundant!

To see how the RM resolves this paradox, it is helpful to consider explicitly the idea of the discrimination of an item. **Figure 6** shows ICCs of two items and the observed proportions in five class intervals. Even though the observed proportions in both items are relatively small, and unlike item 22 in the left panel of **Figure 5**, they show a systematic deviation from the ICC. For item 4 they are flatter than the ICC, and for item 5 they are steeper. The former item is said to discriminate less than the ICC and the latter to discriminate more than the ICC, which, as noted earlier is a kind of average of the discriminations of all items.

The attenuation paradox can be related to the RM by recognizing that a set of items shows greater reliability as the discrimination of the items increases; also because reliability is to be maximized in TTT, items with lower discrimination are eliminated and those with higher are retained. For example, item 4 of **Figure 6** might be eliminated, but item 5 would be retained. As this process is continued with item selection, the items will have a greater average slope, be more correlated with each other, and lead to higher reliability but eventually to lower validity. In contrast, with the ICC as the criterion for the observed proportions in the RM, those items with a very high discrimination as well as those with a very low discrimination would be considered for modification or elimination. Thus, the average discrimination of the original items is retained rather than being successively increased.

It is stressed that when an item is identified as having an anomalous discrimination, it should be studied for a substantive explanation of the anomaly. If it is understood, then the item can be modified or discarded. In general, an

item with a discrimination greater than the ICC suggests that it has very much in common with the other items and is possibly redundant; items with a discrimination less than the ICC suggests that a construct other than the dominant one is being assessed by the item.

A second paradox of TTT has two related aspects: first that a difference between a pair of total scores has the same numerical effect across the continuum; and second that the standard errors of measurement are calculated to be identical for every score. However, it is commonly appreciated that at the extremes, where persons have most items correct or incorrect and a floor or ceiling effect operates, the same difference in scores must have a different meaning from the middle scores of the test and that the standard error of measurement must be greater than in the middle scores of the test. The linearization of raw scores in the RM, as is shown in **Figure 4**, takes account of the former paradox; higher standard errors at the extreme scores than in the middle scores follow directly from eqn [12].

Controversy and the RM

The construction of a model based on *a priori* requirements and independent of data involves a different paradigm from the traditional one in the data-model relationship. In the traditional paradigm, if the model does not fit the data, then the task is to find a model which accounts for the data better. In the case of dichotomous items, this generally means applying a model which has a discrimination parameter, in addition to the location parameter, for an item (Birnbau, 1968). In the Rasch paradigm, the emphasis is on whether the data fit the chosen model, and if not, then the task is to understand the features of the items or the test administration that might have created this misfit with a view to modifying the data collection in some way. This modification of the data collection is intended to increase the validity of the data as well as the fit to the model.

Advocates of the RM consider that its properties of fundamental measurement (Brogden, 1977; Wright, 1997), which are found in physics, are important in improving the quality of social measurements. The disclosure of problems in data as anomalies in measurement that are to be considered qualitatively and substantively, rather than as a problem with the model, is consistent with the role of measurement in physical science (Andrich, 2004). However, some controversy, which has also been discussed in Andrich (2004), has arisen from this distinctive paradigm of the RM.

Short biographies of Rasch can be found in Andersen and Olsen (2000), Andrich (2005), and Wright (1980) and a more extensive study of Rasch's contributions to statistics is provided in Olsen (2003).

The RM for Polytomous Ordered Categories

The RM for polytomous ordered categories is a direct extension of the model for just two categories. It is often termed the polytomous RM. Following a sequence of derivations in Rasch (1961), Andersen (1977), Andrich (1978), and Wright and Masters (1982), the model may be expressed in the form

$$\Pr\{X_{ni} = x\} = (\exp(x\beta_n - \sum_{k=0}^x \delta_{ki})) / \gamma_{ni} \quad [17]$$

where (1) $X_{ni} = x \in \{0, 1, 2, \dots, m_i\}$ is an integer random variable characterizing $m_i + 1$ successive categories, (2) δ_{ki} , $k = 1, 2, 3, \dots, m_i$ represents m_i thresholds for item i which divide the continuum into $m_i + 1$ ordered categories with $\delta_{i0} \equiv 0$, and (3) $\gamma_{ni} = \sum_{x=0}^{m_i} (\exp x\beta_n - \sum_{k=0}^x \delta_{ik})$ is again the normalizing factor that ensures that the probabilities in eqn [17] sum to 1. The thresholds are points at which the probabilities of responses in one of the two adjacent categories are equal. If $m_i = 1$, which is the case of two categories, then eqn [17] specializes to eqn [1].

The Latent Structure of the RM

In a response design, such as that with four categories shown in **Table 1**, there is only one response in one of the categories. However, it can be envisaged that a judge making a decision into one category considers adjacent categories in pairs, decides in which of the pairs the object of assessment is to be categorized, and then settles on a decision between one final pair, and gives a response. These pairwise decisions must be latent as they are never observed. To understand this latent structure, let $X_{ni} = x \in \{0, 1, 2, \dots, m_i\}$ be an integer random variable characterizing the response in category x , and let the probability P_x of a response x in the higher of the two successive categories $x - 1$ and x be deemed a successful response. Then this probability can be written as

$$P_x = \frac{\Pr\{X_{ni} = x\}}{\Pr\{X_{ni} = x - 1\} + \Pr\{X_{ni} = x\}} \quad [18]$$

and its complement

$$Q_x = 1 - P_x = \frac{\Pr\{X_{ni} = x - 1\}}{\Pr\{X_{ni} = x - 1\} + \Pr\{X_{ni} = x\}} \quad [19]$$

It can be shown readily that

$$\Pr\{X_{ni} = x\} = P_1 P_2 P_3 \dots P_x Q_{x+1} Q_{x+2} \dots Q_m / D \quad [20]$$

where $D = Q_1 Q_2 Q_3 \dots Q_m + P_1 Q_2 Q_3 \dots Q_m + P_1 P_2 Q_3 \dots Q_m + P_1 P_2 P_3 \dots P_m$.

From eqn [20] it is evident that the particular response $X_{ni} = x$ implies successes at the first x thresholds and failures at all the remaining thresholds. This is compatible with the decision in one of a number of ordered categories, in the sense that if there is a failure at a particular threshold, then there is an implied failure at all the thresholds that are further along the continuum; and if there is a success at a particular threshold, then there is an implied success at all the preceding thresholds. The response in a category is between a success and a failure at two successive thresholds. It is also evident from eqn [20] that the probability of a response in any category is a function of the latent probabilities of success and failures at all categories.

Equation [18] does not rest on the application of the dichotomous RM. However, let the

$$P_x = \exp(\beta_n - \delta_{ni}) / \eta_{ni} \quad [21]$$

where $\eta_{ni} = 1 + \exp(\beta_n - \delta_{ki})$. This is simply the RM of eqn [1] at each threshold x . Then inserting eqn [21] in eqn [18] gives the RM of eqn [17].

The left panel of **Figure 7** shows the probabilities of responses in each score x , which corresponds to the successive categories of **Table 1**, fail (F), pass (P), credit (C), and distinction (D), known as category characteristic

curves (CCCs). It also shows the location of the thresholds between adjacent categories which are points of equal probability of response between two adjacent categories.

The right panel shows the expected value curve, given by $E[X] = \sum_x x \Pr\{X = x\}$, which specializes to $E[X] = \Pr\{X = 1\}$ in the dichotomous case. $E[X]$, whose values can range from 0 to m_i , can be used analogously to the ICC curve in tests of fit, including DIF. The panel shows observed means of persons in five class intervals.

The Latent Thresholds

It is instructive to consider the interpretation of the latent dichotomous responses of eqn [18]. These are shown as dotted lines in the left panel of **Figure 8** superimposed on the CCCs of **Figure 7**, and then on their own in the right panel of **Figure 8**. Importantly, these latent dichotomous responses are equivalent in theory to observed dichotomous responses at the same thresholds which are independent. Specifically, they are equivalent to three different judges deciding independently whether the performance was successful or not in meeting the standards of pass, credit, and distinction, respectively. In the case of independent dichotomous judgments of that kind, and with the dichotomous RM applied, the thresholds are

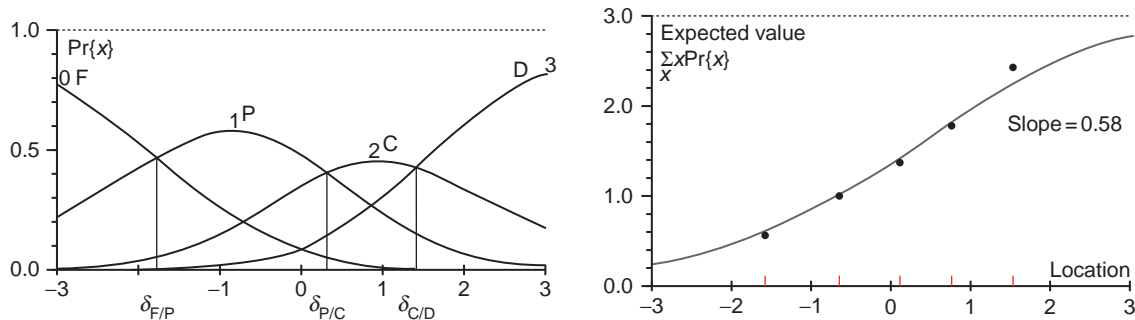


Figure 7 (Left) Category characteristic curves and (right) the expected value curve for an item with four ordered categories.

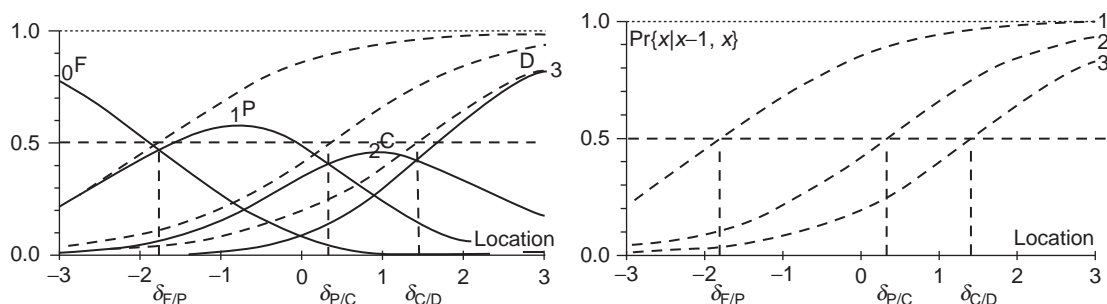


Figure 8 (Left) Category characteristic curves and (right) latent dichotomous responses for an item with four categories.

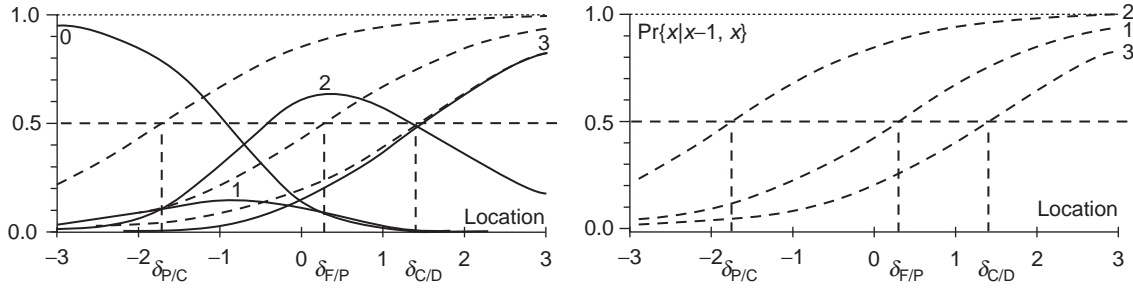


Figure 9 (Left) Category characteristic curves and (right) latent dichotomous responses for an item with four categories with reversed threshold estimates.

better notated as $\delta_{\tilde{P}/P}, \delta_{\tilde{C}/C}, \delta_{\tilde{D}/D}$ where $\tilde{P}, \tilde{C}, \tilde{D}$ indicates not meeting the respective standards. This notation indicates that the decisions are in just two categories, success or lack of success, at each of the three different standards. In terms of the two notations, the equivalence of the responses gives

$$\{\delta_{F/P}, \delta_{P/C}, \delta_{C/D}\} \equiv \{\delta_{\tilde{P}/P}, \delta_{\tilde{C}/C}, \delta_{\tilde{D}/D}\} \quad [22]$$

If the judgments were indeed independent, for example, three different judges made judgments independently at each of the standards of the same performances, it would be required that the threshold estimates show natural difficulty ordering, that is $\delta_{\tilde{P}/P} < \delta_{\tilde{C}/C} < \delta_{\tilde{D}/D}$. For example, it would be untenable that the judge at pass had a higher standard than a judge at credit. However, from eqn [22], this means that in the standard design, it is required that $\delta_{F/P} < \delta_{P/C} < \delta_{C/D}$, that is, that the thresholds in the RM are in their natural order.

Reversed Thresholds

If in the above example three judges did respond independently with a judgment at each of the grade levels, in principle, it would be possible for the data to show that $\delta_{\tilde{P}/P} > \delta_{\tilde{C}/C}$ or $\delta_{\tilde{C}/C} < \delta_{\tilde{D}/D}$. That is, if the judge at pass operated at a higher standard than the judge at credit, there is nothing in the analysis that would preclude the estimates $\delta_{\tilde{P}/P} > \delta_{\tilde{C}/C}$ to be observed. In this case, the threshold estimates are reversed relative to their natural order.

Interestingly, and importantly, if the data are collected in the standard ordered category format of Table 1, then the threshold estimates can show reversals. Figure 9 shows the same information as Figure 8 but for an item with reversed thresholds. The threshold estimates of Figure 9 violate the requirements of eqn [22]. Therefore, it would be necessary to examine the reasons for the inconsistency of the application of the relative standards. Any number of factors can contribute to such anomalous responses, with the factors depending on the item and the context of its administration. If the categories are not working as intended, it calls into question the understanding what it means to have more or less of property as operationalized

by the categories. The property of the RM to disclose whether or not the empirical ordering of the categories is working as intended is a major justification for using the model. These interpretations have arisen from the mathematical derivations that follow from Rasch's formalization of the requirement for invariant comparisons and are unlikely to have been arrived at in any other way.

See also: Classical Test Theory Reliability; Differential Item Functioning; Educational Measurement: Overview; Item Response Theory.

Bibliography

- Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika* **42**, 69–81.
- Andersen, E. B. and Olsen, L. W. (2000). The life of Georg Rasch as a mathematician and as a statistician. In Boomsma, A., van Duijn, M. A. J., and Snijders, T. A. B. (eds.) *Essays in Item Response Theory*, pp 3–24. New York: Springer.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika* **43**, 357–374.
- Andrich, D. (2004). Controversy and the Rasch model: A characteristic of incompatible paradigms? *Medical Care* **42**, 7–16.
- Andrich, D. (2005). Georg Rasch: Mathematician and statistician. In Kempf-Leonard, K. (ed.) *Encyclopedia of Social Measurement*, vol. 3, pp 299–306. Amsterdam: Academic Press.
- Andrich, D. and Luo, G. (2003). Conditional estimation in the Rasch model for ordered response categories using principal components. *Journal of Applied Measurement* **4**, 205–221.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord, F. M. and Novick, M. R. (eds.) *Statistical Theories of Mental Test Scores*, pp 397–545. Reading, MA: Addison-Wesley.
- Brogden, H. E. (1977). The Rasch model, the law of comparative judgement and additive conjoint measurement. *Psychometrika* **42**, 631–634.
- Embretson, S. E. (1996). Item response theory models and inferential bias in multiple group comparisons. *Applied Psychological Measurement* **20**, 201–212.
- Engelhard, G., Jr. and Wilson, M. (eds.) (1996). *Objective Measurement: Theory into Practice*, vol. 3. Norwood, NJ: Ablex Publishing.
- Fischer, G. H. and Molenaar, I. W. (eds.) (1995). *Rasch Models: Foundations, Recent Developments, and Applications*. New York: Springer.
- Guttman, L. (1950). The basis for scalogram analysis. In Stouffer, S. A., Guttman, L., Suchman, E. A., et al. (eds.) *Measurement and Prediction*, pp 60–90. New York: Wiley.
- Lord, F. M. and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

- Olsen, L. W. (2003). *Essays on Georg Rasch and His Contributions to Statistics*. Copenhagen: University of Copenhagen, Institute of Economics.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research. [Expanded edition (1980) with foreword and afterword by Wright, B. D. Chicago, IL: The University of Chicago Press.].
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In Neyman, J. (ed.) *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. IV, pp 321–334. Berkeley, CA: University of California Press.
- Smith, E. V., Jr. and Smith, R. M. (eds.) (2004). *Introduction to Rasch Measurement*. Maple Grove, MN: JAM Press.
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology* **33**, 529–554.
- Van der Linden, W. and Hambleton, R. (eds.) (1997). *Handbook of Modern Item Response Theory*, pp 139–152. New York: Springer.
- Wright, B. D. (1997). A history of social science measurement. *Educational Measurement: Issues and Practice* **16**(4), 33–45.
- Wright, B. D. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago IL: University of Chicago Press. [Reprint 1960. Rasch G. (ed.).].
- Wright, B. D. and Masters, G. N. (1982). *Rating Scale Analysis: Rasch Measurement*. Chicago, IL: MESA Press.

Relevant Websites

<http://www.rasch.org> – Institute for Objective Measurement.
<http://www.jampress.org> – Journal of Applied Measurement.

Scoring Rubrics

J A Arter, Educational Testing Service, Portland, OR, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Analytic rubric – A type of rubric that provides ratings of a performance or product along several different, important dimensions. The term is used to describe one type of general rubric. This is also called analytical trait. Contrast this with holistic rubric.

Assessment of learning – Student assessment, the goal of which is to provide information to judge the adequacy of the level of student learning at a point in time. This is also called summative assessment. Contrast this with assessment for learning.

Assessment for learning – Student assessment, the goal of which is to provide information used to improve student learning while the learning is occurring. Assessment for learning includes actions on the part of teachers such as using assessment information to plan the next steps in instruction and actions on the part of students such as student self-assessment and goal setting. This is sometimes called formative assessment. Contrast this with assessment of learning.

General analytic rubric – A general rubric that results in several judgments about the quality of a product or performance, one for each important dimension of the product or performance.

General holistic rubric – A general rubric that results in a single overall judgment about the quality of a product or performance.

General rubric – A scoring guide designed so that it can be used across tasks measuring the same student learning goal(s). This contrasts with a task-specific scoring guide, which can be used for a single task only. A general rubric is useful for reasoning, performance skills, and product-learning goals. General rubrics can be holistic or analytical. See definitions of the terms general holistic and general analytical.

Holistic rubric – A type of rubric that results in a single, overall rating for an entire performance or product. The term is used to describe one type of general rubric. Contrast this with analytic rubric.

Knowledge-learning goals – The learning goals for students that require them to learn information, including bodies of knowledge and how information interconnects.

Performance assessment – An assessment based on observation and judgment. One of several methods of assessment useful in the classroom. Particularly apt for assessing student reasoning proficiencies, performance skills, and products. Rubrics are used to help make this essentially subjective form of assessment as objective as possible.

Performance-skills learning goals – The learning goals for students which require a behavioral demonstration of competence. For example, reading fluently, giving an oral presentation, playing a musical instrument, and working productively in a group.

Product-learning goals – The learning goals for students that require them to make something. Examples are graphs, laboratory reports, visual arts, and maps.

Reasoning-learning goals – The learning goals for students that require them to use information to solve a problem, analyze, compare, infer, make a decision, and so forth.

Rubric – The written criteria that define the characteristics of quality reasoning, performance skills, or products. Good rubrics have levels defined using indicators and/or descriptors. A rubric can be used across tasks assessing the same learning goal(s). See also general rubric, general holistic rubric, and general analytic rubric.

Scoring guide – A written set of statements that describe how student performance on a task will be evaluated. All rubrics are scoring guides, but not all scoring guides are rubrics. For example, task-specific scoring guides are not rubrics.

Task – The activity, assignment, or exercise given to students to do, the results of which will be evaluated using a rubric. The role of the task is to elicit the desired reasoning, performance skill, or product, so that evaluation using the rubric is possible.

Task-specific scoring guide – A scoring guide designed to be used with only a single task. This contrasts with a general rubric, which can be used across tasks measuring the same learning goal(s). Task-specific scoring is best used for extended written response assessments that focus on student knowledge.

The Role of Scoring Guides

Performance assessment is assessment based on observation and judgment (Stiggins *et al.*, 2006: 92): an evaluator observes a performance or product and makes a judgment as to its quality. The challenge with performance assessment is to make this essentially subjective form of assessment as objective as possible.

Promoting objectivity is the traditional role of scoring guides and rubrics. Good ones define the characteristics of various levels of quality of a product or performance so clearly that different judges can independently agree on an evaluation of quality. However, good scoring guides and rubrics have come to play other important roles as well. They can do the following:

- Define complex student learning goals so that teachers have a common vision that guides instruction. Examples are oral presentations, writing, mathematical problem solving, laboratory reports, literary and music criticism, research reports, music production, group collaboration/teaming, and speaking a foreign language.
- Help students understand characteristics of high-quality products and performances so that they, too, can incorporate those features into their own work.

This evolving use of scoring guides and rubrics has paralleled the evolving notions of assessment of and for learning. Assessment of learning occurs when assessment information is used to judge the adequacy of the level of student proficiency at a point in time, as with a grade or meeting a standard. Assessment for learning occurs during instruction to provide feedback to teachers and students on how to improve performance prior to the next assessment of learning event.

Assessment of learning is summative in nature while assessment for learning is formative. Formative assessment traditionally has meant actions on the part of adults to improve student learning. Groups of teachers might, for example, examine student work to determine who is getting it and who is not so that remediation can take place. However, assessment for learning includes students as data-based decision makers too; they use assessment information to take actions to improve their own learning.

There is growing body of research that shows that when assessment for learning strategies are used – both those involving teachers and those involving students – learning dramatically improves (e.g., Black and Wiliam, 1998; Black, 2003; Black *et al.*, 2002; Brookhart, 2005; Hattie and Timperley, 2007; Dochy *et al.*, 1999; Shepard, 2001; Guskey, 2005). Rubrics are the classic example of tools to promote assessment for learning because they assist students to (paraphrased from Sadler, 1989: 119): see where they are going, discover where they are now, and

close the gap between the two. The impact of rubrics on student learning forms a prominent subset of the research cited above.

Types and Uses of Scoring Rubrics

Scoring guides and rubrics come in several different forms, all of which are viable options depending on the learning goals to be assessed and the purpose for doing the assessment.

Task-Specific versus General Scoring Guides

One important distinction is the difference between task-specific and general scoring and when to use each.

Task-specific scoring guides provide a list of specific facts, features, or information to look for in student responses to a task. This results in either assigning points for the presence of specific facts (as in **Figure 1**) or determining a level of quality based on the presence of certain features specific to the task (as in **Figure 2**). Note that the task in **Figure 1** asks students to detail their understanding of the carbon cycle. Responses are scored on the presence of specific pieces of information. The scoring guide cannot be used for any other task, it is task specific. The task in **Figure 2** asks students to write a story based on a picture prompt. The quality of the writing is largely judged based on features specific to this prompt. Most of the scoring guide could not be used to judge any other piece of writing; thus, major portions are, again, task specific.

General, or generic, rubrics can be used with any task designed to assess the same student-learning goal. Examples are in **Figure 3** (the 6+1-trait writing-assessment rubric), **Figure 4** (analytical trait rubric for mathematical problem solving), **Table 1** (analytical trait rubric for oral presentation), and **Table 2** (holistic rubric for inferences). The 6+1-trait rubric can be used for any writing, the inference rubric can be used for any inference, the math rubric can be used for any math problem, and the oral presentation rubric can be used for any oral presentation.

Task-specific scoring is a good choice only when assessing knowledge using extended written response forms of assessment, as in **Figure 1**. Such scoring can summarize the current level of student understanding (assessment of learning) or pinpoint areas of student understanding and where continued study is needed (assessment for learning). When assessing products (as with writing), performance skills (as with oral presentations), or reasoning (as with mathematical problem solving or making inferences) – where it is necessary to judge the level of quality of work rather than the presence or absence of specific knowledge – general rubrics should be used, especially if the purpose is assessment for learning.

Task:

Based on your understanding of the carbon cycle, please describe why we need to know about it and how it works. Be sure to include:

- Why it is important to understand the carbon cycle: 5 points, one for each reason.
- The four major places (reservoirs) we studied where carbon is stored: 4 points.
- At least six ways that carbon gets transferred from one place to another: 6 points.

Task-specific scoring guide:

Why it is important to understand the carbon cycle. One point for any five of the following:

- Carbon is stored in the atmosphere in the form of carbon dioxide.
- Carbon dioxide in the atmosphere can cause greenhouse warming of the planet.
- In recent times geologically, carbon has stayed in balance: the amount of carbon being added to the atmosphere is balanced by the amount of carbon lost from the atmosphere.
- A balanced carbon cycle keeps the planet from warming up or cooling down.
- Humans could cause this system to become unbalanced through burning fossil fuels and other activities.
- A carbon cycle out of balance would add more carbon to the atmosphere than is being removed through natural means and could lead to global warming.
- The extent to which global warming due to human activity is actually occurring is currently being researched and debated among scientists.

The major reservoirs of carbon. One point for each of the following:

Atmosphere, oceans, land/sediments, and plants/animals

How carbon moves from one place to another. One point for any six of the following:

- Carbon moves from the atmosphere to plants through photosynthesis.
- Carbon moves from the atmosphere to oceans by dissolving in places it is cold.
- Carbon moves from the oceans to the atmosphere by evaporation where it is hot.
- Carbon moves from plant/animals to the atmosphere through breathing/respiration.
- Carbon moves from land to the atmosphere through
 - ✓ Fires
 - ✓ Volcanic eruptions
 - ✓ Burning fossil fuels
- Carbon moves from the land into the oceans through erosion.
- Carbon moves from plants/animals to oceans when animals die.

Figure 1 Example of assessing a body of knowledge with extended written response – the carbon cycle. Adapted from Arter, J. A. and Chappuis, J. (2006). *Creating and Recognizing Quality Rubrics*, p. 21, Princeton, NJ: Educational Testing Service, with permission from ETS.

Writing prompt:

Here is a picture of a kangaroo in Australia. Look at the picture for a while. What do you think is happening? Where do you suppose the kangaroo came from? Where do you think he is going? Look how high he jumps! Why do you suppose he is jumping over the fence? Write a story about what is happening in the picture.

Score-point categories:

(please note that the task-specific portions of the scoring guide are highlighted)

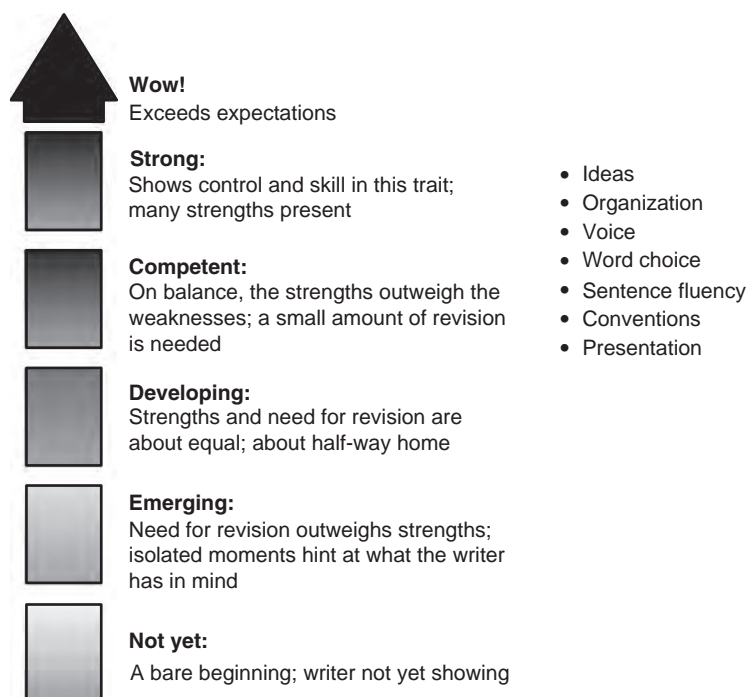
Score of 4 = These writers will enter into the picture imaginatively using such devices as dialogue and character names. There will be clear structure and a sense of drama, which will answer the what and why of the question. The where may or may not be described, but the paper will have a sense of substance. The writing will usually have some spelling and sentence construction errors but will show a clear sense of English syntax.

Score of 3 = These papers will leave out parts of the picture, such as the fence or the jump, and will have less completeness or drama than the 4 paper. They may give commentary or explanation rather than a story. They will have substance, some sense of drama, and some imagination, and they will deal with the what and the why of the question.

The writing need not have much subordination or syntactic fluency, but it will not be so filled with errors as to be difficult to read.

Etcetera ...

Figure 2 Kangaroo – task-specific scoring guide for writing. Adapted from Educational Testing Service (1990), with permission from ETS.



Ideas and content (development)

5. This paper is clear and focused. It holds the reader's attention. Relevant anecdotes and details enrich the central theme.
 - A. The topic is narrow and manageable.
 - B. Relevant, telling, quality details give the reader important information that goes beyond the obvious or predictable.
 - C. Reasonably accurate details are present to support the main ideas.
 - D. The writer seems to be writing from knowledge or experience; the ideas are fresh and original.
 - E. The reader's questions are anticipated and answered.
 - F. Insight – an understanding of life and a knack for picking out what is significant – is an indicator of high level performance, though not required.
3. The writer is beginning to define the topic, even though development is still basic or general.
 - A. The topic is fairly broad; however, you can see where the writer is headed.
 - B. Support is attempted, but does not go far enough yet in fleshing out the key issues or story line.
 - C. Ideas are reasonably clear, though they may not be detailed, personalized, accurate, or expanded enough to show indepth understanding or a strong sense of purpose.
 - D. The writer seems to be drawing on knowledge or experience, but has difficulty going from general observations to specifics.
 - E. The reader is left with questions. More information is needed to fill in the blanks.
 - F. The writer generally stays on the topic but does not develop a clear theme. The writer has not yet focused the topic past the obvious.
1. As yet, the paper has no clear sense of purpose or central theme. To extract meaning from the text, the reader must make inferences based on sketchy or missing details. The writing reflects more than one of these problems:
 - A. The writer is still in search of a topic, brainstorming, or has not yet decided what the main idea of the piece will be.
 - B. Information is limited or unclear or the length is not adequate for development.
 - C. The idea is a simple restatement of the topic or an answer to the question with little or no attention to detail.
 - D. The writer has not begun to define the topic in a meaningful, personal way.
 - E. Everything seems as important as everything else; the reader has a hard time sifting out what is important.
 - F. The text may be repetitious, or may read like a collection of disconnected, random thoughts with no discernable point.

Figure 3 Continued

Organization

5. The organization enhances and showcases the central idea or theme. The order, structure, or presentation of information is compelling and moves the reader through the text.
 - A. An inviting introduction draws the reader in; a satisfying conclusion leaves the reader with a sense of closure and resolution.
 - B. Thoughtful transitions clearly show how ideas connect.
 - C. Details seem to fit where they're placed; sequencing is logical and effective.
 - D. Pacing is well controlled; the writer knows when to slow down and elaborate, and when to pick up the pace and move on.
 - E. The title, if desired, is original and captures the central theme of the piece.
 - F. Organization flows so smoothly the reader hardly thinks about it; the choice of structure matches the purpose and audience.
3. The organizational structure is strong enough to move the reader through the text without too much confusion.
 - A. The paper has a recognizable introduction and conclusion. The introduction may not create a strong sense of anticipation; the conclusion may not tie up all loose ends.
 - B. Transitions often work; other times connections between ideas are fuzzy.
 - C. Sequencing shows some logic, but not under control enough that it consistently supports the ideas. In fact, sometimes it is so predictable and rehearsed that the structure takes attention away from the content.
- D. Pacing is fairly well controlled, though the writer sometimes lunges ahead too quickly or spends too much time on details that do not matter.
- E. A title (if desired) is present, although it may be uninspired or an obvious restatement of the prompt or topic.
- F. The organization sometimes supports the main point or story line; other times, the reader feels an urge to slip in a transition or move things around.
1. The writing lacks a clear sense of direction. Ideas, details, or events seem strung together in a loose or random fashion; there is no identifiable internal structure. The writing reflects more than one of these problems:
 - A. There is no real lead to set up what follows, no real conclusion to wrap things up.
 - B. Connections between ideas are confusing or not even present.
 - C. Sequencing needs lots and lots of work.
 - D. Pacing feels awkward; the writer slows to a crawl when the reader wants to get on with it, and vice versa.
 - E. No title is present (if requested), or if present, does not match well with the content.
 - F. Problems with organization make it hard for the reader to get a grip on the main point or story line.

Voice

5. The writer speaks directly to the reader in a way that is individual, compelling and engaging. The writer crafts the writing with an awareness and respect for the audience and the purpose for writing.
 - A. The tone of the writing adds interest to the message and is appropriate for the purpose and audience.
 - B. The reader feels a strong interaction with the writer, sensing the person behind the words.
 - C. The writer takes a risk by revealing who he or she is consistently throughout the piece.
 - D. Expository or persuasive writing reflects a strong commitment to the topic by showing why the reader needs to know this and why he or she should care.
 - E. Narrative writing is honest, personal, engaging and makes you think about and react to the author's ideas and point of view.
3. The writer seems sincere, but not fully engaged or involved. The result is pleasant or even personable, but not compelling.
 - A. The writer seems aware of an audience but discards personal insights in favor of obvious generalities.
 - B. The writing communicates in an earnest, pleasing, yet safe manner.
- C. Only one or two moments here or there intrigue, delight, or move the reader. These places may emerge strongly for a line or two, but quickly fade away.
- D. Expository or persuasive writing lacks consistent engagement with the topic to build credibility.
- E. Narrative writing is reasonably sincere, but doesn't reflect unique or individual perspective on the topic.
1. The writer seems indifferent, uninvolved, or distanced from a the topic and/or the audience. As a result, the paper reflects more than one of the following problems:
 - A. The writer is not concerned with the audience. The writer's style is a complete mismatch for the intended reader or the writing is so short that little is accomplished beyond introducing the topic.
 - B. The writer speaks in a kind of monotone that flattens all potential highs or lows of the message.
 - C. The writing is humdrum and risk-free.
 - D. The writing is lifeless or mechanical; depending on the topic, it may be overly technical or jargonistic.
 - E. The development of the topic is so limited that no point of view is present – zip, zero, zilch, nada.

Figure 3 Continued

Word choice

5. Words convey the intended message in a precise, interesting, and natural way. The words are powerful and engaging.
 - A. Words are specific and accurate; it is easy to understand just what the writer means.
 - B. The words and phrases create pictures and linger in your mind.
 - C. The language is natural and never overdone; both words and phrases are individual and effective.
 - D. Striking words and phrases often catch the reader's eye and linger in the reader's mind. (You can recall a handful as you reflect on the paper.)
 - E. Lively verbs energize the writing. Precise nouns and modifiers add depth and specificity.
 - F. Precision is obvious. The writer has taken care to put just the right word or phrase in just the right spot.
3. The language is functional, even if it lacks much energy. It is easy to figure out the writer's meaning on a general level.
 - A. Words are adequate and correct in a general sense; they simply lack much flair and originality.
 - B. Familiar words and phrases communicate, but rarely capture the reader's imagination. Still, the paper may have one or two fine moments.
 - C. Attempts at colorful language show a willingness to stretch and grow, but sometimes it goes too far (thesaurus overload!).
 - D. The writing is marked by passive verbs, everyday nouns and adjectives, and lack of interesting adverbs.
 - E. The words are only occasionally refined; it's more often the first thing that popped into my mind.
 - F. The words and phrases are functional – with only a moment or two of sparkle.
1. The writer struggles with a limited vocabulary, searching for words to convey meaning. The writing reflects more than one of these problems:
 - A. Language is so vague (e.g., it was a fun time, she was neat, it was nice, we did lots of stuff) that only a limited message comes through.
 - B. Blah, blah, blah is all that the reader reads and hears.
 - C. Words are used incorrectly, making the message secondary to the misfires with the words.
 - D. Limited vocabulary and/or frequent misuse of parts of speech impair understanding.
 - E. Jargon or clichés distract or mislead. Persistent redundancy distracts the reader.
 - F. Problems with language leave the reader wondering what the writer is trying to say. The words just don't work in this piece.

Sentence fluency

5. The writing has an easy flow, rhythm and cadence. Sentences are well built, with strong and varied structure that invites expressive oral reading.
 - A. Sentences are constructed in a way that underscores and enhances the meaning.
 - B. Sentences vary in length as well as structure. Fragments, if used, add style. Dialog, if present, sounds natural.
 - C. Purposeful and varied sentence beginnings add variety and energy.
 - D. The use of creative and appropriate connectives between sentences and thoughts shows how each relates to, and builds upon, the one before it.
 - E. The writing has cadence; the writer has thought about the sound of the words as well as the meaning. The first time you read it aloud is a breeze.
3. The text hums along with a steady beat, but tends to be more pleasant or business like than musical, more mechanical than fluid.
 - A. Although sentences may not seem artfully crafted or musical, they get the job done in a routine fashion.
 - B. Sentences are usually constructed correctly; they hang together; they are sound.
 - C. Sentence beginnings are not all alike; some variety is attempted.
 - D. The reader sometimes has to hunt for clues (e.g., connecting words and phrases like however, therefore, naturally, after a while, on the other hand, to be specific, e.g., next, first of all, later, but as it turned out, although, etc.) that show how sentences interrelate.
 - E. Parts of the text invite expressive oral reading; others may be stiff, awkward, choppy, or gangly.
1. The reader has to practice quite a bit in order to give this paper a fair interpretive reading. The writing reflects more than one of the following problems:
 - A. Sentences are choppy, incomplete, rambling or awkward; they need work. Phrasing does not sound natural. The patterns may create a sing-song rhythm, or a chop-chop cadence that lulls the reader to sleep.
 - B. There is little to no sentence sense present. Even if this piece were flawlessly edited, the sentences would not hang together.
 - C. Many sentences begin the same way – and may follow the same patterns (e.g., subject–verb–object) in a monotonous pattern.
 - D. Endless connectives (and, and so, but then, because, and then, etc.) or a complete lack of connectives create a massive jumble of language.
 - E. The text does not invite expressive oral reading.

Figure 3 Continued

Conventions	
<p>5. The writer demonstrates a good grasp of standard writing conventions (e.g., spelling, punctuation, capitalization, grammar, usage, paragraphing) and uses conventions effectively to enhance readability. Errors tend to be so few that just minor touch-ups would get this piece ready to publish.</p> <p>A. Spelling is generally correct, even on more difficult words.</p> <p>B. Punctuation is accurate, even creative, and guides readers through the text.</p> <p>C. Thorough understanding and consistent application of capitalization skills present.</p> <p>D. Grammar and usage are correct and contribute to clarity and style.</p> <p>E. Paragraphing tends to be sound and reinforces the organizational structure.</p> <p>F. The writer may manipulate conventions for stylistic effect and it works! The piece is very close to being ready to publish.</p> <p>Grades 7 and up only: The writing is sufficiently complex to allow the writer to show skill in using a wide range of conventions. For writers at younger ages, the writing shows control over those conventions that are grade/age appropriate.</p> <p>3. The writer shows reasonable control over a limited range of standard writing conventions. Conventions are sometimes handled well and enhance readability; at other times, errors are distracting and impair readability.</p> <p>A. Spelling is usually correct or reasonably phonetic on common words, but more difficult words are problematic.</p>	<p>B. End punctuation is usually correct; internal punctuation (commas, apostrophes, semicolons, dashes, colons, parentheses) is sometimes missing/wrong.</p> <p>C. Most words are capitalized correctly; control over more sophisticated capitalization skills may be spotty.</p> <p>D. Paragraphing attempted; may run together or begin in the wrong places.</p> <p>E. Problems with grammar or usage are not serious enough to distort meaning but may not be correct or accurately applied all of the time.</p> <p>F. Moderate (a little of this, a little of that) editing would be required to polish the text for publication.</p> <p>1. Errors in spelling, punctuation, capitalization, usage and grammar, and/or paragraphing repeatedly distract the reader and make the text difficult to read. The writing reflects more than one of these problems:</p> <p>A. Spelling errors are frequent, even on common words.</p> <p>B. Punctuation is often missing or incorrect.</p> <p>C. Capitalization is random and only the easiest rules show awareness of correct use.</p> <p>D. Errors in grammar or usage are very noticeable, frequent, and affect meaning.</p> <p>E. Paragraphing is missing, irregular, or so frequent (every sentence) that it has no relationship to the organizational structure of the text.</p> <p>F. The reader must read once to decode, then again for meaning. Extensive editing (virtually every line) would be required to polish the text for publication.</p>
Presentation	
<p>5. The form and presentation of the text enhances the ability for the reader to understand and connect with the message. It is pleasing to the eye.</p> <p>A. If handwritten (either cursive or printed), the slant is consistent, letters are clearly formed, spacing is uniform between words, and the text is easy to read.</p> <p>B. If word-processed, there is appropriate use of fonts and font sizes which invites the reader into the text.</p> <p>C. The use of white space on the page (spacing, margins, etc.) allows the intended audience to easily focus on the text and message without distractions. There is just the right amount of balance of white space and text on the page. The formatting suits the purpose for writing.</p> <p>D. The use of a title, side heads, page numbering, bullets, and evidence of correct use of a style sheet (when appropriate) makes it easy for the reader to find the desired information. These markers allow the hierarchy of information to be clear to the reader.</p> <p>E. When appropriate to the purpose and audience, there is effective integration of text and illustrations, charts, graphs, maps, tables, etc. There is clear alignment between the text and visuals. The visuals support and clarify important information or key points made in the text.</p> <p>3. The writer's message is understandable in this format.</p> <p>A. Handwriting is readable, although there may be discrepancies in letter shape and form, slant, and spacing that may make some words or passages easier to read than others.</p> <p>B. Experimentation with fonts and font sizes is successful in some places, but begins to get fussy and cluttered in others. The effect is not consistent throughout the text.</p>	<p>C. While margins may be present, some text may crowd the edges. Consistent spacing is applied, although a different choice may make text more accessible (e.g. single, double, or triple spacing).</p> <p>D. Although some markers are present (titles, numbering, bullets, side heads, etc.) they are not used to their fullest potential as a guide for the reader to access the greatest meaning from the text.</p> <p>E. An attempt is made to integrate visuals and the text although connections may be limited.</p> <p>1. The reader receives a garbled message due to problems relating to the presentation of the text.</p> <p>A. Because the letters are irregularly slanted, formed inconsistently, or incorrectly, and the spacing is unbalanced or not even present, it is very difficult to read and understand the text.</p> <p>B. The writer has gone wild with multiple fonts and font sizes. It is a major distraction to the reader.</p> <p>C. The spacing is random and confusing to the reader. There may be little or no white space on the page.</p> <p>D. Lack of markers (title, page numbering, bullets, side heads, etc.) leave the reader wondering how one section connects to another and why the text is organized in this manner on the page.</p> <p>E. The visuals do not support or further illustrate key ideas presented in the text. They may be misleading, indecipherable, or too complex to be understood.</p>

Figure 3 6 + 1 Trait writing-assessment rubric. Adapted from Northwest Regional Educational Laboratory (1999), with permission from NREL.

Trait 1: Mathematical concepts and procedures

5. I completely understand the appropriate mathematical operations and use them correctly.

- I understand which math operations are needed.
- I have used all of the important information.
- I did all of my calculations correctly.

3. I think I understand most of the mathematical operations and how to use them.

- I know which operations to use for some of the problem, but not for all of it.
- I have an idea about where to start.
- I know what operations I need to use, but I'm not sure where the numbers go.
- I picked out some of the important information, but I might have missed some.
- I did the simple calculations right, but I had trouble with the tougher ones.

1. I wasn't sure which mathematical operation(s) to use or how to use the ones I picked.

- I don't know where to start.
- I'm not sure which information to use.
- I don't know which operations would help me solve the problem.
- I don't think my calculations are correct.

Trait 2: Mathematical problem solving

5. I came up with and used a strategy that really fits and makes it easy to solve this problem.

- I knew what to do to set up and solve this problem.
- I knew what math operations to use.
- I followed through with my strategy from beginning to end.
- The way I worked the problem makes sense and is easy to follow.
- I may have shown more than one way to solve the problem.
- I checked to make sure my solution makes sense in the original problem.

3. I came up with and used a strategy, but it doesn't seem to fit the problem as well as it should.

- I think I know what the problem is about, but I might have a hard time explaining it.
- I arrived at a solution even though I had problems with my strategy at some point.
- My strategy seemed to work at the beginning, but did not work well for the whole problem.
- I checked my solution and it seems to fit the problem.

1. I didn't have a plan that worked.

- I tried several things, but didn't get anywhere.
- I didn't know which strategy to use.
- I didn't know how to begin.
- I didn't check to see if my solution makes sense.
- I'm not sure what the problem asks me to do.
- I'm not sure I have enough information to solve the problem.

Trait 3: Communication in mathematics

5. I clearly explained the process I used and my solution to the problem using numbers, words, pictures, or diagrams.

- My explanation makes sense.
- I used mathematical terms correctly.
- My work shows what I did and what I was thinking while I worked the problem.
- I've explained why my answer makes sense.
- I used pictures, symbols, and/or diagrams when they made my explanation clearer.
- My explanation was clear and organized.
- My explanation includes just the right amount of detail not too much or too little.

3. I explained part of the process I used, or I only explained my answer.

- I explained some of my steps in solving the problem.
- Someone might have to add some information for my explanation to be easy to follow.
- Some of the mathematical terms I use make sense and help in my explanation.
- I explained my answer, but not my thinking.
- My explanation started out well, but bogged down in the middle.
- When I used pictures, symbols, and/or diagrams, they were incomplete or only helped my explanation a little bit.

1. I did not explain my thinking or my answer, or I am confused about how my explanation relates to the problem.

- I'm not sure how much detail I need in order to help someone understand what I did.
- I don't know what to write.
- I can't figure out how to get my ideas in order.
- I'm not sure I used math terms correctly.
- My explanation is mostly copying the original problem.
- The pictures, symbols, and/or diagrams I used would not help someone understand what I did.

Figure 4 Analytical trait rubric for mathematical problem solving. Adapted from Central Kitsap School District (2001), with permission from Julie Goldsmith.

Table 1 Analytical trait rubric for oral presentation**Oral presentation criterion 1: Content****5: Strong**

- My presentation had a clear main topic.
- All of the information in my presentation related to and supported my topic.
- The information I included was important to understanding my topic.
- I chose facts, details, anecdotes, and/or examples to make my topic come alive for my audience.

3: Part-way there

- My topic was fairly broad, but the audience could tell where I was headed.
- Most of my details related to and supported my topic, but some might have been off-topic.
- Some of my information was important, but some details might have been too trivial to be included. Maybe I should have left some details out.
- Some of my information may not have been interesting or useful to my audience.

1: Just beginning

- I was not sure what the focus of my presentation was, or I got mixed up and changed topics during my presentation. I think I wandered through a few topics.
- I did not really know how to choose details to share, so I just used whatever came into my mind.
- I forgot to think about what information might be most interesting or useful to my audience.

Oral presentation criterion 2: Organization**5: Strong**

- The opening of my presentation introduced my topic in a way that caught the audience's interest.
- I chose a sequence for the content of my presentation so that it was easy to follow. My audience could easily make a mental outline of the content.
- I used transition words to guide the audience. I do not think anyone got lost listening to me.
- My conclusion wrapped up my topic and left the audience feeling satisfied.

3: Part-way there

- My presentation had a recognizable opening, but it may have been a little plain.
- Most of my ideas were in an order that was easy to follow, but there may have been a place or two where ideas seemed out of place.
- In some places I may have jumped from one idea to the next without helping the audience follow me.
- I had a conclusion. My audience knew when my presentation was over, but I could have done a better job of leaving them with a feeling of satisfaction.

1: Just beginning

- I just plunged into my ideas without setting the audience up to hear about my topic.
- I was not sure what order to put my ideas in, so they came out in a jumble. I think my audience would have had trouble making a mental outline of the content.
- I left out transitions. I did not help the audience follow along with my thoughts.
- When I finished, the audience did not know it was the end. I forgot to make a closing statement.

Continued

Table 1 Continued**Oral presentation criterion 3: Delivery****5: Strong**

- I maintained eye contact with the audience throughout my speech.
- My voice was loud enough for the audience to hear.
- I varied my voice level and intonation to emphasize meaning.
- I articulated clearly so the audience was able to understand every word.
- I spoke at a pace that kept the audience engaged without racing through my speech.
- I avoided repeatedly using filler words between my ideas (e.g., and, uh, um, you know, like, and well).
- I used gestures and movement to enhance the meaning of my words.
- I knew my speech well enough so that I could just glance at my notes to help me remember what to say.
- If I used visual aids or props, they helped make my meaning clearer.

3: Part-way there

- I made eye contact with my audience part of the time. Or, I only made eye contact with a few people in the audience and I forgot to look at everyone.
- My voice was loud enough for the audience to hear part of the time, but it also was too quiet at times.
- I varied my voice level and intonation a few times to emphasize meaning, but I may have spoken in a monotone part of the time, too.
- I articulated clearly some of the time, but some of the time I mumbled.
- I spoke at a fairly good pace, but there were times when I spoke too quickly.
- Sometimes I used filler words between my ideas (e.g., and, uh, um, you know, like, and well).
- My gestures and movement might have been a little stiff or unnatural, but they did not distract the audience from the meaning of my presentation.
- I gave parts of my presentation without having to read my notes, but had to read them quite a bit in places.
- If I used visual aids, they were understandable, but they may not have added much to my meaning.

1: Just beginning

- I had a hard time making eye contact with my audience. I mostly looked up, away, or down.
- My voice was too quiet for everyone to hear me.
- I may have spoken in a monotone, with no variance in intonation. Or I may have tried to vary my voice level and intonation on certain words, but I was not sure which ones to emphasize.
- I mumbled frequently, so the audience had a hard time understanding.
- I had a hard time with the speed of my talking – I either raced or dragged through my presentation.
- I used a lot of filler words between my ideas (e.g., and, uh, um, you know, like, and well).
- My gestures and movement seemed stiff or unnatural, or I moved around so much it distracted the audience from the meaning of my presentation.
- I had to read my notes for most or all of my presentation.
- If I used visual aids, they were confusing. I wasn't sure how to explain them or how to link them to the ideas I was talking about.

Adapted from Arter, J. A. and Chappuis, J. (2006). *Creating and Recognizing Quality Rubrics*, pp. 244–246, Princeton, NJ: Educational Testing Service, with permission from ETS.

Table 2 Holistic rubric for inferences

Score	Criterion
3	Strong inference: A conclusion based on sufficient evidence to be defensible. Or, if there are problems with the evidence, they are described.
2	Inference: A conclusion based on evidence, but not enough evidence to be really sure of the conclusion. Or, there are unnoted problems with the evidence.
1	Wild guess: A conclusion, not a statement of fact, but it is not based on any evidence or the evidence does not support the conclusion. Not an inference: The answer is a statement of fact or a restatement of the problem or issue, not a conclusion based on evidence.

Adapted from Arter, J. (2008) Educational Testing Service. Rubrics Main Handout.

While task-specific scoring may be faster than general scoring, there are several dangers lurking in its use to evaluate products, performance skills, and reasoning. In the kangaroo writing task (Figure 2), for example, there are many ways that a piece might be good. Is it possible that a student will produce a strong response to the kangaroo prompt and yet not fulfill the requirement to enter into the picture imaginatively using devices such as dialog and character names? Similarly, the scoring guide requires the inference that if the work matches the scoring guide, that is, it enters into the picture imaginatively using devices such as dialog and character names, it in fact indicates good writing. Is it possible that a student might fulfill the requirements in the task-specific scoring guide and still not have produced a strong piece of writing? The issue with these considerations revolves around consistency versus accuracy.

Problems with task-specific scoring of reasoning, performance skills, and products are even more dramatic when the purpose is assessment for learning: teacher use of resulting information to plan instruction, student use of information to determine what they already do well, and what needs to be reviewed, or student use of the scoring guide is to learn the nature of quality in general and bring it to bear across tasks. Exactly what in the kangaroo scoring guide will support these uses? The only portions that might assist in these uses are the general portions that can be generalized to the next writing task: sense of drama, imagination, literary devices, and sound conventions.

Compare the kangaroo scoring guide to the 6+1-trait writing rubric in Figure 3. The systematic use of the latter is what will help students to internalize criteria for quality writing and bring them to bear across writing tasks. This is true because a general rubric attempts to directly define the learning target (in this case writing) – no inferences required. This definition gives students and teachers a vocabulary to talk about quality and makes it easier to generalize across assignments.

The problems with task-specific scoring for writing are easily seen; the parallel problem with other

reasoning, performance skill, and product targets, such as math problem solving is more subtle, but equal. Consider this example of task-specific scoring (Arter and Chappuis, 2006: 18): students were shown a picture of a complicated three-dimensional stack of blocks. (Some of the blocks could not be seen because of the two-dimensional rendering of the three-dimensional stack.) The task asked students to answer three questions:

- Part A: How many blocks are in the stack?
- Part B: How many blocks would be found in a similar stack having four sides and a height of six?
- Part C: Write an equation for determining the number of blocks given any value of number of sides (X) and height (Y).

A specific number was given as the answer to Part A. “The answer in the task-specific scoring guide for Part B was: Each side has 15 blocks; there are four sides, $15 \times 4 = 60$.” The specific generalized equation was given for Part C.

The problem is that it’s impossible to tell, from looking at the scoring guide what this item is supposed to assess. Is it problem solving? Communication in math? Spatial reasoning? Algebraic thinking? Something specific to stacking blocks? With task-specific scoring it is impossible to tell. Yet, the inference is being made that correct performance on the task implies that the student is strong at whatever it is that the item is designed to measure. Compare this task-specific scoring guide to the general math problem solving rubric in Figure 4. If the goal is learning what strong problem solving and communication in math looks like so that it can be brought to bear on the next problem, a general rubric is more powerful than one that is task specific.

Students do not enhance their understanding of, for example, quality problem solving, through statements such as, “Each side has 15 blocks; $15 \times 4 = 60$.” They enhance their understanding of quality problem solving through statements like, “The way I worked the problem makes sense and is easy to follow,” and, “I may have shown more than one way to solve the problem,” coupled with examples across tasks that show what is meant.

Based on these considerations, Table 3 shows recommendations for when to use task-specific and general scoring guides. To distinguish between them, the term rubric only is used when referring to general scoring guides.

Having established that task-specific scoring guides reside in the province of extended written response assessments of bodies of knowledge, the rest of this article addresses general rubrics used to assess reasoning, performance skill, and product-learning goals for students.

Holistic versus Analytical Rubrics

General rubrics come in two formats, as described in Table 3: holistic and analytical. Holistic rubrics result in a single overall judgment of quality. Examples are in Table 2 (holistic rubric for inferences), and Table 4

Table 3 Types of scoring guides and when to use each

Type	Description	When to use	Good example of use	Weak example of use
Task-specific scoring guide	A list of specific facts or information to look for in a single task. This results in either assigning points for the presence of specific facts (Figure 1) or determining a level of quality based on the presence of certain features specific to the task (Figure 2).	<ul style="list-style-type: none"> • Can be used to judge student performance only on a single task. • Best used with knowledge targets to identify the pieces of information students are to exhibit in extended written response assessments. • Sometimes used for reasoning, performance skill, and product-learning goals when the goal is speedy scoring and the use is solely summative. 	Figure 1 Carbon cycle. Extended written response assessment of knowledge.	Figure 2 Kangaroo – quality judgments of writing (a product) require a general rubric, especially if the use is instructional.
General holistic rubric	A general description of levels of quality that results in a single overall judgment for a product or performance.	<ul style="list-style-type: none"> • Can be used with any task that measures the same student learning goal. • Best used with reasoning, performance skill, and product-learning goals when there is only a single dimension of quality. • Sometimes used with any complex learning goal if the purpose is to get a single overall summary of student status. 	Table 2 Inferences. This is a reasoning goal, and there is probably only a single dimension of quality.	Table 4 Holistic rubric for informational writing. For instruction, it would be more useful to determine student strengths and weaknesses using an analytical trait rubric.
General analytical trait rubric	A general description of levels of quality along several different important dimensions of a product or performance that results in a separate judgment for each dimension.	<ul style="list-style-type: none"> • Can be used with any task that measures the same target. • Best used with reasoning, performance skill, and product targets when there are several important, distinguishable dimensions of performance. • Especially useful when the purpose is formative. 	<p>Table 1 General analytical trait rubric for oral presentation (a performance skill). Identifying the important dimensions of performance and evaluating them separately for quality is useful, especially if the use is formative.</p> <p>Figure 3 6+1 Trait writing-assessment rubric.</p> <p>Figure 4 Analytical trait rubric for math problem solving (reasoning and communication in math).</p>	

(holistic rubric for writing). Analytical rubrics describe levels of quality along several different important dimensions of a product or performance; this results in separate judgments for each dimension. Examples are in **Table 1** (general analytical trait rubric for oral presentation), **Figure 3** (6+1-trait writing-assessment rubric), and **Figure 4** (analytical trait rubric for math problem solving).

Purpose and the complexity of the learning goal to be assessed determine which is best. Holistic rubrics are best used when the reasoning proficiency, performance skill, or product being assessed is relatively simple and only has a single dimension of quality, as is the case for inferences (**Table 2**). Holistic rubrics are also used with more complex learning goals when the purpose is to make a single overall judgment of current student status – assessment of learning. When the purpose is formative, an analytical rubric for multidimensional learning goals is better because it allows the user to generate a profile of student strengths and weaknesses. A quote from Graves (1983) explains why:

One of the best examples of good teaching I have ever encountered was with a golf professional. On my first lesson, he said, “Here is a bucket of balls . . . hit ‘em.’” A few minutes later he wandered back and quietly said, “Keep hitting them, only this time keep your head down, eye on the ball.” By the next bucket of balls he had introduced one more skill for the day . . . no more. Before a few weeks were out, he had quietly attended to my feet,

grip, shoulder level, and follow through. A few years later I realized with a start that every single one of my problems was visible on the first lesson. If I had attended to all of them that first day, I would probably have missed the ball entirely and resigned in disgust from ever playing golf again (pp 314–315).

During learning it is useful to break a complex performance into manageable parts and practice each separately.

Characteristics of High-Quality Rubrics

To serve their purposes of consistent scoring and defining complex learning outcomes, rubrics need to have certain features, captured in the rubric for rubrics (**Table 5**; Arter and Chappuis, 2006).

For easy reference, we call the two major dimensions of quality in the rubric for rubrics criteria or traits, the lettered subheads under each criterion indicators, and the numbered phrases under each indicator descriptors.

The descriptors under each indicator are not meant to function as a checklist. Not everything has to be present (or missing) for the rubric under consideration to be judged to be at a particular level of quality. Rather, look for the constellation of indicators that best describe the rubric under consideration.

Table 4 Holistic rubric for writing

<i>Excellent</i>	<i>Good</i>	<i>Needs work</i>
1. The paper is clear and focused. It holds the reader's attention. Relevant details enrich the central theme.	1. The writer is beginning to define the topic, even though development is still basic or general.	1. As yet, the paper has no clear sense of purpose or central theme. The reader must make inferences from sketchy or missing details.
2. The organization enhances and showcases the central idea. The order, structure, or presentation of information is compelling and moves the reader through the text.	2. The organizational structure is strong enough to move the reader through the text without too much confusion.	2. The writing lacks a clear sense of direction. Ideas, details, or events seem strung together in a loose or random fashion; there is no identifiable internal structure.
3. The writer speaks directly to the reader in a way that is individual, compelling, and engaging.	3. The writer seems sincere but not fully engaged or involved; the result is pleasant or even personable, but not compelling.	3. The writer seems indifferent, uninvolved, or distanced from the topic.
4. Words convey the intended message in a precise, interesting, and natural way. The words are powerful and engaging. Technical words, when used, are used correctly.	4. The language is functional, even if it lacks much energy. It is easy to figure out the writer's meaning on a general level. Technical words, when used, are sometimes used incorrectly.	4. The writer demonstrates a limited vocabulary or has used vocabulary incorrectly.
5. The writing has an easy flow, rhythm, and cadence. Sentences are well built, with strong and varied structure.	5. The text hums along with a steady beat, but tends to be more pleasant than fluid.	5. Sentences are choppy, incomplete, rambling, or awkward.
6. The writer demonstrates few errors in conventions and may even have used conventions to enhance meaning.	6. Sometimes conventions are handled well and enhance readability; at other times, errors are distracting and impair readability.	6. Errors in spelling, punctuation, capitalization, usage, and grammar repeatedly distract the reader and make the text difficult to read.

Adapted from *6+1 Trait Writing Assessment Scoring Guide*, Northwest Regional Educational Laboratory (1999).

Table 5 Rubric for rubrics

5 – Strong	3 – Medium	1 – Weak
Criterion 1: Coverage/organization		
A. Covers the right content		
1. The content of the rubric represents the best thinking in the field about what it means to perform well on the skill or product under consideration.	1. Much of the content represents the best thinking in the field, but there are a few places that are questionable.	1. You cannot tell what learning target(s) the rubric is intended to assess, or you can guess at the learning targets, but they do not seem important, or content is far removed from current best thinking in the field about what it means to perform well on the skill or product under consideration.
2. The content of the rubric aligns directly with the content standards/learning targets it is intended to assess.	2. Some features do not align well with the content standards/learning targets they are intended to assess.	2. The rubric does not seem to align with the content standards/learning targets it is intended to assess.
3. The content has the ring of truth – your experience as a teacher confirms that the content is truly what you do look for when you evaluate the quality of a student product or performance. In fact, the rubric is insightful; it helps you organize your own thinking about what it means to perform well.	3. Much of the content is relevant, but you can easily think of some important things that have been left out or that have been given short shrift, or it contains an irrelevant criterion or descriptor that might lead to an incorrect conclusion about the quality of student performance.	3. You can think of many important dimensions of a quality performance or product that are not in the rubric, or content focuses on irrelevant features. You find yourself asking, Why assess this? or Why should this count? or Why should students have to do it this way?
B. Criteria are well organized		
1. The rubric is divided into easily understandable criteria as needed. The number of criteria reflects the complexity of the learning target. If a holistic rubric is used, it is because a single criterion adequately describes performance.	1. The number of criteria needs to be adjusted a little: either a single criterion should be made into two criteria, or two criteria should be combined.	1. The rubric is holistic when an analytic one is better suited to the intended use or learning targets to be assessed – or – the rubric is an endless list of everything; there is no organization; the rubric looks like a brainstormed list.
2. The details that are used to describe a criterion go together; you can see how they are facets of the same criterion.	2. Some details that are used to describe a criterion are in the wrong criterion, but most are placed correctly.	2. The rubric seems mixed up – descriptors that go together do not seem to be placed together. Things that are different are put together.
3. The relative emphasis on various features of performance is right – things that are more important are stressed more; things that are less important are stressed less.	3. The emphasis on some criteria or descriptors is either too small or too great; others are all right.	3. The rubric is out of balance – features of more importance are emphasized the same as features of less importance.
4. The criteria are independent. Each important feature that contributes to quality work appears in only one place in the rubric.	4. Although there are instances when the same feature is included in more than one criterion, the criteria structure holds up pretty well.	4. Descriptors of quality work are represented redundantly in more than one criterion to the extent that the criteria are really not covering different things.
C. Number of levels fits targets and uses		
1. The number of levels of quality used in the rating scale makes sense. There are enough levels to be able to show student progress, but not so many levels that it is impossible to distinguish among them.	1. Teachers might find it useful to create more levels to make finer distinctions or to merge levels to suit the rubric's intended use. The number of levels could be attuned easily.	1. The number of levels is not appropriate for the learning target being assessed or the intended use. There are so many levels it would be impossible to readily distinguish among them, or too few to make important distinctions. It would take major work to fix the problems.

Continued

Table 5 Continued**Criterion 2: Clarity****A. Levels defined well**

1. Each score point (level) is defined with indicators and descriptors. A plus: There are examples of student work that illustrate each level of each trait.	1. Only the top level is defined. The other levels are not defined.	1. No levels are defined; the rubric is little more than a list of categories to rate followed by a rating scale.
2. There is enough descriptive detail in the form of concrete indicators, adjectives, and descriptive phrases that allow you to match a student performance to the right score. A plus: If students are to use the rubric, there are student-friendly versions, and/or versions in foreign languages for English language learner (ELL) students. A plus: There is information on rater agreement rates that shows that raters can exactly agree on a score 65% of the time, and within one point 98% of the time.	2. There is some attempt to define terms and include descriptors, but some key ideas are fuzzy in meaning.	2. Wording of the levels, if present, is vague/confusing. You find yourself saying such things as, "I'm confused," or "I don't have any idea what this means." Or, the only way to distinguish levels is with words such as extremely, very, some, little, and none; or completely, fairly well, little, and not at all.
3. If counting the number or frequency of something is included as an indicator, changes in such counts really are indicators of changes in quality.	3. There is some descriptive detail in the form of words, adjectives, and descriptive phrases, but counting the frequency of something or vague quantitative words are also present.	3. Rating is almost totally based on counting the number or frequency of something, even though quality is more important than quantity.
4. Wording is descriptive not evaluative.	4. Wording is mostly descriptive of the work, but there are a few instances of evaluative labels.	4. Wording tends to be evaluative rather than descriptive of the work; e.g., work is mediocre, above average, or clever.

B. Levels parallel

1. The levels of the rubric are parallel in content – if an indicator of quality is discussed in one level, it is discussed in all levels. If the levels are not parallel, there is a good explanation why.	1. The levels are mostly parallel in content, but there are some places where there is an indicator at one level that is not present at the other levels.	1. Levels are not parallel in content and there is no explanation of why, or the explanation does not make sense.
---	---	---

Adapted from Arter, J. A. and Chappuis, J. (2006). *Creating and Recognizing Quality Rubrics*, pp. 183–186, Princeton, NJ: Educational Testing Service, with permission from ETS.

An odd number of levels is used because the middle level represents a balance of strengths and weaknesses. A strong judgment does not necessarily mean that the classroom rubric under consideration is perfect; rather, it means that it would require very little work to get it ready for use. A middle judgment means that some work would be required to make the rubric under consideration usable, but it probably is worth the effort. A weak judgment means that the rubric under consideration needs so much work that it probably is not worth the effort – it is time to find another one or begin from scratch.

In addition, a medium score does not mean average. This is a criterion-referenced scale, not a norm-referenced one. It is meant to describe levels of quality in rubrics, not to compare those currently available. It could be that the typical currently available rubric is closer to weak than to medium.

Although three levels are defined, it is in fact a five-level scale. Think of level 4 as a combination of characteristics

from levels 5 and 3. Likewise, level 2 combines characteristics from levels 3 and 1.

Using the Rubric for Rubrics to Analyze Sample Rubrics for Quality

The following analyses serve two purposes. First, they illustrate the criteria, indicators, and descriptors in the rubric for rubrics. Second, they highlight current issues and considerations about rubrics.

The 6+1 Trait Writing Assessment rubric is strong in all the areas described by the rubric for rubrics. It covers the features of student writing generally thought to be important, it is organized well, the number of levels is appropriate, it attempts to be descriptive and tries to avoid use of vague relative-quantity words, and the levels are strictly parallel. It helps teachers plan instruction, diagnose student strengths and weaknesses, identify students requiring additional assistance, and plan remediation.

It sends a strong and accurate message to students about what high-quality writing looks like.

To illustrate common problems with rubrics, **Table 6** (research report rubric) and **Table 7** (book report rubric) are used.

Rubric for rubrics criterion (trait) 1: Coverage/organization

Covers the right content

While many of the features in the book report rubric in **Table 7** represent important learning outcomes for students, there are some features that do not quite ring true or represent best thinking about what students need to know and be able to do. For example, why does the title page need an illustration (feature 9)? Is it possible to have a good book report without an illustration? Or why is there a requirement for typed, 12-point font, etc.? While this might make it easier for the teacher, it is not an indicator of quality in general. Format is represented by two features: feature 2 (types, font size, length, line spacing) and feature 9 (title page plus illustration). Are there other features of format that are more important for the student to learn?

The book report rubric illustrates the bigger issue of confusion about purposes of rubrics. Is the goal to enable students to get a high grade on an assignment, task, or project, or is the goal to define important complex learning goals for students? If the goal is just to enable students to score well, anything can be put on a rubric – length, type

face, illustrations, and so forth. If the goal is to define important learning outcomes, then not anything goes. In fact, why have a rubric for a book report anyway? Is not the book report just the context for teaching and assessing learning goals such as (1) identifying and summarizing plots, and (2) writing? It would be better to take the requirements for the assignment (length, cover, etc.) out of the rubric and focus instead on defining the complex learning goals that should be the real target of the assignment.

In this regard it is pertinent to ask what message is sent to students by the rubric about what the real learning goals of the work are and where they should focus effort and thought.

The tension between enabling students to score well and defining the important characteristics of a learning outcome can arise when students are involved in developing the criteria by which their work will be judged. Sometimes students include irrelevant criteria. We have to be ready and able to guide students to the important dimensions of work; not anything goes.

Criteria (traits) are well-organized

The book report rubric in **Table 7** could be better organized. Instead of 15 features, might the rubric better communicate what is important if the 15 features were organized into three traits: ideas/content (features 4, 5, 6, 7, 8, and 10; format (features 1, 2, and 9); and conventions (features 11–15). Feature 3, meets the requirements of the task, might be left out entirely because it is repetitive – all the other features specify the requirements of the task. The reorganization might serve to direct thinking about balance. Are ideas/content more important than format and conventions or should they all be weighted the same? The decision might go either way, but it appeared balance was not a consideration at all in the original. The reorganization might also serve to direct thinking about the content of the rubric – the extent to which it represents important learning targets for students.

Number of levels fits targets and purposes

Consider the research report rubric in **Table 6**. The trait of content/organization is worth 50 points. What is the difference between, say, 40 and 41? It is unlikely that teachers could independently be able to assign the same number of points or that students would understand what it takes to earn that extra point. Such rubrics may be used by teachers to make it easier to assign a grade, but they do not promote accuracy or assist with formative interpretations.

Other current issues and considerations about number of levels are these:

1. *Odd or even number of levels.* For formative uses of rubrics, a better consideration is the number of levels that can be usefully distinguished, and how each of the levels is defined. For example, in the rubric for rubrics a 3 is

Table 6 Research report rubric – too many levels

Score	Criterion
50 points	Content/organization: A theme is established and developed with the most pertinent details; content is accurate; technical words, when used, are used correctly; information is organized to easily lead the reader through the ideas; pacing is good – the author takes his/her time on important or confusing points, and speeds up when detail is not necessary; there is an inviting opening; and the ending sums up the ideas.
10 points	Style and voice: The style shows that the writer is interested in the topic and wants the reader to be as well. Sentences are varied in beginnings and lengths and enhance the flow of the work.
20 points	Format: The writer has used the format usual in this type of writing.
10 points	Conventions: Spelling, grammar, punctuation, and paragraphing show very few errors, none of which interfere with the meaning or the style of the report. In fact, conventions are sometimes used purposefully to enhance the meaning of the text.
10 points	References: At least 10 references that are relevant and comprehensive.
_____/100	Total

Table 7 Book report rubric

	<i>Exceeds standard (5 points for each)</i>	<i>Meets standard (4 points for each)</i>	<i>Approaches standard (2 points for each)</i>	<i>Below standard (0 points for each)</i>
1.	Includes the following information: title of the book, author, copyright, publisher, and date	Includes most of the following information: title of the book, author, copyright, publisher, and date	Includes some of the following information: title of the book, author, copyright, publisher, and date	Missing most or all of the following information: title of the book, author, copyright, publisher, and date
2.	Typed, 12 point Times Roman font; two pages; 1.5 spaced	Typed, 12 point Times Roman font; a page and a half; 1.5 spaced	Not typed; one page	Not typed, less than one page
3.	Meets the requirements of the assignment	Meets most of the requirements of the assignment	Meets some of the requirements of the assignment	Does not meet the requirements of the assignment
4.	Information shared shows insightful understanding of the book	Information shared shows adequate understanding of the book	Information shared shows partial understanding of the book	Information shared does not show understanding of the book
5.	Conclusion clearly states your opinion about the book	Conclusion states your opinion about the book	Conclusion is present, but does not state your opinion about the book	Conclusion is missing
6.	Main characters are thoroughly described	Main characters are described	Some main characters are described	Main characters are not described
7.	Setting is clearly identified	Setting is identified	Setting is incompletely identified	Setting is not identified
8.	Plot is summarized completely	Plot is summarized adequately	Plot is partially summarized	Plot is not summarized
9.	Title page plus illustration	Title page plus illustration	Title page plus illustration	Title page plus illustration
10.	Timeline includes 10 or more important events	Timeline includes 7–9 important events	Timeline includes 4–6 important events	Timeline includes three or fewer important events
11.	All spelling is correct	3–5 spelling mistakes	6–10 spelling mistakes	More than 10 spelling mistakes
12.	All punctuation is correct	1–2 punctuation mistakes	3–5 punctuation mistakes	More than 6 punctuation mistakes
13.	All capitalization is correct	1–2 capitalization mistakes	3–4 capitalization mistakes	More than 5 capitalization mistakes
14.	Correct grammar	One grammar mistake	2–3 grammar mistakes	More than 3 grammar mistakes
15.	Correct subject/verb agreement	One subject/verb agreement problem	Two subject/verb agreement problems	Three or more subject/verb agreement problems

Adapted from Arter, J. A. and Chappuis, J. (2006). *Creating and Recognizing Quality Rubrics*, p. 53, Princeton, NJ: Educational Testing Service, with permission from ETS.

defined as a balance of strengths of weaknesses. This is less likely to make a middle score a dumping ground for indecision.

2. *Four levels: exceeds standard, meets standard, almost meets standard, and below standard.* The argument here is that such rating scales make standards-based evaluations and summaries easier. While this might be true for summative uses of rubrics, there are dangers when using rubrics formatively. There is a considerable body of research on the type of feedback to students most likely to produce learning. In general, descriptive feedback is better than evaluative feedback, such as grades. Statements such as below standard are also evaluative. Placing value judgments on how much is good enough too early in the learning process can be counterproductive.

Rubric for rubrics criterion (trait) 2: Clarity **Levels defined well**

One consistent problem with rubrics is the use of vague relative-quantity words like extremely, very, some, and

not much, or almost always, frequently, sometimes, and never. Examples of this are in the book report rubric in **Table 7**. For example, consider the seventh indicator, setting is clearly identified. What is the difference between clearly identified, identified, incompletely identified, and not identified? Even if adults could consistently rate level of identification, would students understand what it takes to clearly identify a setting?

Descriptive detail in a rubric is the opposite of vague quantity words. For example, consider the trait of ideas in the 6+1-trait writing-assessment rubric (**Figure 3**). This rubric does not say that the topic is very clear and focused, somewhat clear and focused, marginally clear and focused, and not focused at all. Rather, it tries to describe what makes writing clear and focused by using phrases like the topic is narrow and manageable, the topic is fairly broad; however, the reader can see where the writer is headed, and the writer is still in search or a topic, is brainstorming, or has not yet decided what the main idea of the piece will be. If rubrics are to

communicate, descriptive words and phrases are necessary. Incorporating the most telling words and phrases into a rubric is a continual process of refinement as users come to more clearly understand what actually does make a performance or product good.

Another common problem relating to the clarity of rubrics is the use of counts to define levels. If one goal of a rubric is to make an essentially subjective judging process as objective as possible, it would seem that counting something would help. The problem with this is that quantity is not always an indicator of quality. For example, it is submitted that three highly relevant and comprehensive references in a research paper are better than ten marginal ones; or in the case of the book report rubric in **Table 7**, under indicator 11, perhaps six spelling mistakes on different words are more of a problem than six spelling errors all on the same word. An important issue is the message sent to students about the nature of quality when counts are used.

Levels parallel

In general, it is desirable to have each level of a rubric describe the same feature. For example, if the quality of the conclusion for a laboratory experiment is described at one level of quality, it would be useful to describe it all levels. All the rubrics illustrated herein are strictly parallel – they are either lined up across the page, or features are numbered or lettered.

There may be cases when levels do not need to be strictly parallel. For example, if the levels of a rubric are thought of as stages in development in becoming competent in some dimension of performance, perhaps these stages are well indicated by different aspects of performance.

Conclusion

Scoring rubrics can be powerful tools for assessment and learning if designed with purpose and the type of learning goal to be assessed in mind. This article has argued for the use of general, analytic rubrics when the use is assessment for learning.

Good assessment-for-learning strategies and examples using rubrics can be found in Stiggins *et al.* (2006, ch. 7); Arter and Chappuis (2006, ch. 6); Arter and McTighe (2001, ch. 6); Spandel (2005); and Kimpton and Harnisch (2005).

Bibliography

Arter, J. A. and Chappuis, J. (2006). *Creating and Recognizing Quality Rubrics*. Portland, OR: Educational Testing Service.

- Arter, J. A. and McTighe, J. (2001). *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*. Thousand Oaks, CA: Corwin Press.
- Black, P. (2003). The nature and value of formative assessment for learning. *Paper Presented at the American Educational Research Association annual meeting*, Chicago.
- Black, P., Harrison, C., Lee, C., Marshall, B., and William, D. (2002). *Working Inside the Black Box: Assessment for Learning in the Classroom*. London: King's College Press.
- Black, P. and William, D. (1998). Assessment and classroom learning. *Educational Assessment: Principles, Policy and Practice* 5(1), 7–74. (Also summarized in inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.)
- Brookhart, S. M. (2005). Research on formative classroom assessment. *Paper Presented at the American Educational Research Association annual meeting*, Montreal.
- Dochy, F., Segers, M., and Sluijsmans, D. (1999). The use of self-peer and co-assessment in high education: A review. *Studies in Higher Education* 24(3), 331–350.
- Graves, D. (1983). *Writing: Teachers and Children at Work*. Portsmouth, NH: Heinemann.
- Guskey, T. R. (2005). Formative classroom assessment and Benjamin S. Bloom: Theory, research, and implications. *Paper Presented at the Annual Meeting of the American Educational Research Association*, Montreal, Canada.
- Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of Educational Research* 77(1), 81–112.
- Kimpton, P. and Harnisch, D. (2005). *Scale Your Way to Musical Assessment: The Ultimate Guide to Creating a Quality Music Program*. Downers Grove, IL: Music Performance Assessments Experts.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* 18(2), 119–144.
- Shepard, L. A. (2001). The role of assessment in a learning culture. *Educational Researcher* 29(7), 4–14.
- Spandel, V. (2005). *Creating Writers through 6-Trait Writing Assessment and Instruction*. San Francisco: Pearson.
- Stiggins, R. J., Arter, J. A., Chappuis, J., and Chappuis, S. (2006). *Classroom Assessment for Student Learning: Doing it Right – Using it Well*. Portland, OR: Assessment Training Institute.

Further Reading

- Loacker, G. (ed.) (2000). *Self Assessment at Alverno College*. Milwaukee, WI: Alverno College Institute.
- Stiggins, R. J. (2008). *An Introduction to Student-Involved Assessment FOR Learning*, 5th edn. Upper Saddle River, NJ: Merrill/Prentice-Hall.

Relevant Websites

- <http://www.publications.gov.bc.ca> – British Columbia Performance Standards, Government Publication Services.
- <http://www.exemplars.com> – Exemplars Assessment Rubrics.
- <http://www.webs1.uidaho.edu> – Idaho State Engineering Department.
- <http://www.isbe.state.il.us> – Illinois Department of Education.
- <http://www.ncamichigan.org> – North Central Association of Colleges and Schools.
- <http://nwrel.org> – Northwest Regional Educational Laboratory.
- <http://www.edu.gov.on.ca> – Ontario Provincial Ministry of Education.
- <http://www.ncsmentor.com> – Pearson Mentor.

Second Language Assessment

L F Bachman, University of California, Los Angeles, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Second-language testing is concerned with measurement and ethical issues that arise in settings where individuals' proficiency in a language that is not their mother tongue or first language (L1) is being assessed, or where their knowledge and skills are being assessed in a language that is not their L1. Examples of the former include assessments of the second- or foreign-language (L2) proficiency of English-language learners (ELLs) in schools, international students seeking admission to universities in which the language of instruction is for them an L2, individuals seeking professional certification for or employment in positions that require proficiency in an L2, or individuals seeking to emigrate to a country where the dominant language is for them an L2. In such settings, where we need to make inferences about individuals' L2 proficiency, language is both the instrument and object of the assessment. That is, we present individuals with assessment tasks that engage them in L2 use (instrument) so that we can interpret their performance as an indication of their L2 proficiency (object).

Examples of settings in which individuals' knowledge and skills in various domains are assessed in an L2 include assessments of the academic achievement and aptitude of students whose L1 is not the language of instruction, the professional competency of individuals whose L1 is not the language of their intended professional practice or employment, and individuals who intend to emigrate to a country where the dominant language is for them an L2. In such settings, L2 proficiency may be either an ancillary skill that facilitates individuals' demonstration of their knowledge or skills in another area, such as science, mathematics, healthcare, law, welding, or plumbing, or considered part of the construct to be assessed.

The primary measurement issue that arises in both theory and in practice in second-language assessment is understanding and defining the construct to be assessed and understanding the nature of assessment tasks and the interaction between test-takers and these tasks. Ethical issues that arise from these settings include the consequences of and responsibility for assessment use.

Language testing as a field draws on information and knowledge in both applied linguistics and educational measurement. Relevant knowledge and information from applied linguistics include the nature of language as a formal system of signs, for example, sounds, utterances,

letters, words, texts, and the cognitive and social correlates of these signs which give them meaning in situated discourse. Relevant knowledge and information from educational measurement include the nature of theoretical and mathematical models of measurement, and of the practical argumentation and reasoning and empirical methods for supporting the validity of interpretations and the appropriateness and consequences of the uses of assessments.

Measurement Issue: The Construct to be Assessed

Historical Perspective

In the past 50 years, L2 language ability or proficiency has been defined in a number of different ways.

Skills and elements

The first explicitly defined model of language ability for language testing was the skills-and-elements approach that was articulated by Lado (1961) and Carroll (1961, 1968). In their formulations, there was a clear distinction between skills and abilities, on the one hand, and approaches/methods/test types on the other. The skills aspect of this model comprised auditory comprehension, oral production, reading, and writing, while the elements comprised pronunciation, grammatical structure, vocabulary, and cultural meanings.

The skills-and-elements approach was the first in the history of language-testing to explicitly draw upon both current linguistic theory and views of language learning and teaching and psychometrics. This approach was extremely influential, and informed a generation of large-scale assessments of foreign or second language in the United States.

Language ability as unitary

In contrast to the skills-and-elements approach, other language testers have viewed language proficiency as a unitary entity. However, two very distinct views of the nature of that entity have emerged: one as a trait or ability that individuals have, and one as what individuals can do (Upshur, 1979).

Language ability as pragmatic expectancy grammar

Drawing on research into the nature of intelligence, particularly Spearman's (1904) theory of general intelligence,

Oller (1979) proposed what he called the “unitary trait hypothesis”, which stated that language proficiency is essentially a single unitary ability, rather than separate skills and elements. In the most extensive discussion of this research and the theory that underlay it, Oller (1979) identified the general factor from his empirical research as “pragmatic expectancy grammar”, which he defined as “the psychologically real system that governs the use of a language in an individual who knows that language” (Oller, 1979: 6). Oller argued that this ability could best be assessed by integrative assessments, such as the dictation, cloze, oral interview, and written composition.

Language ability as performance: Direct testing/performance assessment

Another view of the construct to be tested emerged in the 1970s, particularly in North America, as a result of an intense interest, in both research and practice, in the testing of oral proficiency. Proponents of this view, who called it “direct testing” (Clark, 1972, 1979) or “performance assessment” (Jones, 1979, 1985; Wesche, 1987), argued that it constituted a major sea change from the skills-and-elements approach. The published discussions of direct testing/performance assessment tended to focus primarily on the nature of the test tasks, which were claimed to mirror, or approximate real-life language use outside of the test itself (Clark, 1975). In defining the construct to be tested, proponents of direct testing/performance assessment referred to real-life language performance which they viewed as the criterion for test tasks (Clark, 1972, 1979). Performance on a direct test was seen essentially as a predictor of the language performance that could be expected of the test-taker in real-life settings.

The most influential practical realization of this approach in practice is the Oral Proficiency Interview (OPI) that originated in the Foreign Service Institute of the US Department of State and that has since been adapted for use by other agencies in the US government, as well as for assessing oral proficiency in foreign- and second-language programs in academic settings (American Council on the Teaching of Foreign Languages, 1983) and in other countries (Wylie and Ingram, 2006). While this approach to assessing oral language proficiency has been criticized from a variety of perspectives in applied linguistics, (Bachman and Savignon, 1986; Kramsch, 1986; Lantolf and Frawley, 1985; Young and He, 1998), the OPI continues to be widely used.

Both Oller’s unitary trait view and the OPI were debated heatedly among language testers and other applied linguists in the 1980’s. Although the former debate drew to a close as more and more research supported a multicomponential view of language ability (Bachman and Palmer, 1982; Carroll, 1983; Oller, 1983; Vollmer and Sang, 1983), concerns with the OPI continue

to the present (O’Connell and Norwood, 2007). The view of language ability as performance reemerged in the 1990s under the label of “task-based performance assessment” (Brown *et al.*, 2002; Norris *et al.*, 1998; Skehan, 2001). This approach draws on research from second-language acquisition and language pedagogy (Skehan, 1996, 1998). Nevertheless, proponents of task-based assessment share with the direct testers the views that the construct to be assessed performance on language-assessment tasks that replicate is real-life tasks, and that the primary use of this performance is to predict performance on future real-life language use tasks.

Language ability as multicomponential, communicative competence

At about the time the debate surrounding the unitary trait hypothesis was drawing to a close, applied linguists in the North America and the UK began exploring a much broader view of language ability, drawing on a wide range of research in functional linguistics, sociolinguistics, discourse analysis, psycholinguistics, and language acquisition as well as developments in communicative syllabus design and communicative language teaching.

One of the earliest, and still most influential, discussions of this broadened view of language ability for language testing was by Canale and Swain (1980). In their article, they adopted the term ‘communicative competence’ to describe the ability that is of interest in both language teaching and testing. They defined communicative competence as “the relationship and interaction between grammatical competence, or knowledge of the rules of grammar, and sociolinguistic competence, or knowledge of the rules of language use” (Canale and Swain, 1980: 6). In addition to grammatical and sociolinguistic competence, Canale and Swain posited a third component which they called strategic competence, and defined it as “verbal and non-verbal communication strategies that may be called into action to compensate for breakdowns in communication due to performance variables or insufficient competence” (Canale and Swain, 1980: 30). Canale and Swain’s view of the construct to be measured was much richer than those that preceded, and initiated a major shift in the way language testers viewed the construct.

Canale and Swain’s paper opened a new avenue of research and debate in language testing, leading to several independently convened symposia being held in the UK (Alderson and Hughes, 1981; Hughes and Porter, 1983) and the US (Palmer *et al.*, 1981; Rivera, 1984) to address issues that had been raised by them in their paper, along with the research that supported a multicomponential view of language proficiency.

Building on Canale and Swain’s work, Bachman (1990) conceived of performance on language tests as a function

of the interaction between an individual's language ability and the characteristics of the test method (Bachman, 1990: 113). To address these issues in both the design and development of language tests and the interpretation and use of assessment results, Bachman proposed two frameworks: (1) communicative language ability and (2) test method facets. Communicative language ability was essentially an extension of the Canale and Swain model, in which their notion of strategic competence was expanded from one that functioned essentially in accommodation and compensation to one that he hypothesized underlies all language use. Bachman saw test method facets as "analogous to the features that characterize the context of situation, or the speech event, as this has been described by linguists" (p. 111), and argued that these could "be seen as restricted or controlled versions of these contextual features that determine the nature of language performance that is expected for a given test or test task" (p. 112).

Bachman's two frameworks were subsequently incorporated into an approach to practical test development by Bachman and Palmer (1996), who renamed 'communicative language ability' simply 'language ability' and 'test method facets task characteristics'. Bachman and Palmer argued that in order for score-based interpretations to generalize beyond the test itself, the characteristics of the assessment tasks needed to correspond to the characteristics of tasks in test-takers' target language use (TLU) domains. They argued that by analyzing the characteristics of tasks in the TLU domain, test developers could use these sets of characteristics as templates for generating assessment tasks that would be representative of tasks in the TLU domain. The framework of task characteristics was thus seen as a way to solve the sampling problem of performance assessment and to thus provide a stronger basis for making inferences to domains beyond the test itself.

Bachman's and Bachman and Palmer's frameworks provided richer descriptions of both the construct and assessment tasks than previous approaches to language testing, but they were not without limitations. Even though their approach views language use in terms of interactions between ability and context as discourse is co-constructed, their frameworks are essentially descriptive and provide little explanation of how ability and context interact with each other in language use. Thus, while this approach may provide practical guidance for the design, development, and use of language tests, it does not solve the issue of how abilities and contexts interact in test-takers' performance.

Current Issues in Defining the Construct

Currently, the dominant view in the field is that language ability consists of a number of interrelated areas, such as

grammatical knowledge, textual knowledge, and pragmatic knowledge, and that these areas of language knowledge are managed by a set of metacognitive strategies that also determine how language ability is realized in language use or the situated negotiation of meaning (Bachman, 1990; Bachman and Palmer, 1996; Chapelle, 1998; Chapelle, 2006). Of all these areas, grammatical and textual knowledge have been the most extensively researched and developed in language testing (Purpura, 2004). The area that has been researched the least, pragmatic knowledge, has recently begun to be investigated extensively. Researchers working essentially within Bachman's (1990) definition of pragmatic competence as knowledge of "the relationships between utterances and the acts or functions that speakers intend to perform through these utterances" (p. 89) and drawing on cross-cultural research in language use have developed prototype assessments for use as instruments in applied linguistics research as well as operational measures of this construct (Hudson *et al.*, 1992, 1995; Roever, 2006; Yamashita, 1996).

Other researchers have challenged this view on a number of counts. Working largely within the area of assessing interactive speaking and drawing on a variety of research literatures outside of language assessment, these researchers have identified a number of problems and lacunae in current conceptualizations of the construct, oral language ability, and how we go about assessing it. These researchers who focus on the nature of the interactions in oral face-to-face language use have argued that the view of language ability solely as a cognitive attribute of language users ignores the essentially social nature of the interactions that take place in discourse. They argue that language ability resides in the contextualized interactions or discursive practices that characterize language use (Chalhoub-Deville, 1995, 2003; Chalhoub-Deville and Deville, 2005; McNamara, 2003; McNamara, 1997; Young, 2000).

In a critical review of this debate Bachman (2007) identifies three different approaches to defining language ability: (1) ability-focused, (2) task-focused, and (3) interaction-focused. He concludes that the theoretical issues raised by these different approaches to defining the construct, language ability, present challenging questions for both empirical research in language testing and practical test design, development, and use. For language testing research, these questions imply the need for a much broader methodological approach, involving both so-called quantitative and qualitative perspectives. For language testing practice, they imply that focus on ability, task, or interaction, to the exclusion of the others, will lead to weaknesses in the assessment itself or to limitations on the uses for which the assessment is appropriate.

A closely related issue is the extent to which language ability includes content knowledge. The effect of test-takers' content knowledge on language test performance

is well documented in the language assessment literature (Alderson and Urquhart, 1985; Clapham, 1996; Douglas and Selinker, 1993; Pappajohn, 1999) and the dominant view has been that this is a source of bias in language tests. (Interestingly, in the research on the assessment of educational achievement, it is the reverse: content knowledge is considered the construct to be assessed, while language ability is typically viewed as a source of bias). An alternative view has been articulated in the area of languages for specific purposes assessment. According to this view, what we want to assess 'is specific purpose language ability' (Douglas, 2000) which is a combination of language ability and background knowledge. While a number of assessments of language for specific purposes have been developed and are being used (e.g., by Australian Federal Government, European Organisation for the Safety of Air Navigation, University of Cambridge ESOL Examinations), there is still considerable debate in the field concerning the theoretical basis for such assessments (Hamp-Lyons and Lumley, 2001).

Recently, the issue of the relationship between language ability and content knowledge in test performance has been the focus of considerable research with respect to assessing the academic achievement of ELLs. On the assumption that the construct to be assessed is knowledge and skills in some specific area such as mathematics or science, a widely held view is that ELLs lack the English proficiency necessary to demonstrate their content knowledge on assessments of academic achievement (Koenig, 2002; Koenig and Bachman, 2004). (Bachman (2002) presents an alternative view, in which the construct to be assessed is something like "the ability to use language to communicate knowledge in an academic discipline"). This view has spurred considerable research into the effects of accommodations – changing aspects of the assessments themselves or in the way they are administered – that are intended to compensate for ELLs' lack of English proficiency (Abedi and Gandara, 2006; Abedi *et al.*, 2004). Although the results of this research are often conflicting and inconclusive (Sireci *et al.*, 2003; Thompson *et al.*, 2002), accommodations for ELLs on assessments of academic achievement continue to be widely used (Rivera *et al.*, 2000).

Ethical Issues in Language Assessment

While research into the nature of language ability and performance on language assessments has been historically and continues to be a major area of focus in language assessment research, this is no longer the sole, or even the dominant, concern of the field. Language testers are investigating the difficult questions about the consequences of assessment use, fairness, and the ethical responsibilities of test developers and users.

Consequences of Assessment Use

Language testers have considered the consequences of assessment use in two domains, that of society at large and language instruction. Viewing language assessment from the perspective of critical social theory, Shohamy (2001) discusses the ways in which language assessments are used to promote unstated objectives aimed at benefiting enfranchised groups in a society. Similarly, McNamara (2001) argues that the very constructs we assess are politically constructed; McNamara and Roever (2006) discuss the societal values that underlie such use and larger sociocultural contexts in which language tests are used. Drawing on moral philosophy, Kunnan (2000a, 2004) has outlined a framework for considering fairness in language assessment. His framework includes several qualities: (1) validity, (2) absence of bias, (3) access, (4), administration, and (5) social consequences. These researchers all argue that greater attention needs to be paid, in both the development of language assessments and in the ways these are used, to the social and political milieu in which language assessment takes place and the consequences of the ways these are used.

Looking at the narrower domain of language instruction, a number of language testing researchers have investigated the impact of language assessments on instruction and instructional processes, referred to in the language testing literature as 'washback' (Alderson and Wall, 1993, 1996; Cheng, 1997; Cheng *et al.*, 2004; Green, 2007; Qi, 2005; Wall and Alderson, 1993). The net result of this research is that washback is not just a simple matter of better assessments producing better instruction. Rather, it consists of a network of complex interactions among all stakeholders in the instructional setting (Hawkey, 2006; Wall, 2005).

Professionalism

Language testers have also grappled with the professionalization of the field and the extent to which language test developers are responsible for the ways in which their tests get used. Davies (1997) and Stansfield (1993) have discussed professional ethics in terms of exemplary practice or creating an ethical milieu for language testing practice. Others have talked about the issues involved in developing a code of professional ethics for language testers throughout the world (Bishop, 2004; Boyd and Davies, 2002). In 2000, the International Language Testing Association adopted a code of ethics.

Justifying Assessment Use

These two strains of concern, the validity of assessment-based interpretations of language ability and ethical issues of assessment use, are coming together in a growing body

of research that investigates both the validity of score interpretations and the consequences of assessment use (papers in Kunnan, 2000b; Reath, 2004). Drawing on argument-based approaches to validation in educational measurement (Kane, 2006; Mislevy *et al.*, 2003), Bachman (2005) describes an assessment use argument (AUA) which he argues can provide the conceptual basis for explicitly linking concerns with validity to those of consequences. Building on this conceptualization, Bachman and Palmer (2010) describe 'assessment' justification as the process of providing a rationale and evidence to justify the use of a particular assessment. Since it is the use of a specific assessment that needs to be justified, justification is inherently local. In other words, the AUA for a particular assessment provides a local theory that makes explicit claims about the roles of consequences, decisions, interpretations, and assessment reports in the assessment, and identifies the evidence that needs to be collected to support these claims. The purpose of an AUA is thus to provide and empirically support a coherent argument that is convincing to stakeholders that using the assessment will help promote the intended beneficial consequences.

The Future: Challenges and Opportunities

Second-language assessment faces two interrelated challenges in the future: bringing together its concerns with issues of validity and use, and addressing the growing need worldwide for the active engagement of language testers in the design, development, and use of practical language assessments for real-world uses.

McNamara (2006) argues that the constructs that underlie real-world language assessments are both socially and politically constructed. From this he concludes, "the political imposition of language testing constructs means that language test validation research will have an impact only to the extent that it is politicized." (p. 38). The challenge will be to move beyond the validation of language ability constructs and the rhetoric of postmodern critical theory. As suggested earlier, one mechanism for explicitly linking the measurement concerns about reliability and validity with the ethical concerns about consequences and professionalism is provided by an AUA such as that proposed by Bachman and Palmer (2010). A conceptual framework provided by an AUA and a critical epistemology that admits a wide range of evidence, from research (qualitative and quantitative) to documentation, regulations, laws, community values, and at the same time critically questions these presents language testers with a daunting challenge. This challenge also presents an opportunity for language testers to bring together their concerns about validity and consequences.

The greatest challenges language assessment as a field faces, however, are in the arenas where language tests are

being used to make decisions about individuals and institutions. Two of these are in the areas of classroom language assessment and assessing foreign languages.

Classroom Assessment

There is a huge demand, worldwide, for greater involvement of individuals with expertise in language testing in the areas of classroom and accountability assessment. Although classroom language assessment is one of the most exciting areas in our field (papers in Broadfoot, 2005; Davison, 2007: 1172; Rea-Dickins, 2000; Rea-Dickins, 2004), this is still not considered mainstream language testing by many. In the past quarter of a century, language testers have been only marginally involved in issues of accountability assessment for K-12 and adult education. This has been and continues to be the case worldwide, where the action in large-scale accountability assessment has been the domain of psychometricians and educators, with language testers providing occasional advice from the periphery. Finally, while the vast majority of published research in language testing over the past half century has focused on learners/users of English as a second or foreign language, there is a growing body of research and experience, again worldwide, in the assessment of languages other than English, as these are being learned both as second and as foreign languages.

The assessment demands of No Child Left Behind (United States Congress, 2001, 2002) in the US have greatly increased the pressure on states to develop more useful assessments both for accountability and in the service of classroom language learning. Similar pressures continue to be seen in other countries as well (Brindley, 1998). Of particular concern to language testers and other applied linguists are issues of assessing the English-language development and academic achievement of ELLs (Hakuta and Beatty, 2000; Koenig, 2002; Solano-Flores and Trumbull, 2003). Similar concerns and issues surround the assessment of ELLs in adult schools (Mislevy and Knowles, 2002; US Department of Education, 2001)

Assessing Foreign-Language Proficiency

Recent initiatives on the part of the US government to increase the nation's capacity in foreign languages are also placing increased demands for useful assessments of foreign languages, particularly the less commonly taught ones (US Department of Defense, 2005; US Department of Education *et al.* 2006). As increasingly larger amounts of government resources at all levels – federal, state, and local – are likely to be going into foreign-language instruction in the coming years, there will most likely be a concomitant need for greater accountability (O'Connell and Norwood, 2007). In K-12 education, there is already

an accountability mechanism in place and it can be expected that as the federal government invests more heavily in foreign language instruction at this level, an accountability mechanism will be required and this will necessitate the development of assessments of foreign-language proficiency that meet the accepted professional standards for validity and impact.

Language testing as a field has grown over the past 30 years in its use of sophisticated statistical and measurement models and ever-refined approaches to naturalistic observation to make advances in understanding the nature of language ability, and will continue to advance in understanding how to deal with the ethical issues of the uses and consequences of language assessments in the real world.

The immediate and long-term prospects for language testing as a field are filled with opportunities and challenges. Turning these opportunities and challenges into accomplishments will depend upon the willingness and capability of language testers to apply the knowledge and skills acquired over the past half a century to the urgent practical assessment needs of our education system, from kindergarten to adult school, and of our society. It will also depend upon our willingness to leave the comfortable confines of the academy and join our colleagues in education and measurement to toil in the fields of practice. It is believed that language testers have a unique combination of knowledge and skills as well as a growing understanding of the issues involved in addressing the validity of interpretations and the consequences of test use. If we can but apply this expertise to the practical problems of assessment in our education systems and society, we are in a position to provide leadership and contribute greatly to making our meritocracy fair and equitable.

See also: Bilingualism and Learning; English Language Learners with Special Needs; Fairness; Foreign Language Learning; Impact of Assessment on Classroom Practice; Second Language Learning; Validity of Educational Indicators.

Bibliography

- Abedi, J. and Gandara, P. (2006). Performance of English language learners as a subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurement: Issues and Practice* 26(5), 36–46.
- Abedi, J., Hofstetter, C., and Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy based research. *Review of Educational Research* 74(1), 1–28.
- Alderson, J. C. and Hughes, A. (eds.) (1981). *Issues in Language Testing*. London: The British Council.
- Alderson, J. C. and Urquhart, A. H. (1985). The effect of students' academic discipline on their performance on ESP reading tests. *Language Testing* 2(2), 192–204.
- Alderson, J. C. and Wall, D. (1993). Does washback exist? *Applied Linguistics* 14, 115–129.
- Alderson, J. C. and Wall, D. (1996). *Special Issue: Washback*. *Language Testing* 13(3).
- American Council on the Teaching of Foreign Languages (1983). *ACTFL Proficiency Guidelines*. Hastings-on-Hudson, NY: Author.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2002). Alternative interpretations of alternative assessments: Some validity issues in educational performance assessments. *Educational Measurement: Issues and Practice* 21(3), 5–18.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly* 2(1), 1–34.
- Bachman, L. F. (2007). What is the construct? The dialectic of abilities and contexts in defining constructs in language assessment. In Fox, J., Wesche, M., and Bayliss, D. (eds.) *What Are We Measuring? Language Testing Reconsidered*, pp 41–72. Ottawa: University of Ottawa Press.
- Bachman, L. F. and Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly* 16(4), 449–465.
- Bachman, L. F. and Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Bachman, L. F. and Palmer, A. S. (2010). *Language Assessment in Practice*, 2nd edn. Oxford: Oxford University Press.
- Bachman, L. F. and Savignon, S. J. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *Modern Language Journal* 70(iv), 380–390.
- Bishop, S. (2004). Thinking about a professional ethics. *Language Assessment Quarterly* 1(1 and 2), 109–122.
- Boyd, K. and Davies, A. (2002). Doctors' orders for language testers: The origin and purpose of ethical codes. *Language Testing* 19(3), 296–322.
- Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programmes. *Language Testing* 15, 45–85.
- Broadfoot, P. M. (2005). Dark alleys and blind bends: Testing the language of learning. *Language Testing* 22(2), 123–141.
- Brown, J. D., Hudson, T., Norris, J. M., and Bonk, W. J. (2002). *An Investigation of Second Language Task-Based Performance Assessments, Vol. SLTCC Technical Report 24*. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawai'i at Manoa.
- Canale, M. and Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1(1), 1–47.
- Carroll, J. B. (1961). Fundamental considerations in testing English proficiency of foreign students. In *Testing the English Proficiency of Foreign Students*, pp 30–40. Washington, DC: Center for Applied Linguistics.
- Carroll, J. B. (1968). The psychology of language testing. In Davies, A. (ed.) *Language Testing Symposium: A Psycholinguistic Approach*, pp 46–69. London: Oxford University Press.
- Carroll, J. B. (1983). Psychometric theory and language testing. In Oller, J. W. Jr (ed.) *Issues in Language Testing Research*, pp 80–105. Rowley, MA: Newbury House.
- Chalhoub-Deville, M. (1995). A contextualized approach to describing oral language proficiency. *Language Learning* 45(2), 251–281.
- Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing* 20(4), 369–383.
- Chalhoub-Deville, M. and Deville, C. (2005). A look back at and forward to what language testers measure. In Hinkel, E. (ed.) *Handbook of Research in Second Language Teaching and Learning*, pp 815–831. Mahwah, NJ: Earlbaum.
- Chapelle, C. (1998). Construct definition and validity inquiry in SLA research. In Bachman, L. F. and Cohen, A. D. (eds.) *Interfaces between Second Language Acquisition and Language Testing Research*, pp 32–70. New York: Cambridge University Press.
- Chapelle, C. A. (2006). L2 vocabulary acquisition theory: The role of inference, dependability and generalizability in assessment. In Chalhoub-Deville, M., Chapelle, C. A., and Duff, P. A. (eds.) *Inference and Generalizability in Applied Linguistics: Multiple Perspectives*, pp 47–64. Dordrecht: John Benjamins.

- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education* 11(1), 38–54.
- Cheng, L., Watanabe, Y., and Curtis, A. (eds.) (2004). *Washback in Language Testing: Research Contexts and Methods*. Mahwah, NJ: Erlbaum.
- Clapham, C. (1996). *The Development of IELTS: A Study of the Effect of Background Knowledge on Reading Comprehension* vol. 4. Cambridge: University of Cambridge Local Examinations Syndicate/Cambridge University Press.
- Clark, J. L. D. (1972). *Foreign Language Testing: Theory and Practice*. Philadelphia, PA: Center for Curriculum Development.
- Clark, J. L. D. (1975). Theoretical and technical considerations in oral proficiency testing. In Jones, R. L. and Spolsky, B. (eds.) *Testing Language Proficiency*, pp 10–24. Arlington, VA: Center for Applied Linguistics.
- Clark, J. L. D. (1979). Direct vs. Semi-direct tests of speaking ability. In Brière, E. and Hinofotis, F. B. (eds.) *Concepts in Language Testing: Some Recent Studies*, pp 35–49. Washington, DC: TESOL.
- Davies, A. (1997). *Special Issue: Ethics in Language Testing*. *Language Testing* 14(3).
- Davison, C. (2007). Views from the chalkface: English language school-based assessment in Hong Kong. *Language Assessment Quarterly* 4(4), 37–68.
- Douglas, D. (2000). *Assessing Language for Specific Purposes: Theory and Practice*. Cambridge: Cambridge University Press.
- Douglas, D. and Selinker, L. (1993). Performance on a general versus a field-specific test of speaking proficiency by international teaching assistants. In Douglas, D. and Chapelle, C. (eds.) *A New Decade of Language Testing Research*, pp 235–256. Arlington, VA: Teachers of English to Speakers of Other Languages.
- Green, A. (2007). Watching for washback: Observing the influence of the international English language testing system academic writing test in the classroom. *Language Assessment Quarterly* 3(3), 333–368.
- Hakuta, K. and Beatty, A. (eds.) (2000). *Testing English-Language Learners in U. S. Schools*. Washington, DC: National Academy Press.
- Hamp-Lyons, L. and Lunley, T. (2001). Special issue: Assessing language for specific purposes. *Language Testing* 18(2).
- Hawkey, R. (2006). *Impact Theory and Practice*. Cambridge: Cambridge University Press.
- Hudson, T., Detmer, E., and Brown, J. D. (1992). *A Framework for Testing Cross-Cultural Pragmatics*. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawai'i at Manoa.
- Hudson, T., Detmer, E., and Brown, J. D. (1995). *Developing Prototypic Measures of Cross-Cultural Pragmatics*. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawai'i at Manoa.
- Hughes, A. and Porter, D. (eds.) (1983). *Current Developments in Language Testing*. London: Academic Press.
- Jones, R. L. (1979). Performance testing of second language proficiency. In Brière, E. J. and Hinofotis, F. B. (eds.) *Concepts in Language Testing: Some Recent Studies*, pp 50–57. Washington, DC: TESOL.
- Jones, R. L. (1985). Some basic considerations in testing oral proficiency. In Lee, Y. P., Fok, A. C. Y., Lord, R., and Low, G. (eds.) *New Directions in Language Testing*, pp 77–84. Oxford: Pergamon Press.
- Kane, M. (2006). Validation. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn, pp 17–64. New York: American Council on Education and Praeger Publishers.
- Koenig, J. A. (ed.) (2002). *Reporting Test Results for Students with Disabilities and English-Language Learners*. Washington, DC: National Academy Press.
- Koenig, J. A. and Bachman, L. F. (eds.) (2004). *Keeping Score for All: The Effects of Inclusion and Accommodation Policies on Large-Scale Educational Assessment*. Washington, DC: National Research Council, National Academies Press.
- Kramsch, C. (1986). From language proficiency to interactional competence. *Modern Language Journal* 70(4), 366–372.
- Kunnan, A. J. (2000a). Fairness and justice for all. In Kunnan, A. J. (ed.) *Fairness and Validation in Language Assessment*, pp 1–14. Cambridge, UK: Cambridge University Press.
- Kunnan, A. J. (ed.) (2000b). *Fairness and Validation in Language Assessment: Selected Papers from the 19th Language Testing Research Colloquium, Orlando*. Cambridge: Cambridge University Press.
- Kunnan, A. J. (2004). Test fairness. In Milanovic, M. and Weir, C. (eds.) *European Language Testing in a Global Context*, pp 27–48. Cambridge, UK: Cambridge University Press.
- Lado, R. (1961). *Language Testing*. New York: McGraw-Hill.
- Lantolf, J. P. and Frawley, W. (1985). Oral proficiency testing: A critical analysis. *Modern Language Journal* 69, 337–345.
- McNamara, T. F. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics* 18, 446–46.
- McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing* 18(4), 333–349.
- McNamara, T. (2003). Looking back, looking forward: Rethinking Bachman. *Language Testing* 20(4), 466–473.
- McNamara, T. (2006). Validity and values: Inferences and generalizability in language testing. In Chalhoub-Deville, M., Chapelle, C. A., and Duff, P. A. (eds.) *Inference and Generalizability in Applied Linguistics: Multiple Research Perspectives*, pp 27–45. Dordrecht: John Benjamins.
- McNamara, T. and Roever, K. (2006). *Language Testing: The Social Dimension*. Malden, MA: Blackwell.
- Mislevy, R. J. and Knowles, K. (eds.) (2002). *Performance Assessments for Adult Education: Exploring Measurement Issues*. Washington, DC: National Academy Press.
- Mislevy, R. J., Steinberg, L. S., and Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives* 1(1), 3–62.
- Norris, J. M., Brown, J. D., Hudson, T., and Yoshioka, J. (1998). *Designing Second Language Performance Assessments, Vol. SLTCC Technical Report #18*. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawai'i at Manoa.
- O'Connell, M. E. and Norwood, J. L. (eds.) (2007). *International Education and Foreign Languages: Keys to Securing America's Future*. Washington, DC: National Academies Press.
- Oller, J. W., Jr. (1979). *Language Tests at School*. London: Longman.
- Oller, J. W., Jr. (1983). A consensus for the eighties? In John, J. and Oller, W. (eds.) *Issues in Language Testing Research*, pp 351–356. Rowley, MA: Newbury House.
- Palmer, A. S., Groot, P. J. M., and Trosper, G. A. (eds.) (1981). *The Construct Validation of Tests of Communicative Competence*. Washington, DC: TESOL.
- Pappajohn, D. (1999). The effect of topic variation in performance testing: The case of the chemistry teach test for international teaching assistants. *Language Testing* 16(1), 52–81.
- Purpura, J. E. (2004). *Assessing Grammar*. Cambridge: Cambridge University Press.
- Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing* 22(2), 142–173.
- Rea-Dickins, P. (2000). *Special Issue: Assessing Young Learners*. *Language Testing* 1.
- Rea-Dickins, P. (2004). *Special Issue: Exploring Diversity in Teacher Assessment*. *Language Testing* 21(3).
- Reath, A. (2004). Language analysis in the context of the asylum process: Procedures, validity and consequences. *Language Assessment Quarterly* 1(4), 209–233.
- Rivera, C. (ed.) (1984). *Communicative Competence Approaches to Language Proficiency Assessment: Research and Application*. Clevedon: Multilingual Matters.
- Rivera, C., Stansfield, C., Scialdone, L., and Sherkey, M. (2000). *An Analysis of State Policies for Inclusion and Accommodation of ELLs in State Assessment Programs During 1998–1999*. Center for Equity and Excellence in Education, The George Washington University.
- Roever, K. (2006). Validation of a web-based test of ESL pragmalinguistics. *Language Testing* 23(2), 229–256.
- Shohamy, E. (2001). *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. London: Pearson.
- Sireci, S., Li, S., and Scarpati, S. (2003). *The Effects of Test Accommodations on Test Performance: A Review of the Literature*. Amherst, MA: University of Massachusetts.

- Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics* 17(1), 38–62.
- Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.
- Skehan, P. (2001). Tasks and language performance assessment. In Bygate, M., Skehan, P., and Swain, M. (eds.) *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing*. Harlow: Longman.
- Solano-Flores, G. and Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher* 32(2), 3–13.
- Spearman, C. E. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology* 15, 201–293.
- Stansfield, C. W. (1993). Ethics, standards and professionalism in language testing. *Issues in Applied Linguistics* 4(2), 15–30.
- Thompson, S., Blount, A., and Thurlow, M. (2002). *A Summary of Research on the Effects of Test Accommodations: 1999 through 2001*. Minneapolis, MN: University of Minnesota.
- US Department of Defense (2005). *Defense Language Transformation Roadmap*. Arlington, VA.
- US Department of Education (2001). *Measures and Methods for the National Reporting System for Adult Education: Implementation Guidelines*. Washington, DC: Author.
- US Department of Education, US Department of State, US Department of Defense, and Office of the Director of National Intelligence (2006). National security language initiative. www.ed.gov/about/inits/ed/competitiveness/nsli/index.html (accessed September 2009).
- United States Congress (2001). H.R. 1, no child left behind act of 2001.
- United States Congress (2002). Public law 107-110, no child left behind act of 2001.
- Upshur, J. A. (1979). Functional proficiency theory and a research role for language tests. In Brière, E. and Hinofotis, F. B. (eds.) *Concepts in Language Testing: Some Recent Studies*, pp 75–100. Washington, DC: TESOL.
- Vollmer, H. J. and Sang, F. (1983). Competing hypotheses about second language ability: A plea for caution. In John, J. and Oller, W. (eds.) *Issues in Language Testing Research*, pp 29–74. Rowley, MA: Newbury House.
- Wall, D. (2005). *The Impact of High-Stakes Examinations on Classroom Teaching: A Case Study Using Insights from Testing and Innovation Theory*. Cambridge: Cambridge University Press.
- Wall, D. and Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing* 10(1), 41–69.
- Wesche, M. B. (1987). Second language performance testing: The Ontario test of ESL as an example. *Language Testing* 4(1), 28–47.
- Wylie, E. and Ingram, D. E. (2006). *International Second Language Proficiency Ratings (ISLPR): General Proficiency Version for English*. Brisbane: Centre for Applied Linguistics and Language, Brisbane University.
- Yamashita, S. O. (1996). *Six Measures of JSL Pragmatics*. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawai'i at Manoa.
- Young, R. F. (2000). Interactional competence: Challenges for validity. *A Joint Symposium, "Interdisciplinary Interfaces with Language Testing", of the Language Testing Research Colloquium and the American Association for Applied Linguistics*. Vancouver, BC.
- Young, R. F. and He, A. W. (eds.) (1998). *Talking and Testing*. Amsterdam: John Benjamins Publishing.

Relevant Websites

- <http://www.occupationalenglishtest.org> – Australian Federal Government: Occupational English Test.
- <http://www.eurocontrol.int> – EUROCONTROL, European Organisation for the Safety of Air Navigation: Proficiency Test in English Language for Air Traffic Controllers (PELA).
- <http://www.bulats.org> – University of Cambridge ESOL Examinations: Business Language Testing Service (BULATS).

Self Adaptive Testing

J Revuelta, Autonoma University of Madrid, Madrid, Spain

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Self-adaptive test – An adaptive test in which the examinee is allowed to choose the difficulty category of each administered item.

Self-adaptive tests (SATs) were proposed in the context of psychological and educational measurement as an alternative to computerized adaptive tests (CATs). The purpose of a CAT is to obtain a precise estimate of ability with a small number of items. CATs are efficient tests tailored to a specific examinee. The item difficulty is matched to the examinee's ability. CATs apply more difficult items to more competent individuals, and easier items are applied to those less proficient. The psychometric basis for CAT is item response theory (Hambleton and Swaminathan, 1985), which provides the mathematical justification for this operation mode.

Because in CAT item difficulties roughly match an examinee's ability, the percentage of items passed is only about 50% for any individual, irrespective of ability. Some authors have questioned whether CATs are appropriate from a motivational point of view (Andrich, 1995). As a result, variants of CATs have been proposed to ease examinee's comfort and reduce anxiety. For instance, easy CATs are adaptive tests that match item difficulty to a smaller ability than the estimated one, with the purpose of increasing the proportion of items passed (Lunz and Bergstrom, 1994; Revuelta *et al.*, 2003).

A SAT is an adaptive test in which the examinee is allowed to choose the difficulty category of each administered item (Rocklin and O'Donnell, 1987; Rocklin, 1994; Johnson *et al.*, 1991). The item bank is divided into five or seven difficulty levels. Prior to the administration of each item, the examinee chooses one of the difficulty levels and the computer program selects the most informative items for that group of items based on the most recent ability estimate. In most applications of SAT, previous to the selection of a difficulty level for an incoming item, the computer program provides feedback on the correctness of the last response (Roos *et al.*, 1997; Vispoel, 1998).

The idea behind SAT is that allowing the examinee to play some role in the item-selection process may be beneficial from a motivational point of view. In particular, the possibility of selecting difficulties may be helpful for

reducing anxiety during the test. The consequence would be a purer and more valid estimate of ability, that is, a test score that reflects better the true ability level and that has not been contaminated by anxiety during the test.

Figure 1 depicts the flowchart for a SAT administration session. The test begins with a provisional ability estimate, which may be a value close to the mean population value. Step 2 is specific for SAT and consists of selecting one of the difficulty levels; this step is missing during a CAT. Step 3 consists of selecting an item within the group of items with required difficulty. Item selection is based on the statistical information and the content constraints inherent to adaptive testing. During a CAT, the item bank is not divided into groups of homogeneous difficulty, and the testing algorithm is not subjected to difficulty constraints. The selected item is administered to the examinee and the response is recorded in step 4. Step 5 consists of providing feedback on the correctness of the response; this step is usually missing in CAT. The estimate of ability and the standard error are computed in step 6. The termination criteria involved in step 7 are the test length, standard error of ability, and content constraints. The test ends when the individual has received enough items from any content category and the standard error of the estimated ability is smaller than some arbitrary criterion, such as 0.3.

Item response theory is needed at two points of the flowchart: in step 3, where items are selected partially based on the statistical information that they provide about the estimated ability, and in step 6, where the abilities are estimated on the same measurement scale for individuals that have responded to different items.

Investigations on Anxiety and Comfort

A body of applied psychometric literature has focused on the benefits of SAT for increasing motivation and reducing anxiety during the test. These investigations are based on two hypotheses (Revuelta *et al.*, 2003):

1. *Perceived control hypothesis.* The test is considered a source of stress. Because the stress is reduced when the individual has some control over the aversive situation, the perceived control over the test is the cause of a reduction in anxiety (Wise, 1994).
2. *Perceived performance hypothesis.* Because in SAT the individuals obtain feedback on the number of errors, the selection of difficulty may be used to increase the

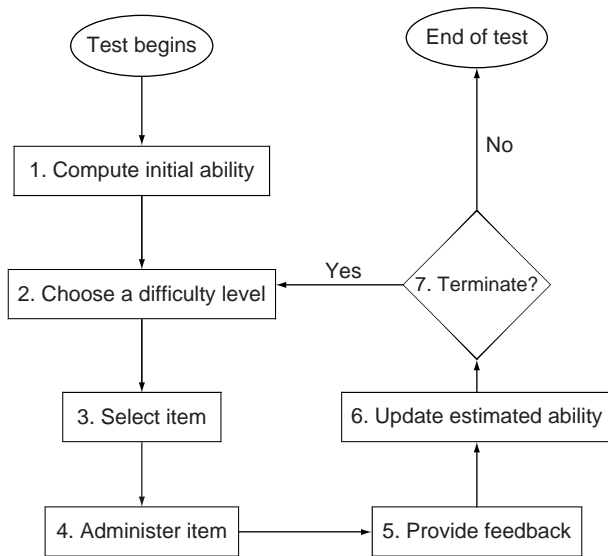


Figure 1 Flowchart describing a SAT.

number of items passed. Anxiety is reduced because difficulties can be adjusted to increase the number of items passed, which is interpreted by the examinee as a better performance (Ponsoda *et al.*, 1999).

The perceived control hypothesis provides an explanation on the effects of SAT that is based on the investigations on anxiety by Averill (1973): the opportunity to modify the aversive stimuli (the test) reduces the stress in comparison to those situations where the control is absent. However, Wise *et al.* (unpublished) showed that not all the individuals who respond to a SAT wish to have that control. If an individual does not desire to control difficulty, SAT may produce an increase of anxiety and a reduction of ability. In summary, SAT would be beneficial only for those individuals with high anxiety and low self-conception on the ability measured by the test (Wise *et al.*, 1994).

The perceived control and the perceived performance hypotheses do not necessarily exclude each other. In fact, the control provided to the examinee is a control of the difficulty levels, which can be adjusted to increase the perceived performance on the test. The perception of control depends also on the instructions given to the examinees. Typically, the instructions indicate that correct responses are weighted by item difficulty to obtain the test score. A high number of correct responses to easy items produce a lower score than a smaller number of correct responses to harder items. In this way, difficulties may be selected to increase the test score, and the control on difficulty may be perceived as an indirect control on the test score or the performance on the test.

The empirical studies have reported some important differences between CAT and SAT and provided support for these hypotheses. In general, SAT yields higher mean

ability estimates and a lower posttest anxiety than CAT (Shermis *et al.*, 2001; Vispoel and Coffman, 1994; Vispoel *et al.*, 1994; Wise *et al.*, 1992; Ponsoda *et al.*, 1997; Rocklin, 1996). Pitkin and Vispoel (2001) conducted a meta-analysis to integrate the different published studies and to gain a better understanding of the differences between CAT and SAT. The result is that estimated ability is higher in SAT, although the estimated effect size is modest (0.11). Moreover, SAT reduces the negative effects of anxiety on test performance, and the standardized effect size for anxiety reduction is 0.18.

The benefits of SAT are not attained without cost; in particular, testing time is longer because the individuals have to spend some time selecting difficulties. Regarding measurement precision, the standard error of ability is higher and reliability is lower than in CAT. Moreover, the instructions given to the examinee are more complex and important in a SAT than in a CAT. If an individual selects difficulty widely apart from the true ability level, the measurement precision would be very low. For these reasons, it is very important that the instructions of the test must very clearly state that the individual should avoid those difficulty levels that do not involve a challenge, that is, those items that are routinely passed or failed. SAT is also very sensitive to the pattern of difficulties, and there is no way to obtain a precise ability estimate if the individuals ignore the instructions.

SAT may also present difficulties related to the development and security of the item pool. Even though most adaptive tests measure a unidimensional ability, they apply some type of control balance of the test content. The item pool is divided into several content groups, and the final test should have a balance number of items from each content category. In SAT, each difficulty level should have a sufficient number of items from each content category because some individuals select all the items from the same difficulty level. Moreover, item exposure control strategies shall be developed specifically for SAT to avoid the occurrence of individuals selecting the same difficulties and receiving the same items, and to obtain a balanced item exposure rate within each difficulty group.

Because of the problems encountered with a pure SAT testing strategy, some hybrid testing strategies have been proposed between CAT and SAT. For example, restricted self-adapted tests (RSAT; Wise *et al.*, 1993) prevent examinees from choosing difficulty levels that differ widely from the most informative one. In an RSAT, the range of choices is restricted to a region around the provisional estimate of ability. Results on RSAT regarding measurement precision and anxiety are more similar to CAT than to SAT (Roos *et al.*, 1998; Plake *et al.*, 1995), and the supposed benefits of SAT are not attained.

Assisted self-adapted testing (ASAT; Hontangas *et al.*, 2004) provides advice to the examinee regarding the difficulty level that best matches the ability estimates, and

yields results similar to CAT. A third variant is the SCAT (self-adaptive supplemented with computerized adaptive test), which acts as a SAT at the beginning of the testing session and as a CAT in the middle and the end of the test (Ponsoda *et al.*, 1997). The idea behind SCAT is that in the first part of a CAT the provisional ability estimate may differ widely from the final one; and the difficulties selected by examinee may be better matched to ability. Finally, the easy SAT recommends the examinee to select easier items than the difficulty level that best matches the ability estimate (Ponsoda *et al.*, 1999).

Investigations on Item-Selection Strategy

The examinee's control over the administration of items gives rise to a number of item-selection strategies that may be related to ability or some other personal characteristics. The item-selection strategy is the association pattern between the response given to an item (pass/fail) and the difficulty selected for the next one. A body of applied research has focused on this problem and has identified two strategies (Hontangas *et al.*, 2000; Johnson *et al.*, 1991; Rocklin, 1989):

1. The rigid strategy is followed by those examinees that select the same difficulty level throughout the test.
2. The flexible strategy consists in selecting a harder level after a correct response and an easier one after an incorrect one.

Figure 2 illustrates the strategies by showing the difficulties selected by two individuals during a 10-item SAT. The item bank is divided into five difficulty levels. The black dots denote the passed items, whereas the white dots are for the failed ones. The dotted line corresponds to an individual who follows the rigid strategy. This individual selects the same difficulty throughout the test, with some minor changes, irrespective of the response to the previous item. The solid line corresponds to an individual who follows the flexible strategy. This individual selects the same difficulty or a higher one after a correct response,

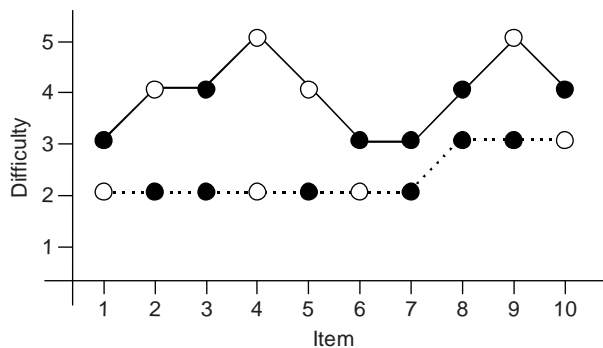


Figure 2 Difficulties selected by two individuals during a SAT session.

and the same difficulty or an easier level after an incorrect response. Then, the flexible strategy consists of searching the more challenging difficulty level, those with a higher uncertainty on the correctness of the response.

The flexible and the rigid strategies can be considered as theoretical patterns for selecting difficulties. In real SAT applications, many individuals exhibit strategies that combine these two or change their strategy throughout the test. For example, the failure-tolerant strategy is followed by those individuals that select the same difficulty level after failing an item and a harder one after a correct response.

Theoretical Investigations on the Statistical Properties of SAT

A theoretical investigation on the statistical properties of SAT was conducted by Revuelta (2004). Bradlow and Thomas (1998) had demonstrated that missing responses are not missing at random data in choice-based examinations (those that permit an examinee to select the items from a designated list). The consequence is that the maximum-likelihood estimate of ability commonly used in item response theory is not valid for this type of test. Conversely, Mislevy and Wu (1996) showed that this problem does not apply to CAT. In SAT, the examinee has more control over the item-selection mechanism in CAT but less than in choice-based examinations. For these reasons, the possibility of estimating ability in SAT remained problematic.

Revuelta (2004) showed that in SAT, the population of examinees is divided into several subpopulations that differ in the item-selection strategy. Moreover, he proposed a latent class model to estimate the number of strategies and identify the strategy followed by each examinee, and applied it to a real data sample. More importantly, this investigation showed that the strategy may be related to ability, which implies that the strategy should be taken into account when estimating abilities.

Consider step 6 of the flowchart depicted in **Figure 1**. Let θ be the ability of the individual and x_i the response given to the i th item. The probability of observing a response pattern $x = (x_1, \dots, x_I)'$ is given by

$$f(x; \theta) = \prod_{i=1}^I f(x_i) b(i), \quad [1]$$

where $b(i)$ represents the probability of selecting item i from the bank and $f(x_i)$ is the probability of the response x_i . In a CAT, $b(i)$ depends on the properties of the item-selection algorithm. The function $f(x_i)$ is computed by using one of the models of item response theory, for example, the one-, two-, or three-parameter logistic models (Hambleton and Swaminathan, 1985). During a CAT, the maximum-likelihood estimate of θ is obtained by maximizing $f(x; \theta)$ with respect to θ .

Consider a population of SAT-taker individuals that is divided into several strategies. Let S be the number of strategies and p_s be the probability of sampling an individual that follows strategy s . The probability of the observed responses is given by

$$g(\mathbf{x}; \theta) = \sum_{s=1}^S p_s \prod_{i=1}^I f(x_i) b_s(i), \quad [2]$$

where $b_s(i)$ is the probability of selecting item i for those individuals that follow strategy s . The probability $g(\mathbf{x}; \theta)$ is a marginal probability function because the strategy followed by the individual is unknown, and the parameters of $b_s(i)$ and p_s are estimated from the data.

The result of the empirical application of the model described in eqn [2] is that there is an association between the strategy and the ability of the individual, and the same strategies are not used for individuals of different ability. In more formal terms, the probability p_s depends on θ . The consequence is that p_s contains information about θ . Therefore, in a SAT the maximum-likelihood estimate of θ is obtained maximizing eqn [2] instead of [1], which implies that the pattern of difficulties is used for estimating ability.

From a purely measurement perspective, this is undesirable because the ability estimate depends not only on the pattern of correct and incorrect responses but also on the item-selection strategy. Moreover, this finding could be exploited by the examinees to obtain a positively biased ability estimate by using the most optimal strategy. For these reasons, SAT appears to be more appropriate for investigating personality characteristics, such as risk-taking behavior or anxiety reduction, than as a tool for obtaining a precise ability estimate.

The theoretical investigations on SAT suggest that the item-selection strategy affects the scores and may be the cause of biases on the ability estimate. The existence and magnitude of these biases have not yet been clearly shown and there is still another unresolved problem: the robustness of the traditional maximum-likelihood estimate to violations of the assumption that the strategy is unrelated to ability. That is, the consequences of estimating ability by using the eqn [1] instead of [2] during a SAT session. Surely, the response would depend on the strength of such a relation, as formalized by the relation between θ and p_s .

Conclusions

The purpose of SAT is to obtain a more valid ability estimate by reducing the effect of anxiety on the test score. Empirical research has shown that not all individuals are highly anxious and benefit from SAT, and that highly anxious individuals show differential tendencies to cope with anxiety and do not necessarily benefit from SAT.

SAT can be viewed as a strategy of test accommodation devoted to some of the individuals that exhibit high-state anxiety related to the testing situation. From this viewpoint, anxiety is considered a disability that prevents the individuals from performing their best during the test. Chapter 10 of the Standards for Educational and Educational Testing, published by the American Psychological Association *et al.* (1999), is devoted to test individuals with disabilities, where the test modifications are designed to accommodate the needs of these individuals. In the face of these standards, there is still a long way to go before SAT can be routinely applied in a real testing scenario.

To apply SAT in a real evaluation, an important concern should be to clearly identify the group of individuals that present this disability. Moreover, the theoretical mechanisms that underlie test anxiety, coping strategy, and ability need to be investigated further, and more empirical evidence should be provided on the properties of SAT scores for groups of individuals with different anxiety and personal characteristics (Rocklin *et al.*, 1995).

Another important unresolved problem is the validity of the SAT scores. Supposedly, SAT may provide a purer measure of ability than CAT because the influence of anxiety is removed and the individuals may perform at his/her best. Research has shown that mean scores are higher for SAT than CAT, although the explanation of this finding in relation with construct validity is still unclear. Moreover, not all the individuals deal with anxiety in the same manner and the benefits of SAT may depend on personal characteristics, even for the population of highly anxious individuals. Thus, it has not yet been established whether this kind of test accommodation is legitimately applied only to the subgroup of highly anxious individuals, to some of them or to any examinee. Finally, if the hypotheses about the underlying mechanisms of SAT are correct, the consequence should be a purer measure of ability and an increased predictive validity. However, this hypothesis is yet to be tested in empirical studies.

Given the current state of research, SAT appears to be a useful tool to investigate stylistic characteristics of the individuals, such as risk-taking behavior or locus of control. SAT can be also applied in low-stakes evaluations such as diagnosing learning difficulty, in which the examinees themselves may wish to obtain a realistic ability estimate. However, the validity of SAT for measuring ability in high-stakes settings constitutes an open problem that requires further investigation.

Bibliography

- American Psychological Association, American Educational Research Association and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

- Andrich, D. (1995). Review of the book computerized adaptive testing: A primer. *Psychometrika* **4**, 615–620.
- Averill, J. R. (1973). Personal control over aversive stimuli and its relationship to stress. *Psychological Bulletin* **80**, 286–303.
- Bradlow, E. T. and Thomas, N. (1998). Item response theory models applied to data allowing examinee choice. *Journal of Educational and Behavioral Statistics* **23**, 236–243.
- Hambleton, R. K. and Swaminathan, S. (1985). *Item Response Theory. Principles and Applications*. Boston, MA: Kluwer.
- Hontangas, P., Olea, J., Ponsoda, V., Revuelta, J., and Wise, S. (2004). Assisted self-adapted testing: A comparative study. *European Journal of Psychological Assessment* **20**, 2–9.
- Hontangas, P., Ponsoda, V., Olea, J., and Wise, S. (2000). The choice of item difficulty in self-adapted testing. *European Journal of Psychological Assessment* **16**, 3–12.
- Johnson, P. L., Ross, L. L., Wise, S. L., and Plake, B. S. (1991). Correlates of examinee item choice behavior in self-adapted testing. *Mid-Western Educational Researcher* **4**, 25–28.
- Lunz, M. E. and Bergstrom, B. A. (1994). An empirical study of computerized adaptive test administration conditions. *Journal of Educational Measurement* **31**, 251–263.
- Mislevy, R. J. and Wu, P.-K. (1996). Missing responses and IRT ability estimation: Omits, choice, time limits and adaptive testing. *Research Report RR-96-30-ONR*. Princeton, NJ: Educational Testing Service.
- Pitkin, A. K. and Vispoel, W. P. (2001). Differences between self-adapted and computerized adaptive tests: A meta-analysis. *Journal of Educational Measurement* **38**, 235–247.
- Plake, B. S., Wise, S. L., and Roos, L. L. (1995). Effects of informed item selection on test performance and anxiety for examinees administered a self-adapted test. *Educational and Psychological Measurement* **55**, 736–742.
- Ponsoda, V., Olea, J., Rodríguez, M. S., and Revuelta, J. (1999). The effects of test difficulty manipulation in computerized adaptive testing and self adaptive testing. *Applied Measurement in Education* **12**, 167–184.
- Ponsoda, V., Wise, S., Olea, J., and Revuelta, J. (1997). An investigation of self adapted testing in a Spanish high school population. *Educational and Psychological Measurement* **57**, 210–221.
- Revuelta, J. (2004). Estimating ability and item-selection strategy in self-adapted testing. A latent class approach. *Journal of Educational and Behavioral Statistics* **29**, 379–396.
- Revuelta, J., Ximénez, M. C., and Olea, J. (2003). Psychometric and psychological effects of item selection and review on computerized tests. *Educational and Psychological Measurement* **63**, 791–808.
- Rocklin, T. R. (1989). Individual differences in item selection in self-adapted testing. *Annual meeting of the American Educational Research Association*. San Francisco, CA.
- Rocklin, T. R. (1994). Self-adapted testing. *Applied Measurement in Education* **7**, 3–14.
- Rocklin, T. R. (1996). Self-adapted testing: Improving performance by modifying tests instead of examinees. *Anxiety, Stress and Coping: An International Journal* **10**, 83–104.
- Rocklin, T. R. and O'Donnell, A. M. (1987). Self-adapted testing: A performance improving variant of computerized adaptive testing. *Journal of Educational Psychology* **79**, 315–319.
- Rocklin, T. R., O'Donnell, A. M., and Holst, P. M. (1995). Effects and underlying mechanisms of self-adaptive testing. *Journal of Educational Psychology* **87**, 103–116.
- Roos, L. L., Wise, S. L., and Finney, S. L. (1998). Comparing restricted and unrestricted self-adapted testing as alternatives to computerized adaptive testing. *Paper Presented at the Annual Meeting of the National Council on Measurement in Education*. San Diego, CA.
- Roos, L. L., Wise, S. L., and Plake, B. S. (1997). The effect of feedback in computerized adaptive and self-adaptive testing. *Educational and Psychological Measurement* **57**, 85–98.
- Shermis, M. D., Mzumara, H. R., and Bublitz, S. T. (2001). On test and computer anxiety: Test performance under CAT and SAT conditions. *Journal of Educational Computing Research* **24**, 57–75.
- Vispoel, W. P. (1998). Psychometric characteristics of computer-adaptive and self-adaptive vocabulary tests: The role of answer feedback and test anxiety. *Journal of Educational Measurement* **35**, 155–167.
- Vispoel, W. P. and Coffman, D. D. (1994). Computer-adaptive and self-adaptive music listening tests. Psychometric features and motivational benefits. *Applied Measurement in Education* **7**, 25–52.
- Vispoel, W. P., Rocklin, T. R., and Wang, T. (1994). Individual differences and test administration procedures. A comparison of fixed-effects, computerized adaptive and self-adapted testing. *Applied Measurement in Education* **7**, 53–79.
- Wise, S. L. (1994). Understanding self-adapted testing: The perceived control hypothesis. *Applied Measurement in Education* **7**, 15–24.
- Wise, S. L., Kingsbury, G. G., and Houser, R. L. (1993). An investigation of self adapted testing. *Paper Presented at the Annual Meeting of the National Council on Measurement in Education*. Atlanta, GA.
- Wise, S. L., Plake, B. S., Johnson, P. L., and Ross, L. L. (1992). A comparison of self-adapted and computerized adaptive tests. *Journal of Educational Measurement* **29**, 329–339.
- Wise, S. L., Roos, L. L., Plake, B. S., and Nebelsick-Gullet, L. J. (1994). The relationship between examinee anxiety and preference for self-adapted testing. *Applied Measurement in Education* **7**, 81–91.
- Wise, S. L., Roos, L. L., Schraw, G. J., Plake, B. S. (unpublished). *Comparing Self-Adapted and Computerized Adaptive tests. The Influence of Examinee Locus of Control and Desire for Control*.

Further Reading

- Frederiksen, N., Mislevy, R. J., and Bejar, I. I. (1993). *Test Theory for a New Generation of Tests*. Hillsdale, NJ: LEA.
- Mislevy, R. J. and Chang, H. H. (2000). Does adaptive testing violate local independence? *Psychometrika* **65**, 149–156.
- Olea, J., Revuelta, J., Ximénez, M. C., and Abad, F. J. (2000). Psychometric and psychological effects of review on computerized fixed and adaptive tests. *Psicológica* **21**, 157–173.
- Roos, L. L., Wise, S. L., Yoes, M. E., and Rocklin, T. R. (1996). Conducting self-adaptive testing using MicroCAT. *Educational and Psychological Measurement* **56**, 821–827.
- Sands, W. A., Waters, B. K., and McBride, J. R. (eds.) (1997). *Computerized Adaptive Testing: From Inquiry to Operation*. Washington, DC: APA.
- van der Linden, W. J. and Glas, C. (2000). *Computerized Adaptive Testing. Theory and Practice*. Boston, MA: Kluwer.
- Vispoel, W. P. (1998). Reviewing and changing answers on computer-adaptive and self-adaptive vocabulary tests. *Journal of Educational Measurement* **35**, 328–347.
- Wainer, H., Dorans, N., Eignor, D., et al. (eds.) (2000). *Computerized Adaptive Testing. A Primer*, 2nd edn. Hillsdale, NJ: LEA.
- Wainer, H., Wang, X.-B., and Thissen, D. (1994). How well can we compare scores on test forms that are constructed by examinees' choice? *Journal of Educational Measurement* **31**, 183–199.
- Weiss, D. J. (ed.) (1983). *New Horizons in Testing. Latent Trait Theory and Computerized Adaptive Testing*. New York: Academic Press.

Student Test Results in School Accountability

L S Hamilton, RAND Corporation, Pittsburgh, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Accountability index – Measure that uses specific rules to combine test scores into a single number or label for use in accountability decisions.

Status measure – Score or aggregate of scores collected at a single point in time.

Successive cohort measure – An index that compares student performance in a school (or grade) in one year to the performance of students in that same school (or grade) in the previous year.

Quasilongitudinal measure – An index that compares the performance of a school's students in a specific grade to the performance of students in the previous grade in the previous year.

Value-added models – Complex statistical models that attempt to isolate the contribution of a teacher or school to a student's achievement.

Score inflation – Gains in scores on high-stakes tests that are larger than the gains in the underlying construct the test is designed to measure.

Test-based accountability (TBA) systems that hold educators accountable for students' scores on standardized achievement tests are prominent in many school systems throughout the world. In the United States, such systems became the cornerstone of federal education policy as a result of the No Child Left Behind Act (NCLB) of 2001, and many states had TBA systems in place for several years prior to NCLB. Most TBA systems report scores for schools or school systems, although some programs are designed to hold individual teachers and principals accountable through pay-for-performance plans in which salary increases or one-time bonuses are awarded on the basis of student test scores.

Large-scale, standardized achievement tests are the central component of these systems. These tests have traditionally served a variety of purposes in education and have been used in both high-stakes and low-stakes contexts. Standardized tests are often used to monitor the education system as a whole, and in many schools they are intended to play an important role providing feedback to help educators make decisions about instruction. High stakes for individual examinees are sometimes associated with tests: scores may be used to make decisions about course assignment, high-school graduation, professional certification, licensure, and other decisions that have

consequences for the person taking the test. The distinguishing characteristic of the TBA systems discussed in this article, as opposed to other ways of using test scores, is the linking of students' scores to consequences for the adults who have responsibility for teaching those students.

This article explores the use of tests in accountability systems. It describes different metrics that have been constructed from test scores for use in TBA, and discusses some of the strengths and limitations associated with each. It also describes a variety of other issues that arise when test results are used to hold individual educators or schools accountable.

Accountability Indices

The use of test scores for teacher or school accountability requires the construction of an aggregate measure, such as an average score for all of the students in a teacher's class or in an entire school. A variety of methods exist for creating these aggregates, or performance indices, but most of them fall into one of several types. This section describes the major types of indices and the relative benefits and limitations of each. For simplicity, the text primarily refers to schools as the units of accountability; however, most of the points are relevant to other units such as classrooms.

Status Measures

Perhaps the most straightforward approach to assigning a single number to a school is to calculate a simple average or other aggregate of scores collected at the same point in time, such as the percentage of students scoring at or above the proficient level from a spring test administration. These are often called status measures to distinguish them from measures based on change or growth. These single-point-in-time measures are typically intended to provide information about the current performance of a school's students or to compare a school's performance against an established target. The calculation of these measures is generally simple, although the need to combine information across different tests, subjects, or grade levels sometimes leads to the adoption of fairly complex rules for constructing these measures. The data demands for most status measures are minimal; for example, they do not require the linking of individual student scores over time, although they do require rules for determining

which students to include in the calculations when students transfer schools midyear. However, the inferences that these measures support are limited. Because they do not adjust for students' prior achievement, they provide limited information to inform inferences about the effectiveness of education a school provides. Their use in TBA systems has sparked intensive debate about the appropriateness of holding schools accountable for students' performance without recognizing the other factors contributing to their achievement.

There are some measurement issues that arise in the calculation of status measures. If scores need to be combined across different grade levels, they need to be placed on a comparable scale. Norm-referenced scores such as normal curve equivalents (NCEs) or percentile ranks are often used for this purpose, and, more recently, the percent-above-cut metric (e.g., the percentage of students scoring at the proficient level or above) has become widespread. Although these types of scales provide a straightforward way to combine results from different grades, they are likely to mask some differences in the meanings of scores across grades. For example, the proficient cut score might not be equally rigorous in every grade. Similar issues arise when combining results across different subjects. Although these measurement issues are important to address, arguably, the most significant drawback of status measures is the absence of adjustment for student characteristics including prior achievement.

Successive Cohort and Quasi-Longitudinal Measures

One way to address the variability in the prior achievement of students served by different schools is to compare the current performance of a school's students with the performance of the students the school served in previous years. An increase in the scores from one year to the next is often interpreted as indicating that a school is improving its effectiveness. This interpretation, however, requires an assumption that the background characteristics of the students, including their prior achievement, remain constant over time. Research on successive cohort measures has shown that substantial year-to-year variability in student characteristics leads to noisy and inaccurate estimates of schools' effectiveness. Another drawback of these measures is that they do not directly measure students' growth in achievement and can provide misleading information about schools' performance. For example, a school that achieves higher-than-average gains in individual students' achievement from the entry year to the exit year, and that does so consistently year after year, would attain a gain of zero on a successive cohort change index. Therefore, these measures cannot support strong inferences about school or teacher effectiveness.

One approach that is sometimes used to address variability in student characteristics across cohorts is to follow an individual cohort over time, such as by comparing the scores of a school's fourth graders with the prior year scores for that school's third graders. These measures, sometimes called quasi-longitudinal change scores, typically require a testing program that uses a vertical scale, which is a scale that links test scores across grades to reflect their increasing difficulty. When a vertical scale is used, differences in scores for consecutive grades can be interpreted as indicating the amount of achievement growth attained over the course of the year, although there are a number of psychometric challenges associated with creating valid vertical scales.

Quasi-longitudinal measures address some of the sampling and attribution problems that are inherent in successive cohort measures, but they are not immune from them, especially when student mobility is high. Both successive-cohort and quasi-longitudinal measures are appealing because of they are relatively simple to calculate and, like status measures, do not require data systems that follow individual students over time. However, their inability to provide information about individual student gains makes them less appropriate for inclusion in TBA systems that impose high stakes.

Growth or Value-Added Measures

Approaches that follow individual students over time are generally called growth models or measures. They include simple change-score models, in which each student's prior-year score is subtracted from that student's current-year score and the change scores for all students in a school are aggregated to produce a school-wide measure. They also include more complex, value-added measures. All of them have the advantage of directly examining individual students' test-score growth over time rather than relying on group averages of single-point-in-time measures. Several categories of growth measures are discussed briefly here.

Average gain score

The simplest growth measure is an average of individual-student gain scores that involves subtracting a student's prior-year score from his or her current-year score. This measure is easy to calculate and straightforward to explain to stakeholders. However, gain scores have some properties that make them less than ideal for use in TBA systems. They are often correlated with student background characteristics; simply subtracting out prior achievement does not adequately control for student background. They rely on an assumption of linearity in the test-score scale, which may not be warranted. In addition, students with missing data in either year are

excluded, and these scores can only be calculated when there is a vertical or developmental scale on the test.

A modification of the simple gain score is an average residual. This measure involves fitting a simple linear model that regresses current-year score on prior-year score, calculating the residual from this model for each student, and obtaining the average of these residuals for the whole class or school. This approach may be modified to include other covariates, such as student background characteristics. Like the simple gain score, this type of index is relatively transparent and easy to implement, but it may be affected by measurement error, which can limit its ability to remove the effects of student background variables from the estimated effects. Moreover, it cannot be used for students with a missing score in the prior year.

Value tables

When the achievement test is scored according to performance levels (e.g., basic, proficient, and advanced) and when there is a desire for transparency, the value table approach may be desirable. Value tables assign points based on students' movement across performance levels. The set of rules embodied in the table is applied to each student, and the points for students are aggregated to the appropriate level (teacher or school). A simple example is provided in **Figure 1**.

Figure 1 illustrates how value tables can be used to communicate priorities and create incentives to focus educators' behaviors. Schools receive points when students move from their current performance level to a higher level, and they receive more points for students who move more than one level. They also receive more points for moves at the lower end of the achievement distribution than at the higher end, a rule that might be designed to promote extra focus on low-achieving students. Schools receive points for keeping students at their current level only when those students are already scoring proficient or above. This table could be altered in a variety of ways, such as by assigning negative values to moves from one level to a lower level or assigning a small number of positive points for students who fall from advanced to proficient to allow for regression to the mean due to measurement error in the test.

As noted above, value tables may be desirable in situations for which transparency is particularly important,

Year 1 achievement level	Year 2 achievement level			
	1	2	3	4
1 – Below basic	0	150	225	300
2 – Basic	0	0	125	200
3 – Proficient	0	0	75	150
4 – Advanced	0	0	0	75

Figure 1 Sample value table.

such as when test scores are used to determine individual teachers' salaries or bonuses. They also provide a way for the designers of the TBA system to communicate their priorities and assign rewards or penalties in a way that is consistent with those priorities. However, value tables have some drawbacks. Like the gain scores discussed above, they only include students who have scores at both relevant time points. Moreover, the rules for assigning points, no matter how carefully considered, are essentially arbitrary and there is typically little effort to explore the sensitivity of decisions to the specific numbers used. Limited empirical studies of the estimators have shown that they can be sensitive to student background characteristics and are unlikely to fully remove the confounding effects of these variables from performance measures. Because value tables rely on performance levels, information about changes within levels is lost, and the measures are affected by other technical issues that arise in the context of standards-based reporting of scores.

Random-effects models

This class of models includes a variety of approaches to modeling the joint distribution of test scores for each student. One example is the layered model used at one time by the Tennessee Value-Added Assessment System (TVAAS), one of the first and most prominent value-added models used in the United States. These models address many of the shortcomings of the other approaches discussed in this section. They allow students who have incomplete test-score data to be included. Also, some of these models can be used with a variety of test-score scales, unlike the gain score models which require a vertical scale. Extensive research has been conducted to explore the properties of these models in the context of estimating teacher effects, and has shown that if enough scores are included for each student (the minimum number is generally considered to be three), they can remove most of the effects of individual student characteristics from estimates of teacher performance. One of the key criteria is that the multiple measures from students are linked with different teachers and that there are no isolated groups of teachers with any students in common. These conditions are less likely to hold for schools than for teachers, and thus the desirable properties of these models maybe less likely to apply to estimates of school performance. The models can be extended in a number of ways, such as by incorporating terms that allow teacher or school effects to diminish over time. Significant drawbacks of this class of models are their complexity and lack of transparency.

Student fixed-effects models

Finally, student fixed effects may be included in models as a way to address confounding that results from individual student characteristics. The dependent variable in these

models is often a gain score, but a single test score may also be used. The model used in estimating school or teacher effects includes an indicator variable (or fixed effect) for each individual student in the data as well as indicators for teachers or schools. By including students fixed effects in the model, the school effects are estimated by variations in outcomes within students, so that persistent effects of individual students do not contribute to the estimated performance measures. Consequently, only students who switch schools contribute to estimates of school effects that are not grade- or year-specific. This restriction of the sample to students who switch schools can greatly reduce the data available for estimating school performance and has been cited as one of the limitations of fixed-effects methods. These models are also challenging to implement on a large scale, are not transparent, and typically require vertically scaled scores.

Challenges in using growth models for TBA

All of the models discussed in this section have clear advantages over the status and cohort-based measures presented earlier. The more-complex models tend to support more valid inferences about teacher or school contributions to achievement than do the simpler models, although no model completely isolates these effects in all cases. Moreover, for some accountability purposes, the need for transparency might lead to a decision to use a less-complex model; trade-offs between transparency and validity must be considered.

All of these measures require longitudinally linked, student-level data, with multiple measures per student and with links between the student's test score and the appropriate unit of accountability (teacher or school). Even when the basic infrastructure is in place to support the collection of this type of data, there are challenges associated with ensuring its accuracy. Determining which teacher or school should receive attribution for a particular score is especially difficult when students move or when more than one school or teacher provide instruction in a particular subject. Clear and consistent rules must be developed to deal with these problems, and applying these rules can be time consuming and expensive. Those responsible for designing TBA systems must decide whether these costs are justified by the quality of the resulting information.

Measurement Issues Raised by Test Use in School Accountability

Regardless of the specific model used, policies that hold schools or educators accountable for student test scores raise a number of issues that need to be considered and addressed. The accountability index is only as good as the

testing program that underlies it, and existing large-scale testing systems face several challenges that are likely to affect their validity. This section does not provide a comprehensive discussion of these issues, but briefly discusses some of the more salient ones.

Test Preparation and Score Inflation

The research on high-stakes testing points to two interrelated concerns that arise when consequences are related to test scores. First, scores tend to become inflated, which means that they overstate the actual level of achievement the test score is intended to represent. This inflation probably occurs in large part as a result of the second finding, which is that teachers and other educators often respond to high-stakes tests by increasing their focus on tested material and decreasing their focus on material not included in the test. Teachers, for example, have responded by reallocating instructional time from non-tested subjects to tested subjects and from nontested topics to tested topics within a subject area. Teachers also tend to focus more on specific item formats and styles, and on test-preparation activities. The use of performance levels in accountability systems has been associated with an increased emphasis on students who are performing near the cut point. Although less common, cheating and other clearly negative responses can also occur. Many of these responses may be mitigated through the design of TBA systems – for example, by including a variety of item formats, considering gains at all points in the score distribution, and by instituting security measures that reduce the likelihood of narrow test preparation or cheating.

Inclusion of All Students

TBA systems are usually intended to include accurate measures of performance for all students served by a school or teacher. A number of policy decisions around inclusion need to be made, such as how to treat students who attend a school for only a portion of the year. Perhaps the most challenging issues, however, pertain to students with disabilities and students whose first language is different from the language in which the test is written. Many TBA systems have explicit policies addressing the inclusion of these groups. For instance, rules may specify the percentage of students with disabilities who may be excluded from regular testing and given alternative assessments or the number of years in which a student must have received instruction in the tested language to be included in testing.

The validity of the entire TBA system will be affected by the validity of scores for students in these groups. Accommodations or modifications to testing are sometimes

made to enhance test validity, but some research suggests that these changes often lead to inaccurate inferences about student achievement.

Alignment

The theory of action behind many of today's TBA systems is that published content standards will communicate the goals students are expected to achieve, and the tests will accurately measure student attainment of these standards. In addition, there is generally an assumption that schools will adopt curriculum programs that provide students with opportunities to learn the material embodied in the tests and standards. This theory requires that tests, standards, and curriculum be aligned with one another. A number of approaches to measuring alignment have been developed, and most jurisdictions that adopt TBA systems conduct an alignment analysis. The methods for evaluating alignment are imperfect; however, in many cases, they have shown that while each test item may be found to be aligned with one or more standards, the tests often fail to capture the breadth and depth that the standards were intended to convey. Higher order reasoning and problem solving are especially likely to be neglected, perhaps in part because of the difficulty of developing test items that tap these skills and that can be administered and scored in a cost-effective way. Aligning tests with curriculum is even more challenging, especially when schools or districts have latitude over the curriculum they adopt and when there is extensive local variation in curriculum.

Reporting in Terms of Performance Levels

The final issue addressed here stems from the nature of reporting that is often used in accountability systems. TBA systems are frequently built on the assumption that all students should be able to achieve a specific level of performance, and that tests should be scored in a way that clearly communicates what knowledge and skills the student has demonstrated. In many cases, these considerations have led to decisions to report according to performance levels such as basic, proficient, and advanced. Each level is associated with a specific cut score on a test. NCLB, for example, requires states to report the percentage of students exceeding the proficient threshold. States must also develop other levels such as basic and advanced, but accountability decisions are based exclusively on whether or not students score at or above proficient.

The use of performance levels is intended to promote better communication, but there is little evidence that most stakeholders truly understand what it means to be classified as proficient or advanced. Moreover, in the United States, where states under NCLB have been permitted to set their own cut scores, the meaning of proficient varies widely across states. Several methods have

been developed for setting these cut scores, but all of them rely heavily on professional judgment, and different methods often result in different cut scores.

Moreover, because these types of scores only capture differences in performance when those differences cross one of the thresholds, they can provide misleading information about student performance. Many of the categories cover a broad range of test scores; a student scoring at the high end of the proficient range may be more similar to an advanced student than to a student scoring at the low end of that category. The use of performance levels is especially problematic for the interpretation of test score gains or group differences – whether the gap between two racial/ethnic groups increases, decreases, or remains constant over time, for example, sometimes depends on where the cut score is set. These drawbacks suggest that caution is warranted when interpreting results based on performance levels.

Summary

Scores on student achievement tests are frequently used to hold educators accountable. The use of test scores in this way assumes that those scores provide reasonably valid information about the quality of education teachers or other school staff provide. A number of methods for creating accountability measures from test scores were described, and differences in the extent to which they support valid inferences about school or teacher effectiveness were discussed. All of the methods presented above, however, are subject to measurement issues that are inherent in large-scale achievement testing. Some of these were discussed in the previous section, but this is by no means a comprehensive list. The use of test scores in accountability systems must be accompanied by careful considerations of threats to validity, and by efforts to minimize those threats to the extent possible.

Many of the concerns raised earlier can be mitigated, although probably not eliminated completely, through careful design of TBA systems. Several sets of guidelines for appropriate test use have been published; the best known is probably the *Standards for Educational and Psychological Testing* published in 1999 by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. Although the standards do not focus exclusively on accountability testing, many of the guidelines included in that document are relevant. Perhaps the most important advice found in this document is that a single test score should not be used to make high-stakes decisions about individuals. The meaning of this seems fairly obvious in the case of tests with stakes for individual students (though even in that case the implications are widely debated); however, it also pertains to accountability decisions that rest on an aggregate score. The standard implies the need for a

compensatory system, in which low scores on one measure can be offset by good performance on other criteria, but that has not always been followed. Providing examinees with multiple opportunities to take the test (or equivalent forms of it) is consistent with the standards but is not generally considered sufficient to meet the requirement for multiple measures.

The use of tests in accountability systems should be subjected to ongoing monitoring that documents the responses of stakeholders to the system. Whether TBA leads to desirable or undesirable responses will depend in large part on the extent to which educators support the system and believe that it provides accurate information about their efforts and effectiveness. Future developments in value-added methodology and in assessment design will lead to new policies and practices surrounding TBA.

See also: Educational Measurement: Overview; Impact of Assessment on Classroom Practice; Validity of Achievement Gains on High-Stakes Tests.

Further Reading

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association/American Psychological Association/National Council on Measurement in Education.
- Baker, E. L., Linn, R. L., Herman, J. L., and Koretz, D. (2002). *Standards for Educational Accountability Systems (Policy Brief #5)*. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing.
- Center on Education Policy (2006). *From the Capitol to the Classroom: Year 4 of the No Child Left Behind Act*. Washington, DC: Center on Education Policy.
- Goldschmidt, P., Roschewski, P., Choi, K., et al. (2005). *Policymakers' Guide to Growth Models for School Accountability: How Do Accountability Models Differ?* Washington, DC: Council of Chief State School Officers.
- Hamilton, L. S. (2003). Assessment as a policy tool. *Review of Research in Education* **27**, 25–68.
- Herman, J. (2005). Making accountability work to improve student learning. *CSE Technical Report 649*. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing.
- Heubert, J. P. and Hauser, R. M. (1999). *High Stakes: Testing for Tracking, Promotion, and Graduation*. Washington, DC: Committee on Appropriate Test Use/National Research Council/National Academies Press.
- Koretz, D. (2003). Attempting to discern the effects of the NCLB accountability provisions on learning. *Paper Presented at the Annual Meeting of the American Educational Research Association*, Chicago.
- Koretz, D. and Hamilton, L. S. (2006). Testing for accountability in K-12. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn, pp 531–578. Westport, CT: American Council on Education/Praeger.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher* **29**(2), 4–16.
- Linn, R. L. (2003). Performance standards: Utility for different uses of assessments. *Education Policy Analysis Archives* **11**(31). <http://epaa.asu.edu/epaa/v11n31> (accessed July 2009).
- Lockwood, J. R., McCaffrey, D. F., Mariano, L. T., and Setodji, C. M. (2007). Bayesian methods for scalable multivariate value-added assessment. *Journal of Educational and Behavioral Statistics* **32**(2), 125–150.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., and Hamilton, L. S. (2003). *Evaluating Value Added Models for Teacher Accountability. MG-158-EDU*. Santa Monica, CA: RAND.
- Rothman, R., Slattery, J. B., Vranek, J. L., and Resnick, L. B. (2002). Benchmarking and alignment of standards and testing. *CSE Technical Report 566*. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing.
- Sanders, W. L. (2006). Comparisons among various educational assessment value-added models. *National Value-Added Conference*, Columbus, Ohio.

Test Development

S M Downing, University of Illinois at Chicago, Chicago, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Test development or test construction refers to the science and art of planning, preparing, administering, scoring, statistically analyzing, and reporting results of tests. This article emphasizes a systematic process used to develop tests in order to maximize validity evidence for scores resulting from those tests.

The point of view of this article is large-scale, high-stakes cognitive testing, scored using classical measurement theory (CMT), which also provides an adequate model for most classroom tests. The selected-response test item format, exemplified by the prototypic multiple-choice item, is an excellent format for testing student cognitive achievement at most levels of schooling and at all levels of the cognitive taxonomy. The 12-step process model is also useful for other testing formats, including performance-type tests using item formats such as the constructed-response form, simulations, structured oral examinations, and all other blended or combination formats.

This test development process model applies equally well to nearly all types of cognitive tests, including those intended to measure achievement or ability, tests used for academic or job placement, tests intended to diagnose learner's strengths and limitations, tests intended to provide learners with formative feedback on progress as well as a final summative measure of cognitive achievement in a domain, and occupational-type tests used to credentialize individuals for a profession or occupation.

The emphasis on some steps will depend on the exact purpose of the test being developed. Some steps or tasks may be subsumed into other tasks or dismissed as unnecessary for a particular testing situation. This process model, summarized in **Table 1**, may be reordered or reorganized, given special requirements for the proposed test, but most tasks will be required to be carried out for most tests, whatever format is used.

Process Model: 12 Steps for Effective Test Construction

This model emphasizes documenting validity evidence for test score data. All test scores require validity evidence to support or refute intended interpretations. In general, the higher the stakes or consequences associated with test scores, the more the sources of scientific validity evidence

required. However, multiple sources of scientific validity evidence are desirable for all test score data.

Step 1: Overall Test Plan

The first planning stage is important since many decisions made at the outset will have an ultimate impact on the quality of the test and the validity evidence for scores resulting from the test.

A clear statement of the proposed purpose of the test scores is critical to the validity of test scores. This statement of purpose determines the hypothesized construct intended to be measured. All validity documentation will seek to refute or support this hypothesis concerning the construct thought to be measured.

The purpose statement may be as simple and narrow as: "...to measure student achievement on the content of instructional objectives in biochemistry, units 2 to 4." Or, as comprehensive as: "...to measure the summative cognitive achievement, across all subspecialties of internal medicine, of physicians in post-graduate medical training in internal medicine, year 2." The statement of clear intent or purpose for using test scores is essential to guide all development activities and suggest sources of validity evidence that will be needed. The test's purpose statement and hypothesized construct, if clearly stated, will assist test developers throughout the cycle in determining which process steps are most essential, when each step must be accomplished, the logical order of the steps, what level of sophistication is required, how the process might best be documented, and so on.

This first planning stage requires agreed-upon timelines, especially for group test development projects, as well as role and task designations. If there is a fixed deadline for test administration, working back from that date will help provide a realistic timeline for accomplishing all the essential tasks.

The complexity of the proposed test and the consequences of the resulting test scores will generally determine the length of the test development timeline, the number and qualifications of staff or personnel required, and so on. Large-scale, high-stakes test development projects often take 1–2 years to complete and require the input of staff with many different capabilities – subject matter experts (SMEs), test development specialists, test editors, psychometricians, production specialists, information technology specialists, and so on. Classroom-type tests, on the other hand, are usually developed by a

Table 1 Twelve-step process model for test development

Steps	Example test-development tasks
1. Complete plan	Systematic plan to guide all other test-development activities: clear statement of test's purpose and definition of the construct; desired test-score interpretations or inferences; planned validity evidence; rationale for test format(s); psychometric model; timelines; security; quality control plan
2. Definition of content	Exact sampling plan relating items to domain; rational or empirical foundation; major source of content-related validity evidence
3. Test specifications	Content operationally defined and specifically related to Step 2; documents content definition; provides content-related validity evidence for score inferences to domain
4. Developing items	Creation of effective test stimuli or item formats; principles of effective item writing; item writer training; content and editorial review to reduce bias
5. Test design and assembly	Designing and operationalizing test forms; item or prompt selection; operational sampling by planned blueprint; pretesting, item tryout
6. Producing test forms	Publishing or producing tests; security, validity, quality control issues
7. Administration	Validity issues concerned with standardization; fairness issues; proctoring, security, timing issues
8. Response scoring	Validity issues; psychometric model; quality control; key validation; item analysis
9. Establishing passing scores	Establishing defensible cut or passing scores; relative vs. absolute methods; validity issues; comparability of standards and constancy of score scale
10. Reporting test results	Validity issues: accuracy, quality control; timely; misuse issues
11. Item banking	Effective item banking; security; flexibility
12. Documentation – technical report	Process documentation: thorough, detailed documentation for validity evidence

Reproduced (in modified form) from Downing, S. M. (2006). Twelve steps for effective test development. In Downing, S. M. and Haladyna, T. M. (eds). *Handbook of Test Development*, pp 3–25. Mahwah, NJ: Lawrence Erlbaum Associates, Copyright 2006, with permission of Lawrence Erlbaum Associates (Taylor & Francis Group, LLC).

teacher or a small group of instructors, with far fewer resources and time but smaller-scale test projects also benefit from an overall planning step.

Quality control issues should also be addressed with a comprehensive plan for assuring accuracy and quality at every stage of test development. The ultimate success of all test development processes is determined by the quality and quantity of the validity evidence for the test scores. This first comprehensive planning step helps assure a successful outcome.

Step 2: Definition of Content

Test content is the single most important feature for most tests, especially achievement tests. Content-related validity evidence is an important type of validity evidence required for achievement tests of all kinds. The manner in which the content to test is determined is key to this type of validity evidence.

Many different methods are used to delineate the content for tests. Generally, these methods use empirical, rational, or a combination of empirical and rational methods.

For many high-stakes tests, such as most credentialing examinations, empirical techniques, generally referred to as practice analysis, task analysis, or job analysis, are used to establish the appropriate content to sample in the test, which in turn serves as a major foundation for content-related validity. These empirical procedures typically employ survey research methods to systematically

sample the opinions of qualified individuals who are knowledgeable about the content domain. After empirical data are collected and analyzed, content-expert decision makers use the results of these studies to formulate an exact sampling plan that reflects the relative importance of various content areas and their criticality for safe or competent practice in the discipline.

Empirical and rational methods are combined to delineate test content for some tests. For example, accountability tests (state-wide assessments), which may be used to evaluate student progress in mastery of essential learning objectives or skills and to evaluate school success, often define content through complex methods using both empirical data from stakeholders (students, parents, teachers, administrators) and rational methods from SMEs in order to arrive at group consensus judgments about the appropriate content to test.

Rational content-defining methods are used for almost all lower-stakes tests. These methods require judgments by SMEs as to the appropriate content to sample and the relative emphasis of the content for measurement of the construct of interest given the purpose of the test. For most classroom achievement tests, content decisions are made by the teacher or instructors, based on the instructional objectives, the teaching/learning experiences, and subjective judgments about the relative importance of topics and content. At minimum, such rational content-defining methods should strive to be unbiased and yield test content that fairly represents instruction and student learning expectations.

The process of content definition is fundamental to test validity. The construct of interest begins to be operationalized through systematic efforts to delineate the exact content domain or universe. The ultimate success of the test as a measure of the construct depends greatly on the soundness of the methods used to define content.

Step 3: Test Specifications and Blueprint

This step specifies all the essential details of the proposed test and all the processes to be used to develop that test. This test-specifications step guides the remaining test-development steps and provides essential documentation and rationale as validity evidence. One important component of this step is to outline the sampling plan which serves as a blueprint or detailed guide for test developers to assure a one-to-one correspondence of the test stimuli or items to the total content domain, universe, or population.

Test stimuli, items, or performance prompts are samples – theoretically random samples from an infinite number of all possible stimuli – used to make inferences to the population or total domain of content. Examinee performance on the particular test items or performance prompts of any given test is of interest primarily to provide a scientifically sound basis for inference to the whole content domain. Thus, the logical basis for these inferences – and a fundamental aspect of validity evidence – relies on the reasonableness of the sampling plan for items and item content.

Test blueprint

The test blueprint operationalizes a sampling plan and ensures that the test is a representative, reasonable, and fair sample of the content domain. Many different procedures can be used to develop this blueprint or sampling plan. For achievement tests, especially classroom tests, a simple matrix of item content by cognitive level may be sufficient. This matrix crosses content topics or areas (to whatever level of specificity desired), with cognitive process levels (such as recall or recognition of facts; application of facts, concepts, and principles; problem solving and other higher-order cognitive processes) and specifies the proportion of items in each cell of the matrix. This type of blueprint is typically prepared by the instructors, based on subjective but informed and rational judgment, and forms the basis for item or prompt development. Blueprinting of content ensures systematic, comprehensive, and representative sampling of the domain, at cognitive levels judged to be appropriate, and functionally aligns the test with instruction. Fair, unbiased, and reasonable blueprinting of content provides a major source of content-related validity evidence, especially when independently verified after the test has been developed.

Large-scale test development programs may use more sophisticated and more complex methods to blueprint the test content and relate the content of test items to the domain. At minimum, test blueprints or other types of exact sampling plans for high-stakes tests require some consensus of representative groups of SMEs.

Test specifications

Every important detail of the test must be specified, together with a rationale for all critical choices. For example, for a cognitive achievement test, what item formats (constructed-response, selected-response, both) will be used? What is the rationale? If selected-response items are to be used, what type (multiple-choice, matching, extended matching, multiple true–false, and so on). Why? How many items will be used on the test and what is the justification for the number of items? Who will create the test items? What is the rationale for the selection of the item or prompt writers? What training will the item or prompt writers receive? How will the test be administered (performance, paper-pencil, oral, computer-based)? What is the rationale?

Each of these 12 steps of effective test development has many individual decision points, each with a myriad of details to be decided. Each test development choice-point is an opportunity to enhance or lessen validity evidence.

Step 4: Item or Prompt Development

Creating effective test items or performance prompts is a challenging task. Writing effective items or prompts is a major test-development task, requiring application of principles based on many years of research; this task also requires specific training, practice, experience, and feedback.

The principles of writing effective selected-response items, summarized here, are well documented in the research literature. Most of these principles may apply to creating performance prompts as well.

Principles of effective items

Well-written and effective selected-response items have the following characteristics. Items focus on a single important idea, concept, or principle asking a specific question at a cognitive level that is aligned to the instructional activities of the learners. Item wording is unambiguous and well edited for content accuracy, language clarity, and style. The stem or lead-in of the item contains most of the information and both the stem and the answer options are worded positively, avoiding negative words. A direct question is usually best. Trick questions and questions dealing with trivial information should be avoided.

The answer options should contain one and only one correct or best answer to the stem question, and most experts must agree that this is the best or most correct answer. Answer options should be homogeneous in content such that all options represent the same general class. Effective selected-response items use as many total options as necessary, but three options (one correct and two plausible incorrect) are usually sufficient. All incorrect answer options must be plausible such that some uninformed examinees choose these distractors. Avoid using *all-of-the-above* and *none-of-the-above* answer options, as well as specific determiner words such as always, never, and all, which clue an incorrect answer choice. There should be no grammatical cues to the correct answer.

Some common item writing problems: Item flaws

Some item writers attempt to test too much material in a single item. Usually, an item that focuses clearly on a single idea works better than an item that is overly complex in its form. Generally, the straightforward single-best answer multiple-choice item works best, so that complex formats – such as formats that require multiple answers or complex answer choices using combinations of options – should be avoided. Open-ended, nonfocused item stems such as, *which of the following statements is true (or not true)?* are usually followed by a set of answer options that require the examinee to answer true or false to each separate option. While the multiple true–false item form is a legitimate format in which each option is separately scored as true or false, it is not effective to use this form for multiple-choice items with one correct or best answer.

Specialized training of item writers, with practice and constructive feedback, is often lacking. Instructors at all levels of teaching frequently lack the specialized training, experience, and practice needed to become effective item writers. Furthermore, teachers often lack sufficient time to prepare effective test items and lack the resources of larger-scale testing programs to carry out the content and editorial reviews which tend to greatly improve the overall quality of newly created test stimuli.

Quality control and security controls for items must be maintained throughout the item-writing stage. Content and editorial review and rewriting is an important means to improve the quality and effectiveness of test items.

Step 5: Test Design and Assembly and Step 6: Producing Test Forms

For most classroom tests, assembling the items into a final form of the test is somewhat trivial. For large-scale tests, with complex multiple-form designs, which may be computer delivered, the final assembly step is far more complex and fraught with quality control challenges. For

all types of written tests, there are validity issues associated with test design, assembly, and production. For classroom-type tests, producing test forms is straightforward, usually involving simple reproduction methods under the direct control of the instructor. For high-stakes, large-scale test programs, production of printed or computer-based tests can be extremely complex.

Quality control of the final test product is the most important issue at these steps. The design of the test and the items selected for the final form of the test must be a representative sample of the content so that these particular items can generalize to the domain of interest.

Pretesting or item tryout is a concern at the test design and assembly step. Large-scale tests typically imbed newly created items on final test forms in order to evaluate the statistical performance of new items. A variety of pretest or tryout designs may be used, but in all cases pretested items should comprise a relatively small proportion of the total test. Classroom tests generally cannot pretest items because of logistic and security issues, so all newly created test materials must be evaluated for effectiveness at the first test use, making key validation (see Step 8) all the more critical.

For high-stakes tests, security issues abound at this stage of development since extremely secure test items or performance prompts may be out of the direct control of the test developers for the first time in the test-development cycle. Any compromise of quality or security at this stage can spoil the entire test-development effort by invalidating some or all of the test scores. Careful documentation of the chain of security must be maintained for high-stakes tests, including secure transport or shipping of all materials to and from production sources or the security controls for computer-based tests.

If there are production-type errors or problems, such that the final test is difficult or impossible to read or use, the validity of the scores resulting from the test may be compromised. Thus, production issues are validity issues.

Step 7: Test Administration

Administration represents the culmination of all test-development steps up to this point. The items or prompts developed to sample the domain are presented as stimuli to examinees in such a way as to control or standardize as many extraneous variables as possible.

Major validity concerns about test administration revolve around security and standardization issues such as:

- proctoring for standard test administration conditions;
- reduction or elimination of test irregularities, such as cheating;
- timing issues such that all examinees have identical amounts of time; and

- security of all test materials before, during, and after the test.

Lax, insecure, or unstandardized test-administration conditions can invalidate score interpretations for some or all examinees and spoil all the test-development work.

Step 8: Response Scoring

Scoring test responses is the starting point for many psychometric concerns and validity issues. CMT models are used for most classroom-type tests. Item response theory (IRT) models are applied to many large-scale tests. The choice of psychometric model has some impact on test development, but this impact may be minor except for specific test design and data collection methods. Whatever psychometric model is used for scoring, examinee responses are compared to the scoring key and typically awarded one point if the response agrees exactly with the scoring key (or rubric) and zero points if the response does not match. Other scoring systems can be used, but right–wrong (0, 1) scoring is the most parsimonious and the starting point for almost all scoring methods.

Item analysis

Item analysis is a process of counting examinee responses to test items and compiling statistical data on the number of examinees who answer the item correctly in order to evaluate the difficulty and the discrimination of the test item. Item discrimination refers to how sharply the item differentiates between examinees who obtain high scores on the test and those who have low scores on the test and is usually estimated as a correlation between the item score and the total test score.

The purpose of item analysis is to evaluate the performance or functioning of test items or prompts. Such data can be used for several purposes including the improvement of future tests and the enhancement of test validity evidence through the elimination of poorly performing test items. Item analysis data can be compiled for all types of tests, including performance tests.

Item analysis data is also used to evaluate the patterns of responses to the correct and incorrect options of the item. Option item analysis permits the evaluation of the functioning of both correct and incorrect options.

For most achievement tests, the ideal selected-response test item will be of middle or medium difficulty (about 55–75% correct) since such an item provides maximum information about examinee achievement. Also, ideal items will sharply discriminate high- and low-scoring examinees, having high positive-correlation coefficients with the total score on the test.

Key validation

For test development, one of the major uses of item analysis is for key validation of test items prior to their final scoring. Key validation refers to the process of performing a preliminary item analysis on tests in order to evaluate the performance of test items prior to final scoring. Key validation allows the test developer to eliminate items that are problematic and may be invalid before students receive final scores on the test. Key validation is especially valuable for tests with many newly written items which have not been pretested.

Step 9: Establishing Passing Standards

Many tests require assignment of a pass or fail status from the test scores or some other classification system, such as grades. All methods used to establish pass or fail cut scores on tests require judgment and are therefore somewhat arbitrary and subjective. Passing scores answer the question: How much is enough to know in order to be classified as passing the test? All passing scores represent policy decisions, but should be informed by unbiased expert judgments about the content tested.

There are three major methods commonly used to establish passing scores on tests: normative, absolute, or combination methods (**Table 2**). Normative passing score methods use the relative ranking of examinee scores to determine the cut score or grade to be assigned. In normative methods, the relative location of the examinee's score in the distribution of scores is the key feature. Absolute methods use the judgment of experts concerning the subject-matter content of the test items to establish the passing score. Combination methods have some normative and some absolute characteristics.

Absolute passing score methods are widely used, especially for high-stakes tests. Such methods emphasize examinee competency and directly address the issue of what the examinee knows or can do. The normative methods, in their pure form, address only the issue of relative standing on whatever is measured by the test, not what content has been mastered.

Table 2 Examples of passing score methods

<i>Method</i>	<i>Judgment</i>	<i>General type</i>
Angoff	Items	Absolute, item centered
Ebel	Items and relevance	Absolute, item centered
Hofstee	Test	Blended: normative/absolute
Contrasting groups	Examinee performance	Examinee centered
Borderline group	Examinee performance	Examinee centered

For large-scale tests using absolute passing scores, statistical equating of test scores across different forms or administrations of the test is usually required to maintain the identical meaning or interpretation of the passing score. Most tests are slightly harder or easier across forms or administrations. Equating levels these slight differences in difficulty and maintains the constancy of the score scale over time so that the passing score has the identical interpretation.

For many classroom-type achievement tests, passing scores or grading scales are determined administratively by policy outside the direct control of the teacher or instructor. Ideally, these policy decisions are based on some types of informed, unbiased judgments concerning the content tested, the nature and quality of the instruction and the tests, and the actual achievement of the students.

Step 10: Reporting Test Results

For many classroom tests, reporting the results of tests is straightforward. Feedback on student performance may be limited to a total test score, with information on the grade or passing status associated with the score. Feedback that is instructionally useful may be an important feature of classroom achievement test score reporting. For high-stakes tests, the reporting of test results may be more complex and detailed.

Generally, the reporting of all types of test scores should be accurate, timely, clear, and meaningful to examinees and other users of the information, and, if possible, should provide specific useful feedback on performance strengths and limitations. If diagnostic-type feedback is provided, the sample of items or behavior upon which subscale performance is based must be sufficiently large to be reliable. Some indicant of the errors of measurement inherent in the scores should be provided.

The score report should clarify legitimate uses of scores and indicate what constitutes a misuse of scores on the test.

Step 11: Item Banking

Item banking refers to the process of filing or storing test items or prompts, together with identifying performance information about them, such that such test materials can be reused on future tests. Security, usability, and flexibility of item-classification systems are key elements of a successful item-banking system. For classroom-type tests, item banking may consist of a secure paper-filing system or secure computer files containing the test materials and minimal identification and item/prompt performance information.

Large-scale testing programs usually require sophisticated computer-based item-banking systems, which maintain large numbers of test items, together with many identification and performance variables about the item. Such systems allow maximum flexibility in storing and retrieving test items/prompts to meet all the content requirements and item-characteristic requirements specified by the testing plan and specific blueprint.

Desirable item banking characteristics include:

- security;
- ease of operation – both input of text and data and output of items/prompts;
- flexibility – the ability to change classification systems and details;
- strong data manipulation capabilities; and
- ability to use multiple item characteristics to sort and select items/prompts

Step 12: Summary Documentation – Technical Report

Step 12 is a summary step, both for all the previous test-construction processes and this article. Documentation of validity evidence is important for all tests. For high-stakes tests, validity evidence should be clearly and completely documented in a written technical report. Some examples of technical report content, organized around the 12-steps of test development include:

- purpose of the test and clear statement of how scores will be interpreted;
- how test content was determined;
- how those who prepared test stimuli were selected;
 - training, experience, demographics if relevant;
- how security was maintained throughout the test-development cycle;
- rationale for method of administration: paper-pencil or computer-based;
- complete information on test scoring;
 - rationale for psychometric model
 - statistics summarizing test results: mean difficulty, discrimination, passing rates
 - item-analysis results
 - information on pretesting
- documentation on equating methods used, if any; and
- documentation on the methods used and the results of any standard setting study

Test development, as process and methodology, consists of multiple well-defined steps or processes. In this article, the test-development process has been outlined in 12 major steps, all of which are generally required for nearly all tests. Test validity evidence is the key organizing principle for each of these steps.

Further Reading

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Cizek, G. J. (2006). Standard setting. In Downing, S. M. and Haladyna, T. M. (eds.) *Handbook of Test Development*, pp 225–258. Mahwah, NJ: Lawrence Erlbaum Associates.
- Downing, S. M. (2006). Selected-response item formats in test development. In Downing, S. M. and Haladyna, T. M. (eds.) *Handbook of Test Development*, pp 287–301. Mahwah, NJ: Lawrence Erlbaum Associates.
- Downing, S. M. (2006). Twelve steps for effective test development. In Downing, S. M. and Haladyna, T. M. (eds.) *Handbook of Test Development*, pp 3–25. Mahwah, NJ: Lawrence Erlbaum Associates.
- Downing, S. M. and Haladyna, T. M. (eds.) (2006). *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Downing, S. M., Tekian, A., and Yudkowsky, R. (2006). Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. *Teaching and Learning in Medicine* **18**(1), 50–57.
- Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items*, 3rd edn. Mahwah, NJ: Lawrence Erlbaum Associates.
- Haladyna, T. M., Downing, S. M., and Rodríguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education* **15**(3), 309–334.
- Kane, M. T. (2006). Validation. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 17–64. New York: American Council on Education/Greenwood.
- Linn, R. L. and Miller, M. D. (2005). *Measurement and Assessment in Teaching*, 9th edn. Upper Saddle River, NJ: Pearson Prentice Hall.
- Messick, S. (1989). Validity. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn., pp 13–104. New York: American Council on Education/MacMillan.
- Millman, J. and Greene, J. (1989). The specification and development of tests of achievement and ability. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn., pp 335–366. New York: American Council on Education/MacMillan.
- Raymond, M. and Neustel, S. (2006). Determining the content of credentialing examinations. In Downing, S. M. and Haladyna, T. M. (eds.) *Handbook of Test Development*, pp 181–223. Mahwah, NJ: Lawrence Erlbaum Associates.
- Schmeiser, C. B. and Welch, C. J. (2006). Test development. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 307–353. New York: American Council on Education/Greenwood.

Test Translation and Adaptation

A Allalouf and P Hanani, National Institute for Testing and Evaluation, Jerusalem, Israel

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Test adaptations have become common in recent years for a variety of reasons. There have been an increasing number of international studies that deal with cross-national, cross-cultural, and cross-lingual comparisons of attitudes, traits, abilities, and educational achievements, for example, the Third International Mathematics and Science Study (TIMSS), which has been translated into 45 languages for use in 30 countries. There is a growing demand for adapted tests for use in selection and certification. In addition, there is a need for cross-lingual and cross-cultural personality and intelligence measures (e.g., the widely used Wechsler test that has been adapted and normed in many languages and countries). As a rule, tests should be administered in the examinee's native language, unless the aim is to test foreign- or second-language proficiency. If a valid and reliable test exists in one language, and there is a need for the same test in other languages, translating and adapting the original test makes good sense. The term test adaptation is preferable to test translation, as it includes all the activities needed to ensure that the intended meaning is retained across languages (Geisinger, 1994). The adaptation process should follow a detailed guide, which usually prescribes a series of consecutive stages such as double-translation, forward and backward translation, and independent reviews, so that the end result is in natural and fluent language. There should be a high degree of similarity in terms of psychometric characteristics between the translated items and the original items to ensure fairness in the measurement process. A fair and reliable comparison of test performance among examinees from different linguistic backgrounds is possible if several steps (i.e., differential item functioning (DIF) analysis between the different languages versions) are taken after the test is administered. These steps will enable score comparisons between the different language groups.

The Adaptation Process

Preparing a test that is constructed in one language and culture for use in a second language and culture requires an adaptation process that takes into account the linguistic and cultural differences between speakers of the two different languages. This process – sometimes referred to as transadaptation (Zucker *et al.*, 2005) – consists of accurately translating those items in the source language which can be adequately rendered in the target (translated) language,

and changing other items which, in simple translation, could lead to linguistically, culturally, or psychometrically inappropriate measurement. The translation and adaptation process should follow specified steps (see Hambleton and Patsula, 1999; Hambleton, 2005; Maxwell, 1996):

1. *Ensuring construct equivalence.* Persons familiar with both languages and cultures must agree that the same measured construct (i.e., content domain, attribute, etc.) exists both in the source language and culture and in the target language and culture.
2. *Determining the adaptation design.* There are two main standard designs used in producing an adaptation: forward-and-backward-translation and forward-only translation.

With the forward-and-backward-translation design, one or more translators adapt a test from the source language to the target language. If the translation is done by a number of translators (working independently), they meet to resolve differences in their translations and produce a single version. Other translators then take the agreed-upon version of the test and adapt it back to the source language. Again, this results in several versions which are merged into a single final version. The original and back-translated versions of the test are then compared and judged with regard to their equivalence (see step 5). This procedure focuses mainly on the source-language version. There is a rather low probability of detecting errors through this procedure, as the back translator may be able to produce a good translation into the source language, even if the original translation was poorly done and resulted in a nonequivalent target-language version of the test.

With the forward-only-translation design, the test is adapted from the source language to the target language by one or more translators with different areas of expertise or from different cultural backgrounds, who work independently. The different versions of the translation are then merged into a single version. The advantage of the double-(or multiple-)translation procedure is that a high degree of equivalence of source and target languages is obtained as all the translators involved focus on both the source and target versions. Another type of forward translation is one in which two parallel source versions are produced in two different languages, and then two independent translations (one from each source language) are produced and merged into a single version in the target language. This design was implemented by the Programme for International Student Assessment

(PISA) International Coordination Centre (Grisay, 2003; Grisay *et al.*, 2006). The advantage of this procedure is that it avoids giving too much importance to cultural forms and linguistic characteristics that are tied to one specific source language.

3. *Building the adaptation team.* The team should include translators and reviewers who are fluent in both languages, familiar with both cultures, and who have some knowledge of both the test's subject matter and principles of test development. A team is preferable to a single translator, as it offers a variety of approaches and perspectives, and thus ensures that the right decisions will be made and a good adaptation produced.
4. *Initial translation and adaptation.* Translators should follow guidelines aimed at ensuring that:
 - a. the cultural context of the item is familiar to all examinees who will take the test in the target language;
 - b. none of the items is provocative or offensive in its adapted version;
 - c. the conceptual equivalent of a word or phrase is kept;
 - d. the essential meaning of the item does not change;
 - e. the difficulty level of the translated item is equivalent to that of the source language;
 - f. the language used is natural and fluent; and
 - g. measurement units, seasons, places, animals, monetary currency, etc. are equally familiar to all examinees.
5. *Reviewing.* After the target language version is completed, additional reviewing is needed. It is important that the reviewers are from different backgrounds and countries of origin, since special attention must be given to different dialects or nuances of a language (e.g., the many dialects of Arabic; British as opposed to American English; and Spanish spoken in Argentina as opposed to Spain). It is recommended that the reviewers first check the adapted version of the test without looking at the source version, to see whether the items stand on their own merits. Only then should the adapted version be compared to the original, with special attention given to the quality of the adaptation. Reviewers note their comments and suggest alternatives. A panel of experts which includes the translator, experts in the subject matter, and experts in test development discuss the comments and revise the adapted version. A helpful step in reviewing is asking reviewers to answer the question in the adapted version. If they give a wrong answer, it may indicate that the adaptation is flawed, thereby indicating a need for revision of the item.
6. *Conducting field tests.* The review of the adapted version of a test should involve a number of translators and reviewers to ensure that most of the problems in the adapted version will be identified and fixed. Nevertheless, some problems go unnoticed until test items are field tested. Item revisions are made where necessary.

7. *Constructing the test final version and documenting the adaptation process.* The final target-language version of the test should be the result of all the above-mentioned steps, in accordance with the required test content and difficulty specifications. All stages in the test adaptation process should be documented – including reviewer comments and the changes made.

Achieving Equivalency

Once the test has been constructed and administered, a three-step procedure should be implemented to ensure a high degree of similarity between the scores of the adapted version and those of the source-language version: (1) detecting construct bias by checking that the dimensional structure of the test versions is equivalent, (2) detecting item bias by applying DIF analysis, and (3) linking the scores of the source- and target-language test versions.

1. *Detecting construct bias – dimensionality assessment.* The purpose of this analysis is to determine whether the construct measured exists in all groups, and if so, whether it has been measured in the same manner in all groups. The goal, in other words, is to examine whether the total score calculated for each test version represents the same construct in each language group. Dimensionality assessment can be done using factor analysis, weighted multidimensional scaling, and other methods (Sireci and Allalouf, 2003; Georgas *et al.*, 2003).
2. *Detecting item bias.* According to the commonly accepted definition of DIF, an item functions differently across groups when examinees of equal ability but from different groups (in this case, different language groups) are not equally likely to respond correctly to that item. When DIF is detected, the next step is usually to investigate the reasons for it, so as to determine whether the DIF item is also biased against one of the language groups. A comprehensive distinction between the two terms, DIF and bias, was made by Zieky (1993):

It is important to realize that DIF is *not* a synonym for bias. . . judgment is required to determine whether or not the difference in difficulty shown by the DIF index is *unfairly* related to group membership. The judgment of fairness is based on whether or not the difference in difficulty is believed to be related to the construct being measured (p. 340).

DIF detection can take place before test administration, using a judging-type procedure (Hambleton and Jones, 1995), but is usually carried out statistically, using empirical data, after the test versions have been administered. Among the main methods employed for this purpose are: delta plot/transformed item difficulty (TID) technique, Mantel–Hanszel, logistic regression,

the item response theory (IRT)-based test, and the simultaneous item bias test (SIBTEST); each method has its advantages (see Holland and Wainer, 1993).

Findings from DIF studies. Generally, translated items tend to vary in terms of the degree of DIF detected. Angoff and Cook (1988) analyzed the equivalence between the Scholastic Assessment Tests (SAT, used for admission to higher education institutions in the USA), and its Spanish-language counterpart, the *Prueba de Aptitud Académica* (PAA). They found that verbal items containing a large amount of text have lower DIF than items containing a small amount of text – where every word is critical and every translation problem has a considerable impact on item characteristics. For example, reading comprehension items have lower DIF than verbal analogies. Conversely, no DIF is expected in nonverbal items such as math or figural items, as noted by Gafni and Melamed (1991). Some studies have listed the possible causes for DIF existing across test versions in different languages. For example, Allalouf *et al.* (1999) studied the translation from Hebrew into Russian of the verbal reasoning domain of the Psychometric Entrance Test (PET, which is used for admission to universities in Israel). They found that DIF is likely to occur if there are differences between source and target language in: (i) word difficulty, (ii) item format, (iii) cultural relevance, and (iv) content.

Implications of DIF studies. Test constructors and translators should be made aware of the results of such studies, and take them into account when adapting a test from one language to another. Such caution will help them to avoid translating items that may have a cultural bias, to seek target-language words of comparable difficulty, to preserve the original item format, and to avoid using content that gives an advantage to one of the language groups.

3. Scoring and linking

Scoring. If the DIF analysis reveals that the psychometric characteristics of certain items have been altered by the translation, these items should not be used for scoring both test versions (source and target). This will make the source and target versions more similar psychometrically, and their scores more comparable.

Before the DIF items are omitted, they should be examined carefully for bias by experts in the subject area, since in very rare cases, as was pointed out by Ellis (1989), there may be real differences in the way a certain language group performs on specific test items – differences rooted in experience, knowledge, or culture. These differences may be considered relevant, and in such cases, the DIF items are not to be omitted from the scoring process.

Linking. An adapted test can be linked to the source-language test even if some items in the target language function differently from the source-language

items. This linking can be achieved by using only the non-DIF translated items as common items (anchor). However, this should be done carefully since deleting too many DIF items can result in an anchor that is too short for equating (see Sireci, 1997).

Perspectives for the Future

In a thought-provoking paper titled ‘Comparing the incomparable’, Wainer (1999) discusses a comparison between groups who took different language versions of a test. In particular, he looks at cases in which there is little overlap in content between the test versions. In such cases, Wainer claims, a fair comparison is actually impossible. Professional test adaptation should avoid such situations. Indeed, sometimes certain test items in one source language cannot be adapted into another language due to irreconcilable cultural or lingual differences. In these cases, some items should be developed in the target language directly. Advances in test adaptation procedures and methodology – based on a large number of studies, including very recent research (e.g., International Test Commission, 2006) – have also helped create new and updated test adaptation standards that take into account new linguistic quality-control methods. Software tools for translation and translation-quality assessment will definitely play an important role in the near future – initially as a rough tool for human translators and later as a more sophisticated aid.

Bibliography

- Allalouf, A., Hambleton, R. K., and Sireci, S. G. (1999). Identifying the causes of DIF in translated verbal items. *Journal of Educational Measurement* **36**, 185–198.
- Angoff, W. H. and Cook, L. L. (1988). Equating the scores of the *Prueba de Aptitud Académica* and the scholastic aptitude test. *College Board Report No. 88-2*. New York: College Entrance Examination Board.
- Ellis, B. B. (1989). Differential item functioning: Implications for test translation. *Journal of Applied Psychology* **74**, 912–921.
- Gafni, N. and Melamed, E. (1991). Equating different language versions of a psychometric test. *Research Report No. 148*. Jerusalem: NITE.
- Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment* **6**, 304–312.
- Georgas, J. W., Weiss, L. G., van de Vijver, F. J. R., and Saklofske, D. H. (eds.) (2003). *Culture and Children's Intelligence: Cross-Cultural Analysis of the WISC-III*. San Diego, CA: Academic Press.
- Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment. *Language Testing* **20**(2), 225–240.
- Grisay, A., de Jong, J. H. A. L., and Berezner, A. (2006). Item/country interactions in groups of countries using a same test language. *Paper Presented at the International Test Commission 5th International Conference*, Brussels, Belgium.
- Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests in multiple languages and cultures. *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*, pp 3–38. London: Erlbaum.

- Hambleton, R. K. and Jones, R. W. (1995). Comparison of empirical and judgmental procedures for detecting differential item functioning. *Educational Research Quarterly* **18**, 21–36.
- Hambleton, R. K. and Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Applied Testing Technology* **1**, 1–16.
- Holland, P. and Wainer, H. (eds.) (1993). *Differential Item Functioning*. Hillsdale, NJ: Erlbaum.
- International Test Commission (2006). *The ITC 5th International Conference on Psychological and Educational Test Adaptation across Language and Cultures*. Brussels, 6–8 July 2006. http://www.intestcom.org/itc_conference_2006.htm (accessed July 2009).
- Maxwell, B. (1996). Translation and cultural adaptation of the survey instruments. In Martin, M. O. and Kelly, D. L. (eds.) *Third International Mathematics and Science Study (TIMSS) Technical Report*, vol. I, pp 159–169. Chestnut Hill, MA: Boston College.
- Sireci, S. G. (1997). Problems and issues in linking assessments across languages. *Educational Measurement: Issues and Practice* **1**, 12–29.
- Sireci, S. G. and Allalouf, A. (2003). Appraising item equivalence across multiple languages and cultures. *Language Testing* **20**, 148–166.
- Wainer, H. (1999). Comparing the incomparable: An essay on the importance of big assumptions and scant evidence. *Educational Measurement: Issues and Practice* **18**, 10–16.
- Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In Holland, P. and Wainer, H. (eds.) *Differential Item Functioning*, pp 337–347. Hillsdale, NJ: Erlbaum.
- Zucker, S., Miska, M., Alaniz, L. G., and Guzman, L. (2005). *Transadaptation Publishing Assessments in World Languages*. San Antonio, TX: Harcourt Assessment.

Testing Creativity

M A Runco, California State University, Fullerton, Fullerton, CA, USA; Norwegian School of Economics and Business Administration, Bergen, Norway

© 2010 Elsevier Ltd. All rights reserved.

The study of creativity has become highly objective and scientific over the past 25 years. This increased objectivity has resulted in part from improvements in testing and measurement. These improvements, along with a dramatic increase in the scientific interest in creativity, has been motivated by the nature of creativity: It has innumerable benefits, including those suggested by correlations with various indications of psychological and physical health, and in education, as an objective that can lead to skills (e.g., adaptability, flexible, and original problem solving) that transfer easily to the natural environment. Then there are the benefits for society as a whole. Creativity fuels technological and social progress and evolution. Creativity is associated with discovery, invention, and innovation. It has replaced knowledge and information as the most important commodity for contemporary society and the future.

The fulfillment of creative potential and the enhancement of creative talent, both depend on accurate measurement. The accurate assessment of creative potential and talent is, however, challenging to say the least. Creativity presents several unique measurement problems. There is, for example, no one universally accepted definition of creativity. This has led to a plethora of different kinds of tests and assessments, some of which do not agree with one another. In traditional psychometric terms, the lack of one accepted definition also undermines the construct validity of creativity tests. This is the label given to the interpretation of test results as indicative of a trait, skill, or capacity – in the present case creative talent. As is the case in many definitions of validity, the definition is often formulated as a question. Validity addresses the question, “Are you measuring what you think you are measuring?” The more specific question specifically addressing construct validation is, “Which construct is explained by the test scores?”

The concept of construct validation was broadened in 1999 such that, when possible and directly relevant, information in addition to test scores is taken into account. In fact, the 1999 Standards for Educational and Psychological Testing (APA, 1999) broadened the definitions of all forms of validity. Specifics from particular test questions and content are now to be taken into account, in addition to test scores, for example, and validity decisions are now focused more on the interpretation of scores than on the test scores themselves. It has even been suggested that validity should be treated as a global and unitary concept, which takes what was formerly separated into

construct, content, and criterion-related validities, all into account at once when making psychometric decisions. Clearly, psychometric theory has shifted such that the inferences drawn from tests are the most important and the test scores themselves only one source of information.

Measuring Originality

There is wide agreement on one aspect of creativity, namely originality. Yet, while originality is necessary for creativity, it is by itself not sufficient. Truly creative actions and products also require some sort of effectiveness, fit, appropriateness, or value. Creative things are, at a minimum, both original and fitting. The originality requirement makes measurement especially difficult because it implies that an examinee’s responses on a test may be the most creative when they are unique, novel, unusual, and contrary to norms. In other words, original and therefore creative things are dissimilar to what has come before. Instead of fitting into norms, then, original and creative responses and answers to test questions are uncommon or even unique. This obviously is dramatically different from a test which assesses whether or not an examinee knows what are supposed to be the correct answers. One breakthrough in the field of creativity testing took place when originality was operationalized in terms of statistical infrequency, rarity, or falling outside norms. In this fashion, originality can be measured.

Tests as Samples of Behavior

There are a large number of creativity tests. None of them is perfect, though of course the same thing could be said of just about any test. Assessments and tests are, after all, samples of behavior, and as such they are limited. The more accurate tests have content validity; they provide the most representative samples of behavior, ability, or aptitude. The better the sample, the better is the test. For creativity, the difficulty is that a large range of processes may be relevant. Creativity is typically defined as a syndrome, or complex, for this reason. Most tests focus on one part of the creativity complex (e.g., personality or cognitive capacity), or at most a small set of traits or capacities. They do not, then, sample the range of relevant skills and traits. Not surprisingly, the predictive validity of these tests is moderate at best.

A second difficulty is that creative behavior is typically intrinsically motivated. Interest comes from within the individual rather than from rewards, incentives, bonuses, or other extrinsic factors. Both intrinsic and extrinsic motives can sometimes play a role in creative performances, but the former is typically much more important than the latter. In fact, extrinsic factors can actually inhibit creative performance. They are sometimes distractions and frequently pull the individual's attention away from the tasks and direct that attention towards expectations, schedules, or the reward itself. Sometimes intrinsic motivation diminishes when extrinsic factors are presented to the individual. When this happens, creativity tends to suffer. All of this creates problems for measurement because tests, and their rationale and the feedback provided by the scores, are all extrinsic. Thus an individual can be highly creative when following his or her own intrinsic interests, but the examiner will not uncover that talent because the test is not an intrinsically motivated activity.

Divergent Thinking Tests

Divergent thinking tests are probably the most commonly used estimate of the potential for creative thinking. They can be used to measure originality and can be used in such a way that intrinsic interests are tapped. These are the tests alluded to earlier that take advantage of the rarity of original and creative responses, answers, solutions, and ideas. They are characterized by open-ended questions, which allow an examinee to produce numerous ideas. This is one of the most obvious ways that tests of divergent thinking differ from intelligence tests (IQ tests) and most academic tests. Those tests rely on convergent thinking; there is one or very few correct or conventional answers. The question, "where do we get turpentine" is convergent and typical of IQ tests. (Points are awarded for the answer "a fir tree" but not for any other answer, including "a hardware store".) The question, "in what ways can a brick be used" allows divergent thinking. Many responses can be given. That is a question for the alternative uses test of divergent thinking. Other questions ask about uses for a shoe or coat hanger. The instances test, asks the individual to list things that move on wheels, strong things, or things which are square. The similarities test asks how a grocery store and restaurant are alike, or how a potato and carrot are alike. There are also visual or figural tests, which ask the examinees to list their ideas concerning what an abstract line drawing could represent.

Performance on these tests is usually statistically independent of performance on tests of convergent thinking. The implication is quite significant: If schools rely on tests of convergent thinking, which most do, they will not identify students with notable potential for original

problem-solving and creative talent. Also quite significant is the point that the independence of divergent and convergent thinking seems to depend on the testing environment. There is clear independence when divergent thinking tests are administered in permissive testing environments (which emphasize through test instructions that the students can list multiple ideas and need not worry about spelling, grades, or time limits) but some correlation between divergent and convergent thinking in more traditional testing ones (where there may be time limits and clear expectations for remembering correct answers or factual information). Permissive testing environments minimize the impact of extrinsic constraint and allow intrinsic involvement by the students.

Divergent thinking tests have been modified since they were first developed, and some of the newer tests are designed specifically to allow intrinsic interests. There are, for example, realistic tests of divergent thinking that ask about situations which students might actually encounter in the natural environment. One of these asks students to list possible solutions to a problem of forgotten homework or getting in to trouble by sitting next to a loud uncooperative peer. Other tests allow problem-finding, problem-identification, and problem-generation. These represent processes which have been associated with actual creative performance many times over. In fact, it has been proposed that the quality of a solution depends upon the quality of a problem. Therefore individuals must first find a good problem or they will not find good and creative solutions. Problem-generation tasks present questions such as, "Can you list problems which might be faced by students at your school?" Another is, "List as many problems as you can that might arise between you and your peers." As is the case with all estimates of divergent thinking, these allow multiple ideas and solutions.

The directions given with such tasks will depend on the age of the examinees. Otherwise a verbal bias might keep a creative individual from understanding the task well enough to perform at his or her true level of ability. And that is one goal of testing: to minimize measurement error and thereby obtain scores that are indicative of true ability or true capacity. Recall here the idea in the 1999 Standards for Educational and Psychological Testing that the objective of good testing is to draw useful, accurate, and appropriate inferences.

Tests of divergent thinking have been used with nearly all age groups. Even preschool children can be tested, though not with paper-and-pencil measures. They can be given a tangible stimulus, such as a plastic object, and asked to list all of the ways it can be used or all of the things it might be. Their responses are recorded and, if necessary, transcribed for scoring. Students of all ages can also be tested easily, as can special adult populations (e.g., entrepreneurs) or individuals in late adulthood. Divergent thinking tests are not the final answer in creativity-testing,

but they do have a broad application. They are best viewed as estimates of the potential for creative thinking. They are certainly not synonymous with, nor do they guarantee, actual creative performance.

Creativity, Problem Solving, and Bias

This brings us to another issue in creativity-testing. It involves a relationship between creativity and problem-solving. Divergent thinking tests assume that creativity can be estimated by asking individuals to find and solve problems. Yet, not all creativity involves problem-solving. Some creativity involves self-expression and some is proactive rather than reactive (i.e., a response to a problem).

Then there is the problem of experiential bias. This is a concern with realistic tests. Their realism may make these tests more intrinsically interesting than the more artificial tests, and individuals tend to give more ideas to realistic tests than the other ones. Unfortunately, realistic tests elicit relatively low originality scores. This is probably because of the same kind of experiential bias which plagues IQ testing. With an experiential bias, individuals who have particular experiences rely on their memory of those experiences instead of, or perhaps more often than, using their imagination and generating ideas anew. In an IQ test, for example, an individual might be asked how far Honolulu is from Los Angeles. If the examinee has made that particular trip, he or she may remember how far it was and know the answer. But what if the same examinee, with the same level of intelligence, has not made that trip? Of course, questions like this draw heavily from long-term memory, but that is just another way of saying that there is an experiential bias; it might also be called a memory bias. It is a potential problem for construct validity. However, there is another perspective, a rebuttal which notes that individuals must be intelligent in order to benefit from experience.

All divergent thinking tests are scored for ideational fluency and ideational originality. The first of these is operationally defined as the number of ideas produced. The second is defined in terms of the number of unique or original ideas produced. Occasionally, ideas are judged subjectively for their originality, but of course this introduces the issue of interjudge or interrater reliability. Divergent thinking tests are sometimes scored for ideational flexibility, which represents the number of categories or themes in an individual's ideational pool. There is a fair amount of overlap among these three indicators, but this can be statistically controlled using regression techniques if predictions are being calculated using correlational statistics, or ratio scores might be used. These take fluency or originality and divide by an individual's fluency. The resulting ratio score controls for an individual's productivity. This is especially important because

chances are that the more ideas an individual produces, the more likely it is that he or she will be flexible and find unique or unusual ideas.

The theory of remote associates also justifies ratio scores. It suggests that the best ideas are remote. According to this theory, humans generate ideas along with change, one idea leading to the next, and then to the next. The first ideas we tend to think of are the most obvious, and original and unusual ideas are only found later. They are remote associates. Of course ratio scores have some statistical problems. In particular, ratio scores are often unreliable. They are also hypothetical because they hide an individual's actual ideation. In other words, an individual who produces two ideas, one of which is highly original, will have the same ratio scores as an individual who produces 20 ideas, ten of which are original.

Occasionally, only fluency is used with divergent thinking tests, but this is probably inappropriate given what has been said earlier about originality being the necessary component of creativity. It is probably best to use a profile that includes fluency, originality, and flexibility for any examinee. Other indices have been used, including elaboration and appropriateness. These are not nearly as well understood as fluency, originality, and flexibility, nor as often used. They do have particular applications.

Domain-Specific Assessments

Creativity potential and talents are sometimes assessed within domains, such as art, music, mathematics, and so on. This is an attractive option, given the corpus of research showing that eminent creators tend to work within particular domains and the research, often with students, that they tend to have quite specific creative interests. This can be done in two ways. First, a checklist of activities and accomplishments can be administered. These list different activities within domains and ask the respondent how often they have been involved in each (e.g., painting pictures, taking photographs, and using mathematics in an original way to solve a problem). These checklists often sample several domains, and include a section asking the respondent to briefly describe his or her most impressive project. Although the essay format of this description requires determination of interrater reliability, it does allow a qualitative score to complement the quantitative score provided by the checklist. Checklists such as these may best be viewed as criteria rather than as predictors of creativity since they look back at previous activities and achievements.

A second domain-specific assessment method involves actual creative products. Examinees can be asked to write a poem, for example, to assess linguistic or poetic creativity. They might also be asked to construct a simple collage. Of course, products such as these require an evaluation by

judges, so once again the issue of interrater reliability arises. The consensual assessment technique, developed by Teresa Amabile, is often used with products and tends to have highly reliable ratings. It relies on appropriate judges (who are experts in the domain in question) who each examine and rate products on originality, technical skill, and creativity. Importantly, the assumption is that appropriate judges know what to look for and therefore are not given any definition of creativity.

Personality and Biographical Indicators

Personality tests are sometimes used as estimates of creative talent. Even the MMPI now has a creativity scale; a creative personality scale has long been used as part of the California Psychological Inventory. Several creative personality scales have been developed for the 300-item Adjective Check List. Other tests of personality, including those focused on the Big Five, are sometimes used in creativity research. Of particular relevance within the context of the Big Five is the Openness scale. One advantage of most of these personality assessments is that they can take into account both indicative traits (e.g., openness to experience), which support creative efforts, as well as the contraindicated traits (e.g., conformity), which make creative performance unlikely.

Admittedly, personality assessments are probably not the most impressive estimates of creative talent, in part because the important characteristics vary from creative domain to domain. Additionally, the State X Trait interaction, which characterizes so much of personality psychology, is an issue, the crux of which is that traits vary from situation or context to situation (the state). Therefore, an individual may show the relevant traits in one context, but not in another. For this reason, predictions from personality tests are limited. Of course, personality assessment is one useful source of information. As a matter of fact, this may be the best way to approach the measurement of creativity: recognize that it is a complex syndrome and multifaceted, and therefore, avoid relying on any one assessment.

Biographical measures of creative potential are usually developed by identifying the common background variables of unambiguously creative individuals, and then asking other individuals (on a paper-and-pencil measure) if they have had similar experiences. The Biographical Inventory of Creativity, for example, asks individuals if they had imaginary friends as children, or if their parents took them to museums and other cultural activities. These sorts of things may be relevant; highly creative individuals often do have imaginary friends as children and their parents typically do expose them to diverse cultural activities. The limitation is probably obvious: not all creative people may have the same backgrounds. Additionally, an individual may have an ideal background for creativity but

not be interested in expending the effort to be creative. When this is the case, someone with a creative personality or creative background may not behave or perform in a creative fashion. In psychometric terms, any index of creative personality or background is in this particular situation unlikely to have acceptable predictive validity.

Ratings by Parents, Teachers, or Supervisors

Sometimes it is useful to obtain ratings from parents, teachers, or supervisors. The rationale for this approach is that these groups often have a great deal of information about the individual being rated. The raters may also be more objective than the person himself/herself. Self-reports are open to memory, honesty, and socially desirable responding errors. Of course, raters may not be perfectly objective. Parents, for example, may have a slanted perspective and ratings could easily suffer from halo effects or similar biases. In this example, a halo effect may occur when parents' feelings about their child influence all ratings, across the board. Even if the child is not often original, the parents may think so highly of the child that they give a high rating to originality – and everything else. This may explain why ratings given by parents and teachers do not agree all that much. This is not to say that teachers are entirely objective. They seem to look mostly at characteristics that fit well with the context of a classroom (e.g., being polite and considerate), which makes perfect sense given that they may have 20, 30, or even more children to monitor at any one time.

The Criterion Problem

One of the longstanding difficulties in the area of creativity-testing is referred to as the criterion problem. This is a huge problem because tests must have criterion-related validity. They should, for example, be able to predict certain criteria (those positively related to the construct in question) but also be unrelated, or perhaps negatively related, to constructs, which should be independent. As an example of the latter, which is discriminant validity, creativity test scores should not be strongly related to traditional intelligence and the IQ. If they are strongly related, it is likely that the creativity test is merely testing traditional intelligence.

There are no widely accepted criteria of creative potential and creative talent. As a matter of fact, there is a kind of circularity here because if there was a widely accepted criterion of creative performance, it might be used instead as a predictor! That line of reasoning does not apply to one kind of criteria, namely those that are unambiguous, such as the actual creative performances of

established creators (e.g., Nobel laureates or award-winning artists). Some measures start by selecting a particular domain and then identifying individuals who are performing at a very high level and are ambiguously creative within this domain. Their backgrounds may then be examined with the hope that predictors can be identified and used to select up-and-coming creative persons. There is, however, yet another kind of sampling concern with this method. If this kind of methodology is used to find a predictor of creative talent, only mature and fulfilled potential will be measured. In other words, the predictors used in the biographical or other predictive measures will only reflect that which was important for individuals who have everything going for them and already perform at a highly creative level. What predicts their performance may not help us to identify individuals who are merely potentially creative. There is a big difference between mere potential and actual manifest performance. For this reason, measures developed with this kind of retrospective method may not be useful when working with children or in gifted and talented educational programs. In both of those cases, the interest may be a measure of creative potential rather than creative productivity.

Conclusions

Not surprisingly, several psychometricians have concluded that creative thinking must be used by examiners when attempting to develop creativity tests or attempting to

measure creative talent and creative potential. Standard measurement techniques often do not apply to creativity. It poses several unique challenges for psychometricians, especially because originality is necessary but not sufficient and thus norms cannot be used as they are with other tests.

As is the case with all good scientific work, it is best to start with a sound theory. In the case of creativity, this means accepting the notion that it is a syndrome or complex and that no one measure captures all kinds of performance (e.g., all domains). It is probably best to triangulate and employ more than one measure. Even then, the intrinsic interests of examinees must be taken into account, as should the idea that individuals who have manifest creative talents may have something that differs from what we see in individuals who have mere potential and need encouragement for its fulfillment. As a matter of fact, it is these individuals with potential in need of fulfillment who probably will benefit the most from accurate testing.

Further Reading

American Educational Research Association (1999). *Standards for Educational and Psychological Testing*, Second Impression 2004. Washington, DC: American Educational Research Association.

Hennessey, B. and Amabile, T. (1999). Consensual assessment. In Runco, M. A. and Pritzker, S. (eds.) *Encyclopedia of creativity*, pp 347–359. San Diego, CA: Academic Press.

Runco, M. A. (2006). *Creativity: Theories and Themes: Research, Development, and Practice*. San Diego, CA: Academic Press.

Testing in History

S Wineburg and B Fogo, Stanford University, Stanford, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

In 1917, the year the United States went to war, history erupted onto the pages of the American Psychological Association's *Journal of Educational Psychology* (Bell, 1917). J. Carleton Bell – the *Journal's* managing editor and professor at the Brooklyn Training School for Teachers – began his tenure with an editorial entitled 'The historic sense'. Bell claimed that the study of history provided an opportunity for thinking and reflection – qualities lacking in many classrooms.

In a companion article, Bell and his colleague, David F. McCollum, presented the first empirical report of achievement testing in history. The authors put forth five ways in which the elusive historic sense might be defined:

1. The ability to understand present events in light of the past.
2. The ability to sift through the documentary record – newspaper articles, hearsay, partisan attacks, contemporary accounts – and construct "from this confused tangle a straight forward and probable account" of what happened, a goal they pointed out of many "able and earnest college teachers of history" (Bell and McCollum, 1917).
3. The ability to appreciate a historical narrative.
4. "Reflective and discriminating replies to 'thought questions' on a given historical situation" (Bell and McCollum, 1917).
5. The ability to answer factual questions about historical personalities and events.

The authors conceded that this fifth aspect was "the narrowest, and in the estimation of some writers, the least important type of historical ability" (Bell and McCollum, 1917). At the same time, they acknowledged it was this aspect that was "most readily tested." In a move that would seal the fate of achievement testing in history in North America for much of the twentieth century, Bell and McCollum elected the path of the "most readily tested." Of their five possibilities only one – the ability to answer factual questions – was chosen for study. While perhaps the first instance, this was not the last in which ease of measurement – not priority of subject matter understanding – determined the shape and contour of an achievement test in history.

Bell and McCollum administered their test to 1500 Texas students in 1915–16. The overall score at the elementary level was a dismal 16%. In high school – following a year of history instruction – students scored a shabby

33%, and, in college – following then a third exposure to history – scores barely approached the halfway mark (49%). The authors concluded that studying history in school led only to "a small, irregular increase in the scores with increasing academic age." Bell and McCollum blamed the educational system and its charges: "Surely a grade of 33 in 100 on the simplest and most obvious facts of American history is not a record in which any high school can take pride" (Bell and McCollum, 1917).

By the next world war, hand-wringing about students' historical benightedness had moved from the back pages of the *Journal of Educational Psychology* to the front pages of the *New York Times*. "Ignorance of U.S. History Shown by College Freshmen" trumpeted the headline on 4 April 1943 – a day when the main story reported that Patton's troops had overrun Field Marshal Erwin Rommel at El Guettar. Providing support for Allan Nevins's claim that "young people are all too ignorant of American history" (*New York Times Magazine*, May 3, 1942), the survey showed that a scant 6% of the 7000 college freshman could identify the 13 original colonies. Often, students were simply confused. Abraham Lincoln "emaciated the slaves" and, as first president, was father of the Constitution. A *Times* editorial excoriated these "appallingly ignorant" youth. "Either the college freshman, recently out of high school, were poorly prepared on the secondary level," surmised *Times* reporter Benjamin Fine, "or they had forgotten what they learned about United States history" (*New York Times*, 4 April 1943). In either case, the survey revealed a "vast fund of misinformation on many basic facts" among American youth (Fine, 1943).

Hand-wringing over students' lack of historical knowledge resurfaced in time for the US bicentennial celebration in 1976. With the aid of the Educational Testing Service (ETS), the *Times* tested nearly 2000 freshmen on 194 college campuses. On 2 May 1976, the results rained on the bicentennial parade: "*Times* Test Shows knowledge of American History Limited." Of the test's 42 multiple-choice questions, students averaged 21 correct – a failing score of 50%.

Results from the 1987 and 1994 administrations of the National Assessment of Educational Progress (NAEP; known informally in the US as the nation's report card) have shown little deviation from these earlier trends (Ravitch and Finn, 1987; NAEP, 1996). The 2001 NAEP administration serves as a case in point. In its wake came the same stale headlines: "High School Seniors Flunk History," (Fletcher, 2002; *Washington Post*, May 9,

2002); “Kids Get ‘Abysmal’ Grade in History: High School Seniors Don’t Know Basics,” *USA Today* (2002) (May 10–12, 2002); the same refrains of cultural decline (“a nation of historical nitwits” wagged the Greensboro [North Carolina] *News and Record*, 13 May 2002); the same holier-than-thou indictments of today’s youth (“dumb as rocks” hissed Lee Bockhorn in the *Weekly Standard*, 13 May 2002); and the same boy-who-cried-wolf predictions of impending doom (“when the United States is at war and under terrorist threat” young people’s lack of knowledge is particularly dangerous, according to *Palm Beach Post*, 10 May 2002). Scores on the 2001 test – significant in that they came after a decade of the Standards Movement – were virtually identical to their predecessors. Six in ten seniors “lack even a basic understanding of American history,” wrote the *Washington Post* – results that NAEP officials castigated variously as “awful,” “unacceptable,” and “abysmal” (reported in *USA Today*, 10–12 May 2002).

Pointing to the 2001 NAEP results, the Albert Shanker Institute claimed in their blue-ribbon report that “something has gone awry . . . We now have convincing evidence that our students are woefully ignorant of who we are as Americans,” indifferent to “the common good” and “disconnected from American history” (Albert Shanker Institute, 2003, emphasis added). All told, such trends among youth point to a perilous “loosening from our heritage” (Albert Shanker Institute, 2003).

One wonders what evidence this committee now possessed that had not been gathering moss since 1917 when Bell and McCollum hand-tallied 1500 student test-sheets. Explanations of today’s low scores wither when applied to results from 1917 – history’s apex as a subject in the school curriculum. No one can accuse those Texas teachers of teaching process over content or serving up a tepid social studies curriculum to bored students – the National Council for the Social Studies did not even exist at the time. Instead of being poorly trained and laboring under harsh conditions with scant public support, these Texas pedagogs were among the most educated members of their communities and commanded wide respect.

There is something almost comic in the ritual display of young people’s historical ignorance – a phenomenon not unique to the US, with abundant parallels in Canada, Great Britain, Australia, Israel, and many other nations. Each generation feels compelled to wag its finger at its youth, all the while forgetting that only a few years back it too was the target of adult scorn. Indeed, Dale Whittington has shown that when results from the early part of the twentieth century are put side by side with the most recent tests, US students do about as well as their parents, grandparents, and great-grandparents (Whittington, 1991). That is a remarkable finding when we compare current near-universal enrolments of today with the elitist composition of the high school in their teens and early 20s.

Young people’s knowledge hovers with amazing consistency around the 40–50% mark despite radical changes in the demographics of test-takers across the centuries.

Instead of rehearsing the ritual of giving kids a test only to find out that they do not know what adults think they should know, we might, instead, raise basic questions about the ability of 17- and 18-year-olds to memorize and retain volumes of discrete, decontextualized factual information – the stuff which, since 1917, has largely defined the substance of achievement testing in history in the US. With the exception of the Advanced Placement examination – geared to a small proportion of college-bound students – multiple-choice tests have dominated large-scale achievement testing in history, despite the fact that written expression – the arrangement of facts into narrative and argument – is at the heart of the historical process.

What happened to the once-prized historical essay? To some extent, it fell victim to a psychometric culture that privileged exactness and objectivity of scoring over deep reflection on the nature of thinking and achievement in the discipline (cf., Stake, 2007). Consider a 1930 study by F. R. Gorman and D. S. Morgan, who argued that having students write essays in history produced “as much harm as it does good.”

The study was conducted in three US history classrooms. These classes – all taught by the same teacher – were assigned different amounts of written homework. Class I was assigned three units, Class II one unit, and Class III none at all. Indeed, Class III did best on the factual outcome measure (181 points vs. 175 for Class I), but the authors failed to account for the wide disparity in the entering achievement levels among students. Moreover, the researchers’ homework assignments often looked more like directions for busy work than for composing thoughtful written responses (e.g., “List Lincoln’s cabinet with the offices held by each”; “List the states which seceded in order, with the dates of secession”). When Gorman and Morgan (1930) concluded that “the popularity of written work with teachers may result from a confusion of busy work with valid learning procedures,” one wonders where the confusion truly lay: with muddled teachers or with zealous researchers determined to demonstrate the ineffectiveness of written assignments?

Advances in psychometrics fueled the movement toward objective testing, as did the spirit of Taylorism that swept American schools between the world wars (Callahan, 1962). However, it would be wrong to see the focus on objective testing as a movement restricted to education and advances in psychometrics. The fact-based image of historical knowledge fit cozily with prevailing views of knowledge in the discipline of history in the 1920s and 1930s. As educational psychologists worked to produce reliable and objective history scales, university historians tried to extricate themselves from their humanistic roots

in order to emerge as scientists who would, as the saying went, “cross an ocean to verify a comma” (Novick, 1988). This doggedly factualist approach – as Peter Novick argued in *That Noble Dream* (1988) – helped distinguish professional historians from their amateur colleagues – a distinction necessary if history was to become a full-fledged member of the academic community. It is no coincidence, then, that, at nearly the same time, Sackett (1919) was presenting his refinement of a world history scale in the pages of the *Journal of Educational Psychology* – a scale that would “nearly eliminate the subjective factor in grading history,” (p. 348), the *American Historical Review’s* editorial policy was being formulated to exclude from its pages “matters of opinion” in favor of “matters of fact capable of determination one way or another” (cited in Novick, 1988).

The Standards Movement

Forms of testing that focused on low-level knowledge and discrete facts were supposed to become obsolete with the advent of standards – a movement that, according to one of its founders, Mark Tucker, would bring us back to asking the most fundamental questions with regard to what students should learn (Tucker and Coddling, 2002). The new goal was not to shake out a grade, to invent items that discriminated among students, but to establish a set of core understandings that all students should attain. In the remainder of our article, we provide an overview of what has happened to the grand promise of standards – and the changes they were supposed to effect on achievement testing – in the most populous state in the union, California.

Indeed, it was in California that the first standards in history/social science were adopted. They originated in the *History–Social Science Framework for California Public Schools*, passed by the State Board of Education in 1987. The *Framework* initiated a new era in state-mandated history–social studies curriculum. It departed from traditional social studies curricula by presenting the study of chronological history in order to develop literacy across the social sciences, to promote civic values, and to nurture critical thinking. The *Framework* increased the amount of history in the elementary grades and mandated sequences of American history for grades 5, 8, and 11 and world history in grades 6, 7, and 10. It was more than twice the size of the state’s previous social science framework and consisted primarily of course descriptions for each grade level – narrative summaries detailing the different units, themes, events, people, and skills covered during the year. As noted in the *Framework’s* preface, these guidelines were designed to help align course content with textbooks, instructional materials, staff development, and testing throughout the state.

In 1998, as a result of national and state legislation calling for the creation of content standards and assessments in core academic subjects, the California Department of Education adopted the *History–Social Science Content Standards for California Public Schools*. As Delaine Eastin, state superintendent of Instruction, and Yvonne Larsen, president of the State Board of Education, explained on page iv: these standards aligned with the *Framework* and “explicitly” defined “the content that students need to acquire at each grade level from kindergarten to 12th grade” (California State Department of Education, 2001). The content standards supplemented the *Framework’s* course descriptions with hundreds of specific historical skills and items. For example, in the 7th grade, students would, in accordance to standard 7.6.8:

Understand the importance of the Catholic church as a political, intellectual, and aesthetic institution (e.g., founding of universities, political and spiritual roles of the clergy, creation of monastic and mendicant religious orders, preservation of the Latin language and religious texts, St. Thomas Aquinas’s synthesis of classical philosophy with Christian theology, and the concept of “natural law”). (California State Department of Education, 1998: 30)

In a nod to local control, the Department of Education stressed that: “The standards include many exemplary lists of historical figures that could be studied. These examples are illustrative. They do not suggest that all of the figures mentioned are required for study, nor do they exclude the study of additional figures that may be relevant to the standards” (California State Department of Education, 1998: vi).

In addition to specifying historical content, the standards also defined analysis skills for students to develop across grade levels. These standards were divided into three categories: Chronological and Spatial Thinking, Research, Evidence, and Point of View, and Historical Interpretation, and split into three levels: 12 analysis standards for grades K–3, 14 for grades 6–8, and 14 for grades 10–12. In 2001, the State Board of Education integrated the content and analysis standards into the *History–Social Science Framework*.

As part of the Standardized Testing and Reporting (STAR) Program implemented in 1998, the California Department of Education initially used the national, norm-referenced Stanford 9 tests to assess the history–social science knowledge of the state’s 9th, 10th, and 11th graders. In the spring of 2001, the Department began introducing the California Standards Tests (CSTs) for history–social science, developed by the Educational Testing Service (ETS) and aligned specifically with the *Framework’s* content and analysis standards. At present, students take these exams at the end of 8th, 10th, and 11th grades. The 8th-grade exam consists of 75 multiple-choice questions covering the standards for grades 6–8,

while the 10th- and 11th-grade tests each include 60 multiple-choice questions to assess the 10th- and 11th-grade standards, respectively. These tests are relatively low stakes. They do not figure into a school's accreditation or a student's graduation status; however, their results are published and do account for approximately 10–15% of a school's Academic Performance Index (API) – a state measure of school progress that is tied to certain rewards and consequences.

The CSTs exemplify many shortcomings of relying exclusively upon multiple-choice tests to measure student historical knowledge. For example, their alignment with the state's history standards is questionable. All of the items on the CSTs relate to specific standards, yet few of the questions begin to capture the breadth of each standard's historical content. In 2005, for example, content standard 6.7.8 – “Discuss the legacies of Roman art and architecture, technology and science, literature, language, and law” (California State Department of Education, 1998: 26) – was measured by the following question:

The origins of checks and balances in the United States can be traced to:

- a. the French Republic
- b. the Roman Republic
- c. the Greek Aristocracy
- d. the Aztec Empire. (California State Department of Education, 2007a: 18)

Furthermore, where the vast majority of the content standards call for students to analyze, explain, understand, discuss, describe, and examine, most of the test questions push students to do little more than recognize and associate. For example, consider content standard 11.5.2:

Analyze the international and domestic events, interests and philosophies that prompted attacks on civil liberties, including the Palmer Raids, Marcus Garvey's “back-to-Africa” movement, the Ku Klux Klan, and immigration quotas and the responses of organizations such as the American Civil Liberties Union, the National Association for the Advancement of Colored People, and the Anti-Defamation League to those attacks. (California State Department of Education, 1998: 49)

Compare this standard with a test-item used to measure student knowledge of it on the 11th-grade CST in 2003:

Marcus Garvey's program in the 1920s emphasized:

- a. vocational training
- b. a back-to-Africa movement
- c. integration into mainstream society
- d. separate-but-equal doctrines. (California State Department of Education, 2007c: 14)

Such examples illustrate the pitfalls of using multiple-choice test questions to assess the complex thinking

processes called for by California's history standards. This tenuous alignment is perhaps most evident in how the CSTs measure the *Framework's* analysis skills standards.

In 2004, analysis skill HR4 (Grade 10) – “Students construct and test hypothesis; collect, evaluate, and employ information from multiple primary and secondary sources; and apply it in oral and written presentations” (California State Department of Education, 1998: 40) – was assessed on the 10th-grade test with the following multiple-choice item:

The streets were hot and dusty on the summer day. Stokers emerged from low underground doorways into factory yards, and sat on steps, on posts, and palings, wiping swarthy visages, and contemplating coals. The whole team seemed to frying in oil. There was the stifling smell of hot oil everywhere.

–Charles Dickens, *Hard Times*, 1854

The historical era most likely referred to in this quotation is the:

- a. Industrial Revolution
- b. Great Awakening
- c. French Revolution
- d. Enlightenment. (California State Department of Education, 2007b: 11)

These questions – like many others on the CSTs – appear to suffer from two sources of invalidity defined by Messick (1995): “construct under-representation” and “construct irrelevant variance.” Where the former features an assessment that “fails to include important dimensions or facets of a construct,” the latter describes test-items that are too general and may measure something other than the target skill or understanding. Both sources of invalidity increase the likelihood of false positives where students are deemed proficient on standards that they may not begin to comprehend.

Beyond the skill mismatch and content under-representation of individual test-questions and standards, other alignment and validity issues characterize the problematic relationship between the CSTs and the *History–Social Science Framework*. Many of the questions on the CSTs feature the historical figures and events that the *History–Social Science Content Standards* claimed were merely “illus-trative” of what teachers might choose to focus on when teaching the standards. For example, in 2006, the 11th-grade test measured content standard 11.7.3 – “Identify the role and sacrifices of individual American soldiers, as well as the unique contributions of the special fighting forces (e.g., the Tuskegee Airmen, the 442nd regimental Combat team, the Navajo Code Talkers)” (California State Department of Education, 1988: 50) – with this question:

During World War II, what was the primary duty of the Navajo Code Talkers?

- a. interpreting confiscated German battle plans
- b. transmitting secret messages to U.S. forces during combat
- c. translating confidential Japanese communications
- d. informing the press about the number of Allied war casualties. (California State Department of Education, 2007c: 16)

Such questions make the history standards far less suggestive and much more prescriptive. They hold teachers and students accountable for the hundreds of figures, groups, events, and phenomena included in the standards and increase the pressure to cover wide expanses of history in short periods of time – a dilemma that approaches the absurd in the case of the 8th-grade test which – by covering the 6th, 7th and 8th-grade standards – measures student knowledge of history from the Paleolithic era to World War I. In so doing, the CSTs run the risk of reducing California's *History–Social Science Framework* to little more than a long list of historical facts for memorization. In this sense, little has changed in the 100 years between Bell and McCollum's test and the 2007 administration of California's STAR test. Then, as now, the real losers are the students.

None of the Above

The dilemmas we face today in assessing young people's knowledge differ little from earlier efforts. Few historians would argue that large-scale multiple-choice tests capture the range of meanings we attribute to the historic sense. Indeed, we continue to use these tests, and will do so in the future, not because they are historically sound or because they predict future engagement with historical study. They endure for the simple reason that they can be read by machines that produce easy-to-read graphs and bar charts. These tests comfort us with the illusion of systematicity – not to mention that scoring them costs a lot less than the alternatives.

A century of achievement testing has taught us time and again that light-rail excursions through mounds of factual information may be entertaining, but such dizzying tours leave few traces in memory. The human mind demands pattern and form, and both are built up slowly and require repeated passes – with each pass going deeper and probing further. If we want young people to know more history, we need to draw on a concept from medicine: triage. As the University of Tennessee's Wilfred McClay (cited in Albert Shanker Institute, 2003) explains,

Memory is most powerful when it is purposeful and *selective*. It requires a grid, a pattern of organization, a structure within which facts arrange themselves in a particular way, and thereby take on significance. Above all, it requires that we possess stories and narratives that link facts in ways that are both meaningful and truthful, and

provide a principle of selection – a way of knowing what facts are worth attending to We remember those things that fit a template of meaning, and point to a larger whole. We fail to retain the details that, like wandering orphans, have no connection to anything of abiding concern The design of our courses and curricula must be an exercise in *triage*, in making hard choices about what gets thrown out of the story, so that the essentials can survive We need to be willing to identify those things that every . . . student needs to know and insist upon them . . . while paring away vigorously at the rest.

Mechanical testing tempts us with the false promise of efficiency – a lure that whispers that there is an easier, less costly, more scientific way. However, the truth is that the blackening of circles prepares students only to blacken more circles in the future. The future belongs to those who know how to think, how to write, how to analyze, and how to critique. When those responsible for writing our history achievement tests realize this, perhaps achievement will truly take place.

Bibliography

- Albert Shanker Institute (2003). *Education for Democracy*. Washington, DC. <http://www.ashankerinst.org/Downloads/EfD%20final.pdf> (accessed August 2009).
- Bell, J. C. (1917). The historic sense. *Journal of Educational Psychology* 8, 317–318.
- Bell, J. C. and McCollum, D. F. (1917). A study of the attainments of pupils in United States history. *Journal of Educational Psychology* 8, 257–274.
- California State Department of Education (1988). *History–Social Science Framework for California Public Schools*. Sacramento, CA: California State Department of Education.
- California State Department of Education (1998). *History–Social Science Content Standards for California Public Schools*. Sacramento, CA: California State Department of Education.
- California State Department of Education (2001). *History–Social Science Content Framework for California Public Schools*. Sacramento, CA: California State Department of Education.
- California State Department of Education (2007a). *California Standards Test: Released Test Questions, History–Social Science, Grade 8*. Sacramento, CA: California State Department of Education. <http://www.cde.ca.gov/ta/tg/sr/documents/rtqgr8history.pdf> (accessed August 2009).
- California State Department of Education (2007b). *California Standards Test: Released Test Questions, History–Social Science, Grade 10*. Sacramento, CA: California State Department of Education. Retrieved from <http://www.cde.ca.gov/ta/tg/sr/documents/rtqgr10history.pdf> (accessed August 2009).
- California State Department of Education (2007c). *California Standards Test: Released Test Questions, History–Social Science, Grade 11*. Sacramento, CA: California State Department of Education. <http://www.cde.ca.gov/ta/tg/sr/documents/rtqgr11history.pdf> (accessed August 2009).
- Callahan, R. (1962). *Education and the Cult of Efficiency*. Chicago, IL: University of Chicago.
- Fine, B. (1943). Ignorance of U.S. history shown by college freshman. *New York Times* pp 1, 32–33.
- Fletcher, M. A. (2002). Students' history knowledge lacking, test finds. *Washington Post*. <http://www.washingtonpost.com/ac2/wp-dyn/A60096-2002May9?language=printer> (accessed August 2009).

- Gorman, F. R. and Morgan, D. S. (1930). A study of the effect of definite written exercises upon learning in a course of American history. *Indiana School of Education Bulletin* **6**, 80–90.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist* **50**(9), 741–749.
- NAEP (1996). *U.S. History Report Card: Findings from the National Assessment of Educational Progress*. NCES Number 96085, 1994. <http://nces.ed.gov/nationsreportcard/pubs/main1994/96085.asp> (accessed August 2009).
- National Research Council (1999). *How People Learn: Brain, Mind, Experience and School*. Washington, DC. <http://books.nap.edu/books/0309065577/html/index.html> (accessed August 2009).
- New York Times (1976). Times test shows knowledge of American history limited (May 2, 1976), **1**, 65.
- Novick, P. (1988). *That Noble Dream: The "Objectivity Question" and the American Historical Profession*. New York: Cambridge University Press.
- Ravitch, D. and Finn, C. E. (1987). *What Do Our 17-Year-Olds Know? A Report on the First National Assessment of History and Literature*. New York: HarperCollins.
- Sackett, L. W. (1919). A scale in United States history. *Journal of Educational Psychology* **10**, 345–348.
- Stake, R. (2007). NAEP report cards and education: An essay review. *Education Review*. <http://edrev.asu.edu/essays/v10n1index.html> (accessed August 2009).
- Tucker, M. S. and Coddling, J. B. (2002). *Standards for Our Schools: How to Set Them, Measure Them, and Reach Them*. San Francisco, CA: Jossey-Bass.
- USA Today (2002). Kids get 'abysmal' grade in history: High school seniors don't know basics (May 10–12 2002), p 1.
- Whittington, D. (1991). What have 17-year-olds known in the past? *American Educational Research Journal* **28**, 759–780.

Further Reading

- Bruner, J. (1960). *Process of Education*. Cambridge, MA: Harvard University Press.
- California State Department of Education (2005). *History–Social Science Content Framework for California Public Schools*. Sacramento, CA: California State Department of Education.
- Eikenberry, D. H. (1923). Permanence of high school learning. *Journal of Educational Psychology* **14**, 463–481.
- Hertzberg, H. W. (1989). History and progressivism: A century of reform proposals. In Gagnon, P. and the Bradley Commission on History in Schools(eds.) *Historical Literacy: The Case for History in American Education*, pp 69–102. New York: Houghton-Mifflin.
- Meyers, G. C. (1917). Delayed recall in history. *Journal of Educational Psychology* **8**, 275–283.
- Paxton, R. J. (2003). Don't know much about history. *Phi Delta Kappan*.
- Salisbury, H. E. (1957). What Americans don't know about Russia. *McCall's Magazine* **84**, 40–41.
- Tucker, M. S. and Coddling, J. B. (2002). *Standards for Our Schools: How to Set Them, Measure Them, and Reach Them*. San Francisco, CA: Jossey-Bass.
- Wesley, E. (1942). History in the school curriculum. *Mississippi Valley Historical Review* **28**, 567.
- Wesley, E. B. (1944). *American History in Schools and Colleges*. New York: Macmillan.
- Whittington, D. (1991). What have 17-year-olds known in the past? *American Educational Research Journal* **28**, 759–780.

Validity

R L Linn, University of Colorado at Boulder, Ouray, CO, USA

© 2010 Elsevier Ltd. All rights reserved.

Validity is the most important consideration in the evaluation of the interpretations of educational assessment results and the uses that are made of those results. Although references are often made to the validity of an assessment, it should be understood that it is the uses, interpretations, and claims about assessment results that are validated.

Validity is a matter of degree rather than an all-or-none characteristic and the degree of validity varies from one interpretation or use to another. One use of assessment results may have a high level of validity, while another use could have modest validity, and a third use might have little or no validity. For example, an assessment might be a relatively good predictor of grades in college and thus yield predictions that have good validity; however, the same assessment might have little validity as a measure of student learning in high school. Another assessment might be a highly valid measure of the content covered in a course, yet have inadequate validity for determining whether or not students should graduate from high school. Thus, it is inappropriate to say that an assessment is valid or invalid. Instead, validity should be evaluated for the particular uses and interpretations that are made of the results; and judgments must be made to determine whether the degree of validity is adequate for a particular purpose.

Guidance on the conduct of a validation program is available from several sources. Perhaps the most notable of these is the *Standards for Educational and Psychological Measurement* (American Educational Research Association *et al.*, 1999) jointly developed by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, hereafter referred to as the *Test Standards* (The three associations responsible for the *Test Standards* recently decided that it is time to revise the *Test Standards* once again and have appointed Barbara Plake and Laurens Wise as co-chairs of the revision committee. Although the revision will introduce changes and new ideas, the conceptualization of validity has evolved gradually from one edition of the *Test Standards* to the next. Hence, it seems unlikely that the revised *Test Standards* will make fundamental changes in the conceptualization of validity.) Several revisions of the *Test Standards* have been published over the last half century. Each revision of the *Test Standards* provides a statement of a broad professional consensus regarding expectations for the evaluation of the adequacy of interpretations of assessments under a wide range of uses at the time a given edition of the *Test Standards* was published.

Evolution of Conceptualizations of Validity

As can be seen either by reviewing the various editions of the *Test Standards* or by an analysis of the writing of leading measurement experts and validity theorists in major publications such as the four editions of *Educational Measurement*, the concept of validity has evolved over time. Until the early 1950s, validity was generally equated either with the relationship of test scores to an external criterion measure or with the degree to which a test measured what it was intended to measure. Correlations between test scores and criterion measures were used to support the first of these interpretations of validity. Analyses of the content covered by the test in relationship to the content domain that the test was intended to measure were used for the second interpretation.

The two notions of validity that were prevalent in the early 1950s were referred to as predictive validity and content validity. More recent conceptualizations of validity have rejected the notion of different types of validity and, instead, have espoused a unified approach to validity that distinguishes different categories of evidence that are relevant to reaching an overall evaluation of validity of particular interpretations and uses of assessment results.

An expanded view of validity was articulated in the first edition of the *Test Standards* which were then called *Technical Recommendations for Psychological Tests and Diagnostic Techniques* (American Psychological Association, 1954). These recommendations still referred to types of validity, but went beyond the two types of predictive and content. Four types of validity – content, concurrent, predictive, and construct – were described in the *Technical Recommendations for Psychological Tests and Diagnostic Techniques*. Concurrent validity was only a slight variation of the prevailing predictive validity which distinguished between correlations of a test with a criterion measure where the latter measure was obtained at nearly the same time rather than at some later point in time and, subsequently, concurrent and predictive validity got subsumed under the heading criterion-related validity. Content validity also represented a perspective on validity that was familiar at that time, particularly with regard to educational achievement tests. Construct validity, on the other hand, was a relatively new idea when introduced in the 1954 *Technical Recommendations for Psychological Tests and Diagnostic Techniques*. Notions of construct validity that were introduced in the *Technical Recommendations*

were elaborated more fully by the following year in a classic article by Cronbach and Meehl (1955), both of whom were members, with Cronbach as chair, of the committee that developed the *Technical Recommendations for Psychological Tests and Diagnostic Techniques*.

The work by Cronbach and Meehl (1955) had a lasting impact on subsequent thinking about and discussions of validity. Although construct validity was initially thought to be more relevant to the assessment of psychological characteristics, such as anxiety or extroversion, than to the assessment of educational achievement, it gradually came to be widely accepted as relevant to assessments of educational achievement. The acceptance of construct validity came about, in part, because leading measurement theorists continued to promote the concept of construct validity (see, e.g., Cronbach, 1971) and, in part, because they illustrated how it could be made more practical and how construct validity considerations necessarily entered into an evaluation of the uses and interpretations of any assessment. Cronbach's (1971) emphasis on the validation of interpretations of assessment results laid the foundation for more of a unitary view of validity and the subordination of content-related and criterion-related evidence to a construct validity perspective.

By the time the fourth edition of the *Test Standards* (American Educational Research Association *et al.*, 1985) was published, this view of validity had become dominant. "The concept [of validity] refers to the appropriateness, meaningfulness, and usefulness of the specific inferences that are made from test scores. Test validation is the process of accumulating evidence to support each inference" (American Educational Research Association *et al.*, 1985: 9). Shortly after the publication of the 1985 *Test Standards*, however, Messick (1989) argued that the *Test Standards* had not gone far enough in moving toward a unified view of validity in which notions of construct validation were paramount.

Messick's (1989) chapter began with the following definition. "Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the **adequacy and appropriateness of inferences and actions** based on test scores or other modes of assessment" (Messick, 1989: 13, emphasis in original). Messick elaborated his definition of validity in a two-by-two table corresponding to the adequacy/appropriateness and inferences/actions distinctions of the definition. The rows of the table distinguished two types of support for validity claims – the evidential basis and the consequential basis that are used to support claims of adequacy and appropriateness. The two columns of the table distinguished between interpretations of assessment results (e.g., 55% of the fourth-grade students are proficient in mathematics) and uses of results (e.g., award of a high school diploma). Although uses generally related to interpretations, implicitly

if not explicitly, the uses involve actions of some type whereas interpretations may not.

Contemporary Conceptualizations of Validity

The chapter on validity in the 1999 *Test Standards* continued the trend apparent in the earlier *Test Standards* and in the chapters on validity in the second and third editions of *Educational Measurement*. The discussion of validity in the *Test Standards* begins with a concise definition of validity. "Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed in the uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests" (American Educational Research Association *et al.*, 1999: 9). Identifying the types of evidence that are most relevant to evaluating interpretations and uses of an assessment logically depends on the proposed uses and interpretations of the assessment results. Thus, it is important to begin with "an explicit statement of the proposed interpretation of test scores, along with a rationale for the relevance of the proposed interpretation to the proposed use" (American Educational Research Association *et al.*, 1999: 9). The 1999 *Test Standards* do not identify different types of validity. Rather they conceive of validity as a unitary concept, but acknowledge that several types of evidence may be germane to the development of a coherent validity argument.

Kane's (2006) chapter on validity in the most recent edition of *Educational Measurement* presents a conceptualization of validity that builds upon the evaluative approach that was articulated by Messick (1989) and in the 1999 *Test Standards*. Kane conceptualized validation as the process of developing two types of argument that he calls the interpretive argument and the validation argument. The "interpretive argument specifies the proposed interpretations and uses of test results by laying out a network of inferences and assumptions leading from the observed performances to the conclusions and decisions based on the performances" (Kane, 2006: 23). The evaluation of the interpretive argument is called the validation argument. The validation argument brings evidence and logical analysis together for the purpose of evaluating the claims and propositions of the interpretive argument.

A systematic analysis of the various uses, interpretations, and claims that are made about assessment results is needed to develop a comprehensive validation program for an assessment. An interpretive argument needs to be developed for each interpretation and the uses entailed in a particular interpretation. Kane (2006) proposes that interpretive arguments be judged in terms of their clarity, coherence, and plausibility. The interpretive argument

can be used to guide the collection of evidence relevant to the particular interpretations, and claims used to support uses of the assessment results, and that evidence needs to be accumulated and organized into relevant validity arguments (Kane, 2006).

Validity Evidence

The 1999 *Test Standards* identify five major “sources of evidence that might be used in evaluating a proposed interpretation of test scores for particular purposes” (American Educational Research Association *et al.*, 1999: 11). These are:

1. evidence based on test content,
2. evidence based on response processes,
3. evidence based on internal structure,
4. evidence based on relations to other variables, and
5. evidence based on consequences of testing.

Evidence Based on Content

An important source of validity evidence is based on the content of an assessment. It begins with the definition of the construct to be measured and the specification of the content for the assessment. Logical analyses and empirical evidence regarding the alignment of the assessment with the content domain that the assessment is intended to measure can provide critical support for a validity argument. Judgments regarding the content of the assessment and its relationship to the content specifications made by independent experts can buttress a validity argument. Experts can also be useful in evaluating the relationship of both the content specifications and the content of the assessment to the construct.

Although content-based evidence is valuable, it rarely, if ever, provides a sufficient basis for a validity argument. Content-based evidence addresses questions of content representativeness and relevance, but does not address questions about assessment scores or the interpretation of those scores. Thus, in most cases, content-based evidence will play only a limited role in the overall validity argument (Kane, 2006: 19).

Evidence Based on Response Processes

Interpretations of assessment results often involve inferences about the mental processes that individuals use when responding to an assessment. For example, an assessment may be interpreted as a measure of a person’s reasoning or problem-solving skills. An issue is the degree to which such interpretations are valid. One approach to obtaining evidence about response processes is to question respondents to assessments about the strategies they used to respond to problems posed by the assessment.

Think-aloud protocols are also used as a means of obtaining evidence about the nature of the response processes during the assessment.

Evidence Based on Internal Structure

Internal analyses can provide evidence of the degree to which the internal structure of the assessment is consistent with proposed interpretations. If several distinct components comprise the construct, then an internal analysis should provide evidence that those components can be distinguished and that they can be combined in ways that are consistent with the hypothesized construct. For example, a mathematics assessment might be intended to measure number sense, algebraic operations, measurement concepts, and geometric relations. An internal analysis that provided evidence that those four dimensions could be reliably distinguished would support the validity of the intended interpretation of the assessment scores. On the other hand, an internal analysis of an assessment of reading that was conceived of as measure of a single dimension should provide evidence of a single major dimension with a high level of internal consistency.

Evidence Based on Relations to Other Variables

Correlations with criterion measures that an assessment is intended to predict are obviously relevant to an evaluation of the validity of an assessment. If an assessment is intended to predict the success in college, for example, then correlations with the grades that students earn in college as well as other measures of success in college provide direct evidence regarding the degree to which the assessment is functioning as intended. However, relations to criterion measures are not the only relations that are relevant.

Relations to other measures of the construct that an assessment is intended to measure provide convergent evidence that supports the validity of the assessment. Relations to measures that are designed to measure different constructs are also relevant. For example, finding that scores on an assessment of mathematical reasoning have higher correlations with other measures of mathematical skills and reasoning than with measures of reading would provide both convergent and divergent evidence of validity.

Evidence Based on Consequences of Testing

The *Test Standards* address the issue of consequences in validation of interpretations and uses of assessment results. After noting that assessments are “commonly administered in the expectation that some benefit will be realized from the intended use of the scores” the *Test Standards* go on to conclude that a “fundamental purpose of validation is to indicate whether these specific benefits are likely to

be realized” (1999: 16). The role of consequences as in the validation has been a source of controversy among measurement experts for a number of years.

Although the inclusion of consequences as part of validation is controversial, Kane (2006: 54) has recently noted there is, in fact, nothing new about giving attention to consequences in investigations of validity. What is relatively new is the salience of the topic and the breadth of the reach that is no longer limited to immediate intended outcomes (e.g., students perform better in classes following the use of a placement test). The inclusion of broader social consequences and the inclusion of unintended as well as intended consequences led to objections by some measurement experts.

The objections that some measurement experts have to the inclusion of consequences in the evaluation of validity have more to do with the question of whether consequences should be a part of validity than they do with the question of whether or not consequences are relevant to an evaluation of a use or interpretation of assessment results. There is broad consensus regarding the importance of investigations of consequences as part of the overall evaluation of particular interpretations and uses of assessment results; however, some authors have maintained that such an evaluation is outside the scope of validity.

Intended consequences of educational assessments often include such outcomes as improved teacher and student motivation and improved student achievement. The collection of evidence regarding such intended consequences of a use of an assessment may involve the use of teacher and student questions, observations, interviews, and/or historical records of student achievement. It is important to look for plausible unintended consequences, such as increased teacher stress and increased student dropout, as well as intended consequences.

An Alternate View

Although the conceptualization of validity articulated in the *Test Standards* represents the prevailing view among measurement experts, there is no universal agreement. As was already mentioned, there is disagreement among experts about the treatment of consequences of the uses of assessment results. A few measurement experts also object to the notion that it is the particular uses and interpretations of assessment results that are validated, rather than the assessment itself.

The November 2007 issue of *Educational Researcher* featured an article by Lissitz and Samuelsen (2007) that proposed an alternative conceptualization of validity that departs in radical ways from the view espoused in the last two editions of the *Test Standards* and in the chapters on validity in the last three editions of *Educational Measurement* by Cronbach (1971), Messick (1989), and Kane (2006).

Lissitz and Samuelsen argue that content considerations are the core of validity. Rejecting the idea that validity depends on the uses and interpretations of assessment results, they argue that validity is (1) a property of an assessment and (2) is determined by evidence related to internal characteristics of content, reliability, and latent processes. Relationships to other variables and considerations of the consequences of assessment uses and interpretations are seen as relevant to an evaluation of the assessment; however, Lissitz and Samuelsen believe that such external considerations are outside the realm of validity.

The Lissitz and Samuelsen article is useful for calling attention to the importance of content, response process, and internal structure considerations; however, as is clear from the analyses by five well-known measurement specialists who wrote responses to that article, considerations of content, response processes, and internal structure do not provide a sufficient basis for evaluating the validity of assessments. Evidence based on content, response processes, and internal structure that are emphasized by Lissitz and Samuelson is clearly important, but the meaning, and hence the interpretation, of assessment results requires a more unified approach to validity that includes the consideration of constructs and relationships to other variables. Validity is best thought of as dependent upon the uses and interpretations of assessment results rather than as a property of the assessment independent of the uses that made be the results or the ways in which assessment results may be interpreted.

Summary

Validity is the foremost consideration in the evaluation of the uses and interpretations of assessment results. It is the interpretations and uses that are validated rather than the assessment itself. Validation begins with the identification of the intended interpretations and uses of assessment results. The identified interpretations can be used to develop interpretive arguments that guide the accumulation of evidence needed to support those interpretations. The evidence can then be organized into a coherent and persuasive validity argument that supports the interpretations and uses of assessment results.

A variety of types of evidence may be relevant to the development of a coherent and comprehensive validity argument. Evidence based on content, on response processes, the internal structure of the assessment, the relation of assessment scores to other measures, and the consequences of assessment uses are all potentially relevant to the development of a validity argument. The most relevant mix of types of evidence depends on the interpretations and uses that are made of assessment results.

Bibliography

- American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education (1985). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association, and the National Council for Measurement in Education (1999). *Standards for Educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Psychological Association (1954). *Technical Recommendations for Psychological Tests and Diagnostic Techniques*. Washington, DC: American Psychological Association.
- Cronbach, L. J. (1971). Test validation. In Thorndike, R. L. (ed.) *Educational Measurement*, 2nd edn., pp 443–507. Washington, DC: American Council on Education.
- Cronbach, L. J. and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin* **52**, 281–302.
- Kane, M. (2006). Validation. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 17–64. Westport, CT: American Council on Education/Praeger.
- Lissitz, R. W. and Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher* **36**(8), 437–448.
- Messick, S. (1989). Validity. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn., pp 13–103. New York: Macmillan.

Validity of Achievement Gains on High-Stakes Tests

D Koretz, Harvard Graduate School of Education, Cambridge, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Audit test – A second test used to validate gains on a high-stakes test.

Focal test – The high-stakes test scores on which are compared to the audit.

Score inflation – An increase in scores larger than warranted by real improvements in the domain measured.

Target of inference – The set of skills and knowledge about which a test-based inference is made.

VIHS – Validity of inferences under high-stakes conditions.

- Scores on high-stakes tests often – but not always – become inflated.
- The inflation of scores is sometimes very large and rapid.
- Score inflation does not require concrete sanctions or rewards for test scores.
- Inflation is highly variable across schools, and we have as yet little evidence relating its severity to characteristics of schools, students, or test-based accountability programs.

One of the most important elements of education policy today is the growing use of achievement tests to monitor the performance of schools and hold educators accountable. This use of tests has been a cornerstone of education policy in the US and England for some time and is now the focus of reform efforts in numerous other countries.

These uses of tests raise two critically important questions. How valid are inferences about performance when students or educators are pressurized to raise scores on a particular test? How can we evaluate the validity of these inferences? Research to date suggests that scores on tests used for this purpose can become seriously inflated – that is, they can increase markedly more than actual gains in student learning warrant. Addressing this possibility raises difficult theoretical and methodological issues.

Terminology is used inconsistently in the literature addressing these issues. In this article, the term high-stakes testing refers to any testing program that creates strong pressure to raise scores on a particular test, regardless of whether educators or students face concrete sanctions or rewards attached to scores. Test preparation is used to refer to any methods, either desirable or undesirable, designed to prepare students for a particular test.

Evidence of the Problem

A modest number of studies in the United States have addressed the validity of gains in scores on high-stakes tests. The results of these studies are consistent and suggest the following conclusions:

The extant studies of this issue all examine the generalizability of gains. Scores on a test are used to support inferences about students' mastery of a domain or target of inference (Koretz and Hamilton, 2006). If performance on the high-stakes test (often called the focal test) generalizes to the larger domain, it should also generalize to other, lower stakes tests, often called audit tests, provided that the focal and audit tests are intended to support inferences about a similar target. Note that this logic does not require that the audit test be a superior measure in any sense when first used. Rather, it requires only that the focal and audit are intended to support similar inferences about achievement. If high stakes lead to inflation of scores on the focal test, the audit test will become a better gauge of achievement as a result, but this is not a function of the initial quality of the measures.

Evaluations of the testing program Kentucky Instructional Results Information System (KIRIS) implemented in Kentucky during the 1990s illustrate the sometimes dramatic failure of generalization of gains on high-stakes tests. The KIRIS program established targets for score increases for every school, provided financial rewards for schools that substantially exceeded their targets, and set sanctions for those that fell short. This case is a particularly useful example because the state designed the frameworks for its reading and mathematics tests to be consistent with the frameworks for the National Assessment of Educational Progress (NAEP), a periodic, low-stakes assessment administered on a sample basis to nationally and state-representative samples of students. This consistency of frameworks implies a similarity in intended inferences and makes the NAEP a clearly appropriate audit measure for the KIRIS tests.

Immediately upon implementation of the high-stakes testing program, KIRIS scores began rising very rapidly, but scores of Kentucky's students on the NAEP assessment did not keep pace. During the first 2 years of the

program, fourth-grade reading scores on KIRIS increased a staggering three-fourths of a standard deviation. (By way of comparison, the decline in test scores during the 1960s and 1970s that helped generate enthusiasm for test-based accountability in the US was typically about 0.03–0.04 standard deviation per year; Koretz, 1986.) During this same period, the scores of Kentucky's fourth-grade students on the NAEP reading assessment did not increase at all (Table 1; see also Hambleton *et al.*, 1995). The pattern in mathematics was a bit less extreme: very large gains on the state test, accompanied by gains on the NAEP that were roughly one-fourth as large (e.g., Table 2; see Koretz and Barron, 1998). For similar findings from other testing programs that entailed concrete sanctions, see Jacob (2005) and Klein *et al.* (2000).

An earlier study by Koretz *et al.* (1991) provided evidence that score inflation does not require concrete sanctions. This study examined the generalizability of performance in a large, urban district in which teachers and principals were pressured to raise scores on a norm-referenced, multiple-choice achievement test but in which there were no concrete sanctions or rewards for either teachers or students. By current US standards, one might call this a moderate-stakes testing program. In the mid-1980s, the district switched from one test to a very similar competing product, and performance dropped markedly. In third-grade mathematics, for example, mean performance dropped by approximately half an academic year (Figure 1). Four years later, mathematics performance on the new test had returned to its previous, high level. The researchers then administered the old test to a random sample of classrooms. While performance on the new test had climbed half an academic year, performance on the old test had declined by a like amount. In effect, teachers were substituting mastery of the details of one test for mastery of details of the other.

Table 1 Gains on high-stakes test Kentucky Instructional Results Information System (KIRIS) and audit test National Assessment of Educational Progress (NAEP) in Kentucky, 1992–1994, fourth-grade reading

	KIRIS	NAEP
Gain in scale scores	18.8	–1
Standardized gain	0.76	–0.03
Average annual divergence (SD)		0.40

Table 2 Gains on high-stakes test (KIRIS) and audit test (NAEP) in Kentucky, 1992–1996, eighth-grade mathematics

	KIRIS	NAEP
Gain in scale scores	23.7	4
Standardized gain	0.52	0.13
Ratio, KIRIS to NAEP		4.10
Average annual divergence (SD)		0.10

Inadequacy of Traditional Validation

Testing with serious consequences for individuals has a very long history. For example, the imperial examination system in China was instituted more than 1400 years ago (Niu, 2007), and matriculation examinations with serious consequences for students, such as the abitur in German Länder and the baccalaureate in France, became widespread in Europe during the nineteenth century. While the large-scale use of tests to monitor the effectiveness of schools or educational systems is new in many contexts, this function of testing is also not entirely novel, dating back at least to the 1840s in the United States, when a test was implemented for this purpose in Boston (Resnick, 1982).

It has long been recognized that high-stakes uses of tests might induce behaviors that undermine validity. For example, E. F. Lindquist, one of the most influential figures in the development of standardized achievement testing, wrote more than half a century ago:

The widespread and continued use of a test [that is a proxy for an unattainable measure of criterion behaviors] will, in itself, tend to reduce the correlation between the test series and the criterion series for the population involved. Because of the nature and potency of the rewards and penalties associated in actual practice with high and low achievement test scores of students, the behavior measured by a widely used test tends in itself to become the real objective of instruction, to the neglect of the (different) behavior with which the ultimate objective is concerned (Lindquist, 1951: 152–153).

The undermining of outcome measures by behavioral responses to measurement is by no means limited to educational testing. It is often labeled Campbell's law in the social sciences: "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures" (Campbell, 1979).

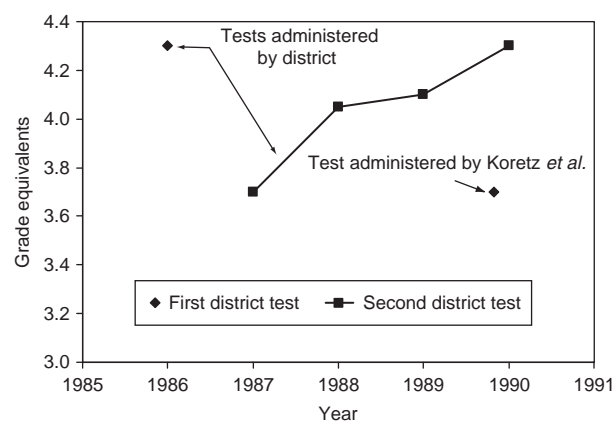


Figure 1 Performance on a moderate-stakes test and an audit test in third-grade mathematics.

Nonetheless, much of the core work of psychometrics, including test design, the development of analytic methods (such as procedures for linking), and empirical validation have proceeded until very recently without much consideration of this problem (Koretz, 2008). Of particular importance here is the inadequacy of traditional approaches to validation for evaluating the validity of inferences under high-stakes conditions (VIHS).

There are several reasons why traditional approaches to validation are insufficient to evaluate VIHS (see Koretz and Hamilton, 2006). With the exception of predictive validity evidence, conventional empirical validity evidence is cross-sectional and primarily correlational and is therefore insensitive to both changes over time and differences in mean levels of performance. This is not problematic when three conditions are met: the location of the scale is arbitrary, there is no reason to expect bias in mean scores, and inferences are cross-sectional. For example, validity evidence of this sort is not problematic when using a low-stakes diagnostic test to determine the normative level of a student's performance. However, these characteristics of traditional validation are critically important when evaluating VIHS. First, many of the most important inferences in high-stakes testing programs are about change over time. Moreover, a key validity question when stakes are high is the possibility of artifactual elevation of scores from inappropriate test preparation, and this bias can arise both in trends (exaggerated improvement from prior years) and in cross-sectional inferences (e.g., upwardly biased estimates of achievement relative to performance standards). Correlations, which are calculated from deviations from means, are insensitive to this bias. This was shown concretely in an evaluation of Kentucky's KIRIS testing program. Scores of high-school students on the state's KIRIS tests were compared to their scores on the ACT college-admissions test. In mathematics, the mean KIRIS score increased by about 0.7 standard deviation in 3 years, while the mean score on the ACT declined trivially. Nonetheless, the student-level and school-level correlations between the two tests remained quite stable over this period (Koretz and Barron, 1998).

Many of the empirical methods traditionally used for validation rest on the implicit assumption of stability in the relationships between the tested sample of material and the broader domain from which it is sampled. Achievement tests are created by defining the domain, creating a blueprint for the test, and then sampling the domain according to the blueprint. The resulting sample of content and skills is intended to represent the larger domain, much as the sample drawn for a political poll is intended to represent the larger electorate. Traditional validation methods are intended to evaluate the adequacy of the tested sample as a representation of the domain. However, as Lindquist (1951) warned, once the test becomes known, there is pressure to focus instruction on the tested sample

rather than the domain of skills and knowledge it is supposed to represent. This undermines the representativeness of the tested sample and causes performance on the tested sample to improve more rapidly than mastery of the domain it should represent. That is, it inflates scores. Traditional approaches to validation do not address this problem. They are well suited to evaluating the initial representativeness of the tested sample, but as typically applied, they are not sufficient to evaluate erosion of that representativeness over time.

Due to the tacit assumption of the stability of the relationship between tested and untested parts of the domain, traditional approaches to evaluation largely ignore behavioral responses to testing. Leaving aside cheating and failures to follow administrative and scoring procedures, behaviors by teachers or students that improve performance on the tested sample are assumed to improve mastery of the domain. This assumption becomes progressively less tenable as increasing stakes intensify the incentive to focus unduly on the particulars of the tested sample rather than on the broader domain the test is intended to represent. The challenge of establishing VIHS is to differentiate the meaningful, generalizable improvements from test-specific gains resulting from high stakes.

A Framework for Evaluating VIHS

Koretz *et al.* (2001) suggested a framework for conceptualizing the problem of VIHS (see also Koretz and Hamilton, 2006). It differs from a traditional view of test construction and validation in several respects – in particular, by highlighting incidental aspects of test construction and omissions from tests, both of which can be keys to score inflation.

A test can be viewed as a sample of performance elements, an intentionally general term that refers to all of the components of a test that influence performance on it. Some of these elements are substantive in that they are related to the intended inference. However, in building a test, it is also necessary to sample nonsubstantive elements that are not directly related to the intended inference. Choice of format is often, although not always, a nonsubstantive choice in this sense. Deciding whether the domain of algebra relevant to a given test includes factoring quadratics is a substantive decision; deciding whether to represent factoring with multiple-choice items is generally not. Other aspects of presentation and response requirements may also be largely or entirely nonsubstantive. Traditional test construction focuses primarily on the sampling of substantive elements, and traditional validation does even more so.

Performance elements are conceptually distinct, but they need not be empirically independent, particularly

in cross section. However, they should be treated as separate if differences among them are relevant to the inference and if they have the potential to vary independently over time. For example, even though algebra and geometry elements are usually highly correlated in cross section, it is straightforward to make them change somewhat independently over time – say, by teaching only algebra and not geometry.

The influence of each performance element on the test score is called its effective test weight. More formally, the test weight is the partial derivative of scores with respect to the performance element. These weights are partly a function of deliberate choices, such as decisions about the proportion of items allocated to a given topic and the discrimination of items selected for use. However, weights also can also be partly unintentional, and even nonsubstantive elements can have considerable weights. For example, a test author may repeat certain styles of presentation or fine details of content without a substantive rationale for doing so, and students' proficiency with these details will then have an appreciable impact on scores.

Which elements should be considered substantive depends on the inference based on scores. Consider the classic example of failure of generalizability offered by Shepard:

When students were asked [on the state's test] to add decimals in vertical format, the state percent passing was 86 percent. In horizontal format for identically difficult decimals, the percent passing was 46 percent. For subtraction of decimals in the two formats the passing rates were 78 percent and 30 percent, respectively (Shepard, 1988: 4).

If one purpose of the test is to compare students' relative mastery of vertical and horizontal formats, this aspect of presentation would be a substantive performance element. However, if the intent is simply to assess competence in addition and subtraction of decimals, the format difference would not be substantive, and if the test author had used only the vertical format, the result would have been a biased estimate of competence.

Scores are used to support inferences about a target of inference. The target of inference is another collection of performance elements, typically a bundle of knowledge and skills. The relative importance of an element to the inference is its inference weight. Inference weights are typically poorly specified. In the case of achievement testing, the target typically comprises a much larger set of performance elements than those sampled for construction of the test. Elements from the target that are not tested by a particular assessment are called implicit elements.

This can be represented as two vectors of weighted performance elements, where π_i represents a performance element, λ_i represents the effective test weight of that element, and ω_i represents its inference weight (Figure 2).

Test	Inference
$\lambda_1 \pi_1$	$\omega_1 \pi_1$
$\lambda_2 \pi_2$	$\omega_2 \pi_2$
\vdots	\vdots
$\lambda_j \pi_j$	$\omega_j \pi_j$
<hr/>	
$0 \cdot \pi_{j+1}$	$\omega_{j+1} \pi_{j+1}$
$0 \cdot \pi_{j+2}$	$\omega_{j+2} \pi_{j+2}$
\vdots	\vdots
$0 \cdot \pi_k$	$\omega_k \pi_k$
<hr/>	
$\lambda_{k+1} \pi_{k+1}$	$0 \cdot \pi_{k+1}$
$\lambda_{k+2} \pi_{k+2}$	$0 \cdot \pi_{k+2}$
\vdots	\vdots
$\lambda_n \pi_n$	$0 \cdot \pi_n$

Figure 2 Schematic of a test and a target of inference.

The top section in Figure 2 represents performance elements that are substantively important for the inference and are represented on the test. Both their test weights and their inference weights are therefore nonzero, although these two sets of weights need not be similar. The middle section, starting with π_{j+1} and ending with π_k , has implicit elements: elements that are important for the inference but are not represented on the test. In the case of most large-scale assessments, this second set is very large because a small sample of test items must represent a large domain of achievement. The final section, beginning with π_{k+1} , comprises performance elements represented on the test but not important for the inference. Many nonsubstantive elements fall into this section.

Validity can then be seen as the extent to which changes on the tested set of performance elements, as weighted by the test, justify inferences about changes on the target of inference, as weighted by users' inferences.

Behavioral Responses to Testing and VIHS

In a traditional framework that assumes low stakes, the impact of the large set of implicit elements can be minor, assuming that the tested elements are a representative sample of the entire domain. If the domain is truly unidimensional, the effects of the omissions are simple measurement error, and the tested sample need not be very large to keep this error reasonable in size. If the domain is modestly multidimensional, the impact of implicit elements is still likely to be modest because of the high cross-sectional correlation among related subdomains under low-stakes conditions.

In contrast, under high-stakes conditions, the impact of incomplete sampling can be very large. Behavioral

responses to testing can make the tested sample unrepresentative of the domain and can undermine the correlations among tested and untested parts of the domain. This can happen in several ways.

As teachers, principals, or students begin to learn what is and is not tested, they can shift instructional resources, such as class time, to focus on the tested elements at the expense of the untested. This entails shifting instructional resources from the elements in the middle of **Figure 2** to elements in the top section. As those in the middle section have nonzero inference weights, this process undermines the ability of performance on the tested sample to represent the target of inference. This is the process about which Lindquist warned years ago. Koretz *et al.* (2001) refer to this process as within-subject reallocation.

Educators and students can also reallocate resources between tested and untested domains – for example, cutting back on music and science, if they are not tested, to free up more time for mathematics and reading. This sort of between-subjects reallocation may be either desirable or not, but it is generally not a matter of VIHS. For example, suppose that a school serving low-performing students responds to a testing program that measures on mathematics and reading by eliminating science instruction in the fourth grade to allow an increase in instructional time in mathematics and reading. Suppose also that mathematics and reading test scores increase as a result. One might argue that the loss of science instruction is an undesirable outcome, or one might argue that mathematics and reading are so essential to the students' later well-being that this is a reasonable price to pay. However, in neither case does this shift in resources necessarily imply anything about the validity of inferences about improved proficiency in mathematics and reading.

Educators and students can also target their efforts on the third region in **Figure 2**, the elements that are given appreciable weight on the test but are not important to the inferences based on scores. Many of these are nonsubstantive elements, such as item format and other modes of presentation. An example is teaching students to use process of elimination to solve problems presented in multiple-choice format, which can produce gains in performance that appear only on multiple-choice tests. When these responses focus on nonsubstantive elements, Koretz *et al.* (2001) refer to them as nonsubstantive coaching. These too can generate increases on the tested vector that are larger than are warranted by increases on the target.

Yet another type of behavioral response that can inflate scores takes advantage of unimportant, sometimes unintentional, substantive details of the focal test. Koretz *et al.* (2001) refer to instruction that targets these details of the test as substantive coaching. One can think of the substantive sampling required for construction of a test as having multiple stages, although they need not be distinct in practice. For example, a state's standards or curriculum

framework may first specify that students should have familiarity with the properties of common polygons and then may add detail, such as this from the Massachusetts mathematics framework for grades 5 and 6:

Students engage in problem solving, communicating, reasoning, connecting, and representing as they: Identify polygons based on their properties, including types of interior angles, perpendicular or parallel sides, and congruence of sides, e.g., squares, rectangles, rhombuses, parallelograms, trapezoids, and isosceles, equilateral, and right triangles (Massachusetts Department of Education, 2000: 42).

Constructing a test, however, requires even further sampling. Which polygons should be represented? Should the test include only regular ones, with six or fewer sides? Irregular but symmetrical polygons? Asymmetrical ones? Should the items require calculation of areas and perimeters? Estimation? Discussion of the relationship between the number of sides and the magnitude of internal angles?

Thus, there can be layers of increasingly fine-grained sampling from within a broad substantive element, eventually reaching the point at which the choices are no longer important for the inference. One could label the finest level of detail relevant to the target as a conceptual element. Often, these conceptual elements can be tapped by a number of more finely grained measurable elements, both substantive and nonsubstantive. This is illustrated in **Figure 3**, which displays the first element in **Figure 2**, π_1 , as encompassing six smaller elements, two represented on the test and the target, three represented in the target but not the test, and one represented in the test but not the target.

If the choices among these fine-grained elements are repeated over time, they too provide an opportunity to tailor instruction to the specifics of the tested sample and to undermine its representation of the target of inference. For example, a Princeton Review publication marketed to prepare students for one state's mathematics assessment notes that "one triangle rule that is often tested [on the state's test] is the *third side* rule: the sum of every two sides of a triangle must be greater than the third side" (Rubinstein, 2000: 52). There is nothing unreasonable about selecting this particular measurable element for inclusion, but its recurrent use – in lieu of other measurable elements

Test	Inference
$\lambda_{11}\pi_{11}$	$\omega_{11}\pi_{11}$
$\lambda_{12}\pi_{12}$	$\omega_{12}\pi_{12}$
$0\cdot\pi_{13}$	$\omega_{13}\pi_{13}$
$0\cdot\pi_{14}$	$\omega_{14}\pi_{14}$
$0\cdot\pi_{15}$	$\omega_{15}\pi_{15}$
$\lambda_{16}\pi_{16}$	$0\cdot\pi_{16}$

Figure 3 Schematic of a single conceptual element, π_1 .

representing the same concept – provides an opportunity for substantive coaching and score inflation.

Finally, educators and students can simply cheat. Whatever the method of cheating chosen—for example, revealing items in advance, modifying answers after the testing period, etc. – the result is necessarily score inflation.

New Directions for Research, Evaluation, and Test Design

It is clear that the validity of inferences about proficiency based on tests used for high-stakes purposes cannot be taken for granted, even after careful validation using traditional methods. To be sure, the sparse research to date does not indicate that score inflation is inevitable. Despite the pervasiveness of test-based accountability in the US, trends on some state tests have been reasonably consistent with those on NAEP (Ho and Haertel, 2006), and research has not clarified the factors – possibly, characteristics of students, schools, test, or accountability programs – that predict the degree of inflation. Nonetheless, the extant research does suggest that this problem is not rare. Moreover, extant research indicates that score inflation can be both rapid and severe, in some instances dwarfing biases one might expect from other sources more commonly addressed in traditional validation studies.

These findings suggest the need for more widespread and routine investigation of VIHS. However, the field has only begun exploring how this investigation should be done. The obvious choice used in most studies to date is to compare trends on focal and audit tests, and this should be done when it is practical. However, this is only feasible when there is a suitable audit test in place – one that is administered to a representative sample or on a census basis in the right grades in the same years, that is intended to support similar inferences, and that can be assumed to be relatively free of inflation itself. A suitable audit measure is often lacking, and in such cases, the burden of evaluating VIHS would greatly increase because it would require either administering an additional audit test or adding an audit component to an ongoing operational assessment.

Moreover, while a simple comparison of trends on focal and audit tests is sufficient to identify instances in which large disparities in trends suggest score inflation, it is a blunt tool. Modest disparities in trends may arise even in the absence of any inflation if the focal and audit tests are designed to support somewhat different inferences. Note, however, that differences between tests are only relevant in this context to the extent that they imply different targets of inference; performance gains, if meaningful, should generalize across substantively irrelevant details of test construction (Koretz, 2007). Differences in intended inferences could bias estimates of score inflation

in either direction (Koretz and Barron, 1998). Moreover, if the focal and audit test happen to share unnecessary particulars – such as substantively irrelevant aspects of presentation or substantively unimportant details of content – there is a chance that inflation generated by behavioral responses to the focal test may contaminate the audit test as well, leading to an underestimate of score inflation. Therefore, simple comparisons of trends may be insufficient to detect modest score inflation, and they are not sufficient for estimating precisely the mix of real gains and inflation when both occur.

The limitations of simple comparisons with extant audit tests pose a design challenge: the need for better and less burdensome audit measures. Ideally, an audit test should be designed to maximize similarities in the intended inferences while minimizing unnecessary similarities in the particulars of test design. In several respects, adding an audit component to an ongoing focal test might be superior to administering an independent audit test. It would remove motivational differences, lessen disruption, and most likely cost less. This might be done, for example, by adding a matrix-sampled block of audit items to the focal test. However, as yet there has been little discussion of these design issues and no empirical evaluation of alternative approaches to this problem.

The crudeness of simple comparisons between focal and audit tests also poses an analytical challenge: the need for analytical methods to be better able to distinguish between meaningful gains and score inflation when they co-occur. Differential item functioning (DIF) analysis might be adapted for this purpose (Koretz and McCaffrey, 2005). One approach might be to categorize items in a manner relevant to VIHS – for example, items that have familiar particulars and audit items that have fewer – and then use DIF analysis to compare performance gains on the sets. One might also do the reverse: use DIF analysis to identify clusters of items showing unusually small or large performance changes and then examine their characteristics *post hoc*. However, these analytical challenges have so far received very little attention.

The challenge to validity posed by high-stakes uses of tests has been recognized since the early days of the science of achievement testing, but to date, it has garnered relatively little attention. Ongoing changes in the uses of large-scale achievement tests have made this issue increasingly important, and recent research has shown the severity of the resulting threats to validity. However, empirical evaluation of VIHS remains rare, and the measurement field still faces the challenge of finding more effective and less burdensome ways of conducting the needed validation.

See also: Ability Testing; Educational Measurement: Overview; Impact of Assessment on Classroom Practice; Impact of Assessments on Classroom Practice; Student Test Results in School Accountability; Testing in History.

Bibliography

- Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning* 2, 67–90.
- Hambleton, R. K., Jaeger, R. M., Koretz, D., et al. (1995). *Review of the Measurement Quality of the Kentucky Instructional Results Information System, 1991–1994*. Frankfort, KY: Office of Education Accountability, Kentucky General Assembly.
- Ho, A. D. and Haertel, E. H. (2006). Metric-free measures of test score trends and gaps with policy-relevant examples. *CSE Technical Report 665*. Los Angeles, CA: Center for the Study of Evaluation, University of California.
- Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics* 89, 761–796.
- Klein, S. P., Hamilton, L. S., McCaffrey, D. F., and Stecher, B. M. (2000). What do test scores in Texas tell us? *Issue Paper IP-202*. Santa Monica, CA: RAND. <http://www.rand.org/publications/IP/IP202> (accessed July 2009).
- Koretz, D. (1986). *Trends in Educational Achievement*. Washington, DC: Congressional Budget Office.
- Koretz, D. (2007). Using aggregate-level linkages for estimation and validation: Comments on Thissen and Braun & Qian. In Dorans, N. J., Pommerich, M., and Holland, P. W. (eds.) *Linking and Aligning Scores and Scales*, pp 339–353. New York: Springer.
- Koretz, D. (2008). Further steps toward the development of an accountability-oriented science of measurement. In Ryan, K. E. and Shepard, L. A. (eds.) *The Future of Test-Based Educational Accountability*, pp 71–91. Mahwah, NJ: Erlbaum.
- Koretz, D. and Barron, S. I. (1998). *The Validity of Gains on the Kentucky Instructional Results Information System (KIRIS)*. MR-1014-EDU. Santa Monica, CA: RAND.
- Koretz, D. and Hamilton, L. S. (2006). Testing for accountability in K-12. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn., pp 531–578. Westport, CT: American Council on Education/Praeger.
- Koretz, D., Linn, R. L., Dunbar, S. B., and Shepard, L. A. (1991). The effects of high-stakes testing: Preliminary evidence about generalization across tests. In Linn, R. L. (chair) *The Effects of High Stakes Testing*, symposium presented at the annual meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago, April.
- Koretz, D. and McCaffrey, D. (2005). Using IRT DIF methods to evaluate the validity of score gains. *CSE Technical Report 660*. Los Angeles, CA: Center for the Study of Evaluation, University of California.
- Koretz, D., McCaffrey, D., and Hamilton, L. (2001). *Toward a Framework for Validating Gains Under High-Stakes Conditions*. CSE Technical Report 551. Los Angeles: Center for the Study of Evaluation, University of California.
- Lindquist, E. F. (1951). Preliminary considerations in objective test construction. In Linquist, E. F. (ed.) *Educational Measurement*, pp 119–158. Washington, DC: American Council on Education.
- Massachusetts Department of Education (2000). *Massachusetts Mathematics Curriculum Framework*. Malden, MA: Massachusetts Department of Education.
- Niu, W. (2007). Western influences on Chinese educational testing. *Comparative Education* 43(1), 71–91.
- Resnick, D. P. (1982). History of educational testing. In Wigdor, A. K. and Garner, W. R. (eds.) *Ability Testing: Uses, Consequences, and Controversies, Part II*, pp 173–194. Washington, DC: National Academy Press.
- Rubinstein, J. (2000). *Cracking the MCAS Grade 10 Math*. New York: Random House.
- Shepard, L. A. (1988, April). Should instruction be measurement driven? A debate. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Relevant Website

<http://ipea.hmdc.harvard.edu> – International Project for the Study of Educational Accountability Systems (IPEA).

Validity of Educational Indicators

B P M Creemers, University of Groningen, Groningen, The Netherlands

L Kyriakides, University of Cyprus, Nicosia, Cyprus

© 2010 Elsevier Ltd. All rights reserved.

The Meaning of Educational Indicators

Recent years have witnessed an explosion of interest in educational indicators (Bryk and Hermanson, 1993). This topic has been the focus of a number of reports, papers, study commissions, and national and state panels (e.g., Fitz-Gibbon, 1996; Gray and Wilcox, 1994; Nuttall, 1994; OECD, 1995; Salganik, 1994; Seltzer *et al.*, 2003; Smith, 1988). Most commonly, an educational indicator is expected either to assess or even to be related to a desired outcome of the education system. In this article, we try to clarify the meaning of educational indicators by taking into account the variety of indicators used by several organizations such as the Organization for Economic Cooperation and Development (OECD), the US Department of Education, and the National Science Foundation. The most basic and elementary definitions contend that indicators should describe key aspects of schooling (Selden, 1994).

This definition implies that indicators are statistics that are expected to provide information about significant features of an educational system. For example, indicators often describe how students perform on a national test, what is being taught, or how much is being spent on education (OECD, 1995). Others note that for a statistic to be an indicator, there must be a standard against which it can be judged. For example, comparisons over time offer dates as reference points; comparison among institutions or nations have built-in reference points as well. However, many indicators require more than one reference point. For example, a decline in scores of a national test over a number of years tells us more than a simple statement of the magnitude of the decline. With the addition of reference points, indicators may become more meaningful, more precise, and more useful.

Other researchers claim that indicators should act as vital signs regarding the health of the educational system (e.g., Nuttall, 1994; Oakes, 1989). There are two aspects embedded in this conceptualization. First, by serving as the vital signs, indicators permit the monitoring of education which is typically understood as an ongoing activity. Second, by providing an assessment of the health of a system, indicators provide a basis for making value judgments. Value judgments can be understood as overall assessments that are rather subjective, for there exist no definitive guideposts to determine the health of a system.

Other definitions go even a step further by invoking a policy criterion. Specifically, it is claimed that a statistic

can be considered an indicator only if it is policy relevant (Shavelson *et al.*, 1989). This definition implies that policy-makers want to know what policy levers they can pull in order to improve performance. Thus, this definition provides an instrumental usage of indicators (Bryk and Hermanson, 1993).

In this section, five characteristics of educational indicators have been identified. As a generic synthesis of the available definitions, we support that educational indicators are statistics that describe key aspects of schooling which permit the evaluation and monitoring of schools, teachers, programs, and students. From these activities, value judgments of the health of an educational system or parts of it can be derived and policy-relevant information which contributes to the improvement of the effectiveness of the system should emerge. This definition implies that educational indicators should provide information about the quality of education associated with the effectiveness and efficiency of the system or parts of it such as teachers, schools, and educational regions. However, although many educational indicators are used in research, practice, and policy, their ability to provide valid information about the quality of education could be questioned.

In this article, it is not only argued that validation of educational indicators is needed but it is also claimed that our definition of indicators implies that the validation process should mainly be concerned with the extent to which information provided by an indicator is able to demonstrate effects of the factors addressed upon outcomes of schooling. For example, when we treat the amount of teachers with a certified degree as an indicator, we may assume that subject and/or pedagogical knowledge is important for the quality of teaching and is related to student outcomes. However, unless we demonstrate through either empirical studies and/or quantitative syntheses of studies that there is a relationship between the degree awarded and student outcomes, this indicator cannot help policymakers take decisions on how to improve the effectiveness of the educational system. Such an approach may produce fewer indicators but may help policymakers have a clear understanding on how each of them affects achievement standards.

Since our proposed methodology for testing the validity of prospective indicators places a lot of emphasis on the impact of indicators upon student outcomes of schooling, we clarify below how we conceptualize quality of education. It is supported that ultimately, the criterion

for success in education is the attainment of the aims, goals, and objectives of the educational system. Since schools are places where primarily learning takes place, the objectives of education are primarily student-learning outcomes. These can be found in the cognitive domain and also in other areas such as the affective, the social, and the esthetic domains. It is recognized that student learning in noncognitive areas is determined overwhelmingly by other actors in the society whereas the cognitive domain is determined less by other social agents. This implies that schools have a specific role in the cognitive domain and that consequently, objectives in this area are crucial for the educational system in general. Moreover, achievement of cognitive outcomes determines to some extent achievement in other domains, like motivation and well-being (Knuver and Brandsma, 1993). However, this does not mean that education should be restricted to cognitive objectives since only a partial relationship exists between achievement of cognitive and noncognitive domains (Kyriakides, 2005). Moreover, schools may act as social agents and, being a part of society, they can also contribute partially to these areas. They can provide a social and esthetic environment in which social behavior and esthetic attitudes can be developed. Therefore, schools and teachers should be supported in such a way that objectives are reached and educational quality becomes a fact. In this context, research can offer insight into which indicators contribute to student results and help policymakers take decisions on how to improve educational standards.

The Validation of Indicators: A Proposed Methodology

In the first section of this article, the importance of investigating the impact of indicators upon achievement of both cognitive and noncognitive aims of schooling has been stressed. In this section, suggestions on building a methodology for testing different types of validities of prospective indicators are provided:

1. *Demonstrate the relation of prospective indicators with student achievement gains.* Our first suggestion is concerned with the importance of investigating the impact of prospective indicators on student-achievement gains. This approach implies that fair indicators of school performance need to measure the contribution of teaching to progress made by students at school (Fitz-Gibbon, 1996). The use of value-added approach is attributed to the fact that variations in the final test results of schools reflect partly the educational attainment of students when they enter the school (Kyriakides, 2002). Therefore, information gathered from value-added assessment is more valid in exploring the effectiveness of an educational system or parts of it (e.g., school units).
2. *Search for educational indicators which operate at different levels.* The last point associated with the advantages of using multilevel approaches to identify the effect of indicators on student-achievement gains reveals that we could generate educational indicators which operate at different levels. However, it cannot be claimed that any factor associated with student achievement should be treated as an educational indicator. For example, school-admittance policy may be found to be related with student achievement but this factor refers mainly to the input of the school since the existence of an admittance policy has a direct and strong effect on the prior knowledge of students entering the school. Therefore, the school-admittance policy cannot be treated as an educational indicator since it does not reveal how the functioning of schools can contribute to student-achievement gains. On the other hand, an educational indicator could be generated based on the fact that empirical studies reveal that opportunity to learn at student level is associated with student achievement (Creemers and Kyriakides, 2008). Such a finding reveals that schools could take actions to improve opportunity to learn. Therefore, to generate a relevant educational indicator we need

The use of multilevel analysis to search for the impact of each prospective indicator upon student achievement gains is also recommended. This is due to the fact that multilevel analysis explicitly models the manner in which students are grouped within classes or schools and, therefore, has several advantages in relation to our attempt to search for the impact of an indicator on student-achievement gains. First, multilevel analysis takes into account the existence of hierarchically structured data and the variability associated with each level. One may draw erroneous conclusions if any of these sources of variability is ignored (Opdenakker and Van Damme, 2000). Second, multilevel analysis provides a means of partitioning the outcome variable's variance into different levels (within and between units). Third, it yields better-calibrated estimates for the variance of standard errors (Goldstein, 2003). Fourth, it offers a single framework that combines the information within and across units to produce more accurate explanations and outcomes. Finally, clustering information provides correct standard-error confidence intervals and significance tests, which are more conservative than the traditional ones that are obtained simply by ignoring the presence of clustering (Snijders and Bosker, 1999). By allowing the use of covariates measured at any of the levels of a hierarchy, we can explore the extent to which the differences in average-achievement results between schools can be accounted for by factors related either to the school, to class, or to student characteristics (i.e., to specific prospective indicators).

to demonstrate through empirical studies that this indicator contributes in increasing opportunity to learn at student level and due to that, student achievement is also increased.

3. *Search for both direct and indirect effects of each prospective indicator on student achievement.* Given that an indicator is expected to help policymakers take decisions on how to improve the effectiveness of the educational system, it is considered important not only to provide information associated with an indicator but also to help policymakers and other stakeholders understand whether, how, and why an indicator is associated with student-achievement gains. In order to achieve this aim, we should search for both direct and indirect effects of each indicator on student achievement. For example, some context-level and school-level indicators may influence the teaching and learning situation which is associated with student achievement and thereby such indicators may have indirect effect on student achievement. Searching for both direct and indirect effects of educational indicators upon student achievement helps policymakers take decisions on how to design intervention programs to improve effectiveness by making sure that changes in the functioning of an indicator will affect intermediate variables and due to that student achievement could be improved. For example, a reform program aiming at the reduction of the class size may not have any effect on student achievement unless it is designed in such a way to help teachers improve their teaching practice (e.g., increase teacher–student and/or student–student interactions). Similarly, providing resources for learning could improve teaching if it is designed in a way to help the teachers and the school to provide more learning opportunities to students or to contribute in the professional development of teachers. Thus, our suggestion for searching for both direct and indirect effects of educational indicators on student achievement has as its ultimate aim to help policymakers understand how and why an indicator is associated with achievement.
4. *Find out whether there are nonlinear relations between indicators and student achievement.* In order to expand our understanding of the impact of an indicator on student achievement, we should also find out whether a nonlinear relation between an indicator and achievement exists. This argument is supported by the fact that meta-analyses of the effect of some effectiveness factors upon student achievement revealed that although they have been perceived as factors affecting teacher or school effectiveness and/or have been used as educational indicators, the research evidence is problematic. For example, teacher subject knowledge is widely perceived as a factor affecting teacher effectiveness (Scriven, 1994) but teachers' subject knowledge,

regardless of how it is measured, has rarely correlated strongly with student achievement (Darling-Hammond, 2000). The explanation may be, as Monk (1994) reported, that the relationship is nonlinear: a minimal level of knowledge is necessary for teachers to be effective, but beyond a certain point, a negative or no relation may occur. Similar findings have been reported for the impact of classroom emotional climate and teacher management upon effectiveness. A negative emotional climate usually shows negative correlations, but a neutral climate is at least as supportive as a warm climate.

Our suggestion to search for nonlinear relations does not only imply that more complex statistical techniques should be used in analyzing data concerned with the relation of an indicator and student achievement but also that the optimal points of the functioning of an indicator should be identified (Creemers and Kyriakides, 2008). In the case of education, we give emphasis to the existence of inverted-U curvilinear relations since such relations reveal that there is an optimal point for the function of a specific factor. After the optimal point, a negative relation with achievement exists, and thereby the identification of the optimal point has important implications for improving educational practice. Obviously, other nonlinear relations might exist. In case that there is more than one optimal point, the question can be raised about the efficiency of application of a certain factor after the first optimal point.

5. *Search for relations between indicators which operate at the same level.* There is also a need to carefully examine the relationships between the various indicators which were found to be related with student achievement and operate at the same level. The identification of relations between indicators may reveal a grouping of factors that make teachers and schools effective. Therefore, providing policymakers with information on how educational indicators operating at the same level are related to each other may generate a more comprehensive picture of the effective characteristics of education. As a consequence, comprehensive strategies for improving effectiveness could be developed (Creemers and Kyriakides, 2006). The results of a study attempting to test the validity of Walberg's (1984) educational productivity model reveal more complex indirect relationships between factors affecting student learning (Reynolds and Walberg, 1990) and provide some support to our suggestion to search for relations between indicators.
6. *Find out whether each indicator is generic or differential.* Although educational indicators are sometimes treated as generic in nature, their impact on different groups of students/teachers/schools may vary. Therefore, a critical question for each prospective indicator is associated with the extent to which it can be treated

as generic or differential. Although we could use different dimensions to investigate whether an indicator is differential (see Campbell *et al.*, 2004), we give more emphasis to the extent to which indicators are able to explain variation on student achievement across countries. This is attributed to the fact that the educational debates that are currently in evidence in many societies concerning the appropriate means of raising educational standards are often based upon simplistic transplants of knowledge from one educational system to another by making use of information gathered by educational indicators held by different countries without any detailed acknowledgment in the educational-policy debate as to the possible context specificity of the apparently effective policies in the original societies utilizing them.

Such an approach in testing the validity of educational indicators reveals the importance of conducting comparative studies in order to understand much more about why some indicators explain effectiveness across countries while others do not. For example, variables associated with inputs such as money spent in education may be associated with the improvement of effectiveness in some countries but not in others. This variation in what works is likely to in its turn generate a need for more sensitive theoretical explanations than those at present on offer, since the variation by national contexts forces the development of more context-specific explanations of how indicators can be used by policymakers to improve practice. Therefore, there is a need to identify effectiveness factors that are present in different educational contexts and generate relevant generic educational indicators. Moreover, factors that are unique to specific countries as well as factors that operate differently in different educational settings should be highlighted.

Beyond conducting comparative studies to find out whether a prospective indicator is generic, meta-analyses of national studies could be undertaken. Using multilevel approaches we can analyze the observed effects of each study and the sources of variances among the findings emerging from different studies (Raudenbush and Bryk, 1985). Therefore, differences in reported effect sizes of an indicator on student achievement can be modeled as a function of study characteristics such as the differences in the countries where the study was conducted or the type of outcomes used to measure student achievement. As a consequence, the extent to which an indicator has a differential effect could be identified.

7. *Investigate the interpretive validity of each prospective indicator.* Traditional conceptualizations of validity treated validity in terms of three distinct facets, or evidential areas: (1) construct validity, (2) criterion validity, and (3) content validity. Each of these facets remains important for contemporary testing. Moreover, all of them involve

scientific generalization. However, contemporary discussion of test validity emphasizes that validity does not refer to a property of a test; rather it refers to the meaningfulness and appropriateness derived from test scores (American Educational Research Association *et al.*, 1999). Although Popham (1997) argued for a limited definition of validity and treated validation as an objective, scientific concern, separate from disputes over the consequences of testing, Linn (1997) and Shepard (1997) favored a broader conception of validity which would include the consequences of test use and the descriptive interpretation of test scores. An emphasis on content and criterion-related questions as well as the strong program of construct validity (Cronbach, 1988) may lead us to underestimate the importance of issues dealing with the consequences of using an educational indicator. However, in real-world applications, we want the desirable consequences of using an indicator or any other instrument to be more important than its negative consequences. Therefore, teachers and policymakers, as end users of educational indicators, should contribute a distinctive perspective on validity of indicators, referred to as inferential validity (Kyriakides, 2004). This implies that the validation of a prospective educational indicator should be concerned not only with demonstrating that it has an effect on student achievement but also with finding out whether statistical figures associated with an indicator are clear and informative to different stakeholders and help them take decisions on how to improve educational practice. Survey studies looking at the perceptions of different stakeholders about the meaning of each indicator and on how each indicator can be used to improve educational effectiveness should be conducted. In addition, results of experimental/evaluation studies concerned with the use of prospective indicators to improve the effectiveness of an educational system or parts of it could help us test the inferential validity of an indicator.

Conclusions: The Main Features of the Proposed Methodology

The proposed methodology for testing the validity of prospective indicators attempts to give answers to three basic questions. First, our attempt to search for relations between an indicator and student-achievement gains reveals whether an indicator is important in predicting the quality of education in relation to student-learning outcomes. Second, beyond demonstrating that an indicator is associated with student achievement, it is considered important to explain why the indicator is able to predict the quality of education. This question reveals the role that educational research and theories associated with educational effectiveness can play in generating and testing the validity of indicators. Unless theoretical explanations of

the impact of indicators are given, neither can their construct validity be tested, nor can the use of information gathered from each indicator to improve policymaking be ensured. In this article, it was claimed that in order to provide a clear answer to this question, we should search for both direct and indirect effects of each indicator on student achievement. We should also find out whether linear or nonlinear relations between indicators and student achievement exist and whether each indicator is generic or differential. Finally, relations among indicators operating at the same level could be identified in order to help us establish comprehensive strategies for improving effectiveness.

The third question raised here is concerned with the actions that teachers and policymakers take as a consequence of the information provided by an indicator. Since indicators are expected to help schools, policymakers, and other stakeholders to design action plans for improving the quality of the system at different levels, it is argued that we should find out how information emerging from an educational indicator is used.

It is finally pointed out that the proposed methodology for testing the validity of prospective indicators does not imply that a specific sequence in providing answers to the above three questions is more desirable than others. It is possible that empirical results providing answers to the first question may help us establish a theoretical explanation of the impact of a prospective indicator and thereby give answers to the second question. We may also initially provide answers to the second question by developing or using specific theories on how some prospective indicators are associated with student achievement and then generate evidence to test these theories and demonstrate the relation of these indicators with student achievement. The latter approach is in line with the typical way used to test the construct validity of measurement instruments but the possibility of using this approach is mainly based on the extent to which educational research can provide us with strong theories explaining why some teachers/schools/systems are more effective than others. Finally, we should consider the possibility that initially studies looking on how schools or systems make use of evaluation data to improve their practice provide answers to the third question and at a later stage we may conduct research to answer the first two questions concerned with the construct validity of indicators.

Establishing Criteria for Valid Indicators in Education

Throughout the literature on educational indicators, scholars provide lists of criteria for choosing indicators. The single most frequently cited criterion is in regards to the technical quality of the data including the issues

of reliability and validity (Blank, 1993; Nuttall, 1994; Shavelson *et al.*, 1989; Smith, 1988). However, our attempt to provide suggestions on testing the validity of each prospective indicator implies that valid educational indicators should meet the following five criteria. First, our emphasis to testing the interpretive validity of each prospective indicator implies that a valid indicator should be informative to the public and help policymakers and other stakeholders develop strategies to improve educational standards. It is easy enough to spot an educational indicator, though interpreting it accurately is often more difficult and thereby emphasis on testing the interpretive validity of indicators is given in this article. It is apparent, as Selden (1994) observed, that defining educational indicators and indicator systems lies principally in determining how they are used. Different uses of indicators offer different forms of control over educational systems but we give more emphasis to the extent to which indicators help policymakers and other stakeholders develop an evidence-based approach to improve quality of education.

Another important criterion of a valid indicator that emerged from our methodology of testing the validity of indicators is concerned with the extent to which empirical support to the ability of an indicator to influence quality of education has been provided. For this reason, we emphasize the importance of demonstrating that an indicator has an impact (direct or indirect) upon student-achievement gains.

Third, beyond the empirical support, an indicator should also get theoretical support that helps policymakers and other stakeholders understand how and why an indicator is associated with achievement (Benjamin, 1996). With this criterion we are facing with the challenge of developing theoretical models illustrating why some teachers/schools/educational systems are more effective than others. Although at the moment there is no single agreed-upon model of educational effectiveness, it has to be acknowledged that during the last 10 years, attempts to generate comprehensive models of educational effectiveness have been made (e.g., Creemers, 1994; Scheerens, 1992; Stringfield and Slavin, 1992) and some empirical support to the validity of these models has been provided (e.g., de Jong *et al.*, 2004; Kyriakides, 2005). Moreover, a dynamic model of educational effectiveness addressing some of the major weaknesses of the current models has been recently developed (Kyriakides, 2008). By making use of these models we may generate valid indicators which are not only informative to policymakers but are supported both empirically and theoretically as well.

A fourth criterion for valid educational indicators is that they are not only policy relevant but they also include aspects that can be manipulated by policies or programs (Nuttall, 1994; Shavelson *et al.*, 1989; Smith, 1988; Visscher, 2001). This implies that indicators should not only refer to input variables especially since part of

the improvement of modeling educational effectiveness is the development of models which give more emphasis to process rather than simply to input and output variables.

Finally, it was argued that we should find out whether an indicator is generic or whether it has a differential impact on achievement. Making clear whether an indicator is generic helps policymakers make better use of the results of comparative studies to improve effectiveness. In addition, information emerging from differential indicators could be useful for establishing reform policies on providing equal opportunities and helping schools and teachers reduce unjustifiable differences in outcomes of schooling (Campbell *et al.*, 2004).

Implications for Policy and Research

The last section of this article is concerned with implications of generating valid indicators for policy and educational research. Although policymakers have their own agenda which is not necessarily fully determined by empirical facts, in view of rational decision making, policymakers are expected to take into account empirical validated knowledge about the political issues at stake. Therefore, the development of valid indicators may help policymakers develop an evidence-based approach for building educational policy. Moreover, the understanding of the impact of an indicator on student achievement is essential not only for the content of a reform policy associated with this indicator but also with the process of disseminating a reform policy. As a consequence, the proposed methodology for testing the validity of prospective indicators is expected to provide information concerned with whether, how, and why an indicator is associated with student achievement. Specifically, our understanding of effective education could be expanded if we answer questions associated with the impact of an indicator on student achievement such as whether it has direct and/or indirect impact on student achievement and on whether its relation with outcomes is linear or nonlinear. Essential for improving educational practice is also to find out whether indicators operating at the same level are not only related to student outcomes but are also related to each other. Finally, finding out whether an indicator is generic or differential may help policymakers understand that results achieved in one country cannot be replicated in another country by a simplistic transplant of some factors, which seem to be imported without any detailed knowledge of possible contextual factors that might explain how factors that work in one country may be ineffective in another country. What is needed is not borrowing policies but generating valid generic and differential educational indicators which reveal those characteristics of the educational systems that explain variation in achievement within and across countries.

Such an approach may allow us to build an evidence-based approach in introducing reform policies.

The proposed methodology for testing the validity of educational indicators has implications for conducting studies which could generate evidence about the relation of each indicator with student-achievement gains. We also draw suggestions on conducting quantitative syntheses to test the validity of prospective indicators especially since meta-analyses help us find out whether an indicator has any differential impact. Therefore, the proposed methodology has significant implications for conducting systematic research which can help us develop the theoretical framework of educational effectiveness. But the proposed methodology does not have implications only for conducting studies concerned with the validity of prospective indicators. Generating valid indicators may also have significant implications for redesigning comparative studies such as those conducted by the International Education Association (IEA) and the OECD. Secondary analyses of Trends in International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA) studies revealed that although a large number of variables at student, classroom, and school level were taken into account, these variables were able to explain only a small percentage of student-achievement variance (Kyriakides, 2006). It could be assumed that if the instruments of these studies had provided more information concerned with valid indicators, more variance would have been explained. This holds true especially at the country level, since variables associated with different educational policies that may affect the school or teacher level were absent. Therefore, the establishment of valid educational indicators could contribute to extending the scope of IEA and OECD studies and at the same time could aid in developing a better understanding of education by which all countries participating in these studies would benefit.

See also: Educational Measurement: Overview; IEA: Globalization and Assessment; National Assessment Programs: The Example of the U.S. National Assessment of Educational Progress; Studies of School Improvement in Developing Countries; The International system for Teacher Observation and Feedback: A Theoretical Framework for Developing International Instruments.

Bibliography

- American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.
- Benjamin, M. (1996). The design of performance indicator systems: Theory as a guide to relevance. *Journal of College Student Development* 37(6), 623–630.
- Blank, R. K. (1993). Developing a system of education indicators: Selecting, implementing, and reporting indicators. *Educational Evaluation and Policy Analysis* 11(2), 65–80.

- Bryk, A. S. and Hermanson, K. L. (1993). Educational indicator systems: Observations on their structure, interpretation, and use. *Review of Research in Education* **19**, 451–484.
- Campbell, R. J., Kyriakides, L., Muijs, R. D., and Robinson, W. (2004). *Assessing Teacher Effectiveness: A Differentiated Model*. London: Routledge/Falmer.
- Creemers, B. P. M. (1994). *The Effective Classroom*. London: Cassell.
- Creemers, B. P. M. and Kyriakides, L. (2006). A critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement* **17**(3), 347–366.
- Creemers, B. P. M. and Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice and Theory in Contemporary Schools*. London: Routledge.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In Wainer, H. and Braun, H. (eds.) *Test Validity*, pp 3–17. Hillsdale, NJ: Erlbaum.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives* **8**(1). <http://epaa.asu.edu/epaa/v8n1> (accessed August 2009).
- De Jong, R., Westerhof, K. J., and Kruiter, J. H. (2004). Empirical evidence of a comprehensive model of school effectiveness: A multilevel study in mathematics in the 1st year of junior general education in the Netherlands. *School Effectiveness and School Improvement* **15**(1), 3–31.
- Fitz-Gibbon, C. T. (1996). *Monitoring Education: Indicators, Quality and Effectiveness*. London: Cassell.
- Goldstein, H. (2003). *Multilevel Statistical Models*, 3rd edn. London: Edward Arnold.
- Gray, J. and Wilcox, B. (1994). Performance indicators: Flourish or perish? In Riley, K. A. and Nuttall, D. L. (eds.) *Measuring Quality: Education Indicators – United Kingdom and International Perspectives*, pp 69–86. Bristol, PA: Falmer.
- Kruver, J. W. M. and Brandsma, H. P. (1993). Cognitive and affective outcomes in school effectiveness research. *School Effectiveness and School Improvement* **4**(3), 189–204.
- Kyriakides, L. (2002). A research-based model for the development of policy on baseline assessment. *British Educational Research Journal* **28**(6), 805–826.
- Kyriakides, L. (2004). Investigating validity from teachers' perspective through their engagement in large-scale assessment: The emergent literacy baseline assessment project. *Assessment in Education: Principles, Policy and Practice* **11**(2), 143–165.
- Kyriakides, L. (2005). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement* **16**(2), 103–152.
- Kyriakides, L. (2006). Using international comparative studies to develop the theoretical framework of educational effectiveness research: A secondary analysis of TIMSS 1999 data. *Educational Research and Evaluation* **12**(6), 513–534.
- Kyriakides, L. (2008). Testing the validity of the comprehensive model of educational effectiveness: A step towards the development of a dynamic model of effectiveness. *School Effectiveness and School Improvement* **19**(4), 429–446.
- Linn, R. L. (1997). Evaluating the validity of assessments: The consequences of use. *Educational Measurement: Issues and Practice* **16**(2), 14–16.
- Monk, D. H. (1994). Subject matter preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review* **13**(2), 125–145.
- Nuttall, D. L. (1994). Choosing indicators. In Riley, K. A. and Nuttall, D. L. (eds.) *Measuring Quality: Education Indicators – United Kingdom and International Perspectives*, pp 17–40. Bristol, PA: Falmer.
- Oakes, J. (1989). What educational indicators? The case for assessing the school context. *Educational Evaluation and Policy Analysis* **11**(2), 181–199.
- OECD (1995). *Indicators of Education Systems: Measuring the Quality of Schools*. Paris: OECD.
- Opdenakker, M. C. and Van Damme, J. (2000). Effects of schools, teaching staff and classes on achievement and well-being in secondary education: Similarities and differences between school outcomes. *School Effectiveness and School Improvement* **11**(2), 65–196.
- Popham, W. L. (1997). Consequential validity: Right concern – wrong concept. *Educational Measurement: Issues and Practice* **16**(2), 9–13.
- Raudenbush, S. W. and Bryk, A. S. (1985). Empirical bayes meta-analysis. *Journal of Educational Statistics* **10**, 75–98.
- Reynolds, A. J. and Walberg, H. J. (1990). *A Structural Model of Educational Productivity*. DeKalb, IL: Northern Illinois University.
- Salganik, L. (1994). Apples and apples: Comparing performance indicators for places with similar demographic characteristics. *Educational Evaluation and Policy Analysis* **16**(2), 125–141.
- Scheerens, J. (1992). *Effective Schooling: Research, Theory and Practice*. London: Cassell.
- Scriven, M. (1994). Duties of the teacher. *Journal of Personnel Evaluation in Education* **8**, 151–184.
- Selden, R. W. (1994). How indicators have been used in the USA. In Riley, K. A. and Nuttall, D. L. (eds.) *Measuring Quality: Education Indicators – United Kingdom and International Perspectives*, pp 41–48. Bristol, PA: Falmer.
- Seltzer, M., Choi, K., and Thum, Y. M. (2003). Examining relationships between where students start and how rapidly they progress: Using new developments in growth modelling to gain insight into the distribution of achievement within schools. *Educational Evaluation and Policy Analysis* **25**(3), 263–286.
- Shavelson, R. J., McDonnell, L. M., and Oakes, J. (eds.) (1989). *Indicators for Monitoring Mathematics and Science Education: A Sourcebook. (Report No. R-3742-NSF/RC.)* Santa Monica, CA: RAND Corporation.
- Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice* **16**(2), 5–8, 13, and 24.
- Smith, M. S. (1988). Educational indicators. *Phi Delta Kappan* **69**(7), 487–491.
- Snijders, T. and Bosker, R. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.
- Stringfield, S. C. and Slavin, R. E. (1992). A hierarchical longitudinal model for elementary school effects. In Creemers, B. P. M. and Reezigt, G. J. (eds.) *Evaluation of Educational Effectiveness*, pp 35–69. Groningen: ICO.
- Visscher, A. J. (2001). Public school performance indicators: Problems and recommendations. *Studies in Educational Evaluation* **27**, 199–214.
- Walberg, H. J. (1984). Improving the productivity of America's schools. *Educational Leadership* **41**(8), 19–27.

Further Reading

- Campbell, R. J., Kyriakides, L., Muijs, R. D., and Robinson, W. (2003). Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education* **29**(3), 347–362.
- Cuttance, P. (1994). Monitoring educational quality through performance indicators for school practice. *School Effectiveness and School Improvement* **5**(2), 101–126.
- Kyriakides, L. and Creemers, B. P. M. (2008). Using a multidimensional approach to measure the impact of classroom level factors upon student achievement: A study testing the validity of the dynamic model. *School Effectiveness and School Improvement* **19**(2), 183–205.
- McDaniel, O. C. (1996). The theoretical and practical use of performance indicators. *Higher Education Management* **8**(3), 125–139.
- Meyer, R. H. (1997). Value-added indicators of school performance: A primer. *Economics of Education Review* **16**(3), 283–301.
- Muijs, D., Campbell, R. J., Kyriakides, L., and Robinson, W. (2005). Making the case for differentiated teacher effectiveness: An overview of research in four key areas. *School Effectiveness and School Improvement* **16**(1), 51–70.
- Scheerens, J., Glas, C., and Thomas, S. (2003). *Educational Evaluation, Assessment and Monitoring*. London: Routledge.
- Teddlie, C. and Reynolds, D. (2000). *The International Handbook of School Effectiveness Research*. London: Falmer.

This page intentionally left blank



GLOBALIZATION OF UNIVERSITIES AND INSTITUTIONS

Contents

Globalization, Universities and Medium of Instruction
The Washington Accord on Engineering Curriculum

Globalization, Universities and Medium of Instruction

J Lo Bianco, University of Melbourne, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

AD 1088: Ancient Globalization

With the benefit of a millennium of hindsight, it appears ironical the University of Bologna came into being offering trade in secular expertise, including the *ius gentium*, the Roman law of peoples. These legal principles were useful in advancing the incipient right of nations, and defending nation-states against the predatory claims of their contemporary rivals, the transnational polities of Empire and Church. Bologna's self-narration (Czarniawska, 1997) today revolves around its special historic claim as *Alma Mater Studiorum*, the oldest continuously operating degree-awarding autonomous institution (Rüegg, 2003), not a college of scholars but a teaching institution independent of state power and religious authority (Huff, 2003: 179; Makdisi, 1981). Its origins demonstrate the truism that knowledge (its creation and transfer) and the skilled competence that knowledge generates are intimately tied to power. Although the social effects of such knowledge–power links are usually indirect and long term, they can be world changing.

The mother university emerged from scholarly effort to assemble and disseminate extranational authoritative knowledge. An ironical outcome was to promote national institutions and even nationality itself. Today, the name Bologna is a code word for procedures and structures in support of global higher education, returning it to its extranational origins. This is appropriate rather than ironical, since from its inception, organized higher learning has been international in content and aspiration. It is

fitting too that the *Constitutio Habita* of 1158, Bologna's academic constitution (1155, according to Rüegg, 2003: 12), enshrines scholastic privilege, later venerated as the documentary origin of academic freedom (Watson, 2005: 373) in the struggle between the Pope and the Holy Roman Empire. This too became international, when on 18 September 1988, 430 university rectors signed the *Magna Charta Universitatum*, marking the 900th anniversary of the university's foundation.

All institutions engage in storytelling to solidify their constituents and define their distinctiveness (Gabriel, 2000; Boje, 2001). Bologna's self-narration tells a story of universalism, principles of institutional autonomy, and academic freedom – a narrative that is today widely endorsed.

The training imparted at Bologna during the eleventh and twelfth centuries arose because new markets emerged (incipient nations and institutions) demanding skilled competence in law and legal rhetoric. Universities today, both Western and non-Western, are enmeshed in rapidly integrating markets for competence as they supply these markets with skills. Universities also reinforce the existence of these markets and their reliance on universities. These markets, then as now, transcend boundaries of nation and culture, though most are still grounded in national traditions and all are marked by stratifications of power and inequalities of wealth.

Bologna's quest for origins (D'Epiro and Pinkowish, 2001), despite serving nationalist claims, is largely confirmable (Rüegg, 2003). In AD 1088 (AD 1087, according to Huff, 2003) a certain Irnerius (and Pepo) commences

teaching Emperor Justinian's sixth-century codification of Roman law, the *Corpus Iuris Civilis*, recently discovered at Pisa (Malagola, 1888). It is a confirmation of the irony that Irnerius' original academic activity was simply to gloss the ancient text, underscoring the indispensability of literate practices to all scholarly activity (Lo Bianco, 2004) while inadvertently adding value.

Lay students arrive in the city from many lands contracting to purchase this glossed knowledge. They organize themselves into collectivities, noncitizen guilds, or Learning Nations comprising Hungarians, Greeks, North Africans, Arabs, Franks, Germans, Iberians, etc. These became the locomotive of the educational economy, since students "had all the power . . . and dominated the masters" (Kerr, 2001: 16, 145; Rüegg, 2003: 12). The Learning Nations serve many purposes: they enhance the purchasing power of students, impose lecturer diligence, require regular timetables, and provide rent control and other practical support.

In 1888, in the aftermath of Italy's national independence, a festival of studies was organised at Bologna to celebrate its eighth centennial (UniBo, 2008) as the mother of universities. The late nineteenth century, an era of internationalism marked by European imperial expansion, jars with the festival's theme of cooperative internationalism. However, while universities inevitably reflect their cultural and national setting, an internationalist impulse runs deep. Knowledge cannot easily be confined within bounded geopolitical spaces, and the technical skill and philosophical reflection that universities produce stimulates cross-border markets for their exchange. The marketplace for competence is international because technical efficiency and substantive reflection are endlessly mutable, applicable, and exchangeable.

Islam's earliest institutions of learning, *madrāsas*, and higher education alike, were dispersed internationally, yet gathered around a shared textual tradition and modes of interpretation, all essentially forms of religious scholarship, but teaching and scholarship were also applied to science and law (Makdisi, 1981; Talabani, 1996; Daun and Arjmand, 2005). In diverse ways, universalism is found in all regions of the world, during all ages and across all civilizations (Hopkins, 2002), even in light of persistent arguments for the intellectual and moral roots of globalization, as distinct from economic and pragmatic ones, originating in European struggles (Headley, 2007). Today, the characteristic mode of higher education universalism applies an early 1800s' German conception of a research university adapted by American-inspired organizational and teaching models.

A shared medium of instruction (MOI) and standardized literacy are indispensable for the possibility of international education, even when the language concerned does not retain a home base in any nation.

Latin functioned in this way for some 14 centuries after Rome, making possible a Republic of Letters (McNeely and Wolverton, 2008); "almost unaffected by political and economic collapse" (Ostler, 2007: 18). This place-less Republic permitted elites to communicate and forge bonds of identity as "a community of scientists and scholars" (Ostler, 2007: 280). Today's Latin, on a historically unprecedented scale, is English, and perhaps because it does retain several home bases, English facilitates both the rise of international education and invests it with dilemmas, issues, and problems. It is important to note that it is not general spoken English that performs this facilitative function, but a range of identifiable registers of literate and persuasive academic English. To retain its privileged position, this kind of English is buttressed by a veritable industry of certification, assessment, and codification practices.

These practices are self-replicating and work to standardize, codify, and disseminate a teaching and written standard in scientific English discourse. Contemporary globalization is fueled by such communication practices making possible "the widening, deepening and speeding up of world wide interconnectedness" (Held *et al.*, 1999: 2). Since the *Lingua Mundi* has not supplanted other languages of scholarship and science, these share the communicative load involved in research, publication, teaching, explanation, and dissemination. Some might ultimately pose a threat to the privileges and domination accorded to scientific English. The precedent of Latin is instructive; ancillary scholarly languages initially shared a minor, mostly spoken, role in academic communication with Latin, then expanded to share literate functions with Latin, and then under impulses of nationalism and pedagogical effectiveness rose to challenge and eventually displace Latin altogether. National vernaculars achieved this only after modeling their rhetorical and literary practices on Latin's example, and therefore today, in specialized vocabulary and modes of reasoning, perversely assure it some kind of permanence.

AD 1999: Contemporary Globalization

The education reform commonly known as the Bologna Declaration was signed on 19 June 1999 by 29 European countries aiming to establish a European Higher Education Area by 2010. Naming new semantic fields, codifying names, standardizing codes, and disseminating their use are all indicators of emergent fields. These processes permit calibration and standardization; comparable nomenclature accompanies degree simplification, permitting standard divisions between forms and levels of study, standard degree requirements, and a standard grading scale with common transcript. In its implementation, Bologna

generates greater administrative coherence and procedural operations. The self-conscious aim is continent-wide comparability.

The expanding number of signatories, 45 at the time of writing, has meant that Bologna standardization has produced catalytic effects beyond its zones of application. Although ostensibly a move for administrative uniformity, Bologna will engender slow-acting but important long-term cultural consequences. One example is the European Credit Accumulation and Transfer System, leading to conversion from longer to shorter first degrees. More commonly designated bachelor graduates will facilitate masters-level expansion as the 3-year regime replaces most 5-year first degrees. Shorter duration first-degree standardization will accelerate study abroad, reduce first-degree government costs, and make master programs the major international study option.

Structural changes such as these fertilize the graduate education market, student behaviors, and study patterns, and make institutional dynamism more likely. In turn, these developments stimulate further adoption of English to attract nondomestic enrolments. More European countries are likely to offer price-based competition for international students in disciplines unconnected to local cultural or linguistic study, for example, English-medium business, technology, and science programs.

From their inception, universities have been the key global institution, far outlasting practically all other ancient institutions – functioning in a shared language, enrolling international students, engaging international academics, and disseminating knowledge beyond national borders. According to Altbach (2004), all universities today, other than Cairo's ancient Al-Azhar, originate from the European medieval university, with non-Western countries having European models "imposed on them by colonial masters. Even those countries not colonized by Western powers – such as Japan, Thailand, Ethiopia . . . – adopted the Western academic model . . . even where, as in China, well-established indigenous academic traditions already existed." (p. 4).

The 1999 *Magna Charta Universitatum*, now available in 47 languages, attracts increasing Asian, American, and African adherents, with signatories now totaling 579 (IPPT, 2008; UniBo, 2008) in 78 countries. It originated in the medieval recognition of rights to unhindered travel to study or teach; but fidelity to its principles of freedom and autonomy appears to be independent of setting and time.

Universalism in higher education therefore, has many sources and modalities: secular and religious, geographic and civilizational, technical and economic, and both imposed and voluntary. The modalities include cooperative charters of principle, competitive ranking of prestige, colonization and domination, and trade in commodified language and certified study (Tan and Rubdy, 2008).

Globalization and *Lingua Mundi*

A recent study by Cha and Ham (2008) documents historic patterns of foreign-language study. During 1875–99, only 6% of secondary-school curricula worldwide nominated English as the first foreign language, increasing to 70% of primary and 80% of secondary curricula by 1990–2000. In Asia, English was represented in only 33% of primary curricula during 1945–69, growing to 83% in primary and 100% for secondary by 2005. By 2006, practically all instances in which foreign languages were employed to teach mainstream subject matter in Asian universities involved English. Recent growth in English is mostly independent of whether countries were British colonies or under American political influence, leading the authors to argue that English is no longer a foreign language, but rather a kind of foundational knowledge, or basic skill used for globalization. Supporting such claims are projections of English use as well as learning, particularly those made by Graddol (2006) who claims that English students and users might in a period of only two decades equal a third, or even half, the world's population.

The diverse sources of demand for English produce innumerable local cultural and political meanings and effects. Global English defies temptations to depict its role in today's world according to a simple binary choice between imperial instrument and unproblematical asset. Concern often arises about English when it replaces teaching-through national languages, even in countries with long-standing national scientific-education traditions, for example, the Netherlands and Sweden. Recent debates about teaching in English had unpredictable effects, pushing policymakers to retreat from multilingual policies concerned about long-term vitality of Swedish as a result of exposure to globalization (Boyd, 2007: 175). This is also true in China, where obligatory English alongside Putonghua results in a tri-lingual and tri-literate burden on non-Han minority populations (Feng, 2007).

Human capital theorizations dominate education policies practically everywhere, and English-delivered instruction reflects and helps to strengthen a distinctive marketplace for competence, certification, and exchange. Internationally compared degrees, international universities, and student mobility are not exclusive to English, but it is the most consistent element in such arrangements. Representing English as a basic skill has the effect of reducing its foreignness and even of representing English as purely an instrument of communication, or a postidentity language (Lo Bianco, 2005). These labels and ways of talking about English are the discursive accompaniment to its material attractiveness as a commodity (Lo Bianco, 2007b; Tan and Rubdy, 2008).

The nature of this commodity is decreasingly Anglo-American and increasing number of Asian institutions

position themselves as suppliers of general English and increasingly for specialized education through English. For some observers, a single global society (Cha and Ham, 2008) follows in the wake of English-mediated mass education with its standardized qualifications, increasingly portable form, and traded in an interlinked marketplace of skills. Perhaps, this is premature given the recency of domination by English as demonstrated in the Cha and Ham data; however, what is clear is the expanding number of extranational communities of communication founded in professional identities and facilitated by English-knowing bilingualism. These are evident in professional associations, professional conferences, and in the academic work and professional conversations that sustain these communities of communication. As these communities are elites within their own societies, this association of English as a medium is making possible horizontal linkages, that is, of elites across national boundaries, associates knowing English with social stratification and inequality.

As the medium of most international scientific literature and the most disseminated journals, both in print and electronic form, English clearly dominates the conversations of academic life. Access to its elevated registers is therefore, for individuals and institutions, a predictor of advantage, such as when competitive promotions rely on reputable publication in approved journals. English capability is increasingly associated with reward systems in higher education as reputation is a critical ingredient of systemic advancement. The reality is, however, that access to English, especially in developing countries, is distributed according to preexisting social, ethnic, and geographic divisions. Schooling, much less higher education, cannot adequately compensate for inherited advantages in access to the powerful written and spoken styles of English, which are often informally acquired in family and affinity networks that socialize privileged groups in cultural capital (Bourdieu, 2005). Education systems which favor English-delivered instruction, therefore compound existing inequalities for minority and poor populations who are precluded by prevailing social structures from acquiring prestigious forms of English.

The effort to reproduce and conform to the norms of academic life which prevail in Anglophone settings requires institutions and societies more generally to make adjustments that go beyond what occurs in universities alone, impacting deeply on social networks and relationships. Additionally, Altbach (2004) notes several ways in which English constitutes advantages for American and British authors and institutions, such as the cost of English-language databases, products and resources, direct experience in editing and housing scholarly journals, and procedural advantages in peer review and academic writing. Higher education institutions in developing

countries find access to this array of advantage systems, through which critical information is negotiated and reputations forged, cost-prohibitive. Restrictions on ownership and usage, such as stringent copyright procedures, entrench existing hierarchies of control and quality and quantity of access.

However, there are other and more subtle cultural and ideological questions which arise, such as command of persuasive argument styles inherent in particular disciplines and their surrounding traditions of academic life (Lo Bianco, 2004). As the reasoning and rhetorical practices valued in the Anglophone tradition are associated with privilege and success within societies, indigenous versions of these practices can be displaced or challenged. While it is true that languages contain low ideology and relatively culture-free transactional styles, higher-order functions of language are hardly neutral. An important recent exploration of this is Wierzbicka's (2006) study of the persisting layers of culture in English. In this work, Wierzbicka finds semantic bias within the grammatical core and discursive styles of English contributing a significant installment in what is a long debate about how language and thought intersect. At the very least, this work serves to render problematic the often naive assumption that any language can be rapidly removed from its historic and contextual origins and shorn of traces of these, as occurs when English is considered a totally neutral tool of communication, an instrument without imprint of history and culture.

Formal university rankings represent a clear example of the language-prestige hierarchy. In the 2006 edition of the Shanghai Jiao Tong Index (SJTI), 19 of the world's 20 top-ranked universities were American or British and 66% of the top 100 were located in English-language settings. In his analysis of the SJTI, Marginson (2007a) argues that such rankings "... favour universities ... from English language nations because English is the language of research (non English language work is published less and cited less); and universities from the large US system as Americans tend to cite Americans" (p. 133).

Reflecting on this palpable advantage bias, Van Parijs (2007: 85) proposes a kind of native-speaker tax on Anglophones, payable to English learners to defray the costs of acquisition. Van Parijs calculates the subsidies involved, concluding that Chinese people would receive €32.20 per capita from the Anglo population; French speakers would be granted a €24.24 subsidy; and Danes would each receive €23.76. While Coulmas (2007) adopts a more accommodating stance he also acknowledges the extra burden imposed on non-native users of English but implies a possible future shift when he has an interlocutor in simulated conversation on scientific publishing conclude: "I registered for an intensive Chinese course yesterday" (p. 13).

Gigantic Peripheries

China, along with India, constitutes what Altbach (1998) has called gigantic peripheries to American and English-centered contemporary globalization. Since 1998, they remain gigantic but have moved close to the center of where knowledge and human capital economic power intersect. In very different ways, Chinese and Indian education systems reveal immense penetration by English.

In 2001, the Chinese Ministry of Education issued a stipulation that universities should provide 5–10% of undergraduate instruction in foreign languages (Pan, 2007), leading to considerable growth of English–Chinese bilingual teaching, and expanded the scope of English teaching at other levels (MinEd, 2001). An evaluation of these initiatives at East China University of Science and Technology in Shanghai found predictable problems of implementation and uneven readiness, but by 2006 bilingual education was flourishing; producing results superior to traditional English teaching (Pan, 2007: 212).

In such environments, English has two principal roles, as MOI and as object of instruction for English majors. Exploring this accelerating globalization in Chinese higher education, Chang (2006) also looks at adjustments to national English language policy.

Documenting a surge of use of English as MOI, Chang's study identifies the ripple caused into the basic university-training models, in revisions to the national curriculum and in textbook design and content. No fewer than six new training models had emerged: English major plus courses in other specialisms, English major plus an orientation toward other disciplines, English major plus a minor, a major plus English language, English language plus another foreign language, and dual degree – BA degree of English language and literature plus another BA degree.

It is unlikely that such an array of English-centered program designs and training models producing composite-type graduates, can be quarantined within English training or language departments in universities. Instead, it is more likely to cause organization adjustments throughout education, as what English training is, becomes meshed with basic academic preparation in a growing number of academic disciplines. English and Chinese will share the communicative load in some programs, such as collaborative degree programs discussed by Pan (2007) which are conducted in English with components of Chinese-language facilitation or support to enhance learning or to overcome predictable gaps. In other settings, such as Malaysia, which have a more established local history of English, there is growing use of English-only (Altbach, 2004; Kaur, 2004) or English-dominant programs.

The transnational status of institutions extends such intra-institutional changes brought about by English

medium education. In a dramatic display of such developments, *Science Magazine* recently proclaimed that Chinese universities are now the top feeders for US doctoral programs. Reporting research showing that Tsinghua and Peking Universities “have topped the list of undergraduate schools whose graduates go on to earn a doctorate from an American university” (Mervis, 2008: 185) the story underscores the transnational dimension of contemporary higher education, the facilitative function of English, and the extent of mobility.

This integrated and global marketplace fosters trade in expertise with practically instantaneous exchange of information. The two prestigious Chinese institutions beat the University of California, Berkeley, to third place with Seoul University next. This cross-border transfer and burgeoning interdependence between US and Chinese higher education is accompanied by massive Chinese investment in high-scale research capability and increasing attraction to large numbers of in-bound students and academic staff in Chinese-language programs.

Small Peripheries

More asymmetrical is the relationship between globalized education products, processes, and certification, and small societies. Cambodia, with its history of multiple colonization and dependency, is an instructive and disturbing case. Exemplifying this is the case of the Royal University of Phnom Penh, founded on 13 January 1960 (Man and Luong Chan, 2002), incidentally a signatory to the *Magna Charta Universitatum*.

Cambodian language policy (Clayton, 2002) has been buffeted by colonization, immersion in global markets, and now by the influence of development assistance, often with linguistically tied programs. Influencing the communicative patterns of Cambodian society are strategic, cultural, ideological, and economic interests from outside, intertwined with domestic interests, and the choices that result from their interaction on questions of MOI, minority-language rights and foreign-language preferences. All these points are in evidence in the recent history of one institution. Since its inception, the Royal University of Phnom Penh (originally Khmer Royal, then Phnom Penh University), taught mostly in French during the 1960s, underwent gradual Khmerization during the 1970s, conserving some use of French, but was closed due to political upheaval and violence between 1975 and 1978. During the 1980s, teaching was initially in Russian, later Vietnamese was added, and still later Khmer was included. In the late 1980s and early 1990s, first French then English were used alongside Khmer. From the early 2000s, teaching has been mainly in Khmer but with booming interest in English and

residual use of French, often promoted as a condition of assistance (Man and Luong Chan, 2002; Clayton, 2006).

Despite the obvious disruptive effects of such changes, MOI policy is often taken to be a straightforward question of cost and efficiency. In this unproblematic depiction, a country or a single institution would simply calculate the costs of different language choices and opt for the cheapest and most feasible alternative. Informing the calculation of cost would be the availability of academic literature in the different language choices, the expense and feasibility of translation, and the availability and training costs of lecturers. Opposition to such efficiency-based reasoning is often represented as backward-looking nationalism or sentimentality. However, reducing MOI choices entirely to such a binary between cost and sentiment, or technical efficiency versus national-language ideologies, is crude, obscuring some very complex issues and problems, which extensive documentation of MOI policies reveal. Not least among these complicating factors are the potentially corrosive effects on the internal efficiency and vitality of national languages when they are excluded from high-level scientific discourses. Affecting people, rather than languages, are the ways in which MOI policies preferring foreign over local languages compound inequalities already faced by minority or disadvantaged groups. Neither of these arguments relates to sentiment, but to the practical effects, on people and on their languages, of favoring external or foreign languages over local ones.

In the light of this, it is not surprising that Tollefson and Tsui's (2004) survey of Asian experiences with MOI policy leads them to argue that there are always agendas involved in choosing a language of instruction, that these agendas align with particular interests, and are subject to change over time in response to shifting economic and political circumstances. We can see these shifts clearly in the case of Sri Lanka (Ceylon), Malaysia, and Singapore. In the postcolonial context of the mid-1950s, these countries all ejected English from their education systems for reasons of national reconstruction, nationalism, or local administrative consolidation; however, by the early 2000s all three had reinserted English both as an object of instruction and as MOI. In some cases, English has been restricted to teaching only some disciplines, especially science, mathematics, technology, and business studies, while in others it was used more widely (Lo Bianco, 2007a). In the mid-1950s, when English was evicted from education, there were protests in its defense, and today when English has been either fully or partially restored as a teaching as well as a taught language, there are varying levels of concern and agitation against its new status.

Kaur's (2004) examination of the Malaysian case isolates the contribution of higher education bifurcation to undermining Malaysia's previous ability to resist the

incursions of English. Bifurcation refers to reforms permitting English-medium private universities to operate alongside Malay-medium public universities. These were the catalysts provoking a wider dismantling of the Malay-only MOI stance, eventually also in schools. Mon-lingual Malay-educated majority-population interests were made vulnerable by the marketplace advantages of English-knowing bilinguals, and in 2009, there continues to be vigorous debate about both the efficacy and desirability of using English to teach mainstream subjects and the ripple effects onto different ethnic and social groups. The public/private divide highlights how the asymmetrical acquisition of English compounds internal social stratification. A related situation typifies Sri Lanka (Lo Bianco, forthcoming) where language learning aligns with social opportunity, ethnic and racial backgrounds, geography, and social class.

Language planning in universities at a time of rapid and profound globalization can only be properly examined as choices situated within preexisting sociopolitical contexts, reflecting divergent interests and agendas as much as opening up new opportunities in emergent markets. Tollefson and Tsui (2004) rightly identify sociopolitical processes as key drivers of MOI choices, rather than unproblematical technical assessments of costs in delivery of content, or access to scientific literature. Even when cost and efficiency considerations predominate, debates about language policy cannot be credibly separated from inherited historical arrangements. Both the discourse, which accompanies MOI policy and the agendas various interests advance are underscored by Tsui's analysis of Hong Kong, in which she argues that MOI is:

... shaped by an interaction between political, social and economic forces. . . among these agendas, it is always the political agenda that takes priority. Other agendas. . . come to the fore only if they converge with the political agenda. Yet it is always these [other] agendas that will be used as public justification for policy making. (Tsui, 2004: 113)

Conclusion

Globalization transforms universities directly making research-based higher education, like finance services, a sensitive and exposed sector (Marginson, 2007b). Critical to national economic positioning, cultural production, and international communication, higher education achieves its capability through intense and endemic globalization. Knowledge-intensive sectors, industries and services, are aggregated around and within universities, making them large business corporations organizing the collection of knowledge and its dissemination in publishing, seminars, conferences, and applications of various

kinds. These interactions bring to the surface an intense connection between English and technical knowledge and help make the language itself a commodity (Tan and Rubdy, 2008) traded in its distinctive marketplaces.

Although positioned in national states, knowledge economies are interlinked across national boundaries, and are organized by these global relations. Opportunities for mobility and employment opened up by English are not as well distributed and organized as is often supposed. English does not just mediate access to knowledge and skills, but is a substantial component of knowledge and skills. Languages and their associated literacies impose their own hierarchies and can compound endemic inequalities even when they offer new opportunities and prospects.

The contemporary era of global education recalls its ancient precedents because it underscores the nonconfineability of knowledge and the internationalist impulse of the university. Education systems both celebrate and defend academic freedom and institutional autonomy and accommodate peripheries, whether gigantic or small, to norms that privilege arrangements set by the dominant centers of education. These patterns are linked to economic privilege and power, and align closely with the world's current language regime. While current language arrangements might not be assured or permanent, their entrenched power contains elements, which are self-fulfilling and which are deeply rooted in systems of reward, ranking, and comparison.

On the positive side, language-based globalization also promotes multiple channels of communication and increasingly uses technologies based on multiple literacies, combining digital, electronic, and visual with the print literacies of the past. These potentially expand the range of individuals and social groups able to participate in and benefit from higher education. No education will be immune from such epoch-making changes and even the most technical subjects and fields of research will have to deal with audiences, which are multicultural and highly mobile and communication technologies that threaten to disrupt the very basis of written literacy (Crossman, 2004). Scientific literate communication is as ancient as the institutions which rely on it, and its future forms appear to be as dynamic as their contemporary functions. As all such changes are located within endemically unequal arrangements, the ancient globalization of institutions sheds light on patterns that too often we imagine are unique in time.

Bibliography

- Altbach, P. (1998). Gigantic peripheries: India and China in the world knowledge system. In Altbach, P. (ed.) *Comparative Higher Education*, pp 133–146. Greenwich, CT: Ablex.
- Altbach, P. (2004). Globalization and the university. *Tertiary Education and Management* 10(1), 3–25.
- Boje, D. (2001). *Narrative Methods for Organizational and Communication Research*. London: Sage.
- Bourdieu, P. (2005). *Language and Symbolic Power*. Malden, MA: Polity Press.
- Boyd, S. (2007). Communication and community. In Pauwels, A., Winter, J., and Lo Bianco, J. (eds.) *Maintaining Minority Languages in Trans-National Contexts*, pp 141–179. Basingstoke, UK: Palgrave.
- Cha, Y.-K. and Ham, S.-H. (2008). The impact of English on the school curriculum. In Spolsky, B. and Hult, F. M. (eds.) *Handbook of Educational Linguistics*, pp 313–327. Malden, MA: Blackwell.
- Chang, J. (2006). Globalization and English in Chinese higher education. *World Englishes* 25(3–4), 513–525.
- Clayton, T. (2002). International languages in education in developing countries. In Lo Bianco, J. (ed.) *Voices from Phnom Penh: Development and Language*, pp 87–102. Language Australia: Melbourne.
- Clayton, T. (2006). *Language Choice in a Nation under Transition*. New York: Springer.
- Coulmas, F. (2007). English monolingualism in scientific communication and progress in science. In Carli, A. and Ammon, U. (eds.) *Linguistic Inequality in Scientific Communication Today*, pp 5–14. AILA Review, Amsterdam: Benjamins.
- Crossman, W. (2004). *Vivo*. Oakland, CA: Regent Press.
- Czarniawska, B. (1997). *Narrating the Organization*. Chicago: University of Chicago Press.
- Daun, H. and Arjmand, R. (2005). Islamic education. In Zajda, J. (ed.) *International Handbook of Globalisation, Education and Policy Research*, pp 377–388. Netherlands: Springer.
- D'Epiro, P. and Pinkowish, M. D. (2001). *Sprezzatura*. New York: Anchor Books.
- Feng, A. (ed.) (2007). *Bilingual Education in China*. Clevedon, UK: Multilingual Matters.
- Gabriel, Y. (2000). *Storytelling in Organizations*. Oxford: Oxford University Press.
- Graddol, D. (2006). *English Next*. London: The British Council.
- Headley, J. (2007). *The Europeanization of the World*. Princeton: Princeton University Press.
- Held, D., McGrew, A., Goldblatt, D., and Perraton, J. (1999). *Global Transformations*. Stanford: Stanford University Press.
- Hopkins, A. G. (ed.) (2002). *Globalization in World History*. London: Pimlico.
- Huff, T. (2003). *The Rise of Early Modern Science*. Cambridge: Cambridge University Press.
- IPPT (2008). *Institut Penyelidikan Pendidikan Tinggi Negara USM*. <http://www.usm.my/ipptn/v2/documents/Globalupdates/Magna%20Charta%20Universitatum-No%2040.pdf> (accessed October 2009).
- Kaur, G. S. (2004). Medium of instruction policy in higher education in Malaysia: Nationalism versus internationalization. In Tollefson, J., Tsui, J., and Tsui, A. B. M. (eds.) *Medium of Instruction Policies*, pp 135–152. Erlbaum Publishers: Mahwah, NJ.
- Kerr, C. (2001). *The Uses of the University*. Cambridge: Harvard University Press.
- Lo Bianco, J. (2004). Widening access and literacy. In Osborne, M., Gallacher, M. J., and Crossnan, N. (eds.) *Researching Widening Access*, pp 65–76. London: Routledge.
- Lo Bianco, J. (2005). No longer a (foreign) language: Rhetoric of English as post-identity language. *Journal of Chinese Sociolinguistics* 2(5), 17–41.
- Lo Bianco, J. (2007a). Advantage plus identity: Singapore's MOI policy. In Vaish, V., Gopinathan, S., and Liu, Y. (eds.) *Language, Capital, Culture*, pp 5–22. Amsterdam: Sense Publishers.
- Lo Bianco, J. (2007b). Protecting English in an Anglophone age. In Cummins, J. and Davison, C. (eds.) *Handbook of English Language Teaching*, pp 169–183. Norwell, PA: Springer.
- Lo Bianco, J. (forthcoming). A friendly knife? English in the context of Sri Lankan language politics. In Farrell, L., Singh, U. N., and Giri, R. A. (eds.) *English Language Education in South Asia*. Cambridge University Press: India.
- Makdisi, G. (1981). *Rise of Colleges: Institutions of Learning in Islam and the West*. Edinburgh: Edinburgh University Press.
- Malagola, C. (1888). *Statuti delle Università e dei Collegi dello Studio Bolognese*. Bologna: Zanichelli.

- Man, S. and Luong Chan, S. (2002). Building institutional capacity. In Lo Bianco, J. (ed.) *Voices from Phnom Penh: Development and Language*, pp 75–87. Melbourne: Language Australia.
- Marginson, S. (2007a). Global university rankings. *Journal of Higher Education Policy and Management* **29**(2), 131–142.
- Marginson, S. (2007b). *Prospects of Higher Education*. Rotterdam: Sense Publishers.
- McNeely, I. and Wolverson, L. (2008). *Reinventing Knowledge*. New York and London: W.W. Norton.
- Mervis, J. (2008). US Graduate Training: Top PhD feeder schools are now Chinese. *Science Magazine* (July 11) **321**(5886), 185.
- MinEd (2001). Standard of English Courses for 9-Year Compulsory Education and General Senior High Schools. <http://www.tefl-china.net/2003/ca13821.htm> (accessed October 2009).
- Ostler, N. (2007). *Ad Infinitum, A Biography of Latin*. London: Harper Press.
- Pan, J. (2007). Facts and considerations about bilingual education in Chinese universities. In Feng, A. (ed.) *Bilingual Education in China*, pp 200–218. Clevedon, UK: Multilingual Matters.
- Rüegg, W. (2003). Mythologies and historiography of the beginnings. In De Ridder-Symoens, H. (ed.) *A History of the University in Europe; Vol 1, Universities in the Middle Ages*, pp 4–34. Cambridge: Cambridge University Press.
- Talabani, A. (1996). Pedagogy, power, and discourse: Transformation of Islamic education. *Comparative Education Review* **40**(1), 66–82.
- Tan, P. K. W and Rubdy, R. (2008). *Language as Commodity*. London and New York: Continuum.
- Tsui, A. B. M (2004). MOI in Hong Kong. In Tollefson, J., Tsui, J., and Tsui, A. B. M. (eds.) *Medium of Instruction Policies*, pp 97–116. Mahwah, NJ: Erlbaum Publishers.
- UniBo (2008). <http://www.eng.unibo.it/PortaleEn/University/Our+History/default.htm>. <http://www.eng.unibo.it/PortaleEn/Students/User+guide+to+Campus+Branches/User+Guide+to+Bologna/history.htm> (accessed October 2009).
- Van Parijs, P. (2007). Tackling the Anglophones' free ride. In Carli, A. and Ammon, U. (eds.) *Linguistic Inequality in Scientific Communication Today*, pp 72–87. AILA, Amsterdam: Benjamins.
- Watson, P. (2005). *Ideas*. London: Weidenfeld and Nicolson.
- Wierzbicka, A. (2006). *English: Meaning and Culture*. Oxford: Oxford University Press.

Relevant Website

<http://www2.unibo.it> – Magna Charter of Universities, University of Bologna.

The Washington Accord on Engineering Curriculum

E Ko, The Hong Kong University of Science and Technology, Hong Kong, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Constructive alignment – The alignment of teaching and assessment to the intended learning outcomes.

Engineering education research – An academic discipline engaged in scholarly enquiries into the principles, methodologies, and practices in engineering education.

Graduate attributes – A set of individually assessable outcomes that are indicative of the graduate's competencies.

Professional competencies – A set of knowledge, skills, and attitudes which a practicing professional should possess and be able to demonstrate in a holistic manner.

Student learning outcomes – What a student should know or be able to do as the result of a learning experience.

Student mobility – The ability of students to study or work after graduation across national boundaries.

Substantial equivalence – A principle used in establishing the mutual recognition of qualifications by focusing on a set of graduate attributes that are necessary for entry to the profession, rather than mandating identical content and outcomes.

Washington Accord – An international agreement which recognizes the accredited engineering degree programs of its signatories and recommends that graduates from these programs be recognized as being prepared to practice engineering at the entry level.

Introduction

Globalization of Higher Education

The higher education sector worldwide has seen many significant changes in recent years. The evolution from elite to mass education has led to a much larger student population and, along with it, a more diverse student body in terms of age, gender, academic, and cultural backgrounds. There are also more modes of delivery and different types of learning experiences, many of which are made possible by the advent of technology. Along with an increasing demand of accountability for public funds

spent on higher education, universities everywhere are under pressure to manage these changes in an efficient and effective way.

At the same time, employers worldwide are placing new demands on university graduates. The inevitable progress of globalization has changed the ways companies operate and the attributes they look for in new hires. In order to compete internationally in a knowledge economy, multinational companies will need to be more agile and adaptive. They want to build a workforce of people who can work productively in a multidisciplinary and cross-cultural environment and are willing to recruit from all over the world.

An important consequence of globalization then is the increasing mobility of students entering and graduating from universities. In Globalization 3.0, a term coined by Friedman (2006), individuals of high mobility can choose to study, work, collaborate, and compete anywhere in the world. Many universities have taken advantage of this development to expand their international student population by recruiting globally. University graduates also enjoy a much wider choice of geographical locations in which they can pursue their career paths. With the diminishing importance of national boundaries, there is a need for international norms and standards by which academic qualifications from different countries can be meaningfully compared. Furthermore, with the emphasis on lifelong learning, an international framework for capturing learning experiences beyond formal education is necessary.

Globalization of Engineering Education

The engineering profession, in terms of its practice, research, and education, has always had a global presence. In fact, engineers have contributed to globalization through technological advances in areas such as communication, energy, environment, logistics, materials, and biotechnology. Engineering companies are often multinational, operating in multiple sites and employing engineers worldwide.

Ironically, despite the contributions engineers have made to the well-being of modern society, engineering as a profession is becoming less attractive to young people, especially in developed countries. An engineering education is often perceived to be narrow, boring, and demanding. It is thus incumbent on engineering schools to come up with innovative curricula that, as put by Vest (2006), arouse engineering students' passion, curiosity, engagement, and

dreams, and thereby prepare them better to face the exciting opportunities in engineering and related disciplines. Furthermore, these graduates will need to have a set of attributes, which can be benchmarked internationally, to ensure their global mobility. The Washington Accord is an international effort to this end, by encouraging institutions that offer engineering degree programs to focus on student learning outcomes in an accreditation process that leads to mutual recognition.

Washington Accord

History

The Washington Accord was an agreement signed in 1989 by six engineering education accreditation bodies. These bodies, having examined the accreditation processes, policies, and procedures in their respective countries, concluded that they are comparable. The original six signatories therefore recognize the substantial equivalence of their engineering degree programs in satisfying the academic requirements for the practice of engineering at the entry level of the profession.

It should be noted that the agreement covers only accredited educational qualification but not professional registration, which rests with individual countries or regions. However, each signatory will encourage the licensing body in its own country or region to accept the substantial equivalence of engineering education programs accredited by other signatories.

The administration of the Accord is supported by a secretariat and the signatories meet biennially. In order to become a signatory, an applicant must be nominated by two of the existing signatories and supported by at least two-thirds of the existing signatories. A successful applicant will be given a provisional status and a prescribed period during which the applicant must demonstrate that it has an accreditation system that is comparable to those of the existing signatories. The change from a provisional status to a full signatory will require the unanimous approval of the existing signatories.

Table 1 shows the founding, other, and provisional signatories of the Washington Accord. In recent years many countries or regions have expressed an interest in joining the Washington Accord and, as part of that process, developing accreditation systems for engineering degree programs.

Related International Partnerships

Based on the principle of substantial equivalence established in the Washington Accord, the Sydney Accord and the Dublin Accord came into being in 2001 and 2002 to recognize accredited qualifications for engineering technologists and engineering technicians, respectively.

Table 2 summarizes the founding and provisional signatories of these two accords.

Another system for a common recognition of accredited engineering degree programs for entry to the profession is the European Network for Accreditation of Engineering Education (EUR-ACE), which established framework standards for comparing educational qualifications in the European Higher Education Area.

The Engineers Mobility Forum agreement and the Asia Pacific Economic Cooperation (APEC) Engineer agreement have been introduced to promote the mutual recognition of professional competence in engineering, both using the International Professional Engineers Register. There are also many international engineering organizations, such as World Federation of Engineering Organizations (WFEO), European Federation of National Engineering Associations (FEANI), and Pan American Federation of Engineering Associates (UPADI), for the sharing of good practices in global engineering activities.

Graduate Attributes and Professional Competencies

Graduate attributes are a set of statements of what graduates from an engineering degree program are able to do

Table 1 Signatories of the Washington Accord

Founding signatories (1989)	Australia, Canada, Ireland, New Zealand, United Kingdom, and United States	
Other signatories	Hong Kong China (1995), South Africa (1999), Japan (2005), Singapore (2006), Chinese Taipei (2007), Korea (2007), and Malaysia (2009)	
Provisional signatories	Germany (2003), India (2007), Russia (2007), and Sri Lanka (2007)	

Table 2 Signatories of the Sydney Accord and Dublin Accord

	<i>Sydney Accord originally signed in 2001</i>	<i>Dublin Accord (originally signed in 2002)</i>
Founding signatories	Australia, Canada, Hong Kong China, Ireland, New Zealand, South Africa, and United Kingdom	Canada, Ireland, South Africa, and United Kingdom
Other signatories	United States	
Provisional signatories		New Zealand (2006) and United States (2007)

and form the basis of outcome-based accreditation criteria. These statements should be clear, succinct, and assessable. The graduate attributes in the Washington Accord, Sydney Accord, and Dublin Accord are all defined in 13 areas: academic education, knowledge of engineering sciences, problem analysis, design/development of solutions, investigation, modern tool usage, individual and teamwork, communication, the engineer and society, ethics, environment and sustainability, project management and finance, and lifelong learning.

For consistency and ease of comparison, a common stem is used in the three accords for each attribute, with the characteristics of engineers, engineering technologists, and engineering technicians differentiated by range information. For example, the common stem for individual and teamwork is: function effectively as an individual, and as a member . . . ; the range information is:

- Engineer: or leader in diverse teams and in multidisciplinary settings.
- Engineering technologist: or leader in diverse technical teams.
- Engineering technician: in diverse technical teams.

There are no differentiating characteristics in three of the 13 attributes (ethics, environment and sustainability, and lifelong learning).

Whereas graduate attributes are meant for academic accreditation, professional competencies are for professional registration. As such, professional competencies define a set of knowledge, skills, and attitudes which a practicing engineer should possess and be able to demonstrate in a holistic manner. The overall performance standard is defined in terms of 13 areas: comprehend and apply universal knowledge, comprehend and apply local knowledge, problem analysis, design and development of solutions, evaluation, responsibility for decisions, manage engineering activities, ethics, protection of society, communication, lifelong learning, judgment, and legal and regulatory. Again, stem and range information are used to differentiate the professional competencies of engineers, engineering technologists, and engineering technicians. For example, the common stem for problem analysis is: define, investigate/clarify/state, and analyze; the range information is:

- Engineer: complex problems.
- Engineering technologist: broadly defined problems.
- Engineering technician: well-defined problems.

There are no differentiating characteristics in four of the 13 areas (ethics, communication, lifelong learning, and legal and regulatory).

Graduate attributes and professional competencies provide a common language for the comparison of educational and professional systems in different parts of the world. In the spirit of substantial equivalence, graduate attributes should not be rigidly interpreted, but instead should be

viewed as an indication of the expected outcomes of an engineering degree program worthy of accreditation.

Impact of the Washington Accord on Engineering Education

Outcome-Based Accreditation

The conceptual framework of outcome-based education is to start with clear intended learning outcomes, then to use the most appropriate curriculum, pedagogy, and assessment to ensure that students succeed in achieving such outcomes. Simply put, a learning outcome is what a person should know or be able to do as a result of a learning experience; it should state clearly the conditions under which a learner can do a particular task and the expected level of performance.

Learning outcomes can be classified into three categories: knowledge, skills, and affective (or attitudinal) outcomes. According to the taxonomy proposed by Ewell (1984), knowledge outcomes include breadth of knowledge and depth of knowledge (in a specific discipline); skills outcomes include general competencies and professional/occupational skills; and affective outcomes include personal goals and aspirations, general attitudes, values, and satisfaction, attitudes toward self, and attitude toward others.

At a time when traditional educational processes have become more diverse in terms of the input and process variables, learning outcomes provide a common language by which graduates from different programs can be meaningfully compared. For this reason many accreditation systems that have an international presence have adopted an outcome-based approach in recent years. The idea is not to ensure identical learning outcomes across national boundaries but, in the spirit of substantial equivalence, spell out clearly what graduates from an accredited program should be able to do for entry to the profession.

In outcome-based accreditation, the focus is on program-level learning outcomes. It is important that these outcomes are (1) consistent with the mission of the program and that of the university, (2) determined based on inputs from all relevant stakeholders, and (3) defined by appropriate performance indicators. The evaluation of the program can then be made by reliable criteria and benchmarks, with an emphasis on providing specific feedback for improvement purposes. The essence of outcome-based accreditation is a commitment to continuous quality improvement based on credible evidence.

A Case Study: ABET, Inc

Engineering and technology programs in the United States are accredited by ABET, Inc. In response to growing concerns that ABET accreditations for engineering degree programs were too bureaucratic and not conducive to educational innovations, ABET introduced in the mid-1990s

Engineering Criteria (EC) 2000. The key focus of EC 2000 is to measure the outcomes of engineering degree programs, in terms of what graduates can actually do.

The EC 2000 has eight components, with criterion 3 being program outcomes and assessment. Criterion 3 states that engineering degree programs must demonstrate that their graduates have the following capabilities:

1. An ability to apply knowledge of mathematics, science, and engineering approaches to the discipline.
2. An ability to design and conduct experiments and analyze and interpret data.
3. An ability to design a system, component, or process to meet desired needs.
4. An ability to function on multidisciplinary teams.
5. An ability to identify, formulate, and solve engineering problems.
6. An understanding of professional and ethical responsibility.
7. An ability to communicate effectively.
8. The broad education necessary to understand the impact of engineering solutions in a societal context.
9. A recognition of the need for an ability to engage in lifelong learning.
10. A knowledge of contemporary issues.
11. An ability to use the techniques, skills, and modern engineering tools necessary for engineering practice.

Every engineering degree program in the United States has now gone through the EC 2000 process and a vast majority of programs have been accredited. A study of the impact of EC 2000 (Lattuca *et al.*, 2006) showed that the new accreditation criteria had led to a positive, and sometimes substantive, impact on engineering programs, student experiences, and student learning. In particular, these researchers noticed a greater emphasis on professional skills and active learning after EC 2000. Their data showed that engineering students who graduated in 2004 differed significantly from their predecessors in eight of ten experiences inside and outside the classroom. Furthermore, there was a slight rise in the knowledge and skills in science, mathematics, and engineering, despite the focus on professional skills. The researchers also reported higher levels of faculty support for continuous improvement, notwithstanding a mixed emphasis on teaching in the faculty reward structure. Results of this study provided the most comprehensive evidence thus far for the effectiveness of an outcome-based approach in accreditation.

Impact on Curriculum, Pedagogy, and Assessment

Curriculum

Many capabilities in EC 2000's criteria 3a–3k for the accreditation of engineering degree programs are related to traditional values related to liberal education. This means

an engineering degree program needs to place increasing emphasis on skills and affective outcomes. At the same time, EC 2000 stipulates the following curriculum requirements in engineering degree programs: (1) 1 year of university-level mathematics and basic sciences, (2) 1.5 years of engineering topics, and (3) humanities and social studies. It also makes explicit the requirement of a major design experience that not only draws on professional knowledge and skills acquired from coursework, but also incorporates engineering standards and realistic constraints.

The need to strike a balance between professional preparation and liberal education in engineering degree programs is not unique to ABET and, in fact, is common to all signatories in the Washington Accord as is evident in the descriptors for the 13 graduate attributes. In order to meet this challenge, engineering educators have developed innovative approaches in curriculum design, pedagogy, and assessment. In curriculum design, there has been a shift from a compartmentalized approach to a holistic approach, resulting in integration in the following areas:

- Integration of different disciplines (e.g., a multidisciplinary approach to problem solving).
- Integration of theory and practice (e.g., incorporating more real-world examples in the curriculum; creating internships and cooperative programs).
- Integration of professional skills across the curriculum (e.g., teamwork, communication).
- Integration of curricular and co-curricular learning (e.g., service learning).

As an illustration of the first integration (of different disciplines), many engineering schools have introduced curricula aimed at combining technology and management. Instead of the traditional model of engineers pursuing MBA degrees, now engineering students can choose from a wide range of options such as minor in business, major, double major, dual degree, or BS/MS combined degree in technology and management. The intent is to produce graduates who will become managers and leaders in technological companies.

One important curriculum issue is the global competence of engineering graduates. As Lohmann (2003) observed, there are three common elements that require attention: foreign language proficiency and cultural acclimation, discipline-related study abroad, and professional experience in an international setting. Besides offering language and intercultural training, many engineering schools in recent years have increased the number of international students and exchange students. A curriculum needs to offer the flexibility for students to explore meaningful international experiences, be they academic or professional.

Pedagogy

The best-designed curriculum will not work unless its implementation is supported by appropriate learning and

teaching activities and assessment approaches. The consistency among learning outcomes, pedagogy, and assessment is known as constructive alignment (Biggs and Tang, 2007), which is essential in outcome-based education.

The major shift in pedagogy is from a teaching perspective to a learning perspective, drawing from research on how people learn. Even though many engineering courses are still taught in a classroom setting, more active learning approaches such as problem-based learning and collaborative learning have been introduced. These approaches engage students to take more responsibility for their learning, usually by working in teams to solve problems. Problem-based learning (PBL), in particular, has been adopted widely by engineering educators; an entire curriculum based on PBL has been offered by Aalborg University since 1974.

Technology has also led to many pedagogical innovations, including content creation and dissemination, asynchronous learning and teaching, learning management systems, and social networking via the World Wide Web. It is now possible for students from all over the world to work on a team project via web conferencing, thus making it easier to inject international experiences into the curriculum.

The adoption of innovative pedagogies requires a systemic effort on faculty development. Universities worldwide are incorporating pedagogy training as part of the orientation of new faculty. For serving faculty, support for teaching is usually provided in three different formats:

1. short courses or workshops on specific topics;
2. learning communities which enable teachers to share their experiences; and
3. formal training programs leading to certification or qualification.

Some European countries are requiring their university teachers to be formally certified or qualified. As discussed further below, everywhere in the world engineering educators are increasingly engaged in a scholarly discussion of learning and teaching matters, guided by the latest findings in engineering education research.

Assessment

Assessment has always been the most challenging aspect in the outcomes–pedagogy–assessment alignment. With a shift in focus from knowledge outcomes to skills and affective outcomes, engineering educators have responded by developing new assessment strategies (Olds *et al.*, 2005; Shuman *et al.*, 2005). In particular, assessment practices have evolved from the assessment of learning, to assessment for learning, then to assessment as learning.

In assessment of learning, the emphasis is on finding out what a student has learned. Typical assessment tools include examinations, term papers, reports, presentations, etc. The assessment usually leads to a grade, based on norm-referencing (drawing a curve) or criterion-referencing (using a set of criteria).

In assessment for learning, the emphasis is on using assessment to facilitate learning. Students are given clear expectations of the intended learning outcomes and their roles in managing their learning. Then students are given appropriate learning activities, along with specific and timely feedback for improvement throughout the learning process. The assessment is often formative rather than summative in nature, in that, it does not involve a grade.

In assessment as learning, students are actively engaged in the assessment process to the extent that they assess their own work and the work of their peers. Assessment becomes a key feature of learning in this approach. When students are aware of the criteria upon which their performances will be assessed, they usually do a conscientious evaluation of their own and others' performances. Learning portfolios (including e-portfolios) can be an effective tool for capturing this type of self-learning and assessment.

It appears that whereas knowledge outcomes can be reasonably assessed with conventional tools, skills and affective outcomes are best assessed with an integrative approach that collects a body of evidence from different perspectives (self, peers, teachers). Furthermore, an overall assessment strategy should be developed to facilitate learning and not just to grade. At its best, assessment serves three important purposes: (1) to promote the intended learning outcomes, (2) to produce credible evidence of having attained those outcomes, and (3) to help students become self-regulated learners in the process.

Increasingly, universities worldwide are under pressure to produce assessment results that demonstrate not only student achievement, but also institutional effectiveness in its educational delivery. So even though learning-oriented assessments demand a lot of human and financial resources, it is incumbent upon engineering educators to develop coherent and effective assessment methodologies. Both ABET and the Assessment of Learning Outcomes in Engineering (ALOE) Working Group in the United Kingdom have provided useful assessment resources. The Council for Higher Education Accreditation (CHEA) Award for Institutional Progress in Student Learning Outcomes, first introduced in 2006, has identified institutions in the United States that are exemplary in addressing student learning outcomes (Eaton, 2008).

Engineering Education Research

Inherent in an outcome-based approach is a commitment to continuous improvement based on credible evidence. In order to obtain such evidence, engineering educators have to address many complex yet fundamental questions on learning and teaching. Education reforms can no longer be piecemeal and empirical, but need to be supported by a holistic and research-based framework. An important consequence of implementing outcome-based accreditation in engineering degree programs then is the emergence of engineering education as a discipline.

The American Society for Engineering Education (ASEE) launched in 2006 an initiative titled Advancing the Scholarship of Engineering Education (ASEE, 2006). Its intent is to transform engineering education through educational scholarship to better prepare graduates for the twenty-first century. This initiative has since led to a project funded by the National Science Foundation, the main purpose of which is to provide a blueprint for and initiation of actions to rapidly transform engineering education (ASEE, 2008). A special report was also published in 2006 in the *Journal of Engineering Education* (Special Report, 2006) to summarize the research agenda for engineering education. To address the questions of what content (knowledge and skills in context) future engineers must possess; how the said content is being learned; and how learning of the content should be assessed, five research areas have been identified as follows:

1. Engineering epistemologies – research on what constitutes engineering thinking and knowledge within social contexts now and into the future.
2. Engineering learning mechanisms – research on engineering learners' developing knowledge and competencies in context.
3. Engineering learning systems – research on the instructional culture, institutional infrastructure, and epistemology of engineering educators.
4. Engineering diversity and inclusiveness – research on how diverse human talents contribute solutions to the social and global challenges and relevance of our profession.
5. Engineering assessment – research on, and the development of, assessment methods, instruments, and metrics to inform engineering education practice and learning.

Even though the above initiatives are in the United States, similar developments in engineering education research are taking place worldwide. In 2007, an International Conference on Research in Engineering Education (ICREE) was held for the first time; it was renamed as the Research in Engineering Education Symposium (REES) in 2008. These events brought together a global community of scholars to share their works and to discuss future research directions in engineering education.

Summary

Globalization has redefined the meaning of international competitiveness. For companies, it means the ability to operate anywhere in the world and be able to attract talent from everywhere. For individuals, it means the ability to choose the company and location to work for. For universities, it means the ability to attract international faculty and students, to build global alliances in

education and research, and to produce graduates that are highly sought after both locally and overseas.

This increasing mobility of people calls for international standards by which academic qualifications can be meaningfully assessed and compared. The Washington Accord is one such example in defining the attributes engineering graduates should possess for entry into professional practice. The descriptions of graduate attributes usher in an outcome-based approach in accreditation, which is increasingly adopted in other professional fields. Furthermore, in the spirit of facilitating lifelong learning, an outcome-based approach has also been used in the drafting of qualifications frameworks (e.g., the European Qualification Framework), for which diverse learning experiences need to be properly recognized.

The essence of an outcome-based approach in accreditation is its focus on student learning and achievement. By examining critically what a student should be able to do after a learning experience and how best to make that happen, engineering educators have introduced innovations in curriculum design, pedagogy, and assessment. Above all, the move toward making learning explicit has built a culture of collecting research-based evidence. This in turn, has led to the emergence of engineering education research as a discipline, one that brings people together to have a scholarly discourse on how best to improve engineering education and, in turn, engineering practice.

The Washington Accord serves as an excellent example of how the evolution of an accreditation system in response to global trends can positively impact educational delivery. Quality assurance itself, of course, is a continuous process for improvement, and the system should continue to evolve and improve based on partnerships among the stakeholders.

See also: Curriculum and Globalization: Higher Education; Innovation in Teaching and Curriculum Design; Professions – Engineering.

Bibliography

- ASEE (American Society for Engineering Education) (2006). Advancing the scholarship of engineering education: A year of dialogue. *Main Plenary, 2006 ASEE Conference & Exposition*, Chicago, IL, USA.
- ASEE (American Society for Engineering Education) (2008). *Engineering Education for the Global Economy: Research, Innovation, and Practice. Project Prospectus*, American Society for Engineering Education, August 2008.
- Biggs, J. and Tang, C. (2007). *Teaching for Quality Learning at University*, 3rd edn. Berkshire: SRHE/Open University Press.
- Eaton, J. (2008). Attending to student learning. *Change* July/August 22–27.
- Ewell, P. T. (1984). *The Self-Regarding Institution: Information for Excellence*. Boulder, CO: National Center for Higher Education Management Systems (NCHEMS).
- Friedman, T. L. (2006). *The World is Flat: The Globalized World in the Twenty-First Century*. London: Penguin Books.

- Lattuca, L. R., Terenzini, P. T., and Volkwein, J. F. (2006). *Engineering Change: A Study of the Impact of EC2000*. Baltimore, MD: ABET.
- Lohmann, J. R. (2003). Will our graduates be global players? *Journal of Engineering Education* **92**(3), 195–196.
- Olds, B. M., Moskal, B. M., and Miller, R. L. (2005). Assessment in engineering education: Evolution, approaches and future collaborations. *Journal of Engineering Education* **94**(1), 13–25.
- Shuman, L. J., Besterfield-Sacre, M., and McGourty, J. (2005). The ABET 'professional skills' – Can they be taught? Can they be assessed? *Journal of Engineering Education* **94**(1), 41–55.
- Special Report (2006). Special Report: The Research Agenda for the New Discipline of Engineering Education. *Journal of Engineering Education* **95**(4), 259–261.
- Vest, C. M. (2006). Educating engineers for 2020 and beyond. *The Bridge* **36**(2), 38–44.
- Ewell, P. T. (2001). *Accreditation and Student Learning Outcomes: A Proposed Point of Departure*. Washington, DC: Council for Higher Education.
- Lucena, J. C. (2006). Globalization and organizational change: Engineers' experiences and their implications for engineering education. *European Journal of Engineering Education* **31**(3), 321–338.
- McKeachie, W. J. and Svinicki, M. (2005). *McKeachie's Teaching Tips*, 12th edn. New York: Houghton Mifflin.
- National Academy of Engineering (2004). *The Engineer of 2020: Visions of Engineering in the New Century*. Washington, DC: The National Academies Press.
- National Research Council (2000). *How People Learn: Brain, Mind, Experience, and School*, (expanded edn.). Washington, DC: National Academy Press.
- Ollis, D. F., Neeley, K. A., and Luegenbiehl, H. C. (eds.) (2004). *Liberal Education in the Twenty-First Century Engineering*. New York: Peter Lang.
- Strayhorn, T. L. (2006). *Frameworks for Assessing Learning and Development Outcomes*. Washington, DC: Council for the Advancement of Standards in Higher Education.

Further Reading

- Augusti, G. (2006). Transnational recognition and accreditation of engineering educational programmes in Europe: Perspectives in a global framework. *European Journal of Engineering Education* **31**(3), 249–260.
- Baillie, C. and Moore, I. (eds.) (2004). *Effective Learning and Teaching in Engineering*. New York: RoutledgeFalmer.
- Duderstadt, J. A. (2008). *Engineering a Changing World: A Roadmap to the Future of Engineering Practice, Research, and Education*. Ann Arbor, MI: The Millennium Project, The University of Michigan.

Relevant Websites

- <http://www.abet.org> – ABET.
- <http://www.abet.org> – ABET, Assessment Planning.
- <http://www.aloe.ac.uk> – ALOE.
- <http://www.washingtonaccord.org> – International Engineering Agreements.

This page intentionally left blank

HIGHER EDUCATION

Higher Education: An Overview

P Scott, Kingston University, Kingston upon Thames, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The development of modern higher education has been one of the most profound but least celebrated social revolutions of the past half-century. The development of much more extensive, and also more open, systems of higher education was not only the culmination of a long educational revolution which began in the nineteenth century when national systems of elementary education first emerged in Europe and North America, it also empowered, and reflected, a more far-reaching revolution in occupational structures, individual identities, gender relations, and even lifestyles. The impact of higher education on society, the economy, and culture continues to increase. In his work on clever cities the American scholar Richard Florida has ably described the nexus of creative freedom, social fluidity, and economic dynamism characteristic of the most dynamic twenty-first-century societies – and which can be substantially linked to the development of mass higher education. No single phenomenon perhaps has done more to determine the feel (or, in a more profound sense, the ethos) of contemporary society than the presence within them of millions of citizens who have experienced higher education. No single phenomenon has had a greater influence on the life-chances of individuals. No single phenomenon has had a greater impact on the economy through science and technology and also through the diffusion (and redefinition) of expertise. In retrospect, the creation of mass higher education systems may come to be seen as the greatest, and best, achievement of the twentieth century – overshadowing its more obvious legacies of genocide, war, and social conflict – and the best hope for the twenty-first century.

Moreover the development of contemporary higher education systems has been a global phenomenon – in two different senses. First, it has been global in its reach.

The expansion (and diversification) of higher education has not been confined to the so-called “West” – the north Atlantic world of the United States/Canada and Western Europe, with its outliers in the Antipodes. It has had an equal, if not greater, impact in the Middle East, South and East Asia, Latin America, and even Africa despite the difficulties that the continent has encountered. Only two regions were initially less affected by the development of mass higher education systems – Central and Eastern Europe, and the People’s Republic of China, where more restricted systems persisted until the last decade of the twentieth century. Second, the development of modern higher education systems has been intimately linked with the phenomenon of globalization. These systems have been shaped by globalization as the increasing flows of students across national frontiers testify – and as the ideologies and practices of global capitalism have eroded older statist conceptions of higher education as a public space insulated to some degree from the pressures of the market. But the same systems have also been vital contributors to globalization through the pivotal role played by universities in the emergence of virtual/global scientific communities and their more diffuse role in the development of global cultures embodied in mass graduate populations.

This overview is divided into two main sections bracketed between this brief introduction and a concluding summary:

1. The first section looks back. It examines in turn (a) the development of mass higher education systems since the 1960s; (b) the major policy drivers for this massification (and different policy responses to it); and (c) the various conceptualizations of this phenomenon.
2. The second section looks forward. It considers three major clusters of characteristics of contemporary

higher education systems: (a) their relationship with the economy; (b) interactions between higher education and politics and society; and (c) linkages with wider issues of identity and culture.

Looking Back

The Development of Mass Higher Education

The most significant phenomenon associated with the development of higher education has been expansion – of the number of students enrolled and also of the number of institutions. In 1900, it was estimated there were fewer than 500 000 students enrolled in higher education in the entire world, a tiny fraction of 1% of the relevant age group. By the end of the twentieth century, according to UNESCO, there were more than 100 million university and college students, one-in-five of young adults. In absolute terms, therefore, higher education expanded by more than 200 times, and in real terms by more than 20 times, during the century. In most developed countries, typically more than half of the young adults are now enrolled in higher education (although definitions of higher/tertiary education remain fluid, which make precise trans-national comparisons difficult). But expansion has not been confined to developed countries. Algeria now has more students than the entire world at the beginning of the twentieth century. Although participation levels are not the same – there are approximately 350 students per 10 000 people in developed countries compared with only 45 per 10 000 in sub-Saharan Africa (in most other regions the figures vary between 200 (Asia) and 250 (central and south America)) – the pattern of expansion has been remarkably similar, consistent and accelerating growth.

In most developed countries, the development of different levels of education was a sequential process beginning with primary education, then moving on to secondary education, and culminating with higher education. In the United Kingdom, for example, more than two generations separated the establishment of universal primary education in 1870 from the creation of compulsory secondary education in 1944, followed less than a generation later by the rapid expansion of higher education in the 1960s. In many developing countries, in contrast, the development of various levels of education has been a near-concurrent process, with expanded higher education systems being created almost in step with the growth of universally (or widely) available primary and secondary education (a phenomenon which has been criticized by the World Bank concerned about skill imbalances and graduate unemployment). With the (hopefully temporary) exception of sub-Saharan Africa, there appears to be a worldwide process of leveling-up in higher education enrolments.

The expansion of higher education, therefore, is a pervasive phenomenon, an expression of the *longue durée* of modern society rather than of transient political or economic circumstances. The number of university and college students continued to increase throughout the twentieth century, despite revolutions and world wars. Expansion also took place in a wide range of countries with different social structures, political and economic systems, and cultural traditions. The differences in growth rates that can be observed are best regarded as secondary phenomena. Decolonization in Asia and Africa clearly encouraged the more rapid expansion of higher education, but on foundations laid down during the colonial era. In Central and Eastern Europe, during the Communist period, growth rates lagged behind those in Western Europe, but still broadly followed the same pattern (and then accelerated rapidly after 1989). Clearly emancipatory episodes, such as the decline and fall of colonial and Communist empires, have played a role in stimulating the growth of higher education – as has the development of more comprehensive democratic cultures in Western societies; there were intimate links between higher education expansion and the New Left during the 1960s, as there have been more recently between expansion of new social movements, notably feminism. Also, changes in occupational structures and definitions of expertise have generated new demands for skills which only higher education can meet.

The expansion in the number of institutions has been equally remarkable. The great majority of universities and colleges has been established since 1900, and the majority since 1960. After the medieval period and nineteenth century in Europe, the past 50 years has been the third great age of university foundations, this time on a global scale. Typically, this growth in the number of institutions has taken two forms – the establishment of brand-new institutions; and the promotion of existing technical and professional schools:

1. The establishment of new campuses of the University of California at Santa Cruz, Santa Barbara, and San Diego in the 1960s was a good example of the first form of growth in the number of institutions – as was the near-contemporaneous creation of new universities in the United Kingdom (also sometimes called the Shakespearean universities because of the cities in which they were established or the names they chose – York, Lancaster, Warwick, Essex, Sussex, and so on). In most developed countries, similar new universities were established between the 1950s and 1980s, with teaching and residential accommodation and research facilities grouped together on (often expensively) designed campuses rather than sprawling incoherently across cities; animated by similar values, the integration of teaching and research and a holistic vision of

higher education combining traditional disciplinary and radical interdisciplinary elements; organized according to similar principles, essentially those of the so-called academic guild; and funded discreetly but generously by the burgeoning welfare state. Existing universities were heavily influenced by this model, as they attempted to create academic precincts out of the urban clutter they had inherited and adopted the academic and organizational values of the newly established universities. Since the 1990s, new institutional types have emerged, notably in Central and Eastern Europe, and South and East Asia where the number of private higher education institutions based on corporate (and managerial) rather than academic (and collegial) principles has rapidly increased. In a similar way, these new institutions have influenced the behavior of existing universities.

2. The second form of growth in the number of higher education institutions displayed a less-consistent pattern. In broad terms, as the educational and so-called administrative scope of higher education systems was extended (and, indeed, the very idea of systems of higher education comprising more than traditional universities came to be accepted), the effect was to re-designate institutions variously labeled technical, professional, or extended-secondary as part of these new systems. Two approaches were taken (which are discussed in greater detail in the next section). The first was simply to re-categorize as part of higher education, existing institutions such as the *Fachhochschulen* in Germany, now usually described in English as universities of applied science. The second was a more deliberate attempt to create larger non-university institutions through a process of mergers to act as a counter-weight to the traditional universities. The best example was perhaps the establishment of the polytechnics in England and Wales, although similar approaches were also taken in Australia and the Netherlands. These institutions placed greater emphasis on professional formation and applied research and less on providing a holistic higher education and undertaking traditional academic research – at any rate, initially. Distinctions between different institutional types – whether traditional universities already established by the middle of the twentieth century, new universities established between 1950 and 1980, polytechnic-like institutions created through restructuring processes, or unadorned higher vocational/professional schools – tended to be eroded as all institutions have gravitated toward a more instrumental and corporate model.

Policies, Structures and Funding

To summarize the evolution of higher education policies across several decades and many countries, developed and

developing, is clearly a difficult task. However, a small number of pervasive trends can be identified:

- the shifting balance between policies reflecting push factors, predominantly the result of deeper social change, and pull factors, predominantly reflecting the needs of the economy as identified by governments, as the major driver of higher education policy;
- the establishment of better-articulated binary systems of higher education under state direction followed by greater, but more customized, differentiation of institutional missions in response to the emergence of quasi-markets in higher education; and
- a parallel drift from predominantly state funding of universities and colleges to more mixed-funding regimes in which students, and other users of academic services, are expected to make more substantial contributions.

Push and pull

During the second-half of the twentieth century, and up to the present, the main engine of higher education expansion has been increasing demand from potential students, even if the extra resources required to fund that expansion have often been justified by governments in terms of the need to develop a more highly skilled (and increasingly graduate) workforce. Certainly, in developed countries, the growing number of qualified secondary school leavers has been the major factor in increasing participation in higher education, while attempts to steer this growing demand into subjects regarded as more important for economic development (often reduced, misleadingly, to the so-called STEM subjects – science, technology, engineering, and mathematics) have produced disappointing results. In the UK, for example, the two periods of most rapid higher education growth can be attributed to wider changes in educational policy which had the effect of increasing the pool of available candidates – in the late 1960s following the introduction of comprehensive secondary schools; and in the late 1980s and early 1990s following the introduction of a common system of secondary school examinations. Similar reforms in other countries had the same effects. In that sense, it is fair to say that push factors have predominated over pull factors, market forces (or, more accurately, social and cultural pressures) over workforce planning.

However, this conclusion must be qualified in two respects. The first is that student demand has become increasingly heterogeneous. Until the 1960s, in Europe (earlier in the United States and later in many developing countries), student demand was largely an expression of the aspirations of young adults, predominantly male and from comparatively privileged social backgrounds already possessing considerable cultural capital. These aspirations

could readily be translated into traditional patterns of higher education provision. Today student demand is an expression of the aspirations of much more diverse social groups – women (the revolution in gender relations inspired by feminism has been a major factor in the development of nearly all higher education systems) and older adults (shifting career patterns and new occupational structures have combined with notions of higher education as an agent of socio-cultural credentialization as well individual enlightenment/enjoyment to promote ideas of lifelong learning), as well as students from less-privileged social backgrounds and minority cultures (reflecting the strengthening emphasis, in many developed countries at any rate, on equal opportunities). These much more diverse, and even conflicting, aspirations have proved to be more difficult to translate into stable patterns of higher education provision.

The second respect in which this conclusion must be qualified is that there has been a trend for both push and pull factors to strengthen in recent years – the former because access to higher education has become more crucial as a determinant of middle-class culture as older class, gender, and ethnic determinants have become less decisive (or as a means of transcending these older determinants, which helps to explain the rapid expansion of higher education in some developing countries even in the face of significant levels of graduate under/unemployment); the latter because in post-industrial society knowledge resources, whether in the form of highly skilled technical experts or of research and innovation capacities, are regarded as increasingly crucial for economic success. As a result governments have placed more emphasis on the instrumentality of higher education, urging universities to forego traditional academic priorities and to work more closely with business and industry. They have also redoubled their efforts to steer higher education, establishing special programs to promote STEM subjects or employer engagement. This shriller rhetoric has tended to mask the underlying driver of higher education development, social change as mediated through student demand.

Systems and markets

The second pervasive trend has been structural reform, as the number of higher education institutions has increased and organized systems of higher education have been established. Structural reform has taken two main forms:

1. The first has been the creation of well-articulated systems in which institutions are assigned clearly defined roles, usually arranged in a hierarchy with explicit arrangements for student progression between tiers. The best examples remain the state systems in many individual US states, with the California Master Plan as an archetype, but they have been emulated in

many other countries with different social systems. The Chinese pattern of key universities, both at national and provincial levels, is another example of a planned stratified system. These systems are both open – in the sense that they typically embrace the totality of post-secondary education and encourage student progression – and planned, in the sense that they restrict masters and doctoral programs (and, by extension, research) to top-tier institutions.

2. The second has been the, often more chaotic, development of binary systems which preserve the distinction between academic (or scientific) and professional higher education. The best example was in England and Wales where polytechnics were established in the 1970s; current examples include the distinction between universities and *Fachhochschulen* in Germany. Typically, the development of binary systems has passed through various phases – starting with an initial recognition that higher education was more than universities and leading to efforts to articulate alternative, more vocational, forms of higher education. Binary systems were emancipatory initially, because they enabled new groups of students (and new disciplines and professions) to access higher education (albeit on subordinate terms). More recently, while struggling to contain academic drift, they have tended to be perceived as educationally rigid and socially restrictive.

Both stratified and binary systems have adapted in different ways to the strengthening of the market in higher education. Stratified systems, on the Master Plan model, have made it easier to identify, and to sustain, major research universities – the top universities which feature in global league tables – but at the cost perhaps of restricting the adaptability and responsiveness of non-elite institutions. Binary systems, in contrast, have made such identification and sustenance more difficult because of their inbuilt reluctance to discriminate between traditional universities and their failure to contain academic drift. But their categorization of institutions in terms of broad educational mission rather than by means of tight administrative classification has perhaps made it easier for individual institutions to respond to market opportunities, especially when binary systems have tended to fray (or, as in Australia and the UK, be superseded by unified but differentiated systems).

However, both stratified and binary systems are ideal types. There are many different variants which may even appear to have more in common with systems of the other type than others systems of their own type. For example, some apparently binary systems have more than two sectors while some stratified systems have become more permeable. Also both ideal types have been subverted by three factors:

1. The first, and more familiar, is the impact of private institutions which has been analyzed by Philip Altbach

and David Levy. Although in the United States the most visible part of the private sector has played a relatively stable, and elite, role, the past two decades have seen the rapid proliferation of more dynamic, but also much more populist, private universities and colleges – for example, in Central and Eastern Europe and in India (often concentrating on subjects such as business and management and information technology). The influence of the private sector has been compounded by the semi-privatization of key elements within publicly regulated institutions.

2. The second factor is globalization. Even in the less intense form of internationalization, this factor enabled institutions to escape from their nationally designated roles in either stratified or binary systems. For example, higher professional schools in the Netherlands have used international partnerships to allow them to engage in activities formally reserved for traditional universities in the Dutch system. The impact of full-blooded globalization on higher education systems is likely to be similar to its impact on economic systems. It has even been argued that the days of national higher education systems are numbered.
3. The third factor is the emergence of regional groupings. The best example is the European Higher Education Area (EHEA) established as a result of the Bologna process which has led to increasing coordination of structures and policies across more than 20 European countries. But universities themselves, especially research universities, have taken the initiative and formed trans-national associations, for example Universitas 21.

Funding and fees

The third pervasive trend has been the, still unresolved, debate about how mass higher education systems should be funded – by the state through taxation, the individual through fees, or by some combination of the two (and other income sources such as endowments or user payments). The standard assumptions, strongly promoted by the World Bank and more tentatively supported by the OECD, have been that it is no longer possible to rely on state funding and that students therefore must make an increasing contribution to the cost of their higher education. Fee regimes have been progressively introduced into some higher education systems which were previously funded almost entirely by the state – beginning with Australia where the Higher Education Contribution Scheme (HECS) was introduced in the mid-1980s, followed by the UK between 1998 and 2004. Tuition fees have also been introduced in the Netherlands and, provisionally because the legal basis for charging fees remains unclear, and in some German *Länder*. At the same time, the low-fee regimes that prevailed in many US state

systems between the 1950s and 1980s have been progressively replaced by moderate-fee regimes, while American private universities have substantially raised their fees. Charging for tuition has also been adopted by several African countries to ensure minimum funding levels for their universities. Only in Nordic and Mediterranean Europe, the Middle East, and Latin America has the introduction of student fees been successfully resisted. The progressive introduction of student fees appears to amount to a global, and inescapable, trend.

However, this noisy debate about fees needs to be seen within a wider context – in two important senses:

1. First, charging tuition fees has been simply one element, and not the most important element, in both the funding of higher education institutions and financial support for individual students. In the case of institutional funding, questions have been raised about the extent to which charging more for tuition might have skewed university and college priorities because it has encouraged students to adopt a more pronounced consumerist mentality – for example, toward student services and away from basic research. There is some evidence this has happened in both US private and state universities which have increased fees substantially. In the case of student support, many other factors apart from tuition fees have had to be taken into account – including the availability of loans (or grants) to cover living expenses, other social benefits for which students have been eligible, and the impact of tax regimes (in particular, the scope for off-setting the cost of loan repayments against tax liabilities). It is also important to recognize that fee regimes have differed substantially in their details. Some have been variants of a graduate tax – for example, when the state has provided subsidized loans for students to pay their fees. In practice, there have been very few examples of pure-fee regimes in public higher education systems which require students to pay the fees either out of their (or family) resources or to take out commercial loans.
2. Second, the debate about fees has often served as a proxy for other important debates. One debate concerned the overall resourcing of higher education (whether those resources were provided by individuals in their roles as taxpayers, fee-payers, or consumers, was a secondary question). The issues included whether universities and colleges have been receiving an adequate share of national GDPs, and the extent to which they should have been expected to take advantage of new learning technologies, and to exploit global economies of scale in research, to improve their efficiency. Another important debate has been about access – what proportion of the population should be educated to higher education level, and on what terms?

Some advocates of fees have seen charging for tuition as a rationing device, to restrict access to the truly worthy and motivated (and, in a few cases, simply to restrict access). Others have justified charging fees in terms of social equity, because by abolishing indiscriminate subsidies to the more privileged (the effect, they argue, of funding higher education through taxes because tax systems have become less progressive since the waning of the welfare state), it has become possible to target financial support on students who are in genuine need. A third, and more ideological, debate has been about the marketization of higher education. Advocates have often regarded the introduction of fees as the first stage in a much wider process of modernization (in a domestic context) and liberalization (in an international context).

Theories of Higher Education Development

Until the middle of the twentieth century, little was written about higher education from a sociological or economic perspective. Before the advent of mass higher education, the bulk of scholarly writing about universities was historical or philosophical in orientation – or merely rhetorical. Nor did mainstream sociologists or economists devote much attention to higher education as a field of enquiry. Only in exceptional circumstances, for example, the Nazi period in Germany and its aftermath, were urgent questions about the role of universities in society seriously explored. As a result, higher education remained a largely under-theorized domain before the 1960s. It was the development of mass higher education systems in this and subsequent decades that encouraged the more systematic development of theories of higher education development. First, the establishment of new universities, often on eye-catching campuses which gave them a high public (and political) profile, demanded a more sophisticated articulation of their rationale – and, in particular, whether this rationale was different from that for the existing, more traditional, universities. Second, the extension of higher education systems to embrace non-university institutions also encouraged more systematic reflection on the purposes of these wider systems – and, in the case of binary systems, whether these institutions were intended to represent a new paradigm of higher education.

Two broad bodies of theory have developed which can be, approximately but not exactly, related chronologically to different phases in the development of mass higher education. The first was broadly functionalist in approach. A pioneering work by Talcott Parsons on the American university set the tone, a strand that continued in the work of younger colleagues such as Neil Smelser. But the most frequently quoted conceptualization was that developed by Martin Trow in the late 1960s and 1970s. He offered a three-stage evolutionary model – beginning with elite

higher education dominated by traditional universities and characterized by restricted access (with up to 10% of the relevant age group enrolled); moving on to mass higher education with much higher levels of participation (up to 40% of the age group) and with a greater variety of institutional types; and culminating in universal higher education embracing, in effect, the whole of post-secondary education. Two comments were made about Trow's typology – first, that it could be seen as a theoretical analogy of the three-tier structure of Californian higher education institutionalized in that state's Master Plan; second, that it appeared almost to be an endogenous theory in which wider social-economic (and scientific/cultural) change was either absent or treated as a given (as was ideology, despite the contemporary emergence of the New Left). The latter deficit was remedied to some extent through the work of a fourth American scholar, Burton Clark, on the connections between the academic system, that is, higher education and research, and the social, political, cultural, and other major systems of modern society.

It is not perhaps an accident that in this first phase, from 1960 until the mid-1980s, much of the theorizing about higher education development originated in the United States. America had been first to experience mass higher education; and the expansion of higher education was seen as closely related to the development of the post-war welfare state. But in the second phase, since the mid-1980s, two new factors have been encountered which tended to invalidate some of the assumptions prominent in these theories. First, the actual development of higher education systems, as empirically observed, appeared to have diverged in key respects from this theoretical path. For example, although most developed countries now enrol more than 50% of the age group in higher education (and, therefore, in Trow's typology, had reached the universal phase), these systems do not appear typically to exhibit the characteristics of universal systems which had been predicted. One reason for this divergence is that the knowledge societies which emerged in the first decade of the twenty-first century have not followed the model of post-industrialism imagined by Daniel Bell and others, a quarter of a century ago. Instead of developing into techno-rational and emancipatory forms, they display more disturbing characteristics, combining emotive ephemerality with control and surveillance.

Second, the rationalistic/positivistic styles that characterized intellectual life in the third-quarter of the twentieth century have been challenged by new ideas. These ideas have been extremely heterogeneous (indeed, this heterogeneity has been their defining characteristic). They have ranged from post-structuralism which later broadened out into more radical expressions of post-modernism; through new definitions of authoritative

knowledge which revived older notions of social constructivism and/or emphasized the contextualization of scientific production, the proliferation of knowledgeable actors, and the reflexivity of scientific exchanges; to more accessible accounts of risk, globalization, and markets as generators of chronic uncertainties. At the same time, new ideas in political science and organization theory have called into question the stability and predictability of the higher education landscape. As a result, higher education has come to be seen as terrain that, both academically and organizationally, is volatile and contested, turbulent, and unpredictable.

The shift from systems to markets in the organization of higher education, therefore, has been mirrored in a similar shift in theories of higher education development, from the linear to the reflexive and from the simple to the complex. Examples include Ron Barnett's work. This second broad body of theory also emphasizes the increasingly intense, but also unstable and transgressive, exchanges between the academic system and other systems. Although more intellectually stimulating, this body of work may be of less immediate help to policymakers and institutional leaders than the theoretical accounts it displaced. However, alongside these meta-theoretical accounts of higher education development, a second strand of academic work has emerged that combines empirical research evidence with more detailed theory – for example, in economics and psychology. Sophisticated rate-of-return analyses have been developed, stimulated to a large extent by the debate about fees and markets in higher education. Theoretically well-informed studies of student learning have also proliferated. This second strand of theoretical work has been much more relevant to the development of public policy; indeed, its direct relevance has occasionally raised questions about degrees of critical distance (if not its scientific objectivity).

Looking Forward

Few people are expecting higher education to expand as fast in the next century as it expanded in the past century. As participation rates currently exceed 50% in most developed countries, even near-universal levels of participation would require substantially less than a doubling in the size of most higher education systems. Also, in many developed countries, the number of young adults, still higher education's core constituency, will decline in the years leading up to 2025 – although, there will be significant differences between socioeconomic and ethnic groups and the overall size of the population will continue to increase because of longer life spans (which could have a significant impact on those higher education systems that adopt a genuinely lifelong learning orientation). Demographic trends beyond 2025, of course, are only

projections subject to revision. It is also important to note that migration patterns could have an important effect on population profiles even before that date. Nevertheless, the consensus is that, in developed countries at any rate, the scope for rapid quantitative expansion is limited. In many developed countries, the scope is greater. But even here, substantial progress has already been made toward developing mass higher education systems with relatively high levels of participation, and predictions suggest that global population levels will stabilize as a result of changes in fertility by the end of the century.

However, the past century has also witnessed a fundamental transformation in both the scale and the purposes of higher education. This mirrored an equally fundamental transformation in wider society – the national state was transformed into the welfare state which is now evolving into the so-called market state; the industrial economy based on manufacturing has been succeeded first by a services economy (with a large public, or social, sector) which is now being subsumed into the post-industrial knowledge-based economy, elite culture has been supplemented by a mass culture, and individual lives have been transformed by affluence (or, more pejoratively, *affluenza*), radical shifts in gender relations, and (not least) participation in mass education. The dominant consensus is that social and economic change will accelerate as the twenty-first century advances. So, even if the scope for quantitative change in higher education systems is limited, the scope for qualitative change could be unbounded.

Higher Education and the Knowledge Economy

Of the three relationships which are discussed in this section, the key one is seen as the relationship between higher education and the knowledge economy. In the eyes of most politicians and other policymakers, and also of an increasing number of university leaders (although not so many of their academic colleagues), this relationship is seen as central to the case for continuing public investment in higher education (and also the primary justification for private investment by individuals when they are paying fees). However, even if the dominance of this relationship is accepted, the implications for universities are far from clear. The central argument is that higher education has become the powerhouse of the new knowledge economy, producing not only highly skilled graduates but also the primary research which powers this new economy. But there are counter-arguments:

1. The first is that the knowledge economy is not as novel as is often supposed. Symbolic knowledge of all kinds has always been crucial to the working of human societies; what is new is the explosion of technical knowledge (and especially of technical processes for data management). It can be argued that the role of the university in

creating symbolic knowledge was (and is) at least as great as its role in producing technical knowledge – a role which can be undertaken by other knowledge organizations. In other words, Apple or Google are the leading institutions of the twenty-first-century knowledge economy, not Harvard or Oxford.

2. The second counter-argument is that in the knowledge economy/society, all organizations have to be suffused with knowledge – by active producers as much as by passive consumers of knowledge. Many of the most successful private-sector organizations value not only the technical knowledge most relevant to their operations, the intellectual property which they generate, and the technical experts they employ, but also the entire knowledge resources represented by their wider workforces; this explains the increasing emphasis on organizational development. These organizations often regard the nurturing of these wider knowledge resources as crucial in terms of securing competitive advantage. The implications of such thinking for universities are far-reaching. Instead of being privileged primary producers in the new knowledge economy, they must work with other knowledge producers in collaborative networks.

There are four main ways in which universities contribute to the knowledge economy: (1) as primary producers of basic research which can be translated into commercially valuable or socially useful goods and services; (2) as agents through which this translation process can take place, by means of technology transfer, knowledge management, and other entrepreneurial activities; (3) as producers of skilled people, from future researchers through a mass graduate workforce to skilled technicians; and (4) as agents of socialization whereby large numbers of young adults can be inducted into optimistic rationalizing ideologies – science, democracy, national development, perhaps the free market. It is often assumed that the first is the special responsibility of research universities (an unproblematic category in the US, but a more contentious – and politically divisive – one in many other countries). The second is seen as the special responsibility of regional universities (or other higher education institutions) with more direct experience of engagement with business and industry, and local communities. The third is assumed to be the responsibility of all institutions – but segmented by level, that is, future researchers are most appropriately trained in research universities, while technician education is the primary responsibility of higher vocational schools. The fourth is similarly the responsibility of all, but the reproduction of elites is still regarded as the quasi-monopolistic responsibility of leading universities.

But this apparently neat categorization of responsibilities may become increasingly difficult to sustain. There

are likely to be two flash-points. The first will be over the future funding of research. In the past two decades, research assessment systems have proliferated – often with the explicit intention of concentrating this funding in a smaller number of research-intensive universities (although in the process of their operation these systems have often, unwittingly, revealed how widely distributed high-quality research has become in mass higher education systems). Efforts have also been made in some countries to encourage regional universities to concentrate on (in the terminology adopted in the UK) third leg activities – in other words, applied and translational research, technology transfer, and other forms of dissemination. But both have produced disappointing results – as well as significant political friction. The main reason is that, in the context of mass higher education and of a pervasive knowledge economy, it no longer makes sense to regard the generation of knowledge as a restricted activity. Instead, it has become a reflexive rather than a linear process, with multiple actors rather than primary producers. Fundamental discoveries can be made in near-market, or socially embedded, environments (just as would-be applied research can be undertaken in academic environments as Michael Gibbons and his colleagues have discussed in two recent books). In other words, not only the boundaries between research-intensive universities and other higher education institutions have become more permeable, the frontiers between the academic and the economic systems have also been transcended within a broader system of innovation.

The second flash-point is likely to be over the differentiation of institutional missions within mass systems, often coded for maintaining (or re-imposing) a more categorical hierarchy of institutions rather than a recognition of the need for all institutions to diversify to respond to a widening variety of social and market demands. But the dynamics of the new knowledge economy, with its proliferation of symbolic and expert knowledge(s), hard and soft forms of knowledge, tend to subvert and frustrate all attempts to produce a clear taxonomy of institutions. As a result, many universities, whether research intensive or socially engaged, have had to extend their missions to embrace new roles and activities. It is this incorporation of once-peripheral activities into the core that has characterized the development of most higher education institutions, more so in systems that have abandoned traditional distinctions between universities and higher professional education and less so in systems which have continued to be more stratified or segmented. This phenomenon of mission stretch, rather than so-called academic drift, is likely to become more significant in coming years. This may place more emphasis on the resourcefulness of institutional leaders and managers and less on the resilience of systems.

Higher Education and Political Economy

The development of twenty-first-century societies has also transformed the links between higher education and the wider political and social environment. It is argued that a fundamental shift has taken place in the dominant social form – from the state (and, especially, in the second-half of the last century, the welfare state) to the market (and, increasingly as the twenty-first century unfolds, the global market). In practice, this broad assessment needs to be qualified in a number of ways. First, the state has not withered away; despite the popularity of political rhetoric about the need for a small state, in practice, the state is as pervasive as ever. Second, markets and globalization are hardly new phenomena; from its birth, the university, as its name suggests, was imagined as transcending national frontiers, and global trading empires have existed for at least half a millennium. However, the shift from state to market may still serve as a good enough narrative, a metaphor which captures a key characteristic of twenty-first-century society.

This shift is often aligned with another transition – from modernity to post-modernity. As a narrative, the welfare state, with its Enlightenment values and its Weberian practices, displayed many of the features associated with modernity – progress defined in normative as well as materialist terms, regularity (even standardization), and scientific objectivity. The market, also as a grand narrative, displays many of the features associated with post-modernity – a moral fluidity, volatility, and subjectivity (but also, it is argued, creativity and innovation). However, there are two ways of thinking about the shifting balance between the state and the market. The first is a literal interpretation which suggests that the state is in actual decline; and, therefore, that many of its responsibilities (e.g., for providing higher education) must now be shouldered by the market. The second is a more nuanced account which suggests that both the state and the market are being infiltrated by new values, which can be labeled a “post-modern.” The state is behaving increasingly as an entrepreneurial agent, the prime responsibilities of which are not only to create the conditions for economic growth but also to create quasi-markets in the public sector, through a combination of privatization and the so-called new public management. The market now is not only characterized by new (and outlandish) trading instruments, notably in financial services, but also by the intersections between the state and the market. These extensive borderlands of regulation and privatization are becoming both more complex and more significant – and arguably the key sector of advanced twenty-first-century political economies.

The implications for higher education of this shift from the state to the market are potentially far reaching:

1. If a literal interpretation is preferred, it may imply that the educationally liberal and socially progressive orientation, and academic (or scientific) values, of the twentieth-century university are likely to become less powerful. The post-War university was clearly closely aligned with the welfare state – not only quantitatively, in the sense that the expansion (and diversification) of higher education systems after 1945, and especially after 1960, was seen as the culmination of long social and educational revolutions that reached back into the nineteenth century, but also qualitatively, in the sense that critical intellectual values were seen as at the heart of the modern university experience (in parallel with the liberal/social-democratic values that suffused post-War European societies, even when ruled by conservative governments). Maybe in the twenty-first century, this will change – both the governance of universities, where academic guilds and state bureaucracies will be superseded by more entrepreneurial models, and the funding of both students and institutions, where fees (and other quasi-market mechanisms) will replace state grants. Both trends appear already to be well established.
2. However, if the second more-nuanced interpretation of the shift from state to market is interpreted in more metaphorical terms, a rather different future for universities may emerge:
 - First, the state may retain, and even strengthen, its control over higher education. There are few signs that the state is even prepared to reduce its detailed operational control over universities. Instead, this control is being exercised in different ways – no longer through traditional administrative regimes, which in practice gave universities considerable room for tactical (and strategic) maneuver, but instead, through increasingly elaborate accountability regimes, from which there is no escape. The cult of targets is already a dominant feature of the new audit state as discussed in books by Michael Power and Marilyn Strathern. The state has also reinvented itself as a contractor with quasi-autonomous (but, in practice, heavily regulated) universities as providers of academic goods and services, the skills of graduates and the outputs of research. For example, in England and Wales, politicians talk in terms of “something for something” – in other words, the government will provide increased funding for universities only in return for specified deliverables. This heightened state intervention is justified by a new discourse that emphasizes the competitiveness of the global knowledge economy.
 - Second, markets too are changing. The production of goods and services, organized through corporate bureaucracies, is no longer the dominant form in the post-modern markets that flourish in many developed countries; these markets are instead characterized by new patterns of behavior, both ultra-creative

in terms of the generation of novelty (including the invention of ever-more complex and inventive financial instruments) and also ultra-volatile (and ephemeral) because of the instantaneities of modern information and communication technologies. Classical capitalism of the industrial age, or even of the social markets that flourished in the era of the welfare state, might have offered an alternative model for the organization of higher education systems to the dominant state-control model; they possessed their own patterns of stability and regularity. But it is more difficult to imagine how the hyper-capitalism of the twenty-first century can provide such an alternative model. In practice, higher education systems are likely to be stranded in the state-market borderlands of targets and regulation that are the second (and perhaps less dynamic) form of twenty-first-century political economy.

Higher Education and Identity and Culture

The third set of relationships is between higher education and wider sociocultural issues – of identity, for individuals; and culture, for society at large. In terms of the *longue durée* of human civilization these relationships are potentially the most significant; but they are also the most difficult to describe. There have been comparatively few studies of the impact of the expansion of higher education on the cultural characteristics of contemporary societies, their ethos, although empirical studies have demonstrated that graduates live longer, enjoy better health, participate more in the political process, and have other positive attributes. Developed countries now possess mass graduate populations where once they only contained small university elites, a change of historical significance in terms of democratic culture and economic capacity. However, the impact of mass higher education has often been regarded as an essentially secondary phenomenon, a subordinate element within much larger transformations such as the erosion of traditional elites, the decline of class-based politics, and revolutions in family structures and gender roles.

However, it is possible to argue that the expansion of higher education over the past half-century has been a phenomenon of primary significance. The impact of expansion has been felt in two distinct ways:

1. The first is familiar. There has been a rapid increase in the number and proportion of highly skilled jobs, for which academic and professional credentials are often required, to keep pace with changes in the occupational structure of the knowledge economy. But it may be a mistake to concentrate exclusively on traditional forms

of scientific, technical, and expert credentialization. In post-modern societies, not only has the number of occupations that require university-level qualifications increased – for example, nurses and other healthcare workers as well as doctors, or policemen as well as lawyers. More subtle and diffuse, but equally significant, forms of social credentialization may also have been at work.

2. This social credentialization is the second way in which mass participation in higher education has had an impact on twenty-first-century societies. Two or more generations of expansion have not only created mass graduate workforces servicing the needs of a knowledge economy; they have also generated a graduate culture, which may be the twenty-first-century equivalent of the bourgeois culture of the nineteenth and twentieth centuries. Today, it is impossible to be middle class, the dominant social class in most advanced societies, without some experience of higher education. However, this assertion must be qualified in two ways:

- First, there are powerful links between the narrower credentializing, or entrepreneurial, functions of higher education and its wider role in the development of a graduate culture. This relationship between business success, particularly in the newer creative industries, and the rich social and cultural milieus which are characteristic of cities with large student and graduate populations, provides an important link between expert and social forms of credentialization. It may also revive, and reinforce, the argument about the links between modernity and modernization, between economic development and a socially progressive culture.
- Second, the graduate culture of the twenty-first century is different from the bourgeois culture of the nineteenth and twentieth centuries. Access to higher education may continue to be socially inequitable (and, in particular, access to elite universities). But contemporary graduate culture is much more open – to women (which is why feminism and mass access to higher education are so closely aligned) and to ethnic and religious minorities. This culture is also more dependent on higher levels of formal education (and the cultural breadth that is associated with experience of higher education). It is not entirely implausible to link the solidity of democratic institutions in post-War Europe, and also the liberalism of most European societies, to the advance of higher education.

However, there may be other dimensions of post-modern societies which relate less well to higher education. It is true that traditional, and fixed, social identities

based on class and gender (and also national affiliations) have gradually been eroded – and even superseded – by newer, more open, and more fluid, social identities based on a combination of expertise (the credentializing functions of universities) and on enlightenment (their wider educational, and civilizing, potentialities). But other, even more novel, forms of identity are also emerging – for example, identities that are based on patterns of consumption and lifestyle choices, which are volatile and ambiguous; or on ethno-religious categories which may be equally contrived, post-modern constructions despite their fundamentalist labels. Higher education's links to these (quintessentially) post-modern identities are much less straightforward. Universities certainly play a key role in the socialization of those who are most able to experiment with new social identities by expressing distinctive lifestyle choices (as well as giving them the credentials which provide them with the economic resources to make such choices). It is also true that higher education systems play a key role in socializing ethnic and religious minorities, so they have some responsibility for (and leverage over) any anti-socialization tendencies among these communities. But in both cases, the influence of higher education is muted.

Conclusion

In this article, the development of modern higher education systems has been discussed mainly in terms of its broader relationship with society, the economy, politics, and culture. This could lead to the conclusion that the emergence of mass higher education is essentially a social phenomenon, the outcome of the shaping of higher education by these larger social and economic forces. Higher education, to a certain extent, itself is an active agent, because the expansion of the system has transformed it into a social force in its own right. But this would be a misleading conclusion in two different senses. First, higher education has always had a powerful social presence. In some respects, the elite universities of the past were even more influential social agents with a crucial role in the reproduction of political and administrative elites, which was more important than their responsibilities for the socialization of cultural elites or the training of scientists, scholars, and researchers until well into the twentieth century. The erosion of elites, and the parallel problematizing of technical and professional expertise, has complicated the links between mass higher education and wider society – which may now be more pervasive but are also more diffuse.

Second, it is important to emphasize that higher education is an academic system. It is shaped as much by the internal dynamics of academic disciplines as it is by external social and economic forces. There is an inside story as well as an outside story. Admittedly, teaching and

research, the two main expressions of academic disciplines, are shaped by social forces (but they also shape them); many disciplines, in particular, professional and applied disciplines, are themselves socially (and economically) as well as cognitively constituted; and in a knowledge society, researching and learning are no longer so clearly constrained within disciplinary boundaries (or even within the boundaries of specialized knowledge organizations) but instead, have become socially distributed activities. But, it remains true that it is impossible to write about higher education without taking into account what Martin Trow called its private life of academic enquiry.

There is no space in this article to attempt to sketch out the academic landscape. But a major force in shaping higher education over the past century has been the emergence of new disciplines – their ceaseless re-configuration as specialties have come and gone – and the decline and fall of old disciplines. The transformation of the humanities from aspects of civilized discourse into a taxonomy of expert subjects, the uneasy cohabitation of basic scientific enquiry with techno-science, the efforts to build (and then defend) a critical base in the social sciences – these were only the most colorful threads within a finely woven academic narrative. It is difficult to predict the major intellectual and scientific developments that will shape the continuing private life of higher education in the twenty-first century: more radical forms of inter-disciplinary enquiry in the humanities and social sciences perhaps; in the natural sciences, the explosive potential of the development of remarkable new experimental technologies; in some more applied fields, the attenuation of the distinctively academic sphere within a wider knowledge enterprise without clear organizational boundaries. But, if exact prediction is impossible, two ambitions are not only possible but essential. The first is to assert the continuing importance of intellectual values that are normatively as well as cognitively grounded within mass higher education systems; these values are more not less essential in mass systems than they were in the elite systems of the past (Fallis, 2007). The second, more fancifully, is to let the imagination roam – to conjure up these new academic landscapes. In considering the future of higher education, this is just as important as predicting new social forms – the knowledge economy, post-modern society, or the market state. Together, they will shape that future.

See also: Diversity of Higher Education; Higher Education Crossing Borders; Internationalization of Higher Education; The Bologna Process in European Higher Education; Transforming Higher Education in Developing Countries: The Role of the World Bank; Unitary, Binary and other National Systems of Higher Education; Universities and Regional Development; Viewing Private Higher Education: How Much, Where, Why, and What?.

Bibliography

Fallis, G. (2007). *Multiversities, Ideas, and Democracy*. Toronto, ON: University of Toronto Press.

Further Reading

Altbach, P. and Levy, D. (eds.) (2005). *Private Higher Education: A Global Revolution*. Rotterdam: Sense.

Barnett, R. (2000). *Realizing the University in an Age of Supercomplexity*. Buckingham: Society for Research into Higher Education/Open University Press.

Clark, B. (1983). *The Higher Education System: Academic Organisation in Cross-National Perspective*. Berkeley, CA: University of California Press.

Florida, R. L. (2005). *Cities and the Creative Class*. London: Routledge.

Froment, E. (ed.) (2006). *EUA Bologna Handbook: Making Bologna Work*. Berlin: Raabe Academic Publishers.

Gibbons, M., Camille, L., Helga, N., et al. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: Sage.

Nowotny, H., Scott, P., and Gibbons, M. (2001). *Re-thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge: Polity Press.

Parsons, T. and Gerald, M. P. (1973). *The American University*. Cambridge, MA: Harvard University Press.

Power, M. (1996). *The Audit Society: Rituals of Verification*. Oxford: Clarendon Press.

Schofer, E. and Meyer, W. J. (2005). *The World-wide Expansion of Higher Education*. Stanford, CA: Stanford Institute for International Studies: Center on Democracy, Development and The Rule of Law, Number 32.

Scott, P. (1995). *The Meanings of Mass Higher Education*. Buckingham: Open University Press.

Strathern, M. (2000). *Audit Cultures*. London: Routledge (European Association of Social Anthropologists series).

Teichler, U. (2004). *Changing Structures of Higher Education Systems: The Increasing Complexity of Underlying Forces. The Diversification of Higher Education and the Changing Roles of Knowledge and Research*. Paris: UNESCO.

Trow, M. (1973). *Problems in the Transition from Elite to Mass Higher Education*. Berkeley, CA: Carnegie Commission on Higher Education.

UNESCO (2004). *UNESCO Online Database*. Paris: United Nations Educational, Scientific and Cultural Organisation.

Weber, L. and Duderstadt, J. (2008). *The Globalization of Higher Education*. London: Economica.

HIGHER EDUCATION – MANAGEMENT, LEADERSHIP AND GOVERNANCE

Contents

Academic Managers – Heads of Departments/Schools
Leading and Managing the University – Presidents and their Senior Management Team
Management of and in Higher Education Institutions
Managerialism and Collegialism in Higher Education Institutions
Strategic Planning in Higher Education
The Changing Role of University Presidents, Vice-Chancellors and Rectors
The Changing Roles of University Governing Boards and Councils

Academic Managers – Heads of Departments/Schools

S Marshall, Leadership Foundation for Higher Education, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Development center – A structured series of activities aimed at bringing about greater self-awareness of leadership strengths and areas for improvement. Such activities include: 360° diagnostics, simulated interactions, and in-tray activity.

Myers Briggs type indicator (MBTI) – A personality instrument, based on Jungian theory, which suggests that individuals have innate psychological or mental dispositions.

Introduction

The position of department chair is one of leadership, charged with the challenges of developing the department's future and of building faculty vitality. As we move into the new millennium, we face a time of major change ... changing student clientele, disintegrating college curricula, growing technological changes, and shifting attitudes and practices of faculty represent some of the many forces currently shaping higher education ... Anticipating your changing environments, developing a future-oriented statement of department mission, and providing leadership to unify department activity toward a 'planned' future is the mark of effective academic leaders ... The time of 'amateur administration' where professors temporarily step into the administrative role

of department chair has lost its effectiveness. The call for leadership is real. Department chairs are critical to the future of the college, department, and faculty (Gmelch and Miskin, 1993: 1–2).

The above quote, written over a decade ago, stills holds true, and holds true for the range of academic management and leadership roles in higher education (HE). Focusing on the role of head of department, or, department chairs, or heads of school (these three terms will be used interchangeably, as representing a key strategic academic management post in HE, in the section which follows), is to focus on what is recognized as one of the most crucial roles in the university (Ramsden, 1998; Shattock, 2003). Without heads of departments who are effective leaders and managers, there is a greatly reduced chance of the university vision and strategy becoming a reality, as they are a key factor in terms of creating an appropriate environment in which the academic endeavor can thrive (Ramsden, 1998; Gibbs *et al.*, 2007). It is really only over the past one to two decades that the cruciality of this role has been properly recognized, and measures taken to first, more accurately articulate just what this role entails, and, second, offer support in terms of development and training in preparation for this role.

As Benoit and Graham (2005) note:

Most academic chairs starting out today will find their role much more complex than they would have just a decade ago. Chairs are involved in a wide variety of tasks, including defining the strategic missions of their units, curriculum planning, performance reviews, fiscal oversight, and advocating for and representing their departments

to the university at large and to external constituents. Academic chairs need skills different from those that originally attracted them to the independent life of the scholar. No one explains their new role to them, and most new chairs learn through personal experiences and by watching what other chairs do (p. 1).

Given the recognition that the head of department role is both crucial and complex, this section sets out to explore current thinking on, first, the range of factors that have led to the increasing complexity of the role; second, the contractual obligations of the role to include its key components and recruitment issues; third, expectations for delivery of the role and appropriate support; fourth, training and development for the role; and, finally, issues for the future.

The Growth of Complexity of the Role of Head of Department/School

In the UK, post the Jarrett Report (1985), the need for better management of higher education (HE) was highlighted. Greater accountability came about with the advent of the quality-assurance agency, the introduction of the research-assessment exercise, the funding councils, and the professional and statutory bodies. Department heads have been delegated more and more responsibilities and accountabilities with respect to various operations and functions (Shattock, 2003). Similar trends have taken place globally (Ramsden, 1998). Indeed, globally, even more performance measures have now to be taken into account (Cuthbert, 2002) for example, leagues tables (Shanghai and Jiao Tong), student surveys, and funders' more rigorous approach to performance assessment. Alongside a reduction in the unit of resource, overall financial health is assessed, in order to categorize higher education institutions (HEIs) according to perceived risk of failure. The business of leadership in HEIs is increasingly in the spotlight as institutions first, engage in global competition for limited resources – most particularly the student market; second, endeavor to respond to the rapidity of technological development and its impact on all aspects of a university's operations; third, address the so-called commodification of knowledge (Bargh *et al.*, 2000); and fourth, explore alternative means of funding further to the state reduction in the unit of resource. Heads of department need to be not only highly effective managers to respond to these drivers, but also entrepreneurial in their thinking in terms of taking their vision forward to ensure survival, sustainability, and global positioning of their departments. No longer is the role of head of department one of steward, or caretaker because to maintain the *status quo*, or stand still, is no longer an option. As Ramsden (1998) notes:

Higher Education is going through a revolution ... The future success of our universities depends on academics' capacity to respond energetically to change. To help academics face new and uncertain demands, we need an entirely different approach to their management and leadership" (Introduction).

So, alongside all the above developments has been an evolving realization that the role of the head of department requires both leadership and management capabilities – not something that many academics are assisted in developing over the course of their academic careers as noted by Benoit and Graham (2005) earlier. As Stevenson and Howlett (2007) state about the appointment of heads of school in the chartered universities of the UK:

Historically, ... leaders were selected 'papally' using soundings from school/faculty members and a nomination system where appointments were often made without interviews" (p. 114).

There now exists recognition that to take such a lax approach to the appointment of a key leadership and management role within an institution is to risk alienating and losing key staff who are crucial to the realization of universities' missions (Murray, 1999; Stevenson and Howlett, 2007). Heads of departments are increasingly required to navigate a complex and turbulent environment, in order to ensure their staff are provided with the best possible environment in which to deliver teaching and research (Gibbs *et al.*, 2007). Attempts to codify such a role are explored more in the section which follows.

The Contractual Obligations of the Role

Further to the realization among researchers globally that the role of head of department was becoming increasingly complex (Gmelch and MIskin, 1993; Shattock, 2003; Benoit and Graham, 2005), there has been a concurrent move to unpack and articulate what exactly the role entails. Once this was established, HEIs have been in a better position to design job descriptions that not only articulated the various requirements, responsibilities, and accountabilities, but provided a conceptual framework for those contemplating taking on such a role. Among early studies was Tucker's (1993) edition of *Chairing the Academic Department* which cataloged 54 different duties. Carroll and Gmelch (1994) undertook a study of US department chairs to focus on roles and duties, articulating ten priority duties (see **Table 1**).

Dyer and Miller (1999) suggest that the investigation of department chairs in the US has tended to focus on their "roles and responsibilities, needed skills, and challenges and coping strategies" (p. 20). Murray (1999), in her article, 'Department chairs call for leadership training,' refers to

Table 1 Ten priority duties of US department chairs

1. Recruit and select faculty
2. Represent the department to administration
3. Evaluate faculty performance
4. Evaluate faculty research and publication
5. Maintain conducive work climate
6. Manage departmental resources
7. Encourage professional development efforts of faculty
8. Develop and initiate long-range departmental goals
9. Provide informal faculty leadership
10. Remain current with academic discipline

From Carroll, J. and Gmelch, W. (1994). Department chairs' perceptions of the relative importance of their duties. *Journal for Higher Education Management* 10, 51.

many studies, to include that of Joseph Steinmetz, chair of psychology at Indiana University. Steinmetz's view is that "good scientists don't necessarily make good chairs... People with management experience, however, often do" (p. 3). He suggests that aspiring chairs should be provided with different departmental responsibilities to bring about awareness of and the development of managerial skills.

Knight and Trowler (2001), further to their extensive research into the role of departmental head, offer the following quotation from one of their US interviewees, illustrating the benefits of engaging in various roles outwith the department prior to taking on a head of department role:

Before becoming chair, I served in leadership positions in the faculty senate. I was viewed as a voice of reason in many situations and as someone who was not afraid to take an unpopular position on an issue, in other words, couldn't be bullied into submission. This has given me substantial credibility with the upper level administration. I also developed very strategic alliances with the business community. They are big supporters of me and the upper level administrators are reluctant to cross me. When they have, the business community has rallied to my position. These trump cards are used VERY sparingly, though. In fact, it was only needed once in the last four years. Skills that help: good reputation in previous on campus leadership spots... ability to set up key alliances with influential contributors... sense of humor that can be used to diffuse difficult situations" (p. 143).

Stevenson and Howlett (2007) offer a heads-of-school framework (see **Table 2**), which identifies and unpacks six behaviors required of the role:

1. working at the strategic level,
2. leadership,
3. credibility,
4. communicating with others,
5. working with people, and
6. embracing change (p. 121).

A host of studies now exist which have generated broadly similar lists (Bareham, 2004; Benoit and Graham,

2005; Aziz *et al.*, 2005; Creswell *et al.*, 1990). How not to put off potential heads of department with these long lists of responsibilities and behaviors required for the role is considered in the next subsection.

Expectations of Staff and Post Holder with Respect to the Role

A number of studies have analyzed the various aspects of the head of department role which today underpin many in-house leadership development programs. However, fewer studies have examined the transition required to make an effective head of department and what will be required in terms of change of self-concept and ways of working. Deem (2007) and Henkel (2000, 2002) have explored the notion of academic identities. Deem (2007) refers to career-track routes, reluctant managers, and good-citizen routes to academic management. Henkel (2000) is more concerned with the reprofessionalization of academics in the midst of the growing complexity of the environment which HE inhabits, which she sees as requiring a re-articulation of core values. Gmelch and Miskin (1993) go beneath the surface of these concepts, unpacking just what is entailed in terms of the transition into an academic management position.

Gmelch and Miskin's (1993) nine transitions that individuals face when moving from a faculty position to that of a department chair still hold good today (see **Table 3**), articulating the transition as being from:

- solitary to social,
- focused to fragmented,
- autonomy to accountability,
- manuscripts to memoranda,
- private to public,
- professing to persuading,
- stability to mobility,
- client to custodian, and
- austerity to prosperity (pp. 11 – 12).

An issue identified in a number of studies focuses on an expectations disconnect between what an incumbent head of department expects of the role, and what they actually experience – often leading to either great stress or an inability to make the professional transition required of the post. This notion is explored in Achtenberg's (2004) article, Look before you leap: Transitions from faculty to administration, which offers itself as a useful summary of how becoming a chair will change your life, with the suggestion that it should be read at the stage when a career move to head of department is being contemplated.

Once in the role, a number of studies suggest that the anxiety and stress can be all consuming. Murray (1999), calling for universities to bolster their guidance of chairs by offering leadership training, quotes a chair of a psychology

Table 2 University of Newcastle: Heads-of-school framework

<i>Behavior</i>	<i>Indicators</i>
Working at strategic level	Sees the big picture and thinks strategically. Is able to manage complex problems and issues. <ul style="list-style-type: none"> • Able to articulate strategic vision for organization • Identifies opportunities for enhancing reputation, kudos, and financial security of organization • Identifies appropriate structures for making strategic vision a reality • Comfortable working with complexities • Can apply numerical problem-solving skills
Leadership	Takes control to ensure that objectives are met. <ul style="list-style-type: none"> • Provides clear direction to others • Delegates tasks to others without relinquishing control and responsibility • Is comfortable with conflict and prepared to confront individuals who stand in the way of organizational achievements • Acts decisively
Credibility	Inspires the trust and confidence of others by the way they handle themselves and others. <ul style="list-style-type: none"> • Energizes others • Possesses insight; particularly into the needs and motivations of others • Self-confident in manner and able to take unpopular decisions • Is straightforward, open, and fair in dealings with others • Motivated to set individual interests aside and work hard for future success of organization
Communicating with others	Communicates with colleagues effectively. Is able to persuade and influence others. <ul style="list-style-type: none"> • Listens to the views and opinions of others • Disseminates key messages to others • Is able to influence by persuasive and articulate communications • Responds to communications from others in clear, appropriate, and timely fashion
Working with people	Exceptional interpersonal skills. <ul style="list-style-type: none"> • Able to relate well to all types of people • Able to alter style depending on people or situation • Behaves with diplomacy • Interested in others and motivated by helping them to feel valued and to succeed • Motivates others so that they wish to contribute
Embraces change	Recognizes the need for change and is forward looking. Promotes the benefits of change to others. <ul style="list-style-type: none"> • Is proactive in ideas generation • Accepts the reality and requirements of change • Uses initiative and presents new ideas and approaches • Is able to achieve change at individual and organizational level • Is supportive of others during change
Planning and organizing	Manages time and resources by prioritizing and organizing effectively. <ul style="list-style-type: none"> • Effective at organizing and managing heavy workload • Anticipates future demands for staff, information, etc. • Identifies quickly what is important and prioritizes and acts accordingly
Management of pressure	Copes with criticism and the demands of the job. <ul style="list-style-type: none"> • Is able to separate him/herself from the job in hand so that negative feedback does not become overwhelming • Able to bounce back in the face of setbacks • Driven to succeed even when problems seem intractable • Remains positive and optimistic when going gets tough

From Stevenson, T. and Howlett, L. (2007). The leadership succession challenge for higher education: A pilot of leadership development centres at Newcastle University. In Marshall, S. (ed.) *Strategic Leadership of Change in HE: What's New?* pp 114–128. London: Routledge Taylor Francis.

department at a university in Indiana: “The worry was pervasive At lunch, in the shower, at home, with my family. It kept me up at night” (p. 1) but he does go on to say that it would not have been so hard if the university had introduced him to what to expect: “I wasn’t prepared for the stress . . . Training would have helped” (p. 1).

Studies of the characteristics of effective chairs are numerous. There are a good number that examine characteristics of a specific discipline areas, for example, chairs of business schools (Bolton, 1996; Bareham, 2004; Bellamy

et al., 2003; Clott and Fjortoft, 2000); others examine characteristics of chairs within a certain institution (Aziz *et al.*, 2005; Meredith and Wunsch, 1991; Lindholm, 2003), and there are those that generate more generic typologies as highlighted throughout this section.

Creswell *et al.* (1990) undertook a qualitative survey of 200 US department chairs who had been identified as excelling in the role as chair. They identified 15 factors as contributing to this excellence as articulated in **Table 4.**

Table 3 Transitions to leadership

1. *From solitary to social.* College professors typically work alone on research, teaching preparation, and projects. Now, as chair, your responsibility forces you to work with and through others. For example, department goals cannot be achieved alone; they must be achieved in concert with your faculty.
2. *From focused to fragmented.* While professors must have long, uninterrupted periods to work on scholarly pursuits, your work as department chair, like other management positions, is characterized by brevity, variety, and fragmentation.
3. *From autonomy to accountability.* Professors generally enjoy control over their time and the feeling of autonomy over activity and movement in their working environment. As you move from your role of professor to administrator, you tend to lose this sense of autonomy and become accountable to upper administration and the faculty for your time and accessibility in the office, as well as for your actions and activities.
4. *From manuscripts to memoranda.* The scholar and researcher labors over a manuscript for a long period of time. Before finding printer's ink, the work goes through many revisions and critiques. As department chair, you quickly must learn the art of persuasion and precision through memos. Thus, chairs report less stress from manuscripts and more from completing paperwork on time.
5. *From private to public.* The professor may block out long periods of time for scholarly work, but as chair you have an obligation to be accessible throughout the day to the many publics you serve. In essence, you move from the privilege of a closed door to the obligation of an open-door policy.
6. *From professing to persuading.* In the academic profession, the professor disseminates information in a manner that will meet the learning objectives of others. As you turn from professor into chair, you profess less and practice more the art of persuasion and compromise.
7. *From stability to mobility.* While always growing and exploring new concepts and ideas, faculty generally experience movement within the stability of their discipline and circle of professional associations. As a chair, you also attempt to retain your professional identity but must become mobile within the university structure. In order to be at the cutting edge of educational reform and to implement needed programmatic changes within, you must be more mobile, visible, and political.
8. *From client to custodian.* In relation to university resources, the professor is a client, requesting and expecting resources to be available to conduct research, classes, and service activities. As chair, you represent the custodian and dispenser of resources and are responsible for the maintenance of the physical setting as well as providing material and monetary resources.
9. *From austerity to prosperity.* While in actuality the pay differential between professor and chair may not be significant, the perception of more control over departmental resources creates the illusion of greater prosperity as chair.

From Gmelch, W. and Miskin, V. (1993). Understanding the challenges of department chairs. In *Leadership Skills for Department Chairs*, pp 11–12. Bolton, MA: Anker.

Table 4 Fifteen factors contributing to chair excellence, according to Creswell *et al.* (1990)

1. Learn about their roles and responsibilities in both the department and institution;
2. Create a balance between professional and personal lives;
3. Prepare for their professional futures;
4. Establish a collective department vision or focus;
5. Develop faculty ownership of the vision;
6. Initiate changes slowly (as a caution against making changes too soon after taking over);
7. Allocate resources of time, information, and assignment to encourage the vision;
8. Monitor progress in realizing the vision;
9. Create a climate open to the flow of ideas to build trust;
10. Listen to the needs and interests of academic staff;
11. Helps academic staff set professional goals;
12. Ensure academic staff receive feedback;
13. Provide representation of academic staff to colleagues and senior administrators;
14. Serve as a role model and mentor;
15. Encourage and support faculty.

From Creswell, J., Wheeler, D., Seagren, A., Eg, N., and Beyer, K. (1990). *The Academic Chairperson's Handbook* Lincoln, NE: University of Nebraska Press.

Benoit and Graham's (2005) study of successful chairs revealed that there were three types of experiences which were felt to be both rewarding and satisfying: obtaining resources for the department, implementing their vision, and creating a positive environment. Lorange's (1988)

study identifies strategic vision as crucial to departments, depicting it as a "rallying point in its strive (sic) to create academic value" (p. 301).

Hecht *et al.* (1999) identified a number of attributes which made for head of department effectiveness, to include being seen as a person with integrity, capable of politicking on behalf of the department, and ensuring adequate marketing of the achievements of the department. Similarly, trust and integrity issues were found to be important in a US study conducted by Murray and Stauffacher (2001). In Moses and Roe's (1990) study of Australian heads of department, however, trust and integrity were less prominent.

Harris *et al.* (2004) conducted a study of leadership effectiveness of heads of department across five US research universities. Their conclusions included a consideration of what made for effective leadership behavior from differing perspectives and identified five factors, not differing significantly from other typologies other than articulating the requirement for an ethical approach (see **Table 5**).

Additionally, they noted that praising staff for good performance, being seen as someone who keeps promises, and acting as a role model for others were deemed to be important.

Stevenson and Howlett (2007) undertook a series of interviews with key senior staff at the University of Newcastle in the UK, which generated eleven specific leadership needs (see **Table 6**). Their findings resonate well with factors raised in other studies, but a distinct need

Table 5 Effective leadership behavior

1. An ethical approach whereby people are treated with respect;
2. Effective communication;
3. Developing a shared vision – particularly in connection with the development of teaching programs and of curricula – in conjunction with others in the department;
4. Networking with the central administration and other departments on behalf of the department;
5. Empowerment of academic staff.

From Harris, J., Martin, B., and Agnew, W. (2004). The characteristics, behaviours and training of effective educational/leadership chairs. In Thompson, D. and Crampton, F. (eds). *The Changing Face(s) of Educational Leadership: UCEA at the Crossroads*. Kansas City, MO.

Table 6 Specific leadership needs

1. Financial/budget awareness
2. Planning and organizing to achieve goals
3. Handling conflict
4. Delegating and letting go
5. Managing and prioritizing an increasing workload
6. Thinking more broadly about faculty/university/sector issues
7. Strategic visioning skills
8. Assertiveness with challenging colleagues
9. Leading others through change
10. Decision making to make progress
11. Staying focused and positive in the face of challenge and pressure (p. 124).

identified in their study was “staying focussed and positive in the face of challenge and pressure” (p. 124).

There is evidence to suggest that this wealth of research as explored in this section has helped to inform the sort of training and development now offered for heads of department (Hecht *et al.*, 1999; Stevenson and Howlett, 2007; Murray, 1999; Rhodes, 2007), despite Bargh *et al.*'s suggestion that:

... an adequate theory of management and/or leadership in academic has yet to be developed, and needs to be firmly grounded in the specific features of the HE sector” (Bargh *et al.*, 2000: 162).

Training and Development for the Role

In response to the concern about preparing academics in a better manner for a possible future role as head of department, the notion of development centers has been introduced in the UK – a notion imported from other areas of the public sector most notably the National Health Service and State Schools, and blue-chip companies such as the Royal Bank of Scotland and Bayer. The aim of development centers is to help both organizations and individuals alike determine (further to the deployment of assessment-center technology), diagnose, and formulate an appropriate personal-development plan

for prospective heads. The purpose of this plan is to assist the individual move toward the realization of his/her leadership potential. Stevenson and Howlett's (2007) account of the setting up of a development center at the University of Newcastle in the UK was further to the university's concern to develop a pool of potential heads of school. Such an initiative had the added benefit of assisting certain ambitious individuals realize, prior to putting themselves forward for the heads-of-school role, that such a role was not actually one that they were either comfortable with or capable of undertaking. As Stevenson and Howlett (2007) note, “investing early can prevent expensive mistakes in the Heads of School appointment process” (p. 125).

Many research studies explore the range of support for heads of department, suggesting various tactics that they have identified as working, in terms of better preparing and supporting heads of department. For example, Murray (1999) considered a range of training options, which included co-opting potential leaders onto various department and university committees, filling administrative roles, or directing various activities. This technique (referred to by Steinmetz in Murray's 1999 article quoted at the beginning of this section) of what constitutes an induction is used quite extensively for prospective heads of department in UK HEIs. However, for individuals to gain maximum benefit from such opportunities, it is suggested that training and development be run in parallel (Rhodes, 2007). An exploration of what research studies suggest effective training and development might look like as what follows.

Little empirical research has focused on the process of determining the training and development needs of heads of department. Two such recent studies, however, provide a useful starting point. The first, undertaken by Aziz *et al.* (2005) in the USA, revealed four key primarily cognitive areas: (1) budgets and funding, (2) academic staff issues, (3) legal issues, and, finally, (4) professional development. A UK study conducted by Ramsden (2006) sought to explore the benefits sought from a heads of department program, identifying six primarily affective areas: (1) self-knowledge, (2) environment knowledge, (3) style/reflection, (4) learning new techniques, (5) sharing experiences, and (6) gaining confidence.

Wolverton *et al.* (2005), identifying a similar set of complex needs for chairs and prospective chairs at the University of Nevada, subsequently designed a developmental program of seven 3-h sessions over 1 year, based on:

- conceptual understanding (tools and techniques to improve performance);
- skill development (communication-related issues); and
- reflective practice (journal based on Myers Briggs Type Indicator diagnostics and small group discussions).

Increasingly, leadership development for academic managers is seen as a cognitive process, moving into the

affective domain, as opposed to solely a training process. As Weick (2001) notes, leadership development is partly about sense making which comes through a journey of self-discovery, understanding, and reflecting on relationships with others, and an awareness of the environment.

Few studies reveal development programs which attempt to develop both the competences as well as the competencies for the role. McLeod and Jennings (1990) have documented a course they ran for International Business Machines (IBM) (UK) which addressed the latter, suggesting effective development programs including:

Gestalt work, personal construct theory, guided affective imagery, Chinese philosophy, psychosynthesis, and using the right –hand side of the brain” (p. 69).

The incorporation of a number of these approaches is now used in the programs offered by the Leadership Foundation for Higher Education in the UK, with the program for heads of department (or those aspiring to such a role) including:

- knowledge, understanding, and skill development;
- case-study analysis;
- self-awareness diagnostics;
- personal construct theory;
- mythodrama; and
- reflective practice.

Rhodes (2007), writing about developing leaders at the University of Chester in the UK, refers to the view that the most effective learning for managers takes place in the workplace. He developed a three-phase in-house program (see **Table 7**) at the University of Chester which moved from the concrete to the abstract conceptualization of the requirements of the role. The third phase:

developed and adapted to take account of the rapidly growing trend towards more personal forms of learning support: mentoring, coaching, 360 degree feedback, secondments, and directed learning relevant to the working context” (p. 69) (**Table 8**).

Despite all the support that can be offered to heads of department in the form of training and development, the

stress of the job remains a major concern (Sarros *et al.*, 1997: 284). Juggling the constant stream of administrative work with that of leading, teaching, and research and scholarship, flags up a growing problem with prioritization and time management. Despite the growth in complexity, and the constant juggling act to ensure attendance to the range of areas requiring an input (e.g., strategic visioning and planning, teaching, curriculum development, research leadership, monitoring resources, and supporting staff), a significant number of committed academics view the role of head of department as offering an opportunity to make a difference. So, what further research remains to be done to better prepare and support heads of department of the future?

Issues for the Future

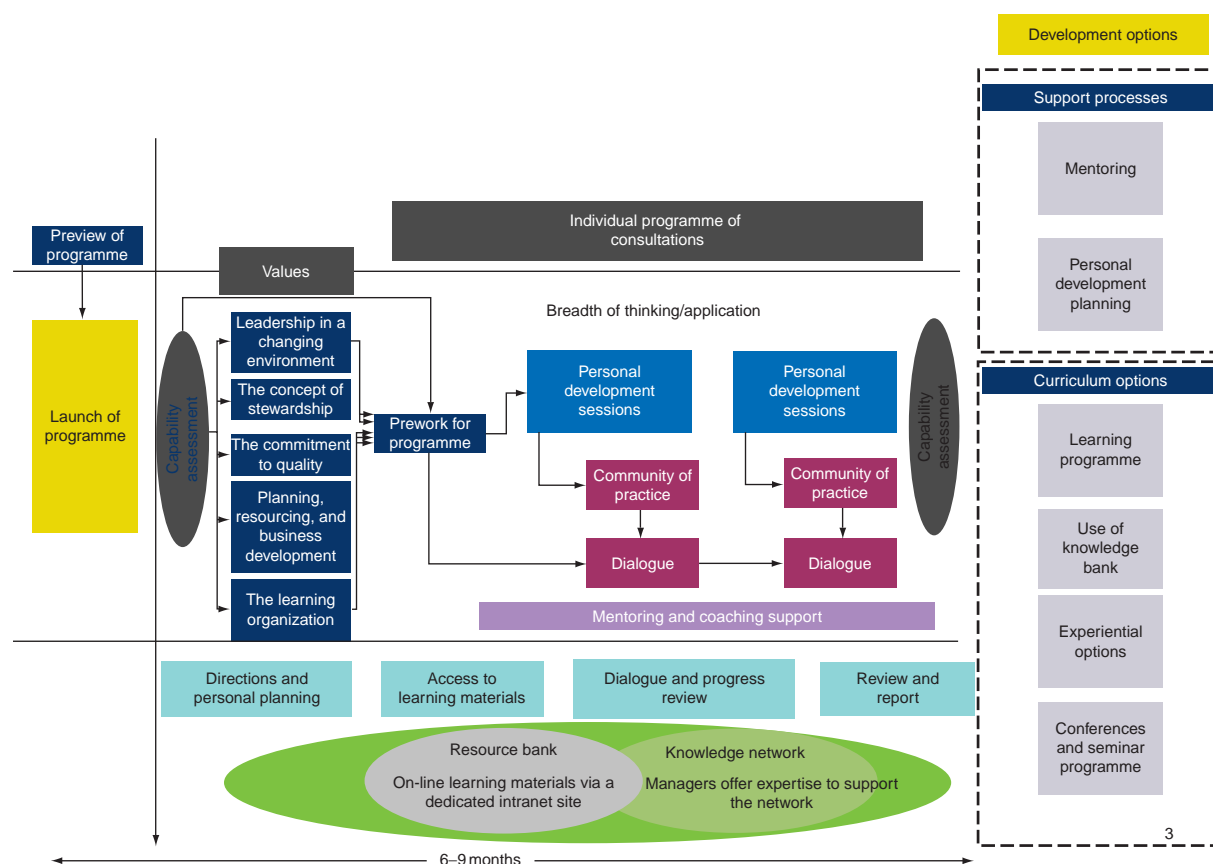
Increasingly, it is recognized that higher education should be more outward looking in terms of seeking and sharing best practices, to include scanning initiatives globally and cross-sectorally. As the world of higher education becomes more competitive, with recognition that knowledge workers are HEI's most important asset, modernizing issues such as those below need to be addressed.

There appear to be three key issues facing HEIs looking to prepare and support heads of department in a better manner. First, despite studies such as Stevenson and Howlett's (2007), there are few studies of effective interventions to help address the issue of succession planning in terms of encouraging academic staff to view this key leadership and management role as a credible career pathway. Exploring talent management and succession planning in other sectors may provide some clues as to what might be trialed and hopefully formally critiqued via publication. The second area that still appears to cause concern in terms of heads of department is their recruitment and the appointment process. Having a pool of credible and well-motivated potential heads of department would be the ideal; however, the harsh reality in the UK is that often soundings have to be taken as to who might be prepared to put themselves forward for the role and, hopefully, not only be a good role model for the rest of the department, but be sufficiently strategic, entrepreneurial, and business minded to be able to take the department forward. As a result, the appointment procedure, most particularly in the chartered universities of the UK, does not always appear as transparent as it should, nor as open as it could be, if an advertisement were placed with a diverse pool of applicants, able to be drawn from, to come forward for formal interview. Finally, despite a host of diversity and equality initiatives, the lack of females and ethnic minorities in these roles remains an issue – certainly in the UK. More studies of initiatives in these three areas would undoubtedly provide some insights as

Table 7 The University of Chester management development program

Phase 1	Generic management development: e.g., strategic management, organizational effectiveness
Phase 2	Action learning sets: to promote the alignment of individuals in support of corporate objectives
Phase 3	Role modeling effective leadership: an emphasis on leadership behaviors

From Rhodes, M. (2007). Developing leaders: A structured approach to the enhancement of organisational and individual performance. In Marshall, S. (ed.) *Strategic Leadership of Change in HE: What's New?* pp 66–78. London: Routledge/Taylor & Francis.

Table 8 Leadership Development Programme

From Rhodes, M. (2007). Developing leaders: A structured approach to organisational and individual performance. In Marshall, S. (ed.) *Strategic Leadership of Change in HE: What's New?* pp 72. London: Routledge/Taylor & Francis.

to how the sector might be assisted in moving further forward than it has done since the last volume of this publication.

See also: Leading and Managing the University – Presidents and their Senior Management Team; Management of and in Higher Education Institutions; Managerialism and Collegialism in Higher Education Institutions.

Bibliography

- Achtenberg, C. (2004). Look before you leap: Transitions from faculty to administration. *Academic Leader* 20(8), 1, 7.
- Aziz, S., Mullins, M., Balzer, W., et al. (2005). Understanding the training needs of department chairs. *Studies in Higher Education* 30(5), 571–593.
- Bareham, J. (2004). *The Leadership and Management of Business Schools*. Brighton: Brighton Business School.
- Bargh, C., Bocock, J., Scott, P., and Smith, D. (2000). *University Leadership: The Role of the Chief Executive*. Buckingham: SRHE and Open University Press.
- Bellamy, A., Morley, C., and Watty, K. (2003). Why business academics remain in Australian universities despite deteriorating working conditions and reduced job satisfaction: An intellectual puzzle. *Journal of Higher Education Policy and Management* 25, 13–28.
- Benoit, P. and Graham, S. (2005). *The Secrets of Successful Chairs*, ACE Departmental Online Resource Centre, www.acenet.edu//AM (accessed August 2009).
- Bolton, A. (1996). The leadership challenge in universities: The case of business schools. *Higher Education* 31, 491–506.
- Carroll, J. and Gmelch, W. (1994). Department chairs' perceptions of the relative importance of their duties. *Journal for Higher Education Management* 10, 49–63.
- Clott, C. and Fjortoft, N. (2000). Culture and strategy in business schools: Links to organizational effectiveness. *Mid-Western Educational Researcher* 13, 27–33.
- Cresswell, J., Wheeler, D., Seagren, A., Eg, N., and Beyer, K. (1990). *The Academic Chairperson's Handbook*. Lincoln, NE: University of Nebraska Press.
- Cuthbert, R. (2002). The impact of national developments on institutional practice. In Ketteridge, S., Marshall, S., and Fry, S. (eds.) *The Effective Academic*, pp 32–48. London: Kogan Page.
- Deem, R. (2007). Managing contemporary UK universities – manager academics and new managerialism. *Academic Leadership: The Online Journal* 1(3). www.academicleadership.org/emperical_research/aManaging_contemporary_UK_universities_Manager-academics_and_new_Managementism.shtml.
- Dyer, B. and Miller, M. (1999). A critical review of literature related to the departmental chair position. ERIC. <http://coe.ksu.edu/ecea/2004/04uceall.pdf>.
- Gmelch, W. and Miskin, V. (1993). Understanding the challenges of department chairs. In *Leadership Skills for Department Chairs*, pp 11–38. Bolton, MA: Anker.

- Gibbs, G., Knapper, C., and Picinnin, S. (2007). *Departmental Leadership for Quality Teaching – An International Comparative Study of Effective Practice*. LFHE commissioned research, www.lfhe.ac.uk (accessed August 2009).
- Harris, J., Martin, B., and Agnew, W. (2004). The characteristics, behaviours and training of effective educational/leadership chairs. In Thompson, D. and Crampton, F. (eds.) *The Changing Face(s) of Educational Leadership: UCEA at the Crossroads*. Kansas City, MO.
- Hecht, I., Higgerson, M., Gmelch, W., and Tucker, A. (1999). *The Department Chair as Academic Leader*. Phoenix, AZ: American Council on Education/Oryx Press.
- Henkel, M. (2000). *Academic Identities and Policy Change in Higher Education*. London: Jessica Kingsley.
- Henkel, M. (2002). Emerging concepts of academic leadership for intra-institutional roles and relationships in HE. *European Journal of Education* **37**, 29–41.
- Jarrett Report (1985). *Report on the Steering Committee for Efficiency Studies in Universities*. London: CVCP.
- Knight, P. and Trowler, P. (2001). *Departmental Leadership in Higher Education*. Buckingham: SRHE and Open University Press.
- Lindholm, J. (2003). Perceived organizational fit: Nurturing the minds, hearts and personal ambitions of university faculty. *Review of Higher Education* **27**, 125–149.
- Lorange, P. (1988). Stimulating strategic direction setting in professional groups: The case of an academic department. *Advances in Strategic Management* **5**, 299–230.
- McLeod, S. and Jennings, J. (1990). Fit for the future. In Pedler, M., Burgoyne, J., Boydel, T., and Welshman, G. (eds.) *Staff Development in Organisations*, pp 67–83. London: McGraw Hill.
- Meredith, G. and Wunsch, M. (1991). The leadership of department chairpersons: Time/resource management, rewards, frustrations and role satisfaction. *Psychological Reports* **68**, 451–454.
- Moses, I. and Roe, E. (1990). *Heads and Chairs: Managing Academic Departments*. St Lucia, QLD: University of Queensland Press.
- Murray, B. (1999). Departmental chairs call for leadership training. *APA Monitor* **30**, 8.
- Murray, J. and Stauffacher, K. (2001). Department chair effectiveness: What skills and behaviours do deans, chairs and faculty in research universities perceive as important? *Arkansas Educational Research and Policy Studies Journal* **1**, 62–75.
- Ramsden, M. J. (2006). *Leadership Development for Heads of Department in UK Higher Education Institutions*. Unpublished MBA thesis, London: Institute of Education, University of London.
- Ramsden, P. (1998). *Learning to Lead in HE*. London: Routledge.
- Rhodes, M. (2007). Developing leaders: A structured approach to the enhancement of organisational and individual performance. In Marshall, S. (ed.) *Strategic Leadership of Change in HE: What's New?* pp 66–78. London: Routledge/Taylor & Francis.
- Sarros, J. C., Gmelch, W. H., and Tanewski, B. G. A (1997). The role of the department head in Australian universities. *Higher Education Research and Development* **16**(3), 283–292.
- Shattock, M. (2003). *Managing Successful Universities*. Berkshire: SRHE and OUP.
- Stevenson, T. and Howlett, L. (2007). The leadership succession challenge for higher education: A pilot of leadership development centres at Newcastle University. In Marshall, S. (ed.) *Strategic Leadership of Change in HE: What's New?* pp 114–128. London: Routledge Taylor Francis.
- Tucker, A. (1993). *Chairing the Academic Department*. Phoenix, AZ: American Council on Education/Oryx Press.
- Weick, K. (2001). *Making Sense of the Organization*. Oxford: Blackwell.
- Wolverton, M., Ackerman, R., and Holt, S. (2005). Preparing for leadership: What academic department chairs need to know. *Journal of HE Policy and Management* **27**(2), 227–238.

Further Reading

- Marshall, S. (ed.) (2007). *Strategic Leadership of Change in HE: What's New?* London: Routledge/Taylor & Francis.
- Thompson, D. and Crampton, F. (eds.) (2004). The changing face(s) of educational leadership: UCEA at the crossroads. Kansas City, MO.

Relevant Websites

- www.academicleadership.org – *Academic Leadership*, the online journal.
- www.acenet.com – American Council on Education.
- www.herdsa.org.au – Higher Education Research and Development Society of Australasia.
- www.lfhe.ac.uk – The Leadership Foundation for Higher Education.

Leading and Managing the University – Presidents and their Senior Management Team

R Middlehurst, Leadership Foundation for Higher Education, London, UK

T Kennie, Ranmore Consulting, Ranmore, UK

S Woodfield, University of Surrey, Guildford, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction: Framing the Focus of Enquiry

Organization, Roles and Context

In some countries and some centuries the idea of leading and managing a university when linked to the responsibilities of a president and his or her senior management team might be regarded as something of an oxymoron. There are several reasons for this. The first concerns conceptions of the university as an organization. Traditional conceptions of the university as a community of scholars, collegial institution, or organized anarchy suggest different expectations of leadership and the role of leaders from those associated with concepts of the university as a business or entrepreneurial venture. The second reason is associated with the nature of a president's role. Election to this position as first among equals suggests a different view of how the leadership role is conceived and executed from a situation in which the role is filled through internal or external selection and appointment to the top rank in a hierarchy. A third reason is linked to the term, management, and the role of the senior management team in leading and managing the university. Use of this terminology locates our topic broadly in the latter half of the twentieth and early twenty-first century and positions it in relation to the political, economic, and social contexts of that era. This period, then, defines the parameters of our enquiry.

Diversity of Perspectives

The variables identified above, namely, the university as an organization, the locus and source of authority for university leadership, and the context and time in which leadership occurs are important in framing how leadership is conceived, organized, and delivered. These variables are all reflected in the analytical framework adopted by Bargh and colleagues in their empirical and theoretical study of University Leadership in the UK and in the 5-year study of Institutional Leadership Project (ILP) conducted by Birnbaum and colleagues in the US. Another important variable is the national focus of studies – as Bargh and colleagues recognize and explore – there remain organizational and cultural differences in the

role of rector of a continental European university, president of a university in the USA, and vice chancellors in UK. The interplay between processes of leadership and management within institutions and the wider national and political context within which universities operate also features in many studies from different parts of the world. These studies provide an opportunity to examine the changing practices, as well as the structures and balances of power between the state and the institution.

There is diversity, as well, between universities of different types, historical origin, and legal and financial status within countries. Such factors also impinge on expectations and analyses of the task of leading and managing universities. To quote Madeline Green, a US scholar of institutional leadership, “With more than three thousand institutions. . . Diversity is the hallmark of the American system. Just as there is no recipe for institutional effectiveness, there is no single model of leadership. And definitions of effectiveness and of leadership requirements inevitably change over time.”

Leading and Managing the University

The literature associated with leading and managing the university can be separated, in part, from the literature associated with the role of president and the nature and responsibilities of a senior management team. The unit of enquiry in the first of these domains is the institution and, in the second, the focus is on individual actors. However, there are also studies that bridge the gap by examining actual or potential links between an institution and its performance and the profiles of leaders.

Within the domain of leading and managing the university, terminology again comes into play. In some studies, leadership is a dominant concept embracing management; in other cases, managing is the primary focus, but embracing leadership. In continental European studies, the broader term, governance, is widely used to include the role of rectors and boards within university decision-making processes. Some studies also specifically disaggregate concepts of leadership and governance or leadership and management as well as the associated roles and responsibilities of each domain. In other cases, the discussion is of shared responsibilities for leadership

between governors or trustees and presidents. Such different approaches facilitate an examination of the dynamics of relationships between actors and the critical interdependencies of leadership, governance, management and administration, and indeed, politics, in the running of an institution.

An important topic in the corporate literature is the link between leadership and organizational performance. A recent literature review in the UK has examined research on leadership and institutional effectiveness in higher education and has found a dearth of rigorous empirical work. The most important studies identified are included below.

University Presidents

The nature of the presidential role and the factors that impinge both on expectations surrounding the role and the way the role is exercised clearly offer a rich vein for research. UK and Australian studies, for example, focus on the individuals occupying or seeking presidential roles and examine their profiles, career paths, and remuneration. These studies shed light on sustained or divergent trends over time, the social and disciplinary profiles of incumbents, the characteristics and capabilities sought for the role, rewards, and recruitment and selection processes. A number of hypotheses are posited or tested in these studies including the relationship between disciplinary and professional backgrounds and social and academic status, as well as the power and influence of presidents within the university or with external constituencies such as government, funding bodies, and sponsors. Other studies in this territory explore barriers to entry, such as gender or ethnicity, and examine strategies for development and recruitment that seek to eliminate or reduce barriers. The development routes and learning opportunities offered to or taken by individuals provide another field for research linked to entry to the role of president or member of the senior management team. The lens is also turned on leadership style and approach in studies that concentrate on how the role of leader and management team is exercised in practice.

Senior Management Teams

The amount of published work on senior (or top) management teams in higher education is limited. In the USA, Bensimon and Neumann provide the most comprehensive analysis based on several focused dialogs with 15 presidents of US higher education institutions. Gioia and Thomas undertook an extensive study of how senior management teams make sense of important issues that affect strategic change in modern universities. More recent studies have emerged in the UK: Jarzabkowski and Wilson provide insights into how one senior team formulated and implemented strategy at the University of

Warwick, tracking the process of strategy formulation through the interplay of various structures within the institution. In a later article, Jarzabkowski extends the analysis to three UK institutions: Warwick, the London School of Economics (LSE), and Oxford Brookes University, exploring some of the micro-practices of strategy. Further work in the UK by Taylor explores the relationship between the overall organizational structure and institutional leadership through a case-study analysis of four UK institutions (all research-led premier league institutions) that had undergone major institutional level restructuring in recent years. Most recently, another UK-based empirical study is investigating the challenges and changes associated with the composition of top management team structures in higher education institutions. This research identifies changes in roles and responsibilities and examples of good practice, or good processes, within institutions in the public and private sectors of higher education.

Types of Literature

The types of available literature in relation to our topic are wide ranging, which include field research of various kinds, scholarly reflections and analyses, personal accounts, policy reports, and media articles. Both quantitative and qualitative approaches are represented in the research studies and research is undertaken in a variety of disciplinary fields as part of the general social science and management literature and the specific literature relating to studies in higher education. On the positive side, there is a rich array of knowledge and insights available. On the negative side, however, there is an absence of recognized and cumulative approach to knowledge generation and many studies lack rigor either in terms of theoretical underpinnings or methodological soundness. There is also an imbalance between normative studies and reflective accounts built on practice and experience, and theoretical and investigative research, with fewer studies in the second category.

Focus of the Enquiry

In the next section, we focus more deeply on two main areas: the roles and responsibilities of presidents in leading and managing the university and senior management teams in order to identify findings, concepts, and issues arising before concluding with suggestions concerning future avenues for research. A few key texts are sampled. These have been chosen for their empirical base, practical insights, and their connections to the wider leadership literature so as to give an indication of the types of study available and their value for different purposes and audiences.

Leading Universities: The Role and Responsibilities of Presidents

Two empirical studies of presidential leadership undertaken in the last 40 years stand out because of the breadth of methods used, the range of data collected, the theoretical perspectives presented, and the new knowledge and insights achieved from the research. The first was undertaken in the US over a 5-year period from 1988 to 1992 (the ILP); the second was undertaken in the UK over a 2-year period. An earlier and smaller-scale empirical study in the UK by Bargh and colleagues has also influenced both the practice of leadership and leadership development in the UK and Europe. In the US, the work of Green and Green as well as of McDade has had similar impact.

The US ILP is important for several reasons. First, the authors examined theories of leadership from outside higher education and tested them in the context of universities. Second, they adopted an approach to studying leadership that was closely linked to universities as organizations. By combining the two perspectives, the researchers were able to offer an integrated conceptual approach to strategy formulation and execution that made sense of the complexity of an institution in its particular context and the stage of development. Building on initial work on the cybernetics of academic organization, the project concluded with Birnbaum's analysis of success and failure in the college presidency.

This project is one of the few higher education research studies that attempt to tackle the difficult question of how leadership effectiveness can be depicted and evaluated. Birnbaum and his colleagues drew on the three criteria defined by Cameron, namely, goal achievement, resource acquisition, and smooth internal functioning, to arrive at a definition of good leadership in academic institutions based around constituent support. The studies conducted by Birnbaum and his colleagues are collectively valuable in emphasizing the importance of context for the exercise of leadership and in placing presidential leadership within a wider framework of institutional players, culture, and history that places limits on the influence of a single individual. The individual studies undertaken within the project are equally important. They focus on different stages of the presidential role, sampling new and experienced presidents. The social and cognitive processes that shape the symbolic and substantive aspects of leadership and how presidents learned their job and assess their effectiveness are further themes. The researchers also explored leadership and follower-ship as well as leadership, culture, and change. Each of these avenues of enquiry is represented in the wider literature on leadership, beyond higher education.

The UK study by Bargh and colleagues is significant for different reasons. First, the research achieves greater depth and more comprehensive coverage than many other

studies by combining several types of data. The biographical data are longitudinal, examining the changing career experiences and backgrounds of vice chancellors for the period 1960–97. A total of 341 cases were collected in two linked databases. The study also includes comparative international data from four countries: the UK, US, Sweden, and the Netherlands. These data comprised a literature review examining executive leadership in the US and Europe and three questionnaire surveys of university presidents, yielding more than 100 usable responses. The third element of the research – aimed at getting a more grounded view of the day-to-day work of vice chancellors – involved institutional case studies. These were based on semi-structured interviews in ten institutions (covering vice chancellors and other members of staff and governors) as well as nonparticipant observation of three further vice chancellors at work over a period of 2–3 weeks in their institutions.

A second reason to value this work is that it focuses on the rhetoric and the reality of a historical transition in the role of institutional leaders – from academic to managerial leadership. This transition is still in progress in many countries as they experience governance reforms that reflect continuing change in the relationship between higher education and the state. Bargh and colleagues demonstrate that the assumed trajectory from one form and style of academic and collegial leadership to managerial and corporate leadership is not so clear-cut. This is partly because of the diversity of institutional missions, histories, and cultures and partly because the transitional state encompasses both forms, as represented in new titles of vice chancellor and chief executive. The evidence drawn from their comparative research suggests that the British and European leaders and, to a lesser extent, their US counterparts shared common experiences. Each in different ways were coping with a complex and often conflicting set of pressures including the “the contradictions of free-market rhetoric, the apparent retreat of the state from planning and the competing demands of corporate control versus creativity and entrepreneurship.” The authors conclude that the role relies on a set of implicit and negotiated powers that align more readily with political organizations and processes than overtly commercial and corporate ones.

The study examines university leadership in detail, tracing the evolution of the role from 1945 in five successive periods, all of which are closely connected to the prevailing sequential political contexts of consolidation, growth and innovation, as well as crisis and challenge. Through the empirical data and analysis, the authors highlight the impact of this context on the changing structures of universities as well as the internal and external role of their leaders. External accountability features particularly strongly as part of the new demands of the role.

This includes traditional representational functions as well as managing and redefining the boundaries of the institution with government and an array of new agencies set up to assess, guide, or support the internal functions of the institution. External accountability also encompasses relationships with the governing body. Since the research was completed, the role of the governing body has become more prominent, and the range of business at board level heavier, more complex, and demanding. This is also true of the range of external partnerships and collaborative ventures with which universities engage, from business and the community to international consortia. The external role of the vice chancellor or president is increasing in most countries, and the balance of time spent on internal and external roles remains a preoccupation for institutional leaders. However, Bargh and co-workers found much variation in practice with no agreement among their respondents about the appropriate balance at any one time.

The study provides valuable insight into the internal role of vice chancellor. Classic leadership responsibilities are examined including developing the mission, strategy building and strategic planning, creating structures in support of strategy, and balancing continuity and change in relation to the values of the organization. The vice chancellor's role in academic leadership is also analyzed. This is seen as central to strategy and mission and, in a modern university, involves managing the interface between governors and senior academic colleagues. The unique features of the vice chancellor's role are described as combining academic and symbolic leadership with managerial expertise. Further fine-grained detail is provided of how the social practice of leadership is exercised. There are clear parallels here to the earlier US study – both sets of researchers, drawing on their data – that emphasize the need to think of university leadership in terms of the language and metaphors of complex political systems, “rather than engineering and simple linear models of command and control.”

The comparative data collected in the Bargh study highlight further points of convergence. For presidents, rectors, and vice chancellors there are shared preoccupations: with strategy, with finding appropriate mechanisms to link control with entrepreneurship and competitive advantage, and in dealing with the consequences of dwindling state resources for higher education. However, beyond being subject to similar environmental forces, there remains great diversity of leadership style and behavior reflecting both situation-specific agendas as well as the individuality of the incumbent. To provide a further check on similarities, differences, and possible convergence in presidential roles and responsibilities, we need to look outside Europe, North America, and Australasia. Transforming and transition countries are an important source of data. Cloete and co-workers provide

an overview of the contexts and challenges facing South African institutional leaders, particularly those in historically disadvantaged institutions (HDIs). Some presidential roles are the same: developing mission, strategic planning, mobilizing support, increasing efficiency, and financial planning and management. However, the contexts are significantly different in terms of resources, political, social, and economic challenges, leading in a strongly unionized environment and dealing simultaneously with a variety of crises.

Further useful contributions from this region have focused on the role of South African vice chancellors in institutional transformation. Bell explores an African perspective on transformational or transformative leadership through interviews with ten vice chancellors of HDIs. No commonality of perspective on leadership, or an African style of leadership, emerged. Instead, Bell's study highlights several conflicts that illustrate the tensions experienced in moving from scholarly notions of leadership (Scott's donnish phase of governance) to a managerial phase with strong expectations of democracy built into it. Kulati's study is also about tensions; this time between reliance on governance transformation at the structural level while neglecting the role of institutional leadership in shaping change. He too explores the shifting balance between collegial and managerial forms of leadership, arguing the case for transformative leadership that facilitates effective institutional management in the context of a professional culture that eschews being. At the end of his paper, Kulati raises important questions for further research, namely, the agency versus structure tension and the personality versus environmental fit challenge. In other words, to what extent are individual leaders and their senior teams critical to achieving success in challenging times?

Senior Management Teams

Both the studies cited above not only consider the role, style, and impact of individual leaders in leading and managing the university but also the part played by senior staff in the collective process of leadership. Other studies also contribute to the picture: Gardiner surveyed a small focused sample of presidents to explore their role in team building and the characteristics of an effective presidential team. Middlehurst takes a wider view of group dimensions of leadership by examining the contribution of senior administrators, lay officers, and committees, as well as senior management groups, describing the interactions, relationships, and dimensions of symbolic, stylistic, and substantive institutional leadership as a complex drama. Major studies of senior management teams in higher education are, however, more limited than studies of presidents. Two are selected: the first arose from the ILP in the US and examines teams and teamwork as part

of leadership, whereas the second is a UK-based study recently completed by Woodfield and Kennie.

Bensimon and Neuman report the difficulties of language (e.g. “a team is not always a team”) and summarize the advantages and disadvantages of teamwork. The authors also offer valuable insights into how presidents do or do not make use of their teams. They find that effective team builders construe the work of their teams in terms of three functions: utilitarian, expressive, and cognitive. The first is formal and task oriented, assisting the president in achieving a sense of rationality and control over institutional functioning. The second is social and integrative, reinforcing a sense of connectedness between individuals involved in joint ventures. The third involves sense-making with the purpose of enabling the group to behave as a creative or corrective system able to see problems from multiple perspectives, to question and challenge. Bensimon and Neumann differentiate between real teams that focus on all three dimensions and illusory teams that focus only on the utilitarian function. They note that the key difference between the two is how the team thinks together and again offer a useful conceptualization of eight thinking roles for a presidential team. These include five core roles, namely definer, analyst, interpreter, critic, and synthesizer, as well as three roles that support the core, disparity monitor, task monitor, and emotional monitor. Understanding and balancing different roles is an important contributor to effectiveness in teams’ thinking, learning, and acting together. The authors suggest that while they observed the process of team learning at the micro-level, it was likely also to be reflected in the wider organization. These insights locate this study in relation to wider research on the learning organization and also to the burgeoning field of shared leadership and distributed leadership.

Woodfield and Kennie’s study of top team structures is similarly empirically based and practically orientated toward helping institutional leaders and their teams analyze their role and performance. Two conceptual frameworks have been created from evidence drawn across the UK sector. The authors relate these frameworks to the range of functions that top teams typically undertake and suggest their suitability for different institutional contexts. From their desk research and interviews (to be supplemented by questionnaire data), the researchers highlight key issues related to the formation and operation of top teams that are likely to impinge on their operational effectiveness. These include:

1. clarifying the criteria for membership of the top team;
2. understanding the meaning of team work in a top team;
3. deciding on the focus of top team activities and functions;
4. outlining agreed working practices;
5. dealing with conflicts and competing priorities;

6. evaluating both personal and team effectiveness;
7. relationship building and interaction;
8. reviewing succession planning, training, and development; and
9. providing relevant support and resources for top team activities.

The findings from the UK study are already beginning to identify not only different modes of team operation but also the range of challenges that top teams face in different types of institutions. The research also provides further evidence of tension, transition, and the need for alignment in the ways that universities are led and managed through a variety of formal and informal structures and groupings.

Future Research

There are many gaps and avenues for further work in the general territory covered above. First, the higher education literature is more limited in range and scope than the corporate literature on chief executive leadership and the operations of senior teams. Equally, bespoke research in higher education would benefit from wider and deeper comparisons with other sectors, by including interdisciplinary perspectives and by examining institutions as organizations. The dynamics of change, transition, and transformation (as well as continuity) could usefully be examined more fully, both in terms of stages of organizational and historical development and from the perspective of different cultures and traditions.

At a macro-level, the interplay between university leadership and the wider political, economic, and social system are worth exploring as institutions and their leaders are drawn more deeply into the rhetoric and reality surrounding the role of universities in developing or transforming societies. However, micro-level treatment of this agenda is also needed for a deeper understanding of the part played not only by the president and his or her team, but also by other leaders, including professional managers, senior academic managers and professors, and teachers and researchers. While the career routes of presidents continue to be charted and examined, the wider career and developmental experiences of members of senior teams are not so well researched. In this context, the particular experiences of those leaders who move into and out of higher education should be explored. At the individual level – and at the level of structural and cultural processes – there remains a pressing need to research the experiences of would-be or could-be leaders, including women and ethnic minorities; Eveline’s recent work is a welcome contribution in this territory.

There are also disconnects that require bridging. There is a gap between: (1) theory and practice, (2) policy

interests and the agendas of researchers, (3) macro- and micro-level studies, and (4) empirical and normative literature. There are also fruitful areas for combining investigation and evidence, for example, between the world of practitioners and consultants working in higher education. Ultimately perhaps, research should reflect the landscape in which the subjects operate and make their contribution to the development of higher education; it should more fully examine the interplay between academic, symbolic, managerial, and societal terrains.

Further Reading

- Asking, B. (2001). In search of new models of institutional governance: Some Swedish experiences. *Tertiary Education and Management* 7, 197–210.
- Association of Governing Boards of Universities and Colleges (2006). *The Report of the Task Force on the State of the Presidency in American Higher Education: The Leadership Imperative*. Washington, DC: Association of Governing Boards of Universities and Colleges.
- Baimbridge, M. and Simpson, C. (1996). Rewards to academia: The remuneration of vice chancellors and principals. *Applied Economics* 28, 631–639.
- Bargh, C., Bocock, J., Scott, P., and Smith, D. (2000). *University Leadership: The Role of the Chief Executive*. Buckingham: Society for Research on Higher Education/Open University Press.
- Bauer, M., Asking, B., Marton, S. G., and Marton, F. (1999). *Transforming Universities: Changing Patterns of Governance, Structure and Learning in Swedish Higher Education. Higher Education Policy Series 48*. London: Jessica Kingsley.
- Bell, C. and Mitchell, B. (1991). Vice-chancellorships since Collison and Millen: A research note and comment. *Sociology* 25, 521–522.
- Bell, D. (2001). Leading change: African conceptions of leadership and transformation in higher education in South Africa. *Paper Presented at the Higher Education Close-Up 2 Conference*. Lancaster University, UK, 16–18 July 2001.
- Bensimon, E. M. (1989). Five approaches to think about: Lessons learned from experienced Presidents. In Bensimon, E. M., Gade, M., and Kauffman, J. (eds.) *On Assuming a College or University Presidency: Lessons and Advice from the Field*, pp 1–12. Washington, DC: American Association for Higher Education.
- Bensimon, E. M. (1990). The new president and understanding the campus as a culture. In Tierney, W. G. (ed.) *Assessing Academic Climates and Cultures. New Directions for Institutional Research*, No. 68. San Francisco, CA: Jossey-Bass.
- Bensimon, E. M. (1991). The social process through which faculty shape the image of a new resident. *Journal of Higher Education* 62, 637–660.
- Bensimon, E. M. and Neumann, A. (1993). *Redesigning Collegiate Leadership: Teams and Teamwork in Higher Education*. Baltimore, MD: Johns Hopkins Press.
- Bensimon, E. M., Neumann, A., and Birnbaum, R. (1989). Making sense of administrative leadership: The “I” word in higher education. *ASHE-ERIC Reports on Higher Education*, No. 1. Washington, DC: ASHE.
- Birnbaum, R. (1986). Leadership and learning: The college president as intuitive scientist. *Review of Higher Education* 9, 381–395.
- Birnbaum, R. (1988). *How Colleges Work: The Cybernetics of Academic Organization and Leadership*. San Francisco, CA: Jossey-Bass.
- Birnbaum, R. (1989). Leadership and followership: The cybernetics of university governance. In Schuster, J. H. and Miller, L. (eds.) *Governing Tomorrow's Campuses: Perspectives and Agendas*, pp 27–41. Washington, DC: American Council on Education/Macmillan.
- Birnbaum, R. (1989). The implicit leadership theories of college and university presidents. *Review of Higher Education* 12, 125–136.
- Birnbaum, R. (1990). How'm I doin'?: How college presidents assess their effectiveness. *Leadership Quarterly* 1, 25–39.
- Birnbaum, R. (1992). *How Academic Leadership Works: Understanding Success and Failure in the College Presidency*. San Francisco, CA: Jossey-Bass.
- Bolden, R., Petrov, G., and Gosling, J. (2008). *Developing Collective Leadership in Higher Education*. London: Leadership Foundation for Higher Education.
- Bryman, A. (2007). *Effective Leadership in Higher Education*. London: Leadership Foundation for Higher Education.
- Cameron, K. S. (1981). Domains of organizational effectiveness in colleges and universities. *Academy of Management Journal* 24, 25–47.
- Clark, B. R. (1998). *Creating Entrepreneurial Universities*. Oxford: Pergamon.
- Cloete, N., Kulati, T., and Phala, M. (2000). Leadership and institutional change in higher education. *Papers from a Seminar in Gordon's Bay*, South Africa, 16–19 May, 1999, Pretoria: Centre for Higher Education Transformation.
- Coaldrake, P. (1999). *The changing climate of Australian higher education: An international perspective*. *Higher Education Management* 11, 117–134.
- Coaldrake, P. and Stedman, L. (1998). *On the Brink: Australia's Universities Confronting Their Future*. St. Lucia: University of Queensland Press.
- Cohen, M. D. and March, J. D. (1986). *Leadership and Ambiguity: The American College President*, 2nd edn. New York: McGraw-Hill.
- Collison, P. and Millen, J. (1969). University chancellors, vice-chancellors and college principals – social profile. *Sociology* 3, 77–109.
- Conraths, B. and Trusso, A. (2007). *Managing the University Community: Exploring Good Practice*. Brussels: European University Association.
- Eggs, H. (ed.) (1997). *Women as Leaders and Managers in Higher Education*. Buckingham: Society for Research on Higher Education/Open University Press.
- Eveline, J. (2004). *Ivory Basement Leadership: Power and Invisibility in the Changing University*. Crawley: Western Australia Press.
- Gardiner, J. J. (1988). Building leadership teams. In Green, M. F. (ed.) *Leaders for a New Era: Strategies for Higher Education*, pp 137–153. New York: American Council on Education/Macmillan.
- Gioia, D. A. and Thomas, J. B. (1996). Identity, image and issue interpretation: Sensemaking during strategic change in academia. *Administrative Science Quarterly* 41, 37–403.
- Goodall, A. H. (2006). Should top universities be led by top researchers and are they? A citations analysis. *Journal of Documentation* 62, 388–411.
- Green, M. F. (ed.) (1988). *Leaders for a New Era: Strategies for Higher Education*. New York: American Council on Education/Macmillan.
- Green, M. F. and McDade, S. A. (1991). *Investing in Higher Education: A Handbook of Leadership Development*. Washington, DC: American Council on Education.
- Hambrick, D. C. and Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *Academy of Management Review* 9, 193–206.
- Jarzbowski, P. (2003). Strategic practices: An activity theory perspective on continuity and change. *Journal of Management Studies* 40, 23–55.
- Jarzbowski, P. and Wilson, D. C. (2002). Top teams and strategy in a UK university. *Journal of Management Studies* 39, 355–381.
- Johnson, R. (2002). Learning to manage the university: Tales of training and experience. *Higher Education Quarterly* 56, 33–51.
- Kulati, T. (2003). From protest to challenge: Leadership and higher education change in South Africa. *Tertiary Education and Management* 9, 13–27.
- McNay, I. (1995). From the collegial academy to corporate enterprise: The changing cultures of universities. In Schuller, T. (ed.) *The Changing University?*, pp 105–115. Buckingham: Society for Research on Higher Education/Open University Press.
- Middlehurst, R. (1993). *Leading Academics*. Buckingham: Society for Research on Higher Education/Open University Press.
- Middlehurst, R., Pope, M., and Wray, M. (1992). The changing roles of university leaders and managers: Implications for preparation and development. *CORE* 16, Fiche 5 E 11.

- Mignot-Gerard, S. (2003). "Leadership" and "governance" in the analysis of university organisations: Two concepts in need of de-construction. *Higher Education Management and Policy* **15**, 135–163.
- Neumann, A. (1990). Making mistakes: Error and learning in the college presidency. *Journal of Higher Education* **61**, 386–407.
- Neumann, A. (1990). On the making of 'good times' and 'hard times': The social construction of resource stress. *Report OP 90:18*. College Park, MD: Center for Higher Education Governance and Leadership.
- Neumann, A. and Bensimon, E. (1990). Constructing the presidency: College presidents' images of their leadership roles: A comparative study. *Journal of Higher Education* **61**, 678–701.
- O' Meara, B. and Petzall, S. (2005). Vice chancellors for the 21st century? A study of contemporary recruitment and selection practices in Australian universities. *Management Research News* **28**, 18–35.
- Pearce, C. L. and Conger, J. A. (eds.) (2003). *Shared Leadership: Reframing the Hows and Whys of Leadership*. London: Sage.
- Ramsden, P. (1998). *Learning to Lead in Higher Education*. London: Routledge.
- Scott, P. (1996). *The Meanings of Mass Higher Education*. London: Wiley.
- Senge, P. (1990). *The Fifth Discipline: The Art and Practice of Leading the Learning Organization*. New York: Doubleday.
- Shattock, M. (2003). *Managing Successful Universities*. Maidenhead: Society for Research on Higher Education/Open University Press.
- Shavlik, D. and Touchton, J. (1988). Women as leaders. In Green, M. F. (ed.) *Leaders for a New Era: Strategies for Higher Education*, pp 98–117. New York: American Council on Higher Education/Macmillan.
- Sloper, D. W. (1984). A social characteristics profile of Australian vice-chancellors. *Higher Education* **14**, 355–386.
- Sloper, D. W. (1989). The selection of vice-chancellors: The Australian experience. *Higher Education Quarterly* **43**, 247–265.
- Smith, D., Bargh, C., Bocock, J., and Scott, P. (1999). New leaders at the top? The educational and career paths of UK university vice-chancellors (1960–1996). *Higher Education Management* **11**, 113–135.
- Taylor, J. (2006). "Big is beautiful". Organisational change in universities in the United Kingdom: New models of institutional management and the changing role of academic staff. *Higher Education in Europe* **31**, 251–273.
- Taylor, J. and Machado, M. (2006). Higher education leadership and management: From conflict to interdependence through strategic planning. *Tertiary Education and Management* **12**, 137–160.
- van den Bosch, H. and Teelken, C. J. (2000). Organisation and leadership in higher education: Learning from experiences in the Netherlands. *Higher Education Policy* **13**, 379–397.
- Watson, D. (2000). *Managing Strategy*. Buckingham: Open University Press.
- Weil, S. (ed.) (1994). *Introducing Change from the Top in Universities and Colleges: 10 Personal Accounts*. London: Kogan Page.
- Wilson, R. and Melendez, S. E. (1988). Strategies for developing minority leadership. In Green, M. (ed.) *Leaders for a New Era: Strategies for Higher Education*, pp 118–136. New York: American Council on Education/Macmillan.
- Woodfield, S. and Kennie, T. (2007). Top team structures in UK higher education institutions: Composition, challenges and changes. *Tertiary Education and Management* **13**, 331–348.

Management of and in Higher Education Institutions

B Sporn, WU (Vienna University of Economics and Business), Vienna, Austria

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Adaptive university – An adaptive university is an institution that needs to work with its environment in order to thrive and survive and is able to respond flexibly. Structures and processes of adaptation involve a committed and supporting leadership. Differentiation of structure is an important prerequisite for adaptation.

Administration – Administration is one of three distinct features (i.e., governance, leadership, and administration) of higher education management and refers to the operational side of running an institution, that is, the structures and processes by which decisions are planned, organized, and controlled.

Board – College or university boards represent very important stakeholders of the individual institution who can range from business managers, politicians, professors to international experts. Board members tend to come from outside the institution and tend to be accountable to the ministry of local government as well as trustees of the institution.

Entrepreneurial university – A university that acts in a more market-oriented way and is run by committed leaders who act based on a coherent strategy and who integrate different constituencies.

Governance – Governance is one of three distinct features (i.e., governance, leadership, and administration) of higher education management and refers to the structure and process of decision making.

Idiosyncratic nature of HE¹ – The idiosyncratic nature of higher education includes ambiguous goals, multiple constituencies, unclear technologies, the pivotal role of experts in the organization, and the environmental vulnerability of colleges or universities.

Leadership – Leadership is one of three distinct features (i.e., governance, leadership, and administration) of higher education management and implies the role of top-level positions taking responsibility for the overall institution.

New public management – New public management is a leading principle in higher education management and dominant in Europe due to the large number of European public institutions. It emphasizes competition, hierarchical and

professional management, and output funding (Schimank, 2005).

Performance contracts – Performance contracts are contracts that were developed as a management tool which define service levels and performance, including indicators such as research productivity or number of entering students or graduates.

Professionalization of HE – Professionalization of university management emphasizes the fact that colleges and universities need to be managed by higher education experts rather than by lay faculty members. Know-how such as leadership, human resource development, financial control and marketing, as well as conflict management are necessary for professional HE managers.

Institutions of higher education have experienced a period of change and transformation since the early 1990s. Colleges and universities have been confronted with an ever-increasing competitive environment. As a response, market mechanisms have intensified and institutions of higher education have been expected to become even more efficient and effective. With this, issues concerning the management of colleges and universities are a critical element in the success of institutions.

Management can be defined as the structures and processes of leadership, governance, and administration. When thinking about management, it is important to keep the idiosyncratic nature of higher education institutions in mind (Baldridge *et al.*, 1977; Clark, 1983; Weick, 1976). Colleges and universities are known for their ambiguous goals, multiple constituencies, unclear technologies, the pivotal role of experts in the organization (i.e., professors), and their environmental vulnerability. Models describing this specific nature range from organized anarchy (Cohen and March, 1974), loosely coupled systems (Weick, 1976) collegial, bureaucratic, and political approaches (Birnbaum, 1989) to adhocracy (Mintzberg, 1979). Management has to take these characteristics into account in order to be successful. Questions of shared governance, the role of leadership, and the shift from bureaucratic compliance to professional management often arise.

Across the continents, higher education management developed new approaches to make colleges and universities more efficient and effective. In Europe, new public management has been used to introduce institutional autonomy and accountability (Schimank, 2005; Shattock, 2003).

¹ higher education.

In the US, increased marketization led to the commercialization of management with special emphasis on revenue generation and professional administrators (Bok, 2003; Rhoades, 1998; Slaughter and Leslie, 1997). In Asia, most of these models were copied and related to the idiosyncratic situation of the respective country (Kennedy, 2003). Australia followed most of the Anglo-Saxon tradition with an increased market-driven approach (Marginson and Considine, 2000).

Key Management Issues

The management of higher education involves three distinct features: governance, leadership, and administration. Governance refers to the structure and processes of decision making. Leadership implies the role of top-level positions taking responsibility for the overall institution. Administration refers to the operational side of running an institution, that is, the structures and processes by which decisions are planned, organized, and controlled (Sporn, 1999).

Governance

One major issue of higher education management refers to governance and decision making inside the institution. Going back to the work of the renowned higher education researcher Burton Clark (1983), three levels constitute the governance structure of systems and institutions alike. Clark refers to this phenomenon as the triangle of coordination. In a nutshell, the levels include the overall system, the institution, and the academic community. Depending on the power and strength of any given level, three coordination styles can be differentiated: state authority, market, and academic oligarchy. Accordingly, certain rules apply to governance depending on the character of the system. They could be rules and regulations, mechanisms of supply and demand, or peer pressure. Accordingly, Clark positioned different countries and their higher education system within the triangle. For example, the US turned out to be market oriented while Italy was basically governed by professors (academic oligarchy). Former Russia (USSR) as well as Sweden and France were defined as a state-controlled system.

With the rise of neo-liberal policies globally focusing on effectiveness and accountability, institutional autonomy empowered central administration and leadership (i.e., rectors, presidents, and vice chancellors of institutions, and deans and heads of institutes or departments). External stakeholders (as board members) started to play an important role in governance representing general societal interests (Birnbaum, 2000; Maassen, 2000; Slaughter and Leslie, 1997). Overall, market mechanisms of supply and demand changed the notion of higher education from

a social system to an industry with important implications for institutional governance (Gumport, 2000).

Governance structures define the role of institutional leaders, academic faculty senates, individual faculty, central administration, and external stakeholders. Leaders in colleges and universities have an ever-increasing importance as they are responsible for the overall success of the institution. They are accountable to an external board. The power of the faculty senate has often been reduced to either a consultancy function (e.g., US, the Netherlands) or to a concentration on issues of academic nature, that is, teaching and research (e.g., many European countries; De Boer, 2001).

A stakeholder approach in higher education governance is most visible through external boards at colleges and universities (e.g., US: board of regents, board of trustees; Europe: university boards, board of trustees). Board members represent very important stakeholders of the individual institution who can range from business managers, to politicians, professors to international experts. The board members come from outside the institution and tend to be accountable to the ministry or local governments as well as trustees of the institution. Appointments to such boards are an important and sometimes contested issue. They can range from purely ministerial to a combination of institutional and governmental nominations (Gumport and Pusser, 1999). Boards in that sense have an important external control and linkage function with the external institutional environment.

Another important form of governance structures are executive boards. They often consist of the top leadership team, that is, rector and vice-rectors, dean and associate deans, or presidents, provost, and vice presidents. They make decisions at the highest level of the institution. Members are often appointed by the board of trustees. There is a functional divide with responsibilities ranging from research to international affairs and board members decide, for example, upon research funds allocation, promotion procedures, and general infrastructure.

The senate is the third element of institutional governance. At this level, major areas for decision making concern the core function of universities in teaching and research. Members can include all faculty members or a subset as delegates of different fields or schools. In the Anglo-Saxon model, senates have mostly advisory function and consist of professors and/or academics with faculty-rank only. The continental European model sees the senate as a decision-making body with membership of all groups, that is, representatives from the group of faculty, students, and administration (Weiler, 2005).

Leadership

The leadership of colleges and universities has been strengthened over the last decades in most higher education

systems. The rector, president or chancellor as well as the vice rectors and deans are the ones running the institution at the top. They are supported by an administrative structure which is also moving from a bureaucratic to a managerial model (Birnbaum, 2000; Mora, 2001).

Institutional leaders have been established as the key figures and major players within the steering of colleges and universities. Their role can be compared with that of a chief executive officer (CEO) or head of the board of directors. Responsibilities encompass the planning and implementing of all major areas ranging from budget and space to teaching and research. He or she is accountable for the success or failure of the institution. The president would often be responsible for external relations and fundraising while a provost would lead the academic side of the enterprise. Leaders are no longer necessarily hired from the faculty of the specific institution. Comprehensive searches are run to find candidates with adequate qualifications. Still in order to guarantee credibility, some academic/faculty background or an equivalent expertise like management of a research center is necessary.

Leadership structures have to combine responsibility and accountability (Weiler, 1998). In many countries, the management practice has been to divide strategic and operational issues. The consequence has been that committees decided (sometimes without the necessary background information, data, or preparation) on issues for which they then have not been held accountable. For example, a senate could decide on the introduction of a new study program without the necessary budget plan. With accountable forms of leadership in place, institutions of higher education define responsibilities based on the specific areas of expertise. Hence, a senate would look at curricula and quality-control issues, whereas a leadership team would fund programs and provide necessary infrastructure.

The leadership level at colleges and universities includes a whole set of growing positions. Vice rectors, deans, and program directors take over certain areas ranging from finance, research and teaching, and international relations to undergraduate and graduate programs. Candidates are often found within the faculty of the institution. For full-time positions, special financial incentives are necessary to attract the right individuals. Deans and vice rectors are normally members of the leadership team with their budgets and areas of responsibility. Often they function as an important liaison between the faculty and the administration. Trust between groups and identification with institutional goals seems to be one of the most important success factors in today's colleges and universities. From European universities, it can be learnt that a faculty-dominated administration or top-level leadership team can help the institution to thrive and overcome major obstacles (Rhoades and Sporn, 2002a; Sporn, 1995).

In the US, administration is sometimes twice as large as the faculty. With this, it is a pivotal task of university leaders to pay close attention to the integration of academic and administrative cultures (Clark, 1997; Dill, 1982; Sporn, 1996).

Administration

The management of colleges and universities is defined as the structures and processes by which decisions are implemented. In a sense, structures imply the role of administration, and processes mean instruments and tools of management. Looking back over the last 25 years reveals that a tremendous change occurred within colleges and universities. In the 1970s, European universities and colleges were based on a bureaucratic model of management with a divided organizational structure, that is, a state bureaucracy at individual institutions responsible for compliance with legal requirements and an academic guild providing the core performance in teaching and research. Following this period, a collegial form of management evolved. This type can be characterized as institutional management and administration firmly rooted in the faculty. Professors have taken over major administrative positions for running the institution in this manner. With the rise of a new policy in higher education, more market pressure, and the call for professionalism, a managerial approach to administration developed. This referred to the use and adoption of private industry tools such as performance indicators, personnel development, and standard reporting. Another major trend can be observed toward more entrepreneurial forms of management. Here institutional leadership sets up certain incentive structures to create opportunities for individual members of the faculty to create new ways of income and generally doing business. Hence, accountability and responsibility for most activities are decentralized. The top-level management solely provides an infrastructure and controls for results (Clark, 2004; Shattock, 2000, 2003; Sporn, 1999).

Problems of management arise from the dualism of controls (Birnbaum, 1989) at most colleges and universities. Compared to business firms, universities have two parallel systems based on very different goals. While the faculty is mainly concerned with the pursuit and dissemination of knowledge, administration has to aim at efficiency and effectiveness of the whole institution. This tension increases with resource scarcity and uncertainty (Cohen and March, 1974). As mentioned earlier, a shift of authority from faculty to administration can be observed.

In a more competitive environment of scarce resources, clarity and agreement on the organizational mission are considered a fundamental basis of success. Higher education management responds by developing a mission, establishing strategic plans, and defining common objectives.

However, with the complexity of academic and administrative goals and little agreement on priorities and measures of goal achievement, it is hard to develop coherent and consistent mission statements. Teaching, research, and service – the dominant three overall goals in higher education – are too broad to serve as starting points for mission statements. Higher education institutions could be managed more effectively if their missions were clarified but this has proven to be difficult to do in larger and more complex organizations. Hence, administration and management have to learn to function within the context of conflicting objectives (Birnbaum, 1989; Kezar and Eckel, 2002).

Yet another problem of higher education management is the distribution of institutional power. Organizations such as colleges and universities rely on expert power (i.e., faculty). Professors are less motivated by salary incentives than by internalized principles of academic freedom and ethical behavior, and by exchange with colleagues with similar values. Hence, faculty behavior cannot be managed successfully by business firm standards. Administration needs to find ways to address this issue by responding to the intrinsic motivation of faculty through recognition, involvement, and respect.

Overview of Management Approaches

Taken together, governance, leadership, and administration form the management in and of higher education. As many authors have noted, a changing environment of colleges and universities has called for a change in management approaches. Given the rising skepticism of society with its institutions, legitimacy and accountability of higher education have become prominent issues (Gumport, 2000). Accordingly, different approaches emerged which are outlined as follows: new public management, strategic management, and professionalization of university management. Their application differs depending on institutional type – especially regarding public and private colleges and universities.

In Europe with its dominance of public institutions, new public management and management by objectives based on performance contracts are leading principles in higher education management. This is tightly connected to the notion of performance funding. New public management emphasizes, for example, competition, hierarchical and professional management, and output funding (Schimank, 2005). Competition between and within institutions based on the principles of supply and demand (in the form of proposals and negotiations for funding) has been introduced in order to spur quality. Management had to respond by becoming more business-like with clear reporting lines and experienced managers involved (and not faculty members). Funding has been based on transparent target goals. In many countries, contracts

were developed as a management tool which defined service levels and performance, including certain indicators ranging from research productivity or number of entering students to graduates.

Ex-post steering is a distinct element of new public management. Institutions can be evaluated regarding their input they are able to process or the output they produce with any given resources. Trends show that output orientation gains importance at colleges and universities. Colleges and universities need to demonstrate the number of graduates, patents, publications, and services to the community in order to receive funding and respect. A key challenge is the definition of indicators. Faculty members believe mostly in the concept of peer review for an objective evaluation of their performance in teaching and research. They find it hard to accept other forms of feedback such as student ratings, rankings by journals, or government indicators. Professors rely more on referred journal articles, citation indexes, or groups of peers visiting their classes. A clear conceptual framework should be part of the total quality management of higher education institutions which takes these differences into consideration (Dill, 1993; Rhoades and Sporn, 2002b).

Strategic management emphasizes the role of the environment and the types of changes which affect colleges and universities. The major goal of strategic management is to keep institutions successful in a changing environment based on flexible and adaptive structures together with a clear strategy. As Shattock (2000, 2003) pointed out successful strategic management needs to take a couple of elements into consideration: competition, opportunism, income generation and cost cutting, relevance, excellence, and reputation. Environments for higher education have turned into competitive markets for talent. Institutions have to be able to assess opportunities and turn them into entrepreneurial activities, thereby providing the basis for acquiring additional funds. At the same time, costs need to be understood and reduced wherever possible. Programs are offered on this market with the understanding to provide value for clients. This should enhance institutional reputation and overall excellence.

The professionalization of university management emphasizes the fact that colleges and universities need to be managed by trained experts rather than lay faculty members. The US has seen a long tradition of university and college managers. Only presidents and provost often need to have some faculty background. Many institutional leaders come to the job ill-prepared and with wrong assumptions. Necessary experience and know-how include, for example, leadership, human resource development, financial control and marketing, as well as conflict management. Academic training programs in higher education management, competitive salaries, and international searches are standard instruments to increase the professional management of colleges and universities. With this, managers

become powerful players who use instruments of business to run the institution. Conflicts arise as values between academic and administrative staff clash. A new compromise seems to be underway where faculty concentrates on teaching and research whereas managers handle the administration of the institution (Rhoades and Sporn, 2002b).

Concluding Perspectives

There are different concepts which have been coined by higher education researchers to describe trends for education management. Although they have been developed in the late 1990s, they still seem to be valid: the learning university, the adaptive university, and the entrepreneurial university.

The model of a learning university resonates with the core function of universities, that is, teaching and research, and applies the concept of organizational learning to institutions of higher education. The underlying principles regarding the learning universities are twofold. On the one hand, the university must provide an atmosphere where student learning can be enhanced. On the other hand, universities as learning organizations have to restructure in order to better react flexibly and proactively to a dynamically changing environment. Some factors play a critical role for university management (Senge, 2000): line managers as promoters, academic units as change agents, university leaders learning from private industry, self-organized restructuring, faculty involvement, and internal networks. The learning approach enables higher education managers to take a fresh look at the structures and processes of innovation. Accordingly, university management needs to develop structures which enhance learning by allowing the building of internal and external networks, by providing sufficient data on performance and environmental trends, and by finding innovative forms of collaboration between faculty and administration.

The adaptive university is based on the assumption that institutions need to work with their environment in order to survive and thrive. Structures and processes of adaptation involve a leadership which is committed and has a supportive role, that is, providing resources to initiatives and to projects initiated by the core units (i.e., academic institutes and departments). Management should be professional, and shared governance involving both faculty and administration should be in place. Differentiation of structure is an important prerequisite for adaptation. In order to respond flexibly, different units can help to respond to a diverse set of external demands. Clear goals will then provide sufficient direction of all university activities. An entrepreneurial culture can prepare the grounds for a climate where individual projects get rewarded and institutional interests (and finances) are protected (Sporn, 1999). Especially the value system and

the institutional culture need to reinforce adaptation. The orientation toward change, risk-taking, and flexibility is an important prerequisite for successful adaptation (De Zilwa, 2007; Kezar and Eckel, 2002).

The entrepreneurial university acts in a more market-oriented higher education system. An entrepreneurial university management is run by committed leaders who act based on a coherent strategy and who integrate different constituencies. Linkages with outside organizations and groups help to reach across traditional institutional boundaries. Differentiated funding based on public spending as well as corporate and private giving enhance the independence of the institution and increase the number of initiatives. The academic units need to be integrated and respected for their role as the core providers of teaching and research. A common value system should be in place which reconciles managerial groups and academics. A culture which supports change will form the basis of the entrepreneurial university (Clark, 1998). From an international perspective, Clark continued his work around the globe to conclude with the concept of the self-reliant university, that is, an institution operating in a highly competitive market and achieving success through tailor-made solutions and collegial entrepreneurialism. As he points out: "The mantra for reform becomes: complex universities operating in complex environments require complex differentiated solutions. One hundred universities require 100 solutions." (Clark, 2004: 183)

The management of and in higher education is a complex task. The nature of colleges and universities, the right balance of leadership, governance and administration, and the use of management instruments need to be chosen with great care in order to fit with the very nature of the institution. Once a culture of trust and respect has been established, higher education institutions can succeed and thrive in an environment which has become increasingly turbulent and competitive.

Bibliography

- Baldrige, J. V., Curtis, D. V., Ecker, G. P., and Riley, G. L. (1977). Alternative models of governance in higher education. In Riley, G. L. and Baldrige, J. V. (eds.) *Governing Academic Organizations*, pp 2–25. Berkeley, CA: McCutchan.
- Birnbaum, R. (1989). *How Colleges Work: The Cybernetics of Academic Organization and Leadership*. San Francisco, CA: Jossey-Bass.
- Birnbaum, R. (2000). *Management Fads in Higher Education: Where They Come from, What They Do, Why They Fail*. Chichester: Wiley.
- Bok, D. (2003). *Universities in the Marketplace: The Commercialization of Higher Education*. Princeton, NJ: Princeton University Press.
- Clark, B. R. (1983). *The Higher Education System: Academic Organization in Cross-National Perspective*. Berkeley, CA: University of California Press.
- Clark, B. R. (1997). Small worlds, different worlds: The uniqueness and troubles of American academic professions. *Daedalus* 126(4), 21–42.
- Clark, B. R. (1998). *Creating Entrepreneurial Universities: Organizational Pathways of Transformation*. Oxford: Pergamon.
- Clark, B. R. (2004). *Sustaining Change in Universities: Continuities in Case Studies and Concepts*. Berkshire: Open University Press.

- Cohen, M. D. and March, J. G. (1974). *Leadership and Ambiguity: The American College President*, 2nd edn. Boston MA: Harvard Business School Press.
- De Boer, H. (2001). On limitations and consequences of change: Dutch University governance in transition. *Tertiary Education and Management* 7(2), 163–180.
- De Zilwa, D. (2007). Organisational culture and values and the adaptation of academic units in Australian universities. *Higher Education* 54, 557–574.
- Dill, D. D. (1982). The management of academic culture: Notes on the management of meaning and social integration. *Higher Education* 11(3), 303–320.
- Dill, D. D. (1993). Quality by design: Toward a framework for quality management in higher education. In Smart, J. (ed.) *Higher Education: Handbook of Theory and Research*, vol. VIII, pp 37–83. New York: Agathon.
- Gumport, P. J. (2000). Academic restructuring: Organizational change and institutional imperatives. *Higher Education* 39(1), 67–91.
- Gumport, P. J. and Pusser, B. (1999). University restructuring: The role of economic and political contexts. In Smart, J. (ed.) *Higher Education: Handbook of Theory and Research*. New York: Agathon.
- Kennedy, K. J. (2003). Higher education governance as a key policy issue in the 21st century. *Educational Research for Policy and Practice* 2, 55–70.
- Kezar, A. J. and Eckel, P. D. (2002). The effect of institutional culture on change strategies in higher education: Universal principles or culturally responsive concepts? *Journal of Higher Education* 73(4), 435–460.
- Maassen, P. (2000). The changing roles of stakeholders in Dutch University governance. *European Journal of Education* 35(4), 449–464.
- Marginson, S. and Considine, M. (2000). *The Enterprise University: Power, Governance and Reinvention in Australia*. Cambridge: Cambridge University Press.
- Mintzberg, H. (1979). *The Structuring of Organizations*. Englewood Cliffs, NJ: Prentice-Hall.
- Mora, J. -G. (2001). Governance and management in the New University. *Tertiary Education and Management* 7, 95–110.
- Rhoades, G. (1998). *Managed Professionals: Unionized Faculty and Restructuring Academic Labor*. Albany, NY: State University of New York Press.
- Rhoades, G. and Sporn, B. (2002a). New models of management and shifting modes and costs of production: Europe and the United States. *Tertiary Education and Management* 8(1), 3–28.
- Rhoades, G. and Sporn, B. (2002b). Quality assurance in Europe and the U.S.: Professional and political economic framing of higher education policy. *Higher Education* 43(3), 355–390.
- Schimank, U. (2005). New public management and the academic profession: Reflections on the German situation. *Minerva* 43, 361–376.
- Senge, P. M. (2000). Die Hochschule als lernende Gemeinschaft. In Laske, S., Scheytt, T., Meister-Scheytt, C., and Scharmer, C. O. (eds.) *Universität im 21. Jahrhundert: Zur Interdependenz von Begriff und Organisation der Wissenschaft*, pp 509. München: Rainer Hampp.
- Shattock, M. (2000). Strategic management in European universities in an age of increasing institutional self reliance. *Tertiary Education and Management* 6(2), 93–104.
- Shattock, M. (2003). *Managing Successful Universities*. London: Open University Press.
- Slaughter, S. and Leslie, L. L. (1997). *Academic Capitalism: Politics, Policies, and the Entrepreneurial University*. Baltimore, MD: Johns Hopkins University Press.
- Sporn, B. (1995). Adaptation process at universities: Organizational implications of a complex environment. *Tertiary Education and Management* 1(1), 72–75.
- Sporn, B. (1996). Managing university culture: An analysis of the relationship between institutional culture and management approaches. *Higher Education* 32, 41–61.
- Sporn, B. (1999). *Adaptive University Structures: An Analysis of Adaptation to Socioeconomic Environments of US and European Universities*. London: Jessica Kingsley.
- Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly* 21, 1–19.
- Weiler, H. N. (1998). Different ways of decision making in higher education and higher education policy in Germany and the United States. In Müller-Böling, D., Mayer, E., MacLachlan, A. J., and Fedrowitz, J. (eds.) *University in Transition*, pp 287–293. Gütersloh: Bertelsmann Foundation.
- Weiler, H. N. (2005). Ambivalence and the politics of knowledge: The struggle for change in German higher education. *Higher Education* 49(1–2), 177–195.

Managerialism and Collegialism in Higher Education Institutions

M Shattock, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

New public management (NPM) reforms have been a characteristic of most developed countries in an effort to contain the rising costs of public services, although Pollitt and Boukaert suggest that the intensity has been greater in countries where governments comprise single-party majorities, such as in the UK, Australia, Canada, and New Zealand, than in countries where coalition or more consensual governments operate, such as in Finland, the Netherlands, or Germany (Pollitt and Boukaert, 2000); it is generally agreed that the stimulus for many countries was the Thatcher revolution in the UK. The impact of NPM or the new managerialism, as it is sometimes called, on the public services has created a new orthodoxy which governments have sought to reflect in the management of higher education. *The Report of the Steering Committee on Efficiency Studies in Universities* (the Jarratt Report) (CVCP, 1985) which was prompted and funded by the UK government and the Lambert Report (HMSO, 2003) funded by HM Treasury, as well as the *Higher Education Management Review* (the Hoare Report) (AGPS, 1995) in Australia are all examples of governments seeking to transfer reforming styles of management used elsewhere in the public services into the universities.

Managerialism as an Ideology?

Some scholars argue that new managerialism should be seen in ideological terms. Tapper writing about the governance of higher education sees the creation of the UK funding council system as:

part of a much broader range of thinking about the control of social policy in Britain in the light of economic failure and political ineptitude. (Tapper, 2006)

Deem and Brehony (2005) looking at new managerialism at the institutional level see it “very clearly as an ideological rather than a technical reform” and quote as evidence manager academics who “assert their right to manage both academics and other staff, thus suggesting that as a social group, such manager academics are very interested indeed in maintaining relationships of power and domination.” However, not all scholars would go this far. Kogan and Hanney (2000) define managerialism as:

the shift in power from senior academics and their departments to the central institution and the dominance of systems over academic values. (Kogan and Hanney, 2000)

However, they make it clear that this partly came about “from institutions’ need to meet new demands from fewer resources.” Marginson and Considine, writing of the parallel Australian scene paint a similar picture:

Those senior executives controlling the future of universities have both more power and less room to manoeuvre than before, while those dependent on them for leadership have fewer alternative means to define their futures than previously was the case. (Marginson and Considine, 2000a)

They define the enterprise university as having “strong executive control,” “a distinctly corporate character” where academic bodies have been “supplemented (and sometimes supplanted) by vice-chancellors’ advisory committees and private ‘shadow’ university structures,” and as operating in a situation where underfunding drives “a pseudo-market in fee incomes, soft budget allocations for specific purposes and contested earnings for new enrolments and research grants” and where some market elements, for example, the recruitment of overseas students, are driven by openly commercial motives (Marginson and Considine, 2000b). These characteristics derive very largely from the imposition of government (Dawkins)-led reforms, which imposed institutional mergers and marketization on universities involving a much more rapid institutional adjustment than was faced in the UK. However, the processes they list are internationally recognizable and have also been described as the growth of academic capitalism (Slaughter and Leslie, 1997) or of university entrepreneurialism (Clark, 1998). If we look for parallels in the application of NPM reforms across the public services generally, we see many of the same approaches: tightening controls, cutting budgets, freezing new hirings, campaigning against waste, squeezing the system, and modernizing it through faster and more flexible approaches to budgeting, managing, accounting for a delivering services, and through marketization, that is, increasing user responsiveness and introducing competition, as we see in higher education (Pollitt and Boukaert, 2000).

The Changing Context of University Management

The context of university management has also changed in some particular ways. All major higher education systems have expanded enormously leading to rapid institutional growth. In 1960, the University of London was the

only university in the UK that had more than 30 000 students but it was a collegiate university and the students (and the staff) were managed in self-governing colleges. Almost 50 years later, at least 15 universities in the UK have over 30 000 students managed on a unitary institutional basis. The other UK institutions have grown proportionately: 15 universities (not necessarily the same ones) have turnovers of over £300 million, an unimaginable figure even 20 years ago. Universities have become big business and require business-like management. In a similar period, the financial contribution of the state has gone up many times prompting much more rigorous public accountability regimes. Public accountability has been extended beyond simply finance – into the quality of teaching monitored through quality-assessment regimes. In addition, the UK, Australia, and New Zealand have introduced formal research-assessment processes.

Competition was essentially reputational until the first published league tables began to appear. The later appearance of global league tables merely added a further dimension to the nature of the competition. Competition breeds stress, demands shorter decision-making timescales, and discounts equity in favor of more *dirigiste* approaches to resource allocation. At the same time, the recognition of the dependence of the knowledge economy on university outputs, both nationally and regionally, has led to creation of a third mission in most countries which requires universities to maximize their community and regional contributions. These contributions need clear management structures if they are to be effective and often entail the recruitment of *animateurs* from the private sector whose first responsibility is to their customers and not to the traditional mores of teaching and research (Goddard and Chatterton, 1999). Their instincts and role are different to those of the academic community.

Financial stringency, combined with marketization, has also been a forcing house for change. Rapid growth in student numbers, financed at marginal cost levels, and the impact of increasing marketization, together with a greater concentration of research monies into fewer universities have led to a steady worsening of staff–student ratios to levels which would have been considered unacceptable in university systems 20 years before, and to repeated rounds of academic restructuring where student demand has fallen. In the early 1980s in the UK, the period of the sharpest cuts imposed by the Thatcher government, questions were raised by, for example, the chairman of Unilever whether universities had “the organisational structures nor as yet the management skills” to cope with change of this nature (Durham, 1982): the Jarratt Report argued that vice-chancellors should be recognized as chief executives and that governing bodies should assert themselves over senates which were endemically liable to resist change and to act out of natural conservatism (CVCP, 1985), propositions that were given

legal validity in the constitutions of the new universities in 1992; the Lambert Report condemned management in some universities as “slow moving, bureaucratic and risk averse” and commended “dynamic management in an environment that cannot wait for the next committee meeting” (Lambert, 2003).

Managerialism and Collegiality in Practice

Thus, in most higher education systems in the UK, size, increasing competition, accountability, the demands of a broader mission, and financial stringency have in combination changed the management task in universities, not just at the institutional level but in the basic academic units of the faculties and academic departments. The solution for many university systems has been the adoption of a much more executive style of management, and a great deal less reliance on collegial decision making. Consensual appointment processes to fill senior management positions were changed on the argument put forward by one UK vice-chancellor that:

academics are not going to vote for a manager who advocates cutting their programmes. The result is that only those people with no plans or those who promise everyone whatever they wish to hear are elected. University managers, including deans and heads of departments should not be elected but should be selected for their management ability. This will give them dignity and authority. Also managers who are hired can be fired – a key element of accountability. (Schwarz, 2003)

In matters of governance too there was pressure, from the Jarratt Report onward to make universities more business like. In the UK, the Committee of University Chairmen, following a recommendation of the Lambert Report, adopted a code of governance for universities which was modeled on the Combined Code of Governance required of major companies by the Financial Services Authority. Many continental European university systems, which had previously operated without any lay (i.e., external members drawn largely from industry and the professions), established boards of governors broadly on the Anglo-Saxon model (Braun and Merrien, 1999; Kehm and Lanzendorf, 2006). Perhaps at the extreme end of this process, a survey of university governance in the UK described DeMontfort University, a post-1992 institution, as having:

a small Board of Governors working closely with a Chief Executive ... who really is the Chief Executive and has considerable delegated powers; the Chief Executive runs the institution with the Senior Executive including four Pro-Vice-Chancellors, two Associate Vice-Chancellors, the Director of Finance, the academic Registrar and the Director of Personnel”. (North Report, University of Oxford, Commission of Inquiry, 1997)

Both the Schwarz statement and the description of DeMontfort reflect managerialist approaches as to how universities should be run. To emphasize how far this represents a shift of opinion, it is worth comparing them with a quotation from a senior and successful vice-chancellor of a major UK civic university from the 1960s:

It is clear that the optimal distribution of responsibility for academic management is not the most economical of time and effort. A much smaller and more authoritarian oligarchy with a tight hierarchy of subordinates could reduce the size and complexity of the committee system. It would, however, be unacceptable for the valid reason that under it academic freedom would be restricted and academics would carry out research and teaching less well. The academic does not produce best performance to order. (Aitken, 1966)

It is worth asking whether the values to which Aitken subscribes still carry force. We can attempt to answer this in a number of ways. The first is to try to evaluate university performance. Here, the evidence looks pretty clear that whether we use the international (Shanghai Jiao Tong or the *Times Higher Education Supplement*), those universities which are toward the top end of the ranking lists place a high priority on participative management structures, while those at the bottom, such as DeMontfort, do not. It could be argued that in the specifically UK league tables the inherited wealth of Cambridge and Oxford overbalances the evidences were not York and Warwick, both 1960s universities, and both strongly collegial in management style not so high up (Shattock, 2003a). Similar evidence is provided by the USA with highly collegial universities at the top of the rankings and vest-pocket institutions at the bottom. Ehrenberg, writing about the top US universities says that: "It is hard to think of any decision made by the university in which faculty members do not feel they have a legitimate interest" (Ehrenberg, 2002). The decision by the president of Harvard to resign in the face of a critical vote by the College of Arts and Sciences, paralleled in the UK by the immediate resignation of the provost (vice-chancellor) of University College, London, on receipt of a critical round robin signed by 40 professors, provides illustration that in leading research-intensive universities the head of an institution feels himself/herself to be heavily dependent on the confidence of the academic community.

Universities and Business-Like Organizations

University management is often judged by external commentators through analogies with business. However, even by this yardstick the evidence we have is that collegiality

works well and particularly in crisis situations. In spite of the strictures of the chairman of Unilever (quoted above) no UK university went bankrupt, or even looked like going bankrupt in the period of severe cuts in 1981–1984. The university which faced the heaviest cuts (44%) dealt with them by asking every professor to tender his/her resignation to the vice-chancellor (not a single one demurred) and when the restructuring was agreed, by the senate as well as by the council, some letters were handed back to their authors and some were accepted. The university even bounced back, and created space to appoint some young new staff. Hardy, in case studies of how Canadian universities handled similar sharp budgetary reductions in the 1980s, identified different institutions' approaches as bureaucratic, technocratic, political, and collegial; she concluded that collegiality remained the most effective mechanism to manage competing pressures and was "more likely to encourage creativity and innovation" (Hardy, 1996).

Creativity and innovation are precisely the sought-after qualities by most successful companies, which is why leading business researchers argue that strategy is best formed by bottom-up rather than top-down methods (Ghoshal and Bartlett, 1993), that the role of senior management is more a retroactive legitimizer than in providing charismatic leadership (Quin, 1985) or that the task of top management "is less to spot and solve problems than to create an organisation that can spot and solve its own problems" (Hayes, 1985: 116). Professional and partnership organizations similarly work more on the basis of consensus than direction: Clifford Chance, the world's largest law firm, manages some of its corporate decision making on the basis of referenda among its 5000 partners worldwide; McKinsey's, the world's top management consultants, elect their top post by a ballot among the partners. The belief that business or accountancy or law firms operate best on a hierarchical managerialist basis is a notion that dates back to the age of manufacture rather than to the knowledge industries of the present. Kets de Vries notes that trust is an important element in leadership and "with trust comes candour, the willingness of people to speak their minds. When people are reluctant to discuss their ideas and thoughts openly realism disappears and the quality of decision-making deteriorates" (Ket de Vries, 2002).

Collegiality and Organizational Culture

Burgan, basing her argument on Drucker's concept of the place of 'knowledge workers' in the postindustrial society organization, suggests that:

Sophisticated, independent minded and highly trained workers require modes of participation within organisations if the benefits of their initiative and imagination are

to be realised. Further, the complexity of the organisations requires more rather than less discussion from within ... Thus decisions that seem right in the short run may never be assessed in the light of long range developments. (Burgan, 2006)

An organizational study of Bletchley Park, the wartime base for 'Enigma' which was crucial to winning World War II concludes, contradicting the argument that it was 'anarchic and amateurish,' that:

whilst the multiple lines of command and employing institutions and the highly compartmentalised organisation were in some sense problems, they gave rise to meritocratic, problem solving groups with a strong sense of group identification. Intriguingly, Bletchley Park seems to conform to many of the tenets of what are nowadays called 'knowledge intensive' or 'post-bureaucratic organisations'. (Grey and Sturdy, 2007)

This could just as easily be a description of a highly ranked twenty-first century university.

If managerialism is ineffective in managing institutions whose core business is teaching and research, where creativity and innovation are at a premium, does collegiality offer an alternative? The answer lies in the organizational culture of the particular university. In European countries where universities have operated as part of state bureaucracies, such as France and Spain, collegiality is often a force for conservatism and resistance to change. However, in universities which recognize the need to compete for talent, research rewards, and for the best students, management which combines the virtues of collegiality and the politics of consent with the disciplines of timeliness in decision making can create an organizational culture that is true to the institution's key functions. Clark, for example, sees no contradiction between collegiality and entrepreneurialism, and indeed would argue from his case studies that university transformation will occur only when "managerial and faculty values become intertwined and then expressed in daily operating procedures" (Clark, 1998). In modern conditions, however, senates/academic boards and governing bodies meet too infrequently and are usually too large and too inexperienced to manage institutions effectively. Many institutions have therefore created what Clark has called "a strengthened steering core"; this may either be a senior management team – that is, a group of full-time managers answerable to the chief executive (common in universities which have gone down the managerialist route) – or a body comprising both academics and administrators which is constitutionally answerable to senate and the governing body. A critical element in this latter model is whether the deans are members of the group; if they are, academic management becomes a central concern and the deans have the task of both reflecting faculty opinion to the centre and representing central

decisions to their faculties; if they are not, the group's decisions tend to become more centrist and managerial, and less academic, and the deans tend to become merely advocates for their faculties against central management, rather than key participants in institutional policymaking. The interrelationships between such a steering group and the senate and the governing body on the one hand, and the faculties and/or academic departments, on the other, will define the organizational culture of the institution.

Where does the vice-chancellor/chief executive's role sit in this new context of university management? Leadership can be expressed in various styles, but Collins, examining a group of companies which had radically improved their performance found that their chief executives were not charismatic leaders in the conventional sense and had no public profile at all (Collins, 2001). Goodall has found that: "The higher the global ranking of a university the more likely it is that the citations of its president will also be high ... better universities appoint better researchers to lead them" (Goodall, 2006). This ensures that such presidents possess what Ramsden sees as a key university leadership skill: "an understanding of how academics work and an ability to enter into their world" (Ramsden, 1998). Leadership in collegially managed universities will be distributed and not concentrated in a single person and the task of a vice-chancellor will be to build robust structures and strong teams not to lead the charge (Shattock, 2003b). "It's about listening, then deciding and then leading forward" and not about managerial direction and confrontation (Midgley and Macleod, 2003). Indeed, the ideal is perhaps summed up in the Chinese Taoist saying:

The wicked leader is he who the people despise.
The good leader is he who the people revere.
The great leader is he who the people say, 'We did it ourselves'. (Lao Zhi, quoted in Yihong, 2004)

Kogan and Hanney were equivocal about how far managerialism had become the dominant culture in UK universities but quoted one system leader as saying that "scholarly values are being to some extent suppressed in favour of business attitudes and criteria and ... if we're not careful, that will mean the downfall of something that is rather precious" (Kogan and Hanney, 2000). Marginson and Conside writing about the Australian scene were more critical:

In recent years there has been a concentration of decision-making at the point of institutional management and leadership. Certain decisions once made by national or state governments, about research deployment, for example, have been transferred to the universities themselves. Other decisions once made by academic units are now determined from above by professional managers and technicians. Many see this concentration of nodal powers as overdue, as essential to the effective running of universities in the manner of government departments or business

firms. Others see it as the primary cause of what they perceive as a crisis of university purposes and values. (Marginson and Considine, 2000c)

There can be little objection to the idea that, in the modern age, universities must be managed but by whom and in what manner are deeply contested questions. Managerialism, if it is defined as top-down nonconsensual management, is not an effective approach to running universities but collegiality without a corporate commitment to academic and other forms of success is not a reliable substitute.

Bibliography

- Aitken, R. (1966). *Administration of a University*, p 77. London: University of London Press.
- Braun, D. and Merrien, F. -X. (eds.) (1999). *Towards a New Model of Governance for Universities? A Comparative View*. London: Jessica Kingsley.
- Burgan, M. (2006). *Whatever Happened to the Faculty?* p 108. Baltimore, MD: Johns Hopkins University Press.
- Clark, B. R. (1998). *Creating Entrepreneurial Universities Organisational Pathways of Transformation*, p 137. Oxford: Pergamon.
- Collins, J. C. (2001). *Good to Great*. London: Century.
- Commission of Inquiry (1997). *Commission of Inquiry Supplementary Volume (The North Report)*, p 90. Oxford: Oxford University Press.
- Committee of Vice-Chancellors and Principals (CVCP) (1985). *Report of the Steering Committee on Efficiency Studies (The Jarratt Report)*, p 24. London: CVCP.
- Deem, R. and Brehony, K. J. (2005). Management as ideology: The case of 'new managerialism' in higher education. *Oxford Review of Education* 31(2), 217–235.
- Durham, K. (1982). Foreword. In Oldham, G. (ed.) *The Future of Research. Leverhulme Programme of Study into the Future of Higher Education*. Guildford: Society for Research in Higher Education.
- Ehrenberg, R. G. (2002). *Tuition Rising: Why College Costs So Much*, p 22. Cambridge, MA: Harvard University Press.
- Ghoshal, S. and Bartlett, C. (1993). *The Individualised Corporation*. London: Heinemann.
- Goddard, J. B. and Chatterton, C. (1999). Regional development agencies and the knowledge economy: Harnessing the potential of universities. *Environment and Planning C: Government and Policy* 17(6), 665–699.
- Goodall, A. (2006). The leaders of the world's top 100 universities. *International Higher Education* 42, 3–4.
- Grey, C. and Sturdy, A. (2007). *The Organisation of Bletchley Park 1939–45*. Unpublished Report.
- Hardy, C. (1996). *The Politics of Collegiality*, p 183. Montreal, QC: McGill-Queen's University Press.
- Hayes, R. H. (1985). Strategic planning – Forward or reverse? Are corporate planners going about things the wrong way round? *Harvard Business Review* 63, 111–119.
- Kehm, B. M. and Lanzendorf, U. (eds.) (2006). *Reforming University Governance*. Bonn: Lemmens.
- Ket de Vries, M. F. R. (2002). Beyond Sloan: Trust is at the heart of corporate values. *Financial Times* 2(October).
- Kogan, M. and Hanney, S. (2000). *Reforming Higher Education*, p 186; 200. London: Jessica Kingsley.
- Lambert, R. (2003). *Lambert Review of Business – University Collaboration, Final Report*, paras 7.2; 7.6. London: HMSO.
- Marginson, S. and Considine, M. (2000a). *The Enterprise University*, p 3. Cambridge: Cambridge University Press.
- Marginson, S. and Considine, M. (2000b). *The Enterprise University*, p 4. Cambridge: Cambridge University Press.
- Marginson, S. and Considine, M. (2000c). *The Enterprise University*, p 8. Cambridge: Cambridge University Press.
- Midgley, S. and Macleod, D. (2003). Vice-squad. *Guardian Education* 1(April).
- Pollitt, C. and Boukaert, G. (2000). *Public Management Reform – A Comparative Analysis*. Oxford: Oxford University Press.
- Quin, J. B. (1985). Managing innovation: Controlled chaos. *Harvard Business Review* 63(3), 73–84.
- Ramsden, P. (1998). *Learning to Lead in Higher Education*, pp 126–127. London: Routledge.
- Schwarz, S. (2003). Can you pinpoint your boss? *Times Higher Education Supplement* 7(March).
- Shattock, M. L. (2003a). *Managing Successful Universities*, p 9. Maidenhead: Open University Press.
- Shattock, M. L. (2003b). *Managing Successful Universities*, p 92. Maidenhead: Open University Press.
- Slaughter, S. and Leslie, L. L. (1997). *Academic Capitalism, Politics, Policies and the Entrepreneurial University*. Baltimore, MD: Johns Hopkins University Press.
- Tapper, T. (2006). *The Governance of British Higher Education – The Struggle for Policy Control*, p 23. Dordrecht: Springer.
- Yihong, F. (2004). *From Integrative World View to Holistic Education Theory and Practice*, p 32. Chengdu: Southwest Jiaotong University Press.

Further Reading

- Deem, R. (1998). New managerialism in higher education – The management of performance and cultures in universities. *International Studies in the Sociology of Education* 8(1), 47–70.
- Deem, R. (2004). The knowledge worker, the manager-academic and the contemporary UK university: New and old forms of public management. *Financial Accountability and Management* 20(2), 107–128.
- Deem, R. and Johnson, R. J. (2000). Managerialism and university managers: Building new academic communities or conducting old ones? In McNay, I. (ed.) *Higher Education and Its Communities*, pp 65–84. Buckingham: Open University Press.
- Exworthy, M. and Halford, S. (eds.) (1999). *Professionals and the New Managerialism in the Public Sector*, p 6. Buckingham: Open University Press.
- Ferlie, E., Ashburner, L., Fitzgerald, L., and Pettigrew, A. (1996). *The New Public Management in Action*. Oxford: Oxford University Press.
- McNay, I. (1999). Changing cultures in UK higher education: The state as corporate market bureaucracy in the emergent academic enterprise. In Braun, D. and Merrien, F. -X. (eds.) *Towards a New Model of Governance for Universities?* pp 34–58. London: Jessica Kingsley.
- National Committee of Inquiry into Higher Education (NCIHE) *Higher Education in the Learning Society Supplementary Report 3 Academic Staff in Higher Education: Their Experiences and Expectations (the Dearing Report)*, pp 121–124. London: HMSO.
- Neave, G. (1998). The evaluative state reconsidered. *European Journal of Education* 33(3), 265–284.
- Pollitt, C. (1993). *Managerialism and the Public Services. The Anglo American Experience*. Oxford: Basil Blackwell.
- Reed, M. (2002). New managerialism, professional power and organisational governance: A review and assessment. In Amaral, A., Jones, G. A., and Karseth, B. (eds.) *Governing Higher Education: National Perspectives on Institutional Governance*, pp 163–185. Dordrecht: Kluwer.
- Shattock, M. L. (2006). *Managing Good Governance in Higher Education*. Maidenhead: Open University Press.

Strategic Planning in Higher Education

L Zechlin, University of Duisburg-Essen, Duisburg, Germany

© 2010 Elsevier Ltd. All rights reserved.

For the last third of the twentieth century, higher education institutions around the world have been struggling under two different types of pressure: the first is financial, as the provision of public funds has been decreasing. The second refers to the political, economic, and social demands on higher education, which have been increasing. Meanwhile, state governments are withdrawing from the direct regulation of and responsibility for the higher education sector. Under the paradigms of new public management and public governance, higher education institutions have been endowed with greater autonomy for a new beginning. It is now up to the institutions to bridge the gap between increasing internal and external demands for societal services and the limited available resources by improving the efficiency and effectiveness of their work. The task requires professional management for all types of higher education institutions, from the Humboldtian universities, with their formerly state-controlled budgets, personnel, and organizations which are forced to create first their own management system, to the Anglo-Saxon and US-American universities that have to professionalize their management system. In terms of the triangle of coordination used by Clark (1983) to classify the national higher education systems according to the major influences of the state, market, or academic oligarchies, the global trend clearly indicates the increasing importance of the market and competition. Consequently, the model of the entrepreneurial university has emerged, which then necessitated the creation of methods and instruments appropriate for institutional leadership in the higher education sector. In this context, the issue of strategic planning takes on new importance for higher education institutions around the world.

The Basic Model for Strategic Planning

The basic model for strategic planning consists of several consecutive steps, beginning with a strategic analysis in which changes in an institution's environment and the resulting external demands are mapped out against the institution's internal potential. On the basis of the analysis, strategies with long-term goals are formulated and then the actions necessary for realizing the strategies are planned. These three steps, which are often accompanied by expertise from external advisors, fall within the narrow field of planning. The predetermined actions are then put into practice and lead to results on the basis of which the

actions, strategies, and analysis can be evaluated (**Figure 1**). Unexpected deviations come about either because of mistakes in the plans (inexact analysis, insufficient strategy derived from the analysis, or actions for change undertaken without sufficient reference to the strategy) or from mistakes in implementation. The main focus remains on the first three steps of the basic model: the rational, linear linkages between analysis, strategy for and planning of action for change. According to the original proponent of this approach, Ansoff (1965), planning should be based on expertise; mistakes in planning must therefore be rectified through more, or better, expertise.

Goals and Performance Areas

Most strategic plans in higher education are based on this primary model. They usually describe goals to be achieved across three distinct, hierarchical levels. The top level, for the normative management, contains statements referring to the long-term social gains to which higher education should contribute (vision and mission), and the associated underlying values by which to proceed. The middle level, the strategic-management area, includes mid- to long-term goals and objectives, as well as the strategies to be followed in order to reach them. Finally, the operative management area is found at the third level, where the actual actions for change (projects) are taken in order to accomplish the plans within the 5–10-year reference period.

A similar hierarchy, although with slightly different terminology, is characteristic of the planning logic behind the new public management approach (**Figure 2**). At the upper level is a strategic plan consisting of long-term goals for impact or outcomes. The higher education institution uses the plan to enumerate the kind of influence it aims to have on its social and economic environment; for example, contributing as a knowledge organization to the improvement of economic prosperity, political democracy, and social cohesion. Outreach to the community means that products resulting from teaching and research are conveyed beyond the institution to the surrounding environment. The strategic plan entails performance or output goals specifically for this purpose. In the area of teaching and learning, for example, the plan describes graduates in terms of quantity (e.g., the number in each study field) and quality (e.g., qualifications related to subject area, as well as methodological and social skills; international profile; and gender balance). For the areas of

research and knowledge transfer, the plan could refer to the number of patents, publications, citation frequency, etc. In end effect, the operative-management level steers the process intended to generate the preidentified desired outcomes or results.

In addition to the hierarchy of levels of objectives, there is also further differentiation between various performance areas. The purpose of higher education institutions as knowledge organizations is to make accessible to the public the knowledge existing within society and newly created through research. Higher education institutions accomplish this through teaching and learning as the means to preserve and convey existing knowledge, through research as the means to generate new knowledge and to prepare the next generation of scientists for further knowledge building. Most universities in UK, US, NZ, Australian, etc. would add to research and teaching/learning, service to or engagement with the community. Higher education institutions are not self-sufficient ivory towers; in fact, they exist precisely because of their impact on the external environment. These externally oriented performance areas are therefore essential to every higher education institution. In economic terms, they are the main business areas where higher education institutions define their policies in order to remain competitive within

the sector. It is important to distinguish between the externally oriented key processes and the internal services that the university provides in order to support these processes. In particular, such services include the make up of the internal financing system, as well as staff recruitment, information services, student services, technology transfer offices, and of course the strategic-planning system. The purpose of the internal services is to provide for internal clients/stakeholders; however, they remain closely related to the university's key processes and are certainly no less important than the externally oriented performance areas. In fact, the opposite can be argued: the performance capacity of an entrepreneurial university is more and more dependent on its internal management system, which is responsible – through appointment policies, for example – for the quality of the externally oriented key processes.

In order to reconcile all the various different demands placed on higher education today, strategic plans (see **Figure 3**) are often complicated undertakings. Normally, they consist of a general vision or mission statement followed by descriptions of the current situation in the respective area, strategic objectives, and the intended course of action for change. Many higher education institutions have published their strategic plans either in printed format or on the Internet, for example, all Australian universities, Radboud University Nijmegen (Netherlands), University of Graz (Austria), or University of Wisconsin-Madison (USA).

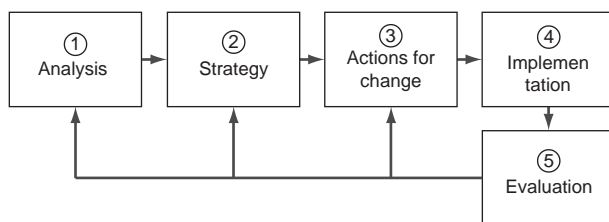


Figure 1 Planning cycle.

Changing the Basic Model

The basic model for rational planning has come under heavy criticism for years, primarily because it places so

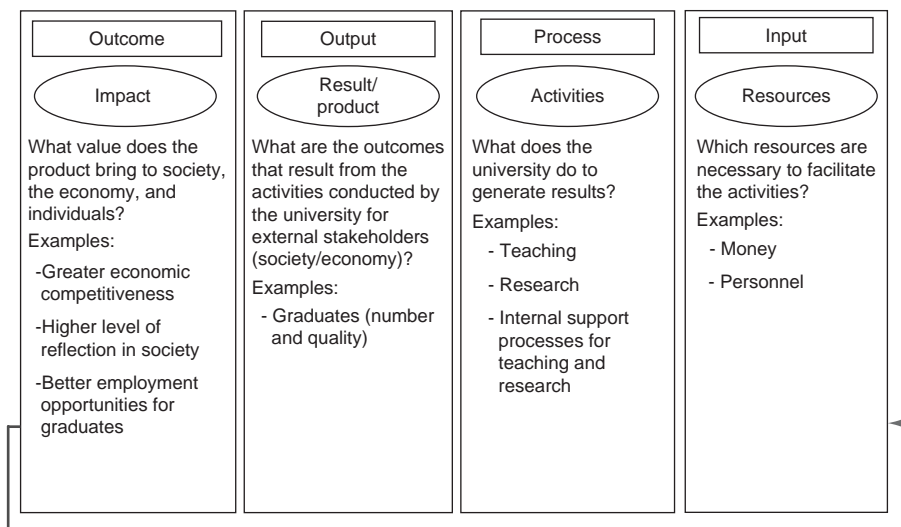


Figure 2 New public management.

1. University mission statement, normative objectives
2. Research
 - a. Situation analysis
 - b. Strategic goals
 - c. Projects and operational goals
3. Development of junior academics (a. – c.)
4. Teaching and learning (a. – c.)
5. Engagement with the community (a. – c.)
6. Budget strategy (a. – c.)
7. Infrastructure (a. – c.)
8. Etc.

Figure 3 Outline of a strategic plan.

much emphasis on planning while neglecting issues of implementation – the very issues that usually come to the fore during the execution phase of the plans. An analysis steered by experts and implemented by management, the critics claim, assumes a division between thought and action that is not consonant with real life. Planning does not happen in isolation, but should be based on actions and deliberate reflection. In fact, empirical studies in management research have shown that the planning model did not determine the strategic performance of many successful enterprises. Instead, it was shown that the most successful enterprises remain extremely flexible and retain a high level of responsiveness with respect to the often-unpredictable external environment. What is more, they place great importance on experience and the implicit knowledge accumulated within the organization, and they rely on their ability to improvise.

There are numerous reasons for why more flexible organizations are the most successful: The outside world is so dynamic that only limited predictions can be made about what might happen – which renders long-term planning rather tenuous. The external environment is an elusive object for a strategic analysis because it is also determined by strategic players who analyze things from their own perspectives and change their behavior accordingly. Insights from systems theory and strategic theory indicate that the reciprocal observations from various analysts and their interaction make it necessary to adopt a high level of flexibility, and that alternating strategies need adjustments depending on the situation. It is also often the case that the actual members of an organization have quicker and more direct access to relevant information on changes in strategies by their competitors through their own external contacts than experts can provide for with their analyses. In order for this information source to be used to the best advantage, the available (but usually implicit) knowledge within the organization has to be mobilized to flow directly into the formation of strategically determined behavior – not only during the implementation phase of a predetermined plan. All of which leads to the point that the internal environment of an institution is equally as important as the external environment. It is also equally complex and can

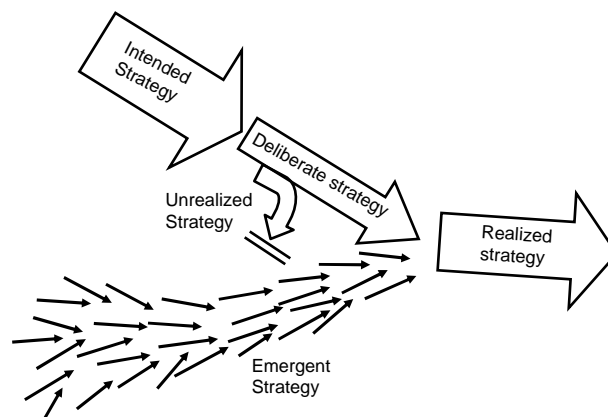


Figure 4 Strategy deliberate and emergent (Mintzberg).

be characterized through various subsystems, perspectives, and usually its own brand of logic as well. Under such circumstances, patterns can emerge that would be recognized retroactively for their inherent logic and possibly serve as organizational plans for the future; however, they do not follow the rational-linear method of planning based on expert analysis followed by separate implementation (specifically Schreyögg, 1999, 2002).

This theory of the breakdown in planning is particularly well developed in Henry Mintzberg's work. Mintzberg defines intentions that become implemented as deliberate strategies. Meanwhile, plans that were identified during the initial phases but remained unrealized must be left aside because of the focus on the first three steps in planning (see Figure 1). He further distinguishes between deliberate strategies and emergent strategies that develop as a result of the system. Realized strategies are based on the combined sum of all of these approaches (Figure 4).

In his book *Strategy Safari* (1999), Mintzberg used Honda's conquest of the American motorcycle market as a good example of emergent strategies. The planned strategy called for the production of heavy machines, but successful penetration of the market only became possible when the players involved happened to notice that the greater demand was in fact for small machines. Mintzberg goes on to cite an impressive study covering 100 of the 500 fastest growing American businesses. According to the study's findings, 41% of the companies had absolutely no business plan at all, 26% worked from a sketch scribbled on a piece of paper, 5% just used a financing plan, while only 28% had a fully formulated business plan.

It is important to note that higher education institutions are special kinds of organizations that are fundamentally different from economic enterprises. Similar to accounting and consulting companies, hospitals, or schools, higher education institutions are knowledge organizations, otherwise known as expert organizations or professional organizations. Knowledge, their most important capital,

does not belong to the organization but remains in the hands of the experts. The experts are the actual owners of the strategically most crucial resource – the very resource on which the university's performance and reputation depend. Even the assessment of the quality of this knowledge can only be conducted by experts, most of who are affiliated with other universities, that is the scientific community. In organization theory, universities are among the loosely coupled systems (Weick, 1976), and their fundamental strengths come directly out of the unplanned, curiosity-driven development of their own systems. Hierarchical strategies for steering such organizations risk being less than effective or can even – in the worst-case scenario – destroy productive-development potential.

There are very few empirical studies on the forms and effects of strategic planning in higher education. However, it appears safe to assume that the universities that are successfully positioned with regard to the competitive market have made good use of the findings in management research on emergent strategies and of the available information on the particularities of professional organizations. These models indicate that leadership should place great value on observation, communication, flexibility, and reflection with regard to changes within and outside the university. Leadership has to be able to cope with ambivalence and contradiction without neglecting the necessity of a strategic orientation for the organization. Leadership duties do not entail prescribing or pre-setting a strategy as much as they refer to the steering of a process framing strategic change. This process is not necessarily a matter of imposing normative, strategic, and operative goals upon different levels of hierarchy and thereby executing and controlling a strategic program that was devised in this manner. It is actually more about developing a rough strategic framework that promotes a common orientation for the entire institution. The framework encourages mutually beneficial behavior, but also allows possibilities for self-determined actions by organizational subsystems (departments, institutes, etc.) and their individual members, who can use their respective specific expertise productively within the semi-autonomous units. In this context, the basic model that included situation analysis, targeting goals, actions for change, and evaluation should in fact be preserved. However, the reciprocity between thinking and action will be reinforced through frequent and repeated loops of reflection on the process. This kind of development is visible in the literature as a new-found emphasis on strategic management rather than strategic planning.

Different Perspectives on Strategy

Thus, there is an abundance of different approaches to strategic planning. In some cases, they intersect with each

other or complement each other, in other cases they cancel each other out. In *Strategy Safari* (1999), Mintzberg differentiates between prescriptive approaches (what he calls the design school, planning school, and positioning school), descriptive methods where certain aspects come to the fore (entrepreneurial school, cognitive school, learning school, cultural school, power school, and environmental school), and the configuration school, which combines these approaches. Taking the perspective of organizational development, Morgan (1986) identifies different organizational pictures depicted by strategy developers, leading to different methods of procedure. They portray organizations as a machine, as an organism, as a brain, as culture, as a political system, as a psychological prison, or as an instrument of power. It is impossible to develop an abstract, standard formula for determining the direction a higher education institution should follow in its strategic development. The answer depends on the institution's identity, on the current context and environment in which it operates, and of course on its various potentials. Deciding which methodological approach to take is already the first step in developing a strategy that must be carried forward by the institution and in which it is also subject to the interplay between intentionally planned and emergent elements.

Although there is no single, right approach to strategic planning, certain basic types can be identified that can function (more or less) singly or in combination with each other, depending on the context, timing, and current status of the organization. Whittington (2001) makes a simple but definite distinction that can easily be adapted to the higher education system by asking two questions: The first is regarding the objectives of the strategy (What is strategy for?) and addresses whether the targeted outcomes are aligned one-dimensionally toward maximizing profits or if multiple, diverse objectives can be targeted. The second refers to the processes of developing strategy (How is strategy done?), with which Whittington distinguishes between planned strategies and emergent developing strategies. Meanwhile, this taxonomy can be combined with the one adopted by Nagel and Wimmer (2002) in which the formulation of strategic objectives is identified according to whether they are developed by the leadership or external experts, or whether they emerge out of the system itself. This approach also distinguishes between processes developed implicitly or explicitly. The combined methodologies lead to a four-field matrix, which can be used as an analytical grid and applied to the higher education sector in order to produce the four distinct approaches to strategic development in higher education (Figure 5).

The upper left quadrant shows the classic approach to strategy formulation and implementation, in which the institutional leadership determines the objectives and designs the necessary implementation processes. Such an approach includes, for example, fundamental reorganizing

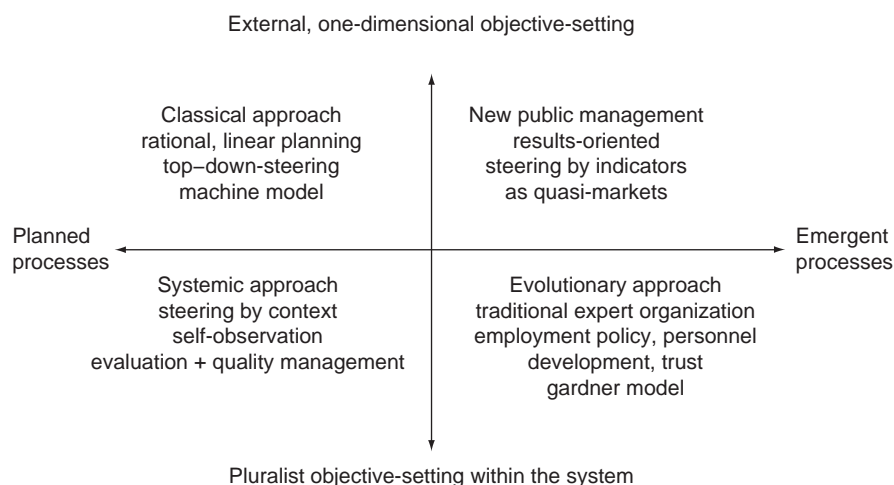


Figure 5 Strategic fields in higher education development.

processes, such as consolidating separate departments into larger faculties or implementing mergers that rely on the central authority to take responsibility for decision making within a given timeframe. The upper right quadrant also shows an approach that uses objectives set through external input, but here the conception and implementation of the requisite transfer measures are undertaken within the system itself. This is the main activity field for new public management, which as a consequence of its product logic is focused on steering by output. It thus requires results, such as increased fundraising, the production of more graduates, or a higher number of doctorates; however, the way in which these results are achieved is determined by the system itself. The lower right quadrant shows the evolutionary approach that aligns with the paradigm of the expert organization. This is the management model of the classic Humboldtian university; it focuses on attracting good scientists who cannot be steered but who are entrusted by the organization to do their work within the recognized limitations of the institution. The lower left quadrant shows the systemic approach of higher education development. This emphasizes processes organized through self-observation of the system, from which conclusions or consequences are then drawn. The main fields of activity here are evaluation and quality management.

It is important that the matrix not be mistaken as a grid for rating good or bad approaches to strategy. All approaches are legitimate and can work. The adoption of any approach can be justifiable according to the specific situation, time constraints for decision making, culture and life cycle of the higher education institution, and the purpose of the decision. There is no right approach to strategic planning; there are many approaches to be considered for each context. Strategic development occurs in each of these fields, accomplished through collective effort by the players in each field. What follows is then a model that aligns with Mintzberg's (1999) ideas on

configuration. This is where the issues of leadership and the competencies of the leadership team and their relationships to objective, rational planning come to the fore. People in leadership positions have to have access to an entire array of approaches and methodology for strategic planning and development, and they have to be able to apply them together appropriately. Then leadership has less to do with the conceptualization of the strategy, than with the conceptualization of processes through which strategy is developed as an achievement of the system. The entrepreneurial university also requires just this kind of entrepreneurial thinking and acting leadership team.

When the basic model for strategic planning (the sequence of analysis, setting objectives, actions for change, and evaluation) is decentralized through such flexibility in the objectives, methodology, and participants, it remains both meaningful and, for a well-ordered strategic process, even indispensable. It shifts from being a linear progression and becomes a reflective, learning cycle that feeds back into the organization (Argyris and Schön, 1996), where the decisions that are made remain alterable and new options are opened. In order to allow the cycle of reflection to work, however, the university's loosely linked experts and subunits all have to be involved with the development, implementation, and revision of the strategy.

Methods and Instruments

Certain methods and instruments have emerged in strategic management as particularly useful for those involved in the four steps of analysis, strategy formulation, actions for change, and evaluation. Of course, such instruments can generate certain routines that run the risk of taking on a life of their own; however, they also contribute to anchoring the change process to the very structure of the institution rather than allowing it to remain merely as an agenda of a

certain group of individuals. This is necessary in order to ensure stability within the process of reform. Three of the instruments that are relevant to the fields of analysis and objective setting are described briefly below.

Strengths, Weaknesses, Opportunities, and Threats Analysis

Strengths, weaknesses, opportunities, and threats (SWOT) analyses are instruments commonly used in higher education today. The intention behind the tool is to compare the institution's internal potential, or its current strengths and weaknesses, with the opportunities and threats that emerge from an analysis of the external environment. The comparison reveals what goals and actions for change can be developed. For example, in most continental European higher education systems, universities display a relatively strong homogeneity. In the coming years, the trends toward entrepreneurialism and competition could result in greater diversity of institutions within the higher education sector. They may profile themselves as research universities, access-providing institutions targeting broader population groups, or as higher education institutions that specialize in certain areas of research but otherwise hone their strengths in teaching and learning. SWOT analyses take the external context into consideration (What are the strengths and weaknesses of the most important competing universities and colleges? What are the foreseeable social needs?) in order to assess separate areas of the institution and determine if they are more appropriately positioned for a profile in research or teaching. The results of the analysis then serve as the foundation for creating a competitive profile and for the institution's strategic positioning in a more diverse higher education landscape.

Working on the assumption that the people who know most about an institution's development potential are the actual members of the respective institution, workshops, group presentations, and moderated podium discussions should be convened in order to make the implicit knowledge explicit (Nonaka and Takeuchi, 1995) and to effectively compile all the necessary information. In order to assemble information on developments in the university's external environment, it is best to draw on expert panels, surveys, and other sources. When the SWOT analysis is condensed into strategic objectives, they should be discussed in top-down/bottom-up planning sessions with the various subunits, departments, etc. before they are adopted at the central level as binding goals for strategic development. In this way, the entire exercise should become an institution-wide strategic process; it should be informed by expert knowledge and advice, but the outcomes should occur as a result of the collective work by all members of the university.

In order for the SWOT analysis to reach the desired depth, it should be oriented toward separate performance fields of an institution or faculty. These are the above-mentioned key processes of research, development of junior academics, and teaching and learning, as well as the central services (budget, staff development, information technology (IT) services, etc.). With these in mind, goals should then be formulated for the respective performance areas. Meanwhile, it is important that the objectives should not be worked out in too much detail. They should remain as a robust but nonprescriptive catalog of goals that can serve as an orienting framework for the university, not a constrictive directive.

Portfolio Analysis

Besides the more discursive, communicative approaches, a more quantitative, figure-oriented instrument can also be used to describe the actual status of an institution. One such instrument is the portfolio (**Figure 6**), which depicts separate features of the university as compared to the respective features of competing institutions. The analysis for this instrument should again be focused on the separate performance areas of research, development of junior academics, teaching and learning, etc., or the results will be too unspecific to be useful. However, unlike the SWOT-analysis, the portfolio technique is dependent on quantifiable measures, which means the various performance areas have to be converted into measurable dimensions. By using two parameters, the *x*- and *y*-axes of the portfolio for instance, it is possible to plot the relative competitive position in teaching (graduates and students per professor, graduates' success on the labor market), research (external funding, concluded promotions per professor), and preparation of junior academics (number of doctoral students and postgraduates who receive offers as professors or assistant professors at other institutions). Since this approach is mainly concerned with the positioning of the institution in a competitive market, comparative data from an institution with a similar profile are required. The data can be obtained either through benchmarking clubs founded for this purpose, or from available official statistics sources.

Internal Contract Management

When higher education institutions are above a certain size, it becomes problematic that strategy formulation is a process to be undertaken by the entire university, across various and sundry subunits. This can be a problem not only during implementation, but during the very process of creating strategy as well. Meanwhile, a purely top-down approach from the center to the various departments would be just as counterproductive as a purely

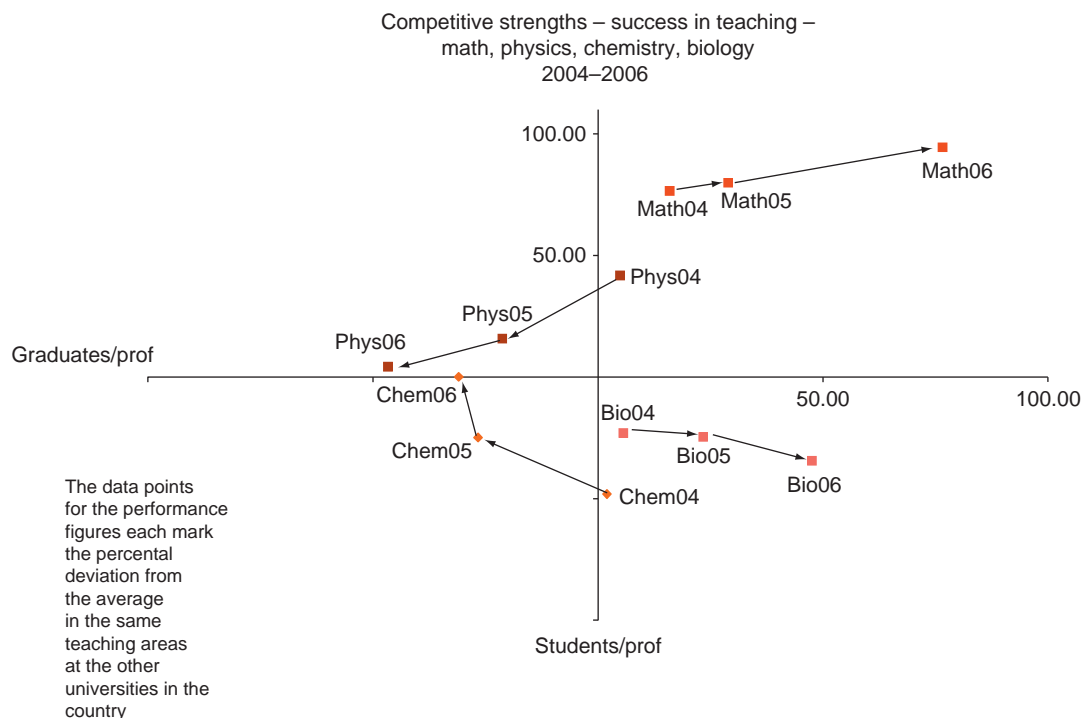


Figure 6 Portfolio to ascertain relative strengths in research. The reticle represents the average of the respective subjects measured across all universities.

Performance area: teaching and learning				
Institutional goal	Services by sub-unit/department	Success criteria	Service needs	Budgetary needs

Figure 7 Goal and performance card.

bottom-up approach in which the central powers waited for the subunits to come up with some strategies. This is where an approach taken from new public management can be useful, namely contract management. This instrument uses collective negotiations on goals and activities between the central-university level and the different departments to arrive at a mutual agreement or contract. Thus, the desired institutional profile can be established as the strategic framework while it can still be made more compatible with the profiles and strategies of the respective departments. In fact, contract management serves as an instrument for both implementation and strategy, all at the same time.

Nevertheless, without a clear system and structure, and lacking strict financial and operational controls, contract management can easily become a merely futile exercise. The goal and performance cards have often been found

useful in the past. The cards are similar to the well-known balanced scorecard in business planning (Kaplan and Norton, 1996), but are designed to reflect the key processes and service provision in higher education.

By matching them to a specific type of performance area, the university's goals can be correlated with the results of different actions for change (services) produced by a faculty or department, and determined via negotiations between the institutional central leadership and the heads of the respective subunit. In order to enter into the negotiations, faculties should prepare their own strategy papers prior to the negotiations round. The institutional goals and the goals set by the subunit are then compared, and after two or three rounds can be consolidated into concrete objectives to be realized by the subunit. In order to assess outcomes fairly, it is very important to establish the criteria by which performance and success are to be

measured upon conclusion of the contract (usually 2 years) at the very beginning in the initial agreement. Meanwhile, potential service needs or financial needs encountered by the subunits while undertaking the activities also have to be enumerated. For example, a department that has agreed to increase the number of foreign exchange students may depend on certain services provided by the student secretariat or the international office; in order to provide the services, the office may then need new software, which has to be made available by the IT department. Such examples reveal glimpses into the complex interdependence of performance objectives within a higher education institution. It quickly becomes clear that contract management helps lower the dividing line between big-picture strategy and a policy of continual change by keeping the service providers very closely involved in the process. The results of the negotiations are set down in the contract between the central level and the subunit; the important outputs that result from the agreements are essential parts of the future strategic plan.

Contract management requires a large amount of time investment. Each round of negotiations between the institutional management and the respective subunits usually requires 2–3 h. There are normally at least two, if not three, meetings of that kind per subunit. The first round is mainly used to assemble the different expectations and possibilities, and to prepare for further discussions. Some results may already be recognizable during the second round; however, a number of points will probably remain open and in need of revision, so the results can usually only be confirmed after the conclusion of the third round. In between the different discussion rounds, time is also needed for internal consensus building and bilateral clarifications. In the end, the entire process normally takes about 3 months. In the instances where the discussion rounds have been used as opportunities for mutual reflection and exchange, and when they have succeeded in making the lurking, implicit institutional knowledge manifest and useful, the productivity of the planning process has paid for itself as a major contribution toward making the university a truly integrated higher education organization.

Bibliography

- Ansoff, H. I. (1965). *Corporate Strategy*. New York: McGraw-Hill.
Argyris, C. and Schön, D. (1996). *Organizational Learning II. Theory, Method, and Practice*. Reading, MA: Addison-Wesley.

- Clark, B. R. (1983). *The Higher Education System. Academic Organization in Cross-National Perspective*. Berkeley, CA: University of California Press.
Kaplan, R. S. and Norton, D. P. (1996). *The Balanced Scorecard*. Boston, MA: Harvard Business Press.
Mintzberg, H. (1999). *Strategy Safari: A Guided Tour Through the Wilds of Strategic Management*. New York: Free Press.
Morgan, G. (1986). *Images of Organization*. Newbury Park, CA: Sage.
Nagel, R. and Wimmer, R. (2002). *Systemische Strategieentwicklung: Modelle und Instrumente für Berater und Entscheider*. Stuttgart: Klett-Cotta.
Nonaka, I. and Takeuchi, H. (1995). *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York: Oxford University Press.
Schreyögg, G. (1999). Strategisches Management – Entwicklungstendenzen und Zukunftsperspektiven. *Die Unternehmung. Schweizerische Zeitschrift für betriebswirtschaftliche Forschung und Praxis* 6, 387–407.
Schreyögg, G. (2002). Strategie folgt Struktur – Lektionen aus einem empirischen Befund für eine neue Theorie der Unternehmenssteuerung. In Böhler, H. (ed.) *Marketing-Management und Unternehmensführung*, pp 35–50. Stuttgart: Schäffer-Poeschel.
Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Sciences Quarterly* 21, 1–19.
Whittington, R. (2001). *What is Strategy – and Does It Matter?* 2nd edn. London: Thompson Learning.

Further Reading

- Birnbaum, R. (2000). *Management Fads in Higher Education: Where They Come from, What They Do, Why They Fail*. San Francisco, CA: Jossey-Bass.
Cohen, M. D., March, J. G., and Olsen, J. P. (1971). A garbage can model of organizational change. *Administrative Science Quarterly* 17, 1–25.
Keller, G. (1983). *Academic Strategy. The Management Revolution in American Higher Education*. Baltimore, MD: Johns Hopkins University Press.
Lewin, K. (1951). *Field Theory in Social Science*. New York: Harper and Row.
Lindblom, C. E. (1959). The science of muddling through. *Public Administration Review* 2, 79–88.
Lockwood, G. and Davies, J. (1985). *Universities: The Management Challenge*. Philadelphia, PA: Nfer-Nelson.
Mintzberg, H. (1994). *The Fall and Rise of Strategic Planning*. New York: Free Press.
Tabatoni, P., Davies, J., and Barblan, A. (2002). *Strategic Management and Universities' Institutional Development*. Brussels: European University Association.
Taylor, J. and De Lourdes Machado, M. (2006). Higher education leadership and management: From conflict to interdependence through strategic planning. *Tertiary Education and Management* 2, 137–160.
Taylor, J. and Miroiu, A. (2002). *Policy Making, Strategic Planning and Management of Higher Education*. Bucharest: European Centre for Higher Education.
Weick, K. E. (1995). *Sense Making in Organizations*. Thousand Oaks, CA: Sage.

The Changing Role of University Presidents, Vice-Chancellors and Rectors

M F Green and P D Eckel, American Council on Education, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

In this age of the growing knowledge economy, pressures on higher education around the world are mounting. Governments, industry, social agencies, and citizens at large are increasingly relying on colleges and universities to educate the young and old, prepare the workforce, and contribute to the nation's economic growth; make technological and scientific breakthroughs; and address society's social ills. These pressures, in turn, heighten the demands and challenges of leading and governing colleges and universities. This article focuses on the changing environment worldwide for higher education and its impact on the role and challenges of campus administrators, specifically the president, rector, and vice-chancellor, referred to collectively as institutional heads. The first section examines two major forces for change shaping institutional management, leadership, and governance: the changing relationships between institutions and their governments, and the intensified market environment. The second section discusses the effects of these forces on institutional management and governance, focusing on the move from collegial governance to managerialism and the resulting tensions. The final section explores the changing role of institutional heads, describing four key factors shaping that role and exploring the tensions arising from the role of institutional head as change agent.

The Context for Change

The role of institutional head is shaped strongly by two important and interrelated external institutional-change drivers – the policy environment and market forces. Certainly, these are not the only drivers of change – changing demographics, advances in technology, pressures for accountability, advances in knowledge, and increasing reliance on higher education to stimulate and support economic development are also factors. However, the policy environment and market forces represent two broad and powerful sets of conditions that play a major role in determining the current challenges and future directions of higher education.

The Policy Environment: The Changing Relationship with Government

Governments around the world are changing their relationships with their public universities, albeit along different

dimensions and to different degrees. In some instances, the key government agencies are at the national or federal level, such as South Africa, Japan, and England. In other instances, such as Switzerland, Canada, the US, and Germany, the state or provincial governments are the key players. Driving many of the changes is the growing philosophy of new public management (NPM), a trend in which the public agencies concerned with higher education adopt the economic market as the dominant model, which is noted by deregulation, the creation of (quasi-) markets for public services, performance-based contracting, policies that foster competition, the use of market incentives and levers, and the rise of performance indicators and auditing systems (Middlehurst, 2004; Mok and Lo, 2002; Sporn, 2006). Singh (2001: 10) summarizes the effects of these beliefs on higher education:

- the requirement of higher education to demonstrate efficiency, effectiveness, and value for money through business re-engineering drives, integration into public finance and accounting systems, external quality assurance, and other accountability frameworks;
- declining investments of public funds, costs passed to students, and the requirement to do more with less (e.g., massification of access at existing or reduced levels of funding); pressure to diversify funding, thus reducing the primary responsibility of the state for public higher education and allowing other funders to exert pressures;
- the dominance of managerial and entrepreneurial approaches resulting in running higher education institutions like income-generating business;
- the privatization of higher education encouraging competition;
- the increasing development curriculum reforms intended to appeal to employers and students as customers and clients; and
- shift of public and private funding from basic to applied research, increased emphasis on academic/industry links, greater concern with issues of intellectual property rights, and the prioritization of research for product development and commercialization.

The effects of this emergent philosophy are being experienced around the world. For example, in Finland, the 1999 development plan for education and research sought to improve quality through evaluation and enhanced competition and is characterized by a management-by-results

principle. The Danish 2003 University Act created greater institutional autonomy, including the ability of institutions to appoint external board members, and fostered increased managerialism as rectors, deans, and department heads now are appointed rather than elected (Schmidt, 2006). Swiss universities have been granted greater degrees of autonomy, provided with global budgets that have attached to them performance metrics (Schenker-Wicki and Hürlimann, 2005). Australian universities are also quickly adapting to market-driven policies (Margison and Considine, 2000) and US universities are increasingly becoming even more sophisticated market actors (Slaughter and Rhoades, 2004) in a system that has traditionally been market sensitive.

Changing Market Environment

Three major aspects of the changing market environment are prominent in shaping the course of higher education: the rising demands and opportunities of the knowledge economy, increased competition among universities, and the push and pull of globalization.

The knowledge economy

As the knowledge economy becomes more deeply established, universities find themselves, as the European Union explicitly states “at the crossroads of research, education and innovation” (Commission of the European Communities, 2003: 5). They are primary producers of advanced knowledge, which is an essential raw material of the new economy, and they play an essential role in creating both the highly educated workers and consumers on which the knowledge economy relies (Slaughter and Rhoades, 2004). The knowledge economy furthermore favors (and is willing to pay for) what universities have in abundance. Former Harvard President Derek Bok argues that regardless of other drivers that pushed US campuses to be more entrepreneurial (such as declining state support, changing public policies, or unclear academic values), “none of these stimuli would have borne such abundant fruit had it not been for the rapid growth of money-making opportunities provided by a more technologically sophisticated, knowledge-based economy” (Bok, 2003: 15).

Competition

Market activity further intensifies competition among higher education institutions within and across borders. This competition takes many forms. Institutions compete not only for public and private funding, but also for students, academic talent, prestige, and media attention. They must determine their market niche, and distinguish themselves from other institutions through branding. In addition, the competitive field has widened. For-profit universities, corporate providers, educational companies, and distance-learning institutions are now a significant part of the competitive landscape in some countries.

For example, nearly 90% of private higher education institutions in Malaysia are for profit, as are approximately two-thirds of Brazil's. In the Philippines, 47% of students attend for-profit institutions, as do 18% of Chilean students (Kinser and Levy, 2006). Furthermore, with the rise of information technology and online education, competitors are now no more than a click away, making geographic location less important. Governments in Singapore, Germany, and China, to name a few, are investing significantly to create world-class universities (Altbach, 2000; Labi, 2006a; Morhman, 2005). A major aim of the Bologna process is to create a European higher education space that will be more competitive in research as well as more attractive to foreign students.

Globalization

The influences of globalization and internationalization further shape the competitive market for higher education. Cross-border or transnational education, in which providers and programs move across national boundaries to offer their degrees (Knight, 2002), is one response to globalization whose long-term impact is as yet undetermined. To extend their reach to students in foreign countries, institutions are establishing degree programs either on their own or with partner universities, governments, foundations or private businesses. Some governments, often under the auspices of the World Trade Organization (WTO) are developing policies and initiatives that encourage (or do not prevent) foreign providers from offering degrees. Countries such as Congo, Jamaica, and Sierra Leone are actively encouraging the entry of foreign institutions (Knight, 2002). Other countries such as Singapore, Qatar, and the United Arab Emirates are going even further by providing foreign institutions with financial incentives, tax-free zones, and infrastructure (including rent-free or below-market leases) to encourage and support foreign providers (Green *et al.*, 2007). At the same time, the globalization of trade and business and rising predominance of English also help to make foreign degrees, particularly from universities in Anglophone countries, increasingly desirable.

Institutional Management and Governance: From Collegial to Managerial to Entrepreneurial Institutions

The effect of new relationships with governments, and greater market pressures have been accompanied by greater institutional autonomy and heightened need for institutions to be more managerial – that is, compete and be more businesslike, efficient, strategic, responsive to stakeholders, and entrepreneurial. Gumpert and Sporn (1999) attribute the rise of managerialism to three interdependent dynamics: resource dependency motivated by

organizational survival; institutional isomorphism motivated by legitimacy concerns; and professional authority motivated by a struggle for professional identity.

Although few scholars or practitioners dispute that managerialism is on the rise worldwide, it is important to note that there are significant differences not only among nations, but also within them, according to institutional mission and type. For-profit institutions are more corporate than nonprofits in terms of culture and management and governance practices. US community colleges and new universities in the UK (i.e., institutions that were polytechnics prior to 1992) have weaker traditions of faculty dominance in decision making than 4-year institutions in the US or old universities. In South Africa, historical institutional differences shaped the extent to which transformational initiatives engaged faculty and other stakeholders or were top-down efforts (Kulati and Moja, 2006). The significant differences among nations are widely recognized. For example, as a result of many European and Latin American countries' traditions of strong faculty bodies and elected rectors, managerialism developed later and is less pervasive than in the United States, Canada, the United Kingdom, or Australia.

Managerialism in the Collegium

Managerialism is often seen as conflicting with traditional notions of collegial governance. However, there is no single conception of managerialism; it can be more or less strident and doctrinaire, more or less adaptive and useful. Cloete and Kulati (2003) distinguish between strategic managerialism – a soft approach in which management techniques are applied to run institutions more efficiently and effectively, and unwavering entrepreneurs, who see higher education institutions as businesses, as distinct from being run like a business. The latter model is characterized by top-down decision making, disempowerment of the faculty – who are viewed as employees rather than partners – and a profound cultural shift toward corporate rather than academic values.

In contrast with managerialism, collegial governance emphasizes the traditional role that academic staff play in making key decisions about all aspects of institutional functioning and directions. It values scholarly engagement, expertise, negotiation and consensus, collaboration, shared governance and decision making, and rationality. (It is appropriate to note, however, that the rise of managerialism has caused some to romanticize the collegiate model of governance, conjuring an idealized academic world many agree never existed.)

Although some argue that traditional models of shared governance are too slow and cumbersome to enable institutions to respond to a rapidly changing environment, others maintain that the corporatization of university management is dangerous for the social purposes and

long-term well-being of higher education (Duderstadt, 2001; Burgan, 2006; Johnson and Cross, 2004; Currie and Vidovich, 1998; Kirp, 2003; Scott, 2001; Cloete and Kulati, 2003). The more abrupt the rise of market orientation in a particular country – and increase in the power vested in institutional administrators – the greater the potential for tension (Green, 1997; Edwards, 1997; Amaral *et al.*, 2003; Shattock, 2003, 2006). Thus, a key challenge of higher education leadership and governance today is to foster governance processes that are deliberative and inclusive, and at the same time responsive, agile, and opportunistic – two seemingly contradictory sets of demands. Effective governance incorporates tradition, academic values, and scholarly community as well as new ways of doing business, leveraging institutional resources, and quickly marshalling institutional strengths (Eckel, 2006).

Emerging Organizational Models

Clark (1998) proposes a model of the entrepreneurial university that enables institutions to combine elements of both managerialism and traditional academic culture (which, he asserts, cannot be ignored) so as to enable institutions to increase their capacity to respond to demands that are both increasing in number and intensifying and be more active in charting their courses. The entrepreneurial university is characterized by five elements: a strengthened steering core; an enhanced development periphery, a discretionary funding base; a stimulated heartland; and an entrepreneurial belief. The strengthened steering core consists of a strong administrative backbone throughout the entire institution at all levels. There are strong collegial connections between faculty and administrators and shared governance. The enhanced development periphery includes various outreach administrative units research centers that may be closely or loosely linked to the administrative core and the academic (heartland) departments. These units provide important links to external groups and enable the university to be more flexible and responsive. A diversified funding base enhances institutional discretion and allows the university greater independence from any single source. The traditional academic departments in the heartland of the institution must be brought into the change process. The interest in and ability of different fields and departments to embrace an entrepreneurial spirit will vary; for example, science and technology will likely move faster, it is important that habits of change be taken up across the institution. Finally, institutions will need to undergo a cultural shift to reconfigure themselves as entrepreneurial universities; this is a long-term undertaking that permits lasting change.

Australian researchers Margison and Considine (2000) are less sanguine than Clark about the ability of a fruitful blend of managerialism and academic traditions. They doubt the capacity of institutions to transform themselves

in ways that do not threaten collegial forms of decision making. Their model of the enterprise university – and all Australian universities, they assert, are to some degree enterprise universities – “joins a mixed public–private economy to a quasi-business culture and to academic traditions partly reconstituted, partly republican, and partly broken” (p. 236). There are many variations on the enterprise university, but they note that the common characteristic is a weakening of traditional academic processes.

Structural Changes and the Rise of Governing Boards

Changing demands and new ways of doing business are also driving institutional structural changes. Heightened demands for connections to external stakeholders and intensified pressures for accountability have led to the rise of governing boards around the world. A distinguishing feature of these boards is that they draw many, if not most, members from outside the institution. Their functions are generally to create ties to the local community, provide fiscal oversight, foster and support entrepreneurial activities, and ensure that institutions are accountable to the larger society. Governing boards strengthen the role of institutional leaders through their oversight of institutional strategy and management. Shattock (2003) outlines seven contributions of governing boards: technical and professional advice; taking the long view; acting as the referee for internal arguments; acting as critical friends; participating in technical aspects of governance, such as audit and remuneration committees; reading the environment; and appointing the vice-chancellor.

Governing boards in the United States, Canada, the United Kingdom, South Africa, and Australia have these general responsibilities. Other countries have moved to the adoption of different types of governing boards or placing external members on decision-making bodies that formerly had internal institutional membership. For example, reforms in Sweden led to the creation of university councils of directors, chaired by the rector and most of whose members are appointed by the government. Legislation enacted in the Netherlands in 1999 created a board for each university, whose five members are appointed by the minister of education in consultation with the university. Since 1983, Spanish universities have had social councils, a majority of whose members are from outside the university. Italy has created similar bodies (Amaral and Magalhães, 2002). Austrian legislation in 2002 created institutional governing boards of five to nine external members that select the rector from a proposal by the academic senate (Pechar, 2003). Australian councils have moved to greater reliance on representatives of business as members (Margison and Considine, 2000).

Lay governing boards – that is, boards whose members are external to the institution – can contribute positively

to the university. They bring expertise, connections to the community, an ability to see the long-term picture for the institution, and have a broad view of the environment in which the institution functions. However, they are not without their drawbacks and dangers. Governing board members may have little understanding of how institutions function, or may be unclear about the role of the board. Infrequent meetings make it difficult for board members to have an in-depth knowledge of the institution. Thus, they may defer uncritically to management or be tempted to dwell on the familiar aspects of operation, finance, and investment (Chait *et al.*, 1996).

As the voices of the lay members of governing boards become more numerous and prominent, and their call for institutions to be more market sensitive, fast acting, and generally businesslike, the potential for conflict with faculty values and academic traditions increase. Institutional effectiveness is strengthened when there is a healthy balance of a strong academic senate or academic board and a strong governing board characterized by true participation of the academic community and open communications between the bodies.

The Role of the Institutional Head

Just as institutional management and governance are being reshaped by the forces described in the first section above, so is the role of the institutional head. The role is highly varied and complex, and includes administrative, political, and entrepreneurial components (Birnbaum and Eckel, 2005). As a leader, the institutional head envisions the future, serves as an institutional symbol, facilitates the hard work of change, and sets the tone (Green, 1997). As administrator or manager, he or she develops and carries out policies, creates plans, determines budgets, supervises and evaluates subordinates, monitors systems, and designs and implements accountability schema. As a politician, that individual responds to multiple constituencies and interest groups, negotiates differing perspectives and priorities, and forms coalitions to get things accomplished. As entrepreneur, the institutional head develops and exploits markets, secures new resources, fosters innovation, and, in many countries, participates in fund raising.

The balance among these roles and the relative time the institutional head devotes to different aspects of the job varies considerably within and across nations. In the United States, presidents delegate most academic matters to the chief academic officer, who bears the title of provost, vice-president for academic affairs, or dean of instruction (in community colleges), or dean of the college or faculty (in liberal arts colleges). According to American Council on Education survey data, US presidents report that their primary use of time is, in descending order, fund raising, budget and financial management, community relations,

strategic planning and governing board relations, and a mix of fiscal and external issues (ACE, 2007). In the United Kingdom, vice-chancellors in old universities play an important role in leading the academic enterprise; at the same time, they focus on external relations (Smith *et al.*, 1999). In other countries with strong government steering and less market orientation, the institutional head generally focuses more on academic matters and less on external duties.

Forces Shaping the Role of the Institutional Head

The four main forces shaping the role of the institutional head are:

1. *Culture.* Institutional heads operate within their national cultures, as well as in the academic culture. Cultural values underlie a society's assumption about the role of higher education in that country – for example, whether higher education acts autonomously from government direction or is an extension of civil service. The norms regarding individual and collective behaviors in an academic community are cultural creations. In some societies, such as the United States, with its strong individualistic and market orientation, leadership is usually equated with assertive behaviors; leaders are exhorted to demonstrate their individual leadership. In others, such as Japan, it is unseemly for the rector to appear to lead too vigorously (Ishikawa, 1997). However, social norms can change, especially in response to important shifts in the external environment or strong pressures on higher education institution. The demise of communism and the end of apartheid in South Africa had profound consequences for the belief systems shaping higher education and its leaders in those respective countries.
Leaders are also shaped by academic cultures. In spite of notable difference of national cultural norms, traditional academic institutions seem to have more cultural similarities that derive from their academic mission than differences that can be attributed to the country in which they are located. The academic values of unfettered inquiry, the pursuit of knowledge for its own sake, the quest for freedom from external interference, and the dedication to preparing the next generation of leaders, professionals, and citizens transcend countries and cultures. In addition, the academy values shared decision making and has a tradition of selecting leaders from within the academic culture. Academic institutions are historically loosely coupled institutions composed of highly independent schools and faculty members who are more connected and loyal to their disciplines than their institutions. Several factors further shape the role of the institutional leader.
2. *Institutional mission.* The top job varies according to whether the college or university is in a system defined as elite – shaping the mind and character of a relatively small ruling class for elite roles, or mass – transmitting the skills and preparation for an involvement in technical and elite roles by a broad population (Trow, 2006). Trow notes that universities in elite systems tend to resemble each other, be more homogenous in their students and academic staffs, and narrower in scope and mission. They are run by small staffs and governed by powerful senior academics. Their administrations consist of academics who leave their scholarly posts for brief periods. For example, the institutional head is typically appointed or elected to a fixed term. In contrast, universities in mass systems are more differentiated from one another, having a wide range of programs and offerings and, in turn, having a broader set of interested stakeholders. Their administrations are large and differentiated with professionals fulfilling specialized roles, including a more professionalized and experienced institutional head. The drive to widen participation around the world makes the distinction of elite and mass institutions more applicable within nations than across them.
3. *The role of the state, the market, and the academic oligarchy.* A set of factors shaping the role of the institutional head and his or her level of authority is the relative weight of state authority, the market, and the academic oligarchy, three elements that Clark (1983) combines into a triangular model of interdependence and coordination. In the dirigiste or state-controlled model, countries have strong ministries, broad regulatory powers, high dependence on government funding, and low institutional autonomy. The combination of strong ministries and low institutional autonomy work to limit the role and influence of the institutional head. Examples include many African nations, Gulf-region states, China, Russia, and former Soviet republics, as well as some Western European nations. An extreme example can be found in some African nations that vest great authority in the university chancellor (usually the head of state) who appoints rectors and in some cases their deputies and other administrative officers, and approves the appointment of elected faculty and students to councils. In these countries, presidential authority to steer institutional direction is curtailed. Ironically, while the vice-chancellor or rector may have limited say in the larger strategic and policy issues, he or she may have significant involvement in and power over many minor matters (Mwiria, 2003).
Institutions that are highly market driven tend to see institutional heads as having a relatively high degree of authority as compared to campus leaders in other situations. They are selected by the governing board

or council, and generally move into the top positions through progressive moves up the administrative ranks, becoming professional administrators at some point in their career. Other countries that are transitioning toward more market-oriented systems are strengthening the role of the institutional head and their ascensions are following similar paths. However, this model is not a single one, varying much by institutional type and country. The term chief executive officer (CEO) is often used to describe this type of position and to suggest a growing similarity to corporate leadership. However, even when the CEO has relatively high formal power and authority – which may include control over resource allocation and appointments of the senior management team and other administrators – her or his influence is often limited by faculty-governance mechanisms and institutional culture. This is the case in the United States, where in spite of the perception that US presidents have a great deal of power, in reality, they must negotiate with the faculty, build consensus among various stakeholders about directions for change, and take care not to alienate internal or external constituencies with high-handed decision making.

Finally, in some countries the academic oligarchy, rather than administrative or government/ministerial authority, plays the dominant role in shaping institutional decision-making processes, thus limiting the rector's power. The push to democratize Eastern and Central European universities immediately after the fall of communism resulted in greater autonomy of the schools within a university and enhanced faculty participation, but left the rectors weak and unable to exercise their formal powers (Scott, 2006). In Western European countries such as France, Spain, and Greece, as well as some developing countries, the rector's authority is constrained compared to their counterparts in the United States or the United Kingdom, since the individual faculties hold a great deal of power and many other institutional bodies must approve new initiatives (Taylor *et al.*, 2007). When academics play a central role in the selection of their leader, they further hamper the institutional head's authority because of pressures on the institutional head to remain in their stakeholders' favor (Bloom and Rosovsky, 2006).

4. *Method of appointment and career patterns.* A fourth set of factors that affect how the institutional head functions is the selection process. Whether the institutional head is appointed or elected affects legitimacy, formal authority, and relationships with stakeholders. The appointed head (often the CEO model) tends to have more formal authority, and arguably less legitimacy, with the faculty than the elected rector. The US president is appointed by the governing board, with highly variable levels of input from faculty, staff, and students.

He or she serves at the pleasure of the board, and under formal agreements that are increasingly codified in a contract (renewable) of 3–5 years. Newly appointed US presidents most often come from senior administrative positions outside the institution (64%), and 21% have served previously as presidents. Furthermore, not all presidents are academics. Thirteen percent of the presidents held their most recent position outside of academe, and more than 60% had some experience outside higher education sometime during their careers (ACE, 2007). In contrast, a study of 16 UK vice-chancellors appointed between 2002 and 2004 indicated that 20% were internal promotions from the position of deputy or pro-vice-chancellor and another 14% internal promotions from a senior academic position; 29% were in those senior management positions in other institutions; and 11% were vice-chancellors in their immediate prior position (Breakwell, 2006). These data parallel those in a larger data set on 341 vice-chancellors appointed since 1960 (Smith *et al.*, 1999). That database indicated that more than 90% had careers in academia. Twelve percent made horizontal moves from other vice-chancellorships, and a third were internal promotions.

The Latin American and European rector, in comparison to the US president or British vice-chancellor, tends to be elected to specific terms, often renewable, directly by senior faculty or from a representative governance body and usually comes from the senior academic ranks of that institution. They expect to return to teaching after serving as rector. Unlike the US president, rectors often campaign for the post. As rectors may not move up the administrative ladder in a career progression from dean, to vice-president or deputy vice-chancellor, they are likely to be less experienced in leadership and management and less well prepared for the job (Taylor *et al.*, 2007). However, several European countries have moved from elections to appointments, including the Netherlands, Sweden, and Austria. Denmark and Norway are in the process of making such changes (Taylor *et al.*, 2007).

The third pathway to the presidency is via government appointment. This ministerial model may be prevalent in systems where universities are subject to administrative law and university administrators are civil servants or where ideology dictates national policy, particularly in nondemocratic countries (Green, 1997). It is most often associated with governments that exercise considerable control over higher education, often in an undemocratic context. Until approximately 10 years ago, the ministerial model was operating widely in Central Europe and many parts of Asia and Africa. However, the push for democracy in some nations and many of the changes discussed throughout this article are pushing some countries to rethink the ministerial appointment model.

The Institutional Head as Change Leader

In hiring institutional heads, the expectation of the governing body is generally that the president or vice-chancellor will move the institution forward, that is, lead positive change. Scholars disagree on the impact that leaders actually have on institutions. While some scholars assert that leaders have a minimal impact in the ambiguous and anarchic culture of higher education (Cohen and March, 1986; Bennis, 1976), most others believe that leadership matters. Even if leadership does matter, it is difficult to gauge how much it matters, making the job of the rector frustrating as well as difficult. Higher education institutions have inconsistent and ambiguous goals, making timely feedback difficult. Furthermore, most universities do not really know how students learn or understand the essential processes involved in creating civically minded students or globally competent citizens. Different stakeholders have different indicators of success. Thus, a successful institution/leader is a contested concept (Birnbaum, 1988).

Different approaches to leading change work under different circumstances. In the absence of crises or very strong external mandates for change, leaders generally need to work within the culture to change it (Kezar and Eckel, 2002). This is a slow and long-term approach. The costs are the slowness and the gradual nature of the change process, the benefits are establishing a culture of change, establishing ownership of the new directions, and working through the inevitable conflicts and suspicions (Eckel *et al.*, 1999). Leaders who are perceived as pushing too hard against established institutional norms or moving too quickly with a change agenda generally meet resistance from academics. In some instances, the resulting loss of mutual trust can cost the institutional head his or her job. The efforts of a new vice-chancellor at Oxford University to change the governance structure angered professors, who viewed the proposed new structure of an academic board as unnecessary and intrusive and reducing opportunities for faculty participation in collegial decision making (Labi, 2006b; North, 2006). The controversy is emblematic of a larger debate in the UK and in other countries with relatively strong institutional heads, about who initiates change, at what pace, and the balance of power between vice-chancellors or presidents and faculty. Tension with the faculty can be costly, as is illustrated by the high-profile case of the resignation of former Harvard University president Lawrence Summers. US presidents commenting on recent presidential oustings noted that what gets presidents into trouble are relationships and process issues, rather than the goals (Fain, 2006). Three recent departures of Canadian presidents suggest that the pressures on university presidents to perform are powerful around the world. The pressure to enhance institutional performance and meet many different stakeholders' needs without creating firestorms

of controversy are key challenges for institutional heads (Alphonso, 2007).

There is no single model of leading change. Effective change leadership will vary with the institutional climate and traditions, the level of turbulence in the external environment, and the intensity of external pressures for change, and the strengths and proclivities of the individual leader. The different approaches that were taken in the volatile and contested environment of South Africa provide a good case in point. Kulati and Moja (2006) describe different leadership responses to the challenges of transformation in the postapartheid era. One set of approaches, transformational leadership combines strong leadership from the top with consultation and respect for the institution's intellectual agenda and prevailing institutional culture. In one variation, called reformed collegialism, the emphasis is on understanding and supporting the academic heart of the institution as the driver for change. A second variation, called transformative managerialism emphasizes driving transformation from the center. In this case, power is centralized, decentralized, and recentralized. The top leadership group makes key decisions, which are implemented by the deans at the school level and then taken back to the center for approval. Other leaders took a more managerial approach, characterized by a strong center aiming to make the institution competitive and market oriented. There were also variations within the managerial approach, with some leaders applying managerial techniques to make the institution more efficient and effective while acknowledging the centrality of academics (strategic managerialism), and others seeking to reshape the institution in more profound ways so that it is competitive and market oriented (unwavering entrepreneurialism). Kulati and Moja also point to failures of leadership, caused largely by institutional conditions resulting from a historical legacy, inexperienced new leadership, and external policy pressures. Crisis leadership occurred mostly, but not entirely at the South African historically black universities, which experienced sudden declines in enrolment that threatened their existence. The crisis was characterized by a breakdown in governance, poor management, and conflict. Leadership at the top proceeded by crisis management, making strategic steering impossible. The South African case illustrates the difficulty of prescribing approaches to leading change. National context and institutional history and culture play have a significant impact on the leadership strategies that fit a particular place and time.

Conclusion

This article has demonstrated that the evolving relationship between government and higher education and

the growing pressures of the market have resulted in a growing managerialism in higher education institutions and strains between collegial governance and decision making and external demands placed on institutions. Regardless of the institutional starting point being highly market-driven (such as in the US or Australia) or being in a centralized system, such as France or Japan, growing managerialism is characterized by greater executive power, a professionalized administration, a diminished role for faculty in decision making, and a shift in values. Increased institutional autonomy and accountability and greater competitiveness are also putting greater pressures on institutional heads to provide strategic leadership for their institutions and to be externally focused. In some nations, institutional heads are clearly CEOs with those specific roles; in others, institutional heads are combining more traditional roles as academic leaders who are first among equals with features of the CEO position.

Bibliography

- ACE (American Council on Education) (2007). *The American College President: A 2007 Edition*. Washington, DC: American Council on Education.
- Alphonso, C. (2007). Trio of resignations shock university system. *The Globe and Mail*, February 6.
- Altbach, P. (2000). Asia's academic aspirations: Some problems. *International Higher Education* **19**, 7–8.
- Amaral, A., Fulton, O., and Larsen, I. (2003). A managerial revolution? In Amaral, A., Meek, V. L., and Larsen, I. M. (eds.) *The Higher Education Managerial Revolution?* pp 275–296. Dordrecht: Kluwer Academic Publishers.
- Amaral, A. and Magalhães, A. (2002). The emergent role of external stakeholders in European higher education governance. In Amaral, A., Jones, G. A., and Karseth, B. (eds.) *Governing Higher Education: National Perspectives on Institutional Governance*, pp 1–21. Dordrecht: Springer.
- Bennis, W. G. (1976). *The Unconscious Conspiracy: Why Leaders Can't Lead*. New York: AMACOM.
- Birnbaum, R. (1988). *How Colleges Work*. San Francisco, CA: Jossey-Bass.
- Birnbaum, R. and Eckel, P. D. (2005). The dilemma of presidential leadership. In Altbach, P. G., Berdahl, R. O., and Gumpert, P. J. (eds.) *American Higher Education in the Twenty-First Century*, 2nd edn., pp 323–346. Baltimore, MD: Johns Hopkins University Press.
- Bloom, D. E. and Rosovsky, H. (2006). Higher education in developing countries. In Altbach, P. G. and Forest, J. F. (eds.) *International Handbook of Higher Education*, pp 443–460. Dordrecht: Springer.
- Bok, D. (2003). *Universities in the Marketplace: The Commercialization of Higher Education*. Princeton, NJ: Princeton University Press.
- Breakwell, G. M. (2006). Leadership in education: The case of vice-chancellors. *Perspectives* **10**(2), 52–58.
- Burgan, M. (2006). *What Ever Happened to the Faculty?* Baltimore, MD: Johns Hopkins University Press.
- Chait, R. P., Holland, T. P., and Taylor, B. E. (1996). *Improving the Performance of Governing Boards*, 2nd edn. Phoenix, AZ: ACE/Oryx Press.
- Clark, B. (1983). *The Higher Education System: Academic Organization in Cross-National Perspective*. Berkeley, CA: University of California Press.
- Clark, B. (1998). *Creating Entrepreneurial Universities: Organizational Pathways of Transformation*. Oxford, UK: Elsevier Science.
- Cloete, N. and Kulati, T. (2003). Managerialism within a framework of cooperative governance? In Amaral, A., Meek, V. L., and Larsen, I. M. (eds.) *The Higher Education Managerial Revolution?* pp 229–251. Dordrecht: Kluwer Academic Publishers.
- Cohen, M. D. and March, J. G. (1986). *Leadership and Ambiguity*, 2nd edn. Boston, MA: Harvard Business School Press.
- Commission of the European Communities (2003). *The Role of the Universities in the Europe of Knowledge*. COM (2003) 58 final.
- Currie, J. and Vidovich, L. (1998). The ascent towards corporate managerialism in American and Australian universities. In Mart, R. (ed.) *Chalk Lines: The Politics of Work in the Managed University*, pp 112–144. Durham, NC: Duke University Press.
- Duderstadt, J. (2001). Fire, ready, aim! University decision-making during an era of rapid change. In Hirsch, W. and Weber, L. (eds.) *Governance in Higher Education: The University in a State of Flux*, pp 26–51. London: Economica.
- Eckel, P., Hill, B., Green, M., and Mallon, B. (1999). *On Change. Reports from the Road: Insights on Institutional Change*. Washington, DC: American Council on Education.
- Eckel, P. D. (ed.) (2006). *The Shifting Frontiers of Academic Decision Making: Responding to New Priorities, Following New Pathways*. Westport, CT: ACE/Praeger.
- Edwards, K. (1997). The United Kingdom. In Green, M. (ed.) *Transforming Higher Education: Views from Leaders around the World*, pp 223–237. Phoenix, AZ: American Council on Education and the Oryx Press.
- Fain, P. (2006). Crisis of confidence: Three current and former college president discuss the recent spate of failures at the top. *Chronicle of Higher Education* **52**(42), A28.
- Green, M. (1997). Institutional leadership in the change process. In Green, M. (ed.) *Transforming Higher Education: Views from Leaders around the World*, pp 27–53. Phoenix, AZ: American Council on Education and the Oryx Press.
- Green, M., Eckel, P. D., Calderon, L., and Luu, D. (2007). *Venturing Abroad: Delivering U.S. Degrees through Overseas Branch Campuses and Programs*. Washington, DC: American Council on Education.
- Gumpert, B. and Sporn, B. (1999). Institutional adaptation: Demands for management reform and university administration. In Stuart, J. and Tierney, W. (eds.) *Handbook of Theory and Research*, vol. xiv, pp 103–145. New York: Agathon Press.
- Ishikawa, T. (1997). Japan. In Green, M. (ed.) *Transforming Higher Education: Views from Leaders around the World*, pp 294–308. Phoenix, AZ: Oryx Press and the American Council on Education.
- Johnson, B. and Cross, B. (2004). Academic leadership under siege: Possibilities and limits of executive dean ship. *South African Journal of Higher Education* **18**(2), 34–58.
- Kezar, A. and Eckel, P. D. (2002). The effect of institutional culture on change strategies in higher education: Universal principles or culturally responsive concepts? *Journal of Higher Education* **73**, 435–460.
- Kinser, K. and Levy, D. C. (2006). For-profit higher education: U.S. tendencies, international echoes. In Altbach, P. G. and Forest, J. F. (eds.) *International Handbook of Higher Education*, pp 141–158. Dordrecht: Springer.
- Kirp, D. L. (2003). *Shakespeare, Einstein, and the Bottom Line*. Cambridge, MA: Harvard University Press.
- Knight, J. (2002). Trade in higher education services: The implications of GATS. *Observatory on Higher Education* March. <http://obhe.ac.uk> (accessed June 2009).
- Kulati, T. and Moja, T. (2006). Leadership. In Cloete, N., Maassen, P., Fehnel, R., et al. (eds.) *Transformation in Higher Education: Global Pressures and Local Realities*, pp 153–170. Dordrecht: Springer.
- Labi, A. (2006a). German universities learn which of them have a shot at elite status. *Chronicle of Higher Education*, January 24.
- Labi, A. (2006b). Reform and resistance at Oxford: A controversial leader stirs up a university in search of new ideas. *Chronicle of Higher Education*, June 30.
- Margison, S. and Considine, M. (2000). *The Enterprise University: Power, Governance and Reinvention in Australia*. New York: Cambridge University Press.

- Middlehurst, R. (2004). Changing international governance: A discussion of leadership roles and management structures in U.K. universities. *Higher Education Quarterly* **58**(4), 258–279.
- Mok, J. K. H. and Lo, E. H. C. (2002). Marketization and the changing governance in higher education: A comparative study. *Higher Education Management and Policy* **14**(1), 51–82.
- Morhman, K. (2005). World-class universities and reform in China. *International Higher Education* **39**, 22–23.
- Mwiria, K. (2003). University governance and university–state relations. In Teferra, D. and Altbach, P. G. (eds.) *African Higher Education: An International Reference Book*, pp 32–43. Bloomington, IN: Indiana University Press.
- North, M. (2006). Stronghold of academic rule besieged by federal camps. *Times Higher Education Supplement* **8**, 16.
- Pechar, H. (2003). In search of a new profession: Transformation of academic management in Australian universities. In Amaral, A., Meek, A., and Larsen, I. M. (eds.) *The Higher Education Managerial Revolution?* pp 109–130. Boston, MA: Kluwer Academic Publishers.
- Schenker-Wicki, A. and Hürlimann, M. (2005). Performance funding of Swiss universities – success or failure? An ex post analysis. *Higher Education Management and Policy* **18**(1), 45–61.
- Schmidt, K. E. (2006). Higher education in Scandinavia. In Altbach, P. G. and Forest, J. F. (eds.) *International Handbook of Higher Education*, pp 517–539. Dordrecht: Springer.
- Scott, P. (2001). Universities as organizations and their governance. In Hirsch, W. and Weber, L. (eds.) *Governance in Higher Education*, pp 125–142. London: Economica.
- Scott, P. (2006). Higher education in Central and Eastern Europe. In Altbach, P. G. and Forest, J. F. (eds.) *International Handbook of Higher Education*, pp 423–442. Dordrecht: Springer.
- Shattock, M. (2003). *Managing Successful Universities*. Berkshire: Society for Research into Higher Education/Open University Press.
- Shattock, M. (2006). *Good Governance in Higher Education*. Berkshire: Open University Press.
- Singh, M. (2001). Reinserting the public good in higher education transformation. *Kagisano CHE Higher Education Discussion Series*, no.1, November.
- Slaughter, S. and Rhoades, G. (2004). *Academic Capitalism and the New Economy*. Baltimore, MD: Johns Hopkins University.
- Smith, D., Scotgt, P., Bocock, J., and Bargh, C. (1999). Vice-chancellors and executive leadership in U.K. universities: New roles and relationships? In Henkel, M. and Little, B. (eds.) *Changing Relationships between Higher Education and the State*, pp 280–306. London: Jessica Kingsley.
- Sporn, B. (2006). Governance and administration: Organizational and structural trends. In Altbach, P. G. and Forest, J. J. F. (eds.) *International Handbook of Higher Education*, pp 141–158. Dordrecht: Springer.
- Taylor, J. S., Amaral, A., and Machado, M. (2007). Strategic planning in U.S. higher education: Can it succeed in Europe. *Planning for Higher Education* **January–March**, 5–17.
- Trow, M. (2006). Reflections on the transformation from elite to mass to universal access: Forms and phases of higher education in modern societies since WWII. In Altbach, P. G. and Forest, J. F. (eds.) *International Handbook of Higher Education*, pp 243–280. Dordrecht: Springer.

The Changing Roles of University Governing Boards and Councils

J Fielden, CHEMS Consulting, Hook, UK

© 2010 Elsevier Ltd. All rights reserved.

The Changing Context of University Governance

Factors that Drive Changes in Governance

With very few exceptions governments around the world are faced with a growing demand for higher education from an increasing proportion of their population. In countries where mass higher education is available to the traditional age group of 18–24, there is also a demand from the mature adult population wishing to catch up on what it had missed or to learn new skills. The implications of these trends for the way that the higher education sector is governed are that:

- The scale of the number of institutions makes central direction difficult; a central ministry might be able to have coherent managerial supervision of 50 institutions, but central control is difficult when there are 300.
- The growth in demand from different types of clientele inevitably leads to different types of institution emerging; the sector becomes more diverse and again harder to manage.
- The funding required for an expanded sector is too much for the state to fund; the options for overcoming this involve encouraging cost-sharing and income-generation or sharing the burden with the private sector. Both these raise administrative complications, often requiring new agencies of the state for functions such as regulating the private sector or managing loan schemes, etc.
- The increasing size and complexity of the higher education sector requires greater specialist knowledge than is usually found in civil service structures that are based on the rotation of generalists from one ministry post to another.

These factors have several structural and procedural consequences:

- Ministries of education hive off their higher education activities into separate ministries of higher education (sometimes linked with science, research, or the development of skills).
- New specialist agencies (which we shall call buffer bodies) are created to take on the complex role of funding or managing higher education institutions for the minister responsible.
- These buffer bodies have a scope that usually covers public and private provision – although some countries

prefer to regulate private providers through separate agencies.

- As higher education systems expand, the quality agenda assumes a growing importance and governments set up quality agencies to monitor and inspect the quality of provision. These are often focused on the private sector where the risks of poor quality are high.
- The role of the central state ministry is to focus on developing strategy and setting policy rather than managing institutions in detail.
- Where the state and its agencies withdraw from detailed supervision of institutional affairs, institutions are granted greater autonomy to manage their own affairs.
- The state in consequence develops different methods of monitoring and management. Institutions are held accountable not for the details of their expenditure but for their achievement of agreed strategic plans.

There are two consequences of these changes: a reconsideration of the balance between autonomy and accountability and an increase in the managerial pressures on university leaders.

The key question that governments face is how much control they should exercise over universities. In **Table 1** Fielden (2008) has described the various models of ownership which influence the degree of control.

In the universities that are seen as a core part of the public sector with financial and staffing regulations drawn directly from the civil service, as in some European countries, the scope for granting autonomy in financial and staffing matters will inevitably be limited. Thus, we find that in countries where academics are civil servants, they have to seek approval before they travel overseas and decisions on their appointment and promotion are taken at a national level. The statutory basis of universities can also limit freedoms; thus, in New Zealand universities are designated as crown corporations and do not own their assets.

It is a generally accepted principle that universities operate more effectively, if given autonomy in key decisions. The debate centers on what powers and controls should be retained by the ministry, statewide governing board, or buffer body that has supervision of institutions. The areas where there is the greatest difference in practice internationally are:

- *Setting a cap on overall student numbers.* Since most countries fund the largest part of public higher education

from taxation, they consider themselves justified in limiting the number of students. However, some countries have a more relaxed view when there is a mixed public/private funding.

- *Setting the level of tuition fees.* In view of the intense political interest in student fees, almost no country leaves the setting of undergraduate tuition fees in public universities to the decision of individual boards. However this rule does not apply to fees for postgraduates, external or open/distance learning students, and international students, where the institution can make the decision. For example, in Sri Lanka public universities can generate considerable income from charges to external degree students, who can gain

Table 1 Various models of ownership which influence the degree of control

<i>Institutional governance model</i>	<i>Status of public universities</i>	<i>Examples in</i>
A. State control	Can be agency of the MOE, or a state-owned corporation	Malaysia
B. Semi-autonomous	Can be agency of the MOE, a state-owned corporation or a statutory body	New Zealand, Italy, and France
C. Semi-independent	A statutory body, a charity or a nonprofit corporation subject to MOE control	Singapore
D. Independent	A statutory body, charity or nonprofit corporation with no government participation and control linked to national strategies and related only to public funding	Australia and United Kingdom

From Fielden, J. (2008). Global trends in university governance. *Education Working Paper Series No 9*. Washington, DC: World Bank.

university awards, but with limited tutorial input from university staff. This income is outside government control.

- *Choosing what programs or courses to offer.* Developed countries set limits on the numbers in high-cost strategic disciplines such as medicine and dentistry and consequently control the opening of medical schools or new medical programs. In some other countries the state seeks to approve the creation of all new programs or, at a higher level, of new faculties and broad program areas.
- *Appointing and then promoting new academic staff.* Even where institutions are largely autonomous, the State may want some say in the numbers of academic staff. For many years the United Kingdom had a ruling that the number of professors could not exceed a certain percentage of the total staff numbers; this control has been abandoned.
- *Selling or acquiring assets.* States are usually very nervous at giving publicly funded institutions the power to dispose of assets. As a result they seek to be involved in the decision in some way.
- *Borrowing money.* There is a similar reluctance to allow institutions the freedom to borrow money from the commercial sector as this usually requires the creation of mortgage charges over publicly funded assets.

As a general rule we find that control is the tightest in developing countries, which is often due to the inheritance of colonial systems and control cultures. An illustration of how extreme this can be is in the area of financial control (Table 2).

Once the state has decided on the extent of freedom, it must devise appropriate procedures for ensuring accountability. The motives for this are obvious: to ensure as far as is possible that higher education achieves the strategic goals set for it by the state; to maintain the quality of the provision offered to students; and to ensure that public money is spent wisely with no fraud, corruption, or

Table 2 Different approaches to financial control

<i>Topic</i>	<i>Centralized control</i>	<i>Full autonomy</i>
Annual budgets	Agreed in detail by MOE or the funding body	Agreed by the Board (but possibly reported to MOE or the buffer body)
Expenditure	Line item control so that institutions cannot switch expenditure between the agreed budget headings	Total freedom to allocate and spend as required within the overall total grant or budget awarded by the MOE
Underspending at the end of an accounting period	Surrender of all underspent sums to MOE/Ministry of Finance	Freedom to carry forward underspending (and to absorb any overspendings from future funds within limits)
External earnings from nongovernment sources	Surrender to the Ministry of Finance or MOE of all external earnings	Freedom to retain and spend freely all sums earned from nongovernment sources
Tuition fees for domestic and international students	Fees cannot be charged or, if they are, have to be set at a fixed rate and then surrendered to the Ministry of Finance	Fee levels can be set freely and the money retained without affecting the budget allocation from the government

From Fielden, J. (2008). Global trends in university governance. *Education Working Paper Series No 9*. Washington, DC: World Bank.

unethical practices. (Salmi, 2008) describes the type of corruption or fraudulent practice that is becoming increasingly common.

In the same paper Salmi (2008) describes the many forms that accountability can take

legal requirements such as licensing, financial audits and reports, quality assurance procedures such as program or institutional accreditation, benchmarking exercises to compare programs across institutions, professional qualification examinations, budget allocation mechanisms that reward performance, and oversight structures such as governing boards with representation from external stakeholders.

The most significant control mechanism is a review of an institution's strategic plan and the extent to which it is achieving any national policy goals or targets set by the ministry. In recent years governments have been adopting the approach of consulting the sector over the development of a national strategy for higher education and, once this has been agreed, requiring institutions to prepare their own strategic investment plans for meeting the national goals. New Zealand adopts this model as the tertiary Education Commission expects institutions to submit investment plans and Australia has followed a similar path.

Accountability is achieved not only through establishing regulatory control mechanisms and monitoring procedures, but also through information requests. While most governments aim to ensure as light a touch as possible, there are pressures from politicians and the public for more and more information on how higher education operates and reviews its effectiveness. In addition, there is a tendency for ministers of education to achieve office on the basis of election promises which usually involve the development of policies that require new forms of regulation and monitoring. The risk therefore is that, as formal control systems are removed, they are replaced by rigorous reporting requirements. In England, for example, the scale of monitoring and reporting has become so heavy that the buffer body itself has established a target for lightening the accountability burden and created a formal entity called the Higher Education Regulation Review Group for reviewing any new accountability and reporting burdens that are planned. **Table 3** illustrates the scale of regulation and reporting in 2008. Some would consider this excessive in a country that has an image of giving autonomy to universities.

From this analysis, we can draw some conclusions that are relevant to institutional governance and the role of university councils or boards:

- The state is increasingly reliant on institutions governing themselves effectively, once it decides to abandon detailed control of university operations.

Table 3 Reporting and regulation in England

-
- Detailed targets set for domestic student numbers, the social class of students and their home location with rewards for achievement and penalties for failure.
 - Semimandatory requirements on management good practice in risk management, procurement, human resource management, etc.
 - Audits by the funding body of governance and management processes.
 - Production of strategies for property, human resources, teaching, and learning.
 - An expectation that the institutions will also have strategies for widening participation (greater access), research, finance, ICT, e-learning, and development of links with industry.
 - Extensive external audits of finance, teaching quality, and achievement of all the national strategies.
 - Special funding pockets for *ad hoc* policy priorities which have conditions attached.
 - National laws on health and safety, racial, sexual, or age discrimination that require formal annual reports.
-

- This places a big accountability burden on university councils and the institutional leadership and leads inevitably to greater use of managerial mechanisms and techniques by the academic leadership.
- Once the state frees institutions from detailed monitoring, they are able to develop unique strategies that are often related to their region or community and this brings with it the need for local and regional representation in decision-making.

University Councils or Boards

The Changing Role of Councils

Traditional collegial models of university government are under attack. A controversial example of this was the attempt in 2006–2007 by a new vice-chancellor at Oxford to limit the powers of the 3500-strong Congregation of the University and its 39 autonomous colleges by strengthening the central authority of a new university council. He was only partially successful, since both Oxford and Cambridge have a deeply engrained and democratic approach to governance and are wholly atypical of the British system as a whole. However, in other countries, legislation has been introduced with the same aim, for example:

- In Austria, legislation introduced a university council with the power to appoint the rector and prepare institutional plans.
- As part of reforms creating national university corporations, in 2004 Japan established a three-tier system with an administrative council, an academic council and an executive board.
- When the United Kingdom gave its polytechnics university status in 1989, the legislation stated that overall

authority should rest with a board of governors which would have authority over an academic board.

- In the Netherlands and Sweden, legislation has established governing or supervisory boards chaired by external members.

Thus, the dominant pattern that is emerging is of a governance structure in which there are three main players: the board or council, an academic board or senate, and the executive. Where there is a Higher Education Act, it is common for the roles of the council and the senate to be defined broadly in the legislation and this also usually defines the relationship between the council and the vice-chancellor/rector. In the latter case it is usual for the legislation to state the accountability of the rector to the board as head of the executive.

The relationship between the council and the academic board is not always easy. In the United Kingdom, for example, it is now accepted that the council “has legally unambiguous authority for the governance of the university,” meaning that the academic board is legally subservient to it (Committee of University Chairs, 2006). But this legal position can be ignored when financial and academic values conflict. In a recent UK example, the senate at the University of Warwick rejected a proposal approved by the university council to establish a campus in Singapore on the grounds that it feared that its academic freedom in Singapore would be constrained. Such a relationship is sometimes referred to as one of creative tension.

Academic boards are usually large bodies and thus not ideally suited to detailed decision-making. Many of them began with the collegial principle that every professor was entitled to be a member and this still holds in some jurisdictions. Their size and the infrequency of their meetings (usually three times a year) has restricted their ability to participate in the immediate and responsive decisions that are now required. As a result an increasing number of academic decisions are made by senior academic members of the executive rather than by academic board. The general trend is therefore towards academic boards losing some of their powers to councils and the executive, but this depends to a great extent on the way decision-making structures operate within the institution. If the rector is the chair of the academic board, there is a greater chance of the decisions of both bodies being coordinated rather than duplicated or contradictory.

The executive, headed by the rector or vice-chancellor, is acquiring a greater say over key decisions largely due to the growing intensity and complexity of managing an institution. No council can master all the technical intricacies of university management and must focus on the key strategies and performance indicators. As all councils devolve full responsibility for management to their executive, they must develop performance indicators that can

assess whether the university is meeting the strategic goals they have set and whether the board itself is being effective. The chief executive is the key player in making this happen and the job specifications for the post are becoming increasingly demanding. Breakwell and Tytherleigh (2008) found that very few vice-chancellors in the United Kingdom had professional management qualifications and there was little evidence of nonacademic experience and yet they concluded that “VC’s are expected to combine academic credibility with all the other competences expected of the leaders of any large commercial businesses that have political significance.”

The Composition of Councils

There is a consistent trend toward a reduction in the size of university councils. In the United States, the board of trustees usually has no more than 12–15 members, although in the model followed in the United Kingdom and the British Commonwealth, it was a larger body numbering between 30 and 50. Australia recommends a number in the range from 15 to 22, while Denmark and New Zealand have recently set a maximum number of less than 15, whereas the post-1992 universities in the United Kingdom can appoint up to 24 members. A recent survey of governance in 20 African countries (Lao and Saint, 2008) has shown that in 70% of the countries, councils had less than 30 members. The managerial demands on councils have tended to increase the number of their meetings as well as reduce their size.

The composition of councils is also changing with a general trend toward the appointment of more members from outside the university system. Sweden, the post-1992 sector in the United Kingdom, Australia, Denmark, Norway, and Tanzania all require a majority of council members to be appointed from external constituencies. In Africa also, this was true for over 50% of the 20 African countries surveyed by Lao and Saint. However, there is a wide variation in practice in designating the origin of external members and who can appoint them. There are various models:

- election of some or all members by various internal constituencies inside the university;
- appointment by the national or state government;
- appointment by the Minister of the recommended people nominated by the board;
- appointment by the board without any outside involvement; and
- nomination by the government or external representative bodies of persons to fill an allotted place.

Many universities have one or more representatives of their central or local government on their council and some institutions in small countries specify that a certain number of members shall be from another country in order to ward against academic isolation. In the United

States, the practice in the composition of boards of state universities varies greatly but a database of the composition, term of service, modes of appointment, and degree of State control over nominations has been compiled by the Association of Governing Bodies (AGBs).

Membership of a university council has long been considered a prestige appointment, which the AGB has designated “the Crown Jewel of public service,” as it is usually unpaid. However, it is becoming increasingly onerous as the weight of accountability shifts on to the council. This is beginning to spawn the provision of training programs for members of council as well as formal requirements for appraisal and performance review. The leading proponents in this are the AGB and the Committee of University Chairs (CUC) in the United Kingdom. In Australia a requirement to offer management development programs to board members is under legislation. The UK’s Leadership Foundation and the CUC have developed a formidable array of training and briefing materials as well as role descriptions, and templates and guidelines for performance review. The message that emerges is that membership of university boards now demands a professional approach and is not to be undertaken lightly.

The frequency of board meetings varies greatly between countries. In the United Kingdom, only three meetings a year were considered necessary but this number is gradually increasing. The AGB provides advice to new trustees in the United States and suggests that in public universities there be ten meetings a year of about 4 h each, while in private institutions there could be only four meetings but lasting an average of 7 h each.

A consequence of this professionalization is that in some countries, vacancies on university boards are carefully planned and skills audits of the existing membership are undertaken to decide what particular competencies are needed in any new appointments. Boards increasingly feel the need to be able to draw on legal, financial, marketing, human resource, and property expertise in their decision-making. This has inevitably led to the use of advertising or external search agencies to find board members, rather than traditional methods of word-of-mouth or personal contact. One other factor in appointments is the need to have a board that represents the society or region in which the university is located with appropriate ethnic and gender balance. These points make the management of the board’s recruitment and membership an increasingly complex matter.

The Chair of the Board or Council

The relationship between the chair and the vice-chancellor/rector is also a key element in an institution’s success. If that relationship does not work, there will be problems, since the chair of the council has to have full confidence in his chief executive and is in any event responsible for appraising his/

her performance in many countries. In legal terms, the chair is responsible to the government for the performance of the university and has to have an appropriate executive to ensure that the strategies and plans are put into effect. A direct consequence of making the university board responsible for the university’s performance is that the government will hold the chair accountable and not the vice-chancellor.

A key issue is how the chair of a board is selected, since so much responsibility now rests on the board. There is a wide variation in the method of selecting and appointing chairs of a council which are broadly similar to the ways of selecting board members. In Sweden and the Netherlands the government appoints the chair for a 3-year term, whereas in Ireland and the United Kingdom it is usual for the members of the board to agree on the appointment of one of them. Again, the role is unpaid and part-time, but since the time required can be significantly more than that of a normal member (up to 1 day a week in some circumstances), this represents a considerable commitment on the part of the chair.

The Search for Good Governance

A key issue for governments is whether they intervene at all to promote good governance by the university council and its executive. As the members of boards increasingly come to be drawn from all walks of society and the complexity of university operations grows, there is every reason for the state (or some buffer agency) to set up support systems for board members. There is consequently a general trend to the development of guidance and codes of practice for boards; recent examples of this have been produced in the United States, Denmark, Australia, Ireland, and the United Kingdom.

- In Ireland the Irish Universities Association and the buffer body, the Higher Education Authority, published a *Governance Code of Legislation, Principles, Best Practice and Guidelines* which was developed jointly by both parties and goes well beyond the realm of monitoring financial accountability.
- Another example is the set of National Governance Protocols for Higher Education Institutions developed by the Australian government in 2003.
- A code of voluntary good practice guidelines for boards and board members was developed by the Committee of University Chairs in the United Kingdom. This also sets out what it calls a *Statement of Primary Responsibilities* that describes the role of the board.
- In Denmark a committee on university boards produced *Recommendations for Good University Governance in Denmark* in 2003, which defined what the role of boards was, since they were a new feature of the system under the University Act 2003.

- The AGB in the United States offers a consultancy service for university boards and over 800 have turned to it for help; it also publishes a regular flow of advisory booklets aimed at trustees.

The state is becoming increasingly interested in how boards and universities exercise their autonomy and it seems very likely that their regulatory instincts will begin to make guidelines and codes of practice mandatory rather than optional.

Conclusions

The increasing complexity and scale of higher education provision is making it much harder for governments to exercise direct control over institutions and there is a global trend towards granting them greater autonomy in return for holding them accountable for the achievement of strategic goals.

This places responsibility for achieving the national goals for higher education squarely on the universities' boards and councils. In parallel, structural and process changes are occurring within these boards: they are becoming smaller; their membership is changing to contain much greater representation from external stakeholders; they are comprised of people with specialist skills rather than philanthropic amateurs; and together with their executive they are expected to master the setting of performance targets backed by strategic and operational plans. Being a member of a board is becoming a demanding and professional role, requiring specialist staff development and support. Governments are increasingly seeing it as their role to set guidelines for the performance of boards.

Bibliography

- Anderson, D. and Johnson, R. (1998). University autonomy in twenty countries. EIP Publication No 3 for 1998. Department of Education, Science and Training. <http://www.dest.gov.au/archive/highered/eippubs/eip98-3/eip98-3.pdf> (accessed September 2009).
- Breakwell, G. M. and Tytherleigh, M. Y. (2008). The characteristics, roles and selection of Vice Chancellors; *Summary Report*. Research and Development Series. London: Leadership Foundation for Higher Education.
- Committee of University Chairs (2004). *Good Practice in Six Areas of University Governance, a Report by CHEMS Consulting*. Sheffield, UK. <http://www.shef.ac.uk/cuc/pubs/CUC-CHEMS-Final-Report.doc> (accessed September 2009).
- Committee of University Chairs (2006). Template for role description of the Chair of Governing Body and template for role description of members of university governing body. *Sheffield, UK*. http://www.shef.ac.uk/cuc/pubs/A-Role_Chair-of-Uni.doc. http://www.shef.ac.uk/cuc/pubs/A-Role_Member-of-GovBody.doc (accessed September 2009).
- Fielden, J. (2008). Global trends in university governance. *Education Working Paper Series No 9*. Washington, DC: World Bank. http://siteresources.worldbank.org/EDUCATION/Resources/278200-1099079877269/547664-1099079956815/Global_Trends_University_Governance_webversion.pdf (accessed September 2009).
- HEA and IUA (2008). *Governance of Irish Universities: A Governance Code of Legislation, Principles, Best Practice and Guidelines*. Dublin: Higher Education Authority.
- Lao, C. and Saint, W. (2008). Legal frameworks for tertiary education in sub-Saharan Africa: The quest for institutional responsiveness. Pending publication.
- McGuiness, A. C. (2002). Reflections on postsecondary governance changes. Education Commission of the States. <http://www.ecs.org/clearinghouse/37/76/3776.htm> (accessed September 2009).
- Ministry of Science, Technology and Innovation, (VTU) (2003). Recommendations for good university governance in Denmark. *Report by the Committee*. Copenhagen: University Boards in Denmark.
- National University Corporation Law, Japan (2003). A new image of National University corporations. Details of the reform of Japanese state universities giving them autonomy as independent administrative corporations. <http://www.mext.go.jp/english/news/2003/07/03120301/004.htm> (accessed September 2009).
- OECD (2003). *Education Policy Analysis, Chapter 3: Changing Patterns of Governance in Higher Education*. <http://www.oecd.org/dataoecd/0/20/35747684.pdf> (accessed September 2009).
- OECD (2008). *Tertiary Education for the Knowledge Society, Volume 1: Governance, Funding, Quality*. <http://www.oecd.org/dataoecd/36/18/41303688.pdf> (accessed September 2009).
- Salmi, J. (2008). The growing accountability agenda: Progress or mixed blessing? *Paper Based on an Article in International Higher Education*. Washington, DC: The World Bank.
- Veld, R., Fussel, H. -P., and Neave, G. (1998). *Relations between State and Higher Education*. The Hague: Kluwer Law International.

Further Reading

- Bargh, C., Bocock, J., Scott, P., and Smith, D. (2000). *University Leadership: The Role of the Chief Executive*. London: Society for Research into Higher Education and Open University Press.
- Braun, D. and Merrien, F. (eds.) (1999). *Towards a New Model of Governance for Universities*. London: Jessica Kingsley.
- Edwards, M. (2003). *Review of New Zealand Tertiary Education Institution Governance*. Wellington, NZ: Ministry of Education.
- Richardson, G. and Fielden, J. (1997). *Measuring the Grip of the State: A Study of the Relationships between Governments and Universities in Selected Commonwealth Countries*. London: Commonwealth Higher Education Management Service/Association of Commonwealth Universities.
- Van Vught, F. (1994). Autonomy and accountability in government/university relationships. In Salmi, J. and Verspoor, A. M. (eds.) *Revitalising Higher Education*. Oxford: Pergamon Press.

Relevant Websites

- <http://www.agb.org> – Association of Governing Bodies of Universities and Colleges (USA).
- <http://www.shef.ac.uk> – Committee of University Chairs (UK), The University of Sheffield.
- <http://www.ecs.org> – Education Commission of the States (USA), Post secondary Governance Structures Database.
- <http://www.lfhe.ac.uk> – Leadership Foundation for Higher Education (UK).

HIGHER EDUCATION – RESEARCH AND SCHOLARSHIP

Contents

Funding of University Research
Mobility of PhD Students and Scientists
National Science Policy and Universities
Research Commercialization
Research Quality Assessment
The Management of University Research
The Training of New Scientists

Funding of University Research

G Harman, University of New England, Armidale, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Basic research – Research that is largely theoretical in emphasis and not focussed on particular industrial or social problems. Basic research is carried on mainly in major research intensive universities.

Block grant funding – Government funding to universities which comes in large allocations where universities have considerable say on its use. Increasingly block research funds are allocated based on assessment of institutional research quality and impact.

Innovation systems – Clusters of institutions involved in research and in R&D that together generate wealth based on the commercialization of intellectual property.

Intellectual property – Research outputs of commercial value that are protected by patents or copyright.

Patents – Legal instrument dealing with the ownership of intellectual property. Patents may be sold or owners may charge a licensing fee.

Priority setting – A process of establishing priorities with high ranked priorities attracting much large financial support.

Public research institutions – Government operated research bodies which are largely funded by Government. They carry out both basic and applied research.

Research commercialization – Process of transforming university discoveries and inventions in processes and products of commercial value.

Research quality frameworks – New approaches to the assessment of national research quality and impact.

Research universities – Universities that are strongly committed to research activities. Typically research universities have developed expensive facilities and employ large numbers of research only staff.

University rankings – Systems such as the Times Higher Education QS and Shanghai Jiang Tong that rank universities international primarily on the basis of the quality of their research.

Introduction

Universities are key elements in national innovation systems, carrying out extensive scientific and academic research, training future researchers and other skilled personnel, and generating and communicating new knowledge to students and the wider community. University research covers a wide range of academic fields and topic areas, adding to the stock of knowledge that guides new research agendas and feeds into PhD training. In addition, today substantial applied research is frequently undertaken by universities to serve the direct needs of industry by developing new products and industrial processes, and contributing solutions to social and economic issues (Harman, 2006).

A large proportion of university research activity is highly expensive, involving considerable staff time and high-level expertise, and often requiring the use of highly

expensive equipment and materials. For this reason, the level of internal and external research funding that universities can attract is of utmost importance in the production of high-quality research outputs, including scientific papers and discoveries and inventions that can be patented or otherwise protected under intellectual property legislation.

Research and the Research Functions of Universities

Scientific and academic research carried out in universities can be defined as critical and creative investigation undertaken on a systematic and rigorous basis, with the aim of extending knowledge or solving particular practical or theoretical problems. Extension of knowledge can be aimed at the discovery of previously unknown phenomena, development of explanatory theory and its application to new situations, and academic work that provides significant contributions to particular disciplines, tackles problems of social and economic significance, or produces original works of intellectual merit.

Since research activity varies considerably in terms of its disciplinary orientations, objectives, methodologies employed, and end products and their use, policymakers and academics find it convenient to make various distinctions, such as between basic and applied research, and between curiosity-driven and problem-driven research. Many countries now use Organization for Economic Development and Cooperation (OECD) categorization of research into pure basic research, strategic basic research, applied research, and experimental development (OECD, 2002), especially for data collection and reporting. While such categorization has utility in monitoring research performance and in discussions of research policy issues, it has been subject to criticism, especially as traditional boundaries are becoming increasingly blurred in many new multipartner and trans-disciplinary university research centers (Gibbons *et al.*, 1994).

Universities undertake research activities for a variety of reasons, but particularly important are strong academic commitments to the value of research and scholarship that are highlighted commonly in university charters and mission statements, and are integral elements of academic and disciplinary cultures. In addition, universities engage in research because of the recognition it attracts, its value in supporting teaching particularly at advanced levels, and as a means for providing service to the wider society.

Over the past half-century, most industrialized countries have defined the functions of universities in terms of threefold responsibilities to teaching, to research and scholarship, and to service (Clark, 1983). But over time, this definition of responsibilities has been somewhat redefined and elaborated. Particularly important has been a growing acceptance that in large, comprehensive, and

diversified higher education systems not all teachers should be required to conduct research as part of their employment conditions and not all higher education institutions should have an equal commitment to research. This has helped push increasing concentration of expensive and large-scale research in relatively small numbers of major research-intensive universities.

Apart from universities, other important research producers are public-sector research institutions (PRIs) and research laboratories operated by business firms. The balance of activity between these three different types of research performing bodies varies considerably between different countries and over time, depending particularly on private-sector industrial strength and research and development (R&D) capacity, and the extent of government support for public research. Funding for university research comes largely from governments, industrial partners, and various donors including alumni and philanthropic foundations.

Nations today generally face major challenges in a global environment where the role of knowledge is becoming increasingly important. Knowledge accumulation and application have become major factors in economic development and are increasingly being recognized as being central to the competitive advantage of nations. The ability of countries to select, adapt, and use knowledge is critical for sustained economic growth and improved living standards (World Bank, 2002). However, this capacity is not equally shared between nations, while scientific and technological knowledge yields its greatest benefits when there are strong linkages between leading universities and innovative companies.

The research functions of universities have changed to a major extent over the past two centuries. The classical European university concept of research-based teaching developed following the establishment by Wilhelm von Humboldt of the University of Berlin in 1810 and continues to be influential today, while the idea of the modern research university developed in the United States in the second-half of the nineteenth century. In the early part of the twentieth century, the research-university idea spread widely to many other industrial countries, with a strong emphasis being placed on the role of the university focusing primarily on basic research and research training, with some commitment to applied research but little to developmental research. The mission and fundamental values of the university at this stage were only moderately tied to the economy and employment of graduates. However, from World War II onward, demands made on scientific research for reasons of national defense and economic and social development brought universities more directly into contact with research users, leading to ongoing efforts to reform and redirect university research. While the classical-university idea retains considerable appeal within academia, increasingly modern universities are being forced to accept a wider research role and to become more directly involved with business, industry, and government (Geiger, 1993; OECD, 1998).

Research and the Role of Governments

Governments today play increasingly important roles in funding, stimulating, and directing research activity in universities. In doing so, they use a variety of policy instruments to achieve particular objectives, including allocation of block grants and specific purpose funds to institutions; funding research centers and individual researchers/research groups; establishment of major research centers of excellence; investment in large-scale research equipment and infrastructure; provision of economic incentives and disincentives (including subsidies, pricing structures, taxation concessions, and charges); regulation (such as legislation relating to intellectual property); the provision of information; and persuasion and advocacy.

Governments face new pressures and challenges related to their role in research funding. They need to respond to demands from a more diverse set of stakeholders, including not only the research community but also research users and interest groups representing particular types of consumers of research knowledge. Governments also need to take account of important changes that are taking place in research enterprises with new efforts to capture research benefits to meet social and economic needs while research costs are rapidly increasing with increased use of highly sophisticated and expensive equipment and other infrastructure support. A further challenge is a general shift away from an almost exclusive emphasis on disciplinary research toward more multidisciplinary research, directly responsive to societal needs and carried out with more interaction between different research performers. This trend has been described by Gibbons *et al.* (1994) as the shift from mode 1 to mode 2 research. Mode 1 research takes place within traditional disciplinary settings whereas mode 2 research emphasizes the importance of knowledge application and usually takes place in multidisciplinary settings. Other challenges include new accountability demands, the struggle to ensure long-term sustainability of research enterprises, and the need to maintain supply of highly trained PhD graduates. Supply of trained personnel depends on capacity in PhD training, the ability of individuals to secure PhD training abroad or attract foreign-trained scientists, and success in addressing problems of brain drain to other countries.

Governments have responded to these new challenges in different ways. According to a major report entitled, *Governance of Public Research* (2003), OECD member countries are responding to new challenges by enhanced stakeholder involvement in priority setting; restructuring research-funding arrangements by redefining responsibilities; combining agencies or developing new mechanisms of coordination; reviewing and renewing R&D funding mechanisms with a strong emphasis on use of competitive arrangements based on performance and merit;

undertaking major funding initiatives to strengthen infrastructure support; encouraging enhanced partnerships between universities, public research institutes (PRIs) and private firms; and reforming and restructuring PRIs.

Public Funding of R&D

Public funding of university research and R&D is a major instrument used by governments to steer science systems and to capture economic and social benefits more effectively. Many countries have embarked on reforms in funding in response to new demands and opportunities, enhancing their strategic-planning capacity and paying more attention to the social and economic environment and to the evolving patterns of relationships between stakeholders. Overall, the volume of R&D funding has increased, although public funding is generally increasing at a lesser rate than private funding (OECD, 2003).

Traditionally, governments financed a high proportion of university research in industrialized countries as a public good, but in the 1990s such funding declined with the result that universities were increasingly forced to seek new sources of financial support. Meanwhile, government funding increased for mission-oriented and contract-based research that are more dependent on output and performance criteria. This forced universities to perform more short-term and market-oriented research.

Almost all OECD countries in recent years have increased their public funding of R&D, although generally such increases have achieved little more than keeping pace with the expansion of national economies (OECD, 2003). As a share of gross domestic product (GDP), funding for R&D in universities and other PRIs remained flat at about 0.61% between 1981 and 2000 for OECD countries generally, although there were some major variations between countries (OECD, 2003). Nearly all countries reported their intention to increase funding for the future, but generally increases are expected to be mainly in priority areas and in new programs where funding is on a highly competitive basis.

Different types of government funding mechanisms are widely used, particularly institutional or block grants, project funding, and special programs funding, but in each case, there is increased use of competitive mechanisms and funding allocations based on performance. Each of these is discussed in some detail below:

- *Institutional or block grant funding.* This takes different forms, although in most countries traditionally it was based on student enrolments or the number of research units (or chairs such as in Japan). Generally, such funding came without conditions, although more recently a clear trend has been to separate institutional funding for research from institutional funding for teaching, and

for allocations for mission-oriented research funding on a competitive basis to increase while long-term general institutional research funding further declines.

Many OECD countries including the UK, Australia, and New Zealand use separate streams of institutional funding for research, with allocations being based on quality and/or performance. The provision of government funding for both institutions as well as individual researchers/research groups is often referred to as a dual system of research funding. While Australia has used simple performance indicators (external research grants, higher degree completions, and publication outputs) to allocate institutional research funding to universities, the UK higher education funding authorities have used data generated by the research assessment exercises (RAEs) run at intervals of 4–5 years. More recently, New Zealand has opted for a modified version of the UK RAE based on assessments of research quality conducted every 4 years by some 70 different panels of experts. While throughout the 1990s, the Australian research quantum scheme allocated about 5% of total operating grant funding to institutions for research support on the basis of performance indicators, the Higher Education Funding Council of England allocated some 20% of total government funding on the basis of RAE assessments, resulting in leading research universities (e.g., Oxford, Cambridge, and Imperial College) gaining much larger amounts from their research allocations than their teaching allocations (Harman, 2000). Currently, both the UK and Australian systems are under review. While the UK intends to abandon the RAE after the 2008 assessment, Australia plans to move to a research-quality framework, which will involve separate assessments of the quality and impact of research in each university in 12 different major disciplinary areas.

- *Project funding.* This involves competitive allocations on the basis of applications submitted in response to notifications or calls for tenders, with evaluation using some kind of peer-review process. In some senses, project funding is similar to business funding of R&D in that it tends to be contract based, with specific objectives and milestones. In a number of countries, project funding has been expanded to include projects with joint government and industry support.
- *Special programs.* These are becoming increasingly common. In general, they are linked to priority areas with funding being allocated on a competitive basis, often for centers of excellence, or special research centers involving universities and other partners. Centers of excellence have been established in many Asian Pacific countries including Japan, Australia, and New Zealand. Japan launched a new university resource-allocation-prioritization program in 2002 called the twenty-first-century COE (center of excellence)

program with the aim of promoting research units of world-class excellence in selected fields. The fields supported in 2002 were life science, chemistry and materials science, information, electrical and electronics, humanities, and interdisciplinary subjects. Each research unit selected was allocated resources around JPY 100–500 million for 5 years. In November 2002, some 113 research units at 50 institutions were selected out of 464 applications from 163 institutions. Australia has programs supporting special research centers, key centers of teaching and research, cooperative research centers (multisite centers jointly funded by government and industry) and a small group of mega-centers in strategic areas such as biotechnology and information and communications technology. In 2001, New Zealand established a center of research excellence fund to support world-class centers expected to be involved in both research and knowledge-transfer activities (Ministry of Education, 2001).

In a number of advanced countries, an important trend is for increased proportion of R&D to be financed and performed by business. However, in Japan, business support for higher education and public research institutions has increased only slightly but is still relatively small, while in Korea business funding for higher education research has decreased. However, this reduction has been compensated by increased funding from the government, with an increase over last two decades of about 100%. Other funding for research comes from universities' own resources, endowments, and patent licensing fees. In Japan and Korea, 5% or more of research funding comes from other sources.

Across advanced regional economies governments are increasingly linking funding allocations to evaluation and assessment. Detailed assessments sometimes are made prior to new initiatives while ongoing assessments of performance are increasingly common. Traditionally, evaluation procedures were mainly based on the use of peer review of project applications but governments now are using in-depth reviews and various performance indicators, such as total external funding attracted, and numbers of publications, patents, start-ups, awards, and prizes.

Concentration, Selectivity, and Priority Setting

Developed nations generally are putting increased emphasis on concentration of research funding among a limited number of institutions, departments, and researchers in order to develop larger groups of highly qualified and productive researchers who have access to superior equipment, facilities, and technical and financial support. This process is often referred to as achieving a critical mass of

researchers. Even wealthy countries recognize that research resources are limited, and therefore pursue such policies of selectivity and concentration.

Priority setting refers to processes of strategic choice with the aim of increasing returns on investment in research (Harman, 2000). Some fields of research or particular research centers or research projects are selected over others to receive preferential funding. Many governments are moving increasingly to setting national research priorities. In some cases, priority setting is driven by reductions in government budgets but more commonly priority setting is favored to make decisions when governments face strong pressures to provide larger allocations to particular areas, such as biotechnology, nanotechnology, or information and communications technology.

Governments use a variety of priority-setting mechanisms, including national science and technology plans, advisory bodies, and foresight processes and public consultation, while universities tend to depend on strategic plans, research management plans, and particular competitive-funding mechanisms. Some countries (such as Japan, Norway, and Hungary) use a top-down approach where central governments adopt explicit strategies, policies, or plans that specify priority areas for research. In these countries, plus others such as Korea, Germany, Netherlands, and Denmark, central advisory bodies make recommendations about priorities. At the other end of the spectrum, is the bottom-up decentralized approach such as in United States, Canada, and Sweden where a number of advisory bodies provide recommendations to different government agencies.

Quality Research and Research-Quality Frameworks

Governments are becoming increasingly concerned not only to build up their R&D capacity but also to ensure that university research outputs are at the highest possible international standards, and are also highly relevant to local industry and business. Assessments of quality are based mainly on peer review or the use of quantitative data or performance indicators. Peer review is a generic term that includes the notions of merit review, refereeing by anonymous referees, expert judgment, peer advice, and peer evaluation. However, in making judgments about research, it is often desirable to take into account the judgments of a range of different research users who should be able to verify claims by researchers about quality and impact. Quantitative data on research output and quality can simply rely on lists of publications, conference papers and presentations, and patents. However, for comparative analysis, the measures used most commonly are bibliometrics comprising publication numbers and associated citations. Citations are a measure of the

impact of particular research papers on the disciplines involved by assessing the number of times a particular article is cited (or referred to) by other authors.

A number of governments in developed countries have established research-quality frameworks that serve a number of purposes, but most important they provide consistent and comprehensive approaches to assessing the quality and impact of research activities in universities, particularly publicly funded research. Such assessments can: inform governments and the wider public about the quality and impact of research and its standing internationally; guide resource allocation, especially the allocation of major block research grants; and provide inputs into policy development, especially with regard to research priorities. While many countries use data from quality assessments to determine funding allocations, some countries do not. Examples of the latter are the Netherlands national evaluations system for publicly funded research, and the German research foundation rankings.

Apart from the quality frameworks developed by governments, both research groups and major newspapers have developed methodologies to rank universities internationally. Rankings developed by the Institute of Higher Education of Shanghai Jiao Tong University rely largely on quantitative research indicators, whereas others such as those of the UK *Times Higher Education Supplement* and the Melbourne Institute of Applied Economic and Social Research use wider definitions of quality and include other performance measures related to teaching and student characteristics.

University Research Links with Industry

Universities in many countries are establishing closer and more effective links with other research providers and stakeholders, particularly PRIs, industry laboratories, business firms, and government laboratories. These links take a variety of different forms, including joint research centers and research appointments, shared use of facilities, industry funding of university research, and consultancy arrangements between universities and research users.

In the United Kingdom, for example, a rapid increase in university-industry collaboration since the 1980s has led to a variety of different partnership arrangements with many positive outcomes including an impressive increase in the number of joint scientific publications. By the late 1990s, joint university-industry scientific papers accounted for about half of all industrial scientific output (Calvert and Patel, 2002). These new arrangements have increasingly broken down traditional arrangements in modern economies where universities and PRIs were viewed as being responsible for basic scientific and precommercial research, while industrial firms performed the bulk of applied research and product development.

On the whole, the new university–industry research links in many leading industrialized countries are working well to the mutual benefit of various partners. University links with industry provide universities with substantial research support, consulting opportunities, financial and project support for postgraduate students, opportunities for graduate employment, and opportunities for academics to gain insights into new technical developments within industry. On the other hand, industry benefits substantially from partnerships through access to university expertise and facilities, access to university intellectual property, and the supply of well-trained graduates.

Particularly important partnerships in a number of countries are based on new research centers with multiuniversity partners as well as partners from PRIs, government departments, and business firms. An example in the Asia Pacific region is the Australian Cooperative Research Centre (CRC) program that has resulted in establishment of some 70 multisite centers. Funding is provided by the Australian Government as well as by partners including universities, government research organizations, government ministries, and industry and business.

While governments, universities, and the researchers involved in partnerships are generally supportive of university–industry partnerships, critics allege that such partnerships threaten traditional academic values, distort the balance between basic and applied research, and corrupt academics with commercial values to the extent that in some cases there is serious neglect of teaching and research responsibilities. It is also alleged that industry contracts lead researchers to withhold scientific information from colleagues and delay publication, thus adversely affecting the free flow of scientific information. While some academic studies have highlighted what they see as dangers in close relationships between universities and industry, particularly the impact on academic work and values, forcing scientists to abandon the traditional cooperative mode of research (Dickson, 1984; Kenney, 1986), other studies have provided strong evidence that many academic researchers increasingly accept the concept that profit generated from research need not corrupt academic work, and conclude that, to date, any adverse affects on academic behavior have been limited. Particularly important in addressing criticisms of the new commercialism have been the American studies by Blumenthal and colleagues. One study (Blumenthal *et al.*, 1986) reporting on a survey of 1200 biotechnology researchers in 40 major American universities found that researchers with industrial support publish at higher rates, patent more frequently, and participate more substantially in administrative and professional activities. At the same time, researchers with industry funds are much more likely than other biotechnology researchers to report that their research has resulted in trade secrets and that commercial considerations have influenced their choice of research projects. Similar findings were found in

an Australian study (Harman, 1999) which reported that compared to other colleagues, industry-funded science and technology academics in leading Australian universities had better publication records, spent longer hours at work each week, and allocated more time on postgraduate teaching, administration, committee work, and interaction with colleagues.

Research Commercialization

Since the early 1980s, first in America, and more recently in many other developed countries, governments and research-intensive universities have put considerable effort into the protection and commercialization of intellectual property of commercial value arising from university-generated inventions and discoveries. The process of commercialization is usually referred to as research commercialization or technology transfer. Intellectual property is a broad term used to refer to the various rights that the law gives for the protection of creative effort, especially the protection of economic investment in creative effort through experimentation, discoveries, or activities of the mind.

Rapid growth in research commercialization efforts has been driven by recognition of the value of intellectual property and by the desire of universities to generate additional income. In addition, governments and universities have become increasingly concerned with national competitive advantage in the age of globalization while universities seek to demonstrate their commitment to the public interest and through enhanced relationships with business firms generate increased political support for public investment in research. Governments, on the other hand, seek to capture the benefits of university research to facilitate economic and social development, and to generate national wealth.

Laws and policies governing the ownership of intellectual property vary considerably between countries. In summary, a number of countries follow United States practice where generally intellectual property arising from federally funded projects is held by universities, but accompanied by arrangements for academics and their departments to share in the financial returns from commercialization after all costs incurred have been met. The other model is that found in Sweden and a number of continental European countries, where individual academic creators own the intellectual property resulting from their university research except for work covered by contractual agreements with firms.

A small number of the leading universities in industrialized countries generate considerable revenue each year through research-commercialization efforts, particularly licensing of patents and creation of university start-up companies based on university discoveries or

inventions. For example, in the US, the University of California system generates about US\$180m per annum and Stanford University about US\$50m per annum in research-commercialization revenue. However, for many universities, licensing of patents and income from spin-off companies is likely to constitute no more than relatively small proportions of total university research financial support.

See also: Higher Education and the Knowledge Society; Internationalization of Higher Education; National and International Rankings of Higher Education; National Science Policy and Universities; Research Commercialization; The Economics and Finance of Higher Education.

Bibliography

- Blumenthal, D., Gluck, M., Seashore Louis, K., Stoto, M. A., and Wise, D. (1986). University–industry research relationships in biotechnology: Implications for the university. *Science* **232**, 1361–1366.
- Calvert, J. and Patel, P. (2002). *University–Industry Research Collaboration in the UK*. Brighton: SPRU, University of Sussex.
- Clark, B. R. (1983). *The Higher Education System: Academic Organization in Cross-National Perspective*. Berkeley, CA: University of California Press.
- Dickson, S. (1984). *The New Politics of Science*. New York: Basic Books.
- Geiger, R. (1993). *Research and Relevant Knowledge: American Research Universities since World War II*. New York: Oxford University Press.
- Gibbons, M., Limoges, C., Nowotny, H., et al. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: Sage.
- Harman, G. (1999). Science and technology academics and university–industry research links in Australian universities. *Higher Education* **38**(1), 83–103.
- Harman, G. (2000). Allocating research infrastructure grants in post-binary higher education systems; British and Australian approaches. *Journal of Higher Education Policy and Management* **22**(2), 111–126.
- Harman, G. (2006). Research and scholarship. In Forrest, J. F. and Altbach, P. (eds.) *International Handbook of Higher Education*, pp 309–328. Dordrecht: Springer.
- Harman, K. (2000). Priority setting and resource allocation in Australian biomedical research: Muddling through with some skill. *Prometheus* **18**(4), 373–390.
- Kenney, M. (1986). *Biotechnology: The University Industrial Complex*. New Haven, CT: Yale University Press.
- Ministry of Education (2001). *The Centres of Research Excellence Fund*. Auckland: Ministry of Education.
- OECD (1998). *University Research in Transition*. Paris: OECD.
- OECD (2002). *Frascati Manual*. Paris: OECD.
- OECD (2003). *Governance of Public Research: Towards Better Practices*. Paris: OECD.
- The World Bank (2002). *Constructing Knowledge Societies: New Challenges for Tertiary Education*. Washington, DC: World Bank.

Further Reading

- Bozeman, B. (2000). Technology transfer and public policy: A review of research and theory. *Research Policy* **29**, 627–655.
- Etzkowitz, H. and Peters, L. (1991). Profiting from knowledge: Organisational innovations and the evolution of academic norms. *Minerva* **XXIX**, 133–166.
- Nowotny, H., Scott, P., and Gibbons, M. (2001). *Rethinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge: Polity.
- Slaughter, S. and Leslie, L. (1997). *Academic Capitalism: Politics, Policies, and the Entrepreneurial University*. Baltimore, MD: The John Hopkins Press.
- Ziman, J. (1994). *Prometheus Bound: Science in a Dynamic Steady State*. Cambridge: Cambridge University Press.

Relevant Websites

- <http://www.arc.gov.au> – Australian Research Council.
- <http://www.dest.gov.au> – Department of Education, Employment and Training (Australia).
- <http://www.rcuk.ac.uk> – Higher Education Research Councils, UK.
- <http://www.mext.go.jp> – Ministry of Education, Culture, Sports, Science and Technology, Japan.
- <http://www.morst.govt.nz> – Ministry of Research, Science and Technology (MoRST).
- <http://www.nhmrc.gov.au> – National Health and Medical Research Council (Australia).
- <http://www.nih.gov> – National Institutes of Health.
- <http://www.nsf.gov> – National Science Council.
- <http://www.nserc-crsng.gc.ca> – Natural Sciences and Engineering Research Council of Canada.
- <http://www.sshrc-crsh.gc.ca> – Social Sciences and Humanities Research Council of Canada.

Mobility of PhD Students and Scientists

S Avveduto, National Research Council, Institute for Research on Population and Social Policies, Rome, Italy

© 2010 Elsevier Ltd. All rights reserved.

Mobility of the highly skilled can be approached from different points of view: as a human resource policy question, as an equity problem between world regions at different levels of development, or as a migration-policy topic, just to mention the most relevant ones. Human resources are the crucial element for development, and all investments in their education and positive utilization are considered essential for the general well-being of society and the economy. In this framework, the internationalization of education and research-career patterns are a basic component in order to enhance the development of a better workforce – a workforce particularly devoted to research, science, and technology, and specifically made up of PhD graduates and scientists. Mobility is unanimously considered an efficient means of knowledge diffusion and technology transfer. The border between mobility and migration for the highly skilled often becomes less evident.

According to most scholars, comparing the mobility of brains to an escape from one's country of origin, is no longer considered in our globalized society. Rather, we refer to it as the free circulation of people and ideas and not as a drain of intelligence.

Mobility of skilled people and specifically of scientists, researchers, and PhD students and graduates is a complex phenomenon and can be considered from different perspectives. It is a typical two-sided question. Positive and negative aspects can be found in an event that could apparently be considered neutral. An excess of mobility is likely to cause as many problems as its absence.

The need to confront, cooperate, and work with other scientists is somehow innate in the research activity itself. The question is to determine the optimal share of mobility in a career, in an institution, or in a country. If this share is either exceeded or not reached, problems may arise. Of course, there is no standard tool that can help in quantifying the optimum level. However, the analysis of positive and negative effects of mobility can help in choosing its right balance.

At an individual, institutional, or country level, the advantages of having a mobile highly skilled workforce and of cooperating, studying, and working in a different environment with foreign partners, are rather obvious and the positive potentialities of this experience are self-explanatory. The risks lie essentially in missing the opportunity to fully develop the positive spillover of this experience or in the changing nature of mobility – from a temporary experience to a permanent escape.

When considering scientists' and PhD students' mobility, we refer to an occurrence which has its roots well into the Middle Ages. The very first mobility in this sector may be found in university students who were in reality, the intellectual elite of that period. They were known as *clerici vagantes* and they used to move within and across nations in order to attend the best *universitates*. Such universities, each had their own area of specialty where the elite students could find the best knowledge available. Much, if not all, has changed since then, but the essence of moving around to find the best options, to work and learn, to come across new ways of thinking, to confront other cultures and ideas, and to understand what is happening outside the borders of our own intellectual village, is the essence of modern science and higher education. This desire for good research and knowledge diffusion has remained virtually unchanged throughout the centuries. The mobility of highly skilled personnel has taken place with relative continuity, in different forms and with different consistencies, since the mid-nineteenth century into the early twentieth century, when a proper science and research career could be acquired. A number of elite migrants, including professionals and scholars, left their home countries for foreign destinations in order to start new businesses or enter cultural and scientific activities. We can recognize the importance of skilled immigrants in several countries and fields of activities – in firms, research laboratories, and universities. The flow of people has been influenced by many factors that, before and during World War II, also included political or racial persecutions. Such conditions, for instance, forced many scientists to migrate to the United States from Europe. In this case, the term brain drain rather than mobility was used. This expression first appeared in the late 1950s and was used for the first time in an official document by the Royal Society in the 1960s in order to explain the movement of British scientists from the country (Royal Society, 1963). Brain drain refers to an excess of flow of talented people who are induced to leave their home country to find better scientific working conditions.

Nowadays, the percentage of mobility is considered one of the indicators of a healthy and innovative system of higher education and science. Many international agencies and supranational bodies (UNESCO, 1998; OECD, 2001; European Commission, 2007) state that both government and institutions should promote national and international mobility of PhD students and scientists as an essential element in enhancing the quality and relevance

of those systems. However, the problems associated with brain drain are still not over. This is especially the case if applied to the drain of talent from less-developed countries, whose return to home countries after a mobility experience is very problematical (Thorn and Holm-Nielsen, 2006). From 1990 to 2000, 5 million tertiary-educated adults moved from a less-developed nation to a more developed one, while 2 million moved between more developed nations (OECD, 2007a). The International Organization for Migration (IOM) estimates that over 300 000 professionals from African countries live and work in Europe and North America, and a large number of professionals, especially health workers, intend to migrate from their home countries (WHO, 2004).

In absolute terms, the largest stock of educated emigrants are from Europe, Southern and Eastern Asia, and, to a lesser extent, Central America. Nevertheless, as a proportion of the local educated labor force, the highest rates are found in Central America and Africa. Between 1990 and 2000, these regions also experienced the strongest increase in brain drain (Docquier and Marfouk, 2004).

The magnetism of some countries for researchers and PhD students and graduates over others is indisputable. The United States has always been and still is an important pole of attraction for skilled people. In fact 40% of its foreign-born, adult population have a tertiary-level education. Since the early 1990s, over 1 million highly skilled professionals, from India, China, Russia, and a few Organization for Economic Cooperation and Development (OECD) countries such as Canada, the United Kingdom, and Germany, migrated to the United States under the H-1B temporary-visa program.

The direction of flows of highly skilled people benefits most of the OECD countries, and recent figures (Figure 1) show that only a few countries: Mexico, Korea, Ireland, Poland, Finland, and some Central and Eastern European countries, experienced a net loss of people. The rest of OECD countries experienced a net inflow (Dumont and Lemaitre, 2005).

As for mobility of PhD students, the figures on foreign participation in science and engineering (S&E) doctoral education in the United States are often used to obtain indicators of global flows, the United States having the strongest attraction for PhD students wishing to complete their education abroad.

The National Science Foundation (Oliver, 2007) states that after 2 years of decline, US enrolment of foreign graduate students in S&E fields increased in 2006. The increase was largely due to first-time, full-time enrolment of foreign students, which grew 16% over the 2005 level.

From 1989 to 2003, foreign students earned nearly 40% of US S&E doctorates, with Asian students representing about 55% of this group. Students from European Union (EU) countries have totaled about 10% of all foreign doctorate recipients in the United States.

The largest Asian sources were China with about 34 000 S&E doctorates from 1989 to 2003, Taiwan with 14 800, and South Korea and India with about 14 500 each. These four sources accounted for nearly 90% of all Asian recipients of US S&E doctorates. Asian recipients of US S&E doctorates are far more likely than EU students to earn an engineering doctorate: 35–38% compared with 17–21% of EU students, while the US average is 13–15% (NSF, 2007).

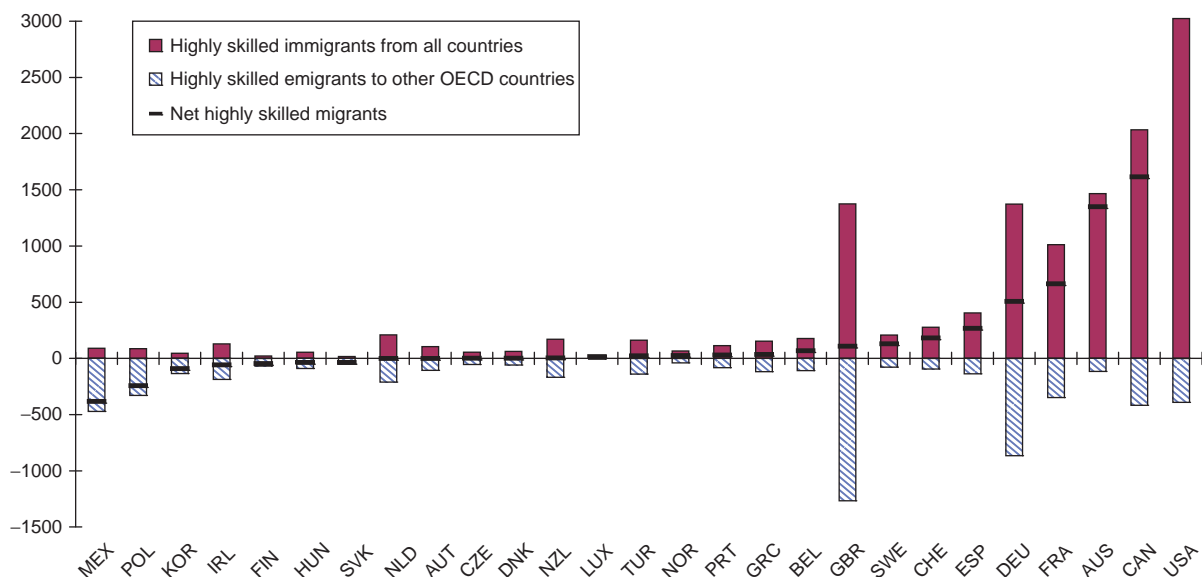


Figure 1 Immigrant and emigrant population 15+ with tertiary education in OECD countries – thousands. From Dumont J. C. and Lemaitre, G. (2005). Counting immigrants and expatriates in OECD countries: A new perspective. *OECD Social Employment and Migration Working Papers*, N. 25. Paris: OECD.

PhD Students' and Scientists' Mobility: Main Questions

Mobility occurs essentially as an individual choice but can be encouraged and favored by institutional or national programs and regulations that are aimed at promoting the circulation of researchers and scholars.

All OECD countries have a variety of policies to encourage PhD students and researchers to be mobile either at a national or institutional level. Such policies include monetary incentives and are targeted toward specific groups and at different scales.

Mobility of PhD Students

The mobility of PhD students faces questions of opportunity and feasibility of international research training, within a research-education process. Objectives and benefits of international mobility are to be set and barriers to internationalization considered.

The opportunity of having an experience abroad is often considered, both by students and professors, as highly desirable, and perceived as a value per se. The research program they carry out may require a visit to a foreign university to collect data or perform experimenting in structures not available at home. But to these formal attainments, some informal, and not easily quantifiable, ones should also be added.

The main achievements are linked to personal, cultural, and scientific growth that PhD students can acquire when exposed to a different scientific and training milieu. Objectives and benefits of international mobility therefore expose PhD students to new scientific and cultural contexts, allowing broader professional and human experience and consequently improved competences and capacity to integrate in new job situations.

Converging research results point out that both concrete results, such as help with dissertation work, and intangible ones, such as personal development, are regarded as valuable outcomes for PhD students studying abroad (Kyvik *et al.*, 1999; OECD, 2001).

The Organization and Structure of Exchange: Motivations and Procedures

A wide variety of different procedures affects the PhD students' choice to go abroad and organize a study period in a foreign university. It is not a single decision-making process which leads the student to a chosen destination. A well-organized process can make doctoral students well aware, from the very beginning of their PhD course, of the opportunity to carry out part of their research abroad. However, it may also happen that only personal initiatives of professors, in advising and guiding students, or of

students themselves, in making applications and independent arrangements, prevail.

Although organizational models differ greatly, the motivations and procedures for choosing the country to study in, and the best period to go there, tend to converge. An international experience is perceived to be most fruitful if it takes place not at the very beginning of the doctoral education but after at least the first year. It may be advisable to spend the first year at home to develop acquired knowledge, to prepare the research proposal properly, and start working under the supervision of a supervisor.

At the final stage of preparation of the dissertation, it is as well perceived that staying in the home country university may be preferable. During the intermediate year(s) of a doctoral course after having consolidated the initial stage, students can benefit from coming to terms with an international context.

As for the country of destination, in many cases, the choice points toward a university which has a tradition of joint research with the *alma mater*, which may also simplify arrangements and eliminate some difficulties.

Regarding the optimal duration of the study period abroad, a variety of different opinions may be considered. Respondents to direct surveys agreed that a period of 6 months is the best one, since it allows for an initial period of acclimatization, while leaving a reasonable amount of time to draw academic benefit from the experience.

The average period of the stay abroad varies from 3 to 6 months for scientific and art subjects. Medicine presents a longer rate of staying (around a year), while engineers tend stay abroad shorter periods (around 3 months).

The various choices are consistent from different perspectives if the study period abroad is either considered useful to strengthen research work already advanced, or part of a less-developed study and research activity maybe more learning oriented, which would thus in many cases include things such as coursework thought in sessions with a supervisor, private tutorials, if appropriate, and lectures.

Problems and Obstacles May Hamper a Proper Mobility Experience

The problems to face may be objective – that is, mainly of an economic, bureaucratic, and logistic nature – or subjective – that is, knowledge of languages, adapting to the new environment, interaction with different subjects and styles of study. Difficulties also arise in delays in work on the thesis, in relationships with professors of the hosting university, or at home with a lack of motivation and support in a too nationally oriented environment.

The role of the university can be very important in fostering international mobility; helping students to cover the expenses and disseminating information widely, are the first relevant objectives to be met. The value of an international experience is to be considered highly relevant in a

proper research training, and so universities and research institutions are increasingly committed in ensuring that the students have this opportunity.

Mobility of Scientists

Programs for researchers and scientists are typically directed at young researchers with the aim of providing them with a specific mobility program that covers part of their relevant expenses. During the middle of an expatriate's research career, such policies are used to attract them back to their home countries. Many policy tools are set up to provide direct incentives to foreign researchers to encourage them to choose a specific country or institution as the preferred place for their mobility experience. Countries such as Korea, for instance, have provided their scientists and researchers living abroad with attractive conditions for returning home, creating positions for them, and offering adequate structures and salaries. The know-how acquired abroad can be beneficially used to serve the home country, and a positive circle of knowledge transfer can be set up. The main flows consist of information and communication technology (ICT) specialists, engineers, and researchers.

At a supranational level, we can point out, for instance, the Marie Curie Actions of the EU in providing European Reintegration Grants (ERG) and International Reintegration Grants (IRG) to researchers. Within the EU, special norms have been developed in order to ease the entry of foreign researchers who wish to carry out scientific research within the EU. The European Commission (EC), for instance, introduced a directive (Directive 2005/71/EC) to regulate the procedures of special scientific visas.

The researchers' visa package contains admission procedures of and visa facilitation for third-country researchers entering the EC to obtain a special permission to enter, stay, and work in the EU in order to carry out scientific research. This allows researchers to exceed the foreigners' quotas in a single state and to pass quickly through the immigration authorities' procedures.

In the United States, a series of special visas (the H-1B visa) allow the entrance of temporary specialty workers, among which scientists or immigrants with extraordinary ability in sciences, arts, education (and also business or athletics, the O-1 visa), are included. The figures on immigration through these visas are currently used to check the flow of highly skilled workers among countries and particularly toward the United States (Regets, 2007).

Universities have played a main role in promoting mobility and fostering what has been called academic migration (Altbach, 2004). Each single university may have its own policies and programs addressed to different kinds of graduate and postgraduate or professors and scholars and offer many types of mobility opportunities.

These may vary from the short-term leave to a year-round commitment.

Mobility schemes per se do not necessarily encompass a high level of mobility to be obtained by a country. However, it is clear that without these schemes, entering into a mobility experience proves to be quite difficult. Studies have proved that the absence of a mobility scheme is an obstacle to mobility. As an EC Green Paper stated: "It should be stressed that the obstacles to mobility are necessarily more difficult to overcome for those who 'spontaneously' seek to undertake training in another Member Country" (European Commission, 1996).

Mobility Indicators and Measurements

Indicators of stock and flow of mobility include the following: (1) occupation mobility, such as a demand for specific skills or flexibility of skills; (2) sector mobility, that is, changes from one sector of activity to another such as from a university to the private sector; and (3) international mobility, from one country to another. The last one can encompass the two previous types of mobility because changing country can also mean changing occupation or one's sector of employment. It is likely, for instance, that a researcher working in a university in his own country moves to the private sector in another.

The measurement of mobility of scientists is not an easy task. This is particularly the case for international mobility, as the indicators to track the outgoing personnel are far weaker than those to track ingoing ones. We should rely mostly on the latter as, when leaving an institution or a country, it is not at all a common practice to inform the former employer or a central agency of the directions of the next move either in terms of employment destination or the foreign country which the employee is planning to travel to. Barring a few countries (e.g., the European Nordic ones that can rely on a very detailed and updated register of all employment changes for each citizen), most countries do not have a central register that can be used to track scientists and PhD graduates' movements across professions, institutions, sectors, and countries. The relevant information is kept by each source separately and a cross-comparison is always very difficult to make.

As for international mobility, it is more likely that one will find information and figures on scientists in the incoming country rather than the country of origin. The country of destination, due to immigration regulations, might have a more detailed record of all skilled immigrants where relevant data on scientists can be found.

Tracking mobility of PhD students is, in principle, an easier task especially if they move within a specific structured program. Each university and research institution they are studying in, keep track of PhD students' movement, and the trajectories under their choices can

be visible at this level. A thorough picture of the phenomenon is, also in this case, more difficult to draw.

Reasons to Move

International mobility is marked by different reasons that encourage movement, push-and-pull factors inducing people to leave their institution, sector, or country and driving them toward different places that are thought to be more attractive.

Pull factors may be identified in the offer, available in the countries, of grants and scholarships that multiply the number of positions. Push factors include direct government or institutional policies created by the sending countries that stimulate the demand as well as programs specifically directed to provide funding and facilities.

To measure mobility reliability and validity and to fully develop the right indicators to do this, it is necessary to analyze the various dimensions relevant to individuals, institutions, and nations. These include: the number of people involved; level of qualification; gender; age profile; length of time, duration, and time interval between the different experiences; type of motivations; push-and-pull factors; barriers and incentives; economic and scientific impact; benefits and losses in personal and institutional terms; and policies to allow individuals, institutions, and countries to benefit most from the mobility experience.

A number of different factors either pertaining to the public or the private sphere influence international mobility. Why people decide to move and why they choose one destination over another remains largely part of individual choice but some motivations are common and some incentives can be determined.

The general framework is influenced by legislative regulations concerning immigration in the host country and general migration policy of the country of origin that can facilitate or hinder the mobility choice.

Working conditions in the country of destination have an important influence on the decision to move. Countries with poor working and/or studying conditions are, of course, less attractive. However, countries with good working/studying conditions may also become unattractive due to factors such as high cost of living, difficulties in settling (finding adequate housing, etc.). The entire set of factors to be considered should certainly take into account personal reasons, such as family ties, a milieu of reasons, such as cultural circumstances, or standards to be met, such as language requirements and qualifications needed. But wider facets such as policies and accessibility to mobility schemes; supply and demand of scientists and researchers in the country of destination; salary and pay-scale levels; qualifications required; and career prospects are also highly relevant.

To acquire a more precise picture of a PhD graduate's profile across countries, and to track their moves, either

as employees or geographically, OECD, in collaboration with United Nations Educational, Scientific and Cultural Organization (UNESCO) and European Statistics (EUROSTAT), launched an international review. This review desires to contribute to the understanding of the scale, directions, and drives of international mobility of PhD graduates and scientists. The international project on Careers of Doctorate Holders (CDH) has thus been set up and should provide comparable data on these groups and their international mobility. The project is at present gathering figures and data on researchers in all stages of their career (OECD, 2007a). The OECD–Group on Steering and Funding of Research Institutions (SFRI) have collected data and policy information at the institutional level, including institutional practices, and on foreign (temporary and permanent) researchers working in the public research sector to complement the picture in member countries (OECD, 2007b).

Many attempts have been made to come up with a reasonable picture of mobility by using direct survey tools that can give a more accurate picture of trends. Such direct surveys allow for an in-depth investigation of the reasons and outcomes of mobility.

As a study carried out for the EC on mobility of scientists points out (Hansen, 2003), the professional factors that drew migrants to foreign countries (i.e., pull factors) have been defined as follows: a large proportion of interviewees stated the possibility of finding better opportunities to work at the forefront of scientific research in their field or to acquire new scientific training and to specialize in a field that was insufficiently developed in their home country. But other factors may be considered pertaining to the personal sphere. Many researchers mentioned the desire to have more freedom in their work and personal life as an important factor in the decision to move (Figure 2).

<i>Reasons</i>	<i>Country EU 25</i>	<i>Country USA</i>
Career advancement	87.9	51.9
Employer reputation	74.0	61.1
Access to leading edge technologies	73.3	29.6
R&D funding	69.8	24.1
Professional networking	67.7	46.3
Employment/business opportunities	56.1	25.9
Salary	54.0	18.5
Adventure	48.8	53.7
Education	45.6	20.4

Figure 2 European and US Scientists' main reasons to move abroad from their home country (percentage of positive answers to the survey questions). From Hansen, W. (2003). Merit survey on flows of qualified scientists. *Final Report Project: Brain Drain Migration Flows of Qualified Scientists*. Brussels: European Commission.

The scientists share both positive and negative evaluations of their mobility experience. The positive ones prevail because mobile scholars feel they are in high-level working environments that produce good-quality results and offer resources and access to scientific equipment. The negative evaluation is commonly due to obstacles linked to bureaucratic red tape and difficulties in obtaining work permits and residence visas.

Postdoctorals and young researchers also complain of the fear of losing opportunities at home and of leaving them to people who preferred staying in the home institution even if at a lower level.

The Timing and Effects of International Mobility: Positive and Negative Outcomes

International mobility includes both short- and long-term experiences and its timing and duration are relevant to a scientist's career.

In the initial stage, mobility commonly takes place just after the completed PhD and may be the first step in a research career.

Immediately after graduation, the possibilities of considering moving out of the home country for a longer period, are higher than for graduates leaving the home country at a later stage. These graduates already present in the host country may be interested in pursuing or accepting a job offer and thus starting a career in the host country. As stated in the NSF studies, the propensity to stay in the host country (the United States in this case) and accept a solid job offer is relevant (NSF, 2006). The mobility experience may, however, take place at a later stage and its meaning and influence on the career choices may be far different.

One obvious benefit of receiving one's PhD abroad is also the much higher probability of getting employment in the country where the PhD has been awarded. In the

United States, one of the most attractive countries to complete a PhD, EU citizens account for 25% of doctorates awarded. This trend has continued since 1991 to the early years of 2000 and places them between 3% and 4% of all those awarded in the United States. The share of employment offers that European PhD graduates obtained is very relevant. In the mentioned period, 45% of graduates planned to stay on in the United States as they received a firm employment offer in the country (Figure 3).

Direct and indirect effects of international mobility can be registered on science, technology, and higher education systems, on human capital, and labor markets of the sending and receiving countries, and at an individual level.

The main positive and negative effects have been studied and cataloged (Regets, 2001). As for the possible positive effects on science, technology, and higher education systems of sending countries, it should be mentioned that this has led to an opening of knowledge flows and collaboration between countries and an increase in ties and networking between foreign research institutions. Specifically for the science system, the return of natives with a foreign education and, in general, highly skilled human capital may prove, directly or indirectly, to be of benefit. The return of mobile researchers means that the country of origin will acquire this new knowledge and experience gained, due to it being brought back upon the citizen's return.

The returnees may in fact bring back valuable entrepreneurship or management skills and give the home country a better insight into export opportunities for technology or into access to global networks.

A recent study of 127 developing countries conducted in 1990 and 2000, respectively, suggested that doubling the emigration rate of the highly skilled led to a 5% increase in gross human capital formation among the native population (Beine *et al.*, 2006). This may mean that scientists' migration may turn out to be a positive

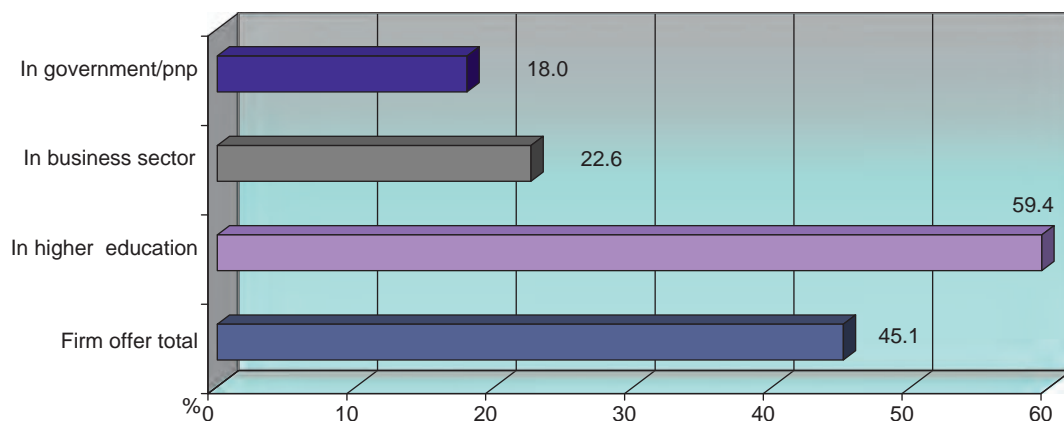


Figure 3 European Union with PhD obtained in the United States with firm employment offer planning to locate in the United States (from 1991 to 2000). Firm offer total: 45.1; in higher education: 59.4; in business sector: 22.6; and in government/pnp: 18.0. From SRS, NSF.

Table 1 Foreign students enrolled in doctoral programs, 2002

<i>Country</i>	<i>Number of foreign doctoral students</i>
Switzerland	5359
Belgium	1990
United Kingdom	22 824
United States (2001)	78 884
Australia	8033
Denmark	872
Austria	2705
Canada	4655
Norway	727
Sweden	3147
Spain	8677
Czech Republic	1367
Hungary	496
Finland	1350
New Zealand	341
Portugal	718
Iceland	2
Korea	649
Turkey (1999)	378
Slovak Republic	110
Mexico	99
Italy	218

From OECD, Education Database, 2006, 2008.

asset for the less-developed countries as well. Researchers produce public goods that can also be shared across borders. They are particularly important as innovation and development agents for the economy, and beneficial effects of their work outcome may spread over different sectors of the economy.

Beneficial effects may also arise when using foreign facilities not available in the country of origin to further that country's research and development (R&D) agenda (i.e., big science facilities).

The receiving countries may benefit from hosting researchers and scientists. This may also turn into an increased R&D and economic activity, due to the availability of additional highly skilled workers, which favors entrepreneurship, especially in high-growth areas. Receiving countries also benefit from stronger knowledge flows and collaboration with sending countries. Diversity and creativity are certainly further promoted by the melting pot of the intellectual mix brought about by the inflow of people with different backgrounds.

A number of negative effects on science, technology, and higher education systems can also be envisaged. For the sending countries, the most-studied effect of highly skilled labor leaving its country of origin is the well-known and mentioned brain drain. This departure may mean losing part of the national ability to generate high-level activities that thus may lead to a weaker intellectual and productive capacity. Similarly, the country of origin will miss out if the outcome of the research, not the

process, is of value, as, for example, in a postdoctoral's research. Patents, for example, may provide instant benefit to the host country with little or no likely benefits for the country of origin.

A direct loss can be envisaged also in purely economic terms. The investment made in tertiary education becomes a benefit for another nation and may be perceived waste of national resources that gives poor returns from public investment.

Summary

The simple and traditional model brain drain/brain gain is no longer adequate to explain the complex movements of the highly skilled population, to describe and explain international mobility patterns and, even more, the consequences that these movements generate on the creation of knowledge and knowledge investments. The attention given by policy analysts and international organizations to these issues has been rapidly growing in recent years in order to develop a common framework for analysis (OECD, 2001, 2002).

Potential effects of mobility have been recognized to be positive or negative, both for the sending and receiving countries. Science and education policies must carefully take these potential effects into consideration to prevent possible drawbacks or, at the other end, to favor the positive spillover.

The envisioned global effect of increased mobility in higher education, science, and technology may be found in the best use of highly qualified people. But the analysis of the effects of highly qualified international migrations on investments in human capital is still to be fully developed. There are not yet consolidated indicators that can ease the measurement of the various effects of international mobility of highly skilled human resources.

Many questions are raised about the meaning and reasons for mobility. What helps and hinders international mobility? What are the driving and resisting forces? What barriers are still to be removed and what actions can ease the mobility process? A wise mix of incentives and schemes seems to be the best way of promoting mobility. The value of an international mobility experience is generally highly considered both by PhD students and scientists. It is part of the personal scientific and cultural growth and may result in being the first step in a research career for PhD graduates. The difficulties in organizing a stay abroad may result in personal and objective problems needing to be overcome, concerning the financing, location, etc. A supranational intervention is a suitable means to simplify international relations and overcome differences in local legislation, through a lightweight normative role, acting as a catalyst and leaving institutions and individuals free to organize mobility.

Bibliography

- Altbach, P. G. (2004). Globalization and the university: Myths and realities in an unequal world. *Tertiary Education Management* **1**, 3–25.
- Beine, M., Docquier, F., and Rapoport, H. (2006). Brain drain and human capital formation in developing countries: Winners and losers. *Discussion Paper 2006-23*. Louvain-la-Neuve: Département des Sciences Économiques de l'Université catholique de Louvain (forthcoming in *Economic Journal*).
- Docquier, F. and Marfouk, A. (2004). Measuring the international mobility of skilled workers (1990–2000). Release 1.0. <http://www-wds.worldbank.org> (accessed June 2009).
- Dumont, J. C. and Lemaitre, G. (2005). Counting immigrants and expatriates in OECD countries: A new perspective. *OECD Social Employment and Migration Working Papers*, N. 25. Paris: OECD.
- European Commission (1996). *Green Paper Eliminating Obstacles to Transnational Mobility*. COM(96) 462. Brussels: European Commission.
- European Commission (2007). *Green Paper the European Research Area: New Perspectives*. COM(2007) 161 final. Brussels: European Commission.
- Hansen, W. (2003). A web-based e-survey on the international mobility of scientists and engineers. *Final Report Project: Brain Drain Migration Flows of Qualified Scientists*. Brussels: European Commission.
- Kyvik, S., Karseth, B., and Blume, S. (1999). International mobility among Nordic doctoral students. *Journal of Higher Education* **38**(4), 379–400.
- NSF (2006). *Science and Engineering Indicators, 2006*. Washington, DC: National Science Foundation.
- NSF (2007). *Asia Rising Science and Technology Strength: Comparative Indicators for Asia, the European Union and the United States*, NSF 07-319. Arlington, VA: National Science Foundation, Division of Science Resources Statistics.
- OECD (2001). *Innovative People: Mobility of Skilled Personnel in National Innovation Systems*. Paris: OECD.
- OECD (2002). *International Mobility of the Highly Skilled*. Paris: OECD.
- OECD (2007a). Labour market characteristics and international mobility of doctorate holders: Results for seven countries. *STI Working Paper 2007/2*. Paris: OECD.
- OECD (2007b). International mobility of researchers. *DSTI/STP(2007)25*, Mimeo. Paris: OECD.
- Oliver, J. (2007). *First-Time, Full-Time Graduate Student Enrollment in Science and Engineering Increases in 2006, Especially among Foreign Students*, NSF 08-302. Arlington, VA: National Science Foundation, Division of Science Resources Statistics.
- Regets, M. (2001). Research and policy issues in high-skilled international migration: A perspective with data from the United States. *Innovative People: Mobility of Skilled Personnel in National Innovation Systems*, pp 243–260. Paris: OECD.
- Regets, M. C. (2007). Research issues in the international migration of highly skilled workers: A perspective with data from the United States. *Working Paper SRS 07-203*. Arlington, VA: Division of Science Resources Statistics, National Science Foundation.
- Royal Society (1963). *Emigration of Scientists from the United Kingdom*. London: Royal Society.
- Thorn, K. and Holm-Nielsen, L. B. (2006). International mobility of researchers and scientists policy options for turning a drain into a gain. *Research Paper No. 2006/83*, United Nations University. Helsinki: UNU-WIDER.
- UNESCO (1998). Declaration on higher education for the 21st century: Vision and action. *Proceedings World Conference on Higher Education*. <http://portal.unesco.org/education> (accessed June 2009).
- WHO (World Health Organization) (2004). *Report by the Secretariat on the Recruitment of Health Workers from the Developing World*, WHO Document EB114/5. Geneva: WHO.

Further Reading

- Aweduto, S., Brandi, M. C., and Todisco, E. (2004). High skilled migrations, opportunity and brain drain. *Migration Studies/Studi emigrazione* **156**, 775–796.
- Gaillard, J. and Gaillard, A. M. (1997). The international migration of brains: Exodus or circulation? *Science Technology and Society* **2**, 2.
- Lowell, L. (2001). Policy responses to the international mobility of skilled labour. *International Migration Papers*. International Labour Office, Geneva.
- Salt, J. (1997). International movements of the highly skilled. *OECD Occasional Papers No 3*.
- Saxenian, A. L. (2006). *The New Argonauts: Regional Advantage in the Global Economy*. Cambridge, MA: Harvard University Press.

National Science Policy and Universities

M Henkel, Brunel University, London, UK

U Kogan[†]

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Research quantum – In the early stages of the development of Australian competitive funding policies, a small component of institutional grants was based on their “research quantum,” assessed by a set of quantitative performance indicators.

Triple helix – The metaphor adopted by Etzkowitz and Leydesdorff to characterise the increasingly complex relationships between universities, industries, and governments, in which they may take on each other's roles.

Science policy has tended, in practice, to denote policy for research and development in the natural sciences, even in national contexts where science is a more comprehensive concept. In recent years, this tendency has been reinforced as the aims of national science policies center increasingly on innovation.

From Benign Neglect to Managed Instrumentalism

At one time, there was no equivocation about national science policies as far as they affected universities. The Humboldtian creed may not have been adopted in its full extravagance everywhere but its key assumptions held good in most Western European and Anglophone countries. As Neave observed, it established the research university legend, with its protective barrier ensuring separation of the world of interests from the world of knowledge. Universities would not be involved in the consequences of the knowledge that they created. Research would be driven by internal dialog between individuals in disciplinary communities.

In the science-push model, too, as enunciated by Vannevar Bush in 1945, science was assumed to be a largely unquestioned public good which would serendipitously yield new discoveries and technologies that would, in turn, enhance defense capabilities, economic development, and social well-being. In many countries, therefore, science policy in general and its application to university finance, governance, and evaluation enjoyed a form of benign neglect. Universities received funding for research accommodation, equipment, and materials from national

authorities mainly on the proposals made by the academics who would use them, without much, if any, reference to national priorities or objectives. These proposals would be driven by academic curiosity. Such criteria that applied were those of the perceived research excellence of the applicants, perhaps interrogated by systems of peer review. Academic research staffing was usually funded by what was required for their other key function – the education of students.

The fact that there were no declared policies did not mean that academic work made no contribution to economic and social development, and two world wars helped to accelerate the demands for relevance. In the UK, for example, a national physics laboratory was founded in order to bring scientific knowledge to bear upon industrial and commercial practice, and, in 1918, the Haldane report asserted the need for research which could support the activities of government. However, the idea of a deliberate science policy was explicitly and repeatedly rejected.

The 1970s brought a turning point. Governments became more directive on science policy. Science became an increasingly important but less distinctive component of public policies, in consequence of which scientists lost some of their authority and autonomy and universities some of their exceptionalism.

The drivers of these developments were multiple and interconnected. Conceptions of the nature of science, its modes of production, and its functions in societies shifted with the emergence of knowledge societies and knowledge economies. A new generation of industries appeared, characterized by synergy between research universities and rapid technological advance. A wider set of changes generated new conceptions of the state and its administration. The 1970s saw radical reappraisals of Keynesian economics and the growing influence of neoliberalism. Arguments for limiting public expenditure and increasing the role of the private sector became powerful forces. They are evident in new forms of public management, involving a blurring of the boundaries between the state and the market. All these developments had implications for relationships between universities and both the state and other sectors of society, notably industry and commerce.

In the last two decades, the internationalization of higher education and research, in the form of cross-border collaborations between academics and their institutions, has accelerated. The influence upon national policies of economic globalization has grown, challenging the power of the nation-state and, ultimately, national policies, too.

[†] Deceased

Changing Assumptions About Science and Science Policies

There were intimations of change in the 1960s. De Solla Price's pioneering work in scientometrics, demonstrating the exponential growth of science, laid the ground for governments to introduce more rational methods of resource allocation in this field. In 1971, the Organization for Economic Cooperation and Development (OECD)'s landmark Brooks report, asserted the need for governments to set the direction of science and its funding within a framework of economic and social goals.

With the emergence of knowledge societies, science has come to be seen as a prime strategic resource in the race for economic competitiveness but it has also been exposed to wider social criticism. Scientists, indeed the academic profession as a whole, began to be regarded as an interest group whose demands might be measured against other contending interests.

Distinctions between basic or curiosity-driven research and applied research were blurred in the introduction and growing importance in public funding of strategic or potentially exploitable research.

Foresight policies were adopted in a number of countries in the 1980s. They entailed bringing together scientists, industrialists, government officials, and others to identify strategic research areas and emerging technologies with most potential for commercial exploitation or social benefit (Martin, 1996).

Clearly, long-established boundaries and distributions of power in the governance of science were being broken down. Boundary blurring was evident in new conceptions of knowledge production. Technological innovation was perceived as more directly dependent on and often intertwined with scientific practice. The science-technology relationship began to be characterized as a dynamic system involving connections and feedback loops between multiple parties, academic and industrial, users as well as producers of innovation. Gibbons and his colleagues discerned a radical shift between two modes of knowledge production, from mode 1 (knowledge generated within disciplines) to mode 2, transdisciplinary knowledge, created within a context of application. This formulation has been criticized on the grounds that such a polarized framework is oversimplifying and fails to recognize scientific diversity and the past importance of contexts in many fields. However, it has been an influential reference point both in the academic literature and among policymakers.

National Science Policies

First, a theme now common to developed and developing countries across the world is the belief on the part of

governments that research and development (R&D) are essential to national economies.

The new surge of economic instrumentalism on the part of governments has several concomitants. While the USA has in the past been distinctive in its unequivocal commitment to competitive allocation systems for research monies, such systems have now been widely embraced, including by some mainland European countries, traditionally committed to equality of status for universities. The purpose is, in part, efficiency but, even more, the support, through concentration and selectivity, of research excellence, is seen now as the key to technological innovation.

The movement toward competitive allocation emerged earliest and most evidently in the UK's research assessment exercise (RAE) but was followed by Australia's competitive assessment of the research quantum. Research councils were historically the providers of resources for researcher-inspired work. Now in both of these countries, they specify priority areas which accord with those determined by central government. Australian Research Council funding is directed largely toward four areas. Hard science priorities predominate and social sciences and humanities are incorporated in them (Meek *et al.*, 2010). A recurrent theme is that funding of applied and strategic R&D has increased, at the expense of pure or basic research.

An almost universally evident policy has been to encourage close working relationships between universities and businesses. Major objectives here have increasingly crystallized round the advancement of innovation. Policies emphasize the role of universities in the development of small- and medium-sized enterprises (SMEs), widely seen as sources of technological innovation, and encourage university-industry partnerships in educational as well as research activities. Governments seek to endorse substantial changes of beliefs in universities and business about the purposes and production of research in societies and the need to enhance research capacity in industries: the functions of doctoral education are now generally seen as extending well beyond the reproduction of the academic profession. Governments want more rapid technology transfer, other forms of knowledge transfer, more exchange relationships, and more collaborative production of knowledge between businesses and academic institutions.

A second major set of objectives in policies designed to bring universities and business closer together has been to encourage industries to take a greater share of the financing of research and universities to engage actively in markets and reduce their dependence on public funds.

The context of all these policies has been an overall decline in the public funding of universities, simultaneous with the increase in the demands upon them. Financial incentives and sanctions have been probably the most potent of the instruments adopted by governments in the science policy arena. Output-based performance measurement, national systems of evaluation, and quality

assurance have also been increasingly important tools for change in universities, even if the direct linkage between evaluation and funding established in the UK RAE has not been widely followed, at least until recently.

The growing use of such policy instruments is one dimension of new models of public sector management. Governments have established clearer direction for higher education and science policies: many, for example, require universities to submit research plans. However, they have sought, too, not only to reduce universities' financial dependence on the state but also, at least in policy rhetoric, to eschew mechanisms of detailed state control. Universities have been encouraged to enhance their capacities for strategic decision making, albeit within strong national priorities, and to replace collegial with managerial modes of governance, although this trend has been taken further in Anglophone countries than in mainland Europe.

The developments outlined represent the loosening of boundaries between the functions of universities and industry, and the blurring of distinctions between public and private funding and the roles of the state and the market in science policy. However, governments have also encouraged academics and universities to move further into academic capitalism. Slaughter and Leslie adopted this term in the mid-1990s not only to denote academic movement into markets to seek external funds but also to signal the encroachment of the profit motive into the academy and the potential it created for contradictions between public and private sector cultures in the research environment.

This is most clearly demonstrated in the widespread changes in the laws governing Intellectual Property Rights (IPR), led by the USA (Bayh Dole Act 1980) but followed by the UK, Australia, and some mainland European countries. As a result, universities were freed to claim IPR over new knowledge developed in state-funded research projects and significantly enhance their patenting and licensing activities. Individuals and institutions have also been encouraged by governments to form their own spin-out companies to capitalize on their discoveries, although, again, this trend was established earlier in Anglophone countries.

Science policies based on the idea of the knowledge economy are increasingly widely diffused across the world, including in the rising economies of Latin America, India, China, and others in the Asia-Pacific region. China, for example, has seen an exponential growth in the funding of research and development by government and, even more, by industry. It is now ranked fifth in the world on the measure of its share of scientific publications and is moving rapidly toward its goal of a successful knowledge-based economy.

In the space available, it is not possible to examine movements in higher education research in the developing world. Certain features are, however, common to such

systems. None can draw on resources comparable to those available to the countries of the north and west and they devote significantly smaller proportions of their gross national product on R&D. India, for example, allocates .69%, Brazil, .82%, and China 1.4%. As Zakri indicates, capacity development is a prime task for such countries. The overriding objective is to produce R&D that is relevant to economic and social needs. This may leave little or no research and scholarship in the humanities or social sciences.

It is increasingly difficult to discuss national science policies without reference to transnational organizations and policymaking. The European Union (EU) is a prime example. It is now a growing influence upon the national science policies of all European countries, including non-members of the EU, as it seeks to align EU and national science policies and create a Europe of knowledge through an integrated innovation agenda, incorporating higher education and research (van Vught, 2010).

State–University Relations

Science policies have been major triggers of change in the functions of universities in nation states. Key issues are their meaning for academic autonomy and for authority and dependency relationships between states and universities, which are complex and sometimes paradoxical.

Theorists of the postindustrial society predicted that universities would become axial structures of the emerging knowledge societies. National policies have reflected this view in the roles allocated to universities and in the more explicit requirements that they help bring about economic transformation.

In some countries where higher education and research activities are subject to both a federal and a state or provincial authority, shifts in the balance of that authority reflect their growing national strategic importance. In Australia and Canada, the federal authorities have increasingly occupied a more dominant role in the financing of and policies for research, and some have given tax concessions to industry engaged in research. Soft federalism has given way to direct funding and a panoply of guidance and instruction. In the USA, while the system remains loosely tied together, both federal and state funding has increased; an important feature is the extent to which some states seek economic advantage by financing research, most often in the burgeoning biomedical fields. Germany has both enhanced the federal financing role and yielded detailed powers to the *Länder*.

An increasingly influential conceptualization of changed relationships among universities, government, and industry, particularly in regional policies, is Etzkowitz and Leydesdorff's (1997) triple helix. The development of regions of knowledge is seen by national governments, as

well as by the EU, as critical to the growth of national economies in a global context. It constitutes a strong example of universities, governments, and businesses combining to construct a common goal, which they could not achieve by operating separately. Universities are often the initiators of such region-oriented ambitions and invest substantially in their realization.

Prima facie, universities' decreased financial dependence on state funding gives them more autonomy through greater diversity of funding and the freedom to compete in a range of markets, according to their chosen missions and specific strengths.

However, most remain substantially dependent on public funding. As such, their choices are structured by national priorities that privilege particular types of research, notably strategic research, in particular fields, such as life sciences, nanotechnology, and information and communication technologies.

Universities are more publicly accountable and while evaluations of their work are still carried out largely by peers, their achievements are measured against criteria reflecting the needs of the state as well as academic judgment. In some countries, notably the UK and Australia, they are subject to performance indicators and given guidance backed in differing degrees by persuasion and incentives.

Several components of public policies combine to support inequalities in the power and dependency relationships between the state and universities. In some instances (e.g., mainland Europe), they create and in others (e.g., the UK and the USA), they reinforce the stratification of universities in terms of reputation and command of resources. Larger and older institutions are more heavily represented among the beneficiaries of national and EU selective and competitive funding. High performers in national evaluations not only enhance their reputations but also are more successful in securing funding and partnerships in the private and nonprofit sectors.

Policies promoting university–industry links may have the opposite types of effect for financially weaker and less-established universities. They may have to accept funds for developing routine contract research that do not cover their full costs and then to recoup these from their public funding. This is likely to weaken them further: partly because research strength needs money and partly because it shifts the balance of their research effort away from more prestigious forms of research.

Meanwhile, as national and transnational policies contribute to the growing power of the strongest universities, they also loosen the relationships between these institutions and their nation-states. Most European states encourage their universities to take advantage of the EU framework funding programs, which reward generation of international networks and other forms of international collaboration between universities and also between universities and innovative industries. These funds, undoubtedly have a

powerful leverage effect on universities' research planning, despite being relatively small as a proportion of total funding. (The sixth framework program, for example, represented only 5% of the total expenditure on research by EU member states.) Advantageous in themselves, they may open up further international funding possibilities and orient institutions toward global markets. Similarly, national policies fueling some of their institutions' ambitions to join the ranks of newly defined world-class universities may encourage them to become more distanced from their national identity.

The University: Institutional and Epistemic Issues

Changes in national policies since the late 1970s have implications for university functions, structures, values, and modes of knowledge production.

There has been a quantum leap in the number and types of links between universities and industry, beginning in the USA, although both there and in the UK, big increases in research collaborations anticipated the major national policy measures (Dill, 2010; Calvert and Patel, 2002). The concept of a third task, additional to research and teaching and broadly defined in terms of knowledge transfer beyond their own boundaries, is increasingly prominent in university missions. Policy changes stimulated a large, if differential, rise in licensing and patenting activities and substantial growth in the formation by individuals and institutions of spin-out companies.

These have been powerful influences in the adoption by universities of more corporate and managerial regimes. There is a general trend toward a concentration of authority at the center, but universities with strong research reputations sustain more collegial or interactive styles of management than others.

However, academics no longer have a monopoly of influence on university culture and organization. All universities have incorporated a more diverse workforce, including an enlarged group of (in Rhoades' terms) managerial professionals, staff with advanced degrees and qualifications who are neither faculty nor senior administrators. Faculty were no more than 30% of all personnel on US campuses and 55% of all professionals in 2006, down from 65% just two decades earlier.

Related phenomena are the growth and diversification of interstitial structures in universities to cope with new functions connected with income generation, knowledge transfer, academic capitalism, and mediation between public, private, and nonprofit bodies at local, regional, national, and international levels.

Academic work, too, has been subject to substantial structural change. National policies for research selectivity and concentration have been replicated by universities,

resulting in the abolition of some departments, the reconfiguration of others and, most notably, the proliferation of different types of basic units superimposed on departmental structures. Indeed, increasingly, universities regard these interdisciplinary and transdisciplinary research institutes and centers, with their associated graduate schools, as their flagships.

These developments are regarded by some as fragmenting discipline-based structures and relationships within universities, undermining the knowledge base and exacerbating already strong pressures for separating research from teaching, particularly at first-degree level, responsibility for which may remain with departments. Research suggests that academic values and identities remain strongly rooted in disciplines, which are still widely seen as providing the foundations of innovative research. However, against this, it is argued, first, that many successful researchers see no contradiction between strong disciplinary identities and agendas and interdisciplinary collaborations; second, that the latter generate new forms of integration and new epistemic communities, reflecting both changing internalist constructions of knowledge and externally defined problems or needs.

There is no doubt that universities have embraced activities and working relationships which challenge long-held assumptions that academic and industry-based scientists belong to two knowledge-production systems with, according to Dasgupta and David (1994), distinct, if not contradictory sets of goals, norms, rules, and rewards, those of academic or open science and those of for-profit or proprietary science. Academic science operates a reputation-based reward system, geared to the production of public knowledge. Academic scientists compete for authority in their fields, through primacy in defining and solving scientific problems. They have seen the university as according them the freedom to choose their own agendas and the prestige associated with curiosity-driven research. These views remain strongly represented among university researchers, even where their leaders are promoting commercialization and social engagement. However, in industry-based science, the rewards are financial, firms set the agendas, and the goal is private knowledge. Hence the norm of secrecy is paramount.

Dasgupta and David argue that a shift toward such norms in universities might undermine the existing (not perfect but functional) complementarity between academic and proprietary science. A reduction in academic freedom and opportunities for curiosity-driven research, combined with greater emphasis on financial rewards, might seriously erode the attractions of academic careers and so threaten not only new research directions but also the flow to industry of well-trained scientists perceived as essential for industry's capacity to exploit new knowledge.

Others take a different view about the collapse of well-established boundaries, between academe and industry and public and private knowledge regimes. They argue that revolutionary changes are occurring in

the production of knowledge but that it is possible for universities to manage the resulting contradictions. Research on the implications for academic research standards of university entry into the commercial arena shows that academic and commercial success can be mutually reinforcing: the most active collaborators with industry are also, largely, the highest ranking research universities, although lower ranked universities may work with industry in limited specialist areas. However, according to Owen-Smith (2005), universities have to learn how to make commercialization work so as to sustain academic freedom and creativity and make money. Over time, the number of academic publications declines with continued pursuit of high-volume patenting. A minority of universities have realized they need more sophisticated strategies, focused on high-impact patents, which are positively associated with high-impact academic publications.

This study reinforces but sharpens the research indicating that national policies may have exaggerated the financial rewards to universities from academic capitalism. In most universities in the US and the UK, the costs of administering technology transfer offices (TTOs) outweigh the income generated, although in the US this is less true of well-resourced and private universities. High returns depend heavily on a few highly marketable discoveries.

These findings seem to support an argument put forward by Geuna in 1998 that has profound implications: that the challenges of a new era are opening up an unbridgeable gap between universities; only a few elite research universities will fully adapt to the new demands and also manage to retain some of the assumed defining features of universities; and many will be marginalized and little influenced by international changes in the production of knowledge.

As yet, nevertheless, significant uncertainties remain about what such changes may mean for the future of universities. More systematic analysis of emerging categories of university is now beginning to take us beyond rather crude observations of enhanced stratification. But most research on the consequences of new modes of working still concentrates on highly ranked institutions and basic units. Not enough is known about the others, although, according to Mansfield and Lee, significant contributions to innovation in major US firms came from them in the mid-1990s.

Given that the major advances into commercialization are confined to a handful of scientific fields, that national and university commitment to curiosity-driven science, the humanities, and the social sciences has declined and that the status of mode 1 knowledge production is uncertain, there is important research to be done on the consequences for what might be called the epistemic ecologies, as well as the institutional identities, of all but possibly the most well-endowed and academically prestigious institutions. There is now a growing sense of the need for more collaboration between higher education studies and the social studies of science, if significant progress is to be made on such research questions.

A final research issue arises from the fact that the balance of academic literature on science policy remains heavily weighted toward the OECD countries, although science is now a key component of most national policies.

Conclusion

The story related here is of great change. Professional autonomies have been confronted by determined policies of what has been conceived as the public utility. Much academic freedom may remain but unequally distributed and conditioned by priorities, mediated largely through funding and established in the polity rather than academe.

Policy is substantially based on neoliberal concepts of economic progress. The age of Keynesian deficiency funding has given way to the imperative to act as if in a market place. Yet, there is a paradox here. Along with liberal declarations about institutional freedom and the need for diversity have come increasingly tight policy frameworks, making for greater uniformity and imitation.

Intensified attempts to engage universities with the private sector, together with reduced support for basic science, the humanities, and the social sciences, have raised large questions about the university as an institution and the nature of its contributions to societies.

See also: Higher Education and the Knowledge Society; Internationalization of Higher Education; Managerialism and Collegialism in Higher Education Institutions; National and International Rankings of Higher Education; Strategic Planning in Higher Education; Universities' Engagement with Society.

Bibliography

- Calvert, J. and Patel, P. (2002). *University–Industry Collaborations in the UK*. Brighton: Science Policy Research Unit/University of Sussex.
- Dasgupta, P. and David, P. (1994). Towards a new economics of science. *Research Policy* **23**, 487–521.
- Dill, D. (2010). The United States. In Dill, D. and van, Vught, F. (eds.) *National Innovation and the Academic Research Enterprise: Public Policy in Global Perspective*. Baltimore: The Johns Hopkins University Press.
- Etzkowitz, H. and Leydesdorff, L. (eds.) (1997). *Universities and the Global Economy. A Triple Helix of University–Industry–Government Relations*. London: Pinter.
- Martin, B. R. (1996). Technology foresight: Capturing the benefits from science-related technologies. *Research Evaluation* **6**, 158–168.
- Meek, L., Goedegebuure, L., and van der Lee, J. (2010). Australia. In Dill, D. and van Vught, F. (eds.) *National Innovation and the Academic Research Enterprise: Public Policy in Global Perspective*. Baltimore: The Johns Hopkins University Press.
- Owen-Smith, J. (2005). Trends and transitions in the institutional environment for public and private science. *Higher Education* **49**, 91–117.
- van Vught, F. (2010). The European Union. In Dill, D. and van Vught, F. (eds.) *National Innovation and the Academic Research Enterprise: Public Policy in Global Perspective*. Baltimore: The Johns Hopkins University Press.

Further Reading

- Bush, V. (1945). *Science: The Endless Frontier*. Washington, DC: National Science Foundation.
- De Solla Price, D. (1963). *Little Science, Big Science*. New York: Columbia University Press.
- Dill, D. and van Vught, F. (forthcoming). *National Innovation and the Academic Research Enterprise: Public Policy in Global Perspective*. Baltimore: The Johns Hopkins University Press.
- Etzkowitz, H. and Leydesdorff, L. (1998). The endless transition: A 'triple helix' of university–industry–government relations. *Minerva* **36**, 203–208.
- Geuna, A. (1998). The internationalization of European universities: A return to medieval roots. *Minerva* **36**, 253–270.
- Geuna, A. (1999). *The Economics of Knowledge Production, Funding and the Structure of University Research*. Cheltenham: Edward Elgar.
- Gibbons, M., Limoges, C., Nowotny, H., et al. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: Sage.
- Haldane Report (1918). *Report of the Machinery of Government Committee*. Cmd 9230. London: HMSO.
- Laredo, P. and Mustar, P. (eds.) (2001). *Research and Innovation Policies in the New Global Economy. An International Comparative Analysis*. Cheltenham: Edward Elgar.
- Mansfield, E. and Lee, J. -Y. (1996). The modern university: Contributor to industrial innovation and recipient of industrial R&DD support. *Research Policy* **25**, 1047–1058.
- Mowery, D., Nelson, R., Sampat, B., and Ziedonis, A. (2004). *Ivory Tower and Industrial Innovation: University–Industry Technology Transfer before and after the Bayh-Dole Act*. Stanford, CA: Stanford University Press.
- Neave, G. (2006). On time and fragmentation: Sundry observations on research, the university and politics from a waveringly historical perspective. In Neave, G., Bluckert, K., and Nybom, T. (eds.) *The European Research University: An Historical Parenthesis?* New York: Palgrave Macmillan.
- OECD (1971). Science, growth and society: A new perspective. *Report of the Secretary-General's Ad Hoc Group on New Concepts of Science Policy*. Paris: OECD.
- Policy Research in Engineering, Science and Technology (PREST) (2002). *Impact of the Research Assessment Exercise and the Future of Quality Assurance in the Light of Changes in the Research Landscape*. Manchester: University of Manchester.
- Reichert, S. (2006). *The Rise of Knowledge Regions: Emerging Opportunities and Challenges for Universities*. Brussels: European University Association.
- Rhodes, G. (2006). The higher education we choose: A question of balance. *Review of Higher Education* **29**, 381–404.
- Slaughter, S. and Leslie, L. (1996). *Academic Capitalism: Politics, Policy and the Entrepreneurial University*. London: The Johns Hopkins University Press.
- Turk-Bicakci, L. and Brint, S. (2005). University–industry collaboration: Patterns of growth for low and middle-level performers. *Higher Education* **49**, 61–89.
- Zakri, A. H. (2006). Research universities in the 21st century: Global challenges and local implications. *UNESCO Colloquium on Research and Higher Education Policy*. December 2006.

Relevant Websites

- <http://www.eua.be> – European University Association.
- <http://www.mbs.ac.uk> – Manchester Business School, Policy Research in Engineering, Science and Technology (PREST).
- <http://webcaspar.nsf.gov> – National Science Foundation, USA, Statistical Data.
- <http://www.oecd.org> – Organization for Economic Cooperation and Development (OECD).
- <http://www.sussex.ac.uk> – University of Sussex, Science Policy Research Unit (SPRU).

Research Commercialization

S Slaughter, University of Georgia, Athens, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

As more and more students attend post-secondary education and state support per student declines, faculty, administrators, trustees, regents and state officials are searching for ways to expand university revenues. One strategy is commercialization of research. To maintain or expand resources, faculty are increasingly encouraged to compete for external dollars tied to market-related research. Market-related research is referred to variously as applied, commercial, strategic, and targeted research. The external monies at which commercial research is aimed come in the form of partnerships with industry and government, technology transfer, royalties and licenses, equity holdings in faculty start-up companies, and income related to incubators and research parks. Another strategy for securing external dollars is through the marketing and sales of educational products and services, ranging from learning enhancement tools to digital education. Generally, institutional and professorial market or market-like efforts to secure external monies can be conceptualized as academic capitalism.

This article focuses on a single aspect of academic capitalism: commercialization of research. The following topics are covered: broad policy changes that made research commercialization possible; patents; start-up or spin-off companies; research parks and incubators; copyright; and the infrastructure required by universities to manage research commercialization. Although the article is comparative, the focus is on the United States because it has led research commercialization.

Policy Changes

Academic capitalism depends on the rise of the neoliberal state, which now characterizes English-speaking industrialized countries such as Australia, Canada, the United States, and the United Kingdom. The European Union (EU) does not embrace neoliberalism as closely as the English-speaking countries, but is in the process of adopting many of its practices with regard to higher education generally and research commercialization, specifically. In the United States, research universities came to maturity after World War II, during the great expansion of post-secondary education in the 1950s and 1960s, in the era of embedded liberalism. Embedded liberalism was a reaction to classic liberalism and sought to constrain capitalism to avoid depression, poverty, and social unrest. To achieve these ends, social and political oversight, regulatory and

planning functions were embedded in the state. The common goals of embedded liberal states were full employment, economic growth, and the welfare of the citizenry. If necessary, the state intervened in market processes to reach these goals. Keynesian monetary policies were characteristic of the embedded liberal state.

In contrast, neoliberalism is:

...a theory of political economic practices that proposes that human well-being can best be advanced by liberating individual entrepreneurial freedoms and skills within an institutional framework characterized by strong private property rights, free markets and free trade. The role of the state is to create and preserve an institutional framework appropriate to such practices (Harvey, 2005).

Although neoliberal theory minimizes the role of the state, in practice state subsidies and oversight are not minimized; rather they shift to new areas. In particular, and important with regard to higher education, subsidies shift from broad general appropriations for the public good – for example, low tuition – to user taxes and fees – for example, high tuition – which emphasize individual rather than social gains accrued as a result of higher education. In general, higher education shifts from a public good knowledge/learning regime to an academic capitalist knowledge/learning regime. The academic capitalist knowledge/learning regime would not be possible without the neoliberal state. However, the relationship is bidirectional and higher education is in many respects the exemplar neoliberal state agency in that it is flexible, entrepreneurial, and essential to the maintenance and expansion of a knowledge economy.

The history of public policy in the United States illustrates the rise of the neoliberal state in research commercialization. Prior to the Bayh-Dole Act (1980), federal policy in the United States placed discoveries made with federal grant funds in the public domain. Although universities were permitted to secure patents from the federal agency that funded their research, a relatively small number of universities patented by engaging in this process. The 1980 Bayh-Dole Act signaled a more general encouragement of university research commercialization. It gave universities title to inventions discovered with federal research and development funds. Bayh-Dole made university technology transfer to the private sector an important outcome of federally financed research, and allowed exclusive licensing of publicly funded research to achieve that policy goal.

The Bayh-Dole Act gave new and concrete meaning to the phrase “commodification of knowledge.” As potential patent holders, university trustees and administrators require faculty to disclose all research that can be protected intellectual property. Faculty begin to conceptualize their discoveries as products or processes, private, valuable, and licensable, not necessarily as knowledge to share publicly with a community of scholars.

The Bayh-Dole Act, for which representatives of research universities and business lobbied intensively, was only one piece of a broader bi-partisan competitiveness strategy developed by the US Congress. Beginning in the 1980s, an array of legislation – ranging from research tax credits to technology transfer – was passed that laid the ground work for the competitiveness coalition. For example, the Federal Courts Improvements Act (1982) created a new Court of Appeals for the Federal Circuit (CAFC) which handled patent appeals from district courts, thereby ending forum shopping in intellectual property cases, creating a more uniform approach to patents. The new court led the way for a greatly strengthened approach to intellectual property, which eventually allowed patents on genetically engineered bacteria, genetically altered mice, particular gene sequences, surgical methods, computer software, financial products, and methods for conducting auctions on the World Wide Web.

Similar changes occurred in the United Kingdom, beginning under Margaret Thatcher’s government, and in Australia, under the Hawke government. Canada lagged behind, but in the 1990s neoliberal policy at both the provincial- and federal-level policy heavily promoted research commercialization. In the 2000s, the EU began a research commercialization program, drawing heavily on the models provided by the English-speaking countries.

Patents

Prior to 1981, fewer than 250 patents were issued to US universities per year. The number of new patent applications more than quintupled between Fiscal Year 1991 and Fiscal Year 2003, indicating the growing effort and increasing success of universities in obtaining patent protection for their technology. The number of institutions awarded patents increased by more than 60% (to 198) between the late 1980s and 2003. Both public and private institutions participated in this rise. Success in obtaining patents seems to depend more on other resources, which are discussed below under infrastructure.

Despite the increase in institutions receiving patents, the distribution of patenting activity has remained highly concentrated among a few major research universities. Among the top 100 research and development (R&D) institutions, the top 25 recipients between 1994 and

2003 accounted for 55% of all academic patents in 2003, a share that has remained constant for two decades. Including the next 75 largest recipients increases the share to more than 80% of patents granted to all institutions since 1987. The growth in academic patents occurred primarily in the life sciences and biotechnology. Of the top 100 US research universities, ten institutions generated 66% of profits received from technology licensing. Less research-intensive institutions with small technology licensing programs or institutions that are new to technology licensing generally experience annual losses.

The EU demonstrates a similar pattern. Although the growth of university owned and invented patents has increased, university licensing is not profitable for most universities. The success of a small number of institutions may have exacerbated the differences among universities within the EU with regard to resources for research and development.

US universities responded to the changing policy opportunities by developing intellectual property policies that simultaneously offered powerful incentives for faculty to patent and expanded institutional control over intellectual property.

Royalties

The various US university patent and/or intellectual property policies offer a wide range of royalty splits among faculty, department and/or college, and university. The most generous policies split royalties with faculty 50–50. At the bottom of range of royalty splits, faculty receive one-third of the royalty income. Private universities tend to be less generous than public ones, with many offering faculty one-third of the income from their licenses. When policies change over time, they usually give faculty a lower percentage of royalties. Whether it is one-third or one-half, the incentives for faculty are powerful.

Personnel Coverage

In 1970s and 1980s, a number of patent policies covered only inventors. By the mid-1990s, they included faculty, staff, graduate students, post-doctoral fellows, non-employees who participate in university research projects, visiting faculty and, occasionally, undergraduates. If these personnel engage in research that might lead to an invention with market potential, they are required to disclose it, so an institutional decision about ownership and patenting can be made.

Exceptions

US universities had long claimed ownership of discoveries made by faculty; the decisive court cases were heard in

the 1950s. However, initially, there were exceptions to universities' ownership claims to intellectual property patented by faculty. If faculty made the discoveries on their own time, using their own resources, and not availing themselves of university facilities, they could claim a patent for themselves – for example, if they invented something in the summer, in their garage work room. As the academic capitalist knowledge/learning regime developed, definitions of time, resources, and facilities use were specified to the point where it became very difficult for faculty to assert any claims to ownership of invention. For example, policies frequently have guidelines that indicate that if researchers or other personnel use anything but routinely available office equipment and commercially available software or library materials generally available in nonuniversity locations, they are making substantial use of university resources and therefore the university owns the intellectual property.

Initially, state system and institutional policies addressed only patents. Over the years, the forms of intellectual property covered multiplied. Among those included were: licensing income, milestone payments, equity interest, mask work, which charts the topography of a semiconductor chip product, material transfer agreements, tangible property (cell lines, software, and composite matter), trade secrets, and copyright.

The multiple forms of market activity pursued by universities together with faculty's close involvement in them created many opportunities for conflict of interest. Factors that increase the possibilities of conflict of interest for faculty are: increased magnitude of personal compensation; growing numbers of financial relationships between a creator and a company; greater commitment of a faculty's time to a company; faculty or administrators holding equity in a company; involvement of trainees or students in a company; and involvement of patients or human subjects in company research trials. In other words, the risk of conflict of interest increases the more closely faculty members or creators of intellectual property are involved with market activity. Conflict-of-interest policies have proliferated and become more detailed. However, few call for close monitoring of faculty or institutional administrators, suggesting the primary purpose of such policies is to protect institutions from liability in legal cases.

Start-Up Companies, Research Parks, and Incubators

As universities became involved in research commercialization, they began to be regarded by state and federal policymakers and institutional leaders as engines of economic development. Universities became the center of a cluster of economic activities directed toward economic

growth in the state or region in which they were located. Licensing of inventions to local and regional companies was at the core of economic development activities, but start-up companies, research parks, and incubators also grew rapidly. These latter activities were seen as a way to spur regional economic development, particularly growth of high-paying jobs.

Start-Up Companies

Start-up companies began to increase rapidly among research universities in the 1980s. Start-ups are new companies dependent on licensing institutions' technology for their formation. Sometimes universities take an equity position in these companies. The investment can be limited to free use of the license, or can involve cash, or arrangement of venture capital partners to fund the company.

The number of equity deals spread rapidly among research universities. In 1994, the first year the Association for University Technology Managers reported such data, there were 175 start-up companies in US universities; in 2001, 402 new start-ups were reported. Equity provides several advantages over licensing: it gives universities options or financial claims on companies' future income; equity deals align interests of university and firm with regard to rapid commercialization of technology; and it signals to interested investors the universities' certainty about the worth of the technology. If the start-up company is successful, it moves from being privately held to being opened to public investment through an initial public offering (IPO), which greatly increases the worth of the stock of the company, and the monetary returns to the university. On the downside, taking equity involves risk of failure and loss of funds, while start-ups also often take up large amounts of faculty time.

Research Parks

University research parks generally house technology-based organizations or companies that seek to benefit from the host university's knowledge base and researchers. By housing the research park, the university encourages technology transfer. In addition to financial gain from technology transfer in which they are participants – for example, the research park may house university start-up companies – research parks often provide opportunities for students and faculty to participate in entrepreneurial science. The community and region gain if technology transfer results in economic development.

Research parks grew rapidly in the mid-1980s and a large number are currently in the planning states. Currently, there are 81 research parks associated with universities, and an additional 27 are being planned.

Information technology and biosciences are often the focus of research parks.

The contribution of research parks to economic development is unclear. The only studies of research parks to date account only for growth of employment within the park itself – about 8.4% over the life of a park. Approximately 50% of the parks were initially funded with public moneys, of which the public sector supported about 70% of initial park cost. Many research parks house activities unrelated to university research and development, in effect serving as a real estate operation on the part of the university, receiving income from rents.

Incubators

About 50% of research parks include incubators. In general, incubators foster start-up companies. As in the case of research parks, there are not a large number of studies of incubators. One study indicated that approximately half of start-up companies failed unequivocally in 3 years, while approximately 30% left the incubator and went out of their own. Start-ups with technology licenses from the sponsoring university and linkages to university faculty were less likely to fail, but less likely to leave the incubator at the end of 3 years.

Other Forms of Commercialization

Technology is arguably more reliably transferred to the market by means other than patents and other forms of university-owned intellectual property. Managers of industrial R&D repeatedly stress that the most important contribution of universities to innovation is training of scientists and engineers. They stress the importance of public research, highlighting transfer channels such as public papers and reports, conferences, informal exchange, consulting, and other forms of university–industry partnerships that do not depend on intellectual property. Patents are most important for the biological sciences, which account for the largest amount. They are less important in fields such as engineering, materials sciences, and computer science.

Copyright

The Telecommunications Act of 1996 dramatically altered the industry regulatory framework. Prior to 1996, the 1934 Communications Act, as implemented through the Federal Communications Commission, authorized separate monopolies: broadcast, cable, wire, wireless, and satellite. The 1996 Telecommunications Act deregulated these various industries, creating a competitive climate that favored

growth of the Internet, World Wide Web, and electronic (e)-business, all of which utilized previously separated communications media in new patterns. Deregulation of telecommunications created numerous possibilities for an academic capitalism knowledge/learning regime, ranging from software to distance education.

The Digital Millennium Copyright Act (DMCA) of 1998 protects digital property by prohibiting unauthorized access to a copyrighted work as well as unauthorized copying of a copyrighted work. The DMCA is far-reaching and covers an array of technologies, from web casting to hyperlinks, online directories, search engines, and the content of the materials made available by these technologies. Not only are citizens (and students) penalized for unauthorized access, but devices and services that circumvent copyright are also prohibited. The law very deliberately seeks to develop electronic commerce and associated technologies by strengthening protections of all forms of digital property. There are some exceptions, the broadest being for law enforcement and intelligence. The other exceptions are quite narrow.

US universities and colleges copyright some forms of software as well as educational products, tests and measurements, and other marketable products able to generate external revenues for colleges and universities. Examples are the Minnesota Multiphasic Personality Inventory, developed by the University of Minnesota and a video demonstrating minimal incision aortic surgery, developed at the University of Wisconsin. Most US universities now have copyright policies that claim institutional ownership of products created with university resources, all products developed by administrative personnel, and products stemming from work-for-hire contracts. Over time, university protection of intellectual products by copyright has become more inclusive.

Infrastructure

Commercialization of research entails considerable infrastructure development on the part of universities. Most universities now fund technology transfer offices, many also fund research parks and incubators and economic development offices. The managerial capacity in US colleges and universities has greatly expanded. The new functions are many: surveilling institutional employees' intellectual property activity to ensure capture by the system or institution through faculty disclosures; reviewing and evaluating faculty disclosures; technology licensing; supervision of royalty flows, including distribution of funds within institutions; reinvestment of funds in new market activities; litigation to defend intellectual property; evaluation of intellectual property for institutional equity investments; monitoring and occasionally administering corporations in which the institution holds equity;

overseeing IPOs; developing and monitoring market activity for conflict-of-interest issues. As colleges and universities become more involved in academic capitalism, they hire more managerial professional staff. Expanded managerial capacity institutionalizes business activity in colleges and universities by allowing segments to directly engage the market.

Studies of university income from intellectual property suggests that the following contributes to success: past interactions with industry; hiring and retaining star faculty; greater numbers of experienced technology licensing officers; close ties to venture capital prior to initiation of a start-up company; and a culture supportive of entrepreneurial endeavor.

Some related costs which are rarely included in cost-benefit analyses of commercialization of research are: establishment or expansion of arms-length organizations, usually foundations, that distance market ventures from public institutions, institutional research boards (IRBs), which approve research with human subjects, essential to clinical trials of pharmaceutical and medical devices products; administrative policy development, ranging from conflict-of-interest policies to ownership policies; and faculty time. Faculty time is consumed in several ways. For example, faculty participate on committees such as IRB advisory committees, intellectual property committees, and conflict-of-interest committees. When faculty are involved in start-up companies, or in incubator activities, they very often work closely with the new firm, taking time away from other forms of research, including grants writing, as well as teaching and committee work.

As the infrastructure for commercialization at universities develops in the United States, it articulates with dense networks in the broader society that support research. These networks range from the National Institutes of Health, and the other federal mission agencies that provide research funding, to hospitals, venture capitalists, state economic development agencies, and to small firms. Over time, these networks become more closely interwoven, promoting rapid commercialization through regional and national links. The union of graduate education and research in US research universities promotes interdisciplinarity that fosters innovation. In general, the US system is more heterogeneous than others. Europe, for example, has a less diverse group of public research institutions that works in narrower scientific areas, and lacks upstream public funding agencies such as the National Institutes of Health.

Conclusion

In general, academic capitalism promotes research commercialization that shifts the emphasis of universities toward graduate-level science and technology. This

occurs because of the infrastructure costs discussed above. Rather than becoming a source of funding for undergraduate liberal arts and professional programs, in which the large majority of students are enrolled, research commercialization usually calls for greater expenditure on infrastructure and perceived high-opportunity product investments. Although funds from research commercialization are returned to universities to support education and research, many universities have moved away from defraying general support programs and instead invest these funds in further commercialization efforts.

Through shifts in incentive systems, emphasis is placed on entrepreneurial research. One indicator is the dramatic rise in patenting by universities. Another is the flattening of scientific journal publishing rates. The number of science and engineering (S&E) articles by authors based in the United States has remained flat since 1992, even though real R&D expenditures and the number of researchers continued to rise. A similar shift was also seen in Canada, the United Kingdom, and the Netherlands. At the same time, a rapid growth of article citations in patents occurred. The growth of citations of scientific research in patents points to the close ties between research and products that are research driven, primarily in the life sciences. Taken together, the rise in patenting, the flattening of article production, and the growth of citations of science in patents suggest a realignment of the research system to promote commercialization.

Most of the literature on research commercialization appears in economics journals or in specialty journals, such as *Research Policy* and *Management Science*. This literature is focused primarily on S&E fields, especially activity in the life sciences, and on increasing innovation and economic development, and on expanding the revenue from university intellectual property. Generally, it takes the position that for economic development to continue, innovation to occur, and revenue to increase more institutional funds should be invested into the infrastructure described above. This literature does not address the university as an organization engaged in undergraduate education across an array of fields other than science and technology, nor does it directly address the critique of research commercialization, which has appeared in a variety of disciplines and also reaches across disciplines.

The critique of research commercialization points out that most research universities do not make money on technology transfer even though they expend substantial sums. Even among the top-rated US universities, only a small number generate a substantial revenue stream, and many of the infrastructure costs discussed above are not figured in the net costs. Even when research universities do generate substantial revenues, these are only a fraction of the grant and contract support from public and industrial sources. Patent and licensing income was \$850 million for

academic institutions in 2004, as compared to \$42 billion in support for grant and contract support for academic research and development.

Another strand of critique points to the erosion of the scientific commons through the expansion of private rights in intellectual property. This critique notes that rapidly published open science contributes to scientific advance. As the number of patents and copyrights increases, access to the scientific commons decreases, and at some point in the not-too-distant future, scientific advances may be slowed and innovation stifled. A related critique notes that as greater numbers of university scientists become involved in commercial research, the university loses its claim to represent disinterested knowledge. If the majority of molecular biologists serves on the boards of biotechnology corporations, how can these scientists and the universities that have invested in their intellectual property impartially judge the merit of the science? This critique makes the case that commercialization undercuts a central public function of universities: disinterested assessment of knowledge.

The broadest level of critique suggests that commercialization of research contributes to undermining public trust in universities. This critique suggests that commercialization of research may be funded by undergraduate tuition rises. As tuition has dramatically increased, so has the amount of funds research universities spend on research: institutional expenditures for research rose from 11% in 1972 to approximately 20% in 2004. High tuition for undergraduates without concomitant attention to them may cause the public to lose trust in higher education, which undercuts support.

Patents dramatically illustrate the growth of the academic capitalist knowledge/learning regime. Patents, licensing and running royalties, start-up companies, and universities' equity positions in corporations built on faculty patents are market behaviors that involve nonprofit institutions in profit taking. Yet, colleges and universities are not market entities because they do not disburse profits to shareholders. Instead, funds from external market revenues are plowed back into the institutions. In some ways, colleges and universities that patent are able to cross the traditional borders between public and private, engaging in practices that best meet their needs for generating external revenues. If patenting, which is expensive, fails to lead to licenses and royalties, the state bears the cost in the case of public institutions. Similarly, the start-up corporations initiated by universities are in many ways a form of state-subsidized capitalism. Start-ups are risky undertaking: only one in ten is likely to succeed. However, faculty and universities are spared the discipline of the market in that there are relatively few direct penalties for failure. Income from royalties and licenses are tax free, so long as the profits are returned to the university. Although patenting and technology transfer are generally portrayed

as a win-win endeavor, a relatively small number of large research universities are the only ones to generate substantial external revenues. For many smaller colleges and universities, the cost of maintaining a technology transfer office exceeds any revenues.

See *also*: Funding of University Research; Research Quality Assessment; The Management of University Research.

Bibliography

Harvey, D. (2005). *A Brief History of Neoliberalism*. Oxford: Oxford University Press.

Further Reading

- Benneworth, P. and Charles, D. (2005). University spin-off policies and economic development in less successful regions: Learning from two decades of policy practice. *European Planning Studies* **13**(4), 537–557.
- Bollier, D. (2002). *Silent Theft: The Private Plunder of our Common Wealth*. New York: Routledge.
- Chan, A. and Fisher, D. (2006). Academic culture in Canadian universities: The contexts of change. In Gingras, Y. and Roy, L. (eds.) *Les Transformations des Universités du X111 au XX1 Siècle*, pp 175–203. Québec: Presses de l'Université du Québec.
- Cohen, W. M., Nelson, R. R., and Walsh, J. P. (2002). Links and impacts: The influence of public research on industrial R&D. *Management Science* **48**(1), 1–23.
- Feldman, M., Feller, I., Bercovitz, J., and Burton, R. (2002). Equity and the technology transfer strategies of American research universities. *Management Science* **48**(1), 105–121.
- Geuna, A. and Nesta, L. J. J. (2006). University patenting and its effects on academic research: The emerging European evidence. *Research Policy* **35**, 790–807.
- Kirp, D. (2003). *Shakespeare, Einstein, and the Bottom Line: The Marketing of Higher Education*. Cambridge, MA: Harvard University Press.
- Krimsky, S. (2003). *Science in the Private Interest: Has the Lure of Profits Corrupted Biomedical Research?* Lanham, MD: Rowman and Littlefield.
- Link, A. N. and Scott, J. T. (2006). U.S. university research parks. *Journal of Productivity Analysis* **25**, 43–55.
- Marginson, S. and Considine, M. (2000). *The Enterprise University: Power, Governance, and Reinvention in Australia*. Cambridge: Cambridge University Press.
- Mark, B. (2002). *Great Transformations: Economic Ideas and Institutional Change in the Twentieth Century*. Cambridge: Cambridge University Press.
- Mowery, D. C., Nelson, R. R., Sampat, B. N., and Ziedonis, A. A. (2004). *Ivory Tower and Industrial Innovation: University–Industry Technology Transfer before and after the Bayh–Dole Act*. Stanford, CA: Stanford Business Books.
- National Science Board (2006). *Science and Engineering Indicators 2006*, two volumes. Arlington, VA: National Science Foundation (volume 1, NSB 06-01; volume 2, NSB 06-01A).
- O'Shea, R. P., Allen, T. J., Chevalier, A., and Roche, F. (2005). Entrepreneurial orientation, technology transfer and spin-off performance of U.S. Universities. *Research Policy* **34**, 994–1009.
- Owen-Smith, J., Riccaboni, M., Pommolli, F., and Powell, W. W. (2002). A comparison of U.S. and European university–industry relations in the life sciences. *Management Science* **48**(1), 24–43.

- Powers, J. B. (2003). Commercializing academic research: Resource effects on performance of university technology transfer. *Journal of Higher Education* **74**(1), 26–50.
- Powers, J. B. The profitability of technology licensing. Ms prepared for submission to *Science*.
- Richard, P. (2003). *Unholy Trinity: The IMF, World Bank and WTO*. London: Zed Books.
- Rothaermel, F. T. and Thursby, M. (2005). Incubator firm failure or graduation? The role of university linkages. *Research Policy* **43**, 1076–1090.
- Slaughter, S. and Rhoades, G. (2004). *Academic Capitalism and the New Economy*. Baltimore, MD: Johns Hopkins University Press.
- Slaughter, S. and Leslie, L. L. (1997). *Academic Capitalism: Politics, Policy and the Entrepreneurial University*. Baltimore, MD: Johns Hopkins University Press.
- Soley, L. (1995). *Leasing the Ivory Tower: The Corporate Takeover of Academia*. Boston, MA: South End Press.
- Washburn, J. (2005). *University Inc.: The Corporate Corruption of Higher Education*. New York: Basic Books.

Research Quality Assessment

I McNay, University of Greenwich, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Quality of research is assessed at various stages: in reviewing proposals for project funding; within a team as work progresses; at output stage when reports are presented, or articles submitted for publication, or new drugs or inventions seek development funding; when outcomes are clear after research-based changes are made in policy or practice. Major programs by research councils, foundations, or governments can be reviewed.

Over the past 20 years, national processes for research-quality assessment have been introduced across the world, led by the United Kingdom and other countries with an Anglo-Saxon influence. The drivers for this development included massification, marketization, new public management, and the development of the evaluative (Neave, 1998) or regulatory (King, 2007) state. Growth in student numbers caused a shift in the concept of university academic staff from combining roles in both teaching and research, specifically where, as in the United Kingdom and Australia, binary systems were merged into unitary ones, or where professional-education institutions were funded by the same body as more purely academic ones, as in New Zealand. There was a perceived need to identify research active (= productive) staff so that funding followed function and performance. This was seen as part of a competitive strategy in a globalized knowledge economy where research becomes a commodity and outputs become assets (Broadhead and Howard, 1998). The results of the assessment then informed funding strategies or, indeed, broader strategic steering. Publication of results provided comparative quality data to inform customer choice and created competitive imperatives to improve performance. Such competition, either in bids or in the research-assessment game (Lucas, 2006), was followed by concentration of funds and a belief in a minimum critical mass of research groupings to attain world-class status. Research, therefore, became better managed and administered (McNay, 1997, 2007), though there has been continuous debate about the evaluative and regulatory frameworks for the assessment exercises, and even about definitions of research and therefore what should be included in any assessments.

This article covers only work done within academic contexts; work in the commercial world has a different set of conditioning factors and criteria. Different national approaches to recognition of research excellence in higher education (HE) are identified. It then considers more fully those HE systems with extensive evaluation

exercises – the processes used and the criteria applied. It records recent moves to international comparative exercises. Finally, it summarizes some of the critical issues arising from research on assessing the assessment and the responses by academic staff subjected to evaluation.

National Models

In the United States, public universities are not funded at federal level for research infrastructure. Branch *et al.* (2001) offer a primer on 11 different evaluation methodologies in use by different agencies and states (Michelson, 2006: 555). The only nation-wide evaluation is by proxy, through reputational rankings of research doctoral programs. Michelson identifies trends in research evaluation across major federal agencies, such as the National Science Foundation (NSF) and National Institutes for Health (NIH), that fund research and development (R&D), linked to a further purpose of such evaluations – justification of further, increased government funding to those agencies and their programs. That is a driving force for similar programs such as the European Union (EU) Framework or the agendas of the UK research councils as part of the government's comprehensive spending review. At the European level, Luukkonen (2003) identifies some difficulties in evaluating the EU Framework program, applicable also at lower levels: there are changing political agendas and the objectives are fuzzy. They embrace promoting competitiveness, and advancing cohesion, quality of life, job creation, and sustainable development in the member states. For particular programs, evaluation may even start by negotiating objectives and desirable outcomes, which suggests poor criteria setting at an earlier stage. The impact studies cited by Luukkonen suggest that many outcomes are intangible, and related to learning and networking, not to commercial or economic effects.

In the United States, the political agenda is also an influence, but, within a constantly evolving and shifting context for the evaluation task, set by the Government Performance and Results Act, Michelson detects moves to standardization and harmonization of methodologies across agencies with increasing use of quantitative data, particularly bibliometric analysis, but paired with a “renewed focus on utilizing qualitative indicators . . . to create more appropriate hybrid methodologies that can capture a wider range of variables related to a programme's performance” (p. 558). The European approach is more qualitative, using surveys

and interviews with participants on their perceptions of impact. Hybrid approaches will help avoid the trap of false precision (Lawrence, 2003: 261) and allow the most appropriate and relevant measures to be used. So, fitness for purpose becomes the quality criterion for *post hoc* evaluation methods; that should also, surely, be a major criterion for assessing projects at the proposal stage.

Key terms in these last two paragraphs – programs, projects, proposals, purposes, perceptions of impact – indicate an emerging framework for the focus of research-quality assessment. There are two other elements that come through in country studies: people and products, with the latter concerned mainly with output – publications, patents, licences, and other such artifacts – rather than outcomes – the impacts on the real world – which are harder to assess in terms of isolating cause/effect relationships and which may only be visible beyond the timeframe within which evaluation reports have to be prepared. The other key variable in many countries is the organization level on which the assessment focuses, when this relates to overall performance, rather than being program or project specific. Different approaches use a combination of product–each piece of output–people, using a portfolio approach, group–team, academic unit–field of enquiry/discipline, or institution.

The people-based processes range from Mexico and Spain to New Zealand. In the first two, individual academics put forward their portfolio for evaluation and accreditation after judgment by peers. This is organized through a national body. In Spain, this is the *Agencia Nacional de Evaluación y Prospectiva*. Submissions, which are voluntary, are made every 6 years and success means a small salary increment: 100 euros a month for all successful *sexenios*. Although the results are formally confidential, professors become labeled as *tres sexenios*, or however many they have completed – the process began in 1986.

In New Zealand, performance-based research funding was introduced in 2003, to replace funding related to student numbers, though student numbers still condition the total quantum of research funds available for distribution to institutions. It is compulsory for all academic staff to submit a portfolio, which is assessed within the institution with judgments submitted to a set of external peer-review panels for validation, resulting in a personal grade – A, B, C, R – where the last implies reject as not research active, or not at a recognized level. These personal-quality grades, again supposedly confidential, are aggregated at subject and institutional level and weighted in a funding formula that also takes into account quantitative data on research-degree completions and external research income in a ratio of 60:25:15. There has been an open debate on the first round of the exercise (Bakker *et al.*, 2006) but the second round of judgments proceeded before a formal review was completed.

In the smaller European countries, systematic research-quality assessment is usually done, if at all, at

institutional level. Sorensen (2003) notes the gentler touch in Scandinavia compared to the United Kingdom, as well as the greater emphasis in Denmark on innovation and interaction with the user community. For Norway, Larsen (2000) records the continuing tradition of institutional autonomy and academic freedom such that evaluation is at institutional level. The major concern among his respondents was about the use of research programs as a policy instrument, following the establishment of a single research council and encouragement from the ministry for research results to be documented and evaluated, as the governance of research responds to pressure for a more coherent policy frame.

Presentations to the United Nations Educational, Scientific and Cultural Organization (UNESCO) Forum on Research Management showed that across major world regions – Latin America and the Caribbean, Africa, Arab countries in the Middle East, and much of Central and Eastern Europe – research-quality assessment is done at input/bid stage or is conducted at institutional level related to staff promotions and internal strategic development and resource allocation. Little systematic assessment is conducted at the national level (Braddock and Neave, 2002; UNESCO, 2006).

In emergent economies in Southeast Asia, there is concentrated funding to achieve world-class competitive levels. In South Korea, this is the Brain Korea 21 project (BK21). This focused funding on selected departments with a resultant increase in publication productivity. In the People's Republic of China, two projects – Project 211 and Project 985 – had a similar aim, and process but with selected institutions, not departments. Full-impact studies are not yet available for either country, and both have had a surge in student numbers putting extra pressure on the system. There are concerns about the criteria for selecting the favored institutions, with Kim (2007) recording the Matthew effect in Korea – to those who have will be given, from those who have not will be taken away the little they have. For China, Lai and Lo (2007) have concerns about the impact on decisions of *guanxi* – social and professional network connections and contacts, so that who you know is more important than how good you are. That, of course, is true in many other countries where an established cadre of researchers controls who can join the club.

Institutional audit locates research evaluation in a whole institution context, and so can link it to teaching, enterprise, knowledge transfer, and the environment in terms of facilities. Often, such linkages are not made (McNay, 2009: 38), and research is evaluated in isolation from its contribution to teaching, to the student experience, and to economic, technological, or social improvements, or even to research-related activities such as collaborative links, capacity development beyond the university, or developmental engagement with practice. There is a growing body of work giving evidence of

the essential relationship between research and teaching (Southampton Solent University, 2007).

Campbell (2003) examined four countries and classified them as types A and B. The type A two – United Kingdom and Netherlands – have comprehensive ex-post research evaluations, covering and addressing all disciplines at national level (p. 98). In Germany and Austria – the type-B pairing – he found a contrasting approach: pluralized and situational with a variety of individual and *ad hoc* evaluations (p. 113). He put Finland and Switzerland close to the type-B cluster. There was no correlation between approaches and the percentage of research funded from general university funds, which in Austria and the Netherlands was over 70% and in the United Kingdom was below 40%. In Germany and Switzerland, as in the United States and Canada, infrastructure funding comes from the canton, province, or state, not the federal governments, which fund major projects. Von Tunzelmann and Mbula (2003) detected an emergent interest in federal-level assessment in Canada. The situation in Germany is also changing. The German Research Foundation (DFG) has published periodically a ranking of research institutions based on funding approvals. There is now to be a rating exercise, every 5 or 6 years using a seven-point scale decided by expert panels, and influencing future funding. This comes close to the UK model. In Pakistan, there is a ranking exercise, with research output as one criterion. Results are not made public, but individual heads of institutions are given their position against each criterion, in an attempt to promote development through competition based on benchmarking.

Campbell noted a trend toward increasing funding through programs and projects, allowing research to be tied more closely to strategic national priorities. That is true across most countries and is manifest in the development of institutional contracts – Finland, Denmark, and Germany – or in designation of centers/networks of excellence – Switzerland, Norway, Japan, Australia – where continued funding is dependent on satisfactory evaluation. In South Africa, most research funding is from national-level state agencies, linked to nation building and renewal. Sweden is one country still resisting this close alignment, and does not tie funding to output-based evaluation. Ireland, despite a tenfold increase in R&D spending over the past decade, is another country not yet embedded in performance-based funding. Student numbers are still a dominant conditioner of funding in countries such as Italy and Hungary, and it is worth noting that a major element in the Finnish contract is the number of completed research degrees.

Trends noted by von Tunzelmann and Mbula (2003) include:

- A shift to formative evaluation linked to prospective funding.
- The increased use of international comparators – benchmarking or bibliometrics, though both have underdeveloped theoretical and processual bases, with only the Flanders region of Belgium having developed any level of sophisticated use of bibliometrics, and then only for a limited range of science-based subjects. Norway is moving to make increased use of them for research funding decisions.
- Greater emphasis on institutional self-evaluation. This often has a three-stage process of internal evaluation, external peer assessment based on that self-evaluation, and a final public report agreed between the institution and the external government agency. That is the model in the Netherlands, France, Switzerland, and Norway.

Those trends may be a balance to shifts to performance-based funding (PBF). The Netherlands and New Zealand are moving to join countries in which PBF is well established, such as the United Kingdom, Poland, Slovakia, Hong Kong, and Australia. The difficulty, then, is to separate the impact of esteem based on the quality rating from the impact of the economics of selective funding. A study of South Africa by Auf der Hyde and Mouton (2007) shows a decline in output per person and a withdrawal of some staff from assessment after the link of funding to performance was discontinued. Himanen *et al.* (2009) used longitudinal Organization for Economic Cooperation and Development (OECD) data on five countries – Netherlands, United Kingdom, Norway, Finland, and Australia – to conclude that:

There is no straightforward relation between the competitive funding environment and research performance . . . The impact of incentives on research performance appears to be quite short-term and sometimes even negative. However, the model of state steering does have an effect. The institutional steering model, which emphasises university independence from the state, seems to be the most beneficial to research performance. (p. 429)

The next sections provide greater detail on four countries with long-established systems and processes.

The Netherlands

In the Netherlands the national evaluation system for publicly funded research aims to improve research quality, to improve research leadership and management, and to provide accountability. Institutions – universities and other research organizations – conduct a self-evaluation every 3 years with external evaluation every 6 years. The self-evaluation documents look back 5 years and also forward 5 years, giving data on funding, staffing, research context and culture, researcher training (capacity building), esteem factors, collaborations, and external validations. The criteria used by the external assessment are

quality – indicators of recognition and identification of future potential for innovative work; productivity (output); relevance – impact on the scientific community and broader society; and vitality – quality of leadership and management. The resultant profiles are reputational and do not (yet) have any formal, direct impact on funding – the emphasis is on formative assessment. The process is now conducted by an independent foundation – Quality Assurance Netherlands Universities (QANU) – set up by the Association of Dutch Universities (VSNU). There are still five grades of quality, but with a shift to have four positives and only one negative, from a previous balance of 2 + 1 + 2.

Hong Kong

In Hong Kong, there has been a research-assessment exercise (RAE) since 1993 (Hong Kong University Grants Committee, 2006). It is conducted through panels of expert reviewers, based on cost centers and aims to inform decisions on funding. It applies a single threshold of quality – a level of excellence appropriate to the discipline in Hong Kong and showing evidence of international excellence. That threshold has been raised over the four exercises, the latest in 2006. The evidence presented includes an institutional self-assessment including a statement of research strategy including research philosophy, vision, and priorities in relation to the role and stage of development of the institution (the Hong Kong University Grants Committee (UGC) wants to promote institutional role diversity). Academic staff submit up to six outputs, identifying the best four, and one can be enough to satisfy the threshold requirements. They are assessed against all four of the Carnegie scholarships – discovery, integration, transmission/teaching, and application (Boyer, 1990). The quality criteria demand an element of innovation; a contribution to scholarship; public accessibility; generalizability; and an interest to peers. Research should support and illuminate teaching and learning, and outputs with social relevance are encouraged. Institutions should demonstrate sustainability of research work and evidence of deep collaboration. Results are confidential and no individuals are rated.

United Kingdom

In the United Kingdom, the sixth RAE took place in 2007/8, with changes from previous exercises and further changes to come after proposals presented as part of the 2007 budget speech by the Chancellor of the Exchequer (Minister of Finance), which points to the driving force behind the exercise. Previously, expert panels – over 60 – had rated four outputs per staff member submitted as being of international, national, or subnational quality, and arrived at an average grade for a unit of assessment,

usually an academic department. There was a seven-point scale depending on the balance of quality of output across those three categories. Funding was gradually withdrawn from all but the top two grades in successive exercises. Each unit also submitted a statement of research strategy and environment, and data on research degree students and research income, but the major element in judgment was the output. So, judgment was on individual output, grading was at unit level, and funding was to the institution with no requirement to allocate it to units or to staff who had generated it. The amount allocated was on a formula of V (number of staff, i.e., volume) \times SQ (the price per subject adjusted on a scale reflecting quality rating). Those last two elements were not revealed until after submission of bids, and the criteria for quality judgment were fuzzy; so much of the exercise was seen as a game (Lucas, 2006; McNay, 2003). If fewer staff were submitted, the balance of output could be shifted – the quality rating might be higher, but the volume multiplier was lower. Different panels operated different standards, and so choice of which panel to use was also part of the game. To add to complexity, funding decisions vary across the four constituent countries of the United Kingdom (McNay, 2007).

For the 2007/8 exercise, there were still panels, but output was given a grade profile, not a single average, and the overall unit profile was conditioned by separate ratings of environment and esteem indicators. Work was judged against three main criteria – originality, significance, and rigor. There were four grades (and an unclassified rating), ranging from world leading, through internationally excellent, but . . . , recognized internationally to recognized nationally. Every panel gave its own interpretation of those terms and also decided on the balance between output, environment, and esteem factors, with the first counting for between 50% and 90% of the final profile, and the other two between 5% and 40% each. Much time was invested in divining meanings of the policy texts. The assumption that funding would be restricted to the top two grades and further concentrated in a very small number of universities proved to be unfounded, with a commitment to fund excellence wherever it was found. The shift to profiling output, and away from a single grade for an academic unit showed that 17% of outputs were deemed world leading (4*) and 37% were internationally excellent. That total of 54% can be compared to the 2001 result, when 55% of individual academics were in units rated at 5 or 5*. The change was that, for 2008, 150 out of 159 institutions had work in the top category (Corbyn, 2008). There was a consequent shift in funding away from concentration on a few elite institutions, where four institutions received 30% of the resources allocated to a flatter, more widespread distribution.

A review of the 2001 exercise had found only four respondents from over 400 in favor of a move to using

an exclusively metrics-based funding system, mainly matching research income from external sources. A Treasury-led initiative subsequently proposed moving to such a metrics-based system for the future. Much agitated discussion has moderated that somewhat to a hybrid model, but there are hints of more emphasis on impact beyond academic circles (Corbyn, 2009b) – the political pressure, based on respect for the market is very evident. It is worth noting that, in the United Kingdom, the research councils were based previously within the government's department of trade and industry; the whole of HE is now within a super-ministry of business, innovation, and skills, with obvious implications for the priority functions of the sector.

Australia

Australia was moving in the opposite direction, after an expert group, chaired by the person who had reviewed the UK 2001 exercise, recommended adoption of something close to the UK model. The previous approach had resulted from a rejection of that approach in the light of evidence from McNay's work (Bourke, 1997; McNay, 1997). It had used three metrics as proxy for quality – number of publications, external research income, and research student numbers and completions. The new system (DAG, 2006) was to use panels to assess quality and impact. The quality descriptors echo those of the United Kingdom, with four acceptable grades ranging from world leading through meeting world standards of excellence, recognized internationally as excellent to methodologically sound . . . and of high originality, significance and rigor. Those three terms apply within all grades, and the top two also add reference to work of particular significance to Australia. Impact would be assessed separately and at longer intervals, also using a five-point scale. The top grade – A – was where “adoption of the research has produced an outstanding social, economic, environmental and/or cultural benefit for the wider community, regionally within Australia, nationally or internationally.” Next down, adoption has produced significant benefit. Level C has produced “new policies, products, attitudes, behaviours, and/or outlooks in the end-user community,” and the lowest recognized grade has “engaged with the end-user community to address . . . an issue.” However, after the politicians changed the policy, the people changed the politicians in 2008 and a further review is underway as this is being written.

International Initiatives

Since many universities now aim to be world class (Lang, 2005), a number of international comparative exercises have developed in recent years. Internal institutional evaluations may be informed by voluntary audits of

whole institutional performance through the European Association of Universities, or participation in benchmarking exercises through the European Centre for Strategic Management of Universities (ESMU). The Association of Commonwealth Universities also conducts benchmarking (Kirkland *et al.*, 2006). The OECD has a program on research management (Connell, 2004), as does UNESCO.

Shanghai Jiao Tong University produces an academic ranking of world universities (Ranking Group, 2004; Liu and Cheng, 2005), based on research indicators, such as prize winners among academic staff and alumni (including dead ones), articles in (only) *Science* and *Nature*, citations, using Science Citation Index (SCI) and the Social Science Citation Index (SSCI). Arts and humanities are excluded from the exercise. So, there are several criticisms of methodology (O'Leary, 2004; Harfi and Mathieu, 2007). UNESCO has set up an international rankings expert group to promote dialog on this topic among stakeholders (Merisotis and Sadlak, 2005; Merisotis and Leegwater, 2007) but there are many problems of comparability as well as methodology to overcome.

Key Issues

Those extended descriptions of diverse approaches serve to show the complexity of the task and the degree of prescription that can be imported into guidelines for making a complex judgment. In many cases there are concerns about criteria, processes, and evidence. The final section turns to some issues to consider for any extended process of research-quality assessment.

The first issue is the framework for judgment. What are the boundaries to research? How far should it be judged in relation to other functions that might be seen as discrete? Approaches based on whole institutional audit as in Ireland, Finland, or the Netherlands can make the linkages clearer in the self-assessment and allow fuller integration into corporate strategies. In the Netherlands, the process covers research units outside the universities. In Hong Kong, research evaluation links all four of Boyer's scholarships – of discovery, integration, transmission, and application. In the United Kingdom, the emphasis is on the first. There is specific separation of quality assessment of teaching from assessment of research. They are done by separate processes and neither informs the other. By contrast, in the United States, the NSF has a specific criterion in merit review of projects requiring integration of research and education; in New Zealand, the PBRF aims to enhance the quality of research-based teaching. In the United Kingdom, there have also been criticisms by the Education Ministry, Parliamentary Committees, and some research councils that the RAE approach gives too little weight to interdisciplinary work, applied work, and

to research derived from professional and practice contexts, despite the presence on panels of user representatives (McNay, 2007). In Scandinavia, more importance is given to innovation and links with practice related to application and knowledge transfer. The Australian approach, hitherto, has given greater emphasis to income from the market as a proxy measure for quality, linked to relevance and applicability, but this risks transferring control of the research agenda to one of short-term pressures demanding immediate results, and undermining basic, blue skies work. Harman (2005) suggests that academics do accommodate to such an environment and negotiate degrees of autonomy within it. The proposals for assessing impact, now overtaken by political change, had been considered an improvement in recognizing research users as legitimate stakeholders in assessing quality.

Autonomy is a second issue, especially when linked to funding. Research locates within an institutional framework and can be judged against its contribution to institutional strategic objectives. Projects funded within national or international programs will have fitness for purpose as a dominant criterion with the purpose defined within strategic priorities set within a wide socioeconomic policy framework. The twin issues, then, are how far such a criterion should also apply to other research, and to what degree the research agenda should be set by researchers themselves. There is, now, considerable evidence (MacGregor *et al.*, 2006; Morris and Rip, 2006; McNay, 2007) that assessment exercises affect the behavior both of researchers in their choice of projects and of research leaders/managers who set priorities within institutions and select staff for sponsorship and submission to the processes of assessment. Any link of assessment outcomes to PBF increases the impact on behavior (Liefner, 2003). Geuna and Martin (2003), having reviewed approaches in 12 countries, warn that PBRF may have initial benefits outweighing costs, but over time produces diminishing returns. They tabulate 15 factors under both advantages and drawbacks.

Among the claimed advantages are:

1. *Efficiency.* It is low cost; cuts out poor research; encourages discipline in completion and publication; and resource concentration through competition gives an incentive for improvement, with a meritocratic reward system for individuals that encourages inclusive engagement.
2. *Policy fit.* The accountability framework encourages a more explicit and coherent institutional strategy which can be linked to changing priorities, while allowing autonomy to researchers, diversity of practice, room for longer-term work, and support to links with teaching.

Their list of drawbacks challenges some of those items:

1. Resources may not flow to the best, with allocative power held by a small number of officials, which reduces incentives and motivation.

2. Rewarding past performance discourages innovation and can promote homogeneity, not diversity.
3. Links with teaching or policy priorities may be notional at best, and so may provoke further, excessive government interference to achieve objectives.
4. The process can be staff intensive and expensive.

A third set of issues revolves around the balance between formative and summative assessment, particularly if the latter has punitive financial consequences. There is some evidence from the United Kingdom, Australia, and New Zealand that lower-rated and lesser-funded units give better value for money (Grichting, 1996; Hazledine and Kurniawan, 2006), that their gearing between state funding and entrepreneurial income is better than for the elite few (Adams, 2005a, b), and that investment in improvement, with incentives for achievement, would yield greater leverage for quality enhancement across national systems.

The fourth issue is the level at which assessment is targeted, and results published. Most data collection starts with single items of output, but, increasingly, weight is now put on organization-context factors and a robust research culture. Paradoxically, outside project/program evaluations, the research team is rarely a focus. A review by Avital and Collopy (2001) found that evaluative studies were conducted at unit level, possibly because those levels were where the controllable enablers of productivity operate. Panels in the United Kingdom, Australia, and New Zealand relate to disciplines, but most research is interdisciplinary. The structural units awarded a grade are often built around the discipline in the context of teaching, but synergy with teaching may not be a quality criterion. In the United Kingdom, at least, members of interdisciplinary research teams, working across departmental boundaries, are separated from their research environment in the RAE submissions. Published results will always be used by the media to create league tables, to identify winners/losers, or the best/the rest binaries (Valadkhani and Worthington, 2006). More nuanced, formative feedback, with its narrative of strengths and weaknesses and potential, escapes the mass media, which may be why institutional audits have had less publicity than exercises where a score is awarded.

Those conditioning factors affect the evidence submitted for assessment. If methodological rigor is dominant, then the primary evidence is in project descriptions and published output. If relevance to policy and practice is a key concern, then input from users and evidence of outcomes assumes greater importance. Some outcomes may not be in published form and may take a considerable time to be evident, with causality between intervention and outcome difficult to audit. If international standing and competitiveness are strategic aims, esteem factors from the international community need to be collated. If institutional or departmental potential for improvement are key to

investment decisions, internal infrastructure, staff qualifications and continuing development, capacity building, and nurturing of new talent need to be demonstrated.

All of those then feed in to the most disputatious issue – what process to use. The Australian Expert Advisory Group (2005) dismissed the known benchmarking exercises, leading to league tables, because they are based on a narrow range of performance measures and have no input from end users. The Group's main debate, as with others in countries covered in this review, was between a peer-review process and a model based on performance indicators (PIs), labeled metrics in some contexts. They noted the debate in New Zealand recorded by the OECD (2004). The listings that follow draw on the Group's work and also on Scott (2006), Sastry and Bekhradnia (2006a, 2006b), and McNay (2003, 2007).

Advantages Claimed for Metrics/PIs

Some of the advantages of the PIs are:

- PIs can provide incentives for productivity, for recruitment, and retention of high-performing staff, for quality supervision to enhance student completions, and for generation of external funds;
- they can provide clear criteria for reward systems;
- data exist or are easy to collect;
- there are low transaction costs; and
- measures are transparent, formulaic, and so easily communicated and understood.

Possible Disadvantages in the Use of PIs/Metrics

Some of the disadvantages of the use of PIs are:

- the choice of metrics is not value free, so some work may gain and other work lose, depending on choice and weighting;
- if income is used as a metric, the control passes to those commissioning work, who may not support open publication of critical or unwelcome findings, or commission work with that risk;
- many more unsuccessful bids will be made, leading to inefficient use of time, and to reduced cooperation in a competitive context;
- some areas are more suited to commercial funding than others;
- quantity is not quality, nor is income necessarily an indicator of quality output;
- citations are not yet a reliable indicator, and are susceptible to manipulation;
- staff recruitment and selection for submission would favor established researchers, not newer ones without a record against the indicators, nor women taking career breaks;

- efforts to meet PI targets may displace commitment to teaching, student support, or community service; and
- if optimizing income is a main aim, there is potential for corruption of values and mission drift.

Advantages Claimed for Expert/Peer Review

Some of the advantages claimed for expert/peer review are:

- there is direct engagement with research output;
- a wide range of factors and criteria can be considered;
- the dynamic of development can be assessed to indicate potential;
- formative feedback can be provided;
- international expertise can be drawn upon through panel membership;
- centers of excellence can be identified;
- quantitative and benchmarking data are not excluded;
- bids can locate performance within interpretation of research and strategic priorities, allowing some element of made to measure in relation to fitness for purpose;
- trust of fellow professionals makes acceptance of judgments easier;
- a fuller, more rounded judgment will emerge; and
- the process can be located within differing disciplinary frameworks and fields of practice to recognize different research traditions and approaches.

Possible Disadvantages of Peer-Review Approaches

Some of the disadvantages of the peer-review approaches are:

- it is staff intensive and has high costs;
- it institutionalizes and legitimizes collective prejudice, thereby underrating new and nonmainstream approaches favoring, say, econometrics over feminist economics, and promoting convergence on a conservative agenda, entrenching academic traditionalism;
- little weight is given to the end users, so there is dominance of a perspective from the closed academic community, using academic journals, downplaying interaction with the wider community of practice, and communication/dissemination through professional or popular journals;
- panels are inconsistent across disciplines and over time;
- criteria may be obscure and the process opaque rather than transparent;
- there are concerns about equity of treatment on gender grounds, and of newer institutions, given the base of most panel members; and
- it has led to recruitment practices favoring high-performing stars who may not be those best suited to nourishing and mentoring others.

Conclusion

Finally, given the complexity of the process, the contentiousness of some approaches, as listed above, constant changes for political reasons (which means a short shelf life for some of the descriptions above), and the ingenuity of academics within an organic, fluid and flexible system, there will always be unexpected, unintended, and, perhaps, undesirable consequences, to balance the positive responses that arise from acceptance of the legitimacy of accountability and the striving for excellence, efficiency, and effectiveness that it may promote. In the United Kingdom, for instance, teaching programs have been closed down following less-than-excellent grades for research. Use of a hybrid process, clarity of objectives and criteria, transparency of approaches, robust data, linking of research strategy to that for other areas – teaching, enterprise, and community development – and, above all, sensitivity to consequences, will optimize the acceptance and validity of research-quality assessment.

See also: Funding of University Research; National Science Policy and Universities; Research Commercialization; The Management of University Research.

Bibliography

- Adams, J. (2005a). Never mind the quality, feel the width. *Research Fortnight* 28 September, 18–19.
- Adams, J. (2005b). Mirror, mirror on the bench. . . *Research Fortnight* 23 November, 18–19.
- Auf der Hyde, T. and Mouton, J. (2007). *Review of the NRF Rating System: Synthesis Report*. Cape Town: National Research Foundation.
- Avital, M. and Collopy, F. (2001). Assessing research performance: Implications for selection and motivation. *Sprouts: Working Papers on Information Environments, Systems and Organizations* 1(1), 40–61. <http://sprouts.case.edu/2001/010303.pdf> (accessed September 2009).
- Bakker, L., Boston, J., Campbell, L., and Smyth, R. (eds.) (2006). *Evaluating the Performance-Based Research Fund: Framing the Debate*. Wellington: Institute of Policy Studies: Victoria University of Wellington.
- Bourke, P. (1997). *Evaluating University Research: The British RAE and Australian Practice*. Canberra: ARC/HEC.
- Boyer, E. L. (1990). *Scholarship Reconsidered. Priorities of the Professoriate*. New York: Carnegie Foundation for the Advancement of Teaching/Jossey Bass.
- Braddock, R. and Neave, G. (2002). Research management in higher education: Overview and conclusion of a debate. *Higher Education Policy* 15, 217–224.
- Branch, K. M., Peffers, M. S., Ruegg, R. T., and Vallario, R. W. (2001). *The Science Manager's Resource Guide to Case Studies*. Seattle, WA: Department of Energy. Cited in Michelson (2006) q.v.
- Broadhead, L. -H. and Howard, S. (1998). "The art of punishing": The research assessment exercise and the ritualisation of power in higher education. *Education Policy Analysis Archives* 6(8), 1–13.
- Campbell, D. F. J. (2003). The evaluation of university research in the United Kingdom and the Netherlands, Germany and Austria. In Shapira, P. and Kuhlmann, S. (eds.) *Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe*, pp 98–131. Cheltenham: Edward Elgar.
- Connell, H. (2004). *University Research Management: Meeting the Institutional Challenge*. Paris: OECD.
- Corbyn, Z. (2008). RAE 2008 proves UK research is world class. *Times Higher Education* 1876(4–5), 18/25 December.
- Corbyn, Z. (2009). Impact may account for 30% of researchers' marks in REF. *Times Higher Education* 1906(6), 23 July.
- Development Advisory Group (DAG) (2006). *Research Quality Framework: Assessing the Quality and Impact of Research in Australia. The Recommended RQF*. Canberra: Department of Education, Science and Training.
- Expert Advisory Group (2005). *Research Quality Framework: Assessing the Quality and Impact of Research in Australia. Issues Paper*. Canberra: Department of Education, Science and Training.
- Geuna, A. and Martin, B. R. (2003). University research and funding: An international comparison. *Minerva* 41, 277–304.
- Grichting, W. L. (1996). Do our research unis give value for money? *Campus Review* 6, 29.
- Harfi, M. and Mathieu, C. (2007). Shanghai rankings and the international image of French universities. *IAU Horizons* 13(2–3), 8.
- Harman, G. (2005). Australian social scientists and transition to a more commercial university environment. *Higher Education Research and Development* 24(1), 79–94.
- Hazledine, T. and Kurniawan, C. (2006). Impact and implications of the 2003 Performance-Based Research Fund Research Quality Assessment exercise. In Bakker, L., Boston, J., Campbell, L., and Smyth, R. (eds.) *Evaluating the Performance-Based Research Fund: Framing the Debate*, pp 270. Wellington: Institute of Policy Studies: Victoria University of Wellington.
- Himanen, L., Aurunen, O., Puuska, H.-M., and Nieminen, M. (2009). Influence of research funding and science policy on university research performance: A comparison of five countries. *Science and Public Policy* 36(6), 419–430.
- Hong Kong University Grants Committee (2006). *Research Assessment Exercise 2006*. Hong Kong: University Grants Committee.
- Kim, Y. J. (2007). *The Impacts of Evaluation Based Funding on Academic Work: A Korean Perspective*. Unpublished doctoral thesis, University of Bristol.
- King, R. P. (2007). Governance and accountability in the higher education regulatory state. *Higher Education* 53(4), 411–430.
- Kirkland, J., Bjarnason, S., Stackhouse, J., and Day, R. (2006). *International Research Management: Benchmarking Programme*. Bristol: Higher Education Funding Council for England.
- Lai, M. and Lo, L. N. K. (2007). The changing work lives of academics: The experience of a regional university in the Chinese mainland. *Higher Education Policy* 20(2), 145–167.
- Lang, D. W. (2005). "World class" or the curse of comparison? *Canadian Journal of Higher Education* 35(3), 27–55.
- Larsen, I. M. (2000). University research policy in Norway – walking the tightrope between internal and external interests. *European Journal of Education* 35(4), 385–402.
- Lawrence, P. A. (2003). The politics of publication. *Nature* 422(20), 259–261.
- Liefner, I. (2003). Funding, resource allocation and performance in higher education systems. *Higher Education* 46, 469–489.
- Liu, N. C. and Cheng, Y. (2005). The academic ranking of world universities. *Higher Education in Europe* 30(2), 127–136.
- Lucas, L. (2006). *The Research Game in Academic Life*. Maidenhead: Society for Research into Higher Education/Open University Press.
- Luukkonen, T. (2003). Challenges for the evaluation of complex research programmes. In Shapira, P. and Kuhlmann, S. (eds.) *Learning from Science and Technology Policy Evaluation: Experiences from the United States and Europe*, pp 132–154. Cheltenham: Edward Elgar.
- MacGregor, R., Rix, M., Aylward, D., and Glynn, J. (2006). Factors associated with research management in Australian commerce and business faculties. *Journal of Higher Education Policy and Management* 28(1), 59–70.
- McNay, I. (1997). *The Impact of the 1992 RAE on Institutional and Individual Behaviour in English Higher Education: The Evidence from a Research Project*. Bristol: Higher Education Funding Council for England.
- McNay, I. (2003). Assessing the assessment: An analysis of the UK Research Assessment Exercise, 2001, and its outcomes, with

- special reference to research in education. *Science and Public Policy* **30**(1), 47–54.
- McNay, I. (2007). Research assessment; researcher autonomy. In Kayrooz, C., Akerlind, G. S., and Tight, M. (eds.) *Autonomy in Social Science Research: The View from UK and Australian Universities. International Perspectives on Higher Education Research*, vol. 4, pp 183–218. Oxford: Elsevier.
- McNay, I. (2009). Research quality assessment: Objectives, approaches, responses and consequences. In Brew, A. and Lucas, L. (eds.) *Academic Research and Researchers*, pp 35–53. Maidenhead: SRHE/Open University Press.
- Merisotis, J. and Leegwater, L. H. (2007). The UNESCO-CEPES/IHEP rankings initiative: Berlin principles on ranking of higher education institutions. *IAU Horizons* **13**(2–3), 6–7.
- Merisotis, J. and Sadlak, J. (2005). Higher education rankings: Evolution, acceptance and dialogue. *Higher Education in Europe* **30**(3), 97–101.
- Michelson, E. S. (2006). Approaches to research and development performance assessment in the United States: An analysis of recent evaluation trends. *Science and Public Policy* **33**(8), 546–560.
- Morris, N. and Rip, A. (2006). Scientists' coping strategies in an evolving research system: The case of life scientists in the UK. *Science and Public Policy* **33**(4), 253–263.
- Neave, G. (1998). The evaluative state reconsidered. *European Journal of Education* **33**(3), 265–268.
- OECD (Organisation for Economic Co-operation and Development) (2004). *A Performance-Based Research Funding for Tertiary Education Institutes: The New Zealand Experience*. Paris: OECD.
- O'Leary, J. (2004). Top performers on the world stage take a bow. *Times Higher Education Supplement* 5 November.
- Ranking Group (2004). *Academic ranking of world universities*. Shanghai Jiao Tong University. <http://ed.sjtu.edu.cn/ranking.htm> (accessed September 2009).
- Sastry, T. and Bekhradnia, B. (2006a). *Using Metrics to Allocate Research Funds*. Oxford: Higher Education Policy Institute. <http://www.hepi.ac.uk/466-1236/Using-metrics-to-allocate-research-funds.html> (accessed September 2009).
- Sastry, T. and Bekhradnia, B. (2006b). *Using Metrics to Allocate Research Funds: Response to the Government's Consultation Proposals*. Oxford: Higher Education Policy Institute. <http://www.hepi.ac.uk/466-1253/Using-metrics-to-allocate-research-funds-initial-response-to-the-Government%E2%80%99s-consultation-proposals.html> (accessed September 2009).
- Scott, A. (2006). Peer review and the relevance of science. *SPRU Electronic Working Paper Series*, vol. 145. Brighton: Science Policy Research Unit. <http://www.sussex.ac.uk/spru> (accessed September 2009).
- Sorensen, C. (2003). Scandinavian versus UK research: The importance of institutional context. *Scandinavian Journal of Information System* **15**, 95–99.
- Southampton Solent University (2007). Research and teaching colloquium: International policies and practices for academic enquiry. http://portal-live.solent.ac.uk/university/rtconference/2007/colloquium_papers.aspx (accessed September 2009).
- UNESCO (2006). Invitational workshop on the comparative analysis of national research systems. *Final Report*. Paris: UNESCO.
- Valadkhani, A. and Worthington, A. (2006). Ranking and clustering Australian university research performance, 1998–2002. *Journal of Higher Education Policy and Management* **28**(2), 189–210.
- Von Tunzelmann, N. and Mbula, E. K. (2003). Changes in research assessment practices in other countries since 1999. *Final Report to the Higher Education Funding Council for England*. Bristol: HEFCE.

Relevant Websites

- <http://www.hea.ie> – Higher Education Authority.
- <http://www.hefce.ac.uk/research> – Higher Education Funding Council for England (HEFCE).
- <http://www.innovation.gov.au> – Innovation Internet.
- <http://www.iuqb.ie> – Irish Universities Quality Board.
- <http://www.knaw.nl> – Koninklijke Nederlandse Akademie van Wetenschappen (KNAW).
- <http://www.nwo.nl> – NOW.
- <http://www.niad.ac.jp> – National Institution for Academic Degrees and University Evaluation.
- <http://www.qanu.nl> – Quality Assurance Netherlands Universities.
- <http://www.aka.fi> – Suomen Akatemia – Akatemia.
- <http://www.tec.govt.nz> – Tertiary Education Commission, Performance-Based Research Fund.
- <http://www.ugc.edu.hk> – The University Grants Committee.
- <http://www.rae.ac.uk> – Research Assessment Exercise (RAE).
- <http://www.unesco.org/education/researchforum> – UNESCO Forum on Higher Education, Research and Knowledge.
- <http://www.vsnul.nl> – Vereniging van universiteiten.

The Management of University Research

J Kirkland, The Association of Commonwealth Universities, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Background

The definition and scope of university research management is changing rapidly. While research has always required some degree of management in order to achieve its objectives, historically this was the province of the research team itself, together with such administrative assistance as may be available within the academic department. In many countries, such decentralized mechanisms were tied to the perceived right of tenured academic staff to set their own research agendas, utilizing time and often other resources provided by their employing institutions.

Most of this process took place with little or no regulation by the central university, or in some cases even the academic department concerned. In an age where the majority of research needs could be met from within the university system, external funding was seen an 'optional extra,' to be sourced and negotiated by the individual research team. In this classical model, the role of the university was one of very minimal regulation – for example, the provision of a basic accounting mechanism to satisfy the demands of audit. There was little expectation that additional resources would make a significant difference to the overall finances of the institution.

The Changing Environment

In essence, the main change in recent years has been the development of a much sharper distinction between research methodology and research management. The former relates to the means of conducting the research itself, ensuring that the results are of an appropriate standard. The expertise required for this is clearly academic, in most cases discipline based, and responsibility remains within the academic staff structure. Research management concerns broader issues, such as ensuring that staff are able and motivated to conduct such work, that it meets the needs of stakeholders (including the host university) and that the results are effectively utilized. The academic research team remains critical to this process, but it is increasingly supported (and monitored) by others. Research management has increasingly been seen as an issue to be conducted at the level of the institution.

Although the pace of change has varied, abandonment of this laissez-faire policy has been a common feature to most regions over the past 25 years. The pace of this change has not been even. The longest established

tradition of research administration can be found in the United States, whilst a survey by Stackhouse and Day (2005) amongst Commonwealth countries confirmed that developed countries generally were much further advanced. The same work also demonstrated, however, that the trend was becoming increasingly evident in developing countries also. Hazelkorn (2005) also demonstrated differences between types of institution. Her study of 25 new universities, many without strong research traditions, in 16 countries showed a capacity to align institutional competences to the external environment. In each of these cases, a primary motivation for change came from outside the academic institution, and to some extent outside the higher education system as a whole. Whilst the precise nature of change varies, it has been possible to advance a model of how research management systems develop in three (possibly overlapping) stages – awareness raising, the development of structures at institutional level, and the development of external support mechanisms which help ensure the sustainability of such structures (Kirkland, 2008).

Several interrelated themes can be identified. The first of these is the desire of governments for more accountability from universities. This has been pursued through a combination of incentives and threats and, although universities retain formal autonomy over the allocation of their own resources, has required the central university at least to collect a great deal more information about its own activities. The South African system tied university research funding directly to the publications by its researchers and the citations of their work published in designated peer review journals. Research Assessment Exercises in the United Kingdom have sought to allocate core research funding according to performance on a range of criteria, including publications, research income, research students, and 'environment and esteem.' In each case, although performance is broken down by unit of assessment, reporting is coordinated by the institution. Thus, the university has a need to know what research is being conducted, and develop its own evaluation mechanisms.

Linked to accountability has been a move toward more competition and selectivity. Research evaluation exercises, in the United Kingdom and elsewhere, have sought to ensure that core research funds are increasingly allocated according to criteria of quality and performance. The proportion of research funding allocated from 'core' sources has declined, with that from sources such as government departments, charitable bodies, business and industry, and

international bodies that allocate their funds on a competitive, project-by-project basis showing strong growth in the first half of this decade. In Australia, external research income almost trebled in the decade to 2003, allowing the Australian Vice-Chancellors Committee to point out that for every dollar provided by Commonwealth grants in 2001, universities earned just under AUD 1.40 from other sources. By 2003, this figure had risen to almost AUD 1.60. In Canada, the sponsored research income of the leading 50 research universities rose by between 12% and 24% in each of the five years to 2004. The New Zealand Department of Education reported that over two-thirds of university research income is now competitively won, and that the income from such contracts more than doubled in the six years to 2003. In the United Kingdom, income from competitively won grants and contracts increased by almost 90% in the 8 years to 2004. There is no certainty that this growth will remain, in Canada the *Research Infosource* report for 2008 showed only 'pedestrian' growth, concluding that "the heady days of 10–20% growth have clearly ended, and at the time of writing it remains uncertain how vulnerable university research income will be to recession." However, the increasingly competitive nature of research is likely to remain a permanent feature, requiring the development of new skills for academic teams and those who support them.

Increased accountability and competition have also led to a growth in the transparency of institutional finances. More than ever before, it is possible for institutions to identify which staff are responsible for earning the bulk of their research income. More importantly, this information is available to staff themselves. This has led to a shift in the balance of power between institutions and staff, with increased pressure on institutions to allocate earnings back to those responsible for creating them. Moves toward more decentralized financing systems, in which areas of universities receive a higher proportion of their indirect earnings from research, but are required to be responsible for costs such as space, heating, and lighting, which would previously have been seen as a central charge, have furthered this process.

Three external factors have also encouraged the more active institutional management of research. The first is the increased recognition by governments that universities have a key role to play as a catalyst for wider social and economic development, and that the dissemination of research activity, particularly through knowledge transfer to users in the private sector, has an important role to play in this.

To policy makers, the argument that better utilization of research which has already been financed – much of it through the public funding – could be facilitated at relatively little expense has proved attractive. Universities have embraced these arguments, while stressing that such expenditure should supplement, rather than replace,

existing income sources. A 2006 report from the Australian Vice-Chancellor's Committee argued that expenditure on knowledge transfer should be 2% of total university expenditure. Although such income has expanded significantly, progress has been slower than many anticipated. Annual surveys of North American institutions by the Association of University Technology Managers since 1996 confirm that the majority recover less than 2% of their research expenditure through such income. In an African context Heher (2005) argues that an understanding of the likely returns is critical to policy making, with unrealistic expectations being likely to lead to dysfunctional policy decisions. One implication of this may be that less-formal mechanisms of knowledge transfer, such as consultancy, continuing education, and outreach activity may have more impact than formal licensing and transfer arrangements, although they are less visible and capable of being quantified. Another is that spin-off companies, which may return less directly to the system but have wider benefits in terms of job creation and economic development, might have more long-term impact than licensing arrangements that derive direct revenue to the university.

The second external factor has been the strong growth in collaborative and interdisciplinary research. Collaboration has increased both between and within sectors, and internationally, partly as a result of better communications and the increased need to be competitive, but partly, too, as a result of developments in funding. Examples of nationally based programs include the linkage schemes of the Australian Research Councils, LINK, and Knowledge Transfer Partnerships in the United Kingdom and the opportunities for large-scale consortia under the European Union Framework program, where almost two-thirds of the EUR 50.5 billion budget agreed for Framework Program 7 in December 2006 is listed under the heading of collaborative research. Many national governments have also devised schemes with the specific intention of bringing university and private-sector research teams together. The growing complexity of these relationships has led to the development of a significant body of academic work analyzing the relationships between universities, industry, and government using a 'triple-helix' model (e.g., Etzkowitz, 2002).

All of the above factors have contributed to a significant increase in the complexity of the environment surrounding university research, which can be described as a third external factor. External research contracts are now not only critical to the research base of universities, but to the core business of those funding them. As a result, the relatively friendly and informal relationships that once governed projects have been replaced by a more formal, legal one, in which universities have onerous obligations to deliver. These obligations are supplemented by wider expectations of society – for example, the requirement to observe increasingly complex expectations in the field of research ethics, the need to ensure that laboratories are

kept secure and confidential, and that valuable intellectual property is not compromised by premature disclosure. Each of these areas has required universities – who typically sign research contracts as corporate bodies and have the obligation to deliver on such obligations – to more closely regulate the activities of individual staff and research teams who deliver the research.

An Emerging Profession

Against this background, contemporary research management can be defined as including activities through which an institution seeks to add value to its research activity, but which do not form part of the research process itself. This does not imply that all such activity take place at the center. Despite their rapid growth in recent years, central research management or support units represent only a small fraction of research-related personnel within the typical university. Given the remoteness of central administrators from the research process, there has been renewed emphasis on the training of researchers themselves in basic management issues, such as project management, reporting and the identification of intellectual property. As central research support units get bigger, progressive universities are seeking to decentralize them. In a recent international benchmarking study commissioned by the Higher Education Funding Council for England, Canadian institutions in particular reported a trend to physically locate research administration staff within faculties, both as a way of increasing credibility and of getting closer to the research process, and any emerging problems. Personal skills – such as the ability to forge good relationships – also featured highly among the attributes identified as being important by research managers themselves in an international survey by Stackhouse (2008).

The biggest growth of activity has, however, been at the level of the central institution, where it is no exaggeration to describe the rapid growth of personnel as leading to the establishment of an entirely new research management profession. The extent of this development can be seen in the survey by Stackhouse, in which 82% found that ‘research manager’ was a suitable title for their work, and 71% were full time on research management activities, the largest minority being in Africa, where the majority still combined research management with other, typically academic roles. The new profession combines elements of traditional administrative responsibilities – such as finance, human resources, and registry – with relatively new ones, such as targeted marketing, customer relations, intellectual property management, and licensing. Key elements of the profession are gradually being put in place, covering professional standards, training and organizations. In most developed countries, a clear network of university research managers exists, and in some cases

more than one. Developing countries are beginning to catch up on this process, as evidenced by the formation of such networks in Southern and Western Africa in recent years. While no internationally recognized qualification exists in the area, the Society of Research Administrators International has operated an accreditation scheme for its members for the past 15 years, and the Higher Education Funding Council for England established a project in 2008 to investigate the establishment of a professional institute for research managers (Green and Langley, 2009).

The Functions of Research Management

The precise boundaries of the new profession are still emerging – as can be seen by areas of overlap between this article and others in this encyclopedia. Most commonly, it is regarded as a continuum which follows the research process itself – from identifying a project and source of funding, through presenting proposals to sponsors, recruiting appropriate staff, negotiating contractual terms and budgets with external sponsors to project management and reporting, financial controls, identification and protection of intellectual property, its commercial exploitation, and dissemination to wider society. There remains unevenness in the extent to which these are addressed comprehensively. Stackhouse and Day (2005) found four models of research management provision – ranging from the one-stop-shop – now increasingly prevalent in developed country universities, through multiple, partial, and even no specific provision, the latter being more likely in developing countries.

Where comprehensive provision is available, functions are commonly divided into ‘pre’ and ‘post award’ activities, a distinction that probably originated in the United States and is increasingly reflected in the organization of professional activity at university and national level. It is noticeable that, as the size and complexity of research support offices increases, distinctions of this nature are becoming increasingly prevalent. Technology transfer and licensing offices, for example, are increasingly separated from those that administer research projects, while separate professional organizations tend to exist for different stages of the process. In the United States, this can be seen through the distinction between the Society of Research Administrators (SRA) and Association of University Technology Managers (AUTM), and in the United Kingdom between the Association of Research Managers and Administrators (ARMA) and Association for University Research and Industry Links (AURIL). The distinction is increasingly justified in terms of the different skill sets required, although this argument requires some caution. Communication between ‘pre’ and ‘post’ award functions continues to be important since, for

example, if contracts are not negotiated thoroughly from the outset, the capacity of a university to own and exploit its intellectual property subsequently may be limited.

There is no universally accepted definition of what constitutes pre and post award activity. However, an international benchmarking exercise commissioned by the Higher Education Funding Council for England (HEFCE) in 2004–05 divided pre-award activity into four areas – the development of internal research strategy, retention and support for academic staff, promotion of university research capacity and the submission and authorization of externally funded work. Post award topics were identified as project management and control, the commercial exploitation of research and dissemination of research to wider society. The intensive benchmarking process which resulted involved fifteen institutions from nine countries – Australia, Canada, China, India, Japan, New Zealand, South Africa, United Kingdom, and United States. The participants went on to identify a list of 87 institutional good practice statements (Bjarnason *et al.*, 2006).

On the evidence from these institutions, it seems clear that the relationship between research management offices and the development of internal research strategy is still emerging. In many universities, the research management function is regarded as being exclusively concerned with externally funded research projects, and is located within the administrative structure (although sometimes formally supervised by a ‘Deputy Vice-Chancellor – Research’ of other senior academic). Distribution and management of internal research funds, by contrast, is often seen as a function of the academic committee system. Although there were examples of overlap – such as cases where the Head of the Research Management office was invited to serve on the university research committee, the common separation of the two strands of activity raised questions about how far both strands of funding were being effectively integrated into a common research strategy for the whole institution.

The retention and support of research staff represents a further area where the boundaries of research management are unclear. In many universities, this is regarded as the province of human resources, staff development, or academic departments. Research managers, however, are becoming involved in several ways. An understanding of current staff expertise and aspirations is vital to their role of promoting the university, discussed below. Training is being provided in areas such as research ethics, publication, and intellectual property, where the distinction between research methodology and research management is a fine one. Some research managers are working with their institutions to devise and implement incentives packages to motivate staff to seek external funds, and ensure loyalty from those who are successful. Value judgments are also required from research managers about where to devote their time and resources. One participant in the HEFCE program reported that

his office had made an (unpublicized) decision to focus on no more than 60% of academic staff. The remainder were either unlikely to develop an externally funded research profile, however, much support was offered, or so successful already that the central office was unlikely to add value.

External promotion of university research capacity involves the identification of funding opportunities, matching these to available expertise and supporting staff in their applications. In addition, there is a wider promotional role (sometimes shared with a public relations, communications, or marketing office) which ensures that the university research is generally well known and respected among target groups, and maintains contact with key funding agencies to identify potential trends in policy before these emerge publicly.

The role of most research management offices in this area is one of adding value to the activities of academic staff. It remains the case that the main source of information, development of consortia, and proposal writing is through academic networks. Thus, effective mentoring and support within the home department is likely to be as important a factor in the development of staff as any central mechanisms.

The increased complexity of the research environment, however, has strengthened the need for professional research support. The increasing number of funding bodies allocating research grants on a competitive basis, often combined with a desire to see collaboration across disciplinary boundaries and increasing pressure on teaching and administrative functions, makes the funding environment difficult to track within traditional academic departments. The same also applies to professional support at institutional level. Increasingly, the tracking of funding opportunities is undertaken through ‘buying in’ to commercially operated subscription services, such as the ‘Research-Research,’ and the ‘Community of Science,’ which offer increasingly sophisticated facilities to filter opportunities and target these to the needs of specific academics.

Two other significant trends have been the increased specialization of staff time within the research support office, and an increase in internal peer review mechanisms. Initially, research staff time was more likely to concentrate on specific sources of funding, such as government departments, charitable bodies and the business and industry – itself increasingly segmented into sectors. As research functions have grown further, it has increasingly been possible for individuals to concentrate on specific departments or groups of individuals within the institution. The increased tendency for universities to formally review applications before submission reflects not only the increased complexity of funder needs, but a concern for reputation and quality control. Although such review is still largely academic based, its emergence could mark the start of a further phase in the development of

institutional research management – from a model predominantly of supporting the research activity of individual academics, to one of regulating it.

The need for regulation is clearer in the area of contractual terms and conditions. Given the increasingly commercial nature of contracts, and the obligations and potential liabilities that universities assume in them, it is surprising that relatively few research offices contain qualified lawyers, instead calling on specialist external companies of the wider legal expertise of their universities as required, and also that the amount of litigation to result from university research contracts remains relatively modest. An explanation sometimes advanced for this is that, although complex, the key legal issues in such agreements are repeated from contract to contract. Nonetheless, universities assuming obligations such as the need to observe strict confidentiality need both to pass these on to individual staff at laboratory level, and ensure that appropriate procedures are in place. Those who do not do so remain vulnerable.

The issue of contractual obligations is related to that of costing and pricing research projects. Where universities were willing to undertake research at less than the full economic cost it may have been reasonable to require that sponsors accept less than commercial standards of delivery, intellectual property, or confidentiality. Increasingly, however, universities have sought to recover such costs, for example, through the Transparent Approach to Costing (TRAC) developed by the higher education funding bodies in the United Kingdom since 2001.

As more attention has been focused on post award management, universities have been forced to take a realistic view on the balance between the roles of center, department and individual investigator. As one Asian participant in the HEFCE benchmarking exercise stated:

We have about 3,600 full and associate professors, but only about fifty staff working for research management at college and university level ... about 10,000 research projects are going on every year. It is impossible for administrators to manage the projects directly.

In these circumstances, strategy has concentrated on finding a balance between central monitoring and the training of those responsible for compliance on a day to day level. Ensuring that obligations are brought to the attention of departmental heads and principal investigators is a prerequisite, but unlikely to be sufficient. Other approaches have included more systematic training programs for academics on research management issues, and systems of incentives and penalties that shift the responsibility for noncompliance or ever spending downwards. The area is, however, still evolving and one in which many universities remain vulnerable.

Finally, post-award management requires procedure for the closure of projects and dissemination of results. Emphasis in this area has too often been placed on

commercial exploitation, through the sale or licensing of intellectual property or establishment of spin-out companies. This can be explained by pressure from universities and governments to see tangible returns that can be quantified. As noted above, however, activity of this type is likely to result from a small minority of projects only. In recent years, there has been increased recognition that innovation is a more incremental process than previously imagined, rather than being driven by sudden leaps forward, and that what is transferred between universities and society (in a two-way process) is better classified in terms of broad knowledge than more specific technology.

University research management will increasingly need to reflect this reality. As the profession expands, one might expect to see rigorous procedures for the identification and protection of intellectual property, with others for systematic analysis of research results, customer satisfaction, and wider dissemination and strategies. These in turn will require consideration of the boundaries between research management offices and those for external relations, media, and continuing education, which tend to have some responsibility for such areas at present. The issue of incentives may also need to be addressed; at present academics and their departments are more likely to respond to those measures which are more quantifiable. Often this involves seeking their next research project rather than disseminating the results of the last one.

Conclusion

The exceptional rate of change in research management in recent years has largely been a reaction to the external environment in which university research take place, rather than the result of decisions taken within the sector. Since this environment is still not a stable one, it is reasonable to assume that further change and development will take place over the next decade. The establishment of a discrete research management profession, with its own norms, values, and communication channels, provides a useful extra dimension to the debate over the future of academic research, which universities and policy makers should increasingly value.

See also: Funding of University Research; Higher Education and the Knowledge Society; Research Commercialization; Strategic Planning in Higher Education; Universities' Engagement with Society.

Bibliography

- Bjarnason, S., Kirkland, J., Stackhouse, J., and Day, R. (2006). International research management: Benchmarking programme. Report to HEFCE by the Association of Commonwealth Universities.

- Bristol, HEFCE. http://www.hefce.ac.uk/pubs/rereports/2006/rd11_06 (accessed August 2009).
- Etzkowitz, H. (2002). Networks of innovation: Science, technology and development in the triple helix era. *International Journal of Technology Management and Sustainable Development* 1(1), 7–20.
- Green, J. and Langley, D. (2009). *Professionalizing Research Management, Report to the Higher Education Council for England*. <http://www.professionalresearchmanagement.com/results.html>.
- Hazelkorn, E. (2005). *University Research Management: Developing Research in New Institutions*. Paris: OECD.
- Kirkland, J. (2008). Research management: An emerging profession in the developing world. *Technology Analysis and Strategic Management* 20(6), 717–726.
- Stackhouse, J. and Day, R. (2005). Global and regional practices in university research management: Emerging trends. *International Journal of Technology Management and Sustainable Development* 4(3), 189–205.

Further Reading

- Clark, B. (1998). *Creating Entrepreneurial Universities: Organisational Pathways of Transformation*. (Issues in Higher Education Series, 12). Oxford, Pergamon.
- Clark, B. (2004). *Sustaining Change in Universities: Continuities in Case Studies and Concepts*. Buckingham: Open University Press/SRHE.
- Connell, H. (2004). *University Research Management: Meeting the Institutional Challenge*. Paris: OECD.
- Gibbons, M., Limoges, C., Nowotny, H., et al. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: Sage.
- Juma, C. and Yee-Cheong, L. (2005). *Innovation: Applying Knowledge in Development (UN Millennium Project Task Force on Science, Technology and Innovation Report.)*, UN Millennium Development Library. UN Millennium Project. London: Earthscan.
- Martin, M. (2000). *Managing University-Industry Relations: A Study of Institutional Practices from 12 Different Countries (Improving the Managerial Effectiveness of Higher Education Institutions Series)*. Paris: IIEP (UNESCO).
- Reichert, S. (2006). *Research Strategy Development and Management at European Universities*. Brussels, EUA.
- Vigdor, M., Satter, M. A., Pumwa, J., et al. (2000). *The Management of University-Industry Relations: Five Institutional Case Studies from Africa, Europe, Latin America and the Pacific Region (IIEP) (Improving the Managerial Effectiveness of Higher Education Institutions Series)*. Martin, M. (ed.), Paris: IIEP (UNESCO).

Relevant Websites

- <http://www.arma.ac.uk> – Association of Research Managers and Administrators(United Kingdom).
- <http://www.aurl.org.uk> – Association for University Research and Industry Links(United Kingdom).
- <http://www.researchadmin.org.au> – Australasian Research Management Society.
- <http://www.caura-acaru.ca> – Canadian Association of University Research Administrators.
- <http://www.earma.org> – European Association of University Managers and Administrators.
- <http://www.globalrmn.org> – Global Research Management Network (Hosted by the Association of Commonwealth Universities).
- <http://www.inorms.org> – International Network of Research Management Societies.
- <http://www.ncura.edu> – National Council of University Research Administrators(United States).
- <http://www.srainternational.org> – Society of Research Administrators International(North America).
- <http://www.sarima.co.za> – Southern African Research and Innovation Management Association.

The Training of New Scientists

R Neumann, Macquarie University, Sydney, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Discipline – A defined body of knowledge for learning, research and instruction.

Interdisciplinary – Two or more disciplinary bodies of knowledge interacting in research, learning or instruction.

PhD – Doctor of Philosophy (or DPhil), formal, advanced degree study at university which applies to a wide array of disciplines in the humanities and sciences.

Postdoctoral fellow – Research appointment for someone who has completed a doctoral degree.

S&E – Science and engineering.

S&T – Science and technology (incorporates computer and information science).

STEM – Science, technology, engineering and mathematics.

Transdisciplinary – Transcending disciplinary boundaries in research, learning or instruction.

Introduction: The Role of Science in a Modern Society

Science is fundamental to the fabric of modern societies and their economic prosperity. Scientific knowledge and advancement are acknowledged as crucial to industrial modernity and technological advancement, forming the powerhouse of a modern world.

A nation's ability to generate fundamental knowledge which can enable innovation and transformation into technological applications stems directly from its scientific capacity. This capacity rests in two areas. The first is the level of national scientific literacy. A high level of science literacy within a country's population provides the skills and insights needed to inform discussion, policy, and decisions about the interpretation of scientific discoveries and their appropriate use to advance social and material progress (see e.g., FASTS, 2006). The second area is a nation's scientists. Their research and its outcomes are central to economic competitiveness, underpinning a nation's social well-being and material progress.

The exponential growth of knowledge and consequently, scientific disciplines, together with advanced research techniques and often expensive equipment for experimentation

make science a major national investment. However, science is not confined to national boundaries. It is international in character and language and there is a global market for scientists.

In particular from the mid-twentieth century, the USA has been recognized internationally as dominating scientific research and the science labor market. Its research universities are major attractors of international students to undertake PhDs and postdoctoral training. The European Research Area was created in 2000 by the European Union as a means of harnessing and extending the European research and science capacity to attract scientists and industry investment in European scientific research and hence make Europe more competitive with the USA (EU, 2007). In 2005, European universities awarded 43% of all Organization for Economic Cooperation and Development (OECD) university degrees in science and engineering, compared to 22% for the USA, and 57% of all doctorates (OECD, 2007a). This increasing international competitiveness from Europe challenges the US preeminence in science research and development. Other challenges from science and education investment and development are seen to come from Japan and the Republic of Korea and strongly developing countries such as China and India. Overall, in the decade 1996–2005 most countries have significantly increased their research investment (UNESCO, 2007).

A competitive global market for science raises important questions for individual countries about the recruitment, retention, and career structures of the scientists it trains, as well as forcing it to consider fundamental aspects of the education and training of its scientists.

The Nature of Science and a Scientist's Role

Science is the systematic study of natural phenomena of the living and nonliving world. It is defined as “a branch of study that deals either with a connected body of demonstrated truths or with observed facts systematically classified and more or less comprehended by general laws and which includes reliable methods for the discovery of new truth in its own domain” (Shorter Oxford English Dictionary, 2002).

Within the taxonomies of knowledge fields, science is categorized as a hard knowledge area, reflecting the cumulative, atomistic structure of knowledge. Scientific

knowledge develops in a linear progression and current discoveries can trace their lineage in a hierarchical fashion, not unlike a family tree. Within this process of steady accretion, there is a clarity of the criteria for establishing or refuting claims of new knowledge and boundaries are clearly defined and circumscribed.

The scientific disciplines are numerous and varied, ranging from theoretical, experimental, observational, and technological subject areas. This spectrum covers what have been termed pure and applied fields where the former investigate fundamental phenomena and the latter focus more on the practical application of scientific research to specific human needs, such as the various forms of engineering. Central to science, in particular in the experimental and observational disciplines, is the scientific method. This is a process of investigation through experimentation, observation, and modeling of natural phenomena in a reproducible way which can lead to predictions, theories, and laws.

Thus, scientists are professionals engaged in the conception and creation of new knowledge, products, processes, methods, and systems and are directly involved in the management of research projects (OECD, 2007a).

Originally one of the seven liberal arts (natural philosophy) studied in the medieval university, science in the twenty-first century spans a wide array of ever-increasing disciplines, subdisciplines, and interdisciplinary fields. Scientists are the creators of increasingly specialized scientific knowledge and the complexity of scientific knowledge is reflected in increased diversity of the roles scientists play. They work in a number of different public- and private-sector contexts and organizations, spanning universities, research and scientific institutes, industry, and government. There is a steady trend in most developed Western countries for science graduates to be employed by government and industry with less than half of all science PhDs moving into academic positions (Enders and de Weert, 2004a).

The work of a scientist includes the generation and discovery of new knowledge, the application of new discoveries, and the communication of scientific advancements to a wide variety of audiences. Scientists span a huge array of disciplines across the physical and life sciences, encompassing many subdisciplines and an increasing array of interdisciplinary fields. The discussion here is not on an examination of the range of scientific specialties, rather it is an acknowledgment of the breadth and diversity of scientific fields and an examination of the key issues shared across these disciplines in training new scientists. Thus, the discussion below is inclusive of these vast disciplinary fields. Studies of various aspects of science and national policies also use a number of different categorizations. These can be quite simply science or they may be science and technology (S&T) to specifically indicate the inclusion of computer and information science, science and engineering (S&E), and science, technology, engineering, and mathematics

(STEM). The discussion here is inclusive of all these categorizations.

The Training of New Scientists

To become a scientist is to embark on an increasingly specialized course of study through school and university, culminating in the most advanced form of scientific research training and apprenticeship, a PhD. Following this the scientists' career generally embarks on a period as a journeyman through postdoctoral experience, through independent researcher and leader.

Early Science Education and Training

Given the sequential, cumulative nature of scientific knowledge the training of a scientist arguably commences during the school years, continuing into initial degree study before a person is sufficiently grounded to embark on specialized advanced study and research. Consequently, the most important themes in the literature are the attraction of sufficient and the best students into science at all education stages, the quality and relevance of science curricula, and, the supply of well-trained science teachers.

An emerging benchmark is the outcomes of education systems through student achievement in various academic literacies undertaken by the OECD through its Programme for International Student Assessment (PISA) studies. The 2006 PISA study examined scientific literacy by assessing student knowledge, skills, and attitudes at the end of their compulsory schooling years in OECD member and nonmember countries. It also examined system characteristics as well as school and human resources impacting on science education. This represented 400 000 students in 57 countries representing 90% of the world economy (OECD, 2007b).

National performance in scientific literacy varies greatly across the countries surveyed. Particularly important for national capacity in science, only 37% of students stated that they would like a career involving science with only 21% interested in doing advanced science. Reinforcing the findings of research studies in many countries about the lack of sufficient qualified science teachers, 62% of schools had vacant science teaching positions. In these schools with science teaching vacancies, 65% of principals reported that instruction was hampered by a lack of qualified science teachers.

The literature on the postsecondary study of science continues the themes of student numbers, student quality, and the quality of the science experience and curriculum. Having sufficient numbers of good quality undergraduate science students is crucial to the pipeline effect of feeding advanced research-degree science training. Various studies of student subject selection indicate that students find

science difficult and believe that they will be more academically successful in nonscience disciplines (Convert and Gugenheim, 2005). Consequently, there are national concerns about the numbers of students entering and graduating with science degrees. In 2005, the OECD averaged 7.4% of graduates in science, 5.4% in mathematics and computer science, and 12.2% in engineering (OECD, 2007c).

There are discussions about the structure and relevance of university science curricula for the early twenty-first century given the growth and complexity of scientific knowledge and the increasing need to have science graduates work in transdisciplinary and interdisciplinary contexts (Bill *et al.*, 2001). Science faculties at undergraduate level are developing education and research opportunities for science students to gain experience in research projects, interdisciplinary research, and also industry-supported research. It has been found that undergraduate student participation in research results in greater likelihood of pursuit of PhD study and a career in science (Hancock *et al.*, 2007; Hunter *et al.*, 2007). Overall, the aims of undergraduate science curricula are to provide a rich science experience which includes real research involvement and enables interaction with researchers in a variety of science disciplines as well as with postdoctoral fellows and graduate students.

The Role and Nature of the PhD

The PhD is the essential training ground for future scientists, followed by a period as a postdoctoral fellow. It represents the highest level of formal education and is a degree involving independent and original research.

In most countries, initial training is essentially within universities or similar higher education institutions commencing from bachelor degree to PhD levels of training. Depending on the national system, the bachelor's degree may include an additional honors year in which research training commences, followed by a PhD, or it may progress through bachelor to master's degree and then PhD. However, irrespective of national education system, for many decades, the PhD has been the accepted and expected route of training for a scientist in all developed countries. Attainment of a PhD is then often followed by postdoctoral employment prior to entry into an academic or scientific career in the public or private sectors.

The PhD has developed from its origins in the Humboldtian university of nineteenth-century Germany as an apprenticeship model of gaining research independence and certification as a master researcher or master scientist. Within this framework the trainee researcher, or emerging scientist, learns under guidance the tools of the trade, establishing credibility among peers as being able to identify significant questions arising from the discipline based on the cumulative advancement of the specialization.

It is considered the essential training in the culture, language, values, and methods of the discipline in particular and of science in general. By the end of the PhD, the apprentice or trainee has become an independent researcher and scientist who is a specialist in a particular field of knowledge. The research is predominantly discipline-focused, and pure research, aimed at advancement of the discipline, although increasingly applied, interdisciplinary and transdisciplinary research projects are also undertaken.

While the specific requirements for a PhD reflect national variations, common to all is the requirement of original academic research submitted as a thesis or dissertation which is examined by a group of experts. The examination process, again reflecting national variations, may be either as written expert assessment of the dissertation or it may involve an oral defense of the research by the candidate before a panel of expert examiners.

Spanning international interests the current literature on the PhD and training of new scientists can be summarized into five main themes: (1) the appropriate duration of the PhD; (2) achieving a balance between the in-depth specialization required by the research and social and business demands for breadth; (3) tensions between the traditional pure research approach to the PhD and the development of new, diversified modes of knowledge production; (4) the development of varied organizational structures and arrangements for PhD research compared with the disciplinary or department-based programs; and, (5) reinvigorating the notion of a steward of science.

Duration of the PhD

There is much national variation in the duration of PhD programs. However, on average a PhD requires 3–5 years of full-time research, but durations of 4–8 years may be more typical in some countries, such as in the USA, where the US Department of Education records 7 years as the median time to obtain the degree (USNEI, 2007). In recent years, governments in most English-speaking nations and member countries of the European Union have been examining their PhD structures with a view to realizing comparability of quality and standards cross nationally as well as promoting mobility for newly trained science PhDs. Underlying these reviews are also state concerns with efficiency, relevance, and the strengthening of national economic and scientific competitiveness. A key consideration in PhD duration is, however, the degree of research preparation that science PhDs have prior to entry. In systems where there is practice of a research honors year following undergraduate study or a research masters, then PhD completion within a 3–4 year time span may be more achievable than where such prior preparation is not customary. However, where such prior research preparation does not exist, then science PhD programs need to incorporate research method and

skill-development courses, hence extending the duration of the doctorate and the nature of the science training.

Balancing specialization and breadth

The second issue is the balancing of advanced specialization with breadth of knowledge and skills required of scientists by industry and society. The literature reflects the views of scientists, policymakers, and researchers that a broader, generic skill set to accompany research specialization is required (Balatti *et al.*, 2004; DEST, 2005; Enders and de Weert, 2004; Getzoff *et al.*, 2003; Teitelbaum, 2006). This has seen the development of taught courses within the PhD, which many attribute to the American model of the PhD. The purposes of such courses is to provide the broadening experience through learning about new knowledge developed, as well as in some countries, to provide additional training in scientific-research methods. The requirement of original scientific research and the development of an independent research scientist remain the central purposes of the PhD. However, such courses are seen to complement and enhance the PhD training through fostering breadth of knowledge and developing general skills and flexibility alongside the specialization necessitated by the conduct of original research. External pressures to reconceptualize PhD training from the traditional preparation for an academic career reflect that in many countries, significant numbers of science PhDs in the past two decades do not embark on academic careers, instead gaining employment in government and private organizations. Given this trend, additional courses need to cover interpersonal and professional skills training as well as career-development courses. Thus PhD programs in general, as well as in science more specifically, include the three key components of mastery of a specialization within the discipline or subdiscipline, broad knowledge of associated fields, and the development of professional skills.

Diversified modes of knowledge production

The third theme is the impact of new, diversified modes of knowledge production on doctoral training. The traditional view has been that a science PhD addresses problems arising from the discipline of a pure research nature. However, the distinctions between pure and applied research have been steadily blurring and the increased interest and involvement of industry in scientific research heighten the focus on problem solving arising in the context of application. Further, alongside the disciplinary imperative, and in part resulting from it, there is the emergence of interdisciplinary and transdisciplinary research problems. Such problems draw on knowledge and skills across a range of disciplines, subdisciplines, and specializations. This development can be seen as a fundamental shift within science and research more generally in the mode of knowledge production from what have been described as inward looking, disciplinary paradigms to more outward-looking

approaches, in which education and research are shaped in a continuous interaction with professions and business outside the university environment. The most influential model to capture this change is that of mode 1 and mode 2 forms of knowledge production (Gibbons *et al.*, 1994, 2001). The terms highlight the shift in emphasis between pure and applied research and research questions generated from within the discipline to ones shaped by social and economic influences and activity. Alongside pure research problems generated from the discipline are added interdisciplinary and transdisciplinary research oriented toward social and economic imperatives rather than disciplinary ones. The temptation is to see mode 1 and mode 2 as a dichotomy; more realistic, however, is to acknowledge that there is a spectrum of modes of knowledge production. There are implications for both the nature and type of research questions undertaken for the PhD as well as the range of inquiry methods and the location or site of production of the PhD (see theme 4 below).

Varied organizational structures and arrangements for PhD research

Connected with the third theme of the diversification of the mode of knowledge production are the more varied organizational structures and arrangements for PhD research (Enders and de Weert, 2004). There are two key developments in this regard. The first is the growth of flatter university structures to accommodate interdisciplinary research problems. Such structures may reside within existing faculties or departments in the form of centers or institutes and are normally of fixed duration to enable adaptation to the constantly changing disciplinary and interdisciplinary knowledge. The second development is that such structures and arrangements involve the inclusion of research partners external to the university or research institute through the establishment of science research programs in centers and consortia combining university and industry resources (see e.g., Harman, 2004). This may involve joint funding of scientific research often including PhD scholarships. The involvement of external partners enables an experience with real-world problems experienced in industry with access to testing and application in real contexts. Potentially, the innovation and development of new products in competitive international science markets are strengthened. Hence, increasingly there may be joint arrangements between universities and the private sector in the training of science PhDs, leading to an increased blurring of the boundaries between public and private sector. Such arrangements are often encouraged by national science and research policies where governments in many countries over the past two decades have promoted closer involvement and interaction between universities and the private sector. Japan, however, may represent a counterexample where its tradition of industrially trained PhDs have been found to be poor at contributing to research

in new and emerging fields. Here, the effort is to encourage future innovation through the promotion of university science PhD training with a focus on basic research (NSF, 2000).

Stewardship: The responsibilities of a scientist

A consistent theme in the writings of scientific professional associations is professional ethics or professional code of conduct. While the issues of professionalism, ethics, and codes of conduct are perpetual, the more recent discussions are framed in terms of social and disciplinary expectations of a scientist. Most recently, this notion of professional responsibility and ethics has been termed stewardship in the Carnegie studies of the doctorate (Golde *et al.*, 2006). Their investigations re-affirm the important role of PhD graduates in relation to the ethical and moral aspects of their work as leaders in their discipline. The discussions emphasize the role of the PhD in identity building and transformation from student to professional. This transformation involves an appreciation of a larger sense of purpose than just the discipline and individual careers to encompass a set of roles and skills, as well as a set of principles. Thus, to graduate from a PhD as a new scientist, there is not just an obligation to discover and generate new knowledge but also to transmit, critique, and use knowledge in moral, ethical, and responsible ways. Hence, fundamental in the training of new scientists is the development of a set of qualities and values that ensure “the continuing health of the discipline and how to preserve the best of the past for those who will follow. Stewards are concerned with how to foster renewal and creativity. Perhaps most important, a steward considers how to prepare and initiate the next generation of stewards” (Golde *et al.*, 2006: 13).

Following the PhD there are prospects for employment as a scientist within government agencies, industry, or as independent researchers in a university or research institute. For employment in the latter, further training at postdoctoral level is considered a prerequisite as an independent scientist.

Gaining Employment as a Scientist: Postdoctoral Training and Career Paths

The postdoctoral stage forms the critical final step in the training of new scientists prior to appointment as a scientist in a university or research institute. It represents a period of independent but loosely supervised research training, post the PhD. There is no formal qualification since during postdoctoral training these new scientists are no longer students but employees. The postdoctoral position is dedicated to research rather than teaching and may be funded by the university or research institute in which it is offered, or, funded independently through external

research-granting agencies. The position has multiple aims: to diversify technical skills; to obtain a deeper understanding of the cutting edge of the discipline; and to enable continued development of research networks.

The laboratory where the postdoctoral fellow resides provides mentorship to assist in the move to a long-term position as an independent scientist, since to progress from a PhD directly to a tenured academic position would be the exception rather than the norm. Over recent decades, a period as a postdoctoral fellow has become the accepted route to an independent scientific career and the appointee is expected to publish the results of his/her own research during the appointment period. Postdoctoral appointments for periods of 3–5 years are common before appointment as an independent scientist. Mobility is an important factor in the postdoctoral stage, since a postdoctoral position is likely to be in a different institution from the PhD stage and even in a different country.

Since the 1990s, there have been numerous surveys of holders of postdoctoral positions with recurring findings of dissatisfaction. A major issue has been that new scientists have held numerous sequential postdoctoral positions without gaining employment as an independent scientist in a permanent position. Postdoctoral appointments have become holding bays rather than stepping stones to continuing appointments. Consequently, the literature documents increasing dissatisfaction with the lack of career paths, employment conditions, and salaries attached to postdoctoral positions together with the need for ongoing career and professional development (Singer, 2004; Association of American Universities Committee on Postdoctoral Education, 1998).

Within the scientific community and for national governments, the underlying concern rests with the long-term attractiveness and capacity of the scientific labor market. The evidence from many countries is that there is an unacceptably high rate of attrition in early scientific careers (Preston, 2004). Of particular concern is the emerging evidence of the aging scientific workforce and the increasing age at which new scientists obtain their first independent research grant (Bravo, 2006).

The situation is further complicated by differing reports on supply and demand in the science labor market. There are contradictory reports of either a surplus or a shortage of scientists with discussions often reflecting snapshots of a particular period of time, selected science fields, and specific countries. Recurring themes, however, are the unattractiveness of scientific careers in terms of remuneration and advancement, insufficient academic positions within universities, and the personal costs associated with the continued need for mobility in establishing a scientific career. Of particular national concern should be that the continued growth in science PhD graduates and postdoctoral fellows since the late 1990s in Western, developed countries such as the USA, the UK,

Canada, and Australia, has resulted from an expansion of international doctoral growth rather than domestic growth (Moguerou, 2005). Science doctorate holders from Europe and South East Asia are most likely to move to, or remain in, the USA (Auriol, 2007; Cerny and Nerad, 2002). The long-term concern is that once these international graduates and postdoctoral students return to their countries, the host nation will be left with a shortage of qualified scientists within a competitive international market. As more nations establish themselves as education hubs and with the continued scientific development in countries such as India and China, global competition for new scientists will become much fiercer.

See also: Human Resource Issues in Higher Education; Mobility of PhD Students and Scientists; National Science Policy and Universities; Productivity of University Faculty Staff.

Bibliography

- Association of American Universities Committee on Postdoctoral Education (1998). Report and Recommendations. <http://www.aau.edu/reports/PostDocRpt.html>. (accessed June 2009).
- Auriol, L. (2007). Labour market characteristics and international mobility of doctorate holders: Results from seven countries. *OECD Working Paper*.
- Balatti, J., Gilbert, R., Turner, P., and Whitehouse, H. (2004). The generic skills debate in research higher degrees. *Higher Education Research and Development* **23**, 375–388.
- Bill, A., Grossenbacher-Mansuy, W., Haberli, R., et al. (eds.) (2001). *Transdisciplinarity: Joint Problem Solving among Science, Technology and Society*. Basel: Birkhauser.
- Bravo, N. R. (2006). Fostering new investigators in a community effort. *NIH Extramural Nexus* **September**, 1–3.
- Cerny, J. and Nerad, M. (2002). Postdoctoral appointments and employment patterns of science and engineering doctoral recipients ten plus years after Ph.D completion: Selected results from the 'Ph.Ds – ten years later' study. *Council of Graduate Schools Communicator* **XXXV**, 1–2; 10–11.
- Convert, B. and Gugenheim, F. (2005). Scientific vocations in crisis in France: Explanatory social development and mechanisms. *European Journal of Education* **40**, 417–431.
- DEST Department of Education, Science and Training (2005). *Australian Government Audit of Science, Engineering and Technology Skills*. Canberra: Commonwealth of Australia.
- Enders, J. and de Weert, E. (2004a). Science, Training and Career – Editorial. *Higher Education* **17**, 129–133.
- Enders, J. and de Weert, E. (2004b). Science, training and career: Changing modes of knowledge production and labour markets. *Higher Education Policy* **17**, 135–152.
- European Commission (EU) (2007). *The European Research Area: New perspectives. Green paper 04.04.2007*. Luxembourg: Office for Official Publications of the European Communities.
- FASTS (Federation of Australian Scientific and Technological Societies) (2006). *A 21st Century Role for Australian Science and Innovation*. Canberra: FASTS.
- Getzoff, E. D., Gordon, J. I., Kron, S. J., et al. (2003). Educating future scientists. *Science* **301**, 1485.
- Gibbons, M., Limoges, C., Nowotny, H., et al. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: Sage.
- Gibbons, M., Nowotny, H., and Scott, P. (2001). *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge: Polity Press.
- Golde, C., Walker, G., and Associates (2006). *Envisioning the Future of Doctoral Education: Preparing Stewards of the Discipline*. Carnegie Essays on the Doctorate. San Francisco, CA: Jossey-Bass.
- Hancock, M., McCullough, J., and Russell, S. (2007). Benefits of undergraduate research experiences. *Science* **316**, 548–549.
- Harman, K. (2004). Producing 'industry-ready' doctorates: Australian cooperative research centre approaches to doctoral education. *Studies in Continuing Education* **26**, 387–403.
- Hunter, A., Laursen, S. L., and Seymour, E. (2007). Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. *Science Education* **91**, 36–74.
- Moguerou, P. (2005). Doctoral and postdoctoral education in science and engineering: Europe in the international competition. *European Journal of Education* **40**, 367–392.
- NSF (National Science Foundation) (2000). Graduate education reforms in Europe, Asia, and the Americas and international mobility of scientists and engineers. *Proceedings of an NSF Workshop*, Arlington, VA.
- OECD (Organization for Economic Co-operation and Development) (2007a). *OECD Science, Technology and Industry Scoreboard 2007*. Paris: OECD.
- OECD (Organization for Economic Co-operation and Development) (2007b). *Programme for International Student Assessment, PISA 2006: Science Competencies for Tomorrow's World*. Paris: OECD.
- OECD (Organization for Economic Co-operation and Development) (2007c). *Education at a Glance 2007 OECD Indicators*. Paris: OECD.
- Preston, A. (2004). Plugging the leaks in the scientific workforce. *Issues in Science and Technology* **20**, 69–74.
- Singer, M. (2004). The evolution of postdocs. *Science* **306**, 232.
- Teitelbaum, M. (2006). A new science degree to meet industry needs. *Issues in Science and Technology* **20**, 69–74.
- UNESCO (United Nations Educational, Scientific and Cultural Organization) (2007). A global perspective on research and development. UNESCO. Institute for Statistics.
- USNEI (US Department of Education) (2007). Structure of US Education – Graduate and Post Education Levels. <http://www.ed.gov>.

Further Reading

- European Institute of Education and Social Policy (2005). *European Journal of Education* **40**, 4. Special edition on scientific studies in Europe and the European Research Area, December 2005.
- Trumble, W. and Brown, L. (2002). *Shorter Oxford English Dictionary*, 5th edn. Oxford: Oxford University Press.

Relevant Websites

- <http://www.aau.edu> – Association of American Universities.
- <http://www.ed.gov> – US Department of Education.
- <http://www.nationalpostdoc.org/> – National Postdoctoral Association (U.S.).
- <http://www.nsf.gov/> – National Science Foundation (U.S.).
- <http://www.nsfgrfp.org/> – National Science Foundation Graduate Research Fellowships.

HIGHER EDUCATION – SOCIETY

Contents

Academic Freedom

Access and Equity in Higher Education

Boundaries of Institutional Autonomy and their Impact on Higher Education

Diversity of Higher Education

Employability of University Graduates and Graduate Outcomes

Higher Education and the Knowledge Society

Higher Education and the Labor Market

Higher Education and the Transformation of Society

Quality Assurance in Higher Education – Practices and Issues

The Role of Accreditation of Higher Education Institutions

Universities and Regional Development

Universities' Engagement with Society

Academic Freedom

S F Macintyre, The University of Melbourne, Parkville, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Academic freedom is a freedom distinctive to academic institutions and those who work in them: it safeguards their freedom to undertake academic activities and fulfill the academic mission. This freedom derives from the academic commitment to the pursuit of knowledge by means of open inquiry. It is not an absolute freedom, nor is there unanimity on its interpretation and application, which is subject to the changing circumstances of academic life. Proponents also appeal to broader human rights such as freedom of thought, conscience, religion, expression, assembly and association; however, academic freedom has a more restricted meaning – it enshrines such rights only as they apply to the performance of academic duties.

A statement from the first Global Colloquium of University Presidents in 2005 sets out the core principle: “At its simplest, academic freedom may be defined as the freedom to conduct research, teach, speak and publish, subject to the norms and standards of scholarly inquiry, without interference or penalty, wherever the search for truth and understanding may lead.” Such declarations typically rely on two complementary premises. First, that the pursuit of knowledge is dependent on freedom of inquiry and the university accordingly needs to be protected from political, religious, or other forms of interference. Second, that the university is a core institution of

civil society and best performs that role if it is able to conduct its own affairs.

Academic freedom is codified and promulgated in various statements by universities, academic associations, governments, and international organizations. The Lima Declaration of the World University Service in 1988 defined it as the “freedom of members of the academic community, individually or collectively, in the pursuit, development and transmission of knowledge, through research, study, discussion, documentation, production, creation, teaching, lecturing and writing.” In 1997, the *United Nations Educational, Scientific, and Cultural Organization* (UNESCO) adopted a recommendation concerning the status of higher education personnel, and Article 27 defines academic freedom as “the right, without constriction by prescribed doctrine, to freedom of teaching and discussion, freedom in carrying out research and disseminating and publishing the results thereof, freedom to express freely their opinion about the institution or system in which they work, freedom from institutional censorship and freedom to participate in professional or representative bodies.”

The conventional understanding of academic freedom is one of freedom from coercion, principally outside coercion from the state or other agencies, though extending to improper interference by university authorities in the performance of academic duties. There is also an element of positive freedom, especially in aspects of activity such as research that depend upon institutional support.

History

Academic freedom had its historical origins in the universities of medieval Europe. The freedom of the academic institution was established first. It arose as these early universities sought greater autonomy from church and state, and advanced as conformity gave way to religious difference. From the status of the university as a self-governing corporation came the principle of institutional autonomy, which carried over to the national systems of higher education that were established in the nineteenth and twentieth centuries; from toleration and religious liberty came the idea of an independent intellectual community; from doctrines of political and economic liberalism came the principle of freedom of thought and the metaphor of free competition among ideas; and from a new emphasis on education as a process of discovery, articulated most influentially by Wilhelm von Humboldt in early nineteenth-century Germany, came the idea that teaching and research were interdependent activities joined in an untrammelled pursuit of truth.

The modern understanding of academic freedom was shaped decisively by the rise of science. The branches of scientific research that were established in universities in the nineteenth century were based on the understanding that knowledge was produced by a continuous process of inquiry, that hypotheses had to be verified, and that verification must follow certain rules in a procedure that required a trained expertise. From these principles came the custom of peer judgment guided by scientific criteria of honesty (a commitment to truth), neutrality (the subordination of beliefs in the search for truth), and testifiability (the publication of findings so that their truth could be tested).

Here already, the meaning of academic freedom was extending beyond the institutional freedom of the university to the intellectual freedom of those who worked within it. It was the application of the methods of natural science to social inquiry that led to a further extension: the formation of the academic disciplines of economics, politics, sociology, and history, and the conviction of practitioners that their investigations could and should improve institutions and social practices, created a new dimension of academic freedom – freedom of commentary.

This idea of professional privilege was associated with the special status of the university as a self-regulating intellectual community. It found expression in the German concepts of *Lehrfreiheit*, denoting the distinctive prerogatives of the academic profession, and *Lernfreiheit*, the freedom of students to determine their studies, attendance at classes, and living arrangements. This latter aspect of academic freedom, the entitlements of students, has since been extended significantly to encompass principles of equal opportunity and affirmative action. Thus,

the International Covenant of the United Nations on Economic, Social and Cultural Rights commits signatory nations both to the provision of higher education (“in particular by the progressive introduction of free education”) and to the elimination of “discrimination of any kind as to race, color, sex, language, religion, political or other opinion, nation or social origin, property birth or other status.”

Institutional Autonomy

Arrangements for safeguarding university autonomy vary widely. In continental Europe, where the modern university developed as a state institution, it has been codified in legislation. In Latin America, many universities include the word *autonoma* in their titles. In the United States, Britain, and its dominions, where civic universities supplemented the original religious foundations, autonomy was inherent in the independent status and self-governing character of the institution. None of these arrangements were secure. State systems of higher education were vulnerable to government interference (and even the legislative safeguards were omitted from many of the university systems created in the new nation-states of Africa and Asia after World War II), while civic universities were subject to intervention from lay members of their governing bodies.

As the result of changes in higher education over the past three decades, such formal differences have narrowed and institutional arrangements have tended to converge. The rapid expansion of higher education and increased public investment in teaching and research brought corresponding expectations of accountability. More than this, universities became a key determinant of competitive advantage in the globalized information economy. They were subjected to institutional devolution as systems of market competition were introduced to maximize their efficiency and effectiveness, and far more intrusive interference as governments redirected their procedures and priorities. These developments led one commentator to question the received wisdom of institutional autonomy: “far from being naturally ‘outside’ government,” Simon Marginson wrote, “the modern university is a product of government.”

Nevertheless, it remains a distinctive institution with a large measure of formal independence and practical control over its activities. The institutional autonomy of the university typically includes determination of the appointment of staff and selection of students; the establishment of courses and award of degrees; and the initiation of research and dissemination of knowledge. However many universities lack at least some of these powers: some governments determine policies for selecting students, some continue to control the appointment of

staff, and some intervene directly in the choice of the senior executive officer. Other restrictions on institutional autonomy are discussed below.

Freedom of Academics

The freedom of academics is distinct from institutional autonomy, for it is possible for a university to have a large measure of autonomy but restrict the freedom of its members. This freedom also found formal expression in many European countries, while it remained implicit in the Anglophone universities until associations of academics attempted to secure recognition of their entitlements.

The impetus in the United States was the dismissal of some prominent social scientists for public statements critical of big business. The formation of the American Association of University Professors (AAUP) in 1915 allowed the publication of a Declaration of Principles by its Committee on Academic Freedom and Tenure, asserting freedom of research, teaching, and extramural utterance and action. It argued that academic freedom was an indispensable attribute of a university if it was to advance knowledge, succeed in its educational mission, and contribute to social progress. More than this, the authors asserted that while a university teacher accepted a responsibility to the university authority, “in the essentials of his professional activity his duty is to the wider public to which the institution itself is morally amenable.”

Although the college presidents who formed the Association of American Colleges in the same year rejected this Declaration of Principles, their own Academic Freedom Commission produced its own report in 1922. After a series of joint conferences, the two parties finally agreed in 1940 upon the Statement of Principles on Academic Freedom and Tenure. The statement accepted that “institutions of higher education are conducted for the common good” and “the common good depends upon the free search for truth and its free exposition.” Each of the three freedoms advanced by the AAUP was affirmed, and each with its corresponding responsibilities. Teachers were entitled to full freedom in research, subject to adequate performance in other duties and a suitable arrangement concerning financial benefits. They were entitled to freedom in the classroom, but should be careful not to introduce into their teaching controversial matter which has no relation to their subject. They were also entitled to speak or write as citizens free of institutional censorship or discipline, so long as they recognized that their special position in the community imposed special obligations of accuracy, restraint, and respect for the opinions of others.

The Statement, which has since been amended and reaffirmed, is endorsed by more than 140 learned bodies, and frequently cited by courts, distinguishes the approach to academic freedom followed in the United States from other countries. In Britain, Australia, Canada, New Zealand, and

South Africa academic staff associations concerned themselves with academic freedom, and in some cases secured agreements with university authorities, but nowhere with the same level of detail or measure of acceptance. Where the American approach used voluntary associations to specify the entitlements of academics, these other countries relied far more on customary practice – though they also have resorted to the courts and in some cases moved toward the legislative approach of other countries.

Historically, the freedom of academics was circumscribed by the religious nature of the university. David Hume’s atheism precluded him from appointment to Edinburgh University in the eighteenth century, and well into the nineteenth century; only men in holy orders of the Church of England were eligible for fellowships of Oxford and Cambridge. Heresy and immorality continued to be grounds for dismissal in Britain and elsewhere into the twentieth century, but by then the most common reasons for refusing or terminating an academic appointment were secular. The cases in the United States that stimulated the formation of the AAUP in 1915 arose from the expression of unpopular views on controversial issues, and similar cases of victimization occurred in many other countries.

Beyond these restrictions on freedom of public commentary, academic freedom was curtailed by increased demands of loyalty to the state. Aliens were removed from many universities in World War I; the Soviet Union purged its universities in the 1920s, and in 1933, German universities were required to dismiss all persons of Jewish ancestry or socialist belief. During the Cold War, a number of American universities imposed loyalty tests, and many dismissals followed. Other Western countries followed similar practices, and the German *Berufsverbot* (a policy of banning individuals whose loyalty to the constitution of the German Federal Republic was in question) remained in the 1970s, while communist regimes in Eastern Europe, China, and elsewhere applied the Soviet insistence on political conformity to their universities. Surveys of academic freedom in many countries reveal the persistence of such expectations and practices.

The protection of academic freedom in teaching, research, and public commentary is based on appropriate procedures for appointment, tenure, promotion, and reward. It is expected that appointments will be made on merit by due process that allows all qualified applicants to be considered; similarly, confirmation of tenure, promotion, and other decisions affecting academics are expected to involve peer assessment and be based on performance of duties. The AAUP’s 1940 Statement of Principles places particular emphasis on tenure as a protection against victimization.

Such academic freedom is not absolute. As Edward Shils observed, it does not extend to charlatans, plagiarists, or cranks, but is reserved for academics to enable

them to discharge their academic duties. Hence, the introduction of political or other convictions inside the classroom should be consistent with the nature of the class (it would be appropriate to express a judgment on global warming in a geography lecture but not one on literature) and should allow for alternative judgments. It allows academics to determine their research projects and decide how they are to be conducted, provided they are consistent with the disciplinary expertise of the researcher and conducted honestly. It accepts the legitimacy of public commentary as part of the free exchange that is integral to the university's civic role, with the expectation that they will uphold the standards of their profession and the reputation of their institution. Recent controversies over this aspect of academic freedom are considered below.

Freedom of Students

The freedom of students embodied in the nineteenth-century German understanding of *Lernfreiheit* was a freedom to learn. It marked out the privileged status of the student in a hierarchical society, denoting an independence from parental supervision and an absence of administrative coercion in the pursuit of learning. Once qualified to undertake university studies, the student was entitled to move from one university to another, choosing which classes to attend and when to sit examinations.

Both then and later, this understanding of student freedom did not have the same force in other countries. The undergraduate enjoyed a privileged status in older universities but it derived from the exclusiveness of the institution rather than a prerogative to pursue learning as he chose. If their supervision of studies was light, the collegiate foundations imposed residential requirements and exercised a pastoral responsibility for their charges. The growth of civic universities in the twentieth century brought increasingly detailed programs of study, tuition, and assessment, and continuing restrictions on the extracurricular activities of students: representative student associations were allowed a limited autonomy but controls were imposed on their publications and political activities.

These constraints were challenged in the 1960s as part of the upsurge in student radicalism that wracked campuses in Europe, America, and beyond. The principal concern of student radicals was to contest government policies, especially the Vietnam War, and drew on the idea of the university as a distinctive institution of civil society with a particular capacity and entitlement to challenge orthodoxies. This claim extended to the expanded role of the university and its complicity in military and industrial activity, and soon involved demands for greater student control of the university – although these demands called into question the university's autonomy as an independent and politically neutral institution.

From the turmoil on the campuses of the United States came a joint statement on rights and freedoms of students in 1967, which spelt out the freedom of association, expression, publication, and participation in institutional government. The joint statement also laid down procedural standards for disciplinary proceedings against students. It also recognized that students "should be free to take reasoned objection to the data or view offered in any course of study," an aspect of student freedom that has since become a particular source of contention.

Student protests in the late 1960s and early 1970s played a major role in Europe, Latin America, and several Asian countries. Especially where civil rights were weak and freedom of speech curtailed, the university provided opportunities for discussion, criticism, and political opposition. Some of the restrictions on students' freedom of association derive from this fact. The destruction of the World Trade Center on 11 September 2001 was followed by new restrictions on student activities regarded as linked to terrorism in the United States, the United Kingdom, and other countries.

Restrictions on Academic Freedom

Academic freedom was never absolute and remains contested. The freedom it demarcates is restricted to particular pursuits, and the meanings attached to it have emerged through a constant process of contestation. As with other freedoms, it establishes protection from interference but in doing so imposes restrictions. Moreover, since academic freedom applies to different aspects of academic activity, the protections and restrictions operate within the university: rights claimed by academics have implications both for institutional autonomy and student freedom. The present arguments over restrictions on academic freedom arise partly from these differences and partly from disagreements over the purpose of the university as it has expanded in size, function, and importance.

Institutional Autonomy

The older university was a training institution for a small number of learned professions and a custodian of higher learning. The university sector now undertakes the education and vocational preparation of the majority of the population of industrialized countries, and is itself a major industry, while the expansion of higher education in developing countries has been even more rapid. The university is the principal site of knowledge production in the information economy, and the source of expertise that enters into every corner of social life. The growth of universities, and the increased competition among them as well as with alternative providers, exceeds the capacity

or willingness of governments to fund them, so that they are increasingly dependent on alternative sources of income.

One consequence of a mass system of higher education is a much greater variety of participants. The inclusion of women, ethnic and religious minorities, and other groups brought new procedures for equal and affirmative action, new curricular demands, and new norms of behavior. Speech codes, in particular, brought allegations that universities were imposing political correctness at the expense of freedom of expression. These arguments are particularly marked in the United States but have arisen in other countries where multiculturalism and identity politics operate.

A less discussed but more significant effect of the present circumstances of the university has been on its management and procedures. As the higher education sector has become more dependent on nonpublic sources of income, it has commercialized much of its teaching and research, and adopted new methods of management. University administration has expanded in size and complexity, employing business models to plan, finance, manage, and measure academic activity. Business partnerships and contracts often impose restrictions on intellectual property that inhibit the free exchange of information. The enterprise university pays particular attention to its corporate image, and seeks to restrict adverse publicity. Universities have never been immune from laws of supply and demand, and one justification for academic freedom was to insulate intellectual judgments from market forces; this separation has become less clear.

These restrictions of academic freedom are mostly self-imposed, but can be traced to a diminution of institutional autonomy. While governments have reduced their contribution to operating budgets, they have intervened to make universities serve national objectives. The governing bodies of formally self-governing universities have been required to conform to the model of the company board; changes in management and staff policies have in some cases been required as a condition of government funding; specific-purpose funding is used in place of general grants, and competitive performance is used to determine allocations; and public research agencies commonly identify areas of commercial opportunity or national interest.

Finally, universities are now expected to meet much greater expectations of public accountability. If they were once neglected, they are now of direct interest to a mass clientele and a constant source of news and media commentary. As they became central to the operations of government, economy, and society, many of their older immunities have been withdrawn. The university is a public institution that depends upon public support and is expected to serve its public; but in order to do so it requires respect for its distinctive nature. The shifting

boundaries of institutional autonomy are formed by this interrelationship of dependence and independence.

Academic Freedom

Many of the recent developments summarized above have implications for the freedom of academics. The establishment of academic freedom was linked to the growing self-confidence and self-esteem of the academic profession, and to the collegial forms of academic life as a vocation. The growth of the higher education industry, the emergence of alternative providers, closer links with industry, and new forms of management have weakened that special status. Academics are now far more closely regulated in the performance of their duties. Tenure is no longer the norm; the connection between teaching and research is weaker; research is more likely to be dependent upon obtaining external funding, and also to be subject to restrictive codes of ethics; especially in the humanities and social sciences, public commentary has become more hazardous.

In 1975, the United States Senator William Proxmire invented the Golden Fleece Award for wasteful public expenditure and conferred the first award on the country's principal research agency, the National Science Foundation, because it had provided a grant of US\$84 000 for a study of why people fall in love. In the following decade, the Thatcher government waged a similar campaign against the British agency, the Social Sciences Research Council, for funding left-wing researchers and supporting arcane projects – the example singled out for ridicule was a study of village life in Poland.

Similar invigilation of research and publications spread to other countries, and foreshadowed the culture wars waged by neo-conservatives. Commonly associated with allegations of political correctness, betrayal of the Western canon and abandonment of truth for postmodernist nihilism, these media campaigns have done considerable damage to individual reputations and academic disciplines. They clearly impinge on the principle that academics should be free to conduct teaching and research without vilification or intimidation.

Student Freedom

A similar development arose in 2001 with the formation of Students for Academic Freedom in the United States, with the purpose of protecting students from an alleged liberal bias in the classroom. The organization drafted an academic bill of rights, defining academic freedom as the freedom to teach and learn in an atmosphere of intellectual diversity, and claiming that liberal arts faculties

at most universities are politically and philosophically one-sided. It claims to have over 150 chapters and its website coordinates publicity for members' reports of partisan indoctrination, anti-Americanism, refusal to allow conservative perspectives and profanity. Two states have passed legislation based on its bill of rights.

Further Reading

- Aby, S. H. and Kuhn, J. C. IV (eds.) (2000). *Academic Freedom: A Guide to the Literature*. Westport, CT: Greenwood Press.
- American Association of University Professors (2001). *Policy Documents and Reports*, 9th edn. Washington, DC: American Association of University Professors.
- First Global Colloquium of University Presidents, Columbia University (2005). Statement on Academic Freedom. www.aft.org/higher_ed/pubs-reports/AcademicFreedomStatement.pdf.
- Hofstadter, R. and Metzger, W. P. (1955). *The Development of Academic Freedom in the United States*. New York: Columbia University Press.
- International Association of Universities (1998). *Trends in Higher Education: Academic Freedom*. Bologna: International Association of Universities.
- Lima Declaration on Academic Freedom and Autonomy of Institutions of Higher Learning* (1988). Lima: United Nations.
- Marginson, S. (1997). How free is higher education? *Higher Education Research and Development* **16**, 359–369.
- Poch, R. K. (1993). *Academic Freedom in American Higher Education: Rights, Responsibilities and Limitations*. ASHE-ERC Higher Education Report no. 4. Washington, DC: The George Washington University, School of Education and Human Development.
- Shils, E. (1991). Academic freedom. In Altbach, P. G. (ed.) *International Higher Education: An Encyclopedia*, pp 1–22. New York: Garland Publishing.
- Students for Academic Freedom (2005). *Students for Academic Freedom. Mission and Strategy, and Second Year Achievement Report*. <http://cms.studentsforacademicfreedom.org>.
- World University Service (1990). *Academic Freedom: A Human Rights Report*. Atlantic Highlands, NJ: Zed Books.

Access and Equity in Higher Education

G Baldwin and R James, The University of Melbourne, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

The rates and patterns of access to and participation in higher education are significant policy issues in most countries. Social equity in the patterns of access to higher education is often depicted to be one of three criteria for the quality of a national higher education system, alongside effectiveness and efficiency. The literature on access and equity in higher education is thus vast, including national analyses and related policy documents, research reports, and comparative international reports by bodies such as the Organization for Economic Co-Operation and Development (OECD), United Nations Educational, Scientific, and Cultural Organization (UNESCO), and the World Bank. Contemporary research into access and equity is tightly interwoven with wider policy issues and research questions, including the purposes of higher education, the funding of higher education, the assessment of academic achievement, and the changing character of higher education delivery and patterns of teaching and learning.

An analysis of the patterns of access and equity in higher education systems across the world produces a paradoxical picture of great diversity on the one hand and remarkable uniformity on the other. The rates of participation vary enormously, yet significant expansion has been virtually global over the last 50 years, driven by public demand, labor-market demand for higher skill levels, and government policies designed to improve national economic competitiveness. Despite massive growth, inequities persist in all societies, if the measure of equity is the extent to which the demographic profile of students in higher education matches the profile of the community as a whole. However, the bases of disadvantage vary according to social, economic, and political contexts. The widespread and seemingly intractable source of disadvantage is low socioeconomic status (SES), but that disadvantage takes various forms in different countries. In developing countries, poorer students have little chance of gaining entry into higher education, while in many of the developed countries they will tend to find places in the less-prestigious institutions and fields of study. The extent to which expansion has ameliorated disadvantage is a subject of intense debate, with some commentators arguing that it has actually widened the gap between rich and poor, and others taking a more positive view of recent progress. Regardless, policies and programs designed to create more equitable access appear to have had mixed success.

The Availability of Data on Higher Education Participation

Much of the research into access and equity has focused on solving the daunting problems of data gathering, so that countries have accurate information on the progress they are making toward the extension of higher education to all capable of benefiting from it. The work done in recent years by international bodies to develop reliable data systems for monitoring higher education is discussed in detail later. It is evident that many problems remain and there will continue to be much debate about the validity of data.

Many individual countries have developed extensive databases for monitoring access and equity in their higher education systems. Australia, for example, collects and analyzes detailed information on access, participation, and completion rates, considering a wide range of factors including age, gender, SES, locality, ethnicity, and disability. This data collection has become increasingly detailed and sophisticated over the last 30 years and time-series analyses are published every year, allowing for detailed investigation of trends (e.g., Coates and Krause, 2005). Australia has the advantage of dealing with a centralized, almost entirely public system. In contrast, while there is extensive data collection in the United States, the national picture is not as easy to analyze because of the highly decentralized nature of the higher education system.

The extent and quality of data vary considerably among different countries, again in largely predictable ways: poorer countries generally do not have the resources to devote to extensive data collection and analysis. A further problem is that, as different systems have developed, there has been little agreement on methodologies and definitions. Perhaps, the most basic disagreement has been over the meaning of the term higher education and particularly the place of technical education, which takes a variety of forms in different countries.

Much complex, useful, work has been done by the OECD, UNESCO, and the World Bank, to establish common definitions and procedures. Thus, for example, the categorization of tertiary education into type A (theoretically based programs which are designed to provide qualifications for entry into advanced research programs and professions with high-skill requirements and which lead to the equivalent of bachelor, master, or diploma

degrees) and type B (more occupationally oriented programs which prepare for direct entry into the labor market and are typically of shorter duration) has been widely accepted as a meaningful distinction, although there are still uncertainties about where particular systems should be placed. Further, the data relating to type B systems tend to be much patchier and less reliable than those for type A. In general, higher education is taken to coincide with tertiary type A programs. It has been argued by Clancy and Goastellac (2007) that entry rates should not be used at all, since, as the OECD points out that there is overlap between types A and B, the two rates cannot be added to give an overall rate. This is true if one is considering tertiary as opposed to higher education, but this distinction is not universally agreed upon.

The Global Expansion of Higher Education Participation

Despite the problems associated with the lack of comparable international data, some overall participation trends are evident. Martin Trow's seminal paper, *Problems in the Transition from Elite to Mass Higher Education*, written in 1973, has provided a valuable theoretical framework for understanding the rapid changes in higher education systems since the middle of the twentieth century. The framework defines three ideal types of elite, mass, and universal systems of higher education, defined as those in which 0–15%, 16–50%, and over 50% of the relevant age groups,

respectively, participate in higher education. In an updated version of this paper (Trow, 2006), Trow argues that the United States, as the country which has led the way toward universal access, offers the best model for others to follow. Other developed, industrialized countries are at the mass stage, moving toward the universal. Although he does not deal specifically with developing countries, most of these, according to his schema, are still attempting to move from the elite to the mass stage. However, there are some notable exceptions – Russia, Chile, and Argentina – which are discussed further below.

Despite this variation, the overall picture is clear. In the high-income countries (those in the OECD), expansion of higher education started after World War II, gathered pace in the 1960s and 1970s, slowed a little through the 1980s, and took off again in the 1990s. At the other end of the income scale, in poorer countries, such as those in sub-Saharan Africa and parts of Asia, the expansion started later and still lags well behind the growth in the wealthiest countries. In relative terms, this expansion appears quite dramatic but it started from a very low base, so the gap between rich and poor countries would seem to be widening. It may well be that, in attempting to increase the opportunities for their own citizens, developed countries are increasing the relative disadvantage of their less-developed neighbors. **Figure 1** presents World Bank gross tertiary enrolment (GER) data which highlight the regional discrepancies in growth rates, expressed as the number of students participating in tertiary education as a percentage of the number of people in the age cohort.

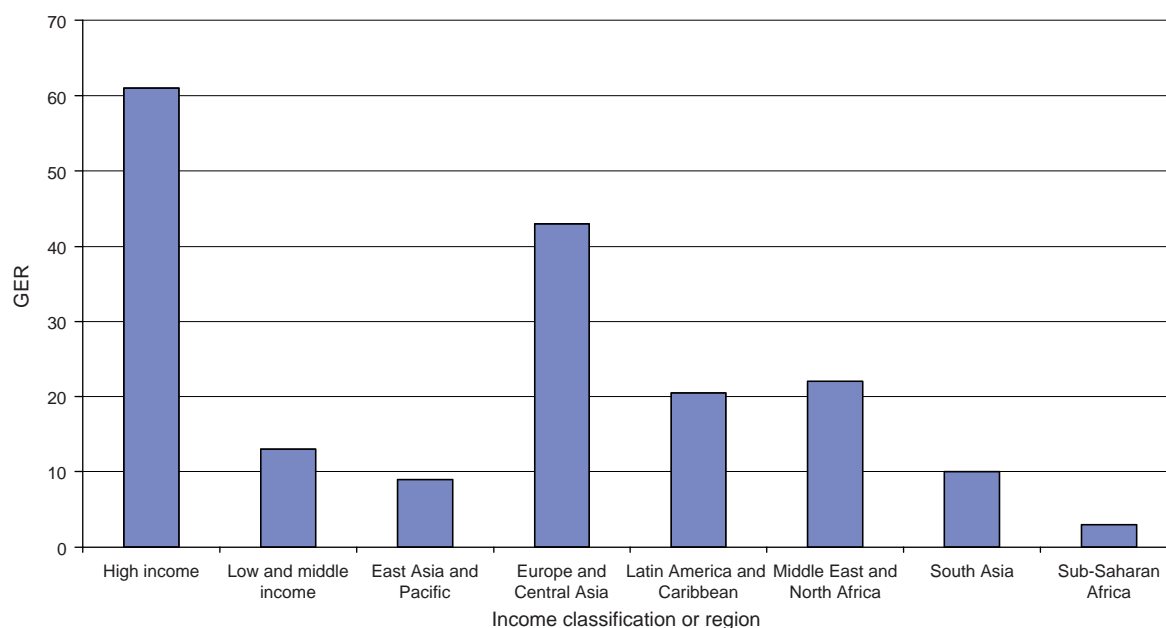


Figure 1 Tertiary gross enrolment ratio (GER) by region (%). The data are for 2000 or most recent year. The regional classification does not include any high-income countries. From World Bank – *International Handbook of Higher Education*, p. 448.

However, for those committed to reducing the global participation imbalances in higher education, there are grounds for optimism in the figures produced by UNESCO and the OECD, which suggest that, between 1995 and 2003, the group of middle-income nations analyzed as the world education indicator (WEI) countries saw greater expansion in their higher education systems than the high-income OECD countries – 77% as opposed to 43%. These countries are: India, China, Peru, Zimbabwe, Egypt, Indonesia, Chile, Russia, Jordan, Thailand, Paraguay, Jamaica, Argentina, Tunisia, Malaysia, the Philippines, and Brazil. Further, the UNESCO report points out that, in terms of entry rates, a number of these countries had better results than OECD countries, with Argentina, Chile, and Russia above the OECD average. These impressive rates suggest that Trow's analysis may be too narrowly focused on Western Europe and North America. One of the problems in interpreting these data, however, is that some systems (e.g., such as Argentina) have high noncompletion rates, which raises the question of whether entry into higher education is a useful measure or whether completion rates are more revealing. This is but one of the complex methodological problems which hinders research and analysis in this area.

Higher Education Access and the Underrepresentation of Population Subgroups

Despite the extraordinary expansion of access to higher education, social inequities in participation exist in all nations regardless of the stage of development of their systems. In a number of nations, the expansion of higher education systems from elite to mass access has placed renewed attention on the demographic composition of the student population. To some extent, the level of overall participation defines the characteristics of the equity considerations.

The concern for equitable participation in higher education is underpinned by the goal of achieving individual social justice and the belief that increased participation in higher education of people from educationally disadvantaged groups is important for the long-term social and economic integration of these groups. The concept of equity is open to a number of interpretations. Equality of opportunity is often referred to as fair access without discrimination on gender, class, ethnicity, or other personal characteristics and without explicit or implicit barriers to participation, such as the cost of tuition fees and living expenses while studying. Increasingly, in the diverse systems in the mass stage, equality of outcomes has become a more salient objective, in recognition that greater equity in access to higher education may not lead to greater equality in outcomes should the personal and career

outcomes differ markedly across universities or fields of study. A comprehensive consideration of equity might therefore look not only at broad access but also the type of higher education in which people participate – the nature of institution, the field of study, and the level of study – as well as retention rates, completion rates, and graduate outcomes.

Though equity is often depicted in terms of individual social justice, one of the most common measures of educational advantage and disadvantage has been the patterns of representation of subgroups in national populations. The social and demographic factors associated with subgroup underrepresentation vary across nations and cultures. The community groups underrepresented include indigenous people, ethnic minorities, women, people from lower socioeconomic backgrounds, people from rural or remote areas, and people with disabilities. The position of ethnic minorities ranges from gross underrepresentation to significant overrepresentation, depending on the historical, social, and demographic circumstances. The extent of disadvantage associated with rural locations depends to some extent on the size of the country (James *et al.*, 1999). Access for students with a disability has advanced much more in some nations than others (Ebersold and Evans, 2003).

Women's participation in higher education remains an issue. In some countries, women are still underrepresented but in others they are clearly in the majority, although not necessarily in all fields of study or at all levels of awards. **Table 1** shows the sizable variation in the UNESCO gender parity index (GPI), which is the ratio of males to females (a GPI of 1.00 indicates parity, and greater than 1.00 indicates more females than males). All of these GPIs rose between 1999 and 2004 with the exception of sub-Saharan Africa.

In most nations, social class or SES is the most reliable predictor of the likelihood that an individual will participate in higher education. One German study in the 1990s (Schnitzer *et al.*, 1999) reported that, of 100 lower social background children only 33 reached upper secondary school and only eight entered higher education. In contrast, of 100 upper social background children, 84 reached upper secondary and 72 entered higher education.

Table 1 UNESCO gender parity index for all tertiary enrolments, 2004 (ratio of males to females, including type A and type B)

Arab states	0.95
Central and Eastern Europe	1.25
Central Asia	1.05
East Asia and the Pacific	0.89
Latin America and the Caribbean	1.17
North America and Western Europe	1.32
South and West Asia	0.70
Sub-Saharan Africa	0.61

The Higher Education Funding Council for England has reported that young people from the most prosperous areas are 5–6 times more likely to go to university than young working-class people in particular areas of disadvantage (HEFCE, 2007). Finland is an important example because despite a highly egalitarian public school system, higher education participation is skewed according to parental educational backgrounds. The OECD thematic review of tertiary education (Davis *et al.*, 2006) reported that:

...large inequalities in access to tertiary education by social origin still persist in Finland. Participation rates in university education among young students (aged 20–24) differ considerably according to the educational background of their parents. The relative chance of entering university education has remained at least ten times higher during the last decades for those coming from academic home background compared to students from less educated families. The expansion of the tertiary system appears to have narrowed the relative advantage of an academic home background to seven-fold.

In many nations, the expansion of participation has done little to influence the participation share for low socioeconomic students. In Australia, for example, the proportion of low socioeconomic background students in higher education remained virtually static at close to 14.5% (against a community reference point of 25%) in the period 1996–2005, during which time the overall growth in enrolment of domestic students was 17% (DEST – Department of Education, Science and Training, 2006). In the USA, some commentators have argued that continued growth has created greater social inequality in the nature of higher education participation (Astin and Oseguera, 2004), resulting from standardized entry testing used for selection purposes and rising tuition costs.

Despite the significance of SES in determining the likelihood of higher education participation, by its nature it is a difficult concept to define and measure. This has a range of implications, including for efforts to broaden access and for the collection of data policy for comparison and monitoring purposes. Various approaches have been used, all with limitations, including parental occupations and educational levels, family income levels, and geographical indicators such as the zipcode or postal code of home address.

Numerous factors underlie the underrepresentation of certain groups. The interrelationships between these factors are not well understood. Nonetheless, the underrepresentation of certain subgroups in higher education is a result of the combined effects of lower school-completion rates, lower levels of educational attainment in schools – thus limiting opportunities in the circumstances of competitive entry based on academic achievement – lower levels of educational aspiration, and lower perceptions

of the personal and career relevance of higher education, alienation from the culture of universities, and a range of interrelated financial factors: the expense of university fees, the lack of availability of income support while studying, and the loss of potential income while studying. Some of these factors are explicit barriers to participation; others are not literally barriers but reflect that absence of family, school, or community environments that lead to the development of higher education aspiration. The imbalances in higher education participation often reflect endemic educational disadvantage that begins in the earliest years of schooling.

The underrepresentation of certain groups of people has led to policy responses in many developed nations. Policies and programs designed to achieve more equitable access to higher education have been implemented at both national and institutional levels. Equity initiatives have taken a number of forms, including the removal or reduction of perceived barriers, such as scholarships to fully or partially meet the cost of fees, compensatory admission for students with lower levels of school achievement as a result of family or personal hardship of educational disadvantage, and programs described as affirmative action that focus on subgroup membership as the basis for admission. Policies for affirmative action or positive discrimination have been contentious for they conflict with conceptions that entry to higher education should be based on merit. The admission of people with lower levels of educational attainment has been seen by some to lower academic standards and to take places away from those who deserve them. Specific affirmative action programs in the United States have been successfully challenged in legal cases that have significant ramifications for public and institutional policy (Allen, 2005).

The effects of equity policies are not clear. Almost paradoxically, expansion has made it difficult to assess whether gains have been made with equity. Certainly, at the aggregate level, policies appear not to have significantly reduced the persistent, proportional underrepresentation of certain subgroups of people. However, it is the case that higher numbers of people from previously underrepresented groups are participating in higher education in the nations with expanded systems. Whether this is an effect simply of expansion or a result of equity initiatives, or both, is difficult to discern.

It is increasingly evident that there are significant limits to the capacity of policies centered in higher education to address the problems of group underrepresentation for the higher socioeconomic groups are more successful in gaining access to higher education, and to the most prestigious universities and fields, partly because of higher school-completion rates and higher levels of school achievement. Investment in schooling, including in the earlier years, is an essential element of a coordinated policy approach to widening higher education access.

The Issues and Areas for Research in the Context of Increased Demand for Higher Education

The Limits to Expansion

Against the background of great complexity and difficulty in monitoring the rapid expansion of higher education around the globe, the reasons for it are starkly clear and simple. It is a general perception that postindustrial societies require higher levels of education, and the heightened aspirations of individuals and their families in a globalized world are seen as depending on education. One of the most interesting aspects of this field is the extent to which the goal of universal higher education seems to have been accepted by most governments around the world. The rhetoric associated with the knowledge economy seems to be virtually unchallenged. Any doubts that what we may be witnessing is largely creeping (or perhaps galloping) credentialism are silenced by research which demonstrates that more education is strongly associated with increased productivity. Whether there is any limit to this acceleration remains unclear.

Collecting Data to Track Access and Equity

A range of indicators has been developed to address the issue of what is the most meaningful way of gauging higher education participation in societies. As Trow points out, the use of a young-adult age group as a reference point for participation rates is becoming increasingly problematic, as more mature students enter higher education and the ideal model moves toward the concept of lifelong learning. Therefore, in addition to entry and graduation data, the OECD and UNESCO calculate the average number of years of higher education completed by citizens and projected future levels based on growth at lower educational levels. Clancy and Goastellac (2007) propose that a composite measure is needed, one which incorporates three participation and two outcome indicators.

Another area in which new measures and collection systems are urgently required is transnational enrolments. As more students enter higher education programs in other countries, they can slip through the data-gathering net. In Australia, for instance, although information about international students is collected, it is not analyzed with the data relating to domestic students and is seen, probably appropriately, as not relevant to the monitoring of equity in participation within the host country. There is as yet no comprehensive system for monitoring these patterns internationally. The need is clear: a large proportion of citizens of low-income countries who gain entry into higher education do so in countries other than their own. Issues of access and equity in higher education need to be addressed

globally as well as nationally and for this more comprehensive and refined databases are essential. Further, present conceptions of SES are firmly grounded in national economic, social, and cultural frameworks and a new understanding is needed of how participation rates vary by social class in a global higher education environment.

The Concept of Meritocratic Entry

Merit and equity continue to be in tension. The concept of equity in elite systems of higher education has been partly based on the meritocratic principle that certain people are deserving of higher education on the basis of untapped intellectual potential. Yet, Trow (1973) speculated that as systems moved from mass to universal participation, access would move from being a right to an obligation, and that meritocratic admissions and compensatory programs for equity would be replaced by open access and equality of group achievement, on the basis of either class or ethnicity. There is little evidence of this occurring on a wide scale, even in the most expanded systems. Instead, there is the coexistence of meritocratic and open-access admissions, with institutions adopting differing stances toward student recruitment and selection in relation to their hierarchical position within highly diversified systems – emphasis in admissions policy for some institutions shifts from the challenge of fair, meritocratic selection toward the active recruitment of students, including people with lower levels of academic achievement.

Within the most expanded higher education systems, institutional diversity heightens competition for the places in the most prestigious universities: students compete for entry to what are perceived to be the best universities, while in turn the more prestigious universities compete for students with the highest level of school achievement. The importance of higher education as a positional good to some extent prevails.

Open Access

A relatively new layer of the higher education hierarchy has been added in recent decades, with the development of open institutions. The concept of openness applies primarily to admissions policies. To varying extents in different countries, entry is open to any student with minimal educational qualifications or, in some cases, to any individual who wishes to attempt a course, without any prerequisites. Increasingly, this concept has become associated with distance-learning procedures which, again to varying degrees, are used to minimize costs. Although supporters of distance education in its more traditional forms argue that it should be no cheaper to fund than face-to-face teaching, there is no doubt that the forms it is now taking (including increasing use of online learning) are intended to be significantly less expensive. In some

developing countries, these institutions have developed into mega-universities with the largest in China (the China Central Radio and Television University) enrolling approximately 2 million students. While such programs offer valuable opportunities to many people who would not be able to engage in higher education in any other form, the status of these qualifications is generally lower than those from traditional universities and the outcomes for students patchy. High proportions of students fail to complete their courses in open institutions. Again, although the data are quite inadequate in this area, it is clear that students in these programs are more likely to be from disadvantaged groups.

Funding Expansion

If there is general agreement about the desirability of universal higher education, no country seems to have found a very satisfactory answer to the obvious question of how this is to be funded. The United States, the richest nation in the world, has probably come closest, although numerous commentators, politicians, and university leaders have drawn attention to the signs of strain even in that system. There have been four main responses, which again have been strikingly consistent across the range of national systems, and which have important implications for access and equity. These are: ongoing reductions in funding per unit of student load (i.e., average expenditure per student), increasing stratification of institutional types, the development of open university programs, and privatization. All of these powerful trends are analyzed elsewhere in this encyclopedia. They will be considered here only in relation to equity issues. For it is a striking irony that, in order to increase access to higher education, governments have had to adopt policies which have the potential to exacerbate existing inequities. Decreased average funding rates would not in themselves have this effect, if applied equally. However, in many countries distribution is unequal, with some sectors of higher education being much less expensive to fund than others. This inevitably coincides with a hierarchy of prestige and, in turn, with different outcomes for students in terms of employment and earnings. Evidence from a number of countries indicates that members of disadvantaged groups are much more likely to attend less-prestigious institutions.

The Impact of Privatization

A further effect of the inability of governments to fund higher education adequately is a very rapid process of privatization, which takes a variety of forms. Privatization involves students paying increasing proportions of the costs of their education, either directly to private

institutions or back to governments in the form of upfront or delayed payments. The term can also be applied to the process of raising substantial funds from private enterprise or individuals. Ability to pay will obviously disadvantage individuals from lower socioeconomic groups, although the Higher Education Contribution Scheme (HECS) devised in Australia and adapted by some other countries does offer the option of delayed payment which is only required when a certain income threshold has been reached – that is, in theory, when the individual has acquired the ability to pay. Another form of privatization is subsidization of a set number of higher education places, while allowing individuals who do not win one of these places, to pay the full costs of their education. Unfortunately, another universal pattern seems to be that academic results tend to reflect socioeconomic differences, so that the low SES students who are more likely to miss out on the subsidized places are those least likely to be able to afford full fees. In a familiar vicious cycle, the most prestigious institutions are those most likely to benefit from privatization, in terms of attracting full fee-paying students, business support, and private donations, so the gap in quality and outcomes is likely to widen further.

Conclusion

The expansion of higher education systems from elite to mass to universal systems has not resolved inequities in access. Of course, in this context, equity is a relative concept. Just as with poverty, the definition cannot be absolute, but is determined by comparisons between groups. It is undeniably true that, in every country in the world, more citizens are receiving the benefits of higher education than ever before. The fact that comparative inequities persist should not cloud recognition of the achievement involved in the massive expansion of the last half-century.

Access and equity in higher education remain important issues for public policy in most countries. In the more expanded systems, institutional stratification has invited growing consideration of whether equity should be considered at the level of system, institution, or field of study. A major issue for research and policy during the next decade is the development of a new conceptualization of equity that takes into account the globalization of higher education and international student flows.

Bibliography

- Allen, W. R. (2005). A forward glance in the mirror: Diversity challenged – access, equity and success in higher education. *Educational Researcher* 34, 18–23.

- Astin, A. W. and Oseguera, L. (2004). The declining “equity” of American higher education. *Review of Higher Education* 27, 321–341.
- Clancy, P. and Goastellec, G. (2007). Exploring access and equity in higher education: Policy and performance in a comparative perspective. *Higher Education Quarterly* 61(2), 136–154.
- Coates, H. and Krause, K. (2005). Investigating ten years of equity policy in Australian higher education. *Journal of Higher Education Policy and Management* 27, 35–47.
- Davies, J., Weko, T., Kim, L., and Thulstrup, E. (2006). *OECD Thematic Review of Tertiary Education. Finland Country Note*. Paris: OECD.
- DEST (Department of Education, Science and Training) (2006). *Students 2005: Selected Higher Education Statistics*. www.dest.gov.au. (accessed May 2009).
- Ebersold, S. and Evans, P. (2003). *Disability in Higher Education*. Paris: Organisation for Economic Co-operation and Development.
- HEFCE (2007). *Strategic Plan 2006–2011*. Bristol: Higher Education Funding Council for England.
- James, R., Wyn, J., Baldwin, G., et al. (1999). *Rural and Isolated Students and Their Higher Education Choices: A Re-Examination of Student Location, Socioeconomic Background, and Educational Advantage and Disadvantage*. Canberra: Australian Government Publishing Service.
- Schnitzer, K., Isserstedt, W., Mussig-Trapp, P., and Schriebe, J. (1999). *Student Life in Germany; the Socio-Economic Picture, Summary of the 15th Social Survey of the Deutsches Studentenwerk (DSW)*. Bonn: Bundesministerium für Bildung und Forschung.
- Trow, M. (1973). *Problems in the Transition from Elite to Mass Higher Education*. Berkley, CA: Carnegie Commission on Higher Education.
- Trow, M. (2006). Reflections on the transition from elite to mass to universal access: Forms and phases of higher education in modern societies since WWII. In Forest, J. and Altbach, P. (eds.) *International Handbook of Higher Education*, pp 243–280. Dordrecht: Springer.

Further Reading

- Douglass, J. (2007). *The Conditions for Admission: Access, Equity and the Social Contract of Public Universities*. Palo Alto, CA: Stanford University Press.
- HEFCE (2006). *Widening Participation: A Review*. Report to the Minister of State for Higher Education and Lifelong Learning by the Higher Education Funding Council for England.
- James, R. (2001). Participation disadvantage in Australian higher education: An analysis of some effects of geographical location and socioeconomic status. *Higher Education* 42, 455–472.
- Layer, G. (ed.) (2005). *Closing the Equity Gap: The Impact of Widening Participation Strategies in the UK and the USA*. Leicester: National Institute of Adult Continuing Education.
- Naidoo, R. (2007). *Higher Education as a Global Commodity: The Perils and Promises for Developing Countries*. London: The Observatory on borderless higher education.
- OECD (2006). *Education at a Glance*. Paris: Organisation for Economic Co-operation and Development.
- Skilbeck, M. (2000). *Access and Equity in Higher Education: An International Perspective on Issues and Strategies*. Dublin: Higher Education Authority.

Boundaries of Institutional Autonomy and their Impact on Higher Education

V L Meek, University of Melbourne, Parkville, Victoria, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Institutional autonomy and the related concepts of academic or scientific freedom are defining characteristics of academe. "The university is the corporate realization of man's basic determination to know" (Jaspers, 1960: 20), and it is this classic notion of knowledge as sacrosanct upon which arguments for the autonomy of the university are ultimately grounded. Traditionally, it was assumed that the purpose or idea of the university was best served by its separation from government, on the one hand, and society, on the other. The state served as the guardian rather than the determiner of academic principles, with the issue being "how far higher education institutions are licensed to be free to set their own norms, or even to be in conflict or tension with the society that sponsors them to be its antibodies" (Kogan, 1984: 67). For most contemporary higher education institutions, however, the boundaries of institutional autonomy have become exceedingly porous.

Institutional autonomy is a relative concept, conditional on a variety of historical, cultural, political, and economic factors. Institutional autonomy is not in itself a guarantee of academic freedom, and the degree of genuine institutional autonomy has waxed and waned over the years in accordance with large-scale political changes in national systems of higher education. For example, within the context of the so-called European continental model of higher education governance, considerable enhancement of university autonomy may be observed over recent decades, while it is arguable that higher education institutions operating under the Anglo-Saxon tradition have moved in the opposite direction (Trow, 2005). It appears that the autonomy of Japan's national universities is being considerably strengthened with their transformation into "national university corporations" (Arimoto, 2006). In every political jurisdiction, the degree of autonomy enjoyed by higher education institutions varies according to type and function of the institution, with universities at the top of most autonomy hierarchies. Looked at historically and cross-culturally, the autonomy of higher education institutions and the principles underpinning it are in a state of constant flux.

One approach to better understanding how the boundaries of institutional autonomy shift over time and space is to view the phenomenon from both external and internal perspectives. The external environment in which

universities operate and various powerful stakeholders in that environment – politicians, employers, parents, etc. – place limits on what higher education institutions can and cannot do. However, autonomy is also bound by the university's own internal dynamics – by the ways in which it is managed and governed and the mechanisms it uses to distribute power and decision making. Moreover, autonomy itself can be defined from a number of perspectives.

Definitions

Much of the recent changes in higher education in many jurisdictions involve questions of institutional autonomy and academic freedom. However, the autonomy debate is not new nor is it restricted to any one country, except in its details (see Ashby, 1966; Berdahl, 1988). Moreover, debates about autonomy are often more emotive than analytically rigorous, and, from an analytical perspective, a distinction needs to be made between academic (maybe better phrased as scientific or academic freedom) and institutional autonomies. Drawing on Ashby (1966), Berdahl (1988: 7) defines academic freedom as the "freedom of the individual scholar in his/her teaching and research to pursue truth wherever it seems to lead without fear of punishment or termination of employment for having offended some political, religious, or social orthodoxy." On the other hand, "institutional autonomy... is typically viewed as the right to self-determination in the appointment of academic staff, student admission, teaching content and methods, standards control, priority setting and future development" (Kayrooz, 2006: 4).

In its literal sense, no higher education institution has complete autonomy; autonomy is not an all-or-nothing issue. There is an important distinction between the autonomous university as a self-governing scientific and educational community in a dynamic relationship with society, and the university as autarchy – totally self-sufficient and having absolute sovereignty (Zgaga, 1997). Higher education institutions will always be subject to some demand to be publicly accountable, whether the institutions themselves are public or private. Society has too much of an interest in higher education to allow autarchy or pure autonomy to prevail. According to Ashby (1966: 296), what is important is to examine the "essential ingredients" of institutional autonomy:

1. Freedom to select staff and students and to determine the conditions under which they remain in the university.
2. Freedom to determine curriculum content and degree standards.
3. Freedom to allocate funds (within the amounts available) across different categories of expenditures.

Drawing again upon Ashby, Berdahl (1988:7) further subdivides autonomy into substantive and procedural issues: “*substantive autonomy* is the power of the university or college in its corporate form to determine its own goals and programs . . . ; *procedural autonomy* is the power of the university or college in its corporate form to determine the means by which its goals and programs will be pursued” Interferences in procedural autonomy – for example, control over purchasing – “can be an enormous bother to Academe, and often even counter-productive to efficiency, but still usually do not prevent universities or colleges from ultimately achieving their goals. In contrast, governmental actions that affect substantive goals affect the heart of Academe” (Berdahl, 1988: 8–9).

Generally, in many countries, substantive autonomy has either been maintained, as in Australia and Britain, or is being substantially increased, as in many continental European nations. However, simultaneously, institutions are being held more accountable to governments and a variety of other external stakeholders for achievement of their goals and missions. The nature and detail of procedural autonomy wax and wane in all countries, to the great annoyance of institutional administrators.

Rather than viewing autonomy as an absolute, it should be viewed as a relational issue involving the balance of power between institutions and government, on the one hand, and between management and the academic profession within institutions, on the other. Direct threats to academic freedom are more closely associated with the internal balance of power between executive and collegial governance than with external intervention, although the executive arm of the institution may act as a proxy for government bureaucrats.

External Boundaries of Institutional Autonomy

What is important to understand is not the strengthening or weakening of the boundaries of institutional autonomy *per se*, but the demands of changing social and political contexts that set the limits on the degree to which higher education institutions can behave as independent actors. Historically, the most important external political factor determining the autonomy of the university was the church – now almost everywhere replaced by the state.

Although, nearly everywhere, legislative frameworks set the broad boundaries within which autonomy is exercised, the autonomy of the university as a legal

entity varies in its detail from country to country and over different historical periods. In Australia, for example, what constitutes a university is defined by legislation, as are a set of protocols that specify its functions. In the USA, the term university is used much more freely, while in the former eastern European block of now newly emerging states – such as the Czech Republic and Slovenia – the legal autonomy of the university as a corporate body is being created from a collection of disparate state-controlled faculties and research institutes.

However, more important than the legislative framework, in which higher education institutions operate, is the impact on institutional autonomy of the demands of government, consumers, and other stakeholders for the enhanced accountability, relevance, and contribution to the knowledge economy of higher education institutions. Neoliberal reforms of higher education systems, the imperatives of New Public Management, and international trade agreements all have an impact on institutional autonomy. Supranational agreements, such as the Bologna process in Europe, and global initiatives, such as the WTA General Agreement on Trade in Services (GATS) and International Monetary Fund (IMF) funding stipulations in developing countries, all have the potential to curtail institutional autonomy. “Institutional Autonomy becomes a contingent item in a broader political and economic agenda” (IAU, 2006: 2). Moreover, as Zgaga (1997: 124) notes, “the discussion of the external dimension of university autonomy includes not only the relationship to the state – that is, the regulation of the legal system and funding – but also to civil society, its heterogeneous needs, value orientations, individual and group aspirations, etc. And this relationship is not necessarily always harmonious.”

In relation to the external environment, two powerful and interrelated concepts presently reshaping university autonomy are that of the Service University – the shift of higher education from a public to a consumer good – and market steering of higher education systems. The Service University is responsive and relevant to the needs of clients and stakeholders, and with it “comes the notion of ‘conditional autonomy’, this is institutional autonomy subject to rendering accounts to public and to financial providers” (IAU, 2006: 2). Supposedly, the ideals of the Service University are best achieved under market-like conditions that demand competitive and responsive behavior from institutions.

The notion of the Service University is tightly bound to those of government deregulation, accountability, and market competition – a belief in the hidden hand of the marketplace as a powerful mechanism for bending the goals of higher education institutions to the needs of society. “In the new market context of higher education, policy makers are seeking to redesign systems of higher education to address the newly conflicting demands of institutional autonomy and public accountability” (Dill, 2001: 23).

It is argued, in several jurisdictions, that market competition – as opposed to centralized state control – is better able to produce innovative, adaptive, and responsive higher education institutions. Underpinning this shift in government steering of higher education from what Van Vught (1989) termed the state-control model to the state-supervising model is the ideology of economic rationalism. Apparently, universities and colleges must become more economically efficient through such means as greater competition in a more open academic market, through the strengthening of institutional management along the lines employed by major corporations, and through improved public accountability. The shift from state to consumer control reflects a belief that market forces, rather than state intervention, will make institutions more cost effective and better managed, as well as making higher education systems more fluid and responsive to client needs and demands. Somewhat paradoxically, as discussed in more detail below, while in many nations the direct legislative control of higher education institutions has been relaxed, academic leaders and rank-and-file staff feel that the autonomy of their institutions has diminished significantly.

Universities have always competed with one another for students, resources, and prestige. However, it has only been since the late 1990s that the concept of the market has helped regulate the relationship between institutions and government, on the one hand, and between institutions and the broader community, on the other. Government is not totally disinterested in the regulation of higher education as is evidenced by increasing emphasis on quality control and other system-wide accountability measures. However, in order to ensure that higher education will cost less while better serving national economic priorities, the concept or metaphor of the market has become “central to a number of discourses which constitute the current policy agendas of governments . . . and educational institutions. Education is currently being thought of in market terms and markets of various sorts are guiding priorities and funding” (Kenway *et al.*, 1993: 2).

There may be a relationship between institutional autonomy and the multiplicity of funding sources upon which the institution is dependent – the less dependent on any single source of funding, the greater the degree of autonomy. However, what is at issue here is not the multiplicity of funding sources *per se*, but rather the structures of control embedded in particular types of funding mechanisms. In this sense, market forces may be more of a threat to institutional autonomy than government policy and centralized control ever was. There is a potential rigidity to the market, for “market forces create more, not less bureaucracy which by its very nature needs to see, taste and manipulate *objective* performance indicators so that the operators of bureaucracy . . . can evaluate what constitutes value for money” (McElwee, 1992: 191).

Universities compete with one another in attempting to interpret how best to take advantage of external and other financial incentives, and, in so doing, have been caught in a continuous process of trying to second guess both the market and government policy. While the relationship between universities and government may nominally be centered on autonomy, it is actually based on funding rewards to institutions that support governmental objectives which result in university competition for funding and enforced conformity to governmental goals. Somewhat paradoxically, universities are given the freedom to conform. Neave (1996) encapsulates this seeming paradox in what he terms as the “law of anticipated results.” The law of anticipated results operates at the institutional level – giving the impression of autonomous institutional action to what is, in fact, an institutional reaction to actual or anticipated external forces, directives, or events. Institutions interpret what is or will be required by governmental policy and act accordingly, making it considerably difficult to determine whether change is institutionally driven or externally imposed. Thus, “the rhetoric of public policy is often at odds with the institutional behaviour it elicits” (Neave, 1996: 30). Many institutional leaders report that their autonomy has been restricted. Possibly, it may be the general competitive environment that governmental policies help create, rather than the policies *per se*, that institutional leaders perceive as unduly constraining their activities and freedom.

The market – as a form of social coordination – has in many jurisdictions replaced state bureaucracy as the primary regulator of higher education. In a totally free market, no one is in control. However, governments everywhere regulate markets and “a mix of factors, including values and legal systems, determines at any moment the balance between market coordination and other forms of social coordination” (Geiger, 2004: 233). In many countries, the higher education market is balanced by a number of accountability measures, for example, external quality assurance frameworks for both teaching and research. While accountability may be seen as a threat to the self-referential autonomy of the university, it is only one aspect of the institution’s relationship with society and “other relationships, if cultivated alongside accountability, can work to strengthen university autonomy rather than weaken it” (Gibbons, 1998: 58). Several students of higher education have argued that in the last few decades what has been lost in the relationship between higher education institutions and society is not so much autonomy as trust (Trow, 2005).

Internal Boundaries of Institutional Autonomy

The higher education institution itself, incorporating its governance and management control structures, is an

important determinant of institutional autonomy. The university as an organization or institution in itself provides boundaries to its autonomy. In addition, internal organizational design is often directly related to external political, social, and economic pressures.

Philosophers of higher education have long recognized the tension between the idea of the university and the institutional form within which that idea is realized: “The institution is simultaneously indispensable and a standing threat to the idea of the university”, as Jaspers (1960: 51) pointed out shortly after World War II. Half a century later, Marginson and Considine (2000: 8) make a similar observation: “It has become all too easy to imagine the university without one or another of the academic disciplines; it is impossible to imagine the academic disciplines without their institutional setting, and its governance.” Alternatively, as Jaspers (1960: 83) puts it: “The university exists only to the extent that it is institutionalized. The idea becomes concrete in the institution . . . Yet ‘institution’ necessarily implies compromises. The idea is never perfectly realized. Because of this a permanent state of tension exists at the university between the idea and the shortcomings of the institutional and corporate reality.”

The realities of the external environment in which the modern university operates have changed, some would say transformed utterly, its present internal organizational form. In many countries, market-steering at the sector level has, in turn, changed the way in which universities are managed, decisions made, and power distributed within the organization.

There is one school of thought that considers management principles as universal: good management practice can be transferred from organization to organization despite the line of business of a particular organization. Coupled with this is the belief that the university – like any large organization – requires corporate style management, strong leadership, and clear lines of authority. Others, however, argue the unique features and professional character of higher education institutions – higher education institutions are professional organizations, and in all professional organizations there is an inherent tension between professional autonomy and management prerogative, but with the balance of power leaning toward the professional community.

According to Clark (1983), academic groups and disciplines in competition with one another create boundaries between themselves in order to avoid direct conflict. However, the current emphasis on market competition and corporate management coupled with increasing demands for accountability and economic relevance are making professional group insulation increasingly difficult, with some claiming that we are now experiencing a basic re-norming of the academic profession (Slaughter and Leslie, 1997). The movement from a collegial to a managerial approach in higher education is challenging

traditional academic professional loyalties. Trow (1983: 122) wrote, in the early 1980s, that “the chief protection of the autonomy of the universities lies in the expertise of their staff . . . [A]cademics’ monopoly on expertise and their need for freedom to direct their own advanced instruction and research, place limits respectively on governmental control . . . and the bureaucratisation of the organization of university work.” A quarter-century later, academe’s monopoly of knowledge and complete control over setting the research and teaching agenda are severely contested (Gibbons, 1998). In the neoliberal corporate university, strong management – rather than academic excellence – is the “lynchpin of university ‘performance’ and competitive position” (Mollis and Marginson, 2002: 321).

Despite policy statements at various political levels to the contrary, there have been a number of reports indicating that many in the academic community believe that higher education institutions have lost autonomy and the freedom to determine their own directions (Meek, 2003). In several jurisdictions, this has created a strong them/us attitude between universities and government and severely eroded academic morale; in addition, within institutions, there are serious tensions among management levels as to how higher education institutions should be run. “Presidents and rectors in many countries . . . now find themselves engaged in extensive negotiations with ministries and governments over issues of internal governance, managerial flexibility, and authority for program approval” (Dill, 2001: 22).

The political context in which universities operate has always moderated the degree of institutional autonomy. Furthermore, the way in which the political context prescribes boundaries to institutional autonomy, in turn, determines the way in which the institution determines the autonomy of its members. The market steering of higher education and the competitive environment in which institutions must operate have strengthened the role of the executive and central management to the detriment, some would argue, of academic freedom. Tapper and Salter (1995: 59), for example, state that “the link between institutional and individual autonomy within the British university systems has been broken. A decline in the autonomy of the dons has been matched by an actual enhancement of the autonomy of the universities as institutions.”

The movement of the university away from strict state bureaucratic control toward enhanced institutional autonomy is a characteristic of many national higher education systems. Even in countries with an historical predilection toward strong university autonomy, there is a concerted effort on the part of the state to strengthen governance and management at the institutional level, possibly with an implied assumption that it is the job of management to bring recalcitrant academics into line.

Several higher education institutions themselves have called for greater autonomy from state control in order to more effectively compete, be more entrepreneurial, and develop a competitive culture.

Several factors have led to greater market-like competition among universities, for example, the presumption of the global ascendancy of capitalism following the end of the Cold War; the movement from the welfare state to the competitive state with an emphasis on maximizing returns on investment more so than general welfare; government as service purchaser rather than service provider; movement of the state from manager of public values to manager of scarce resources; and the ascent of neoliberal ideology and its emphasis on lower taxes, smaller government, and the presumption that higher education is more of a private, than a public, good.

The shift of higher education from a public to a consumer good has several consequences for institutional autonomy. A shift toward consumerism and the student as client, for example, has not only enhanced the notion of user-pays, but has given the client in a consumer's market a good deal of control over what they pay for. The greater orientation to student as consumer and education as product has changed the balance of power within many universities. Theoretically, consumers purchase what they want and are always right. It has been argued that, in the USA (Geiger, 2004), through student consumerism, power has shifted from universities to students, and the university has largely lost control over the design of the undergraduate curriculum. In a market driven by student demand, student choice coordinates the curriculum. This relationship significantly detracts from student learning and limits the autonomy of the university to provide what is, in the opinion of its professional staff, the most appropriate educational product: "The market power of students has consistently shifted the balance in the direction of 'customers'" (Geiger, 2004: 246).

There appears to be a paradox embedded in the relationship between market coordination and institutional autonomy: while universities with sufficient autonomy from the state are best able to adapt to national markets through enhancing quality and competitiveness, responding to these incentives may detract from providing direct service to state citizens (Newman *et al.*, 2004). "The marketplace has . . . brought universities greater resources, better students, a far larger capacity for advancing knowledge, and a more productive role in the . . . economy. At the same time, it has diminished the sovereignty of universities over their own activities, weakened their mission of serving the public, and created through growing commercial entanglements at least the potential for undermining their privileged role as disinterested arbiters of knowledge" (Geiger, 2004: 265).

Another paradox in the autonomy debate is that, with the enhancement of institutional autonomy and the

corresponding loosening of strict state control, the rights of individual scholars to set their own individual research and teaching agendas free from outside direction appear to have diminished. "In a remarkable accomplishment of living with contradiction, the university . . . has for an extended part of its history been able to reconcile [a] firm insistence on autonomy with the persistence of an institutional relationship between the university and the state that was both utterly non-autonomous and characterized by more or less total dependence on both the regulatory and providential tutelage of the state over the university – with the one notable exception, of course, of not directly interfering with the individual professor's 'autonomy'" (Weiler, 2003: 7). A more autonomous relationship between the state and the university has strengthened institutional autonomy, self-governance, and self-determination in the belief that this will make the institution more effective in setting goals and strategic directions and in mobilizing resources. The right to decide basic institutional objectives, and how their achievement will be resourced, has seemingly shifted from academic professionals to institutional management: "Institutional autonomy competes with individual autonomy" (Weiler, 2003: 8).

It seems that the twenty-first century Service University has shifted its orientation from social knowledge to market knowledge and that the "development of a market oriented university supersedes academic decision making" (Buchbinder, 1993: 335). According to Newson (1993: 298): "These new forms of decision making fundamentally undermine a conception of the university as an autonomous, self-directing, peer-review and professional-authority based institution, and thus changes the politics of how academic work is accomplished."

Conclusion

Clearly, the boundaries of institutional autonomy in higher education have become more porous. However, academics and the institutions they inhabit are far from being totally directed by the forces of managerialism or the market. Maton (2005: 699), drawing on Bourdieu's field theory, makes a distinction between positional autonomy – power distribution between positions within the field of higher education and those external to it – and relational autonomy: relative control over work practices, aims, measures of achievement, etc., *vis-a-vis* external influences. Relational autonomy has been the most weakened of the two. "The notion that social and economic goals are served by enabling universities to create and transmit knowledge as ends in themselves has been replaced within policy discourse by an utilitarian view of higher education as an instrument for achieving politically desirable outcomes" (Maton, 2005: 699). However, with respect to positional autonomy, "the agents responsible

for enacting these principles remain overwhelmingly located within higher education.” In addition, there is evidence to suggest (Meek, 2003) that academe is more adroit, than often assumed, in turning policy discourses concerning external accountabilities and social and economic relevance to its own ends. The autonomy of the university was never absolute, and its boundaries will remain contested.

See also: Academic Freedom.

Bibliography

- Arimoto, A. (2006). National research policy and higher education reforms in Japan. In Meek, V. L. and Suwanwela, C. (eds.) *Higher Education, Research, and Knowledge in the Asia-Pacific Region*, pp 153–174. New York: Palgrave-MacMillan.
- Ashby, E. (1966). *Universities: British, Indian, African*. London: Weidenfield and Nicolson.
- Berdahl, R. (1988). Academic freedom, autonomy and accountability in British universities. *Paper Prepared for the Conference of the Society for Research into Higher Education*, University of Surrey, 19–21 December.
- Buchbinder, H. (1993). The market oriented university and the changing role of knowledge. *Higher Education* 26(3), 331–347.
- Clark, B. (1983). *The Higher Education System: Academic Organization in Cross-National Perspective*. Berkeley, CA: University of California Press.
- Dill, D. (2001). The regulation of public research universities: Changes in academic competition and implications for university autonomy and accountability. *Higher Education Policy* 14(1), 21–35.
- Geiger, R. (2004). *Knowledge and Money: Research Universities and the Paradox of the Marketplace*. Stanford, CA: Stanford University Press.
- Gibbons, M. (1998). Higher education relevance in the 21st century. *Paper Prepared for the UNESCO World Conference on Higher Education*, Paris: UNESCO, 5–9 October.
- IAU (International Association of Universities) (2006). Background. IAU Presidents' Symposium, "Institutional Autonomy Revisited: National Dimensions, Cross Regional Experiences" Chiang Mai, Thailand, 8–9 December. http://www.unesco.org/iau/conferences/chiang_mai/pdf/background_doc.pdf (accessed August 2009).
- Jaspers, K. (1960). *The Idea of the University* (translated from the German *Die Idee der Universität* 1946). London: Beacon Press.
- Kayrooz, C. (2006). The role of autonomy in social science research: The view from UK and Australian universities. *Paper Presented at the Colloquium on Research and Higher Education Policy, "Universities as Centres of Research and Knowledge Creation: An Endangered Species?"* Paris: UNESCO, 29 November – 1 December.
- Kenway, J., Bigum, C., Fitzclarence, L., and Collier, J. (1993). Marketing education in the 1990s. *Australian Universities' Review* 36(2), 2–6.
- Kogan, M. (1984). The political view. In Clark, B. R. (ed.) *Perspectives on Higher Education*, pp 56–75. Berkeley, CA: University of California Press.
- Marginson, S. and Considine, M. (2000). *The Enterprise University*. Melbourne: Cambridge University Press.
- Maton, K. (2005). A question of autonomy: Bourdieu's field approach and higher education. *Journal of Education Policy* 20(6), 687–704.
- McElwee, G. (1992). How useful are performance indicators in the polytechnic sector? *Educational Management and Administration* 20(3), 189–192.
- Meek, V. L. (2003). Introduction. In Amaral, A., Meek, V. L., and Larsen, I. (eds.) *The Higher Education Managerial Revolution?* pp 1–30. Dordrecht: Kluwer.
- Mollis, M. and Marginson, S. (2002). The assessment of universities in Argentina and Australia: Between autonomy and heteronomy. *Higher Education* 43(3), 311–330.
- Neave, G. (1996). Homogenization, integration and convergence: The Cheshire cats of higher education analysis. In Meek, V. L., Goedegebuure, L., Kivinen, O., and Rinne, R. (eds.) *The Mockers and Mocked: Comparative Perspectives on Differentiation, Convergence and Diversity in Higher Education*, pp 26–41. Oxford: Pergamon.
- Newman, F., Couturier, L., and Scurry, J. (2004). *The Future of Higher Education: Rhetoric, Reality, and the Risks of the Market*. San Francisco, CA: Jossey-Bass.
- Newton, J. (1993). Constructing the 'post-industrial' university: Institutional budgeting and university corporate links. In Altbach, P. and Johnstone, B. (eds.) *The Funding of Higher Education: International Perspectives*, pp 285–304. New York: Garland Publishing.
- Slaughter, S. and Leslie, L. (1997). *Academic Capitalism: Politics, Policies and the Entrepreneurial University*. Baltimore, MD: Johns Hopkins University Press.
- Tapper, E. and Salter, B. (1995). The changing idea of university autonomy. *Studies in Higher Education* 20(1), 59–71.
- Trow, M. (1983). Defining the issues in university–government relations: An international perspective. *Studies in Higher Education* 8(2), 115–128.
- Trow, M. (2005). *An American Perspective on British Higher Education: The Decline of Diversity, Autonomy and Trust in Post-war British Higher Education*. Berkeley: University of California eScholarship Repository. <http://repositories.cdlib.org/igs/WP2005-3> (accessed August 2009).
- Van Vught, F. (1989). *Government Strategies and Innovation in Higher Education*. London: Jessica Kingsley.
- Weiler, H. (2003). Ambivalence and the politics of knowledge: The struggle for change in German higher education. *Paper Prepared for a SCANCOR Conference on "Universities and the Production of Knowledge"*. Stanford University, 25–26 April.
- Zgaga, P. (1997). *Autonomia not autarkia*. In Wolff, K. (ed.) *Autonomy and External Control*, pp 111–126. München: ludicum.

Further Reading

- Amos, S. K., Keiner, E., Proske, M., and Radtke, F. (2002). Globalisation: Autonomy of education under siege? Shifting boundaries between politics, economy and education. *European Educational Research Journal* 1(1), 193–213.
- Beck, J. (1999). Makeover or takeover? The strange death of educational autonomy in neo-liberal England. *British Journal of Sociology of Education* 20(2), 223–238.
- Kayrooz, C., Akerlind, G., and Tight, M. (eds.) (2007). *Autonomy in Social Science Research: The View from UK and Australian Universities*. London: Elsevier.
- Neave, G. (2001). On being economical with university autonomy: Being an account of the retrospective joys of a written constitution. *Higher Education Policy* 14(1), 1–5.
- Wilson, B. (1992). Studies in the conceptualisation of a university. *Studies in Higher Education* 17(3), 295–303.

Relevant Websites

- <http://www.mihelm.unimelb.edu.au/> – LH Martin Institute for Higher Education Leadership and Management.
- <http://www.utwente.nl/cheps> – Centre for the Study of Higher Education Policy Studies, University of Twente.
- <http://cshe.berkeley.edu> – Center for Studies in Higher Education, University of California at Berkeley.
- <http://www.unesco.org/iau/> – International Association of Universities.
- <http://www.unesco.org/en/higher-education> – UNESCO Forum on Higher Education, Research and Knowledge.

Diversity of Higher Education

U Teichler, University of Kassel, Kassel, Germany

© 2010 Elsevier Ltd. All rights reserved.

Major Distinctions

The key activities of higher education – teaching and learning, research, and, possibly, service – are undertaken in diverse institutional settings. The settings vary according to the tasks and functions; the composition as well as the level and substance of activities; the stages of study programs, etc.

In referring to the macro-level of society, the term higher education system is employed as a rule for all higher education institutions within a country. This is because higher education, in spite of the universalistic nature of some disciplines, international cooperation and exchange, and cosmopolitan attitudes of many scholars, is strongly shaped by modes of supervision and funding, organizational rules, curricular practices, etc. usually determined within individual countries.

The most obvious single element of diversity is that of disciplines. Theories, methods, and areas of knowledge develop their specific territories of discourse, and institutions of higher education, as a rule, are subdivided into disciplinary units.

The major debate on diversity, however, addresses variations between institutional segments in which teaching, learning, and research are undertaken – it might be called the shape, pattern, or structure of the higher education system, whereby attention is paid notably to institutional and program types, levels of study programs, or levels and profiles of the individual institutions of higher education or the individual departments. Some studies, though, include other dimensions as well, such as organizational differences (e.g., Birnbaum, 1983; Huisman, 1995), the financial regime of the institutions (notably public vs. private higher education), and the modes of educational delivery (such as distance education).

The debate on diversity is popular and controversial, because policies in this domain are widely viewed as being very powerful: the shape and size of the system, though continuously changing incrementally, are often regulated by external actors, for example, governments and parliaments, and these externally driven regulations might have a strong influence on the internal activities and thus on the services higher education actually provides to society.

Classifications

Descriptions of higher education systems employ a bewildering diversity of concepts, terms, and dimensions. This

even holds true for the term higher education. Until about 1950, the term university was mostly employed. The term higher education became popular notably, when politicians began to point out that other institutions, characterized by a predominant teaching function, have much in common with universities. Since about the 1980s, the term tertiary education gained popularity by underscoring that all education for students beyond the typical age of secondary education can be viewed as sharing somewhat of a common function, though the level of intellectual ambition and the link to scientific knowledge might vary.

Higher education systems are described concretely (see Teichler, 2007a) in most cases according to:

- types of higher education institutions or types of programs,
- levels of programs, and
- variations in reputation and prestige, and, possibly, according to the substantive profiles of institutions and study programs.

In characterizing the differences named, most actors and observers:

- sort differences between institutional settings to a varying extent as vertical (in terms of levels of quality, reputation, selectivity, etc.) and horizontal (in terms of substantive profiles, conceptual schools, etc.);
- refer to a varying extent about formal elements employed in laws, other regulations, and any official descriptions (e.g., institutional types, levels of programs, official functions of study programs) and informal elements (e.g., profiles and reputation);
- make a distinction between diversity within higher education institutions (internal, intra-institutional, or program diversity) or between higher education institutions (external, inter-institutional, or intra-institutional diversity).

Often, these terms aimed to describe the overall character of the higher education system according to the kinds and degrees of variety of the institutional setting. A higher education system might be characterized as:

- unitary, if one institutional type clearly dominates: in the past, Italy was often taken as an example of a system where more or less all institutions were universities and most study programs were long;
- binary, if two major types of higher education are seen as characteristic, for example, polytechnics, colleges of advanced education, *Fachhochschulen*, etc.;

- multilevel, if the type of system can be sorted better according to the levels of the study programs than according to the types of institutions of higher education and according to institutional types. For example, the French system was often classified according to programs of the first 2 years of study (baccalaureate +2; e.g., vocational programs at *écoles*, the first half of university programs, and preparatory classes for *grandes écoles*), programs leading eventually to the university degrees and degrees of the *grandes écoles*, and finally advanced programs (baccalaureate + 6 and more); and
- comprehensive systems, if diversity is predominantly taken care of intra-institutional differences rather than through differences between institutions: This was envisaged in the 1970s but eventually never fully realized through the establishment of a few *Gesamthochschulen* in Germany and through the incorporation of some higher education institutions with short study programs into universities in Sweden.

In some countries, specific classifications are undertaken to characterize their higher education system. For example, the Carnegie classification in the US groups institutions of higher education notably according to their involvement in research and doctoral training: research universities, comprehensive universities (with selective areas of high-quality research and doctoral programs), 4-year colleges, etc.

Finally, some classifications characterize the functions of institutional diversification. The US higher education researcher Martin Trow (1974) coined the words elite higher education, mass higher education, and universal higher education. His widely known concept often is misunderstood as directly depicting historical stages of the overall higher education system (mass higher education substituting elite higher education). According to Trow, however, elite higher education is supplemented by mass higher education when about 15% of the corresponding age group enrol, and eventually by universal higher education, when about 50% enrol. Mass higher education serves best those students additionally enrolling in higher education institutions and is at the most, moderately involved in research, thereby protecting elite higher education which serves a select student population and is strongly involved in research. This functional diversification, according to Trow, does not necessarily correspond to an institutional diversification.

Types of Higher Education Institutions and Study Programs

In the early 1950s, about 10% of the corresponding age group were enrolled at institutions of higher education

in the US, while the enrolment rate was about 3–5% in most other economically advanced countries. Thereafter, the view spread soon that a growing proportion of highly educated persons in the labor force contributes substantially to economic growth. In the 1960s, educational expansion was increasingly interpreted as serving both economic growth and the reduction of unequal opportunity in access to advanced levels of education. Moreover, educational systems had become more open in general, and a rising percentage of the population wanted to enrol.

In the wake of increasing enrolment rates, the view spread in Europe that a substantial expansion was not feasible in the relatively costly setting of universities characterized by a close link between teaching and research and by relatively long study programs. Furthermore, the notion dominating in the US that the growing number of students was bound to be more diverse in terms of motivations, academic talents, and future job prospects, and would be served best by a diversification of higher education was accepted in Europe as well. However, in contrast to the US, clear distinctions between types of higher education institutions and possibly between the overall duration of study programs became the most popular approach in Europe from the 1960s onward for a considerable period (see Teichler, 1988). Three early national reforms are named most frequently in this context:

- Polytechnics were established in the United Kingdom in the early 1960s. They required in principle the same entry qualifications as universities and had the same types of degree programs, but their level of academic ambition was more moderate and many of the programs had a stronger vocational emphasis. Polytechnics have striven more strongly to become similar to universities than to strengthen their specific profile, and they eventually were named universities in 1992.
- During the 1960s as well, *Instituts Universitaire de Technologie* (IUTs) were established in France as specific units within universities offering 2-year programs which turned out to be more select and demanding than the existing programs in post-secondary *écoles*, but clearly were more applied than university programs. Various IUT programs were extended to 3 years when a convergent system of study programs and degrees was advocated in Europe in the late 1990s.
- Since 1970, *Fachhochschulen* were established in the Federal Republic of Germany as a second institutional type. One year less prior schooling than at universities was the required entry quality qualification, possibly including vocational training. The study programs were slightly shorter and more applied than at universities. Since about 2000, these institutions (recently called in English universities of applied sciences) began to introduce a bachelor and master structure, whereby

majority of master programs have an applied emphasis, while some aim to be similarly theoretical as those at universities.

Efforts, for example by the Organization for Economic Co-Operation and Development (OECD), to formulate generally acceptable terms for these new institutions and programs were not successful. Nonuniversity higher education was viewed as too prerogative to that sector. Short-cycle higher education was not applicable in all cases, and alternatives to universities (OECD, 1991) were viewed as too general. Eventually, the international organizations used abstract terms such as tertiary type A and tertiary type B. The three most often named characteristics of the second type of higher education institutions (see Teichler, 1988) do not apply in all cases:

- The second type was expected to concentrate on teaching and learning. The teaching load of academics was substantially higher than that of those at universities. However, applied research was seen as an important function in some countries.
- The study programs were in most cases shorter, but not consistently across all countries.
- The study programs had an applied curricular emphasis, but again, this distinction was more strongly accentuated in some countries than in others.

Views differed in Europe whether the sectors should be primarily differentiated according to types of higher education institutions or according to types of study programs. In some instances, varied types of programs under a single institutional roof were advocated as superior because they could offer students a more open choice during the course of study and because one could expect a cross-fertilization of the various approaches.

Actually, the second type of higher education institutions constantly tried to redefine its role through partially underscoring its specifics and concurrently by trying to move closer to the universities in various respects (teaching load, teachers' salaries, research function, etc.).

Almost silently, as far as the overtones of higher education policy debates are concerned, a third type of institutions and study programs emerged. According to OECD statistics, the entry rates to tertiary education increased in economically advanced countries from less than 40% on average in the 1990s to more than 60% in the middle of the first decade of the twenty-first century. This includes in some countries up to about one-third of persons enrolled in tertiary programs not being considered higher education. These programs require in some countries the same number of years of prior schooling as university programs, in some countries somewhat less of prior schooling, and in some others, these are advanced vocational training programs, often following a first

vocational training program, for example, technician training programs subsequent to, for example, car mechanic training programs.

Levels of Study Programs

Intra-institutional diversity according to teaching and learning, though, had a long tradition in a substantial number of countries as regards levels of study programs. In contrast to countries such as Germany, the Netherlands, or Italy, where long university programs existed, Anglo-Saxon countries had a second level of study programs, that is, master programs subsequent to bachelor programs. Northern European countries had upper *kandidat* programs subsequent to lower *kandidat* programs, and France had the *maîtrise* subsequent to the licence in some fields of study. These stage systems of study programs and degrees had different functions in the various countries: In France and Northern European countries, one could leave the system earlier if one wanted to become a teacher outside the sector of the academic secondary schools. In the United Kingdom and Japan, the first degree was generally viewed until recently as the normal entry qualification to all high-level occupations. In the US, the masters or professional degrees gradually became the entry qualification for high-level careers while the bachelor opened up only the second-ranking career avenues.

In the late 1990s, ministers in charge of higher education – initially from four countries in the Sorbonne Declaration in 1998 and thereafter from 29 countries in the Bologna Declaration of 1999 – called for the establishment of a convergent stage system of study programs and degrees in Europe within about a decade. In fact, many institutions of higher education established bachelor programs – often for a period of 3 years, but some up to 4 years – and subsequent master programs – most often for a period of 2 years, but some for 1 or 1.5 years. Universities were called to arrange the bachelor programs in a way that they turn out to be a meaningful entry qualification for employment rather than a *de facto* interim qualification for studying up to a masters program.

As a consequence, stages of study programs and degrees certainly became the single most important formal dimension of diversification more or less all over the world. It remained an open question, though, whether diversification by types of higher education institutions would persist thus, underscoring horizontal differences by the substance of the study programs and the competences striven for, or even be strengthened, or whether diversity by institutional type of higher education would gradually erode as a consequence of the increased emphasis on levels.

The various official policy documents published in the so-called Bologna Process do not address explicitly the

issue whether efforts to establish a convergent system of stages of study programs and degrees were expected to have any implication for horizontal diversity according to profiles of study programs. These documents only point out that the policy in favor of structural convergence should not exert pressure for curricular convergence among national higher education systems.

Reputational Diversification and Rankings

Higher education systems traditionally differed substantially according to the extent of informal vertical diversity. In the US and Japan, the range of quality and reputation between the universities at the top and at the bottom was always extreme as compared to European countries. Thereby, research, graduate education, and to some extent the quality of entering undergraduate students were seen as the most salient dimensions of vertical distinction in the US. In Japan, in contrast, most emphasis was placed on the entry level of students and the careers of the graduates. In the UK, a few universities were viewed as clearly excelling both in research and in the quality of study programs. In France, the top *grandes écoles* were viewed as far superior to the other institutions as far as graduate careers were concerned, while hardly any research was undertaken traditionally at these institutions.

Other higher education systems were characterized by some degree of hierarchy according to institutional reputation, though clearly flatter than the above-named countries, for example, Australia, Canada, or Sweden. Finally, Germany is generally viewed by experts as a large economically advanced country with the flattest reputational hierarchy among universities.

Since about the late 1970s, we note in Europe a gradually rising popularity of a steeper stratification of the higher education system. The belief began to spread among experts that top quality in research was achieved more easily through a higher degree of concentration of resources and top scholars in a small number of institutions. Since the 1990s, the *Zeitgeist* clearly moved toward a stronger popularity of steep differences between higher education institutions as far as academic reputation is concerned. In various countries, the funding system and other steering mechanisms were rearranged in order to facilitate a steeper vertical diversification among the universities. Moreover, international ranking lists of world-class universities gained popularity.

The advocates of a fierce race for ranks usually claim that economic success of a country increasingly becomes dependent on cutting-edge breakthroughs in knowledge, which can be fostered by financial investments into the apex of the higher education system. Moreover, ranking lists are viewed as both contributing to competition

beneficial for quality enhancements and to transparency essential for rational and quality-reinforcing choices (see various arguments in Sadlak and Liu, 2007).

The critics, first, question the validity of indicators employed. Second, they question the rationales of the stratification policies, for example, that success in learning depends on homogenous environments and that the quality of work of individual scholars and research institutions largely depends on the university as a whole. Third, they point out negative side effects, for example, increasing advertising activities of higher education institutions, data faking, and efforts to improve in terms of indicators instead of quality as such; reinforcement of past prestige hierarchies thus support labelism and credentialism rather than meritocratic reward, discourage those who are not the most successful ones in the race, and undermine horizontal diversity according to profiles of research approaches and study programs (see the arguments collected in Teichler, 2007a).

Major Explanatory Concepts

The public discourse as well as the literature in this domain do not merely address the configuration of the system, but also try to explain its causes. Three documents show the variety of explanations.

The organizers of an expert conference on diversity and convergence in higher education held in the early 1990s identified three major theoretical perspectives on diversity (Meek *et al.* 1996):

- *The internal perspective.* According to Burton Clark, “it is the academic discipline engine that invariably drives higher education institutions and systems to differentiation.” The “uncontrolled drive towards ever-increasing disciplinary specialisation is not always obvious; it often results in incremental or ... ‘unannounced cumulative change.’ But it is these changes ... that ultimately shapes higher education institutions and systems, making them evermore complex in an environment where operational diversification is far more important than nominal integration” (Meek *et al.*, 1996: 207). The authors argue that, according to Clark, “subject partitioning, program affiliation, subject dignification and subject dispersion” contribute to increasing diversity (Meek *et al.*, 1996: 213).
- *The systemic perspective.* According to Guy Neave, patterns of the higher education system are strongly affected by actors, notably those at national and supra-national levels. Neave names various “forces that work for and against homogenization or integrity”, and “no higher education institution or system moves inevitably towards either homogenization or diversification” (Meek *et al.*, 1996: 207). Neave does not only name

contradicting forces, but also names ambivalent realities rather than a clear trend toward a single direction: “whether a particular system is ‘diverging or converging is largely a function of where we focus our attention’” (Meek *et al.*, 1996: 208).

- *The environment perspective.* According to Frans van Vught, higher education institutions are “located within a supra-system consisting of the social, political and economic environment” (Meek *et al.*, 1996: 209, 210). Institutions of higher education “constantly survey the environment to identify opportunities and risks with respects to obtaining the resources . . . Those institutions that ‘read’ the environment correctly survive, those who do not perish” (Meek *et al.*, 1996: 210). In principle, a varied environment, thus, leads to increased diversity, while an isomorphic and uniform environment encourages a decrease of diversity. Van Vught views national governments as strong actors mostly opting for homogenous systems or types of higher education and thus contributing to a low degree of diversity.

According to the authors of the above study, the various countries analyzed “are themselves substantially divergent in terms of national approaches to diversity” (Meek *et al.*, 1996: 234). The various theoretical efforts help to “provide a better understanding of the complexity surrounding questions of diversity and convergence in higher education” (Meek *et al.*, 1996), but there might be more questions raised than answered.

In a synthesis of debates and analyses from the 1970s to the early 1990s, Ulrich Teichler (2007b: 124–125) points out that various scholars have opted for various developmental theories in their efforts to explain the dynamics of structural change in higher education.

- “The ‘*expansion and diversification*’ theory . . . had the strongest impact on the public debate. Accordingly, expansion of higher education creates pressures for diversification because the needs of the learners and other potential users become more varied and because, as many actors believe, these varied needs might be more readily met through a certain ‘division of labour’ among institutions.
- A second type of theory, ‘*drift*’ theories, also became very popular. Types of higher education institutions are not necessarily very faithful in pursuing the goals they were expected to pursue when they were initially established. According to these theories, different types of higher education institutions are eager to pursue their initial mission at most for a short period after they had been newly founded or upgraded. After some period, they begin to consider themselves as competitors to other types of higher education institutions. . . . the almost universal ‘academic drift’ . . . Similar reasoning, although with different expectations, was frequently employed in the 1980s. Many experts suggest that the tight labour market for graduates

from purely academic fields had triggered a trend of ‘vocational drift,’ ‘vocationalism’ or ‘professionalism.’

- A third type of approach might be called ‘flexibilization’ theories. . . . In contrast to the first theory, they point at weaknesses in segmented institutional types serving clearly distinct needs. Over time, soft models and broad ranges might be superior . . . Accordingly, late selection in pre-career education, permeability of educational careers, compensatory measures for the disadvantaged, soft diversified structures of higher education, and the establishment of a life-long education system contribute to a soft system in three respects: no decision in the educational career would be considered as definite, the model could satisfy both the advocates and critics of educational expansion, and it would finally facilitate rapid adaptations, if major problems occurred.
- Finally, we note ‘*cyclical*’ theories of the structural development of higher education. According to these theories, certain structural patterns and policies come and go in cycles. For example, opening up of educational avenues and a reduction of the differences between varied types of institutions and programs might be on the agenda at times when a shortage of graduates is felt, whereas segmentation and hierarchization of higher education is favored or just taking place, when fears of over-supply or ‘over-education’ dominate the scene”

In a recent analysis, Frans van Vught (2008: 153–154) argues that organizational theories are the most crucial ones in explaining policies and trends regarding diversity in higher education. He points out three theoretical perspectives:

- “The population ecology approach is based on the Darwinian evolutionary point of view. . . . (it) concentrates on the sources of variability and homogeneity of organizational forms. . . . In doing so, it pays considerable attention to population dynamics, especially the processes of competition among diverse organizations for limited resources such as membership, capital and legitimacy.”
- “The resource dependency perspective stresses the mutual processes of interaction between organizations and their environments. According to this approach, organizations on the one hand are dependent on their environments (which primarily consists of other organizations) but on the other hand these organizations are also able to influence their environments”
- “The institutional isomorphism approach stresses that in order to survive, organizations have to adapt to the existence of and pressures by other organizations in their environment. These adaptation processes tend to lead to homogenization, as organizations react more or less similarly to uniform environmental conditions. Isomorphism is a constraining process that forces organizations to resemble other organizations that face the same set of environmental conditions.”

Drawing from these concepts, van Vught (2008: 20) formulates two propositions which he considers relevant for the current developments of higher education systems: (1) the larger the uniformity of the environmental conditions of higher education organizations, the lower the level of diversity of higher education systems and (2) the larger the influence of academic norms and values in a higher education organization, the lower the level of diversity of the higher education system.

Obviously, all the conceptual frameworks named are fruitful in pointing out elements possibly contributing to the changing extent of vertical and horizontal diversity in higher education. Not a single one, however, can claim to come close to a theory of diversity in higher education.

Concluding Observations

National higher education systems cannot be completely uniform. They have to serve different motives, talents, and job perspectives of their students. Not all institutions of higher education or their subunits can serve teaching and research equally. There are quality differences, and there has to be room for a diversity of aims and paradigms in teaching and research.

The increase of student enrolment and the growth of research activities seem to reinforce the view that diversification is the best response, that is, a constant growth of variety both according to the level of quality and reputation and to substantive profiles of research and study programs. Already in the 1960s and 1970s, many experts and actors considered a highly diverse higher education system such as in the US as the modern response. However, deficits of such a model were pointed out as well; various options of diversification were chosen, for example, varied types of higher education institutions in some European countries, and incremental trends did not necessarily move toward increased diversification.

National systems of higher education seem to keep certain characteristics for a long period, but are characterized as well by a constant unrest of those sectors or institutions not at the apex of the reputational hierarchy. In highly regulated systems, we noted collective efforts of upward mobility, while in less-regulated systems individual institutions try to upgrade their status.

Obviously, the growth of student numbers and research activities seems to call for increased vertical and horizontal diversity. In contrast, competition within higher education tends to undermine horizontal diversity and reinforce vertical diversity only as a consequence of imitation behavior according to the success dimensions of the most highly reputed individual institutions or institutional types. In contrast to Martin Trow's initial expectation that elite higher education needs protection in the

process of higher education expansion, horizontal diversity and programmatic profiles other than those popular at top institutions need protection in order to secure diverse services of higher education to society.

Obviously, a certain degree of segmentation within higher education is viewed as necessary to ensure pride in and care of distinct profiles of study programs. However, segmentation is viewed as creating barriers for permeability if students want to correct their educational choices and if academics want to embark on new activities. Therefore, the political debate on the strengths and weaknesses of different models of diversity is driven by a permanent search for a soft system, that is, a creative mix of stability and flexibility.

The views about the strengths and weaknesses of certain patterns of the higher education systems depend strongly on the emphasis placed either on the research or on the teaching and learning function. By and large, steep vertical stratification of higher education is more often claimed to be beneficial for research than for teaching and learning.

In recent years, the variety of different national options as regards the pattern of their higher education systems seems to diminish. On the one hand, we note that most European countries move toward the establishment of a convergent system of study programs and degrees, thereby aiming to make the levels of study programs and degrees the single most important dimension of institutional diversity. On the other hand, institutions of higher education increasingly view themselves as competitors in the race for high quality and reputation amid trends and policies to create more steeply stratified national higher education systems or even move toward an even more steeply stratified world higher education system.

During any period of major changes and reforms of the institutional patterns of the higher education system, sweeping statements are made about the miraculous strength of certain models. In contrast, Neave (2006: 246) pointed out that in various European countries, the policies of advocating diversity primarily through varied institutional types were woefully unsubstantiated. Similarly, current efforts to reinforce competition for world-class universities and to create more steep vertical diversity are based on empirical badly founded claims and even of disinterest in asking key questions about the strengths and weaknesses of this model as compared to alternative models, that is, a lower extent of vertical diversity and a stronger element of horizontal diversity.

See also: Access and Equity in Higher Education; Bologna Process: On the way to a Common European Higher Education Area; Higher Education and the Labor Market; National and International Rankings of Higher Education; Unitary, Binary and other National Systems of Higher Education.

Bibliography

- Birnbaum, R. (1983). *Maintaining Diversity in Higher Education*. San Francisco, CA: Jossey-Bass.
- Huisman, J. (1995). *Differentiation, Diversity and Dependency in Higher Education: A Theoretical and Empirical Analysis*. Utrecht: Lemma.
- Meek, V. L., Goedegebuure, L., Kivinen, O., and Rinne, R. (1996). Conclusion. In Meek, V. L., Goedegebuure, L., Kivinen, O., and Rinne, R. (eds.) *The Mockers and Mocked*, pp 207–236. Oxford: Pergamon/IAU Press.
- Neave, G. (2006). Mass higher education system and the research university: A post-modern revival of Ockham's razor. In Neaman Institute, S. (ed.) *Transition to Mass Higher Education: International Comparisons and Perspectives*, pp 241–264. Haifa: S. Neaman Press.
- OECD (1991). *Alternatives to Universities*. Paris: OECD.
- Sadlak, J. and Liu, N. C. (2007). *The World-class University and Rankings: Aiming beyond Status*. Bucharest: UNESCO-CEPES and Presa Universitara Clujeana.
- Teichler, U. (1988). *Changing Patterns of Higher Education Systems*. London: Jessica Kingsley.
- Teichler, U. (2007a). Changing patterns of the higher education systems in Europe and the future tasks of higher education research. In European Science Foundation *Higher Education Looking Forward: Relations between Higher Education and Society*, pp 79–103. Strasbourg: ESF.
- Teichler, U. (2007b). *Higher Education Systems: Conceptual Frameworks, Comparative Perspectives, Empirical Findings*. Rotterdam, Taipei: Sense Publishers.
- Trow, M. (1974). Problems in the transition from elite to mass higher education. In OECD (ed.) *Policies for Higher Education*, pp 51–101. Paris: OECD.
- van Vught, F. (2008). Mission diversity and reputation in higher education. *Higher Education Policy* 21(2), 151–174.

Further Reading

- Altbach, P. G. (2004). The costs and benefits of world-class universities. *Academe* 90(1), 20–23.
- Altbach, P. G. (2006). The dilemmas of ranking. *Higher Education Digest* (CHERI, London) 55, 9–10.
- Clark, B. R. (1996). Diversification of higher education, viability and change. In Meek, V. L., Goedegebuure, L., Kivinen, O., and Rinne, R. (eds.) *The Mockers and Mocked*, pp 17–25. Oxford: Pergamon/IAU Press.
- Clark, B. R. and Neave, G. (eds.) (1992). *The Encyclopaedia of Higher Education*. Oxford: Pergamon.
- De Moor, R. A. (1979). *Changing Tertiary Education in Modern European Society*. Strasbourg: Council of Europe.
- Dill, D. D. and Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher Education* 49(4), 495–533.
- OECD (ed.) (1973). *Short-cycle Higher Education: A Search for Identity*. Paris: OECD.
- Scott, P. (1996). Unified and binary systems of higher education in Europe. In Burgen, A. (ed.) *Goals and Purposes of Higher Education in the 21st Century*, pp 37–54. Bristol, PA: Jessica Kingsley.
- Teichler, U. (2009). *Higher Education and the World of Work*. Rotterdam, Taipei: Sense Publishers.
- van Vught, F. (1996). Isomorphism in higher education? Toward a theory of differentiation and diversity in higher education systems. In Meek, V. L., Goedegebuure, L., Kivinen, O., and Rinne, R. (eds.) *The Mockers and Mocked*, pp 42–58. Oxford: Pergamon/IAU Press.
- van Vught, F., van der Wende, M., and Westerheijden, D. (2002). Globalisation and internationalisation: Policy agendas compared. In Enders, J. and Fulton, O. (eds.) *Higher Education in a Globalising World*, pp 103–121. Dordrecht: Kluwer.

Employability of University Graduates and Graduate Outcomes

J Brennan, The Open University, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Graduate employment is a subject of frequent political and research attention these days. This is because presumed high levels of employability of graduates provides a large part of the justification for the massive expansion of higher education which has occurred in most developed and many developing countries over the last 20 years or so. This justification is twofold. First, the promise of well-paid employment fuels the social demand for higher education. Young people and their parents see higher education as the essential route toward the achievement or the maintenance of a good social position and lifestyle. Second, governments see the expansion of higher education and the consequent increase in the numbers of graduates in the labor market as essential prerequisites for a successful economy. Hence, state investment in higher education becomes a public good with high rates of social return. Claims in recent years about the arrival of the knowledge economy with its high-level skills requirements have further accentuated the attention given to the second of these justifications and government targets for numbers of graduates in the labor force have been on the rise again.

The two justifications rest on somewhat different assumptions about the importance of higher education qualifications in the labor market. The social demand justification requires only that there is a positional advantage gained from possession of a degree. All that is required is evidence that employers prefer – for whatever reasons – to recruit and reward graduates more generously than people with lesser or no qualifications. However, the justification for greater state investment requires the additional assumption that graduates are recruited by employers because they will bring greater productivity to the enterprise and therefore to the economy and hence greater wealth to the nation. The greater productivity may be the result of the knowledge and skills acquired within higher education – the human capital argument; or it may be the result of the selectiveness of higher education institutions in identifying able young people – the screening argument. Or there may be no productivity at all with selection criteria for higher education entrance reflecting personal, social, and educational characteristics of little relevance to the workplace. The latter would still provide positional advantage to the graduates in terms of getting a job, but it would not bring the economic growth associated with greater productivity in doing the job.

In expanded systems of higher education, with in many cases well over 30% of the age group entering higher education, the selection or screening hypothesis becomes less plausible. Graduates can no longer be considered to constitute an elite. Thus, their greater employability must be seen to rest in the knowledge, skills, and competences they have acquired through higher education rather than through some set of innate qualities.

Yet, the elite conceptions of graduates die hard. Much research has been commissioned to determine whether graduates are getting jobs at all and, if so, whether these jobs are of a suitable graduate level. Since expansion of higher education implies that graduates will be entering new areas of the labor market, it becomes difficult to define graduate level and several typologies have been produced to represent different levels of employment, with some clearly regarded as being superior to others.

The next section of this article summarizes what we know about the kind of employment obtained by graduates. The following section considers the effects of increasingly differentiated higher education systems and new kinds of relationship between them and the labor market. The subsequent section considers the division of labor between higher education and employers in the development of employment-related knowledge and skills. The final section considers the effects of globalization, features of the knowledge society, and the need for education and training throughout the life course.

Graduates in Employment

Notwithstanding the concerns and scares about graduate unemployment that periodically surface in most countries from time to time, the evidence is pretty clear and consistent that graduates are more likely to be in work than nongraduates and are likely to earn more. Recent data published by the Organization for Economic Co-Operation and Development (OECD) show, for example, that 89% of male graduates and 79% of female graduates aged between 25 and 64 years participate in the labor force compared with the total participation rates of 82% men and 63% women. Unemployment rates among men are 3.5% for graduates compared with 5.7% overall, and for women 4.3% compared with 6.8%. Graduate earnings compared with people with upper secondary qualifications are typically 50% greater, the highest benefits apparently

going to British women (a 98% advantage compared with upper secondary education) and American men (an 88% advantage) (OECD, 2004).

A few years ago, the author summarized some of the main features of graduate employment as follows: graduates

- were unlikely to experience long-term unemployment;
- were likely to earn substantially more than people with an upper secondary education;
- were likely to experience high levels of job satisfaction and responsibility in the long term;
- were increasingly likely to experience a transitional period of several years between leaving higher education and entering “long-term” employment;
- have different experiences in the labour market according to what and where they studied, as well as according to a wide range of other educational and sociobiographical characteristics. (Brennan, 2001)

Further expansion of higher education does not appear to have changed this picture. A recent comparative study of graduates from 13 countries (Schomburg and Teichler, 2006) reported an average unemployment rate among graduates of only 4%, 4 years after graduation. Only Spain (with 13%) and France (with 10%) were significantly higher, although these latter cases do underline the point that graduate employment is not immune to the effects of general economic problems in particular countries and to changes over time. In most countries, graduate unemployment rates – though consistently low – do tend to rise and fall in line with the general economic state of the country.

Data on job quality are notoriously difficult to interpret but from the above study, we can note that 78% of employed graduates had permanent contracts of employment (again Spain was the exception with 50% of graduates on temporary contracts), that 6% were self-employed and the average annual gross income was 28.2 thousand euros. Seventy percent were in professional, managerial, and other high-level positions and a further 18% were employed as technicians and associate professional positions. Only 12% considered that they were in jobs which did not require advanced levels of educational attainment (Schomburg and Teichler, 2006: 136). Job satisfaction seemed quite reasonable with 64% of the sample indicating they were either satisfied or very satisfied with their current jobs.

Of course, aggregate figures can hide some important differences between groups of graduates. The subject of study is important in many countries, although sharp differences at the time of graduation tend to lessen with time. In the study mentioned above, the highest level of unemployment by subject was found among natural scientists (at 7%) and humanities graduates (at 6%) (Schomburg and Teichler, 2006: 82), illustrating the point

that it tends to be the graduates from nonvocational courses, irrespective of field, who are more likely to encounter difficulties in finding suitable employment. Other factors that can influence the employment prospects of particular groups of graduates include the status of the higher education institution attended, the social background, age, ethnicity and gender of the graduate, and experiences such as work placements and other employment-enhancing activities within higher education, although the effects of the latter tend not to be very great over the long term (Mason *et al.*, 2003).

All of this is indication that graduates are successful in getting jobs but not how well they are doing them. This second question – and a crucial one to the human capital argument – is less easy to answer. Some studies have seen graduate employment as largely reflecting processes of elite reproduction (e.g., Brown and Hesketh, 2004) following Bourdieu (1996). Others have indicated considerable skepticism about the claimed relationships between higher education and economic growth (e.g., Dore, 1976; Wolf, 2002).

Hard evidence of the ways in which higher education improves worker productivity is hard to come by. For example, we do not really know whether graduates are more productive than other workers, what makes some graduates more productive than other graduates, whether graduates make use of what they have learned in higher education when they enter the workplace, why employers prefer to recruit graduates – or certain types of graduates, and what is it – for example, ability, expertise, and ambition – that appears to give graduates the edge in the labor market. Answers to such questions are, at best, incomplete. They tend to be based on the perceptions of key actors – employers and graduates themselves. One can certainly point to voices of employers (though how representative one can never be sure) pointing to the deficiencies and limitations of the graduates they recruit; and surveys of graduates report a mixed picture:

Most graduates appreciated their studies and believed that learning in higher education was useful for coping with their job tasks. Yet, there was widespread critique of many aspects of higher education – certainly to a varying degree across countries and fields of study In sum, many graduates who were surveyed did not see close links between study and employment. Only 38% considered their field of study as the only one possible or the best for their area of work, and only half noted a frequent use of their specific area of knowledge. Rather, they perceived a broad range of job requirements which were to a certain extent served by the study programme and by a broad range of experiences prior to and alongside their studies. (Schomburg and Teichler, 2006: 139)

In short, most graduates get good jobs but we cannot be sure why.

Differentiation

One of the limitations of much of the research on graduate employment has been that it has dealt with aggregations, that is, it has told us about a mythical average graduate without distinguishing the effects of factors such as field of study, type of institution, age at time of study, gender, social and ethnic background, and, indeed, of many more factors. As higher education expands, it becomes increasingly differentiated, both in terms of its institutional forms and in the kinds of students it recruits. The social and educational factors interact to help determine the employment outcomes for the individual. As far as the graduate labor market itself is concerned, it is possible to distinguish between public and private enterprises, between small, medium and large employers, between different employment sectors and types of work, and to identify regional differences, all of which produce demand-side variations that we need to understand.

There have also been attempts to differentiate graduate jobs themselves. Kate Purcell and Peter Elias have elaborated a fivefold categorization of graduate jobs as follows (Purcell and Elias, 2004):

- “Traditional” graduate occupations
- “Modern” graduate occupations
- “New” graduate occupations
- “Niche” graduate occupations
- “Non”-graduate occupations

Distinctions of this sort are entirely valid but extremely time and context bound. After all, there was a time when graduate jobs were largely restricted to the clergy and running empires. They also tend to reflect different national traditions and labor-market characteristics. Such distinctions also reflect getting-a-job factors rather than doing-a-job factors. Clearly, as graduates have become more plentiful in the labor market, they have come to take a wider range of jobs. The question of the usefulness of their education to the job in question applies to all the categories. Is it helpful? Is it necessary? Is it sufficient? Even in something as self-evidently a graduate job as medicine, there is research to suggest that the most useful knowledge is gained on the job rather than at university (see Becker *et al.*, 1961).

So far, we have been skirting around one of the most fundamental why questions of all, that is, why is it that employers prefer to recruit graduates and, on the whole, elect to pay them more? Besides the screening and productivity answers, there is also the answer that employers want to get the best people, if only for reputational reasons, and are prepared to accept higher education's definition of who are the best. An expanded higher education system of course eventually fails to deliver this – it produces too many people who cannot

all be best – and so it becomes necessary to differentiate the system in status terms, to ensure the identification of an elite within a mass system.

It is in connection with the increasing differentiation of higher education, its student population, and the labor market itself that issues of employability begin to confront issues of social equity. This is another major variant of the why question. Is higher education principally about matching skills and talent to economic and employment need in order to generate economic growth (the human capital approach) or is it principally about ensuring that social classes and groups maintain their social positions and status (social reproduction approaches) or have greater and more equitable opportunities to improve their positions and status (social transformation approaches)? It may well of course do all of these things.

There are also important international differences to be borne in mind. National higher education systems differ in the extent of their vertical and horizontal differentiation, the former characterized by reputational range and the latter by functional or mission-related differences. There are also international differences in the role played by educational credentials in determining movement through the labor market. The two sorts of differences seem to be linked. Where possession of a specialist credential is a normal pre-entry requirement to a particular job, the educational system is likely to be structured to take account of functionally relevant labor-market differences. Where there is greater formal openness in the labor market – concerning who is theoretically eligible for appointment to a particular job – differentiation is likely to be more reputational and hierarchical. Broadly speaking, Anglo-Saxon-influenced systems tend to accord to the latter characteristics, while continental European-influenced systems tend toward the former. However, some convergence around the greater flexibilities and skills requirements of modern knowledge economies may also be identified.

Enhancing Employability

Higher education institutions in many countries have in recent years been making more explicit the ways in which their academic programs enhance the employability of their students. In many respects, this is in response to pressures from governments and to public expectations that investment in higher education should have real economic pay-off, for both the nation and the individual. Besides making existing activities more explicit, employability has also been the frequent object of new initiatives at both national, institutional, and program levels.

The focus on employability tends to shift attention from academic content and skills toward the more generic

and transferable things such as teamwork, communication skills, leadership, time management, initiative taking, and setting priorities. However, it can also include enhanced curriculum offerings in subjects such as languages and information technology.

Employability-enhancing initiatives vary according to whether they are optional or compulsory, institution-wide or program specific, and whether they are geared to getting a job or to doing one. Where they are optional, questions of take-up become important and the support or otherwise of academic staff can play a decisive role. Where they are compulsory, generally at the program level, whole programs may become vocationalized – geography becomes tourism, sociology becomes criminology, and English becomes creative writing. Particularly in Anglo-Saxon countries, provision of career-preparation modules within the curriculum may be designed to equip students for the job search, providing them with presentational skills with which to impress prospective employers. Such assistance is less likely to be provided in countries which have a much closer link between the subjects studied in higher education and future employment. In the latter, it is still subject knowledge and skills which are decisive for employment rather than the more generic skills.

These developments have not been without controversy, with some academics seeing them as threatening academic freedom and student choice. Acquiring the skills and dispositions necessary for working life should, from this perspective, be left until after graduation. However, the requirements of modern labor markets may call for a different approach. One American commentator notes that whereas new graduates used to have

extended time in their first position to develop interpersonal, applied reasoning, and self-management skills. . . . (but) the complex demands and pace of the new economy require college students to attain a better balance between their academic skills and people-related applied competencies before graduation. (Gardner, 1998: 60)

What may be at issue here is change in the balance of responsibilities between higher education and employers in the preparation of young people for working life.

Division of Labor between Higher Education and Employers

All jobs require some knowledge and skills but jobs differ in terms of where these knowledge and skills are acquired. Where graduates lack necessary knowledge and skills on entry to employment, it becomes the employer's responsibility to equip them. There are considerable differences in national traditions in how this division of

labor is worked out. For example, the French emphasis on professional education gives its higher education institutions – especially the *Grandes Ecoles* – a key role in professional formation, whereas this role has tended to be assumed by employers in the UK. In general, the longer first-degree courses in continental European higher education systems allow the inclusion of a greater volume of occupationally relevant preparation. The greater specificity of employment outcomes means that both students and their teachers have a much clearer picture of the kinds of employments they are destined for. Hence, greater preparation for them can be attempted. The Anglo-Saxon tradition has been more about providing an educational base of some breadth as a foundation for subsequent professional education and training often provided by employers. One hears the phrase “oven-ready” graduates used as a term of criticism for attempts to load too much of the responsibility for employment preparation onto higher education.

Transition between Higher Education and Employment

One of the reasons why there is periodic concern in many countries about graduate employment is that data are often available for only the first few months after graduation. This for many graduates is a period of transition, especially for graduates from the more nonvocational academic disciplines. Decisions about the future have to be made, job applications drafted, and, in many cases, additional qualifications secured. Initial temporary employment might be obtained as a short-term expedient, while longer-term decisions are thought through. Many graduates appear to start the job-search process rather late, perhaps putting off the evil day of commencing working life or concentrating on the successful completion of their higher education studies.

The study by Schomburg and Teichler referred to above notes that on average, graduates spend 6 months on their search for their first job after graduation. Around half had started their job search prior to graduation and, on average, about 25 contacts with employers had to be made before a final choice of job (Schomburg and Teichler, 2006: 135).

The study notes that the timing of initial regular employment after graduation varies significantly among graduates.

This reflects different traditions of transition to employment in the countries included, differences in the employability of graduates, different labour traditions according to country, economic sector and field of study and occupational area, as well as different options and strategies of graduates in the transition period. (Schomburg and Teichler, 2006: 135)

In this study, 40% were regularly employed in potentially long-term jobs immediately after graduating and a further 25% 6 months after, and a further 10% after 12 months. This left 25% of the graduates still looking for suitable work at the end of the first year after graduation. However, 4 years on, 84% were employed and a further 7% were in full-time graduate education and professional training (Schomburg and Teichler, 2006: 136). However, the first few years after graduating are a period of transition in another sense. Besides seeking a first suitable job, job changes during the first few years are common as is time spent in further study. A short period of unemployment is not all that uncommon, although this might be largely voluntary, for example, allowing time for the graduate to travel.

All of this indicates that it is unwise to view information on the early employment experiences of graduates as an indicator of graduate employment overall. However, various studies have suggested that a late transition into regular employment can be associated with relatively poor career prospects in the long term for particular graduates.

Globalization and the Knowledge Society

Much has been written about the new knowledge economies and their needs for greater supplies of well-qualified knowledge-workers. As has already been noted, it is this belief that has justified the rapid expansion of higher education internationally, despite the occasional voices of skeptics (e.g., Wolf, 2002; Brown and Hesketh, 2004). An ongoing international research project (the flexible professional in the knowledge society, a European Union (EU)-funded framework VI project) is addressing questions concerning the requirements made of graduates within knowledge societies and the following paragraphs draw on its conceptualization.

The term knowledge society or knowledge economy suggests not only the expansion of higher education and knowledge-intensive or high-technology sectors of the economy but also that the characteristics of the vast majority of work organizations are changing under the influence of the increasing importance of knowledge. This is important to acknowledge as it alters materially our notions of graduate jobs. Moreover, it has implications for the kinds of demands made of higher education graduates. The project distinguishes four such demands: expertise, functional flexibility, knowledge management, and mobilization of human resources.

Expertise refers to the possession of specialist knowledge, related to workplace productivity.

Functional flexibility is required because of rapid developments in technology, markets, organizations, and relevant knowledge. These make work dynamic and require graduates to be able to respond to diverse challenges and be

willing to rapidly acquire new knowledge. Knowledge management refers to the whole process of developing ideas and implementing them. Mobilization of human resources refers to the capacity to take responsibility for change, of oneself and of others. It is not only partly about leadership but also partly about self-steering and having the flexibility and skills to change the working environment.

Of course, not all graduates face all four demands and the reality of employment is of considerable diversity. The project has conceptualized this diversity in the following way, distinguishing between elite and mass and between specialist and generalist positions to give a four-fold classification as follows:

1. Elite specialists – the lawyers, doctors, accountants – need high status based on specialist expertise with recruitment highly regulated and based on possession of educational credentials, generally enjoying a protected position in the labor market.
2. Elite generalists – high-level managers, civil servants, top politicians, and advisors – need some expertise but functional flexibility and knowledge management are more important; recruitment sometimes is based on institutional prestige (Oxbridge, *Grandes Ecole*, Ivy League) more than credentials.
3. Mass specialists – engineers, nurses, and teachers – need specialist knowledge but do not need to be experts; strict entry requirements are based on credentials but supply much less rationed than elite specialists; therefore, these are more subject to labor-market fluctuations – possibility of status congestion (see Brown and Hesketh, 2004).
4. Mass generalists – marketing, sales, support staff, and administrators – for whom expertise/knowledge is much less important than an optimal mix of knowledge management, ability to mobilize available human resources, and functional flexibility. Recruitment is much less regulated than for the other categories.

These are positions in the knowledge society rather than subjects of study, types of higher education, or even types of graduates. Clearly, the expansion of higher education has given rise to a huge growth in the mass categories and most graduates in future are likely to find themselves in these sorts of positions.

Another current project suggests that the demand for certain elite specialists will increasingly be met within the global labor market with jobs transferring to low-wage economies possessing high-skills labor – China and India most commonly mentioned (Brown, 2004). However, this seems likely to affect only a limited number of occupations and employment fields.

A linked issue concerns the mobility of students and graduates as they use their higher education and its resultant qualifications in order to be geographically mobile. However, while 17.6% of the graduates surveyed in the

Schomburg and Teichler study had some international work or study experiences abroad during their higher education, only 2.7% were currently working abroad at the time of the survey (4 years after graduation). A further 5.2% had previously been employed abroad and 10.7% had occasional need to work abroad for short periods (Schomburg and Teichler, 2006: 127).

A further caveat must be made about the nature of the globalized knowledge society by pointing to the considerable national differences that exist in the nature of the relationships between higher education and employment. The Schomburg and Teichler study concludes that:

By and large, the . . . survey demonstrates so striking disparities of graduate employment and work in the 11 European countries and Japan that common elements seem to be at most secondary. (Schomburg and Teichler, 2006: 139)

Beyond the incontrovertible fact that graduates are more likely to be employed than nongraduates and will probably have better jobs, much else is going to depend on local labor-market conditions and on traditions of higher education's relationship with it. While there may be some signs of convergence between different national traditions – for example, in the Bologna process in Europe – it seems that multinational employers are also adapting themselves to different local circumstances in different parts of the world. While notions of knowledge economy and society provide useful conceptual tools with which to view graduate employment in different national and local contexts, it also seems likely that it will continue to be unwise to generalize empirically much beyond national (and in some places regional) settings.

Conclusion

In a paper based on work done for the United Nations Educational, Scientific, and Cultural Organization (UNESCO) a few years ago, Ulrich Teichler listed what he saw were some important biases in available information on graduate employment. These were:

- quantitative information is more extensive in the traditional graduate employment sectors (that are shrinking) and relatively weak in the newly emerging, atypical (and expanding) sectors;
- an imbalance between quantitative-structural and qualitative data (i.e. plenty of information on the whereabouts of graduates and their income, but much less on types of work tasks, extent to which knowledge acquired in HE is actually utilised);

- employers' statements are often taken as objective and factual;
- researchers assume that practitioners surveyed (employers, graduates) are 'experts' on links between skills and work tasks;
- research findings in individual countries are often over-interpreted as universal truth (Teichler, 2000).

A further bias in the research literature to the ones listed above is the focus on the early years of employment. Even if we are beginning to move beyond the fixation with first destinations, we still rely too heavily on information gained only during the first few years of employment. Evidence on transition is too often interpreted as evidence on employability.

Another source of bias is that most research – and contributions to debate – is undertaken by interested parties. There is sometimes a spin put on research findings to draw attention to the positives and the successes. Research that is focused on informing policy or enhancing practice may fail to challenge existing assumptions or raise questions that might produce uncomfortable answers.

Bibliography

- Becker, H., Geer, B., Hughes, E., and Strauss, A. (1961). *Boys in White: Student Culture in Medical School*. Chicago, IL: University of Chicago Press.
- Bourdieu, P. (1996). *The State Nobility: Elite Schools in the Field of Power*. Cambridge: Polity Press.
- Brennan, J. L. (2001). *The Employment of UK Graduates*. Bristol: HEFCE (Higher Education Funding Council for England).
- Brown, P. (2004). *The Social Construction of Graduate Employability*. Swindon: Economic and Social Research Council.
- Brown, P. and Hesketh, A. (2004). *The Mismanagement of Talent: Employability and Jobs in the Knowledge Economy*. Oxford: Oxford University Press.
- Dore, R. P. (1976). *The Diploma Disease*. Berkeley, CA: University of California Press.
- Gardner, P. D. (1998). Are college seniors prepared for work? In Gardner, J. N. and van der Veer, G. (eds.) *The Senior Year Experience-Facilitating Integration, Reflection, Closure and Transition*, pp 60–78. San Francisco, CA: Jossey-Bass.
- Mason, G., Williams, G., Cranmer, S., and Guile, D. (2003). *How Much Does Higher Education Enhance the Employability of Graduates?* Bristol: Higher Education Funding Council for England.
- OECD (2004). *Education at a glance: OECD Indicators 2004*. Paris: Organisation for Economic Co-operation and Development.
- Purcell, K. and Elias, P. (2004). *Seven Years On: Graduate Careers in a Changing Labour Market*, p 15. Manchester: HECSU.
- Schomburg, H. and Teichler, U. (2006). *Higher Education and Graduate Employment in Europe*. Dordrecht: Springer.
- Teichler, U. (2000). New perspectives on the relationship between higher education and employment. *Tertiary Education and Management* 6(2), 79–92.
- Wolf, A. (2002). *Does Education Matter? Myths About Education and Economic Growth*. London: Penguin Books.

Higher Education and the Knowledge Society

J Välimaa, University of Jyväskylä, Jyväskylä, Finland

© 2010 Elsevier Ltd. All rights reserved.

Knowledge Society as a Discourse and as a Social Phenomenon

As a concept, knowledge society is very popular in modern societies, used by popular media, politicians, and academic researchers. Internet search through Yahoo! (in February 2007) gave about 59 800 000 entries for the concept knowledge society. Therefore, a useful starting point for understanding the many dimensions of knowledge society is to define it as a discourse which is based on certain intellectual starting points in the analyses of social realities of modern societies, while as a concept it tends to create its own images, expectations, and narratives. In this sense, knowledge society discourse describes the current situation in which knowledge society is both the objective of policies and debates and an agent promoting policies and debates concerning its potentials. This may be one of the reasons why knowledge society is and has constituted a set of lucrative ideas in modern societies, used by popular media, politicians, and academic researchers. The knowledge society discourse takes place in the context of globalization where higher education institutions are more important than ever as mediums in global knowledge economies. In a globalized world, higher education institutions are integral to the continuous flows of people, knowledge, information, technologies, products, and financial capital (Marginson, 2006).

Knowledge itself and the uses of knowledge are nothing new for mankind which understands itself through languages which are by themselves symbolic systems for cultivating and transferring knowledge. What makes the present situation of knowledge society exceptional is the quantity of knowledge (and information) produced daily and the use of information technologies in data-intensive processes. According to Bell (1973), the postindustrial society is a knowledge society for two reasons: "first, the sources of innovation are increasingly derivative from research and development (and more directly, there is a new relation between science and technology because of the centrality of theoretical knowledge); second, the weight of the society – measured by a larger proportion of Gross National Product and a larger share of employment – is increasingly in the knowledge field." The same notion has been repeated by Castells (1996) when analyzing the difference between previous modes of development with the mode of development of the digital world.

Knowledge Society as an Intellectual Device

As a concept, knowledge society, however, has its own history. The term knowledge society was first used by Lane (1966) whose concept of knowledgeable society describes the great optimism of the early 1960s, suggesting that science would allow for the possibility of a society in which common sense would be replaced by scientific reasoning. Drucker (1969) saw, in turn, that knowledge was central to society as the foundation of economy and social action. The use of the term knowledge society began to expand, in turn, with the studies of researchers such as Mansell and When (1998) and Stehr (1994) in the 1990s.

There can be found two main traditions in understanding and defining knowledge societies. The first emphasizes the importance of information technologies as the driving force of knowledge society or the information society, whereas the other focuses on knowledge and its importance in changing societies. The tradition of information societies continues the intellectual tradition begun by Machlup (1962) who analyzed the production and distribution of knowledge in the US. In this tradition, people have discussed the major transformations that are possible through harnessing electronic information processing technologies to the social and economic priorities of industrial societies. In this intellectual context, knowledge society is often understood as emerging from the simultaneous growth of the Internet, mobile telephony, and digital technologies with the third industrial revolution, which has seen much of the working population migrate to the service sector and has revolutionized the role of knowledge in societies (see UNESCO, 2005: 18).

However, according to a number of writers (Stehr, 1994; UNESCO, 2005), information society as a concept or notion tends to give a more limited and technically oriented description of the challenges in a modern society, even though intellectually it belongs to the same trend aiming to analyze social changes in postindustrial societies. The main critique against this perspective to changes in societies acknowledges the fact that knowledge always has a social function which is rooted in its production, distribution, and reproduction. The nature of these issues is not technical; rather it is political, because these qualities of information and knowledge are related to social structures and the uses of power in a society.

The distinction between information society and knowledge society is also rooted in the difference between knowledge and information. According to Bell (1973: 175), knowledge is “a set of organized statements of facts or ideas, presenting a reasoned judgment or an experimental result, which is transmitted to others through some communication medium in some systematic form.” Following the argument presented by Castells (1996), information is data that have been organized and communicated. Therefore, information remains a fixed stabilized form of knowledge, tied to time and users. For this reason, information may be used as a commodity and, therefore, bought or sold.

The second tradition of knowledge society aims to analyze the societal importance of knowledge responding to the sociological need for a theory of society that resonates with the new social realities. According to Stehr, conceptualizations of society based on the relationships between labor and property (capital) no longer provided the intellectual insight necessary to describe, understand, and explain modern societies. Stehr does not argue that labor and capital dynamics have disappeared; however, he argues that societal relationships cannot be explained without integrating the primacy of dynamics related to knowledge. In creating his own theory of modernization, knowledge society, Stehr (1994) suggests that modernization essentially involves multiple and necessarily unilinear processes of extension and enlargement. The question is: Does the nature of knowledge production change societies, cultures, and economics? The popularity of the term knowledge society is evidence in and by itself that understanding modern society as knowledge based indicates that the more conventional traditional understandings within societies are changing.

The idea of social change based on extension and enlargement is also familiar to higher education researchers. Martin Trow's assumption that the social role of higher education changes with the expansion of the student body has been accepted as an insightful conceptualization of mass higher education (Trow, 1974). Following this reasoning, it can be said that mass higher education is the dominant social form of higher education in the knowledge society. A similar trend has been noted by Clark (1983) who maintains that the main source of social dynamics in higher education is the expansion of knowledge, which leads to new research fields creating a demand for new chairs and professorships to be established for emerging disciplines. It also creates the need to establish new training programs and new higher education institutions, thus supporting the development of mass higher education systems. The emergence of knowledge societies and the expansion of higher education are historically simultaneous and may also have a causal relationship because knowledge production in itself supports growth in industrial production and creates new business activities.

Associated Concept of Learning Society

The discussion on learning societies coincided with the emerging knowledge society discourse aiming to describe potential relationships between knowledge and change in society from the perspective of teaching and learning. Originally, the concept of learning society referred to a new kind of society in which the old distinctions between formal and nonformal education were no longer valid (Hutchins, 1968). In this new context, lifelong learning becomes indispensable because there is a need to change workplaces and often professions and update knowledge during one's career. Crucial new skill in a learning society is also the ability to learn how to learn. Furthermore, learning is no longer the privilege of an elite or a same-age cohort, but it tends to cover the whole community and the whole lifetime of an individual (UNESCO, 2005). The notion of learning society supports many discourses of knowledge society. Both emphasize the centrality of knowledge production and lifelong learning of the labor force. These perspectives on the social role of education have also supported the expansion of higher education into mass higher education systems.

Higher Education and Knowledge Production

The importance of higher education in the knowledge society discourse emerges from the notion that changes in science, research, and technology have changed the social role of universities. There are two main perspectives concerning the recent transformation. The first asserts that a radical metamorphosis is taking place in the relationship between knowledge production and university institutions. Authors such as Gibbons *et al.* (1994), Nowotny *et al.* (2001), and Etzkowitz *et al.* (2000) propose that governments are promoting national prosperity by supporting new lucrative technologies together with universities which become engines of their regions. Gibbons *et al.* (1994) argue that a new form of knowledge production (mode 2) is replacing the traditional one (mode 1). Mode 1 knowledge had been produced within autonomous disciplinary contexts governed mainly by academic interests of a specific community, whereas mode 2 is produced in the context of application. Mode 2 is transdisciplinary research, characterized by heterogeneity and is more socially accountable and reflexive than mode 1 knowledge. In addition, they argue that universities are losing the monopoly of knowledge production, because knowledge may be produced in a variety of organizations and institutions.

The other variant of the metamorphosis thesis is the triple helix thesis which states that the university can play an enhanced role in innovation in increasingly knowledge-based societies. Etzkowitz and Leydesdorff

(2000) assert that the previously isolated institutional social spheres of university, government, and industry have become increasingly intertwined. This has brought academic, economic, and wider networks of social actors together in new constellations in the triple helix knowledge dynamics. Taking support from system theory thinking, Etzkowitz *et al.* (2000: 4) assert that four processes describe the major changes in the production, exchange, and use of knowledge in the triple helix model. These are internal transformations in each of the helices (academia, state, and industry), followed by the influence of one institutional sphere on another. The third process is the creation of a new combination of trilateral linkages, networks, and organizations among the three helices, while the fourth describes the effect of these interinstitutional networks both on their originating spheres and the larger society.

The characterization of knowledge production dynamics between mode 1 and 2 have been, however, criticized as being one-eyed and reductionistic, focusing on a relatively small domain of the vast diverse landscape of science in society. It has also been argued that the dichotomy of mode 1 and 2 presents two discrete ideal types that probably never existed in the real world. The ideological connection between this discourse and political neoliberalism has also been pointed out. The triple helix thesis is, in turn, problematic in its the assumption that all twenty-first-century universities will follow the developmental path of entrepreneurial universities (Tuunainen, 2005: 278–279).

A second, more moderate view of the transformation of knowledge production and universities holds that academic capitalism is challenging the traditional values found in higher education institutions, where an attempt is made to strengthen neoliberal values and management practices. Academic capitalism is a process which promotes competition and market-like behavior both inside higher education institutions and with their relations with business and society. Universities become fertile ground for entrepreneurial universities and academics. However, there are significant differences between disciplines in the academic world as regards their relationship with society (Slaughter and Leslie 1997; Slaughter and Rhoades, 2004).

All these different theoretical assumptions assert that the nature of universities as an organization is changing. It has been asserted that universities are becoming hybrid organizations (Slaughter and Leslie, 1997), or mode-2 institutions (Nowotny *et al.*, 2001) or entrepreneurial universities (Etzkowitz, 2003). According to Clark (1998), entrepreneurial universities should be able to capitalize on genuine connections to the academic heartland or central missions of the university, that is, teaching, research and service.

According to an empirical study by Marginson and Considine (2000), it is indeed evident that there is a general

pattern of modeling universities along the lines of enterprises. This new form of enterprise university has a strategically centralized leadership which is highly responsive to the external setting, it makes wide use of corporate and business forms, empties out academic governance, and weakens disciplinary identity. However, it does not appear as if mode 2 or the triple helix models constitute global trends, because knowledge production plays out differently in distinct types of universities. Old, established universities with strong academic and disciplinary cultures possess more power to resist and generate change, while other types of higher education institutions are more vulnerable to the new management ideas (Marginson and Considine, 2000). Furthermore, empirical research findings suggest that commercialization of academic research turns out to be in conflict with other university activities, most apparently, with publicly funded research and university teaching (Tuunainen, 2004). Empirical academic research has revealed that generalizing abstractions like mode 2 or triple helix model do not necessarily take into account the fact that academic work and universities are complex and often contradictory entities whose development is shaped by a combination of historical, political, and cultural characteristics.

Intellectual Property Rights and Public Good

Questions about the ownership of knowledge are typical examples of the emerging problems of higher education in the era of the knowledge society. There are two interrelated issues here. The first issue is related to the ownership of innovations. A number of countries have tried to solve this problem through legislation which regulates intellectual property rights of academics and universities. The first such act to be passed was the 1980 Bayh-Dole Act in the United States which gave ownership of intellectual property, arising from federally funded research, to universities (Etzkowitz *et al.*, 2000). However, the idea of intellectual property rights is challenged by the ethical basis of the open source development process, which is based on a vision of information and communication technologies as a public good, in which anyone is welcome to participate and all are invited to benefit. The second issue is related to student tuition fees. The question of who benefits from higher education is often translated into the question of who should pay for education. These fundamental topics are related to debates on private and public goods in higher education, because when these questions are combined with budget reductions in higher education, they easily produce debates on the problems of public higher education institutions.

Knowledge Society as a Political Goal

Governments virtually in all parts of world see the potential of the university as a resource to enhance innovation

environments and create a regime of science-based economic development. However, it is quite difficult to show empirically what this means in practice even though it is clear that governments all over the globe are willing to promote this third mission of universities in order to encourage universities to be active producers of innovations. In addition to nation-states, regional (such as European Union (EU)) and global dimensions of knowledge society discourse need to be addressed in order to have an overall picture of different political uses of knowledge society as concept.

Nation-States

At the level of nation states, knowledge society can be seen to have taken on distinct forms. Castells and Himanen (2002) forecast three alternative routes to knowledge society: (1) Silicon Valley – a market-driven, open society (USA), (2) Singapore – an authoritarian model of the knowledge society, and (3) the Finnish model – which describes an open, welfare-state-based knowledge society. This typology highlights the variety of possibilities to organize a knowledge society and that the notion can be defined and approached in different ways. The examples of Finland and the UK reveal the different approaches taken by national governments.

A Nordic example is provided by the Finnish experience of creating a national innovation strategy to promote cooperation between private companies and public authorities. The nation-state plays a significant role through various social actors which bring researchers and business companies together in order to focus resources on problems deemed to be of economically strategic importance. These are either development agencies which support cooperation between businessmen and researchers or public organizations which promote cooperation between the world of business and academic worlds. It is in this context that the role of higher education policy becomes important. In Finland, universities are seen and defined as part of the national innovation system aiming to increase the capacities of Finnish enterprises and the nation-state in general with regard to the international market place. The Finnish example also suggests that knowledge society is organized in and through networks. A network-based social system is a highly dynamic, open system, susceptible to innovating without threatening its balance.

In the UK, the purposes of universities in the learning society have been defined in the Dearing Report (1997) as inspiring and enabling individuals to develop their capabilities to the highest levels; increasing knowledge and understanding; serving the needs of the economy; and shaping a democratic and civilized society. These aims describe the importance of higher education to the personal development of the individual in contrast to short-term employment and education provided by corporate

training programs. It also emphasizes the functions of teaching and research in the development and dissemination of knowledge and, furthermore, expresses the economic value of both of these activities. Finally, this political report aimed at setting goals for the development of British higher education, also pays specific attention to the cultural and political value of higher education in maintaining and developing civil society (Laurillard, 2002).

Regional Level

In Europe, it is easy to find support for the idea that higher education should support national economic competitiveness in the global market place. In addition to European nation-states, knowledge society discourse has created its own imaginary social space in the EU itself. This argument is underlined on reading the European Commission's knowledge society homepage, which begins with the central objective of the Lisbon strategy: "to become the most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion." This citation, in and by itself, indicates the importance of the topic for the EU. In order to reach this objective, Europe's education and training systems need to adapt both to the demands of the knowledge society and to the need for an improved level and quality of employment.

Developing Global Information Society with the Help of Higher Education

In a global perspective, the knowledge society may be characterized as a global information society (UNESCO, 2005). In the global context, the use of information technologies, the access to knowledge resources, and the political aspects of knowledge society are raised. In this perspective, a global information society emerges as one of the main challenges for development, because it is evident that the global information society is a political goal which is far from being reached. Specifically, in the beginning of the twenty-first century, only 11% of the world's population had access to the Internet and 90% of these connections were in leading industrialized economies in North America (30%), Europe (30%), and in the Asia-Pacific region (30%). In addition, 82% of the world's population accounted for only 10% of the Internet connections in the world (UNESCO, 2005). This disparity has been called the digital divide, which cuts across the globe, following the contours of social and economic capital distribution. The problem was not only due to lack of proximity to electricity grids but also due to the fact that interactive computers and Internet connections were expensive for the majority of mankind. This form of exclusion unites urban slum dwellers and the homeless as

well as remote villagers in developing countries. Digital divide is also related to contents, as sources of knowledge in several languages are not a characteristic of information and communication technologies. One should not forget, therefore, that knowledge society discourse is dominated by and focused on the Western cultural hemisphere and even there it is focused on the conditions of the relatively young, well-educated working-age citizens geographically located in the urban areas of a few rich countries.

The role of higher education is, however, seen crucial in the development of global information societies. The UNESCO (United Nations Educational, Scientific, and Cultural Organization) World Conference on Higher Education emphasized that the relevance of higher education means: (1) being politically responsive, (2) being responsive to the world of work, (3) being responsive to other levels of the education system, (4) being responsive to culture and cultures, (5) being responsive to all, (6) being responsive everywhere and all the time, and (7) being responsive to students and teachers. (UNESCO, 2005: 97).

Summary

This impressive list of social responsibilities expected from higher education clearly indicates that world communities have high hopes regarding higher education. It also indicates that the social role of higher education in the global information society is seen as crucial for the development of societies. Furthermore, the list of expectations highlights the central roles universities as producers of knowledge and educators of experts are assumed to play in knowledge societies. It also reveals the fact that knowledge society is a discourse which contains multiple and often conflicting expectations concerning the roles expected to be played by higher education institutions.

Bibliography

- Bell, D. (1973). *The Coming of Post-Industrial Society: A Venture in Social Forecasting*. New York: Basic Books.
- Castells, M. (1996). *The Information Age: Economy, Society and Culture. Vol I: The Rise of the Network Society*. Oxford: Blackwell.
- Castells, M. and Himanen, P. (2002). *The Information Society and the Welfare State: The Finnish Model*. Oxford: Oxford University Press.
- Clark, B. R. (1983). *The Higher Education System*. Berkeley, CA: University of California.

- Clark, B. R. (1998). *Creating Entrepreneurial Universities: Organizational Pathways of Transformation*. Oxford: Published for the IAU Press by Pergamon Press.
- Dearing Report (1997). *Higher Education in the Learning Society*. Norwich: HMSO.
- Drucker, P. (1969). *The Age of Discontinuity. Guidelines to Our Changing Societies*. New York: Harper and Row.
- Etzkowitz, H. (2003) The norms of entrepreneurial science: Cognitive effects of the new university-industry linkages. *Research Policy* 27(8), 823–833.
- Etzkowitz, H. and Leydesdorff, L. (2000). The dynamics of innovation: From national systems and “Mode 2” to a triple helix of university–industry–government relations. *Research Policy* 29(2), 109–123.
- Etzkowitz, H., Webster, A., Gebhardt, C., and Terra, B. R. C. (2000). The future of the university and the university of the future: Evolution of ivory tower to entrepreneurial paradigm. *Research Policy* 29(2), 313–330.
- Gibbons, M., Limoges, C., Nowotny, H., et al. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: Sage.
- Hutchins, R. (1968). *The Learning Society*. London: Penguin.
- Lane, R. E. (1966). The decline of politics and ideology in a knowledgeable society. *American Sociological Review* 31, 649–662.
- Laurillard, D. (2002). Rethinking teaching for the knowledge society. *EDUCAUSE Review* 37(1), 16–25.
- Machlup, F. (1962). *The Production and Distribution of Knowledge in the United States*. Princeton, NJ: Princeton University Press.
- Mansell, R. and When, U. (1998). *Knowledge Societies: Information Technology for Sustainable Development*. New York: UN Commission on Science and Technology for Development. Oxford University Press.
- Marginson, S. (2006). Putting ‘Public’ back into the public universities. *Thesis Eleven* 84, 44–59. London: Sage.
- Marginson, S. and Considine, M. (2000). *The Enterprise University*. Cambridge: Cambridge University Press.
- Nowotny, H., Scott, P., and Gibbons, M. (2001). *Re-Thinking Science. Knowledge and the Public in an Age of Uncertainty*. Cambridge: Polity Press.
- Slaughter, S. and Leslie, L. L. (1997). *Academic Capitalism-Politics, Policies and the Entrepreneurial University*. Baltimore, MD: Johns Hopkins University Press.
- Slaughter, S. and Rhoades, G. (2004). *Academic Capitalism and the New Economy. Markets, State and Higher Education*. Baltimore, MD: Johns Hopkins University Press.
- Stehr, N. (1994). *Knowledge Societies*. London: Sage.
- Trow, M. (1974). Problems in the transition from elite to mass higher education. *Policies for Higher Education. Conference on Future Structures of Post-Secondary Education*, Paris.
- Tuunainen, J. (2005). Hybrid practices? Contribution to the debate on the mutation of science and university. *Higher Education* 50, 275–298.
- UNESCO (2005). *Toward Knowledge Societies. UNESCO World Report*. <http://unesdoc.unesco.org/images/0014/001418/141843e.pdf> (accessed Jun. 2009).

Relevant Website

<http://ec.europa.eu> – European Commission.

Higher Education and the Labor Market

J Enders, University of Twente, Enschede, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

The relationship between higher education and the labor market is among the key issues of scholarly debates whenever the qualification, certification, and status-distributive functions of higher education are at stake. While higher education was traditionally thought of as a means for the self-reproduction of the elite in society, massive educational expansion has stimulated growing interest in the study of change and stability in higher education institutions, curricula, and students; in the labor market and the employment system; and in the relationship between changes in higher education and changes in the labor market. As with many other concerns about (higher) education, the sources of growing theoretical and empirical insights in this area mainly lie in the social sciences. Most studies find their roots in business studies, economics, educational science, policy studies, public administration, and sociology. Quite a few studies have also been undertaken in interdisciplinary teams of scholars that tend to undermine strong assumptions in disciplinary perspectives on the principal nature of the relationship between higher education and the labor market.

This article offers an overview on the development of theoretical work and empirical insight on the relationship between higher education and the labor market in times of educational expansion. Structural–quantitative aspects in the development of the match and mismatch between higher education and the world of work as well as qualitative aspects of the demand and supply of competencies and skills are addressed. Further attention is given to growing insights into the institutional embeddedness of higher education and the labor market in cross-national comparative perspectives and their change and stability over time. Comparison is made between the utility and limitations of different models explaining the ongoing adaptation between higher education and the labor market by providing international examples of empirical research.

Quantitative Aspects

In the course of massification of higher education, a lot of concern was and still is about the quantitative match or mismatch between higher education and the labor market. Public debates and scholarly reflections in this area started in the 1960s with high hopes for more equal opportunities and economic prosperity through

investment in education and training. The expansion of higher education occurred as part of a larger societal development that was accompanied by a flourishing public sector. Its expansionist logic proceeded on a dual track, offering new educational opportunities as well as new employment opportunities in education, health, and welfare in a kind of self-vindicating system (Nowotny, 1995). The 1970s faced the end of the dream of everlasting economic prosperity. The pessimistic view spread that expansion of higher education had gone too far and that graduates' skills no longer matched the needs of the employment system. The debate was marked by sharp disagreements over a presumed overeducation or underemployment of the many more graduates for whom not sufficient or not sufficiently well-qualified jobs would be available. In the 1980s, expectations and empirical findings adjusted to a somewhat blurred state of affairs which neither supported the high hopes of the 1960s nor reinforced the deep sense of crises of the 1970s (Teichler, 1999). Despite the ongoing expansion of higher education, labor markets still are able to absorb graduates and most of them find work and careers providing substantial material and immaterial rewards.

What emerged from further analyses was a mixture of vertical and horizontal adjustments in job placement, changing values, and expectations of what was considered a desirable job for the highly qualified and what was meant by a proper link between higher education and work. Career studies and employment statistics show that the absorption of the growing number of graduates by the labor market has been achieved in a variety of ways: the growth of occupations traditionally filled by graduates, increasing employment in the private sector and the nonprofit sector, increasing numbers of graduates entering mid-level positions and semi-professions, a growth in the (still) small number of self-employed and the (still) small number of graduates in blue-collar jobs, as well as an extension of job-search periods after graduation and initial unstable employment and unemployment. In developed countries, an important component in such adaptations between an expanding higher education system and the labor market has been due to the massive inclusion of women into higher learning. This has led to an increase in the labor market participation of women. Nonetheless, the careers of female graduates still show remarkable differences from those of male graduates with a lower return of investment in education for women.

In the wake of massification, both policy debates and scholarly work thus became more interested in the dynamics of the relationship between higher education and the labor market. It became obvious that certain gaps could not be avoided. Shortages or oversupply of graduates overall or in specific occupations and labor market segments readjust only after some time; adaptations of student choices as well as of the labor markets proceed with time lags (Brennan *et al.*, 1996). All this undermines strong hopes for a perfect forecasting and planning of supply and demand.

Such economic perspectives on the forecasting of the contribution of higher education to the labor market often draw on human capital theory (Becker, 1964) that assumes that highly qualified employees will be preferred by employers and will earn more because of their qualifications that make them more productive. Some empirical evidence could be found for such a claim that continues, however, to be contested. Human capital theory suffers certain empirical shortcomings. It can neither rely on appropriate measures of graduates' qualifications nor on appropriate measure of the qualification demands of the workplace. Instead, research in the tradition of the human capital approach works with proxy variables such as the level of educational attainment (for qualifications acquired) and wages (for contributions to productivity). Further difficulties derive from the fact that higher education also sorts and selects graduates, and that employers hold imperfect information on graduates' abilities and are thus heavily dependent on this function of higher education in labor-allocation decisions. Scholars such as Spence (1973) and Arrow (1973) have pointed out that selection and rewarding of employees take place on the basis of such signals that are assumed to point to the productive capacities of the employee. The labor queue theory (Thurow, 1975) has added to this debate the interesting argument that part of the qualifications required on the job will not even be acquired in (higher) education but will be acquired by training on the job or further training alongside the job. This argument points to the unavoidable fact that higher education will not deliver graduates that perfectly match the needs of the labor market overall or of a specific job. Higher education rather provides the more or less developed foundations for development and refinement of qualifications on the job and throughout the career. These scholarly debates have already pointed to what became in our days a more commonly held insight into the relationship between higher education and the labor market: the match between supply and demand of qualifications is by definition imperfect, only partly transparent and fluid. Graduates thus should hold qualifications of flexible knowledge workers and lifelong learners being able to make contributions to shifting and ever-changing demands on the workplace. Finally, some

scholars have raised doubts on to what extent higher education has an effect on the capabilities and the productivity of its graduates at all. They provide a variation of a theme that credentialists such as Collins (1979) developed earlier. Credentialists argue that higher education is used as a gatekeeper to social mobility and elite reproduction, legitimizing access to privileged positions in economy and society. Again, there is certain empirical evidence to this claim that does, however, not necessarily exclude that higher education also has an impact on the capabilities of its graduates. Research thus points to a mixed state of affairs with elements of empirical evidence for the theories addressed above.

Qualitative Aspects

Since the 1990s, a new process of adjustment and restructuring seems to be underway, that tends to undermine the whole notion of a quantitative match. The perils are no longer seen in diminishing return of investment due to growing competition or in the labor markets being swamped by overqualified and dissatisfied applicants. It is nowadays more frequently underlined that it is the occupational structure and social stratification system itself that has become mobile. This is accompanied by deep structural changes in the way the economy works as well as a perceived individualization of the life-course regime (Beck, 1986). The characteristics of occupations and jobs, the vertical as well as the horizontal division of work, and the needs and reward structures of the employment system continue to be restructured. Learning–working pathways through education, training, and employment tend to be de-institutionalized and re-institutionalized. Quality thus stands for possessing a mixture of skills and knowledge for new and changing configurations. Graduates are expected to be trained for what is increasingly seen to become a market for knowledge workers in constant flux. In addition, the student body has become more heterogeneous due to ongoing massification in terms of social background, age, levels of preparation and work experience, patterns of studying and learning, aspirations, and life chances.

The issue of the relationship between higher education and the labor market and between demand and supply thus gains again in importance. However, the emphasis is less on structural–quantitative relationships and more on the qualitative match between a changing body of students and graduates and ever-changing job requirements and labor markets. In consequence, competences beyond cognitive knowledge play a stronger role in political debates as well as in scholarly work. Terms such as soft skills, key qualifications, practice orientation, and employability signal that higher education is increasingly expected to take more explicit care of competences that go

beyond the codified body of knowledge related to certain disciplines, fields of study, occupations, or professions (Teichler, 1999). Moreover, research on employers' recruitment processes and selection criteria show in fact that they pay less attention to curricula details than to the reputation of certain higher education institutions or fields of studies, professional experiences of graduates during their studies, and the training and socialization of soft skills.

From this point of view, higher education is too youth centered and program oriented as far as the actual and desired clientele is concerned, and relies on the old-fashioned idea that full-time students can appropriately be piggybacked with consistent stocks of knowledge prior to the transition from education to work. Higher education is, thus, expected not only to continue considering fair access according to sociobiographic background, but also to strengthen the overall supply with a highly trained workforce in the sense of the old regime. It is also expected to further diversify structurally and, in terms of conditions of study and courses provided, to devote greater attention to generic competencies and social skills and to reshape its function for a society of lifelong learning, to prepare students for a growing internationalization, and to serve practical learning beyond classroom teaching. In other words, higher education is expected to move from a front-end model to a life-span model of education and training, from curricula to learning pathways.

Certainly, such a vast and inclusive concept of higher education in the knowledge society holds strong appeal but goes by no means uncontested. This is partly so because such an integrated view of the role of higher education in the knowledge society can serve to obscure attempts to define clearly what educational goals should be pursued and who should be responsible for which provisions and actions. It is also unclear what an appropriate balance between cognitive skills and soft skills, between workplace-related competencies and transferable skills would look like (Tuijman, 1999).

Studies among employers point, for example, to the fact that growing expectations as regards soft skills do not necessarily imply that the cognitive domain of knowledge becomes less important. Further signals for the employability rather grow in importance on top of traditional expectations in the cognitive domain. In addition, various studies show that different countries provide distinctive mixtures of qualifications, and that the returns of investment for further specialization or greater flexibility are a topic of considerable debate.

Last but not least, the lack of any satisfactory measure of the abilities and competencies of higher education graduates has thus found increasing attention. The outcomes and policy impacts of research on school effectiveness

and of studies on the qualifications of students in school (e.g., the Programme for International Student Assessment (PISA) study of the Organization for Economic Co-Operation and Development (OECD)) are also stimulating this search for appropriate measures and methods to study abilities and competencies of higher education graduates. This is an area where further conceptual and analytical work is of growing interest. Various projects are undertaken to overcome this situation and to find ways for an objective assessment of the value added by higher education training and experience. This is by no means an easy undertaking but would help to learn more about suggested key qualifications or soft skills, and how they are related to the mix of hard skills and tacit knowledge traditionally provided. Further research on learning outcomes could also link up to labor market studies in order to learn more about the trainability of employability, and the evidence for the recognition, transfer, and portability of skills to and within the labor market. Further insights in this area are also likely to revitalize traditional debates between the human capital approach and the screening hypotheses that were suffering from a lack of reliable measures of graduates' qualifications.

Institutional Embeddedness

Major contributions to the further development of our insights into the (changing) quantity and quality of the relationship between higher education and the labor market have been made by institutionalist approaches. Such a perspective emphasizes the matching problem between higher education and the labor market as the outcome of individual and organizational decision making, shaped by institutional characteristics. Such institutional characteristics determine the opportunity structures for new and old graduates (macro-level) and shape the interactions between graduates and employers (micro-level). Insights into the institutional embeddedness of the relationship between higher education and the labor market and their cross-national variations have benefited a lot from a life-course perspective. Life-course research analyzes the pathways of individuals, social groups, and cohorts as a process rather than as a state. Concepts such as trajectories, transitions, duration, and sequences are employed to identify life-course regimes in cross-national perspectives as well as their change and stability over time.

There is, for example, evidence that the signaling function of formal educational qualifications is more important in job-allocation processes in countries characterized by a high degree of selectivity, stratification, and standardization of higher education (Allmendinger, 1989; Shavit and Müller, 1998). Selectivity points to the reliability of degrees due to earlier selection processes for

access to higher education or certain parts of higher education. Stratification of higher education points to a vertical or horizontal differentiation between institutions or institutional types providing further signals to the employer. Standardization points to the nationwide recognition of curricula and degrees. All these characteristics may allow employers to invest more trust into the reliability of degrees than in countries with a low degree of selectivity, stratification, and standardization of higher education. Systems also differ in the extent to which they have a more academic or a more vocational orientation, the extent to which general education or more specialized training is held as important, the existence of institutionalized links between higher education and employers, or the flexibility of pathways from education to work.

The transition of graduates into work and their further careers are also enabled or constrained by institutional patterns of the labor market (Shavit and Müller, 1998; Müller and Gangl, 2003). Countries differ in the extent to which they offer positions on occupational labor markets that are in principle open for all highly qualified or positions on internal labor markets that are usually filled by internal candidates. The degree of labor market regulation by public authorities, the influence of trade unions on labor market conditions, and the degree of (un)employment protection differ between countries as well and impact on the employment relations of young and old graduates.

The relationship between the demands of the labor market and the supply from higher education is also influenced by structural changes in the labor market and by the way work and career are organized in employment organizations. The most important development concerns the tertiarization of the labor market shifting away from manufacturing and toward service provision. The rise of multinational companies and the flourishing of small- and medium-sized enterprises have changed the labor market for graduates, and they provide a specific mix of demands for their highly qualified employees. Many changes of labor market regulation went in the direction of deregulation and flexibilization even though with different speeds and intensities in different countries. The public service sector traditionally sheltered from market competition became the subject of scrutiny in the wake of new public management approaches. The concept of the transitional labor market points to the assumption that overall expectations on mobility and flexibility of employees are increasing resulting in an overall focus on the employability of graduates and a de-standardization of the life course.

Various lessons can be drawn from empirical investigation into the cross-national variations of the relationship between higher education and the labor market and their change and stability over time. First, transition to the

labor market and the quality of work and career obtained by graduates differ strongly cross-nationally (Teichler, 2007) and differences can partly be explained by the different institutional regimes in higher education and in the labor market. Second, many national systems of higher education have become more stratified in the wake of educational expansion. Growing variation according to different types of institutions, different fields of study programs, different degree levels, and reputation and prestige within formally equal institutions and programs have been observed. Such differentiation has been instrumental in stratifying the growing number of graduates on the labor market. As a consequence, occupational rewards of graduates differ more than in the past and the status-selection function of higher education has become even more important due to its massification. Third, stability and ongoing adaptation rather than dramatic change in the relationship between higher education and the labor market can be observed over time (Shavit and Müller, 1998; Heinz, 1999). Fourth, many studies show that early careers have a lasting effect on later career. They confirm the early career hypotheses that expect that the relationship between early and later career success has changed to a lesser extent than what the proponents of an overall flexibilization of the labor market and a de-standardization of the life course would assume. However, as we all know: what might be a sign of the future to come may not be representative today.

Conclusion

Several insights can be highlighted on looking at the development of research on higher education and the labor market. One is that an automatic and stable link between higher education and certain positions in the employment system cannot be taken for granted, and also that it has never existed from a systematic point of view. In societies where certification plays an important role in job placement, career trajectories, and life chances, the mechanisms through which this occurs and the expectations over how this should occur change. Second, differences among national traditions of training and selecting those who are able to run society are, however, likely to be important and may change only over long periods. Third, when access to higher education becomes more widespread, the growing opportunities for higher education become, at the same time, something of an obligation. The requirements of knowledge economies exert demand for more, and not for less education. In effect, the gap between the highly educated and the lowly educated in terms of income, status, and life chances is likely to grow, while the rewards for the highly educated are likely to become more diversified and negotiable. Educational expansion leads to a further stratification of higher

education that is instrumental in screening and sorting the increasing number of graduates for the labor market. Fourth, the importance of higher education for the labor market is in the combination of a qualification, a certification, and a status-distributive function. Which function prevails is not a question of principle, but an empirical question to specify the institutional contexts under which one or the other function prevails.

Research on the relationship between higher education and the labor market has traditionally been dominated by strong theoretical assumptions built on relatively simplistic models and a limited set of issues taken into account. Such approaches still play their role but are increasingly challenged by more complex research approaches. Growing complexity has been called for by developments on the research frontier as well as by changes in the relationship between higher education and the labor market. The trend toward more complex studies has also been enforced by new methodologies and improved conditions for quantitative research. Last but not the least, beliefs built into certain theories about universal relationships between higher education and labor and universal megatrends are undermined by growing insights into the relevance of time-specific and country-specific institutional settings and dynamics. Growing awareness of the complexity and variability in the relationship between higher education and the labor market has also led to a growing eclecticism as far as theoretical debates are concerned. Theories strongly adhered to in the past have partly lost their simple beauty, but so far, new holistic frameworks, that would sufficiently cover the territory of the more complex research approaches chosen today, have not been able to replace them. This may be disappointing for those who call for a clear-cut and simple diagnosis of the relationship between higher education and work. However, at the same time, it opens windows of opportunity for a stepwise improvement of our theoretical frameworks and empirically validated insights. Research has improved our insights into the institutional embeddedness of individual choices and collective pathways within the educational system and the world of work, the changing structural-quantitative relationships and qualitative dimensions in the relationship between higher education and the labor market. The further systematization and enhancement of this knowledge is unlikely to generate perfect solutions for problems discussed in policy and practice. However, it will provide better information on the opportunities and limits for attempts in shaping the relationship between higher education and the labor market.

See also: Employability of University Graduates and Graduate Outcomes.

Bibliography

- Allmendinger, J. (1989). Educational systems and labour market outcomes. *European Sociological Review* 5, 231–250.
- Arrow, K. (1973). Higher education as a filter. *Journal of Political Economy* 3, 192–216.
- Beck, U. (1986). *Risikogesellschaft. Auf dem Weg in eine andere Moderne*. Frankfurt am Main: Suhrkamp.
- Becker, G. S. (1964). *Human Capital. A Theoretical and Empirical Analysis with Special Reference to Education*. New York: Columbia University Press.
- Brennan, J., Kogan, M., and Teichler, U. (eds.) (1996). *Higher Education and Work*. London: Jessica Kingsley Publishers.
- Collins, R. (1979). *The Credential Society: An Historical Sociology of Education and Stratification*. New York: Academic Press.
- Heinz, W. R. (ed.) (1999). *From Education to Work: Cross-National Perspectives*. New York: Cambridge University Press.
- Müller, W. and Gangl, M. (eds.) (2003). *Transitions from Education to Work in Europe: The Integration of Youth into EU Labour Markets*. Oxford: Oxford University Press.
- Nowotny, H. (1995). Mass higher education and social mobility: A tenuous link. In Dill, D. D. and Sporn, B. (eds.) *Emerging Patterns of Social Demand and University Reform: Through a Glass Darkly*, pp 72–89. Oxford: Pergamon Press.
- Shavit, Y. and Müller, W. (eds.) (1998). *From School to Work. A Comparative Study of Educational Qualifications and Occupational Destinations*. Oxford: Clarendon Press.
- Spence, M. (1973). Job market signalling. *Quarterly Journal of Economics* 1, 355–374.
- Teichler, U. (1999). The contribution of education and training to employability of youth: Changing concerns, debates and measures. In OECD (ed.) *Preparing Youth for the 21st Century: The Transition from Education to the Labour Market*, pp 215–261. Paris: OECD.
- Teichler, U. (ed.) (2007). *Careers of University Graduates. Views and Experiences in Comparative Perspectives*. Dordrecht: Springer.
- Thurrow, L. C. (1975). *Generating Inequality*. New York: Basic Books.
- Tuijman, A. (1999). Research agenda for lifelong learning: A report by the Task Force of the International Academe of Education. In Tuijman, A. and Schuller, T. (eds.) *Lifelong Learning Policy and Research*, pp 1–22. London: Portland Press.

Further Reading

- de Weert, E. (1999). Contours of the emergent knowledge society: Theoretical debate and implications for higher education research. *Higher Education* 38(1), 49–69.
- Elder, G. H., Kirkpatrick, M. J., and Crosnoe, R. (2003). The emergence and development of life course theory. In Mortimer, J. T. and Shanahan, M. J. (eds.) *Handbook of the Life Course*, pp 3–19. New York: Kluwer/Plenum.
- Enders, J. and Fulton, O. (eds.) (2002). *Higher Education in a Globalising World. International Trends and Mutual Observations*. Dordrecht: Kluwer.
- Esping-Andersen, G. and Regini, M. (eds.) (2000). *Why Deregulate Labour Markets*. Oxford: Oxford University Press.
- Teichler, U. (2002). Diversification of higher education institutions and the profile of the individual institution. *Tertiary Education and Management* 8, 199–216.

Higher Education and the Transformation of Society

P Scott, Kingston University, Kingston upon Thames, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The expansion of the number of students in higher education and the development of mass higher education systems during the past 50 years have been accompanied by a number of profound socioeconomic (and scientific and cultural) changes. These included the intensification of the democratic revolution (and a deeper process of democratization which has transformed the basis of class society); the completion of the educational revolution (which began in the nineteenth century and reached its climax with the development of near-universal secondary school systems after 1945); individualization (a complex and contradictory phenomenon combining the extension of civil rights – e.g., to universal welfare-state services – with the accentuation of market-led consumerism); the triumph of (mainly scientific and technological) expertise and the more pervasive phenomenon of professionalization which has reached into every nook and cranny of modern society (already apparent in the nineteenth century but reaching its apotheosis in the second-half of the twentieth century); and, most crucially, the development of a post-industrial and services-intensive economy in the many developed countries (which tended to be the same countries that produced the most mature mass higher education systems). The list can be greatly extended – popular culture, mass literacy/media, urbanism, post-modernism, the emergence of risk society.

However, as with all social phenomena, the balance between causality, contingency, and coincidence is difficult to assess. Whether the expansion of higher education has been an important driver of these profound social and cultural changes or simply their consequence is an interesting question – but ultimately no more susceptible of a clear answer than another, more direct, question which is often asked “do rich and developed countries need extensive higher education systems to sustain their economic development, or does economic growth provide a surplus which allows these countries to afford such systems which are desirable for other, non-economic, reasons?” Economists may be obliged to continue to try to answer this question, if only to enable politicians to make rational choices between competing priorities within higher education. But other social scientists may have to limit themselves to the conclusion that the relationship between social change and the development of higher education is open and reflexive, multi-linear as well as multi-lateral, probably irreducible to a series of unambiguous causalities.

One reason for this is simply that on the wider canvass of social transformation, in practical terms, there is not the same pressure to make rational choices between competing priorities as there is in the narrower domain of economic development. But a second reason is that the expansion of higher education has been surprisingly poorly articulated in more theoretical terms. By and large it is a phenomenon that has failed either to develop or to attract a convincing body of social theory. Sociologists, cultural historians, anthropologists, and others have typically preferred to generate their theories, to seek to explain the larger and wider connections between social transformation and cultural and political change, in domains other than their own home territory, the academic system. Only a few, such as Talcott Parsons and Gerald Platt in the 1970s (Parsons and Platt, 1973) and, in a narrower sense, Burton Clark in the 1980s and 1990s (Clark, 1983), have attempted to interrogate the university itself; typically, such efforts have not been sustained in any systematic and scholarly tradition but have remained one-offs, however intriguing. As a result, the big pictures of the links between the expansion of higher education and wider social change have been painted either by policymakers, institutional leaders, and other practitioners or – once again – by economists largely preoccupied by the application of concepts of human capital to understanding the explosive growth of universities and colleges. In recent years, these concepts of human capital have been incorporated within wider theories of economic development within the so-called knowledge society, which may have blunted their sharp economic edge but nevertheless remain limited as general explanations. Higher education, the *habitus* of social theorists, remains a largely theory-free zone.

In this article, the links between higher education and social transformation are considered under five headings: (1) democratic participation – the relationship with the rest of the education system, and the rise of a mass graduate class; (2) civil society – and, in particular, the links between the university and the city; (3) clever cities – the role of modern higher education systems in developing creative communities; (4) knowledge society – the fundamental changes in occupational, and therefore social, structures as a result of the development of post-industrialism; and (5) new patterns of knowledge production, which have reshaped notions of expertise and scientific/technical literacy.

Democratic Participation

In most of Europe, North America, and many other developed regions of the world universal systems of elementary education were first established in the second-half of the nineteenth century. Half a century or approximately two generations later, universal systems of secondary education were established. A quarter of a century (or one generation) later, after 1960, there was in most of these countries a decisive shift away from closed and elite university systems to more open and mass systems of higher educations (Scott, 1995). Although these mass systems did not become truly universal systems until the end of the twentieth century, this sequence depicts the quickening pace of the educational revolution. In some less well-developed regions of the world, the sequence was a little different, as mass higher education systems developed precociously before universal secondary education was achieved.

Two major processes appear to have been at work – a push process which meant that the establishment of much more extensive systems of education for younger age groups inevitably fueled demand for the expansion of opportunities for older age groups (an iron law of chronology so-to-speak); and a pull process because the capacity to participate fully in an increasingly sophisticated society came to depend on higher levels of educational achievement (in terms both of cultural capital and of formal skills and qualifications). These two processes are linked by a third, the complex (and controversial) phenomenon of up-skilling and/or dumbing-down whereby individuals require enhanced skills and superior qualification merely to retain their comparative positions in the occupational structure (and social hierarchy). The extent to which higher education remains, in Hirsch's classification, a positional good remains unresolved – and irresolvable (Hirsch, 1976).

The first process, the iron law of chronology, is well understood. Indeed, it can be measured in the case of the transition from secondary to higher education; the number of secondary school leavers with qualifications that entitle them to enter higher education (whether as of right, as in many continental European countries or as candidates for consideration, as in the United Kingdom, the United States, and Australia) has increased. The major driver of increased participation in higher education in most developed countries has been this increasing pool of eligible candidates. However, this has been a qualitative as well as a quantitative phenomenon. For example, a growing proportion of this increasing pool of higher education entrants has been women; indeed the equalization of participation rates between men and women has been not only a major cause of the expansion of higher education but also a manifestation of a fundamental shift in

gender identities in these societies (the graduate woman became an iconic figure in late-twentieth-century society). Another example of qualitative change has been the changing nature of the higher education systems to which this increasing number of secondary school leavers have gained access; a core of traditional universities has been supplemented by a growing penumbra of other post-secondary education institutions (*Fachhochschulen* in Germany, HBO (higher professional) schools in the Netherlands, as well as the former colleges of advanced education in Australia and polytechnics in England and Wales) while these universities themselves have been transformed into more open institutions.

The second process, the acquisition of the cultural capital necessary to function effectively in a democratic society, is more complex. It has been argued by some scholars that mass, or universal, systems of education make a reduced contribution to social mobility because they provide less scope for the co-option of the best and brightest, of whatever social origins, into elite structures. As a result, their emergence has had conservative, even regressive, effects. However, the contrary (and more convincing) view is that the threshold of active civic participation has been raised as surely as the threshold for effective participation in the labor market, and economy more generally. Mass higher education systems are necessary to provide the means for such threshold participation; without the expansion of universities and colleges over the past half-century, a growing proportion of young (and older) people would have been disenfranchised in sociopolitical terms as well as disempowered in economic terms. However, the growth of a (perhaps predominant) graduate class cannot be regarded as a largely reactive phenomenon, a response to the need to accumulate increasing stores of cultural (and social) capital which has been determined by other forces of social change; mass participation in higher education has itself transformed notions of such capital, changing the rules of engagement between the academic system and society at large for ever.

Civil Society and Urban Culture

Civil society, that crucial layer of institutions which mediates between the citizen/consumer and the state/market, has been to some degree de-populated in contemporary society, although social scientists and commentators are divided as to the causes. Some, typically on the left, blame the *anomie* of the market while others, generally on the right, blame the dependency culture produced by an intrusive welfare state. But they appear to agree that traditional civil-society institutions such as local government agencies, trade unions, and churches (in most but not all countries;

the United States is a notable exception) have been attenuated. Moreover, it is not clear that the civil-society institutions which may have replaced them – the mass media with its celebrity culture on the one hand or new social movements with their shifting allegiances (and unstable institutionalization?) on the other – have the same mediating influence. However, mass higher education systems may be an exception, although universities and colleges are not conventionally regarded as elements within civil society. The emergence of mass graduate populations, bound together by something resembling a common culture (or, at the very least, shared lifestyle experiences) may represent a crucial mediating layer in contemporary societies.

This is most evident in the interaction between the mass university and the modern city, whether the mega-cities of the developing world or the world cities of the developed world (Scott, 2005). Both are key sites of social transformation – and both universities and cities have spread far beyond their spatial boundaries; their normative influence is now pervasive. Although some twentieth-century campus universities were established in semi-rural or small-town locations (in, perhaps unconscious, celebration of the visual and spatial utopianism of nineteenth-century American liberal arts colleges), the university has always been a predominantly urban institution. Nor has this association, historical and contemporary, between the university and the city been purely accidental. The city has always been a site of radical experimentation – in social, economic, technological, and cultural terms. This was true in the classical world, the medieval and early modern periods, the Industrial Revolution – but never more so than under the hyper-urbanizing conditions of contemporary society with its mega-cities/world-cities where the impulses of the market and of mass culture are most keenly and intensely felt (and, consequently, where the need for novel forms of social mediation – such as the presence of large-scale graduate populations – is most necessary). The university has been an equally radical source of innovation – in scientific and intellectual terms. Although there have been quieter, and more quiescent, periods – the eighteenth century perhaps – in the long history of universities, the dynamism of the modern university is not in doubt. It has come to dominate the social landscape through the development of mass higher education systems (and the mass institutions that are so prominent in contemporary cityscapes); and it has an equally commanding role in the determination, and reproduction of, technical and professional expertise.

The histories of both the university and the city (in its widest sense), therefore, demonstrate remarkable affinities – which over time have become more intense (and continue to intensify). Academic life in all its manifestations – from the street life of students, through donnish society and the influence of universities on cityscapes, to their impact on the formation of expert professions and the generation of scientific and technical knowledge – has

always been a major contributor to the emergence of broader urban cultures (which both determine and reflect the social structures of our now predominantly urban, even hyper-urban, societies). The normative structures of city and university were, and are, also similar. Both encouraged the emergence of individualism and secularism, skepticism and rationality which became the key values of modernity (Bender, 1993). Arguably, this dynamic engagement between the city and the university continues even more intensely in the interaction between the post-industrial and consumerist social structures characteristic of the contemporary city and the post-modern (or, at any rate, more febrile and heterogeneous) intellectual cultures on display in the modern university.

Yet, important changes have taken place in the respective positions of the city and the university. The twenty-first-century city, particularly the mega-cities/world-cities, lacks the coherence, organizational and normative, even of the metropolitan cities of the nineteenth and earlier twentieth centuries. A favorite, and revealing, metaphor is of the shopping mall, a post-modern version of the arcades through which the *flâneurs* of the pre-industrial city wandered (Harvey, 1990; Benjamin, 2002). As with civil society, more broadly traditional civic cultures have been attenuated – first, by the rise of the national bureaucratic state which reached its apotheosis in the welfare states of the second-half of the twentieth century; and, more recently, by the impact of free-market and mass-media globalization. As a result, the role played by the city in social transformation, although even more powerful, is also more ambiguous and volatile. Once the university was one among a number of civic institutions, but, by no means the most powerful – and one of a range of ingredients of civil society (the two words, of course, come from the same root). Today, in contrast, the university has become in many cities the dominant civic institution, its role enhanced rather than attenuated by the rise of national bureaucracies, the welfare state, and globalization.

The strength of the university is apparent in pragmatic terms. First, it is typically among the largest employers, and makes a decisive contribution to the economy of the city and its region. Second, it is often an important locus for civic culture, both traditional and experimental – as the provider of theaters, galleries, and museums in its own right; as a corporate patron of the city's own cultural institutions; and as a sponsor, or source, of the audiences that sustain cultural activities and events. Third, the university is an intermediary, or more active mediator, between national (and now global) markets and cultures on the one hand and local contexts and civic environments on the other – or, perhaps more accurately, it has as important a role as financial institutions or mass-media corporations in these interactions (sometimes clumsily labeled glocalization). But the dominant role played by the university in the modern city is also normative. First,

although it plays an increasingly central role in market exchanges, the university is the most powerful intermediary between global and local in the non-market domain, in civil society. Second, the city itself having abandoned its traditional role of articulating the civic, or the public, in the pursuit of global economic success (or lacking the institutions to articulate them), the university has become the primary agent for such articulations. In both pragmatic and normative terms, therefore, the mass university plays an absolutely crucial role in social transformation.

Clever Cities

A particular manifestation of that role is to be found in a phenomenon labeled clever cities, after the work of Richard Florida. He argues that entrepreneurs, wealth-makers of all kinds, social experimenters, and cultural innovators flourish best in creative communities. This is no longer a Bohemian phenomenon largely restricted to artists, although the creative and cultural industries are among the fastest-growing, and most distinctive, sectors in the most dynamic contemporary economies; it also embraces designers, scientists, engineers, bankers, even the public at large (Florida, 2005). So these communities have a double impact. They not only provide a social and cultural milieu, an environment, in which other forms of enterprise, both business and social, flourish, but also create, through the cultural industries, their own powerful forms of wealth generation.

Universities are key to the building of creative communities – and so of clever cities. The influence of universities is felt directly – through the generation of new technologies and the production of skilled graduates. These technologies and graduates, in turn, produce intellectual property of various kinds – patents, spin-out (and spin-in) companies, consultancy capacities – and also cultural property (plays, music, novels, and other cultural artifacts). But the influence of universities is also felt indirectly – because they create the cultural ambience and social infrastructure which creative people value (for example, theaters, museums, schools, even restaurants); in short university cities are also clever cities.

However, the ways in which mass higher education is transforming the social base of these cities are more complex than the socioeconomic (and cultural) impact of traditional universities on cities of the industrial age. In both periods, universities educated future elites, both political and cultural elites, but in particular professional and technical elites. As the urban middle-class in the course of the nineteenth and twentieth centuries succumbed to the twin forces of credentialization and professionalization (or, to put it more positively, came to be determined more in terms of educational attainments and professional skills

and less in terms of social origins), the modern city took on distinctive characteristics – university quarters with a raffish, even Bohemian, quality as well as solid suburbs inhabited by the new university-trained professional classes. Both continue to be characteristic of the social structure of post-modern cities, although the raffishness of university quarters has intensified and professional structures have become both more open and less secure.

But the social impact of the mass university is also more far reaching – and also more volatile. First, mass graduate populations have penetrated more deeply into the urban environment, influencing not only elites but also the general population. As a result, a distinctively graduate culture has become pervasive – and may have ceased to provide such a significant social demarcation (the bartender may well be a graduate like her/his customers). Second, these graduate populations are less rooted. Not simply the elites (international bankers or designers) but the masses are often highly mobile (adding to the ephemeral quality of the post-modern city). This mobility, attributable in part to the development of mass higher education systems, has had a profound effect on housing patterns, employment structures, even the patterning of social life, consumption habits, and leisure pursuits. Third, the creativity which Florida celebrates is produced not simply (or even mainly) by traditional forms of scholarly enterprise, scientific research, and professional practice; it is also produced, often serendipitously, by the contingencies of intellectual playfulness and post-modern lifestyles. Silicon Valley and its worldwide clones are characterized by highly unconventional social and economic structures. The college dropout, the alternative entrepreneurs, the camp-followers of mass higher education – such are the creative people that power some of the world's most significant clever cities.

Knowledge Society

The idea of a knowledge society has gone through a number of intriguing iterations. It began in the 1970s under the label post-industrial society, a relatively unproblematic characterization of the shift from an economy dominated by extractive and manufacturing industries to one in which services, and increasingly symbolic goods, made a more prominent contribution – although Daniel Bell attempted to highlight the social and cultural consequences of this shift (Bell, 1976). This first iteration coincided with the emergence of the first mass higher education systems, initially in the United States and later, and more tentatively, in Western Europe, Latin America, Australasia, and some Asian countries. The expansion of universities reflected, and reinforced, these new occupational structures – both meeting the demand for new kinds of professional workers, predominantly in

services (and at this stage, largely in public services rather than commercial services, which explains the tight association between the welfare state, at its apogee in the 1970s, and the first stages of mass higher education) and also reducing the supply of now obsolescent industrial workers by keeping young people out of the labor market for longer (and, of course, producing the technician and higher-level engineers required for the functioning of increasingly high-technology manufacturing industry).

Later, the knowledge society took on new forms – for example, the network society of Manuel Castells (Castells, 1996, 1997, and 1998) and the risk society of Ulrich Beck (Beck, 1992). The first iteration emphasized the importance of connections and connectivity (powered, of course, by the information and communication technology revolution) – and, at any rate by implication, the relative decline in importance of the institutions which were being connected in these novel and powerful ways; to that extent the idea of a network society captured the process of de-institutionalization, the so-called hollowing-out of government (through privatization), of industry and business (through outsourcing) and even established professions (because of competing notions, and alternative locations, of expertise). The second iteration, the risk society, highlighted the accumulation of uncertainties in a society in which the intensification, and acceleration, of knowledge produced alarming new forms of social and cultural volatility. These more recent iterations of the knowledge society coincided with the maturity of many mass higher education systems, now outputting large-scale graduate populations defined as much by their lifestyle, and social and cultural capital, as by their technical or professional expertise. It is probably not accidental that this period, the 1990s onward, also saw the absorption of formerly distinct systems of higher technical education into these mass systems of higher education (and, in some countries such as the United Kingdom and Australia, the abandonment of the distinction between traditional universities and institutions of professional education).

These policy shifts may reflect significant changes in the articulation between higher education and the social base. Once discrete groups of graduates – social, political and cultural elites; higher professional workers; technical experts; the now more highly educated populace – have tended to coalesce into a less well-differentiated mass (or, perhaps more accurately, once discrete roles have become fuzzier). As a result, the need for special-purpose institutions of higher education, such as universities or (in England and Wales) polytechnics, has diminished. Changes in the structure of higher education, therefore, mirror wider changes in social structures in the knowledge society. Mass higher education systems and socioeconomic structures now touch at several different points – the traditional role played by universities in the

formation of future elites, the re-production of the established professions, and the development of technical expertise (although all three have become more heterogeneous, and even contested, categories for reasons that are discussed in the next section of this article), has been both reinforced and subverted by the growth of graduate culture, that unstable combination of lifestyle preferences and the (at times defensive) accumulation of social and cultural capital needed to maintain positional advantage in a knowledge society.

The conventional view, therefore, that there is a straightforward and comparatively linear relationship between the development of mass higher education and the emergence of knowledge society – essentially, that in a society powered, and pervaded, by knowledge there is a corresponding need for an increase in the number of not only knowledge, that is, technical, experts but also of knowledgeable actors (and that modern higher education systems are near-monopoly producers of both) – is perhaps too simple. Instead, there are multiple, ambiguous, and maybe conflicting relationships. It is as a total phenomenon that mass higher education is contributing to social transformation, as a powerful social fact and cultural presence in its own right, as much as a series of inputs that are contributing to social change such as better-educated citizens or more highly skilled workers – being as much as becoming. Its presence is perhaps more normative than instrumental.

New Patterns of Knowledge Production

One of the most enduring classifications of higher education systems was that devised by Martin Trow (Trow, 1973). He suggested a three-stage process of evolution – elite higher education systems that enrolled up to 10% of the eligible population; mass systems that enrolled up to 40%; and universal systems that enrolled more than 40% (and came to embrace a majority of young people). This classification has been criticized, partly on the grounds that it suggested a tidier linear sequence than was actually the case and partly because it focused on the initial higher education of young adults and therefore failed to take account of the growing phenomenon of lifelong education available to all age groups. Nevertheless, it has remained a useful classification. Indeed, it is tempting to align elite systems with the education of Robert Reich's symbolic workers (Reich, 1991, 2001), the old social and cultural elites dressed in new knowledge society clothes perhaps; mass systems with the training of the burgeoning cadres of knowledge workers (not all necessarily with highly technical expertise), for which the new economy of the late twentieth and early twenty-first century has an insatiable appetite; and universal systems with the development of an

even more extensive graduate culture in which all citizens, and all workers, have become knowledgeable actors.

However, to describe the social impact of modern higher education systems in such a tidy, simple, and (by implication) hierarchical taxonomy may be misleading. Studies of new patterns of knowledge production suggest the need for a more nuanced account of the transformative effects of higher education. For example, it has been argued that the more closed and hierarchical knowledge production systems of the past have been overlaid, if not replaced, by more open and fluid systems (Gibbons *et al.*, 1994; Nowotny *et al.*, 2001). Consequently, it has become more difficult to define the primary (and, therefore, privileged) producers of new knowledge, partly because the complex, interactive, and even reflexive characteristics of innovation in the new economy are now better understood and partly because of the playful and serendipitous quality of creativity in a knowledge society, which has already been discussed. This has important implications for the role of higher education – and, in particular, for traditional research universities. Their primacy may be less marked than is often assumed, for example, in the politically dominant discourse about world-class universities.

Conversely, mass higher education institutions may flourish in more open knowledge production systems and make a greater social impact – first, because they develop the socially distributed expertise needed to empower workers and to enfranchise citizens and transform both into knowledgeable actors; and second, because these institutions are more familiar with notions of diversity and difference (and, therefore, may promote an intellectual flexibility need to manage this socially distributed expertise which is inherently problematical – and often contested as rival expertises clash). However, another implication of the emergence of more open knowledge production systems is that the (social) contextualization of knowledge has become a key attribute not only of its relevance but also its novelty. As a result applied research has been replaced as a category by knowledge produced in the context of application (or even implication – which hints at the creative, even visionary, qualities of post-modern knowledge).

The result could be to erode the organizational base of all higher education institutions, even the most open (in terms of intellectual traditions and research practices) and accessible (in terms of their social constituencies), as they confront more competition from other knowledge institutions (indeed, in a knowledge society it becomes axiomatic that all institutions must become to some degree both learning and knowledge-producing organizations). In turn, this could compromise their capacity to shape the social landscape as decisively in the future as in the past. At the very least, the relationship between higher education and social transformation (which, of course, was always mediated through other institutions, notably the

labor market) is likely to be dependent even more on higher education's interdependences with other knowledge institutions – thereby reflecting the importance Castells attaches to networks and connectivity.

Conclusion: Looking to the Future

The relationship between higher education and social transformation – and, in particular, between mass higher education systems that first emerged in the 1960s in the United States and have now become the dominant global model and the emergence of a knowledge society (and the new economy, equally a global phenomenon) – is difficult to reduce to a small number of simple formulas. But two general conclusions may perhaps be possible conceptually as well as institutionally:

1. The first is that the relationship has become more intense and more direct. Mass higher education systems help to shape their societies more powerfully than the elite systems of the past – not simply quantitatively, because many more people are embraced within these systems, but also qualitatively, because social structures are both more finegrained and more volatile (and higher education plays a more active role in determining not only the division of labor but also the distribution of cultural capital).
2. The second conclusion is that the relationship is more complex, even problematical. Straightforward linear producer models no longer adequately describe the social impact of higher education, whether in research or teaching. More reflexive models of the engagement between higher education and society are needed. Mass graduate culture, socially distributed expertise, playful creativity, open innovation systems, and other key characteristics of contemporary society, most certainly can be traced back to the influence of modern higher education systems. But attempts to delineate these relationships more precisely are likely to fail – not simply, for empirical reasons, because of the lack of research and paucity of data (there is a super-abundance of both, particularly the latter), but mainly for structural reasons because these systems are now so interwoven that they cannot be disentangled.

See also: Higher Education and the Labor Market; Universities and Regional Development.

Bibliography

- Beck, U. (1992). *Risk Society: Towards a New Modernity*. London: Sage.
- Bell, D. (1976). *The Cultural Contradictions of Capitalism*. London: Heinemann.

- Bender, T. (1993). *Intellect in Public Life: Essays on the Social History of Academic Intellectuals in the United States*. London: Johns Hopkins University Press.
- Benjamin, W. (2002). *The Arcades Project*, Eiland, H. and McLaughlin, K. (trans.). Harvard, MA: Belknap Press for Harvard University Press.
- Castells, M. (1996, 1997, and 1998). *The Information Age: Economy, Society and Culture* (Vol. 1: *The Rise of Network Society* [1996], Vol. 2: *The Power of Identity* [1997], Vol. 3: *End of Millennium* [1999]). Oxford: Blackwell.
- Clark, B. (1983). *The Higher Education System: Academic Organisation in Cross-National Perspective*. Berkeley, CA: California University Press.
- Florida, R. L. (2005). *Cities and the Creative Class*. London: Routledge.
- Gibbons, M., Camille, L., Helga, N., et al. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: Sage.
- Harvey, D. (1990). *The Condition of Postmodernity*. Oxford: Blackwell.
- Hirsch, F. (1976). *Social Limits to Growth*. Cambridge, MA: Harvard University Press.
- Nowotny, H., Peter, S., and Michael, G. (2001). *Re-thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge: Polity Press.
- Parsons, T. and Gerald, M. P. (1973). *The American University* (with the contribution of Neil M Smelser). Cambridge, MA: Harvard University Press.
- Reich, R. (1991). *The Work of Nations: Preparing Ourselves for 21st-Century Capitalism*. London: Simon and Schuster.
- Reich, R. (2001). *The Future of Success*. London: Heinemann.
- Scott, P. (1995). *The Meanings of Mass Higher Education*. Buckingham: Open University Press.
- Scott, P. (2005). The university and civil values. In Robinson, S. and Clement, K. (eds.) *Values in Higher Education*, pp 8–23. Llantwit Major: Aureus Press and the University of Leeds.
- Trow, M. (1973). *Problems in the Transition from Elite to Mass Higher Education*. Berkeley, CA: Carnegie Commission on the Future of Higher Education.

Quality Assurance in Higher Education – Practices and Issues

D Dill, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Academic accreditation – Assessment of an institution's or a program's capacity for quality with a binary judgment about the attainment of threshold academic standards.

Academic audits – Assessment of the processes institutions use to assure themselves that their chosen standards are being achieved.

Academic standards – The specific levels of knowledge, skills, and abilities that students achieve as a consequence of their engagement in a particular academic program.

Culture of compliance – Institutions' investment of time and effort to satisfy external assessors rather than on active efforts to assure and improve academic standards.

External examining – Assessments by external peers of the performance of students on subject examinations.

External academic quality assurance – Supra-institutional policies and practices whereby the quality of higher education institutions and programs are assured.

National graduate examinations – Nationally designed and administered exams in all major subject fields for university graduates.

Internal academic quality assurance – Policies and practices whereby academic institutions themselves monitor and improve the quality of their education provision.

Market regulation – Consumer sovereignty in which informed student choice of academic programs helps assure academic standards.

National qualifications frameworks – Broad descriptors of learning outcomes specific to each level of academic degrees.

Professional regulation – Producer sovereignty in which academics themselves are principally responsible for defining and enforcing academic standards.

State regulation – State sovereignty in defining and enforcing academic standards.

Student experience surveys – Surveys of students' or graduates' experiences in their academic programs known to be associated with effective learning.

Subject assessments – Systematic evaluations of the quality of delivered performance of study programs with emphasis on curriculum, teaching, and program relevance to graduates and the economy.

Subject benchmarks – Publicly defined academic content and threshold standards for each subject field.

The Emergence of New Quality-Assurance Practices

At the close of the twentieth century, the national policy frameworks for higher education institutions underwent substantial reforms. In the emerging global economy, advanced human capital has become a crucial factor in economic development and a central component of a nation's competitive advantage. Consequently, many countries are shifting from elite to mass systems of higher education (Trow, 2005).

The combined impacts of globalization and massification have radically altered the traditional relationship between the state and institutions of higher education and motivated policymakers to seek new means for assuring academic quality in higher education (World Bank, 2002). First, the global demand for skilled human capital encouraged changes in the degree frameworks of many countries as policymakers sought international recognition of the credentials granted by their country's higher education institutions. These new degree frameworks also encouraged a rapid proliferation of new academic programs in many countries, thereby testing established national practices for assuring academic standards. Second, the rapid growth of higher education systems has provided incentives for the development of private institutions, including cross-border franchise and virtual universities, which have posed novel challenges to national systems of external quality assurance (QA), particularly those based upon central control of public institutions. Third, the competitive forces unleashed by globalization and massification have required institutions of higher education to become more responsive to rapidly changing labor markets and to student program interests. Consequently, institutions in many countries

have sought increased flexibility and autonomy from traditional state QA regulations so that they can react more swiftly to changing social demands by establishing new academic programs, reconfiguring existing programs, and eliminating outdated programs. Finally, the rapidly expanding social demand for higher education has been caused in large part by students' desire to achieve the increasing private benefits available to individuals with higher degrees. The empirical reality of the growing private benefits of academic degrees has altered the traditional debate about higher education finance, encouraging many countries to require students and their families to pay a larger share of higher education costs. Consequently, the new practices of external QA also seek to respond to public concerns that institutions provide educational value for money.

This dramatically altered environment of institutions of higher education helped reveal the inadequacy of both the traditional internal and external practices for assuring academic standards (Brennan and Shah, 2000). Therefore, in their search for a national framework that will encourage innovation in academic programs while maintaining and improving academic standards, policy-makers are experimenting with many innovative forms of academic QA.

The first government experiments with new QA practices occurred predictably in the US, an early exponent of mass higher education. Concerned with evidence of declining academic standards in public education, the majority of US states, in the early 1980s, adopted regulations requiring that publicly supported universities develop explicit plans for assessing student teaching (Dill *et al.*, 1996). Subsequently, new national QA policies were also introduced in France (1984), the United Kingdom (1985), and the Netherlands (1985) (van Vught and Westerheijden, 1993). The French government was primarily interested in reducing its dysfunctional QA bureaucracy, the government in Great Britain sought to achieve a better linkage of higher education with the labor market, while the Netherlands adopted a new QA framework in association with an innovative approach to

steering universities. The developments in these pioneering countries were then diffused to other countries in Europe, Asia, and eventually around the globe.

The traditional national frameworks for external QA varied from country to country, but had generally followed three modal forms: the European model of central control of QA by state educational ministries, the US model of decentralized QA combining limited state control with market competition, and the British model in which the state essentially ceded responsibility for QA to self-accrediting universities (Dill, 1992). In the UK, up until the actions by the Thatcher government in 1981, the assurance of academic quality in the publicly supported university sector was delegated to the academic profession itself, which monitored and assured the standard of university degrees through collective mechanisms such as the external examiner system. In contrast, ministries of education on the continent were much more active in setting standards for universities. They established and monitored regulations on university admissions, academic appointments, program curricula, and endpoint examinations. In the US, as higher education rapidly expanded following World War II, the federal Congress explicitly adopted a market-based approach to academic QA as a supplement to the existing tradition of state licensing and voluntary institutional as well as program accreditation. During the 1972 re-authorization of the Higher Education Act, members of the Congress argued that providing federal, financial assistance directly to students rather than to institutions was the most efficient and effective means to both equalize opportunities in higher education and harness market forces for enhancing academic quality.

Table 1 outlines representative external QA practices organized by the locus of authority over academic QA. Professional or self-regulation clearly assumes producer sovereignty in which academics themselves are principally responsible for defining and enforcing the rules and norms assuring the quality of academic provision. As noted, this places greatest emphasis on traditional voluntary practices carried out by professional bodies

Table 1 External assurance of academic quality

<i>Locus of authority</i>	<i>Professional (self) regulation</i>	<i>State (direct) regulation</i>	<i>Market regulation</i>
Practices	Professional accreditation and licensure Voluntary institutional accreditation External examining	National qualifications frameworks Subject assessments State-conducted accreditation Academic audits Performance-based funding or contracting National examinations or surveys	Commercial information provision, for example, institutional or program performance data, assessments, and rankings

including accreditation of academic programs and institutions by professional associations as well as collective professional practices such as external examining. State or direct regulation of academic quality assumes the sovereignty of the state in defining and enforcing academic standards. The new QA instruments emphasized by the state include the definition of academic degree frameworks, policies introducing new assessment practices such as academic audits and/or subject assessments designed to maintain and improve internal QA practices, state-conducted accreditation of programs and/or institutions, performance funding and contracts, and finally, regulations influencing the public provision of academic information such as state-mandated exams or surveys. For the market to work effectively as a means of assuring academic standards, it is necessary for students and their families to achieve effective consumer sovereignty through informed choice of academic programs. QA practices associated with this perspective include commercially produced rankings and student guides designed to provide academic quality information to students.

Several key points can be derived from these simple distinctions. First, a number of QA practices such as accreditation or academic audit are essentially generic processes that can be conducted voluntarily under the auspices of academic professional organizations such as the American Board of Engineering and Training (ABET) or the European University Association (EUA), or can be a requirement of national policy carried out by agencies established by or affiliated with the state. Similarly, quality rankings can be produced by the academic profession as in the world university league table published by the Shanghai University, by the state as in the UK subject assessments, or by the private sector as in the commercially produced rankings of the *London Times* or the *US News and World Report*. Second, while it is often argued that professional self-regulation or market forces represent serious alternatives to state regulation of academic standards, the reality is that professional or market-based QA practices are usually highly dependent upon the state for their effective functioning. If professional self-regulation or market forces are to successfully protect the public interest in the assurance of academic standards, they must be reinforced by law or formally recognized and/or subsidized by the state. For example, the current influence of voluntary accreditation in the US derives almost entirely from the fact that the national government utilizes institutional accreditation to determine college and university eligibility for federal student aid. Similarly, more valid commercial rankings such as those of the *Guardian* in the UK or the *Good University Guide* in Australia are highly dependent upon government subsidized or produced data on universities. In short, effective professional self-regulation and/or market regulation are best understood as alternative state instruments for

assuring academic quality. The challenge confronting all nations is to design a policy framework that effectively and efficiently utilizes the forces of the state, the academic profession, and the market to assure academic standards.

An Assessment of External QA Practices

The sections that follow review the primary new practices of external academic QA and the issues they raise for higher education.

National Qualifications Frameworks

In response to the changing environment of higher education, a number of countries, such as Australia and Ireland, have adopted new national qualifications frameworks (Young, 2003). The cross-national Bologna framework of bachelors, masters, and research doctoral degrees, the Dublin Descriptors, and the UK Graduate Standards Program can also be understood in these terms. The initial rationale for these frameworks was to provide international recognition for academic degrees to aid in attracting foreign students as well as to help in placing program graduates in the global market. However, by providing broad descriptors of learning outcomes specific to each level of academic degrees, academic qualification frameworks also provide some potential reference points for external QA practices (McInnis, 2005).

A more refined example of degree frameworks is the UK subject benchmarks program (Williams, 2005). The massification of higher education has also produced many new professional or more vocational university programs for which clear agreement on academic content and student learning outcomes is often lacking. In response to a growing concern about the comparability of academic standards across academic programs, the UK systematically created subject benchmark committees to publicly define appropriate academic content and threshold standards at the level of each subject field. Over a 5-year period, the UK committees developed subject benchmarks for programs enrolling the vast majority of first-level students.

Some policymakers clearly hoped and some academic staff clearly feared that national qualifications frameworks and subject benchmarks would become a regulatory device for assuring the fitness of purpose of academic degrees. While the new frameworks do play a role in external QA, the fact that the higher education component of these frameworks and benchmarks is usually defined by academic staff means that the frameworks' impact on academic standards is limited and indirect. In the UK, the impact of subject benchmarking proved more formative and developmental than regulatory (Williams, 2005) – helping to define and legitimize new academic

subjects, generating discussion about appropriate academic standards at the subject and university level, as well as helping to strengthen internal university processes for new course approvals and academic QA. Similarly, the most significant contribution that qualifications frameworks make to an overall national QA system is to help encourage a focus on student learning outcomes rather than course content in national debates about academic standards.

Quality Assessments

One of the most significant changes in national QA frameworks at the end of the twentieth century was the emergence of the evaluative state (Neave, 1988). Many national governments initiated and/or subsidized the creation of new agencies and practices designed to assess quality in existing higher education programs and institutions. The new assessment practices included academic audits, subject assessments, and new forms of academic accreditation.

During the 1980s, the majority of the US states adopted assessment regulations designed to encourage public institutions to place greater attention on the improvement of student learning. However, these regulations delegated implementation responsibility to institutions, the institutional responses were not externally assessed, and the regulations had little measurable impact on academic quality (Peterson *et al.*, 1999). Similarly, traditional institutional accreditation in the US was criticized for being too comprehensive in its scope, overly focused on input criteria rather than policymakers' concerns about academic standards, and insufficiently transparent in that accreditation reviews are not required to be made public (Dill *et al.*, 1996). Existing voluntary program accreditations did not assess the education programs in which the majority of students were enrolled, accrediting primarily professional fields. Under mounting pressure from policymakers concerned about declining academic standards, all of the institutional accreditation agencies adopted student assessment as one of their standards for review and several accreditation agencies introduced new accreditation review processes. However, after several decades of experimentation, a United States Department of Education (2006) report concluded that US-style voluntary institutional and program accreditation were still inadequate mechanisms for assuring the quality of student learning.

Outside the US, new forms of quality assessment, including subject assessments, academic audits, and innovative approaches to accreditation, were mandated by many countries for all publicly supported institutions of higher education. Each of these practices adopted a similar sequence of activities – an institutional self-study, an external peer review, and a public report of findings – but the focus of each practice differed. Subject

assessments, as implemented in countries such as the UK, the Netherlands, and Denmark, involved systematic evaluations of the quality of delivered performance of study programs with emphasis on curriculum, teaching, and program relevance to graduates and the economy. In contrast, academic audit as implemented in countries such as Australia, the UK, and Hong Kong, focused on the processes that institutions use to assure themselves that their chosen standards are being achieved, what in Sweden is usefully termed education quality work. Accreditation, as implemented in Europe, is similar to US accreditation in its assessment of a program's capacity for quality with a binary judgment about the attainment of threshold academic standards. However, the European practices differed in their focus on study programs rather than institutions, in their comprehensive coverage, and in their attention as well to the effectiveness of program QA activities.

All three of these new external assessment practices had the effect of encouraging dialog and collaboration among academic staff regarding the improvement of student learning and assurance of academic standards within academic institutions (Kehm, 2006; Massy, 2005; Stensaker, 2004). This is not a negligible impact given the increasing incentives in all of higher education for academic staff to invest time and effort in research. However, if overly focused on external control rather than institutional responsibility for improvement, these assessments can encourage a culture of compliance in which institutions invest time and effort on developing policy documents and erecting quality infrastructure to satisfy external assessors rather than on active efforts to assure and improve academic standards. External subject assessments and program accreditation are also very costly to mount and sustain over time. Their focus is exclusively on the subject level and, therefore, these assessments also provide limited incentives for the overall institution to develop an effective internal QA process. Audits by contrast are much less costly, applicable to all types of institutions, and provide some of the same incentives for communication and collaboration on the improvement of teaching and learning. However, the potential positive impacts of academic audits may be limited if poorly designed, for example, by too comprehensive an assessment that includes other than core educational processes, or by focusing too much on QA documentation rather than on empirical evidence regarding the validity and reliability of internal processes for assuring academic standards.

The evolution of external examining in the UK, a professional external QA practice, illustrates a number of issues involved with quality assessments (Lewis, 2005). Of the many external QA practices, external examining, as conducted in the UK and some Scandinavian countries, most clearly assesses academic standards. External examiners traditionally assessed the actual performance of

students on subject examinations used to award honors degrees in UK universities. External examining emerged in early nineteenth-century England as a professional practice and was explicitly encouraged in subsequent university charters. However, the practice was not regulated or codified until, under pressure from the government to assure academic standards, the committee of vice chancellors and principals published the first code of practice in 1986. Subsequent research revealed that only a minority of universities were following these professional standards, challenging academic assertions on the effectiveness of the practice (Warren Piper, 2004). Government regulations were then introduced requiring summaries of university external examining reports to be made public and the governmentally supported UK Quality Assurance Agency (QAA) incorporated specific precepts on external examining into their academic audits of each university, thereby encouraging greater equivalency of practice across institutions. However, the widespread adoption of modular instruction and continuous assessment in much of UK higher education over the last decades has lessened the ability of external examiners to actually compare academic standards within a subject field. Instead, external examiners now perform the function of assessing the validity and reliability of the overall assessment regime in a particular subject.

Information

A third means of external QA is information provision. The development of performance indicators to help assure academic standards has become an important external QA practice in many countries (Cave *et al.*, 1997). Initially performance indicators were designed by policymakers to inform government funding, often in association with new higher education financial instruments such as performance-based funding or university performance contracts. As higher education has become more competitive nationally and internationally, indicators of academic quality have increasingly been published by government, academic, nonprofit, and commercial entities to better inform student choice of an educational program. Many policymakers also believe that informed student choice is an influential means of external QA. While government subsidies of higher education in every country are predicated on the belief that the educational value added by higher education programs produce beneficial outcomes for society, these social benefits are difficult to gauge directly, therefore external QA utilizes proxy measures of academic quality. These include: more immediate outcome measures such as graduate placement, salaries, and satisfaction with education program; output measures such as student test scores, completion rates, and marks; and various process measures such as student engagement.

Outcome measures, such as graduate placement and salaries, are informative and generally valid quality

information for potential students and could also be valuable general indicators of effectiveness for academic programs if used by institutions. Readily available output measures such as student marks or graduation rates may be unreliable indicators of academic quality because they can be increased by lowering academic standards. For this reason, performance-based funding or contracts, which are usually based upon available input, process, and output measures, have proven to be an inadequate instrument for assuring academic standards (Jongbloed and Vossensteyn, 2001) and need to be supplemented by government mandates on quality information provision and external quality assessments. While common graduate exams exist in certain professional fields, such as medicine or teaching, and have been utilized in association with external assessments by professional accrediting agencies in the UK and the US to improve academic standards, common exams do not exist in all subjects. Brazil attempted the most ambitious experiment in mandated common subject exams (Schwartzman, 2004), developing nationally designed and administered exams in all major subject fields for graduates of both public and private universities. The exams provided influential performance information, but could not effectively assess the actual value added by an academic program. Under strong political pressure from the universities, the exams were modified to a voluntary survey of a sample of graduates, thereby diminishing the validity and reliability of the performance information provided. Both Australia (graduate skills assessment – GSA) and the US (collegiate learning assessment – CLA) are experimenting with less costly and more easily administered tests of generic skills for first-level students. While these tests could provide a general indicator of the educational value added by institutions, the generic content of such exams might encourage the lowering of academic standards and in any event is unlikely to help inform academic staff about the means of improving academic standards in specific educational programs.

The limitations of traditional output measures and common exams or tests necessitated the development of common surveys of student experience in Australia and the US as a means of academic QA and these are spreading to other countries. The Australian course experience questionnaire (CEQ) surveys graduates' perceptions of teaching quality, skills learned, and their satisfaction with their education in their academic program (Harris and James, 2006). The US national survey of student engagement (NSSE) asks currently enrolled students to report on experiences in their educational program known to be associated with effective learning (Ewell, 2005). While generic instruments, both these surveys are based upon systematic research on effective teaching and student learning and therefore offer more valid and informative indicators of academic quality for potential students as well as academic staff.

Overall, the validity of public information on academic quality has become a controversial issue in higher education. Quality rankings by commercial publications often rely upon information gleaned from reputational surveys, input measures such as student test scores or financial resources, and indicators of research quality all of which have questionable validity as predictors of effective student learning (Dill and Soo, 2005). These rankings have become highly influential on academic behavior, often encouraging institutions and programs to invest time, resources, and effort in improving their rated reputations rather than in the challenging education quality work necessary to actually improve academic standards (Brewer *et al.*, 2002). More valid and useful academic quality information and rankings have been produced by not-for-profit entities such as the Center for Higher Education Development (CHE) in Germany, as well as the NSSE and National Research Council's (NRC) assessment of research doctoral programs in the US. A critical determinant of the legitimacy of academic quality information is government policy, which may subsidize the development and provision of more valid quality information as in the case of the NRC doctoral surveys in the US and the Australian graduate survey, and/or mandate publicly supported universities to participate in more valid surveys and rankings as in the case of the NSSE and CHE rankings.

Publicly available information on academic quality plays an important role in assuring academic standards, but ironically its import may be greater for internal than external QA. International research on student choice suggests that quality rankings and ratings influence the educational decisions of a relatively small segment of the student population, primarily those of high ambition and achievement (Dill and Soo, 2005). The education choices of most students are influenced by a wide variety of educational, social, and personal factors, which suggests that the individual decisions of even well-informed potential students are unlikely to be a strong influence on the assurance of academic standards within academic institutions. The more significant role of quality information is therefore likely to be its development and use by academic staff as part of institutional efforts to assure and improve academic standards.

Summary

The growing importance of human capital to the economic and social development of contemporary societies mandated the massification of higher education and subsequently motivated a search for new methods of assuring and improving academic standards. All countries, therefore, are seeking an effective national framework for academic QA that will likely include an appropriate classification of academic degrees, the provision of valid and reliable

information on academic quality, and some efficient means of external assessment designed to assure that institutions of higher education have in place effective internal QA processes. Degree frameworks and subject benchmarks can contribute to external QA by helping to redirect public and academic debate about academic quality from curricular issues to socially beneficial learning outcomes. In addition, some type of state-mandated external assessment of internal QA practices seems warranted, since traditional professional practices appear inadequate in the new environment of higher education. Centralized control of academic quality by state education ministries is impractical in mass systems. Professional practices such as voluntary academic accreditation in the US and external examining in the UK have noted limitations, and academic audits of internal QA in a number of countries have clearly revealed that these institutional practices can be improved. Comprehensive program accreditation practices, such as those being experimented with in Europe, can help assure the attainment of threshold academic standards following the adoption of a new degree framework. However, over time, the high cost and limited focus of regular program accreditations as well as subject assessments appear difficult to justify and sustain. National quality-assessment practices seem to be evolving out of necessity toward the adoption of some form of external academic audit of internal QA processes. Developing effective academic audit processes will therefore likely be a continuing and important challenge for the field of academic QA (El-Khawas, 2005). Finally, the public provision of valid and reliable information on academic quality will be a critical component of national QA frameworks. While commercial publications provide an increasing amount of information to the public on academic quality, they have limited incentives to publish rankings that will actually help maintain and improve academic standards. The development and public provision of valid and reliable indicators of academic quality therefore is best understood as a pure public good, which must be subsidized by the government.

See also: Employability of University Graduates and Graduate Outcomes; International Accreditation in Higher Education; National and International Rankings of Higher Education; Promoting Effective Student Learning in Higher Education; The Role of Accreditation of Higher Education Institutions.

Bibliography

- Brennan, J. and Shah, T. (2000). *Managing Quality in Higher Education: An International Perspective on Institutional Assessment and Change*. Buckingham: OECD, SRHE and Open University Press.
- Brewer, D., Gates, S. M., and Goldman, C. A. (2002). *In Pursuit of Prestige: Strategy and Competition in US Higher Education*. New Brunswick, NJ: Transaction Press.

- Cave, M., Hanney, S., Henkel, M., and Kogan, M. (1997). *The Use of Performance Indicators in Higher Education: The Challenge of the Quality Movement*, 3rd edn. London: Jessica Kingsley.
- Dill, D. D. (1992). Academic administration. In Clark, B. R. and Neave, G. (eds.) *The Encyclopedia of Higher Education*, pp 1318–1329. Oxford: Pergamon Press.
- Dill, D. D., Massy, W. F., Williams, P. R., and Cook, C. M. (1996). Accreditation and academic quality assurance: Can we get there from here? *Change* 28(5), 16–24.
- Dill, D. D. and Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher Education* 49(4), 495–533.
- El-Khawwas, E. (2005). *Accreditation of Teacher Education in the US: An Audit Approach to Subject Assessment*. Public Policy for Academic Quality Research Program. Chapel Hill, NC: University of North Carolina.
- Ewell, P. T. (2005). *The National Survey of Student Engagement (NSSE)*. Public Policy for Academic Quality Research Program. Chapel Hill, NC: University of North Carolina.
- Harris, K. -L. and James, R. (2006). *The Course Experience Questionnaire, Graduate Destinations Survey and Learning and Teaching Performance Fund in Australian Higher Education*. Public Policy for Academic Quality Research Program. Chapel Hill, NC: University of North Carolina.
- Jongbloed, B. and Vossensteyn, H. (2001). Keeping up performances: An international survey of performance-based funding in higher education. *Journal of Higher Education Policy and Management* 23(2), 127–145.
- Kehm, B. M. (2006). *The German System of Accreditation*. Public Policy for Academic Quality Research Program. Chapel Hill, NC: University of North Carolina.
- Lewis, R. (2005). *External Examiner System in the United Kingdom: Fresh Challenges to an Old System*. Public Policy for Academic Quality Research Program. Chapel Hill, NC: University of North Carolina.
- Massy, W. F. (2005). *Education Quality Audit as Applied in Hong Kong*. Public Policy for Academic Quality Research Program. Chapel Hill, NC: University of North Carolina.
- McInnis, C. (2005). *The Australia Qualifications Framework*. Public Policy for Academic Quality Research Program. Chapel Hill, NC: University of North Carolina.
- Neave, G. (1988). On the cultivation of quality, efficiency and enterprise: An overview of recent trends in higher education in Western Europe, 1986–1988. *European Journal of Education* 23(1/2), 7–23.
- Peterson, M. W., Augustine, C. H., Einarson, M. K., and Vaughan, D. S. (1999). Revolution or evolution? Gauging the impact of institutional student assessment strategies. *Change* 31(5), 53–56.
- Schwartzman, S. (2004). *The National Assessment of Courses in Brazil*. Public Policy for Academic Quality Research Program. Chapel Hill, NC: University of North Carolina.
- Stensaker, B. (2004). *Subject Assessments for Academic Quality in Denmark: A Review of Purposes, Processes and Outcomes*. Public Policy for Academic Quality Research Program. Chapel Hill, NC: University of North Carolina.
- Trow, M. (2005). Reflections on the transition from elite to mass to universal access: Forms and phases of higher education in modern societies since WWII. In Altbach, P. (ed.) *International Handbook of Higher Education*, pp 243–280. Dordrecht: Kluwer.
- United States Department of Education (2006). *A Test of Leadership: Charting the Future of U.S. Higher Education*. Washington, DC: United States Department of Education (USDOE).
- van Vught, F. A. and Westerheijden, D. F. (1993). *Quality Management and Quality Assurance in European Higher Education: Methods and Mechanisms*. Luxembourg: Office for Official Publications of the Commission of the European Communities.
- Warren Piper, D. J. (2004). *Are Professors Professional? The Organisation of University Examinations*. London: Jessica Kingsley.

Williams, G. (2005). *Subject Benchmarking in the United Kingdom. Research Program for Academic Quality*. Public Policy for Academic Quality Research Program. Chapel Hill, NC: University of North Carolina.

World Bank (2002). *Constructing Knowledge Societies: New Challenges for Tertiary Education*. Washington, DC: World Bank.

Young, M. F. D. (2003). National qualifications frameworks as a global phenomenon: A comparative perspective. *Journal of Education and Work* 16(3), 223–237.

Further Reading

- Brennan, J., de Vries, P., and Williams, R. (eds.) (1997). *Standards and Quality in Higher Education*. London: Jessica Kingsley.
- Brown, R. (2004). *Quality Assurance in Higher Education: The UK Experience since 1992*. London: Routledge Falmer.
- Burke, J. C. Associates (2002). *Funding Public Colleges and Universities for Performance: Popularity, Problems, and Prospects*. Albany, NY: Rockefeller Institute of Government, State University of New York.
- Dill, D. D. (2000). Designing academic audit: Lessons learned in Europe and Asia. *Quality in Higher Education* 6(3), 187–207.
- Massy, W. F. (2003). *Honoring the Trust: Quality and Cost Containment in Higher Education*. Bolton, MA: Anker Publishing.
- Santiago, P., Tremblay, K., Basri, E., and Arnal, E. (eds.) (2008). *Tertiary Education for the Knowledge Society*, vol 1: Governance, Funding, Quality. Paris: OECD.
- Scheele, J. P., Maassen, P. A. M., and Westerheijden, D. F. (eds.) (1998). *To be Continued: Follow-Up of Quality Assurance in Higher Education*. Maarssen: Elsevier/de Tijdstoom.
- Schwarz, S. and Westerheijden, D. (eds.) (2004). *Accreditation and Evaluation in the European Higher Education Area*. Dordrecht: Kluwer.
- Westerheijden, D. F., Brennan, J., and Maassen, P. A. M. (eds.) (1994). *Changing Contexts of Quality Assessment: Recent Trends in west European Higher Education*. Utrecht: Lemma.

Relevant Websites

- <http://www.qualityresearchinternational.com> – Analytic Quality Glossary.
- <http://www.che.de> – Center for Higher Education Development (CHE).
- <http://www.bcu.ac.uk> – Centre for Research into Quality, Birmingham City University.
- <http://www.chea.org> – Council for Higher Education Accreditation (CHEA), US.
- <http://www.enqa.eu> – European Association for Quality Assurance for Higher Education (ENQA).
- <http://www.graduatecareers.com.au> – Graduate Careers Australia (GCA) – Australian Graduate Surveys.
- <http://www.inqaahe.org> – International Network for Quality Assurance Agencies in Higher Education (INQAAHE).
- <http://www.jointquality.nl> – Joint Quality Initiative.
- <http://nsse.iub.edu> – National Survey of Student Engagement (NSSE).
- <http://www.unc.edu/ppaq> – Public Policy for Academic Quality Research Program (PPAQ).
- <http://www.qaa.ac.uk> – Quality Assurance Agency for Higher Education (QAA), UK.
- <http://www.teac.org> – Teacher Education Accreditation Council (TEAC).
- <http://portal.unesco.org> – UNESCO: Higher Education/Quality Assurance and Recognition.
- <http://www.worldbank.org> – World Bank – Tertiary Education.

The Role of Accreditation of Higher Education Institutions

J S Eaton, Council for Higher Education Accreditation, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

Defining and Describing Accreditation

Accreditation refers to examining and affirming the quality of a higher education institution. While there is significant variation in the use of the term among countries, many would agree with the description of accreditation offered by the United Nations Educational, Scientific and Cultural Organization (UNESCO): “(Accreditation is)... the outcome of a process by which a governmental, parastatal or private body (accreditation agency) evaluates the quality of a higher education institution as a whole, or a specific higher education programme/course, in order to formally recognize it as having met certain predetermined criteria or standards and award a quality label” (Sanyal and Tres, 2007: 6).

Whether called audit, evaluation, quality assurance, or accreditation, five characteristics are common to the quality-affirming efforts across many countries. These are the development and application of accreditation standards, self-reviews undertaken by colleges and universities, peer reviews of institutions, judgments by an accreditation body using the standards that have been developed, and awards of accredited status.

Accreditation standards are established either by an accrediting organization itself or in cooperation with the institutions and programs that are accredited. The standards focus on the resources, processes, and results associated with institutional efforts with regard to teaching and learning, governance and finance, among other areas. Standards are periodically reviewed and revised.

Self review by an institution is undertaken in the early stages of an accreditation review. It is usually based on the accreditation standards or, in some instances, more generally on individual institutional mission and goals. The self-review and accompanying documentation is sent to the accrediting organization for its consideration.

Peer review refers to an examination of the self-review by colleagues in higher education or a specific profession. This takes place through scrutiny of the self-review documents, usually accompanied by a visit to the institution. Review teams are composed of higher education colleagues and, increasingly, also include individuals from, for example, the business sector, students, or the international community. The peer-review process culminates in a report to the accrediting organization.

The accrediting organization, through its decision-making body, uses the self-review and the peer report to make a judgment about whether and to what extent an

institution or program meets the accreditation standards. These decision-making bodies are usually composed of academic administrators and faculty, sometimes accompanied by public members.

With regard to awards of accreditation, a decision-making body may act to grant accreditation, deny accreditation, or ask the institution to undertake some additional work to achieve accreditation. The length of accreditation periods varies from a few years to ongoing accredited status, provided that the institution sustains an effective system of internal quality review.

Accreditation status may be granted for an indefinite period or require periodic renewal.

Most accreditation bodies are owned and operated by governments, although they exhibit degrees of freedom from the government in a number of ways. They may, for example, be only partially funded by government and receive funds from other sources such as universities. Although established by government, these bodies may be governed, at least in part, by the university community. The United States is a conspicuous exception, where all accreditation organizations are private (nongovernmental) organizations.

The Values That Drive Accreditation

Accreditation is based on a core group of values about how colleges and universities are to operate. While this varies in some measure from country to country, commitment to these certain values is frequently expressed. They buttress the work of accreditation that, in turn, reinforces these values in an institutional setting.

A core value related to accreditation is the importance of institutional mission. An institution's statement of mission describes the range of its activity, for example, degree level, selectivity, curricula, governance, and intended student population. Most systems of accreditation use mission as the point of departure for judging the quality of institutional performance. In the accreditation review, the institution is typically held accountable for carrying out its mission, but not for activities that fall outside this range.

A second value is institutional autonomy, the importance of responsible freedom of action at the institutional level as essential to assuring quality. This value is grounded in the view that the faculty and academic administrators of a college or university are best positioned to make academic decisions such as determining curricula and setting

learning standards. While a number of countries are still sustaining systems in which government decides curricula, admission criteria, and faculty hiring, there is a growing international dialog about the need and desirability of institutional autonomy.

A third value is academic freedom or the importance of sustaining independent intellectual inquiry in the teaching and research of faculty. The tradition of faculty as maintaining and defining this academic space as they carry out their responsibilities has been essential to knowledge development throughout the world. As with institutional autonomy, the commitment to academic freedom is often carried out alongside considerable government authority for academic decisions affecting colleges and universities.

The Roles of Accreditation

Accreditation carries out several roles that are vital to students and society. Among the most important are accreditation's work to (1) assure threshold quality and quality improvement in colleges and universities, (2) serve as a reliable authority on quality for students, government, and the public, and (3) assist with student mobility through easing both recognition of degrees and transfer of credit.

For colleges and universities, accreditation is a useful external review that enables an institution to affirm its quality and create paths to improving quality. If successful in achieving accreditation, an institution has a quality stamp, a significant indicator to students and society that it has undergone external review which confirms that it meets threshold-quality expectations.

Accreditation serves as a reliable authority for both the public and private sectors that fund higher education institutions (e.g., government, private foundations, and alumni donors). Accreditation plays the role of confidence-builder. Without the accredited status, it is quite difficult, if not impossible, for an institution to either obtain sustained funding for its operation or maintain the confidence of these funders in the work of an institution.

Accreditation assists with student mobility and recognition of qualifications. Accreditation is one assurance of the value of a degree as well as a factor in the recognition of degrees across institutions and countries. Accreditation is also a significant factor in transfer of credit or students attempting to move credits earned at one institution to another, usually in the course of earning a degree.

There are other roles that accreditation plays. For example, quality assurance is sometimes seen as a means to assure equity in higher education, providing for high quality along a range of institutions, whatever the income level of students who attend. The accreditation process is sometimes considered from a social-justice perspective, with concerns that the process not be used to reinforce

undesirable social stratification. Higher education is often viewed as essential to reinforcing social responsibility, ethical awareness, and citizenship and these characteristics are included in concepts of quality that accreditors review.

Change in Higher Education and Implications for Change in Accreditation

The significant changes that higher education is experiencing in many countries are having a considerable impact on accreditation. Three changes are especially dominant: the expansion of access, internationalization of higher education, and heightened accountability to the public.

Expansion of Access

Both developed and developing countries are experiencing significant growth in higher education attendance. Many countries have constructed mass higher education systems, enrolling anywhere from 25 to 40% of the college-going age cohort (Altbach, 2007). The Organization for Economic and Cooperative Development (OECD) countries saw a doubling of the proportion of adults with higher education qualifications between 1975 and 2000 (*The Economist*, 2005). An estimated 133 million students were enrolled in higher education in 2004 (Sanyal and Tres, 2007), depending on whether part-time as well as full-time students are included and how individual countries define higher education. This is up from an estimated 90 million in 2000 (Garrett, 2007).

The unprecedented expansion of access is driven by a range of factors. The diminution of manufacturing employment and the rise of knowledge and service employment dominate more and more national economies, intensifying the need for higher education. Students need at least some higher education in order to obtain jobs and assure upward mobility in the workforce. In many countries, the level of education is closely tied to social success and perceptions of personal accomplishment. Beyond education for work, it is vital that countries also sustain the important traditions of knowledge development and intellectual inquiry.

International competition and cooperation are putting pressure on more and more countries to assure their populations measure up to workers in other countries. Finally, the increase and expansion of trade agreements as well as technology that enables businesses to quickly develop a worldwide workforce has also put pressure on countries to improve access to higher education.

Access through distance learning

Distance learning has been around for many years, mainly in the form of correspondence courses. The nature of distance learning varies among countries for reasons

related to geography, culture, history, and current capacity for technology. In all instances, however, students are separated, either by time, location, or physical proximity, from the institutions and faculty offering instruction. Twenty-first-century distance learning is an electronically driven, synchronous, or asynchronous and web-based enterprise. It has emerged as a significant enabler of access. The Internet has had a most profound impact on the provision of distance learning, allowing providers to invent themselves at will.

An estimated 7 million students are enrolled in distance-learning offerings (Garrett, 2007). Some countries, such as Kenya and Jordan, import distance learning; others such as Malaysia have significant in-country capacity. Yet others, such as the United Kingdom, Australia, and the United States, are major exporters of distance learning (Middlehurst and Woodfield, 2005). The growth of reliance of distance learning has been particularly significant in countries such as Turkey, China, Japan, India, and Israel, where some of the largest distance-learning providers are located. As distance learning is readily accessible even without the development of campuses or a faculty, it has become part and parcel of mass higher education, instantly international and capable of rapid changes (Altbach, 2007).

Access through for-profit providers of higher education

For-profit enrolments are a subsection of the complex privatization of higher education that is going on worldwide. The emerging for-profit activity is large scale, career oriented, degree granting, aimed at a mass market, and competitive with traditional higher education. At this time, there are few countries in which for-profit, degree-granting higher education exceeds 5% of total enrolments (Garrett, 2007a). In the United States, a traditionally hospitable for-profit higher education environment, 2 million of the 16–17 million students in colleges and universities attend for-profit institutions (JBL Associates and Wilson, 2007). An estimated 5 million students are enrolled in for-profit operations worldwide (Garrett, 2007).

For-profit providers are expanding rapidly in countries such as China, India, and regions such as Latin America. Particularly, the US-based publicly traded for-profit operations such as the Apollo Group, DeVry, Laureate, and Kaplan are actively engaged in international expansion. The for-profit focus is primarily on training, credentials, and employment. International for-profit higher education relies on both local and electronic delivery, cooperative initiatives with local higher education providers in other countries, and special initiatives for short periods of time such as summer programs (Levy, 2006).

Degree mills, dubious or bogus providers of higher education, are the dark side of for-profit higher education. The rogue operations, offering degrees and other credentials for considerable sums of money but requiring little if

any educational experience, are causing considerable concern for many countries, especially those that import higher education. A provider is likely to be a degree mill if, in addition to the characteristics mentioned above, it employs few if any faculty, has a name that is similar to a well-known traditional institution, sells credentials online, and provides no evidence of accredited status.

While there is considerable discussion about the problems caused by degree mills, few countries to date claim to have adequate laws and effective practices to fully protect students, employers, and the public, although considerable efforts are underway to address this. In the United States, for example, federal legislation has been introduced that would define degree mills and make the use of fraudulent credentials a crime. A number of states, as well, are introducing or strengthening their laws with regard to degree mills. Central to the effectiveness of these efforts is the capacity to distinguish degree mills from other, quite legitimate for-profit providers.

Impact on accreditation

The expansion of access to higher education in country after country has resulted in increased pressure to establish, expand, or restructure accrediting bodies. Accrediting organizations are being asked to assure the quality of a significantly massified and more diverse higher education enterprise. Since 1991, the number of countries with fully operational accreditation systems has grown from 11 to 60 (Lewis, 2005). As traditional institutions develop new programs, students and the public want the same kind of assurances about the quality of these new offerings as with more mature programs. As branch campuses are developed to serve more and more students, the public wants to be similarly assured. The internationalization of higher education as a means to expand access has been part of this pressure as well.

A number of countries have passed new legislation or revised current law to address the quality of their expanding and diversifying higher education systems. For example, Japan amended its School Education Act in 2002 and initiated a new accreditation scheme in 2004. All public and private universities, junior colleges, and colleges of technology have to be accredited by an organization authorized by the national government (Yonezawa, 2005). In the same year, China's Ministry of Education merged accreditation, regular institutional assessment, and selective assessment into a single policy package and in 2004, a quasi-governmental quality-assurance agency was established (Sanyal and Tres, 2007).

Portugal's new degree law, launched in 2006, provides a legal framework for a national accreditation system (European Network for Quality Assurance in Higher Education, 2006). In South Africa, the higher education quality committee was established in 2001 to carry out three mandated roles of accreditation of programs, audit of institutions,

and quality promotion and capacity development. This has resulted in a strengthened focus on quality and the application of a qualifications framework across a coordinated system (Council on Higher Education, 2004).

With regard to distance learning, more and more accrediting bodies are developing capacity to review these operations. They have put in place new standards or applied current standards to address synchronous and asynchronous learning environments, free-standing distance-learning providers, or blended instruction that is a combination of distance and site-based delivery. India, for example, has established a distance-education council specifically for its open universities and distance-education system. The quality assurance agency (QAA) for higher education in the United Kingdom has a statement of good practice that is applied to distance learning as well as site-based institutions (Sanyal and Tres, 2007). In the United States, review of distance learning has been incorporated in the ongoing accreditation reviews of colleges and universities and one US accreditor, the distance education and training council, reviews only distance-learning providers.

For-profit providers of higher education are, in some countries, subject to the same scrutiny as nonprofit operations. In other countries, there is little capacity or infrastructure to acknowledge and thus address these providers, the result being an essentially unchecked and blossoming enterprise, operating alongside government-regulated higher education. Legitimate apprehension and concern about degree mills are driving out careful attention to legitimate for-profit providers. To the extent that for-profit operations are accredited, the apprehension and concern are diminished. In the United States, for example, more than 2500 for-profits have earned accreditation from a reliable accrediting organization.

Accreditation is often a first-line defense against for-profit degree mills in a number of ways. Institutions that have achieved accredited status from reputable accreditors constitute a *de facto* positive list of responsible institutions on which students and the public can rely. A number of degree mills seek accreditation and fail to achieve this status, thus remaining outside the sphere of what are generally considered legitimate higher education institutions. Accrediting organizations, working with organizations such as UNESCO or OECD, are playing a significant role in creating greater international awareness and enhancing capacity to identify degree mills. Going forward, accrediting organizations, working with institutions, government agencies, and business, can develop effective practices and other means to enhance public awareness and understanding of how to identify and avoid degree mills.

Internationalization of Higher Education

According to one UNESCO estimate, 3.84 million students were studying abroad in 2002 (Sanyal and Tres, 2007).

In the 1990s, the United States has seen growth in international students from Korea, China, and Japan, (Altbach, 2006). As of 2005–06, 564 766 international students were enrolled in US colleges and universities. (Institute of International Education, 2007). More recently, there has been considerable growth in the number of foreign students in England, Germany, and France as well (*The Economist*, 2005).

In response to this enrolment growth, the international activity of colleges and universities is expanding rapidly and has taken several forms. More institutions are entering into partnerships with colleges and universities outside a home country. More institutions are moving unilaterally to operate internationally, either through offering programs in another country or establishing new campuses or even freestanding institutions. Faculty exchange programs, a staple of international activity, have grown significantly. In some instances, developing countries are inviting foreign institutions to come to them as an alternative to sending students abroad. This is the case in Singapore, Dubai, and Qatar (*The Economist*, 2005). Distance-learning operations, as well, are increasing the extent of international activity.

Impact on accreditation

All of this activity has resulted in pressure on accreditation to operate internationally as well as nationally. In country after country, accrediting bodies are developing capacity to review programs established for student exchange, partnerships among institutions that cross national boundaries, and institutions establishing operations in other countries. The QAA of the United Kingdom, for example, reviews all overseas activity of UK institutions (Middlehurst and Woodfield, 2005). In Australia, all offerings of providers are examined, wherever they are carried out, by the Australian universities quality agency (Sanyal and Tres, 2007).

US private, nongovernmental accrediting organizations are now exporting accreditation. As US institutions expand internationally, accreditors are required to expand their work to other countries. Some US accreditors actively seek significant additional international engagement through accrediting non-US institutions and programs. As of 2005, 43 of the 81 recognized accrediting organizations were active in 96 countries, accrediting 710 operations (Council for Higher Education Accreditation, 2006). While growing slowly, the international activity of US accreditors raises a number of issues. As the only country that is, at present, free to export accreditation, what are the appropriate means by which to do this? Can the fundamentals of US accreditation, closely intertwined with US higher education that is different in a number of ways from higher education in many other countries, avoid a kind of academic colonialism – shaping academic policies in other countries (Altbach, 2006)?

Accreditation is taking on a regional character in the service of an internationalizing of the higher education enterprise. The European Bologna Process, a comprehensive initiative involving 46 countries, provides one example of attention to assuring quality through a regional regulatory framework. European standards and guidelines for quality assurance were adopted in 2005. A register of European higher education quality-assurance agencies has been developed and is in the implementation stage (European University Association, 2007). The purpose of the register is to provide for peer review of quality-assurance bodies.

Other regional efforts include the Ibero-American Network for the accreditation of quality in higher education established in 2003. The network is made up of quality-assurance agencies, government organizations, and associations of higher education institutions in Latin America and Spain. Its efforts are focused on development of national quality-assurance mechanisms, including assistance with development of technical staff for quality-assurance agencies. It also promotes regional efforts (Lemaitre, 2006). The Arab states have been promoting quality assurance at regional and national level, with regional efforts underway since 2003 to establish the Arab states quality-assurance network and the Arab society for quality assurance in education. At a national level, this includes mechanisms for quality assurance and accreditation (Sanyal and Tres, 2007).

Multinational organizations have been active in this area as well. UNESCO initiated a global forum on international quality assurance, accreditation, and recognition of qualifications in 2002, establishing a significant arena in which governments, institutions, and accreditation and quality-assurance bodies can address quality-related issues. The forum's work includes a major capacity-building effort, especially for developing countries as well as efforts to establish an information portal to provide a central international source for information about reliable higher education institutions and accreditation and quality-assurance providers in each country. UNESCO and OECD published, in 2005, *Guidelines for Quality Provision in Cross-Border Higher Education* to assure quality across national boundaries as international activity in higher education continues to expand.

Heightened Accountability to the Public

Across the globe, governments are concerned to make higher education institutions more accountable to the public. Two expectations define their efforts: increased calls for evidence of student achievement and demands for greater transparency. The similarity of expectations across countries is striking and related to the internationalization of higher education as well as expanded access.

In many countries, evidence of student achievement is being addressed through the development of student

learning-outcome indicators and identification of expected student competencies. Australia, various provinces in Canada, England, South Africa as well as Europe, and a number of other nations and regions, for example, have developed qualifications frameworks, seeking to link degree levels with student achievement in specific skill areas. The European effort, the Dublin descriptors, is multinational and identifies capacities that are to be associated with degrees at the bachelors, masters, and doctoral levels (Lewis, 2005). In Hong Kong, incorporating additional attention to student achievement has been central to policy discussions related to the ongoing work of its publicly funded universities and a comprehensive qualifications framework has been developed as well.

Accountability has dominated federal policy discussion in the United States during the past 2 years. There is little current interest in qualifications frameworks, but there is increasing emphasis from both state and federal governments to develop a common set of institutional performance indicators or expectations of student achievement. Institutions themselves are pressed to provide more information about performance and student achievement on websites and in official print documents. This has been accompanied by a number of federal and state associations of higher education institutions that have recently taken significant steps to expand their attention to both performance indicators and student achievement.

Transparency, demands for readily accessible and easily understandable information to the public about quality, is routinely part of accountability conversations. Institutions are called upon to provide information, for example, about graduation rates, job placement, transfer, and entry to graduate school. There is much discussion about rankings and mechanisms that allow for comparison among institutions as additional means of addressing the public interest in greater accountability. Less of this work has taken place through accreditation bodies, but has been pursued by governments, universities, or the commercial sector.

There is considerable debate about whether or not rankings provide useful information about quality; nonetheless, these lists attract a good deal of attention. These include the national rankings of *US News and World Report* in the United States and the *Times Higher Education Supplement* in the United Kingdom and the international academic rankings of world universities at Shanghai Jiao Tong University. To address some of the issues and concerns surrounding rankings, an international group of professionals issued, in 2006, the Berlin principles that provide a framework to approach and improve ranking systems (Institute for Higher Education Policy, 2007).

Impact on accreditation

A number of accrediting organizations in various countries have standards that require institutions to

develop expectations of student-learning outcomes and provide evidence that these expectations are realized. In the United Kingdom, the QAA has subject benchmarks for more than 40 program areas (Lewis, 2005). In the United States, expectations to address student outcomes have been embedded in the standards of all recognized institutional accreditors since 1992.

Accrediting organizations in many countries are expected to enforce qualifications frameworks. In addition, accreditors are pressurized to provide comprehensive information about their reviews of institutions. As of 2005, 46% of accrediting organizations made their reports public, with 52% keeping the reports private (Lewis, 2005).

There are also international calls for accrediting bodies to address both outcomes and transparency. In addition to the UNESCO/OECD *Guidelines*, two statements on sharing higher education quality across borders have been developed by the American Council on Education, the Association of Universities and Colleges of Canada, the Council for Higher Education Accreditation, and the International Association of Universities, urging institutions and accreditors to address both of these areas (Council for Higher Education Accreditation *et al.*, 2005, 2006). The *Guidelines of Good Practice* of the international network of quality assurance agencies in higher education includes attention to transparency as well (2006).

Summary

Accreditation of higher education is a quality-affirming process that, in many countries, includes the creation of accreditation standards, institutional self-review, institutional peer review, judgment about achieving accreditation, and award of accredited status. Just as the process of accreditation is similar across countries, both the values and roles of this enterprise are similar as well. Accreditation builds on the key values of institutional mission, institutional autonomy, and academic freedom. It plays major roles with regard to assuring threshold quality and quality improvement, serving as a reliable authority on quality to students and society and assisting with student mobility through recognition of qualifications and transfer of credit.

Major changes in higher education in many countries are also driving changes in accreditation. Higher education is experiencing a significant expansion of access with the internationalization of colleges and universities and demands for heightened accountability to the public. To accommodate these changes, accrediting bodies are expanding in numbers as well as enhancing and diversifying their capacity in order that these emerging key features of higher education are effectively examined for quality and quality improvement.

Bibliography

- Altbach, P. G. (2006). *International Higher Education: Reflections on Policy and Practice*. Chestnut Hill, MA: Center for International Higher Education, Boston College.
- Altbach, P. G. (2007). *Tradition and Transition: The International Imperative in Higher Education*. Chestnut Hill, MA: Center for International Higher Education, Boston College.
- Council on Higher Education (2004). *South African Higher Education in the First Decade of Democracy*. Pretoria, South Africa.
- Council for Higher Education Accreditation (2006). *The Condition of Accreditation: U.S. Accreditation in 2005*. Washington, DC. http://www.chea.org/pdf/condition_of_accr.pdf (accessed June 2009).
- Council for Higher Education Accreditation, The American Council on Education, The Association of Universities and Colleges of Canada and The International Association of Universities (2005). *Sharing Quality Higher Education across Borders: A Statement on Behalf of Higher Education Institutions Worldwide*. <http://www.chea.org/international/index.asp#statement> (accessed June 2009).
- Council for Higher Education Accreditation, The American Council on Education, The Association of Universities and Colleges of Canada and The International Association of Universities (2006). *Sharing Quality Higher Education across Borders: A Checklist for Good Practice*. <http://www.chea.org/international/index.asp#statement> (accessed June 2009).
- European Network for Quality Assurance in Higher Education (2006). *Quality Assurance of Higher Education in Portugal*. Helsinki, Finland. <http://www.enqa.eu/pubs.lasso> (accessed June 2009).
- European University Association (2007). *EUA Newsletter Number 9, 25 May 2007*. http://www.eua.be/File/Admin/USER_upload/files/newsletter/COM_NL_9-2007_1.pdf (accessed June 2009).
- Garrett, R. (2007). *Personal Correspondence*. August 1, 2007.
- Institute for Higher Education Policy (2007). *College and University Ranking Systems: Global Perspectives and American Challenges*. Washington, DC. <http://www.ihep.org> (accessed June 2009).
- Institute of International Education (2007). *Search results*. <http://www.iie.org/template.cfm?Section=Search&Template=/CustomSource/Search> (accessed June 2009).
- International Network for Quality Assurance in Higher Education (2006). *INQAAHE Guidelines of Good Practice*. <http://www.inqaahe.org> (accessed June 2009).
- JBL Associates and Wilson H. (2007). *Economic Impact of America's Career Colleges*. Washington, DC: The Imagine America Foundation.
- Lemaitre, M. J. (2006). *Developing Trust: Towards Mutual Recognition Agreements – MERCOSUR-RIACES*. The World Bank Learning Seminar: Quality Assurance in Tertiary Education, Sevre France, June 18–20, 2006. http://siteresources.worldbank.org/EDUCATION/Resources/QA_Seminar_Agenda.pdf (accessed June 2009).
- Levy, H. O. (2006). The commercialization of higher education: A for-profit perspective. *Presentation to the Council for Higher Education Accreditation International Commission*. January 26, 2006, San Francisco, CA. http://www.chea.org/international/commission2006/HLevy_ICO12606.pdf (accessed June 2009).
- Lewis, R. (2005). International trends in the quality assurance of higher education. *Presentation to the Korean Educational Development Institute*, Seoul, Korea, October 28, 2005.
- Middlehurst, R. and Woodfield, S. (2005). *International Quality Review and Distance Learning: Lessons from Five Countries*. Washington, DC: Council for Higher Education Accreditation.
- Sanyal, B. C. and Tres, J. (eds.) (2007). *Higher Education in the World 2007: Accreditation for Quality Assurance: What Is at Stake?* Global University Networks for Innovation Series on the Social Commitment of Universities, 2nd edn. New York: Palgrave Macmillan.
- The Economist* (2005). Survey: Higher education – the brains business. September 10, Volume 376, Number 8443.
- Yonezawa, A. (2005). *The Reintroduction of Accreditation in Japan: A Government Initiative*. International Higher Education, Number 40. http://www.bc.edu/bc_org/avp/soe/cihe/newsletter/Number40/p20_Yonezawa.htm (accessed June 2009).

Universities and Regional Development*

J Goddard, Newcastle University, Newcastle upon Tyne, UK

J Puukka, OECD Directorate for Education, Paris, France

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Across the globe, universities and local communities are discovering each other and more and more partnerships are being established based on a growing appreciation of shared interests. This article explores the drivers behind such engagement from both the university and local-community perspectives, the barriers to effective working, and how these barriers can be addressed at the institutional, local, and national levels.

The article draws heavily on an Organization for Economic Cooperation and Development (OECD) review on supporting the contribution of higher education institutions to regional needs and which involved case studies of 14 regions across 12 countries. Following the OECD approach, this article links two hitherto separate areas of enquiry, namely that concerned with the economic, social, and cultural development of localities in the round (cities and regions) and the university as a key institution in civil society. It therefore considers how groupings of local interests in business and the wider community can reach in to universities in their area and how universities can reach out and engage with these groupings.

Inevitably, a short high-level article cannot pay adequate attention to important differences between countries, universities, and localities. The nature of engagement between universities and their local communities is highly contingent on institutional and local-development trajectories and national policies toward higher education and the governance of cities and regions. The reader is therefore referred to the OECD case study reports which reveal universities and regions at various stages on this journey of discovery (or re-discovery) and how in the process they are seeking to overcome all manner of barriers to engagement.

The Regional-Development Drivers behind Community Engagement

Post World War II, regional policy in many OECD countries emphasized the need for intervention by the nation state to reduce disparities between regions. Public

intervention took the form of financial support for established industries and the attraction of mobile investment in order to absorb surplus labor. There were also measures to equalize living standards between regions, including standards of primary and secondary education.

Significantly, higher education did not enter into the panoply of regional-policy interventions. Many European universities which had developed to serve traditional industries during the later part of the nineteenth and first-half of the twentieth century were subsequently incorporated during the 1960s into national systems of higher education. In this process, their local ties were weakened. However, in the United States, individual states did support public universities in serving the needs of their territories, building on the land-grant tradition established in the nineteenth century. Indeed, state investment in higher education to tackle industrial decline in New England and to attract new federal investment in areas facing structural adjustment in agriculture in California laid the foundation for subsequent high-technology corridors such as Route 128 and Silicon Valley. In Canada and Australia, where a federal structure of government was established, higher education played a key role in the development of the cities which were the gateways to the individual states, for example, laying the foundations for the so-called sandstone universities in each of the state capitals of Australia. In Australia, regional problems were (and remain) essentially problems of underdeveloped city hinterlands and rural areas. Outside of the so-called developed world, the priority of nation building around national capitals contributed to rising regional disparities with national universities being one of the magnets for internal migration.

The European post-war consensus around the need for state intervention to reduce core/periphery regional disparities broke down during the 1970s. This was associated with the onset of structural adjustment problems and the rejection of the post-war Keynesian models of economic regulation. The emergence of so-called rust belts linked to traditional industries such as coal and steel, heavy engineering, and textiles which were now facing competition from newly industrialized countries and the related decline of mobile investment seeking lower-cost sites within industrialized countries, undermined the basis of redistributive regional policy.

In response to the crisis, the emphasis on territorial and industrial policy switched toward indigenous development

* Based on Goddard, J. and Puukka, J. (2008). The engagement of higher education institutions in regional development: An overview of the opportunities and challenges. *Journal of the Programme on Institutional Management in Higher Education* 20(2). OECD (2008) *Special Issue: Higher Education and Regional Development*. With permission from OECD.

focused on small and medium-sized enterprises (SMEs) with a particular emphasis on the role of innovation in raising their competitiveness.

This shift of emphasis opened the way for links into the research base in local universities. It also coincided in the US with the passing of the Bayh-Dole Act in 1980 which empowered universities to commercialize their own intellectual property. During the 1980s, a growing body of academic literature underpinned the case for local or bottom-up public intervention in the supply side of the local environment supporting (or inhibiting) business innovation. Studies of the so-called third Italy indicated that networks of traded and untraded interdependencies between SMEs could provide a fertile environment for innovation in traditional industries outside established urban agglomerations. Whereas in Italy these networks did not involve universities, the experience of Silicon Valley in California and Route 128 in New England assumed totemic significance in relation to the possibility of creating new industrial districts or regenerating older districts through strong links with research-intensive universities.

Moving into the 1990s, the range of supply-side factors that regional policymakers deemed to be influencing economic performance widened. Most significantly, education and skills and the tacit knowledge gained through work-based learning became embodied in the concept of the learning region. This had resonances with the growing appreciation that innovation is not necessarily a linear process and could involve close interaction between producers and users, interactions which were best conducted face to face. Moreover, the role of students and graduates in what might be termed knowledge transfer on legs and establishing the social relations between researchers and the business in which they work became increasingly apparent.

During the 1990s, these perspectives began to be formally adopted in public policies to foster the development of industrial clusters rooted in particular places. The concept of the industrial cluster recognized that innovation is seldom isolated but systemic with the industrial cluster acting as an innovation system. Clusters, in this instance, could encompass strategic alliances of universities, research institutes, knowledge-intensive business services, bridging institutions and customers. Cluster success required and encouraged flows of talented individuals, including students and graduates, and the creation of vibrant and exciting places.

Within the cluster, universities could assume an entrepreneurial role while firms developed an academic dimension. Emphasis was placed on a spiral model of interaction where a number of channels feed into the innovation process including research links (the creation of new knowledge), information transfer (selling existing knowledge), and people-based transfer (students and staff) as well as spin-offs. In this model, specialized centers and cluster discourse could provide a focus for both

higher education institutions (HEIs) and the business community. It could involve embedding engagement in the core business processes of both HEIs and industry.

Throughout the OECD, there is now a convergence of innovation and territorial-development policy. This is placing new demands on universities as innovation policy becomes more comprehensive. There is increased emphasis on education and training, employability, the quality and skills of the workforce, and lifelong learning. People and human resources are being brought into focus. There is recognition that initiatives to foster innovation and competitiveness need to take account of challenges of urban and regional variations in unemployment, poverty, and exclusion in a multicultural society. There are also aspirations to establish and foster creative and enterprising places where people and companies want to locate. Thus, many towns and cities have been inspired by reflections on the new creative class and the global competition for talent which has led to increasing investment on place marketing and the branding of cities as nice places to live.

In summary, regional policy which was redefined and narrowed down to technological innovation policy is now in the process of being ever broadened as other fields of policy are given an innovation signature and more agents and levels of government (city, regional, national, and international) are drawn into the process of building innovative capabilities. From a rather narrow focus on high technology and manufacturing industry and the private sector, attention has been widened to include social and organizational innovations and business, consumer, and public services.

This broadening of regional policy has wide-ranging implications for the expectations placed on universities by cities and regions. They are now expected to participate in public and private partnerships and contribute to balanced region building. Whereas previously attention was focused on universities as a source of high-technology innovations and new knowledge-based industries, these are now beginning to be regarded in a broader perspective, encompassing the whole social fabric of which universities are part. For example, the new emphasis on social innovation, tourism, the creative industries, and welfare widens the academic domains relevant to community engagement from science and technology and medical faculties to the arts, humanities, and social sciences.

In summary, and for those agencies charged with city and regional development, universities are: major businesses generating tax and other revenues; global gateways in terms of marketing and attracting inward investment in the private sector; generators of new businesses and sources of advice to existing businesses; enhancers of local human capital through graduate retention and professional updating of the existing workforce; and providers of content and audience for local cultural programs.

Universities, particularly in highly centralized states, can also be key local agents able to bring together within the territory different national interests in science and technology, industrial performance, education and skills, health, social inclusion, and culture.

The Higher Education Drivers behind Engagement

The longevity of universities as key institutions in the evolution of civil society is linked to their adaptability to changing circumstances, while maintaining key elements of continuity (such as the global connections which characterized the medieval foundations). The emergence of the Humboldtian University in nineteenth-century Prussia was linked to the professionalization of science, the requirements for specialized infrastructure to support it and to underpin, at a distance, the development of the state.

The principle of at a distance is important because in many respects, the research university that evolved in continental Europe during the nineteenth century can be described as a denial of place. This is because the ideal of scientific enquiry embodied in the modern university is to strive for universalism. Scientific claims to truth were deemed to be irrespective of time and place and the university had to have a mission that transcended its actual location. Indeed, the notion of the university as a detached site for critical enquiry, exchange of ideas, and advancement of knowledge for its own sake has been of vital importance to the creditability and legitimacy of the institution.

The nationalization of science and education during the twentieth century further enhanced the detachment of universities from places. Due to their importance to nation building, universities were no longer expected to rely on the patronage of churches, town councils, and local elites. They now received their core funding from national governments and in return trained the cadres for the civil service and national corporations and the professions such as law, medicine, engineering, and architecture. They were to contribute to new national identities and the cultural spirit which underpinned the nation-building process. All of this was based on a compact whereby the university rendered services to the state in return for a degree of institutional autonomy in terms of internal governance.

Part of the American higher education system, however, developed in a different direction. Land-grant universities, which in the first instance promoted agricultural development, were regionally embedded people's universities based upon widening access to education and service to the community.

The second part of the twentieth century witnessed a massive expansion of public investment both in research

and development and higher education. This had a profound impact on the universities that had emerged in the previous century and their engagement with regions. The expansion of higher education typically took place outside the established universities which were regarded as too inflexible to meet the demands for new skills emerging in the workplace and from communities where they were not present. Thus, we now speak of HEIs not just universities. The higher education map of most countries has been colored in incrementally with a diverse set of institutions. Many of the new institutions built on previous foundations, typically with limited tradition of research (such as teaching and nurse-education colleges), and many of them have been given a specifically regional mission.

In some countries, this geographical dispersal of higher education has formed part of a conscious policy seeking to preserve the spatial distribution of the population and to achieve balanced regional development by addressing regional disparities. It has included also the objective of improving regional access to higher education. This has translated into policies to establish new HEIs in various regions, for example, in Norway, Sweden, Finland, Japan, and Mexico. However, in many countries dispersal of higher education has followed a simple logic of higher education expansion modified by political lobbying. This is not just a top-down phenomenon. Towns and cities have lobbied for their own university.

The consequence is that many OECD countries have a highly diversified system of higher education with complex mixes of universities, polytechnics, regional colleges, and vocational-training institutions. The regional role has sometimes served to differentiate among the various types of institutions. In Finland and Portugal, for example, universities are considered to have a stronger national and international role while polytechnics are assumed to focus on their regional role.

In many countries, the distribution of universities is not necessarily structured to meet the challenge of balanced regional development in a highly competitive global economy. So, while disadvantaged regions may possess locally orientated HEIs, these are often more geared toward upgrading the existing industry and less equipped to build a new knowledge-based economy.

The expansion of public investment in research in science and technology inside and outside of universities has likewise had an impact on the issue of regional engagement. This expansion has largely been driven by ministries of science and technology and in many cases has taken place in public research laboratories outside higher education, characteristically in the hinterlands of capital cities. At the same time, universities were able to compete for research funding from research councils operating at arm's length from government. In these councils, the academic community has had a major influence

via peer review in a manner that has preserved the autonomy of their institutions and their distance from the state. This peer-review process has often reinforced the position of the longest established institutions, typically in capital cities, thereby reinforcing regional disparities.

During the 1990s, this model for the organization of public research began to break down as governments began to demand a more immediate economic return for investment in the science base. A key challenge has been to remove barriers and bottlenecks between scientific research and industrial innovation. The institutional division of labor which implied that research was carried out in isolation from the context of application was perceived as a problem when science policy was morphing into innovation policy. In this process, HEIs as institutions, as well as the individual academics who work within them, have been expected to become more active players in the so-called triple helix of government, business, and HEI relations.

Industrial policy and science and technology policy have thus been converging toward a common innovation policy which in some countries explicitly or implicitly embodies a strong territorial dimension. Research-intensive universities have been surrounded by science parks and a host of special-purpose organizations established to support close cooperation with industry. In some instances, these have served to buffer the institution from external pressures and instead of facilitating links, these have operated as filters or merely served as display windows toward the universities' political environment. But increasingly, universities are expected to take the lead and to rearrange the structures so that the promotion of entrepreneurship and knowledge exchange form part of the academic heartland of research and teaching. HEIs are now expected to contribute to economic development by: creating new sectors through spinning out businesses backed by research; attracting and retaining global businesses in the region through the availability of quality research links and the supply of well-trained graduates; assisting with the diversification of established businesses in their production of new products and services; and upgrading existing mature industries through assistance with incremental product/service innovation and improvement in industrial/business processes.

This science-driven model nevertheless overlooks many features of regional development to which universities can directly and indirectly contribute. It neglects the contribution of broad-based teaching and learning to the enhancement of regional human capital. Service industries provide most regional jobs and the majority of graduates take up employment in financial, legal, and other professional services. Some regional services will be traded nationally and internationally and use the skills of graduates to develop new products, some of which will also be provided to regional high-technology-based

businesses. These businesses may in turn require non-scientific graduates, for example, with a business-school background to assist in activities such as marketing. Other important nonmanufacturing sectors recruiting graduates are the cultural industries and tourism which can serve to attract and retain creative people within the region, including those working in high-technology businesses and HEIs themselves. Moreover HEIs are creators of, and venues for, cultural and social activity.

HEIs also make a considerable contribution to public services, particularly health and education; these services play a role in economic development not least as regions with wide internal social disparities less likely to be attractive to leading-edge investors in the global knowledge economy. Finally, as environmental sustainability moves up the political agenda, it is becoming increasingly apparent that HEIs could have a key role to play through research, teaching in public education, and in building sustainable communities. All of these latter roles highlight the public service responsibility of HEIs as distinct from the more private focus of the science-driven model.

In summary and in terms of economic drivers, HEIs are seeking: local support for their global aspirations in research and student recruitment; increased student enrolments from the local population; additional income from services provided to local businesses through consultancy and professional training; and last but not least the indirect benefits of a local environment that can attract and retain creative academics and motivated students. At a higher level, regional engagement is an outward and visible sign of the third or public-service role of higher education and through which the institution can demonstrate its contribution to civil society. Through such endeavors, HEIs are able to provide concrete evidence of the value that higher education and research adds to public investment in it.

National Policy Barriers to Regional Engagement

In most OECD countries, higher education policy does not include an explicit regional dimension. Ministries of Education characteristically act as champions of the role of higher education and research in meeting national aspirations in terms of scientific excellence and advanced education of high quality for its own sake.

The seemingly more mundane task of applied research and development and meeting skill needs in the local labor market may be left to lower tiers in the education system, such as tertiary/community colleges. In some countries, the boundaries between the levels of higher education have become blurred. Examples include the designation in 1992 of polytechnics in the United Kingdom as universities, the designation of selected colleges in

the Netherlands as universities of professional education (now universities of applied sciences), and the creation of universities of applied science in Finland (formerly polytechnics).

Characteristically, the newer institutions do not have a well-established tradition in research or the infrastructure to support it and have to work hard with limited resources to build a national, let alone an international, profile traditionally associated with university status.

An important point to note in relation to regional engagement is that longer-established HEIs have developed and grown in locations that broadly follow the national-settlement hierarchy. These locations are quintessentially larger cities with the most prestigious institutions located in or around the capital city. In contrast, the newer institutions, often with a specific remit to serve particular territories, tend to be more geographically dispersed.

To what extent has the process of rolling out of higher education across national territories been part of conscious national policies to use higher education as an instrument in regional development? The answer depends on the definition of development and the extent to which this has been a task laid upon HEIs by their funders in central government. It is widely accepted that the challenge of raising competitiveness via research-led innovation is now at the heart of regional policy. However, it is clear that supporting excellent research in all regions has not been an objective of higher education policy. Even when engagement with business and the community has been recognized and laid upon HEIs as a duty as in all the Nordic countries, it has been very much a third task, not explicitly linked to the core functions of research and teaching. Nor, in most instances, is this task specifically funded or linked to regional development.

Turning to science and technology policy, there are growing pressures to ensure that public investment in this area has an economic impact. Consequently, there is an increasing convergence between research policy and other policies designed to support business innovation. But notwithstanding the growing recognition of the importance of organizational and social barriers to innovation, most top-down science and innovation policies continue to have a high-technology and manufacturing-industry focus and neglect the contribution of the arts, humanities, and social sciences to new ways of working and servicing the creative industries. Recent decades have therefore witnessed the establishment of centers of science-research excellence to support industrial innovation and which focus on newly emerging high-technology fields such as biotechnology, nanotechnology, and information and communication technology (ICT). But the policies underpinning such centers often fail to recognize that much innovation is neither science based nor radical, but incremental in nature and taking place in SMEs.

National innovation policy driven by ministries of science and technology also does not as a rule pay regard to the role of teaching and learning in knowledge transfer from the research base. Work-based learning schemes which usually involve regional links between employers and HEIs are designed to enhance graduate employability and not as specific tools to improve regional business competitiveness. A notable exception in this regard is the UK's knowledge transfer partnership scheme under which postgraduates undertake projects in companies which are usually local. However, this is not explicitly a regional scheme.

While most OECD countries have active national-labor-market policies led by the ministries of labor or their equivalent, the focus is chiefly on intermediate and lower-level skills and the unemployed, not those associated with higher education. At this level, it is assumed that the market (i.e., demands from students and employers) will work effectively without intervention. National employer-led associations for particular professions (e.g., lawyers, architects, and civil engineers) often play a key role in regulating supply and maintaining quality. Only in areas where the state remains a major provider of public services, most notably health, does the government undertake a planning role. While the market for intermediate and lower-level skills may be local and therefore require a strong spatial dimension, it is assumed that the market for high-level skills is national and international. Therefore, there is no case for intervention at the intermediate or regional level.

For these reasons, there appears to be little engagement by research-intensive universities in the development of human capital at the regional level, particularly as it relates to the skills required by knowledge-intensive businesses growing on the support of links with the research base. In contrast, newer and vocationally oriented institutions are usually committed to upgrading skills in the established industrial base.

The cultural domain is another area where the role of universities in contributing to city and regional development is not widely acknowledged in national policy. Universities are often owners or custodians of cultural assets displayed in their own museums and galleries. Their music, arts, and drama departments directly and indirectly contribute to the vibrancy of their cities through performance and related activities. In some countries, support for the arts and heritage does have a regional dimension and embrace higher education, but this is an exception rather than a general rule. Increasingly, universities are finding it difficult to support such activities out of their core teaching and research budgets and are seeking support from regional sources to maintain expensive facilities and activities. At the same time, the fast growth of the creative industries is shifting the focus to new enterprise formation by graduates of creative arts, design, and media.

Barriers in Regional Structures and Governance

Although many regions across the OECD are looking to universities to contribute to their economic, social, cultural, and environmental development, the capacity of city and regional public authorities to reach into higher education is often constrained by a wide range of factors. At the most general level, the public governance of territory operates within closed boundaries. Local and regional governments are responsible for administratively defined areas and these are usually linked to unambiguous political mandates. By contrast, research-intensive universities cannot have a mandatory geographical sphere of influence; indeed, such institutions operate at the local, regional, national, and international scales. Some lower-tier HEIs do have a specific regional mandate but it is increasingly less likely to be enforced by national, regional, and local governments with the institutions competing for students and contracts wherever these can be obtained. So, the delimitation of its region is a challenge for many universities.

Local government in many OECD countries is highly fragmented with individual municipalities having limited powers and resources to engage in economic development generally, let alone with higher education. In some countries, municipalities pool resources across several units and/or establish joint development agencies with a responsibility to work with universities in the combined area. At the next level of aggregation (or disaggregation of the national governance system), some countries have politically powerful regional authorities with a specific mandate to support higher education in their region. This is the case in the Spanish autonomous regions and the Provinces of Canada. In highly centralized countries like the UK, the national government has devolved some powers to the countries of Scotland and Wales including some aspects of higher education. Within England, special development agencies in each of the ten regions have been established by the central government. These agencies have some autonomy and are increasingly seeking to mobilize HEIs in support of economic development even though higher education remains a central function.

In attempting to engage with some level of government between the national and local and even when there is a specific regional administrative structure in place, universities often face challenges of intraregional competition for their attention. Relating to the specific municipality in which they are located is one thing – serving a multitude of locations across a broader region with several centers of population is another. Multi-campus solutions raise questions of dilution of resource and partnerships between several HEIs across a region and can be very demanding of senior-management time and energy.

Finally, identifying who speaks for the private sector in relation to what higher education has to offer can be

challenging, especially in regions without a strong private sector research and development base. In strong and dynamic regions, there are often well-developed private sector networks that are plugged into higher education and articulated through chambers of commerce. But in weaker regions, the SME sector is often inchoate and there are not well-developed industrial clusters. In such regions, branches of national and international companies can lack the autonomy to engage with universities for the development of new products and services and to provide placements for students and jobs for graduates.

In summary, the environment for universities to engage in regional development across OECD countries is highly variable. Where the governance and business structure is poorly developed and where there is no strong regional leadership, it is often necessary for HEIs to not simply respond to regional needs but to set the development agenda. Whether the HEIs are able to do this depends on their own governance, leadership, and management.

Barriers in the Governance, Leadership and Management of HEIs

Regional engagement is a challenge for universities, particularly for longer-established institutions organized around academic disciplines and along a supply-driven agenda. Most universities recognize the importance of teaching quality and research excellence and link these qualities to the cross-cutting roles of vice rectors (as distinct from the disciplinary roles of deans and heads of department). However, the integration of teaching and research within the disciplines to deliver regional impact is seldom recognized.

The all-embracing nature of regional engagement implies that this is a task for the head of the university. He/she can integrate the function and disciplinary areas and represent the corporate view of the institution externally. In many cities and regions, rectors and vice chancellors are key members of local elites, participating in many forums. At the same time, individual academics or other staff members may be active as business or social entrepreneurs in projects supported by the city and region. But in many instances, there is little connection between the high-level engagement of the senior management and the actions of individual academics. Indeed, the customs and practices of the institution may act as a barrier to more systematic engagement across the institution.

There are numerous institutional barriers particularly within research-intensive universities. First and foremost is the lack of incentive to individuals. Few institutions recognize regional engagement as one of the grounds for academic promotion; this is characteristically based around research excellence as reflected in peer-reviewed

publications with an occasional nod toward innovative teaching or academic management.

Second, resources to support the development of ideas (proof of concept) into products, services, or public policies are often not available let alone translational research facilities to build prototypes or test drugs. Third, intellectual property can also be a major source of conflict between the academic and his/her institution even where the national legislative environment is favorable.

Fourth, continuing professional development for small businesses and the community does not easily fit into conventional full-time teaching programs and can require evening and weekend teaching, eating into time for research and scholarship. Finally, problem-solving research and development for local SMEs (who may have difficulty in formulating their needs) can be very time consuming and diversionary from what are regarded as core activities.

How far are these barriers to institutional mobilization, in support of regional development, a function of traditional forms of institutional governance and how far are they a matter of the underfunding of the third task? The evidence from the OECD countries suggests that it is a combination of both factors.

Enhancing the development of more entrepreneurial universities is becoming an objective of new higher education policies in many countries. Some OECD member states, for example, the Netherlands, Austria, the United Kingdom, and Denmark, which have embraced new public management principles, have replaced collegial forms of governance and management (i.e., elected rectors, deans, and heads of departments) by a system of stronger and more overt managerial roles undertaken by appointed vice chancellors or rectors and heads of faculties. However, while governments recognize that more leeway needs to be granted to higher education managers, reducing the burden of regulation does not necessarily proceed at a fast pace.

In many OECD countries, universities still have limited autonomy (in contrast to the autonomy of the academic staff) in terms of their mission, academic profile, program offer, and management of human resources and infrastructure. In particular, the ability of the university to exercise control over its estate can be a key asset in supporting engagement with city and regional development but as this is a significant financial-resource responsibility, it is often retained by the central government.

Where governance of universities has not been changed, the national government has often looked to new institutions, notably polytechnics, to address the regional-development task. Such institutions characteristically are strongly managed. The internal mechanisms which mobilize the institutions to support the region are well tuned using a variety of performance measures. However, these institutions characteristically lack a strong research base capable of transforming a regional economy

as distinct from improving the existing industrial base. In these instances, delivering the higher education capacity that has both global reach and local engagement may require strong collaboration with research-intensive universities – a further challenge for the leadership.

Reference to the entrepreneurial approach is not to imply that this is the appropriate model to ensure that a university actively engages in regional development. An institution with greater freedom of action may well pursue the achievement of international status rather than local utility. The challenge for academic leaders wishing to build a locally engaged and globally competitive institution is to manage the tensions arising from the different rationalities embedded within higher education. The key task is to produce a synthesis through which the institution not only responds to regional needs but also becomes a motor for regional development by mobilizing a strongly independent academic heartland. Indeed, with the right institutional and civic leadership, partnership with the local community can become a crucible within which more dynamic and open universities can be forged, both responding to and shaping developments in the wider society.

Further Reading

- Boucher, G., Conway, C., and Van Der Meer, E. (2003). Tiers of engagement by universities in their region's development. *Regional Studies* **37**, 887–897.
- Castells, M. and Hall, P. (1994). *Technopoles of the World: the Making of 21st Century Industrial Complexes*. London: Routledge.
- Charles, D. (2003). Universities and territorial development: Reshaping the regional role of UK universities. *Local Economy* **18**, 7–20.
- Clark, B. R. (1998). *Creating Entrepreneurial Universities: Organizational Pathways of Transformation*. Oxford: Pergamon.
- Economic Geography of Higher Education: Knowledge Infrastructure and Learning Regions. In Rutten, R., Boekema, F., and Kuijpers, E. (eds.) (2003). London: Routledge.
- Etzkowitz, H. and Leydesdorff, L. (2000). The dynamics of innovation: From national systems and mode 2 to a triple helix of university-industry-government relations. *Research Policy* **29**, 109–123.
- Fritsch, M. and Schwirten, C. (1999). Enterprise-university co-operation and the role of public research institutions in regional innovation systems. *Industry and Innovation* **6**, 69–83.
- Geiger, R. L. and Sá, C. (2005). Beyond technology transfer: US state policies to harness university research for economic development. *Minerva* **43**, 1–21.
- Harloe, M. and Perry, B. (2004). Universities, localities and regional development: The emergence of the 'mode 2' university? *International Journal of Urban and Regional Research* **28**, 212–223.
- Higher Education Management and Policy (2008). *Special Issue: Higher Education and Regional Development*. *Higher Education Management and Policy* **20**, 9–183.
- Lawton-Smith, H. (2007). Universities, innovation, and territorial development: A review of the evidence. *Environment and Planning C: Government and Policy* **25**, 98–114.
- Marginson, S. (2002). Nation-building universities in a global environment: The case of Australia. *Higher Education* **43**, 409–428.

Mowery, D. C. and Sampat, B. N. (2005). Universities in national innovation systems. In Fagerberg, J., Mowery, D. C., and Nelson, R. R. (eds.) *The Oxford Handbook of Innovation*, pp 209–239. Oxford: Oxford University Press.

OECD (2007). *Higher Education and Regions: Globally Competitive, Locally Engaged*. Paris: OECD.

OECD (2008). *OECD Review of Tertiary Education: Final Report*. Paris: OECD.

Public Universities and Regional Development. In Mohrman, K., Shi, J., Feinblatt, S. E., and Chow, K. W. (eds.)(2009). Chengdu: Sichuan University Press.

Slaughter, S. and Leslie, L. L. (1997). *Academic Capitalism: Politics, Policies, and the Entrepreneurial University*. Baltimore, MD: Johns Hopkins University Press.

Relevant Websites

<http://www.oecd.org> – OECD: Project on supporting the contribution of higher education institutions to regional development 2004–07.

Universities' Engagement with Society

D Watson, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Until the advent, in the late twentieth century, of company or for-profit universities, all university institutions grew in some way from the communities that originally sponsored them. These acts of foundation varied according to a range of local circumstances, in time and location. Many such founding commitments have been transformed – positively and perversely – over the ensuing years, but the familiar image of a university as somewhat separate from its community – as, for example, an ivory tower – is curiously unfaithful to the historical record.

This article explores the foundation and subsequent history of types of institution around the world from the perspective of their relationship with the community, civil society, and the state, before attempting a contemporary assessment of who can claim to own the university enterprise, and with what implications.

Foundations

Most university foundations had an immediate element of service to the community in their agreed mission and purpose. The idea of responsiveness to social priorities was much more central to the founding goals of their institutions than many subsequent generations of university leaders and members have come to believe. There is a pattern here, as set out in the following stages of establishment of universities. Stephen Lay's *The Interpretation of the Magna Charta Universitatum* provides an excellent overview of these developments culminating in the proposal that "an expectation of public service could be written into the Magna Charter and thus become a guideline for university administration" (Lay, 2004: 18–71, 109).

The early foundations were specialist communities such as the late medieval colleges for poor scholars in England (Oxford and Cambridge), and for urban professionals (such as Bologna and Paris in continental Europe). Three centuries later, a similar trajectory was followed by the American colonial seminaries (many of which subsequently became expensive private schools in the United States, including the heart of the Ivy League). Lay points out that what distinguished all of these foundations from their ancient predecessors was the presumption of independence from the state, or what has subsequently become termed autonomy.

After a further fallow period, the next significant wave of foundations took place in the mid- and late-nineteenth century. These grew similarly out of perceived social and

economic needs, but in the radically different context of industrializing societies. Examples are the great Victorian and Edwardian Civic universities in the UK and the Morrill Act-inspired Land Grant universities of the American West and mid-West, leavened by specific, primarily research-based institutions on the German Humboldtian model, such as Johns Hopkins.

It is often helpful to go back to the founding acts or charters to see what was intended. The illustration below is of extracts from the royal charter of what is now the University of Sheffield. Note the emphasis on practical knowledge, as well as the sense of place (**Case study 1**).

In the next wave of development, the mid-twentieth century saw the establishment of local authority-based public systems of higher education, as in the English polytechnics, the Scottish central institutions, and American state systems (of which the archetypes are Wisconsin and the Californian Master Plan). These were equally specifically tied to expectations about relevant education and training, with a new element of ensuring both access by groups previously underrepresented, and progression. In many societies, the result was to create what came to be known as binary systems of higher education – a group of traditional university institutions contrasted with a more local, apparently more locally accountable, and apparently more responsive pattern of provision. Around the turn of the twenty-first century, this juxtaposition posed real dilemmas for policymakers dealing with the advent of mass higher education. Those with binary systems felt that they had run their course; those without them felt that the only way to re-inject mission diversity was to try to create a polytechnic-style counterpoint to unresponsive autonomous universities; others who had tried the change decided they needed to change back.

These were followed by late-twentieth-century experiments in curriculum, pedagogy, and a further drive toward accessibility (such as, notably, the pioneering of open access, or admission of adults without formal qualification by the UK's Open University and New York's City College system, and their imitators around the world). At the same time, nations began to establish the mega-universities, as analyzed by John Daniel, making use of open and distance learning (ODL) technologies to speed up participation, and to cut costs. The Indira Gandhi National Open University, founded in 1985 had 1.4 million enrolments in 1996, and the Islamic Azad University had 1.2 million. The notion of community interest is thereby dramatically expanded.

Case study 1

Extracts from the Charter granted by Edward VII in 1905, to convert the University College of Sheffield (founded 1836) into the University of Sheffield

To provide for:

- Instruction and teaching in every Faculty.
- Such instruction in all branches of education as may enable students to become proficient in and qualify for degrees diplomas associateships and certificates in arts pure science applied science commerce medicine surgery law and all other branches of knowledge.
- Such instruction whether theoretical technical artistic or otherwise as may be of service to persons engaged in or about to engage in education commerce engineering metallurgy mining or in other industries or artistic pursuits of the city of Sheffield and the adjacent counties and districts.
- Facilities for the prosecution of original research in arts pure science applied science medicine surgery law and especially the applications of science. (University College Charter, 1905: Paragraph 14).

However, even the experiments in ODL built upon traditional foundations. In 2007, the University of London's external degree scheme (which celebrated 150 years of such business in 2008) supports 43 000 students in 183 countries.

Finally, the latter part of the twentieth and beginning of the twenty-first centuries have seen significant action on the frontier activity between compulsory education, optional tertiary provision, and the initial rungs of higher education. Examples are the UK phenomenon of higher education in further education and the vitally important American Community College network: the former, especially in the provision of intermediate qualifications such as the higher national certificates and diplomas and foundation degrees, and the latter through 2-year (when taken full-time) associate degrees. The latest descriptor of activity in this borderland is that of dual sector provision.

These latter two waves of developments illustrate that as communities have changed – most recently in response to global communications – not only have existing universities had to respond, but also the acts and intentions of foundation of new institutions have also adapted.

Communities, Civil Society, and the State

The big question that then arises is about how universities as institutions fit into the structure and function of public life. Various theoretical models are available, and one of the most influential conceptualizations in modern political philosophy can be used. Writing in Prussia in the early nineteenth century (a decade after Humboldt's founding of the University of Berlin), Hegel saw a rising standard of

collective identity, behavior, and self-realization through three circles: the family and community (or what he called ethical life), civil society, and the state. The intermediate point – of civil society – he described as follows in *The Philosophy of Right* (1821):

In the course of the actual attainment of selfish ends – an attainment conditioned in this way by universality – there is formed a system of complete interdependence, wherein the livelihood, happiness, and legal status of one man is interwoven with the livelihood, happiness, and rights of all. On this system, individual happiness etc. depend, and only in this connected system are they actualized and secured. . . . (para. 183) (Knox, 1952: 123).

For Hegel, the Prussian State was an ethical end-point, a proposition which it would be hard to defend today:

The state is the actuality of the ethical idea. It is ethical mind *qua* the substantial will manifest and revealed to itself, knowing and thinking itself, accomplishing what it knows and in so far as it knows it. The state exists immediately in custom, mediately in individual self-consciousness, knowledge and activity, while self-consciousness in virtue of its sentiment towards the state finds in the state, as its essence and the end and product of its activity, its substantive freedom (para. 257) (*Ibid.*: 155).

The contention here is more modest: the university occupies a critical role within civil society. According to Michael Edwards of the Ford Foundation, “as a concept ‘Civil Society’ speaks to the best of us, and calls upon the best of us to respond in kind” – and so should the university (Edwards, 2004: 3). There is a corollary: when it becomes over-identified with the political interests of the state, it has probably lost its way. The sticking point is well-articulated by Michael Daxner, former president of Oldenburg University and post-war European Union (EU) education commissioner in Kosovo. “East of Vienna,” he has said, “the role of universities is in society-making, not state-making.” Universities are needed, he says “because of our dangerous knowledge.” At a conference of the European University Association and the American Council of Education in 2004, he went on to explain how this priority can easily become masked:

No wonder that most of the harmonizing structures in higher education refer to pure scholarship, administration, government, and institutional autonomy, whereas the basic notions of the university as the ‘lead institution’ in civil society – republican legitimacy, democracy, and citizenship – are rarely included in modern concepts of academic freedom, or treated only nominally in the mission statement of universities (EUA/ACE 2004: 64).

The notion of dangerous knowledge – that is of being critical as well as supportive of activities across civil

society leads to moral injunctions for both states and their universities. How do these work in practice? In a work on *Managing Civic and Community Engagement*, it has been suggested that universities relate to their communities in three main ways (Watson, 2007: 132–141).

First-Order Engagement

First-order engagement arises from the university simply being there. One of the primary roles for universities is to produce graduates who go to work (perhaps in areas completely unconnected with those they have studied); who play their parts in civil society (where the evidence suggests they are likely to contribute more than if they had not been to university); who have families (and read to their children); who pay their taxes (and return a proportion of their higher-than-average incomes as graduates through progressive taxation); and who support their universities through gifts and legacies. An analysis of national cohorts in the UK summarized in the work by Schuller *et al.* (2004) on *The Benefits of Learning* has securely established that graduates are not only wealthier, but also happier, healthier, and more democratically tolerant than their non-graduate peers.

Also, in this first domain, universities guard treasures (real and virtual) in their museums, galleries, and archives. They provide a safe place for the exploration of difficult issues or challenging ideas. They also supply material for a branch of popular culture (the campus novel, film, and television series).

Together these features add resonance to the university as a social institution in its own right: at its best, a model of continuity and a focus of aspiration for a better and more fulfilled life; at its worst, a source of envy and resentment.

First-order considerations also imply that universities should strive to behave well, to be ethical beacons. Universities can choose to behave well or badly in a number of different directions, in relation:

- to applicants (and their families);
- to students (and their sponsors);
- to staff (of all kinds);
- to the local community (or neighbors);
- to the institutions of civil society (as above);
- to investors and supporters (the stakeholders discussed below);
- to government (in their role as a commentator and contributor to policy, as well as a deliverer of a public service);
- to global citizenship (e.g., by progressive engagement with political, economic, social, and environmental issues); and
- to groups of other higher education institutions (locally, regionally, nationally, and internationally).

Some of the sticking points include the following. Universities can offer misleading promotion and advice, to staff, students, and potential students, about their real performance and intentions. As powerful institutions they can undermine and intimidate their members, their partners, and their clients. They can perpetuate self-serving myths. They can hide behind specious arguments (narrow constructions of academic freedom, *force majeure* and the like). They can displace responsibilities, and blame others. They can fail the stewardship test (e.g., by not assessing and responding to risk, by cutting corners, or by failing to safeguard their assets). They can be bad neighbors. Above all, they can fail to tell the truth to themselves at least as easily as failing to tell the truth to others.

Second-Order Engagement

Second-order engagement is generally structured and mediated by contracts. In this domain, the university produces graduates in required disciplines and professional areas (whether directly or indirectly required to do so). It responds to perceived needs for particular skills, or for professional updating, or to more general consumer demand for courses in particular subjects. It supplies services, research and development, and consultancy at either a subsidized or a for-profit rate. The university may run subsidiary businesses – some as spin-outs or joint ventures, others in the service sector of entertainment, catering, conference-organizing, or the hotel business.

Also, in this domain, the university is often an important local and regional economic player. It supplies employment – from unskilled occupations to the highly skilled. It provides an expanded consumer base, as students and staff are attracted to the institution and its locality. The university offers a steady, well-indemnified customer for goods and services. It is a source of development, such as of buildings, amenities, office space, and green spaces, although this has its downsides, like controversy over planning, car-parking, congestion, or studentification (the perceived takeover by temporary student residents of streets and neighborhoods, as well as the potential displacement of local residents from low-paid jobs).

The first domain affects the second in some complex and significant ways. The university, as a kind of moral force, is expected to behave better than other large organizations (which are similarly concerned about the bottom line).

Third-Order Engagement

Third-order engagement relates to commitments between the university and its members.

Universities are voluntary communities: around the world, they are rarely part of the compulsory educational

infrastructure of the state. Thus, they should not be regarded as agents of the state in creating citizens or subjects. This is not to say, following the precepts of first-order relationships, that universities do not play a role in ensuring social cohesion, in promoting community solidarity, and in problem-solving for policymakers and practitioners of all kinds.

University members have a similar set of obligations as individuals; this is the dimension of academic citizenship. To be a full member of a university requires more than completing basic, obvious tasks. For traditional academics, this has meant collective obligations: to assessment, to committee membership, and to strategic scoping. There is a growing body of literature about such professional academic practice.

Since the late twentieth century, such practice has been recognized as no longer belonging exclusively to the ranks of the faculty. The teaching, research, and service environments are increasingly recognized as being supported and developed by university members with a variety of expertise (e.g., finance, personnel, estates, libraries, and information and communications technology), each with their own spheres of professional competence, responsibility, and recognition.

At the heart of academic citizenship is the concept of membership. As consumers, students have entitlements and expectations. Both students and staff have responsibilities, along with all of their rights, within the community. Such responsibilities include the following:

- a special type of academic honesty, structured most clearly around scientific procedure;
- reciprocity and honesty in expression, for example, avoiding plagiarism by accurately and responsibly referring to other people's work within one's own;
- academic manners, such as listening to and taking account of other people's views;
- self-motivation and the capacity for independent learning, along with learning how to learn;
- submission to discipline (most clearly in the case of assessment – for both assessors and the assessed);
- respect for the environment in which members of the college or university work; and
- adherence to a set of collectively arrived at commitments and policies (on equalities, grievances, harassment, etc.).

Who Owns the Modern University?

According to the conventions of corporate governance, organizations are governed in the interests of either shareholders (the institutions, groups, or individuals who own the shares – and expect dividends) or stakeholders (the individuals and groups, including the staff – whose interests might

be affected by aspects of the organization's performance). To the frustration of several commentators, most universities are neither shareholder nor stakeholder institutions. On one end of a spectrum, institutions with the university title may be wholly for-profit institutions, especially in jurisdictions where the title is relatively unprotected by law. At the other end, they may be unmediated emanations of the apparatus of the state. However, in the vast center-ground, they are unashamedly *sui generis*, with, as Lay says, institutional autonomy lying at the heart of the conception of the modern university.

In these circumstances, who owns the university (or pieces of it) or thinks that they do? There are several potential candidates.

The state, directly and indirectly, is invariably a major funder. It will also claim to represent the people's share by investing the proceeds of taxation. However, attempts to co-opt universities into politically influenced national priorities is, as Daxner warns, dangerous. Nor is it likely to work. Early results from a 15-country project on the role of universities in the transformation of societies would urge caution on the more aggressive advocates of higher education and the national interest. Brennan *et al.* (2004) find, in general, a relatively weak role for higher education in stimulating economic change, complex and contradictory influence on political change, and a social role that is at least as much about reproduction as about transformation.

Other big investors may be other public services, the professions, business, and employers, including through sponsorship and purchase of student places. The professions are a particularly interesting case. They were in at the beginning of the modern European university (law at Bologna; theology at Paris), and they played their part in the nineteenth- and twentieth-century expansion of the system (science and technology in the civics – especially engineering throughout the Commonwealth, and beyond – and more recently the addition of health professions to the traditional formation of doctors of medicine).

Then there is the public more generally, especially as refracted through the media. There are contrasts between cultural roles of universities and colleges in different national contexts: in the United States, they are more loved and respected than may be deserved; in Australia and the UK, they stimulate more opprobrium than is objectively fair. This picture may, however, be changing, as US higher education is hitting – almost for the first time – a combination of cuts in public subsidy, consumer resentment, and consumer debt.

So far, the various communities which interact with and within the university have been outlined.

But probably most important in the historical sense are the members of the university. Here a degree of detachment and self-knowledge is required. As Gordon Graham concludes his study of *The Institution of Academic Values*,

“when it comes to debates and disagreements about their own affairs, universities are as prone to self-protecting flights of unreason as any other institution” (Graham, 2005: 280).

To describe all these as stakeholders is by no means straightforward. Stakeholder, is one of those words which has almost exactly the opposite meaning now, since when it was originally coined. The stakeholder used to be the person who held the coats – and the prize-money – while the fight was on; the notion was one of scrupulous disinterest. Stakeholders need to understand that if they are to live up to the modern designation (as having invested something themselves), they have to put something at risk. A rigorous stakeholder analysis from the perspective of the university would throw up some surprising results. Questions that arise include the following: whose are the stakes on the table (really) in the sense of sharing risk? And who can most effectively (i.e., legitimately as well as logically) claim to hold the third party stake (the celebrated people’s money) on behalf of the community as a whole? The politicians would like to claim the latter is theirs – through democratic validation – although they too can fail the stewardship test.

The Challenge of Engagement

What are the implications of this analysis for universities in the early twenty-first century? One response, as set out in Iacobucci and Touhy’s *Taking Public Universities Seriously* has been to try to establish systematic compacts between universities and all of their stakeholders, including the government (Iacobucci and Tuohy, 2005: 11–19). The UK’s National Committee of Inquiry into Higher Education (the Dearing Committee of 1996–97) tested this approach – some would say to destruction (Watson, 2007: 18–28).

Another response has been to rediscover the role of universities and colleges as fully engaged members of their communities. An emphatic lead has been taken here by the Association of Commonwealth Universities (ACU). According to the Association of Commonwealth Universities (ACU), “engagement is [now] a core value for the university.” In the Association’s widely circulated consultative paper (published as *The Idea of Engagement: Universities in Society*), this proposition was revealed as follows:

Engagement implies strenuous, thoughtful, argumentative interaction with the non-university world in at least four spheres: setting universities’ aims, purposes and priorities; relating teaching and learning to the wider world; the back-and-forth dialogue between researchers and practitioners; and taking on wider responsibilities as neighbours and citizens (Bjarnson and Coldstream, 2003: 312–313).

Case study 2

The Talloires Declaration (2005)

- Expand civic engagement and social responsibility programs in an ethical manner, through teaching, research and public service.
- Embed public responsibility through personal example and the policies and practices of our higher education institutions.
- Create institutional frameworks for the encouragement, reward and recognition of good practice in social service by students, faculty, staff and their community partners.
- Ensure that standards of excellence, critical debate, scholarly research and peer judgement are applied as rigorously to community engagement as they are to other forms of university endeavor.
- Foster partnerships between universities and communities to enhance economic opportunity, empower individuals and groups, increase mutual understanding and strengthen the relevance, reach and responsiveness of university education and research.
- Raise awareness within government, industry, charitable, not-for-profit and international organizations about higher education’s contributions to social advancement and well-being. Specifically, establish partnerships with government to strengthen policies that support higher education’s civic and socially responsible efforts. Collaborate with other sectors in order to magnify impacts and sustain social and economic gains for our communities.
- Establish partnerships with primary and secondary schools, and other institutions of further and higher education, so that education for active citizenship becomes an integral part of learning, at all levels of society and all stages of life.
- Document and disseminate examples of university work that benefits communities and the lives of their members.
- Support and encourage international, national and regional academic associations in their efforts to strengthen university civic engagement efforts and create scholarly recognition of service and action in teaching and research.
- Establish a steering committee and international networks of higher education institutions to inform and support all their efforts to carry out this Declaration.

In similar vein, in September 2005 Tufts University brought together leading figures from universities across the world at their conference center in Talloires, southwest France. The meeting resulted in a draft declaration on the civic roles and responsibilities of higher education which has attracted signatories from around the world, and which is now being used to design systematic interventions, for example, on literacy (**Case study 2**).

These are serious challenges for the leaders and managers of higher education institutions. They can absorb some lessons from history. Without a sense of obligation to the spirit of its foundation, university communities will find it hard to develop and adhere to a core set of values underlying strategic choices.

There are some lessons from the needs of civil society, especially through partnerships of various kinds. This is the necessary moving picture against which the institutions

need to test their plans. It is also the context for what Donald Kennedy has eloquently described as academic duty.

And there are some lessons about restoring the traditional role of universities in the special circumstances of a globalized knowledge economy. The post-Enlightenment project of a liberal higher education requires constant reinvention to meet new needs. In the process of its early twenty-first century reinvention, engagement is indeed a core value.

See also: Academic Freedom; Higher Education and the Transformation of Society; Universities and Regional Development.

Bibliography

- Bjarnson, S. and Coldstream, P. (2003). *The Idea of Engagement: Universities in Society*. London: Association of Commonwealth Universities.
- Brennan, J., King, R., and Lebeau, J. (2004). *The Role of Universities in the Transformation of Societies: An International Research Project (Synthesis Report)*. London: Centre for Higher Education Research and Information (CHERI).
- Edwards, M. (2004). *Civil Society*. Cambridge: Polity.
- European University Association/American Council on Education (EUA/ACE) (2004). *Charting the Course between Public Service and Commercialisation: Prices Values and Quality Conference. Proceedings*, Turin, 3–5 June.
- Graham, G. (2005). *The Institution of Intellectual Values: Realism and Idealism in Higher Education*. Charlottesville, VA: Imprint Academic.

- Iacobucci, F. and Tuohy, C. (eds.) (2005). *Taking Public Universities Seriously*. Toronto, ON: University of Toronto Press.
- Knox, T. M. (1952). *Hegel's Philosophy of Right: Translated with Notes*. Oxford: Clarendon Press.
- Lay, S. (2004). *The Interpretation of the Magna Charta Universitatum and its Principles*. Bologna: Bononia University Press.
- Schuller, T., Preston, J., Hammond, C., Brassett-Grundy, A., and Bynner, J. (2004). *The Benefits of Learning: The Impact of Education on Health, Family Life and Social Capital*. London: Routledge Falmer.
- Talloeires (Talloeires Conference on the Civic Roles and Social Responsibilities of Higher Education) (2005). *Strengthening the Civic Roles and Social Responsibilities of Higher Education; Building a Global Network*. Tufts University. <http://www.tufts.edu/talloeiresnetwork/conference.shtml> (accessed May 2009).
- Watson, D. (2007). *Managing Civic and Community Engagement*. Maidenhead: McGraw-Hill/Open University Press.

Further Reading

- Daniel, J. (1996). *The Mega-Universities and Knowledge Media*. London: Routledge.
- Kennedy, D. (1997). *Academic Duty*. London: Harvard University Press.
- National Committee of Inquiry into Higher Education (NICHE) (1997). *Higher Education in the Learning Society (The Dearing Report)*. London: HMSO.

Relevant Websites

- <http://www.shef.ac.uk/calendar> The University of Sheffield.
- www.president.tufts.edu Tufts University.

HIGHER EDUCATION – TEACHING AND LEARNING IN HIGHER EDUCATION

Contents

Distance Education and Open Universities

Education Development in Higher Education

Effective Use of Technology in Teaching and Learning in HE

Facilitating Students' Success – Student Support and Learning Resources in Universities

Faculty Values and Expectations and Their Implications for Teaching, Learning and the Organization of Higher Education Institutions

Flexible Learning in Higher Education

Innovation in Teaching and Curriculum Design

Productivity of University Faculty Staff

Promoting Effective Student Learning in Higher Education

Student Attrition from Higher Education Institutions

University Students' Experiences of Higher Education

What Do University Students Learn?

Distance Education and Open Universities

A Kanwar and J Daniel, Commonwealth of Learning, Vancouver, BC, Canada

© 2010 Elsevier Ltd. All rights reserved.

Introduction

There are more entrants to higher education today than at any time in the past. Higher education (HE) has, in fact, evolved from an elitist pursuit into a mass system. By 2020, 40% of the global workforce will be knowledge workers, with a need for tertiary qualifications. The World Bank, which in the 1980s and the 1990s privileged basic education, is now of the view that for countries to achieve sustainable economic development, the age-participation rates (APRs) in HE must be in the region of 40–50%. With access to HE being less than 10% of the relevant age group in South Asia and sub-Saharan Africa, this key concern is being addressed by most developing countries. Malaysia plans to raise its APRs to 40% by 2010. The Government of Trinidad and Tobago plans to see an APR of 60% by 2015, while the Jamaican government expects to raise the APR in tertiary education to 30% by the same time. India has since announced the establishment of one central university for each of its 28 states and its intention to raise the APRs to 15% by 2012 (Daniel *et al.*, 2006: 17–23). Can existing institutions cater to the rising demand? Dhaka University, Bangladesh could only enrol 10 000 of the 80 000 applicants in 2000, while in Kenya only 9000 of the 40 000 qualified students could be accommodated in the public university system (Kapur and Crowley, 2008: 16).

As building brick-and-mortar institutions is not a viable option for most countries, especially developing ones, the search for alternative approaches will continue. Even though distance education has been in existence for over 150 years, it is only in the last 40 years that its growth has been phenomenal and it has emerged as a viable supplement and serious alternative to the formal system. In Africa, there was only one open university in 1988; today there are four. India has nearly six million students in its 14 open universities and 126 dual-mode institutions.

The emergence of new information and communication technologies (ICTs) offered the right tools to reach out to people irrespective of their location and specific circumstance. Time and space ceased to be barriers to education. This unprecedented demand for growth has led to the evolution of new institutional structures, flexible and innovative teaching and learning styles that have transformed the traditional approaches to education. The role played by open and distance learning (ODL) is a key element in this metamorphosis.

Distance Education and Open Learning

Distance education is the delivery of learning or training to learners who are separated, mostly by time and space, from those who are teaching and training. As learners and

teachers are separated by time and space, some kind of technology or media must be used for communication between them.

The term open learning describes policies and practices that permit entry to learning with as few barriers as possible. For example, when the Open University in the UK was created in 1969, it based its claim to be open by abolishing all academic prerequisites for entry. Potential learners did not have to show any evidence of having been to school or passed any examinations. Once they start, they have to work hard to pass courses and continue – but there are no barriers to starting.

But the Open University was not open on some other dimensions. Students could not enrol unless they were aged over 21, so there was an age barrier. They had to study at a prescribed pace, between fixed dates for starting and finishing, so there was a barrier of time. On the other hand, when seen from the perspective of Athabasca University in Canada, that did not have an age barrier and allowed students to start a course at the beginning of any month and take as long as required, the UK Open University did not appear to be particularly open.

Open learning, then, is a relative term but it does express the ambition of an institution to remove those barriers to learning that are particularly restrictive in its own environment.

While distance education refers to the methodology by which the learners are reached, open learning refers to the philosophy underlining open entry and access to learning opportunities. Though the two are conceptually distinct, distance education and open learning are often complementary. Opening up learning without introducing some elements of distance education is not always possible and introducing distance education inevitably opens up learning in new ways for many people.

Hilary Perraton has pointed out that the term distance education captures the economic imperative of gaining benefits of scale, low-cost, and consistent quality, whereas open learning evokes the political imperative of widening access.

For this reason, it is customary to bring these two terms together in the expression open and distance learning or ODL. New terms, however, have evolved to designate forms of ODL that include more than just print and paper. For instance, e-learning, online learning, and virtual learning refer to learning with a variety of media, not just computers. Other terms in currency include m-learning; multimedia education, technology-mediated learning, and technology-enhanced learning. Another recent name that reflects the growing trend of mixing ODL with conventional face-to-face teaching is flexible learning.

The Evolution of ODL

Initially, distance education catered to children in remote locations with no access to school and to adult learners

who had missed the opportunity for higher education for one reason or another. Today, distance education is seen as a cost-efficient and effective tool for providing education and training at all levels and for a variety of disciplines and professions. Developments in ICTs have made it possible to make effective pedagogic interventions for providing unlimited flexible learning options to several constituencies of learners. In the process, distance education is emerging as an effective tool not just for education but for development generally.

When Britain introduced a universal postal system, the penny post, in the middle of the nineteenth century, Isaac Pitman offered postal tuition for a course in shorthand almost immediately and correspondence education was born. It began as a private commercial venture and correspondence education has always retained a strong commercial component, even though governments and public universities also began offering traditional education programs through correspondence in the late nineteenth century. The UK University Correspondence College (1887) was set up to support external students to prepare for the University of London's external degree. Early correspondence-course providers included Illinois Wesleyan University (1874), Queen's University in Kingston, Ontario, Canada (1889), the university extension department of the University of Chicago (1891), and the University of Queensland in Australia (1911). In the 1920s, a range of universities and other tertiary-level institutions offered programs solely at a distance, including a number of Soviet all-union correspondence polytechnics (Rumble and Koul, 2007: 11).

As the name implies, correspondence education usually involved mail going in two directions: the school mailed instructional texts to the students; each student mailed back assignments which would be corrected and commented on by a tutor and returned to the student. Holmberg (1977), called this process a guided didactic conversation, a pillar of good distance-education practice.

For over a century, until the second-half of the twentieth century, correspondence education brought great benefits to large numbers of people. This contribution was never properly recognized, partly because it involved individuals learning privately and partly because the correspondence schools were mostly private enterprises. This combination, private individuals dealing with private companies, was also the explanation for the abuses that gave correspondence education a bad name and caused a backlash against it in the 1960s.

The Rise of the Open University

The University of South Africa (UNISA) started its distance-education operations in 1946 and was interracial even

during the apartheid years. It is highly rated in South Africa and is one of the world's mega-universities.

But it was the establishment of the UK Open University (UKOU) in 1969 that led ODL to expand phenomenally in the 1970s and 1980s. Two trends came together in the 1960s to initiate a revolution that brought together open learning and distance education in an explicit manner and dramatically increased the impact and effectiveness of ODL. First, the 1960s was a time when most governments gave high priority to expanding education at all levels. Second, it was a time of effervescence and enthusiasm for communications technology. Television was seen as far too important to be limited to entertainment alone and the UKOU used it creatively to lead a revolution in bringing education to the masses.

Throughout history, education has been constrained by the iron triangle of quality, access, and cost. If access is increased, there is the danger of lowering quality. If this is to be avoided, then the costs would have to be raised. ODL is revolutionary because it does allow, through division of labor, specialization, and economies of scale, to reconfigure the access–quality–cost triangle. Access can be increased, quality can be improved, and costs can be cut, all at the same time.

But this is by no means an automatic process. While research shows that there is no significant difference between distance education and traditional classroom instruction in relation to student outcomes, there are still the issues of quality and the recognition of qualifications that need to be addressed constantly.

Harnessing the mass media to higher education enabled the UKOU to excel in all three dimensions and reshape the triangle. It immediately became the largest university in Britain. Independent rankings of the quality of the teaching programs in UK universities put the Open University in fifth place out of 100 institutions. Cost studies conducted by the UK government show that whichever way the calculations were done, the total cost of the Open University degree was substantially less than that in conventional institutions.

ODL is revolutionary because it does allow, through division of labor, specialization, and economies of scale, to reconfigure the access–quality–cost triangle. Access can be increased, quality can be improved, and costs can be cut, all at the same time. That was the message of the Open University; something that had never happened before in the history of education.

Mega-Universities

It was not surprising that numerous open universities came to be established thereafter in many countries across the world. Most of them soon became popular enough to register enrolments of over 100 000 students each. Notable among them are those in India, Indonesia, Iran, Korea,

Pakistan, Spain, Thailand, and the US. These new open universities, along with some older institutions in China, France, and South Africa came to be known as mega-universities. In 1999, there were 11 such universities in the world with a collective enrolment of about three million students (Daniel, 1996). Today, the number of mega-universities has doubled, among them figure the following:

- Allama Iqbal Open University, Pakistan;
- Andalu University, Turkey;
- Athabasca University, Canada;
- Bangladesh Open University, Bangladesh;
- *Centre National d'enseignement à distance*, France;
- China Central Radio and TV University, P.R. China;
- City College of San Francisco, USA;
- Dr. B. R. Ambedkar Open University, India;
- Indira Gandhi National Open University, India;
- Korea National Open University, South Korea;
- Madhya Pradesh Bhoj Open University, India;
- Netaji Subhash Open University, India;
- Open University, UK;
- Payame Noor University, Iran;
- Shanghai TV University, P.R. China
- Sukhothai Thammathirat Open University, Thailand;
- *Universidad Nacional de Educacion a Distancia*, Spain;
- *Universitas Terbuka*, Indonesia;
- University of Maryland University College, USA;
- University of Phoenix, USA;
- University of South Africa, South Africa;
- Yashwantrao Chavan Maharashtra Open University, India.

Dual-Mode Institutions

In all countries, conventional educational institutions, most notably at university level, are adding ODL activities to their face-to-face teaching and becoming dual-mode institutions. Australia has many such institutions: Deakin University, Monash University, and University of Southern Queensland. The University of Delhi has a large campus of open learning with student enrolments in the range of 150 000. The University of South Pacific and the University of West Indies are multimodal institutions and cater to both face-to-face and distance students on a regional basis.

However, it is not only campus institutions which are introducing distance education. Dedicated distance-education institutions such as the Open University of Hong Kong are offering face-to-face classes. In Canada, the single-mode ODL institutions in Québec and British Columbia, the *Télé-université* and the Open University of British Columbia, have been merged with conventional institutions.

E-Learning is gradually seeping into education at all levels and contributing to the trend of flexible education.

By 2006, in a typical American college or university, one fifth of all continuing and professional-development courses were being conducted online (Kapur and Crowley, 2008: 32–33). With 81% of face-to-face institutions in North America offering blended learning options, it is clear that there is an increasing convergence of distance and face-to-face education.

Key Features of ODL

Compared to campus-based institutions, open universities are large enterprises that require careful planning, efficient organization, and effective management to carry out a variety of operations that include the design and development of learning materials, their production, distribution and delivery, and the mobilization of physical and intellectual resources across wide geographical areas. Due to this basic feature, distance-education systems are sometimes referred to as an industrial form of education. They can involve thousands of faculty members, students, academic counselors and tutors, and staff at headquarters and in distant and far-flung learning centers. Necessary organizational and administrative components include clear statements of vision, purpose, and objectives; an integrated strategy for program development that includes curricula, teaching and learning, policies for students, tutors, and other staff; materials production and distribution; learner support systems and student assessment. Some of the critical success factors are:

1. *Self-instructional learning materials.* The key to the academic strength of an open and distance-education system is the learning materials around which its programs and courses are designed and delivered. As the learning process takes place at a distance from the teacher or the institution, interactive pedagogic processes are built into them and the materials are therefore, called self-instructional. Preparation of self-instructional materials is a complex professional skill that requires specialized training. Teams of specialists that comprise subject experts, media professionals, graphic designers, editors, and reviewers generally develop these materials. Extensive participation of both in-house and external experts ensures the quality and authenticity of the learning materials on which most open universities and dual-mode institutions build their academic reputation.
2. *Effective student support services.* Excellence of materials alone does not assure the quality of the outcomes. Just as important as the preparation of the materials is the effectiveness of the delivery systems that connect the students with their institutions. Most open universities and distance-education institutions need to develop extensive and intensive support services, both to help

students with the process of learning and with the new experience of being a distance learner. The challenge is that student support is inherently more labor intensive than simply using technology to allow students to learn independently. Student support through people does not have the same potential for economies of scale. Therefore, the fundamental dilemma in designing a distance-learning system is to ensure a good balance between using technology so that students can learn independently and using tutors so that they can be well supported.

3. *Logistics and administrative support.* The success of ODL systems depends mainly on the learning environments in which they operate. Creation of an effective learning environment is based on several factors – technology choices, learner preparedness, the design and development of the learner support systems, and the commitment of the leadership. Technology choices must be made only after scanning the environment to determine a variety of factors that include: the form of telecommunication that is within the reach of potential learners; the policies of governments that support and strengthen the applications of emerging technologies in educational development; the technology infrastructure that the ODL provider is able to create and maintain; as well as the type of programs that the institution can provide within the context of its mission and resources. An understanding of the learner's needs and characteristics can help integrate the chosen technology with the design of the program and is quite central to the success of any ODL operation. Any weaknesses in any of these links in the chain can have major consequences for the system as a whole leaving it with frustrated students, demotivated staff, and the loss of credibility.
4. *Networking.* A critical success factor in most ODL systems today is the ability of the provider institution to mobilize both physical and intellectual resources from a variety of sources for the design, development, and delivery of its programs. Networking with suitable institutions that are well-known in their areas of competence not only reduces institutional costs, but also ensures the quality of the teaching–learning transaction. The networking arrangements often extend to the acquisition and adoption or adaptation of learning materials, collaboration with major media houses to produce and deliver technology-supported and technology-enhanced learning packages, and at the purely local level, the support of the existing educational institutions to function as learning centers. As the theory–practice integration assumes critical importance, collaboration with the world of work is fundamental to the success of all education. For a variety of reasons, ODL institutions are well suited to implement industry–academy partnerships to foster this collaboration.

New Trends

The following trends will impact the development of ODL in the twenty-first century:

1. *ODL in the time of e-learning.* The most important trend today is e-learning. A feature of the history of education, going right back to the invention of the blackboard in 1850, is that each new technology is hailed as the harbinger of an educational revolution. Such claims were made for radio, film, television, programmed learning, and computers, and have been made abundantly for online communication.

Education, like many other areas of human life, was swept up in the dot.com frenzy of the late 1990s when some prophets argued that the Internet would be the only vehicle for education in the future. Studies show that e-learning has so far failed to deliver on these extravagant claims.

E-learning shares some of the features of correspondence education and, like correspondence education, attracts many private sector players. At its best, it enables a much more sustained guided didactic conversation than was possible with the postal service. At its worst, it allows digital diploma mills to take students' money before closing down their website and putting their profits in the bank.

The Internet is not, and probably never will be, a mass medium – even if the day comes when everyone is connected to it. Telephone is not radio, even in well-connected communities. Internet learning alone cannot have the mass impact on access that was achieved with the mass media of broadcast and print. This partly explains why the pure Internet learning operations that were launched during the peak of the dot.com frenzy have either disappeared or broadened out to take a multimedia approach.

In the last 10 years, the most exponential growth in ICTs has been in the area of mobile telephony. By 2010, there will be 2.5 billion mobile users in the developing world (Atkins *et al.*, 2007). A recent study at the University of Pretoria (UP) indicates that of about 14 000 teachers, mostly located in rural South Africa, who are enrolled in its distance-learning programs, only 1% have access to e-mail; but 99% own cell phones. The UP first began by using cell phones to provide administrative support and then academic messages. Through this personalized intervention from the university, students became more motivated than before and were satisfied with both the administrative and academic support provided (Hendrikz *et al.*, 2006). The implications of this development for the way we teach and learn are enormous.

2. *Open education resources (OERs).* The last 5 years have seen tectonic shifts in how technology is being used in education. Thanks to the prestige of the Massachusetts

Institute of Technology (MIT), the open courseware movement, based on the principle of sharing the knowledge contained in faculty lecture notes, marked the first generation in a movement to make knowledge our common wealth. The online course materials of the UKOU were the second generation as existing self-instructional materials were made freely available in online format. The third generation is collaborative course development as exemplified by the wikiEducator, a course-authoring tool being used to develop materials for the Virtual University for Small States of the Commonwealth (VUSSC) and many other projects. In this phase, the focus is shifting from this courseware is mine to this courseware is for (open) mining (Atkins *et al.*, 2007). The open education resources (OER) movement is largely based on four principles: (1) encouraging mass ownership rather than elitism; (2) acknowledging faith in everyone's inherent capability to self-organize; (3) enlisting amateurs as producers of content; and (4) promoting collaboration for the common good.

Content development is resource intensive and the OER movement provides a unique opportunity to developing countries to access global knowledge flows. Bernard and his colleagues carried out a meta-analysis of hundreds of studies in which distance-education students were treated in different ways. They distinguished three types of interactions: student–content; student–student; and student–teacher. They then analyzed all the studies to find which type of interaction made the greatest difference when it was increased. The results were very clear. Increasing student–content interaction had much the greatest effect; with student–student interaction coming next and student–teacher interaction last. Within this context, the importance of content cannot be underestimated. In the coming years, there will be a greater need to collaborate on free content development and sharing resources (2008). The open university of the twenty-first century will adopt and adapt existing content rather than create it.

3. *Cross-border higher education (CBHE).* In 1975, there were 600 000 international students globally. This figure quadrupled to 2.7 million in 2004. Existing unmet demand opens the door for cross-border tertiary education and there are at least 50 providers, registered and unregistered, in Jamaica alone. The number of cross-border providers in India increased from 27 in 2000 to 114 in 2004 (Daniel *et al.*, 2005). This is a growing business because of the general growth in interest in international education, because of demand from working adult students, who cannot travel abroad to study except for short periods, and because countries like Australia, as they approach saturation in imported students, are exporting the courses instead. The future of distance learning

across borders lies in partnership. Partnership is a sound principle because it helps develop an indigenous capacity for distance learning in the country concerned and because partnering with a local institution, that is credible or committed to becoming credible, can provide access to much larger numbers of students and can facilitate relationships with the national authorities.

National governments are grappling with the issue of diploma mills, online providers, and the recognition of foreign qualifications. Quality will be a key requirement of cross-border distance education.

4. *The diversity of the new learner.* The characteristics of the twenty-first century student have changed. Half the world's population (6.5 billion) is under 20 years with two billion teenagers in the developing world. In countries such as Malaysia and Pakistan, approximately 65% of the population is under the age of 30, while over two-thirds of the tertiary-education students in Singapore are over the age of 25 (Kapur and Crowley, 2008: 7). Today, a HE student may be anywhere between 18 and 50 years of age. Then there is the traditional young learner between 18 and 24 years.

There are relatively more women in HE today. Ghana, Kenya, Uganda, and Tanzania offer incentives for women candidates by offering lower admission cut-off points even as female enrolments in Africa compare unfavorably with other developing regions. In Bangladesh, women account for 34% of all enrolments in public and 17% in private institutions. In Brunei Darussalam, women at the tertiary level outnumbered males by 32% in 2004 and the trend continues (Kapur and Crowley, 2008: 53).

The new learner is a digital native, a twenty-something, who takes to technology as a fish to water. This is in contrast to the digital migrant (Prensky, 2005), the adult who has adopted technology relatively late in life. The digital native is a multitasker who can perform several tasks at the same time. The new learner can be the adult who needs continuing professional development combined with full-time employment. S/he would have little time for synchronous instruction. Such an academic customer is on the rise in both developed and developing contexts. The new learner belongs to a very diverse constituency and has a range of needs that the traditional institution thus far has never known.

Distance education has always dealt with a diverse range of learners and as conventional institutions grapple with this challenge, open universities and distance-education institutions will assume a leadership role in pedagogic design and delivery.

Key Issues

From the experience of the last four decades, it is clear that ODL will be an important element of future education and

training systems across the world irrespective of the stages of the economic, political, and social developments within countries. While in the educationally advanced and resource-rich countries, ODL provides a more convenient learner-driven model of education, in the developing countries, challenged by inadequate resources, ODL is a cost-effective option and can reach out to larger numbers without major constraints of time and place. The following issues, however, need to be addressed:

1. *Role of ODL in national systems.* There is a lack of clarity about the role of ODL in national systems. In several cases, it is perceived to be an alternative to the formal system leading to comparisons with respect to policies, processes, resource allocations, and performance outcomes. ODL suffers in such comparisons as the formal system usually dominates most national systems with ODL relegated to a secondary status. Is it a supplementary system? Yes, because the formal system does not have a real alternative. New colleges and universities will continue to be established throughout the world, and students will continue to enrol in them. The key challenge is that they will never be able to meet the rising demand. ODL supplements the efforts of the formal system.
2. *Graduation rates.* There are several instances of ODL systems that have taken off with great promise and potential. Some of them have indeed emerged as leading mega-universities measured by the size of the enrolments. However, there have also been instances of a disconnect between promise and performance. Wider access is not synonymous with success. Success is reflected in the outcomes measured in terms of the ratio of graduates to enrolment, the time taken to complete a program, the ratios of retention and drop-out, and the response of the market to the graduates. Graduation rates from some of the high-enrolment open universities are disproportionately low.
3. *Recognition of qualifications.* Recognition of qualifications is just as important as the performance levels. ODL degrees and certificates need to enjoy the same status and recognition for the purposes of employment and further studies as conventional systems both within national jurisdictions as well as globally. Provisions need to be made in the national policies for according the necessary equivalence to ODL qualifications and to ensure that students who have graduated from this system are treated at par with students from conventional institutions. Fortunately, sustained efforts are being made by international agencies to build a consensus among countries on the recognition of ODL qualifications. While research shows that there is no significant difference between distance education and traditional classroom instruction in relation to student outcomes, there is still a perception barrier to be

overcome, in order for this mode of education/training to be utilized to its fullest potential universally.

4. *Research.* What is the nature of research that open universities conduct? There are open universities that are engaged in disciplinary research with considerable success. The UKOU is rated high among British universities for excellence in research. But there are many open universities that struggle with a variety of challenges. Some of the more commonly reported impediments are inadequate technology infrastructures, lack of human capacity, deficiencies in planning and management and inadequacies of teaching and learning resources. It would be worthwhile for the ODL institutions systematically to record their experiences in all aspects of the implementation of their programs and activities to create a strong database that could support in-depth studies to draw useful lessons from. Such an effort could also help develop a body of best practices, identify strategic issues, analyze learner response and experiences, design effective support systems for different learning environments, and appropriate student-assessment procedures and practices. Research into these aspects would give ODL institutions a leadership position in improving the quality of mainstream tertiary institutions.
5. *Technology and the developing world.* The development of ICTs has been uneven across the globe. The increasingly symbiotic relationship between ODL and technology is likely to disadvantage developing countries in the short term. While technologically advanced countries will benefit from advances in technology and develop more cost-effective pedagogies, developing countries will continue to use older technologies. There will be a need to focus on the ways to close this digital and pedagogic divide.

Conclusion

Crisis often generates creativity and innovation. Innovations are not necessarily spectacular initiatives. They can be minor modifications or adaptations of known practices and methods to get over challenges and obstacles in specific situations. It has now become possible to transform verbal information into visual communication. It is possible to study synchronously or asynchronously. The consequences of this transformation are more far-reaching in education than in most other areas of human endeavor. They impact the traditional notion of the classroom, the teaching profession, the pedagogy, the design of the curricula, and the learning experience. Engaged as they are with every aspect of these emerging trends, the ODL systems can, through systematic and

systemic research, provide valuable insights into the unfolding universe of a new, but not fully defined, world-education order.

Bibliography

- Atkins, D., Brown, J. S., and Hammond, A. (2007). A review of the OER movement: Achievements, challenges and new opportunities. *Report to the William and Flora Hewlett Foundation*. CA: San Francisco.
- Daniel, J. (1996). *Mega-Universities and Knowledge Media: Technology Strategies for Higher Education*. London: Kogan Page.
- Daniel, J., Kanwar, A., and Uvalic-Trumbic, S. (2005). 'Who's afraid of cross-border higher education? A developing world perspective. *Higher Education Digest*, 52. London: Open University.
- Daniel, J., Kanwar, A., and Uvalic-Trumbic, S. (2006). A tectonic shift in global higher education. *Change* 38(4), 17–23. Washington, DC: Heldref Publications.
- Hendrikz, J., Viljoen, J. M., and Adams, L. (2006). *Text Messaging (SMS): Perceptions and Usage Patterns*. Pretoria: University of Pretoria.
- Holmberg, B. (1977). *Distance Education: A Survey and Bibliography*. New York: Nichols.
- Kapur, D. and Crowley, M. (2008). *Working Paper No 139. Beyond the ABC's: Higher Education and Developing Countries*. Centre for Global Development, University of Pennsylvania.
- Prensky, M. (2005). The future is now: Strategies for reaching today's students. *A Presentation Made at the WCET Conference at San Francisco, USA*.
- Rumble, G. and Koul, B. (2007). *Open Schooling for Secondary and Higher Secondary Education: Costs and Effectiveness in India and Namibia*. Vancouver: Commonwealth of Learning.

Further Reading

- Bates, A. W. (2005). *Technology, E-learning and Distance Education*. London: Routledge.
- Creed, C., Allsop, T., Mills, R., and Morpeth, R. (2005). The art of the possible: Issues of learner support in open and distance learning in low income countries. *Report Prepared by the International Research Foundation for Open Learning (IRFOL) for Commonwealth of Learning*, Vancouver.
- Daniel, J. (1996). *Mega-Universities and Knowledge Media: Technology Strategies for Higher Education*. London: Kogan Page.
- Hope, A. and Gupton, P. (eds.) (2006). *Strategies for Sustainable Open and Distance Learning*. London: RoutledgeFalmer/Commonwealth of Learning.
- Koul, B. and Kanwar, A. (eds.) (2006). *Towards a Culture of Quality, Perspectives in Distance Education*. Vancouver: COL.
- McIntosh, C. and Zeynep, V. (eds.) (2005). *Lifelong Learning and Distance Higher Education, Perspectives on Distance Education*. Vancouver: UNESCO and COL.
- Moore, M. G. and Anderson, W. G. (eds.) (2003). *Handbook of Distance Education*. Mahwah, NJ: Erlbaum.
- Moran, L. and Rumble, G. (eds.) (2004). *Vocational Education and Training through Open and Distance Learning*. London: RoutledgeFalmer/Commonwealth of Learning.
- Robinson, B. and Latchem, C. (eds.) (2003). *Teacher Education through Open and Distance Learning*. London: RoutledgeFalmer/Commonwealth of Learning.
- Rumble, G. (1997). *The Costs and Economics of Open and Distance Learning*. London: Kogan Page.
- Yates, C. and Bradley, J. (eds.) (2000). *Basic Education at a Distance*. London: RoutledgeFalmer/Commonwealth of Learning.

Education Development in Higher Education

Y Ryan, Australian Catholic University, Virginia, QLD, Australia

K Fraser, Australian Catholic University, Melbourne, Victoria, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This article discusses the discipline of professional development for the learning and teaching roles of staff in the higher education sector. Internationally, the terms used to describe this activity include academic development, learning and teaching development, education development (ED), and faculty development. The terminology is contested and somewhat divides along US and UK/Commonwealth lines with faculty development the preferred US term. In this article, we use the term ED to refer to “developmental activities informed by the discipline of teaching and learning in higher education. This discipline is underpinned by research into university teaching and learning” (Fraser, 2005: 5). We have chosen to use this term rather than the older term of academic development because it reflects the reality for those working in the area from both central development units and within faculties. While ED staff have the remit to work with academics to improve the quality of teaching, we are as likely to be working with staff from the library, computing technology, student services, and the disability and equity office, as with academic staff. We also prefer the term because it encompasses the work that we do to develop academic policy, implement legislative requirements, and carry out pedagogic research, as well as the work done to develop the quality of curricula, assessment, evaluation, and teaching delivery.

In this article, we discuss ED from the perspective of our combined experiences of working in the discipline in Australia, North America, and the United Kingdom.

The Context

ED takes place in what has been described by Barnett (2000) as the age of supercomplexity. The pace of change in higher education is extraordinarily rapid and futures are uncertain. Internationally, drivers for change in the sector are: the shrinking of government funding sources while accountability measures increase; the increase in student diversity and numbers; demands by Generation Y (Gen Y) students for more flexibility in their access to education and the speed of that access; globalization adding increasing complexity and competition; government agendas driving curricula increasingly toward

vocationalization; information and communication technologies (ICTs) which are finally ubiquitous within universities with their potential to store and create information and foster new forms of communication; and changing conceptions of the nature of knowledge itself (Gibbons *et al.*, 1994). All these factors have disrupted traditional conceptions of the academic role, away from content transmission to learning processes. The consequence is a necessary expansion in the knowledge base of academic (and professional) staff: they must be ‘double professionals’ – discipline and education specialists. ED’s concerns are with this second profession, spanning psychology, sociology, systems theory, communications theory, cross-cultural studies, curriculum and education theories, and epistemology as well as legislative requirements, for example, in relation to international students.

The seeds of ED were sown internationally in the 1960s and 1970s, coinciding with both the beginning of mass participation in higher education in the developed world and the student activism of that time. In North America “students demanded the right to exercise some control over the quality of their undergraduate learning experience, by such means as evaluating their teachers’ performance in the classroom. At the same time, economic recession and the concomitant decline in [academic] mobility and opportunities for renewal contributed to new interest in [education development] and instructional improvement” (Sorcinelli *et al.*, 2006: 2–3). Barbara Falk (University of Melbourne), credited with founding ED in Australia, recalls that “it was a time [in the later 1960s] when students were turning against their university. They had moved from turning their indignation on public issues to criticising the institution Actually the student movement was instrumental in focusing people’s attention on teaching and learning” (Lee *et al.*, 2008: 28). Universities were called to task “about the curriculum and quality of teaching” (Lee *et al.*, 2008: 10, 11). In the 1970s, dedicated ED positions and units emerged in several universities in the UK, North America, Europe, and Australasia, although it was not until the 1990s that significant numbers of ED units were established. By the 1970s and 1980s, higher education teaching and learning (T&L) societies and associations were founded, such as the UK Staff and Education Development Association (SEDA), the Canadian Society for Teaching and Learning in Higher Education, the American Professional and Organizational Development Network in Higher Education,

and the Higher Education Research and Development Society of Australia (HERDSA). Annual higher education T&L conferences and journals dedicated to higher education T&L were established.

Yet ED remained piecemeal across the higher education sector. Ling *et al.* (2009) report that the organization of ED “occurred idiosyncratically within individual institutions, resulting in a diverse, context-specific and volatile set of models and practices” (Ling *et al.*, 2009: 17). The 40-year history of ED allows us to add another observation: as with any emergent field of knowledge, individual and often inspirational university staff such as Barbara Falk in Australia, Ruth Beard in the UK, Chris Knapper in Canada, and William Bergquist and Steven Phillips in the USA shaped the form that ED took in specific institutions, adding to the idiosyncratic nature of ED approaches and models.

By the late 1990s however, the UK Dearing Report (NCIHE, 1997), the Australian West Report (1998), and the crisis in funding for postsecondary education in the USA ushered in a resurgence in government interest in the quality of teaching and learning in higher education. This interest coincided with the seminal work of Boyer and colleagues on the scholarship of teaching and paralleled the burgeoning of research in organizational learning and organizational change processes, in itself a consequence of another emergent discipline, that of management (Rossiter, 2007). With this came accountability and quality-assurance demands, generally tied to funding initiatives. As students contributed more to the costs of their education, government concern shifted strongly to the consumer of education, the student. Universities needed to demonstrate systematic improvement in support for learning and teaching, and at least minimal development programs for staff. Performance review schemes, derived from the literature in management, and setting achievement goals in teaching as well as research and service, further prescribed academic roles.

The new millennium witnessed the introduction of national quality-assurance agencies (UK Quality Assurance Agency, the Australian Universities Quality Agency); national teaching quality indicators such as the Course Experience Questionnaire (CEQ) in Australia, and the US National Survey of Student Engagement; the development of national standards for higher education teaching, such as the UK National Professional Standards Framework for Teaching and Supporting Learning in Higher Education; the development of national higher education T&L institutes such as the UK Higher Education Academy, the Australian Learning and Teaching Council and the New Zealand Ako Aotearoa; stronger teaching reward and recognition schemes, for example, the Canadian Society for Teaching and Learning in Higher Education 3M Fellowships, and the Ako Aotearoa Teaching Excellence Awards; and the introduction of contestable national T&L

funding schemes, for example, the Australian Learning and Teaching Council grants and the UK Teaching Quality Enhancement Fund (TQEF).

With international interest in higher education T&L growing, ED became more systematic at institutional and sector levels. In the next section, we explore ED models and approaches.

ED Models and Approaches

By 2000, the majority of universities in the UK, Australia, and North America had constituted a central ED unit, although models and approaches varied, both between institutions and between countries.

Models

Hicks (1999) analyzed education development units across Australia in terms of their location within the university (centrally or faculty-based), and their activities in terms of disciplinary-based or generic. His paper describes four models of academic development in late twentieth century Australian universities:

1. the central model, characterized by a strong central unit and usually with minimal local discipline activity accompanied by little activity and responsibility at the local level;
2. the dispersed model with more activity at the local level;
3. the mixed model, of central and discipline-specific activities “with separate resourcing, potential duplication and a significant lack of coordination” (Hicks, 1999: 17); and
4. the integrated model with systematic and established collaborative processes.

The predominant model was the central unit providing generic programs and activities. In the UK, Gosling (2006) reported that universities tended to use a hub-and-spoke model in which a central ED unit worked closely with ED staff distributed through the faculties. The arrangements varied significantly between universities in terms of relationships, secondment, workload allocations, and role definition.

Approaches to ED

ED can be described in terms of both models and approaches. The latter refers to the work of ED staff: their activities, interventions, relationships, and networks. Sorcinelli *et al.* (2006) describe US ED approaches

(referred to as faculty development) from 1950 to the turn of the century in terms of four ages: scholar; teacher; developer; and learner. According to Sorcinelli *et al.*:

1. "In the Age of the Scholar (1950s and early 1960s) the term *faculty development* referred primarily to practices for improving scholarly competence" (Sorcinelli *et al.*, 2006: 27).
2. In the age of the teacher (mid-1960s through 1970s), ED focused on teaching development in which an "... earlier interest in behavioristic research on college-level teaching was superseded by an interest in research and practice related to the development of teaching skills and competencies, as well as the design of teaching development and evaluation programs" (Sorcinelli *et al.*, 2006: 3).
3. "In the Age of the Developer (1980s), faculty development broadened to address curricular issues, faculty needs at different career stages, and collective as well as individual faculty growth" (Sorcinelli *et al.*, 2006: 3).
4. In the age of the learner (1990s) "student learning rather than teaching took center stage ... Student diversity ... also called for a greater range and variety in teaching and learning methods, skills and sensitivities" (Sorcinelli *et al.*, 2006: 3, 4).

The UK and Australian literature, summarized by the Ling *et al.* (2009), suggests two main approaches to ED from the 1960s to 2000: teacher-focused and learner-focused.

Teacher-focused

ED staff worked primarily with voluntary academics, individual staff with a particular interest in teaching, via university-wide workshops or one-to-one-consultations. As in North America from the mid-1960s through the 1970s, the focus of ED work during this period was primarily on the improvement of individual teaching skills, "... teaching in different teaching contexts (e.g., lectures, tutorials, laboratories and so on), and for assessing students' work" (Ling *et al.*, 2009: 23).

Learner-focused

During the 1980s and 1990s the seminal phenomenographic research of Marton *et al.* (1984) and Bigg's (1999) work on enhancing student learning led to a focus on student learning research:

A deeper understanding of this interplay of teacher, context, student and curriculum has led to the development of more integrated conceptual models of teaching and learning that now play a major role in shaping the efforts of academic developers (Ling *et al.*, 2009: 23).

ED in the Twenty-First Century

While teacher and learner-focused approaches still form a significant part of ED work, different approaches to ED have emerged since the turn of the century. Sorcinelli *et al.* (2006: 4) in US argue that "with the new millennium faculty development has ... entered ... the Age of the Network". They argue that with the expansion of academic roles and the increase in sessional staff which in turn puts more administrative and management pressure on fewer full-time staff, more collaboration between all staff, academic and professional, is required.

Ling *et al.* (2009) report that since the turn of the century two other ED approaches have evolved: organization-focused and sector focused.

Organization-Focused

The turn of the century saw significant government regulation of higher education teaching quality. Universities were required to develop university teaching and learning strategic and implementation plans and performance indicators. Publications such as Ramsden's (1998) *Learning to Lead in Higher Education* focus "on the quality of organisational leadership, organisational structures that develop quality, scholarship that includes the ideas of reflection, informed critique, evaluation and development ... He squarely places the responsibility for academic development onto academic leaders" (Reid, quoted in Ling *et al.*, 2009: 23).

Sector-Focused

With specific reference to the UK and Australia, the Ling *et al.* (2009) that the first decade of the millennium saw "sector-wide initiatives stemming from government investigations into higher education and established and sustained by hitherto unprecedented amounts of government funding" specifically focusing on learning and teaching (Ling *et al.*, 2009: 23). These initiatives include:

1. National quality-assurance agencies, which audit individual institutional processes to ensure higher education standards and quality, such as the UK Quality Assurance Agency, and the Australian Universities Quality Agency.
2. National T&L agencies, such as the UK Higher Education Academy (HEA), the Australian Learning and Teaching Council (ALTC), and the New Zealand Ako Aotearoa, have unprecedented funding for supporting higher education T&L.
3. National funding bodies. The Australian Learning and Teaching Performance Fund, while controversial, rewards institutions for learning performance outcomes

measured against national benchmarks. The UK Teaching Quality Enhancement Fund supports T&L development at the institution, sector, and individual level.

4. National Teaching and Learning Standards, for example, the UK National Professional Standards Framework for Teaching and Supporting Learning in Higher Education.
5. National peak representative bodies and professional associations. In Australia, the Council of Australian Directors of Academic Development (CADAD) and the Australasian Council of Open, Distance and Elearning (ACODE) operate at the sector level, hosting and commissioning projects of significance to the sector and working with other national organizations to promote and support higher education T&L. The UK Heads of Educational Development Group and the Canadian Educational Developers Caucus play a similar role to CADAD.

The Ling *et al.* conclude that:

Academic development units typically concurrently engage with all four approaches – teaching oriented, learning-oriented, organisationally-directed, and response to system-level initiatives – in complex and dynamic configurations and arrangements peculiar to their institution (Ling *et al.*, 2009: 25).

All these factors have converged to position teaching professional development for higher education staff as critical in the adaptation of universities to the new millennium.

In his review of the ALTC in Australia, Lee Dow (2008: 4) argues “there is little doubt that the learning and teaching function of Australian universities has gained greater prominence in recent years, there is more respect for good teaching among many academic staff on most campuses of most universities, and the measures of student satisfaction with their learning and teaching experiences appear to have increased overall”.

Despite national initiatives, and the call by some for the professionalization of higher education teaching (Dearn *et al.*, 2002), academics in 2009 are still typically employed on the basis of disciplinary research strengths. ED is carried out in an ambivalent environment: national policy and programs promote and value good teaching, yet the prestige of institutions is judged on research performance and productivity, and research grants gain greater recognition for academics at the institutional level. The next section considers the challenges for successful ED.

Challenges

Aging Demographic and Workforce Structures

The aging demographic of academic staff in Northern America, the UK, and Australia represents a significant

challenge to the sector in the coming decades (Mills *et al.*, 2006). Hugo's (2005) studies on the aging demographic of Australian academic staff, and belated responses to succession planning by management suggest two immediate challenges for ED: the reluctance of existing staff in later age brackets to engage fully with less transmission-based pedagogies and the potential of Web 2.0 technologies in teaching, and the imperative to expand ED for the influx of new staff expected within the next decade. A further challenge is the dramatic increase in numbers of sessional staff (Ling *et al.*, 2009; Sorcinelli *et al.*, 2006), who are often also practising professionals in nonteaching fields such as accounting, nursing, and management. Such staff are employed for several reasons: they permit a university to decrease labor costs; increase staffing flexibility when enrolments fluctuate; and, because they bring professional and real-life experience to their teaching, give relevance to students in their educational programs. However, they pose difficulties in terms of ED, since they often work outside usual academic contact hours, making workshop timing a major issue, and they may be less inclined to consider development necessary in what is essentially a poorly paid and part-time work role.

Vocationalization of Curricula

The issue of the relevance of curricula to engage students poses a further dilemma for ED: work-integrated learning (WIL) has emerged in the past decade as a means of improving the work-readiness of graduates, and as a corollary of the trend to more vocational programs in universities. In consequence, a larger share of teaching is done at the workplace, under the supervision of experienced professionals who do not often see mentoring or student supervision as central to their role. Universities have not responded quickly to the additional complication of this off-campus teaching, in terms of their quality control over workplace supervisors, or the ED of such ancillary staff.

New Technologies

Although new technologies have given affordances to teaching approaches unimaginable even 30 years ago, such as videoconferenced lectures, podcasting, and teaching blogs, all concerning the physical separation of teacher from student and student from class, a major contributor to the complexity of ED in the contemporary and future university is the very nature of new technologies, its capacity to fragment the academic role and dissipate the authority of the teacher. Many ED units include training in technologies as a major component of their work, infusing technical skills with pedagogy. This places ED staff in the awkward position of dancing with the devil of technology as Katz (1999) argues, since they are by fiat supporting

extended use of technologies which many staff see as subverting their role.

Students themselves appear ambivalent about the use of newer technologies in teaching. Prensky's (2001) arguments regarding digital natives (Gen Y students who have grown up with ubiquitous computing and mobile devices) versus digital immigrants (those who came late to computer devices) seem convincing: the natives demand that academics use these new tools with as much ease as they themselves do. Yet, more robust research (such as Krause *et al.*, 2005) reveals both that natives are less competent and confident in their use of ICTs for learning than Prensky argues, and that they often resent the use of their spaces (FaceBook and MySpace) for educational purposes.

Further, the very ease of cut-and-paste research for assignments has created yet another domain of activity for ED: educating staff about assessment that discourages/prevents plagiarism of this sort, and educating students about academic integrity in a surrounding culture which places no virtue on originality.

Theoretical Perspectives

One of the principal challenges in learning and teaching in higher education is a lack of consensus on its theoretical underpinnings. At all levels of education studies, multiple disciplinary inputs (from sociology, psychology, systems theory, history, and philosophy) are crucial to understanding and practice. In higher education, however, the influence of these disciplines in teaching practices has persistently clashed with discipline-centered arguments, around tribes and territories. Nor has education as a broad discipline been assisted by the evolution of, and overlap in, psychological theories pertinent to learning and teaching. Behaviorist approaches still dominate our classroom practice, as expressed through observable learning outcomes, at the same time as new discoveries in neuroscience inform our understanding of learning, and as constructivism urges a teaching approach that builds on students' existing knowledge, skills, and motivation. The edubabble jargon peculiar to education studies – and of postmodernist and critical theory in the social sciences – has also been unhelpful in professional development for learning and teaching. The lack of an authoritative coherent theoretical core for learning and teaching frustrates many discipline-strong academics, especially those in the hard sciences, when they engage in ED.

Accredited Higher Education Learning and Teaching Programs

One component of the efforts to focus more intently on learning and teaching was the establishment of formal award programs in higher education teaching generally

at graduate certificate level. The efficacy of such courses in improving student learning has been widely contested, most particularly by Knight *et al.* (2006), and as heatedly defended by ED staff (Dearn *et al.*, 2002). In practice, they attract few academic staff unless mandated, yet are essential to credibility in the development of the double professional.

Structural Instability

As the idea of the university has shifted over the past 20 years, and ED encompasses an ever-increasing repertoire of responsibilities, the models adopted by individual institutions have fluctuated between those outlined by Hicks (1999). One consequence of what appears to be regular restructuring of ED units has been high turnover among directors of ED units, with resultant instability in learning and teaching leadership. A more intense focus on learning and teaching systematic policies and processes has, paradoxically, made ED a tenuous profession, occupying an uncertain space between senior management and staff, developing and promulgating policy designed to standardize quality in learning and teaching, to a workforce historically resistant to uniformity. The disparate nature of ED within institutions, and this instability, has made it difficult to measure the value of ED units in enhancing learning and teaching performance, as discussed below.

ED Futures

The futures of ED are as varied as the ways in which ED operates and is configured in individual universities. In spite of that qualification, we believe that ED success over the next decade depends in large part on the achievement of the following: development and adoption of key ED performance indicators; the national and international collaboration of higher education staff on projects to address sector-wide ED issues and challenges as is the norm in other disciplines; the distribution of ED roles throughout institutions and the development of staff in these roles; and the professionalization of ED staff and leaders.

Key ED Performance Indicators

The work of ED staff is mediated through their colleagues who work directly with students, or through the development of policies and procedures designed to improve learning environments and student engagement with the university. Because of this mediation it has been difficult to directly measure the impact of ED work on student learning.

In spite of these difficulties, the demonstration of the impact of ED is of critical importance to the sector. Just as the demonstration of the quality of teaching was inevitable with greater accountability demands from government, the sector is increasingly demanding that the impact of the work of ED is measured. A project to develop ED key performance indicators was started in Australia in late 2008 and is expected to be completed in 2009. It is likely that international collaboration on the development of ED key performance indicators would be of value to the sector.

International and National Collaboration in ED

While there has been a long tradition of staff exchange between the UK and Australia in ED leaders (Ramsden, Trigwell, and Prosser), ensuring some commonality in sector emphases in promoting learning and teaching, and some cross-fertilization via associations such as the International Society for the Scholarship of Teaching and Learning (ISSOTL) and the International Consortium for Educational Development (ICED), globalizing trends will demand more collaboration at national and international levels.

National collaborations are well underway through the auspices of agencies such as ALTC and the HEA. Embryonic international relationships are also being established by the same agencies, but it will require robust effort to foster initiatives that will meet ED needs in transnational issues such as professional development for sessional staff, and training in information technology (IT) for teaching. The question of academic standards at a global level will also require international collaboration of an order of magnitude above the efforts of Bologna.

Distributed ED Roles

To effect the changes in teaching quality and standards and student access that twenty-first-century societies and governments demand, the work of ED needs to be the responsibility of many positions distributed across the university in collaboration with a central ED unit. This hub-and-spoke model allows for central (hub) expertise in all areas of ED to be accessed by (usually) discipline-focused staff (spokes) located in the faculties. Discipline-based ED roles include faculty deans/associate deans T&L, T&L champions or advocates, course coordinators, and externally funded teaching project staff.

While the distributed configuration of ED provides a useful approach to achieve institution-wide T&L improvement, the issue of a lack of ED experience and expertise undermines this approach, unless more systematic initiatives are undertaken to professionalize ED.

Professionalization of ED

The quality of ED is as good as the experience and expertise of the staff fulfilling the roles. While the sector increasingly relies on staff with ED skills and knowledge to lead disciplinary, interdisciplinary, institutional, and sector improvement of T&L, an urgent issue for universities is the struggle to find staff who are experienced and skilled in ED, teaching scholarship, and leadership. Whether in full- or part-time roles, many learn on the job (frequently as managers on short-term contracts) and yet the ED role is a highly specialized profession in its own right.

Currently, relatively few university teaching staff in distributed ED roles have a professional qualification in T&L or in T&L leadership. Most make the transition from being a discipline specialist to a role that includes ED on the basis of their interest and enthusiasm for learning and teaching. These staff have considerable experiential knowledge but many lack a theoretical understanding to underpin their practice. They are often expected to lead significant changes and innovations in their university/faculty/school but have little strategic planning experience and the expertise and management skills needed to effect change.

The processes by which staff in ED leadership roles learn to be strategic and to take a broad sectoral view of teaching and learning are not yet well documented. It is critical for the sector that this role is supported and developed through a scholarly and evidence-based approach to their professional development. Further research into pathways into the profession, the theoretical understanding needed for working successfully in ED roles, and effective models for ED professional development are needed.

Conclusion

The next decade will see yet more structural changes in the university sector in response to financial stringency, digital technologies, and government demands for more flexible pathways between technical and further education, and higher education. We expect that the role of ED will continue to be to work through all levels of the institution (individual, department/school/faculty, and across the institution) and potentially across the sector to improve T&L quality, learning environments, and student engagement with the university. We have argued that to do this work effectively staff with ED roles need to be both centrally based and distributed throughout the institution and that those staff need to engage in professional development specific to the profession of ED.

In addition to addressing these imperatives, we believe that T&L leadership at the deputy and pro-vice-chancellor levels is critical to the success of institutional ED.

Just as, accountability and role-specific professional development are needed for teaching and ED, the same is true for those in Deputy Vice Chancellor (Academic/Education/T&L) roles. Further research is needed to determine ways in which leadership at this level can be most effective.

Bibliography

- Barnett, R. (2000). *Realizing the University in an Age of Supercomplexity*. Buckingham: Open University Press.
- Biggs, J. (1999). *Teaching for Quality and Learning at University*. London: Open University Press.
- Dearn, J., Fraser, K., and Ryan, Y. (2002). *Investigation into the Provision of Professional Development for University Teaching in Australia*. Canberra: Department of Education, Science and Training. http://www.dest.gov.au/NR/ronlyres/D8BDFC55-1608-4845-B172-3C2B14E79435/935/uni_teaching.pdf (accessed August 2009).
- Fraser, K. (ed.) (2005). *Education Development and Leadership in Higher Education: Developing an Effective Institutional Strategy*. London: Taylor & Francis.
- Gibbons, M., Limoges, C., Nowotny, H., et al. (1994). *The New Production of Knowledge – The Dynamics of Science and Research in Contemporary Societies*. London: Sage.
- Gosling, D. (2006). Educational development in 2006. *Report of the Heads of Educational Development Group Survey of Educational Development Units in the UK 2006*.
- Hicks, O. (1999). Integration of central and departmental development – reflections from Australian universities. *International Journal for Academic Development* 4(1), 43–51.
- Hugo, G. (2005). Demographic trends in Australia's academic workforce. *Journal of Higher Education Policy and Management* 27(3), 327–343.
- Katz, R. (ed.) (1999). *Dancing with the Devil*. San Francisco, CA: Jossey-Bass.
- Knight, P., Tait, J., and Yorke, M. (2006). The professional learning of teachers in higher education. *Studies in Higher Education* 31(3), 319–339.
- Krause, K., Hartley, R., James, R., and McInnis, C. (2005). *The First Year Experience in Australian Universities: Findings from a Decade of National Studies*. Melbourne: Centre for the Study of Higher Education, University of Melbourne.
- Lee, A., Manathuga, C., and Kandlbinder, P. (2008). *Making a Place*. Milperra, NSW: HERDSA.
- Lee Dow, K. (2008). *An Evaluation of the Australian Learning and Teaching Council 2005–2008*. <http://www.altc.edu.au/system/files/documents/D08%201652%20%20An%20Evaluation%20of%20the%20ALTC%202005-2008%20%20Professor%20Kwong%20Lee%20Dow%20AM%20130808.pdf> (accessed 27 October 2009).
- Ling, P. (2009). Council of Australian Directors of Academic Development. *Development of Academics and Higher Education Futures, Vol.1, Report*. Sydney: Australian Learning and Teaching Council.
- Marton, F., Hounsell, D., and Entwistle, N. (1984). *The Experience of Learning*. Edinburgh: Scottish Academic Press.
- Mills, D., Jepson, A., Coxon, T., et al. (2006). Demographic review of the UK social Sciences. ESRC. http://www.esrc.ac.uk/ESRCInfoCentre/Images/Demographic_Review_tcm6-13872.pdf (accessed August 2009).
- Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon* 9(5), 1–2. www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20Digital%20Immigrants%20-%20Part1.pdf (accessed August 2009).
- Ramsden, P. (1998). *Learning to Lead in Higher Education*. London: Routledge.

- Rossiter, D. (2007). Whither e-learning? Conceptions of change and innovation in higher education. *Organisational Transformation and Social Change* 4(1), 93–107.
- Sorcinelli, M. D., Austin, A. E., Eddy, P. L., and Beach, A. L. (2006). *Creating the Future of Faculty Development: Learning from the Past, Understanding the Present*. Bolton, MA: Anker.

Further Reading

- Alstete, J. (2000). Post-tenure faculty development: Building a system for faculty improvement and appreciation. *ASHE-ERIC Higher Education Report*, 27(4). San Francisco, CA: Jossey-Bass.
- Boyer, E. (1990). *Scholarship Reconsidered: Priorities of the Professoriate*. Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching.
- Briggs, J. (1999). *Teaching for Quality and Learning at University*. London: Open University Press.
- Fraser, K. (1999). Australasian academic developers: Entry into the profession and our own professional development. *International Journal for Academic Development* 4(2), 89–101.
- Fraser, K. (2001). Australasian 'academic developers' conceptions of the profession. *International Journal for Academic Development* 6(1), 54–64.
- Ipsos MORI (2007). Student expectations study. <http://www.jisc.ac.uk/media/documents/publications/studentexpectations.pdf> (accessed 27 October 2009).
- Marginson, S. and Considine, M. (2000). *The Enterprise University: Power, Governance and Reinvention in Australia*. Cambridge: Cambridge University Press.
- Oblinger, D. (2004). Boomers, gen-exers and millennials: Understanding the new students. *EDUCAUSE Review* 38(4), 37–47.
- Pascarella, E. and Terenzini, P. (2005). *How College Affects Students: A Third Decade of Research*, Vol. 2. Indianapolis, IN: Jossey-Bass.
- Patrick, K. and Lines, R. (2005). Assuring and improving teaching quality. In Fraser, K. (ed.) *Education Development in Higher Education: Developing an Effective Institutional Strategy*, pp 30–49. London: Taylor & Francis.
- Prebble, T., Hargraves, H., Leach, L., et al. (2004). *Impact of Student Support Services and Academic Development Programmes on Student Outcomes in Undergraduate Tertiary Study: A Synthesis of the Research*. Wellington: Ministry of Education.
- Race, P. (1999). *2000 Tips for Lecturers*. London: Routledge.
- Radloff, A. (2005). Decentralised approaches to education development: Supporting quality teaching and learning from within a faculty. In Fraser, K. (ed.) *Education Development in Higher Education: Developing an Effective Institutional Strategy*, pp 72–75. London: Taylor & Francis.
- Ramsden, P. (1992). *Learning to Teach in Higher Education*. London: Routledge.
- Reid, A. (2002). Is there an 'ideal' approach to academic development? In Goody, A. and Ingram, D. (eds.) *Spheres of Influence: Ventures and Visions in Educational Development*. Perth: 4th World Conference of the International Consortium for Educational Development (ICED). <http://www.osds.uwa.edu.au/about/conferences/iced2002/publications/toc> (accessed August 2009).
- Rowley, D., Lujan, H., and Dolence, M. (1998). *Strategic Choices for the Academy: How Demand for Lifelong Learning will Re-Create Higher Education*. San Francisco, CA: Jossey-Bass.
- Ryan, Y. (2004). Taking it to the world: The US private sector model. In Moran, L. and Rumble, G. (eds.) *Vocational Education and Training through Open and Distance Learning, World Review of Distance Education and Open Learning*, vol. 5, pp 147–163. New York: RoutledgeFalmer.
- Ryan, Y., Dearn, J., and Fraser, K. (2003). Preparing academics for teaching in Australian universities. HERDSA Conference, June, Christchurch, New Zealand. <http://surveys.canterbury.ac.nz/herdsa03/pdfsref/Y1156.pdf> (accessed August 2009).

- Ryan, Y. and Fitzgerald, R. (2009). Exploring the role of social software in higher education. *Handbook of social software*, pp 165–180. Hershey, PA: IGI Global.
- Smith, B. (2005). The role of national UK organisations in enhancing the quality of teaching and learning. In Fraser, K. (ed.) *Education Development in Higher Education: Developing an Effective Institutional Strategy*, pp 16–29. London: Taylor & Francis.

Relevant Websites

- <http://www.abs.gov.au> – Australian Standard Classification of Education.
- <http://www.ipsos-mori.com> – Ipsos-MORI: National Student Survey.

Effective Use of Technology in Teaching and Learning in HE

D Laurillard, Institute of Education, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Blended learning – A mix of conventional and digital technologies for learning and teaching.

Conversational framework – Laurillard's representation of the minimum requirements that any learning environment should provide to support the complete learning process, embracing the learning experiences of instruction, construction, experimentation, discussion, collaboration, production, and articulation, as a unified way of challenging both conventional methods and learning technologies.

E-learning – The use of any digital technology in support of teaching and learning.

Learning Activity Management System (LAMS) – An online interactive tool for designing sequences of learning activities using standard individual and group learning tasks.

Learning object repositories – Websites offering access to peer-reviewed online teaching and learning materials, for mainly college-level and HE curricula.

Technology enhanced learning (TEL) – A more recent term than 'e-learning', which has the same meaning.

Virtual learning environment (VLE) – An institutional-level software application to be developed software application to support teaching transactions online, for example, curriculum plans over a period, access to content, scheduling, information on assignments, assignment-handling.

Effective Use of Technology in Teaching and Learning in HE (D3)

Introduction

Until the widespread availability of personal computing in the 1980s, the general technology used for teaching and learning in higher education (HE) had been confined to audio-visual formats such as overhead projectors, audio-cassettes, video, slides, and tape-slide systems. The central aim was to improve the delivery of lectures. Toward the end of this earlier period, educational technology as a discipline began to develop a wider scope to include the

study of teaching methods more generally (McKenzie *et al.*, 1970), and bring a critical methodology to ways of improving all forms of teaching and learning. This brought a greater emphasis on the importance of the evaluation of effectiveness, and a focus on how educational technology affected the learners' work and performance. The discipline was eclectic in its methodologies, borrowing from the natural sciences to set up control studies of the differential effects of a technology or method in comparison with traditional approaches, and from qualitative methods in the social sciences to carry out interviews and observations of the learners' experience of the technology.

In the mid-1980s the development of personal computers and institution-based digital communications networks gave individual faculty members their own access to computer power. Students could rarely afford personal access, but benefited from the IT suites set up in many university departments, libraries, or central resource areas. The surge in business and domestic use of personal computing toward the end of the twentieth century, fuelled by lower hardware costs, the arrival of the worldwide web, cheaper communications technologies, and consumer-oriented entertainment software such as games and music, meant that, increasingly, students had personal and portable access to computers and the Internet.

The interactive personal computer, offering multimedia and online access to communications and content, brought all forms of traditional educational technology and teaching methods together in digital form. The development, in the space of two decades or so, of the extraordinarily rich opportunities for the technological enhancement of teaching and learning was too rapid for educational research to keep pace with it. The R&D methodologies and pedagogic theories of traditional educational technology did not have time to mature to the robust and rigorous research discipline needed for such an explosion of innovation. The funding required for installation and maintenance of hardware, the development of software, and the need to ensure that it was a universal amenity, not just for an elite few, was high. This meant the funding for research and development, and for the training of faculty staff, was rarely commensurate with the expenditure on hardware.

The result was rapid but uneven development, across all the developed countries, of a wide range of uses of digital technology to support teaching and learning in HE, as educationists struggled to work out how best to exploit each new technological opportunity for the benefit

of learners. The discipline of educational technology began to focus on the potential advantages of each new digital technology that became available – multimedia, hypertext, artificial intelligence, the Internet, the web, mobile/cell phones, email, computer-based conferencing, simulations, animations, multiplayer games, virtual worlds, blogs, wikis, podcasts, user-generated design tools, social networking sites – exploring what uses might be made of them in the service of different aspects of teaching and learning. Research methodologies and pedagogic theories drew on ever wider antecedents, as the field developed and members of more discipline areas became engaged in the problem of how best to use digital technologies to improve teaching and learning in HE.

To provide some focus in this nascent, but rapidly growing field, an analysis of the effective use of technology for teaching and learning can be structured around its role in the curriculum, in learning, and in teaching, complementary to the other articles in this section, but providing also a contrasting perspective on teaching and learning in HE.

The Effective Use of Technology in the Curriculum

The academic curriculum in many universities is determined in part by the faculty staff, with courses derived from their areas of expertise. As research responds to innovation and change, so courses follow. However, faculty also respond to a variety of other influences on the curriculum.

There has always been a close link between higher education and the requirements of professional work, and this tends to influence the inclusion of high-level transferable skills, such as problem-solving, information analysis, learning to learn, and communication in general, rather than specific content. Where students personally pay a high proportion of the costs of teaching, HE behaves more like a consumer-oriented business, and works to attract students by offering high-demand courses, especially in business and IT, thereby creating a further pressure on the curriculum. The role of government varies in different countries but, wherever it does play a role, universities are urged to respond to the requirements of the knowledge society and provide graduates with the IT skills they will need in all walks of life, no matter what their discipline qualification. The expectations of graduate skills are therefore strongly affected by the increase in technology in research, in culture, and at work.

Technology also affects the content of what is learned. It has such a powerful effect on how knowledge is being developed, negotiated, represented, and disseminated, that it will continue to affect the curriculum, whether it is academically, vocationally, or personally motivated. The online availability of data sets in science and social science enables students to analyze more authentic and complete

data; software used as tools of the trade in research and business, such as data analysis, computer-aided design, text analysis, and data mining, come into the curriculum of courses whose graduates need to compete in the modern workplace.

Global communications technologies have now begun to move the curriculum beyond the institution toward a potential global curriculum. In 2002 MIT made the first web-based offering of their curriculum with the Open Courseware Initiative, by making lecture notes available to all (Lynch, 2003). While this does not make the MIT learning experience accessible to all, as this is defined primarily by the on-campus experience, it does make public the detail of the curriculum being learned by their students. The material is available to other universities or individuals to use as they wish, but its mere availability helps to define what the curriculum is in a particular field. The UK Open University launched a similar initiative, OpenLearn, in 2006, with two significant developments from the original idea: (1) the courseware made available was designed for distance learning students, not as lecture materials and (2) an additional lab space provided the means for the material to be negotiated and adapted, making the curriculum organic, rather than static (Lane, 2008). In both cases, an important use of open teaching resources is likely to be through teachers using them to enhance their own courses and teaching. Students will use them for individual study, and many will no doubt benefit in a way that would not otherwise have been possible, but mere access to the material does not yield a university education. An open teaching resource is the modern equivalent of the public library. The effectiveness of the online learning experience is highly dependent on the interactive nature of the material and the personal support also provided (Littlejohn and Pegler, 2007).

There has been some speculation that the internet would make it possible for students to study at many different universities, simply by accessing their materials online, a truly market approach. However, the learning experience provided by a university is not simply the transmission of information given – the relationships between teachers and students are critical, as is the coherence of the curriculum. This would be lost in a supermarket model of higher education. Students are not able to act like consumers, precisely because they are learners, putting their trust in the quality of the institution and its staff.

In the early years of the twenty-first century, there have already been several examples of universities making curriculum materials available through the Internet. The trend will probably continue and, as teachers increase their use of other teachers' materials, the process will tend to develop a common global curriculum, with much greater commonality between the courses on offer. However, as the textbook market demonstrates, an increasing volume of material tends to create pressure for a more clearly differentiated curriculum as universities compete more

aggressively for students in a global market for higher education. An online curriculum, providing economies of scale across a widespread student audience, can enable specialist subjects to survive despite relatively low local demand. At a time when many universities were closing classical language courses, the UK Open University was able to aggregate the interest across the country to build a viable distance learning course. Global online access to higher education should enable scholars and teachers to build a highly differentiated online curriculum.

Technology affects what students need to learn, independently of the academy, because its impact on society, work, and research affects what must be studied in all disciplines, defining the skills and knowledge graduates need. It also provides access to extensive digital libraries that constitute the information and ideas that are available to be explored and studied. Technology also enables teachers to exchange and use each others' materials, and this in itself will affect the development of the curriculum. The effective use of technology in any context, however, is primarily dependent on the reflective practice of the academy, and its willingness and ability to innovate.

The Effective Use of Technology in Learning

Learning through instruction

Digital technologies emulate a very wide range of traditional educational technologies and delivery channels, all of them relevant to higher education because they concern the engagement with and manipulation of information, and the social processing of information as knowledge and practice. The personal computer enables the user to engage with text, photography, sound, film, video, and computational models of real-world systems. Together with the Internet it also provides the digital equivalent of the printing press, postal services, public libraries, the telephone, radio, and television. Software tools and environments provide the digital equivalent of writing, debating, literature searching, data analyzing, system modeling, film editing, painting, and designing – there is a long list, and tools relevant to university study continue to be further developed.

Throughout the history of higher education, all the traditional forms of information and communications media and technologies have been used to improve the experience of student learning – blackboards, books, slides, laboratories, libraries – as supplements to the delivery of the subject by the teacher, and the interpersonal dialogs between teachers and students. Digital technologies are capable of emulating all these, but add further value by bringing them together in a personal machine, with personal control over many aspects of their use and organization. Since the advent of wide access to personal computing, the higher education community has looked for ways to exploit the opportunities offered by digital

technologies. The expectation is that they should be capable of enhancing student learning to a far greater extent than the traditional forms.

The early work in the United States on instructional design, analyzing the conditions of learning (Gagné, 1970, 1997), sought to improve on traditional teaching by building interactive programs that could be used by individual students, going at their own pace, rehearsing them in concepts, examples, and illustrative exercises, using the limited static graphics of the time (Merrill, 1994). The early learning theorists had emphasized the importance of knowledge of results, or feedback as an essential component of the learning process, and this has always been embedded within teaching and learning of any kind, as fundamental to pedagogy. The interactive computer offered exactly this capability. However, the type of feedback on offer was usually limited to multiple-choice questions, easily handled by a computer, and commonplace as an assessment method. This essentially instructional use of learning technologies has been the most prevalent, especially in professional training, and in foundational courses in higher education. Didactic forms of teaching continue to dominate e-learning, through many of the open teaching resources, digital libraries, and learning object repositories, together with the continuing widespread use of multiple choice questions as the main form of interactive feedback. An alternative instructional approach used computer modeling to represent the behavior of a human tutor, as a branch of artificial intelligence, and built intelligent tutoring systems (Brown and Burton, 1978). This approach fell out of favor as it could not compete with the explosion of web-based resources, which were easier to develop and closer to the experience of most academics, being similar to writing texts.

Learning through construction

Around the same time a more experiential approach to the use of technology exploited the capability of computers to model systems, and used interactive simulations, where learners could input parameters to a computer model and see the effect on graphical or pictorial output (McKenzie *et al.*, 1978). Intelligent tutoring systems and simulations were both pursued as strands of development within the field, and were represented in the national R&D programs set up to develop the field of technology enhanced learning, although they never dominated.

An importantly different approach within this genre, which gained wide acceptance from its early inception, was the idea of constructionism within a digital environment. Pioneered by Seymour Papert at MIT, and influenced by Piagetian psychology, it embodied the theory that we learn complex concepts and ideas best by constructing representations that use them (Papert, 1980; Papert and Harel, 1991). The best-known example is turtle geometry, where learners have to guide a digital turtle using turn, move, and repeat instructions, and angle and

distance parameters. The trace of the turtle's journey illustrates what it takes to draw a triangle, and other more complex geometric forms. The student learns about angles and the geometrical properties of triangles through giving the turtle the sequence of instructions that construct the triangle. An important part of the theory is the production of a public representation of the learner's idea that can be shared with other learners. The idea was applied to other areas of higher education, such as physics, but it had most impact at school level, in many different countries (Hoyle and Noss, 2003). The fundamental concept of learning through construction is applicable across a wide range of discipline areas, at all levels of learning. Thirty years on, the concept has been implemented as NetLogo, a modeling tool that allows learners to set up and investigate models of the behavior of systems such as population growth, electrical circuits, and climate change, wherever a computational model is possible (Gilbert and Troitzsch, 2005).

Learning through discussion

Before the arrival of communications technologies, individualized instruction – making the learner active, and providing individualized feedback – was seen as the principal value of digital technology. During the 1990s, the Internet and the Web spread quickly within higher education to provide a different kind of value – offering both wide access to digital assets (text, audio, images, animations, simulations, videos), and communicative access to people (synchronous and asynchronous, one-to-one and one-to-many, text, audio, and video). These new capabilities offered the possibility of digital emulation of a much wider range of learning activities – information searching and handling, presentation of ideas, dialog, discussion, and debate. The clear beneficiaries were the distance learning universities, now able to emulate academic discussion at a distance (Harasim *et al.*, 1995; Mason, 1994), and the first major software applications for higher education, such as the First Class conferencing system, were developed. Campus universities were also quick to adopt, as it enabled them to reach part-time and work-based students. Networking was quickly seen as one of the most important applications of the technology (Castells, 2001). As it relied on the construction of knowledge through discussion among the participants, this use of the technology was most obviously aligned with the idea of the social construction of knowledge (Scardamalia and Bereiter, 1994).

Digital communications environments, developed to support group and community discussion in business and leisure, could emulate aspects of face-to-face communication quite successfully. However, the structure of the communication made an important difference to the success of the communication for teaching purposes. The form of discussion in a university refectory is different from the form of discussion in a tutorial – both can be

highly educational, but they have different kinds of learning outcomes. Educators had to rediscover the rules of good pedagogy in these new kinds of communication environments. One clear strand of development focused on the online version of the tutorial, and developed guidelines on how to prepare and moderate asynchronous group discussion (Salmon, 2003). Another focused on the online version of the seminar, where learners collaborated to develop knowledge, and the online environment is structured to encourage the sharing, critiquing, and negotiation of ideas (Crook, 1996; Koschmann *et al.*, 2002; Scardamalia and Bereiter, 2006). This focus on the social aspects of learning led researchers to return to the work of Vygotsky, and the socio-cultural theory of learning as a form of social action (Vygotsky, 1978).

Learning through collaboration

In more recent years, with wider access to broadband in the home and the workplace, users began to want to do things with the technology themselves, not just receive information from others. This led to a much greater emphasis on the development of user tools, enabling individuals to create their own web pages, computer-aided designs, and virtual environments, providing the means for construction as a learning method, to be more widely available. With all the digital capabilities now available, it is possible to emulate practically any traditional teaching method – the seminar, the guided field-trip, the experimental laboratory, the drama studio, the role-play simulation – and developers of e-learning can potentially call on a much wider range of pedagogic theory. The emergence of what is often called social networking enables a community to build, share, and re-use each others' materials – in the leisure context: text, music, photographs, and videos. This means that digital environments now enable learners to collaborate in the building and sharing of representations of their ideas in a variety of media.

Commercial online games have taken the lead in developing highly interactive virtual reality environments. The capability of the technology now goes well beyond the capacity of the educational community to keep pace with exploiting the software developed for other communities. Virtual reality, serious games, and dynamic modeling environments are being explored by commercial companies for educational applications, but the market tends to be sought at school level, rather than from the more specialist communities in higher education. It is unlikely that universities will be able to exploit the full potential of digital technologies for learning, until the development of such games becomes affordable outside the leisure market. University students will certainly continue to use them for their own purposes, and possibly, sometimes, for self-directed learning.

E-learning Pedagogies

Much of the recent research and development literature on e-learning has appealed to pedagogic theory to support the argument for a particular technology application (Alexander *et al.*, 2006; Conole and Oliver, 2006; Garrison and Anderson, 2003; Jonassen and Land, 2000). The field of student learning has in recent decades developed an analysis of what it takes to learn that focuses on learning as an activity by the student, rather than learning as an effect induced by the act of teaching. Since the initial link between behaviorism and programmed learning, and the early development of instructional design, the key theoretical developments in student learning have continued in the Dewey tradition of experiential learning (Dewey, 1938), emphasizing the importance of the learner being active, learning through making, experimenting, exploring, and the social aspects of learning through structured forms of discussing, articulating, and reflecting (Entwistle and Tomlinson, 2007). Every learning technology application can be linked to a pedagogic theory that demands what it provides. **Table 1** provides some examples.

Technology has something to offer every pedagogic theory, from instructional design to constructionism, to situated learning, to social constructivism. The relationship is essentially a post hoc rationalization, however. Technologies are developed for markets other than education – defense, manufacturing, commerce, leisure – and education finds uses for them. Therefore, this technology-led approach to the enhancement of learning is inevitable and productive, as innovations and new opportunities inspire teachers to invent new ways of using them in their teaching. However, alongside this enthusiasm for exploiting new opportunities there has also been a concern to lead innovation from an understanding of what it takes to learn, bringing a critical approach to how technology should be used, and a concern with ensuring that it be effective (Laurillard, 2002). The conversational framework built on research on student learning, and

borrowed the ideas of Conversation Theory (Pask, 1976), to represent the minimum requirements that any learning environment should provide to support the complete learning process. It was designed to embrace the learning experiences of instruction, construction, experimentation, discussion, collaboration, production, and articulation, as a unified way of challenging both conventional methods and learning technologies. Blended learning is the useful term used to describe the judicious mix of the two (Garrison and Vaughan, 2008).

The effectiveness of these different uses of technology for the quality of the learning experience has usually been evaluated with respect to the particular application in a particular learning context, such as a course or subject topic. Use of control groups is not very common, partly because when used, they continued in the educational technology tradition of demonstrating no significant difference, and partly because the value of the technology was often precisely in its innovation, which had no counterpart in conventional methods (Ehrmann, 2002). Evaluation of effectiveness typically takes the form of demonstrating through learner feedback that the application is seen to be enjoyable and beneficial, but the most compelling evidence of effectiveness is found in those studies that use qualitative data, such as comparisons of learners' written output, or quotes from learners about their experience of the learning environment afforded by the technology (Conole and Oliver, 2006).

Use of Technology for Flexible Study

There are two distinct types of role that technology can play in enhancing higher education from the students' point of view. One is its role in supporting the learning experience, as discussed above. The other is its role in extending the ways in which learning is undertaken, which could be referred to as the study experience, that is, enabling flexible study.

Table 1 The relationships between pedagogic theory and technology applications

<i>Pedagogic theory</i>	<i>Learning technologies</i>	<i>Examples</i>
Instructional design	Interactive programs able to select learning material according to performance, graphical displays, automated checking of multiple choice questions	Interactive teaching systems (Merrill, 1994; O'Shea and Self, 1983)
Constructionism, Experiential learning, Situated learning	Simulations, system models, and virtual environments offer an emulation of investigative or experimental learning experiences	Simulations and virtual worlds (Gilbert and Troitzsch, 2005; Harper <i>et al.</i> , 2000; Luckin and du Boulay, 1999)
Social constructivism, activity theory – that learning is a product of social interactions and the use of tools	Social software, such as conferencing systems, wikis, where contributors build a consensual representation of knowledge, bringing students into the activity of negotiating knowledge	Computer-Supported Intentional Learning Environment, CSILE (Engeström, 1999; Scardamalia and Bereiter, 1994) and HE applications of activity theory (Jonassen and Rohrer-Murphy, 1999; Scanlon and Issroff, 2005)

The traditional modes of study in higher education are campus-based, face-to-face, time-limited, teacher-paced classes. These constraints, which effectively excluded part-time or working students, were overcome by the distance teaching universities to provide home-based, correspondence, asynchronous, self-paced study (Bates, 2005; Daniel, 1996). The distance teaching universities could not easily match the library and laboratory facilities of the campus, so this provision was either limited to what could be delivered at a distance, or covered by occasional short study periods on a campus. As innovations in digital technology offered ways of emulating all these aspects of study – synchronous and asynchronous communication, access to digital libraries, and virtual laboratory environments – both distance and campus universities took advantage of them. All universities could now offer flexible forms of study to attract part-time, home-based, and work-based learners. This was a valuable development at a time when higher education was being pressed to find ways of widening participation (Bates, 2005; Castells, 2001; Collis and Moonen, 2001).

With the much wider availability of computing technology in all universities by the beginning of the twenty-first century, and the realization that it could emulate many aspects of the transactions between teachers and learners, the next major software application to be developed for HE was the virtual learning environment (VLE). Modeled on the traditional description of learning and teaching transactions – teacher plans curriculum over a period, delivers content, arranges tutorials, specifies assignments, returns assignments – the approach of the VLE was teacher-oriented, rather than being based on the requirements of learners, and was valued mostly for its administrative features, enabling faculty staff to post assignments and course documents (Dutton *et al.*, 2004).

The effectiveness of information and communications technologies in support of students' study experience is highly dependent on the extent to which the organizational context has adjusted to the presence of technology. It should not simply be an add-on, but properly embedded in the planning and conduct of courses, and the information, advice, and guidance to students. All the studies mentioned in this section above stress the importance of organizational structures. In particular, Bates proposes the SECTIONS framework (Bates and Poole, 2003) involving the following criteria by which institutions can test their implementation of e-learning:

Students – who are they, what do they need, and to what extent do they have technology skills?

Ease of use – how accessible is the technology for all groups of learners?

Costs – what is the cost structure (fixed and variable costs) of each technology, and the unit cost per learner?

Teaching and learning – what kinds of learning and instruction are needed? What are the best technologies for supporting T&L?

Interactivity – what kind of interaction does the technology enable?

Organizational issues – what are the organizational requirements, and barriers? What changes in the organization need to be made?

Novelty – can the use of the technology be sustained over the long term? Will initial capital costs be followed through with recurrent maintenance costs?

Speed – how quickly can courses be mounted and materials changed?

In this and other books on implementation, effectiveness of the technology is not evaluated simply as a general property of the technology, which would make no more sense than it would to test the effectiveness of books or lectures in general. Empirical studies are used instead to clarify the contextual conditions and design features that ensure the effectiveness of particular technologies, for particular educational ends.

Effective Use of Technologies for Teaching

The teacher's perspective on the use of technology derives to a great extent from their approach to teaching. If teaching is understood as a way of transmitting their own knowledge to the mind of the learner, to caricature one extreme, then faculty staff will use digital technology to deliver resources, much as they would deliver lectures and books. These teachers are well served by the VLE, described above. Unfortunately, although it has been widely implemented as a solution to the problem of making e-learning available, the VLE does not typically challenge higher education staff to develop new approaches to teaching and learning (Zemsky and Massy, 2004).

On the other hand, to the extent that teaching is seen as a way of facilitating the development of students' high-level skills through the vehicle of a particular field of study, then lecturers want to take advantage of the interactive, communicative, and collaborative capabilities of the technology. However, innovative teachers, even those well supported by a university's learning technology unit, find little time in a busy schedule to learn how to innovate with e-learning. It is least time-consuming to make use of communication environments, where students manage their own peer learning, and digital libraries, where students manage their own investigation and discovery learning. These teachers need learning design tools that make it easy for them to build a learning environment in which their learners can be active in pursuit of their study and learning. Few such tools are currently available. One example that is attracting attention in several countries is the Learning Activity Management

System (LAMS), an online tool for designing sequences of learning activities using standard individual and group learning tasks (McAndrew *et al.*, 2006).

The use of technologies for teaching will be more effective when teachers can more easily engage in designing their application to learning. The tools of teaching they are familiar with are blackboards/whiteboards, slide presentations, lecture notes, books, moveable desks, field trips, laboratories, and libraries. Here they have a reasonable degree of control over what and how their students learn, sometimes supported by technical staff. Designing their teaching with digital tools and resources requires either technical support of the kind that takes away their control, or a steep learning curve. One current trend may be an indication of how this could change. The development of repositories of learning objects is providing faculty staff with the means to share and build on each others' teaching materials. One example, the Multimedia Educational Resource for Learning and Online Teaching (MERLOT) provides peer-reviewed online teaching and learning materials, for most of the HE curriculum, through a website available internationally (MERLOT). While the majority of early submissions were primarily text-based materials, the number of more interactive examples is growing. This kind of development alongside the open education movement, dominated by MIT in the United States and the Open University in the United Kingdom, brings to the academic teaching community the means to become true scholars of teaching, able to critique, improve, and share teaching ideas. Their best pedagogy is no longer an evanescent abstraction, but is captured for the benefit of others in digital form.

This supply of teaching materials is not yet matched by demand. Faculty staff prefer to build their own, without reference to others' ideas, in stark contrast to their behavior in their research fields. However, if the learning object repository trend is matched by the further development of the learning design tools trend, then there may yet be a flourishing exchange of ideas and innovations in the practice of e-learning by the academic community.

Studies of online courses have used surveys to try to discover the extent to which e-learning achieves cost-effectiveness, but the results are largely inconclusive, due to the immaturity of most implementations, and the complexity of the analysis (Bates, 2005; Boyer, 1997; Garrett and MacLean, 2004; OBHE, 2003; Twigg, 2003). The best evidence comes from institutional studies where the organization has set out to achieve cost effectiveness through the innovation, and has managed and monitored costs of all aspects of the implementation (Bates, 2005; Twigg, 2003).

Concluding Points

Developments in digital technologies over the last few decades have repeatedly offered opportunities to emulate

traditional forms of teaching and learning and, by integrating them, and operating at scale, have extended their potential to contribute to all aspects of higher education. However, the changes have been so rapid that scarcely has the potential of one technology been understood than another offers even greater promise. The main developments in e-learning have therefore tended to be technology driven.

Educators and scholars have found ways of using digital technologies to facilitate

- learning through instruction;
- learning through construction;
- learning through discussion; and
- learning through collaboration.

However, typically, the pedagogic theory is brought in to fit what the technology offers. It may then direct the design of the application, but it never drives technology development.

This will continue, because technology attracts massive investment, and responds to the powerful drivers of the business and leisure markets, whereas pedagogy develops slowly, unnourished by investment. The field of e-learning, whether research, development, or practice, is immature and fragmented. It needs a clear and robust pedagogic theory, capable of withstanding the allure of technology developments that do not serve learner understanding, and of challenging the technology to develop what learners actually need.

There is some promise in the trends of development of tools and resources by higher education, for higher education, that now go beyond the administrative tasks to generate learning object repositories and learning design tools. These new capabilities could generate a genuine scholarship of teaching, emulating the rigor of scholarship in research.

Technology development will continue, and students in HE will continue to use it extensively in all parts of their lives. Some of this use is likely to extend to their learning, but in a self-directed way. Faculty staff will use technology extensively in their research, and to the extent that research and teaching are synergistic, this will benefit their teaching.

However, innovation in e-learning, driven by pedagogic imperatives, and focused on improved learner understanding, is not likely to be a major force in HE until it is linked to the broader strategic aims for the quality and reach of teaching and learning, at the highest levels in universities.

See also: Distance Education and Open Universities; Faculty Values and Expectations and Their Implications for Teaching, Learning and the Organization of Higher Education Institutions; Flexible Learning in Higher Education; Innovation in Teaching and Curriculum Design.

Bibliography

- Alexander, S., Harper, C., Anderson, T., et al. (2006). Towards a mapping of the field of e-learning. *Paper Presented at the World Conference on Educational Multimedia, Hypermedia and Telecommunications*. Chesapeake, VA: AAC.
- Bates, A. W. T. (2005). *Technology, E-Learning and Distance Education*, 2nd edn. London: Routledge.
- Bates, A. and Poole, G. (2003). *Effective Teaching with Technology in Higher Education*. San Francisco, CA: Jossey-Bass.
- Boyer, E. L. (1997). *Scholarship Reconsidered: Priorities of the Professoriate*. Jossey-Bass: San Francisco, CA.
- Brown, J. S. and Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science* **2**, 155–192.
- Castells, M. (2001). *The Internet Galaxy: Reflections on the Internet, Business and Society*. New York: Oxford University Press.
- Collis, B. and Moonen, J. (2001). *Flexible Learning in a Digital World: Experiences and Expectations*. London: Routledge.
- Conole, G. and Oliver, M. (2006). *Contemporary Perspectives in E-Learning Research*. London: Routledge.
- Crook, C. (1996). *Computers and the Collaborative Experience of Learning*. London: Routledge.
- Daniel, J. S. (1996). *Mega-Universities and Knowledge Media: Technology Strategies for Higher Education*. London: Kogan Page.
- Dewey, J. (1938). *Experience and Education*. New York: Kappa Delta Pi.
- Dutton, W., Cheong, P. H., and Park, A. (2004). An ecology of constraints on e-learning in higher education: The case of a virtual learning environment. *Prometheus* **22**(2), 131–149.
- Ehrmann, S. C. (2002). Evaluating (and improving) benefits of educational uses of technology. *Paper*, WCETo.
- Entwistle, N. and Tomlinson, P. (eds.) (2007). *Student Learning and University Teaching*. Leicester: The British Psychological Society.
- Gagné, R. M. (1970). *The Conditions of Learning*. New York: Holt, Rinehart and Winston.
- Gagné, R. M. (1997). *The Conditions of Learning*, 3rd edn. New York: Holt, Rinehart and Winston.
- Garrett, R. and MacLean, D. (2004). Pricing online learning: Practice, rationale and transparency. *Briefing Note*, Observatory on Borderless Higher Education.
- Garrison, D. R. and Anderson, T. (2003). *E-Learning in the 21st Century*. London: RoutledgeFalmer.
- Garrison, D. R. and Vaughan, N. D. (2008). *Blended Learning in Higher Education: Framework, Principles, and Guidelines*. Wiley: San Francisco, CA.
- Gilbert, N. and Troitzsch, K. G. (2005). *Simulation for the Social Scientist*. McGraw-Hill Education: Maidenhead.
- Harasim, L., Hiltz, S. R., Teles, L., and Turoff, M. (1995). *Learning Networks: A Field Guide to Teaching and Learning Online*. Cambridge, MA: MIT Press.
- Hoyles, C. and Noss, R. (2003). What can digital technologies take from and bring to research in mathematics education? In Bishop, A. J., Clements, M. A., Keitel, C., Kilpatrick, J., and Leung, F. K. S. (eds.) *Second International Handbook of Research in Mathematics Education*, pp 323–349. Dordrecht: Kluwer.
- Jonassen, D. H. and Land, S. M. (2000). *Theoretical Foundations of Learning Environments*. Erlbaum: Mahwah, NJ.
- Koschmann, T., Hall, R. P., and Miyake, N. (eds.) (2002). *CSCL 2: Carrying Forward the Conversation*. Erlbaum: Mahwah, NJ.
- Lane, A. (2008). Widening participation in education through open educational resources. In Iiyoshi, T. K. and Vijay, M. S. (eds.) *Opening up Education: The Collective Advancement of Education through Open Technology, Open Content, and Open Knowledge*, pp 149–164. Cambridge, MA: MIT Press.
- Laurillard, D. (2002). *Rethinking University Teaching: A Conversational Framework for the Effective Use of Learning Technologies*, 2nd edn. London: RoutledgeFalmer.
- Littlejohn, A. and Pegler, C. (2007). *Preparing for Blended Learning*. London: Routledge.
- Lynch, C. A. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *Libraries and the Academy* **3**(2), 327–336.
- Mason, R. (1994). *Using Communications Media in Open and Flexible Learning*. London: Routledge.
- McAndrew, P., Goodyear, P., and Dalziel, J. (2006). Patterns, designs and activities: Unifying descriptions of learning structures. *International Journal of Learning Technology* **2**(2/3), 216–242.
- McKenzie, J., Elton, L. R. B., and Lewis, R. (1978). *Interactive Computer Graphics in Science Teaching*. New York: Ellis Horwood.
- McKenzie, N., Eraut, M., and Jones, H. C. (1970). *Teaching and Learning: An Introduction to New Methods and Resources in Higher Education*. Paris: UNESCO.
- Merrill, M. D. (1994). *Instructional Design Theory*. Englewood Cliffs, NJ: Educational Technology Publication.
- OBHE (2003). Redesigning teaching and learning in higher education using ICT: Balancing quality, access and cost. A Report to the Department for Education and Skills. London: The Centre for Policy and Change in Tertiary Education/University of Surrey/The Observatory on Borderless Higher Education.
- Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. Brighton: Harvester Press.
- Papert, S. and Harel, I. (1991). Situating constructionism. In Harel, I. and Papert, S. (eds.) *Constructionism: Research Reports and Essays, 1985–1990*, pp 1–11. Ablex: Norwood, NJ.
- Pask, G. (1976). Conversational techniques in the study and practice of education. *British Journal of Educational Psychology* **46**(1), 12–25.
- Salmon, G. (2003). *E-Moderating: The Key to Teaching and Learning Online*, 2nd edn. London: Taylor and Francis.
- Scardamalia, M. and Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences* **3**(3), 265–283.
- Scardamalia, M. and Bereiter, C. (2006). Knowledge building: Theory, pedagogy and technology. In Sawyer, K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 97–118. Cambridge, UK: Cambridge University Press.
- Twigg, C. A. (2003). Improving learning and reducing costs: New models for online learning. *Educause* **38**(September/October), 28–38.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Zemsky, R. and Massy, W. F. (2004). Why the e-learning boom went bust. *Chronicle of Higher Education* **50**(44), B6.

Further Reading

- Engeström, Y. (1999). Activity theory and individual and social transformation. In Engeström, Y., Miettinen, R., and Punamaki, R. (eds.) *Perspectives on Activity Theory*, pp 19–38. New York: Cambridge University Press.
- Harper, B., Hedberg, J. G., and Wright, R. (2000). Who benefits from virtuality? *Computers and Education* **34**(3/4), 163–176.
- Jonassen, D. H. and Rohrer-Murphy, L. (1999). Activity theory as a framework for designing constructivist learning environments. *Educational Technology Research and Development* **47**(1), 61–79.
- Luckin, R. and du Boulay, B. (1999). Ecolab: The development and evaluation of a Vygotskian design framework. *International Journal of Artificial Intelligence in Education* **10**, 198–220.
- O'Shea, T. and Self, J. (1983). *Learning and Teaching with Computers: Artificial Intelligence and Education*. Brighton: Harvester Press.
- Scanlon, E. and Issroff, K. (2005). Activity theory and higher education: Evaluating learning technologies. *Journal of Computers and Learning* **20**(6), 430–439.

Relevant Website

<http://www.merlot.org> – Multimedia Educational Resource for Learning and Online Teaching (MERLOT).

Facilitating Students' Success – Student Support and Learning Resources in Universities

E Woodberry and C Barnes, University of New England, Armidale, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

Context

This entry examines the role of student support services in promoting success in university study. Such services have been a feature of higher education systems worldwide for decades. However, their critical importance has not always been recognized. This situation has changed as a consequence of the transformation of the higher education sector since the early 1990s.

The international shift from elite to mass higher education has greatly increased the demand for student support services. Newer entrants to universities often lack the academic literacy skills higher education programs require. The rapid growth in distance education programs has also affected demand, as learners at a distance typically require greater support than those in traditional programs. This need is most acute in online learning, where additional assistance is often necessary to allow students to cope with new technologies.

Growing emphasis on accountability has also played a part in the renewed interest in student support services. Employers, taxpayers, and governments are now more demanding in terms of educational outcomes, attrition rates, and graduate skills. Other factors have been the commercialization and internationalization of higher education in many industrialized countries. Competitive pressures and the growing importance of foreign student income have made many institutions far more sensitive to the needs of students as consumers. It is no longer possible for universities in Western countries to be complacent about student outcomes.

Equally significant has been the change in teaching philosophies. Even those students enrolled in traditional on-campus programs are exposed to fresh challenges as newer forms of learning grow in popularity. Although intended to mitigate the perceived inadequacies of traditional pedagogies, newer active-learning methods, particularly those based on advanced technologies, demand much more of students, particularly in terms of information literacy. Student services have an important role to play in facilitating student adjustment to state-of-the-art teaching environments.

What Does Student Support Entail?

Student support usually involves a range of nondiscipline-based services that provide guidance and tuition tailored to

each student and at every point in the student's academic progress. Effective student support programs include opportunities for students to access: pre-enrolment guidance, career counseling, work-ready programs, course advising, orientation sessions, basic literacy and numeracy programs, study skills tuition, academic literacy instruction, library skills tuition, information literacy training, technical support, personal counseling, disability support, and advocacy.

This list is not exhaustive, representing only some of the range of student support programs typically provided in a major university in an industrialized country.

Academic Support

History

Academic support programs in higher education have a much longer history than is often recognized. Modern programs are based on a tradition that dates back at least as far as the early college preparation courses offered at the University of Wisconsin in 1849. By the 1950s, writing courses for freshmen were common in community colleges in the United States of America. From these early beginnings, effective academic support has emerged as a critical element in institutional and program success. Academic support is now seen as crucial to improved learning outcomes, student fulfillment, and higher completion rates.

Early research into academic support programs is often associated with the work of John E Roueche in the United States community college sector during the late 1960s and 1970s. Roueche and his colleagues at the University of Texas at Austin emphasized mastery of learning as a core concept in academic support. This idea took the form of highly structured, performance-based instruction in which frequent testing was used to measure learning and to provide feedback. Although a behaviorist, Roueche favored varied teaching approaches. These included class discussions and group projects. Roueche also argued that academic support programs should involve a personal counseling component. Many of Roueche's insights have been confirmed by later investigators.

Although early research provided a significant body of knowledge, progress in the field since the 1970s has been uneven. Academic support programs, particularly in the United States, initially took the form of remedial

education: tuition in reading, writing, and basic mathematics skills. The initial prevalence of deficit models meant that such services were often accorded low priority as a research topic and were starved of resources in practice. As delivery was often contingent on a range of local circumstances, much research was in the form of 'how to do it' case studies. This emphasis hampered efforts at the development of generalized theory. However, considerable progress was made during the 1980s in the United States in terms of developmental education, an approach that introduced topics such as study skills and time management.

New Theoretical Perspectives

In more recent decades, deficit models have lost ground. Research outside the United States has led to a richer understanding of the importance of academic support programs. The success of first-year students in adjusting to university study is less often taken for granted. Stresses associated with an unfamiliar learning environment, social isolation, and inevitable disappointments are recognized as potentially affecting all new learners, not only weaker or under-prepared students. Another factor has been growing awareness that students require more explicit instruction in academic culture. Academic literacies research has challenged the assumptions that the discipline-specific genres are self-evident, or that student acculturation through simple exposure to university practice is easy or natural. There has also been a growing acceptance of the point that learning outcomes and completion rates are affected by teaching quality. Academic support services are required to remedy deficiencies in teaching as well as flaws in learning. Although this point was made by Roueche in the 1970s, it has taken three decades for the decision-makers in the higher education sector to come to terms with realities long apparent to generations of students.

Approaches

Academic support services are currently delivered through a range of delivery mechanisms. These include traditional lectures, self-paced workbooks, peer-led study groups, group tutorials, one-to-one counseling, computer-based training (CBT) packages, Web-based tutorials, audio tapes, video tapes/DVDs, podcasts, and blogs and wikis.

Although each of these approaches has yielded positive results, some should not be used in isolation. When used alone, CBT packages do not generally result in satisfactory learning outcomes. The effectiveness of traditional lecture format relative to more student-centered, active learning approaches has also been questioned in the academic support context.

Information Literacy

Importance

Student training in information literacy (IL) has, for many years, been recognized as a crucial element in student support. This is an area in which librarians and other information specialists can offer a significant contribution, particularly when working in conjunction with specialist academic staff. The teaching of information skills (particularly when tailored to a specific subject or discipline) provides a long-term benefit to students and assists in reducing attrition and promotes life-long learning.

History

The work of the American scholar Zurkowski (1974) is usually seen as the starting point for modern concepts of information literacy. Growing US interest in IL during the late 1980s led to the *Final Report* (1989) of the American Library Association (ALA) Presidential Committee on Information Literacy. This report is now seen a milestone in the development of IL and it is still highly influential. By the mid-1990s, IL concepts had spread well beyond the United States, leading to curriculum changes in universities in the United Kingdom, The Netherlands, South Africa, and Australia.

The 1990s also saw the development of a range of theoretical frameworks. Kuhlthau's Information Search Process (ISP) model emerged in 1993 (Kuhlthau, 1993). This has proved to be one of the most significant theoretical interventions in the IL field, and has been applied to IL research and practice in universities in many countries. Other crucial theories include Bandura's (1997) theory of self-efficacy, and Kapitzke's critical IL model (Kapitzke, 2001). Librarians and educationalists in this decade have also explored the implications of older theoretical perspectives, such as Bloom's taxonomy of educational objectives (Bloom, 1956), to IL instruction.

The release of a number of comprehensive IL standards marked the growing maturity of the IL field in higher education. The most important of these was the *Information Competency Standards for Higher Education* (2000) issued by the US Association of College and Research Libraries (ACRL, 2000). The ACRL document provided the basis for the *Information Literacy Standards* (2001) published by the Council of Australian University Librarians (CAUL, 2001). In its position paper *Information Skills in Higher Education* (1999), the UK Society of College, National and University Libraries (SCONUL, 1999) provided a list of seven headline skills and a model for IL. These three documents have provided a framework for the development of nation-level IL initiatives in many countries. The ACRL standards alone have been translated into upwards of 15 languages, including French, German, Italian, Finnish, Japanese, Spanish, and Chinese.

The strategic importance of IL on the Anglo-American model of information skills is currently acknowledged by universities across Europe, North America, and other parts of the industrialized world. This is the case even in Eastern Europe, where IL has taken a different trajectory for many years. The systematic incorporation of IL instruction into curricula across all levels of university study is increasingly seen as the ideal. There is growing recognition that effective IL instruction cannot be separated from discipline-related content or from assessment. The implication is that successful IL requires collaboration between academic staff, university administrators, librarians, and other information professionals. In addition, agreement grows that IL is best taught through problem-based learning and other forms of authentic, student-centered instruction. This consensus is far from complete, but few educators in the industrialized world would now see IL instruction in isolation or give it a low priority.

Progress in higher education institutions in the developing world has been uneven. IL initiatives are often hampered by the lack of technology, resources, trained personnel, and infrastructure. Despite efforts by UNESCO and the International Federation of Library Associations (IFLA) to raise awareness of IL issues worldwide, progress has been limited. South Africa is a notable exception. Recent developments in other developing nations, such as the People's Republic of China, also give grounds for optimism.

Approaches

In the early 1990s, most libraries taught IL using more-or-less traditional methods. These included library tours, printed pathfinders or subject guides, self-paced workbooks, subject-specific lectures, face-to-face tuition, and telephone support for distance learners.

Whatever the mode of instruction, the emphasis was on library orientation and specific tools, such as the online public access catalog (OPAC). Course-related instruction was often limited to students in specific subjects or those faced with particularly difficult assessment tasks. Little weight was placed on transferable skills.

These older methods have not been superseded or entirely replaced. In universities in many parts of the world, they remain standard. They are now only part of a wider range of instructional modes. Some modes, like the greatly emphasized on hands-on instruction (as opposed to the traditional lecture format), or the use of Web-based subject guides, represent evolutionary, rather than revolutionary, change. More striking has been the proliferation of technology-driven modes, many of which are in their infancy.

Universities, and in particular university libraries, are testing the use of technologies such as the following to assess their usefulness for information literacy: podcasting,

screencasting, bulletin boards and email lists, virtual chat, instruction by SMS, blogs and wikis, and Web-based tutorials.

Emphasis has also shifted from tools-based instruction toward tuition designed to ensure that students make effective use of information they find. The Web has resulted in an information-rich environment in which responsibility for assessment and evaluation now falls primarily on the user.

There is also greater recognition of the fact that information literacy is a worthwhile goal in itself, as demonstrated in the Graduate Attributes programs prevalent in Australia at the turn of the twenty-first century. While instruction based on textbooks or collected readings still has a role in university education, particularly in developing countries, the increasing availability of online information has shaped teaching strategies. The trend in many industrialized countries is toward learning strategies in which students learn through engagement with an ever-increasing range of information resources, often those which they would use in the workplace. Enthusiasm for problem-based learning (PBL) is one factor in this development.

The ideal is articulated as a situation where IL learning is embedded into the curriculum at all levels. In this model, IL instruction is designed through collaboration between librarians and academic staff.

Instruction takes the form of specific assessment tasks that involve a structured, systematic search for information. Although widely recognized as desirable, this approach has often proved to be difficult to realize in practice.

The creation of Web-based packages providing generic IL instruction is often seen as a cost-effective alternative. These are relatively easy to build and do not depend on academic collaboration. They can also incorporate interactive exercises, self-assessment tools, and online tests. Although such packages are useful, their generic nature means that they fall short of the ideal. Despite this drawback, they are increasingly popular and a handful have achieved an international reputation, having been translated or adapted into a number of languages. Examples of Web-based IL programs with an international profile include TILT (USA) and eSKILLS UNE (Australia).

Learning Spaces

The revolution in information technology in the 1990s, simultaneous with the availability of information resources in electronic form, had a major impact on university libraries. The necessity to provide the technology to support access resulted in university libraries redeveloping themselves as institutional access points for digital resources. The inevitable convergence of these trends

led to the emergence of the concept of Information Commons, which included not only equipment but also new services required to support students in access and utilization of the technology and the resources. Lippincott (2006) defines the underlying philosophy of the Information Commons as one of providing users with a technology-rich space where students can access, manage, and produce information all at the same workstation.

History

One of the earliest examples of an Information Commons was the facility established in 1994 by the Leavey Library at the University of Southern California. In the intervening period, the Information Commons concept has been employed at university libraries in countries such as Australia, Canada, Malaysia, New Zealand, South Africa, and the United Kingdom. Early best-practice examples were at the University of Calgary and the University of Guelph in Canada, with later developments such as the Kate Edgar Information Commons at the University of Auckland and the University of Otago in New Zealand, significantly expanding the concept.

Developments

Recognition that the design of space impacts on learning emerged as a significant issue in the late 1990s. Bennett (2003) identifies the fundamental changes that occurred during this period as recognition of the social dimensions of learning, and the revolution in information technology. Design of learning spaces was challenged by a change in student requirements, the pervasiveness of information technology, the emergence of e-learning technologies and an increased understanding of learning. Research carried out by Oblinger (2006) for Educause in the United States, the Joint Information Systems Committee (JISC, 2006) in the United Kingdom, and the Canadian Association of Research Libraries (CARL, 2005) develop the concept of learning space design and its application for academic support services.

Responding to the demand from students for increased variety in spaces and technologies, the Information Commons concept was further developed to that of a Learning Commons. Lippincott (2006), in *Linking the Information Commons to Learning* defines the purpose of the Learning Commons as being to enhance student learning and scholarship. The Learning Commons concept recognizes that libraries need to respond to changes in teaching practices and student demand not only for social space and technology access but also for access to academic support personnel, technology-enabled shared facilities for project work and group study purposes. In practice, the Learning Commons provides an integrated work environment for students

with technologies and spaces that support learning theory principles; they are student centered, support a range of flexible learning methods, and provide a single point of access to support services.

The commons personnel can include advisors in writing, numeracy, academic literacy, library, information skills and information literacy, together with assistance in the utilization of the technologies and associated software. The technology-rich environment of the commons requires that students have immediate access to technical support staff for troubleshooting and advice. The scope of the learning commons can also include consultation spaces, rooms for group work (such as peer-led study sessions), and electronic classrooms to support the delivery of academic support programs.

Early examples of Learning Commons developments are the University of Queensland Ipswich campus in Australia and the Learning Centre at the Glasgow Caledonian University in Scotland.

The Learning Commons ideal extends beyond the idea of a dedicated physical space, which supports on-campus students, to online support services. An example is the Learning Commons at the University of Guelph, which incorporates a very extensive Web site including self-paced instruction modules, multimedia resources, and interactive self-assessment tools. The future of student support services lies with strategies that make best use of the possibilities inherent in the latest generation of educational technologies within a physical space and online.

Student Support in Distance Education

History

Student support programs for distance learners have a long history. The first universities to provide correspondence tuition to their own learners were in North America, beginning with the Illinois State University (1874). The University Correspondence College (1887), a British provider, was perhaps the first to offer distance learners a well-thought out program of student support. The University Correspondence College was a private institution providing support to nonresident students intending to sit for University of London degrees. It ran classes in London and Cambridge, offered residential schools, published its own textbooks and offered tuition by correspondence. The University Correspondence College anticipated the comprehensive student support systems associated with higher education in the late twentieth century.

The methodical approach of the University Correspondence College was very much the exception. For most of the twentieth century, distance learners in higher education were largely dependent on informal support from partners,

families, and friends to overcome the frustrations and loneliness of study by correspondence. This situation changed in the early 1970s, when the Open University in the United Kingdom revolutionized distance higher education through the concept of supported open learning.

Open University learners enjoyed study skills advice, telephone support, face-to-face tutorials, and a network of local resource centers, in addition to tuition by correspondence. One of the strengths of the Open University program was 'continuity of counseling'. Students were assigned a personal tutor-counselor who provided support and advice throughout their degree. The Open University also made use of broadcast media as well as high-quality print materials.

During the 1980s and 1990s, the Open University provided a model for many single-mode distance providers. Although the Open University system remained a standard of excellence, not all of the 'second-generation' universities had the resources to provide the same level of support to distance students. In fact, the Open University itself was forced to modify its program in 1997, in part due to the need to control costs.

Another competing model in these decades was the University of New England (1955), an Australian institution that pioneered dual-mode distance education in the 1960s. Dual-mode instruction was based on the principle that on-campus and distance students should be taught by the same academic staff. Student support mechanisms at the University of New England included telephone counseling, face-to-face tuition at residential schools, audio tapes, and broadcast radio. However, the standards of student support at dual-mode institutions were dependent on a range of institutional factors. In parts of the developing world, dual-mode institutions found their ability to provide effective support constrained by inadequate resources and the often rudimentary state of national infrastructures.

New Approaches

As in other areas in higher education, the 1990s represented a period of renewal for student support services. Universities had previously experimented with a range of technologies for supporting distance students as independent learners. These included telephone counseling, audiotapes, television, radio, video, computer disk, CD-ROM, and written materials. Technological changes and cheaper telecommunications made it possible for distance education providers to go beyond this goal: to facilitate the involvement of distance students in collaborative, social learning. Initially, the creation of online learning communities involved the use of email or bulletin boards, often through commercial learning management systems such as WebCT and Blackboard.

More recently, universities have looked toward Web 2.0 technologies to provide an improved level of student support for distance learners. A growing number of universities have built student portals, which offer rich functionality and access to a personalized range of electronic resources. Many administrative functions have been moved online, reducing delays and providing students with instant access to information on enrolment and course progression. New technologies – blogs, wikis, and podcasts – provide opportunities for the innovative delivery of student support services.

See also: Student Attrition from Higher Education Institutions.

Bibliography

- ACRL (2000). *Information Competency Standards for Higher Education*. Chicago, IL: Association of College and Research Libraries. <http://www.ala.org/ala/mgrps/divs/acrlstandards/standards.pdf> (accessed August 2009).
- ALA Presidential Committee on Information Literacy (1989). *Final Report*. Chicago, IL: American Library Association.
- Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York: Freeman.
- Bennett, S. (2003). *Libraries Designed for Learning*. Washington, DC: Council on Library and Information Resources.
- Bloom, B. S. (ed.) (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. New York: David McKay.
- CARL (2005). *Libraries and e-Learning: Final Report of the CARL E-Learning Working Group*. Ottawa: Canadian Association of Research Libraries. http://www.carl-abrc.ca/projects/e_learning/e_learning-e.html (accessed August 2009).
- CAUL (2001). *Information Literacy Standards*. Canberra: Council of Australian University Librarians. <http://www.caul.edu.au/caul-doc/InfoLitStandards2001.doc> (accessed August 2009).
- JISC (2006). *Designing Spaces for Effective Learning: A Guide to 21st Century Learning Space Design*. Bristol: Higher Education Funding Council for England. http://www.jisc.ac.uk/uploaded_documents/JISC%20learning%20spaces.acc.pdf (accessed August 2009).
- Kapitzke, C. (2001). Information literacy: The changing library. *Journal of Adolescent and Adult Literacy* **44**(5), 450–456.
- Kuhlthau, C. C. (1993). *Seeking Meaning: A Process Approach to Library and Information Services*. Norwood, NJ: Ablex.
- Lippincott, J. (2006). Linking the information commons to learning. In Oblinger, D. G. (ed.) *Learning Spaces*, pp 7.1–7.18. Washington, DC: Educause.
- Oblinger, D. G. (ed.) (2006). *Learning spaces: An Educause e-Book*. Washington, DC: Educause. <http://www.educause.edu/LearningSpaces/10569> (accessed August 2009).
- SCONUL (1999). *Information Skills in Higher Education: A SCONUL Position Paper*. London: Society of College, National and University Libraries.
- Zurkowski, P. G. (1974). *The Information Service Environment: Relationships and Priorities*. Washington, DC: National Commission on Libraries and Information Sciences.

Further Reading

- McCartin, M. and Feid, P. (2001). Information literacy for undergraduates: Where have we been and where are we going? In

- Lynden, F. C. (ed.) *Advances in Librarianship*, pp 1–27. San Diego: Academic Press.
- Spencer, M. E. (2006). Evolving a new model: The information commons. *Reference Services Review* **32**(2), 242–247.
- Tait, A. and Mills, R. (eds.) (2003). *Re-Thinking Learner Support in Distance Education: Change and Continuity in an International Context*. London: RoutledgeFalmer.

Relevant Websites

- <http://tilt.lib.utsystem.edu> – Texas Information Literacy Tutorial TILT (USA).
- <http://www.une.edu.au> – University of New England: eSKILLS UNE.

Faculty Values and Expectations and Their Implications for Teaching, Learning and the Organization of Higher Education Institutions

K Harman, University of New England, Armidale, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Academic culture – Represents the deeply embedded, historically transmitted traditions, myths, values, and expectations that are peculiar to academic occupational life and work.

Academic freedom – Typically refers to the unencumbered search for truth, freedom for academics to undertake research and publication in their specialist fields and to teach without any restrictions in their area of expertise.

Academic values – Are those strongly held beliefs and principles that guide academic work and life. Three such values or 'tenets of faith' in academia are academic freedom, collegiality and professional autonomy.

Collegiality – In academia refers to the active participation of faculty in academic and democratic decision-making, the professional sharing of ideas, as well as the mutual support and sociability it offers. Collegiality derives from the 'community of scholars' ideal which reflects the dominant ethos of the medieval guild – collegial decision making by a body of equals in an unhierarchical and collaborative enterprise.

Research-led teaching and learning – Implies a close relationship exists between research, teaching, and student learning. It holds that as teachers are at the forefront of knowledge they need to be actively engaged in scholarship and research, teaching, and learning must be evidence-based and curricula designed that reflect a research-based approach to learning.

Introduction

Like all social entities, the world of academia possesses a symbolic dimension – an academic culture that, unlike institutional structures, cannot be visibly or diagrammatically depicted, but can only be inferred from the values and expectations of academic faculty who reside there. The potency of academic culture and its impact on the life

and work of faculty and their institutions should never be underestimated. Three integral tenets of faith peculiar to this culture, which influence significantly the core business of higher education and its organizational arrangements, are those of academic freedom, collegiality, and professional autonomy.

Academic freedom and professional autonomy in teaching, research, and publication are two forcefully defended values of academic culture. These are, however, in danger of being eroded as faculty values and norms in many countries come under considerable pressure to conform to changes wrought by demographic, economic, social, and technological developments (Honan and Teferra, 2001: 183). Traditional collegial practices are also under threat as progressive corporatization and managerialism take stronger hold in higher education institutions (Kreber, 2006: 6). Measures taken to conform to these pressures impact on the core business of higher education institutions and the way teaching is carried out. Growing competition and financial constraints mean that professional decisions are now becoming financial ones, forcing institutions to increase the attractiveness of academic programs to potential clients (Anderson *et al.*, 2002: 54). Competition also leads to institutions increasingly defining their distinctive approaches to teaching and learning (Zubrick *et al.*, 2001: ix).

In addition, pushing for change are the significant influences of massification and diversification of the student body in higher education, the changing role of research in knowledge economies, societal demands for lifelong education, and changing labor-market needs (Evans, 2002; Enders, 2004; Kehm, 2007; Harman, 2008). In such fast-changing environments, themes of growing interest in higher education are how best to safeguard and sustain the integrity of academic culture while simultaneously providing quality teaching and learning.

Attention, in this article, is directed to core values and expectations enshrined in a deeply entrenched academic culture that influence the teaching role of faculty, student learning, and the organization of higher education institutions. Of further interest are external pressures impacting on higher education institutions that have modified the teaching role and placed the integrity of deeply rooted faculty values and expectations under increasing strain.

Faculty Values and Expectations: Core Components of Academic Culture

Academic faculty represent the embodiment of a distinctive culture as they uphold and sustain deeply held traditions, myths, values, and expectations that are peculiar to their occupational life and work. These cultural traits are historically transmitted, cumulative, deeply embedded and not easy to turn-off at will. The thicker the culture, the more potent is the culture's influence (Harman, 2002: 97).

The most deeply embedded values and expectations shared by academic faculty are those of academic freedom, collegiality, and professional autonomy. Creativity, critical appraisal by peers (a sacred cow of academia), and integrity in teaching, research, and scholarship also form integral elements of academic culture. The influence of this culture on teaching, learning, and the organization of higher education institutions is pervasive.

Academic freedom – the most forcefully defended value – refers to the unencumbered search for truth, freedom to undertake research and publication in scholars' specialist fields, and to teach without any restrictions in their area of expertise. Another interpretation sees academic faculty as public intellectuals who should have the freedom to express their views in any public forum without fear of retribution or censure (Altbach, 2007: 24) – a value that has been eroded significantly over the last 20 years with the “crushing of the dissident voice” in heavy managerialistic institutional contexts (Anderson *et al.*, 2002: 57).

Collegiality – the active participation of faculty in academic decision making, the professional sharing of ideas, as well as mutual support and sociability (Anderson *et al.*, 2002: 47–48) – is another strongly defended value which does not sit easily alongside managerial styles and structures that exclude faculty from decision making in nonacademic areas. Collegial, democratic decision making is based on the principles that intellectual authority derives from the disciplines and truth is no respecter of status and hierarchy. As a value, collegiality derives from the community-of-scholars ideal which reflects the dominant ethos of the medieval guild – collegial decision making by a body of equals in an unhierarchical and collaborative enterprise.

Professional autonomy – closely related to academic freedom – represents the ability of academics to exercise their professional judgment about what they teach and research and is another strongly defended value. Growing competition, financial constraints, and accountability are, however, eroding this autonomy, as noted earlier. This autonomy–accountability tension is a ‘perennial concern’ in higher education as external bodies demand tighter control over higher education (Altbach, 2007: 20).

Faculty Values Guiding Teaching and Learning

The teaching role typically encompasses communicating via traditional face-to-face interactions and technological means, preparation of lesson content, curriculum development, individual consultations with students (real and virtual), postgraduate supervision, provision of feedback to students, and assessment of their learning. Faculty values and expectations that have a marked influence on this teaching role are identified in the sections that follow.

Research-Led Teaching and Learning

Good teaching which draws on disciplinary expertise, professional experience, and is continually revitalized by research and scholarship, is valued for fostering quality learning outcomes (Australian Vice-Chancellors' Committee (AVCC), 2002a: 1). This value implies a close relationship exists between research and teaching. It holds that teachers are at the forefront of knowledge and are actively engaged in scholarship and research, teaching and learning must be research-led, and curricula designed that reflect a research-based approach to learning (Anderson *et al.*, 2002; Kreckel, 2002; Jenkins *et al.*, 2003; Harman, 2005). Research-led teaching is seen to inform practice, encourage inquiry/problem-based learning, provide up-to-date knowledge on curriculum content and development, and inform teachers better on how students learn effectively and how particular teaching methodologies encourage deep learning. This implies that knowledge imparted by faculty needs to be evidence based and teachers must make decisions about curriculum, teaching pedagogy, and modes of assessment on the basis of well-informed research and scholarship. Thus, teachers must be intellectually engaged as learners who acquire knowledge and improve practice through research and scholarship that augments research findings (Brown, 2002: 5).

While arguments persist over how close the relationship is, or needs to be, between teaching and research (Hattie and Marsh, 1996; Elton, 2001; Scott, 2007) and while some teachers are clearly not active researchers, a strongly held value is nevertheless that students should be taught by experts in their respective fields of research. Problems typically experienced by institutions with teachers who are nonresearchers are that those who do not engage in research often use outdated knowledge, are limited to merely using knowledge generated by others, and are constricted in scope and range of teaching skills. These problems are highlighted in the report, the Boyer Commission on Educating Undergraduates in the Research University (1998: 1–2), which argues that the teacher-as-researcher or scholar-teacher moves from a base of original inquiry

and that students should be taught by those who discover, create, and transmit insights about subjects in which they are the experts.

Deep Learning Versus Surface Learning

Another strongly held value is that good teachers encourage deep learning which builds on the intrinsic interests of the learner, is inquiry/problem-based, and seeks the interrelationship of new knowledge with existing knowledge. Surface learning (memorizing or rote learning) is discouraged as this method merely encourages passively storing and processing information rather than synthesizing, evaluating, and analyzing. Work that values teacher as facilitator, learner, and educative leader and adopts an interactive mode of teaching that encourages problem solving, will ultimately be more successful in ensuring deep student learning. The Boyer Commission on Educating Undergraduates in the Research University (1998: 1) adds that inquiry-based learning possesses an element of reciprocity – faculty learn from students (the learner–teacher concept) as students learn from faculty.

Higher education reform agendas, in many transition economies, place significant value on improving teaching, learning, and assessment methods. In Vietnam, lecturers are being urged to move from passive to interactive teaching modes that encourage active participation of students and to ensure their teaching is research-led. The former Minister of Education and Training emphasized recently that interactive teaching and problem-based learning increase learners' participation and their ability to work in teams, and adapt to their jobs and future careers. The teaching principle is promoted as “teaching how to learn” and “learning how to learn” (Ministry of Education and Training (MOET), 2005: 18).

The Move to the Learning Paradigm

As the workplace that graduates will enter is vastly different now than even a decade ago, the modern college or university is increasingly encouraged by external constituencies to strengthen undergraduate and postgraduate education by shifting from an emphasis on instruction to an emphasis on learning (Barr and Tagg, 1995; Rice, 2006). The greater value placed on learning devalues the instructional (didactic) paradigm exemplified by the passive transmission of information in lecture-discussion format where faculty talk and students listen – a form that does not encourage deep learning.

Focusing on the learner rather than the teacher leads to new expectations for how faculty will enact their roles. Rice (2006: 16) talks of the need for the unbundling of the traditional role – a change that poses a serious challenge to the holistic conception of the teacher–scholar ideal rooted in a collegial culture where peer review and

self-monitoring are the norm. Faculty members now need to know how to support and advise students, how to facilitate learning through applying a range of innovative learning processes, and be able to link learning with life experiences and service in the community.

Stress on learning opposed to teaching is seen mostly as a problem in research universities. In Germany, for example, it is argued that the teaching function has remained faithful to the Humboldtian tradition, but the learning function is less clear. The traditional Humboldtian concept of concentrating on teaching the elite students that would somehow elevate the pedestrian rest is an idea that is not seen to be viable any more (Kreckel, 2002). It is now argued that effective teaching not only assists students less well prepared for university, but also provides those imbued with the consumer-rights ethic better value for their investment.

One implication for institutions of higher learning still holding to the instructional paradigm is that in the context of diminishing resources alongside the ever-increasing quest for quality, they are faced with a real dilemma – how to increase quantity and quality of outputs without a corresponding increase in resources.

External Pressures Impacting on Values and Expectations Guiding Teaching and Learning

The New Political Economy of Higher Education

A number of major changes impacting on teaching and learning in many higher education systems are encapsulated in the literature under ‘the new political economy.’ According to Rich (2006: 37), these changes include:

- erosion of public funding;
- escalating costs in the face of public fiscal stringency;
- restraint sought on increasing fees and tuition;
- increase of political and ideological intrusiveness, for example, managerialism and growing corporatization in a more profit-driven environment;
- changing expectations of students influenced by user choice and the emphasis on students as customers/consumers;
- professional standards dictated by outside control, for example, by government regulation rather than by disciplinary norms;
- globalization of economies and research systems, intensification of international competition (free trade), and the rise of new information technologies;
- erosion of academic freedom to determine curriculum content and standards and methods of teaching; and
- greater accountability on access, cost containment, and learning outcomes.

In this climate, faculty feel the pressure from both top-down and bottom-up; from administrators reacting to market forces and from students demanding new styles of presentation (Hagner and Schneebeck, 2001: 2). Moreover, the core value of academic freedom is seriously threatened (Altbach, 2001: 205).

Technological Advances

The impact of information and communication technologies (ICTs) that liberate learning from constraints of time and space has revolutionized traditional forms of teaching. Teaching is continually being challenged with emphases on borderless education using e-learning, with online interactive teaching modes, to meet the developing needs of a more diverse student population globally and locally. Traditional face-to-face delivery modes are now challenged by more flexible multimedia approaches which add value to lifelong and distance learning and expand the webs of learning available in a global context.

As technology advances and increases in off-campus teaching and flexible delivery modify the teaching role, teaching faculty are now expected to be up to speed with these developments and adapt accordingly. Implications are that teaching must adapt and ongoing training for faculty needs to be provided if quality teaching and learning, curriculum design, and assessment practices are to result.

Resulting from the impact of the digital revolution, a typology of faculty is offered by Hagner and Schneebeck (2001: 3–5). The predominant types comprise:

1. The *entrepreneurs* characterized by a high level of commitment to quality teaching and learning, work that tends to be idiosyncratic, and a preference to use their expertise to solve their own instructional problems.
2. The *risk averse* characterized by a commitment to quality teaching but a lack of technical expertise, fear that current success in teaching will not translate into new teaching environments, and a need for significant levels of instructional support.
3. The *reward seekers* whose motivation is tied to reward structures and who see adoption of technology-based teaching as a way to advance careers.
4. The *reluctants* characterized by computer illiteracy or a firm belief that traditional learning models are superior, that new technology may have adverse effects on teaching evaluation, and change is fearful.

Implications of the above four categorizations are that institutions need to know their faculty-mix before beginning any transformation or e-skilling processes (Hagner and Schneebeck, 2001: 5). The fact that so much more new information is being generated and disseminated more quickly than ever before highlights the need for faculty to continue to acquire new skills, develop new

mindsets and practices, expand their knowledge bases throughout their careers, and possess a capacity for continuous learning.

Massification and Diversification: Demand for New Skills

Apart from signaling the need for greater resources and different organization and governance modes, mass education access has meant a marked growth of nontraditional students oriented toward gaining useful skills and knowledge. This growth demands adjusting teaching to serve a greater range of students with diverse needs, aspirations, and academic talents (Trow, 2000: 323). Concerns are that outdated knowledge and ivory-tower curricula repeatedly stifle creative thinking and fail to produce students who can apply academic knowledge to real-life situations.

With increasing diversity of students' needs (especially of mature-age learners) has come new education packages that offer convenient, quality, and low-cost courses. In response to these demands, many higher education providers now offer certificates and short courses that are particularly attractive to adults who face changing employment demands or wish to explore personal interests (Austin, 2002: 122).

Importantly, the knowledge economy is generating a need for emerging skills and knowledge to replace outdated knowledge that has not, earlier, been part of higher education curricula – skills such as initiative and enterprise, information literacy, planning and management skills, capacity for lifelong learning and self-learning (learning how to learn), flexibility in order to adapt to jobs yet to be envisaged, and skills useful for working in multidisciplinary contexts. As the integration of work and learning is now so critical in knowledge economies, employers are increasingly demanding researchers who can integrate knowledge across traditional disciplinary boundaries and demonstrate industry readiness when they graduate. Many universities have now adopted integrative doctoral programs that provide graduates with skills that are transferable to a range of different work contexts (Harman, 2004: 387–388).

The Economic Imperative

Government economic agendas and user-choice ideology, in many countries, have impacted profoundly on teaching values and expectations in higher education systems as the relationship between learner and teacher is becoming more like that of consumer and producer (Bjarnson, 2001: 97). A vice-chancellor of a leading British university argues that university teaching is changing from a tradition based on craft to a model based on industry, where the emphasis is more on developing skills for economic purposes (Bjarnson, 2001: 86). Likewise, students who develop

skills for more instrumental ends have been characterized by Cunningham *et al.* (2000: 3) as earner-learners and just-in-time learners (similar to just-in-time manufacture) who will be able to apply what they have learned immediately in the workforce.

The possibility of mix-and-match courses from a variety of institutions and corporate providers gives the student-as-consumer (Bjarnson, 2001: 89) tremendous power to determine the content and direction of study. Such practices tend to erode the long-held traditional value of teaching for intrinsic worth and self-development. However, the emphasis on continuing professional development and postgraduate provision in professional areas such as law, computing, information technology (IT), business and management, engineering, and languages suits the concept of borderless education.

Building Learning Societies

Emphasis on the value of lifelong learning indicates that teaching and student learning should be focused on the needs of tomorrow (Commonwealth Department of Education, Science and Training (DEST), 2002: ix) as well as on the short-term future. Lifelong learning is particularly valued in developing countries. The Vietnamese Government, for example, acknowledges that there is a need to build a learning society where every effort should be made by all educational institutions to create opportunities for lifelong learning (Ministry of Education and Training (MOET), 2005: 8).

That building learning societies ultimately leads to citizens enjoying a better quality of life and that developing new learning and research skills enhance considerably a nation's development and standard of living, are reinforced by the World Bank. Knowledge and advanced skills are seen to comprise critical determinants of a country's economic growth and standard of living. Learning outcomes are transformed into goods and services and greater institutional capacity, a more effective public sector, a stronger civil society, and a better investment climate result. High-quality, merit-based, equitable, efficient higher education teaching and research are stressed as essential components of this transformation. All countries, whether industrially based or in transition, are seen to benefit from the dynamic of the knowledge economy (The World Bank Group, 2005: 1).

Accountability Demands

Indicators that measure the quality of teaching and learning outcomes are now well established in many higher education systems. Accountability has also become more prominent as new providers with little experience in higher education enter the scene, especially where quality is seen to be threatened.

Implications of accountability are that teaching faculty need to be aware of more formalized procedures that institutions have put in place for measuring learning outcomes and comply with these. Graduate attributes resulting from skills learnt during the undergraduate years are one form of measures used to assess institutional teaching and learning capacity. In Australia, the Review Committee on Higher Education Financing and Policy (West Committee, 1997) suggested seven attributes that a first-degree graduate should have on exit: the capacity for critical, conceptual, and reflective thinking; technical competence and an understanding of the conceptual and theoretical elements of their fields of specialization; intellectual openness and curiosity and an appreciation of the interconnectedness and areas of uncertainty in current human knowledge; effective communication skills in all – reading, writing, speaking, and listening; research, discovery, and information-retrieval skills, and a capacity to use information; multifaceted problem-solving skills and the capacity for team work; and high ethical standard in personal and professional life, underpinned by a capacity for self-directed activity. Encouragingly, the Committee went on to add, "Above all else, however, the successful graduate must be captivated by a love of learning" (West Committee, 1997: 47).

In the UK, in the same year, the Dearing Committee (National Committee of Inquiry into Higher Education (The Dearing Report, UK), 1997) came up with four graduate attributes: communication skills, numeracy, the use of IT, and learning how to learn. In such contexts, expectations are that faculty will work toward ensuring that these standards are met.

Implications for Higher Education Institutions: New Initiatives in Teaching and Learning

It is well acknowledged that good teachers impart the art of and the thirst for learning and unlock the potential of their students who, in turn, will ultimately have an enormous impact on the future cultural and economic life of a nation. In this respect, aspirations of institutions of higher learning have changed little over the centuries (Commonwealth Department of Education, Science and Training (DEST), 2002: v). What has changed, however, is the way that many institutions are now set up to ensure that quality teaching and student learning prevail and are rewarded. Teaching and learning plans are now firmly embedded in most institutional plans and the requirement to demonstrate quality teaching now forms a key element of most tenure and promotion-seeking faculty. Teaching and learning centers that aim to facilitate the highest quality of student learning by instilling into faculty the ability to manage effectively the design, presentation, teaching, and assessment of the subjects for which they are responsible are now part and

parcel of most universities. These centers are typically based on values that support scholarship, research-led teaching, lifelong learning, problem-based learning, e-learning, self-learning, and professional and personal growth through critical reflection and self-evaluation (Australian Vice-Chancellors' Committee (AVCC), 2002a: 2).

Other initiatives now also acknowledge that quality teaching in higher education needs to be identified, recognized, and rewarded. For Australian universities, there is now an explicit separation of teaching and learning funding streams from research streams – changes that will have long-term significance (Dearn, 2006: 39). These changes have been marked by establishment of the Carrick Institute for Learning and Teaching in Higher Education (now the Australian Learning and Teaching Council) to support excellence in teaching and learning, the Awards for University Teaching Scheme, and the Learning and Teaching Performance Fund. To be eligible for the last-mentioned benefit, universities must satisfy the Commonwealth Government that they have met certain minimum standards for learning and teaching.

British higher education has also seen funding initiatives to introduce accredited teaching programs, e-learning, and centers of excellence in teaching and learning. For example, in 2004, the Higher Education Academy was created with the specific intent of providing institutional support for enhancing the learning experience. It acts to provide the sector with access to research, and develop and fund pathfinder projects and networks which bring faculty from different disciplinary groups together with senior institutional managers (Ramsden, 2007: 37). Additionally, some universities in the UK are creating different kinds of teaching and learning centers as the production and imparting of knowledge created within the academy, typically characterized as Mode 1, is no longer seen to be solely the responsibility of universities. Mode 2 knowledge – generated from a variety of sources which transcend traditional boundaries and is more applications generated – now questions the monopoly of Mode 1 (Bjarnson, 2001: 94).

In some transition economies, various policy instruments have been used by the World Bank (2002: 104) to stimulate performance and enhance the quality and relevance of teaching in higher education institutions. These include well-designed competitive funds, accreditation mechanisms, and management and information systems.

A good example provided by the Bank of a successful competitive fund to develop teaching programs is The Engineering Education Fund in Egypt which transformed the traditional engineering degree into more applied programs that link closely with industry. In Jordan, a similar fund is available to teachers in science and technology fields and an operations manual has been devised. In addition, Brazil, Mexico, and Venezuela are developing human capital in science and technology through

competitive funding. In all these funding programs, international peer review experts play a prominent role in assessing the quality of proposals.

In Vietnam, strategies to upgrade the knowledge of teaching staff in universities have been adopted by the Ministry of Education and Training. These include training-abroad programs and relationship building between universities in the same region and overseas so that teachers might experience practices in other universities and other countries. The World Bank is also establishing different kinds of cooperative ventures between universities in transition economies. This encourages universities to expand bilateral cooperation in their regions by co-sponsoring courses, student exchanges, and lecturers sharing their experiences across national and international boundaries (Harman, 2005).

Concluding Comments

While a healthy new culture of professionalism in teaching – similarly termed the scholarship of teaching and learning (Hatch, 2005) – has emerged in higher education, many factors aiding this development have impacted on the values and expectations to which academic faculty traditionally subscribe. The modern academic workplace is now characterized by the move to near-universal student access, greater student diversity, the impact of the digital revolution, globalization, changing economic and societal expectations, tighter competition, greater demands for accountability, new forms of governance with growing corporatization, expanding faculty workloads, industrial demands, and a new labor market for faculty. Austin (2002: 123) adds that fiscal constraint, the emergence of new institutional types, the introduction of postmodern ways of knowing, and dramatic shifts in the nature of faculty appointments have all impacted on the core business of higher education. Most of these changes have been met with institutional responses that now place greater emphasis on the quality of and more distinctive and innovative approaches to teaching and learning at both undergraduate and postgraduate levels.

Implications are that the newly developed teaching and learning culture needs to be sustained with a coordinated effort by institutions in areas of quality control, staffing and teaching quality, and more innovative curricula. Research activity of teaching faculty also needs to be strengthened and their knowledge continually updated. As one professional higher education body stresses, if high-quality learning outcomes are to be achieved, institutions of higher learning need to possess a teaching and learning culture that acknowledges the value of teaching, promotes scholarly inquiry and reflection, provides opportunities for professional development of faculty in their disciplines as well as in teaching methodology,

has supportive policies and practices that attract faculty with a strong commitment to effective teaching, and is able to promote or reward faculty for teaching excellence (Australian Vice-Chancellors' Committee (AVCC), 2002b: 10).

Importantly, however, as higher education institutions respond to changing environmental pressures that continue to threaten traditionally held faculty values and expectations and, ultimately, the core business of higher education, institutional managers have a special responsibility to uphold the integrity and independence of teaching faculty and to foster, nurture, and safeguard those enshrined values and expectations that undergird the very essence of the academic enterprise.

See also: Academic Freedom; Higher Education and the Knowledge Society; Promoting Effective Student Learning in Higher Education; Quality Assurance in Higher Education – Practices and Issues; Transforming Higher Education in Developing Countries: The Role of the World Bank.

Bibliography

- Altbach, P. G. (2001). Academic freedom: International realities and challenges. *Higher Education* **41**(1–2), 205–219.
- Altbach, P. G. (2007). Peripheries and centers: Research universities in developing countries. *Proceedings of the 8th International Conference on Education Research*. 23–25 October, Seoul National University, Korea, 3–38.
- Anderson, D., Johnson, R., and Saha, L. (2002). *Changes in Academic Work: Implications for Universities of the Changing Age Distribution and Work Roles of Academic Staff*. Canberra: Commonwealth of Australia.
- Austin, A. (2002). Creating a bridge to the future: Preparing new faculty to face changing expectations in a shifting context. *Review of Higher Education* **26**(2), 119–144.
- Australian Vice-Chancellors' Committee (AVCC) (2002a). *Guidelines for Effective University Teaching*, July, Canberra.
- Australian Vice-Chancellors' Committee (AVCC) (2002b). *Universities and their Students: Principles for the provision of education by Australian universities*, Canberra, AV-CC, December.
- Barr, R. B. and Tagg, J. (1995). From teaching to learning – A new paradigm for undergraduate education. *Change* **27**(6), 13–25.
- Bjarnson, S. (2001). Managing the changing nature of teaching and learning. *Minerva* **39**(1), 85–98.
- Boyer Commission on Educating Undergraduates in the Research University (1998). Reinventing undergraduate education, Stony Brook, State University of New York, 1–5. <http://naples.cc.sunysb.edu/Pres/boyer.nsf/webform/1> (accessed August 2009).
- Brown, S. (2002). Linking teaching and research in higher education. *Keynote Address to the Faculty of Education Research Conference*, University of Strathclyde, September, 3–7.
- Commonwealth Department of Education, Science and Training (DEST) (2002). Striving for quality: Learning, teaching and scholarship. *Issues Paper Prepared for the Government Review, Higher Education at the Crossroads*, Canberra, Commonwealth of Australia.
- Cunningham, S., Ryan, Y., Stedman, L., et al. (2000). The business of borderless education in brief. *Report No. 38, Higher Education Division*, Canberra, Department of Education, Training and Youth Affairs.
- Deam, J. (2006). Enhancing the quality of teaching: An Australian perspective. *New Directions for Higher Education* **133**, 33–41.
- Elton, L. (2001). Research and teaching: Conditions for a positive link. *Teaching in Higher Education* **6**, 43–56.
- Enders, J. (2004). Research training and careers in transition: A European perspective on the many faces of the PhD. *Studies in Continuing Education* **26**(3), 419–430.
- Evans, T. (2002). Part-time research students: Are they producing knowledge where it counts? *Higher Education Research and Development* **21**(2), 155–165.
- Hagner, P. R. and Schneebeck, C. A. (2001). Engaging the faculty. In Barone, C. A. and Hagner, P. R. (eds.) *Technology-Enhanced Teaching and Learning*, ch. 1, pp 1–12. San Francisco, CA: Jossey-Bass, Higher and Adult Education Series.
- Harman, K. (2002). Merging separate campus cultures into a coherent educational community: Challenges for higher education leaders. *Higher Education* **44**(1), 91–114.
- Harman, K. (2004). Producing 'Industry-ready' doctorates: Australian cooperative research centre approaches to doctoral education. *Studies in Continuing Education* **26**(3), 388–404.
- Harman, K. (2005). Strengthening the links between teaching, learning and research in higher education. *Paper No. 2 Prepared for the Vietnam Ministry of Education and Training (MOET) Higher Education Project No. 2 (HEP 2)*, Hanoi, 25 June, 1–14.
- Harman, K. (2008). Challenging traditional research training culture: Industry-oriented doctoral programs in Australian cooperative research centres. In Välimaa, J. and Ylijoki, O.-H. (eds.) *Cultural Perspectives on Higher Education*, ch. 12, pp 174–190. Dordrecht: Springer.
- Hatch, T. (2005). *Developing the Scholarship of Teaching and Learning*. San Francisco, CA: Jossey-Bass The Carnegie Foundation for the Advancement of Teaching.
- Hattie, J. and Marsh, H. W. (1996). The relationship between research and teaching: A meta-analysis. *Review of Educational Research* **66**, 507–542.
- Honan, J. P. and Teferra, D. (2001). The US academic profession: Key policy challenges. *Higher Education* **41**(1–2), 183–203.
- Jenkins, A., Breen, R., and Lindsay, R. (2003). *Reshaping Teaching in Higher Education: Linking Teaching and Research*. London: Kogan Page.
- Kehm, B. M. (2007). The changing role of graduate and doctoral education as a challenge to the academic profession: Europe and North America compared. In Kogan, M. and Teichler, U. (eds.) *Key Challenges to the Academic Profession*, pp 111–124. Paris/Kassel: UNESCO Forum on Higher Education Research and Knowledge/International Centre for Higher Education Research Kassel.
- Kreber, C. (2006). Setting the context: The climate of university teaching and learning. *New Directions for Higher Education* **133**, 5–11.
- Kreckel, R. (2002). 'Teaching' and 'learning' in Germany: Structural continuities, changing conditions and necessities of change, DAAD (German Academic Exchange Service). Conference on Universities of the Future, Bonn, June, 105–111.
- Ministry of Education and Training (MOET) (2005). Conference conclusions: 'Renovation in higher education in Vietnam – integration and challenge'. Speech delivered by the Minister of MOET, Mr Nguyen Minh Hien.
- National Committee of Inquiry into Higher Education (The Dearing Report, UK) (1997). Higher Education in the learning society. *Report of the National Committee into Higher Education* (Sir Ron Dearing, Chairman), Leeds, NCIHE.
- Ramsden, P. (2007). Getting better all the time. *Australian* **10**(37).
- Rice, R. E. (2006). Enhancing the quality of teaching and learning: The US experience. *New Directions for Higher Education* **133**, 13–22.
- Rich, D. (2006). Academic leadership and the restructuring of higher education. *New Directions for Higher Education* **134**, 37–48.
- Scott, P. (2007). The distinctive contribution of universities to research and innovation. *Paper Presented to Seminar Four – Supporting and Developing Research and Knowledge Transfer*, Langbank, Scotland, 1–11 November.
- The World Bank (2002). *Constructing Knowledge Societies: New Challenges for Tertiary Education*. Washington, DC: The World Bank.
- The World Bank Group (2005). Tertiary Education, pp 1. <http://www.worldbank.org/education/tertiary> (accessed August 2009).

- Trow, M. (2000). From mass higher education to universal access: The American advantage. *Minerva* **37**(4), 303–328.
- West Committee (Review Committee on Higher Education Financing and Policy) (1997). Learning for life: Review of higher education financing and policy, *Final Report* (Rod West, Chairman), Canberra, Department of Education, Employment and Youth Affairs.
- Zubrick, A., Reid, I., and Rossiter, P. (2001). Strengthening the Nexus Between Teaching and Research, Evaluations and Investigations Programme (EIP) Higher Education Division, Canberra, DETYA.

Further Reading

- Becher, T. (2001). *Academic Tribes and Territories: Intellectual Enquiry and the Cultures of Disciplines*, 2nd edn. Milton Keynes: Open University Press.
- Dill, D. (1982). The management of academic culture: Notes on the management of meaning and social integration. *Higher Education* **11**(3), 303–320.
- Henkel, M. (1997). Academic values and the university as corporate enterprise. *Higher Education Quarterly* **51**(2), 134–143.

Flexible Learning in Higher Education

S Alexander, University of Technology Sydney, Sydney, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

The term flexible learning was largely unknown prior to the 1990s, became a buzzword in higher education during the 1990s, and has largely been replaced by other terms, such as blended learning, over the course of the 2000s. This article provides an overview of the context in which flexible learning became one of the driving forces of higher education during the 1990s, its successes and consequences, and the changing language of the discourses over the last decade.

In this article, the term flexible learning is taken to represent a range of approaches to meet the varied needs of students in contemporary social contexts. The needs of students include independence in terms of time and location of learning, and the availability of some degree of choice in the curriculum (including content, learning strategies, and assessment). The approaches taken to meet these needs include the use of contemporary information and communication technologies to support a range of learning strategies, and other learning strategies such as learning contracts.

Flexible Learning and the Context of Higher Education

Although the term flexible learning began to emerge in higher education discourses during the early part of the decade, its high watermark was the mid- to late 1990s. It was no accident that time period coincided with the release of the first popular web browser, NCSA Mosaic, in 1994. The subsequent rapid adoption of the Internet gave rise to predictions of the end of higher education as we knew it, and in its place, the emergence of new providers such as private, online, for-profit institutions, and media and communication organizations. It was widely believed that these providers had ready access to content which could be beamed out to students all over the world who would learn about theories firsthand from those who had conceived them.

Others, such as Cunningham *et al.* (1997), studied the potential for this new era of higher education to be realized, but concluded that “the ‘threat’ of communication networks had been overstated” (p. 207). Despite this, the apprehension remained and indeed underpinned much of the activity of the particular discourse where the term flexible learning was used interchangeably with the use of information and communication technologies.

During that period there were many who believed that higher education was in crisis, based on a number of factors including:

- the move to mass higher education;
- escalating costs of provision; and
- lack of access to education.

Thus, the arrival of the Internet was very attractive to university administrators such as Daniel (1997) who wrote that “technology provides the most fertile ground for growing these key ingredients of university renewal: lower costs and unique attractions” (p.14).

Flexible Learning: Toward a Definition

The *Oxford Dictionary* (2005) reports the word flexible as being derived from the Latin word *flexibilis*, from *flectere* meaning to bend. Its definitions include “to be easily modified to respond to altered circumstances” and “capable of bending easily without breaking.” Both of these meanings are relevant to the descriptions in this article about the ways in which flexible learning has evolved as a consequence of the changing circumstances in which higher education found itself operating during the 1990s, and the ways in which it has been implemented without breaking the higher education system.

Despite the fact that there has been much discussion about flexible learning in the higher education community, this discussion has invariably focused on the meaning of the term itself, rather than its implementation. The consequence has been that a shared understanding of what it means in practice has not evolved. This imprecision has contributed to its status as a mantra in higher education (Cloonan, 2004), where different stakeholders ascribe meanings to suit their own purpose. Vice-chancellors, for example, were said to be enthusiastic about flexibility in learning and teaching because of “a perception that flexible delivery is more effective and efficient in terms of getting teaching resources to students” (Bigum and Rowan, 2004). Managers and politicians are said to focus on “effectiveness and efficiency and cut price solutions to the delivery of a service” (Nunan, 1996: 2), while some faculty see it as a path to improving the experience for their students and/or as a path to certain promotion, as its use is deemed to have the approval of management. Thus, as expressed very well by Edwards and Nicoll (2000: 130)

“flexibility in higher education is not a single phenomenon but has a range of possible meanings and practices.”

Notwithstanding the lack of clarity around a shared meaning, there has been almost universal acceptance of flexibility as a positive attribute – good for students as well as good for the university. As a result, references to flexible learning in the strategic plans of universities and/or its teaching and learning plans and policies became common during the 1990s.

The next section discusses the range of meanings of flexible learning, the first of which is of flexible learning as synonymous with the delivery of education through the use of new information and communication technologies.

Flexible Learning as Flexible Delivery

This definition, the emergence of which coincided with the development and release of multimedia-capable computers and the Internet in the early 1990s, focuses on the mode of delivery as the most important aspect of flexible learning. At its most minimal, the term flexible delivery is taken to mean any form of delivery where students have a choice of study mode (Roberts, 2001: 22). This is said to give students choice of location of study (as the course materials are delivered to them), and time at which they study, thus widening access to higher education. The design of the materials may also facilitate students’ control of learning through choice of media (e.g., text vs. video), and pathway through the content.

It was also common to associate the use of flexible modes of delivery with traditional distance-education practices where flexible learning has replaced distance education as a means of “servicing the needs of geographically distant or remote students” (Kirkpatrick, 1997: 160).

As well as the focus on the delivery of course materials, for some there has been an assumption that flexible delivery means the use of advanced learning technologies (Kirkpatrick, 1997; Roberts, 2001; Sadler-Smith and Smith, 2004; Collis and Moonen, 2002).

Information and Communication Technologies

The development and release in the early 1990s of multimedia-capable computers and the Internet came at the very time when the crises of higher education were being widely discussed and thus became the catalyst for many of the changes that ensued. University administrators and academics alike were attracted, as many commentators noted, like moths to the flame to use of the new technologies which were seen as the panacea to the contemporary issues they faced.

In one the largest initiatives of its kind in the world in higher education, the UK government provided

significant funding for the Teaching and Learning Technology Programme (TLTP) with the aim of “making teaching and learning more productive and efficient by harnessing modern technology” (Teaching and Learning Technology Program, 1996). In Australia, successive national government committees, such as the Committee for the Advancement of University Teaching (CAUT) and the Committee for University teaching and Staff Development (CUTSD), offered grants to institutions for the development of projects to improve teaching and learning. Although not specifically targeting the use of new information and communication technologies, the reality was that a significant proportion of the projects submitted involved the use of the new technologies.

Early developments in the use of new technologies for flexible learning included the development of CD-ROMs (CD-ROM, compact disk, read-only-memory) containing course materials in the form of lecture notes, databases, animations, simulations, and collections of still and moving images, and sounds. Some of the learning materials were said to be interactive when there was a degree of learner choice through, for example, following hypertext or hypermedia links, and through multiple-choice selections on screen. Other programs used more sophisticated interactive multimedia software to provide that user choice. Although still in use today, especially in locations where the Internet is unavailable, slow, or unreliable, CD-ROMs have largely been replaced by the Internet as a delivery mechanism.

From the late 1990s until the present day much of the focus of higher education’s use of the Internet has been through the deployment of specialized software variously referred to in different countries as Course Management Systems, Virtual Learning Environments, and Learning Management Systems. These tools have become almost ubiquitous in higher education institutions in First World countries. Although this software had its origins in developments in computer-mediated conferencing in the mid-1970s, it was rarely used outside conventional distance-education higher education institutions until the mid-1990s. Early commercially available software included FirstClass, Lotus Notes, Top Class, Blackboard, and WebCT. More recently, there has been a significant move toward the adoption of open-source software such as Moodle and SAKAI.

Those tools have been supplemented by more recent technological developments such as the mobile phone (including the iPhone), Personal Digital Assistants (PDAs), and Personal Access Devices (PAD), which, when used with ubiquitous wireless communications with high bandwidth, enable what is now known as m-learning or mobile learning. Students may use these devices to access new mobile versions of Learning Management Systems, and to download and listen to or watch podcasts (audio recordings) and vodcasts (video recordings) of lectures

available on Internet sites such as iTunesU and YouTube, thus facilitating access to content at a time and location convenient to them.

Some universities have opened up the content of their courses to anyone with Internet access. The most famous of these is the Massachusetts Institute of Technology (MIT) OpenCourseWare initiative which, at the time of writing, has 1900 courses available online complete with syllabi, course materials, assignments, and in some cases, videos of lectures.

The degree of interactivity afforded by the earlier technologies was minimal compared to that available in the late 2000s. The rapid adoption of social networking tools such as Facebook, LinkedIn, Twitter, and MySpace have provided opportunities for the proliferation of user-controlled networks of 'friends' sharing photos, videos, blogs, and personal profiles (Mason and Rennie, 2008). Third-party tools, such as Flickr (photos) and del.icio.us (social bookmarking), can be integrated with personal profiles. Students form special-interest groups (which share common pages and message boards and may be private or public) within social networking sites which, in the higher education context, are used for diverse activities such as group homework problem solving, and general academic support. However, tempting it is for universities to make use of these sites, students express a strong preference to be left alone within them (Aleman and Wartman, 2009). Each of these tools does, however, support learning which is time and location independent.

In addition to the growing use of these Web 2.0 tools, there has been the increased sophistication of virtual worlds software (such as Second Life) which enables learners to inhabit computer-generated worlds using a computer-generated self known as an avatar. While some higher education institutions have reproduced real-life classrooms in which lectures take place, others have taken the opportunity to design qualitatively different learning experiences for students.

Effectiveness of Information and Communication Technologies for Flexible Delivery

Early evaluations of the effectiveness of these innovations found some evidence of benefits of the use of these new technologies to higher education. In a large-scale national study of 104 projects, the majority of which had developed computer-based learning/interactive multimedia products, Alexander and McKenzie (1998) found that for students there was some evidence in some projects of:

- improved quality of learning;
- improved productivity of learning;
- improved access to learning; and
- improved student attitudes to learning.

Examples of positive learning outcomes which resulted from students' use of information-technology (IT)-enabled projects included:

- opportunities for students to interact with others to gain a more sophisticated and global understanding of complex international political issues;
- improved understanding of concepts which students are known to have difficulty with in a range of disciplines, through the use of interactive multimedia animations, simulations, and microworlds;
- enhanced communication between part-time students and their lecturer, through the use of the Internet;
- the acquisition of information, such as language learning, where a high component of factual recall is required; and
- the facility for students to assess their own learning of concepts, through computer-based qualitative and quantitative assessment modules.

The benefits for faculty involved in the projects were found to include: job satisfaction flowing from the improved learning of their students; an improved understanding of student learning, student needs, and difficulties; an improved understanding of their own discipline area; enhanced enthusiasm for teaching; and, for some, an increase in their own personal profile.

For departments or faculties, the major benefit was the faculty development opportunities afforded by faculty member's participation in the projects. This sometimes led to significant change in teaching approach in areas other than the designated project, as faculty developed enhanced understanding of learning and teaching. For some departments, the project outcome helped faculty to cope with decreased resources, without a commensurate decrease in the quality of teaching. Finally, the teaching profile of some departments was raised as a result of external recognition of the innovation.

The study emphasizes the fact that it is not the presence of information and communication technologies by itself that accounts for enhanced learning, rather it is the design of the learning experience, the support of the learners undertaking that experience, and the students' perception of the context in which they learn.

In a more recent study of the use of lecture recording technologies used across four universities, Gosper *et al.* (2008) report that students perceive the following benefits of lecture recording technologies in improving their learning: picking up on things they missed in class; to revise for exams; to revisit complex material, ideas, and concepts; and to work through material at their own pace.

Faculty have expressed two main concerns around the use of these lecture recording technologies: that students would not pace their use of these technologies, resulting in the need to cram learning at the end of the semester,

and/or that they would stop coming to lectures. This fear was somewhat confirmed by the Gosper *et al.* study in their report that only approximately 50% of students listened to the recordings on a regular basis and almost 40% listened to several weeks at the one time. Of the students who no longer attended lectures, as reported in the study, 68% said they had chosen not to attend because they could learn as well from the recorded lectures as they could from the face-to-face lectures.

Critique of Flexible Delivery

Alongside the benefits ascribed to flexible delivery using the information and communication technologies discussed above, there are also significant challenges. First and foremost, there remain, despite the rhetoric about the high levels of IT literacy of generations X and Y, whole cohorts of students who do not have sufficient levels of IT literacy to access the course materials. Adequate technical support is essential to these learners in achieving successful learning outcomes.

A second factor is that of students reporting insufficient time to devote to the course. Mason (2001) famously proclaimed that “time is the new distance,” since lack of time, rather than long distance, has become one of the primary reasons that students withdraw from courses. The self-discipline required to undertake a course that has been delivered to the door is far greater than that required by students whose attendance requirement at face-to-face lectures serves as a time-management strategy. It is often only the experienced learners who have the time-management skills to undertake such self-directed study, yet the majority of students in the higher education system are undergraduates, who have come directly from high school, and who are studying full time.

A third critique has centered around the degree to which information and communication technologies are in fact innovative. Many authors have pointed to the long-term existence of books and papers which have for many years facilitated learners’ access to the content of learning in a way that is time and location independent. These authors also point to the fact that learners have control over the order in which they read printed material, thus facilitating learner control over the content of learning that is held up as a unique feature of information and communication technologies. Others, such as Alexander and Boud (2001), argue that the potential for the use of information and communication technologies to enhance learning has been lost because faculty have, by and large, simply used the new technologies to automate existing didactic practices. Much of what passes for innovation in learning is little more than lectures that have been turned into podcasts and vodcasts, and textbooks which have been repurposed as websites with electronic page turning for example.

The view of flexible learning described above which equates flexible learning with the use of technologies for flexible delivery of teaching/learning has been criticized for being a particularly narrow view of flexible learning. It does, nonetheless, remain a prominent discourse in both everyday discussions and promotional literature.

There are, however, other views of flexible learning which are discussed below.

Flexible Learning as an Educational Philosophy and a Political Position about Education

Another view of flexible learning differs from that of flexible delivery because it foregrounds the students’ learning and their experience of education as paramount, rather than the features of the delivery mechanisms as have been described above. As Nunan states:

Part of the attraction of flexible learning for many educationists is the belief that it enables students to use desired learning strategies which involve them in constructing meaning, and thereby educationists link the term with student centred learning, self-directed learning, activity learning, problem based learning . . . and the like. (Nunan, 1996: 5–6)

While the above definition focuses on particular views of learning as underpinning the design of flexible-learning strategies, others focus on providing options for students. In this view, students, rather than faculty, choose the dimension of learning from a range of options.

Flexible Learning as Providing Access and Democratization of Processes of Teaching and Learning

This view of flexible learning relates to democratization of learning by improving access to higher education, thus challenging the power relationships said to exist between teachers and learners (Nunan, 1996: 1). Those who now have increased access to higher education include people who live in geographic locations which do not have physical universities. It also includes a significant group for whom, despite living in reasonable proximity from a university, work and family commitments or illness and disability previously precluded their attendance on campus. A positive consequence of increased access is a corresponding increase in diversity of the student body.

Having gained access to higher education, this group is said to no longer be subject to learning which is teacher centered, having opportunities instead to control and direct their own learning.

Flexible Learning as a Set of Strategies in Response to New Challenges in Teaching and Learning and Changing Student Needs

The challenges of higher education that are said to be new include: a decrease in traditional sources of funding for many higher education institutions; evidence that students are undertaking longer hours of paid work to finance their studies; and increased numbers of students seeking to change careers.

The response of many higher education institutions to the decrease in traditional sources of funding has been to increase class sizes. This does not necessarily lead to a less personalized learning experience for students, however, as faculty draw upon many contemporary innovations in learning and teaching strategies to offer a degree of flexibility in content and assessment of learning, and to provide individualized feedback.

There is mounting evidence that the availability of podcasts and vodcasts has made it easier for students who are unable to attend class because of work commitments to continue their enrolment. The same students also make use of the range of tools (e.g., discussion boards) which form part of the learning management systems which have now become ubiquitous.

For those students seeking a change of career, flexible learning can be designed to provide an experience which is tailored to their individual needs, which takes account of their background and existing knowledge, and which can be undertaken in a workplace. The majority of these students are more mature in age, more experienced as learners, and bring with them much existing knowledge. These learners may take advantage of flexible strategies such as individual learning contracts (affording a degree of learner choice of both content and assessment), and work-integrated learning opportunities.

A Critique of Flexible Learning

The major criticism of flexible learning centers on its alleged flexibility. Releasing students from the need to attend regular on-campus classes, and instead making it possible for them to learn at their own pace using flexible learning strategies transfers the time-management responsibility entirely to students, some of whom are not yet able to cope with this degree of flexibility. Less-experienced (and usually younger) students have not yet learned to pace their learning, often leaving the majority of activities to the last week/s of semester. This results at best in learning which is not as deep as it might otherwise have been, and at worst in stress and failure. Some argue the case for more inflexibility in learning for such students (Edwards and Nicoll, 2000),

where this term is not necessarily assigned a negative connotation.

Similarly, giving students the responsibility for negotiating the content of learning places them in a situation where they are said to require the symbolic-analytic literacies of “interpretation, analysis and communication of information – as well as capacities of critique, problem-solving, flexibility, teamwork, communication and self-direction” (Edwards *et al.*, 2002: 202). Learners who develop these are said to go on to become flexible lifelong learners.

A second significant area of critique lies in the particular discourse that equates flexible learning to the use of information and communication technologies. For some, the mere presence of these technologies somehow magically improves the learning experience. Often, however, they merely automate existing teaching practices such as lectures in the same way that earlier innovations were seen as automations of previous technologies (e.g., the telephone was conceived of as an automated telegraph). As noted earlier however, it is the design of the learning experience using the technologies that accounts for any improvements in learning rather than the presence of technologies.

Although the problems have now largely been resolved, the early years of flexible learning using information and communications technologies involved significant technical problems for many aspiring students. Problems included lack of bandwidth, unreliable networks, difficulties loading specialist software, and printing failures, to name just a few. For those who did not have access to adequate technical support, access to higher education was inflexible.

From Flexible Learning to Blended Learning

In recent times there has been an increase in use of the term blended learning such that its boundary with flexible learning has now become blurred. Early uses of the former suggested a simple combination of face-to-face learning experiences with online activities. More recently, however, the term has evolved to include dimensions which embrace both individual self-paced and collaborative learning activities, structured and unstructured learning experiences, and blending of both custom and off-the-shelf learning content (Khan, 2007: 7). The goal is to use these blends to create flexible learning experiences for students.

Exhibiting further similarity to the evolution of flexible learning however, there appears to be no single agreed definition of blended learning. In their review of over 300 studies of blended learning Sharpe *et al.* (2006) noted that “Although difficult to define, the term ‘blended learning’ is finding acceptance among higher education staff. We suggest that the advantages of the term include its poor definition – which allows staff to negotiate their own

meaning – the implication of the protection of face to face teaching, and the implication of designing for active learning.”

Conclusion

Students have much to gain from the new order of higher education where they need no longer sit passively in lecture theaters and tutorials at a time that suits the university, learning content that has been determined in the absence of considerations of what they already know, using approaches that lead to memorization rather than deep understanding of content, nor be assessed in ways that have no parallels with the world of work they are about to enter. Likewise, students, who have previously been unable to access higher education for a range of reasons, are now able to do so.

The significant expertise that has developed in the design of learning experiences, both face to face and online, means that there is the potential for a high-quality learning experience for all who seek out higher education.

See also: Distance Education and Open Universities; Effective Use of Technology in Teaching and Learning in HE; Innovation in Teaching and Curriculum Design; Pathways and Articulation into Higher Education; Promoting Effective Student Learning in Higher Education.

Bibliography

- Aleman, A. M. and Wartman, K. (2009). *Online Social Networking on Campus: Understanding What Matters in Student Culture*. New York: Routledge.
- Alexander, S. and Boud, D. (2001). Learners still learn from experience when online. In Stephenson, J. (ed.) *Teaching and Learning Online: Pedagogies for New Technologies*, pp 3–15. London: Kogan Page.
- Alexander, S. and McKenzie, J. (1998). *An Evaluation of Information Technology Projects in University Learning*. Department of Employment, Education and Training and Youth Affairs. Canberra: Australian Government Publishing Services.
- Bigum, C. and Rowan, L. (2004). Flexible learning in teacher education: Myths, muddles and models. *Asia-Pacific Journal of Teacher Education* **32**(3), 213–226.
- Cloonan, M. (2004). Notions of flexibility in UK higher education: Core and periphery re-visited? *Higher Education Quarterly* **58**(2–3), 176–197.
- Collis, B. and Moonen, J. (2002). Flexible learning in a digital world. *Open Learning* **17**(3), 217–230.
- Cunningham, S., Tapsall, S., Ryan, Y., et al. (1997). New media and borderless education: A review of the convergence between global media networks and higher education provision. *EIP Report 97/22*. Higher Education Division, Department of Employment, Education, Training and Youth Affairs, Canberra.
- Daniel, J. S. (1997). Why universities need technology strategies. *Change* **29**(4), 11–17.
- Edwards, R. and Nicoll, K. (2000). Flexible learning for adults. In Foley, G. (ed.) *Understanding Adult Education and Training*, pp 127–139. St Leonards: Allen and Unwin.
- Edwards, R., Nicoll, K., and Lee, A. (2002). Flexible literacies: Distributed learning and changing educational spaces. In Lea, M. R. and Nicoll, K. (eds.) *Distributed Learning: Social and Cultural Approaches to Practice*, pp 196–209. London: Routledge.
- Gosper, M., Green, D., McNeill, M., Phillips, R. A., Preston, G., and Woo, K. (2008). *Final Report: The Impact of Web-Based Lecture Technologies on Current and Future Practices in Learning and Teaching*. Sydney: Australian Learning and Teaching Council. http://www.altc.edu.au/carrick/webdav/site/carricksite/users/siteadmin/public/grants_project_webbasedlecture_report_aug08.pdf.
- Khan, B. H. (ed.) (2007). *Flexible Learning in an Information Society*. Hershey, PA: Information Science Publishing.
- Kirkpatrick, D. (1997). Becoming flexible: Contested territory. *Studies in Continuing Education* **19**(2), 160–173.
- Mason, R. (2001). *Time Is the New Distance?* An Inaugural Lecture. The Open University, Milton Keynes, 14 February. http://kmi.open.ac.uk/projects/stadium/live/berrill/robin_mason.html.
- Mason, R. and Rennie, F. (2008). *E-Learning and Social Networking Handbook*. London: Routledge.
- Nunan, T. (ed.) (1996). Flexible delivery – What is it and why is it part of educational debate. *Proceedings HERDSA Conference 1996*, Perth Western Australia, 8–12 July. <http://herdsa.org.au/confs/1996/nunan.html> (accessed August 2009).
- Roberts, T. S. (2001). From distance education to flexible learning. *Proceedings of the 16th Annual Conference for the International Academy for Information Management*, New Orleans, LA, 21–25 December.
- Sadler-Smith, E. and Smith, P. E. (2004). Strategies for accommodating individuals' styles and preferences in flexible learning programmes. *British Journal of Educational Technology* **35**(4), 395–412.
- Sharpe, R., Benfield, G., Roberts, G., and Francis, R. (2006). *The Undergraduate Experience of Blended E-Learning: A Review of UK Literature and Practice*. Higher Education Academy. https://www.heacademy.ac.uk/assets/York/documents/ourwork/research/literature_reviews/blended_elearning_full_review.pdf (accessed August 2009).
- Teaching and Learning Technology Program (1996). *TLTP: The Teaching and Learning Technology Programme*. <http://www.ariadne.ac.uk/issue4/tltp/> (accessed August 2009).
- The Oxford Dictionary (2005). The Oxford Dictionary of English, (Soanes, C. and Stevenson, A. (rev. edn.)) *Oxford Reference Online*. Oxford University Press. University of Technology, Sydney. <http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t140.e28284> (accessed August 2009).

Further Reading

- Bates, A. W. (1997). Restructuring the university for technological change. *Paper Presented at What Kind of University?* 18–20 June, The Carnegie Foundation for the Advancement of Teaching, London. <http://bates.cstudies.ubc.ca/carnegie/carnegie.html> (accessed August 2009).
- Bonk, C., Graham, C. R., Cross, J., and Moore, M. G. (2005). *The Handbook of Blended Learning: Global Perspectives, Local Designs*. San Francisco, CA: Pfeiffer.
- Guest, T. (2007). *Second Lives: A Journey through Virtual Worlds*. London: Hutchison.
- Laurillard, D. (2002). *Rethinking University Teaching: A Conversational Framework for the Effective Use of Learning Technologies*, 2nd edn. London: Routledge.
- Mason, R. (1994). *Using Communications Media in Open and Flexible Learning*. London: RoutledgeFalmer.
- Nunan, T., George, R., and McCausland, H. (2000). Rethinking the ways in which teaching and learning are supported: The Flexible Learning Centre at the University of South Australia. *Journal of Higher Education and Policy Management* **22**(1), 85–98.

- Rowntree, D. and Lockwood, F. (1994). *Preparing Materials for Open, Distance and Flexible Learning: An Action Guide for Teachers and Trainers*. London: RoutledgeFalmer.
- Sharpe, R. (2006). *Why Blend? Rationales for Blended E-Learning in Undergraduate Education*. The Higher Education Academy. www.heacademy.ac.uk/assets/York/documents/ourwork/research/literature_reviews/blended_elearning_why_blend.pdf (accessed August 2009).
- Trigwell, T. and Oliver, M. (2005). Can 'blended learning' be redeemed? *E-Learning* 2(1), 17–26.
- Twigg, C. A. and Oblinger, D. G. (1996). The virtual university. *Report from a Joint Educom/IBM Roundtable, Nov 5–6*. Washington, DC.
- Wade, W., Hodgkinson, K., Smith, A., and Arfield, J. (1994). *Flexible Learning in Higher Education*. London: Kogan Page.

Innovation in Teaching and Curriculum Design

J M Dearn, University of Canberra, Canberra, ACT, Australia

© 2010 Elsevier Ltd. All rights reserved.

Innovation in Higher Education

Higher education is usually known more for its traditional approaches to teaching and curriculum than for innovation. However, recent years have seen a widespread change, both in what is taught in universities and colleges and how teaching takes place. The pressure for change has come from two main drivers. First, a growing realization that traditional teaching methods are relatively ineffective for many students, and second, the changing demands being made on higher education resulting from the changing nature of work and knowledge, the changing nature of the student population, and the emergence of new information and communication technologies. Notable developments have been the growth of academic-development units in colleges and universities which have been established to support staff in their teaching and promote educational innovation, the proliferation of books and resources on teaching methods, and the emergence of national institutions devoted to supporting innovation in higher education.

Nevertheless, despite the diversity of approaches to both teaching and curriculum design now seen in higher education, the view persists, that higher education has been slow to respond, at least in relation to changes that have taken place in other community sectors.

Recent national reviews of higher education systems have made a clear connection between higher education, and both economic growth and social development and have called on higher education to greater embrace innovation. The Australian 2002 national review of higher education, *Higher education at the Crossroads*, stated “We need a system that produces graduates who can think critically and have adaptable skill sets as well as technical expertise. In many ways, that will require significant transformation of approaches to curriculum and pedagogy to stimulate and stretch students and accommodate their varying needs” (Department of Education, Science and Training, 2002: 14).

The USA 2006 national review of higher education, *A Test of Leadership: Charting the Future of U.S. Higher Education*, also commented on the importance of higher education saying “And for the country as a whole, future economic growth will depend on our ability to sustain excellence, innovation, and leadership in higher education” (Spellings, 2006: 1). The review concluded that “Too many of our colleges and universities have not embraced opportunities to be entrepreneurial, from testing new

methods of teaching and content delivery to meeting the increased demand for lifelong learning” (Spellings, 2006: 4–5) and “Institutions as well as government agencies have failed to sustain and nurture innovation in our colleges and universities. The commission finds that the results of scholarly research on teaching and learning are rarely translated into practice” (Spellings, 2006: 15). Any understanding of innovation in teaching and curriculum design in higher education must consider not only the changing demands being made on higher education by government and employers, but also the way these forces are interpreted, acted on, and disseminated within the academy.

Innovation and the Importance of Context in Higher Education

While there is evidence of considerable change occurring in higher education, the concept of innovation is complex. For example, what is designated an innovation could be a general educational concept, such as the social construction of knowledge, which could be implemented in a variety of different ways in different teaching contexts. On the other hand, it could be a new procedure, such as a laboratory process in a science class, which needs to be implemented prescriptively. Furthermore, it needs to be agreed whether it is enough for an idea to be new for it to be called an innovation or whether it also has to be shown to be effective in achieving its stated objectives. Even if a new idea is effective in achieving its objectives, it can be argued that it needs to be adopted by somebody other than the person who first introduced it, to be designated an educational innovation. Higher education is replete with examples of innovations that have not gone beyond the original developer. A particular problem is that many educational innovations are dependent for their effectiveness on the particular context within which they are implemented, be it the individual teacher, the discipline, or the type of institution. Even innovations that have been shown to be highly effective in one particular context may not work in another.

Even the concept of new is problematical. For example, a teaching practice such as case-based teaching would be considered innovative in an undergraduate science program but is a commonplace approach to teaching in graduate law programs. Similarly, what would be considered innovative teaching in a traditional university might

be a common practice in a newer institution. To add yet another layer of complexity, for each of these types of innovation, the change can be regarded as novel or an adaptation or modification to an existing innovation.

The importance of context for innovation is also seen in cross-cultural comparisons where approaches to teaching and what is regarded as appropriate pedagogy and curriculum vary markedly between different countries. Finally, it is worth noting the importance of context at the level of the individual where the adoption of a familiar approach to teaching might be seen for some individuals as highly innovative. In part, this will reflect their underlying beliefs or perspectives about teaching and learning. There is a diversity of different models of teaching perspectives but the typology developed by Daniel Pratt will serve as an example. His model recognizes five different teaching perspectives based on the underlying beliefs and values held, implicitly in most cases, by teachers, which he labeled as transmission, apprenticeship, developmental, nurturing, and social reform. What approaches to teaching and curriculum design are implemented by individual teachers, and consequently which will be regarded as innovative, will depend in part on the particular perspective of their own teaching.

The contextual nature of innovation in higher education does not negate the usefulness of the concept but it does mean that judgments about whether a change is innovative can only be made in the context of the discipline, the type of institution, the culture, and the teaching perspective of the individual teacher as well as what the innovation was attempting to achieve. However, despite these problems, the persistence of the concept of innovation in discussions about higher education is indicative of both the level of change occurring in our colleges and universities and, more importantly, the need for change.

Types of Innovation in Higher Education

It is useful to recognize different types of innovation in higher education. First, some innovations attempt to improve the quality of student learning outcomes based on new ideas emerging from our understanding about the nature of learning. Second, innovations may be aimed at developing qualitatively new types of learning outcomes, reflecting, for example, the changing nature of work and society. Third, innovations may be attempting to increase access to higher education by different groups of students. Fourth, innovations may have as their aim the amelioration of adverse external factors such as the reduced availability of funding or increased class size. Finally, innovations may be focused on exploiting the potential of new information and communication technologies, the Internet being a prime example. Indeed, the impact that technology has on higher education is so great that for

many, educational innovation is synonymous with using educational technology.

These different types of innovation are not, of course, mutually exclusive. For example, an attempt to cope with increased class size by making students more independent of the teacher for their learning might, if somewhat fortuitously, improve the quality of student-learning outcomes. Similarly, teaching innovations using information and communication technologies are opening access to higher education to students who had previously been unable to attend regular classes on campus.

Innovations in higher education take many forms, each with their own issues related to support dissemination. Andrew Hannan and Harold Silver in their book *Innovating in Higher Education: Teaching, Learning, and Institutional Cultures* proposed a number of different types of innovation including those initiated by individuals, those initiated by discipline associations or professional bodies, institutional initiatives, and systemic initiatives responding to system-wide policy changes. The way innovations are perceived, disseminated, taken up, and implemented will differ in each case.

Innovation, Lone Rangers and Systemic Change

Innovation in higher education has for a long time been the province of individuals, the lone rangers, to quote Tony Bates' memorable term, a reflection of the particular culture of higher education where individual academic staff typically enjoy a relatively high level of autonomy over both what they teach and how they teach. However, the recognition-and-reward systems found in colleges and universities typically favor achievements in research rather than teaching, and staff wishing to introduce teaching and curriculum innovations frequently encounter significant barriers and even opposition.

There have, of course, been notable examples of successful innovations being introduced by individuals and as Hannan and Silver (2000: 32) note "It seems that innovators will take on extra work, learn new skills, court unpopularity with other staff and take risks with their own careers so long as they feel that by doing so they can improve the quality of their teaching, and/or, if they feel that circumstances are such that they have no choice but to depart from their old methods to cope with new demands."

However, even effective innovations have a poor record of being adopted by other people and being sustained in the absence of the original innovator. This has led in recent years to a recognition that in order to bring about sustained change, innovations need to be adequately funded, linked to the strategic goals of institutions, have the support of senior management, and be

supported by effective leadership and change management processes. The need for a more systemic approach to innovation was recognized in the influential 2002 US report *Greater Expectations: a New Vision for Learning as A Nation Goes to College* by the Association of American Colleges and Universities which noted “The academic world . . . as the repository of accumulated knowledge, functions as a conservator, slow to change in fundamental ways. Thus, new ideas appear as pockets of innovation, taking root at the margins of institutions, and maintaining themselves by the patient effort of a few dedicated individuals. These innovations only slowly permeate the mainstream” (American Association of Colleges and Universities, 2002: 19). The report concluded “The next step is to create from these isolated innovations a comprehensive movement for change across the higher education landscape” (American Association of Colleges and Universities, 2002: 12).

Initiatives such as the Australian Learning and Teaching Council, the Higher Education Academy in the UK, and Pew Foundation in the USA are examples of attempts to introduce large-scale system-wide sustainable change.

The Scholarly Basis for Teaching Practice in Higher Education

While there exists a considerable body of research literature on learning and teaching in higher education, it is largely disconnected from classroom practice, a reflection, in part, of the lack of any required systematic preparation for the teaching role in higher education. A notable recent development in higher education is the emergence of discipline-based initiatives such as those of the UK’s Higher Education Academy, the Carnegie Foundation, and the Australian Learning and Teaching Council where there is a recognition of the critical role played by disciplinary cultures in teaching and the different approaches to knowledge creation and inquiry in different disciplines and professions. There is no doubt that university or college teachers steeped in the traditions of their disciplinary field, be it physics or law, will respond more positively to initiatives aimed at improving student learning if these are framed within the culture and language of their discipline.

However, despite the many initiatives that are taking place to better support the development of teachers and teaching in higher education through staff-development programs and reward schemes such as fellowships and prizes, much teaching practice is still based on the individual teacher’s own personal experiences and falls far short of what might be regarded as evidence-based practice. In particular, there is a dearth of studies that critically evaluate the effectiveness of different teaching practices or interventions on student learning. One of the most

interesting and effective recent developments in higher education has been the introduction of what is called classroom assessment developed by Tom Angelo and Patricia Cross. Classroom assessment involves teachers investigating what and how their students are learning as teaching takes place, typically through short questions given out at the end of each class, as opposed to seeing the outcomes of student learning at the end of teaching when there is no longer the opportunity to change teaching practice. An interesting recent variant of classroom assessment is the use of electronic feedback or polling systems in the classroom which can provide immediate feedback to the lecturer as the class progresses. Combined with other teaching approaches, such as peer instruction, this can make for a very powerful learning environment focused on student learning that can be used even in large introductory undergraduate classes. A good example of this approach is the work of Eric Mazur in introductory physics classes.

Classroom assessment can be seen as a part of a growing interest in classroom research and the scholarly underpinning of teaching in higher education, reflecting Boyer’s designation of the scholarship of teaching as one of the four types of scholarship within the academy. It is still early days, but the growing interest in what is being called the scholarship of teaching and learning (SoTL) has the potential to strengthen and enhance teaching practice in universities and colleges and, while not being pedagogic research as such, offers an effective approach for improving student learning.

A Focus on Student Learning

Perhaps the most significant idea that has influenced innovation in teaching and curriculum design has emerged from the realization that teaching as an activity does not always lead to learning, or at least the learning that was intended by the teacher. As Thomas J. Shuell explained “If students are to learn desired outcomes in a reasonably effective manner, then the teachers’ fundamental task is to get students to engage in learning activities that are likely to result in their achieving those outcomes . . . It is helpful to remember that what the student does is actually more important in determining what is learned than what the teacher does” (Shuell, 1986: 429).

The implications of this proposition are significant and together with the work of educational researchers, on student approaches to learning, such as Keith Trigwell, Michael Prosser, Paul Ramsden, Noel Entwistle, Roger Säljö, and Ference Marton, underlie the major shift of focus from teaching to learning that underpins much of the innovation that has been undertaken in higher education in recent years.

A work that has been particularly influential in changing educational practice in higher education is *From Teaching to Learning - a New Paradigm for Undergraduate Education* by Robert Barr and John Tagg. In it, they describe the shift from what they call an instruction paradigm of higher education to a learning paradigm. Under the instruction paradigm, knowledge is seen as being transferred from teachers to students and there is a focus on covering content. Under the learning paradigm, the role of the teacher is to facilitate students construct their own knowledge and there is a focus on student learning and understanding.

The term active learning, above all others, best represents the type of innovations that reflect the shift in perspective from the instruction paradigm to the learning paradigm of teaching. The term, which was popularized by Charles Bonwell and James Eison, is used to describe anything that students do in the classroom other than passively listening to an instructor talk and is one of the principles in the very well-known *Principles of Good Practice in Undergraduate Education* of Arthur Chickering and Zelda Gamson.

Active learning activities are many and varied and have been implemented in a wide diversity of contexts including traditional large lecture theaters. Typical approaches include introducing into the classroom activities like writing, brainstorming, think-pair-share, problem solving, role playing, games, and debates, approaches which not only require a high level of student interaction but also provide students with immediate feedback on how effectively they are learning. The systematic use of these techniques is seen in teaching approaches such as inquiry-based learning, case-based learning, problem-based learning, and scenario-based learning. These learner-centered approaches to teaching allow students to integrate theory and practice and apply their knowledge and skills while engaging with issues and problems. While initially popular in medical curricula, inquiry-based approaches to student learning are gaining wide acceptance across the disciplines.

Student Approaches to Learning

While much of the research that has been conducted into student learning in higher education remains unknown to university and college teachers, the research on student approaches to learning has had a major impact. Studies on how students approach their learning has revealed one of two basic patterns. Students who adopt a surface approach to their learning focus only on the memorization of information, treat assessment tasks as an external imposition, and are unreflective about their purpose or strategy for learning. On the other hand, students who adopt a deep approach to learning have an intention to understand,

examine the logic of arguments, vigorously interact with the content, and relate new ideas to previous knowledge. The consequences of these two approaches are profound since a surface approach to learning is associated with a limited retention of knowledge and a failure to recognize key ideas, distinguish principles from examples, or gain an overview of a topic.

These two approaches to learning clearly describe characteristics that most teachers in higher education would wish to either enhance or reduce. However, in contrast to what is sometimes thought, surface and deep are not characteristics of students themselves but describe the way students interact with teaching in a particular context. A given student may adopt a surface approach to his/her learning in one teaching context and a deep approach in another. Significantly, research has shown which characteristics of a teaching environment are associated with students adopting one approach or another. For example, the characteristics associated with students adopting a deep approach to their learning include a strong motivational context where there is ownership of learning by students, where there is a high level of learner activity, and where there are interactions with other students through discussion. These findings can be used as the basis for developing a wide range of innovative and effective approaches to teaching.

The Curriculum in Higher Education

Depending on how the curriculum is defined, it can be seen to have a number of dimensions – not only content but also structure and planned student experiences. However, in marked contrast to teaching, the concept of curriculum, *per se*, has received relatively little attention in higher education. One of the few recent books on the curriculum in higher education by Ronald Barnett and Kelly Coate begins “Higher education is ever more important to increasing numbers of people. And yet, despite all this growth and debate, there is very little talk about the curriculum. What students should be experiencing is barely a topic for debate. What the building blocks of their courses might be and how they should be put together are even more absent from the general discussion. The very idea of the curriculum is pretty well missing altogether” (Barnett and Coate, 2004: 1).

In 1985, the Association of American Colleges report, *Integrity in the College Curriculum*, put it even more bluntly, “As for what passes as a college curriculum, almost anything goes” (Association of American Colleges, 1985: 2). The *Imaginative Curriculum Project* of the UK’s Higher Education Academy is a notable exception. This project takes a broad view of the curriculum and sees the concept as including what is to be learned, why it is to be learned, how it is to be learned, and when it is to be learned.

However, irrespective of what is included under curriculum, the most important distinction to be made is the curriculum as planned versus the curriculum as experienced by the learner.

The notion of a set of overall learning outcomes applying to the extended period of study typical of the undergraduate degree is widely held in higher education. Learning in universities and colleges has traditionally taken place over an agreed time period, typically 3 or 4 years for an undergraduate degree. Within this time period, learning occurs with a rhythm determined by the institution consisting of classes during the week, teaching weeks over a semester or term, and semesters or terms over a year. In Europe, there is currently an attempt to standardize the time periods over which learning in higher education takes place under what is known as Bologna process where the bachelor, masters and doctoral cycles are to take 3, 2, and 3 years respectively.

At the same time, we are currently seeing the emergence of very different models of teaching where learning can be compressed or fragmented and more under the control of the learner often facilitated by information and communication technologies. One can see these developments, often described as flexible learning, as a move to a model of student-centered just-in-time learning and away from the traditional just-in-case model of higher education. Even within traditional models of higher education, the increased modularization of units of study and the increasing tendency for students to combine study with work is resulting in a more fragmented and a less-integrated learning experience. The impact of these developments on the achievement of the traditional learning outcomes of higher education remains to be seen.

Curriculum Design in Higher Education

If the notion of the curriculum in higher education is problematical, this is even more the case for curriculum design where whole-degree programs and their constituent parts are planned. As the influential American Association of Colleges and Universities 2002 report *Greater Expectations* put it “The fragmentation of the curriculum into a collection of independently “owned” courses is itself an impediment to student accomplishment, because the different courses students take, even on the same campus, are not expected to engage or build on one another. Few maps exist to help students plan or integrate their learning as they move in and out of separately organized courses, programs, and campuses. In the absence of shared learning goals and clear expectations, a college degree more frequently certifies completion of disconnected fragments than of a coherent plan for student accomplishment” (American Association of Colleges and Universities, 2002: 10).

A powerful idea that is influencing teaching and curriculum design in higher education is that of constructive alignment proposed by John Biggs. The basic concept is that there should be alignment of (1) the intended learning outcomes, (2) the learning and teaching activities designed to facilitate those learning outcomes, and (3) the ways in which the learning is assessed. This might sound like common sense, but in many cases the particular pedagogy used by a teacher, such as a lecture, is inappropriate for facilitating development of the intended learning outcomes and the assessment tasks given to students do not address the learning outcomes. For example, if an intended learning outcome is related to the skills of learning to work as a member of a team, then the teaching must involve practice at working in a team and the assessment of this learning must directly involve an assessment of capacity to work in a team. Ensuring alignment between learning outcomes, pedagogy, and assessment, is emerging as one of the most effective approaches to improving learning and teaching in higher education which has traditionally put great emphasis on didactic instruction and written exams.

The increasing role of information and communication technologies in both the preparation and delivery of teaching materials has introduced an element of enforced curriculum design though not always within the context of a coherent explicit educational model.

Innovation at the Level of the Whole-Degree Program

Teaching models such as problem-based learning, case-based learning, and collaborative learning used in individual units of study can become the central feature of teaching for the whole-degree program or even institution. However, teaching innovations at the level of whole-degree program are much harder to implement than innovations at the level of the unit of study. Degree-program initiatives require considerable coordination and planning, something often hard to do given the high level of autonomy and independence that characterizes the work of most teachers in higher education who typically have a higher degree of ownership of individual subjects. Yet, it is the learning outcomes of the whole-degree program that are the most significant aspects of higher education from the perspective of the student. Alverno College’s pioneering work on integrating curriculum and assessment at the level of the whole institution is a notable exception and shows what is possible.

The learning outcomes of whole-degree programs are usually expressed as a set of graduate attributes including those relating to generic communication, critical thinking, and interpersonal skills as well as discipline-specific knowledge. However, there is often a considerable gap between the rhetoric and reality in terms of ensuring

that the graduate attributes are actually developed and assessed. In addition to teaching models *per se*, whole of degree curricula have characteristics such as depth and breadth, integration, and the level of general education and interdisciplinarity. Concerns about the lack of integration across undergraduate programs have led to initiatives such as capstone courses where students have an opportunity to integrate and apply knowledge gained over previous years of study.

Employability, and specifically the curricula and pedagogies that support the development of the knowledge, skills, and attitudes needed for employment, have proved to be powerful models for teaching and curriculum design. However, the skills and attributes considered to be needed in the modern workplace comprise a long list. Peter Knight and Mantz Yorke's book *Learning, Curriculum and Employability In Higher Education* lists a total of 39 such skills attributes that should underpin an undergraduate curriculum focused on employability:

1. *Personal qualities.* Self-awareness, self-confidence, independence, emotional intelligence, adaptability, stress tolerance, initiative, willingness to learn, and reflectiveness.
2. *Core skills.* Reading effectiveness, numeracy, information retrieval, language skills, self-management, critical analysis, creativity, listening, written communication, oral presentations, and global awareness.
3. *Process skills.* Computer literacy, commercial awareness, political sensitivity, ability to work cross-culturally, ethical sensitivity, prioritizing, planning, applying subject understanding, acting morally, coping with complexity, problem solving, influencing, arguing for and/or justifying a point of view or a course of action, resolving conflict, decision making, negotiating, and team work.

Clearly mapping of all of these to an undergraduate curriculum and ensuring they are all aligned with both pedagogy and assessment, together with required disciplinary knowledge, is a daunting task. Recent developments in this area have included introducing work-integrated learning into undergraduate degree programs, such as internships, and putting greater responsibility on students themselves to document their acquisition of their knowledge, skills, and attitudes, including extracurricular activities, through portfolios.

In contrast to programs that focus on employability, a significant curriculum innovation in undergraduate education in recent years has been the development of research-led teaching, together with similar approaches such as research-aligned teaching and research-enhanced teaching. These models attempt to create a strong nexus between teaching and research in undergraduate teaching by introducing students not only to current research in the discipline but also ensuring their learning is aligned with the methods of inquiry and research characteristic of

the discipline. This approach to the curriculum has been taken up, in particular, by research-intensive universities and is consistent with the recommendations of the well-known 1998 report of the Boyer Commission *Reinventing Undergraduate Education: a Blueprint for America's Research Universities* and can involve, in some teaching models, undergraduate students engaging in real research themselves. Engaging with real research problems using disciplinary-based inquiry is not only an effective way for students to learn the inquiry skills of their disciplinary area, but is also highly motivational.

Innovations Aimed at Exploiting the Potential of New Information and Communication Technologies

New information and communication technologies have been the focus of much of what is viewed as innovation in higher education in recent years. Developments such as the Internet are providing access to higher education to students who would otherwise be unable to attend universities and colleges in ways that have transcended the traditional models of distance education. However, the new technologies, in particular the capacity to store, access, and share vast quantities of information, are also transforming what it means to learn and to teach in higher education. Universities and colleges no longer have a monopoly on the possession and distribution of information and this presents a major challenge to the authority of knowledge and the concept of authorship. Moreover, mobile technologies mean students can access information from just about anywhere, which is challenging the traditional role of the classroom and lecture hall and is giving a new meaning to the concept of internationalization.

Aside from the issue of information access, information and communication technologies are impacting on the process of learning itself, giving rise to the concept of e-learning. The ability of students to customize the information they receive, process it in a way and at a speed they themselves control, and the ability to review it when they want, stands in marked contrast to the one-size-fits-all approach of traditional higher education where students are forced to learn at the pace and style of the particular teacher they have. Moreover, technologies are making it possible for students to explore their ideas through dynamic simulations and receive immediate feedback on their learning. Many of the different technologies that facilitate learning are being integrated into commercial learning-management systems that have become the foundation for many university and college teaching programs.

Information and communication technologies such as the Internet are relatively recent in terms of their application to higher education and there is a tendency to equate any use of technology in learning and teaching as

an innovation. However, the effectiveness of many of these approaches to teaching remains to be determined and indeed there is a considerable debate about how such effectiveness can or should be determined. Some concerns center on the lack of face-to-face interactions between both teachers and their students and among students. Recent years have seen a growing interest in online collaboration and social-networking tools and Web 2.0 technologies such as blogs and wikis offer the potential for a high level of interactivity and a rich learning experience. It must be remembered that students now entering higher education, the so-called digital natives, see the Internet in all of its manifestations as an integral part of their lives. Nevertheless, for many students and teachers, face-to-face interactions remain an important, if not the central, feature of the higher education experience and recent years have seen the emergence of the concept of blended learning which combines features of both face-to-face and online learning.

Bibliography

- American Association of Colleges and Universities (2002). *Greater Expectations: A New Vision for Learning as a Nation Goes to College*. Washington, DC: American Association of Colleges and Universities.
- Association of American Colleges (1985). *Integrity in the College Curriculum*. Washington, DC: Association of American Colleges.
- Barnett, R. and Coate, K. (2004). *Engaging the Curriculum in Higher Education*. Buckingham: SRHE/Open University Press.
- Department of Education, Science and Training (2002). *Higher Education at the Crossroads: An Overview Paper*. Canberra: Commonwealth of Australia.
- Hannan, A. and Silver, H. (2000). *Innovating in Higher Education: Teaching, Learning and Institutional Cultures*. Buckingham: SRHE/Open University Press.
- Shuell, T. J. (1986). Cognitive conceptions of learning. *Review of Educational Research* **56**, 411–436.
- Spellings, M. (2006). *A Test of Leadership: Charting the Future of U.S. Higher Education*. Jessup, MD: Education Publications Centre, U.S. Department of Education.

Further Reading

- Angelo, T. A. and Cross, K. P. (1993). *Classroom Assessment Techniques: A Handbook for College Teachers*, 2nd edn. San Francisco, CA: Jossey-Bass.
- Ashwin, P. (ed.) (2006). *Changing Higher Education: The Development of Learning and Teaching*. New York: Routledge.
- Barkley, E., Cross, K. P., and Major, C. H. (2004). *Collaborative Learning Techniques: A Handbook for College Faculty*. San Francisco, CA: Jossey-Bass.
- Barr, R. B. and Tagg, J. (1995). From teaching to learning – a new paradigm for undergraduate education. *Change* **November/December**, 13–24.
- Bates, A. W. and Poole, G. (2003). *Effective Teaching with Technology in Higher Education: Foundations for Success*. San Francisco, CA: Jossey-Bass.
- Biggs, J. and Tang, C. (2007). *Teaching for Quality Learning at University*, 3rd edn. Buckingham: SRHE/Open University Press.

- Bonwell, C. and Eison, J. (1991). *Active learning. ASHE-ERIC Higher Education Report, No. 1*. Washington, DC: George Washington University.
- Boud, D. and Solomon, N. (eds.) (2001). *Work-Based Learning: A New Higher Education?* Buckingham: SRHE/Open University Press.
- Boyer Commission on Educating Undergraduates in the Research University (1998). *Reinventing Undergraduate Education: A Blueprint for America's Research Universities*. Washington, DC: Carnegie Foundation for the Advancement of Teaching.
- Chickering, A. W. and Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *American Association for Higher Education Bulletin* **39**(7), 3–7.
- Cowan, J. (2006). *On Becoming an Innovative University Teacher: Reflection in Action*, 2nd edn. Buckingham: SRHE/Open University Press.
- Diamond, R. M. (2008). *Designing and Assessing Courses and Curricula: A Practical Guide*, 3rd edn. San Francisco, CA: Jossey-Bass.
- Fink, L. D. (2003). *Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses*. San Francisco, CA: Jossey-Bass.
- Gaff, J. G. and Ratcliff, J. L. (1997). *Handbook of the Undergraduate Curriculum – A Comprehensive Guide to Purpose, Structures, Practices, and Changes*. San Francisco, CA: Jossey-Bass.
- Garrison, D. R. and Vaughan, N. D. (2007). *Blended Learning in Higher Education: Framework, Principles, and Guidelines*. San Francisco, CA: Jossey-Bass.
- Jaques, D. and Salmon, G. (2006). *Learning in Groups: A Handbook for Face-to-Face and Online Environments*. Abingdon: Routledge.
- Jenkins, A., Breen, R., and Lindsay, R. (2003). *Reshaping Teaching in Higher Education. Linking Teaching with Research*. London: RoutledgeFalmer.
- Knight, P. and Yorke, M. (2004). *Learning, Curriculum, and Employability in Higher Education*. New York: RoutledgeFalmer.
- Mazur, E. (1997). *Peer Instruction: A User's Manual*. Saddle River, NJ: Prentice-Hall.
- Mentkowski, M. and associates (2000). *Learning that Lasts: Integrating Learning, Development, and Performance in College and Beyond*. San Francisco, CA: Jossey-Bass.
- Pratt, D. D. and associates (1998). *Five Perspectives on Teaching in Adult and Higher Education*. Malabar, FL: Krieger.
- Prosser, M. and Trigwell, K. (1999). *Understanding Learning and Teaching: The Experience in Higher Education*. Buckingham: SRHE/Open University Press.
- Tagg, J. (2003). *The Learning Paradigm College*. Bolton, MA: Anker Publishing.
- Walker, M. (2005). *Higher Education Pedagogies*. Buckingham: SRHE/Open University Press.
- Weimer, M. (2002). *Learner-Centered Teaching: Five Key Changes to Practice*. San Francisco, CA: Jossey-Bass.

Relevant Websites

- <http://www.aishe.org> – All Ireland Society for Higher Education.
- <http://www.altc.edu.au> – Australian Learning and Teaching Council.
- <http://www.carnegiefoundation.org> – Carnegie Foundation for the Advancement of Teaching (USA).
- <http://www.educause.edu> – Educause.
- <http://www.heacademy.ac.uk> – Higher Education Academy (UK).
- <http://www.hersa.org.au> – Higher Education Research and Development Society of Australasia.
- <http://www.issotl.org> – International Society for the Scholarship of Teaching & Learning (ISSOTL).
- <http://www.mcmaster.ca> – Macmaster University, Canada.
- <http://www.seda.ac.uk> – Staff and Educational Development Association (UK).

Productivity of University Faculty Staff

S Kyvik, NIFU STEP, Oslo, Norway

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Publishing of papers and books is the principal means of communication of new ideas and results of scientific and scholarly enquiry. It is through the exchange of published knowledge that science makes progress. However, publication is also the principal means of career advancement and scientific reputation for individual university staff, and, increasingly, a means for funding of new research projects and research environments. In the competition for research funding and individual rewards, the number of publications has become an important parameter of scientific productivity, quality, and success. Through the introduction of performance-based funding of universities and university departments, being a productive researcher is no longer only an individual matter, but also a matter for the collectivity of staff members.

Over time, three trends in publication patterns stand out as particularly visible: the number of publications per academic staff member has increased, co-authorship has become much more common, and, worldwide, the scientific article in an English-language journal has enhanced its position as the dominating type of publication.

The increasing publication pressure has led to a significant rise in the number of published papers and books by individual faculty members. This is not only due to more productive researchers, but probably also due to a tendency to fragmentation of data presentation in several shorter, but strongly related papers, and to the publishing of slightly different versions of basically the same paper. The frequent complaints about these side effects from journal editors and referees have not had much impact on the publication system. The number of scientific journals is constantly growing, publishers invent new book-series to earn more money, and report-series are established by universities themselves, so that virtually every manuscript eventually might be published somewhere.

The large increase in the rate of co-authored publications as well as in the number of co-authors is related to fundamental changes in the science system itself as well as to changes in the steering and funding of research. The augmentation of joint authorship of scientific papers is a reflection of the trend toward collaboration in research; a development that can be traced back to the beginning of the twentieth century in the natural sciences, and which later accelerated in those disciplines where research projects are dependent on contributions from different researchers with complementary knowledge and skills to

make progress. The trend toward more co-authorship is also determined by another aspect of the organization of research. Over the last decades, more and more weight has been put on research programs, initiated by national and cross-national authorities and research councils, and in all fields of science. This way of organizing research usually presupposes collaboration between scientists that, in turn, leads to more co-authored papers by an increasing number of authors per paper. Finally, the increase in co-authorship is related to the pressure on universities and individual staff members to document results in the form of published output. In the competition for research, funding, and individual rewards, the number of publications has become an important parameter of success. The social importance of becoming a co-author of scientific papers has subsequently increased.

Measurement of Publication Productivity

The number of journal articles is the most widely used measure of publishing productivity. The reason is simple: journal articles are indexed in various databases and are relatively easily accessible for measurement or assessment purposes. Moreover, in most disciplines, they are presumed to reflect the production of new knowledge or research productivity. There are, however, four major problems related to this measure.

First, each article does not represent an equal contribution in quantitative terms. The information contents of journal articles vary enormously. This is not so much due to variation in the number of pages or words, as to differences in the density of information (Seglen, 1996).

Second, few databases include all relevant scientific and scholarly journals. In those databases that aim at international coverage, there is an English-language bias in the journals that are indexed, and a considerable number of national journals as well as newly established journals are not included.

Third, the author list of most journal articles counts several names. It is not evident how this problem should be dealt with. The main problem is how much credit each co-author should be attributed. The various authors are seldom equal contributors to an article. On the one hand, plain paper counts obviously favor those who have their names attached to many articles. On the other hand, attributing fractional counts to each author according to the number of authors might disfavor those contributors

that are first authors. One way of dealing with this problem has been to give credit to first authors only. However, customary practices within or across different disciplines as to the ordering of names among joint authors do not exist. In some articles, the first author is the primary author; in others, the authors are listed alphabetically; still others use different conventions.

Fourth, the publication pattern differs considerably between fields. The number of publications per academic staff member is higher in the natural and medical sciences and technology than in the social sciences and the humanities. This is partly due to the concentration on article production in the former fields and the larger emphasis on books in the latter fields, and partly due to less frequent co-authorships in the social sciences and in particular in the humanities. In order to deal with the problems of skewed coverage of publications in the journal-indexing databases, simple publication counts, or more refined weighting procedures have been developed to account for different types of publications. Scanning the research literature in the field reveals however that no agreement exists on the appropriate weighting that should be assigned to various types of publications.

In spite of these methodological problems, publication counts make sense as a rough measure of the output of scientific enquiry. Publication productivity would, however, be a dubious measure on scientific output if high productivity could be proven to correlate with low quality and vice versa. There is, however, general agreement among scholars that if one examines larger groups of scientists as opposed to individuals, there is a positive – albeit not perfect – correlation between high numbers of publications and the significance and impact of the authors as deemed by their colleagues. The conclusion is, therefore, that the more productive researchers will, in general, contribute more to the stock of knowledge than the less productive scientists. This tendency can, moreover, be explained theoretically; those scientists who receive recognition for their publications are encouraged to continue to publish, while those whose work is not regarded as important are likely to become discouraged and less productive. Nevertheless, there are differences between academic staff members; some manage to publish a lot of low-quality work, while others publish few, but excellent, papers or books. Productivity measures are accordingly insufficient means to gauge the scientific standing of individual scientists and scholars.

Productivity Differences Between Faculty Members

There are large individual productivity differences among faculty members in all university systems. A relatively

small percentage of staff accounts for a disproportional large share of the published output. During the time since Lotka (1926) formulated his inverse square law on productivity, which says that the number of scientists producing n papers is proportional to $1/n^2$, several attempts have been made to test the generality of this law. According to Lotka, about one-half of all papers are produced by 6% of all researchers and nearly one-quarter by 1%. This law obviously overestimates the number of papers produced by the most prolific researchers. Nevertheless, many studies have confirmed that highly skewed patterns of productivity exist in scientific publishing. This skewed pattern is shown to apply to all university disciplines (Kyvik, 1991), and seems to be insensitive to a variety of time frames (Pao, 1985). A study of tenured faculty members in research universities reports that the most prolific 20% of the staff produced 50% of the total number of publications, and that this pattern was stable over a 20-year period (Kyvik, 2003). This skewed distribution applied to the natural and medical sciences, technology, the social sciences, and the humanities, indicating a remarkable regularity in individual productivity differences independent of time and discipline.

Explanations for Productivity Differences

Many explanations have been put forward to explain the large differences in publishing activity among academic staff members, of which the following are the most common.

Abilities

There seem to be only few studies on the relationship between individual abilities and scientific productivity, and these inquiries indicate a very weak, if any, positive correlation. The conclusion which may be drawn on the basis of present knowledge is that high intelligence quotient (IQ) may be a prerequisite for becoming a scientist, but that differences in measured ability within this group of people do not determine subsequent levels of performance.

Time and Energy

A number of studies report only a weak correlation between time spent on research and scholarly output. However, the splitting up of a work day such that it is difficult to have uninterrupted time for research appears to be a widespread problem among tenured faculty. Some individuals appear to cope with this more easily than others. Differences in the organization of a working day and the ability to concentrate on research in spite of interruptions can, therefore, be just as important as the total time spent on research.

Resources

Research is often expensive with regard to equipment, materials, travels, and assistance. Those scientists who have large resources at their disposal will be in a better position to be productive researchers than those who have scarce resources. A number of studies report a positive, although relatively weak, significant correlation between financial support and published output.

Communication

It seems to be widely accepted that having close contacts with other researchers stimulates productivity. There are informal networks, or so-called invisible colleges that have the leading scientists within disciplinary specialties as members, and where information is exchanged, people are invited to take part in collaborative projects, and invited to submit papers for publication as book chapters or journal articles. Communication may create intellectual synergy between individuals through the cross-fertilization of ideas. This process may, in turn, generate new insights or perspectives that individual researchers – working on their own – would not have grasped. There is some scattered evidence that networking with colleagues enhances publication productivity.

Collaboration

It is a well-known fact that collaboration in research has become gradually more common as evidenced by the increase in co-authored publications. The reasons for the enhancement of research collaboration are a mixture of related factors; internal scientific reasons, science-policy initiatives, individual motivations, and improvements in communication facilities. Many scholars and policymakers seem to assume that research collaboration enhances publication productivity, because division of labor and the joining of specialized skills may enable researchers to increase their efficiency. In addition, much collaborative research of academic staff is done as supervisors of doctoral students and materializes in co-authorship. On the other hand, there are transaction costs associated with working with others, and, for a variety of reasons, collaboration does not always live up to expectations (Lee and Bozeman, 2005). Despite this, there is some evidence suggesting that collaboration in itself enhances productivity in research, even when corrections are made for the number of co-authors.

Environmental Location

Even though we know relatively little about under what circumstances organizational variables will affect research productivity, there is still some scattered evidence that organizational context influences research performance. The average quality of academic staff in a university

department, and values and norms related to the level of publishing activity seem to affect an individual staff member. There is some evidence that academic staff who change institutional location after some time conform with the characteristics of that context, independently of previous productivity. On the other hand, the effect of group size and department size on publication productivity is unclear. Different studies report various tendencies, indicating that size in itself has no significant effect on publishing productivity.

Cumulative Advantage and Disadvantage

These various explanations of productivity differences are not mutually exclusive but have been integrated in the theory on cumulative advantage and disadvantage in science. According to this theory, the skewed pattern of publication productivity is the result of the way the scientific reward system operates (Merton, 1968; Cole and Cole, 1973). Minor differences between scientists early in their careers accumulate to produce substantial differences in productivity over time. This process can be viewed as consisting of two feedback loops in which recognition and resources are intervening variables (Allison and Stewart, 1974). Scientists who have been recognized as having made significant advances will be motivated to maintain or increase their recognition by additional publications, and will be influenced by their colleagues' expectations that they repeat or exceed those achievements. Beyond these direct effects, recognition usually implies increased access to resources which facilitate research. In contrast to this picture of spiraling success, the scientist who publishes little or whose work is not recognized is likely to become discouraged with research.

The cumulative advantage theory has been widely accepted as the best explanation of why there are such large productivity differences among individual scientists. The theory has, however, obvious weaknesses (Kyvik, 1991). One objection is that accumulative advantage (resources, rewards, etc.) will not necessarily lead to higher productivity. For some academic staff, the utility of using more time on academic administration or professional work will be regarded as more advantageous for their future career prospects than persistent efforts in research. Another objection is that accumulative disadvantage will not necessarily lead to lower productivity. Some faculty members find scientific work personally enjoyable and continue publishing even though their publications are not recognized as worthwhile by their colleagues. For others, publishing papers is a self-legitimizing act as researchers, even in the absence of rewards. Nonetheless, the cumulative advantage and disadvantage theory is attractive as an explanation of the large differences in productivity between scientists.

Age and Productivity

Many studies report that, in general, publication activity declines with increasing age and that old academic staff are less productive than their middle-aged colleagues. There appears to be a curvilinear relationship between age and number of publications. Productivity reaches a peak when scientists are in their 1940s, depending on scientific field, and then declines. We can distinguish between six different hypotheses which all suggest reduced publishing activity with an aging academics staff (Kyvik and Olsen, 2008).

The Utility-Maximizing Hypothesis

According to this hypothesis, academic staff will pay less attention to research with increasing age because the expected utility of time dedicated to research diminishes (Diamond, 1984). Investing time in other university tasks – for example, administration – or in work external to universities can be seen as more effective ways of increasing one's prestige and/or income.

The Seniority Burden Hypothesis

Another possible explanation for the decline in research productivity with increasing age is that the more experienced academic staff are, the more duties they are expected to engage in – administrative work, evaluation tasks, supervision of postgraduate students, professional work in relation to the discipline and the community, etc. Seniority and experience bring about an accumulation of tasks and duties that reduce the time available for research (Zuckerman and Merton, 1972).

The Cumulative Disadvantage Hypothesis

According to this explanation, the average publishing productivity of a cohort declines with age because those who are not rewarded for their research gradually lose the motivation for new achievements (Cole, 1979a). As they continue to publish, some find their work rewarded and go on to be even more productive. In addition, professional recognition leads to greater possibilities for access to economic resources, research assistants, and membership in international research networks – all of which have positive effects on productivity. Others, who are treated indifferently by their colleagues, and who receive negative responses to their work, may gradually lose their motivation and the possibilities to keep their productivity at a high level. The older a cohort of researchers grows, the less motivated an increasing share of the cohort will be to keep up high publication rates.

The Age Decrement Hypothesis

Another hypothesis is that older scientists, on average, function on a lower intellectual and physical level than their younger colleagues. While there is probably very little age decrement up to the age of 60 in intellectual abilities and skills that do not require fast responses or are not affected by reaction time (Schaie, 1975), older scientists are less creative than younger researchers, and they are also exposed to age-related physical problems and illnesses which may hamper or slow down their engagement in research.

The Obsolescence Hypothesis

According to this explanation, the speed of the introduction of new research technology has had different consequences for the younger and older generations. While the oldest cohorts of scientists were trained in methods and equipment neither sufficient nor relevant for new scientific breakthroughs, the younger generations have access to the latest theory, methodology, and equipment in their doctoral work. Thus, the skills of old scientists do not necessarily decline, but rather become obsolete (Price *et al.*, 1975).

The Intellectual Deadlock Hypothesis

A final assumption is that older academic staff are not able or willing to reorient their research toward new scientific or social problems. There is a tendency among scientists to pursue research problems which were important in the early days of their career, making their research irrelevant to new generations of academics who possess a fresh point of view on how to solve a problem as well as on what problems ought to be studied (Stephan and Levin, 1992).

There are, however, several methodological problems related to the study of age and productivity. In general, it is extremely difficult to measure the impact of aging on human behavior because observed results can be influenced by three different effects (Riley *et al.*, 1972): First, by pure age effects – that is, of physical and mental changes related to the aging of the body and the mind. Second, by generation effects – that is, that different age groups have had different life experiences, and the reason why people of different ages differ on a given characteristic may be due to the fact that they belong to different generations, rather than because they differ in age. Third, by period effects – that is, that norms for appropriate behavior might have changed over time, and resources available for job performance may be quite different at different time periods. In practice, these effects intermingle, making it difficult to disentangle the impact of aging from generational and situational effects.

Gender and Productivity

Virtually all studies that have examined scientific publishing among men and women scientists have found that men are more productive than women. Cole and Zuckerman (1984) refer to more than 50 studies in various fields which show that men on average publish 40–50% more papers than their female counterparts. Even though a number of studies have attempted to explain the large differences in scientific productivity between male and female researchers, there is disagreement about the reasons for this difference. Various explanations have been proposed to account for this difference. One important reason is differences in academic rank. Full professors are more productive than associate professors, who in turn are more productive than assistant professors. As there are relatively fewer women in senior positions, the difference in productivity between ranks has consequences for the average productivity of male and female staff. However, even when controls are made for rank, productivity differences persist. Thus, it has been claimed that women are less integrated into professional networks than are men, and that they are accordingly less motivated or enjoy fewer opportunities to be productive. It has been contended that women have less academic confidence than men, and that they have poorer access to economic resources for doing research. Some argue that women use less time for research than do men, either because they are less motivated to spend most of their time on research, or because they are not able to use as much time for research as are men due to stronger commitments in teaching, supervision, and administrative work – duties that have clear parallels with women's roles as caretakers. In addition, women often interrupt their research careers because of childbirth and have more caring responsibilities than do men. These different explanations have also been integrated into the cumulative advantage and disadvantage theory. If women do less well than men on a number of such indicators, productivity differences are likely to reflect different cumulative effects for men and women over time.

The two most commonly used explanations for gender differences in publication productivity when controlled for academic position are that women are not as well integrated in scientific networks and collaborative research as are men, and that women are hampered by childbirth and child-care.

Networking, Collaboration, and Productivity

The probably most widely used explanation for gender differences in productivity is that women in male-dominated universities have problems in integrating into important informal networks (Reskin, 1978; Cole, 1979b;

Fox, 1991). This is thought to be the case in the department as well as in relation to the research community as a whole. The reasons are that women are a minority, and that men communicate best with men. Many men might also avoid working closely with women – either because they or their spouses fear that a sexual relationship could develop, or because they want to avoid the gossip of their colleagues. Several studies show that researchers who have wide professional contacts with colleagues publish more than those who do not. There also is some evidence that teamwork in itself enhances productivity in research. If women are excluded from important networks so that their opportunities for collaboration in research are thus restricted, their publishing activity may be adversely affected. The research literature has not, however, produced consistent evidence that women are less integrated in informal professional networks or in collaborative research than men.

Motherhood and Productivity

The other principal explanation for gender differences in publication productivity is that women have interruptions in their research career due to childbirth and also have more caring responsibilities than do men. Various studies report inconclusive results. Some studies indicate that women with children publish more than do childless women and thus suggest that women's lower productivity cannot be due to maternal responsibilities (Cole, 1979b; Cole and Zuckerman, 1984). Other studies show that when children's age is introduced as a variable, maternity leave and child-care turn out to be important determinants for women's publishing activity. One study reports that while women tenured faculty with children under age 6 were almost 60% less productive over a 3-year period than were their male counterparts, and women with children in the age range 6–10 were 30% less productive, there were virtually no differences between men and women whose children were all older than 10 years (Kyvik and Teigen, 1996). This indicates that maternity leave greatly reduces women's scientific productivity temporarily, but child-care in itself also affects the publishing activity of women. Other studies report that gender, family characteristics, and publication productivity are complex relations that go beyond being married or not married, and the presence or absence of children (Fox, 2005).

Conclusion

Research done on scientific and scholarly publishing indicates that productivity patterns among university staff are remarkably similar across different countries and time-periods. There are large individual differences; a relatively small share of a given researcher population

accounts for the majority of publications. Men publish on average more than women, there are differences between various age groups, and the higher the rank, the more productive is the average faculty member. No single explanation can account for these differences. Even the theory of cumulative advantage and disadvantage – which aims at integrating the various inequality-causing processes into a coherent explanation – falls short in explaining the skewed productivity patterns, although it is a fruitful theoretical abstraction.

Bibliography

- Allison, P. D. and Stewart, J. A. (1974). Productivity differences among scientists: Evidence for accumulative advantage. *American Sociological Review* **39**, 596–606.
- Cole, S. (1979a). Age and scientific performance. *American Journal of Sociology* **84**, 958–977.
- Cole, J. (1979b). *Fair Science. Women in the Scientific Community*. New York: The Free Press.
- Cole, J. and Cole, S. (1973). *Social Stratification in Science*. Chicago, IL: University of Chicago Press.
- Cole, J. and Zuckerman, H. (1984). The productivity puzzle: Persistence and change in patterns of publication of men and women scientists. In Maehr, P. and Steinkamp, M. W. (eds.) *Advances in Motivation and Achievement*, vol. 2, pp 217–258. Greenwich, CT: JAI.
- Diamond, A. M. (1984). An economic model of the life-cycle research productivity of scientists. *Scientometrics* **6**, 189–196.
- Fox, M. F. (1991). Gender, environmental milieu and productivity in science. In Zuckerman, H., Cole, J., and Bruer, J. (eds.) *The Outer Circle: Women in the Scientific Community*, pp. 188–204. New York: Norton.
- Fox, M. F. (2005). Gender, family characteristics, and publication productivity among scientists. *Social Studies of Science* **35**, 131–150.
- Kyvik, S. (1991). *Productivity in Academia*. Oslo: Norwegian University Press.
- Kyvik, S. (2003). Changing trends in publishing behaviour among university faculty, 1980–2000. *Scientometrics* **58**, 35–48.
- Kyvik, S. and Olsen, T. B. (2008). Does the aging of tenured academic staff affect the research performance of universities? *Scientometrics* **76**, 439–455.
- Kyvik, S. and Teigen, M. (1996). Child care, research collaboration, and gender differences in scientific productivity. *Science, Technology, and Human Values* **21**, 54–71.
- Lee, S. and Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science* **35**, 673–702.
- Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences* **16**, 317–323.
- Merton, R. (1968). The Matthew effect in science. *Science* **159**, 56–63.
- Pao, M. L. (1985). Lotka's law: A testing procedure. *Information Processing and Management* **21**, 305–320.
- Price, R. L., Thompson, P. H., and Dalton, G. W. (1975). A longitudinal study of technological obsolescence. *Research Management* **18**, 22–28.
- Reskin, B. (1978). Scientific productivity, sex, and location in the institution of science. *American Journal of Sociology* **83**, 1235–1243.
- Riley, M. W., Johnson, M., and Foner, A. (1972). *Aging and Society – Vol. 3: A Sociology of Age Stratification*. New York: Russel Sage Foundation.
- Schaie, K. W. (1975). Age changes in adult intelligence. In Woodruff, D. S. and Birren, J. E. (eds.) *Aging, Scientific Perspectives and Social Issues*, pp. 111–124. New York: D. Van Nostrand.
- Seglen, P. (1996). Quantification of scientific article contents. *Scientometrics* **35**, 355–366.
- Stephan, P. and Levin, S. (1992). *Striking the Mother Lode in Science. The Importance of Age, Place, and Time*. New York: Oxford University Press.
- Zuckerman, H. and Merton, R. K. (1972). Age, aging, and age structure in science. In Riley, M. W., Johnson, M., and Foner, A. (eds.) *Aging and Society – Vol. 3: A Sociology of Age Stratification*, pp. 292–356. New York: Russel Sage Foundation.

Further Reading

- Blackburn, R. T. and Lawrence, J. H. (1995). *Faculty at Work. Motivation, Expectation, Satisfaction*. Baltimore: The Johns Hopkins University Press.
- Creamer, E. G. (1998). Assessing faculty publication productivity. *ASHE-ERIC Higher Education Report*, vol. 26. No. 2. Washington, DC: The George Washington University.
- Fox, M. F. (1983). Publication productivity among scientists. A critical review. *Social Studies of Science* **13**, 285–305.
- Moed, H. F., Glänzel, W., and Schmoch, U. (eds.) *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of Science & Technology Systems*. Kluwer: Dordrecht.
- Ward, K. B. and Grant, L. (1996). Gender and academic publishing. In Smart, J. (ed.) *Higher Education: Handbook of Theory and Research*, vol. 11, pp 172–212. New York: Agathon Press.

Promoting Effective Student Learning in Higher Education

K Trigwell, University of Sydney, Sydney, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Conceptions of learning – Students ways of thinking about what learning is – e.g. memorising or developing personal meaning.

Deep approach to learning – Students main intention in learning is to develop personal understanding.

Effect size – A number between 0.0 and 1.0 used to gauge effectiveness, with numbers around 0.2 being small, 0.5 being medium and 0.8 being large.

Interventions – Introduction of education practises designed to improve teaching/learning.

Learning environment/climate – The context in which learning is experienced.

Phenomenography – A way of mapping variation in ways of experiencing a phenomenon.

Self-efficacy – The degree of confidence students have in being able to tackle requirements.

Self regulation – Students ability to monitor and adapt their learning.

Surface approach to learning – Students main intention in learning is to satisfy assessment requirements.

Threshold concepts – Those areas of subject content that, having been mastered, open the field to greater understanding.

Effective student learning is learning that prepares students for the present as well as for an unknown future. While it must include passing courses and the acquisition of skills and key concepts, it needs to be more than that. In a major UK study on student engagement and high-quality learning, the Enhancing Teaching–Learning (ETL) Environments research team describes desired student outcomes more in terms of ways of thinking and practicing in the discipline or professional area (Hounsell *et al.*, 2005). Ways of thinking and practicing “capture the richness, depth and breadth of what the students could learn through engagement with a given discipline or subject area. It rests on a secure foundation of subject knowledge and understanding and could also encompass subject-related skills, conventions and practices for communicating within the subject, and salient values and attitudes.”

Research on what promotes effective learning in higher education has, for the last 30 years, been dominated by

two quite different schools of thought. One has its origins in Europe, is now known as the Student Approaches to Learning (SAL) perspective, and is focused on seeing learning from the point of view of the experience of the students (Marton *et al.*, 1997). From this perspective, promoting effective learning is about helping students experience a supportive learning environment. The other has its origins in the USA, is now known as the Self-Regulated Learning (SRL) perspective and has developed principally from top-down psychological conceptualizations (Pintrich, 2004). From this perspective, promoting effective learning is about changing students’ ability to teach themselves.

A recent variant on the two dominant research conceptualizations is the work on observations from the widely adopted National Survey of Student Engagement (NSSE) in North America (e.g., Kuh, *et al.*, 2005). In focusing on broad environmental factors associated with student engagement these studies complement the two research areas outlined above. Finally, while the student is at the center of these three areas, teaching and the teacher are seen to play a crucial role in promoting effective learning in all of them, including the minimizing of barriers to learning.

In the notes that follow, the key elements from each of four research fields (SRL, SAL, student engagement, and higher education teaching) are presented. There is overlap between the fields. Indeed they share two concepts considered to be fundamental in effective learning. The first is the assumption that students construct their own knowledge and that this construction may not always be consistent with teachers’ ways of knowing. Therefore, effective learning is more likely when students are provided with the opportunity to be active learners, and both students and teachers are able to monitor learning outcomes through the learning process. The second is that students’ prior academic learning is one of the strongest predictors of academic achievement. Students who have had previous success tend to continue to have success. Promoting effective learning means being aware of students’ prior learning. Despite these commonalities, each of the four fields also address a different aspect of what is a very complex education system, and an understanding of what promotes effective learning in higher education would not approach completeness without the inclusion of all four.

One area of research that is considered to be important in understanding student learning has not been included in this entry. Students’ personal and contextual characteristics

are related to academic achievement, but they are not normally seen as amenable to much change after late adolescence.

SAL Perspective: Promoting Effective Learning Through Changes to Students' Perceptions of Their Learning Environment

From the relational perspective of this research, students adopt qualitatively different approaches to learning according to their perceptions of their learning situation. This situation is a relationship between the learning context and the student (Marton *et al.*, 1997). An approach, either deep or surface, is not fully either a characteristic of the student or the context. A student's approach to learning can change with their changing perceptions of their learning situation (Ramsden, 2003).

In adopting a deep approach to learning, the students' intention is to understand ideas and seek meanings. They have an intrinsic interest in the task and an expectation of enjoyment in carrying it out. They adopt strategies that help satisfy their curiosity, such as making the task coherent with their own experience, relating and distinguishing evidence and argument, looking for patterns and underlying principles, integrating the task with existing awareness, seeing the parts of a task as making up the whole, theorizing about it, forming hypotheses, and relating understanding from other parts of the same subject, and from different subjects. Overall, they have a focus on the meaning in the argument, the message, or the relationships but they are aware that the meanings are carried by the words, or the text, or the formulae (Prosser and Trigwell, 1999).

In adopting a surface approach to learning, students see tasks as external impositions and they have the intention to cope with these requirements. They are instrumentally or pragmatically motivated and seek to meet the demands of the task with minimum effort. They adopt strategies which include: a focus on unrelated parts of the task, separate treatment of related parts (e.g., principles and examples), a focus on what are seen as essentials (factual data and their symbolic representations), the reproduction of the essentials as accurately as possible, and rote memorizing information for assessment purposes rather than for understanding. Overall, they would appear to be involved in study without reflection on purpose or strategy, with the focus of that study being on the words, the text, or the formulae (Prosser and Trigwell, 1999).

Students who adopt a deep approach to learning are more likely to engage in learning that lasts than students who adopt a surface approach to learning. The early research studies showing that students who adopt a deep approach to learning are more likely to have a higher-quality learning outcome than those adopting more

surface approaches (Marton *et al.*, 1997) has been repeated in numerous contexts. Deep approaches to learning are effective learning processes.

Promotion of effective learning, from this relational perspective, is about establishing a learning environment perceived by students to afford more of a deep approach to learning and less of a surface approach to learning. In studies of the relations between perceptions of the learning environment and approaches to learning, five factors associated with the variation in approach have been identified:

1. teaching that fosters active and long-term engagement with learning tasks;
2. academic goals and standards that are clear to students;
3. assessment that discourages recall or the application of trivial procedural knowledge;
4. workloads perceived by students to be appropriate (not so heavy that in-depth exploration is stifled, but not so light that the work is not challenging); and
5. choice of learning tasks.

To assist curriculum designers and course administrators to continuously improve the learning context, a quantitative instrument (Course Experience Questionnaire, Ramsden, 2003) has been developed to gauge students' perceptions of their learning context.

In more recent studies involving analyses using structural equation modeling, and in the ETL project (Hounsell *et al.*, 2005), the associations between these environmental factors and approaches to learning have been confirmed. Several new factors have also emerged. For example, the drop in frequency of small-group meetings brought on by financial savings and increased participation in higher education has affected students' approaches to learning. In the current context, supporting individual learning and studying may enhance effective learning. While feedback to students on their progress was originally seen as being a part of a good teaching environment, there is a growing literature that suggests this as a major factor in its own right, and that immediate feedback, given in a personalized way, such as through discussions, can assist in promoting effective learning.

In designing supportive teaching–learning environments, however, it is also crucial to see these as interacting systems, in which even one element that interferes with high-quality learning can have damaging effects.

SRL Perspective: Promoting Effective Learning Through Changes to Students' Ability to Teach Themselves

A central tenet of the SRL perspective is that learners can potentially monitor, control, and regulate certain aspects

of their own cognition, motivation, and behavior as well as some features of their environment (Pintrich, 2004). Students who do self-regulate their learning are more effective learners than those who do not. The characteristics of those who self-regulate their learning include:

1. A familiarity with and knowledge of how to use a series of cognitive strategies (e.g., repetition, elaboration, and organization) which help them to attend to, transform, organize, elaborate, and recover information.
2. A knowledge of how to plan, control, and direct their mental processes toward the achievement of personal goals (meta-cognition).
3. Evidence of a set of motivational beliefs and adaptive emotions, such as a high sense of academic self-efficacy, the adoption of learning goals, the development of positive emotions toward tasks (e.g., joy and enthusiasm), as well as the capacity to control and modify these, adjusting them to the requirements of the task and of the specific learning situation.
4. Ability to plan and control the time and effort used on tasks and knowledge of how to create and structure favorable learning environments, such as finding a suitable place to study and help-seeking from teachers and class mates when they have difficulties.
5. Evidence of greater efforts to participate in the control and regulation of academic tasks, classroom climate, and structure (e.g., how one will be evaluated, task requirements, the design of class assignments, and organization of work teams) when the context allows it.
6. Ability to put into play a series of volitional strategies, aimed at avoiding external and internal distractions, in order to maintain their concentration, effort, and motivation while performing academic tasks (Montalvo and Torres, 2004).

In a framework for classifying the different phases and areas for regulation, Pintrich (2004) describes a temporal matrix involving four phases, from forethought, planning and activation, to monitoring, to control, and, finally, to reaction and reflection. He sees each of these phases as being applicable to cognition, motivation/affect, behavior, and context. In cognition, for example, the students set specific goals for learning, are aware of progress toward those goals, have the ability to change directions if the strategies are not working, and make cognitive judgments on how they did. In motivation/affect, regulation would include attempts to regulate beliefs such as self-efficacy, goal orientation, task value, perceptions of task difficulty, and personal interest, as well as controlling emotions such as fear and anxiety. In behavior, regulation is focused, for example, on effort, control, time management, and help-seeking strategies, while regulation of context is about control of the environment. While the latter would not have been thought possible even 30 years ago in higher education, the

growth of student-centered teaching approaches, where students are asked to make a greater input into processes, has meant that students do have influence over some aspects of the environment.

Self-regulators treat learning as an activity that they develop proactively rather than mere reactive processes stimulated by their reactions to teaching. Each process of self-regulated behaviors can be taught by parents, teachers, classmates, etc.

From this perspective then, the process of promoting effective student learning involves teaching the students to become more effective self-regulators of their learning. This is normally achieved through interventions embedded in the context of normal subject-matter teaching, and in which teachers play a crucial role.

Different types of intervention have been reported. Since the main objective is meta-cognition, the current emphases are on self-reflective practices and on instruction designed to scaffold the students' development. To develop awareness and control of their learning, students must engage with and practice using a range of skills, know how they are used, know why one strategy is best used at one time rather than another, and know what effects the adopted strategy is having on learning. Modeling by the teacher of the key processes – planning, controlling the process, outlining the relevant sources, and reviewing what has been done – provides students with observable examples, and direct teaching of the self-regulation strategies are both fundamental starting points. However, it is the students who must engage in practice, in self-monitoring, and with the social support systems to achieve the levels of competence needed to self-regulate their own learning.

The development of ways of measuring the effectiveness of the interventions has not kept pace with the development of the explanatory frameworks behind the current thinking. Both the Learning and Study Strategies Inventory (LASSI) and the Motivated Strategies for Learning Questionnaire (MSLQ) were designed before the many dimensions now considered to be a part of the SRL framework were identified. In including many of these dimensions, the latter is probably still the best quantitative instrument available.

While many learning-to-learn programs fit comfortably within the descriptions in this section, there are some, for example, those focused on developing specific skills such as reading, or on the development of mnemonic skills that do not. Simpson *et al.* (1997) note in a review of academic assistance programs that these programs have continued to flourish for a variety of reasons, one being that many university students are not self-regulated learners. They describe program models as being based on either a generic or embedded approach. Generic interventions are usually conducted outside of the teaching context (e.g., development of note-taking skills).

In a comprehensive meta-analysis of the effects of learning skills interventions on student learning Hattie *et al.* (1996) use effect size to gauge the effectiveness of a range of intervention programs in terms of students' academic performance, change in study skills, and affective changes. The conclusions they draw from the meta-analysis are consistent with the conclusions in the literature at the time: the interventions are associated with increases in student performance, in study skills, and in affect and are effective most of the time. The effects are greatest on performance (effect size 0.57) and least for study skills (0.16). They conclude that the typical study skills package is not as effective as metacognitive and contextualized intervention, but it is significantly better than nothing. However, they also found that these effects are not uniform across the educational sectors. Of the 51 studies analyzed by Hattie and coworkers, 21 were from university or adult-learning contexts. Effect sizes from the university subsample, for performance, study skills, and affect are 0.27, 0.19, and 0.68, respectively. In terms of performance and study skills, these are small effect sizes, and such interventions cannot be considered to be major contributors to enhancing student learning. The moderate effect size for affect suggests that the interventions may change students' attitudes to their work, and as noted in the notes above on self-regulated learning, students' attitudes are related to aspects of their performance.

Student Engagement: Promoting Effective Learning Through Alignments of Student Support within the Educational System

Kuh *et al.* (2005) have used the idea of students' self report of their engagement in their college life as a proxy for effective learning. Drawing on results from the National Survey of Student Engagement (NSSE), their major conclusion is that what is needed for the promotion of student success (greater engagement) is a holistic approach, as the responsibility lies with just about everybody in the college system, including the students. This includes the governing board, president, provost, teachers, and others in the classroom context, the context outside the classroom, and the administration. The key, however, is not the inclusiveness of this array, but an overall plan aimed at organizing for student success that involves alignment or congruence between the various elements that are included and the sustainability of these initiatives.

Kuh and colleagues describe nine key areas, and strategies within each, identified from the practices of colleges where students report high levels of engagement. Taken together as an aligned and sustained process, adoption of these nine suggestions may lead to more effective learning:

1. *Feature student success in the institutions enacted mission.* Have a vision of what is to be accomplished with undergraduate programs, clarify and translate the mission into plain language, ensure that the espoused mission is enacted, senior leaders champion undergraduate education, and that multiple missions are balanced.
2. *Make talent development a central tenet of philosophy.* Establish high expectations, know your students, set performance standards for students consistent with their preparation, provide generous amounts of constructive feedback, balance academic challenge with adequate support, use pedagogic approaches to complement learning approaches, and encourage student-faculty interaction that benefits student learning.
3. *Cultivate an ethic of positive restlessness.* Steer organization toward continuous improvement, use data to inform decision making, remove less-effective programs and support high-priority initiatives, and put someone in charge.
4. *Put money where it will make a difference to student engagement.* Invest in: (a) activities that contribute to student success, (b) faculty members who are doing the right thing, (c) teaching and learning centers, and (d) opportunities that allow students to apply what they are learning in ways that benefit others, as well as consider a budget model that privileges student learning processes.
5. *Feature diversity inside and outside the class.* Use a multi-faceted, aggressive approach to diversify personnel, and ensure that diverse perspectives are represented in the curriculum.
6. *Attract, socialize, and reward competent people.* Align the reward system with the mission, pick leaders who are right for the times, recruit faculty who are committed to student learning, emphasize student-centeredness, make room for difference, and ensure high-quality student support services.
7. *Encourage collaboration across functional lines and with community.* Encourage cross-functional activity that leads to student success, tighten links between academic and student affairs, harness expertise of other resources, make governance a shared responsibility, and form partnerships with local community.
8. *Lay out the path to student success.* Draw a map for student success, front load resources to smooth the transition, align the physical environment with institutional priorities and goals for success, teach newcomers about the college culture, create a sense of specialness about being a student here, and develop interventions for under-engaged students.
9. *Reculture the institution for student success.* Identify cultural properties that are obstacles to student success, and expand the number of cultural practitioners on campus.

Promoting Effective Learning Through Teaching

Teaching strongly features as a medium through which effective learning can be promoted in both the SAL and SRL perspectives and alignment of teaching conceptions with the rest of the educational system is essential for student engagement.

A set of teaching qualities generally agreed by researchers to be essential for the promotion of student learning have been summarized by Ramsden *et al.* (1995) as follows:

1. Good teachers are also good learners, for example, they learn through their own reading, by participating in a variety of professional development activities, by listening to their students, by sharing ideas with their colleagues, and by reflecting on classroom interactions and students' achievements. Good teaching is therefore dynamic, reflective, and constantly evolving.
2. Good teachers recognize the importance of context, and adapt their teaching accordingly; they know how to modify their teaching strategies according to the particular students, subject matter, and learning environment.
3. Good teachers encourage deep-learning approaches, rather than surface approaches, and are concerned with developing their students' critical thinking skills, problem-solving skills, and problem-approach behaviors.
4. Good teachers set clear goals, use valid and appropriate assessment methods, and provide high-quality feedback to their students.
5. Good teachers show respect for their students; they are interested in both their professional and their personal growth, encourage their independence, and sustain high expectations of them.
6. Good teachers demonstrate an ability to transform and extend knowledge, rather than merely transmitting it; they draw on their knowledge of their subject, their knowledge of their learners, and their general pedagogical knowledge to transform the concepts of the discipline into terms that are understandable to their students. In other words, they display what Shulman has termed "pedagogical content knowledge."
7. Good teachers display enthusiasm for their subject, and a desire to share it with their students.

The last two of the seven focus on an area which has to a large extent been overlooked in learning enhancement – the subject-matter understanding of the teacher.

An essential element in promoting effective learning is the minimizing of barriers to learning. Reducing students' perceptions of the environmental factors affording a surface approach, and developing students' self-regulated learning skills to enable them to take control of their learning, are ways this is achieved in the SAL and SRL perspectives,

respectively. However, there are three other types of barriers that teachers must address; conceptions of learning, conceptions of knowing, and conceptions of understanding.

Adult students vary in the ways they conceive of learning (Marton *et al.*, 1997). Six qualitatively different conceptions have been constituted from interview studies:

1. learning as a quantitative increase in knowledge;
2. learning as memorizing;
3. learning as acquiring facts, skills, and methods that can be retained and used as necessary;
4. learning as making sense or abstracting meaning;
5. learning as interpreting and understanding reality in a different way; and
6. learning as changing as a person.

Students who do not conceive of learning as described in conceptions 4, 5, or 6 have difficulty adopting a deep approach to learning. Promoting effective learning means raising students' awareness of the higher-level conceptions.

Students' conceptions of knowledge, or ways of knowing, have also been found to be barriers to learning (Perry, 1970). Many students, when first encountering a new study area, see knowledge in absolute terms. Something is right or wrong, or black or white. A diversity of views may be acknowledged, but the diversity exists because insufficient study has been done to find the right answer. As students develop, they recognize that because knowledge is relative and socially constructed there will be a diversity of views. In the final phase of development, students make a commitment to a view in preference to other views, based on their own analysis of the evidence available. Students may be prevented from engaging in effective learning if they do not conceive of knowledge in the more relativistic ways.

Early research in science education focused on identifying the misconceptions that students might have about science. The presence of misconceptions was considered to be a factor preventing many students from achieving the scientific forms of those conceptions, and effective teachers were considered to be those who could change those conceptions. Much of the focus of science-discipline-based education journal articles has been on ways of organizing complex scientific knowledge in ways to make it easier for students to learn. More recently, phenomenographic studies of subject matter in a range of disciplines have revealed the variation in the way a group of students experience that subject matter. Phenomenography is the empirical study of the limited number of qualitatively different ways in which a group experiences a phenomenon. These differing experiences are characterized in terms of categories of description, logically related to each other, and forming hierarchies in relation to given criteria. The categories of description corresponding to those differing understandings and the logical relations that can be established between them constitute the main

results of a phenomenographic study (Marton and Booth, 1997). Some categories or ways of experiencing subject matter are found, through this approach, to be more a part of effective learning than others. The logical relations between the categories that constitute an outcome of this process also provide a means through which variation in the quality of student learning might be addressed.

Teachers also have a crucial content-related role in identifying and dealing with troublesome knowledge and threshold concepts (Hounsell *et al.*, 2005). Threshold concepts seem to serve as portals, opening up previously inaccessible ways of thinking about certain aspects of a subject – a transformed way of understanding without which the learner cannot progress. While they are considered to be transformative, and, once understood, provide a gateway to previously hidden interrelatedness, they are, like troublesome knowledge, a barrier to learning until that threshold is crossed.

Bibliography

- Hattie, J., Biggs, J., and Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research* **66**, 99–136.
- Hounsell, D., Entwistle, N., Anderson, C., *et al.* (2005). Enhancing teaching–learning environments. *Final Report to the Economic and Social Research Council on TLRP Project L139251099*. <http://www.ed.ac.uk/etl> (accessed August 2009).
- Kuh, G. D., Kinzie, J., Schuh, J. H., *et al.* (2005). *Student Success in College: Creating Conditions That Matter*. San Francisco, CA: Jossey-Bass.
- Marton, F. and Booth, S. (1997). *Learning and Awareness*. Mahwah, NJ: Erlbaum.
- Marton, F., Hounsell, D., and Entwistle, N. J. (eds.) (1997). *The Experience of Learning: Implications for Teaching and Studying in Higher Education*, 2nd edn. Edinburgh: Scottish Academic Press.
- Montalvo, F. T. and Torres, M. C. G. (2004). Self-regulated learning: Current and future directions. *Electronic Journal of Research in Educational Psychology* **2**, 1–34.
- Perry, W. (1970). *Forms of Ethical and Intellectual Development in the College Years*. San Francisco, CA: Jossey-Bass.
- Pintrich, P. (2004). A conceptual framework for assessing motivation and self-regulation in college students. *Educational Psychology Review* **16**, 385–407.
- Prosser, M. and Trigwell, K. (1999). *Understanding Learning and Teaching: The Experience in Higher Education*. Buckingham: Society for Research into Higher Education and Open University Press.
- Ramsden, P. (2003). *Learning to Teach in Higher Education*. London: RoutledgeFalmer.
- Ramsden, P., Margetson, D., Martin, E., and Clarke, S. (1995). *Recognising and Rewarding Good Teaching in Australian Higher Education*. Canberra: Australian Government Printing Services.
- Simpson, M. L., Hynd, C. R., Nist, S. I., and Burrell, K. I. (1997). College academic assistance programs and practices. *Educational Psychology Review* **9**, 39–87.

Student Attrition from Higher Education Institutions

M Hayden, Southern Cross University, Lismore, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Student attrition – This concept is variously measures, but most often by student persistence from year to year and from course commencement to course completion.

Student survival – The OECD calculates this rate by determining the ratio of the number of students who graduate from an initial degree during a reference year to the number of new entrants to that degree n years beforehand, with n being the number of years of full-time study required to complete the degree.

Although estimates of its incidence and perceptions of its gravity vary from one national higher education system to another, student attrition from higher education is universally regarded as a matter of serious educational concern. Some indication of its extent is provided by a recent estimate of the Organization for Economic Cooperation and Development (OECD) that as many as 30% of all students who commence undergraduate degree programs fail to complete them within a reasonable period of time (OECD, 2006: 51). In some countries, most notably the United States, the rate is closer to 50%.

The empirical literature on the topic is voluminous. Not surprisingly, in view of the high rates of student attrition experienced, much of this literature comes from the United States. The United States is also the main source of relevant theoretical models. Preeminent among these is Tinto's model of student departure from higher education in which a student's failure to integrate with the social and academic environment of a higher education institution is identified as the main cause of student attrition.

This article addresses various aspects of student attrition: how it is measured, how it is explained, how it might be addressed, and what it costs. It focuses on attrition from undergraduate degree programs, this being the main area of attention in most of the relevant literature.

Measuring Student Attrition

Student attrition is variously measured. These measures address student persistence from year to year and from course commencement to course completion. Ideally, they would take into account the incidence of deferments

and of transfers between courses and institutions, but often they do not. Estimates of attrition may be inflated if deferments, that is, periods of approved interruption of enrolment, and if transfers, that is, leaving one course for the purpose of undertaking another at the same or at a different institution, are not taken into account.

The most common estimates of student attrition are those based on measures of year-to-year persistence at the same institution. These are typically calculated for students in their first year of undergraduate studies, when student attrition is highest. In Britain, a higher education statistics agency calculates year-to-year continuation rates for all full-time first-degree entrants to higher education. These rates are relatively high: as many as 87.7% of all full-time first-degree entrants to higher education in 2003/2004 continued their studies at the same institution in 2004/2005, and a further 2.8% of entrants transferred to another higher education institution in 2004/2005 (HESA, 2006, Table 3a). The year-to-year attrition rate from higher education for this cohort was, therefore, 9.5%, although even this estimate may be inflated because it does not take deferments into account. Of note from the British data is that some higher education institutions, generally the more selective ones, had year-to-year attrition rates of less than 5%, while others had rates of attrition between 20 and 30%. This kind of institutional variation is replicated in most countries.

In Australia, the Department of Education, Science, and Training also calculates year-to-year persistence rates for entrants to undergraduate courses. Although not as high as in Britain, these rates are relatively high: 78.8% of all commencing undergraduate domestic students in 2002 continued their studies at the same institution in 2003 (Lukic *et al.*, 2004). When allowance was made for the impact of 1-year deferments, the annual continuation rate increased by about 10 percentage points. The year-to-year attrition rate when adjusted for 1-year deferments was, therefore, about 12%. This estimate would be further reduced if institutional transfers were taken into account.

It is interesting to compare the British and Australian rates with persistence rates for the United States. Berger and Lyon (2005: 25) report that, based on estimates for 2000 produced by the American College Testing Program, only about three-quarters of all freshmen at 4-year institutions return to their studies the following year. At more selective institutions, the average continuation rate was 92%, while at less-selective institutions it was only 65%. The availability of open-access higher education opportunities in the

United States impacts considerably on these rates. Year-to-year persistence rates for freshmen at these institutions were in the order of only 50%.

International comparisons of attrition rates are interesting, but there are many data limitations. Apart from the fact that these comparisons are possible only for those countries that publish attrition data, the main problem is that the measures adopted are rarely exactly the same. Estimates of attrition are also significantly affected by whether or not part-time students have been considered, and, if they have been, by the relative size of the part-time student population. Attrition among part-time students is generally well above that for full-time students.

Student attrition rates may also be estimated on the basis of course-completion data. An example is the OECD's annual estimate of student survival rates for member countries. These rates are calculated, using cross-sectional data, by determining the ratio of the number of students who graduate from an initial degree during a reference year to the number of new entrants to that degree n years beforehand, with n being the number of years of full-time study required to complete the degree (OECD, 2006: 57). In 2004, most OECD countries recorded survival rates close to the average of 70%, but some, notably Japan, Ireland, and Korea, had rates that were markedly above the average. Others, notably Mexico, New Zealand, and the United States, had rates markedly below it. Student-survival rates for systems with stable enrolment patterns, that is, those that show neither expansion nor contraction in total enrolments, provide a reasonable basis for international comparisons, but there is a significant error component in these estimates because of (1) possible double-counting of course commencements, due to transfers; and (2) underestimation of completions, due to extended deferments.

Completion rates calculated using longitudinal data, that is, data derived from tracking higher education entrants through to course completion, generally provide a more realistic impression of student attrition. Astin and Oseguera (2005) calculated the award-completion rates of a large national sample group of students in the United States who were surveyed in 1994 when they were in their first year of college. By 1998, only 36% of students had completed a bachelor's degree award. By 2000, however, 58% of students had completed an award. Completion rates need, therefore, to be considered over realistic periods of time that allow for the normal incidence of interruptions to studies.

Completion rates calculated using longitudinal data also allow for consideration of the impact of institutional transfers. In an Australian investigation by Martin *et al.* (2001), entrants to undergraduate programs in 1992 were tracked through to 1999, by which time 64% of students had completed an award from their university of first enrolment. When their progress in terms of whether or

not they had ever completed an award from any university was calculated, however, 72% of students were estimated to have completed an award. The difference was attributable to the incidence of institutional transfers.

Student attrition is assuming increasing political significance in many countries, and so too, therefore, is its measurement. Any government subsidizing higher education is likely to interpret any instance of student attrition as representing inefficiency in the use of scarce resources. However, the fact is that not all student attrition represents inefficiency. Students often transfer from courses and institutions for sound reasons, and interruptions through deferment may in many instances be beneficial to future academic progress. It is attrition that results in long-term dropout from higher education that should be of most concern. Curiously, this form of student attrition is not well indicated by most conventional measures of student attrition.

Explaining Student Attrition

Tinto's (1975, 1987, 1993) interactionist theory of student attrition dominates the various theoretical formulations advanced to explain student attrition. Developed during the 1970s and 1980s, this theory proposes that premature student departure from a higher education institution is a voluntary behavior prompted by feelings of isolation as a consequence of the failure to integrate with the new social and academic environment of the higher education institution. This process is said to be longitudinal in nature. At the time of admission, a student's decision to withdraw may be directly affected by individual characteristics, including motivation, ability, family background, and prior experiences of education, but, more likely, these characteristics will impact on two types of commitment, the student's commitment to the institution and the student's commitment to the goal of graduating. These commitments may, however, be strengthened or weakened over time as the student embarks on the process of assimilating to a new social and academic culture. Difficulties experienced in integrating with the social culture will weaken the student's commitment to the institution, while difficulties experienced in integrating with the academic culture will weaken the student's commitment to the goal of graduating. Over time, the combined strength of these commitments will largely determine whether the student persists or withdraws.

Although remarkable in terms of its influence, Tinto's model has been found to be deficient in a number of important respects. Empirical support for its key propositions is reported from the United States to be partial in the case of residential universities, and problematic in the case of commuter universities (Braxton *et al.*, 2004: 20). Its inability to accommodate the influence of significant external community groups (e.g., families, neighborhoods,

and secondary schools), the organizational attributes of colleges and universities, individual psychological processes, and general economic forces, including student finances and financial aid, on student attrition has also been identified (Braxton, 2000: 259). The model remains, however, a landmark in the relevant literature.

Other significant theoretical formulations include those of Bean and Astin. Bean's model, which is explicitly psychological in nature, proposes that student attrition is effectively the product of attitudes and psychological processes, the impact of which may be intensified as a consequence of interactions with the academic, social, and organizational conditions of a particular higher education institution environment (Bean and Eaton, 2000). In this model, incentives and opportunities for persistence play an important role through their impact on individual attitudes and psychological dispositions.

Astin's model, also psychological in orientation, focuses more on the specific importance of student involvement with studies and academic life (Astin, 1984). According to Astin, what matters is the amount of time spent by a student attending to studies, being on campus, interacting with academic staff, and engaging with student organizations and activities. As in Bean's model, academic, social, and organizational conditions are seen to impact on student persistence, but, in Astin's theory, their importance derives from the extent to which they stimulate and support a student's propensity to be fully involved in the academic experience.

These formulations, and Tinto's in particular, have been characterized as assigning too much responsibility for persistence to the ability and willingness of students entering higher education to assimilate what for many will be a completely new and quite challenging environment. In recent literature, calls have been made for higher education institutions to adapt more to the cultural attributes of their new students. Thomas (2002: 431), for example, argues that: "If a student feels that they do not fit in, that their social and cultural practices are inappropriate and that their tacit knowledge is undervalued, then they may be more inclined to withdraw early." Zepke *et al.* (2006: 588) invite consideration of the ways in which "institutions accept and recognize diverse learners' goals and cultural capital, and adapt their mores and practices to accommodate these in a learner-centred way." These arguments are underpinned conceptually by Bourdieu's conception of the importance to academic success of cultural capital: "the importance of cultural experiences is regarded as a driver of inequalities which ultimately lead to limited and restricted opportunities for higher education" (Longden, 2006: 176).

A substantial proportion of the literature on student attrition is not, however, overtly concerned with theory. The range of possible explanatory variables considered in many empirical studies is often eclectic in nature, reflecting local interests, as well as characteristics and

constraints of particular data sets and methodological approaches. An implication of this diversity is that generalization is more difficult.

In one investigation (referred to earlier), Astin and Oseguera (2005) analyzed data on degree completions by more than 50 000 undergraduates at 262 colleges and universities across the United States who commenced undergraduate studies in 1994. Using multivariate statistical methods, they examined the ability of selected individual characteristics (e.g., gender, high school grades, standardized test scores, parental-education levels, involvement with athletics), environmental contingencies (e.g., living in a residence hall, financial support, working off-campus), and institutional characteristics (e.g., institutional selectivity, public, or private status) to predict degree completion within 6 years of course commencement. The investigation pointed to the strong independent predictive ability of high school grades, institutional selectivity, and residency in a student hall. Oversleeping and a tendency to miss classes and appointments were also found to be independently significant in affecting degree completion – in these cases in a negative way.

In another investigation, Yorke and Thomas (2003) sought to identify the conditions most likely to impact on institutional success in reducing student attrition. They employed a case-study approach involving interviews with senior management at six higher education institutions in Britain that had performed unexpectedly well in achieving high-retention and completion rates for students from backgrounds not strongly disposed to academic success. Their investigation was informed by a comprehensive literature review (Thomas, 2002) that identified seven areas of interest: academic preparedness, the academic experience (teaching, learning, and assessment), institutional expectations and commitment, academic and social match, finance and employment, family support and commitments, and institutional support services. No blueprint for success was identified, but the conditions considered most likely to impact positively on student retention were: a positive and supportive institutional climate, an emphasis on support leading up to and during the first year of studies, an emphasis on formative assessment in the early phase of programs, recognition of the importance of the social dimension in learning activities, and a willingness to respond to changing patterns of students' engagement in higher education.

In yet another empirical investigation, Long *et al.* (2006) employed multivariate statistical analysis of national survey data, together with a thematic analysis of students' written comments, to examine the reasons for first-year undergraduate students in Australia in 2004 not proceeding to the second year of their studies in 2005. Various possible contributing factors were considered. A striking feature of the results was that, although 28.7% of the students surveyed did not continue in 2005 in the course they had commenced

in 2004, only 13.7% of all students had, in fact, withdrawn from studies in 2005, with almost one-half of these indicating an intention to return to studies after a short period of interruption. Most first-year course attrition was, therefore, simply the result of course transfers, within and between institutions. The investigation also found that students who dropped out from university studies (even if in some cases for no more than a 6-month period of deferment) were likely to be older, to be studying part-time, and to be unlikely to nominate a clear reason for having enrolled at university in the first place. Older students who dropped out were more likely to report conflict in balancing study requirements with full-time jobs and family commitments, while younger students who dropped out were more likely to report a change in direction or the need for a break from study.

Student attrition from higher education is a complex phenomenon. As Long and colleagues report, even the students themselves find it necessary to point to combinations of factors when trying to explain why they do not continue their studies. However, enough is known to be able to identify some key conditions that affect the incidence of student attrition. These include: a student's involvement with and commitment to a particular course of study, a student's ability to commit sufficient time to studies, a higher education institution's commitment to making its courses exciting and rewarding for students, a higher education system's academic selectivity, and, for groups less well represented in higher education, the generosity of provisions for financial support to students.

Addressing Student Attrition

It is widely agreed that policies seeking to address student attrition should have as their focus the creation of conditions for success by all students, rather than being concerned solely with improving retention. Yorke and Longden (2004: 132) argue, for example, that: "A policy focus on student success in higher education through teaching, learning and assessment, and through institutional support services, is likely to lead to better retention than a focus on retention itself." The problem is that an exclusive focus on retention is too limiting, and may even result in outcomes that are incompatible with broader policy objectives. One of the surest ways of improving student retention would be by restricting access to higher education to students with superior levels of academic ability and motivation. Yet, in most systems, such a policy would be inconsistent with a commitment to increasing youth access to higher education and improving the social representativeness of the student population.

The literature on student attrition contains an impressive array of suggestions on how to create conditions for success by all students. Yorke and Longden (2004: 134–147) provide what is possibly the most succinct account of the

various measures proposed. They group these measures according to whether responsibility for their implementation rests with the individual student, the higher education institution, or the higher education system. Students are said to have responsibility for being adequately informed about their course of studies, for developing effective ways of learning, for responding appropriately to any initial lack of academic success, and for anticipating reasonably the extent to which their studies may be adversely affected by financial and lifestyle risks. Higher education institutions are said to have responsibility for selecting students wisely, for informing students fully about course requirements, for giving students a sense of belonging to the institution, for ensuring that students have appropriate and effective support services, for ensuring that academic staff are capable in performing their teaching duties, for creating an academic culture that is supportive of a high quality of student learning, for ensuring that first-year students receive particular support, and for responding sympathetically to adverse events in students' lives. Higher education systems are said to have responsibility for ensuring that good teaching is highly valued, for implementing system-wide quality-assurance processes, for seeking to ensure that financial support systems are as straightforward as possible, and for ensuring that policies intended to support participation by designated student groups are not contradicted or undermined by other policy initiatives.

Tinto (2005) has recently proposed a model of institutional action to enhance student success in higher education. He identifies five conditions in universities that, in the light of insights from research, are of greatest importance. These are: the seriousness of an institution's commitment to the importance of student success, especially among low-income and underrepresented student groups; the extent to which an institution creates an educational setting in which students are required to achieve high academic standards; the adequacy of an institution's provision of adequate and effective forms of academic and social support for students; the effectiveness of an institution's monitoring and feedback arrangements for informing students about progress in their learning; and the institution's success in engaging students with the academic and social culture of the institution.

It is evident that Tinto attaches great importance to the role higher education institutions might play in affecting the conditions for student success. Over recent years, higher education institutions do appear to have become primarily responsible for student success in higher education, but they are not solely responsible. As Tinto (2005: 321) himself points out, first-year students typically spend less time on their studies than is considered necessary for successful learning. Their commitment and expectations in relation to their studies are not always amenable to influence by higher education institutions. Student success is affected also by external conditions beyond the control

of individual higher education institutions. It is evident in many countries that, on average, students feel compelled to engage for longer periods than ever before in paid employment during study semesters. Beyond a certain level, this engagement impacts adversely on student success, regardless of anything higher educational institutions can do (Long and Hayden, 2001).

Costs of Student Attrition

Student attrition generally incurs cost. The most costly form of attrition is long-term dropout from higher education, but even course transfers have a cost impact on individual students, their higher education institution, and the public (if public subsidies are involved). Cost for the most part can be calculated in financial terms, but some forms of cost, particularly opportunity costs and psychological costs, are not readily measurable in financial terms.

Various attempts have been made to calculate the cost of student attrition. Yorke and Longden (2004: 11) provide examples from Britain of different ways in which these calculations can be performed. Using data from 1999, they estimated the cost of student attrition to be of the order of 3–4% of the public funding allocated to undergraduate higher education. However, whether this cost represented wastefulness depends on what the prior expectations were. These expectations are often not well articulated by governments, except in response to a public outcry.

Students and higher education institutions also incur financial costs as a consequence of student attrition. The scale of these costs will depend on the extent to which public subsidies have been provided. Costs incurred by students who discontinue are generally in the form of fees and study-related expenses incurred, while costs for institutions as a result of discontinuation are generally in the form of upfront expenses incurred in the expectation that students would complete their courses within a reasonable time period.

Students, higher education institutions, and the public also incur opportunity costs as a consequence of premature departure from higher education. These costs are rarely calculated in financial terms, but their impact may well be significant. Full-time students who do not realize the benefits of graduating from higher education generally forego possible paid employment during the period of their enrolment in higher education. Higher education institutions and the public also incur an opportunity cost in that the student will in many cases have tied up a place in higher education that another student could possibly have utilized more effectively.

The psychological costs of student attrition are the most difficult to measure, and they are also the most emotive. The disappointment and loss of self-esteem associated with

dropout may be overwhelming, as in the case of a student who reported that: "My family did not talk to me for a long time; they said I'd let them down. They had huge expectations of me; I was the brightest one; the first in the family to go to university" (quoted in Yorke and Longden, 2004: 7). However, it is also the case that students may perceive certain benefits in having participated in higher education, even if they did not achieve course completion.

Concluding Remarks

Student attrition takes various forms, including course and institutional transfer, deferment and long-term dropout. There is rarely a single reason for it, and its consequences can vary greatly in terms of their seriousness. Responses to it require a focus on the broad range of conditions that affect student success.

Student attrition provides an early warning sign of possible problems in a higher education institution, and, more broadly, in a higher education system. Adverse changes in its incidence may well indicate the need for increased attention to be given to key conditions affecting the quality of student experience. These conditions include the extent to which students are engaged by and committed to their courses, have been effectively selected for the courses they seek to undertake, and have sufficient financial support to enable them to have the time required for their studies. Higher education institutions inevitably bear much of the responsibility for addressing these conditions, but students also have responsibilities, particularly in terms of the earnestness of their personal commitment to academic success. So too does the public, particularly concerning financial support, because it is the public that ultimately benefits from having a highly educated population.

The form of student attrition that is of most concern is long-term dropout. The incidence of this form of attrition is lower than headline estimates of attrition based on persistence and completion rates would suggest. Its seriousness should not, however, be diminished. The individual and social costs of long-term dropout may be substantial, and the fact that it occurs more often among student groups whose backgrounds are characterized by social and educational disadvantage should be of additional concern.

Finally, to return to an earlier theme, the differences between systems in terms of the incidence of student attrition are striking. The reasons for these differences have not been able to be explored in this article, but then neither are they examined in much detail in the literature on student attrition. This area is one worthy of further exploration, particularly concerning how policies seeking to widen access to higher education can be effectively integrated with policies that seek to reduce the incidence of student attrition.

Bibliography

- Astin, A. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel* **25**, 297–308.
- Astin, A. W. and Oseguera, L. (2005). Pre-college and institutional influences on degree attainment. In Seidman, A. (ed.) *College Student Retention: Formula for Student Success*, pp 245–276. Westport, CT: Praeger.
- Bean, J. P. and Eaton, S. B. (2000). A psychological model of college student retention. In Braxton, J. (ed.) *Reworking the Student Departure Puzzle*, pp 48–61. Nashville, TN: Vanderbilt University Press.
- Berger, J. B. and Lyon, S. C. (2005). Past to present: A historical look at retention. In Seidman, A. (ed.) *College Student Retention: Formula for Student Success*, pp 1–29. Westport, CT: Praeger.
- Braxton, J. (2000). Conclusion: Reinvigorating theory and research on the departure puzzle. In Braxton, J. (ed.) *Reworking the Student Departure Puzzle*, pp 257–274. Nashville, TN: Vanderbilt University Press.
- Braxton, J. M., Hirschy, A. S., and McClendon, S. A. (2004). *Understanding and Reducing College Student Departure. No. 3. ASHE-ERIC Higher Education Research Report Series*. San Francisco, CA: Jossey-Bass.
- Higher Education Statistics Agency (HESA) (2006). *Performance Indicators*. Cheltenham: HESA.
- Long, M., Ferrier, F., and Heagney, M. (2006). *Stay, Play or Give It Away? Students Continuing, Changing or Leaving University Study in First Year*. Melbourne: Centre for the Economics of Education and Training, Monash University/Australian Council for Educational Research.
- Long, M. and Hayden, M. (2001). *Paying Their Way: A Survey of Australian Undergraduate University Student Finances, 2000*. Canberra: Australian Vice-Chancellors' Committee.
- Longden, B. (2006). An institutional response to changing student expectations and their impact on retention rates. *Journal of Higher Education Policy and Management* **28**, 173–187.
- Lukic, T., Broadbent, A., and Maclachlan, M. (2004). *Higher Education Attrition Rates 1994–2002: A Brief Overview. Research Note No. 1, Higher Education Analysis Section*. Canberra: Department of Education, Science and Training.
- Martin, Y. M., Maclachlan, M., and Karmel, T. (2001). *Undergraduate Completion Rates: An Update. Occasional Paper Series 01/F*. Canberra: Department of Education, Science and Training.
- Organisation for Economic Cooperation and Development (OECD) (2006). *Education at a Glance*. Paris: OECD.
- Thomas, L. (2002). Student retention in higher education: The role of institutional habitus. *Journal of Education Policy* **17**, 423–442.
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research* **45**, 89–125.
- Tinto, V. (1987). *Leaving College: Rethinking the Causes and Cures of Student Attrition*. Chicago, IL: University of Chicago Press.
- Tinto, V. (1993). *Leaving College: Rethinking the Causes and Cures of Student Attrition*, 2nd edn. Chicago, IL: University of Chicago Press.
- Tinto, V. (2005). Epilogue: Moving from theory to practice. In Seidman, A. (ed.) *College Student Retention: Formula for Student Success*, pp 317–333. Praeger: Westport, CT.
- Yorke, M. and Longden, B. (2004). *Retention and Student Success in Higher Education*. Maidenhead: Open University Press.
- Yorke, M. and Thomas, L. (2003). Improving the retention of students from lower socio-economic groups. *Journal of Higher Education Policy and Management* **25**, 64–74.
- Zepke, N., Leach, L., and Prebble, T. (2006). Being learner-centred: One way to improve student retention. *Studies in Higher Education* **31**, 587–600.

Further Reading

- Braxton, J. (ed.) (2000). *Reworking the Student Departure Puzzle*. Nashville, TN: Vanderbilt University Press.
- Seidman, A. (ed.) (2005). *College Student Retention: Formula for Student Success*. Praeger: Westport, CT.

Relevant Website

<http://www.hesa.ac.uk> – Higher Education Statistics Agency.

University Students' Experiences of Higher Education

C McInnis, PhillipsKPA, Richmond, Australia

© 2010 Elsevier Ltd. All rights reserved.

The student experience refers generally to the broad range of academic, extracurricular, and social activities that involve students during their time at university. The nature and extent of involvement varies for groups and individuals. The term is often used in a more specific sense to refer to aspects of the student experience of teaching and learning such as course design, curriculum organization, instruction, and learning with others, inside and outside the classroom. The first-year experience is a distinctive area of study in its own right that considers issues such as the transition of students into higher education, especially the adaptation and adjustment of those moving directly to university from secondary school. Universities and researchers also refer to the total student experience. This emphasizes the interrelationship between the various elements of university life that result in an experience, and outcomes, for students that is assumed to be something greater than the sum of the parts.

This article reviews developments in the study of the student experience over the last decade or so, and is focused on undergraduates. Among the issues discussed are: whether, and to what extent, the student experience as traditionally defined is still meaningful or indeed viable; the ways in which research and policy is responding to the changes in the nature of the student experience; and the extent and significance of global convergence in student lives.

The student experience is more difficult to define now than it was in the 1960s and 1970s, not least because of the far greater number of students now enrolled in universities. This inevitably brings with it considerably more diversity in student backgrounds, daily lives, and expectations. Renewed interest as to what constitutes the student experience and how it is changing has been further generated by the pressures of market competition and accountability on universities, major changes in the ways students go about learning, and new external social and economic pressures on student lives. Moreover the dimensions of the student experience multiply when national contexts are examined and compared.

In Europe, research and policy in this area is more strongly focused on the social and economic conditions of student life. For example, the Euro Student project gives little attention in its surveys to student perceptions of their experience or the patterns of their behavior related to study habits and engagement (Euro Student, 2004). It emphasizes the relationship between the culture of different disciplines, the time demands or workload, and the extent to which

courses are regimented. Synthesizing culturally bound surveys from eight countries inevitably smooths out some major points of distinction in the assumptions and beliefs about the ways students live and the nature of their experience. A view implicit in the surveys is that student perceptions of their personal experience are unreliable and of limited interest.

In France, the National Observatory of Student Life developed by the Ministry of National Education aims to examine all aspects of student life. Its comprehensive triennial survey is focused on the impact of living conditions, experiences, and perceptions on study patterns and outcomes. The elements of the French student experience considered significant include: the level of independence from parents; the relationship between paid work and study; student expectations as to what they will gain from the experience; and factors that might inhibit learning.

The complexity and multiplicity of student types and roles are currently challenging established research, policy, and practice for governments, universities, and academics. Traditional notions of a campus-based experience and the university as a cohesive learning community preparing undergraduates for citizenship and the workplace are being tested in most developed countries. Universities around the world are under pressure to adapt to significant changes in student needs and expectations while maintaining core values concerned with the holistic development of students.

Types and Purposes of Studies

Studies of the student experience range considerably in scope and methodology. They include: investigations of national trends by way of large-scale surveys quantifying student self-reported behaviors; qualitative strategies aimed at uncovering the everyday realities of student life; and, less common, longitudinal studies of cohorts as they move through their undergraduate programs. The purposes of the studies include contributing to theory and practice in higher education, informing national policy, providing indicators of institutional performance, quality assurance, and marketing. There is a substantial gray literature from institutional research (Harvey and Drew, 2006). These studies gather data for strategic planning and improvement purposes although they are often the basis for journal publications and continue to add to what Clark characterized in the 1970s as a relatively massive but trivial literature

(Clark, 1973: 9). Thirty years later Terenzini and Reason observe that the literature overall is: "... highly segmented, even atomistic, and virtually atheoretical. With rare exceptions (e.g., Astin, 1998) studies focus narrowly on individual programmatic interventions or overlook the wide variety of influences shaping an outcome" (Terenzini and Reason, 2006: 1).

The expectation that institutions should enhance the student experience has generated new demands to monitor student behavior and satisfaction. There has been a proliferation of studies from government, research institutes, and institutions. In addition, research is routinely commissioned to assess the quality of student life for marketing purposes. Studies of student life have also grown at the institutional level as part of the quality-assurance process. Partly in anticipation of external pressures to provide evidence of student satisfaction, and partly to anticipate and deflect criticism, many universities are now generating quite sophisticated pictures of student lives to inform their policy and planning. Until recently, this has been more prominent in the Anglophone countries, but has now become an almost inescapable feature of university strategic-planning initiatives in most developed systems.

The main thrust of the research considered in the discussion here has been toward establishing whether the university contribution to student life is more substantial than providing the basic elements of instruction and delivery of curriculum content. Despite evidence of diversity, the prevailing models of student life from the US assume an optimal level of student engagement with the university or college in a campus-based environment. These include the assumptions that a positive student experience is unlikely to occur in a social vacuum; that learning in a group is critical to the quality of student life; and that there are important learning outcomes from university beyond the mastery of subject matter.

The US is the most significant source of systematic and comprehensive scholarly work on the student experience. The substantial US research literature is in part a product of the deeply embedded belief that undergraduate education ought to be intensively student centered. What students do in college and what that means for them and America's future is a common theme. The key issues generating research and policy most recently have been focused on the need to improve the quality of the student experience set against an apparent drift away from the levels of student engagement that characterized the undergraduate experience for previous generations. Over the last 5 years or so, there has been a wave of internal reviews of undergraduate programs in US universities and colleges to counter the perceived loss of focus and direction. This is particularly the case for large research-intensive institutions. Most institutions involved in the reviews have attempted to clarify and reassert core values about the

importance of the undergraduate years while adapting to the changed realities of student lives and priorities.

Models and Theories

There is a strong tradition of psychological and sociological theory and research on the student experience. The origins of this work can be found especially in models of adolescent and institutional socialization. In their comprehensive review of US studies, Pascarella and Terenzini (2005) examine the range of theories and models of student change in two broad families: developmental theories and models that focus on individual growth; and, college impact models that study change that result from institutional characteristics. Research in the first group draws from fields such as psychosocial theories of student development and cognitive-structural theories. In the second group, the somewhat eclectic sociologically oriented studies tend to identify and evaluate several sets of variables presumed to influence one or more aspects of change (Pascarella and Terenzini, 2005: 18). The variable sets include environmental factors such as the academic and social climate created by faculty and students on a campus.

A typical model works from a chronological sequence of major points in student life. Pre-entry attributes of students are included to account for diversity in student readiness and relative fit with the institution. In these models, the formation of a student identity through interaction with peers and teachers sits alongside academic performance as central to the experience, and the overall attributes gained on graduation combine to represent a holistic outcome. Notwithstanding the view that institutions should bear considerable responsibility for the quality of the student experience, most models assume that students share responsibility for their choices and level of commitment.

Pascarella and Terenzini (2005) chose to organize their review around outcomes for students rather than, for example, attempting to isolate the discrete influences on the student experience. Of the six critical questions they pose regarding available evidence on the impact of university on students, three are particularly pertinent to this discussion: the evidence that different kinds of institutions have a differential influence on student change; the evidence on the effects of different experiences in the same institution; and the evidence that a collegiate experience produces conditional, as opposed to general, effects on student change or development. The latter concerns the extent to which these experiences are conditioned in their influence by the different characteristics of students (Pascarella and Terenzini, 2005: 9). While much of the research in this area has focused on differences in gender, racial, or social background, there has been an increasing

interest in key characteristics of students in relation to their personal priorities and choices about the nature and extent of their commitment to university study.

The broad findings that have informed the search for optimum undergraduate learning environments include the following: the greater the engagement in academic work/academic experience the better the knowledge acquisition and general cognitive growth; effective strategies for engagement are those that are based on increasing active learning, including peer teaching and individualized systems of instruction; wide cognitive impact comes from academic experiences that purposefully provide for challenge and integration; and student learning shows an unambiguous link to teacher classroom behaviors. The two most salient dimensions are the skill levels of teaching, and the nature of the course structure and organization. In summary, effective learning in the undergraduate experience occurs when educational settings are designed and delivered on the assumptions that learning is holistic rather than segmented, and that opportunities are created for active student involvement.

Social Dimensions of the Student Experience

Beyond these fundamental elements of effective learning environments, there is increasing acknowledgment that the social aspects of the student experience inside and outside the classroom, whether constructed or accidental, mediate learning in ways that have a profound influence on student outcomes. Pascarella and Terenzini (2005) demonstrated that the interaction between students, and between academics and students, outside the classroom makes a highly significant contribution to the cognitive and personal growth of students. Likewise, the overall nature of the learning climate of an institution has direct and positive outcomes on student achievement. One of the unequivocal conclusions from their review of the literature is that the outcomes of undergraduate education are largely determined by individual effort and involvement in the academic, interpersonal, and extracurricular offerings on a campus. However, the authors also make the crucial point that it is therefore all the more important to focus on the ways in which an institution can shape its academic, interpersonal, and extracurricular offerings to encourage student engagement (2005: 602). This is explored by a UK project on the organizational mediation of university learning (2005–2007) that represents a significant shift in the established UK research on teaching and learning (Brennan and Jary, 2005). Supported by the Higher Education Academy (HEA), the study aims to show how university learning “is affected by the way courses are organised, by the places in which it is happening, by the people one is

learning alongside, by the reasons people have for studying, by the other things that are going on in their lives” (Brennan and Jary, 2005: 3). As with a number of recent studies, it is partly motivated by the need to balance the tendency in the research on teaching and learning against the overemphasis on the significance and impact of instruction and curriculum content on student outcomes.

Importantly, the UK project distinguishes between the collective experiences of students within an institutional/organizational setting (the curriculum, student culture, and so on) and the personalized experiences of individual students in those settings (e.g., study method and reasons for study). A third dimension, the parallel experience, refers to the amount and nature of paid work, domestic life, and other commitments (Brennan and Osborne, 2005: 4). Focusing on diversity in higher education, the project provides new insights into how institutions, by the ways they are organized, provide opportunities for students to create a collective experience. The differentiated student experience in the UK is revealed through the analysis of student choices and expectations in specific fields of study. It identifies the tensions and issues raised in the interplay between social and academic cultures in fragmented higher education landscapes where university culture is not uniformly accessed or experienced (Houston and Lebeau, 2006: 12).

First-Year Experience

The first-year experience has perhaps attracted as much attention outside the US in recent years as the studies of the student experience in general. Students' initial experiences of higher education are pivotal in establishing attitudes, expectations, motivation, and approaches to learning (McInnis, 2001). It is now widely understood that investing in the quality of first year brings substantial benefits for institutions (Tinto, 1998). Research into the first-year experience provides critical insight into wider issues of student engagement and the development and improvement of learning and teaching. The sheer volume of research from the US has generated academic journals and specialist centers dedicated to understanding and improving the first-year experience. For 40 years, US first-year students from around 700 colleges and universities have been surveyed annually by the Higher Education Research Institute at the University of California. This has produced a rich database of change in the norms and profiles of students (Astin, 1998; HERI, 2007).

A decade of national trend studies in the first-year experience in Australian universities has revealed trends in patterns of student engagement, their use of information and communication technologies, and the ways they manage their everyday lives beyond the classroom and in

paid employment. The study monitored changes in student time devoted to academic endeavors, including class attendance and time spent on campus (McInnis and James, 1995; McInnis *et al.*, 2000; Krause *et al.*, 2005). The impact of the increasing numbers of students working in part-time paid work has only recently been considered of national significance in many countries. In many instances, changes in student financial support systems have put pressure on students to work to sustain themselves and to pay fees. However, there is also clear evidence that students are increasingly working to maintain a level of financial independence from their families for discretionary spending. This represents an international phenomenon of a shift in student expectations about their lifestyles and their commitment to university life.

The first year has also generated interest in the UK recently where the first major UK study of students' first-year experiences in higher education produced many findings in common with the Australian research. The study, commissioned by the HEA aimed particularly to find out what aspects of the student experience may affect students' decision to withdraw from their studies – an area that has been underresearched in the UK (Yorke and Longden, 2006). It surveyed students on their experiences of learning, teaching, and assessment, and on other aspects of their first-year experience such as travel, finance, and social life. At the same time, the HEA commissioned a substantial review of the literature on the first-year experience (Harvey and Drew, 2006).

Student Engagement and the Social Experience of Learning

A growing number of studies have been prompted by a widely perceived decline in the level of engagement in the face of competing pressures on student time and interest (McInnis, 2002; McInnis and Hartley, 2002). The decline is indicated by factors such as the number of days students spend on campus and the amount of contact they have with fellow students. Students are now more likely to study in multiple settings: in large lecture theaters, in groups on collaborative exercises, in computer laboratories with two or three others in an online tutorial, or simply work at home alone. They are less likely to spend significant time in small-group tutorials, or to have one-to-one consultation with their lecturers. On the other hand, they often have access to the personal homepages of their lecturers and easy access to comprehensive learning resources through learning-management systems.

The National Survey of Student Engagement (NSSE) reports on the levels of engagement of large numbers of universities and colleges across the US. The initial findings confirm the pattern of declining engagement on many fronts, for example, "First-year students on average

reported only occasional direct contact (once or twice a month) with their teachers" (Kuh, 2001: 13). The purpose of the NSSE is to provide benchmark and diagnostic information to individual institutions that will assist them to improve practice. It is quite deliberately limited in scope to identifying those aspects of the student experience that can be addressed directly by institutions.

In Australia, the major conclusions with respect to engagement from the first-year studies (Coates, 2005) have recently been supported by a systematic and comprehensive analysis of open-ended comments from graduates in a government funded annual survey of graduates. The study confirmed that it is students' total experience of university – not just what happens in the traditional classroom – that shapes their judgments of quality, promotes retention, and engages them in productive learning (Scott, 2005: vii).

Diversity in the Student Experience

Studies on the whole continue to work from the assumption that undergraduate education is primarily concerned with the transition of individuals from adolescence to adulthood. This does not reflect the reality of declining populations of young people in many developed countries, the expansion of lifelong learning opportunities provided by universities, and the subsequent shift in student profile toward mature age and part-time students. Notwithstanding this widely acknowledged observation, the literature, government policy, and public debate persist in presenting undergraduate education as synonymous with young full-time students. Harvey and Drew (2006) make the point with respect to the first-year experience that there is a multiplicity of experiences contingent on type of institution and student characteristics.

There are at least four major subgroups of students that should be considered in thinking about diversity in the student experience: traditional domestic school leavers (residential halls and commuter); traditional international school leavers; deferred entry under 25 years age; and, mature-age return-to-study over 25 years age. The evidence is clear: the 18-year-old school leaver, the 23-year-old delayed-entry student, and the 35-year-old return-to-study students are different in significant respects in terms of their approaches to study, and their expectations of university (McInnis *et al.*, 2000). Students who take up university places even after just a few years of work experience are markedly different in the ways they experience university life. Their expectations, aspirations, and sense of identity vary significantly from those who go straight from school to university. The experience of each group is qualitatively different and the outcomes vary accordingly. Although the groups have a great deal in common as a consequence of their shared experience, it is a mistake to

assume homogeneity in terms of outcomes and attributes. They have not simply arrived at the same endpoint by different paths: their experience is qualitatively different with distinctive patterns of time and space in their connection to the university. The evidence base for the efficacy of the social dimension is quite strong and generally convincing with respect to young full-time undergraduates, but there remains far less understanding of the impact of the social dimensions of university on mature-age students.

While there is a compelling case for treating young undergraduates as a key focus of research and policy related to the student experience, it should be recognized that being a student no longer has the transitional and marginal status it once had when students were a relatively small minority of the late adolescent population who went to university before they entered the real world. A European Commission project titled 'Students as Journeymen between Communities of Higher Education and Work' illustrates the persistence of the ideal type of undergraduate experience shaping current research (EC, 2004). However, students are now less different from the mainstream society simply because, by virtue of their numbers in mass higher education systems, they are the mainstream of their generation. The values, behaviors, and outlooks of students are now much more in alignment with those of the society in which they live.

Moreover, the traditional divide of full- and part-time enrolment has become blurred for many undergraduates. Likewise, there is pressure on universities from governments and universities to redraw the traditional academic calendar. The trend to concurrent college-vocational education enrolment in the US, and reflected in some respects in the Australian experience, as well as the popularity of combined and double degree programs, means that students in many systems are in effect creating their own collage of university-study programs.

Here lies the central challenge for the organization of undergraduate education. Where once it was assumed that students would naturally form peer-support groups and make regular informal contact with faculty, it is now clear that the mix of part-time work, idiosyncratic timetables, and the accessibility of web-based resources requires lecturers and course designers to create learning experiences that encourage students to develop informal networks. Students are now highly strategic in the choices they make (McInnis, 2004).

Rapid advances in communications technologies have increased options students have for configuring their approaches to learning. Indeed, separating the two is entirely superficial. It is difficult to predict the impact of technology change on the student experience in the long term. Even the most enthusiastic observers of the impact of technological change, consider that a 4–5-year time-to-adoption perspective is about as far as they can reasonably

go in their predictions (Educause, 2006). New challenges emerge with every major innovation. In the current environment, the impact of instant messaging, pod casting, and social-networking websites require universities to manage multiple ways of communicating with students to reconfigure and promote the collective experience of learning. While the new media forms embraced by young students can complement their engagement with universities, they add another layer of social networks independent of the university that have the potential to compete with it.

Patterns of Convergence in the Student Experience

There is no reason why the student experience should be immune from the impact of globalization. The increase in the numbers of students combining full-time study with paid work is recognized internationally. Similarly, most developed systems are placing high priority on students taking part of their undergraduate program in another country. Many are actively encouraging students to include substantial periods of time in the workplace during their studies. Combined and dual-degree programs have proliferated in response to student demand to broaden their employment options. These widely shared patterns of change in the student experience suggest considerable convergence that will be accelerated by the globalization of quality-assurance processes, and a changed emphasis away from institutional inputs and toward what students actually do in the process of learning.

The global shift in high levels of accountability for universities has placed responsibility for the management of the student experience squarely with institutions. The capacity of universities to respond to the changing needs of undergraduate students, and the desirability or efficacy of doing so, is a significant issue both for governments and universities. For many universities, providing a quality experience beyond classroom instruction is a matter of legitimacy and survival; for others, it is not a significant consideration. While much of the current research and scholarly discussion is directed toward measuring, monitoring, and managing the student experience, notable work is also in progress on the impact of these changes on institutional responses, adaptation, and student outcomes.

The theoretical and conceptual frameworks that shape models of the student experience are currently being reassessed and overhauled. The focus of research is shifting away from formal academic and institutional processes to the relatively unknown and sometimes competing influences of the informal or unsystematized student experience. Understanding these experiences and how they might be harnessed to enhance new forms of undergraduate education will continue to be a major focus for higher education research.

See also: Promoting Effective Student Learning in Higher Education; What Do University Students Learn?.

Bibliography

- Astin, A. (1998). The changing American college student: Thirty year trends, 1966–1996. *Review of Higher Education* 21(2), 115–135.
- Brennan, J. and Jary, D. (2005). What is learned at university? *SOMUL Working Paper 1* (May). The Higher Education Academy and Open University/CHERI. <http://www.open.ac.uk/cheri>.
- Brennan, J. and Osborne, M. (2005). The organisational mediation of university learning. *SOMUL Working Paper 2* (December). The Higher Education Academy and Open University/CHERI. <http://www.open.ac.uk/cheri> (accessed June 2009).
- Clark, B. R. (1973). Development of the sociology of higher education. *Sociology of Higher Education* 46, 2–14.
- Coates, H. (2005). The value of student engagement for higher education quality assurance. *Quality in Higher Education* 11(1), 25–36.
- Educause (2006). *The Horizon Report 2006*. The New Media Consortium. <http://www.nmc.org> (accessed June 2009).
- European Commission (EC) (2004). *Students as Journeymen between Communities of Higher Education and Work*. HPSE CT-2001-00068. <http://www.hewl.net> (accessed June 2009).
- Euro Student (2004). *Social and Economic Conditions of Student Life in Europe*. Hanover: Higher Education Information System.
- Harvey, L. and Drew, S. (2006). The first year experience: A review of literature for the Higher Education Academy. <http://www.heacademy.ac.uk> (accessed June 2009).
- HERI (2007). *The American Freshman – National Norms for 2006* (Press Release, Research Summary). Higher Education Research Institute, University of California, Los Angeles. <http://www.gseis.ucla.edu/heri/heri.html> (accessed June 2009).
- Houston, M. and Lebeau, Y. (2006). The social mediation of university learning. *SOMUL Working Paper 3*, Higher Education Academy. The Higher Education Academy and Open University/CHERI. <http://www.open.ac.uk/cheri>.
- Krause, K., Hartley, R., James, R., and McInnis, C. (2005). *The First Year Experience in Australian Universities: Findings from a Decade of National Studies*. Canberra: DEST.
- Kuh, G. (2001). Assessing what really matters of student learning. *Change* May/June, 10–17.
- McInnis, C. (2001). Researching the first year experience: Where to from here? *Higher Education Research and Development* 20(2), 105–114.
- McInnis, C. (2002). Signs of disengagement: Responding to the changing work patterns of full-time undergraduates in Australian universities. In Enders, J. and Fulton, O. (eds.) *Higher Education in a Globalising World*, pp 175–189. London: Elsevier.
- McInnis, C. (2004). Studies of student life: An overview. *European Journal of Education* 39(4), 383–394.
- McInnis, C. and Hartley, R. (2002). *Managing Work and Study*. Canberra: AGPS.
- McInnis, C. and James, R. (1995). *First Year on Campus: Diversity in the Initial Experiences of Australian Undergraduates*. Canberra: AGPS.
- McInnis, C., James, R., and Hartley, R. (2000). *Trends in the First Year Experience in Australia*. Canberra: AGPS.
- National Observatory of Student Life (OVE) <http://www.ovenational.education.fr/>
- Pascarella, E. and Terenzini, P. (2005). *How College Affects Students*. San Francisco, CA: Jossey-Bass.
- Scott, G. (2005). *Accessing the Student Voice: Using CE Query to Identify What Retains Students and Promotes Engagement in Productive Learning in Australian Higher Education*. Canberra: DEST.
- Terenzini, P. and Reason, R. (2006). *Parsing the First Year of College: Rethinking the Effects of College on Students*. Centre for the Study of Higher Education, The Pennsylvania State University.
- Tinto, V. (1998). Colleges as communities: Taking research on student persistence seriously. *Review of Higher Education* 21(2), 167–177.
- Yorke, M. and Longden, B. (2006). The first-year experience survey. Higher Education Academy. <http://www.heacademy.ac.uk> (accessed June 2009).

Further Reading

- Kuh, G. (1998). How are we doing? Tracking the quality of the undergraduate experience from the 1960s to the present. *Review of Higher Education* 21, 2.
- Kuh, G. (2003). What we're learning about student engagement from NSSE. *Change* March/April, 24–32.
- Pascarella, E. and Terenzini, P. (1998). Studying college students in the 21st century: Meeting new challenges. *Review of Higher Education* 21, 151–165.
- Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The course experience questionnaire. *Studies in Higher Education* 16, 129–150.
- Yorke, M. and Longden, B. (eds.) (2004). *Retaining Students in Higher Education*. Buckingham: Open University Press.

Relevant Websites

- <http://www.deewr.gov.au> – Australian Government, Department of Education, Employment and Workplace Relations.
- <http://www.eair.nl> – EAIR: The European Higher Education Society, SIGS Student.
- <http://www.lfhe.ac.uk> – Leadership Foundation for Higher Education.
- <http://www.ovenational.education.fr> – National Observatory of Student Life (OVE).
- <http://www.ed.psu.edu/cshe> – Pennsylvania State University College of Education, Center for the Study of Higher Education.
- <http://www.srhe.ac.uk> – Society for Research into Higher Education, Student Experience Network.
- <http://www.heacademy.ac.uk> – The Higher Education Academy.
- <http://www.open.ac.uk/cheri/index.htm> – The Open University, Cheri.
- <http://www.gseis.ucla.edu> – University of California Los Angeles, Graduate School of Education and Information Studies, Higher Education of Research Education, Findings.

What Do University Students Learn?

G D Kuh, Indiana University Bloomington, Bloomington, IN, USA

B K Bridges, American Council on Education, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

This article summarizes some of the more important investigations into the impact of postsecondary education on desired learning and personal development outcomes. We draw on summaries done by Bok (2006), Kuh *et al.* (1994, 2006), and especially Pascarella and Terenzini (2005) – who produced the most recent definitive volume on how university postsecondary education affects student development. The vast majority of the studies reviewed by these scholars were done in the US. Economic and market forces are shaping curricular design and educational goals throughout the world (Yorke and Knight, 2006). With this in mind, we also report findings – when available – on student learning in other countries and research on study abroad.

We focus primarily on studies that take into account the influence of different types of institutional conditions, such as selectivity, and differences in learning and personal development for students with different characteristics such as race and ethnicity, gender, and socioeconomic status. These distinctions are important because many of the effects of postsecondary education are conditional and indirect – meaning that some students gain more than others and their participation in certain activities mediates (or makes possible), but does not contribute directly to, certain outcomes. For example, living on campus does not by itself contribute to desirable outcomes, but it creates opportunities for other experiences that do contribute directly to these outcomes, such as interacting more frequently with faculty members and peers. We use the terms college and university interchangeably to denote postsecondary educational enrolment eligible for baccalaureate degree credit, whether at a 2- or 4-year institution.

Learning and Personal Development Outcomes

To organize the review of learning and personal development outcomes associated with college attendance, we use the five outcome domains in **Figure 1** distilled by Kuh (1993), which encompass the range of skills, competencies, and abilities addressed in most studies on this topic.

Cognitive Complexity

The development of cognitive skills and competencies is one of the most important intended outcomes of university study. Developing intellectual and practical skills – particularly critical and creative thinking – has

never been more important than when the pace of change escalates in almost every dimension of human endeavor (AAC&U, 2007). The cognitive complexity domain consists of two outcome categories – reflective judgment and application of knowledge, abilities needed to think critically and to evaluate logically or assess the quality of one's own thinking and experience by exercising independent judgment.

Students become more critical, analytical, reflective, and intellectually flexible in their thinking in college (Baxter Magolda, 2001). Institutions that emphasize close relationships and feature high levels of student–faculty contact tend to foster greater gains in critical thinking, analytical competencies, and general intellectual development (Pascarella and Terenzini, 2005). Students who spend more time with faculty outside of class also exhibit the greatest gains in cognitive outcomes, such as comprehension, interpretation, evaluation, and application abilities. Those who engage more in such areas as intellectual, vocational, athletic, political, or social activities typically gain more in abstract, analytical, and evaluative thinking than those who are less involved. In general, students reporting greater gains in cognitive development are those who (1) perceive faculty as being concerned with teaching and student development, (2) have developed a close, influential relationship with at least one faculty or staff member, and (3) report that their peers have had an important influence on their development.

Gains in cognitive development linked to out-of-class experiences tend to be related to the amount of effort students expend in educationally purposeful activities, such as service-learning, studying or talking with peers and faculty about substantive matters, or other issues related to their studies such as paper topics or career aspirations. Students who work off campus tend to benefit in similar ways as those who work on campus or who do not work during the first year of college (Pascarella *et al.*, 1994).

Students who live on campus usually gain more in critical thinking than those who commute. Living on campus may not directly affect desired outcomes; rather, it is a mediating factor because it puts students in more regular contact with peers and faculty. Some residential environments are more powerful than others. For example, first-year students in living–learning residences exhibit significantly greater gains in cognitive development compared with their counterparts in other types of campus residences.

African-American students attending historically Black colleges or universities (HBCUs) and predominantly

- *Cognitive complexity*: cognitive skills including reflective thought, critical thinking (e.g., ability to summarize information accurately and perceiving logical coherences and discernable themes and patterns across different sources of information), quantitative reasoning, and intellectual flexibility (i.e., openness to new ideas and different points of view).
- *Knowledge acquisition and application*: understanding knowledge from a range of disciplines and physical, geographic, economic, political, religious, and cultural realities, and the ability to relate knowledge to daily life including using information presented in one class in other classes or other areas of life.
- *Humanitarianism*: an understanding and appreciation of human differences including an increased sensitivity to the needs of others.
- *Interpersonal and intrapersonal competence*: a coherent, integrated constellation of personal attributes (e.g., identity, self-esteem, confidence, integrity, appreciation for the esthetic and spiritual qualities of life and the natural world, sense of civic responsibility) and skills (e.g., how to work with people different from oneself).
- *Practical competence*: skills reflecting an enhanced capacity to manage one's personal affairs (e.g., time management, decision making), to be economically self-sufficient, and to be vocationally competent.

Figure 1 Outcome domains associated with college attendance. From Kuh, G. D. (1993). In their own words: What students learn outside the classroom. *American Educational Research Journal* 30(2), 277–304.

White institutions (PWIs) perform comparably on standardized measures of reading comprehension, mathematics, critical thinking, and composite success measures. Some research shows that African-American students at HBCUs report greater intellectual gains than their counterparts at PWIs (DeSousa and Kuh, 1996; Fleming, 1984) and interact more frequently with their faculty (Pascarella and Terenzini, 2005). These findings corroborate the widespread perception that HBCUs provide more supportive learning environments for their students.

Institutional mission and culture shape campus environments – which, in turn, are associated with intellectual development to varying degrees. For example, although attending an academically selective institution has a negligible impact on general cognitive development, students attending small, private, liberal arts colleges more frequently reported gains in cognitive complexity, ostensibly fueled by their engagement in educationally purposeful academic activities, and with peers and by the institutional ethos (Kuh, 1993, 1995). In contrast, participating in student organizations and other out-of-class activities had a negative effect on cognitive outcomes and the grade point average (GPA). The conflicting findings may be due to campus-specific differences; that is, some institutions are more effective than others in structuring learning opportunities to encourage students to integrate what they learn in class with their lives outside the classroom. Such activities seem to be especially important for African-American students and older, part-time students in terms of cognitive development.

Knowledge Acquisition and Academic Skills

The outcomes in this domain are primarily a function of course-related learning and the amount of effort students devote to their studies and independent research. At the same time, engaging in certain kinds of out-of-class activities also contributes to knowledge acquisition – with men and women benefiting to a comparable degree (Kuh, 1993, 1995).

Students learn more when their instructors use effective educational practices, such as providing feedback and inducing students to engage in active and collaborative learning activities – both inside and outside the classroom. Learning is also fostered when assignments and class activities complement students' preferred learning styles, and when faculty and peers positively reinforce student effort directed toward educationally purposeful activities. In addition, the more time students spend studying, the more they learn; because full-time students have more time to devote to their studies, they typically acquire more knowledge in college. This is generally true for all students – men and women, adult learners, traditional-age students, students from historically underserved minority groups, distance learners, and so forth (Astin, 1993; National Survey of Student Engagement, 2006).

Students gain substantially in subject-matter knowledge, verbal and quantitative skills, and oral and written communication – especially in those subjects connected with their major (Pascarella and Terenzini, 2005). For example, natural science and engineering majors report

greater gains in scientific and quantitative reasoning than humanities and social science majors. As expected, full-time students gain more than part-time students; and black students tend to report greater gains than Asian students (Horn and Ethington, 2002).

As with cognitive complexity, where one attends university has a modest effect on knowledge and academic skill acquisition. Students attending small, private liberal arts colleges more frequently tend to report greater gains in knowledge and academic skills attributable more often to classroom, laboratory, and studio activities than to out-of-class experiences (Kuh, 1993), while attending an academically selective institution seems to have only a negligible effect on knowledge acquisition (Pascarella and Terenzini, 2005).

On balance, first- and second-generation students benefit to comparable degrees as to how much they learn, after controlling for student background characteristics and engagement during college. When differences exist in terms of academic skill development, they tend to be in mathematics, science, and foreign languages (Chen, 2005) with first-generation students gaining less; these differences may be due more to individual student characteristics, such as educational aspirations, and college experiences – such as living off-campus – than their first-generation status (Pike and Kuh, 2005). Intercollegiate athletes – especially men's football and basketball players – appear to gain less in reading comprehension and mathematical problem solving compared with other students (Pascarella and Terenzini, 2005). Working (on campus, off campus, or not at all) does not seem to matter in terms of gains in reading comprehension or mathematics, at least during the first year of college. Students who study abroad learn a substantial amount about the culture, politics, and society of their host country (Carlson *et al.*, 1991; Oppen *et al.*, 1990).

As engagement increases, so do gains in literacy – especially prose and document literacy. The number of analytical courses completed also affects gains in literacy with relationships being stronger for students at 2-year colleges than their counterparts attending 4-year schools. Other activities that show a consistent pattern between engagement and literacy are asking questions in class, frequently receiving prompt feedback, and perceiving the institution emphasizes studying and academic work and provides support to succeed academically (Baer *et al.*, 2006). Taken together, these findings suggest that students at 2-year schools marked by challenging classroom activities and a success-oriented campus climate gain more in literacy.

Humanitarianism

The two outcome categories within this domain – altruism and estheticism – represent interests and actions that advance the welfare of others – especially for people

from diverse backgrounds – and appreciation for the arts. These sensibilities and habits of the mind and heart that prepare students to engage in civic affairs and work effectively with people from diverse backgrounds have never been more important than in the twenty-first-century global economy (AAC&U, 2007).

As they near graduation, students are more likely to support gender equality and are more likely to be tolerant of the political, social, and religious views of others compared to their views when they started college (Pascarella and Terenzini, 2005). Among the factors associated with these changes are living on campus, participating in cultural awareness workshops, undertaking social leadership activities, and perceiving that their institution places an emphasis on diversity and multiculturalism. In addition, faculty attitudes and values along with peer interactions consistently and positively influence changes in a range of sociopolitical and civic attitudes and behaviors. For example, the frequency with which a student volunteers in college predicts similar behavior following college. Student interactions with other students who are different ethnically or culturally have a strong, positive effect on cultural awareness. Interacting with faculty members is also positively correlated with promoting racial understanding, as are voluntarism and socializing with persons from different racial/ethnic groups.

Race and gender are significant, consistent predictors of student involvement in humanitarian efforts, which is linked to promoting racial understanding and an appreciation for diversity. Women tend to engage more frequently in civic pursuits compared with men, who are more likely to be aware of current events.

Students who live or spend time with others from different racial and ethnic backgrounds gain in appreciation for the aesthetic qualities of life and sensitivity to, and appreciation of, human differences. Other university experiences linked to these outcomes are enrolling in ethnic studies courses, attending racial awareness workshops, engaging in cross-racial interactions, and participating in community service. In addition, exposure to people with diverse perspectives and interaction with people who have more advanced stages of moral reasoning, such as discussions between first-year students and senior students or faculty members or staff, and work-related experiences appear to enhance moral reasoning abilities. Finally, students from the US who study abroad tend to become more civic-minded, more committed to international issues, and become more appreciative and aware of cultural differences. International students studying in the US tend to benefit in terms of general education outcomes as well as personal and social development.

Community-college students evidence greater gains than their counterparts at 4-year institutions in becoming more open to intellectual and racial/ethnic diversity (Pascarella and Terenzini, 2005). Students attending small,

private liberal arts colleges more frequently reported changes in altruism and estheticism (Kuh, 1993).

The relatively homogeneous environment of HBCUs does not seem to inhibit the development of humanitarianism in African-American students (Pascarella and Terenzini, 2005), even though they tend not to report as many experiences with diversity (Umbach and Kuh, 2006). Attending an HBCU also appears to have an impact on civic orientation as their graduates are typically much more likely to participate in community service activities (Redd, 2000). Compared with their counterparts at PWIs, HBCU African-American students report more contact with faculty, a greater belief that their institutions contribute to their spiritual growth, and a greater sense that their campus experience has contributed to their community involvement and civic engagement (National Survey of Student Engagement, 2004). Students at HBCUs also report greater increases in self-understanding and the likelihood that they will vote.

Interpersonal and Intrapersonal Competence

This domain consists of five attributes considered indispensable to living a meaningful, self-regulating, fulfilling life. They are self-awareness, autonomy, confidence, social competence, and sense of purpose. These outcomes contribute to one's sense of self-worth and well-being, and affect the quality of interactions with others.

Educationally purposeful out-of-class experiences such as peer interactions, work experiences, and meaningful leadership activities are linked with positive social self-esteem, self-confidence, and other aspects of personal development – with women typically gaining more in self-esteem than men; this may be, in part, because of a ceiling effect with men scoring higher on such measures when they start college. Similarly, experience in a paraprofessional role, such as tutoring or being a residence hall assistant, is related to gains in self-confidence, self-awareness, skills in interpersonal communication, and awareness of group dynamics. Attending an institution with large numbers of students from different racial and ethnic backgrounds is also related to enhanced social self-concept. Service-learning courses help students clarify and define their identities and strengthen their self-esteem, internal locus of control, and interpersonal skills. Living on campus, taking diversity courses, helping teach a course, working with a faculty member on a research project, and being involved in collaborative group projects are all positively linked to varying degrees with academic self-concept, self-esteem, and self-directed behavior (Terenzini *et al.*, 1996).

Student-faculty interaction beyond the classroom is positively correlated with personal growth in the areas of leadership, social activism, and intellectual self-esteem, and academic as well as social self-concept. Such interactions include the hours per week spent talking with faculty

outside of class, assisting faculty in teaching a class, working on a professor's research project, and being a guest in a professor's home.

Students of color tend to report making greater gains in personal and social development compared with white students. African-American students at HBCUs tend to be more confident than their counterparts who attend PWIs (Fleming, 1984; Watson and Kuh, 1996). Because African-American students attending PWIs have to devote more effort to maintaining supportive social relationships due to their minority status, they tend to gain more in areas such as interpersonal skills. African-Americans at HBCUs also appear to benefit more in terms of desired outcomes of college from course-related experiences than out-of-class engagement. Hispanic community-college students reported gaining more in communication skills and personal and social development compared to their Asian counterparts, and greater gains in personal and social development compared to Caucasian students.

Practical Competence

Practical competence represents students' capacity to perform effectively in and after college in a variety of areas. This domain includes the acquisition of time-management skills and other attitudes and competencies, such as interpersonal communication, group process, teamwork, decision making, and understanding and demonstrating sensitivity to workplace culture – needed to manage one's own affairs and perform well in post-college employment settings. Employers and policy-makers are increasingly interested in this arena, stating that while students are well-prepared in their major field, many lack the skills and abilities needed to be successful in the workplace. Although practical competencies can be obtained in classrooms, laboratories, and studios, the nature of many out-of-class activities often requires that students become competent in these areas, demanding that students examine and test their skills and values in a variety of situations not unlike those they will encounter following college. For example, experiencing diversity and voluntarism and community service in particular are linked with student-reported gains in competence or the belief that their college experience provided good preparation for work.

Educationally purposeful out-of-class experiences such as voluntarism, community service, and holding office in student government or other organizations contribute positively to leadership development, decision-making skills, and feelings of personal competence. Although the research is mixed as to whether leadership experiences in student government or other organizations has a direct effect on career choice, some studies suggest that extra-curricular activities may positively affect career aspirations

(especially for African-American students) and career mobility. Although the influence of co-curricular student involvement on career development is inconsistent, college graduates think such activities are important to their success after college; that is, college graduates typically refer to leadership roles as being important to the development of practical competence and later achievements. However, it is likely that other variables (personality, motivation, etc.) may be more important in explaining such effects on practical competence (Pascarella and Terenzini, 2005).

Involvement in intercollegiate athletics also shows mixed effects in relation to occupational status. Some research shows that athletic participation was related to occupational status attainment for African-American men and had a positive indirect effect on occupational status for Caucasian men, after controlling for race, socioeconomic background, occupational aspirations, college grades, and educational attainment. Other studies found negligible and statistically nonsignificant effects when comparing athletes and nonathletes.

Working during college – particularly in a job related to one's major or vocational goal – is related to such practical competencies as decision making and time management as well as subsequent career attainment.

Student–faculty informal contacts outside the classroom appear to significantly influence career choice, career interest, and eventual career selection. Relationships with faculty members had meaningful and positive effects for female students on self-reported growth in career-related skills. Interactions with faculty members positively affected self-reports of career preparation for 2-year college students, particularly those in trade and industry fields. Student–faculty interactions also appear to positively influence the likelihood of students choosing careers in academic and scientific research. However, other aspects of practical competence such as decision-making and time-management do not seem to be influenced by contact with faculty.

Conclusion

On average, a wide range of benefits are associated with postsecondary education for everyone – men and women, students from various racial and ethnic backgrounds, international students, residential and commuter students at 2- and 4-year colleges, and so forth. These effects are cumulative and mutually shaping, with cognitive growth influenced by a variety of experiences and conditions on a campus, particularly when out-of-class climates and experiences complement and encourage students to integrate what they learn in class with their lives outside the classroom. In addition, out-of-class activities that impact the development of cognitive skills may also impact the

development of ethical and moral reasoning abilities. More specifically:

- Institutional type influences development of personal values.
- Peer interactions are the single most powerful influence on students' academic and personal development.
- Interacting with faculty is positively associated with persistence, practical competence, and other measures of success and desired outcomes.
- Community college students have different goals and outcomes than their counterparts at 4-year institutions.
- Student effort and engagement in educationally purposeful activities have a cumulative positive effect on values development, cognitive development, and life-long learning outcomes.
- Students who participate in study abroad benefit in a number of ways, including enhanced cultural awareness and civic involvement.

See also: Employability of University Graduates and Graduate Outcomes; Facilitating Students' Success – Student Support and Learning Resources in Universities; Promoting Effective Student Learning in Higher Education; University Students' Experiences of Higher Education.

Bibliography

- Association of American Colleges and Universities (AAC&U). (2007). *College Learning for the New Global Century*. Washington, DC: Author.
- Astin, A. W. (1993). *What Matters in College? Four Critical Years Revisited*, 1st edn. San Francisco, CA: Jossey-Bass.
- Baer, J. D., Cook, A. L., and Baldi, S. (2006). *The Literacy of American's College Students*. Washington, DC: American Institutes for Research.
- Baxter Magolda, M. B. (2001). *Making Their Own Way: Narratives for Transforming Higher Education to Promote Self-Development*. Sterling, VA: Stylus.
- Bok, D. (2006). *Our Underachieving Colleges*. Princeton, NJ: Princeton University Press.
- Carlson, J., Burn, B., Useem, J., and Yachimovicz, D. (1991). *The Experience of American Undergraduates in Western Europe and the United States*. New York: Council on International Educational Exchange.
- Chen, X. (2005). *First Generation Students in Postsecondary Education: A Look at Their College Transcripts*. (NCES 2005-171). U.S. Department of Education. Washington, DC: US Government Printing Office.
- DeSousa, J. D. and Kuh, G. D. (1996). Does institutional racial composition make a difference in what Black students gain from college? *Journal of College Student Development* **37**(3), 257–267.
- Fleming, J. (1984). *Blacks in College: A Comparative Study of Students' Success in Black and White Institutions*, 1st edn. San Francisco, CA: Jossey-Bass.
- Horn, R. A. and Ethington, C. A. (2002). Self-reported beliefs of community college students regarding their growth and development: Ethnic and enrollment status differences. *Community College Journal of Research and Practice* **26**(5), 401–413.
- Kuh, G. D. (1993). In their own words: What students learn outside the classroom. *American Educational Research Journal* **30**(2), 277–304.

- Kuh, G. D. (1995). The other curriculum: Out-of-class experiences associated with student learning and personal development. *Journal of Higher Education* **66**(2), 123–155.
- Kuh, G. D., Douglas, K. B., Lund, J. P., and Ramin-Gurnek, J. (1994). Student learning outside the classroom: Transcending artificial boundaries. *ASHE-ERIC Higher Education Report No. 8*. Washington, DC: The George Washington University, Graduate School of Education and Human Development.
- Kuh, G. D., Kinzie, J., Buckley, J., Bridges, B., and Hayek, J. C. (2006). What matters to student success: A review of the literature. *Final Report for the National Postsecondary Education Cooperative and National Center for Education Statistics*. Bloomington, IN: Indiana University Center for Postsecondary Research. <http://nces.ed.gov/npec/papers.asp> (accessed August 2009).
- National Survey of Student Engagement (2006). *Engaged Learning: Fostering Success of All Students*. Bloomington, IN: Indiana University Center for Postsecondary Research.
- Opper, S., Teichler, U., and Carlson, J. (1990). *Impacts of Study Abroad Programmes on Students and Graduates*. London: Jessica Kingsley.
- Pascarella, E. T., Bohr, L., Nora, A., Desler, M., and Zusman, B. (1994). Impacts of on-campus and off-campus work on first-year cognitive outcomes. *Journal of College Student Development* **35**(5), 364–370.
- Pascarella, E. T. and Terenzini, P. T. (2005). *How College Affects Students: A Third Decade of Research*. San Francisco, CA: Jossey-Bass.
- Pike, G. R. and Kuh, G. D. (2005). First- and Second-Generation College Students: A Comparison of Their Engagement and Intellectual Development. *Journal of Higher Education* **76**(3), 276–300.
- Redd, K. E. (2000). *HBCU Graduates, Employment Earnings and Success after College*. Indianapolis, IN: USA Groups.
- Terenzini, P. T., Pascarella, E. T., and Blimling, G. S. (1996). Students' out-of-class experiences and their influence on learning and cognitive development: A literature review. *Journal of College Student Development* **37**(2), 149–162.
- Umbach, P. D. and Kuh, G. D. (2006). Student experiences with diversity at liberal arts colleges: Another claim for distinctiveness. *Journal of Higher Education* **77**, 169–192.
- Watson, L. W. and Kuh, G. D. (1996). The influence of dominant race environments on student involvement, perceptions, and educational gains: A look at historically Black and predominantly White liberal arts institutions. *Journal of College Student Development* **37**(4), 415–424.
- Yorke, M. and Knight, P. T. (2006). Curricula for economic and social gain. *Higher Education* **51**, 565–588.

Further Reading

- Association of American Colleges and Universities (AACU). (2005). *Liberal Education Outcomes: A Preliminary Report on Student Achievement in College*. Washington, DC: Author.
- Astin, A. W. (1977). *Four Critical Years: Effects of College on Beliefs Attitudes, and Knowledge*. San Francisco, CA: Jossey-Bass.
- Bowen, H. (1977). *Investment in Learning: The Individual and Social Value of American Higher Education*. San Francisco, CA: Jossey-Bass.
- Chickering, A. W. and Gamson, Z. F. (eds.) (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin* **March**, 3–7.
- Education Commission of the States (ECS) (1995). *Making Quality Count in Undergraduate Education*. Denver, CO: ECS Distribution Center.
- Flowers, L. A. and Pascarella, E. T. (1999). Does college racial composition influence the openness to diversity of African-American students? *Journal of College Student Development* **40**(4), 405–417.
- Hoachlander, G., Sikora, A. C., and Horn, L. (2003). *Community College Students: Goals, Academic Preparation, and Outcomes*. (NCES 2003-164). U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- Kuh, G. D. (2003). What we're learning about student engagement from NSSE: Benchmarks for effective educational practices. *Change* **35**(2), 24–32.
- Kuh, G. D., Kinzie, J., Buckley, J., Bridges, B., and Hayek, J. C. (2007). Piecing together the student success puzzle: Research, propositions, and recommendations. *ASHE Higher Education Report* **32**(5). San Francisco, CA: Jossey-Bass.

HIGHER EDUCATION – TERTIARY EDUCATION IN A GLOBALIZED WORLD

Contents

Associations of Universities
Collective Bargaining in Higher Education
Financing of Higher Education
Higher Education Crossing Borders
Higher Education in Post-Conflict Societies
Higher Education in the New Economies
Human Resource Issues in Higher Education
International Accreditation in Higher Education
Internationalization of Higher Education
National and International Rankings of Higher Education
Steering of Higher Education Systems – The Role of the State
Student and Faculty Transnational Mobility in Higher Education
The Bologna Process in European Higher Education
The Economics and Finance of Higher Education
The Role of the OECD in the Development of Higher Education in a Globalized World
Transforming Higher Education in Developing Countries: The Role of the World Bank
UNESCO's Role in the Development of Higher Education in a Globalized World

Associations of Universities

G F Moodie, Griffith University, Brisbane, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

This article considers national associations of universities, national subgroups of universities, university consortia or associations of universities to fulfill a closely defined purpose, and international associations of universities. While these groups are very different in their specific purpose and organization, they are all associations of whole institutions to further their institutional interests and often a broader social purpose. Notwithstanding the breadth of the article, several important types of academic associations remain beyond its scope. The article does not cover various types of disciplinary associations such as the American Society for Engineering Education, Chinese Society for Engineering Education, Conference of European Schools for Advanced Engineering Education and Research, Engineering Professors' Council, Indian Society for Technical Education, International Federation of Engineering Education Societies, and the Société Européenne pour la Formation des Ingénieurs. Neither does the article cover learned academies such as the Royal Swedish Academy of Sciences (which awards the Nobel prizes for chemistry, economics, and physics), the Royal Society (the national academy of science of the UK and the Commonwealth), the US National Academy of Sciences, and the International Council for Science.

National Associations of Universities

National associations of universities have been established in several countries. Many important characteristics such as their age, scope, legal status, size, and role reflect the country in which they are established. For example, one of the earliest national associations of universities, the *Österreichische Universitätenkonferenz* (Universities Austria), was established in 1911 as the rectors' conference of higher education institutions in the Austro-Hungarian Empire. During much of its history, Universities Austria was a statutory corporation, an institution of public law, but at other times it has been a nonprofit association under private law, its current legal status. Higher Education South Africa (no date) describes itself as “the unified body of leadership in a transforming, dynamic and diverse system of higher education in South Africa,” reflecting the government's restructuring of higher education as well as South Africa's general political environment.

Notwithstanding their important differences, national associations of universities have some characteristics in common. Many originated as informal meetings of university leaders. Thus, the Association of Universities and

Colleges of Canada evolved from a meeting of university presidents in 1911 to discuss issues they planned to raise at a meeting of the Congress of Empire Universities which was to be held the following year in London. Most have comprehensive if not complete membership of the institutions within their scope, which are usually either universities, or, more broadly, higher education institutions in their country. Many have roles similar to those of the *Hochschulrektorenkonferenz* (German Rectors' Conference) and the *Finlands Universitetsrektorers Råd* (Finnish Council of University Rectors) of formulating and presenting common positions to government and the public, and advancing cooperation between their members.

Some national associations of universities have distinctive roles in addition to the roles commonly discharged by such associations. Thus, the New Zealand Vice-Chancellors' Committee is responsible for institutional quality assurance and for approval, accreditation, and moderation of university programs. The *Vereniging van Universiteiten* (Association of Universities in the Netherlands) is the universities' employers' organization. The American Council on Education mounts senior managers' development programs, and the Association of Universities and Colleges of Canada administers more than 150 scholarship, fellowship, and internship programs on behalf of governments, foundations, and private companies.

National and Regional Subgroups of Universities

Some national associations of universities circumscribe their scope and membership criteria, leaving other higher education institutions to be represented by other bodies. Thus, France has both the *Conférence des Présidents d'Université* and the *Conférence des Grandes Ecoles* (conference of professional academies), and the Netherlands has both the *Vereniging van Universiteiten* (Association of Universities in the Netherlands) and the HBO-raad (Netherlands Association of Universities of Applied Sciences). Other countries have both a comprehensive universities association and various subgroups of universities to pursue more specific interests than those represented by the comprehensive national association of universities. An illustrative example is the UK.

The UK has a comprehensive national association of universities and colleges of higher education, Universities UK. Universities UK originated in the nineteenth century from informal meetings of several university vice chancellors and principals of university colleges. The first meeting of all vice chancellors was held in 1918 and a standing Committee of Vice-Chancellors and Principals was established in 1930. In 1992, the UK Government dismantled the binary divide between universities and polytechnics, allowing polytechnics to become universities. The Committee of Vice-Chancellors and Principals

admitted the newly designated universities as members, doubling in size and becoming much more diverse in membership.

The attempt by the UK Committee of Vice-Chancellors and Principals to represent the diverse interests of its newly expanded and diverse membership soon led to the establishment of subgroups of universities to advance more specific interests. The 20 big research-intensive universities formed the Russell Group in 1994 and later the same year, 17 of the smaller research-intensive universities formed an association, calling itself the 1994 group. The 29 newer or modern universities established as universities since 1992 formed themselves into a group now known as Million+. This left a diverse mixture of pre- and post-1992 universities without a subgroup affiliation and in 2006, some 24 such universities formed themselves into a group now known as the University Alliance.

While these UK subgroups were formed mainly to represent different interests to the national government, at the same time, another collection of subgroups was formed, mainly to promote higher education and institutional cooperation in their region. Examples of these regional subgroups include Universities for the North East (of England) established in 1983, East Midlands Universities Association (1998), West Midlands Higher Education Association, and London Higher. There is in addition a national subgroup of UK universities based on a special function. The Council of Validating Universities was formed in 1982 to promote good practice and standards of awarding institutions' validation of their programs offered by other institutions, called collaborative provision by the Council but known more generally in the UK as franchising. The Council of Validating Universities has 78 members, ranging across all the other subgroups.

The US has numerous diverse national and regional subgroups of universities. The American Council on Education is the comprehensive representative body of all types of accredited, degree-granting institutions in the US. It was founded in 1918. Several national university subgroups were established before then. The US's oldest higher education association is the National Association of State Universities and Land-Grant Colleges, which originated in 1887 and now promotes public higher education. The Association of American Universities was founded in 1900 to promote the standing and acceptance of their members' doctoral degrees among European universities. The Lutheran Educational Conference of North America was established in 1910 and the Association of American Colleges and Universities was founded in 1915 to promote undergraduate liberal education.

The most prominent subgroup of US universities is the Ivy League, which is an athletic conference or sports competition among eight private higher education institutions in the Northeastern United States. Its members are Brown University, Columbia University, Cornell

University, Dartmouth College, Harvard University, Princeton University, University of Pennsylvania, and Yale University. While these colleges had competed against each other informally since 1900, the first Ivy Group Agreement was not signed until 1945 and covered only football. The Ivy League now sponsors conference championships in 33 men's and women's sports. While the Ivy League is an athletic conference, the term is also used to refer to its members which are some of the most elite and oldest colleges in US, seven of its members having been established before the start of the American War of Independence in 1775.

The US is distinctive in that its most authoritative accreditation of education from kindergarten to doctoral level is conducted by six regional associations. The oldest such body, the New England Association of Schools and Colleges, Inc., was founded in 1885. It accredits more than 2000 public and private schools, colleges, and universities in the six states of the New England region. While regional accrediting associations are now established as independent corporations, they originated as associations of universities and colleges which remain strongly represented on their governing bodies.

Other regional US university associations range greatly in geographic coverage. A number of associations cover several states, such as the Associated Colleges of the South which comprises 16 liberal arts colleges and universities across 12 states in the US south. Some associations cover a region within a state, such as the Colleges of Worcester Consortium which comprises 13 public and private colleges and universities located in central Massachusetts. The Associated Colleges of the Twin Cities is a group of five private liberal arts colleges located in either Minneapolis or Saint Paul, Minnesota. The Atlanta Regional Council for Higher Education comprises 19 public and private colleges and universities in one city, Atlanta, Georgia. There are several associations of universities in parts of cities, such as Rochester Area Colleges which comprises 18 higher education institutions in Rochester, New York. Perhaps the regional association with the smallest geographic coverage is the Claremont Colleges, a group of five undergraduate and two graduate schools of higher education located in Claremont, California. The Claremont College campuses are adjoining and within reasonable walking distance of one another, covering roughly one square mile or 260 hectares.

University Consortia

University consortia are associations of universities to fulfill a closely defined purpose. The expansion of information and communication technologies has stimulated the formation of several university consortia to offer programs online, often in combination with other study modes. The Great Plains Interactive Distance Education

Alliance was established in 1994 by ten western and Mid-western US universities to offer graduate programs in human services by distance education and online. Students enrol in an alliance program through one of the partner institutions. The revenue-sharing formula provides for 12.5% of the student fee to be retained by the admitting or home university to cover student records and services, 75% is passed to the teaching university, and 12.5% of the fee is allocated to the alliance to support joint activities such as promotion and maintaining its website. Students graduate with a degree from their home institution.

Western Governors University is a prominent cooperative online university in the US, having been founded in 1997 and enrolling 10 000 students in 2008. However, it is perhaps not strictly a university consortium since it was founded and is governed by 19 state governors. University Alliance involves nine US universities offering certificate, bachelor, and masters programs online through a system operated by Bisk Education. Bisk Education's system includes instructional designers who support teachers at the participating universities to develop online materials, media production to record and edit teachers' classes in video and audio, a course management system, and marketing and student support.

An example of an international university consortium is the online university U21Global. U21Global was formed in 2001 by Universitas 21 and a private partner. Universitas 21 was in turn founded in 1997 and now has 21 members in 13 countries. U21Global offers masters programs in business management and information-technology management. Programs and their subjects are reviewed and accredited by U21pedagogica, a wholly owned subsidiary of Universitas 21. Degree certificates awarded by U21Global bear the crests of Universitas 21's members.

International Associations of Universities

Kirkland (2004: 49) observes that the development of the Association of Commonwealth Universities, the oldest international association of universities, prefigures the development of many other such associations. The Association of Commonwealth Universities originated in 1903 when universities of the then British Empire met in the then premises of the Royal Society in Burlington House, London. The ensuing interest encouraged the University of London to initiate a congress of universities of the Empire which was held in London in 1912 and attended by representatives of 53 universities. They formed the Universities Bureau of the British Empire in 1913. The Bureau was a federation founded on sentiment according to Ashby (1988: 27 (1963)). Its membership was and remains both voluntary and based on subscriptions. It grew to 100 members by 1953, almost 400 by 1994, and it now has some 500 members.

As Kirkland (2004: 52) notes, many of the interests of the participants in the precursor and earliest meetings of the Association of Commonwealth Universities remain salient now. Participants in the 1903 meeting discussed student and staff mobility among other matters of common interest. What is now the *Commonwealth Universities Yearbook* originated in 1912 and is now published as an online database as well as a book of over 2500 pages. The initial aims of the Universities Bureau of the British Empire set out in its constitution of 1919, included collecting and disseminating information, both about university members and about topics of shared interest, holding conferences, and facilitating the exchange of students and teachers. In time, the Association established a regular information bulletin for members.

The Association of Commonwealth Universities is like numerous other international university associations in having developed from an informal meeting of heads of institutions which share an interest. The common interest of the Association of Commonwealth Universities is based on political geography. This is in common with a third of 50 international associations of universities, such as the *Unión de Universidades de América Latina* (founded 1949), Association of African Universities (1967), Inter-American Organization for Higher Education (1979), *Communauté des universités méditerranéennes* (1983), and the Circumpolar Universities Association (1989). Many other international associations of universities have goals similar to the Association of Commonwealth Universities of sharing information and facilitating staff and student exchange which they likewise advance by holding regular conferences and occasional workshops, publishing newsletters, conducting research, and publishing occasional reports. The Association of Commonwealth Universities is, however, unusual in having been founded at least 50 years before most other international university associations.

One of the most important international bodies, the International Association of Universities, was first proposed in 1948 by the government of the Netherlands and the United Nations Educational, Scientific and Cultural Organization (UNESCO). Since its foundation in 1950, the association has grown to over 580 higher education institutions. The association holds conferences, runs staff-development programs, and publishes a journal *Higher Education Policy*. The association has also published the *International Handbook of Universities* since 1959. It now contains entries for 17 000 higher education institutions published in a print directory of two volumes and now also online as the world higher education database. The International Association of Universities also importantly issues policy statements, on: sustainable development (1993), funding (1994), academic freedom (1998), internationalization (1998), information technologies (2004), quality higher education across borders (2005), and equitable access (2008).

Complementing the International Association of Universities is the International Association of University Presidents which was founded in 1964 and now has some 600 members. The International Association of University Presidents holds triennial conferences of its whole membership and regional conferences; it convenes working groups on various matters and it publishes a quarterly newsletter.

In the mid-1960s two associations were formed on common cultural interests: the Association of Arab Universities and the International Federation of Catholic Universities. The Hispanic Association of Colleges and Universities is similar to the Association of Arab Universities in being based on common sociocultural interests, but was formed 26 years later in 1986. There are at least three other international university associations based on shared religion: the Council for Christian Colleges & Universities (founded in 1976), the International Council of Universities of Saint Thomas Aquinas (1993), and the International Association for the Promotion of Christian Higher Education (1975).

The number of international associations of universities has exploded since the 1970s. In the 35 years from 1972 to 2007 at least 38 international associations of universities were established, more than one each year. Neither is the pace slackening. In 2006 alone, three associations were formed: the Global U8 Consortium, the EuroMed Permanent University Forum, and the International Alliance of Research Universities. The newer associations continue to follow the pattern set by the Association of Commonwealth Universities almost a century ago. Many like the Association of Commonwealth Universities are formed at the initiative of a single university, often of an individual. The Inter-American Organization for Higher Education was initiated in 1979 by Gilles Boulet, president of the *Université du Québec*; the *Communauté des universités méditerranéennes* was formed in 1983 following the lead of the *Président de l'Université de Bari* and a group of lecturers; the *Europaeum* was conceived in the early 1990s by George Weidenfeld and Ronald Grierson, and many other examples can be given.

The newer associations also follow the Association of Commonwealth Universities in seeking to share information and promote student and staff exchange, and most undertake much the same activities: holding conferences and publishing newsletters and occasional papers. We have seen that many are also based on a shared political geography. Common cultural interests have also continued to stimulate the foundation of new international associations of universities, with newer associations being based on a common religion, and others on a common language, such as the *Associação das Universidades de Língua Portuguesa* (1988) and *L'Agence universitaire de la Francophonie* (1989).

Some associations have been founded on a different principle: status, almost always based on claims of research distinction. For example, Consortium Linking Universities of Science and Technology for Education and Research

(CLUSTER) (founded in 1990) describes itself as 12 leading universities of science and technology working together; the Association of East Asian research universities (1996) describes itself as a forum for leading research-oriented universities in East Asia; the League of European Research Universities (2005: slide 5) notes in its promotional presentation that it was founded in 2002 by a group of 12 European universities with the objective of creating a common policy forum of top research universities in Europe (it has since expanded to 20 universities); the Worldwide Universities Network (2003) says that it is a partnership of 16 research-led universities from Europe, North America, South East Asia, and Australia; and the International Alliance of Research Universities (2006) seeks to be the premier alliance of universities in the world.

International university associations differ greatly in size, falling into three groups. There are at least eight associations with 400 or more members. One of the biggest associations is the European Association of Institutions in Higher Education which was founded in 1990 and now has 830 members. This association's central interest is occupational education and its members include all types of higher education institutions. Another very big association with 775 members is the European University Association. This association is the representative body for European universities, most importantly recently in the Bologna process. Other big associations are *L'Agence universitaire de la Francophonie* (692 members), Magna Charta Observatory of Fundamental University Values and Rights (622 members), International Association of Universities (580 members), Association of Commonwealth Universities (500 members), and the Inter-American Organization for Higher Education (400 members).

The second group has 100–200 members. Two of the bigger associations in this group are based on religion: the International Federation of Catholic Universities (200 members) and the Council for Christian Colleges & Universities (181 members). One of the smaller associations in this group is based on language: *Associação das Universidades de Língua Portuguesa* (117 members). Most of the other universities in this group are based on political geography, such as the *Asociación Universitaria Iberoamericana de Postgrado* (Latin American Association of Postgraduate Universities) which was formed in 1998 and has 119 members.

The third-sized group of associations of universities has from 10 to around 60 members. Some of these smaller associations were established before the explosion: the *Unión de Universidades de América Latina* (established in 1949), the Association of Arab Universities (1964), and the Women's College Coalition (1972). As with the other-sized groups, many smaller associations are founded on political geography, such as the *Asociación de Universidades Grupo Montevideo* (founded in 1992) and the Association of Pacific Rim Universities and Øresund University, both of which were founded in 1997.

Analysis

Whetten (1981) offers a useful theoretical analysis of inter-organizational relations. Whetten (1981: 2–4) notes that interorganizational relations have been analyzed from four distinct research traditions. Public administration researchers seek to improve coordination within a service-delivery system, such as health. More specifically, they examine lateral coordination of agencies at the local level. In contrast, marketers examine vertical links between businesses in channels of distribution between suppliers of raw materials, manufacturers, wholesalers, retailers, and customers. Some economists study the relations between businesses in a sector and the relations between sectors, positing industry clusters and innovation systems. The aim of this research is to determine the conditions for forming clusters and fostering innovation. And sociologists study the formation and dynamics of groups and organizations to understand their power relations. Most analyses of associations of universities are in the sociological tradition, but the other perspectives could usefully be applied to universities' associations with other organizations, both within the sector and between sectors.

An early analysis of international associations of universities was offered by Neave (1992). Neave (1992: 52) posits five stages along a continuum of international links between higher education institutions: (1) monodisciplinary bilateral links, (2) exchange partnerships, (3) network partnerships, (4) multidisciplinary networks, and (5) consortia. It will be noted that most university associations considered in this article are of the last three of Neave's types. Other analysts have followed Neave's lead and developed typologies of university associations (Wächter, 2000; de Wit, 2002; Beerkens, 2002). de Wit (2004: 34–36) distinguished between academic associations which are loose associations most commonly established to advance interest in a discipline, academic consortia which are groups formed to fulfill a contract, and institutional networks which are groups established to advance multiple purposes.

Future

Whetten (1981: 15) posits five conditions for organizations' voluntary coordination. A positive attitude to coordination may be established by a cosmopolitan ethos, strong shared background and values, or rewards, such as may be provided by research grants or special government grants. Organizations may see the need to coordinate with others on which they partly depend or with which they share goals, and Whetten (1981: 17) found that organizations are likely to perceive partial interdependence if they have broad goals and provide diverse services to a wide range of clients. Organizations may become aware of prospective partners from informal contact, geographic

proximity, or formal communication. Organizations assess the compatibility and desirability of potential partners by their status congruity, compatible ideology and definition of problems, consensus over domain or the allocation of roles, complementary organizational structures and procedures, and goal compatibility. Whetten's fifth condition of capacity to maintain the coordination process depends on adequate resources and staff, adequate communication channels, and flexible rules and procedures.

These conditions are satisfied for many universities; hence, their associations of universities seem to have a strong future. This is borne out by their dynamism. Well-established associations continue to be active, many are still growing in size and activities, and numerous new associations continue to be formed. There are approximately 9,60 universities in the world; so if, as some argue, international associations are formed in response to globalization which affects most universities (de Wit, 2004: 29), there is considerable scope for the formation of new associations and the considerable expansion of associations that seek comprehensive membership. In 2004, Kansas State University established an institute for academic alliances to study and support inter-institutional academic program development, implementation, and management. Associations may be further encouraged and strengthened by sharing information on their operation in international meetings of associations of universities convened by the International Association of Universities. The first such meeting was held in 2005, the second in 2007, and the third in 2009.

On the other hand, as Teather (2004: 1) observes, "Some international networking alliances and consortia of universities started with a media fanfare, but have not lived up to their founders' expectations. They now limp along with a low level of activity, or have disbanded altogether." Some, perhaps many, associations are formed by universities responding to similar conditions. A change in universities' environment or international conditions may result in the disbandment of some associations, or perhaps their amalgamation or consolidation into overarching associations, as various associations of European universities have done. The associations of disparate universities formed on the initiative of an individual seem most vulnerable to withering once the initial champion leaves his/her current position or his/her enthusiasm wanes.

See also: Internationalization of Higher Education; Student and Faculty Transnational Mobility in Higher Education.

Bibliography

Ashby, E. (1988) (1963). *Community of Universities: An Informal Portrait of the Association of Universities of the British Commonwealth 1913–1963*. London: Association of Commonwealth Universities.

- Beerens, E. (2002). International inter-organisational arrangements in higher education: Towards a typology. *Tertiary Education and Management* 8, 297–314.
- de Wit, H. (2002). *Internationalisation of Higher Education in the United States of America and Europe: A Historical, Comparative and Conceptual Analysis*. Westport, CT: Greenwood Press.
- de Wit, H. (2004). Academic alliances and networks: A new internationalisation strategy in response to the globalisation of our societies. In Teather, D. C. B. (ed.) *Consortia: International Networking Alliances of Universities*, pp 28–48. Carlton: Melbourne University Press.
- Kirkland, J. (2004). Change and continuity: The case of the Association of Commonwealth Universities. In Teather, D. C. B. (ed.) *Consortia: International Networking Alliances of Universities*, pp 49–68. Carlton: Melbourne University Press.
- League of European Research Universities (2005). General presentation of LERU and its members. <http://www.leru.org> (accessed August 2009).
- Neave, G. (1992). Institutional management of higher education: Trends, need and strategies for cooperation, *unpublished reference document*. International Association of Universities, Paris, 27 February.
- Teather, D. C. B. (2004). Introduction. In Teather, D. C. B. (ed.) *Consortia: International Networking Alliances of Universities*, pp 1–7. Carlton: Melbourne University Press.
- Wächter, B. (2000). *Handbook of European Associations in Higher Education*. Bonn: Lemmens.
- Whetten, D. (1981). Interorganizational relations: A review of the field. *Journal of Higher Education* 52(1), 1–28.

Further Reading

- Association of Commonwealth Universities (2005). What is the Policy & Research Unit? <http://www.acu.ac.uk> (accessed August 2009).
- Organisation for Economic Co-operation and Development (2004). *Internationalisation and Trade in Higher Education: Opportunities and Challenges*. <http://www.oecd.org> (accessed August 2009).

Relevant Websites

- <http://www.auf.org> – Agence universitaire de la Francophonie.
- <http://www.aaru.edu.jo> – Association of Arab Universities.
- <http://www.acu.ac.uk> – Association of Commonwealth Universities.
- <http://www.asaihl.asia> – Association of Southeast Asian Institutions of Higher Learning.
- <http://auap.sut.ac.th> – Association of Universities of Asia and the Pacific.
- <http://www.eurashe.be> – European Association of Institutions in Higher Education.
- <http://www.eua.be> – European University Association.
- <http://www.hesa.org.za> – Higher Education South Africa.
- <http://www.hacu.net> – Hispanic Association of Colleges and Universities.
- <http://www.oui-iohe.qc.ca> – Inter-American Organization for Higher Education.
- <http://www.iaruni.org> – International Alliance of Research Universities (2008).
- <http://www.iapche.org> – International Association for the Promotion of Christian Higher Education.
- <http://www.unesco.org> – International Association of Universities, UNESCO.
- <http://www.iaups.org> – International Association of University President.
- <http://www.fiuc.org> – International Federation of Catholic Universities.
- <http://www.leru.org> – League of European Research Universities.
- <http://www.magna-charta.org> – Magna Charta Observatory of Fundamental University Values and Rights, Observatory.
- <http://www.universitas21.com> – Universitas 21: About Universitas 21.
- <http://www.wun.ac.uk> – Worldwide Universities Network (2005), About Us.

Collective Bargaining in Higher Education

C Allport, K McAlpine, and S Roberts, National Tertiary Education Union (Australia), South Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Beamte – A civil servant.

Beihilfe – A subsidy or economic assistance.

Fachhochschule – A university of applied sciences.

Grande Ecole – A higher education establishment outside the mainstream framework of the public universities system.

Kunsthochschule – A university of applied sciences focusing on art.

Musikhochschule – A university of applied sciences focusing on music.

Polytechnique – A technical institute.

Technikon – A university of technology.

Universität – A university.

Higher education teaching and research personnel increasingly work in an international environment, and there are international standards that are generally expected by United Nations Educational, Scientific, and Cultural Organization (UNESCO) under its Recommendation on the Status of Higher Education Teaching Personnel, passed by its general conference in 1997. These standards specifically highlight the importance of academic freedom and institutional autonomy as core standards that are expected of all higher education institutions. In like vein, there are also International Labour Organization (ILO) conventions which, while not specific to higher education, also support the rights of higher education staff to collective bargaining. In general, these standards have been seen to be necessary in order to ensure that all higher education institutions are able to assure freedom of expression, and maintain an independent voice without government suppression. In some countries, these freedoms are defined in national or regional legislation, but for others, it is the right to collectively bargain that provides a solid foundation for the roles and rights held by higher education teaching and research personnel. In New Zealand, the Education Act specifically speaks about the role of universities as ‘critic and conscience’ of society.

Higher education teaching and research personnel also have rights and responsibilities in their teaching, research, and public commentary roles. In order to ensure that academic freedom is not stifled, such staff retain the right to participate in academic associations, including trade

unions. The UNESCO recommendation sees these freedoms as essential in the maintenance of academic freedom.

27. Higher education teaching personnel are entitled to the maintaining of academic freedom, that is to say, the right without constriction by prescribed doctrine, to freedom of teaching and discussion, freedom in carrying out research and disseminating and publishing the results thereof, freedom to express freely their opinion about the institution or system in which they work, freedom from institutional censorship and freedom to participate in professional or representative academic bodies. All higher education teaching personnel should have the right to fulfill their functions without discrimination of any kind and without fear of repression by the state or any other source.

However, over the last decade or so, higher education has become both more internationalized and more connected to the market. Governments have instituted reforms of their higher education systems that have had the effect of propelling higher education institutions to engage in academic capitalism. Throughout the world, academics are now expected to become involved in the market, and to search for increasing private income to sustain the core public work of teaching and research. This tendency, referred to in the Singapore context as ‘jumping into the sea’, pervades many higher education institutions and has been a powerful transformative force. The impact of embracing the market has been profound and has led to an unprecedented expansion of private higher education institutions, including private operations of public universities from the developed world. It is in these contexts that collective bargaining has become more important as a tool for the protection and enforcement of rights articulated in the 1997 UNESCO recommendation.

Introduction to Bargaining Systems

All higher education systems or institutions employ academic/faculty and other staff. Therefore, in each case the following questions must be asked.

- What are the terms and conditions of employment?
- How are they to be determined?

In higher education, as in other spheres of voluntary employment, these questions are determined in one of the following ways:

- By direct state regulation. The state may be the employer and use its legal power to directly prescribe conditions of employment, or it may empower an employer unilaterally to determine employment conditions.
- By individual contract between each employer and employee, usually subject to a regime of minimum legal conditions.
- By a system of binding arbitration by state agencies. These can range from a genuinely independent quasi-judicial tribunal, to *de facto* political determination.
- By collective bargaining.

Collective bargaining refers to negotiations between employees, through representatives which are usually labor unions, and an employer or group of employers, with a view to establishing agreement, usually with legal force, on working conditions and terms of employment, and regulating the relations between the relevant employer(s) and the employees and their representatives.

The right to collective bargaining is enshrined in a number of conventions of the ILO, which is the agency of the United Nations charged with the promotion of internationally recognized human and labor rights. The right to collective bargaining and the right to take strike action in pursuit of bargaining claims are acknowledged as core labor standards.

In the specific context of higher education, the UNESCO recommendation states in part:

52. Higher education teaching personnel should enjoy the right to freedom of association, and this right should be effectively promoted. Collective bargaining or an equivalent procedure should be promoted in accordance with the standards of the International Labour Organization (ILO) set out in the appendix.

53. Salaries, working conditions and all matters related to the terms and conditions of employment of higher education teaching personnel should be determined through a voluntary process of negotiation between organizations representing higher education teaching personnel and the employers of higher education teaching personnel, except where other equivalent procedures are provided that are consistent with international standards.

Nevertheless, there are a number of countries in which there is no right to collective bargaining in the general law or where such rights are in practice unavailable due to repression by the state or the absence of the rule of law in labor relations. A list of such countries includes Brunei, Burma (Myanmar), Peoples' Republic of China, North Korea (DPRK), Laos, Maldives, Oman, Qatar, Saudi Arabia, Syria, Djibouti, Libya, Sudan, Togo, and Cuba. Moreover, in countries and territories including Afghanistan, Somalia, Democratic Republic of the Congo and Haiti, Iraq, and Western Sahara, civil strife renders collective bargaining and/or higher education inoperative.

In many other countries, laws preclude employees and/or employers in higher education from collective bargaining, most commonly as part of exclusions of varying classes of public employment or those engaged in essential services, which often include education.

Country Analysis

This brief survey of collective bargaining, discussed in the following sections, concentrates its attention only on a selection of countries. While such a selection is necessarily arbitrary, it includes many of the highly developed and the largest national sectors, as well as other countries representing the diversity of higher education, and demonstrates how collective bargaining operates in different constitutional, cultural, and economic settings across six continents.

United States of America

The strongest and perhaps most diverse higher education sector is that found in the United States of America. There are large public university systems established by state governments, although employees are rarely direct employees of the state, as well as many strong private universities.

Collective bargaining developed gradually during the twentieth century, but was boosted greatly by a 1970 decision which found that universities were engaged in interstate trade and commerce, bringing them within the ambit of the National Labor Relations Act (NLRA), which provides the opportunity for collective bargaining. However, a Supreme Court decision in 1980 declared that most full-time academic staff were managerial employees and were therefore excluded from the NLRA. This major setback to collective bargaining has been offset by state legislation providing collective bargaining rights to employees in public universities in some states, and by private universities voluntarily agreeing to bargain with academic employees.

The result of these legislative changes is that collective bargaining has been widespread among nonacademic staff and among adjunct and sessional staff, including those who work as contingent (or casual) staff. On the other hand, among academic staff, depending on definitions, about 40% of full-time academic staff in the USA in 1998 were covered by collective bargaining arrangements, with the figure being around 50% in the public sector. In recent years, separate bargaining among graduate assistants has become a feature, with some bitter strikes and disputes, particularly about union recognition, as well as legal disputes about whether graduate assistants are employees or students.

Although fragmented to the level of each institution, and often broken into different bargaining units representing academic staff, part-time academic staff, and nonacademic

staff, academic bargaining is coordinated to a limited extent by the three national organizations representing academic staff.

United Kingdom

In the United Kingdom, the public sector dominates higher education, but universities remain separate autonomous employers. Nevertheless, collective bargaining on pay and related matters occurs at a national level between unions and employers for all academic and nonacademic staff, through the Joint Negotiating Committee for Higher Education Staff (JINCHES). The tradition of free collective bargaining in the UK is less formal than in most other countries, with national agreements coexisting with many local arrangements and agreements. None of these requires formal registration nor are they, strictly speaking, enforceable, except by industrial action, which can occur at any time, subject to an employee ballot. In 2006–2007 a new union, the University and College Union was formed, with over 100,000 members, representing all academic staff. This may lead to changes in the way collective bargaining will operate for the future.

France

The state has a dominant role in higher education in France, which has a multilayered system of *grand ecoles*, universities, research institutes, and *polytechniques*. Most university staff in the public system are direct employees of the state, and are generally represented by unions affiliated to one of several major labor federations. The right to strike is guaranteed in the French constitution, and is not limited to disputes over collective bargaining.

University salaries are set by the state, and the tenure enjoyed by most university staff arises directly from their relationship with the state, leading to the unusual position that staff unions tend to oppose increased university autonomy – at least in relation to employment matters. Unions have agitated for agreements with the government about job security, the ratio of tenured to untenured staff, and workloads – which have been affected by rising student/staff ratios – as well as the usual issues of salaries and pension arrangements.

Germany

German higher education institutions consist of three types of *Hochschulen*: *Universitäten* (universities), *Fachhochschulen* (universities of applied sciences), and *Musik- und Kunsthochschulen* (colleges of music and art). Higher education institutions are either state run or state recognized, and all are subject to higher education legislation.

University professors and lecturers are *Beamte*, or civil servants. As *Beamte*, such staff are entitled to reasonable

pay, access to the *Beihilfe* health plan, an indexed pension, and pension insurance.

While German Basic Law guarantees freedom of association and protects the right of unions to exist and to be active, *Beamte* may not conclude collective agreements and their working conditions are unilaterally established through laws rather than through collective bargaining with a union. Such university staff also do not have any right to strike. However, some lecturers employed on contracts by universities do enjoy the right to collective bargaining.

The *Gewerkschaft Erziehung und Wissenschaft* (GEW) teachers' and faculty staff union states that employment conditions are better for university staff in research institutes outside the university system.

Sweden

In Sweden, the public sector dominates higher education, and plays a coordinating role. University staff have the same rights to collective bargaining as other employees. Employees have the right to strike in support of collective bargaining claims and employers have the right to lock-out. As in the rest of the economy, staff are organized into separate union federations of manual, administrative and technical and professional workers. Collective bargaining is organized at various levels, and agreements can be at an industry level, or at the level of the higher education institution. Bargaining in Swedish higher education rarely leads to industrial action and the academic union SULF supports some individually determined salary arrangements. However, tensions have existed over the employment consequences of short-term research funding and the employment status of doctoral students.

Russia

A new labor code (NLC) for Russia was ratified in 2003 amid mass protests, weakening some union rights and making union organization more difficult for smaller unions. Collective bargaining is permitted through the NLC, and employees can walk off the job if their pay is more than 15 days overdue. Employees cannot be terminated for taking subsequent strike action.

To date, despite the existence of a strong trade union representing education employees in the public sector, a mature system of negotiations about salaries and conditions of employment in the dominant state-run university sector is yet to emerge. Unions since the mid-1990s have engaged, however, in many protests against the nonpayment of salaries in education.

West Africa and East Africa

In West Africa, there has been a serious decline in higher education since the 1970s, with expansion in the number

of students but a relatively sharp decline in public resources, particularly as several countries carried out Structural Adjustment Programs (SAP) prescribed by the World Bank. Meager salaries have led many qualified staff and potential staff to seek employment in other sectors of the economy, with many staff also remaining university employees. Authoritarian postcolonial regimes have proscribed collective bargaining either generally in the economy or specifically for university staff and other public sector employees. Nevertheless, academic staff unions in Ghana, Nigeria, *Côte d'Ivoire*, Burkina Faso, Senegal, and Cameroon have all been able to operate independently, and to varying degrees pressure governments, which generally set conditions and salaries by decree.

In East Africa, Kenyan academic staff and successive governments have been engaged in years of bitter industrial disputes over salary levels, without reaching agreement.

Nigeria

In Nigeria, the political turmoil and economic crises of recent decades have seriously affected higher education. These factors and the very low salaries paid to university staff have contributed to the so-called brain drain, with many of the best-qualified staff seeking employment in other countries. Labor relations within the university system, which is public and overseen by the federal government, have been seriously affected by, and contributed to, these problems. Negotiations on pay traditionally occur at the national level, between university unions and the government, although the universities themselves are separate employers. National collective agreements have been the subject of prolonged strike action by academic and non-academic staff in recent years. These disputes and the dismissal of union members have been the subject of repeated complaints to, and adverse comment by, the ILO.

Cameroon

In neighboring Cameroon, political liberalization since the early 1990s has allowed the formation of unions representing university staff. However, it has only been in the face of repeated strike action that, early in this century, the government has been willing to deal with unions on conditions of employment, governance, and academic freedom. University salaries remain fixed by government decree, rather than being a result of formal collective bargaining.

South Africa

With Nigeria, South Africa holds a dominant position in sub-Saharan Africa, and this is reflected in the level of development of its higher education system and its labor

market institutions. Collective bargaining and the right to strike are guaranteed by law and practiced widely in the higher education system, which consists of a dominant public sector of universities and *Technikons*, and a small private sector. Collective bargaining takes place in the context of the transformation of South Africa after apartheid. This involved major growth in enrolments to combat the legacy of apartheid, a redistribution of resources within the sector toward both historically disadvantaged students as well as primary and secondary sectors, in conjunction with outsourcing of many support functions, institutional amalgamations, and the use of market solutions to make up for the shortage of public investment.

Union representation is highly fragmented, particularly among academic staff, and despite the campaigns of some national unions for national collective bargaining, bargaining occurs at an institutional level, with the state not directly participating. Centralized bargaining, and the harmonization of pay rates to which it might lead, were opposed by employers and by some employees and local unions at those institutions where salaries and conditions of work are more favorable.

India

India has a large higher education sector, with over 6 million students and nearly 200 general and specialist universities, as well as thousands of other colleges. Collective bargaining around pay and pensions does not occur directly between the universities and the unions representing university staff. Pay for university staff and other public employees is set by the central (union) government on the recommendation of national pay commissions, which are established from time to time. Six have been established since independence in 1947. Despite this bureaucratic mode for salary determination, higher education staff generally have a legal right to form unions and take strike action.

South Korea

While there is official freedom of association, Korean governments have consistently attempted to suppress union organization. When accepted into the Organization for Economic Cooperation and Development (OECD) in 1996, Korea pledged to bring labor relations into line with international standards. Since then there have been active movements toward recognizing collective bargaining rights for government-employed teaching staff. However, the International Confederation of Free Trade Unions (ICFTU) (now International Trade Union Confederation (ITUC)) says that the position of trade unionists in South Korea is still worse in relation to freedom of association than in any other OECD country, with consistent state repression of union demonstrations and strikes.

Korea's universities have faculty-based unions, although there has been a move toward national unionization. The Korean Federation of Teachers Associations covers all education workers, including university staff, and a new professors' union was established in Korea in 2001. The Korean Professors Union has indicated that it wishes to use collective bargaining as a tool to enshrine greater participation in university governance and intellectual freedom.

Since 1998, teaching staff have been free to form new unions, and academic freedom has improved. However, there is a legislative ban on political activities, strikes, and sabotage in the Teachers Union Act, and infringements of academic freedom still occur regularly. In 2003, Korean Teachers Union leaders were imprisoned for participating in protests against a national online educational data system. In 2004, an academic who publicly called the Korean War North Korea's war of unification was questioned by police, leading to demonstrations by university staff.

Japan

The Japanese constitution enshrines freedom of association and a right to organize and bargain collectively. It also provides a guarantee for academic freedom. Japan has a large higher education sector, in which the most prestigious universities are in the public sector, but in which the private sector dominates numerically. The public sector universities in recent years have been corporatized, and are now expected to raise a far greater proportion of their income from private sources.

The law previously deemed all university staff to be civil servants, who are precluded from collective bargaining rights and who may not go on strike. This absence of a bargaining culture has, however, pervaded both public and private universities and colleges, where local unions representing staff have traditionally lacked influence. It is too early to tell whether the corporatization of the public sector institutions will lead to more fundamental changes.

China

Freedom of association and the right to strike are not recognized in the People's Republic of China. China's All-China Federation of Trade Unions (ACFTU) effectively acts as a wing of government, and includes 16 industrial unions and 30 provincial trade unions, with 91.3% of all employees being members.

The 1994 Labor Law reforms failed to guarantee some basic rights, although enterprise-level collective contracts are now permitted. Minimum conditions of employment, including maximum hours, leave provisions, and occupational health and safety, are set by law, but these minima are often not enforced.

Article 50 of China's Higher Education Law states "the state protects the legitimate rights and interests of teachers and other educational workers of institutions of higher learning, takes measures to improve the working conditions and living conditions of teachers and other educational workers of institutions of higher learning."

Other than this statement, there is little information on the conditions of employment for Chinese workers in higher education and their access to collective bargaining. However, we know that in the 1990s and 2000s significant reforms have been undertaken in higher education, including massification of higher education and decentralization and marketization of higher education institutions. These reforms will inevitably present new challenges for the organization of labor in Chinese higher education institutions.

Malaysia

Malaysia's labor laws are widely acknowledged as being oppressive. The criteria for legal registration of trade unions are restrictive and trade union membership is low, at less than 10% during the 1990s. Collective bargaining (and strike action) is not permitted over selected matters and, in any case, public servants including teaching staff do not have the right to bargain collectively.

Large-scale reform of higher education commenced in the 1990s, with significant corporatization of universities. However, most universities still rely on public funding for survival.

Academics employed by public universities are prohibited from forming a union. Staff associations exist but they cannot perform the full range of functions of a union. However, academics in private tertiary colleges are permitted to form a union if the proposed union wins a ballot of staff, or has over 50% of staff as members.

Indonesia

In the mid-1960s, under President Soekarno, Indonesia had a strong union movement, led by the communist union organization *Sentral Organisasi Buruh Seluruh Indonesia* (SOBSI). Soekarno's fall in 1966 ended this period, and unions were effectively outlawed. In the 1990s, the Ministry of Manpower regulations granted a tightly restricted right to strike and introduced penalties for violating workers' rights. However, the military continued to be involved in labor disputation and union activists were imprisoned and notoriously subjected to human rights abuse.

The 2003 labor legislation, while entrenching freedom of association, failed to guarantee a right to strike (particularly for public servants, including in higher education) and does not protect collective bargaining. In practice, collective bargaining agreements rarely exceed statutory

minimum employment standards, and labor laws are routinely ignored.

There are currently 51 public universities, 26 state/public polytechnics, and 1328 private higher education institutions.

Australia

Australia for many decades had a system of arbitration of wages and conditions by a quasi-judicial commission. This was replaced in the 1990s by a system of enterprise-level collective bargaining with a limited right to strike, which allowed employers to bypass unions in making collective agreements. In 2006, a controversial new law gave legal primacy to statutory individual contracts and further limited the rights of unions. In higher education, the federal government has used funding incentives to require universities to offer these individual contracts and to further restrict collective bargaining rights. These actions have been the subject of complaints to the ILO and UNESCO by the National Tertiary Education Union.

Mexico

The Mexican constitution guarantees the right to strike, but public sector employees may only strike if there has been a systematic abuse of their constitutional rights. Education workers are often employed under civil contracts for providing professional services, under which they are not permitted to join a trade union. Under these contracts (not legally employment contracts), education workers do not have the right to take strike action or to negotiate a collective agreement.

While there is no official collective bargaining in the public sector, unions have used discussions about general working conditions to negotiate unofficial collective agreements. Notice must be given of a strike, and while actual strike action is not common, if held to be legal, the company must close and may not hire replacement workers. If the strike is illegal, the workers can be ordered to return to work immediately.

Chile

The higher education system in Chile is decentralized, and tertiary institutions have a great deal of autonomy from central government. While public sector funding declined in the 1980s (and increased again in the 1990s), Chile's higher education institutions benefit from strong contributions from the private sector.

A new labor code was passed in 2001, including the right to organize and bargain collectively. Public servants do not have the right to strike (although government teaching staff have organized strikes in the past). There is no provision for compulsory arbitration in the public

sector. Rights to collective bargaining exist at the enterprise level, but national collective bargaining is only voluntary.

Argentina

All workers have the right to form unions and the right to strike. The constitution gives unions the right to bargain collectively and to access conciliation and arbitration; however, in 1990, the right to strike in essential industries, including education, was restricted.

In 2002, Argentina's financial crisis led to a virtual freeze in collective bargaining and widespread strikes in the education sector. Teaching staff and government workers protesting the late payment of their salaries in 2002 were dispersed with tear gas and pellets.

Collective agreements at anything over and above the enterprise level require the approval of the Ministry of Labor.

In 2004, a reforming labor law was passed which is expected to bring labor rights in Argentina closer to ILO standards.

Brazil

The 1988 constitution provided for freedom of association and the right to strike, but strikes are not permitted in the public sector. However, in 2002, teachers and university professors joined other unionists in widespread strike action over pay, which was repressed severely by the state. The police used tear gas and truncheons to prevent protesters from entering the Education Ministry buildings.

The government may cancel collective agreements that are not consistent with its wages policy. Bargaining rights were extended from municipal to state and national level through labor law reforms in 2004.

Themes and Issues in Bargaining

Although no trend is universal in higher education, there are a number of issues which appear repeatedly in any world survey of collective bargaining. Many countries still lack basic labor rights, and in others civil society is underdeveloped, or the state is so hostile, that higher education employees have no effective collective representation.

One area of contention is the scope of bargaining. The claims of management prerogative, and of public policy, almost inevitably clash with the claims of employees and unions for the right to bargain about matters which affect the employment relationship. For example, the length of the academic year may be claimed by the state or by management as a question exclusively of educational policy, yet will be seen as central to employment conditions by staff. The law, state policy, and accepted conventions,

all will affect how these issues are resolved, and these vary greatly. For example, in most public universities in those US states which allow bargaining, there are statutory restrictions excluding or limiting the right of unions to bargain about such matters as faculty workloads, the academic year, or course offerings.

In Australia, recent funding policies of the national government have proscribed that there must be no bargaining about the use of precarious employment modes and a range of other matters.

In Malaysia as well, the subject matter that may be included in a collective agreement is severely curtailed, with no proposal for a collective agreement being lawfully able to include matters such as promotion, transfer, appointments to vacancies, termination, dismissal, reinstatement, or assignment of duties. Such matters are considered to be within the purview of managerial prerogative.

A further issue is the relationship between the state, the institutions, and the higher education institution. In many countries across the developing and industrialized world, the state plays a central role. This sometimes involves the state directly regulating salaries, with unions playing what is essentially a lobbying or protest role. In a minority of countries the state generally remains at arm's length from the process, with even public higher education providers seen as autonomous employers. Changes in the role of the state over time, as legislator for education or labor laws or as holder of the purse strings, can reflect the policy of the state or interests of one of the industrial parties, and there is no universal pattern. However, the survey indicates that interventions by the state on behalf of employees or unions are rare.

Collective bargaining in higher education since the 1990s has also regularly seen the intersection, and arguably the contradiction between, some of the traditions of a university and radical changes taking place in the industry. Academic freedom has always been seen as one of the core values of the university. According to organizations of academic staff, the protection of such freedom has generally required a high level of job security, lest staff be laid-off or disciplined arbitrarily. Such discipline may ostensibly be for operational reasons, but in fact may actually be for expressing unpopular views or undertaking research which opposes dominant paradigms or offends funding bodies. This is reflected, for example, in the major policy statements of the American Association of University Professors (AAUP), which stress the importance of tenure of academic staff as a bulwark of academic freedom.

Unions representing faculty staff internationally have sought to use collective bargaining as a means to make this a legal reality. At the same time, since the 1960s, higher education in nearly all countries has undergone rapid growth. Higher education has grown from a small, relatively elite, and, in many countries, self-governing system of teaching and research to its present status as a large

industry providing mass education. Higher education now constantly has to respond to the perceived needs of government economic, labor market, and educational policies, which have been mediated by the increased use of market mechanisms and private contributions from both students (tuition fees) and corporations (to research funding).

In the eyes of many government and university employers, the traditions of career employment for academic and nonacademic staff are often seen as relics of a bygone age, as are the special protections for academic freedom. However, in many countries, existing academic freedom protections are similar to those operating in many associated professions.

These challenges have hit much of the developing world in quite different ways. Beginning in the 1980s, the World Bank encouraged many indebted governments to cut public expenditure and encourage more private provision of education. Whatever the economic merits of these policies, many countries, particularly in Africa, are now suffering a brain drain caused by the steep decline in salaries and resources.

Further Reading

- Allport, C. and Murphy, T. (2003). *Academic Freedom in the Asia-Pacific Region: A Study Undertaken for UNESCO and Education International*.
- American Association of University Professors (2006). *Policy Documents and Reports*, 10th edn. Baltimore, MD: John Hopkins Press.
- Anungwum, E. (2002). Cogs in the wheel: Academic trade unionism, government, and the crisis in tertiary education in Nigeria. *African Studies Review* 45(2), 141–155.
- Benjamin, E. (2006). Faculty bargaining. In Benjamin, E. and Mauer, M. (eds.) *Academic Collective Bargaining*, pp 35. New York: Modern Language Association.
- Borgnas, K. (2006). *Collective Bargaining, Wage Formation and Unemployment in Russia*. Undergraduate Theses Collection, Uppsala University, Uppsala, Sweden.
- Chevallier, T. (2000). French academics between the professions and the civil service. In Altbach, P. G. (ed.) *The Changing Academic Workplace: Comparative Perspectives*, pp 49–75. Boston, MA: Center for International Higher Education.
- Cornell University v National Labor Relations Board, 183, NLRB 329, 1970.
- Education International (2003). *Annual Report*. Brussels: Education International.
- Education International (2004). *Barometer of Human and Trade Union Rights in the Education Sector*. Brussels: Education International.
- European Council/European Union/UNESCO (2000). *National Higher Education System, Germany. Section 8 Diploma Supplement*.
- Hansen, T., Holm-Nielsen, L., and Zúñiga, P. G. (2005). *Human Resources for the Knowledge Economy*. Chile: World Bank.
- Harvey, M. (2006). Graduate employee bargaining. In Benjamin, E. and Mauer, M. (eds.) *Academic Collective Bargaining*, p 139. New York: Modern Language Association and the American Association of University Professors.
- International Centre for Trade Union Rights (2005). *Trade Unions of the World*. Biggleswade: John Harper Publishing.
- International Confederation of Free Trade Unions (2004). *Annual Survey of Violations of Trade Union Rights*. Brussels: International Confederation of Free Trade Unions.

- International Labor Organisation (1948). *Freedom of Association and Protection of the Right to Organise Convention* (No. 87). Geneva: International Labor Organisation.
- International Labor Organisation (1949). *Right to Organise and Collective Bargaining Convention* (No. 98). Geneva: International Labor Organisation.
- International Labour Organisation (1994). *Committee on Freedom of Association Report Case 1820 on the Collective Bargaining Rights of Teachers*. Geneva: International Labour Organisation.
- International Labour Organisation Committee of Experts Report (2006). The Application of Conventions Number 343, vol. LXXXVIX, Series B, no. 3, on the Complaints by the Nigerian Academic Staff Union of Universities (ASUU).
- Kohler, G. (GEW) (2005). *Strengthening the Doctoral Phase: A Trade Union Statement' in Doctoral Programmes for the European Knowledge Society Conference Papers*. Salzburg.
- Konings, P. (2004). Trade-union activism among university teachers during Cameroon's political liberalisation. *Nordic Journal of African Studies* **13**(3), 289–301.
- Lee, M. H. and Gopinathan, S. (2003). Reforming university education in Hong Kong and Singapore. *Higher Education Research and Development* **22**, 177.
- Mok, J. H. K. and Lo, E. H. C. (2002). Marketization and the changing governance in higher education: A comparative study. *Higher Education Management and Policy* **14**, 57.
- National Labour and Economic Development Institute (NALEDI) (2005). *Higher Education Sector Bargaining and Working Conditions: Factors Favouring or Inhibiting Organisation in the HE Sector*. Johannesburg.
- NLRB v Yeshiva University, Case No 444, US 672, 1980.
- Savage, D. C. (2004). *Academic Tenure and Its Functional Equivalent in Post Secondary Education*. Geneva: International Labour Organisation.
- UNESCO (1997). *Recommendation Concerning the Status of Higher-Education Teaching Personnel*. Paris: UNESCO.
- van der Walt, L., Bolsmann, C., Johnson, B., and Martin, L. (2002). *Globalisation and the Outsourced University in South Africa: The Restructuring of the Support Services in Public Sector Universities in South Africa, 1994–2001*. Centre for Higher Education Transformation: University of South Africa.
- Varghese, N. (2004). *Institutional Restructuring in Higher Education in Asia: Trends and Patterns, Policy Forum on Institutional Restructuring in Higher Education in Asia Conference Papers*. UNESCO.
- Zhu, Y. (2001). Economic reform and labour market regulation in China. In Cooney, S., Lindsey, T., Mitchell, R., and Ying, Z. (eds.) *Law and Labour Market in Asia*, pp 157–184. London: Routledge.

Relevant Websites

- <http://www.ei-ie.org> – Education International.
- <http://www.eurofound.europa.eu> – EIRO (European Industrial Relations Observatory).
- <http://www.ilo.org> – International Labour Organization.
- <http://www.oecd.org> – Organization for Economic Cooperation and Development.
- <http://www.seameo.org> – Southeast Asian Ministers of Education Organisation.
- <http://www.sulf.se> – Sveriges Universitetslärarförbund.
- <http://www.unesco.org> – United Nations Educational, Scientific and Cultural Organization.
- <http://www.ucea.ac.uk> – University and Colleges Employers Association.
- <http://www.state.gov> – US Department of State.

Financing of Higher Education

B Chapman, Australian National University, Canberra, ACT, Australia

P Tulip, Board of Governors of the Federal Reserve System, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

There are very many possible ways to finance higher education and this is illustrated in what follows with respect to the systems of a large number of countries. We consider in detail the conceptual issues related to the costs and benefits of different approaches to student loans. Income contingent loans (ICL) are emerging as a mechanism to deliver financial assistance to students, and international reforms in this area are examined in some detail.

Descriptions of Higher Education Financing Systems

International Comparisons of Tertiary Education Expenditure

The financing of higher education varies greatly from country to country. Among the developed economies of the Organization for Economic Co-operation and Development (OECD), spending on higher education institutions averaged 1.4% of the gross national product (GDP) in 2003, which is about the same as a decade ago. Expenditures range from less than 1% of GDP in Italy up to almost 3% of GDP in the USA (see **Figure 1**). The variation in spending reflects large differences in both enrolment rates and in expenditure per student.

On average, 24% of tertiary expenditures in the OECD were financed privately in 2003. This is a marked increase from 1995, when the average was only 19%, although this partly reflects changes in the composition of countries within the sample. In some countries (the United States, Korea, Japan, and Australia), most spending is from private sources (see **Figure 1**). In others, including most European countries, almost all spending is by the government.

Notwithstanding these variations, government subsidies are substantial in every industrialized economy. The rationale for this government support for higher education is controversial. For example, it is often said that tertiary education is a right, not a privilege. However, the same might be said of food or clothing, which governments do not heavily subsidize. A more concrete argument is that subsidies permit children from less-wealthy families to undertake higher education, promoting equality of opportunity. This argument is powerful when the only alternative is the free market. However, as discussed

below, other government policies, such as provision of student loans, can facilitate access without large subsidies. Perhaps the most common argument in the academic literature is that higher education provides benefits to others, not just the student – so it is fair and efficient that society pays a share of the cost. Examples of these benefits or externalities include technological innovation and civic participation. In practice, estimating the size of these benefits is difficult. Therefore, while there is widespread agreement in principle that governments should contribute to some of the costs of higher education, there is also disagreement as to how large that contribution should be.

International Comparisons of Tuition Fees

As might be expected, countries that finance higher education privately tend to have high tuition fees, and this is illustrated in **Figure 2** which also shows differences between private and public institutions. Estimates across countries are not precisely comparable but the ranking of countries is not very sensitive to measurement issues. In particular, tuition is relatively expensive in the United States, Korea, and Australia, and tuition is low, or even free, in most European countries.

Other Costs

The direct education costs described above typically represent a small share of the total cost of attendance. Students also need to pay for room, board, and other living costs. Estimates of these tend to be highly judgmental and are difficult to compare. As one indicator, governments and educational institutions in the United States typically base financial aid packages on the assumption that students need about \$US10,000 to \$11,000 per annum to live on. That compares with an average tuition fee of around \$US8000. In other countries, estimates of both living costs and tuition fees tend to be lower, but the former outstrip the latter by even wider margins. The sum of tuition and living costs (less scholarships and grants, discussed below) represents the amount that needs to be financed by the typical tertiary student.

Students with ready access to finance will recognize that living costs would be incurred regardless of whether or not they study and hence are not relevant to the choice as to whether to enrol. The opportunity cost – the resources that are sacrificed in order to enrol – is foregone income. An

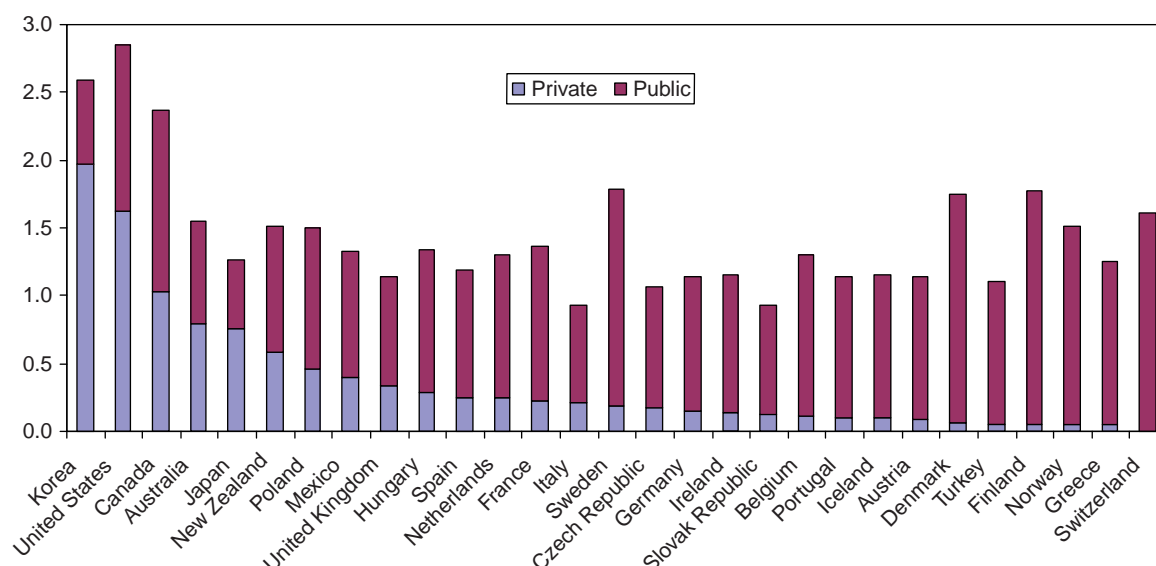


Figure 1 Expenditure on tertiary education institutions: OECD countries percentage of GDP, 2003. Countries are ranked by level of private spending. In Switzerland, private spending data are not available. For further notes and qualifications, see source. From table B2.1b OECD (2006). *Education at a Glance*. Paris: OECD.

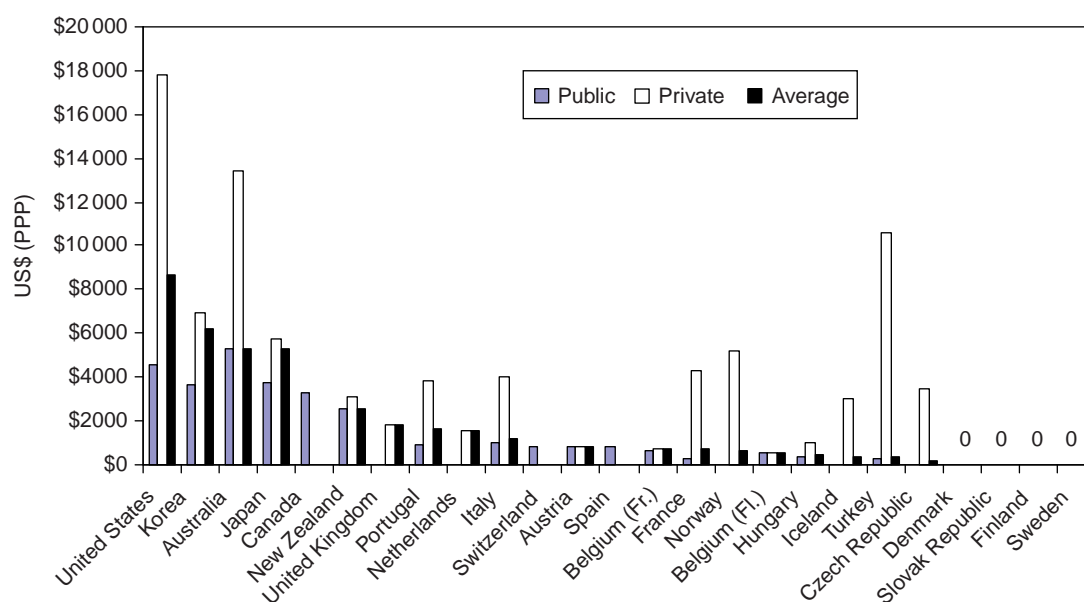


Figure 2 Average tuition fees: tertiary type-A institutions, 2003. Countries are ranked by average tuition fee, weighted by full-time enrolment. Tertiary type-A institutions mainly comprise bachelor's degree programs, and also include masters and professional degrees and some academically oriented associates degrees. Where data are provided as a range, the midpoint is used. For further notes and qualifications, see source. From table B5.1 OECD (2006). *Education at a Glance*. Paris: OECD.

illustrative estimate is the median earnings of a 20-year-old high-school graduate in the United States – around \$22,000 a year. This exceeds the cost of tuition at almost all institutions, a difference that is even greater in other countries. Hence, although public discussion often emphasizes fees as a barrier to education, the bigger disincentive is foregone earnings and the need to finance living costs. Furthermore, although tertiary education is often described as free in

many countries, the size of living costs and foregone earnings means that large contributions are nevertheless required on the part of the student.

Different Kinds of Public Support

The form of public support for higher education also differs widely. In all OECD countries, direct spending

on institutions accounts for the majority of public spending on tertiary education. Indeed, in many countries (e.g., Switzerland, Poland, and Portugal), it is almost the only form of support. Elsewhere, however, substantial support is provided through households in the form of grants and loans. Grants may be based on academic performance (when they are often called scholarships) and/or means testing, either of family or individual income. Loans may be provided directly from the government, or indirectly through a government guarantee of a commercial bank loan. However, increasingly financial assistance from governments is taking the form of an ICL, and this reform is considered in detail in the following sections.

From **Figure 3** it is apparent that New Zealand is the OECD country that relies most heavily on loans, while Denmark relies most on grants. The OECD loan estimates shown represent the gross amount directly lent by government, plus subsidies to private lenders. For some purposes, this overstates the budgetary cost of loans, as repayments are excluded. Meanwhile, it may understate the importance of guaranteed private loans. In principle, it would be desirable to also include tax concessions; however, comparable international estimates are not available.

The relative merits of grants, loans, direct institutional funding, and tax concessions are strongly debated. Researchers have endeavored to measure the costs and benefits of these alternative policies but effects are not always clear. One simple piece of evidence is presented in **Figure 4**, which shows that countries with large loan programs tend to have above-average enrolments. There are several possible explanations for this, including reverse causation. However, the most usual interpretation is that the ready availability of finance facilitates access to higher education.

The following sections consider the conceptual bases of different policy approaches to student loans.

Conceptual Issues Concerning Student Loans

The Need for Student Loans

A significant financing issue for higher education discussed above is that there is a case for both a contribution from students and a taxpayer subsidy. The next important question is: Is there a role for government beyond the provision of the subsidy?

An understanding of the issue is facilitated through consideration of what would happen if there were no higher education financing assistance involving the public sector. That is, a government, convinced that there should be a subsidy, could simply provide the appropriate level of taxpayer support to higher education institutions, and then leave market mechanisms to take their course. Presumably, this would result in the institutions charging students upfront on enrolment for the service.

However, there are major problems with this arrangement, traceable in most instances to the potent presence of risk and uncertainty. This critical point was first raised by Friedman. The argument can be best understood with reference to the nexus between labor markets and human capital investments. The essential point is that educational investments are risky, with the main areas of uncertainty being as follows as discussed by Barr (2001), Palacios (2004), and Chapman (2006a):

1. Enrolling students do not know fully their capacities for (and perhaps even true interest in) the higher

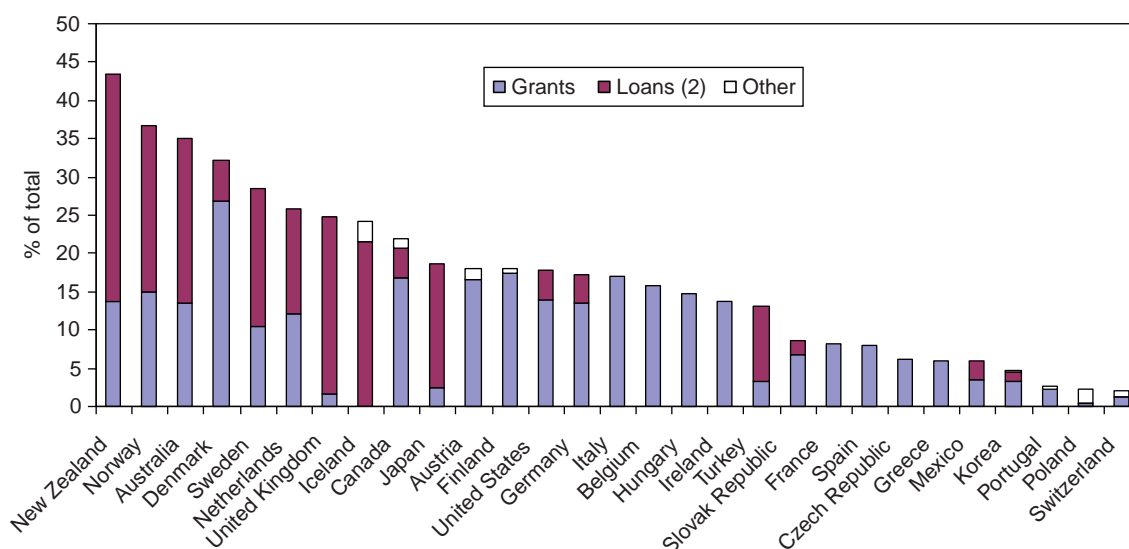


Figure 3 Public subsidies to households as a percentage of total public expenditure on tertiary education (2003). Countries are ranked by share of public spending flowing through households. Loans are measured as gross amount directly lent by government, plus subsidies to private lenders. For further notes and qualifications, see source. From table B5.2 OECD (2006). *Education at a Glance*. Paris: OECD.

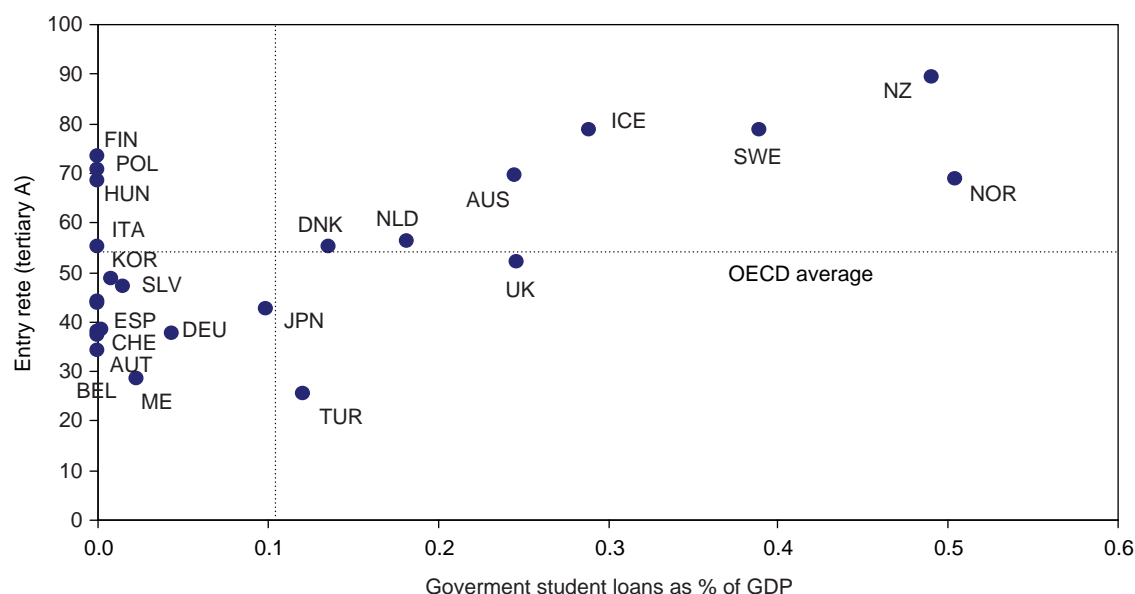


Figure 4 Student loans and entry rates to tertiary education, 2003. The entry rate is the proportion of people entering tertiary type-A programs for the first time. It differs from enrolment rates in that comparability between countries is not distorted by different course lengths. Loans are valued as the gross amount directly lent by government, plus subsidies to private lenders. From table B5.2 OECD (2006). *Education at a Glance*. Paris: OECD.

education discipline of their choice. This means in an extreme they cannot be sure that they will graduate with, in Australia for example, around 25% of students ending up without a qualification.

2. Even given that university completion is expected, students will not be aware of their likely relative success in the area of study. This will depend not just on their own abilities, but also on the skills of others competing for jobs in the area.
3. There is uncertainty concerning the future value of the investment. For example, the labor market – including the labor market for graduates in specific skill areas – is undergoing constant change. What looked like a good investment at the time it began might turn out to be a poor choice when the process is finished.
4. Many prospective students, particularly those from disadvantaged backgrounds, may not have much information concerning graduate incomes, due in part to a lack of contact with graduates.

These uncertainties are associated with important risks for both borrowers and lenders. The important point is that, for example, if the future incomes of students turn out to be lower than expected, the individual is unable to sell part of the investment to refinance a different educational path. For a prospective lender, a bank, the risk is compounded by the reality that in the event of a student borrower defaulting on the loan obligation, there is no available collateral to be sold,

a fact traceable in part to the illegality of slavery. Moreover, even if it was possible for a third party to own and sell human capital, its future value might turn out to be quite low taking into account the above-noted uncertainties associated with higher education investments.

It follows that, left to itself – and even with subsidies from the government to cover the value of externalities – the market will not deliver propitious higher education outcomes. Prospective students judged to be relatively risky, and/or those without loan repayment guarantors, will not be able to access the financial resources required for both the payment of tuition and to cover income support. There would be efficiency losses (talented but poor prospective students would be excluded), and distributional inequities (the nonattainment of equality of educational opportunity). Government intervention of some form is thus required.

The capital market failure with respect to higher education financing is apparently understood by the governments of most countries, given that public sector loan interventions are commonplace internationally. Until recently, government intervention often took the form of public sector guarantees for commercial bank provision of education loans, but over the last decade or so has increasingly involved ICL. While quite different in practice, both approaches are motivated in part by the recognition that, left alone, higher education markets will function poorly.

The Costs and Benefits of Conventional Student Loans

A possible solution to the capital market problem described above, used in many countries, is the provision of student loans – either directly by the government or indirectly through guarantees to banks. Typically, and most simply, these loans involve fixed repayments, as, for example, with a house mortgage. While this seems to address the capital market failure, it raises problems of default.

Students face an important default issue; that is, some may be reluctant to borrow for fear of not being able to meet future repayment obligations. Not being able to meet repayment obligations has the potential to inflict significant damage to a person's credit reputation (and thus access to future borrowing, e.g., for the purchase of a house). These concerns imply that there will be less borrowing than there would be in the absence of this default concern.

A reluctance to borrow due to the uncertainty of repayment constitutes what might be labeled an *ex ante* default problem for prospective students. There is also an *ex post* problem, which is that a proportion of those students who took the credit risk of borrowing for a human capital investment will end up not being able to repay because of low incomes. In these circumstances, default imposes a potentially large cost on those unlucky borrowers who do poorly in the labor market. Significantly, research suggests that members of the default group are predominantly those who ultimately experienced relatively high unemployment rates and relatively low earnings.

The prospect and consequences of a student defaulting on a loan obligation is a potentially critical issue for borrowing to finance human capital investments, due to the uncertainties noted above. A consequence is that some eligible prospective students will not be prepared to take bank loans. This problem can be traced, in part, to the fact that bank loan repayments are insensitive to the borrower's financial circumstances.

The bottom line is that, even though government-assisted conventional loans are a common form internationally of public sector involvement in higher education financing, such an approach has several apparently very significant weaknesses.

The Costs and Benefits of ICL

A final approach to student financing involves income contingent loans, such as Australia's Higher Education Contribution Scheme (HECS), introduced in 1989. The attraction of these schemes is that they can be designed to avoid the problems associated with alternative financing policies.

First, there is no concern with intrafamily sharing so long as the scheme is universal. That is, no student would

be denied access through the imposition of means-testing arrangements that could exclude some whose parents or partners are unwilling to help.

Second, given an efficient collection mechanism, there is no default issue for the government. That is, if the tax system is used to collect the debt (at least for Australia, this is essential because the Australian Taxation Office is the only institution with reasonably good information on a former student's income), it is extremely difficult for the vast majority of graduates to avoid repayment. There is a small default issue in that some students will not pay back in full, because income contingent systems are designed to excuse some former students' payments because their lifetime incomes are too low. Harding (1995) calculates that the total repayments remaining uncollected because of the nature of HECS would be of the order of 20% for the original scheme (when the repayment conditions were much more generous for the student (before the 1996/1997 changes)). Other reasons loans may not be repaid include death and emigration.

Third, because repayments depend on income, there should be no concerns by students with respect to an incapacity to repay the debt. That is, once an individual's income determines repayment, and so long as the repayment parameters are sufficiently generous, it is not possible to default because of a lack of capacity to pay. This is the critical practical advantage of income contingent collection schemes – unlike any other form of assistance, there is insurance against default.

Income contingent schemes have significant advantages over alternative financing arrangements, in that they can be designed to avoid the major problems of their alternatives. This, however, does not make such approaches a panacea generally: for an income contingent scheme to be made operational, it is essential that there is an efficient administrative collection mechanism.

The matter of collection is of great importance for the introduction of ICL in countries without the necessary institutional apparatus. Chapman argues that the minimum conditions for a successful income contingent loan seem to be:

1. accurate record keeping of the accruing liabilities of students;
2. a collection mechanism with a sound, and if possible, a computerized record-keeping system; and
3. an efficient way of determining with accuracy, over time, the actual incomes of former students.

While most OECD countries will have income tax systems that enable efficient collection of income contingent debts, it is very unlikely that developing countries have the capacity to meet the third requirement mentioned above.

International Reforms in Student Loan Systems

Background

Friedman's suggested policy response to the capital market problem involved a type of graduate tax in which former students would repay their debts as a percentage of their incomes for a given period of time. It was not until the 1970s that a more conventional ICL arrangement was first introduced, unsuccessfully, by Yale University. The policy approach received a major boost involving the use of the national income tax system for collection with the introduction of the HECS in Australia in 1989, followed by quite similar policy reforms in other countries. These experiences are now reviewed briefly.

The Yale Plan

Yale University introduced an ICL in 1972, which was extended in 1976 but discontinued several years later. Apart from loans being repaid depending on income, the scheme had the feature of the borrowing being of the form of a group loan, in which there was mutual responsibility between members with respect to the repayment of the total debt. Chapman (2006a, b) categorizes the Yale scheme as a risk-pooling ICL.

Individual repayments were not unlimited, however, with a cap being defined at 150% of the borrower's loan. This then became a buyout option for former students wishing to discontinue in the program (Palacios, 2003). Even so, risk pooling necessarily meant that high-income earners covered the unpaid debts of low-income earners and those who defaulted for other reasons. High initial default rates eventuated, predicted by Nerlove in 1975, because of moral hazard, and the scheme was discontinued. One of the major problems with the Yale scheme was that the university was ill-equipped to act as the collection agency.

Sweden

Prior to 1988, Sweden had a student loan system with fixed repayments and heavily subsidized interest rates. In 1988, repayments were made a flat 4% of income, with less-subsidized interest rates. In 2001, following concern about interaction between these income contingent repayments and the country's already high marginal income tax rates, this scheme was replaced by one in which repayments are initially low, then increase based on a formula that takes into account the students' outstanding debt, the prevailing interest rate, and an annual escalator (Usher, 2005). There is little available evidence of the effect of the scheme.

Australia

As noted, in 1989 Australia instituted a broadly based risk-sharing ICL charging system for higher education, known as the HECS, which seeks to recover a part of tuition costs, and the system does not involve student income support. In Australia, student income support takes the form of means-tested grants. Students incur a debt which is repaid according to future incomes, there being at the time a first threshold of repayment of around average Australian earnings. The Australian Tax Office is the collection agency.

Payments are progressive and, after the debt is incurred, there is a real rate of interest of zero. However, the interest rate regime is more complicated than this, because if a student chooses to pay upfront they (currently) receive a 20% discount. This means that HECS implicitly has a rough form of a real rate of interest, in that those choosing to pay later initially incur a higher level of debt, although the difference obviously reduces in present value terms over time. The discount could be seen to be Australia's way of encouraging what are called parental contributions.

There has by now been considerable research on the effects of the introduction of HECS on a critical issue for policy, the consequences of the scheme for the access of relatively disadvantaged prospective students. The conclusions from the Australian research with respect to socioeconomic mix and access are as follows:

1. the introduction of HECS was associated with aggregate increases in higher education participation; and
2. HECS did not result in decreases in the participation of prospective students from relatively poor families, although the absolute increases were slightly higher for relatively advantaged students.

Both findings raise some important discussion points. With respect to the first, it does not follow that HECS *per se* resulted in an increase in the demand for higher education. Indeed, if this were the case it would constitute a curiosity for economic theory, since the result would suggest that increasing the price of a service increases also the quantity demanded. Indeed, an important finding from many disparate case studies is that the socioeconomic mix of higher education students seems fairly insensitive to funding regimes. That is, marked changes in the levels, incidence, and nature of grant, as well as loan support systems (and tax and other fiscal incentives) do not seem to affect significantly the proportion of enrolments of students from different family wealth backgrounds.

The other important finding from HECS is that the collection of the debt is apparently quite efficient in administrative terms. That is, the Australian Tax Office estimates put the collection costs at around \$25 million

annually, or less than 3% of yearly receipts. Administratively, the system seems to have worked well.

New Zealand

Third, after Sweden and Australia, New Zealand adopted an ICL scheme in 1991 that shares several features of HECS. Specifically:

- loan repayments depend on an individual's income, and are collected through a tax system which made this simple in operational terms; and
- a first income threshold of repayment, after which there is a progressive percentage rate of collection.

The New Zealand arrangements differ importantly to those introduced in Australia. In particular:

- the loans are designed to cover both university fees and some living expenses, although there is also a system of means-tested grants for students from poor backgrounds;
- initially the loans carried a market rate of interest, but now the nominal rate is zero so there is a negative real rate of interest; and
- until recently, universities were free to set their own fees.

A potential advantage of the New Zealand scheme is that universities have discretion over the prices charged, and receive the tuition revenues directly. This implies that in New Zealand there is the prospect of resource allocation effects within the higher education system as a result of the direct nexus between the prices charged and the revenue received. In 2005, the Australian government suggested reforms that were more along these lines. There is little direct evidence of the effects of the New Zealand ICL on the access of disadvantaged prospective students.

The Republic of South Africa

The Republic of South Africa introduced an ICL in 1991, known as the National Student Financial Aid Scheme (NSFAS). NSFAS was motivated essentially by a concern that without assistance the marked racial skewing of the higher education system away from non-white students would remain. While bursaries could have been used instead of ICL, it was considered that the costs involved "... would not be financially sustainable" (Jackson, 2002: 83). The scheme initially provided resources to about 7500 students, but by 2002 this number had risen to over 100,000, or more than 20% of South Africa's higher education students.

Resources are distributed through the universities, with preference going to prospective students who are

both poor and academically able. That is, unlike other national schemes, the South African ICL involves means-testing on the basis of family income at the point of entry to higher education.

There are two major differences between the South African approach and those used in both Australia and New Zealand. The first concerns the first income level of repayment, which at about \$US5000 is very much lower than the thresholds used in other countries' ICLs. Second, in the first instance, the student repays directly to the lending institution. That is, the taxation system is not the first port of call, but is instead a last resort. Employers are required to be involved only when a student is apparently not maintaining expected debt repayments. It is unclear how much this adds to administrative costs, but it would seem to suggest that collection would necessarily be relatively expensive. Jackson argues that the annual administrative costs are less than 2% of the total value of loans distributed. The more important figure, however, would be costs as a proportion of revenues collected – data not reported.

The UK

Higher education financing policy over the last 15 years or so in the UK has been characterized by considerable change. In 1990, a loan scheme was introduced, but collection was not based on a former student's income. In 1997, an ICL was initiated which took the following form: students from poor backgrounds were excused from paying any tuition, while students from rich families incur the entire debt. In between, the debt obligation is determined by means of a sliding scale. This decision seemed to reflect a concern by the government that relatively disadvantaged students would be more likely than others to find an ICL a deterrent to higher education participation, a view at variance with the evidence from the HECS experience reported above. In 2006, the UK government announced further reforms to higher education financing. The major changes are:

- the introduction of price discretion for universities, but with a cap of 3000 pounds per full-time student year; and
- the introduction of tuition for all students, but with the poorest being provided with subsidies.

As with the Australian and New Zealand schemes, the UK ICL policy is likely to be relatively inexpensive to administer. This is because income tax arrangements in these countries greatly facilitate the operation of ICL. As pointed out by Chapman, this is far from the case with respect to developing countries, where public administrative challenges related to the collection of ICL loom large.

The US

In 1993, the Clinton administration introduced broadly based reforms to student loan programs. One noteworthy aspect of the reforms included an option for students to adopt income contingent repayments for some part of their loan obligations, with the ICL obligation being 20% of disposable income.

The ICL reforms introduced in the US have not worked well. With respect to takeup, for example, in 1999 only 7% of the eligible student population initially chose to repay their loan obligations through the ICL option, although 26% of borrowers consolidating (bundling) their loans choose ICL repayments. The basis for low takeup of ICL in the US seems to have two, arguably closely related, explanations. In broad terms, these are: the poor design characteristics of the scheme; and the government's ineffectiveness in accurately explaining and publicizing the scheme's implications for student debt and repayment obligations. It is possible that both weaknesses reflect a lack of ICL policy commitment on the part of those with US policy influence. As of mid-2007, the US Congress was considering legislation that would substantially expand the income contingent component of its existing loan programs.

Other Countries

Other countries have introduced ICL systems, but in general it is not obvious that these have been successful. Chile's 1994 scheme ran into significant collection problems (the tax system was not used), and the Thailand ICL reforms of 2006 have now been suspended arguably in part because of concerns with both the extent of taxpayer subsidies and matters of collection. Israel is close to implementing an ICL, and there are active debates concerning the costs and benefits of such arrangements underway in a large number of countries.

Conclusion

As variations between countries and over time may indicate, there are no definitive answers to how higher education should best be financed. Nevertheless, as indicated above, there is a trend toward both decreases in direct taxpayer support and increased reliance on student loans, particularly those with income contingent repayments. International organizations such as the OECD, International

Monetary Fund (IMF), and World Bank routinely recommend their members move in this direction, and the OECD, for example, describes income contingent student loans as international best practice. This assessment, however, is based in large part on economic reasoning rather than solid evidence, although data consistent with the predictions of the conceptual arguments are accumulating. A critical issue concerns the collection of ICLs, and it is difficult to believe that for many developing countries that this approach can currently work.

Bibliography

- Harding, A. (1995). Financing higher education: An assessment of income-contingent loan options and repayment patterns over the life cycle. *Education Economics* **3**, 173–203.
- Jackson, R. (2002). The South African student assistance scheme. *Welsh Journal of Education* **12**(1), 63–79.
- Palacios, M. (2004). *Investing in Human Capital: A Capital Markets Approach to Student Funding*. Cambridge: Cambridge University Press.
- Usher, A. (2005). *Global Debt Patterns*. Toronto, ON: Educational Policy Institute.

Further Reading

- Barr, N. (2001). *The Welfare State as Piggy Bank*. Cambridge: Cambridge University Press.
- Brody, E. (1994). Paying back your country through income-contingent student loans. *San Diego Law Review* **31**, 449–518.
- Chapman, B. (1997). Conceptual issues and the Australian experience with income contingent charges for higher education. *Economic Journal* **107**(442), 738–751.
- Chapman, B. (2006a). *Government Managing Risk: Income Contingent Loans for Social and Economic Progress*. London: Routledge.
- Chapman, B. (2006b). Income contingent loans for higher education: International reforms. In Hanushek, E. A. and Welch, F. (eds.) *Handbook of the Economics of Education*, vol. 2., pp 1435–1503. Amsterdam: North-Holland.
- Freidman, M. (1955). *Capitalism and Freedom*. Chicago, IL: University of Chicago Press.
- Jonhstone, B. D. and Aemero, A. (2001). *The Applicability for Developing Countries of Income Contingent Loans or Graduate Taxes, with Special Consideration of an Australian HECS-Type Income Contingent Loan Program for Ethiopia*. New York: The International Comparative Higher Education Finance and Accessibility Project.
- OECD (2006). *Education at a Glance*. Paris: OECD.
- OECD (2007). *Economic Survey of the United States*. Paris: OECD.
- Schrag, P. G. (2001). The federal income-contingent repayment option for law student loans. *Hofstra Law Review* **29**, 733–862.
- US General Accounting Office and Department of Education (2001). *Alternative market mechanisms for the student loan program*. Washington, DC, GAO-02-84SP.

Higher Education Crossing Borders

J Knight, University of Toronto, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Growth and Complexity of International Academic Mobility

A fascinating but very complex world of cross-border education is emerging. The last decade has been a hotbed of innovation and new developments. For instance, Phoenix University has become the largest private university in the US (owned and operated by Apollo Group) and is now present or delivering courses in Puerto Rico, Netherlands, Mexico, and Canada. Other Apollo companies are offering courses in Brazil, India, and China. The Netherlands Business School (Universiteit Nijenrode) has recently opened a branch campus in Nigeria and Harvard is planning to develop two branch campus initiatives in Cyprus and the United Arab Emirates. Furthermore, Jinan University will be the first Chinese university to open a branch campus outside China and will do so in Thailand. Laureate Education (formerly Sylvan Learning Systems) has purchased whole or part of private higher education institutions in Chile, Mexico, Panama, and Costa Rica and owns universities in Spain, Switzerland, and France. Dubai has developed a knowledge village in the Dubai Technology and Media Free Zone and to date the London School of Economics, Indian Manipal Academy of Higher Education, and the University of Wollongong from Australia are offering courses through franchising agreements and branch campuses. The University of Westminster (UK) is the key foreign academic partner in the new private Kingdom University of Bahrain and plays a similar advisory/provision role with new institutions in Nigeria, Uzbekistan, and Kazakhstan. (OBHE, 2004)

These are only a few examples of hundreds of new initiatives that have developed in the last few years. They involve higher education providers (including institutions and companies) delivering their courses and programs to students in their home countries. It is convincing evidence that it is no longer just the students who are moving around the world. Different kinds of providers including private companies, traditional academic institutions, and professional associations are now taking academic programs to students where they live. Various types of program delivery methods such as franchising, twinning, joint degrees offer a wide range of courses. New types of partnerships including public/private, nonprofit/for-profit, local/foreign, and institutions/corporations are being formed to respond to the burgeoning demand for access to higher education, and in many cases, the appeal of foreign academic qualifications.

In short, the world of education crossing borders is escalating exponentially and it is likely to continue based on the prediction that the demand for international education will increase fourfold from 1.8 million students in the year 2000 to 7.2 million students in 2025. (Bohm *et al.*, 2002) These are staggering figures. They help to explain and forecast the growth in worldwide academic mobility and sharpen the focus on the movement of programs and providers across borders.

Given the expansion of education crossing borders it is important to discuss emerging issues such as the implications for licensing, quality assurance, access, funding, research, intellectual property, trade policies, recognition of qualifications, and joint ownership. While these issues are definitely central to the analysis of the new forms of international academic mobility, it is difficult to address them, especially in terms of models, policies, and regulations, if there is confusion about what cross-border means and how it relates to internationalization, globalization, and trade in education services.

Evolution of Terminology

It is both fascinating and revealing to see how the vocabulary of international education has evolved. International cooperation, international relations, and international education were the most common terms used 40 years ago. These concepts were usually defined in terms of activities such as development projects, foreign students, and international academic and cultural agreements. In the 1990s, the discussion centered on differentiating the term 'international education' from 'comparative education', 'global education', and 'multicultural education'. But, today, international academic mobility is the focus and there are a number of terms such as cross-border, transnational, and borderless education to describe this phenomenon. **Table 1** illustrates how the international education vocabulary has evolved in the last four decades.

Clearly, the growing interest in the international dimension and delivery of higher education is accompanied by an increase in the number of terms used to describe the changes. It is important to be clear at the outset how related concepts are used in this article and their link to higher education crossing borders:

1. Globalization. It is described as a process that is increasing the flow of people, culture, ideas, values,

Table 1 Evolution of international education terminology

<i>New terms Last 15 years</i>	<i>Existing terms Last 25 years</i>	<i>Traditional terms Last 40 years</i>
<i>Generic terms</i>		
–Globalization	–Internationalization	–International education
–Borderless education	–Multicultural education	–International development cooperation
–Cross-border education	–Intercultural education	–Comparative education
–Transnational education	–Global education	–Correspondence education
–Virtual education	–Distance education	
–Internationalization abroad	–Offshore or overseas education	
–Internationalization at home		
<i>Specific elements</i>		
–Education providers	–International students	–Foreign students
–Corporate universities	–Study abroad	–Student exchange
–Liberalization of educational services	–Institution agreements	–Development projects
–Virtual universities	–Partnership projects	–Cultural agreements
–Branch campus	–Area studies	–Language study
–Twinning programs	–Double/joint degrees	
–Franchise programs		
–Networks		
–Global education index		

From Knight, J. (2005a). *Borderless, Offshore, Transnational and Crossborder Education: Definition and Data Dilemmas*. London: Observatory for Borderless Higher Education.

knowledge, technology, and economy across borders resulting in a more interconnected and interdependent world. Globalization affects each country in different ways. It can have both positive and/or negative consequences, according to a nation's individual history, traditions, culture, priorities, and resources. (Knight, 2006a) Education is one of the sectors impacted by globalization.

2. Internationalization of higher education. This is also a process, albeit a different process than globalization. Internationalization of higher education is described as the process of integrating an international, intercultural, and global dimension into the purpose, functions (teaching, research, and service), and the delivery of higher education at the institutional and national levels (Knight, 2006b).
3. Internationalization strategies. These include international cooperation and development projects; institutional agreements and networks; the international/intercultural dimension of the teaching/learning process, curriculum, and research; campus-based extra-curricular clubs and activities; mobility of academics through exchange, fieldwork, sabbaticals and consultancy work; recruitment of international students; student exchange programs and semesters abroad; joint/double-degree programs; twinning partnerships; branch campuses, etc. Internationalization strategies apply to both campus-based activities and cross-border initiatives. The cross-border aspects of internationalization are the focus of this discussion.
4. Cross-border education refers to the movement of people, knowledge, programs, providers, ideas, curricula,

projects, research, and services across national or regional jurisdictional borders. Cross-border education is a subset of internationalization and can be part of development cooperation projects, academic exchange programs, and commercial initiatives.

5. Trade-of-education services is a term that is primarily used by the trade sector. It focuses on those cross-border education initiatives that are commercial in nature and are usually intended to be for-profit in nature – though this is not always the case. This term coincides with the advent of the general agreement on trade in services (WTO, 1999) which includes the education sector as a tradable service.

Relationship of Cross-Border to Transnational, Borderless, and Offshore Education

The knowledge society, information and communication technologies (ICTs), international trade agreements, market economy – all important aspects of globalization – have had a profound influence on the shape and substance of the higher education sector. As a result, there is a new emphasis on international academic mobility and new terms have been introduced to describe international academic mobility.

Transnational and Offshore Education

Australia was one of the first countries to use the term transnational education in the early 1990s as it wanted to

differentiate between international students recruited to Australian campuses and those who were studying for Australian degrees offshore. (Davis *et al.*, 2000) Hence, the term transnational education became used to simply describe offshore international student enrolments regardless of whether the offshore students were studying through twinning, franchise, distance, or branch campus arrangements. It is interesting to note how the use of terms in Australia has evolved in such a way that international education usually refers to foreign students studying in Australia and transnational education refers to those studying offshore. In this conceptualization of the term transnational, the focus is on where the student is studying.

UNESCO and the Council of Europe in their code of practice on transnational education describe transnational education as all types and modes of delivery of higher education study programs, or sets of courses of study, or educational services (including those of distance education) in which the learners are located in a country different from the one where the awarding institution is based. Such programs may belong to the education system of a state different from the state in which it operates, or may operate independently of any national education system (UNESCO and the Council of Europe, 2005). This definition differs from the Australian term as it includes all situations where programs move across a border or where the provider is virtual and delivering by distance. It is unclear whether they cover new types of providers, especially those that establish a physical presence in the country and obtain permission from the receiving country to offer recognized qualifications. In this scenario, the providers are clearly awarding foreign degrees but they are not necessarily located in a different country than the student.

Offshore education is usually used to denote education delivered abroad, but its use is decreasing due to the more recent introduction of the term cross-border. Furthermore, those countries without a shoreline find this term irrelevant to their situation.

Borderless Education

The term borderless education first appeared in an Australian report (Cunningham *et al.*, 2000) and was followed by a similar type of study in the United Kingdom. Basically the term borderless education refers to the blurring of conceptual, disciplinary, and geographic borders traditionally inherent to higher education (CVCP, 2000). The innovative feature of this term is that it goes beyond geographic and jurisdictional boundaries to include temporal, disciplinary, and conceptual borders. This is the strength of the term, but also a possible weakness. Its contribution is at a conceptual level as it remains fairly

abstract and challenging to use in concrete and applied situations.

It is interesting to juxtapose the concepts of borderless education and cross-border education. The former term acknowledges the disappearance of borders while the latter term actually emphasizes the existence of borders. Both approaches reflect the reality of today. In this period of unprecedented growth in distance and e-learning education, geographic borders seem to be of little consequence. Yet, on the other hand, we can detect a growing importance of borders when the focus turns to regulatory responsibility, especially related to quality assurance, accreditation, funding, joint ownership, or intellectual copyright. (Verbik and Jokivirta, 2005) Borderless education does not seem to be extensively used in operational or applied settings. Nevertheless, it is a useful concept to capture the notion of the blurring and erosion of traditional academic boundaries and for that reason has played an important role in raising awareness about changes in the education sector.

Cross-Border Education

As noted above, cross-border education includes the movement of people, knowledge, programs, providers, ideas, curricula, projects, research, and services across national or regional jurisdictional borders. In many ways it seems ironic that the role of borders is actually stronger in a globalized world that encourages the free flow of people, ideas, goods, services, knowledge, capital, and technology. The introduction of new multilateral trade rules or immigration laws are examples where crossing borders into a different jurisdiction has major implications. Hence, the notion of jurisdictional boundaries has increasing significance for many sectors including education. During the last 5 years, partially in response to the importance of borders, the term cross-border education has emerged and is becoming more widely used as a comprehensive term.

Both UNESCO and OECD have cross-border education higher on the policy agenda in response to member state interests. They have developed, with key education stakeholders, the guidelines for quality provision in cross-border higher education (UNESCO/OECD, 2005). The definition of cross-border education in the proposed guidelines is as follows: higher education that takes place in situations where the teacher, student, program, institution/provider, or course materials cross national jurisdictional borders. Cross-border education may include higher education by public/private and not-for-profit/for-profit providers. It encompasses a wide range of modalities in continuum from face-to-face taking various forms from students traveling abroad and campuses abroad to distance learning using a range of technologies and including e-learning.

Whether higher education is conceptualized as moving across borders or across nations may seem to be insignificant, but the difference between crossing jurisdictional borders (cross-border education) and the separation of learner and the awarding provider in different locations (transnational education) does involve consequences in terms of regulations. There needs to be clarity and consistency of terms within a country for national regulations to work and for a common understanding of regional/international policies. That said, one is optimistic that the new cross-border guidelines developed jointly by UNESCO and OECD will bring some leadership and convergence to the current confusion about terminology.

Elements of Cross-Border Education

Table 2 provides a framework to understand the nature of cross-border education and illustrates two significant trends. The first trend is the vertical shift downward from student mobility to program and provider mobility. It is important to note that numbers of students seeking education in foreign countries is still increasing; but, more emphasis is currently being placed on delivering foreign academic courses and programs to students in their home country. The second shift is from left to right signifying substantial change in orientation from development cooperation to competitive commerce, or in other words, from aid to trade.

Growth in Number and Diversity of Cross-Border Providers

The increase in worldwide demand for higher education has resulted in a diversity of providers delivering education across borders. The providers are classified into two categories: (1) the traditional higher education institutions (HEIs) which are normally oriented to teaching, research, and service/commitment to society, and (2) the new or alternative providers which primarily focus on teaching and the delivery of education services.

Traditional Higher Education Institutions

This category includes public nonprofit, private nonprofit, and private for-profit institutions. Many countries have a mixed system of publicly and privately funded HEIs. There is a definite blurring of the boundary between public and private institutions as public universities are now finding it necessary to seek private financing and are charging a tuition or service fee. On the other hand, in many countries private institutions are eligible for public funds and engage in social nonprofit activities.

An important factor is whether the HEI is part of a home national education system and recognized by a national bona fide licensing/ accrediting body. In cross-border education recognition/registration is critical to

Table 2 Framework for cross-border education

Category	Forms and conditions of mobility		
	Development cooperation	Educational linkages	Commercial trade
People			
Students		Semester/year abroad	
Professors/scholars		Full degrees	
Researchers/		Field/research work	
Experts/consultants		Internships	
		Sabbaticals	
		Consulting	
		Twinning	
		Franchised	
Programs			
Course, program			
subdegree, degree,			
postgraduate			
		Articulated/Validated	
		Joint/double award	
		Online/distance	
Providers			
Institutions		Branch campus	
Organizations		Virtual university	
Companies		Merger/acquisition	
		Independent institutions	
Projects			
Academic projects		Research	
Services		Curriculum	
		Capacity building	
		Educational services	

From Knight, J. (2005b). Crossborder education: An analytical framework for program and provider mobility. In Smart, J. and Tierney, W. (eds.) *Higher Education: Handbook of Theory and Practice*, pp 345–395. Dordrecht: Springer Academic Publishers.

ensuring the legitimacy of the institution and the qualifications provided. The majority of traditional universities are bona fide institutions that comply with domestic and foreign regulations (where they exist). But, there is also an increase in rogue or low-quality providers who are not recognized by bona fide accreditation/licensing bodies in either the sending or receiving countries. Rogue providers are often accredited by self-accrediting groups or by agencies that sell accreditation (accreditation mills). In addition, there is a worrisome increase in the number of degree mills operating around the world. (Garrett, 2005) These are often no more than Web-based companies that are selling certificates based on life experiences and are not delivering actual education programs.

New or Alternative Providers

The new providers are diverse in nature, but are typically described as a company or organization that provides education programs and/or services on a commercial basis. They are more oriented to delivering education and training programs than to undertaking research and scholarly activities. The new providers include publicly traded companies such as Apollo (USA), Informatics (Singapore), and Aptech (India), corporate universities such as those run by Motorola and Toyota, professional associations, and organizations. These new types of cross-border providers can be brick-and-mortar institutions or virtual universities and can complement, compete, collaborate, or simply co-exist with domestic higher providers (and other cross-border providers).

Typology of Program Mobility

Cross-border mobility of programs can be described as the movement of individual education/training courses and programs across national borders through face to face, distance or a combination of these modes. Credits towards a qualification can be awarded by the sending foreign country provider or by an affiliated domestic partner or jointly. Franchising, twinning, double/joint degrees, and various articulation models are the more popular methods of cross-border program mobility (Knight, 2005). A short description of each follows:

1. Franchise. It refers to an arrangement whereby a provider in the source country A authorizes a provider in country B to deliver their course/program/service in country B or other countries. The qualification is awarded by provider in country A. Arrangements for teaching, management, assessment, profit-sharing, awarding of credit/qualification, etc. are customized for each franchise arrangement and must comply with national regulations (if they exist) in country B.

2. Twinning. It describes a situation where a provider in source country A collaborates with a provider located in country B to develop an articulation system that allows students to take course credits in country B and/or source country A. Only one qualification is awarded by provider in source country A. Arrangements for twinning programs and awarding of degree usually comply with national regulations of the provider in the source country A.
3. Double/joint degree. It is an arrangement where providers in different countries collaborate to offer a program for which a student receives a qualification from each provider, or a joint award from the collaborating partners. Arrangements for program provision and criteria for awarding the qualifications are customized for each collaborative initiative in accordance with national regulations in each country.
4. Articulation. It refers to various types of articulation arrangements between providers situated in different countries permit students to gain credit for courses/programs offered by all of the collaborating providers. This allows students to gain credit for work done with a provider other than the provider awarding the qualification.
5. Validation. Validation arrangements between providers in different countries allow provider B in the receiving country to award the qualification of provider A in source country. In some cases, the source-country provider may not offer these courses or awards themselves which may raise questions about quality.
6. Virtual/distance. It refers to arrangements where providers deliver courses/program to students in different countries through distance and online modes. It may include some face-to-face support for students through domestic study or support centers.

A critical factor in program mobility is who awards the course credits or ultimate credential for the program. As the movement of programs proliferates, there will undoubtedly be further changes to national, regional, and even international regulatory frameworks. The question of who grants the credits/awards will be augmented by who recognizes the provider and whether or not the program has been accredited or quality assured by a bona fide body. Of central importance is whether the qualification is recognized for employment or further study in the receiving country and in other countries as well (OECD, 2004). The perceived legitimacy and recognition of the qualification at home and abroad, are fundamental issues yet to be resolved.

Given that several modes for program mobility involve partnerships, there are questions about who owns the intellectual property rights to course design and materials. What are the legal roles and responsibilities of the participating partners in terms of academic, staffing,

recruitment, evaluation, financial, and administrative matters? While the movement of programs across borders has been taking place for many years, it is clear that the new types of providers, partnerships, awards, and delivery modes are challenging national and international policies.

Typology of Traditional and New Provider Mobility

Cross-border mobility of providers can be described as the physical or virtual movement of an education provider (institution, organization, and company) across a national border to establish a presence in order to offer education/training programs and/or services to students and other clients. The difference between program and provider mobility is one of scope and scale in terms of programs/services offered and the local presence (and investment) by the foreign provider. A distinguishing feature between program and provider mobility is that with provider mobility, the learner is not necessarily located in a different country than the awarding institution, which is usually the case in program mobility. Credits and qualifications are awarded by the foreign provider (through foreign, local, or self-accreditation methods) or by an affiliated domestic partner. (Knight, 2005) Different forms of cross-border provider mobility are as follows:

1. Branch campus. A provider in country A establishes a satellite campus in country B to deliver courses and programs to students in country B. This may also include students from country A taking a semester/courses abroad in country B. The qualification awarded is from the provider in country A.
2. Independent institution. A foreign provider A (a traditional university, a commercial company, or an alliance/network) establishes in country B a stand-alone HEI to offer courses /programs and awards. There is usually no home institution in country A.
3. Acquisition/merger. A foreign provider A purchases a part of or 100% of a local HEI in country B.
4. Study center/teaching site. A foreign provider A establishes study centers in country B to support students taking their courses/programs. Study centers can be independent or in collaboration with local providers in country B.
5. Affiliation/networks. Different types of public and private, traditional and new, local and foreign providers collaborate through innovative types of partnerships to establish networks/institutions to deliver courses and programs in local and foreign countries through distance or face-to-face modes.
6. Virtual university. It refers to a provider that delivers credit courses and degree programs to students in different countries through distance education using

predominantly the Internet technology mode, generally without face-to-face support services for students.

The virtual and physical movement of providers to other countries raises many of the same registration, quality assurance, and recognition issues that program mobility does, but there are additional factors to consider if a network or local/foreign partnerships are involved. Setting up a physical presence requires attention being paid to national regulations regarding status of the entity, total or joint ownership with local bodies, tax laws, for profit or nonprofit status, repatriation of earned income, boards of directors, staffing, granting of qualifications, selection of academic programs and courses, etc. For some countries, it means that strict regulations are being developed to closely monitor, and in some case restrict, new providers coming into the country. In other instances, incentives are being offered to attract high-quality institutions/providers to set up a teaching site or full campus. This is especially true where knowledge parks or technology zones or education cities are being developed to attract foreign companies and education/training providers.

Challenges

In a field that is changing as much as international education is, the challenge of finding common meanings for the same and similar terms will continue. Each country, association, or government agency will use terms that make sense from their perspective. It is not the intention to propose one universal definition but it is useful to have a common understanding of what is being debated and more importantly, what is being regulated. The emphasis is on developing a common understanding of the terms for policy, regulations, research, and data-collection purposes, not necessarily a common language, though the latter does facilitate the former.

The mobility of students, professors, knowledge, and values has been part of higher education for centuries but it has only been in the last two decades that there has been a significant growth in the mobility of programs and providers through physical and virtual modes of delivery. This presents many new opportunities – for increased access to higher education; for strategic alliances between countries and regions; for the production and exchange of new knowledge; for the movement of graduates and professionals; for human resource and institutional capacity building; for income generation; for the improvement of academic quality; and for increased mutual understanding. The list of potential benefits is long and varied. But so is the list of potential risks. Risks can include: an increase in low-quality or rogue providers; a potential decrease in public funding if foreign providers are providing

increased access; unsustainable foreign provision of higher education if profit margins are low; foreign qualifications not recognized by domestic employers or education institutions; elitism in terms of those who can afford cross-border education, overuse of English as the language of instruction; and national higher education policy objectives not being met.

Risks and benefits vary between sending and receiving countries, between developed and developing countries, for students, institutions, companies, and employers. In the light of the fast pace of cross-border growth and innovation, it is important that the higher education sector be informed and vigilant about the risks and benefits and more importantly, the need for appropriate policies and regulations to guide current and future developments.

See also: Distance Education and Open Universities; International Accreditation in Higher Education; Quality Assurance in Higher Education – Practices and Issues; Student and Faculty Transnational Mobility in Higher Education.

Bibliography

- Bohm, A., Davis, D., Meares, D., and Pearce, D. (2002). *The Global Student Mobility 2025 Report: Forecasts of the Global Demand for International Education*. Canberra: IDP.
- Cunningham, S., Ryan, L., Stedman, S., et al. (2000). *The Business of Borderless Education*. Canberra: Australian Department of Education, Training and Youth Affairs.
- CVCP (2000). *The Business of Borderless Education: UK Perspectives*. London: Committee of Vice-Chancellors and Principals.
- Davis, D., Olsen, A., and Bohm, A. (2000). *Transnational Education: Providers, Partners and Policy. A Research Study*. Brisbane: IDP.
- Garrett, R. (2005). Fraudulent, sub-standard, ambiguous – the alternative borderless higher education. *Briefing Note*. July 2005. London: The Observatory on Borderless Higher Education.
- Knight, J. (2005a). *Borderless, Offshore, Transnational and Crossborder Education: Definition and Data Dilemmas*. London: Observatory for Borderless Higher Education.
- Knight, J. (2005b). Crossborder education: An analytical framework for program and provider mobility. In Smart, J. and Tierney, W. (eds.) *Higher Education: Handbook of Theory and Practice*, pp 345–395. Dordrecht: Springer Academic Publishers.
- Knight, J. (2006a). *Higher Education Crossing Borders: A Guide to the Implications of Gaps for Cross-Border Education*. Paris: Commonwealth of Learning and UNESCO.
- Knight, J. (2006b). Internationalization: Concepts, complexities and challenges. In Forest, J. and Altbach, P. (eds.) *International Handbook of Higher Education*, pp 207–228. Dordrecht: Springer Academic Publishers.
- OBHE (2004). *Breaking News Stories from 2002–2004*. London: Observatory on Borderless Higher Education.
- OECD (2004). *Internationalization and Trade of Higher Education – Challenges and Opportunities*. Paris: Organization for Economic and Community Development.
- UNESCO and Council of Europe (2001). *The UNESCO-CEPES/Council of Europe Code of Good Practice for the Provision of Transnational Education*. Paris: UNESCO.
- UNESCO/OECD (2005). *Guidelines for Quality Provision in Cross-Border Higher Education*. Paris: UNESCO and OECD.
- Verbik, L. and Jokivirta, L. (2005). *National Regulatory Frameworks for Transnational Higher Education: Models and Trends. Briefing Notes Part 1 and Part 2*. London: Observatory on Borderless Higher Education.

Further Reading

- De Wit, H., Jaramillo, J. C., Gacel-Avila, J., and Knight, J. (eds.) (2006). *Higher Education in Latin America: The International Dimension*. Washington, DC: World Bank.
- Huang, F. (2006). *Transnational Higher Education in Asia and the Pacific Region*. Hiroshima: Research Institute for Higher Education.
- Kishun, R. (ed.) (2006). *The Internationalization of Higher Education in South Africa*. Durban: International Association of South Africa.
- Knight, J. (2006). *Internationalization of Higher Education: New Directions, New Challenges. 2005 IAU Global Survey Report*. Paris: International Association of Universities.
- Larsen, K., Momii, K., and Vincent-Lancrin, S. (2004). *Cross-Border Higher Education: An Analysis of Current Trends, Policy Strategies and Future Scenarios*. London: Observatory on Borderless Higher Education.
- OECD (2004). *Quality and Recognition in Higher Education: The Cross-Border Challenge*. Paris: Organization for Economic and Community Development.
- UNESCO (2004). *Higher Education in a More Globalized World. Position Paper*. Paris: UNESCO.

Higher Education in Post-Conflict Societies

M Daxner, Carl-von-Ossietzky Universitaet Oldenburg, Oldenburg, Germany

© 2010 Elsevier Ltd. All rights reserved.

Glossary

FRY – Federal Republic of Yugoslavia. A federal state which was systematically dismantled and fell into parts after 1991.

Secondary conflicts – Conflicts that stem from the intervention into a territory that has been subject of an external intervention in order to stop original conflicts between two or more parties both internal and cross-border.

Societies of intervention – Postconflict societies where the interveners form a kind of ‘new society’ together with the intervened people. Both sides contribute to this new society by bringing their traditions, culture and customary laws as well as competing lifestyles into a joint structure, which will later be disentangled again. In the meanwhile the society of intervention develops its own culture of intervention.

The World has Changed

Following the Cold War and the end of the bipolar hegemony by two superpowers, the world has become more global and less stable; only for a few countries – for example, in Eastern Europe – is there a clear peace dividend; for other countries – their situation has become more volatile, even precarious, because the well-established elements of balancing economic, political, military, and cultural structures have become shaky. In addition, for quite a few countries, there are no sustainable fundamentals of society and state due to the effects from wars and violent conflicts. It is this group of countries which are in the focus of interest for this article.

Postconflict Societies

Since 1989 – that is, the year of the Communist hemisphere disintegrating – the only remaining superpower is the United States of America. However, new superpowers are emerging, which are likely to bypass established major players in global policies, such as India, China, Brazil, and some more in a second line – among them Australia, Iran, Turkey, South Africa (SA), and South Korea. The old masters of the game, for example, united by their alliances in the North Atlantic Treaty Organization (NATO), the European Union (EU), Association of Southeast Asian

Nations (ASEAN), etc., have become accustomed to deal simultaneously with the USA – one or more of the older powers and at least one of the new players – be it at the level of G 20, the United Nations Security Council, or the International Atomic Energy Agency (IAEA). It is important to start from this viewpoint, because it allows an understanding of the importance of the new conflict scenarios as different from previous wars and belligerent conflicts, such as the two world wars or the series of liberation wars during decolonization. Consequently, these new conflicts create new postconflict constellations. A few qualities of the new scenarios should catch our attention, before we look at the impact those conflicts have on higher education and at the influence of higher education in creating conflicts and contributing to peace-keeping and peace building.

Many conflicts are no longer between (nation-) states, but rather between imperfect states and nation-states or among imperfect states. Imperfect states – also known as failed states, sometimes as rogue states – lack the qualities of recognizable states with most of ingredients such an organization requires in order to be accredited with the international community. The causes of conflicts may be traditional ones, for example, on borders, resources, land for settlement, traffic routes. More often, they are based on ethnic, religious, and other ideological justifications, which may or may not tell the true reason for conflict. We know of cases, where oil or diamonds are the true cause of a rift, but – in the course of conflict – other reasons have won the power of dominating the discourse, such as religious or national discord. Conflicts can start over single issues, such as water resources or access to the sea; they may be the result following the secession of an ethnic entity which had been previously integrated, if not well respected. Many conflicts have very strong partners taking sides – though these may not be active participants in combat, their financial and diplomatic support and their role in supplying one side with arms and logistics may be decisive for a conflict.

The results are many wars – which increasingly share two paradoxical characteristics: The conflicts occur in a zone where the international law (e.g., conventions of warfare, combatant status, prisoner of war (POW), and refugee status) does not apply or is questioned or neglected by at least one party of the conflict; and the conflicts take place in a domain – which can be called enhanced global domestic policy – to the disadvantage of traditional foreign policy and external relations among states. The traditional

prerogative of national prefixes to treaties and agreement is fading. Within the orderly part of the world, supranational agreements and structures replace national institutions and structures; this may be in the course of globalization, or even anticipate global development. The conflicts have become borderless, they transgress traditional limits of warfare, they have become asymmetric, and the outcome has become almost unpredictable. Military success is often twinned with the inability to win peace for the people.

This introductory consideration is necessary in order to make clear how education and higher education fit into the new constellation. Conflicts do not happen among the military alone. They must have been prepared by a complex interplay of institutions and interest groups, of carefully orchestrated provocation and heating up the wars of words and gestures. Since education and culture play a pivotal role in forming a discourse of conflict and inclusion or exclusion, they are also absolutely crucial to prepare a society for a violent conflict – or to restrain their constituency from such a confrontation. The same crucial role is destined for higher education in any postconflict situation.

What does postconflict mean? There are legal consequences from ceasefires and mutual agreements between antagonists, with or without any interlocutor or pairing force. However, in this instance, we must distinguish between the systems level and the level of life-world. This is not simply the distinction between governments and military rulers high up and the rest of the people deep down. The systems level indicates the place where the norms, laws, and rules for the essential infrastructure of the society to emerge and its organization are being designed and decided upon. In many cases, the normative texture of a postconflict society is written outside the country – for example, at the Petersberg/Bonn-Conference for Afghanistan, or at a special body such as the Security Council. In other cases, the system is being re-established inside the country, but far away from the respective society. The life-world level is the domain of traditions – of internalized rules and norms for settling personal conflicts and of arrangements with everyday life, experience, and common sense, and also strong genealogical allegiance, and routines, which are not well understood outside the very cycles of local life. The two normative levels – system and life-world – exist in all societies. However, the importance of the gap and relation between the two is more significant in societies which have only a few civic traditions, or which are totally torn apart by preceding wars and violent conflicts – especially when they lasted for a long period of time. As a rule, the relation between the two levels is heavily damaged during conflicts. In addition, as a consequence, the relation remains deficient for a long period in the postconflict era. Even if there are signs of normalization in a postconflict or reconstruction period, this deranged relation can be detrimental to any civic progress.

Furthermore, it should not be forgotten that postconflict does not mean a complete ending of a preceding conflict. In many cases the conflicts prevail, but are reduced or controlled by new measures of power or by an emerging effect from the rule of law; in some cases, a short period of peace is mistaken as the beginning of a period of low conflict, when the antagonism just regroups and the adversaries resurface after a while. In cases of true reconciliation, we find a good ground for society building and reforms; in such cases, the emerging democratic rule and the impact of civil society help to overcome typical postconflict situations.

The Crucial Role of Higher Education

Higher education has always been a catalyst for the emergence of conflicts and social confrontation. We even have examples, where the universities (in brief, for all higher education institutions) were instrumental to shape and fuel a secessionist insurgency, like in Kosovo (former Federal Republic of Yugoslavia (FRY)). In many cases, students are at the core of the attacking and the defending units, which may be paramilitary, guerrilla, or regular military. More important for both conflict and postconflict is the tight linking of universities with teacher training and with a certain ability to shape the mindset of major actors. Their authority often attributes credibility to them, which most of the politicians and other peers are lacking.

Some elements are often neglected or underestimated, when it comes to the pivotal role of higher education in the process of society-building following a war or other conflict.

1. The student masses are one of the most consistent social strata in a society; they represent a typical passage of status and are united by certain comprehensible political or ideological ideas; more often than not, they take part in movements calling for change. Less often, students are split along the partisan lines of the conflict itself, for example, in cases of ethnic or tribal clashes.
2. Universities have the most effective access to social multipliers of ideas and messages. Both intellectuals and teachers are the perfect bodies to spread information, messages, information, and interpretation.
3. It is within the academic world that the most significant clash of any postconflict society is likely to take place. It is the confrontation between those returning from exile, temporary migration, and foreign employment, and those who have stayed as residents, combatants, active participants in the conflicts, or as hostages and victims of the conflict. Both sides may share a history of political oppression and detention – even torture and other hardship – though it may be this experience which continues the division between the

sides, as its sources and effects may be totally in contrast with each other.

4. Universities often own assets and valuable land, building, and equipment, which may attract the attention of rivaling postconflict groups.
5. Academia is most likely to have international connections and both personal and institutional links with the world outside the conflict zone. Together with the media, universities are likely to determine whether a conflict becomes known to the international community and to form a first – and quite often sustained – image of the conflict.

It is a truism – which does not need an open conflict to be confirmed – that universities are dangerous institutions. They are dangerous for any dictatorship or authoritarian rule, especially when it is based on faith, a narrow ideology, or an exaggerated nationalism. Therefore, universities are first attacked or even shut down in any case of emergency or violent clashes, or they may be instrumental for one of the parties in the conflict.

The Postconflict Scenario

What does a country or a society need when it has just passed through a period of war and destruction? There are two answers – which concur and antagonize nevertheless. One answer is that security and safety – together with a re-establishment of political representation and a strong monopoly of power by the new government or military force – should be established. This would enable the economic and infrastructure to recover. The other answer is pointing at the necessity to create a spirit of ownership and responsibility with the people. The priorities are clearly set in the soft sectors, that is, the reconstruction of education, public health, and social security. Of course, it is evident that both sides must become intertwined and should develop in a complementary way. However, their implementation regularly creates secondary conflicts, and it is likely that the competition between hard and soft sectors is a major impediment to proper reconstruction. The formula which is almost unanimously accepted, if not applied in most cases, is that the rule of law should precede any other major institution, and that security must prevail in order to let civil society and public institutions develop.

How to Restart Universities – Some Case Studies and Examples

Postconflict societies are not to be compared to normal societies: people are traumatized, they have suffered severe losses, and many have lost next of kin or partners; businesses out of function, there is no social security,

displacement is a major problem, refugees compete for scarce resources, etc.

One aspect should be considered with special emphasis. Postconflict societies are stressed with quite a few secondary conflicts – which are different in any case – but share one element altogether: the secondary conflicts merge in the comprehension of the people with the preceding conflicts that were the initial cause of the situation. For example, if there is an occupation regime, its behavior is often compared to the authoritarian style of the prewar regime, and thus creates analogous reactions under totally different auspices. Another factor is the presence and interference of military. Civil and military affairs mix, but are rarely tuned.

There is a special postconflict culture in the making, which can turn – in case of external intervention – into a culture of intervention.

The diverse roles of higher education under such circumstances and their interdependence with the society-building process are described by a few paradigmatic examples. There is no best-practice approach, but a few lessons-learned scopes and good-practice examples may depict the multiple problems. (The intention is not to go into a comprehensive review of all concerns, but to highlight a few with special significance).

South Africa (SA)

Following the fall of the Apartheid Regime and the long-lasting conflicts, the country has achieved a rapid return into the international community of higher learning. Notwithstanding residual conflicts from earlier, such as ethnic and class-related discrimination, SA has adopted a system of higher education – which aims at utmost compatibility with international standards. The most prominent initiator of higher education reforms, Jairam Reddy, was called to the National Higher Education Commission, in 1995, and inspired major reforms and legislation. He recommends this form of reforms to the reconstruction of Iraqi higher education. When it comes to accreditation, quality assurance, and curriculum development, the system has accepted the pull of global standards, while the push is for more participation, equality of access, and an end to the residual conflicts from the past. The Higher Education Quality Committee (HEQC) is now chaired by an internationally recognized scholar, Mala Singh, who – after Reddy – has already established a kind of reform routine, which is essential for the sustainability of postconflict reconstruction. The point here is that the higher education system has adapted to the new situation – which is democratic and free – without relinquishing all features from the past. The system is reformist in many ways, while the society has undergone a revolution. Reconciliation, truth commissions, and a diversified civil society are helping to

stabilize a still fragile system – and all challenges can be coped with. This does not imply that there are no severe postconflict potentials and frictions; but, SA is one example for an inclusive role of higher education. Certainly, the stark international attention the country – and its role in Africa and worldwide – enjoys adds to the success of the system.

South-East Europe (SEE)

The FRY dissociated gradually after 1989. Wars and secession have dismantled one formerly large federal nation-state into six sovereign states and one (Kosovo) to be, while one other (Bosnia-Herzegovina) is split into three ethnical entities and some ten ethnic cantons with different features of higher education. The former FRY higher education was a modified socialist state-centered system with some rather well-reputed universities, that is, Belgrade, Zagreb, and Novi Sad, and quite a few outstanding research accomplishments. The universities became instrumental in the process of dissolving a Yugoslav identity in favor of national resurfacing (Croat, Serbian, Bosnian, etc.). The wars of secession (Slovenia and Croatia), the cruel liberation war in Bosnia – which ended with the Dayton Agreement (1995) – and the war on Kosovo (1999) dissociated a formerly moderately centralized system with very high autonomy of the institutions. This autonomy was, however, embedded in the authoritarian rule of a party regime and an increasingly ineffective state-ridden economy. The system was structurally more corrupt than it was individually. While Slovenia is now a member of the EU and likely to be entirely integrated within a few years, the other countries still struggle on their way to the new Europe: there are many losers and there is certainly nostalgia from those who cannot adapt to the competitive standards of the new supranational situation. We can take a closer look at a typical postconflict entity: Kosovo – which was never a full fledged Republic in the FRY – developed a quasi-state parallel system following 1989, when the autocratic rule of Milosevic began to tyrannize the province. The official institutions (one university and some other tertiary colleges) expelled all students from the Albanian majority and became mainly Serbian, only admitting few non-Albanian minorities. The parallel system hatched a parallel school and university system – which enjoyed a lot of support from expatriate Albanians and the international community. While it is true that this system was idealized and did not meet high academic standards, it was able to function as an ideological center for the guerrilla war of the Albanians against the Serbian dictatorship, and, at the same time, prepared for a purely Albanian structure following the conflict. This was the situation encountered by the United Nations interim government (UNMIK), which had to set up a fully developed civil administration under the mandate of the Security Council (Resolution 1244). While the

Albanians, following the liberation, wanted to continue the old, basically socialist system under altered ethnic terms, UNMIK insisted on reform legislation and succeeded in promulgating a new law on higher education – written by the UNMIK administration, local experts, and the Council of Europe. This law is typically postconflict and represented the new Europe versus a national – that is, local – solution. Most of it is still in effect, but the interethnic perspectives and liberation of the higher education system from its traditional ties has only been realized partially. Thus, the university is still an incubator for some political parties and a source of utmost concern for the international working group, which prepares a state-building solution for the conflict. In addition, local and international corruption and a remarkable inconsistency of the UNMIK administration following 2003 have created a difficult legacy for a failed state-to-be.

Serbia – which was still the heart of the residual FRY in 1999 – was heavily hit by the NATO bombardment in order to rescue the Kosovo Albanians from the ethnic purging. While its domestic policy changed into a phase of fragile Europe-oriented reformism under Prime Minister Djindjic, its higher education system could immediately connect with the civil society from the prewar underground or opposition. Milosevic had narrowed the autonomous leeway of universities till he destroyed all freedom by the legislation of 1998. During his dictatorship, many higher education associations and also some governmental and nongovernmental organizations (NGOs) supported the major opposition network, the Alternative Academic Education Network (AAEN). It was from this network that the major reform impulses came into the administration and new legislative efforts. European standards and ongoing reforms – for example, the Bologna process – enjoyed some support, while the core element of the socialist system – the autonomy of faculties under a weak rector's rule – prevailed. Following the assassination of Djindjic and the establishment of a nationalist government, the old structures could partially regain in power, though the reforms and the new spirit did not entirely disappear. One university – Novi Sad – has managed to use the relative autonomy of its provincial structure – The Province of *Vojvodina* – to become a truly integrated European institution; Belgrade still struggles with the postconflict inconsistencies; and the other universities lose ground because of the ambivalence of the government – whether to connect more closely with European reforms or to follow nationalist isolationism. It is hoped that that a final decision on Kosovo will clearly push the country into a European allegiance with hopes to join the EU. One very typical effect of postconflict situations can be studied in Serbia: as there are no longer stark ethnic frictions, the mere existence of a traditional state bureaucracy and administration is still an asset of the respective government, although it slows down radical reforms.

Iraq

Iraq is certainly no postconflict country, but it is the extreme case of peaceful reconstruction of one sector while war is pending. The postwar situation is certainly worse than the years under the dictatorship of Saddam Hussein – at least with regard to Sunni female students and to scholarship. Apart from the Kurdish regions – which always had a particular structure – the dominating intervention power – that is, the USA – has given way to a re-establishment of fundamental religious impact on education and lifestyle. On the other hand, the former high level of higher education and scientific standards allows, at present, a much faster reorientation to the future than in other postconflict areas, for example, Afghanistan. We can study some effects from a forced and alienating imposture of foreign values, which are evidently not shared by major parts of the population. On the other side, the availability of sufficiently enough well-trained school teachers and academics allows for a rapid redevelopment of international ties and scholarly partnership. As one journalist put it: there is a war within the war. The USA itself and its allies have pledged a vast amount of money for the reconstruction of higher education. There is the danger that these peaceful segments of the present policy will fall prey to a general anti-American and anti-Western resentment and that the country will increasingly become divided into three cultures – that is, Kurd, Shiite, and Sunni. One lesson to be learned from the present status of higher education is that formerly highly qualified faculty should be given a chance to actively take part in the reconstruction, and to use their old ties with the international community of learning. Since Iraq is far from any South African approach toward reconciliation, it is dangerous to wait for the moment when justice can determine who will be part of the future academia; this is a situation before the rule of law can also determine the future of universities.

Afghanistan

This country is the paradigm for what we can call peace in wartime. Afghanistan had enjoyed its last strong reform period in the 1920s. Following World War II, the system was capital-centered (Kabul), with a small urban upper class and sufficiently developed higher education for the upper-middle classes, while the elite were trained in foreign countries – mainly in the West. The last legislation under preconflict conditions stems from 1976. Following the Soviet occupation in 1978, the country has suffered, until recently, an endless succession of violent conflicts, liberation wars, civil wars, and dictatorships, of which the last one – the Taliban regime, until 2001 – was the worst. Consequently, the establishment of the Karzai Government, in 2002, by international intervention and

recognition by the international community of states gave the postconflict commonwealth a strong legitimacy. However, a system of barely 40 000 students in a country of 27 million inhabitants and an infrastructure which has been totally destroyed cannot be rebuilt by one consistent act. Moreover, the secondary conflicts as described above cumulate in the country. While the war against the Taliban regime was eventually won by the victorious Northern Alliance with the massive support of international military, it goes on in many areas. It is not clear whether this is part of the global War on Terrorism, or whether the resurfacing of Taliban is a regional conflict – only concerning Afghanistan and Pakistan.

From 2002 to 2004, a treacherous progress occurred: the Minister of Higher education was then a secular, intelligent typical returnee from the USA – a scholar and liberal patriot, who enjoyed high reputation internationally. In this time, a review of the system resulted in plans for new legislation, a stark expansion of the system, and a broad admittance of women and formerly neglected parts of the population, mostly rural. Community colleges were in the focus of the policy. A rectors conference was founded and asked for more academic freedom and institutional autonomy. Following the elections, however, the balance shifted away from the returnees toward the local residents – who were more traditional, had a stronger inclination toward religion, and certainly followed stronger ethnic imperatives. The ideas then, until 2006, were to re-establish a slightly modernized version of the 1976 law, and to restart a society – or academic community at least – which was, given the horrible years to follow, relatively calm and stable. Today, the new draft law resembles an authoritarian, state-oriented master plan of the government. This is a reaction to the typical postconflict frustration – when the call for a strong state and a speeding up of reforms and improvement is no longer entrusted to the people, but to power. The situation is literally one of peace in wartimes. In many parts of the country, there is a full-fledged war – conducted mainly by international intervention forces – and reconstruction continues throughout the country, with some effect. Higher education is the most sensitive area for the immediate future. If it does not happen that many more young people can enrol in useful and educating programs, the men will likely defect to more radical groups than earlier, and the women will fall back into their depreciated, oppressed status. Furthermore, if the universities do not deliver tens of thousands of well-educated schoolteachers every year, the basic reforms of the school system in an overpopulated country will fail: 50% of the people are below 21 years of age.

Another element is also critical to the success of the reforms. Afghanistan has a small, well-educated elite – who had returned at one time or another from exile, either in the USA and Western Europe, or from Iran and Pakistan. They know about good universities. They are the advocates

for adopting global standards in the system, also being aware of the benefits of diversification and quality control. The mainstream policy is slowing down the reforms and hopes for an autochthonous solution of problems, thus alienating international assistance.

Thus, there is a competition between two different elites that is presently underway.

Conclusion and Lessons to be Learned

All postconflict societies share the fate of being torn apart as a consequence of the incoherent pace and structure of reconstruction. The vicious circle is hard to dismantle: there is a devastated society, there are traumatized people, and what are needed are education and the security and safety to lead a decent life. Education does depend much more on universities than on the reform of primary education. Universities and other tertiary institutions are the hinge between the upcoming elite and the masses of people who are in dear need of practically everything. As higher education is hardly ever factored into the initial steps to restore civic life – that is, immediately following ceasefires and the provisional establishment of government and administration – an impact from higher learning on society-building is neglected or even unwanted. Thus, the important segment of students and future leaders is given too little attention as to generate quick impact loyalty to the new situation; many students only become open to extremist propaganda because they are not embedded in the new ownership spirit. Second, universities do not get quickly enough influence on the mindset of teachers – who are the major multipliers in any new society, apart from clergy and strongmen. Third, higher education is the place where ideally reconciliation can start. It is a place where the global exchange of ideas does follow more civic rules than economy and foreign policy; this is important because the soft power of international cooperation can be evidenced as one of the most motivating and sustainable impacts on regaining self-attainment and identity. In all countries described as examples, international organizations such as the International association of Universities (IAU), the International Association of University Presidents (IAUP), UNESCO, or national rectors conferences have moved minds and money in order to create islands of stable development. The Magna Charta Observatory – which is an organization monitoring academic freedom – has extended its range of action beyond the borders of Europe into countries like Afghanistan and the Arab world, thus showing the universal aspects of higher education reforms. The main obstacles to all these changes are the local governments, the old vested interests among the academic elites, and the prevailing insecurity – which makes it less attractive to invest in higher education. Governments give in too easily

to claims from fundamentalist religious leaders, and they allow an unorganized privatization to take stock where a strong and liberal governmental leadership is needed. Finally, the civil–military cooperation also gains in importance in the academic sectors. Not only do institutions and individuals very often need a kind of protection which only the military can provide, there is also a militarization in post-conflict societies, where the military – especially foreign intervention forces – play a role in creating a new culture, with or without their explicit intention. From the classic postconflict region in Europe, we know how important the military of the allies was for re-education. In many cases – especially in Afghanistan, where there is sovereign state – the cooperation between foreign military and local forces has even created a model for such cooperation – the provincial reconstruction teams (PRT).

Conflicts must be regulated. They cannot be solved like other social and political problems. Regulation always depends on power. In addition, each legitimate power needs to understand the basis of its own functionality. That is why, from the beginning of a postconflict situation, universities and research into the environment of the diverse powers and social groups are needed. This also gives the universities the chance to influence the new societies better than they usually had been able to do in the past. We should not forget that postconflict regimes need time, resources, and sufficient brain gain in order to survive.

Further Reading

- ARC (Afghan Rectors Conference) (2004). *Statuten für die Afghanische Rektorenkonferenz und Grundordnungen für afghanische Hochschulen*. (Statutes for the Afghan Rectors Conference. Drafted and translated from German). Commissioned by the Afghan Ministry of Higher Education and the German Rectors Conference (HRK). Kabul: Ministry of Higher Education.
- Daxner, M. (2003). *A Coherent Policy for South East Europe*. Vienna: Federal Ministry of Higher Education.
- Daxner, M. (2004). *Academic Freedom and University Institutional Responsibilities in the Balkans (1989–2003)*. Bologna: Magna Charta Observatory, Bononia University Press.
- Daxner, M. (2006). *Bildung in Nicht-Zivilgesellschaften*. In Gerlof, K and Ulrich, A. (eds.) *Die Verfasstheit der Wissensgesellschaft*. Münster: Westfälisches Dampfboot.
- Daxner, M. and Palandt, K. (2004). *Afghan higher education law, with commentary (Draft)*. Translated by Minister Sharif Fayeze. Commissioned by the Afghan Ministry of Higher education and the German Academic Exchange (DAAD). Kabul: Ministry of Higher Education.
- Holsti, K. (1996). Peace, war, and the state of the state. *International Political Science Review* 16(4), 319–339.
- Holsti, K. (1998). *The Strength of State*. Cambridge, UK: Cambridge University Press.
- Kaldor, M. (2004). *New and Old Wars*. Cambridge, UK: Polity Press.
- Lempinen, E. W. (2006). High-ranking Iraqi education and science leaders to visit AAAS. AAAS News Archives, 29th July, 2006, p.1.
- Reddy, J. (2005). *UNU Report on Iraq*. UNU, ILI (International Leadership Institute), Amman, Jordan: UNU.
- UNU-ILI (2006). *Iraq higher education*. www.unu.edu/news/ili/Iraq.doc (accessed August 2009).

Higher Education in the New Economies

J Fielden, CHEMS Consulting, Hampshire, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Franchised programs – The programs designed and awarded by one institution but delivered under license by another.

Gross enrolment ratios – The percentage of the age group enrolled in tertiary-education institutions of all kinds.

Incubator centers – The offices or proto-manufacturing facilities usually located close to a university where the outcomes of university intellectual property or enterprise are piloted.

Introduction

Since this article has to cover a massive field, in order to distil the range of countries and the key issues into a manageable framework, the main focus is on China and India and the issues that arise within them as two of the world's largest higher education systems. Three cross-cutting issues are then considered in the context of the Indian subcontinent and other emerging nations.

The rankings from the United Nations Development Program (UNDP)'s human development index for five emerging economies are shown in **Table 1** and are compared with Norway and Australia, two countries at the top of the index. The literacy rate is shown to be low in India and gross tertiary-enrolment ratios in China, India, and South Africa are well below the two developed countries. The share of public expenditure taken by tertiary education is also low, although this figure can often be distorted by private-sector participation and funding.

China

The Chinese higher education system has expanded rapidly since 1980, growing from 1 million students in 1980 to 15 million in 2005 with the gross enrolment ratio shown above of 19.1%. China may have overtaken the USA to become the largest higher education system in the world, as reports from the Ministry of Education in March 2008 claimed that enrolments exceeded 25 million.

This expansion has been faster than targeted in the 5-year plan and has involved growth across the board in

both public and private-sector provision. The key features of the expansion have been:

- Structural reforms with the centrally managed mergers of institutions. Between 1993 and 2001, 708 universities were merged to become 302 and as a result, several institutions (such as Sichuan or Jilin Universities) have student populations exceeding 50000.
- The removal of responsibility for specialist universities from parent ministries (under a Soviet model) to place them under provincial governments.
- Managerial reforms which delegated more powers to university presidents and in the provinces, particularly, allowed them greater autonomy. The Ministry of Education agreed to share the strategic management of some of its institutions with the provincial administration where it was based.
- The introduction of tuition fee in the 1990s which increased gradually and is now capped at RMB 6000 by the Ministry of Education. In 1992, such fees accounted for less than 5% of universities' total income, but now represent up to 20% of income across the board.
- The generation of income from other sources such as consultancy and commercial enterprises. This policy backfired in some instances in the late 1990s, and some universities lost money on ill-advised commercial ventures. Government is now more cautious in its encouragement of entrepreneurial activity.
- A general welcome for a suitably regulated private sector as part of the national system. In 2001, there were more than 1200 private higher education institutions with over 1.1 million students.

The sudden growth in enrolments and numbers of graduating students is affecting employment prospects and reports of graduate unemployment have been circulating – even among those who have paid to receive a Western-university education. Consequently, some universities have been developing very close links with employers in neighboring business parks in order to guarantee jobs for their students.

The dramatic transformation of the income profile of Chinese universities can be seen in **Table 2**.

Universities are also being transformed academically. Until the mid-1990s, many institutions had outdated equipment, aging facilities, and curricula. Students gained honors by rote learning and memorization. This has now changed with a massive investment in equipment and buildings as well as a transformation in the approach to

Table 1 Relevant statistics from four emerging economies

Country	Overall HDI ranking (2005)	Adult literacy rate	Gross tertiary enrolment ratios	Public expenditure on tertiary education, 2002–05 (% of all levels)
Brazil	70	88.6%	22.3%	19%
Russian Federation	67	99.4%	68.2%	n/a
China	81	90.9%	19.1%	21%
India	128	61.0%	11.8%	18%
South Africa	121	82.4%	15.6%	16%
Norway	2	99.0%	80.5%	33%
Australia	3	99.0%	72.2%	25%

Extracted from UNDP Human Development country reports for 2007–08.

Table 2 Sources of income for Chinese public universities

	1978	1992	2005 ^a
Allocation from central or provincial governments	95.9%	81.8%	30%
Net income from university-generated enterprises	4.1%	8.9%	10%
Research grants, contracts, etc.	-	4.7%	30%
Tuition fees, domestic and international students	-	4.6%	30%
	100.0%	100.0%	100.0%

From (1997 and 1992) China: Higher education reform. Washington, DC: World Bank.

^aEstimate based on an analysis of a major public institution.

learning. Visitors from other countries are astounded at the scale and speed with which large new campuses (and in many cases, associated business parks or incubator centers) are being developed all around the country.

Policymakers see the growth of a group of elite universities as being central to national goals of strengthening economic competitiveness. The national aim is that a select number of Chinese universities should be accepted as world class by 2010 and considerable investments are being made by the ministry in the two categories of elite institution: the 100 or more in the 211 grouping and the 35 in the most prestigious category of 985 universities. The criteria for admission to these two groupings are not published, but their numbers include most of the large comprehensive universities as well as selected specialist ones in medicine, education, and agriculture.

Chinese universities are encouraged to develop strategic partnerships with leading universities in the USA and Europe and internationalization is a strategic goal for all the major institutions. In the first 7 years of the twenty-first century, hundreds of such partnerships, teaching and research collaborations, have been established

and, after taking advantage of a new law relating to private and international providers, several offshore campuses were set up by UK and US institutions. There are many hundreds of twinning, joint degree, and franchised programs with overseas universities, and in 2007, a queue of 378 applications for such programs was awaiting approval by the Ministry of Education. This global strategy is beginning to work: in 2007, The Times Higher World University rankings included five universities from China, while Peking University was ranked 18th in the world for life sciences and for arts and humanities. Tsinghua University was listed as number 16 for technology.

China is the world's largest supplier of internationally mobile students with as many as 67 000 in the USA and 50 700 in the UK. However, these flows are gradually being reversed as another element in the internationalization picture is the recruitment of international students to China itself. In March 2008, the Chinese Ministry of Education announced that 195 000 overseas students were studying in China, making it the sixth largest destination country. In order to attract such students, many courses are delivered in English and tuition fees are set well below Western levels, thus ensuring a flow of students from the Asian market.

Chinese policymakers are frank about the shortcomings of their system. They acknowledge for example:

- that quality and quality-assurance processes are in need of attention;
- that most development has been from the center and that more needs to be done to encourage local initiatives; and
- that university presidents lack management competences.

Despite these concerns, the Chinese higher education scene has come a very long way in the last 15 years and is moving rapidly toward a fully developed system similar to that in North America and Europe. All this has been achieved by a clear sense of strategic direction from the center, allied to a substantial investment of funds.

India and the Subcontinent

The Indian subcontinent presents a microcosm of most of the issues confronting tertiary education in developing economies; it also illustrates very different responses to the problems from those seen in China.

The gross enrolment ratio in tertiary education in the subcontinent is well below that of regional competitors such as Malaysia, Thailand, and China (see **Table 3**). Figures for Sri Lanka are not available.

Despite their low starting point India, Pakistan, and Bangladesh have aspirations to enter the knowledge economy and to provide higher education to a much

Table 3 Gross enrolment ratios in the Asia region

Malaysia	32%
Thailand	43%
China	19%
India	12%
Pakistan	5%
Bangladesh	6%

Extracted from UNDP Human Development country reports for 2007–08.

higher proportion of the relevant age group than at present. National planning targets require significant expansion in public-sector institutions. In 2007, this was recognized in India since the 5-year plan for 2007–12 aims for a participation rate of 15% by 2012 and could involve a doubling of the number of institutions and another 6 or 7 million students. There are some doubts as to whether the number of graduates that would emerge could find jobs in an economy that is currently dominated by public-sector employment.

The Indian tertiary education system is complicated and has the following elements (see **Table 4**).

The Indian government's strategy is to create 20 new institutions of national importance (which will include five new entities called Indian Institutes of Science Education and Research), 30 new universities, and hundreds of new colleges. In addition, existing universities would increase their student numbers by 50% over the next 5 years. For example, Delhi University would grow from 120 000 to 180 000 students in parallel with increasing the proportion of its students coming from other backward classes to 27%.

India's resurgent economy should make this expansion possible and the plan suggests a dramatic quintupling of higher education funding. Until this decision, however, public funding for higher education had remained static at less than half a percent of gross domestic product (GDP) for two decades. One-third of institutions received no government funding at all, relying on tuition fees and surpluses from providing distance education (Agarwal, 2007).

However, funding is not the main concern of Indian policymakers. In November 2006, the National Knowledge Commission produced a report for the prime minister that painted a very dismal picture of the state of Indian universities in the state sector (with a few notable exceptions). (National Knowledge Commission, 2006). It reported that the curriculum has not changed for years, buildings and equipment are in poor condition, academic staff are poorly motivated, promotion is not based on merit, there is little good research, and campuses are often politicized. Much of this was due to low funding, as 90% of recurrent budgets are absorbed by salaries and there are no incentives to generate income from other sources. The scope for entrepreneurial management is limited.

Table 4 Key statistics for higher education institutions in India

Institutions of national importance (e.g., IITs, IIMs)	13
Central universities funded by the UGC	18
State-funded universities funded by their state	211
Deemed universities, funded by the UGC or other public bodies	38
Private deemed universities	72
Private universities	10
Government colleges	4,250
Private and private-aided colleges	7,964
Students enrolled (in millions)	11.2

From Agarwal, P. (2007). *Private Higher Education in India*. London: Observatory on Borderless Higher Education and University Grants Commission (2006). *2004–05 Annual Report*. New Delhi.

The other major issues that remain to be resolved in the Indian higher education system are:

- The very uneven quality of higher education, which ranges from the world class institutions such as the Institutes of Indian Technology to very poor-quality provision in the bulk of the government colleges affiliated to universities. There is no compulsory quality-assurance regime and a body set up to undertake accreditation; the National Academic Accreditation Council (NAAC) has limited powers as its reviews are voluntary.
- The ambivalent and confusing attitude of government to private institutions, which again range in quality terms from many that are outstanding and far superior to the majority of the state universities to some that are of very low quality indeed.
- Government attempts to introduce some regulation and quality assurance over the private sector and to encourage collaboration with international providers have been thwarted by the legislature and the courts.

Despite the absence of any clear policy steers from government, North American, European, and major Asian universities have been clamoring to establish partnerships and teaching and research links with the better Indian institutions. An Indian government survey in 2007 reported that there were 130 twinning programs between foreign providers and private institutions; the majority of the foreigners being American universities. Unlike China, there have been no foreign universities establishing their own campuses in India, since this is not yet allowed. Research collaborations between Indian institutions and foreign partners are mushrooming and some Western governments have established funds to foster research links. An example of this is the UK India Education Research Initiative, which uses government and UK private-sector funding to support joint research projects.

Distance learning is a significant player in the Indian system with a network of 11 state open universities and

the Indira Gandhi National Open University. Together with distance education institutes within conventional universities, these deliver higher education to over 2.8 million students. However, as with the state universities, the quality of the provision is uneven.

Other Countries on the Indian Subcontinent

In Bangladesh, the 2006 higher education strategy took a similar approach as that taken by India in suggesting that existing public institutions would grow substantially. However, the bulk of the expansion would be handled by the private sector and the Open University in view of the government's inability to fund more than a certain percentage.

In India, Sri Lanka, and Bangladesh the consequence of a small tertiary public sector is that many students leave secondary education qualified to enter tertiary education, but cannot get a place in public institutions. The figures are particularly striking in Sri Lanka:

- secondary school leavers qualified to enter higher education: 117 000;
- annual entrants to public universities: 16 000;
- students who study overseas (estimated): 12 000;
- possible entrants to the private sector: 15 000; and
- unaccounted for: 74 000.

As we shall see below, the place of the private sector in expansion is crucial. It offers governments an alternative way to achieve the skilled manpower they seek. Yet, its status in many national higher education systems is still unclear, uncharted, and suspected. Despite this, the private sector is filling the gap and parents are paying to get their children an education that the state cannot provide. Given that many already pay for private nursery, primary, and secondary education, they see little wrong in the principle of paying for tertiary education.

The planned Indian expansion of higher education funding follows the example of Pakistan where, after a long series of task-force reports and strategic reviews, it was decided in 2003 to invest very large sums in a major reform of the sector. The agency for this was the Higher Education Commission (HEC), a new powerful body, independent of the ministry that was given the authority to embark on an ambitious investment program covering:

- staff-development scholarships, fellowships and exchanges with over 15 countries;
- curriculum reform;
- installation of a national academic information and communication technology (ICT) network;
- a digital library offering open access to all staff and students;

- management training for leaders;
- a quality-assurance agency;
- research grants and development grants for industrial linkages.

The work of the HEC is regarded as the most wide-ranging and ambitious sector-reform program in the world and owes its success to presidential support and energetic leadership of the HEC.

Cross-Cutting Themes

In emerging economies, there three common issues that dominate the agenda of every minister responsible for higher education. They are:

- how to finance expansion;
- the role of the private sector; and
- maintaining quality, relevance, and employability.

Financing Expansion

Funding the expansion of higher education is beyond the purse of most governments. There are only four ways in which the money can be found from within the public sector:

- *Increasing tuition fees.* This issue is always political rather than financial or procedural. In many countries, fee levels have often remained unaltered for 20 years. In Bangladesh, India, and Sri Lanka tuition fees of US\$10–US\$15 a year are not uncommon, but the suggestion of an increase to this figure is likely to lead to student riots and political retreat. Yet, there are massive contradictions in reality; Bangladesh has recently launched a new public institution that charges near-market fees and even state-funded universities such as the Bangladesh University of Engineering and Technology are able to levy realistic charges for accommodation and lecture facilities which result in massively improved and visible premises. In Sri Lanka, students in public universities continue to receive stipends as well as access to food and facilities at almost no cost. In India, a large number of state-run institutions only survive because of the fees that they can charge for distance-education programs. In Kenya and Uganda, the government encourages public universities to charge market-level fees to separate streams of students in parallel with state-sponsored low-fee-paying students.
- *The flows of repayments from student-loan schemes.* Such schemes allow poorer parents and students to borrow the full cost of their fees and costs (or receive a grant toward them) and are an essential concomitant to increases in tuition fees. Yet, they suffer from real barriers; no money is saved in the short term since the funding for the loan (or financial guarantees) must still come from government initially; very few developing

countries have found ways of collecting loan repayments efficiently; the mechanisms for assessing needs and managing the loan processes are themselves expensive and prone to fraud and inefficiency. As a consequence, it will take some time for the trickle of repayments to represent any substantial share of government income in most countries.

- *Improving the economic efficiency of institutions.* This is often cited as another way to produce more from the same inputs of resources. There are several candidates for improvement:
 1. Worsening the student–staff ratios above the current levels in the developed world of around 15:1–20:1 to 30 or 40:1. This presents a risk to quality and is more likely to occur as a result of financial neglect rather than a clear strategy.
 2. Reducing the numbers of manual support staff, which are often bloated and hard to justify. Ratios of four support staff for each academic staff member are not uncommon.

Since remedying both inefficiencies involves a reduction in staff numbers, they are extremely hard to achieve in societies such as India and Bangladesh, where employment is regarded as for life and where leadership responsibilities are held by politically sensitive appointees.

- *Generating income from nongovernment sources.* This is a useful avenue, but it requires the support of responsive internal procedures and incentives as well as entrepreneurial skills. These are both rare but not unknown, but tend to be found in the more market-related disciplines such as engineering, law, or business in universities in capital cities. There are exceptions; the Department of Archaeology at the University of Peradeniya in Sri Lanka, for example, draws a significant income from its contracts with the United Nations Educational, Scientific and Cultural Organization (UNESCO).

The Role of the Private Sector

The role of the private sector is increasingly important in those countries where the state is unable to finance the necessary expansion in public institutions. We saw the

large numbers of private institutions in China and India and this is reflected in most other emerging and developing countries (Table 5). An emerging pattern is for the number of public universities to remain constant, while expansion occurs in the private sector.

Governments' policies toward the private sector can be classified as follows:

- *laissez faire*, with no regulation or quality assurance, as a result of which poor-quality providers can flourish at the expense of an ill-informed public;
- hostile, with a regulatory regime that is aimed at controlling the worst providers (but which does not apply equally to public institutions) and a presumption that all private-sector universities are of poor quality;
- neutral, with licensing, accreditation, and program-review regimes that apply equally to both public and private-sector institutions; and
- positive, with government incentives such as the award of land, tax relief, and access to funding for suitable sponsors.

It is common for the vested interests of the established public institutions to dominate decision making in ministries of education and their buffer bodies and this leads in many countries to a reluctance to acknowledge the role of the private sector and to support it. However, this attitude often sits alongside a willingness by academic staff in the public sector to accept well-paid part-time teaching appointments with private institutions and by former public vice chancellors to take up posts in the private sector after their retirement.

While the private sector does contribute to meeting the national targets for higher education, there are concerns about its reliance on market-orientated disciplines and the lack of research activity which can inform and strengthen teaching. Another major concern is that of equity, since the students come from social classes with the ability to pay (and usually from the wealthier urban areas), since very few private institutions can afford to set aside funding for scholarships and bursaries. In countries with low participation rates in higher education, this criticism can also be levied at the public sector, since entrance to universities is always very competitive (as in

Table 5 Numbers of private institutions of higher education (HE) in selected countries

	<i>Numbers of public HE institutions</i>	<i>Numbers of private HE institutions</i>	<i>Private foreign branch campuses</i>	<i>Comment</i>
Malaysia	55	27	4	Both public and private are subject to the Malaysian qualifications agency.
Kenya	7	18	nil	Private institutions are at various stages of accreditation
Nigeria	27 Federal 31 State	34	nil	The National Universities Commission oversees both public and private institutions.

Source: Compilation by the author from websites of the national agencies in the countries named.

the Sri Lankan example mentioned above) and only those with access to the best urban private secondary schools have much chance of gaining a university place.

If a foreign university wishes to establish a campus in an emerging economy, it is regarded as a private-sector entity, but is also subject to additional scrutiny because of its origin. Most countries are uncertain about how to treat enquiries from overseas institutions, although some such as Singapore welcome them within their long-term strategy. In Asia, the Chinese government has temporarily stopped approvals of overseas campuses, after granting three to the UK, and the Indian government has faced political difficulties in getting any legislation on international providers through parliament. Malaysia has encouraged the creation of foreign-branch campuses for some time and Vietnam has made an open invitation to foreign providers, following the success of Royal Melbourne Institute of Technology (RMIT), Vietnam.

Maintaining Quality, Relevance, and Employability

National strategies for higher education have to face the question of how to maintain the quality of provision as the system expands. It is all too easy for more to mean worse and for the growth in the academic profession to have to rely on recruiting new staff at lower levels of talent. This inevitably leads to a vicious circle of lower standards of learning, poor quality of graduates, and increased unemployment as they are rejected by employers.

The programs adopted by governments to improve quality include:

- Establishing accreditation and quality-assurance regimes that follow best international practices and are validated where possible by regional or international peers. China is seeking to do this and India is yet to build on the foundations of NAAC.
- Restoring practices such as the use of external examiners to prevent insularity. This will depend on institutional budgets meeting the cost unless subsidies are provided by government.
- Investment in academic-development programs such as research fellowships with overseas institutions and exchanges. The programs of the HEC in Pakistan illustrate the scale of investment needed, if standards have been allowed to fall.
- Subject reviews of curricula by panels with academic members from other countries.
- Encouraging twinning or franchising arrangements with reputable overseas partners, as has been the policy of Malaysia.
- Promoting the establishment of high-quality private institutions (backed by international partners if possible) that will provide benchmarks for the public sector.

- Investing in regular staff development and training programs for academic staff, both prior to appointment and mid-career.

The important point for governments to acknowledge is that academic quality is an international issue. While there are many hundreds of institutions aspiring to be world class, only a limited number can achieve this and questions of reputation, brand, and image are often based on subjective assessments between peers. These are rooted in regular exchanges and interchange between institutions. As both India and China show, a country cannot isolate itself from the global academic community and, once it embraces regular flows of academic staff, qualitative benefits will become apparent.

Closely related to quality is the issue of the relevance of the learning experience. If universities are not providing graduates with the skills that make them employable, much of the national investment will be wasted. A common complaint of graduates in many countries is that the curricula they study in public universities are out of date and irrelevant to today's job market. By comparison, private universities are seen to be better able to modernize and make their graduates employable. In India, 45% of the public-sector enrolment is in arts and humanities subjects, while over 80% of enrolments in professional disciplines are offered by the private sector (Aggarwal, 2007).

Conclusion

This article has shown that the issues facing higher education policymakers in emerging countries are very similar to those in the more developed world. All the countries concerned are keen to expand the numbers of their citizens enjoying higher education; some like China and India have rapidly growing economies and are well positioned to make the extra investment needed. India's decision in 2008 to make a fivefold increase in funding for tertiary education proves that point. Others do not have the resources and have to face either the tricky political question of cost sharing by introducing tuition fees or develop their higher education systems in partnership with the private sector and foreign institutions.

The benefits of such investments are already clear from countries that are further down the road such as Malaysia. Previous generous support for public institutions and a liberal attitude to foreign providers has positioned Malaysia at number 40 in the World Bank's Knowledge Economy Index, compared with China at number 75 and India at 101. Its gross enrolment ratio is well on the way to reaching the levels in the most developed nations.

Bibliography

Agarwal, P. (2007b). Higher education in India; growth, concerns and change agenda. *Higher Education Quarterly* **61**, 197–207.
National Knowledge Commission (2006). *Note on Higher Education*. Government of India New Delhi. <http://knowledgecommission.gov.in/downloads/recommendations/HigherEducationNote.pdf> (accessed August 2009).

Further Reading

Agarwal, P. (2007). *Private Higher Education in India*. London: Observatory on Borderless Higher Education.
Bhushan, S. (2006). *Foreign Education Providers in India: Mapping the Extent and Regulation*. London: Observatory on Borderless Higher Education.
(1997). China: Higher education reform Washington, DC: World Bank.
de Burgh, H. (ed.) (2008). *The China Impact: The Westminster Hearings on China's Economic Development and the UK*. London: University of Westminster, China Media Centre.

Kynge, J. (2006). *China Shakes the World*. Boston, MA: Houghton Mifflin.
Recommendations on Higher Education (2006). National Knowledge Commission. New Delhi: Government of India.
Salmi, J. and Hauptman, A. M (2006). Innovations in tertiary financing: A comparative evaluation of allocation mechanisms. *Education Working Paper*. World Bank.
University Grants Commission (2006). *2004-05 Annual Report*. New Delhi.
World Bank (2007). Knowledge Economy Index produced by the Knowledge for Development programme. www.worldbank.org (accessed August 2009).

Relevant Websites

www.bc.edu – *International Higher Education* Quarterly publication of the Center for International Higher Education, Boston College.
kolds@wisc.edu – Globalhighered from Kris Olds.
www.obhe.ac.uk – Observatory on Borderless Higher Education.
www.universityworldnews.com – University World News.

Human Resource Issues in Higher Education

E El-Khawas, George Washington University, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

Universities and other higher education institutions address the human resource issues faced by all large organizations. Their rules and procedures provide an orderly structure of roles and appropriate conditions of employment, and arrange for recruitment, promotion, and retention of workers as well as for their retirement or dismissal at some point.

Higher education institutions also face distinctive human resource issues related to their academic staff, which makes up the largest and most critical part of a university's workforce. Institutions of higher education must adjust their general personnel rules to the special nature of tenure and other forms of job security, to the multilayered structure by which academics are evaluated, and also to the constraints imposed by norms of academic governance and academic freedom. As academics often account for a significant part of a university's workforce, appropriate and up-to-date human resources policies are vitally important to all universities and systems of higher education.

This article on human resource issues in higher education focuses on the professoriate, and is organized into two main themes. First, it examines issues arising during the different stages in the life cycle of the academic career, from recruitment into initial positions to advancement in one's career. Second, the conditions of employment for academic work are discussed, with special attention to the impact of recent pressures for greater productivity. The article begins with a brief outline of recent challenges facing universities and higher education systems. Over the last few decades, significant external pressures have been felt, with potentially transformative implications for the shape and responsibilities of the academic workforce. It must be acknowledged that, because of space constraints, the article focuses on the general trends and patterns widely found among countries today. Readers may consult the sources provided in the section titled 'Further reading' for discussion of country-specific approaches.

A Changing Context

Higher education institutions in almost all countries around the world have faced new challenges during the last few decades, with the outlines of significant impacts on university life already emerging. Rising enrolment, spurred by demands for wider access, has placed higher education in the public spotlight and has led to a heightened climate of expectations and closer public scrutiny. At the same time, new technology, constrained financing, and

increased competition have exerted other pressures on academe. Business models have been applied at many institutions of higher education, resulting in greater attention to financial efficiency and managerial accountability. Increasing globalization of business and communication, along with new imperatives to adapt to a knowledge-based economy, have placed new burdens on universities to aspire to international standards for academic research and institutional reputation.

Taken together, these secular trends have impacted most institutions and systems of higher education around the world, with significant repercussions for the academic profession. As several studies have shown (Altbach and Lewis, 1996; Henkel, 2001; Kogan *et al.*, 1994), the academic work force today, in most countries, involves a greater diversity in positions, settings, and expectations. To be sure, the classic model of academic work still remains a model in which persons with advanced degrees in a specialized field have lifelong university appointments that combine research and teaching (cf. Clark, 1987). In many countries, however, other models have emerged, with different combinations of academic tasks for those taking appointments in nonuniversity institutions, or different terms of employment being offered to those accepting temporary or part-time academic positions. Even for established academics, there are new pressures to secure research funding, to develop productive partnerships with industry, or to modify their teaching and the curriculum to respond to changing student interests. According to some analysts (e.g., Slaughter and Leslie, 1997; Henkel, 2001; Harman, 2003), significant reshaping of the academic profession has resulted, in large part due to the specific policy choices made by government and by university administrators.

Does this trend apply only to certain universities and countries, or does it depict a long-term scenario likely to affect all types of higher education institutions across most settings? Currently, there is no systematic information to answer this question, especially with respect to the experience of developing countries. To be sure, dramatically new models of academic work have emerged in only a few settings. Substantial uncertainty remains as to whether a fundamental change is taking place. Is the long-dominant paradigm of traditional academic careers merely shifting to accommodate new challenges by offering new, more efficient models at the margin? Instead, is the nature of academic work undergoing deep structural shifts that will create significantly new forms? Have policies creating a second-tier of academic positions

become widespread? Will the classic model disappear from contemporary practice? This review of changing policies and practices in a wide range of countries suggests that a serious transformation in the nature of academic life is underway, despite uncertainty about its long-term effects.

Recruitment and Hiring for the Academic Career

A distinctive aspect of the academic workforce is that almost all positions are specific to a scholarly discipline and often to an area of specialization within the discipline. To recruit and hire members of academic staff, countries around the world generally follow a universalistic or meritocratic model that acknowledges the importance of specialized knowledge and skills. Available positions, with duties and necessary qualifications described in some detail, are formally announced and qualified individuals are given an opportunity to apply for the positions. Selection and appointment to each available position are governed by procedural rules in which the quality of the candidate's qualifications weigh most heavily.

Countries differ in whether recruitment and hiring are conducted centrally, for example, at the ministry level, or can be conducted on a decentralized basis. In some settings, the university or the department arranges the recruitment. Typically, general rules set terms of employment, in some countries based on formal regulations. Even in countries characterized by more open, market-based recruitment and hiring patterns, informal norms provide a general framework. Data on average salaries and other terms of employment for each academic rank, for example, are widely available in the United States and other market-oriented countries.

Planning issues sometimes arise as to whether a country can depend on an adequate supply of candidates for new positions, especially in more specialized or demanding fields. The main source of recruitment is found in a country's own production of individuals holding the doctorate or other highest degree in each field. For some countries, the supply of new doctorates routinely exceeds the number of academic positions available in universities, with the excess finding positions outside of academe or in other higher education institutions. Other countries have difficulty filling positions, including some that suffer from brain drain as their best graduates take academic positions in other countries. International competition in recruitment has grown over the last decade, putting pressure on some higher education institutions to introduce stronger incentives to attract highly qualified candidates. Higher pay, provision of research support, or lighter workloads are among the incentives offered. Some countries have modified the types of entry positions they offer, as illustrated by Germany's adoption of 3-year contracts, renewable once, for junior professorships.

The overall constraints on university funding have introduced other issues affecting recruitment patterns. One general response has been to meet staffing needs with part-time or temporary appointments. While practice varies greatly across different countries and types of higher education institutions, with nomenclature varying widely as well, financial considerations are a major reason that large numbers of part-time and temporary appointments are being made.

Most countries have witnessed an expansion in part-time teaching positions in which qualified individuals receive contracts on a term-to-term basis, according to the university's needs. Individuals in these positions typically teach a specific course or for a certain number of hours but have few other duties, and may not have offices. Traditionally, such part-time or adjunct positions were used occasionally, in response to specific needs or opportunities, and were justified as a way to bring specialized or real-world knowledge to teaching (e.g., accomplished artists, writers, or businesspersons). Increasingly, higher education institutions use part-time positions to cover a significant proportion of the institution's teaching needs, and do so for pragmatic reasons; such positions avoid the commitment to regular, permanent positions and may allow lower salaries or fewer benefits to be offered. In many countries, such part-time positions are undertaken by persons moonlighting in these roles after completing their obligations tied to their regular employment positions. In other countries, individuals may accept part-time positions at several different institutions after failing to secure a full-time university position.

There has also been increased use of full-time temporary appointments, contracted for a year or other limited term. Such positions are sometimes termed as visiting positions. Persons in these positions take on the full range of academic duties – teaching, research, and service – but the institution does not incur long-term employment obligations.

These two trends – toward greater use of part-time and temporary appointments – raise many questions of how systems of higher education plan for their human resource needs over the long term. In settings where short-term appointments are not renewed, considerable institutional effort is sometimes needed to find and hire capable individuals for these appointments, with last-minute decisions, based on enrolment, sometimes requiring compromises with what skills are sought. In other settings, individuals on temporary appointments provide good service and have their appointments renewed several times, raising questions of why they are kept in temporary arrangements. These appointments, which function largely as teaching-only appointments, may fill short-term needs and offer cost savings for universities; however, academic departments and programs must balance the benefits of substantial, continuing reliance on this type of staffing

with the need to maintain sufficient stability and long-term strength in the areas of expertise they need.

Another issue for systems and institutions is to consider the implications of a two-tier system of appointments, with differing employment terms and benefits, especially in job security. A limited, second-tier career line is emerging, in which individuals who have held temporary positions may find it impossible to gain a permanent academic appointment because, with heavy teaching obligations, they have not been able to develop a strong record of scholarly productivity. In some countries, inequities in gender and minority status are strongly linked to this emerging two-tier system, with less-favorable academic appointments – positions in nonuniversity institutions, in part-time or temporary positions – being occupied more often by women or minority individuals. The question for systems of higher education is how to determine the right mix and how to keep the negative consequences of using temporary positions to a tolerable level.

Advancement in the Academic Career

Countries have differing practices in the types of academic appointments or ranks that are available, and how formal the requirements are for gaining promotion to a higher rank. Some models are quite specific, including the well-known progression from assistant professor to associate professor to full professor. Typically, a formal set of procedures and guidelines have been developed to guide each step in the promotion process. At each stage, the individual academic must submit documentation showing that his or her accomplishments meet expectations for moving to the next rank. Academic self-governance is a distinctive part of the process for reviewing this documentation. In some countries, academics review the individual's record as a preliminary step before government officials make decisions; in other countries, a committee of academics makes the decision (promotion or not), which is then affirmed by university or governmental officials as long as formal guidelines have been followed.

The first phase of the academic career is a lengthy probationary period. Countries vary on how this is regulated but, in general, new appointees are under continuing pressure over a 4- to 6-year period to demonstrate that they can meet high expectations for research and teaching, develop good relationships with other academics and, generally, can build their scholarly reputation. In some countries, probation takes place within one university and can lead to permanent appointment within that university. In other countries, a young academic, after the initial appointment, must formally apply for a permanent or senior position at another university. Some countries require further specific accomplishments (e.g., a habilitation or other

formal evidence of research) before a young academic can seek a permanent senior post.

These models, reflecting different ways to obtain evidence of scholarly productivity, represent attempts by universities or higher education systems to control the overall quality and shape of their academic workforce. The more flexible models sometimes create imbalances in the supply of academics at each career stage. If most academics are in senior positions, few openings may be available for beginning academics; at another time – after increased enrolment, for example – most academics may be relatively new, with few senior academics available to share their experience. Recently, a number of countries are aware that their current academic workforce is aging, with sizable numbers likely to retire within the same decade.

In light of the increased hiring of part-time and temporary appointments that is found in many countries, new issues have arisen about promotion or renewal of individuals with such appointments. Sometimes, strict policies affirm that persons with temporary appointments cannot be renewed or promoted, but policies in other settings allow for renewed contracts, sometimes with terms of 2 or 3 years instead of 1-year renewals. Whatever the details, individuals with part-time and temporary appointments face differential treatment compared to those with full-time academic appointments.

After they have gained permanent appointments, academics are expected to continue to be accomplished in research and teaching. In some cases, university officials require regular reports of one's scholarly productivity and confirmation that contracted obligations for hours of teaching or other service have been met. In other settings, an elected committee of academics reviews annual or biannual reports. From the viewpoint of the individual academic, however, this type of continuing review is not generally difficult, and at best may be the basis for an increase in pay level.

In most academic systems, after achieving promotion to the rank of professor, academics encounter a typically lengthy period in which formal expectations are relatively steady – to continue with research, to meet teaching, as well as service obligations. Notably, individuals differ greatly in how hard they work on their research, how prolific they are in publishing and sharing their work, and how well known they become in their discipline. While the ideal of high productivity is quite clear, surveys have shown that large proportions of academics do quite well in teaching and service but have only modest research accomplishments (e.g., Altbach and Lewis, 1996).

In most countries, retirement for academics generally falls between 60 and 65 years of age, based on rules applied to all academic staff, regardless of their own interests or level of performance. In some countries, academics are sometimes awarded emeritus status upon retirement, a largely honorary title but one that sometimes provides access to

university offices, library, and other services. Some academics take retirement but stay active with consultancy, with short-term administrative assignments at their institution, or with the support of continued research funding.

Conditions of Employment

The classical model of academic work remains as a valued and desired view of academic life. Under this model (cf. Clark, 1987), an academic carries out a bundle of interrelated tasks – teaching in one's areas of expertise, conducting research and producing scholarship on a continuing basis, providing service through committee work and limited administrative duties related to his/her departmental unit, as well as contributing to one's discipline on editorial boards, committees, and through presentations at academic conferences.

Many observers contend that this classical model has been eroded, and that the nature of academic work has undergone significant change over the last several decades. Some general trends have affected the academic workplace in most countries, especially the necessity of adapting to constrained funding. Higher education systems have found that, as enrolment and participation rates have grown, higher education's share of a nation's budget increases, moving it into competition with other broad areas of public spending. Recent experience has been that higher education's share has not grown in such countries, even as enrolments have increased.

Another general trend affecting academic work has been a shift toward managerial concerns in the administration of higher education institutions. Human resource issues have become focused on discussions of efficiency, measuring outcomes, and strategic planning. In some settings, academic deans are expected to raise funds and to generate certain levels of revenue on a continuing basis. In other settings, entrepreneurial actions are expected, sometimes to develop partnerships with business and, at other times, calling for the development of new educational offerings specifically designed to yield net income.

As to be expected, many academics have been highly critical of these changed demands. Studies conducted in several countries have documented increased levels of dissatisfaction and disillusionment among academics but considerable adaptation has been found as well (Kogan *et al.*, 1994; Harman, 2003). While it is unclear whether perceptions will improve over time, and dissatisfaction has certainly not led large numbers of academics to leave their positions, universities, and systems of higher education need to be mindful of possible long-term damage to the academic profession.

The combined effect of constrained funding and managerial perspectives has undoubtedly led to changes in how academics work. Several analysts (e.g., Harman, 2003;

Amaral *et al.*, 2003) have shown that, in a wide range of countries, the traditional bundle of tasks – teaching, research, and service – has become more demanding. Academics are teaching many more hours, and many more students, in response to greatly increased enrolments. In some settings, academics have been compelled to commit more time to administrative and committee assignments, as their universities have engaged in more evaluations of their operations, reorganized or introduced new structures or program offerings, or responded to an array of mandated external reporting and assessment. For some countries, entrepreneurial pressures have introduced additional work, with academics participating in the planning and delivery of revenue-producing partnerships and similar activities, often in other countries.

Another trend involves a fragmentation of academic tasks (Slaughter and Leslie, 1997). Teaching-only positions have emerged in some countries as a matter of government policy. In other settings, institutions impose heavy teaching requirements and give no systematic support for research, creating positions that are, in effect, teaching only. Especially in the sciences, research-only positions have proliferated as large-scale research grants and contracts have become a larger part of university activity. So also, a small number of high-profile academic positions have emerged, in which academic stars achieve much scholarly productivity, are active contributors to policy debate, widely quoted for their views, and called to serve on important projects and commissions. Typically, such stars have very limited teaching and service obligations.

Still another trend involves a sharp differentiation between the work done by permanent staff and that done by temporary staff. Some systems and institutions have allowed their staffing to evolve in a way that separates academics into core and secondary roles. This pattern first emerged in the United States with instruction based on distance learning and other electronic technologies for course delivery. Under this model, a large number of academic staff work under temporary contractors carry out various teaching duties (e.g., as facilitators of online or actual discussions, and as graders of student papers), while a small permanent academic staff offers continuity by designing, evaluating, and implementing academic programs but engaging in very little direct teaching.

In addition, these trends have led to greater disparities in the compensation and benefits accorded to academic staff. Traditional academic values emphasized a general level of equality among academic staff, with small gradations in salary and benefits according to rank and length of service. In many settings, a central decision body, usually government ministries, determined salary levels for each academic rank, and established the terms for pay increases that applied to all academic staff. Merit-based or performance-based pay increases were unusual, tied to promotion in rank. However, in today's changing

academic world, the picture has become more complicated. Some countries have allowed separate, and sharply lower, pay scales for temporary and part-time academics, compared to those on permanent appointments. In other settings, academics judged to be more productive have received higher pay. Overall salary levels, however, have stagnated in some countries, with academic pay falling well below levels in comparable professions.

Similar differentials have emerged with regard to other benefits of academic employment. Opportunities for travel to conferences or for research funds and space are increasingly distributed according to the new hierarchies, with reduced opportunities for second-tier academics.

An unusual aspect of academic work is the ability to move temporarily to other settings. Unlike other professions, academics have several mechanisms allowing them to engage in new activities without giving up their long-term appointments. Typically, an academic arranges to be at another higher education institution during a sabbatical leave (a release from all obligations for one term or 1 year, after completing some years of service, intended to allow uninterrupted attention to scholarly research). Established academics may take positions as visiting professors at another university (e.g., in order to work on a different research project) while on temporary leave from their own institutions. A similar opportunity, possibly increasing in frequency, is for academics to leave their university temporarily to take positions in government or industry. In the past, such appointments were seen primarily as a benefit to the individual academic but, recently, as partnerships become more strategically important to universities, these temporary appointments help enhance university relationships with current or future partners.

Finally, mobility between institutions – changing from one permanent position to another – is also distinctive to academic work. In some systems of higher education, it is necessary to move in order to gain a position at a higher rank. In more market-oriented systems, established academics may seek other postings as a way to advance their careers more rapidly. There can be substantial differences in mobility across countries and even from one time period to another in the extent to which such changes are made. When the entire higher education system is expanding rapidly, established academics have the opportunity to move to newer institutions. In other periods, with tight financial circumstances, academics may experience very little mobility.

Systems and institutions of higher education, in turn, need to respond to such shifting opportunities for mobility. In more competitive periods, institutions have typically needed to improve the conditions of work for their own academics, as a way to preserve a stable staff. When systems are feeling competitive pressures for productivity, bidding wars can develop for certain high-performing academics. The Research Assessment Exercise in the United Kingdom,

which rewarded those departments with highly productive academics, led to significant institutional attempts to improve their profile by hiring strong academics within certain assessment time periods.

On the whole, it must be acknowledged that the academic profession is quite stable. Some early career academics may need to move when it appears that they will not gain a long-term appointment in their initial setting, perhaps simply because there are no vacancies in senior positions but, once they achieve a permanent appointment, academics in the past have tended to spend the rest of their career at one institution (e.g., Altbach and Lewis, 1996). In the changing context of academic work, however, this pattern may be changing.

Conclusion

This article on the recent trends in the conditions of academic life has relied on reports from a limited number of countries, along with a few broad, multi-country studies. Given the substantial change that appears to be underway, judgments have been necessarily tentative about the overall impact of trends. Fortunately, readers can soon look to the findings from a significant new international study on the changing academic profession, now underway (Brennan, 2006), which should reveal much about the status of the professoriate on a worldwide basis, clarifying the effects of many changes over the last decade.

While the broad direction of change is relatively clear, there is much uncertainty. This new international study should help establish whether the classic academic model is still dominant or, instead, how prevalent the trend has been to a more differentiated and demanding model. It should also shed light on how current members of the academic profession view their circumstances. Some have argued that morale and levels of commitment to academic work have declined in response to recent trends, but others believe that the overall picture shows a mutual accommodation, as policies have been modified where they were too harsh and academics have adjusted to new circumstances.

See also: Academic Freedom; Productivity of University Faculty Staff.

Bibliography

- Altbach, P. G. and Lewis, L. S. (eds.) (1996). *The International Academic Profession: Portraits of Fourteen Countries*. San Francisco, CA: Jossey-Bass/Carnegie Foundation.
- Amaral, A., Meek, V. L., and Larsen, I. M. (eds.) (2003). *The Higher Education Management Revolution*. Dordrecht: Kluwer.
- Brennan, J. (2006). A changing academic profession: International perspectives. In *Higher Education Digest*, vol. 56, pp 1–5. London: Centre for Higher Education Research and Information/Open University.

- Clark, B. R. (1987). *Academic Life: Small Worlds, Different Worlds*. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching/Princeton University Press.
- Harman, G. (2003). Australian academics and prospective academics: Adjustment to a more commercial environment. *Higher Education Management and Policy* **15**(3), 105–122.
- Henkel, M. (2001). *Academic Identities and Policy Change in Higher Education*. London: Jessica Kingsley.
- Kogan, M., Moses, I., and El-Khawas, E. (1994). *Staffing Higher Education; Meeting New Challenges*. Paris: OECD.
- Slaughter, S. A. and Leslie, L. L. (1997). *Academic Capitalism: Politics, Policies, and the Entrepreneurial University*. Baltimore, MD: Johns Hopkins University.

Further Reading

- Altbach, P. G. (ed.) (2003). *The Decline of the Guru: The Academic Profession in Developing Countries*. New York: Palgrave.

- Higher Education Funding Council for England (2006). *The Higher Education Workforce in England: A Framework for the Future*. HEFCE, 2006/21.
- Enders, J. (ed.) (2001). *Academic Staff in Europe: Changing Contexts and Conditions*. Westport, CT: Greenwood Press.
- Musselin, C. (2005). *Le marche des universitaires (The Academic Labor Market)*. Paris: Presses de Sciences Politiques.

Relevant Websites

- <http://www.aaup.org> – American Association of University Professors.
- <http://www.bc.edu/cihe> – Boston College, Center for International Higher Education.
- <http://www.hefce.ac.uk> – Higher Education Funding Council for England.
- <http://www.nces.ed.gov> – National Center for Education Statistics, National Survey, National Study of Postsecondary Faculty.

International Accreditation in Higher Education

A Stella and D Woodhouse, Australian Universities Quality Agency, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Over the last 20 years, many countries have created quality assurance (QA) bodies to assist, check, or regulate their higher education institutions (HEIs). These bodies exercise their QA responsibility through various modes and methods that essentially fall under one of the three basic approaches – accreditation, assessment, and quality audit. Of these modes and their combinations, accreditation is the most widely used term in many countries around the world and many QA bodies have some form of accreditation practices. Accreditation being the predominant mode of QA, the term accreditation is used in the following pages to denote all forms of QA.

A recent phenomenon in accreditation is that institutions are looking beyond their national borders, and they voluntarily undergo multiple accreditation processes from QA bodies in different countries. What can a foreign accrediting body offer that is not provided by an institution's local agency? The answer lies in addressing the emerging challenges and changes in the higher education (HE) sector. The HE sector in many countries is changing dramatically characterized by increasing international operations of HEIs, demand for value for money from the various stakeholders, increasing academic and professional mobility across national borders, demand for recognition of qualifications, and growth of regional and global trade initiatives. These are the new rationales that drive international engagement and put pressure on HEIs to take an international approach to accreditation.

Changing Rationales of International Engagement

Internationalization has now come to be defined in very broad terms. From an initial focus on bringing international students to the domestic campus, it has expanded to include the international engagement of HEIs through collaboration in curriculum development, teaching, research, and outreach. There are two dimensions to the international engagement of HEIs. Institutions may integrate an international or intercultural aspect into the teaching and research in their domestic campus and encourage international students and staff to join the domestic campus; this is called internationalization at home. HEIs may cross national borders to offer their services in other countries and that is internationalization

abroad. Both these aspects of international engagement are increasing rapidly and this affects various stakeholders in different ways. Stakeholders such as governments and the public who once emphasized accountability and quality improvement in the national context now extend their attention to the international operations of the HE sector.

Furthermore, the drivers of internationalization have changed. Until recently, cultural, educational, and mutual understanding motives were the common historical basis of internationalization policies for HE. For one or more of these reasons, every country finances staff and students to study or research abroad, via university bursary schemes, bilateral or multilateral agreements, and policies to promote mobility. These rationales and the attendant policies are still present today, but they have been complemented by new trends and rationales. Although mutual understanding and international cooperation in teaching and research rank high on many countries' internationalization agenda, economic and revenue-generation rationales have become much more important recently and have sometimes become primary (OECD/Norway, 2003). The economic rationales have made the international clientele scrutinize the value for money spent on study. To be able to remain competitive, especially in the overseas marketing of the programs, HEIs are under pressure to affirm their quality in an internationally comparable way.

In parallel, mobility of students and staff has grown considerably. Regional and global trade agreements are encouraging the movement of professionals across national borders. One of the most important consequences of this development is the demand for academic and professional recognition of qualifications. Discussions on the World Trade Organisation's (WTO) general agreement on trade in services (GATS) have already identified problems in recognition of qualifications as an indirect barrier to liberalizing trade; as countries join GATS, issues related to recognition issues are gaining greater momentum, and HEIs look for ways to enhance the portability of the qualifications they offer.

Closely related to the recognition of qualifications is the issue of accreditation. Credential evaluators, recognition bodies, employers, and professional bodies show interest in determining the quality of an institution, program, or qualification, and they are inclined to take account of the quality of the institution in their decision making. Many of the problems in this context are around determining whether a program or institution awarding a certain qualification meets certain threshold levels of

quality. This attention to levels of quality and standards exerts pressure on HEIs to assure the quality of their services to clientele spread across national borders.

In recent years, there have also been criticisms about the quality of internationalization itself. The QA arrangements and academic oversight an institution has for its domestic offerings may be robust, but it may not be possible to generalize that to the international activities because of variations in academic oversight and coordination. This situation has resulted in severe criticism, especially of the marketing of offshore programs in receiving countries. To address this, HEIs take a strategic approach to their planning and development; today, internationalization and assurance of its quality are given an institution-wide systemic and strategic attention.

Thus, the interplay between various internal and external pressures has resulted in HEIs exploring strategic approaches to their QA processes with an international orientation. This has triggered developments in three major streams and they are outlined in this section. First, HEIs have opted for multiple accreditation and international accreditation. Second, associations of HEIs have started approaching QA issues that cross national borders through codes of practices and agreements. Third, QA agencies and their networks have expanded their scope and approach to QA in response to the internationalized HE sector.

HEIs in the Internationalized Landscape

National QA bodies often focus on assuring the quality of programs offered in their respective countries by their domestic institutions. In such systems, HEIs active in international engagement realize that they have outgrown their national contexts, including their national QA systems. Acknowledging that the national QA systems are not yet ready to serve the cause of international engagement of the order the HEIs need, many HEIs have opted for QA strategies that go beyond the national approaches.

Of the various factors that have motivated HEIs to take an international approach to QA issues, the demand for academic and professional recognition of qualifications ranks very high. Often the recognition of qualifications across borders is complicated by the lack of transparent national criteria for quality and standards. Even in Europe, which is known for student mobility programs such as European region action scheme for the mobility of university students (ERASMUS), supported by the network of information centers and recognition bodies, incomplete recognition for the credits earned in study abroad and delay in getting the recognition was a problem till recently. The situation has been improved by the wider introduction of the European credit transfer scheme and the Bologna process. However, in other parts of the world,

the recognition of qualifications beyond national borders is not very encouraging and the impact that national QA systems can have to address this issue is still weak.

To get past the limitations of national QA systems in this respect, some HEIs opt for certification by international organizations such as the International Organisation for Standardisation (ISO) for their quality management systems. ISO 9001 is an international standard which relates to the achievement of quality in an organization. Although some institutions find it helpful to use ISO 9001 to demonstrate their quality in an internationally comparable way, there is a general criticism that ISO standards originate from the manufacturing context and are not therefore well matched to the educational sphere. Few HEIs find the effort required to achieve full ISO 9001 certification provides adequate returns, although many institutions have ISO 9001 for part of their operations, usually administrative areas.

More commonly, HEIs attempt to overcome the barriers to recognition of qualifications by undergoing QA processes of organizations that operate internationally. Often, one HEI undergoes multiple reviews or accreditations. This is particularly found in professional areas of studies where student and staff mobility is high. The European Foundation for Management Development (EFMD) under its European Quality Improvement System (EQUIS), the Association to Advance Collegiate Schools of Business (AACSB), and the Association of MBAs (AMBA) are professional associations that accept international applications for their accreditation process in business studies.

Associations and Networks of Higher Education Institutions

In addition to such direct methods for addressing academic and professional recognition of qualifications, institutions also assure themselves and others of their quality by adherence to codes of practice and guidelines established by associations and networks of HEIs. Codes of practice are common at the national level and for the most part they are statements of principles related to moral imperatives in international engagement. The code of practice and guidelines for Australian universities on provision of education to international students by the Australian vice-chancellors committee; the code of practice for the assurance of academic quality and standards in HE: collaborative provision, QAA, UK; principles of good practice for educational programs for non-US nationals by the council for HE accreditation; and the code of ethical practice in international education developed by the Canadian Bureau for International Education are a few examples.

Making a commitment to adhere to codes of practice serves as a generic type of QA and it can work well in mature HE systems. If the code is framed as a guideline

rather than as a policy, it may not be possible to have a system in place to monitor and assess compliance. Instead, such a code appeals to the ethics and conscience of the institutions and the staff who are involved in international engagement and it tries to develop a set of values and principles to guide the process.

In addition, associations and networks of HEIs have initiated projects on quality that cut across national borders. The institutional evaluation program (IEP) initiated by the European University Association (EUA; formerly CRE) is an example that has a focus on quality management. In 1994, EUA launched the individualized education plan (IEP) for its member universities to assess their strengths and weaknesses in quality management. The review is based on a self-evaluation and external peer review conducted by senior international institution leaders. So far, over 150 universities in Europe and worldwide have participated in the evaluation. EUA reports that the major benefits derived are an increased strategic capacity and strengthened internal quality culture – two essential attributes for dealing with current and future challenges in HE. Since 2001, EUA has also carried out sector-wide evaluations to identify and make recommendations on the systemic challenges and the common issues shared by all institutions in a given sector.

Other examples include the projects and benchmarking initiatives of the association of commonwealth universities on issues that are common to universities across national borders; the task force on accreditation established by the international association of university presidents; and a working group on accreditation set up by the confederation of European rector's conferences.

As internationalization becomes more widely understood and the process of internationalization matures, it is increasingly important that institutions of HE address the issues of quality assessment and assurance of the international aspects of their operations. This realization galvanized international and intergovernmental bodies with an interest in HE to pay attention to the international aspects of HEIs. For example, the OECD program on institutional management in higher education (IMHE) has taken a close interest in the international dimension of HE over a number of years. Since 1994, it has led an activity focussing on a cross-country analysis of institutional level strategies, the internationalization quality review process (IQRP). IQRP aims to help individual institutions of HE to assess and enhance the quality of their international dimension according to their own stated aims and objectives.

Several organizations of HEIs have produced position papers and declarations on issues related to quality in cross-border education, triggered by the GATS debate. They too serve as guidelines in shaping the international approach of the HEIs toward QA. Examples are the Accra Declaration on GATS and the Internationalisation of

Higher Education in Africa (2004) by the Association of African Universities; and the document Sharing Quality Higher Education Across Borders: A Statement on Behalf of Higher Education Institutions Worldwide (2005) jointly prepared by the International Association Of Universities, the Association Of Universities and Colleges of Canada, the American Council on Education, and Council for Higher Education Accreditation.

Response of the Quality Assurance Agencies

As HEIs collaborate with their counterparts across borders, it is but natural that they expect the quality assurance agencies to follow a similar approach and collaborate with QA agencies across borders. If QA agencies do not develop this capacity to cooperate with their counterparts, the serviceability of QA agencies to HEIs and the effectiveness of QA services will be seriously limited. While looking at the way the QA agencies have responded to the international engagement of HEIs, it is necessary to see the general QA developments as well as the developments specific to professional areas of studies where the professional bodies and professional accreditors do the gate keeping and act as QA bodies in their areas of specialization. Professional bodies are a type of quality agency, using many of the common methods, although their authority is confined to one discipline area.

Professional Bodies and Their Networks

Attention to the mobility of professionals has given a fillip to professional accrediting associations, whose members want international collaboration between associations in order to achieve international recognition of their professional qualifications. Systems of licensure, certification, and accreditation are emerging to support professional mobility. This gives rise to international definitions of quality of education in professional areas of studies. Professional organizations, accrediting bodies, and certification and licensure bodies are interested in developing mutually acceptable standards in cooperation with their counterparts in other countries and are considering mutual recognition (MR) agreements. There are instances of professional bodies accepting international applications and recognition agreements among a group of countries with mutually acceptable standards. The way the accreditation board for engineering and technology (ABET), a federation of the US professional engineering societies, expanded its scope is an interesting example.

As the demand for engineering mobility has increased, the assessment of the quality of education in engineering programs at institutions outside US has become increasingly important. To meet these needs, ABET has become

involved internationally through mutual recognition agreements, program evaluations, educational consultancy visits, assistance in developing accreditation systems in other countries, and in the accreditation of engineering programs outside US. The Washington Accord that developed out of this international approach of ABET is a successful example of a mutual recognition agreement. It is an agreement between the bodies responsible for accrediting professional engineering programs in each of the signatory countries. It recognizes the substantial equivalence of programs accredited by those bodies, and recommends that graduates of accredited programs in any of the signatory countries be recognized by the other countries as having met the academic requirements for entry to the practice of engineering. As of May 2007, there are ten signatories to this accord with four more in the provisional list.

The procedures underpinning this accord include circulation and consideration of any revisions made by any one of the associations to its procedures, and intermittent attendance at accreditation events by teams of representatives of the other members of the accord. It is to be noted that these agreements are between English-speaking countries or countries that are willing to operate in English for this purpose. Despite the wide attention to this accord, it has not spawned many copies.

Professions more tightly controlled by law or statute find the mutual recognition task more difficult. However, some progress is being made in a number of other areas. In architecture, the international union of architects (UIA) has developed the UIA Accord on Recommended International Standards of Professionalism in Architectural Practice and nine related accord policy guidelines. It is the first time the profession of architecture has adopted a global standard. Further, an international charter for architectural teaching was agreed in 1996, and work on this has resulted in the UNESCO–UIA validation system for architectural education.

The International Federation of Nurse Anaesthetists (IFNA) has been working on this for a number of years, and IFNA is the first international nursing or medical organization to have developed international standards and guidelines for education, practice, and ethics. The International Council of Nurses and IFNA have started addressing the concept of international accreditation, but it will be several years before it is fully implemented. There are also some bilateral MR agreements of more limited extent. The actions of the professional associations in this area overlap with the regional and global trade initiatives and consequently with the GATS requirements.

Although professional associations are themselves quality agencies, they do not commonly contribute to the design and development of the wider QA framework. On the other hand, general purpose QA agencies often engage in the QA developments of the professional sector. Most regional accreditors in US, QAA (UK), New Zealand

Universities Academic Audit Unit NZUAAU (New Zealand), and Australian Universities Quality Agency (AUQA) (Australia) collaborate with professional associations to varying degrees, from two-way information sharing, to collaborative review events.

General QA Developments

Just as HEIs are internationalizing, so is QA itself. Similar to the HEIs offering their services across borders, some QA agencies extend their service to international clientele or HE sectors across borders. One reason is the need for capacity development in QA in some countries. The second reason is the unsustainability of a dedicated QA system in some situations such as small island nations. It might be more practicable for these countries to approach another reliable QA agency(ies) for QA services. Auditing of the universities of Mauritius and the South Pacific, the latter covering 12 small island nations, by the QA agencies of neighboring countries, is an example. Another example is the accreditation carried out by the regional accreditors of the US for overseas non-US institutions. The current volume of this activity is low, but is increasing.

The networks of QA agencies have made a significant impact in creating awareness among the QA agencies on common issues related to QA of international activities. There is a growing acknowledgment by quality agencies that they should whenever possible recognize each others' decisions, and that if possible this recognition should be mutual between the agencies. Two agencies that are both operating at the internationally accepted level of competence may recognize each other's judgments about the institutions in their jurisdiction, and jointly provide assurance to governments, employers, students, and other institutions about the quality and standing of the institutions and their qualifications.

The International Network of Quality Assurance Agencies in Higher Education (INQAAHE) has discussed the mutual recognition of QA decisions among its members. The Asia-Pacific Quality Network (APQN) runs a project on mutual recognition and has a target of all its full members recognizing each others' judgments by 2010. The European Association for Quality Assurance (ENQA) has also been active in pilot projects on mutual recognition. The European Consortium for Accreditation, which is a sub-network of ENQA, and whose members are those agencies in Europe whose main evaluation activity is program accreditation, have committed themselves to develop a system for mutually recognizing each other's accreditation decisions. ENQA has also committed itself to developing a European register of QA agencies, and agency compliance with the European standards for external QA agencies will be one criterion for placement in the register. This register has the potential to facilitate mutual recognition developments.

MR depends on the trust and confidence QA agencies can have in each other's QA decision making. In recent years, the networks of QA agencies have paid attention to the question: what is a good QA system? Discussions addressing this question have resulted in the description of features that can be expected of ideal QA systems, and in the identification of ways in which QA systems may attain these characteristics. Notable developments include the INQAAHE guidelines of good practice (GGP), and standards and guidelines for quality assurance in the European HE area (ESG), 2005.

In addition to the networks of QA agencies, pointers and guidelines have also been developed by intergovernmental bodies such as UNESCO and OECD, drawn up by various stakeholders as well as experts in the field.

Role of Intergovernmental Bodies

Intergovernmental bodies such as UNESCO and OECD have also contributed to these developments. UNESCO regional conventions on recognition of qualifications are a notable example. UNESCO Conventions are legal agreements between countries to recognize academic qualifications issued by other countries that have ratified the same agreement. A number of recognition tools such as a diploma supplement and the code of good practice in the provision of transnational education (adopted by the Lisbon recognition convention committee) have been developed in the framework of the UNESCO Conventions. It was the Lisbon convention that established an explicit link between QA and the recognition of qualifications. Specifically, the diploma supplement, designed to facilitate international recognition of qualifications is required to show the QA regime under which the qualification was awarded. Unfortunately, it continues to be difficult to give effect to this link, and QA agencies and recognition bodies tend to operate in different spheres.

UNESCO–OECD guidelines on quality provision in cross-border HE (2005) propose good practices that can assist countries in assessing the quality and relevance of HE provided across borders and to protect students and other stakeholders in HE from low-quality HE provision.

Issues for the Future

Capacity Development

Traditionally, the national QA systems were not oriented to cover the international engagement of the HEIs. As the international engagement of HEIs increases, the expectation that national QA agencies be able to expand their scope is emerging as an imperative for all countries irrespective of the stage of development of their HE systems.

Surveys conducted by networks such as APQN, CINDA (in Latin America), AAU (in Africa), and INQAAHE have indicated that there is a need for capacity development in quality-related issues of the international activities. Similar results have been highlighted in the surveys conducted among the APEC economies as well as in the reports produced by UNESCO on mapping the QA systems in different regions. Consequently, many capacity-development initiatives are in progress, supported by national governments and intergovernmental and international bodies such as the World Bank and the UNESCO. In spite of attention to capacity development, unequal stages of development will continue to be a problem in the future. Till the capacity of the national QA systems to adequately cover the international engagement of their HEIs is raised, HEIs might be burdened with multiple and international accreditation to facilitate recognition of their qualifications across national borders.

Lack of Convergence among Diversity

Policies and practices of QA of the international activities of the HEIs vary even among countries of similar type, especially regarding internationalization abroad. QA agencies take various approaches to assuring the quality of this cross-border education. For example, in the UK, the Quality Assurance Agency (QAA) from time to time convenes a panel to visit a group of overseas operations of the British HEIs in one country or region. In Australia, AUQA's audits address the transnational operations of the university sector institution by institution. The regional accreditors of the US carry out visits to any operation that is established more than 40 miles from the institution's main campus, including across borders. If there is such a variation among the three major exporters, one can understand how difficult it would be in other situations to look for a common framework. Now QA agencies have realized that they need to ensure convergence toward some common framework and some QA agencies have taken initiatives to map the common framework. Developments in Europe through the Bologna process have the potential to facilitate this convergence, but how much convergence can be achieved in the QA of international activities in the near future is a big question.

Lack of Dialog and Cooperation among Quality Assurance Agencies

There is a need for strengthened dialog among QA agencies at two levels. First, communication among the QA agencies of major sending countries needs to improve. There are many instances where an institution of the receiving country that has partnership with HEIs from different countries is burdened with the QA regime of a number of those countries, embracing a number of separate reporting requirements and review visits. This could

be avoided if QA agencies of the major sending countries can establish an appropriate mechanism to work collaboratively. Collaboration is emerging, but it will take many years for the cooperation to materialize in ways that will reduce the QA overload of institutions.

Second, the dialog between the QA agencies of the sending and receiving countries is very weak. Differences in the capacity of the QA systems to consider international activities itself may be a barrier to facilitate a meaningful dialog among the QA agencies.

Lack of capacity, diversity in practices, and absence of a productive dialog among QA agencies lead to mistrust and stereotypic understandings about the QA of international activities. That perception is gradually changing in some countries due to the evidence that international engagement in HE can respond to human and social development needs, provide new opportunities, and increase the possibilities for improving workforce skills if managed appropriately. Teacher-education programs offered in Africa, when there was shortage of trained teachers, with the support of the Commonwealth of Learning through open and distance learning, is an example and India had an active role in this. Malaysia, Singapore, Hong Kong, Korea, China, Maldives, and Indonesia are examples of countries in the Asia-Pacific that have benefited from cross-border educational services. Today, countries have realized that there are ways to protect the interests of both the sending and receiving countries and they see cooperation as the best way to do that.

As the quantity of international activity grows, and generates corresponding QA attention, there is a danger of too much QA for international operations. In the long term, this might be a threat to innovation and creative learning models and promote a compliance culture, as HEIs have to undergo multiple QA regimes leading to overload of QA. This realization has resulted in many pilot initiatives for cooperation among QA agencies and their networks. Regional initiatives with political will such as the Bologna process and the Brisbane Communiqué are likely (eventually) to make a significant impact on these initiatives.

Conclusion

There are massive changes in the international engagement of institutions, both in nature and in quantity. This poses challenges for the institutions as they operate in different cultures and their students cross national boundaries. Maintaining high quality of these activities is the responsibility of the institutions (QA of internationalization) and also of the QA agencies (internationalization of QA). However, most QA agencies were established with national authority and national remits. They must work

(and indeed are working) on the mechanisms for assuring quality across boundaries and for collaborating with other national QA agencies. Networks of QA agencies are playing a central role here, in capacity-building for their member agencies, and advocating the development and use of good practice. Ensuring some convergence among diverse QA policies and practices, and strengthening cooperation among QA agencies, emerge as issues for the future. Until these developments take hold, there will be a mismatch between what the HEIs need in terms of international QA and what the various QA bodies are able to provide. The current widespread attention to these needs gives cause for optimism that a better alignment will gradually be achieved.

See also: The Bologna Process in European Higher Education; UNESCO's Role in the Development of Higher Education in a Globalized World.

Bibliography

OECD/Norway (2003). *Papers Presented at the Forum on 'Trade in Educational Services'*, 3–4 November 2003, Norway.

Further Reading

- Campbell, C. and Van der Wende, M. C. (2000). International initiatives and trends in quality assurance for European higher education: Exploratory trend report. *ENQA Occasional Paper*, Helsinki.
- Larsen, K. and Momii, K. (eds.) (2004). *Quality and Recognition in Higher Education: The Cross-Border Challenge*. Paris: OECD.
- OECD (1999). *Quality and Internationalisation in Higher Education*. Paris: Institutional Management in Higher Education (IMHE).
- Stella, A. (2006). Accreditation of higher education in the Asia-Pacific: What is at stake? In Sanyal, B. (ed.) *Higher Education in the World 2007*. Barcelona: GUNI.
- Stella, A. and Woodhouse, D. (2007). International dimensions of quality assurance in higher education. *Paper Presented in the INQAAHE Bi-Annual Conference on 'Quality Assurance: Coming of Age – Lessons from the Past and Strategies for the Future'*, 2–5 April 2007, Toronto, ON.

Relevant Websites

- <http://www.abet.org> – Accreditation Board for Engineering and Technology (ABET).
- <http://www.apqn.org> – Asia-Pacific Quality Network (APQN).
- <http://www.mbaworld.com> – Association of MBA (AMBA).
- <http://www.aacsb.edu> – Association to Advance Collegiate Schools of Business (AACSB).
- <http://www.auqa.edu.au> – Australian Universities Quality Agency (AUQA).
- <http://www.avcc.edu.au> – Australian Vice-Chancellors' Committee (AVCC).
- <http://www.coneau.edu.ar> – Comisión Nacional de Evaluación y Acreditación Universitaria (CONEAU).
- <http://www.chqa.org> – Council for Higher Education Accreditation (CHEA).

<http://www.enqa.eu> – European Association for Quality Assurance in Higher Education (ENQA).
<http://www.eqa.consortium.net> – European Consortium for Accreditation (ECA).
<http://www.efmd.be> – European Foundation for Management Development (EFMD).
<http://www.eua.be> – European University Association (EUA).
<http://www.iau.org> – International Astronomical Union (IAU).
<http://www.icn.ch> – International Council of Nurses (ICN).
<http://www.ifna-int.org> – International Federation of Nurse Anesthetists (IFNA).

<http://www.inqaahe.org> – International Network for Quality Assurance Agencies in Higher Education (INQAAHE).
<http://www.iso.org> – International Organization for Standardization (ISO).
<http://www.aau.ac.nz> – New Zealand Universities Academic Audit Unit (NZUAAU).
<http://www.qaa.ac.uk> – Quality Assurance Agency for Higher Education (QAA).
<http://www.unesco.org> – United Nations Educational, Scientific and Cultural Organization (UNESCO).
<http://www.washingtonaccord.org> – Washington Accord.

Internationalization of Higher Education

M van der Wende, University of Twente, Enschede, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

The Shaping and Motivation of Internationalization in Higher Education

The international dimension of higher education is widely recognized. Knowledge transcends borders, research has an intrinsically international character and wandering students were already a feature in medieval Europe. Arguments of this type usually provide the basis for claiming the international character of higher education, which is generally acknowledged. This does not, however, substantiate the often advanced position that higher education has always had, and has inherently, an international character. In fact, nation-states have played a crucial role since the nineteenth century in the development of the modern university. Neave (2001) refers to two centuries of nationalism in higher education. Very few higher education institutions can lay claim to a centuries-old international tradition for the simple reason that most of them were established after 1900 or even after World War II. The modern university is a national institution, a creation of the nation-state (Scott, 1998). Furthermore, the extent of international activity and orientation in teaching and research varies enormously depending on the discipline and professional area concerned. Finally, the proportion of students and staff who actively participate in international activities is usually quite limited and internationalization may therefore remain marginal on many campuses.

Nevertheless, in the second-half of the twentieth century, a gradual change in the extent to which higher education was nationally determined and orientated can be observed. International cooperation between nations in the field of higher education emerged following major political and economical developments (Blumenthal *et al.*, 1996; De Wit, 2002). The post-war reconstruction of Europe was supported by promoting peace and mutual understanding through the stimulation of scientific cooperation between the USA and Europe. The Fulbright program is a prime example of this so-called North–North cooperation. During the Cold War, the focus of internationalization was on achieving foreign policy goals through aid and technical assistance. The process of decolonization in the 1950s and 1960s resulted in new forms of mobility and cooperation aimed at the development of a new intellectual stratum in the former colonial nations (South–North mobility and North–South cooperation). From the 1980s on, processes of regional integration notably in the European Union (EU), the Asia Pacific

Region, and Latin America (Mercosur) resulted in new types of multilateral exchange and cooperation such as the European Region Action Scheme for the Mobility of University Students (ERASMUS) and the University Mobility in Asia and the Pacific (UMAP) programs. After the terrorist attacks in 2001, internationalization became in the USA more clearly tied to diplomacy and national security, with new exchange programs with Muslim countries and heightened awareness of the need for language and area studies (Green and Knight, 2004).

In the context of these political and economical developments, internationalization of higher education became a way for governments to support the related processes of reconstruction, nation building, economic and democratic reform, technological development, regional integration, and international dialog. Public subsidies were made available to achieve these aims. Implementation was shaped through student and staff mobility, scientific cooperation, capacity building, knowledge transfer, the training of intellectual cohorts to international quality standards, and the promotion of mutual understanding and knowledge of different languages and cultures. Rationales for internationalization in higher education may thus include political, economic, cultural, and academic dimensions (Knight and de Wit, 1995; Blumenthal *et al.*, 1996). While from a government perspective, foreign policy and economic development have dominated the agendas, higher education institutions traditionally focused on internationalization as a way to improve the quality of teaching and research as well as to help institutions in other countries build capacity.

At the start of the twenty-first century, with globalization and the knowledge economy as major contextual factors, the various categories of activities and rationales can still be recognized in policy approaches to internationalization (OECD, 2004, 2006). Mutual understanding is an approach especially related to the social, cultural, and linguistic aspects of regional integration, international diplomacy, and more generally to the creation of international networks of elites and the enhancement of human capital. Student mobility programs are the main policy instrument in this approach and are still by and large publicly funded. Capacity building refers to the growing demand for higher education in an increasing range of developing countries and emerging economies, which can be met by study abroad (and return), but is increasingly paralleled by the establishment of foreign providers, that is, higher education institutions from

abroad that offer programs and services either on their own or in cooperation with domestic institutions. Since the new context is characterized by more and global competition in which knowledge is a prime factor for economic growth, access to higher education has become a key issue of national competitiveness and a more market-oriented model in which import and export categories can be distinguished emerged in the 1990s. In this context, internationalization may also be undertaken as a revenue-generating approach considering higher education as an export industry and using the revenues to finance the domestic higher education sector. Full-cost tuition fees and for-profit branches abroad play an important role in this approach and private sources of funding are often substantial. Finally, global competition urges national economies not only to educate their own population to the best, but especially in cases with national shortages in certain fields, also to attract highly skilled foreign people for the knowledge economy. Skilled migration is therefore another approach that may characterize the internationalization of higher education. Foreign graduate students and academic staff are attracted to enhance the competitiveness of higher education and the research and development sector in the host country. Internationalization policies of this kind need to be coordinated with immigration regulations, as visa and work permits are conditional.

Obviously, reality will usually reflect a combined blend of these various approaches, while focus and priorities differ across types of institution, countries, and regions, as discussed below. Overall it is clear, however, that economic rationales have become more dominant for internationalization over the last decade (Van der Wende, 2001a, 2001b; Van Vught *et al.*, 2002; Teichler, 2004). This is clearly linked to the process of globalization and its impact on higher education, notably the global competition for talented students and highly skilled workers as key resources for the knowledge economy and the emergence of an international market where demand for access to higher education is being met across borders. Responses from the higher education sector have in many countries been conditioned by national reforms that draw on the techniques of new public management (Marginson and Considine, 2000), including the modeling of national systems as economic markets and government-steered competition between institutions. In this spirit, governments apply steering models that grant institutions considerable autonomy and encourage them to become entrepreneurial also in the international higher education market. This may be complemented by active policies to lower barriers to such economically driven internationalization activities (i.e., international trade in higher education) through trade negotiations in educational services under the General Agreement on Trade in Services (GATS) of the World Trade Organization or other trade agreements.

Definitions and Related Conceptual Issues

Internationalization is understood in the literal sense as inter-national and refers to any relationship across borders between nations, or between single institutions situated within different national systems. It assumes that societies defined as nation-states continue to function as bounded economic, social, and cultural systems even when they become more interconnected as activities crossing their borders increase. In contrast, globalization puts emphasis on an increasing convergence and interdependence of economies and societies and a de-nationalization and integration of regulatory systems as a well as a blurring role of nation-states are taken for granted (Huisman and Van der Wende, 2004). The point that internationalization is predicated on a world order dominated by nation-states while globalization is more agnostic on this point (Scott, 1998) and the issue of whether national systems become more integrated as suggested by globalization, or more interconnected as with internationalization (Beerkens, 2004) can be seen as central distinctions between the two concepts. Furthermore, internationalization can involve as few as two units, whereas globalization takes in many nations and is a dynamic process drawing the local, national, and global dimensions more closely together (Marginson and Rhoades, 2002).

Globalization cannot be regarded simply as a higher form of internationalization. The relationship between the two concepts is not linear or cumulative but of a different order. Scott argues that the relationship is in fact a dialectical one in the sense that “not all universities are (particularly) international, but all universities are subject to the same process of globalisation – partly as objects, victims even, of these processes, but partly as subjects, or key agents of globalisation” (1998: 122). Globalization and internationalization in higher education may thus be potentially conflicting or rival, while at the same time interactive, mutually generative, and continually reinforcing each other. Globalization goes directly to the economic, cultural, and political core of nations, while also refashioning the larger higher education environment and thus creating a dynamic impact on higher education. In a networked environment in which every higher education institution is visible to every other, and the weight of the global dimension is increasing, it is no longer possible for nations or for individual institutions to seal themselves off from global effects. In this respect, internationalization can be seen as one possible response to globalization, that is, as a way to make higher education institutions more effective in response to the globalization of societies, cultures, economies, and labor markets (Van der Wende, 1997); as by definition, internationalization is a process more readily steerable by governments than is globalization. By the

same token, single governments have only a partial purchase on global developments through the medium of internationalization (Marginson and Van der Wende, forthcoming).

Internationalization: Main Elements and Strategies

Conceptually, internationalization was for a long time mainly seen as concentrating on the cross-border mobility of individual students and scholars and not as a strategy that affected higher education institutions or systems. This, however, changed over the last two decades. A broadening of the range of activities associated with internationalization has taken place – from an almost exclusive focus on individual mobility to more elaborate strategies encompassing curriculum development, research cooperation, staff development, and quality enhancement. Yet, international mobility remains a very important and visible element.

In 2003, 2.12 million students were enrolled in higher education outside their country of origin, of which 93% studied in the Organization for Economic Co-Operation and Development (OECD) area. Comparison of these numbers over the last 5 years reflects an annual increase of 8.3% on average. Five major countries of destination enrol the vast majority (70%) of these students: the USA (28%), the UK (12%) Germany (11%), France (10%), and Australia (9%). Substantial numbers of students are also attracted to Japan (4%) and the Russian Federation (3%). Malaysia is playing an increasing role in receiving students from China and India. On average, foreign students represent 6.4% of the total student population in OECD countries in 2003, up from 4.5% in 1998, indicating that in most countries, the foreign student population is growing faster than the overall domestic participation in higher education. A closer look at the numbers show that although the USA attracts the largest number of foreign students (close to 600 000), this represents less than 4% of its overall student population. A comparison with Europe indicates that the EU attracts a higher number of foreign students than the USA, but that more than half of the students studying in Europe are from within the region. Finally, comparisons between foreign students and domestic students studying abroad point to important differences between countries: these indexes may differ between 23.0 (Australia), 15.8 (USA), 8.9 (UK), and –18.3 (Korea) (OECD average = 2.4 students received for one student going abroad; data for 2001), indicating important imbalances between incoming and outgoing students in especially the main receiving countries (OECD, 2004, 2006).

A particular category of mobile students study abroad in an exchange program for a relatively short period (up to 1 year). The EU's ERASMUS program celebrated its

twentieth anniversary in 2007 with a total of around 1.5 million exchange students. The annual number of ERASMUS students has risen steadily and was over 144 000 in the year 2004–05. The largest EU member states are obviously the main sending and receiving countries and reciprocity is an important principle of the program. Nevertheless, also in this context, the UK demonstrates the greatest imbalance between incoming and outgoing students (2.25). ERASMUS students are found mostly in business studies (21%), languages (15%), engineering (11%), and social sciences (11%) (EC, 2006). Other examples of exchange programs are the Fulbright program, sponsored by the US Department of State. Since its establishment, more than 44 000 students from the US and 147 000 students from other countries have benefited from this program. The Fulbright US student program is now the largest US exchange program offering approximately 1200 grants annually in all fields of study (IIE, 2006).

High expectations were placed on the virtual mobility of students when the use of new information and communication technologies in higher education (notably the Internet and digital learning environments) increased in the 1990s. Electronic learning (e-learning) was expected to widen global access to higher education, to lead to pedagogic innovation, and decreased costs. Virtual universities were seen to become serious competitors for campus-based institutions, challenging the centrality of the face-to-face classroom setting. Expectations on the contribution of e-learning to internationalization still envisage opportunities for enhanced access, exchange, and collaboration. However, the challenges have also become clear. They are related to learner support and guidance systems that can effectively operate across national, cultural, and linguistic borders, to international systems for accreditation and for the recognition of qualifications, and to ensure affordable prices for international e-learning. (Helios, 2006). Market research shows that Australia's higher education export programs solely delivered through e-learning only represent a small proportion of total offerings and that virtual for-profit providers in the Asia-Pacific region reported more often losses than brick-and-mortar providers (OBHE, 2005). Further reflections on the experiences indicate that distance-education provision will lose ground in favor of prestigious providers – perhaps in consortium arrangements – with a commitment to a major brick-and-mortar presence, and a traditional academic focus on research and furnish students with a real (rather than virtual) campus experience (McBurnie, 2006: 66). In the US, only 3.2% of the students took all their courses online (in 2002–03) and they were mostly adult learners in professional areas. Generally, students prefer blended models of learning (Douglas, 2005). So far, e-learning has generally failed to emerge as a significant global market. The complex possibilities of international e-learning are typically left to small-scale,

department-led experiments (OECD, 2005). They show that organizational conditions, such as different academic calendars and time zones, student and staff competences in areas as intercultural communication and foreign languages, and technological challenges and issues of ownership related to the coupling of institutional learning environments, cannot be taken for granted.

Data on the mobility of academic staff show that the mobility of teachers in the ERASMUS program rose from less than 8000 in 1997–98 to almost 20 000 in 2004–05 (EC, 2006). Yet, there are no clear signs of a process of Europeanization in academic recruitment and careers, as each national academic labor market is still characterized by salaries, status, recruitment procedures, workloads, career patterns, and promotion rules being very different from one country to another (Musselin, 2004). But while in Europe, postdoctoral mobility is stable (Enders and de Weert, 2004), in the USA, a high and increasing proportion of postdoctoral personnel holding a US doctoral degree are foreign born: 41% in 2001 compared to 21% in 1985 (NSB, 2006). The USA followed by the UK also draws the largest number of visiting faculty, at least 20% coming from Europe. In 2005, US higher education institutions hosted 89 600 foreign scholars to conduct teaching or research activities. Two-thirds were in science and engineering. In most OECD countries, two to four scholars and researchers hold positions in the USA for every 100 at home. In 2003–04, the ratio of visiting scholars to those at home was highest for Korea (13 per 100), the Russian Federation (8), and China (6). About half the foreign doctoral graduates stay in the USA after graduation, many in faculty positions, augmenting the capacity of the United States as a global knowledge economy (OECD, 2006). Obviously, these flows create serious concerns regarding brain drain in other, especially non-OECD, countries.

In the area of academic research, international collaboration is clearly on the rise. The growth of internationally co-authored scientific articles (with at least one international co-author) went from 8% in 1988 to 18% of all scientific articles in 2001. In the USA, they represent 23% and in Europe as much as 33% of all articles, although in Europe the collaboration has a strong intra-regional character. Foreign scientific articles are increasingly cited, representing 55% of all citations in 1992 compared to 62% in 2001. The average number of collaborating countries in scientific activities increased from 89 in 1994 to 102 in 2001 (Vincent-Lancrin, 2006).

Student mobility and the international collaboration in research have been two important driving forces in the development of institutional strategies for internationalization. Such strategies typically include provisions for mobility and exchange, including the related support structures, for example, international offices and cooperation agreements with foreign partner institutions.

In the area of curriculum development, many institutions undertake activities to ensure that the nonmobile students also benefit from an international dimension in their training. Internationalized curricula can be described as curricula with an international orientation in content, aimed at preparing students for performing (professionally/socially) in an international and multicultural context, and designed for domestic students and/or foreign students (Van der Wende, 1996). Initiatives in this area include joint degree programs (offered by two or more institutions in different countries), programs that involve substantial foreign-language learning and/or are taught in English as a second language, and that involve an international classroom, which is seen as a key to acquiring intercultural skills and competences (Hudson and Todd, 2000).

The focus of institutional strategies for internationalization is linked to the institution's aims in this area (Huisman and Van der Wende, 2005). Institutions that seek to become recognized global players understand internationalization as being related to worldwide competition among elite universities (encouraged by global university rankings) for the recruitment of bright, talented students, young researchers, and renowned teaching staff. Their student recruitment strategy is very explicit and highly selective. Other institutions undertake internationalization in the context of cooperation and networking for mutual benefit and want to strengthen their regional profile. Their strategies are focused on collaborative research, exchange of students and staff, joint study programs, etc. In other cases, internationalization may be necessary for survival, for instance, related to decreasing domestic demand, and institutional income generation. Recruitment of (full) fee-paying students and the establishment of overseas programs and branches are used to this end. Finally, institutions may see internationalization as a way to enhance their local or national status. Especially certain nonuniversity or teaching-oriented type of institutions employ international activities in order to offer, for instance, degree programs at higher levels, to develop research activities, or even to achieve university status. Institutional aims and strategies for internationalization are closely connected to institutional profiles. Research-intensive universities are clearly driven by international collaboration and competition in research, and therefore focus more on attracting talented graduate students, whereas teaching-oriented institutions usually focus more on undergraduate students. Specialized institutions, such as business schools or arts schools, are real niche players that attract very specific international clientele.

Internationalization strategies at institutional level reflect the wider set of policy approaches as discussed before. Clearly, global competition and the knowledge economy have advanced, especially the revenue-generating and skilled-migration approaches. This coincides with

what is seen by many as a shift from cooperation to competition as the main paradigm for internationalization. Especially in continental Europe, perceptions indicate that cooperation is seen to be the inherent model of internationalization, associated with academic exchange, quality and excellence, and cultural notions of mutuality and intercultural learning. More generally, the conception of the university as a competitive (global) enterprise is countered with the public service model and the notion of higher education as a public good (Olson 2005; Van Vught *et al.*, 2002). However, both cooperation and competition are important strategic options for internationalization and in many cases, internationalization strategies are indeed composed by a combination of cooperative and competitive options. Examples of cooperation for competition can be found in global university alliances (consortia). The complexity of such strategies, however, is demonstrated by the fact that despite the high expectations of the potential role of these consortia in the global higher education scene, whether they are successful or not seems to be largely defined by the extent to which the institutions concerned are embedded in their national systems (Beerkens, 2004). More generally, it can be said that: “Despite all the research demonstrating the growing importance of internationalisation, and even more the rhetoric in this respect, higher education institutions’ behavior (including their internationalisation strategies) are (still) mostly guided by national regulatory and funding frameworks. For internationalisation in particular, historical, geographic, cultural and linguistic aspects of the national framework are of great importance” (Luijten-Lub *et al.*, 2005: 238).

System-Level Changes, Concerns, and Perspectives

The international mobility of students and scholars requires systems for the recognition of diplomas and qualifications, in terms of academic recognition (a decision that allows a person to pursue or continue studies, or to use an academic title or degree) and professional recognition (a decision to grant professional rights, license to practice, or status to a graduate). Recognition methodology developed over the second-half of the twentieth century from an approach aimed to establish equivalence to one based on acceptance (of the nearest comparable degree) as expressed in the Council of Europe/UNESCO Recognition Convention of Lisbon (1997). The latter approach requires mutual trust in each other’s education system. The related need for more transparency was one of the main rationales for the Bologna Declaration, which was signed by 29 European countries in 1999. With the aim to promote employability within the EU and to enhance the international attractiveness of European higher education, a process of convergence toward a

two-cycle (undergraduate–graduate) degree system was undertaken (Reichert and Tauch, 2005; Alesi *et al.*, 2005; Witte, 2006). The process now includes 45 countries and is increasingly influential on internationalization agendas also beyond Europe (Australian Government, 2006). Besides a framework for recognition, the Bologna Process also aims to set international standards for quality assurance.

In the context of internationalization, quality is an issue of concern as much as this context offers opportunities for improvement. On the one hand, learners should be protected from low-quality provision and qualifications. To this end, guidelines on quality assurance of cross-border education were developed (UNESCO/OECD, 2005). On the other hand, international benchmarking, perhaps ranking (provided that methodological principles are seriously respected, see IREG, 2006), and especially the comparison of actual learning outcomes can offer higher education useful insights into the quality and effectiveness of systems and institutions. The further process of internationalization is expected to offer many new and challenging opportunities in this respect and will be beneficial if governments and institutions manage to combine competitiveness with quality improvement and enhancement of opportunity for all.

See also: Associations of Universities; Flexible Learning in Higher Education; Higher Education and the Knowledge Society; Higher Education and the Labor Market; Higher Education Crossing Borders; Mobility of PhD Students and Scientists; National and International Rankings of Higher Education; Student and Faculty Transnational Mobility in Higher Education; The Bologna Process in European Higher Education.

Bibliography

- Alesi, B., Burger, S., Kehm, B., and Teichler, U. (2005). *Bachelor and Master Courses in Selected Countries Compared with Germany*. Bonn: Federal Ministry of Education and Research.
- Australian Government. (2006). *The Bologna Process and Australia: Next Steps*. Canberra: Department of Education, Science and Training.
- Beerkens, H. J. J. G. (ed.) (2004). *Global Opportunities and Institutional Embeddedness: Higher Education Consortia in Europe and Southeast Asia*. Enschede: Center for Higher Education Policy Studies, University of Twente.
- Blumenthal, P., Goodwin, C. D., Smith, A., and Teichler, U. (eds.) (1996). *Academic Mobility in a Changing World*. London: Jessica Kingsley.
- Council of Europe (1997). *Convention on the Recognition of Qualifications concerning Higher Education in the European Region*.
- De Wit, J. W. M. (2002). *Internationalization of Higher Education in the United States of America and Europe. A Historical, Comparative, and Conceptual Analysis*. Westport, CT: Greenwood Press.
- Douglas, J. A. (2005). *All Globalization is Local: Countervailing Forces and the Influence on Higher Education Markets*. Berkeley, CA: University of California.
- Enders, J. and de Weert, E. (2004). Science, training and career: Changing modes of knowledge production and labour markets. *Higher Education Policy* 17, 135–152.

- European Commission (EC) (2006). *Erasmus – Statistics. Erasmus Student and Teacher Mobility 2004–2005*. http://ec.europa.eu/education/programmes/socrates/erasmus/stat_en.html (accessed Jun. 2009).
- Green, M. and Knight, J. (2004). Cross-border post-secondary education in North America. In OECD (ed.) *Internationalization and Trade in Higher Education: Opportunities and Challenges*, pp 39–87. Paris: OECD.
- Helios (2006). *Is E-Learning Contributing to Internationalisation of Education and Training?* (draft). <http://www.education-observatories.net/helios> (accessed Jun. 2009).
- Hudson, B. and Todd, M. (eds.) (2000). *Internationalizing the Curriculum in Higher Education*. UK: Sheffield Hallam University Press.
- Huisman, J. and van der Wende, M. C. (eds.) (2004). On cooperation and competition. National and European policies for internationalisation of higher education. *ACA Papers on International Cooperation*. Bonn: Lemmens.
- Huisman, J. and van der Wende, M. C. (eds.) (2005). On cooperation and competition II. institutional responses to internationalisation, Europeanisation and globalisation. *ACA Papers on International Cooperation*. Bonn: Lemmens.
- Institute of International Education (IIE) (2006). *Fulbright Program*. <http://www.iie.org> (accessed Jun. 2009).
- IREG International Ranking Expert Group (2006). *Berlin Principles on Ranking of Higher Education Institutions. Internationalisation of Higher Education*, pp 35–91. Paris: OECD.
- Knight, J. and de Wit, H. (1995). Strategies for internationalisation of higher education: historical and conceptual perspectives. In de Wit, H. (ed.) *Strategies for Internationalisation of Higher Education. A Comparative Study of Australia, Canada, Europe and the United States of America*, pp 5–32. Amsterdam: EAIE.
- Luijten-Lub, A., Huisman, J., and van der Wende, M. C. (2005). Conclusions, reflections and recommendations. In Huisman, J. and van der Wende, M. C. (eds.) *On cooperation and competition II. Institutional responses to internationalisation, Europeanisation and globalisation. ACA Papers on International Cooperation*. Bonn: Lemmens.
- Marginson, S. and Considine, M. (2000). *The Enterprise University: Power, Governance and Reinvention in Australia*. Cambridge: Cambridge University Press.
- Marginson, S. and Rhoades, G. (2002). Beyond national states, markets, and systems of higher education: A glonacal agency heuristic. *Higher Education* **43**, 281–309.
- Marginson, S. and van der Wende, M. C. (2009). *Globalization and Higher Education*. Paris: OECD.
- McBurnie, G. (2006). 'Leveraged footprints', 'dark clouds' and 'bleeding millions': Perspectives on Australian universities' offshore campuses. *Perspectives in Education* **24**(4), 57–68.
- Musselin, C. (2004). Towards a European academic labour market? Some lessons drawn from empirical studies on academic mobility. *Higher Education* **48**, 55–78.
- National Science Board (NSB) (2006). *Science and Engineering Indicators 2004*. <http://www.nsf.gov/statistics/seind04> (accessed Jun. 2009).
- Neave, G. (2001). The European dimension in higher education: An excursion into the modern use of historical analogues. In Huisman, J., Maassen, P., and Neave, G. (eds.) *Higher Education and the Nation State. The International Dimension of Higher Education*, pp 13–73. Oxford: Pergamon Press.
- OBHE (Observatory on Borderless Higher Education) (2005). *Mapping Borderless Higher Education: Policy, Markets and Competition*. London: OBHE.
- OECD (2004). *Internationalization and Trade in Higher Education: Opportunities and Challenges*. Paris: OECD.
- OECD (2005). *E-learning in Tertiary Education*. Paris: OECD.
- OECD (2006). *Education Policy Analysis. Focus on Higher Education*. Paris: OECD.
- Olson, J. P. (2005). The institutional dynamics of the (European) university. *Arena Working Paper*. <http://www.arena.uio.no> (accessed Jun. 2009).
- Reichert, S. and Tauch, C. (2005). *Trends IV: European Universities Implementing Bologna*. Brussels: EUA.
- Scott, P. (ed.) (1998). *The Globalization of Higher Education*. Buckingham/Philadelphia, PA: Society for Research into Higher Education/Open University Press.
- Scott, P. (1998). *The Globalization of Higher Education*. Buckingham: The Society for Research into Higher Education and Open University Press.
- Teichler, U. (2004). The changing debate on internationalization of higher education. *Higher Education* **48**, 5–26.
- UNESCO/OECD (2005). *Guidelines on Quality Provision in Cross-Border Higher Education*. Paris: OECD.
- Van der Wende, M. C. (1996). Internationalizing the curriculum in higher education. *Tertiary Education and Management* **2**(2), 186–195.
- Van der Wende, M. C. (1997). Missing links. In Kälvermark, T. and van der Wende, M. C. (eds.) *National Policies for Internationalisation of Higher Education in Europe*, pp 10–42. Stockholm: National Agency for Higher Education.
- Van der Wende, M. C. (2001a). Internationalisation policies: About new trends and contrasting paradigms. *Higher Education Policy* **14**(3), 249–259.
- Van der Wende, M. C. (2001b). The international dimension in national higher education policies: What has changed in Europe in the last five years? *European Journal of Education* **36**(4), 431–441.
- Van Vught, F. A., van der Wende, M. C., and Westerheijden, D. F. (2002). Globalisation and internationalisation. Policy agendas compared. In Enders, J. and Fulton, O. (eds.) *Higher Education in a Globalizing World. International Trends and Mutual Observations*, pp 103–121. Dordrecht: Kluwer.
- Vincent-Lancrin, S. (2006). What is changing in academic research? Trends and futures scenarios. *European Journal of Education* **41**(2), 169–202.
- Witte, J. (2006). *Change of Degrees and Degrees of Change. Comparing Adaptations of European Higher Education Systems in the Context of the Bologna Process*. Enschede: CHEPS, University of Twente.
- Hudson, B. and Todd, M. (eds.) (2000). *Internationalizing the Curriculum in Higher Education*. Sheffield: Sheffield Hallam University Press.

Further Reading

National and International Rankings of Higher Education

S Marginson, University of Melbourne, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Institutional classifications – Listing of higher education institutions in groups on the basis of nominated characteristics such as mission, size, and/or level of study.

University ranking – Ordinal listing of individual higher education institutions, or groups of institutions, based on performance indicators and/or surveys of reputation.

Introduction

Ordinal rankings have become installed as a prominent feature of higher education in many countries and also on the global scale. With one important exception discussed below (CHE, 2006), rankings take the form of vertically ordered league-tables. They operate both at the level of institutions, and also at the level of disciplines and specializations, such as business, law or medicine schools, or aspects of institutional mission. Rankings systems draw on a variety of methods to derive rank order, including quantitative metrics and absolute and comparative judgments expressed in surveys and by expert panels. Rankings are often conducted by publishing companies that draw on the techniques of annually updated consumer guides (Usher and Savino, 2006).

Although higher education rankings have roots as far back as the nineteenth century in the United States, their modern origin was the annual survey and ranking of higher education institutions by the US News and World report (USNWR, 2006) from 1983. Most rankings systems have been established after 2000 (Salmi and Saroyan, 2006). The trend to rankings can be understood as part of the larger evolution of techniques of evaluation and accountability; the focuses on competition between institutions, outcomes, quality, consumer choice making, and institutional management in higher education; and the growing public discussion about the sector. In turn, the rankings process contributes to these secular tendencies, especially by fostering intranational and global competition between institutions. National rankings function as an often potent shorthand accountability mechanism for institutional leaders. Global rankings serve a similar function and, by extending scrutiny to the performance of

systems, they extend the accountability pressure to national governments.

Notwithstanding the public prominence of rankings systems, and the apparent authority of their findings, these systems use varied, sometimes eclectic, and often contested methods to reach those findings. On the whole, research and publications performance, with established metrics recognized on the global scale, has proven to be the aspect of higher education most amenable to rankings. There is much debate about techniques used for combining different measures or judgments about quality into single indices for rank-ordering purposes. There is less debate about rank ordering *per se*. Rankings data are widely publicized and utilized, more than any other data on higher education; however, media organizations and public authorities are more comfortable with rankings than are higher education institutions themselves.

National Rankings

Most countries with large higher education systems have rankings of one kind or another. Countries with rankings devised by newspapers and magazines include China and Hong Kong China, Japan, India, the Ukraine, Romania, Poland, Portugal, Italy, Spain, Germany, Sweden, Switzerland, France, UK, USA, and Canada. In Thailand, Malaysia, Pakistan, India, Kazakhstan, Korea, Tunisia, Nigeria, the Netherlands, the UK, Brazil, and Argentina, rankings have been instigated by ministries of education, grants councils, or accreditation agencies. In China, Japan, Australia, Kazakhstan, Slovakia, Romania, Russia, the Ukraine, Germany, Spain, Switzerland, the UK, and Canada, rankings have been initiated by universities, professional associations, or other organizations (Salmi and Saroyan, 2006). Usher and Savino (2006) review rankings compiled in Australia (Williams and Van Dyke, 2006); Canada; China, with several rankings systems (Liu and Liu, 2005), and Hong Kong; Germany; Italy; Poland; Spain; the UK, where four different newspapers have developed rankings; and the United States. For the most part, national rankings consist of a single table; however, in the USA and Canada, higher education institutions have been divided into groups according to mission and other characteristics, creating a set of mini league tables within which the category of comprehensive research universities has highest status.

Specialist rankings focus on characteristics ranging from research output, to student services, to MBA programs, to Yahoo magazine's ratings of connectivity, the contribution to social diversity, and other features. These rankings are not further discussed here but it can be noted that ratings of disciplines and other aspects feed into the larger schemes of Shanghai Jiao Tong University (SJTU) and the German Centre for Higher Education Development (henceforth, CHE).

USNWR Rankings

In the United States, the annual USNWR survey focuses on aspects of institutions seen to contribute to the quality of teaching and the student experience, with research and scholarship playing a minor role. The categories of institutions are drawn from the classification in 2000 by the Carnegie Foundation for the Advancement of Teaching (USNWR, 2006: 81ff). The most important grouping is national universities – 248 universities (162 public and 86 private) with a wide range of fields of study and offering degrees to doctoral level. Most of them are research intensive. Other categories are liberal arts colleges, master's-level universities, comprehensive bachelor-level universities, business programs, undergraduate engineering programs, and specialty schools mostly in the fine and performing arts. In national universities and liberal arts colleges, the largest part of the index (25%) is comprised by a survey of undergraduate academic excellence sent to university presidents, provosts, and deans of admissions. Two items constitute 20% of the index: student retention and graduation rate; and faculty resources which reward small classes, high academic salaries, high academic qualifications, and a high proportion of full-time staff. Student selectivity at entry, a proxy for market demand, is worth 15%. The lesser items are: spending per student (10%), the proportion of alumni who donate back to the institution, and the graduation rate after controlling for spending and student aptitude (each 5%) (USNWR, 2006: 77–79).

The annual US News survey shapes institutional prestige and influences the flows of high-scoring students, academic labor, and resources. Individual institutions usually make strenuous efforts to improve their market standing, focusing on measures designed to enhance reputation and data management of those indicators open to discretionary action. Some employ highly remunerated enrolment managers to manipulate student entry so as to maximize student scores and the rate of refusal of applications; for example, limiting the size of the initial intake, taking in some students through nonselective means, and offering incentives to high-scoring students. These trends have been criticized for enhancing merit-based student aid at the expense of needs-based aid and fostering social elitism and institutional interest at the expense of the public interest (Kirp, 2004). Comparative retention and

graduation rates can be enhanced by dropping academic standards or even by poaching students from rival institutions. Other indicators open to gaming and variable reporting techniques include alumni donation rates as well as staffing and spending levels.

Despite criticisms of the US News survey, its functioning seems to be accepted. The parallel Canadian survey by MacLeans has achieved lesser consensus. In 1994 and 1995, there were large-scale institutional boycotts; subsequently, there were selective boycotts of data provision in specific areas, and in 2006 11 universities withdrew *in toto*. MacLeans stated that it would use freedom-of-access laws to secure the data necessary to maintain its rankings (Salmi and Saroyan, 2006: 12–13).

The CHE rankings

The CHE, located in Gutersloh in the state of north Rhine-Westphalia in Germany, in conjunction with the German Academic Exchange Service, which assists international students, and the publisher *Die Zeit*, has developed a distinctive approach to national rankings. This dispenses with holistic (summative) rank ordering of institutions in league tables. CHE surveys 130 000 students and 16 000 academic faculty in almost 250 higher education institutions, focusing on student experiences, student satisfaction, and academic recommendations on the best places in each field of study. It supplements the surveys with independent sources comprising one-third of the total database. No data are taken from institutions. CHE ranks institutional departments according to each separate indicator of academic and service quality, assigning them to the top third, middle third, or the bottom third of all institutions. It provides neither discipline-based league tables nor institutional league tables. It refuses to integrate the different indicators into a single indicator for each institution because there is no one best university across all areas, and minimal differences produced by random fluctuations may be misinterpreted as real differences in holistic rankings systems (CHE, 2006).

The CHE data are made available to prospective students and the public free of charge though an interactive web-enabled database. Any person can interrogate this database by investigating and rank identified disciplines and administrative services, using their own combination of criteria (CHE, 2006), thereby creating weightings and rankings themselves. CHE acknowledges that the definition of quality is purpose driven and open to variation and passes power over the definition from the ranker to the consumer. It also provides by far the most comprehensive set of comparative data devised so far. A difficulty is that the CHE approach requires sufficient homogeneity between institutions and their programs and services to enable the comparisons. Arguably, it would be necessary to develop several sets of institutions for CHE-style

comparisons in more varied environments such as that of the United States.

This process of data collection has been extended from all higher education institutions in Germany to Switzerland and Austria. The Netherlands and Belgium (Flanders) are preparing to join the system, and some Nordic countries may follow. The CHE ranking system is well positioned to develop into a European-wide system, and has attracted attention in other parts of the world as well (Usher and Savino, 2006; see also Van Dyke, 2005), including Canada and the USA. The Commission on the Future of Higher Education in the US is working on a comparable concept that would allow consumers to rank colleges based on variables of their choice (Field, 2006).

Global University Rankings

Increasingly, national policymakers and institutional leaders respond to a global higher education environment in which cross-border linkages and flows of knowledge and people are essential to strategy, and international comparisons are constantly being made. The advent of global university rankings suggests that higher education is now constituted by an open worldwide competition between nations, and between individual institutions as global actors in their own right. The most influential global rankings, prepared annually by the Shanghai Jiao

Tong University Institute of Higher Education (SJTUIHE), began in 2003. The second main global rankings, prepared by The Times Higher Education Supplement (THES), commenced in 2004.

Table 1 sets out the top 20 universities in each ranking. Although these ranking systems use quite different measures of university quality, both confirm the reputations of the leading American and British universities, including Harvard, Stanford, Yale, Berkeley, Massachusetts Institute of Technology (MIT), Caltech, Princeton, Cambridge, and Oxford.

Global comparisons are possible only in relation to one kind of institution, the comprehensive research-intensive university. This is the only model sufficiently widespread and homogeneous to lend itself to the formation of a single competition. Research and doctoral education are the most globalized of all dimensions of higher education within the medium of a singular English-language system of publication. This is not to say that global comparisons of research performance, and the exclusion of knowledge produced in languages other than English, are unproblematic.

SJTU Rankings

In compiling an academic ranking of world universities, the SJTUIHE focused exclusively on broadly available and internationally comparable data of measurable research performance (Liu and Cheng, 2005: 133). It was considered impossible to compare teaching and learning

Table 1 The world's leading universities as measured by the Shanghai Jiao Tong University, 2006, and The Times Higher Education Supplement, 2006

Shanghai Jiao Tong research university rankings				The Times Higher university rankings			
	University	Points	Nation		University	Points	Nation
1	Harvard U	100.0	USA	1	Harvard U	100.0	USA
2	U Cambridge	72.6	UK	2	U Cambridge	96.8	UK
3	Stanford U	72.5	USA	3	U Oxford	92.7	UK
4	U California, Berkeley	72.1	USA	=4	Massachusetts IT	89.2	USA
5	Massachusetts IT	69.7	USA	=4	Yale U	89.2	USA
6	California IT (Caltech)	66.0	USA	6	Stanford U	85.4	USA
7	Columbia U	61.8	USA	7	California IT ('Caltech')	83.8	USA
8	Princeton U	58.6	USA	8	U California, Berkeley	80.4	USA
8	U Chicago	58.6	USA	9	Imperial College London	78.6	UK
10	U Oxford	57.6	UK	10	Princeton U	74.2	USA
11	Yale U	55.9	USA	11	U Chicago	69.8	USA
12	Cornell U	54.1	USA	12	Columbia U	69.0	USA
13	U California, San Diego	50.5	USA	13	Duke U	68.3	USA
14	U California, Los Angeles	50.4	USA	14	Beijing U	67.9	China
15	U Pennsylvania	50.1	USA	15	Cornell U	65.9	USA
16	U Wisconsin-Madison	48.8	USA	16	Australian National U	64.8	Australia
17	U Washington (Seattle)	48.5	USA	17	London S. of Economics	63.9	UK
18	U Calif., San Francisco	47.7	USA	18	E. Normale Supérieure Paris	63.3	France
19	Tokyo U	46.7	Japan	=19	National U of Singapore	63.1	Singapore
20	Johns Hopkins U	46.6	USA	=19	Tokyo U	63.1	Japan

U, university; IT, institute of technology.
From SJTUIHE (2006) and THES (2006).

on a worldwide basis “owing to the huge differences between universities and the large variety of countries, and because of the technical difficulties inherent in obtaining internationally comparable data.” The SJTUIHE also decided not to use subjective measures of opinion or data source from institutions. It has consulted widely and its annual data are increasingly robust.

The main part of the SJTU index is determined by publication and citation performance in the sciences, social sciences, and humanities: 20% by citation in leading journals; 20% by articles in *Science* and *Nature*; and 20% by the number of HiCi researchers in the institution. HiCi researchers are those named in the Thomson/Institute for Scientific Information (ISI) classification of the leading 250–300 researchers by field of study, nearly all of which are science-based disciplines, on the basis of citation performance (ISI, 2006). Another 30% is determined by the location of the winners of Nobel prizes and Fields medals in mathematics, during their training (10%) and in their current employment (20%). The remaining 10% is determined by dividing the total derived from the above data by the number of academic staff.

Factors determining performance in the Jiao Tong rankings

The SJTUIHE rankings favor universities that are large and comprehensive enough to amass research performance over a broad range of research fields, especially those strong in the hard sciences, and with relatively few research-inactive staff. They also favor English-language universities because non-English-language work falls outside the main journals and is cited less within them; and they favor universities from the United States because Americans tend to cite Americans (Altbach, 2006). Of the HiCi researchers, 3614 are located in the USA compared to 224 in Germany, 221 in Japan, 162 in Canada, 138 in France, 101 in Australia, 94 in Switzerland, 55 in Sweden, 20 in China, and none in Indonesia (ISI, 2006). Among the US universities, Harvard and its affiliated institutes alone have 168 HiCi researchers; Stanford has 132, UC Berkeley 82, and MIT 74. There are 42 at the University of Cambridge in the UK.

One criticism made of the SJTU rankings is that Thomson, the publishing company that comprises the Thomson/ISI database and determines which journals should receive an ISI rating, has a vested interest in selecting Thomson-published journals ahead of other journals. The Nobel prize criterion is perhaps the most controversial, as the prize is submission based and, at times, an element of politicking appears to enter the decisions; scientific merit is not the only factor.

According to the SJTUIHE, 54 of the world's leading research universities are located in the United States. The UK provides 11 of the top 100. When Canada (four) and Australia (two) are added, the English-speaking nations

between them constitute 71% of the top 100 group. A further 22 are located in Western Europe, six in Japan, and one in each of Israel and Russia. The principal Western European nations are Germany (five), France and Sweden (four each), Switzerland, (three) and the Netherlands (two). China and India have none of the top 100 research universities. India has just two in the top 500. China, including Hong Kong, has 19 of the top 500.

As **Table 2** indicates, there is a broad overall correlation between a nation's gross national income per head and its proportion of universities in the SJTUIHE top 100 and top 500 when the size of the national economy is taken into account. When SJTUIHE research performance is compared to overall economic size and wealth, Israel, Sweden, Switzerland, UK, the Netherlands, Canada, Finland, Denmark, and Australia appear to perform better than expected. The USA performs very well in its share of the world's top 100 but less well in the top 500, indicating its internal stratification. Germany does well at the level of the top 500, indicating a broad-based research capacity across the national system, but less well in its share of the top 100. Japan underperforms relative to its economic capacity on both measures.

There is a closer correlation between the strength of public research systems and SJTUIHE research performance. Aside from the USA, in nations that over-perform relative to economic capacity, the independent private sector plays a relatively minor role. Correspondingly, several nations that underperform at the top 100 level have large private sectors and a highly stratified research effort, including Japan, Korea, Poland, Brazil, and Mexico. Moreover, even though it is obvious that these are all non-English-speaking nations, some non-English-speaking nations in Europe have developed strong basic research systems. These findings underline the dependence of basic research capacity on public investment, given the public good character of research (Stiglitz, 1999).

Times Higher Rankings

The THES sets out to provide the best guide to the world's top universities (THES, 2005). Altogether, 40% of the index is comprised by an international opinion survey of academics and 10% by a survey of global employers. There are two internationalization indicators: the proportion of students that are international (5%) and the proportion of staff (5%). Another 20% is determined by the student-staff ratio, treated as proxy for teaching quality. The remaining 20% of the Times index is comprised by research citations per staff member using the Thomson/ ISI database.

The reputational surveys and the student internationalization indicator appear to boost the standing of universities active in the global market in cross-border students. The THES rankings boost the number and

Table 2 Nations' share of the top 500 and 100 research universities as measured by Shanghai Jiao Tong University Institute of Higher Education, 2006, compared to total gross national product, population, and gross national income per head, 2005

<i>Nation</i>	<i>Gross national product 2005 (\$b USD PPP)</i>	<i>Population 2005 (millions)</i>	<i>Gross national income per head 2005 (\$USD PPP)</i>	<i>Share of top 500 research universities 2006 (%)</i>	<i>Share of top 100 research universities 2006 (%)</i>
United States	12 410	296.5	41 950	33.4	54.0
United Kingdom	1927	60.2	32 690	8.6	11.0
Japan	3944	128.0	31 410	6.4	6.0
Germany	2418	82.5	29 210	8.0	5.0
Canada	1061	32.3	32 220	4.4	4.0
France	1830	60.7	30 540	4.2	4.0
Sweden	280	9.0	31 420	2.2	4.0
Switzerland	256	7.4	37 080	1.6	3.0
Australia	643	20.3	30 610	3.2	2.0
Netherlands	538	16.3	32 480	2.4	2.0
Italy	1668	57.5	28 840	4.6	1.0
Israel	177	6.9	25 280	1.4	1.0
Finland	164	5.2	31 170	1.0	1.0
Denmark	183	5.4	33 570	1.0	1.0
Norway	186	4.6	40 420	0.8	1.0
Russian Federation	1560	143.1	10 640	0.4	1.0
China ^a	8573	1305.5	6600	1.8	0.0
Korea	1056	48.3	21 850	1.8	0.0
Spain	1134	43.4	25 820	1.8	0.0
Belgium	337	10.5	32 640	1.4	0.0
Austria	276	8.2	33 140	1.4	0.0
New Zealand	93	4.1	23 030	1.0	0.0
China Hong Kong	214	6.9	34 670	1.0	0.0
Taiwan	n.a.	n.a.	n.a.	1.0	0.0
Brazil	1627	186.4	8230	0.8	0.0
South Africa	558	45.8	12 120	0.8	0.0
Ireland	170	4.2	34 720	0.6	0.0
India	3816	1094.6	3460	0.4	0.0
Poland	534	38.2	13 490	0.4	0.0
Hungary	182	10.1	16 940	0.4	0.0
Greece	262	11.1	23 620	0.4	0.0
Singapore	130	4.4	29 780	0.4	0.0
Egypt	330	74.0	4400	0.2	0.0
Mexico	1052	103.1	10 030	0.2	0.0
Chile	206	16.3	11 470	0.2	0.0
Argentina	559	38.7	13 920	0.2	0.0
World total ^b	61,007	6437.8	9420	100.0	100.0

^aExcludes China Hong Kong and China Taiwan, listed separately.^bIncludes nations without universities in top 500.

From World Bank (2006) and SJTUIHE (2006)

placement of British universities in the top 20, while reducing the number of US universities in the top 100 from the 54 in the SJTUIHE rankings to 32. The number of Australian universities in the top 100 rises from two in the SJTUIHE rankings to seven in the THES rankings.

The THES rankings are open to a number of criticisms. The surveys are nontransparent: it is not specified who has been surveyed or what questions were asked. The student internationalization indicator rewards volume building rather than the quality of student demand or the quality of programs. Teaching quality cannot be adequately assessed

using a resource quantity indicator such as the student–staff ratio. The THES rewards an institution's marketing division better than its researchers. Further, the rankings can be highly volatile, raising questions about whether there is a positive incentive-based relationship between performance and ranking. To cite some examples among many: between 2004 and 2006 the University of Malaya oscillated between rankings 89 and 192; Fudan in China between 72 and 195; Seoul National between 63 and 118; and the Royal Melbourne Institute of Technology (RMIT) in Australia between 55 and 146. In the USA, Emory has risen from

173 to 56, and Purdue has fallen from 59 to 127. It is highly unlikely that these fluctuations reflect actual changes in performance. They are more likely to reflect changes internal to the THES instruments and methodologies.

Newsweek Rankings

Newsweek (2006) prepared a composite system of rankings that combined the research publication and citation measures from SJTIHE with the measures of internationalization, student staff ratios, and citations per head from the Times, adding also a metric for library holdings; thus removing the least plausible aspects of both the Times index (the reputational surveys) and the SJTUIHE index (the Nobel indicators). Newsweek's leading universities were Harvard, Stanford, Yale, Caltech, and Berkeley. Cambridge, UK was at six, MIT at seven, Oxford at eight, Tokyo 14, Toronto in Canada at 15, and the Federal Institute of Technology in Zurich at 18. There were five universities from Switzerland in the first 41, the same number as the UK.

Issues of Method and Interpretation

Comparing ten rankings systems Van Dyke (2005) concludes that although the rankings share broad principles and approaches, they differ considerably in aims, systems, cultures, as well as in availability and reliability of data. Usher and Savino (2006: 12–26) examine 19 league tables and ranking systems. The number of raw indicators in use varies from 1 to 71. Indicators are selected from five broad categories:

1. the starting characteristics of students;
2. learning inputs (financial and other resources, staff, and the learning environment);
3. learning outputs (student retention and completion, skills and attributes of graduates, employment outcomes, and citizenship attributes);
4. research (resources invested in research, doctoral programs, publications, citations, and awards); and
5. indicators of institutional reputation.

Half of all rankings systems use surveys to collect data, the other half use data provided by universities, and nearly all use data collected from other sources. The different rankings systems are driven by different purposes and associated with different notions of what constitutes university quality, a highly contested notion (Usher and Savino, 2006: 9). Some look at outputs, others focus on inputs. "Among both inputs and outputs, there is very little agreement as to what kind of inputs and outputs are important"; "rankings have an element of capriciousness to them: with a large enough sample, choosing a different set of indicators does indeed create a different

set of ordinal rankings" (pp. 29–31). Rankings systems that are below the very top produce widely divergent results (Salmi and Saroyan, 2006: 19).

Holistic Reputational Rankings

Usher and Savino also remark on the arbitrary character of the weightings used to construct composite indexes uniting the different aspects of quality or performance. Weighting systems are rarely grounded theoretically (Salmi and Saroyan, 2006: 9) and many institutional practices are omitted. As Rocki (2005: 180) concludes, "the variety of methodologies, and thus of criteria used, suggest that any single, objective ranking could not exist". But "the fact that there may be other legitimate indicators or combinations of indicators is usually passed over in silence. To the reader, the author's judgment is in effect final" (Usher and Savino, 2006: 3).

A recurring problem is that rankings are called on to do more than is mandated by the data. Rank order becomes an end in itself. In some cases, institutions are rank ordered when the differences between them are not statistically significant. Typically, rankings particular to certain aspects of inputs or performance become used as general indicators of reputation. The Jiao Tong research rankings are widely interpreted as holistic indicators of the best universities for students despite the warnings of the SJTUIHE group. No ranking or quality-assessment system has been able to generate data based on measures of value added during the educational process; few focus on teaching and learning as such (Dill and Soo, 2005: 503–507), although these data would be most useful for prospective students. There is no necessary connection between the quality of teaching, and student selectivity or research outputs. "The correlation between research productivity and undergraduate instruction is very small and teaching and research appear to be more or less independent activities" (p. 507).

Reputational surveys like that of the Times tend to recycle reputation rather than reward quality (Guarino *et al.*, 2005: 149). "Raters have been found to be largely unfamiliar with as many as one third of the programs they are asked to rate" (Brooks, 2005: 7). Well-known university brands generate halo effects. One American survey of students found that Princeton was ranked in the top 10 law schools in the country. Princeton did not have a law school (Frank and Cook, 1995: 149)!

A still more fundamental problem is that higher education institutions have different goals and missions, vary profoundly throughout the world, and are internally differentiated. By establishing the Anglo-American science university as the dominant model, the main rankings systems tend to weaken diversity by undermining the prestige of institutions with different purposes: technical and vocational institutions, such as the German *Fachhochschulen* and institutions in Finland, Switzerland, and

France; specialist colleges of all kinds; and institutions everywhere that are primarily oriented to teaching. Even among comprehensive research universities there is much variation in mission, size, scope, and languages of use. This category varies from small establishments focused on graduate education and research in relatively expensive scientific fields to the very large access-based national universities of Mexico and Argentina. There are many institutions in between for which preparation in the professions is at least equally important as the research mission. However, once a rankings system is understood as holistic and reputational, it fosters the illusion of a level playing field, obscuring the specific criteria used to compile the rankings and the identity of the institutions and nations favored by the chosen criteria.

Conclusions

Rankings have discernable effects on behaviors. They generate a strong drive to improve the comparative position. Within national systems, global rankings have prompted a greater emphasis on research concentration and in some cases greater institutional stratification, for example, in China, Germany, and Australia. The European Union has proposed the European Institute of Technology which would draw together existing research bases in a mega-university or network capable of challenging the superior SJTUIHE rankings position of the US universities. Global rankings have stimulated global competition for HiCi researchers and the best younger talent, with likely upward effects on the price of research labor. These responses have both cemented the role of the rankings themselves and intensified competitive pressures in the sector (Marginson and van der Wende, in press).

However, there are continuing doubts about whether interpretations of rankings data are sound and are contributing to ongoing improvements in quality, especially teaching quality. Most problems of interpretation are avoided when it is recognized that all rankings in higher education are partial in coverage, and purpose driven, and rankings data are interpreted in the light of those purposes and limits. However, these restraints are mostly not observed in the public discussion of rankings, where holistic reputational rankings, however flawed, carry greater weight than more valid data.

In 2004, an expert group was established to develop principles on good practice in rankings. This International Ranking Expert Group (IREG) was founded by the United Nations Educational, Scientific and Cultural Organization-European Centre for Higher Education (UNESCO-CEPES) in Bucharest and the Institute for Higher Education Policy in Washington. Together with the CHE, IREG published its Principles on

Ranking of Higher Education institutions in May 2006 (see UNESCO/IHEP, 2006). The Berlin Principles focus on the purposes and goals of ranking, the design and weighting of indicators, the collection and processing of data, and on the presentation of ranking results. They are meant to set a framework for the elaboration and dissemination of rankings, to support the continuous improvement and refinement of methodologies, and to guide those producing rankings in holding themselves accountable for the quality of their data collection, methodology, and dissemination. However, the Berlin Principles do not distinguish good rankings from bad. Perhaps the more promising sign is the emergence of the CHE rankings, which sidestep the problems of reputational league tables and provide the most informative data.

See also: Funding of University Research; National Science Policy and Universities; Research Commercialization.

Bibliography

- Altbach, P. (2006). The dilemmas of ranking. *International Higher Education* 42, 2–3.
- Brooks, R. (2005). Measuring university quality. *Review of Higher Education* 29(1), 1–21.
- Center for Higher Education Development (CHE) (2006). *Study and Research in Germany*. University Rankings, published in association with Die Zeit. <http://www.daad.de/deutschland/studium/hochschulranking/04708.en.html> (accessed June 2009).
- Dill, D. and Soo, M. (2005). Academic quality, league tables, and public policy: A cross-national analysis of university rankings. *Higher Education* 49, 495–533.
- Field, K. (2006). Another accountability idea: A new database that would customize college rankings. *Chronicle of Higher Education*, 31 March 2006.
- Frank, R. and Cook, P. (1995). *The Winner-Take-All Society*. New York: Free Press.
- Guarino, C., Ridgeway, G., Chun, M., and Buddin, R. (2005). Latent variable analysis: A new approach to university ranking. *Higher Education in Europe* 30(2), 147–165.
- Institute for Scientific Information, Thomson-ISI (ISI) (2006). Data on highly cited researchers, ISIHighlyCited.com. <http://isihighlycited.com> (accessed June 2009).
- Kirp, D. (2004). *Shakespeare, Einstein and the Bottom-Line: The Marketing of Higher Education*. Cambridge, MA: Harvard University Press.
- Liu, N. and Cheng, Y. (2005). The academic ranking of world universities. *Higher Education in Europe* 30(2), 127–136.
- Liu, N. and Liu, L. (2005). University rankings in China. *Higher Education in Europe* 30(2), 217–227.
- Marginson, S. and Van Der Wende, M. (in press). Globalisation and higher education. *Paper to be Published by the Organisation for Economic Cooperation and Development*. Paris: OECD.
- Newsweek (2006). The world's most global universities, 36–69.
- Rocki, M. (2005). Statistical and mathematical aspects of ranking: Lessons from Poland. *Higher Education in Europe* 30(2), 173–181.
- Salmi, J. and Saroyan, A. (2006). League tables as policy instruments: Uses and misuses. *Paper Prepared for OECD/IMHE and Hochschulrektorenkonferenz, Germany, Workshop 4–5 December 2006, Bonn*.
- Stiglitz, J. (1999). Knowledge as a global public good. In Kaul, I., Grunberg, I., and Stern, M. (eds.) *Global Public Goods: International Cooperation in the 21st Century*, pp 308–325. New York: Oxford University Press.
- Times Higher Education Supplement (THES) (2006). World University Rankings. <http://www.thes.co.uk> (accessed June 2009).

- UNESCO/IHEP (2006). International partnership issues groundbreaking principles on ranking of higher education institutions. http://www.che.de/downloads/Berlin_Principles_IREG_534.pdf (accessed June 2009).
- US News and World Report (USNWR) (2006). *America's Best Colleges*, 2007 edition. Washington.
- Usher, A. and Savino, M. (2006). *A World of Difference: A Global Survey of University League Tables*. <http://www.educationalpolicy.org> (accessed June 2009).
- Van Dyke, N. (2005). Twenty years of university reports cards. *Higher Education in Europe* 30(2), 103–124.
- Williams, R. and Van Dyke, N. (2006). *Rating Major Disciplines in Australian Universities: Perceptions and Reality*. Melbourne: Melbourne Institute of Applied Economic and Social Research.

Further Reading

- Marginson, S. (2006). Dynamics of national and global competition in higher education. *Higher Education* 52, 1–39.
- Shanghai Jiao Tong University Institute of Higher Education, SJTU IHE (2006). Academic Ranking of World Universities. <http://ed.sjtu.edu.cn/ranking.htm> (accessed June 2009).
- Times Higher Education Supplement (2006). World University Rankings. <http://www.thes.co.uk> (accessed June 2009).
- Turner, D. (2005). Benchmarking in universities: League tables revisited. *Oxford Review of Education* 31(3), 353–371.
- Yonezawa, A., Nakatsui, I., and Kobayashi, T. (2002). University rankings in Japan. *Higher Education in Europe* 27(4), 377–382.

Relevant Websites

- <http://www.daad.de> – Center for Higher Education Development (CHE), Germany, including interactive rankings.
- <http://www.educationalpolicy.org> – Educational Policy Institute, Washington and Toronto.
- <http://www.oecd.org> – Organization for Economic Cooperation and Development, Institutional Management in Higher Education program, including meetings concerning rankings in higher education.
- <http://ed.sjtu.edu.cn> – Shanghai Jiao Tong University Institute of Higher Education, including world university rankings.
- <http://www.thes.co.uk> – Times Higher Education Supplement.
- <http://portal.unesco.org> – UNESCO program of activities on higher education.
- <http://www.unesco.org> – UNESCO/International Association of Universities Information Centre.

Steering of Higher Education Systems – The Role of the State

K Yokoyama, University of California, Berkeley, California, USA

V L Meek, University of Melbourne, Victoria, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Neoliberalism – A type of capitalist political philosophy, which emphasizes markets, choice, and competition. It is associated with the theories of Friedrich Hayek, who criticized socialism, statism, and Keynesianism, arguing that the free market mechanism is superior to collectivist planning.

Policy instrument – A method or mechanism used by government, political parties, or individuals to achieve a desired effect through various means.

The Role of the State in Higher Education

The role of the state in higher education is not fixed, but contextual and contingent. It is related to different political, ideological, economic, sociocultural, and historical contexts. Analyses of state/higher education steering relationships are concerned with such concepts as power, legitimacy, influence, ideology, and autonomy.

One of the best-known conceptualizations of the role of the state in higher education coordination is that of the model by Clark (1983). Clark's triangle of coordination of higher education is a heuristic model, incorporating three contrasting forces: state authority, academic oligarchy, and the market. Each corner of the triangle in the model indicates the extreme form of one force and the marginal form of the other two forces. According to Clark, the state is but one force that shapes higher education coordination. Academic oligarchy and the market are also influential – although the ways in which the three forces interact vary according to contexts. The power of the state is, therefore, for Clark, not absolute, but moderated by academic oligarchy and the market.

While Clark's model has been very influential in the study of the coordination of higher education, it tends to ignore two important aspects: the interaction between actors and the state's involvement in the market; and a more pluralistic perspective involving multiple stakeholders. Clark approaches the state, academic oligarchy, and the market as independent actors. However, the market cannot be seen as an actor that is independent of the state since the state controls and conditions the market – for example, by putting a cap on student numbers or

removing it (Olssen and Peters, 2005; Williams, 1995; Yokoyama, 2003).

State actors include central bureaucracies (such as Treasury), specific government agencies, and intermediate bodies (such as former funding councils – Commonwealth Tertiary Education Committee in Australia and University Grant Committee in the UK). The power and influence of these actors wax and wane over time – note the demise of buffer bodies in some countries, mainly due to severe criticism of the efficiency of central state control and the public funding crisis in recent decades. The pattern of state involvement in the coordination of higher education and the policy instruments available to the state through which to effect its involvement (funding, legislation, evaluation, planning, etc.) are influenced by politics, ideology, and the history of specific higher education sectors. For a better understanding of the dynamics of the power relationships between the state and higher education, the way in which particular policy instruments are employed needs to be addressed.

Policy Instruments

Policy instruments are related to the effectiveness of an actor's influence over other actors' decisions and behavior. These are often thought of in terms of degrees of restrictiveness.

The idea of policy instruments has been developed in the areas of public administration and public policy. There is no consensus on the categories of policy instruments in these disciplines. Two scholars of higher education policy, Maassen and van Vught (1994), in reviewing the literature, identify various types of policy instruments (p. 45):

- Prescription, enabling, positive incentives, and deterrence (Bardach, 1979)
- Mandates (providing constraining rules), inducements (providing financial resources to encourage certain activities), capacity (providing financial resources to enable actors to take certain actions), and instruments of system-change (Elmore, 1987; McDonnell, 1988)
- Authority tools, incentive tools, capacity tools, symbolic and hortatory tools, and learning tools (Schneider and Ingram, 1990)
- Information, treasure, authority, and action as government's administrative tools (Hood, 1983).

In the context of higher education, Goedegebuure *et al.* (1994) identify four policy instruments: funding, planning, evaluation, and regulation. Veld *et al.* (1996) highlight the following most usual instruments: regulation, money, and persuasion (pp. 45, 46).

The degree of restrictiveness between instruments is not clear in the context of higher education. Some argue that funding is the most effective instrument. However, funding can be linked to evaluation, as seen in the case of the Research Assessment Exercise (RAE) in England, which indicates the possibility of a combination of different policy instruments. Moreover, the effectiveness of policy instruments is often contextual, influenced by socio-cultural, historical, and political conditions, and, to some degree, contingent. Accordingly, it is a complex task to identify the most restrictive or least restrictive instruments. Nonetheless, understanding the way in which policy instruments are used is useful in the identification of patterns in state steering of higher education systems.

State Steering Models in Higher Education

Governments steer higher education through the use of policy instruments. State steering refers to the “approaches governments use to control and influence specific public sectors, such as higher education” (Gornitzka and Maassen, 2000: 268).

State steering models relate, on the one hand, to state roles and functions, and, on the other, to university autonomy. These models reflect “the efforts of government to steer the decisions and actions of specific societal actors according to the objectives the government has set and by using instruments the government has at its disposal” (van Vught, 1994: 3). They also highlight the importance of institutional context in the policy processes (Gornitzka and Maassen, 2000), which is often related to collegiality and university autonomy. These models largely rely upon theories of organizational governance.

State steering models, according to Gornitzka and Maassen (2000), elucidate the differences between two underlying sets of rules: interaction rules and context rules. The former refers to the “structure [of] the interaction behaviour of actors in a public sector.” The latter is “the way the context in which the interaction takes place is regulated” (p. 268).

There are various ways of depicting how the state goes about coordinating higher education, such as state control models, state regulatory models, or state supervisory models. In this article, the relationships embedded in state steering of higher education are conceptualized as: (1) bilateral; (2) multilateral; and (3) contextual. These models and approaches have been developed mainly at

the Center for Higher Education Policy Studies, University of Twente, the Netherlands.

Bilateral Relationships

Bilateral steering models focus upon two actors: the state and academic professionals or faculty guilds. The state’s actions are often understood as those performed on behalf of society, while self-governance of the academic estate is often linked to the idea of university autonomy. State power is the opposite of universities’ power in the models. The reality is found in the balance between the two. Bilateral steering models are based upon a snapshot of the state/higher education relationship and often lack a historical perspective concerning the transformation of higher education systems.

Van Vught (1989), later with (Neave and van Vught, 1994) and with (Maassen van Vught, 1994), proposes two fundamental models for government regulation of higher education: the model of rational planning and control or the state control model; and the model of self-regulation or the state-supervising model. The former model refers to governmental actors and agencies controlling through stringent rules and an extensive control mechanism, while the latter model relies upon institutional self-regulation. The former is based upon full confidence in the government’s capability to control. It assumes that governmental actors and agencies are capable of acquiring comprehensive knowledge and making the best decisions. The model relies upon the normative ideal of the rationalist perspective of decision making. It was a model found primarily in continental Europe.

The model of self-regulation is in opposition to the state control model. It is based upon the assumption that the devolution of complex decision-making processes to the institutions will bring about robustness, flexibility, and innovation to a significant degree, simultaneously limiting information, administration, and other transaction costs. This model relies upon a cybernetic perspective of decision making and a strategy of self-regulation. The government regards itself as an arbiter who “watches the rules of a game played by relatively autonomous players and who changes these rules when the game is no longer able to lead to satisfactory results” (Maassen and van Vught, 1994: 40). In other words, in this model, the government supervises institutions, which are relatively autonomous, and changes the rules if the result of institutional actions does not fulfill the government’s aims. Self-regulation, however, is not the same as pure market coordination. Van Vught and Maassen argue that the model of rational planning and control tends to go with highly restrictive instruments, while the model of self-regulation fits with the less-restrictive instruments.

Neave and van Vught (1991) suggest bilateral steering models that incorporate contrasting notions of the nation-state: facilitatory states and interventionary states.

The idea of facilitatory states and interventionary states are related to the state-supervising model and the state control model, respectively. They argue that the former suits the characteristics of higher education and promotes diversification and innovation at the institutional level. On the basis of such a theoretical framework, they argue that higher education in Western European countries has shifted from an interventionary state to a facilitatory state, in which procedural autonomy has been extended.

Some of the literature, however, questions the extent to which these dichotomous models reflect the real world. "Especially in Continental Europe the traditional bilateral relationship between higher education and the state is rapidly becoming a multilateral relationship between higher education and various external actors, including the Ministry of Education" (Maassen, 2000: 9).

Multilateral Relationships

Multilateral models of state steering pay attention to the complexity of higher education and not only encompass the rationality-bounded model and institutional self-regulation model, but also highlight the importance of market mechanisms and corporate pluralism in terms of the number of stakeholders involved in the policymaking structure. The elements in multilateral models, therefore, overlap with those in bilateral steering models to a significant degree.

Gornitzka and Maassen's (2000) hybrid steering approach is archetypical of multilateral models. It is based upon Olsen's (1988) four models of state governance and agency autonomy, reinterpreting them in the higher education context. Accordingly, the central concern in their approach relates to the degree of autonomy which government is prepared to allow.

The four models originally proposed by Olsen are based upon organizational theories: the sovereign, rationality-bounded model, institutional state model, corporate-pluralist state model, and supermarket state model. Olsen argues that the relationship between organizational forms and outputs is not simple because organizational forms interact with other variables, modifying the effect of an organizational form from one context to another.

The sovereign, rationality-bounded state model concludes that rationality is problematic because of the limitation of political leaders' time and energy. In this model, leaders choose between tight control of agencies or loose coupling. The former case makes agencies formalistic and lacking in initiative. The latter case enables agencies to pursue self-interest. Higher education in this model is, according to Gornitzka and Maassen, viewed as a political instrument for achieving government goals, and the model emphasizes universities' accountability to political authorities.

The institutional state model sees legitimacy as problematic, assuming that the role of the state is to maintain

norms that are shared by the population and autonomous institutional spheres of society. Agency autonomy is based upon a shared norm about the noninterference of the state. In the higher education context, the model is related to institutional responsibility for the protection of academic values.

The corporate-pluralist state model views power negatively, based upon the assumption that the state is "an area for resolving conflicts between self-interested public and private actors" (Olsen, 1988: 237). The public is considered as one of a variety of self-interested formal organizations. Reorganization is the result of struggles over the control of organizations in constellations of interests, resources, and alliances. Gornitzka and Maassen explore this model in the context of higher education, highlighting a pluralistic point of view which includes multiple stakeholders that shape higher education, such as professional associations, student and staff unions, and industrial and regional authorities.

The supermarket state model is market oriented and involves minimal state control. In this model, the state is viewed as a service provider, with the public as sovereign consumers.

The models outlined above are useful for categorizing state/higher education relationships at a particular point in time. However, they have little or nothing to say about the factors that bring these relationships into being. For causal models of state steering of higher education, we need to look more carefully at contextual variables, including the history of particular higher education systems.

Contextual Relationships

In contrast to the above models, historical, contextual approaches to state steering rely upon theoretical analysis in context. They highlight the transformation from bureaucratic control of higher education to market or market-like coordination, and in the case of Meek's (2002) approach, new coordination approaches in the post-marketization era. Market coordination is based upon the assumption that market forces lead to cost effectiveness, better management, and responsiveness to clients' demands.

Neave's (1988, 1998) evaluative state thesis offers a comparative, historical perspective, emphasizing the importance of how various policy instruments are used. The concept of an evaluative state subsumes Neave's previous work on state control and university autonomy. He argues that government regulation of universities in Western European countries has, since the mid-1980s, stressed efficiency and client responsiveness through increasing use of performance-based evaluation and output-based funding. According to Neave, the shift in the mode of state regulation toward a lighter form of surveillance and the rise of the evaluative state are the result of the government's increasing demands for greater quality and

institutional efficiency, coupled with the introduction of market mechanisms as the supreme regulating principle for higher education.

The underlying philosophy of Neave's evaluative state thesis is neoliberalism, which emphasizes market supremacy within an overall context of continuity of state power. The concept of the evaluative state is antithetical to that of bureaucratization.

The difference between the evaluative state thesis and the bilateral and multilateral state steering models introduced above relates to their respective emphasis on the patterns of macro-management defining the relationship between the state and individual institutions. The difference between them is that the evaluative state highlights the contractual relationship between the state and the universities. In addition, rather than focusing on the degree of state power, the evaluative state thesis identifies particular types of policy instruments and how they have changed from a routine verification and maintenance mode, which was based upon the instruments of guideline laws, decrees, and circulars, to a *a posteriori* evaluation and funding (e.g., performance indicators, quality indices, and standards through benchmarking).

Some scholars have examined the feasibility of the evaluative state thesis as well as contextualized it. For instance, Dill (1998), in evaluating the evaluative state thesis through focusing upon agency theory points out two technical problems in the evaluative state model in relation to the function of public authorities: (1) the principal or evaluative body may be inefficient, caused by ambiguity in the quality measurement of academic output; and (2) the high cost of the state's performance measurement in the evaluatory mode, compared to other instruments such as market competition and professional control.

Meek (2002) questions the usefulness of drawing a strict dichotomy between market steering and centralized state control of higher education. He stresses that the reality is found in the balance between bureaucratic control and market coordination, or between centralized funding and planning, and institutional autonomy. Meek also identifies the negative effects of market coordination and argues that the market may be a greater threat to academic freedom and university autonomy than state bureaucratic control.

Meek identifies two stages in the recent changes to higher education coordination. The first is the shift from state regulation to markets, competition, and entrepreneurship as the result of failure in public policy and bureaucratic controls. The second, which offers a new perspective in the study of state steering in higher education, is the return to state regulation as the result of market failure in achieving the goals and aims of higher education in many countries. New Zealand is a case where the state has backed away from market coordination, reintroducing a buffer body.

There appears to be a movement back and forth between state regulation and market coordination brought about by the continuing attempt to find a proper balance between the two.

The significance of Meek's argument can also be found in his clear identification of the roles of intermediary bodies. He argues that governments have stopped using intermediary bodies – which used to protect the interests of higher education institutions – to regulate higher education in countries such as Australia (not in India and Hong Kong) in favor of market-oriented coordination. The United Kingdom also falls into this category since the nature of the funding councils has changed from being a buffer to being a government agency. It appears that the main reason for this change was the government's desire to directly manage higher education, making higher education institutions more accountable to the state and society as a whole, as well as making higher education contribute more directly to economic growth. In continental European countries and Japan, opposite trends are observed, moving in the direction of steering at a distance, emphasizing deregulation and quality control. Important national differences in state steering models remain.

Conclusion

Neoliberal political ideology underpins much of the recent change to state steering of higher education. For example, the neoliberal political philosophy, with its emphasis on the market and light state surveillance, is clearly evident in both the supermarket and evaluative state models. The supermarket and evaluative state models emphasize the state's supervisory function without tight, detailed control of institutions, as well as the self-regulatory nature of higher education institutions based on strong executive leadership and the efficiency of resource utilization and management. These models and approaches are to a significant degree illustrative of current higher education trends in Australia, Western Europe, and Japan. The state's roles in the evaluative state model, in theory, encapsulate the idea of the separation of funder and provider, and policy and delivery.

The evaluative state thesis identifies evaluation – in particular, output rather than input and process quality control – as the government's main policy instrument in the postroutine evaluation regime and in the context of scarce public resources. A *a posteriori* evaluation as well as financial incentives are common policy instruments currently in use in Western European countries and Japan. In this respect, it can be argued that the higher education systems in these countries are converging on the evaluative state model.

The higher education system can be understood as a subsystem of the nation-state system. It can also be

understood as a part of a state education system – which is “a nationwide and differentiated collection of institutions devoted to formal education, whose overall control and supervision is at least partly governmental, and whose component parts and processes are related to one another” (Archer, 1984: 19) – regardless of whether individual institutions are public or private. However, challenge to traditional state steering models and approaches in higher education have also arisen from new global and regional trends, such as the Bologna process. Meek (2002) identifies the global dimension – including international higher education consortia such as Universities 21 (a network of 18 leading research universities in ten countries) and supranational coordination authorities such as the United Nations Educational, Scientific and Cultural Organization (UNESCO), the World Bank, and the European Union – and attempts to incorporate it into the concept of higher education coordination. These supranational actors and agreements affect the choices which government can make, which in turn changes the mode of higher education coordination.

In the final analysis, the steering of higher education is observed within the nation-state system. The role of the state has been modified by various forces and trends and is not the same in all countries. However, in the end, it is the state that is supreme. For example, education and training fall under the principle of subsidiarity in the European Union. The Bologna Declaration, which is a multinational agreement initiated by education ministers of individual countries, relies upon its implementation by the government of each nation-state. The steering of higher education remains the state’s prerogative, with the exact form that steering assumes being dependent on a variety of national political and ideological factors.

See also: Higher Education and the Transformation of Society.

Bibliography

- Archer, M. (1984). *Social Origins of Educational Systems*. London: Sage.
- Bardach, E. (1979). *The Implementation Game*. Cambridge, MA: MIT Press.
- Clark, B. R. (1983). *The Higher Education System: Academic Organization in Cross-National Perspective*. Berkeley, CA: University of California Press.
- Dill, D. (1998). Evaluating the ‘evaluative state’: Implications for research in higher education. *European Journal of Education* 33(3), 361–377.
- Elmore, R. F. (1987). Instruments and strategy in public policy. *Policy Studies Review* 7(1), 174–186.
- Goedegebuure, L., Kaiser, F., Maassen, P., et al. (eds.) (1994). *Higher Education Policy: An International Comparative Perspective*. Oxford: Pergamon Press.
- Gornitzka, A. and Maassen, P. (2000). Hybrid steering approaches with respect to European higher education. *Higher Education Policy* 13, 267–285.

- Hood, C. (1983). *The Tools of Government*. London: MacMillan.
- Maassen, P. (2000). *Governmental Policies and Programmes for Strengthening the Relationship between Higher Education Institutions and the Economy*. Utrecht: Center for Higher Education Policy Studies.
- Maassen, P. and van Vught, F. (1994). Alternative models of governmental steering in higher education: An analysis of steering models and policy-instruments in five countries. In Maassen, P. and van Vught, F. (eds.) *Comparative Policy Studies in Higher Education*, pp 35–64. Utrecht: Center for Higher Education Policy Studies.
- McDonnell, L. (1988). Policy design as instrument design. *Paper Presented at the 1988 Annual Meeting of the American Political Science Association*. Washington, DC.
- Meek, V. L. (2002). Changing patterns in modes of co-ordination of higher education. In Enders, J. and Fulton, O. (eds.) *Higher Education in a Globalising World*, pp 53–72. Dordrecht: Kluwer Academic Publishers.
- Neave, G. (1988). On the cultivation of quality, efficiency and enterprise: An overview of recent trends in higher education in Western Europe, 1986–1988. *European Journal of Education* 23, 7–23.
- Neave, G. (1998). The evaluative state reconsidered. *European Journal of Education* 33, 265–284.
- Neave, G. and van Vught, F. (1991). *Prometheus Bound: The Changing Relationship between Government and Higher Education in Western Europe*. Oxford: Pergamon Press.
- Neave, G. and van Vught, F. (1994). *Government and Higher Education Relationships across Three Continents*. Oxford: IAU Press/Pergamon.
- Olsen, J. P. (1988). Administrative reform and theories of organization. In Campbell, C. and Peters, B. (eds.) *Organizing Governance: Governing Organizations*, pp 233–254. Pittsburgh, PA: University of Pittsburgh Press.
- Olssen, M. and Peters, M. (2005). Neo-liberalism, higher education and the knowledge economy: From the free market to knowledge capitalism. *Journal of Education Policy* 20(3), 313–345.
- Schneider, A. and Ingram, H. (1990). Behavioral assumptions of policy tools. *Journal of Politics* 52(2), 510–530.
- Van Vught, F. (ed.) (1989). *Governmental Strategies and Innovation in Higher Education*. London: Jessica Kingsley.
- Van Vught, F. (1994). Policy models and policy instruments in higher education: The effects of governmental policy making on the behaviour of higher education institutions. In Smart, J. C. (ed.) *Higher Education: Handbook of Theory and Research*, pp 88–125. New York: Agathon Press.
- Veld, R., Fussell, H. P., and Neave, G. (eds.) (1996). *Relations between State and Higher Education*. The Hague: Kluwer Law International.
- Williams, G. (1995). The ‘marketization’ of higher education: Reforms and potential reforms in higher education finance. In Dill, D. and Sporn, B. (eds.) *Emerging Patterns of Social Demand and University Reform: Through a Glass Darkly*, pp 170–193. Oxford: IAU Press.
- Yokoyama, K. (2003). *Ideologies, Policies, and the Control of the University Systems in England and Japan*. Doctoral Thesis, London: Institute of Education.

Further Reading

- Foucault, M. (1991). Governmentality. In Burchell, G., Gordon, C., and Miller, P. (eds.) *The Foucault Effect: Studies in Governmentality*, pp 87–104. Chicago, IL: University of Chicago Press.
- Kogan, M. and Marton, S. (2000). The state and higher education. In Kogan, M., Bauer, M., Bleiklie, I., and Henkel, M. (eds.) *Transforming Higher Education: A Comparative Study*, pp 89–108. London: Jessica Kingsley.
- Marginson, S. (1997). Steering from a distance: Power relations in Australian higher education. *Higher Education* 34, 63–80.
- Teichler, U. (1996). The changing nature of higher education in Western Europe. *Higher Education Policy* 9(2), 89–111.
- Van Vught, F. (1997). Combining planning and the market: An analysis of the government strategy towards higher education in the Netherlands. *Higher Education Policy* 10(3/4), 211–224.

Student and Faculty Transnational Mobility in Higher Education

U Lanzendorf and B M Kehm, University of Kassel, Kassel, Germany

© 2010 Elsevier Ltd. All rights reserved.

Introduction – History and Typology

Student and faculty transnational mobility is defined as the crossing of national borders for the purpose of studying or teaching in higher education or engaging in research abroad. Academic mobility will be used here as a generic term referring to both groups of participants (students and faculty).

Historically, the golden age of academic mobility was the Middle Ages when nation-states did not exist. There were only a limited number of cosmopolitan universities across Christian Europe which adhered to similar study programs taught in Latin. This enabled students and faculty to gain academic experience at different institutions across the region. One of the best-known wandering scholars of that time, Erasmus of Rotterdam, later gave his name to the largest European support program for academic mobility and cooperation (ERASMUS which is also the acronym for European Community Action Scheme for the Mobility of University Students).

Contemporary forms of academic mobility go back to the 1920s (United States) and to the period after World War II (Europe). Different development phases can be distinguished for the United States and Europe, which are the traditional geographical foci of transnational student flows (Baron, 1993; Briggs and Burn, 1985; de Wit, 2002).

The first phase, which lasted until the mid-1970s, was dominated by one-way student flows from countries with less-developed higher education systems to countries where higher education was well developed. Mainly, students from the South went to a limited number of countries in the North, namely the United States, France, the USSR, Germany, the UK, Canada, and Italy. Teichler refers to this type of student mobility as vertical (Teichler and Jahr, 2001: 443). Since the objective of vertically mobile students is usually to obtain a degree abroad, this type of mobility is also referred to as degree mobility.

Among the countries with well-developed higher education, study experiences abroad were actively promoted primarily in the US. Already since the 1920s, the so-called junior year abroad provided third-year undergraduate students with an intercultural experience and helped to strengthen their foreign-language proficiency. It is organized by the home universities of participating students, which often have branch campuses abroad so that mobile students do not necessarily have to be affiliated to a foreign institution of higher learning.

National education and immigration policies have been important enabling or limiting factors for student mobility from its inception. During the first development phase, host countries valued foreign students as future contacts for the promotion of their political or economic interests abroad. Their policy of open doors' was an element of foreign policy and initially enabled high growth rates of inbound student flows. Later, host countries brought this trend to a halt by strategically regulating and partly restricting student influx, for example, by setting quotas by field of study, imposing high entrance qualifications, increasing study fees for foreign students, or making visa regulations stricter.

In the second development phase of academic mobility (mid-1970s to 1987), political strategies to guide transnational student mobility became more elaborate and students from countries with well-developed higher education systems increasingly participated. In Europe, the promotion of study abroad displaced the regulation of foreign-student inflow at the top of the political agenda. The ideas of student exchange and intercultural learning gained ground and complemented the traditional search for learning opportunities by students from underprivileged world regions as determinants of transnational student flows. The most important destination of intra-European mobility at that time was the United Kingdom. Student mobility between similarly developed national higher education systems has been called horizontal (Teichler and Jahr, 2001: 443) or temporary, the latter term referring to the fact that students did not want to obtain a degree abroad, but only to spend part of their study program at a foreign higher education institution. With the expansion of this type of mobility, the recognition of learning abroad by the home institutions of mobile students became a relevant topic.

Another new element of the second phase was the systematic promotion of faculty mobility in Europe. In fact, a support scheme for the mobility of university teaching staff was the first element of supranational mobility policies introduced under the objective of regional integration. In 1976, the European Commission established the Joint Study Programme to support international faculty contacts and stimulate joint curriculum development. Soon after, however, financial support was focused on the promotion of student and teacher mobility because these activities seemed to be more visible for a broad public. The idea developed to offer students the possibility to participate in organized study-abroad programs – a strategy which

resembled the longer-standing practice of American universities. Finally, in 1984, a specific budget line for student mobility grants was introduced into the Joint Study Programme. With this measure, the European Commission took the European lead in the promotion of academic mobility.

The third phase in the development of academic mobility started with the introduction of the European support program for horizontal academic mobility called ERASMUS in 1987 and lasted until 1999. ERASMUS stands for the strategic supranational support of student and faculty mobility within the context of regional integration in Europe (see, e.g., Smith, 1996) and complements national schemes for academic mobility. With respect to student mobility, its objective was to make study periods at foreign universities attractive by providing grants that covered the additional costs of study abroad and promoting the support of mobile students with respect to admission procedures and accommodation abroad, health insurance, and academic recognition upon return to the home university. As a new subcategory of transnational student mobility, self-organized versus organized mobility became relevant. Students who did not participate in a particular scheme but organized their study abroad period themselves came to be called free movers. The length of ERASMUS study abroad periods is between 6 months and 1 year. They are explicitly intended to serve higher education in addition to linguistic and intercultural learning. Apart from student mobility, ERASMUS also supports teaching periods at foreign partner universities for faculty. It is expected that faculty mobility provides new learning experiences to students and stimulates joint curriculum development and internationally integrated study programs of which student mobility is an essential part. The immediate objective of ERASMUS was the promotion of mutual understanding and interaction between national higher education systems in Europe. This, however, contributed to a so-called provocation of national systems – discussions on alternative ways of organizing higher education were stimulated – and can therefore be assumed as indirectly having paved the way for policies of structural convergence in European higher education (Wächter, 1996).

Especially after the Treaty of Maastricht had created a legal basis for the European Commission's activities in higher education in 1992, academic mobility was the nucleus of comprehensive Europeanization policies in higher education. As from 1996, ERASMUS was integrated into the newly set up umbrella program SOCRATES which, up to 2006, promoted Europeanization in the fields of higher education, school education, and continuing/adult education. Concomitantly, the support for student mobility and cooperation was substantially increased and the development of internationalization strategies by universities was introduced as a new element of the ERASMUS program. Another new policy objective was to make

the nonmobile higher education students benefit from Europeanization by stimulating internationalization at home. Successively, the number of countries participating in SOCRATES increased beyond the European Union (EU) member states. In 2007, 31 countries were eligible for program funding.

For Europe, a fourth phase can be identified, starting with the Bologna Declaration in 1999 and running until the present. The so-called Bologna Process was initiated by the self-commitment of the ministers of education of 29 EU member states to work toward a structural harmonization of their national higher education systems. The core element of reform is the change to a tiered system of study programs and degrees (bachelor and master) in all signatory states of the Bologna Declaration (by 2007 their number had grown to 45). Thus, national governments may be regarded as finalizing the process of integration in European higher education which had initially been started by the European Commission. The Bologna Process aims to create a European Higher Education Area by 2010 to enable European students to change countries during their course of study without facing problems of recognition of prior learning. In this context, the development of joint or double-degree programs organized in cooperation with universities located in different countries receives particular attention. In recent years, the topics of quality assurance, academic recognition, and doctoral education have been added to the Bologna policy agenda. The European Higher Education Area is expected to significantly increase the visibility and attractiveness of European universities to students from the Third World countries.

In the US, the terrorist attacks of September 2001 triggered substantial changes in the conditions for student and faculty mobility. Government regulations regarding visas and reporting obligations to government agencies became stricter, and international students were charged new types of fees (Altbach, 2004). Recently, US policy toward transnational student mobility has outlined a future focus on study abroad (Senator Paul Simon Study Abroad Act of 2007). The aim is to set up a national study-abroad program in order to significantly increase the transnational mobility of national students so as to make study abroad the routine, rather than the exception, and reach participation of at least 1 million American students each year within 10 years. Particular emphasis is laid on the necessity to direct more students to nontraditional destination countries.

Statistical Overview – Patterns and Trends

The global pattern of student mobility is relatively well documented by the United Nations Educational, Scientific

and Cultural Organization's (UNESCO) annual compilation of national enrolment data which covers nearly all countries in the world. With respect to faculty mobility, however, data are scarce. The only internationally comparative statistical reference is provided by figures on the participation in short-term ERASMUS teaching assignments abroad. Further data are available from national surveys, regularly conducted, for example, in the US or Germany. They are, however, not comparable between countries and primarily address the mobility of scholars for purposes of research.

Student Mobility

According to UNESCO statistics, there were some 2.5 million mobile tertiary education students worldwide in 2004 (UNESCO, 2006). This figure excludes students who are enrolled outside their country of permanent residence for up to one academic year. However, participation figures in temporary, that is, nondegree, mobility are comparatively small. In 2004/05, just over 200 000 study-abroad students from the US received academic credit after returning to their home institution (Koh Chin and Bhandari, 2006). In addition, there were about 145 000 ERASMUS students, and a modest number of temporary mobile students supported by programs funded by individual European countries (Maiworm and Wächter, 2006).

Since 1975, when internationally comparative data became available, worldwide participation in student mobility has more than tripled. The sharpest rise ever in foreign student numbers took place at the beginning of the twenty-first century. Comparing figures for 1999 with those of 2004, a 41% increase can be observed (UNESCO, 2006: 34). However, the growth rates of student mobility have always been roughly in line with the overall increase in participation in higher education. Only recently has the actual share of transnationally mobile students among the total number of tertiary education students slightly increased. Scenarios on the future growth of the worldwide demand for international higher education have been presented by an Australian study (Böhm *et al.*, 2002). International higher education involves student mobility and may also be provided by foreign institutions in the students' home countries. The study came to the conclusion that the demand for international education will at least double between 2000 and 2015 and double again by 2025 to reach more than 7 million students.

What is the current structure of student mobility? In 2004, the pattern of outbound student flows was as follows:

- UNESCO statistics show that almost 30% of students spending more than 1 year abroad came from East Asia and the Pacific. Almost half of these (nearly 350 000) came from China alone. The second most important

region of origin is Western Europe, sending nearly 20% of worldwide mobile students (roughly 400 000). Student mobility from East Asia and the Pacific is increasing, whereas participation from Western Europe is stagnating. The third most important sending region is Central and Eastern Europe with 12% of all mobile students on a worldwide scale (just below 300 000 students).

- The major individual sending country is China (sending 14% of worldwide mobile students), followed by India (5%), the Republic of Korea (4%), Japan (2.5%), and Germany (2%).
- The overwhelming majority of mobile students from Western Europe (77 %) remains within this region. With about 40%, the proportions of students from North America or East Asia and the Pacific staying within their regions are also high.
- Notably, the share of outgoing students among all tertiary students is lowest in the United States (0.2%), whereas the countries of sub-Saharan Africa on average have by far the highest share of tertiary students learning abroad for more than 1 year (6 %) (UNESCO, 2006: 37). Yet, the fact that in Western Europe the share of tertiary students enrolled abroad also surpasses the worldwide average shows that it is not only a low development status of higher education which makes large numbers of students decide to study abroad. Further, it is interesting to note that in East Asia and the Pacific the proportion of students enrolled abroad among all tertiary education students is close to the global average. Obviously, Asian students are not particularly inclined to study abroad, although a very high absolute number of transnationally mobile students come from Asia.

For inbound mobility, the following aspects are noteworthy:

- Western Europe and North America stand out as the most important destinations of transnationally mobile students, currently attracting 44% and 25% of worldwide mobile students, respectively. Another 15% decide to study in East Asia and the Pacific. (It must be noted that data for Western Europe tend to overestimate inbound student mobility because they refer to students with foreign nationalities, a relevant number of who live permanently in their country of study (about 32% in Spain, 23% in the UK, and 16% in Germany. See Lanzendorf, 2006a).)
- The most important individual target country of mobile students in the world is the United States, with 23% of mobile students. The next four most frequent study destinations each account for a maximum of about half the US American share in worldwide student flows: The United Kingdom (12%), Germany (11%), France (10%), and Australia (7%). Together, the five

major destination countries attract 62% of mobile students worldwide. Among the developing countries, South Africa and Malaysia host substantial numbers of incoming students. They benefited from the recent expansion of mobile student numbers, whereas international student enrolment in the US has not grown since the academic year 2002/03 (Koh Chin and Bhandari, 2006). This is probably due to stricter access regulations after the events of 11 September 2001.

- For the United States, the structure of inbound student mobility is rather well documented in the annual Open Doors report (Koh Chin and Bhandari, 2006). Data show that incoming graduate students have slightly outnumbered incoming undergraduate students since the beginning of this century. The most frequent field of study of inbound mobile students is business and management (about 18% of all international students), closely followed by engineering (about 16%).
- In Europe, many national higher education systems traditionally have a single-tier study structure. It will therefore only be possible to distinguish undergraduate from graduate inbound mobility after the implementation of the Bologna goals which envisage a common two-tier structure. ERASMUS statistics show that most temporary mobile students in Europe come from business studies (more than one-fifth). The second and third most highly represented fields of study are the social sciences and engineering/technology (about 11% each).
- In Germany, students from economic fields comprised 16% of nonpermanent resident students in 2005 (Deutscher Akademischer Austauschdienst and HIS 2006). The next most frequent subjects are German and informatics (about 9% of non permanent resident students each). They are closely followed by electrical engineering and mechanical and process engineering. The social sciences do not figure among the ten most often represented fields of study.
- For some target countries, foreign students – who do not necessarily have to be inbound mobile students – make up significant shares of all tertiary students. Particularly high percentages of foreign students among all tertiary students can be observed for Switzerland (18%), Australia (17%), Austria (14%), New Zealand, the UK (13% each), France, and Germany (11% each) (UNESCO, 2006: 48).

Faculty Mobility

Faculty mobility is statistically documented for the US, Germany, and for the ERASMUS program. In the US, available data refer to inbound mobile academics at doctoral-degree-granting institutions (Koh Chin and Bhandari, 2006). Of almost 97 000 incoming scientists

and scholars in 2005/06, the overwhelming majority primarily engaged in research. Only less than 20% (nearly 20 000 scholars) came either for teaching (12%) or for teaching and research purposes (7.3%). The overall number of incoming academics has steadily increased since the beginning of the 1990s, and the trend shows a slight increase in the share of scholars with teaching only functions. Of all incoming academics, most originate from China (nearly 20%), Korea, and India (9% each). In 2005/06, the life and biological sciences were represented most often, followed by the health sciences (23% and 20%, respectively). The two fields, physical sciences and engineering, accounted for over 10%. Other fields of specialization had only marginal relevance.

For Germany, data on inbound and outbound research mobility supported of (post-)doctoral researchers and senior faculty are available (Deutscher Akademischer Austauschdienst and HIS, 2006). They are, however, limited to the participation in major national support programs. In 2004, nearly 21 000 incoming researchers and over 4000 outgoing researchers received support. Among the incoming researchers, about 50% were doctoral students or graduate students pursuing a research project, and among the outgoing researchers almost two-thirds belonged to this group. The absolute number of outgoing researchers has clearly been decreasing in recent years. The number of incoming researchers, however, has shown a slight increase. The majority of mobile researchers came from science and engineering fields.

Within ERASMUS, more than 20 000 teachers taught temporarily in another of the 31 countries participating in the SOCRATES program in 2004/05. The largest number of ERASMUS teachers came from Germany (about 13%, corresponding to about 2% of academic faculty at German universities). Spain and France were the next two major sending countries, accounting for about 11% each of overall faculty participation in SOCRATES/ERASMUS. Germany, France and Italy, followed very closely by Spain, were also the major destinations of ERASMUS teaching staff mobility. Teaching periods abroad lasted on average about 1 week with at least 8 h teaching. Teaching staff mobility is most frequent in languages and philological sciences as well as in engineering/technology with about 15% and 14% of participants, respectively. Business studies come third. However, this field of study comes first among outward faculty flows from Central and Eastern European countries (about 19%).

Perspectives of Students and Governments and Impacts of Academic Mobility

Students studying for a degree abroad usually expect that the foreign degree will open up opportunities for

obtaining a good job in their home country or abroad. The selection of a host country is based on criteria such as language, admission practices, study costs, reputation of a higher education system or specific subjects, and general living conditions abroad, including attitudes toward foreign students. As demonstrated by large student flows between countries with identical or similar languages or neighboring countries, many mobile students prefer to study in cultural contexts with which they are familiar (see, e.g., Lanzendorf, 2006b).

A recent study carried out on behalf of the European Commission provides insights into the criteria applied by students in Brazil, China, India, Mexico, Russia, and Thailand when planning to study abroad (Wächter and Maiworm, 2006). In 2004, a survey was conducted among nearly 20 000 students from these six countries. They were asked to provide information about their motivation for planning international mobility and their perception of European higher education. The results of the survey show that they consider the master's level to be the most appropriate period for going abroad. Their most frequent general motivation was to experience new ways of thinking and acting in their field of study (88%). Furthermore, they hoped to improve their opportunities for an international career or a career in their home country, to improve their language skills (81% each), and to have an opportunity for personal development (76%). In contrast, limited access to high-quality education at home or a lack of opportunities to specialize in a particular subject area were of minor importance (Wächter and Maiworm, 2006: 169). The students surveyed considered the worldwide recognition of degrees from a country and the prestige and quality of an institution as the major criteria in selecting the host country for their study-abroad period (Wächter and Maiworm, 2006: 180). Students notably looked for specific and high-quality provisions in their area of specialization and an up-to-date and well-managed institution with a high standing and affordable fees. As major obstacles and problems, more than 50% of the students surveyed quoted insufficient financial resources. Immigration regulations and language preparation were also mentioned as problematic aspects.

It is widely assumed that a substantial proportion of internationally mobile students intend to live and work abroad, possibly in the host country of study, after graduation. However, little is known about the share of foreign graduates who do not go back to their home countries. Suter and Jandl (2006: 8) provide some information. According to them, for Canada, survey data suggest that between 15% and 20% of students from abroad on average do not return to their home countries after graduation. In Norway, about 18% of non-European students stay in the country. In the UK, the retention rate for students from European countries is particularly high, that is, about 27% in 2005.

Students' motivations and expectations with respect to nondegree mobility have been relatively well investigated. Comprehensive information is available from evaluation studies on the ERASMUS program. The most recent survey among former ERASMUS students from all countries participating in the program – the ERASMUS 2000 study – referred to mobile students and faculty of the 1998/99 academic year. With respect to students' motives for spending a study period at a foreign university, it was found that a number of different aspects were important, that is, improvement of foreign-language proficiency, self-development, academic learning experience, the wish to travel, and better opportunities for future professional development (Maiworm and Teichler, 2002: 87). Some variations could be observed for different fields of study. Academic and educational matters were especially salient for students in the fields of agriculture, architecture, fine arts, and mathematics. For students of education, cultural reasons were particularly relevant. Students of economics had a particular interest in career advancement. The overwhelming majority of participating students had already spent a minimum of one period abroad of at least 1-month duration before going abroad with ERASMUS (Maiworm and Teichler, 2002: 87).

With respect to the perceived impact of temporary ERASMUS study abroad, a general finding is that students regard academic matters highly, but observe more impressive advancement in foreign-language proficiency, personality development, and career enhancement. This picture differs only moderately according to students' home countries. The relatively cautious ratings of academic progress could be explained by students taking this kind of success largely for granted and being more excited about other learning experiences. This interpretation is substantiated by the finding that more than 50% of the students who were surveyed in the ERASMUS 2000 study rated their academic progress abroad more highly than what they would have expected for a corresponding period at home. Only 18% of the 1998/99 students thought that they made less academic progress abroad than at home (Maiworm and Teichler, 2002: 110). In addition, according to the ERASMUS 2000 study, about 80% of mobile students were either granted credit for study abroad or their study achievements were otherwise considered equivalent to requirements of their home program by their home institution. Almost 60% of the courses recognized upon return were recognized as equivalent to mandatory courses at home (Maiworm and Teichler, 2002: 109). Irrespective of this positive finding, more than 50% the 1998/99 ERASMUS students surveyed expected to need more time to graduate than would have been necessary had they not gone abroad. They estimated that their overall period of study would be prolonged by more than half the duration of their study-abroad period.

Recent research on the professional impact of ERASMUS-supported study-abroad periods across

graduates from all participating countries suggests that the relevance of temporary study abroad to obtain the first job and its perceived positive impact on the type of work and income had decreased during the last decade (Bracht *et al.*, 2006). In a first survey among former ERASMUS students who had been employed up to 3 years, in 1993, 71% of respondents believed that the period abroad had been helpful in obtaining the first job. In a subsequent survey of the year 2000, this figure decreased to 66%, and in the most recent study conducted in 2005, only 54% of graduates surveyed perceived a positive outcome in this respect. A positive impact on the type of work and the income level has always been recorded by a minority of former ERASMUS students. In 1993, 49% perceived a positive impact on the type of work and 25% on income; the respective figures 12 years later were 39% and 16%. These findings could reflect that international experience was becoming a less-exceptional asset with globalization. The authors of the survey called it a “gradual decline of the uniqueness of the ERASMUS experience” (Bracht *et al.*, 2006: xxiv).

A survey of about 20 000 academic staff in 14 countries conducted in the mid-1990s revealed that highly research active faculty were more inclined toward international activities than those with a preference for teaching (Welch, 1997). Surveys of former ERASMUS teachers point in a similar direction. They come to the conclusion that academic activity abroad is particularly relevant for expanding research contacts. In the most recent one, the improvement of research contacts was more often stated as a positive impact of temporary teaching abroad than the improvement of teaching and the development of new teaching methods (Bracht *et al.*, 2006). In general, former ERASMUS teachers found a particularly positive impact of short-term teaching stays abroad on their intercultural understanding and the breadth of their academic knowledge. With respect to individual career prospects, however, they did not have an impression of benefit. Overall, the most recent study comes to the conclusion that the professional impact of 1-week faculty assignments abroad is impressively high. It is particularly high for faculty from Central and Eastern European countries (Bracht *et al.*, 2006).

National governments and international organizations support student mobility for various reasons. At the national level, human capital development plays a major role. Mobile students are expected to acquire competences which they would not develop if they did not go abroad. Whereas brain drain is a major concern of countries sending large numbers of students abroad, host countries increasingly develop brain-gain strategies. As Suter and Jandl (2006) report, European countries and, for example, New Zealand have introduced policies to facilitate the integration of international graduates in the home labor market. In the US and Canada, providing such options is already a long-standing practice. On the contrary, for other

industrialized countries, it can be observed that scholarships for students from the South increasingly require that students must return to their home country or else reimburse the scholarship after finishing their studies. As stated above, at the European level, student mobility is viewed as contributing to regional integration.

Outlook

Recent policy initiatives in the US and in Europe suggest that major developments can be expected in the field of academic mobility, especially concerning student mobility. The US initiative to reach 1 million study-abroad students per year has the potential to fundamentally alter the global picture of student flows. If it reaches its ambitious objectives, it will substantially increase the weight of short-term, nondegree mobility and of student flows from the North to the South.

In Europe, the introduction of a system of tiered study programs and degrees according to the bachelor and master model has led to concerns that mobility could decrease because the new study programs are shorter, are structured in a more rigorous way, and – at least at the bachelor level – frequently have very dense curricula. A recent study (Bürger *et al.*, 2006) tried to answer the question for German students whether the new tiered study programs would discourage them from temporary study abroad with ERASMUS. The results of the study showed that the new study programs did not discourage mobility *per se* but that greater efforts had to be made in the accompanying measures, that is, information and advice, integration of study abroad into the curriculum, learning agreements for study abroad, recognition, and contacts between home and host institution. The study was also able to show an emerging trend toward more degree mobility, that is, going abroad for a whole program (mostly for a master’s degree) rather than just spending a temporary period abroad within an ongoing program. In order to substantially increase the number of European higher education graduates who have experienced foreign teaching and learning cultures, the feeling is that alternative means to student mobility must be further developed; especially the internationalization of teaching at home deserves more attention.

On a global scale, discussions on the future of academic mobility are increasingly shaped by issues of transnational education. Universities from countries with well-developed higher education systems export their study programs to countries where the national demand for study places cannot be met by local institutions. Thus, students in these countries do not necessarily have to go abroad any more to study for a foreign degree. The spread of transnational education is expected to complement rather than to lower participation in transnational student

mobility. However, it does lead to a new need for faculty mobility insofar as mother universities are involved in the teaching and quality assurance of their programs abroad.

A further trend to be noted is that international competition becomes increasingly complementary to intercultural learning and cooperation as rationales for the political support of academic mobility. In the context of globalization, national higher education policies increasingly stress the relevance of generating an additional institutional income by promoting the enrolment of foreign students and gaining competitive advantages by attracting and keeping best talent from abroad. In most European countries, for example, universities are keen to develop English-language-taught degree programs, particularly at the master and doctoral level to attract students from abroad. Through its openness and – in many countries – comparatively modest tuition fees, Europe as a study destination is becoming more and more of a competitor to the US. In parallel, however, new destination countries of transnational student mobility such as South Africa or Malaysia emerge.

See also: Higher Education Crossing Borders; Internationalization of Higher Education.

Bibliography

- Altbach, P. (2004). Higher education crosses borders. *Change* 36(2), 18–25.
- Baron, B. (1993). The politics of academic mobility in Western Europe. *Higher Education Policy* 6(3), 50–54.
- Böhm, A., Davis, D., Meares, D., and Pearce, D. (2002). *Global Student Mobility – Forecasts of the Global Demand for International Higher Education*. Sydney: IDP Education Australia.
- Bracht, O., Engel, C., Janson, K., et al. (2006). *The Professional Value of ERASMUS Mobility*. http://ec.europa.eu/education/programmes/llp/erasmus/evalcareersum_de.pdf (accessed June 2009).
- Briggs, A. and Burn, B. (1985). *Study Abroad. A European and an American Perspective*. Paris: European Institute of Education and Social Policy.
- Bürger, S., Günther, M., Kehm, B. M., et al. (2006). International study on transnational mobility in bachelor and master programmes. In German Academic Exchange Service (DAAD) (ed.), *Transnational Mobility in Bachelor and Master Programmes*, pp 1–65. Bonn: German Academic Exchange Service (DAAD).
- Deutscher Akademischer Austauschdienst (DAAD) and HIS Hochschul-Informationssystem (eds.) (2006). *Wissenschaft weltoffen 2006 – Daten und Fakten zur Internationalität von Studium und Forschung in Deutschland*. Bielefeld: Bertelsmann.
- Koh Chin, H. and Bhandari, R. (2006). *Open Doors 2006: Report on International Educational Exchange*. New York: Institute of International Education.
- Lanzendorf, U. (2006). Inwards and outwards mobile students. In Kelo, M., Teichler, U., and Wächter, B. (eds.) *Eurodata – Student Mobility in European Higher Education*, pp 54–77. Bonn: Lemmens.
- Lanzendorf, U. (2006). Foreign students and study abroad students. In Kelo, M., Teichler, U., and Wächter, B. (eds.) *Eurodata – Student Mobility in European Higher Education*, pp 7–53. Bonn: Lemmens.
- Maiworm, F. and Teichler, U. (2002). The students' experience. In Teichler, U. (ed.) *ERASMUS in the SOCRATES Programme – Findings of an Evaluation Study*, pp 83–115. Bonn: Lemmens.
- Maiworm, F. and Wächter, B. (2006). Student mobility in national programmes. In Kelo, M., Teichler, U., and Wächter, B. (eds.) *Eurodata – Student Mobility in European Higher Education*, pp 182–192. Bonn: Lemmens.
- Smith, A. (1996). Regional cooperation and mobility in a global setting – the example of the European Community. In Blumenthal, P., Goodwin, C., Smith, A., and Teichler, U. (eds.) *Academic Mobility in a Changing World*, pp 129–146. London: Jessica Kingsley.
- Suter, B. and Jandl, M. (2006). *Comparative Study on Policies towards Foreign Graduates – Study on Admission and Retention Policies towards Foreign Students in Industrialised Countries*. Vienna: International Centre for Migration Policy Development and Advisory Committee on Aliens Affairs.
- Teichler, U. and Jahr, V. (2001). Mobility during the course of study and after graduation. *European Journal of Education – Research, Development and Policies* 36(4), 443–458.
- UNESCO Institute for Statistics (2006). *Global Education Digest 2006 – Comparing Education Statistics across the World*. Montreal: UNESCO Institute for Statistics.
- Wächter, B. (1996). Probleme der europäischen Bildungspolitik. Vielfalt und offener europäischer Bildungsraum. *Das Hochschulwesen* 1996/1, 6–10.
- Wächter, B. and Maiworm, F. (2006). Perceptions of European Higher Education in Third Countries. <http://ec.europa.eu/education/programmes/mundus/acareport.pdf> (accessed June 2009).
- Welch, A. R. (1997). The peripatetic professor: The internationalisation of the academic profession. *Higher Education* 34(3), 323–345.
- de Wit, H. (2002). *Internationalization of Higher Education in the United States of America and Europe: A Historical, Comparative, and Conceptual Analysis*. Westport, CT: Greenwood Press.

The Bologna Process in European Higher Education

D Rich, University of Tasmania, Launceston, TAS, Australia

© 2010 Elsevier Ltd. All rights reserved.

The overall aim of the Bologna Process is to modernize European higher education and encourage the development of a high-quality, internationally competitive system. By 2010, it will create a European higher education area (EHEA) that will:

1. facilitate mobility of students, graduates, and higher education staff;
2. prepare students for their future careers and for life as active citizens in democratic societies, and support their personal development; and
3. offer broad access to high-quality higher education, based on democratic principles and academic freedom.

The intention is to allow the diversity of national systems and universities to be maintained while the EHEA increases comparability and compatibility between systems and institutions and enhances their quality. The approach is not to mandate harmonization in the sense of imposing a single common approach, but to provide a means of connecting varying national systems.

Launched as the Bologna Process in 1999 by ministers for education and university leaders from 29 countries, by 2008, some 46 countries participated, extending from Portugal and Iceland in the west to the Russian Federation and Azerbaijan in the east (see **Table 1**). All are party to the European Cultural Convention and are committed to the goals of the EHEA. Importantly, the Bologna Process is not a European Union (EU) process: the EU does not have authority over higher education and the EHEA extends well beyond EU boundaries. Participation in the process is a voluntary decision made by each country and its higher education stakeholders, and there is no legally binding treaty or regulation. All stakeholders, including national governments, universities, employers, students, and quality agencies, are involved in the decision-making process. A key to success is the close cooperation between all parties.

The Bologna Process has stimulated sweeping reforms in higher education across much of Europe at the level of national systems and individual institutions. Due to the sheer size of the enterprise – involving 46 countries, 5600 tertiary-education institutions, and 31 million students – as well as deliberate attempts to engage with higher education outside of Europe, there is the potential, over time, to induce major changes in higher education across most of the globe.

Evolution

The Bologna Process has long historical underpinnings, with its roots in post-World War II efforts to create institutions to guarantee freedom and democracy in Europe. It builds on fundamental principles laid down in the *Bologna Magna Charta Universitatum Europaeum*. This document was signed by the heads of 430 universities in 1988 at a ceremony marking the 900th anniversary of the University of Bologna, widely recognized as the progenitor of the modern university. The Magna Charta, since signed by another 400 universities, affirmed the autonomy of universities, the essential link between teaching and research, and the transcendence of limits imposed by geographical and political boundaries.

A more immediate precursor was the Sorbonne Declaration, adopted by France, Germany, Italy, and the United Kingdom in 1998. Underpinned by a belief in the importance of education and educational cooperation in the development and strengthening of stable, peaceful, and democratic societies and universities' central role in the cultural development of Europe, the Declaration proposed the creation of an EHEA as a key means to promote the mobility and employability of citizens and the overall development of the continent.

The Bologna Declaration of 19 June 1999 reaffirmed support for the values reflected in the Magna Charta and the general principles laid down at the Sorbonne. It committed signatories to achieve, within the first decade of the third millennium, six objectives considered to be of primary relevance in establishing the EHEA and in promoting the European system of education worldwide:

1. adoption of a system of easily readable and comparable degrees;
2. adoption of a system based essentially on two cycles of higher education, undergraduate and postgraduate. The former would require at least 3 years of study and would be a qualification relevant to the European labor market; access to the second would require successful completion of the first and would lead to a master and/or doctoral degree;
3. establishment of a system of credits to indicate studies completed as a means of supporting student mobility;
4. promotion of mobility by removing obstacles to free movement of students, teachers, researchers, and administrative staff;

Table 1 Participating countries and organizations*Since 1999*

Austria
Belgium
Bulgaria
Czech Republic
Denmark
Estonia
Finland
France
Germany
Greece
Hungary
Iceland
Ireland
Italy
Latvia
Lithuania
Luxembourg
Malta
The Netherlands
Norway
Poland
Portugal
Romania
Slovak Republic
Slovenia
Spain
Sweden
Switzerland
United Kingdom

Since 2001

Croatia
Cyprus
Liechtenstein
Turkey

Since 2003

Albania
Andorra
Bosnia and Herzegovina
Holy See
Russian Federation
Serbia (and Montenegro)
The former Yugoslav Republic of Macedonia

Since 2005

Armenia
Azerbaijan
Georgia
Moldova
Ukraine

Since 2006

Montenegro

Additional member

European Commission

Consultative members

Council of Europe
UNESCO European Centre for Higher Education
European University Association
European Association of Institutions in Higher Education
European Students' Association
European Association for Quality Assurance in Higher Education
European International Pan-European Structure
BUSINESSEUROPE

5. promotion of cooperation in quality assurance with a view to develop comparable criteria and methodologies across Europe; and
6. promotion of European dimensions of higher education, particularly with respect to curriculum development, interinstitutional cooperation, mobility schemes, and integrated programs of study, training, and research.

Since the signing of the Bologna Declaration, the Bologna Process has evolved in many directions: modification and extension of the initial objectives; increases in the number of signatory countries and expansion of participants to include a number of European-level institutions; development of links with other European processes, particularly those related to vocational education and training; and efforts to build links beyond Europe.

Supporting Structures

Decision making in the Bologna Process is carried out intergovernmentally; it is not administered by the European Commission. Decisions are reached by the consensus of the ministers for higher education of the countries involved, an approach that is designed to recognize the diversity of higher education systems across Europe.

The key forum in the decision-making process is the 2-yearly ministerial review meetings, where ministers in charge of higher education meet to assess progress and determine next steps. Meetings have occurred successively in Prague (2001), Berlin (2003), Bergen (2005), London (2007), and Leuven/Louvain-la-Neuve (2009). An extraordinary summit will occur in Vienna and Budapest in 2010 to mark the culmination of the process and the creation of the EHEA.

Ministerial meetings are supported by two groups, the Bologna Follow-Up Group and the Bologna Process Board. Both are administered and supported by the Bologna Secretariat, drawn largely from the next nation to host the ministerial meeting.

The Bologna Follow-Up Group consists of ministerial representatives from all participating countries as well as representatives of other European-level organizations, including: the European Commission, European University Association (EUA), European Students' Association (ESA), European Association of Quality Assurance in Higher Education (ENQA), European Association of Institutions in Higher Education (EURASHE), the Council of Europe, the United Nations Educational, Scientific and Cultural Organisation European Centre for Higher Education (UNESCO-CEPES), the Education International Pan-European Structure and the Union of Industrial and Employers' Confederations (renamed BUSINESSEUROPE in 2007). The group's role is to follow up on the recommendations made at ministerial

meetings. It produces a work program of events, projects, and working-group activity on specific aspects of the Bologna Process.

The Bologna Board is smaller. It consists of representatives of the previous and forthcoming ministerial summits, and representatives of the previous, current, and succeeding European presidencies, the European Commission, the European-level organizations participating in the Follow-Up Group, and of two non-European countries. The board is chaired by the host of the next ministerial meeting.

In addition, numerous seminars, conferences, and meetings are convened, carrying the unofficial title of Bologna Seminars. These discuss various aspects of the Bologna Process, obstacles to implementation, and opportunities for cooperation. These occur at a variety of levels from pan-European, through national, down to institutional scales. While not formally part of the Bologna support structure, in many cases they play an important part in translating European objectives and priorities into local actions via planning and implementation at the national and institutional levels.

Key Priorities and Outcomes

After a decade of debate, negotiation, and reflections on practical experience, the Bologna Process identified ten so-called action lines required to establish an EHEA in 2010. These build on, and in some cases modify, the original six priorities:

- adoption of a system of easily readable and comparable degrees;
- adoption of a system based on three cycles of higher education, undergraduate, master, and doctoral;
- establishment of a system of credits;
- promotion of mobility;
- promotion of European cooperation in quality assurance;
- promotion of the European dimension in higher education;
- focus on lifelong learning;
- inclusion of higher education institutions and students;
- promotion of the attractiveness of the EHEA; and
- doctoral studies and the synergy between the EHEA and the European Research Area.

The Bologna Process has brought particularly significant changes in three areas, which are detailed below.

Three Cycles of Higher Education

Until the Bologna Process, every European country had its own degree structures, duration of study, and approaches to assessing student achievement. A key objective of the

Bologna Process is to converge higher education structures in Europe toward a common framework, more in line with international standards, without enforcing strict adherence to a single pattern. The complexity and diversity of national systems made it hard to attract foreign students, despite the worldwide growth in student mobility in recent decades. Giving effect to the European vision of greater mobility of students, portability of qualifications and migration of labor required new frameworks and also systems for mapping the equivalence of learning outcomes and qualifications. In 2005, ministers adopted an overarching framework for qualifications in the EHEA comprising three cycles and agreed to develop national qualifications frameworks compatible with the Europe-wide schema.

National qualifications frameworks describe the qualifications available in an education system and how they interrelate. They describe what learners should know, understand, and be able to do on the basis of a given qualification, as well as how learners can move from one qualification to the next within the national framework.

Within the European framework, the first cycle (bachelor) program should involve discipline-based study relevant to the European labor market. Access to the second (master) cycle usually requires successful completion of the first cycle and again involves discipline-based study. The bachelor degree requires 3–4 years of full-time study (or equivalent part-time) and the master's degree 1–2 years, depending on the discipline and the national context. There is no differentiation between countries with 11, 12, or 13 years of pretertiary study; whatever the duration of primary and secondary education, it is deemed to provide the general educational background needed to complete a first cycle of university study.

Many countries have made substantial changes to their degree structures in response to the Bologna Process. The first two cycles have been adopted by every country participating in the Bologna Process, sometimes in parallel with existing degrees during a transition period, sometimes replacing them completely. Introduction of new degree structures and implementation at the institutional level has required an enormous effort in reviewing and reconstructing curricula and learning outcomes. As well as widespread changes in the structure and duration of degree programs, many countries have used the reform process as an opportunity to implement a more student-focused approach and new quality procedures.

The third (doctoral) cycle was not explicitly included in the Bologna Process until the Berlin ministerial meeting in 2003. Subsequently, there has been a continuing policy-level focus on the need to build closer links between higher education and research. However, adoption of a structured approach to the third cycle has been much slower and more uneven than is the case with the first and second cycles.

New Means of Reporting Student Achievement

Despite the implementation of a common qualifications framework across Europe, a significant level of diversity will remain between and within countries. Within a context of broad adherence to the principles of the framework, such diversity is seen as highly desirable. Consequently, reliable and transparent tools are required for measuring and reporting student achievement to allow for adequate understanding and appropriate recognition of those qualifications between institutions and between countries.

In this context, the European Credit Transfer and Accumulation System (ECTS) has been developed as a standard measure of student achievement. In this system, credits reflect the total workload required to achieve the objectives of a program; those objectives are defined in terms of the learning outcomes and the competences achieved. Under the ECTS, 60 credits measure the workload of a full-time student during one academic year. The actual workload may vary between 1200 and 1800 h of various forms of study. Credits are only obtained after successful completion of the work required and appropriate assessment of the learning outcomes achieved.

The ECTS now plays an important part in curriculum design and in validating learning achievements. It makes study programs easy to comprehend and to compare for all students, and so facilitates academic recognition and student mobility. Its design reflects the desire to switch from an inputs-based approach to reporting study achievements (usually defined in terms of hours spent) to a focus on learning outcomes.

The other key tool for reporting student achievement is the Diploma Supplement. Attached to the higher education diploma (credential certificate), the Diploma Supplement describes the qualification achieved by a student in a standardized and easily understood way, as well as relating it to the higher education system and institution within which it was awarded. It provides a standardized description of the nature, level, context, content, and status of the studies successfully completed by the graduate. It is not intended as a resume/curriculum vitae or a substitute for the original credential, but rather as a means of providing more detailed information about any qualification in a way that will aid interpretation and understanding of the nature of the qualification.

Taken together, the Diploma Supplement and information on the number of ECTS credits attained reflect the achievements of a graduate. The information can be used in a systematic way to evaluate what a student has achieved and to assist in comparing those achievements with those of other students in interpreting them in other higher education contexts. It is therefore seen as a key instrument to support mobility of students and labor between diverse environments.

Quality Assurance

The term quality assurance refers, in higher education, to all the policies, review processes, and actions designed to ensure that institutions, programs, and qualifications meet specified standards of education, scholarship, and infrastructure. Quality assurance aims to guarantee and further enhance the quality of higher education provision. The Bologna Process is actively encouraging the development of a quality culture within European universities and a more systematic approach to quality assurance at the national and European levels.

Individual universities, which are primarily responsible for maintaining and improving their quality, have commonly seen Bologna-driven structural and curriculum reforms as an opportunity to reflect on their management practices, review their programs and teaching methods and, in many cases, implement more systematic approaches to quality assurance. In addition, the last decade has seen the rapid emergence of national quality-assurance systems in many countries; as a result, Europe-wide requirements for national systems have been defined to improve consistency of national quality-assurance schemes.

European standards have also been developed for internal and external quality assurance to provide institutions and quality-assurance agencies with common reference points. The resulting standards and guidelines for quality assurance in European higher education were adopted by ministers in 2005. The European Quality Assurance Register for Higher Education (EQAR) was set up in 2005 to list those agencies that operate in accordance with those standards and guidelines and the relevant national legal provisions. The overall intention is to increase transparency and confidence regarding quality assurance in higher education and therefore ultimately in higher education itself and the qualifications it provides.

Wider Dimensions

The Bologna Process of modernization and reform of course lies within a wider context that extends well beyond higher education. From a EU perspective, it is part of the Lisbon Strategy for Growth and Jobs, launched by EU member states in 2000 and relaunched in 2005. The overall intention of the strategy is to enable Europe to maintain and increase its prosperity and therefore to preserve and enhance its social models. To achieve this, the European Council called for the EU to become the most competitive and knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion. Key elements of the strategy include budgetary sustainability, better regulation, improved tax and benefit systems, and

improvements in education and training systems, and investment in research to maintain comparative advantage over other parts of the world. Achieving the ambitious objectives was seen to require a radical transformation of the European economy and a challenging program for the modernization of the social welfare and education systems, including higher education.

The Lisbon Strategy also involves strengthened cooperation in vocational education and training, the so-called Copenhagen Process. The European Commission has sought to establish synergies between the Bologna and Copenhagen processes, for example, via the development of the European Qualifications Framework (EQF) for lifelong learning and vocational education and training equivalents of the ECTS and ENQA.

A continuing theme in the Bologna Process has been the need to build closer links between higher education and research. The European Commission has sought to develop close links between the EHEA and the European Research Area (ERA). Unlike the EHEA, the ERA is an EU initiative, taken in 2000, to create a unified area across Europe in which researchers can move and interact seamlessly, build and benefit from world-class research infrastructure, and work with excellent networks of research institutions. As research institutions, universities are closely involved, particularly through their third-cycle doctoral programs. With its focus on investment in research, the Lisbon Strategy builds intimately upon the ERA.

Beyond Europe

A key objective of the Bologna Declaration was to increase the international competitiveness of the European higher education system. A continuing theme of the Bologna Process has been to ensure that European higher education becomes an attractive worldwide destination for students. The Bergen Communique from the 2005 ministerial meeting identified the EHEA as a partner to higher education systems in other parts of the world, stimulating balanced student and staff exchange and cooperation between higher education institutions. It is therefore no surprise that the Bologna Process has increasingly engaged with, and impacted upon, higher education beyond Europe.

The Bologna Process and development of the EHEA has prompted discussion between European and international partners on a range of higher education policy issues and areas of potential cooperation, particularly quality assurance and mutual recognition of qualifications. To provide a framework for cooperation at European, national, and institutional levels, in 2007, ministers adopted a strategy entitled *The European Higher Education Area in a Global Setting*. This strategy incorporates five priorities:

1. improving information on the EHEA;
2. promoting European higher education to enhance its worldwide competitiveness and attractiveness;
3. intensifying policy dialog;
4. strengthening cooperation based on partnership; and
5. furthering the recognition of qualifications.

The EHEA has thus opened dialog with educators and educational administrators in many parts of the world, but there are widely varying understandings of and responses to the Bologna Process.

There is momentum for increased investment in education, educational reform, and a degree of alignment across international boundaries in the Asia-Pacific region. In 2006, 27 ministers for education met in Brisbane (Australia) in 2006 to discuss how to respond to the challenges of the Bologna Process and create stronger regional links. The resulting Brisbane Communique laid down Bologna-like objectives such as increasing student and academic mobility, transferability of qualifications, and greater interchangeability or integration of education frameworks. Ministers agreed to collaborate on quality-assurance frameworks, recognition of qualifications, common competency-based standards for teachers, and mechanisms for the common recognition of technical skills. Moreover, Bologna-like, they agreed to establish biennial ministerial meetings that will consider and plan for educational issues throughout the Asia-Pacific region. Subsequently, South-East Asian ministers for education adopted a long-term project titled *Raising Awareness: Exploring Ideas of Creating a Higher Education Common Space in South-East Asia*.

Some countries are beginning to adopt specific Bologna-inspired features. For example, in 2008, the Australian government announced its intention to provide an Australian Higher Education Graduation Statement for all graduates, a document similar in purpose and content to the European Diploma Supplement. Elsewhere, for example, in parts of Africa and South America, attempts are being made to align higher education with elements of the Bologna Process.

Perhaps at the other extreme, knowledge of and official responses to Bologna in the United States were very limited during the George W. Bush administration. While all indicators demonstrate that the US accounts for the majority of the world's best universities and it still attracts a large proportion of the world's best scholars, there are concerns that its dominant position is under threat. For example, while it clearly remains the leader in research, its share of international students has fallen, while that of Europe has risen sharply since 2000. Nevertheless, growing concern about the standing of American higher education has not translated into national-policy responses to the Bologna Process. For example, the 2006 final report of the Secretary of Education's Commission on the Future of Higher Education (the *Spellings Report*) made little mention of international trends in higher

education and no direct reference to the Bologna Process or the EHEA.

Despite this limited attention at the level of national policy, influential educators have begun to argue that the United States should pay more attention to the implications of the Bologna Process and perhaps adopt some of its features. For example, given the highly diversified system of higher education in the United States, it has been suggested that there should be efforts to build common understandings of what bachelors degrees represent in terms of learning outcomes and skills developments. Others have pointed to the need to come to grips with the implications of the three-cycle system if the United States is to attract a share of the most able postgraduate students from Europe.

European Futures

As the Bologna Process moves toward its target date of 2010, three outcomes are becoming clear. First, and most obviously, there have been very substantial changes in the structure of higher education and the surrounding policies and processes. Many countries have made substantial progress toward implementing two- or three-cycle models consistent with the Bologna model; quality-assurance processes and systems have been more widely embedded. Second, there has been widespread curriculum reform with an increasing emphasis on defining explicit learning outcomes. Third, the objective of improving the quality and attractiveness of European higher education seems to be at least partly achieved, as reflected in the EHEA's growing share of in-bound international student movement.

However, it is important to note that there remains substantial diversity in both the nature and speed of change in all these dimensions. Besides, in many cases there remains a contrast between the adoption of Bologna-compliant legal and policy frameworks and limited changes in educational practice or the experience of individual students. There is an enormous complexity of detail, which can be tracked through annual reports provided by each participating nation.

Much will remain to be done to complete implementation of the ten action lines after 2010. A continuing theme of particular importance is likely to be the embedding of more sophisticated and robust quality-assurance processes as a basis for building confidence in and understanding of educational qualifications that are essential to underpin the goal of increasing the mobility of students, staff, and labor.

Nevertheless, there will clearly be progressively less emphasis on structural and process change. It is expected that a new agenda will be defined at the 2009 ministerial meeting in Leuven. It seems likely that three emphases

will dominate the forthcoming decade. The emergent emphasis on the learning process and the creation of a student-focused, rather than teacher-driven, education is likely to become a dominant feature.

Second, there will probably be a renewed focus on the purposes of higher education and its role in both individual personal development and societal change. In the wake of the global economic crisis of 2008–2009, it is almost certain that this will involve a reinforced focus on skills, employment, and economic growth, although there will be at least lip service to the long-held belief in the value of education as a key underpinning of democracy and freedom.

A third and final element will certainly be a renewed emphasis on the external dimension. As a minimum, there is little doubt that the EHEA will strengthen its programs for reaching out to other nations and regions to increase global awareness of the Bologna Process, build partnerships, and increase mutually beneficial cooperation. It is also quite likely that the next decade will see increasing competition between three global powers in higher education and research, the EHEA, the United States, and the Asia-Pacific, dominated by China and India.

See also: Curriculum Reform; Higher Education: An Overview; Internationalization of Higher Education; Quality Assurance in Higher Education – Practices and Issues; Student and Faculty Transnational Mobility in Higher Education.

Further Reading

- Adelman, C. (2008). *The Bologna Club: What U.S. Higher Education Can Learn from a Decade of European Reconstruction*. Washington, DC: Institute for Higher Education Policy.
- Australian Government (2006). *The Bologna Process and Australia: Next Steps*. Canberra: Department of Education, Science and Training.
- Austrian Federal Ministry of Science and Research (undated). *The European Higher Education Area*. www.ond.vlaanderen.be (accessed August 2009).
- Bologna Follow-up Group Stocktaking Working Group 2005–2007 (2007). *Bologna Process Stocktaking Report London 2007*. London: Department for Education and Skills. http://www.eurashe.eu/FileLib/maindocs/bologna/bologna_process_stocktaking_london_2007.pdf (accessed August 2009).
- Corbett, A. (2006). Higher education as a form of European integration: How novel is the Bologna Process? *Working Paper No. 15*. Oslo: University of Oslo, Centre for European Studies. www.arena.uio.no (accessed August 2009).
- Crosier, D., Purser, L., and Smidt, H. (2007). Trends V: Universities shaping the European higher education area. *An EUA Report*. Brussels: European Universities Association.
- Europe Unit (2007). *Guide to the Bologna Process*, 2nd edn. London: United Kingdom Higher Education Europe Unit. www.europeunit.ac.uk (accessed August 2009).
- European Higher Education Area (1999). *The Bologna Declaration of 19 June 1999*. http://www.bologna-bergen2005.no/Docs/00-Main_doc/990719BOLOGNA_DECLARATION.PDF (accessed August 2009).
- European Universities Association (2008). *Europe's New Higher Education Landscape*. Brussels: European Universities Association.

Marginson, S. and van der Wende, M. (2007). Globalisation and higher education. *OECD Education Working Paper No. 8*, Paris: OECD.
Zgaga, P. (2006). *Looking Out: The Bologna Process in a Global Setting. On the 'External Dimension' of the Bologna Process*. Oslo: Norwegian Ministry of Education and Research.

Relevant Websites

<http://www.ond.vlaanderen.be> – Bologna Process, London 2007-Benelux 2009.
<http://www.buinesseurope.eu> – BUSINESSEUROPE.
<http://www.coe.int> – Council of Europe.
<http://www.eurashe.be> – European Association of Institutions in Higher Education.

<http://www.enqa.eu> – European Association for Quality Assurance in Higher Education.
<http://ec.europa.eu/education> – European Commission Education and Training.
<http://www.ei-ie.org/europe/en/index.php> – Education International Europe.
<http://www.esib.org> – European Students Union.
<http://www.eua.be> – European Universities Association.
<http://www.heacademy.ac.uk> – Higher Education Academy.
<http://www.magna-charta.org> – Magna Charta Observatory of Fundamental University Values and Rights.
<http://www.cepes.ro> – UNESCO European Centre for Higher Education.
<http://www.europeunit.ac.uk> – United Kingdom Higher Education Europe Unit.

The Economics and Finance of Higher Education

A Wolf, King's College, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Agency theory – Deals with situations where a principal hires or depends on an agent, who may have different interests, and where there is imperfect (or asymmetric) information: typically the agent knows things the principal does not or cannot know.

Economies of scale – Occur when, the more you produce, the cheaper it is to produce any one unit.

Positive externalities – Exist when individuals and firms making decisions or selling/buying goods and services do not get the full benefit of the decision. Some of it is received by others. (Examples include beekeepers who pollinate others' field, people ensuring their children receive vaccinations, and creators of beautiful private gardens on view from the public highway.)

Public choice theory – Studies the behavior of politicians and government officials and employees using the theoretical apparatus of economics to explain their actions, and using the same assumptions of rational self-interest.

A Public good – One where consumption by one individual does not 'use up' the good (it remains available for others) and where you cannot exclude people from using it and preserve it only for those who pay. Examples include the rule of law, and national defence forces.

Skill-biased technological change – A change in productive technology and techniques that favors and increases the demand for highly skilled workers.

The economics of higher education is normally understood to cover both analyses of higher education's impact on the wider economic system, and the use of economic theories and methods of analysis to examine the higher education system as a system: for example, the demand and supply of higher education, the efficiency of higher education institutions, and the drivers of innovation within the system. This sounds like a potentially enormous field of enquiry, and indeed it is. In practice, however, academic work in this area has been concentrated in rather a limited number of areas.

Far and away the most important current area of study, at least in terms of volume of output, is concerned with the economic impact of higher education on people's earnings and on the economy. This work is generally

approached from the perspective of human capital theory. As recently as the previous edition of this encyclopedia, it was possible for the writer of the economics of higher education article to refer to human capital theory as highly controversial. Today, it is the overwhelming orthodoxy of both academic economics and political policy-making worldwide.

Its influence derives largely from the work of Gary Becker, the Nobel-prize-winning economist, although the basic idea is very old. Using the normal definition of capital as something which is not entirely used up in the production period under examination, but remains as a source of future production, human capital is seen as something embodied in individuals which contributes to production and economic growth. It is created through "activities that influence *future* monetary and psychic income by increasing the resources in people. These activities are called investments in human capital" (Becker, 1993: 11, emphasis mine). Discussion of higher education's impact on the economy is now carried out overwhelmingly from this perspective, and is the topic of this article's opening section.

It is followed by a discussion of economic work related to the efficiency of higher education. Efficiency is a core topic for economics but research is extremely difficult to carry out in this area because of the intrinsic problems in measuring higher education's output: namely, that universities are multioutput organizations, that there is no agreed relative importance attached to different outputs by society, and that only some outputs are measurable across the system as a whole. The volume of publications in recent years has been correspondingly small, but includes some very interesting empirical work. We then turn to two emerging fields: the application of economic theories of organization to higher education, and work using concepts from the economics of information. The final section addresses the financing of universities. This is largely and of necessity descriptive, although good theoretically based work on what determines public, and not just private, willingness to pay for education would be very useful in a rapidly expanding and changing sector.

The Economic Impact of Higher Education

Work on the economic impact of higher education is, as noted above, dominated by work which examines this in terms of human capital, built up by individuals during

their studies, and used across a lifetime of work and employment. Much of this work deals with rates of return: specifically, how much more people earn as a result of attending university, and how much this contributes to productivity, growth rates, and prosperity.

Rates of Return

In volume terms, the economics of higher education is dominated by empirical studies of the individual rate of return to years in college and to formal degrees: that is, the impact of university attendance on the individual's future earnings. This is part of a more general growth in such studies, which use large data sets on individual earnings and prior education to examine the financial benefits of formal education, and cover an ever-expanding list of countries. This literature, termed a veritable tsunami by Krueger and Lindhal (1999), is reviewed by Psacharopoulos and Patrinos (2004), who list over 100 studies, and cover over 80 countries in what they emphasize to be a constantly expanding field.

The estimates which can be made with some confidence are of private returns: namely, how much more individuals earn as a result of an additional year of education, compared with those who are otherwise like them (and allowing for costs borne directly by the individual plus income foregone as a result of studying rather than being at work). Such private returns are almost universally highly positive, but the absolute levels vary by country and level of education (primary/secondary/higher). In some parts of the world, primary education confers much higher rates of return than higher education, in others, the opposite. Returns are usually higher for poorer/developing countries than for wealthier ones. The status of the university attended, age at graduation, and the subject studied all affect the result (see e.g., Grogger and Eide, 1995; Egerton, 2000; Egerton and Parry, 2001; Hoxby, 2004; Zhang, 2005). However, Hamermesh and Donald (2004) indicate that the differences associated with degree subject chosen may be overstated, since that choice is itself highly correlated with other factors which influence later earnings (such as prior ability and hours worked once employed).

In the Organization for Economic Cooperation and Development (OECD) countries, private returns to primary education have historically averaged out at slightly above those for secondary and higher education. However, with primary and lower-secondary education now universal, the question of most interest to policymakers has been the extent to which returns to higher education remain high in a context of ever-expanding enrolments. The question is important because very few university students, anywhere in the world, pay the full cost of their education, with tuition fees, in particular, being highly subsidized by most governments.

The high private returns to college and university are themselves enough to explain the huge increases in enrolments which have taken place in recent years (Kosters, 1999; Wolf, 2002). Other things being equal, one would expect that an increased supply of graduates would tend to lower their average price (and hence the private returns to be gained), compared with other parts of the labor force. The composition of the job market has changed in recent decades, so in that sense, things are not equal: developed countries register a marked increase in the proportion of technical and professional jobs, and a sharp fall in the proportion of skilled manual jobs. However, studies also conclude consistently that there is a good deal of over-education or graduate underemployment, with graduates employed in jobs for which their level of education is not a necessity (see, e.g., Patrinos, 1997; Alpin *et al.*, 1998; Battu *et al.*, 2000; Green *et al.*, 2002; Kler, 2005; Duru-Bellat, 2006). While most detailed studies of over-education employ data sets from developed countries, the phenomenon is also common and highly visible in many developing nations (see, e.g., Dore, 1997).

One of the major puzzles in this field is, therefore, the continuing high level of private returns to higher education and, even more specifically, the fact that, in developed countries, these have tended to grow, relative to other forms of education and, especially, compared with an education which does not include a formal upper-secondary certificate (e.g., a North American high school diploma, or a baccalaureate). Discussion of this phenomenon involves, inevitably, the debate over more general trends in income distribution, including increases in earned income inequality which were especially marked in some countries (notably the United States and United Kingdom in the 1990s) but generally widespread (Atkinson, 2003). The current consensus – for which returns to university education provide some of the evidence – is that the main driver has been skill-biased technological change, and it is this rather than, for example, outsourcing of jobs to developing countries or widespread low-skilled immigration that has kept the wages of the less educated depressed relative to those of graduates (Heckman *et al.*, 1998; Moore and Ranjan, 2005).

Although there can be little doubt that private returns to a university education are high, calculating the precise impact of that education is extremely difficult, and the subject of a large literature in itself (Card, 2001; Blundell *et al.*, 2005). One problem is that those who enter higher education may differ from their peers in ways over and above the simple fact of university study. If they are more able, or harder workers, later returns to education may in fact be rewarding skills and effort as well, or instead. Data sets differ in the degree to which they allow controls for other variables; in general, the finding is that, while returns decrease somewhat with controls for ability, they remain significant. Randomized or controlled experiments are, of

course, impossible, and natural ones very rare at this level of education. One such occurred in France, however, when the 1968 student riots resulted in many formal examinations being shelved. The result was that a number of individuals entered university who in the previous or later periods would not have done so. Maurin and McNally (2005) demonstrate that the benefits were significant and strong for this marginal group – the wage returns to a year of university study are, for them, about 14% – and that the benefits were also passed onto their children, who were in turn more likely to enter higher education than if their parents had not done so (but been otherwise unchanged).

Another problem, which a natural experiment such as this cannot resolve, is the extent to which graduates are being rewarded for substantive skills gained at university. As Spence argued in his classic work on signaling (Spence, 1973), formal qualifications may be used to signal underlying qualities, such as innate ability, hard work, and perseverance, rather than as evidence of specific, valued skills learned on the courses which the qualification accredits, and may be used by employers for general screening purposes. It is extremely difficult to estimate how much qualifications (of any sort) operate as signals of this type, and how much their associated wage premia are a reflection of specific acquired skills. The evidence suggests that both factors are important, but in varying proportions depending on time, place, subject studied, etc. (Cameron and Heckman, 1993; Murnane *et al.*, 1999; Wolf, 2004).

The difficulty of estimating how much the private returns to higher education are returns to new and valuable skills complicates the question of how much there is a social, as well as private, return to university education; and, therefore, how much governments should subsidize higher education. Funding of higher education – both teaching and research – is a major item in the budget of almost every national government in the world, and, as discussed further below, has been under increasing strain in recent years. The dominant rationale for these subsidies is economic: more so, indeed, than in previous centuries, when social, cultural, and religious arguments were relatively more prominent. Reflecting the influence of Becker (1964), governments have taken on an explicit responsibility for developing their countries' human capital.

This implies that – on economic grounds – public funding and subsidy for higher education should be undertaken up to, but not beyond, the point at which the social return on the investment reaches the cutoff level conventionally adopted by the country's finance ministry when evaluating public investment opportunities. However, the social returns are very hard to calculate (Wolf, 2002; Psacharopoulos and Patrinos, 2004). Conventionally, they simply involve adjusting the private rate to allow for the full cost of the education, and not just that part which the student pays. The result is assumed to reflect

the benefit to society, on the grounds that individuals' increased earnings reflect their increased productivity as a result of their education. However, many graduate wages, especially in the public sector, are not set by market forces and so may not accurately reflect additions to economic output; and any part of the private return which is determined by signaling rather than skills is similarly problematic.

Externalities

Most of the economic literature on university outcomes emphasizes private returns, and their presumed reflection of individuals' increased skills (human capital), with social returns treated as simply a more fully costed version of these. However, economics also suggests that higher education may influence the economy in rather less-obvious ways. The most important economic argument advanced to justify the public funding of university research, especially in science, is that knowledge should be treated like a public good, available to all. (If a private good is consumed by someone, then it cannot also be consumed by others: we cannot all eat the same apple or own and occupy the same house. Knowledge produced by universities and in the public domain remains there for others to use; however, many other people have used it beforehand. Knowledge is not, however, naturally a public good, unlike, say, defense, law and order, street lighting, and public drains.)

The alternative, which is to have private companies and individuals finance research, and own its results, is likely to be highly suboptimal, because it means that the owners monopolize information to their own benefit and (excess) profit, and militates against continuing innovation and scientific progress, because of the inaccessibility of important findings (Aranson, 1995). While this is difficult to prove, it is noteworthy that developed countries rarely are successful in industrial sectors where their universities are weak in the corresponding underlying sciences (Porter *et al.*, 2000; but see also Kealey, 2008).

A related argument is that there are general spillover effects (positive externalities) from having a highly educated workforce with many graduates. The productivity effects will be greater than the sum of the individual parts because they will have a generally beneficial effect on the performance of co-workers and on technological progress. This belief has underpinned the recent enthusiasm of governments for increased public subsidies for higher education, and for raising the proportion of a cohort graduating (both absolutely and as in relation to other OECD countries). However, the supporting economic evidence is slim at present. At a global level, higher education participation levels do not seem to have any consistent, positive relationship with wealth or growth (Pritchett, 2001; Wolf, 2004; Lewis, 2004). A review conducted for a major UK government review of higher

education in the 1990s found no clear evidence of externalities (Gemmell, 1997). More recently, however, there have been some studies which indicate genuine cluster or spillover effects, not only for industries operating at the technological frontier, which benefit from closeness to universities, but also at the within-plant level (Audretsch, *et al.*, 2003; Moretti, 2004a, 2004b).

Although the relationship between higher education and wealth continues to dominate the literature, a number of other themes have also emerged, or been revisited, in recent years. These can be grouped broadly under the headings: (1) studies of institutional efficiency; (2) applications of general theories of economic organization to the higher education sector; and (3) analyses of student and other consumer behavior in terms of information search and related transaction costs.

Education Production Functions: Efficiency and Productivity

The growing size and expense of universities has created corresponding demands from government funders for efficiency gains, and for accountability and performance indicators. The fact that universities have multiple outputs, and the difficulty of measuring any of these outputs with any precision, means that economics is unable, with any ease, to examine the efficiency of universities, in the sense of their ability to generate given output levels with more or fewer resources. In the school sector, externally set assessments often are available as a more or less comparable, albeit imperfect, measure of outputs at the within-country level; but university degrees and grades cannot be treated as standardized in this way, and there is also no agreement on how best to measure research outputs.

A good deal of the rather limited literature in this area looks at either teaching, or research, in isolation. In the case of teaching, one possible approach is to look at changes in graduation rates over time, and relate them to resources, although this assumes that grading standards remain constant at the within-institution level. Bound and Turner (2006) use a very large US data set to argue, on this basis, that lower levels of resource per student (using levels of public subsidy per student as proxy) have a direct and significant effect on graduation rates: that is, that reduced resources per student reduce the likelihood of graduation.

However, this finding needs to be evaluated with care. An extensive literature on the impact of resources on primary and secondary schools has underscored the complexity of the relationship (see e.g., Hanushek, 2007; Chevalier *et al.*, 2005). There appears to be no clear, let alone linear, relationship between levels, or changes, in total resource per pupil and learning outcomes, even when the measure of resource accurately reflects spending per pupil at the classroom, rather than the system, level. It seems unlikely that

universities (where large amounts are spent in ways which have no direct impact on students' classroom experiences) will display a simple link between resources and graduation rates, let alone resources and substantive learning. For example, Andras *et al.* (2007) argue that the productivity of science faculties in UK universities (in terms of quantity and quality) increased between 1975 and 2004 relative to the United States, even though the resource gap was actually widening during this same period.

Studies at a more micro-level, where learning outcomes can be treated as standard, are potentially more fruitful, although the issue of what makes for best practice in teaching, and the efficient organization of teaching, remains little understood, and little researched by economists. One exception is a study by Bettinger and Long (2004) of standardized college remediation courses, which provide assistance to students whose skills, on entry, are deemed inadequate for coping fully with the college curriculum. Such courses are controversial. They not only may assist students in coping with regular courses, but may also discourage them from persevering at all. The research indicates that both outcomes exist – those who complete remediation are more likely to persevere in college than otherwise comparable students, but being placed in such a course increases the likelihood that the student will drop out or transfer.

Another recent study using detailed administrative data from a Canadian university indicates that, among first-year students taking large courses (with multiple instructors but common assessments), instructors have a small but significant effect on students' grade, likelihood of dropping out and likelihood of taking the same (elective) subject in later years of the degree (Hoffman and Oreopolous, 2006). The effects are not attributable to instructors' rank or salary, but are positively correlated with students' pooled subjective rankings of instructor quality. Such studies have the potential to provide findings which can be applied directly to instructional practice, and it is to be hoped that, in spite of the difficulty, research of this type will increase in future years.

There have also been some innovative attempts to examine whether different institutions are more or less efficient in producing research outputs, notably in the sciences and economics, where international citation indices provide an acceptable measure of quality. Kim *et al.* (2006) use data from economics and finance faculties in over 800 universities to demonstrate that, while absolute levels of high-status publications are higher in elite universities, the latter do not seem to add value to academics' performance. In the 1970s and 1980s, being at a top university generally had a positive impact on individuals' productivity. Today, young scholars in these disciplines are publishing high-impact papers without having to be in elite institutions when they do so. Conversely, when and if their publications win them better jobs, the move does not

affect their performance; the fact of being in an elite institution, among established or other rising stars does not have any obvious positive effect. This may, however, be because the move does not actually change who they collaborate with. Over the period studied, the authors also show a big increase in the number of co-authored papers involving both elite and nonelite institutions, which would be consistent with this interpretation.

At the sectoral level, the measurement problems associated with higher education's multiple outcomes and objectives remain formidable. A number of studies have used data envelopment analysis (DEA) or multiproduct cost function to address the issue. DEA examines how far, within a group of similar organizations, individual organizations diverge from best-practice performance. (The technique requires that some units be identified as representing best practice, for comparison purposes, even though none of them may in fact be operating at anything close to the best-practice frontier.) These studies are concerned both with technical efficiency, in the sense of asking whether institutions could generate more with given resources; and with scale efficiency, or whether institutions might be more efficient if they were smaller, or larger, because returns to scale are not constant. Abbott and Doucouliagos (2003) find, on this basis, that there are few major differences in the operating efficiency of Australian universities. Johnes and Johnes (1995) apply the technique to an analysis of the technical efficiency of UK university economics departments as producers of research. They argue that the technique not only has a positive contribution to make, but also underline how susceptible the results are to the choice of inputs and outputs.

Other researchers examine economies of both scale and scope, looking at teaching and research, although again the number of studies is small. Dundar and Lewis (1995) used peer ratings of graduate program quality, in conjunction with quantitative measures of student credit-hours and publications to look at economies of both scale and scope; and their results give clearer indications of the existence of both, in American public research universities, than most previous studies. Hashimoto and Cohn (1997) similarly look at undergraduate and graduate teaching plus research, and find evidence of ray economies of scale, and both global and product-specific economies of scope, in Japanese private universities.

While the evidence of returns to scale may have some important policy implications, there are, as Cohn *et al.* (1989) emphasize, major problems associated with any attempt to analyze multiproduct organizations, and especially when – as in higher education – the measurement of both the teaching and the research outputs is highly problematic, and there is no agreed weighting for the relative importance of teaching, pure research, and applied research/consultancy as outputs. This has not prevented governments from pursuing measures of quality and productivity which can be placed on a scale and so used for

accountability and funding purposes. The UK government, for example, wishes to move to a metrics-based approach to distributing its research funding for universities, even though empirical research indicates that quite small changes in the measures used, and the weightings they are given, can lead to major changes in both scores and rankings (Sastry and Bekhradnia, 2006).

The Organization of Higher Education

Classical economic theory has concerned itself in large part with the way in which resources are allocated through price mechanisms in more or less competitive markets. Higher education, however, has for many years been an overwhelmingly publicly funded, and, in many cases, publicly run activity in which governments determine the number of students, provide most or all of the expenditure for the sector, and allocate it between institutions which are themselves generally directly created by, or licensed by, the state. This has been the case even in the United States, which has a large private not-for-profit higher education sector, as well as many proprietary schools.

This is likely to change in the future, because of the rapid growth of private institutions, notably in ex-Communist and developing countries, and the emergence of a successful for-profit sector in the United States. The circumstances under which such institutions emerge and succeed, alongside subsidized public-sector provision; the types of degree they offer; their pricing strategies, and the often highly segmented nature of their target student populations, would all appear to be areas in which economics should be able to provide useful insights (Ortmann, 2001; Kraft and Vodopivec, 2003; Tooley and Stanfield, 2007). To date, however, the published literature in this area is still small.

Economics has paid rather little attention to the forces determining levels of public subsidy across time and jurisdictions (though see Carpentier, 2005). However, the analysis of how university staff and students' behavior is affected by different funding and incentive patterns has been addressed, intermittently, by economists, most of whom draw more or less explicitly on the pioneering analyses of Smith (1776/1976). Smith argued that the reason why the quality of teaching was low in the English universities of his day, as compared with the Scottish, was that in the former, faculty lived off endowments while in the latter they had to attract fee-paying students. Later economic analyses (Buchanan and Devletoglou, 1970; Garvin, 1980; West, 1995; Raines and Leathers, 2003) have advanced similar arguments which draw broadly on public choice or agency theory: students for whom tuition is free will act as uninformed consumers, while faculty whose jobs are paid for by block grants from state revenues will prioritize their own interests (maximize their own utilities) rather than the objectives of students or, indeed,

the funding governments, especially when protected by tenure, and not-for-profit governance structures. For many, this will involve prioritizing research and related activities which bolster their own, and their institutions', prestige, and, therefore, their future career opportunities.

This analysis would lead one to expect changes in recent years, as students' fees have come to account for rapidly growing proportions of university expenditure in many countries. This has indeed been predicted, but Zemsky *et al.* (2005) find less change than expected. They ascribe this to students being most interested in their diploma's reputation, or brand, rather than what they actually learn; and therefore having interests which are more aligned with those of faculty than has generally been assumed.

This approach to understanding universities has also been elaborated using theories of property rights. Meiners and Staaf, for example, note the wide-ranging and recurrent criticisms made of higher education in the United States, and argue that most are "implicit normative arguments for changing property rights" (Meiners and Staaf, 1995: 197). As already noted, a recurrent theme in both academic analysis and public comment on universities is the importance of research, and research publications, for individuals' career success, and the consequent devaluing of teaching, and especially undergraduate teaching (see, e.g., Ehrenberg *et al.*, 2003). The property rights approach elaborates on the traditional economists' explanation by noting that the value of research outputs is relatively easily measurable, and that a good deal of it can be captured by the researcher directly. For successful researchers, it adds directly to their opportunity cost; the salary they could obtain by moving to another institution, forego by staying where they are, and which is in fact therefore very likely to be paid in full or almost in full by their current institution, which generally does not wish to lose them. (Good researchers bring in grants, which are shared with the employing institutions, and increase a university's reputation, which in turn attracts both students and other faculty.)

Teaching, by contrast, is successful to the degree that students learn. As noted earlier, compared with primary and secondary education, very few data are available on the extent to which higher education teacher quality affects learning. However, whatever the impact of the teacher, learning requires joint inputs, by students as well as by teachers. It is, in practice, impossible to calculate routinely the degree to which it is the teacher's input rather than the student's effort which matters; moreover, while students who learn effectively because they are well taught may later earn much more as a result, the teacher concerned has no way of capturing any of this gain for her or himself. (If a given institution is effective in raising students' later incomes more than others in the same system, it may be able to raise tuition fees; but these benefits will accrue to the institution as a whole, and,

while they may lead to higher average salaries, cannot be captured by individually excellent teachers.) The tension between research and teaching is thus endemic in higher education, and there is little or no market for good teachers.

A more general principal-agent perspective is used by a number of analysts to explain the relative importance given to teaching or research in different institutions and systems, and the effort expended by academic staff on either or both. Changes in the financing of higher education have provided an opportunity to see whether universities respond in the directions predicted within this framework (Slaughter and Leslie, 1997; Webster and Etzkowitz, 1998; St. John and Parsons, 2004). These changes are discussed in more detail below, but may be summarized briefly as involving a reduction in the proportion of funding received in the form of governmental grants, and increases in the importance of fee income, research grants, and contracts.

These changes have been characterized by Slaughter and Leslie as creating a new form of academic capitalism, by which they mean that universities, including public universities, are now directly involved, at all levels, in attempting to secure external monies in a competitive marketplace, and do so on the basis of whether or not given activities are likely to be profitable. Countries greatly vary in the extent to which this model reflects their current form of organization; the United States can be seen as one end of the spectrum, whereas in some other countries, developed and developing, there has been rather little such development. (Italy is a case in point.) However, movement is generally in the direction of this model (CERI, 2004).

The effects vary as a result not only of the relative size of the different income streams, but also of the way that universities are organized internally, and the extent to which this also involves market-type mechanisms, with individual departments and staff benefiting directly from successful income generation.

The most obviously market-oriented and entrepreneurial activities are those which involve direct collaboration between academics and industries, including cases where universities launch their own start-ups, and these have been extensively studied (Etzkowitz, 2007). Webster and Etzkowitz (1998) point out that a situation in which universities generate income by capitalizing the knowledge they produce involves a genuine structural change in the way higher education is financed. Although the amount of money raised in this way remains relatively small compared with tuition fees, research grants, and block grants, it changes incentives for academics in ways which are highly discipline dependent, and also has led to far more differentiated contracts of employment for entrepreneurial staff (Slaughter and Leslie, 1997). For example, among successful faculty in the relevant fields – mostly

science and engineering – many work largely within centers or institutes which are as much engaged with commercial as with academic cultures and networks, have little engagement with departmental duties or teaching, and negotiate terms and conditions directly with central university administrators.

The impact of other changes in university financing very strongly depends on the internal organization of universities. It might appear that a growing dependence on tuition income would increase the influence of the student body, and go some way to redress the imbalance between incentives to do research and incentives to prioritize teaching. However, in some systems (including the United States), the internal allocation of both block grant and tuition revenues is often the prerogative of central administrations, and resources do not follow students in any formula-based fashion. Instead, for example, undergraduate fees may be used to subsidize graduate teaching; and science and engineering infrastructure may be supported at levels which cannot be covered from funds dedicated to those areas (Ehrenberg *et al.*, 2003).

In contrast, in some other systems – notably the United Kingdom and Australia – a large proportion of fee income is commonly devolved directly to the departments which recruit the students, creating a tuition–profit-sharing system. Both these countries, in common with many others, operate with a dual-fee policy, in which home students and overseas students pay very different levels of fee (with home students, but not overseas ones, also attracting matching funds from government). This creates strong incentives for departments to attract and recruit students, especially from lucrative overseas markets, and to concern themselves with their reputation as teaching, as well as research, institutions. Both the United Kingdom and Australia (and New Zealand) have become highly dependent on overseas fee income, and have developed qualifications – notably the 1-year master's degree – in direct response to these financial incentives (CERI, 2004). The ways in which new departmental incentives to recruit students and maintain teaching quality interact with continuing individual incentives to prioritize research and consultancy have not, however, been studied empirically in any detail to date.

Information Issues

A major focus of much contemporary economic analysis, especially in fields relating to public-sector provision, is information: its accessibility, for consumers (and other principals); the costs of obtaining it; the degree to which consumers engage in rational information search; and the extent to which people depend on particular signals (including brand names) as a way of managing the costs of information search.

Access to and use of information about quality and outcomes are acknowledged to be important issues in higher education, as are the strategies which institutions use to establish and maintain reputation in an increasingly global market (De Burgh *et al.*, 2007; Keep and Mayhew, 2004; Zemsky *et al.*, 2005). The growing size of the university sector, both within countries, and worldwide, and of the increasing numbers of students studying outside their own countries, means that universities which are competing for students and faculty must establish name recognition; something which was much less problematic at a time when students had very little choice of institution and studied almost entirely within their countries of birth.

One result of this is the emergence of a winner-take-all syndrome, echoed in other spheres such as sport and the arts, in which a few institutions which have an established reputation for very high quality are able to command increasingly high relative prices from fee-paying students, and greater, and increasing, shares of renowned faculty and research revenue (Frank and Cook, 1995; Wolf, 2002). Competition for entry into these institutions is commensurately intense, with applicants using both traditional and new ways of distinguishing themselves from each other, especially in systems where the application process is not tightly controlled by government (Bakker and Wolf, 2001). Another consequence is that employers more or less consciously sort institutions and qualifications into a very limited number of categories for selection and screening purposes (Jenkins and Wolf, 2005).

The number of higher education institutions which can become global winners is limited by the costs and difficulty of processing large amounts of information. For example, the international league tables showing university rankings, which have become increasingly numerous and widely reported in recent years, sometimes include several hundred institutions – by itself a tiny fraction of the worldwide number. Most reports, however, focus on the top 50, 100, or at most 200 institutions.

Developments at the elite end of the sector demonstrate that some students are engaged in very active information search leading to discriminating application decisions. It is less clear, *a priori*, whether such behavior characterizes the bulk of applications for university. A large body of work has questioned whether the concept of rational, utility-maximizing consumers is consistent with empirical observations of human behavior; for example, people often appear to use discount rates in making decisions which are very different from market rates, and researchers have argued that young people from poorer and less-educated backgrounds do not act as informed rational consumers in making decisions about education and training (Chapman, 1996; Ball *et al.*, 1999). A small number of detailed studies now exist of information search and university applications, and their relation to both students' labor-market expectations and their expectations of obtaining admission

to different institutions. These indicate that, for university entrance, young people's behavior is in fact largely, although not entirely, consistent with the type of information search and application strategies that would be predicted by traditional economic theories of utility-maximizing behavior (Cam, 2001; Rochat and Demeulemeester, 2001; Menon, 2004; Varga, 2006).

Financing Higher Education

For most of the twentieth century, higher education in developed countries, including the United States, whose system is the most studied, received much of its funding in the form of block grants made to public or quasi-public institutions. Tax breaks are also very important in some countries (e.g., exemption from property taxes, allowing donors to deduct gifts from taxable income). The justification for public funding and support was partly economic, and also cultural, social, and religious (Thelin, 2004).

In recent decades, a number of developments have placed this system under pressure. Higher education enrolments have soared, as a result of both student demand and governmental enthusiasm, the latter reflecting a belief that universities are a potent tool for promoting economic growth (see above). However, the increased size of the sector (absolutely and as a proportion of the eligible population) has made it increasingly difficult to secure funding increases from governments which maintain funding on a per-student basis, even though governments are committed to general expansion.

Figures for public spending on education are published annually for all member countries by the OECD. The average proportion of tertiary spending which comes from public sources has fallen since the mid-1990s for both the OECD and the European Union (EU), although not by huge amounts (five and two percentage points, respectively), and the large bulk of funding remains public, although with large differences among countries. It is, for example, extremely high (well over 90%) in most of Scandinavia, Greece, and Turkey; but well under 50% in Japan, Korea, and the United States. Looking at trends over the last decade (OECD, 2006), it emerges that about half the OECD countries have seen a fall in the proportion of funding that comes from public sources, although these falls greatly vary in scale – major in Australia, the United Kingdom, and Italy, and very small indeed in a number of other cases. Conversely, only four countries have seen increases in the proportion of funding from public sources, although in two of these (the Czech Republic and Ireland) the rise has been quite substantial. Remaining countries have seen no effective changes. These trends are, it must be emphasized, for the proportion of funding from public sources. They do not in themselves imply anything about levels of real spending.

In some countries, higher education has also attracted falling shares of public education budgets, even when the latter as a whole are protected or growing as a share of gross domestic product (GDP) or public spending (Rizzo, 2003; St. John, 2003; St. John and Parsons, 2004). However, in others, and for the OECD as a whole, the growth in the size of tertiary education has meant that it captures a larger percentage of total GDP than 10 or 20 years ago, and maintains or increases its share of the total education budget. It is at institutional and per-student level that falls in real spending are most likely to occur.

At the same time, the sector has been unable to realize significant efficiencies in the use of labor or find effective technological means to increase productivity to any significant degree. Indeed, small class sizes remain a key selling point for many private universities. In this situation, universities are constantly affected by what Baumol has characterized as the cost disease of the public services (Baumol and Bowen, 1966; Baumol, 1996). The problem is that some industries achieve above-average increases in productivity (e.g., through new machinery which allows the production of a given output with much less labor), whereas in others – notably education, health, and the performing arts – there are no such advances. Unless the latter also raise their wages, however, skilled and high-ability labor will gradually leach away. Some widening in the earnings gap may be tolerated, but there is also, inevitably, pressure to raise prices or reduce quality (e.g., in education, by increasing class sizes or reducing the number of taught hours). Even in countries which appear to have maintained per-student funding, in the sense that expenditure per student keeps pace with inflation, this creates serious pressure on the quality of higher education.

In countries where universities are unable to access funding other than governmental grants, the main impact of reduced block grant levels has simply been on class sizes (and, presumably, teaching quality), and on workloads and salaries (Capano, 1999). However, in many countries, universities have been able to change their funding mix, by increasing tuition fee income, research income, and/or consultancy income from industry. This has been particularly important for those universities which are part of a competitive system either internally or – increasingly – globally, and for whom the maintenance of prestige and brand recognition involves the hiring of high-status staff who are themselves part of a global labor market. A growing dispersion in salaries for professors is noticeable in such systems, which include the United States, United Kingdom, and Australia (see e.g., AAUP, 2007).

Increased fees, or the introduction of fees into systems which were originally free at the point of use, have been the most visible, and the most politically controversial aspect of these developments. User charges for lodging and food, and smaller student grants are also in evidence. For example, China and Britain introduced tuition fees in

the late 1990s; the value of maintenance grants shrank in Russia, and Central and Eastern Europe; tuition fees rose at rates well above inflation in the United States. In Germany, the states secured a legal victory over the federal government in 2006 which allows them to charge tuition fees, something their universities see as necessary to reestablish quality provision. Even in Scandinavia, where there is a very strong political commitment to free higher education, tuition fees are under active discussion, and there is very limited subsidy of students' living costs.

Another strategy is to maintain public funding for a limited part of the higher education sector, while directing excess demand toward courses and institutions which levy higher, or full-cost, fees. In Russia and eastern Europe, this often takes the form of two tracks within the same institution, with competition for the free places. Elsewhere, the distinctions take place at institutional level, between public and private colleges and universities. The public institutions may be the best resourced and most prestigious, but with entry highly circumscribed (as in, say Japan and much of Latin America); or, rather more commonly, public institutions may expand, but with rapidly declining expenditure per student, so that private institutions emerge as more desirable and higher prestige. The latter group includes much of Southeast Asia and Italy (see e.g., Ehrenberg, 2002; Johnstone, 2004; Karkkainen, 2006).

Although fiscal pressures provided the major impetus for reforms in higher education financing, a growing awareness of the very high private returns to years in college have been important in changing the willingness of politicians to increase cost sharing (by raising fees) rather than simply holding down expenditures at the expense of quality. This has been especially evident in the political debates surrounding finance reforms in the United Kingdom, New Zealand, Australia, and Hungary, all of which introduced comprehensive income-contingent loan systems for student fees and living expenses in the 1990s or 2000s (Barr, 2004; Barr and Crawford, 2005). Such systems require students to repay loans only when their earned income reaches a certain threshold. In the United States, there has also been a steady move toward student loans rather than grants (Hoxby, 2004; Hearn and Holdsworth, 2004; Lyall and Sell, 2006), but the system is far more complex, with multiple student-aid programs, plus various forms of tax relief for families and individuals paying college fees (and attendant information-processing costs).

A shift away from free tuition to fees plus a student loan system is attractive to governments because they save money. However, such systems do still, as a rule, involve extensive public subsidy and regulation. They also mean that support is not tied to a particular institution: students, having taken out a loan, can decide which institution to attend more freely than under a system whereby grants are made to universities, and education is free at the point of use. Some economists support such a move because it is

likely to increase competition within the sector, and also makes it easier to target support to the more needy.

Block grants which are used to keep tuition low for home students – whether within-state, as in the United States, or within-EU, as in European member states of the EU – tend to be regressive in their impact, since most students come from the wealthier segments of society (Kosters, 1999; Wolf, 2002). However, economists also, for that reason, caution against the idea that any policies for the delivery of higher education funding can provide an effective method of counterbalancing the growing wage inequalities which make a degree so attractive in the first place. With universities throughout the world recruiting from a predominantly advantaged population, government education policies which are primarily concerned with equalizing opportunities and skills should, logically, concentrate on prioritizing expenditures at pre-university level rather than on levels of funding, or funding methods, for higher education as such (see e.g., Hoxby, 1999; Cameron and Heckman, 1999).

See also: Economic Outcomes of Adult Education and Training; Education and Economic Growth; Education Production Functions: Concepts; Employability of University Graduates and Graduate Outcomes; Financing of Higher Education; Funding of University Research; Higher Education and the Labor Market; Human Capital; Productivity of University Faculty Staff; Signaling in the Labor Market; The Economics of Tuition and Fees in Higher Education; Training Finance.

Bibliography

- Abbott, M. and Doucouliagos, C. (2003). The efficiency of Australian universities: A data envelopment analysis. *Economics of Education Review* 22, 89–97.
- Alpin, C., Shackleton, J. R., and Walsh, S. (1998). Over- and under-education in the UK graduate labour market. *Studies in Higher Education* 23(1), 17–34.
- AAUP (American Association of University Professors) (2007). *Financial Inequalities in Higher Education*. Washington, DC: AAUP.
- Andras, P., Herald, N., and Charlton, B. (2007). Science production in top 20 UK universities is catching-up with the US, *Technical Report CS-TR 1053*. University of Newcastle.
- Aranson, P. H. (1995). Normative and positive theories of science and technology policy. In Sommer, J. W. (ed.) *The Academy in Crisis: The Political Economy of Higher Education*, pp 225–256. New Brunswick: Transaction Publishers.
- Atkinson, A. B. (2003). Income inequalities in OECD countries: Data and explanations, *CESifo Working Paper Series No. 881*.
- Audretsch, D., Kehmann, E., and Warning, S. (2003). University spillovers: Strategic location and new firm performance. *Centre for Economic Policy Research Discussion Paper 3837*. London: CEPR.
- Bakker, S. and Wolf, A. (2001). Special issue: Examinations and entry to higher education: Pressure and change in a mass system. *Assessment in Education* 8, 3.
- Battu, H., Belfield, C. R., and Sloane, P. J. (2000). How well can we measure overeducation and its effects? *National Institute Economic Review* 171, 82–93.

- Ball, S., Macrae, S., and Maguire, M. (1999). Young lives, diverse choices and imagined futures in an education and training market. *International Journal of Inclusive Education* **3**, 195–224.
- Barr, N. (2004). Higher education funding. *Oxford Review of Economic Policy* **20**(2), 264–283.
- Barr, N. and Crawford, I. (2005). *Financing Higher Education. Answers from the UK*. London: Routledge.
- Baumol, W. J. (1996). Children of “Performing arts: The economic dilemma.” The climbing costs of education and health care. *Journal of Cultural Economics* **20**, 183–206.
- Baumol, W. J. and Bowen, W. G. (1966). *Performing Arts: The Economic Dilemma*. New York: The Twentieth Century Fund.
- Becker, G. (1964). *Human Capital*. Chicago, IL: University of Chicago Press.
- Bettinger, E. and Long, B. T. (2004). Do college instructors matter? *NBER Working Paper Series No. 10370*. Cambridge, MA: National Bureau of Economic Research.
- Blundell, R., Dearden, L., and Sianesi, B. (2005). Measuring the returns to education. In Machin, S. and Vignoles, A. (eds.) *What's the Good of Education? The Economics of Education in the UK*, pp 117–146. Princeton, NJ: Princeton University Press.
- Bound, J. and Turner, S. (2006). Cohort crowding: How resources affect collegiate attainment. *NBER Working Paper Series No. 12424*. Cambridge, MA: National Bureau of Economic Research.
- Buchanan, J. M. and Devletoglu, N. E. (1970). *Academia in Anarchy: An Economic Diagnosis*. New York: Basic Books.
- Cam, P. (2001). The French *Baccalauréat* since 1985: Level of qualification or type of diploma? *Assessment in Education* **8**(3), 291–314.
- Cameron, S. V. and Heckman, J. J. (1993). Determinants of young male schooling and training choices? *NBER Working Paper Series No. 4327*. Cambridge, MA: National Bureau of Economic Research.
- Cameron, S. V. and Heckman, J. J. (1999). Can tuition policy combat rising wage inequality. In Koster, M. H. (ed.) *Financing College Tuition. Government Policies and Educational Priorities*, pp 76–124. Washington, DC: AEI Press.
- Capano, G. (1999). Italy: The endless transition. In Braun, D. and Merrien, F. -X. (eds.) *Towards a New Model of Governance for Universities?*, pp 195–219. London: Jessica Kingsley.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica* **69**(5), 1127–1160.
- Carpentier, V. (2005). Cycles longs et financement universitaire: Une perspective historique sur les réformes actuelles au Royaume-Uni. *Economies et Sociétés, Cahier de l'ISMEA Hors Série* **40**, 1607–1634.
- CERI (2004). *Internationalism and Trade in Higher Education*. Paris: Centre for Educational Research and Innovation, OECD.
- Chapman, G. (1996). Temporal discounting and utility for health and money. *Journal of Experimental Psychology* **22**, 771–791.
- Chevalier, A., Dolton, P., and Levacic, R. (2005). School and teacher effectiveness. In Machin, S. and Vignoles, A. (eds.) *What's the Good of Education? The Economics of Education in the UK*, pp 39–56. Princeton, NJ: Princeton University Press.
- Cohn, E., Rhine, S. L. W., and Santos, M. C. (1989). Institutions of higher education as multi-product firms: Economies of scale and scope. *Review of Economics and Statistics* **71**, 284–290.
- De Burgh, H., Fazackerley, A., and Black, J. (eds.) (2007). *Can the Prizes Still Glitter? The Future of British Universities in a Changing World*. Buckingham: University of Buckingham Press.
- Dore, R. (1997). *The Diploma Disease: Education, Qualification and Development*, 2nd edn. London: Institute of Education.
- Dundar, H. and Lewis, D. R. (1995). Departmental productivity in American universities: Economies of scale and scope. *Economics of Education Review* **14**(1), 119–144.
- Duru-Bellat, M. (2006). L'inflation scolaire. Les disillusions de la méritocratie. Paris: Seuil.
- Egerton, M. (2000). Pay differentials between early and mature graduate men: The role of state employment. *Journal of Education and Work* **13**, 3.
- Egerton, M. and Parry, G. (2001). Lifelong debt: Rates of return to mature study. *Higher Education Quarterly* **55**, 1.
- Ehrenberg, R. G. (2002). *Tuition Rising: Why College Costs so Much*. Cambridge, MA: Harvard University Press.
- Ehrenberg, R. G., Rizzo, M. J., and Jakubson, G. H. (2003). Who bears the growing cost of science at universities? *Cornell Higher Education Research Institute Working Paper WP 35*. Ithaca, NY: Cornell University.
- Etzkowitz, H. (2007). The University as crucible of enterprise: The MIT/Stanford model. In De Burgh, H., Fazackerley, A., and Black, J. (eds.) *Can the Prizes Still Glitter? The Future of British Universities in a Changing World*, pp 132–141. Buckingham: University of Buckingham Press.
- Frank, R. and Cook, P. (1995). *The Winner-Take-All Society*. New York: Free Press.
- Garvin, D. A. (1980). *The Economics of University Behaviour*. New York: Academic Press.
- Gemmell, N. (1997). Externalities to higher education: A review of the new growth literature. *Report 8 of Higher Education in the Learning Society: The Dearing Report*. London: HMSO.
- Green, F., McIntosh, S., and Vignoles, A. (2002). ‘Overeducation’ and skills – clarifying the concepts. *Discussion Paper 435*. London: LSE/CEP.
- Grogger, J. and Eide, E. (1995). Changes in college skills and the rise in the college wage premium. *Journal of Human Resources* **30**, 280–310.
- Hamermesh, D. S. and Donald, S. G. (2004). The effect of college curriculum on earnings: Accounting for non-ignorable non-response bias. *NBER Working Paper Series No. 10809*. Cambridge, MA: National Bureau of Economic Research.
- Hanushek, E. A. (2007). School resources. In Hanushek, E. A. and Welch, F. (eds.) *Handbook of the Economics of Education*, pp 865–908. Amsterdam: North-Holland.
- Hashimoto, K. and Cohn, E. (1997). Economies of scale and scope in Japanese private universities. *Education Economics* **5**(2), 107–115.
- Hearn, J. C. and Holdsworth, J. M. (2004). Federal student aid: The shift from grants to loans. In St. John, E. P. and Parsons, M. D. (eds.) *Public Funding of Higher Education: Changing Contexts and New Rationales*, pp 40–59. Baltimore, MD: Johns Hopkins Press.
- Heckman, J., Lochner, L., and Taber, C. (1998). Explaining rising wage inequality: Explorations with a dynamic-general equilibrium model of earnings with heterogeneous agents. *Review of Economic Dynamics* **1**(1), 1–54.
- Hoffman, F. and Oreopolous, P. (2006). Professor qualities and student achievement. *NBER Working Paper Series No. 12596*. Cambridge, MA: National Bureau of Economic Research.
- Hoxby, C. M. (1999). Where should federal education initiatives be directed? In Koster, M. H. (ed.) *Financing College Tuition. Government Policies and Educational Priorities*, pp 28–52. Washington, DC: AEI Press.
- Hoxby, C. M. (2004). *College Choices: The Economics of Where to go, When to go and How to Pay for it*. Chicago, IL: University of Chicago Press.
- Jenkins, A. and Wolf, A. (2005). Employers’ selection decisions: The role of qualifications and tests. In Machin, S. and Vignoles, A. (eds.) *What's the Good of Education? The Economics of Education in the UK*, pp 147–168. Princeton, NJ: Princeton University Press.
- Johnes, J. and Johnes, G. (1995). Research funding and performance in UK University departments of economics: A frontier analysis. *Economics of Education Review* **14**(3), 301–314.
- Johnstone, D. B. (2004). The economics and politics of cost sharing in higher education: Comparative perspectives. *Economics of Education Review* **23**, 403–410.
- Karkkainen, K. (2006). *The Emergence of Private Higher Education Funding in the OECD Area*. Paris: CERI, OECD.
- Kealey, T. (2008). *Sex, Science and Profits*. London: Vintage.
- Keep, E. and Mayhew, K. (2004). The economic and distributional implications of current policies in higher education. *Oxford Review of Economic Policy* **20**(2), 298–314.
- Kim, E. H., Morse, A., and Zingales, L. (2006). Are elite universities losing their competitive edge? *NBER Working Paper Series No. 12245*. Cambridge, MA: National Bureau of Economic Research.
- Kler, P. (2005). Graduate overeducation in Australia: A comparison of the mean and objective methods. *Education Economics* **13**(1), 47–72.
- Koster, M. H. (ed.) (1999). *Financing College Tuition. Government Policies and Educational Priorities*. Washington DC: The AEI Press.

- Kraft, E. and Vodopivec, M. (2003). The new kids on the block: The entry of private business schools in transition economies. *Education Economics* **11**(3), 239–257.
- Krueger, A. and Lindhal, M. (1999). Education for growth in Sweden and the world. *NBER Working Paper Series No. 7190*. Cambridge, MA: National Bureau of Economic Research.
- Lewis, W. W. (2004). *The Power of Productivity, Wealth, Poverty, and the Threat to Global Security*. Chicago, IL: University of Chicago Press.
- Lyall, K. C. and Sell, K. R. (2006). *The True Genius of America at Risk: Are We Losing Our Public Universities to de facto Privatisation?* Westport, CT: Praeger.
- Maurin, E. and McNally, S. (2005). Vive la revolution! Long term returns of 1968 to the angry students. *Discussion Paper 48*. London: Centre for the Economics of Education.
- Meiners, R. E. and Staaf, R. J. (1995). Property rights in academe. In Sommer, J. W. (ed.) *The Academy in Crisis: The Political Economy of Higher Education*, pp 197–217. New Brunswick: Transaction Publishers.
- Menon, M. (2004). Information search as an indication of rationality in student choice of higher education. *Education Economics* **12**(3), 267–283.
- Moore, M. and Ranjan, P. (2005). Globalisation versus skill-based technological change. Implications for unemployment and wage inequality. *Economic Journal* **115**(503), 391–422.
- Moretti, E. (2004a). Worker education, spillovers and productivity: Evidence from plant-level production functions. *American Economic Review* **94**(June), 656–690.
- Moretti, E. (2004b). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-section data. *Journal of Econometrics* **121**(1–2), 175–212.
- Murnane, R. J., Willett, J. B., and Tyler, J. H. (1999). Who benefits from obtaining a GED? Evidence from high school and beyond. *NBER Working Paper Series 7172*. Cambridge, MA: National Bureau of Economic Research.
- OECD (2006). *Education at a Glance*. Paris: OECD.
- Ortmann, A. (2001). Capital romance: Why Wall Street fell in love with higher education. *Education Economics* **9**(3), 293–311.
- Patrinos, H. A. (1997). Overeducation in Greece. *International Review of Education* **42**(2–3), 203–223.
- Porter, M. J., Hirotaka, T., and Mariko, S. (2000). *Can Japan Compete?* Basingstoke: Macmillan.
- Pritchett, L. (2001). Where has all the education gone? *World Bank Economic Review* **15**(3), 367–391.
- Psacharopoulos, G. and Patrinos, H. A. (2004). Returns to investment in education: A further update. *Education Economics* **12**(2), 111–134.
- Raines, J. P. and Leathers, C. G. (2003). *The Economic Institutions of Higher Education: Economic Theories of University Behaviour*. Cheltenham, UK: Edward Elgar.
- Rizzo, M. J. (2003). A (less than) zero sum game? State funding for public education: How public higher education institutions have lost. *Job Market Paper, September 2003: Cornell Higher Education Research Institute*. Ithaca, NY: CHERI.
- Rochat, D. and Demeulemeester, J. -L. (2001). Rational choice under unequal constraints: The example of Belgian higher education. *Economics of Education Review* **20**, 15–26.
- Sastry, T. and Bekhradnia, B. (2006). *Using Metrics to Allocate Research Funds*. Oxford: Higher Education Policy Institute.
- Slaughter, S. and Leslie, L. (1997). *Academic Capitalism: Politics, Policies and the Entrepreneurial University*. Baltimore, MD: Johns Hopkins University Press.
- Smith, A. (1776/1976). *The Wealth of Nations*. London: Methuen.
- Spence, A. M. (1973). Job market signaling. *Quarterly Journal of Economics* **2**(3), 355–374.
- St. John, E. P. (2003). *Refinancing the College Dream: Access, Equal Opportunity and Justice for Taxpayers*. Baltimore, MD: Johns Hopkins University Press.
- St. John, E. P. and Parsons, M. D. (eds.) (2004). *Public Funding of Higher Education: Changing Contexts and New Rationales*. Baltimore, MD: Johns Hopkins Press.
- Thelin, J. R. (2004). Higher education and the public trough: A historical perspective. In St. John, E. and Parsons, M. D. (eds.) *Public Funding of Higher Education: Changing Contexts and New Rationales*, pp 21–39. Baltimore, MD: Johns Hopkins Press.
- Tooley, J. and Stanfield, J. (2007). Private universities around the world. In De Burgh, H., Fazackerley, A., and Black, J. (eds.) *Can the Prizes Still Glitter? The Future of British Universities in a Changing World*, pp 116–122. Buckingham: University of Buckingham Press 2007.
- Varga, J. (2006). The role of labour market expectations and admission probabilities in students' application decisions on higher education: The case of Hungary. *Education Economics* **14**(3), 309–327.
- Webster, A. and Etzkowitz, H. (1998). Toward a theoretical analysis of academic-industry collaboration. In Etzkowitz, H., Webster, A., and Healey, P. (eds.) *Capitalizing Knowledge: New Intersections of Industry and Academia*, pp 47–72. Albany, NY: State University of New York Press.
- West, E. G. (1995). The economics of higher education. In Sommer, J. W. (ed.) *The Academy in Crisis: The Political Economy of Higher Education*, pp 135–170. New Brunswick: Transaction Publishers.
- Wolf, A. (2002). *Does Education Matter? Myths about Education and Economic Growth*. London: Penguin.
- Wolf, A. (2004). Education and economic performance: Simplistic theories and their policy consequences. *Oxford Review of Economic Policy* **20**(2), 315–333.
- Zemsky, R., Wegner, G., and Massy, W. (2005). *Remaking the American University: Market-Smart and Mission-Centred*. Piscataway, NJ: Rutgers University Press.
- Zhang, L. (2005). Do measures of college quality matter? The effect of college quality on graduates' earnings. *Journal of Higher Education* **28**(4), 571–596.

Further Reading

- Bok, D. (2003). *Universities in the Marketplace: The Commercialization of Higher Education*. Princeton, NJ: Princeton University Press.
- Carpentier, V. (2006). A Historical perspective on the relationship between funding in higher education and economic growth in France and the UK 1921–2003. *Journal of Higher Education Management and Policy* **18**(3), 1–26.
- Heckman, J. (2005). Lessons from the technology of skill formation. *NBER Working Paper Series 11142*. Cambridge, MA: National Bureau of Economic Research.
- Symposium on the Economics of Higher Education (1999). *Journal of Economic Perspectives* **13**(1).
- Krueger, A. (2000). *Education Matters: A Selection of Essays on Education*. Cheltenham: Edward Elgar.
- Machin, S. and Vignoles, A. (eds.) (2005). *What's the Good of Education? The Economics of Education in the UK*. Princeton, NJ: Princeton University Press.
- Weinberg, B. A., Fleisher, B. M., and Hashimoto, M. (2007). Evaluating methods for evaluating instruction: The case of higher education. *NBER Working Paper Series No. 12844*. Cambridge, MA: National Bureau of Economic Research.

Relevant Websites

- <http://cee.lse.ac.uk> – Centre for the Economics of Education London School of Economics.
- <http://www.ilr.cornell.edu> – Cornell Higher Education Research Institute, Cornell University ILR School.
- <http://www.gse.buffalo.edu> – ICHEFAP-International Comparative Higher Education and Finance Project, UB Graduate School of Education.
- <http://www.oecd.org> – Organization for Economic Cooperation and Development (OECD) site, and especially the 'Policy determinants of investment in tertiary education' project part of the site.

The Role of the OECD in the Development of Higher Education in a Globalized World

R Yelland, OECD Directorate of Education, Paris, France

© 2010 Elsevier Ltd. All rights reserved.

Origins of the Organisation for Economic Cooperation and Development

The Organisation for Economic Cooperation and Development (OECD) was created in 1961 as the successor to the Organisation for European Economic Cooperation (OEEC), which had itself been created to manage the reconstruction of Europe following World War II. During the 14 years since George Marshall's 1947 Harvard University address and the launch of the plan that bore his name, European prosperity had been restored, and the European Community had been established. The United States remained economically dominant but the Cold War was at its height. The new organization was headquartered in Paris, like its predecessor, and shared its principles and goals.

The Organization's Aims and Purpose

The OECD describes itself as: "a unique forum where the governments of 30 market democracies work together to address the economic, social and governance challenges of the globalizing world economy, as well as to exploit its opportunities. . . . OECD countries between them produce almost 60% of the world's goods and services, but in an interdependent global economy the OECD does not and cannot work alone. It shares expertise and exchanges views with more than 100 countries worldwide on topics of mutual concern from improving health and education systems' performance to encouraging innovation and ensuring accountability of governments to their citizens." (OECD, 2005: 7)

Article 1 of the OECD Convention says that the aims of the organization

shall be to promote policies designed:

- a. to achieve the highest sustainable economic growth and employment and a rising standard of living in Member countries, while maintaining financial stability, and thus to contribute to the development of the world economy;
- b. to contribute to sound economic expansion in Member as well as non-member countries in the process of economic development; and
- c. to contribute to the expansion of world trade on a multilateral, non-discriminatory basis in accordance with international obligations.

OECD work covers a wide range of government activity – from aging society to transport; from entrepreneurship to agriculture. The process by which priorities are set, specific outputs determined, and budgets allocated is led by the Council, which is made up of representatives of the member countries and the European Commission.

Roles and Methods of Work

The paradigm for OECD activity in pursuit of its objectives runs from data collection and analysis, through discussion and decision making, to surveillance through peer review of policy implementation.

One of the key roles of the organization is therefore as a collector and provider of reliable and comparable data on economic and social issues. To underpin its analytical and policy work, the OECD compiles statistics for the 30 member countries and selected nonmember economies, and contributes to the development of international standards for statistics. OECD statistics are made public in a range of statistical publications and databases.

Developing cross-cutting, high-quality, shared, accessible data about how a society is doing is crucial to ensure that decision making is both responsive and responsible at all levels. However, in an age of unprecedented and overwhelming information flows, the common understanding necessary for informed public discourse is often inadequate. In particular, quality of life is a concern common to all societies and a consensus is growing around the need to develop a more comprehensive view of progress – one that takes into account social, environmental, and economic concerns.

The OECD, in cooperation with other international organizations, has embarked on a global project to measure and foster the progress of societies. The work is built around regional conferences culminating in a world forum, launched in Palermo in 2004, and the second of which on statistics, knowledge, and policy took place in Istanbul in June 2007.

The organization also provides a setting in which member countries can compare and assess their own policies and practices. It is frequently the initiator and incubator of soft law – nonbinding instruments – that is, recommendations and decisions, which on occasion lead to formal agreements, treaties, or conventions. Perhaps the best-known example is the 1997 Convention on Combating Bribery of Foreign Public Officials in International

Business Transactions which has been ratified by a number of nonmember countries as well as all OECD members. More recently in 2005, the Council accepted the OECD recommendation concerning guidelines on earthquake safety in schools, which seeks to apply peer pressure and review to improve national policy and practice. This was the first recommendation in the field of education and was quickly followed by the Guidelines for Quality Provision in Cross-border Higher Education described more fully below.

Growth in Membership in OECD

The organization grew quickly to the point where by 1974 its 24 members “produced more than two-thirds of the world’s goods, and accounted for more than four-fifths of its trade” (Sullivan, 1997: 7). The 20 founding members of the OECD are: Austria, Belgium, Canada, Denmark, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States. They have since been joined by Japan (1964), Finland (1969), Australia (1971), New Zealand (1973), Mexico (1994), the Czech Republic (1995), Hungary, Poland, Korea (all 1996), and the Slovak Republic (2000).

Further enlargement is on the agenda. Since 2004, the European Union includes countries which are not OECD members. Moreover, attention is increasingly paid to the major global players: the 2006 OECD Ministerial Council Meeting encouraged the major emerging economies to participate more actively in OECD’s work and to move closer to OECD standards and disciplines and they called on the organization to support this process and to strengthen its capacity to deal with global issues and the impact of important new players in a comprehensive, systematic, and forward-looking manner.

On 16 May 2007, the OECD Council, meeting at the ministerial level, took a landmark decision, agreeing to:

offer enhanced engagement programmes, with a view to possible membership, to Brazil, China, India, Indonesia and South Africa; to invite Chile, Estonia, Israel, the Russian Federation and Slovenia to open discussions for membership of the Organisation; and to expand the OECD’s relations with selected countries and regions of strategic interest to the OECD, priority being given to South East Asia, with a view to identifying countries for possible membership.

Funding OECD Work

The organization is funded primarily by assessed and voluntary contributions from its member countries,

within the framework of a biennial program of work and budget. Through the budget, the Council accords the necessary commitment authorization and makes the necessary appropriations for the functioning of the organization and the delivery of its agreed output results. It determines the amount of assessed contributions to be paid by members after taking into account other resources of the organization.

All the OECD member countries fund the Part I Budget (M EUR 167 in 2007), which accounts for about 50% of the overall budget (M EUR 340 in 2007). Their contributions are essentially based on a scale of contributions proportional to the relative size of their economies with a cap of 24.975% (USA). Part II programs (around M EUR 67 in 2007) include a significant number of non member participants in addition to those member countries. Part II programs are funded according to a scale of contributions or other agreements among the participating countries. The other main budget items are pensions (M EUR 86 in 2007) and publications (M EUR 12).

The approval of the budget by the Council empowers the secretary-general to make all payments to be borne by the organization, and to receive the income entered in the budget, together with any other resources accruing to the organization in respect of its activities.

Engagement with Nonmember Economies

From its beginning in 1960 global relations have been an integral part of the OECD program. The organization’s global nature has been expressed in the OECD Convention, which states in Article 1, that OECD Members “should contribute to sound economic expansion in Member as well as non-Members countries in the process of economic development.”

However, until the mid-1990s, most of the work with nonmembers was either done with Central and East European transition economies, or it was undertaken by separate bodies of the organization. The work with transition economies was unidirectional, that is, aimed at assisting transition economies in their development toward market economies. Things have changed since 1997 when the OECD progressively evolved from a traditional concept of outreach in its relations with the rest of the world to a two-way flow where giving access to and participation in core OECD work and processes is as important as disseminating OECD best practices. Today, besides the OECD’s 30 members, many more countries and economies outside its membership are involved in the organization’s work. In 2007, 25 nonmembers participated as regular observers or full participants in at least one committee or working group. Altogether, there are about 70 nonmembers engaged in at least one OECD working party, scheme, or program.

The OECD's Global Relations Programme includes policies to enhance the investment climate, improve both public and corporate governance, build stronger national institutions for trade policy, and, more generally, to encourage the range of domestic reforms required in all countries to reap the benefits of globalization. The program consists of global forums, regional approaches, and country programs, which together propose solutions at the global, regional, and country levels.

Development of Work on Education

Even in the 1950s, before the establishment of the OECD, one office within the OEEC addressed issues of education and provided analyses about how to increase the number of students, as well as scientific and technical personnel. This was a response to the need to improve the scientific and technical potential of European countries in the face of competition from the US and – following the Sputnik launch – the USSR. The 1960s were a decade of rapid educational growth throughout the OECD area, especially in secondary and higher education. Consequently, the OECD focused on issues of managing – and paying for – expansion, using a combination of manpower planning and an investment approach to education.

The formal recognition of the educational policy role of OECD came with the creation of the Education Committee in 1970, with a mandate to address “prospects and policies for educational growth and development to meet social and economic objectives, both in relation to the general problems of allocating resources and for the efficient management of resources in education as such.” Both the Programme on Educational Building and the Programme on Institutional Management in Higher Education have their origins in the early 1970s, while other key themes of work in that decade include equality of opportunity and the related debate about selection for secondary school, and early work on recurrent education which – renamed lifelong learning – has been a leitmotif of OECD work ever since. The economic outlook for the late 1970s and 1980s was dominated by the oil shocks and the ensuing recession. The OECD's work on education put a greater emphasis on the labor market relevance of education through vocational education and training, and to an examination of the funding of higher education.

The dominant feature of the 1990s was the developing capacity and confidence of the organization in collecting quantitative information on education, using it to develop indicators, and not just statistics, and the gradual acceptance that policy development could be supported by international comparative analysis of outputs and outcomes, not only inputs. The annual publication in September of Education at a Glance, first published in 1992 has become a fixture on the educational scene, while the Programme on

International Student Assessment (PISA) in secondary education has had an even more explosive impact. In 2002 a separate Directorate for Education was created which brings together most the OECDs work on educational issues.

OECD Work on Higher Education

Unlike UNESCO the OECD Directorate for Education does not have a division of higher education; and unlike the Council of Europe it does not have a standing Committee overseeing work on higher education policy. Analysis of higher education policy and practice runs across the four functional divisions of the directorate: Education and Training Policy Division, Centre for Educational Research and Innovation, Indicators and Analysis Division, and Education Management and Infrastructure Division. The aim is to bring a set of skills and expertise to bear on the analysis of the sector in appropriate combinations.

It has been recognized by all the OECD countries that higher education plays a vital role in driving economic growth and social cohesion, and it is now understood that a high-quality system of higher education is central to the ability of nations to participate successfully in the global knowledge economy. With more and more players getting involved in higher education, and with the great expansion and diversification of higher education system itself, this sector is facing challenges to meet the needs from knowledge economy, labor market, local community, and individual learner. This context has led the OECD to focus on higher education in its work related to education and innovation. In 2006, the biennial meeting of the OECD Education Committee at Ministerial level at Athens, Greece focused specially on higher education at the first time. “Higher education: Quality, equity and efficiency” was the theme of the meeting where issues like purpose, governance and sustainable provision of higher education, who should pay for higher education, measuring the quality and impact of higher education, internationalization and higher education's contribution to research and innovation were addressed. Participants from different countries all agreed on a new task: to go beyond growth, by making higher education not just big but also better.

The OECD is the main collector and provider of reliable, international comparative data on a wide array of aspects of education. On higher education, the OECD develops statistics and indicators about access and participation, finance and investment, labor force attainment and on internationalization. The latter will seek to capture more comprehensive data on mobility, including data by field of studies. In addition, the OECD will continue to develop indicators on labor market earnings and the returns to education for the individual and for society. These statistics and indicators are widely used to inform policy decision making. The OECD will also assess the

skills and competencies of the adult population in various countries with the new Programme on International Assessment of Adult Competencies (PIAAC). Since 2007, the OECD has been exploring possibilities for an international assessment of the general and subject specific competencies of higher education graduates known as AHELO – the Assessment of Higher Education Learning Outcomes.

The thematic review of tertiary education, which has been conducted between 2004 and 2007, aims to examine how the organization, financing and management of tertiary education can help countries achieve their economic and social objectives. The focus of the review is primarily upon national policies for tertiary education systems, rather than upon policies and practices at the institutional level. Key questions include the economic and social objectives of tertiary education; sustainability, structures, links and mechanisms to ensure quality; mobilizing adequate funding resources; and national policies and mechanisms to ensure effective governance. More specifically, the review synthesizes research-based evidence on the impact of tertiary education policies, identifies innovative and successful policy initiatives and practices, facilitates exchanges of lessons and experiences among countries, and identifies policy options for countries. Twenty-four countries take part in the review. In addition, some countries have chosen to take part in a ‘Country Review’. This involves an external review team undertaking a country visit. The panel produces a ‘Country Note’ containing an analysis of national tertiary education policies and policy recommendations. The final comparative report published in 2008 brings together the evidence collected and presents key findings and policy messages.

Internationalization and Trade in Higher Education

Internationalization and trade in higher education have also been a recent focus of the work of the OECD’s directorate for education. The increasing international mobility of students, faculty and employees, and new forms of cross-border delivery of education call for an international policy approach because internationalization challenges the regulatory capacity of national and regional education authorities. Cross-border tertiary education presents opportunities and challenges in sending as well as receiving countries for quality, access, cost, and capacity building. It has become a significant economic and commercial stake in some countries and educational services are included in the current negotiations under the General Agreement on Trade in Services in the World Trade Organization. The Center for Educational Research and Innovation (CERI) has brought together the education and trade communities to discuss the opportunities and

challenges of these new trends through three fora on trade and education in 2002, 2003, and 2004.

The policy implications of cross-border tertiary education have been analyzed in two publications: *Internationalization and Trade in Higher Education: Opportunities and Challenges*; and *Quality and Recognition in Higher Education: The Cross-border Challenge*. Furthermore, the OECD develops conceptual foundations and a data strategy to improve international statistics and indicators on the internationalization of tertiary education.

The OECD has collaborated with UNESCO on the development of guidelines for cross-border tertiary education that will enhance learner protection while respecting countries’ rights to regulate the quality of their systems. These guidelines are based on an analysis of the ways in which OECD countries deal with international quality assurance, accreditation, and recognition of higher education qualifications. The guidelines have been endorsed by the OECD Council in 2005 and their implementation is overseen by the OECD Education Committee.

Future of University and Tertiary Education Systems

Another strand of OECD’s work in higher education examines options for the future of university and tertiary education systems over the next 15 years. To this end, CERI elaborated four scenarios about the future of higher education systems. These scenarios inform about the socioeconomic changes affecting higher education and about the challenges and opportunities offered by these trends. They are intended to stimulate policy debate about strategic and long-term plans, and to help stakeholders in proposing adequate responses to socioeconomic changes.

To develop the scenarios, the OECD and higher education experts undertook in-depth thematic reviews of how six major socioeconomic changes affect higher education systems. The changes reviewed were:

1. demographic change;
2. avenues opened by new information and communication technologies for learning, teaching and research;
3. the rise of market forces and quasi market forces;
4. the globalization of higher education;
5. changes in university research; and
6. labor market demands.

The four scenarios about the future of higher education systems were developed based on these thematic reviews and on the discussions in various expert meetings and stakeholder seminars. The OECD program on Institutional Management in Higher Education (IMHE) was also involved in this project and looked at the implications of socioeconomic changes for higher education institutions.

The future scenarios were presented and discussed at the Meeting of OECD Education Ministers in Athens in 2006.

The Program for Institutional Management in Higher Education (IMHE)

Issues around the governance and management of higher education institutions have for many years been a feature of the work of the OECD's Directorate of Education. These topics are addressed by the Programme on Institutional Management in Higher Education (IMHE). This program differs from other OECD programs in that it is financed by the contributions of more than 200 members, including higher education institutions (the majority), associations of institutions and Government departments. Its mission is to provide members with analysis and advice on issues of higher education governance and institutional strategy. Since its creation in 1969 the original focus on management information tools and capacity building has broadened to include the wider interface between government, academia, and the market. IMHE is essentially a collaborative network. Member institutions meet in a variety of settings to discuss and to develop their idea on issues of the day. The research activities are supported by OECD experts and external consultants. A governing board made up of representatives of the members determines the Programme of Work, within the broad framework of the OECD.

IMHE addresses a variety of issues about institutional management and typically runs three to four projects at a time. One major current focus has been the investigation of innovative solutions for improving the contribution of higher education institutions to regional development. This project is a response to the finding that higher education can make significant contributions to regional development but that potential synergies are often thwarted by failures of communication and conflicting agendas in higher education institutions. Other IMHE projects have addressed institutional strategic management in a changing policy environment, the impact of ranking an institutional behavior, university research management, quality management, and higher education internationalization strategies.

IMHE supports its members through the publication of the journal *Higher Education Management and Policy*, which addresses higher education administrators and managers. The journal covers issues such as quality assurance, human resources, funding, and internationalization. In addition, IMHE helps its members to build capacity in institutional management through leadership development seminars, the biennial general conference, international meetings on specific issues and the IMHE newsletter and website. Topics that have been addressed in recent years

include human resource management in higher education, public relations of higher education institutions, or the role of governing boards. The Centre for Effective Learning Environments (CELE) works with IMHE to examine trends in the provision of tertiary education infrastructure.

Directorate for Science, Technology and Industry (DSTI) Work on Higher Education

The DSTI is responsible for work on research and innovation. Innovation and R&D are pillars of a knowledge-based society, as well as fundamental to the dynamism and effectiveness of education. They focus on questions relevant to education like: is education sufficiently innovative? Does the education system invest sufficiently in R&D? A number of relevant indicators suggest that there is a long way still to go. The lessons of this work are of interest to those at very different levels in the overall educational enterprise, from the 'macro' level of government right through to 'micro' level of practice. It also brings in the 'meso' level of the wide range of intermediaries in between who are critical to the innovation and R&D process.

With its agenda set by Ministries responsible for science, technology and Industry, DSTI has a strong interest in educational innovation and R&D, and the development of human resources. The essential part of DSTI's work deals with higher education and how this sector can be more effectively developed and used as a lever of change within knowledge-based societies. It focuses on policies in support of private-sector innovation: financing R&D and education; using intellectual property rights and competition policy to create innovation-friendly environments; and fostering links between science and industry. Management of public research, including R&D in higher education is also a key area of DSTI's analyses. Work focuses on: priority setting and funding of public research and questions relating to development of human resources for the knowledge economy, including innovation in the education system.

The Global Dimension of OECD's Work on Higher Education

Given the increasing importance of higher education in emerging economies and the growing internationalization of higher education, the OECD undertakes projects with a global dimension and includes nonmember countries in their studies. Global issues are frequently pursued in co-operation with UNESCO and the World Bank. Together with UNESCO, IMHE has begun to develop an International Higher Education Policy Portal (IHEPP) intended to inform policymakers, institutional leaders, and other stakeholders about policy developments and

experiences in higher education. Other examples of projects with a global perspective are the aforementioned guidelines for Quality Provision in Cross-border Higher Education (see above), the forum on trade in educational services, and the study 'Financing Education – Investments and Returns', which investigated the link between the level of education and economic growth in 16 emerging economies. All these activities were undertaken in cooperation with UNESCO.

The OECD also frequently cooperates with the World Bank in order to make its analyses valuable to member as well as nonmember countries. IMHE supported the project Internationalization of Higher Education in Latin America which was led by the World Bank and the identified risks, opportunities and challenges of internationalization in this specific context. Another example is a seminar on quality assurance in tertiary education which had the goal of moving toward a collection of general principles of good practice, as well as to set a research agenda to better understand what makes for effective quality assurance in developing countries. CERI is working on the possible uses of cross-border tertiary education for capacity building in developing countries with Netherlands Organisation for International Cooperation in Higher Education (NUFFIC) and the World Bank, and is preparing a joint publication with the latter.

Conclusion

Over the past 40 years, the OECD has made important contributions to the development of higher education around the world. In cooperation with policymakers, institutional leaders, and researchers, the OECD has produced influential analyses of important issues such as lifelong learning, internationalization, or the contribution of higher education to socioeconomic development. This sharing of good practices has supported policymakers and institutional leaders in preparing higher education systems for a global knowledge economy. Due to the OECD's provision of international comparative statistics and indicators on the performance of higher education systems, policymakers now have the necessary information for making sustainable and strategic policy decisions. As an international organization, the OECD is uniquely positioned for collecting international comparative data, promoting the international exchange on good practices and developing international guidelines. This role as an international disseminator of data and good practices is likely to become ever more important as higher education continues to expand, diversify, and internationalize.

Acknowledgment

I would also like to acknowledge the assistance of Wen Wen, Department of Education, University of Oxford, in preparing this article.

Bibliography

OECD (2005). *OECD Annual Report 2005*. Paris: OECD.
Sullivan, S. (1997). *From War to Wealth: 50 Years of Innovation*. Paris: OECD.

Further Reading

OECD (2004). *Internationalisation and Trade in Higher Education: Opportunities and Challenges*. Paris: OECD.
OECD (2004). *Quality and Recognition in Higher Education: The Cross-Border Challenge*. Paris: OECD.
OECD (2005). *E-learning in Tertiary Education – Where Do We Stand?* Paris: OECD.
OECD (2005). *From Education to Work – A Difficult Transition for Young Adults with Low Levels of Education*. Paris: OECD.
OECD (2005). *Learning a Living – First Results of the Adult Literacy and Life Skills Survey*. Paris: OECD.
OECD (2005). *Promoting Adult Learning*. Paris: OECD.
OECD (2005). *University Research Management – Developing Research in New Institutions*. Paris: OECD.
OECD (2006). *21st Century Learning Environments*. Paris: OECD.
OECD (2006). *Education Policy Analysis: Focus on Higher Education 2005–2006*. Paris: OECD.
OECD (2006). *ICT and Learning Supporting Out-of-School Youth and Adults*. Paris: OECD.
OECD (2006). *Programme on Educational Building: Compendium of Exemplary Educational Facilities*. Paris: OECD.
OECD (2006). *Skills Upgrading: New Policy Perspectives*. Paris: OECD.
OECD (2007). *Cross-Border Tertiary Education: A Way towards Capacity Development* (Co-edition with the World Bank). Paris: OECD.
OECD (2007). *Education at a Glance*. Paris: OECD.
OECD (2007). *Higher Education and Regions: Globally Competitive, Locally Engaged*. Paris: OECD.
OECD (2007). *Knowledge Management: Evidence in Education*. Paris: OECD.
OECD (2007). *OECD Annual Report 2007*. Paris: OECD.
OECD (2007). *Understanding the Social Outcomes of Learning*. Paris: OECD.
OECD (2008). *Tertiary Education for the Knowledge Society*. Paris: OECD.
OECD IMHE Journal: *Higher Education Management and Policy*.
Papadopoulos, G. (1995). *Education 1960–1990: The OECD Perspective*. Paris: OECD.

Relevant Website

<http://www.oecd.org>– Organisation for Economic Co-operation and Development, Chair's summary of the OECD Council at Ministerial Level, Paris, 23–24 May 2006 – Delivering Prosperity; OECD (forthcoming), Thematic Review of Tertiary Education, Comparative Report; The OECD's Global Relation Program 2007–2008.

Transforming Higher Education in Developing Countries: The Role of the World Bank*

J Salmi and R M Bassett, The World Bank, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The time when higher education was almost a proscribed subject at the World Bank is long past. The 2002 publication of *Constructing Knowledge Societies* marked the culmination of a decade of growing attention to and investment in higher education in all regions of the world where the Bank is active. Not only is higher education seen today as a crucial element in the Bank's development strategy, but the Bank's thinking and actions in this area have also begun to influence a number of bilateral donors into reconsidering their own focus on basic education as the only priority for financial and technical assistance in the education sector.

Higher education's contribution to economic and social development is multifold. It exercises a direct influence on national productivity, which largely determines living standards and a country's ability to compete in the global economy. Higher education institutions support knowledge-driven economic growth strategies and poverty reduction by (1) training a qualified and adaptable labor force, including high-level scientists, professionals, technicians, teachers in basic and secondary education, and future government, civil service, and business leaders; (2) generating new knowledge; and (3) building the capacity to access existing stores of global knowledge and to adapt that knowledge to local use. Higher education institutions are unique in their ability to integrate and create synergy among these three dimensions. Sustainable transformation and growth throughout the economy are not possible without the capacity-building contribution of an innovative higher education system. This is especially true in low-income countries with weak institutional capacity and limited human capital.

This article begins by presenting the progression of the World Bank's position with respect to higher education and its contribution to development. It then reviews the different ways in which the Bank works with developing and transition countries in support of their higher education reform efforts, before concluding with a few

observations about higher education's prospects within the World Bank's agenda.

The World Bank's Policy Framework: From Taboo Topic to Mainstream Theme

Even though the first World Bank project in education, launched in 1963 in Tunisia, included a teacher-training component, for many years the institution did not consider support for higher education among its priorities, and decades passed before the Bank developed a fully articulated and supportive position regarding the importance of higher education in the development agenda. In fact, during the era between 1963 and the 1994 publication of the first policy paper on higher education, *Higher Education: The Lessons of Experience*, the education sector within the Bank was guided by rate-of-return formulas that discouraged investment in higher education in favor of primary and secondary education. Skepticism over the comparative value of investment in and reform of higher education in borrower/client countries led, in parts of the Bank, to the purposeful exclusion or, at best, minimization of higher education within lending for the education sector.

Later Bank-supported research and policy papers, including those analyzed in this document, cast doubt on the validity of rate-of-return analysis as the main approach for measuring the value of investment in tertiary education. By focusing exclusively on the private returns of government spending, the methodology excluded broad social benefits such as research externalities, entrepreneurship, job creation, good economic and political governance, and the effect of a highly educated cadre of workers on a nation's health and social fabric (Bloom *et al.*, 2005: 20).

With the acknowledgment that higher education "is of paramount importance for economic and social development" (p. 1), the first policy paper marks a significant turning point in the Bank's support for initiatives designed to improve higher education capacity in its client countries. This paper identifies the crisis in higher education emerging around the world: with diminished public funding coupled with dramatically increasing need and demand for higher education and operational inefficiencies at the institutional level.

In this policy paper, the Bank outlined four key directions aimed at reform for higher education in developing

*Respectively Tertiary Education Coordinator and tertiary education consultant the World Bank. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and should not be attributed in any manner to the World Bank, the members of its Board of Executive Directors or the countries they represent.

countries: diversification of institutions, including private institutions; diversification of funding sources, including student fees, and linking funding with performance; reexamining the links between government and higher education; and focusing policy developments on quality and equity objectives (World Bank, 1994: 4). However, this paper does not indicate so much a transition of focus from primary/secondary to higher education as it does a realization that higher education serves a separate and equally important function in broad socioeconomic development, requiring a distinct level of attention and expertise.

Following the first policy paper and the subsequent emergence of higher education as a stronger area of interest for the Bank's education endeavors, the Bank and the United Nations Educational, Scientific and Cultural Organization (UNESCO) convened an independent task force to examine the specific challenges facing higher education in developing countries at the turn of the twenty-first century. *Higher Education in Developing Countries: Peril and Promise* (2000) emerged from the efforts of this group, which was headed by professors Henry Rosovsky (Harvard University) and Mamphela Ramphele (University of Cape Town), and included researchers, policy analysts, and practitioners of higher education from 13 countries.

Peril and Promise emerged from 2 years of detailed research as well as discussions and interviews of relevant issues with actors around the world. The focus of the report is the crises that higher education in developing countries must manage in order to fulfill its mission of promoting sustainable cultural, social, and economic development. Major areas of concern included: privatization, access (particularly for women), diminished government funding, and increased demand. Of greatest consequence in the short term is the decline of funding at a time of dramatically increased demand as well as the understanding that effective development depends on becoming an active participant in the international knowledge economy.

The report expounds upon four main areas in which immediate action is needed, including: funding (focusing on mix-source models), resources (using physical and human capital to its greatest advantage), governance (promoting structures for good governance and effective management techniques for environments with limited resources), and curriculum development (with focus on complementary elements such as general education and investment in science and technology). These issues are, by and large, reminiscent of the issues of focus in *Higher Education: The Lessons of Experience*, exposing the inherent challenge of effectively addressing deeply embedded issues, regardless of awareness of or commitment to the necessity of taking such action.

Peril and Promise provides both qualitative, anecdotal evidence of the problems plaguing higher education development, and quantitative data supporting comparative analyses on the demographics of higher education around the world. The report takes a stance supporting

comprehensive investment in higher education as a vital component of sustainable development in a global knowledge-based economy and political environment. It concludes that policymakers and the higher education community must focus on two, highly broad issue areas: increasing resources for higher education and managing those resources better and more efficiently.

Finally, in 2002, the Bank published its most comprehensive examination of higher education as a tool for poverty reduction, development, and participation in the global knowledge economy in *Constructing Knowledge Societies: New Challenges for Tertiary Education*, which emerged from a thorough data collection and research process that not only focused on the efforts of the Bank directly, but also examined the experiences of relevant actors outside the Bank. By bringing together these different perspectives, *Constructing Knowledge Societies* offers a broad analysis of the current issues facing the higher education sector across developing and transitional countries. It also presents predictions and recommendations for the future of higher education in these countries.

Between the publication of *Higher Education: The Lessons of Experience* in 1994 and *Constructing Knowledge Societies: New Challenges for Tertiary Education* in 2002, the list of the most challenging issues facing higher education reform had changed very little. Unresolved challenges at the time of the publication of *Constructing Knowledge Societies* included: the need to expand the higher education sector to meet the rapidly growing demand, inequality of access and outcomes, quality assurance concerns, and the need for more effective and relevant governance and management structures. The range of findings of *Constructing Knowledge Societies* spans from more traditional concerns such as promoting higher education as a tool for human (cultural), economic, and social development to more modern challenges such as utilizing higher education as a means of participating in an increasingly service- and technology-driven world and facing the myriad challenges of globalization.

The Bank, as a lending and knowledge-sharing institution, is in a unique position to promote mechanisms to assist countries in addressing all of these issues within higher education reform policies. *Constructing Knowledge Societies: New Challenges for Tertiary Education* presents the basic principles that the Bank follows in supporting specific activities within any one country and notes that the support should: "be appropriate to each country's circumstances; based on strategic planning at the national and institutional levels; promote autonomy and accountability in the higher education sector; focus on capacity enhancement and regional sharing of successful experiences and models; sequence activities in a manner consistent with the entire development agenda for any one country; and acknowledge and account for the political nature of higher education reform" (pp. 119, 120).

In its conclusion, *Constructing Knowledge Societies* does not attempt to present readymade solutions, per se, to the challenges originally outlined in *Lessons* and reiterated in *Peril and Promise*. Indeed, in *Constructing Knowledge Societies* the Bank explores more thoroughly than in previous publications the need for locally driven higher education initiatives that can be supported and enhanced by the international expertise offered by a multilateral organization like the Bank.

At the same time, however, *Constructing Knowledge Societies* identifies significant global public goods that countries on their own may not be able to handle effectively, including human capital migration (brain drain), intellectual property concerns, the challenge of quality assurance for borderless higher education, the digital divide, and the impact of the global trade in services on higher education.

In sum, *Constructing Knowledge Societies* makes the case that merely modernizing the higher education system will no longer be enough. To play their role effectively, higher education systems must be locally relevant yet globally aware, adaptive and evolving, flexible, and of high quality. *Constructing Knowledge Societies* underscores the continuum of engagement in higher education development that the Bank has undertaken and the direction in which the Bank anticipates a long future of related efforts.

The latest World Bank *Education Sector Strategy Update* (December, 2005) confirms the importance of tertiary education within a holistic view of the education sector, recognizing the key role of education advancement at all levels and the significant linkages between education and the rest of the economy. The Education Sector Strategy Update (ESSU) emphasizes three themes: “integrating education into a country-wide perspective, broadening the strategic agenda through a system-wide approach (from pre-school through tertiary education), and becoming more results-oriented” (p. vi). These themes underscore the broader focus on linkages between education and labor markets, with more systematic attention to secondary and tertiary as pillars of the knowledge economy, to complement efforts made in expanding basic education through Education for All (EFA). The 2005 ESSU ties together the Bank’s successive and evolving statements about the significance of tertiary education with *Lessons* (1994) and continued through *Perils and Promise* (2000) and *Constructing Knowledge Societies* (2002) and translates the foundation of these works with continued support for investment in tertiary education development around the world.

Table 1 summarizes the main messages and implications of the various policy papers reviewed in this article.

World Bank Activities in Higher Education

Though the World Bank is known, essentially, for the loans given to countries in support of their development

efforts, an equally significant contribution is the policy dialog and analytical work conducted by the Bank, as a knowledge-sharing institution, to help governments consider options about possible higher education reforms and set the stage for their implementation. **Table 2** presents the list of countries where higher education studies were prepared in recent years.

Through its projects across four continents, the World Bank has supported countries’ efforts to expand higher education and improve its quality and relevance. Even in earlier times when higher education was not a priority subsector, the Bank financed a significant number of projects in response to specific requests by countries such as China, Kenya, or Tunisia. World Bank projects in tertiary education have amounted to approximately one-third of total lending in education in the era since the 1970s. **Table 3** details the lending for tertiary education within total education lending between 2001 and 2006, providing important data to highlight fluctuations in education lending as a whole and tertiary education lending in particular. Specific commitments to tertiary education during this 5-year period amount to only 14.5% of new education commitments. It is important to note, however, that commitments in other areas (e.g., lifelong learning initiatives and lending for teacher training for primary and secondary school teaching) may be classified under a different education subsector but are, in fact, directly tied to tertiary education.

Table 4 provides a more detailed illustration of the scope of Bank lending for higher education in recent years, with total lending figures broken down by region and highlighted with examples of the largest borrower countries during this period.

The main types of activities supported by Bank projects come under one or more of the following headings, tailored to the needs of the country and the specific requests of the national authorities and the higher education community:

- vision development, strategic planning, and consensus building at both the national and institutional levels;
- finance reforms (e.g., allocation of recurrent budget; competitive funding; cost sharing; student loans; and scholarships)
- governance and management reforms (creation of policy bodies; mergers; adoption of academic credit systems; and management information systems);
- quality improvement (strengthening of existing programs; evaluation and accreditation systems; innovations in program content and delivery; innovations in academic organization; and information and communication infrastructure);
- institutional diversification (establishment or strengthening of polytechnic or technical institutes); and
- science and technology development (strategy development; capacity for monitoring and evaluation; reform

Table 1 Key publications/policy frameworks

<i>Publication</i>	<i>Challenges</i>	<i>Main messages</i>	<i>Bank role/strategy</i>
<i>Higher Education: Lessons of Experience</i> (1994)	Need to re-examine rate-of-return priority that de-emphasized higher education in favor of primary and secondary education	<ul style="list-style-type: none"> • Higher education is of paramount importance for economic and social development • A crisis in higher education is emerging around the world, caused by diminished public funding, rapidly increasing demand for higher education, and operational inefficiencies at the institutional level 	<ul style="list-style-type: none"> • Diversification of institutions, including private institutions • Diversification of funding sources, including cost-sharing, and linking funding with performance • Re-examining the links between government and higher education institutions • Focusing policy developments on quality and equity objectives • Funding (focusing on mixed source models) • Resources (effectively using physical and human capital) • Governance (promoting structures for good governance and effective management) • Curriculum development (with focus on complementary elements such as general education and investment in science and technology)
<i>Higher Education in Developing Countries: Peril and Promise</i> (2000)	Crises in higher education in developing countries must be managed to ensure HE fulfils its mission of promoting sustainable cultural, social, and economic development	<p>Major areas of concern include:</p> <ul style="list-style-type: none"> • privatization • access (particularly for women) • diminished government funding • increased demand for HE 	<ul style="list-style-type: none"> • Curriculum development (with focus on complementary elements such as general education and investment in science and technology)
<i>Constructing Knowledge Societies: New Challenges for Tertiary Education</i> (2002)	Remaining challenges include expanding the HE sector to meet increasing demand, inequality of access and outcomes, quality assurance concerns, and the need for more effective governance and management structures	<p>Programmatic support for HE should be:</p> <ul style="list-style-type: none"> • appropriate to each country's circumstances • involve planning for development at institutional, local, and national levels • promote autonomy and accountability • focus on capacity enhancement and regional sharing of successful experiences and models • sequence activities in a manner consistent with the entire development agenda for any one country • acknowledge political nature of higher education reform 	<p>Asserts the need for locally driven higher education initiatives that can be supported and enhanced by the international expertise offered by a multilateral organization like the Bank</p> <p>Identifies significant global public goods that countries on their own may not be able to handle effectively, including:</p> <ul style="list-style-type: none"> • human capital migration (brain drain) • intellectual property concerns • the challenge of quality assurance for borderless higher education • the digital divide • the impact of the global trade in services on higher education
<i>Education Sector Strategy Update</i> (December 2005)	Recognizes the key role of education advancement at all levels and the significant linkages between education and the rest of the economy	<p>Emphasizes three themes:</p> <ul style="list-style-type: none"> • integrating education into a country-wide perspective • broadening the strategic agenda through a system-wide approach (from early childhood development through tertiary education) • becoming more results oriented 	<p>Support Bank's broader focus on education-labor market linkages, with more systematic attention to secondary and tertiary as pillars of the knowledge economy, to complement efforts made in expanding primary education through Education for All</p>

of resource-allocation mechanisms; competitive funding; promotion of research in priority areas; joint public-private sector technology development; capacity for metrology, standards, and quality testing; and intellectual property rights).

The combination of policy dialog, analytical work, and financial assistance has facilitated the implementation of comprehensive reforms in the higher education sector in countries as diverse as Argentina, Chile, China, Vietnam, Egypt, Tunisia, Ghana, or Mozambique. Often, governments use the resources made available through multilateral

loans as incentives for institutions willing to challenge old modes of operation and be innovative new ground after thorough strategic planning and/or self-evaluation efforts.

The competitive innovation funds that several countries have established with World Bank support have been among the most effective channels to stimulate the participation of higher education institutions in meaningful transformation efforts. Under such funds, institutions are typically invited to formulate project proposals that are reviewed and selected by committees of peers according to transparent procedures and criteria. One of the main benefits of competitive funding mechanisms is that they

Table 2 Recent higher education studies (2001–2006)

Region	Countries
Eastern Europe and Central Asia	Russia (2002), Georgia (2003), Kazakhstan^a (2006)
East Asia and Pacific	Vietnam (2003), Malaysia (2006)
Latin America and the Caribbean	Colombia (2002), Venezuela (2004)
Middle East and North Africa	Yemen (2001), Palestine (2002), Morocco (2004), Tunisia (2005), Egypt (2001)
South Asia	Sri Lanka (2004), Pakistan (2006)
Sub-Saharan Africa	Uganda (2004), Mauritius (2004), Namibia (2004), Niger (2005), Nigeria (2006)
Regional Studies	Africa (2004, 2006), South Asia (2006)

^aStudy carried out jointly with OECD.

Notes: Names in bold indicate studies exclusively dedicated to higher education, in the other cases higher education is part of a sector-wide education study.
From World Bank data.

Table 3 New commitments for education by subsector (fiscal year 01–06)

Subsector	IBRD+IDA new commitments (millions of current US\$)					
	FY01	FY02	FY03	FY04	FY05	FY06
Adult liter./nonformal ed	56	18	4	11	5	40
General education sector	435	442	639	355	507	457
Pre-primary education	32	32	102	25	88	147
Primary education	315	406	780	883	565	552
Secondary education	124	133	285	250	376	449
Tertiary education	41	268	524	62	361	263
Vocational training	91	85	15	98	50	82
Total	1095	1385	2349	1684	1951	1991

General education sector includes more than one subsector. About 50% of financing under general education is for primary education.

From World Bank EdStats. Retrieved April 2, 2007, from http://devdata.worldbank.org/edstats/wbl_A.asp

encourage higher education institutions to adopt a forward-looking strategic planning approach which helps them formulate well-conceived projects that are consistent with the overall direction of the institution.

Oftentimes, the Bank is able to act as a bridge builder, bringing to the discussion table stakeholders who do

Table 4 Lending for higher education between 2001 and 2006 (US\$ million)

Country	Total lending (% of total)	Top 10 largest borrower countries
Eastern Europe and Central Asia	37 (2%)	
East Asia and Pacific	109 (7%)	Indonesia (\$96m)
Latin America and the Caribbean	728 (48%)	Mexico (\$409m); Colombia (\$194m); Chile (\$87m)
Middle East and North Africa	88 (6%)	Tunisia (\$42m)
South Asia	335 (22%)	India (\$247m); Sri Lanka (\$30m); Afghanistan (\$38m);
Sub-Saharan Africa	223 (15%)	Ethiopia (\$37m), Ghana (\$30m)

From World Bank data.

not routinely talk to each other. In several countries, for example, the Bank has been instrumental in initiating a dialog between public and private universities, between universities and technology institutes, or between universities and employers. Similarly, in countries where the relationship between the government and the university sector is tense or even conflictive, the Bank sometimes manages to facilitate a constructive policy dialog on key issues, as happened in 2003 in Bolivia around the themes of quality enhancement and accreditation.

The Bank has also played a convening role at the regional level, as evidenced by the success of the global business school initiative in the Africa region and the launch of several regional quality-assurance networks (Asia and Latin America). Ideally, the regional policy seminars that the Bank organizes on a regular basis provide an effective forum for South–South knowledge sharing. In 2002–2003, for example, a series of seminars in South Asia helped stimulate reform efforts in Sri Lanka and Nepal. In 2005 and 2006, the Bank organized seminars in sub-Saharan Africa that helped disseminate successful reforms in individual countries and/or institutions.

While it is difficult to measure the direct impact of global publications such as *Constructing Knowledge Societies*, sometimes these policy documents can serve as catalysts for initiating reforms. In Pakistan, for instance, after the publication of *Peril and Promise*, the government established its own higher education task force whose findings and recommendations guided the launch of a comprehensive reform in 2003. In Yemen, the government received a small technical assistance loan (Learning and Innovation Loan) from the Bank in 2004 that facilitated the launch of

a nationwide consultation effort and the preparation of a long-term strategy for higher education reform.

The capacity of the World Bank to play these roles of convener and facilitator is due to several factors: the institution is able to rely on direct experience across a wide range of countries and situations; it interacts with client countries from a multisectoral perspective; it has learned to cooperate with multiple stakeholders; and it seeks to integrate its higher education work into the overall economic and social development framework of the concerned countries.

Conclusion

It is impossible to have a complete education system without an appropriate and strong higher education system . . . You have to have centers of excellence and learning and training if you are going to advance the issue of poverty and development in developing countries . . . the key . . . is higher education, not just on the technological side, but to create people with enough wisdom to be able to use it.

James D. Wolfensohn, former President of the World Bank (March 1, 2000).

In the past two decades, the World Bank has adopted a more holistic view of education, and higher education has become an essential part of the Bank's work in the sector, representing between 15% and 25% of the lending program in education depending on the year. Today, even countries that no longer need or want to borrow from the World Bank, such as China, Kazakhstan, or Malaysia, actively seek policy advice on various aspects of higher education reform, often guided by aspirations of becoming full-fledged knowledge economies.

In the coming years, the evolution of the World Bank's higher education program will likely be shaped by the following factors. First the Bank's higher education work will be increasingly framed within the overall context of the knowledge economy strategy first articulated in the 1998 World Development Report. Recent analytical work on innovation and competitiveness in Malaysia and Brazil, focusing on the role of higher education for human capital formation and the construction of strong national innovation systems, is symptomatic of this trend.

Second, in accord with a renewed Bank-wide focus on results, the success of higher education reforms will be assessed by looking at outcomes rather than considering only the reform measures themselves. This will require extensive work on the determinants of quality and the effectiveness of various quality assurance mechanisms, progress in measuring actual learning outcomes of students, and impact evaluations of innovative schemes

such as the new voucher funding in the US State of Colorado and Kazakhstan or the ProUni program in Brazil, which provides scholarships for low-income students financed through tax exemptions for private higher education institutions.

Third, the World Bank recognizes that, in higher education more than in any other area of education development, reform programs and innovative approaches have little chance of success unless careful attention is paid to the political economy of change. Stakeholder consultation and consensus building are as important as a technically sound design. Mozambique provides a powerful illustration of how a new minister of higher education managed to implement a comprehensive reform based on a strategic vision elaborated after extensive consultations and debates. Bank experience in several countries underscores the need for continuous policy dialog and vision sharing with all stakeholders to ensure the sustainability of reforms. Increasingly, the preparation of new Bank projects in support of higher education reform includes a social assessment exercise to identify the concerns of key stakeholders and inform the design of the project, as happened, for example, in Jordan, China, and Colombia.

Last but not least is the increased attention paid to global public good issues that affect higher education in all countries but that are often beyond the control of any one government. For example, the World Bank is supporting efforts to improve connectivity in Africa and to develop distance education opportunities through the African Virtual University. It is also helping establish regional networks of quality-assurance agencies in Asia, Latin America, and Africa and working toward consolidating these regional initiatives into a global network that would help developing countries enhance the quality and relevance of their tertiary education institutions through capacity building and experience-sharing activities. In attempting to address these global public good issues, the World Bank works increasingly in partnership with other multilateral agencies (OECD, UNESCO, AUF, etc.) and with international and regional associations of universities (IAU, AAU, SADEC, etc.).

Bibliography

- Bloom, D., Canning, D., and Chan, K. (2005). *Higher Education and Economic Development in Africa*. Washington, DC: World Bank.
 World Bank (1999). *World Development Report 1998/1999: Knowledge for Development*. New York: Oxford University Press.

Further Reading

- Saint, W. (2004). *From Aid to Global Sharing of Knowledge: Research Excellence and Commitment to Development*. Washington, DC: World Bank.

- Salmi, J. (1994). Shaping the higher education study: The dynamics of consultation. *Norrag News* **16**, 1–6.
- World Bank (1994). *Development in Practice Series: Higher Education: The Lessons of Experience*. Washington, DC: World Bank.
- World Bank (2002). *Constructing Knowledge Societies: New Challenges for Tertiary Education*. Washington, DC: World Bank.
- World Bank (2005). *Education Sector Strategy Update*. Washington, DC: World Bank.
- World Bank and UNESCO (2000). Higher education in developing countries: Peril and promise. *Report of the Independent World Bank/ UNESCO Task Force on Higher Education and Society*. Washington, DC: World Bank and UNESCO.

Relevant Websites

- <http://www.bc.edu> – Boston College, USA, Center for International Higher Education.
- <http://www.oecd.org> – OECD, Tertiary Education.
- <http://www.albany.edu> – State University of New York at Albany, Program for Research on Private Higher Education.
- <http://www.worldbank.org> – The World Bank, Education; Tertiary Education.
- <http://portal.unesco.org> – UNESCO, Higher Education.
- <http://www.utwente.nl> – University of Twente, The Netherlands, Center for Higher Education Policy Studies.

UNESCO's Role in the Development of Higher Education in a Globalized World

S Uvalić-Trumbić, UNESCO, Paris, France

© 2010 Elsevier Ltd. All rights reserved.

UNESCO and Its Global Mission

The United Nations Educational, Scientific and Cultural Organization (UNESCO) was created in 1945 in the aftermath of World War II. Although it was written several decades before world globalization entered the vocabulary, its Constitution describes a mission that balances global and national concerns:

That since wars begin in the minds of men, it is in the minds of men that the defenses of peace must be constructed; (. . .)

That a peace based exclusively upon the political and economic arrangements of governments would not be a peace which could secure the unanimous, lasting and sincere support of the peoples of the world, and that the peace must therefore be founded, if it is not to fail, upon the intellectual and moral solidarity of mankind; (. . .)

For these reasons, the States Parties to this Constitution, believing in full and equal opportunities for education for all, in the unrestricted pursuit of objective truth, and in the free exchange of ideas and knowledge, are agreed and determined to develop and to increase the means of communication between their peoples and to employ these means for the purposes of mutual understanding and a truer and more perfect knowledge of each other's lives; (. . .)

These principles are firmly embedded in UNESCO's Higher Education program. The formal establishment of the program goes back to a Resolution adopted by the 2nd General Conference of UNESCO in Mexico (6 November–3 December 1947) which mentioned higher education explicitly as one of the six main areas of work. Working with universities was perceived as a useful vehicle for promoting international understanding. Along with the other educational activities identified in these early statements, such as primary education, adult education, and teacher education, higher education is still at the core of UNESCO's strategy today.

A recent UNESCO world report underlines the fundamental role that institutions of higher education are destined to play in knowledge societies following radical changes in the traditional patterns for the production, diffusion, and application of knowledge. It reaffirms the unique position of UNESCO "as the only one in a position to carry out the tasks needed to ensure quality and

relevance of systems of higher education, while at the same time furthering the development of international cooperation in this field" (Towards Knowledge Societies, UNESCO, 2005).

This article will first look at the context of global higher education before giving examples of some of UNESCO's work.

Globalization and Higher Education: The Context

The trend to mass higher education is now the key driver in the worldwide development of the sector. Student enrolments are burgeoning and there are some 138 million tertiary students worldwide if part-time students are included (UIS, 2005). China and India have doubled enrolments in the past 10 years and China now has the largest higher education system in the world with some 27 million students in 2008.

Most countries want to join the knowledge society by following the example of developed countries where age participation rates in higher education of 40–50% are now perceived as necessary for sustainable development. Unfortunately, many countries in the developing world are far from reaching this percentage and government action alone cannot satisfy the rising demand. Indeed, government budgets for higher education are declining steadily on a per capita basis. An example from Africa demonstrates that due to the massive enrolment growth, public expenditure per student fell from US\$6300 in 1980 to US\$1241 in 1995 (Martin and Sanyal, 2006).

Burgeoning demand has encouraged a diversification in the ways higher education is provided and funded. Open universities, community colleges, and vocational schools are already making higher education more accessible, especially to working adults and those in remote rural areas.

One aspect of this trend is that private higher education (including for-profit institutions) is now the fastest growing element of the sector worldwide which already accounts for one-third of global enrolments. Although in Western Europe only 10% of students enrol in private higher education institutions (PHEIs), in other parts of the world they are now well established. For example, in Asia Pacific countries such as Japan, South Korea, Philippines, and Indonesia, 70% of students are enrolled in private institutions, although governments exercise

control and regulation over this private sector (Altbach and Levy, 2005). In Africa there are often partnerships between private and public institutions and IT academies like Cisco deliver certificates (Varghese, 2006). In Latin America 45% of enrolments are in private higher education, a figure that has risen from only 3% in 1950 (Levy, 2008), with Brazil, Colombia, and Argentina showing the highest enrolments in PHEIs.

Open and distance learning (ODL) and e-learning are growing in order to provide access to the wider student population now seeking higher education. Where there were only ten open universities in the world in 1988, by 2005 there were more than ten in India alone.

Growing student mobility is another dominant trend worldwide: 2.4 million students went abroad in 2004 – a threefold increase since 1980. African students are proportionately the most mobile, with one out of every sixteen studying abroad. The Global Student Mobility 2025 Report predicts that demand for international education will increase to 7.2 million students in 2025. However, some of these students will study with a foreign institutions in their own country. For example in 2004, 33% of all international students enrolled in Australian institutions studied in their own countries, a figure which was up from 24% in 1996.

This is an example of the growing mobility of programs and institutions. China saw a ninefold increase in foreign programs between 1995 and 2003, and in 2000 more Singaporean undergraduate students accessed a foreign program in Singapore than went abroad. With the rapid deployment of information and communications technology (ICT), ICT-enhanced Cross Border Higher Education (CBHE) is likely to become a significant development (Böhm *et al.*, 2003).

These trends in global higher education raise a number of policy issues for governments and institutions. Are private higher education, distance education, and cross-border provision the answers to expanding access to higher education in the developing world? What policies can governments and institutions adopt to ensure that these new providers make a positive contribution? How do countries address the challenges of Quality Assurance and the Recognition of Qualifications? What about the growing threat of degree mills and bogus institutions to the integrity of national tertiary systems? How can we draw a line between institutions providing substandard programs that are susceptible to improvement and the out-and-out degree mills?

UNESCO's Responses to These Trends

Policy debate

The World Conference on Higher Education (WCHE) that convened in Paris in 1998 gave a new momentum to policy debate on higher education. It brought together 4000 participants from 182 countries and provided a

comprehensive forum for policy debate. Education ministers were joined by other stakeholders, recognizing that governments could no longer manage alone the most radical renewal of higher education they have ever been required to undertake.

The 1998 conference adopted a World Declaration on Higher Education for the Twenty-first Century: Vision and Action, supported by an action plan (UNESCO, 1998). It outlined a conceptual framework based on a set of basic principles: broadening access to higher education as a key factor of development, both as a public good and also as a human right; promoting reforms of higher education at system and institutional levels to enhance quality, relevance and efficiency; and securing adequate resources and funding – both public and private – to cope with the increased demands placed on higher education by its different stakeholders.

As the twenty-first century dawned, globalization became another buzz-word in higher education: globalization as a form of greater interdependence as well as globalization as contrasted to internationalization. For some, globalization, in the sense of a global market, is seen as a threat to the traditional values of higher education, notably to the notions that higher education is a public good and therefore primarily a public responsibility, and that access to higher education based on merit is a basic human right.

As the nation-state loses its monopoly over higher education provision, there are concerns, for example, about how an agenda of social inclusion can be pursued in the context of the increasing commercialization of higher education. This has generated heated political debate about national sovereignty over higher education. The most common response is for states to focus their attention on quality, assessment, and standards, through quality assurance mechanisms and accreditation systems.

International organizations have had to position themselves in this worldwide debate, particularly after the General Agreement on Trade in Services (GATS), under the World Trade Organization (WTO), included higher education as one of the 12 tradable services under Education. A key event in the process was “OECD/US Forum on Trade in Educational Services” held in Washington in May 2002. Acrimonious discussions, pitting the academic and trade communities against each other, put the issue of the nature of higher education firmly on the table.

At the Conference, however, keynote speakers and leading government representatives, although stressing the opportunities for trade in education, also emphasized quality assurance, accreditation, and qualifications recognition as key elements in the trade-in-education agenda. They highlighted the importance of UNESCO's role in this global debate and urged that the nexus between trade and education be seen in a common and wider perspective, so as to maximize its benefits and minimize

possible disadvantages. It was also proposed that OECD and UNESCO should work together.

Following an expert meeting in September 2001, UNESCO pursued this agenda by launching the UNESCO Global Forum on International Quality Assurance, Accreditation and the Recognition of Qualifications in Paris in October 2002. This took the debate forward and strengthened UNESCO's mandate to address these issues. The opening address by Sir John Daniel, Assistant-Director General for Education, set the tone of the debates with a positive approach to globalization in higher education, based on the UN position on globalization, enshrined in UNESCO's Medium Term Strategy (globalization with a human face; globalization to be made to work for all). He put forward certain basic principles, combining the notion of higher education as a public good with the conviction that "new need not be bad" and that globalization promotes competition, which in turn creates diversity (Daniel, 2002).

The Global Forum was conceived as a response to the ethical challenges and dilemmas facing higher education in an era of globalization. Its mission was to provide a platform for exchange between different partners and to initiate debate on the social, political, economic, and cultural dimensions underpinning globalization and higher education.

Compared to the frustration felt by civil society and academia at the OECD Forum in Washington, the launch meeting of UNESCO Global Forum was more harmonious. It was successful in bringing together a wide variety of stakeholders ranging from for-profit providers such as the Apollo Group's Phoenix University, corporate universities such as Cap Gemini Ernst & Young, and the private sector such as the University Relations Division of Hewlett-Packard, to the UNESCO's traditional partners such as public higher education institutions, teachers' and students' associations. Both developed and developing countries were well represented.

The forum ended with unanimous support for an action plan. The participants agreed on the need to build bridges between education (i.e., academic values and principles) and trade in higher education services and that UNESCO, the WTO, and the OECD could complement each other in providing a joint forum to assess both the cultural and commercial aspects of trade in higher education.

The two subsequent meetings of the Global Forum (UNESCO, Paris in 2004 and Dar es Salaam, Tanzania in 2007) took these discussions further. They addressed respectively the issues of capacity building in higher education and responding to learners needs in new higher education spaces.

A third OECD Forum on Trade in Education was hosted by Australia in Sydney in October 2004 as a collaborative effort of the OECD, UNESCO, and Australia. The event was a good demonstration of unity and the

value of partnerships and showed that in the 2 years since the Washington event, the world had moved on in a constructive manner.

To make UNESCO's position clear an education sector position paper on "Higher Education in a Globalized Society" was prepared as one in a series of position papers that aimed to give UNESCO's views on some key issues in contemporary education. Based on reviews of trends, issues, global debates, and regional concerns through case studies, as well as UN and UNESCO standard-setting instruments and principles, UNESCO stated its position as follows:

Higher education in a globalized society should assure equity of access and respect cultural diversity as well as national sovereignty. In addition, UNESCO is committed to assuring the quality of global provision of higher education in an increasingly diverse higher education arena and raising the awareness of stakeholders, especially students, on emerging issues in this field. This position aims to establish the conditions under which the globalization of higher education benefits all.

The 1998 World Conference generated the establishment of two powerful, independent and self-financing networks, supported by extra-budgetary funding: the UNESCO Forum on Higher Education, Research and Knowledge with a Secretariat in the UNESCO Division of Higher Education but with a wider interdisciplinary approach, and the Global University Network for Innovation (GUNI) network with a Secretariat at the Technical University of Catalonia (UPC) in Barcelona. Both networks generate research and policy debate on issues related to globalization and higher education. However, the impact and outreach of these two networks deserve a more in-depth analysis which is beyond the scope of this article.

Standard setting

As part of its core functions of setting standards, building capacity at national and regional levels, and serving as a clearinghouse, UNESCO has grounded its responses to the new challenges in global higher education in the existing legal framework that is supported by the six regional conventions on the recognition of qualifications that have been ratified by over 100 member States worldwide. Although the ultimate objective of the conventions was to develop a universal convention covering the whole world, this is still an unrealistic objective, so the regional conventions are now being revised and regional frameworks are being strengthened in response to the rising challenges of globalization.

An important result of the Washington OECD/US Forum on Trade and the first meeting of UNESCO's Global Forum was the decision by the Education sectors of the two organizations to work together to develop

Guidelines for Quality Provision in Cross-Border Higher Education. The Guidelines address six groups of stakeholders in higher education: governments, higher education institutions, student bodies, recognition bodies, quality assurance bodies, and professional bodies. They recommend actions based on collaboration, foster mutual trust and confidence, and encourage access to reliable and transparent information. The overall aim is to promote quality in the growing phenomenon of cross-border higher education.

Global initiatives for capacity building in quality assurance

The 2002 Global Forum on International Quality Assurance, Accreditation and the Recognition of Qualifications created a space for policy debate. Its third meeting, held in Tanzania in 2007, brought together stakeholders around the theme, Guiding Learners in New Higher Education Spaces: Challenges for Quality Assurance and the Recognition of Qualifications. Since the key theme of the Global Forum was empowering learners, it discussed issues ranging from mobility and migration to academic fraud.

The meeting urged a strong focus on capacity building for quality assurance at all levels and the 2006 Guidelines for Quality Provision in Cross-border higher education jointly developed by UNESCO and the OECD were held up as a useful tool for this work as CBHE spreads. The Guidelines have been acknowledged as such within the Bologna Process in Europe. They featured at the last Bologna Ministerial Conference in 2007 and in the 2007 the London Communique, as part of the global context of the Bologna process.

In early 2008, UNESCO and the World Bank launched a new partnership, the Global Initiative for Quality Assurance Capacity (GIQAC). This aims to create a global framework to support capacity development in developing and transition countries. In the first year, GIQAC will support regional Quality Assurance networks in Africa, the Arab States, the Asia/Pacific region and Latin America and the Caribbean, as well as the International Network for Quality Assurance Agencies in Higher Education, INQAAHE. Capacity-building activities within GIQAC are also based on sharing experiences across regions and promoting South–South cooperation.

Clearinghouse: The UNESCO portal of higher education institutions

To further its responsibility for protecting and empowering learners, UNESCO launched a pilot project for a portal of recognized higher education institutions in 2006 and presented the first results in 2008. This work was a logical extension of its earlier work with the OECD on the Guidelines for Quality Provision in Cross-Border Higher Education as well as of its Study Abroad publication that includes information tools for students, directories and listings of

ODL institutions, financial aid and information on bogus providers.

The aim of the UNESCO Portal is to make up-to-date, accurate, and comprehensive information on recognized higher education institutions/providers available at the international level. It provides authoritative data on the status of higher education institutions and quality assurance systems in countries around the world in order to help students make informed decisions about undertaking higher education (including cross-border higher education), and to protect them from misleading information, biased guidance, low-quality provision, rogue providers, and qualifications of limited validity.

The pilot phase of the portal included access to information on institutions and systems in Argentina, Australia, Canada, China, Egypt, Jamaica, Japan, Kenya, Malaysia, Nigeria, Norway, Switzerland, the United Kingdom, and the United States of America. The portal was launched in April 2008, now counts 20 countries and is growing rapidly. The country information on the portal is managed and updated by the competent authorities in the participating countries.

Information on the national processes for recognizing or otherwise sanctioning institutions is available on the country pages. These define key terms, indicate where information can be found, describe the national higher education system, list recognized institutions and programs, and provide information for students planning to study in the country (including details on foreign credential assessment and recognition, opportunities for financial assistance, and provisions for cross-border higher education).

What about degree mills?

Fraudulent and bogus providers, popularly known as degree mills, threaten the credibility of the online provision of courses internationally by offering costly credentials and degrees of no educational value. Electronically delivered degrees are largely unregulated and pose serious challenges around the world. Although comprehensive and reliable data on degree mills are not available, thousands of degree mills are estimated to be in operation. The growing and often unsatisfied demand for higher education has created a significant market for such providers but few governments or organizations are positioned to do what is necessary to educate and protect the public.

These fraudulent providers often misuse the names of international organizations, in particular UNESCO, in various ways. These range from claiming a connection with UNESCO where none exists to exaggerating a real link with UNESCO to give the impression that UNESCO accredits them. UNESCO is not an accrediting agency.

The US Council on Higher Education Accreditation (CHEA) is working with UNESCO to address the

problem of degree mills and accreditation mills, beginning with the development of some general international principles. The aim is to develop suggestions for effective practice to assist the international community in confronting bogus providers. A possible outcome is the establishment of an ongoing, reliable network of higher education, quality assurance/accreditation bodies, and international organizations that can identify degree mills and share information and ideas.

The Next Milestone: The 2009 World Conference "The New Dynamics of Higher Education"

Is higher education today really a driver for development – for sustainable development – nationally and internationally? Does the sector live up to the expectations that it will foster change and progress in society and be a key architect of knowledge societies? How does higher education contribute to the development of education systems as a whole? What are the most significant trends that will shape the new higher education and research spaces? How are learners and learning changing? What are the new challenges for quality?

These issues will be on the agenda of the World Conference on Higher Education that is to be held from 5 to 8 July 2009 in Paris. The conference will take stock of changes in higher education since the 1998 World Conference on Higher Education and will address the new dynamics that are likely to shape the strategic agenda for the development of higher education policies and institutions for the foreseeable future.

The 1998 World Conference on Higher Education was a special moment for the international higher education community. However, the policy agenda has evolved considerably since 1998 and today the sector is a greater priority than ever. Higher education faces many challenges, both new and ongoing. Understanding them better will help to shape action at the global, regional, national, and institutional levels.

The 2009 world conference will provide a global platform for future-focused debate and thinking about the rapidly changing higher education and research spaces in order to identify the actions necessary to ensure that the sector meets both national development objectives and individual aspirations. The conference will bring together key stakeholders in a new commitment to the development of higher education. In preparation for this event, a series of regional conferences will be organized to bring specific regional concerns, expectations, and proposals to the 2009 world conference.

At the time of the preparation of this article, several significant regional and subregional events have already taken place as part of the build-up to the 2009 event: The Regional Conference on Higher Education and the Caribbean (CRES, 2008) held in Cartagena de Indias,

Columbia, on 4–6 June 2008; The Southeastern European Forum and High-Level Roundtable on Science, Higher Education and Innovation Policy in Budva, Montenegro on 1–3 July 2008; the Asia-Pacific Sub-regional Preparatory Conference in Macao, SAR China on 24–25 September 2008, and the Regional Conference on Higher Education in Africa (CRESA, 2008) in Dakar, Senegal (10–13 November 2008). These conferences revealed some convergence of thinking on the key challenges faced by higher education and also identified specific regional concerns.

This suggests that the key challenge for the 2009 WCHE will be to establish a proper balance between the global and local missions of higher education and reveal how UNESCO can best play its role in leading the world toward a new commitment to Higher Education and Research by sharing responsibilities with the different stakeholders.

Conclusion

For many years the primary educational focus of the international community and of UNESCO as its lead agency for education has been the achievement of the goals of Education for All (EFA), most especially Universal Primary Education. So far higher education has seemed somewhat marginal to that work but this is now changing rapidly for two main reasons.

First, the growing success of the campaign for Universal Primary Education has led to large increases in the numbers of children seeking education beyond that level. Ministers in many countries are grappling with the challenge of expanding the provision of both secondary and tertiary education to meet the coming surge in demand.

Second, there is increasing acknowledgment that "at no time in human history did the welfare of nations depend in such a direct manner on the quality and outreach of their higher education systems and institutions" (UNESCO, 2004). Many Higher Education Ministers are coming to UNESCO with requests for help with policy development to strengthen and sustain their national higher education systems. As Professor Mahmood Mamdani told the 17th Conference of Commonwealth Education Ministers: "If your object is to transform general education, you have to begin with higher education. For higher education is the strategic heart of education; it's where choices are developed" (Commonwealth Secretariat, 2006).

International organizations are repositioning themselves accordingly. The Secretary-General of the UN, Ban Ki-moon, recently announced that the UN is developing an initiative called Academic Impact aimed at building stronger ties with institutions of higher learning, believing that they can foster the development of global

citizenship and encourage higher education to respond to the global challenges of this millennium.

We hope that these global political realities will stimulate the renewal of UNESCO's original mission of intellectual solidarity and empower it to continue its successful tradition of support of higher education by playing a leading role in the new circumstances of a globalized world.

Bibliography

- Altbach, P. G. and Levy, D. C. (2005). *Private Higher Education: A Global Revolution*. Boston, MA: Sense Publishers/Boston College Center for International Higher Education.
- Böhm, A., Meares, D., et al. (2003). *Global student mobility 2025: Analysis of global competition and market share*, Sydney, IDP Australia.
- Commonwealth Secretariat (2006). *The Cape Town Communiqué. 16th Conference of Commonwealth Education Ministers*, 14 December.
- Daniel, J. S. (2002). *Proceedings, First Global Forum on International Quality Assurance, Accreditation and the Recognition of Qualifications in Higher Education*, UNESCO, Paris, France.
- Levy, D. C. (2008). *Private Higher Education's Global Surge: Expanse and Types. Presentation for UNESCO Expert Group on Private Higher Education*. UNESCO, Paris, 20th November, 2008.
- Martin, M. and Sanyal, B. (2006). *Financing Higher Education in World Higher Education Report 2006*. (GUNI).
- UIS (2005). UNESCO Institute of Statistics.
- UNESCO (1998). *Higher education in the twenty-first century: Vision and action. World Declaration on Higher Education for the Twenty-first Century*.
- UNESCO (2004). *Higher education in a globalized society. UNESCO Position Paper ED-2004/WS/33*. UNESCO, Paris, France.
- UNESCO (2005). *Towards Knowledge Societies, UNESCO World Report*. Paris, France.
- UNESCO (2006). *Guidelines for Quality Provision in Cross-border Higher Education*. UNESCO, Paris, France.
- Varghese, N. V. (ed.) (2006). *Growth and Expansion of Private Higher Education in Africa*. IIEP, UNESCO.

HIGHER EDUCATION – THE ORGANIZATION AND FINANCING OF HIGHER EDUCATION

Contents

Higher Education in Federal Systems

Pathways and Articulation into Higher Education

Unitary, Binary and other National Systems of Higher Education

Viewing Private Higher Education: How Much, Where, Why, and What?

Higher Education in Federal Systems

M A M Cortés and U Teichler, University of Kassel, Kassel, Germany

© 2010 Elsevier Ltd. All rights reserved.

Relationships Between Central and Decentral Powers and Responsibilities

Of the nearly 200 countries worldwide, about 20 are characterized by a federal political system, and about 20 others with a dominant allocation of the political power at the central level have some federal features. With the exception of China, federal systems prevail in the largest countries of the world, but can also be found in small countries, for example, Switzerland.

In centralized or unitary systems, various governmental levels might exist as well, but the provincial and local levels are subordinate to the central government. Federal systems are based on a share of ultimate power; in other words, the central institutions ultimately derive their authority from the approval of the constituent decentral units. Federal systems vary according to the degree of rights and powers allocated centrally or decentrally as well as to the degree of overlapping and shared responsibilities and activities. Moreover, modes of funding vary substantially.

In federal systems, international relations and defense tend to be national prerogatives. Often, national governments have a strong say in economic matters as well. In contrast, powers for education are often allocated decentrally. Higher education and research are treated in some countries as a subsector of education, while in others, they might be placed separately. In almost all federal systems, both the central government and decentral governmental units (provinces, states, *Länder*, *Kantone*, etc.) are involved in higher education. Even in the Federal Republic of Austria, where the coordination, supervision, and most public funding of higher education rest with the central

government, the provinces might establish and fund universities of applied sciences (*Fachhochschulen*).

In most analyses of higher education in federal systems, a distinction is made regarding the central and decentral role with respect to:

- the overall coordination of the system (e.g., safeguarding similarities of study programs and credentials);
- the supervision and possibly the establishment of individual institutions; and
- funding of higher education.

Some analyses lay prime emphasis on the allocation of rights and powers as well as on the respective political debates. Others aim to establish the extent of homogeneity or heterogeneity of systems with respect to key features of higher education, for example, types of study programs, qualifications of academic staff, and modes of administration (Brown *et al.*, 1992; Heidenheimer, 1992).

Most federal systems are characterized by a certain degree of fuzziness of the central and decentral territories of power and influence. Decentral governments often try to defend or increase their influence by pointing at cultural and educational diversity and at the needs of communities, while central governments try to stabilize or enlarge their powers by pointing at international dimensions of higher education and its economic relevance. Often, national governments ensure their influence through their superior financial resources.

As the political roles of the central and decentral governments as well as the extent of similarity or variety differ substantially by country, their characteristics can be best described through the presentation of cases. Hence, five countries are presented subsequently.

United States of America

In the United States of America, the individual states have the legislative and supervisory power of higher education. Public universities, as a rule, are state universities. The decentral emphasis of public higher educational policies is visibly underscored by the fact that citizens of the state, where the public university is located, pay substantially lower tuition fees than out-of-state students. Private universities, enrolling altogether more than one-third of students, dominate among the most prestigious universities, but are also represented on other levels of quality and reputation.

No formal interstate coordination exists. The federal government runs colleges only for its own administrative purposes (e.g., military colleges). It does not have any coordination rights as regards the higher education system, but it exerts a substantial influence by the power of the purse (see Heidenheimer *et al.*, 1990). It funds student aid, including tuition fees, thus making students the carriers for institutional support. The federal government is the key funding source of research in many areas, and it provides funds for innovation programs in higher education. As these funds are often conditional as regards institutional policies (e.g., affirmative action), they are more influential than mere financial support schemes.

Nationwide coordination in the US is to a large extent based on informal communication and cooperation among diverse actors, for example, umbrella organizations of higher education institutions, regional accreditation agencies, academic and professional agencies, and more or less visible lobbies. System changes, such as the emergence of a postdoctoral career stage, often are initiated not through legislation, but through communicative processes, and funding initiatives and experiments.

Altogether, the role of system coordination, legislative steering, and state supervision for the overall character of the higher education system in the US is viewed as relatively weak. Most states steer and supervise higher education more loosely than in other countries; the individual institutions have enormous room for maneuver, but are expected to be accountable in providing evidence that public support was used effectively and efficiently. Moreover, private donors, industry, churches, communities and other organizations, as well as individuals play a major role in funding and often claim legitimate stakeholder influence regarding the development of the individual higher education institutions.

As a consequence, an enormous degree of vertical diversity among institutions, often combined with a variety of institutional profiles, is more highly accepted in the US as a virtuous element of higher education than anywhere else. This reinforces the view that strong political powers of coordination are not needed. The federal political system keeps top-down coordination and supervision

of higher education in the US in bounds and reinforces diversity of higher education.

Canada

In Canada, the public right of supervision and of establishing higher education institutions rest with the individual provinces. Many experts suggest that there has never been any higher education system in Canada. The federal government is not involved in any formal mechanism of coordination and supervision of higher education. It is only directly in charge of military colleges. However, it has been holding a strong power for a few decades in the overall coordination of research policies and in research funding, and it takes initiatives here and there for funding innovation programs in higher education (see Fischer *et al.*, 2006; Jones, 2006).

Canadian higher education has been relatively homogeneous for a long time. The university sector is highly uniform in spite of substantial cross-border interaction with the diverse US higher education system, although some differences are noteworthy between the Anglophone provinces and in the Francophone province Québec. In the process of higher education expansion, however, this uniformity was gradually blurred, and the newly established nonuniversity institutions vary substantially between the provinces.

The gradual increase of diversity is linked to the fact that the provincial governments had favored the existence of autonomous higher education institutions being steered more strongly through funding and casuistic regulation mechanisms rather than through heavy-handed legislation and homogeneity-oriented reform concepts.

From the late 1980s onward, Canadian higher education began to suffer serious financial constraints caused by cuts in the federal transfer of tax revenues to the provinces. Although the so-called Established Program Financing inaugurated in 1976 ensured the full discretion of the provinces to determine how the funds made available by the federal government actually are spent, the universities were affected by these cuts. The Canada Assistance Plan of 1995 practically ended the role of federal transfers being the major source of basic funding for Canadian post-secondary education. However, concurrently, the federal role grew through enlarged funding of research and the physical infrastructure. Curiosity-based research was largely funded by federal monies transferred to research councils in charge of quality-based competitive distribution of these funds. Overall, the federal government created a variety of other financial support schemes, some of which could exert substantial influence on the characteristics of the higher education system.

For example, the Canada Foundation for Innovation was created in the mid-1990s to provide massive infrastructure support for programs linked to the private

sector. The Canada Research Chairs Program, established in 2000, provided salary and some infrastructure support for the creation of new research chairs at universities in order to help them attract leading international scholars and retain leading domestic professors.

The increase of federal incentive programs for higher education was accompanied by the establishment of varied similar schemes of the individual provincial governments. Combined with the expectation that universities have to seek increasingly for nongovernmental sources, a low-key supervisory attitude of the Canadian provincial governments *vis-à-vis* higher education emerged, which contributed to an increasing diversity of higher education both within and across provinces. Within the provinces, the diversity is most pronounced with respect to the intensity of research and the character of nonuniversity higher education. Across provinces, it is worth noting that, for instance, Alberta, Ontario, and British Columbia increased their tuition fees, created performance-funding mechanisms, and other initiatives to strengthen market forces, and took various measures to support linkages between higher education and industry. In contrast, Québec and Manitoba maintain low tuition fees as a mechanism to encourage access.

Germany

When the Federal Republic of Germany was established in 1949, the *Länder* exclusively were in charge of higher education, notably of regulation and supervision, the establishment of individual institutions, and the funding of the higher education system. National system coordination was the task of the Permanent Conference of the Ministers of Education and Culture of the *Länder* (KMK), which actually was active in areas of access and admission, formal structures of degree programs, etc. The federal government supported international cooperation in higher education and research, and, through its responsibility for the national coordination of the civil service, shaped the employment conditions for the staff.

In the course of the 1950s, the federal government got involved in research promotion. Mechanisms were established through which the federal government and all *Länder* governments together each funded 50% of the major research institutes outside the university and a research-promotion program for projects in higher education institutions. Similarly, the federal government took over 65% of the costs for a newly established need-based student-aid system, and it established its own promotion system of applied research.

In the late 1960s, a change of the constitution led to the establishment of joint responsibilities of the federal and the *Länder* governments in the coordination of the higher education. In addition to the funding schemes

previously established, this comprised the inauguration of the Federal-State Commission for Educational Planning and Research Promotion (BLK). Both governmental levels cooperated and shared the costs for construction of new university buildings. In addition, they became jointly responsible for setting frameworks for study programs, institutions, staff, and decision making as well as other similar matters: The two chambers of the national parliament – the representatives elected nationally and the representation of the governments of the individual *Länder* – eventually agreed on a Framework Act of Higher Education (HRK, for the first time enacted in 1976) serving as a binding guideline for the laws of the individual *Länder*, that is, those actually regulating higher education (see Teichler, 1992).

The complex system of federal–*Länder* interaction in the coordination and funding was in part due to the fact that the federal government could take over the rising costs more easily. Moreover, federal–state cooperation could contribute to a balance between two conflicting constitutional principles affecting higher education: to be embedded in regional cultural surroundings and to safeguard a homogeneity of living conditions in Germany as a whole; the latter, for example, with respect to the entry qualifications and admission, the financial conditions for study, any element of higher education relevant for the freedom of occupational choice. A relatively strong nationwide coordination of higher education was widely viewed as necessary to counteract centrifugal powers in the wake of higher education expansion, because a more or less even quality among universities had prevailed, whereas intrainstitutional diversity among staff and among students as well as the two-type structure of universities and universities of applied sciences (*Fachhochschulen*) were seen as appropriate arenas of diversity.

Over the years, the *Länder* repeatedly tried to reduce the power or even exclude the federal government, when new needs for system-wide coordination arose. They tried to set up new mechanisms of inter-*Länder* coordination instead of federal–*Länder* coordination. In reverse, the federal government often took new needs for funding and innovation as a starting point for the establishment of new federal or federal–*Länder* promotion schemes. At the beginning of the twenty-first century, many federal and *Länder* reform initiatives triggered off time-consuming fundamental debates about the rights of the central level and the *Länder* level. Moreover, the federal government and the *Länder* governments came to the conclusion that increasing joint legislation and joint funding schemes in many political sectors have led to overly complex decision-making processes. Consensus was reached in 2005 to reduce the number of areas of joint decision making (see Kehm, 2006). Actually, most legal rights of the federal government with respect to higher education were discontinued, while the systems of joint funding

of research institutions, research promotion, student aid, and construction of buildings were retained. Even new schemes of joint funding were agreed upon, notably the so-called excellence initiative for preferential funding of top universities, research clusters, and graduate programs, as well as a program to provide funds to cope with the expected temporary rise of student numbers for a limited number of years. The funding of the latter task, traditionally solely in the domain of the *Länder*, was justified as emergency measure only for a period of exceptional enrolment growth eventually offset by a demographic decline thereafter.

Switzerland

Switzerland is a country where the national identity is deeply rooted in the belief that the ultimate political power rests with the 23 cantons, while the national government clearly derives its power either through agreements on the part of the cantons or from national popular votes on individual political motions. This belief is reinforced by Switzerland's multiculturalism and the existence of four linguistic regions: the German-speaking, the French-speaking, the Italian-speaking, and, the small, Romanisch-speaking regions. Actually, some experts point out that the Swiss universities in the French-speaking area have more in common with French universities than with the Swiss universities in the German-speaking area which are in various respects similar to German universities.

The individual cantons are in charge of establishing, funding, and supervising educational institutions including universities. Actually, only 10 of cantons run universities, that is, those in Basel, Berne, Fribourg, Geneva, Lausanne, Neuchâtel, Zurich, St. Gall, and Lucerne as well as the *Università della Svizzera Italiana*; this created quite an uneven financial burden. In 1979, an intercantonal agreement was signed according to which major parts of the costs for study places are paid by the student's home canton.

Cantonal governance promotes an increase of differences between cantons. The structure and duration of secondary education varies, and participation rates differ enormously. Student aid provided by the cantons is heterogeneous. For many years, the motivation for higher education was different among language regions; it remains higher in French- and Italian-speaking regions than in German-speaking areas because people in the latter have more nonacademic routes of social and economic mobility.

The autonomous university sector is not the only one in Switzerland; another important part is that of federal institutes of technology. The *Eidgenössische Technische Hochschule* Zürich (ETHZ) and the corresponding French-speaking institution in Lausanne are supervised and financed by the federal government. This is based on article 34 of the constitution, where the responsibility for general regulation of vocational training in industry, crafts,

commerce, agriculture, and domestic economy is attributed to the Confederation.

In the 1990s, former advanced vocational training institutions were upgraded to universities of applied sciences (*Fachhochschulen* or *hautes écoles spécialisées*). In 1999, the constitution was revised to allocate the responsibility for these new institutions and advanced vocational training institutions to the Confederation. While the Department of Home Affairs of the Confederation is in charge of matters of universities and science, the responsibility for universities of applied sciences rests with the Department of Economic Affairs.

The Confederation plays a major role through funding. It funds the research-promotion system, and provides the public means for covering the regular costs of the institutes of technology and the public universities of applied sciences. With respect to the universities not under federal supervision, the Confederation covers minor parts of the operational costs and provides substantial subsidies for capital costs and equipment. Moreover, the so-called extraordinary subsidies are made available for activities such as training and research in informatics, continuing education, ecology and environmental sciences studies, and national and international university mobility.

The cantons cooperate in the Conference of Cantonal Ministers of Education (CDIP) which, for example, coordinates teacher training. The major nationwide coordination bodies are the Swiss University Conference (CUS), a cooperative common agency of the Confederation, the cantons and the universities, and the Council for Universities of Applied Sciences (Conseil UAS) (see OECD, 2003).

Since 2002, the Swiss Confederation stimulates nationwide discussions and initiatives for constitutional reforms in order to strengthen the overall system coordination of the various sectors of higher education. This was considered necessary in the wake of increasing international cooperation and global competition in higher education.

Australia

Most higher education institutions in Australia are public. They have been established by an act of the state government in charge, and they had been funded traditionally by state governments, and even now they are required to present an annual report to their state government. The most widely known exception is the Australian National University which is under direct supervision of Commonwealth, that is, the national government. Australia, however, moved relatively early to a hard federalism with respect to higher education. Since 1974, the financial responsibility of all public universities rests with the national government, and it established over the years changing advisory or buffer institutions not only for the sake of allocating basic funds to higher education, but also for many other issues, for example, research promotion, the institutional patterns of the

higher education system, and improving the links between higher education and social and economic priorities of the country. Over the years, however, state government became more strongly involved in the funding of their institutions of higher education, whereby they often tried to ensure a closer link between the activities of higher education institutions and the specific policies and perceived needs of the individual state.

There are three types of higher education institutions: universities, self-accrediting providers (other than universities), and non-self-accrediting providers. The terms self-accrediting and non-self-accrediting refer to the freedom with which these institutions are able to accredit their own awards; universities are also self-accrediting providers. Non-self-accrediting providers are authorized by state or territory legislation, and they offer at least one course that is accredited as a higher education award. These three types of institutions are present at the private sector as well. There are 39 universities – 37 of them are public and two are private.

Since the 1980s, successive federal governments attempted successfully to change universities into corporate-style entities and to wean them from federal funding. This reduced the financial burden of the federal government, and close and detailed supervision of public funds gave way to a steering from a distance. As a consequence of these changes, universities had an increased *de facto* autonomy from governmental policies but, according to some experts, initiatives and new strategies to raise funds increased the influence of some stakeholders, notably that of industry. Moreover, the Australian universities increased their revenues through taking in vast numbers of fee-paying foreign students.

Nowadays, the Australian higher education system is strongly shaped by market principles. The federal government has structured the rules for competition for funds and prestige, it shows the assumption that this competition drives efficiency, forces accountability to industry and student-clients, and secures production of public goods. The former binary system of universities and colleges of advanced education was replaced from 1989 on by a unitary system somewhat stratified as a consequence of increased competition and additional income generation (see Marginson, 2006).

The rules for competition and a call to align to national priorities are set in *Our Universities Backing Australia's Future*, a document elaborated in 2003 on federal level by the minister of education, science, and training (Nelson, 2003). The Commonwealth also created a new grant system spelling out the federal contributions to the numbers of study places annually negotiated between each institution of higher education and the Commonwealth and state governments.

The Commonwealth also provides student aid, for example, in subsidization of loans and help for fee-paying students in Australia and abroad. It funds special initiatives such as the establishment of the National Institute

for Learning and Teaching in Higher Education and awards for centers of excellence. Moreover, research funding is a federal task, whereby competitive funding is customary and the individual institutions are encouraged to be more flexible and responsive in developing a strategic portfolio of research activities and training programs.

Altogether, the Australian universities have to report to state governments in many respects. However, the allocation of study places, student loans, and research funding and promotion, including quality audit, is done on a federal level. Moreover, the federal level is influential through various ways of guiding the system, for example, as regards access and equity, and various initiatives, for the promotion of excellence, for example, in teaching and learning.

Mexico

Mexico is officially a federal system, however, its higher education system is highly centralized and, at the same time, it has fostered heterogeneous and unequal conditions among its states and among the subsystems of higher education. For many years, the Mexican federal government, after public protests against the repression of student protests in 1968, avoided interference into university matters. In the 1980s and at the beginning of the 1990s, however, new rules for the regulation of interactions between the state and higher education were created.

The Mexican higher education system is constituted by six subsystems, in four of them, the federal government has functions of direct coordination, namely the subsystem of public universities, which includes the three federal universities (*Universidad Nacional Autónoma de México* (UNAM), *Universidad Autónoma Metropolitana* (UAM), and *Universidad Pedagógica Nacional* (UPN)) and the state universities; the subsystem of technological education, which is coordinated at the federal level by the subsecretary of the Ministry of Public Education; the subsystem of other public institutions, such as navy and army academies and the National Institute of Fine Arts; and the subsystem of higher schools for teacher training called *Normales*. The other two subsystems are the subsystem of private higher education institutions, which does not receive any federal funding and is regulated only for the official recognition of its programs; and the subsystem of technological universities, which was created in the 1990s following a joint initiative by the federal government and the state governments.

The centralism of the Mexican higher education system is most obvious in the modes of funding. The federal government funds the federal universities completely, among them the large UNAM and the UAM, and technological institutions are financed by more than 50%. Altogether, the major public sources for higher education in Mexico are: (1) the ordinary and extraordinary federal subsidy,

(2) the ordinary state subsidy, and (3) the subsidy for enlarging educational facilities, with federal and state funds.

The creation of an extraordinary federal subsidy in 1984 strengthened the federal role. Previously, higher education institutions were free to spend the funds for salaries, research, facilities, etc. Under conditions of economic crisis, extraordinary federal subsidies were viewed as necessary, but these were provided for specific purposes.

The National System of Researchers (SNI) was created in 1984 providing incentive funds through the National Council of Science and Technology (CONACYT) for the production of publications and for obtaining a PhD degree. In 1990, the Fund for the Modernization of the Higher Education (FOMES) was established to finance projects of institutional development as well as ESDEPEP, a system of incentives for teaching performance, which is run by the Ministry of Treasury and Public Loan (SHCP). In 1996, the Program for Improvement of Professors (PROMEP) was created, providing professors of public universities with scholarships to carry out high-quality postgraduate studies. Since 1998, the Program of Support for University Development (PROADU), also coordinated by the Ministry of Public Education, provides funds for various purposes of institutional development. All these funds aim to increase competition between universities (see Kent and Richardson, 2002; Casanova, 2006).

The creation of the – rapidly growing – subsystem of technological universities at the beginning of the 1990s, although funded jointly by federal and state governments and operated under the coresponsibility of federal and state authorities, is generally viewed as a step toward decentralization. These institutions are expected to promote regional development, and many of them were established in localities without other higher education institutions. In the late 1990s, the State Commissions for the Planning of the Higher Education (COEPES) were reactivated: they delegate responsibilities to state governments and aim to develop the strategic administration of the state institutions, approve new proposals of academic programs and educational models, elaborate, disseminate, and evaluate state policies of higher education, and implement actions and projects for the development of higher education in a specific state. Concurrently, the federal government decided to subsidize public programs only if they were previously approved by COEPES.

Concluding Observations

In most federal systems, the supervisory power as regards higher education does not rest at the central (federal) level. This provides room for the regional (state, provincial, etc.) governments in charge to have a substantial influence in many respects. However, the coordinators or actors of higher education in some countries are in favor of

guaranteeing a system-wide homogeneity or similarity of various features of the system and aim to realize that either through central powers, through the coordination between central and regional powers, or through cooperation among regional powers. Moreover, the central government in most federal systems has the power of the purse to a certain extent, notably regarding research promotion, student aid, etc.; in some countries, federal innovation programs provide central powers even with a stronger say.

Governmental policies in many countries have strengthened the formal autonomy of universities and the managerial power of university leadership in recent years. Expert views vary whether governments altogether have strengthened their role of strategic target setting alongside or have delegated part of the strategic target setting to the individual institutions of higher education. Notwithstanding this, governmental target setting in the majority of federal systems aims increasingly for nation-wide coordination.

See also: National Science Policy and Universities; Steering of Higher Education Systems – The Role of the State.

Bibliography

- Brown, D., Cazalis, P., and Jasmin, G. (eds.) (1992). *Higher Education in Federal Systems*. Kingston, ON: Institute for Intergovernmental Relations, Queen's University.
- Casanova, H. (2006). Educación Superior y Sociedad en México: Los Retos del Siglo XXI. Website of the Red de Investigadores sobre Educación Superior – RISE (Network of Researchers of Higher Education). http://www.riseu.unam.mx/documentos/acervo_documental/txtid0023.pdf (accessed June 2009).
- Fischer, D., Rubenson, K., Clift, R., et al. (2006). *Canadian Federal Policy and Postsecondary Education*. Vancouver, BC: University of British Columbia, Centre for Policy Studies in Higher Education.
- Heidenheimer, A. J. (1992). Government and higher education in unitary and federal systems. In Clark, B. R. and Neave, G. (eds.) *The Encyclopedia of the Higher Education. Vol. 2: Analytical Perspectives*, pp 924–934. Oxford: Pergamon Press.
- Heidenheimer, A. J., Hecl, H., and Adams, C. T. (1990). *Comparative Public Policy: The Politics of Social Choice in America, Europe and Japan*. New York: St. Martin's Press.
- Jones, G. A. (2006). Canada. In Forest, J. J. F. and Altbach, F. G. (eds.) *International Handbook of Higher Education. Part Two: Regions and Countries*, pp 627–645. Dordrecht: Springer.
- Kehm, B. M. (2006). Germany. In Forest, J. J. F. and Altbach, F. G. (eds.) *International Handbook of Higher Education. Part Two: Regions and Countries*, pp 729–746. Dordrecht: Springer.
- Kent, R. and Richardson, R. C. (2002). *Políticas Federales en la Educación Superior Mexicana*. New York: New York University/AIHEPS.
- Marginson, S. (2006). Australia. In Forest, J. J. F. and Altbach, F. G. (eds.) *International Handbook of Higher Education. Part Two: Regions and Countries*, pp 587–611. Dordrecht: Springer.
- Nelson, B. (2003). *Our Universities: Backing Australia's Future*. Canberra: Department of Education, Science and Training.
- OECD (2003). *Reviews of National Policies for Education: Tertiary Education in Switzerland*. Paris: OECD.
- Teichler, U. (1992). Higher education in federal systems: Germany. In Brown, D., Cazalis, P., and Jasmin, G. (eds.) *Higher Education in Federal Systems*, pp 141–168. Kingston, ON: Institute for Intergovernmental Relations, Queen's University.

Pathways and Articulation into Higher Education

G F Moodie, Griffith University, Brisbane, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Adult and continuing or community education –

The informal sector of post compulsory education that offers mainly unaccredited studies for personal development or interest.

Articulation – A relationship that educational institutions establish between programs and normally links curriculum to provide for the accumulation or transfer of credit.

Grade point average – The average of grades awarded in a program which is weighted by the size of each subject.

Horizontal transfer – The transfer of students from one program or institution to another at the same level.

Modularization – The division of the curriculum or teaching program into discrete but linked parts with specific opportunities for students to make choices which may or may not be linked to specific assessments.

National qualifications framework –

A classification and description of the relations between types of qualifications often depicted as a table where the rows group qualifications by educational level and the columns group qualifications by broad field, occupational group or educational sector.

Nested awards – The qualifications that are conceived and developed as a coherent whole but offer a number of entry and exit points.

Numerus clausus (Latin, closed number) – An entry quota.

Pathway – An arrangement for students to transfer from one program to another which may be either designed by educational institutions or developed by students themselves.

Transnational education – The education that involves a student or teacher traveling to another country to study or teach.

Twinning – An arrangement between two institutions which allows students to undertake the initial years of their study in one institution and complete their study at the institution that awards the qualification.

Unitization – The division of a qualification into discrete parts or units, normally of the same size, which are assessed and awarded credit separately from other units.

Vertical transfer – The transfer of students from a lower to a higher level of education, most commonly higher education.

As higher education changes from admitting a small elite of typically around 5% of the group that provides most of its students to a mass system admitting more than 15% of the relevant age group (Trow, 1974: 63), it not only greatly expands but also diversifies its intakes and intake routes. This may be observed in the increasing popularity of nontraditional pathways or routes into higher education.

Pathways into Bachelor Programs

Broadly, the transition from elite to mass higher education is associated with a shift from admitting almost all students through traditional pathways shown in the early numbers of the typology set out below to admitting a higher proportion of students from less-conventional routes listed lower in the typology:

1. Entry from school.
 - a. Direct entry from an academic or general school.
 - b. Entry from a vocational or specialist school.
2. Transfer from a program at the same level of higher education.
3. Transfer from a program or institution in a lower tier or sector of tertiary education.
4. Second chance or mature-age entry.
 - a. By (modified) open entry.
 - b. After completing a higher education transitional program.
 - c. After completing an adult or continuing education program.

Direct entry from an academic or general secondary school remains the most populous, but not necessarily the majority, pathway into even mass higher education systems. This is typically based on performance in a university entrance examination established by the middle of the twentieth century, such as England's A-levels, France's *Baccalauréat*, Germany's *Abitur*, and the *Matura* administered in Italy as well as many central European countries. Matriculation or entering into the register (*matricula* in Latin) of a university has been automatic upon completion of the university entrance examination in

some countries, particularly those in continental Europe such as France, Germany, Italy, and Spain, and countries in this tradition such as some in Latin America. However, many countries have since introduced *numerus clausus* (closed number in Latin) or entry quotas, particularly for entry to programs in high demand such as medicine. Most United States do not have a university entrance examination and competitive entry to higher education is based on a personal application which is assessed individually but often with heavy reliance on the applicant's performance in a national standardized aptitude test such as the Scholastic Aptitude Test or the ACT, formerly American College Testing.

As secondary education became mass and then near universal from the middle of the nineteenth century, many countries established separate schools or streams for preparing students for advanced vocational study. Some countries established specialist schools or secondary academies for a specific vocation. Initially, higher education institutions did not accept graduates from vocational and specialist schools for admission but they started to do so as higher education expanded. Thus, it can be seen from the results of a survey conducted by the Organization for Economic Cooperation and Development (OECD) shown in **Table 1** that tertiary education institutions admitted 55–85% of their students from general secondary education and from 2% to 15% of students from vocational education.

Table 1 Students' qualifications on entry to tertiary education, selected countries

	%
<i>Australia commencing bachelor's (1991)</i>	100
Complete final year of secondary education	55
Some tertiary education	26
Mature age or employment experience	6
Other	13
<i>Denmark modeled cohort analysis all participants (1993)</i>	100
General upper secondary	75
Vocational upper secondary	15
Other	10
<i>France first-year entrants (1995–96)</i>	100
New baccalauréat – general	62
New baccalauréat – technological	23
New baccalauréat – vocational	2
Other	13
<i>Japan first-year entrants from high school (1995)</i>	100
General secondary	85
Vocational secondary	15
<i>United Kingdom first-degree and diploma students (1994–95)</i>	100
A-levels	65
Vocational qualifications (NVQ, GNVQ)	8
Other	27

Reproduced from OECD (1997). Responding to new demand in tertiary education. In OECD (ed.) *Education Policy Analysis 1997*, pp 79–96. Paris: OECD, with permission from OECD Publishing.

Conversely, not all students who complete general secondary education and proceed to tertiary education proceed to higher education. Another OECD analysis shown in **Table 2** found that 45–80% of students who completed a general secondary education proceeded to tertiary education, with many proceeding to institutes of technology or universities of applied science such as *Fachhochschulen* in Germany.

Graduates from secondary vocational or specialist schools were initially restricted in their higher education studies or were required to undertake transitional studies, but these requirements were gradually relaxed as it became clear that they performed as well as graduates from academic or general secondary schools. More recently, some jurisdictions have established academically selective senior secondary schools to prepare students specifically for transfer to higher education. In the United States, these are called middle college high schools when they are located on higher education college campuses.

For many students, higher education is as much about learning about themselves as it is about acquiring high-level knowledge and skills; therefore, it is not surprising that some seek to transfer to a different program or institution after their first year of study. Higher education institutions have long accepted students transferring from other programs or institutions at the same level, what is sometimes known as horizontal transfer. Some students know their preferred program well, but are unable to gain direct entry because their entry score or other qualification is not as competitive as other applicants' scores. In some countries, these students may be admitted to their preferred program if they perform well in their initial program. This is often measured by the grade point average, the average of grades awarded in the initial program which is weighted by the size and, in some instances, the importance of the subjects in the initial program.

Since the early twentieth century the United States has established a sector of higher education called successively junior, community and 2-year colleges with the explicit aim of providing lower-cost and low-risk opportunities for students to complete a short program, typically a 2-year associate degree in arts or sciences. The associate degree was designed to be both a valuable qualification in its own right and also a qualification for entry into the third year of the normally 4-year bachelor degree offered by senior colleges and universities. This is often known as vertical transfer. As nearly 50% of higher education students in the United States are enrolled in 2-year colleges and since admission to the preferred 4-year colleges and universities is highly selective and hence very competitive, the effectiveness of the vertical transfer system is crucial to the fairness and vitality of United States higher education.

There is, therefore, a small industry of studies of higher education student transfer in the United States which is

Table 2 Destination of students entering tertiary education by entry qualification, selected countries, percent of students

Country	General entry qualification	Vocational entry qualification
France (1995)	Baccalauréat – general	Baccalauréat-technological
University	72	30
Preparatory classes for <i>grandes écoles</i> (CPGE)	13	1
University institutes of technology (IUT)	8	13
Advanced technician sections of <i>lycées</i> (STS)	7	56
Germany (FTFR) (1992)	Higher education entrance qualifications	Fachhochschule entrance qualifications
University	79	8
Fachhochschule	21	92
Japan (1995)	General high-school graduate	Vocational high-school graduate
University	45	21
Junior college or special training school	55	79

Reproduced from OECD (1997). Responding to new demand in tertiary education. In OECD (ed.) *Education Policy Analysis 1997*, pp 79–96. Paris: OECD, with permission from OECD Publishing.

conducted by statisticians and policy analysts in state higher education coordinating boards, policy think tanks and charitable foundations, staff of 2-year colleges, as well as researchers in 4-year institutions. Long-standing issues are: how transfer is measured and whether rates are going up or down (Grubb, 1991), how well 2-year colleges prepare and advise their graduates for more advanced work (– California Postsecondary Education Commission, 2002: 15), the alignment of core subject and distribution requirements of 2- and 4-year colleges (Prager, 1993), and 4-year colleges' selection practices and granting of credit for prior studies (US Department of Education, 2006: 15).

Student transfer is also becoming increasingly common in other countries, typically from institutions established initially for vocational training but which have acquired broader and higher-level programs as they have developed. Thus, there is an increasingly important pathway into higher education from further education colleges in the United Kingdom, *Instituts Universitaires de Technologie* (IUTs) in France, *Hogescholen* in the Netherlands, vocational education and training institutes in Australia, polytechnics in New Zealand, and from Germany's *Berufsakademien*.

The final important pathway into higher education is by various second-chance or mature-age entry provisions. These open opportunities to applicants who finished their schooling, often without completing secondary education, several years earlier when persistence to the final year of secondary education and transition to higher education was more restricted than it is in many countries now. Initially traditional universities had rather modest intakes of 'second chance' students, but these have been expanded considerably as these students have shown that their maturity and motivation lead them to success in higher education similar to more traditional students. Second chance entry has also been encouraged by the noteworthy success of students attending open universities that do not have entry requirements such as the University of South Africa, the UK's Open University, Germany's FernUniversität, and India's Indira Gandhi National Open University.

One of the early and most prominent, national, modified, open-entry schemes is Sweden's 24:4 rule which reserves places for people aged at least 24 years who have 4 years' work experience. The scheme was introduced in 1977 and has since been further liberalized. Some campus-based higher education institutions now admit significant numbers of students by various schemes for which the only entry qualification is being adult or mature age, often over 25 or even 21 years of age. In other countries, higher education institutions offer foundation, preparatory, or bridging programs of 6–12 months to prepare students who are not otherwise prepared for higher education. In many countries, adult and continuing or community education – the informal sector of post compulsory education that offers mainly unaccredited studies for personal development or interest – provides an important pathway into further study which often leads to higher education. However, the role of adult and continuing education in providing access to higher education is unsystematized and is largely unacknowledged.

The discussion so far has assumed that pathways into higher education are into the bachelor program, which is typically of 3–4 years' duration. In many countries, higher education formally includes short-cycle programs of a 2-year duration and so pathways into higher education should also include pathways into short-cycle programs. However, short-cycle higher education programs often overlap or are at least confused with advanced vocational education programs which are also often of up to 2 years' duration. With the implementation of the Bologna process in Europe, the bachelor degree of 3–4 years' duration is becoming an increasingly common initial higher education qualification.

Pathways into Postgraduate Programs

There are also multiple pathways into postgraduate programs, which again may conveniently be considered in descending order of frequency:

1. an undergraduate program in the same field;
2. an undergraduate program in a different field;
3. a lower-level postgraduate program;
4. work experience.

The most conventional route to postgraduate study is through an undergraduate program in the same field as the postgraduate program. This assumes that the postgraduate program builds more or less sequentially on undergraduate studies. However, institutions have developed a number of programs that do not assume specific prior knowledge, although they may assume a general intellectual ability and knowledge similar to an undergraduate degree. The most ubiquitous of the nonsequential postgraduate degrees is the master of business administration, which is often studied by technical specialists who seek to develop broad business management expertise as they move from their specialization into middle management. The entry requirements for MBAs vary widely; however, while they are likely to require a degree, this is unlikely to be in business administration. Just as important for many MBA schools is the applicant's work experience and seniority. Higher education institutions have developed several other nonsequential masters degrees.

In some countries, higher education institutions have developed other postgraduate qualifications, such as graduate certificates, which typically require 6 months' equivalent full-time study or 30 credits in the European Credit Transfer System (ECTS), and graduate diplomas of 60 ECTS credits. While some graduate certificates and diplomas are sequential qualifications, others assume no specific prior educational knowledge and admit students with appropriate work experience. These in turn may qualify students for progression to specific masters programs, which introduces the notion of linked or articulated higher education qualifications.

Articulation into Higher Education

While pathways are arrangements for students to transfer from one program to another which may be either designed by educational institutions or developed by students themselves, articulation is a relationship that educational institutions establish between programs that normally links curriculum to provide for the accumulation or transfer of credit. A student following a pathway into higher education will normally be required to compete for a place in a program with *numerus clausus* or an entry quota. Students following an articulated route into higher education may be required to compete for admission to the higher-level program but, particularly if the programs are integrated, students may also be guaranteed progression into the higher-level program subject to satisfactory academic progress.

While there is commonality in the concepts underlying articulation into higher education, there is considerable variation in the terminology. Articulation into higher education may be:

1. sequential;
2. concurrent
 - a. multiple enrolment,
 - b. multiple credit;
3. integrated
 - a. nested awards,
 - b. incorporated awards.

The simplest instance of sequential articulation into higher education is where a lower-level program is designed to prepare graduates for progression to a higher-level program, for example, where a secondary school or specialist preparatory academy prepares graduates for admission to specific undergraduate programs or where an undergraduate program prepares graduates for admission to a sequential postgraduate program. Articulation from upper secondary to higher education is common in many countries but United States states generally have much weaker articulation between senior secondary school and higher education.

More sophisticated sequential articulation is common. In the United States, associate degrees offered by 2-year colleges may be, but are not always, articulated with bachelor degrees offered by collaborating 4-year colleges. In other countries, a college may offer the initial years of an undergraduate degree that is awarded by a collaborating higher education institution. This is increasingly common in transnational education where students undertake the initial years of their undergraduate study in their home country and complete their degree overseas at the campus of the awarding institution. This is often known as twinning and is described in shorthand by the number of years students study at each institution. Thus, a 1 + 2 is a twinning arrangement whereby students undertake the first year of their degree at the home institution and complete the final 2 years of a 3-year degree at the awarding institution, whereas 2 + 2 would represent the standard (but not necessarily majority) study pattern of a United States transfer student.

There is considerable variety in articulation which provides for students to study multiple, but normally only two, programs concurrently. Thus, students undertaking the final years of their secondary education may study concurrently a higher education subject which is often known as advanced placement programs, or they may study concurrently for a vocational education qualification. Where there is considerable overlap in the curriculum of the concurrent programs, some subjects may be credited toward both programs. A survey conducted by the US Department of Education National Center for Educational Statistics (Waits *et al.*, 2005) found that 71% of United States public high schools offered subjects for

dual credit and 67% offered advanced placement subjects in 2002–2003.

Although dual credit articulation is a more recent development in articulation arrangements involving different institutions in different sectors, it is reasonably common among students studying at the same level in the same institution where a student may be described as studying for a double major or joint degrees. While sequential articulation normally involves programs in the same discipline but in different sectors, concurrent articulation may involve programs in the same or different sector or in the same or different discipline. However, some of the most interesting concurrent articulations are of programs in complementary disciplines, such as accounting and information systems.

A further integration of articulated programs can develop into nested awards which are conceived and developed as a coherent whole but offer a number of entry and exit points. An example of nested awards is:

- certificate of bookkeeping,
- diploma of financial services,
- associate in accounting systems,
- bachelor of accounting.

At the end of each section of the nested sequence, students are granted an award whereupon they may choose to exit or continue to the next stage of the sequence. Students are not required to negotiate entry to each program because progression through the sequence of programs is guaranteed on condition of satisfactory academic progress. Nested programs are normally within the same field of study. A considerable advantage of nested awards is that they offer students early accreditation so that they may work part time as a credentialed worker in their field while continuing their study toward a higher qualification, thus gaining valuable experience from integrating work and study (Wheelahan, 2000: 19). An even closer form of integrating articulated awards is to incorporate one within another. Thus, a bachelor of science may incorporate a laboratory technician's certificate. Incorporated awards offer similar advantages to nested awards.

Measures to Broaden Pathways and Strengthen Articulation

Kintzer (1973) propounds a useful typology of articulation agreements and transfer policies: state formal agreements or legal prescriptions requiring the adoption of specified measures, state policies encouraging articulation and transfer, and voluntary agreements between individual institutions or systems. United States have adopted a variety of measures to promote articulation and transfer from 2- to 4-year institutions, most exhortatory or

voluntary rather than mandatory despite often being expressed in legislation. The Education Commission of the States (2001) found that of the 50 United States, 30 had legislation supporting transfer, 40 had statewide cooperative transfer agreements, 33 states regularly collected and reported transfer data, 18 states offered incentives and rewards to either transfer students or sending or receiving institutions, and 26 states maintained a statewide guide to transfer. While there is some evidence that state policy can influence the effectiveness of student transfer (Wellman, 2002: 45), others (Anderson *et al.*, 2006) have doubted the effectiveness of state articulation agreements. Other jurisdictions adopt the range of measures to promote articulation described in Kintzer's typology, with the extent of state prescription, encouragement, or quiescence on articulation generally related to the extent of their involvement in other aspects of higher education.

Thus far the discussion has considered pathways and articulation largely based on arrangements between institutions and their occupational and academic communities. These arrangements are strong in East Asian and continental European countries such as France, and in Germany where the social partners—employees, employers, and the federal as well as *Länder* (state) governments—cooperate closely in maintaining the links between tertiary education and work. In contrast, some countries have sought to broaden pathways and strengthen articulation into higher education by establishing a national qualifications framework above institutions. National qualifications frameworks typically place qualifications in a table where the rows sort qualifications by educational level and the columns sort qualifications by broad fields, occupational groups, or educational sectors. The aim is to provide students the means for moving from one qualification to another so that in principle there remain few if any terminal qualifications.

National qualifications frameworks have been adopted by many of the Anglophone countries of the Commonwealth of Nations that have structured their tertiary education in a neoliberal market, such as New Zealand, Scotland, Australia, South Africa, England, and Ireland. They may also be considered in Kintzer's typology, with those founded in New Zealand and South Africa being prescriptive and those of Scotland, Ireland, and Australia being more enabling (Young, 2005: 12) and almost voluntary for some sectors. Some argue that a qualifications framework is part of the neoliberal program since it commodifies education and thus “contributes to the creation of educational markets by providing a common qualifications currency. This common currency, like money in an economy, is viewed as promoting greater competition between the providers of educational qualifications because all institutions are recognising and rewarding learning in the same way” (Strathdee, 2003: 157). However, even in the countries that have heavily marketized

their tertiary education, such as New Zealand, Australia, and, to a lesser extent, England, national qualifications frameworks are promoted as a way to minimize barriers to vertical and horizontal transfer and to “maximise access, flexibility and portability between different sectors of education and work and different sites of learning” (Young, 2003: 224). However, there is relatively little evidence that national qualifications frameworks achieve these goals (Young, 2005: 1).

Most national qualifications frameworks are associated with the unitization of programs, although in some countries programs have been unitized independently of national qualifications frameworks. Until the mid-twentieth century, higher education students’ performance was not formally assessed until they reached the end of their program, when the assessment was to determine whether the student met the requirements for the award of the qualification. Therefore, students may not have been formally assessed until three (UK), four (US), or, in continental Europe, up to 7 years after they started their program. While end-of-year assessment was introduced in many countries, this was still a long and, for some, an unwieldy learning cycle; hence, teachers or institutions divided the curriculum or teaching program into discrete but coherently linked parts or modules with specific opportunities for students to make choices. Modules may or may not be linked to specific assessments (Young, 2005: 25). Young (2005: 25) distinguishes modularization of the curriculum from unitization of qualifications.

Young characterizes unitization as a more radical departure from traditional approaches to the curriculum since it refers to the break up of qualifications, not the curriculum, and it is concerned with assessment rather than teaching. Unitization divides programs into discrete parts or units, normally of the same size, which are assessed and awarded credit separately from other units. While institutions had previously assumed that all students in a cohort began together and studied common subjects, unitized programs allow students to treat units as the building blocks of many different qualifications. Whereas previously the teacher had determined the study sequence, students determine their sequence in a unitized program (Young, 2005: 25). Unitization greatly increases students’ flexibility and, at least in principle, facilitates the construction of pathways and articulation into higher education.

Perhaps the most significant influences on pathways and articulation into higher education are broad economic, technological, and social changes. Increases in national wealth allow governments to expand higher education and establish facilities in smaller population centers, greatly expanding access. Increases in personal and family wealth give people and their children the resources and, importantly, the aspiration to pursue higher education. Improvements in and the reduced cost of transport and communication make it much more feasible for

students to study at institutions distant from their home. These are some of the reasons for the great expansion of transnational higher education over the last two decades, which has spawned the twinning arrangements noted earlier as well as the many pathways and articulation agreements that are being developed between institutions in different countries. The techniques of distance or resource-based education and, more recently, the great expansion of information and communication technologies have supported a variety of new pathways and articulation arrangements. This suggests that future developments in pathways and articulation into higher education are likely to be influenced as much as by these general economic, technological, and social developments as by measures specifically directed to pathways and articulation into higher education.

See also: Access and Equity in Higher Education; Student and Faculty Transnational Mobility in Higher Education.

Bibliography

- Anderson, G. M., Sun, J. C., and Alfonso, M. (2006). Effectiveness of statewide articulation agreements on the probability of transfer: A preliminary policy analysis. *Review of Higher Education* 29(3), 261–291.
- California Postsecondary Education Commission (2002). *Student Transfer in California Postsecondary Education. Commission Report 02-3*. Sacramento, CA: California Postsecondary Education Commission.
- Education Commission of the States (2001). *Transfer and Articulation Policies*. Denver, CO: Education Commission of the States.
- Grubb, W. N. (1991). The decline of community college transfer rates: Evidence from national longitudinal surveys. *Journal of Higher Education* 62(2), 194–222.
- Kintzer, F. C. (1973). *Middleman in Higher Education*. San Francisco, CA: Jossey-Bass.
- Prager, C. (1993). Transfer and articulation within colleges and universities. *Journal of Higher Education* 64(5), 539–554.
- Strathdee, R. (2003). The qualifications framework in New Zealand: Reproducing existing inequalities or disrupting the positional conflict for credentials. *Journal of Education and Work* 16(2), 147–164.
- Trow, M. (1974). Problems in the transition from elite to mass higher education. In Organisation for Economic Co-operation and Development (OECD) (ed.) *Policies for Higher Education*, pp 51–101. Paris: OECD.
- US Department of Education (2006). *A Test of Leadership: Charting the Future of U.S. Higher Education*. Washington, DC: US Department of Education.
- Waits, T., Setzer, J. C., and Lewis, L. (2005). *Dual Credit and Exam-Based Courses in U.S. Public High Schools: 2002–03* (NCES 2005–009). Washington, DC: US Department of Education National Center for Education Statistics.
- Wellman, J. V. (2002). *State Policy and Community College-Baccalaureate Transfer. National Center Report #02-6*. San Jose, CA: The National Centre for Public Policy and Higher Education.
- Wheelahan, L. (2000). *Bridging the Divide: Developing the Institutional Structures that Most Effectively Deliver Cross-Sectoral Education and Training*. Leabrook: National Centre for Vocational Education Research.
- Young, M. (2005). National qualifications frameworks: Their feasibility for effective implementation in developing countries. *Skills Working Paper No. 22. In Focus Programme on Skills, Knowledge and Employability*. Geneva: International Labour Office.

Young, M. F. D. (2003). National qualifications frameworks as a global phenomenon: A comparative perspective. *Journal of Education and Work* 16(3), 223–237.

Further Reading

Moodie, G. (2005). *Student Transfers*. Adelaide: National Centre for Vocational Education Research.

OECD (1997). Responding to new demand in tertiary education. In OECD (ed.) *Education Policy Analysis 1997*, pp 79–96. Paris: OECD.

Young, M. (2002). Contrasting approaches to qualifications. In Evans, K., Hodkinson, P., and Unwin, L. (eds.) *Working to Learn: Transforming Learning in the Workplace*, pp 44–62. London: Kogan Page.

Relevant Websites

<http://www.cpec.ca.gov> – California Postsecondary Education Commission, Issues, Transfer.

<http://www.ecs.org> – Education Commission of the States, Education Issues, Transfer/Articulation.

<http://www.ncver.edu.au> – National Centre for Vocational Education Research, VOCED an International Database of Research Abstracts for Technical and Vocational Education and Training.

<http://www.oecd.org> – Organisation for Economic Co-operation and Development, Education Topic, Tertiary Education.

<http://unesco.org> – United Nations Educational, Scientific and Cultural Organization, Education, Higher Education.

Unitary, Binary and other National Systems of Higher Education

L Goedegebuure, University of New England, Armidale, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Binary system – A higher education system consisting of two groups of institutions, normally universities and nonuniversity institutions.

Stratification – A division of the higher education system in distinct groups with different status.

Unitary system – A higher education system without a formal distinction in types of institutions; also known as a unified system.

Introduction

Policymakers and researchers alike for many years have struggled with the question of system design and structure. Teichler's observation that "[t]he structure of the higher education system has been a major topic of debate and of reform efforts in most Western industrial societies since about the mid-1960s" (Teichler, 1988: 96) is as valid 20 years later as it was then. This points to the fact there is no right answer to the question about what is the ideal structure of a higher education system. It is dependent upon a combination of factors that include the developmental stage of the system in terms of participation rates, political priorities, the structure of the secondary-education system, and, increasingly, its international connectedness. In this article, the primary focus is on the issue of system dynamics: what are the tensions that particular structural arrangements evoke and how do they play out? This brings us into the realms of status and hierarchies and the question whether multidimensional classification schemes are a possible way forward. First, however, the most common structural arrangements that can be found in higher education systems are presented.

Structural Arrangements in Higher Education Systems

Throughout the higher education policy literature a number of structural typologies can be found to categorize systems. The earliest one probably is Trow's distinction of elite–mass–universal higher education systems, based on the participation rates of a system (Trow, 1974). In his typology, elite systems are those with participation

rates up to 15%, mass systems have a 15–40% participation, and universal systems enrol over 40% of students from the standard age group. Technically, this typology does not say anything about the structure of a system, but since it is so commonly used, it needs to be included here. However, it should also be noted that the concept of an age group is increasingly becoming contentious with lifelong learning becoming the rule rather than the exception. Scott (1995) identifies four broad types of higher education systems: the dual system, with separate positions for universities and other postsecondary-education institutions that are treated differently; the binary system in which "alternative institutions are deliberately established to complement (and rival?) the universities" (Scott, 1995: 35); the unified system in which there is no formal differentiation between types and all institutions are part of a common system; and the stratified system, which also comprises all institutions, but with specific roles allocated to the institutions in it. As Scott himself acknowledges, the distinction between the first two types is not really watertight (blurred) and the typology may be "too simplistic to capture the complexity of existing higher education systems" (Scott, 1995: 36). This leads him to rearrange the building blocks in an evolutionary typology where higher education systems move from university-dominated to dual to binary to unified and finally stratified systems (see also Kyvik, 2004). We return to this evolutionary path later in this article, which in 2007 Scott modified to a pre-binary, binary, post-binary/unified, and differentiated/market-system trajectory (Scott, 2007). Teichler (1988, 2004) has a different take on distinguishing systems by grouping them in two categories: diversified and integrated systems. According to him, diversified systems are characterized by substantial differences in quality, status, and content of the programs provided by the institutions in the system, while in integrated systems, those differences are much more contained. Important to the further discussion in this article, he notes that "some kind of consensus seems to have emerged that borderlines between various sectors of the higher education system ought to be blurred, and that a certain degree of permeability of educational ladders ought to be ensured" (Teichler, 2004: 7). Binary systems, with clear distinctions between university and nonuniversity type of institutions are the most common example of diversified systems, although there also are examples of more complex systems such as the French system. Clear examples of integrated systems would be Australia and the United Kingdom. Both countries decided

to abandon their binary systems in favor of a single sector, university system. We return to this later, but first a few more words on the most common of formally diversified systems, the binary system.

Binary Systems: Rise and Demise

There is strong agreement in the higher education-policy literature that binary systems originated out of the desire of policymakers to keep an elite university system while also allowing for greater participation (e.g., Pratt, 1997; Goedegebuure and Meek, 1988). Nonuniversity institutions were set up as basically teaching-only institutions with a predominant vocational orientation, as policymakers realized that meeting increased demand through expansion of the university sector would be financially unsustainable. The perception of expansion at the cheap and thus a lower status was politically countered by the introduction of the equal but different rhetoric. Although these newly created institutions at the outset indeed were different, the equality concept in practice never really got off the ground. It should also be noted that the underpinning academic-vocational distinction from the outset has been contentious, with disciplines such as law and medicine remaining firmly located within the university sector. This has been exacerbated by universities further entering the so-called vocational domain to secure student numbers and nonuniversity institutions entering the postgraduate domain to secure both students and academic reputation, although many would deny that this is what they are after.

There is no denying the occurrence of academic and vocational drift, but we should be wary of treating binary systems in a dichotomous or black-and-white manner.

First of all, since the rise of the early binary systems in the 1960s, our world, including the world of work, has moved on and has become much more complex. In response to this, the programmatic offerings of higher education institutions have become more diverse, both in terms of content, levels, and mode of delivery, competency-based curricula being a good case in point. This has resulted in a certain blurring of boundaries.

Second, with the ever-growing number of students, both traditional and mature, higher education institutions have grown in size as well as complexity, and many more people now find employment in this sector than was the case 40 years ago. A substantial number of staff working in nonuniversity institutions would have obtained their training in universities and hence have been exposed to a university culture. This in turn would have resulted in a blending of what at the outset were distinct institutional cultures. Once again, care should be taken how far this argument can be stretched, but certainly in the more mature binary systems such as Germany and the Netherlands, increasing emphasis

is placed on the importance of staff with doctorates to support both research and postgraduate training activities.

Third, it should be realized that not all binary systems find their roots in the 1960s and that their developmental trajectories therefore are quite different. Binary structures have become a common feature in the post-communist higher education systems in central and Eastern Europe; they have been introduced in the Nordic countries in Europe much later, and also in countries like Austria, Portugal, and Switzerland. As such, many of these institutions find their substantive roots in the 1990s, and many have been established not for reasons of protecting universities, but rather because universities in these countries were not meeting the demands of a diversifying labor market and hence of industry.

Fourth, one should be aware of the different sizes of both sectors. As Kyvik (2004: 404) demonstrates, in Europe there is a wide variety in enrolment patterns of first-year students across the two sectors, with the Netherlands seeing 75% of enrolments in the nonuniversity sector, Flanders 70%, Finland 60%, and Ireland and Norway 55%. On the other side of the scale, we find Austria with a 15% first-year enrolments in the nonuniversity sector, Switzerland and Germany with 30%, and Sweden and Portugal with 40%. As is the case with the introduction of bachelors and masters under the banner of the Bologna Process, what this teaches us is that one binary system is not identical to another binary system: much diversity resides within these broad container concepts (see also: Taylor *et al.*, 2008). It also very clearly points to the role played by the nonuniversity sector in terms of the transition from secondary to higher education and expanding access to higher education (for an elaboration of this, see: Moodie, 2008). In this respect, it is not the case that there is a clear link between the structure of secondary education in terms of vocational/professional tracks and general/pre-university tracks and entry patterns to higher education. Three examples may suffice to illustrate this point. In the Netherlands, secondary education is clearly segmented or tracked with 6 years of preuniversity education providing for access to university studies and 5 years of more vocationally oriented studies allowing access to the polytechnic sector. Flanders, however, only has general secondary education upon competition of which students can enrol in either a university or a polytechnic program. Finally, Finland has a segmented secondary-education system with general upper-secondary education traditionally being the route to university entrance. However, the vocational upper-secondary education and training track provides eligibility to both university and polytechnic programs. This mixture of preparatory tracks leads to the hypothesis that in those systems with large enrolments in the nonuniversity sector, this sector has been able to establish itself as a viable alternative to the university sector, with a less-academic orientation and in

general lower entry levels. Although testing of this hypothesis is beyond the scope of this article, there is plenty of supportive documentation available to not reject the hypothesis out of hand. In both the Netherlands and Flanders, it is common for students to try a university program, find out that they have difficulty in coping, and then cascade down to the polytechnic sector where the vast majority completes their studies successfully. In Finland, the university sector is organized very traditionally and the creation of the polytechnic sector in the early 1990s intended “overcoming the functional shortcomings in the system [...] and a means of clearing a vocational and matriculation backlog” (OECD, 2006: 8). The review report then goes on to argue that since then the polytechnic sector has “continually matured to equal status with a very specific character to universities...” (OECD, 2006: 8).

The notion of maturing or coming of age logically brings back the concept of evolutionary pathways: is there a natural progression from one system state to the other? Once again, we need to be careful. First of all, evolution is a random process of adaptation and selection, and not a process of progression despite the commonly used iconography from ape to man. Taking progress and stages into our higher education world is problematic. As argued elsewhere: “(...) the jury is still out with respect to the applicability of concepts like natural selection and evolution in the social sciences” (Goedegebuure *et al.*, 1996: 7). Second, there appears to be an empirical problem to substantiate the argument. As a matter of fact, there only are two well-documented cases of binary systems being replaced by unified systems: Australia and the United Kingdom. This, therefore, begs the question whether these are the exception rather than the rule. To answer this question, we need to examine the rationale behind these cases more closely. For Australia, Meek and Goedegebuure (1989) point to the failure of the binary system to deliver on its promises of furthering equity and contributing to economic growth through expansion of participation in combination with the structure’s internal contradictions and inconsistencies. As they argue “College students were very much like their university counterparts, slightly less financially well-off, but taking similar subjects and having, on the whole, similar aspirations. [...] College academics had the same predilections for research and postgraduate teaching as their university colleagues” (Meek and Goedegebuure, 1989: 31) and certainly the larger nonuniversity institutions had developed the capacities to deliver on this. Therefore, basically, the binary system became untenable because the differences between the constituting sectors on the important indicators such as program level, subject distribution, entry levels, and research intensity no longer constituted significant differences. Much the same picture can be painted for England. As argued by Clark (2006: 14): “In the 1991 White Paper ‘Meeting the Challenge’, the Government

announced the abolition of the binary line on the grounds that there was significant overlap in the academic work of the two sectors (with many polytechnics offering research degrees). It was generally agreed that a single sector of universities (and colleges) with diverse aims and objectives would better serve the interests of students and society. It was expected that direct competition between a wider range of institutions would lead to increased efficiency and effectiveness.” Since the Australian higher education system has been modeled to a large extent on the British system, the question thus becomes whether these two binary systems lost their *raison d’être* because they had existed for some 25 years and had evolved to the next stage, to use the Teichler and Scott terminology, or whether it was their location in systems of similar structural and cultural order that resulted in their ultimate demise.

The above is not an easy question to answer, but the following observations may provide some further insights. If the evolutionary paradigm holds true, then over the next decade one would expect to see the demise of all binary systems in Europe. All countries that have binary systems in Europe now also have a similar degree structure as a resultant of the Bologna Process, which also was the case for Australia and the UK. Unified transnational quality assurance processes will further reduce sectoral differences, and continued staff and student mobility schemes will further integrate academic norms and value systems. In 10 years, this hypothesis can be tested, but now it has to remain in the realm of speculation. Extrapolating from empirical observations, however, could also lead one to hypothesize that binary systems can viably be maintained on the condition that firm boundaries are set and not transgressed. The most well-known case in point would be the Californian higher education system where the master plan, originally developed under the leadership of Clark Kerr in 1960, stipulates clearly differentiated and legislated roles and mission for the California State University (CSU) system and the University of California (UC) system. As such, the system clearly is a stratified system with a formal hierarchy and has remained so for almost 50 years. The point to be made here is that the structure has not been without tension, in particular when it comes to the doctoral-granting status that has been the preserve of the UC institutions, much to the chagrin of ambitious CSU institutions, but it has remained in place because of strict mission-oriented legislation and higher education policy (for an elaboration, see Douglas (2003); Fox (1994); formally, of course, the system also includes the community colleges sector). Despite the fact that the Californian system is the best known, these types of nonpermeable stratified systems are quite common across the US. Taking this to Europe, one could equally hypothesize that binary systems will remain conditional on national governments willing to put in place and keep in place regulation and policies that keep

both sectors clearly apart in terms of mission, programs, and stakeholders. This, however, is no mean task as there are inherent tensions embedded in segmented systems (as, *inter alia*, there are in all systems for that matter but they come to the fore more easily in segmented systems).

Inherent Tensions in Formally Diversified Systems

The main problem that policymakers and institutions alike find with working in formally diversified systems such as binary systems, is that most of the times the crucial underpinning philosophy of equal but different does not work. As summarized by the Organization for Economic Cooperation and Development (OECD), this is because of: “the social and cultural status attributed to older universities and their members (staff and students); the more generous resourcing available to elite and research-oriented universities; and the ‘trickle-down’ effect of academic staff recruitment: most staff in all but the most prestigious institutions are likely to have obtained their qualifications from an institution higher in the academic hierarchy than their present place of work” (OECD, 2008: 97). University staff salaries on average are higher than in the nonuniversity sector, and research by and large remains more valued than teaching. So basically, we do not have a level playing field when it comes to binary systems. In fact, they can well be conceptually conceived of as stratified systems. As Riesman has already noted in 1956, in stratified systems lower-status institutions will try to gain status by imitating higher-status institutions, resulting in what he calls a reptilian procession (Riesman, 1956; see also: Neave, 1979; van Vught 2008). The overall result, of course, is that over time, differences between institutions diminish and sectoral boundaries become more difficult to maintain. These processes of academic – and its corollary vocational – drift not only are stimulated by perceptions of status, they are equally instigated by uniform policy environments with singular reward and/or incentive structures. A policy environment can be characterized as “a political arena in which institutions and interests groups within institutions vie with one another to maximize status, prestige and financial rewards. Diversity and convergence become part and parcel of the political process, for one or the other state of affairs may privilege one set of institutions or interest groups over others, setting in train campaigns by the ‘disadvantaged’ to alter their circumstances” (Goedegebuure and Meek, 1997: 315). Therefore, in order to create and maintain relatively stable stratified systems, it is paramount that distinct policy environments are put in place that reward the different missions of the institutions in these sectors at a relatively comparable level. In this way, legitimacy for both sectors can be ensured. But once again, this is easier said than done. Mission-based

steering requires a solid policy capacity at the central government level which is not something that should be taken for granted. In addition, many governments the world over have moved away from this type of hands-on governance approach to more market-like arrangements. Consequently, existing boundaries between sectors become increasingly blurred as institutions are provided with the opportunity to imitate perceived higher order institutions. But it is not only imitation behavior that blurs boundaries. As has been argued before, our world is increasingly becoming complex and in response our tertiary education institutions are becoming increasingly complex, simultaneously operating on what can best be described as multilayered chessboards. While on certain of these boards, they may be engaged in the same play for reasons of national or regional needs, on others they may be engaged in quite different sets of games. It is this complexity that makes sector boundaries at times more like artifacts than a reflection of empirical realities, and it makes the afore mentioned mission-based steering arrangements so difficult to implement for one central government agency. In this respect, we need to remember that the overarching rationale for the move from state control to state supervision exactly was this notion of increased complexity (Neave and van Vught, 1991). A possible way out of this dilemma might be found in recent initiatives to introduce multilevel classification schemes that reflect the multifaceted nature of contemporary higher education institutions and allows for self-selection by individual institutions on the dimensions they want to be held accountable for.

From Stratification to Classification?

Probably the best-known institutional classification is the so-called Carnegie Classification in the US. Originally developed in 1970 to support the research and policy analysis program of the Carnegie Commission on Higher Education, the classification “has been the leading framework for describing institutional diversity in U.S. higher education” (Carnegie Foundation for the Advancement of Teaching, 2008). In 2005, a new system was introduced based on a set of multiple, parallel classifications which “provide different lenses through which to view U.S. colleges and universities (...) organized around three fundamental questions: what is taught (Undergraduate and Graduate Instructional Program classification), who are the students (Enrolment Profile and Undergraduate Profile), and what is the setting (Size and Setting)” (Carnegie Foundation for the Advancement of Teaching, 2008). The new classification scheme also allows for a set of elective classifications that depend on voluntary participation by institutions – all other sets are constructed on the basis of publicly available empirical data – to highlight specific mission elements such as community engagement. The new classification scheme has been developed

for a better reflection of the institutional fabric existing in US higher education and to address some of the negative aspects that have come to the fore through the use of a single classification framework in, for example, newspaper rankings (McCormick and Zhao, 2005).

In similar vein, a project to build a European classification of higher education institutions, supported by the European Commission, has been ongoing since 2005. The aim of the project is to create a level of transparency in the European higher education area which would support the various stakeholders in this area:

- Business and industry will be able to have a better identification of the institutions they wish to relate to with respect to hiring graduates, commissioning research, organizing knowledge transfer, etc.
- Policymakers (at various levels) will be able to target policies and programs better.
- Students will be able to easily identify their preferred higher education institutions and make better choices regarding their study programs and labor-market perspectives.
- Higher education institutions will be able to develop their missions and profiles in a better manner and engage in partnerships, benchmarking, and networking (CHEPS, 2008: 11).

Like the Carnegie Classification, the European classification under construction is multidimensional and uses *a posteriori* information. The 2008 draft version has been developed through an extensive and encompassing stakeholder-consultation process and is explicitly nonhierarchical in terms of the dimensions, criteria, and categories used. It consists of 14 dimensions on the basis of which institutional profiles can be created: types of degrees offered; range of subjects offered; orientation of degrees; involvement in lifelong learning; research intensiveness; innovation intensiveness; international orientation: teaching and staff; international orientation: research, size, and mode of delivery; public/private character; legal status; cultural engagement; and regional engagement.

Both the US and the European classification schemes allow for a range of profiles, making it possible for institutions to identify with a group of similar institutions to be compared with on the basis of actual performance. It can help institutional policymakers to set aspirational goals and missions and allow them to clearly state what the institution wishes to be funded for and to be judged on (and even ranked on – thus avoiding the current apples-and-oranges problem). It also allows for the identification of diversity within a system and for national policymakers to fund institutions in terms of their missions and to be far more assured that this will generate the intended policy outcomes rather than the unintended ones that so often occur due to copycat behavior. It can create parity of esteem as sets of missions are agreed upon by institutions

and national governments and each of the sets (or classification groups) has an intrinsic value since it has been arrived at through an inclusive consultative stakeholder process. As such, it may avoid the rigidities and hierarchies that formally stratified systems encompass. By having institutions in the driver's seat in terms of selecting the dimensions they consider important and want to be central in the national policy processes, it avoids the capacity problems at the national policy level identified above. It is still early days in terms of the actual viability of these schemes, particularly in Europe; yet, their multilayered principles may be instrumental in reducing the negative status and hierarchy connotations that accompany formally stratified systems such as binary systems, and may provide a better support for higher education institutions in the pursuit of multiple missions attuned to the needs of an ever-evolving society.

See also: Access and Equity in Higher Education; Bologna Process: On the way to a Common European Higher Education Area; Boundaries of Institutional Autonomy and their Impact on Higher Education; Diversity of Higher Education; Higher Education and the Transformation of Society; Higher Education: An Overview; National and International Rankings of Higher Education; Pathways and Articulation into Higher Education; Returns to Education in Developed Countries; Steering of Higher Education Systems – The Role of the State; The Bologna Process in European Higher Education; The Changing Role of the State in Higher Education.

Bibliography

- Carnegie Foundation for the Advancement of Teaching (2008). The Carnegie classification of institutions of higher education. <http://www.carnegiefoundation.org/classifications> (accessed August 2009).
- CHEPS (2008). *Mapping Diversity; Developing a European Classification of Higher Education Institutions*. Enschede: Center for Higher Education Policy Studies.
- Clark, T. (2006). *Country Report: United Kingdom; OECD Thematic Review of Tertiary Education*, Paris: OECD.
- Douglas, J. A. (2003). The Dynamics of massification: A Comparative Look at Higher Education Systems in the UK and California', OxCHEPS/Perspectives Manuscript, mimeo.
- Fox, W. (1994). Higher education policy in California. In Goedegebuure, L., Kaiser, F., Maassen, P., et al. (eds.) *Higher Education Policy: An International Comparative Perspective*, pp 49–82. Oxford: Pergamon.
- Goedegebuure, L. and Meek, V. L. (eds.) (1988). *Change in Higher Education; The Non-University Sector. An International Perspective*. Culemborg: Lemma.
- Goedegebuure, L. and Meek, V. L. (1997). On change and diversity: The role of governmental policy and environmental influences. *Higher Education in Europe* 22(3), 309–319.
- Goedegebuure, L., Meek, V. L., Kivinen, O., and Rinne, R. (eds.) (1996). On diversity, differentiation and convergence. In *The Mockers and Mocked: Comparative Perspectives on Differentiation, Convergence and Diversity in Higher Education*, pp 2–13. Oxford: Pergamon.
- Kyvik, S. (2004). Structural changes in higher education systems in Western Europe. *Higher Education in Europe* 29(3), 393–409.

- McCormick, A. C. and Zhao, C. (2005). Rethinking and reframing the Carnegie classification. *Change* **37**(5), 51–57.
- Meek, V. L. and Goedegebuure, L. (1989). *Higher Education: A Report*. Armidale: DAHES, University of New England.
- Moodie, G. (2008). *From Vocational to Higher Education: An International Perspective*. Berkshire: McGraw-Hill.
- Neave, G. (1979). Academic drift: Some views from Europe. *Studies in Higher Education* **4**(2), 143–159.
- Neave, G. and van Vught, F. A. (eds.) (1991). *Prometheus Bound. The Changing Relationship between Government and Higher Education in Western Europe*. Oxford: Pergamon.
- OECD (2006). *Finland: Country Note, Thematic Review of Tertiary Education*. Paris: OECD.
- OECD (2008). *Tertiary Education for the Knowledge Society, Vol. 1*. Paris: OECD.
- Pratt, J. (1997). *The Polytechnic Experiment: 1965–1992*. Buckingham: SRHE and Open University Press.
- Riesman, D. (1956). *Constraint and Variety in American Education*. Lincoln: University of Nebraska Press.
- Scott, P. (1995). *The Meanings of Mass Higher Education*. Buckingham: SRHE and Open University Press.
- Scott, P. (2007). Back to the future? The evolution of higher education systems. In Kehm, B. M. (ed.) *Looking Back to Look Forward; Analyses of Higher Education after the Turn of the Millennium*, Werkstattberichte 67, Kassel: INCHER.
- Taylor, J. S., Brites Ferreira, J., Lourdes Machado, M. de, and Santiago, R. (eds.) (2008). *Non-University Higher Education in Europe*. Dordrecht: Springer.
- Teichler, U. (1988). *Changing Patterns of Higher Education Systems*. London: Jessica Kingsley.
- Teichler, U. (2004). Changing structures of the higher education systems; The increasing complexity of underlying forces. In *Diversification of Higher Education and the Changing Role of Knowledge and Research*, pp 3–15. *UNESCO Forum Occasional Paper Series*. Paris: UNESCO.
- Trow, M. (1974). Problems in the transition from elite to mass higher education. In OECD (ed.). *Policies for Higher Education*. Paris: OECD.
- van Vught, F. A. (2008). Mission diversity and reputation in higher education. *Higher Education Policy* **21**, 151–174.

Further Reading

- Bleiklie, I. (2003). Hierarchy and Specialisation: on the institutional integration of higher education systems. *European Journal of Education* **38**(4), 341–355.
- Bleiklie, I. (2005). Organizing higher education in a knowledge society. *Higher Education* **49**, 31–59.
- Guri-Rosenblit, S., Sebkova, H., and Teichler, U. (2007). Massification and diversity of higher education systems: Interplay of complex dimensions. *Higher Education Policy* **20**, 373–389.
- Kogan, M. (1997). Diversification in higher education: Differences and Commonalities. *Minerva* **35**, 47–62.
- Pratt, J. and Burgess, T. (1974). *Polytechnics: A Report*, London: Pitman.

Viewing Private Higher Education: How Much, Where, Why, and What?

D C Levy, University at Albany (SUNY), Albany, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Overview

This article aims at a broad sketch of private higher education (PHE). PHE now holds roughly 30% of total global enrolment. Yet there has been woefully little scholarship about the sector. Fortunately, the amount and quality has been improving in the last few years, with recent cross-national volumes including Altbach and Levy (2005), Slantcheva and Levy (2007), Gupta *et al.* (2008), Mabizela *et al.* (2008), and Levy and Zumeta (forthcoming). Thus, one goal of this article is to promote the adequate entry of PHE into analysis of important higher education topics.

This article proceeds to three parts: (1) the astonishing growth of PHE; (2) major types of institutions that constitute the sector; and (3) private–public contrasts in finance, accountability, and governance. Throughout, comparisons will be drawn to the public sector. Most often there is considerable intersectoral distinctiveness, but there is also blurring. Contrasts are especially sharp where the private sector is largely for profit (Kinser and Levy, 2006).

Reality differs greatly within and across regions and countries, and across time. PHE is a wide-ranging and multifaceted phenomenon. Most generalizations about it are simplifications, usefully qualified, including with counter examples, and made with risk. Nonetheless, we can now make a number of warranted and weighty generalizations about PHE.

Emergence and Size

Leaving aside precursors including colonial institutions, borderline training and secondary higher education institutions, specialized colleges, and old private–public hybrids, indeed leaving aside almost everything prior to the twentieth century, our focus is contemporary, basically on first-degree higher or tertiary education. The growth of graduate (Landoni, forthcoming) and research PHE (Levy, 1996) merits attention, as does comparison between PHE and private primary and secondary schools (Scheker, 2007).

Outside the US, for most of the twentieth century, higher education was usually public, religious roots receding, and until mid-century PHE remained limited in scope. However, the surge from mid-century, especially since the 1980s, has been almost unabated to the present (Levy, 2006, 1992).

PHE growth is mostly about the developing and transitional post-communist world. The only clear-cut example of a developed country with majority private enrolment has been Japan, although currently South Korea and Taiwan could reasonably be cited as well.

Western Europe continues to be the main region with only very limited PHE. Its higher education system has been mostly state centered. Private sectors often have been only marginal (Geiger, 1986). Even in this region, however, recent years show privatization. The most noteworthy is the partial privatization of the public sector, but that goes beyond the scope of our article. MBAs exemplify private emergence tied to marketization and globalization. There is very recent major philanthropy for two private institutions – Germany and Italy. Still, Portugal remains the region's sole example of PHE holding a substantial share (27%) of total enrolment.

Until 1989, even more nonprivate was communist Europe, but that picture was transformed quickly and powerfully (Slantcheva and Levy, 2007). Most dramatic until the mid-1990s, the growth produced up to 30% private shares (Paschuavili, 2007 on Georgia and Jablecka, 2007 on Poland), although others of only a few percent (e.g., Croatia and Slovak Republic). Moreover, PHE growth may convert to stagnation and even decline given the aging demographic context, just as it may in Japan. Decline is also a threat for particular sectors of PHE, as shown by Catholic PHE in the US (Collier, in progress) and elsewhere.

Yet if continued private growth is not inevitable everywhere, it has prospects of further ascent elsewhere. The newest regional sites of sudden private development from near zero to noteworthiness are the Middle East and North Africa. The private breakthrough is partly a byproduct of the still-very-low cohort enrolments in higher education overall. There is much room to grow, if economies develop, with corresponding intensification of demand for higher education. Much of the Gulf, Morocco, Egypt, Syria, and Jordan (and Israel, if we concentrate on colleges) are among the noteworthy examples, usually with government initiative and support. Meanwhile, the surge in sub-Saharan Africa (Mabizela *et al.*, forthcoming) has been a bit longer but dramatic nonetheless, more often following the typical global development pattern of an unplanned and unanticipated private surge (Levy, 2006). Although uneven, the private growth is now widespread in Africa.

A longer experience with PHE characterizes Latin America. Cuba is alone in having no PHE. From only about 3% private around 1950, the regional private share catapulted to near 40% by around 1980, despite unprecedented public growth in absolute numbers (Levy, 1986) and may be around 45% now, according to the United Nations Educational, Scientific and Cultural Organization (UNESCO) data. Brazil, Chile, Colombia, the Dominican Republic, and El Salvador are examples where most of the system's enrolments are private.

Largest in population, Asia shows enormous variation with striking cases of over 70% private (Indonesia, Japan, the Philippines, South Korea, and Taiwan). A set of countries more in the 30–40% range includes India, Kazakhstan, and Malaysia. However, the largest present growth, with huge potential, comes in countries with private shares still under 15%. Most important is China, but Vietnam, Cambodia, and Thailand are other examples.

This overview of PHE growth and size of course tells us much about one of the chief societal concerns about higher education overall: access. Among developed countries, possibly only Japan massified through the private sector (again now reasonably adding South Korea and Taiwan). Yet the situation is radically different for much of the developing world. No surprise then that PHE promoters prominently cite the provision of access as a major achievement. Even societies not ideologically keen on privateness in education have faced the stark reality of lack of funds and commitment to provide mass access through the public sector. Soaring demand for higher education has been a characteristic of recent decades and is projected to further intensify rather than slacken in the coming decades. “You can’t do it without us,” could be the PHE rallying point about access.

As a corollary, expanded access usually means access for groups less-socioeconomically privileged than those already in the system – including those in the extant public sector. Of course, access for the less-privileged is one of public higher education’s most passionately proclaimed aims and there are indeed relevant accomplishments. Yet the public sector is usually less concerned with competition from the private sector on higher education’s low end, than it is at higher rungs.

Types of PHE

PHE varies tremendously. Differences among institutions can be greater than those within the public sector. Therefore, most questions and answers about PHE should fruitfully be attentive to types of private institutions. Different types can be widely encapsulated into three categories, which are discussed below. Each has a fairly defining set of characteristics, most intertwined with factors driving their emergence.

Group Identity

Historically, initial private emergence has frequently, perhaps usually, been tied to religious beliefs or grouping. A particular catalyst was often the secularization of pre-existing institutions, along with the fresh creation of markedly secular public institutions. Thus, public higher education today is overwhelmingly secular. In contrast, much of the private sector remains discernibly (though not overwhelmingly) religious. Religious higher education has been a way to promote – although largely just to protect – the identity of a religiously based group.

Two principal changes have recently modified the picture. One is the increasing mix of religions. In Latin America, most markedly, but in other regions as well, Catholic predominance yields to an evangelical Protestant emergence as an expression of the growing evangelical population. A quite contrasting culture arises where Islamic higher education also emerges. Where Muslims are a minority, PHE is an option; where they are a majority, religion may find expression in the public sector. Perhaps the region with the most dramatically surging PHE religious proliferation (Catholic, evangelical, and Islamic) is Africa.

The other change has been the shrinking share of the religious within the cultural identity category. This may or may not reflect falling religious enrolment but it certainly reflects a rise of outlets for nonreligious identities. The most important of these identities is ethnic. Naturally, the phenomenon may be particularly strong in countries with quite heterogeneous populations. As on the religious side, so also on the ethnic side, PHE is often both protective and promotional in intent and practice. It has played those roles even in societies of relative religious tolerance and pluralism, as in the Protestant majority in the US. In more divided or intolerant societies, the minority cultural niche is often especially defining. South Asia and most of the European and Asian post-communist world provide notable examples.

Sometimes, overlapping a group or religious orientation is a set of values at odds with perceptions of what dominates in the public mainstream. Where these values stress authority, safety, and the like, parents have special PHE interest for their daughters. This may link up with concentration of study in traditionally female fields such as nursing.

Semi-Elite PHE

Although the type of PHE that often generates the greatest interest, elite PHE is very rare. The main exception is the Americas, particularly the US. However, outside the US the private sector is rarely the pinnacle, rarely the top choice for students who have choice. Although some Asian PHE institutions are reasonably seen as elite, they almost never match the public top; Japan has a few

privates in the top ten (e.g., Keio University and Waseda University) but only in South Korea are privates dominate the top ten. The much publicized global rankings produced in China and the United Kingdom confirm the only paltry appearance of non-US PHE. Accordingly, here we emphasize semi-elite PHE. In Africa, Asia, and post-communist Europe, a set of public universities continues to dominate the top, so that semi-elite private universities may compete principally against public universities in the next tier.

Important institutions in Latin America may well merit an elite label, but we must highlight limitations. For one thing, in countries such as Brazil and Argentina, the most prepared students continue disproportionately to choose leading public universities. Elsewhere, even where PHE holds an overall lead and there is disenchantment with quality and disorder in the public sector, the best publics usually remain the top choice in certain scientific and medical fields as well as in other traditional professions (e.g., law). Moreover, full-time teaching, graduate education, and research remain the exception even toward the top of the region's PHE. In both Latin America and beyond, it is mostly in niches that semi-elite concentrates and even approaches the elite level. Business fields are the leading example, including at the master's level.

Criticism and praise for the semi-elite subsector partly echo that associated with the elite subsector. A major example is the lament that students tend to come from privileged families; another is that they are too non-national, too tied to US models, English, or fitting into the global economy. At the same time, semi-elite institutions lack the quality and breadth of the elite counterparts. Conversely, they are serious institutions providing good opportunity below the pinnacle but ahead of the bulk of the sector or system. They often pursue success within high-demand areas associated with lucrative employment. Witness the 1990s enrolment surge in academies of economics (business), rising from zero to nearly 300 000 (Kraft and Vodopivec, 2003). Many semi-elite institutions are entrepreneurial and competitive, demonstrating capacity to improve and enlarge, sometimes using gains from fields with heavy demand (yielding substantial income) into more costly fields, facilities, and faculties. However, it is difficult to rise above semi-elite. The already-established top places, public and private, boast the highest level of students, faculty, competitive grants, donors, political muscle, and alumni.

Nonelite Demand Absorbers

Were importance defined only by numbers, the clearly nonelite subsector would warrant the bulk of our attention. Here, most students are not choosing their institutions over other institutions as much as choosing them over nothing. The crux of the nonelite subsector is demand

absorption as the soaring demand for higher education – fueled by economic growth and an enlarging middle class, as well as by copying the examples of others – exceeds the public supply of slots as governments conclude they are no longer able to finance most new enrolment. Further contributing to private nonelite proliferation has been a lax regulatory environment at least for some initial period of proliferation (Levy, 2006), although this factor has been weaker in East Asia and the Middle East than in Africa and Latin America. In every region in which PHE becomes the majority sector (and in many where it becomes a large minority sector), it is this demand-absorbing subsector that has been numerically dominant, usually increasingly so.

The private nonelite subsector often includes institutions labeled universities but the label is questionable. Moreover, this is the subsector most concentrated in institutions not even labeled universities. Indeed, many private institutions are technical or vocational institutions, ambiguously perched on the definitional borderline marking higher education as well as on a for-profit, non-profit border (Atchoarena and Esquieu, 2002).

Evaluation of the nonelite sector is often extreme, especially on the negative side, reflecting a mix of real observation, traditional prejudice, and stereotype. Unfortunately, the negative critique is often apt. Fly-by-night institutions operate in search of financial reward, shamelessly profiting from the large demand–supply gap. They are most often quite small, sometimes family owned, with much of what they do quite nontransparent. Here is where we mostly find institutions that are legally nonprofit yet functionally for profit; in addition, among the legally for-profit institutions the overwhelming number is in the nonelite private subsector. Attempts to justify these demand absorbers are strained but become serious where we turn to demand absorbers that pursue aims beyond financial gain alone. The best of these institutions tend to be focused on useful training and teaching (Education Commission of the States and Kelly, 2001; Cao, 2007). As the whole nonelite subsector is composed of less-privileged clientele than found in much or most of higher education, it makes a substantial contribution to access – in often egregiously unequal societies. Moreover, the serious institutions may well strive for improvement, even hoping eventually to border the semi-elite subsector. In contrast, the fly-by-night organizations are vulnerable to national tightening of licensing standards or introduction of accreditation systems.

The Finance, Governance, and Accountability Nexus

Issues of finance and governance as well as accountability are also addressed elsewhere in the encyclopedia. How does

PHE fit? We explore this question with distinctions by type of private subsector. In general, private–public differences are formidable, notwithstanding overlap and blurring.

Finance

PHE operates basically on private money. Across all regions, and types of PHE, as well as for all the decades of private growth, a salient characteristic of PHE is its huge reliance on private income. Most private institutions are fully or almost fully tuition dependent.

Exceptions go all the way to predominantly public finance, but these exceptions have been few. Belgium and the Netherlands are the long-standing cases, matched by Chile around the 1960s and 1970s, where private–public higher education differences overall have been limited. The limitations stem from government coming to treat private institutions the way it treats the public ones. This means public funding of both public and private sectors. In India, private colleges became basically public institutions after the nation's independence; they are affiliated to universities (Agarwal, 2006). Moreover, even these cases do not suggest private–public near uniformity in funding the whole system. Additionally and strikingly, quite private institutions may emerge alongside the publicized private predecessors. Perhaps the most dramatic case is Chile, where the military government and neoliberal economists hatched the idea of a whole wave of new private institutions (Bernasconi, 2003). The prior, publicized institutions are now referred to as old privates to distinguish them from the markedly private institutions. Similarly, India's neoliberal policies and economic boom have contributed to the rise of a quite private sector, privately funded.

Much less dramatic but more and increasingly common is government funding of some, but not the bulk, of PHE income. Provision is linked to some desired criteria other than private–public status. Following US precedent, two prominent modes are funding to students (usually based on need), rather voucher like, and competitive funding of research grants. Such funding remains an idea more than a reality for the bulk of the world's PHE but the reality is expanding. In Israel, students at private colleges are eligible for public funds. In New Zealand, private institutions are eligible for public money. In Thailand, as in other parts of East Asia, PHE can receive grants in aid, income-contingency loans, and the like. Added to the increasing share of public higher education funding that comes from private sources, this public funding of PHE makes for an intersectoral blurring. However, convergence would be too strong a term, since the key comparison remains stark: most public higher education funding is public, most PHE funding is private.

A few PHE subsectoral tendencies merit mention. Religious institutions have been recipients of public funding in some cases where they are long-standing, have shown a

degree of academic seriousness, and largely emulate the public sector except for religious mission. However, cultural institutions built around ethnic minorities may present particular political difficulties in regard to potential public funding. Naturally, any sort of religious or ethnic institution may receive private contributions, thus diversifying the private funding base beyond tuition and fees alone.

Elite privates have a substantial claim on competitive public or private (e.g., foundation, domestic or international) funds in terms of quality, including ability to compete for sector-blind funding pools – but elite PHE is rare. Semi-elite PHE shares some of that claim and particularly highlights the possibilities of contracts and applied research funding. In addition, semi-elite institutions can charge high tuitions and have some realistic prospect for future alumni giving. Corporate funding is another possible source, including where for-profit PHE is tied to corporations.

Where tuition and fees are usually the nearly complete source of funding is in the demand-absorbing subsector. Yet even there the upper-end, serious institutions may be able to broaden their income profiles, including that if public funds are available for PHE it is often for accredited institutions.

In the financial blurring of sectors, however, weightier than additional public funding of PHE is additional private funding of public higher education. As with so much of the global transformations of higher education, the US has been the leader both by its less statist and more market-oriented traditions and by its intensification of those tendencies; the market penetrates very deeply into even the (public and nonprofit private) high-end research universities (Geiger, 2004). The introduction of tuition public-sector fees remains heavily opposed for a variety of reasons; however, it is also increasingly overcome – sometimes not only by introducing nontuition fees, or by introducing tuition at the master's level, but also by direct and widespread imposition at the undergraduate level. A particularly striking African and East European manifestation – reflecting our earlier point about private–public competition above the low end of the system – is public institutions' admission of private, paying students alongside their subsidized ones, sometimes in semi-elite beachheads in market-oriented fields. Another intriguing recent development (as in South Africa, China, and Russia) is partnership between public universities and private colleges, the former bringing power, academic resources, and standing, the latter bringing in tuition-paying students to the public universities. Yet the main twentieth-century and persisting public norm is state funding of public institutions.

Accountability and Governance

Much of what can be said about PHE accountability relates to the nature of funding, as well as to parallel

points about governance. However, whereas finance is mostly clear and measurable, accountability and even governance are vaguer concepts, more subject to interpretation.

Self-financed, most PHE institutions are self-governed. At least this tends to be the case with regard to government, whereas autonomy may be limited by private funders to whom accountability may be owed (e.g., churches, businesses, foundations, and families).

However, as by far the main funders are the students' families, one could reason that PHE is accountable to them. True in many cases, the proposition is dubious where we are dealing with fly-by-night demand absorbers in situations of demand vastly exceeding supply, an easy market for private institutions. Of course, they still need to attract students but many students have little alternative between no higher education and this private education. Thus, these private institutions may have to deliver much less than they may promise, that is, the accountability is low; or it may be enough for institutions to offer up a loose copy of what public or other private universities provide, thus garnering sufficient credibility and legitimacy to survive. Yet, when demand slackens, low-status private institutions face a much greater threat and may have to compete more and be much more accountable than they had been. This is the present case in Japan and there are echoes in Eastern and Central Europe, also owing much to changing demographics. Similarly, US Catholic and other nonelite institutions find themselves sinking or swimming in a more difficult market than before. It is likely that semi-elite institutions need to be more accountable than demand-absorbing ones when it comes to satisfying paying clients, given that fees are higher, clientele more privileged, and institutional aspirations clearer and more ambitious. Accountability to pertinent businesses (funders and/or employers) is also a logical reality for semi-elite institutions.

To whatever extent PHE institutions might provide accountability, and to whom, the accountability is often tied to comparatively hierarchical management at the institutional level. These institutions usually make little pretence to internal democracy or broad participation in governance. In other words, the risk of institutions being forced internally to change their mission and profile of accountability is limited. Student and faculty power pales compared to what is found not infrequently in the public sector. Faculty are mostly part time (including many who are full time at public universities), unionization minimal, and professors lack high academic training and status. On the student side, the classic formulation of student choice (of institutions) without much voice inside holds. The formulation is especially defining in for-profit institutions, where faculty power is especially weak (Kinser, 2006).

PHE accountability tends to be rather targeted to particular groups, especially those that pay. Key is that accountability need not be widely spread out. A typical

PHE institution need not be responsive to broad wishes of a broad public. Nor need it be responsive to private actors that are crucial at another private institution. By the same token, however, its accountability to its few lords may have to be deep and substantial. A common for-profit claim is that they prove themselves through measurable results, usually in the job market. These results can be crucial to their legitimacy, whether they also pursue legitimacy through accreditation (Kinser, 2007).

In contrast, public university accountability tends to be more diffuse, often at least with a claim of responsiveness to the general public. This tendency also suggests a more prominent place for government than found for private institutions, although this tendency is blunted where public universities enjoy traditions and guarantees of substantial autonomy, a reality that today fuels widespread government proclamations that public institutions must become much more accountable, including measurably so.

The extreme of PHE nonaccountability to government relates back to the unplanned explosion of PHE. Yet the deleterious effects of this lack of accountability, combined with operation in nontransparent and easy markets, have increasingly led to more imposition of rules, what we might call delayed regulation (Levy, 2006). Some rules are system wide, others targeted at PHE. Accreditation is a major global example normally tied to system-wide rules. Private-public contrasts may further diminish as public universities are urged, including by government, to institute more management techniques associated with private sectors. The new contract of greater autonomy in exchange for clearer accountability in terms of results can be seen as a form of partial privatization in the public sector.

Conclusion

PHE growth has substantially reshaped the higher education landscape. Once the norm, public monopolies and near monopolies have become uncommon. When we refer to higher education we must be sensitive to the reality of two sectors and the differences between them.

PHE growth has been spectacular in recent decades, mainly in the developing and transitional world, where there still appears to be much potential for further expansion. Across regions and within them, the strength and extent of PHE growth have varied, but no region has been unaffected.

Also connected to variation is that there are several different principal types of PHE. Numerically and increasingly the dominant form is nonelite, absorbing demand unaccommodated by the public sector. Within this subsector, in particular, it is crucial to try to distinguish between fly-by-night institutions and serious ones usually oriented to the job market. On the other end of the PHE subsectoral spectrum are elite universities but unless

the term is used liberally, few such institutions exist, although of course the exceptions are noteworthy. Much more common are semi-elite institutions, sharing some characteristics with high-end, nonelite counterparts, and often doing well in status and results in the entrepreneurial realm. Finally (and sometimes overlapping the semi-elite), religiously and culturally oriented institutions maintain a strong presence. Growing in important respects, receding in others, they contribute considerable diversification to the private sector and to higher education overall.

PHE growth is significant not just for the numbers alone but for the fundamental differences that generally characterize the two sectors. Of course, the differences are not uniform and we can identify tendencies of increased public roles in PHE and increased privatization within public higher education. However, for the most part, the public sector remains quite public, whereas private institutions are privately financed and have governance and accountability profiles consistent with their funding and purpose.

Bibliography

- Agarwal, P. (2006). Higher education in India: The need for a change. Indian Council for Research on International Economic Relations, New Delhi *Working Paper* #179.
- Altbach, P. G. and Levy, D. C. (eds.) (2005). *Private Higher Education: A Global Revolution*. Boston, MA: Sense Publishers, Centre for International Higher Education/Boston College and PROPHE (Program for Research on Private Higher Education)/University at Albany.
- Atchoarena, D. and Esquieu, P. (2002). *Private Technical and Vocational Education in Sub-Saharan Africa: Provision Patterns and Policy Issues*. Paris: International Institute for Education Planning (IIEP)/UNESCO.
- Bernasconi, A. (2003). *Organizational Diversity in Chilean Higher Education: Faculty Regimes in Private and Public Universities*. PhD Dissertation. Boston University, Boston, MA, USA.
- Cao, Y. (2007). *Chinese Private Colleges and the Labor Market*. PhD Dissertation. University at Albany, NY, USA.
- Collier, M. B. (in progress). *Student Choice of Non-Elite Private Colleges*. PhD Dissertation, University at Albany, NY, USA.
- Education Commission of the States and Kelly, K. F. (2001). *Meeting the Needs and Making Profits: The Rise of the For-Profit Degree-Granting Institutions*. Denver, CO: Education Commission of the States.
- Geiger, R. (1986). *Private Sectors in Higher Education: Structure, Function, and Change in Eight Countries*. Ann Arbor, MI: University of Michigan Press.
- Geiger, R. (2004). *Knowledge and Money: Research Universities and the Paradox of the Marketplace*. Stanford, CA: Stanford University Press.
- Gupta, A., Levy, D. C., and Powar, K. B. (eds.) (2008). *Indian Private Higher Education: International Perspectives*. Delhi: Shipra.
- Jablecka, J. (2007). Legitimation of nonpublic higher education in Poland. In Slantcheva, S. and Levy, D. C. (eds.) *Private Higher Education in Post-Communist Europe: In Search of Legitimacy*, pp 179–200. New York: Palgrave/MacMillan.
- Kinser, K. (2006). *From Main Street to Wall Street: The Transformation of For-Profit Higher Education*. San Francisco, CA: Association for the Study of Higher Education/Jossey-Bass.
- Kinser, K. (2007). Sources of legitimacy in U.S. for-profit higher education. In Slantcheva, S. and Levy, D. C. (eds.) *Private Higher Education in Post-Communist Europe: In Search of Legitimacy*, pp 257–278. New York: Palgrave/MacMillan.
- Kinser, K. and Levy, D. C. (2006). For-profit higher education: United States tendencies, international echoes. In James, F. and Altbach, P. (eds.) *International Handbook on Higher Education*, pp 107–120. New York: Springer.
- Kraft, E. and Vodopivec, M. (2003). The new kids on the block: The entry of private business schools in transition economies. *Education Economics* 11(3), 239–257.
- Landoni, P. (forthcoming). How do private higher education concepts and configurations play out at the graduate level? A case study of Uruguay. *PROPHE Working Paper*, University at Albany.
- Levy, D. C. (1986). *Higher Education and the State in Latin America: Private Challenges to Public Dominance*. Chicago, IL: University of Chicago Press.
- Levy, D. C. (1992). Private institutions of higher education. In Clark, B. and Neave, G. (eds.) *The Encyclopaedia of Higher Education*, vol. 2, pp 1183–1195. New York: Pergamon Press.
- Levy, D. C. (1996). *Building the Third Sector: Latin America's Private Research Centers and Non-Profit Development*. Pittsburgh, PA: University of Pittsburgh Press.
- Levy, D. C. (2006). The unanticipated explosion: Private higher education's global surge. *Comparative Education Review* 50(2), 217–240.
- Levy, D. C. and Zumeta, W. (eds.) (forthcoming). Private higher education and public policy around the world: Influences and implications. *Journal of Comparative Policy Analysis*.
- Mabizela, M., Levy, D. C., and Otieno, W. (eds.) (2008). *Special Issue: Private Surge amid Public Dominance: Dynamics in the Private Provision of Higher Education in Africa*. *Journal of Higher Education in Africa*.
- Paschuavili, M. (2007). Exploring private higher education growth across post-communist countries: The case of Georgia. *PROPHE Working Paper*. University at Albany, NY, USA.
- Scheker, A. (2007). Comparing school-level to private higher education: Using the Dominican Republic as a pioneer study. *PROPHE Working Paper*. University at Albany, NY, USA.
- Slantcheva, S. and Levy, D. C. (eds.) (2007). *Private Higher Education in Post-Communist Europe: In Search of Legitimacy*. New York: Palgrave/MacMillan.

Relevant Websites

<http://www.albany.edu> – University at Albany-SUNY Program for Research on Private Higher Education, with Publications, Data, Laws, and Global News Features.

This page intentionally left blank

IMMIGRATION

Contents

Children of Migrant Populations

Third Culture Kids

Children of Migrant Populations

M M Suárez-Orozco and C Suárez-Orozco, Immigration Studies at New York University, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Ours is the age of movement. Approximately 200 million transnational migrants are living beyond the countries of their birth – and hundreds of millions more are internal migrants within the confines of ever-changing nation-states. What drives global migration? First, the global integration of production, distribution, and consumption of goods and services stimulate migration because where capital flows, immigrants follow. Second, new information, communication, and media technologies enable global economic production and also stimulate migration by generating new desires, tastes, consumption practices, and lifestyle choices. Third, globally integrated economies increasingly rely on foreign workers (in north-western Europe and Japan due to demographic considerations: low fertility rates and rapidly aging populations) both in the well-remunerated knowledge-intensive sector and in the least desirable sectors of the economy. Fourth, the affordability of mass transportation has put the option of migration within the reach of millions who, heretofore, could not do so. Fifth, globalization stimulates new migration because it has produced uneven results – wage differentials, when controlled for cost-of-living differences, are on the order of 4 –1 in many North–South migration corridors.

In the field of global migration, remarkably, the well-being and future of the children of immigrants are all too often neglected. The children of immigrants are a fast-growing sector of the child and youth population in countries such as Australia, Canada, Germany, Italy, the Netherlands, Spain, Sweden, and Spain. In the United States, the country with the largest number of immigrants

in the world, approximately 25% of all youth, are of immigrant origin and it is projected that by 2040 over one-third of all children will be growing up in immigrant households. Schools in cities, large and small, the world over, from New York to Beijing, from Barcelona to Toronto, from Sydney to Reggio Emilia, are being transformed by growing numbers of immigrant children. Just as schools face the challenge of educating linguistically, culturally, and racially diverse students, globalization imposes yet another challenge on education – nurturing ever more complex skills, competencies, and sensibilities on students to equip them to engage in the globally interlinked economies and societies and to become globally conscious and competent citizens in the twenty-first century.

Immigrant-origin youth arrive in new destinations with distinct social and cultural resources: high aspirations for education, meta-cognitive advantages afforded by a dual frame of reference, well-honed skills for developing relationships to help them negotiate unfamiliar territories, and the abilities needed to navigate across difficult circumstances. Their optimism, high aspirations, dedicated hard work, and positive attitudes toward school and ethic of family support for advanced learning contribute to the fact that some immigrant youth in some countries educationally outperform their native-born peers of comparable backgrounds. Many immigrant youth also encounter such a myriad of challenges – xenophobia, economic obstacles, language difficulties, family separations, under-resourced neighborhoods and schools – that they struggle to gain their bearings in an educational system that often puts them on a downward trajectory.

Immigrant youth arrive from multiple points of origin at multiple destinations often adding new threads of cultural, linguistic, religious, and racial difference to the social tapestry of their new homes. Some are the children of educated professional parents (such as Indians in California), while others may have illiterate parents (such as Kurds in Stockholm). Some received excellent schooling (such as the children of emigrants from Hong Kong in Vancouver), while others left educational systems that were in shambles (such as the children of emigrants from Haiti). Some escaped political, religious, or ethnic persecution (such as Iraqi children in Jordan), others are motivated by the promise of better jobs and better educational opportunities (such as Uzbek children in the Russian Federation). Some are documented migrants, while others are unauthorized young migrants. Some join well-established communities with robust social supports, while others move from one migrant setting to another, forcing students to often change schools. The social and educational outcomes of immigrant youth will vary substantially depending upon the specific constellation of resources and the settlement context.

The global economy is largely unforgiving to those who do not achieve formal education. Indeed, schooling processes and outcomes shape socioeconomic mobility. Recent studies suggest that, while some children of immigrants are successfully navigating the educational system in their new societies, large numbers struggle academically, leaving schools without acquiring the tools that will enable them to function in the globally competitive labor market.

Academic trajectories and performance are multiply determined by an alchemy of family-background variables, the quality of schools that immigrant students encounter, second-language-acquisition challenges, student engagement, and relational supports. Taken together, these factors serve to bolster or, conversely, undermine academic adaptation and success. Next, we review the variables that have the strongest implications for the schooling performance and social adaptation of immigrant-origin youth in new societies: (1) family background; (2) quality of schools; (3) academic language of the new country and general second-language-acquisition challenges; (4) individual level factors; and (5) social supports.

Family Background

Parental Education

Immigrant youth arrive in new countries with varied levels of human capital. On one end of the spectrum, we find youth from upper-status urban backgrounds. Their parents are typically highly literate, and the children arrive with high levels of literacy, and well-developed study skills. These parents are well equipped to guide the children in how to study, access, and make

meaning of data and information, and can provide resources, including additional books, a home computer, Internet access, and tutors. In contrast, there are immigrant youngsters whose parents have little or no formal education. These children may arrive having missed critical years of classroom experience and often cannot read and write adequately in their native language. Such varied experiences and backgrounds will have important implications for the transition to their new countries. Unsurprisingly, youth arriving from families with lower levels of education tend to struggle academically, while those who come from more literate families and with strong educational skills typically flourish.

Poverty

Poverty has long been recognized as a significant risk factor for poor educational outcomes. Children raised in circumstances of socioeconomic deprivation are vulnerable to an array of distresses, including difficulties concentrating and sleeping; anxiety and depression; as well as a heightened propensity for delinquency and violence. Poverty frequently coexists with a variety of other factors that augment risks – such as single parenthood, residence in neighborhoods plagued with violence, gang activity, and drug trade, as well as school environments that are segregated and overcrowded. High poverty is also associated with high rates of housing mobility and concurrent school transitions. All of these stresses are highly disruptive to educational engagement and performance.

There is evidence that large numbers of immigrant-origin students in many contexts suffer from the challenges associated with poverty. The United States offers a worrisome example. In 1999, 22.8% of Latinos, two-thirds of them immigrants or the children of immigrants, were living in poverty, compared with 7.7% of whites. In 2006, the poverty rate for Latino students in the United States (28%) was almost double of that of white children (16%). For immigrant Latino families, poverty reaches much higher percentages. Fully 35% of foreign-born Latino students live in poverty, compared to 27% of native-born counterparts. Immigrant children are more than 4 times as likely as native-born children to live in crowded housing conditions and 37% of Latino immigrant families report difficulties affording food. A large proportion of these children are raised in families where parents are working but many are employed in very low-paying professions with erratic working conditions. Similar patterns of immigrant poverty can be found in other countries of immigration.

Such impoverished family conditions may lead immigrant students to shoulder financial responsibility for the family; many immigrant students contribute to their family income by working after school. According to a recent US survey in 2007, 17% of immigrant-origin Latino students between the age of 16 and 18 held jobs.

Research suggests that working long hours can distract from concentration on school work and is related to lower grades. After-school work is also associated with lower achievement on math and science scores of the National Educational Longitudinal Study (NELS) data, and has a particularly negative effect on boys. Poverty and other heightened stressors have negative implications for educational outcomes for immigrant origin youth.

Undocumented Status

Millions of immigrant children live in new countries without proper documentation. In the United States, for example, nearly 2 million children and youth are unauthorized immigrants. Research suggests that undocumented youth often arrive after multiple family separations and traumatic border crossings. Immigrant children and youth in households with unauthorized members often experience fear and anxiety around being separated from family members if they or a relative is apprehended or deported. Such psychological and emotional duress can take a heavy toll on the academic experiences of children growing up in these homes. Further, while unauthorized youth may in some settings legally have equal access to basic education, they may not have equal access to health, social services, or to jobs. In some countries, undocumented students with dreams of graduating from high school and going on to college may find that their legal status stands in the way of their access to post-secondary educational opportunities. Thus, immigrant students who are unauthorized or who come from unauthorized families suffer from particular burdens.

Seasonal Migrant Youth

Being a child whose parents do seasonal migrant work presents additional challenges – multiple moves, frequent interruptions in schooling, poverty, as well as harsh working and living conditions are realities in their lives. For example, in a representative national survey in the United States, Latinos comprised 83% of the migrant farm workers interviewed, with 75% born in Mexico. Many families lived in conditions of poverty. It is estimated that over 750 000 children travel with their seasonal migrant parents in the United States each year. The lack of continuity in their schooling trajectories (because of interruptions during the school year, the difficulty of transferring school records, health problems, and lack of English-language skills) contributes to an acute risk of low attendance, and high dropout rates.

School Factors

Where immigrant families settle will shape the experiences and adaptations of the children, particularly with

regard to the kinds of schools their children access. Many immigrant-origin students attend segregated schools where the majority of their peers are other immigrant students – segregation is a notable problem in many advanced post-industrial societies with large numbers of immigrant youth, including the United States, Germany, and Sweden. Immigrant segregation often involves race, poverty, and linguistic isolation – so-called triple segregation. Triple segregation is inexorably linked to negative educational outcomes, including academic underperformance, linguistic isolation, in many cases reduced school resources, lower achievement in standardized tests, greater levels of school violence, and higher dropout rates.

Second-Language Learning

While some immigrant children, like Ecuadorians in Spain speak the same language at home as in school, many immigrant youth the world over are second-language learners. Many are receiving suboptimal, if not erratic, bilingual and English (or French, or German, or Italian, or Spanish, etc.) as a second-language instruction. Research has established that it takes 5–7 years of optimal academic instruction to develop academic second-language skills comparative to native-born samples. Yet, for many immigrant students robust second-language-acquisition educational infrastructures and adequate supports are rarely in place. In the United States, the National Assessment of Educational Progress data showed that 70% of English-language learners (ELLs) immigrant-origin students tested in the fourth grade scored below basic in reading, while 1% scored at the advanced level and 71% of the ELLs in the eighth grade scored below basic, while 1% scored at the advanced level.

Academic language difficulties present particular challenges for optimal performance for immigrant students in the current context of educational tracking systems. Even when immigrant students are able to participate in mainstream classrooms, their academic language skills may still be developing and they may struggle, missing subtleties in lectures and discussions, reading more slowly than native speakers, and experiencing difficulty in academic writing. The Program for International Student Assessment (PISA) study quantified the comparative disadvantage of immigrant youth across 17 countries, showing that language-minority immigrant youth are on average 1 year behind their peers in various measures of academic achievement. In a number of European countries of immigration, for example, immigrant-origin children are currently being placed into the lowest academic and vocational tracks before they have sufficient time to learn the academic language needed to keep up with their peers.

For second-language learners in the United States, high-stakes tests have become *de facto* language policy shaping their curriculum and daily instruction in a myriad

of ways; many are indeed tested before their skills are adequately developed – dropping out of the educational system prematurely as a result.

Second-language-acquisition issues can mask the actual skills and knowledge of immigrant youth in both oral and written displays. With more time to develop academic language skills, and more flexibility between educational tracks, however, immigrant students have been shown to sustain high-achievement trajectories. The nearly exclusive focus on learning the majority language and performance on high-stakes tests developed for and normed on native-born speakers has led researchers and policymakers to neglect the immense opportunity immigrant youth have to contribute to a multilingual and multicultural society in the age of global interdependence. As a consequence of this neglect, immigrants in many settings lose their native-language competencies by the second generation in the new society.

Individual-Level Factors

For many immigrant students, the educational journey begins with their migration to the new country. Migration is a transformative process with profound implications for the family as well as the potential for lasting impact on socioemotional development of the child. By any measure, immigration is one of the most stressful events a family can undergo. Immigration removes individuals from predictable contexts – community ties, jobs, and customs, and stripping them of significant social ties, extended family members, best friends, and neighbors. New arrivals who experienced trauma may remain preoccupied with the violence and may also feel guilty about having escaped when loved ones remained behind. Those who are undocumented face an ever-present threat of apprehension that can lead to further traumatic and sudden separations. For some immigrants, the dissonance in cultural expectations, the cumulative stressors, together with the loss of social supports lead to affective and somatic symptoms. Some immigrant parents are relatively unavailable psychologically due to their own struggles in adapting to a new country, thus posing developmental challenges to their children. The immigrant parents, whether their children are of the first or second generation, often turn to them for help in navigating the new society. Children of immigrants are asked to take on parentified roles, including translation and advocacy. Such tasks often fall more on the shoulders of daughters, which have both positive and negative consequences for their development.

Immigrant children and youth also face the challenges of forging an identity and sense of belonging to a country that may reflect an unfamiliar culture, while at the same time honoring the values and traditions of their parents. Acculturative stress has been linked to high levels of

intergenerational conflict as well as psychological and academic problems. Immigrant children are often asked to take on responsibilities beyond their years, including sibling care, translation, and advocacy, which at times undermine parental authority. These often highly gendered roles may have both positive and negative consequences for development.

First-generation immigrant youth face their parents' challenges of adjusting to a new context. Often, the children migrate not simply to a new country but also to new family structures as many are separated for long periods of time from their parents during the course of their migration. During long family separations, divorces can take place, new partners can enter the family system, and new siblings can be born. In such cases, the reunited immigrant child will have to adapt to new family constellations. Further, the first-generation immigrant child must learn a new language going through a difficult transition when they are unable to communicate their thoughts with ease; while some acquire competency over time, most are marked by accents, and others never gain full proficiency. The significant time it takes to acquire academic language in the new country presents significant educational as well as social challenges for immigrant students.

The second-generation students may have limited facility in their parents' native language, which presents other challenges in maintaining communication at home. While the second generation may share a lack of access to those who can guide them through the institutions of the unfamiliar dominant society, they are typically spared the challenges of premigratory trauma, undocumented-status-related stress, and family separations. On the other hand, they often face the stressors of poverty – especially if they live in segregated urban contexts – without the protection of immigrant optimism and a dual frame of reference.

Data examining the well-being of immigrant origin populations in various countries and across generations and ages reveal mixed results according to country of origin, destination, developmental group, cohort, and age of arrival as well as developmental outcome. While there is a fairly consistent immigrant paradox showing a decline across generations with greater length of residency for physical health outcomes and engagement in risk behaviors, the results are inconsistent with regard to the risk to psychological health. Further, the body of evidence on immigrant health has focused on adults and families rather than on adolescents. Immigrant youth of refugee origin appear to be at greatest risk for affective disorders. Immigrant adolescents show patterns of progressive risk-taking behaviors the longer they are exposed to the culture of their new societies. This is also the case for academic engagement – an increasingly important indicator of well-being in the knowledge-intensive economy – which also decreases across time and across generation (in some

cases, after peaking in the second generation). Given the limited and mixed evidence on the developmental trajectories of this growing population of immigrant adolescents, more research on a variety of indicators of their well-being is needed using both qualitative and quantitative lenses.

Social Contexts of Learning

Healthy social support networks are linked to better adjustment and psychosocial outcomes. Interpersonal relationships provide a number of functions. Social companionship, a basic human need, serves to maintain and enhance self-esteem as well as provides acceptance and approval. Instrumental social support provides individuals and their families with tangible aid (such as running an errand or making a loan) as well as guidance and advice (including information, and job and housing leads). These instrumental supports are particularly critical for disoriented immigrant newcomer youth.

The Immigrant Family and Motivation

For many immigrants, social relations play a critical role in initiating and sustaining motivation in a new land. While for mainstream nonimmigrant students achievement is often motivated by an attempt to gain independence from the family, immigrant students are typically motivated to achieve for their families. Family cohesion and the maintenance of a well-functioning system of supervision, authority, and mutuality are perhaps the most powerful factors shaping the well-being and future outcomes of children. For immigrant families, extended family members – godparents, aunts, uncles, older cousins, and the like – are often critical sources of tangible instrumental and emotional support.

Families can support children's schooling by maintaining a value of education and establishing a standard of expectation. Families establish expectations about appropriate behaviors and attitudes vis-à-vis school authorities and peer interactions. They can also actively scaffold children as they complete school assignments. Immigrant parents who work long hours and may have limited schooling are at a distinct disadvantage in this regard. Immigrant parents may be unable to tangibly support their children in ways that are congruent with the new country's cultural models and expectations. Many come from traditions that revere school authorities and expect parents to keep a distance from the day-to-day workings of their child's education.

Communities and Community Organizations

As no family is an island, family cohesion and functioning are enhanced when the family is part of a larger community displaying effective social supports. When immigrants settle in cohesive communities and when adults

within the community are able to effectively monitor, sanction, and support youngsters' activities, they will tend to display better social adaptations and psychological adjustment. Children who live in such communities are less likely to become involved with gangs and delinquency and are likely to be more focused on their academic pursuits.

Youth-serving organizations can enrich immigrant communities and foster healthy development through the support they provide to children, parents, and families. Such urban sanctuaries, often affiliated with neighborhood churches, religious groups, and nongovernmental organizations (NGOs), provide immigrant youth out-of-school time that is not spent in isolation, unsupervised, or on the streets with one's peers. These programs can provide safe havens from the pressures of the streets. Community program staff can serve as culture brokers for youth, bridging the disparate norms in place in children's homes and those in place at school. Adults who work in community programs can provide tutoring, educational guidance, advice about the college-application process, and job-search assistance – information that is often inaccessible to immigrant youth whose parents have not navigated the academic system in the new land. Such programs can aid in counteracting embittered school personnel and toxic schools' impact on the educational trajectories and academic achievement of immigrant youth.

Mentoring Relationships

In nearly every story of immigrant success, there is a caring adult who took an interest in the child and became actively engaged in his/her life. Connections with non-parent adults – a community leader, teacher, church member, or a coach – are important in the academic and social adaptation of immigrant adolescents. These youth are often undergoing profound shifts in their sense of self and are struggling to negotiate changing circumstances in relationships with their parents and peers. Protective relationships with adults can provide immigrant youth with compensatory attachments, safe contexts for learning new cultural norms and practices, and information that is vital to success in schools.

Mentoring relationships may have special implications for immigrant youth as during the course of migration, loved ones are often separated from one another and significant attachments are ruptured. Mentoring relationships can give immigrant youth an opportunity to be involved in reparative relationships engendering new significant attachments. In some cases, the guidance and affection of a mentor may help to fill the void created by parental absence. The mentor, as an adult who has been in the Netherlands longer than the *protégé*, can also provide information about and exposure to Dutch cultural and educational institutions, and help as the adolescent negotiates developmental transitions. If the mentor is of

the same immigrant background as the *protégé*, he or she can interpret the rules of engagement of the new culture to parents and, hence, help to attenuate cultural rigidities. Furthermore, bicultural mentors can serve as role models in the challenging process of developing a bicultural identity, exemplifying the ways in which elements of the ethnic identity can be preserved and celebrated even as features of the more mainstream culture of the new country are incorporated into youth's lives.

Peer Relationships

Peers can provide emotional sustenance that supports the development of significant psychosocial competencies in youth. Peers can specifically serve to support or detract from academic engagement. By valuing (or devaluing) certain academic outcomes and by modeling specific academic behaviors, peers establish the norms of academic engagement. Peers can tangibly support academic engagement by clarifying readings or lectures, helping one another in completing school assignments, and by exchanging information (about examinations, helpful tutors, volunteer positions, and college pathway knowledge). However, as immigrant youth often attend segregated schools, they may have limited access to knowledgeable networks of peers.

Taken together, networks of relationships can make a significant difference in immigrant youths' lives. They can serve to help immigrant youth develop healthy bicultural identities, engender and sustain motivation, and provide specific information about how to best navigate schooling pathways. When successful, these relationships help immigrant youth and their families overcome some of the barriers associated with poverty and discrimination that prevent full participation in the new country's economic and cultural life.

Immigrant-origin youth are the fastest-growing student population in a number of countries. They often arrive with an optimism and hope in the future that must be cultivated and treasured – almost universally they recognize that schooling is the key to a better tomorrow. Over time, however, too many immigrant youth, especially those enrolling in impoverished, segregated, and mediocre schools, face negative odds and uncertain prospects. Too many leave schools without developing and mastering the kinds of higher-order skills needed in today's global economy and society. The future of our world will in no small measure be tied to the fortunes of these new young immigrants.

Further Reading

- Banks, J. (ed.) (2007). *Diversity and Citizenship Education: Global Perspectives*. San Francisco, CA: Jossey Bass.
- Cohen, J. E., Bloom, D. E., and Malin, M. B. (eds.) (2007). *Educating All Children: A Global Agenda*. Cambridge, MA: MIT Press.
- Crul, M. (2007). Globalization and education: The integration of immigrant youth in Europe. In Suárez-Orozco, M. (ed.) *Learning in the Global Era: International Perspectives on Globalization and Education*, pp 277–286. Berkeley, CA: University of California Press.
- Crul, M. and Doornik, J. (2003). The Turkish and Moroccan second generation in the Netherlands: Divergent trends between and polarization within the two groups. *International Migration Review* **37**, 1039–1064.
- Cummins, J. (1991). Language development and academic learning. In Malavé, L. M. and Duquette, G. (eds.) *Language, Culture, and Cognition*, pp 161–175. Clevedon: Multilingual Matters.
- Fuligni, A. (1997). The academic achievement of adolescents from immigrant families: The roles of family background, attitudes, and behavior. *Child Development* **69**, 351–363.
- Fuligni, A. J. and Yoshikawa, H. (2003). Socioeconomic resources, parenting, and child development among immigrant families. In Bornstein, M. and Bradley, R. (eds.) *Socioeconomic Status, Parenting, and Child Development*, pp 107–124. Mahwah, NJ: Erlbaum.
- Hugonier, B. (2007). Globalization and education: Can the world meet the challenge? In Suárez-Orozco, M. (ed.) *Learning in the Global Era: International Perspectives on Globalization and Education*, pp 212–229. Berkeley, CA: University of California Press.
- Kasitz, P., Mollenkopf, J., Waters, M., and Holdaway, J. (2008). *Inheriting the City: The Children of Immigrants Come of Age*. Cambridge, MA: Harvard University Press/Russell Sage Foundation.
- Lacy, V. (2008). Diversity as an expression of the modern Canadian education system. In Meier, J. and Morehouse, C. (eds.) *Immigrant Students Can Succeed: Lessons from Around the Globe*, pp 37–45. Berlin: The Bertelsmann Foundation.
- Lansford, J. E., Deater-Deckard, K., and Bornstein, M. (eds.) (2007). *Immigrant Families in Contemporary Society*. New York: The Guilford Press.
- McAndrew, M. (2007). The Education of immigrant students in a globalized world: Policy debates in a comparative perspective. In Suárez-Orozco, M. (ed.) *Learning in the Global era: International Perspectives on Globalization and Education*, pp 255–277. Berkeley, CA: University of California Press.
- Meier, J. and Morehouse, C. (eds.) (2008). *Immigrant Students Can Succeed: Lessons from Around the Globe*. Berlin: The Bertelsmann Foundation.
- Menken, K. (2008). *English Language Learners Left Behind: Standardized Testing as Language Policy*. Clevedon, UK: Multilingual Matters.
- Organization for Economic Cooperation and Development (2003). *Where Immigrant Students Succeed*. http://www.oecd.org/document/44/0,3343,en_32252351_32236173_36599916_1_1_1_1,00.html (accessed September 2009).
- Portes, A. and Rumbaut, R. (2001). *Legacies: The Story of the Second Generation*. Berkeley, CA: University of California Press.
- Sirin, S. and Fine, M. (2008). *Muslim American Youth: Understanding Hyphenated Identities through Multiple Methods*. New York: New York University Press.
- Suárez-Orozco, C., Pimentel, A., and Martin, M. (2009). The significance of relationships: Academic engagement and achievement among newcomer immigrant youth. *Teachers College Record*. <http://www.tcrecord.org/Content.asp?ContentId=15342> (accessed September 2009).
- Suárez-Orozco, M. and Qin, D. (eds.) (2004). *Globalization: Culture and Education in the New Millennium*. Berkeley, CA: University of California Press.
- Suárez-Orozco, C. and Qin, D. (2006). Psychological and gendered perspectives on immigrant origin youth. *International Migration Review* **40**(1), 165–199.
- Suárez-Orozco, C. and Suárez-Orozco, M. (2002). *Children of Immigration*. Cambridge, MA: Harvard University Press.
- Suárez-Orozco, C. and Suárez-Orozco, M. (2007). Moving stories: Immigrant youth adapt to change. *Du Bois Review: Social Science Research on Race* **4**, 351–359.
- Suárez-Orozco, C., Suárez-Orozco, M., and Todorova, I. (2008). *Learning a New Land: Immigrant Students in American Society*. Cambridge, MA: Harvard University Press.

- Suárez-Orozco, C. and Todorova, I. (eds.) (2003). *Understanding the Social World of Immigrant Youth*. San Francisco, CA: Jossey-Bass.
- Suárez-Orozco, C., Todorova, I., and Louie, J. (2002). Making up for lost time: The experience of separation and reunification among immigrant families. *Family Process* 41(4), 625–643.
- Süssmuth, R. (2007). On the need for teaching intercultural skills: Challenges for education in a globalizing world. In Suárez-Orozco, M. (ed.) *Learning in the Global Era: International Perspectives on Globalization and Education*, pp 195–212. Berkeley, CA: University of California Press.
- United Nations Global Commission on International Migration (2005). *Migration in an Interconnected World: New Directions for Action*. Geneva: Switzerland GCIM Secretariat.

Third Culture Kids

R E Van Reken, Families in Global Transition, Indianapolis, IN, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Adult cross-cultural kid (ACCK) – A person who has grown up as a cross-cultural kid (CCK).

Adult third culture kid (ATCK) – A person who has grown up as a third culture kid (TCK).

Cross-cultural kid (CCK) – A child who lives in – or meaningfully interacts with – two or more cultural environments for a significant period of time during the first 18 years of life.

Hidden diversity – A diversity of experience that shapes a person's life and worldview but is not readily apparent on the outside, unlike the usual diversity markers such as race, ethnicity, nationality, etc.

Hidden immigrant – A child who has moved to host or passport culture and, while physically resembling members of the surrounding culture, has had a different life experience shaping his or her deeper layer of culture.

Third culture kid (TCK) – A child who accompanies parents into another culture, usually for a career's sake. This person spends a significant period of time in developmental years outside the parent's passport culture.

Before World War II, most communities – whether in the city of Chicago, the suburbs of London, or the village of Timbuktu – shared one thing in common: they were basically monocultural societies. In such an environment, children easily learned their sense of personal and cultural identity because who was 'us' and who was 'them' were clear. Role models for what life would be like in the future abounded. In general, the cultural environment in which educators had grown up was similar to the one their students knew.

Since the end of World War II, however, global shifts have had major, but often unrecognized, ramifications for educators around the globe. As ease of transportation and communication has exponentially increased the number of international businesses and organizations, millions of children worldwide are now being raised amid a plethora of cultural worlds. While recent increases in immigrant populations have been obvious to educators, the growing population of children identified by sociologists as *third culture kids* (TCKs) has sometimes been more invisible. For educators to remain maximally effective in today's globalizing world, it is important that they understand not

only what the common characteristics of this population are, but more importantly, why these characteristics occur.

Defining Third Culture Kids (TCKs)

In the mid-1970s, a sociologist from Michigan State University, Ruth Hill Useem, traveled to India to complete a research project. While there, she became fascinated with the dynamics she observed among internationally mobile families who had gone from their home, or first culture, to live in a host, or second culture, because of a career choice, not as permanent immigrants. Useem noticed that these expatriates were not living as they would have lived in their home culture nor the way members of the local host culture lived. They had, however, developed a basic lifestyle they shared as expatriates. Useem called this an interstitial, or third, culture (see **Figure 1**) – one lived out between the first and second (Useem, 1973).

Useem named the children who grew up in this internationally mobile experience third culture kids (see **Figure 2**) and defined them simply as "children who accompany their parents into another culture" (Useem, 1993: 1).

In the mid-1970s another intercultural pioneer, David C. Pollock, expanded the definition and described TCKs as individuals who spend "a significant part of their developmental years outside the parents' passport culture(s)." (Pollock, 1989: 1). Again, his focus was on internationally mobile children, not immigrants. Pollock's emphasis on the developmental years differentiates a cross-cultural experience for TCKs from that of the adults who grew up in a more traditional, monocultural upbringing and move cross-culturally for the first time as adults. While they, too, are definitely changed by cross-cultural sojourns, when these adults move to a new place, they have already developed a strong sense of their personal and cultural identity. For children, this is not so. The key developmental tasks of childhood, such as forming a sense of identity, belonging, and cultural competency, become more complicated when the cultural mirror by which children define themselves is forever changing.

Pollock also describes common outcomes of a global childhood.

The TCK frequently builds relationships to all of the cultures, while not having full ownership in any. Although elements from each culture may be assimilated into the TCK's life experience, the sense of belonging is most often in relationship to others of similar background. (Pollock, 1989: 1)

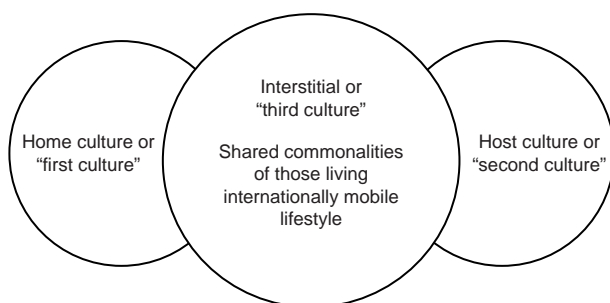


Figure 1 Third culture model. From David C. Pollock/Ruth E. Van Reken, used by permission.

Related Terms

Those who have grown up as TCKs were defined as adult TCKs (ATCKs) during the mid-1980s. Ruth Van Reken conducted the first ATCK research project in 1986 (Van Reken, 1989). Useem and Ann Baker Cottrell followed with a comprehensive project among ATCKs in 1993 (Useem and Cottrell, 1993).

Norma McCaig, a business ATCK, coined the term *global nomads* as another way to identify TCKs. Her definition reads,

Global nomads are persons of any age or nationality who have spent pre-adult years living in one of more countries outside their country of passport because of a parent's occupation. (McCaig, 1987: 2)

During this same time frame, other researchers, including Momo Kano Podolsky, were studying Japanese children raised abroad. In Japanese, *kaigai* means overseas, *kikoku* means return to home country, and *shijo* is the word for children. Thus, they identified those children living with their parents outside of Japan as *kaigai-shijo* (literally, overseas children) and named them *kikoku-shijo* (literally, returned home children) when they returned to Japan (Podolsky, 1994).

Understanding the TCK Experience

While TCKs may not share similar traditions, ethnicity, nationality, or race, the external experience they share of growing up cross-culturally in a highly mobile lifestyle creates many common internal and external responses. Culture is also a system of shared concepts, beliefs, values, and a worldview from which people interpret and make sense of life and the world around them (Hiebert, 1983). In fact, Ximena Vidal used the TCK experience to demonstrate a new way some sociologists were beginning to define belonging and identity through a shared experience rather than through the traditional models of race, tribe, and nationality (Vidal, 2000).

What TCKs Share

Virtually all TCKs share three things which are different from children raised in the more traditional monocultural environments of the past.

1. TCKs grow up in a genuinely cross-cultural world. During their formative years, TCKs don't just observe various cultures, they interact deeply with them – moving back and forth between the cultural milieus, negotiating relationships on a daily basis between those of different backgrounds.
2. The highly mobile nature of the third culture lifestyle is obvious for TCKs who have lived in many countries. But even when TCKs live in one place for an extended time, they still travel back and forth between home and host cultures at regular intervals and others in their expatriate community constantly come and go.
3. Expected repatriation is another norm for TCK families. Whether or not all do repatriate is another question, but most adults working internationally expect that one day they will return to live in their home country again. Parents make educational choices based on this expectation.

Common Benefits and Challenges

It is from these shared experiences that the often paradoxical characteristics of the TCK profile emerge and become apparent by young and ongoing adulthood. Obviously, not every TCK has the same experience or reaction to it. Personality, family of origin, home and host cultures, all play a role. Through the years, however, well-documented themes occur. The following are some of the benefits and often corresponding challenges (Pollock and Van Reken, 2009).

Large worldview versus ignorance of own culture

Many TCKs develop a large worldview from first-hand interactions with people of many cultures which can help TCKs develop strong negotiating skills on a personal and group level.

Conversely, TCKs may be ignorant of cultural nuances in their passport country. This can lead to a sense of uneven maturity. In some ways, TCKs seem wise beyond their years. In other ways, they can be socially slow.

3-D view of the world versus painful view of reality

TCKs have seen places that others only read or watch on TV.

With that privilege, they have also often witnessed first hand the pain and suffering that poverty, bad governance, natural disasters, or war bring. Increasingly, TCKs may themselves be involved in traumatic situations because of political unrest in areas where their parents are working.

Some may suffer depression or post-traumatic stress syndrome as a result of such experiences.

Large network of friends worldwide versus avoidance of close relationships locally

Most TCKs have friends all over the world and try to keep up with them via incessant text messaging, Facebook, or Skype.

This can, however, keep some TCKs from moving on in the current scene. They don't want to become close to others one more time only to say goodbye again. Others feel they are being disloyal to old friends if they move on.

Multilingual versus potential confusion

The TCKs often learn more than one language. For most, this is a great gift which frequently can be used in later careers.

For some TCKs, however, the issue of multiple language learning may create special challenges during their school years. Many TCKs attend school where their mother tongue is not the language of instruction. While the child may learn the language used at school, if parents are not fluent in this language, the child may have little help when projects are assigned as homework (Van Reken and Rushmore, 2009). In addition, while such TCKs may converse in their mother tongue at home, they often lack the academic vocabulary needed to continue higher education in that language after they repatriate (Pollock and Van Reken, 2009). One other particular challenge related to language often occurs when TCKs move to a country where their mother tongue is the one used at school. While they and their educators presume language will be no barrier, both may soon discover that spellings can differ from country to country or a particular word can have vastly different meanings (Bethel and Van Reken, 2003). In addition, Meneses demonstrates the importance of language in forming a sense of personal and cultural identity. If students have lost their mother tongue, this can have significant effect on their lives in these critical areas (Meneses, 2007).

Rootlessness versus restlessness

For many TCKs, the questions *Where is home?* or *Where are you from?* are most difficult to answer. In some ways, they feel connected to each place they've lived, but in other ways they may never feel a total sense of belonging to any of them.

Many have 'itchy feet syndrome'. Since they have moved so frequently, they see a transitional lifestyle as the norm. Rather than facing conflicts or struggles, the instinct is to deal with any stress by moving and starting over. This can affect their academic life significantly if this syndrome kicks into high gear during university years. Too often, they will keep changing schools or majors and have

a hard time finishing. In the end many find their roots in relationships rather than a particular geographical site.

Adaptability versus sense of cultural imbalance

The fact that they regularly change cultural worlds means TCKs must learn to stop and pay attention to how people interrelate in this new environment. This often leads to a sharper capacity to notice the nuances or behind the scenes messages occurring. If TCKs see puzzling behavior, many want to understand it rather than dismiss it. Like chameleons changing color to match their environment, many TCKs learn to change their behavior to blend in with the new surrounding cultures. The TCKs carry that basic sense of adaptability into adulthood where it translates as a strong sense of confidence and the ability to think outside the box or to start something new.

Sometimes, however, such chronic adaptation can leave TCKs with the sense that they are always just a bit off balance culturally and never quite at ease because they don't know for sure how the game is played here. In some ways they can fit everywhere. In other ways, they feel they never quite fit anywhere.

Multifaceted identity versus identity crisis

Ultimately, many TCKs become comfortable with the idea that there are many facets to their cultural identity. As one ATCK said, "My life is like windows [on a computer screen]. I know that all the [cultural] windows are open, but I have to operate in the one that is on the screen at the moment" (Pollock and Van Reken, 2009).

On the other hand, because they have grown up among many cultural worlds TCKs face particular challenges to developing a clear sense of their personal or cultural identity.

Although biracial or bicultural children also often face very real challenges to their identity as they try to sort how to which race or culture do they belong, TCKs often face an even greater challenge (than biracial or bicultural children). Their parents may be from one culture, they may live in a second one, and their peers or school groups may belong to a third one. They are connected to two or more cultures, without a real sense of roots (Meneses, 2007).

This search for identity congruence (Schaetti, 2000) is one of the biggest challenges TCKs face as they seek to answer these fundamental questions: Which of my many selves am I? Is it all right to be all of the above in a world which often asks me to choose one of the above?

Resilience versus unresolved grief

The TCKs, and often those closest to them, move back and forth between home and host cultures with great regularity (White, 1983). Some international schools report that a 30% turnover in their student population each year is normal (Jehle-Caitcheon, 2009). Each move precipitates a normal transition cycle. In any transition, there

is loss as well as gain. Rosalea Cameron discusses how these chronic transitions affect the developmental ecology of TCKs (Cameron, 2003). The TCKs who learn how to process transitions in a psychologically healthy way can develop a remarkable resilience in dealing with change throughout their lives (Pollock and Van Reken, 2009). In fact, many come to view future moves positively, often with anticipation and excitement.

Many TCKs, however, do not find healthy ways to process the losses inherent in any transition process. This leaves them with the second major challenge TCKs and ATCKs often face: unresolved grief. There are several reasons for this:

- *Many losses are invisible.* It is not easy to recognize how significant such things as the loss of a sense of cultural balance, or loss of the sights, sounds, and smells of a world can be when TCKs switch nations and cultures overnight with one simple airplane ride.
- *TCKs may have little permission to grieve.* The gifts of their lifestyle are so many, the losses of leaving friends, family, and place are often discounted by themselves and others.
- *There is often little time.* In the fast-paced twenty-first century there is little time to grieve appropriately (Pollock and Van Reken, 2009).

All of these factors can impede the process of appropriately moving through the various stages of grief Dr. Elisabeth Kübler-Ross identified as denial, anger, bargaining, depression, and acceptance (Kübler-Ross, 1997). Some common patterns educators might see as evidence of this unresolved grief are students who refuse to enter into the activities of the new school, excessive demeaning of the current school or surrounding culture, or a withdrawal into their own silent world (Pollock and Van Reken, 2009).

Expanding Our Language

In 1984, sociologist, Ted Ward, predicted that TCKs were the prototype citizens of the future. By that he meant that the experience of growing up among different cultural worlds would one day be the norm rather than the exception (Ward, 1984). That time is nearly here. As more and more people who did not grow up in a traditional TCK experience say they relate to the benefits and challenges of the TCK profile (such as large worldview, speaking several languages, questions of identity, and unresolved grief), it has become useful to expand the traditional paradigm for TCKs (Van Reken and Bethel, 2005). Just as military, corporate, missionary, and diplomatic children are subsets of the TCK model (see **Figure 2**), TCKs can now become a subset of a larger group called cross-cultural kids (CCKs) – children who live in – or meaningfully interact with – two or more cultural environments for a significant period of

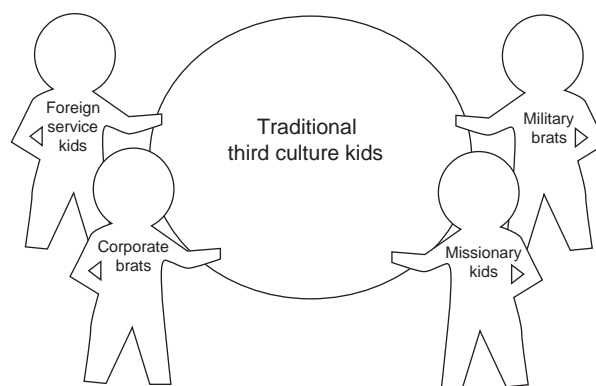


Figure 2 Traditional third culture kids. From David C. Pollock/Ruth E. Van Reken, used by permission.

time during the first 18 years of life (Van Reken, 2006). An adult cross-cultural kid (ACCK) is one who grew up as a CCK (see **Figure 3**).

This group of CCKs includes:

- Traditional TCKs – children who move into another culture with parents due to a parent's career choice.
- Children from bi/multicultural homes – children born to parents from at least two cultures. May or may not be of the same race.
- Children from bi/multiracial homes – children born to parents from at least two races. May or may not be of the same culture.
- Children of immigrants – children whose parents have made a permanent move to a new country where they were not originally citizens.
- Educational CCKs – children who may remain in their home or passport country but are sent to a school (e.g., an international school) with a different cultural base and student mix than the traditional home culture or its schools.
- Children of refugees – children whose parents are living outside their original country or place due to circumstances they did not choose, such as war, violence, famine, or natural disasters.
- Children of borderlanders – children who cross borders frequently, even daily, as they go to school, or whose parents work across national borders.
- Children of minorities – children whose parents are from a racial or ethnic group that is not part of the majority race or ethnicity of the country in which they live.
- International adoptees – children adopted by parents from another country other than the one of that child's birth.
- Domestic TCKs – children whose parents have moved in or among various subcultures within that child's home country.

This new model demonstrates not only the many ways children grow up in culturally diverse situations but also

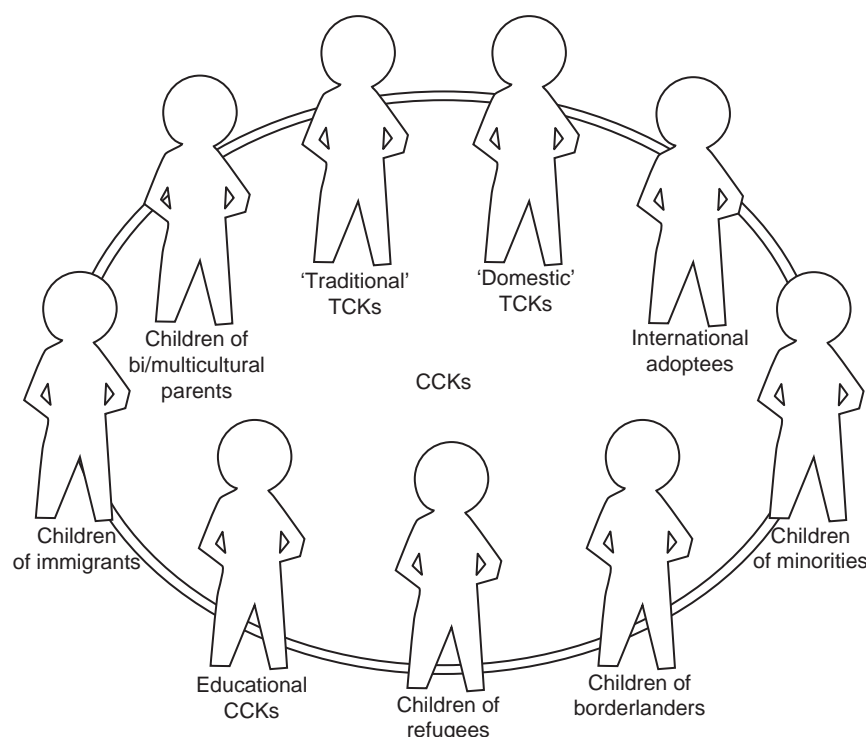


Figure 3 Cross-cultural kid (CCK) model. From David C. Pollock/Ruth E. Van Reken, used by permission.

demonstrates the increasingly complex worlds in which some children live. Many TCKs are now in bicultural or biracial families. Instead of living between their passport country and one host country, countless TCKs have lived in three, four, five, and even more countries. Some TCKs are from minority communities. Others have parents who immigrated to a new land before or during the time they embarked on their international career.

In addition, this new paradigm also provides a way to look at why so many who live in, or have grown up among, many cultural worlds in ways that are different from the traditional TCK model, relate strongly to many common characteristics of the TCK profile. Using the lens of the traditional TCK experience to look at deeper issues that relate to a cross-cultural childhood will not only give educators a deeper understanding of the TCK experience itself, but offer beginning insight into some core factors many other CCKs also share (Van Reken, 2007).

Understanding Hidden Diversity

The concept of culture as an iceberg is a key place to begin looking at the effects of a cross-cultural childhood. Anthropologist, Gary Weaver, explains that just as there are seen and unseen layers in an iceberg, so there are seen and unseen layers in culture. The visible places of culture, such as food, dress, language, and customs are where

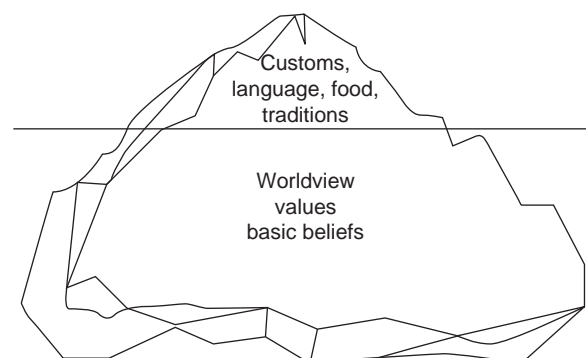


Figure 4 The cultural iceberg. From Weaver, G. (2006). *The American cultural tapestry*, Washington, DC: eJournalUSA, used by permission.

people express the unseen places of who they are (see **Figure 4**). Based on what is seen, people decide when meeting one another if they will relate from likeness or difference.

Weaver warns, however, that in today's globalizing world, the visible layers of culture will begin to look more similar among people of many backgrounds. People will then assume the deeper layers of culture are also alike and the usual allowances given to someone known to be different will be gone. This, he says, will only lead to more cultural clashing and confusion.

Such disparity between the visible and invisible layers of culture is at the heart of many challenges TCKs and

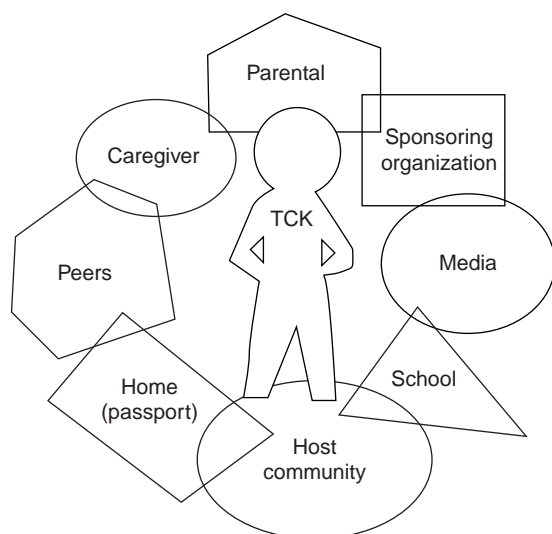


Figure 5 Possible spheres of culture influence in a TCK/CCK's world. Adapted from chart for GNI's. From Norma McCaig (1988), used by permission.

other CCKs face. The visible and invisible aspects of their lives often do not match according to traditional expectations of others. The TCKs and other CCKs, like everyone else, learn cultural practices, values, and worldviews from their environment. The difference between TCKs/CCKs and others from a traditional monocultural background isn't how they learn culture, but the world in which they learn it. **Figure 5** shows some of the markedly different cultural communities with which TCKs/CCKs may interact.

In the predominantly monocultural environments of the past, it was relatively easy for children and others in the community to develop a clear sense of who they were. Children saw themselves reflected in their parents, extended family, teachers, and peers who generally both looked and thought like them and vice versa.

For TCKs and many CCKs, this simply is not so. Because they have grown up interacting closely with many who do not share their national or ethnic culture, they often wind up with a hidden diversity – “a diversity of experience that shapes a person's life and world view but is not readily apparent on the outside, unlike the usual diversity markers such as race, ethnicity, nationality, etc.” (Bethel and Van Reken, 2003: 1). Ultimately, social scientists may have to relook at traditional models of developmental and diversity patterns and see which ones still apply to children in our changing world and where we may need new models (Meneses, 2007; Cameron, 2003).

Not only do TCKs interact with many cultural worlds while living in a particular place, but the high mobility patterns of their lifestyle mean the actual world around them is constantly changing as well. **Figure 6** demonstrates basic patterns for how children raised among many cultures can relate to a surrounding dominant culture (Pollock and Van Reken, 2009). Of course, a single

Foreigner Look <i>different</i> Think <i>different</i>	Hidden immigrant Look <i>alike</i> Think <i>different</i>
Adopted Look <i>different</i> Think <i>alike</i>	Mirror Look <i>alike</i> Think <i>alike</i>

Figure 6 The PolVan cultural identity box. From David C. Pollock/Ruth E. Van Reken, used by permission.

airplane ride can completely shift them from one pattern of relating with the surrounding culture to another.

Figure 6 helps explain some challenges these global shifts of culture and mobility are creating for educators. When students are in the foreigner or mirror box, who they are inside, matches what others expect them to be when looking from the outside. Educators will make allowances for language differences, or possible curriculum discrepancies if students are clear foreigners. For students who are mirrors, educators will assume correctly that these students know the language and cultural practices operative in this school. For both foreigners and mirrors, their identity is clear and life is relatively simple. When, however, students are in the hidden immigrant or adopted box, life can become quite complicated, particularly at school. Educators may assume that a student who is a hidden immigrant knows the historical or cultural references in the curriculum and not realize this child has no idea of the information others expect them to know. Conversely, educators may assume a student in the adopted category will have knowledge about the language or cultural practices of the ethnicity or nationality he or she appears to be when, in fact, that student knows no more about those things than anyone else in the classroom. These are examples of the hidden diversity mentioned above: who others expect them to be is not who they are because they have learned their cultural cues amid various cultural groups. Often these experiences can leave TCKs feeling as if they are always on the margins, or even outside, of the current culture in which they live (Schaetti, 1996).

Typical Responses to Identity Challenges

There are three common ways TCKs and other CCKs often react when they sense that others are assigning them a different cultural identity from how they see themselves. They may choose to be cultural chameleons, screamers, or wallflowers (Pollock and Van Reken, 2009). Cultural chameleons try to make the visible layers of themselves look and act like others in the place, no matter how differently they may feel inside. Screamers may adopt some form of extreme behavior, dress, or attitude to let everyone around know that, despite external appearances, they are

not like their peers inside. Wallflowers choose to withdraw until they know how life operates in the place. Others have used the terms homecomer, cosmopolite, and stranger to refer to similar types of reactions when they occur specifically during repatriation to the passport country (Meneses, 2007). In her studies, Schaetti found the most common reaction during repatriation was that of withdrawal (i.e., becoming a wallflower) because TCKs “expected to find some measure of homecoming and found instead that they were strangers in an unfamiliar land” (Schaetti, 2000).

Summary

The TCKs are students who, during their formative years, lived in one or more cultures which were different from the passport, or home culture(s) of their parents. These students have adopted many feelings, thought patterns, relationships, and traditions of the various culture(s) in which they have lived, but the outcome is more than an amalgamation of these different worlds. The very experience of growing up among an assortment of cultural worlds leads to well-documented shared characteristics with others of similar experience. The gifts of this experience are many, potentially developing skills of the capacity to be effective cultural bridges and negotiators in our globalizing world. The challenges are also real, particularly issues related to identity and unresolved grief. These students may feel like foreigners when in their passport culture even though they look like the majority of their classmates or they may feel at home when they don't look like most of their classmates. Because of varying degrees of looking and thinking alike and differently, TCKs may be misunderstood or even ridiculed by other students for things they do not understand in the present culture. Most TCKs struggle with answering questions like “Where are you from?” because of the connectedness they feel to each culture that others would see as foreign. The mobility of their lifestyles lends itself to feelings of unresolved grief which can be expressed in ways. Lessons learned from the traditional TCK model can be useful in helping educators understand other types of CCKs and the growing personal complexity of cultural mixing of many students.

Bibliography

- Bethel, P. and Van Reken, R. (2003). Third culture kids and curriculum issues in the international school system: Recognizing (and dealing effectively with) the hidden diversity of third culture kids (TCKs) in the classroom. Presented at *A Conversation on Educational Achievements Globally, Comparative and International Education Society Annual Conference*, 12–16 March, New Orleans, Louisiana.
- Cameron, R. J. (2003). *The Ecology of “Third Culture Kids”: The Experiences of Australasian Adults*. EdD Dissertation, Perth, Western

- Australia: Murdoch University. <http://wwwlib.murdoch.edu.au/adt/browse/view/adt-MU20041014.111617> (accessed September 2009).
- Hiebert, P. G. (1983). *Cultural Anthropology*, 2nd edn, pp 28–29. Grand Rapids, MI: Baker Book House.
- Jehle-Caitcheon, N. (2009). *International Schools*. <http://www.caitcheon.com/schools/school1.htm#is>.
- Kübler-Ross, E. (1997). *On Death and Dying*. New York: Touchstone Books.
- McCaig, N. (1987). Brochure for Global Nomads, International.
- Meneses, L. (2007). *Homesick for Abroad: A Phenomenological Study of Third Culture Identity, Language, and Memory*. United States – District of Columbia: The George Washington University.
- Podolsky, M. (1994). Kaigai-shijo socialization in Toronto. *IJJS* 3, 11.
- Pollock, D. (1989). “The TCK profile.” Seminar brochure for Interaction, Inc., p 1.
- Pollock, D. and Van Reken, R. (2009). *Third Culture Kids: Growing Up Among Worlds*. London: Nicholas Brealey.
- Schaetti, B. (1996). *Phoenix rising: A question of cultural identity. Perspectives*. Washington, DC: Global Nomads International.
- Schaetti, B. (2000). *Global Nomad Identity: Hypothesizing a Developmental Model*. Ann Arbor, MI: Bell and Howell.
- Useem, R. (1973). Third cultural factors in educational change. In Brembeck, C. and Hiler, W. (eds.) *Cultural Factors in School Learning*, pp 122. Lexington, MA: D.C. Heath.
- Useem, R. and Cottrell, A. (1993). Third culture kids: Focus of major study. *Newslinks. Newspaper of the International School Services* 12(1).
- Van Reken, R. (1989). Caring for the adult MK. In Echerd, P. and Arathoon, A. (eds.) *Understanding and Nurturing the Missionary Family*. Pasadena, CA: William Carey Library.
- Van Reken, R. (2006). The internationally mobile child: Understanding the experience of growing up among worlds. Presented at *Council of British Independent Schools in the European Communities Conference*, 6–8 May, 2006, London, England.
- Van Reken, R. (2007). The internationally mobile child: Why a cross-cultural childhood matters. Presented at *European Council of International Schools Annual Conference*, 22–25 November, 2007, Madrid, Spain.
- Van Reken, R. and Bethel, P. (2005). Third culture kids: Prototypes for understanding other cross-cultural kids. *Intercultural Management Quarterly*. Washington, DC: Intercultural Management Institute, American University. www.imquarterly.org (accessed September 2009).
- Van Reken, R. and Rushmore, S. (2009). Thinking globally when teaching locally. *Kappa Delta Pi Record* 45(2).
- Vidal, X. (2000). Third culture kids. A binding term for a boundless identity. Senior Essay, Professor Clarke.
- Ward, T. (1984). Plenary talk at *International Conference on Missionary Kids*. Philippines: Manila.
- Weaver, G. (2006). The American cultural tapestry, Washington, DC: eJournalUSA. <http://www.gmfus.org/doc/mmf/The%20American%20Cultural%20Tapestry.pdf>.
- White, F. (1983). Some reflections on the separation phenomenon idiosyncratic to the experience of missionaries and their children. *Journal of Psychology and Theology* 11(3), 181–188.

Further Reading

- Copeland, A. and Bennett, G. (2008). *Understanding American schools*. Boston, MA: Interchange Institute.
- Eakin, K. (1999). *According to My Passport I'm Coming Home*. Washington, DC: Family Liaison Office.
- Eidse, F. and Sichel, N. (2004). *Unrooted Childhoods: Memoirs of Growing up Global*. London: Nicholas Brealey Publishing/Intercultural Press.
- McCluskey, K. (ed.) (1994). *Notes from a Traveling Childhood: Readings for Internationally Mobile Parents & Children*. Washington, DC: Foreign Service Youth Foundation.
- Park, E. and Rumrell-Teese, K. (2005). *Here Today, There Tomorrow: A Training Manual for Working with Internationally Mobile Youth*. Washington, DC: Foreign Service Youth Foundation.

- Pascoe, R. (2006). *Raising Global Nomads: Parenting Abroad in an On-Demand World*. Vancouver, BC: Expatriate Press.
- Taber, S. (1997). *Of Many Lands: Journal of a Traveling Childhood*. Washington, DC: Foreign Service Youth Foundation.
- Van Reken, R. (2005). Third culture kids: Prototypes for understanding other cross-cultural kids: Part 1 – recognizing a hidden diversity. *Intercultural Management Quarterly* 9, 3–4.
- Van Reken, R. and Rushmore, S. (2009). Thinking globally when teaching locally. *Kappa Delta Pi Record*.

Relevant Websites

- <http://www.TCKid.com> – A Home for Third Culture Kids.
- <http://www.expatexpert.com> – Expat Expert.
- <http://www.tckworld.com> – TCK World: The Official Home of Third Culture Kids.
- <http://www.interchangeinstitute.org> – The Interchange Institute.

INTERNATIONAL COMPARISONS

Contents

IEA Studies in ICT

IEA Studies in Mathematics and Science

IEA Study in Civic Education

IEA: Globalization and Assessment

IEA Studies in ICT

N Law, University of Hong Kong, Hong Kong, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

Glossary

ICT – The abbreviation for information and communication technology. This term is often used interchangeably with information technology (IT) in the literature on technology in education.

IEA – The abbreviation for International Association for the Evaluation of Educational Achievement.

IT – The abbreviation for information technology. Also used interchangeably in the literature on technology in education with ICT and refers both to computer hardware and software, as well as networking infrastructure to enable the user to connect to the Internet via the computer hardware.

The design of International Association for the Evaluation of Educational Achievement (IEA) studies generally involves collecting data from different levels: education system, school, teacher, and student levels. Until 2008, IEA had completed two comparative studies of information and communication technology (ICT) in education – the first study, Computers in Education Study (CompEd), conducted between 1987 and 1993 and the second study, Second Information Technology in Education Study (SITES), conducted between 1997 and 2008. In addition to these two major studies specifically devoted to ICT in education, IEA has also gathered information about access and use of computers and the Internet through background questionnaires in some of its other studies such as Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS). The rationale for ICT in education worldwide has seen changes over the 20 years during which the CompEd and SITES studies were conducted, and hence the specific research questions and designs for

these two sets of studies also differ. This article provides an overview of the research design and key findings from these two ICT-specific studies to shed light on the status and role of computers in school education, the changes that have taken place over the two decades since the late 1980s, and the factors that have influenced ICT use in schools in different countries around the world.

The First Computers in Education Study (1987–93)

The CompEd study was designed as a two-stage investigation. The first stage, with data collection in 1989, aimed to provide an overall picture about how computers were used, the extent and availability of computers in schools, the nature of instruction about computers, and estimates of the effects that computers are having on students, the curriculum, and the school as an institution. Stage 2 with data collection in 1992 comprised two parts: a repetition of the Stage 1 survey to examine the rate of development of computers in education over time, and a survey of students to find out the level of access to computers, ways in which computers were used in schools and outside of schools, the extent of students' information technology (IT) competencies, and their attitudes toward and perceptions of computers.

The Stage 1 Study

The Stage 1 study results are reported in Pelgrum and Plomp (1991, 1993). The survey was conducted on a random sample of schools and teachers at primary, lower-secondary, and upper-secondary levels. Altogether 21 systems participated, 15 of which were from Europe, two from North America, three from Asia, and New Zealand was the only participating country from outside these three continents.

Schools were sampled with probabilities of selection proportional to the size of the school. Surveys were distributed to principals, technical coordinators (or anyone who was most familiar with the technical aspects of the computer infrastructure in the school), computer-education teachers, and computer-using and non-computer-using teachers in mathematics, science, and the mother tongue.

The Stage 1 survey conducted in 1989 found great diversity across countries in the availability of computers in schools, and access was generally highest in upper-secondary schools and lowest in primary schools. In 1989, in a few countries such as Canada (British Columbia), France, and the US, the access was almost 100% throughout all three levels of education, while the access was much lower even in some highly developed countries such as Japan. An examination of the history of computer use in countries found that typically a period of about 4 years was needed to equip all schools with computers and a longer history was associated with more favorable student-computer ratios. Computers were located in special computer rooms for a majority of computer-using lower- and upper-secondary schools in almost all countries. In computer-using elementary schools, there was a comparatively higher probability of computers used in classrooms compared to secondary schools, though computers were still more likely to be located in special computer rooms.

The dominant use of computers in the curriculum differed between primary and secondary schools. In secondary-school education, computers were used mainly for teaching students how to use computers as a separate subject, and the two subject areas with the highest frequency of computer use were mathematics and commercial subjects. In primary schools, computers were most frequently used for drill and practice. Different locations of computers tend to favor different kinds of computer use. Availability of computers in classrooms was more closely associated with the integration of computers into the teaching and learning of school subjects rather than for computer literacy.

In terms of the responsibility for introduction of computers in schools, school authorities were the most mentioned group. The most prominent reason given by principals in all three populations for the use of computers in the school curriculum was to develop students' computer literacy. Where improving the teaching and learning process was given as a reason for using computers, there was also more reported use of computers in the teaching of school subjects. The lack of hardware and software were identified as the two main obstacles to computer integration in the teaching of school subjects. It was also found that the integration of computers into subject teaching was most consistently associated with the knowledge and skills of teachers. The amount of training teachers received and the topics covered were found to have positive influence on teachers' attitudes toward the

educational impact of computers and the likelihood of computer integration in their teaching. Further, the use of computers appeared to be male dominated and, in most countries, a majority of schools did not have special policies to promote gender equity.

The Stage 2 Study

The Stage 2 survey was conducted in 1992. In addition to a repeated administration of the Stage 1 survey to gauge the rate of development of computers in education over the intervening years, a survey of students sampled from three populations was conducted: students in grades in which the modal age was 10 and 13 years, respectively, and students in the next to final year of secondary education. Only 12 systems participated in this second stage of the CompEd study.

Results of this study are reported in Pelgrum and Plomp (1996) and Pelgrum *et al.* (1993). It was found that in most countries, there was considerable improvement in access to hardware and software over the 3 years since the Stage 1 survey. However, only very small numbers of students were using computers regularly and still mainly for learning about computers. Hence, the educational impact of increased computer access was not evident. A Functional Information Technology Test (FITT) was developed to assess students' ability to function effectively with respect to practical information-related tasks. In most countries, there was a large achievement gap between the mean scores for the lowest 25% and highest 25% of students surveyed. Though the reported opportunity for students to learn the assessed knowledge and skills was extremely low, the students' results were well above chance level, indicating that many students did possess significant functional knowledge and skills about computers. Further, the use of computers in a school did not appear to add much value to students' FITT results, indicating that students' functional knowledge about computers was probably learnt outside of schools. Exposure to computers was found to be positively correlated with students' level of knowledge and their attitudes toward computers.

In all countries, boys scored higher than girls, and the difference is highest in upper-secondary education, though the difference was not significant in some countries. Girls also tend to enjoy using computers less than boys. However, it was rare for schools to have gender-related policies related to computer use. A large number of students agreed with unethical practices related to computer use, such as illegal copying of software.

The most available types of professional development at the school level were introductory courses on computing and application programs. Technical coordinators were very often regular teachers and they did not have much time available to help with the training of other teachers.

The Second Information Technology in Education Study

The motivation behind the conceptualization of the Second Information Technology in Education Study (SITES) was rather different from that of the CompEd study, although the time gap between data collection in CompEd Stage 2 and SITES Module 1 was only 6 years. SITES was not interested in studying the teaching of computer-related subjects, but geared rather to address the increasing policy focus on ICT use in the curriculum as an integral part of preparing students for adequate and effective functioning in the information society. The assumptions of the SITES study are that education has to prepare and support citizens for lifelong learning and that ICT can play an important role in this process of change (Anderson *et al.*, 1997; IEA, 1999).

SITES was initially designed as a research program consisting of three modules. Module 1 (SITES M1), was a survey of a random sample of schools conducted in 1998 to provide a snapshot picture of the situation regarding ICT in education, mainly with respect to access (hardware, software, and Internet availability), pedagogical vision and practices according to the principals, as well as the availability of professional development and support for teachers and students for the use of ICT in teaching and learning. Module 2 (SITES M2) was an international comparative study of case studies of innovative pedagogical practices using technology, with data collection conducted in 2000–01 to gain insight into whether and how ICT use may contribute to changes in teaching and learning that will achieve the new education goals described in many policy documents around the world regarding the preparation of twenty-first-century citizens. Module 3 (SITES M3) was initially intended as a third module of the study that will build upon findings from the leading-edge innovative classrooms of Module 2 and utilize the emerging pedagogical practices as the central themes for a teacher survey and a student assessment. The specific goal of this module was to measure the ICT-supported knowledge-management competencies of students, including their abilities to retrieve, organize, critically evaluate, communicate, and produce knowledge (Kozma, 2002). Unfortunately, the IEA General Assembly, which makes all key decisions on IEA studies, found the methodological issues and hurdles that need to be overcome for the study to meet all the stringent requirements for a rigorous international comparative study to be too challenging and rejected the proposal. A further proposal to conduct a survey of schools and teachers to study pedagogy and ICT use in schools was subsequently accepted for data collection in 2006. This study focused on how teaching and learning took place in mathematics and science classrooms and the role of ICT in them, and is referred to as the SITES 2006 study. Each of the three completed SITES study components are described briefly below.

SITES M1

At the time the SITES M1 study was conceptualized, many countries had announced system-wide education-reform efforts involving changes in both the goals and pedagogies of school education that aimed to prepare citizens for the knowledge society. Hence a major question that this study addressed was the extent to which such changes have actually taken place and the role of ICT in the school curriculum. The study also aimed to find out which ICT infrastructure was available in schools, what staff development and support services exist with regard to ICT use in schools, and the roles played by the school leadership. This survey study was administered between November 1998 and February 1999 to principals and technical coordinators in randomly sampled schools at three levels of the education system: primary, lower secondary, and upper secondary. Altogether 26 countries participated in the study, mostly from Europe and Asia. Only 12 countries took part in the survey at all three education levels, and the numbers of countries participating at the primary, lower-secondary, and upper-secondary levels were 14, 24, and 21, respectively. The key results of this study are published in Pelgrum and Anderson (1999).

The SITES M1 results found, from responses of the national research coordinators in participating countries, that governments generally acknowledge the need for education reforms. In most countries, there were programs that aimed to improve the ICT infrastructure in schools, though the financial commitments differed widely across countries. In order to find out whether the kinds of teaching and learning practices promoted in many policy documents on education reform were actually taking place in schools, the study constructed indicators for emerging and traditionally important pedagogical practices based on principals' responses on the extent to which different kinds of learning and teaching activities were present in their schools.

The survey findings indicate considerable variation in the presence of emerging pedagogical practices across countries, as well as between schools within the same country. While it was not clear whether schools with a higher emerging pedagogical practice indicator score had a stronger reform orientation, it was found that these schools tended to have more favorable student–computer ratios. There appeared to be a correlation between the use of IT, and the school and classroom cultures of a country (Law *et al.*, 1999). In Hong Kong, Belgium, the Czech Republic, France, Israel, Italy, Japan, Singapore, South Africa, Thailand, and Slovenia, there was a strong policy goal for teachers to use computers for instructional purposes. On the other hand, in Canada, the Scandinavian countries, and most of the European countries participating in the study, the policy focus was more on enabling students to use computers as supportive learning aids to

cultivate in them the ability to process and analyze information, use e-mail for communication, and gain access to external databases via the Internet.

In terms of ICT infrastructure, the computer–student ratio, as an indicator of access opportunity, has improved considerably compared to the same statistics collected 3.5 years earlier in the TIMSS-1995 study. However, there were very large variations within and across countries. Even when the comparison is made across computer-using schools, the mean computer–student ratio in schools at the lower-secondary level in the participating countries varied from 1:7 to 1:216. The kinds of computers available are also important as these have great implications on the kinds of activities that can be conducted. In some countries, almost all equipment was suited for multimedia purposes while in other countries, particularly in East European countries, substantial numbers of early generation (8-bit) computers were in use, which cannot practically be used to support activities beyond learning the basics of programming. There were also considerable differences in schools' access to the Internet and World Wide Web for teaching and learning. In a few countries such as Singapore, Iceland, and Canada, almost 100% of schools had Internet access for instructional purposes, while the statistic was 10% or less in some others, and was almost nil at the primary level in a few countries.

Staff development is generally identified as an important factor influencing ICT use in classrooms, particularly for integration in everyday educational practice to support learning and teaching. Teachers' lack of ICT-related knowledge was seen as a major obstacle to realizing the ICT-related school objectives by a large number of principals in many of the participating countries and is the second most-reported obstacle, next to the lack of computer hardware. It was also found that the majority of schools had a policy goal that all teachers should receive training for the instructional use of ICT, but the achievement of this goal and the availability of courses reported were both low. Furthermore, the technical coordinators' self-rated ICT competence was much higher for general applications than for instructional uses. In terms of school leadership, great variation was found across countries in terms of the school principals' attitudes toward the contribution of ICT to lifelong learning.

SITES M2

Altogether 174 case studies of technology-supported pedagogical innovations were collected in 2000–01 and reported by research teams in the 28 countries/regions participating in the SITES M2 study. A basic assumption of the study was that new pedagogical practices were emerging in schools and that in-depth comparative case studies would provide a better understanding of what kinds of pedagogical innovations have developed around the world where technology

plays a substantial role, and what kinds of school factors contribute to the emergence and sustainability of these innovations. The study aims to inform national and local policymaking, contribute to improved classroom practice, and to advance knowledge and build theory about the contexts and factors within and across countries that contribute to the successful and sustained use of innovative technology-based pedagogical practices through providing rich case descriptions. The research design and key findings are reported in Kozma (2003).

Cases in SITES M2 may be selected from one or more of the elementary, lower-secondary, and upper-secondary levels based on a set of selection criteria and process established by the International Coordinating Consortium. Each participating country followed a guideline to establish an expert panel comprising a range of stakeholders that might include policymakers, teachers, administrators, technology experts, and researchers to provide a local definition for innovation and to review and select the cases for study. National research teams were responsible for conducting the case studies based on the study guidelines, procedures, and instruments and for writing an approximately 5000-words report on each case. Each case report comprises a summary, a description of the school background and culture, the history of the innovation, the technological infrastructure available in the school, the national and regional policies that affected the innovation, as well as details of the innovation in terms of the curriculum and assessment goals, and the teachers' and students' practices and outcomes, according to a common template.

It is noteworthy that there was great similarity in the criteria for innovation decided by the different national expert panels. The case studies collected in this study reflect a very different picture than the traditional classroom where the teacher lectures in front of the classroom and students take notes or do worksheets, indicating that technology was supporting significant changes in classroom teaching and learning in these cases. Students were often involved in searching for information, designing products, and publishing or presenting the results of their work; collaboration among students (within and outside of the same classroom) was common. The most frequently used technology were productivity tools such as word processors and presentation software, as well as World Wide Web resources, e-mail, and multimedia software. Many of the case studies were ICT-based innovations involving multidisciplinary projects. Innovations pertaining to a single subject area were found in only 29% of the cases, and only a small minority of the cases involved the study of computer literacy, computer science, or informatics as a subject area.

Law (2003) reports on the use of a six-dimensional framework for comparing innovativeness of the case studies: (1) the curriculum goals, (2) the roles played by the teacher, (3) the roles of the learner, (4) types of products indicative of learning outcomes, (5) the complexity and sophistication of

the ICT used, as well as (6) connectedness – the extent to which the classrooms are connected to the outside world through external participants such as teachers and students from other classes or even other countries, parents, alumni, and experts. While the six dimensions were interrelated, the teacher's role was found to have the strongest impact on the students' roles, and the kinds of learning outcomes observed. Examination of the case studies reveals large diversities across these cases, and cases with highly innovative characteristics on all the 6 dimensions were rare. It was found that cases from Asian countries were less connected compared to cases collected from European countries (Law, 2004).

Despite the innovativeness of the case studies collected, only 18% reported a change in curriculum goals or content that was supported by technology, and only 41% provided evidence that the innovation had been disseminated to other classrooms or schools (Kozma, 2003). The energy and commitment of teachers, student support, the perceived value for the innovation, the availability of teacher-professional-development opportunities, and administrator support were important school factors contributing to the sustainability of the innovations. Support from external sources, innovation champions, funding, and connection with national technology plans that provided resources often enabled the innovations to succeed.

SITES 2006

SITES 2006 built on the previous two SITES studies within a global context of increasing policy interest in the use of ICT in schools to help students develop twenty-first-century skills, such as the ability to engage in collaborative knowledge creation and problem solving with peers and experts around the world. It was a comparative study of ICT use within the context of the overall pedagogical practice of the teacher in mathematics and science classrooms. The contextual factors examined in this study also include those found pertinent to supporting pedagogical change and innovation. In 2006, the study collected information from principals and technology coordinators from roughly 9000 schools and over 35 000 grade 8 mathematics and science teachers in 22 education systems. The national research coordinator from each country completed a questionnaire related to system-level conditions in relation to education and ICT use in schools. The research design and findings of this study are reported in Law *et al.* (2008).

Twenty out of the 22 systems reported having a system-wide ICT in education policy. In nearly all of the participating systems there was some level of government funding for the provision of hardware and software in schools and at least slightly increased ICT spending during the previous 5 years. Across the 15 systems that participated in both SITES M1 and SITES 2006, great

improvements in access to computers and the Internet since 1998 were observed. With the exception of one country, all participating systems reported all or almost all schools having access to computers, although there are wide diversities in terms of ICT infrastructure available in schools. There are also wide variations in the extents to which technical support and pedagogical support were available, as well as in the relative priorities given to these two kinds of support.

Many educational systems do not have active, centralized policies on teachers' professional development related to ICT use or new pedagogies, and only a very small minority of schools required teachers to be trained in a variety of areas dealing with new pedagogy and ICT. In a few systems, re-allocation of workload to allow for collaborative planning was reported in 80% or more of the schools, but the percentage was much lower in other systems.

Less than half of the mathematics teachers and science teachers in 12 and 4 systems, respectively, reported having used ICT with their target class. Generally, ICT use was more prevalent among science teachers compared to mathematics teachers in most systems. On the other hand, the extent of ICT adoption by teachers differs enormously across systems, varying from below 20% to over 80%. Further, no clear relationship was found between the mean student-computer ratio of a country and the mean percentage of its mathematics teachers or science teachers reporting instructional use of computers.

The teacher survey constructed indicators for traditionally important and twenty-first-century-oriented pedagogical activities. The study found that pedagogical practices of the mathematics and science teachers surveyed were still mainly traditional. In most systems, there was a stronger twenty-first-century orientation in the ICT-using teaching and learning practices compared to pedagogical practices in general. However, ICT adoption *per se* does not determine pedagogical orientation as evidenced by the observation in a few systems that ICT-using practices exhibited a stronger traditional orientation. While the teachers' perceptions of the impacts of ICT use on themselves and on their students were generally positive, the perceived learning gains were not necessarily associated with higher levels of reported ICT use. Increased impacts on students were reported in the area of twenty-first-century abilities, such as inquiry and collaboration skills, and the ability to work autonomously at the student's own pace, though the perceived extents of such impacts vary greatly across systems. Further, the impact of ICT use on students was found to be highly dependent on the teachers' pedagogical orientation adopted in ICT use – lifelong-learning-oriented pedagogical uses of ICT were positively correlated with the perceived gains in students' learning outcomes, the correlations being highest for inquiry and collaboration skills and the lowest for traditional outcomes. No significant correlation was found for

traditionally oriented uses of ICT and perceived students' learning outcomes.

While the CompEd study found a much higher proportion of male teachers using computers compared to female teachers, the SITES 2006 study showed a very different picture. In a few countries, particular gender or age groups of teachers were found to be more likely to report having used ICT in their teaching. However, there are no general gender or age-related trends with respect to ICT use across all participating countries. Hence it can be concluded that age and gender effects, where they exist, are most likely due to age- and gender-related contextual factors rather than the reflection of an inherent relationship between these teacher characteristics and ICT use.

Summary

The CompEd and SITES studies conducted under the auspices of the IEA have provided important comparative data on the changes that have taken place over two decades since the late 1980s in terms of the status and role of computers in education. These studies clearly show that the policy focus of ICT in education has moved from computer/information literacy to leveraging ICT use to realize the education-reform goals of preparing learners for the twenty-first century. Great improvements in computer and Internet access in schools have been observed, but the actual adoption of ICT in classroom practices by teachers in 2006 was still relatively low in many countries. Furthermore, general pedagogical practices were still largely traditional and ICT use was not necessarily associated with more twenty-first-century-oriented practices.

While the IEA studies have contributed much to our understanding of ICT use in schools, there is still a lot of scope for researchers to conduct further secondary analysis on the datasets of these studies accessible from the IEA website.

See also: IEA: Globalization and Assessment.

Bibliography

- Anderson, R., Haider, G., Pelgrum, W., and Wanatabe, R. (1997). *Second Information Technology in Education Study: Revised Proposal to the IEA General Assembly 1997*. Amsterdam: IEA.
- IEA (1999). *SITES M1 Press Release*.
- Kozma, R. (2002). *Second Information Technology in Education Study: Module 3 (SITES M3) Prospectus*. Amsterdam: IEA.
- Kozma, R. (ed.) (2003). *Technology, Innovation, and Educational Change: A Global Perspective*, p 301. Eugene, OR: ISTE.
- Law, N. (2003). Innovative classroom practices and the teacher of the future. In Dowling, C. and Lai, K. W. (eds.) *Information and Communication Technology and the Teacher of the Future*, pp 171–182. Dordrecht: Kluwer.
- Law, N. (2004). Teachers and teaching innovations in a connected world. In Brown, A. and Davis, V. (eds.) *Digital Technology,*

- Communities and Education*, pp 145–163. London: Routledge Falmer.
- Law, N., Pelgrum, W. J., and Plomp, T. (eds.) (2008). *Pedagogy and ICT in Schools Around the World: Findings from the SITES 2006 Study*, p 296. Hong Kong: CERC/Springer.
- Law, N., Yuen, H. K., Ki, W. W., Li, S. C., and Lee, Y. (1999). *Second International Information Technology in Education Study: Hong Kong SAR Report*, p 147. Hong Kong: CITE, University of Hong Kong.
- Pelgrum, W. J. and Anderson, R. E. (eds.) (1999). *ICT and the Emerging Paradigm for Life Long Learning*, p 361. Amsterdam: IEA.
- Pelgrum, W. J. and Plomp, T. (1991). *The Use of Computers in Education Worldwide: Results from the IEA "Computers in Education" Survey in 19 Educational Systems*, 1st edn., p 179. New York: Pergamon Press.
- Pelgrum, W. J. and Plomp, T. (1993). *The IEA Study of Computers in Education: Implementation of an Innovation in 21 Education Systems*, 1st edn., p 332. New York: Published for the International Association for the Evaluation of Educational Achievement by Pergamon Press.
- Pelgrum, W. and Plomp, T. (1996). Information technology and children from a global perspective. In Collis, B. A., Knezek, G. A., Lai, K. W., et al. (eds.) *Children and Computers in School*, pp 23–42. Mahwah, NJ: Erlbaum.
- Pelgrum, W. J., Reinen, J. I., and Plomp, T. (1993). *Schools, Teachers, Students and Computers: A Cross-National Perspective*, p 239. The Hague: International Association for the Evaluation of Educational Achievement (IEA).

Further Reading

- Law, N. (2007). Comparing pedagogical innovations. In Bray, M., Adamson, B., and Mason, M. (eds.) *Comparative Education Research: Approaches and Methods*, pp 315–337. Hong Kong: CERC and Springer.
- Law, N., Lee, Y., and Chow, A. (2002). Practice characteristics that lead to "21st century learning outcomes" *Journal of Computer Assisted Learning* 18(4), 415–426.
- Law, N., Yuen, H. K., Ki, W. W., et al. (eds.) (2000). *Changing Classrooms and Changing Schools: A Study of Good Practices in Using ICT in Hong Kong Schools*. Hong Kong: CITE, University of Hong Kong.
- Mioduser, D., Nachimias, R., Tubin, D., and Forkosh-Baruch, A. (2003). Analysis schema for the study of domains and levels of pedagogical innovation in schools using ICT. *Education and Information Technologies* 8(1), 23–36.
- Nachimias, R., Mioduser, D., Cohen, A., Tubin, D., and Forkosh-Baruch, A. (2004). Factors involved in the implementation of pedagogical innovations using technology. *Education and Information Technologies* 9(3), 291–308.
- Plomp, T., Anderson, R. E., and Kontogiannopoulou-Polydorides, G. (1996). *Cross National Policies and Practices on Computers in Education*. Dordrecht: Kluwer.
- Plomp, T., Anderson, R. E., Law, N., and Quale, A. (eds.) (2003). *Cross-National Information and Communication Technology Policies and Practices in Education*. Greenwich, CT: Information Age.
- Plomp, T., Pelgrum, W. J., and Law, N. (2007). SITES2006–International comparative survey of pedagogical practices and ICT in education. *Education and Information Technologies* 12(2), 83–92.
- Yuen, H. K., Law, N., and Wong, K. C. (2003). ICT implementation and school leadership: Case studies of ICT integration in teaching and learning. *Journal of Educational Administration* 41(2), 158–170.

Relevant Websites

- <http://www.sitesdatabase.cite.hku.hk> – Hong Kong SITES M2 website.
- <http://www.mscp.edte.utwente.nl> – IEA SITES M1 website.
- <http://www.sitesm2.org> – IEA SITES M2 website.
- <http://www.sites2006.net> – IEA SITES 2006 website.
- <http://www.iea.nl> – IEA The Computers in Education Study website.

IEA Studies in Mathematics and Science

F K S Leung

© 2010 Elsevier Ltd. All rights reserved.

Introduction

International Association for the Evaluation of Educational Achievement (IEA) is an independent, international organization with members comprising national research institutions and governmental research agencies from nearly 70 countries. It was founded in 1959 to coordinate international studies among countries to determine the level of students' achievement within an educational system. IEA conducts cross-national studies of educational achievement in different subject areas. Other than mathematics and science, areas of study include reading literacy (Progress in International Reading Literacy Study or PIRLS), civic education (CIVED and ICCS), information technology in education (SITES-M1, SITES-M2, SITES 2006), and pre-primary education (PPP). In the area of mathematics and science, four studies were conducted before the 1990s, namely, the First International Mathematics Study (FIMS), the First International Science Study (FISS), the Second International Mathematics Study (SIMS), and the Second International Science Study (SISS). In 1995, the Third International Mathematics and Science Study (TIMSS) tested both mathematics and science, and the study was repeated in 1999 (TIMSS-R). From then on, it was decided that the study be conducted in 4-year cycles, and from the 2003 cycle, the study has been renamed Trends in International Mathematics and Science Study, while keeping the acronym TIMSS. The latest testing took place in 2007, and the results were released in December 2008. Planning for the next cycle of the study in 2011 is underway. In addition to this regular cycle of studies in mathematics and science achievement, IEA conducted a video study of mathematics classroom practices in 1995 (Stigler *et al.*, 1999), and another one involving mathematics and science classrooms in 1999 (Heibert *et al.*, 2003), and is conducting a study on teacher education and development in mathematics (TEDS-M).

Early IEA Studies of Mathematics and Science

The first IEA study was the so-called 12-country survey for which mathematics and science were two of the five areas under study. The study, conducted between 1959 and 1962, was in essence a piloting one, testing the feasibility of conducting large-scale cross-national studies. Although difficulties were identified, the survey concluded that conducting such large-scale studies was viable. On that

basis, IEA launched its first formal study in early 1960s, the First International Mathematics Study (FIMS), and 12 countries participated in the study.

The First International Mathematics Study

FIMS was more than a study of the mathematics curriculum. It was meant to be a comparative study of the outcomes in different education systems around the world using mathematics achievement as the indicator for output (McLean *et al.*, 1986). The main objective of the study was "to investigate the 'outcomes' of various school systems by relating as many as possible of the relevant input variables" (Husén, 1967a: 30). Mathematics was chosen because of the general interest in the mathematics curriculum around the early 1960s and because mathematics was considered a universal language, relying on symbols that were more or less international and hence easier for comparison.

FIMS was a comprehensive study, collecting information on the educational systems, school structures, teacher attitudes and classroom practices, as well as student attitudes toward and achievement in mathematics. Achievement was measured by tests that were specially designed for the purpose through a thorough procedure that involved national centers comprising mathematicians and mathematics educators. Other information was collected through questionnaires, one for students, one for teachers, one for schools (school heads), and one for nations (completed by a representative in that nation). The population of students studied comprised of 13-year-olds and pre-university students.

The FIMS report was written by Husén (1967a,b) and published in two volumes. For a study of this nature and scale, it is not surprising to find that many of the hypotheses posed at the outset of the study could be neither confirmed nor rejected. In the chapter on 'Summary of major findings' in the second volume of the report, for example, the author admitted that "neither the 'productivity' of an educational system of a country, nor the effect of the instruction given, can be assessed from national means" (Husén, 1967b: 288). The only consistent finding was the great variability in student achievement. The report listed the correlations of achievement with many other variables and discussed problems related to school organization, instruction and curriculum, and societal factors, and some rather obvious conclusions were drawn.

Perhaps the most important contribution of FIMS is that these obvious findings were tested empirically for the first time in the international setting.

The First International Science Study

FISS took place in the late 1960s and early 1970s, and was part of the Six-Subject Survey (science, reading comprehension, literature education, English as a foreign language, French as a foreign language, and civic education). Eighteen countries participated in the study, and the target populations were 10-year-old students, 14-year-old students, and students in the final grade of the secondary school. The study encompassed three domains of science: biology, chemistry, and physics. Similar to FIMS, data were also collected on student and teacher attitudes, methods of teaching, the development of students' practical skills in science, and understanding of the nature of science.

The major findings of the study were that there were high intercorrelations between the subscale scores in biology, chemistry, physics, and practical performance, and so it warranted the calculation and usage of a total score in science achievement. Boys outperformed girls in science, and there was significant improvement in science scores between age levels. Time given to the study of science had significant effects on student achievement, and the variance of science scores explained by school and classroom factors was sizable. New teaching method, such as controlled practical work, did not lead to better performance in science, although secondary school students tended to perform better if their teachers were "specialist science teachers or had received more post-secondary education, had participated in science curriculum reform, had spent more time on preparation, or were members of a subject association."

The SISS report admits that "limitations of time and money imposed serious constraints which prevented the Committee from attaining their goals completely" (Walker, 1976: 78). However, the report still concludes that schooling did matter, although the effect was not as marked as could be expected.

The Second International Mathematics Study

SIMS was conducted in the early 1980s, and 20 countries or systems participated in the study. Like the first study, a large number of variables were studied in addition to mathematics achievement, but the organizers emphasized that, unlike the first study where mathematics was chosen for convenience (Travers, 1980), the second study focused on the teaching and learning of mathematics in schools.

Its purpose was "to compare and contrast, in an international context, the varieties of curricula, instructional practices, and student outcomes (both affective and cognitive) across the schools of twenty countries and educational systems" (Travers and Westbury, 1989: 1).

SIMS conceived the curriculum as being made up of three levels: the intended, the implemented, and the attained levels. The first level was measured by survey questionnaires as well as an international grid completed by educators in different countries. The implemented curriculum was measured by, among other instruments, an opportunity to learn (OTL) index which asked teachers to rate each item in the tests given to their students according to whether that topic had been taught to the students or not. Student achievement was measured by a battery of tests, and attitudes measured by questionnaires, with some of the items identical to those in the first study for comparison purposes.

The major findings of SIMS were that, for the education systems, "there is a fair degree of structural similarity between the various target populations even though there are significant differences in the organization of curricula and courses and in the grade levels being investigated" (Travers and Westbury, 1989: 78). For the intended curriculum, there was a common core of mathematical content for all countries, but there was great variety in geometry content across the systems (p. 110). For the implemented curriculum, teachers' general approaches to teaching were found to be rather similar, namely largely chalk and talk and relying heavily on a prescribed textbook (Robitaille and Garden, 1989: 235). Furthermore, there were substantial overall discrepancies between the intended content coverage and implemented coverage (OTL), so that it was "difficult to make confident claims about the validity of either or both of these indices of coverage" (Travers and Westbury, 1989: 116). For the attained curriculum, "performance on items involving . . . straightforward applications of basic concepts was generally good; however, performance fell off sharply on items calling for the use of higher order thinking skills" (Robitaille and Garden, 1989: 238).

The Second International Science Study

SISS was also conducted in the 1980s, and 23 countries or education systems participated in the study. For each science topic in SISS, three types of activities were tested: knowledge, understanding of a principle, and application of information and principles to solve a practical problem. The science test was complemented by questionnaires for students, teachers, and school principals. In addition, countries participating in the study prepared a National Science Education Case Study Report on teaching science within the country.

The key findings were that there was a general consistency in all countries concerning the content of science taught in the fields of biology, chemistry, and physics, but not in earth science. Science was seen as a practical subject requiring proficiency in the manipulation of apparatus and in experimentation, but emphasis on the applications of science was rather low. At all grade levels, the opportunity to learn provided in the curriculum as well as time given to the study of science were related to the average achievement level of a country, and students perform better when they are taught by teachers who are experienced and competent in science.

Between SISS in the 1970s and SISS in the 1980s, the majority of countries saw a significant improvement in students' science achievement, and science had assumed a more prominent place in the curriculum. At the upper-secondary school level, although the average level of achievement in science dropped in those schools where a higher proportion of students remained at school and continued studying science, the performance of the better students did not decline. In several countries, the difference between boys' and girls' achievement at both middle- and upper-secondary school levels lessened, particularly in the physical sciences.

Third International Mathematics and Science Study – TIMSS 1995

By late 1980s and early 1990s, when IEA considered the third wave of mathematics and science studies, it was decided that the two fields of studies be combined to form the Third International Mathematics and Science Study, or TIMSS. Data were collected in 1995, and altogether more than half a million students from more than 40 countries or systems participated in the study, the largest international comparative study up to that time. TIMSS was an extremely ambitious study. Not only did it encompass the two subjects of mathematics and science, but it also targeted three populations of students: the 9-year-olds (known as population 1), the 13-year-olds (population 2), and students in the final year of secondary education (population 3). To capture students of the target ages in populations 1 and 2, two adjacent grades with the highest number of students of the relevant age were sampled for study.

TIMSS investigated the mathematics and science curricula of the participating countries through an analysis of curriculum guides, textbooks, and other curricular materials. For the format of the test, questions requiring composed answers, in addition to multiple-choice items, were used. Furthermore, a subsample of students was selected to be tested on their hands-on skills (known as Performance Assessment). All these were done on top of

the usual questionnaires for students, teachers, schools, and system personnel.

The key findings of TIMSS were that, for most countries, gender differences in mathematics achievement were small or essentially nonexistent. In science, the gender differences in the primary school years were much less pervasive than in the middle school years. However, boys had significantly higher achievement than girls at both levels in about 50% of the countries, particularly in earth science and physical science. One spectacular finding of TIMSS is that the East Asian countries performed very well, especially in mathematics.

For student attitudes, the majority of students in nearly all the countries indicated that they preferred mathematics and science and that they did well in the two subjects. Interestingly, some of the countries with the highest achievement also were those whose eighth-grade students had the most negative perceptions of success.

For home and school factors, having educational resources at home was strongly related to student achievement. In school, for both mathematics and science, small-group work was used less frequently than other instructional approaches. Across countries, teachers reported that working together as a class with the teacher teaching the whole class, and having students work individually with assistance from the teacher were the most frequently used instructional approaches. At grade 8, teacher demonstrations of experiments were common in science lessons.

Third International Mathematics and Science Study Repeat (TIMSS-R)

Because of the great impact of TIMSS on the international community, a repeat study (TIMSS Repeat or TIMSS-R) was conducted in 1999 to study trends in mathematics and science achievements. In TIMSS-R, only eighth-grade students were tested. Thirty-eight countries participated in the study, 26 of which had participated in the eighth-grade test and 17 of them had participated in the fourth-grade test in 1995.

The key results of TIMSS-R are that, for most of the countries that participated in both TIMSS and TIMSS-R, there was not much difference in terms of their relative performance in the two studies. In mathematics, East Asian students continued to outperform their counterparts in other parts of the world, and in science, boys had significantly higher average achievement than girls in 16 of the 38 countries. The latter was attributable mainly to significantly higher performance by boys in physics, earth science, chemistry, and environmental and resource issues. The gender gap in science achievement is especially apparent among high-performing students, but the average gender difference showed a decrease from 1995 to 1999.

In 21 of the 38 countries, science was taught as a single general subject, while separate courses were offered in the different science subjects in other countries. The trend data from 1995 to 1999 show a small but significant increase for integrated science.

For variables related to achievement, there was no clear relationship between the wealth of countries and their students' achievements. Within the TIMSS-R countries, students with a higher level of educational resources at home and in school did better in the test.

As far as student attitudes are concerned, eighth-grade boys generally had a more positive self-concept in science than girls, and this difference was most pronounced in countries where the sciences are taught as separate subjects. Moreover, although student attitudes toward science were generally positive in countries where eighth-grade science is taught as a single subject, they were less positive in separate-science countries.

In mathematics, students' positive attitudes toward the subject were related with higher achievement within a country, but the same relationship did not hold across countries. In fact, with the exception of Singapore, all the top-performing countries had relatively negative attitudes toward mathematics. In particular, it is noticeable that students from all the five top-performing East Asian countries had very low self-image of mathematics.

As far as instructional practices are concerned, science teachers reported devoting substantial percentages of their class time to student experiments and teacher-guided student practice. In mathematics, teachers reported that the two most predominant activities in their classrooms were teacher lecture and teacher-guided student practice. Although solving nonroutine problems was mentioned in the intended curriculum of nearly all countries, teachers reported that they put relatively low emphasis on mathematics reasoning and problem solving.

Trends in International Mathematics and Science Study – TIMSS 2003

The publication of the TIMSS-R results again drew the attention of the education community and beyond, worldwide. IEA therefore decided to continue the study in 4-year cycles, with the following cycle to take place in 2003. TIMSS was renamed Trends in International Mathematics and Science Study.

More than 360 000 students in 49 countries participated in TIMSS 2003. In mathematics, at both fourth and eighth grades, gender difference was negligible in many countries, and the trend results at the eighth grade showed that a few more countries showed improvement in girls than boys. In science, the majority of eighth-grade boys outperformed girls, often by a substantial margin.

This was attributable mainly to higher performance by boys in physics and earth science, although girls had higher achievement in life science. Nevertheless, girls had greater improvement than boys, especially since 1999. At the fourth grade, gender difference in science achievement was negligible.

For factors related to student achievement, the home context for learning was important in helping to foster higher achievement, including having more highly educated parents, having more books at home, frequently using a computer, etc. Besides, students expecting to finish university had substantially greater achievement. Students with higher achievement attended schools with positive climates for learning, with fewer students from disadvantaged homes, and where teachers and students felt safe.

TIMSS 2007

TIMSS 2007 is the fourth and latest round of assessment in TIMSS. Together with results from previous rounds of TIMSS, it provides achievement information at four time points over a 12-year period. The key findings of TIMSS 2007 included the following:

- On trends of achievement, in both mathematics and science, more countries showed improvement in 2007 than declines in the fourth grade. At the eighth grade, the pattern was less pronounced. Differences in achievement between boys and girls were negligible in approximately 50% of the countries in both mathematics and science at the fourth grade. At the eighth grade, the differences in achievement between boys and girls were negligible in about one-third of the countries. In the remaining countries, girls had higher achievement than boys in more countries, especially in mathematics.
- Similar to findings in previous rounds of TIMSS, at the eighth grade, better home environment was associated with higher mathematics and science achievement. Students with more positive attitudes toward these subjects, who reported a higher level of self-confidence in learning mathematics and science, and placed a higher value on them as important to future success, also had higher achievement. Student achievement was highest where there were high expectations for student achievement and parental support.

Observations

Few people would dispute the idea that a high-quality education is extremely important for the individual and for society, and so there is a pressing need to gauge the

quality of education that our school system offers. But how do we find out how well our school system is doing? Are there ways in which our policies, our schools, and our teachers can do better? Many school systems around the world have measures in place to monitor the output and value-addedness of their system (e.g., through national testing). However, since many conditions are constant within a country, it is hard to objectively evaluate the effectiveness of the system. Many 'what if' questions simply cannot be answered through measures and studies within the system, since it is difficult or impractical to alter some conditions within the system. International studies conducted by IEA aim exactly at providing answers to these important questions, using the world as a laboratory. Through these IEA studies, we can measure how well students from our own system perform compared to students from other countries, especially countries which have a similar background. We can look at conditions which do not exist in our own system, and investigate how these conditions are affecting student achievement in other countries, and how conditions in our own system may be changed to bring about higher student achievement.

For 50 years, IEA has gathered detailed information on the mathematics and science achievements of students in participating countries, as well as a host of data on student, teacher, school, curriculum, and system backgrounds. The rich data set generated allows explanations of achievement to be explored, and thus may throw light on the answers to some of the questions mentioned above. The TIMSS series of studies also offers powerful information to study trends in mathematics and science learning. Results of these international studies help clarify the relationship between system background variables and student achievement and increase our understanding of our education system, and will no doubt provide important information for both policymakers and practitioners in improving student learning and achievement.

The IEA studies, however, do not elucidate all the factors affecting achievement or offer unambiguous answers to all our questions on quality of the school system. In particular, as can be seen from some of the findings above, many factors that explained differences in performance within a system failed to account for differences across systems. This failure should perhaps be expected for studies of the nature and scale of IEA. For example, notwithstanding all the precautionary measures taken by IEA, there are limitations in the use of questionnaires as an instrument in getting at the factors that influence student achievement in a cross-cultural setting. The same term in a questionnaire may mean different things in different languages, and the different cultural values may render some of the comparative results dubious. In particular, it is hard for instructional practices, a potentially important explanatory variable for achievement, to be captured by a self-reporting questionnaire

(Stigler *et al.*, 1999), and that is why the associated TIMSS video studies are so important in this regard (Heibert *et al.*, 2003). What these IEA studies do tell us is that there exist vast differences in mathematics and science achievement across a large number of countries. Hopefully, the recognition of the differences will fuel a search for the ways of enhancing our education system rather than a race to top the league tables generated by such studies.

Because of the limitations of large-scale international studies discussed above, care should be taken in interpretation of the results, especially in understanding the meaning and contexts behind the ranking of countries. This is especially important for policymakers, since IEA studies, such as TIMSS, have already had major impact on educational policies in many countries. Nevertheless, it should be pointed out that these studies have been conducted in very rigorous manners, and extreme care has been taken on the side of IEA to ensure that the results are as reliable as such large-scale studies can afford. Therefore, the results of such studies should not be brushed aside lightly.

The most important thing is perhaps that such international studies are meant for countries to learn from each other, and before learning from another culture, one should first understand one's own cultural values, take them as a given starting point, and design and improve educational practice on that basis. Complicated factors might have affected classroom practices and student achievement, and drastic changes should not be undertaken until such factors are thoroughly examined. Simple transplant of policies and practices from high-achieving countries to low-achieving ones would not work because one cannot transplant the practice without regard to the cultural differences, and any change in educational policy must ensure that the strength in a country is not lost in the process. What is needed is to identify not only superficial differences in educational settings and practices as identified by these international studies, but also the intricate relationships between these practices and the underlying culture. Through a careful study of these relationships, one may then determine how much can or cannot be borrowed from another culture.

Concluding Remarks

Public interest in IEA studies usually focused only on the ranking of countries in the league tables generated, and it is understandable that attention is drawn to this aspect of the findings. However, there is much more to the IEA studies than a league table. As argued above, these studies for which participating countries have invested so much in terms of money and manpower deserve the attention beyond a competition of student achievement.

Furthermore, results of IEA studies often provide for countries' impetus toward educational changes, but sometimes, such changes are made without a careful consideration of the complex context within which the achievement and classroom instructions are situated.

To make better use of the IEA studies, the study results should serve as mirrors for us to better understand our system. Education is an extremely complex enterprise, and one cannot expect these international studies to produce quick answers for all national problems in education. Instead, the major contribution of IEA studies is that they provide a rich data set based on which individual countries may engage in the process of seeking solutions to improve their education system. In this regard, we need discernment and wisdom, and not just data and a league table!

See also: IEA: Globalization and Assessment.

Bibliography

- Husén, T. (ed.) (1967). *A Comparison of Twelve Countries: International Study of Achievement in Mathematics*, vols. 1–2. Stockholm: Almqvist and Wiksell.
- Robitaille, D. F. and Garden, R. A. (eds.) (1989). *The IEA Study of Mathematics II: Context and Outcomes of School Mathematics*. Oxford: Pergamon Press.
- Stigler, J. W., Gonzales, P., Kawanaka, K., Knoll, S., and Serrano, A. (1999). *The TIMSS Videotape Classroom Study, Office of Educational Research and Improvement*. Washington, DC: US Department of Education.
- Travers, K. J. and Westbury, I. (eds.) (1989). *The IEA Study of Mathematics I: Analysis of Mathematics Curricula*. Oxford: Pergamon Press.
- Walker, D. A. (1976). *The IEA Six-Subject Survey: An Empirical Study of Education in Twenty-One Countries*. Stockholm: Almqvist and Wiksell.
- Burstein, L. (ed.) (1992). *The IEA Study of Mathematics III: Student Growth and Classroom Processes*. Oxford: Pergamon Press.
- Comber, L. C. and Keeves, J. P. (1973). *Science Education in Nineteen Countries*. Stockholm/New York: Almqvist and Wiksell/Wiley.
- Foshay, A. W., Thorndike, R. L., Hotyat, F., Pidgeon, D. A., and Walker, D. A. (1962). *Educational Achievement of Thirteen-Year-Olds in Twelve Countries*. Hamburg: UNESCO Institute for Education.
- Hiebert, J., Gallimore, R., Garnier, H., et al. (2003). *Teaching Mathematics in Seven Countries. Results From the TIMSS 1999 Video Study*. Washington, DC: National Center for Education Statistics.
- Keeves, J. P. (ed.) (1992). *The IEA Science Study III: Changes in Science Education and Achievement: 1970 to 1984*. Oxford: Pergamon Press.
- Martin, M. O., Mullis, I. V. S., Beaton, A. E., et al. (1997). *Science Achievement in the Primary School Years*. Chestnut Hill, MA: International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., and Foy, P. (2008). *TIMSS 2007 International Science Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., et al. (2000). *TIMSS 1999 International Science Report*. Chestnut Hill, MA: International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., and Chrostowski, S. J. (2004). *TIMSS 2003 International Science Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Beaton, A. E., et al. (1997). *Mathematics Achievement in the Primary School Years*. Chestnut Hill, MA: International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., et al. (2000). *TIMSS 1999 International Mathematics Report*. Chestnut Hill, MA: International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., and Foy, P. (2008). *TIMSS 2007 International Mathematics Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., and Chrostowski, S. J. (2004). *TIMSS 2003 International Mathematics Report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Postlethwaite, T. N. (ed.) (1967). *School Organization and Student Achievement: A Study Based on Achievement in Mathematics in Twelve Countries*. Stockholm/New York: Almqvist and Wiksell/Wiley.
- Postlethwaite, T. N. and Wiley, D. E. (eds.) (1992). *Science Achievement in Twenty-Three Countries: The IEA Study of Science II*. Oxford: Pergamon Press.
- Rosier, M. J. and Keeves, J. P. (1991). *Science Education and Curricula in Twenty-Three Countries: The IEA Study of Science I*. Oxford: Pergamon Press.

Further Reading

- Beaton, A. E., Martin, M. O., Mullis, I. V. S., et al. (1997). *Science Achievement in the Middle School Years*. Chestnut Hill, MA: International Study Center, Boston College.
- Beaton, A. E., Mullis, I. V. S., Martin, M. O., et al. (1997). *Mathematics Achievement in the Middle School Years*. Chestnut Hill, MA: International Study Center, Boston College.

Relevant Websites

- <http://www.iea.nl> – International Association for the Evaluation of Educational Achievement.
- <http://timss.bc.edu> – TIMSS and PIRLS International Study Center.

IEA Study in Civic Education

J Torney-Purta and J-A Amadeo, University of Maryland, College Park, MD, USA
J Schwille, Michigan State University, East Lansing, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Civic dispositions – Attitudes toward various kinds of civic participation and attitudes toward social justice issues, such as immigrants' and minorities' rights.

Civic knowledge – Ability to recognize and understand concepts in the civic and political domain such as democratic practices, constitutions, political rights, and press freedom.

Civic skills – The ability to apply civic knowledge to understand political communications such as cartoons and election leaflets.

Concepts of citizenship – Beliefs about the characteristics and norms of good adult citizenship

National Assessment of Educational Progress (NAEP) – Test administered to samples of U.S. students in various subject areas including civic education.

Open classroom climate – The perception that the classroom is a place for respectful discussion of students' views about civic issues, including controversial issues.

Qualitative research – Uses data from document analysis, interviews, and/or observations often gathered into case studies or other descriptive formats.

Quantitative research – Uses data from a test, survey, or experimental performance analyzed with statistical techniques.

Introduction to International Association for the Evaluation of Educational Research International Comparative Studies

The International Association for the Evaluation of Educational Research (IEA) celebrated its 50th anniversary in 2008. It was started by a small number of well-known educational researchers from Europe and the United States, who decided in 1959 to see if it was possible to design and use a cross-national test of educational achievement (to supplement the data on which national comparisons had previously been based, such as

enrolment figures). Since the 1960s, IEA has pioneered in carrying out influential large-scale, cross-national studies in education, the Third International Mathematics and Science Study, now Trends in Mathematics and Science Study (TIMSS), and Progress in Reading Literacy (PIRLS). From being a small group of distinguished researchers with an informal decision-making structure, it has developed into a consortium of institutions in over 60 countries.

IEA's formal research agenda started with the first international mathematics study. Mathematics was chosen because it was thought to be relatively easy to measure and was taught universally. Thus, measures of achievement in mathematics could be considered an important outcome of education suitable for international comparison. The aim of IEA from its beginning to the present has been not only to measure outcomes as national educational indicators, but also to attempt to explain why some students, schools, and countries have better outcomes on these tests than others. From the start, data were collected about potential explanatory factors at the individual, school, and country levels. Some of the studies have emphasized country differences, but some of the studies have placed relatively equal emphasis on individual or school-level predictors of achievement as on analyzing country differences.

Following the success of the initial mathematics study in the late 1960s, IEA launched one of its most ambitious efforts, a study which addressed six subject-matter domains, including science, reading, literature, French as a foreign language, and English as a foreign language. Civic education was added as the sixth subject, but not without controversy. It is difficult to recapture today the concerns that surrounded this domain in the midst of the Cold War. What counted as legitimate civic education in one country was not what counted in countries with different ideologies. A few communist countries were starting to participate in IEA studies in mathematics and science. Second, measurement was daunting in the civic-education domain, where attitudes were important as desired outcomes of civic education, and where model standards for measurement of knowledge were rare. In other words, this study was a bold move with risks both for the researchers and for IEA as an organization. Nevertheless, under the leadership of J. Torney-Purta, A. N. Oppenheim, R. Farnen, and J. Schwille, the study was designed and implemented successfully, testing in 1971, though in fewer than a dozen countries. By the time the results were published

(Torney *et al.*, 1975), civic education was recognized as a solid part of the IEA six-subject survey.

Nevertheless, compared to the IEA core subjects of mathematics, science, and reading/literacy, civic education remained at the margins of the IEA effort, along with foreign languages, literature, and written composition. There were no timely efforts to follow up with a second study. It was 20 years before the second IEA Civic Education Study began (with a planning committee chaired by J. Torney-Purta starting in 1994 and testing in 1999–2000). This study, called IEA Study of Civic Education (CIVED), established the civic domain as a part of the IEA core. The second IEA Civic Education Study capitalized on interest in preparing young people for life in newly created democracies and recruited more than two dozen countries to participate (11 post-Communist countries, 12 other European countries, two Latin American countries, Australia, Hong Kong SAR, and the United States). The methodological procedures and standards that had been established by IEA for the TIMSS and PIRLS studies were helpful in allowing the second civic education study to meet high standards of methodological rigor. With widely disseminated results of the CIVED study and a third study (International Civic and Citizenship Education Study (ICCS) testing in 2008–09), civic education has found a niche in providing respected research evidence for international policy discussions of education.

The Roots of the CIVED Study

In 1993, when IEA's general assembly first expressed interest in a second study of civic education, a basic question formed the core of interest: how were young people prepared to undertake their roles as citizens in a range of countries? How was this different in new democracies (where teachers and parents had been educated under a different value system) and in older democracies (which were seeing declining proportions of young people interested in civic and political participation)? Further, how would it be possible to encompass both the meaning of civic education for a nation (e.g., the need to transmit basic knowledge about legislative processes, national attachment, and belief in the rule of law and attitudes toward civic participation) and its meaning for individual young people of different levels of cognitive development or different ethnic groups? An earlier educational encyclopedia entry for civic education noted, "The civic values that individuals internalize often differ substantially from those that representatives of the polity [including educators] have attempted to inculcate, either because of individual differences in the processing of information or because of experience and reflection, which leads to challenging the status quo" (Torney-Purta, 1992: 158).

In the years between the first and second civic-education studies, substantial research had been conducted on the topic of political socialization, much of it by political scientists and primarily using surveys measuring the attitudes of young people of secondary-school age in the United States and Europe (Jennings, 2007). There was a diminution in this research in the late 1970s to 1980s, with some studies by psychologists and social studies educators filling the gaps (Torney-Purta, 1990). The National Assessment of Educational Progress (NAEP) in the United States tested the civic knowledge of representative samples of fourth, eighth, and twelfth graders starting in the 1970s and continuing to the present. A reanalysis of the 1988 NAEP data suggested that educators should again give attention to this area (Niemi and Junn, 1998). This analysis indicated that civic-education courses in upper-secondary school were important, but also pointed to the existence of considerable learning before that age (corroborating studies by psychologists, who had also studied attitudes and intentions to participate using multimethod studies). At about this time, improving civic education was also being discussed in Australia and England.

Design of the CIVED Study and the Associated Collaborative Process

The CIVED Study was designed and implemented in two phases, the first phase more qualitative in nature, and the second more quantitative (see Table 1). During the first phase, national researchers from 24 countries collected information on the circumstances, content, expectations, and processes of civic education in their countries including students' civic engagement and political interest. These national case studies provided insight into the political, social, economic, and educational context within which civic identity develops and civic education occurs (Torney-Purta *et al.*, 1999). The information collected during the first phase also contributed to the conceptual framework which then guided the design of a test of civic knowledge and skills and a survey of civic attitudes, concepts, and expected behaviors administered to students during the second phase of the study.

During the second phase, approximately 90 000 14-year-old students from 28 countries were administered a test and a survey, each taking about 40 min. These data were collected in 1999; findings were released in 2001, and reported in the volume titled *Citizenship and Education in Twenty-Eight Countries: Civic Knowledge and Engagement at Age Fourteen* (Torney-Purta *et al.*, 2001). School heads and teachers also provided information.

In the following year, over 50 000 upper secondary school students from 16 countries (15 of which also tested 14-year-olds) received a similar test of civic knowledge and skills and the same survey of civic attitudes and

Table 1 CIVED study: Focus of two-phased study^a

	<i>CIVED phase 1 (1999)</i>	<i>CIVED phase 2 (2001/2002)</i>
Number of countries	24	28 (14-year-olds)
Primary method	Qualitative	16 (16–19-year-olds) Quantitative
Primary focus	Context within which civic education occurs Countries' expectation for civic education outcomes for 14-year-olds Collaboration and consensus about the study among national researchers Development of conceptual framework: Democracy and its definition; democratic institutions; citizenship rights and duties. National identity; international relations Social cohesion and diversity.	Contents of test and survey for age 14: Civic knowledge and cognitive skills: 38 items and 3 IRT scales Concepts of democracy, citizenship, and government: 52 items and 4 IRT scales Attitudes of trust in institutions, attitudes toward one's country and government, political efficacy: 24 items and 5 IRT scales Attitudes toward immigrants, women, minorities: 34 items and 4 IRT scales Expected political activities: 12 items and 4 IRT scales School conditions for learning: 19 items and 2 IRT scales Student background: Demographics and organizational participation 31 items

^aTest for upper-secondary students contained similar items as well as economic literacy items. Survey was identical to survey given to 14-year-olds with a few additional items about contextualized efficacy and intervention added.

Table 2 Countries participating in CIVED study (2001 and 2002)

Australia	Finland	Poland ^a
Belgium (French)	Germany	Portugal ^a
Bulgaria	Greece	Romania
Chile ^a	Hong Kong (SAR) ^a	Russian Federation ^a
Colombia ^a	Hungary	Slovak Republic
Cyprus ^a	Italy	Slovenia ^a
Czech Republic ^a	Latvia ^a	Sweden [*]
Denmark ^a	Lithuania	Switzerland ^a
England	Norway ^a	United States
Estonia ^a		

^aCountries that tested both 14-year-old students and upper-secondary students. Israel tested only upper-secondary students.

behaviors (Amadeo *et al.*, 2002). In sum, during the second phase of the study, quantitative data were collected from over 140 000 students to provide an assessment of the civic knowledge and engagement of both younger and older adolescents across a diverse set of democratic societies (see **Table 2** for a list of participating countries).

Both groups of students were tested on their civic knowledge and skills. They were also surveyed on their concepts of democracy and government; their attitudes relating to trust in institutions, political efficacy, and opportunities for immigrants and women; and their expected participatory actions relating to politics. A final part of the student survey assessed the students' perceptions of the climate for discussion in their classrooms as well as other background variables. In addition, an internationally relevant list of organizations was developed and students were asked to indicate to which groups they

belonged. This list includes organizations such as student government, environmental groups, and other youth organizations (see **Table 1**).

The theoretical framework of the CIVED conceptualizes the ways in which "the everyday lives of young people in homes, with peers and at school serve as a 'nested' context for young people's thinking and action in the social and political world" (Torney-Purta *et al.*, 2001: 20). This theoretical model, found in **Figure 1**, has its roots in ecological theory (Bronfenbrenner, 1988) and situated cognition (Lave and Wenger, 1991).

Selected Results

The first major result of these cross-national studies of civic education has been to demonstrate the feasibility and utility of focusing on civic education as a domain for cross-national assessment and research. When the first civic-education study was started by IEA in the late 1960s, it was met with skepticism for attempting precise measurement of a politically contentious sphere. It continues to be the one area of IEA research, where noncognitive outcomes receive as much or more emphasis than cognitive outcomes. In short, over the years the IEA work in civic education has continued to be a pioneering effort in defining and measuring what it means for young people to learn about democracy and related matters such as national identity and social cohesion. Already in the first study in the 1970s, the Federal Republic of Germany had attained remarkable support among young people for attitudes considered important to democracy by Western

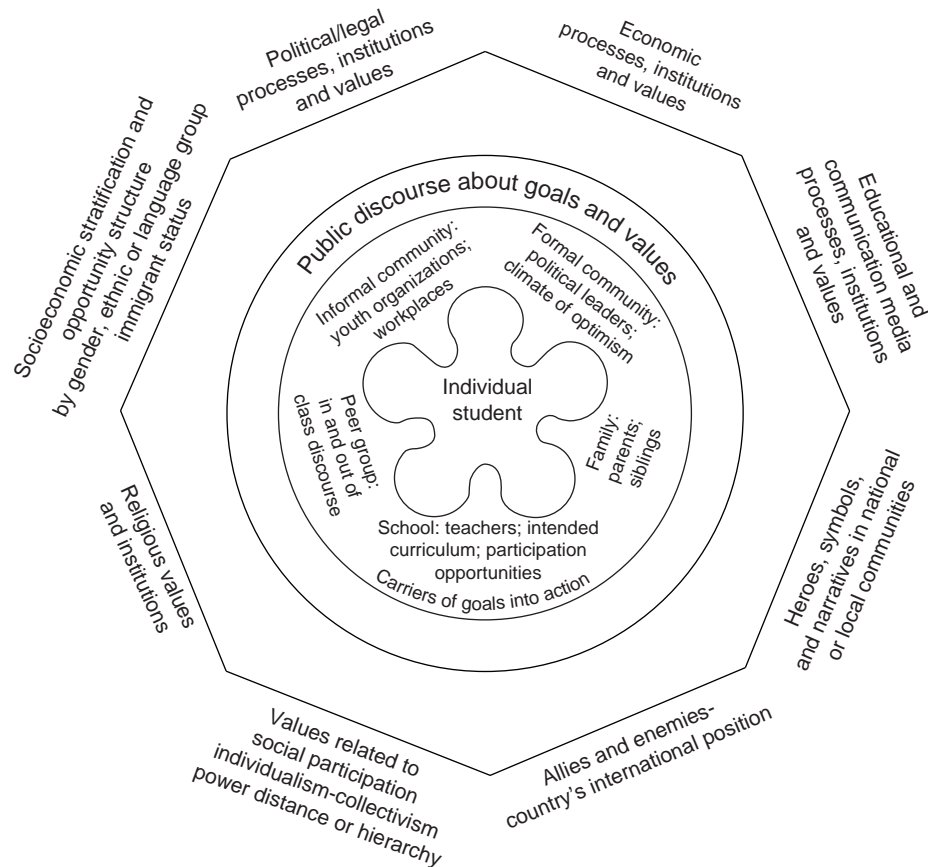


Figure 1 The octagon model of Civic Education Study (CIVED). Torney-Purta, J., Schwille, J., and Amadeo, J. (1999). *Civic Education across Countries: Twenty-Four National Case Studies from the IEA Civic Education Project*. Amsterdam: IEA.

theorists whereas nations with longer traditions of representative democracy did not all reach the same level. By the early years of the twenty-first century, the cross-national study of civic education had progressed from a marginal and controversial part of the IEA research agenda to a domain that receives systematic attention.

Second, while civic education is still generally considered a difficult area to assess and deal with in complex multivariate and psychometric models, in fact it has been possible to develop robust results with considerable implications for education using data in this subject area. One of the most significant findings to emerge from IEA cross-national civic achievement studies in the last 40 years is the relationship, controlling for potentially confounding variables, between the existence of a climate that encouraged open discussion in classrooms on the one hand, and scores on civic-education knowledge tests (as well as on attitude measures) on the other. In cross-national studies of other subject matters such as mathematics, science, and literacy, generalized cross-country relationships between pedagogical or classroom-process variables and achievement have proved elusive. What seems important in the case of civic education is that the

empirically demonstrated importance of encouraging open expression by students in classrooms is consistent with and supportive of much theorizing about the importance of discussion and deliberation to desired outcomes in civic education (Hess and Avery, 2008).

Third, the qualitative national case studies of civic education in the 1990s collected in the first phase of CIVED provided in-depth information about what sort of role schooling can and should have in civic education, challenging the conventional wisdom that much of civic education takes place outside schools. These case studies (Torney-Purta *et al.*, 1999) revealed a strong consensus that civic education was not sufficiently responsive to the alienation from political life observed among young people. Issues confronting civic education in school were examined, including the status of civic education as a separate subject (vs. integration with other subjects), the needs of civic education pedagogy, the importance of school organizations as well as student rights and student government for civic preparation, and how civic education can respond to forces outside the school such as the media. Although a very ambitious role for schooling in civic education has been proposed, few countries in the 1990s claimed to have

been successful in implementing this challenging agenda. Assessment and teacher preparation have been especially neglected. In fact, realizing the aspirations of civic education as documented in these case studies would require systemic reforms of education far beyond the capacity of civic educators to carry out without support from members of the community (including parents) and those who shape the policy of educational, political, and economic institutions (Schwille and Amadeo, 2002).

Fourth, the CIVED survey results interpreted in light of both observed country similarities and differences show a multidimensional picture of the outcomes of the civic-education process. For example, there was considerable consensus among 14-year-olds across countries about important norms for political engagement such as obeying the law and voting. In most of the countries, testing in CIVED the less conventional but nonprotest types of activities, such as joining an environmental or volunteer organization, were more positively viewed than conventional political activities such as running for political office or writing a letter about a political issue. Even young people who believed strongly in social justice were unlikely to say that they would actively protest for a political cause. Very small proportions across countries planned to join a political party. Even among upper-secondary students (aged 16–19), when many students are eligible to vote and welcome to join partisan organizations, the preference is still for community volunteering and support of human rights and the environment (Torney-Purta and Amadeo, 2003).

Fifth, some widely held beliefs about civic education were challenged by these results. Many argued that the most important reason for civic-related classes is that young people need historical and contemporary civic knowledge. However, the CIVED results indicate that knowledge of civic and political processes and concepts by itself is insufficient to ensure participation. Civic knowledge was strongly related to the students' expectations that they would vote but was not related to expectations of other forms of adult engagement such as political activities extending beyond voting or engagement in volunteer activities. Knowledge of civic topics was related to students' support for the rights of immigrants and minorities. There was a curious lack of willingness to move beyond expressing attitudes to action in support of equality among many students who express beliefs in social justice (Torney-Purta, 2009).

Sixth, the need for a balanced emphasis on skills in addition to content knowledge was suggested by the CIVED results. Distinct measures of conceptual civic knowledge and civic skills (assessed by questions asking for the interpretation of an election leaflet or political cartoon) were developed. Students in the United States had the highest score of the 28 countries on these cognitive civic skills. Other countries where hands-on teaching

methods and political media are regularly used in class, such as England, Australia, and Sweden, also had relatively high scores on civic skills. Students in the post-Communist countries performed much less well on the portion of the test requiring mastery of skills. However, students in some Eastern and Central European countries scored well on questions asking about democratic concepts, such as what role constitutions play in democracy. In contrast, students in the United States ranked tenth out of the 28 countries in this conceptual knowledge of democracy.

Seventh, in democratic societies, civic knowledge, civic skills, inclinations to participate, and civic-education opportunities are supposed to be equally distributed across economic groups. In fact, the IEA study shows a substantial equality gap in civic knowledge as well as in expectations of voting and some other types of activity that is evident when comparing students from home backgrounds with more and fewer educational resources and when comparing those who plan to attend higher education with those who expect to drop out before completing secondary school (Baldi *et al.*, 2001; Torney-Purta *et al.*, 2001; Wilkenfeld, 2009). Furthermore, a small group of students in the United States has highly negative attitudes toward minorities and toward obeying the law; a substantial proportion of ninth graders who expect to drop out before completing high school are in this alienated group (Torney-Purta *et al.*, 2008). Schools with many students from disadvantaged home backgrounds also experience curricula that lack major positive civic-education features (Baldi *et al.*, 2001; Kahne and Middaugh, 2008).

Dissemination of CIVED Results and Influence on Policy Resulting from the Research

The IEA organization prescribes a pattern for its publications. After the technical committee has approved a study's instrument and plans including scaling, the basic analysis by country is completed, resulting in tables or figures of results, usually concentrating on country comparisons. A draft report is prepared, usually by the International Coordinating Center (for CIVED, the Humboldt University of Berlin) and with the close involvement of the international steering committee (for CIVED coordinated through the Department of Human Development, University of Maryland). This draft was circulated to the national research coordinators for their suggestions. The volume reporting the data collected from nearly 90 000 14-year-olds in 28 countries (Torney-Purta *et al.*, 2001) was released at the Comparative and International Education Society's meeting in March 2001, while the report of the 50 000 upper-secondary students surveyed in 16 countries

(Amadeo *et al.*, 2002) was released at the meeting of the International Society for Political Psychology in August of 2002. An executive summary for each volume and a webpage, which includes publications and the released survey questions were prepared. Activities to stimulate the awareness of the public, of professional educators, and of policymakers about the results of the CIVED study and its implications for policy and practice were also undertaken. Numerous articles tailored to particular audiences of educators and policymakers, as well as presentations at conferences to audiences with interests in policy or professional practice stimulated a broadened debate about young people's preparation for civic engagement.

The study was released, and there was a resurgence of interest in civic learning among educators and foundations in many countries. Empirical evidence about the knowledge and attitudes of young people received new attention. For example, in the United States, a consensus process involving researchers and practitioners supported by the Carnegie Corporation of New York and the Pew Charitable Trusts resulted in a report entitled *The Civic Mission of Schools*; the IEA CIVED research is a frequently quoted source of evidence in this influential report. The report has served as the basis of a major public-awareness campaign on civic learning. Similar reports were formulated in England and Australia in the 1990s and the CIVED research evidence played a role in their implementation.

The Education Commission of States (a nongovernmental organization (NGO) in the US) mounted an effort to assist state policymakers to clarify the meaning of competencies in the areas of civic knowledge and skills in order to enhance policies in the 50 states to promote these competencies (Torney-Purta and Vermeer, 2006). An associated effort collected and disseminated high quality and innovative measures for assessing students and schools, in order to encourage the assessment of students' civic knowledge, skills, and dispositions as part of educational accountability. The resulting website of juried items included many from the CIVED. Further examination of the CIVED data prepared for scholarly audiences in the United States has included an analysis of differences between the civic knowledge and attitudes of Latino/a and non-Latino/a students in the United States (Torney-Purta *et al.*, 2007) as well as estimating the proportion of US students who were conventionally civically engaged, engaged in support of social justice, indifferent, and alienated (Torney-Purta *et al.*, 2008).

The Council of Europe launched the year of education for democratic citizenship in Europe in 2005, and a report based on the CIVED data served as their background paper on policies regarding the practice of democracy at school. The emphasis in this paper was on the relation between measures of the existence of democratic practices at school and student outcomes within and among

countries. Additional analysis for scholarly audiences has taken a similar direction, looking at the impact of the participation of governments in intergovernmental dialog about human rights on the knowledge of and attitudes toward human rights among adolescents in 27 countries (Torney-Purta *et al.*, 2008).

The Organization of American States supported a specialized analysis of the IEA data from Chile, Colombia, Portugal, and the United States published in *Strengthening Democracy in the Americas through Civic Education: An Empirical Analysis of the Views of Students and Teachers* (2004). The analysis identified problems in this region with students' lack of awareness of threats to democracy and poor understanding of citizens' rights that had been obscured when all 28 countries were analyzed. The National Commission on the Reform of Citizenship in Chile reflected these research findings in their recommendations for curricular reform. A similar effort has addressed the results from the Pacific Rim countries that participated in CIVED (Kennedy *et al.*, 2008). A special issue of the British journal *Citizenship Teaching and Learning*, was entitled *Reflections on the IEA Civic Education Study (1995-2005)*, and included contributions covering Latin America, the post-Communist countries, the United States, and Australia.

All IEA studies release their data for use by other scholars. The first step was the issuing of a technical report (Schulz and Sibbers, 2004). Civic Education Data and Researcher Services (CEDARS) was established at the University of Maryland, US to provide assistance to researchers, graduate students, and policymakers' staff. This culminated in the acceptance of the data for systematic dissemination by the Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan, US. Political scientists have used the CIVED data to study, among other things, the extent to which women political leaders serve as models for adolescent girls' political engagement (Campbell and Wolbrecht, 2006).

The International Civic and Citizenship Education Study (ICCS 2009) builds on the previous CIVED. The studies have a similar purpose, to investigate the ways in which young people are prepared to undertake their roles as citizens in a range of countries. ICCS includes student achievement on a test of conceptual understandings and competencies in civic and citizenship education, but does not distinguish between content and skills as CIVED did. ICCS has also collected data about student dispositions and attitudes relating to civic and citizenship education. The population tested includes students enrolled in the grade that represents 8 years of schooling, provided the mean age at the time of testing is at least 13.5 years. ICCS teacher and school questionnaires gathered information about teaching and class-management practices, school governance, and climate. An innovation in this study is the inclusion of regional modules for Asia, Europe, and Latin America.

Summary

The first IEA study of civic education began its planning in the late 1960s. Since that initial study, which tested in 1971, civic-education research has become an integral part of cross-national study. While it is still considered a difficult domain to assess, the IEA civic education studies demonstrate that it is possible to develop research models and assessment instruments relevant to countries with different ideologies. The major findings (particularly those from CIVED) have had substantial implications for education policy and teaching. Further, the research has emphasized that the distinction between civic knowledge and skills is an important one, and that civic knowledge alone is not sufficient for democratic participation. Young people's attitudes, activities, and concepts are also of vital importance. The findings from CIVED have been disseminated widely in numerous countries and secondary analysis of the data is ongoing.

See also: IEA: Globalization and Assessment.

Bibliography

- Amadeo, J., Torney-Purta, J., Lehmann, R. H., Husfeldt, V., and Nikolova, R. (2002). *Civic Knowledge and Engagement among Upper Secondary Students. Citizenship and Education in Sixteen Countries*. Amsterdam: IEA.
- Baldi, S., Peries, M., Skidmore, D., Greenberg, E., and Hahn, C. (eds.) (2001). *What Democracy Means to Ninth Graders: U.S. Results from the IEA Civic Education Study*. Washington, DC: National Center for Education Statistics.
- Bronfenbrenner, U. (1988). Interacting systems in human development. In Bolger, N., Caspi, C., Downey, G., and Moorehouse, M. (eds.) *Persons in Context: Developmental Processes*, pp 25–30. Cambridge, UK: Cambridge University Press.
- Campbell, D. E. and Wolbrecht, C. (2006). See Jane run: Women politicians as role models for adolescents. *Journal of Politics* **68**(2), 233–247.
- Hess, C. and Avery, P. (2008). Discussion of controversial issues as a form and goal of democratic education. In Arthur, J., Davies, I., and Hahn, C. (eds.) *The Sage Handbook of Education for Citizenship and Democracy*, pp 506–518. Los Angeles, CA: Sage.
- Jennings, M. K. (2007). Political socialization. In Dalton, R. J. and Klingemann, H. D. (eds.) *The Oxford Handbook of Political Behavior*, pp 29–44. New York: Oxford University Press.
- Kahne, J. and Middaugh, E. (2008). Democracy for some: The civic opportunity gap in high school. *Working Paper #59*. College Park, Maryland: The Center for Information and Research on Civic Learning and Engagement (CIRCLE).
- Kennedy, K., Hahn, C., and Lee, W.-O. (2008). Constructing citizenship: Comparing the views of students in Australia, Hong Kong, and the United States. *Comparative Education Review* **52**(1), 53–59.
- Lave, L. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Niemi, R. and Junn, J. (1998). *Civic Education: What Makes Students Learn*. New Haven, CT: Yale University Press.
- Schulz, W. and Sibbels, H. (2004). *IEA Civic Education Study Technical Report*. Amsterdam: IEA.
- Schwille, J. and Amadeo, J. (2002). The paradoxical situation of civic education in schools: Ubiquitous and yet elusive. In Steiner-Khamsi, G., Torney-Purta, J., and Schwille, J. (eds.) *New Paradigms and Recurring Paradoxes in Education for Citizenship: An International Comparison*, pp 105–139. Amsterdam: Elsevier.
- Torney, J. V., Oppenheim, A. N., and Farnen, R. F. (1975). *Civic Education in Ten Countries: An Empirical Study*. New York: Wiley.
- Torney-Purta, J. (1990). Youth in relation to social institutions. In Feldman, S. and Elliot, J. (eds.) *At the Threshold: The Developing Adolescent*, pp 457–478. Cambridge, MA: Harvard University Press.
- Torney-Purta, J. (1992). Civic education. In Alkin, M. (ed.) *Encyclopedia of Educational Research*, pp 158–161. New York: MacMillan.
- Torney-Purta, J. (2009). International psychological research that matters for policy and practice. *American Psychologist* **64**(8), 822–837.
- Torney-Purta, J. and Amadeo, J. (2003). A cross national analysis of political and civic involvement among adolescents. *Political Science and Politics* **36**, 269–274.
- Torney-Purta, J., Barber, C., and Wilkenfeld, B. (2007). Latino adolescents' civic development in the United States: Research results from the IEA Civic Education Study. *Journal of Youth and Adolescence* **36**(3), 111–126.
- Torney-Purta, J., Barber, C., Wilkenfeld, B., and Homana, G. (2008). Profiles of civic life skills among adolescents: Indicators for researchers, policymakers, and the public. *Child Indicators Research* **1**, 96–106.
- Torney-Purta, J., Lehmann, R., Oswald, H., and Schulz, W. (2001). *Citizenship and Education in Twenty-Eight Countries: Civic Knowledge and Engagement at Age Fourteen*. Amsterdam: IEA. <http://www.terpconnect.umd.edu/~iea> (accessed September 2009).
- Torney-Purta, J., Schwille, J., and Amadeo, J. (1999). *Civic Education across Countries: Twenty-Four National Case Studies from the IEA Civic Education Project*. Amsterdam: IEA.
- Torney-Purta, J. and Vermeer, S. (2004, updated version 2006). Developing citizenship competencies from kindergarten through grade 12: *A Background Paper for Policymakers and Educators*. Denver, CO: Education Commission of the States 32 pages. <http://www.ecs.org> (accessed September 2009).
- Torney-Purta, J., Wilkenfeld, B., and Barber, C. (2008). How adolescents in twenty-seven countries understand support and practice international human rights. *Journal of Social Issues* **64**(4), 857–880.
- Wilkenfeld, B. (2009). Does context matter? How the family peer, school, and neighborhood contexts relate to adolescents' civic engagement: *Working Paper #64*. Medford, MA: Tufts University Center for Information and Research on Civic Learning and Engagement. <http://www.civicyouth.org> (accessed September 2009).

Further Reading

- CIRCLE and Carnegie Corporation of New York (2003). *The Civic Mission of Schools*. New York: Carnegie Corporation of New York.
- Torney-Purta, J. and Amadeo, J. (2004). *Strengthening Democracy in the Americas through Civic Education: An Empirical Analysis of the Views of Students and Teachers*. Washington, DC: Organization of American States.

Relevant Websites

- <http://civicyouth.org> – Center for Information and Research on Civic Learning and Engagement (CIRCLE).
- <http://www.ecs.org/qna> – Education Commission of the States.
- <http://www.terpconnect.umd.edu> – IEA Civic Education Study (CIVED).
- <http://iccs.acer.edu.au> – International Civic and Citizenship Education Study (ICCS).
- <http://dx.doi.org> – University of Michigan ICPSR.
- <http://www.civicmissionofschoools.org> – Campaign for the Civic Mission of Schools.

IEA: Globalization and Assessment

H Wagemaker, International Association for the Evaluation of Educational Achievement, Amsterdam, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The International Association for the Evaluation of Educational Achievement (IEA) is an independent, not-for-profit cooperative of national research institutions and governmental research agencies that collaborate to conduct large-scale comparative studies of educational achievement. The association was legally incorporated as a scientific society in Belgium on 21 September 1967. Its legal and financial affairs are conducted through the Stichting IEA Secretariaat, Nederland. The secretariat is situated in Amsterdam and is a registered foundation under the Chamber of Commerce in The Hague, the Netherlands. A standing committee in association with the Amsterdam secretariat functions as a board of directors. The IEA Data Processing and Research Center (DPC), based in Hamburg, Germany, provides ongoing organizational and operational support for all studies.

Membership in IEA is open to all countries, each of which is required to nominate a national research center from which to conduct IEA studies. The member countries of IEA constitute the association's General Assembly, which meets on an annual basis to review project progress and develop new initiatives. In 2008, 66 countries from five continents were members of the organization. To date, approximately 100 countries have participated in IEA's various studies.

Through its comparative research and assessment projects, IEA aims to achieve the following:

- Provide international benchmarks that may assist policymakers to identify the comparative strengths and weaknesses of their educational systems.
- Provide high-quality data that increase policymakers' understanding of the key school- and non-school-based factors that influence teaching and learning.
- Provide high-quality data that serve as a resource for identifying areas of concern and action, and for preparing and evaluating educational reforms.
- Develop and improve the capacity of educational systems to engage in national strategies for educational monitoring and improvement.
- Contribute to developing the worldwide community of researchers engaged in educational evaluation.

Although IEA did not become a legal entity until 1967, its origins date back to 1958 when a group of scholars, educational psychologists, sociologists, and psychometricians met at the United Nations Educational, Scientific

and Cultural Organization (UNESCO) Institute for Education in Hamburg to discuss problems related to school and student evaluation. In an era when the quality of schooling was largely judged on the basis of school inputs or of data-free assertions about the relative merits of one educational system over another, those attending the meeting concluded that effective evaluation of schooling also requires a focus on student outcomes, including knowledge, attitudes, and participation.

Despite IEA's first meetings being held under the auspices of UNESCO, the association's roots were firmly tied to leading academic institutions through its first scholars, who included Benjamin Bloom from the University of Chicago, Arnold Anderson and Robert Thorndike from Teachers College, Columbia University, and Torsten Husén from the University of Stockholm. During the early years of the association, these leading academics were primarily responsible for organizing its work. Motivated by a desire to understand and identify those factors that might have meaningful and consistent influences on educational outcomes, they contributed their time and resources on a voluntary basis. The international comparative focus was, for them, a context from which to view the policy- and practice-related differences that occur naturally across countries and thereby provide insight for policy reform and improvement. They argued that evidence from a wide range of educational systems would be of sufficient variability to permit the revelation of important relationships that would otherwise escape detection.

Foshay *et al.* (1962), the authors of one of the first publications to arise from the work of IEA, eloquently expressed this focus when making the case for international comparisons: "If custom and law define what is educationally allowable within a nation, the educational systems beyond one's national boundaries suggest what is educationally possible" (p. 2). This desire to explore the limits and possibilities of schooling and the quest to identify and understand those factors capable of providing insights into and explanations for differences in outcomes became the key drivers of the design of IEA studies.

IEA Studies

Organization, Principles, and Development

Since its inception in 1958, IEA has conducted more than 24 research studies of cross-national achievement. Each

IEA study is managed through an international coordinating center, which is responsible for carrying out all key international developmental and administrative tasks. International advisory and expert committees guide the countries participating in the respective studies. Their aim is to ensure that the highest quality expertise is brought to bear on key aspects of each study and, in particular, on the more complex issues relating to study design, subject-matter content, and analysis. National research coordinators and their national centers are responsible for conducting the studies within each participating country. All participants contribute to decision making in the respective studies.

Contemporary IEA studies, particularly those with a focus on student achievement at the primary and secondary levels of schooling, are fundamentally concerned with measuring learning outcomes and their antecedents after a fixed period of schooling. This broad organizational principle along with a curriculum model that links the intended curriculum (i.e., what educational policy dictates) with the implemented curriculum (what is taught in classrooms and how) and with the achieved curriculum (what students know and can do) is central to the design and conduct of IEA studies.

IEA has recently moved beyond the study of student achievement and has engaged in a study of the preparation of mathematics teachers. This study, called the Teacher Education Development Study-Mathematics (TEDS-M), examines the antecedents and outcomes of the preparation of mathematics teachers. As for the studies of student achievement, it is organized around the curriculum model described above.

The first IEA study, known as the Pilot Twelve-Country Study, was implemented as a feasibility study and conducted between 1959 and 1962. It examined the performance of 13-year-olds in mathematics, reading comprehension, geography, science, and nonverbal abilities. While many of the findings of this study had practical value, its significance was that it showed that constructing a common set of achievement tests and questionnaires that work cross-culturally is possible and that some of the significant methodological challenges related to language differences can be mitigated through careful translation procedures.

The First International Mathematics Study (FIMS), which followed, with data collection taking place in 1964, explored what 13-year-old students and students in their pre-university year had learned after a fixed period of schooling. In addition to identifying factors influencing the teaching and learning of mathematics, this study introduced and demonstrated the importance of the concept of opportunity to learn for understanding learning outcomes. Moreover, FIMS, in building on the experience of the Twelve-Country Study, established the aforementioned relationship between what is prescribed by policy and the curriculum (the intended

curriculum), what is taught in classrooms (the implemented curriculum), and what students know and can do (the achieved curriculum), as a key design feature of future IEA studies.

The FIMS also raised the question of whether and to what extent factors positively associated with achievement in mathematics could be generalized to other subject-matter areas. This question provided much of the impetus for the group of studies that came to be known as the Six Subject Survey, which examined the performance of 10- and 14-year olds in science, reading comprehension, literature, French as a foreign language, English as a foreign language, and civic education.

The manner in which IEA's studies have evolved since the early years (see **Table 1**) reflects not only the interests of IEA's early pioneers but also the findings and experiences of the first studies and the ongoing shaping effects of a variety of factors across time. The latter include the demands of increased globalization and concerns for greater local accountability, the needs of an increasing and more diversified membership and participant group of countries, and the development of the technology and science related to large-scale assessment.

Funding

During IEA's early history, limited project funding meant that individuals who volunteered their time and expertise carried out much of the association's work. As the demand for greater regularity of studies increased, and the need for higher standards of technical quality and expertise rose, so too did the need to secure more stable financial support. The decision in 1990 to establish a permanent secretariat with an agreement to charge membership and (later) project fees helped secure the financial base of the organization and its projects.

In 1995, IEA decided to make mathematics and science (Third International Mathematics and Science Study (TIMSS), later renamed Trends in Mathematics and Science) its core study. In 1999, reading (Progress in Reading Literacy Study (PIRLS)) was also made a core study. These events, which marked a significant turning point in the development of IEA, were made possible by securing core funding for the studies from the US Department of Education, through the National Center for Educational Statistics, and establishing an international study center at Boston College. The latter assumed overall management of the two series of studies, and in so doing ensured continuity of expertise.

Although the intention was for the balance of project costs to be met through country participation fees, these presented a significant barrier to participation for many developing countries. From 1999 onward, IEA was successful in securing partial support for many developing countries through partnerships with such organizations as

Table 1 Chronological listing of IEA studies

<i>Name of study</i>	<i>When conducted</i>
<i>Completed studies</i>	
Pilot Twelve-Country Study	1959–62
First International Mathematics Study (FIMS)	1963–67
Study of Literature Education ^a	1968–73
First International Science Study (FISS) ^a	1968–72
Study of Reading Comprehension ^a	1968–72
Study of English as a Foreign Language ^a	1968–73
Study of French as a Foreign Language ^a	1968–73
Study of Civic Education ^a	1968–73
Second International Mathematics Study (SIMS)	1977–81
Classroom Environment Study	1980–85
Second International Science Study (SISS)	1982–86
Written Composition Study	1983–88
Reading Literacy Study (RLS)	1985–94
Computers in Education Study (COMPED)	1987–93
Language Education Study (LES)	1993–97
Third International Mathematics and Science Study 1995 (TIMSS 1995) (Re-named Trends in Mathematics and Science Study 1995)	1993–97
Second Information Technology in Education Study Module 1 (SITES-M1)	1997–99
Trends in Mathematics and Science Study 1999 (TIMSS 1999)	1997–2001
Civic Education Study (CIVED)	1994–2002
Pre-Primary Project (PPP)	1986–2003
Second Information on Technology in Education Study Module 2 (SITES-M2)	1999–2002
Progress in International Reading Literacy Study 2001 (PIRLS 2001)	2000–03
Third International Mathematics and Science Study Repeat Video Project	1998–2004
Trends in Mathematics and Science Study 2003 (TIMSS 2003)	2001–04
Progress in International Reading Literacy Study 2006 (PIRLS 2006)	2004–08
Second Information on Technology in Education Study 2006 (SITES 2006)	2005–08
Trends in International Mathematics and Science Study 2007 (TIMSS 2007)	2003–07
Trends in International Mathematics and Science Study Advanced 2008 (TIMSS Advanced 2008)	2006–09
<i>Current studies</i>	
Teacher Education and Development Study in Mathematics 2008 (TEDS-M 2008)	2005
International Civic and Citizenship Education Study 2009 (ICCS 2009)	2006
Progress in International Reading Literacy Study 2011 (PIRLS 2011)	2008
Trends in Mathematics and Science Study (TIMSS 2011)	2009

^aPart of the Six Subject Survey.

the World Bank, the United Nations Development Program (UNDP), the Inter-American Development Bank, and several private foundations.

Transformation and Expansion

Examination of the transformation that has occurred over the years in the design and scope of IEA studies allows us to observe the transition from the early studies, where the focus of the research was primarily descriptive, to the later investigations, where more sophisticated attempts have been made to conduct more explanatory analyses. The successful completion of four cycles of TIMSS and two cycles of PIRLS represents a significant advancement in study design, scaling, data analysis, and reporting.

Whereas in the late 1970s the more tactical approach to participation in IEA studies led to the repetition of studies in mathematics and science on a somewhat *ad hoc* basis, the period at the end of the 1980s and the beginning of the 1990s was characterized by an increasing demand

for regular assessments, particularly in the areas of mathematics, science, and reading. This development reflected the growing recognition that regular monitoring of educational outcomes is a necessary strategic investment in quality assurance. The release of the results of IEA's Reading Literacy Study and, more importantly, the results of the Third International Mathematics and Science Study (TIMSS 1995), contributed significantly to this recognition and also brought to public attention advantages and concerns associated with international competitiveness in the area of education. The TIMSS 1995 was the first of IEA's regular cycle of studies of learning in basic school subjects. Like the PIRLS that followed in 2001, these studies were designed as trend studies – studies that allow the performance of students from different cohorts to be examined for changes over time.

For many developed countries, particularly those that were members of the Organization for Economic Cooperation and Development (OECD), the late 1980s were characterized by an increased concern for greater public

accountability for the outcomes of public schooling. The more general concerns about educational quality were accompanied by the premise that it was no longer sufficient for educational systems to be nationally relevant. Instead, countries needed to develop educational systems that would prepare their respective citizens to meet the demands of global competitiveness and thereby ensure economic and social wellbeing. In many of these countries, which included, for example, Canada, the United Kingdom, and New Zealand, significant public sector and educational reforms accompanied these demands. In this context, the role of assessments, both national and international, were central to the discourse about quality accountability.

As participation in IEA studies has expanded in terms of the absolute number of countries and the diversity of their economic development, IEA has increased its investment in capacity building and training. Training has always been an integral part of IEA projects, but many of the countries with more limited experience in large-scale assessment require additional support. These demands and changing expectations related to the need for more in-depth analysis of data have precipitated several initiatives to facilitate not only the successful completion of studies but also the use of data for policy evaluation and reform. Regional and country-specific training seminars in particular have begun to address training needs that go beyond the specific demands of participation.

In the beginning of the first decade of the new century, IEA introduced two important developments in response to the increasing demand for secondary analysis of IEA study data. The first was the association's establishment in 2004 of a biennial international research conference (IRC). The second, which took place in 2006 in cooperation with the US Educational Testing Service (ETS), was the association's establishment of a joint research institute based at the IEA Data Processing and Research Center in Hamburg. The IEA/ETS Research Institute (IERI) provides twice-yearly training for academics in a broad range of research and analytical areas. The institute also supports and conducts secondary analysis of data (particularly in support of the IRC) and publishes a monograph series with a specific focus on the science of large-scale assessment.

IEA's Impact on Educational Policy, Research, and Practice

The work of IEA, perhaps more than any other organization involved in similar work, has brought into sharp relief the outcomes of schooling. As noted above, the public visibility of the results of the first TIMSS study was arguably a key factor in shaping the public policy debate.

The outcomes of the major IEA studies, including TIMSS, PIRLS, Civic Education, and Second Information Technology in Education Study, have all captured media attention not only in the international media such as the Associated Press, CNN, and USA Today, but also in many – if not most – of the major newspapers of participating countries. One prominent example of how visible the work of IEA has become relates to comments made in 2004 by Alan Greenspan, Chair of the United States Federal Reserve, during his testimony before the Committee on Education and the Workforce of the United States House of Representatives. Greenspan appealed to data from TIMSS to express his concern about educationally low performing sectors of the United States population and the assumed consequent detrimental impact on the United States economy.

Another indication of the influence of IEA studies is that dissemination of study findings is not restricted to OECD countries. Independent evaluations (Elley, 2002; Gilmore, 2005) of the impact of IEA studies in the developing countries that participated in TIMSS and PIRLS note that although public release of information was politically difficult in some of these countries, most, if not all, produced national reports that were released and, in some cases, were the focus of intense public scrutiny.

In more recent years, countries such as England and Norway have used the release of study results, especially those from TIMSS and PIRLS, to stimulate public discussion and debate with respect to educational reform and improvement.

Evidence that the work of IEA has entered public discourse and the public policy debate is but a first step in arguing the case for the marked impact that IEA studies can have on educational discourse and educational policy, particularly that related to systemic and curricular reform. Again to take TIMSS as an example, the release of the TIMSS 1995 data led to countries as diverse as Iceland, Kuwait, New Zealand, Norway, Romania, and South Africa serving as a catalyst for curricular review and change. In Iceland, the information collected during TIMSS resulted in a recommendation (and subsequent implementation) to increase teaching hours for mathematics and science instruction at the elementary school level. In New Zealand, the TIMSS results precipitated the establishment of a taskforce charged with addressing issues that had emerged from the data on science and mathematics education. These issues included the low expectations of success for New Zealand students held by many teachers and parents, underachievement among Maori and Pacific Island students, and teachers' lack of confidence in their ability to teach some aspects of the mathematics and science curricula. In Israel, the periodical, *Education Week*, published a substantial article headed Down in rankings: Israel seeks changes in education,

featuring an account of the reforms being implemented in that country largely as a result of what it had learned from TIMSS.

While TIMSS and PIRLS, focusing as they do on core subjects, tend to attract particular interest from educational policymakers and planners, the outcomes of IEA's many other studies have also made and continue to make important contributions to educational policy and practice. Findings from Second Information Technology in Education Study Module 1 (SITES-M1), for example, drew the attention of policymakers worldwide to at least three major issues. The first concerned the fact that although, in participating countries, the earlier challenge of providing schools with sufficient numbers of computers had generally been met, a critical gap still existed – that of teachers not being sufficiently familiar with and able to use technology for instructional purposes. The second issue was that, with few exceptions, the promise or expectation that computers would transform curricula and pedagogy had not been realized. The third issue centered on the difficulties schools were beginning to have in managing access to the Internet and protecting children from inappropriate materials. The availability of information about the changes that had taken place in terms of the penetration of computers into classrooms also helped change the discourse relative to information technologies in education from questions about, for example, whether to invest in computers to questions about how to facilitate their integration into curricula and instruction. The SITES-M1 showed educational policymakers, planners, and practitioners that while technology has the capacity to shape the teaching–learning environment, efforts to implement it into schools and integrate it in classroom programs tend to create considerable turmoil for schools, teachers, and students.

The findings of the Civic Education Study (CIVED) has also featured strongly in educational policy debate both nationally and internationally. Conducted at a time of significant change in Central and Eastern Europe, CIVED addressed concerns related to the processes of civic education. It asked, for example, such questions as how countries manage the process of preparing students for citizenship and membership of participatory democracies. The findings of this study revealed important differences across and within countries in terms of knowledge about and understanding what is meant by democracy, as well as of student attitudes toward nation, government, immigrants, and women's political rights.

The IEA studies have also had a tangible impact on research acumen. For most countries, participation in various studies and the training opportunities they afford have significantly enhanced ability to carry out research and assessment activities. Independent evaluations (Elley, 2002; Gilmore, 2005) provide compelling evidence of the

positive impact that participation has on developing national assessment capacity. For example, in countries like Macedonia, TIMSS 1995 was the first instance of a national assessment of student achievement being conducted. Furthermore, all countries that participated in TIMSS 1995 were reported (Elley, 2002) to have benefited in some way from the training in large-scale student assessment that accompanies participation in IEA studies. For many countries, direct outcomes of the TIMSS training activities included capacity building. Some countries built on this experience by establishing dedicated research institutes that used the skills and knowledge developed through the TIMSS experience to enhance their national educational and curricular policy and assessment development. Even in the more economically advanced countries, centers of excellence have developed in those institutions responsible for the execution of IEA projects. While perhaps already in possession of a group of skilled researchers, these institutions generally have been exposed to new techniques and technologies as a result of the evolving methodological and analytical sophistication of IEA studies.

The IEA studies (with TIMSS and PIRLS again to the fore) have furthermore been instrumental in spawning a large number of research investigations, both national and international, directed at key policy issues as well as at developing fundamental understandings of subject-based learning (notably mathematics, science, and reading). An increasing body of research involving secondary analyses of data from other IEA studies is also evident, as is research prompted by questions emerging from the findings of the more recent IEA assessments.

The impact of IEA studies on curriculum development and educational practice within schools is, as yet (given time lags between implementation of policy arising out of research findings and evidence of those developments on practice), less strongly articulated. However, several countries (notably Australia, Canada, Spain, Japan, and the Philippines) used TIMSS 1995 results to bring the teaching of mathematics and science into sharp focus. Canada (Ontario), for example, developed instructional materials as an outcome of its participation in the study. In the Philippines, government resolve was reflected in a program to train 100 000 mathematics and science teachers over a period of 5 years.

In other countries, TIMSS findings have pointed to particular practices and curricular resources in need of reconsideration, remedy, or recognition. In England, for example, the results of TIMSS focused attention on lower-achieving students and were also the catalyst for a guidance document that drew attention to the frequency with which students were using calculators and accordingly stressed the importance of having students do mental calculations. In Iran, the TIMSS results led to major changes to the science curriculum and textbooks. In New Zealand,

information from TIMSS contributed to the development of resource materials and professional development programs (designed to address the perceived areas of relative weakness) for mathematics and science teachers. In the United States, the TIMSS findings saw the structure and content of the United States approach to curriculum coming under intense scrutiny and debate, a state of affairs that continues today.

The PIRLS has also forced examination and reappraisal of reading-related teaching practices and curriculum content and guidelines. The data from the PIRLS questionnaire for parents, for example, highlighted, for nearly all participating countries, the importance of parental involvement in the reading acquisition process and raised the question of how parents might be better involved in their children's learning to read.

The SITES assessments have provided important insights into the role that technology does and can play in teaching and national curricula. As noted above, the first module of the study identified the gap between the provision of technology and teacher readiness in the use of that technology for instructional purposes. In Norway – a country that has invested significantly in the provision of computer technology – this gap was a particular area of concern.

Summary

The growth and development of IEA has been stimulated by ongoing and growing concerns related to public accountability in education and the imperative for improved quality of educational outcomes associated with increased globalization. During the last 10–15 years, IEA has experienced a significant increase in both its membership and in terms of participation in its studies. As reflected by the change in focus of the organization over the last 50 years, this changing membership has placed increasing demands for training and support for secondary analysis of project data. It is through its commitment to make educational reform and improvement a primary goal that IEA's influence can be seen and measured. The influence is clearly evident not only in terms of the impact that the studies have had on the public discourse on educational quality but also in terms of the very tangible impact that IEA studies have had on educational policy, research, practice, and curriculum.

See also: Comparative Education: Societies and Associations; IEA Studies in ICT; IEA Studies in Mathematics and Science; IEA Study in Civic Education; NGOs and Globalization of Education.

Bibliography

- Elley, W. B. (2002). Evaluating the impact of TIMSS-R (1999) in low- and middle income countries. *An Independent Report on the Value of World Bank Support for an International Survey of Achievement in Mathematics and Science. Report to the IEA (Amsterdam) and the World Bank*, Washington, USA.
- Foshay, A. W., Thorndike, R. L., Hoyt, F., Pidgeon, D. A., and Walker, D. A. (1962). *Educational Achievement of Thirteen-Year-Olds*. Hamburg: UNESCO Institute for Education.
- Gilmore, A. (2005). *The Impact of PIRLS (2001) and TIMSS (2003) in Low and Middle Income Countries*. Washington DC: World Bank.

Further Reading

- Burkhardt, H. and Shoenfeld, A. H. (2003). Improving educational research: Toward a more useful, more influential, and better-funded enterprise. *Educational Researcher* 32(9), 3–14.
- Degenhart, R. E. (ed.) (1990). *Thirty Years of International Research: An Annotated Bibliography of IEA Publications (1960–1990)*. The Hague: International Association for the Evaluation of Educational Achievement.
- Down in rankings: Israel seeks changes in education (2004). *Education Week*. <http://www.edweek.org> (accessed August 2009).
- Greenspan, A. (2004). *Address to the Committee on Education and the Workforce, United States House of Representatives March 11, 2004*. <http://www.federalreserve.gov/boarddocs/testimony/2004/20040311/default.htm> (accessed August 2009).
- Keeves, J. P. (1995). *The World of School Learning: Selected Key Findings from 35 Years of IEA Research*. The Hague: International Association for the Evaluation of Educational Achievement.
- Osborne, D. and Gaebler, T. (1993). *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*. New York: Plume.
- Pelgrum, W. J. and Anderson, R. E. (1999). *ICT and the Emerging Paradigm for Lifelong Learning*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Postlethwaite, T. N. (1974). Introduction. *Comparative Education Review* 18(2), 157–163.
- Robitaille, D. F. and Beaton, A. E. (2002). *Secondary Analysis of the TIMSS Data*. Dordrecht: Kluwer.
- Robitaille, D. F., Beaton, A. E., and Plomp, T. (2000). *The Impact of TIMSS on the Teaching and Learning of Mathematics and Science*. Vancouver: Pacific Educational Press.
- Schmidt, W. H., McKnight, C. C., Cogan, L. S., et al. (1999). *Facing the Consequences: Using TIMSS for a Closer Look at United States Mathematics and Science Education*. Hingham, MA: Kluwer.
- The World Bank (1999). *Education Sector Strategy (Report # 79)*. Washington DC: The International Bank for Reconstruction and Redevelopment World Bank.
- Torney-Purta, J., Lehmann, R., Oswald, H., and Schulz, W. (2001). *Citizenship and Education in Twenty-Eight Countries: Civic Knowledge and Engagement at Age Fourteen*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Tuijnman, A. C. and Postlethwaite, N. T. (1994). *Monitoring the Standards of Education*. Oxford: Pergamon.

Relevant Website

http://www.iea.nl/brief_history_of_iea.html – International Association for the Evaluation of Educational Achievement.

INTERNATIONAL ORGANIZATIONS IN EDUCATION

Contents

Education and the European Union

National University on Educational Planning and Administration, New Delhi

On School Management – International Council on School Effectiveness and Improvement, Commonwealth Council on Educational Administration and Management

UNESCO's International Institute for Educational Planning

Education and the European Union

S Hegarty, International Association for the Evaluation of Educational Achievement (IEA)

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The European Union (EU) is a partnership of some, though not all, European countries. Its mission is to “provide peace, stability, and prosperity” for its citizens. Founded as the European Coal and Steel community with six countries in 1951, it currently comprises 27 member countries with a number of candidate countries seeking admission. The three countries of the European Economic Area and Switzerland also participate in some EU programmes. The intervening decades have seen a succession of treaties and constitutional developments. These have been matched by a growth in the EU's sphere of operations: it has a substantial budget (€129 billion – about 1% of EU Gross National Income – in 2007) and concerns itself with a wide range of policy matters in the economic, social, regulatory, and financial areas. It has major programmes in agriculture, energy, the environment, and research. It launched the euro as a common currency in 2002; this is now used in 16 member countries (plus some others outside the EU) and is already established as a major world currency.

Support for research is organized within framework programmes which run for a number of years. The current programme, which is the seventh, has a budget of €50 billion for the period 2007–13. The topics covered include energy, food and agriculture, information and communications technology (ICT), the environment, transport, security, space, and the socioeconomic sciences.

The European Commission (EC) is the executive arm of the EU. Political authority is vested in the Council of Ministers as the top-level decision-making body and the European Parliament which is directly elected across all member states every five years, but operational responsibility falls to the EC. The Commission currently comprises one Commissioner from each of the 27 member

states. It is assisted by a civil service made up of 46 directorates-general (DGs) and services, which are mainly based in Brussels and Luxembourg.

Education and Culture is one of the Commission's 46 DGs. Its mission is “To reinforce and promote lifelong learning, linguistic and cultural diversity, mobility and the engagement of European citizens, in particular the young”. It engages in a range of actions to improve the quality of learning systems and provide learning opportunities for people at all stages of their lives. While each member state is in charge of its own education and training system, the Commission has the specific role of supporting co-operation between countries so as to achieve common goals in education and training. Its work focuses on two areas in particular: cooperation with national authorities and European stakeholders in improving policies and exchanging good practice; and the development and administration of funding programmes.

Early Educational Cooperation in Europe

The history of educational cooperation in Europe is a checkered one. Despite the much-quoted dictum of Jean Monnet, one of the founding fathers of European cooperation, “If I could start again, I would begin with education,” education did not feature at all in the early years of the European Community. Indeed, one author describes the time from 1957 to 1969 as “a period when education remained a taboo subject within the corridors of the European Community” (Neave, 1984: 6). This was because of a visceral attachment to curriculum, school organization, and so forth at national level, on the one hand, and a widespread belief, on the other, that action in the political, economic, and labor market spheres was both more necessary and more likely to be efficacious.

The beginnings of European cooperation in education were, in fact, achieved outside the European Community, in particular, through the Council of Europe. The latter is quite separate from the EC, having been founded in 1949 and comprising 47 member countries. The Council of Europe was in fact the main player in European cooperation in education for many years, because of its early commitment to education, the active networks of policy makers, experts, and nongovernmental organizations it facilitated, and the fact that the European Community did not at the beginning concern itself with education. From its early years it stimulated important initiatives in areas such as language learning, research and documentation, and adult and higher education.

Despite some structural similarities with the EC – it has a Council of Ministers and a Parliamentary Assembly – the Council of Europe has a distinctive set of programmes (and a very much smaller budget). Its principal concerns are with human rights, the rule of law, democratic development, and intercultural cooperation. In education, the initial focus was on information sharing through activities such as EUDISED (European Documentation and Information System for Education), a multilingual directory of ongoing and completed projects of European educational research and development (R&D), conferences, and expert meetings. For many years, it convened regular meetings of the directors of European educational research institutions. Its current emphases are on education for democratic citizenship and human rights, language learning, intercultural education, history teaching, aspects of higher education, and the education of Roma/gypsy children.

The practice of collaboration and the associated networks established by the Council of Europe provided a fertile seed bed for the more substantial programmes that the EC's greater political and budgetary clout made possible. The two bodies collaborate in areas of common interest such as higher education, language learning, and the promotion of democratic citizenship. Thus, they jointly organized the European Year of Languages 2001 and collaborate in supporting the Bologna process aimed at creating a European higher education space.

Emergence of Educational Cooperation in the EU

The first meeting of ministers of education within the EU took place in 1971. This was followed 2 years later by the establishment of an administrative division specifically responsible for education. The first Community-wide programme in the field of education was agreed in 1976 and encompassed six fields for priority action:

- education of migrant workers and their children
- closer relations between educational systems in Europe
- compilation of up-to-date documentation and statistics

- cooperation in higher education
- teaching of foreign languages,
- equal opportunities.

While the rationale for these action fields was clear, if not compelling, in the context of greater European cooperation, progress was slow for a number of years. This was partly because of budget constraints – less than 0.1% of the Community budget was committed to the programme – and partly because of political sensitivities and (some) countries' opposition to anything that smacked of European harmonization of national educational systems. The aim, therefore, was to share information and strengthen mutual understanding through pilot projects, studies, and study visits.

Despite this low-key start, some progress was made. With regard to migrant workers, whose children numbered an estimated 3.5 million in EC schools in the 70s, a Directive adopted in 1977 required all member states to provide mother tongue teaching. This led to new pedagogical approaches to language learning, and, in later years, to a focus on the development of intercultural education.

An information network on education in Europe, Eurydice, was established in 1980 to share information on policies and systems in a structured way. Along with Eurostat, established in 1978 to collate and publish statistical data on education, it continues to be a key tool in creating links between European education systems. Now established as part of the EC's umbrella Lifelong Learning Programme, it provides a wide range of information to policy makers in education at European, national, and local levels. At the European level it maintains a range of databases that give comprehensive descriptions of European education systems and conducts studies to provide comparative information on topics of current interest. It is supported in this work by a network of national units which provide inquiry and current awareness services. All of its information is provided free of charge.

Despite some work on student mobility and mutual recognition of qualifications, progress in higher education was modest. Likewise with foreign languages, there was little progress until the Lingua programme in 1989, though the European Bureau for Lesser Used Languages was established in 1982 to support regional and minority languages within the EU. Regarding equal opportunities, the main thrust was the large programme on preparation for working life; explicit attention to gender issues was delayed till 1985 and to disabilities and learning difficulties till 1987.

Parallel developments were taking place in vocational training. Structures and activities similar to those found in education were created. While there was some linkage with education, for example, in relation to the programme on preparation for working life, the emphasis was firmly on the labor market. Key drivers of activity were the growing levels of youth unemployment and the need to improve the technological skills of the workforce.

The late 1980s saw significant developments, which culminated in 1992 in the establishment for the first time of a legal basis for European Community cooperation in education. These were driven, in part, by the appointment of Jacques Delors as president of the EC in 1985 and, in part, by juridical rulings. Delors' success in relaunching the social dialog between employer and trades union representatives went along with a greater recognition of the importance of education and training. This led to a strengthening of administrative support for education and training activities within the Commission, culminating in the establishment of a separate Directorate-General for education in 1995, which became the current DG for Education and Culture in 1999.

A judgment by the European Court of Justice in 1985 (that all European students were entitled to undertake vocational training in any country on the same basis regardless of their nationality; and, very importantly for Commission purposes, that the interpretation of vocational training should be expanded to include higher education) opened the way for substantial programmes in education and training, as described below. In particular, it created the momentum which would lead to the establishment of a legal basis, in Articles 126 and 127 of the Maastricht Treaty of 1992, for European cooperation in education and training. These articles allowed the European Community to take actions and commit resources to 'contribute to the development of quality education' and to 'implement a vocational training policy'. Apart from being an important symbolic gesture which liberated education and training from the uncertain status they had long endured in the EC, this enabled substantial action in the following years.

The major programmes rolled out in this period were: Comett (university-business cooperation); Erasmus (student mobility and cooperation between universities); Tempus (university cooperation with central and eastern Europe); PETRA and Youth for Europe (initial training for young people and youth exchanges); FORCE (continuing vocational training); and Lingua (language learning). These programmes became the main thrust of Community cooperation in education and training between member states. In comparison with previous activities, they were highly resourced and some had a substantial impact.

Education and Training Post-Lisbon

The current framework for education and training in the EU derives from a wide-ranging economic and social programme agreed by Heads of State meeting in Lisbon in 2000. The so-called Lisbon Strategy set the target for 2010 that the Union "should become the most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion" (European Council, 2000). The

economic dimension remained paramount but it was now accompanied by an explicit recognition that the economic goals would not be met without greater attention to the modernization of education (and social welfare) systems. Indeed, the Council was to assert in 2002 that Europe should be the world leader in terms of the quality of its education and training systems.

This new recognition of the role of education and training led ministers of education to adopt a report on the future objectives of education and training systems (European Commission, 2002a) and a 10-year work programme (European Commission, 2002b), now commonly known as the Education and Training 2010 process. This covered school education, vocational education and training, higher education and lifelong learning, and marked the first time that ministers had agreed on a common set of objectives in education across the EU.

Countries retain control over their education and training systems but much is made of the so-called 'open method of coordination' as a means of achieving common goals. This is a governance process that depends on persuasion and voluntary agreement as opposed to regulation and the use of sanctions. It reflects the 'soft' approach to governance characteristic of much of the work of bodies such as the United Nations and the Council of Europe, but in its EU version the process is relatively more developed and data-driven. Broadly, there are a number of stages: strategic goals and objectives are agreed at a senior political level; these are then translated into appropriate policies at national and regional levels; indicators by which performance can be measured in respect of these are developed; and statistical information is collected in the light of these so that countries can compare themselves with EU averages and with specific other countries. Besides statistical data, information on good practice is also gathered and disseminated. The process relies, therefore, on information gathering and the publication of information – and the associated public pressure – rather than on regulation or other legal instruments.

The goals agreed at the political level were necessarily rather general, and are concerned with issues of effectiveness, quality, and access. These were developed into 13 objectives: improving the training of teachers and trainers; developing key competences; ensuring ICT access for everyone; increasing the number of science and technology graduates; making best use of resources; creating an open learning environment; making education and training more attractive; working on active citizenship, equal opportunities, and social cohesion; strengthening links with the worlds of work, research, and society; developing the spirit of enterprise; improving foreign language learning; increasing mobility and exchanges; and strengthening European cooperation.

The work programme that flowed from this marked progress in two significant respects. First, it capitalized on the new recognition given to education and training, and

went further – by asserting that education and training were important and deserving of investment in their own right, and not simply as instruments for employability. This led to an explicit focus on matters such as social skills, citizenship, and intercultural understanding. Second, it introduced regular monitoring of performance and progress: indicators and benchmarks were established to provide a framework for detailed information gathering across EU countries.

The initial proposal was to base monitoring on 29 indicators and five benchmarks agreed in 2003. Annual reports were produced from 2004 onward to set out the evidence on progress and provide guidance on the progress of the Education and Training 2010 Programme. As a result of these early reports, the framework was modified and in 2007 a new framework comprising 16 core indicators was agreed.

The five EU-level benchmarks agreed in 2003 continued to be relevant, not least because of the uneven progress toward them. They are:

- No more than 10% of early school leavers
- Decrease of at least 20% in the percentage of low-achieving pupils in reading literacy
- At least 85% of young people to have completed upper secondary education
- Increase of at least 15% in the number of tertiary graduates in mathematics, science, and technology, with a simultaneous decrease in the gender imbalance
- 12.5% of the adult population to participate in lifelong learning.

The 16 core indicators agreed in 2007 are concerned with the following:

1. Participation in preschool education
2. Special needs education
3. Early school leavers
4. Literacy in reading, mathematics, and science
5. Language skills
6. ICT skills
7. Civic skills
8. Learning to learn skills
9. Upper secondary completion rates of young people
10. Professional development of teachers and trainers
11. Higher education graduates
12. Cross-national mobility of students in higher education
13. Participation of adults in lifelong learning
14. Adult skills
15. Educational attainment of the population
16. Investment in education and training.

The 2008 report (European Commission, 2008b) indicates how challenging these benchmarks are for education and training systems in Europe. The report draws on a rich array of statistical data from EU agencies and beyond

(sources also include IEA – the International Association for the Evaluation of Educational Achievement and OECD – the Organization for Economic Cooperation and Development) to indicate the directions in which European educational systems are moving and where they stand in relation to meeting the EU's Lisbon objectives.

The core message is that, while the provision of education and training in Europe is improving and some aspects match the best in the world, performance across Europe is very uneven and most of the benchmarks developed within the Education and Training 2010 programme will not be met by 2010.

Nine countries (Finland, Denmark, Sweden, the United Kingdom, Ireland, Poland, Slovenia, Norway, and Iceland) exceed the composite objective of the five benchmarks set for 2010 and are progressing well. Other countries, however, fall below one or more benchmarks and are not on track to meet the targets. Of the individual targets, the only one that has been met is that relating to increasing the number of graduates in mathematics, science, and technology (though not in respect of decreasing the gender imbalance): all countries are increasing the number of graduates relative to 2000, and the majority are already close to or above the 2010 target.

Regarding early school leavers, eight countries have achieved the benchmark (of not more than 10%) but the EU as a whole falls below this with 14.8% of early school leavers in 2007. Eleven countries reached the benchmark of 85% upper secondary completion rate, although again the average EU rate falls below this; at 78.1% it has increased little from 2000.

Literacy levels represent the most challenging target area. Only ten countries have hit the target (and three of these are losing momentum), and the EU as a whole is falling back from the target of a 20% reduction in low achievement in literacy. On the measure used, 24.1% of 15-year-olds were deemed to have had low literacy levels in 2006, an increase over the figure of 21.3% in 2000. These figures mask very wide discrepancies across Europe, from fewer than 5% of low performers in reading in Finland to more than 50% in some countries.

Nine countries have reached the benchmark of 12.5% participation in lifelong learning. Despite the high levels of participation in some countries – more than 25% in Sweden, Denmark, and the United Kingdom – the EU average is just below 10%, and the 2010 target is unlikely to be met.

Following consultation with member states and others in 2008, the Commission issued a Communication in 2008 (European Commission, 2008a) reflecting on the achievements to date, outlining urgent priorities in the short term and looking to the long term. The document proposes four strategic challenges for the years to 2020: make lifelong learning and learner mobility a reality; improve the quality and efficiency of provision and outcomes;

promote equity and active citizenship; and enhance innovation and creativity, including entrepreneurship, at all levels of education and training.

Lifelong Learning Programme

The EC's various spending programmes in education and training have been pulled together into a single overarching programme, the Lifelong Learning Programme. This is scheduled to run from 2007 to 2013 and is resourced with a budget of €7 billion. This replaces or subsumes all previous programmes in education, vocational training, and e-learning and aims to support individuals at all stages of life to pursue stimulating learning opportunities across Europe. The Programme has four sub-programmes focusing on the different stage of learning.

Comenius

The Comenius programme is concerned with the first phase of education, from early years and primary education to secondary schools, and is open to staff and parents as well as pupils. It seeks to develop knowledge and understanding among young people and staff of the diversity of European cultures, languages, and values. It promotes partnerships between schools in different countries and facilitates the mobility of pupils and staff. It has particular objectives in relation to language learning, innovative use of ICT and the European dimension of teacher education.

Target: Comenius should involve at least 3 million pupils in joint educational activities over the life of the project.

Erasmus

Erasmus is the EC flagship programme in higher education. Its initial *raison d'être* was to support student mobility, and 1.9 million students – from 90% of European universities – have participated since it started in 1987. In addition to student mobility, it supports staff travel and professional development, facilitates networking, and underpins multilateral projects. Erasmus aspires to be the main driver of the modernization of European higher education and indeed to help create a European higher education area.

Target: Erasmus should reach a total of 3 million individual participants in student mobility actions by 2012.

Leonardo da Vinci

The Leonardo da Vinci programme links policy to practice in the field of vocational education and training. Projects range from those giving individuals the chance to improve their competences, knowledge, and skills through a period abroad, to Europe-wide cooperation between training organizations. The programme enables training organizations to work with European partners to exchange best practice and increase the expertise of their staff, thereby helping to make vocational education more

attractive to young people. By enabling European citizens to acquire new skills, knowledge, and qualifications, the programme also aims to bolster the competitiveness of the European labor market.

Target: Leonardo da Vinci should increase placements in enterprises to 80 000 a year by 2013.

Grundtvig

The Grundtvig programme focuses on the teaching and study needs of those in adult education and alternative education streams, as well as the institutions and organizations delivering these services. In addition to the learners, it also covers the teachers, trainers, education staff, and facilities. These include relevant associations, counseling organizations, information services, policy-making bodies and others involved in lifelong learning and adult education at local, regional, and national levels. There are also actions supporting adult learner mobility such as exchanges and, in the 2007–13 programme, 'European Assistantships,' as well as initiatives to support collaboration between adult education providers.

Target: Grundtvig should support the mobility of 7000 individuals involved in adult education each year by 2013.

Transversal programme

While these four sub-programmes are directed at different stages of education, they have a number of elements in common. A crosscutting or 'transversal' programme targets these common elements in order to ensure that the sub-programmes achieve the best results possible. This has four key activities:

1. *Policy cooperation:* This activity has two strands: study visits; and studies and comparative research. It is open to policy makers, experts and a wide range of education practitioners. Priorities for support are decided on an annual basis. For 2008, priorities for visits include: educational systems and their values; school environments; vocational education and training; adult education; and language teaching and learning. Studies and comparative research are intended to strengthen the evidence base for policy and practice in education and training. Current priorities include: the promotion of excellence in higher education, especially student access and retention; quality in adult learning; weaknesses in early years and compulsory education, particularly in the acquisition of key competences; promoting vocational education and training; and the role of creative activities in the learning process.
2. *Languages:* This activity aims to raise awareness of the importance of linguistic skills, boost access to language learning resources and develop teaching materials. It has two strands that operate across the different stages of education: multilateral projects to develop new language

learning materials, including online courses, instruments for language testing, the promotion of language awareness, and access to language learning resources; and networks to promote languages policies and disseminate good practice. Any language is eligible for funding under this activity, including European regional and minority languages. Other initiatives include publicity campaigns, conferences, studies, and the development of statistical indicators.

3. *ICT*: Effective use of ICT is seen as critical to learning in each of the sub-programmes, and this activity seeks in general to develop innovative educational practices, improve access to lifelong learning, and help develop management systems. It is not concerned with developing the technology itself but rather with the use of ICT tools to enhance learning. At present, priorities include: multilateral projects to encourage innovation in learning and teaching and boost the use of new ICT tools and trends, particularly for groups at risk of exclusion, such as early school leavers, ethnic minorities, and elderly people; and networks to promote greater linking up between learning communities, and foster creativity through the use of ICT.
4. *Dissemination*: All project findings arising from the Lifelong Learning Programme should be made as widely known as possible to potential users. This activity is concerned with boosting dissemination and exploiting initiatives by funding complementary work. The latter could include: improving accessibility of project results; actions to translate project results into policy; using networks to exploit results; developing models and practical tools for sharing and exploiting project results; and supporting think tanks geared to user needs.

Jean Monnet programme

A final strand within the Lifelong Learning Programme is the Jean Monnet programme. Originally launched in 1990, it seeks to stimulate excellence in teaching and research on European integration in higher education institutions throughout the world. The programme is concerned with the construction of the European Community and its institutional, legal, political, economic, and social development. It has projects in 61 countries across five continents and reaches up to 250 000 students a year. The programme supports three distinct activities: university-level projects on European integration, including university chairs, centers of excellence, and research activities; six specific academic institutions pursuing an aim of European interest, such as the College of Europe and the European University Institute; and associations active at the European level (in at least 12 member states) in the field of education and training. Other activities include Jean Monnet conferences and thematic groups focused on topics of contemporary political interest.

Conclusion

The European project in education is an extraordinary one. Given the historical divisions, fierce attachment to national educational systems, and the enormous diversity across Europe, it is remarkable that such a comprehensive set of programmes has secured political and fiscal support. The Lisbon process is still relatively young, however, and many challenges remain to be addressed. Some of these are sketched below in terms of dilemmas that remain to be resolved.

The first tension is the traditional one between education for economic competitiveness and education for personal fulfillment and social benefit. While European rhetoric has long acknowledged the imperatives of democracy and peace, the principal driver of EC action has been the push for economic development. Education, when considered, was viewed in its relation to the labor market. The Lisbon process accords an important role to personal and social goals, and the explicit regard for active citizenship and intercultural understanding bodes well. The main thrust of action, and funding, however, remains strongly on education as preparation for the labor market. This tendency may well increase if, as seems likely, the economic recession of 2008 lasts for some time. If this were to happen and the EU reverted to being mainly an economic and labor market player, it is difficult to see the broader vision of Europe being realized.

A further tension relates to the place of Europe, and European countries, in a global context. The attachment to Europe as a guiding concept is challenged by some countries' orientation toward other parts of the world, for example, Britain toward the USA (and the Commonwealth), Portugal and Spain toward Latin America. (These are the most obvious examples, but there are many others such as the orientation of many countries in north-west Europe toward North America.) There may well be economic and cultural advantages for Europe in the links, say, between Portugal and Brazil, but there is no gainsaying the fact that countries are not alike in their understanding of or attachment to the European ideal. This plays out in schools and educational systems in relation to differences in curriculum, language learning, school and system organization, and core values.

Language brings many of these issues to a head. There is a strong tension between the well-known educational benefits of providing school instruction in the mother tongue and the pragmatic demand for competence in a regional language. (For many, the latter will be English but, depending on the part of Europe, it could be Russian, German, or French.) The EC's principled commitment to linguistic diversity requires it to spend considerable resources on Europe's many languages. (The total cost of translation and interpretation within EU institutions across the 23 official languages was

estimated at €1.1 billion in 2006 – close to 1% of the total EU budget (Europa, 2006). To this must be added the substantial expenditure on language learning within the Lifelong Learning Programme.) This is in sharp contrast with those many commercial enterprises who conduct their business in English regardless of whether they are in Barcelona, Berlin, or Birmingham. Cost apart, Europe's linguistic diversity raises extremely difficult issues relating to national (and sub-national) identity and the extent to which it is dependent on language, cultural participation and exchange, personal mobility, the emergence of technocratic elites who have mastery of a regional language, and so on.

The concept of a European educational space is complex, if not problematic. This is not simply because of the extraordinary diversity residing in European families and schools. The diversity is real, and a challenge, but the bigger difficulty is in developing this concept in sufficient detail for it to be useful *and* securing agreement to it. The many actions of the Lifelong Learning Programme and the structures furnished by the open method of coordination are undoubtedly giving many Europeans a better understanding and appreciation of each other, and helping education and training systems to benefit from each others' good practice. This is some way short of a European educational space, however. It may be a necessary step along the way but there is still a distance to travel.

Bibliography

- Europa (2007). *Multilingualism: More Translators and Interpreters for the Three New Languages are Arriving*. (Press release 23 February 2007.) Brussels: Europa.
- European Commission (2002a). Education and training in Europe: Diverse systems, shared goals for 2010. *The Work Programme on the Future Objectives of Education and Training*. Brussels: EC, Education and Culture DG.
- European Commission (2002b). *Detailed Work Programme on the Follow-Up of the Objectives of Education and Training Systems in Europe, OJ C 142*. Brussels: EC, Education and Culture DG.
- European Commission (2008a). An updated strategic framework for European cooperation in education and training (*Communication from the Commission to the European Parliament and Other Bodies*, 18.12.08). Brussels: EC.
- European Commission (2008b). Progress towards the Lisbon objectives in education and training. *Indicators and Benchmarks – 2008*. Brussels: EC, Education and Culture DG.
- European Council (2000). *Conclusions of the Lisbon European Council*, 23 and 24 March 2000. Brussels: EC.
- Neave, G. (1984). *The EEC and Education*. Trentham: Trentham Books.
- European Communities (2006). *The History of European Cooperation in Education and Training. Europe in the Making – an Example*. Luxembourg: Office for Official Publications of the European Communities.

Further Reading

Relevant Websites

- <http://ec.europa.eu> – EC Education and Training Policies and Programmes, European Commission.
- http://ec.europa.eu/education/lifelong-learning-programme/doc78_en.htm.
- <http://www.coe.int> – Education in the Council of Europe, Council of Europe.
- <http://ec.europa.eu> – European Union and the European Commission.
- <http://ec.europa.eu> – The 2008 Report on Progress Toward the Lisbon Objectives, European Commission.
- <http://ec.europa.eu> – Translation and interpretation in the EC, European Commission.

National University on Educational Planning and Administration, New Delhi*

J B G Tilak, National University of Educational Planning and Administration, New Delhi, India

© 2010 Elsevier Ltd. All rights reserved.

The National University of Educational Planning and Administration (NUEPA), New Delhi, India, is a specialized public institution of higher education that is involved in research, training, consultancy, and teaching in the area of educational policy, planning, and administration, operating at the national and international levels.

The Genesis and the Growth

The National University assumed its present form and name after 45 years of existence, initially as a training institution and later as a research and training institute in the area of educational planning and administration. Its origin dates back to 1962 when it was established, in pursuance of a resolution passed at the Eleventh Session of the General Conference of the United National Educational, Scientific and Cultural Organization (UNESCO) to help the member states in the Asian region. The member states were: Afghanistan, Burma (now known as Myanmar), Ceylon (now known as Sri Lanka), Republic of China, India, Indonesia, Iran, Japan, Khmer Republic (now known as Cambodia), Republic of Korea, Laos, Malaysia, Mongolia, Nepal, Pakistan, the Philippines, Singapore, Thailand, and Republic of Vietnam. It was established with an objective of preparing the key personnel needed for implementation of the Karachi Plan for the provision of compulsory primary education of at least 7 years' duration by 1980. It was established as the Asian Regional Centre for Educational Planners, Administrators and Supervisors by the UNESCO in February 1962 under an agreement signed between the UNESCO and the Government of India. The main functions of the Asian Regional Centre for Educational Planners, Administrators, and Supervisors were to conduct research on problems connected with educational planning, administration, and school supervision, to conduct short-term training programs for educational planners, administrators, and supervisors in Asia, and to offer technical assistance to member states. Besides launching short-term, including 3-month, training programs for educational planners and administrators, one of the important outputs of this organization was a report of long-term educational projections for India, prepared in 1963. With effect from 1 April 1965, the Asian Regional Centre for Educational Planners,

Administrators, and Supervisors was renamed as the Asian Institute of Educational Planning and Administration and continued to be funded primarily by the UNESCO. The institute functioned under the legal and constitutional framework of the Government of India and its director and other staff were appointed by the Government of India in consultation with UNESCO. The Asian Institute provided, *inter alia*, valuable research inputs to the Indian Education Commission, constituted by Government of India 1964. The Commission, chaired by Dr. D.S. Kothari, submitted its voluminous report in 1966 (Education Commission, 1966), which formed the basis for the first National Policy on Education (1968). Several recommendations of the Commission are found to be still relevant for educational development in India.

At the end of the 10-year agreement between the UNESCO and the Government of India, the Asian Institute was taken over by the Government of India and subsequently the National Staff College for Educational Planners and Administrators was established in December 1970. The institute was registered as an autonomous institution in 1970, and was re-registered as National Institute of Educational Planning and Administration (NIEPA) on 31 May 1979. The primary focus of attention of the institute changed in a sense, from Asian region to India, and more importantly from being a staff college more intended for training, to an academic institution with an added emphasis on research. Compared to the earlier days, the research agenda of the organization after it became the institute in 1979, was widened and deepened as well and the institute was begun to be seen almost for the first time as a major research institution. Immediately after it became a national institute, it launched a series of studies on surveys in educational administration in various states in India, and a large number of research studies on various other issues. Although it had been an important objective to help the Government of India, as the National Staff College or the National Institute, and it was mandated, besides conducting and promoting research, to provide services to the union (central) Government of India and state governments in India and other government bodies, agencies, institutions, and personnel engaged in educational planning and administration, it continued to meet training, research, and consultancy needs of other countries and international organizations as well.

In recognition of the important work being done by the organization in the area of educational planning and

* The views expressed here are personal and are not necessarily shared by the organization with which the author is associated.

administration, the institute was conferred the status of a university in 2006 – an institution deemed to be university under Section 3 of the University Grants Commission Act, 1956. Thus, it traversed a long distance over a period of four-and-a-half decades from a UNESCO regional center to a national university and grew from strength to strength. Thus, it is a young university, although an old research and training institution.

Nature and Activities

Some of the special characteristics of this institution are worth noting. First, with an exclusive focus on educational policy, planning, and administration, the institute is widely recognized as a unique institution of special significance. There is no comparable institution of the kind in India so exclusively focused in an area and, also at the same time, performing four different functions, namely, research, training, consultancy, and teaching in addition to dissemination of information; and very few exist in other neighboring countries of the Asian region. Even as a university, it is visualized as a single faculty research university with students enrolled in doctoral (and predoctoral) programs only. Second, the institute considers education as a holistic system and it recognizes that all levels of education – primary, secondary, and higher – are important and interlinked branches of the educational edifice. Accordingly, the institute's activities focus on all levels as whole as well as each level separately. Third, it also considers educational planning as an integral part of overall development planning. This is true from the very beginning. The Asian Institute, for example, emphasized the need to study educational planning as an integral part of overall planning and to visualize educational development with an integrated approach. Fourth, the institute recognizes the importance of multilevel planning in education, and accordingly focuses on planning at central and decentralized levels – global, regional, national, and subnational – state (provincial), and below state levels. The institute's national diploma program of 6 months duration, introduced in 1982 has helped a lot in capacity building at local (district) levels and helped in strengthening mechanisms of decentralized planning in education. For planning to be effective, the flow of information has to be in both directions – from the lowest unit to the highest (central) level, and vice versa. Fifth, it also notes that educational planning means long-term planning. Several activities of the institute with respect to training and research focused on long-term planning – developing methodologies for long-term planning and research on long-term planning in education. Sixth, although it is specialized in the area of educational planning and development, it views educational planning as an interdisciplinary subject both for research and planning. Accordingly, its faculty are drawn from various social sciences – economics,

sociology, anthropology, psychology, education, politics, law, history, management, etc., besides statistics, mathematics, and computer sciences – but deeply interested in issues related to educational policy, planning, and administration. The recently introduced Doctoral program is also multidisciplinary and students are selected from various disciplines primarily in social sciences. Finally, although it is mostly funded by the Government of India, it is registered as an autonomous organization.

The present functions of the institute also reflect the special nature of the organization. Research, training, better known as capacity building, consultancy, essentially to the governments, and teaching are the four important functions of the organization. Research is largely, but not exclusively, policy oriented and empirical in nature. The institute's capacity-building activities involve training the educational policymakers, planners, and administrators at all levels of education – ranging from school-level principals or headteachers, to the top-level civil servants in the Union government, majority being administrators and planners at state (provincial) level; coverage ranging from primary school education managers to university vice-chancellors, registrars, etc. Although most of the trainees are from the state sector, private-sector people are not necessarily excluded. Training programs also meet the needs of many developing countries. The organization provides the much-needed technical and professional assistance to the government at the central and state levels and other public bodies in India, such as the University Grants Commission, All-India Council for Technical Education, Planning Commission, Finance Commission, State Councils of Higher Education, and the Central Advisory Board of Education, by providing research and professional inputs into the formulation of educational policies (e.g., National Policies on Education 1968, and 1986, 1992), plans (e.g., 5-year plans formulated by the Planning Commission), and other development programs, projects, and schemes (e.g., Operation Blackboard and the scheme of mid-day meals). The institute's contributions did change the nature and quality of policymaking and planning processes (e.g., District Primary Education Project, and Sarva Shiksha Abhiyan). The institute also helps in the development and implementation of plans and programs of the government and in setting up other institutions in educational planning and management (e.g., State Institutes of Educational Management and Training). Discussions on policy issues at the institute contribute to research, training, and consultancy. In a sense, the institute serves as an important, rather the only, think tank and also more than a think tank on educational issues in the country. The institute's faculty also render consultancy services to international organizations in the area of educational policy, planning, and development. Presently as a research university, it has an additional program of doctoral research in the area of educational policy, planning, and administration. This fourth important activity recently added, involves

teaching and research, which aims at producing certified professional educational planners and administrators. Another important function that the institute took up was collection, processing, and dissemination of huge amount of data on schools in India and developing an educational management information system. All these functions are visualized as closely related and to be contributing to strengthening of each other. Thus, analysis of policy issues and development, application and dissemination of planning methodologies, and management techniques form the main content of the institute's activities. The very names of the academic departments also indicate the area of focus of the university. The departments, some of which are thematic, some level specific, and some geographically specific, are: educational policy, educational planning, educational finance, educational administration, school and nonformal education, higher education, comparative education and international cooperation, inclusive education, foundations of education, and educational management information system, the later four of which were added in the recent years. By very design, the departments are so planned to be overlapping in their work.

In its activities, the NUEPA is often compared with the International Institute for Educational Planning (IIEP) set by UNESCO in 1963, which also has exactly the same four functions – research, training, consultancy, and teaching, teaching being the most recent addition, as in the case of NUEPA. However, the IIEP's activities are mainly at the international level, while the activities at NUEPA are both national and international in coverage, although the national content forms a major part. NUEPA's International Diploma in Educational Planning and Administration of 6 months duration is often compared with IIEP's Advanced Training Programme of 8 months duration. NUEPA and IIEP do not however, compete, but actually have close collaboration with regard to quite a few of their activities; they jointly conduct training programs and research studies and bring out joint publications. The collaboration has been mutually beneficial. From its inception, the institute has an active international collaboration and its activities spread beyond the national borders. Presently, it has international collaboration for research and training with a large number of universities (e.g., Hiroshima, Sussex, and Stanford) and institutions of higher education (e.g., Korean Educational Development Institute, Shanghai Institute of Human Resource Development, and Australian Council for Educational Research, to name a few), and also international development agencies, such as the World Bank, the Department For International Development (DfID), the United Nations Children's Fund (UNICEF), the United Nations Educational, Scientific and Cultural Organization (UNESCO), Danish International Development Agency (DANIDA), and European Union (EU). Collaboration with some of these institutions has been under the

framework of a network called Asian Network of Training and Research Institutions in Educational Planning (ANTRIEP), in the formation of which, the institute along with IIEP played an important role, and the focal point of which is housed at NUEPA.

In short, as an apex-level organization, the institute serves the society with a strong research base, as a major think tank, as a training institution focusing on building capacities of manpower at different levels, helping in skill formation and building of human capital among the educational planners and administrators in India and abroad, and as a consulting and advisory body, contributing to sound policymaking and efficient planning at international, national, and subnational levels in the area of education.

Governance

The University is governed by a council, which is the highest-level governing body of the organization. This is presently headed by the Minister for Human Resource Development in the Government of India. The University has also a Chancellor. The Board of Management, headed by the vice-chancellor, is the main executive body that guides and supervises the functioning of the organization. Other management and other committees include the Finance Committee, Academic Council, Board of Studies, etc. The annual report of the institute is presented in the national Parliament every year.

Funding

Except for the first decade of its existence, when it, as the Asian Regional Centre or the Asian Institute when it was completely funded by the UNESCO, it as a National Staff College or National Institute and now as a National University is funded primarily by the Government of India. The University, however, generates a small amount of resources through student/trainee fees and miscellaneous user charges, and also through projects from international organizations.

Regarding the fee for trainee, it may, however, be added that most of the training programs are offered free. No fees are charged. The participants, in most cases, however have to meet their travel and subsistence costs, and in some special needy cases the travel and subsistence costs are also met by the institute. Since the institute's core budget comes from public sources, and since the trainees are mostly public servants, such an arrangement is found to be somewhat socially efficient. Travel and subsistence costs of trainees of the 6-month national diploma program are also met by the institute. In addition, a stipend/fellowship is also given to the trainees. A small token fee is charged for programs specially meant for private sector in India. A major exception to the free

training is the international diploma program, for which a fee is charged. The fees, however, do not cover the total costs of the program; the program is also partly subsidized by the University.

Although a small token fee is charged from the students enrolled in the University's doctoral program, every full-time student is also provided with a generous fellowship. This is to attract talented graduates from universities to pursue serious research in educational policy, planning, and administration.

Challenges and Future Perspective

From a small training and research center, over the last four-and-a-half decades, the institute has undergone significant transformation and has grown into a national university. The young university faces new challenges and tasks, with the rapidly changing socioeconomic, political, and educational milieu. Although the institute continues to remain focused on educational policy, planning, and administration, it has diversified in its areas of interest and extended to new areas. For example, for a long time it was felt that traditional disciplines of study which were offered in conventional universities need not be the focus of the institute. Hence, areas such as foundations in education and comparative education were not considered relevant for the institute for a long time. However, given that inadequate attention has been given to these areas in the other universities, the National University has taken them on as additional areas of concern for itself.

Second, the institute has also diversified its activities. Although its core activities revolve around research, training, consultancy, and teaching, several other activities also gradually entered into the work program of the institute, which include funding and management of research, collection and dissemination of data on a large scale, helping the government in executing policies and implementing programs/schemes, serving as a clearing house of ideas and information, conducting extension activities, organizing briefing sessions, in addition to conferences, workshops, meetings, and seminars, and also serving even as an advocacy group.

Given the small size of the faculty, and accordingly in the limited number of areas of specialization, spreading of the interests thinly into various areas of operation may lead to different kinds of problems. The principle of comparative advantage suggests that although one can do many activities better than many others, it would be socially efficient if one focuses on the areas in which he or she or the institution has a more relative advantage. NUEPA may have been at fault, to some extent, in its approach with respect to this principle.

Third, research is a very important function of the institute. For obvious reasons, the whole research is policy

centric. Although it can be argued that every research has a policy use in the long term, the term policy here is narrowly interpreted so as to refer to immediate, current, or contemporary policy. Evaluation exercises and also developmental work on implementing government policies and programs all tend to be treated as policy research. Any research which is not of direct relevance to the present context is regarded as other or academic research. However useful it could be in a long-term perspective, such research might not get a place in policy-research organizations. Nonetheless, there are discussions about according due priority to academic research as well, which is not only of long-term significance, but also long term in duration of research, which can also be termed as delayed returns research in contrast to immediate/instant returns research (Beteille, 2007).

Given (1) the role of the government in funding and in the governance of the institute on the one hand, and (2) the predominance of activities that support and/or are related to the government programs – policy research and training the government personnel, NUEPA runs the risk of being branded as a government department, by the government, by the organization, and by the society at large, which may turn out to be detrimental in the long run for developing it as a strong independent academic organization. There are, therefore, calls for making strong efforts to develop, through its various activities, critical thinking on various issues, going beyond and even against the governmental thinking.

An important issue that NUEPA faces is shortage of faculty. Given the specialized nature of faculty and its areas of work, one does not find manpower in abundance. The shortage of manpower has been severely felt by the university as whole as well as by certain specific areas. There is a demonstrated necessity of securing a critical mass in the development of a specialized branch of study. The interdisciplinary nature also makes it a more complicated issue. The shortage of manpower also makes it difficult not only to fully develop the institute, but also to produce replicas or its centers in a big country like India, the need for which is being felt.

Over the years, the institute emerged as a strong, vibrant, and unique institution of its own kind not only at the national level, but also in the Asian region and the developing world, and even at the international level at large in research and training in educational planning, administration, and development. Both as NIEPA and NUEPA, the institute is different from typical research/training institutions and conventional universities. The contribution of NUEPA, including in its earlier forms, has been remarkable, although not vast in quantum. Its contribution is confined to research, capacity building, and lending support to the government in the area of educational policy, planning, and administration, but its contributions are unique. Its training programs have vastly

improved the quality of planning and administration in India and in quite a few other countries. The research and professional contributions have helped in the improvement in the information base, in informed policymaking, and created knowledgeable discourses on policy issues. Given the changing and challenging circumstances NUEPA faces daunting challenges in the near future, to make the institute's academic programs qualitatively better, sustainable, and more effective, and to produce a significant multiplier effect on the educational research environment.

Bibliography

- Béteille, A. (2007). The School the Community. *Journal of Educational Planning and Administration* **21**(3), 191–202.
- Education Commission (1966). *Education and National Development: Report of the Education (Kothari) Commission 1964–66*. New Delhi: Government of India.

Relevant Website

<http://www.nuepa.org> – National University of Educational Planning and Administration (NUEPA).

On School Management – International Council on School Effectiveness and Improvement, Commonwealth Council on Educational Administration and Management

A Walker, Hong Kong Institute of Education, Hong Kong SAR, People's Republic of China

T Townsend, The University of Glasgow, Glasgow, UK

© 2010 Elsevier Ltd. All rights reserved.

The purpose of this article is to outline the development and activity of two successful international professional networks focusing on educational management and organizational change. Between them the Commonwealth Council for Educational Administration and Management (CCEA/M; the CCEA was renamed as the Commonwealth Council of Educational Administration and Management (CCEAM) in 1994 (Sapre, 2000). For a description and history of the Commonwealth foundation see their website at: <http://www.commonwealthfoundation.com/>) and the International Congress for School Effectiveness and Improvement (ICSEI), have successfully stimulated meaningful cross-national and cross-cultural dialog between scholars and practitioners since the early 1970s. As such, both organizations have worked successfully to fuse closer links among school leaders and others attempting to improve education across the globe.

The article has three sections. The first section describes the foundation and ongoing activity of the CCEA/M; the second section does the same for the ICSEI. Knowledge of why these organizations were formed, who was involved, and the activities they engaged in is necessary to understand why they have become successful networks. A brief recount of their histories holds some useful messages for the establishment and sustenance of international networks even though these largely pre-date the electronic era. The third section, therefore, identifies a number of key messages that emerge from the histories.

The Commonwealth Council for Educational Administration (and Management)

In 1955, the infant University Council of Educational Administration (UCEA) was established at Teachers College, Columbia University with a grant from the W. K Kellogg Foundation (see Culbertson (1995) for a history of UCEA). The US-based council's purpose was to improve the professional preparation of what were then labeled 'administrative personnel in education' (Culbertson, 1976). It also sought to build and broadly disseminate empirical findings, emerging concepts, and worthwhile teaching materials to activate their agenda.

In 1966, as part of its strategy to collaboratively build and share knowledge, UCEA brought together key educators interested in educational administration from Australia, Canada, Great Britain, and the USA to discuss research and practice in the area. This ground breaking gathering was funded by the Kellogg Foundation and was called the International Inter-Visitation Program (IIP). At the conclusion of the foundation, IIP delegates expressed a desire to continue their work together and meet regularly in order to advance international exchange and development. Part of the conversation around this theme focused on the benefits of closer formal links with and between British Commonwealth countries.

Conversation in this direction ensued over the next three or so years and reached fruition at the 1969 UCEA conference in Atlantic City. Partly facilitated by Jack Culbertson, then president of UCEA, an agreement was reached to hold a second IIP in 1970, this time in Australia. Accord was also made "to attempt to establish a Commonwealth Council for Educational Administration (CCEA) as soon as possible" (Walker, 1972: 20).

The second IIP was jointly planned by the UCEA and the University of New England in Australia. It was generously funded by the Commonwealth Foundation. The key role of the UCEA in the early foundation of the CCEA prompted Walker (1972) to call the fledgling organization "UCEA's bright son (*sic*) at Morning." However, despite the close linkage between the two bodies, CCEA diverged in some important ways from its US-based backer. As Culbertson (1995) noted, the understanding was "that the projected new organization should *not* be an international UCEA. Rather, it (would) be defined by its differences – not its similarities to UCEA" (p. 185, emphasis in original).

As planned, the CCEA was officially formed during the 1970 IIP. Fourteen developed and developing Commonwealth countries attended the conference (Culbertson, 1995). The infant CCEA was established as a nongovernmental organization (NGO) within the Commonwealth. As noted, CCEA differed fundamentally from UCEA in that it opened membership to non university-based educators – to everyone and anyone interested in the practice of educational administration (UCEA enrolled university-based units only, not individual members). The first CCEA newsletter (CCEA, 1971) stated this clearly:

(Membership is) open to all those interested in the administration of education. This prescription enables educationists from all levels – pre school to tertiary – to participate in Council activities. Similarly, it permits membership for practising administrators, for scholars and professors involved in the study and teaching of educational administration, for teachers aspiring to administrative status, for politicians and civil servants” (p. 1).

The purposes of the fledgling council further reinforced the inclusive nature of its establishment.

1. To foster close links between those concerned with the improvement of educational administration in Commonwealth countries.
2. To foster a high standard in the practice and study of administration at all levels.
3. To hold Commonwealth-wide and regional conferences on various aspects of educational administration.
4. To facilitate the dissemination of knowledge about research and practice in educational administration.
5. To foster high standards in the preparation of educational administrators.
6. To facilitate the exchange between member countries of teachers, students, and practitioners of educational administration
7. The establishment in Commonwealth countries of national associations of those concerned with the improvement of educational administration. (Thomas, 1971: 129)

The aims of the organization have changed somewhat in terms of nomenclature and the number of ‘points’, depending on the source, but they have held steadfastly to the founding values. Although predominantly comprising scholars from the ‘old Commonwealth’, the composition of the first CCEA Board highlighted its intention to become a true international network. Board members came from Australia (3), England, Canada, The Gambia, New Zealand, and East Pakistan (now Bangladesh). The background of the presidents over the years shows a similar, though stronger propensity toward the traditional Commonwealth powers. In chronological order, the presidents were: William Walker, Australia (1970–82), Robin Farquhar, Canada (1982–86), Meredydd Hughes, UK (1986–90), Bill Mulford, Australia (1990–94), Angela Thody, UK (1994–2000), Jo House, New Zealand (2000–04), and Petros Pashiardis, Cyprus (2004–08).

The formation of the CCEA as an international network was unusual for its time, at least within the Commonwealth. The normal track for international network formation through the Commonwealth foundation was for diffuse national bodies to be brought together into a multinational network. In contrast, CCEA followed what Harris (1976) called an ‘upside down’ approach in that it actually began as a Commonwealth-wide structure and then proceeded to foster the establishment of national

and/or intra-national professional bodies dedicated to educational administration – this became one of the major activities of the CCEA. As Culbertson recalled, “During the decade of the 1970s, it (CCEA) nurtured the founding of numerous national associations” (p. 186).

The CCEA’s effort to support the creation of professional bodies dedicated to educational administration was remarkably successful. For example, in 1973 it was involved in the formation of the Australian Council of Educational Administration (ACEA – later to become the Australian Council for Educational Leadership) and associated state affiliates. By 1976, CCEA had played a role in supporting similar professional associations in Cyprus, Malaysia, New Zealand, Singapore, Sri Lanka, Trinidad and Tobago, and the UK. In early 2009, it had 23 affiliated bodies, including seven in Africa, five in Australasia, three each in the Americas and Europe, and five (regionally based) in India.

Since its establishment, the CCEA has instituted strategies intended to increase and tighten links between Commonwealth societies and national and more provincial associations. The Council’s main activities include promoting dialog through face-to-face contact, including different forms of conferences, publications, and a range of other connective activities.

The spark behind and formation of the CCEA transpired during the first IIP. This connection was maintained even while the program evolved through the 1980s, 1990s, and 2000s. Involvement in various IIPs put CCEA members in touch with cutting-edge ideas and debates in the field, including the landmark meeting in Bristol in 1974. One which Jacobson (in Lee, 2008) described as “arguably the most notable” in that it “triggered the famous T.B. Greenfield–Dan Griffiths debate” (p. 11). Discussion and involvement with IIP positioned CCEA members at the heart of issues, many of which shaped the meaning and definition of the field (Gunter, 2001; Sapre, 2000; Lee, 2008).

The organization, however, did more than co-sponsor the four-yearly IIP; it also played a key role in encouraging, supporting, and promoting numerous regional conferences and workshops throughout the Commonwealth. In its early years, these included conferences in Fiji in 1973, Penang in 1995, Dacca in 1977, and Cyprus 1980 (Harris, 1976; Hughes and Bush, 1991). The function of these conferences was to focus on areas of concern relevant specifically to educational leaders across the Commonwealth. Perhaps just as importantly, the CCEA/M became heavily involved in people and idea exchange through loosely connected sets of workshops. Commenting on the CCEA/M, Culbertson (2005) captured the influence of such activities.

Going beyond idea exchange, CCEA’s leaders developed national associations and new training programs which helped advance educational management in

many countries. Seeing much potential in CCEA's initiatives, leaders in non Commonwealth countries adopted most of CCEA's innovations and applied them for their own practices. Culbertson (p. 192)

The formation and operation of the CCEA/M also spawned a number of enduring formal communication mechanisms, predominantly through publications. These emerged both within the states comprising the organization and directly from the CCEA itself. The latter includes the *CCEA Newsletter* (currently called *Managing Education Matters*) and *International Studies in Educational Administration*, now in its 36th year of operation and still one of the few truly international journals in the field.

The CCEA/M has also played an influential connective role beyond the Commonwealth, forming strong working relationships with non Commonwealth developing nations, a range of nongovernment organisations (NGOs), the European Forum for Educational Administration (EFEA), UCEA, the European School Heads Association (ESHA), other Commonwealth-based groups and international bodies such as UNESCO (Pashiardis, 2008; Weeks, 1991).

Despite the tremendous effort committed to formal knowledge networks, perhaps CCEA/M's greatest connective contribution cannot be quantified through formal occasions, affiliations, or outputs, but rather through the personal and personal/professional relationships formed between educators across the globe keenly interested in further understanding and improving the practice of educational administration. This phenomenon, which endures still, was neatly captured by the one-time executive director of CCEA/M, Harry Harris:

The role of CCEA is not to be the focal point of all activity. Rather its function is that of diffusing, encouraging, multiplying through collective effort. . . . We need to encourage: improvisation, imagination, divergent ways of thinking, how to be positive, to say 'how can it be done?' to strengthen leadership qualities in everyone. (Harris, (1980: 273), emphasis in original, cited in Weeks (1991))

As the CCEA/M rapidly approaches its 40th anniversary, it continues to function as a conduit for international networking across the field of educational administration. Although the Council has changed somewhat in terms of shape and activity, the observations made by Walker (1991) continue to broadly define the Council:

- CCEA was never anglocentric, with its headquarters located 12 000 miles from London, the base of most other professional associations, and with the spread of linkages from Australasia into the Pacific and South-east Asia matched by encouragement from the United Kingdom of initiatives in Africa, and links from Canada into the Caribbean.
- There was little order or rationality in the Council's growth, a good reason being that expansion usually

depended, in days prior to teleconferencing and the Internet, on the enthusiasm of a committed individual and his or her leadership and *mana* or prestige, a less acceptable reason being the great variety in teachers' conditions and salaries across countries so that a universal subscription rate was quite impracticable.

- Leadership of successive presidents was drawn from across the Commonwealth and the regional representation on the Board and the selection of vice presidents emphasized the sense of a Commonwealth 'Education family' (pp. 9–10).

The formative values which fashioned CCEA/M's character appear as relevant today as in its early developmental years. Evidence of this can be found in the outgoing president's 2008 annual report (Pashiardis, 2008). In the report, Pashiardis welcomed six new affiliates (Cameroon, Kenya, the Seychelles, and three from India – Assam, Gujarat, and Uttar Pradesh), reported funding for conferences in Kenya and South Africa and for visiting fellows in three African countries, noted the continued publication of ISEA and MEM and announced new (or renewed) partnerships with EFEA, UCEA, and ESHA.

The International Congress for School Effectiveness and Improvement

In January 1988, a group of educators met in London for the first International Congress for Effective Schools (ICES). According to Judith Chapman's report in the first issue of the *Australian Network News* (Chapman, 1989: 1):

The initiative for ICES was taken by Dale Mann, former Chairperson (1976–85) of the Department of Educational Administration, Teachers' College, Columbia University, who served as the first Chairperson (1984–85) for the National Council for Effective Schools in the United States ...[who] felt it timely to bring policy-makers, researchers and planners together.

By mid-1987, eight countries, the USA, England, Wales, Scotland, Australia, Sweden, Canada, and South Africa had shown sufficient interest for an international congress to be conducted in late 1987 or early 1988.

The planning group at Columbia was interested in a Congress in two parts: (1) a conference on school effectiveness open to all with an interest and with papers presented in the normal fashion for such events, and (2) a decision-making meeting at which the organisation would be formally constituted and decisions made. (Chapman, 1989: 1)

In January 1988, the first Congress was held at the University of London. Policy makers, practitioners and scholars from 14 countries, including the initial eight, together with Germany, Hungary, Ireland, Israel, the

Netherlands, and Norway, attended the Congress and adopted the name 'International Congress for School Effectiveness.' However, it was clear to those involved that this did not really capture the current mood for change in education. The effective schools research had the underlying purpose of developing practical means for school improvement, but there were some important distinctions and relationships between school effectiveness and school improvement that needed to be identified and highlighted. As Smink (1991) pointed out:

School effectiveness is concerned with results. Researchers try to describe certain variables for school success in measurable terms. On the other hand, school improvement places the accent on the process; here one finds a broad description of all the variables that play a role in a school improvement project. Both approaches need the other to successfully modernize the system. (p. 3)

So the organization became the International Congress for School Effectiveness and Improvement when it was formally constituted at the third congress in Israel in 1990. The Congress's constitution states that:

- The purpose of ICSEI is to enhance the quality and equity of education for all students in elementary (primary) and secondary schools.
- It is intended that this purpose be achieved through the acquisition and dissemination of information in a co-ordinated effort among policy makers, practitioners and scholars, using the Annual Conference, the Congress's related journals, the website and any other appropriate methods.
- It is acknowledged that efforts to achieve this purpose are made in diverse settings by many individuals and organisations employing a variety of perspectives. These efforts have been variously described, reflecting among other things a concern for school effectiveness, school improvement, excellence in education, quality in education, equity in education, school differences and school effects. (www.icsei.net)

In the early years, the organization was overseen by country representatives who made recommendations to the Annual General Meeting held during the congress. Congress program chairs covered the administration of the organization, supported by the publisher Swets & Zeitlinger. By the early 1990s, a constitution had been developed and it was agreed that ICSEI would be better served by having a formal committee, a president, and a formal structure. Hedley Beare from Australia served as the first ICSEI president (1994–95) followed by Larry Sackney, Canada (1995–97); Janet Chrispeels, USA (1997–99); Tony Townsend, Australia (1999–2001); Louise Stoll, England (2001–05); Leif Moos, Denmark (2005–2007); John MacBeath, Scotland (2007–2009), and currently Tony Mackay, Australia (2009–11). To ensure

international perspectives are always considered, the ICSEI Board members must come from different countries, with no single country having more than one general member (except for the elected office bearers). Key decisions made by the Board still need to be accepted by the Annual General Meeting, ensuring that all members have the opportunity to contribute. Membership of ICSEI is fee based, but there are different fees for different categories of membership, with people from UNESCO-designated low-income or medium-income countries either paying a substantially reduced membership fee, or no fee at all.

The two dominant driving forces-cum-communicative mechanisms for ICSEI have been the Annual Congress and the journal *School Effectiveness and School Improvement*. These two vehicles were seen as the prime means of bringing people together and the key to ICSEI's success as an international network.

In January 2009, in Vancouver, ICSEI celebrated its 22nd year of bringing over 500 people together from more than 50 nations. Conferences have been held in many parts of the world, in Rotterdam, the Netherlands (1989); Jerusalem, Israel (1990); Cardiff, Wales (1991); Victoria, Canada (1992); Norrköping, Sweden (1993); Melbourne, Australia (1994); Leeuwarden, the Netherlands (1995); Minsk, Belarus (1996); Memphis, USA (1997); Manchester, UK (1998); San Antonio, USA (1999); Hong Kong (2000); Toronto, Canada (2001); Copenhagen, Denmark (2002); Sydney, Australia (2003); Rotterdam, the Netherlands (2004); Barcelona, Spain (2005); Fort Lauderdale, USA (2006); Portoroz, Slovenia (2007); and Auckland, New Zealand (2008).

Each year, key educational researchers, practitioners, and policy makers have been brought together to consider ways of making schools effective for all students who enter them. All of the key thinkers in the field of school effectiveness or school improvement have been keynote speakers for these congresses; these have been attended by researchers, policy makers, and practitioners from more than a hundred countries.

Substantial progress has been made from the early 1980s, when the five-factor model of school effectiveness (leadership, instructional focus, climate conducive to learning, high expectations, and consistent measurement of pupil achievement; Edmonds (1979)) was paramount, to a time in the 1990s when it was widely acknowledged that the effectiveness of any school must be considered within the context in which that school operates rather than simply on the various 'ingredients' that help to make up the school's operations.

Despite its strong proactive stance on the importance of ongoing research into school effectiveness and school improvement, ICSEI and other organizations recognized that this was not universally accepted by education researchers (e.g., Slee *et al.*, 1998; Thrupp, 1999). To address this, ICSEI was invited by the American Education

Research Association (AERA) to present a symposium on international developments in the field, which brought the proponents of school effectiveness research face to face with the critics – the challenge was readily accepted.

On Wednesday 26 April 2000, the session entitled ‘School effectiveness comes of age: 21 years after Edmonds and Rutter; has school effectiveness had a positive or negative effect on school reform?’ was offered at the AERA conference in New Orleans. Four papers were offered and a lively debate ensued. The four papers (Slee and Weiner, 2001; Reynolds and Teddlie, 2001; Thrupp, 2001; Teddlie and Reynolds, 2001) proved to be so insightful and popular that they were published in the *Journal of School Effectiveness and School Improvement* (Vol. 12, No. 1, March 2001) as a means of expanding the debate. This became a landmark issue because of the broader critical position, previously not heard, as well as the more traditional perspective was brought to the field – this soon became another characteristic of the Congress and moved their activities beyond the annual conference and the journal.

In order to encourage ongoing debate both between congresses and for people unable to access the journal (or journals, now that ICSEI has entered partnerships with leadership journals around the world to enable members to access the literature they need), ICSEI has now established a series of networks. These include the *Educational Leadership Network*, *The 3P Network* (policymakers, politicians, and practitioners), and the *MORE Network* (methodology of research in effectiveness), where members can access information, exchange ideas, and blog with each other.

Another initiative emerging from ICSEI was the regular exchange of international research through country reports. This strategy was designed to stimulate research and debate across national cultural boundaries. Country reports became a major part of ICSEI’s development and comprised a significant contribution to the Congress (a selection of the reports from this first meeting was published in Reynolds *et al.* (1989)). Creemers and Osinga (1995: 1) explain that the major studies (Brookover *et al.*, 1979; Mortimore *et al.*, 1988; Rutter *et al.*, 1979) were well known but almost nobody had a full picture of the studies and the improvement projects going on in the field in all the countries participating in this first meeting.

The second meeting in Rotterdam in 1989 started the tradition of country reports. Creemers *et al.* (1989) demonstrated clearly that the search for more effective schools no longer remained a North American and European pursuit. However, it also became clear that the time required to translate research into practice meant that it was preferable to produce ICSEI country reports every few years rather than each year. As it stood, there was much new, more general research, and activity and this deemed to take precedence in ICSEI’s formative years.

Consequently, the next major attempt to collate a series of country reports was made for the Leeuwarden

conference in 1995, where nine countries from Europe, North America, Asia, the Middle East, and the Pacific region joined to become part of the ICSEI reporting network. The major aim of the conference was to establish links between school effectiveness and school improvement. Scholars were invited to comment on international developments. Their comments provided a context within which international development in school effectiveness and school improvement, in the areas of research, policy, and practice, could be judged (some of the country reports were subsequently published in *School Effectiveness and School Improvement*, vol. 7, no. 2, 1996).

At the 1998 Manchester conference, the ICSEI country reports were reactivated, but with the special brief of trying to increase both the number and the diversity of the countries included. The purpose here was to encourage educators from a greater range of countries to consider development falling within the purview of school effectiveness and improvement, while maintaining contact with countries already involved. The result was a publication by Townsend *et al.* (1999), which contained a total of 20 country reports, including those of some nations not represented previously; these included Scandinavian, Pacific, Asian, African, and South American countries.

This publication helped ICSEI to provide accessibility to and understanding of school effectiveness research and improvement policies and practices to developing as well as developed societies. Although this was not complete, it did lay a foundation which allowed the Congress and individual countries to chart the progress of school effectiveness research within and outside their contexts. Through this endeavor, ICSEI publicly acknowledged that developing countries faced issues that were often very different from those in more established economies. For example, some countries were forced to make judgments about the meaning of effectiveness even when educative opportunities could not be provided for all students. Others were trying to interpret effectiveness while struggling to rebuild after suffering under oppressive regimes.

The *International Handbook of School Effectiveness Research* (Teddlie and Reynolds, 2000) and *Improving Schools and Educational Systems: International Perspectives* (Harris and Chrispeels, 2006) provided further evidence of the interest in, and developing understanding of, the international perspective of school effectiveness and school improvement, a tradition continued in the two-volume set entitled *The International Handbook of School Effectiveness and School Improvement* (Townsend, 2007). The handbook contained over 50 chapters from around the world and considered both the history and future of research into these two important fields.

In its short history, the ICSEI has provided the research, the opportunity for interaction, and the leadership that has taken the object of its concern to the forefront of political

and educational thinkers. It has been influential, sometimes more than it perhaps should have been, in changing the face of education in many parts of the world.

International Networks

A number of 'rough' commonalities can be drawn from the histories of the CCEA/M and ICSEI. These patterns emerge from the histories of the two successful international networks built around a dedication to education management and school change. The patterns, however, are based on observations from the two bodies only and may not apply to other organizations or networks. Other tentative conclusions may also be drawn – the list below is not intended to be comprehensive.

- International networks begin when a small collection of like-minded energetic people purposefully gather around a well-defined purpose. This purpose can be a variant of an existing if differently shaped formalized network and/or from a 'seed' concept linked to new areas of interest and research.
- To take root, successful international networks require funding, vision, energy, and concerted political action. These resources and activities come from already formalized structures, different-leveled institutional (established) support, personal energy and belief structures, and purpose-driven collective beliefs that are fed, rather than restricted, by national, ethnic, or cultural traditions.
- When growing, intra- and inter-network energy and values feed off each other to produce personal and collective commitment and enthusiasm. This also drives the acquisition of the physical resources needed for initial sustenance.
- Successful international networks are inevitably 'hooked' to other formal and informal networks. They do not appear to work in isolation but rather cross-fertilize within and across boundaries. As such, at least in their formative years, they appear more complementary than competitive.
- To remain effective (and relevant), networks not only accept evolution but recognize and work within it to change form and adjust activity – but within a defined purpose only. In other words, the skills needed to spark and found a network must be applied throughout the network's life span to keep it useful and relevant – without this, it may cease to exist.
- Successful international networks must explicitly and implicitly embrace diversity and difference. (This is a complex issue. As recounted in the histories, both networks sought to engage diversity, which they appeared to do well. However, their top leadership was drawn almost exclusively from 'developed,' normally English-speaking countries. The fact that this appeared to work

may be related to resource availability and the location of more established institutional and societal structures in these countries.) In fact, without this, international networks cannot claim relevance.

- International networks organize regular and 'movable' high-profile events to maintain their usefulness. These (conferences) seem to be important collective devices for the core network as they act as gathering points for affiliation, sharing, debate, and coherence.
- International networks produce regular formal outputs in their area(s) of focus in order to maintain linkage. Outputs take the form of written publications, usually reporting empirical work, scholarly commentary, emerging knowledge and recent network activities.
- International networks ultimately survive and flourish on the personal and personal–professional relationships forged between members – these loosely attached relationships spread to sustain the networks over time.
- To become truly international, organizations need to recognize and identify strategies to support colleagues from countries whose resources would otherwise prevent them from being actively involved.

This article described the foundation and ongoing activity of the CCEA/M and ICSEI. It also attempted to identify some key messages that emerged from these histories. The assumption was that knowledge of why these networks were formed, who was involved, and the activities they engaged in is necessary if we are to further understand why they have become successful.

Bibliography

- Brookover, W., Beady, C., Flood, P., Schweitzer, J., and Wisenbaker, J. (1979). *School Social Systems and Student Achievement: Schools Can Make a Difference*. East Lansing: Institute for Research on Teaching, Michigan State University.
- CCEA (1971). *CCEA Newsletter No. 1*. http://www.commonwealth.edu.org/cgi-bin/items.cgi?_item=static&_article=200611201634343702 (accessed October 2009).
- Chapman, J. (1989). Australian network grows from international beginning. *Network News* 1(1), 1.
- Creemers, B. and Osinga, N. (1995). *ICSEI Country Reports*. Leeuwarden: GCO.
- Creemers, B., Peters, T., and Reynolds, D. (1989). *School Effectiveness and School Improvement*. Amsterdam: Swets and Zeitlinger.
- Culbertson, J. (1976). The University Council for Educational Administration. *International Review of Education* 22(4), 515–520.
- Culbertson, J. (1995). *Building Bridges: UCEA's First Two Decades*. Pennsylvania, PA: UCEA.
- Edmonds, R. (1979). Some schools work and more can. *Social Policy* 9(4), 28–32.
- Gunter, H. M. (2001). *Leaders and Leadership in Education*. London: Paul Chapman.
- Harris, H. T. B. (1976). The Commonwealth Council for Educational Administration. *International Review of Education* 22(4), 521–524.
- Harris, H. T. B. (1980). Some aspects of educational administration in the Commonwealth. *Paper Presented at the National Institute of Educational Planning and Administration*, New Delhi.

- Harris, A. and Chrispeels, J. H. (eds.) (2006). *Improving Schools and Educational Systems: International Perspectives*. London: Routledge.
- Hughes, M. and Bush, T. (1991). Theory and research as catalysts for change. In Walker, W., Farquhar, R., and Hughes, M. (eds.) *Advancing Education: School Leadership in Action*, pp 232–264. London: Falmer Press.
- Lee, P.-L. (2008). Interview with Steve Jacobson, UCEA's newly inducted president. *UCEA Review* 46(1), 10–12.
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D., and Ecob, R. (1988). *School Matters*. Somerset: Open Books.
- Murphy, J. (1991). *Restructuring Schools: Capturing and Assessing the Phenomena*. New York: Teachers College Press.
- Pashiardis, P. (2008). *CCEAM Annual Report 2008*. Retrieved November 17, 2009, from <http://www.emasa.co.za/node/37>.
- Reynolds, D., Creemers, B. P. M., and Peters, T. (eds.) (1989). *School Effectiveness and Improvement: Proceedings of the First International Congress*. Groningen: Rion.
- Reynolds, D. and Teddlie, C. (2001). Reflections on the critics, and beyond them. *School Effectiveness and School Improvement* 12, 99–112.
- Rutter, M., Maughan, B., Mortimore, P., and Ouston, J. (1979). *Fifteen Thousand Hours: Secondary Schools and Effects on Children*. Boston, MA: Harvard University Press.
- Sapre, P. M. (2000). Realizing the potential of management and leadership: Toward a synthesis of western and indigenous perspectives in the modernization on non-western societies. *International Journal of Leadership in Education* 3(3), 293–305.
- Slee, R. and Weiner, G. (2001). Education reform and reconstruction as a challenge to research genres: Reconsidering school effectiveness research and inclusive schooling. *School Effectiveness and School Improvement* 12, 83–98.
- Slee, R., Weiner, G., and Tomlinson, S. (1998). *School Effectiveness for Whom?* London: Falmer Press.
- Smink, G. (1991). The Cardiff Conference. ICSEI 1991. *Network News International* 1(3), 2–6.
- Teddlie, C. and Reynolds, D. (eds.) (2000). *International Handbook of School Effectiveness Research*. London: Falmer Press.
- Teddlie, C. and Reynolds, D. (2001). Countering the critics: Responses to recent criticisms of school effectiveness research. *School Effectiveness and School Improvement* 12, 41–82.
- Thomas, A. R. (1971). The Commonwealth Council for Educational Administration: A new centre for educational leadership. *Journal of Educational Administration* 9(2), 128–134.
- Thrupp, M. (1999). *Schools Making a Difference: Let's Be Realistic! School Mix, School Effectiveness and the Social Limits of Reform*. Buckingham: Open University Press.
- Thrupp, M. (2001). Sociological and political concerns about school effectiveness research: Time for a new research agenda. *School Effectiveness and School Improvement* 12, 7–40.
- Townsend, T. (ed.) (2007). *The International Handbook of School Effectiveness and School Improvement*, pp 996. New York: Springer.
- Townsend, T., Clarke, P., and Ainscow, M. (1999). *Third Millennium Schools: A World of Difference in Effectiveness and Improvement*. Lisse: Swets & Zeitlinger.
- Walker, W. G. (1972). U.C.E.A.'s bright son at morning: The Commonwealth Council for Educational Administration. *Educational Administration Quarterly* 8, 16–25.
- Walker, W. G. (1991). The changing Commonwealth. In Walker, W., Farquhar, R., and Hughes, M. (eds.) *Advancing Education: School Leadership in Action*, pp 3–13. London: Falmer Press.
- Weeks, J. (1991). CCEA and the Commonwealth: Living organizations for development. In Walker, W., Farquhar, R., and Hughes, M. (eds.) *Advancing Education: School Leadership in Action*, pp 65–278. London: Falmer Press.

Further Reading

- Cuban, L. (1989). The 'at-risk' label and the problem of school reform. *Phi Delta Kappan* 71(8), 780–781.
- Hickcox, E. (2003). Friends by any other name: Around the World with CCEAM, or CCEA, or IIP. *Managing Education Matters* 6(1), 4–5.

Relevant Websites

- <http://www.cceam.org> – Commonwealth Council for Educational Administration and Management.
- <http://www.commonwealthfoundation.com> – Commonwealth Foundation.
- <http://www.icsei.net> – ICSEI.

UNESCO's International Institute for Educational Planning

S J Hite, Brigham Young University, Provo, UT, USA

M Bray, International Institute for Educational Planning, Paris, France

© 2010 Elsevier Ltd. All rights reserved.

Introduction

UNESCO's International Institute for Educational Planning (IIEP), which was founded in 1963, is headquartered in Paris, France. The Institute also has an office in Buenos Aires, Argentina, which was opened in 1998. The Paris office mainly operates in English and French, and has a global focus, while the Buenos Aires office mainly operates in Spanish and chiefly focuses on Latin America.

In its 2008–13 medium-term plan, IIEP indicates that it “is mindful that it is a small organization with a huge mandate” (IIEP, 2007: 5). While the actual or relative size of IIEP is open to interpretation, the extent of its mandate in the core functions of training, research, and technical assistance is indeed huge. Conceived and established during a time of significant geopolitical change, IIEP has been both a mainstay and an innovator in the field of educational planning. Geopolitical contexts and trends have played a significant role in defining the need for IIEP's activities.

Contextual Precedents: Global Expansion, Expectations, and Imperatives

At the first session of the 1963 IIEP Governing Board, C. E. Beeby, chairman of the UNESCO Executive Board (quoted in IIEP, 2003: 4), declared:

I know of no human activity that is much more important at this moment in the world's history than research and teaching in educational planning, and I know none that is more difficult. . . . [W]e know little as yet about the planning of education on the scale, at the speed and in the economic conditions that now face many of the countries we set out to advise. Few of us have had to face problems as complex as these that now beset the Ministers of Education in developing countries. It is for this reason that I am happy at the emphasis you are placing on research as well as on teaching. There is, as yet, no ready-made body of knowledge to impart, and the Institute must learn as it teaches.

Nested within this statement are salient ideas that reflect the complex conditions into which IIEP was thrust. First, IIEP was created in a context defined by a specific set of world conditions, and was designed to address particular issues within those conditions. Second, not only was educational planning seen as difficult, but little

was known (if anything was substantially known at all) about educational planning of the type which IIEP was expected to assist. Planning of course had a long history in the Soviet Union and then China, but this was in a very different political context from that of the newly independent countries in Africa and elsewhere. Third, the scale and speed at which the Institute was expected to operate was unprecedented. Fourth, the focus was intended to be on the needs of ministries of education in developing countries, which has indeed remained IIEP's main focus albeit with some diversification. Fifth, IIEP was expected not only to deliver training but also to conduct research – and in the process to find synergies between them.

IIEP was thus indeed given a huge mandate by the United Nations Educational, Scientific and Cultural Organization (UNESCO). To understand this mandate better, the global context that led to the establishment of the Institute needs to be outlined. This will be assisted by commencing with the creation of the United Nations (UN).

The United Nations

The UN was established in 1945, very soon after the end of World War II (WWII). The impact of the war on geopolitical structures was profound. Empires and nations collapsed either during the war or as a result of it, and the UN was created to address the world transformation that occurred at this time. Four powerful sets of ideas had been generated: peace, independence, development, and human rights. This last idea, human rights, carried the bold assertion that every individual in every country had an equal claim to not only civil and political rights but also to economic and social freedom.

Flowing from this post-WWII idea of human rights, the UN's 1948 Universal Declaration of Human Rights asserted in Article 26 that:

Everyone has the right to education. Education shall be free . . . [and] shall be equally accessible to all on the basis of merit. Education shall be directed to the full development of the human personality and to the strengthening of respect for human rights and fundamental freedoms. It shall promote understanding, tolerance and friendship among all nations, racial or religious groups, and shall further the activities of the United Nations for the maintenance of peace.

Education, then, emerged as not only a fundamental component of human rights but also a mechanism by

which the UN could help achieve its mandate in the post-WWII era. The challenge remained, however, as to how the UN would facilitate this education-oriented mandate. The creation of UNESCO provided the institutional solution, and the evolution of UNESCO's membership in the early years dictated the institutional context giving rise to the need for IIEP.

UNESCO

UNESCO is an agency of the UN, and all UN member states have a right to belong to it. At its inception in 1946, UNESCO was formed with 20 signatory states. As former colonies became independent, this number grew substantially and balances changed. By the time of its 60th anniversary in 2006, UNESCO had 193 member states. As more states joined, most from the developing regions of the world, clear needs emerged for specific types of institutional responses and services. In particular, a focus on the human right of education emerged as critical, particularly in Africa and Asia.

By the early 1960s, UNESCO clearly recognized the need for a focused institutional response to the expanding educational imperatives of emerging member states and the growing geopolitical complexities. The most pressing need for educational assistance focused on countries where the capacity for planning educational systems was clearly the lowest. From the early post-WWII origins of the UN and UNESCO, a global conviction of the twin ideas of education as a human right and the potential power of education in national economic development took hold. However, no proven basis existed upon which the national systems of the newly independent countries could be planned.

Thus, the need for IIEP within UNESCO's institutional boundaries was generated, and the initial elements of IIEP's services were defined. Educational planning, particularly for member states in the developing regions, became a critical but uncharted institutional imperative for UNESCO. A significant component of UNESCO's response to this imperative was the formation of IIEP.

IIEP's Early Years

Government representatives from Brazil, France, Germany, Italy, Nigeria, Sweden, the United Kingdom, the USA, and the USSR, as well as agency representatives from the Food and Agriculture Organization (FAO), the International Labor Organization (ILO), the World Bank, the World Health Organization (WHO), and the UN met in 1962 to discuss the creation of an institute within UNESCO that would focus on training and research in educational planning. As the committee discussed the mission, structure, and management, it determined that

the institute should be multidisciplinary, a place for practitioners and potential practitioners to be instructed and to gather useful experience, and a body with significant autonomy within UNESCO's administrative and funding structures.

Institutional Autonomy

IIEP was formally established in July 1963. The rationale for an autonomous institute was rooted in the belief that autonomy would better allow IIEP to create a strong cooperative and operational network between international institutions that might otherwise be reluctant to converge under the larger UNESCO umbrella with its administrative constraints. At the same time, with IIEP as a formal institute within UNESCO's legal framework yet outside the UNESCO secretariat, those same international institutions could draw upon the political capital of the UNESCO brand. Ultimately, the autonomy of IIEP would allow UNESCO the benefit of being viewed as an active agent in educational planning without impinging on the Institute's ability to create the necessary network of partners. IIEP's autonomy would also enhance its flexibility to define itself and its programs within the emerging domain of educational planning.

IIEP has its own statutes and an independent governing board with full authority over the Institute's policies, programs, and budget. The structure of the governing board is distinctive in the UNESCO system, and brings with it both strengths and potential challenges. Its 12 members are drawn from a broad global spectrum, including eight elected members from among economists, educators, and other discipline-specific specialists, with at least one member each from Latin America, Asia, Africa, and the Arab States. Four additional members are designated by the UN, the World Bank, a UN agency, and a UN economic commission. The strengths of this composition are that the members provide broad disciplinary and institutional representation in the governance, evaluation, and maintenance of IIEP programs and services. When compared to some other UNESCO agencies and institutes, the lack of members of the governing board with voting rights in UNESCO's general conference is one possible drawback. Nonetheless, the level and structure of IIEP's institutional autonomy has served its purpose well.

Training and Research

Training and research emerged early as IIEP's two core functions. In his first report to the governing board, Philip H. Coombs, the Institute's first director, promoted a critical mindset that guided IIEP in its formative years:

In a broad sense the Institute's establishment at this time expresses the growing recognition by economists, educators, general planners, and national leaders that more

emphasis must be placed on the human factor in economic and social development. Shortages of competent manpower, reflecting educational inadequacies, have become in many countries a serious handicap not only to economic growth, but to the strengthening of crucial social institutions and advancement generally. (1 GB/5 \$4)

The Advanced Training Program

For IIEP, the human factor, shortages of competent manpower, strengthening of crucial social institutions, and advancement generally were all parts of how the Institute went about creating and implementing what was initially called the Annual Training Program and was later renamed the Advanced Training Program (ATP). The ATP remains the centerpiece around which IIEP has delivered its various training programs. Lasting 10 months, the ATP chiefly serves participants from developing countries around the world. IIEP's first medium-term plan defined the purposes of the ATP:

- to familiarize participants with the concepts, analytic tools, and techniques necessary for developing competence in the field of educational planning; and
- to enable participants to take advantage of their stay at the Institute in order to think more deeply about their country's educational problems and the ways in which improved planning and administration could help to solve them.

The ATP was created to focus on experienced senior- and middle-level education professionals at the national, regional, or provincial levels. A high priority has been placed on the return of ATP graduates to their home countries for assignment to key posts in the planning and monitoring of national education systems.

Research

Research in a broad sense draws of course from different perspectives, is directed at various needs, and generates knowledge aimed at accomplishing very different types of goals. Since educational planning was a new field in the early 1960s and was directed at the practical needs of national educational contexts in developing countries, the chosen IIEP research modality was applied and practical.

To a significant degree, the early research of the Institute served three basic purposes: (1) to help IIEP set its priorities; (2) to clarify what areas should serve as focal points for the Institute's subsequent research; and (3) to provide the bases for its training courses. The first research studies focused on the USSR, France, and various Latin American and African countries.

From the outset, significant publications have helped to define and guide the field of educational planning. Particularly important is the series *Fundamentals of Educational Planning*, in which all volumes are published in English and

French and of which many volumes have been translated into other languages. As explained in the cover notes for the early booklets in the series, the works:

are written primarily for two groups: those engaged in – or preparing for – educational planning and administration, especially in developing countries; and others, less specialized, such as senior government officials and civic leaders, who seek a more general understanding of educational planning and of how it can be of help to over-all national development. They are devised to be of use either for self study or in formal training programmes.

Geopolitical Events and IIEP Responses

As an institute born in response to the needs of particular institutional and geopolitical realities, IIEP evolved according to changing global events, trends, and priorities. While these events cannot all be detailed here, and the consequent influences and responses cannot be extensively chronicled, some major events and resulting changes in IIEP's programming deserve note.

Among the significant geopolitical events contextualizing IIEP's history have been:

- decolonization and its ongoing effects;
- persistent political turmoil, institutional instability, and economic distress in many parts of the world, particularly Africa;
- the sweeping post-Cold-War era of democratization, with consequent movement toward decentralization, localization, and grass-roots participatory systems;
- the impact of the human immunovirus/acquired immunodeficiency syndrome (HIV/AIDS) pandemic on social services, national economies, and cultural structures, especially in Africa;
- World Bank policies and conditionalities, such as the structural adjustment programs;
- the World Conference on Education for All (EFA) held in 1990 in Jomtien, Thailand, which led to UNESCO being designated as the principal coordinating and monitoring agency for EFA;
- the 2000 Dakar World Education Forum, leading to the Dakar Framework for Action for EFA; and
- The United Nations' Millennium Development Goals (MDGs) resulting from the 2000 Millennium Summit.

Each of the above events has had a major impact on IIEP work. For example, the EFA movement and the MDGs were specific reference points in both training and research, and IIEP became a leader in identification of the implications for education of the HIV/AIDS pandemic. By the end of the first decade of the twenty-first century, the focus of IIEP had transformed significantly in many domains. At the broadest level, the two initial institutional areas of activity – training and research – had evolved into a

three-pronged strategy focusing on training, research, and technical assistance. The domain of technical assistance was officially added by UNESCO's Executive Board in 1993. Much of this assistance is grouped with training in what IIEP calls organizational support.

Training and Organizational Support

Training is widely accepted, both within and outside IIEP, as the most important of the Institute's functions. While IIEP delivers a number of training modalities, the uncontested centerpiece of IIEP's training and organizational support domain is the ATP. Well over 1500 trainees have attended the ATP since its creation in 1965. ATP participants have come from over 150 countries, with most coming from Africa, Asia, and the Arab States.

Perhaps the three most significant changes to the original ATP have been the creation in 1998 of the Spanish-language Regional Training Course (RTC) taught in the Buenos Aires office, the addition of a master's option in 2002, and the launch in 2008 of an education sector planning (ESP) course for Anglophone African countries through distance education supplemented by face-to-face meetings. Both the RTC and the ESP are equivalent to the common core of the ATP, and their graduates can join the program in Paris for the specialized phase of study and completion of the full ATP.

In addition to the ATP are other forms of capacity development. IIEP offers short courses, strategic seminars, intensive training courses, distance courses, Internet discussion forums, and a summer school. These programs serve a very broad spectrum of educationists from around the world. In many cases, they are either tailored to particular countries and held in those countries, or provided on a regional basis. Within these courses, attention has been given to gender balances. From a low of 23% female participants in the ATP class of 1988/89, female participation rose, albeit with fluctuations, to 52% in the 2008/09 class. The RTC has usually had more females than males, with 71% of participants in 2008 being female. Other courses have been dominated by males, and the overall balance in 2008 was 42% female.

IIEP also ensures much capacity development through on-the-job coaching in forms of technical assistance. Some of this has been through relatively short assignments while others have had multiyear arrangements. The latter category has been work in Afghanistan, Angola, Cambodia, and Egypt, for example. IIEP's technical assistance focuses on capacities in:

- preparation of planning documents;
- costing of plans and projects, preparation of medium-term expenditure frameworks, and budgeting;
- educational management information systems at national and subnational levels;
- school mapping and micro-planning;

- educational project design and management;
- organizational audits of ministries of education and other institutions;
- assessment of training needs in decentralization schemes; and
- monitoring and evaluation of educational reforms.

Producing New Knowledge

Following the publication of Coombs' (1968) book entitled *The World Education Crisis: A Systems Analysis*, the interdisciplinary nature of educational planning became more obvious. Economists, sociologists, psychologists, and political scientists, among others, could each demonstrate significant contributions. With the onset of the HIV/AIDS pandemic, healthcare professionals also emerged as central to the endeavor of educational planning.

In the early period, research priorities included:

- the performance of educational systems;
- the mutual adaptation of educational and economic systems;
- the internal effectiveness of educational systems;
- resources employed in education;
- lifelong learning; and
- governance, planning, and management of education.

Global imperatives, such as those driven by EFA, the MDGs, and HIV/AIDS, have generated appropriate responses and realignment within IIEP's research agenda. For example, the medium-term plan for 2008–13 designates the following key areas for the production of new knowledge:

- equity, access and quality;
- costs and financing; and
- governance and management.

While overlap is clearly seen between the two lists, the particular focus and nuanced implications of terminology reflect the influences of contextual and geopolitical realities. All IIEP research is required to have conceptual and/or practical relevance to the real concerns of practitioners, which naturally evolve over time. Research in the contemporary era has given much more attention to themes, such as decentralization, internationalization, and privatization, which were not strong thrusts in the early years of the Institute's existence. IIEP has been able to use its functional autonomy to address some themes which are sensitive and which could not easily be addressed within the more political arena of UNESCO headquarters. An example is IIEP's work on corruption and education which was initially viewed externally with some caution, but is now viewed as having been pathbreaking.

Sharing Knowledge for Informed Decisions

Findings from IIEP's research are shared in three ways. First are the publications, of which the Institute has

produced over 1500 since 1963. The majority of publications from the Paris office are in English and/or French, while most publications from the Buenos Aires office are in Spanish. In addition, works have been translated into Arabic, Armenian, Azeri, Chinese, Japanese, Portuguese, Russian, and other languages. All print publications produced in recent years, and some produced during earlier periods of history, are also available in electronic format. All electronic publications can be downloaded free of charge from the website.

A second mechanism for sharing knowledge is through the documentation center, which contains over 33 000 items of both historical and contemporary value, including national planning documents and the findings of the Institute's research. The documentation center also operates an electronic portal of national plans and related documents, entitled Planipolis. In 2009, this portal attracted 20 000 visits per month, which demonstrated that it was indeed a valued resource. Another team within the documentation center operated the HIV and AIDS Education Clearinghouse in conjunction with other UNESCO offices. The Clearinghouse issued a regular newsletter to call attention to new resources and to explain how existing resources could be accessed electronically and/or in hard copy.

Third, IIEP undertakes policy dialog and advocacy through its website, training, policy forums, and technical assistance. IIEP personnel participate in many significant policy-related and academic events, particularly at the international level and frequently also at the national level. The Institute collaborates for such events with other parts of UNESCO, with the United Nations Children's Fund (UNICEF) and other UN bodies, with nongovernmental organizations (NGOs), and with various national bodies. Three specific networks receive particular IIEP support, namely the Asian Network of Training and Research Institutions in Educational Planning (ANTRIEP), the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), and *Educación, Trabajo, Inserción Social* (RedEtis). IIEP also works closely with bodies such as the Inter-Agency Network for Education in Emergencies (INEE), and provides the secretariat for the International Working Group on Education (IWGE).

The nature and role of ANTRIEP deserves elaboration because it permits introduction of a sister institution which operates mostly at the national level and has a long history of collaboration with IIEP. This institution is the National University of Educational Planning and Administration (NUEPA), in New Delhi, India. NUEPA commenced its life in 1962 with UNESCO support as the Asian Regional Centre for Educational Planners and Administrators (ARCEPA). In due course, it evolved into the National Institute of Educational Planning and Administration (NIEPA), and then in 2006 NUEPA. Particularly in the early decades, IIEP had significant partnerships with this body. ANTRIEP remains among these partnerships.

NUEPA is the focal point for ANTRIEP with significant IIEP support.

IIEP also of course shares knowledge through its alumni network. Graduates from the ATP and other programs are located in ministries of education, international organizations, and development bodies in many countries. The Institute endeavors to remain in touch with the alumni, and continues to receive many requests for information and support. Some of the alumni occupy top positions in government as ministers and senior advisers.

Finally, IIEP welcomes many visitors, especially in the Paris office. Some visitors are seasoned professionals who wish to use the documentation center and dialog with counterparts, while others are junior colleagues who are at the outset of their careers. In the latter category are various groups of postgraduate students whom IIEP welcomes on an annual basis on study tours organized by their universities in France, Germany, the United Kingdom, and elsewhere. The Institute is glad to host such visitors and to promote the types of interchange which are so important for the learning of lessons and for continuing innovation.

Conclusion

IIEP represents an ideal, as well as being an organization; it is comprised of people as well as having a physical location; it offers services globally and in its pair of offices; it has significant autonomy, but chooses to operate in partnerships and networks; it leads and influences the field of educational planning, but is responsive to global events and priorities. In short, IIEP is a complex entity and can be legitimately considered from a number of perspectives. The Institute indeed has a huge mandate, as declared in the 2008–13 medium-term plan. In pursuing its mandate, IIEP has been both responsive and adaptive while remaining central and influential in both practical and academic global efforts to train, research, and deliver technical assistance to receptive individuals, groups, organizations, nations, and regions throughout the world.

See also: History of Educational Leadership/Management; Leadership in Diverse Cultures; National University on Educational Planning and Administration, New Delhi; Teachers in Developing Countries; UNESCO's Role in the Development of Higher Education in a Globalized World.

Bibliography

- Beeby, C. E. (ed.) (1969). *Qualitative Aspects of Educational Planning*. Paris: IIEP-UNESCO.
- Bray, M. (2008). *Evolution in the Field of Educational Planning: Four Milestones in the Work of IIEP*. Paris: IIEP-UNESCO.

- Bray, M. and Varghese, N. V. (2009). *Directions in Educational Planning: Report on an IIEP Symposium*. Paris: IIEP-UNESCO.
- Caillods, F. (ed.) (1989). *The Prospects for Educational Planning: A Workshop Organized by IIEP on the Occasion of its XXVth Anniversary*. Paris: IIEP-UNESCO.
- Coombs, P. H. (1968). *The World Educational Crisis: A Systems Analysis*. New York: Oxford University Press.
- Coombs, P. H. (1970). What is educational planning? *Fundamentals of Educational Planning 1*. Paris: IIEP-UNESCO.
- Emmerij, L., Jolly, R., and Weiss, T. G. (2001). Ahead of the curve? UN ideas and global challenges. *United Nations Intellectual History Project*. Bloomington: Indiana University Press.
- Hallak, J. and Poisson, M. (2007). *Corrupt Schools, Corrupt Universities: What Can be Done?* Paris: IIEP-UNESCO.
- Hanhimäki, J. M. (2008). *The United Nations: A Very Short Introduction*. New York: Oxford University Press.
- IIEP (1995). *Medium-Term Plan: 1996–2001*. Paris: IIEP-UNESCO.
- IIEP (2003). *International Institute for Educational Planning: 40 Years Planning for Change in Education, 1963–2003*. Paris: IIEP-UNESCO.
- IIEP (2007). *Medium-Term Plan: 2008–2013*. Paris: IIEP-UNESCO.
- Valderrama, F. (1995). *A History of UNESCO*. Paris: UNESCO.

Relevant Websites

- <http://www.iipe-buenosaires.org.ar> – UNESCO: Instituto Internacional de Planeamiento de la Educacion.
- <http://www.iiep.unesco.org> – UNESCO: International Institute for Educational Planning.
- <http://www.unesco.org> – United Nations Educational, Scientific and Cultural Organization.

This page intentionally left blank



LEADERSHIP AND MANAGEMENT

Leadership and Management Overview

B Mulford, University of Tasmania, Hobart, TAS, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Educational leadership and management is a relatively young field of study; however, in recent times it has entered a golden age in Anglo-American societies (Day and Leithwood, 2007). At the same time the field is faced with the fact that new managerialism, which embraced managerial efficiency and effectiveness through bureaucracy and accountability as key levers for reforming schools, cannot sustain improvements. What is argued throughout the articles forming the leadership and management section of this encyclopedia is that the time has come for the professionals and educational leaders to strive to ensure what happens now and in the future is what they want to happen in schools and that they are professionally accountable for their choices. However, it is recognized that overcoming the gap between dependence on, or a feeling of the inevitability of, system or school bureaucracies as the means of achieving what they want and their preferred model of seeing schools as communities of professional learners remains a challenge (Mulford, 2008). Part of the challenge involves a shift from what is to what ought to be.

A very brief summary of the major recurring points from the articles forming the leadership and management section of this encyclopedia starts with a reaffirmation of the importance of school leadership (in contrast to school administration or management), especially in its distributed form, for continued student learning. This leadership is then seen to work best when it is indirect, working through the building organizational capacity (such as in communities of professional learners or organizational learning), has agency (such as professional identity or professional autonomy), and uses networks to build social capital, including in and with the wider community. Strong arguments are presented for a greater

focus on what ought to be, including the ethical, equality, social justice, and the democratic. There are also calls for greater use of quality evidence by school leaders and the generation of more quality research in the field to support this evidence.

In order to achieve the greater professionalism that has been called for, it is clear that educational leaders need to understand and be able to act concurrently on the context, organization and leadership of the school, as well as the interrelationship among these three areas. Successful educational leadership involves being contextually literate, organizationally savvy, and leadership smart, as well as being the prime vehicle for linking all three areas. As a single input by a leader can have multiple outcomes, success will depend on which areas and in what sequence the educational leader chooses to spend time and attention over time. This article is organized around these three broad themes.

Contextually Literate

Context matters. School leaders need to be contextually literate. A context involving rapid advances in science and technology, increased globalization, changes in demography, including changes in the nature of work, and pressures on the environment sees educational leaders both having to choose between competing forces and broadening of what counts for good schooling.

Articles forming the leadership and management section indicate that choices between competing forces make the most sense when they foster stability for change (in the form of a school's collective capacity to learn), independence rather than dependence, community rather than individualism, and heterogeneity rather than homogeneity. Broadening what counts for good schooling is seen to

include excellence and equity as well as the cognitive and noncognitive (especially personal and social skills).

Choosing between Competing Forces

Continuity or constant change

In contrast to past continuity, recent times have been witness to constant change, a stream of new movements, new programs, and new directions. Unfortunately, some in education seem to be forever rushing to catch the next bandwagon that hits the scene; unfortunately because there is increasing evidence that many a school and school system and its children have been badly disillusioned by the itinerant peddlers selling the new movements.

The main challenge in such a situation, a world of massive and constant change, is how to foster enough internal stability in people and the organization in which they work and study in order to encourage the pursuit of change. Stability for change, moving ahead without losing our roots, is the challenge (Peters, 1987).

It is quite incorrect to assume that a school is effective only if it is undergoing change. Change may be in an inappropriate direction, for example, toward a facade of orderly purposefulness (Sergiovanni, 1990). It may also involve the use of inappropriate measures of success, especially when they are merely procedural illusions of effectiveness (Meyer and Rowan, 1978) such as program planning budgeting systems, school-based management, charters/partnership agreements, strategic plans, standards, benchmarks, and so on. Such processes contribute an illusion of effectiveness and become desired outputs in themselves, thus deceiving outside observers and many of those in schools as well. Such deception should have no place in the provision of quality education.

In a changing world it might be more helpful for school leadership and management to remember Noah's principle: one survives not by predicting rain (change) but by building arks. Amid uncertain, continually changing conditions, many schools are constructing arks comprising their collective capacity to learn, striving to become communities of professional learners.

Dependence or independence

A second contextual area relates to the balance between the competing factors of dependence and independence and what is believed to be the current imbalance favoring dependence. This situation is most easily seen in the over dependence many of those in schools place on leaders outside schools (such as ministers of education, heads of departments of education), often engendered by the overconfidence of these leaders in their own abilities or importance.

There seem to be a lot of these people who want to tell those in schools what to do. This situation is unfortunate

because many of those doing the telling do not seem to want to accept responsibility for their advice, including not being around long enough to take responsibility for their directions. A different approach is required involving greater independence, empowerment, partnership, and consistency for those in schools. Hyman, who left 10 Downing Street after many years as speech writer and advisor to the prime minister to work as an assistant to the headteacher at London's Islington Green School, concurs:

Perhaps the biggest eye-opener for me on my journey has been how the approach I had been part of creating, to deal with 24-hour media and to demonstrate a decisive government, was entirely the wrong one for convincing front-line professionals, or indeed for ensuring successful delivery. Our approach to political strategy has been based on three things: momentum, conflict and novelty, whereas the frontline requires empowerment, partnership and consistency. (Hyman, 2005: 384)

Individualism or community

Although schools do have the responsibility of care for students, at the same time debate continues as to whether schools can deal with broad social issues (Bernstein, 2000). It may be unreasonable to expect the schools to pick up the slack if the home and community cannot. But who will counter, for example, the pressure inherent in much of our modern society to act alone rather than with, or for, the community? We are reminded in the articles forming the leadership and management section that change for the sake of change, including technological change, is not necessarily good; it must be tempered with other values such as wisdom, compassion, and justice.

A different generation, those born from the 1980s onwards, the new millennial learner (NML), now populate our schools – as students and, increasingly, as staff. The NML is the first generation to grow up surrounded by digital media, and much of their activity involving peer-to-peer communication and knowledge management is mediated by these technologies (Howe and Strauss, 2000). This generation, also called Homo Zappiens (Veen, 2003), has made popular the less controllable socially oriented technologies such as blogs, wikis, tagging, and instant messaging (Pedro, 2006).

In this individualistic, technology-mediated world, a skills crisis would indeed be bad enough but a values crisis would be devastating. For example, turning back the tide of a virtual, computer-based cyberspace existence, with its stress on individualism and encouragement to dissociate oneself from an increasingly challenging world, is vital for our future survival. For, as Peck (1987) has reminded us, a community is a place where conflict can be resolved without physical or emotional bloodshed and with wisdom as well as grace. A community is a group that fights gracefully.

A generation unable to feel for others is incapable of creating the social trust that is so essential to maintain culture; also, as it is in the broader culture, so it is in schools. For example, it has been demonstrated that where teachers' trust in principals is undermined by perceptions of principal co-option of top-down system change initiatives, especially when unsupported by teachers, it results in teacher alienation and feelings of disempowerment, which can then bring teacher strategies of resistance to the fore (Bishop and Mulford, 1999).

Homogeneity and/or heterogeneity

Articles in the leadership and management section argue that a common denominator in successful schools is to find a way to get some of the staff and students to do radical things, to take initiatives, and to take risks. If a system is too tight for risk taking there will be no search, and no development and if there is no search and no development there is no learning.

One lesson here is that reductionist approaches in education, to the complexity that is the world of the teacher and the student, should not go unchallenged. Uniformity for schools and education systems in aims, in standards, and in methods of assessment is a complexity-reducing mechanism. It is far tidier to have a single set of aims for all, a single curriculum for all, a single set of standards for all, and a single array of tests for all than to have locally developed approaches to school improvement.

Homogeneity of outcome for the future of our schools and society is not necessarily the highest pinnacle and attempts to reach it may have backfired in terms of student attitudes to school. International research (OECD, 2004) shows, for example, that more than a quarter of students agree or strongly agree that school is a place where they do not want to go. Researchers in UK are:

beginning to encounter students expressing doubts about the genuineness of their school's interest in their progress and well-being as persons, as distinct from their contributions to their school's league table position. [The result is that] contract replaces community as the bond of human association. (Fielding, 1999: 286)

Broadening What Counts as Good Schooling

Pressures from competing forces, such as those outlined above, suggest that in order for schools to fully achieve their purposes there is a need to broaden what counts for good schooling. Measures of successful student achievement in a knowledge society are increasingly being seen as wider than the cognitive/academic, more personalized and involving both excellence and equity (DfES, 2005; OECD, 2001; World Bank, 2005). For example, if we stress only scientific and technological knowledge, or only

literacy and numeracy, we could languish in other areas, including physically, esthetically, morally, and spiritually.

Consistent with this argument to broaden what counts is a range of impressive research using data from the British Cohort Study. This database follows all children born in the UK in the first week of April 1970 and surveyed them again in 1975, 1980, 1986, 1991, and 1996. At age 10, in 1980, over 12 000 children were tested for mathematics and reading ability and the psychological attributes of self-esteem and locus of control. The children's teachers were questioned about their behavioral attributes of conduct disorder, peer relations, attentiveness, and extraversion. In 1996, at age 26, information was collected on highest qualification attained, earnings, and periods of unemployment.

The author of one of these studies, Leon Feinstein, an economist, summarizes his findings as follows:

...attentiveness in school has been shown to be a key aspect of human capital production, also influencing female wages even conditioning on qualifications. Boys with high levels of conduct disorder are much more likely to experience unemployment but higher self-esteem will both reduce the likelihood of that unemployment lasting more than a year and, for all males, increase wages. The locus of control measure ... is an important predictor of female wages ... Good peer relations are important in the labour market, particularly for girls, reducing the probability of unemployment and increasing female wages. ...

[These results] suggest strongly that more attention might be paid to the non-academic behaviour and development of children as a means of identifying future difficulties and labour market opportunities. It also suggests that schooling ought not be assessed solely on the basis of the production of reading and maths ability. (Feinstein, 2000: 22, 20)

These results have been confirmed in other longitudinal research by Carneiro et al. (2006) where it was found that 7- and 11-year-old children who exhibited social maladjustment were less likely to stay on at school post-16 (after taking into account cognitive ability and other family background factors), did less well in terms of performance in higher education, were more likely to display negative adolescent outcomes, such as trouble with the police by age 16 and teenage motherhood, and even conditioning on schooling outcomes were more likely to have both lower employment probabilities and lower wages at age 33 and 42.

Other research (e.g., Cunha *et al.*, 2005) shows that non-cognitive skills are more malleable than cognitive skills, thus suggesting that while both areas are important, schools can have a greater effect on students' noncognitive than cognitive outcomes. Cunha *et al.* (2005: 1) also reminds us that "remediation of inadequate early investments [in such areas of social skills] is difficult and very costly."

Organizationally Savvy

The articles forming the leadership and management section of this encyclopedia show that school organization also matters. Educational leaders need to be organizationally savvy. Broadening the way schools are organized and run would see a move from the mechanistic to living systems, from thin to deep democracy, from mass approaches to personalization through participation, and from hierarchies to networks. The emphasis here is very much on building capacity, on social capital, learning organizations, and collective teacher efficacy, or, in brief, communities of professional learners.

From Mechanistic to Organic, Living Systems

In her book, *Finding Our Way: Leadership for an Uncertain Time*, Wheatley (2005) employs two competing metaphors – organizations as machines and organizations as living systems – as explanation for both organizations and leadership that differ radically in their functioning and outcomes. The machine metaphor encourages a view of organization as a fixed structure of some sort, a structure consisting of parts that need to be oiled if they are to function together smoothly. From this view, organizations require effortful monitoring, coordination, and direction by someone, typically a leader.

Wheatley (2005: 4) notes that “in the past few years, ever since uncertainty became our insistent twenty-first century companion, leadership strategies have taken a great leap backward to the familiar territory of command and control.” Such leadership, aiming to increase employees’ certainty about their work (and increase the school’s level of accountability to government and the public), is mostly transactional. This means that, in the case of school organizations, teachers are assumed to be motivated by the promise of such extrinsic, positive rewards as money and status and opposing, extrinsic, negative impacts such as school reconstitution and public shaming through the publication of league tables.

Transactional, command and control, forms of leadership on the part of principals further manifest themselves in the close supervision of teachers, specification of the one best model of instruction which all teachers must use, centralized decisions about how time in the classroom is to be used together with very long lists of curriculum standards or expectations which teachers are required to cover with students. Teachers are allowed little autonomy over their work in classrooms, their voices are heard weakly, at best, in school-wide decision making and yet they are held almost entirely accountable for student achievement (Day and Leithwood, 2007).

An organic, or living systems, metaphor encourages a view of organization as a process, one of constant

adaptation, growth, and becoming that occurs naturally and inevitably in response to a strong desire for learning and survival. As Wheatley describes it:

the process of organizing involves developing relationships from a shared sense of purpose, exchanging and creating information, learning constantly, paying attention to the results of our efforts, co-adapting, co-evolving, developing wisdom as we learn, staying clear about our purpose, being alert to changes from all directions. (Wheatley, 2005: 27)

A description of organization-as-living-system bears a strong resemblance to accounts of organizational learning in schools, professional learning communities (OECD (2001, 2006) scenarios for future schools as social centers and learning organizations.

From Thin to Deep Democracy

Furman and Shields (2003) argue that there is a need to move our schools from thin conceptions of democracy based on the values of classical liberalism, and its concern with the right of the individual to pursue his or her self-interest and the resolution of conflict through democratic majority voting, to a notion of deep democracy. Dewey (in Furman and Shields, 2003) saw deep democracy as involving respect for the worth and dignity of individuals and their cultural traditions, reverence for and the proactive facilitation of free and open inquiry and critique, recognition of interdependence in working for the common good, the responsibility of individuals to participate in free and open inquiry, and the importance of collective choices and actions in the interest of the common good. Furman and Shields (2003) state that deep democracy needs to be practiced in schools. However, as a consequence of risk of chaos and loss of control from the forces on schools, the typical pattern they perceive is that students

are expected to conform to hierarchically imposed decisions about what they study and teach and when, what the outcomes of instruction should be, how to behave and talk, and even how they look. ... [In fact,] learning democracy may be one of the least experiential aspects of K–12 curricula. (Furman and Shields, 2003: 10)

The results of a recent analysis of school principal training in the Australian state of Tasmania (Mulford, 2004) leads one to suggest that the same could be said about the adults in schools within bureaucratically designed systems. Deep democracy needs to be practiced by them but it may be the least experienced aspect of their working world, especially when it comes to their own professional development.

Personalization through Participation

A major debate taking place in the UK about the future shape of public services picks up on the confused organizational situation for those in schools. This debate is pitched into the chasm between the way public institutions work and how users experience them. For example, in the education sector it has been argued by Leadbeater that efficiency measures based on new public management as reflected in:

[t]argets, league tables and inspection regimes may have improved aspects of performance in public services. Yet the cost has been to make public services seem more machine-like, more like a production line producing standardised goods. [And, I would add, increasingly create dependence on the system.] ... It is ... clear that the State cannot deliver collective solutions from on high. It is too cumbersome and distant. The State can only help create public goods – such as better education – by encouraging them to emerge from within society. ... That is, to shift from a model in which the centre controls, initiates, plans, instructs and serves, to one in which the centre governs through promoting collaborative, critical and honest self-evaluation and self-improvement. (Leadbeater, 2004a: 81, 83, 90)

It is further argued (Leadbeater, 2004a 2004b, 2005) that public services can be improved by focusing on what is called personalization through participation. The pay off of personalization is believed to be increased knowledge, participation, commitment, responsibility, and productivity. Thus, personalization can be seen to be both a process and outcome of effective public organizations, including schools.

A personalized public service is seen as having four different meanings:

- providing people with a more customer-friendly interface with existing services;
- giving users more say in navigating their way through services once they have access to them;
- giving users more direct say over how the money is spent; and
- users are not just consumers but co-designers and co-producers of a service.

As we move through these four meanings, dependent users become consumers and commissioners then co-designers, co-producers, and solution assemblers. As Townsend *et al.*, point out, the change in focus is from one where teachers teach, students learn, and parents support, to one where everyone contributes what they know (teach), and works toward common goals (support), one of which is the ongoing improvement of what the school does (learns).

In schools, learners (students and staff) become actively and continually engaged in setting their own targets, devising their own learning plan and goals, and choosing among a range of different ways to learn. As we move through these four meanings, the professional's role also changes from providing solutions for dependent users to designing environments, networks, and platforms through which people can together devise their own independent and interdependent solutions.

From Hierarchy to Networks

Authors in the leadership and management section of this encyclopedia and others (e.g., Leadbeater, 2005) believe that personalized learning will only become reality when schools become much more networked, collaborating not only with other schools, but also with families, community groups, and other public agencies. Arguably, one of the best-funded and continuous school networks – The Network Learning Group (NLG) with its hub at the UK's National College for School Leadership (NCSL) – summarizes its learning about the advantages networks in comparison to traditional hierarchically designed organizations as greater sharing, diversity, flexibility, creativity, risk-taking, broadening of teacher expertise and learning opportunities available to pupils, and improved teaching and pupil attainment. They point out that while there is no blueprint for an effective network, it is possible to identify factors that successful networks have in common:

- design around a compelling idea or aspirational purpose and an appropriate form and structure;
- focus on pupil learning;
- create new opportunities for adult learning; and
- plan and have dedicated leadership and management.

However, Leadbeater (2005: 22) warns that the collaboration needed for effective networks “can be held back by regulation, inspection and funding regimes that encourage schools to think of themselves as autonomous, stand alone units.” Levin (NCSL, 2005: 6) agrees, pointing out that there “are inevitable tensions between the idea of learning networks, which are based on ideas of capacity building as a key to reform, and ... reform through central policy mandate.” Rusch (2005), in fact, concludes that networks cannot be controlled by the formal system. She questions the role of the system in effective school networks, identifying competing institutional scripts between what is likely to be required by networks as opposed to the system as follows:

- structures are seen as malleable in networks but fixed and hierarchical in the system;
- conflict is open and valued in networks while it tends to be hidden and feared in the system;

- communication is open and unbounded in networks but controlled and closed in the system;
- leadership tends to be fluid in networks while it is hierarchical and assigned in the system;
- relationships are egalitarian in networks but meritocratic in the system; and
- finally, knowledge and power are based on inquiry and learning is valued in networks while expertise and knowing are valued in the system.

Social Capital and Communities of Professional Learners

Arguably, the two organizational concepts that underpin schools as living systems and deep democracies and involving personalization through participation and networking, are social capital and communities of professional learners.

Social capital

The idea of social capital has enjoyed a remarkable rise to prominence. By treating social relationships as a form of capital, it proposes that they are a resource, which people can then draw on to achieve their goals. It also serves alongside other forms of capital (such as economic, human, cultural, identity, and intellectual) as one possible resource and accepted contributor to our individual, community, and national well-being. International bodies such as United Nations Educational, Scientific and Cultural Organization (UNESCO), Organization for Economic Co-operation and Development (OECD), and World Bank have engaged in extensive conceptual, empirical, and policy-related work in the area.

What does social capital mean? The World Bank (Grootaert *et al.*, 2004: 3) concludes that social capital “is most frequently defined in terms of the groups, networks, norms, and trust that people have available to them for productive purposes.” In addition to this generally accepted definition, Grootaert *et al.* (2004: 4) point out that common distinctions are made among bonding, bridging, and linking forms social capital. Bonding social capital refers to “ties to people who are similar in terms of their demographic characteristics, such as family members, neighbours, close friends and work colleagues.” Bridging social capital is also horizontal in nature but refers to “ties to people who do not share many of these characteristics.” However, it continues to connect “people with more or less equal social standing.” Linking social capital operates across power differentials and thus is seen vertical in nature. It refers to “one’s ties to people in positions of authority such as representatives of public (police, political parties) and private (banks) institutions.”

Knowing the definition of social capital and its different forms is helpful, but it does little to assist educational

leaders with the challenges in building social capital in schools. A way through this situation is for the educational leader to see bonding social capital as that occurring among work colleagues within schools. It is the most developed area in the research literature. Bridging social capital can be taken as that occurring between schools. This area is a recent but growing one in the research literature, especially in the area of networking. Linking social capital can be understood as that occurring between a school and its community. While there is a long research tradition in this area, it tends to be unidirectional, concentrating on what the community can do for the school, rather than the other way around.

The research evidence is clear in its strong support for all three forms of social capital. The outcomes are impressive, not the least of which being improved student engagement, academic performance and later life chances, improved teaching and learning, reduced within school variation and retention of teachers in the profession, and increased individual and community capacity to influence their own futures.

However, the research also points to many challenges to overcome at the contextual, organizational, and individual levels including the current accountability press, especially system preoccupation with a limited number of academic performance outcomes, the micro-politics of schools such as contrived collegiality, groupthink, and conflict avoidance, differences between policy development and its implementation, dedicated leadership, large, secondary, high-poverty schools, and professional autonomy.

Communities of professional learners

Where do we take this research evidence on the importance of and challenges to social capital? The way forward is to see the task as establishing communities of professional learners (CPLs) and to see it as developmental starting with the building of social capital. A message arising from the other articles in the leadership and management section is that those in schools must learn how to lose time in order to gain time. Awareness of, and skill development in, group and organizational processes is a first step in any effective change. Instead of others trying to insert something into a school’s (or community’s) culture, the school, and especially its leadership, should first be trying to help that culture develop an awareness of and responsiveness to itself (Scribner *et al.*, 2002).

In brief, there are three major, sequential, and embedded elements in successful school reform. It takes the two elements in the definition of social capital: groups, networks, norms, and trust and for productive purposes, and extends them to include a third element, learning. The first element in the sequence relates to the community, how people are communicated with and treated. Success is more likely where people act rather than are always reacting, are empowered, involved in decision-making

through a transparent, facilitative, and supportive structure, and are trusted, respected, encouraged, and valued. It is a waste of time moving to the second element until such a community is established. The second element concerns a community of professionals. A community of professionals involves shared norms and values including valuing differences and diversity, a focus on implementation and continuous enhancement of quality learning for all students, de-privatization of practice, collaboration, and critical reflective dialog, especially that based on performance data. However, a community of professionals can be static, continuing to do the same or similar thing well. The final element relates to the presence of a capacity for change, learning, and innovation; in other words, CPLs (see Mulford, 2007).

Each element of a CPL, and each transition between them, can be facilitated by appropriate leadership and professional development. Furthermore, each element is a prerequisite for the other; they are embedded within each other with only the emphasis changing. For example, when learning is occurring there is still a need to revisit the social community and the professional community, especially where there has been a change of personnel and/or a new governmental direction announced.

Using this analysis of bonding, bridging, and linking social capital to understand the importance of, challenges to and developmental nature of CPLs can assist the educational leader in better translating the research into policy and practice. It can help him or her to:

- Understand better and be able to take action on the intricacies involved in moving a school, or part of a school, from where it is now to becoming truly a place of ongoing excellence and equity without those in schools being bowled over by the demands for change that surround them.
- Target appropriate interventions to ensure more effective progression through the stages; in targeting interventions, recognition will need to be given to the fact that it is a journey and that actions at one stage may be inappropriate, or even counterproductive, at another stage.
- Support the position that a school will need to be evaluated differently depending on the stage it has reached.

As Stoll concludes, what colleagues do together outside of classrooms is critically important in supporting what they do inside to promote student learning, as well as to enhance their own and their school's development. However, changing the organization and leading of schools and school systems so they become CPLs will not be for the faint of heart. It will require schools and their leaders to radically rethink how they operate. However, and as Leadbeater (2005) points out, many of the basic building blocks of traditional education, the school,

the year group, the class, the lesson, the blackboard, and the teacher standing in front of a class of 30 children, could be seen as obstacles. All the resources available for learning – teachers, parents, assistants, peers, technology, time, and buildings – will have to be deployed more flexibly than in the past. School leadership in such organizations will certainly be less lonely and more collaborative and professionally interactive than ever before (NCSL, 2007).

Leadership Smart

Educational leadership and management matters, is changing, and, given the changing context and organizational response, needs to be smart. Unfortunately, in this situation, the plethora of advice about strong, adjectival, one-size-fits-all school leadership – instructional, transformational, transformative, distributed, democratic, authentic, teacher, strategic, and so on – seems anachronistic. Successful educational leadership and management is more complex; it needs to be able to see and act on the whole and well as the individual elements and the relationships among them over time (i.e., in a developmental manner). As Hargreaves and Fink (2006) point out, it is a meal not a menu, with all pieces needing to fit together in different ways at different times.

A lack of time and professional isolation are major barriers to collaborative endeavors. Donaldson (2001: 11) describes some major attributes of schools that contribute to what he calls a leadership-resistant architecture reflected in a conspiracy of business. There is, according to Donaldson, little time for the school leader to convene people to plan, organize, and follow through. Contact and the transaction of business often take place catch-as-catch-can. Opinion setting and relationship building in schools, he argues, are mostly inaccessible and even resistant to the principal's formal attempt to guide and structure the direction of the school.

Given that the effects of leadership on a wide range of student learning are largely indirect, it may be that strong, visible, visionary leadership is dysfunctional. Barnett *et al.*'s (2001) research is key here, showing as it does the positive effects of principals demonstrating individual concern and building relationships but the negative effects of being inspirational and visionary. While one leadership style or approach may work well for some leaders, most have a range of leadership styles. As schools develop and change, different leadership approaches will inevitably be required and different sources of leadership will be needed that the development work keeps moving. A one-size-fits-all, adjectival style or approach to leadership, or checklists of leadership attributes, may seem superficially attractive but can often limit, restrict, and

distort leadership behavior in ways that are not always conducive to school development and improvement.

Regarding this point, it is interesting to note that proponents of instructional, transformational, and distributed leadership have, over time, moved well away from the one-size-fits-all, charismatic, heroic model of school leadership and expanded their understandings to include aspects of the context, antecedent conditions (e.g., school level, size, SES) and school mission, culture and a reinforcing structure (especially developing people, collaboration, and monitoring), and instructional program.

Conclusion

As we have seen over the course of this overview article, a great deal of a school's success depends on those areas of school life the educational leader chooses to spend time and attention. As a single input by a leader can have multiple outcomes, a leader needs to be able to see and act on the whole, as well as on the individual elements, and the relationships among them over time.

Recent developments in the field of educational leadership and management demonstrate that it is more complex, nuanced, and subtle than previously portrayed (Mulford 2008). The article moved through evidence on three elements: context, organization, and leaders. Context is related to the forces currently pressing on schools and the implications of these forces for schools and their leaders. School organization focused on evolving models that moved beyond the outmoded and ineffective bureaucratic model to communities of professional learners. Evidence on leaders questioned whether one type of leadership fits all contexts and organizations. A great deal of promise was found in the evidence on successful leaders building school capacity, or communities of professional learners, and doing this in a developmental way.

To be successful on all these fronts and how they interrelate in the interests of continuing improvement in student learning is the biggest challenge for the field. Within such complexity, school leaders need to be part of ongoing conversations about context and its implications for schools. Leaders need to understand and be able to act on the evolving and preferred organizational models for schools; in addition, they need to be able to understand and act on the quality evidence that is now accumulating on being a successful school leader and what it means for their professional learning and standards.

With the eminent retirement of a larger-than-normal proportion of school leaders (for Australia, see Anderson *et al.*, 2007), there is no better time to act on these challenges. Will education systems and, more importantly, the profession take up the challenges? Will they actually use quality evidence, such as that provided in the articles in the leadership and management section of this

encyclopedia, in schools and school systems to enable us to continually move forward? Inherent in this question is a plea for us to move beyond mere technical competence in school leadership. Galton makes this point well in terms of teachers:

By making certain techniques mandatory you run the danger of turning teachers into technicians who concentrate on the method and cease to concern themselves with ways that methods must be modified to take account of the needs of individual pupils. As we face the demands of a new century, creating a teaching profession which while technically competent was imaginatively sterile would be a recipe for disaster. (Galton, 2000: 203)

As it is for teachers, so it is for school leaders (PricewaterhouseCoopers, 2007).

Gronn concludes that it is the latitude for choice and discretion that provides educational leaders with their sources of work satisfaction, fulfillment, and self-actualization. It is when the incentives to perform their work begin to dry up, and the constraints and demands cease to be enabling and become burdensome, that leaders come to experience their work as greedy and all consuming; at that point, some form of disengagement may become an option.

There is clearly a need to achieve better balances in our world, including between learning what the political and bureaucratic systems require of individual educational leaders and what practicing professionals require of themselves and their colleagues. On the basis of the available research, we believe this balance can best be achieved by groups of educational leaders, or professional collectives and alliances, setting, negotiating, and delivering their own agendas. This position is also consistent with the emerging priorities for successful educational leadership detailed in this article. After all, participation in context, organization, and leadership, including policymaking, not only enhances efficiency in implementation, but also contributes to the creation of more pluralistic and democratic political systems (Lecomte and Smillie, 2004).

See also: Contexts for Leadership at the Beginning of the 21st Century; Educational Leadership and Social Capital; Educational Leadership: Philosophical Issues; Equity and Educational Effectiveness; Esthetics, Educational Leadership and Management; Ethical/Moral Issues in Educational Leadership; History of Educational Leadership/Management; Leader Mistreatment of Teachers; Leaders for Productive Schools; Leadership and Urban Education; Leadership for Learning; Leadership in Diverse Cultures; Leadership in the Implementation of Innovations; Leadership: Authentic; Leadership: Democratic; Leadership: Instructional; Leadership: Transformative; Leading/Managing Schools in Communities Made Poor; Networked Learning Communities School-to-School Collaboration as an Essential Component of a System

Reform Strategy; Networks and Communities of Knowledge; Organizational Learning in Schools; Principals Role in Restructuring Schools; Principals Role in Restructuring Schools; Professional Learning Community; Research for Leaders: The National College for School Leadership; Research on Educational Leadership – Approaches/Promising Directions; School Community Relations; School Reform and Restructuring: Self Managing School; Social Capital, Educational Institutions and Leadership; Supervision and Personnel Administration; System Leadership; Teacher Leadership and Organizational Development; The Educational Leader's New Work; Transformational School Leadership; Understanding How Leadership Influences Student Learning.

Bibliography

- Anderson, M., Gronn, P., Ingvarson, L., *et al.* (2007). Australia: Country background report. OECD improving school leadership activity. *A Report Prepared for the Australian Government Department of Education, Science and Training*. Melbourne: ACER.
- Barnett, K., McCormick, J., and Connors, R. (2001). Transformational leadership in schools: Panacea, placebo or problem? *Journal of Educational Administration* **39**(1), 24–46.
- Bernstein, B. (2000). *Pedagogy, Symbolic Control and Identity*, 2nd edn. London: Taylor and Francis.
- Bishop, P. and Mulford, B. (1999). When will they ever learn? Another failure of centrally imposed change. *School Leadership and Management* **19**(2), 179–187.
- Carneiro, P., Crawford, C., and Goodman, A. (2006). Which Skills Matter? London: Centre for the Economics of Education, London School of Economics. <http://cee.lse.ac.uk/cee%20dps/ceedp59.pdf> (accessed April 2009).
- Cunha, F., Heckman, J., Lochner, J., and Masterov, D. (2005). Interpreting the evidence of life cycle skill formation. London: NBER Working Paper 11331. <http://papers.nber.org/papers/w11331> (accessed April 2009).
- Day, C. and Leithwood, K. (eds.) (2007). *Successful Principal Leadership in Times of Change: An International Perspective*. London: Springer.
- DfES (2005). Every Child matters: Change for Children. <http://www.everychildmatters.gov.uk/aims> (accessed April 2009).
- Donaldson, G. (2001). *Cultivating Leadership in School: Connecting People, Purpose and Practice*. New York: Teachers College Press.
- Feinstein, L. (2000). *The Relative Economic Importance of Academic, Psychological and Behavioural Attributes Developed in Childhood*. London: Centre for Economic Performance, London School of Economics and Political Science, University of London.
- Fielding, M. (1999). Target setting, policy pathology and student perspectives: Learning to labour in new times. *Cambridge Journal of Education* **29**(2), 277–287.
- Furman, G. and Shields, C. (2003). How can leaders promote and support social justice and democratic community in schools? *Paper Presented to the Annual Meeting of the American Educational Research Association*. Chicago, IL.
- Galton, M. (2000). Big change questions: Should pedagogical change be mandated? Dumbing down on classroom standards: The perils of a technician's approach to pedagogy. *Journal of Educational Change* **1**(2), 199–204.
- Grootaert, C., Narayan, D., Jones, V., and Woolcock, M. (2004). *Working Paper No. 18: Measuring Social Capital*. Washington, DC: The World Bank.
- Hargreaves, A. and Fink, D. (2006). *Sustainable Leadership*. San Francisco, CA: Jossey-Bass.
- Howe, N. and Strauss, W. (2000). *Millennials Rising: The Next Great Generation*. New York: Vintage Original.
- Hyman, P. (2005). *1 out of 10: From Downing Street to Classroom Reality*. London: Vintage.
- Leadbeater, C. (2004). *Learning about Personalisation: How Can We Put the Learning at the Heart of the Education System?* London: UK Department for Education and Skills.
- Leadbeater, C. (2004). *Personalisation through Participation*. London: DEMOS.
- Leadbeater, C. (2005). *The Shape of Things to Come*. London: DfES Innovation Unit. <http://www.standards.dfes.gov.uk/innovation-unit> (accessed April 2009).
- Lecomte, H. and Smillie, I. (2004). *Ownership and Partnership: What Role for Civil Society in Poverty Reduction Strategies?* Paris: OECD.
- Meyer, J. and Rowan, B. (1978). Notes on the structure of educational organisations: Revised version. *Paper Prepared for the Annual Meeting of the American Sociological Association*. Reported in Hannaway, J. Administrative structures why do they grow? *Teachers College Record* **79**(3), 416–417.
- Mulford, B. (2004). Congruence between the democratic purposes of schools and school principal training in Australia. *Journal of Educational Administration* **42**(6), 625–639.
- Mulford, B. (2007). The challenge of building social capital in professional learning communities: Importance, challenges and a way forward. In Stoll, L. and Louis, S. K. (eds.) *Professional Learning Communities: Divergence, Detail and Difficulties*, pp 166–180. London: Open University Press/McGraw Hill.
- Mulford, B. (2008). *Review 53: The Leadership Challenge: Improving Learning in Schools*. Camberwell: Australian Council for Educational Research. www.acer.edu.au/research_reports/AER.html.
- NCSL (2005). *Establishing a Network of Schools*. Cranfield: Network Learning Group, National College for School Leadership.
- NCSL (2007). *What We Know about School Leadership*. Nottingham: National College for School Leadership.
- OECD (2001). *What School for the Future?* Paris: OECD.
- OECD (2004). *Learning for Tomorrow's World: First Results from PISA 2003*. Paris: OECD.
- OECD (2006). *Think Scenarios, Rethink Education*. Paris: OECD.
- Peck, S. (1987). *The Different Drum: Community Making and Peace*. New York: Simon and Schuster.
- Pedro, F. (2006). The new millennium learners: Challenging our views on ICT and learning. Paris: Paper for the OECD-CERI New Millennium Learners activity. http://oecd.org/document/56/0,2340,en_2649_35845295_38358584_1_1_1_1,00.html (accessed April 2009).
- Peters, T. (1987). *Thriving on Chaos*. New York: Alfred A. Knopf.
- PricewaterhouseCoopers (2007). *Project RB818: Independent Study into School Leadership*. London: Department of Education and Science.
- Rusch, E. (2005). Institutional barriers to organisational learning in school systems: The power of silence. *Educational Administration Quarterly* **41**(1), 83–120.
- Scribner, J., Hager, D., and Warne, T. (2002). The paradox of professional community: Tales from two high schools. *Educational Administration Quarterly* **38**(1), 45–76.
- Sergiovanni, T. (1990). *Value-Added Leadership: How to Get Extraordinary Performance in Schools*. London: Harcourt Brace Jovanovich.
- Veen, W. (2003). A new force for change: Homo Zappiens. *Learning Citizen* **7**(1), 5–7.
- Wheatley, M. (2005). *Finding Our Way: Leadership for an Uncertain Time*. San Francisco, CA: Berrett-Koehl.
- World Bank (2005). *Expanding Opportunities and Building Competencies for Young People: A New Agenda for Secondary Education*. Washington, DC: World Bank.

LEADERSHIP AND MANAGEMENT – AS A FIELD OF STUDY

Contents

Contexts for Leadership at the Beginning of the 21st Century

Educational Leadership: Philosophical Issues

Ethical/Moral Issues in Educational Leadership

History of Educational Leadership/Management

Educational Leadership as a Field of Study: Professional Journals 1994–2006

Research on Educational Leadership – Approaches/Promising Directions

Contexts for Leadership at the Beginning of the 21st Century

M Bottery, University of Hull, Hull, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The last couple of decades have seen a proliferation of approaches in education which attempt to either extract the essence of individual achievements, and apply these universally, or devise policies almost as an ideological thought experiment, and then construct a series of aims, objectives, and targets, to be engineered into particular institutions. Both of these approaches ignore a critical mediating factor – the individual context. This is a fatal mistake, as it ignores the personal understandings of local situations, and therefore the humanity and appropriateness of the educational exercise. This article then argues that context is critical to a proper appreciation of the challenges of education leaders at the beginning of the twenty-first century. It however suggests that context is not a simple term, as there are both different meanings and different levels to the term. It therefore discusses the meanings of context before arguing that there are at least four levels of context. In so doing, it suggests that an appreciation of current contexts suggests the need for four leadership dispositions – what are called ironic, autonomous, ecological, and public dispositions.

What is Context?

The Concise Oxford Dictionary defines context as the circumstances relevant to something under consideration. This is useful as far as it goes, but leaves many issues open, one of which is the problem of selecting circumstances as relevant. After all, Carr (1982), in describing historical facts, said that they were like fish swimming in

some huge ocean, and which ones were actually caught depended in part upon where one was fishing, what equipment one was using, and what one wanted to catch. The same kind of thing can be said of contexts, for they are in part defined by how people define them, and that is usually influenced by a number of prior agendas. This can lead down a relativist road: context is then how you want to define it. However, it need not be: one can require as much justification for relevance in the definition of a context as for any other activity, even if such justification may never be fully objective in some positivist sense. As Popper (1982: 111) argued with respect to scientific proof, it doesn't rest upon absolute foundations of truth, but is rather like a structure being erected on piles on swampy ground. One may never then actually drive the piles down to some permanent solid base, but as long as the structure is sufficiently stable, it can be used, even if one knows that it will almost certainly need further adjustments in the future.

One can, then, get better and worse descriptions, even if they cannot be proved better or worse in some ultimate way. This applies to the descriptions of context given by Seddon (1993) when she argues that there are at least three different frames through which leadership can be viewed. The first, and most simple, frame is the categorical – where particular individuals are not seen as real people, with all their personal idiosyncrasies, but as examples of particular categories of individuals – transformational leaders, for instance. This description of context neither provides a rich description of individuals, nor of the particular circumstances within which their lives are conducted. A richer description is given by Seddon's second frame, the interpretive context, where not only are real people with their passions and peculiarities described, but so are

their relationships, and their different perceptions of these. Yet, a more complete understanding of context would describe all those things of which individuals may not be conscious, but which affect them and their relationships. This is the relational context, described by Gronn and Ribbins (1996: 456) as “networks of hierarchically stratified, material institutional arrangements” which may be unique to different societies or cultures, and which may then “both provide opportunities for the expression, and constraint, of human agency.” In thus describing this movement from categorical, through interpretive, and on to a relational meaning for context, Seddon deepens and broadens our appreciation of the range of factors that must be included.

Levels of Context and Their Interactions

A particularly useful way of developing an appreciation of current contexts when it comes to educational leadership is to examine different levels of context at the micro-, meso- and macrolevels, and of the interactions between these levels. It is important to appreciate not only the issues at these levels, but also the interactions between these different levels. Each is examined in a little detail.

From the Micro to the Meso

The microlevel is viewed here as that of the individual, the interactions between individuals, and the effects that individuals have upon each other. At a time when rationalist approaches to management and leadership focus on aims, outcomes, and processes, and the structures needed to implement and measure these, the influence of this level of context can be underestimated or even ignored. Yet, individuals differ in how they approach and deal with issues, and this can have dramatic effects upon particular events. Take, for example, the extensive literature on burnout, early retirements, and the lack of applications for the post of educational principal in the public sector across the Western world (see Gronn (2003); Hargreaves (2003); Fullan (2004)). However, even though similar pressures may apply to an entire population of headteachers, individuals may react very differently. So while *The Independent* (19/01/07) notes that New Labour, since coming to power, has imposed 58 new responsibilities on headteachers in England (*The Independent*, (19/01/07)), research (e.g., Bottery, 2007) has shown that the effect of such pressures is neither uniform nor predictable, for the personality and experience of the headteacher, as well as the relationships they have with significant others, radically affect the impact of such a legislative deluge.

As one moves from the purely micro- to the mesocontext, one moves to the influence of particular groups within an organization. Bryk and Schneider (2002),

through their work in Chicago elementary schools, provide a good example of this when they argue that trust, as well as other kinds of relationships, within educational institutions should be understood not as simply between individuals, but rather as between individuals taking on roles and responsibilities as a member of a group within the institution. The relationships then need to be seen as transcending the purely personal, for now individuals join with others to take on particular roles. The kind of interaction that occurs then, is what Bryk and Schneider describe as relational trust. In the schools they studied, the significant groups were parents, teachers, local government officials, and the principal. Each had a role to play, a particular objective to achieve. When each of these groups carried out their respective role obligations within the larger vision of institutional priorities, and each group saw that the other groups were doing likewise, the development of relational trust was strongly correlated with school academic attainment. Bryk and Schneider thus demonstrate that relational trust helped sustain a focus by all groups upon advancing the best interests of the children, rather than one which concentrates upon personal interests. Such an example indicates that a context can exist, which shares the characteristics of both the personal microlevel, and the institutional mesolevel.

The mesocontext of organizations can be conceptualized in a variety of ways. Morgan (1997), for instance, suggests that there are at least eight different images in which they may be viewed. As organizations have more of the attributes of one image than another, the dominance of particular characteristics can have major effects upon individuals within them. Morgan's image of the organization as a machine, for instance – described at various times by Weber (1970), Taylor (1911), and by Ritzer (1993) – emphasizes the characteristics of predictability, efficiency, calculability, and control, and tends to reduce individuals to performers of particular operations within a bureaucratized pattern of management. In so doing, however, it can reduce the capacity to think creatively, or ethically. Yet, one needs to avoid painting a too simplistic picture. No organization conforms completely to one image; it contains elements of many other images as well. Moreover, organizations will always contain a variety of different interests, and therefore will never be totally functional, totally rational in some simplistic hierarchical sense. Instead, Hoyle and Wallace (2005) argue that as organizations will always generate incommensurable values and demands, which in turn produce dilemmas and ambiguities, no amount of rationality, efficiency, or control will completely eliminate these. If this is the reality of organizations, then an awareness and acceptance of such ambiguity makes great sense, and becomes an essential leadership-survival tool. Through such appreciation, leaders will be able to better interpret and mediate macro-agendas into specific locations and contexts.

Hoyle and Wallace suggest that the last two decades of centrally imposed directives in the UK have exacerbated the generation of ambiguity and dilemmas, as they have prevented the local interpretation of legislation. In so doing, more ambiguities and dilemmas have been created, making the longer-term, sustainable practice of leadership that much harder. Greater professional and leadership autonomy is then not always a self-serving desire: it follows logically from the need to interpret policies or legislation into the local. Without an ironic disposition, and without the autonomy to make best sense of the local, institutional success is much more difficult. These then are powerful arguments for the recognition and significance of personal and institutional contexts.

The Macro

While the personal and institutional are local contexts, the national is a macrocontext for educational leadership practice, encompassing historical, social, and cultural backgrounds, as well as current legislative determinants of practice. Indeed, Grace (1995) argues that one can only understand the situation and challenges of English headteachers by appreciating this shared national background. The more recent periods of such experience, he suggests, are:

- the nineteenth century assumptions that headteachers would be leaders of educational hierarchies, concerned with control, and moral and cultural transmission;
- the mid-twentieth-century role of patriarchal leadership at a time of professional domination, with a remit to pursue a social democratic vision of the public good; and
- the late twentieth-century experience and expectations of headteachers in a quasi-market situation, where service competition, entrepreneurialism, and responsiveness were seen as central qualities.

It would seem that a further period can now be added to Grace's analysis:

- the early twenty-first century situation of being largely an outcomes headteacher, having to mediate a centralized, short termist vision of standards, value-added targets, benchmarking self-evaluation, and an integration with other services.

No other country has had quite the same experience. Thus, while US principals experience some of the current pressures of English headteachers, they come to their role with a very different tacit knowledge and understanding. Goodwin *et al.* (2005) describe the history of the American principalship as:

- a nineteenth- and early twentieth-century duty of producing an American citizen out of a melting pot of a culturally diverse immigrant population;

- a mid-century emphasis on the use of education as a response to perceived increasing Soviet competition;
- from the 1960s onward as having to deal with the implications of the civil rights movements for issues of equality, diversity, and tolerance; and
- most recently, in having to deal with the implications and fall-out from the No Child Left Behind Legislation of 2002 (Meier *et al.*, 2004).

Such historically different backgrounds suggest that while countries may share the same language, they do not necessarily share the same leadership assumptions. Indeed, such difference is brought out even more markedly when one examines the context of the Chinese principal. Thus, Ribbins and Zhang's (2006) analysis of the role of headteachers in rural China, allied to wider work by writers like Hutton (2006), suggests that at least four strands need to be recognized in understanding the role of the principal:

- the continued influences of the millennia of Confucian and patriarchal culture, which emphasize hierarchy, respect, and rote learning;
- the lingering effects of the cultural revolution, which is still seen by a few as a necessary counterblast to increasing Western capitalist influences;
- the influence of Communist ideology and of the power of the party, which still essentially controls the official political and educational culture of the nation; and
- the expanding impact of the new economic culture, which challenges the ideology of communism, and of its influence upon the individual.

In these, as in all other cases, it has to be asked if these are descriptions of the evolution of the role, or rather an accumulation of expectations. If it is the former, the role is likely to be more understandable, more manageable. There must, however, be doubt here; certainly in the English context, it seems to be less about smooth transition, and more about the jerky movement from one set of ideas and legislation by policymakers desperate to get their ideas implemented and evaluated as a success before the next election, to another group with similar short-term concerns. Indeed, this may well be part of the explanation for the increased stress and early retirements of educators seen throughout the Western world.

Moving to the Global

While contexts may be different, Levin (2003), in discussing the commonalities and differences between English-speaking countries' educational policy experiences, suggests that they share some similarities. Thus, while some may be less wedded to assessment and accountability, and there is differential adoption of policies within countries with federal structures, they nevertheless all share the same

economic rationales for educational change, possess a more critical climate toward professional practice than previously, declare similar pressures on additional educational funding, are engaged in forms of financial decentralization, and in the publication of results, and all see markets as vehicles to generate greater competition and choice. This, then, is a policy context, beyond any particular nation-state, which in an age of policy borrowing cannot be ignored.

It is a context which points to a final level which has been under-appreciated by educators. This is the global context, where a number of forces exert an influence transcending any particular nation-state, or grouping of states. There may be as many as nine of these forces (see Bottery, 2005), steering nation-states toward the adoption of practices and values, and from there down to organizations and individuals. This article examines three of these which have particular effect upon educational activities – technological, economic, and demographic globalizations.

Technological Globalization

Technology has huge influence on the current nature of globalization, partly because of the massive increase in the ability of people around the world to communicate and access information. It is however also in part because of the speed of such technology, which permits individuals to access, make decisions, and take action more quickly than ever before. Such trends then tend to democratize access to information. Indeed, as the cost of technology hardware drops, Friedman (2005) argues that another democratizing tendency takes place: the poor begin to compete on more level terms with the rich, and the level of competition moves from one between nations, to one between individuals. If the recent history of global economic competition has been in part that of Asian tiger economies catching up with their Western counterparts, the present situation is increasingly one where even at the highest technical level, competition is between individuals in rich and poor countries. They can now remain in Bangalore, India, or Dalian, China, and yet compete with the best of the West. The result, globally, is an even greater rate of change, and an intensification of global economic competitiveness. In such situations, Western governments are even more keen than previously to keep their workforces ahead in this game by re-engineering their educational systems to meet these demands.

As technological access, speed of interaction, and economic competition increase, they generate international education comparisons. Government advisors now talk of education as needing to be “world class. . . which matches the best anywhere on the planet” (Barber, 2000). When the vision becomes global rather than national, governments begin to judge their educational systems less by

comparing internal performances, and more by using international comparators. As policy and practice are then seen as international, national and local standards are expected to adapt to international benchmarks. As the vision changes, so does the purpose of education.

Economic Globalization

For some people, economic globalization is what globalization actually means. Certainly, it is an immensely powerful force which impacts on all other globalizations. It consists of a number of different elements. One is the acceptance of the use of free-market approaches as the best means of generating wealth, a view which has migrated to affect the structure and practice of many public sectors around the world. Thus, socialist prime ministers like Tony Blair talk of the choice between private and public provision of a service as a purely pragmatic decision: “what matters is what works” (quoted in Ainley, 2004: 507). Yet, so do Chinese communist premiers like Deng Xiaoping, who argues that the choice of communist or free-market economics in developing China’s economic base is no more important than choosing between a black or a white cat to catch mice: it doesn’t matter as long as it succeeds in achieving the ultimate objective. Free-market economics is then the overwhelmingly dominant global economic standard.

Yet, such adoption is not entirely a nation-state option, for, a second element of economic globalization is the influence of supra-national bodies like the World Trade Organization, the International Monetary Fund, and the World Bank, which facilitate the entry of nation-states into international agreements which prevent them from firewalling their economies against global financial movements. In so doing, nation-states become increasingly vulnerable to global financial market trends, a situation exacerbated by the speed of such transactions. Nation-states, however, are also weakened by a further force, – their increasing dependency upon the investments of multinational corporations, who search out the most profitable national locations for their operations, moving these as and when they see greater profitability elsewhere (see Korten, 2001).

The acceptance of the free-market paradigm globally, the global nature of financial transactions, the national vulnerability to international financial transactions, and the actions of multinational corporations, are large elements in what Bobbitt (2003) argues is the transmutation of the nation-state into the market state. According to him, nation-states, unable to provide the historic degree of protection and welfare to its citizens, are aligning themselves with neoliberal economic policies through the provision of an increasingly privatized range of different educational mechanisms. Their role then is reduced to providing opportunities for citizens to generate individual

wealth. This, Bobbitt argues, has been most strongly seen in the regimes of Bush in the US and Blair in the UK, and likely underpins policy moves in England over the last few years toward greater consumer choice, through the creation of specialist schools with a degree of private investment and considerable private control. While such effects may be rather more parochial than Bobbitt might believe, and vulnerable to changes in political leadership, this does seem a good example of how difficult it is to appreciate national policies without understanding the global context within which such ideas are generated.

Finally, because markets reduces educational activity to that of an individual good, such globalization effects prevent the articulation of the kind of vision for education which Grace (1995) suggests it must have. For him, education can develop in citizens “a moral sense, a sense of social and fraternal responsibility for others”(p. 214), and by doing so, provides the basic conditions for making democracy possible. He argues that, because of this quality, education should be accorded the status of a public good. Given such a global economic context, a third essential leadership disposition, then, in addition to those of irony and autonomy, is a belief in education as an activity transcending the satisfaction of the purely personal, and of the re-statement of education as a good for society as a whole.

Demographic Globalization

One further significant global force impacting on education systems currently is a general increase in the age of populations, at the same time as there is a decrease in fertility (Business week (2005)). Where this occurs, there is a threefold result. First, there is a decline in the proportion of children in a population, and therefore a decline in the number of schools and teachers required. Second, there is a decrease in the percentage of a population who work, and who therefore contribute taxes. As governments then recognize the financial implications of lower tax revenues, of increased spending on pensions, and of having to devote more resources to treating the chronic diseases characteristic of the elderly, policies are generated which are aimed at reducing pensions, or raising the age of retirement. It also means that not only are there less pupils, less schools, and less teachers, but as tax revenues decline, there is less money for remaining educational activities. In such situations, governments are very careful with the money that they do have, wanting more of its control and direction, and greater value for what is spent. The effect on educators' work has been apparent for a couple of decades at least. While the economic globalization of neoliberal policies has led to greater marketization, choice, and competition,

demographic factors have paradoxically led to greater control and direction of professional work, and greater accountability for what they do. In the process, notions of professional autonomy are seriously eroded.

The result highlights once again the ambiguity and irony pointed up by Hoyle and Wallace (2005); this time, contradictory global pressures are the ultimate causes, as government policies of both control and fragmentation are visited upon institutions, which have to be managed, mediated, and implemented by individuals within them. A lack of appreciation of such a global context, and of the forces emanating from that level, leaves professional educators parochial and myopic, unable to see why things are happening the way they are, and therefore unable to articulate for themselves, their pupils, and their communities, why such things are happening, and what can be done about these. If knowledge is power, a lack of appreciation of the global context leaves professionals powerless. This then suggests a fourth and final necessary leadership disposition – the need to develop an ecological view of leadership – one which appreciates the range of contexts that impinge on professional practice and identity, which recognizes that this range extends to the global, and which can understand how such forces affect the work of individuals and institutions.

Conclusions

Context is normally seen as an appreciation of the local, yet, if one is to truly understand the effects of standardization and control, or the marketizing tendencies in education over the last 20 years, one needs to understand the wider contextual situation within which professionals work. Such an ecological perspective then enables educational leaders to better utilize both an ironic pragmatism and an autonomy in translating such macroforces within the particular meso- and micro-contexts they inhabit. However, these three dispositions or qualities, while essential to a more successful and more humane educational practice, also need to be informed by a heightened valuing of the public nature of education, and of the appreciation of its erosion by elements of national and supra-national contexts. In such ways do educational leaders understand that context extends from the personal, through the institutional, past the national, and all the way to the global, defining, determining, and directing their work. Without such an appreciation, they are blind, and the danger is, as Matthew, verse 14 states, “...if the blind lead the blind, both shall fall into the ditch.” The next generation will face enough problems in the twenty-first century without being led by those who cannot, or will not, see.

Bibliography

- Ainley, P. (2004). The new 'market-state' and education. *Journal of Education Policy* 19(4), 497–514.
- Barber, M. (2000). High expectations and standards for all – no matter what. Edited version of a speech delivered to the Smith Richardson Foundation in Washington <http://www.tes.co.uk> (accessed May 2009).
- Bobbitt, P. (2003). *The Shield of Achilles*. London: Penguin.
- Bottery, M. (2005). Education and globalisation: Redefining the role of the educational professional. *Educational Review* 58(1), 95–114.
- Bottery, M. (2007). Reports from the front line – English headteachers' work in an era of practice centralisation. *Educational Management, Leadership and Administration* 35(1), 89–110.
- Bryk, A. and Schneider, B. (2002). *Trust in Schools*. New York: Russell Sage Foundation.
- Business week (2005). Global Ageing. 8th February.
- Carr, E. H. (1982). *What is History?* Harmondsworth: Penguin.
- Friedman, T. (2005). *The World is Flat*. London: Penguin.
- Fullan, M. (2004). *The Moral Imperative of School Leadership*. Thousand Oaks, CA: Corwin Press.
- Goodwin, R., Cunningham, M., and Eagle, T. (2005). The changing role of the secondary principal in the United States: An historical perspective. *Journal of Educational Administration and History* 37(1), 1–19.
- Grace, G. (1995). *School Leadership: Beyond Education Management*. London: Falmer.
- Gronn, P. (2003). *The New Work of Educational Leaders*. Oxford: Oxford University Press.
- Gronn, P. and Ribbins, P. (1996). Leaders in context: Postpositivist approaches to understanding educational leadership. *Educational Administration Quarterly* 32(3), 452–473.
- Hargreaves, A. (2003). *Teaching in the Knowledge Society*. Milton Keynes: Open University Press.
- Hoyle, E. and Wallace, M. (2005). *Educational Leadership: Ambiguity, Professionals and Managerialism*. London: Sage.
- Hutton, W. (2006). *The Writing on the Wall*. London: Little Brown.
- Korten, D. (2001). *When Corporations Rule the World*. Bloomfield, CT: Kumarian Press.
- Levin, B. (2001). *Reforming Education*. London: RoutledgeFalmer.
- Meier, D., Kohn, A., Hammond, L. D.,Sizer, T., and Wood, G. (eds.) (2004). *Many Children Left Behind*. Boston, MA: Beacon Press.
- Morgan, G. (1997). *Images of Organisations*, 2nd edn. London: Sage.
- Popper, K. (1982). *The Logic of Scientific Discovery*. London: Hutchinson.
- Ribbins, P. and Zhang, J. (2005). Culture, societal culture and school leadership – a study of selected head teachers in rural China. Unpublished paper.
- Ritzer, G. (1993). *The McDonaldization of Society*. London: Sage.
- Seddon, T. (1993). *Context and Beyond: Reframing the Theory and Practice of Education*. London: Falmer.
- Taylor, F. W. (1911). *Principles of Scientific Management*. London: Harper and Row.
- Weber, M. (1970). *From Max Weber*. (Gerth, H. and Mills, C. W. (eds.)) London: Routledge and Kegan Paul.

Further Reading

- Bottery, M. (2004). *The Challenges of Educational Leadership*. London: Sage.
- Ritzer, G. (2004). *The Globalization of Nothing*. Thousand Oaks, CA: Pine Forge Press.

Educational Leadership: Philosophical Issues

C W Evers, The University of New South Wales, Sydney, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Coherentism – A pattern of justifying theories, or sets of knowledge claims, that appeals to a range of criteria, including empirical adequacy, consistency, simplicity, comprehensiveness, explanatory unity, and fecundity.

Epistemology – A branch of philosophy that deals with the nature and justification of knowledge claims.

Ontology – The branch of philosophy concerned with the nature of being and existence.

Postmodernism – A philosophical doctrine that claims that particular epistemologies have only the status of meta-narratives, or stories, about knowledge and its justification, and are not themselves justified.

Theory – A body of interconnected claims related to some topic or set of topics.

Introduction

Although discussions of educational leadership presume stances on many philosophical issues, these issues are often hidden or regarded as background, and despite some notable exceptions, are rarely discussed explicitly. Yet, they exert an enormous influence over substantive positions taken with regard to how educational leadership is to be understood, theorized, and practiced.

In an effort to render these influences more transparent, and therefore open to critical commentary, the article is structured according to major philosophical categories that bear on different aspects of educational leadership. Because philosophical ideas have most bearing on conceptions of leadership theory, this is the main focus of this discussion.

Educational Leadership Theory and Science

What can we reasonably expect of a theory of educational leadership? One approach asserts that such a theory should have a normative function, offering advice on what ought to be done in educational contexts, concerning not just the means to achieve certain ends, but also the ends themselves (Begley and Johansson, 2003; Hodgkinson, 1978,

1983, 1991, 1996; Samier, 2003; Sergiovanni, 2006; Starratt, 2004). However, many writers place a premium on leadership theories that purport to be scientific where this sense of the term is thought to exclude norms or values as elements of theory (Hoy and Miskel, 2005). In this approach, leadership theories are descriptive, serving as guides to practice only when it comes to means for achieving goals that have been determined by appeal to normative resources that lie outside these theories. Here, the divide between these perspectives is portrayed as a contrast between leadership as a science and leadership as primarily an ethical and political discipline. Views about the nature of leadership theory are thus shaped by views about the nature of scientific theory.

Within the so-called Theory Movement at the center of attempts, during the 1950s in North America, to reconceptualize educational administration as a more rigorous academic discipline, a good scientific theory of leadership was claimed to possess the following properties, thought central to behavioral science:

1. A hypothetico-deductive structure of empirical claims, with more general claims at the top of the structure and more particular derivable claims descending downward from the top.
2. A procedure of justification based on empirical testability having two components. If empirical claims derived from the theory are actually observed to be the case, then the theory is confirmed. If they are refuted by empirical evidence, then the theory is said to be disconfirmed. A theory is more justified than its rivals if it has more confirmations and fewer, or no, disconfirmations.
3. All theoretical concepts should be operationally defined in terms of some measurement procedure or instrument.

The above-mentioned properties, typical of logical empiricism, found their way into the dominant systems theoretic accounts of educational administration that continue to exert influence today (Evers, 2003a). When applied directly to leadership theory, logical empiricism yields what purports to be a science of leader behavior. An early influential example is Halpin's (1966: 81–130) study, 'How leaders behave', where he describes the methodology in the following terms:

By measuring the behavior of leaders ... we can determine by objective and reliable means how specific leaders differ in leadership style, and whether these differences

are related significantly to independent criteria of the leader's effectiveness and efficiency. In sum, the Leader Behavior Description Questionnaire offers a means of defining these leader behavior dimensions *operationally*, making it possible for us to submit to empirical test additional specific hypotheses about leader and group behavior (Halpin, 1966: 88).

Notice that within this framework of assumptions, ethics is excluded from leadership theory because what functions as evidence for such theory is observable behavior. That is, the evidence will always be about what leaders do, not what they ought to do. As Halpin (1966: 8) puts it, "the immediate purpose of research is to enable us to make more accurate predictions of events, not to prescribe preferential courses of human action."

Also excluded is methodological recourse to a leader's unobservable thought processes such as having certain intentions and beliefs, or holding a particular interpretation of events. We may, of course, define these operationally from patterns of observable behavior; however, the subjective inner episodes of cognition and affect cannot themselves constitute empirical evidence.

These philosophical ideas also influence the formulation of research designs, and even research questions, in the empirical testing of theories of leadership. For example, one issue of much interest is the matter of the effect school leaders have on student learning outcomes. The sorts of statistical analyses reported in meta-analyses conducted by Hallinger and Heck (1998) or Witziers *et al.* (2003) attempt to measure correlations among a range of observed leader behaviors and learning outcomes, or to calculate effect sizes of leadership activities. Evidence of small or no effect sizes in turn figures in assessing, for example, the merits of transformational leadership over instructional leadership (Hallinger, 2003).

The demise of logical empiricism in philosophy in the 1950s and 1960s, at the hands of Quine (1951, 1960), Hanson (1958), Popper (1959), and Kuhn (1970), among others, did not lead to its similar demise in educational administration; however, it did usher in a range of alternative perspectives during the 1970s.

Generality in Leadership Theory

In addition to philosophical disputes over what is admitted into the content of a theory of educational leadership, there is a more fundamental dispute over what it means for a theory of leadership to describe anything. There are two main strands of argument. Concerning the first, even if there is agreement that a theory of leadership can somehow represent phenomena in the world, debate exists over how much generality we can expect such representations to express. Traditional views of science place a premium on law-like generalizations, a reasonable-enough emphasis

where the phenomena under investigation are relatively invariant with respect to contextual variables. However, in the study of leadership in particular, and social science in general, the boundary between context and the phenomena to be described is difficult to draw and, in some cases, may be impossible to draw in principle. Hence, it is claimed, a theory of educational leadership will be about particulars, not generalizations, and making good inferences from this sort of theory to guide practice will be an art, or perhaps a craft, not a science.

This point is made in an incisive way by Donmoyer (2001), following some ideas developed by Toulmin (1972, 1993). His claim is that in meeting the challenge of deciding what to do in social contexts, the sort of theory that decision makers need to appeal to is not that given in scientific theory, but rather that which pertains to a policy field. For example, in deciding where to build a dam, the answer that engineers might give, based on a fairly narrow range of technical considerations, could be different from the answer that policymakers give.

The latter group must balance the engineers' technical concerns with a host of other concerns and values. They might, for example, need to consider such things as economic impact, the toll of neighborhood disruption, and even whether particular decisions will affect their own or their party's re-election. (Donmoyer, 2001: 566)

Scientific theory is thus the wrong model to apply to these sorts of multidimensional, multicriterial decisions since it is the particularities of each case that can be crucial. So if a theory of leadership is centrally a theory of contingent leader behaviors, the difficulty is that both leaders and the circumstances of their actions are so immersed in the particularities of individual differences and the unique features of contexts of action that the scientific notions of generalization, prediction, and deductive explanation from more inclusive empirical claims becomes untenable.

In replying to this argument, observe, first of all, that it assumes something similar to a logical empiricist view of the nature of science. However, more recent philosophies of science also reject this characterization of science (Godfrey-Smith, 2003). Take the quest for law-like generalizations, reasonable in, say, physics, but hardly possible in social science. Nevertheless, there exists a range of orderliness in the world between the ideal of exceptionless laws in physics on the one hand and randomness on the other. The possibility of social life is only sustained by a vast number of patterns in social behavior. We catch planes, drive on busy roads, turn up for work, and are participants in a thousand social activities that are accomplished by organizations whose existence and conduct is sustained by both the expectation and the accomplishment of regularities, or patterns, in the social world.

This range or orderliness is captured by the scientific concept of compression. Following Chaitin (1975) and

Dennett (1991), we say that the more the data can be compressed, the greater pattern they exhibit, where compression means that we can find an expression for the data that is shorter than that required to describe every individual item. Very high degrees of compression can be found in the natural sciences where simple formulas compress vast amounts of data. However, even in the social sciences, the practice of explaining, predicting, and understanding human behavior in terms of the rational coordination of beliefs and desires achieves significant levels of compression.

Second, decision making that involves the satisfaction of multiple soft constraints is actually one of the proposed methodologies for choosing scientific theories, according to recent approaches to theory choice. Not only is empirical adequacy required, but so also are simplicity, comprehensiveness, consistency, explanatory unity, and fecundity. These criteria are more extensive than the traditional logical empiricist criterion of empirical adequacy and are known as coherence criteria (Evers, 1999; Churchland and Hooker, 1985).

Recently, a number of writers on leadership have argued that the traditional science model is compromised further by the unpredictable or even chaotic behavior of some complex systems, including social systems. It is argued that the whole idea of making inferences from generalities in theory should be replaced by the development of learning strategies that take us beyond known particulars (Morrison, 2002).

Hallinger (2003: 346) raises the less exotic possibility that no one model of leadership may be suitable because of the wide range of variation in the contingencies of its application and that “it is virtually meaningless to study principal leadership without reference to the school context.” Once differences in students and their backgrounds, schools and their social and economic location, resources, staffing expertise, and the bureaucratic and organizational arrangement have been taken into account, it is at least an open question whether any particular theory of leadership will be suitable.

One way of combining the effect of radical contingency with the idea that we can build warranted theories of leadership is to locate the task of theory building at the level of the school leader. Suppose that principals approach the tasks of leadership from the perspective of being critical learners. In dealing with problems, they apply their knowledge in the form of a tentative theory, engage in testing their theory against the demand that problems be solved, and improve their theory in the light of this experience by making coherent adjustments to it so that it better provides solutions. Assuming that something similar to this sequence of epistemic procedures makes for some growth in knowledge and allows principals to do better than chance in problem solving, decision making, and acting in the world, we end up with a lot of good

theories of leadership tailored to the contingencies of particular circumstances, but with the possibility that they have little in common (Evers and Katyal, 2007).

The second main strand of argument, associated with postmodern views of educational leadership, by denying the assumption of epistemically progressive procedures that enable the growth of knowledge, denies that theories are known to represent. Rather, theories are best seen as narratives, or stories that we tell ourselves, and are not to be judged on their truth content, because the categories of truth and falsehood are thought not to apply, but on their esthetic qualities, or perhaps their role in giving voice to interests that are silenced (English, 2003; Maxcy, 1994, 2001). According to Maxcy (2001: 585), for contemporary school administrators “theory’s role becomes one of reinterpreting the meanings that attach to practice, transforming them as part of an interpretive rather than justificatory enterprise that is at once critical aesthetics and practical artistry.” This is a position that draws heavily on skeptical arguments taken from epistemology. These arguments are believed to show that knowledge has no foundations and that, as a consequence, theories cannot be said to refer to things in the world. (See Donmoyer *et al.* (1995) for a range of applications of this view.)

As with all major philosophical disputes, debates over skepticism are both complex and enduring. However, in replying to postmodern critics of leadership theory, the broad strategy can easily be sketched. In general, a defense of doubt either requires some foundational premises to start the argument, or doubt needs to emerge from a more coherent nonjustificationist viewpoint. If the former is the case, then the attack on foundations of knowledge seems to require the very thing it is attacking. If the latter is the case, then coherence is functioning as a form of epistemic justification that does not require foundations. This latter alternative opens the door to the possibility of administrative and leadership theories that are justified by appeal to coherence epistemic criteria (Evers and Lakomski, 1991).

Leadership and Ontology

Ontology is the study of what exists, including criteria for positing existence and the natures or categories of posited entities. The link between epistemology – the machinery used to justify theories – and ontology is particularly close where it is believed that what exists is something that is known from what our most warranted theories say exists (Quine, 1948). Nevertheless, other considerations also can determine ontological issues, particularly where the epistemological enterprise is thought to be mistaken.

One longstanding dispute in the field concerns whether the explanatory apparatus of natural science is the appropriate framework for conducting social science and building theories in social science. This issue was first

raised is a particularly clear and incisive way by Thomas Greenfield, initially in 1974, and then, increasingly in a series of publications in which he sought to defend a radically subjectivist view of epistemology and its ontological corollaries (Greenfield and Ribbins, 1993). At issue is whether a theory of administration should try to describe the world as it really is, or whether that is a mistaken goal and such a theory should be concerned with how different people construe the world in different ways. Greenfield takes the latter view, with the result that, for him, theory is not sets of law-like generalizations about society and human conduct, but rather, is “sets of meanings which people use to make sense of their world and action within it” (Greenfield and Ribbins, 1995: 7). However, if reality is how people see things and a matter of their interpretations, understandings, feelings, intentions, reasons, and will, then the ontology of the world cannot be separated from these various subjective acts of interpretation, understanding, and the like. Hence Greenfield’s claim that “organizations are invented social reality” (Greenfield and Ribbins, 1995: 7). It follows from this that so also are leaders and leadership roles.

Although Greenfield’s approach takes its cue from skepticism about the objectivity of traditional science when applied to human affairs, skepticism does not automatically spill over into explanations in terms of human subjectivity, especially where these are to do with reasons (as opposed to causes), because these are subject to their own normative constraints, not everything counting as a good reason. Moreover, a measure of intersubjectivity is required to sustain the kind of collective intentionality necessary for organizational and social life. For someone to be seen as a leader in organizational contexts requires a lot of others sharing that interpretation. Searle (1995, 1998), for example, has developed a view of social explanation that combines a rigorous defense of objectivity in epistemology with a thoroughgoing interpretivist perspective on the understanding of social life. In addition, as we saw earlier, the regularities that underpin the very possibility of social engagement, although not law like, may turn out to be explicable in terms of a science of patterns after the fashion of Chaitin’s (1975) or Dennett’s (1991) work. (See Macmillan (2003) for critical perspectives on Greenfield’s ideas.)

In leadership studies, ontological issues traditionally arose over the question of whether leadership is best characterized in terms of a set of special attributes, or properties, or traits that leaders possess. Although the quest for a stable set of traits possessed by those who are clearly leaders has been carried on extensively as an empirical research program by many scholars, the results have been disappointing. There does not seem to be any obvious nontrivial trait, or set of traits, that all leaders possess, or which is essential for the performance of leadership tasks.

Some writers have suggested that the absence of manifest leadership performance is not sufficient evidence for the absence of essential leadership capacities; instead, context matters. For example, Winston Churchill’s leadership capacities manifested in the context of war, but arguably, not in peace. Under this common contingency view of leadership, leadership behaviors are the result of a combination of leaders possessing certain capacities, together with the existence of appropriate contexts in which leaders can act suitably (Hoy and Miskel, 2005: 375–409). One problem with this view is that it presupposes that the relationship between purported traits and their contexts for manifestation is well understood. However, this is not the case. At the center of the issue is the need to understand the relationship between individual agency and the structures that contexts impose. Generally speaking, social science does not always permit an allocation of causal responsibility to leaders concerning particular outcomes in causally complex situations.

A further problem for a distinctive ontology of leadership is that there is evidence that leadership is a relational property and, therefore, not intrinsic to leaders. To give an example of this sort of property, being a husband is not an intrinsic property of a man, but rather a relational property that depends on having a wife. That is, husbandhood is not some kind of latent capacity that is manifest by becoming married, as a contingency explanation would have it. Rather, one acquires the property of being a husband when one marries. In the same vein, it is argued that leadership cannot be conceptualized in the absence of followership. The parallel with husbandhood is particularly close where we are dealing with leadership in the context of constitutively defined roles, that is, roles defined by contracts, legislation, or some network of social rules. Leadership will thus only make sense within the relational leader–follower nexus.

Recent work on distributed leadership makes an ontology of leadership even more problematical by combining causal complexity with relational conceptualizations. Lakomski (2005) has argued in a very detailed way that much of the evidence we have for making attributions of leadership is, in fact, evidence for no more than coordinated behavior attributable to very local cues. To give an example, the flocking behavior of birds looks like it critically depends on there being a lead bird that all the others follow. However, a closer analysis reveals that flocking can be simulated in models where birds are merely responding to what their immediate flockmates do (Reynolds, 1987). They do not get too close or too far away from each other, maintain the same speed, and head in the average direction of those that are nearest (Lakomski, 2005: 122–123). Lakomski’s point is not that all organizational leaders are ontologically similar to lead birds in terms of leadership attributions. Rather, she is objecting to the methodological presumption, a kind of

commonsense consensus, that they are not. This is the sort of consensus that Resnick calls the centralized mindset and says that it “is not just a misconception of the scientifically naïve. It seems to affect the thinking of nearly everyone” (Resnick, 1997: 4).

The upshot of these arguments is that the ontology of leadership may fragment in much the same way as causal attributions of leaders’ contributions to the achievement organizational goals in the presence of variable contingencies. That is, leaders may require some attributes in some contexts and other attributes in other contexts. They may be similar to lead birds in some organizational settings and like leaders in others. There may be no general theory of leadership from which we can read off the required ontological posits.

Epistemology, Ethics, and Leadership

In discussing the nature of leadership theory earlier, it was clear that epistemology had a significant part to play in shaping the structure of these theories. However, epistemological considerations also figure in determining, or adjudicating, the content of leadership theories. For example, theories that require of leaders knowledge that outruns the resources posited by good epistemology imply that leaders are required to know things that they cannot know.

As an example of this kind of critique, consider the case of transformational leadership. Proposed in many places as an adjunct to implementing rapid change, particularly varieties of school-based management which are subject to the twin demands of autonomy and accountability, transformational leadership purports to promote organizational goals by transforming the followership in four distinctive ways. Such leaders act as an idealized influence, being trusted and respected, and indeed charismatic; they exercise inspirational motivation, changing expectations and forging a shared vision; they provide intellectual stimulation for problem solving and creativity; and they attend to the needs of individuals (Leithwood *et al.*, 1996). In examining this theory, epistemologists would ask how someone could ever acquire the knowledge required to be an effective transformational leader. Consider the matter of promoting a shared vision. Apart from the trivial sense where vision is analytically related to the nature of the organization, visions can sometimes be mistaken, and when they are not, the means chosen to achieve them, such as outcomes from problem solving, can be. In the absence of leader infallibility, organizational mechanisms for error detection and correction become a vital epistemic tool for achieving organizational outcomes.

This raises the more fundamental question of how an organization goes about facilitating error detection and

correction. The difficulty is that on the one hand, any kind of leadership that influences others on a course of action contributes to decisive decision making and efficient implementation, while on the other hand, leadership that has the power to influence the actions of others also has the power to privilege a leader’s ideas, thus contributing to confirmation bias, a form of groupthink where potential error signals are construed as confirming a leader’s vision and strategy, even when these are wrong, hence impeding organizational learning. Hutchins (1995: 260–261) regards the trade-off between effective decision making and reducing confirmation bias as a basic organizations issue, varying in terms of an organization’s goals and its value commitments. This epistemic argument thus pits a preferred model of fallible leadership against the demands for organizational learning.

Also raised is the question of the place of values in leadership theory. As we saw earlier, traditional science approaches to theory located values beyond the scope of theory, owing to a belief in a sharp distinction between fact and value (Simon, 1976: 45–60). The logical empiricist epistemology that underwrites the traditional science of administration sees theories as being justified by appeal to facts, but value justification lies beyond appeal to facts. Values reflect an attitude to facts; they are not justified by facts. Or so the story goes. In response to this argument, Hodgkinson (1978, 1983, 1996), in his many writings, has concluded that administration is therefore not a science at all, but rather a humanism. Furthermore, he claims that there is an isomorphism between epistemic access to levels in a hierarchy of the types of moral decision making and one’s position in an organizational hierarchy. Administrative leadership, as opposed to management, invokes awareness of the highest level of moral knowledge, the type I values that are transrational and intuited as the right courses of action (Hodgkinson, 1996: 157).

Most writers in the field accept some distinction between facts and values. However, those who defend coherentist models of justification blur the distinction, seeing the justification of moral claims as on a par with the justification of other theoretical empirical claims; namely, they are justified if they are part of a more coherent rival theory (Evers, 2003b). Those of more Aristotelian persuasions defend ethical claims by appeal to the good character of leaders (Sergiovanni, 2006). Those in the critical theory tradition defend ethical claims by appeal to the extent to which the normative requirements of maintaining an ideal speech situation in organizational life are attained (Foster, 1986). In addition, those within the Deweyan tradition defend such claims by appeal to their role in describing the conditions that best serve to promote the growth of knowledge (Willower, 1994). Perhaps the most explicit appeals to the ideas of philosophy are to be found in discussions over the place of ethics in leadership theory and practice.

Conclusion

Philosophical ideas about the nature of leadership theory, of its ontology, epistemology and ethics, or axiology, are at the center of theoretical discussions of educational leadership, and presumed by other types of discussions. The characteristic methodologies of philosophers therefore have the potential to make a decisive contribution to thinking about educational leadership.

Bibliography

- Begley, P. T. and Johansson, O. (eds.) (2003). *The Ethical Dimensions of School Leadership*. Dordrecht: Kluwer.
- Chaitin, G. (1975). Randomness and mathematical proof. *Scientific American* **232**(5), 47–52.
- Churchland, P. M. and Hooker, C. A. (eds.) (1985). *Images of Science*. Chicago, IL: University of Chicago Press.
- Dennett, D. C. (1991). Real patterns. *Journal of Philosophy* **87**(1), 27–51.
- Donmoyer, R. (2001). Evers and Lakomski's search for leadership's holy grail (and the intriguing ideas they encountered along the way). *Journal of Educational Administration* **39**(6), 554–572.
- Donmoyer, R., Imber, M., and Scheurich, J. J. (eds.) (1995). *The Knowledge Base in Educational Administration: Multiple Perspectives*. Albany, NY: State University of New York Press.
- English, F. (2003). *The Postmodern Challenge to the Theory and Practice of Educational Administration*. Springfield, IL: Charles C. Thomas.
- Evers, C. W. (1999). From foundations to coherence in educational research methodology. In Keeves, J. P. and Lakomski, G. (eds.) *Issues in Educational Research*, pp 264–279. Oxford: Pergamon.
- Evers, C. W. (2003a). Philosophical reflections on science in educational administration. *International Studies in Educational Administration* **31**(3), 29–41.
- Evers, C. W. (2003b). Naturalizing ethical judgment: A neurocomputational perspective. In Samier, E. (ed.) *Ethical Foundations for Educational Administration*, pp 235–252. London: Routledge/Falmer.
- Evers, C. W. and Katyal, K. R. (2007). Paradoxes of leadership: Contingencies and critical learning. *South African Journal of Education* **27**(3), 377–390.
- Evers, C. W. and Lakomski, G. (1991). *Knowing Educational Administration*. Oxford: Pergamon.
- Foster, W. (1986). *Paradigms and Promises*. Buffalo, NY: Prometheus.
- Godfrey-Smith, P. (2003). *Theory and Reality: An Introduction to the Philosophy of Science*. Chicago, IL: University of Chicago Press.
- Greenfield, T. B. and Ribbins, P. (eds.) (1993). *Greenfield on Educational Administration*. London: Routledge.
- Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education* **33**(3), 329–351.
- Hallinger, P. and Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement* **9**(2), 157–191.
- Halpin, A. W. (1966). *Theory and Research in Administration*. New York: Macmillan.
- Hanson, N. R. (1958). *Patterns of Discovery*. Cambridge: Cambridge University Press.
- Hodgkinson, C. (1978). *Towards a Philosophy of Administration*. Oxford: Blackwell.
- Hodgkinson, C. (1983). *The Philosophy of Leadership*. Oxford: Blackwell.
- Hodgkinson, C. (1991). *Educational Leadership: The Moral Art*. Albany, NY: State University of New York Press.
- Hodgkinson, C. (1996). *Administrative Philosophy*. Oxford: Pergamon.
- Hoy, W. K. and Miskel, C. G. (2005). *Educational Administration: Theory, Research and Practice*, 7th edn. New York: McGraw-Hill.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Kuhn, T. (1970). *The Structure of Scientific Revolutions*, 2nd edn. Chicago, IL: University of Chicago Press.
- Lakomski, G. (2005). *Managing without Leadership: Towards a Theory of Organizational Functioning*. Oxford: Elsevier.
- Leithwood, K., Tomlinson, D., and Genge, M. (1996). Transformational school leadership. In Leithwood, K., Chapman, J., Corson, D., Hallinger, P., and Hart, A. (eds.) *International Handbook of Research on Educational Leadership and Administration*, pp 785–840. Dordrecht: Kluwer.
- Macmillan, R. (ed.) (2003). *Questioning Leadership: The Greenfield Legacy*. London: Althouse.
- Maxcy, S. J. (ed.) (1994). *Postmodern School Leadership: Meeting the Crisis in Educational Administration*. Westport, CT: Praeger.
- Maxcy, S. J. (2001). Educational leadership and management of knowing: The aesthetics of coherentism. *Journal of Educational Administration* **39**(6), 573–588.
- Morrison, K. (2002). *School Leadership and Complexity Theory*. London: Routledge/Falmer.
- Popper, K. R. (1959). *The Logic of Scientific Discovery*. London: Hutchinson.
- Quine, W. V. (1948). On what there is. *Review of Metaphysics* **2**(5), 21–38.
- Quine, W. V. (1951). Two dogmas of empiricism. *Philosophical Review* **60**, 20–43.
- Quine, W. V. (1960). *Word and Object*. Cambridge, MA: MIT Press.
- Resnick, M. (1997). *Turtles, Termites, and Traffic Jams: Explorations in Massively Parallel Microworlds*. Cambridge, MA: MIT Press.
- Reynolds, C. (1987). Flocks, herds and schools: A distributed behaviour model. *Computer Graphics*, **21**. <http://www.red3d.com/cwr/boids/> (accessed May 2009).
- Samier, E. (ed.) (2003). *Ethical Foundations for Educational Administration*. London: Routledge/Falmer.
- Searle, J. (1995). *The Construction of Social Reality*. New York: Free Press.
- Searle, J. (1998). *Mind, Language and Society*. New York: Basic Books.
- Sergiovanni, T. J. (2006). *The Principals'hip: A Reflective Practice Perspective*. Boston, MA: Pearson/Allyn and Bacon.
- Simon, H. A. (1976). *Administrative Behavior*, 3rd edn. New York: Macmillan.
- Starratt, R. J. (2004). *Ethical Leadership*. San Francisco, CA: Jossey-Bass.
- Toulmin, S. (1972). *Human Understanding*. Princeton, NJ: Princeton University Press.
- Toulmin, S. (1993). The construal of reality: criticism in modern and postmodern science. In Mitchell, W. (ed.) *The Politics of Interpretation*, pp 99–117. Chicago, IL: University of Chicago Press.
- Willower, D. J. (1994). Values, valuation and explanation in school organizations. *Journal of School Leadership* **4**(5), 466–483.
- Witziers, B., Bosker, R. J., and Kruger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly* **39**(3), 398–425.

Further Reading

- Busher, H. (2006). *Understanding Educational Leadership*. Maidenhead: Open University Press.
- Caldwell, B. J. (2006). *Re-imagining Educational Leadership*. Thousand Oaks, CA: Sage.
- Davies, B. (ed.) (2005). *The Essentials of School Leadership*. London: Paul Chapman.
- Davies, B. and West-Burnham, J. (eds.) (2003). *A Handbook of Educational Leadership and Management*. London: Pearson Education.
- English, F. W. (2003). *The Postmodern Challenge to the Theory and Practice of Educational Administration*. Springfield, IL: CC Thomas.
- English, F. W. (ed.) (2005). *The Sage Handbook of Educational Leadership: Advances in Theory, Research and Practice*. Thousand Oaks, CA: Sage.
- Fullan, M. (2006). *Turnaround Leadership*. San Francisco, CA: Jossey-Bass.

- Gronn, P. C. (2002). *The New Work of Educational Leaders: Changing Leadership Practice in an Area of School Reform*. London: Paul Chapman.
- Leithwood, K., Jantzi, D., and Steinbach, R. (eds.) (2006). *Changing Leadership for Changing Times*. Maidenhead: Open University Press.
- Murphy, J. and Louis, K. S. (eds.) (1999). *Handbook of Research on Educational Administration*. San Francisco, CA: Jossey-Bass.

Relevant Websites

- <http://www.gu.edu.au> – Asia Pacific network of principals' and leadership Centres: Centre for leadership and management in education.
- www.edfac.unimelb.edu.au – Australian Council of Educational Administration.
- <http://www.ec.tased.edu.au> – Australian Principals Association Professional Development Council.
- <http://www.apcentre.edu.au> – Australian Principals Centre Ltd.
- <http://www.sofweb.vic.edu.au> – Australian Principals Centre. (It is an alliance of the Department of Education, The University of Melbourne and principals associations.)
- www.sagepub.co.uk – British Educational Management and Administration Society.
- <http://www.edfac.unimelb.edu.au> – Centre for Applied Educational Research, The University of Melbourne.
- <http://www.trinity.edu> – Center for Educational Leadership, Trinity University.
- <http://www.udel.edu> – Center for Educational Leadership and Policy, University of Delaware.
- <http://www.qed.qld.gov.au> – Centre for Leadership Excellence, The University of Queensland.
- <http://www.edfac.unimelb.edu.au> – Centre for Organizational Learning and Leadership (COLL), The University of Melbourne.
- <http://www.qut.edu.au> – Centre for Policy and Leadership Studies, Queensland University of Technology.
- <http://www.unc.edu> – Center for School Leadership Development.
- <http://curry.edschool.virginia.edu> – Center for the Study of Leadership and Ethics, University of Virginia.
- <http://eric.uoregon.edu> – Eric Clearinghouse on Educational Management (ERIC/CEM), College of Education, University of Oregon.
- <http://www.edlchange.org> – International Center for Educational Leadership and Social Change.
- <http://www.daggett.com> – International Center for Leadership in Education, Inc.
- <http://www.ulh.ac.uk> – International Educational Leadership Centre, Lincoln School of Management.
- <http://www.jbp.com> – International Network of Principals' Centers, Harvard Graduate School of Education.
- <http://www.ioe.ac.uk> – London Leadership Centre, Institute of Education, University of London.
- <http://www.ioe.ac.uk> – Management Development Centre, Institute of Education, University of London.
- <http://www.ncrel.org> – National Center for Educational Leadership (NCEL), Harvard University.
- <http://www.ncsl.org.uk> – National College for School Leadership.
- <http://www.iteachnet.com> – Principals' Training Center.
- <http://www.leadership.sa.edu.au> – South Australian Centre For Leaders in Education.
- <http://www.shu.ac.uk> – The Centre for Education Management and Administration (CEMA), Sheffield Hallam University.
- <http://www.ucea.org> – UCEA (University Council For Educational Administration) Center for International Development in Educational Administration; UCEA Center for the Study of Leadership and Ethics; UCEA Center for the Study of Leadership in Urban Schools.
- <http://www.vanderbilt.edu> – Vanderbilt International Institute for Principals, Vanderbilt University.

Ethical/Moral Issues in Educational Leadership

R J Starratt, Boston College, Chestnut Hill, MA, USA

L Langlois, Laval University, Sainte-Foy, QC, Canada

P Duignan, Australian Catholic University, Sydney, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

The Societal Context of Ethical Leadership

Increased technological advances as well as the relentless rationalization of layers of administrative control of corporate and governmental institutions have led many to question the significance, for human beings, of progress made in the name of economic rationality. This questioning is notably visible in ethical issues surrounding, for example, moral accountability for illegal corporate decisions, when the complex maze of decisions appears to relieve individual administrators from responsibility for their actions; moral accountability for the degrading of the natural environment; moral accountability for the suppression of civic participation of cultural, linguistic, and religious minorities; and moral accountability of commercial and governmental institutions for enormous disparities of wealth and social privilege within their own communities, as well as for fellow humans facing dramatic problems of starvation, displacement, crippling diseases, and tribal violence in the Third World. In a rapidly interdependent, globalizing world, these issues inevitably become moral issues for everyone. The market-economy era, which was supposed to bring with it increased wealth, freedom, and prosperity for all, appears to have delivered its promises to a few at the expense of aggravated inequality and injustice for many. The world is currently mutating under the influence of globalization and increased social and political interdependence and these mutations seem to imply major modifications in society's moral bearings.

Within this globalizing and interdependent social context, educators and schools face a variety of ethical concerns, varying in intensity from country to country and culture to culture. The context of economic development or underdevelopment; of race, gender, language, and cultural inequalities; of differing religious communities and political ideologies; of the level of professionalization and social status of educators; the epistemologies behind various learning theories – these and other contextual factors tend to shape and weight the ethical issues that one finds at the forefront of public debates. What follows in this section are verbal reports from educators and scholars in various countries, intended to provide a small sample of ethical issues and debates within and around the field of education and educational leadership.

In Canada, one debate highlights the developing ambiguities and political ambivalence around the professionalization of teaching. Political disagreements between provincial governments and teacher unions around government control over the content and process of teaching; the establishment or imposition of codes of ethics for teachers, whether by the union or by the government or by negotiation between the two; the evaluation of teachers by administrators using government-designed criteria (e.g., standardized tests), or professional criteria (research-based or otherwise) – all these raise issues around the rights, responsibilities, and ethical identities of teachers as professionals. These rights are inescapably connected to the rights and responsibilities of governments – led by politicians of various ideological commitments who are charged with providing the funding for schools – to impose standards. That brings into focus the rights, responsibilities, and ethical identities of educational leaders whose responsibilities simultaneously overlap, and sometimes conflict with responsibilities to the local and provincial governing agencies, to the rights of students, to the rights of teachers, to the rights of parents, and with the fundamental educating mission of the school.

Many groups including racial minorities, women, indigenous peoples (natives), and religious minorities in Canada feel excluded from a common culture not only because they are economically disadvantaged but because they are different. Some groups demand the incorporation and enforcement of their traditional or religious codes within the dominant school system. Muslim communities ask for their girls to be excused from gym classes; Sikh communities request their sons to be allowed to wear their kirpan.

The special accommodations demanded by immigrant minorities, religious groups, gays and lesbians, and other such identity groups, are typically designed to facilitate the integration of their members into the larger society, not their separation (Normand and Kimlicka, 2003: 221).

Education and educational leadership in Australia and the United States is characterized by continuing ideological tensions between the states (constitutionally, the primary governing agency responsible for education) and the federal government, which is increasingly driving the educational agenda with focus on the national interest. Ostensibly, it constitutes a clash between federal political

and economic ideologies and the more socially oriented tendencies of the state governments, influenced as they are by local political pressures. The ethical challenges in education, generally, parallel those in Canada, for example, development of a national curriculum framework, national testing regimes and standardized forms of reporting to parents; national standards for teaching and school leadership; and, at the same time, the contradictory calls for the greater professionalization of teachers (Legault, 1999).

Much of the debate in educational circles on this emerging national agenda tends to be more ideological and political than ethical and moral. While considerations of care and justice influence the critique of the national initiatives, these tend to be subsumed within the political and ideological rhetoric, which surfaces during state and federal elections. The media's moral compass seems to be directed much more toward recent ethical and moral scandals in the corporate world. There is little evidence of a national dialog and debate on the ethical dimensions/contradictions within the emerging federal policy agenda on education and educational leadership (e.g., student, parental, teacher rights; values, other than economic prosperity, involved in educating the young; etc.).

In differing conditions, educational leaders in South Africa face quite different ethical issues. In rural communities, for example, the lack of technical, financial, and professional resources, as well as the poverty of the communities being served, leaves local educational leaders with multiple, more basic dilemmas, for example choosing to devote meager budgetary funds to provide snacks for children or new textbooks, to include adult literacy programs for the parents of the children in order to increase parental collaboration with the school's literacy education of their children, or to devote some funds to AIDS prevention seminars for adults (including teachers, among whom exists an alarming rate of HIV infection). In South African cities, educational administrators face not only the lack of sufficient funds, but also tribal rivalries and allegiances, among groups of parents, teachers, and students, that challenge administrative attempts at an equitable distribution of resources and influence within the school. Educators face similar ethical issues in other African countries, with the additional problem of political instability, displaced populations, and armed, intertribal and religious violence.

In Sweden, other ethical challenges for educational leaders surface around the long-established mission of the schools to educate for democratic participation in the country's civic life, as the cultivation of that mission encounters traditional attitudes about the role of women in public life, and the more recent influence of the corporate community to concentrate on education for economic and technical proficiency.

Thus, when surveying the ethical challenges educational leaders face, one is confronted by a multiplicity of

perspectives, both analytical, ideological, and political, that tend to reflect the differing societal and internal contexts of schools and the differing perspectives – traditional, progressive, research-based, individualistic, or communitarian focused – on the process of educating the young. These realities suggest the necessity of careful analysis of differing perspectives as a necessary methodology of ethical deliberation, as well as a weighing of consequences of various choices.

Theoretical Developments in the Ethical/Moral Agenda in Education

In the past 20 or so years, there have been significant theoretical developments in the area of ethics in education in general. The work of Kohlberg (1981) and his associates at Harvard University highlighted the cognitive development of reasoning about ethical issues of justice, bridging the earlier work of Piaget and Inhelder (1969) on cognitive development with Rawls' (1971) theory of justice. Kohlberg developed an approach to a just community school in which he attempted to encourage a pedagogy in line with his stage-theory of cognitive moral development. One of his students, Carol Gilligan developed an alternative view of women's moral development, one characterized by a primary concern with relationships (Gilligan, 1982; Gilligan *et al.*, 1988). This highlighting of women's ways of knowing and thinking (Belenkey, 1986) was further reinforced by the work of Nel Noddings (1984) on the ethics of caring.

In the meantime, a number of scholars' voices had been raised in various criticisms of schools, criticisms that pointed to deep structural and cultural inequities in schools that, in turn, were rooted in the attitudes and beliefs that underlay inequities in society at large. These critics would agree with Karen Lebacqz (1987: 10) who argued that, "reflections on justice must begin with the realities of injustice." This approach to ethics in education begins with a view of a world already steeped in ethnic and racial injustice, sexual injustice, economic injustice, political injustice, and cultural injustice – each strand of injustice often interwoven with one or several other strands, creating a social fabric of institutionalized, legitimated injustice. As Lebacqz suggests, within such a social fabric, decontextualized discourse about justice cannot help but be distorted by the social context of widespread, if unexamined, injustice. Authors such as Freire (1970), Bowles and Gintis (1976), Apple (1982), Giroux (1983, 1988), Bates (1983), Foster (1987), Spring (1988, 1994), Pagano (1991), Nieto (1992), Purpel (1989) and others provided the analysis of injustice both as it infected the way schools were run as well as the choice of the content and pedagogy of teaching. Their critique implied an ethic, an ethic of resistance and a struggle for emancipation based on the inherent human dignity of those who were institutionally disenfranchised

because of their race, sex, economic class, cultural identity, language, sexual preference, or politics.

In the last decade or so of the twentieth century in the United States, these three ethical discourses seemed at war with each other. Those proposing an ethic of justice upheld a rational, principled approach, in opposition to what they characterized as disjointed, emotional, or subjective responses to moral dilemmas. Those espousing an ethic of care criticized adherents of the ethic of justice as embracing an impersonal, decontextualized, rationalistic form of ethics. The scholars immersed in the ethic of critique accused the proponents, of both an ethic of justice as well as an ethic of care, of naiveté. That is, in specific ethically charged situations in institutions that were already structurally and culturally compromised, a search for justice or affiliative community could not be conducted as though those institutions were ethically neutral.

Emerging Theory of an Ethics Served by Multiple Perspectives

Starratt (1991) argued against the apparent separatism of each ethical position, proposing that each ethic could find its deeper fulfillment only and necessarily in a larger synthesis of the three dominant ethical discourses. If an ethic of justice were to avoid a minimalist approach to estimating what could be considered just, then it would need to be filled out, deepened, or made more humane, by an ethic of care that would ensure a level of generosity in applying the justice formulae. Some morally charged situations calling for the care of a friend or child might also require consideration for the common good, and for complying with the demands of justice. Some situations raise questions not only about the fair application of a school rule, but also about the rule itself, thus bringing the ethic of critique into play. On the other hand, in order to avoid the ranting of the perpetual malcontent or the cynicism of the self-righteous, the ethic of critique needs to be tempered by some sense of compassion for human weakness and moral ineptitude. While some cases will require a heavier emphasis on an ethic of caring, others will require an emphasis on critique or justice. Still other complex cases will require a combination of all three ethics in their fuller resolution.

Subsequently, Strike teamed up with Noddings and Katz to search for common ground between the ethics of justice and caring (Katz *et al.*, 1999). While the essays in that book were too diverse to construct a coherent umbrella for the two ethics, the book pointed to the complementarity between both ethical approaches. Furthermore, as Strike (1999) pointed out in his essay, it would be a mistake to assume that the ethics of justice and caring encompassed the whole field of educational ethics and its many complex problems. That suggestion

echoed the earlier work by Capper (1993) who had stressed the need for a multi-paradigm approach in her collection of essays that reflected the challenges a pluralistic society pose for educational leaders.

Nash's (1996) ecumenical approach for developing legitimate, but differing bases for understanding the real world of ethical situations, also agrees with Strike's argument for using multiple perspectives. His book on the ethical education of educators and other human service professionals suggests that values and beliefs derived from one's upbringing should be taken seriously, not discounted out of hand. Rather, he would have his students test those ethical beliefs and values in explicit applications to various ethically charged situations. He encourages, also, an ethical exploration of moral character based on foundational virtues that seem essential to a fully human life. Finally, he had his students explore the application of principled ethical reasoning to various situations. Throughout his book he refused to provide correct answers, insisting that everyone take responsibility for their choices to respond to each situation and provide reasonable defenses for those choices.

In Australia, the work of Haynes (1998) analyzes multiple ethical theories and perspectives but argues that it is not "... useful to treat them as oppositional theories, but as theories that simply pick up different aspects of morality" (1998: 8). She recommends that educational practitioners should learn to reflect on ethical and moral challenges in a rational way, rather than adopt a stance of unthinking acceptance and they should develop their criteria or standards for moral judgment through dialog on the propriety of their actions and those of others.

Shapiro and Stepkovich (2001) turned their experience of team-teaching a course in ethics to prospective school administrators into a book. They echoed Capper, Strike, and Nash in arguing for multiple paradigms, including both a personal and a professional code of ethics. They call attention to the ethic of the profession which is centered around the best interests of the student, thus bringing deliberations derived from all of the ethical perspectives into focus around this core value.

Emerging Awareness of Ethics and Values for Educational Leaders

Starratt (1991) proposed an ethical perspective specifically for school leaders and administrators. That same year, Hodgkinson (1991) published *Educational Leadership: The Moral Art*, considered by some as his best work. Hodgkinson, steeped in classical philosophy, argued that educational leadership embodied by its very nature various moral commitments, and called forth the virtuous character of the leader. No doubt educational leaders needed professional expertise and a commitment to reasoned

decision making. Nevertheless, as leaders of an educational enterprise that involved adults working carefully to effect the human growth of young people, their leadership involved them in the moral character of that work.

The 1990s saw the growth of interest in the moral/ethical side of school administration among scholars and university professors involved in the preparation of school administrators. Strike *et al.* (1988) published their influential work, *The Ethics of School Administration*, which moved much of the ethical analysis developed in their earlier work on the ethics of teaching (1985) into a focus on the work of administrators. Their analyses of various cases of administrators caught up in morally complex cases tended to employ a Kantian emphasis on the inviolability of persons, the utilitarianism of Mill, and the classical liberal theory of justice espoused by Rawls. Beck (1994) applied the ethic of care to the practice of educational administration, showing in the process that educators acting out of an ethic of care often were led also to tackle issues of justice in their schools. In other words, caring about children and their communities led these educators to critique conditions in their school and seek to promote a clearer sense both of respect and fairness to the students and their parents.

The Willower and Licata (1997) work on values in the practice of education emphasized the growing awareness in the field of educational leadership preparation of the need to employ ethics and values within the practice of leadership in education. Previously, the widespread emphasis on the management side of administration, especially as fulfilling bureaucratic functions – budgeting, goal setting, scheduling, monitoring daily operations, resource management, and so forth – had tended to dominate educational administration. Management practice was considered an applied science; as science, it assumed the fact/value dichotomy in human organizations. Managers were to deal in objective facts, not in subjective and emotion-laden values.

That view of administration and leadership was changing. In 1994 and again in 1997, Beck and Murphy surveyed the field of graduate schools of education in the United States to find out whether ethics played a part in their administrative preparation programs. They found that some university programs had initiated a course on ethics in administration, or had introduced ethical consideration into their cases of problem solving in a variety of courses. Nevertheless, the field was clearly involved in an ambivalent catch-up situation, with the question frequently asked, “Whose ethics shall we teach?” The research of Beck and Murphy (1994, 1997) uncovered promising initiatives, though little consistency in the field.

In Australia, while some consideration has been given to the importance of general ethics for educational practitioners, especially, principals, only passing concern has been shown for professional ethics. Evers (1992: 39) was one of the first to directly address the practical implications

of moral philosophy and ethical theories for educational leaders. He recommended some “. . . moral requirements for the existence and maintenance of organized learning . . . and . . . that these requirements be used to morally evaluate educative leadership.” His approach in developing these requirements reflected a pragmatic realistic tradition drawing on the work of Dewey and Popper.

Recent Empirical Research in Ethical Educational Leadership

Langlois followed up empirical research begun in her doctoral dissertation (Langlois, 1997) on Starratt’s tri-dimensional ethic of justice, care, and critique, with further quantitative studies of superintendents and system directors in Canada. Her findings (2004) show that administrators employ two and three ethics to solve their moral dilemma, thus tending to confirm Starratt’s position about the complementarity of the ethics of justice, care, and critique. Her continuing research has identified the ethical dimensions and delineated the decision-making process among several samples of school administrators and thus has put some empirical meat on the skeleton suggested by Starratt. Her results (Langlois, 2004; Langlois and Lapointe, 2006) indicate that varieties of ethical responses are related very much to the given context. Variables such as age and experience exercise a determining ethical influence whereas gender has less impact. Early on in their careers, administrators tend to refer to a normative framework to justify their decisions. The more administrators gain experience, the more they seem to be able to free up their moral imagination and use multiple ethical lenses to better analyze a situation. Evaluating the consequences of their decisions and making room for reflection on ethics, two aspects close to Jonas’ philosophy, become part of their decision-making process.

Walker and his associates (Walker and Quong, 1998; Walker and Dimmock, 2005; Walker, 2006) have uncovered in their research in Hong Kong that principals face pressures to conform to traditional culture in a city that is changing rapidly. Principals’ ethical leadership in schools characterized by multicultural diversity requires both a pragmatic, ground-up approach to responding to issues of social justice and courage to tackle issues of racism, poverty, and ethnic differences head-on.

Most research-based work on ethical leadership in Australia focuses on principals’ ethical leadership practice, and on ways of preparing educational leaders for contemporary ethical challenges and dilemmas. Duignan (Duignan *et al.*, 2003), based on a three-year research project with school principals, has provided a framework for the analysis of ethical dilemmas and recommended a formation approach for principals to enable them to resolve situations with ethical tensions. Duignan has developed these ideas further

in his most recent book (Duignan, 2007), where he argues that ethical educational leaders need to be, first and foremost, capable and authentic human beings. He proposes a leadership development approach based on leadership capabilities, not just competencies.

In a recent PhD study on ethical leadership and principals (Branson, 2006), echoed Duignan's focus on the person of the principal when he concluded that "arguably, truly ethical leadership might only be possible for those leaders who have the commitment and courage to come to know and understand the full extent of the influential power of their whole inner Self" (p. 2). He recommends that principals "... engage in some professional development experiences incorporating a process of deeply structured self-reflection" (p. 20).

Cranston *et al.* (2004, 2006), focusing their research on the ethical dilemmas faced by seven independent school heads in Australia, reported that ethical dilemmas, usually involving staff and/or student issues, have become daily challenges of educational leaders' lives. They also developed an analytical framework for educational leaders to understand and respond to the dynamics of ethical dilemmas. They recommended problem-based professional development programs in ethics and ethical decision making for principals.

Wildy *et al.* (2001: 12), reporting on parallel research work on the moral dimensions of principals' work in two Australian states, conclude that education reform efforts should focus more on "... the pivotal role of the relationships formed by principals with people inside and outside their schools" (they refer to this as developing social capital) and that professional development programs for principals should engage principals in case analysis that "... is grounded in everyday recognisable [ethical] problems and dilemmas to stimulate reflection on practice." (pp. 10–11).

Research by Dempster (2001), Dempster and Berry (2003), and Dempster *et al.* (2004a, 2004b) concludes that when faced with ethical dilemmas principals, overall, do not display consistent conceptual knowledge of major ethical theories and they require professional development that educate them on the concepts and processes required for consistent, confident, and well-reasoned ethical decision making (2004a: 450). Further, principals are left "... unprepared and often unsupported when they are confronted with decisions that are influenced by macro-political, social and market-oriented forces" (2004b: 163).

Burford (2005: 18), suggested that a major moral challenge and ethical dilemma facing primary principals in Australia is to "... confront and attempt to create authentic and transforming learning for children," in their schools. He refers to an ongoing project engaging key educational leaders in clusters of schools in New South Wales called leaders transforming learners and learning (LTLL), which uses Starratt's (2004) virtues of authenticity, responsibility, and presence as its energizing ethical source.

Burford cautions, however (2006), that there is a lack of clarity in what we mean in our language of ethics and leadership and how we communicate that meaning to the field. In this, he points to a recently recognized need by scholars working in the field of ethical educational leadership to attempt an empirically grounded international effort to develop a more commonly accepted language and perspectives that will prove useful to practitioners and scholars alike.

International Scholarly Networks Focused on Ethical Educational Leadership

The University Council for Educational Administration (UCEA), a national association of research universities in the United States that grant the doctorate in educational administration, has supported scholarly attention to ethical issues confronting school leaders. Themes that ground the annual conferences of UCEA have tended in the last decade or more to bring issues of social justice to the forefront of the agenda of educational leadership, issues such as school inequities based on race, sex, social class, disabilities, immigrant status, anti-democratic school practices, and so forth. UCEA's partnerships with associations of leadership scholars in other countries have stimulated a growing international synergy among scholars around ethical issues as an important area for ongoing research.

Furthermore, UCEA has supported several university-based research centers, one of which, The Center for the Study of Leadership in Ethics, has become an international source and stimulus specifically for research in the ethics of leadership. That center includes other partners in Sweden, Hong Kong, Denmark, Canada, and Australia. The Center has produced in the past 3 years six academic books edited by center associates around themes of ethics and values in educational administration, and six special issues of refereed journals focusing on ethics and values in educational leadership, besides numerous papers delivered by center associates at various national and international conferences (Begley, 2006). The scholarly ferment stimulated by the Center has helped develop a growing network of scholars collaborating on further research in the ethics of leadership, which contains the promise of continuing developments in this field (Begley and Johansson, 2003).

Needed Initiatives/Developments in the Field

Recently, Johansson (2006), Langlois (2006), and Burford (2006) have suggested that the field may be ready, given the groundwork of steadily increasing international collaboration and sharing of ideas and studies, to embark on a more structured form of international collaboration, one

that seeks to establish, with greater scholarly rigor, a significant empirical base across a variety of national and regional contexts that will illuminate a grounded theory for ethical educational leadership. Such a theory cannot suggest, given the significant differences across countries and cultures, a uniform, hierarchical formula for deliberation or for practice. Nonetheless, the empirical studies could begin to suggest various underlying commonalities that, when the context is sufficiently taken into account, point to what might be called a meta-ethics of professional responsibility, or, in more technical terms, professional virtue.

In order to gain scholarly credibility and legitimacy within the academy, however, such international studies would initially have to fit within an overall plan agreed to by the researchers involved. Such a plan would, following the suggestions of Johansson (2006) and Langlois (2006), conform to accepted methodological rigor for empirical studies, allowing for a legitimate meta-analysis of studies that revealed adherence to similar methodological criteria. Those studies would follow, for example, the following methodological structure: (1) study the same research questions (with the understanding that the terminology used in the question might have to differ to convey the similar inquiry but in the vocabulary appropriate to a specific region or culture); (2) employ the same sampling methods; (3) employ the same data-gathering instruments and methods; (4) use the same data-analysis methods; (5) use the same data-reporting methods and formats; (6) seek for comparisons and contrasts with existing studies that resembled their own; and (7) seek for policy and practice recommendations/ implications that might flow from their findings, especially as those might concern the school renewal efforts in the various countries and regions of the study.

Bibliography

- Apple, M. W. (1982). *Education and Power: Reproduction and Contradiction in Education*. London: Routledge and Kegan Paul.
- Bates, R. (1983). *Educational Administration and the Management of Knowledge*. Geelong, VIC: Deakin University Press.
- Beck, L. G. (1994). *Reclaiming Educational Administration as a Caring Profession*. New York: Teachers College Press.
- Beck, L. G. and Murphy, J. (1994). *Ethics in Educational Leadership Programs: An Expanding Role*. Thousand Oaks, CA: Corwin Press.
- Beck, L. G., Murphy, J. and Associates (1997). *Ethics in Educational Leadership Programs: Emerging Models*. Columbia, MO: University Council for Educational Administration.
- Begley, P. T. (2006). *UCEA Program Center Summative Review, 2003–2006*. State College, PA: D.J. Willower Center for the Study of Leadership and Ethics.
- Begley, P. T. and Johansson, O. (eds.) (2003). *The Ethical Dimensions of School Leadership*. Dordrecht: Kluwer.
- Belenkey, M. F. (1986). *Women's Ways of Knowing: The Development of Self, Voice, and Mind*. New York: Basic Books.
- Bowles, S. and Gintis, H. (1976). *Schooling in Capitalist America: Education Reform and the Contradictions of Economic Life*. New York: Basic Books.
- Branson, C. M. (2006). Structured self-reflection as a cornerstone of ethical leadership development. Paper delivered at the 11th Annual Values and Leadership Conference, Victoria, BC, Canada, October 2006.
- Burford, C. (2005). Engaging children in their learning: Engaged to learn what? Moral challenges for leaders and learners in Australian schools. Paper delivered at the Australian Primary Principals Association National Conference, Canberra, ACT, Australia, June 2005.
- Burford, C. (2006). The moral agency of the educational leader: Observations. Paper delivered at the 11th Annual Values and Leadership Conference, Victoria, BC, Canada, October 2006.
- Capper, C. (ed.) (1993). *Educational Administration in a Pluralistic Society*. Albany, NY: State University of New York Press.
- Cranston, N., Ehrich, L., and Kimber, M. (2004). Paper delivered at the Australian Association for Research in Education Conference, Melbourne, VIC, Australia, October 2004.
- Cranston, N., Ehrich, L., and Kimber, M. (2006). Ethical dilemmas: The "bread and butter" of educational leaders' lives. *Journal of Educational Administration* 44(2), 106–121.
- Dempster, N. (2001). The ethical development of school principals. Paper delivered to the International Leadership Institute, Adelaide, SA, Australia, August 2001.
- Dempster, N. and Berry, V. (2003). Blindfolded in a minefield: Principals' ethical decision making. *Cambridge Journal of Education* 33(3), 457–477.
- Dempster, N., Carter, C., Freakley, M., and Parry, L. (2004a). Conflicts, confusions and contradictions in principals' ethical decision making. *Journal of Educational Administration* 42(4), 450–461.
- Dempster, N., Carter, C., Freakley, M., and Parry, L. (2004b). Contextual influence on school leaders in Australia: Some data from a recent study of principals' ethical decision making. *School Leadership and Management* 24(2), 163–174.
- Duignan, P. (2007). *Educational Leadership: Key Challenges and Ethical Tensions*. New York: Cambridge University Press.
- Duignan, P., Burford, C., Cresp, M., et al. (2003). *Contemporary Challenges and Implications for Leaders in Frontline Human Service Organizations*. Sydney, NSW: Australian Catholic University.
- Evers, C. W. (1992). Ethics and ethical theory, in educative leadership. In Duignan, P. A. and Macpherson, R. J. S. (eds.) *Educative Leadership: A Practical Theory for New Administrators and Managers*, 21–43. London: Falmer.
- Foster, W. (1987). *Paradigms and Promises: New Approaches to Educational Administration*. Buffalo, NY: Prometheus Books.
- Freire, P. (1970). *Pedagogy of the Oppressed*. Ramos, M. B. (trans.). New York: Herder and Herder.
- Furman, G. (2003). What is leadership for? *UCEA Review* 55(1), 1–6.
- Giroux, H. (1983). *Theory and Resistance in Education: A Pedagogy of Opposition*. South Hadley, MA: Bergin and Garvey.
- Giroux, H. (1988). *Schooling and the Struggle for Public Life: Critical Pedagogy in the Modern Age*. Minneapolis, MN: University of Minnesota Press.
- Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University Press.
- Gilligan, C., Ward, J., and Taylor, J. (1988). *Mapping the Moral Domain: A Contribution of Women's Thinking to Psychology and Education*. Cambridge, MA: Harvard University Press.
- Haynes, F. (1998). *The Ethical School*. London: Routledge.
- Hodgkinson, C. (1991). *Educational Leadership: The Moral Art*. Albany, NY: State University of New York Press.
- Johansson, O. (2006). Conducting research on values, ethics and moral leadership. Paper delivered at the 11th Annual Values and Leadership Conference, Victoria, BC, Canada, October 2006.
- Katz, M. S., Noddings, N., and Strike, K. A. (eds.) (1999). *Justice and Caring: The Search for Common Ground in Education*. New York: Teachers College Press.
- Kohlberg, L. (1981). *The Philosophy of Moral Development: Moral Stages and the Idea of Justice*. San Francisco, CA: Harper and Row.
- Langlois, L. (1997). *Relever les défis de la gestion scolaire d'après un modèle de leadership éthique: une étude de cas*. Thèse de doctorat, Université Laval, Québec.

- Langlois, L. (2004). Responding ethically: Complex decision-making by school district superintendents. *International Studies Educational Administration Management, ISEA* 32(2).
- Langlois, L. (2006). The goals of ethics and related challenges: A scientific work in progress. Paper delivered at the 11th Annual Values and Leadership Conference, Victoria, BC, Canada, October 2006.
- Langlois, L. and Lapointe, C. (2006). Ethical leadership in Canadian school organizations: Tensions and possibilities. Special Issue of *Educational Management, Administration, and Leadership*, Cross-Cultural Perspectives on Authentic School Leadership. Sage publications.
- Lebacqz, K. (1987). *Justice in an Unjust World: Foundations for a Christian Approach to Justice*. Minneapolis, MN: Augsburg.
- Legault, G. (1999). *Crise d'identité professionnelle et professionnalisme*. QC, Canada: Presses de l'Université du Québec.
- Nash, R. J. (1996). *Real World Ethics: Frameworks for Educators and Human Service Professionals*, 2nd edn. New York: Teachers College Press.
- Nieto, S. (1992). *Affirming Diversity: The Socio-Political Context of Multicultural Education*. New York: Longmans.
- Noddings, N. (1984). *Caring: A Feminine Approach to Ethics and Moral Education*. Berkeley, CA: University of California Press.
- Normand, W. and Kimlicka, W. (2003). Citizenship. In Grey, R. and Wellman, C. H. (eds.) *A Companion to Applied Ethics*, pp 210–223. London: Blackwell.
- Pagano, J. (1991). *Exiles and Communities: Teaching in the Patriarchal Wilderness*. Albany, NY: State University of New York Press.
- Piaget, J. and Inhelder, B. (1969). *The Psychology of the Child*. Weaver, H. (trans.). New York: Basic Books.
- Purpel, D. E. (1989). *The Moral and Spiritual Crisis in Education: A Curriculum for Justice and Compassion in Education*. New York: Bergin and Garvey.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Belknap Press of Harvard University Press.
- Shapiro, J. P. and Stefkovich, J. A. (2001). *Ethical Leadership and Decision Making in Education*. Mahwah, NJ: Lawrence Erlbaum.
- Spring, J. H. (1988). *Conflict of Interests: Politics of American Education*. New York: Longmans.
- Spring, J. H. (1994). *Deculturalization and the Struggle for Equality: A Brief History of the Education of Dominated Cultures in the United States*. New York: McGraw-Hill.
- Starratt, R. J. (1991). Building an ethical school: A theory for practice in educational leadership. *Educational Administration Quarterly* 27(2), 185–202.
- Starratt, R. J. (2004). *Ethical Leadership*. San Francisco, CA: Jossey-Bass.
- Strike, K. A. (1999). Justice, caring, and universality; in defense of moral pluralism. In Katz, M. S., Noddings, N., and Strike, K. A. (eds.) *Justice and Caring: The Search for Common Ground in Education*, pp 21–36. New York: Teachers College Press.
- Strike, K. A., Haller, E. J., and Soltis, J. F. (1998). *The Ethics of school Administration*. New York: Teachers College press.
- Walker, A. (2006). Leadership values in a multicultural setting. Paper delivered at the 11th Annual Leadership & Ethics Conference, Victoria, BC, Canada, October 2006.
- Walker, A. and Dimmock, C. (2005). Leading the multiethnic school: Research evidence on successful practice. *The Educational Forum* 69, 291–304.
- Walker, A. and Quong, T. (1998). Valuing differences: Strategies for dealing with the tensions of educational leadership in a global society. *Peabody Journal of Education* 73(2), 81–105.
- Wildy, H., Loudon, W., Dempster, N., and Freakley, M. (2001). *The Moral Dimensions of School Principals' Work: Standards, Cases and Social Capital*. Deakin West, ACT: The Australian College of Educators.

Further Reading

- Bates, R. (2003). Morals and markets: Adam Smith's moral philosophy as a foundation for administrative ethics. In Samier, E., Mihailova, C., and Hodgkinson, C. (eds.) *Ethical Foundations for Educational Administration*, pp 97–125. London: Routledge Falmer.
- Bates, R. (2006). Educational administration and social justice. *Education, Citizenship and Social Justice* 1(2), 1–141.
- Jonas, H. (1990). *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Chicago, IL: D. Her.
- Mulkeen, T., Cambron-McCabe, N. H., and Anderson, B. (eds.) (1994). *Democratic Leadership: The Changing Context of Administrative Preparation*. Norwood, NJ: Ablex.
- Murphy, J. (1999). *The Quest for a Center: Notes on the State of the Profession of Educational Leadership*. Columbia, MO: University Council for Educational Administration.
- Starratt, R. J. (2006). Cultivating the moral character of learning: A neglected dimension of educational leadership. Paper delivered at the conference of the Commonwealth Council on Educational Administration and management, Nicosia, Cyprus, October, 2006.
- Strike, K. and Soltis, J. F. (1985). *The Ethics of Teaching*. New York: Teachers College Press.

History of Educational Leadership/Management

R Bates, University of Deakin, Waurn Ponds, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

In 1902, John Dewey warned that “it is easy to fall into the habit of regarding the mechanics of school organization and administration as something comparatively external and indifferent to educational ideals” (Dewey, 1902: 22, 23). The history of educational leadership and management as a field of study and practice bears out his misgivings for, despite some exceptions, the field is historically constructed around a strict separation of administrative theory from educational concerns and constituted by a focus on the mechanics of school organization.

The formal study and teaching of educational administration began in the United States in the early years of the twentieth century with the appointment of men such as George Strayer at Teachers College Columbia; Paul Hanus at Harvard; Edward Elliot at Wisconsin; Bobbit at Chicago; and, perhaps most importantly, Ellwood Cubberly at Stanford. Cubberly had little background in the study of education, his own education and profession being that of geology and physical science. This did not prevent him, however, from undertaking to teach “school administration, school problems, school organization, school statistics, secondary schools, history of education, relation of ignorance and crime to education etc.” (Tyack and Hansot, 1982: 124). Partly, no doubt, these topics sprang to mind as a result of the 2 years he spent as superintendent of schools in San Diego, where he came to the conclusion that urban school boards should become nonpolitical. In this, his views coincided with those of the emerging Municipal Reform Movement, a movement that was devoted to the amalgamation of small public enterprises such as schools and school districts into large, hierarchically ordered, and centrally directed corporate systems. The model was that of the large industrial corporation informed by the newly emerging science of administration. When this new science of administration was applied to the public sphere, it isolated administrative questions from political questions: “...administration lies outside the proper sphere of politics. Administrative questions are not political questions” (Woodrow Wilson, 1887: 97).

By 1930, the transformation of American city management, along with the management of schools, had succeeded in displacing local political interests and creating a series of professionalized bureaucracies linked together into national organizations supported by systematic professional education and a national network of influential individuals. Strayer, for example, argued that “significant progress in the administration of city school systems during the past twenty-five years is due primarily to

two causes. First, the application of the scientific method to the problems of administration, and, second, the professional training of school executives” (Tyack and Hansot, 1982: 152).

Carter Alexander, a colleague of Strayer’s at Teachers College, outlined the administrative progressives’ view of the purpose of educational research as being: “to discover, in the light of the purposes of education commonly acknowledged, the most efficient procedures in the organization, supervision, finance and evaluation of the program of educational service” (Tyack and Hansot, 1982: 153). Taken to extremes, this agenda led to what Callaghan (1962) called the cult of efficiency, a cult based on supposed agreement over the purposes of education commonly acknowledged, which were then set aside from the pursuit of efficient administration.

Based upon the *Principles of Scientific Management*, first advocated by Taylor (1911), leadership in education became defined as extracting the maximum efficiency from educational resources, both physical and human. Advocated and publicized widely by leading figures such as Strayer, Cubberly, and Bobbit, this business-managerial conception of educational leadership dominated the emerging university-based programs of training in educational administration. Moreover, while the new professional elite of university-trained administrators set about supplanting the Tammany Hall of local politics and cronyism, they created for themselves and their protégés a national “educational Tammany Hall that made the Strayer–Englehardt Tammany Hall in New York look very weak” (Tyack and Hansot, 1982: 141). This educational trust reformed education from the top down through new university programs, national associations (both public and private), and networks of political influence in key school districts through which the new meritocracy of the administrative progressives was established and sponsorship and patronage exercised in the creation of the new educational elite.

The result was a widespread transformation of leadership in American education that consolidated many small local districts into large systems; imposed the efficiencies of scientific management upon teachers; entrenched central administrative control over curriculum and assessment; and replaced educational ideals with managerial goals. Callaghan’s judgment was that this constituted an American tragedy that was fourfold, one in which: “educational questions were subordinated to business considerations; ... administrators were produced who were not,

in any sense, educators; . . . a scientific label was put on some very unscientific and dubious methods and practices; and . . . an anti-intellectual climate, already prevalent, was strengthened" (1962: 246).

There were, of course, voices raised against this transformation, chief among them being that of John Dewey who insisted that the displacement of educational ideals from the discourse of educational leadership was unfortunate and that the transformation of educational leadership into an oligarchy was inappropriate in a democratic society (Dewey, 1902, 1916).

However, the clash was not a clash of ideas alone. The educational oligarchy developed through the appeal to scientific management and the creation of a professionally trained cadre of managers was overwhelmingly male. Indeed, while one of the objectives of scientific management was the separation of conception from execution and the enhancement of managerial prerogatives (Edwards, 1979), its application in education was designed to separate male managers from overwhelmingly female teachers. This separation was challenged by women such as Ella Flagg Young who, as superintendent of the Chicago School District, provided leadership to women unionists and was a strong champion of female leadership of the profession. One of the battlegrounds for the female challenge to the male oligarchy was the National Education Association (NEA) where Margaret Haley spearheaded an attack on the male old guard and its centralist tendencies arguing that its intention was "to make a despotism of our entire school system" (Tyack and Hansot, 1982: 186). Attempts by women teachers to broaden participatory structures within the NEA met with some success, but were eventually repulsed by male executive members such as Charles Eliot whose patronizing attitude was indicated in his comment that women such as Young and Haley exhibited "a general moral ignorance or incapacity which is apt to be in evidence whenever women get stirred in political, social or educational contexts" (Tyack and Hansot, 1982: 186).

Young and Haley, however, were arguing not simply for greater female participation in administration and leadership, but also for a style of leadership that was aligned with Dewey's observation that in order to sustain a democratic society, a democratic education was fundamental. Their call was for a kind of educational leadership that "would make education the great instrumentality helping children and youths grow into citizenship in a government intended to be of, by and for all" (Young, 1916: 6). In this, they were arguing for the reintroduction of politics into educational leadership and for an acknowledgment that educational ideals were indeed political and an inescapable part of educational leadership.

While the leaders of the Educational Trust were pursuing efficiencies in the management of increasingly large-scale urban school systems and insisting that education was

an activity best left to the professionals, the inescapable links between politics and education were being demonstrated through the leadership of teachers such as Leonard Covello whose political agitation in East Harlem led to the establishment of and his appointment as the first principal of Benjamin Franklin High School in 1934. Covello believed that the school should adapt to the cultures of its students and be a place of mobilization of communities for social justice. Learning was to be through active participation in the community, and politics and learning, democracy, and social justice were to be articulated through action (Covello, 1936, 1958; Peebles, 1978). Leadership was not so much to be demonstrated through the application of principles of efficiency and scientific administration, but through assisting in the articulation of democratic claims through learning and social action in the pursuit of social justice: a very Deweyan perspective.

Meanwhile, in the academy, a new generation of professors, while continuing their studies of school finance and efficient administration through survey techniques, was increasingly concerned with the disparagement of colleagues in the real science departments who looked upon education and educational leadership as unscientific pursuits. The result was a shift toward importing ideas from the social sciences, especially psychology and sociology. By the 1950s, an influential cadre including Jacob Getzels (a psychologist at the University of Chicago), Andrew Halpin (a member of the Personnel Research Board at Ohio State University), and Daniel Griffiths (an iconoclastic professor of administration at Columbia Teachers College) began what became known as the Theory Movement in educational administration. Grounded in a commitment to logical positivism, these professors decried the atheoretical nature of research in educational administration and set about devising an approach both grounded in theory and articulated through research design derived from the emerging social sciences (Culbertson, 1988).

The Theory Movement took the depoliticization of educational leadership one step further through the pursuit of a value-free science of administration which would reveal, through the study of what is, rather than what ought to be, a series of empirically grounded hypothetico-deductive propositions that could form the basis of a proper positivist science of administration. This was seen as a general theory of administration of which educational administration was simply a specific instance.

Supported by the Kellogg Foundation through the Cooperative Program in Educational Administration, eight university-based centers were established at Chicago, Teachers College (Columbia), Harvard, George Peabody, Texas, Oregon, Stanford, and Ohio State in the United States, and at Ontario Institute for Studies in Education (OISE) and Alberta in Canada. These centers were charged with developing scientifically based programs for the preparation of

educational administrators, producing a new literature on educational administration, and supporting the establishment of professional organizations committed to the new approach to educational administration and leadership (Moore, 1964). The effect was to instantiate systems theory and behavioral science as the theoretical foundations of educational administration and leadership.

In essence, the Theory Movement was built on the appropriation of Parsonian systems theory through Getzels' 'A psycho-sociological framework for the study of educational administration' (1952). Here, the abstractions of the social system, nomothetic and ideographic dimensions of social behavior, and those of institution, role, expectation, and individual, personality, and need disposition were elaborated into a theoretical set of systematic relationships for exploration through empiricist methodologies. The 1964 National Survey of Student Engagement (NSSE) yearbook (*Behavioural Science and Educational Administration*; Griffiths, 1964) showed how pervasive systems theory and its associated abstractions had become. Leadership was thus further stripped of its normative concerns and the study of leadership became the study and classification of how leaders behave, divorced from the ethical concerns of what they ought to do (Halpin, 1958: 6).

While the Theory Movement pursued its theoretical abstractions on which a science of administrative and leadership behavior was to be built, educational leaders in the United States were being confronted by the Civil Rights Movement and with the aftermath of Brown versus Board of Education, that is, the landmark legal decision intended to bring to an end the provision of segregated schooling in the United States. The 1964 NSSE Yearbook was published a decade after Brown versus Board of Education; however, there is not a single reference to the impact of Brown versus Board of Education on the field of educational administration and leadership in it. Nor is there a mention of the leadership of education in the Southern states that closed virtually all black schools and sacked 38 000 black teachers and administrators (Hudson and Holmes, 1994; Anderson, 2006). Nor is there the name of a single member of the Theory Movement on the list of social scientists who signed the report prepared by the Black psychologist, Kenneth B. Clark, which was so influential in the Brown versus Board of Education decision.

This disconnection of administrative theory from the world of school leadership was noted as early as 1960 by Andrew Halpin when he observed, with regard to the Theory Movement, that: "There is indeed something missing. The fault is that the scientist's theoretical models of administration are too rational, too tidy, too aseptic. . . we had better examine afresh our current perspectives" (1960: 284). This was a sentiment echoed by Schwab (1964) in his assessment "...that the theory movement

reflected a false model both for inquiry and training" and that hypothetico-deductive systems were viciously abstract (in Culbertson, 1988: 19).

Despite this disconnection, however, there was an increasing demand for university courses in educational administration – at least in part because of the enormous growth in the numbers and size of secondary schools during the postwar period. This was not simply the case in North America, but also throughout the world. In Britain, courses in educational administration were established at the Institute of Education (London University), Bristol, Birmingham, and Oxford, led by men such as Ray Bolam, Meredydd Hughes, Eric Hoyle, Len Watson, William Taylor, and George Baron; in Australia, they were led by Bill Walker at the University of New England and Bill Bassett and Mac Grassie at the University of Queensland; and in Canada, they were led by Robin Farquhar at OISE and Art Reeves at Alberta.

While there was general agreement in the Commonwealth that the social sciences were a useful source of theory and methodology for educational administration (Walker *et al.*, 1973; Baron and Taylor, 1969; Greenfield, 1968), the traditions of analysis in the Commonwealth were somewhat different from those in the United States. For instance, in Great Britain, the historical traditions and political and religious roles of the headmaster were important influences. Given the role played by the great public schools in England, this is not surprising (Baron, 1970; Bernbaum, 1976). Indeed, in many cases, the fate of not only schools, but also whole communities was dependent upon the success or failure of the headmaster (Bamford, 1957). Even today, Marlborough College, for instance, enrolls some 800 pupils and generates some 22 million pounds a year income in a town of less than 8000 inhabitants. Rugby School generates some 18 million pounds a year for the town of the same name.

Historically, the role of the head in such schools (Peters, 1976) provided the foundation for the great man theory of leadership where autonomy, authority, religion, and morals jointly defined an omnipotent role for the headmaster, perhaps best put by Thring of Uppingham: "I am supreme here and will brook no interference" (quoted in Bernbaum, 1976: 34). Issues of religion, class, and politics were brought together in the headship in ways that influenced ideas of educational administration and leadership (Cannon, 1970). Central to this vision was control of the curriculum and its relationship to religion and to class stratification. As Gordon (1974) pointed out, the history of schooling in England was a history of class warfare: "No scheme of education will be accepted as satisfactory by the middle class which does not provide for the entire separation of their children from those of a lower grade, a separation as complete as that which exists between them and the children of the upper classes. . . One of two things only can relieve the pressure felt by the middle class: either

the curriculum of the secondary schools must be raised, or that the Board and Voluntary Schools be reduced to a more elementary standard" (quoted in Gordon, 1974: 292).

The politics of class warfare inevitably, therefore, shaped the study of educational administration in England, as did the policies of successive twentieth century governments directed toward increasing access to both primary and secondary schooling for a greater and greater proportion of the population (Baron, 1969). As Hoyle (1986) pointed out, in such a context, the politics of school management were central to the understanding of educational administration and management. A similar view was taken in Canada (Robinson and Elliott, 1973) and in Australia (Partridge, 1968). In New Zealand, such a perspective had been long established (Webb, 1937; Parkyn, 1954) and influential (Currie, 1962).

It was hardly surprising that, in 1974, when Thom Greenfield delivered his address to the International Intervisitation Program in Bristol, those from the Commonwealth were largely supportive of his view that organizational theory as it stood on the twin supports of behaviorist theory and systems theory was an inadequate and misleading explanation of schools as organizations (Greenfield, 1975). The furor among those from the United States was, however, quite patent. The very foundations of the Theory Movement were under attack and were to be defended in strident terms.

The ensuing decade can be seen both as an attempt to shore up the remnants of the Theory Movement through texts such as Hoy and Miskel (1978) and Silver (1983), which continued to separate education from administration and pursue a science of the latter, and as setting the groundwork for exploration of alternative theoretical approaches as a basis for the teaching of educational administration as well as an attempt to redress the feeling among educational administrators (especially in the United States) that contemporary courses were "...too abstract and remote from real administrative conditions" (Glatter, 1970: 66).

That Glatter's observation was justified was confirmed by the publication, in 1988, of the *Handbook of Research on Educational Administration* (Boyan, 1988) which displayed both an obsession with abstract theorizing and an astonishing parochialism which ignored the increasing volume of work being produced in the Commonwealth. As Bill Walker observed, despite UCEA support for the establishment of the British, Australian, Canadian, and New Zealand educational administration societies and their umbrella organization the Commonwealth Council for Educational Administration and its outreach to Africa, the Caribbean, and Cyprus, and the outpouring of research and writing that ensued, the Americans continued to pursue scholarship in educational administration that was characterized by "narcissism and the tyranny of isolation" (Walker, 1984; in Boyan, 1988: 12).

During the 1980s, a variety of new perspectives made an appearance. Sergiovanni did his best to encourage a cultural approach to educational organizations (Sergiovanni and Corbally, 1984). Greenfield further developed his subjectivist perspective (Greenfield and Ribbins, 1993; Macmillan, 2003). Foster (1987) articulated an approach based on the critical theory which was followed up by Smyth and his colleagues (Smyth, 1989). Bates (1980, 1983, 1987) developed an approach based on the new sociology of education and called for the redress of the separation of educational and administrative ideas along with calling for educational administration to be informed by ideas of social justice. Gronn (1986) was advocating the application of micropolitical theory and psychosocial perspectives to the management of schools. Leithwood (1982) was linking educational leadership to curriculum decision making. Fullan (1982) was linking leadership to educational change. Stephen Jacobsen and his colleagues were arguing for educational leadership to be seen within the context of educational reform (Jacobsen and Conway, 1990).

There was, in fact, an outpouring of ideas and, by the early 1990s, a wholly new postpositivist perspective was being developed in the United States by Donmoyer, Scheurich, and their colleagues, a perspective which sought to build a new approach to theory, construct a quite different knowledge base for educational administration, and include previously excluded voices – particularly those concerned with gender and race (Donmoyer *et al.*, 1995).

The 1990s saw a further broadening of perspective, including considerable work around the local management of schools as a practical response to concerns in educational management (Caldwell and Spinks, 1992; Wallace, 1992) as well as a renewed attempt to reassert the centrality of educational issues as a central focus for educational leadership (Duignan and Macpherson, 1992). Crump (1993) and his associates placed school-centered leadership within the context of turbulent social and political change in both national and international contexts. Hodgkinson (1991) further developed his ideas of educational leadership as a moral art. Ribbins and Marland (1994) renewed biography as a means of understanding leadership. Starratt (1990) applied dramaturgical analysis to leadership. Evers and Lakowski (1995) advocated a coherentist approach to theory. Blackmore, Kenway, and Hall articulated a feminist perspective on educational leadership (Blackmore, 1999; Blackmore and Kenway, 1993; Hall, 1996). Begley and his colleagues articulated a values base for educational administration (Begley and Leonard, 1999). Grace (1995) presented a policy scholarship approach to school leadership. Gronn (1999) presented an analysis of educational leadership located in the demands of the new world order. Sergiovanni (1999) once again rethought about educational leadership and the school as a covenantal community.

Schmuck and Runkel (1995) consolidated an organizational development approach to schools, one earlier pioneered in Australia by Mulford and his associates (Mulford and Kendall, 1975; Mulford *et al.*, 1977; Mulford, 1978).

What is noticeable about this diversity of approaches is the shift of emphasis from administration to management and from management to leadership. Historically, the roots of the field began in administration, moved through management, and now focus on leadership. This transition is exemplified by the changed titles of the British Educational Administration Society and its journals: *Educational Administration* in the 1970s; *Educational Management and Administration* in the 1980s and 1990s, and *Educational Management, Administration and Leadership* now, in the 2000s. The shift of focus has been significant as currently, leadership has become the dominant theme.

While, for instance, Tomlinson's (2004) encyclopedic collection of papers is called *Educational Management*, its categorization of those papers is shaped by four key ideas: educational values, theory, leadership, and change. In this, its organization is markedly different, for instance, from Boyan's *Handbook of Research on Educational Administration* of 1988. While Tomlinson's collection is retrospective, other collections such as those edited by Davies and West-Burnham (2003) and English (2005) herald the entry of significant new voices in the field and a diversity of themes that is markedly different from past decades. Perhaps stimulated by the kind of concerns elaborated by Murphy (2002, 2006), there is a significant concern with the transformation of the context of education; the impact of globalization and the demands of the new economy; the social purposes of schools and the impact of the market economy; the management of learning; school improvement and change; the role of leadership in professional development; the impact of technology; and the micro-politics of school leadership. Other collections, such as Brent-Davies (2005), display the variety of approaches to leadership (strategic, transformational, invitational, ethical, learning centered, constructivist, poetical, emotional, entrepreneurial, distributed, and sustainable), now part of the discourse in the field. New texts such as Starratt (2003) display a focus on meaning, community, and responsibility, absent from earlier work. Marshall (2004), Larson and Murtadha (2002), and Bates (2006) relate educational leadership to social justice. Lingard *et al.* (2003) place a renewed emphasis on leadership and learning as do Mulford and his colleagues (Mulford *et al.*, 2004; Mulford, 2005). Starratt (2004) and Samier (2003) examine anew the ethical foundations of leadership. Gunter (2001, 2005) explores the responsibilities of leading teachers within a Bourdieuan framework. Gronn (2003) examines the idea of designer leadership and the new (greedy) work of teachers. Samier and Bates (2006) renew interest in the role of esthetics in leadership.

Walker and Dimmock (2002) espouse a cross-cultural perspective. Bottery (2004) places educational leadership within the context of the global crisis. Beare (2000) argues for leadership to be concerned with the future of the school as an institution. Ribbins (2006) and his colleagues revisit the place of history in the study of educational leadership.

In all this recent writing, there are several connections of great importance. First, the importance of leadership in the promotion of learning is being reestablished; second, such leadership is increasingly connected with the social, economic, and political changes and demands of societies that are becoming both localized and globalized; third, leadership is seen to be no longer focused on a single person, but distributed throughout educational institutions as a shared responsibility; and fourth, and perhaps most importantly, the mechanics of school organization and administration can no longer be seen to be something comparatively external and indifferent to educational ideals. Dewey would have approved. A 100 years of effort in the study and practice of educational leadership has not been a wasted one after all.

Bibliography

- Anderson, J. D. (2006). A tale of two Browns: Constitutional equality and unequal education. In Ball, A. F. (ed.) *With More Deliberate Speed. 105th Yearbook of the NSSE*, pp 14–35. Malden, MA: Blackwell.
- Bamford, T. W. (1957). The public school town in the nineteenth century. *British Journal of Educational Studies* 6, 25–36.
- Baron, G. (1969). The study of educational administration in England. In Baron, G. and Taylor, W. (eds.) *Educational Administration and the Social Sciences*, pp 3–17. London: Athlone.
- Baron, G. (1970). Some aspects of the headmaster tradition. In Musgrave, P. W. (ed.) *Sociology, History and Education*, pp 183–192. London: Methuen.
- Baron, G. and Taylor, W. (eds.) (1969). *Educational Administration and the Social Sciences*. London: Athlone.
- Bates, R. (1980). Educational administration, the sociology of science and the management of knowledge. *Educational Administration Quarterly* 16, 1–20.
- Bates, R. (1983). *Educational Administration and the Management of Knowledge*. Geelong, VIC: Deakin University Press.
- Bates, R. (1987). Corporate culture, schooling and educational administration. *Educational Administration Quarterly* 23, 79–115.
- Bates, R. (2006). Educational administration and social justice. *Education, Citizenship and Social Justice* 1, 141–156.
- Beare, H. (2000). *Creating the Future School*. London: Routledge Falmer.
- Begley, P. and Leonard, P. (1999). *The Values of Educational Administration*. London: Falmer.
- Bernbaum, G. (1976). The role of the head. In Peters, R. S. (ed.) *The Role of the Head*, pp 9–36. London: Routledge Kegan Paul.
- Blackmore, J. (1999). *Troubling Women: Feminism, Leadership and Educational Change*. Buckingham: Open University Press.
- Blackmore, J. and Kenway, J. (eds.) (1994). *Gender Matters in Educational Administration and Policy*. London: Falmer.
- Bottery, M. (2004). *The Challenges of Educational Leadership*. London: Paul Chapman.

- Boyan, N. J. (ed.) (1988). *Handbook of Research on Educational Administration*. New York: Longman.
- Caldwell, B. and Spinks, J. (1992). *Leading the Self-managing School*. London: Falmer.
- Callaghan, R. (1962). *Education and the Cult of Efficiency*. Chicago, IL: Chicago University Press.
- Cannon, C. (1964). The influence of religion on educational policy, 1902–1944. *British Journal of Educational Studies* **12**, 143–160.
- Covello, L. (1936). A high school and its immigrant community. *Journal of Educational Sociology* **9**, 331–346.
- Covello, L. (1958). *The Heart is the Teacher*. New York: McGraw Hill.
- Crump, S. (1993). *School Centred Leadership*. Melbourne, VIC: Nelson.
- Culbertson, J. A. (1988). A century's quest for a knowledge base. In Boyan, N. J. (ed.) *Handbook of Research on Educational Administration*, pp 3–26. New York: Longman.
- Currie, G. (1962). *The Report of the Commission on Education in New Zealand (The Currie Report)*. Wellington: Government Printer.
- Davies, B. (ed.) (2005). *The Essentials of School Leadership*. London: Paul Chapman.
- Davies, B. and West-Burnham, J. (eds.) (2003). *Handbook of Educational Leadership and Management*. London: Pearson–Longman.
- Dewey, J. (1902). *The Educational Situation*. Chicago, IL: Chicago University Press.
- Dewey, J. (1916). *Education and Democracy*. New York: Macmillan.
- Donmoyer, R., Imber, M., and Scheurich, J. (eds.) (1995). *The Knowledge Base in Educational Administration*. Albany, NY: State University of New York Press.
- Duignan, P. and Macpherson, R. (1992). *Educative Leadership*. London: Falmer.
- Edwards, R. (1979). *Contested Terrain: The Transformation of the Workplace in the Twentieth Century*. New York: Harper.
- English, F. (ed.) (2005). *The SAGE Handbook of Educational Leadership: Advances in Theory, Research and Practice*. Thousand Oaks, CA: Sage.
- Evers, C. and Lakomski, G. (1995). Towards coherence in administrative theory. In Evers, C. and Chapman, J. (eds.) *Educational Administration: An Australian Perspective*, pp 97–110. Sydney, NSW: Allen and Unwin.
- Foster, W. (1987). *Paradigms and Promises: New Approaches to Educational Administration*. Buffalo, NY: Prometheus Books.
- Fullan, M. (1982). *The Meaning of Educational Change*. New York: Teachers College Press.
- Getzels, J. W. (1952). A psycho-sociological framework for the study of educational administration. *Harvard Educational Review* **22**, 235–246.
- Glatzer, R. (1970). *Management Development for the Education Profession*. London: London University Institute of Education.
- Gordon, P. (1974). *The Victorian School Manager*. London: Woburn.
- Grace, G. (1995). *School Leadership: Beyond Educational Management*. London: Falmer.
- Greenfield, T. B. (1968). *Systems Analysis for Educational Management*. Toronto, ON: Ontario Institute for Studies in Education.
- Greenfield, T. B. (1975). Theory about organization: A new perspective and its implications for schools. In Greenfield, T. and Ribbins, P. (eds.) *Greenfield on Educational Administration*, pp 1–25. London: Routledge.
- Gronn, P. (1986). Politics, power and the management of schools. In Hoyle, E. and McMahon, A. (eds.) *The Management of Schools: World Yearbook of Education*, London: Kogan Page.
- Gronn, P. (1999). *The Making of Educational Leaders*. London: Cassell.
- Gronn, P. (2003). *The New Work of Educational Leaders*. London: Paul Chapman.
- Gunter, H. (2001). *Leaders and Leadership in Education*. London: Paul Chapman.
- Gunter, H. (2005). *Leading Teachers*. London: Continuum.
- Hall, V. (1996). *Dancing on the Ceiling: A Study of Women Managers in Education*. London: Paul Chapman.
- Halpin, A. W. (ed.) (1958). *Administrative Theory in Education*. Chicago, IL: Chicago University Press.
- Halpin, A. W. (1960). Ways of knowing. In Cambell, R. F. and Lipham, J. M. (eds.) *Administrative Theory as a Guide to Action*. Chicago, IL: Mid-west Center University of Chicago.
- Hodgkinson, C. (1991). *Educational Leadership: The Moral Art*. Albany, NY: SUNY.
- Hoy, W. and Miskel, C. (1978). *Educational Administration: Theory, Research and Practice*. New York: Random House.
- Hoyle, E. (1986). *The Politics of School Management*. London: Hodder and Stoughton.
- Hudson, M. J. and Holmes, B. J. (1994). Missing teachers, impaired communities: The unanticipated consequences of 'Brown vs Board of Education' on the Afro-American teaching force at pre-collegiate level. *Journal of Negro Education* **63**, 388–393.
- Jacobsen, S. L. and Conway, J. A. (eds.) (1990). *Educational Leadership in an Age of Reform*. New York: Longman.
- Larson, C. and Murtadha, K. (2002). Leadership for social justice. In Murphy, J. (ed.) *The Educational Leadership Challenge: Redefining Leadership for the 21st Century*, pp 134–161. Chicago, IL: Chicago University Press.
- Leithwood, K. (1982). *Studies in Curriculum Decision-making*. Toronto, ON: Ontario Institute for Studies in Education.
- Lingard, R., Hayes, D., Mills, M., and Christie, P. (2003). *Leading Learning*. Buckingham: Open University Press.
- Macmillan, R. B. (ed.) (2003). *Questioning Leadership: The Greenfield Legacy*. London: Althouse.
- Marshall, C. (ed.) (2004). *Educational Administration Quarterly* **40**(1).
- Moore, H. A. (1964). The ferment in school administration. In Griffiths, D. E. (ed.) *Behavioural Science and Educational Administration: NSSE Yearbook*, pp 11–32. Chicago, IL: Chicago University Press.
- Mulford, B. (1978). Organisational development in schools: An octet of dilemmas. *Educational Change and Development* **1**(2), 13–25.
- Mulford, B. (2005). Quality evidence about leadership for organisational and student learning in schools. *School Leadership and Management* **25**, 321–330.
- Mulford, B., Conabere, T., and Keller, J. (1977). Organisational development in schools: Early data from the Australian experience. *Journal of Educational Administration* **15**(2), 210–237.
- Mulford, B. and Kendall, L. (1975). *Organisational Development for Australian Schools*. Canberra, ACT: ACEA Bulletin 3.
- Mulford, B., Silins, H., and Leithwood, K. (2004). *Leadership for Organisational Learning and Student Outcomes: A Problem-based Learning Approach*. Dordrecht: Kluwer.
- Murphy, J. (ed.) (2002). *The Educational Leadership Challenge: NSSE Yearbook*. Chicago, IL: Chicago University Press.
- Murphy, J. (2006). *Preparing School Leaders*. Toronto, ON: Rowman and Littlefield.
- Parkyn, G. W. (ed.) (1954). *The Administration of Education in New Zealand*. Wellington: New Zealand Institute of Public Administration.
- Partridge, P. H. (1968). *Society, Schools and Progress in Australia*. Oxford: Pergamon.
- Peebles, R. (1978). *Leonard Covello: A Study of an Immigrant's Contribution to New York City*. New York: Arno Press.
- Peters, R. S. (ed.) (1976). *The Role of the Head*. London: Routledge Kegan Paul.
- Ribbins, P. (2006). Administration and leadership in education: A case for history? *Special Issue. Journal of Educational Administration and History* **38**(2).
- Ribbins, P. and Marland, M. (1994). *Headship Matters: Conversations with Seven Secondary School Headteachers*. London: Longman.
- Robinson, N. and Elliott, J. (eds.) (1973). *The Politics of Education: Current Issues*. Burnaby, BC: Simon Fraser University.
- Samier, E. (ed.) (2003). *Ethical Foundations for Educational Administration*. London: Routledge.
- Samier, E. and Bates, R. (eds.) (2006). *Aesthetic Dimensions of Educational Administration and Leadership*. London: Routledge.
- Schmuck, R. and Runkel, P. (1995). *The Handbook of Organization Development in Schools and Colleges*. Prospect Heights, IL: Waveland Press.
- Schwab, J. (1964). Theory-art-practice. In Willower, D. and Culbertson, J. (eds.) *The Professorship in Educational Administration*, pp 47–70. Columbus, OH: University Council for Educational Administration.

- Sergiovanni, T. (1999). *Rethinking Leadership*. Arlington Heights, IL: Skylight Professional Development.
- Sergiovanni, T. and Corbally, J. (eds.) (1984). *Leadership and Organizational Culture: New Perspectives on Administrative Theory and Practice*. Urbana, IL: University of Illinois Press.
- Silver, P. (1983). *Educational Administration: Theoretical Perspectives on Practice and Research*. New York: Harper and Row.
- Smyth, J. (ed.) (1989). *Critical Perspectives on Educational Leadership*. London: Falmer.
- Starratt, R. J. (1990). *The Drama of Schooling*. New York: Falmer.
- Starratt, R. J. (2003). *Centering Educational Administration*. London: Erlbaum.
- Starratt, R. J. (2004). *Ethical Leadership*. San Francisco, CA: Jossey-Bass.
- Taylor, F. W. (1911). *The Principles of Scientific Management*. New York: Harper.
- Tomlinson, H. (ed.) (2004). *Educational Management* (4 vols). London: Routledge.
- Tyack, D. and Hansot, E. (1982). *Managers of Virtue: Public School Leadership in America, 1820–1980*. New York: Basic Books.
- Wallace, G. (ed.) (1992). *Local Management of Schools*. Adelaide, SA: Multilingual Matters.
- Walker, A. and Dimmock, C. (eds.) (2002). *School Leadership and Administration: Adopting a Cultural Perspective*. London: Routledge Falmer.
- Walker, W. G., Crane, A. R., and Ross-Thomas, A. (eds.) (1973). *Explorations in Educational Administration*. St Lucia, QLD: University of Queensland Press.
- Webb, L. (1937). *The Control of Education in New Zealand*. Wellington: New Zealand Council for Educational Research.
- Wilson, W. (1887). *The study of administration*. *Political Science Quarterly* **2**, 197–222.
- Young, E. F. (1916). Democracy and education. *Journal of Education* **84**, 5–6.

Further Reading

- Culbertson, J. A. (1983). Theory in educational administration: Echoes from critical thinkers. *Educational Researcher* **34**, 15–22.
- Hoyle, E. and McMahon, A. (1986). *The Management of Schools*. London: Kogan Page.

Educational Leadership as a Field of Study: Professional Journals 1994–2006

A R Thomas, University of Wollongong, Wollongong, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

Developments in Educational Leadership Journals

A professional journal serves to define its discipline. As such, in its immediacy, it reports contemporary practice and scholarship. In the longer term, however, a journal also serves to redefine its field because knowledge begets knowledge and the boundaries of scholarship are inexorably extended. An example of the latter may be seen in the title of this article: were it written a decade ago, educational administration would have appeared. However, within that same decade, leadership has infiltrated the lexicon of practitioners and scholars. In the United States of America and United Kingdom, for example, where administration and management respectively have long served, leadership appears to be gradually replacing these generic descriptors. Further evidence is available. For example, in the United Kingdom, in 2004, *Educational Management and Administration* became a more comprehensive journal – *Educational Management Administration and Leadership* (Vol. 32 No. 1); in Australia in 2005 *The Practising Administrator* was retitled *The Australian Educational Leader* (Vol. 32 No. 1).

At the outset, a rider must be added. Scholars in educational leadership are eclectic in their research and writing. Accordingly, there are many journals not specific to the field from which authors draw material and to which they also contribute. Examples of such may be readily found in *Administrative Science Quarterly*, *Journal of Leadership and Organizational Studies*, *Journal of Social Issues*, *Race, Gender and Class*, *Women in Management*, and *Phi Delta Kappan*, to name but a few.

The purpose of this article is twofold – first, to report developments in the publication of journals in educational leadership since the appearance of the second edition of the *International Encyclopedia of Educational* in 1994 and, second, to analyze the themes of the articles published therein in an attempt to capture, at least in part, the vast growth and change in knowledge in the last 10–15 years.

There are many worthy journals in the field of educational leadership; little did this author realize that, in preparing this article, his reading would be come so extensive and time consuming. Ultimately, a proposed survey of all journals became a passing review of several and a specific concentration on a few. Criteria for selection were that a journal had to be generalist in composition but

with its emphasis on school and system leadership, refereed, well established, and frequently cited. A journal also had to be in English, admittedly a disappointing limitation, particularly when there are available substantial non-English publications such as the Israeli *I'yunim Be'minhal U're'irgun Hachinuch* (*Studies in Educational Administration and Management*). The journals consulted (with some exceptions to the preceding criteria) are displayed in Table 1.

The three journals chosen for detailed analysis satisfy the previous criteria and were also established well before 1994:

- *Educational Administration Quarterly* (EAQ), established 1964 and edited in the United States of America. EAQ is the journal of the University Council for Educational Administration (UCEA) (of which there are approximately 80 institutional affiliates). The journal's editorial board consists of 43 members (no more than ten of whom may be from non-UCEA institutions) "representing a diversity of conceptual and methodological expertise as well as institutional affiliation, gender, and race." During the period 1994–2006 there were five editors. The EAQ appears in five issues per volume (as from Vol. 32, 1996).
- *Educational Management Administration & Leadership* (EMAL), established 1973 and edited in the United Kingdom. EMAL is the journal of the British Educational Management and Administration Society. The journal's international editorial board consists of approximately 20 members (currently drawn from 12 countries). During the period 1994–2006, there were two editors. The journal appears in four issues per volume having undergone a significant increase in pagination (as from Vol. 34, 2006).
- *Journal of Educational Administration* (JEA), established 1963 and edited in Australia. The JEA has no particular institutional affiliation although it is edited from the University of New England for its 44 volumes. Since 2007 (Vol. 45), the journal is being edited from the University of Wollongong. The journal's editorial advisory board consists of 33 members (currently drawn from ten countries). The responsibilities of board members are to referee submitted papers, advise the editor on journal policy, and select the annual outstanding paper award. There has been only one editor throughout the review period. The JEA appears in six issues per volume (Vol. 33, 1995, five issues; Vol. 39, 2001, six issues).

Arguably, and at the time of writing, these are the leaders in the field. Which of the three may be labeled the premier journal is open to debate although editors of two have laid a claim to such distinction. Lindle (2003), editor of *EAQ* has editorialized:

As the oldest (sic) journal in the diverse and applied field of educational leadership, Educational Administration Quarterly holds a venerated spot in the annals of our field. Over the past couple of decades, other journals have arisen to address the vexing problems of practice that persist in educational leadership. Even among the burgeoning claims to immediate answers concerning the complex issues of leadership, *EAQ* persists in delineating the scope and direction of a volatile field of scholarship (p.433).

Pounder (2005), Lindle's successor as editor, was also "committed to promoting and maintaining *EAQ*'s reputation as the premier scholarly journal in the field of educational leadership" (p.3).

Bush (2004b), current editor of *EMAL*, has championed his journal:

Since becoming editor in August 2002 I have been seeking to develop Educational Management Administration and Leadership (*EMAL*) as the leading international journal in the field ... I have continued to encourage international researchers and scholars to regard *EMAL* as the outlet of choice in disseminating research findings and critical comments on all aspects of our field ... this issue of the journal comprises papers from five countries and four continents (p.123).

As revealed in **Table 1**, the number of journals is increasing. The 1990s in particular saw (at least) five new publications – a period as productive as any preceding decade – and since 2000 there have been three more – *Cases in Educational Leadership (CEL)*, *Values and Ethics in Educational Administration (VEEA)* and, most recently, *Research on Leadership in Education (RLE)*. (Two of these – *CEL* and *RLE* – have been produced by UCEA.) Each journal reflects a move to greater specialization, designed to capture material that hitherto has been dispersed among the generalist journals. It may also be argued that these specializations reflect, not just the further maturing and consolidation of educational leadership, but also an expansion and part re-definition of the field.

Cases in Educational Leadership is an electronic journal, possibly the first of its kind in the field. Whether it pre-sages more of this kind is a matter of conjecture. Clearly, however, the most influential factors affecting journals throughout the period under review have been the rapid developments in information technology. Access to most of the journals listed above has been greatly extended through the Internet. University library databases, often enlarged via consortium agreements with other institutions, have facilitated an extraordinary accessibility to

Table 1 Selection of journals in educational leadership (and year first published)

<i>Journal of Educational Administration</i> (1963)
<i>Journal of School Leadership</i> (1991)
<i>Journal of Educational Administration and History</i> (1969)
<i>Journal of Educational Administration and Foundations</i> (1986)
<i>Journal of Educational Policy</i> (1986)
<i>Journal of Research on Leadership Education</i> (2006)
<i>Journal of Cases in Educational Leadership</i> (1998)
<i>International Journal of Leadership in Education</i> (1998)
<i>International Journal of Studies in Educational Administration</i> (1973)
<i>International Journal of Educational Management</i> (1987)
<i>Educational Administration Quarterly</i> (1964)
<i>Values and Ethics in Educational Administration</i> (2004)
<i>Leading and Managing</i> (1996)
<i>Planning and Changing</i> (1970)
<i>School Leadership and Management</i> (1981)
<i>Educational Management Administration and Leadership</i> (1973)
<i>Educational Evaluation and Policy Analysis</i> (1979)
<i>Leadership and Policy in Schools</i> (2002)
<i>Educational Administration: In Theory and Practice</i> (1972)
<i>NASSP Bulletin</i> (1917)
<i>Canadian Administrator</i> (1961)
<i>Canadian School Executive</i> (1982)
<i>Educational Leadership</i> (1944)
<i>The School Administrator</i> (1944)
<i>New Zealand Journal of Educational Leadership</i> (1986)
<i>The Australian Educational Leader</i> (1979) (formerly <i>The Practising Administrator</i>)

contemporary and archived journal articles. Records of downloads of contents pages, abstracts, and full-text articles can now provide data of inestimable value to journal publishers, editors, subscribers, and contributors alike.

This information, challenges in several respects, the hitherto simple meaning of a journal's circulation! Nevertheless, the equation – circulation equals subscribers – persists although not all journals advertise such figures. Of all those reviewed, however, the *EAQ* is the most open and specific in this regard. Details of its circulation are provided annually as part of its statement of ownership, management, and circulation (1200–1300 in 2006). For others, such as *JEAL* and *International Journal of Educational Management*, such information is classified by their publishers.

The application of new technologies to the editorial processes of receipt and evaluation of submitted manuscripts has gathered pace. As shown in **Table 2**, most journals have eschewed paper submissions in favor of an all-electronic process. *EAQ* and *EMAL* are in this category; the *JEAL*'s procedures switch to electronic in 2007. Software systems produced by some of the journals' publishers (e.g., Emerald's JADE) are being installed to provide, *inter alia*, information to contributors about the status of papers under review.

The preferred length of manuscripts varies considerably with papers as brief as 1000 words and as long as 8000 being entertained. Perusal of articles in the three selected

Table 2 Journal submissions

Journal	Edited	Issues per annum	Articles per issue (ex thematic)	Sub-mission	Length (1000 words)	Website
<i>Educational Management Administration & Leadership</i>	UK	4	7	elec	N/A	http://ema.sagepub.com
<i>Journal of Educational Administration & History</i>	UK	2	5–6	3 hc	3–6	www.tandf.co.uk
<i>School Leadership & Management</i>	UK	5	5–6	3 hc + elec	6	www.tandf.co.uk
<i>International Journal of Educational Management</i>	UK	7	6	3 hc	1–4	www.emeraldinsight.com/ijem.htm
<i>Educational Administration Quarterly</i>	USA	5	5	elec	25–40 pp	http://eaq.sagepub.com
<i>International Journal of Leadership in Education</i>	USA	4	5	elec	5–8	www.ingentaconnect.com/content/routledg/led
<i>Journal of School Leadership</i>	USA	6	5	elec	N/A	www.rowmaneducation.com/journals/JSL
<i>Journal of Educational Administration</i>	AUST	6	5	4 hc or elec	4–8	www.emeraldinsight.com/jea.htm
<i>Leading & Managing</i>	AUST	2	9	1 hc	5–7	N/A
<i>International Studies in Educational Administration</i>	HK/Cyprus	3	7	3 hc + elec	NA	www.cceam.org

hc – hard copy; elec – electronic copy.

journals quickly revealed that generous tolerance in wordage is frequently afforded authors!

Technology has also provided the means of reducing turnaround time for contributors. It has not been the custom for journals to advertise their turnaround times but within the past 3 years, such information has started to appear, usually on websites. Two months is seen by some editors as an ideal minimum but all aspire to reduce this period. Queuing time – from receipt of manuscript through the refereeing process to publication – has long been an understandable concern of contributors and editors as well, of course, who strive to maximize the currency of their journals' contents. The editors of *EMAL* and *JEA* both report queuing times of at least 12 months although, with the latter, a paper if accepted promptly and without substantial amendment, may be published 9 months after receipt.

As Henson (2003) states in his biennial contributions to *Phi Delta Kappan*, there is no uniform definition of refereed. Accordingly, in order to categorize the 50 or so educational journals that he reviews, Henson has established his own formula:

I award one point if manuscripts are sent outside the editorial offices for review, one point if the names of authors and their institutions are withheld from the reviewers, and one point if the manuscript is accompanied by a rating instrument of some sort (p. 789).

(Of the few journals relevant to education leadership in Henson's review, *NASSP Bulletin* and *Planning and Changing*

each score 3; *Educational Leadership* and *School Administrator* each score 0.) Using these criteria, the three journals closely examined for this article – *EAQ*, *EMAL*, and *JEA* – all score 3. For each refereeing is customarily conducted blind by three readers. Each of the three *JEA* referees comes from a different country.

Of importance to contributors – and also the tenure and promotions committees that plague editors for such information – is a journal's acceptance rate. These data were once relatively obscured, but most journals now acknowledge then via their websites. Comparison is not always straightforward; some journals calculate acceptance rate as a percentage of all papers submitted; others disregard papers culled initially by their editors and use only those assessed by referees as the basis for calculation some count an amended paper as an entirely new submission. Advertised figures for *EAQ*, *EMAL*, and *JEA* are currently 6%, 23%, and 27% respectively but these are not fully comparable and are, in fact, misleading – different formulae have been used in each case. Likewise, details relevant to citation frequency and impact factor are starting to appear – a direct response to the seemingly insatiable demand that permeates academe to quantify a journal's quality.

It has not been the intention of this article to analyze the range of research methods encompassed by the articles reviewed – notwithstanding the importance of such methods in further illuminating the dimensions of educational leadership as a social science. Suffice it to say that, in this respect, the methodology of research in the field is diverse

and comprehensive. While qualitative research methods have assumed a greater role throughout the period under review, quantitative works continue to remain prominent. Descriptive and inferential statistics still find ready use in educational leadership research, the latter enhanced by the availability of ever-increasing computing power.

In a fashion similar to their requirements for keywords in the 1980s, electronic databases have, understandably, demanded greater commonality in both the content and acuity of article abstracts. In 2005, both *EAQ* and *JEA* prefaced all articles with structured abstracts, which had to address four obligatory fields – purpose, design, findings, and value. The design/methodology/approach field requires authors to indicate how their objectives were achieved, the main method(s) used, their approach to the topic, and the theoretical or subject scope of their papers. Similarly, authors must locate papers within one of seven categories. Of these, five are particularly apposite to the articles reviewed for this article: research paper (which may involve the construction or testing of a model or framework, action research, testing of data ... surveys, empirical, scientific, or clinical research), viewpoint (author's opinion and interpretation), conceptual (...developing and testing hypotheses ... discursive ... philosophical and comparative studies of others' work and thinking), case study (actual interventions or experiences within organizations), and literature review (...annotate and/or critique the literature...).

The standing of a professional journal depends in no small degree on its internationality – the extent to which it conveys its discipline to the world at large and its reciprocal, the extent to which it attracts submissions from scholars throughout the world. Accordingly, at least three journals in educational leadership lay claim to such standing by including international in their titles. The educational objectives of the *International Journal of Educational Management*, for example, make this clear:

To provide those interested in the effective management of the educational process with a broad overview of developments and best practice in the field, with particular reference to how new ideas can be applied worldwide ... seek to contain material relating to innovation in educational management across the spectrum, the development of educational delivery mechanisms, and the creation of an environment in which the management of resources provides the most efficient outputs achievable on an international basis to allow the sharing of new initiatives.

Likewise, the *International Journal of Leadership in Education* expresses its intent. The journal

is an international journal for the publication of theoretical and practical discussions of educational leadership. It provides a forum for researchers and practitioner researchers to consider conceptual, methodological and practical issues in a range of professional and service settings

and sectors ... It offers a broad definition of leadership, including teachers as leaders, shared governance, site-based decision-making and community-school collaborations.

Two of the three journals chosen for close analysis are also avowedly international in outlook. *EMAL*

publishes original contributions on educational administration and management in the wildest sense: on the management of schools of all types, and of further and higher education institutions; on administration and policy at all levels – institutional, local, national and international; and on the study and teaching of educational administration.

Contributions may encompass critical discussions, accounts of new methods, developments and controversial issues as well as research reports. Analytical contributions from those actually involved in the practical management and administration of education – as heads, senior school staff, local education authority advisers and education officers, college and university staff – are particularly welcome.

The *JEA*

is for all interested in the practice and theory of educational administration worldwide. It is designed to meet the needs of principals, inspectors, superintendents, directors of education, administrators in institutions of higher education, and of university teachers and students of educational administration.

In seeking to advance thinking in the field, the Editor believes that there is no aspect of education more deserving of disciplined study and research than the administrative process, on which the efficacy of the teaching – learning process so much depends, and that this will best be achieved through an international approach to the field.

For long assumed to be the domain of the English-speaking world, the study of educational leadership has spread its wings. Encouraged, *inter alia*, by the formation of international bodies such as the Commonwealth Council for Educational Administration and Management (CCEAM), educational leadership has developed – and continues to develop – as a field of scholarship. Reference above to information accrued on journal websites highlights evidence of such growth. The nationalities of authors contributing to journals also lend further substance to this development. In stark contrast to this development, however, a sample of 25 issues of *EAQ* revealed that 81% of articles were by American authors, 19% by authors from elsewhere. Furthermore, of the articles authored by Americans, 70% were submitted from UCEA universities.

One may surmise as to why the *EAQ* publishes relatively few articles from non-American sources. Clearly the sheer size of the United States of America and the many, many universities there are reasons for this. But there may also be a level of insularity therein that serves to discourage international contributors. Nevertheless,

the *EAQ* is integral to the operation of UCEA; that it is seen as the first choice for publication by UCEA-affiliated authors should not come as a surprise. As stated in the journal's preamble:

EDUCATIONAL ADMINISTRATION QUARTERLY invites the submission of manuscripts contributing to scholarly knowledge about educational administration. Submissions may include but are not limited to empirical investigations, conceptual and theoretical perspectives, policy and legal analyses, reviews of research and practice, and analyses of methodological procedures related to broad conceptions of administration in education. Papers on education reform, governance and reform in colleges of education, the teaching of educational administration, and the professional preparation of educational administrators – topics advancing the University Council for Educational Administration – are welcome.

Similar analyses in other journals revealed, somewhat serendipitously, yet in accord with the preceding two paragraphs, that authors tend to favor their local or own-national journals as outlets for their scholarship. This is not altogether surprising, particularly when articles are localized in content. Nevertheless, in spite of their professed internationality, journals seldom include a majority of articles from elsewhere. For example, identification of nationality of principal authors of articles in *EMAL* from a sample of 15 issues (excluding thematic or special numbers) revealed that 60% were from the UK, 40% from other countries. Two other UK-edited journals display this phenomenon. *School Leadership and Management* and the *Journal of Educational Administration and History* publish 67% and 54% local articles respectively. A startling exception to these practices is to be found in the (aptly titled) *International Journal of Educational Management (IJEM)* (also edited in the United Kingdom) wherein 86% (37) of the 43 articles in Volume 20 (2006) were by international contributors. An analysis of 21 issues of the *Journal of Educational Administration* displayed a similar distribution – 22% were by Australian authors, while 78% were submitted from other countries.

Thus, in acknowledging that the study of educational leadership is not the monopoly of a few English-speaking countries (particularly the United States of America, United Kingdom, Canada and Australia), the three journals – but especially *EMAL* and *JEA* – have, throughout the period of review, and particularly in more recent years, published papers submitted by authors in Belgium, Brazil, Brunei, Chile, China, Cyprus, Denmark, France, Germany, Greece, Hong Kong, Israel, Kenya, Kuwait, Malaysia, Netherlands, Papua New Guinea, New Zealand, Nigeria, Norway, Palestine, Singapore, South Korea, Spain, Sweden, Thailand, Turkey, United Arab Emirates, West Indies, and Zimbabwe. Add to these three other

sources – Finland, India, and Malta – in the *IJEM* and a clear picture emerges, one very different from that which prevailed throughout the 1970s and 1980s: the study of educational leadership has assumed international acceptance.

Every issue of the *EMAL* and *JEA* reviewed was introduced by an editorial. In the *EAQ*, the editorial has been an irregular inclusion, variously known as editor's foreword or note, or letter from the editor. Two elements are common to all editorials – an introduction to each of the articles included and obituaries acknowledging the contributions to the field by prominent professors – Foster, Griffith, and Willower in *EAQ*, Baron, Griffith, and Tomlinson in *EMAL*, and Walker and Willower in *JEA* (Dedicated memorial issues for Foster and Willower have been published respectively by *EAQ* and *JEA*). A third common feature may also be tentatively identified – the inclusion of operational details such as acceptance rates and queuing times for each.

In the main, editorials have not sought (as do those of newspapers) to support particular policies or influence readers' opinions on aspects of educational leadership. However, encouraging exceptions to this generalization may be found in *EMAL*, where, on occasions, the editors (Ribbins and Bush) have both authored well-designed, researched, and relatively outspoken statements of policy. Bush (2004a), for example, searches for greater clarity in the terminology of management, administration, and leadership. The subtitle to his editorial – Leadership in the Ascendancy – corroborates one of the significant findings to emerge from the analyses of articles described in the following section.

Lamented in two editorials (*EAQ* and *JEA*) has been the paucity of responses/rejoinders critical of articles published. The current review has identified very few examples. Could this be reflective of a nonargumentative readership and/or the outcome of a refereeing process that, in general, permits the publication of papers only after extended processes of amendment?

Article Themes

A key purpose of this article is to identify the subjects that authors have addressed throughout the period under review. In order to meet this challenge, attention was directed exclusively to the three leading journals identified above. During the period 1994–2006, over 1000 articles were published in these three journals – approximate figures for each: *EAQ* 340; *EMAL* 350; *JEA* 390. Of these, 120, 82, and 116 articles were attributable to thematic or special issues of the journals respectively. Such issues are customarily the outcome of a call for papers addressing a special theme designated by a guest editor. In recent years, it has become a practice for the *EAQ* and *JEA* to publish two thematic issues per volume and for *EMAL* to

publish one. Details of the thematic issues are provided in a later section.

For purposes of the following analysis, it was decided to focus only on the articles published in the nonthematic issues of the journals. This was done in order to better capture the broad, spontaneous nature of research and scholarship in the field. Accordingly, 210 articles in *EAQ*, 265 in *EMAL*, and 272 in *JEA* were considered. Each article was examined primarily by abstract and, where available and necessary, by keywords. The information was crosschecked with the article's title to ensure that the key subject was identified (titles sometimes do not reveal, and may even deliberately obscure, an article's intent!). Subjects were recorded and processed through an iterative series of groupings in order to identify broad themes of content. For example, the resultant theme of teachers included articles that addressed subjects such as recruitment, retention, commitment, motivation, empowerment, effectiveness, collegiality, and so on. Public schools, a major theme in each of the three journals, encompassed subjects such as instruction, effectiveness, development, safety, control, school-based management, and so on. On occasions, articles gave equal prominence to multiple themes. These were categorized accordingly.

To be established, a theme required a minimum of five articles in a journal that could clearly be clustered thereabouts. In total, 39 themes were identified in the nonthematic issues of the three journals. Some were common to the three journals, for example, principals/headteachers; some common to two, for example, organization theory in *EAQ* and *JEA*; some were unique, for example, teaching educational administration in *EAQ*, deputy principals/middle managers in *EMAL*, and stress/anxiety in *JEA*. The ten most-addressed themes for each of the three journals are listed in order of frequency in **Table 3**.

Any exercise of this kind has inherent dangers and, although the author has strenuously endeavored to eliminate them, some elements of subjectivity and arbitrariness must inevitably remain. Although the journals reviewed are published in English, the background and location of each journal bring numerous differences in expression, especially, in labeling. One additional task, therefore, has been to establish common ground on which to locate seemingly different subjects. Thus, for example, schools described variously as private, independent, maintained, religious, and nongovernment in the United States of America, United Kingdom, and Australia are herein identified as nonpublic.

The interpretation of these themes must be approached with caution, although there does seem some justification in claiming that, collectively, their emergence serves to add definition to the field of educational leadership. In particular, the themes common to the three journals – public schools, principals/headteachers, and reform/restructuring/change – vouch for far-reaching and international

Table 3 Most frequently addressed themes in three selected journals

	<i>EAQ</i> (USA)	<i>EMAL</i> (UK)	<i>JEA</i> (Australia)
1	Gender/women	Higher education	Principals
2	Students	Public schools	Public schools
3	Law/legislation/politics	Resources	Comparative/international
4	Teachers	Headteachers	Teachers
5	Public schools	Governance	Higher education
6	Organization theory	Deputy principals/middle managers	Values/ethics
7	Inclusion/special education	Comparative/international	Organization theory
8	Principals	Reform/restructuring/change	Stress/anxiety
9	Reform/restructuring/change	Inclusion/special education	Gender/women and Reform/restructuring/change (eq)
10	Teaching educational leadership	Nonpublic schools	Students

concerns among scholars in the field. Less expansive, but nevertheless important, are the themes that are located in two journals: for example, gender/women, students, teachers, and organization theory in *EAQ* and *JEA*; higher education and comparative/international in *EMAL* and *JEA*; and inclusion/special education in *EAQ* and *EMAL*.

As always with grouping or categorizing, the process in some aspects may be counterproductive, serving, for example, to mask what may be emerging or developing subjects of concern and scrutiny. Thus, for example, subjects such as trust and truth in schools, efficacy, schools' nexus with business, physical conditions of schools (including architecture), school marketing, technology, learning organizations and communities, and chaos/complexity theory may not be identified, either being subsumed in broader categories or isolated as minor groupings.

The reader will note that all of the themes identified are expressed as nouns. Implicit in each, of course, is an activity (or activities) of some kind. That activity (and here differences in English use bedevil the field) is managing or administering or leading. These are the action words that explicitly or implicitly accompany all journal articles. By far the most commonly used (and lending further support to the opening paragraph of this article) is leadership.

Leadership (or derivatives of such) has been used in almost 22% of all article titles uniformly in each of the three journals. Furthermore, the rate of use has been

increasing. In the period 1994–2000 titles containing leadership appeared in 14% of *EAQ*, 10% of *EMAL*, and 16% of *JEA* articles. In the later period 2001–06, these percentages have increased respectively to 26, 30, and 21.

Leadership is not a buzz word; nor is its use simply an attempt to resolve long-standing semantic arguments. Through increasing use of leadership, journals are contributing to a renaming of the field – a field that seeks a more appropriate descriptor of what it is school managers and administrators actually do and, ideally, what they must do in order to achieve the most desirable of outcomes for all who are engaged in the educative process.

As this author has stated elsewhere:

The phenomenon of leadership is, once again, undergoing one of its periodic, sustained examinations: definitions of leadership, components of leadership, correlates of leadership and so on, are occupying more and more journal space and more and more conference time. Yet therein lies an emerging danger. Just as the trait approach to leadership in decades past ‘succeeded’ in identifying a plethora of individual attributes or characteristics ‘fundamental’ to ‘successful’ leadership, contemporary studies threaten to engulf us with their own tidal wave of descriptors (Thomas, 2006: 11).

So from the titles of articles in the leading three journals, the following have emerged to confront this reader: administrative–instructional, authentic, dialogic, democratic, departmental, directive, distributed, educational, educative, effective, female, goal-orientated, headteacher, inclusive, instructional, local, moral, organizational, participative, peer-assisted, principal, school, school-technology, servant, site-based decision-making, small school, strategic, successful, symbolic, synergistic, teacher, technology, transformational, values-driven, and even bastard and principled-bastard leadership!

Again, it seems apposite to make the following observations:

Such a profusion of labels may be viewed with alarm and, paradoxically perhaps excitement: alarm because of the possibility of duplication in conceptualisation, definition and approach, with a resulting potential for confusion and even exasperation; excitement because of the variety of ‘lenses’ being used to examine educational leadership. One may only hope that for the individual presenters of these [articles] valuable guidelines are actually providing bases for their future leadership. Similarly, one may hope for a successful application of their findings beyond the specific institutions in which they collected their data. (Thomas, 2006: 11).

The preceding analysis has been derived from standard issues of the three selected journals. As previously stated, each of the journals published several thematic issues throughout the period under review. A selection of thematic issues for each journal is presented in **Table 4**.

Table 4 Samples of thematic issues

<i>Educational Administration Quarterly</i>			
Vol	No	Year	
30	3	1994	Connecting schools, families, and communities
33	5	1997	What will replace the comprehensive high school?
36	1	2000	Women in the superintendency
38	2	2002	Ensuring the capacity of university-based educational-leadership preparation
39	1	2003	Can the politics of education inform practice in educational leadership?
39	2	2003	Justice in the school – an issue devoted to the law and education
40	2	2004	The legacy of William Foster: Promises for leadership and schooling
40	4	2004	Critical perspectives on mentoring for leadership
41	4	2005	Pushing back resistance: African American perspectives on school leadership
42	1	2006	Educational change over time
<i>Educational Management Administration & Leadership</i>			
Vol	No	Year	
22	4	1994	Greenfield on educational administration
24	3	1996	Primary schools and their management
25	4	1997	Special international edition celebrating the 25th year of the journal
27	3	1999	Redefining educational management and leadership
28	3	2000	Reconstructing the learning society
30	1	2002	Theorists and theorizing in educational administration and management
30	2	2002	Education and its management. Global dimensions: local challenges
32	3	2004	The National College for School Leadership
33	2	2005	Researching leadership. Preview of progress
34	2	2006	Leadership and diversity
<i>Journal of Educational Administration</i>			
Vol	No	Year	
34	5	1996	Culture and educational leadership: An introduction
35	5	1997	Inclusion and school leadership
39	5	2001	Donald J. Willower: appraisal, appreciation, remembrance
41	2	2003	New technologies and educative leadership
41	6	2003	Violence in schools
42	2	2004	Education, ethics, and the cult of efficiency: Implications for values and leadership
42	6	2004	Democracy and leadership
43	1	2005	Educational institutions and leadership through the lens of organization theory
43	6	2005	The international successful school principalship project
44	4	2006	Beginning the principalship – international views

Each journal has explored a wide range of special themes several of which, it will be recognized, relate closely to some of those that emerged from analysis of standard issues.

Summary

The first section of this article reported on developments that had taken place in the production and publication of educational leadership journals since 1994. An extensive sample of English-speaking journals was initially identified. It was observed that the number of journals is increasing, simultaneously adding greater specialization to the educational-leadership library. Three major generalist journals – *EAQ*, *EMAL*, and *JEA* – were selected for intensive analysis.

The extraordinary developments in information technology have exerted the greatest influence on all aspects of journals' operations. Although not fully utilized by all educational-leadership journals, the Internet has paved the way for a more efficient conduct of the submission, refereeing, and publishing process. The first all-electronic journal has been established; others will follow. Offshoots of the Internet – databases bringing a greater array of indexing and abstracting facilities – have in part, brought about a retailoring of articles through demands for keywords and structured abstracts. It is likely that future developments in electronic presentation will induce further changes in articles' structure.

That submitted papers are refereed is no longer sufficient evidence of a journal's quality. Several factors – particularly the requirements of tenure and/or promotion committees – demand added forms of validation for journals. Quantifiable features such as acceptance (or, better for some, rejection) rates, impact factors, and even (as this author/ editor was recently asked) the number of indexing and abstracting agencies reporting a journal, are used increasingly to calibrate a journal's quality.

The influence of the Internet has also contributed substantially to the growing internationality of most educational-leadership journals. Easier and far-reaching access to journals has encouraged and enabled the expansion of educational leadership as a field of study. Data on electronic visits, together with an expanding international authorship, attest to such.

With few exceptions editorials remain policy free and do not seek to influence the field. Concomitant with this practice is a dearth of opposing and/or dissenting rejoinders to published articles.

The latter section of this article reported the analysis of article content in the three selected journals. It did so in order to present a picture of research and writing in the field since 1994. Over 700 articles, published in nonthematic

issues of the journals, provided grounds for establishing 39 content themes. The ten most-identified themes for each of the journals, in tandem with the subjects of their respective special or thematic issues, revealed considerable diversity in article content. Whereas there were some themes in common, the three journals displayed several differences and even elements of uniqueness. Nevertheless, collectively, the themes portrayed a field that encompasses much diversity.

The concept of leadership increasingly pervades the literature of journals but with a diversity of definitions and applications. One can only hope that, come the next edition of the encyclopedia, greater agreement will have been reached on what is perceived by so many as a noble behavior.

Bibliography

- Bush, T. (2004a). Editorial: What's in a name? *Educational Management Administration and Leadership* **32**(1), 5–9.
- Bush, T. (2004b). Editorial. Educational leadership and globalization. *Educational Management Administration and Leadership* **32**(2), 123–127.
- Henson, K. T. (2003). Writing for professional publication: Some myths and some truths. *Phi Delta Kappan* **June**, 788–791.
- Lindle, J. C. (2003). Editor's note. *Educational Administration Quarterly* **9**(4), 433–435.
- Pounder, D. G. (2005). Letter from the editor. *Educational Administration Quarterly* **41**(1), 3–6.
- Thomas, A. R. (2006). New waves of leadership: In retrospect. In Smith, L. and Riley, D. (eds.) *New Waves of Leadership*, pp 10–14. Winmalee, NSW: ACEL.

Further Reading

- Barnett, K. and McCormick, J. (2003). Vision, relationships and teacher motivation: A case study. *Journal of Educational Administration* **41**(1), 55–73.
- Bennett, N., Harvey, J. A., and Anderson, L. (2004). Control, autonomy and partnership in local education: Views from six chief education officers. *Educational Management Administration and Leadership* **32**(2), 217–236.
- Conley, S., Fauske, J., and Pounder, D. G. (2004). Teacher work group effectiveness. *Educational Administration Quarterly* **40**(5), 663–703.
- Conley, S. and Woosley, S. A. (2000). Teacher role stress, higher order needs and work outcomes. *Journal of Educational Administration* **38**(2), 179–201.
- Datnow, A. (2005). The sustainability of comprehensive school reform models in changing district and state contexts. *Educational Administration Quarterly* **41**(1), 121–153.
- Dinham, S. (2005). Principal leadership for outstanding educational outcomes. *Journal of Educational Administration* **43**(4), 338–356.
- Draper, J. and McMichael, P. (2003). Keeping the show on the road: The role of the acting headteacher. *Educational Management Administration and Leadership* **31**(1), 67–82.
- English, F. W. (1995). Toward a reconsideration of biography and other forms of life writing as a focus for teaching educational administration. *Educational Administration Quarterly* **31**(2), 203–223.
- Fennell, H. A. (2002). Letting go while holding on: Women principals' lived experiences with power. *Journal of Educational Administration* **40**(2), 95–117.
- Fink, D. (1999). Deadwood didn't kill itself: A pathology of failing schools. *Educational Management Administration and Leadership* **27**(2), 131–142.

- Francis, L. J. and Stone, E. A. (1995). School governors and the religious ethos of C of E voluntary aided primary schools. *Educational Management Administration and Leadership* 23(3), 176–187.
- Geijsel, F., Slegers, P., and van den Berg, R. (1999). Transformational leadership and the implementation of large-scale innovation programs. *Journal of Educational Administration* 37(4), 309–328.
- Gilbert, M. B. (1999). Why educators have problems with some students: Understanding frames of preference. *Journal of Educational Administration* 37(3), 243–255.
- Hatton, E. (2001). School development planning in a small primary school: Addressing the challenge in rural NSW. *Journal of Educational Administration* 39(2), 118–133.
- Ho, E. S. (2006). Educational decentralization in three Asian societies: Japan, Korea and Hong Kong. *Journal of Educational Administration* 44(6), 590–603.
- Kelly, C. and Protsik, J. (1997). Risk and reward: Perspectives on the implementation of Kentucky's school-based performance award program. *Educational Administration Quarterly* 33(4), 474–505.
- Kruse, S. D. (2003). Remembering as organizational memory. *Journal of Educational Administration* 41(4), 332–347.
- Law, L. Y. S., Walker, A., and Dimmock, C. (2003). The influence of principals' values on their perception and management of school problems. *Journal of Educational Administration* 41(5), 498–523.
- Mayrowetz, D. and Weinstein, C. S. (1999). Sources of leadership for inclusive education: Creating schools for all children. *Educational Administration Quarterly* 35(3), 423–457.
- McTavish, D. (2003). Aspects of public sector management: A case study of further education, ten years from the passage of the further and higher educational act. *Educational Management Administration and Leadership* 31(2), 175–188.
- Pashiardis, P. (1998). Researching the characteristics of effective primary school principals in Cyprus: A qualitative approach. *Educational Management Administration and Leadership* 26(2), 117–130.
- Sergiovanni, T. J. (1994). Organizations or communities? Changing the metaphor changes the theory. *Educational Administration Quarterly* 30(2), 214–226.
- Sharp, S. (2000). Allocating resources for learning support: A case study. *Educational Management Administration and Leadership* 88(2), 211–222.
- Simkins, T. (1995). The equity consequences of educational reform. *Educational Management Administration and Leadership* 23(4), 221–232.
- Smith, B., Roderick, M., and Degener, S. C. (2005). Extended learning time and student accountability: Assessing outcomes and options for elementary and middle grades. *Educational Administration Quarterly* 41(2), 195–236.
- Soutar, G. and McNeil, M. (1996). Measuring service quality in a tertiary institution. *Journal of Educational Administration* 34(1), 72–82.
- Stroud, V. (2006). Sustaining skills in headship: Professional development for experienced headteachers. *Educational Management Administration and Leadership* 34(1), 89–104.
- Uline, C. L., Miller, D. M., and Tschannen-Moran, M. (1998). School effectiveness: The underlying dimensions. *Educational Administration Quarterly* 34(4), 462–483.
- Wilson, M. (2001). Comprehensive school governance and special educational needs provision: Policy, practice and future priorities. *Educational Management Administration and Leadership* 29(1), 49–62.
- Winter, P. A. and Morgenthal, J. R. (2002). Principal recruitment in a reform environment: Effects of school achievement and school level on applicant attraction to the job. *Educational Administration Quarterly* 38(3), 319–340.
- Young, M. D. and McLeod, S. (2001). Flukes, opportunities, planned interventions: Factors affecting women's decisions to become school administrators. *Educational Administration Quarterly* 37(4), 402–502.

Relevant Websites

- <http://www.cceam.org> – Commonwealth Council for Educational Administration and Management.
- <http://eqq.sagepub.com> – Educational Administration Quarterly.
- <http://ema.sagepub.com> – Educational Management Administration and Leadership.
- <http://www.emeraldinsight.com> – Emerald, International Journal of Educational Management.
- <http://www.emeraldinsight.com> – Emerald, Journal of Educational Administration.
- <http://www.ingentaconnect.com> – Ingenta Connect, International Journal of Leadership in Education.
- www.rowmaneducation.com/journals/JSL – Rowman and Little Education, Journal of School Leadership.
- <http://www.tandf.co.uk> – Taylor and Francis Group.

Research on Educational Leadership – Approaches/Promising Directions

W A Firestone, Rutgers University, New Brunswick, NJ, USA

V M J Robinson, The University of Auckland, Auckland, New Zealand

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Backward mapping – A research or planning strategy that starts with the intended outcome and works backwards to identify the most proximate variables or conditions likely to affect that outcome. After clarifying how the proximate variables influence the intended outcome, the process works farther back to other layers of variables until the most distant ones of interest are identified. That is, one starts from the intended effect rather than the intended innovation.

Indirect effects – When a first variable influences a third variable through a second variable. For instance, leadership may influence student achievement through its effect on teaching.

Instructional leadership – A theory of leadership limited to education that focuses on how leaders influence student outcomes.

Leadership – Influence based on some mix of formal authority, individual characteristics, or the perceived “rightness” of the proposed mission such that volunteers want to follow: influence that is at least partly consensual and leads to change.

Leadership content knowledge – Knowledge that leaders have about an area of the curriculum that help them supervise, make decisions about resources, and provide assistance and professional development to teachers.

Pedagogical content knowledge – the special knowledge teachers have about a subject area, different from what other “experts” have that help them teach that specific content effectively.

Transformational leadership – A general that is not education-specific—form of leadership that relies on some mix of charisma, inspiration, intellectual stimulation, and individualized consideration to change the motivations and practices of followers.

compares research on educational leadership with that in other areas of education, the effects are smaller, the designs are often less rigorous, and studies are less likely to be peer reviewed. This can be seen by comparing a recent meta-analysis of leadership studies such as that by Waters *et al.* (2003) with one on phonics (Ehri *et al.*, 2001). The field continues to be plagued by faddism that precludes accumulating findings in a useful way (Ogawa *et al.*, 2000).

Yet, since Bridges made his observation, things have improved considerably. The best studies have become more sophisticated methodologically and conceptually (Hallinger and Heck, 1996). In the early years, researchers in educational administration drew primarily on general theories of leadership to formulate their studies (Lagemann, 2000). After the Coleman report, the effective schools movement created a greater interest in how leaders might improve instructional outcomes in general and help contribute to educational equity in particular (Edmonds, 1979). By now, there is a great interest in understanding how educational leadership contributes to student learning and reducing the variation in outcomes stemming from family background (Elmore, 2000; Firestone and Riehl, 2005).

These developments have generated exciting new trends that will substantially increase the contribution of research on educational leadership policy and practice and reduce the isolation of the field from those branches of educational and social research that are more directly concerned with teaching and learning (Robinson, 2006). The purpose of this article is to introduce these trends and the theoretical and methodological approaches through which they are being pursued.

The first trend is an increasing emphasis on the impact of educational leadership on students' cognitive and non-cognitive outcomes. The second trend involves a shift from a predominantly generic approach to the study of educational leadership to one that is more focused on the knowledge, skills, and practices that are specific to educational leaders. Third, there is a shift from the study of leadership style to leaders' thinking and practices. Finally, research has moved from the study of the single heroic leader to analyzing distributed leadership. In identifying these four trends, it is not implied that the shifts are radical or that there are no rich veins still being explored in the earlier approaches. Rather, these shifts signal changes that will bring leadership research much closer

Concluding a review of leadership research administrator in 1982, Bridges quoted a predecessor to the effect that “The utility of such research is simply underwhelming in relation to its volume” (Bridges, 1982: 26). Moreover, when one

to students, teachers, and classrooms and to the questions now being raised by policymakers about the leadership of school reform.

The Concept of Educational Leadership

These new trends challenge older concepts of leadership and require greater clarity about, for example, the difference between the concept of leader and leadership. This is particularly true of work on distributed leadership which needs to be able to address the question of what it is that is distributed.

There is general agreement that leadership is an influence process – that it causes another person or group to think or act in ways they would not have done otherwise (Leithwood and Riehl, 2005). Additional specification is needed, however, because we would not want to call many types of influence – such as force, coercion, or manipulation – leadership. The distinguishing feature of leadership is that the source of the influence is either positional authority (the right of the person or institution to require things of others is accepted) or the reasonableness of what is proposed (Fay, 1987).

This account of leadership indicates that, unlike other forms of influence, leadership is fundamentally consensual, with followers agreeing to do what was asked because they judge “that the leaders occupy a position which gives them the right to command a course of action or that they possess the requisite personal characteristics of leaders, or that they seek an action that is correct or justifiable” (Fay, 1987: 121). Leadership involves change, and this aspect of leadership enables distinctions to be drawn between leading and managing. Managing maintains operations and routines while leadership garners support for their reconsideration and change. The distinction should not be drawn too sharply, however, for managers need leadership skills (to be influential) and leaders need management skills (so they can assess how well the daily routines serve organizational goals).

Both conventional positional and new forms of distributed leadership are included in the above account of leadership. It can embrace research on all formal position holders in the education system and also recognize that leadership may be exercised by anyone whose ideas or actions are influential in the context of specific tasks and activities (Robinson, 2001).

What is educational leadership? Educational leadership can be distinguished from leadership in general in several ways. One approach is to say that leadership that is exercised by those in educational institutions is by definition educational leadership. This is unsatisfactory because it rules out the possibility that some leadership activities in schools are not directed toward educational ends.

Indeed, one concern of educational leaders and policy-makers in many jurisdictions is just how much of the work of school principals, for example, particularly in self-managing systems, is not directed toward educational ends (Hallinger, 2005). Educational leadership is more fruitfully treated as those leadership acts that improve the quality of teaching and learning, either through direct intervention into teachers’ work or through creating school conditions that do so indirectly. This concept of educational leadership focuses directly on the core business of teaching and learning and strongly associates educational leadership with its improvement (Elmore, 2000). It also foregrounds the empirical work that must be done to identify the nature of the leadership practices (including the leadership values, skills, and knowledge) that are needed to engage in this work.

First Trend: An Increased Emphasis on the Impact of Leadership on Students

At least five compilations of research ask about the nature and extent of leaders’ impacts on students (Bell *et al.*, 2003; Leithwood *et al.*, 2004; Mulford, 2003; Waters, *et al.*, 2003; Witziers *et al.*, 2003). These studies suggest that leaders make a very slight to modest difference to the cognitive and non-cognitive outcomes of their students. This finding is contrary to the political and public expectation that the quality of school leadership is a powerful influence on the quality of teaching and learning. The public may have an overly romantic view of the power of their educational leaders (Meindl *et al.*, 1985). On the other hand, research approaches may not yet be capturing how leaders make a difference. There are several reasons why the latter may be the case.

First, much research on the impact of leadership on students employs theories that are not designed for this purpose, as is described below. Second, there are major methodological challenges in doing quantitative research on the impact of leadership on students. The impacts of leadership can not be reliably detected unless the numerous out-of-school conditions that impact achievement are adequately modeled and statistically controlled (Hallinger and Heck, 1996). In addition, since the path through which educational leaders make their impact on students is usually indirect, a theory is needed to guide the specification and measurement of the in-school variables. Here, there is considerable overlap between research on school effectiveness and research on leadership. The selection and measurement of student outcomes is also a considerable challenge, for while both cognitive and non-cognitive student outcomes are important, and cognitive assessments should tap work of intellectual depth, such measures are expensive to obtain. How leaders are sampled also makes a difference as outlier studies (Heck *et al.*, 1990; Marks and Printy, 2003) typically

find much stronger findings than those which use more representative samples.

In summary, the new research and policy emphasis on the impact of leadership on students brings substantial theoretical and methodological challenges in identifying the leadership practices that make a difference, and the antecedents of those practices.

Second Trend: From Generic to Educational Leadership

While generic leadership skills and knowledge (e.g., staff relationship skills) are applicable to educational leadership, those generic skills must be infused with educational content to promote educational goals. This is the point at which educational leadership theory and research needs to engage deeply with research findings on the pedagogical, assessment, curricular, and student management practices that help diverse students engage in important intellectual work. Instead of asking if leadership theory has any educational impact, an alternative backward-mapping strategy starts with research on effective pedagogy and teacher learning and identifies the role of leadership in creating the conditions that develop and sustain those practices. Theories of educational leadership would be the end result rather than the starting point of such programs of research.

Most research on educational leadership has not used this approach. One of the most frequently researched leadership theories focuses on transformational leadership. Transformational leadership is a form of leadership through which a combination of charisma, inspiration, intellectual stimulation, and individualized consideration is expected to change the motivations and purposes of followers. The theory was first developed by Burns (1978), who wanted to explain why leaders in many settings gain unusually high levels of effort and commitment from followers. It was not designed to predict how leadership affects student learning. Much subsequent empirical work was done in industry and is summarized by Bass and Avolio (1994). A recent review of the research on transformational leadership in education finds that most studies have not looked at how it influenced student achievement (Leithwood and Jantzi, 2005). Those that did found mixed results. It was more common to examine the effects of transformational leadership on student engagement with school; those results tended to be more positive.

In contrast to transformational leadership, instructional leadership theory is more educationally oriented. While there are variations in the concept, the common theme is the close involvement of leaders in establishing and overseeing the academic mission and goals of the school. Specific practices can include setting academic goals, ensuring appropriate resourcing of the teaching program, acting as a resource for solving pedagogical problems, evaluating teaching, and monitoring student progress (Hallinger, 2005). Instructional leadership theory

has its origins in studies of leaders in high-poverty urban schools that succeeded against the odds (Edmonds, 1979). Subsequently, survey instruments were developed to measure dimensions of instructional leadership and its direct and indirect impact on student learning (Hallinger, 2005). While the impacts are still modest, the multivariate research using measures of instructional leadership has produced somewhat stronger impacts on student outcomes than has similar research on transformational leadership (Hallinger *et al.*, 1996; Heck *et al.*, 1990).

These two traditions are starting to converge. Some educational research on transformational leadership has incorporated instructional leadership variables such as “providing instructional support” (Leithwood and Jantzi, 1999). In addition, there is some evidence that leaders who integrate both transformational and instructional leadership qualities achieve greater gains in student learning than leaders who score high on only one of these leadership qualities (Marks and Printy, 2003).

Even though instructional leadership research has a more explicit focus on teaching and learning, it does not draw explicitly on theories of teaching and learning to develop ideas of what leaders should do. For that step to be taken, educational leadership researchers have moved beyond organizational theories to learn from researchers studying classroom phenomena.

Third Trend: From Leadership Styles to Educationally Powerful Leadership Practices and Cognitions

Abstract constructs like transformational leadership and instructional leadership provide only the most general guidance to people in schools about what they need to do in order to help schools improve. While sensitive to varying degrees to school conditions, they are rather insensitive to teaching *per se*. As Stein and Spillane (2005) note, leadership researchers need to take advantage of the cognitive revolution of the 1980s wherein people supplemented behavioral research with a more direct focus on the thinking of teachers and students (Shulman, 1987). This shift has been very fruitful. It opened up new methods through which researchers looked closely as students went about learning and teachers went about teaching. Researchers also asked how both made sense of their day-to-day work. In the process, researchers gained greater access to higher-order thinking processes (Stein and Spillane, 2005).

In addition to identifying the teaching and learning practices that leaders need to target, leadership research needs to expand conceptions of the school setting and the elements of the school and its context on which the leader must act. One conceptual resource for thinking about how to do this comes from Hallinger and Heck’s (1996) idea that the effects of the leader’s work are mediated by school

conditions. More detailed analysis is needed to determine what those school conditions are and the actions through which leaders can influence them. With a better understanding of what teaching practices need to change, what elements in the school setting are available to support the change process, and how leaders think about the change process, researchers will be better able to ask how leaders can actually change teaching and learning.

Backward mapping takes advantage of these conceptual developments to create a conceptual framework that links student learning to leaders' thoughts and actions (see **Figure 1**). In the process, it draws on a great deal of research on teaching and learning to suggest new avenues for research on educational leadership. Here we highlight two of the several possible areas. The first is research on leadership content knowledge, an area that comes directly out of research on teaching and learning. The second is research on teachers' professional communities – a school condition that can promote the teacher learning that improves student learning.

Leadership content knowledge (Stein and Nelson, 2003) refers to leaders' understandings about the content of school subjects. This is different from the knowledge of lay people, professors, students, and teachers. It refers partly not only to the kinds of things that educated people know about a subject, and partly to the kinds of things teachers need to know – that is, the pitfalls of learning the subject and ways to overcome them – but also to special knowledge of administrators about how to assess materials, how to help other educators improve their performance, and how to explain necessary changes to the community for instance (Stein and Nelson, 2003). Leadership content knowledge derives from a very different paradigm from that used in the research on transformational and instructional leadership. Rather than focusing on what leaders do, it asks how they think (Stein and Spillane, 2005). Except for Leithwood's research on leaders' problem solving (Leithwood and Steinbach, 1995), this cognitive shift has been slow to arrive in the study of educational leadership. One of the most useful forays in that direction has been to explore leaders' content knowledge in mathematics (Stein and Nelson, 2003). This research draws on Shulman's

(1987) useful observation that the content that teachers need to know to teach well – that is, pedagogical content knowledge – is different from what researchers in a discipline need to know. Teachers, Shulman suggests, need to know the pitfalls of learning a subject and ways of representing a topic to help students overcome those pitfalls. By analogy, it was suggested, leaders may need to have special knowledge about a subject in order to help teachers become more expert instructors, to make better decisions about curriculum, to be better able to communicate about the subject to parents and so forth.

The methods used to study leadership content knowledge to date have been to examine how leaders learn about specific curricular areas. The most elaborate program of research is underway in mathematics (Nelson and Sassi, 2000). The researchers study how principals make sense of simulated and real mathematics lessons, what they attend to – that is, what teachers are doing or what students are learning – what kind of supervision they provide, and how they manage the process of selecting curricular materials. To date, this research has been largely descriptive. It explores dimensions of leaders' content knowledge and how leaders with different kinds of content knowledge engage in different administrative processes. It has also raised important problems. For instance, how can principals develop content knowledge in all the subjects in the curriculum; and if they cannot, how can they compensate? At some point, as understanding grows about how leadership content knowledge influences leaders' work, it will be important to establish the relationship between such knowledge and student learning, but that relationship will undoubtedly be complex – and like all instructional leadership relationships – indirect.

Professional community is an organizational condition that helps teachers learn to become better at instruction (Louis and Kruse, 1995). It does so by creating a school culture where teachers have the opportunity to learn from each other because there is a collective focus on student achievement, and teachers engage in reflective dialog, so they talk extensively and analytically with each other about teaching and their students. Moreover, teachers de-privatize their practice by observing in each other's classroom or sharing student work. Case studies illustrate how professional communities build commitment to improved teaching and support teacher learning of improved teaching practice (Louis and Kruse, 1995) while correlational research suggests that stronger professional community is associated with higher levels of authentic achievement (Louis and Marks, 1998).

Case study research also illustrates how leadership is important in creating professional community (Louis and Kruse, 1995). Leaders put in place some of the conditions necessary for professional community, like time for reflective dialog and classroom observations. Their actions also create a context of trust and communication that allows for shared

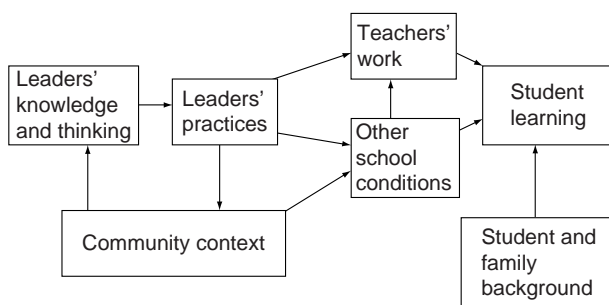


Figure 1 Conceptual framework for understanding how leadership relates to student learning.

values and authentic communication to develop. Leaders' failure to act or negative actions can undermine professional communities as well (Rollow and Bryk, 1995).

Fourth Trend: From Heroic to Distributed Leadership

Until recently the study of educational leadership focused on those who hold the most senior formal position in the organization, as if leadership impacts are attributable to this person alone. This heroic model of leadership is now being challenged by the idea of distributed leadership, which recognizes a division of labor for leadership like the more general division of labor in schools (Gronn, 2003). Rather than a single heroic leader, the distributed approach sees leadership as a set of activities in which the initiators and recipients of influence are constantly changing depending on the task at hand, the available expertise, and the willingness and skill of those involved (Robinson, 2001; Spillane *et al.*, 2004). A distributed leadership approach is critical to capturing the extent and impact of instructional leadership in schools and to discovering the role of principals in developing, coordinating, and evaluating the instructional leadership of teachers. It has also created space for the growing literature on teacher leadership (York-Barr and Duke, 2004) and, in some countries, for considering how leadership is distributed between schools and districts (Firestone and Martinez, 2007).

Research on distributed leadership needs careful theoretical and methodological development if it is to identify how the influence attempts of a wide range of staff make a difference to students. One problem with some initial writing on distributed leadership was the assumption that greater distribution was desirable in itself. Certainly one could argue that a rejection of the heroic in favor of a more distributed model of leadership is desirable on grounds of democracy and teacher empowerment. It does not follow however, that it is better for students (Timperley, 2005). Distributed leadership will serve students' learning if it is coordinated around that purpose. This means that studies are needed not only of the distribution of instructional leadership but also of how it is coordinated so that instructional goals are achieved. There are many coordinating mechanisms in organizations, not all of which require face-to-face interaction. Electronic templates, checklists, and written policies co-ordinate the activities of people who are working on part of a task and who may never meet to discuss their work. The people who design the tools around which people coordinate their work also exercise leadership (Halvorson *et al.*, 2004).

Conclusion

The study of educational leadership has come a long way since Bridges made his remarks about its limited utility.

Since then, the field has become more theoretically complex, has drawn on a greater variety of intellectual sources from the study of organizations to research on student learning, and has become more methodologically diverse and sophisticated as fits a field where the phenomena of interest are complex and interactive. The field still needs to identify a few common problems and develop a critical mass of energy around them. With any luck, the growing interest in leadership for learning will focus on that critical mass. At the same time, the combination of borrowing from other research areas, policy pressures, demands from educators and our own drive, persistence, and imagination will create the mix of rigor and relevance needed to generate the breakthroughs to really influence practice in a serious way.

See also: A Distributed Perspective On School Leadership and Management; Leadership: Instructional; Leadership: School Improvement; Transformational School Leadership; Understanding How Leadership Influences Student Learning.

Bibliography

- Bass, B. M. and Avolio, B. J. (1994). *Improving Organizational Effectiveness through Transformational Leadership*. Thousand Oaks, CA: Sage.
- Bell, L., Bolam, R., and Cubillo, L. (2003). *A Systematic Review of School Headteachers and Principals on Student Outcomes*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Bridges, E. M. (1982). Research on the school administrator: The state of the art, 1967–1980. *Educational Administration Quarterly* **18**(3), 12–33.
- Burns, J. M. (1978). *Leadership*. New York: Harper and Row.
- Edmonds, R. R. (1979). Effective schools for the urban poor. *Educational Leadership* **37**(1), 15–24.
- Ehri, L. C., Nunes, S. R., Stahl, S. A., and Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research* **71**(3), 393–447.
- Elmore, R. F. (2000). *Building a New Structure for School Leadership*. Washington, DC: The Albert Shanker Institute.
- Fay, B. (1987). *Critical Social Science: Liberation and its Limits*. Cambridge: Polity.
- Firestone, W. A. and Martinez, M. C. (2007). Districts, teacher leaders, and distributed leadership: Changing instructional practice. *Leadership and Policy in Schools* **6**(1), 3–35.
- Firestone, W. A. and Riehl, C. J. (2005). *A New Agenda for Research in Educational Leadership*. New York: Teachers college press.
- Gronn, P. (2003). *The New Work of Educational Leaders: Changing Leadership Practice in an Era of School Reform*. London: Paul Chapman.
- Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and Policy in Schools* **4**(3), 221–239.
- Hallinger, P., Bickman, L., and Davis, K. (1996). School context, principal leadership, and student reading achievement. *Elementary School Journal* **96**(5), 527–549.
- Hallinger, P. and Heck, R. L. (1996). Reassessing the principal's role in school effectiveness: A review of the empirical research: 1980–95. *Educational Administration Quarterly* **32**(1), 5–44.
- Halvorson, R., Kelley, C., and Kimball, S. (2004). Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. In Hoy, W. K. and Miskel, C. G.

- (eds.) *Educational Administration, Policy and Reform: Research and Measurement*, pp 153–188. Greenwich, CT: Information Age.
- Heck, R. L., Larsen, T. J., and Marcoulides, G. A. (1990). Instructional leadership and school achievement: Validation of a causal model. *Educational Administration Quarterly* **26**(2), 94–125.
- Lagemann, E. C. (2000). *An Elusive Science: The Troubling History of Educational Research*. Chicago, IL: University of Chicago Press.
- Leithwood, K. and Jantzi, D. (1999). The relative effects of principal and teacher sources of leadership on student engagement with school. *Educational Administration Quarterly* **35**(Supplemental), 679–706.
- Leithwood, K. and Jantzi, D. (2005). A review of transformational school leadership research 1996–2005. *Leadership and Policy in Schools* **4**(3), 177–199.
- Leithwood, K., Louis, K. S., Anderson, S. E., and Wahlstrom, K. (2004). *Review of Research: How Leadership Influences Student Learning*. New York: Wallace Foundation.
- Leithwood, K. and Riehl, C. J. (2005). What do we already know about school leadership? In Firestone, W. A. and Riehl, C. J. (eds.) *A New Agenda: Directions for Research on Educational Leadership*, pp 12–27. New York: Teachers College Press.
- Leithwood, K. and Steinbach, R. (1995). *Expert Problem Solving: Evidence from School and District Leaders*. Albany, NY: SUNY Press.
- Louis, K. S. and Kruse, S. D. (1995). *Professionalism and Community: Perspectives on Reforming Urban Schools*. Thousand Oaks, CA: Corwin.
- Louis, K. S. and Marks, H. M. (1998). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. *American Journal of Education* **106**, 534–575.
- Marks, H. M. and Printy, S. M. (2003). Principal leadership and school performance. *Educational Administration Quarterly* **39**(3), 370–397.
- Meindl, J. R., Erlich, S. B., and Dukerich, J. M. (1985). The romance of leadership. *Administrative Science Quarterly* **30**(1), 78–102.
- Mulford, B. (2003). *School Leaders: Changing Roles and Impact on Teacher and School Effectiveness*. Paris: OECD.
- Nelson, B. S. and Sassi, A. (2000). Shifting approaches to supervision: The case of mathematics supervision. *Educational Administration Quarterly* **36**(4), 513–553.
- Nelson, B. S. and Sassi, A. (2005). *The Effective Principal: Instructional Leadership for High-Quality Learning*. New York: Teachers College Press.
- Ogawa, R. T., Goldring, E. B., and Conley, S. (2000). Organizing the field to improve research on educational administration. *Educational Administration Quarterly* **36**(3), 340–357.
- Robinson, V. M. J. (2006). Putting education back into educational leadership. *Leading and Managing* **12**(1), 62–75.
- Robinson, V. M. J. (2001). Embedding leadership in task performance. In Wong, K. and Evers, C. (eds.) *Leadership for Quality Schooling: International Perspectives*, pp 90–102. London: Falmer Press.
- Rollow, S. and Bryk, A. S. (1995). Catalyzing professional community in a school reform left behind. In Louis, K. S. and Kruse, S. D. (eds.) *Professionalism and Community: Perspectives on Reforming Urban Schools*, pp 105–132. Thousand Oaks, CA: Corwin Press.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Education Review* **57**(1), 1–22.
- Spillane, J. P., Halvorson, R., and Diamond, J. (2004). Theory of leadership practice: A distributed perspective. *Journal of Curriculum Studies* **36**(1), 3–34.
- Stein, M. K. and Nelson, B. S. (2003). Leadership content knowledge. *Educational Evaluation and Policy Analysis* **25**(4), 423–448.
- Stein, M. K. and Spillane, J. P. (2005). What does research on teaching tell us about research about educational leadership? Building a bridge. In Firestone, W. A. and Riehl, C. J. (eds.) *A New Agenda: Directions for Research on Educational Leadership*, pp 28–45. New York: Teachers College Press.
- Timperley, H. (2005). Distributed leadership: Developing theory from practice. *Journal of Curriculum Studies* **37**(4), 395–420.
- Waters, J. T., Marzano, R. J., and McNulty, B. (2003). *Balanced Leadership: What 30 Years of Research Tells Us about the Effect of Leadership on Student Achievement*. Denver, CO: McREL.
- Waters, J. T. and Marzano, R. J. (2006). *School District Leadership That Works: The Effect of Superintendent Leadership on Student Achievement*. Denver, CO: McREL.
- Witziers, B., Bosker, R. J., and Kruger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly* **39**(3), 398–425.
- York-Barr, J. and Duke, K. (2004). What do we know about teacher leadership? Findings from two decades of scholarship. *Review of Educational Research* **74**(3), 255–316.

LEADERSHIP AND MANAGEMENT – LEADER PREPARATION

Contents

Leaders for Productive Schools

New Approaches in Preparing School Leaders

Research for Leaders: The National College for School Leadership

School Leadership Standards

Leaders for Productive Schools

J Murphy, S N Elliott, E B Goldring and A C Porter, Vanderbilt University, Nashville, TN, USA

© 2010 Elsevier Ltd. All rights reserved.

In this article, we develop a portrait of leaders for productive schools, or what we refer to elsewhere as leadership for learning (Murphy *et al.*, 2006). It is this type of leadership that is closely associated with schools in which all youngsters reach ambitious academic targets of performance. We draw our portrait from the school effectiveness and school-improvement literature. We capture the knowledge base of new leadership for new schools in eight dimensions and 30 functions (Table 1).

Vision for Learning

Leaders in high-performing schools devote considerable energy to “the development, articulation, implementation, and stewardship of a vision of learning that is shared and supported by the school community” (Council of Chief State School Officers, 1996: 10). On the development end of the continuum, leaders ensure that the vision and mission of the school are crafted with and among stakeholders. They also ensure that a variety of sources of data that illuminate student learning are used in the forging of vision and goals. In particular, they make certain that (1) assessment data related to student learning, (2) demographic data pertaining to students and their families, and (3) information on patterns of opportunity to learn are featured in the development process.

Effective leaders facilitate the creation of a school vision that reflects high and appropriate standards of learning, a belief in the educability of all students, and high levels of personal and organizational performance. They emphasize ambitious goals, ones that call for improvement over the *status quo*. In particular, instructionally anchored leaders make certain that goals are focused on students, feature student learning and achievement, and are clearly defined.

They ensure that responsibility for achieving targets is made explicit and that timelines for achieving objectives are specified. In short, they make sure that the school vision is translated into specific, measurable, and concrete end results. They also ensure that the resources needed to meet goals are clearly identified – and made available to the school community.

Effective principals and other school-based leaders articulate the vision through personal modeling and by communicating with others in and around the organization. On the first front, they are adept at making the school vision central to their own daily work. They demonstrate through their actions the organization’s commitment to the values and beliefs at the heart of the mission as well as to the specific activities needed to reach goals. On the second issue, communication, instructionally grounded leaders work ceaselessly to promote the school’s mission and agenda to staff, students, parents, and members of the extended school community (e.g., business and religious leaders, district office staff). Indeed, effective leaders are masters in keeping vision, mission, and goals in the forefront of everyone’s attention and at the center of everyone’s work. To accomplish this, they engage a wide array of formal and informal avenues of exchange and employ a variety of techniques (e.g., symbols and ceremonies).

Leaders for productive schools are especially well versed at translating vision into operation and in stewarding the school’s vision. They are careful monitors (1) ensuring a continuous examination of assumptions, beliefs, and values; (2) assessing implementation of goals; and (3) evaluating the impact of school objectives on organizational performance and student learning. These leaders recognize, celebrate, and reward the contributions of community members to the development, the implementation, and, most importantly, the realization of school goals. At the same time,

Table 1 Leadership for Learning

<i>Vision for learning</i>
• Developing vision
• Articulating vision
• Implementing vision
• Stewarding vision
<i>Instructional program</i>
• Knowledge and involvement
• Hiring and allocating staff
• Supporting staff
• Instructional time
<i>Curricular program</i>
• Knowledge and involvement
• Expectations, standards
• Opportunity to learn
• Curriculum alignment
<i>Assessment program</i>
• Knowledge and involvement
• Assessment procedures
• Monitoring instruction and curriculum
• Communication and use of data
<i>Communities of learning</i>
• Professional development
• Communities of professional practice
• Community anchored schools
<i>Resource allocation and use</i>
• Acquiring resources
• Allocating resources
• Using resources
<i>Organizational culture</i>
• Production emphasis
• Learning environment
• Personalized environment
• Continuous improvement
<i>Social advocacy</i>
• Stakeholder engagement
• Diversity
• Environmental context
• Ethics

they are not wanted to overlook shortcomings and failures. Certainly a critical dimension of stewarding is seeing to it that school vision and school goals shape routine school activities and anchor organizational systems and structures. On a personal front, shepherding occurs when leaders act as keepers and promoters of the vision; maintain enthusiasm and a sense of optimism, especially in periods of waning energy; and inspire others to break through barriers to make the school vision a reality.

Instructional Program

Leaders in highly productive schools have a strong orientation to and affinity for the core technology of their business – learning and teaching. In the area of pedagogy, they are knowledgeable about and deeply involved in the instructional program of the school and are heavily invested in instruction, spending considerable time on

the teaching function. They model the importance of teaching by being directly involved in the design and implementation of the instructional program. They pay attention to teaching, visiting classrooms, and working with groups of teachers on instructional issues, both in formal and informal settings.

Leaders in schools where all youngsters reach ambitious learning targets realize that teachers are the keystone of quality education. Therefore, they devote considerable time and undertake much careful planning to guarantee that the school is populated with excellent teachers, and with colleagues whose values and instructional frameworks are consistent with the mission and the culture of the school. Instructionally centered leaders are also diligent in assigning teachers to various responsibilities. These women and men allocate this critical resource based on educational criteria, especially student needs, rather than on less-appropriate foundations such as staff seniority and school politics.

Instructionally grounded leaders devote abundant time to supporting colleagues in their efforts to strengthen teaching and learning in and across classrooms. Foremost, they are aggressive in identifying and removing barriers that prevent colleagues from doing their work well. They provide intellectual stimulation and make certain that teachers have a high-quality stream of job-embedded opportunities to expand, enhance, and refine their repertoires of instructional skills. They also make sure that the materials that teachers require to perform their jobs are on hand in sufficient quantity and in a timely fashion. Consistent with the involvement and investment theme, effective leaders make themselves available to, and demonstrate personal interest in, the staff.

We know from the literature that feedback about performance is essential to the learning process, and leaders in high-performing schools are diligent about providing this information to colleagues on a consistent basis and in a timely manner. In supplying performance feedback, these women and men (1) rely on personal knowledge developed through numerous classroom observations, both informal and formal; and (2) employ a variety of supervisory and evaluation strategies. They make student learning the calculus of the exchange process. Effective leaders are especially expert in opening up a wide assortment of improvement opportunities for teachers. They are relentless in counseling poor teachers to leave the classroom. In a related vein, instructionally grounded leaders aggressively monitor the instructional program in its entirety, assuring alignment between learning standards and objectives and classroom instruction.

Academic learning time is the caldron in which student achievement materializes, and effective leaders work tirelessly with staff to ensure that this precious resource is maximized. They begin by making sure that the great bulk of time is devoted to instructional activities and that noninstructional time is kept to a minimum. They also see to it

that the majority of instructional time is dedicated to core academic subjects. Within this learning space, they work with teachers to accentuate the use of instructional strategies that maximize student engagement at high levels of success. On a parallel track, effective leaders undertake an array of activities that protect valuable instructional time from interruptions, including: (1) assigning academic subject time slots that are least likely to be disturbed by school events; (2) protecting teachers from distractions from the school office; (3) developing, implementing, and monitoring procedures to reduce student tardiness and absenteeism; and (4) ensuring that teachers are punctual. They also foster more productive use of time by coordinating time usage among teachers and across classes (e.g., all language arts instruction unfolding during the first 2 h of the day).

Curricular Program

Effective leaders are also knowledgeable about and deeply involved in the school's curricular program. They work with colleagues to ensure that the school is defined by a rigorous curricular program in general and that each student's program in particular is of high quality. On the first issue, they establish high standards and expectations in the various curricular domains consistent with blueprints crafted by professional associations and learned societies. On the second topic, they ensure that opportunity to learn is maximized for each youngster. These leaders are also diligent in monitoring and evaluating the effectiveness of the school's curricular program.

In the array of factors that define high-performing schools, curriculum alignment enjoys a position of exceptional prominence. Moreover, effective leaders are especially attentive to creating a tightly coupled curriculum throughout the school. This means that they ensure that objectives (standards), instruction, curriculum materials, and assessments are all carefully coordinated. It also means that all special programs (e.g., bilingual education) are brought into the gravitational field of the regular program. Finally, it means that there is a high degree of coordination (1) across subjects within grades, (2) across grade levels and phases of schooling (e.g., from the elementary to the middle school), and (3) among teachers within and across departments and grade levels.

Assessment Program

Assessment systems in schools with effective leaders are characterized by a variety of distinguishing elements. First, they are comprehensive. They address classroom- and school-based activity. They feature the use of a wide variety of monitoring and data collection strategies,

both formal and informal, that is, they ensure that student learning is measured using an assortment of techniques. For example, comprehensive designs often include teacher-record-keeping systems, end-of-level or end-of-unit reports, student work products, criterion-referenced tests, and standardized measures of student performance. They also highlight information gleaned from direct observations in classrooms. Second, they disaggregate information on the important conditions and outcomes of schooling (e.g., program placement of students and test results) by relevant biosocial characteristics of students (e.g., gender, race, and class). Third, they are constructed in ways that foster the triangulation of data from multiple sources in arriving at judgments about the effectiveness of curricular and instructional programs and organizational operations. Finally, as alluded to above, these systems highlight tight alignment between classroom-based and school-based methods of assessing student learning. We close here where we began, that in schools with effective assessment programs, the fingerprints of school leaders are distinctly visible.

The literature also informs us that effective leaders are master craftspersons in the communication and use of the data that is the lifeblood of the assessment system. On the issue of use, instructionally grounded leaders ensure that assessment data is at the heart of (1) mission development, (2) instructional planning, (3) the evaluation of the curricular program, (4) the identification of and the design of services for special needs students, (5) monitoring progress on school goals and improvement efforts, and (6) the evaluation of school staff. On the communication front, effective leaders provide teachers and parents with assessment results on a regular basis. They also unpack the meaning of results with staff – as a body of the whole, in appropriate groups, and individually. They make certain that information about student progress is regularly reported to students and parents in an accessible form, at multiple times, across an array of forums, and in multiple formats.

Communities of Learning

Effective school leaders are especially skillful in creating learning organizations and fostering the development of communities of learning. They are vigorous promoters of professional development, they nurture the growth of communities of professional practice, and they shape school organizations to adhere to the principles of community. In the area of staff development, effective leaders thoughtfully attend to their own growth, modeling a life-long commitment to learning for their colleagues. Unlike many peers, these women and men focus their learning on issues of school improvement. They assume an active role in planning and evaluating specific staff-learning activities and the overall professional-development system of the

school. Attending to professional development is a significant piece of their work portfolios.

In working with colleagues, instructionally centered leaders establish an expectation that the continual expansion of one's knowledge and skills focused on helping students succeed is the norm at the school. These leaders also demonstrate a dedication and a willingness to assist teachers in strengthening their instructional skills. They furnish needed resources to teachers, including support to help teachers gain new knowledge (e.g., they fund workshops, hire coaches, and facilitate intra- and interschool visitations) and they provide the materials teachers require to implement new skills in the classroom. These leaders are committed to ensuring that their colleagues have both direct and indirect, both formal and informal, guidance as they work to integrate skills learned during professional development into their portfolios of instructional behaviors. They are well versed in providing regular incidental interventions – casual conversations and suggestions of ideas – that assist teachers in their efforts to improve instruction. As we outline below, they create systems and procedures that nurture this type of informal learning throughout the school, mechanisms that promote the exchange of professional dialog about strengthening instruction and improving the school.

Instructionally centered leaders forge a structure for professional development from the principles of learning theory and models of best practice. They make certain that a robust system for developing staff expertise is in place and that each staff member has the learning experience necessary to grow his or her instructional skills. They ensure that development opportunities and experiences flow from data on student achievement, link carefully with district and school goals, are integrated into the culture of the school, and focus on student learning. These leaders also make sure that learning activities are scaffolded on the principles of adult learning.

Leaders in effective schools actively promote the formation of a learning organization, the development of staff cohesion and support, and the growth of communities of professional practice. At the broadest level, these leaders endeavor to create a culture of collaboration and the systems, operations, and policies that provide the infrastructure for that collegial culture. At this level, they also are active in building shared beliefs about the importance of community. They nurture collaborative processes (e.g., shared decision making), forge schedules (e.g., common planning time), and create organizational structures (e.g., team leadership) that permit and encourage shared mission and direction, collaborative work, and mutual accountability for school goals and student learning. These leaders are particularly attentive to ensuring that there are a variety of mechanisms for teachers to communicate among themselves. Besides, to be sure, these women and men are active participants in the various

school learning communities, often serving a key linking and pollinating role in the process. They understand and help others understand that communities of professional practice offer the most appropriate caldrons for professional learning and the forging of new instructional skills. Finally, they take advantage of the fact that they are in a unique position to garner and allocate resources to bring communities of professional practice to life.

Leaders in high-performing schools also often promote a shared or team approach to leading the organization. The DNA of this more distributed conception of management – of pushing leadership outward to students, parents, and especially staff and helping others assume the mantle of leadership – is the privileging of expertise, rather than role, in managing the school. Effective leaders are adept in meeting this challenge. They involve others in the crafting and implementation of important decisions. They empower others and provide faculty with voice, both formal and informal, in running the school, not simply their own classrooms. They delegate often and effectively and often form leadership teams to assist in shaping the vision and in managing the operations of the school, especially in and around the core technology.

Resource Allocation and Use

High-performing school leaders have a gift for acquiring, allocating, and using resources to promote student success. Indeed, researchers in the areas of school improvement and instructional leadership consistently report that high-performing school leaders are more successful than their peers in locating and securing additional resources for their schools. In particular, these women and men use the formal and informal channels at their disposal to influence district-level decision making to better the competitive position of their schools in the distribution of resources. They also show adeptness in attracting additional funds and materials from the larger school community. Evidence is also emerging that effective leaders are more skillful than their peers in building up the stock of social capital at the school level.

On the allocation and use issues, two critical dynamics are in play in high-performing schools. First, leaders here assiduously link resource deployment and use it to attain the mission and goals of the school. All requests and all commitments are not equal in effective schools. Financial, human, and material resources are all directed in the service of improved student learning. Second, effective leaders are masters at taking the dimensions of work that have historically occupied centerstage in school administration – management, politics, organization, finance – and ensuring that they are no longer ends in themselves but assume importance to the extent that they strengthen the quality of the instructional and curricular program and enhance student learning.

Organizational Culture

Effective organizations in all sectors, including education, are marked by a strong production emphasis (Bossert *et al.*, 1982: 37). And consistent with the core theme of this article, leadership is a key factor in explaining the presence of this organizational orientation toward outcomes. On the front end of this condition, leaders in high-performing schools work ceaselessly to create an environment of high-performance expectations for self, staff, and students. They model risk taking in the service of attaining important goals. They regularly communicate a concern for and interest in staff performance and student achievement. They establish clearly defined, school-wide academic standards to bring expectations to life. They carefully ensure that these high expectations are translated into school policies and behavioral expectations. These leaders make certain that expectations are decoupled from beliefs about biosocial characteristics of students.

On the other side of production emphasis, effective leaders maintain school accountability. They hold everyone – students, teachers, parents, and school administrators – responsible for achieving school goals and reaching targets in the area of student performance, providing special weight to the contributions of teachers and other professional staff at the school. While these leaders acknowledge the value of hard work, they clearly couple success to performance.

Instructionally grounded leaders are the catalysts in school-based efforts at continuous improvement. They understand and communicate that complacency is the enemy of improvement, that *status quo* is more tightly linked to decline than to growth. These leaders confront stagnation. They ensure that the school systematically reviews and adopts more productive strategies to accomplish important goals. They take risks and encourage others to do so in the quest for better education. They act entrepreneurially to support school-improvement efforts. They encourage initiative and proactiveness. They make sure that the assessment program we described earlier is a driver in the work of continuous school improvement.

As with many of the areas we explore in this article, leaders have a dual role in the domain of learning environment. On the one hand, they demonstrate what is valued through their own behaviors. Thus, effective leaders model appropriate behavior by personally enforcing discipline with students and by confronting problems quickly and forcefully. On the other hand, these leaders are responsible for the creation and operation of systems and structures and the performance of colleagues. In this area, they ensure that operations, rules, and procedures to maintain discipline and order in the school community are developed and monitored on regular basis.

There is a fair amount of research that shows that impersonality reigns in many schools in America, especially

secondary schools. That is, students are neither well known nor particularly well cared for. Since schools have been constructed using institutional and hierarchical blueprints, both of which feature impersonality, this condition should come as a surprise to no one. Yet, the fact that it can be explained is not much consolation to the youngsters in these schools. On the other side of the ledger, we know that in schools where academic and social learning thrive high, academic press is almost always coupled with high personalization. At the broadest level, this indicates that each student is well known and cared for, that each youngster feels valued and important at school.

Recognition and rewards also fill a central cell in the personalization design in high-performing schools. In these communities, an abundance of classroom-based and school-wide recognition systems and mechanisms are in play, systems that are carefully designed to be reinforcing. Rewards are distributed frequently and they reach a high percentage of students. They are seen as meaningful and important throughout the school community, especially to students. They are often public in nature. They highlight the accomplishments of individuals and groups. And while they unquestionably privilege academic accomplishments, rewards are provided for success in a wide array of areas. We close our narrative here with an important reminder: Leadership is the central ingredient in ensuring that these frameworks of meaningful student engagement and widespread rewards and recognition become defining elements of school culture.

Social Advocacy

Working from a research base that overlaps quite extensively with the one we examine here, the framers of the Interstate School Leaders Licensure Consortium (ISLLC) standards for school leaders concluded that one of the defining characteristics of highly productive leaders is that they “understand, respond to, and influence the larger political, social, economic, legal, and cultural context of schooling to promote the success of all students” (Council of Chief State School Officers, 1996: 25). That is, they actively manipulate the environment in the service of better education for youngsters and their families. The central issue here is understanding contextual trends and influences and their potential impacts on the school and the larger community, particularly how these environments support or hinder learning in classrooms. In their role as social advocates, effective leaders proactively respond to external policy initiatives (e.g., speak at public forums, address civic organizations) to ensure that public policy is advantageous to the children and youngsters in their schools – and their families.

Effective leaders demonstrate an understanding of and a commitment to the benefits that diversity offers to the

school. They translate this knowledge and commitment into work that creates educational experiences that honor diversity (e.g., the use of culturally rich educational materials) while strengthening instruction and improving student achievement. As we see below, these leaders are also adept at building and using channels of communication that promote ongoing dialog with diverse groups of stakeholders.

According to the literature, most noticeably the ISLLC standards, effective leaders “act with integrity, fairness, and in an ethical manner” (Council of Chief State School Officers, 1996: 18). On one front, the authors of the standards note that this means that leaders fulfill legal and contractual obligations and apply laws and district and school policies and procedures fairly, wisely, and considerately. It means that they guarantee the privacy rights of students and recognize and respect the legitimate authority of others. At a deeper level, it means leaders treat others fairly, equitably, and with dignity and respect – and they establish the expectation that others in the school community act in a similar manner.

On a personal basis, effective leaders are more cognizant than their peers of their own values and beliefs and they shape their behavior in accord with personal and professional codes of ethics. They are more reflective and self-critical about their own practice and its impact on others in the extended school community. They know the difference between using office and position for one’s own gain and for the benefit of the school community and they honor the latter. These leaders serve as role models in terms of accepting responsibility for what happens to children and families in their school community.

Finally, the research on high-performing schools and instructionally oriented leaders reveals that effective leaders are attuned to and expert at linking the school to parents and others in the extended school community. Much more so than peers, these leaders weigh connections in terms of their value in enhancing the academic and social learning of students. That is, they engage families and other community members in the service of school goals, the learning agenda, and student performance. Inside the school, these women and men model community collaboration for staff, establish norms about the importance of parent connections, and provide opportunities for staff to develop the collaborative skills needed to work effectively with parents. They also ensure that information about family and community concerns, expectations, and interests inform school decisions.

Effective leaders craft and work from a comprehensive design about school–community relations that is anchored by the school’s academic mission. The plan is systematic, not simply a collection of *ad hoc* and unrelated activities. In the wider community, these leaders develop relationships with influential actors in the religious, business, and political sectors. They are actively involved in the school community and communicate frequently with stakeholders therein. They employ multiple channels and a variety of

forums to operationalize these connections. Their objectives are to inform, promote, learn, and link – to ensure that the school and the community serve one another as resources. On the extended community front, effective leaders are also specially attentive to building bridges with (1) other youth and family-service agencies that can promote better lives for youngsters and their families and (2) the media that can help promote the image of the school.

For effective leaders, connections with parents occupy a strategic position in the algorithm of stakeholder engagement. Leaders communicate with families regularly and through a variety of channels. They create programs and strategies that bring parents from the periphery to the inner circle of school operations.

Acknowledgment

This article was developed with the help of a Wallace Foundation grant on leadership assessment to Vanderbilt University. The generous support of the Wallace Foundation is gratefully acknowledged.

Bibliography

- Bossert, S. T., Dwyer, D. C., Rowan, B., and Lee, G. V. (1982). The instructional management role of the principal. *Educational Administration Quarterly* 18(3), 34–64.
- Council of Chief State School Officers (1996). *Interstate School Leaders Licensure Consortium: Standards for School Leaders*. Washington, DC: Council of Chief State School Officers.
- Murphy, J., Elliott, S. N., Goldring, E., and Porter, A. C. (2006). Learning-centered leadership: A conceptual foundation. *Paper Prepared for the Wallace Foundation*. New York.
- ## Further Reading
- Barker, B. (2001). Do leaders matter? *Educational Review* 53(1), 65–76.
- Barnett, K. and McCormick, J. (2004). Leadership and individual principal–teacher relationships in schools. *Educational Administration Quarterly* 40(3), 406–434.
- Beck, L. G. and Murphy, J. (1996). *The Four Imperatives of a Successful School*. Thousand Oaks, CA: Corwin.
- Knapp, M. S., Copland, M. A., Ford, B., et al. (2003). *Leading for Learning Sourcebook: Concepts and Examples*. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.
- Leithwood, K., Louis, K. S., Anderson, S., and Wahlstrom, K. (2004). How leadership influences student learning. *Paper Commissioned by the Wallace Foundation*. Minneapolis, MN: University of Minnesota.
- Marzano, R. J., Waters, T., and McNulty, B. A. (2005). *School Leadership That Works: From Research to Results*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Murphy, J. (1990). Principal instructional leadership. In Lott, L. L. and Thurston, P. W. (eds.) *Advances in Educational Administration: Changing Perspectives on the School*, vol. I, part B, pp 163–200. Greenwich, CT: JAI Press.
- Murphy, J., Beck, L., Crawford, M., Hodges, A., and McGaughy, C. L. (2001). *The Productive High School: Creating Personalized Academic Communities*. Thousand Oaks, CA: Corwin.
- Sammons, P., Hillman, J., and Mortimore, P. (1995). *Key Characteristics of Effective Schools: A Review of School Effectiveness Research*. London: Institute of Education.

New Approaches in Preparing School Leaders

S G Huber, University of Teacher Education Central Switzerland (PHZ) Zug, Zug, Switzerland

© 2010 Elsevier Ltd. All rights reserved.

Glossary

MOE/NIE – Ministry of Education/National Institute of Education

NCSL – National College of School Leadership

NSO – Nederlandse School voor Onderwijs management (Netherlands School for Educational Management)

The pivotal role of the school leader as a factor in effective schools has been corroborated by findings of school effectiveness research for the last decades. (The term school leader, in this article, is used instead of principal, headteacher, administrator, rector, or other terms describing the person who is in charge of an individual school.) Extensive empirical efforts of the quantitatively oriented school effectiveness research – mostly in North America, Great Britain, Australia, and New Zealand, and also in the Netherlands and in the Scandinavian countries – have shown that leadership is a central factor for the quality of a school. The research results show that schools classified as successful possess a competent and sound school leadership (there is a highly significant positive correlation). The central importance of educational leadership is therefore one of the clearest messages of school effectiveness research (see Gray, 1990). In most of the lists of key factors (or correlates) that school effectiveness research has compiled, leadership plays such an important part, so much so that the line of argument starting with the message, schools matter, schools do make a difference, may legitimately be continued as: “school leaders matter, they are educationally significant, school leaders do make a difference” (see Huber, 1997).

Studies on school development and improvement also emphasize the importance of school leaders, especially in the view of the continuous improvement process targeted at an individual school. For all phases of the school development process, school leadership is considered vital and is held responsible for keeping the school as a whole in mind, and for adequately coordinating the individual activities during the improvement processes. Furthermore, leadership is required to create the internal conditions necessary for the continuous development and increasing professionalization of teachers. It holds the responsibility for developing a cooperative school culture. In this connection, the role model function of the school leader is worth mentioning.

Leadership Development

In view of the ever-increasing responsibilities of school leaders for ensuring and enhancing the quality of schools, school leadership has recently become one of the central concerns of educational policymakers. In many countries, the development of school leaders is high on the agenda of politicians of different political wings. At the beginning of the new century, there seems to be a broad international agreement about the need for school leaders to have the capacities required to improve teaching, learning, and pupils’ development and achievement. On looking more carefully, however, it is apparent that a number of countries have engaged in this issue more rigorously than others. While in some countries discussions of school leader development are mainly rhetoric, elsewhere concrete steps have been taken to provide significant development opportunities for school leaders. Hence, a closer examination of school leadership development opportunities in different countries is instructive.

In this article, we first use a comparison drawing on data from an international study of school leadership development programs (see Huber, 2004). This project on school leadership development was based on researching, analyzing, contrasting, comparing, and discussing programs of 15 countries in Europe, Asia, Australia/New Zealand, and North America. The report surveys the development models for school leaders in those countries. It describes international patterns in school leadership development and provides recommendations based on current trends. A broad variety of school leadership development approaches and models became apparent from this project. Second, we use our experiences not only in developing program designs, but also in implementing training and development opportunities.

International Approaches

Table 1 summarizes school leadership development models in 15 countries. It is meant to provide an accessible overview of predominant approaches in use across Europe, Asia, Australasia, and North America.

In spite of the differences in cultural and institutional traditions, the study’s findings underpin a number of international patterns or tendencies in school leadership development seen from a global perspective. While some of them may be viewed as differences in emphasis, others may be so

Table 1 Overview of current approaches to develop school leaders*Europe*

Denmark

Optional offers made by municipalities, universities, and private suppliers without any central framework or delivery system

Sweden

A national preparatory program offered by universities through a basic course plus additional offers by the municipalities

England and Wales

A centrally organized program delivered by regional training centers; combines assessment and training with a competency-based and standards-driven approach; the program is embedded in a three-phase training model

France

A mandatory, centrally designed, intensive, full-time, half-year preparation program with internship attachment for candidates who have successfully passed a competitive selection process; completion guarantees a leadership position on probation (during which further participation in training is required)

Netherlands

A broad variety of different optional preparatory and continuous development programs by different providers (e.g., universities, advisory boards, and school leadership associations) in an education market characterized by diversity and choice

Germany

Courses conducted by the state-run teacher training institute of the respective state, mostly after appointment; differs from state to state in terms of contents, methods, duration, structure, and extent of obligation

Austria

Mandatory centrally designed, modularized courses post-appointment; delivered by the educational institute of each state; required for continued employment after 4 years

Switzerland

Quasi-mandatory, canton-based, modularized programs offered post-appointment; delivered by the respective provider of the canton, most often the teacher training institute, wherein the aim is nationwide accreditation (national standards are currently being developed)

South Tyrol, Italy

A mandatory program for serving school leaders to reach another salary level as becoming Dirigente; delivered by a government-selected provider that combines central, regional, and small group events with coaching attachment

Asia

Singapore

A mandatory, centrally controlled, preparatory, 9-month, full-time program provided through a university; comprised of seminar modules and school attachments

Hong Kong, China

A centrally designed, mandatory, 9-day, content-based induction course immediately after taking over the leadership position

Australasia

New South Wales, Australia

An optional, modularized, three-phase program offered by the Department for Education; centrally designed, yet conducted decentralized through regional groups; besides, there are offers by independent providers

Continued

Table 1 Continued

New Zealand

A variety of programs with variation in contents, methods, and quality; conducted not only by independent providers, but also by institutes linked to universities; no state guidelines, standards, or conditions for licensure

North America

Ontario, Canada

Mandatory, preparatory, university-based, 1-year, part-time program delivered through several accredited universities following a framework given by the College of Teachers (the self-regulatory body of the profession)

USA:

Washington

New Jersey

California

Mandatory, intensive, preparatory, 1-year, university programs that include extensive internship attachments; programs use a broad variety of instructional methods

Huber (2004)

significant as to represent paradigm shifts. The largest differences are evident in those countries with longer experiences in school leadership development and school leadership research. In the following sections, these trends will be explored (for a full account, see Huber, 2004).

International Trends

Central Quality Assurance and Decentralized Provision of Programs

Provision of development opportunities for school leaders varies broadly across the countries (as shown in **Table 1**). There are different degrees of centralization and decentralization with regard to how much choice prospective participants have over available providers and development programs. Here, the interrelation between the approach to school leader development and the educational policy and school system background is of particular interest. The countries can be categorized in terms of these two dimensions (see **Table 2**).

In some centrally organized school systems (see **Table 2**, cell A), there is a centrally regulated development program. It has a standardized approach and its delivery is centrally organized. The program is mandatory for all school leaders. In contrast, in some decentralized school systems (see **Table 2**, cell D), there are a variety of programs offered by competing providers. The choice of which program(s) to attend is up to the individual (aspiring) school leader. Here, the governments abstain from any regulation or control of professional development. Countries with a predominantly centralized school system and with an entrepreneurial approach to school leader development could – not too much surprisingly – not to be found in the study.

Table 2 Centralization and decentralization of school systems and school leader development

		<i>Approach to school leader development</i>	
		Predominantly centralized or using standards or guidelines	Entrepreneurial
<i>Level of central control over school management</i>	Predominantly centralized	A France; South Tyrol; Austria; Germany; Hong Kong; Singapore	B
	Substantially devolved	C Ontario, Canada; USA ^a ; NSW; Australia; Sweden; England and Wales Switzerland	D Denmark; Netherlands; USA ^a ; New Zealand

^aDouble listing is due to differences in the approaches of the different states.

Another existing variant, however, is represented by countries with decentralized school systems (see **Table 2**, cell C), whose programs are designed according to central guidelines, but are not standardized in every detail. Their general approach seems particularly progressive and pioneering. In North American countries, responsibility for designing and conducting qualification programs (preservice training) lies primarily with universities (e.g., Ontario, Canada as well as in the US examples included in the study). However, these universities are not completely independent when setting up their development programs. They must take centrally developed goals and standards into account.

Generally, there is a developing trend in which the responsibility for designing goals and programs, and assuring quality lies with a central institution, while delivery is decentralized. Teachers who want to qualify for a leadership position can choose among various service suppliers, but rely on a fundamental level of quality assurance due to established guidelines or standards, and on nationwide acceptance and recognition of programs by employing bodies.

New Forms of Cooperation and Partnership

New models of partnerships in numerous countries can be viewed as the second trend. These arrangements were created to conceive, implement, supervise, and evaluate school leader development programs. The most striking

feature of this development, however, is the fact that representatives of the recruiting committees (either state or local) of the colleges of education at the universities, and an increasing number of representatives of the profession itself (predominantly from professional organizations, and also from local schools) are now included as well.

It becomes apparent that much of the coherence that characterizes the new programs in these countries is due to this cooperation. These groups contribute a variety of perspectives concerning the essential content of the programs, the teaching strategies and learning methods, and the organizational and chronological conception of the programs; in other words, their conception on a macro- and micro-didactic level. These partnerships have also contributed to the creation of a pool of highly qualified and accredited or certified trainers in some countries.

Dovetailing Theory and Practice

Partnerships such as these have also contributed to the increasing combination of the theoretical and practical aspects of school leadership development, which is an important and yet, difficult task to achieve.

It might sound axiomatically that theory has to be made accessible through practice and vice versa. Seemingly, it has never been easy to achieve both at the same time. In many of the countries investigated, it was perceived that either development programs emphasized theory and were theoretically, respectively academically oriented or were focusing on a more practical approach based on experiences of practitioners. Both models, therefore, seem to suffer from a one-sided approach, and do not seem to be attracting participants or leading to the expected increase in knowledge, understanding, skills, and abilities. Only a more balanced model leads to participant satisfaction and is a suitable method to meet the participants' needs. It is thus safe to assume – although this aspect has not yet been investigated sufficiently – that dovetailing theoretical and practical aspects is essential for designing effective development programs which aim at changes in one's behaviors and dispositions. Admittedly, school leaders themselves seem to prefer what they refer to as practical experience and, at times, regard theoretical and academically oriented topics as less useful. It can be seen (see West *et al.*, 2000), however, that they find it much easier to deduce general knowledge from their experiences and to use effective strategies when they have a theoretical conceptual framework that underpins their decisions and actions.

Theory and practical experiences are interdependent and therefore have to be developed together. The partnerships indicated above appear to be a suitable starting point, since research is conducted alongside the development programs and can affect the development concepts.

Hence, research-based training concepts are realized. This connection requires partnerships between the individuals working at schools and those who research and study schools. This will more effectively link the work carried out in both areas. Mutual respect and collaboration between both groups are essential for this to occur.

Preparatory Qualification

Effective school leadership requires a demanding set of attitudes, attributes, skills, knowledge, and understanding. Thorough training and development, starting with appropriate preparation prior to assuming the position, has been recognized as undoubtedly vital. Hence, in many countries, development opportunities are scheduled before taking over school leadership. Preservice preparation is offered instead of relying solely on in-service training. Moreover, these programs differ as to whether they are optional or mandatory (see **Table 3**).

In countries that have mandatory preparation (see **Table 3**, cell A), taking part in the program is an

important selection criterion for future employment as a school leader.

In countries where preparation programs are optional (see **Table 3**, cell C), there is a tendency among employing bodies toward expecting some preparation for the position. An alternative trend finds the provision of in-service training immediately after appointment and before taking over the leadership position.

Extensive and Comprehensive Programs

Taking the school leaders' role seriously and regarding school leadership as a profession have implications concerning the extent of training and development provision for school leaders. In some countries, the school leader is no longer seen as *primus inter pares*, as being a teacher with only a few additional responsibilities. Hence, some development programs have become quite time consuming, comprising around 100 or more course days.

By way of illustration, some examples are given here from North America, Europe, Asia, and Australia/New Zealand (see **Table 4**). It is important to mention that all of the programs listed here are preparatory, which means that they all take place before appointment (except the offer from the Netherlands, which may also be attended after appointment). This suggests the increasing recognition of school leader professionalization.

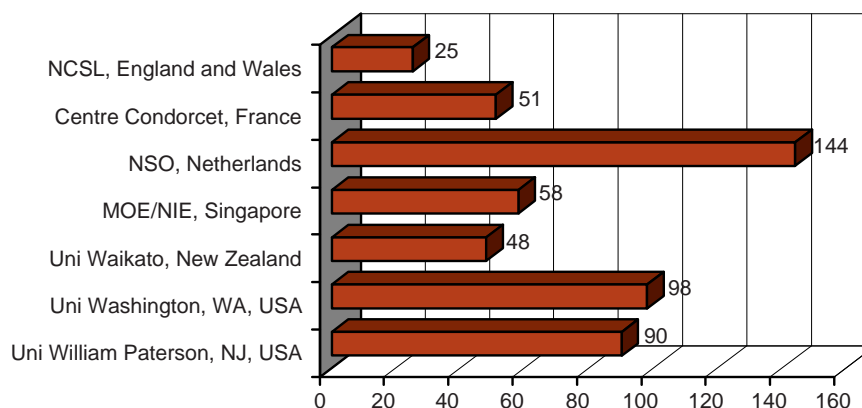
While **Table 4** indicates only the number of course days, the real demands on the time of the participants are apparent when we consider that beyond contact time there is other time committed to preparation. Additionally, activities include individual study time for readings and writing assignments, time for internships or school-based projects, and the documentation of one's progress and reflection by writing a learning journal to mention but a few examples.

Table 3 Timing in participants' career and nature of participation

		Preparatory	Induction
Mandatory	A	Ontario, Canada; USA; France; Singapore	B Germany ^a ; Austria; Switzerland ^a ; South Tyrol; Hong Kong
Optional	C	England and Wales; Netherlands; NSW, Australia; New Zealand	D Denmark; Sweden; Germany ^a ; Switzerland ^a

^aDouble listing due to differences in the approaches of the German Laender or Swiss Kantone.

Table 4 Length of school leader preparation programs (contact time)



Multiphase Designs and Modularization of Programs

Clearly, providers are increasingly moving toward multiphase development models based on a coherent conceptual approach and comprising orientation, preparation, and induction phases for aspiring and newly appointed school leaders, as well as a continuous development phase and a reflective phase for established school leaders.

There is also a trend toward providing professional development through a series of modules. This takes two general forms. In the first form, the models are conceived of as a mandatory sequence of rounded single programs. In the second form, there is no specific sequence for completing the modules. Rather participation in the modules depends on the professional position and development needs of the individual participant. The modules may be collected in a kind of personal portfolio. The individual school leader may well fall back upon them as support in crucial career phases.

In general, there is a tendency away from one-size-fits-all designs and toward programs tailored for the individual participant through modularization. Development should be linked to the career cycle and to specific needs of the leader, both personally and dependent on the particular school. Needs assessment and choice, rather than a flexible curriculum, helps in this matter (Table 5).

Personal Development Instead of Training for a Role

As the role of school leaders is becoming increasingly complex, it becomes more and more evident that it is no longer sufficient to train potential candidates or school leaders for a fixed role, whose model may be quickly outdated. Instead, aspiring school leaders must develop a

vision within the context of their school and adapt their role and responsibilities to that context. To achieve successful adaptive leadership, the programs of some countries include components such as personal vision, personal and professional development, development of fundamental values and of one's ability to reflect, time and self management, developing mental models of the organizational structure, and activities in the school that mirror good leadership activities. Moreover, day-to-day school or internship experiences have become reflective activities that result in constant re-conceptualization.

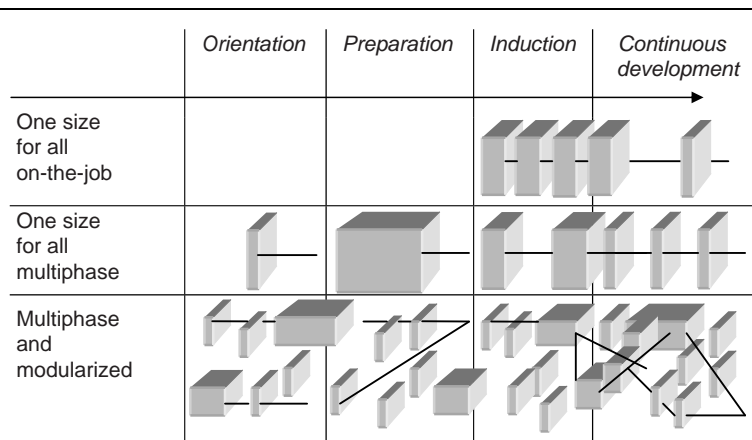
As far as qualifications are concerned, the emphasis has shifted from focusing on a specific role to a broader one that concentrates on personal learning and one's needs in the areas of knowledge, dispositions, and performances that would be useful in a more complex environment. Often, training for a management position has been replaced by offering professional development opportunities for one's leadership style. It is the personality of the (aspiring) school leader that becomes the focal point of any program.

The Communicative and Cooperative Shift

In spite of the increasing stress on school leaders due to the complexity of the role – particularly in countries with more devolved systems – school leadership programs are not preoccupied with administrative topics. On the contrary, the overall focus of school leadership programs is no longer on administrative and legal topics as it used to be in earlier programs, but has shifted to topics that focus on communication and cooperation.

The image of school leaders as experts in administration has shifted to school leaders as experts in communication and cooperation. This trend has become another international paradigm shift. Topics such as communication,

Table 5 Phased models of school leadership development



The individual squares represent courses; the black line signifies the sequence of participation.

motivation, collaboration, collegiality, and cooperation are essential parts of all programs. Internationally, there is the recognition that understanding and effectively using these topics is essential for one to become a successful school leader.

Communication and cooperation, as essential components in leadership development programs, also play an important role as far as the methods applied in those programs are concerned. Realizing that learning processes that take place in groups provides participants with better opportunities for experiential learning, more programs are moving in the direction of small- and large-group interaction. The aim then becomes one of creating reflective practitioners and this will intensify the teaching–learning experiences. In addition to traditional seminars, collegial learning—learning together with other colleagues—is being realized through a variety of strategies, including peer-assisted learning, peer coaching, critical partnerships, acquiring knowledge from experienced peers by shadowing, or through mentoring programs or collegial networks (professional learning communities) that were created (e.g., as a result of experiences from the cohorts that existed during other training programs). When one uses these strategies, learning evolves through mutual reflection and problem-solving processes; it is about learning with and from colleagues.

From Administration and Maintenance to Leadership for Improvement

Schools are more and more viewed systemically as learning organizations, each with their own specific conditions, rules, and cultures. Consequently, leading schools entails developing learning organizations (see Senge, 1990; Fullan, 1993, 1995). Hence, the overall focus is not on managing schools with an emphasis on maintenance, but has shifted to a focus on leadership for improvement.

The central task of school leadership is the continuous improvement of the school in cooperation with all stakeholders and agents. This is mirrored in the choice of contents and methods of development programs, which take into account that school leaders must be educational leaders and that is about initiating, supporting, and sustaining substantive and lasting change as well as continuous improvement in schools for the benefit of pupils. Communication and cooperation, school development and staff development, and evaluation and quality assurance play an important part. The focus is then on a collaborative and collegial style of leadership.

Qualifying Teams and Developing the Leadership Capacity of Schools

The conceptualization of school as a learning organization increasingly shifts the focus away from the development of the individual school leader to the development

of each individual school's leadership capacity. In consequence, the school leadership development program becomes a means of school development. With this in mind, some providers now explicitly focus not only on (aspiring) school leaders, but also on teachers who want to enhance their leadership competencies even if they are not planning to apply for school leader positions. A few programs even target whole school leadership teams, and may include parent and community representatives.

As an additional note, this new focus on developing team leadership capacity has interesting implications for program contents. When a program focuses on a team, development activities must become even more contextualized: it is no longer context-free training, but context-specific applied development.

From Knowledge Acquisition to the Creation and Development of Knowledge

When rapid social and economic change and changes in the educational system are coupled with a global increase in information production, it is insufficient for programs to focus solely on enlarging the quantity of leaders' knowledge. The qualification must prepare for an unknown future environment. This suggests yet another paradigm shift. It is a shift away from imparting a stable knowledge base and toward the development of procedural knowledge that can be applied. The notion of acquiring knowledge is being replaced by the concepts of developing or creating knowledge and by information management. The participants will enhance their ability to learn, understand cognitive processes, and achieve what is referred to as conceptual literacy (see Giroux, 1988). They have to be enabled to act in a complex, sometimes chaotic, work environment (see Murphy, 1992).

Needs-, Experience-, and Application Orientation

There is consensus that delivery methods must address the learning needs and competences of adult learners. Hence, fundamental andragogic principles must be taken into account. This reflects the belief that new knowledge is built on previous experiences and the knowledge of the adult learners. Adults bring personal and professional experiences, prior knowledge, and their own personal ways of seeing themselves to bear on the learning process to a greater degree than children (see Siebert, 1996). Themes that cannot be linked to previously existing cognitive systems are often quickly forgotten. The reality and the experiences of the participants, their needs and problems, should therefore become the starting point of new learning. Consequently, methods of learning tend to favor a problem-centered rather than theme-centered approach. According to Gruber (2000), gaining experience for professional competences means learning in complex

application-relevant and practice-relevant situations (see also Joyce and Showers, 1988/1995). New competences are mostly gained in a process of practice and feedback. For this reason, sufficient theoretical foundations should be imparted as well in order to foster reflection on practice (see also Schön, 1983, 1984).

In many development programs, there is a clear tendency toward experience-oriented and application-oriented methods and toward practice-with-reflection-oriented learning.

New Ways of Learning: Workshops and the Workplace

Increasingly, the participants are placed in a workshop surrounding, and confronted with modeled situations of school leadership work life and carefully constructed cases. They may be involved in teams in problem-based learning (PBL) where learning is cooperative, interactive, participative, and, to a certain degree, group- and self-organized. More consequently than the case studies and simulations often applied in development programs, the PBL approach starts with real-life experiences and then looks for supportive knowledge as a tool. The slogan here is: first the problem, then the content (Bridges and Hallinger, 1995: 8).

Within PBL, team learning is especially critical in order to achieve solutions to problems. Problem solving is an interactive participative process.

Going one step further means using genuine cases that are taken from real schools, either from the schools of the participants or from partnership schools. Within this approach, participants of the project group become external counselors for the leaders of these schools. Through this interaction, both parties benefit.

Some development programs take a step further: leave the workshop and turn to the authentic workplace, using it as a clinical faculty. It is argued that only the authentic working context can assure an adequate complexity and authenticity leading to learning processes required. For the participants of preservice school leader development, internships at one school or several schools are organized parallel to the training. They can observe the school leader by shadowing her/him, can partially take over leadership tasks themselves, and can carry out projects independently. The school leaders at the internship schools then function as mentors or supervisors and will also benefit from this cooperation. In general, new partnership arrangements between universities, other providers of school leadership development, and schools are an important basis for learning opportunities such as these. Thus, certainly the best possible practice relevance is created: exemplary learning processes take place in the reality of school.

As Huber and West (2003) show, the training provision can be conceptualized as being spread across two continua

Table 6 Emphasis of learning opportunities within school leader development programs

<i>Experience-Based Learning</i>		<i>Course-Based Learning</i>	
Centered around experiential methods	Extensive internships	Mixed model	Centered around courses
	France; Singapore; Washington, USA	NSO, Netherlands; New Jersey, USA; Ontario, Canada; England and Wales	Germany; Hong Kong

It has not been taken into account whether the offers are made to teachers aspiring to leadership or to school leaders newly appointed and in position. Besides, the different emphasis could be viewed in reference to the total amount or length of training available; since offering experiential learning opportunities inevitably means expanding the program accordingly.

of course-based and experience-based learning opportunities. Hence, it is possible to distribute the programs worldwide according to the relative emphasis given to these two strategies (see Table 6).

Adjusting the Program to Explicit Aims and Objectives

It becomes increasingly obvious that the process of developing school leaders is becoming more professional. This also includes explicitly stating the program's aims that aspiring leaders must achieve and adjusting the program to these explicit aims and objectives. Until now, programs were not necessarily developed with explicit goals or objectives, especially in the early stages of their development. Instead, generalized statements such as school leader development aims at developing school leaders were used. Contentwise, however, the aims postulated differ greatly at a higher level of explicitness. They can be classified according to their main foci: those with an explicit functional orientation and/or task orientation, those which are distinctly competence oriented or cognition oriented, those with a definite orientation toward school improvement, and some which are clearly vision or value oriented. Moreover, they state definite and explicit expected outcomes.

As new concepts of leadership and schools emerge, based on the values of society, they increasingly begin to impact the programs.

New Paradigms of Leadership

Changes in the schools and their context also have some impact on the role of school leaders. This new role can hardly be filled with old concepts of leadership.

School leader development has to take this into account. Consequently, some of the development programs relate to new and quite specific leadership conceptions.

As schools are no longer seen as static systems, conception such as transformational leadership, are becoming more popular. Transformational leaders view school as a culturally independent organism that is able to develop. Hence, they exercise an active influence on the culture of the school. They are expected not only to manage structures and tasks, but also to concentrate on people and their interpersonal relationships. They make an effort to win their cooperation and commitment. Leadership of this type is considered more suitable for the tasks of school development (see Leithwood, 1992).

If school is to become a learning organization, this implies the active empowerment and cooperative commitment of all stakeholders and agents. Then, the previous division between the positions of teachers on the one hand and learners on the other hand cannot be maintained; nor can the division between leaders and followers. Leadership is no longer statically linked to the hierarchical status of an individual person, but empowers as many staff members as possible as partners in various parts. This is conceptualized by the notion of posttransformational leadership (see Jackson and West, 1999).

Another concept, for example, is integral leadership. It views school leaders primarily as leaders with genuinely educational tasks and emphasizes an integrating perspective, which overcomes the divide of management and leadership for the sake of the educational aims of schools (see Imants and de Jong, 1999).

Instructional leadership is another model cited very often (see, for example, de Bevoise, 1984; Hallinger and Murphy, 1985). This leadership concept focuses most on those aspects of school leadership actions that concern the learning progress of the pupils. They include management-oriented as well as leadership-oriented activities such as a suitable application of resources for teaching, agreeing upon goals, and promoting cooperative relationships between staff (e.g., preparing lessons cooperatively), but, especially, the evaluation and counseling of teachers during lessons through classroom observation, structured feedback, and coaching (see Hallinger, 2003).

Moreover, we can see concepts such as organizational-educational management and leadership (Rosenbusch, 2005; Huber, 2004), moral or democratic leadership (mainly in Europe), and, as a more eclectic model, integrative leadership (see Huber, 2004; Huber and Moos, 2007).

Orientation toward the School's Core Purpose

School leadership development programs are more strongly oriented toward the schools' core purpose, namely teaching and learning, and the specific aims of schools within society today and in the future.

The idea of school as an administered addition of lessons, as the lowest unit in the school system's hierarchy is no longer up to date. Many changes in the school system have occurred in recent years, even in the more centralized school systems. Therefore, it would be quite important to make the aims of school in society and possible goals of schools of the future as central themes in the program. These should then be discussed by the participants and adapted into their thinking framework for aligning their action to these goals.

These reflections on leadership activities, the school, its role, and function are explicitly made in some countries. Increasingly, there is more conceptual elaboration of the aims toward which schools are to be developed. The aims of school leader development programs should answer questions such as these: What is school and schooling about? What is leadership and management about? What is the core purpose and what should be the aims? What kind of training and development opportunities are therefore needed to prepare and support (aspiring) school leaders in adjusting their perspectives, conceptualizing their role and function, developing the necessary competences, and mastering the manifold tasks within the individual school in order to provide conditions and support staff so that effective and efficient teaching and learning take place for the sake of the pupils?

This multilevel adjustment of aims should be essential designing, implementing, and evaluating school leadership development programs and should shape the programs with regard to contents, methods, patterns in terms of timetabling, etc. (see Huber, 2004).

Conclusion

A comparison of school leader development programs gives a dominant impression of global approaches and shifts. What can be clearly stated about school leader development from this international perspective is that there have been many changes during the last years in many countries. In other countries, this process has just started. School leadership and leadership development are high on the agenda of educational policymakers.

Obviously, many of the countries that have enhanced their leadership development programs have increasingly focused on linking leadership development with school development. Here, increasing the leadership competences of an individual is seen as a component of building the leadership capacity of the whole school.

To sum up, we find two new avenues in preparing school leaders: first, new ways explore the development of training and development designs, quality assurance, and the overall organization through the development of central institutions that are in charge or the setting up of standards and accreditation procedures for the providers.

Second, new ways explore the implementation and carrying out of training and development programs based on new macro- and micro-didactic considerations of instruction and learning settings with a focus on putting theory into practice and, vice versa, using experiences to develop subjective theories. (Macro-didactic considerations are about defining not only the target group(s), the timing, the nature of participation, and the professional validity, but also the pattern with the total number of training and development days, the time span, the scheduling, etc. Micro-didactic considerations are about the curriculum, the content, the teaching strategies and learning methods used, etc.)

For successful training and development opportunities, it is necessary to link stakeholders and agents in the school system both vertically and horizontally. The training needs to have a multilevel approach using this vertical and horizontal linking. It is about: (1) cooperation within the school and among schools, either position or theme oriented; (2) learning from and with colleagues, and professional learning communities within the whole school system; and (3) developing a shared language, shared concepts, and a shared culture. This approach is a vehicle for school development and has to be taken into account and mirrored by leadership training and development opportunities.

Bringing theory and practice together seems to become very important as well. The linkage is using a reflective learning approach. Reflection – also together with others (peers, seniors, and experts) – plays an important part. Moreover, training and development have to be seen as a continuous process, and have to be multiphase oriented. Additionally, programs have to be context related (to the country, the society, the school system, the individual school, and to the individual needs of the participants). Besides these premises, training and development opportunities should be needs oriented, practice oriented, and application-oriented, consequently, competence oriented. Finally, development programs – as mentioned above – require a multistage adjustment of aims. (For further concrete recommendations for program designs and their implementation, see Huber, 2004.)

Given the fact that school leadership is getting increasingly complex and that not only the tasks but also the competences are too demanding for one person alone, shared, distributed, or cooperative leadership concepts seem to be solutions discussed internationally not only in the academic community but also increasingly in the profession itself. However, it may be stated that the conception of school leadership in training and development programs, even taken internationally, still is a rather narrow one. Perhaps there does need to be one supreme head in each school. Maybe school leadership development programs are about finding and equipping such individuals. However, perhaps there are other alternatives – collective leadership, the development of whole teams of staff, and the

reconceptualization of the school leader's role as simply one part in a team, a team made up of leaders who all need support, training, and development opportunities. For us, it is this last issue that seems to challenge most forcibly the orthodoxy underpinning current provision, and that offers the most interesting avenue of exploration for the future. Particularly in the last couple of years, we have been able to find that writings about shared, cooperative, or distributed leadership concepts have increased, more policy initiatives which promote these concepts have been put in place, and training institutes implicitly integrated and explicitly offered programs which foster these approaches. Although this has an impact on training and development programs for school leadership, it is still too little, quantitatively and qualitatively speaking.

If change is on the agenda of schools and school leaders, it is crucial to have a vision which gives them a direction. Leaders (of any kind) need to know what goals and aims for real improvement are as well as what is needed is to have criteria to judge the overall leadership approach and the day-to-day decision making. This should be back-mapped against the core purpose of school, namely teaching and learning. As a solid base for what education aims at, in some of the programs an orientation toward a specific value-based attitude is intended. Thus, the understanding of leadership in this context includes moral and political dimensions.

A very final remark should hint at the phenomenon across countries that fewer teachers are interested in leadership functions. Training and development opportunities can and have to take that into account, in terms of fostering potentials. Besides being an individual training measure and a school development initiative, they can also be a measure for personnel marketing. It is about attracting potential leaders to apply for school leadership positions. This can be achieved more easily if training and development programs are less position oriented than competence oriented. Hence, there may be training and development programs, which are not school leader programs with a fixed set of competences in mind, but school and school leadership development programs, which try to enhance the development and leadership capacity of a school in particular and the school system in general. This, as well, fits to the now newly discussed concepts of system leadership.

Increasingly relevant to policy, practice, and research is the issue of the effectiveness and efficiency/efficacy of training and development programs. It is about the resources needed, the output and outcome, and the benefits not only for the participants, but also for the individual school organization and the school system in general. Whether school leadership development programs are successful is still not researched sufficiently. The efficacy and effectiveness of programs are still a research desiderate.

Bibliography

- Bridges, E. and Hallinger, P. (1995). *Implementing Problem-Based Learning in Leadership Development*. Eugene, OR: ERIC Clearinghouse on Educational Management.
- de Bevoise, W. (1984). Synthesis of research on the principal as instructional leader. *Educational Journal* **41**(5), 14–20.
- Fullan, M. (1993). *Change Forces, the School as a Learning Organisation*. London: Falmer.
- Fullan, M. (1995). Schools as learning organizations: Distant dreams. *Theory into Practice* **34**(4), 230–235.
- Giroux, H. A. (1988). *Teachers as Intellectuals: Toward a Critical Pedagogy of Learning*. Granby, MA: Bergin and Garvey.
- Gray, J. (1990). The quality of schooling: Frameworks for judgements. *British Journal of Educational Studies* **38**(3), 204–233.
- Gruber (2000). Erfahrung erwerben. In Harteis, C., Heid, H., and Kraft, S. (eds.) *Kompendium Weiterbildung*, pp 121–130. Opladen: Leske Budrich.
- Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education* **33**(3), S329–S351.
- Hallinger, P. and Murphy, J. (1985). Assessing the instructional management behaviour of principals. *Elementary School Journal* **86**(2), 217–247.
- Huber, S. G. (1997). *Headteachers' Views on Headship and Training: A Comparison with the NPQH*. Cambridge: School of Education, University of Cambridge.
- Huber, S. G. (ed.) (2004). Preparing school leaders for the 21st century. In Chrispeels, J., Creemers, B., Reynolds, D., and Stringfield, S. (eds.) (2004). *The Series Context of Learning*, London: RoutledgeFalmer.
- Huber, S. G. (2004). School leadership and leadership development – adjusting leadership theories and development programs to values and the core purpose of school. *Journal of Educational Administration* **42**(6), 669–684.
- Huber, S. G. and Moos, L. (2007). School leadership and school effectiveness and school improvement: Democratic and integrative leadership. In Townsend, T. (ed.) *International Handbook on School Effectiveness and Improvement*, pp 579–596. Dordrecht: Springer.
- Huber, S. G. and West, M. (2003). Developing school leaders – a critical review of current practices, approaches and issues, and some directions for the future. In Hallinger, P. and Leithwood, K. (eds.) *International Handbook of Educational Leadership and Administration*, pp 1071–1101. Dordrecht: Kluwer Academic Press.
- Imants, J. and de Jong, L. (1999). Master your school: The development of integral leadership. *Paper Presented at the International Congress for School Effectiveness and Improvement*. San Antonio, TX, USA, January.
- Jackson, D. and West, M. (1999). Learning through leading: Leading through learning. Leadership for sustained school improvement. *Paper Presented at the International Congress for School Effectiveness and Improvement 1999*. San Antonio, TX, USA, January.
- Joyce, B. and Showers, B. (1988/1995). *Student Achievement through Staff Development*. New York: Longman.
- Leithwood, K. A. (1992). The principal's role in teacher development. In Fullan, M. and Hargreaves, A. (eds.) *Teacher Development and Educational Change*, pp 86–103. London: Falmer.
- Murphy, J. (1992). *The Landscape of Leadership Preparation: Reframing the Education of School Administrators*. Newbury Park, CA: Corwin.
- Rosenbusch, H. S. (2005). *Organisationspädagogik: Grundlage pädagogischen Führungshandelns*. München: Luchterhand in Wolters Kluwer.
- Schön, D. (1983). *The Reflective Practitioner*. New York: Basic Books.
- Schön, D. (1984). Leadership as reflection-in-action. In Sergiovanni, T. and Corbally, J. (eds.) *Leadership and Organizational Culture*, pp 36–63. Chicago, IL: University of Illinois Press.
- Senge, P. (1990). *The Fifth Discipline*. New York: Doubleday.
- Siebert, H. (1996). *Didaktisches Handeln in der Erwachsenenbildung: Didaktik aus konstruktivistischer Sicht*. Neuwied: Luchterhand.
- West, M., Jackson, D., Harris, A., and Hopkins, D. (2000). Learning through leadership, leadership through learning. In Riley, K. A. and Seashore-Louis, D. (eds.) *Leadership for Change and School Reform*, pp 30–49. London: Routledge Falmer.

Further Reading

- Bridges, E. and Hallinger, P. (1997). Using problem-based learning to prepare educational leaders. *Peabody Journal of Education* **72**(2), 131–146.
- Huber, S. G. (2003). School leadership development – current trends from a global perspective. In Hallinger, P. (ed.) *Reshaping the Landscape of Educational Leadership Development: A Global Perspective*, pp 273–288. London: Swets and Zeitlinger.
- Huber, S. G. (ed.) (2004). Preparing school leaders for the 21st century. In Chrispeels, J., Creemers, B., Reynolds, D., and Stringfield, S. (series eds.) *The Series Context of Learning*, London: RoutledgeFalmer.
- Rosenbusch, H. S. (1997). Die Qualifikation pädagogischen Führungspersonals. In Glumpler, E. and Rosenbusch, H. S. (eds.) *Perspektiven der universitären Lehrerbildung*, pp 147–165. Bad Heilbrunn: Klinkhardt.

Research for Leaders: The National College for School Leadership

G Southworth, National College for School Leadership, Nottingham, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The National College for School Leadership (NCSL) was established in 2000 to be a single national focus for school leadership, be the driving force for world-class leadership in schools, and to coordinate national development programs and other leadership initiatives and activities in England. It is almost wholly funded by the central government.

It has been a founding principle that NCSL would be a learning organization. Therefore, it has always strived to be an evidence-informed institution. The research group plays a key and strategic role in knowledge management within NCSL, aiming to provide access to the latest thinking and research into school leadership for school leaders, policy makers, and staff in the institution, who manage and develop leadership activities. In terms of knowledge management, the role of research and evaluation is to create, collate, and communicate what is known about effective practice. Moreover, as a national organization, we constantly scan research findings within education and beyond (e.g., business schools and the public sector) in the UK and internationally.

It is important to note that NCSL is committed to actionable knowledge. We regard our activities as delivering leadership learning for action, and our research should realize this aim. Hallinger and Heck (1996) demonstrated that there were some persistent blind spots and blank spots in educational leadership research and advised ways of addressing these. NCSL took their advice seriously, particularly in respect of the need to study how leaders make a difference. Thus, while NCSL's research covers a range of leadership topics and issues, one issue which is pursued in depth is how school leaders make a positive difference to the progress and achievements of pupils.

In a relatively short time, NCSL has established a high profile internationally and, while we are never complacent about the scale of the task, it is also fair to say that we have received some very positive feedback on our work. For example, a critical review of leadership development in the USA, published in 2005, which had examined what the NCSL did as an external comparison to work within the USA, concluded that NCSL "blends research and practice better than any American school of education" (Levine, 2005: 59).

This article will draw on NCSL's research evidence, which includes knowledge of others' findings, to provide a high-level overview of what we know about successful school leadership.

What We Know about School Leadership

Between 2001 and 2006, NCSL hosted and sponsored a number of think tanks, seminars, visiting scholars, and practitioner enquiries. We commissioned a high number of research studies, evaluations, and literature reviews. We also scanned research publications and outputs nationally and internationally and attended and presented at major research conferences.

Without doubt, there is a wealth of work now available to practitioners, policy makers, and the research community. Moreover, it is important to bring this work together to build coherence, note differences and inconsistencies, and summarize the key insights. Furthermore, having been in existence for 5 years, in 2006, researchers at NCSL took the decision to synthesize and summarize their work and that of others. That exercise forms the basis of what now follows.

The knowledge base is characterized by the following features:

- There exist a high number of small-scale research studies.
- The evidence tends to be either qualitative or quantitative; there are only a limited number of projects which use mixed methods.
- Often, there is a conflation between description and prescription – with advocacy sometimes a strong feature of the scholarship and empiricism placed in a distant second place.
- There is not always enough building on preexisting knowledge – and, perhaps, there is insufficient integration and too much atomized thinking.
- Generally, there is weak attention paid to examining the impact of leaders. There are some exceptions too, notably Silins and Mulford (2000), who, as part of their study, attempted a critical path analysis and showed the value of trying to trace linkages.
- North American research sometimes dominates; yet, at the same time, we are increasingly aware that context-sensitive studies are needed. Cross-cultural borrowing needs to be treated with caution.

In recent years, there have been a number of attempts to summarize the knowledge base, for example, Leithwood and Reihl (2003), Leithwood *et al.* (2006b), and Ingvarson *et al.* (2006).

From our own work, which includes a number of practitioner studies as well as academic ones, and our

awareness of other project findings, it is possible to identify seven sets of insights about which there is growing agreement. These are described in the following.

Context Matters

There can be little doubt that contingency theories are very popular both with the academic community and with practitioners. The former regularly make much of the need for leaders to be alert to the demands of their specific contexts. For example, Leithwood *et al.* (1999) declare that outstanding leadership is “exquisitely sensitive to the contexts in which it is exercised” (p. 4), while Hallinger and Heck (1996) state that it is “virtually meaningless to study principal leadership without reference to the school context” (p. 14). Practitioners too subscribe to the importance of context by repeatedly describing the communities that their schools serve and explaining their actions as leaders in terms of the needs of the students, their home backgrounds, and the social and economic circumstances of their communities.

Therefore leaders need to be contextually literate: they have to be able to read their contexts like a text, including understanding the subtexts, the meta-messages, and the micropolitics, while not becoming victims of them. Leaders are not passive players in their contexts – indeed, they are influential actors and should be proactive in shaping their organizational settings, ways of working, and cultures (Schein, 1985). It should also be recognized that becoming contextually literate is a major, indeed core, skill and one which has important implications for leadership development. It implies that leaders should be able to analyze and understand their settings, determine priorities, and enact their own and others’ leadership in ways which are needs based. This means using a mix of strengths, weaknesses, opportunities, and threats (SWOT) analyses, school self-evaluation methods, and student outcome data to recognize trends and the brute facts of the school’s levels of performance, progress, and rates of improvement.

In emphasizing the context-specific nature of leadership, this is not to impute particularity. There are some core skills and strategies which leaders need to know about and adopt. According to Leithwood *et al.* (2006b), almost all successful school leaders draw on the same repertoire of basic leadership practices (see section titled ‘Core tasks for leaders’):

- setting directions (motivation);
- developing people (capacity);
- managing the instructional program (stability); and
- redesigning the organization (setting).

Furthermore, it is the enactment of these basic leadership practices – not the practices themselves – that is responsive to the context. In other words, how leaders

read and then act, given their intimate knowledge of their contexts, is the distinguishing issue between them. Expressed more simply, it is not so much what you do as a leader, as how you do it that makes the difference in any given environment.

Leaders of Learning

Schools are places where children and young people are expected to learn whatever a community or society decides should be passed on to them. In turn, this means that school leaders are responsible for the learning which takes place there and, essentially, school leaders are leaders of learning.

The intention of this first paragraph is not to be tautological, but to make the core purpose of school leadership plain. In addition, it needs to be spelt out because there has always been a tendency for some leaders to become leaders of the organization as against exercising leadership of what the school is for. Moreover, the larger the school, the more organizational demands there are, and the more remote school leaders can be from the core business.

None of the above is to imply that headteachers and principals do this single handedly, nor that these are their only responsibilities. Nevertheless, research at NCSL (2004b) has shown that high-performing and rapidly improving schools are characterized by learning-centered leaders. Such leadership is focused on influencing what happens inside classrooms and is geared to improving the quality of teaching and the pupils’ learning. In addition, such a strong focus and interest in students’ learning is the basis for these leaders being role models for teacher and staff colleagues.

Learning-centered leadership is relentlessly focused not only on learning – most of all, student learning processes and outcomes – but also on staff learning and development – especially pedagogic development. The leaders monitor what is going on inside classrooms and across the whole school by using data, observing teaching and learning and identifying strengths and the development needs of teachers, and determine priorities for groups of students and units of the school.

Such leadership is shared, with middle leaders increasingly playing important roles. Headteachers and senior leaders create and sustain leadership at all levels because although distributing leadership matters, what matters most is distributing and developing learning-centered leadership. Therefore, the development of leadership is an important feature of school leadership today – it always has been – however, as headship has intensified and become more complex, so also has the need to share leadership increased, bringing with it the responsibility to develop other leaders.

Part of such school-based leadership development rests on the school being an environment characterized

by relational trust and professional dialog, and leading by example and management arrangements which enable feedback, coaching, and teamwork. Furthermore, collaborative research with NCSL's leadership network shows that when these conditions and behaviors are present and effective, within-school variations in pupil and departmental performances are reduced (NCSL, 2006c).

Having conducted a number of studies examining learning-centered leadership in action, NCSL (2004b) researchers concluded that learning-centered leaders:

- lead by example;
- monitor the achievements and progress of pupils, classroom practices, and the quality of teaching;
- use data to analyze and evaluate performance;
- generate discussion about teaching and learning;
- take steps to sustain school improvement; and
- create school structures, systems, and processes to support learning.

Core Tasks for Leaders

As implied earlier, recent work, while concluding that there is no one way to lead because it is so strongly situational, has identified a core set of leadership practices which form the basics of successful leadership (Leithwood and Reihl, 2003; Leithwood *et al.*, 2006a). The basics are:

1. Setting directions: identify and articulate a vision, create shared meaning and high-performance expectations, foster acceptance of group goals, monitor organizational performance, and communicate.
2. Developing people: offer intellectual stimulation, provide individual support, and an appropriate model – lead by example.
3. Developing the organization: strengthen school culture and modify organizational structures; build collaborative processes and manage the environment.
4. Respond productively to the challenges and opportunities created by accountability.
5. Respond productively to the challenges and opportunities of educating diverse groups of students.

These five practices also imply that leadership rests on certain qualities too, notably optimism and a positive disposition, a developmental orientation (e.g., that people and organizations can and do improve), and a strong moral compass – schools exist to serve children and students and to enable them to grow and improve as learners and social actors.

These qualities are vital because leadership is powerfully motivational (and, by definition, demotivational when it is less effective). Inspiring others and securing their followership is part and parcel of leadership. Yet, while there has been considerable attention paid to moral purpose and the importance of vision, little research has

actually been conducted into what teachers find motivating, although Blasé and Blasé (1998) have explored followers' perceptions of leaders and what teachers' preferences from their principals or headteachers are.

Values also play a part in leadership; indeed, they are inextricably tied up with leadership. There is a considerable body of work to support this claim (e.g., Fullan, 2003). Work by some of NCSL's research associates shows that what sustains school leaders in difficult times is their core values and moral purpose (Flintham, 2003).

It should also be noted that while there is growing consensus on some of the technical competences, there is also a broadening of enquiry in the leadership field to include studies on the passions and emotions of leadership (Sugue, 2005; Day, 2004) as well as the moral and spiritual dimensions of leadership (West-Burnham, 2002; Starrat, 2004). However, there are only limited studies regarding leadership for social justice, that is, research on those leaders who frame their leadership as a quest for equity (Jansen, 2006: 37, 38).

Distributing Leadership Matters

Distributed leadership has already been mentioned as part of the section titled 'Leaders of learning'. For some researchers, distributed leadership is, without doubt, one of the current big ideas which has swirled around during the last 5 years (see Spillane, 2005). There is considerable advocacy for the concept from a number of different perspectives. Some argue for distributing leadership on grounds of social equity and the democratization of schools as organizations; others extend the concept to include students and view it as an advancement of pupil voice and inclusion. Undoubtedly, further work is needed to untangle some of the reasons why distributed leadership matters.

NCSL research shows that distributed leadership has a part to play in school improvement; it does make a difference to school and student performance (NCSL 2004a, 2004b, 2006c). Moreover, NCSL's work demonstrates that distributing leadership is vital if schools are to be places where pools of leadership talent are created and from which tomorrow's school leaders can be drawn.

It should also be acknowledged that distributed leadership supports system leadership: they are two sides of the same coin – you cannot have headteachers and other leaders working beyond their schools unless and until other leaders are willing and able to step up and take on new roles and responsibilities; each supports the other.

Leadership Succession

Across many developed nations, there is a demographic time bomb ticking away. Urgent attention is needed to defuse this challenge, and NCSL's staff have been examining what the

challenge is and how it might be addressed in England (NCSL, 2006a). Essentially, the challenge is demographic – larger-than-average numbers of school leaders are going to retire between 2008 and 2012 and there are lower-than-average numbers of teachers and leaders following them – too few to replace those who will be retiring. Moreover, in England, the time it takes to become a head is now too long for the next bulge in the teacher population to be ready to take up the vacancies anticipated in the coming years.

All of this means two things in England – and in other countries too. First, we are going to need many more school leaders than in recent times. Second, we need more leaders than those that current approaches to promoting staff are presently able to produce.

Furthermore, the demographic challenge is compounded by negative perceptions of the work and role of school leaders – especially regarding accountabilities and workload. Many of those most likely to step up to headship do not like the look of leadership. However, such perceptions tend to be held by those with little or no experience of running a school for a period. Those who have had a spell of running a school report that the experience does two things: shows them that there are significant rewards and job satisfiers which are underreported by some serving school heads; and that having had a go at leadership, the great majority have learned that they can do it and this markedly improves their self-confidence and increases the likelihood of applying for leadership positions.

Therefore, NCSL has advised ministers that, to address the demographic challenge, there needs to be: increased fast-tracking of those with leadership potential, which means early identification of talent, and mentoring and coaching these individuals; providing them with many more opportunities to lead – in their own and in other schools – to broaden their knowledge of school types and contexts, and increase the number of head-teacher role models.

Our advice to ministers recommended that this national challenge is dealt with by local solutions developed by groups of schools taking responsibility for developing their talent pools. Local initiatives will need some additional and external support and, nationally, there needs to be a campaign to talk up headship since we know that the overwhelming majority of headteachers are very positive about their work. By providing more opportunities for middle leaders and deputies to experience periods when they are acting heads, we believe that perceptions of the role can be made more positive and self-confidence to perform the job increased. However, this is an English, context-specific strategy. It has been constructed in the light of England's devolved school system and may not be appropriate for other school systems.

At the same time, we should also appreciate that succession planning is not simply a quantitative issue. It is

vital to ensure there is a supply and flow of high-quality candidates for headship and leadership teams. Attending to quality also means tracking over time that we have the right mix of leaders (e.g., gender; ethnic minorities) and that the recruitment and appointment of headteachers improve (NCSL, 2006b).

School Leadership Is Hard Work

Over the last two decades, research has suggested that the work of leaders has intensified and become more complex, relentless, and accountable. It should therefore be no surprise that leaders become tired, if not exhausted, when working in some of the most challenging environments. Without doubt, concerns about work–life balance and well-being are legitimate. School leaders need to look after their own well-being and spend time on the balcony as well as on the dance floor (Heifetz and Linsky, 2002).

While their work may feel ever more urgent and demanding, their learning, development, and support needs have equally increased and must be met. In part, this means remodeling not only the school's workforce, but also the support mechanisms on which the heads and the leadership team can call upon – in particular secretarial and personal assistant support, and school business managers, as well as access to good-quality human resources advice. Unless these are rethought and redesigned, headship and principalship may continue to look too intensive, demanding, and too accountable.

It is also the case that, in addition to improved support for leaders within their schools, more attention should be paid to their coping strategies (Cooper and Kelly, 1993; Southworth, 1995) to deal with the strains and pressures of their work and workloads. Indeed, there are rewards, and some heads derive a sense of achievement from being able to deal with whatever comes their way, as well as the success of the school, staff, and students. However, it is also the case that a more mature approach to sustaining leaders is needed today; indeed, sustainable leadership might be one of the next big leadership issues (Hargreaves and Fink, 2006).

The work Is Polyphonic

One weakness of many a leadership research is that it takes but one slice of leadership activity and fails to register that leaders engage in several things, often at once. It is for this reason that the metaphors of practitioners for their work tend to be about juggling and plate spinning.

In an early attempt to synthesize research findings and move beyond atomistic analyses, NCSL researchers emphasized two conclusions:

1. One thing leads to another.
2. Leadership is both simple and complex.

With regard to conclusion 1, NCSL researchers noted that a single input by a leader can create multiple outcomes, and that effective leaders use time twice or thrice. For example, one episode of coaching between a more experienced leader and a novice could contribute to strengthening distributed and learning-centered leadership, growing tomorrow's leaders, and creating a culture of learning for students and staff alike.

As for conclusion 2, this insight is based on evidence which showed that effective leaders do not have a repertoire of skills unknown to other leaders. Rather, what distinguishes the highly effective leaders is their ability to deploy seemingly ordinary strategies and tactics and execute them to an extraordinary degree. In addition, they use tactics in combinations and with care and diligence, knowing that one thing leads to another, and therefore concentrate on applying them consistently so that they mutually reinforce one another.

This suggests that rather than tackling each task in a serialized way, or in some checklist manner, what it takes is the ability of leaders to make links between their tasks and tactics, develop their awareness of how one approach links to others, and their capacity to create synergies between all they have to do. Such polyphonic working may lie at the heart of what makes them effective and their work manageable. In turn, this idea suggests that leaders need to be able to see the whole – as well as the individual elements – and to be able to appreciate that the sum total of leadership can be greater than the parts (NCSL, 2005: 43).

Conclusions

These seven sets of findings suggest that the texture of school leadership is changing; therefore, it is possible to say that school leadership today is:

- more data and evidence based than ever before;
- less lonely and more collaborative;
- involving the building of leadership capacity;
- leading by example; and
- intense and varied.

These changes have, in some cases, pushed at the parameters of school leadership, certainly in England, so that relatively new models of headship emerge, such as system leaders, executives who run more than one school, federations of schools, community leadership, multi-agency leaders, and co-headships.

It is also true that leadership in and of networks has taken root in the last 5 years. Networks can be fertile environments for developing leaders since they create new ways of engaging with others and accessing different school contexts, and enable leaders in one school to contribute to the development of many others.

If this points to leaders working in more than one context, it does not mean that leadership is confined to geographical location. In today's world, context includes awareness of policy directions and system needs. In school systems which are highly devolved, as in England, this places a responsibility on the shoulders of leaders to know and understand what a central policy means for them and their schools since the implications of policy have to be understood quickly and acted upon. Hence, strategic leadership has become more important in the last decade. Leaders need to be able to look ahead, especially in a fast-changing world for which we are preparing pupils and future leaders.

In turn, this implies that the development of leaders and leadership is becoming increasingly pivotal to the performance and vitality of their schools and the communities that they serve.

Taken together, these findings show three things. First, the theme of leaders needing to be able diagnosticians and proactive about their contexts should be developed – otherwise they themselves are not leading, but following existing patterns. Second, the importance of context reinforces what we have known for a long time about leaders as cultural designers, builders, maintainers, and (sometimes) destroyers (Schein, 1985) as well as culture mediating leadership and leadership practices needing to change as the stage of a school's development unfolds (Bryk and Schneider, 2002).

Third, if there was one thing above all others that this article throws into sharp relief it is that building leadership capacity and developing leadership are critical to school and system success. At a time when national and international belief in leadership, in all sectors of work and life, has never been higher, it follows that the development of leadership is important. Furthermore, the challenges that the school system is now facing such as the need to grow tomorrow's leaders in greater numbers and ensure that these individuals are able to lead new forms of organizations such as: (1) multidisciplinary schools or campuses; (2) those that provide strong leadership of learning; (3) those that take executive responsibility for federations of schools; or (4) a multisite organization, are effectively developmental challenges.

None of this should mask the fact that the insights here show that school leadership today is more nuanced and diverse than it was formerly. Unquestionably, we not only need strong leaders, that is, resilient and determined individuals and teams, but also need more emotionally intelligent ones too. Indeed, the picture outlined here is of leaders and leadership becoming increasingly intelligent. Intelligent leadership rests on learning to lead at many sites, in a range of settings, and in differing ways.

Therefore, we need to look much more closely at how leaders develop most productively and powerfully, under what conditions, over what time scales, and with

what outcomes; in addition, we need to find out which leadership strategies yield the greatest outcomes and benefits.

Bibliography

- Blasé, J. and Blasé, J. (1998). *Handbook of Instructional Leadership: How Really Good Principals Promote Teaching and Learning*. Thousand Oaks, CA: Corwin Press.
- Bryk, A. and Schneider, B. (2002). *Trust in Schools*. New York: Russell Sage Foundation.
- Cooper, C. and Kelly, M. (1993). Occupational stress in headteachers: A national UK study. *British Journal of Educational Psychology* **63**, 130–143.
- Day, C. (2004). The passion of successful leadership. *School Leadership and Management* **24**(4), 425–437.
- Flintham, A. J. (2003). *Reservoirs of Hope: Spiritual and Moral Leadership in Headteachers*. Nottingham: National College for School Leadership.
- Fullan, M. (2003). *The Moral Imperative of School Leadership*. Thousand Oaks, CA: Corwin Press.
- Hallinger, P. and Heck, R. (1996). 'Reassessing the principal's role in school effectiveness: A review of empirical research 1980–1995'. *Educational Administration Quarterly* **32**(1), 5–44.
- Hargreaves, A. and Fink, D. (2006). *Sustainable Leadership*. San Francisco, CA: Jossey-Bass.
- Heifetz, R. and Linsky, M. (2002). *Leadership on the Line: Staying Alive through the Dangers of Leading*. Boston, MA: Harvard Business School Press.
- Ingvanson, L., Anderson, M., and Gronn, P. (2006). *Standards for School Leadership: A Critical Review of the Literature*. Melbourne: VIC: Australian Institute for Teaching and School Leadership and Australian Council for Educational Research.
- Jansen, J. D. (2006). Leading against the grain: The politics of emotions of leading for social justice in South Africa. *Leadership and Policy in Schools* **5**(1), 37–51.
- Leithwood, K., Jantzi, D., and Steinbach, R. (1999). *Changing Leadership for Changing Times*. Buckingham: Open University Press.
- Leithwood, K., Day, C., Sammons, P., Harris, A., and Hopkins, D. (2006a). *Seven Strong Claims about Successful School Leadership*. Nottingham: NCSL/DfES.
- Leithwood, K., Day, C., Sammons, P., Harris, A., and Hopkins, D. (2006b). *Successful School Leadership: What It Is and How It Influences Pupil Learning*. Nottingham: NCSL/Department for Education & Skills, University of Nottingham.
- Leithwood, R. and Reihl, C. (2003). *What We Know about Successful School Leadership*. Philadelphia, PA: Laboratory for Student Success, Temple University/NCSL.
- Levine, A. (2005). *Educating School Leaders*. New York: Teachers' College, Columbia University.
- NCSL (2004a). *Distributed Leadership*. Nottingham: NCSL. <http://www.ncsl.org.uk/publications> (accessed April 2009).
- NCSL (2004b). *Learning-Centred Leadership*. Nottingham: NCSL. <http://www.ncsl.org.uk/publications> (accessed April 2009).
- NCSL (2005). *Annual Review of Research 2004–05*. Nottingham: NCSL. <http://www.ncsl.org.uk/publications> (accessed April 2009).
- NCSL (2006a). *Leadership Succession: An Overview – Securing the Next Generation of School Leaders*. Nottingham: NCSL. <http://www.ncsl.org.uk/publications> (accessed April 2009).
- NCSL (2006c). *Narrowing the Gap: Reducing within School Variation in Pupil Outcomes*. Nottingham: NCSL. <http://www.ncsl.org.uk/publications> (accessed April 2009).
- NCSL (2006b). *Recruiting Headteachers and Senior Leaders: Seven Steps to Success*. Nottingham: NCSL. <http://www.ncsl.org.uk/recruitingleaders> (accessed April 2009).
- Silins, H., Mulford, W., Zarins, S., and Bishop, P. (2000). 'Leadership for organisational learning in Australian secondary schools'. In Leithwood, K. (ed.) *Understanding Schools as Intelligent Systems*, pp 267–291. Stamford, CT: JAI Press.
- Schein, E. (1985). *Organizational Culture and Leadership*. San Francisco, CA: Jossey-Bass.
- Southworth, G. (1995). *Talking Heads: Voices of Experience – An Investigation into Primary Headship in the 1990s*. Cambridge: University of Cambridge.
- Spillane, J. (2005). *Distributed Leadership*. San Francisco, CA: Jossey-Bass.
- Sugue, C. (ed.) (2005). *Passionate Principals: Learning from the Life Histories of School Leaders*. London: RoutledgeFalmer.
- West-Burnham, J. (2002). *Leadership and Spirituality*. NCSL Leading Edge Seminar Thinkpiece. <http://www.ncsl.org.uk/leadingedge> (accessed April 2009).

Relevant Website

<http://www.ncsl.org.uk> – National College for School Leadership.

School Leadership Standards

L Ingvarson, Australian Council for Educational Research, Camberwell, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This article is based on a review of international developments in leadership standards. The review focused on five countries where leadership standards were being used for purposes such as professional learning and recognition of achievement through some form of certification or accreditation.

The review indicated an increasing number of initiatives in this area over recent years and closer links between leadership standards and research. After summarizing purposes and international developments for leadership standards, the article illustrates the steps involved in developing valid leadership standards for professional learning and certification purposes.

Purposes for School Leader Standards

The context of school leaders' work has become increasingly complex. Expectations of school leaders as agents for school improvement, and even transformation, have increased, as have demands for accountability. As a result, greater attention is being given to effective methods and programs for preparing school leaders and supporting their continuing professional learning.

Standards have an important role to play in clarifying what the central goals of such programs should be. Well-written and valid standards synthesize the implications of research for what school leaders need to know and be able to do to be effective.

Recruitment and retention of effective school leaders are also concerns common to many countries. Greater attention is therefore being given to career paths and conditions of work that attract and retain effective teachers and school leaders. These include methods for identifying and rewarding teachers and school leaders who attain high performance standards.

Standards have a central role to play here also. A complete set of standards not only describes what school leaders should know and be able to do, but it also describes how performance against the standards will be assessed and what levels of performance meet the standards. In other words, well-written standards are assessable as well as valid; they point to observable features of what school leaders know and do.

In summary, the main purposes for school leadership standards are to increase the effectiveness of professional

preparation and development for school leaders, and to provide better systems for recognizing and rewarding effective school leaders.

International Developments in Standards for School Leaders

The growing interest in leadership standards since the early 1990s can be linked to standards-based approaches to educational reform (Smith and O'Day, 1990; Sykes and Plastrik, 1993). The new curriculum standards for students had obvious implications for the kinds of knowledge and skills needed by their teachers. This in turn created pressure for greater clarification about the capacities school leaders needed if they were to enable teachers and students to meet the goals of educational reform.

The National Council of Teachers of Mathematics (NCTM), for example, produced a set of challenging curriculum standards for mathematics in 1989, and followed this with its Professional Standards for Teaching Mathematics (NCTM, 1991). The first major effort to develop leadership standards, initiated by the Council of Chief State School Officers in the USA, led to the Interstate School Leaders Licensure Consortium (ISLLC) standards in the early 1990s (Murphy and Shipman, 1999). The production of standards has proliferated rapidly since that time. Gronn (2003) contrasts the role of standards in the preparation of headteachers in England and school principals in the USA. Huber (2004) compares approaches to preparing school principals in 15 countries.

Ingvarson *et al.* (2006) provide a critical review of recent international developments in leadership standards. They focused on five countries where leadership standards were being used as a part of a system that linked the standards to professional learning and to some form of recognition or certification that the standards had been attained. These countries, and the responsible agencies, included:

- England: the National Standards for Headteachers (National College for School Leadership (NCSL) and Department for Education and Skills).
- USA: Standards for School Leaders (Council of Chief State School Officers ISLLC as applied by the Connecticut State Board of Education).
- Australia: Performance Standards for School Leaders developed by the Department of Education in Western Australia.

- The Netherlands: Professional Standard for Educational Leaders in Primary Education (Dutch Principal Academy, or Nederlandse Schoolleiders Academie).
- Scotland: The Standard for Headship (Scottish Executive).

The research literature on school leadership provides few clear-cut guidelines for agencies charged with the development of standards. Consequently, writers of standards for school leaders face a considerable challenge in developing profession-wide standards. However, although there was some variation in details across the five countries, there was a remarkable degree of similarity in the sets of leadership standards emerging internationally. Standards did not vary markedly according to what might be thought of as very different national and cultural contexts, although it is important to note most in this review were from English-speaking countries.

Although many sets of standards for teachers and school leaders were being developed internationally, most were specific to particular jurisdictions or employing authorities. In other words, few standards were profession-wide. Apart from the ISLLC standards and those of the Dutch Principal Academy, few were developed by a broad range of stakeholders and professional organizations. Unlike most professions, governments and employing authorities usually played the major role in developing standards for teachers and school leaders, not professional associations of school leaders. In this sense, teaching is almost unique among the professions.

Standards were being used increasingly in these five countries to guide the preparation and development of school leaders. In a few instances, they were also being used to assess performance for aspiring and experienced school leaders for certification purposes.

It was noteworthy that recent versions of school leadership standards resist the temptation to scope out the full practice of leadership and management in schools. They focus first on quality student learning, and move outward to identify implications for what school leaders should know and be able to do. In other words, leadership standards are beginning to look more like professional standards rather than lists of micro-competencies or job descriptions as seen in earlier sets of standards. The latter rarely had a clear guiding conception of school leadership, or how the work of school leaders links to quality learning opportunities for students.

The main organizers in recent sets of leadership standards are more parsimonious and interesting, as researchers and school leaders refine and reorganize their conceptions of what effective school leaders know and do. This effort has been greatly assisted by researchers as they synthesize those aspects of school leaders' work that establish the conditions for effective teaching and learning. Leithwood *et al.* (2004), for example, identify three main categories of effective leadership practice:

- developing a deep understanding of how to support teachers;
- managing the curriculum in ways that promote student learning; and
- developing the ability to transform schools into more effective organizations that foster powerful teaching and learning for all students.

Syntheses of evidence-based research such as these are valuable to standards developers. However, it is one thing to write standards and quite another to ensure the standards become embedded in everyday thought and practice. To have an impact, school leadership standards need to be linked to modes of professional learning that enable school leaders to implement effective leadership practices.

The review identified two main approaches toward linking standards to professional learning. The traditional approach is to develop a course, or even a set of courses for school leaders. In contrast, emerging approaches, such as those used by the National Board for Professional Teaching Standards (NBPTS) in the USA, focus on developing a rigorous system for providing certification to teachers and school leaders who demonstrate they can meet the performance standards, and on developing recognition of that certification.

Recent criticisms of the quality of traditional course-based programs for preparing school leaders, particularly in the US (e.g., Levine, 2005), highlight the need for alternative routes and opportunities for professional learning in school leadership. An accumulation of academic credits and courses is no guarantee of capability or achievement in the workplace. Professional associations of school leaders are increasingly providers of a wider range of alternative professional learning activities. Particularly important are the activities, networks, and other forms of support that associations provided locally to support candidates preparing for national professional certification.

An emerging approach to professional learning is to develop a system for providing certification based on evidence of meeting the standards – a standards-based professional learning system.

In this approach, an independent standards agency develops a rigorous assessment and professional certification process based on evidence of performance. Employing authorities who come to regard the certification as credible may choose to provide support and incentives for teachers and school leaders to undertake the certification process. The process of applying for professional certification places teachers and school leaders in a more proactive role in relation to planning and providing their own professional learning, such as seeking courses that will help them directly to prepare for certification. The processes of preparing evidence of leadership initiatives, such as building professional community or leading and managing change projects, necessarily engage teacher leaders

in activities highly consistent with effective modes of professional learning.

Evidence is accumulating that the certification system established by the NBPTS, for example, provides an effective approach toward linking standards to professional learning (Ingvarson and Hattie, 2008). Most teachers who complete the NBPTS certification process, for example, say it was the most significant professional learning they had ever experienced.

Finally, the review of international developments examined approaches used by each of the five national systems to judge whether school leaders had met the standards. Needless to say, it is vital that such a judgment process is rigorous and fair. They found that the validity of the certification process in most systems remains unknown or uncertain, as it is usually based on course completion, not evidence of performance.

Writing Valid Standards for School Leadership

The purpose of this section is to describe the steps that would be involved in the development of a valid set of professional standards for school leaders.

Definition of Standard

The *Shorter Oxford Dictionary* gives two definitions of the word standard:

- distinctive flag (often figure of principle to which allegiance is given or asked; the royal and c-raise the – of revolt; free trade, and c) and
- specimen or specification by which the qualities required of something may be tested, required degree of some quality, levels reached by average specimens (attrib.) serving as test, corresponding to the – of recognized authority or prevalence.

Both definitions can be applied to the development of standards for school leaders. In the first, standards would aim to articulate professional principles and values. Similar to the flag on ancient battlefields, they would provide a rallying point and, perhaps, inspiration.

Standards are also measures, as indicated by the second definition – “the required degree of some quality,” for example. Standards are tools that are used often in making judgments and making improvements in many areas of life and work, whether measuring length, evaluating writing, or critiquing restaurants. Standards provide the context of shared meanings and values that is necessary for fair, reliable, and useful judgment.

Standards as principles

In the first sense, writers of leadership standards would aim to arrive at a consensus on the principles and values

that drive practice and pervade professional relationships. A straightforward example might be: Highly accomplished school principals are committed to their students and their learning.

In the process of developing standards, standards writers endeavor to identify the distinctive features and aspirations of their profession – in this case, the unique things that effective school principals know and do. The process of writing standards for school leadership, understood in this sense, unites people around shared ideals and values, and encourages the reconciliation of divergent approaches to practice. Standards are statements about those features of leadership that are most valued in the profession.

Standards for school leaders, similar to those for classroom teachers, ultimately rest on professional norms and values relating to interpersonal relationships and the kinds of learning that are valued in a society. Education is ultimately and inescapably a moral enterprise. Standards developers need to articulate a vision of quality learning that will guide their more detailed work of describing what teachers and school leaders should know, believe, and be able to do to provide opportunities for that kind of learning.

Among the five systems reviewed by Ingvarson *et al.* (2006), the ISLLC Standards for School Leaders document was the most explicit about the principles that the writers used to guide the development of their standards for school leaders. Members of the consortium decided that standards for school leaders should:

- reflect the centrality of student learning;
- acknowledge the changing role of the school leader;
- recognize the collaborative nature of school leadership;
- be high, upgrading the quality of the profession;
- inform performance-based systems of assessment and evaluation for school leaders;
- be integrated and coherent; and
- be predicated on the concepts of access, opportunity, and empowerment for all members of the school community (Council of Chief State School Officers, 1996).

Reaching a consensus about principles is a necessary part of standards development, but it is a consensus that needs to be justified in terms of research and the wisdom of expert practitioners. Additionally, standards developers must also reach agreement on the scope of school leaders' work – they must set boundaries as it were – if that work is to be realistic and feasible for average mortals.

Standards as measures

To be useful for purposes such as professional learning and recognition, standards must also be understood in the second sense of the dictionary definition, as a valid basis for making assessments of performance and providing useful feedback. Standards should aid discernment and

discrimination. To value something, it helps to know how to evaluate its main attributes. A standard is incomplete if it cannot be used to assess practice, whether that is self-assessment or assessment by peers. If it cannot be used to assess practice, it will be of little use for professional learning or recognition.

Sykes and Plastrik (1993) point out that the word standard carries different usages and nuances. One of these is the idea of a standard as a legally recognized unit, such as that of Greenwich Mean Time, or the Gold Standard, or the Standard Metre in Paris for measuring length. Another is the notion of a standard as “an authoritative or recognised exemplar of perfection,” such as the sacred books of a religious organization. Yet another usage refers to “a definite level of excellence, attainment, wealth or the like” such as standard of living or a particular level of proficiency.

From their analysis, Sykes and Plastrik (1993) provide this useful definition of a standard: “A standard is a tool for rendering appropriately precise the making of judgments and decisions in a context of shared meanings and values” (p. 4).

Developing Standards

Sykes and Plastrik’s definition is a useful reminder that standards are pragmatic tools to achieve certain defined purposes. When standards are used in assessing performance, for purposes such as professional learning, recognition, and certification, there are three essential steps in their development:

- Defining what is to be assessed (i.e., what is school leadership? What are the essential elements of good leadership); these are often called content standards.
- Deciding how it will be assessed; that is, how valid evidence about (leadership) practice will be gathered.
- Assessing the evidence and identifying what counts as meeting the standard (i.e., how good is good enough?). This leads to performance standards, which specify the level of performance that meets the standards.

A complete definition of standards needs all three components above. A full set of standards also needs to describe how evidence about capability and performance will be gathered, and how decisions will be made about whether the standards have been met.

While content standards define the scope of (a school leader’s) work, performance standards are needed to tell us how good a (school leader’s) performance needs to be to meet the standard. The international review found few examples of leadership standards that are complete in this sense. Consequently, few were used extensively for professional learning. Fewer still could (or should) be used in making selection or certification decisions in ways that were valid, reliable, or fair.

Guiding Conception of Leadership

Every set of content standards for school leaders needs a guiding conception of what leadership is. It is not sufficient for a set of standards to spell out or map the territory of school leadership, like a job description or a list of responsibilities, yet this is what many sets of leadership standards tended to be in the past. It was still possible to read some present sets and be none the wiser about the meaning of leadership that underpinned them.

Fullan provides an example of a guiding conception of leadership that might be considered by standards developers:

The litmus test of all leadership is whether it mobilises people’s commitment to putting their energy into actions designed to improve things. It is individual commitment, but above all it is collective mobilisation (Fullan, 2001).

The work of many other researchers could be drawn upon. However, Fullan’s conception is useful because it points to school leadership practices that would be observable. To illustrate, Fullan elaborates his concept of leadership into five components, including:

- having a clear moral purpose;
- relationship building;
- understanding and managing change;
- knowledge creation and sharing; and
- ensuring coherence and alignment of structures.

These five components provide standards writers with a basis for elaborating and operationalizing Fullan’s conception of school leadership. They also provide school leaders with a clear guide in planning their professional development and in preparing an account of their practice. It is possible to imagine, for example, asking a school leader to provide an account of an instance when they led and managed a change initiative in some area of school functioning – one that led to improved teaching and learning. Further, the school leaders might be asked to include evidence related to Fullan’s five components in their account.

The five components could then form the foundation of a rubric for self-assessment, or assessment by peers, for professional recognition or certification. It would then be possible to imagine asking a group of carefully trained peers whether, in their view, the account provides clear and convincing evidence of the five components of Fullan’s concept of leading and managing change. That is, whether the school leader has shown that they have met the standard – that they have demonstrated the capacity to lead and manage change in an educational setting. These would be the steps involved in operationalizing a leadership standard.

Finally, and most important, the process and the experience of completing the assessment task should be

constructed so that it promotes professional learning. The task of putting together such an account in something similar to a portfolio entry should engage people in effective modes of workplace learning. Completing a well-prepared assessment task, with all the steps involved in learning about the standards, preparing the evidence and reflection on one's practice, can provide an excellent vehicle for active, school-based professional development. As a result, assessment and learning then go together.

As measures therefore, standards not only describe what practitioners need to know and be able to do to put these values into practice, but they also describe how attainment of that knowledge is to be assessed, and what counts as meeting the standard. A standard, in the latter sense, is the level of performance on the criterion being assessed that is considered satisfactory in terms of the purpose of the evaluation.

Content Standards

Standards that describe the nature and scope of a professional's work are usually referred to as content standards. As in educational measurement generally, content standards set out the domain of what is to be assessed. Ideally, they set out the main areas of effective practice and provide elaborations on what the standards mean in terms of what practitioners should know and be able to do.

Leithwood and Jantzi (2005) and Leithwood *et al.* (2004) provide a set of four core leadership practices grounded in research on how leadership influences student learning that could form part of the content domain of a set of school leadership standards. Each could be seen as an area within which a school leader could exercise leadership and lead and manage a change initiative:

- building vision and setting directions;
- understanding and developing people;
- redesigning the organization (e.g., building professional community/culture); and
- managing the teaching and learning program.

Each standard is elaborated on in the reports. Together they form a strong set of organizers for the content in a set of leadership standards. As the research reviewed by Leithwood and Jantzi (2005) indicates, these core practices are related (indirectly) to student achievement. It is possible to argue therefore that, as standards, they have content validity; the standards arguably identify the knowledge, skills, and attitudes possessed by effective leaders. Other reviews of effective leadership practices could also be used to define content standards for school leaders (e.g., Mulford, 2003a, 2005b).

Another example of content standards is the Standards for Headteachers from the NCSL in England (NCSL, 2004). In contrast to Leithwood *et al.*, these have six main organizers of the content of the standards:

- shaping the future;
- leading, learning, and teaching;
- managing the organization;
- developing self and working with others;
- securing accountability; and
- strengthening community.

While each of these lists from Leithwood *et al.* and the NCSL provides a powerful group of organizers for a set of school leadership standards, this is only the first step in developing standards. Each standard needs explication and elaboration in a full set of content standards. Standards writers need to drill down deeper to capture what effective school leaders know and do in relation to each standard to promote good teaching and provide quality learning opportunities for students.

For example, most recent sets of standards for school leaders usually have a standard that refers to the importance of building professional community, or culture, along the following lines:

Research indicates that the knowledge and skills of its teachers are the most important educational resource that a school possesses in meeting its mission to provide quality opportunities for students to learn. Highly accomplished principals establish a strong professional culture in their schools that nurtures and develops this resource. . . .

(Ingvarson *et al.*, 2006)

Echoes of such a standard can be found under Leithwood's *Developing People* organizer above, or in the *Developing Self and Working with Others* organizer from the NCSL. The standard can be justified in terms of research over the past 20 years indicating that an active, accountable professional community within a school is important for effective teacher development and high-quality teaching (Little and McLaughlin, 1993; Louis *et al.*, 1996).

Ingvarson *et al.* (2006) provide an elaboration of a Building Professional Community standard developed by the Australian Council for Educational Research (ACER). The elaboration teases out the various facets of the professional community concept, illustrates the areas where school leaders might take action, and points to types of evidence that one should see in, for example, a portfolio entry that a school leader was submitting to show that they had met this standard through leading and managing a project to strengthen professional community in their school.

Characteristics of Well-Written Standards

Several features of a standard such as Building Professional Culture and its elaboration are noteworthy. The first is that it points to a large, meaningful, and significant chunk of school leaders' work – it is an example of the purposes they are trying to achieve. It is not a microlevel competency, or a personality trait. School leaders readily identify this type of standard as an authentic (i.e., valid)

example of the kind of work they do (or should do). The standard is well grounded in research on the characteristics of effective schools.

The second is that the standard is context free, in the sense that it describes a practice, which most agree, accomplished principals should follow, no matter where the school is. By definition, a professional standard applies to all contexts in which professionals work (which is not to say that context does not affect practice). No matter where a school is, building professional culture in that school over time is likely to be regarded as a core responsibility of a school leader.

The third feature is that the standard is nonprescriptive about how to build professional culture; it does not standardize practice or force school leaders into some kind of straightjacket. There are many ways to build a professional culture. While the standard identifies an essential element of good leadership, it does not prescribe how the standard is to be met. In this way, the standard also allows for diversity and innovation.

The fourth is that, as a standard, with its elaboration, it points to something that is measurable, or observable. It is possible to imagine the types of evidence that a principal might assemble over time to show that they have strengthened the various components of professional community in their school and met the standard.

In summary, well-written standards:

- are grounded in clear guiding conceptions of leadership;
- are valid; that is, represent what school leaders need to know and do to play in promoting quality learning opportunities for students;
- identify the unique features of what school leaders' know and do;
- delineate the main dimensions of development the profession expects of its members – what school leaders should get better at over time, with adequate opportunities for PD; and
- are assessable; that is, point to potentially observable leadership actions.

Leithwood and Steinbach (forthcoming) provide a critical review of several existing sets of standards and propose an overarching set of seven meta-level principles, or standards for standards, that might be used to evaluate a set of school leadership standards.

The process by which a set of standards is developed is also a critical issue, not only for the validity of the assessment procedures, but also for their legal defensibility. If the procedures for developing standards are to be regarded as valid and legally defensible, it is important to ensure:

- the integrity and independence of the body responsible for developing the standards;
- that the standards developing body is composed primarily of those who are already highly accomplished practitioners;

- that the diversity of perspectives in the profession is represented;
- that the process of defining the standards is developed on a sound scientific basis and that the process of developing the standards is formally documented; and
- that a wide sampling of agreement is sought for the standards from the major professional groups and other interested parties regarding the appropriateness and level of the standards (Hattie, 2008).

Few developers of standards, apart from those who developed the ISLLC standards, make explicit the procedural principles on which they developed their standards (Ingvarson *et al.*, 2006), but that is likely to change as the teaching profession becomes more involved in developing profession-wide standards that it wants key stakeholders to regard as credible.

Methods of Assessing Performance against the Standards

Although a trend toward performance-based methods for assessing teachers and school leaders against standards is noticeable, most are still based on evidence of course completion. A trend toward more authentic forms of assessment for professional certification is found generally across other professions, both for initial registration and for advanced levels of professional certification. The term performance, as used here, refers to an authentic instance of a school leader's work over time, situated in a specific context (such as Building Professional Community in an English Department around collaborative assessment of student writing).

This trend is closely related to research on more effective modes and designs for learning and program delivery. These emphasize the need to create a wide range of opportunities to gain leadership experience through taking on and documenting small-scale leadership projects, such as leading and managing a change initiative in a particular area of school functioning that improves learning opportunities. This kind of experience creates stepping stones for teachers from the classroom to school leadership.

Here, for example, is a summary of an authentic assessment task referred to as leading and managing change, under development by the ACER, which a teacher might complete and enter into a portfolio of evidence:

This portfolio task invites you to document an initiative in which you led and managed a change effort with colleagues in your school. The project will have aimed to meet an important need related to student learning by enhancing the quality of teaching in your school. In your entry, you will provide evidence for: the need; your project plan to meet that need; what happened during the implementation of your project; and the improvements in

teaching and learning that took place. Finally you will reflect on what you have learned about leading and managing change and yourself as a leader (pp. 8–10).

Teachers and school leaders are very good at creating assessment tasks similar to this, which are authentic examples of what school leaders should know and be able to do. Completing such portfolio tasks provides opportunities for collaborative reflection on practice with colleagues and learning in the workplace. To increase reliability in the assessment, school leaders need to complete several such tasks, and each task should provide evidence relevant to several standards. Rather than engendering more work, authentic assessment tasks should be based on the natural harvest of evidence gathered in the course of conducting normal school functions and responsibilities.

In summary, here is a list of principles to guide the development of valid tasks for gathering evidence about school leader performance:

- tasks should be authentic and, therefore, complex;
- tasks should allow for the variety of forms that sound school leader practice can take;
- tasks should be open-ended, allowing school leaders to show their own practice;
- tasks should be fair; that is, they should give school leaders a fair chance to demonstrate the quality of their practice;
- tasks should provide ample opportunity and encouragement for analysis and reflection;
- research-based knowledge should underlie all performances;
- tasks should encourage school leaders to exemplify good practice;
- each task should provide evidence relevant to a cluster of standards; and
- each standard should be assessed by more than one task (Pearlman, 2008).

Specification of required evidence – spelling out what evidence school leaders need to provide and how to present it – needs to be precise for an evaluation to be valid and feasible, and fair. This requires carefully structured portfolio tasks with guiding questions, not an open-ended invitation to fill a wheelbarrow. Vague or imprecise requirements often result in teachers and school leaders presenting an oversupply of evidence that bears little or no relation to the relevant standards, making reliable and fair judgments difficult or impossible.

Assessing the Evidence and Setting Performance Standards

The final stage in developing a complete set of standards is setting performance standards. While content standards

define the scope of school leaders' work, they do not tell us how good a school leader's performance needs to be in relation to the standards. Or, put another way, content standards alone do not tell us what a satisfactory level of performance is on the assessment tasks. The key questions to be answered in setting performance standards are: How good is good enough? or What counts as meeting the standard?

Setting standards, and training teachers and school leaders to use them in assessing evidence can be just as complex as identifying the content standards. Evidence gathered by ACER over recent years suggests that teachers and school leaders can do this very well. The process involves developing scales and scoring rubrics, weighting different tasks and sources of evidence, identifying benchmark performances, and training assessors. Recent experience indicates that teachers and school leaders can reach high levels of reliability assessing evidence of performance in relation to the standards.

Conclusion

The main reasons for establishing school leadership standards are to:

- increase the effectiveness of professional preparation and development for school leaders, and
- provide better systems for recognizing and rewarding effective school leaders.

Although our review of international developments in leadership standards indicated a high level of standards development activity, few had gone beyond developing content standards to develop performance standards. Few elaborated deeply on what the standards meant in terms of practice; few could be used to assess performance; and few were accompanied by research programs on the validity of the standards or the reliability of assessments based on the standards.

Standards that are observable and assessable are more useful for self-assessment and professional learning. It is primarily by engaging future school leaders and experienced principals in more effective professional learning that standards can make a major contribution to improving student learning. Well-written standards can provide the basis for a standards-based professional learning and recognition system (Ingvarson and Kleinhenz, 2006a, 2006b), which includes four basic components:

1. standards that describe what effective leaders know and do thereby providing long-term goals and direction for professional learning;
2. an infrastructure for professional learning that enables school leaders to develop the attributes and capabilities embodied in the standards;

3. a credible system for assessing school leader performance against the standards that provide certification to school leaders who meet the standards; and
4. selection procedures and career paths that provide recognition and incentives from school authorities for those who gain professional certification.

Taken together, these components form a standards system of interdependent and mutually supportive parts. The four elements of standards, professional learning, certification, and recognition are interlinked. They can be applied to any profession. Take one away and the system loses its capacity to function effectively as an instrument for encouraging and recognizing evidence of professional learning.

Standards can be a means by which a profession exercises more control over its professional learning system. Engagement in standards development builds ownership and can be a powerful form of professional learning in itself. The capacity to develop standards gives the profession the ability to play a stronger role in defining the long-term goals of their own professional learning.

Responsibility for the development and application of professional standards also enables the profession to play a more significant role in providing recognition to members who meet its standards. As writers of standards must synthesize the implications of research on effective leadership practices, responsibility for the development and application of professional standards also enables the profession to play a stronger leadership role in relating research to practice.

Acknowledgment

Valuable assistance in preparing this article was provided by Michelle Anderson and Bill Mulford.

Bibliography

- Council of Chief State School Officers (1996). *Interstate School Leaders Licensure Consortium Standards for School Leaders*. Washington, DC: Council of Chief State School Officers.
- Fullan, M. (2001). *Leading in a Culture of Change*. San Francisco, CA: Jossey-Bass.
- Gronn, P. (2003). *The New Work of Educational Leaders: Changing Leadership Practice in an Era of School Reform*. London: Paul Chapman.
- Hattie, J. (2008). Validating the specification of a complex content domain: Applications to the national board for professional teaching standards. In Ingvarson, L. C. and Hattie, J. (eds.) *Assessing Teachers for Professional Certification: The First Decade of the National Board for Professional Teaching Standards*, vol. 11, pp 55–93. Amsterdam: Elsevier.
- Huber, S. (2004). *Preparing School Leaders for the 21st Century: An International Comparison of Development Programs in 15 Countries*. London: RoutledgeFalmer/Taylor and Francis.
- Ingvarson, L. C. and Hattie, J. (2008). *Assessing Teachers for Professional Certification: The First Decade of the National Board for Professional Teaching Standards*. London: Elsevier.
- Ingvarson, L. C. and Kleinhenz, E. (2006a). *A Standards-Guided Professional Learning System*. Melbourne, VIC: Centre for Strategic Education. <http://www.cse.edu.au> (accessed April 2009).
- Ingvarson, L. C. and Kleinhenz, E. (2006b). Estándares profesionales de práctica y su importancia para la enseñanza. *Revista de Educación* **340**, 265–298.
- Ingvarson, L. C., Anderson, M., Gronn, P., and Jackson, A. (2006). *Standards for School Leadership: A Critical Review of the Literature*. Canberra: Teaching Australia, Australian Institute for Teaching and School Leadership. <http://www.teachingaustralia.edu.au/ta/go/home/projects/standards> (accessed April 2009).
- Little, J. W. and McLaughlin, M. W. (eds.) (1993). *Teachers' Work: Individuals, Colleagues, and Contexts*. New York: Teachers College Press.
- Leithwood, K. and Jantzi, D. (2005). *A Review of Transformational School Leadership Research 1996 to 2005*. Montreal, QC: American Educational Research Association (AERA).
- Leithwood, K., Seashore-Louis, K., Anderson, S., and Wahlstrom, K. (2004). Executive summary: How leadership influences student learning. *Learning from Leadership Project*. The Wallace Foundation.
- Leithwood, K. and Steinbach, R. (forthcoming). Toward a second generation of school leadership standards. In Hallinger, P. (ed.) *Global Trends in School Leadership Preparation*. Lisse: Swets and Zeitlinger.
- Levine, A. (2005). *Educating School Leaders*. Washington, DC: The Education Schools Project.
- Louis, K. S., Marks, H. M., and Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal* **33**(4), 757–798.
- Mulford, B. (2003a). *School Leaders: Changing Roles and Impact on Teacher and School Effectiveness*. Paris: Education and Training Policy Division OECD.
- Mulford, B. (2005b). Quality evidence about leadership for organisational and student learning in schools. *School Leadership and Management* **25**(4), 321–330.
- Murphy, J. and Shipman, N. J. (1999). The interstate school leaders licensure consortium: A standards-based approach to strengthening educational leadership. *Journal of Personnel Evaluation in Education* **13**(3), 205–224.
- NCSL (National College for School Leadership) (2004). *National Standards for Headteachers*. Annesley: Department for Education and Skills Publications. <http://www.ncsl.org.uk/publications/publications-nationalstandards.cfm> (accessed April 2009).
- NCTM (National Council of Teachers of Mathematics) (1991). *Professional Standards for the Teaching of Mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Pearlman, M. (2008). The design architecture of NBPTS certification assessments. In Ingvarson, L. C. and Hattie, J. (eds.) *Assessing Teachers for Professional Certification: The First Decade of the National Board for Professional Teaching Standards*, vol. 11, pp 55–93. Amsterdam: Elsevier.
- Smith, M. and O'Day, J. (1990). Systemic school reform. In Fuhrman, S. and Malen, B. (eds.) *The Politics of Curriculum and Testing*, pp 233–267. Philadelphia, PA: Falmer Press.
- Sykes, G. and Plastrick, P. (1993). *Standard Setting as Educational Reform*. Washington, DC: American Association of Colleges for Teachers of Education.

Further Reading

- Dutch Principals Academy (2005). *Brief Introduction on the Activities of the Dutch Principals' Academy*, pp 5. Utrecht: Dutch Principals Academy.
- Fullan, M. (2003). Leadership and sustainability. *Plain Talk; The Newsletter for the Center for Development and Learning* **8**, 11.
- Getzels, J. W. and Jackson, P. W. (1963). The teachers' personality and characteristics. In Gage, N. L. (ed.) *Handbook of Research on Teaching*, pp 506–582. Chicago, IL: American Educational Research Association.

Scottish Executive (2003). *Continuing Professional Development for Educational Leaders*. Edinburgh: Scottish Executive.

Relevant Websites

<http://www.eddept.wa.edu.au> – Department of Education and Training, Leadership Centre.

<http://isp.ecu.edu.au> – Edith Cowan University.
<http://www.eastconn.org> – Eastconn, Connecticut Administrator Test.
<http://www.nbpts.org> – National Board for Professional Teaching Standards.
<http://www.nsanederland.nl> – Nedelandse Schoolleiders Academie, Dutch Principals Academy.
<http://www.teachingaustralia.edu.au> – Teaching Australia, Australian Institute for Teaching and School Leadership Limited.
<http://www.ccsso.org> – The Council of Chief State School Officers.

LEADERSHIP AND MANAGEMENT – LEADERSHIP ISSUES

Contents

Effective Leadership in Challenging Schools
Esthetics, Educational Leadership and Management
Leader Mistreatment of Teachers
Leadership and Gender
Leadership and Student Voice
Leadership and Urban Education
Leadership for Learning
Leadership in Diverse Cultures
Leading Diverse Schools
Supervision and Personnel Administration
The Big Challenge: Leadership for Inclusion
The Educational Leader's New Work

Effective Leadership in Challenging Schools

A Harris, Institute of Education, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The challenge of improving schools in the most disadvantaged contexts is currently high on the political agenda in many countries. It remains the case that many schools face acute levels of socioeconomic deprivation compounded by sets of external factors that adversely affect their ability to perform (West and Pennell, 2003). Schools located in disadvantaged areas are often the recipients of higher-than-average numbers of students with diverse ethnic backgrounds and low literacy levels. The net result of the powerful amalgam of social and economic problems means that it is more difficult to improve schools in challenging contexts because of the multiple problems they face. As Power *et al.* (2002: 26) conclude, “educational outcomes in deprived areas are worse than those in non-deprived area, whether they are measured in terms of qualification, attendance, exclusions or ‘staying on’ rates. Inner-city areas in particular feature as having low outcomes.” They also point to the need to reduce the “compositional effects that appear to result from high concentrations of disadvantaged students” (p. 65).

Students from disadvantaged backgrounds can challenge teachers’ conceptions of what to teach, what to expect of students, and even how to communicate with them (Knapp, 2001). This is not to suggest a deficit model of teaching in schools in difficult circumstances,

but simply to acknowledge the extent of the task in securing the levels of performance that schools in more affluent areas take for granted. Research has shown that in order to achieve and sustain improvement in schools in challenging contexts, teachers must exceed what might be termed as normal efforts (Maden, 2001). They have to work much harder and be more committed than their peers in more affluent schools. In addition, teachers in challenging schools have to maintain that effort in order to sustain improvement over time. Research has consistently shown that success can be short-lived and fragile in schools in difficult or challenging circumstances (Whitty, 2001: 109).

Is Improvement Possible?

The question of whether sustained improvement is really possible in schools in challenging circumstances remains. The school improvement movement has often been criticized for ignoring powerful socioeconomic influences that impact schools and for offering naive and, sometimes, simplistic solutions to complex social and economic problems (Thrupp, 1999). While there is undoubtedly some truth in this, a number of researchers working in this field have engaged with issues of context and context-specific

improvement (e.g., Hopkins, 2001; Reynolds *et al.*, 2001; Harris *et al.*, 2003; Teddlie *et al.*, 2000; Townsend *et al.*, 2007).

The emerging evidence base points toward the range of difficulties that high-poverty schools face in simply getting to the starting line for improvement (Harris and Chapman, 2002; Stoll and Myers, 1998). High staff turnover, poor facilities, lack of resources, falling student numbers, and a constant stream of supply teachers are pressures that schools in more prosperous areas simply do not face (Reynolds *et al.*, 2004; Harris *et al.*, 2006). Factors such as geographical isolation – particularly of rural schools – weak support at the district level, low levels of formal qualifications in the local adult population, and poor employment opportunities, all serve to compound the problem and make the extent of the educational challenge facing these schools significantly greater (Reynolds *et al.*, 2004).

It is clear that the performance of schools in challenging contexts remains stubbornly low despite successive, determined, and well-intentioned efforts to alter the odds through targeted resources and strategic support. Initiatives aimed at supporting schools in disadvantaged areas have had variable success. Although resource to such schools may have increased, subsequent improvements in educational performance have not materialized (McKinsey, 2007). Some improvement initiatives or interventions have simply underestimated the power of the socioeconomic context or applied a standard formula for improvement in a nonstandard context. Others have failed to fully appreciate the need for differentiated intervention and support in schools in these circumstances. While schools might face similar socioeconomic difficulties, this does not mean that the same improvement strategies can be applied across all schools categorized as challenging (Harris *et al.*, 2006).

The research evidence suggests that while the powerful effect of the socioeconomic conditions cannot be ignored, the fact remains that schools can and do improve despite their context or circumstances. Evidence shows that schools in challenging circumstances can improve the levels of student performance and achievement (Borman, 2000; Harris *et al.*, 2003), and that the quality of leadership is a major contributory factor of this success (Reynolds *et al.*, 2001 and Hopkins, 2001). The importance of leadership in securing school improvement in schools in difficult contexts should not be surprising. The research base is unequivocal about the importance of leadership in securing effective school change, development, and improvement, particularly in challenging school contexts (Maden and Hillman, 1993; Maden, 2001; Leithwood *et al.*, 2004). The evidence reinforces that effective school leaders exercise an indirect but powerful influence on the effectiveness of the school and on the achievement of students, explaining up to a quarter of the school-level variance in student achievement (Mulford, 2007a, 2007b).

Effective Leadership in Challenging Schools

Research studies are beginning to offer some important insights about effective leadership in schools in challenging circumstances. First, many principals and teachers actively choose to work in schools in difficult contexts. Second, the type of leadership found in schools in difficult contexts is driven by a powerful sense of moral purpose and a central concern for the learning of all young people. Third, leadership in challenging schools has to be more centrally focused on raising expectations about performance than in schools in more affluent areas.

Contemporary research studies (Harris and Chapman, 2002; Harris *et al.*, 2006) have identified some common features or characteristics of effective leadership in schools in challenging circumstances. While it is clear that more research is needed, these studies offer an insight into successful leadership practice in schools in challenging circumstances.

Vision and Values

Of central importance to leaders in all schools, but particularly for schools in challenging circumstances, is the cooperation and alignment of others to a shared set of values and vision. Establishing a clear vision and communicating a sense of direction for the school is a critical task for leaders. A lack of direction or common purpose can be a contributory factor to a downward spiral of performance among staff in schools in challenging circumstances. Reestablishing direction and regaining staff confidence is essential if subsequent changes are to be effectively implemented.

The vision and practices of effective principals in challenging schools tend to be organized around a number of core personal values: the modeling and promotion of respect (for individuals); fairness and equality; caring for the well-being and whole development of students and staff; and integrity along with honesty. It is clear that the leadership values and visions of effective principals in challenging schools are primarily moral (i.e., dedicated to the welfare of staff and students (with the latter at the center)), rather than primarily instrumental (for economic reasons) or noneducative (for custodial reasons).

The emphasis upon core values such as respect, fairness, equality, integrity, and honesty is a way of defining or redefining the moral code of the school and setting in place minimum standards of conduct. With a well-defined vision and established values in place, the possibility of raising staff and student expectations of performance is enhanced.

Distributing Leadership

Evidence suggests that highly creative approaches to tackling complex problems are required in challenging schools

and the leadership task is one that cannot be undertaken by an individual. This implies distributed or shared forms of leadership where tasks are not the sole responsibility of one person or, indeed, the leadership team (Harris, 2007). Highly effective principals tend to invest in the leadership of others. The overarching message is one of building the community of the school in its widest sense, that is, through developing and involving others.

In many schools in difficult circumstances, the decision to work with and through teams as well as individuals is a common and effective response to the management of change. Effective principals tend to distribute certain strategic leadership responsibilities or core developmental work to teams or individuals within the school. While all principals clearly recognize the need to take ultimate responsibility for all decisions made, they also acknowledge the importance of empowering staff to lead important initiatives or developments within the school.

Learner-Centered Leadership

Effective principals in challenging schools are quick to dispel the cultural deficit notion and are strongly committed to improving teaching and enhancing learning. They place a strong emphasis upon student achievement and learning and those conditions that can lead to higher student performance. They set high expectations for students, emphasizing consistency in teaching practices, and provide clear rules about behavior and discipline. Their central developmental focus is on improving the quality of teaching and learning. In this sense, they are also effective instructional leaders.

In short, effective leaders in schools facing challenging circumstances ensure that improving the quality of teaching and learning is at the center of all development work and ensure that all teachers focus upon this central priority.

Investing in Staff Development

Effective leaders in all schools recognize the importance of staff development as a powerful lever for change and development. In schools in difficulty, staff development is also a very important means of maintaining staff morale and motivation. The emphasis placed on the continuing development of teachers is not only a clear endorsement that teachers are valued within the school, but is also recognition that their professional learning is important. It also secures and maintains positive relationships between staff as it encourages the collaboration and sharing of ideas. In schools in difficulty, the building of positive relationships between teachers, students, and parents is the most important leadership task. It is clear from the research evidence that schools with a collaborative culture are more likely to be able to improve student performance and achievement.

In challenging school contexts, it is important that time is allocated for staff development and that teachers have opportunities to meet. It is also important that teaching quality is monitored and reviewed. Internal accountability processes need to be firmly in place. It is clear that effective leaders do not ignore or tolerate poor teaching within their schools. Where it does exist, it is confronted, and ways are sought to improve performance. In the majority of cases, a combination of structured support, monitoring, and an individual development program can tackle the problem of ineffective classroom practices. To summarize, effective leadership in challenging schools is primarily about securing high-quality teaching and taking every opportunity to enhance the learning of all students.

Community Building

Effective leaders in schools in difficult contexts are acutely aware of the need to engage their community. This is achieved, in part, by communicating regularly with parents and engendering trust by showing genuine care for young people. They understand the forces within the community that impede learning, are aware of the negative forces of subcultures, and listen to parents' views and opinions regularly. Effective leaders recognized that family, school, and community relationships directly affected student outcomes; hence, the need to connect with the community is of paramount importance to them.

As Hall (1997: 237) notes, "it has become increasingly evident that multiple forms of inequality do not exist as separate or merely additive influences...different forms of disadvantage are experienced more or less as working together." The implication here is that schools cannot operate in isolation from other agencies or the wider community they serve, but must positively engage with the wider community.

Strategies for Improvement

The strategies used by effective principals across a range of challenging school contexts were:

- School leaders identified and pursued an important, yet attainable, first goal. They used this success to set more ambitious goals.
- School leaders redirected time and energy away from conflicts between adults, and appealed to teachers, support staff, and parents to put aside their own interests and focus on teaching and learning.
- School leaders created a collective sense of responsibility for school improvement. The shared sense of responsibility was nurtured by joint planning processes and reinforced by efforts to involve everyone in key components of the school's work.
- School leaders created opportunities for teachers to work, plan, and learn together around instructional issues.

- School leaders created additional time for instructional issues during the school day.
- School leadership purposefully distributed leadership and created opportunities for the staff to lead initiatives.

In summary, effective leadership in schools in difficulty builds sustainable capacity and has a relentless focus on teaching and learning. The evidence points toward leaders who create strong internal accountability processes and set high expectations in terms of attainment and achievement. They are also leaders who recognize the importance of sharing responsibility and generating the cultural conditions where individuals feel valued and are willing to take responsibility.

Context-Specific Improvement

We undoubtedly need to know much more about improving schools in challenging circumstances and, particularly, how such schools sustain improvement over time. We also need to know what particular combinations of external support and internal development are optimum for generating positive change and development. Increasingly, the evidence base is pointing toward the possibilities and potential of professional learning communities to generate the capacity for school improvement (Hargreaves, 2004; Stoll and Seashore Louis, 2007). Such communities offer opportunities for teachers to work together without being dependent upon external initiatives or interventions. However, much depends upon a school's internal capacity to become a learning community in the first place. It is clear that not all schools have this capacity and that those schools which would benefit most from teachers working together are least able to make this happen.

As Elmore (2003) points out, to meet performance targets, "schools must develop successively higher capacities" (p. 14), and that this is achieved by generating internal accountability rather than responding to external accountability. It is clear that many accountability systems require schools with the lowest capacity to improve most through a combination of compliance and sanctions. In the long term, this strategy is destined to fail unless capacity-building measures are simultaneously put in place. Gray (2004) argues that schools that create the capacity for improvement move through three phases of school improvement – "catching up, consolidation and moving ahead" (p. 306). He notes, however, that our ignorance about their starting points serves as a block to improvement and suggests that "we tend not to see these schools accurately because our frameworks are too limited." Consequently, we need more sophisticated ways of diagnosing the needs of schools in challenging circumstances and offering differentiated rather

than standardized approaches to improvement (Chapman and Harris, 2004).

As the long-term pattern of educational inequality looks set to remain, to rely on standardized approaches to school improvement would seem unwise. In schools in challenging contexts, such approaches are more likely to exacerbate the problem rather than solve it. The evidence would suggest that more locally owned and contextually sensitive improvement strategies are more likely to succeed (Teddle *et al.*, 2000). If the goal of raising performance in schools in challenging circumstances is to be achieved, school improvement approaches that ignore the inherent diversity and variability across and within schools will be guaranteed to fail. More context-specific and contextually sensitive approaches to improving these schools are urgently needed.

Nevertheless, how is context-specific school improvement achieved in practice? Lupton's (2004) work argues that there are serious policy issues that need to be embraced to generate a greater alignment between the needs of the school and the forms of improvement interventions employed. In terms of funding, she argues that current funding mechanisms are based on measures that are too crude and therefore mask deep differences between schools. In short, broad poverty or ethnicity measures may not be the optimum way of allocating extra resources to schools in disadvantaged contexts.

Another implication for policymakers is to recognize that the practice in schools in disadvantaged contexts might vary from the practice in other schools and that the aim of replicating the teaching and learning processes in a good school may not be appropriate. As Lupton (2004: 35) argues, "at the level of the school we need a better understanding of effective practice in particular circumstances, in addition to the generic practice lessons that are already available." For example, what is the optimum size of groups for students in difficult school contexts with emotional and behavioral needs? What are the benefits to different groups of students from mixed-ability or streamed groups in such schools? How are parents engaged most effectively in challenging circumstances? More radically, we may need to consider alternative models of schooling for students who find it difficult to learn within the standard organizational framework of a school.

A contextualized school improvement program would undoubtedly need a high degree of flexibility and diversity to meet the needs of the different types of students in different types of school. As Hargreaves (2004) has suggested, there is an "apartheid of professional development and school improvement" which has been generated by default, rather than design. He argues that schools that are performing well enjoy "earned autonomy", while those categorized as "failing or close to failing have prescribed programmes and endlessly intrusive monitoring and

inspection” (Hargreaves, 2004: 190). He notes that, often, schools in difficulty are in receipt of multiple innovations, “while the cruising schools with coasting teachers who ride in the slip stream of middle class academic achievers get off scot free” (Hargreaves, 2004: 190)

The implication of a contextualized approach to school improvement is to reverse this trend and allow schools in the most disadvantaged contexts the opportunities to innovate, experiment, and divert, if necessary, from the educational norms of schools in more favorable settings. It also implies a steady supply of well-qualified teachers who are recruited and retained in poorer areas. As Lupton (2004) notes, this will only be achieved if there are the right financial and career incentives in place to attract the best teachers to some of the most difficult teaching contexts. It is also possible that both initial training and ongoing professional development programs need to be more carefully differentiated to prepare teachers for the particular challenges of disadvantaged schools.

Final Word

Leadership is a complex undertaking in any school, but for schools in challenging circumstances, it presents extra challenges. The core message about effective leadership in schools facing difficult or challenging circumstances is one of building capacity through empowering, involving, and developing others and by providing systems of learning support, guidance, and assistance. Capacity building should be the central aim of all schools, but for those in challenging circumstances, it is imperative to ensure that school improvement is long term rather than short-lived.

While gains in student outcomes are hard fought, and sustainable improvement is particularly difficult to achieve in such challenging contexts, both are possible. However, to succeed in these difficult circumstances requires a school improvement process that fits the individual school context and is underpinned by an unrelenting focus on improving conditions at the learning level. The resilience, sheer capacity for hard work, and continuing adherence of principals and teachers under the most challenging circumstances to provide the best opportunities for the learning and achievement for students have been shown to make a significant difference to a school’s ability to improve.

To conclude, effective leaders in challenging school contexts are constantly managing tensions and problems directly related to the particular circumstances of the school. The main leadership task is one of coping with unpredictability, conflict, and dissent on a daily basis without discarding their moral purpose and core values. Effective leaders of schools in challenging circumstances tend to be, above all, people centered, combining a

willingness to be collaborative and to promote the highest levels of teaching and learning. The evidence suggests that while the difficulties facing such leaders are considerable, the personal and professional benefits far outweigh them and, in the majority of cases, change their school for the better.

Bibliography

- Borman, G. D., Rachuba, L., Datnow, A., et al. (2000). *Four Models of School Improvement. Successes and Challenges in Reforming Low-Performing, High Poverty Title 1 Schools*. Baltimore, MD: Johns Hopkins University, Center for Research into the Education of Students Placed At Risk.
- Chapman, C. and Harris, A. (2004). Strategies for school improvement in schools facing challenging circumstances. *Educational Research* 46(3), 119–228.
- Elmore, R. (2003). *Knowing the Right Thing to Do: School Improvement and Performance-Based Accountability*. Washington, DC: Best Practices Center, NGA, Harvard University.
- Gray, J. (2004). Frames of reference and traditions of interpretation: Some issues in the identification of ‘under-achieving’ schools. *British Journal of Educational Studies* 52(3), 293–309.
- Hall, P. A. (1997). Social capital a fragile asset. In Christie, I. and Perry, H. (eds.) *The Wealth and Poverty of Networks: Tackling Social Exclusion*. London: Demos.
- Hargreaves, A. (2004). Distinction and disgust: The emotional politics of school failure. *International Journal of Leadership in Education Theory and Practice* 7, 27–43.
- Harris, A. (2007). Distributed leadership: Conceptual confusion and empirical reticence. *International Journal of Leadership in Education* 10(3), 1–11.
- Harris, A. and Chapman, C. (2002). *Effective Leadership in Schools Facing Challenging Circumstances*. Final Report, NCSL, Nottingham. <http://www.ncsl.org.uk> (accessed April 2009).
- Harris, A., Muijs, D., Chapman, C., Stoll, L., and Russ, J. (2003). *Improving Schools in Disadvantaged Areas – A Review of the Research Evidence*. Moorfoot, UK: DfES.
- Harris, A., Clarke, P., James, S., Harris, B., and Gunraj, J. (2006). *Improving Schools in Difficulty*. London: Continuum Press.
- Hopkins, D. (2001). *Meeting the Challenge. An Improvement Guide for Schools Facing Challenging Circumstances*. London: Department for Education and Skills.
- Knapp, M. S. (2001). Policy, poverty and capable teaching. In Biddle, B. (ed.) *Social Class, Poverty and Education*. New York: Routledge Falmer.
- Lupton, R. (2004). *Schools in Disadvantaged Areas: Recognising Context and Raising Performance, CASE paper 76*. London: Centre for Analysis of Social Exclusion, London School of Economics and Political Science.
- Maden, M. (ed.) (2001). *Success against the Odds: Five Years on*. London: Routledge.
- Maden, M. and Hillman, J. (eds.) (1993). *Success against the Odds: Effective Schools in Disadvantage Areas*. London: Routledge.
- McKinsey (2007). *How the World’s Best Performing Systems Come out on Top*. London: McKinsey.
- Mulford, B. (2007a). *Overview of Research on Australian Educational Leadership 2001–2005, Monograph No. 40*. Melbourne, VIC: Australian Council for Educational Leaders.
- Mulford, B. (2007b). Successful school leadership: What is it and who decides? *Australian Journal of Education* 51(3), 228–246.
- Power, S., Warren, S., Gillbourn, D., et al. (2002). *Education in Deprived Areas. Outcomes, Inputs and Processes with Perspectives in Education Policy*. London: Institute of Education, University of London.
- Reynolds, D., Hopkins, D., Potter, D., and Chapman, C. (2001). *School Improvement for Schools Facing Challenging Circumstances*:

- A Review of Research and Practice*. London: Department for Education and Skills.
- Reynolds, D., Clarke, P., and Harris, A. (2004). Improving schools in exceptionally challenging circumstances. *American Education Research Association Conference*, 11–16 April, San Diego, CA.
- Stoll, L. and Myers, K. (1998). *No Quick Fixes. Perspectives on Schools in Difficulty*. London: Falmer.
- Stoll, L. and Seashore Louis, K. (2007). *Professional Learning Communities*. New York: Open University Press.
- Teddlie, C., Stringfield, S., and Reynolds, D. (2000). Context issues within school effectiveness research. In Teddlie, C. and Reynolds, D. (eds.) *The International Handbook of School Effectiveness Research*, pp 160–185. London: Falmer.
- Thrupp, M. (1999). *Schools Making a Difference: Let's Be Realistic! School Mix, School Effectiveness and the Social Limits of Reform*. Ballmoor, UK: Open University Press.
- Townsend, T. (ed.) (2007). *International Handbook of School Effectiveness and School Improvement*. Dordrecht: Springer.
- West, A. and Pennell, H. (2003). *Underachievement in Schools*. London: Falmer.
- Whitty, G. (2002). Education, social class and social exclusion. *Education and Social Justice* 1(1), 2–8.

Further Reading

- Harris, A. and Chapman, C. (2004). Towards differentiated improvement for schools in challenging circumstances. *British Journal of Educational Studies* 52(4), 417–431.
- Hopkins, D. and Reynolds, D. (2001). The past, present and future of school improvement: Towards the third age. *British Educational Research Journal* 27(4), 459–475.
- James, C., Connolly, M., Dunning, G., and Elliott, T. (2006). *How Very Effective Primary Schools Work*. London: Paul Chapman.
- Keys, W., Sharp, C., Greene, K., and Grayson, H. (2003). *Successful Leadership of Schools in Urban and Challenging Contexts. Report of the National College for School Leadership*. Nottingham, UK: NCSL.
- Leithwood, K. and Steinbach, R. (2002). Effective leadership for especially challenging schools. *Journal of Leadership in Education* 79(2), 73–82.
- Leithwood, K. and Steinbach, R. (2002). Successful leadership for especially challenging schools. Toronto: OISE/University of Toronto, unpublished paper.
- Trimble, S. (2002). Common elements of high performing, high poverty middle schools. *Middle School Journal* 33(4), 67–81.

Esthetics, Educational Leadership and Management

R Bates, University of Deakin, Waurn Ponds, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

The study of organization, administration, and leadership has largely ignored the important field of esthetics. This is as true in the field of education as in the parent disciplines of organizational, administrative, managerial, and leadership theory. This is surprising, as esthetics, being concerned with the appreciation and exploration of different ways of seeing and with the imaginary exploration of possible worlds and selves is, as Dewey suggested, fundamental to learning. Indeed, an exploration of esthetics as a foundation for organizational leadership in education might well provide grounds for a proper educational theory of administration.

It is possible, however, to trace a variety of attempts to link esthetics with organizational theory within the parent fields as well as in education. These attempts fall generally into six categories:

- esthetics and organizational contexts;
- esthetics and organizational artifacts;
- esthetics and organizational leadership;
- esthetics and organizational culture;
- esthetics and organizational power; and
- esthetics and organizational capital.

Esthetics and Organizational Contexts

The most obvious relationship between esthetics and organization is that of the built environment within which organizations operate. Fundamental to such environments is the envisaging of architectural form through an exploration of the nature of the organization of work. But this is not simply a matter of idiosyncratic design but also a reflection of the needs of the client and the contemporary zeitgeist. Larson, for instance, argues that architecture is a social art, one that is public and useful, one that "... must convince a client, mobilize the complex enterprise of building, inspire the public ... and work with the culture, visual skills, and symbolic vocabulary not of the client but of its time" (Larson, 1993: 16). Leadership, exercised through the design of physical settings is, therefore, both a response to particular needs and a bridging of those needs with the broader culture within which organizations operate.

One of the most interesting examinations of this relationship during the twentieth century is that of Guillen who explores the relationship between organization and esthetics through the specific instance of "the formulation by modernist artists of an aesthetic based on the beauty of

the machine and on the new scientific management methods of the turn of the [twentieth] century" (1997: 682). As he suggests, Taylorist scientific management had a profound impact on organizational studies and the management of work, including educational work, in the United States. This impact, however, was largely the effect of the profession of engineering and regarded as an essentially technical exercise. In Europe and Russia a differing esthetic tradition combined with scientific management to produce quite a different architectural tradition. For instance, while in the United States the professions of engineering and architecture developed separately, with engineering following the principles of scientific management and architecture clinging to "... the old fashioned taste for superfluous ornamentation" (Guillen, 1997: 684), in Europe "... modernist architects found an aesthetic message in scientific management, producing an unlikely synthesis between art and the mechanical world" (one which) "succeeded in combining technology with style, science with history, management with creativity, and functionality with aesthetic" (1997: 683).

Such a perspective was rejected in England and Spain where the traditions of William Morris and Antoni Gaudi and the Arts and Crafts movements held sway but in the rest of Europe, new forms of modernist architecture developed where form followed function. In France, Le Corbusier, Benoit-Levy, and Leroy; in Germany Gropius, Meyer, and Wagner; in Russia Arvatov, Krinsky, and Lissitzky among others articulated an architecture that found inspiration:

... not in nature but in the rationalized or Taylorized world of machine production. They showed us that scientific management contained an aesthetic message emphasizing regularity, continuity and solidity. In their eyes, monotony and standardization had become beautiful. (Guillen, 1997: 698)

In the United States, however, much building had come under the influence of engineers rather than architects. Callaghan (1962) showed how Taylorist engineering principles underlay the cult of efficiency in education, while Tyack and Hansot (1982) showed how regularity, monotony and standardization became fundamental to the practice of education and educational leadership. Indeed an educational esthetic based on engineering and the principles of scientific management but without the esthetic of the European modernist architects had come to dominate the construction and management of schools. In school design as in educational practice, "Thorndike won

and Dewey lost” (Lageman, 1989: 185). The result was, as Klein and Diket suggest, that “Universities and schools were built according to the factory metaphor, not designed to address the needs of their inhabitants” (Klein and Diket, 2006: 99).

Adopting an alternative esthetic for leaders as architects and architects as leaders, Klein and Diket argue for an educational approach to leadership in design where “... visionaries of change ... create portals and paths ... [and] ... alter spaces and beliefs about the way people view themselves, others and society” (2006: 104). This argues for quite a different approach to architecture in general and schools in particular – one where creativity, flexibility, and possibility take the place of regularity, monotony, and standardization. Leadership in their view should be exercised on the basis of a creative approach to education and design that opens up possibilities for learning and imagination.

Esthetics and Organizational Artifacts

A second approach to the esthetics of leadership is that advocated by Gagliardi (1996), and Strati (2001) as well as by Rafaeli and Pratt (2006). Their approach is both a way of studying organizations through the ways in which leaders demonstrate commitments through the employment of particular artifacts, and an assessment of organizations through the artifacts that they produce.

In the first place, organizational leaders are seen to demonstrate particular commitments through the ways in which they display particular esthetic artifacts: such displays announce the culture of the organization preferred by the leaders. As Strati suggests:

... aesthetic objects are not simply appreciated on the basis of their artistic workmanship alone; they are special principally because of the emotions, insights, and feelings that they arouse in the cultures of those who use them or talk about them. (Strati, 2001: 573)

Gagliardi argues further that the ways in which leaders employ esthetic artifacts both affect the perceptions of reality held by organizational members and either facilitate or hinder organizational action (Gagliardi, 1996: 568). In fact “... artefacts are pathways of action in the sense that they structure sensory experience and enlarge or narrow the range of behaviour that is materially possible” (1996: 569).

Rafaeli and Pratt provide a comprehensive analysis of the ways in which artifacts are used by leaders to manage organizational problems, especially those related to the construction of collective identity, strategies of communication with the market and environment, and the creation of work settings that foster behaviors consistent with corporate purposes (Rafaeli and Pratt, 2006; Gagliardi, 2006).

Dean *et al.* (1997) consider not only the ways in which organizations use esthetic objects to manage their internal processes, but also the ways in which organizations generate esthetic artifacts in order to interact with their environment. They show how leadership is focused on the production of artifacts that will generate esthetic appreciation and through such appreciation, commitment to the organization:

... such potential objects of aesthetic appreciation could include products (an automobile) and services (an airline flight), organizational settings (the ambience of a restaurant) and processes (fabrication in a manufacturing cell). It could also include jobs (the role of a team leader), computer programs (the look and feel of a graphic interface), “performances” (resolution of a messy conflict), and even teams (that work together “beautifully”). (1997: 424)

Oddly enough, such esthetic phenomena are often ignored or trivialized in the study of organizational leadership. As Witkin argued:

... the exclusion of the aesthetic from conceptualisations of ... modern organizations ... has meant that phenomena which are clearly recognized to be aesthetic tend to be conceptually trivialized. (Witkin, 1990: 327)

Yet, the communication of purpose through esthetics is a major component of leadership through the vision of organizational mission and the harmony of organizational coherence through the engineering of consent. Similarly, the relationship of the organization to its wider environment (public, market, politics, etc.) is managed through esthetic artifacts in terms of both products and communication systems (such as logos and advertising).

In education, esthetic artifacts are used continuously by leaders as they attempt to generate particular commitments and outcomes both internally within educational organizations and externally in terms of the way in which their organization is perceived.

Esthetics and Organizational Leadership

Perhaps the one enduring tradition of esthetics in leadership is the notion of leadership as performance, or, as Weber (1968) called it, charisma. For Weber, charisma was a form of authority distinct from other types of traditional, legal-rational, or bureaucratic authority in that it engaged leaders and followers in an intensity of relationship and commitment uncommon in these other forms. As Samier suggests:

Charisma potentially exceeds any other authority type. ... in its personal effect through emotional intensity, inspirational capacity, and sacrificial character – all of which are

creative processes taking recognisable and analysable expressive form. (Samier, 2006: 172)

While charisma has been involved in the great-man theory of leadership and its modern version – transformational leadership – most theories of organizational leadership assume a benevolent relationship between charisma and organizational and social outcomes (for instance Deal and Peterson, 1999; Saphier and King, 1995). But as Gronn, (2006) and Bates (2006b) point out, the cult of heroic leadership can serve ill purposes as well as good. However, there seems little doubt that such charisma depends very much on the presentation of self and the rituals of face-to-face encounters (Goffman, 1959, 1972). Indeed, Goffman's dramaturgical analysis was a precursor to the analysis of cult leadership and spin and imagemaking in the second-half of the twentieth century.

As Gardener (1992) points out, impression management is an important part of organizational dramaturgy where "... skilful players ... take great care in defining and playing their roles, because they realize the importance of their performance" (p. 34). Duke explores the importance of such impression management where "The artistry of leadership encompasses leaders' purposeful efforts to foster the impression of leadership [through] dramatics, design and orchestration" (Duke, 1986: 20) and Howard suggests that "leadership is less a process or a thing-in-itself than a matter of performance success" (Howard, 1996: 34).

Starratt (1990, 1993) explores the drama of leadership through consideration of leaders as players, directors, stage managers, critics, and educators (1993) arguing that leadership is both a dramatic performance and a bridge between organizational life and the social dramas of the wider society. Leadership as a performance is therefore a bridge between internal and external dramas.

The question arises therefore as to whether leadership in this sense of performance can be taught. Some, like Howard, don't believe so:

... however many skills of leading we may identify, however much we may learn about leading and leadership from the experience of others, leadership as an achievement, like virtuosity in the deployment of musical skills, cannot be taught. (1996: 34)

Others, like Duke, take a somewhat neutral ground, suggesting that one of the chief virtues of an esthetic approach to leadership is appreciation, in that "like beauty, leadership deserves to be enjoyed for its own sake" (1996: 24).

Greenfield, however, argues that experience of the esthetic is fundamental to the performance of leaders in that, first, esthetic experience gives us insights into the human condition that cannot be captured through rational thought and, second, that esthetic performance speaks

of human involvement in the world. "It gives us the highest form of authentic human expression embedded in the very fabric of human intention, human communication and human understanding" (in Harris 1996: 494). Here, there is an argument that without an appreciation of esthetic traditions through which understanding and intention are communicated most vividly, leadership is likely to be an impoverished, inauthentic activity where performance is divorced from humanity and esthetics divorced from ethics. For Greenfield both leadership and esthetics fundamentally "speak to questions of how to live a life" (in Greenfield and Ribbins, 1993: 430).

Harris develops this thesis with reference not only to Greenfield, but also to Greene (1995), Eisner (2002), and Collingwood (1958), putting the position that artistry both enlarges our experience and provides "an emancipatory potential ... that draws us towards envisioning the world as it could become" (Harris, 2006: 48). It is this emancipatory potential, that shows us how things might be but are not, that is at the heart of both art and leadership.

The issue of how art changes perception of reality and can inform leadership, in terms of imagining and articulating possibilities, is taken up by several commentators. For instance, Samier and Stanley in their exploration of legacy of the Romantic tradition argue that:

Romanticism is both an ideal and a caution for administration and leadership. ... it provides a liberationist ethic for rising above the mundane, the profane, and the conventional in which educational administration is mired, in its promise of heightening the human potential towards self-determination, self-actualisation and authenticity. It also includes a critique of power, authority educational mediocrity and political stagnation. (2006: 42)

Educational leadership is then, at its best, characterized by "the cultivation of a deep and broad sensibility, rather than a narrow rationality or self-interest characterised by the lust for power or money" (Samier and Stanley, 2006: 43). This is an esthetic ideal for the creation of character as a work of art among both educational leaders and the other inhabitants of educational organizations. The Romantic tradition also, however, warns of the slide into the ugly and grotesque through novels such as Mary Shelley's *Frankenstein* and the works of Edgar Allan Poe (see also Strati and Montoux, 2002).

Others speak specifically of the role of literature (Brieschke, 1990; Stanley, 2006), theater (Mangham and Overington, 1987; Meyer, 2001), and cinema (Stockton, 2006) in the preparation of educational leaders and the ways in which such esthetic experiences can both deepen insight into the human condition and bring about an understanding enlivened by the contradictions of culture, not the unities or harmonies (Brieschke, 1990: 389) as well

as a confrontation with the moral nature of decision making in contradictory circumstances (Meyer, 2001; Bebeau *et al.*, 1999).

Esthetics and Organizational Culture

The literature on leadership and organizational culture comes to education through two routes, anthropology and organizational studies. The anthropological perspective sees organizations as cultures, that is, as communities and as manifestations of human consciousness (Smircich, 1983: 347). In this view, the realization of organizational culture is a manifestation of both the commonality and diversity of the knowledge and beliefs of its participants through material artifacts, events, behavior, and emotions. On the other hand, organizational studies have tended to see organizational culture as something that an organization has: a shared system of beliefs and values that conveys a sense of identity, facilitates commitment, and enhances organizational stability (Smircich, 1983: 345–46). In this second sense, culture is seen as another critical lever or key by which strategic managers can influence and direct the course of their organizations (Smircich, 1983: 346).

These alternative approaches are articulated in education by, for instance, Bates (1987, 2005b, 2006a) and Angus (1993), who argue the anthropological view, and Beare (1982); and Deal and Peterson (1999) who argue the managerial view. The difference in view is one of conflict between the anthropological concern for understanding and the managerial concern with manipulation (Bates, 2006b: 161). The implications are not trivial as the managerial view produces a drive toward consensus – contrived through the exercise of authority – while the anthropological view acknowledges the struggle between diversity and coherence and the need in education, as elsewhere, to acknowledge the complexity of the dynamics of cultural interactions both within schools and across their boundaries with various communities (Bates, 2005b, 2006b).

This does not deny the role of cultural entrepreneurs or their role within organizations, or the importance of their role in helping construct ceremonies, myths, and rituals (Pettigrew, 1979), but it does suggest that such an entrepreneurial role is not necessarily restricted to organizational leaders. The anthropological approach also illuminates the cultural dynamics of organizations that provoke resistance as well as compliance and describes rituals that celebrate both identity and difference (Bernstein *et al.*, 1966; Samier, 1997). In short, such a perspective can help to

... reveal heterogeneity in educational organizations, identify the bases of both conflict and a shared or common culture, suggest resolutions for leaders of the adverse effects of cultural conflict, and demystify and clarify educational ethos and values for establishing intellectual

standards and integrity for planning and policy-making. (Harman, 1989 in Samier, 1997: 433)

Again, the power of the esthetic may well lie not solely in its ability to offer us shared cultural experience, but even more so in its power to help us cross cultural boundaries both within organizations and between organizations and other cultures. If the problem of contemporary society is indeed to answer Touraine's question "Can we live together?" (Touraine, 1990; Bates, 2005b) then we might well accept Schein's assertion that "We will not learn about the power of culture unless we cross real cultural boundaries" (Schein, 1996: 239).

Esthetics and Organizational Power

Weber was the great theorist of organizational power in the modern world, arguing that the increasing rationalization of organizational life generated by the spirit of capitalism spilled over and captured every other area of social, ethical, and esthetic activity (Weber, 1930, 1968). Rationalized, bureaucratic, and calculative organizations were, for Weber, instrumental in the displacement of cultured people from positions of leadership and their replacement by technical experts: "specialists without spirit, sensualists without heart" (1946: 182). This new form of organizational power – that of bureaucracy – set up a conflict between specialized expertise and the cultivated personality, a conflict that "is determined by the irresistibly expanding bureaucratisation of all public and private relations of authority and by the ever-increasing importance of expert and specialized knowledge. This fight intrudes into all intimate cultural questions" (Weber, 1946: 243).

As Samier (2002) and others have shown, and in contradiction to the continuing misrepresentation of Weber in organizational and educational studies (e.g., Bush, 2000), Weber increasingly viewed the iron cage brought about by such calculative rationalization with despair. The pessimism of his later work contrasts with the optimism of his earlier work which lauded the demystification of traditional society and the increasing autonomy of science, ethics, and esthetics (see Pusey, 1987, Bates, 2003).

Habermas, in his reinterpretation of Weber argues that, rather than the rationalization of ethical and esthetic values analyzed in Weber's early work being displaced by the rationalization of interests that became the focus of his later work, the process of rationalization in each of these spheres continues alongside each other in the conflicts between what Habermas calls the life-world and the system (Habermas, 1989).

This conflict is played out within organizations where, on the one hand, individuals search for utility "which becomes the predominant value that directs the deployment of technical rationality [and] where the exchange and consumption of commodities becomes the main

object of social action” and, on the other hand, they search for meaning through “substantive values, emotion and authentic self-expression [in] morality, ethics, culture, art and aesthetics” (Milley, 2006: 83). The result, in educational organizations as elsewhere, is “a fine balancing act between substantive cultural work and instrumental action” (Milley, 2006: 91).

Organizational power, and the theorization (or as Foucault (1980, 1998) would call it, the normalization) of organizational power, has, for the most part been justified by appeals to utility and efficiency. However, as Maxcy (2006: 64) points out, leadership that asserts organizational power that is vainglorious, value neutral, and programmatic is deeply problematic, for leading does not occur in a vacuum, but rather, is rooted in our deepest beliefs about humankind, nature, and the real world around us (2006: 65).

Cairns (2002) takes a similar position when he explores the relationships between esthetics, morality, and power in organizations. In his exploration of the redesign of a particular workplace, he examines the multiple, conflicting meanings constructed by individuals and groups with regard to the change. The esthetics of the redesign were supposedly guided, not so much by the principles of scientific management discussed earlier, but by consideration of interpersonal relations as a key driver of organizational design (2002: 800). However, the changed esthetics of the workplace produced a situation where various individuals and groups held divergent and conflicting beliefs simultaneously and without contradiction (2002: 817). These beliefs, about the power relationships embedded in the esthetics of the redesign of the workplace by their masters, were, for instance, argued by staff to be both genuine reforms directed at staff empowerment and the implementation of fashionable management models for managerial self-advancement (2002: 815).

Even at the level of the design of physical environments then, esthetics are capable of holding contradictory meanings in tension. In terms of the exercise of power by different groups in organizations, esthetics may in fact be important in allowing individuals and groups to relate the system and life-world within the organization in ways that produce “. . . not a state of consensus, [but] rather a state of divergent equilibrium” (Cairns, 2002: 815).

This is a particular consideration in education where typically members of such organizations have diverse and multiple goals as well as hold various definitions of reality. They are likely, therefore, to be sites of competing esthetics that need constantly to be rebalanced into Cairns divergent state of equilibrium.

Esthetics and Organizational Capital

Recently, attempts have been made to explore the place of educational management and leadership in the (re)

production and distribution of symbolic (esthetic) capital and the role of education in esthetic self-realization. This is an area of investigation that follows from the work of Bourdieu (1984, 1993) and his exploration of the ways in which symbolic capital is distributed throughout society and the role played by various cultural institutions, such as museums, art galleries, theaters, publishing houses, and foundations, in organizing, adjudicating, and distributing such capital. The effect of such institutions is to produce a hierarchy of taste (Bourdieu, 1984), access to, and ownership of which constitutes symbolic capital that in turn legitimates the position of the owner in social space (Bourdieu, 1984, 1993).

Schools play a vital role in the consecration of symbolic capital and the maintenance of hierarchies of taste and power. Educational leaders, therefore, inasmuch as they articulate particular notions of cultural/symbolic capital (classical as against popular culture, for instance) and, through the assessment systems of their institutions, certify the possession (or absence) of such capital. Typically, schools act to conserve traditional notions of symbolic/cultural capital. In this way the school simply

. . . maintains the pre-existing order, that is, the gap between pupils endowed with unequal amounts of cultural capital. More precisely, by a series of selection operations, the system separates the holders of inherited cultural capital from those who lack it. Differences in aptitude being inseparable from social differences according to inherited capital, the system tends to maintain pre-existing social differences. (Bourdieu, 1998: 20)

As Bates (2006b) suggests, advocates of high culture (Eliot, Leavis, Bloom) and popular (working class) culture (Hoggart, Willams, Hall) articulate very different versions of symbolic capital. Educational institutions that adopt particular versions of symbolic capital therefore label their graduates in particular ways. For many, such distinctions mark off the sacred from the profane (Bourdieu, 1998) and purity from danger (Douglas, 1970; Durkheim, 1971). The placing of individuals in symbolic space therefore depends upon both the differences in the location of institutions and the processes by which they associate particular differences with particular individuals.

The work of educational administrators is, therefore, esthetic, not only in terms of their vision of school culture and their embodiment of that vision in the esthetic performance of the self, but also in the act of consecration of a particular esthetic distinction between purity and danger, between the noble and the mundane, and their classification of individuals according to such categories (Bates, 2006: 212).

In the contemporary world however, the struggle to maintain clear hierarchies of symbolic capital and power is confused by the anarchy of cultures that now rub up against each other in various media. This is not to suggest

that such hierarchies are less important than they once were, but rather that the competition for places within and between hierarchies is more intense (McKee, 2005). Educational institutions and educational leaders are, therefore, increasingly caught up in debates about the production and reproduction of such hierarchies of symbolic capital and arguments over what particular esthetic (as well as ethical and instrumental) values they should transmit.

Conclusion

The application of esthetic theories to educational leadership and to the study of educational institutions has hardly begun. There are some promising leads in both organizational and educational literature, but much work remains to be done at both theoretical and empirical levels. As Samier (2006: 5) suggests, following Hostadter (1965) “art is the means by which the three functions of symbolism, expression, and meaning are integrated.” Such integration is at the heart of the educational process. It is also at the heart of organizational and social processes as Dewey (1916) acknowledged. But, for Dewey, such processes were fundamentally an exercise in consensual value formation across boundaries and always with the potential that things could be other than what they are. The exercise of such an esthetic imagination is important in both organizational and educational leadership.

Bibliography

- Angus, L. (1993). The sociology of school effectiveness. *British Journal of the Sociology of Education* 14(3), 333–345.
- Bates, R. (1987). Corporate culture, schooling and educational administration. *Educational Administration Quarterly* 23(4), 79–115.
- Bates, R. (2003). The bird that sets itself on fire: Thom Greenfield and the renewal of educational administration. In Macmillan, R. (ed.) *Questioning Leadership: The Greenfield Legacy*, pp 211–236. London, ON: Althouse Press.
- Bates, R. (2005). Can we live together? Towards a global curriculum. *Arts and Humanities in Higher Education* 4(1), 95–109.
- Bates, R. (2006a). Culture and leadership in educational administration. *Journal of Educational Administration and History* 38(2), 155–168.
- Bates, R. (2006b). Towards an aesthetics of educational administration. In Samier, E. and Bates, R. (eds.) *Aesthetic Dimensions of Educational Administration and Leadership*, pp 205–220. London: Routledge.
- Beare, H. (1982). Education's corporate language. *Unicorn* 8(1), 12–28.
- Bebeau, M., Rest, J., and Narvaez, D. (1999). Beyond the promise: A perspective on research in moral education. *Educational Researcher* 28(4), 18–25.
- Bernstein, B., Elvin, H., and Peters, R. (1966). Ritual in education. *Transactions of the Royal Society* 251(772), 429–436.
- Bourdieu, O. (1998). *Practical Reason*. Cambridge: Polity Press.
- Bourdieu, P. (1984). *Distinction: A Social Critique of the Judgement of Taste*, Nice, R. (trans.). Cambridge, MA: Harvard University Press.
- Bourdieu, P. (1993). *The Field of Cultural Production*, Johnson, R. (trans.). Cambridge: Polity Press.
- Brieschke, P. (1990). The administrator in fiction: Using the novel to teach educational administration. *Educational Administration Quarterly* 26(4), 376–393.
- Bush, A. (2000). Administration and management on education: Theory and practice. In Moon, B., Ben-Peretz, M., and Brown, S. (eds.) *Routledge International Companion to Education*, pp 273–282. London: Routledge.
- Cairns, G. (2002). Aesthetics, morality and power: Design as espoused freedom and implicit control. *Human Relations* 55(7), 799–820.
- Callaghan, R. (1962). *Education and the Cult of Efficiency*. Chicago, IL: Chicago University Press.
- Collingwood, R. G. (1958). *The Principles of Art*. Oxford: Oxford University Press.
- Deal, T. and Peterson, K. (1999). *Shaping School Cultures*. San Francisco, CA: Jossey-Bass.
- Dean, J. W., Ottensmeyer, E., and Ramirez, R. (1997). An aesthetic perspective on organizations. In Cooper, C. and Jackson, S. (eds.) *Creating Tomorrow's Organizations: A Handbook for Future Research in Organizational Behaviour*, ch. 22, pp 419–437. New York: Wiley.
- Dewey, J. (1916). *Democracy and Education*. New York: Macmillan.
- Douglas, M. (1970). *Purity and Danger*. Harmondsworth: Penguin.
- Duke, D. (1986). The aesthetics of leadership. *Educational Administration Quarterly* 22(1), 7–27.
- Durkheim, E. (1971). *The Elementary Forms of the Religious Life*, Swain, J. (trans.). London: George Allen and Unwin.
- Eisner, E. (2002). *The Arts and the Creation of Mind*. New Haven, CT: Yale University Press.
- Foucault, M. (1980). *Power/Knowledge*, Gordon, C. (trans.). Brighton: Harvester Press.
- Foucault, M. (1998). *Aesthetics, Method and Epistemology*, Hurley, R., et al. New York: New Press.
- Gagliardi, P. (1996). Exploring the aesthetic side of organizational life. In Clegg, S., Hardy, C., and Nord, W. (eds.) *Handbook of Organizational Studies*, pp 565–580. London: Sage.
- Gagliardi, P. (2006). Review of A Rafaeli & Michael Pratt *Artifacts and Organizations*. *Administrative Science Quarterly* September, 505–506.
- Gardener, W. (1992). Lessons in organizational dramaturgy: The art of impression management. *Organizational Dynamics* 21(1), 33–46.
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Harmondsworth: Penguin Books.
- Goffman, E. (1972). *Interaction Ritual*. Harmondsworth: Penguin Books.
- Greene, M. (1995). *Releasing the Imagination: Essays on Essays on Education, the Arts and Social Change*. San Francisco, CA: Jossey-Bass.
- Greenfield, T. B. and Ribbins, P. (1993). *Greenfield on Educational Administration: Towards a Humane Administration*. London: Routledge.
- Gronn, P. (2006). Aesthetics, heroism and the cult of the ‘leader’ In Samier, E. and Bates, R. (eds.) *Aesthetic Dimensions of Educational Administration and Leadership*, pp 191–204. London: Routledge.
- Guillen, M. F. (1997). Scientific management's lost aesthetic: Architecture, organization, and the taylorized beauty of the mechanical. *Administrative Science Quarterly* 42, 682–715.
- Habermas, J. (1989). *The Theory of Communicative Action: The Lifeworld and the System*, McCarthy, T. (trans.). Boston, MA: Beacon Press.
- Harman, K. (1989). Culture and conflict in academic organization: Symbolic aspects of university worlds. *Journal of Educational Administration* 27(3), 30–54.
- Harris, C. (1996). The aesthetic of Thomas B Greenfield: An exploration of practices that leave no mark. *Educational Administration Quarterly* 32(4), 487–511.
- Harris, C. (2006). Collingwood on imagination, expression and action. In Samier, E. and Bates, R. (eds.) *Aesthetic Dimensions of Educational Administration and Leadership*, pp 44–63. London: Routledge.
- Howard, V. (1996). The aesthetic face of leadership. *Journal of Aesthetic Education* 30(4), 22–37.
- Klein, S. and Diket, R. (2006). Aesthetic leadership: Leaders as architects. In Samier, E. and Bates, R. (eds.) *Aesthetic Dimensions of Educational Administration and Leadership*, pp 99–109. London: Routledge.

- Lageman, E. C. (1989). The plural world of educational researchers. *History of Education Quarterly* 29(2), 185–214.
- Larson, M. S. (1993). *Beyond the Post-Modern Façade: Architectural Change in Late Twentieth Century America*. Berkeley, CA: University of California Press.
- Mangham, I. L. and Overington, M. A. (1987). *Organizations as Theatre: A Social Psychology of Dramatic Appearances*. Chichester: Wiley.
- Maxcy, S. J. (2006). The metaphysical sources of a pragmatic aesthetic leadership. In Samier, E. and Bates, R. (eds.) *Aesthetic Dimensions of Educational Administration and Leadership*, pp 64–78. London: Routledge.
- McKee, A. (2005). *The Public Sphere: An Introduction*. Cambridge: Cambridge University Press.
- Meyer, M. (2001). Reflective leadership training in practice using theatre as representation. *International Journal of Leadership in Education* 4(2), 149–169.
- Milley, P. (2006). Aesthetic experience as resistance to the 'Iron Cage' of dominative administrative rationality. In Samier, E. and Bates, R. (eds.) *Aesthetic Dimensions of Educational Administration and Leadership*, pp 79–98. London: Routledge.
- Pettigrew, A. (1979). On studying organizational cultures. *Administrative Science Quarterly* 24(4), 570–581.
- Pusey, M. (1987). *Jurgen Habermas*. London: Tavistock.
- Rafaeli, A. and Pratt, M. (2006). *Artifacts and Organizations*. Mahwah, NJ: LEA.
- Samier, E. (1997). Administrative ritual and ceremony. *Educational Management and Administration* 25(4), 417–436.
- Samier, E. (2002). Weber on education and its administration: Prospects for leadership in a rationalised world. *Educational Management and Administration* 30(1), 27–45.
- Samier, E. (2006). The aesthetics of Charisma: Architectural, theatrical and literary dimensions. In Samier, E. and Bates, R. (eds.) *Aesthetic Dimensions of Educational Administration and Leadership*, pp 161–174. London: Routledge.
- Samier, E. and Stanley, A. (2006). The art and legacy of the romantic tradition: Implications for power, self-determination and education. In Samier, E. and Bates, R. (eds.) *Aesthetic Dimensions of Educational Administration and Leadership*, pp 34–44. London: Routledge.
- Saphier, J. and King, M. (1985). Good seeds grow in strong cultures. *Educational Leadership March*, 72–75.
- Schein, E. (1996). Culture: The missing concept in organizational studies. *Administrative Science Quarterly* 41(1), 229–240.
- Smircich, L. (1983). Concepts of culture and organizational analysis. *Administrative Science Quarterly* 28(1983), 339–358.
- Stanley, A. (2006). The Victorian hangover: Colourful headmasters in the works of Mr Dickens and Ms Bronte. In Samier, E. and Bates, R. (eds.) *Aesthetic Dimensions of Educational Administration and Leadership*, pp 110–127. London: Routledge.
- Starratt, R. J. (1990). *The Drama of Schooling*. London: Falmer.
- Starratt, R. J. (1993). *The Drama of Leadership*. London: Falmer.
- Stockton, J. (2006). A narrative looking glass for leadership studies in administration: Cinema and literature as source and reflective medium. In Samier, E. and Bates, R. (eds.) *Aesthetic Dimensions of Educational Administration and Leadership*, pp 128–144. London: Routledge.
- Strati, A. (1992). Aesthetic understanding of organizational life. *Academy of Management Review* 17, 568–581.
- Strati, A. (2001). *Organization and Aesthetics*. London: Sage.
- Weber, M. (1930). *The Protestant Ethic and the Spirit of Capitalism*. London: Unwin Hyman.
- Weber, M. (1946). *From Max Weber: Essays in Sociology*. New York: Oxford University Press.
- Weber, M. (1968). *Economy and Society: and Outline of Interpretive Sociology*. Berkeley, CA: University of California.
- Witkin, R. W. (1990). The aesthetic imperative of a rational-technical machinery: A study of organizational control through the design of artefacts. In Gagliardi, P. (ed.) *Symbols and Artifacts: Views of the Corporate Landscape*, pp 325–338. Berlin: DeGruyter.

Further Reading

- Bates, R. (1989). Leadership and the rationalization of society. In Smyth, J. (ed.) *Critical Perspectives on Educational Leadership*, pp 131–156. London: Falmer.
- Bates, R. (2005a). An anarchy of cultures. *Asia Pacific Journal of Teacher Education* 33(3), 231–241.
- Samier, E., Bates, R., and Stanley, A. (2006). Foundations and history of the social aesthetic. In Samier, E. and Bates, R. (eds.) *Aesthetic Dimensions of Educational Administration and Leadership*, pp 3–19. London: Routledge.
- Strati, A. and Montoux, P. (2002). Organizing aesthetics. *Human Relations* 55(7), 755–766.
- Touraine, A. (2000). *Can We Live Together?* Stanford, CA: Stanford University Press.

Leader Mistreatment of Teachers

J Blase and J Blase, University of Georgia, Athens, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The problem of school leader workplace mistreatment of teachers has only recently provoked scholarly attention. This article reviews the major conceptual, typological, and theoretical approaches to the general problem of workplace mistreatment/abuse and superordinate mistreatment/abuse of subordinates in particular. This is followed by a presentation of the findings of empirical research focusing on abusive behavior and its effects on the personal–professional lives of workers in organizational settings. Major findings from the two existing published studies of school leader mistreatment/abuse of teachers are also presented. A brief methodological critique of the general workplace mistreatment/abuse literature and discussion of further research conclude this article.

Internationally, research on workplace mistreatment/abuse has increased substantially during the past two decades. Most of the theoretical and empirical work has been published in Western Europe, particularly in countries such as Finland, Sweden, Great Britain, France, and Norway, as well as Australia and Canada. Although slow to start, scholars in the United States have made significant contributions to the rapidly emerging international research base on workplace mistreatment/abuse (Blase and Blase, 2002, 2003a, 2003b; Einarsen *et al.*, 2003; Rayner *et al.*, 2002). In some countries, significant legislation and organizational policies have been created to prohibit abusive conduct by superordinates and co-workers and provide opportunities for redress (Irish Taskforce on the Prevention of Workplace Bullying, 2001; Lynch, 2002; Office of the Employee Ombudsman, 2000; Queensland Government Department of Industrial Relations, 2004). Despite this proliferation of research and legal analyses that support the development of organizational policies and legislation addressing this serious problem, such policies and legislative proposals are only beginning to emerge in the United States (Yamada, 2000, 2004). Unfortunately, only two studies of school leader mistreatment/abuse of teachers have been published, both in the United States (Blase and Blase, 2002, 2003a, 2003b; Blase *et al.*, 2008), despite the fact that several large-scale studies of workplace mistreatment/abuse across occupations in Sweden (Leymann, 1990), Norway (Matthiesen *et al.*, 1989), Great Britain (Hoel and Cooper, 2000), Ireland (Irish Taskforce on the Prevention of Workplace Bullying, 2001), and Australia (Queensland Government Workplace Bullying Taskforce, 2002) indicate

that teaching is one of the highest-risk occupations for mistreatment/abuse.

Concepts of Workplace Mistreatment/Abuse

A range of concepts has appeared in the scholarly literature to define workplace mistreatment/abuse; generally, these concepts refer to the same phenomenon. Björkvist *et al.* (1994) used the term work harassment to describe repeated actions intended to provoke emotional pain against individuals or groups unable to defend themselves. Price-Spratlen (1995) used the more inclusive terms, mistreatment and abuse, interchangeably, to identify actions perceived as excessive, inappropriate, and unwanted by a target. Mobbing was used by Leymann (1990) to refer to communication by one or more individuals acting together toward a target over a long period of time that produces substantial emotional, physical, and social damage. Bullying, a concept used by Einarsen and Skogstad (1996), describes a pattern of harassment of one or more persons by an individual of superior power.

A handful of concepts specifically describe administrator (superordinate and superior) mistreatment/abuse of subordinates. Abusive supervision, as discussed by Tepper (2000), refers to a prolonged use of harmful verbal and nonverbal behaviors as well as physical contact. Ashforth (1994) described petty tyranny as the behavior of one who lords power over others, behaves in self-aggrandizing and arbitrary ways, demeans employees, is inconsiderate, discourages initiative, and arbitrarily dispenses punishment. In both the works of Hornstein *et al.* (1995) and Hornstein (1996), abusive supervision has been described as transgressions, including coercion, selfishness, deceit, cruelty, and disregard, by administrators toward subordinates. Blase *et al.* (2008) produced a broad-based inclusive definition of school administrator mistreatment/abuse drawn directly from studies of abused teachers: typically, a pattern of verbal, nonverbal, and physical behaviors (excluding physical violence) by an administrator over an extended period of time that, in the teacher's perception, causes psychological–emotional, physical–physiological, and profession-life and personal-life harm; racial and sexual harassment are included in this definition of school administrator mistreatment/abuse (Blase and Blase, 2002, 2003a, 2003b; Blase *et al.*, 2008). Under certain circumstances, a single behavior

may be considered mistreatment/abuse, depending on its type, frequency, and severity (Yamada, 2004).

Keashly's (1998) comprehensive definition of workplace emotional abuse identified verbal and nonverbal behaviors used to elicit compliance from others; specifically, abusive behavior is unwanted by the target, intended to harm the target, constitutes a pattern of abuse, violates norms for appropriate conduct and individual rights, produces harm, and occurs in the context of an unequal power relationship. It should be mentioned that organizational scholars generally agree that frequency and duration are critical and strong elements of a definition of workplace mistreatment/abuse; however, a perpetrator's intentions and an unequal power relationship are less critical, weaker elements of such a definition. Put differently, definitions of workplace mistreatment/abuse may and should vary with regard to context. In addition, the inclusion or exclusion of race-based and sex-based harassment in such definitions, although appropriate to teachers' perspectives of mistreatment/abuse, remains inconsistent in the scholarly literature (Blase and Blase, 2002, 2003a, 2003b; Hoel *et al.*, 1999; Einarsen *et al.*, 2003; Einarsen and Skogstad, 1996; Irish Taskforce on the Prevention of Workplace Bullying, 2001; Rayner, 2000; Westhues, 2004).

Typologies of Abusive Behaviors

A handful of typologies of abusive behavior in the work setting that are constructed from empirical research have been aptly employed to classify abusive leader behavior. The applicability of a particular typology would vary according to context. Baron and Neuman (1996) developed a three-factor model of workplace aggression that is based on Buss's (1961) seminal model of human aggression and identifies manifestations of hostility, obstructionism, and overt aggression. Manifestations of hostility include verbal and symbolic behavior, such as obscene physical gestures, dirty looks, staring, silent treatment, ridiculing a target's work, belittling an individual's opinion, giving unfair negative information to higher-level administrators, flaunting status, and delivering unwarranted reprimands. Obstructionism describes conduct that is aggressive but passive in nature; this includes stonewalling, not returning phone calls or responding to e-mails and memos, interfering with the target's work, withholding permission or resources, failing to provide needed information, and soliciting others to delay needed decisions and actions. The third factor, overt aggression, refers to threats or acts of physical violence or destruction or the theft of property. Note that acts of physical violence are excluded from most perspectives of workplace mistreatment/abuse.

Ryan and Oestrich's (1991, 1998) model of leader mistreatment/abuse describes abrasive (i.e., less harmful) and abusive (i.e., more harmful) behavior, including silence,

glaring, abruptness, ignoring or snubbing, insults, blaming, discrediting, aggressive control, threats, shouting and yelling, angry outbursts, and threats of physical harm. Consistent with the workplace mistreatment/abuse literature in general, Ryan and Oestrich argued that the degree of harm caused by a behavior will vary in terms of frequency, place of occurrence (e.g., public vs. private context), and timing. Robinson and Bennett's (1995) typology of workplace deviance conceptualizes various types of behavior by two dimensions, minor versus serious and interpersonal versus organizational. Minor types of deviance are described in terms of two quadrants, including productive deviance (e.g., taking excessive breaks and leaving early) and political deviance (e.g., blaming subordinates, favoritism, and gossiping). More serious types of deviance are described as property deviance (e.g., lying about hours worked and damaging equipment) and personal aggression (e.g., endangering subordinates, stealing from subordinates, verbal abuse, and sexual harassment).

Theories of Workplace Mistreatment/Abuse

Several important theories of leader abuse of subordinates deserve attention. Hornstein *et al.* (1995) developed a perspective on supervisory disrespect based on theoretical work in organizational justice, symbolic interactionism, stress theory, and the general psychological literature. These scholars soundly argued that an individual's feelings of self-worth and security are influenced by how they are treated by others and that these feelings affect one's psychological/emotional well-being. Hornstein *et al.* used the Boss Behavior Questionnaire (BBQ), a valid and reliable measure of respect and disrespect. They have demonstrated that disrespectful behavior by leaders toward subordinates violates an individual's dignity and produces a loss of self-esteem and a range of adverse stress-related conditions (e.g., anxiety and depression).

Likewise, Folger's (1993) theoretical work – which is based on exchange theory and, specifically, referent cognitive theory (RCT) – is another insightful explanation of how subordinates come to feel mistreated by superordinates. This theory defines employment as both the economic and social exchanges of material goods (e.g., wages and benefits) and relational goods (e.g., treatment by superiors with regard to interpersonal processes, such as politeness and respect). Folger contended that in addition to equitable treatment with regard to the application of policies and procedures, superordinates have a moral obligation to treat employees with dignity in accordance with normative standards. During an exchange, respectful conduct by a superordinate enhances a subordinate's feelings of dignity, self-respect, and self-worth, responses essential to the perception that one has been treated as a person, not as

an object. Subordinates can be expected to have a hostile reaction (e.g., moral indignation) when a superordinate's interpersonal treatment has placed their dignity in jeopardy.

By comparison, Ashforth (1994) proposed a model of petty tyranny based on antecedents and effects. He maintained that petty tyranny results from an interaction of the predispositions of superordinates (i.e., beliefs about the organization, such as believing that it is a bureaucratic one), beliefs about subordinates (e.g., in the theory X approach, superordinates believe that the average person dislikes work, lacks ambition, requires direction, and is resistant to change; such beliefs often lead to a close, coercive leadership style (McGregor, 1960)), and situational facilitators (i.e., macro-level factors, such as institutional values and norms, and micro-level factors, such as power). Ashforth hypothesized that such antecedents interact to produce petty tyranny (e.g., arbitrariness and self-aggrandizement, belittling subordinates, lack of consideration, and non-contingent punishment). This approach to leadership causes low leader endorsement; high frustration, stress, and reactance; high helplessness and work alienation; poor work performance; and low work-unit cohesiveness. Ashforth described feedback loops that connect the effects of tyranny with its antecedents and behavior. Subordinate alienation, for example, may enhance negative dispositions toward subordinates (an antecedent) and increase the likelihood of greater use of coercion (behavior). More important, subsequent research by Ashforth (1997) suggested that petty tyranny may be accurately described as a gestalt (i.e., syndrome), a set of mutually reinforcing individual dispositions, situational facilitators, leader behaviors, and subordinate effects that interact to create a coherent cluster.

Several phase models of workplace mistreatment/abuse are particularly valuable. Leymann's (1990) seminal model of mobbing differentiates between and among various phases of workplace mistreatment: briefly, in phase 1, there is a triggering event that initiates a conflict; in phase 2, the individual is stigmatized and subjected to repeated injurious actions over an extended period of time; in phase 3, the management becomes involved, identifies the conflict as official and, because the victim has been stigmatized, predictably assumes the negative attitudes and prejudices of the abusers, which suggests a serious violation of a victim's rights by the management; and in phase 4, the victim is expelled from the organization. Similarly, Björkvist (1992) described a three-phase model of mistreatment that emphasizes the severity and escalation of mistreatment, and Zapf and Gross (2001) discussed the applicability of Glasl's (1994) three-phase model of conflict, including rationality and control, severing the relationship, and aggression and destruction. Although phase-escalation models of workplace mistreatment/abuse include contextual variables (e.g., organizational leadership and the organization of work), such inclusions tend to be limited.

In contrast, prominent causal models view workplace mistreatment/abuse as a more complex phenomenon consisting of a host of personal (perpetrator and victim), social, and organizational variables (i.e., potentially multiple causes (reasons) that explain the mistreatment/abuse phenomenon). Neuman and Baron's (1998) theoretical model contends that the antecedents of aggression, that is, social factors (e.g., privation and unfair treatment) and situational factors (e.g., organizational culture and climate and restructuring) influence personal determinants (e.g., type A behavior – being irritable and impatient (Glass, 1977), control oriented (Miller *et al.*, 1985), emotional, and overly aggressive (Holmes and Will, 1985); self-monitoring problems), which, in turn, affect an individual's internal state (e.g., hostile feelings) and cognitive appraisal (e.g., was I attacked or treated improperly?) and produce an aggressive or nonaggressive response. Notwithstanding this, Zapf's (1999) causes and consequences model of mobbing posits greater complexity and multidirectionality. In brief, Zapf argued that organizational factors (e.g., leadership and culture) and social group factors (e.g., hostility and envy) influence a perpetrator to exhibit mobbing behaviors (e.g., rumors, social isolation, and verbal aggression). Such behaviors affect the victim's actions, and such actions, in turn, affect the perpetrator's actions. In effect, the causes of mistreatment/abuse are multilateral and interactive processes that produce consequences (e.g., depression and anxiety) and such consequences influence the organization, perpetrator, social group, and victim variables. Within this dynamic-interactive and multidirectional mistreatment/abuse process, causes become consequences and consequences become causes. Needless to say, leading-edge theories of workplace mistreatment/abuse suggest that one-sided, simplistic explanations of workplace mistreatment are misguided (Keashly, 1998; Rayner *et al.*, 2002; Zapf, 1999).

Additional theoretical knowledge directly relevant to mistreatment/abuse in organizations is available in the general leadership literature (e.g., Adams and Balfour, 1998; Hodgkinson, 1991; Kets de Vries, 1989, 1993), social psychological literature (e.g., Baumeister, 1996; Haney *et al.*, 1973; Katz, 1993; Kekes, 1990; Kipnis, 1972, 2001; Lee-Chai and Bargh, 2001), human aggression literature (e.g., Bandura, 1973; Berkowitz, 1993; Buss, 1961; Geen, 1999; Giacalone and Greenberg, 1997; Glick and Roose, 1993; Goldstein, 1994; Potegal and Knutson, 1994; Rizzuto *et al.*, 2004), organizational justice literature (e.g., Bies, 1987; Cropanzano, 1993), and work stress literature (e.g., Barling *et al.*, 2005).

Abusive Behaviors

A great deal of empirical research on workplace mistreatment/abuse has focused on determining the types of verbal, nonverbal, and physical behaviors that harm victims. Typical verbal forms of mistreatment/abuse include

scapegoating, threats, put-downs, false accusations, swearing, name calling, unfounded criticism (public and private), gossiping, deceit, dishonesty, favoritism, unfair evaluations, taking credit for another's accomplishments, unwarranted reprimands, unfair reassignments or terminations, and racial and sexual harassment. Nonverbal forms of mistreatment/abuse include ignoring, snubbing, aggressive eye contact (e.g., the silent treatment), staring, and physical gestures (e.g., finger pointing and foot stomping). Withholding essential resources and the destruction of property are examples of physical forms of mistreatment/abuse (Baron and Neuman, 1996; Björkvist *et al.*, 1994; Blase *et al.*, 2008; Blase and Blase, 2002, 2003a, 2003b; Glomb, 2002; Harlos and Pinder, 2000; Hoel and Cooper, 2000; Irish Taskforce on the Prevention of Workplace Bullying, 2001; Keashly and Jagatic, 2000; Leymann, 1990; Namie, 2000; Price-Spratlen, 1995; Queensland Government Workplace Bullying Taskforce, 2002; Salin, 2001; WBTI, 2003; Westhues, 1998, 2004). It is important to emphasize that, by definition, victims of workplace mistreatment/abuse are subjected to a range of the aforementioned behaviors over extended periods of time.

Effects of Abuse

Research on workplace mistreatment/abuse has also emphasized the deleterious effects of such behavior on a victim's psychological-emotional and physical-physiological well-being, work performance and relationships with colleagues, and one's personal and family life well-being. Harmful effects on psychological-emotional well-being include feelings of desperation, incompetence, shame, self-doubt, loneliness; obsessive thinking, distrust, anxiety, disorientation, shock, panic attacks, depression, and posttraumatic stress disorder (PTSD). Examples of the effects on physical-physiological well-being are hair loss, back and neck pain, headaches and migraines, significant weight changes (loss or gain), ulcers, chronic fatigue syndrome, high blood pressure, irritable bowel syndrome, and heart attacks. Effects on work performance and relationships with colleagues include work impairment (e.g., decreases in initiative, creativity, risk taking, and commitment; tardiness), absenteeism, mistakes made at work, impaired group decision making, and withdrawal from extra-role and social involvements. Negative effects on personal and family life include increases in family conflict and deterioration of relationships with spouses, children, and friends (ACTU, 2000; Björkvist *et al.*, 1994; Blase *et al.*, 2008; Blase and Blase, 2002, 2003a, 2003b; Einarsen *et al.*, 1994; Glomb, 2002; Hoel *et al.*, 2001; Hornstein, 1996; Irish Taskforce on the Prevention of Workplace Bullying, 2001; Keashly, 2001; Leymann and Gustafsson, 1996; NNLI, 1993; Pearson *et al.*, 2000; Price-Spratlen, 1995; Queensland Government Workplace Bullying Taskforce, 2002; Tepper, 2000; WBTI, 2003; Westhues, 1998, 2004).

Leader Mistreatment/Abuse of Teachers

To date, only two studies of school principal mistreatment/abuse have been published. Blase and Blase (2002, 2003a, 2003b) conducted a series of in-depth interviews with 50 exemplary teachers in the United States and Canada who had been repeatedly mistreated/abused by their school principals for substantial periods of time (6 months to 9 years). They produced descriptive and conceptual knowledge focusing on abusive principal behaviors and effects on teachers, teaching, teachers' relationships with students and colleagues, and teachers' personal life.

In a subsequent study, Blase *et al.* (2008) administered a comprehensive survey to 172 teachers from the United States and Canada. The results of this study confirmed those of their first study and provided greater detail on abusive principal behavior and its deleterious effects on teachers. Research participants had been mistreated from 1 to 3 years (40.1%) and over 3 years (25.6%).

Blase *et al.* (2008) reported that some of the most frequently occurring and most intensely harmful mistreatment behaviors were failing to recognize or praise teachers for work-related achievements, favoring other teachers, intimidation, and failure to support teachers in difficult interactions with parents and students. Other high-frequency and intensely harmful behaviors were nitpicking about time and micromanaging, unjust criticism, overloading with work, lying, isolating teachers from colleagues, gossiping to other teachers/parents/students, and unwarranted reprimands. Over 70% of the teachers studied disclosed that their principals abused them very frequently. Female teachers, as compared with male ones, experienced principal mistreatment which was more frequent as well as more intensely harmful. Blase *et al.* (2008) found that the ten most frequent effects of mistreatment (in rank order) were stress, resentment, anger, insecurity, a sense of injustice and moral outrage, self-doubt, anxiety, a sense of powerlessness, and maintenance of silence and bitterness. Over 50% of the research participants identified additional effects, including damaged relationships with colleagues, fear, decreases in self-confidence and self-esteem, reduced motivation to teach, and depression. Over 77% of teachers reported that principal mistreatment significantly undermined teaching, that is, reduced the characteristics of motivation, innovation/creativity/risk taking, and caring/patience/tolerance of students and increased the use of rigid, authoritarian, ineffective teaching methods. Of the total number of research participants, 51% indicated that they sought medical or psychological treatment (or both) for illnesses related to their mistreatment/abuse.

Blase *et al.* (2008) also found that 50% of their research participants attributed at least half of their total life harm to a principal's mistreatment and that 50% of them wanted to leave the teaching profession altogether as a result of principal mistreatment. This suggests that school principal

mistreatment of teachers should be considered one of life's major sources of harm. These scholars also described findings related to how teachers cope with principal mistreatment (i.e., teachers' predominant use of passive strategies) as well as perceived reasons (e.g., primarily political reasons as well as demographic factors such as gender and race) for mistreatment.

A Methodological Note

During the 1990s, most research on organizational abuse centered on attempts to clearly define the construct of abuse and its components. Multi-item behavioral scales determined abusive behaviors, frequencies, intensity of harm, effects, and coping strategies (see, e.g., studies with reliability and validity information, including those of Sheehan *et al.* (1990), Björkvist *et al.* (1994), Keashly *et al.* (1994), and Baron and Neuman (1996)); little research was undertaken to determine the responses of victims to abuse and factors that influence these responses (see, e.g., Kahn and Byosièrè's (1992) study of the victims' perceptions of personal and organizational resources that influence their responses to abuse).

In the past decade, researchers have begun to examine the relationship between the intensity and frequency of abuse (Blase *et al.*, 2008); the nature of harm (e.g., Blase and Blase's (2003b) qualitative descriptions of victims' experiences and Tepper's (2000) factor and regression analyses of the consequences of abuse and implications for organizational justice theory); antecedent conditions, or what contributes to the presence of abusive behaviors (e.g., highly hierarchical organization, local norms, and management style; see Glomb's (2002) qualitative interviews on antecedents to abuse and Zapf's (1999) psychometric properties of scales that determine multicausality from the organization, social system, victim, and perpetrator); and the nature of conflict escalation and coping (see Zapf and Gross's (2001) interviews and questionnaire).

Despite nearly two decades of empirical work in the field of organizational abuse, researchers are still using a limited, though slowly expanding, range of methods to investigate, define, and operationalize the construct. Moreover, research samples tend to be self-selected, and some investigators continue to use nonvalid measures of effects and responses to organizational abuse.

Further Research

To date, only two empirical studies of principal mistreatment of teachers have been published. Like most research on workplace mistreatment, both have investigated the victim's perspective on abusive behaviors, deleterious

effects, teachers' coping responses, and causes (antecedents and contributors) to mistreatment. Much more research is needed to understand each of the aforementioned dimensions of this deeply disturbing problem. Research aimed at examining the causes (i.e., antecedents, reasons, and contributors) of teacher mistreatment/abuse, using a comprehensive theoretical approach (i.e., one that subsumes the individual, dyadic, group, and organizational levels) would be particularly valuable (Rayner *et al.*, 2002). At this early stage, case studies using multiple methods would be especially useful in generating needed descriptive, conceptual, and theoretical knowledge. More research on how teachers cope with school leader mistreatment would make an important contribution, given the significance of coping to determine the degree of harm to individuals or groups and their work (Cox, 1980). In addition, large-scale national and international studies are essential to determine the frequency and pervasiveness of the principal mistreatment problem. Most research on workplace mistreatment, in a broad sense, and on existing published research on teacher mistreatment specifically, has examined the victims' perspective. Research focusing on the perpetrator's perspective is critical; however, given the sensitivity of the topic, such research will be difficult to implement. Finally, following the development of deeper understandings of the school leader mistreatment/abuse problem, research centered on both preventative measures (e.g., university preparation for educational leaders and teachers) and protective measures (e.g., development of organizational policies and laws) to ameliorate the problem would be valuable.

Internationally, the problem of workplace mistreatment/abuse has been generally ignored by organizational scholars. This is particularly true with regard to school leader mistreatment/abuse of teachers; indeed, the essential knowledge base about this problem is not yet available for enhancing school leader-/teacher-preparatory programs or challenging the *status quo* in education on a national or international level. Yamada wrote:

...contemporary attitudes toward workplace bullying parallel those toward sexual harassment several decades ago. Long before "sexual harassment" entered into everyday usage, the underlying behavior had been a staple of too many workplaces. Its targets often suffered in silence, and many paid dearly in return for keeping a job. The notion that such conduct should be illegal had little support. ...[I]n view of the growing body of evidence documenting the harm done by bullying to workers and employers, this is a measure whose time has come. (Yamada, 2004: 515–516)

See also: Ethical/Moral Issues in Educational Leadership; Leadership: Democratic; New Approaches in Preparing School Leaders; School Leadership Standards; Understanding How Leadership Influences Student Learning.

Bibliography

- ACTU (Australian Council of Trade Unions) (2000). Being bossed around is bad for your health. http://www.actu.asn.au/public/ohs/bullying/1063694361_23162.html (accessed April 2009).
- Adams, G. B. and Balfour, D. L. (1998). *Unmasking Administrative Evil*. Thousand Oaks, CA: Sage.
- Ashforth, B. (1994). Petty tyranny in organizations. *Human Relations* 47(7), 755–778.
- Ashforth, B. E. (1997). Petty tyranny in organizations: A preliminary examination of antecedents and consequences. *Canadian Journal of Administrative Sciences* 14(2), 126–140.
- Bandura, A. (1973). *Aggression: A Social Learning Analysis*. Englewood Cliffs, NJ: Prentice Hall.
- Barling, J., Kelloway, E. K., and Frone, M. (eds.) (2005). *Handbook of Work Stress*. Thousand Oaks, CA: Sage.
- Baron, R. A. and Neuman, J. H. (1996). Workplace violence and workplace aggression: Evidence on their relative frequency and potential causes. *Aggressive Behavior* 22, 161–173.
- Baumeister, R. F. (1996). *Evil: Inside Human Cruelty and Violence*. New York: W.H. Freeman.
- Berkowitz, L. (1993). *Aggression: Its Causes, Consequences, and Control*. New York: McGraw-Hill.
- Bies, R. J. (1987). The predicament of injustice: The management of social outrage. In Cummings, L. L. and Staw, B. M. (eds.) *Research in Organizational Behavior*, vol. 9, pp 289–320. Greenwich, CT: JAI Press.
- Björkvist, K. (1992). Trakassering forekommer bland anställda vid AA ('Harassment exists among employees at Abo Academy'). *Meddelanden från Abo Akademi* 9, 14–17.
- Björkvist, K., Österman, K., and Hjelt-Bäck, M. (1994). Aggression among university employees. *Aggressive Behavior* 20, 173–184.
- Blase, J. and Blase, J. (2002). The dark side of leadership: Teacher perspectives of principal mistreatment. *Education Administration Quarterly* 38(5), 671–727.
- Blase, J. and Blase, J. (2003a). Breaking the Silence: Overcoming the Problem of Principal Mistreatment of Teachers. Thousand Oaks, CA: Corwin.
- Blase, J. and Blase, J. (2003b). The phenomenology of principal mistreatment: Teachers' perspectives. *Journal of Educational Administration* 41(4), 367–422.
- Blase, J., Blase, J., and Du, F. (2006). The national study of school administrators' mistreatment of teachers. *Paper Presented at the Annual Meeting of the University Council for Educational Administration*. San Antonio, TX, November.
- Blase, J., Blase, J., and Du, F. (2008). The mistreatment of teachers: A national study. *Journal of Educational Administration* 46(3), 263–301.
- Buss, A. H. (1961). *The Psychology of Aggression*. New York: John Wiley.
- Cox, T. (1980). *Stress*. Baltimore, MD: University Park Press.
- Cropanzano, R. (1993). *Justice in the Workplace: Approaching Fairness in Human Resource Management*. Hillsdale, NJ: Erlbaum.
- Einarsen, S., Hoel, H., Zapf, D., and Cooper, C. L. (eds.) (2003). *Bullying and Emotional Abuse in the Workplace: International Perspectives in Research and Practice*. London: Taylor and Francis.
- Einarsen, S., Raknes, B. I., Matthiesen, S. B., and Hellesoy, O. H. (1994). *Mobbing og harde personkonflikter (Bullying and Harsh Interpersonal Conflicts)*. Bergen: Sigma Forlag.
- Einarsen, S. and Skogstad, A. (1996). Bullying at work: Epidemiological findings in public and private organizations. *European Journal of Work and Organizational Psychology* 5(2), 185–201.
- Folger, R. (1993). Reactions to mistreatment at work. In Murningham, J. K. (ed.) *Social Psychology in Organizations: Advances in Theory and Research*, pp 161–183. Englewood Cliffs, NJ: Prentice Hall.
- Geen, R. (1999). *Human Aggression*. Buckingham, UK: Open University Press.
- Giacolone, R. A. and Greenberg, J. (eds.) (1997). *Antisocial Behavior in Organizations*. Thousand Oaks, CA: Sage.
- Glasl, F. (1994). *Conflict Management: A Handbook for Managers and Consultants*, 4th edn. Bern, Switzerland: Haupt.
- Glass, D. C. (1977). *Behavior Patterns, Stress, and Coronary Disease*. Hillsdale, NJ: Erlbaum.
- Glick, R. A. and Roose, S. P. (eds.) (1993). *Rage, Power, and Aggression*. New Haven, CT: Yale University Press.
- Glomb, T. M. (2002). Workplace anger and aggression: Informing conceptual models with data from specific encounters. *Journal of Occupational Health Psychology* 7(1), 20–36.
- Goldstein, A. (1994). *The Ecology of Aggression*. New York: Plenum.
- Haney, C., Banks, C., and Zimbardo, P. (1973). Interpersonal dynamics in a simulated prison. *International Journal of Criminology and Penology* 1, 69–97.
- Harlos, K. P. and Pinder, C. C. (2000). Emotion and injustice in the workplace. In Fineman, S. (ed.) *Emotion in Organizations*, 2nd edn., pp 255–276. Thousand Oaks, CA: Sage.
- Hoel, H. and Cooper, C. L. (2000). Destructive conflict and bullying at work. Unpublished Report, UMIST, UK, November.
- Hoel, H., Cooper, C., and Faragher, B. (2001). The experience of bullying in Great Britain: The impact of organizational status. *European Journal of Work and Organizational Psychology* 10(4), 443–465.
- Hoel, H., Rayner, C., and Cooper, C. L. (1999). Workplace bullying. In Cooper, C. L. and Robertson, I. T. (eds.) *International Review of Industrial and Organizational Psychology*, vol. 14, pp 195–230. Chichester: Wiley.
- Hodgkinson, C. (1991). *Educational Leadership: The Moral Art*. Albany: State University of New York Press.
- Holmes, D. S. and Will, M. J. (1985). Expression of interpersonal aggression by angered and non-angered persons with Type A and Type B behavior patterns. *Journal of Personality and Social Psychology* 40, 723–727.
- Hornstein, H. A. (1996). *Brutal Bosses and Their Prey*. New York: Riverhead Books.
- Hornstein, H. A., Michela, J. L., Van Eron, A. M., et al. (1995). Disrespectful supervisory behavior: Effects on some aspects of subordinates' mental health. Unpublished manuscript, Teachers College, Columbia University.
- Irish Taskforce on the Prevention of Workplace Bullying (2001). *Dignity at Work: The Challenge of Workplace Bullying. Report of the Taskforce on the Prevention of Workplace Bullying*. Dublin: Stationery Office.
- Kahn, R. L. and Byosiè, P. (1992). Stress in organizations. In Dunnette, M. D. and Hough, L. M. (eds.) *Handbook of Industrial and Organization Psychology*, 2nd edn., vol. 3, pp 571–650. Palo Alto, CA: Consulting Psychologists Press.
- Katz, F. E. (1993). *Ordinary People and Extraordinary Evil: A Report on the Beguilings of Evil*. Albany, NY: State University of New York.
- Keashly, L. (1998). Emotional abuse in the workplace: Conceptual and empirical issues. *Journal of Emotional Abuse* 1(1), 85–117.
- Keashly, L. (2001). Interpersonal and systemic aspects of emotional abuse at work: The target's perspective. *Violence and Victims* 16(3), 233–268.
- Keashly, L. and Jagatic, K. (2000). The nature, extent, and impact of emotional abuse in the workplace: Results of a statewide survey. *Paper Presented at the Annual Conference of the Academy of Management*. Toronto, Canada.
- Keashly, L., Trott, V., and MacLean, L. M. (1994). Abusive behavior in the workplace: A preliminary investigation. *Violence and Victims* 9(4), 341–357.
- Kekes, J. (1990). *Facing Evil*. Princeton, NJ: Princeton University Press.
- Kets de Vries, M. F. R. (1989). *Prisoners of Leadership*. New York: Wiley.
- Kets de Vries, M. (1993). *Leaders, Fools, and Imposters: Essays on the Psychology of Leadership*. San Francisco, CA: Jossey-Bass.
- Kipnis, D. (1972). Does power corrupt? *Journal of Personality and Social Psychology* 24(1), 33–41.
- Kipnis, D. (2001). Using power: Newton's second law. In Lee-Chai, A. Y. and Bargh, J. A. (eds.) *The Use and Abuse of Power: Multiple Perspectives on the Cause of Corruption*, pp 3–18. Philadelphia, PA: Psychology Press.
- Lee-Chai, A. Y. and Bargh, J. A. (2001). *The Use and Abuse of Power: Multiple Perspectives on the Causes of Corruption*. Philadelphia, PA: Psychology Press.

- Leymann, H. (1990). Mobbing and psychological terror at workplaces. *Violence and Victims* 5(2), 119–126.
- Leymann, H. and Gustafsson, A. (1996). Mobbing at work and the development of post-traumatic stress disorders. *European Journal of Work and Organizational Psychology* 5, 251–275.
- Lynch, J. (2002). *Workplace Bullying: Implications for Police Organizations*. Australia: Australasian Center for Policing Research.
- Matthiesen, S. B., Raknes, B. I., and Rokkum, O. (1989). Mobbing på arbeidsplassen ('Bullying in the workplace'). *Tidsskrift for Norsk Psykologforening* 26, 761–774.
- McGregor, D. (1960). *The Human Side of Enterprise*. New York: McGraw-Hill.
- Miller, S. M., Lack, E. R., and Asroff, S. (1985). Preferences for control and the coronary-prone behavior pattern: I'd rather do it myself. *Journal of Personality and Social Psychology* 49, 492–499.
- Namie, G. (2000). *U.S. Hostile Workplace Survey 2000*. Benicia, CA: Campaign Against Workplace Bullying.
- Neuman, J. and Baron, R. (1998). Workplace violence and workplace aggression: Evidence concerning specific forms, potential causes, and preferred targets. *Journal of Management* 24(3), 391–419.
- NNLI (Northwestern National Life Insurance Company) (1993). *Fear and Violence in the Workplace*. Minneapolis, MN: Author.
- Office of the Employee Ombudsman (2000). *Bullies not Wanted: Recognising and Eliminating Bullying in the Workplace*. Adelaide: Office of the Employee Ombudsman of South Australia.
- Pearson, C., Andersson, L., and Porath, C. (2000). Assessing and attacking workplace incivility. *Organizational Dynamics* 28, 129–137.
- Potegal, M. and Knutson, J. F. (1994). *The Dynamics of Aggression: Biological and Social Processes in Dyads and Groups*. Hillsdale, NJ: Erlbaum.
- Price-Spratlen, L. (1995). Interpersonal conflict which includes mistreatment in a university workplace. *Violence and Victims* 10(4), 285–297.
- Queensland Government Department of Industrial Relations (2004). *Workplace Health and Safety Act 1995: Prevention of Workplace Harassment Advisory Standard, 2004*. Brisbane, Australia: Queensland Government Department of Industrial Relations.
- Queensland Government Workplace Bullying Taskforce (2002). *Report of the Queensland Government Workplace Bullying Taskforce*. Brisbane, QLD: Queensland Government Department of Industrial Relations.
- Rayner, C. (2000). *Bullying and Harassment at Work: Summary of Findings*. Staffordshire, UK: Staffordshire University Business School.
- Rayner, C., Hoel, H., and Cooper, C. L. (2002). *Workplace Bullying: What We Know, Who Is to Blame, and What Can We Do?* London: Taylor and Francis.
- Rizzuto, A., Meissner, S. J., and Buie, D. H. (2004). *The Dynamics of Human Aggression: Theoretical Foundations, Clinical Applications*. New York: Brunner-Routledge.
- Robinson, S. L. and Bennett, R. J. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal* 38(2), 555–572.
- Ryan, K. D. and Oestrich, D. K. (1991). *Driving Fear out of the Workplace: How to Overcome the Invisible Barriers to Quality, Productivity, and Innovation*. San Francisco, CA: Jossey-Bass.
- Ryan, K. D. and Oestrich, D. K. (1998). *Driving Fear out of the Workplace*, 2nd edn. San Francisco, CA: Jossey-Bass.
- Salin, D. (2001). Prevalence and forms of bullying among business professionals: A comparison of two different strategies for measuring bullying. *European Journal of Work and Organizational Psychology* 10(4), 425–441.
- Sheehan, K. H., Sheehan, D. V., White, K., Leibowitz, A., and Baldwin, D. C. (1990). A pilot study of medical student abuse. *Journal of the American Medical Association* 263(4), 533–537.
- Tepper, B. J. (2000). Consequences of abusive supervision. *Academy of Management Journal* 43(2), 178–190.
- WBTI (The Workplace Bullying and Trauma Institute) (2003). The WBTI 2003 report on abusive workplaces. <http://bullyinginstitute.org/res2003B.html> (accessed April 2009).
- Westhues, K. W. (1998). *Eliminating Professors: A Guide to the Dismissal Process*. Lewiston, NY: Edwin Mellen.
- Westhues, K. W. (2004). *Workplace Mobbing in Academe: Reports from Twenty Universities*. Lampeter, UK: Edwin Mellen.
- Yamada, D. C. (2000). The phenomenon of "workplace bullying" and the need for status-blind hostile work environment protection. *Georgetown Law Journal* 88(3), 477–536.
- Yamada, D. C. (2004). Crafting a legislative response to workplace bullying. *Employee Rights and Employment Policy Journal* 8(2), 475–521.
- Zapf, D. (1999). Mobbing in organizations: Ein überblick zum stand der forschung. ('Mobbing in organizations: A state of the art review'). *Zeitschrift für Arbeits- und Organisations psychologie* 43, 1–25.
- Zapf, D. and Gross, C. (2001). Conflict escalation and coping with workplace bullying: A replication and extension. *European Journal of Work and Organizational Psychology* 10(4), 497–522.

Further Reading

Hodgkinson, C. (1991). *Educational Leadership: The Moral Art*. Albany, NY: State University of New York Press.

Relevant Websites

<http://www.edethics.org> – American Society for Ethics in Education.

<http://www.actu.labor.net.au> – Australian Council of Trade Unions.

<http://www.beyondbullying.co.nz> – Beyond Bullying Association.

<http://www.bullybusters.org> – Bully Busters Stuff.

<http://www.bullyonline.org> – Bully Online.

<http://members.shaw.ca/mobbing/mobbingCA/index.htm> – Mobbing, bullying, and workplace harassment resources.

<http://www.endteacherabuse.org> – National Association for Prevention of Teacher Abuse.

<http://www.nobullyforme.ca> – No Bully for Me.

<http://www.workplacebullying.co.uk> – The Workplace Bullying Site: Employment Law.

<http://www.bullyinginstitute.org> – Workplace Bullying Institute.

<http://www.worktrauma.org> – Workplace Dignity Institute.

Leadership and Gender

J Blackmore, Deakin University, Burwood, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

It is well recognized that conceptualizing leadership is problematic, and that much of the leadership literature is Western centric (Sander, 1993). Gender has also become an ongoing issue in leadership studies as well as in educational policy and practice. Politically, the lack of women and dominance of men in formal leadership positions are troubling for policymakers: symbolically, in democratic societies that seek to be seen to be representative and fair, and, practically, in terms of recruiting and retaining effective leaders in a numerically feminized profession of education. Epistemologically, how the gender and leadership nexus is understood and explained is reliant, first, upon how the problem has been defined and, second, the theoretical and conceptual frames that are utilized to analyze the problem.

Framing the Woman Problem

Two traditions exist within the research and policy literature on gender and leadership that frame research and policy – one in which gender is treated as a category of analysis or a factor to be considered within mainstream educational leadership studies and the second that is a critical strand of theory and research informed by feminist and, later, postcolonial theory that argues that leadership is shaped by wider cultural, social, and political belief systems, sociocultural practices, and discourses that are gendered, each with different policy trajectories (Marshall, 1997).

Mainstream Studies

To understand how gender and leadership has been constructed discursively within the field, as a research issue and a policy problem, it is necessary to understand the emergence of administration as a disciplinary field in education. Scholars and practitioners in educational administration in the early twentieth century sought legitimacy, initially by aligning with business theories of scientific management, and, subsequently, post-1945, with a narrow positivist tradition of science. Both perspectives assumed the Enlightenment binaries representing that which was rational, objective, and universal as masculine, and that which was subjective, emotional, and particular as feminine, further embedding legacies of early twentieth-century psychological and biological theories essentializing women, depicting them deficient as leaders because of their lack of the necessary physical, mental, moral, and emotional attributes (Shakeshaft, 1987; Blackmore, 1989).

This naturalization of the male/female division of labor was reinforced by existing patriarchal social, political, and economic structures and norms delineating the public, work, and leadership as masculine domains and the private, family, and teaching as feminine domains.

The structural functionalist sociology of the 1950s theory movement that dominated well into the 1960s institutionalized these binaries with epistemological claims that fact was separate from value and that the scientific method was objective and the nonscientific method subjective, while simultaneously claiming a universal gender-neutral subject and epistemology. Within this frame, gender was treated as a subcategory, a pattern emerging from large statistical analyses that identified sex differences in behaviors of leaders and outcomes of schooling, or as a variable to be controlled to measure the effects of other variables such as class (e.g., Jones, 1990). The theories of sex differentiation and socialization of this time relied on sociological and psychological studies of male leaders that largely ignored organizational and cultural contexts. Both the theories and the practices of educational administration were largely premised upon hierarchical organizational structures, whether bureaucratic or corporate, equating leadership with formal position and/or individual attributes, thus constructing binaries between leadership and the management, and leaders and followers. Such perspectives treated leadership as a set of generic competencies acquired and required to manage organizational and educational change to produce more efficient and effective schools, including definitions of merit that favored the experiences of men over those of women.

The mainstream tradition during the 1970s and 1980s increasingly drew on guru change management, human resource management, and psychological and economic literatures, but less so on sociological and political literatures or on the teachers' work. Administration was being constructed as different from teaching. Leadership research within this paradigm continued to generalize from the experience of men to women, with the leadership norm being an idealized male type: rational, autonomous, unemotional, and with higher-level intellectual and moral attributes. Both theoretical and social norms justified, and were justified by, the naturalness of the gendered division of educational labor between leading and teaching.

Mainstream research shifted its focus from the individual attributes of leaders to consider the role of context, culture, and community in managing organizational change during the 1980s, and, subsequently, on how to lead learning organizations in the context of emerging knowledge-based

economies and increasing cultural diversity during the 1990s. Whereas historically, leadership featured as a subset of educational management and administration, it now came to subsume administration as the lexicon of reform. The effective schools literature equated the successful principal to the effective school and, in an era of school self-management, the key solution to neoliberal educational reform. Within this context, and under pressure from equal opportunity legislation, policies, and programs, women were moving into middle management in schools. However, again, the dominant school effectiveness, school improvement, and learning organization paradigms ignored gender by either assuming the self-maximizing and autonomous gender-, race-, and class-neutral individual or by treating women as a unitary category who had specific needs, wants, and interests. Discourses produced from feminist research about women's styles of leading were readily and unproblematically appropriated into the mainstream, producing incremental changes (e.g., Celikten, 2005). In the early twenty-first century, emerging research fields, such as internationalization, address cultural diversity as an issue, and either ignore gender or treat it as an issue for other cultures.

Gender within mainstream educational administration has therefore become a factor to be addressed in policy, for example, teacher career paths restructured to attract more women. In general, the tendency is to position gender as particularist, largely irrelevant to the mainstream, often cordoned off in texts in sections categorized under alternative perspectives or equity, or largely ignored (as in the *Second International Handbook of Leadership and Administration* published in 2006). Mainstream leadership research on gender therefore falls into four categories:

- that which ignores gender by universalizing the experience of men;
- that which treats gender as a variable to be explained in studies focusing on other issues (e.g., lack of aspirants for principalship);
- that which draws on feminist (and critical) theory without recognition of the politics of its origins, for example, collapse of emotional/rational binary and transformational leadership; and
- that which treats gender as a problem for/with women.

All such approaches neglect wider structural and cultural factors as well as the gender order that is socially, politically, and economically constructed.

Feminist Studies

The mainstream social theory was challenged epistemologically and politically by the women's, civil rights, and student social movements during the late 1960s. However, such disruption failed to penetrate mainstream educational

and administration literature until the 1980s (Weiler, 1993). While feminist theorists initially drew from the sex-role socialization theory, increasingly, gender and the production of knowledge came to be understood within the field of leadership (as in sociology, philosophy, psychology, and history) to be socially constructed and about relations of power. Leadership (as feminism) was therefore a political and personal issue – how one came to be, to know, and to act, and with what intent. Initially, feminist researchers identified, mapped, and sought to explain the historical patterns of women's unequal representation in leadership (Shakeshaft (1987), Yeakey *et al.* (1986), Marshall and Ortiz (1988) for the USA; Gaskell and McLaren (1991) for Canada; Blackmore (1989) for Australia; Al-Khalifa (1989), Acker (1989), Grant (1989), Deem (1990) for the UK; and Nixon (1987) and Court (1998) for New Zealand). Researching women leader's experiences provided a knowledge base and the missing empirical evidence tracking gender inequality. Working within the individualistic framework of liberal feminism (Schmuck, 1996; Adler *et al.*, 1993) and the traditional notions of professionalism (Glazer, 1991; Strachan, 1993), the focus was on women's continuing underrepresentation in leadership, nurturing women's aspirations, skills, and confidence to take up leadership, and procedural and legalistic approaches to provide access for individual women into leadership through equal opportunity (in Australia and New Zealand) or affirmative action (in the USA) legislation.

Interpretivist traditions of the new sociology of knowledge, and radical and cultural feminist theories that focused on women's ways of doing, being, and thinking emerged during the 1980s. These shifted the focus from the individual to the collective, and from women's lack to what women collectively could contribute to leadership in terms of particular moral positions, an ethics of care, a democratic disposition, and a focus on student learning, while collapsing the public/private binary (Noddings, 1992; Beck, 1994; Marshall and Anderson, 1995). As with the mainstream management theory, this focus on organizational culture explained not only collective resistance to educational reform, but also how women continued to be excluded from masculinist leadership cultures (Blackmore, 1993). While mainstream literature treated culture as homogeneous and static (Hall, 1996), feminist literature (as critical theorists drawing on Habermas) viewed organizational culture as contested in terms of values, belief systems, and practices (du Billing and Alvesson, 1989). Cultural theories also shifted focus onto the social relations of gender and how different masculinities (bureaucratic and entrepreneurial) and femininities (caring and sharing) were constructed in relation to each within a wider gender order, explicating how hegemonic masculinities reinvented themselves over time to subordinate other masculinities (e.g., homosexual) and most femininities (Connell, 1987).

Context and culture were thus seen to simultaneously frame societal expectations and perceptions of women leaders as the other (Ozga, 1993).

Radical (cultural) feminism sat comfortably within the wider politics of recognition of the 1980s in which women, as other marginalized groups, sought both voice and recognition within pluralist Western societies by asserting the positive aspects of different cultures and value systems of specific social groups (Capper, 1993; Marshall, 1993; Dunlap and Schmuck, 1996). However, women leaders in education were repositioned ambiguously – either as outsiders – inside dominant organizational cultures and, therefore, as good change agents, or as complementing and softening the hard-nosed masculine attributes of reason and decision making through women's styles of leadership. Both positionings reproduced the essentializing binaries of rationality/care and hierarchy/democracy. Neither stance challenged the mainstream theory or practice because such approaches were readily appropriated as a complementary add-on (Blackmore, 1999a; Marshall *et al.*, 1996).

The linguistic turn toward poststructuralism in the 1990s synchronized with black and postcolonial feminists contesting the dominance of white middle class feminist perspectives in theory and practice (Sleeter, 1993; Tuhiwa-Smith, 1993). Leadership identity formation now came to be understood in terms of multiple subjectivities discursively constituted, in a constant state of formation, being, and becoming (Mirza, 1993; Dillard, 1995). Women leaders were now seen to have agency and be positioned, but also positioning themselves, within a range of discursive constraints and possibilities as educational change was seen to be uneven, unpredictable, irrational, and emotionally charged (Blackmore, 1995; Boler, 1999). This perspective foregrounded, for women leaders, their feelings of contradiction and ambivalence; collapsed false binaries between reason and emotion; recognized how women leaders were positioned in contradictory ways as women and as leaders according to their gender, race, language, ethnicity, and sexuality (Chase, 1995; Blount, 1994; Skrla and Scheurich, 2000); and challenged the authenticity of mainstream theories and knowledge base in educational administration and leadership (Shakeshaft, 1995; Ikpa, 1995; Ortiz and Ortiz, 1995).

Paradoxically, this poststructuralist focus on identity diverted attention from wider historical and materialist analyses at the time that relations between the individual and the state in education with neoliberal restructuring in both Anglophone and developing nation states were being fundamentally transformed (Blackmore, 1999b; Stromquist and Monkman, 2000). Women became the new source of, and resource for, leadership in the neoliberal restructuring of education in the 1990s as teaching became more feminized and casualized internationally (Goldring and Chen, 1994; Court, 1998). Research indicates how women leaders,

particularly feminists, have been ambiguously situated by the new contractualism of marketized and managerialized school systems, producing significant ethical dilemmas as often, but not always, women's preference due to their experience is for more democratic and collegial rather than contractual and entrepreneurial relationships (Chase, 1995; Blackmore, 1999a; Grogan, 1999; Reynolds, 2002; Young, 2003).

The first decade of the twenty-first century now sees schools confronted with multiple demands: greater cultural and linguistic diversity; the changing social relations of gender and families as more women enter paid work; greater extremes of poverty and wealth; increased student mobility with flows of refugees, migrants, and third-culture kids; and the emerging effects of internationalization of schooling on local provision (Grogan, 2002; Blackmore and Sachs, 2004). In particular, the voices of indigenous feminists (Battiste, 2005; Fitzgerald, 2006; Ahnee-Benham, 2003) and marginalized women from religious and linguistic minorities in the Anglophone nation states (Mendez-Morse, 2003; Shah, 2006) are challenging white middle class women and men in leadership to consider the privilege and advantage that accompanies their whiteness. As black, Muslim, or indigenous women leaders, they are expected to not only negotiate the complexity of localized gender, race, and class politics within their cultural groups and communities, but also mediate between cultures transnationally and intranationally (Ngururwutthun and Stewart, 1996). Their alternative histories, epistemologies, and experiences of, and in, educational leadership confront mainstream administrative theory and practice because many claims reach beyond (and before) the nation state, and dominant (recently reinvigorated) scientific paradigms, as well as outside schools and systems as organizations, drawing on community-based activism and social movements (Mendez-Morse, 2003).

Feminist theorists now struggle with how globalization impacts women, educational management, and leadership at the macro level (Stromquist and Monkman, 2000; Blackmore, 1999b), while addressing the ways in which culture, race, gender, and class interplay within local community and organizational politics in terms of who leads, how women lead, and with what effect. Culture, nationhood, citizenship, and religion are being foregrounded within the feminist literature (e.g., Oplatka and Hertz-Lazarowitz, 2006). In this context, feminist research on leadership is best depicted now as a series of interconnected or transnational inquiries with a familial resemblance, articulated differently within particular cultural, historical, and material contexts in terms of the relationship between education, the individual, the state, other social movements, and equity discourses and strategies (Stromquist and Monkman, 2000; Grogan, 2002). The paradox is that as education becomes critical in terms of promoting human rights internationally

along with social justice for girls and women, and for improving individual and collective life chances, there is significant evidence that the number of women in formal leadership in the Anglophone nation states is on the decline.

Reframing Gender and Leadership

Many feminists have argued that the focus on leadership is itself the problem for research, policy, and practice around the issue of gender and leadership. Treating leadership as the solution and equating it to a formal position individualize what is a collective practice. Research evidence and theories of distributive leadership illustrate that effective organizational change in complex network societies and learning organizations requires multiple modes of leadership across multiple sites, in particular teacher leadership. However, regardless of the type of leadership, feminist theories foreground the purpose of leadership as social justice and equity (Grogan, 2002).

Significantly, the issue of gender and leadership is still constructed as a woman's problem. Focusing on women and leadership fails to recognize how educational organizations produce gender identities and are structured by the wider social relations of gender in culturally specific ways (Aailto and Mills, 2002; Oplatka and Hertz-Laraeowitz, 2006). There is little research that problematizes the ongoing dominance of men in educational leadership from the perspective of the new sociology of masculinity (Lingard and Limerick, 1995; Lingard and Douglas, 1999). A relational sociology of gender would question how different masculinities and femininities are produced in relation to each other from a range of subject positions inflected by class, race, and ethnicity, and question the enduring nature of white masculinism.

Furthermore, class and race, as religion and culture, fragment, fracture, and shore up dominant social relations of gender. The critical organizational theory argues that organizational structures, processes, and cultures are socially constructed to advantage some and not others, and are not gender, race or class neutral. Gender, race, and ethnicity are not variables, problems, or issues, but integral in structuring effective leadership and organizations (Hearn and Parkin, 1992; Aailto and Mills, 2000).

The gender and leadership question is symptomatic not only of the lack of diversity in leadership generally with regard to race, class, and culture, but also of the difference in terms of values and politics (Blackmore, 2006). Women gaining a critical mass in school leadership will not necessarily change values and practices. First, women as a group, similar to men, take different political and value stances (Weiner, 1995). Second, it is generally underestimated how context (confluence of political economy, governance, cultural practices, and organizational culture) shapes the conditions of possibility of,

and for, democratic/inclusionary/emancipatory leadership practices. Democratic nations have to consider leadership not only for social justice (Grogan, 1999), but also for diversifying images, practices, and theories of leadership and management. Western education systems currently face increased disengagement of teachers with the principalship not only because of the complexity and overwhelming nature of the job, but also because women, as have men in the past, no longer view teaching as a lifetime career option but more as an episode in their career portfolio with more lucrative and family-friendly career options available. Thus, diversity in leadership requires wider systemic shifts in governance and value systems (Blackmore, 2006). Yet, research and policy continue to focus on women's underrepresentation in leadership, when they should be encouraging diversity in leadership to meet the challenges of democratic, globalized, cosmopolitan societies.

Bibliography

- Aailto, I. and Mills, A. (eds.) (2002). *Gender, Identity and Culture in Organisations*. London: Routledge.
- Acker, S. (1989). *Teachers, Gender and Careers*. Lewes: Falmer Press.
- Adler, S., Laney, J., and Packer, M. (1993). *Managing Women. Feminism and Power in Educational Management*. Milton Keynes: Open University Press.
- AhNee-Benham, M. (2003). 'In our mothers' voice: A native women's knowing of leadership'. In Young, M. D. and Skrla, L. (eds.) *Reconsidering: Feminist Research in Educational Leadership*, pp 161–178. New York: SUNY Press.
- Al-Khalifa, E. (1989). Management by halves: Women teachers and school management. In De Lyons, H. (ed.) *Women Teachers*, pp 83–96. Milton Keynes: Open University Press.
- Battiste, M. (2005). Leadership and aboriginal education in contemporary education: Narratives of cognitive imperialism reconciling with decolonisation. In Collard, J. and Reynolds, C. (eds.) *Leadership, Gender and Culture in Education, Male and Female Perspectives*, pp 150–156. Buckingham: Open University Press.
- Beck, L. (1994). *Reclaiming Educational Administration as a Caring Profession*. New York: Teachers College Press.
- Blackmore, J. (1989). Educational leadership: A feminist critique and reconstruction. In Smyth, J. (ed.) *Critical Perspectives on Educational Leadership*. Lewes: Falmer Press.
- Blackmore, J. (1993). 'In the shadow of men': Exclusionary theory and discriminatory practice in the historical construction of 'masculinist' administrative cultures. In Blackmore, J. and Kenway, J. (eds.) *Gender Matters in the Theory and Practice of Educational Administration and Policy: A Feminist Introduction*, pp 27–48. Sussex: Falmer Press.
- Blackmore, J. (1995). Doing emotional labor in the educational market place: Stories from the field of women in management. *Discourse* 17 (3), 337–350.
- Blackmore, J. (1999a). *Troubling Women. Feminism, Leadership and Educational Change*. Buckingham: Open University Press.
- Blackmore, J. (1999b). Localization/globalization and the midwife state: Strategic dilemmas for state feminism in education? *Journal of Education Policy* 14(1), 33–54.
- Blackmore, J. (2006). Deconstructing diversity discourses in the field of educational management and leadership. *Leadership, Educational Management and Administration* 34(2), 188–199.
- Blackmore, J. and Sachs, J. (2007). *Performing and Reforming Leaders. Gender, educational restructuring and organizational change*. New York: SUNY Press.
- Blount, J. (1994). One postmodernist perspective on educational leadership. Ain't I a leader? In Maxcy, S. (ed.) *Postmodern School Leadership: Meeting the Crisis in Educational Administration*, pp 47–60. London: Praeger.

- Boler, M. (1999). *Feeling Power: Emotions and Education*. London: Routledge.
- Capper, C. (ed.) (1993). *Educational Administration in a Pluralistic Society*. New York: SUNY Press.
- Celikten, M. (2005). A perspective on women principals in Turkey. *International Journal of Leadership in Education* 8(3), 207–223.
- Chase, S. (1995). *Ambiguous Empowerment: The Work Narratives of Women School Superintendents*. Amherst, MA: Massachusetts University Press.
- Connell, R. W. (1987). *Gender and Power*. Sydney: Allen and Unwin.
- Court, M. (1998). Women challenging managerialism: Devolution dilemmas in the establishment of co-principalships in Aotearo/New Zealand. *School Leadership and Management* 18(1), 35–57.
- Deem, R. (1990). Governing by gender – the new school governing bodies. In Abbott, P. and Wallace, C. (eds.) *Gender, Sexuality and Power*, pp 135–156. London: Macmillan.
- Dillard, C. (1995). Leading with her life. An African American feminist for an urban school principals. *Educational Administration Quarterly* 31(4), 539–563.
- du Billing, Y. and Alvesson, M. (1989). Four ways of looking at women and leadership. *Scandinavian Journal of Management* 5(1), 63–80.
- Dunlap, D. and Schmuck, P. (eds.) (1996). *Women Leading in Education*. New York: SUNY Press.
- Fitzgerald, T. (2006). Walking between two worlds. Indigenous women and leadership. *Educational Management, Administration and Leadership* 34(2), 201–213.
- Gaskell, J. and McLaren, A. (1991). *Women and Education: A Canadian Perspective*. Calgary, AB: Detselig.
- Glazer, J. S. (1991). Feminism and professionalism in teaching and educational administration. *Educational Administration Quarterly* 27(3), 321–342.
- Goldring, E. and Chen, M. (1994). The feminisation of the principalship in Israel: The trade-off between political power and cooperative leadership. In Marshall, C. (ed.) *The New Politics of Race and Gender*, pp 175–182. London: Falmer Press.
- Grant, R. (1989). Heading for the top: Career experiences of women. *Gender and Education* 1(2), 113–125.
- Grogan, M. (1999). Equity/equality issues. *Education Administration Quarterly* 35(4), 518–536.
- Grogan, M. (2002). *Special Issue: Social Justice and Leadership. Journal of School Leadership* 12.
- Hall, V. (1996). *Dancing on the Ceiling, a Study of Woman Managers in Britain*. London: Paul Chapman.
- Hearn, J. and Parkin, W. (1992). Women, men, and leadership: A critical review of assumptions, practices and change in industrialised nations. In Adler, N. and Israeli, D. (eds.) *Women in Management World Wide*, pp 17–40. New York: M. Saharpe.
- Ikpa, V. (1995). Gender, race, ethnicity and the quest for a knowledge base in educational administration. In Donmoyer, R., Imber, M., and Scheurich, J. (eds.) *The Knowledge Base in Educational Administration. Multiple Perspectives*, pp 174–181. New York: SUNY Press.
- Jones, B. K. (1990). The gender difference hypothesis: A synthesis of research findings. *Educational Administration Quarterly* 26, 5–37.
- Limerick, B. and Lingard, B. (eds.) (1995). *Gender and Changing Educational Management*. Sydney, NSW: Hodder and Stoughton.
- Lingard, B. and Douglas, P. (1999). *Men Engaging Feminisms: Pro-Feminism, Backlashes and Schooling*. Buckingham: Open University Press.
- Marshall, C. (ed.) (1993). *The New Politics of Race and Gender*. London: Falmer Press.
- Marshall, C. (ed.) (1997). *Feminist Critical Policy Analysis*, vols. 1 and 2. London: Falmer Press.
- Marshall, C. and Anderson, G. (1995). Rethinking the public and private spheres: Feminist and cultural studies perspectives on the politics of education. In Scribner, J. and Layton, D. (eds.) *The Study of Educational Politics*, pp 169–184. London: Falmer Press.
- Marshall, C. and Ortiz, F. (1988). Women in educational administration. In Boyan, N. (ed.) *Handbook of Research in Educational Administration*, pp 123–142. New York: Longman.
- Marshall, C., Patterson, J., Rogers, D., and Steele, J. (1996). Caring as career: An alternative perspective for educational administration. *Educational Administration Quarterly* 32(2), 271–294.
- Mendez-Morse, S. (2003). Chicana feminism and educational leadership. In Young, M. D. and Skrla, L. (eds.) *Reconsidering: Feminist Research in Educational Leadership*, pp 161–178. New York: SUNY Press.
- Mirza, H. (1993). The social construction of black womanhood in British educational research: Towards a new understanding. In Arnot, M. and Weiler, K. (eds.) *Feminism and Social Justice*, pp 32–57. London: Falmer Press.
- Ngurruwutthun, N. and Stewart, M. P. A. (1996). Learning to walk behind; learning to walk in front. A case study of the mentor program at Yirkala community education centre. *Unicorn* 22(4), 3–23.
- Nixon, M. (1987). Few women in school administration: Some explanations. *Journal of Educational Thought* 21(2), 63–70.
- Noddings, N. (1992). *The Challenge to Care in Schools: An Alternative Approach to Education*. New York: Teachers College Press.
- Oplatka, I. and Hertz-Lararowitz, R. (eds.) (2006). *Women Principals in a Multicultural Society*. Rotterdam: Sense Publishers.
- Ortiz, F. and Ortiz, D. (1995). How gender and ethnicity interact in the practice of educational administration: The case of Hispanic female superintendents. In Donmoyer, R., Imber, M., and Scheurich, J. (eds.) *The Knowledge Base in Educational Administration. Multiple Perspectives*, pp 158–173. New York: SUNY Press.
- Ozga, J. (ed.) (1993). *Women in Educational Management*. Milton Keynes: Open University Press.
- Reynolds, C. (ed.) (2002). *Women and School Leadership: International Perspectives*. New York: SUNY Press.
- Sander, B. (1993). Educational administration and developing countries. In Capper, C. (ed.) *Educational Administration in a Pluralistic Society*, pp 238–266. New York: SUNY Press.
- Schmuck, P. (1996). Women's place in educational administration: Past, present and future. In Leithwood, K., Chapman, J., Corson, D., Hallinger, P., and Hart, A. (eds.) *International Handbook of Educational Leadership and Administration*, vol. 1, pp 337–368. Dordrecht: Kluwer.
- Shah, S. (2006). Leading multiethnic schools: A new understanding of Muslim youth identity. *Educational Management, Administration and Leadership* 34(2), 215–237.
- Shakeshaft, C. (1987). *Women in Educational Administration*. Newbury Park, CA: Sage.
- Shakeshaft, C. (1995). A cup half full: A gender critique of the knowledge based in educational administration. In Donmoyer, R., Imber, M., and Scheurich, J. (eds.) *The Knowledge Base in Educational Administration. Multiple Perspectives*, pp 139–157. New York: SUNY Press.
- Skrla, L. and Scheurich, J. (2000). Sexisms, silence and solutions. Women superintendents speak up and speak out. *Educational Administration Quarterly* 36(1), 44–75.
- Sleeter, C. (1992). Power and privilege in white middle class feminist discussions of gender and education. In Biklen, S. and Pollard, D. (eds.) *Gender and Education*, pp 25–42. Chicago, IL: University of Chicago Press.
- Strachan, J. (1993). Including the personal and the professional: Researching women in educational leadership. *Gender and Education* 5(1), 71.
- Stromquist, N. and Monkman, K. (eds.) (2000). *Globalisation and Education: Integration and Contestation across Cultures*. Lanham, MD: Rowman and Littlefield.
- Tuhiwa-Smith, L. (1993). Getting our from under: Maori women, education and the struggles of Mana Wahine. In Arnot, M. and Weiler, K. (eds.) *Feminism and Social Justice in Education. International Perspectives*, pp 58–78. London: Falmer Press.
- Weiler, K. (1993). Feminism and the struggle for democratic education. In Arnot, M. and Weiler, K. (eds.) *Feminism and Social Justice: International Perspectives*, pp 210–225. London: Falmer Press.
- Weiner, G. (1995). A question of style or value? Contrasting perceptions of women as educational leaders. In Limerick, B. and Lingard, B. (eds.) *Gender and Changing Educational Management*, pp 23–33. Sydney: Hodder.
- Yeakey, C., Johnston, G., and Adkinson, J. (1986). In pursuit of equity: A review of research on minorities and women in

- educational administration. *Educational Administration Quarterly* **22**(3), 110–149.
- Young, B., Staszewski, D., McIntyre, S., and Joly, L. (1999). Care and justice in educational leadership. *Canadian Administrator* **33**(2), 38–51.

Further Reading

- Coleman, M. (ed.) (2005). *Special Issue: Diversity and Leadership. Leadership, Educational Management and Administration* **34**(2).
- Grogan, M. (2000). Laying the groundwork for a re-conception of the superintendency from feminist postmodern perspectives. *Education Administration Quarterly* **36**(1), 117–142.
- Larson, C. and Murtadha, K. (2002). Leadership for social justice. In Murphy, J. (ed.) *The Educational Leadership Challenge: Redefining Leadership for the 21st Century*, pp 134–161. Chicago, IL: University of Chicago Press.
- Murtadha-Watts, K. (1999). Spirited sisters. Spirituality and activism of African American women in educational leadership. In Fenwick, L. (ed.) *School Leadership*, pp 155–167. Lancaster: Technomic.
- Ortiz, F. (1982). *Career Patterns in Education: Women, Men and Minorities in Public School Administration*. New York: Praeger.
- Ozga, J. and Walker, L. (1999). In the company of men. In Whitehead, S. and Moodley, R. (eds.) *Transforming Managers: Gendering Change in the Public Sector*, pp 107–120. London: UCL Press.
- Young, B. (2002). The Alberta advantage: DeKleining prospects for women educators. In Reynolds, C. (ed.) *Women and School Leadership: International Perspectives*, pp 75–91. New York: SUNY Press.

Leadership and Student Voice

C Lanskey, University of Cambridge, Cambridge, UK
J Rudduck[†]

© 2010 Elsevier Ltd. All rights reserved.

Contemporary Thinking and Practice

Earlier Commitments

Educational practices which promote student voice and student leadership are not new: some independent (i.e., nonstate) schools in the early twentieth century developed a view of schooling in which the role of the student was radically re-thought. Despite the diversity of their ideological and political perspectives, the leaders of these schools shared a dislike of what they perceived to be a narrow educational outlook and undemocratic practices in existing schools and they established initiatives to enable students to participate in the collective leadership of the school community and the personal leadership of their own learning. Accounts of their experiences provide a useful reference point for reflection on contemporary thinking and practice.

In these schools, opportunities for students to express their opinions were considered integral to developing leadership skills. The most common arena for the expression of student voice was the school council, parliament, or moot where in some cases, students had an advisory role while in others they had a direct input onto strategic decision making. At Summerhill, perhaps the best known of these schools, every student had the same right to raise agenda items and to vote, as did each member of staff, including the school principal – A. S. Neill himself (Neill, 1992: 16–20). Student voice and student leadership – although not necessarily referred to in those terms – flourished in the protected context of mainly small residential schools with passionately upheld principles and with the autonomy to devise their own curricula. Recently, however, we have witnessed a resurgence of interest in student voice among state schools, despite the strongly performance-oriented climate.

Student Voice and Leadership Today

At one level, the annexing of student voice and student leadership is the outcome of a potentially transformative – or potentially trivializing – bringing together of two high-profile movements in educational policy and research – one supporting student rights and participation and the other exploring new conceptions of leadership, often as a solution to problems of underperforming schools. Some institutions are merely following the fashion in proclaiming

leadership opportunities for students but the small print of definition or description is often unconvincing; in others, it is the outcome of serious reflection on the purpose and possibilities of leadership in a context in which recognition of young people's capacity for insightful commentary and responsible action is central to the contemporary youth agenda. The current positions of leadership and student voice are explained below:

1. *Leadership.* Leadership has recently grown in prominence both as a set of practices and a field of enquiry. Levin (2006: 38) documents its extended remit – the noticeable shift in educational policy from expecting (school) leaders to be effective fiscal, organizational, and political managers to making them accountable for student, staff, and school performance. Mulford (2003: 4) also notes a marked shift from maintenance functions to activities that promote school improvement and explicitly seek to raise standards of achievement. The concept of distributed leadership – leadership that is not located in one person, but can be dispersed within the school (Mulford, 2003: 2), opens the door to the possibility of teacher and student leadership. Maden (2001: 327), in her 5 years on study of schools that appeared to be succeeding against the odds, identifies student participation as one of three crucial conditions for improvement, mentioning in particular the nurturing of leadership opportunities for teachers – and for students.
2. *Student voice.* Student voice has also achieved a high profile although there is an acknowledged gap between rhetorical advocacy and the realities of practice in schools and classrooms. It tends most commonly to take the form of consultation, which is a distinctive feature of the wider move toward greater student participation. The reasons why different groups – governmental and nongovernmental – endorse the idea of consultation and participation vary, but together they appeal to the following:
 - the principle of democracy in school as a way of preparing young people for their role in society;
 - the principle of young people's rights including their rights as members of the school community;
 - the idea that schools need to be more inclusive and to review the status of the student group;
 - concern for students' personal and social development which, it is argued, is nurtured by the respect and confidence that making a contribution and making a difference can bring; and

[†]Deceased

- the possibility of enhanced commitment to learning in that, students can help define a practical agenda for improvement that they themselves could respect and feel some ownership toward.

Although the reasons for supporting the student-voice movement may be different for different groups, all of them contribute to a growing awareness of the importance of listening to young people and strengthening their involvement in the various communities – including the community of the school – that they inhabit.

Different Interpretations

There are two main ways of interpreting the idea of student voice and school leadership and a third which is embryonic. First, it can refer to the ways in which students' voices can inform the understanding of school leaders, both the teacher in the classroom and the teacher(s) responsible, in either a hierarchical or a distributed-leadership model, for the management of the school and the well-being of those who work there.

Second, student voice and school leadership can refer to students' actual experience as leaders within and outside the classroom, whether in relation to their support for their peers or to the development of their own learning.

There is a third interpretation which is not yet well formulated and the examples of good practice which might give it stronger credibility are relatively few: it is about students' learning from experience what is entailed in being members of the school as a democratic working community, where they know that that they can contribute to decision making, and feel that they are sharing in the governance of the school.

All three interpretations demand a radical shift of perspective on the part of both student and nonstudent members of the school, particularly in terms of relationships and in ways of addressing issues of identity and power. Teachers must be ready to recognize young people as expert witnesses in the business of school improvement; and to accept that the traditional power structures of schools need to change in order to reflect new ways of seeing and positioning young people. The three main interpretations are discussed below:

Student Voice and School Leadership

1. *Consulting students: Ways in which students' voices can inform the understanding of classroom teachers.* There is a robust tradition of consultation which was particularly prominent in the 1970s and which continues today where external researchers go into schools, talk to students, write up what they have to say, and make their accounts available through publication. The tradition is an important one: it rests on the analytic expertise of the

professional researcher who must also, of course, be skilful in talking and listening to young people. Although such research provides a legitimate space for students to talk about their learning, there is often no attempt to feed back the outcomes of the enquiry to the students involved and no guarantee that opportunities for students to talk about their experiences as learners will be kept open. Another tradition exists within subject teaching, where consultation claims kinship with constructivist pedagogies and where students' subject knowledge is the main focus.

Giroux (1981: 150) said that what went under the name of educational innovation was often little more than the recycling and repackaging of old ways of doing things; it is important to be clear what is new about the current approach to consultation. Three things stand out:

- students are explicitly invited to comment on frameworks for classroom learning including, for example, teaching strategies and feedback and support systems;
- students have a more consciously analytic stance, knowing that they are contributing to school improvement and that their voices can make a difference; and
- students' expertise as insightful commentators on teaching and learning is recognized and respected.

Students may want to talk about their experiences of learning in a particular lesson or subject and explain what they find engaging, stimulating, satisfying, bewildering, difficult, or off-putting. Teachers may want to learn what students have to say, for instance, about a new teaching strategy or a new way of assessing learning, or the system of incentives and rewards used to motivate students and recognize their progress. Students involved in consultation know that their views are being sought because it is expected that they will have something useful to contribute.

2. *Consulting students: Ways in which students' voices can inform the understanding of school leaders.* Here the focus is not so much on aspects of teaching and learning – unless a school-wide issue has been identified that cannot be appropriately explored through student consultation at classroom level – but instead on aspects of school structures, procedures, regimes, and behaviors.

Student observations may be invited by school leaders or, particularly in schools where student voice and participation are already well established, students may themselves take the initiative in identifying and working on issues that concern them and seeking occasions to present their views/data/recommendations to school managers. In many schools, the student council will routinely offer proposals for action on issues of concern within the school community, and managers may set up task groups of students, possibly with one or more teacher members, to work on such issues.

A willingness on the part of the school principal or senior management team (assuming here a traditional organizational hierarchy) to take serious note of what students have to say and to demonstrate their receptiveness in practical ways, is the foundation of this approach. But student observations can sometimes make for uncomfortable learning for school managers – they tend to say it straight:

The management – they're very isolated from the rest of the school. It's more like there's the school and then there's the management who control the school rather like a business.

Sometimes I feel very much like just a statistic. They just want students to go through with good grades, more for the school's benefit than the students' – to keep the school having a high reputation and stuff.

I think it's easy when you're at somewhere like (school name) to lose you're your identity, just become one of the masses. (Rudduck *et al.*, 2006: 13)

Students as Leaders

In this construction, leadership can mean that students are taking on responsibility for the learning and/or social guidance of some of their peers but it can also be about the serious pursuit and shaping of students' own learning – a responsibility that students in some schools want to, and are being encouraged to, accept for themselves.

Students as leaders: Leadership roles that involve the regulation and/or support of others

Traditionally leadership roles for students have tended to be restricted to those that require them to act as extensions of teachers, whether to monitor conduct (i.e., as prefects) or to support learning (i.e., as reading mentors), or as ambassadors (i.e., talking about the school to outsiders). Recently, possibly as an outcome of the growing interest in student participation, schools have explored different ways of helping students to feel that they have a part to play in the daily running of the school. They may be responsible for displays of student work in different subjects, for designing and carrying out surveys of student opinion (the student-as-researcher role), for planning and leading learning walks so that teachers understand more about the different local environments that students come from, for acting as mediators (usually after receiving some training in conflict resolution). The roles are more about leading people than about leading learning. Students seem to value the chance to do something for the school and also the sense of being trusted; the challenge is to find tasks that are real rather than contrived and to make sure that many students become involved and not just the articulate and virtuous few.

Students as leaders: Leadership roles in relation to students' own learning

Here the concept of leadership is quite different and implies a degree of autonomy and trust that is absent in many classrooms (Rudduck and Flutter, 2000: 84). Brooker and Macdonald (1999: 86) see this form of leadership as requiring students to take a proactive role in the planning, implementing, and evaluating of their own learning. Flecknoe (2004: 406) suggests that student leadership is about young people participating in their own learning and in the organization of their learning context. Mulford (2003: 26) sees students' task, as leaders of their own learning, as being about controlling the learning process, including figuring out what they need to learn, working out as they go what concepts they have not understood, looking for additional information when they do not understand, (and) checking whether they remember what they have learned. The idea of students taking responsibility for their own learning in these ways is not new but is still not convincingly realized in the practice of most schools (Rudduck, 2004: 1).

Students as Co-Leaders in the Governance of the School

This interpretation remains somewhat distant from the current everyday realities of most schools: it may be a utopian vision whose day is coming or it may be a mirage. This, the third way, brings together several strands of thinking: a recognition of young people's maturity and capability; the ideal of partnership between teachers and students; a belief among teachers that students should be able to contribute significantly to decisions that affect the student body; a belief that students should be enabled to understand and critique the organizational and power structures of their school as a community; an acceptance that students have a right to have their voices heard and to ensure, where necessary, that basic community principles, such as inclusion, are respected. The culture of the school would be such that the student body is expected to support and also contribute to the development of democratic practice.

It is essentially about students learning about and exemplifying the principles of citizenship through their experiences in the everyday community of the school. As Hodgkin (1998: 11) argued, democracy is not something which is taught; it is something which is practiced.

There seem to be relatively few concrete examples of school practices in this category and Davies (1999: 39) suggests how tough the democratic agenda is to put into practice: it entails "a continuous political process whereby the operations of the decision making are transparent and open to challenge; whereby all members participate in the organization of the school; whereby rules and laws are consensually drawn up and members agree to abide by

those contractual rules . . . ; and whereby the human rights of all participants are upheld.” Gunter (2001: 129) refers briefly to a school researched by Smyth *et al.* (1998) which may meet the criteria and where, it is claimed, “dialogue about teaching and learning is ongoing and central, where people feel safe to challenge ideas and where teaching as a political process is accepted” (Gunter, 2001: 129). Little seems to be known about how easy it is to sustain such patterns of student involvement.

Impact

There is as yet little reporting of the effects of student leadership on students, teachers, and schools but evidence of the impact of student consultation and participation is reasonably consistent across reports (Davies *et al.*, 2006) and could well be similar in many respects for student leadership. Key outcomes for students appear to be these:

- a stronger sense of membership with students feeling more positive about school and more included in its purposes;
- a stronger sense of respect and self-worth so that students feel positive about themselves;
- a stronger sense of self-as-learner so that students are better able to manage their own learning; and
- a stronger sense of agency so that students see it as worthwhile to contribute to improvements in teaching and learning and wider school matters.

The impact of hearing students talk about aspects of teaching, learning, and schooling can also be considerable for teachers and for the school as an organization. Teachers have said that they gain:

- a more open perception of young people’s capabilities and attitudes;
- the capacity to see the familiar from a different angle;
- a readiness to change in the light of new perceptions;
- a renewed sense of excitement in teaching; and
- a practical agenda for improvement.

The benefits for schools can include the following:

- a basis for self-evaluation;
- an agenda for change that can make a difference;
- a stronger sense of the school as an inclusive learning community; and
- the basis for committing itself to enacting democratic principles in the daily life of the school.

However, such benefits – to students, teachers, and schools – only become fully realized when the process of involving students reflects authentic commitment, when students feel that school and classroom leaders’ interest in what they have to say is genuine, and when both teachers

and students feel that mutual trust is opening up the possibility of a more partnership-oriented way of working together.

Issues

The transformative potential of student voice can, it seems, be considerable but it is often difficult to implement, largely because of the power issues involved; similar issues could well be raised by attempts to introduce opportunities for student leadership. The following issues feature in research reports on student voice.

Inclusivity

A key issue is who is valued and whose voices are heard in the acoustic of the school. In a climate which celebrates high academic achievement, less-successful students can feel marginalized and/or that not much is expected of them. It is not surprising therefore if such students are cautious about believing that opportunities to express a view and to influence decisions are for them. Some predict, from past experience, that teachers only want to listen to or work with those who do well or those who are more articulate in the language of the school. Brice-Heath (2004) argues that many young people who have not participated extensively at home or at school in open discussions or small group conversation and as planners and thinking partners will not be confident in telling their stories or in representing their peers and may therefore miss out. In some circumstances therefore, the consultation process can affirm rather than challenge existing dividing practices in schools and the regimes which lead to some students being valued above others.

Again, where teachers anticipate a negative response from their colleagues to the idea of student voice and student leadership, the only way to build support may be to work with a small group of students and to demonstrate their achievements but this pioneering group can become an elite – and its status is often rooted in competence in talk which may, in turn, be linked to social class.

Uncertainty and Anxiety

Ellsworth (1992: 197) warns against the naivete of acting as if our classroom were a safe space in which democratic dialog was possible. It can take time to build the degree of trust that teachers and students claim is the cornerstone of learning to work together in new ways and there is evidence that both teachers and students new to student voice and student leadership can feel unsure about where the boundaries lie and whether retaliation might follow from getting them wrong. Such anxiety can lead both teachers and students to play it safe – students by saying what they think teachers will want to hear about

their lessons and teachers by finding ways of limiting the agenda of what students can talk about to things that are not professionally or personally threatening for them as teachers.

Credibility

As Fielding and Prieto (2002) warn, students will quickly lose trust in the potential of consultations if their opinion is sought on topics that are only of importance to teachers or that students see as important but which do not lead to change. They added: "We ... regard it as crucial for student perceptions and recommendations to be responded to, not merely treated as minor footnotes in an unaltered adult text" (p. 20). Hart (1997) identifies three forms of inauthentic use of student voice – which might also apply to student leadership. At the bottom of his ladder of participation are three rungs where the process is manipulative, decorative, or tokenistic:

1. *Tokenism*. Children seem to have a voice but have little or no choice in the subject or style of communication and no time to formulate their own opinions.
2. *Decoration*. Children are used to promote a cause but have no involvement in organizing the occasion.
3. *Manipulation*. Adults consciously use children's voices to carry their own message.

If schools are to build open and trusting relationships as a basis for constructive dialog and partnership, then students need, at the very least, to know what is happening as a result of what they have told the visiting researcher or the researching teacher in school: if they have made specific proposals, they need to know what is possible and what is not possible.

Other Issues

In addition to these issues that reflect the deep-rootedness of power-based relationships in schools are four others that deserve mention. First, in a climate where performance is privileged, teachers say that they have difficulty in making space for innovations that do not directly raise standards. Second, in a climate which endorses restless innovation, it is difficult for schools to practice the discipline of ensuring that all their initiatives support the same core values; this can mean that there might, at any one time, be a diversity of initiatives in a school, each reflecting different concerns and values. Third, while there may be a rhetoric of encouragement for student voice and student leadership from the senior management team, the level of practical commitment in terms of resources for effective implementation may remain low. Fourth, while student leadership and voice may appear to have a firm base in the school, they may be contained in designated spaces such as citizenship education or in meetings

of the student council and they may not permeate the everyday exchanges of the school as a community.

Comment

The implications of translating the idea of student voice and student leadership into practice are considerable but the fundamental changes involved are not easily accomplished. As Cook-Sather (2002: 4) makes it clear, "power issues structure the twin challenges to the status quo: '(a) changing the structures in our minds that have rendered us disinclined to elicit and attend to students' voices; (b) changing the structures in educational relationships and institutions that have supported and been supported by this disinclination.'" Orner, taking a similar perspective, urges educators to discard notions of monarchical power and shift to notions of power as productive and present in all contexts' (1992: 83; in Brooker and Macdonald, 1999: 88). These more open notions of power extend the possibilities of leadership and this in turn creates spaces for developing roles for students as expert witnesses and as leaders and co-leaders in matters of school governance and classroom learning. At the moment, however, in this unsettled field, examples of carefully thought-through practices seem to lag behind the urgent patronage of the movement's various supporters. What has been achieved, however, is the placing of the issues firmly on the educational agenda so that teachers, school managers, and researchers can explore forms of leadership that will engage capabilities in young people that often go unrecognized in conventional structures and pedagogies. And, as Smyth (2006: 279) says, we are at last learning to listen to and attend to the voices of the most informed, yet marginalized witnesses of schooling – the students.

See also: Leadership: Democratic.

Bibliography

- Brice-Heath, S. (2004). Risks, rules and roles. In Perret-Clermont, A. N., Pontecorvo, C., Resnick, L. B., Zittoun, T., and Burge, B. (eds.) *Joining Society*, pp 41–70. Cambridge: Cambridge University Press.
- Brooker, R. and Macdonald, D. (1999). Did we hear you?: Issues of student voice in a curriculum innovation. *Journal of Curriculum Studies* 31(1), 83–97.
- Cook-Sather, A. (2002). Authorising students' perspectives: Towards trust, dialogue, and change in education. *Educational Researcher* 31(4), 3–14.
- Davies, L. (1999). Researching democratic practice understanding in primary school. *Research in Education* May (61), 39–48.
- Davies, L., Williams, C., Yamashita, H., and Aubrey, M.-H. K. (2006). *Inspiring Schools: Impact and Outcomes: Taking up the Challenge of Pupil Participation*. London: Carnegie Trust.
- Ellsworth, E. (1992). Why doesn't this feel empowering? Working through the repressive myths of critical pedagogy. In Luke, C. and Gore, J. (eds.) *Feminisms and Critical Pedagogy*, pp 90–119. New York: Routledge.

- Fielding, M. and Prieto, M. (2002). The central place of student voice in democratic renewal: A Chilean case study. In Schweisfurth, M., Davies, L., and Harber, C. (eds.) *Learning Democracy and Citizenship: International Experiences*, pp 19–36. Oxford: Symposium Books.
- Flecknoe, M. (2004). Challenging the orthodoxies: Putting a spoke in the vicious circle. *Educational Management Administration and Leadership* **23**(4), 405–422.
- Giroux, H. A. (1981). *Ideology, Culture and the Process of Schooling*. Lewes: Falmer.
- Gunter, H. M. (2001). *Leaders and Leadership in Education*. London: Paul Chapman.
- Hart, R. (1997). *The Theory and Practice of Involving Young Citizens in Community Development and Environmental Care*. London: Earthscan.
- Hodgkin, R. (1998). Partnership with pupils. *Children UK*, Summer.
- Levin, H. M. (2006). Can research improve educational leadership? (A review of 3 books). *Educational Researcher* **35**(8), 38–43.
- Maden, M. (2001). Further lessons in success. In Maden, M. (ed.) *Success against the Odds – Five Years on: Revisiting Effective Schools in Disadvantaged Areas*, pp 307–339. London: RoutledgeFalmer.
- Mulford, B. (2003). School leaders: Changing roles and impact on teacher and school effectiveness. *A Paper Commissioned by the Education and Training Policy Division*. OECD.
- Neill, A. S. (1992). *The New Summerhill*. Harmondsworth: Penguin.
- Orner, M. (1992). Interrupting calls for student voice in 'liberatory' education: A feminist poststructuralist perspective. In Luke, C. and Gore, J. (eds.) *Feminisms and Critical Pedagogy*, pp 74–89. New York: Routledge.
- Rudduck, J. (2004). Consulting pupils about teaching and learning. *Paper Presented at the 5th Annual Conference of the Teaching and Learning Research Programme*, Cardiff, 22–24 November 2004.
- Rudduck, J. and Flutter, J. (2000). Pupil participation and pupil perspective: 'Carving a new order of experience'. *Cambridge Journal of Education* **30**(1), 75–89.
- Smyth, J. (2006). Educational leadership that fosters 'student voice'. *International Journal of Leadership in Education* **9**(4), 279–284.

Further Reading

- Fielding, M. (2004). Transformative approaches to student voice: Theoretical underpinnings, recalcitrant realities. *British Educational Research Journal* **30**(2), 295–311.
- Finney, J., Hickman, R., Morrison, M., Nicholl, B., and Rudduck, J. (2005). *Rebuilding Engagement through the Arts, Responding to Disaffected Students*. Cambridge: Pearson (This short book includes accounts of 'difficult' students taking leadership roles.)
- McIntyre, D. and Rudduck, J. (in press). *Improving Schools, Consulting Pupils*. London: Routledge.
- Mitra, D. (2005). Student voice in school reform: From listening to leadership. *Paper Presented at Annual Meeting of the American Educational Research Association*, San Francisco, CA.
- Rudduck, J. and Flutter, J. (2004). *How to Improve your School: Giving Pupils a Voice*. London: Continuum.
- Thrupp, M. and Willmott, R. (2003). *Education Management in Managerialist Times*. Maidenhead: Open University Press. (See in particular the discussion of Gunter's work.)

Leadership and Urban Education

K A Riley, Institute of Education, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Context and community shape leadership – but no more so in challenging urban localities. The argument presented in this article is that urban school leaders and policymakers need to understand more about the context, challenges, and opportunities of our cities. Given this underlying premise, the first part of the article explores the broad context for urban leadership, highlighting the ways in which our cities can change rapidly, and conceptualizing the challenges of urban education as well as the characteristics of urban schools.

Urban school leaders face the additional leadership challenges of developing insights into the daily lives of children and young people, and finding new ways of connecting to local communities. These challenges emerge because urban communities are very diverse; they change rapidly; and staff in urban schools rarely live locally to be sufficiently aware of these changes in the community.

The second part of the article focuses on the development of systematic ways that can be adapted by school leaders to understand the nature and complexity of local communities and connect this knowledge and understanding with the school's internal community. It offers a tool, which strengthens the school leaders capacity to read and interpret the unique community context of their own schools: the physical characteristics of a locality in terms of housing and transport; the socioeconomic features; issues about race, ethnicity, and cultural diversity; and patterns of mobility and transformation. These historical and contemporary factors influence the social context and the expectations and beliefs that children and their families bring to their educational experience. The article concludes with a brief discussion about how urban school leaders can apply this knowledge to their leadership.

Making Sense of the Big Picture

Cities Today

Cities across the globe attract large and diverse populations. They may be migrants from within their own countries, or immigrants from countries facing social, political, or military upheaval, or seeking new economic opportunities. Many cities are maelstroms, whirlpools of excitement, energy, and risk. They become the focal points of societal struggles for the distribution of resources, epicenters for the

overthrow of accepted social norms, and seed beds for the creation of new ideas. Typically, they also contain conservative forces which may strive to maintain the influence of the elite and the wealthy. The contrasts between rich and poor raise issues about how societies educate the children of the most deprived in close proximity to those of the most advantaged. Given the complexity of cities, it is important that policymakers and school leaders understand the diversity and the mobility.

In the late nineteenth and early twentieth century, rural populations flocked to the industrial cities of Europe or North America, seeking jobs and food. In the twenty-first century, the rural populations of sub-Saharan Africa and East and Southeast Asia congregate in Dakar or Vientiane. Cities today bear first witness to the social transformations of societies acting as the interface between the poor, the aspirational, the established, and the elite. In South America, cities such as San Paulo and Rio have become the melting pots which attract the creative, as well as the displaced. About one in three of the world's urban population now lives on the edges of cities in developing countries, and the urbanization of poverty is on the increase (UN-Habitat, 2005).

Our cities can become transformed, seemingly overnight, as new immigrant groups move in – bringing traditions which can enrich host communities but which can also generate prejudice (see e.g., Winder, 2004) or result in established communities becoming dispersed. In the 1950s, Michael Young and Peter Willmott documented the impact of redevelopment on East London's indigenous communities in Bethnal Green, criticizing planners and policymakers for breaking up families and communities (Willmott and Young, 1957). In the 1960s, Herbert Gans wrote the parallel American story: the decanting of Italian-Americans from their tenement community in Boston's West End, to make way for luxury housing (Gans, 1962).

Regeneration continues apace, raising contemporary but comparable challenges. For example, central parts of the city of Birmingham England are being redeveloped. Established multiracial urban communities are being replaced by the young and upwardly mobile. In 2006, as part of a project on urban leadership (The headteacher was part of an ongoing research and development project Leadership on the Front-Line which is led by Professor Kathryn Riley. See page x and Riley *et al.*, 2004.) UK headteachers and school principals visited an early years' center located at the center of this redevelopment.

Reflecting on the impact of this development (in terms not unfamiliar to Young, Willmott, or Gans) one London headteacher commented:

It's social engineering, ethnic cleansing in a way... What they're doing there is destroying the community. There are generations of families and the daughter's got a house down there... and the daughter's being moved away. And what they've done is they've separated that link. And sometimes those communities need those links. The structure of the family unit – they've destroyed that. And that's going to create social problems.

Yet, urban regeneration also implies a sense of possibilities and opportunities. For example, children and young people in Newham, North London, are enthralled by the rebuilding and development which is part of preparation for the 2012 London Olympic Games.

The contradictions of our cities seem endless: the rich cultural contexts versus the day-to-day challenges of life in the inner city or in social housing estates; the creativity and vibrancy of the children versus the unremitting pressures on our city schools; or the employment attractions of city life for high flyers versus the difficulties which many city schools face in attracting and retaining staff. These contradictions and contrasts not only create major social and economic problems but also provide many positive features to urban life which have implications for the provision of education.

Cities are key economic centers, with diverse economic and employment opportunities. Cities and capital cities in particular, are centers of government, political, and social activity which attract a wide range of organizations and agencies that require proximity to the center of policymaking. Almost invariably, they are a focal point for education and arts. Their ethnic, linguistic, and religious diversity serve to attract a rich range of internal and external migrants who contribute to social, economic, and cultural developments (Riley and West-Burnham, 2004).

Conceptualizing the Challenges of Urban Education

We can conceptualize the educational challenges of our city schools in radically different ways: solely in terms of malfunction and the collapse of school systems, part of a picture of unremitting failure; or in terms of the successes and richness of our city schools, their potential as seed beds of innovation, within contexts that are challenging and ever changing. Our starting point will influence not only what we are looking for from our schools, in terms of aspirations and outcomes, but also the policy options and strategies we choose to adopt.

While many features of urban education are connected to broader educational, social, and economic issues which extend beyond the immediate urban context (see Riley,

1998; Whitty *et al.*, 1998), to a considerable degree, politicians on both sides of the Atlantic have focused on the negative aspects of urban schooling. The policy discourse, particularly in the United States of America, has tended to center on the perpetual crisis of urban education: teacher shortages and cutbacks in services. For some, the term urban has become synonymous with “poverty, non-white violence, narcotics, bad neighbourhoods, an absence of family values, crumbling housing and failing schools” (Kincheloe, 2004: 2). But are these problems due to schools, or due to governments?

Contributors to a review of education in the United States of America (intended for the President, George W. Bush), then argued that government spin has obscured the nature of the crisis of US urban education, projecting it as a failure of public organizations, rather than as a failure of political will (Glickman, 2004). (In an introduction Bill Cosby argues, “I’m assuming that the President of the United States never went to a poor and neglected public school – where books have missing pages, walls having peeling paint, children have nothing to draw or write with, and where there is no library for reading a story doing homework” (Cosby, 2004: xii.) The authors pointed to major discrepancies in funding across schools in America – between those serving the most advantaged and disadvantaged communities – and argued that high-stakes testing and overreliance on marketplace solutions (such as vouchers, incentives, competitions, and privatization), had done little for the most disadvantaged students (Oakes and Lipton, 2004). (They point out that schools serving African American and Latino students typically offer fewer advanced classes, and have more outdated text books, materials, and science laboratories than other schools. Their school buildings are likely to be older and in greater disrepair, and they are far more likely to be crowded, unhealthy, and uncomfortable. Teachers in those schools are more likely to have less experience, to not have their state’s highest teaching credential, or to not have a degree in the subject they teach. In addition, low-income students and students colored are placed disproportionately in low-level programs (Oakes and Lipton, 2004: 188).) While contributors to a parallel review in the United Kingdom (intended for the newly elected Prime Minister of the day, Tony Blair) welcomed increased investment in education from previous Labor Administrations. (They also welcomed major initiatives in child care, such as Sure Start which focuses on the needs of children and their families in the most deprived areas (Wragg, 2005)). They too, raised major issues about equity and the allocation of resources, drawing attention to the relatively low percentage of school resources allocated to tackling social disadvantage. Prof. Tim Brighouse, senior adviser to London Schools, pointed out that Additional Education Needs (AEN) range between 1% and 8% of the total school budget. He proposes

allocating resources to secondary schools in relation to the performance of children (on national tests) on entering the school (Brighouse, 2005).

Defining the Characteristics of Urban Schools

Although there are many distinctive differences between our city schools, there are probably a number of defining characteristics of the educational landscape. Writing about the urban context in the United States, Joe Kincheloe identifies common features such as high population density, areas of profound economic disparity, and higher levels of ethnic, racial, religious, and linguistic diversity (Kincheloe, 2004). The school population of US urban schools is increasingly black and Hispanic (Olmedo, 1997). The American urban-education landscape, Kincheloe argues, is also likely to include schools that are typically larger than elsewhere (which means that students can be overlooked and may find it difficult to develop a sense of identity); school districts which face a concentration of problems; and school boards which “experience factionalized infighting . . . over resources and influence” (Kincheloe, 2004: 6).

A UK review of schools in urban and challenging contexts concluded that while formal definitions were difficult to find, there appeared to be a “myriad of complex and socially related problems” which characterized the schools (Keys *et al.*, 2003: 7). (Despite the complexity and challenges of urban contexts, initial teacher training does little to prepare teachers to work in urban contexts (Ash *et al.*, 2006).) Schools had a range of negative indicators in terms of educational potential and capacity, including high proportions of pupils on free school meals (an indicator of poverty and social disadvantage). Social and economic factors included ill-health, financial pressures, family stress, and poor housing (Ash *et al.*, 2006). Dislocation, social disaffection, high levels of drug and alcohol abuse, and high crime rates are also common community features. There are also school factors such as management, resourcing, and schools’ interactions with their local community which characterize a number of challenging urban schools (Hopkins, 2001; Learmonth and Lowers, 1998).

Nevertheless, the cultural and community diversity of many of our city schools is often a source of great strength. Schools reap the benefits of multiplicity in our societies: the creativity, the energy, and the resilience and exuberance of the children; and the rich cultural understanding and experiences of their parents. However, cultural and community diversity also has many other implications for schools. This is equally true of many urban schools in the United States which have to respond to significant changes in the demographic terrain (Louis, 2003). There are issues about:

- the ability of newly arrived immigrant groups to engage with society;

- the concentration of socially disadvantaged children, and the demands that this creates on resources;
- how to work with an underclass with very low expectations, high levels of disaffection from society, and a dismissive attitude toward learning; and
- poverty and mental health problems, for both pupils and their families, which serve to increase the complexity and nature of the challenges.

Making Sense of the Locality

Having examined the big picture, school leaders need to focus on their particular locality.

As part of the project, Leadership on the Front-Line, we brought together over 60 headteachers and principals from a range of schools in challenging urban contexts across the United Kingdom and Eire – Belfast, Birmingham, Cardiff, Dublin, London (Greenwich, Hammersmith and Fulham, Newham and Tower Hamlets), Londonderry, Liverpool, and Manchester. (The schools are cross-phase (nursery, primary, secondary, and special) and reflect the range of types of schools in urban contexts: denominational, integrated, county, girls, and boys. The development of Leadership on the Front-Line was stimulated by concerns about the gaps in understanding about the challenges of urban schools, and due to a recognition of the importance of developing a more focused community-based approach to leadership.) The schools have been subject to flux, through changes in the policy environment: housing, policing, and regeneration. They are being rebuilt and restructured: their communities can change fast, with a speedy exodus of the upwardly mobile. When the affluent move in, they rarely boost the local state-school population.

While poverty, deprivation, and high incidences of mental ill-health are features of the community context in which these schools are located, there can be contrasts: a school’s population can include children of doctoral students, as well as those of refugees. Some of the urban schools located in inner-city areas have typically been the most deprived parts of our cities; however, a number of schools are also located on the outer rim of cities and serve white working-class communities housed in postwar social housing, areas of high unemployment which may lack good transport or infrastructure.

As part of the project Leadership on the Front-Line we wanted to enable headteachers and school principals to gain greater insights into the local community context of their schools and we asked them to carry out their own local-community audit based on three core questions:

- How do I read my community?
- What is changing?
- What is my role within it?

Using a tool designed for the project, they mapped their school community on four axes: population (stable

or mobile e.g., with refugees and immigrants); community profile (single or multiple communities); levels of engagement (engaged with education or disengaged); and community identity (integrated with a strong community identity, or fragmented with disparate communities). Once participants had undertaken the mapping exercise they were left with a shape which related to their own community context. This exercise generated considerable discussion and helped create a shared language to discuss the complexities of context.

Based on the mapping exercise, we were able to distinguish four types of urban schools, those that serve:

- an inner-city single, relatively homogenous community;
- multiple and diverse communities within a locality;
- an estate community; and
- multiple and diverse communities over an extended area.

Single/Homogenous Communities

The label single or homogenous community covers a range of possibilities. It can mean that the school serves a predominantly white working-class community, or a single ethnic group, for example, Bengali. In Northern Ireland, this community can be Catholic or Protestant. **Figures 1 and 2** illustrate the complexities of the label single community. **Figure 1** depicts the makeup of a school in Londonderry which serves a white Protestant community which, according to the principal, while not necessarily engaged with education, can become engaged with the school.

Figure 2 has many similarities. It depicts a school in Tower Hamlets, London, which draws its pupils from a predominantly Bengali, Muslim community and which has greater mobility than the Londonderry school. While the social and cultural worlds of the two school communities shown in **Figures 1 and 2** are very different, school

leaders in both contexts share the challenge of working with communities which (on the surface) appear to be unified by one culture and set of beliefs but may, nevertheless, contain many differences in views and perceptions.

Multiple Communities

The community context of schools that draw from multiple communities can be complex and subject to rapid change. A headteacher from Greenwich in London described these changes as follows:

From the time I started as the head in 1992 until about three years ago, we had 32% bilingual pupils and about 46% children from minority ethnic groups. And that has now changed. We've gone up to nearly 47% bilingual pupils and sixty something minority ethnic groups.

Relationships within a community can also be subject to significant shifts, as one headteacher from Tower Hamlets in London commented:

It's a community in which, in the 10 years or so that I've been here, has undergone quite significant change. And 10 years ago the tensions were between the Bangladeshi community and the White and Afro-Caribbean community. To a large degree that's changed. Although our school mirrors [the Borough] almost exactly in its ethnic make-up, so we're about 55% Bangladeshi, 30% English, Welsh, Scottish, Irish, 10% Afro-Caribbean and the rest is a mixture of other groups like Turkish, Somali, Vietnamese. . . although that's the case, touch wood, there isn't a great deal of violence among or between the White and the Bangladeshi community. Where the violence tends to be now, quite often, is between different groups of Bangladeshi males and that gang behaviour among young Bangladeshi males is probably one of the most difficult problems facing [the Borough] at the moment. It's one that's impacting on schools to a quite a significant extent.

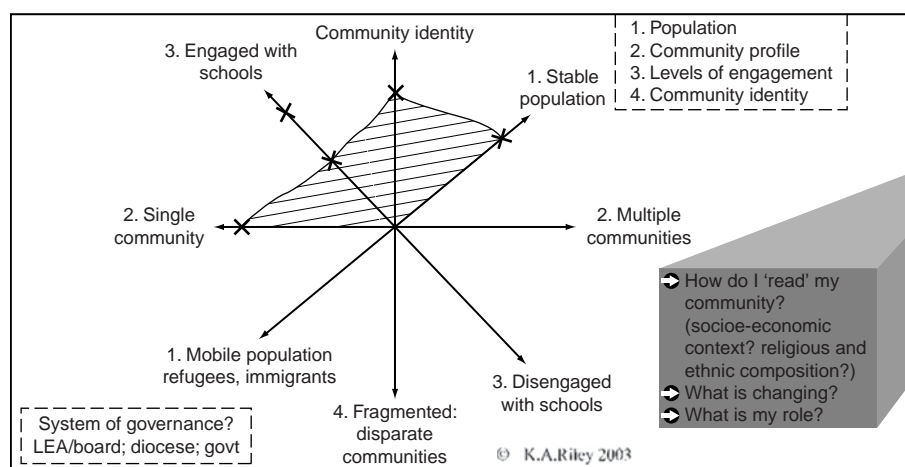


Figure 1 City of Derry school. Single community, predominantly Protestant white working-class.

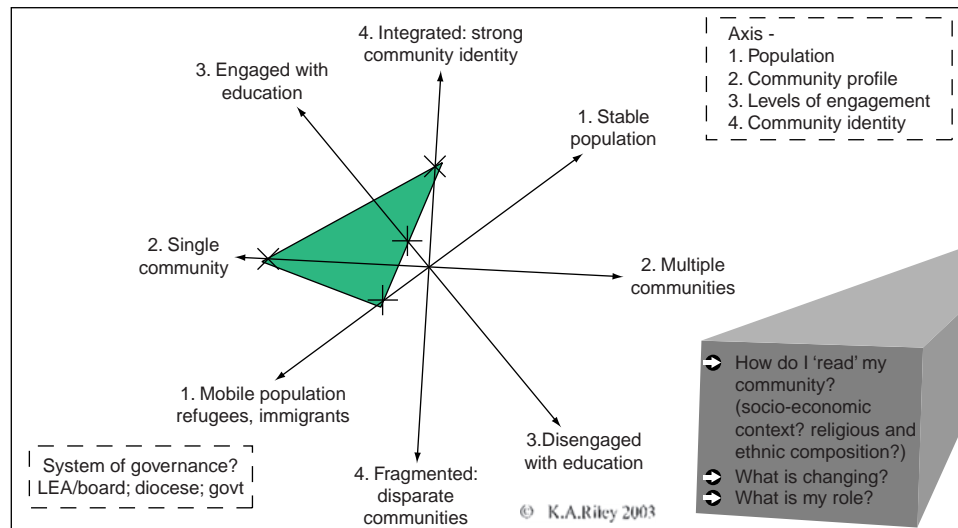


Figure 2 Tower Hamlets School, Phase II. Single community, predominantly Bengali.

Within these varied and multiple communities, some groups may be engaged with education and others not. In the Tower Hamlets, for example, the white working-class families are more likely to be disaffected than immigrant groups, a situation common to other inner-city areas. The Tower Hamlets' headteacher continued her analysis of the context as follows:

...In the white working-class population, often the families have had a negative view of education themselves and they transpose that onto their children and they see that education has done nothing for me, therefore my child doesn't particularly need it. ...And perhaps they live in the past. ...the days when you could leave school on Friday and go to work on a Monday.

The Bangladeshi families ... have a sense that education will somehow give them a way out of this and a value for education, although there is a willingness to try and participate in the educational process, there is a lack of knowledge on how to do that and it's up to us to try and help them to understand the educational process because I think what often happens is, they've had very limited education themselves and what they have had would have been in a very different setting.

In other contexts, whole sections of a community can be struggling (Figure 3), as is described in the following extract from an interview with a Liverpool headteacher:

It's a very dysfunctional community, really I think. It's coming to the bottom of a very deep hole of deprivation, problems with drugs, arms, related crime. A lot of disaffected youth in gangs, children between 10 and 16 or 17 maybe. Causing a lot of problems for the people living here. ... I've been a headteacher here for 9 years, It was a very stable area ... but I've seen it go down and down, possibly due to the

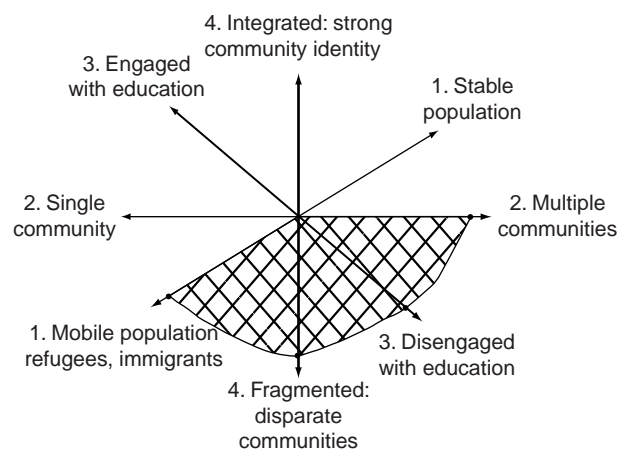


Figure 3 Liverpool School (serving multiple disadvantaged communities). School headteacher's perceptions of the community using the mapping exercise.

housing policy here where housing is being used to decant people who are having difficulties but the effect has been to move people out of this area who were very stable. Into that mix, I think, comes a lot of people just recently, asylum seeking families, who are having quite a lot of difficulty here.

It's quite a racist area, but there is a lot of short-term housing here. So currently in the school, I think, it's 19% of the children are from asylum seeking families. The mobility rate here is 41% – it's very high. ... There are 17 different languages amongst that group, but what we're finding is that people are coming in and going out, some of them because economically, they are moving around the area. That tends to be Czech Roma families, but there are other families who really are finding it very difficult to come to terms with the aggression.

Estate Schools

Estate schools are another facet of urban schooling. Typically, they serve communities that have moved from traditional inner-city areas. This can be through major post-World War II re-housing (as in the cases of Manchester's Wythenshawe Estate, or Cardiff's Ely Estate), or through re-housing related to sectarian violence, as in the case of Belfast, described below:

The school originally would have served the area of T which was a new housing estate developed in the 60s and, at that time, it was an area basically that was to house people who had been burnt out, frightened out of their original area, and certainly during the 70s it was a very fragmented community because there was no ownership, no community feel, as well as that there was no infrastructure shopping centre. So people at that time . . . were very poorly served and it was also on the outskirts of West Belfast. It's now a more settled environment The area would be in the top 10% of wards with multi-deprivation and the social/economic deprivation is severe. Within the school itself, which is the only secondary school in this particular area, although there are quite a number of schools which people can gravitate to. . . . There is over 70% free school meals.

Across all of the cities involved in Leadership on the Front-Line, there are remarkable similarities between the ways in which school leaders of the estate schools have depicted their schools. **Figure 4** depicts a typical drawing. This particular school is located in Cardiff.

The Cardiff headteacher responsible for **Figure 4** described the context of the school in the following terms:

It takes its catchment from one indigenous location; it's predominantly an estate, so it's fairly homogenous . . . predominantly white. There's a strong identity of all

the pupils with this area geographically. . . . There are sections of the community who create difficulties. The school's playing fields were littered with burnt out cars and there was damage done to the building. . .

A Manchester headteacher described the estate community served by his school as follows:

It's 52% free school meals. We serve the poorest, officially the most socially deprived ward in England. . . . So we've got severe urban poverty. . . . but in terms of cleverness, we actually do get some of our brightest children from x (the most deprived ward). They really, really are bright kids. In terms of the estate, its 70,000 housing units.

If you want to talk about an inverted snobbery form of evaluating how tough the context is here . . one of our ex-students was murdered. He left school two years ago last summer and we had his friend who was shot dead just before Christmas. . . One of our parents was killed in a motorcycle accident. So there seems to be lots of deaths in the area. For the last 4 years, every summer I've gone to a funeral, to do with our children.

The crime rate is high, there is a drugs problem, but we do have squirrels on the ground, we do have our own pond. But it is a tough community. It is white working-class, we have about 1% from different ethnic groups, very few, less than 10% but it is on the increase, because they're pulling down flats and building start-up homes. . .

Multiple Communities over an Extended Area

Each of these three types of community discussed so far (i.e., single; multiple or diverse; and estate) has its own distinctive challenges. However, schools without obvious and immediate communities, and which serve a wide

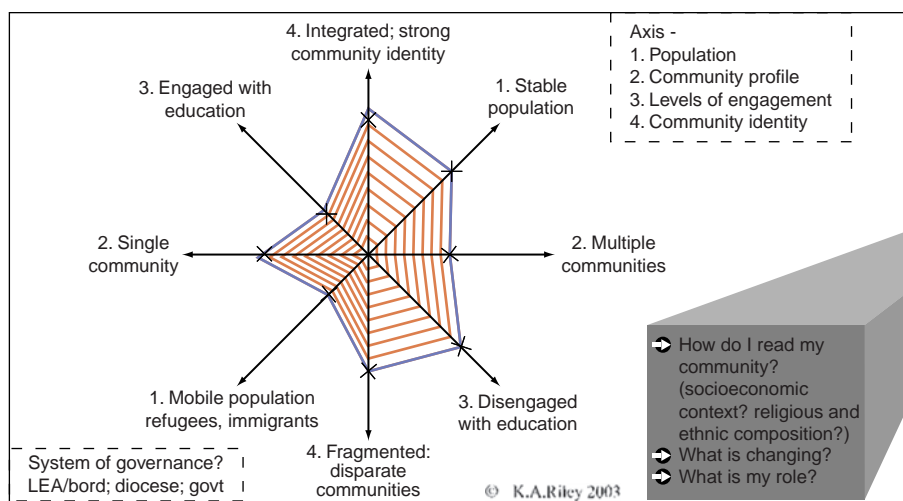


Figure 4 Estate School, Cardiff (serving relatively stable population).

catchment area face a different set of challenges. In the following extract, a Birmingham headteacher describes the complexity of the community context of his school:

We are the first step out as receivers of the urban doughnut so whilst we're in lovely grounds and Mosley which is near by, is one of the more affluent pleasant bits of central Birmingham. 70% of our pupils come from the top 2% areas of deprivation across the European community. . . . 76% of our pupils are in the poorest quintile of society by multiple deprivation figures. . . . We have an incredible diversity in terms of economic background, class. I have the son of a member of the House of Lords, and a government minister and I have 35% Muslim community predominantly Pakistanis but some Bengali. . . . Generally, Muslim families come from backgrounds of very low literacy levels and with a deficit starting point. We have some very needy white working-class families. . . . Significantly, the families come from very different areas so we have an official catchment area of 3.2 km which for a dense urban area is quite big but actually we have quite a lot of youngsters who come from outside that circle. . . . and so there is a lack of identity on our campus.

What Do You Do with What You Know?

So far in this article, the broad urban context as well as the ways in which urban localities can vary have been explored. Having taken stock of their community context – the challenges and the opportunities – what options are then available to school leaders? There appear to be at least three possible responses (Riley and Stoll, 2005). (See Riley (2008) for a more detailed account of how school leaders can draw on children's knowledge and experience to develop connections between schools and communities and create shared beliefs about what can be achieved for, and by, the young people of our cities):

- *Response 1: evade.* Here, the school leader attempts to weaken the links between the school and its immediate catchment area which is seen to be too problematic. The leadership strategy is directed toward attracting more middle-class pupils into the school to help push up test scores. This is not an uncommon response and can have some benefits to the local children remaining in the school, but not to those denied access.
- *Response 2: ignore.* The school leader attempts to insulate the school from the community. There are probably two versions of this approach. The first is the fortress model – an approach favored by some secondary and high schools and designed to keep the local community at bay. The second is the bubble model – a favorite for those working with younger children and designed to

protect them from malign influences. Both models are motivated by worthy intents – to provide children and young people with space, opportunities, and safety. Both models have their success stories but there are also downsides. Children can feel torn between two cultures and communities which never meet.

- *Response 3: engage.* Here the school leader attempts to reach out to understand the complexities of community and bring the school and the community into closer alignment. This notion of alignment is an important one. It implies developing an understanding and reaching an agreement, based on trust and mutual respect, about how schools and communities will work together. This can be a complex undertaking in pluralist and multi-racial societies and school leaders can find themselves mediating with the community on faith, cultural, and value issues. A school's cultural beliefs and expectations may be at odds, not only with those of newly immigrant communities but also with long-established communities.

Urban school leaders have to grapple with issues and pressures which may confront other schools and communities in future years. What school leaders in challenging contexts have to know and be able to do now is what their counterparts (in less pressing contexts) will have to contend with in future years.

What we have learned from the project Leadership on the Front-Line is that successful urban leadership is about passion, beliefs, and possibilities. It is about building connections between schools and communities, and creating some shared beliefs about what can be achieved for and by children and young people. Where communities are dispersed, complex, and fragmented, the leadership challenge is to create and recreate community within the school by bringing young people to work together to appreciate the richness and diversity of each others' lives.

Providing school leaders with approaches that will develop their capacity to understand their local community better, and to work together with that community more effectively, can unleash their creativity and potential. It is not a burden for school leaders but something that can be extremely beneficial. It supports their leadership in contexts that are both challenging and rewarding and enables them to reap the rewards, as one leader, who involved rewarding in the project Leadership on the Front-Line concluded:

Being an urban school leader is like being on a roller coaster. It can be full of highs and lows. It is full of teamwork and feels very supportive. The sense of volatility and the unexpected is addictive. Understanding more about the local area and learning to listen to what pupils have to tell us about their lives has given my leadership an enormous boost. It has given me a clearer focus about how to work with communities.

See also: Effective Leadership in Challenging Schools; Leadership: Democratic; Leading Diverse Schools.

Bibliography

- Ash, A., Hall, D., Raffo, C., Diamantopoulou, S., and Jones, L. (2006). *Urban Initial Teacher Training: Working with Urban Schools in Challenging Contexts: Development Award Report, Teacher Training Agency*. Manchester: Institute of Education, University of Manchester.
- Brighouse, T. (2005). Tackling disadvantage, disaffection and underachievement. In Wragg, E. C. (ed.) *Letters to the Prime Minister*, pp 1–125. London: Central Books.
- Cosby, B. (2004). Where do we start the sweep? In Glickman, C. (ed.) *Letters to the President: What We Can Do About the Real Crisis in Public Education*, pp 1–11. New York: Teachers College Press.
- Gans, H. (1962). *The Urban Villagers*. New York: Free Press.
- Glickman, C. (ed.) (2004). *Letters to the President: What We Can Do About the Real Crisis in Public Education*. New York: Teachers College Press.
- Hopkins, D. (2001). *Meeting the Challenges: An Improvement Guide for Schools Facing Challenges*. London: DfEE.
- Keys, W., Sharp, C., Greene, K., and Grayson, H. (2003). *Successful Leadership of Schools in Urban and Challenging Contexts*. Nottingham: National College for School Leadership. www.ncsl.org.uk.
- Kincheloe, J. L. (2004). Why a book on urban education? In Steinberg, S. R. and Kincheloe, J. L. (eds.) *10 Urban Questions: Teaching in the City*, pp 1–212. New York: Peter Lang.
- Learmonth, J. and Lowers, K. (1998). A trouble shooter calls. In Stoll, L. and Myers, K. (eds.) *No Quick Fixes: Perspectives in Schools in Difficulty*, pp 133–144. London: Falmer.
- Louis, K. S. (2003). Democratic values, democratic schools: Reflections in an international context. In MacBeath, J. and Moos, L. (eds.) *Democratic Learning: The Challenges to School Effectiveness*, pp 15–41. London: Routledge Kegan Paul.
- Oakes, J. and Lipton, M. (2004). And equal education for all. In Glickman, C. (ed.) *Letters to the President: What We Can Do about the Real Crisis in Public Education*, pp 184–192. New York: Teachers College Press.
- Olmedo, I. M. (1997). Challenging old assumptions: Preparing teachers for inner city schools. *Teaching and Teacher Education* **13**(3), 245–258.
- Riley, K. A. (1998). *Whose School is it Anyway?* London: Falmer.
- Riley, K. A. (2008). Improving city schools: Who and what makes the difference? In Sugrue, C. (ed.) *New Directions for Educational Change: International Perspectives*. London: Routledge.
- Riley, K. A., Hesketh, T., Rafferty, S., et al. (2004). *Urban Pioneers – Leading the Way Ahead: First Lessons from the Project Leadership on the Front-Line*. London: Institute of Education, Issues in Practice Series.
- Riley, K. A. and Stoll, L. (2005). *Leading Communities: Purposes, Paradoxes and Possibilities – Professorial Lecture*, pp 1–34. London: Institute of Education.
- Riley, K. A. and West-Burnham, J. (2004). *Educational Leadership in London*. Nottingham: National College for School Leadership. www.ncsl.org.uk/Talk2Learn (accessed July 2009).
- Un Habitat (2005). www.unhabitat.org/whd/2005/background-information.asp (accessed July 2009).
- Whitty, G., Power, S., and Halpin, D. (1998). *Devolution and Choice in Education: The School, the State and the Market*. Buckingham: Open University Press.
- Winder, R. (2004). *Bloody Foreigners: The Story of Immigration to Britain*. London: Little Brown.
- Wragg, E. C. (2005). *Letters to the Prime Minister*. London: Central Books.
- Willmott, P. and Young, M. (1957). *Family and Kinship in East London*. London: Penguin Books.

Further Reading

- Riley, K. A. and Emery, H. (2007). Reconfiguring urban school leadership: Some lessons from London. In Brighouse, T. and Fullich, L. (eds.) *The Global Challenges of the City*, pp 169–196. London: Bedford Way Papers.

Leadership for Learning

J MacBeath, University of Cambridge, Cambridge, UK

© 2010 Elsevier Ltd. All rights reserved.

In the last decade leadership for learning has become commonplace within the educational discourse, yet has been accompanied by little empirical research able to identify what that phrase means either in theoretical terms or in day-to-day school practice. It is perhaps taken simply as self-evident that in an educational milieu this is what leadership is for and is what leaders do. Yet, the connections remain problematic and largely untheorized.

The first systematic attempts to gather data came from researchers working within a school effectiveness paradigm (Brookover, 1979; Mortimore *et al.*, 1988), seeking to identify characteristics that distinguished more-effective from less-effective schools. The proxy measure for learning employed was value-added student attainment aggregated to whole school level, while the proxy for leadership tended to refer to the qualities or behaviors of principal or senior school leaders.

Within this genre one of the most frequently quoted studies is that by Sammons *et al.* (1994), who conducted a meta-analysis of two decades of research, identifying 11 characteristics of more-effective schools. The first of these was professional leadership. This, however, left a great deal unsaid as to the nature of leadership, nor did it (located as it was within a positivist effectiveness paradigm) problematize the nature of learning. As these studies worked within a correlational framework, the connecting tissue between leading and learning remained largely unexplored.

The challenge taken up by researchers in the leadership field was to try to get closer to an understanding of forms of leadership that are most likely to contribute to student learning. Following in the effectiveness tradition, student outcomes were taken as the proxy measure for learning, while leadership was approached through qualitative exploration of what people in positions of leadership actually did or said that they did. Successive studies by Leithwood and colleagues (Leithwood and Riehl, 2003; Leithwood *et al.*, 2004; Leithwood and Jantzi, 1999) identified clusters of leadership practices which were associated with raised student achievement, such as setting directions, developing people, creating strong collaborative communities in school, nurturing the development of educational cultures, redesigning the organization, and building powerful forms of teaching and learning. These effects were described as indirect, or mediated. That is, raised achievement was explained by good teaching, which was in turn made possible by a climate of support and continuing professional development.

In Australia, Mulford's research (Mulford and Silins, 2003; Silins *et al.*, 2002) also identified indirect effects, or intervening variables, describing practices very similar to those found in Canadian and American studies, for example, collaborative climate, shared and monitored mission, and initiative and risk taking supported by appropriate professional development.

These professional cultures were associated with transformational leaders who sought to create conditions within which others would become more committed and motivated to work toward the improvement of the school. This model of leadership, it is argued (e.g., Mulford and Silins, 2003), does not address the principal's responsibility for managing the technical core of the school (teaching and learning). Instead, it is assumed that the principal is empowering others (e.g., teacher leaders and administrators) who directly manage these classroom-based processes (Hallinger, 2003). "You cannot empower people," argue Binney and Williams (1997) but one can, they suggest, create conditions in which people feel empowered to take ownership and initiative.

Underpinning much of the literature on mediation is the assumption that student learning, or more specifically enhanced outcomes, is the singular end purpose of effective leadership. In other words, mediation, as the term suggests, is simply instrumental to the primary goal (the technical core) of student achievement.

This brings sharply into the foreground the nature of learning and the question "Who is learning for?" Knapp *et al.* (2003) offer a tripartite model of learning, referred to as the wedding cake because of its layering of interconnected forms of learning (**Figure 1**).

In this model, student learning is not the singular purpose of schooling. While it is nourished through the iterative process of professional and system learning, these are not simply means to an end but ends in themselves. Indeed, it is argued that professional learning vouchsafes a longer term educational future than by focusing on the achievement of any single cohort of students. A system that is able to go on learning is one that can serve and sustain a much wider constituency of stakeholders.

In Wenger's (1998) conception of communities of practice, the emphasis is on the interchange of craft and professional skill and knowledge, enhancing teacher learning, and building systemic capacity. Capacity is a key idea in theories of organizational learning, the primary goal being the cultivation of a continuing search for

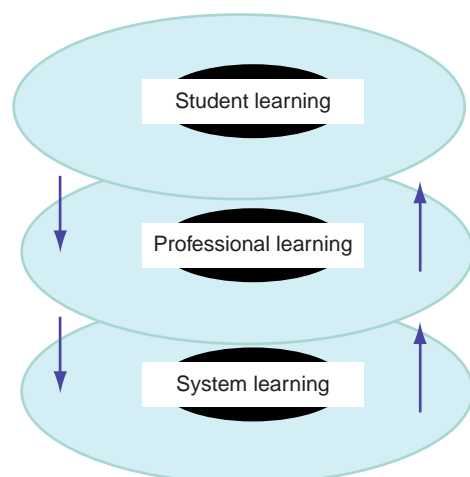


Figure 1 The Learning Wedding Cake.

understanding and meaning, whether in a business enterprise (Senge, 1990), a public agency, or a school (Hargreaves, 2003; 1995). An organization committed to learning is one which challenges its own premises, its structural conventions, its curriculum rationale, and the very nature of learning itself. Argyris and Schön (1978) describe a process of double-loop learning, a form of conceptual escape from the single loop of planning, implementation, and evaluation into a more reflective and self-critical cycle.

Organizational learning is a slippery concept, and difficult to realize within an education policy context counsel. Silins *et al.* (2002) arguing that an organization cannot learn and suggested that it is people who learn. However, the case for sustainability rests on a belief that in a learning organization people build structures and create cultures which prevail beyond the lives of the individuals who created them. Hargreaves (2006) goes on to argue that a preeminent task of leadership is to recognize one's own mortality and to invest in succession planning for a future in which one will personally play no part. Working counter to that ethic, however, is the tendency for a new principal to wish to leave his or her own deep imprint on the school, yet in the process potentially disempowering and disabling the organization. Senge (1990) categorizes a number of organizational learning disabilities among which he cites "I am my position," that is, a disabling of agency by reference to status and institutional authority as encapsulating leadership. The fallibility of heroic and charismatic leadership is documented in a military context by Dixon (1994) who categorized 15 incompetences, all of which are antithetical to individual, group, or organizational leadership for learning and have been found to be applicable in both a business and school context (MacBeath, 2006).

This stream of literature with its focus on systemic and hegemonic factors invites us to revisit conceptions of both leadership and learning and to reconsider whether

researchers may be starting from two false premises – on the one hand, equating leadership with those at the apex of the organizational pyramid (Murphy and Forsyth, 1999) and, on the other, equating learning with what students achieve through the relatively crude measures of school-administered tests (Fielding, 2001).

Leadership for Learning as Distributed

In the last half decade, distributed leadership is a concept that has been much written about although relatively little researched. It pushes the locus of leadership beyond principals, and senior managers, extending leadership roles (Spillane, 2006) to middle management and to teachers both as individuals and groups – Gronn (2003) characterizes this as the additive conception of distribution. It is a formal and essentially hierarchical form of distribution in which leadership is handed out by those in senior positions of responsibility to people in given roles, clearly defined and bounded by individual responsibility and accountability. By contrast, holistic distribution looks for a synergy arising through concertive action, stretching leadership function across the social and situational contexts of the school (Gronn, 2003). In other words, rather than defining leadership by reference to where people are located in the hierarchy, team leadership is essentially different. Rather than downward delegation to specific individuals (Katzenbach and Smith, 2003), it celebrates individual differences and sees these as fostering collective strengths. In a similar vein, Mulford and Silins (2003) argue that distribution is concerned with achieving the right dynamics among people, developing social trust, an essential precondition for development and learning for change.

Teacher Leadership for Learning

Literature on teacher leadership comes at the issue from both an additive and a holistic perspective. In 1986, a seminal document titled *A Nation Prepared: Teachers for the Twenty-First Century* argued that curriculum leadership should lie with those closest to student learning and that the real power to improve achievement lay with teachers, who should be entrusted with new responsibility and accountability for change. While the empowerment of teachers, together with an emphasis on accountability, may be seen as something of a double-edged sword, it is also one which positions teachers as leading and shaping the purpose of learning in a new millennium.

The holistic perspective, as represented by Lieberman and Miller (2004), posits a major transformational shift in respect of three interrelated strands: first, a shift from individualism to professional community; second, a shift from teaching at the center to learning at the center; and,

third, a shift from technical and managed work to inquiry and leadership. Taken together, these represent a pendulum swing from prescription of curriculum to capacity building of teachers, construing teachers as members of professional communities and assuming collective leadership which is learning centered. In Mitchell and Sackney's (2000) description of learning teams, they are initiators in discussion of "tough problems and deep mysteries of teaching and learning."

This unrealized capacity is referred to by Katzenmeyer and Moller (1996) as a sleeping giant, because both leadership and learning potential lie dormant, constrained by inert ideas (Whitehead, 1929) and by the inertia of traditional structures. Mindset and structures are, to an extent inseparable, an example of which is cited by Frost (2005) who describes conventional approaches to professional development as being provided for, or done to, teachers rather than teachers as the prime movers.

If teachers do not see themselves as prime movers, as leaders of learning, in their school and not just their classrooms, it may be due to deeply institutionalized conceptions of upward accountability. An alternative narrative proposes lateral accountability (Fullan, 2001), which comes through teachers focusing collectively on student learning with discussion and planning of what it will take to get there. The induction of new members into such a learning community allows opportunities for rehearsing new roles and taking risks in a community of support and assistance (Lieberman and Miller, 2004). Leadership is open to even the neophyte who may bring in new, perhaps naive but fresh, perspectives to the organization he or she is joining.

Without such opportunities, teaching may become a lonely and disempowering activity. For new entrants to the profession from industry, Johnson (2005) found that they brought with them an expectation of teamwork, seeking opportunities for expanded influence, and sharing leadership responsibility, as they were accustomed to in their previous settings. Yet, they found themselves working in isolated classrooms, robbed of the initiative and latitude of decision making they had previously enjoyed. As Elmore (2003: 18) argues, "these environments can operate to pull the organization toward higher levels of collective action and higher levels of student learning, or they can operate to pull schools apart into highly variable and weak learning environments." This stress on synergetic properties of leadership for learning is echoed in the work of Spillane *et al.* (2001), in which collective cognitive properties of a group working together to enact a particular task, forge out of their work a leadership practice greater than the sum of each individual's practice. "Consequently, to understand the knowledge needed for leadership practice in such situations, one has to move beyond an analysis of individual knowledge and consider what these leaders know and do together" (Spillane *et al.*, 2001: 12).

Student Leadership for Learning

Students are the largest untapped knowledge source in a school, writes SooHoo (1993: 23) "the treasure in our very own backyard," arguing that a school which overlooks that intelligence source is inevitably poorer as a consequence. Historically, students have been the last to be consulted about school quality and effectiveness, the last to be commentators on learning, and the last to assume leadership roles; yet, as a developing literature attests, they may get closest to the heartbeat of the school (Fielding, 2004).

Jorgensen's (2004) ladder of participation offers a practical framework for identifying how student voice and leadership are conceptualized and operationalized within a school. The six rungs of the ladder represent a hierarchy of participation from the lowest level – adults decide and inform students – to the highest level – students decide and adults support. The intermediate rungs ascend in turn from adults use students as decoration (a ritualistic illusion of participation); adults consult, then decide on their own premises; adults decide and take students' views into account; and students decide together. As Fielding and Rudduck (2002) comment, it is a strategy of the fearful to limit student involvement to aspects of school life which are seen as relatively safe or which do not have significant impact on the work of adults in the school. However, each rung may be seen as a successive stage in the maturing of a learning school in which thought is given to the nature of tasks or opportunities which might fall into any one of those six categories. In English primary and secondary schools, it is increasingly common for students to be involved in the interviewing of new staff, to conducting classroom observation, taking a lead in evaluation, as well as in the design and refurbishment of playgrounds, toilets, common rooms, and classrooms. In recent years, it has been increasingly recognized that they have an important contribution to make in evaluating the quality of teaching and learning (MacBeath and Sugimine, 2002).

Such learning and leadership opportunities often reflect Gronn's additive model, a trickle-down conception of leadership, casting students within a respondent or consultative frame, their place clearly demarcated as the lowest rung in the organizational hierarchy with ritual occasions in which their viewpoint may be heard.

Hearing what pupils have to say about teaching, learning and schooling enables teachers to look at things from the pupil perspective – and the world of school can look very different from this angle (Rudduck and Flutter, 2004).

In the holistic model, however, the roles of student, teacher, and senior leader are less neatly circumscribed. There will be occasions when students teach and teachers learn (exploring the scope and potential of information and communication technology (ICT) is an obvious example), times when teachers lead and senior leaders

follow (e.g., around issues of pedagogy), and occasions when there is collective enterprise as, for example, when schools are engaged in projects, productions, community activities, school trips, or residential experiences. Musical and sporting events are often viewed as peripheral extra-curricular activities but are potentially rich in opportunities for student leadership, for new ways of learning and leadership for learning (MacBeath *et al.*, 2001). Collective enterprise of this nature can build social capital founded on establishing trust and mutual respect rather than deference and knowing one's place.

Social capital flows from the endowment of mutually respecting and trusting relationships, which enable a group to pursue its shared goals more effectively than would otherwise be possible. . . It can never be reduced to the mere possession or attribute of an individual. It results from the communicative capacity of a group (Szreter, 2004: 575).

Leadership for Learning, Without Limits

The more leadership for learning is conceptually, and in practice, locked within the black box of the schools, the greater the limitations it places on learning and leading for learning. Not only does most of learning occur outside of the school walls, in family and community (Epstein, 1999; Weiss and Fine, 2000; Castells, 2000), but many young people and teachers are leaders in other out-of-school contexts, in families and extended families, in clubs and sporting activities, in uniformed organizations, such as Guiding and Scouting, and in the gray economy, through gangs and entrepreneurial activities which offer routes back into society through the backdoor, what Castells (2000) refers to as perverse integration.

A recent 3-year study in England of schools in exceptionally challenging circumstances (MacBeath *et al.*, 2006) found that despite heavy government investment and support to turn these schools round, it was the impact of what happened outside of schools that explained rises in student achievement over a 3-year period. The most successful leadership was in schools that built strong community links and invested in interagency collaboration, seizing opportunities for sharing leadership with voluntary community and social workers.

In these schools on the edge, leadership for learning was outward looking, to parents, community, to public opinion, to media influence, and to local and national policy. It raises the question of how far does the responsibility of school leaders extend? Should school leaders have a core role in ensuring young people and their families value what is really important in life? As defining and shaping public opinion and public expectations as to the nature of learning, effects, and outcomes of schooling? For some (Starratt, 1998; Carmichael *et al.*, 2006) this

wider impact is considered to be one of the key tasks and moral responsibilities of school leadership.

A 3-year study in England for the Economic Social and Research Council (ESRC) (James *et al.*, 2006) was able to show that individuals, who had a role in external initiatives, were able to create links through their own personal and social networks outside the school. These intensional networks were not *ad hoc* but created, managed, and sustained deliberately. One conclusion of the research was that effective leaders found ways of creating opportunities for staff to create and exploit those intensional networks, supplementing those of the senior leadership team. "We know from the literature," conclude Carmichael *et al.* (2006) "that these different kinds of links, networks, and network roles are important in knowledge creation and sharing."

Leadership for Learning in International Perspective

Issues of leadership at student, teacher, organizational, and system level were the focus of a 3-year international project (MacBeath, 2006), seeking to explore what leadership for learning meant to schools in eight different sites in seven countries. What became immediately apparent from the outset was that of all of the three words contained in that short phrase – leadership, learning, and "for" were understood very differently in the 24 schools from eight world cities – Athens, Brisbane, Copenhagen, Innsbruck, London, Oslo, Seattle, and Trenton (New Jersey).

There was, for example, no English equivalent to the Germanic concept of *Bildung* or the Danish concept of *Dannelse*, the latter broadly equivalent to the conditions which help children to become citizens in a democratic society, a notion deeply embedded in a Nordic culture. American terminology of instruction and instructional leadership proved problematic not only to non-English speakers but to English and Australian participants as well. For the Austrians, translations of leader and leadership (*Fubrer* and *Fubrung*) were ideas off-limit, while the Danes and Norwegians shared an antipathy to conceptions of individualistic, or heroic, leadership. Working through these differing perspectives to forge a common vocabulary and shared understanding, led to a set of principles, tested by teacher and students in the daily round of school and classroom life and progressively reframed over the last 2 years of the project.

Five Principles of Leadership for Learning

The five principles which have been elaborated elsewhere (MacBeath, 2006) not only arose from grounded work with schools, but also owed much to theoretical work discussed above. A focus on learning was the first and

guiding principle, framed as a process of continuing exploration and deepening understanding as to what learning is most worth, how it plays out in practice, and who the learners are. This main principle was elaborated in a set of subprinciples emphasizing that everyone can be a learner, that the effective interplay of emotional, social, and cognitive processes is highly sensitive to context and to the differing ways in which people learn, and that the capacity for leadership arises out of powerful learning experiences.

The second principle was framed in terms of leadership for learning as creating and sustaining conditions favorable to learning as an activity in which the culture of the school nurtures the learning of all its members; creates physical and social spaces that stimulate and celebrate learning, enabling everyone to take risks, cope with failure, and respond positively to challenges; and developing tools and strategies which enhance thinking about learning and the practice of teaching.

The third principle emphasizes the role played by dialog – the process by which learning and leadership are made explicit, discussable, and transferable, not only conceptually but also in practical application. Collegial inquiry focuses on the link between learning and leadership, addressing factors which inhibit and promote learning, celebrating the differing perspectives that arise through networking with researchers and practitioners across national and cultural boundaries.

The fourth leadership for learning principle practice involves the sharing of leadership by creating structures which invite participation, encouraging all members of the school community to take the lead as appropriate to task and context; drawing on the experience and expertise of staff, students, and parents as resources and promoting collaborative patterns of work and activity across boundaries of subject, role, and status.

The fifth principle holds that leadership for learning implies accountability by taking account of political realities and exercising informed choice as to how the school tells its own story to external audiences. This is founded on developing a shared approach to internal accountability as a precondition of accountability to external agencies, maintaining a focus on evidence and its congruence with the core values of the school; reframing policy and practice when they conflict with core values; embedding a systematic approach to self-evaluation at classroom, school, and community levels and maintaining a continuing focus on sustainability, succession, and leaving a legacy.

The Policy Imperative

The common strand across all five principles is a sense of agency. It is agency that connects leaders and learners and leading and learning as activities. The further school

leaders progressed in their shared understanding of agency as the connecting tissue between leadership and learning, the more acutely it brought to the surface the dilemmas facing practitioners, directly in the line of fire of social and political forces that placed sharply defined limits on their freedom to act according to their professional judgment or conviction. While the degree of constriction of movement differs considerably from one country context to another, as Giddens' (1984) has argued in his theory of action, structures can be modified by the agency of individuals if there is a will, a source of support, and a commitment to a set of uncompromised values.

Both leadership and learning are social activities, in schools rich in opportunities for reflection and celebration and in which learning dispositions and behavior are modeled, made visible, and internalized. This is the hallmark of a learning community (Mitchell and Sackney, 2000) in which there is a transparent openness to learning of all its members and which is accessible from the behaviors and intelligences of people with whom we share and create ideas, who lead, and who follow. Dialog (meaning flow through it; Bohm, 1983) is something qualitatively different from conversation, discussion, and debate. It is a big idea and hard to achieve in a school environment in which deep listening is not commonplace and in which there has traditionally been a low level of tolerance for dissent. In a school devoid of dialog, Elmore argues:

The transfer of agency from teacher to student is minimal because the nature of the task locates the knowledge with the teacher and the obligation to learn with the student—knowledge is transferred, agency over learning is not (Elmore, 2003: 8).

Leadership for learning, with the added edge of international dialog, challenges the assumptions that are brought to how and what we see and the judgments we bring to it. Czarniawska coins the term *outsidedness* to denote a form of knowing – by difference rather than by similarity.

It aims at understanding not by identification ('they are like us') but by the recognition of differences – 'we are different from them and they are different from us; by exploring these differences we will understand ourselves better' (Czarniawska, 1997: 62).

Self-evaluation is the process by which schools make their the intellectual and moral journey, measuring the distance they have traveled not in the simplistic trajectory of aggregated attainment scores, summative tools that say little about deep learning. The tools of authentic, professionally driven self-evaluation, by contrast, are set in a social context. They encourage dialog; serve a primarily formative purpose; are congenial, flexible, and adaptable to new situations and new challenges; are not restricted to what happens in classrooms or to students' learning; apply

to teacher and organizational learning; measure how teachers are progressing in their thinking and practice and how the school is developing as a community of learners; and they relish accountability because it is the platform for telling a story rooted in evidence of the most profound kind.

Bibliography

- Argyris, C. and Schön, D. (1978). *Organizational Learning: A Theory of Action Perspective*. Reading, MA: Addison-Wesley Ball.
- Binney, G. and Williams, C. (1997). *Leaning into the Future*. London: Nicolas Brealey.
- Bohm, D. (1983). *Wholeness and the Implicate Order*. New York: Ark Paperbacks.
- Brookover, W. (1979). *School Social Systems and Student Achievement; Schools Can Make a Difference*. New York: Praeger.
- Carmichael, P., Fox, A., McCormick, R., Procter, R., and Honour, L. (2006). Teachers' professional networks in and out of school. *Research Papers in Education* 21(2), 217–234.
- Castells, M. (2000). *The Rise of the Network Society, Vol. I: The Information Age: Economy, Society and Culture*, 2nd edn. Oxford: Blackwell.
- Czarniawska, B. (1997). *Narrating the Organization: Dramas of Institutional Identity*. Chicago, IL: University of Chicago Press.
- Dixon, N. (1994). *On the Psychology of Military Incompetence*. London: Pimlico.
- Elmore, R. (2003). *Agency, Reciprocity, and Accountability in Democratic Education*. Boston, MA: Consortium for Policy Research in Education.
- Epstein, J. (1999). *School, Family and Community Partnerships: Preparing Educators and Improving Schools*. Oxford: Westview.
- Fielding, M. (ed.) (2001). *Taking Education Really Seriously: Four Years Hard Labour*. London: RoutledgeFalmer.
- Fielding, M. (2004). Transformative approaches to student voice: Theoretical underpinnings, recalcitrant realities. *British Educational Research Journal*, 1469–3518 30(2), 295–311.
- Fielding, M. and Rudduck, J. (2002). The transformative potential of student voice: Confronting the power issues. *Paper Presented at the British Educational Association Annual Conference*, September.
- Frost, D. (2005). Resisting the juggernaut: Building capacity through teacher leadership in spite of it all. *Leading and Managing* 10(2), 83.
- Fullan, M. (2001). *The New Meaning of Educational Change*. Toronto, ON: OISE Press.
- Giddens, A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley, CA: University of California Press.
- Gronn, P. (2003). *The New Work of Educational Leaders: Changing Leadership Practice in an Era of School Reform*. London: Paul Chapman.
- Hallinger, P. (2003). Leading educational change: Reflections on instructional and transformational leadership. *Cambridge Journal of Education* 33(3), 329–351.
- Hargreaves, A. (1995). Renewal in the age of paradox. *Educational Leadership* 52(7), 14–19.
- Hargreaves, D. (2003). From improvement to transformation. *A Paper Presented at the 2003 Conference of the International Congress for School Effectiveness and Improvement (ICSEI)*, Sydney, New South Wales, 5th–8th February 2003.
- Hargreaves, A. (2006). Dimensions of leadership. www.nationalpriorities.org.uk/Resources/Miscellaneous/NCSL/index.html. (accessed June 2009).
- James, M., Black, P., Carmichael, P., et al. (2006). *Learning How to Learn: Tools for schools*. London: Routledge.
- Johnson, S. M. (2005). *Can Professional Certification for Teachers Reshape Teaching as a Career?* Bloomington, IN: Phi Delta Kappan.
- Jorgensen, P. S. (2004). Children's participation in a democratic learning environment. In MacBeath, J. and Moos, L. (eds.) *Democratic Learning: The Challenge to School Effectiveness*, pp 113–131. London: Routledge.
- Katzenbach, J. R. and Smith, D. (2003). *The Wisdom of Teams: Creating the High-Performance Organization*. New York: Harper Collins.
- Katzenmeyer, M. and Moller, G. (1996). *Awakening the Sleeping Giant*. Thousand Oaks, CA: Sage/Corwin Press.
- Knapp, M., Copland, M., and Talbert, J. E. (2003). *Leading for Learning: Reflective Tools for School and District Leaders*. Washington, DC: Center for the Study of Teaching and Policy, University of Washington.
- Leithwood, K. and Jantzi, D. (1999). A review of transformational school leadership research 1996 to 2005. *Paper Presented at AERA*, Montreal, April.
- Leithwood, K. and Riehl, C. (2003). What do we already know about successful school leadership? *Paper Presented at AERA*, Chicago, April.
- Leithwood, K., Seashore Louis, K., Anderson, S., and Wahlstrom, K. (2004). *Review of Research: How Leadership Influences Student Learning*. New York: The Wallace Foundation.
- Lieberman, A. and Miller, L. (2004). *Teacher Leadership*. San Francisco, CA: Jossey-Bass.
- MacBeath, J. (2006). *Leadership as a Subversive Activity*. ACEL Monograph Series. Winmaelee: Australian Council for Educational Leaders.
- MacBeath, J., Gray, J. M., Cullen, J., Frost, D., Steward, S., and Swaffield, S. (2006). *Schools on the Edge: Responding to Challenging Circumstances*. London: Paul Chapman.
- MacBeath, J., Kirwan, T., Myers, K., Smith, I., and McCall, J. E. (2001). The impact of study support. A study into the effects of participation in out-of-school-hours. *Research Report RR273*. London: Department of Education and Skills.
- MacBeath, J. and Sugimine, H. (2002). *Self-Evaluation in the Global Classroom (What's in It for Schools?)*. Oxford: RoutledgeFalmer.
- Mitchell, C. and Sackney, L. (2000). *Profound Improvement*. Lisse: Swets and Zeitlinger.
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D., and Ecob, R. (1988). *School Matters: The Junior Years*. Berkeley, CA: University of California Press.
- Mulford, B. and Silins, H. (2003). Leadership for organisational learning and improved student outcomes. *Cambridge Journal of Education* 33(2), 175–195.
- Murphy, J. and Forsyth, P. B. (1999). *Educational Administration: A Decade of Reform*. Thousand Oaks, CA: Corwin Press.
- Rudduck, J. and Flutter, J. (2004). *How to Improve Your School: Giving Pupils a Voice*. London: Continuum.
- Sammons, P., Hillman, J., and Mortimore, P. (1994). *Key Characteristics of Effective Schools: A Review of School Effectiveness Research*. London: Office of Standards in Education.
- Senge, P. (1990). *The Fifth Discipline: The Art and Practices of the Learning Organisation*. New York: Doubleday.
- Silins, H., Zarlins, S., and Mulford, B. (2002). What characteristics and processes define a school as a learning organisation? Is this a useful concept to apply to schools? *International Education Journal* 3(1), 24–32.
- SooHoo, S. (1993). Students as partners in research and restructuring schools. *Educational Forum* 57, 386–392.
- Spillane, J. P. (2006). *Distributed Leadership*. San Francisco, CA: Jossey-Bass.
- Spillane, J. P., Halverson, R., and Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher* 30(3), 23–28.
- Staratt, R. J. (1998). Grounding moral educational leadership in the morality of teaching and learning. *Leading and Managing* 4(4), 243–255.
- Szreter, S. (2004). The state of social capital: Bringing back in power, politics, and history. *Theory and Society* 31(5), 573–621.
- Weiss, M. and Fine, L. (2000). *Construction Sites: Excavating Race, Class and Gender among Urban Youth*. New York: Teachers College Press.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. Cambridge: Cambridge University Press.
- Whitehead, A. N. (1929). *Aims of Education*. New York: Free Press.

Further Reading

- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist* **41**(10), 1040–1048.
- Egan, K. (1997). *The Educated Mind*. Chicago, IL: University of Chicago Press.
- Hargreaves, A. and Fink, D. (2006). *Sustainable Leadership*. San Francisco, CA: Jossey-Bass.
- James, M. (2005). Insights on teacher learning from the Teaching and Learning Research Programme (TLRP). *Research Papers in Education* **20**, 105–108.
- MacBeath, J. and Cheng, Y. C. (2007). *Leadership for Learning: International Perspectives*. Rotterdam: Sense Publishers.

Leadership in Diverse Cultures

A Walker and P K Yu-kwong, Hong Kong Institute of Education, Hong Kong SAR, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Successful schools, regardless of their location or situation, require good leadership and management. What this entails, and indeed what it looks like, however, differs in important ways across and between different societies and cultures. Awareness that leadership may be conceptualized differently in diverse contexts has prompted calls for more explicit analysis of educational leadership using wider, less restrictive lenses, such as societal culture (Bajunid, 1996; Cheng, 1995a; Dimmock and Walker, 1998a; Hallinger and Leithwood, 1996). Leadership in diverse cultures is taken here to refer to leading in cultures outside Anglo-American societies where the vast majority of widely articulated and disseminated leadership theory emanates. Focusing on diversity between societies does not imply that diverse cultures do not also reside within societies, and indeed schools. Although this article does not delve into leadership within national intercultural settings, it may contribute messages for leaders in such circumstances (Walker and Chen, in press).

The worth of investigating leadership in diverse cultures rests on at least three basic, interrelated assumptions. The first is that leadership makes a difference in schools, even as we remain unsure of exactly how this works. This assumption is founded on international literature that confirms the centrality of school leadership to school improvement and quality schooling; and that it most effectively influences school outcomes indirectly through multiple variables (Hallinger and Heck, 1999; Leithwood *et al.*, 2006; Southworth, 2005). It is also now widely accepted that how leaders make a difference is contingent upon the context within which they lead. The second assumption therefore is that multiple contexts influence how school leaders lead. In other words, what leaders do is mediated and moderated by both their personal internal states as well as the organizational and external milieu of the school (Leithwood *et al.*, 2006; Cheung and Walker, 2006). This context is not only complex, but is also constantly shifting and evolving in response to factors such as personality, ethnicity, gender, politics, history, economics, and culture (Rizvi, 1997). The interplay between these value sets plays out in an assortment of forms in schools as leaders attempt to make sense of what is needed (Bottery, 2004).

Third, given that leadership is centrally concerned with the interpretation and enactment of values, it is reasonable to assume that one influential contextual factor on how

leadership is conceptualized and exercised is the cultural values, norms, and beliefs which help to define the group or society within which they live and work. Although a contested construct, culture can be broadly defined as patterns of shared values, beliefs, and norms held by a particular group and/or society which combine in various ways to influence behavior and action. An acceptance that culture matters, however, as with leader effects, still leaves us unsure just how much it matters – this will probably never be definitively resolved. However, the bottom line appears to be that even though cultural values exist within a complex and vibrant globalizing context, they continue to exert a strong influence on people's lives (House *et al.*, 2004). As such, they form a key element of the environment within which leaders lead and on how they lead. As Dorfman *et al.* state:

While we acknowledge that global communication, technical innovation and industrialisation can create a milieu for cultural change, a convergence among cultural values is by no means assured. In fact, cultural differences among societies may be exacerbated as they adapt to modernisation while simultaneously striving to preserve their cultural heritage. (Dorfman *et al.*, 2004: 709)

Growing Interest in Leadership in Diverse Cultures

Interest in educational leadership across diverse societal contexts has until recently been of peripheral interest only to scholars in the field. The majority of available literature on school leadership relates to the social and organizational structures and values prevalent in Western English-speaking societies. This often fails to adequately recognize that these structures and cultures can look very different across systems and societies. With some exceptions, the same has been true of international studies in both school effectiveness (Teddie and Reynolds, 2000) and school improvement (Reynolds, 2000).

Over the last decade or so researchers shifted some attention to leadership in diverse international settings and cultures (e.g., Beare and Boyd, 1993). This interest has been prompted partly by the heightened visibility of large-scale international comparative studies, such as Program for International Student Assessment (PISA) (Ho, 2006) and Trends in International Mathematics and Science Study (TIMSS) (Baker and LeTendre, 2005), acknowledgement

that good leadership is essential if schools are to improve, regardless of their context, the increasing diverse makeup of Western societies themselves and the vast expansion of electronic, written, and physical access to diverse knowledge.

Investigation has taken a number of forms and traversed various national and cultural terrains. For example, scholars have argued broadly and persuasively for recognition of indigenous approaches to educational administration (Hallinger and Leithwood, 1996), the synthesis of Western with indigenous understandings (Sapre, 2000), the dangers of simplistic knowledge transfer across cultures (Shaw, 2005; Walker and Dimmock, 1999), and the need for cross-cultural comparative frameworks to guide research in the area (Hallinger and Leithwood, 1996; Dimmock and Walker, 1998b, 2005; Heck, 2002).

Others have examined the principalship in specific cultural contexts through the lens of politics (Palomares and Castillo, 2004), philosophy (Begley, 2000; Wong, 2001), international schooling (Jabal, 2006), gender (Celikten, 2005; Oplatka, 2002, 2006; Oplatka and Hertz-Larowitz, 2006), social justice (Jansen, 2006), and democracy. The latter has been used, for example, for analysis of the roles of principals in Spain and Nordic Europe, where it appears to infuse most discussions of educational policy and leadership (Moller, 2006, 2007; Persson *et al.*, 2005). In Spain, Llamas and Serrat (2002) decried the advancement of education leadership as opposed to organizational control in Spanish schools. This finding is mirrored across other more-developed (Shen, 2001) and developing societies (Bush and Oduro, 2006).

Empirical work has also begun to appear in English from diverse societies largely unsighted in the literature a decade ago. Examples included Saitis and Menon's (2004) research in Greece into teacher perceptions of future principals and Al-Hamdan and Al-Yacoub's (2005) investigation into the evaluation of educational leaders in Kuwait. Research into the principalship is also beginning to emerge from the Chinese Mainland (Cheng and Wong, 1996; Ribbins and Zhang, 2004; Wong, 2006), Hong Kong (Law and Walker, 2005; Walker and Dimmock, 2002), Portugal (Pashiardis *et al.*, 2005), Malta (Bezzina, 1998), and into school management and related areas in Cyprus (Pashiardis and Ribbins, 2003).

Studies of leadership in diverse cultures written in English are by nature comparative. An emerging body of literature has set out to more explicitly compare different notions of leadership across a number of societal contexts. Such approaches, or collections of work, include a focus on successful principals (Jacobson *et al.*, 2005), beginning principals (Walker, 2006b), leadership in developing societies (Oplatka, 2004), leadership and values (Begley and Wong, 2001), leadership and ethics (Begley, forthcoming), leadership development (Hallinger, 2003; Huber, 2004), and leadership and reform (Townsend and Cheng, 2000), to name but some. Although comparisons

are yet to match either the scope or sophistication of international business studies such as those by Hofstede (2005), Trompenaars and Hampden-Truner (2000), or the more recent Global Leadership and Organizational Behaviour Effectiveness (GLOBE) Study (House *et al.*, 2004), which investigated the influence of cultural values on leadership across 62 societies, the trend is certainly in motion.

Leadership in Diverse Cultures

Five themes – leadership for change, leadership in developing societies, successful leadership across diverse cultures, leader development, and the influence of societal cultures – can be used to illustrate recent work on school leadership in diverse cultures. Whereas these themes are addressed under separate headings for ease of presentation, and reflect the general state of play in the area, they are somewhat artificial in that equally strong patterns cut across as well as between them. For example, although the influence of culture is discussed separately, it certainly influences how leaders deal with change, frame their own professional development, and lead successfully, or otherwise. It also influences leadership in diverse developing societies. Themes, therefore, should be viewed as interrelated and overlapping (as they are often presented in the literature). What they hold in common is that they seek to capture some of the intricacies of what leadership looks like across diverse cultures, how it is enacted, and why it plays out in particular ways. The task as set by Stambech is:

not one of identifying what is universal or converging, nor to label and minutely specify what is unique about each situation, but to address how locally interpreted narratives give force to universal categories and how universal categories give force to local narratives.

(Stambeck, 2003: 157)

Leadership for Change in Diverse Cultures

The place of the principal in education reform has sparked considerable interest internationally (Hallinger, 2004). Much of this relates to the role of principals in increasingly decentralized systems (Hanson, 2000) and more school-specific changes such as school-based management (Hui and Cheung, 2006).

Although decentralizing reforms come in an assortment of shapes and sizes and follow varying rationales across diverse cultures, they basically aim to transfer greater responsibility from central authorities to the principal and, often simultaneously, teachers and the wider community (Walker, 2004). This has had a marked impact on principals' jobs in societies such as Singapore (Chew *et al.*, 2000), Taiwan, and Hong Kong. Yang (2001) explains that the role of Taiwanese principals has become diverse

and paradoxical as these principals are not only charged with control and responsibility for curriculum, personnel, and budget, but also expected to share decision making with parents, teachers, and communities.

In a study of education reform in Southeast Asia over a 10-year period, Hallinger (2004) identified a list of obstacles for effective change. These included the lack of a systematic perspective, coordination and implementation, and communication and vision. While noting that such blockages also typically emerge from research in the US, he made the key point that, “more significant than differences in the *types* of obstacles encountered were differences in their *character as presented in practice*” (p. 5). The obstacles sprouted from a mixture of instrumental, political, and cultural factors which, on their own, neither enhance nor retard change. Thus, differences in leadership across diverse cultures stem from both deeply held values and their enactment in practice.

Leadership in Developing Societies

Oplatka (2004, 2006) heads a number of scholars focusing specifically on leadership in developing societies – an emerging area of study which digs deeper than traditional single-site, regional, or large-scale cross-national studies conducted by international agencies. Oplatka (2004) defines developing societies as those outside of Europe and North America (and a few other nations such as Australia, New Zealand, and Japan). “These (developing countries) countries were ruled by Europeans for a long time (with the exception of countries such as Thailand), their economy is more agricultural-based, and are usually characterized by high mortality rates, high birth rates, high levels of poverty and large gaps between rich and poor” (p. 428).

Oplatka (2004) analyzed English-language articles from a wide range of journals published since about 1990. Her review produced a rich portrait of principals across diverse cultures and identified a number of common similarities and differences both within developing societies and between these developed societies. One marked similarity between developing societies was the limited authority held by principals, at least beyond their immediate organizations, and a related tendency to adhere strictly to central directions and superiors’ instructions. The same was identified within schools, that is, principals both expect and are expected to practice a strong almost autocratic style of leadership (Cheung, 2000). The research by Hallinger *et al.* (2000) in Thailand supports this:

Sometimes even if we want to involve staff more actively in determining the direction and procedures for change they misunderstand us. If we really spend a lot of time asking their opinions, they even begin to think we don’t

know how to do our job! It’s as if they say, “You’re the principal. If you have to ask me, it must mean that you don’t know what you’re doing.”

(Hallinger *et al.*, 2000: 218)

A further similarity (excluding some East Asian societies) was that principals in developing countries focused predominantly on routine-control mechanisms, rarely involved others in decision making and did not appear to focus on areas such as instructional leadership, vision building, or shared leadership, as are widely espoused in Anglo-American literature (Oplatka, 2004).

A major difference between principals in different developing societies was that leaders in East Asia were perceived as playing a major role in large-scale reform efforts, whereas African, Moslem, and Mexican principals played a rather minor role. Cheng *et al.* (2003) support this finding, “[in the Asia-Pacific region] leaders are often perceived as the key actors mobilising their institutions and members at the site-level to face up with those challenges and make educational services and provision more quality effective and accountable” (p. 922). A similar difference between East Asian and other developing societies was the exercise of instructional leadership and community involvement in schools; this was linked to rapid economic growth.

When concluding her analysis of principals in developed and developing societies, Oplatka summarized:

When role expectations and position are compared, the two groups of nations vary substantially in terms of the image of principalship, its position and status, principals autonomy, attitudes towards change initiation and parental involvement in school life. . . . there were great differences between role behaviours of principals from developed and developing countries. In contrast to the democratic spirit in the former, non-western principals are likely to employ autocratic, non-participative, summative evaluation. . . this may be ascribed to different contextual and cultural features in both group of nations. (Oplatka, 2004: 441,442)

Successful Leadership in Diverse Cultures

Principals’ leadership practices are common across contexts in their general form but highly adaptable and contingent in their specific enactment (Leithwood, 2005: 622).

Another approach to understanding leadership in diverse cultures is being undertaken under project titled International Successful School Principals Project (ISSPP) (see Jacobson *et al.*, 2005). In simple terms, ISSPP set out to “investigate whether ‘success’ in one country was ‘success’ in another”, how customs, ideologies, and practices affect leadership, how success is measured, and how it plays out in the lives of principals (Day, 2005: 533). Although ongoing, the study is beginning to produce insights into leadership

across seven different societies. Leithwood (2005) pulled together and analyzed 63 qualitative ISSPP case studies around four key questions: "What practices are used by successful principals and do these practices vary across contexts; What gives rise to successful principal leadership?; Under what conditions are the effects of such practices heightened or diminished?; Which variables effectively 'link' principals' influence to student learning?" (p. 620).

He reached a number of conclusions. Among the more general of these were that principals across cultures and policy environments used a set of basic leadership practices, for example, developing a shared vision, consensus building, distributing leadership, or demonstrating high expectations, but the way these were enacted was linked in various ways according to their context. In other words, this finding echoed assertions that even though concepts may be similarly named, how they were perceived and operationalized are influenced by a principal's value sets and other internal and external mediating and moderating variables. Leithwood draws on Wong's (2005) Shanghai-nese case as an example:

The two Chinese cases (do) illustrate the range of meanings associated with concepts of collaboration and leadership distribution. . . In such cultures, people demonstrate greater respect for hierarchy, position, age and formal authority than do those in low power distance cultures. Not surprisingly, then, although the Chinese principals involved their staffs in decisions making, theirs were more often decisions about how than what. (Leithwood, 2005: 621).

Cross-societal studies into educational leadership effectiveness and success may be informed by the recent GLOBE study which investigated the influence of cultural values on leadership across 61 societies over 11 years (House *et al.*, 2004). The study aimed to develop an empirically based theory to describe, understand, and predict the influence of cultural variables on leadership and organizational processes. It attempted to show that both individuals and groups of individuals in certain societies possess an implicit leadership theory. In other words, it, "wanted to show that societal and organizational culture influences the kind of leadership found to be acceptable and effective by people within that culture" (Grovewell, 2005: 4). Whereas it is not possible to discuss in detail the study's outcomes, it identified ways in which people worldwide distinguish between leaders who are effective and ineffective, and, perhaps more importantly, the extent to which the differences in leader styles and effectiveness across societal clusters could be explained in terms of the values that prevailed in those clusters. Although certainly not without critics, the study showed quantitative differences in perceptions held about leadership across clusters of societies and, as such, may inform work in school leadership.

Leader Development in Diverse Cultures

If we accept that leadership is conceptualized and exercised differently in diverse cultures, it is axiomatic that it must also impact the values, design, and operation of Education Leader Development Programs (ELDPs) (Walker, 2006a). While describing the state of leader development internationally, recent cross-societal analysis by Wales and Welle-Strand (2005), Huber (2004), Hallinger (2003), Lin (2001, 2003), Walker *et al.* (in press), and individual studies by Slater *et al.* (2003), Brundrett and Derling (2006), Wang (2007) and Wong (2004) have begun to explore this area. Huber (2004) reported a 15-country international comparative study of ELDPs where programs emphasized finding a balance between theory and practice, involving appropriate trainers and facilitators and the importance of relating learning opportunities to diverse contexts. In a similar review of ELDPs internationally, Hallinger (2003) identified a number of themes describing current best practices in 12 (East Asian, European, and North American) societies. The themes included the movement from passive to active learning, connecting training to practice, crafting an appropriate role and tools for using performance standards, supporting effective transitions into the leadership role, evaluating programs, and, importantly, developing and validating an indigenous knowledge base across cultures (also see Walker and Dimmock, 2005).

As alluded to by Hallinger, when taken to a more microlevel, however, much development conducted outside Western contexts is still driven by frameworks, content, and pedagogies that may lack cultural validity because of the absence of an indigenous knowledge base (Walker, 2006a). For example, when discussing content used by the Center for Creative Leadership (CCL) in cross-national LDPs, Jones (2006) highlights this thus:

Simulations are, in effect, prefigured with 'rational' decisions made by 'rationale' actors with whom participants must interact. Such normative assumptions are not culturally neutral; rather, they are distinctly western in their assumptions about human behavior and the psychological backdrop of decision making. (Jones, 2006: 485)

There is considerable movement in this area internationally, and understanding is getting sharper in terms of what EDLPs look like in diverse societies. However, investigation needs to stretch beyond description to deeper understanding, including the implementation of emerging work-based approaches (Lumby *et al.*, forthcoming; Webber and Robertson, 2004). Important philosophical and technical questions remain about how such approaches travel, or the shape they may take in different societies depending on values, political structures, or resource availability. Such questions may include: Are the models and practices of leader development being used by

context-specific organizations and development programs applicable across cultural contexts? What adjustments need to be made in methods, practices, assessment, and philosophies so that they will work in or across cultures? What can be done to successfully transfer Western (or other) leadership development models and practice? And, indeed, should Western (or other) leader development models and practices be transferred? And, if so, how? (Walker (2006a), adapted from Hoppe (2004)).

Influence of Culture on Leadership

As is apparent throughout the preceding discussion, any investigation of leadership in diverse cultures is at least partly underpinned by the seemingly perpetual question of whether, and the extent to which, societal culture influences what leaders do (Emihovich, 2006); this calls specifically for greater recognition of the influence of larger macro-societal cultural values on school leadership across diverse societies arrived in the mid-1990s and followed widespread recognition in the link between school leadership and organizational culture, and the importance of context to successful school leadership. Almost simultaneously, Cheng (1995b) and Hallinger (1995) advocated the analytical properties of societal culture and its potential for expanding understanding in the field. Cheng (1995a) argued that, “the cultural element is not only necessary, but essential in the study of educational administration” (p. 99). He continued to elaborate that the literature on organizational culture could be extended to intercultural understanding, “if (the) external cultural factorsare taken into account” (p. 99). By external cultural factors, Cheng referred to national or societal cultural characteristics.

Four questions posed from outside the Western English-speaking academic world summarize some of the enduring issues driving increased understanding. These were: Is the wholesale acceptance of Western educational practices appropriate to their national goals? Are the educational practices they have adopted from the West consistent with and sustaining of their cultural heritage? What are their own intellectual traditions and indigenous approaches to education and cultural transmission? How does the indigenous knowledge embedded in their culture fit with the theories, assumptions, and practices embedded in our Western-derived educational programs? (Bajunid (1996), cited in Hallinger and Leithwood, 1996).

Recent work has asserted the influence of societal culture on school leadership and organization (Dimmock and Walker, 2005). Hallinger *et al.* (2005), for example, assert that indigenous approaches to school leadership in East Asia have been and continue to be fundamentally concerned with the moral development of individuals and the creation of community. However, in East Asian and Latin American societies the community metaphor

manifests in cultural contexts that differ markedly and systematically from North America and Europe. This contrast between diverse cultures challenges scholars to suspend assumptions as they seek to deepen the understanding of educational leadership. The wealth of richness here appears to lie in the influence of culture on how leadership is understood, constructed, and enacted.

For example, scholars such as Begley (2000) and Walker and Hallinger (in press) point out that current Western notions of democratic learning communities constructed within culturally restricted understandings may be at best puzzling for school leaders in vertically aligned culture systems, where democracy may hold very different meanings (Walker and Hallinger, in press; Jansen, 2006; Leung and Chan, 2001; Hallinger and Kantamara, 2002). Roles in and modes of participation in community building differ in a number of ways. One of these is the more distinct power gaps that exist among people in many Asian and Middle Eastern cultures. These are reflected in the daily cultural practices of schools. For example, vertically aligned cultural systems exert significant influence upon social relations in the workplace. Persons of lower status (i.e., age, position, and seniority) naturally defer to those of higher status, accepting differences in power as a normal feature of social relations. Educational leaders and followers consequently tend to be more conscious of status and hierarchy than colleagues in Western contexts, even as they are more conscious of the need to foster community. Leaders are granted respect by followers based on age, formal position, and seniority. There is tacit acquiescence among followers as long as the leader’s behaviors remain aligned with these cultural norms. This normative reciprocity preserves relationships and promotes surface harmony among the members of social groups (Hallinger *et al.*, 2005).

East Asian, Middle Eastern, and Latin American societies, be it in different ways, are recognized for their collectivist orientation. Although the meaning of collectivism is widely debated, a high collectivist orientation suggests that people form personal perspectives first and foremost in terms of their significant group associations (Hofstede, 1994). In terms of behavior in schools, this implies that individual teachers and mid-level leaders, as well as the principal, subjugate their needs, ambitions, and, if necessary, opinions for the greater good of the school.

A combination of high power distance and collectivism creates a rigorous interplay, for example, when applied to conceptions like distributed leadership and empowerment. While high power distance reinforces the formal authority accorded to principals, collectivist tendencies shape the enactment of that power (Walker and Dimmock, 2005). It is the consensus building in which leaders engage within the values and norms of their school communities that creates the legitimacy needed to act. Within a school, this may mean that teachers interact with their leaders

without engaging in open disagreement. However, this does not mean that discontent does not agitate below the surface, but to emerge it must follow often informally ascribed tracks before it is addressed. Leadership, therefore, may be sometimes more concerned with maintaining illusory rather than actual harmony. When viewed through Western leadership lenses, it can appear as if the desire for smooth relationships interferes with task achievement (Hallinger *et al.*, 2005).

Conclusion

On the back of numerous scholarly and more pragmatic interests, research into educational leadership in diverse cultures is on the move, but still has a long way to go (Heck, 1996). Despite increased output, there is still a dire shortage of studies, especially written in English, on principalship and educational leadership in non-Western cultures. This is not surprising given the difficulties of understanding and interpreting, for example, the influence of culture on leadership, the critical lack of even the most basic resources for schools in many societies, and the political and religious divisions which continue to divide them.

Uncovering cultural nuance is a difficult task, mainly because much of what shapes and reflects how leaders think and what they do is difficult to find or see. This blindness is because people are often not only unaware of the influence of culture on other groups, but also of their own culture on themselves. Hoppe (2004) explains that developing cross-cultural self-awareness is complex because culture, in essence, is internalized patterns of thinking and behaving that are believed to be natural – or merely the way things are. This is compounded by the dynamic state of culture and development internationally, nationally, and locally.

This observation highlights the permeable nature of the cultural and other boundaries that exist between schools and their surrounding societies. This is the nexus at which principals worldwide are located today. Cultural values and other contextual conditions differ between and within societies, and these differences carry over into education. To push research onward, different values must be respected, even if not fully understood. This does not suggest that every culture's values are equally desirable simply because they are the culture's values. For example, in too many locations, cultural, economic, and social stratification continues to legitimate institutionally unacceptable inequities in access to education among the poor, minorities, and females. Identification and elimination of these inequities, however, is near impossible unless we increase the understanding of why they exist across diverse contexts. In terms of educational leadership, this means working across as well as within

cultures in order to better understand how, why, and, indeed, whether, leaders make a difference.

Finally, it is noteworthy that there are many contested issues around whether and how societal culture influences leadership. Such a debate includes issues such as whether there are indeed a set of values, for example, which can be classified as Asian. For discussion of some of these, see Emihovich (2006), Pye (2000), Rizvi (1997), and Walker (2003). While acknowledging ongoing debate in the area and that there is no single set of Asian value that are represented in equal intensity across all Asian nations. We do assert that there is more commonality among the strongest values extant within Asian societies than when compared with Western societies.

Bibliography

- Al-Hamdan, J. M. and Al-Yacoub, A. M. (2005). Evaluations of section heads' performance at Kuwait secondary schools. *International Journal of Educational Management* **19**(3), 233–245.
- Bajunid, I. A. (1996). Preliminary explorations of indigenous perspectives of educational management: The evolving Malaysian experience. *Journal of Educational Administration* **34**(5), 50–73.
- Baker, D. P. and LeTendre, G. K. (2005). *National Differences, Global Similarities: World Culture and the Future of Schooling*. Stanford, CA: Stanford University Press.
- Beare, H. and Boyd, W. L. (eds.) (1993). *Restructuring Schools: An International Perspective on the Movement to Transform the Control and Performance of Schools*. London: Falmer.
- Begley, P. (2000). Cultural isomorphs of educational administration: Reflections on western-centric approaches to values and leadership. *Asia Pacific Journal of Education* **20**(2), 23–33.
- Begley, P. (forthcoming). *Guest Editor of a Special Issue: Cross-Cultural Perspectives on Authentic School Leadership*. *Educational Management, Administration and Leadership*.
- Begley, P. and Wong, K. C. (2001). *Guest Editor of a Special Issue: Multiple Perspectives on Values and Ethical Leadership*. *International Journal of Leadership in Education* **4**(1), 293–296.
- Bezzina, C. (1998). The Maltese primary school principal: An observational study. *Educational Management and Administration* **26**(3), 243–256.
- Bottery, M. (2004). *The Challenges of Educational Leadership: Values in a Globalised Age*. London: Paul Chapman.
- Brundrett, M. and Derling, A. (2006). *Guest Editor of a Special Issue: Leadership Development*. *School Leadership and Management* **26**(2).
- Bush, T. and Odoro, G. K. T. (2006). New principals in Africa: Preparation, induction and practice. *Journal of Educational Administration* **44**(4), 359–375.
- Celikten, M. (2005). A perspective on women principals in Turkey. *International Journal of Leadership in Education* **8**(3), 207–221.
- Cheng, K. M. (1995a). Excellence in education: Is it culture-free? *Keynote Address Presented at the 9th Annual Meeting of the Educational Research Association*. Singapore, 22–24 November.
- Cheng, K. M. (1995b). The neglected dimension: Cultural comparison in educational administration. In Wong, K. C. and Cheng, K. M. (eds.) *Educational Leadership and Change: An International Perspective*, pp 87–102. Hong Kong: Hong Kong University Press.
- Cheng, K. M. and Wong, K. C. (1996). School effectiveness in East Asia: Concepts, origins and implications. *Journal Educational Administration* **34**(5), 32–49.
- Cheng, Y. C., Mok, M. C., and Chow, K. W. (2003). Organization and the management of education: Development and growth. In Keeves, J. P. and Watanabe, R. (eds.) *International Handbook of Educational Research in the Asia-Pacific Region*, pp 915–930. Dordrecht: Kluwer Academic Publishers.

- Cheung, M. B. (2000). Securing a better future: A Hong Kong school principal's perception of leadership in times of change. In Dimmock, C. and Walker, A. (eds.) *Future School Administration: Western and Asian Perspectives*, pp 225–248. Hong Kong: Chinese University Hong Kong Press.
- Cheung, M. B. and Walker, A. (2006). Inner worlds and outer limits: The formation of beginning school principals in Hong Kong. *Journal of Educational Administration* 44(4), 389–407.
- Chew, J., Stott, K., and Boon, Z. (2000). The making of secondary school principals: Some perspectives from Singapore. *Paper Presented at the Australian Council for Educational Administration Conference Community Building in a Global Context*. Tasmania, Australia, September.
- Day, C. (2005). Sustaining success in challenging contexts: Leadership in English schools. *Journal of Educational Administration* 43(6), 573–583.
- Dimmock, C. and Walker, A. (2005). *Educational Leadership, Culture and Diversity*. London: Sage.
- Dimmock, C. and Walker, A. (1998). Comparative educational administration: Developing a cross-cultural conceptual framework. *Educational Administration Quarterly* 34(4), 558–595.
- Dimmock, C. and Walker, A. (1998). Transforming Hong Kong's schools: Trends and emerging issues. *Journal of Educational Administration* 36(5), 476–509.
- Dorfman, P. W., Hanges, P. J., and Broadbeck, F. C. (2004). Leadership and cultural variation: The identification of culturally endorsed leadership profiles. In House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., and Gupta, V. (eds.) *Culture, Leadership and Organizations: The GLOBE Study of 62 Societies*, pp 669–720. Thousand Oaks, CA: Sage.
- Emihovich, C. (2006). The cultural complexities of learning and development in a post-modern world. *Reviews in Anthropology* 35, 201–214.
- Groveswell, C. G. (2005). Introduction to the GLOBE Research Project on leadership worldwide, GROVEWELL, 2005. <http://www.groveswell.com/pub-GLOBE-leadership.html> (accessed April 2009).
- Hallinger, P. (1995). Culture and leadership: Developing an international perspective of educational administration. *UCEA Review* 36(2), 1–13.
- Hallinger, P. (ed.) (2003). *Reshaping the Landscape of School Leadership Development: A Global Perspective*. Lisse: Swets and Zeitlinger.
- Hallinger, P. (2004). Making education reform happen: Is there an 'Asian' Way? *Paper Presented at International Conference 'Making Change Happen: Learning from the Asian Experience'*. Bangkok, Thailand, 22–24 September.
- Hallinger, P., Chantarapanya, P., Sriboonma, U., and Kantamara, P. (2000). The challenge of educational reform in Thailand: Jing jai, jing jung, and nae norm. In Cheng, Y. C. and Townsend, T. (eds.) *Educational Change and Development in the Asia-Pacific Region: Challenges for the Future*, pp 208–226. Lisse: Swets and Zeitlinger.
- Hallinger, P. and Heck, R. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980–1995. *Educational Administration Quarterly* 32(1), 5–44.
- Hallinger, P. and Heck, R. (1999). Can leadership enhance school effectiveness? In Bush, T., Bell, L., Bolam, R., Glatler, R., and Ribbins, P. (eds.) *Educational Management: Redefining Theory, Policy, and Practice*, pp 178–190. London: Paul Chapman.
- Hallinger, P. and Kantamara, P. (2002). Educational change in Thailand. In Walker, A. and Dimmock, C. (eds.) *School Leadership and Administration: Adopting a Cultural Perspective*, pp 123–140. London: RoutledgeFalmer.
- Hallinger, P. and Leithwood, K. (1996). Culture and educational administration: A case of finding out what you don't know you don't know. *Journal of Educational Administration* 34(5), 98–116.
- Hallinger, P., Walker, A., and Bajunid, I. A. (2005). Educational leadership in East Asia: Implications for education in global society. *UCEA Review* 45(1), 1–5.
- Hanson, E. M. (2000). *Guest Editor of a Special Issue: Educational Decentralization Around the Pacific Rim*. *Journal of Educational Administration* 38(5).
- Heck, R. (1996). Leadership and culture. Conceptual and methodological issues in comparing models across cultural settings. *Journal of Educational Administration* 35(5), 74–97.
- Heck, R. (2002). Issues in the investigation of school leadership across cultures. In Walker, A. and Dimmock, C. (eds.) *School Leadership and Administration, Adopting a Cultural Perspective*, pp 77–102. London: RoutledgeFalmer.
- Ho, E. S. C. (2006). Educational decentralization in three Asian societies: Japan, Korea and Hong Kong. *Journal of Educational Administration* 44(6), 590–603.
- Hofstede, G. H. (1994). *Uncommon Sense about Organizations: Cases, Studies, and Field Observations*. Thousand Oaks, CA: Sage.
- Hofstede, G. H. (2005). *Cultures and Organizations: Software of the Mind*. New York: McGraw-Hill.
- Hoppe, M. H. (2004). Cross-cultural issues in development of leaders. In McCauley, C. D. and Velsor, E. V. (eds.) *Handbook of Leadership Development*, pp 331–360. San Francisco, CA: Jossey-Bass.
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., and Gupta, V. (2004). *Culture, Leadership and Organizations: The GLOBE Study of 62 Societies*. Thousand Oaks, CA: Sage.
- Huber, S. G. (2004). *Preparing School Leaders for the 21st Century: An International Comparison of Development Programs in 15 Countries*. London: RoutledgeFalmer.
- Hui, K. F. and Cheung, H. Y. (2006). A re-examination of leadership style for Hong Kong school-based management (SBM) schools. *Asia Pacific Journal of Education* 26(2), 173–187.
- Jabal, E. (2006). Leading from Hong Kong alumni: Lessons for school leadership. *International Journal of Leadership in Education* 9(1), 21–44.
- Jacobson, S. L., Day, W. D., and Leithwood, K. (2005). *Guest Editor of a Special Issue: The International Successful School Principalship Project*. *Journal of Educational Administration* 43(6).
- Jansen, J. D. (2006). Leading against the grain: The politics and emotions of leading for social justice in South Africa. *Leadership and Policy in Schools* 5(1), 37–51.
- Jones, A. (2006). Developing what? An anthropological look at the leadership development process across cultures. *Leadership* 2(4), 481–498.
- Law, Y. S. L. and Walker, A. (2005). Different values, different ways – principal problem solving and education reform. *International Studies in Educational Administration* 33(1), 62–78.
- Leithwood, K. (2005). Understanding successful principal leadership: Progress on a broken front. *Journal of Educational Administration* 43(6), 619–629.
- Leithwood, K., Day, C., Sammons, P., Harris, A., and Hopkins, D. (2006). *Successful School Leadership: What It Is and How It Influences Pupil Learning*. Nottingham: National College for School Leadership.
- Leung, J. Y. H. and Chan, H. M. (2001). The school management reform in Hong Kong: Administrative control in a new cloak of managerialism. In Cheung, B. L. and Lee, J. C. Y. (eds.) *Public Sector Reform in Hong Kong into the 21st century*, pp 229–260. Hong Kong: The Chinese University Press.
- Lin, M. D. (2001). Professional development for principals in Taiwan: The status quo and future needs trends. *Paper Presented at the International Conference on School Leaders Preparation, Licensure, Certification, Selection, Evaluation and Professional Development*. Taipei, Taiwan, March.
- Lin, M. D. (2003). Professional development for principals in Taiwan: The status quo and future needs trends. In Hallinger, P. (ed.) *Reshaping the Landscape of School Leadership Development: A Global Perspective*, pp 191–204. Lisse: Swets and Zeitlinger.
- Llomas, J. C. and Serrat, M. F. (2002). Is there a way out? A critical analysis of participation, leadership and management in Spanish. *International Journal of Leadership in Education* 5(4), 303–332.
- Lumby, J., Bryant, M., Walker, A., Bush T., and Bjork, L. (forthcoming). International perspectives on leadership education. Young, M. D., Crow, G., Murphy, J., et al. eds.) *The University Council for Educational Administration: Handbook of Research on the Education of School Leaders; Research on Leadership Preparation in a Global Context*. Mahwah, NJ: Erlbaum.
- Moller, J. (2006). Democratic schooling in Norway: Implications for leadership in practice. *Leadership and Policy in Schools* 5(1), 53–69.
- Moller, J. (2007). Educational leadership and a new language of learning. *International Journal of Leadership in Education* 10(1), 31–48.

- Oplatka, I. (2002). Women principals and the concept of burnout: An alternative voice? *International Journal of Leadership in Education* 5(3), 211–226.
- Oplatka, I. (2004). The principalship in developing countries: Context, characteristics and reality. *Comparative Education* 40(3), 427–448.
- Oplatka, I. (2006). Women in educational administration within developing countries: Towards a new international research agenda. *Journal of Educational Administration* 44(6), 604–624.
- Oplatka, I. and Hertz-Laraowitz, R. (2006). *Women Principals in a Multicultural Society: New Insights into Feminist Educational Leadership*. Rotterdam: Sense Publishers.
- Palomares, L. B. and Castillo, J. L. A. (2004). Professionalizing the principalship in Spain: Barriers in teachers' minds? *International Journal of Leadership in Education* 7(2), 147–163.
- Pashiardis, P., Costa, J. A., Mendez, A. N., and Ventura, A. (2005). The perceptions of the principal versus the perceptions of the teachers: A case study from Portugal. *International Journal of Educational Management* 19(7), 587–604.
- Pashiardis, P. and Ribbins, P. (2003). On Cyprus: The making of secondary school principals. *International Studies in Educational Administration* 31(2), 127–146.
- Persson, A., Andersson, G., and Lindstrom, M. N. (2005). Successful Swedish headmasters in tension fields and alliances. *International Journal of Leadership in Education* 8(1), 53–72.
- Pye, L. W. (2000). "Asian values": From dynamos to dominoes. In Harrison, L. E. and Huntington, S. P. (eds.) *Culture Matters: How Values Shape Human Progress*, pp 244–255. New York: Basic Books.
- Reynolds, D. (2000). School effectiveness: The international dimensions. In Teedlie, C. and Reynolds, D. (eds.) *The International Handbook of School Effectiveness Research*, pp 232–256. London: Falmer.
- Ribbins, P. and Zhang, J. H. (2004). Headteachers in rural China: Aspects of ambition. *International Journal of Leadership in Education* 7(2), 127–146.
- Rizvi, F. (1997). Beyond the East–West divide: Education and the dynamics of Australia–Asia relations. In Blackmore, J. and Toh, K. A. (eds.) *Educational Research: Building New Partnerships – A Collection of Selected Papers*, pp 13–26. Singapore: Educational Research Association.
- Saitis, C. and Menon, M. E. (2004). Views and future and current teachers on the effectiveness of primary school leadership: Evidence from Greece. *Leadership and Policy in School* 3(2), 135–157.
- Sapre, P. M. (2000). Realizing the potential of management and leadership: Toward a synthesis of Western and indigenous perspectives in the modernization of non-western societies. *International Journal of Leadership in Education* 3(3), 293–305.
- Shaw, K. E. (2005). Researching the trade in knowledge between the West and developing countries. *International Journal of Educational Management* 19(6), 459–468.
- Shen, W. J. (2001). Centralisation or decentralisation: International experiences and the implications for Taiwan. *Paper presented at the International Forum on Education Reforms in the Asia-Pacific Region*. Hong Kong, February.
- Slater, C. L., McGhee, M. W., Capt, R. L., et al. (2003). A comparison of the views of educational administration students in USA and Mexico. *International Journal of Leadership in Education* 6(1), 35–55.
- Southworth, G. (2005). Learning-centred leadership. In Davies, B. (ed.) *The Essentials of School Leadership*, pp 75–92. London: Paul Chapman.
- Stambeck, A. (2003). World-cultural and anthropological interpretations of 'choice programming' in education. In Anderson-Levitt, K. M. (ed.) *Local Meanings of Global Schooling: Anthropology and World Culture Theory*, pp 141–160. New York: Palgrave/St. Martins.
- Teedlie, C. and Reynolds, D. (2000). Context issues within school effectiveness research. In Teedlie, C. and Reynolds, D. (eds.) *The International Handbook of School Effectiveness Research*, pp 160–186. London: Falmer.
- Townsend, T. and Cheng, Y. C. (2000). *Educational Change and Development in the Asia-Pacific Region: Challenges for the Future*. Lisse: Swets and Zeitlinger.
- Trompenaars, A. and Hampden-Turner, C. (2000). *Riding the Waves Culture (Electronic Resource): Understanding Cultural Diversity in Business*. London: Nicholas Brealey.
- Wales, C. and Welle-Strand, A. (2005). *School Management Training Country Report: Norway, Studies in Education Management Research, No 16*. Norway: CEM Centre for Education Management Research, Norwegian School of Management and University of Oslo Institute of Educational Research.
- Walker, A. (2003). Developing cross-cultural perspectives on education and community. In Begley, P. and Johansson, O. (eds.) *The Ethical Dimensions of School Leadership*, pp 145–160. Dordrecht: Kluwer.
- Walker, A. (2004). Constitution and culture: Exploring the deep leadership structures of Hong Kong. *Discourse: Studies in the Cultural Politics of Education* 25(1), 75–94.
- Walker, A. (2006a). Leader development across cultures. *Journal of Research on Leadership Education* 1(1), 1–4.
- Walker, A. (2006b). Guest editor of a special issue of the *Journal of Educational Administration* entitled *International Views of Beginning Principalship* 44(4).
- Walker, A. and Chen, S.-Y. (in press). Leader authenticity in intercultural school contexts. *Educational Management, Administration and Leadership*.
- Walker, A. and Dimmock, C. (1999). A cross-cultural approach to the study of educational leadership: An emerging framework. *Journal of School Leadership* 9(4), 321–348.
- Walker, A. and Dimmock, C. (eds.) (2002). *School Leadership and Administration: Adopting a Cultural Perspective*. London: RoutledgeFalmer.
- Walker, A. and Dimmock, C. (2005). Leading the multi-ethnic school: Research evidence on successful practice. *Educational Forum* 69(3), 291–304.
- Walker, A. and Hallinger, P. (in press). Navigating culture and context: The principalship in East and Southeast Asia. In McLean, R. (ed.) *Learning and Teaching for the Twenty-First Century: Papers in Honour of Professor Phillip Hughes*. Dordrecht: Springer.
- Walker, A., Hallinger, P., and Qian, H. (in press). Leadership development for school effectiveness and improvement in East Asia. In Townsend T. (ed.) *The international handbook on school effectiveness and improvement*. Dordrecht: Springer.
- Wang, T. (2007). Understanding Chinese educational leaders' conceptions in an educational context. *International Journal of Leadership in Education* 10(1), 71–88.
- Webber, C. and Robertson, J. (2004). Internationalization and educators' understanding of issues in educational leadership. *Educational Forum* 68, 264–275.
- Wong, K. C. (2001). Chinese culture and leadership. *International Journal of Leadership in Education* 4(4), 309–319.
- Wong, K. C. (2006). Contextual impact on educational management and leadership: A case of Chinese education. *Journal of Educational Change* 7, 77–89.
- Wong, P. M. (2004). The professional development of school principals: Insights from evaluating a programme in Hong Kong. *School Leadership and Management* 24(2), 139–162.
- Yang, C. L. (2001). The changing principalship and its implication for preparing selecting and evaluating principals. *Paper Presented at the International Conference on School Leaders Preparation, Licensure, Certification, Selection, Evaluation and Professional Development*. Taipei, Taiwan.

Relevant Website

<http://www3.fed.cuhk.edu.hk> – Educational Leadership Development Net (ELDevNet) by Allan Walker.

Leading Diverse Schools

J Ryan, The Ontario Institute for Studies in Education, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Many scholars and educators believe that leadership has an important part to play in our increasingly diverse schools. Contemporary schools are more visibly diverse than they used to be, the consequence, in part, of changing immigration patterns and burgeoning local populations. Many Western countries, for example, are receiving more immigrants from non-Western countries than they did 20 and 30 years ago, just as nonwhite populations in Western countries have steadily increased. However, diversity has more dimensions than just race/ethnicity. School populations also display other kinds of differences that have always been with us. These include those that revolve around gender, sexual orientation, age, social class, religion, and ability, among others. What is significant about this diversity, however, is not so much the differences themselves, but the particular ways in which they are configured. Indeed, the meanings that are often attributed to these differences have the power to influence how schools are run and how students learn. More specifically, they can routinely enable or disable students, educators, and parents. The task for educators then becomes one of recognizing, understanding, and critiquing the deleterious meanings and associated practices that disenfranchise particular groups and, ultimately, acting to change the structures associated with them so that everyone can benefit from education.

Leadership can play a key role in this process. The right leadership practices can help marginalized groups get the most out of their schools. However, not all approaches to leadership are able to address disabling practices. Individualistic and technical approaches, for example, are the least able in this regard. On the other hand, leadership perspectives that emphasize the collective nature of leadership and promote social justice and inclusion are in a better position to counter such disabling practices. This article elaborates on these diversity-related leadership issues. It is organized in the following manner. First, it provides an account of the nature of diversity in contemporary schools and the consequences of the inequitable ways in which the difference is configured. This is followed by a critique of one common approach to leadership and a description of key dimensions of leadership. The last section describes leadership practices that are consistent with the pursuit of social justice and inclusion.

Diversity and Education

Diversity has increasingly become a matter of concern for many in the contemporary world. This is because humanity is now more aware of social difference than it was a mere two or three decades ago. There are many reasons for this increased sensitivity. One is the ease in which goods, communications, images, and people can move around the globe. A person who lives in London, England, can sample genuine Japanese cuisine, purchase a hat manufactured in Singapore, communicate instantaneously with a colleague in New York, and view images of urban life in Johannesburg, South Africa, in his/her living room. While these facilities have most certainly heightened awareness of diversity, it is perhaps the flow of humanity and, in particular, the changing immigration patterns that have sensitized us the most. For example, more people now immigrate to Western countries from developing countries than ever before (Census Canada, 2003; Australian Department Immigration and Multicultural Affairs, 2006; Government of the United Kingdom, 2005; US Census Bureau, 2001). However, this awareness of racial difference does not just emanate from the presence of foreign-born people; native-born nonwhite populations are also burgeoning in Western countries. In the United States, for example, people of African and Hispanic heritage constituted 16.6% of the total population in 2000 (US Census Bureau, 2001). One consequence of these patterns is that communities and schools do not seem to be what they did 20 years ago. This is particularly evident in larger cities where schools may have upward of 60 ethnicities represented in their student populations (Ryan, 2006a). Another consequence is that these changes have altered the way in which educators do, or should do, their work.

While immigration and changing local populations have sensitized us to human diversity, race/ethnicity constitutes only one dimension of difference. Many other kinds of diversity pervade our schools and communities. In addition, even though they have been around for some time now, they have not always been recognized or acknowledged in appropriate ways. These differences include those associated with gender, sexual orientation, age, social class, religion, and ability. To take one example, social class continues to persist as a significant difference within and between communities in a world where poverty shows few signs of disappearing. In Western countries, such

as the United States, Canada, and the United Kingdom, the gap between the rich and the poor has steadily increased since the 1970s. In 1995, the richest 1% of the population of the United States owned 47.3% of the financial assets, while the poorest 40% owned less than nothing –1.3% (Keister, 2000). In 2003, the number of people below the poverty line was 35.9 million (US Census Bureau, 2004). The figures in Canada and the United Kingdom tell the same story (Toronto Star, 2002; Byrne, 1999). This disparity, however, pales in comparison to the difference in wealth between Western countries, such as Canada, the United States, and the United Kingdom, and developing nations. For example, the European Union and the United States have gross national products (GNPs) of 13 502 800 and 12 455 825 (millions of USD), respectively, while African countries such as Liberia, Sierra Leone, and the Central Africa Republic record meager levels of 530, 1213, and 1377, respectively (International Monetary Fund, 2004).

These differences have consequences for student experience. More pointedly, the ways in which these differences are configured – interpreted, valued, and judged – can have a decisively positive or negative affect on how students learn in school. For example, differences associated with culture, ethnicity, social class, sexual orientation, gender, and ability can mean the difference between success and failure, enrichment and impoverishment, and hope and despair for students (Ryan and Rottman, *in press*). In most contemporary schools in the Western world, nonwhite, female, gay, lesbian, poor, and differently abled students frequently do not have as positive an educational experience as their white, male, straight, middle-class, and physically able counterparts. Many of the former tend to achieve at lower levels, drop out in greater numbers, and are less likely to attend post-secondary institutions than the latter (see e.g., Bennett, 2001; Natriello *et al.*, 1990; Paquette, 1990). Of course, not all students in the aforementioned groups follow this path. However, even those who perform well academically suffer from other significant consequences of this differential evaluation. They are systemically subjected to harassment, exclusion, and discrimination both in school and later, when they leave to venture into the working world (see, e.g., Datnow, 1998; Lugg, 2003; Orenstein, 2002; Ryan, 2006a; Stein, 2002). Difference makes a difference for students and their parents. A crucial task for leadership in the twenty-first century then is to recognize, understand, critique, and change the structures that generate the inequities that revolve around such differences.

Leadership Orientation and Diversity

Some leadership approaches are better equipped to meet the challenges of diversity and inequity than others.

More often than not, more traditional orientations are not able to acknowledge and confront the structures that marginalize groups of students and their parents. One approach that is poorly equipped for diversity is the so-called managerial orientation – both contemporary new managerialism (Gewirtz, 2002) and the more traditional forms. The weaknesses of managerial approaches reside in the ways proponents envision and practice (1) relationships among organizational members, (2) the roles that they occupy, and (3) the ends for which leadership is geared. Managerial proponents view administration as a technical practice mediated by an individual who occupies a position of authority and is responsible for putting any number of policies – whatever they may be – into practice (Ryan, 2006a). This view of leadership, however, is unable to promote the interests of the marginalized in the hierarchical relationships that it favors between managers and others, in the way in which it invests leadership power with individuals in particular positions, and in the way in which it promotes efficiency and productivity at the expense of more specific ends, such as social justice or inclusion (Blackmore, 1999; Marshall, 2004). In a managerial world, those who are not in management positions will be excluded from decision-making processes because they supposedly do not merit this inclusion. Vesting the power that comes from formal leadership in single individuals also excludes others who are not in these positions. Finally, the values of efficiency and productivity do little to promote the cause of inclusion or equity if leadership activities do not specifically target the latter.

Other approaches to leadership are better prepared for diverse contexts than managerial ones. This is due, in part, to the fact that their proponents are not content to be idle or neutral bystanders or merely describe what they study in a clinical or detached manner. They care deeply about what is happening to already-marginalized groups in diverse schools and are determined to do something about it. These sentiments figure prominently in their approaches to inquiry. Many of these scholars and practitioners believe that their work in leadership can provide the foundation for action that can rectify these unfair practices. Even though these approaches share many elements, they adopt a number of labels. They include emancipatory (Foster, 1989; Corson, 2000), democratic (Woods, 2005), social justice (Brown, 2004; Furman and Gruenewald, 2004; Marshall, 2004), feminist (Blackmore, 1999; Grogan, 2002), and inclusive (Ryan, 2006a) approaches to leadership.

Many scholars who champion these approaches reject hierarchical views of leadership. They correctly point out that these kinds of arrangements both reflect and reinforce wider social hierarchies and injustices (Corson, 1996). Some contend that these educational

organizational hierarchies themselves display class and gender overtones (Blackmore, 1989; Grace, 1995). Feminists have been the most articulate about this. They have argued that this hierarchical division of labor is masculine in nature and criticized the ideals of power and control that are part of this corporate management view (Blackmore, 1999; Grundy, 1993; Ozga, 1993). The point that they and others make is that hierarchies marginalize those not privileged enough to occupy formal positions of authority and those who do not possess the personal characteristics needed to influence others. For leadership to successfully counter practices that disadvantage some students, it must foster equitable and horizontal relationships that also transcend wider gender, race, and class divisions.

Leadership approaches that serve diverse populations best operate collectively rather than individually. Proponents argue that concentrating leadership power in a single individual – who may more often than not be an administrator – leaves out those who are not considered leaders from leadership-related activities (Foster, 1989; Ryan, 2003; 2006a). Go-it-alone individuals also leave problems in their wake when they move on to other positions (Hargreaves and Fink, 2006). Others make the point that the influence of single individuals on institutions is generally limited, and, as a consequence, it does not make sense to rely so heavily on single leaders. Debunking the heroic view of leadership, these and other scholars point out that individual men and women who occupy formal positions are seldom capable, on their own, of creating fundamental and long-lasting changes. As a consequence, they call not for heroes, but for modest men and women to step forward (Foster, 1989). Schools will serve the needs of their diverse populations not necessarily as the result of individual people doing remarkable things in isolation, but from a variety of people working together in many different ways and roles, using the multitude of different resources that are available to them (Leithwood *et al.*, 1999; Smylie *et al.*, 2002). Leadership that works best in diverse settings then, resides in the hands of the many rather than the few (administrators).

Leadership approaches best suited for diverse settings also emphasize the importance of clearly specifying the ends for which leadership should be directed, which in this case are inclusion, equity, and social justice. These contrast with other approaches to leadership that revolve around vague platitudes about efficiency and productivity. A number of the approaches cited above are organized to pursue the former goals. The proponents of emancipatory and feminist approaches, for example, are first and foremost concerned with social justice (Foster, 1989; Marshall, 2004). They believe that leadership activities need to be organized to pursue these ends on both a local and a global scale, and they view leadership processes as only one element of this enterprise. These scholars see leadership as the means for getting people to recognize these

injustices and for working together to change them. For leadership to be able to work for marginalized students, then, it needs to be seen and practiced as an inclusive, equitable, and collective process that is also organized to promote inclusion, social justice, and equity.

The Practice of Leadership in Diverse Contexts

School communities can do many things to advance collective leadership processes that promote equity and social justice. Among other things, they can encourage alternate ways of thinking about leadership; explore ways to include members of their school communities in school activities; advocate for inclusion and equity; educate students, parents, teachers, and administrators; promote critical consciousness; encourage dialog; emphasize the importance of student learning; promote inclusive decision- and policy-making processes; and ensure that equity initiatives become part of whole-school practices.

There are many ways to think about and practice leadership. Unfortunately, as mentioned above, many administrators follow more or less traditional patterns of leadership, that is, they tend to look at leadership as an individual and hierarchical practice. It is easy to understand why they do this. Among other things, they work in hierarchical bureaucratic organizations. By law, they are responsible for just about everything that occurs in their organizations – from ensuring that the drink machines are working to seeing that the province- or state-wide testing is carried out to guaranteeing that everyone is safe within the school walls. Therefore, many administrators may be unwilling to trust others to perform tasks for which they – and not these others – are responsible. However, if they are interested in ensuring that all students have meaningful educational experiences, they and everyone else in their school communities need to consider other ways of approaching leadership. Understanding and confronting practices that unfairly put students at a disadvantage require that they consider collective forms of leadership that move away from disabling hierarchies. Only when leadership is organized to give everyone a voice can the business of education proceed in a fair way.

School communities ought to also consider the many ways in which members can be included in school activities. If schools are to be equitable places, then students, teachers, and parents alike need to be included in their various influences and learning processes; there are many ways in which this can be done. Some roles can be formal, such as being part of committees or school councils. Others can be more informal, such as sharing expertise, volunteering, and accepting responsibility. Members of school communities can also be included in many areas of responsibility, including more administrative-like areas

such as hiring, teacher evaluation, the budgeting, promotion, and retention, as well as teaching-related areas such as instruction, curriculum, student behavior, and student placement (Ryan, 2006a, 2006b). However, not all areas or forms of inclusion may be appropriate for everyone. The respective dispositions, skills, and positions of parents, students, and parents will dictate that they may be included in different ways and at different times. Moreover, tasks that demand immediate action do not always lend themselves to collective decision making.

One of the reasons that it is not always easy to include students, teachers, and parents in leadership processes or to promote equity is that not everyone may favor them. Some may actively oppose these sorts of initiatives. For example, administrators may be reluctant to admit to the presence of racism in their schools (Ryan, 2003), just as some teachers condone sexism (Datnow, 1998). In order to counteract this opposition, those involved in leadership activities need to advocate for social justice and marginalized groups. In this regard, they can establish alliances with like-minded others (Bishop, 2002), make inclusion non-negotiable (Keys *et al.*, 1999) and create cognitive dissonance, discomfort, and a sense of urgency among members of the school community (Thousand and Villa, 1994). They may also share with others the theoretical, ethical, and research-based rationales for inclusive education and leadership (Thousand and Villa, 1994), trade and bargain with resisters (Gale and Densmore, 2003), stall and maneuver to counter exclusive policies (Gale and Densmore, 2003), establish links between educators and disadvantaged groups (Oakes and Lipton, 2002), and involve school and community stakeholder groups in formulating objectives for supporting all students (Thousand and Villa, 1994).

In order for school communities to embrace inclusion and equity, they need to be educated about these issues and the various groups in their communities. This is important because administrators, teachers, students, and parents generally know too little about each other and about debilitating practices such as racism (Ryan, 2003). In order to acquire the necessary knowledge, understandings, and attitudes, all members of the school community will have to assume the role of both teacher and learner. Educators can help parents and students learn about the school system and community, while parents can assist educators in learning about themselves and their communities. This can be accomplished in both informal and formal ways – in the day-to-day interactions among members of the school community and in more formally organized developmental activities. Parents can learn about the school through newsletters, brochures, handbooks, formally and informally organized meetings, and face-to-face interaction (Ryan, 2006a, 2006b). With regard to sensitive matters such as racism, schools need to find a balance that maintains pressure to be reflective about current assumptions and practices, without being

confrontational in a way that merely reproduces and amplifies current conflicts. The best route to take is one that is positive without being comfortable and that prompts people to not only reflect on the present state of affairs, but also does not produce the fear and guilt that trigger further conflict. This strategy recognizes that people make mistakes, and that these mistakes should be acknowledged and discussed in a constructive manner (Gillborn, 1995).

It is not enough to simply educate members of the school community about equity issues. This education also needs to be critical, that is, leadership needs to promote a critical consciousness in students, teachers, administrators, and parents (Anderson, 1990; Blasé and Anderson, 1995; Foster, 1989; Grundy, 1993). This involves helping people to look at the school and community context through a critical lens so that they can see many of the things, such as racism, which they normally take for granted, in a new light. In order for individuals to break out of their usual patterns of thinking that obscure exclusion, they need to engage others – particularly different others – in critical conversations. These critical conversations can help school communities acknowledge, recognize, critique, and change those invisible practices that impede inclusion (Ryan, 2006a, 2006b). One of the strategies that school communities can do to promote these critical conversations is to provide an atmosphere that supports the posing of critical questions. Questions such as What is happening here? What do we know about this? Who says this is the way things should be? What overall purposes are being served? Whose vision is it? Whose interests are being served? Whose needs are being met? Whose voices are being silenced, excluded, or denied? How come some viewpoints always get heard? Why is this particular initiative occurring now? What kind of prudent and feasible action can we adopt? Who can we enlist to support us? How can we start now? How are we going to know when we make a difference? (Smyth, 1996) can help school communities look at their practices critically and take action to do something about those that disadvantage students.

In order for school communities to engage in critical conversations, they will need to find ways to nurture dialog. These conversations will work only if people are able to communicate with one another on an equal footing, that is, when everyone has a voice. This means that those who have power need to develop strategies to help them communicate with those with less power. Listening is important in this context. It is vital that listeners from dominant groups provide as much comfort as possible so that the latter can speak their minds (Levine-Rasky, 1993). To do this, it helps to displace oneself as a knower and evaluator, abandon a desire to overtly assign a relative worth to observations, reflect on one's privileges, suspend personal authority, be willing to experience vulnerability,

and admit one's ignorance. Dialog can also be nurtured when educators position themselves in ways that bring them into contact with others (Ryan, 2003). Dialog, however, is not a panacea (see, e.g., Ellsworth, 1989; Boler, 2004). There is no guarantee that its use will ensure that the marginalized will be meaningfully included in school activities and processes. Nevertheless, even the critics acknowledge that it can be useful in a number of respects, such as helping groups form alliances (Ellsworth, 1989) or simply continuing conversations (Burbules, 1993).

Inclusion and equity are best achieved in schools that emphasize the importance of student learning and classroom teaching. One important task for leadership then is to find ways to improve the capacities and commitment of professional educators. The latter is best achieved when educators are aware of practices that work in diverse settings and are given the chance to talk critically about their teaching (see e.g., Ryan, 2006a: 123). It is easiest for teachers to adopt these teaching strategies in schools where goals are clearly understood, collaboration is supported, risk taking is encouraged, and the monitoring of progress is fair. Teaching in diverse contexts is best developed when teachers and their supervisors have opportunities to talk about it and critically reflect upon it (Blasé and Blasé, 2000). These critical conversations occur most often in contexts where school goals are developed collaboratively and clearly communicated to everyone. This is particularly important in diverse settings because members of various groups may not understand what others in more homogeneous contexts simply take for granted. All this occurs most often in schools that spend time making sure that student learning is a priority (Kleine-Kracht, 1993).

In order for diverse schools to ensure that everyone has a voice, and that alternate perspectives and worldviews are fairly represented, they need to adopt inclusive policy-making processes. Policy deliberations can work toward inclusion in two ways. The first is by promoting policies that favor inclusive values. The second is by organizing policy-deliberation processes that are themselves inclusive. This requires that all interests in the school community be fairly represented in these processes and that everyone have equally fair opportunities to influence the outcome of these processes. To accomplish this, schools need to know their communities, seek fair representation from all segments of the community to participate, establish an appropriate place for deliberations, pay particular attention to the power dynamics, ensure that everyone is heard in policy deliberations, help policy makers see problems from others' viewpoints, consult with the people affected by the policy, and be prepared to engage in ongoing process of monitoring and adjusting the policies (Corson, 1996; Gillborn, 1995; Ryan, 2006a, 2006b).

Finally, if inclusion and equity are to be successfully promoted in schools, then leadership practices need to be organized to ensure that they become firmly entrenched

in day-to-day activities, that is, in the culture of schools (Gillborn, 1995; May, 1994; Ryan, 2003). In order to accomplish this task, schools need to involve whole-school communities. This entails making equity an essential and routine part of educational practice in ways that ensure its longevity and protect it against wider changes in educational policy. For schools to promote cultures that foster inclusion, they will have to acknowledge the complexities of culture and be sensitive to the ends to which their efforts are directed. An approach to leadership that is suited to diverse settings demands that the school community's efforts should promote everyone's interests – not just those of management or of dominant groups. School vision statements, for example, ought to emanate not only from management or powerful individuals and groups, but also, equitably, from all segments of the school community. The shared values that emerge from whole-school efforts must represent all groups and all groups must benefit equally from these values (Ryan, 2006a, 2006b).

Conclusion

Leadership is more important than ever in today's diverse schools. It can be a powerful tool in confronting and changing the practices that disable certain groups of students. In order to be able to do so, though, it needs to be practiced as a collective, inclusive, and equitable process that is also organized to promote equity and inclusion. However, the strategies associated with this approach should not be seen as a panacea or as prescriptions that apply equally to all contexts. The nature of diversity itself precludes such a conclusion; schools, communities, and the people who attend and live in them will display many differences. Therefore, we need to be cautious when putting the above strategies into practice. The best way to approach them is to view them as part of a general framework that is sensitive to the local context and flexible enough to be adapted to a wide range of circumstances. If educators are able to do this, they will be taking an important step toward making schools places of learning from which all students can benefit.

Bibliography

- Anderson, G. (1990). Toward a critical constructivist approach to school administration: Invisibility, legitimation, and the study of non-events. *Educational Administration Quarterly* 26(1), 38–39.
- Australian Government Department of Immigration and Multicultural Affairs (2006). *Overseas Arrivals and Departure Statistics*. <http://www.immi.gov.au/media/statistics/statistical-info/oad/index.htm> (accessed April 2009).
- Bennett, C. (2001). Genres of research in multicultural education. *Review of Educational Research* 71(2), 171–217.
- Bishop, A. (2002). *Becoming an Ally: Breaking the Cycle of Oppression in People*. Halifax, NS: Fernwood.

- Blackmore, J. (1989). Educational leadership: A feminist critique and reconstruction. In Smith, J. (ed.) *Critical Perspectives on Educational Leadership*, pp 93–130. London: Falmer.
- Blackmore, J. (1999). *Troubling Women*. Buckingham, UK: Open University Press.
- Blasé, J. and Anderson, G. (1995). *The Micropolitics of Educational Leadership: From Control to Empowerment*. New York: Teachers College Press.
- Blasé, J. and Blasé, J. (2000). Principals' perspectives on shared governance leadership. *Journal of School Leadership* 10(1), 9–39.
- Boler, M. (ed.) (2004). *Democratic Dialogue in Education: Troubling Speech, Troubling Silence*. New York: Peter Lang.
- Brown, K. (2004). Leadership for social justice and equity: Weaving a transformative framework and pedagogy. *Educational Administration Quarterly* 40(1), 77–108.
- Burbules, N. (1993). *Dialogue in Teaching: Theory and Practice*. New York: Teacher College Press.
- Byrne, D. (1999). *Social Exclusion*. Philadelphia, PA: Open University Press.
- Census Canada (2003). Canada's ethnocultural portrait: The changing mosaic. http://www12.statcan.ca/english/census01/products/analytic/companion/etoimm/canada.cfm#proportion_foreign_born_highest (accessed April 2009).
- Corson, D. (1996). Emancipatory discursive practices. In Leithwood, K., Chapman, J., Corson, D., Hallinger, P., and Hart, A. (eds.) *International Handbook of Educational Leadership and Administration*, pp 1043–1067. Dordrecht: Kluwer.
- Corson, D. (2000). Emancipatory leadership. *International Journal of Leadership in Education* 3(2), 93–120.
- Datnow, A. (1998). *The Gender Politics of Educational Change*. London: Falmer.
- Ellsworth, E. (1989). Why doesn't this feel empowering? Working through the repressive myths of critical pedagogy. *Harvard Educational Review* 59(3), 297–324.
- Foster, W. (1989). Toward a critical practice of leadership. In Smyth, J. (ed.) *Critical Perspectives on Educational Leadership*, pp 70–81. London: Falmer.
- Furman, G. and Gruenewald, D. (2004). Expanding the landscape of social justice: A critical ecological analysis. *Educational Administration Quarterly* 40(1), 47–76.
- Gale, T. and Densmore, K. (2003). Democratic educational leadership in contemporary times. *International Journal of Leadership in Education* 6(2), 119–136.
- Gewirtz, S. (2002). *The Managerial School: Post-Welfarism and Social Justice in Education*. New York: Routledge.
- Gillborn, D. (1995). *Racism and Antiracism in Real Schools*. Philadelphia, PA: Open University Press.
- Government of the United Kingdom. (2005). Control of Immigration: Statistics United Kingdom. <http://www.official-documents.gov.uk/document/cm69/6904/6904.asp> (accessed April 2009).
- Grace, G. (1995). *School Leadership: Beyond Educational Management*. London: Falmer.
- Grogan, M. (2002). Influences of the discourse of globalisation on mentoring for gender equity and social justice in educational leadership. *Leading and Managing* 8(2), 123–134.
- Grundy, S. (1993). Educational leadership as emancipatory praxis. In Blackmore, J. and Kenway, J. (eds.) *Gender Matters in Educational Administration and Policy*, pp 65–78. London: Falmer.
- Hargreaves, A. and Fink, D. (2006). *Sustainable Leadership*. San Francisco, CA: Jossey-Bass.
- International Monetary Fund (2004). World economic outlook. <http://www.imf.org/external/pubs/ft/weo/2004/02/data/dbaselm.cfm>
- Keister, L. (2000). *Wealth in America: Trends in Wealth Inequality*. Cambridge: Cambridge University Press.
- Keys, M., Hanley-Maxwell, C., and Capper, C. (1999). Spirituality? It is the core of my leadership: Empowering leadership in an inclusive elementary school. *Educational Administration Quarterly* 35(2), 203–237.
- Kleine-Kracht, P. (1993). Indirect instructional leadership: An administrator's choice. *Educational Administration Quarterly* 29(2), 187–211.
- Leithwood, K., Jantzi, D., and Steinbach, R. (1999). *Changing Leadership for Changing Times*. Buckingham: Open University Press.
- Levine-Rasky, C. (1993). Listening as an appropriate response by members of privileged groups. *Paper Presented at CSAM Meeting*. Ottawa: Carleton University.
- Lugg, C. (2003). Sissies, faggots, lezzies and dykes: Gender, sexual orientation, and a new politics of education. *Educational Administration Quarterly* 39(1), 95–134.
- Marshall, C. (2004). Yes, "but"... education leaders discuss social justice. *Journal of School Leadership* 14, 530–563.
- May, S. (1994). *Making Multicultural Education Work*. Clevedon, UK: Multilingual Matters.
- Natriello, G., McDill, E., and Pallas, A. (1990). *Schooling Disadvantaged Children: Racing against Catastrophe*. New York: Teachers College Press.
- Oakes, J. and Lipton, M. (2002). Struggling for educational equity in diverse communities: School reform as social movement. *Journal of Educational Change* 3(4), 383–406.
- Orenstein, P. (2002). Striking back: Sexual harassment at Weston. *Jossey-Bass Reader on Gender in Education*, pp 459–475. San Francisco, CA: Jossey-Bass.
- Ozga, J. (1993). *Women in Educational Management*. Milton Keynes, UK: Open University Press.
- Paquette, J. (1990). Minority participation in secondary education: A fine-gained descriptive methodology. *Education Evaluation and Policy Analysis* 13(2), 139–158.
- Ryan, J. (2003). *Leading Diverse Schools*. Dordrecht: Kluwer.
- Ryan, J. (2006a). *Inclusive Leadership*. San Francisco, CA: Jossey-Bass.
- Ryan, J. (2006b). Inclusive leadership and social justice for schools. *Leadership and Policy in Schools* 5(1), 3–17.
- Ryan, J. and Rottman, C. (2007). Educational leadership and policy approaches to critical social justice. *Journal of Educational Administration and Foundations* 18(1 and 2), 9–23.
- Smylie, M., Conley, S., and Marks, H. (2002). Building leadership into the roles of teachers. In Murphy, J. (ed.) *Educational Leadership Challenge: Redefining Leadership for the 21st Century*, pp 162–188. Chicago, IL: NSSE.
- Smyth, J. (1996). The socially just alternative to the "self-managing school". In Leithwood, K., Chapman, J., Corson, D., Hallinger, P., and Hart, A. (eds.) *International Handbook of Educational Leadership and Administration*, pp 1097–1131. Dordrecht, Netherlands: Kluwer.
- Smyth, J. (1989). A "pedagogical" and "educative" view of leadership. In Smyth, J. (ed.) *Critical Perspectives in Educational Leadership*, pp 179–204. London: Falmer.
- Stein, N. (2002). Bullying as sexual harassment in elementary schools. *The Jossey-Bass Reader on Gender Education*, pp 409–428. San Francisco, CA: Jossey-Bass.
- Thousand, J. and Villa, R. (1994). Managing complex change toward inclusive schooling. In Villa, R. and Thousand, J. (eds.) *Creating an Inclusive School*, pp 5–79. Alexandria, VA: Association for Supervision and Curriculum Development.
- Toronto Star (2002). Rich getting richer: Middle-class poorer, March 15.
- US Census Bureau (2001). Adding diversity from abroad. <http://www.census.gov/population/pop-profile/2000/> (accessed April 2009).
- US Census Bureau (2004). Income stable, poverty up, numbers of Americans with and without health insurance rise. *Census Bureau Reports*. http://www.census.gov/Press-Release/www/releases/archives/income_wealth/002484.html (accessed April 2009).
- Woods, P. (2005). *Democratic Leadership in Education*. Thousand Oaks, CA: Sage.

Further Reading

- Ingram, P., Mayrowetz, D., and Weinstein, C. (1997). Sources of leadership for inclusive education: Creating schools for all children. *Educational Administration Quarterly* 35(3), 423–449.
- Jencks, C. (1972). *Inequality: A Reassessment of Effect of Family and Schooling in America*. New York: Basic Books.
- Marshall, C. and Oliva, M. (2005). *Leadership and Social Justice: Making Revolutions in Education*. Boston, MA: Allyn and Bacon.

- Maxcy, S. (1998). Preparing school principals for ethno-democratic leadership. *International Journal of Leadership in Education* **1**(3), 217–235.
- OECD (2005). First results from PISA 2003. Organization for Economic Cooperation and Development. <http://www.oecd.org> (accessed April 2009).
- Pounder, D., Ogawa, R., and Adams, E. (1995). Leadership as an organization-wide phenomenon: Its impact on school performance. *Educational Administration Quarterly* **31**(4), 564–588.
- Scheurich, J. and Skrla, L. (2003). *Leadership for Equity and Excellence*. Thousand Oaks, CA: Corwin.
- Shields, C. (2003). *Good Intentions Are Not Enough: Transformative Leadership for Communities of Difference*. Lanham, MD: Scarecrow Press.
- US Census Bureau (2003). Poverty, income see slight change: Child poverty rate unchanged. Census Bureau Reports. <http://www.census.gov/Press-Release/www/2003/Cb03-153.html> (accessed April 2009).

Supervision and Personnel Administration

P D Tucker, University of Virginia, Charlottesville, VA, USA

D G Pounder, University of Utah, Salt Lake City, UT, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Centralization – The concentration of administrative functions at a central location.

Compensation – The payment of salary and/or benefits for services rendered.

Decentralization – The transfer of administrative functions to local authorities.

Equity – Fundamental fairness to all parties.

Human resources – A more current term for personnel administration, emphasizing that people are a critical resource to education.

Induction – A means of introducing and socializing an employee to a new position and organization.

Particularism – The practice of focusing on the characteristics of a person that are distinct such as religious, ethnic, or political affiliations.

Professional development – Learning experiences that are intended to increase the knowledge and skills of educators.

Professional learning communities – Groups of individuals who engage in reflection and dialog about their practice to improve their collective performance.

Recruitment – The strategies and process of attracting applicants to an organization.

Retention – The ability to keep or retain employees over time.

Scientific management – The movement of the early twentieth century that emphasized a technical approach to management that was formalized, standardized, and specialized.

Selection – The process of choosing the most qualified person for a position.

Supervision – The process of working with individuals to improve performance.

Universalism – The practice of focusing on the characteristics of a person that match objective criteria for a position.

Introduction

Supervision and personnel administration serve fundamental purposes in the educational experience of 1.5 billion children worldwide. It is with and through people that

organizations both maintain and develop their professional capacity and respond to governmental goals and societal changes. While there is a global commitment to education, it is not universally available, particularly in developing countries where the average secondary enrollment is only 19% (Foskett and Lumby, 2003). Hence, there are substantial differences in the underlying beliefs about and realities of the educational experience throughout the world, but it is the individuals in schools, particularly teachers, who engage and motivate students in the learning process. The manner in which these adults are treated by state or school officials influences the daily experiences of children and adolescents in schools and, ultimately, their economic futures.

The purpose of this article is to provide an overview of personnel administration from an international perspective. Given that much of the literature and research on personnel administration emanates from English-speaking countries, such as the United States, it is necessary to acknowledge that the reported research reflects personnel practices based on a narrow set of cultural values. Where possible, a cross-cultural perspective will be provided. The topics to be addressed are the purposes and functions of personnel administration, current challenges and promising strategies, and future directions.

Purposes, Historical Perspective, and Policy Context

The primary purposes of personnel administration are the recruitment, selection, induction, development, supervision, evaluation, and compensation of the most competent individuals to create learning environments for students. While the functions are largely constant across settings, personnel practices and the concepts which drive them greatly differ based on cultural beliefs, norms, and customs.

Purposes and Significance

Despite differences in personnel practices in different countries and locales, there are commonalities in the types of issues to be resolved, such as: (1) how to define the ideal educator, (2) how to recruit and identify the ideal educator, (3) how best to prepare a sufficient supply of capable educators, (4) how to support the ongoing learning and development of educators, (5) how to evaluate effective job performance, (6) how to structure compensation to

attract, retain, and reward competent educators, and (7) how to provide supportive working conditions (Thomas, 1990). Issues and values such as equity, quality, accountability, decision-making control regarding employment, efficiency, economic conditions, legal constraints, and labor supply conditions influence most employment policies and practices in education across the spectrum of cultures.

An example of cultural influence on actual practices is illustrated by the concept of an ideal candidate. In the United States, a universalistic approach is used and applicants are judged by professional qualifications (e.g., professional licensure or certification, appropriate degree attainment, and classroom teaching performance recommendations). This is viewed as nondiscriminatory and more objective by Americans. In other parts of the world, particularism is valued to a greater extent and an emphasis is placed on particular characteristics, such as ethnicity, religious affiliation, and kinship, which are viewed as more critical to serving community needs. Where a school is situated along this continuum, from universalism to particularism, reflects an intricate weave of cultural and community considerations (Foskett and Lumby, 2003).

Historical Perspective

Personnel administration is deeply rooted in a bureaucratic orientation to school organization and functioning. Growing out of the scientific management movement during the early 1900s, personnel administration became a core function of the central administration of larger school districts in the United States to respond to the growing complexity of schools due to urbanization, immigration, and industrialization. Personnel procedures were characterized by formalization and specialization which led to a standardized approach to filling positions through job analysis, position descriptions, selection criteria, application procedures, and reference checks. In the 1940s, the human relations movement shifted the discourse to one of valuing individuals as human resources and as a means of achieving organizational goals. Greater attention was given to working conditions as well as to psychological and social underpinnings of worker motivation. The application of systems theory in the 1970s suggested interplay between individual needs and institutional goals that could be aligned for optimal organizational effectiveness. More recently, the concept of learning organizations in the business world has laid the groundwork for professional learning communities within schools (Hall and Hord, 2001). From this perspective, people are viewed as integral to shaping the cultural fabric and contributing to the professional capacity of a school and school system.

In many ways, this historical perspective from the United States describes a developmental sequence through which many educational systems travel. While some emerging countries are still struggling to simply manage

school employees, in other places, schools have re-created themselves as high-functioning learning communities. For example, in many parts of Africa, the relationship between supervisor and teacher is described as a master-servant relationship, suggesting an industrial-era mentality in contrast to the push elsewhere for a human resource approach that is responsive to the needs of individuals.

The influence of various schools of thought on personnel administration has contributed to systematic approaches in identifying and meeting staffing needs, understanding motivational workplace practices, the necessary balance between organizational goals and individual needs, and the reconceptualization of schools as learning organizations. Collectively, these trends have shaped both the technical and humanistic aspects of personnel administration. In some parts of the world, such as South Africa, the basic challenge of filling positions with instructors who are familiar with the curriculum remains the primary purpose of personnel administration, while in other countries, the focus of personnel administration is to facilitate and support professional communities of practice.

Today, personnel administration is shaped by a research base that draws from a broad range of theoretical perspectives in education and business. They include social systems theory, organization theory, change theory, communications theory, leadership theory, decision theory, motivation theory, adult learning theory, group dynamics theory, and others. While researchers began to focus on personnel issues specific to educational settings in the 1970s and 1980s, an international perspective was absent until the early 1990s when greater attention was given to comparative international practices (Brewster and Tyson, 1991). Comparative analyses have helped highlight educational commonalities as well as reveal the underlying beliefs and meanings of practices due to cultural differences.

Policy Context for Personnel Administration

People constitute the single, largest, and most expensive resource in schools (70–90% of most educational budgets) and, therefore, are the focus of most current reform efforts around the world. In the United States, the No Child Left Behind Act of 2001, has mandated accountability for student achievement and has identified highly qualified teachers as a critical contributor to increased student performance. Similar top-down accountability efforts have been applied to education in other countries as the business forces of globalization exert pressure on educational systems to compete in the knowledge arena. How policy mandates and incentives affect individual schools and their inhabitants, however, is a complex interaction of national, regional, community, and family expectations as well as legislative, cultural, and economic factors. Often, collective bargaining agreements between employee

unions and employing educational agencies further influence personnel administration policies and practices.

Functions of Personnel Administration

At a minimum, the central administration for schools is responsible for the functions of strategic planning, compensation, and collective bargaining associated with personnel administration. Depending on the level of centralization within an educational system, the functions that are shared sometimes with school heads are employee selection, induction, professional development, supervision, and evaluation. This is certainly the case in many parts of the United States and China. This push for decentralization of personnel functions is evident not only in increased self-determination at the school level, but also at the teacher level with greater empowerment to be active participants in selection efforts and professional development.

One descendent of scientific management is the planned and systematic approach most education entities use to determine future goals for the organization. Strategic planning involves the identification of system goals, strategies for achieving goals, resource requirements, and contextual variables that will enhance or limit goal attainment. Planning encourages the systematic determination of personnel requirements, allows for a proactive focus on programs and students, and targets resources, financial and human, to prioritized goals. By better aligning personnel practices with high-priority organizational goals, school systems are better able to improve performance by evaluating, developing, and compensating individuals in ways that contribute to clearly identified objectives.

Recruitment and Selection

Well-planned recruitment efforts can enhance the defensibility and effectiveness of employee hiring practices. A rich and diverse applicant pool increases the probability of finding candidates who have both the personal qualities and professional expertise that are sought by the employer. Affirmative efforts, however, are required to attract a broader range of candidates and many school systems are actively seeking protected class members to diversify the employees within the organization. This effort is especially critical to teacher hiring in districts enrolling large numbers of students of color.

Significantly, more research has focused on employee selection, particularly the selection interview, than on recruitment. Research findings from industrial/organizational psychology and business management have provided much of the foundational knowledge on personnel selection. It documents the historically weak reliability and predictive validity of the selection interview due to non-job-relevant factors (Delli and Vera, 2003).

This body of research has resulted in employee selection methods designed to increase the reliability, and to a lesser degree the predictive validity, of selection decisions. These methods include clarification of the hiring criteria before conducting the interview, use of a structured interview format, formulation of questions to elicit experience-based responses, use of a scoring rubric for responses, multiple trained interview raters, and review of other data sources (Campion *et al.*, 1988). Greater reliance on more objective evaluation of the person's qualifications and experiential knowledge increases inter-rater reliability and reduces the biases that have been documented in less-structured dyadic interviews.

Selection methods requiring evidence of actual job performance may strengthen the predictive power of selection decisions (Pounder, 1988). Consistent with this thinking, some school systems ask applicants to submit evidence of job performance through videotaped teaching, substitute teaching, or actual teaching observations. Research has not documented the predictive validity of these strategies but they have strong face validity.

It is critical to note that the American concern for objectivity is both a legal and a cultural concern that is not shared with other cultures. China, for example, is much more comfortable with subjectivity in personnel matters and would be more concerned with the ability of the applicant to serve the school and the wider community in an agreeable manner. In addition, commitment to existing political entities is expected. These values reflect China's collectivist nature which contrasts with the emphasis on the individual in most Western European cultures (Walker and Dimmock, 2002).

Induction

Induction programs provide new employees with their first opportunity to acquire context-specific information about the system and school in which they will work and fill the gap between preservice preparation program and ongoing in-service professional development. School districts have been found to play meaningful roles as teacher educators. They determine the assignments given to new teachers, the resources they are provided, and the opportunities for learning that affect the quality of novice educators' teaching (Grossman and Thompson, 2004). For these reasons, the growth of such programs in the United States has been dramatic with 83% of teachers from public schools reporting participation in 2000 (Smith and Ingersoll, 2004). The programs vary in terms of content, characteristics, duration, and intensity but common elements exist such as mentors, beginners' seminars, extra resources, and common planning time.

If effective, induction activities can assist individuals who are new to a school to appreciate the values, professional

expectations, and social networks that define it. The effects of such programs are typically measured in terms of teacher attitudes and teacher retention. Research has confirmed their success in increasing job satisfaction, efficacy, and retention. Retention is a major issue in the United States with 29% of all first-time teachers in 1999–2000 either changing schools or leaving teaching. While only a small percent of beginning teachers in the United States experience a comprehensive induction program, those who do are half as likely to leave teaching or change schools.

Supervision and Evaluation

Supervision and evaluation are often used in tandem in the literature implying that they are reciprocal processes, but effective evaluation subsumes supervision as one of its primary purposes. Supervision is conceptualized as a formative process with the primary purpose being the improvement of instruction and includes classroom observation, group development, and professional development. Evaluation is conceptualized as both formative and summative with the dual purposes of performance improvement and accountability. Summative purposes of evaluation include making decisions regarding retention, teacher tenure, promotion, and sometimes differentiated pay. Most Western school systems develop evaluation systems that emphasize both the improvement and accountability aspects of evaluation; however, in some countries, such as China and Hong Kong, evaluation is used primarily to reward teachers or make promotion decisions. A greater emphasis on the improvement aspects of evaluation are expected in the future.

To enhance both effectiveness and defensibility of evaluation policies and practices, criteria and methods for performance evaluation are typically tied to specific job requirements. In the United States, these job expectations tend to focus on classroom activities but this is not the case in all cultures. As a point of contrast, morality, diligence, abilities, and student performance are major areas of evaluation in the People's Republic of China (Teddlie *et al.*, 2003). Research on teacher evaluation suggests that more highly structured evaluation systems increase the consistency of summative judgments on performance and the level of discussion about instruction. Components of an evaluation system that increase its effectiveness and legal defensibility include a written policy on evaluation that is compatible with existing legal requirements and collective bargaining contracts, prior notification of job-related expectations, trained evaluators who are familiar with both the content and procedural aspects of the evaluation policy, equitable treatment of all employees across schools and locations, collection of a broad sampling of performance across time, documentation of performance that is shared with the employee, and an opportunity to improve in areas of concern (Tucker and DeSander, 2006). These minimal

guidelines reflect an approach to evaluation that is fundamentally fair and one that is ethical and supportive of professional growth. More extensive guidelines for the development of evaluation systems for educators can be found in *The Personnel Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 1988).

Although many countries have elaborate evaluation systems (e.g., India, South Africa, China, New Zealand, and the United States), personnel evaluation practice does not meet these standards consistently and thus may not provide meaningful feedback for performance improvement. In the Netherlands, for example, the high level of school autonomy results in local definitions of teacher evaluation and weakly conceptualized approaches to it. In the People's Republic of China, teacher evaluation is highly structured but focuses on political variables along with teaching duties and does not promote professional development (Liu and Teddlie, 2003). Without meaningful implementation by skilled supervisors, professional growth remains an unrealized goal of supervision and evaluation in many schools.

Professional Development

The provision of in-service professional development acknowledges the need for educators to continue learning once they have been hired. Depending on the quality of preservice preparation, professional development can be used to address training gaps, expand knowledge and skills, explore new strategies to promote student learning, and respond to ubiquitous change. The possible delivery models are numerous but they tend to cluster around four dominant strategies: (1) knowledge transmission, (2) demonstration or modeling of the skill or concept, (3) practice of the skill or application of the knowledge, and (4) peer coaching or collaborative work with other teachers to implement the new skills or knowledge as an integral part of teaching (Joyce and Showers, 2002). Professional development is often delivered through workshops based on the needs determined by the central administration and may have minimal impact on performance. Evidence suggests that teachers learn best from each other through collaborative work around self-identified problems of practice.

The actual content of the professional development can be open-ended, organized around school-wide goals, or based on individual needs as identified by self-assessment or supervisor feedback. Ideally, it should be a natural outgrowth of supervision and evaluation that is intended to improve individual performance. Without systemic support and opportunities to develop new skills, evaluation does not contribute to individual or organizational growth. In the current context of results-driven education, effective professional development should be informed by research on teacher and learning, guided by coherent long-term planning, based at the building level, and structured to be

continuous and ongoing. Most important, however, is the enlistment of teachers as agents, and not as objects, in their own professional development.

Compensation

Although considerable research has been done on the development of effective and defensible compensation policies and practices in the private sector, many of these findings have had only marginal applicability to compensation practices in public schools. Nonetheless, there are some broad principles of equity (and effectiveness) that deserve attention in public school employee compensation systems, including issues regarding internal equity, external equity, and personal equity. First, internal equity, salary comparisons between and among different jobs in the same organization, has applicability to decisions about compensation levels between and among administrators, teachers, professional support staff, and various types of service personnel. Job-evaluation methods identifying appropriate compensable factors (based on job requirements) and their compensation value for each major job cluster in an organization enable school districts to analyze the internal equity of their compensation structures.

Second, external equity, salary comparisons between similar jobs in different organizations, has implications for an organization's ability to attract and retain employees due to competitive salary levels. Salary data comparing teaching salaries to those of other degree-holding occupational groups have been used to argue for increased salary levels in education. Due to the widespread process of collective negotiations in most American public school districts, teacher unions often utilize comparative salary data from other relevant school districts to bargain for salary increases. This influence often results in teacher compensation schedules that are externally competitive with neighboring school districts. Where this influence does not exist, as with nonteaching jobs or with educational organizations that do not bargain collectively, educational organizations may need to conduct market wage surveys to obtain comparative salary data on key jobs within various job clusters. These data may be used to adjust salary levels within the organization to maintain a competitive edge in attracting and retaining employees.

Finally, performance-based pay in teacher compensation has received a considerable amount of attention during the last two decades. This practice is still quite controversial, but it does speak to a final type of equity: personal equity, which addresses an employee's perception of his or her performance and worth relative to that of referent others. Although the single salary scale dominates most school systems, there is mounting pressure based on models from business to consider differentiated pay based on subject-area knowledge, specific skills, student

outcomes, and willingness to teach in difficult-to-staff locations. In the United States, the US Department of Education has committed millions of dollars to funding innovative incentive proposals from states, cities, and individual schools which plan to use funding to enhance recruitment and retention in schools serving poor children. Many of the proposals will reward teachers for better student achievement in their own classrooms or in schools as a whole.

Current Challenges

With the tremendous strides that have been made in access to education for most countries, the current focus for educational organizations is on improvement of the quality of instruction for children. This requires economic support, knowledge about the teaching and learning process, and capable educators in central administrations, schools, and classrooms. The challenges for the field of personnel administration include generating an adequate supply of high-quality and diverse educators, decentralizing educational decision making, and responding to the increasing pressure for greater student achievement.

Teacher and Administrator Shortage Issues

Most countries are struggling to recruit and retain an adequate number of teachers and other professionals for schools (Toole and Louis, 2002). Many countries (e.g., United States, United Kingdom, India, and Philippines) are experiencing notable shortages of trained teachers or shortages in specific subject areas, such as math, science, and special education. There are also staffing gaps between high- and low-poverty schools. Even when trained teachers are hired, issues of retention are rampant due to a lack of satisfaction with existing working conditions and compensation. The quality of leadership in a school and working relationships with colleagues are major factors which influence commitment and motivation to remain in or leave a school (Day *et al.*, 2006).

In the United States, there is also a dramatic shortage of teachers of color – far fewer than the proportion of students of color in many classrooms. Although this shortage has been evident for decades, dramatic increases in the proportion of K-12 students of color and students who are English-language learners have drawn renewed attention to the need for a more diverse teaching staff. Colleges and universities must work collaboratively with K-12 schools to recruit diverse high school graduates into teacher-education programs so that they can earn educator credentials to work in K-12 schools. Historically, educators have been mostly first-generation college graduates and professionals, often coming from working-class families. Thus, recruiting future educators from among

lower income, students of color, or second-language learners may reduce the teacher shortage as well as address the need for greater diversity in the education profession.

In many Western European countries and the United States, long working hours and the increasing demand for student achievement are concerns among teachers and administrators alike but much more common in other countries (e.g., East Asia, Africa, and South America) are the low salaries, large class sizes, and poor facilities. Many educators must work two jobs to make a livable wage. In Mexico and Chile, teachers often teach in two different schools during the day and taxi between them to fulfill their assignments. Classes of 40–60 are the norm in many Asian countries, such as China and the Philippines, in contrast to 20–30 in the United States. Facilities in many places are overcrowded with insufficient materials, such as textbooks and desks. In some rural areas of Africa and other developing countries, there may be no facilities at all by Western standards. Teachers toil in all these circumstances, however, and provide instruction by any means possible.

Efforts are being made to address these problems by offering a whole range of inducements and supports. Signing bonuses have been used in the United States and the United Kingdom; induction programs are becoming commonplace in the United States; salaries are increasing, especially in East Asia due to improving economic conditions; and greater attention is being given to the quality of professional development. None of these is a panacea for the legitimate concerns of teachers worldwide but they reflect a responsive effort on the part of many countries to improve support for education.

Push for Business-Based Forms of Accountability

The forces of globalization that are shaping the business world have permeated the education field through accountability-driven national policies, which have shifted the focus from the processes of service delivery to the outputs of student achievement. These policies are top-down and threaten the professional prerogatives of educators at the local level by yoking accountability demands with prescribed curricula and policies. For example, most Asian countries have a centralized form of personnel administration which is described as a pyramid model with a central ministry of education that promulgates policy, programs, and logistics. Headmasters at individual schools are responsible for implementation of ministerial mandates in addition to school management, school and community relations, and instructional supervision. The pressure to improve student learning is substantial but is often constrained by the very programs and policies which are designed to extract it from the system.

Decentralized Leadership

Simultaneous to the push for greater accountability, there has been the contradictory trend toward greater decentralization with the hope that educators who are closer to the community will be more responsive to local conditions and that there will be greater community participation and support for the schools. This goal depends on the knowledge and skills of school leaders to negotiate these boundaries and, yet, often they have had little preparation for this increased autonomous role. In countries such as Denmark, New Zealand, and the Netherlands, this ideal is closer to a reality than in more hierarchical systems like those in the United States, France, and Germany where local autonomy is still relatively weak in most locations (Mulford, 2003).

Quality Instructional Improvement

Widespread changes throughout the world are taking place due to immigration, globalization, and economic restructuring; South Africa, Eastern Europe, and Asia are a few examples of these shifts. In many Asian countries, there has been increasing access to and quality of education due to improved economic conditions and a strong public commitment to education over the last 30 years. Enrollment of primary students is close to 100% in most countries but this rapid expansion has led to predictable problems with teacher and administrator quality. Central administrations are bloated and services, such as professional development, have not kept pace with the growth. While there are long-standing traditions of professionalism, especially in China and Japan, a greater focus on the understanding and delivery of high-quality instruction is predicted in the coming years for China, Pakistan, Vietnam, Indonesia, and other East Asian countries (Chapman, 2000) not unlike the current push for more effective instructional practices in the United States.

Summary

Personnel administration, as a specialization within educational administration, is a young and dynamic field of study encompassing a host of issues and considerations. The primary purposes of personnel administration, to attract, retain, develop, and motivate personnel to achieve organizational goals, are promoted currently through a complex array of personnel processes, policies, and practices reflecting its bureaucratic roots. Commonalities exist in the purposes and structural components of personnel administration across countries, but a host of converging expectations and pressures, especially cultural, shape the different meanings of those activities for educators. As a result, there are few best practices with universal utility due to the contextual interpretations of these activities but a push for greater decentralization and local autonomy to establish

professional learning communities seems to have wide appeal and shifts responsibility for many of these formerly centralized functions down to the school level so that personnel decisions are made to fit the local context and serve the needs of the existing learning community.

See also: Adult Learning; Characteristics of Adult Learning; Contemporary Approaches to Teacher Professional Development; Evaluation of Teacher Quality and Practice; Leadership for Learning; Organizational Learning in Schools; Professional Learning Community; Research on Educational Leadership – Approaches/Promising Directions; Teacher Induction; Teacher Learning as Workplace Learning; Teachers Career Stages and Professional Development; Trends in Workplace Learning Research.

Bibliography

- Brewster, C. and Tyson, S. (eds.) (1991). *International Comparisons in Human Resource Management*. London: Pitman.
- Campion, M., Persell, E., and Brown, B. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology* **41**, 25–42.
- Chapman, D. W. (2000). Trends in educational administration in developing Asia. *Education Administration Quarterly* **36**(2), 283–308.
- Day, C., Stobart, G., Sammons, P., et al. (2006). *Variations in Teachers' Work, Lives, and Effectiveness*. Research Brief No. RB743. London: Department for Education and Skills.
- Delli, D. A. and Vera, E. M. (2003). Psychological and contextual influences on the teacher selection interview: A model for future research. *Journal of Personnel Evaluation in Education* **17**(2), 137–156.
- Foskett, N. and Lumby, J. (2003). *Leading and Managing Education: International Dimensions*. London: Paul Chapman.
- Grossman, P. and Thompson, C. (2004). District policy and beginning teachers: A lens on teacher learning. *Educational Evaluation and Policy Analysis* **26**(4), 281–301.
- Hall, G. and Hord, S. (2001). *Implementing Change: Patterns, Principles, and Pitfalls*. Boston, MA: Allyn and Bacon.
- Joint Committee on Standards for Educational Evaluation (1988). *The Personnel Evaluation Standards: How to Assess Systems for Evaluating Educators*. Newbury Park, CA: Corwin.
- Joyce, B. R. and Showers, B. (2002). *Student Achievement through Staff Development*, 3rd edn. Alexandria, VA: Association for Supervision and Curriculum Development.
- Liu, S. and Teddlie, C. (2003). The ongoing development of teacher evaluation and curriculum reform in the People's Republic of China. *Journal of Personnel Evaluation in Education* **17**(3), 243–261.
- Mulford, B. (2003). School leaders: Changing roles and impact on teacher and school effectiveness. *Paper Commissioned by the Organization for Economic Cooperation and Development*. Education and Training Policy Division, OECD, Paris.
- Pounder, D. G. (1988). Improving the predictive validity of teacher selection decisions: Lessons from teacher appraisal. *Journal of Personnel Evaluation in Education* **2**(2), 141–150.
- Smith, T. M. and Ingersoll, R. M. (2004). What are the effects of induction and mentoring on beginning teacher turnover. *American Educational Research Journal* **41**(3), 681–714.
- Teddlie, C., Stringfield, S., and Burdett, J. (2003). International comparisons of the relationships among educational effectiveness, evaluation and improvement variables: An overview. *Journal of Personnel Evaluation in Education* **17**(1), 5–20.
- Thomas, M. R. (1990). Teacher supply systems. In Thomas, M. R. (ed.) *International Comparative Education: Priorities, Issues and Prospects*, pp 163–193. Oxford: Pergamon.
- Toole, J. C. and Louis, K. S. (2002). The role of professional learning communities in international education. In Leithwood, K. and Hallinger, P. (eds.) *Second International Handbook of Educational Leadership and Administration*, pp 245–279. Dordrecht: Kluwer Academic Publishers.
- Tucker, P. D. and DeSander, M. K. (2006). Legal considerations in designing teacher evaluation systems. In Stronge, J. H. (ed.) *Evaluating Teaching*, pp 69–97. Thousand Oaks, CA: Corwin.
- Walker, A. and Dimmock, C. (2002). Moving school leadership beyond its narrow boundaries: Developing a cross-cultural approach. In Leithwood, K. and Hallinger, P. (eds.) *Second International Handbook of Educational Leadership and Administration*, pp 167–202. Dordrecht: Kluwer Academic Publishers.

Further Reading

- Bolam, R., McMahon, A., Stoll, L., Thomas, S., and Wallace, M. (2005). *Creating and Sustaining Effective Professional Learning Communities*. Research Report RR637. London: UK Department for Education and Skills.
- Leithwood, K. and Louis, K. S. (eds.) (1998). *Organizational Learning in Schools*. Lisse: Swets and Zeitlinger.
- Mawhinney, H., Hass, J., and Wood, C. (2005). School district design for improving instructional capacity: Assessing teachers' collective efficacy beliefs and perceptions of school professional learning community maturity. *Paper Presented at the Annual Conference of the American Educational Research Association*. Montreal, Canada, April.
- Mulford, B., Silins, H., and Leithwood, K. (2004). *Leadership for Organizational Learning and Student Outcomes: A Problem-Based Learning Approach*. Dordrecht: Kluwer.
- Sheppard, B. and Brown, J. (2000). The transformation of secondary schools into learning organizations. In Leithwood, K. (ed.) *Understanding Schools as Intelligent Systems*, pp 293–314. Stamford, CT: JAI Press.
- Silins, H. and Mulford, B. (2004). Schools as learning organizations: Effects on teacher leadership and student outcomes. *School Effectiveness and School Improvement* **15**(3–4), 443–466.

The Big Challenge: Leadership for Inclusion

M Ainscow and A Sandill, University of Manchester, Manchester, UK

© 2010 Elsevier Ltd. All rights reserved.

The issue of how to develop more inclusive forms of education is arguably the biggest challenge facing school systems throughout the world. In economically poorer countries, the priority has to be with the millions of children who never see the inside of a classroom. Meanwhile, in wealthier countries some young people leave school with no worthwhile qualifications, while others are placed in various forms of special provision away from mainstream educational experiences, and some simply choose to drop out since the lessons seem irrelevant to their lives.

Faced with this challenge, there is evidence of an increased interest in the idea of inclusive education. However, the field remains confused as to what actions need to be taken in order to move policy and practice in a more inclusive direction. Here, we explore possible ways forward, looking at what organizational conditions are needed in order to foster inclusive policies and practices, and what this means for the role of leadership.

Inclusive Education

In some countries, inclusive education is thought of as an approach to serving children with disabilities within general-education settings. Internationally, however, it is increasingly seen more broadly as a reform that supports and welcomes diversity among all learners. The argument developed here adopts this broader formulation. It presumes that the aim of inclusive education is to eliminate social exclusion that is a consequence of attitudes and responses to diversity in race, social class, ethnicity, religion, gender, and ability. As such, it starts from the belief that education is a basic human right and the foundation for a more just society.

More than a decade ago, the United Nations Educational, Scientific and Cultural organization (UNESCO) World Conference on Special Needs Education, held in Salamanca, endorsed the idea of inclusive education. Arguably, the most significant international document that has ever appeared in the special needs field, the Salamanca Statement argues that regular schools with an inclusive orientation are the most effective means of combating discriminatory attitudes, building an inclusive society and achieving education for all. Furthermore, it suggests that such schools can “provide an effective education for the majority of children and improve the efficiency and ultimately the cost-effectiveness of the entire education system” (p. ix).

During the subsequent years, there has been considerable activity in many countries to move educational policy and practice in a more inclusive direction but there is still considerable uncertainty as to how best to proceed. The confusion that exists arises, at least in part, from the fact that inclusion can be defined in a variety of ways. It is also important to remember that there is no one perspective on inclusion within a single country or school.

Given the confusion and uncertainties that exist as policymakers and practitioners seek to make sense of different perspectives, advancing toward the implementation of inclusive education is far from easy. Moreover, it must not be assumed that there is full acceptance of the inclusive philosophy. Consequently, as we consider ways of developing schools that are effective in reaching all children, it is necessary to recognize that the field itself is riddled with uncertainties, disputes, and contradictions. However, what can be said is that throughout the world, attempts are being made to provide more effective educational responses for all children, whatever their characteristics, and that, encouraged by the Salamanca Statement, the overall trend is toward making these responses within the context of general-educational provision (see the special edition of the *European Journal of Psychology of Education*, December 2006).

Inclusive Practice

A recent study, carried out by Ainscow, Booth, and Dyson in the United Kingdom, attempted to throw light on what needs to happen in order to develop inclusive practices in schools. It concluded that the development of inclusive practice is not, in the main, about adopting new technologies of the sort described in much of the existing literature. Rather, it involves social learning processes within a given workplace that influence people's actions and, indeed, the thinking that informs these actions. This led the authors to seek a deeper understanding of what these processes involve, using the ideas of Etienne Wenger, who provides a framework that can be used to analyze the development of practices in social contexts. At the center of this framework is the concept of a community of practice, a social group engaged in the sustained pursuit of a shared enterprise. This suggests that practices are ways of negotiating meaning through social action. Wenger argues, for example, that a particular strategy may be developed as part of a school's

planning activities and summarized in a set of guidance for action, providing a codified reification of intended practice. However, the meaning and practical implications of the strategy only becomes clear as it is used and discussed between colleagues.

The implication is that a methodology for developing inclusive practices must take account of such social processes of learning that go on within particular contexts. It requires a group of stakeholders to look for a common agenda to guide their discussions of practice and, at much the same time, a series of struggles to establish ways of working that enable them to collect and find meaning in different types of information.

Similarly important, therefore, is the development of a common language with which colleagues can talk to one another and indeed to themselves about detailed aspects of their practice. It seems, moreover, that without such a language, teachers find it very difficult to experiment with new possibilities. It has been noted, for example, that when researchers report to teachers what has been observed during their lessons they will often express surprise. Much of what teachers do during the intensive encounters that occur in a typical lesson is carried out at an automatic, intuitive level, involving the use of tacit knowledge. Furthermore, there is little time to stop and think. This is perhaps why having the opportunity to see colleagues at work is so crucial to the success of attempts to develop practice. It is through such shared experiences that colleagues can help one another to articulate what they currently do and define what they might like to do. It is also the means whereby taken-for-granted assumptions about particular groups of students can be subjected to mutual critique.

Research has drawn attention to certain ways of engaging with evidence that seem to be helpful in encouraging such dialog. This can help to create space for reappraisal and rethinking by interrupting existing discourses, and by focusing attention on overlooked possibilities for moving practice forward. Particularly powerful techniques in this respect involve the use of mutual observation, sometimes through video recordings, and evidence collected from students about teaching and learning arrangements within a school. Under certain conditions, such approaches provide interruptions that help to make the familiar unfamiliar in ways that stimulate self-questioning, creativity, and action. In so doing they can sometimes lead to a reframing of perceived problems that, in turn, draws the teacher's attention to overlooked possibilities for addressing barriers to participation and learning.

Here, the role of the school principal in providing leadership for such processes is crucial. Indeed, as a result of their extensive literature review, Leithwood and Riehl contend that developing people by providing intellectual stimulation has to be one of the core practices of effective

leaders. Lambert and her colleagues seem to be talking about a similar approach when they stress the importance of leaders gathering, generating, and interpreting information within a school in order to create an inquiring stance. They argue that such information causes disequilibrium in thinking and, as a result, provides a challenge to existing assumptions about teaching and learning.

All of this underlines the way norms of teaching are socially negotiated within the everyday context of the communities of practice within schools. In this sense, it is evidence of how the culture of the workplace affects how teachers see their work and indeed their students. It underlines the idea that the development of more inclusive approaches does not arise from a mechanical process in which any one specific organizational restructuring, or the introduction of a particular set of techniques, generates increased levels of participation. Rather, as we have argued, the development of inclusive practices requires processes of social learning within particular organizational contexts.

Organizational Factors

What, then, are the organizational conditions that can help foster such social learning? In other words, how can schools become more inclusive? Where writers have addressed these questions, they tend to give particular emphasis to the characteristics of schools which stimulate and support processes of interrogation and reflection. For example, Skrtic argues that schools with what he calls *adhocratic* configurations are most likely to respond to student diversity in positive and creative ways. Such schools emphasize the pooling of different professional expertise in collaborative processes. Children who cannot easily be educated within the school's established routines, are not seen as having problems, but as challenging teachers to re-examine their practices in order to make them more responsive and flexible. Other authors suggest organizational conditions – distributed leadership, high levels of staff and student involvement, joint planning, a commitment to enquiry, and so on – that promote collaboration and problem solving among staff, and which, therefore, produce more inclusive responses to diversity. Literature regarding educating children in multicultural contexts also highlights similar processes of meaning making, such as openness to experiment and innovate when resolving issues as they emerge, focusing on the needs of the individual child, and giving teachers the freedom to take initiative.

These themes are supported by a recent international literature review, carried out by Dyson and his colleagues, that examines the effectiveness of school actions in promoting inclusion. It concludes that there is a limited but by no means negligible body of empirical evidence about the relationship between school action and the participation of

all students in the cultures, curricula, and communities of their schools. In summary, it suggests that:

- Some schools are characterized by an inclusive culture. Within such schools, there is some degree of consensus amongst adults around values of respect for difference and a commitment to offering all pupils access to learning opportunities. This consensus may not be total and may not necessarily remove all tensions or contradictions in practice. On the other hand, there is likely to be a high level of staff collaboration and joint problem solving, and similar values and commitments may extend into the student body, and into parent and other community stakeholders in the school.
- The extent to which such inclusive cultures lead directly and unproblematically to enhanced student participation is not clear. Some aspects of these cultures, however, can be seen as participatory by definition. For instance, respect for diversity from teachers may itself be understood as a form of participation by children within a school community. Moreover, schools characterized by such cultures are also likely to favor forms of organization (such as specialist provision being made within the ordinary classroom, rather than by withdrawal from lessons) and practice (such as constructivist approaches to teaching and learning) which could be regarded as participatory by definition.
- Schools with inclusive cultures are also likely to be characterized by the presence of leaders who are committed to inclusive values and to a leadership style which encourages a range of individuals to participate in leadership functions. Such schools are also likely to have good links with parents and with their communities.
- The local and national policy environment can act to support or to undermine the realization of schools' inclusive values.

On the basis of this evidence, the Dyson review team suggest that attempts to develop inclusive schools should pay attention to the development of inclusive cultures and, particularly, to the building of some degree of consensus around inclusive values within school communities. This leads them to argue that school leaders should be selected and trained in the light of their commitment to inclusive values and their capacity to lead in a participatory manner.

In summarizing the current knowledge base on educational leadership, Leithwood and Riehl conclude that in diverse student environments, particular forms of leadership can be effective in promoting school quality, equity, and social justice through more powerful forms of teaching and learning, creating strong communities of students, teachers and parents, and nurturing educational cultures among families. Such approaches are congruent with the view that inclusion is essentially about attempts to embody particular values in particular

contexts. Unlike mechanistic views of school improvement, they acknowledge that decisions about how to improve schools always involve moral and political reasoning, as well as technical considerations. Moreover, they offer specific processes through which inclusive developments might be promoted. Discussions of inclusion and exclusion can help, therefore, to make explicit the values which underlie what, how, and why changes should be made in schools. Inclusive cultures, underpinned by particular organizational conditions, may make those discussions more likely to occur and more productive when they do occur. A helpful list of indicators in relation to this analysis is provided by the 'Index for inclusion,' a review framework for examining school factors that constitute barriers to learning and participation, developed by Booth and Ainscow.

Culture and Leadership

All of this underlines the importance of cultural factors. Cultures are about the deeper levels of basic assumptions and beliefs that are shared by members of an organization, operating unconsciously to define how they view themselves and their working contexts. The extent to which these values include the acceptance and celebration of difference, and a commitment to offering educational opportunities to all students, coupled with the extent to which they are shared across a school staff, relate to the extent to which students are enabled to participate.

Cultures can be seen as having a reality-defining function, enabling those within an institution to make sense of themselves, their actions, and their environment. Hargreaves argues that a current reality-defining function of culture is often a problem-solving function inherited from the past. In this way, today's cultural form created to solve an emergent problem often becomes tomorrow's taken-for-granted recipe for dealing with matters shorn of their novelty.

Changing the norms that exist within a school is difficult to achieve, particularly within a context that is faced with so many competing pressures and where practitioners tend to work alone in addressing the problems they face. On the other hand, the presence of children who are not suited to the existing menu of the school can provide some encouragement to explore a more collaborative culture within which teachers support one another in experimenting with new teaching responses. In this way, problem-solving activities gradually become the reality defining, taken-for-granted functions that are the culture of a school that is more geared to fostering inclusive ways of working.

The implication of all of this is that becoming more inclusive is a matter of thinking and talking, reviewing and refining practice, and making attempts to develop a more

inclusive culture. Such a conceptualization means that inclusion cannot be divorced from the contexts within which it is developing, nor the social relations that might sustain or limit that development. This suggests that it is in the complex interplay between individuals, and between groups of individuals, that shared beliefs and values and change occur, and that it is impossible to separate those beliefs from the relationships in which they are embodied.

Alternatively, some researchers argue that in order to bring about the cultural change that inclusion demands, it is essential to consider the values underlying the intended changes. Thus, cultural change is directed toward a transformative view of inclusion, in which diversity is seen as making a positive contribution to the creation of responsive educational settings. This involves developing the capacity of those within schools to reveal and challenge deeply entrenched deficit views of difference, which define certain types of students as lacking something. Writers who are involved in facilitating and evaluating such processes in schools repeatedly identify the role of leadership as critical for bringing about and sustaining such changes.

It seems, then, that the principle of inclusion is likely to require challenges to the thinking of those within a particular organization and, inevitably, this raises questions regarding forms of leadership. Lambert and her colleagues argue for constructivist leadership as a strategy for responding to learner diversity. This involves reciprocal processes that enable participants in an educational community to construct common meanings that lead toward a common purpose about schooling. They use this perspective to argue that leadership involves an interactive process entered into by both students and teachers. Consequently, there is a need for shared leadership, with the principal seen as a leader of leaders. Hierarchical structures have to be replaced by shared responsibility in a community that becomes characterized by agreed values and hopes, such that many of the control functions associated with school leadership become less important or even counterproductive.

Leithwood and Riehl refer to two approaches to school leadership, one with an orientation to student achievement and the other with a focus on meeting the needs of individuals. They contend that leaders serving diverse schools need to use both approaches in order to perform their role effectively. Others argue that school leaders need to strike a continual balance between concern for people, including teachers, students, parents, and the community that the school serves, and accountability toward government mandates.

Much of the literature on the role of leadership in relation to school improvement places emphasis on the importance of social relationships. It has been argued, for example, that leaders may structure staff working relationships in one of three ways: competitively, individualistically,

or cooperatively. Within a competitive structure, teachers work against each other to achieve a goal that only a few can attain; an individualistic structure exists when teachers work alone to accomplish goals that are unrelated to the goals of their colleagues; whereas, a cooperative structure exists when teachers coordinate their efforts to achieve joint goals. This means that, to maximize the productivity of a school, principals have to: challenge the *status quo* of traditional competitive and individualistic approaches to teaching; inspire a clear mutual vision of what the school should and could be; empower staff through cooperative team work; lead by example, using cooperative procedures and taking risks; and encourage staff members to persist and keep striving to improve their expertise.

The most helpful theoretical and empirical leads, however, are provided by Riehl, who, following an extensive review of literature, develops a comprehensive approach to school administration and diversity. She concludes that school leaders need to attend to three broad types of tasks: fostering new meanings about diversity; promoting inclusive practices within schools; and building connections between schools and communities. She goes on to consider how these tasks can be accomplished, exploring how the concept of practice, especially discursive practice, can contribute to a fuller understanding of the work of school principals. This analysis leads the author to offer a positive view of the potential for school principals to engage in inclusive, transformative developments. She concludes: "When wedded to a relentless commitment to equity, voice, and social justice, administrators' efforts in the tasks of sensemaking, promoting inclusive cultures and practices in schools, and building positive relationships outside of the school may indeed foster a new form of practice" (p. 71).

The Role of Networking

What emerges from the evidence we have summarized is the importance of forms of leadership that encourage social learning processes within particular contexts. Such processes can, we have argued, be stimulated by inquiry which fosters a greater capacity for responding to learner diversity. Achieving a deeper and more sustainable impact on the culture of schools is, however, much more difficult. This necessitates longer-term, persistent strategies for capacity building at the school level. It also requires new thinking and, indeed, new relationships at the systems level. In other words, efforts to foster inclusive school development are more likely to be effective when they are part of a wider strategy.

This has led to an increasing emphasis on the idea of sharing expertise and resources between schools. Such an approach is consistent with what has been called public

value management, with its emphasis on network governance. The origins of this approach can be traced to criticisms of the current emphasis in some Western countries on strategies drawn from private-sector experience. The formulation of what constitutes public value can only be achieved through deliberation involving key stakeholders and actions that depend on mixing in a reflexive manner a range of intervention options. Consequently, networks of deliberation and delivery are seen as key strategies. In education services, this implies the negotiation of new, interdependent relationships between schools, administrations, and communities. With this in mind, some researchers argue for system leadership. This requires senior staff to widen their sphere of engagement by interacting with other schools in order to develop lateral capacity building toward sustainable development.

Drawing on a series of studies, Ainscow and West present evidence from England suggesting that school-to-school collaboration can strengthen the capacity of individual organizations to respond to learner diversity. However, this does not represent an easy option for schools, particularly in policy contexts within which competition and choice continue to be the main drivers. Recent studies, for the most part, have focused on situations where schools have been given short-term financial incentives linked to the demonstration of collaborative planning and activity. They suggest that collaboration between schools can help to reduce the polarization of schools, to the particular benefit of those students who are marginalized at the edges of the system and whose performance and attitudes cause concern. There is evidence, too, that when schools seek to develop more collaborative ways of working, this can have an impact on how teachers perceive themselves and their work. Specifically, comparisons of practices can lead teachers to view underachieving students in a new light. Rather than simply presenting problems that are assumed to be insurmountable, such students may be perceived as providing feedback on existing classroom arrangements. In this way they may be seen as sources of understanding as to how these arrangements might be developed in ways that could be of benefit to other members of the class.

Concluding Remarks

This article has addressed what we see as the biggest challenge for countries around the world: that of developing more inclusive education systems. The approach we have outlined is not about the introduction of particular techniques or organizational arrangements. Rather it places emphasis on processes of social learning within particular contexts. The use of evidence as a means of stimulating experimentation, and collaboration within

and between schools, are seen as key strategies. Copland argues that inquiry can be the engine to enable the distribution of leadership that is needed in order to foster participation, and the glue that can bind a school community together around a common purpose.

All of this has major implications for leadership practice at different levels within schools and education systems. In particular, it calls for efforts to encourage coordinated and sustained efforts by whole staff groups around the idea that changing outcomes for all students is unlikely to be achieved unless there are changes in the behaviors of adults. Consequently, the starting point must be with staff members: in effect, enlarging their capacity to imagine what might be achieved, and increasing their sense of accountability for bringing this about. This may also involve tackling taken-for-granted assumptions, most often relating to expectations about certain groups of students, their capabilities, and behaviors.

Our argument is, then, based on the assumption that schools know more than they use and that the logical starting point for inclusive school development is with a detailed analysis of existing arrangements. This allows good practices to be identified and shared, while, at the same time, drawing attention to ways of working that may be creating barriers to the participation and learning of some students. However, as we have stressed, the focus must not only be on practice. It should also address and sometimes challenge the thinking behind existing ways of working.

Further Reading

- Ainscow, M., Booth, T., and Dyson, A. (2006a). *Improving Schools, Developing Inclusion*. London: Routledge.
- Ainscow, M. and West, M. (2006). *Improvement Urban Schools: Leadership and Collaboration*. Buckingham: Open University Press.
- Booth, T. and Ainscow, M. (2002). *Index for Inclusion: Developing Learning and Participation in Schools*. Bristol: CSIE.
- Copland, M. A. (2003). Leadership of inquiry: Building and sustaining capacity for school improvement. *Educational Evaluation and Policy Analysis* 25(4), 375–395.
- Dyson, A., Howes, A., and Roberts, B. (2002). A systematic review of the effectiveness of school-level actions for promoting participation by all students (EPPI-Centre Review, version 1.1). Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Hargreaves, D. H. (1995). School culture, school effectiveness and school impact. *School Effectiveness and School Improvement* 6(1), 23–46.
- Lambert, L., Walker, D., Zimmerman, D. P., et al. (2002). *The Constructivist Leader*. New York: Teachers College Press.
- Leithwood, K. A and Riehl, C. J. (2005). What do we already know about educational leadership? In Firestone, W. A. (ed.) *A New Agenda for Research in Educational Leadership*, pp 12–27. New York: Teachers College Press.
- Riehl, C. J. (2000). The principal's role in creating inclusive schools for diverse students: A review of normative, empirical and critical literature on the practices of educational administration. *Review of Educational Research* 70(1), 55–81.

- Skrtic, T. (1991). *Behind Special Education: A Critical Analysis of Professional Culture and School Organization*. Denver, CO: Love.
- UNESCO (1994). *The Salamanca Statement and Framework for Action on Special Needs Education*. Spain: UNESCO/ Ministry of Education and Science.
- Unesco (2001). *Open File on Inclusive Education: Support Materials for Managers and Administrators*. Paris: UNESCO. <http://unesdoc.unesco.org/images/0012/001252/125237eo.pdf> (accessed August 2009).
- (2006). *Educational Management Administration and Leadership* 34(2). (Special edition on Leadership and Diversity.)
- Wenger, E. (1998). *Communities of practice*. Cambridge: University Press.

Relevant Websites

- <http://inclusion.uwe.ac.uk> – Centre for Studies on Inclusive Education.
- <http://www.eenet.org.uk> – Enabling Education Network.
- <http://www.inclusive-solutions.com> – Inclusive Solutions.
- <http://inclusion.uwe.ac.uk> – Inclusion UK.
- <http://www.isec2005.org.uk> – Inclusive and Supportive Education Congress (ISEC) International Special Education Conference. Theme-Inclusion: Celebrating Diversity? Index to ISEC 2005 Presentations.
- <http://www.inclusion.com> – Marsha Forest Centre.
- <http://www.ncsl.org.uk/> – National College for Leadership.

The Educational Leader's New Work

P Gronn, University of Cambridge, Cambridge, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The term educational leader includes a potentially large and diverse category of people who exercise both formal and informal responsibilities for training, development, growth, and learning. Not only may such people be located in a range of educational settings and at a number of points in the life cycle of learning (i.e., from nurseries and preschools through to higher education), but they may also be employed in different workforce sectors of the economy. In the face of this scope and variety, some caution is therefore required to avoid oversimplified understandings of leadership. While the hallmark of leaders who warrant the description educational is probably an overarching commitment to the intrinsic worth of learning and development in the realization of human potential, the particular features that distinguish educationalists as leaders are less straightforward. Traditionally, leadership discourse has attributed the attainment of human purposes through collectively organized activity to the influence of one or a small number of strategically positioned individuals. Such attributions operate both retrospectively and prospectively: they provide explanations for the outcomes of past actions and are also expressed as expectations of people's future activities. Each set of attributions entails implicit theories of causation with regard to human conduct. In recent educational reform, particularly policies and measures concerned with the modernization of school education, where an imperative for change has become so strong, the weight of expectations attached to the agency of leaders and their leadership is large, perhaps even grossly inflated. Moreover, as indicated by a recent interest within the school leadership field in distributed leadership and organizational capacity building, expectations of the exercise of leadership have also spread well beyond the customary preserve of a few individuals. In Scottish schools, for example, it is currently anticipated that leadership will not only be provided by headteachers, but also by depute headteachers and principal teachers. Nationally, there are more than 14 000 educators in these three groups who, together, are officially designated as Scottish educational leaders (Scottish Executive Education Department, 2007: para 5.3.1).

Notwithstanding this heightened expectation of organizational leadership being diffused from senior level to mid-level and beyond in schools, in this article, attention is mainly focused on the work of headteachers

(an interchangeable title with principals). Apart from limitations of space, there are two reasons for this choice of focus. First, the amount of overall responsibility and accountability required of headteachers has become exceedingly high and demanding. Indeed, in a governance era of outcome-based and target-driven service delivery, the leadership anticipated of headteachers is also high-stakes leadership and arguably, the prominence accorded school leadership has never been greater. Second, perversely perhaps, at the same time as these role pressures on headteachers have increased, there is also accumulating evidence of difficulties in recruiting headteachers and principals of schools in both the government and nongovernment sectors in a number of countries; examples are Australia (Anderson *et al.*, 2007: 47–57), and England and Wales (Howson, 2007: 142–145). There may be a paradox in the uneasy coexistence of these two trends. That is, one unintended consequence of a strengthened attachment to leadership may be a diminution of teachers' willingness to become leaders – at least in sufficient numbers to replenish current and anticipated shortfalls of headteachers. In short, insofar as there is evidence of a global leadership conversation in education, this is occurring at two countervailing levels: vociferous exhortations about the salience of leaders and leadership on the one hand, yet nagging evidence of disengagement from leadership on the other. In the remainder of the article, I try to account for this paradox by developing the idea that educational leadership might profitably be understood as a form of greedy work (Gronn, 2003: 147–156).

Nature of Leadership Practice

All leadership practice is contextualized. Contexts for understanding leadership practice range from the global through societal to the immediate or situational. Globally and societally, a large and accumulating body of literature has emerged in the last two decades or so that has promoted the idea of leadership as virtuous conduct. Some of this material has been written for nonspecialist audiences and is widely accessible in magazines and books, while some of it is confined to scholarly publishing outlets. Still more of it has emanated from think piece discussion documents distributed by influential public sector agencies (e.g., Audit Scotland, 2005). Much of this overall corpus of publications, along with published memoirs and autobiographies of leaders, has provided a fertile

source for quotable quotes about leadership, recipe-style advice, and widely circulating and highly appealing claims to the effect that leadership matters, that leaders do the so-called right things, and that leaders turn around ailing and failing organizations. In conjunction with an extensive market for consultancy advice, this literature is evidence of the existence of a leadership industry. What may have commenced life in this industry as the promotion of a simple virtue, however, often ends up fueling a false promise of heroic leader exceptionalism (Gronn, 2004: 352).

There are two main ways in which this wider climate of expectations has filtered through to situational contexts of school leadership, in particular to redefine the role of headteachers. First, as school leaders, the work practices of heads are increasingly governed by performance standards. Second, the productivity of heads and their schools has been radically reshaped by a learning improvement agenda. Both developments reflect the significance that politicians, policymakers, and opinion leaders generally have attached to accountability.

Standards are a means of measuring the quality of performance, although they serve different purposes for different groups. When they are utilized by professional bodies, for example, standards provide a vehicle for professional learning and recognition of quality and in the case of leadership, an assumption is that "prospective and established school leaders should keep up with developments in research and knowledge in their area of teaching and meet standards for accomplished practice" (Ingvarson *et al.*, 2006: 41). If they are adopted by employing authorities and agencies, on the other hand, standards are a means by which leaders fulfill their contractual obligations as employees. In the second sense, standards offer an alternative to satisfactory tertiary course completion as evidence of professional competence. They may be used by employers not merely to accredit and regulate the actions of incumbent school heads, but also to influence their ongoing professional preparation, selection, and appointment. To some extent, these latter developments express the willingness of employers to commit to preferred normative prototypes or profiles of headteachers and embody these in statements of standards. This trend has been characterized as designer leadership (Gronn, 2003: 7).

The notion of an improvement agenda refers to the link that is increasingly being made between leadership and learning. Improvement lends itself to a number of different interpretations. Equally, there are different ways of measuring and providing evidence of improvement. Essentially, improvement is about the benefits and gains, both immediate and long term, that accrue to students as a result of their learning. At the heart of the leadership-learning link, there is a presumption about causality; that is, directly through their own actions and/or through the indirect influence of their actions on colleagues, school

leaders will enhance the learning of students. Typically, performance assumptions and targets in relation to improvement find practical expression in national curriculum statements. An example from Scotland is the *Curriculum for Excellence* (Scottish Government, 2006). Here, the leadership for learning connection finds practical expression in two ways: fostering an environment for learning and providing arrangements for learning that enshrine five core curriculum values, which in turn are consistent with a series of seven curriculum design principles. In regard to these two measures (i.e., culture building and the management of learning), the expectation of school heads by those responsible for monitoring school performance (in Scotland, Her Majesty's Inspectorate in Education) is that heads will initiate site-based changes to facilitate improved learning. These might entail new structures, processes, procurement of additional resources, and other measures that may be calculated to build school-wide capacity.

Determinants of Leadership Practice

To what extent may these kinds of expectations about the causal role of leaders be justifiable? Much of the thinking that underpinned the euphoria associated with the leadership of high-profile individuals in the recent past was grounded in questionable assumptions. The main reason for this claim is that expectations of leaders rest on a cognitive misattribution, namely an exaggerated understanding of the agency of leaders and their scope for autonomous action. (Interesting in this connection is the fact that in Australia, sections 14k and 31k of the Schools Assistance (Learning Together – Achievement through Choice and Opportunity) Act, 2004 (Act No. 135) specify as one of the conditions of federal funding of schools a commitment to strengthened principal autonomy by state governments and nongovernment authorities.) The reality of leadership practice, by contrast, is that the autonomy of leaders is never absolute and is always circumscribed. This is partly because there are limits to individual and collective human capacity. As Simon (1976/1945), an early writer in the field, pointed out, due to their inherent cognitive limitations, executives typically satisfice rather than maximize when they make decisions. That is, because human rationality is bounded and because it operates within an environment of information that is less than optimal, a particular course of action can only ever hope to be "satisfactory or good enough" (Simon, 1976/1945: xxix). Writing at about the same time as Simon, Hayek (1945: 520) reinforced this point when he exposed the fallacy of what he termed omniscient minds. That is, there is no one mind that is capable of knowing all that needs to be known about a particular problem or a range of problems. In addition to these inherent limitations on

capacity, Stewart's (1982) model of managerial practice has highlighted the highly contingent and circumscribed nature of leader autonomy. (Although this model was originally couched in the language of management, in a recent symposium on her work, Stewart (2003: 202) has affirmed that it applies equally to leaders.)

In Stewart's model, the choices made by managers and leaders are determined by the demands and constraints that shape their work. Demands refer to the components of a role that, as it were, go with the territory. These set the framework within which leaders act and are primarily manifest in job descriptions and role prescriptions. They are "what must be done" (Stewart, 1982: 9). Thus, in the early 2000s in the state of Victoria, Australia, there were 52 bullet-point accountabilities in 11 program areas for principals (Department of Education, Employment and Training, 2001: 47–51). Constraints impinge on leaders at a number of contextual levels (e.g., systemic, organizational, and situational). These constraints comprise factors that are both internal and external to an organization, and which impose limits on what jobholders can do. Some constraints, particularly external factors, may be standardized across a system and are experienced uniformly, regardless of particular settings. Examples include legal frameworks (e.g., industrial relations, health and safety) and national and local authority policies and procedures (e.g., personnel, curriculum, evaluation, and finance). Other factors, such as the quality and quantity of a school's overall resources, may be experienced as more or less constraining. Examples of such relativities and their impact might include the collective body of expertise and experience of teachers, rates of teacher retention and turnover, and the strength of community support and social capital, all of which contribute to the improvement (or nonimprovement) of students' learning. For Stewart (1982: 9), choices represent a domain of opportunities. Choices may be understood in two senses. First, there may be discretion for managers and leaders to do different things (i.e., perform self-chosen, nonmandated work). In this way, proactive or innovative individuals may modify their immediate or wider role sets' expectations of what they should do (Fondas and Stewart, 1994: 97). Second, there may be scope for leaders and managers to do things differently. As an example, two executives who are required to perform the same work may do so in sharply contrasting ways (see, e.g., De Roche, 1994).

It is the latitude for choice and discretion that provides leaders with their sources of work satisfaction, fulfillment, and self-actualization. On the other hand, it is when the incentives to perform their work begin to dry up, and the constraints and demands cease to be enabling and become burdensome, that leaders come to experience their work as greedy and all-consuming. At that point, some form of disengagement may become an option.

Satisfaction in Leadership Practice

In addition to the expectations, demands, constraints, and choices that frame the work of leaders, leadership practice requires a significant investment of personal resources. In keeping with other forms of service work, perhaps the most important of these resources is emotional labor (Hochschild, 1983). As with other occupational commitments, people tend not to become leaders by default or to fall automatically into leadership roles. Rather, they make conscious decisions to commit themselves to undertake role responsibilities. For that commitment to occur, they have to engage in the emotion work associated with role transition and the requirements of career mobility.

There are at least two senses in which leadership entails emotional labor. First, to the extent to which leadership is a people-centered activity, the work itself requires the self-aware management of one's own emotions as a leader along with those of one's colleagues and other groups. This requirement may entail the conscious containment (or even denial) of one's own feelings and emotional needs, and it may also mean that one becomes an object of or container for the projected and displaced emotions (both positive and negative) of colleagues. Second, leadership also entails emotional labor because it is about identity work. That is, to be able to make the transition to leadership and undertake the psychological commitment that is required necessitate a depth of self-engagement and identification with a role such that self and role become integrated. It is for this reason that individuals may sometimes be said to lose themselves in their work (Kahn, 1992: 326). Thus, in one study of principals and deputy principals in Australia (Victoria, Department of Education and Training, 2004), 750 principal class respondents were found to have a very high sense of vocationalism (i.e., in the traditional sense of a calling to serve other people). That is, they claimed that they experienced their jobs as a way of life to which they were very strongly committed personally and professionally. Such a depth of commitment, however, was found to leave people potentially vulnerable to emotional overload or stress. With about 80% of the respondents in this same study reporting high levels of stress (Victoria, Department of Education and Training, 2004: 31), there was a strong possibility that their current vocational commitments might be transferred to other occupations (Victoria, Department of Education and Training, 2004: 26). These findings suggest that the point at which sources of dissatisfaction begin to outweigh sources of satisfaction in service work such as leadership, particularly in a high-stakes accountability climate, is very finely balanced. If so, then a key question that arises is: When do self-actualization, the realization of one's values as a leader, and the desire to give expression to one's identity and to serve others begin to give way in the face of emotional wear and tear, a sense of lack of support, and perhaps even abandonment?

Dissatisfaction in Leadership Practice

There are a number of discontents which may cause leaders, particularly headteachers, to experience their work as a form of servitude. In instances where this servitude is sufficient for heads to disengage, their decisions to do so rest on a combination of factors (Lacey and Gronn, 2007). In light of current concerns about headship shortages, however, two factors are especially worthy of comment. These are work intensification and risk.

In relation to Stewart's three-mold model of managerial and leadership work, intensification refers to an increase in the volume and complexity of the demands on leaders. There are three reasons why such an increase tends to be experienced as intensified: first, expansion of the demands that occupy a leader's role space is unlikely to be accompanied by the augmented resources required to complete them; second, the time frames within which these demands have to be accomplished are invariably short; and third, numerous colleagues with expertise relevant to the particular task demands have to be consulted. School heads compensate for the existence of role overload by working longer hours and by encroaching on times otherwise set aside for personal and family matters. This mode of response to greedy work demands has fueled concerns with work-life balance and with occupational well-being. Moreover, experienced heads are conscious of the potentially negative impact of their modeling of work routines on prospective leaders within their own schools (Gronn, 2003: 51–60). Work intensification also has serious consequences for the size and sustainability of headteacher roles. That is, the design of headship may need to be rethought with a view to reducing the level and scale of its demands; otherwise, the widespread negative impact of heads' intensified work may be compounded if and when teachers feel discouraged from becoming leaders. Risks may provide a further disincentive to becoming a leader. With the devolution of responsibilities to schools, headteachers find themselves enmeshed in a set of unresolved tensions concerning risk taking. Encouraged to take risks by apologists with high expectations of leadership, on the one hand, the reality for them, on the other, is that some risks may be more acceptable (politically and in other ways) than others. Their difficulty is in not knowing in advance which those risks are likely to be. With the possibility that actions or initiatives may go pear shaped, and with the onus on a headteacher in such circumstances to show cause, there is a strong incentive in an accountability culture which names, blames, and shames, for school heads to play safe. In short, with proactivity and risk aversion potentially at odds, the kinds of heightened expectations of leadership identified at the outset of this article may prove self-defeating.

Summary

This article has reviewed a range of recent developments which have redefined the work of headteachers. This redefinition process has been a core component of the modernizing of schooling and has occurred primarily as a result of recasting the work of headteachers as leaders in preference to an older, established discourse of managers or administrators. The focus was mainly on some key situational imperatives experienced by headteachers, along with a number of intended and unintended consequences that have emerged from this new work order of leadership. These outcomes have forcefully highlighted a gap between the rhetoric of plausibility invoked by its apologists to justify the reform and reconstruction of site-based activity and the reality of on-the-ground leadership practice experienced by role incumbents. The indications are that this gap is likely to persist into the foreseeable future.

See also: A Distributed Perspective On School Leadership and Management; Contexts for Leadership at the Beginning of the 21st Century; Leaders for Productive Schools; Leadership: Democratic; Leadership: School Improvement; Organizational Learning in Schools; Principals Role in Restructuring Schools; School Leadership Standards; School Reform and Restructuring: Self Managing School; Transformational School Leadership; Understanding How Leadership Influences Student Learning.

Bibliography

- Anderson, M., Gronn, P., Ingvarson, L., et al. (2007). Australia: Country background report – OECD improving school leadership activity. *A Report Prepared for the Australian Government, Department of Education, Science and Training*, 79p. Camberwell: ACER.
- Audit Scotland (2005). *Leadership Development: How Government Works*. Edinburgh: Audit Scotland.
- Department of Education, Employment and Training (2001). *Victorian Government Schools: Principal Class Handbook*. Melbourne: DEET.
- De Roche, C. P. (1994). On the edge of regionalization: Management style and the construction of conflict in organizational change. *Human Organization* 53(3), 209–219.
- Fondas, N. and Stewart, R. (1994). Enactment in managerial jobs: A role analysis. *Journal of Management Studies* 31(1), 83–103.
- Gronn, P. (2003). *The New Work of Educational Leaders: Changing Leadership Practice in an Era of School Reform*. London: Sage/Paul Chapman.
- Gronn, P. (2004). Distribution of leadership. In Goethals, G. R., Sorenson, G. J., and MacGregor, B. J. (eds.) *Encyclopedia of Leadership*, vol. 1, pp 351–355. Thousand Oaks, CA: Sage.
- Hayek, F. A. (1945). The use of knowledge in society. *American Economic Review* 35(4), 519–530.
- Hochschild, A. R. (1983). *The Managed Heart: Commercialization of Human Feeling*. Berkeley, CA: University of California Press.
- Howson, J. (2007). The state of the labour market for senior staff in schools in England and Wales, 2006–7. *13th Annual*

- Report Prepared for the NAHT and ASCL*, Oxford: Education Data Surveys.
- Ingvarson, L., Anderson, M., Gronn, P., and Jackson, A. (2006). *Standards for School Leadership: A Critical Review of Literature*. Melbourne: ACER/Teaching Australia.
- Kahn, W. A. (1992). To be fully there: Psychological presence at work. *Human Relations* **45**(4), 321–349.
- Lacey, K. and Gronn, P. (2007). *Letting Go: Former Principals Reflect on Their Role Exit*. Centre for Strategic Education, Seminar Series, No. 163. Jolimont: CSE.
- Scottish Executive Education Department (2007). *Improving School Leadership – OECD Background Report: Scotland*. Edinburgh: SEED.
- Scottish Government (2006). A Curriculum for Excellence – The Curriculum Review Group <http://www.scotland.gov.uk/Resource/Doc/26800/0023690.pdf> (accessed May 2009).
- Simon, H. A. (1976/1945). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*. New York: Free Press.
- Stewart, R. (1982). A model for understanding managerial jobs and behaviour. *Academy of Management Review* **7**, 7–13.
- Stewart, R. (2003). Woman in a man's world. *Leadership Quarterly* **14**, 197–204.
- Victoria, Department of Education and Training (2004). *The Privilege and the Price: A Study of Principal Class Workload and Its Impact on Health and Well-Being*. Melbourne, VIC: Human Resources Division, DE&T.

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

This page intentionally left blank

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

EDITORS-IN-CHIEF

PENELOPE PETERSON
School of Education and Social Policy
Northwestern University
Evanston, IL
USA

EVA BAKER
UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

BARRY MCGAW
McGaw Group Pty Ltd
Mont Albert Nth, VIC
Australia



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
The Boulevard Langford Lane, Kidlington, Oxford OX5 1GB, UK

Copyright © 2010 Elsevier Ltd. All rights reserved

LATVIA
Copyright © 2010 A Joma

STUDENT DATA SYSTEMS AND THEIR USE FOR EDUCATIONAL IMPROVEMENT
Copyright © 2010 J C Wayman, V Cho, and M P Richards

QUALIFICATIONS FRAMEWORKS AND THEIR ROLE IN THE REFORM OF EDUCATION AND TRAINING
Copyright © 2010 M Young

LANGUAGE OF VOCATIONAL EDUCATION AND TRAINING
Copyright © 2010 J Misko, J MacKenzie, and R-A Polvere

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44)(0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at (<http://elsevier.com/locate/permissions>), and selecting *Obtaining permissions to use Elsevier material*

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Catalog Number
A catalogue record for this book is available from the Library of Congress

ISBN: 978-0-08-044893-0

For information on all Elsevier publications
visit our website at books.elsevier.com

PRINTED AND BOUND IN THE UNITED KINGDOM
09 10 11 12 13 10 9 8 7 6 5 4 3 2 1

PREFACE

A preface usually provides a brief introduction to a work, intended to set the stage, provide some background insight, and whet the appetite of the reader. In our case, however, this preface has to address a fundamental question, one that was in our minds at the time we were recruited as Editors-in-Chief for the International Encyclopedia of Education. The question was “Why do we need an Encyclopedia? Its subtext was inspired by the ever-growing, ever-popular Internet. We believe that *this* Encyclopedia is desperately needed and will become a valued resource in education and associated social sciences and arts. The reasons are intellectual and procedural. Anyone with a modicum of knowledge knows that finding and trusting information gleaned from the Internet are two separate actions. The reliance on browsers to help discover references and comments result in resources based on popularity not quality. Pithy titles catch the eye and references rise in the ranks of browser searchers. Related to this is the “editing” in the Internet realm of populist efforts at encyclopedia, references, and other compilations. Once again, after removing offensive material, the accuracy, completeness, lack of bias, and other provenance for entries simply do not exist. Experienced researchers in education can sort through and make intelligent choices. Novices and many journeyman, or practitioners, parents, and policy makers cannot. Contrast how this Encyclopedia was built. Key domains of educational research were identified, and a tentative list of sub-domains or useful applied areas was posited. Then the Editors-in-Chief (apologies for the awkwardness of the term) identified the leading researcher in a particular domain, and with surprisingly little effort, recruited them to participate. They in turn identified the two best researchers in a sub-domain, such as formative assessment or the training of pre-school teachers. The authors of the sections of the Encyclopedia do not represent a collective group of friends and acquaintances, although friendships have been made. Rather they embody a deep and broad scholarly community. The difference from compiled Internet resources is the built-expertise and intellectual engagement of the authors. The summary of the developments and futures in their personal areas of scholarship have been filtered through their years of experience, both as scholars and communicators. Quality, then, is endemic to each piece, developed through this top-down identification of expertise, and made indelible by the bottom-up application of high standards from people leading the sub-domains – the authors, and the domains themselves, the section editors.

On a procedural level, the publishers early committed to the notion that this Encyclopedia would also be an online resource, and access would be available through print, for those with strong bookcases and the persisting love of turning real pages. The Internet version will allow multiple prisms through which the reader may access articles and provide, as it were, an emulation of the Internet in our field, albeit bounded by expertise and high quality.

What must be underscored in the assessment of this effort are the Editors-in-Chief and the publishers’ commitment to find excellence worldwide. We tried very hard to persuade notable scholars from all parts of the world to make contributions. Less than to fulfill the title of “International,” we were on the hunt for perspectives that would enrich the scope and depth of the sections. Our section editors put in enormous time attempting to find the best in the field, wherever they resided. Yet, not everyone is in the volume. Some were overcommitted. Many were not fully confident of their English, and the automated translation software has not yet met standards for technical writing. We believe that such writing and editing tools will make the outreach to an even broader International group of scholars possible in future revisions, or online updates. Furthermore, the birth of the World Educational Research Association (in 2009) will provide a better set of interlocking networks to find and evaluate scholarship from any place on the globe.

Finally, the scope of the effort must be acknowledged: 28 section editors, 926 articles were commissioned, drafted, reviewed, redrafted, edited, and put together in the space of four years. The publishers underwent some internal changes, and alterations in management. We as Editors-in-Chief, changed roles, moved, and also had to keep our own research and development enterprises afloat. Deadlines wobbled; authors dropped from view and had to be replaced.

Yet, at times frustrating as all development is, we find the final product exhilarating. We are enthusiastic not simply because it came into being at all, but because the collective light of the minds that wrote have left a bright resource for the future, one that will impact the way our colleagues understand and experience the educational knowledge, improvement, and impact in the future.

Eva Baker
Penelope Peterson

EDITORS

EDITORS-IN-CHIEF

Penelope Peterson

School of Education and Social Policy
Northwestern University
Evanston, IL
USA

Eva Baker

UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

Barry McGaw

McGaw Group Pty Ltd
Mont Albert Nth
Melbourne, VIC
Australia

SECTION EDITORS

Walter Allen

HEOC Division
University of California, Los Angeles
Los Angeles, CA
USA

Vibeke Grøver Aukrust

Institute for Educational Research
University of Oslo
Oslo
Norway

Dominic J Brewer

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

Kai-Ming Cheng

Faculty of Education
The University of Hong Kong
Hong Kong

Gregory Chung

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Penny Enslin

Department of Educational Studies
University of Glasgow
Glasgow
UK

Noelle Griffin

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Dennis Hocevar

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

David Hopkins

Institute of Education
University of London
London
UK

Mary James

Faculty of Education
University of Cambridge
Cambridge
UK

Sanna Järvelä

Department of Educational Sciences and
Teacher Education
University of Oulu
Oulu
Finland

Thomas Karmel

National Centre for Vocational Education Research
Adelaide, SA
Australia

Harold G Levine

School of Education
University of California, Davis
Davis, CA
USA

Robert L Linn

University of Colorado at Boulder
Ouray, CO
USA

Allan Luke

Faculty of Education
Queensland University of Technology
Brisbane, QLD
Australia

Ann M Mastergeorge

Human Development and Family Studies
University of California, Davis
Davis, CA
USA

Patrick J McEwan

Department of Economics
Wellesley College
Wellesley, MA
USA

Barbara Means

Center for Technology in Learning
SRI International
Menlo Park, CA
USA

Ingrid Moses

University of Canberra
Canberra, ACT
Australia

Bill Mulford

Faculty of Education
University of Tasmania
Hobart, TAS
Australia

Yu Jing Ni

Faculty of Education
The Chinese University of Hong Kong
Hong Kong

Festus Obiakor

Department of Exceptional Education
University of Wisconsin-Milwaukee
Milwaukee, WI
USA

P David Pearson

Graduate School of Education
University of California, Berkeley
Berkeley, CA
USA

Colin Power

Graduate School of Education
University of Queensland
Brisbane, QLD
Australia

Kjell Rubenson

Faculty of Education
The University of British Columbia
Vancouver, BC
Canada

Sandip Sinharay

Education Testing Services
Princeton, NJ
USA

Collette Tayler

Queensland University of Technology
Brisbane, QLD
Australia

Nico Verloop

ICLON, Graduate School of Education
Leiden University
Leiden
Netherlands

HOW TO USE THE ENCYCLOPEDIA

The International Encyclopedia of Education is intended for use by students, research professionals, and interested others. Articles have been chosen to reflect major disciplines in the study of education and common topics of research by academics in this domain. Each article serves as a comprehensive overview of a given area, providing both breadth of coverage for students, and depth of coverage for research professionals. We have designed the encyclopedia with the following features for maximum accessibility for all readers.

The contents of the encyclopedia are arranged alphabetically by section, and within sections, alphabetically by article. The Subject Index is located in Volume 8. Some topics are covered in a multitude of articles from differing perspectives, while other topics may have only one entry. We encourage use of the index for access to a subject area, rather than use of the Contents list alone, so that a reader has a full notion of the coverage of that topic.

The articles include cross-references to other related encyclopedia articles, suggested further readings where applicable, and many contain relevant websites for additional information. We encourage readers to use the cross-references to locate other encyclopedia articles that will provide more detailed information about a subject.

The Further Reading sections include recent secondary sources to aid the reader in locating more detailed or technical information. Review articles and research articles that are considered of primary importance to the understanding of a given subject area are also listed. These suggested further readings are not intended to provide a full reference listing of all material covered in the context of a given article, but are provided as next steps for a reader looking for additional information.

This page intentionally left blank

CONTENTS

Preface	v–vi
Editors	vii–viii
How to use the Encyclopedia	ix

VOLUME 1

A

Adult Education

Adult Education Overview	<i>K Rubenson</i>	1
--------------------------	-------------------	---

Adult Education – Adult Learning, Instruction and Program Planning

Adult Learning	<i>S B Merriam</i>	12
Adult Learning and Instruction: Transformative Learning Perspectives	<i>P Cranton</i>	18
Adult Learning in a Biographic Perspective	<i>L West</i>	25
Adult Learning, Instruction and Programme Planning: Insights from Freire	<i>P Mayo</i>	31
Characteristics of Adult Learning	<i>K Illeris</i>	36
Informal Learning: A Contested Concept	<i>P Hodgkinson</i>	42
Organizational Learning	<i>P-E Ellström</i>	47
Program Planning	<i>A L Wilson and R M Cervero</i>	53
Workplace Learning Frameworks	<i>S Billett</i>	58

Adult Education – as a Field of Study

Class Analysis in Adult Education	<i>T Nesbit</i>	64
Gender Analysis	<i>J Stalker</i>	70
Race and Ethnicity in the Field of Adult Education	<i>J Johnson-Bailey and D Drake-Clark</i>	77
Trends in Workplace Learning Research	<i>T Fenwick</i>	83

Adult Education – Concepts

Lifelong Learning	<i>J Field</i>	89
Rewriting the History of Adult Education: The Search for Narrative Structures	<i>B J Hake</i>	96

Adult Education – Domains and Provision

Adult Basic Education: A Challenge for Vocational Based Learning	<i>R Höghelm</i>	102
Adult Literacy Education	<i>L Tett and R St.Clair</i>	107
Citizenship and Immigrant Education	<i>D Schugurensky</i>	113
Community Based Adult Education	<i>B Connolly</i>	120
Continuing Professional Education: Multiple Stakeholders and Agendas	<i>R M Cervero and B J Daley</i>	127
Financing of Adult and Lifelong Learning	<i>Å Sohlman</i>	133
Labor Education	<i>K Forrester</i>	139
Learning Cities and Regions	<i>C Duke</i>	144
Museums as Sites of Adult Learning	<i>R S Grenier</i>	150
Overview of Lifelong Learning Policies and Systems	<i>D-B Kwon and C H O Daeyeon</i>	156
Popular Adult Education	<i>S Tøsse</i>	162
Provision of Prior Learning Assessment	<i>P Andersson</i>	169
The Age of Learning: Seniors Learning	<i>P Jarvis</i>	175
University Adult Continuing Education: The Extra-Mural Tradition Revisited	<i>M Slowey</i>	183

Adult Education – Economy and Society

Adult Education and Civil Society	<i>J Crowther and I Martin</i>	191
Adult Education and Nation-Building	<i>S Walters</i>	198
Economic Outcomes of Adult Education and Training	<i>A Ferrer and W C Riddell</i>	204
Health and Adult Learning	<i>P Bélanger and M Robitaille</i>	211
Modernization Processes and the Changing Function of Adult Learning	<i>H S Olesen</i>	217
The Political Economy of Adult Education	<i>G Rees</i>	223
Wider Benefits of Adult Education	<i>R Desjardins and T Schuller</i>	229

Adult Education – Participation

Barriers to Participation in Adult Education	<i>K Rubenson</i>	234
Participation in Adult Learning	<i>R Desjardins</i>	240

C**Comparative Education**

Comparative and International Education: English-Language Scholarly Publications	<i>D Post and M Fedri</i>	251
Comparative Education: Philosophical Issues and Concepts	<i>M Mason and C W Evers</i>	257
Comparative Education: Societies and Associations	<i>M Bray</i>	266

Curriculum Development

Curriculum and Religion	<i>C Kapitzke and P Graham</i>	272
Gender and Curriculum	<i>J Marsh</i>	277
Local Teacher-Based Curriculum Development	<i>M D Osborne and H B Barker</i>	283

Curriculum Development – Contexts

Curriculum and Human Rights	<i>C Lohrenscheit</i>	287
Curriculum and the Education of Cultural and Linguistic Minorities	<i>S May</i>	293
Curriculum in Postcolonial Contexts	<i>A Hickling-Hudson</i>	299
Curriculum, Economic and Cultural Development	<i>A Rampal</i>	307
Globalization and Curriculum	<i>R Dale</i>	312
Popular Culture	<i>K Schultz and R Throop</i>	318

Curriculum Development – Evaluation and Research

An Overview of Research in Curriculum Inquiry	<i>F M Connelly and S J Xu</i>	324
Curriculum Evaluation: Approaches and Methodologies	<i>V Klenowski</i>	335
Curriculum Studies, Discourse Analysis, and the Construction of Historical Time	<i>J Laspina</i>	342
Curriculum Use in the Classroom	<i>M Ben-Peretz and B Eilam</i>	348
The Roots and Routes of Teacher-Based Action Research and Curriculum Inquiry: An Historical Perspective	<i>L G Putney and J L Green</i>	355

Curriculum Development – Planning and Development

Curriculum and Syllabus Design	<i>A Woods, A Luke, and K Weir</i>	362
Curriculum and the Publishing Industry	<i>J Larson, A-R Allen, and D Osborn</i>	368
Curriculum and Teacher Change	<i>J G Ladwig</i>	374
Curriculum Governance and Planning	<i>B Levin</i>	379
Curriculum Planning and Systems Change	<i>Z Deng</i>	384
Curriculum, Digital Resources and Delivery	<i>Y Zhao, G Zhang, and C Lai</i>	390
Textbook Development and Selection	<i>P Shannon</i>	397

Curriculum Development – Subject Matter

Curriculum and Globalization: Higher Education	<i>K H Mok</i>	403
Curriculum Development as Subject Matter: Social Studies	<i>J Brophy and J Alleman</i>	411
Curriculum Development in the Area of Reading	<i>C L Madda and T E Raphael</i>	415
Curriculum in the Arts	<i>C M Thompson, L Bresler, and T E Costantino</i>	423
Curriculum: Foreign Language Learning	<i>Angel M Y Lin</i>	428
First Language Learning	<i>T Rogers and L Moffatt</i>	435
Learning a Second Language in First Language Environments	<i>E Bernhardt and D J Tedick</i>	441
Literature	<i>S Hynds</i>	447
Mathematics	<i>B Kaur and C P Vistro-Yu</i>	453
Physical Education and Sports	<i>D Kirk</i>	459
Primary and Elementary/Middle Grades Reading	<i>D E Alvermann and C A Mallozzi</i>	464
Science	<i>Y Kali and M Linn</i>	468

Curriculum Development – Theory

Bilingual Learning (Learning L1 and L2 in an L1 and L2 Environment)	<i>D Schwinge</i>	475
Curriculum and Complex Systems Theory	<i>W-M Roth and J Thom</i>	481

Curriculum and Constructivism	<i>B Davis and D Sumara</i>	488
Curriculum and Critical Theory	<i>L Yates</i>	494
Curriculum and Poststructuralist Theory	<i>J L Miller</i>	499
Curriculum and Structuralist Sociology: The Theory of Codes and Knowledge Structures	<i>J Muller and J Gamble</i>	505
Curriculum Reform	<i>R Gilbert</i>	510
Early Childhood Curriculum and Developmental Theory	<i>S M-Y Lim and C Genishi</i>	514

D

Demography and Social Change

An Evaluation Framework for E-Learning Effectiveness in the Arab World	<i>W AbuSneineh and M Zairi</i>	521
Globalization and Social Justice in Higher Education	<i>E Kimura-Walsh</i>	536
Racial Inequality and the Economic Payoffs of Higher Education	<i>S A Lachtman</i>	542
School Health Education	<i>J S Yang</i>	547

Demography and Social Change – Demography

A Paradigm of Contradictions: Racism and Science Education	<i>B A Brown and J W Mutegi</i>	554
Africana Studies: Past, Present And Future	<i>B R Hare, A Young, J B Stewart, and D P Aldridge</i>	565
Beyond Black Nationalism to Black Internationalism: The Compelling Case for Examining Black Educational Challenges Globally	<i>K Freeman</i>	573
Beyond Racial, Ethnic, and Gender Bias in Education Statistics	<i>J T Lott and C Bennett</i>	581
Caste and Education in India	<i>G G Wankhede</i>	589
Demography, Community, and the Education of Asian American and Pacific Islander Students	<i>R T Teranishi</i>	595
Educational Attainment and Mortality Differentials	<i>H D Horton, L Hixson, and S M Monnat</i>	602
Educational Opportunity and Latino/Chicano College Choice: New Findings, and Theoretical Perspectives	<i>M A Ceja and P A Pérez</i>	610
Ethnic Minority Identity and Educational Outcomes in a Rising China	<i>G A Postiglione</i>	616
Ethnic Studies and Women's Studies: From the Past to the Present, Providing a Space in the Academy for Scholarship Focused on the Histories and Experiences of Marginalized People	<i>C A Brown</i>	623
Gender Equity in Higher Education: Challenges and Celebrations	<i>L Morley</i>	629
Locating Space and Place in the College Access Debate: New Tools for Mapping and Understanding Educational Inequity and Stratification	<i>B Cantwell and J F Milem</i>	636
Public Policy and Inequality in Postsecondary Opportunity: Educational Statistics and the Failure of Education Reform	<i>E P St. John, K Williams, and K Moronski</i>	649
The Education of Indigenous Students	<i>J W Tippeconnic and S C Faircloth</i>	661
The Health Advantages of Educational Attainment	<i>G Y Dinwiddie</i>	667
To Go or Not to Go: Access and Barriers to Ethnic Minority Participation in Higher Education within the United Kingdom	<i>P Brug and T L Oliver</i>	673
Universal Secondary Education and the Two-Tiered School System in the English-Speaking Caribbean	<i>P Anderson and J Devonish</i>	679
Women of Color in Science, Technology, Engineering and Mathematics (STEM): Refining the Concepts, Reframing the Issues	<i>C B Leggon</i>	686

Demography and Social Change – Social Change

Educational Diversity in U.S. Law Schools	<i>C E Daye, A T Panter, W R Allen, M E Deo, and L F Wightman</i>	691
Educational Reform	<i>F Lopez-Segrera</i>	698
Emergence of For-profit Higher Education	<i>X Li</i>	705
Engaged Education: Experiential Learning, Intensive Field Experiences, and Social Change	<i>T R Rone</i>	711
Higher Education and Social Change in Mexico	<i>A Mungaray, G Estrella, P Moctezuma, and M Ramírez-Urquidy</i>	719
The Changing Role of the State in Higher Education	<i>K Gürüz</i>	728
The Role of the State in School Reform: Responding to Increasing Ethnic Diversity in Britain	<i>P Ratcliffe</i>	737
Transforming Mindsets Through Education for Sustainable Development	<i>V M Segovia</i>	746

Demography and Social Change – Sociology

Achievement Motivation in Ethnic Minority Youth	<i>S Graham and E Bryant</i>	753
Affirmative Action and Higher Education in Brazil	<i>J T dos Santos and D M Queiroz</i>	760
Black in White: Black Students at White and Black Colleges	<i>L Evans and J R Feagin</i>	767
Bologna Process: On the way to a Common European Higher Education Area	<i>A V Rezaev</i>	772
Critical Race Theory in Education Research	<i>R Vue and C B Newman</i>	779
Cultural Diversification and Japanese Education: Social Constructions of the New Diversity	<i>R Tsuneyoshi</i>	787
Education in War Countries: A Disempowerment of Society	<i>J L Nkata</i>	793
Educational Attainments of U.S. Black Males and Females: 1971 to 2003	<i>F B Bonner, M-C Jipguep-Akhtar, and R J Harrison</i>	799
Educational Preparation: Fostering the Self-Efficacy and Resilience of Urban Adolescent Youth	<i>M B Spencer and B Tinsley</i>	806
HIV Education for Low-Literate People: Transforming Students and Communities through Paulo Freire's Praxis and the Pedagogy of Action	<i>N Z Haniff</i>	814
How Are Discussions of Interdisciplinary Studies Linked to Diversity Discourse?	<i>C N Spence</i>	823
Identifying Patterns of Doctoral Attrition Across Academic Fields	<i>A James, K J Fosnacht, E B Stolzenberg, and D Thirumalai</i>	827
Male Underachievement in Education Across the Globe: A Shift in Paradigm for Gender Disparities Regarding Academic Achievement	<i>J F L Jackson, J L Moore III, and R A Leon</i>	838
Noncognitive Measures for Higher Education Admissions	<i>W E Sedlacek</i>	845
Peace Education	<i>J Page</i>	850
Remote Control: Africa's Development and Outflow to Developed Nations of Technical Talent	<i>F Haniff</i>	855
Scholar-Baller®: Student Athlete Socialization, Motivation and Academic Performance in American Society	<i>C K Harrison, S Bukstein, J Mottley, E Comeaux, J Boyd, C Parks, and D Heikkinen</i>	860
Service-Learning and Diversity as Mediums to Foster Civic and Community Engagement in Undergraduates	<i>L Oseguera</i>	866
Success and Failure in Education and Criminal Justice: Identifying Common Mechanisms	<i>S Ramirez and A Harris</i>	874
The History of Education: Race and Education	<i>J O Jewell</i>	881

VOLUME 2**E****Early Childhood Care and Education Theory**

- Early Childhood Care and Education: The Family, The Market, and The State *S B Kamerman and M J Neuman* 1
- Gender Issues in Early Childhood Education and Care *B Pocock* 11

Early Childhood Education and Care Governance

- Early Childhood Risk, Protection and Abuse Prevention *E Khoo* 17
- Evaluating Early Childhood Education and Care Programs *M Janus and S Brinkman* 25

Early Childhood Education and Care Investment

- Investing in Early Childhood Education and Care: The Health and Wellbeing Case *T Moore and F Oberklaid* 32
- Cost-effective Early Childhood Programs from Preschool to Third Grade *A J Reynolds, J A Temple, and B White* 38
- Early Childhood in Post-Modern Cultures: Thoughts and Some Concerns *P Gammage* 49
- Investing in Early Childhood Education and Care: Some Policy Implications *J Bennett* 55
- Investing in Early Childhood Education and Care: The Economic Case *G Cleveland and M Krashinsky* 63

Early Childhood Education and Care Programs

- Early Care and Education Programs for Infants and Toddlers *B Nyland* 69
- Literacies in Early Childhood: The Preschool Period *M Flear and B Raban* 75
- Participation in Early Childhood Education and Care Programs: Equity, Diversity and Educational Disadvantage *M Vandenbroeck* 81
- Teaching in Early Childhood Centers Instructional Methods and Child Outcomes *I Siraj-Blatchford* 86

Early Childhood Education and Care Provision

- Early Childhood Education and Care in the People's Republic of China *Zeng Xiaodong* 93
- Home-Based and Institutional Early-Childhood Education and Care Services *R Fukkink and H Blok* 98
- Integrated Early Childhood Education and Care Services – Care, Upbringing, Education and Health *S Colley* 104
- Issues of Access and Program Quality *J T Hustedt and W S Barnett* 110
- Kindergarten Transitions and Linkages to Primary School – Readiness Re-conceptualized *A Petriwskyj* 120
- Parent Support in Early Childhood – Approaches and Outcomes *K S Howard, K A Beckmann, and J Brooks-Gunn* 126

Early Childhood Education and Care Research

- Development and Implementation of Early Learning Standards in the United States *C Scott-Little* 132
- Global Status of Early Learning and Development Standards *P R Britto and S L Kagan* 138
- Rethinking Early Childhood Education and Care: Implications for Research and Evaluation *K Thorpe, D Cloney, and C Tayler* 144

Early Childhood Education and Care Theory

- Child Rearing and Early Education: Parents and Professionals: Theoretical and Historical Influences *H Penn* 151

Early Childhood Education and Care Workforce

- Early Childhood Education and Care Professional Profiles in the United Kingdom: Multiagency Teamwork and Transdisciplinary Challenges *A Anning and P Calder* 157
- Early Childhood Education and Care Staff Support: Mentoring and Professional Development in Finland *P Lindberg* 162

Economics of Education

- Education and Health *P Muennig* 169
- Data *M Perez and M Socias* 178
- Empirical Research Methods in the Economics of Education *P J McEwan* 187
- Theoretical Concepts in the Economics of Education *D J Brewer, G C Hentschke, and E R Eide* 193

Economics of Education – Costs and Financing of Schooling

- Cost–Benefit Analysis and Cost–Effectiveness Analysis *C Belfield and H M Levin* 199
- Economic Approaches to Adequacy *J Imazeki* 204
- Economic Approaches to School Efficiency *L L Taylor* 210
- School Finance Reform *T Downes* 216
- School Finance: An Overview *J K Rice, D H Monk, and J Zhang* 222
- The Economics of Tuition and Fees in Higher Education *R G Ehrenberg* 229

Economics of Education – External Benefits of Schooling

- Education and Civic Engagement *T S Dee* 235
- Education and Crime *L Lochner* 239
- Education and Economic Growth *E A Hanushek and L Wößmann* 245
- Parental Socioeconomic Status, Child Health, and Human Capital *J Currie and J Goodman* 253
- The External Benefits of Education *W W McMahon* 260

Economics of Education – Labor Markets, Education and Earnings

- Education and Inequality *J Blanden and S Machin* 272
- Human Capital *E R Eide and M H Showalter* 282
- Race Earnings Differentials *M Carnoy* 288
- Returns to Education in Developed Countries *M Gunderson and P Oreopoulos* 298
- Returns to Education in Developing Countries *H A Patrinos and G Psacharopoulos* 305
- School Quality and Earnings *J R Betts* 313
- Signaling in the Labor Market *M E Page* 321
- The Economics of High School Dropouts *R W Rumberger* 325

Economics of Education – Markets, Incentives and School Choice

- Competition and Student Performance *A A Payne* 331
- Educational Privatization *C Belfield and H M Levin* 337

Student Incentives	<i>E R Eide and M H Showalter</i>	342
The Economic Role of the State in Education	<i>D N Plank and T E Davis</i>	347
The Economics of Catholic Schools	<i>W Sander and D Cohen-Zada</i>	354
The Economics of Charter Schools	<i>R Zimmer and R Buddin</i>	360
The Economics of Parental Choice	<i>R Chakrabarti and J Roy</i>	367
The Economics of School Accountability	<i>D N Figlio and H F Ladd</i>	374
The Efficacy of Educational Vouchers	<i>R Zimmer and E P Bettinger</i>	380
Tiebout Sorting and Competition	<i>T J Nechyba</i>	388

Economics of Education – Production of Schooling

Desegregation, Academic Achievement and Earnings	<i>S Rivkin</i>	394
Education Production Functions: Concepts	<i>D N Harris</i>	402
Education Production Functions: Developed Country Evidence	<i>E A Hanushek</i>	407
Education Production Functions: Evidence from Developing Countries	<i>P Glewwe and S Lambert</i>	412
Family Environment in the Production of Schooling	<i>R Rothstein</i>	423
Neighborhoods and Peers in the Production of Schooling	<i>J Vigdor and J Ludwig</i>	431
Teacher Quality in Education Production	<i>I Pelayo and D J Brewer</i>	438
The Economics of Class Size	<i>D W Schanzenbach</i>	443
The Economics of Early Childhood Interventions	<i>M Nores</i>	450

Economics of Education – Teaching Labor Markets

An Overview of Teacher Labor Markets	<i>H Lankford and J Wyckoff</i>	456
Compensating Differentials in Teacher Labor Markets	<i>J Chambers</i>	465
Economic Approaches to Teacher Recruitment and Retention	<i>S Loeb and J Myung</i>	473
Teacher Incentives	<i>L Santibañez</i>	481
Teacher Supply	<i>P Dolton</i>	489
Teacher Training and Preparation in the United States	<i>M Podgursky, M G Springer, and R Hutton</i>	498
Teachers in Developing Countries	<i>F H Rogers and E Vegas</i>	504
The Economics of Teachers' Unions in the United States	<i>K Strunk</i>	512

Education of Children with Special Needs

Adapted Physical Activity for Students with Special Needs	<i>S R Hodge</i>	519
Assessment Accommodations for Children with Special Needs	<i>R J Kettler and S N Elliott</i>	530
Assistive Technology and Educational Practice	<i>H P Parette and G R Peterson-Karlan</i>	537
Augmentative–Alternative Communication	<i>C Mason and K M Chinn</i>	544
Community Integration and Employment of Youth with Special Needs	<i>B A Ford and A W Daviso III</i>	550
Cooperative Learning for Children with Special Needs	<i>E A Dooley and K R Kossar</i>	555
Co-Teaching	<i>E D Cramer</i>	560
Cultural and Linguistic Diversity and Children with Special Needs	<i>J M Patton, N Day-Vines, and A McSwain</i>	565
Curriculum-Based Assessment and Students with Special Needs	<i>D J O'Shea and J D Katsafanas</i>	571

Diagnosis and Classification	<i>M Sapp</i>	577
Direct Instruction and the Education of Children with Special Needs	<i>T Mehring</i>	584
Disabilities and Gifted Education	<i>V I Daniels and M J McCollin</i>	588
Early Childhood Special Education: Birth to Eight Years Old	<i>T S Lovelace, R Gardner III, J B Axe, and D M Sainato</i>	594
Early Intervention for Students with Special Needs	<i>P A Grant</i>	608
Early Transition of Children with Special Needs	<i>S A Fowler</i>	613
Educating Children Who are Deaf-Blind	<i>J Wood and K M Chinn</i>	621
Educating Children with Deafness and Hearing Impairments	<i>G M Mukuria and C J Eleweke</i>	628
Educating Children with Other Health Impairments (ADHD)	<i>G M Mukuria and F E Obiakor</i>	634
Educating Children with Physical Disabilities	<i>Z P Nkabinde, F E Obiakor, M T Ofor, and D J Smith</i>	640
Educating Children with Specific Learning Disabilities	<i>P A Grant and P B Grant</i>	646
Educating Children with Speech and Language Disorders	<i>S L Green and C D Qualls</i>	654
Educating Students with Autism and Related Disorders	<i>T T Dyches</i>	661
Educating Students with Cognitive Disabilities	<i>J P Bakken</i>	669
Educating Students with Emotional and Behavioral Disorders	<i>S W Smith and G G Taylor</i>	678
Educating Students with Special Needs: An Overview	<i>F E Obiakor, M K Harris, P Harris-Obiakor, and G M Mukuria</i>	688
Educating Students with Traumatic Brain Injury	<i>S Burkhardt and A Rotatori</i>	695
English Language Learners with Special Needs	<i>R Rueda and G Ragusa</i>	701
Functional Behavioral Assessment	<i>J P Bakken</i>	708
History of Special Education	<i>L M Garrick Duhaney and S J Salend</i>	714
Inclusion of Students with Special Needs in General Education Classrooms	<i>M A Prater</i>	721
Instructional Accommodations for Children with Special Needs in Inclusive Settings	<i>A L Yurick, S Musti-Rao, and G Cartledge</i>	727
Interagency Agreements	<i>L A Owens</i>	733
Interdisciplinary Collaboration in Early Childhood	<i>N File</i>	739
Language Development in Children with Special Needs	<i>M S L Stewart</i>	745
Large Scale Assessment and Accountability and Students with Special Needs	<i>M L Thurlow</i>	752
Literacy Instruction for Students with Special Needs	<i>L H Mason</i>	759
Math Instruction for Children with Special Needs	<i>B A Bottge</i>	767
Overrepresentation of Culturally and Linguistically Diverse Students in Special Education	<i>S C Trent</i>	774
Paraprofessionals in Special Education	<i>B Y Ashbaker, S C Dunn, and J Morgan</i>	780
Parent and Family Involvement in the Education of Children with Special Needs	<i>L P Prater</i>	788
Peer Relations and Socialization of Children and Adolescents with Special Needs	<i>L M Bullock, M Wong-Lo, and R A Gable</i>	794
Peer-Tutoring	<i>L J O'Shea and D J O'Shea</i>	802
Postsecondary Participation of Students with Special Needs	<i>R A Stodden and K D Roberts</i>	808
Prereferral Strategies	<i>C G Simpson and J P Bakken</i>	815

Preventing Antisocial Behavior and Delinquency: A Comprehensive School-wide Approach <i>P R Hallam, K R Young, P Caldarella, D G Wall, and L Christensen</i>	820
Programs and Instructional Strategies for Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	829
Related Services for Children with Special Needs <i>S O Obi</i>	834
Risk and Resilience Frameworks in Understanding Special Education <i>W-Y Hsieh and M L Donahue</i>	840
School Dropout Prevention <i>F E Obiakor</i>	845
School-Based Services for Children with Special Needs <i>S O Obi</i>	850
Self-Determination <i>M L Wehmeyer</i>	857
Social Competence <i>L A Owens and S Johnston-Rodriguez</i>	865
The Identification of Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	870
Transition from School to Adult Life <i>M J McCollin and F E Obiakor</i>	876
Visual Impairment <i>W Sapp</i>	880
Workforce Issues in Special Education <i>E Carlson and B Billingsley</i>	886

VOLUME 3

Education of Professionals

An Overview of Accreditation, Certification, and Licensure Processes <i>J J Norcini and M van Zanten</i>	1
Educating America's Military Officers <i>M H Adamshick</i>	7
Education for Architecture in the United States and Canada <i>S Schreiber</i>	13
Education for Medicine <i>M B Anderson</i>	19
Education for the Legal Profession <i>J W Wegner</i>	25
Education of Clergy <i>D O Aleshire</i>	31
Framing Professions Education <i>P S O'Sullivan</i>	37
Issues in Accounting/Business Education <i>W H Gijsselaers and R G Milter</i>	39
Professions – Engineering <i>J Heywood</i>	45
Professions – Health-Related <i>G M Jensen and C B Royeen</i>	51
Professions – Pharmacy <i>D E Beck</i>	58
Teaching for the Complex Practice of Nursing <i>P Benner and M Sutphen</i>	65
Veterinary Education <i>K M Edmondson</i>	71

Education Research Methodology: Quantitative Methods and Research

Abductive Research Methods <i>B D Haig</i>	77
Field Experimentation in Education <i>M A Constan</i>	83
Internal Validity <i>M H Clark and S C Middleton</i>	90
Methods for Approximating Random Assignment <i>B Schneider and S-K McDonald</i>	97
Methods for Approximating Random Assignment: Regression Discontinuity and Propensity Scores <i>L M Scheier</i>	104

Multitrait–Multimethod Designs	<i>M Bong and S Hong</i>	111
Multivariate Longitudinal Data Analysis	<i>H Liu and L Chang</i>	120
Quasi-Experimentation: Two Group Design	<i>E Hong</i>	128
Regression Discontinuity Designs	<i>A Kelava, S Rohrmann, and V Hodapp</i>	134
Sampling	<i>R S Brown</i>	142
Statistical Conclusion Validity	<i>P L Busk</i>	147
Survey Research Methods	<i>A Fink</i>	152

Educational Assessment

An Overview of Educational Assessment	<i>M James</i>	161
Assessment and the Evaluation of Institutional Effectiveness	<i>S M Thomas</i>	172
Assessment in Schools – Dispositions	<i>R D Crick</i>	181
Assessment in Schools Related To Literacy: Reading	<i>K J Topping</i>	189
Impact of Assessment on Students' Learning Strategies and Implications for Judging Assessment Quality	<i>M Segers, F Dochy, and D Gijbels</i>	196

Educational Assessment – Assessment Activities

Alternative Assessment	<i>G Stobart and C Gipps</i>	202
Classroom Assessment Tasks and Tests	<i>J Phelan and J Phelan</i>	209
Dynamic Assessment	<i>J G Elliott, E L Grigorenko, and W C M Resing</i>	220
Instructional System Provided Feedback	<i>H F O'Neil, S Chuang, and T Huang</i>	226
Peer and Self-assessment	<i>C Harrison</i>	231
Portfolio Assessment	<i>V Klenowski</i>	236
Records of Achievement: Beyond Traditional Tests	<i>P Broadfoot</i>	243

Educational Assessment – Assessment in Domains

Assessment in Higher Education	<i>D R Sadler</i>	249
Assessment in Schools – Affective Domain	<i>M Ventura, R D Roberts, and M Zeidner</i>	256
Assessment in Schools – Creative Subjects	<i>T E Costantino and L Bresler</i>	262
Assessment in Schools – Learning to Learn	<i>A Hautamäki, J Hautamäki, and S Kupiainen</i>	268
Assessment in Schools – Literacy Writing (Extended)	<i>J M Parr and L Earl</i>	273
Assessment in Schools – Mathematics	<i>M Brown and J Hodgen</i>	279
Assessment in Schools – Oracy	<i>A L Bailey</i>	285
Assessment in Schools – Primary Science	<i>W Harlen</i>	293
Assessment in Schools – Secondary Science	<i>J Shemwell, A Fu, M Figueroa, R Davis, and R Shavelson</i>	300
Assessment in Schools – Technology Education and ICT	<i>A Jones, B Cowie, and J Moreland</i>	311
Assessment in the Early Years	<i>M J Drummond</i>	316
Assessment in the History, Civics and Social Studies Domains	<i>P G Avery</i>	323
Assessment in the Workplace of Performance, Developing Expertise and Competence	<i>A J Brown</i>	330
Assessment in Vocational Education	<i>K Ecclestone</i>	337

Educational Assessment – Concepts and Issues

Assessing Group Work	<i>M Galton</i>	342
Assessment and the Regulation of Learning	<i>L Allal</i>	348
Challenges of Developing and Implementing Formative Assessment Practices in Schools	<i>R Sutton</i>	353
Formative Assessment	<i>P Black</i>	359
Formative Assessment and Instructional Planning	<i>B Marshall</i>	365
Progression and Assessment: Developmental Assessment	<i>M Forster and G Masters</i>	369
Students' Rights in Assessment and Evaluation	<i>J Gardner and B Holmes</i>	378
Summative Assessment by Teachers	<i>R Daugherty</i>	384
The Multiple Purposes of Assessment	<i>P E Newton</i>	392
The Relationship between Assessment and the Organization and Practice of Teaching	<i>L B Resnick, D William, R Apodaca, and E S Rangel</i>	397

Educational Assessment – Enhancing Quality and Use

Assessment Conversations: Reading and Writing Conferences with Students and with Parents	<i>M Barrs</i>	403
Assessment: Pre-service and In-service Teacher Education	<i>A Davies</i>	410
Classroom Assessment in Policy Context (Australia)	<i>J J Cumming</i>	417
Classroom Assessment in Policy Context (England and France)	<i>M Osborn and M Raveaud</i>	425
Classroom Assessment in Policy Context (French Sources)	<i>D Laveault</i>	432
Classroom Assessment in Policy Context (Hong Kong)	<i>D Carless</i>	438
Classroom Assessment in Policy Context (New Zealand)	<i>T J Crooks</i>	443
Classroom Assessment in Policy Context (Sub-Saharan Africa)	<i>S J Howie, L Zimmerman, and K Draper</i>	449
Moderation of Student Work by Teachers	<i>G S Maxwell</i>	457
School Policies and Practices to Support Effective Classroom Assessment for Learning	<i>D Pedder</i>	464
Student Assessment: Policy and Practice in Eastern Europe	<i>G Bethell</i>	472

Educational Assessment – Key Relationships

Assessment Practice in Policy Context: Latin American Countries	<i>F M Rizo</i>	479
Impact of Assessment on Learner Groups (Boys/Girls)	<i>J Elwood</i>	486
Impact of Assessment on Learner Groups (Disabilities)	<i>M Rouse</i>	492
Impact of Assessment on Students' Test Anxiety	<i>P Buchwald and C Schwarzer</i>	498
Impact of Assessments on Classroom Practice	<i>J L Herman</i>	506
Social Practices in School Assessment and their Impact on Learner Identities	<i>A Pollard and A Filer</i>	512

Educational Evaluation – Concepts, Practice, and Future Directions

Educational Evaluation: Concepts, Practice, and Future Directions	<i>Y J Ni</i>	518
---	---------------	-----

Educational Evaluation – Contemporary Themes in Evaluation

Evaluating E-Learning	<i>D D Williams and C R Graham</i>	530
Evaluating Schools as Learning Communities	<i>S M Thomas</i>	539
Evaluation of Adult Education and Training Programs	<i>G McNamara, P Joyce, and J O'Hara</i>	548
Evaluation of Integrated Health Programs in School	<i>T M Franke</i>	555

Formative Assessment in Teacher Education and Teacher Professional Development	<i>H Tillema</i>	563
Informal Education and Evaluation	<i>T Jeffs</i>	572
Methods to Evaluate Technology	<i>W L Bewley</i>	578

Educational Evaluation – Contexts of Educational Evaluation

Cost Analysis in Evaluation Studies	<i>J D Fletcher</i>	585
Cultural Issues that Can Affect the Validity of Educational Evaluations	<i>R P Durán</i>	592
Evaluation and Accountability	<i>E Sato and S N Rabinowitz</i>	600
Evaluation Governance and Planning	<i>R Picciotto</i>	607
Evaluation Use	<i>V G Thomas</i>	614
Moral and Ethical Issues in Evaluation	<i>M Morris</i>	622
The Role of Stakeholders in Educational Evaluation	<i>S Taut and M C Alkin</i>	629

Educational Evaluation – Evaluation Domains

Curriculum Evaluation	<i>T Levin</i>	636
Evaluating Education in Three Policy Eras	<i>M A Peters, R Britez, and S M Weber</i>	645
Evaluation of Mathematics Education Programs	<i>J Cai</i>	653
Evaluation of Science Education Programs	<i>G Berberoğlu</i>	660
Evaluation of Social Studies and Civic Education Programs	<i>S G Grant</i>	667
Evaluation of Teacher Quality and Practice	<i>W Robinson and J Campbell</i>	674
International Evaluations	<i>A Schleicher</i>	681
National Assessment Programs: The Example of the U.S. National Assessment of Educational Progress	<i>M Schneider</i>	690
Principal Evaluation: Concepts, Needs and Realities	<i>N S Glasman and L D Glasman</i>	698
Program Evaluation	<i>C A Christie and L A Fierro</i>	706
School Based Evaluation: Purposes, Protocols and Processes	<i>J MacBeath</i>	713
School Inspection/External School Evaluation	<i>C Chapman and P Earley</i>	719
The International System for Teacher Observation and Feedback: A Theoretical Framework for Developing International Instruments	<i>L Kyriakides, B P M Creemers, C Teddlie, and D Muijs</i>	726

Educational Evaluation – Evaluation Methods

Evaluation Methodology	<i>M Chatterji</i>	735
Evaluation Reporting and Communicating	<i>R T Torres</i>	746
Longitudinal Evaluation Designs	<i>X Ma</i>	757
Meta-evaluation: Purpose, Prescription, and Practice	<i>L A Wingate</i>	765

Educational Evaluation – Purposes and Models of Educational Evaluation

Defining Quality in Evaluation	<i>P Dahler-Larsen</i>	775
Internal and External Evaluation	<i>D Nevo</i>	781
Needs Assessment in Education	<i>J W Altschuld and T L Lepicki</i>	786
The Purpose of Educational Evaluation	<i>S Mathison</i>	792
Understanding Approaches to Evaluation	<i>A J Love</i>	798

VOLUME 4**Educational Measurement**

Ability Testing	<i>P B Tymms</i>	1
Admissions Testing	<i>R Zwick</i>	7
Appropriateness Measurement: Person Fit	<i>R R Meijer and A M Weekers</i>	15
Automated Essay Scoring: Writing Assessment and Instruction	<i>M D Shermis, J Burstein, D Higgins, and K Zechner</i>	20
Classical Test Theory Reliability	<i>M D Miller</i>	27
Criterion-Referenced Measurement	<i>S A Shrock and W C Coscarelli</i>	31
Differential Item Functioning	<i>M D Hidalgo and J Gómez-Benito</i>	36
Educational Measurement: Overview	<i>R L Linn</i>	45
Equating and Scaling	<i>A A von Davier</i>	50
Fairness	<i>C Gipps and G Stobart</i>	56
Generalizability Theory	<i>R L Brennan</i>	61
Impact of Assessment on Classroom Practice	<i>J L Herman</i>	69
Intelligent Systems	<i>V J Shute and D Zapata-Rivera</i>	75
Item Response Theory	<i>W J van der Linden</i>	81
National Assessments	<i>T J Crooks</i>	89
Norm-Referenced Measurement	<i>H D Hoover</i>	95
Progress Monitoring	<i>L S Fuchs and D Fuchs</i>	102
Rasch Models	<i>D Andrich</i>	111
Scoring Rubrics	<i>J A Arter</i>	123
Second Language Assessment	<i>L F Bachman</i>	140
Self Adaptive Testing	<i>J Revuelta</i>	148
Student Test Results in School Accountability	<i>L S Hamilton</i>	153
Test Development	<i>S M Downing</i>	159
Test Translation and Adaptation	<i>A Allalouf and P Hanani</i>	166
Testing Creativity	<i>M A Runco</i>	170
Testing in History	<i>S Wineburg and B Fogo</i>	175
Validity	<i>R L Linn</i>	181
Validity of Achievement Gains on High-Stakes Tests	<i>D Koretz</i>	186
Validity of Educational Indicators	<i>B P M Creemers and L Kyriakides</i>	193

G**Globalization of Universities and Institutions**

Globalization, Universities and Medium of Instruction	<i>J Lo Bianco</i>	201
The Washington Accord on Engineering Curriculum	<i>E Ko</i>	209

H

Higher Education

Higher Education: An Overview	<i>P Scott</i>	217
-------------------------------	----------------	-----

Higher Education – Management, Leadership and Governance

Academic Managers – Heads of Departments/Schools	<i>S Marshall</i>	229
Leading and Managing the University – Presidents and their Senior Management Team	<i>R Middlehurst, T Kennie, and S Woodfield</i>	238
Management of and in Higher Education Institutions	<i>B Sporn</i>	245
Managerialism and Collegialism in Higher Education Institutions	<i>M Shattock</i>	251
Strategic Planning in Higher Education	<i>L Zechlin</i>	256
The Changing Role of University Presidents, Vice-Chancellors and Rectors	<i>M F Green and P D Eckel</i>	264
The Changing Roles of University Governing Boards and Councils	<i>J Fielden</i>	273

Higher Education – Research and Scholarship

Funding of University Research	<i>G Harman</i>	279
Mobility of PhD Students and Scientists	<i>S Avveduto</i>	286
National Science Policy and Universities	<i>M Henkel and U Kogan</i>	294
Research Commercialization	<i>S Slaughter</i>	300
Research Quality Assessment	<i>I McNay</i>	307
The Management of University Research	<i>J Kirkland</i>	316
The Training of New Scientists	<i>R Neumann</i>	322

Higher Education – Society

Academic Freedom	<i>S F Macintyre</i>	328
Access and Equity in Higher Education	<i>G Baldwin and R James</i>	334
Boundaries of Institutional Autonomy and their Impact on Higher Education	<i>V L Meek</i>	341
Diversity of Higher Education	<i>U Teichler</i>	347
Employability of University Graduates and Graduate Outcomes	<i>J Brennan</i>	354
Higher Education and the Knowledge Society	<i>J Välimaa</i>	360
Higher Education and the Labor Market	<i>J Enders</i>	365
Higher Education and the Transformation of Society	<i>P Scott</i>	370
Quality Assurance in Higher Education – Practices and Issues	<i>D Dill</i>	377
The Role of Accreditation of Higher Education Institutions	<i>J S Eaton</i>	384
Universities and Regional Development	<i>J Goddard and J Puukka</i>	390
Universities' Engagement with Society	<i>D Watson</i>	398

Higher Education – Teaching and Learning in Higher Education

Distance Education and Open Universities	<i>A Kanwar and J Daniel</i>	404
Education Development in Higher Education	<i>Y Ryan and K Fraser</i>	411
Effective Use of Technology in Teaching and Learning in HE	<i>D Laurillard</i>	419

Facilitating Students' Success – Student Support and Learning Resources in Universities	<i>E Woodberry and C Barnes</i>	427
Faculty Values and Expectations and Their Implications for Teaching, Learning and the Organization of Higher Education Institutions	<i>K Harman</i>	433
Flexible Learning in Higher Education	<i>S Alexander</i>	441
Innovation in Teaching and Curriculum Design	<i>J M Dearn</i>	448
Productivity of University Faculty Staff	<i>S Kyvik</i>	455
Promoting Effective Student Learning in Higher Education	<i>K Trigwell</i>	461
Student Attrition from Higher Education Institutions	<i>M Hayden</i>	467
University Students' Experiences of Higher Education	<i>C McInnis</i>	473
What Do University Students Learn?	<i>G D Kuh and B K Bridges</i>	479

Higher Education – Tertiary Education in a Globalized World

Associations of Universities	<i>G F Moodie</i>	485
Collective Bargaining in Higher Education	<i>C Allport, K McAlpine, and S Roberts</i>	491
Financing of Higher Education	<i>B Chapman and P Tulip</i>	499
Higher Education Crossing Borders	<i>J Knight</i>	507
Higher Education in Post-Conflict Societies	<i>M Daxner</i>	514
Higher Education in the New Economies	<i>J Fielden</i>	520
Human Resource Issues in Higher Education	<i>E El-Khawas</i>	527
International Accreditation in Higher Education	<i>A Stella and D Woodhouse</i>	533
Internationalization of Higher Education	<i>M van der Wende</i>	540
National and International Rankings of Higher Education	<i>S Marginson</i>	546
Steering of Higher Education Systems – The Role of the State	<i>K Yokoyama and V L Meek</i>	554
Student and Faculty Transnational Mobility in Higher Education	<i>U Lanzendorf and B M Kehm</i>	559
The Bologna Process in European Higher Education	<i>D Rich</i>	566
The Economics and Finance of Higher Education	<i>A Wolf</i>	573
The Role of the OECD in the Development of Higher Education in a Globalized World	<i>R Yelland</i>	584
Transforming Higher Education in Developing Countries: The Role of the World Bank	<i>J Salmi and R M Bassett</i>	590
UNESCO's Role in the Development of Higher Education in a Globalized World	<i>S Uvalić-Trumbić</i>	597

Higher Education – The Organization and Financing of Higher Education

Higher Education in Federal Systems	<i>M A M Cortés and U Teichler</i>	603
Pathways and Articulation into Higher Education	<i>G F Moodie</i>	609
Unitary, Binary and other National Systems of Higher Education	<i>L Goedegebuure</i>	616
Viewing Private Higher Education: How Much, Where, Why, and What?	<i>D C Levy</i>	622

I

Immigration

Children of Migrant Populations	<i>M M Suárez-Orozco and C Suárez-Orozco</i>	629
Third Culture Kids	<i>R E Van Reken</i>	636

International Comparisons

IEA Studies in ICT	<i>N Law</i>	644
IEA Studies in Mathematics and Science	<i>F K S Leung</i>	650
IEA Study in Civic Education	<i>J Torney-Purta, J-A Amadeo, and J Schwillie</i>	656
IEA: Globalization and Assessment	<i>H Wagemaker</i>	663

International Organizations in Education

Education and the European Union	<i>S Hegarty</i>	669
National University on Educational Planning and Administration, New Delhi	<i>J B G Tilak</i>	676
On School Management – International Council on School Effectiveness and Improvement, Commonwealth Council on Educational Administration and Management	<i>A Walker and T Townsend</i>	681
UNESCO's International Institute for Educational Planning	<i>S J Hite and M Bray</i>	688

L**Leadership and Management**

Leadership and Management Overview	<i>B Mulford</i>	695
------------------------------------	------------------	-----

Leadership and Management – as a Field of Study

Contexts for Leadership at the Beginning of the 21st Century	<i>M Bottery</i>	704
Educational Leadership: Philosophical Issues	<i>C W Evers</i>	710
Ethical/Moral Issues in Educational Leadership	<i>R J Starratt, L Langlois, and P Duignan</i>	717
History of Educational Leadership/Management	<i>R Bates</i>	724
Educational Leadership as a Field of Study: Professional Journals 1994–2006	<i>A R Thomas</i>	731
Research on Educational Leadership – Approaches/Promising Directions	<i>W A Firestone and V M J Robinson</i>	740

Leadership and Management – Leader Preparation

Leaders for Productive Schools	<i>J Murphy, S N Elliott, E B Goldring, and A C Porter</i>	746
New Approaches in Preparing School Leaders	<i>S G Huber</i>	752
Research for Leaders: The National College for School Leadership	<i>G Southworth</i>	762
School Leadership Standards	<i>L Ingvarson</i>	768

Leadership and Management – Leadership Issues

Effective Leadership in Challenging Schools	<i>A Harris</i>	777
Esthetics, Educational Leadership and Management	<i>R Bates</i>	783
Leader Mistreatment of Teachers	<i>J Blase and J Blase</i>	790
Leadership and Gender	<i>J Blackmore</i>	797
Leadership and Student Voice	<i>C Lanskey and J Rudduck</i>	803
Leadership and Urban Education	<i>K A Riley</i>	809
Leadership for Learning	<i>J MacBeath</i>	817
Leadership in Diverse Cultures	<i>A Walker and P K Yu-kwong</i>	824

Leading Diverse Schools	<i>J Ryan</i>	832
Supervision and Personnel Administration	<i>P D Tucker and D G Pounder</i>	839
The Big Challenge: Leadership for Inclusion	<i>M Ainscow and A Sandill</i>	846
The Educational Leader's New Work	<i>P Gronn</i>	852

VOLUME 5

Leadership and Management – Leadership Types

A Distributed Perspective on School Leadership and Management	<i>J P Spillane</i>	1
Leadership: Authentic	<i>P T Begley</i>	7
Leadership: Democratic	<i>J Møller</i>	12
Leadership: Instructional	<i>P Hallinger, K Leithwood, and R H Heck</i>	18
Leadership: Transformative	<i>C M Shields</i>	26
Strategic Leadership	<i>B Davies and B J Davies</i>	34
Teacher Leadership and Organizational Development	<i>A Harris</i>	40
Understanding How Leadership Influences Student Learning	<i>K Leithwood and B Levin</i>	45

Leadership and Management – Politics and Governance

Equity and Educational Effectiveness	<i>P Sammons</i>	51
Leading/Managing Schools in Communities Made Poor	<i>P Thomson</i>	58
Principals Role in Restructuring Schools	<i>G M Crow and K D Peterson</i>	66
School Reform and Restructuring: Self Managing School	<i>B J Caldwell</i>	72
System Leadership	<i>D Hopkins and R Higham</i>	78

Leadership and Management – School and Community

Educational Leadership and Social Capital	<i>J West-Burnham</i>	85
Networked Learning Communities School-to-School Collaboration as an Essential Component of a System Reform Strategy	<i>D Jackson</i>	92
Networks and Communities of Knowledge	<i>T Schuller and H Theisens</i>	100
School Community Relations	<i>R L Crowson and E B Goldring</i>	106
Social Capital, Educational Institutions and Leadership	<i>S Kilpatrick, S Johns, and B Mulford</i>	113

Leadership and Management – School Effectiveness and Improvement

Leadership and Technology	<i>C F Webber</i>	120
Leadership in the Implementation of Innovations	<i>I E Bogotch, T Townsend, and M Acker-Hocevar</i>	128
Leadership: School Improvement	<i>R H Heck and P Hallinger</i>	135
Organizational Learning in Schools	<i>B Mulford and H Silins</i>	143
Professional Learning Community	<i>L Stoll</i>	151
Transformational School Leadership	<i>K Leithwood</i>	158

Learning and Cognition – Issues and Concepts

Attention in Cognition and Early Learning	<i>M Heimann, T Tjus, and K Strid</i>	165
Concept Learning	<i>S J Unsworth and D L Medin</i>	172
Cooperative Learning	<i>R E Slavin</i>	177
Intelligence	<i>R J Sternberg</i>	184
Learning from Multiple Information Sources	<i>H I Strømsø and I Bråten</i>	191
Memory	<i>S Magnussen and T Brennen</i>	197
Metacognition	<i>L Baker</i>	204
Personal Epistemology in Education: Concepts, Issues, and Implications	<i>I Bråten</i>	211
Piaget: Recent Work	<i>L Smith</i>	218
The Adult Development of Cognition and Learning	<i>E A L Stine-Morrow and J M Parisi</i>	225
Vygotsky and Recent Developments	<i>J V Wertsch</i>	231

Learning and Cognition – Issues and Concepts – Focus on Cognition

Cognition and Emotion	<i>F Pons, M de Rosnay, and F Cuisinier</i>	237
Cognition: Overview and Recent Trends	<i>J Wiley and B D Jee</i>	245
Development of Creativity	<i>M A Runco</i>	251
Knowledge Domains and Domain Learning	<i>L Maggioni and P A Alexander</i>	255
Problem Solving and Human Expertise	<i>T J Nokes, C D Schunn, and M T H Chi</i>	265
Problem Solving and Reasoning	<i>R E Mayer</i>	273
Reasoning, Moral and Social	<i>N S Jampol, C B Richardson, and M Killen</i>	279
Remembering as Social Activity	<i>D Middleton</i>	285

Learning and Cognition – Issues, Concepts, Types

Computer-Supported Collaborative Learning: Basic Concepts, Multiple Perspectives, and Emerging Trends	<i>S R Ludvigsen and A I Mørch</i>	290
Learning as Inquiry	<i>S R Goldman, J Radinsky, S Tozer, and D Wink</i>	297
Learning Through Narrative Socialization	<i>A Fasulo</i>	303
Neuroscience Bases of Learning	<i>M H Immordino-Yang and K W Fischer</i>	310
The Neuroscience of Reading	<i>U Goswami</i>	317

Learning and Cognition – Issues, Concepts, Types – Focus on Learning

Learning Strategies	<i>C E Weinstein, J Jung, and T W Acee</i>	323
Learning Through Play	<i>K Aronsson</i>	330
Situated View of Learning	<i>A Collins and J G Greeno</i>	335
The Neuroscience of Aging and Cognition	<i>W J Hoyer and B Bessette-Symons</i>	340

Learning and Cognition – Language, Literacy and Subject-Related

An Overview of Language and Literacy in Educational Settings	<i>V G Aukrust</i>	345
Bilingualism and Learning	<i>L S Verplaetse and E Schmitt</i>	355
Classroom Discourse and Student Learning	<i>L Hemphill</i>	361
Early Writing	<i>B E Hagtvet</i>	367

First Language Acquisition	<i>N B Ratner</i>	375
Foreign Language Learning	<i>C Kinginger</i>	382
Learning Science	<i>J Leach and P Scott</i>	387
Learning to Read	<i>M S Burns and J K Kidd</i>	394
Mathematics Learning	<i>L Verschaffel, B Greer, and E De Corte</i>	401
Music Learning	<i>S Hallam</i>	407
Reading Comprehension: Reading for Learning	<i>C E Snow</i>	413
Second Language Learning	<i>J S Arnfast, J N Jørgensen, and A Holmen</i>	419
Writing: Advanced	<i>R Andrews</i>	426

Learning and Cognition – Other Issues

Gender and Learning	<i>L Yates and J McLeod</i>	432
Parent–Child Relationships in Early Learning	<i>C P Edwards, S M Sheridan, and L L Knoche</i>	438
Peer Interaction and Learning	<i>S Blum-Kulka and S Dvir-Gvirsman</i>	444

Learning and Cognition – Places of Learning

Learning in a Cross-Cultural Perspective	<i>L-F Zhang and R J Sternberg</i>	450
Learning Outside of School	<i>T Nunes</i>	457
Reasoning and Explanatory Talk: Learning in Everyday Settings	<i>D E Beals</i>	464

Learning and Cognition – Theoretical Perspectives – Learning

Apprenticeship Approach to Learning	<i>K Nielsen</i>	469
Connectionism and Learning	<i>T R Shultz</i>	476
Constructivism and Learning	<i>S Sjøberg</i>	485
Cultural–Historical Activity Theory	<i>A Havnes</i>	491
Learning in a Sociocultural Perspective	<i>R Säljö</i>	498
Self-Regulated Learning and Socio-Cognitive Theory	<i>P H Winne and A F Hadwin</i>	503

N

National Systems of Education

Australia	<i>C N Power</i>	510
Austria	<i>M J Gaertner</i>	517
Bangladesh	<i>R Z Dewan</i>	524
Barbados	<i>R Rudder</i>	534
Bulgaria	<i>A Chavdarova and D Vesselinov</i>	540
England	<i>C Whatford</i>	550
Estonia	<i>M Keränen and J Kõrgesaar</i>	555
Fiji	<i>I Senibulu</i>	563
Finland	<i>J E Tuovinen</i>	569
France	<i>M Sellier</i>	576

Germany	<i>A Hüfner and K Hüfner</i>	582
Ghana	<i>R Palmer</i>	590
Greece	<i>K Toura and A Karagiannis</i>	599
Iceland	<i>M Hardardottir and M Gunnlaugsdóttir</i>	605
Indonesia	<i>A C Alwasilah and Furqon</i>	611
Israel	<i>A Volansky</i>	618
Jamaica	<i>E Hannam</i>	627
Latvia	<i>A Joma</i>	630
Lebanon	<i>N A Mneimneh</i>	636
Lithuania	<i>J Navickaitė, L Paurienė, R Zablackė, A Tijūnelytė, and V Kožemiakina</i>	656
Malawi	<i>F R Mkandawire and D Mulera</i>	666
Malta	<i>C Farrugia</i>	673
Mongolia	<i>B Yembuu</i>	681
Namibia	<i>J Ellis</i>	687
New Zealand	<i>R Baker and C Wylie</i>	692
Niger	<i>A A I Daouda</i>	699
Norway	<i>H Farstad</i>	709
Oman	<i>S A Issan</i>	718
Pakistan	<i>N Mahmood and R Malik</i>	727
Poland	<i>A Janowski</i>	737
Qatar	<i>S R Salman</i>	744
Russia	<i>V M Filippov and G F Tkach</i>	750
Rwanda	<i>E Bahizi</i>	758
Scotland	<i>D Raffe</i>	770
Singapore	<i>P T Ng</i>	776
Slovenia	<i>T Plevnik and A B Lakota</i>	782
Solomon Islands	<i>D P Malasa</i>	790
South Africa	<i>P Kgobe and J Pampallis</i>	799
Spain	<i>A Tiana</i>	806
Sri Lanka	<i>B M J Balasooriya, M M Wehella, and S U Wijeratne</i>	813
Suriname	<i>H R Ori</i>	826
Taiwan	<i>C-M Lee</i>	834
Tanzania	<i>R Wedgwood</i>	841
Thailand	<i>C Rukspollmuang</i>	847
The Gambia	<i>Y A-M Jobe</i>	858
The Philippines	<i>E A P Valenzuela</i>	865
Uruguay	<i>B Guinovart and C Troncoso</i>	873
USA	<i>B Pusser and M E R Lloyd</i>	888
Zanzibar	<i>I M Hija</i>	896

VOLUME 6**P****Philosophy of Education**

Philosophy of Education: Overview	<i>P Enslin</i>	1
-----------------------------------	-----------------	---

Philosophy of Education – Contemporary Issues

Neoliberalism, the Market and Performativity	<i>M A Peters</i>	11
Parents and Children's Rights	<i>M Clayton</i>	17
Peace Education	<i>I Gur-Ze'ev</i>	22
Philosophy and Educational Research	<i>D Bridges</i>	29
Race, Critical Race Theory and Whiteness	<i>B Applebaum</i>	36
School Choice and the Common School	<i>G Haydon</i>	44

Philosophy of Education – Philosophical Perspective

African Philosophy of Education	<i>K Horsthemke</i>	50
Feminism	<i>L Stone</i>	56
Hermeneutics	<i>C Bingham</i>	63
Islamic Education	<i>Y Waghid</i>	69
Liberalism and Education	<i>W Feinberg</i>	74
Phenomenology and Existentialism	<i>B Levering</i>	80
Pragmatism	<i>B Thayer-Bacon</i>	86
Psychoanalysis and Education	<i>P Smeyers</i>	92
Social and Cultural Capital in Education	<i>M Print</i>	97
The Aims of Education	<i>R Marples</i>	103
The Analytical Tradition	<i>D Carr</i>	108

Philosophy of Education – Philosophical Themes

Adult Learning: Philosophical Issues	<i>D Beckett</i>	114
Autonomy	<i>C A Wringe</i>	121
Bildung	<i>J Masschelein and N Ricken</i>	127
Critical Theory and Pedagogy	<i>M Mason</i>	133
Critical Thinking	<i>H Siegel</i>	141
Environmental Education	<i>M Bonnett</i>	146
Identity	<i>M S Merry</i>	152
Inclusion	<i>R Cigman</i>	158
Justice and Care	<i>M Tjiattas</i>	164
Lifelong Learning	<i>J Halliday</i>	170
Post-Structuralism and Education	<i>M Mason and M Clarke</i>	175
Reason and Rationality	<i>P Hogan</i>	183

Religious and Spiritual Education	<i>J C Conroy and S J McKinney</i>	189
The Capabilities Approach	<i>L Terzi</i>	194

Primary and Secondary Education – Learning and Teaching in School Age Education

Class Size – Arguments and Evidence	<i>P Blatchford and K C Lai</i>	200
Continuing Professional Development of Teachers	<i>P Earley</i>	207
Homeschooling	<i>P Farenga</i>	214
Innovations and Early Intervention in the Teaching of Literacy	<i>R E Slavin</i>	221
Personalized Learning in School Age Education	<i>D Hopkins</i>	227
Perspectives on Schooling in the Middle Years	<i>V J Russell</i>	233
Reform of the School Workforce	<i>H M Gunter and G Forrester</i>	242
Secondary Education in Developing Countries	<i>P Hughes and A Walker</i>	249

Primary and Secondary Education – Social Justice Issues in School Age Education

Community Focused Schools	<i>A Dyson</i>	255
---------------------------	----------------	-----

Primary and Secondary Education – The Effectiveness of School Systems

Reforming the American School System	<i>M Barber</i>	261
School Effectiveness in Developed Societies	<i>D Reynolds</i>	269
School to School Collaboration: Innovation and Improvement	<i>R Higham</i>	274
Studies of School Improvement in Developing Countries	<i>S E Anderson</i>	281
Sustainable Educational Reform	<i>A Hargreaves</i>	289
The Role of the District in Tri Level Reform	<i>M Fullan</i>	295
Whole School Designs for Enhancing Student Achievement	<i>S Stringfield and J Nunnery</i>	303

Q

Qualitative Research

Action Research in Education	<i>E Stringer</i>	311
Classroom Ethnography	<i>F Erickson</i>	320
Computer Assisted Qualitative Data Analysis	<i>C Silver and A Lewins</i>	326
Conversational Analysis	<i>H G Levine</i>	335
Critical Ethnography	<i>D Foley</i>	337
Critical Race Theory	<i>D Gillborn and G Ladson-Billings</i>	341
Critical Theory	<i>J J Ryoo and P McLaren</i>	348
Cultural Studies	<i>R Gaztambide-Fernández and K Cairns</i>	354
Cultural–Historical Activity Theory	<i>Laboratory for Comparative Human Cognition</i>	360
Design Experiments	<i>A Collins</i>	367
Discourse Analysis	<i>P Nikander and S Vehviläinen</i>	373
Document Analysis	<i>T Rapley and K N Jenkins</i>	380
Ethnography	<i>M Hammersley</i>	386

Ethnomethodology in Education Research	<i>D Macbeth</i>	392
Focus Groups	<i>D L Morgan and K Hoffman</i>	401
Grounded Theory	<i>K Charmaz and A Bryant</i>	406
Hermeneutics	<i>J Dyer</i>	413
Interpretive Research	<i>E Rauscher and M E Graue</i>	419
Interviews and Interviewing	<i>A Marvasti</i>	424
Life History	<i>J A Hatch and S K Newsom</i>	430
Narrative Inquiry	<i>D J Clandinin and J Huber</i>	436
Participant Observation	<i>J Schostak</i>	442
Phenomenology	<i>M van Manen and C A Adams</i>	449
Qualitative Case Studies	<i>S B Merriam</i>	456
Qualitative Data Management	<i>C Grbich</i>	463
Semiotics	<i>M Danesi</i>	475
Symbolic Interactionism, Naturalistic Inquiry, and Education	<i>J A Forte</i>	481
The Quality of Evidence in Qualitative Research	<i>S K Biklen</i>	488
Visual Data in Education Research	<i>J Wagner</i>	498

Qualitative Research – Standards of Proof

Validity: Mapping Diverse Perspectives	<i>G B Rossman, S F Rallis, and A M Kuntz</i>	505
--	---	-----

Quantitative and Qualitative Approaches to Research – Integration

Mixed Methods	<i>A Tashakkori and I Newman</i>	514
---------------	----------------------------------	-----

Quantitative Methods and Research Design

External Validity	<i>G E Matt, A Brewer, and M Sklar</i>	521
-------------------	--	-----

R

Role of NGOs in Globalization

NGOs and Globalization of Education	<i>M Ahmed</i>	529
-------------------------------------	----------------	-----

S

Social and Emotional Aspects of Learning

Achievement Goal Theory: Definitions, Correlates, and Unresolved Questions	<i>T Urdan</i>	537
Affect, Mood and Emotions	<i>S Schnall</i>	544
Anxiety	<i>M Zeidner</i>	549
Attribution Theory	<i>B Weiner</i>	558
Children's Friendship	<i>S R Asher and K L McDonald</i>	564
Coping with Stressful Situations: An Important Aspect of Self-Regulation	<i>M Boekaerts</i>	570
Culture in Motivation Research: A Challenging and Enriching Contribution	<i>K Kimmel and S Volet</i>	576
Early Social Development and Schooling	<i>G M van der Aalsvoort</i>	585

Emotion in Educational Contexts	<i>P A Schutz, P D Quijada, S de Vries, and M Lynde</i>	591
Empathy	<i>R Arnold</i>	597
Flow in Education	<i>J A Schmidt</i>	605
Interest	<i>M Ainley</i>	612
Intrinsic and Extrinsic Motivation	<i>C Levesque, K J Copeland, M D Pattie, and E L Deci</i>	618
Motivating Students in Classrooms	<i>J Brophy</i>	624
Motivation Regulation	<i>C A Wolters and S A Mueller</i>	631
Peer Learning in the Classroom	<i>N M Webb</i>	636
Poverty, Effects of on Social and Emotional Development	<i>N Philipsen Hetzner, A D Johnson, and J Brooks-Gunn</i>	643
Seeking Help as an Adaptive Response to Learning Difficulties: Person, Situation, and Developmental Influences	<i>S A Karabenick and R S Newman</i>	653
Self-Concept in Learning: Reciprocal effects model between academic self-concept and academic achievement	<i>H W Marsh and L F Scalas</i>	660
Self-Efficacy Beliefs	<i>D H Schunk and F Pajares</i>	668
Social and Emotional Outcomes of Learning	<i>J Vorhaus</i>	673
Social Aspects of Collaborative Learning	<i>M Arvaja and P Häkkinen</i>	685
Social Development and Schooling	<i>G W Ladd</i>	691
Social Interaction and Learning	<i>K Littleton</i>	698
Social Networks and the Education of Children and Youth	<i>B Schneider, T G Ford, and L Perez-Felkner</i>	705
Sociocultural Issues in Motivation	<i>R A Walker</i>	712
Socioemotional Aspects of Technology-Supported Learning	<i>S Järvelä</i>	718
Volitional Control of Learning	<i>J Husman and L Corno</i>	724
Wellbeing	<i>T Hascher</i>	732

VOLUME 7

Statistics

An Overview of Statistics in Education	<i>S Sinharay</i>	2
Analysis and Interpretation of Multivariate Data	<i>D J Bartholomew</i>	12
Analysis of Covariance	<i>N T Longford</i>	18
Analysis of Extreme Values in Education	<i>J Beirlant, G Dierckx, K Luwel, and P Onghena</i>	25
Analysis of Variance	<i>B M King</i>	32
Bayesian Statistical Analysis	<i>G Kokolakis</i>	37
Bootstrap Method	<i>K Singh and M Xie</i>	46
Canonical Correlation	<i>A Basu and A Mandal</i>	52
Categorical Data Analysis	<i>S E Fienberg and B W Junker</i>	58
Causal Inference	<i>D B Rubin</i>	66
Cluster Analysis: Overview	<i>S Landau and I Chis Ster</i>	72

Cognitive Psychology and Educational Statistics	<i>H S Kim and S Embretson</i>	84
Computational Statistics	<i>J E Gentle</i>	93
Continuous Probability Distributions	<i>S Sinharay</i>	98
Correspondence Analysis	<i>M Greenacre</i>	103
Data Mining	<i>R S J d Baker</i>	112
Decision Theory	<i>M Mendoza and E Gutiérrez-Peña</i>	119
Design of Experiments	<i>C A Albers and T R Kratochwill</i>	125
Discrete Probability Distributions	<i>S Sinharay</i>	132
Discrimination and Classification	<i>R C Neath and M S Johnson</i>	135
Empirical Bayes Methods	<i>E Ip and G Molenberghs</i>	142
Evaluation Research	<i>T Kellaghan</i>	150
Exploratory Data Analysis	<i>A Unwin</i>	156
Factor Analysis: An Overview and Some Contemporary Advances	<i>F A N Alhija</i>	162
Generalized Linear Mixed Models	<i>S Rabe-Hesketh and A Skrondal</i>	171
Generalized Linear Models	<i>J J Faraway</i>	178
Generating Random Numbers	<i>P J Pashley and A Amodeo</i>	184
Goodness-of-Fit Testing	<i>A Maydeu-Olivares and C García-Forero</i>	190
Graphical Models	<i>R G Almond</i>	197
Growth Modeling	<i>I Plewis</i>	203
Hierarchical Linear Models	<i>D Rindskopf</i>	210
Hypothesis Testing and Confidence Intervals	<i>T R Konold and X Fan</i>	216
Instrumental Variables	<i>E P Bettinger</i>	223
Jackknife Methods	<i>S Sinharay</i>	229
Large-sample Statistical Methods	<i>J M Singer and D F Andrade</i>	232
Latent Class Models	<i>J K Vermunt</i>	238
Markov Chain Monte Carlo	<i>A M Johansen</i>	245
Matrix Algebra	<i>H Yanai and H Ishii</i>	253
Measure of Association	<i>G-H Huang</i>	260
Measures of Central Tendency	<i>D Clark-Carter</i>	264
Measures of Dispersion, Skewness and Kurtosis	<i>D Rindskopf and M Shiyko</i>	267
Meta Analysis	<i>J Sánchez-Meca and F Marín-Martínez</i>	274
Missing Data	<i>C A W Glas</i>	283
Model Selection	<i>S J Haberman</i>	289
Monte Carlo Methods	<i>A M Johansen</i>	296
Multidimensional Scaling	<i>Z Zhang and Y Takane</i>	304
Multiple Comparisons	<i>C Lewis</i>	312
Multivariate Analysis of Variance	<i>N T Longford</i>	319
Multivariate Linear Regression	<i>S Ganesh</i>	324
Multivariate Normal Distribution	<i>J Tacq</i>	332

Nonlinear Regression Analysis	<i>H-H Huang, C K Hsiao, and S-Y Huang</i>	339
Nonparametric Statistical Methods	<i>K Sijtsma and W H M Emons</i>	347
Observational Studies	<i>P R Rosenbaum</i>	354
Order Statistics	<i>I Bairamov</i>	360
Point Estimation Methods with Applications to Item Response Theory Models	<i>F Bartolucci and L Scrucca</i>	366
Principal Components Analysis	<i>D J Bartholomew</i>	374
Probability Theory	<i>T Rudas</i>	378
Recursive Partitioning	<i>S Landau and S Barthel</i>	383
Sampling	<i>J Qian</i>	390
Sequential Probability Ratio Test	<i>T J H M Eggen</i>	396
Sequential Testing	<i>H J Vos</i>	401
Signal Detection Theory	<i>H Abdi</i>	407
Small Area Estimation	<i>A Noble</i>	411
Statistical Analysis of Functional Data	<i>F Ferraty, P Sarda, and P Vieu</i>	417
Statistical Inequalities	<i>H M Mahmoud and J L Gastwirth</i>	426
Statistical Paradoxes	<i>Y-K Tu</i>	429
Statistical Power Analysis	<i>L V Hedges and C Rhoads</i>	436
Statistical Significance Versus Effect Size	<i>X Fan and T R Konold</i>	444
Stochastic Processes	<i>F Baudoin</i>	451
Structural Equation Models	<i>S-Y Lee and X-Y Song</i>	453
Survival Data Analysis	<i>A F De Champlain</i>	459
The Normal Distribution and its Applications	<i>J Tacq</i>	467
Time Series Analysis	<i>P J Brockwell</i>	474
Univariate Linear Regression	<i>M L Hazelton</i>	482
Value-Added Models	<i>D F McCaffrey</i>	489

T

Teacher Education – Evaluation in the Teacher Education Program

Accreditation and Standards in Teacher Education	<i>R J Yinger and K L Daniel</i>	495
Determining Long Term Effects of Teacher Education	<i>C N Brouwer</i>	503

Teacher Education – Evaluation of the Prospective Teacher

The Evaluation of Prospective Teachers in Teacher Education	<i>D E H Tigelaar and J van Tartwijk</i>	511
---	--	-----

Teacher Education – Macro Factors

Diversity in Teacher Education	<i>T Wubbels</i>	518
Social and Cultural Influences on Teacher Education	<i>M Ben-Peretz and R Lotan</i>	525
The Role of Governance in Teacher Education	<i>G E Karlsen</i>	532

Teacher Education – Post-Initial Professional Development

Action Research as a Tool for Teachers' Professional Development	<i>P Ponte</i>	540
Contemporary Approaches to Teacher Professional Development	<i>H Borko, J Jacobs, and K Koellner</i>	548
School Development for Teacher Learning and Change	<i>P Sleegers and K Leithwood</i>	557
Teacher Induction	<i>D Beijaard, J Buitink, and C Kessels</i>	563
Teacher Learning as Workplace Learning	<i>J Imants and K van Veen</i>	569
Teacher Learning with Lesson Study	<i>A Murata</i>	575
Teachers Career Stages and Professional Development	<i>R Fessler and E Rice</i>	582

Teacher Education – Preservice Teacher Education

A Pedagogy of Teacher Education	<i>J Loughran</i>	587
Case Methods in Teacher Education	<i>M Sato and C Rogers</i>	592
Conceptions of Teacher Education	<i>J A Whitcomb</i>	598
Inquiry-Oriented Teacher Education	<i>R Schulz</i>	604
Narratives and Biography in Teacher Education	<i>G Kelchtermans</i>	610
Supervision in Teacher Education	<i>J McCarthy and L F Quinn</i>	615
Taking Prospective Teachers' Beliefs into Account in Teacher Education	<i>M Valcke, G Sang, I Rots, and R Hermans</i>	622
Teacher Education and Models of Teacher Reflection	<i>V K LaBoskey</i>	629
Teacher Education as Teaching for Understanding with New Technologies	<i>M S Wiske and D E Spicer</i>	635

Teacher Education – Preservice: The Knowledge Base

Experienced Teachers' Craft Knowledge	<i>P C Meijer</i>	642
Moral Values in Teacher Education	<i>W Veugelers</i>	650
Pedagogical Content Knowledge	<i>J H Van Driel and A Berry</i>	656
Teacher Education and the Educational Foundations Knowledge Base	<i>F B Murray</i>	662
The Relationship Between Theory and Practice in Teacher Education	<i>F A J Korthagen</i>	669

Teacher Education – Teacher Educators

Characteristics, Scholarship and Research of Teacher Educators	<i>M Lunenberg</i>	676
Professional Development of Teacher Educators	<i>P K Smith</i>	681
Self-Study by Teacher Educators	<i>T Russell</i>	689

Teacher Education – Teaching Specific Domains

Arts Teacher Education	<i>L Beudert</i>	695
Mathematics Teacher Education	<i>K Krainer and S Llinares</i>	702
Music Teacher Education	<i>M W-y Wong</i>	706
Science Teacher Education	<i>J H Van Driel and S K Abell</i>	712
Social Studies Teacher Education	<i>A-L Halvorsen and S M Wilson</i>	719
Teacher Education for Elementary Education	<i>K D Brown and G N Carnes</i>	725

Teacher Education for Teaching the Gifted	<i>A Y Baldwin</i>	732
Teacher Preparation in Mother Tongue Education	<i>M Kooy and H Hulshof</i>	738
Vocational Teacher Education	<i>S Johansson</i>	744

Teacher Education – The Cooperating School

Mentoring in Teacher Education	<i>P D Tomlinson, A J Hobson, and A Malderez</i>	749
Partnerships Between Schools and Higher Education	<i>C Day and L Smethem</i>	757

VOLUME 8

Technology and Learning

An Overview of Technology and Learning	<i>B Means and J Roschelle</i>	1
Educational Data Modeling	<i>S Iksal</i>	11
Student Data Systems and Their Use for Educational Improvement	<i>J C Wayman, V Cho, and M P Richards</i>	14
Technology and Learning: Access in Schools Around the World	<i>R Anderson and J Ainley</i>	21

Technology and Learning – as a Subject of Instruction

Conceptions of Technology Literacy and Fluency	<i>O Erstad</i>	34
--	-----------------	----

Technology and Learning – Assessment

Technology and Formative Assessment	<i>J W Pellegrino</i>	42
Technology for Large-Scale Assessment	<i>R E Bennett</i>	48
Technology Supports for Assessment Design	<i>R J Mislevy, F I Winters, I I Bejar, R E Bennett, and G D Haertel</i>	56

Technology and Learning – Education Reform and Economic Development

National Strategies to Build a Technology Workforce	<i>H-M Cheah</i>	66
Public–Private Partnerships for Educational Reform	<i>V Quah, C P Lim, and C Brook</i>	73
Relating Technology, Education Reform and Economic Development	<i>R B Kozma</i>	81

Technology and Learning – Evaluating Technology

Examining the Effects of Educational Technology Programs: Challenges and Strategies	<i>M Russell</i>	88
---	------------------	----

Technology and Learning – Global Trends

Equity in Technology Access and Opportunities	<i>T Matuchniak and M Warschauer</i>	95
Internet-Based Education	<i>M Bakia</i>	102

Technology and Learning – Issues

Media Use and School Achievement	<i>K Roe</i>	109
Technology and Physical and Social Health	<i>K Subrahmanyam</i>	112

Technology and Learning – Meeting Special Students’ Needs

Universal Design for Learning *D H Rose and J W Gravel* 119

Technology and Learning – School Leadership for Technology Integration

Technology as a Support for School–Community Connections *K M Culp* 125

Technology and Learning – Supports for Managing Instruction

Classroom Uses of Technology to Manage Instruction *W R Penuel* 133

Technology and Learning – Supports for Skill Learning

New Media, Learning from *W Schnotz and H Horz* 140

Reading and Technology *N DeFrance, D Khasnabis, and A S Palincsar* 150

Technological Supports for Acquiring Twenty-First-Century Skills *C Dede* 158

Technology Supports for Acquiring Inquiry Skills *T de Jong* 167

Technology Supports for Acquiring Mathematics *M J Nathan* 172

Technology Supports for Lifelong Learning *A Collins and R Halverson* 184

Technology and Learning – Supports for Student Collaboration

Knowledge Communities in the Classroom *J D Slotta and H Najafi* 189

Technology and Learning – Supports for Subject Matter Learning

Technology Supports for Science Learning *C Delgado and J Krajcik* 197

Technology Supports for Second Language Learning *D Nunan* 204

Technology and Learning – Teacher Learning

Teacher Skills and Knowledge for Technology Integration *N Law* 211

Technology in Preservice Teacher Education *N Davis* 217

Technology Resources for Teacher Learning *J E Hinojosa, C Labbé, and L López* 222

V

Vocational Education and Training

An Overview of Vocational Education and Training *T Karmel* 229

Informal and Non-Formal Learning in Vocational Education and Training *F Beddie and S Halliday-Wynes* 240

School-Based Vocational Education and Training *B Knight and R Sweet* 247

The Economics of Vocational Education and Training *P Carneiro, L Dearden, and A Vignoles* 255

Vocational Education and Training – Individuals and Students

Age Equality in Education and Training *P Taylor* 262

Career Guidance in Vocational Education and Training and in the Workplace *P Plant* 269

Displaced Workers, Unemployed and Vocational Education and Training *B Kriechel* 275

Immigrant Investment in Host Country Vocational Education and Training *I Kogan* 280

Indigenous People and Vocational Education and Training	<i>R G Schwab</i>	289
Persons with a Disability and Vocational Education and Training	<i>F Ferrier and C S Smith</i>	294
Prisoner Education and Training	<i>S Dawe, S H Wynes, and J Misko</i>	300
Vocational Education and Training and the School-to-Work Transition	<i>G Hayward</i>	306

Vocational Education and Training – Industry and Employers

Apprenticeships	<i>E Smith</i>	312
Skills Shortages: Concepts, Measurement and Policy Responses	<i>C Shah and G Burke</i>	320
Training and Learning in the Workplace	<i>R L Jacobs</i>	328
Training in Small Businesses	<i>H Matlay</i>	337

Vocational Education and Training – Research on VET

Language of Vocational Education and Training	<i>J Misko, J MacKenzie, and R-A Polvere</i>	343
Research on Vocational Education and Training	<i>J Robson and B Bailey</i>	352
Vocational Education and Training Information Resources	<i>L-A Harris, R Naidu, and M Saunders</i>	358

Vocational Education and Training – Teaching and Learning

Employability Skills	<i>E Smith</i>	368
Learning Styles in Vocational Education and Training	<i>J Vorhaus</i>	376
Modularization in Vocational Education and Training	<i>H Ertl and G Hayward</i>	383
Personalized Learning and Vocational Education and Training	<i>I Duckett</i>	391
Recognition of Prior Learning and Vocational Education and Training	<i>J Keating</i>	397
The Competency Model	<i>A Gonczi and P Hager</i>	403

Vocational Education and Training – VET in Context

Globalization and Vocational Education and Training	<i>P Loveder</i>	411
Learning to Work for the Future	<i>J Fien, R Maclean, and D N Wilson</i>	416
The Relationship between Technical and Vocational Education and Training and Innovation	<i>K Ruth and L Deitmer</i>	423
The Status of Vocational Education and Training	<i>J Lasonen</i>	429
Vocational Education and Training in Post-Conflict Countries	<i>E Lyby</i>	441

Vocational Education and Training – VET System

Dual System	<i>T Deissinger</i>	448
Industry Involvement in the Vocational Education and Training System	<i>M Hughes and C Hughes</i>	455
Non-Completion in Vocational Education and Training: Issues, Patterns and Possibilities	<i>S Lamb and A Walstab</i>	463
Planning and Policy Development for Technical Vocational Education and Training Systems	<i>R Maclean and M Pavlova</i>	469
Qualifications Frameworks and their Role in the Reform of Education and Training	<i>M Young</i>	476
Quality Assurance in Vocational Education and Training	<i>D Gunning</i>	482

Training Finance	<i>A Ziderman</i>	489
Training Markets	<i>P Ryan</i>	496
Vocational Education and Training Teacher Training	<i>S Nielsen</i>	503
Vocational Education and Training Workforce	<i>D L Wallin</i>	513
Subject Index		519
Contributors		747

LEADERSHIP AND MANAGEMENT – LEADERSHIP TYPES

Contents

A Distributed Perspective On School Leadership and Management

Leadership: Authentic

Leadership: Democratic

Leadership: Instructional

Leadership: Transformative

Strategic Leadership

Teacher Leadership and Organizational Development

Understanding How Leadership Influences Student Learning

A Distributed Perspective on School Leadership and Management

J P Spillane, Northwestern University, Evanston, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Distributed leadership has garnered much attention from policymakers, practitioners, philanthropists, and researchers over the past decade (Hammersley-Fletcher and Brundrett, 2005; Storey, 2004). The concept, or at least the term, has easily and rather effortlessly entered the discourse about school leadership and management from Australia to America and many places in between.

The usage of the term varies widely. For some, distributed leadership is a conceptual lens or theoretical framework for studying or diagnosing the phenomena of leadership and management, whereas others use it as a prescription – an improvement strategy for schools (Bennett *et al.*, 2003; Harris, 2005; Spillane *et al.*, 2001; Spillane, 2006). Those who use distributed leadership as a conceptual lens do not do so uniformly: some focus exclusively on who takes responsibility for key organizational functions, whereas others add to this a consideration of the practice of leading and managing. Yet, even when scholars take the practice aspect into account some equate leading and managing practice with the actions or behaviors of individuals (Leithwood *et al.*, 2007), whereas others see this practice as a dynamic system or web of interactions (Spillane, 2006; Spillane and Diamond, 2007). Similarly, as a prescription for school improvement, distributed leadership has come to mean everything from simplistic mantras such as everyone is a leader in the schoolhouse, to more sophisticated models that center on work redesign (Mayrowetz and Smylie, 2004; Mayrowetz *et al.*, in press).

The appeal of distributed leadership lies partially in the ease with which it can mean many things to diverse users, but, therein also lies its Achilles' heel as a concept.

Distributed leadership's user-friendliness may portend trouble and seriously undermine its usefulness to the fields of policy, practice, and research. Regardless of whether distributed leadership is a conceptual framework or strategy for improvement, clarity about what it means is critical. Without such clarity, it is impossible for the field to use a distributed perspective to amass a knowledge base, test hypotheses, replicate studies, develop improvement strategies, and move forward. A distributed perspective will have limited utility in improving practice if policymakers and practitioners use the term to mean different things, talking past one another in their efforts to improve schools and school systems. Variance is inevitable and potentially productive as long as it is recognized and acknowledged – too often it is not.

We examine what it means to take a distributed perspective to school leadership and management. Focusing on distributed leadership as a conceptual or diagnostic frame, we examine some core elements paying attention to variation in usage. This discussion is organized around the leader plus and practice aspects of a distributed frame. Within both subsections, we also consider the empirical knowledge base, though the review is not exhaustive. While the empirical knowledge base is sparse (Bennett *et al.*, 2003), the situation is changing with numerous studies currently underway using some version of a distributed perspective to frame their investigations. Next, we consider the question of effectiveness. Finally, we take up the issue of what a distributed perspective is not by examining some common synonyms for distributed leadership.

A Distributed Perspective: Conceptual and Empirical Work

Over the past decade, scholars have worked at fleshing out a framework utilizing a distributed perspective for examining school leadership and management (Gronn, 2000, 2003; Spillane *et al.*, 2001, 2004; Spillane, 2006). A distributed perspective is a conceptual or analytical framework for researchers involved in investigating school leadership and management. Of course, it can just as easily be used as a diagnostic instrument for practitioners and interventionists working to improve practice, drawing their attention to particular dimensions of school leadership and management (Spillane, 2006). Scanning the theoretical literature, at least two aspects of a distributed perspective, are evident – the leader plus aspect and the practice aspect.

Leader Plus Aspect

Conceptual

A distributed perspective acknowledges that the work of leading and managing schools involves multiple individuals, not just those with formally designated leadership and management positions but also individuals without such designations. A distributed perspective then presses us to take into account the work of all the individuals who may have a hand in leadership and management work. Acknowledging that anyone in the school organization might have a hand in the working of leading and managing the organization is not to argue that everyone does or even should be leaders or managers. Taking a distributed perspective cautions us against investigating school leadership and management by focusing chiefly on the work of formally designated leaders.

While allowing for heroes and heroines, a distributed perspective presses us beyond the heroics of leadership paradigm (Yukl, 1999: 292; also Heller and Firestone, 1995; MacBeath 2003; Spillane, 2006). Thus, the leader plus aspect acknowledges, something that organizational scholars have long recognized, that investigations of leadership and management must move beyond those at the top of the organization and pay attention to shifting coalitions of decision makers (Cyert and March, 1963; Katz and Kahn, 1966; Barnard, 1968). Teacher leaders and other professionals are important in understanding school leadership and management (Smylie and Denny, 1990; Heller and Firestone, 1995). The distribution of responsibility for leadership and management spans across formally designated leaders and informal leaders. It is achieved through at least three arrangements that coexist in schools—division of labor, co-performance, and parallel performance (Gronn, 2003; Heller and Firestone, 1985; Spillane, 2006).

Empirical

Recent empirical work sheds considerable light on the distribution of responsibility for leadership and management

work. Multiple individuals typically perform leadership work, including formally designated leaders and individuals without these designations. One recent study of 120 geographically dispersed US elementary schools found that the responsibility for leadership functions was typically distributed among three to seven formally designated leadership positions (Camburn *et al.*, 2004). Other studies show that teachers with no formally designated leadership positions as well as school district personnel and external consultants take responsibility for leadership and management functions and organizational routines (Frost, 2005; Heller and Firestone, 1995; Portin *et al.*, 2003; Spillane, 2006; Spillane and Diamond, 2007). Even studies that focus exclusively on the school principal's workday point to the important role played by other formally designated leaders and individuals, who have no formal leadership designations in leading and managing the schoolhouse (Spillane *et al.*, 2007). Some recent studies of school leadership also underscore the need to extend investigations beyond the schoolhouse walls to school districts (Mayrowetz and Smylie, 2004; Firestone and Martinez, 2007; Leithwood *et al.*, 2007).

The distribution of responsibility for leadership and management differs depending on the organizational function or routine, school type, the school subject, school size, and a school's stage of development. The responsibilities differ depending on the formally designated position (Camburn *et al.*, 2004; Heller and Firestone, 1995). School administrators tend to take responsibility for a broad array of functions, whereas curriculum and reform specialists focus chiefly on instruction-related functions. District administrators typically were not involved in functions such as monitoring program implementation or handling disturbances in schools.

These patterns of distribution, however, also depend on the school type – public, private, charter, Catholic, or magnet. A study of 21 US schools found that the responsibility for critical areas of leadership (e.g., instruction, culture, management, human resources, and external development) differed depending on the school type (Portin *et al.*, 2003). Principals in private and entrepreneurial schools, for example, were more likely to distribute leadership for culture, strategic vision, and human resources. The distribution of responsibility for leadership and management also depends on the school subject. The number of individuals involved in the performance of organizational routines, and the extent to which formally designated leaders were involved, differed by the school subject (Burch, 2007; Sherer, in press; Spillane, 2005; Spillane *et al.*, 2003a). The developmental stage of the school or school leadership team is yet another factor that appears to influence the distribution of responsibility for leadership and management work (Copland, 2004; Frost, 2005; Harris, 2002).

Research studies are also generating knowledge about how responsibility for leadership and management gets

distributed in schools. The distribution of responsibility for leadership can result by design, default, and by crisis. The design decisions of formally designated leaders – to create a new position or organizational routine, join a particular comprehensive school reform model, hire a particular individual to fill a formally designated leadership position, and to change the responsibilities of a formally designated position – can influence the distribution of responsibility for leadership and management (Camburn *et al.*, 2004; Heller and Firestone, 1995; Spillane, 2006; MacBeath *et al.*, 2004). The distribution of responsibility for leadership and management also evolves over time as individuals understand one another's skills and weaknesses and step up to the plate when they see particular needs (Gronn, 2003; Spillane, 2006; Hargreaves and Fink, 2003); sometimes, it is in response to default. Responsibility also gets distributed through crisis when a school encounters an unanticipated problem or challenge and formal leaders and teachers find themselves working together to address it (Gronn, 2003). Of course, a crisis is in the eye of the beholder, and could be a situation where a new administration wants to fundamentally change business as usual in a school that triggers a group of veteran teachers in opposition to their plans (Hallet, 2007). One recent study of 11 schools in three local authorities in the UK identified six ways in which the distribution of responsibility for leadership and management was conceived and effected. These styles ranged from the formal designation of official top-down hierarchy positions to a more bottom-up opportunistic taking on of leadership and management functions by those without formally designated positions, including students (MacBeath *et al.*, 2004). These patterns of distribution include: formal, pragmatic, strategic, incremental, opportunistic, and cultural distribution. Pragmatic distribution, for example, is frequently in reaction to external stimuli with headteachers asking the staff to take up responsibility in order to ease the burden and to spread the workload (MacBeath *et al.*, 2004).

Practice Aspect

Conceptual

While the leader plus aspect has achieved the lion's share of the attention from scholars and practitioners, who take a distributed perspective, the practice aspect is critical. The practice of leadership and management has not been a major pursuit in research (Hallinger and Heck, 1996). Some readers may find this claim overdrawn, pointing to a larger literature on leadership styles and approaches that characterize collections of practices. While helpful, this literature extracts practice from its temporal context and in so doing ignores the urgency of practice (Bourdieu, 1981: 310). A focus on practice necessitates attention to the immediacy and improvised nature of practice.

A distributed perspective frames practice in a particular way – as a product of the interactions of school leaders, followers, and aspects of their situation. This framing departs from the rather narrow psychological view that equates practice with the actions of individuals. From a distributed perspective, leading and managing practice is stretched over the work of two or more leaders and followers. Interactions are key to understanding practice. Further, aspects of the situation including tools, artifacts, organizational routines, and language (among others) define practice by enabling and constraining interactions among leaders and followers. While scholars have long recognized that the situation is a critical consideration in studying leadership and management (Murphy, 1991; Hallinger and Murphy, 1987a; Bossert *et al.*, 1982; Rosenholtz, 1989), from a distributed perspective, aspects of the situation do not simply affect what school leaders do or moderate the impact of what they do. Rather, the situation is one of the three core constituting elements of practice. At the same time, aspects of the situation are a product of practice (Spillane, 2006; Spillane and Diamond, 2007). Viewed this way, the practice of leading and managing is an emergent phenomenon (Gronn, 2000).

Work on the practice aspect of a distributed perspective draws on various theoretical traditions, including socio-cultural activity theory, distributed cognition, and situated cognition. An emerging emphasis in psychology and related fields is recognizing how social context is an integral component for activity. Work in situated and distributed cognition argues that the study of human cognition must attend to purposeful activity in its natural habitat (Hutchins, 1995; Pea, 1993; Latour, 1987; Leont'ev, 1981). Cognition is not purely a function of mental capacity because it is enabled and constrained by the circumstances in which it happens (Resnick, 1991). This work shifts the unit of analysis from an exclusive focus on the individual's knowledge structure to the practice system (Greeno, 1998). Sociocultural activity theorists argue that language, number systems, theories of action, and interpretive schema provide mediational means that enable intelligent social activity (Wertsch, 1991; Vygotsky, 1978; Leont'ev, 1975, 1981; Rogoff *et al.*, 1995). These material and cultural artifacts form identifiable aspects of the sociocultural context as products of particular social and cultural situations (Cole and Engeström, 1993; Engeström and Middleton, 1998; Wertsch, 1991; Vygotsky, 1978).

Empirical

The empirical knowledge base is thin with respect to applying the practice aspect of a distributed perspective to investigations of school leadership and management. Based on earlier work on interdependencies between activities and resources (Malone and Crowston, 1994; March and Simon, 1958; Thompson, 1967), one line of work

centers on identifying the interdependencies among leaders and followers in interactions. One typology identifies three types of distribution – collaborated, collective, and coordinated (Spillane *et al.*, 2000, 2003b; Spillane, 2006). Collaborated distribution refers to practice that is stretched over the work of two or more people working in the same place and same time to perform the same organizational routine. Collective distribution refers to situations where people work separately in place and time but interdependently to perform an organizational routine or function. Coordinated distribution refers to organizational routines where the tasks or activities have to be performed in a particular sequence; in order for task b to be performed, task a must be performed first. Each distribution type involves different sorts of interdependencies that pose unique challenges for practice. A single organizational routine could involve more than one type of distribution.

Distinguishing between additive models of distributed leadership and distributed leadership as concertive action, Peter Gronn identifies three forms of concertive action – spontaneous collaboration, intuitive relations, and institutionalized practices (Gronn, 2003, 2004). Each of these forms represents a stage in an institutionalization process. Further, the people who make up the group or unit act conjointly in that they synchronize their actions by paying attention to their own goals, their peers' goals, and their sense of belonging to the group or unit. Spontaneous collaboration refers to impromptu collaborations that are frequently transient and motivated by particular challenges. Intuitive relations refer to those working relations that develop over time as two or more people work together and learn to trust and rely on one another. Finally, institutionalized practices are represented in new structures that enable two or more individuals to co-perform.

Recent work has also started to explore how aspects of the situation (e.g., tools, artifacts, and organizational routines) define the practice of leading and managing by framing and focusing the interactions among leaders and followers (Coldren, 2007; Sherer, 2007; Spillane and Diamond, 2007). Further, there have also been some efforts to explore the prevalence of co-performance of leadership and management work in schools. One recent study of 42 school principals in a mid-sized urban district in the US found that school principals co-performed almost half (47%) of administration and instruction as well as curriculum-related activities for which they had some responsibility. These school principals reported co-leading with just one other individual 63% of the time, while they reported co-leading with two or more individuals 37% of the time with classroom teachers, having no formal leadership designation, being the most frequent co-leaders (Spillane *et al.*, 2007). This work, focused exclusively on the school principal's workday, suggests that situations involving collaborated distribution are prevalent in schools.

Effects

Considering that efforts to develop a distributed perspective are relatively new, the thin empirical knowledge base is understandable. Moreover, over the past 5 or 6 years, a growing number of empirical studies have used a distributed perspective to frame some part of the work. Some researchers express concern about the effectiveness of a distributed perspective wanting evidence that supports strong causal inferences; however, such thinking is problematic. To begin with, a distributed perspective is a conceptual frame for studying leadership and management, not a well- or even moderately specified approach to leading and managing schools. Hence, what is most salient in trying to tease out the effects a distributed perspective is not that the work of leading and managing the schoolhouse is distributed, but how it is distributed. Theory-building and hypotheses-generating work that will generate evidence about how the work of leading and managing is distributed and how these arrangements vary across schools is an important precursor to hypotheses-testing research and any serious attempts at tackling questions of effectiveness. Further, understanding the factors that might account for these different arrangements is also important and some of the work reviewed earlier offers important hints.

Based on the available empirical evidence, we might examine the relationship between the prevalence of co-performance of leading and managing and different outcome variables. Similarly, we might examine whether the concentration of responsibility for leadership and management work in a few individuals versus the dispersion over many individuals is related to particular outcomes. It is important to remember, however, that the influence of these arrangements is likely to differ, depending on various factors such as where a school or school leadership team is in some developmental trajectory. An example of this hypothesis generating and testing work is Ken Leithwood and his colleagues' recent research in which they focus on the extent of conscious alignment across different sources of leadership for the performance of organizational functions (Leithwood *et al.*, 2007). They identify four different leadership distribution arrangements – planful alignment, spontaneous alignment, spontaneous misalignment, and anarchic misalignment. Based on these four conceptualizations of leadership distribution patterns, they articulate a series of hypotheses for testing. For example, they hypothesize that planful alignment will contribute more than other patterns to organizational productivity over time.

The Problem with Synonyms

Efforts to generate valid evidence about how leadership and management are distributed in schools will be thwarted

unless scholars are precise about what they mean and do not mean when they say they are studying distributed leadership or taking a distributed perspective. Collaborative leadership, democratic leadership, and co-leadership, among others, are sometimes used interchangeably with distributed leadership. We hope we have made it clear in the preceding pages that a distributed perspective is not synonymous with democratic or collaborative leadership.

By definition, collaborative leadership is distributed; however, a distributed perspective can be used to analyze leadership and management situations that are not necessarily collaborative. While some see a distributed perspective as implying agreement and collegial arrangements among those doing the leading and managing, this is problematic. A distributed perspective applies just as well to situations where leaders do not see eye to eye or get along with one another (Hallett, 2007). Hence, a distributed perspective applies to situations where leaders have different or contrary goals as easily as it does to situations where leaders are striving for the same goal.

Similarly, a distributed perspective on leadership allows for democratic leadership or autocratic leadership or something in between. Examined from a distributed perspective, the leadership and management in a school can be stretched over leaders but not necessarily democratically (Woods, 2004). While co-leadership involves distribution or responsibility among two or more leaders, a distributed perspective urges us to consider how leadership practice takes shape in the interaction of leaders, followers, and aspects of the situation. Finally, a distributed perspective differs from transformational leadership; it applies as easily to situations involving transactional leadership as it does to situations involving transformational leadership.

Conclusion

Distributed leadership is certainly popular with researchers, practitioners, and policymakers. However, popularity is a poor measure of usefulness. A key issue for any conceptual frame is whether it adds value to the field by enabling scholars to generate new insights and helping practitioners think about a familiar phenomenon in new ways. The growing number of studies that are framed from a distributed perspective suggests the frame has value in investigating school leadership and management; however, more empirical work is necessary. Armchair theorizing about distributed leadership is plentiful. While theorizing is necessary for solid empirical work, future theorizing would benefit from a firm grounding in empirical data.

Generating empirical knowledge about school leadership and management from a distributed perspective will also necessitate attention to methodological challenges. These challenges are especially acute with respect to using the practice aspect of a distributed perspective to

frame investigations of leadership and management. Currently, we rely mostly on ethnographic studies, structured observations, and video data collection methods to study the practice of leading and managing. These methods generate rich insights into practice, but they are limited by cost and small sample size that make generalization to a population of schools difficult. Annual surveys address some of these challenges but they face another problem with respect to capturing practice. Research in survey methodology shows that annual surveys often yield flawed estimates of behaviors as respondents cannot accurately remember whether or how often they engaged in a behavior (Tourangeau *et al.*, 2000). In responding to an annual survey, respondents must often consider many different episodes of behavior that may have occurred weeks or months in the past. As one might expect, surveys that are administered closer to when a behavior occurs (e.g., logs and diaries) are more accurate than surveys that are given further away from when a behavior occurs (e.g., annual surveys; Hilton, 1989; Lemmens *et al.*, 1992). Hence, developing logs and diaries for school leaders, both formally designated leaders and informal leaders, offer some promise. Moreover, these methods can be supplemented with more intensive observational, interview, and video methodologies.

Bibliography

- Bennett, N., Harvey, J. A., Wise, C., and Woods, P. A. (2003). *Desk Study Review of Distributed Leadership*. Nottingham, UK: National College for School Leadership.
- Bossert, S. T., Dwyer, D., Rowan, B., and Lee, G. V. (1982). The instructional management role of the principal. *Educational Administration Quarterly* 18(3), 34–63.
- Burch, P. (2007). School leadership practice and the school subject: The Baxter case. In Spillane, J. P. and Diamond, J. (eds.) *Distributed Leadership in Practice*, pp 129–145. New York: Teachers College Press.
- Camburn, E., Rowan, B., and Taylor, J. (2003). Distributed leadership in schools: The case of elementary schools adopting comprehensive school reform models. *Educational Evaluation and Policy Analysis* 25(4), 347–373.
- Coldren, A. (2007). Spanning the boundary between school leadership and classroom instruction at Hillside elementary school. In Spillane, J. P. and Diamond, J. (eds.) *Distributed Leadership in Practice*, pp 16–34. New York: Teachers College Press.
- Cole, M. and Engeström, Y. (1993). A cultural–historical approach to distributed cognition. In Salomon, G. (ed.) *Distributed Cognitions: Psychological and Educational Considerations*, pp 47–87. New York: Cambridge University Press.
- Copland, M. A. (2004). Leadership of inquiry: Building and sustaining capacity for school improvement. *Educational Evaluations and Policy Analysis* 25(4), 375–396.
- Engeström, Y. and Middleton, D. (eds.) (1998). *Cognition and Communication at Work*, 2nd edn. Cambridge: Cambridge University Press.
- Firestone, W. A. and Martinez, C. (2007). Districts, teacher leaders, and distributed leadership: Changing instructional practice. *Leadership and Policy in Schools* 6, 3–35.
- Frost, D. (2005). Resisting the juggernaut: Building capacity through teacher leadership in spite of it all. *Leading and Managing* 10(2), 83.

- Gronn, P. and Hamilton, A. (2004). A bit more life in leadership: Co-principalship as distributed leadership practice. *Leadership and Policy in Schools* 3(1), 3–35.
- Gronn, P. (2003). *The New Work of Educational Leaders: Changing Leadership Practice in an Era of School reform*. London: Paul Chapman Press.
- Gronn, P. (2000). Distributed properties: A new architecture for leadership. *Educational Management and Administration* 28(3), 317–338.
- Hallett, T. (2007). The leadership struggle: The case of costen elementary school. In Spillane, J. P. and Diamond, J. (eds.) *Distributed Leadership in Practice*, pp 85–105. New York: Teachers College Press.
- Hallinger, P. and Heck, R. H. (1996). Reassessing the principal's role in school effectiveness: A review of the empirical research. *Educational Administration Quarterly* 32(1), 27–31.
- Hammersley-Fletcher, L. and Brundrett, M. (2005). Leaders on leadership: The impressions of primary school head teachers and subject leaders. *School Leadership and Management* 25(1), 59–75.
- Hargreaves, A. and Fink, D. (2003). The seven principles of sustainable leadership. *Educational Leadership*.
- Harris, A. (2002). Effective leadership in schools facing challenging contexts. *School Leadership and Management* 22(1), 15–26.
- Harris, A. (2005). Distributed leadership. In Davies, B. (ed.) *The Essentials of School Leadership*. London: Paul Chapman Press.
- Heller, M. F. and Firestone, W. A. (1995). Who's in charge here? Sources of leadership for change in eight schools. *Elementary School Journal* 96(1), 65–86.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge: MIT Press.
- Latour, B. (1987). *Science in Action: How to Follow Engineers and Scientists through Society*. Milton Keynes: Open University Press.
- Lemmens, P., Tan, E., and Knibbe, R. (1992). Measuring quantity and frequency of drinking in a general population survey: A comparison of five indices. *Journal of Studies on Alcohol* 53, 476–486.
- Leithwood, K., Mascal, B., Strauss, T., et al. (2007). Distributing leadership to make schools smarter: taking the ego out of the system. *Leadership and Policy in Schools* 6, 37–67.
- Leont'ev, A. N. (1975). *Activity, Consciousness and Personality*. (Hall, M. J. Trans.). Englewood Cliffs, NJ: Prentice-Hall.
- Leont'ev, A. N. (1981). *Problems of the Development of the Mind*. Moscow: Progress.
- MacBeath, J. and McGlynn, A. (2003). *Self Evaluation: What's in It for Schools?* London: Routledge.
- MacBeath, J., Oduro, G., and Waterhouse, J. (2004). *Distributed Leadership in Schools*. Nottingham: National College of School Leadership.
- Malone, T. W. and Crowston, K. (1994). The interdisciplinary study of coordination. *Computing Surveys* 26, 87–119.
- March, J. G. and Simon, H. A. (1958). *Organizations*. New York: Wiley.
- Murphy, J. (1991). *Restructuring Schools: Capturing and Assessing the Phenomena*. New York: Teachers College Press.
- Murphy, J. and Hallinger, P. (eds.) (1987). *Approaches to Administrative Training in Education*. Albany, NY: SUNY Press.
- Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In Salomon, G. (ed.) *Distributed Cognition: Psychological and Educational Considerations*, pp 47–87. New York: Cambridge University Press.
- Portin, B., Schneidre, P., DeArmond, M., and Gundlach, L. (2003). *Making Sense of Leading Schools: A Study of the School Principalship*. Seattle, WA: Center for Reinventing Public Education, University of Washington.
- Rogoff, B. M. (1990). *Apprenticeship in Thinking: Cognitive Development in Social Context*. New York: Oxford University Press.
- Rosenholtz, S. J. (1989). *Teachers' Workplace: The Social Organization of Schools*. New York: Longman.
- Sherer, J. (in press). The practice of leadership in mathematics and language arts: The Adams case. In *Distributed Leadership in Practice*. New York: Teachers College Press.
- Smylie, M. A. and Denny, J. W. (1990). Teacher leadership: Tensions and ambiguities in organizational perspective. *Educational Administration Quarterly* 26(3), 235–259.
- Spillane, J. P. (2006). *Distributed Leadership*. San Francisco, CA: Jossey-Bass.
- Spillane, J. P., Camburn, E., and Pareja, A. (2007). Taking a distributed perspective to the school principal's workday. *Leadership and Policy in Schools* 6, 103–125.
- Spillane, J. P. and Diamond, J. (eds.) (2007). *Distributed Leadership in Practice*. New York: Teachers College Press.
- Spillane, J. P., Diamond, J. B., and Jita, L. (2000). Leading instruction: The distribution of leadership for instruction. *Paper Presented at the Annual Meeting of the American Education Research Association*. New Orleans, LA.
- Spillane, J., Diamond, J., and Jita, L. (2003a). Leading instruction: The distribution of leadership for instruction. *Journal of Curriculum Studies* 35(5), 533–543.
- Spillane, J., Diamond, J., Sherer, J., and Coldren, A. (2004a). Distributing leadership. In Coles, M. and Southworth, G. (eds.) *Developing Leadership: Creating the Schools of Tomorrow*, pp 37–49. New York: Open University Press.
- Spillane, J. P., Hallett, T., and Diamond, J. B. (2003b). Forms of capital and the construction of leadership: Instructional leadership in urban elementary schools. *Sociology of Education* 76(1), 1–17.
- Spillane, J. P., Halverson, R., and Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. *Educational Researcher* 30(3), 23–28.
- Spillane, J. P., Halverson, R., and Diamond, J. B. (2004). Towards a theory of leadership practice: A distributed perspective. *Journal of Curriculum Studies* 36(1), 3–34.
- Storey, J. (2004). Changing theories of leadership and leadership development. In Storey, J. (ed.) *Leadership in Organizations: Current Issues and Key Trends*, pp 11–38. London, UK: Routledge.
- Thompson, J. D. (1967). *Organizations in Action: Social Science Bases of Administrative Theory*. New York: McGraw-Hill.
- Tourangeau, R., Rips, L. J., and Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.
- Vygotsky, L. S. (1978). *Mind and Society: The Development of Higher Mental Processes*. Cambridge: Harvard University Press.
- Wertsch, J. V. (1991). *Voices of the Mind: A Sociocultural Approach to Mediated Action*. Cambridge: Harvard University Press.
- Woods, P. A. (2004). Democratic leadership: Drawing distinctions with distributed leadership. *International Journal of Leadership in Education* 7(1), 3–26.
- Yukl, G. (1999). An evaluation of conceptual weaknesses in transformational and charismatic leadership theories. *Leadership Quarterly* 10(2), 285–305.

Leadership: Authentic

P T Begley, Pennsylvania State University, University Park, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Authentic leadership is a metaphor for professionally effective, ethically sound, and consciously reflective practices in educational administration. It alludes to a knowledge-based, values-informed, and skillfully executed leadership. Leadership by definition refers to practices that extend beyond the usual procedural context of organizational management. Authentic leadership implies a genuine kind of leadership – a hopeful, open-ended, visionary and creative response to social circumstances, as opposed to the more traditional dualistic portrayal of management and leadership practices characteristic of now-obsolete research literature on effective principal practices (Begley, 2001). Moreover, in recent years, management has been negatively portrayed as mechanistic, relatively shortsighted, and a precedent-focused enterprise. An integrated image of leadership and management more in keeping with current times is a values-informed leadership – a form of leadership that acknowledges and accommodates, in an integrative way, the legitimate needs of individuals, groups, organizations, communities, and cultures – not just the organizational perspectives that are the usual focus of most leadership literature. It is a perspective that has, in recent years, been explored by several other scholars, including Taylor (1991), Duignan and Bhindi (1997), and Starratt (2004). The innovative dimension being proposed here is the adoption and application of a values and valuation process perspective to educational leadership to make the objectives of administration more understandable, compelling, and achievable. It is in this context that authentic leadership is proposed as the outcome of self-knowledge, sensitivity to the orientations of others, and a technical sophistication that leads to a synergy of leadership action (Begley, 2001, 2003).

Relationship of Valuation Processes to Authentic Leadership

There are essentially three ways in which valuation processes relate to authentic leadership. The first is as an influence on the cognitive processes of individuals and groups of individuals. Understanding how values reflect underlying human motivations and shape subsequent attitudes, speech, and actions is essential knowledge for any person in a leadership role. Leaders should know their own values and ethical predispositions, as well as be sensitive to the value orientations of others. The second way in which valuation processes relate to leadership practices is as a guide to action, particularly as supports

to resolving ethical dilemmas. Ethics are highly relevant to school leadership as rubrics, benchmarks, socially justified standards of practice, and templates for moral action. The third way in which valuation processes relate to leadership is as a strategic tool that leaders can employ to build consensus among the members of a group toward the achievement of shared organizational objectives. In this sense, leaders literally use ethics as a leadership tool in support of actions taken. Authentic leadership is therefore grounded in the understanding or interpretation of observed or experienced valuation processes as well as in ethical decision-making processes.

Research on principal valuation processes (Begley and Johansson, 1998) and problem-solving processes (Leithwood and Steinbach, 1995) demonstrates that administrators tend to employ ethics as a guide to action under certain conditions: in situations of high stakes and urgency, when consensus is impossible, when responding to unprecedented situations, and for certain hot-button social issues which tend to quickly escalate debate to a point where people seek refuge within an ethical posture. Ethics may be consciously or unconsciously employed in such situations. A typical application for ethics is as a personal guide to action, particularly as a support to resolving ethical dilemmas. However, there are more strategic and collective applications for ethics. One of the most common in a school or school district setting is using an ethic as a focus for building consensus around a shared social objective. These collective and strategic applications of ethics may be the more common manifestation of this value type in the administration of schools and school districts, as well as in the corporate sector. However, this is yet to be confirmed empirically by any systematic research beyond that conducted by Langlois (2004).

The Valuation Processes of Individuals

In order to understand the relationship between motivation and values, and between values and authentic leadership actions, it is helpful to conceptually situate values within the context of one person's being using a simple onion figure. **Figure 1** (Begley, 2006) is an adaptation of a graphic originally proposed by Hodgkinson (1978, 1991, 1996).

Beginning from the outside, the first ring represents the observable actions and speech of the individual. Leaders and people in general intuitively rely on the clues provided by the actions and attitudes of others to derive

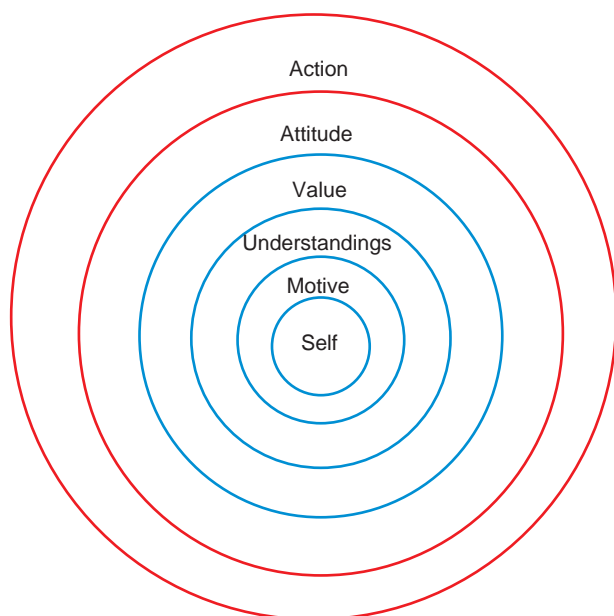


Figure 1 Values syntax. From Begley, P. T. (2006). Self-knowledge, capacity and sensitivity: Prerequisites to authentic leadership by school principals. *Journal of Educational Administration* 44, 570–589.

predictive insights into the nature of the values others hold. This is a generally sound strategy, but has the same limits to its reliability in day-to-day life as it does in a research context. Political leaders, principals, teachers, parents, and children regularly demonstrate through their speech and actions that their observable actions may or may not be accurate indicators of their underlying values. Individuals often articulate or posture certain values while actually being committed to quite different ones. In both the research and the leadership context, the implication is clear. The validity and reliability of interpretation is best enhanced by sustained periods of observation and multiple measures.

The next ring or layer of the figure represents attitudes. Attitudes can be thought of as the membrane between values and the observable actions or speech of an individual, or the permeable boundary of personality that acts as the interface between the psychological and physical world. They are the predisposition to act specifically as a result of values or value systems acquired previously and elsewhere (Begley, 2001). For example, the attitudes of educators toward students encountered in their professional setting may change when they become parents, with young children of their own. Conversely, when we look across a career, we can see that the values of an individual in one role as a teacher, principal, or professor readily spill over as attitudes into other social roles. The strength of this extended influence can be residual in nature, a significant spillover of effect, or intrude to such an extent that it overrides or neutralizes the influence of a

second value or value system. Hodgkinson (1991) also suggests that attitudes can often be detected in the body language of posture, gait, or unconscious muscular tensions. They are outward and visible signs of inner and invisible inclinations.

The next layer represents the actual values held or manifested by an individual. For example, while one person might prefer a glass of beer to a glass of Shiraz, another might prefer a chat with friends in the staff lounge to reading the newspaper, or working independently over working in a group, or a monarchical system of government over a republican system. A school leader might gravitate toward relatively controlled approaches to delegating authority over more open styles of distributed leadership. A teacher might prefer computer-mediated instruction to workbook exercises, or instruction individualized to students' needs as opposed to a teacher-centered curriculum. The important thing to keep in mind is that identifying these values is one thing, while knowing why they are held is quite another. Making that latter judgment requires going deeper into the onion.

Between the values layer and motivational base layer of the figure is a category that can be labeled available knowledge or understandings. The kinds of knowledge referenced here are acquired through life experiences, professional training, and reflection, and provide a linkage and context between the motivational bases and the specific values adopted by the individual. The contention here is that, as a result of experience, training, and/or reflection, an individual responds to basic motivations by adopting particular value positions that will support the fulfillment of that basic motivation in a specific way. These responses are manifested through actions or speech selected by the individual to achieve the valued objective. People vary, of course, in terms of the skills and sophistication they can bring to bear on achieving their objectives. This is generally applicable to all aspects of human enterprise. Consider how an experienced school administrator, consensually motivated as a professional to achieve a complex set of educational objectives, might employ a carefully orchestrated collaborative school-improvement project to achieve these educational objectives. By contrast, a less-experienced administrator, with the same desire to build consensus among the faulty, but responding to different knowledge or the absence thereof, might decide a memo is all that is required to achieve the same objective.

The motivational base layer of the onion figure provides the key to understanding the nature and function of values as influences on leadership. This is the motivating force dimension behind the adoption of a particular value which, working out through the layers of the figure, shapes attitudes and potentially influences subsequent actions. Hodgkinson (1978, 1991, 1996), proposes that there are four basic motivational bases: personal preference

or self-interest, an inclination toward consensus, an inclination toward or concern for consequences, and an inclination toward transrational ethics or principles. These four motivational bases are relatively broad and arbitrary distinctions. In application, individuals can manifest a predisposition toward one motivational base over another, or adopt more than one motivational base when responding to a given situation. Research conducted in several countries on the valuation processes of school administrators (Begley and Johansson, 1998) suggests that the normative motivational bases for administrative decision making are the rational domains of consequences and consensus. Self-interest is infrequently acknowledged as a motivation, predictably because professional activity is publicly accountable. Ethics and principles tend to be employed relatively infrequently as motivational influences on the cognitive processing of individuals. Leaders do, however, regularly employ ethics as strategic supports to a collective leadership process. The distinction being made here is between processes where leaders use ethics for strategic purposes as opposed to being ethical.

The final layer at the center of the figure is the self—the biological self as well as the existential or transcendent self. The formation of the self and its relevance to authentic leadership is something which also warrants some attention.

Arenas of Leadership: Sources of Influence, Values, and Conflicts

The external influences on leadership and education in general can be thought of as coming from multiple social sources. Some of these influences can take on the status of values when they are perceived as conceptions of the desirable with motivating force (Hodgkinson, 1991). Unfortunately, our personal values as well as those of the profession, organization, community, and society are not necessarily consistent or compatible with each other. As a result, these influences and values derived from the various arenas of our environment can generate inconsistencies and conflicts. A second onion figure (see **Figure 2**) is used to illustrate these distinctions. These are the interactive environments within which valuation processes and administration occur. They are also the source of personal, professional, and social values, as well as the source of many of the conflicts people encounter in life.

Within the figure, the individual is represented within the center ring and extending through all the rings. His or her character is the outcome of many transient influences as well as relatively more enduring values acquired from multiple arenas.

The second ring from the center represents the arena of groups and other collective entities including family, peers, friends, and acquaintances. The third ring, profession, represents a more formal arena of administration that is

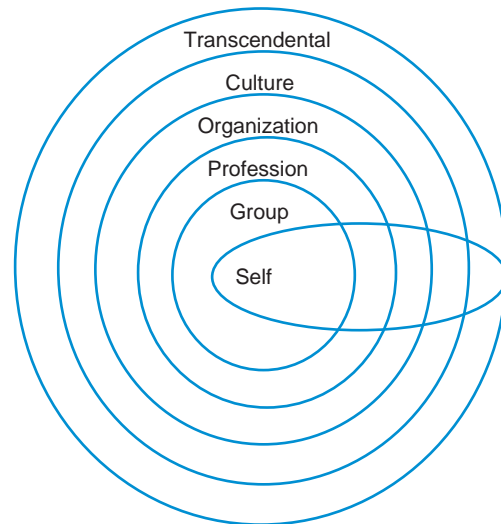


Figure 2 Arenas of influence. From Begley, P. T. (2006). Self-knowledge, capacity and sensitivity: Prerequisites to authentic leadership by school principals. *Journal of Educational Administration* 44, 570–589.

closely related to the second ring, but is given special emphasis here because of its relevance to the professional context that is the focus of this article.

The fourth ring, the organization, represents the arena that is traditionally of most concern to academics and practitioners in the field of educational administration. Much of the literature of educational administration and most of the corporate literature are grounded within the organizational perspective, adopting it as a primary reference point for administrative activity.

Moving further outward in the figure, one encounters the arena representing the greater community, society, and culture. Within recent decades, school administrators have learned that it is necessary to pay a lot more attention to the community as a relevant administrative arena and source of influence on school leadership (Leithwood *et al.*, 1992). The increasing diversity of our societies and a general trend toward globalization have highlighted society and culture as relevant arenas of administrative activity.

A final, sixth ring is included to accommodate notions of the transcendental – God, faith, spirituality, and even extra-sensory perception. Spirituality is of considerable importance to many individuals and has begun to attract the attention of more scholars as an important influence on educational leadership. Even administrators who do not subscribe to a spiritual dimension as a source of influence on their own daily lives are well advised to keep this arena in mind, even if only because others associated with their professional roles do so. A leader who wants to understand the motivations of those he/she is supposed to lead will be sensitive to all potentially significant categories of influence.

Complications

Having made the case for authentic leadership through the adoption of a values perspective on problem-solving and decision-making processes, and having outlined the conceptual framework underpinning these assertions, the rest of the article is devoted to the consideration of several specific complications that emerge as implications.

Inauthentic or False Leadership

As already stated, research findings (e.g., Begley and Johansson, 1998) confirm that the relevance of principles or ethics to a given administrative situation seems to be prompted in the minds of school administrators by particular circumstances. These circumstances include: situations where an ethical posture is socially appropriate (e.g., the role of the arts); situations where consensus is perceived as difficult or impossible to achieve (e.g., an issue involving ethnic bias); or situations when high stakes and urgency require decisive action (e.g., student safety). There is also some evidence to suggest that school leaders use ethics in strategic applications as ways to develop group consensus, and as a basis for promoting compliance with minimum need for evidence (Langlois, 2004). These are all examples of authentic leadership and ethically sound applications of ethics to leadership situations. However, one only has to survey the newspaper, or work in any organization, or live in any community for a time, to readily detect situations where ethic-based postures can be unethical or socially unjust. For example, ethical postures may be unethical in application when they are used to impose a culturally based ethic on people from another culture. It is equally unethical to employ ethics as a justification for otherwise reprehensible action (e.g., flying passenger jets into skyscrapers). The common political or organizational tactic of adopting an ethical posture to veil a less-defensible value is also a violation of ethics. In addition, fundamentally, it is unethical to employ a culturally derived ethic as the justification for trumping basic human rights. The implication is that using ethics is not always ethical. Such is the nature of ethics when they are adopted as guides to action. Transrational values (Hodgkinson, 1978, 1991, 1996) of any sort, and ethics and principles in particular, are rather vulnerable to multiple interpretations in application from one social context to another. In educational leadership situations, the application of an ethical posture in arbitrary ways (e.g., democratic process or the will of the majority) can often produce outcomes that are far from ethical (or democratic). The essential, and often absent, component that would make adherence to a value genuinely ethical is dialog.

Cross-Cultural Issues

Distinctive, unique, or minority-based social conditions may be obscured, veiled, or blurred by the perspectives

and language adopted to describe social processes. In many respects, this is a natural outcome and limitation of language as a means of assigning meaning to concepts and events, or the bounded rationality that occurs when models and frameworks are applied to complex social situations. This is also an outcome of general human inclinations to generalize the specifics of one context to the point that they become automated as a cognitive schema or a set of abstract principles (Begley, 1996). Similarly, a number of scholars, notably Allan Walker and Clive Dimmock, believe that the field of educational administration has developed along ethnocentric lines, dominated by Western perspectives emanating mostly from the United States and United Kingdom (Dimmock and Walker, 1998; Walker and Dimmock, 1999; Walker, 2003). The consequences are a risk that the generalized experiences of one country may be inappropriately assumed to be instructive to practices in radically different contexts. As societies become more globalized, and as the exchange of information among international scholars becomes more widespread, the implications become more urgent. Many administrators are discovering that some of their society's most cherished ethical foundations sometimes must be carefully re-examined in terms of how they are interpreted as well as their appropriateness to social circumstances. As our communities and societies become more diversified, school administrators must become more sophisticated in their leadership and more sensitive to the value orientations of others. For their part, researchers must move beyond their traditional orientation toward generalization and description to also consider the deeper matters of intent and motivational base. Once again, what emerges as the critical implication is the need for dialog and negotiation of meaning among stakeholders in professional educational settings.

Conclusion

It is not enough for school leaders to merely emulate the values of other principals currently viewed as experts. Leaders in schools must become reflective and authentic in their leadership practices. There is no reliable catalog of correct values that school leaders can adopt as solutions for the dilemmas of administration. School leadership situations are much too context bound to permit this kind of quick fix. School leaders need to be reflective practitioners. The first step toward achieving this state is, predictably enough, to engage in personal reflection (Coombs, 2004). The adoption of a values perspective on school leadership can transform this rather idealistic advice into something specific enough for school administrators to act upon. However, once a degree of improved self-knowledge has been achieved through personal reflection, administrators must then take the next step

toward authentic leadership. That is, they must strive to develop sensitivity to the value orientations of others in order to give meaning to the actions of the students, teachers, parents, and community members with whom they interact. This is both a personal quest for the leaders as well as a collective one for the organizational unit for which the leader is responsible. The authentic leader needs to work strategically to engage the school community in a collective and ongoing dialog on the dilemmas of professional practice and the social problems of the community. Authentic leadership occurs when understanding the value orientations of others provides leaders and the professionals as well as the community members to whom they are accountable with information on how they might best influence the practices of others toward the achievement of broadly justifiable social objectives.

Acknowledgment

Portions of this article are based on material taken from a previously published journal article, Begley, P. T. (2006). Self-knowledge, capacity and sensitivity: Prerequisites to authentic leadership by school principals. *Journal of Educational Administration* 44, 570–589, with the permission of the publishers, Emerald Press, as part of their standing agreement with authors.

See also: Ethical/Moral Issues in Educational Leadership; History of Educational Leadership/Management; Leadership in Diverse Cultures; Principals Role in Restructuring Schools; Research on Educational Leadership – Approaches/Promising Directions; School Community Relations; Transformational School Leadership.

Bibliography

- Begley, P. T. (1996). Cognitive perspectives on values in administration: A quest for coherence and relevance. *Educational Administration Quarterly* 32, 403–426.
- Begley, P. T. (2001). In pursuit of authentic school leadership practices. *International Journal of Leadership in Education* 4, 353–366.
- Begley, P. T. (2003). In pursuit of authentic school leadership practices. In Begley, P. T. and Johansson, O. (eds.) *The Ethical Dimensions of School Leadership*, pp 1–12. Dordrecht: Kluwer.
- Begley, P. T. (2006). Self-knowledge, capacity and sensitivity: Prerequisites to authentic leadership by school principals. *Journal of Educational Administration* 44, 570–589.
- Begley, P. T. and Johansson, O. (1998). The values of school administration: Preferences, ethics and conflicts. *Journal of School Leadership* 8, 399–422.
- Coombs, C. P. (2004). Developing reflective habits of mind. *Values and Ethics in Educational Administration* 1, 1–8.
- Dimmock, C. and Walker, A. (1998). Comparative educational administration: Developing a cross-cultural conceptual framework. *Educational Administration Quarterly* 34, 558–595.
- Duignan, P. and Bhindi, N. (1997). Authentic leadership: An emerging perspective. *Journal of Educational Administration* 35, 195–209.
- Hodgkinson, C. (1978). *Towards a Philosophy of Administration*. Oxford: Basil Blackwell.
- Hodgkinson, C. (1991). *Educational Leadership: The Moral Art*. Albany, NY: SUNY Press.
- Hodgkinson, C. (1996). *Administrative Philosophy*. Oxford: Elsevier-Pergamon.
- Langlois, L. (2004). Making the tough calls: Complex decision-making in light of ethical considerations. *International Studies in Educational Administration* 32, 78–93.
- Leithwood, K. A., Begley, P. T., and Cousins, J. B. (1992). *Developing Expert Leadership for Future Schools*. London: Falmer.
- Leithwood, K. A. and Steinbach, R. (1995). *Expert Problem Solving*. Albany, NY: SUNY Press.
- Starratt, R. J. (2004). *Ethical Leadership*. San Francisco, CA: Jossey-Bass.
- Taylor, C. (1991). *The Ethics of Authenticity*. Cambridge, MA: Harvard University Press.
- Walker, A. (2003). Developing cross-cultural perspectives on education and community. In Begley, P. T. and Johansson, O. (eds.) *The Ethical Dimensions of School Leadership*, pp 145–160. Dordrecht: Kluwer.
- Walker, A. and Dimmock, C. (1999). A cross-cultural approach to the study of educational leadership: An emerging framework. *Journal of School Leadership* 9, 321–334.
- Begley, P. T. and Johansson, O. (eds.) (2003). *The Ethical Dimensions of School Leadership*. Dordrecht: Kluwer.
- Branson, C. (2006). Effects of structured self-reflection on the development of authentic leadership practices among Queensland primary school principals. *Educational Management Administration and Leadership* 35, 227–248.
- Stefkovich, J. and Begley, P. T. (2007). Conceptualizing ethical school leadership and defining the best interests of students. *Educational Management and Leadership* 35, 205–226.

Further Reading

- Begley, P. T. and Johansson, O. (eds.) (2003). *The Ethical Dimensions of School Leadership*. Dordrecht: Kluwer.
- Branson, C. (2006). Effects of structured self-reflection on the development of authentic leadership practices among Queensland primary school principals. *Educational Management Administration and Leadership* 35, 227–248.
- Stefkovich, J. and Begley, P. T. (2007). Conceptualizing ethical school leadership and defining the best interests of students. *Educational Management and Leadership* 35, 205–226.

Relevant Website

<http://www.ed.psu.edu> – D. J. Willower Center for the Study of Leadership and Ethics of the UCEA.

Leadership: Democratic

J Møller, University of Oslo, Oslo, Norway

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Democratic school leadership is a relational concept which must be contextualized. It is situated in a democratic governance structure and policies that aim at educating students for citizenship in a democratic society. Depending on the cultural context and the people involved education for democracy can be framed differently, and the exercise of democratic leadership in practice varies greatly from school to school (cf. Blase *et al.*, 1995). Besides, schools have increasingly become sites of disputes with parents claiming the rights of their children to special education programs, to special treatments, to exceptions to school policies, etc. (Starratt, 2004). Some will emphasize democracy-creating, which involves building the conditions for and encouraging democratic participation. Others will highlight democracy-doing, which includes collective decision making, voting, and the initiative by all members in a community (Woods, 2005). Much of the disparity can be linked to the complexities of the social organization of a school and to the fact that images of school leadership must be analyzed in relation to historical distinction and sociocultural contexts (Møller and Schratz, 2008).

My approach to conceptualization is that democratic leadership is a moral activity which cannot be fully grasped without including a discussion of what we mean by an education based on democratic values. It implies that the primary responsibility of education is to create democratic citizens, a conviction that a more democratic and egalitarian organization of society is both possible and desirable, and that education can have an important role to play in attaining this kind of society (cf. Grace, 1995, Banks *et al.*, 2005). Becoming a knowledgeable and engaged citizen, who is responsive to diversity within and across nations, is a process, and education should facilitate the development of students' consciousness and agency. It requires both knowledge about democracy and experiences of democracy through democratic leadership in practice. Neither is sufficient alone.

Education for Democracy

Scholars have theorized extensively about issues of democracy. Dewey's (1937) writings about lived democracy and his vision of a great community, Habermas' (1987) theory of communicative rationality and Arendt's (1958) vision of a

provisional community where the public space is identified with plurality have all figured prominently in these views.

In short, democracy means rule by the people, drawn from the Greek word *demos* which signifies people and *kratos* which implies rule. Democracy consists in two complementary ideals, one involving equal distribution of the power to make collective decisions, the other equal participation in collective judgment (Warren, 2002). In other words, it is not only equality of votes, but also equal and effective opportunity to participate in processes of collective judgments. Many consider deliberation to be the ideal way of making collective judgments because it requires reasons to be developed and justifications to be given and opinions to be cultivated. It seeks to provide space for the political deliberation of moral issues. It is a theory of the procedures that allow moral conflicts to be argued, and disputes over norms are dependent upon deliberation for valid resolutions (Warren, 2002).

Dewey extended the concept by saying that democracy is a way of living together as well as a kind of government. For Dewey (1937), educational institutions have an important role to play in helping the development of the individual's character. His view of democracy rests on a certain faith in the capacity of human nature. The development of truly democratic communities is restricted by our inability to critically examine currently held assumptions, and these habits are learned in schools. Therefore, it is only through considerable political learning that students can recognize their potential and become citizens who are capable of deliberating about what they are doing, set goals and priorities for themselves, and work for the common good. The best way to teach and learn democracy, Dewey says, is to practice it.

In education, critical theorists such as Apple (2000) and Bates (1989) have analyzed the relationship between society and education within the conception of a critical democracy, focusing on who is included within a community, and who is left out. They underscore that a community is always constituted within specific historical conditions and against a background of political interests. According to Apple and Beane (1999), the central concerns of democratic schools include the open flow of ideas that enables people to be as fully informed as possible; a faith in the individual and collective capacity of people for resolving problems; use of critical reflection to evaluate ideas, problems, and policies; concern for the welfare of others and the common good; and concern for the dignity and rights of individuals and minorities.

In a similar way, Furman and Starratt (2002) underscore that establishing a forum for democratic participation is not sufficient. In addition, democracy “requires the *ability* to listen, understand, empathize, negotiate, speak, debate and resolve conflicts in a spirit of interdependence and working for the common good” (p. 118). In other words, education for democracy implies engaging the students in reasoned deliberations. They argue for a new understanding of the idea of democratic community. This understanding should be based on the acceptance and celebration of difference rather than the nostalgic striving for homogeneity, and include key concepts like interdependence and the common good. Moreover, they make a claim about a strong moral sense as a basic component of a democratic community that values sociality itself, a reverence for open inquiry and critique, a respect for individuals, and the interdependence of all. They emphasize the need for a school curriculum that is compatible with these moral dimensions.

In sum, the following themes are common to many definitions of education for democracy: (1) recognizing the basic value and rights of each individual; (2) taking the standpoint of others into consideration; (3) deliberation in making decisions; (4) embracing plurality and difference; and (5) promoting equity and social justice.

Conceptualizing School Leadership

Not only education for democracy is contested, but also our understanding of leadership varies widely. A line may be drawn between those who frame principalship as the key to school development (Hallinger and Heck, 1996; Leithwood *et al.*, 1990), and those who emphasize leadership as taking place in the interactions between people and their situations (Foster, 1986; Hargreaves and Fink, 2006). A relational perspective implies that school leadership is not necessarily synonymous with a particular position; it may come from school principals, teachers, or others; in other words, it is distributed within the organization and stretches over the school’s social and situational contexts (Spillane *et al.*, 2004; Gronn, 2002).

The distributed perspective, which implies that the leadership capacity lies not only within individuals, but is also constituted by the institutional arrangements, culture, and relationships, is currently gaining terrain. However, even though many argue for the potential of distributed leadership, studies often demonstrate a focus on how the principals develop personal styles and courage relative to the perceived and actual pressures (Day, 2007; Leithwood *et al.*, 2006). Moreover, the model of system leadership, which is now emerging and is associated with distributed leadership as a concept, has paradoxically a strong individual and heroic focus. Hopkins (2006), a strong promoter of this model, underlines that system leaders are those head

teachers who are willing to shoulder the system-leadership roles and care about and work for the success of other schools as well as their own. It entails that school leaders should take greater responsibility for neighboring schools as well as their own, and the vision of shared leadership is turned into even higher individual expectations to head teachers.

Critical theorists such as Anderson (1996), Bates (1989), Blackmore (1996), Foster (1986), and others have argued that the dominant traditions of theory and practice in educational administration often serve to justify uncritical patterns of organizations and control in schools that both mirror and reinforce the dominant patterns of inequality in the wider society. For instance, Foster has claimed that

... the concept of leadership has been chewed up and swallowed down by the needs of modern managerial theory. The idea of leadership as a transforming practice, as an empowerment of followers, and as a vehicle for social change has been taken, adapted and co-opted by managerial writers so that now leadership appears as a way of improving organizations, not of transforming our world. (Foster, 1989: 45)

This does not mean that leadership and management are separate, however. If management is viewed as maintaining the school organization, and leadership is seen as developing the school, in practice, they overlap (Harris *et al.*, 2007).

Comparing Democratic Leadership with Distributed Leadership

Distributed and democratic leadership have much in common, but the two concepts also provide different frameworks on leadership practice (cf. Harris *et al.*, 2007). A main distinction can be drawn between a normative and a descriptive way of framing leadership. Democratic leadership implies that one of the main responsibilities of school leaders is to build educational institutions around central democratic values such as promoting equity and social justice in the school as well as in the wider community. It emphasizes that the underlying principles and values of a democratic approach need embedding in both management and leadership practices. Both need to reinforce and uphold a focus on social justice and a concern for the welfare of others, including the dignity and rights of minorities and individuals. Putting empowerment as a central principle involves balancing power and trust whether in the area of leadership or management. It means creating a stimulating learning environment where pupils flourish and develop as citizens.

While democratic leadership requires serious attention to the value base of leadership practice and the processes that create and sustain social justice, empowerment, and

community, distributed leadership is a descriptive way of conceptualizing the phenomenon or a conceptual framework for thinking about the work of leadership. It is a way of conceptualizing how leadership practice takes shape in the interactions between leaders and followers in their situation. As such, it is an analytical perspective which provides a lens for generating knowledge about how leadership is practiced. Similar to democratic leadership, it is about dispersed activity in which initiative circulates widely. However, while distributed leadership can be understood as a helpful descriptive or analytical conception that has a potential of highlighting some ideas across different societies, the value base is not necessarily linked to social justice and an education based on democratic values as is the case with democratic leadership. As Woods (2005: 23) has emphasized, democratic leadership grows from a concern with philosophical, political, and sociological questions that surface with the idea of opening the boundaries of leadership, and translating into practice the ideals which form an integral part of democratic rationalities.

Distributed leadership, as developed by Gronn (2002) and Spillane *et al.* (2004) is helpful as an analytical perspective, but it becomes problematic when it is used a normative concept simply as another way of referring to democratic leadership. Figure 1 demonstrates how distributed and democratic leadership represents different frameworks and also how they overlap and have certain concepts in common (cf. Harris *et al.*, 2007).

The two lenses both focus on leadership as interaction and communicative action and are closely linked to a sociocultural perspective on learning, but the value base is explicitly connected to social justice and equity within a framework of democratic leadership.

Empirical Studies of Democratic School Leadership

Statements about democratic missions for schools have been found in policy statements in most countries since the twentieth century, but it is an empirical question to what degree these missions influence schooling. How do such democratic missions unfold in practice? Empirical studies in support of democratic school leadership tend to

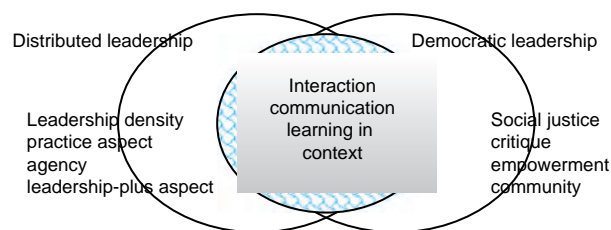


Figure 1 The relationship between distributed leadership and democratic leadership.

be case studies, and such approaches to research will not produce a universal list of what characterizes a democratic leader. However, it may enhance a better understanding of what goes on in schools and provide examples of how the conditions on which a democracy depends might be established in different contexts. For this purpose, a few studies have been selected.

Such examples are, for instance, given in the book *Democratic Principals in Action* (Blase *et al.*, 1995), where the authors have presented a series of portraits of eight principals, who are struggling to move toward more democratic forms of school administration. When the book was published all the school principals were members of the League of Professional Schools in Georgia. This league's purpose was to establish democratic decision-making structures to promote teacher involvement in school-wide instructional and curricular decisions (Blase *et al.*, 1995). The stories of the eight principals did not represent perfect role models. On the contrary, these principals were not perfect, and shortcomings and confusions were apparent among them all. Their stories do however show that democratic leadership can be framed differently, depending on the people involved and the context, and that they were struggling daily to fulfil their vision. However, they shared common orientations:

- They all tried to encourage teachers' involvement in decision making about instruction and were committed to the principle of sharing power with others.
- They were all child centred and strongly committed toward improving teaching and learning and supporting teachers.
- They all had trust in teachers' motives.
- They all had the ability to listen and to communicate openly.

Their personal satisfaction was linked to seeing growth in teachers and programs. They talked about the pride they felt because of what they had achieved from working together. However, the relationships with school boards were problematic. The principals were caught between traditional central-office expectations and teachers' new expectations for participation in decision making. Transforming organizational and cultural structures required a great deal of time over the long term. There was no quick fix.

Several studies have been carried out on the effects of the League of Professional Schools in Georgia, and a summary is provided by Allen and Glickman (2005). In their review of these studies they emphasized that although the studies showed that leadership did not lie only with the individuals, the principal's leadership seemed crucial for the school's development. For instance, interviews with all veteran League principals indicated that these principals had stopped trying to be fixers, and instead they enabled other's efforts to reach solutions to problems. In addition,

they modeled involvement in curriculum, staff development, and student learning. They exhibit trust and respect for teachers and encouraged involvement from a wide range of people. Interviews with teachers reinforced this picture. A survey of principals and teachers in 46 League schools also confirmed that implementing the League's framework had resulted in improvement of student learning and faculty learning.

The task ahead is not just about formulating a vision for democracy and inclusiveness. This is illustrated by a case study based on observations in two Norwegian schools over a period of 9 months. The study explored the negotiation of meaning and manifestations of leadership working for democracy in a multiethnic environment (Vedøy and Møller, 2007). Both schools had a strong commitment to comprehensive education and social justice, inspired by social democratic politics for promoting equity. However, the analysis demonstrated clear differences in creating challenging learning environments for all students, and these differences were closely linked to the leadership approach by the principal at the local school. Through an implicit discourse of pluralism based on goodness and caring, and with a focus on deficits, the principal at one of the schools actually suppressed democratic processes to the development of diversity in the school. The other school provided an example of how the principal, through an explicit discourse of critical multiculturalism based on respect, opened up for democratic processes and learning. Based on the findings, it is argued that the principal plays a pivotal role for including all stakeholders in the work for democratic schooling.

These findings echo Shields' (2006) analysis of how schools work for promoting social justice. Her study explores some of the challenges experienced by both practitioners and researchers attempting to work within political frameworks in which educational discourse is dominated by discourses of accountability, high-stakes testing, and a new standard of scientific research. Her analysis is drawn from interviews with 12 practicing educators in three schools. All three principals had a reputation for promoting social justice. In her investigation she uses three different story lines in which principals and teachers describe some of the challenges presented to them in their roles and how they have addressed them in practice. Her analysis clearly shows how the different story lines used by the participants in the study demonstrate differences in ways in which principals and teachers both incorporate theoretical perspectives and navigate dominant discourses of high-stakes accountability and scientific research. The first story focused on the deficit and neither the principal nor the teachers expected high levels of student achievement. As such, their deficit thinking masqueraded their talk about social justice. They had a negative focus on the students and their inability to improve their performance on test scores. In the second story the educators tried to make a difference

for the individual students, but the school did not seem to address the learning environment in a systemic way. Thus, the teachers remained isolated islands of success. The third narrative demonstrated that there is not necessarily a dichotomy between discourses of social justice and high academic achievement on the part of the students. This school could serve as an example of school-wide transformation and success. According to Shields, the knowledgeable principal was crucial for such an achievement. She had a clear sense of purpose related to how to create a learning environment in which all children may not only feel they belong, but also in which they may be successful. Moreover, she was able to engage her staff in extensive dialog about how to make meaningful change to make sure that the decisions were grounded in theories of social justice.

Other researchers have examined whether the means used to achieve the democratic purpose of schooling by the educational authorities were consistent with the end. For instance, Mulford (2004) explored this aspect in terms of principal training in Australia. In his analysis he concluded that different initiatives and strategies at national level seemed to reinforce this purpose, but this was not the case at state level in Tasmania. He has provided examples which show how it was expected that the school leaders should promote democracy in schools, but the principals were not themselves treated in a democratic way by their employer. In other words, many of those who were telling schools what to do were often unwilling to accept responsibility for their advice, blaming everybody else for lack of success. Common excuses were budget cuts or change of government.

Similar, an analysis of successful school leadership in Norway has shown that despite the democratic rhetoric in policy documents, it often seems difficult to make democratic dreams into reality. Competition and individualism are on the rise in the society and valued over cooperation and interdependence. Schools are expected to fulfil a mission which the wider society has dumped in practice, and learning is talked about as being part of an economic exchange between a provider (teacher) and a consumer (student). However, simultaneously, the study demonstrates that schools can still support activist professionals who counteract the tendency toward an economic way of thinking about education (Møller, 2006, 2007). The study showed how principals, in collaboration with teachers, continually worked to mediate government policy and external changes so that they could be integrated with the school's values about developing a more just society. Specific democratic values such as establishing open communication between staff and students; creating opportunities for student decision making and deliberation; a personal commitment to making a difference in students' lives; and establishing an ethic of care for individuals as well as a concern for the common good were enacted through the leadership at the school. As such, the study provides an example of

democratic school leadership in practice, with particular attention to the distribution of power and leadership in the school, student voice in the decision-making process, their opportunities for open dialogs, and the conditions that must be in place for students to develop as citizens.

Summary

Empirical studies can provide examples of how the conditions on which a democracy depends might be established, and serve as an inspiration for democratic leadership working for democracy in different contexts. However, the task ahead is not just about modeling and developing democratic practices within the school. It is also about challenging the wider power structures in which the school is embedded, and committing oneself to work for social change. Today, there are evidences of a widespread erosion of public trust in representative democracy in many countries, and the September 11 has recast the debate on multiculturalism and led to a review of what we understand by democracy (MacBeath, 2004). We know not only that poverty is closely linked to student achievement, but also that schools can make a difference in students' lives. As a consequence, teaching about and for democracy cannot be a theoretical notion. Teacher and student learning mirror one another, and democratic leadership includes a consideration of the ways in which external social structures are reproduced through the administration of schooling.

More research that helps us to understand leadership for socially just and democratic communities is needed. Democratic leadership is a slippery subject for research because it is so closely related to how we understand democracy in schools. Education based on democratic values is an ideal, a moral purpose toward which teachers and school leaders strive. It is about practice and processes, and it cannot be reduced to a tidy definition because it is necessary to focus on the links between democratic leadership practice and normative concepts of social justice and democratic schooling (Furman and Shields, 2005). It cannot be captured unless context is taken into account. Historical work can explain the past and anthropological work can examine culture, but there is also a need for understanding that what is happening at the school level is embedded and subsumed in wider social structures of power. What occurs in schools is necessarily contextualized. In addition, such research requires a careful examination of one's own beliefs and practices as researchers.

Bibliography

- Allen, L. and Glickman, C. D. (2005). Restructuring and renewal: Capturing the power of democracy. In Hargreaves, A. (ed.) *Extending Educational Change. International Handbook of Educational Change*, pp 225–249. New York: Springer.
- Anderson, G. (1996). The cultural politics of schools: Implications for leadership. In Leithwood, K., Chapman, J., Corson, D., Hallinger, P., and Hart, A. (eds.) *International Handbook of Educational Leadership and Administration*, pp 947–966. Dordrecht: Kluwer Academic Publishers.
- Apple, M. W. (2000). Between neoliberalism and neoconservatism: Education and conservatism in a global context. In Burbules, N. C. and Torres, C. A. (eds.) *Globalization and Education. Critical Perspectives*, pp 57–79. New York: Routledge.
- Apple, M. W. and Beane, J. A. (eds.) (1999). *Democratic Schools. Lessons from the Chalk Face*. Buckingham: Open University Press.
- Arendt, H. (1958). *The Human Condition*. Chicago, IL: The University Press.
- Banks, J. A., Banks, C. A. M., Cortes, C., et al. (2005). Democracy and diversity. Principles and concepts for educating citizens in a global age. *An Invitational Conference*. Center for Multicultural Education, University of Washington, Seattle.
- Bates, R. (1989). Leadership and the rationalization of society. In Smyth, J. (ed.) *Critical Perspectives on Educational Leadership*, pp 131–157. London: Falmer Press.
- Blackmore, J. (1996). 'Breaking the silence': Feminist contributions to educational administration and policy. In Leithwood, K., Chapman, J., Corson, D., Hallinger, P., and Hart, A. (eds.) *International Handbook of Educational Leadership and Administration*, book 2, pp 997–1043. Dordrecht: Kluwer Academic Publishers.
- Blase, J., Blase, J., Anderson, G., and Dungan, S. (1995). *Democratic Principal in Action. Eight Pioneers*. Thousand Oaks, CA: Corwin Press.
- Day, C. (2007). Sustaining success in challenging contexts. Leadership in English schools. In Day, C. and Leithwood, K. (eds.) *Successful Principal Leadership in Times of Change. An International Perspective*, pp 59–71. Dordrecht: Springer.
- Dewey, J. (1937). Democracy and educational administration. In Ratner, J. (ed.) *Education Today*, pp 345–346. New York: G.P. Putnam.
- Foster, W. (1986). *Paradigms and Promises: New Approaches to Educational Administration*. Buffalo, NY: Prometheus Books.
- Foster, W. (1989). Toward a critical practice of leadership. In Smyth, J. (ed.) *Critical Perspectives on Educational Leadership*, pp 39–63. London: Falmer Press.
- Furman, G. and Shields, C. (2005). How can leaders promote and support social justice and democratic community in schools? In Firestone, W. A. and Riehl, C. (eds.) *A New Agenda for Research in Educational Leadership*, pp 119–138. New York: Teacher College Press.
- Furman, G. and Starratt, R. J. (2002). Leadership for democratic community in schools. In Murphy, J. (ed.) *The Educational Leadership Challenge: Redefining Leadership for the 21st Century*, pp 105–134. Chicago, IL: The University of Chicago Press/NSSE.
- Grace, G. (1995). *School Leadership: Beyond Education Management: An Essay in Policy Scholarship*. London: Falmer Press.
- Gronn, P. (2002). Distributed leadership as a unit of analysis. *Leadership Quarterly* **13**, 423–451.
- Habermas, J. (1987). *The Theory of Communicative Action*. Cambridge: Polity Press.
- Hallinger, P. and Heck, R. (1996). The principal's role in school effectiveness: An assessment of methodological progress, 1980–1995. In Leithwood, K., Chapman, J., Corson, D., Hallinger, P., and Hart, A. (eds.) *International Handbook of Educational Leadership and Administration*, bk. 2, pp 723–784. Dordrecht: Kluwer Academic Publishers.
- Hargreaves, A. and Fink, D. (2006). *Sustainable Leadership*. San Francisco, CA: Jossey-Bass/Wiley.
- Harris, A., Moos, L., Möller, J., Robertson, J., and Spillane, J. (2007). Challenging leadership practice. *International Leadership Education Research Network, ILERN*. National College of School Leadership.
- Hopkins, D. (2006). Realizing the potential of system leadership. *Paper to Support a Keynote Presentation at the School Leadership Symposium (SLS) for School Effectiveness, School Improvement, School Management*. University of Erfurt Erfurt, Germany, 26–28 September.
- Leithwood, K., Jantzi, D., and Dart, B. (1990). Transformational leadership: How principals can help reform school cultures. *School Effectiveness and School Improvement* **1**(4), 249–280.

- Leithwood, K., Day, C., Sammons, P., Harris, A., and Hopkins, D. (2006). *Seven Strong Claims about Successful School Leadership*. Nottingham: National College for School Leadership.
- MacBeath, J. (2004). Democratic learning and school effectiveness. Are they by any chance related? In MacBeath, J. and Moos, L. (eds.) *Democratic Learning. The Challenge to School Effectiveness*, pp 19–52. London: RoutledgeFalmer.
- Møller, J. (2006). Democratic schooling in Norway: Implications for leadership in practice. *Special Issue: International Perspectives on Leadership for Social Justice. Leadership and Policy in Schools* 5(1), 53–69.
- Møller, J. (2007). Educational leadership and the new language of learning. *International Journal of Leadership in Education* 9(1), 31–48.
- Møller, J. and Schratz, M. (2008). Leadership development in Europe. In Gary, C., Jacky, L., and Petros, P. (eds.) *UCEA/BELMAS/CCEAM International Handbook on the Preparation and Development of School Leaders*, pp 341–367, Erlbaum.
- Mulford, B. (2004). Congruence between the democratic purposes of schools and school principal training in Australia. *Journal of Educational Administration* 42(6), 625–639.
- Shields, C. (2006). Thinking about social justice: Bridging the gap between theory and educators conceptions. *A Paper Presented at the CCEAM Conference (Commonwealth Council of Educational Administration and Management)*. University of Cyprus, Lefkosia, Cyprus, October.
- Spillane, J. P., Halverson, R., and Diamond, J. B. (2004). Towards a theory of leadership practice: A distributed perspective. *Journal of Curriculum Studies* 36(1), 3–34.
- Starratt, R. J. (2004). Leadership of the contested terrain of education for democracy. *Journal of Educational Administration* 42(6), 724–731.
- Vedø, G. and Møller, J. (2007). Successful school leadership for diversity. Examining two contrasting examples of working for democracy in Norway. *ISEA, International Studies in Educational Administration* 35(3), 58–67.
- Warren, M. (2002). Deliberative democracy. In Carter, A. and Stokes, G. (eds.) *Democratic Theory Today*, pp 173–202. Cambridge: Polity.
- Woods, P. A. (2005). *Democratic Leadership in Education*. London: Paul Chapman.

Further Reading

- Berliner, D. (2007). Our impoverished view of educational research. In Sadovnik, A. (ed.) *Sociology of Education. A Critical Reader*, pp 487–519. New York: Routledge.
- Furman, G. and Gruenewald, D. A. (2004). Expanding the landscape of social justice: A critical ecological analysis. *Educational Administration Quarterly* 40(1), 47–76.
- Jacobson, S. and Leithwood, K. (eds.) (2007). *Special Issue: The Leading Edge of Distributed Leadership Research Leadership and Policy in Schools* 6(1).
- Johansson, O. (ed.) (2004). *Special Issue: Democracy and Leadership Journal of Educational Administration* 42(6).
- MacBeath, J. and Moos, L. (eds.) (2004). *Democratic Learning. The Challenge to School Effectiveness*. London: RoutledgeFalmer.
- Maxcy, S. J. (1995). *Democracy, Chaos, and the New School Order*. Thousand Oaks, CA: Corwin Press.
- Møller, J. (2009). Approaches to school leadership in Scandinavia. *Journal of Educational Administration and History* 41(2), 165–177.
- Shields, C. M. (2007). A failed initiative: Democracy has spoken – or has it? *Journal of Cases in Educational Leadership* 10(1), 14–21.

Leadership: Instructional

P Hallinger, Mahidol University, Bangkok, Thailand

K Leithwood, University of Toronto, Toronto, ON, Canada

R H Heck, University of Hawaii, Honolulu, HI, USA

© 2010 Elsevier Ltd. All rights reserved.

The Evolution of Instructional Leadership

Instructional leadership has its first roots in the teaching role of the head teacher of the school. As schools grew in size, their head teachers or principal teachers took on additional management responsibilities (Cuban, 1988; Tyack and Hansot, 1982). During the first decades of the twentieth century in the United States, the growth of education management as a profession saw conscious efforts to promote the status of school managers by emphasizing their managerial – as opposed to their instructional – responsibilities (Tyack and Hansot, 1982). This resulted in a continuing tension between the key roles of school leaders that persists to this day.

Cuban (1988), for example, identified instructional, managerial, and political roles that he asserted were equally relevant to work of successful school leaders. While acknowledging the importance of the instructional leadership role, Cuban contended that there is a DNA in the principalship that exerts an undeniable force on the occupant of the role to attend to managerial and political responsibilities and often to place them above the instructional role. He further suggested that attempts by principals to focus too tightly on the instructional role alone would be short-lived and potentially counter-productive.

The first significant attempt to clearly define conceptually and study empirically what we today term instructional leadership was undertaken by Gross and Herriott (1965) in a large-scale study of elementary school principals in the United States during the 1960s. These scholars formulated the construct of executive professional leadership (EPL) and developed associated measures for use in empirical research. They also reasoned that school leadership entailed working with educational professionals in the improvement of their schools and that attempts to lead would require a focus on the educational process of the school. Therefore, the foci embedded in their EPL construct and instrument included the principal's attention to educational standards, teachers' professional development, school goals and values, and the principal's role in helping to create meaning in the work of teachers. Gross and Herriott (1965) found positive correlations between EPL and three key outcomes: staff morale, the professional performance of teachers, and pupils' learning (pp. 150, 151).

These empirical findings were among the first systematically derived findings concerning the effects of

instructional leadership in schools. While these findings stimulated further interest in instructional leadership, Gross and Herriott's own discussion of the implications of their study frame tensions in this role in terms that resonate 40 years later.

Our findings bear upon a basic controversy over the role of school principals. For years they have been exhorted by their state and national associations and their superiors to make the most of their position; but recently the leadership conception of their role has been challenged as unrealistic and inappropriate. Principals, say the critics, should not engage in efforts to influence teachers' performance but simply provide routine administrative services; otherwise school administrators invade the teachers' professional autonomy. (Gross and Herriott, 1965: 151)

The EPL findings paved the way for future studies of the role of principal in school change and improvement during the 1970s. However, the true emergence of principal instructional leadership grew out of research on instructionally effective schools conducted during the mid- to late 1970s (Brookover and Lezotte, 1977; Edmonds, 1979; Rutter *et al.*, 1979). These studies found a clear association between instructionally effective schools and evidence of strong instructional leadership carried out by the school principal (Bossert *et al.*, 1982; Dwyer, 1986; Edmonds, 1979; Hallinger and Murphy, 1985a; Leithwood and Montgomery, 1982). Notably, publication of these findings came at a historical moment when policymakers, especially in the USA, were concerned about declining levels of school performance.

Subsequently, these research findings touting the importance of instructional leadership were widely disseminated during the 1980s. With these findings in hand, policymakers in the USA took steps to encourage all principals to strengthen their capacities in carrying out this role in order to make their own schools more effective. Critics identified holes in this logic (e.g., Barth, 1986; Cuban, 1988), as well as important limitations of the underlying research (Bossert *et al.*, 1982; Cuban, 1984). Nonetheless, by 1990 in the United States instructional leadership had become a normatively preeminent role that principals who wished to be effective were expected to fulfill.

Instructional leaders – goal oriented – were described as strong, directive principals who had been successful at turning around schools operating in difficult circumstances (Bamburg and Andrews, 1990; Bossert *et al.*, 1982;

Edmonds, 1979; Hallinger and Murphy, 1985a, 1985b, 1986). They defined a clear vision for the school focused primarily on the improvement of student academic outcomes (Bamburg and Andrews, 1990; Cheng, 1991, 1994; Glasman, 1984; Goldring and Pasternak, 1994; Hallinger and Murphy, 1986; Heck *et al.*, 1990; Leithwood *et al.*, 1990; Leitner, 1994; O'Day, 1983). The effective instructional leader was able to align the strategies and activities of the school with the school's academic mission (Bamburg and Andrews, 1990; Bossert *et al.*, 1982; Cheng, 1994; Cohen and Miller, 1980; Dwyer, 1986; Glasman, 1984; Goldring and Pasternak, 1994; Hallinger *et al.*, 1996; Heck, 1992, 1993; Heck *et al.*, 1990; Jones, 1983; Leitner, 1994).

Instructional leaders led through a combination of expertise and charisma. These were hands-on principals, hip-deep in curriculum and instruction (Cuban, 1984), and unafraid of working directly with teachers on the improvement of teaching and learning (Bossert *et al.*, 1982; Dwyer, 1986; Edmonds, 1979; Hallinger *et al.*, 1996; Hallinger and Murphy, 1986; Heck *et al.*, 1990; Leithwood *et al.*, 1990). Instructional leaders were viewed as a minority of principals who somehow managed to overcome the multiple pressures that push principals away from curriculum, instruction, and the classroom.

It is notable that during the 1980s relatively little reference was made to teachers, department heads, or even to assistant principals as instructional leaders. There was little discussion of instructional leadership as a distributed function to be shared beyond the role of the principal. This reflected a heroic view of this leadership role fostered by the imagery of the effective school literature as well as by intent of policymakers on implementing instructional leadership as a policy solution to problems of school performance.

During the 1990s, an emerging focus on teacher professionalism led to reconsideration of the instructional leadership role in schools. The idea that a single leader could fulfill the complex requirements of the instructional leadership role was questioned. The arguments raised by earlier critics such as Cuban and Barth resurfaced as the rhetoric of broad research findings bumped up against the reality of practicing instructional leadership in real schools. Several core issues pressed for the development of a view of instructional leadership as a shared or distributed role in schools.

- Under the emerging systems of school-based management, principals were being asked to take on ever greater responsibilities which further precluded a unitary focus on instructional leadership.
- The technical expertise required for instructional leadership was often beyond the capacity of any single individual in the school.
- The size and complexity of secondary schools made it difficult, if not impossible, for the principal to carry out

all of the tasks as described in the work of instructional leaders in smaller elementary schools.

- Contingency theories of leadership, selected empirical studies, and the wisdom of practitioners all supported the conclusion that the specific model of instructional leadership derived from research on instructionally effective elementary schools serving low-income children was not generalizable to all schools.

Consequently, during the late 1990s researchers turned toward the development of shared models of instructional leadership (Barth, 2002; Day *et al.*, 2001; Jackson, 2000; Lambert, 2002; Marks and Printy, 2004; Southworth, 2002). We will review the instructional leadership construct in light of this historical evolution.

It is interesting to observe and note that this heroic view of the principal as instructional leader is being resurrected at the present time by those who argue that principals should have or acquire a deep understanding of subject matter and how to teach it. Elmore's (2000) argument for such a view has received considerable attention among policymakers, and a compelling strand of research has demonstrated differences in how leaders interact with their colleagues based on their pedagogical content (usually mathematics) knowledge (e.g., Nelson and Sassi, 2005; Stein and Nelson, 2003).

While these authors acknowledge the possibilities for the distribution of such leadership, most of their cases and illustrations are focused on principals. This resurrection, it should be noted, comes at a time when principals are expected to work much more closely with parents and external community members, manage relationships with increasingly demanding teacher unions, balance shrinking budgets, and respond to a blizzard of accountability-oriented paperwork. This call for greater pedagogical content knowledge also comes as fewer and fewer teachers see much incentive for assuming what is increasingly being viewed as an impossible job.

A Conceptual Definition of Instructional Leadership

Following development of the EPL model by Gross and Herriott (1965), several notable models of instructional leadership were proposed, all grounded in the effective schools research base (see Andrews and Soder, 1987; Bossert *et al.*, 1982; Hallinger and Murphy, 1985; Leithwood *et al.*, 1990; Leithwood and Montgomery, 1982; Van de Grift, 1987; Villanova *et al.*, 1982) during the 1980s. A model proposed by Hallinger and Murphy (1985) has been used most frequently in empirical investigations (Hallinger, 2001; Hallinger and Heck, 1996a, 1996b). This model, similar in general terms to the others noted above, proposed three

dimensions for the instructional leadership role of the principal: defining the school's mission, managing the instructional program, and promoting a positive school learning climate (Hallinger, 2001; Hallinger and Murphy, 1985a). These three dimensions were further delineated into ten instructional leadership functions (see **Figure 1**).

Two functions, framing the school's goals and communicating the school's goals, comprise the first dimension, defining the school's mission. This dimension concerns the principal's role in determining the central purposes of the school. It describes the principal's role in working with staff to ensure that the school has both a broad inspiring mission as well as clear, measurable, time-based goals focused on the academic progress of students. It is also the principal's responsibility to communicate these goals so they are widely known and supported throughout the school community.

Within this model, the process of goal development was considered less critical than the outcome. Goals could be set by the principal or in collaboration with staff. The bottom line, however, was the school should have a clear, inspiring, academic vision that staff support and incorporate into their daily practice. This picture of goal-oriented, academically focused schools contrasted with the typical situation in which schools were portrayed as pursuing a variety of vague, ill-defined, and sometimes conflicting academic and nonacademic goals (March, 1978; Weick, 1976, 1982).

The second dimension, managing the instructional program, focuses on the coordination and control of instruction and curriculum. This dimension incorporates three leadership (or what could also be termed management) functions: supervising and evaluating instruction, coordinating the curriculum, and monitoring student progress. This dimension requires the principal to be deeply engaged in stimulating, developing, supervising, and monitoring teaching and learning in the school. Obviously, these functions demand the principal to have expertise in teaching and learning, as well as a commitment to the school's improvement. It is this dimension that requires the principal to become hip-deep in the

school's instructional program (Blasé and Blasé, 1998; Bossert *et al.*, 1982; Cuban, 1984; Dwyer, 1986; Edmonds, 1979; Marshall, 1996).

The third dimension, promoting a positive school learning climate, includes several functions: protecting instructional time, promoting professional development, maintaining high visibility, providing incentives for teachers, developing high expectations and standards, and providing incentives for learning. This dimension is broader in scope and purpose than the other two. It conforms to the notion that effective schools create an academic press through the development of high standards and expectations for students and teachers (Blasé and Blasé, 1998; Bossert *et al.*, 1982; Purkey and Smith, 1983).

Instructionally effective schools develop a culture of continuous improvement in which rewards are aligned with purposes and practices (Barth, 1990; Blasé and Blasé, 1998; Glasman, 1984; Hallinger and Murphy, 1986; Heck *et al.*, 1990; Leithwood and Montgomery, 1982; Mortimore, 1993; Purkey and Smith, 1983). Finally, the principal must model values and practices that create an environment and support the continuous improvement of teaching and learning (Dwyer, 1986; Hallinger and Murphy, 1985b).

Empirical Research on Instructional Leadership

Several reviews of the educational leadership literature by Hallinger and Heck (1996a, 1996b, 1999) found that instructional leadership was the most frequently studied model of school leadership during the past quarter century. Although theoretical interest in instructional leadership originated in the United States, research on instructional leadership has been extensive and global in scope. Important contributions have been made by researchers outside of the United States in North America, Europe, and Asia (e.g., Cheng, 1991, 1994; Day, 2002; Jackson, 2000; Leithwood and Jantzi, 1999; Mortimore, 1993; Southworth, 2002; van de Grift, 1989, 1990).

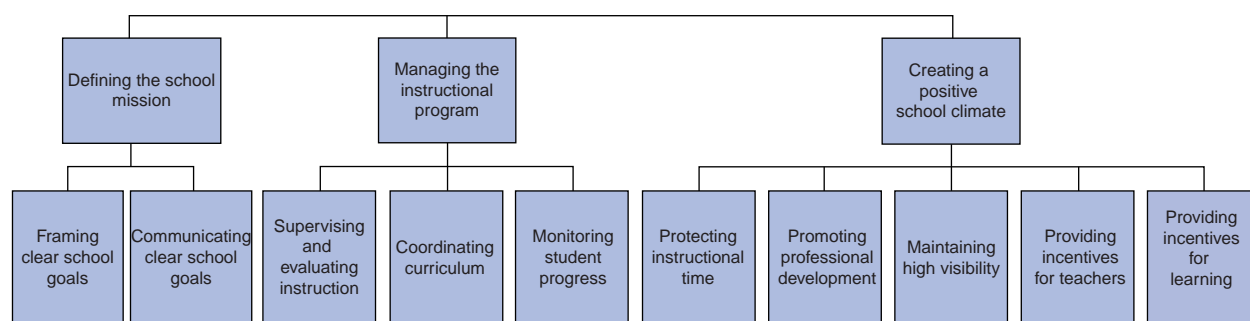


Figure 1 Instructional management framework. From Hallinger and Murphy (1985).

Research Trends

The most frequently used conceptualization of instructional leadership was developed by Hallinger during the early 1980s (Hallinger and Murphy, 1985a). Over 120 empirical studies have been completed using this conceptual model and related instrumentation – the principal instructional management rating scale (Hallinger, 2001 (it should be noted here that this paper has since been updated in 2007 but has not been presented or published. The figures presented here reflect the updated numbers from the most recent analysis in 2007)). This affords a useful perspective on development of interest in this leadership construct.

Although early studies of instructional leadership using the *PIMRS* were conducted almost entirely in the United States, subsequent studies have spanned North America, Europe, and Austral-Asia. If we break the period from 1983 to 2007 review into 5-year periods, it is possible to see the trend of scholarly interest in instructional leadership since the inception of the effective schools movement in the early 1980s. (It should be pointed out that the studies reviewed in the Hallinger papers consisted entirely of doctoral dissertations that used the *Principal Instructional Management Rating Scale* developed by him in 1982: 1983–1988 17 studies; 1989–1994 43 studies; 1995–2000 29 studies; 2001–2006 33 studies.)

This trend demonstrates a strong consistency of interest in the topic of instructional leadership over this 25-year period among scholars in educational administration. The first half of the period reviewed, 1983–1994, shows the growing interest in instructional leadership following the emergence of the effective schools movement. While interest waned somewhat during the mid-1990s, interest in studying this role of the school principal has remained quite stable since then.

Research Findings on Instructional Leadership

Scholars have conducted a substantial body of international research on instructional leadership since 1980. While the quality of the research remains somewhat uneven, the scope far exceeded prior efforts at understanding principal practice in this domain (Bridges, 1982; Hallinger, 2001; Hallinger and Heck, 1996a, 1996b). Consequently, there is a more systematic knowledge base today than existed 30 years ago.

This body of research has yielded findings concerning the:

- effects of personal antecedents (e.g., gender, training, and experience) and the school context (school level, school size, school SES) on instructional leadership;

- effects of instructional leadership on the organization (e.g., school mission and goals, expectations, curriculum, teaching, and teacher engagement); and
- direct and indirect effects of instructional leadership on student achievement and a variety of school outcomes.

Interested readers are referred to other up-to-date sources (Day *et al.*, 2001; Hallinger, 2001; Hallinger *et al.*, 1996; Hallinger and Heck, 1996a, 1996b; Leithwood *et al.*, 1990; Marks and Printy, 2004; Southworth, 2002; Witziers *et al.*, 1983). In brief, the following conclusions from research on instructional leadership warrant specification.

The preponderance of evidence indicates that school principals contribute to school effectiveness and student achievement indirectly through actions they take to influence school and classroom conditions (Hallinger and Heck, 1996a, 1996b, 1999). The size of the effects that principals indirectly contribute toward student learning though statistically significant is also quite small. While a small contribution may be meaningful, it is wise to keep in mind the strength of the treatment in relation to the desired outcomes when policymakers focus on the selection and training of school leaders as a strategy for large-scale change (March, 1978).

The most influential avenue of effects concerns the principal's role in shaping the school's mission (Bamburg and Andrews, 1990; Goldring and Pasternak, 1994; Glasman, 1984; Hallinger *et al.*, 1996; Hallinger and Heck, 1996a, 1996b, 2002; Hallinger and Murphy, 1985b; Heck *et al.*, 1990; Marks and Printy, 2004). This finding is important in that the effect of the vision/mission variable is strongly substantiated by research on leadership outside of education (e.g., Kantabutra, 2003).

The school context has an effect on the type of instructional leadership exercised by principals (Hallinger and Heck, 1996a, 1996b; Hallinger and Murphy, 1985b, 1986; Mulford and Silins, 2003). This finding is consistent with Bridges' (1977) assertion that principal leadership be viewed as both an independent and dependent variable also find empirical support. In particular, the role that principals play in mission building appears to be influenced by features of the school context (Bamburg and Andrews, 1990; Hallinger and Heck, 2002; Hallinger and Murphy, 1985b, 1986; Scott and Teddlie, 1987). Successful instructional leaders work with other stakeholders to shape the purposes to fit the needs of the school and its environment.

Instructional leaders also influence the quality of school outcomes through the alignment of school structures (e.g., academic standards, time allocation, and curriculum) and culture with the school's mission (Barth, 1990, 2002; Dwyer, 1986; Hallinger and Heck, 1996a, 1996b; Leitner, 1994; Southworth, 2002). Instructional leaders both lead through building a mission and manage through activities that increase alignment of activities with those purposes.

This again finds support in the more general leadership literature (e.g., Kantabutra, 2003; Kotter, 2002).

In 1980, popular images of instructional leadership portrayed principals as evidencing active hands-on involvement in classroom. It is interesting, however, to note that relatively few studies have actually found instructional leaders displaying this type of hands-on supervision of classroom instruction (Hallinger and Heck, 1996a, 1996b). Where principals do get actively involved in instructional supervision, it tends to be at the elementary school level (Braughton and Riley, 1991; Heck *et al.*, 1990; Hallinger and Heck, 1996a, 1996b). The preponderance of studies instead suggests that principal effects on classroom instruction operate through the school's culture and by modeling rather than through direct control through the supervision and evaluation of teaching.

Significant progress has been made over the past 30 years in understanding the nature of instructional leadership in schools. As a research topic scholars have studied the role extensively and with a variety of frameworks and methodologies. There is little evidence to support the view that on a broad scale at either the elementary or secondary school level, principals have become more significantly engaged in hands-on directed supervision of teaching and learning in classrooms.

Yet, if we define instructional leadership broadly to focus on the dimensions of defining a school mission and creating a positive school culture, the picture is somewhat different. Research on instructional leadership suggests that these dimensions of the principal's leadership role are becoming integrated more firmly into the principal's role behavior (Hallinger, 2004).

Toward Shared Instructional Leadership

Since the turn of the millennium, a global tsunami of educational reform has refocused the attention of policymakers and practitioners on the question: How can we create conditions that foster the use of more powerful methods of learning and teaching in schools (Caldwell, 2006; Hallinger, 2003; Jackson, 2000; Murphy, 2000)? This global focus on the improvement of learning and teaching has once again brought instructional leadership to the fore (Gewertz, 2003; Hallinger, 2003; Huber, 2003; Stricherz, 2001a, 2001b).

What should these instructional leaders be concerned with? A broad reading of the literature on instructional leadership that has emerged over the past 25 years would have instructional leaders focus on:

- creating a shared sense of purpose in the school, including clear goals focused on student learning;
- fostering the continuous improvement of the school through cyclical school development planning that involves a wide range of stakeholders;

- developing a climate of high expectations and a school culture aimed at innovation and improvement of teaching and learning;
- coordinating the curriculum and monitoring student learning outcomes;
- shaping the reward structure of the school to reflect the school's mission;
- organizing and monitoring a wide range of activities aimed at the continuous development of staff; and
- being a visible presence in the school, modeling the desired values of the school's culture.

However, the experience of policymakers and practitioners who have sought to exercise instructional leadership over the past 25 years provides a more realistic starting point. Contextual and structural constraints embedded in schools as organizations and in the DNA of the principalship suggest the need for a less heroic and more grounded conception of the instructional leadership role. Moreover, both practical experience and research findings suggest suitability of conceptualizing instructional leadership as a role to be shared by the principal with others in the school (Barth, 2002; Ogawa and Bossert, 1995).

This point was captured by Lambert (2002) who contends that, "The days of the lone instructional leader are over. We no longer believe that one administrator can serve as the instructional leader for the entire school without the substantial participation of other educators" (p. 37). Thus, several different writers have attempted to integrate these constructs into a variant referred to as shared instructional leadership (Barth, 2002; Day *et al.*, 2001; Donaldson, 2001; Jackson, 2000; Lambert, 2002; Marks and Printy, 2004; Southworth, 2002).

While several of the scholars cited here have written eloquently about the possible forms this might take, the most ambitious attempt to study shared instructional leadership empirically was undertaken by Marks and Printy (2004). Their study examined both instructional leadership and its major competitor for the attention of school leaders, transformational leadership (e.g., Leithwood and Jantzi, 2006). Their conclusion points the way toward bringing together both of these major approaches to school leadership by addressing both teachers' commitments and engagement (the transformational leadership focus) as well as the focus on student learning (the instructional leadership focus).

Although it is too soon to know whether the findings from the research by Marks and Printy will be replicated by others, their conclusion points the way toward one possible avenue of reconciliation for these constructs:

This study suggests that strong transformational leadership by the principal is essential in supporting the commitment of teachers. Because teachers themselves can be barriers to the development of teacher leadership (Smylie and Denny, 1990), transformational principals are needed to invite

teachers to share leadership functions. When teachers perceive principals' instructional leadership behaviors to be appropriate, they grow in commitment, professional involvement, and willingness to innovate (Sheppard, 1996). Thus, instructional leadership can itself be transformational. (Marks and Printy, 2004)

A second approach to understanding the relationship between these leadership models may lie in contingency theory (Hallinger, 2005). In a review of the literature on principal effects, Hallinger and Heck (1996a, 1996b) concluded that it is virtually meaningless to study principal leadership without reference to the school context. The context of the school is a source of constraints, resources, and opportunities that the principal must understand and address in order to lead. Contextual variables of interest to principals include student background, community type, organizational structure, school culture, teacher experience and competence, fiscal resources, school size, and bureaucratic and labor features of the school organization (Bossert *et al.*, 1982; Hallinger and Heck, 1996a, 1996b).

The contingent characteristic of leadership must be explicitly incorporated into theoretical models (Hersey and Blanchard, 1969). Leadership must be conceptualized as a mutual influence process, rather than as a one-way process in which leaders influence others (Bridges, 1977; Graeff, 1997; Jackson, 2000; Leithwood and Jantzi, 1999). Effective leaders respond to the changing needs of their context. Indeed, in a very real sense the leader's behaviors are also shaped by the school context.

Thus, one resolution of the quest for an integrative model of educational leadership would link leadership to the needs of the school context. Jackson (2000) and Fullan (2002) have observed that school improvement is a journey. The type of leadership that is suitable to a certain stage of the journey may become a limiting or even counter-productive force as the school develops (see also Hersey and Blanchard, 1969). Schools at risk may initially require a more forceful top-down approach focused on instructional improvement. Instructional leaders would typically set clear, time-based, academically focused goals in order to get the organization moving in the desired direction. They would take a more active hands-on role in organizing and coordinating instruction.

The practice of instructional leadership has benefitted from theoretical and empirical developments of the past 25 years. Nonetheless, the same tensions evident in the practice of this role suggested by Gross and Herriott in the 1960s and echoed by Cuban in the 1980s remain relevant today. We suggest that future development of instructional leadership in schools will benefit from recognition of its complexity as well as from understanding leadership as an organizational phenomenon (Ogawa and Bossert, 1995).

Acknowledgment

This article has drawn extensively on the work published by Hallinger (2005). The excerpts were included with permission from the publisher, Taylor and Francis.

Bibliography

- Andrews, R. and Soder, R. (1987). Principal instructional leadership and school achievement. *Educational Leadership* **44**, 9–11.
- Bamburg, J. and Andrews, R. (1990). School goals, principals and achievement. *School Effectiveness and School Improvement* **2**(3), 175–191.
- Barth, R. (1986). On sheep and goats and school reform. *Phi Delta Kappan* **68**(4), 293–296.
- Barth, R. (1990). *Improving Schools from within*. San Francisco, CA: Jossey-Bass.
- Barth, R. (2002). *Learning by Heart*. San Francisco, CA: Jossey-Bass.
- Blasé, J. and Blasé, J. (1998). *Handbook of Instructional Leadership: How Really Good Principals Promote Teaching and Learning*. Thousand Oaks, CA: Corwin Press.
- Bossert, S., Dwyer, D., Rowan, B., and Lee, G. (1982). The instructional management role of the principal. *Educational Administration Quarterly* **18**(3), 34–64.
- Broughton, R. and Riley, J. (1991). The Relationship between Principals' Knowledge of Reading Processes and Elementary School Reading Achievement, ERIC Document Reproduction Service No. ED341952.
- Bridges, E. (1977). The nature of leadership. In Cunningham, L., Hack, W., and Nystrand, R. (eds.) *Educational Administration: The Developing Decades*, pp 202–230. Berkeley, CA: McCutchan.
- Brookover, W. and Lezotte, L. (1977). *Changes in School Characteristics Coincident with Changes in Student Achievement*. East Lansing, MI: Michigan State University Press.
- Caldwell, B. (2006). Re-imagining educational leadership. *Australian Educational Leader* **44**, 10–12; 44–45.
- Cheng, Y. C. (1991). Leadership style of principals and organizational process in secondary schools. *Journal of Educational Administration* **29**(2), 25–37.
- Cheng, Y. C. (1994). Principal's leadership as a critical factor for school performance: Evidence from multi-levels of primary schools. *School Effectiveness and School Improvement* **5**(3), 299–317.
- Cohen, E. and Miller, R. (1980). Coordination and control of instruction in schools. *Pacific Sociological Review* **4**, 446–473.
- Cuban, L. (1984). Transforming the frog into a prince: Effective schools research, policy, and practice at the district level. *Harvard Educational Review* **54**(2), 129–151.
- Cuban, L. (1988). *The Managerial Imperative and the Practice of Leadership in Schools*. Albany, NY: SUNY Press.
- Day, C., Harris, A., and Hadfield, M. (2001). Challenging the orthodoxy of effective school leadership. *International Journal of Leadership in Education* **4**(1), 39–56.
- Donaldson, G. A. (2001). *Cultivating Leadership in Schools: Connecting People, Purpose, and Practice*. New York: Teachers College Press.
- Dwyer, D. (1986). Understanding the principal's contribution to instruction. *Peabody Journal of Education* **63**(1), 3–18.
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership* **37**, 15–24.
- Fullan, M. (2002). The change leader. *Educational Leadership* **59**(8), 16–20.
- Gewertz, C. (2003). N.Y.C. Chancellor aims to bolster instructional leadership. *Education Week* **22**(16), 7; 12.
- Glasman, N. (1984). Student achievement and the school principal. *Educational Evaluation and Policy Analysis* **6**(3), 283–296.
- Golding, E. and Pasternak, R. (1994). Principals' coordinating strategies and school effectiveness. *School Effectiveness and School Improvement* **5**(3), 239–253.
- Graeff, C. (1997). Evolution of situational leadership theory: A critical review. *Leadership Quarterly* **8**(2), 17–26.

- Hallinger, P. (2001). A review of two decades of research on the principalship using the principal instructional management rating scale. *Paper Presented at the Annual Meeting of the American Educational Research Association*. Seattle, WA.
- Hallinger, P. (2004). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education* **33**(3), 329–351.
- Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and Policy in Schools* **4**(3), 221–240.
- Hallinger, P., Bickman, L., and Davis, K. (1996). School context, principal leadership and student achievement. *Elementary School Journal* **96**(5), 498–518.
- Hallinger, P. and Heck, R. (1996a). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980–1995. *Educational Administration Quarterly* **32**(1), 5–44.
- Hallinger, P. and Heck, R. (1996b). The principal's role in school effectiveness: A review of methodological issues, 1980–95. In Leithwood, K., Chapman, J., Corson, P., Hallinger, P., and Hart, A. (eds.) *The International Handbook of Educational Leadership and Administration*, pp 723–784. Dordrecht: Kluwer.
- Hallinger, P. and Heck, R. (2002). What do you call people with visions? The role of vision, mission and goals in school improvement. In Leithwood, K., Hallinger, P., Furman, G., et al. (eds.) *The Second International Handbook of Educational Leadership and Administration*, pp 9–40. Dordrecht: Kluwer.
- Hallinger, P. and Murphy, J. (1985a). Assessing the instructional leadership behavior of principals. *Elementary School Journal* **86**(2), 217–248.
- Hallinger, P. and Murphy, J. (1985b). What's effective for whom? School context and student achievement. *Planning and Changing* **16**(3), 152–160.
- Hallinger, P. and Murphy, J. (1986). The social context of effective schools. *American Journal of Education* **94**(3), 328–355.
- Heck, R. (1993). School context, principal leadership, and achievement: The case of secondary schools in Singapore. *Urban Review* **25**(2), 151–166.
- Heck, R. (1992). Principal instructional leadership and the identification of high- and low-achieving schools: The application of discriminant techniques. *Administrator's Notebook* **34**(7), 1–4.
- Heck, R., Larson, T., and Marcoulides, G. (1990). Principal instructional leadership and school achievement: Validation of a causal model. *Educational Administration Quarterly* **26**, 94–125.
- Hersey, P. and Blanchard, K. (1969). Life cycle theory of leadership. *Training and Development Journal* **23**(2), 26–34.
- Huber, S. (2003). School leader development: Current trends from a global perspective. In Hallinger, P. (ed.) *Reshaping the Landscape of School Leadership Development: A Global Perspective*, pp 273–288. Lisse: Swets and Zeitlinger.
- Jackson, D. (2000). The school improvement journey: Perspectives on leadership. *School Leadership and Management* **20**(1), 61–78.
- Lambert, L. (2002). A framework for shared leadership. *Educational Leadership* **59**(8), 37–40.
- Leithwood, K., Begley, P., and Cousins, B. (1990). The nature, causes and consequences of principals' practices: An agenda for future research. *Journal of Educational Administration* **28**(4), 5–31.
- Leithwood, K. and Jantzi, D. (1999). The relative effects of principal and teacher sources of leadership on student engagement with school. *Educational Administration Quarterly* **35**, 679–706.
- Leithwood, K. and Jantzi, D. (2005). A review of transformational leadership research: 1996–2005. *Leadership and Policy in Schools* **4**(3), 177–200.
- Leithwood, K. and Montgomery, D. (1982). The role of the elementary principal in program improvement. *Review of Educational Research* **52**(3), 309–339.
- Leitner, D. (1994). Do principals affect student outcomes? An organizational perspective. *School Effectiveness and School Improvement* **5**(3), 219–239.
- March, J. (1978). The American public school administrator: A short analysis. *School Review* **86**, 217–250.
- Marks, H. and Printy, S. (2004). Principal leadership and school performance: An integration of transformation and instructional leadership. *Educational Administration Quarterly* **39**, 370–397.
- Marshall, K. (1996). How I confronted HSPS (hyperactive superficial principal syndrome). *Phi Delta Kappan* **77**(5), 336–345.
- Mortimore, P. (1993). School effectiveness and the management of effective learning and teaching. *School Effectiveness and School Improvement* **4**, 290–310.
- Murphy, J. (2002). Reculturing the profession of educational leadership: New blueprints. *Educational Administration Quarterly* **38**(2), 176–192.
- Mulford, B. and Silins, H. (2003). Leadership for organisational learning and improved student outcomes – what do we know? *Cambridge Journal of Education* **33**(2), 175–195.
- Nelson, B. S. and Sassi, A. (2005). *The Effective Principal: Instructional Leadership for High-Quality Learning*. New York: Teachers College Press.
- O'Day, K. (1983). *The Relationship between Principal and Teacher Perceptions of Principal Instructional Management Behavior and Student Achievement*. Unpublished doctoral dissertation, Northern Illinois University, Normal, Illinois.
- Ogawa, R. and Bossert, S. (1995). Leadership as an organizational quality. *Educational Administration Quarterly* **31**(2), 224–243.
- Purkey, S. and Smith, M. (1983). Effective schools: A review. *Elementary School Journal* **83**, 427–452.
- Rutter, M., Maughan, B., Mortimore, P., Ouston, J., and Smith, A. (1979). *Fifteen Thousand Hours: Secondary Schools and Their Effects on Children*. Cambridge, MA: Harvard University Press.
- Scott, C. and Teddlie, C. (1987). Student, teacher, and principal academic expectations and attributed responsibility as predictors of student achievement: A causal modeling approach. *Paper Presented at the Annual Meeting of the American Educational Research Association*. Washington, DC.
- Sheppard, B. (1996). Exploring the transformational nature of instructional leadership. *Alberta Journal of Educational Research* **42**(4), 325–344.
- Southworth, G. (2002). Instructional leadership in schools: Reflections and empirical evidence. *School Leadership and Management* **22**(1), 73–92.
- Stein, M. K. and Nelson, B. S. (2003). Leadership content knowledge. *Educational Evaluation and Policy Analysis* **25**(4), 423–448.
- Stricherz, M. (2001a). D.C. Principal's training designed to boost instructional leadership. *Education Week* **21**(2), 13.
- Stricherz, M. (2001b). Leadership grant aimed at schools in South. *Education Week* **21**(3), 21.
- Tyack, D. and Hansot, E. (1982). *Managers of Virtue*. New York: Teachers College Press.
- van de Grift, W. (1989). Self perceptions of educational leadership and mean pupil achievements. In Reynolds, D., Creemers, B. P. M., and Peters, T. (eds.) *School Effectiveness and Improvement*, pp 227–242. Cardiff/Groningen: School of Education/RION.
- van de Grift, W. (1990). Educational leadership and academic achievement in elementary education. *School Effectiveness and School Improvement* **1**(3), 26–40.
- Villanova, R., Gauthier, W., Proctor, P., and Shoemaker, J. (1981). *The Connecticut School Effectiveness Questionnaire*. Hartford, CT: Bureau of School Improvement, Connecticut State Department of Education.
- Weick, K. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly* **21**, 1–19.
- Weick, K. (1982). Administering education in loosely coupled schools. *Phi Beta Kappan* **63**(10), 673–676.
- Witziers, B., Bosker, R., and Kruger, M. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly* **39**, 398–425.

Further Reading

- Kliene-Kracht, P. (1993). Indirect instructional leadership: An administrator's choice. *Educational Administration Quarterly* **29**(2), 187–212.

- Lambert, L. (1998). *Building Leadership Capacity in Schools*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Leithwood, K. and Louis, K. S. (1999). Organizational learning in schools: An introduction. In Leithwood, K. and Louis, K. S. (eds.) *Organizational Learning in Schools*, pp 1–16. Dordrecht: Kluwer Academic Press.
- Louis, K. S. and Marks, H. (1998). Does professional community affect the classroom? *American Journal of Education* **106**(4), 532–575.
- Marsh, D. (1992). School principals as instructional leaders: The impact of the California School Leadership Academy. *Education and Urban Society* **24**(3), 386–410.
- Rowan, B. and Denk, C. (1984). Management succession, school socioeconomic context and basic skills achievement. *American Educational Research Journal* **21**(3), 517–537.
- Wimpelberg, R., Teddlie, C., and Stringfield, S. (1989). Sensitivity to context: The past and future of effective schools research. *Educational Administration Quarterly* **25**, 82–105.

Leadership: Transformative

C M Shields, The University of Illinois at Urbana–Champaign, Champaign, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

The words promise, liberation, hope, empowerment, activism, risk, social justice, courage, or revolution do not automatically evoke images of educational leaders in charge of schools and systems, working within the dominant political and bureaucratic frameworks of the twenty-first century. Yet, all of these concepts are at the heart of transformative leadership – the concept to be explicated here. Transformative leadership (as opposed to either transactional or transformational leadership) takes seriously Freire's (1998) contention "that education is not the ultimate lever for social transformation, but without it transformation cannot occur" (p. 37). Transformative leadership begins with questions of justice and democracy; it critiques inequitable practices and offers the promise not only of greater individual achievement but of a better life lived in common with others. Transformative leadership, therefore, inextricably links education and educational leadership with the wider social context within which it is embedded.

This article provides an overview of the development of transformative leadership as a theory in its own right, distinct from transformational leadership; it demonstrates how both theories emerged from Burns' (1978) seminal book, *Leadership*; and explains how the terms, although originally often used synonymously, have developed into distinct theories – each with its own advocates, critics, and body of literature. The article concludes with an assessment of some of the strengths and weaknesses of transformative leadership and its potential to move the field of educational leadership forward in a time of conflict, confusion, and chaos.

Origins

Formal studies of educational administration and leadership owe much to the long history of work in other fields; nevertheless, in the middle of the last century, educational administration became a field of study in its own right. The futility of early efforts to identify physical traits, personal abilities, and individual attributes determinant of good leadership (interpreted most frequently as effective rather than moral) is well known, as are the attendant critiques of related great-man studies and rational, technical, positivist theories of leadership. Two early attempts to move beyond individual traits and characteristics of leaders were Burns' (1978) conceptions of transactional and transforming leadership. Transactional

leadership expressed the values and relationships often still associated with rational and technical forms of scientific management. Transforming leadership emphasized the ends of leadership as it sought to unite people around organizational goals.

These theories are well known, although the ways in which the latter concept has led to both transformational leadership and transformative leadership are less explored. The article clarifies this distinction, with particular emphasis on the latter – transformative leadership – as holding the most promise and potential to meet the needs of complex, diverse, and beleaguered education systems. **Table 1** demonstrates clearly, the deep differences among the three theories (transactional, transformational, and transformative) that have, in various ways, dominated the field of educational leadership for the last 30 years.

Ethical Leadership

Burns' concept of leadership requires a desire to "deal with leadership as distinct from mere power-holding and as the opposite of brute power" (p. 4). In Burns' conception, an amoral leader is an oxymoron. For that reason, Burns argues, "naked, power wielding can be neither transactional nor transforming; only leadership can be" (p. 20). Here, we note the centrality of moral and ethical behavior in Burns' conception of leadership. He begins his treatise by calling for a consideration of how both power (composed of motive and resources) and power relationships are central to comprehending the "true nature of leadership" (pp. 11–12). At the same time, neither transactional nor transformational leadership, as it has come to be known through the subsequent work of Bass (1985), Bass and Avolio (1994), Leithwood and Jantzi (1990, 1999), and others is adequate to address the challenges inherent in educating all students to high levels of personal intellectual development or of creating the civic capacity required for full participation in democratic society. Neither explicitly attends to the moral and ethical issues related to power relationships that may perpetuate inequity and inequality in organizations, although the moral dimension is more recognized in transforming leadership than in its transactional partner. To be a leader, according to Burns, is to induce "followers to act for certain goals that represent the values and the motivations – the wants and needs, the aspirations and expectations – *of both leaders and followers*" (p. 19, italics in original). Further, when

Table 1 Distinctions among three theories of leadership

	<i>Transactional leadership</i>	<i>Transformational leadership</i>	<i>Transformative leadership</i>
Foundation	An exchange	Meet the needs of complex and diverse systems	Critique and promise
Emphasis	Means	Organization	Deep and equitable change in social conditions
Processes	Immediate cooperation through mutual agreement and benefit	Understanding of organizational culture; setting directions, developing people, redesigning the organization, and managing the instructional program	Deconstruction and reconstruction of social/cultural knowledge frameworks that generate inequity, acknowledgment of power and privilege; dialectic between individual and social
Key values	Honesty, responsibility, fairness, and honoring commitments	Liberty, justice, equality, and effectiveness	Liberation, emancipation, democracy, equity, and justice
Goal	Agreement; mutual goal advancement	Organizational change; effectiveness	Individual, organizational and societal transformation
Power	Mostly ignored	Inspirational	Positional, hegemonic, tool for oppression as well as for positive action
Leader	Ensures smooth and efficient organizational operation through transactions	Looks for motive, develops common purpose, and focuses on organizational goals	Lives with tension and challenge; requires moral courage, activism
Related theories	Bureaucratic leadership, Scientific management	School effectiveness, School reform, School improvement, Instructional leadership	Critical theories (race and gender), Cultural and social reproduction, Leadership for social justice

he states that “transcending leadership is leadership engaged” (p. 20), he is pointing the way for transformative leadership which is inextricably engaged with the wider society.

which leadership eschews patronage in favor of rules and principles, can be clearly seen. Transactional leadership, for Burns, focuses on the means of leading, while transforming leadership implies a focus on the ends.

Transactional Leadership

Although Burns is often considered the originator of the concept of transactional leadership, it is important to note its origins in Weber’s (1924) earlier discussion of bureaucracy in which he discussed the categories of rational, legal, hierarchical power. Although Weber attempts to counteract the kind of bargaining that resulted in patronage and favoritism, there can be little doubt that he favored a kind of transactional legitimacy based on positional and not personal authority and on a formal code of administrative decision making and practice. To overcome favoritism and capriciousness, Weber advocated transactions based on knowledge and a systematic application of rules including rules of exchange.

Burns’ (1978) transactional leadership is also based on exchanging one thing for another: jobs for votes, subsidies for campaign contributions, and overtime work for increased pay. Its dominant values relate to means – “honesty, responsibility, fairness, and honoring commitments – without which transactional leadership could not work” (p. 426). The continuity between transactional leadership and Weber’s articulation of bureaucratic principles, in

Transforming Leadership: The Root of Confusion

Transforming leadership, as conceived by Burns (1978), occurred when the leader “recognizes and exploits an existing need or demand of a potential follower. . . , looks for potential motives in followers, seeks to satisfy higher needs, and engages the full person of the follower” (p. 4). “Transformational leadership,” he states (using the term to apply to the work of transforming leaders), “is more concerned with *end-values*, such as liberty, justice, equality” (p. 426, italics in original) – all aspects of both transformational and transformative leadership theories today. The focus on the moral purposes or ends of leadership has led to both transformational and transformative concepts. It is patently obvious that both theories of leadership – transformational and transformative – have at their heart the notion of transforming something. Indeed, to transform, according to most dictionaries, is to change in form, appearance, or structure, or in nature, condition, or character, usually with connotations of changing for the better. The Random House dictionary lists, as adjectives related to the verb transform and the noun transformation, both the words transformational and

transformative. As both theories of leadership emerged, it is little wonder that the two terms have frequently been used synonymously to describe educational leadership and that this lack of clarity has led to considerable confusion.

Moreover, although Burns most frequently used the term transforming leadership, he also used the words transformation and transformational; surprisingly, the term transformative, often associated with his work, is markedly absent. Nevertheless, the implications of his conception of transformation are directly related to transformative leadership. For example, he stated that “revolution is a complete and pervasive transformation of an entire social system” (p. 202) and later emphasized the need for “real change – that is, a transformation to the marked degree in the attitudes, norms, institutions, and behaviors that structure our daily lives” (p. 414).

Transformative Leadership: Roots in Other Social Sciences

Interestingly, the word transformative has been used with more clarity in other areas than in educational leadership where it is relatively recent. In healthcare and related social-service areas, transformative approaches are quite common. Duncan *et al.* (2006) advocated transformative curriculum in relation to an interdisciplinary approach to the education of healthcare professionals in South Africa; Evans *et al.* (2007) supported supplementing ameliorative approaches with transformative approaches by human service organizations. The latter distinguish between “incremental, developmental, evolutionary, or ‘first-order’ change” and “transformative, discontinuous, revolutionary, or ‘second-order’ change in human systems” (p. 332). Watkins’ (2000) used the concept of transformative leadership to describe a fourfold approach to nursing administration and healthcare.

Transformation in Education

The notion of transformation has led, in education, to concepts such as transformative teaching, the transformative classroom (Duncan and Clayburn, 1997), transformative curriculum, transformative material activity (Miettinen, 2006), and so forth. One particularly well-developed use of the term, explicated subsequently by numerous writers in the field of adult learning, is Mezirow’s (1991, 1996) transformative learning theory that outlines a process of effecting change in one’s frame of reference. The original focus was on individual learning prompted by self-reflection as a tool for deep and lasting personal change, but the concept has been expanded to emphasize the need to deconstruct and reconstruct knowledge frameworks as well as to “develop an

appreciation of our own culture and the associated privileges and powers” (Taylor, 2006: 92). Davis (2006), for example, emphasizes that transformative learning “involves the acquisition (or manipulation) of knowledge that disrupts prior learning and stimulates the reflective reshaping of deeply ingrained knowledge and belief structures” (p. 1). Franz (2002) reports that her findings related to effective staff partnerships included “eight types of transformative learning” (p. 1). Sterling *et al.* (2007) write about science camp for children of socioeconomically disadvantaged urban citizens as being a transformative experience both for students and their parents – changing the way they see themselves as learners and their self-assessment of their ability to attend college.

Others invoke the notion of transformative learning as it relates to increased gender equity and transformative gender justice. Keddie (2006a) calls for a “transformative approach that seeks to challenge and rework (rather than normalize and reinscribe) boys’ narrow conceptions of gender” (p. 111). She argues (Keddie, 2006b) that transformative gender justice remedies social disadvantage “through problematizing and restructuring the underlying frameworks that generate such disadvantage” (p. 401). Stromquist (2006) argues that

If education is such a powerful tool for economic and social betterment, we would expect that public policy would seek to make education accessible to all, to provide education of high quality and to distribute education equitably at all levels (p. 145).

She continues to say that “the acquisition of intellectual skills and habits by women is conducive to social change” (p. 149). King and Biro (2000) call for transformative learning to start with a disorienting dilemma and for it to “progress through a dynamic pathway of stages . . . [to a] final reintegration of a new frame of reference” (p. 19). MacKinnon (2000) argues that transformative educational practice must be based on a pedagogy of social justice (p. 11).

The common elements in these transformative approaches include the need for social betterment, enhancing equity, and for a thorough reshaping of knowledge and belief structures – elements that reappear as central tenets in the concept of transformative (although not so necessarily in transformational) leadership. Transformative concepts and social justice are closely connected through the shared goal of identifying and restructuring frameworks that generate inequity and disadvantage. Transformative ideals also owe much to the work of Freire (1970, 1998) who used the terms transform, transformation, and transformative to describe the changes that may occur as a result of education. Freire calls for personal, dialogic relationships to undergird education, because without such relationships, he argues, education acts to deform rather than to transform. He states, “Each time the ‘thou’ is

changed into an object, an “it,” dialogue is subverted and education is changed to deformation” (1970: 89).

Transformative Leadership: The Evolution of a Theory

Transformative leadership as a theory has developed in ways that are congruent with these uses of transformation and transformative learning in other fields of social science and education. It has also developed in ways that are consistent with Burns’ understanding of social change, of leaders who “build advocacy and conflict into the planning process in response to pluralistic sets of values” (p. 420). It incorporates Burns’ emphasis on purposeful moral leadership, as well as his acknowledgment that leadership necessarily includes an understanding of historical and social causation, of power wielding, and political power (pp. 433–434).

One of the first writers to discuss transformative educational leadership was William Foster (1986). His belief was that leadership “must be critically educative; it can not only look at the conditions in which we live, but it must also decide how to change them” (p. 185). Although perhaps ahead of his time (as discussed by Starratt (2004) in a retrospective essay commemorating Foster’s life and legacy), Foster’s advocacy of leadership that both transforms and empowers is central to today’s notion of transformative leadership.

In 1986, Bennis wrote an article titled ‘Transformative power and leadership’ in which he identified three factors as components of transformative power: the leader, the intention, and the organization and defined the transformative power of leadership as “the ability of the leader to reach the souls of others in a fashion which raises human consciousness, builds meanings, and inspires human intent that is the source of power” (p. 70). Acknowledgment of the effects of power is increasingly advocated and clarified in the emerging theory of transformative leadership, although writers will soon call for an understanding of power, not simply as an inspiring force, but also as a force that both implicitly and explicitly perpetuates hegemonic and dominating behaviors, cultures, and structures.

Capper (1989), in a paper advocating a more inclusive approach to democratic schooling for severely disabled students, drew on Aronowitz’ and Giroux’s (1985) notion of a transformative intellectual, and identified the need for the school administrator to be a transformative intellectual to encourage social justice and to practice “transformative leadership which can transcend the intellectual bias in democratic schooling to the benefit of all students and staff” (p. 5). Capper also cites Giroux’s and McLaren’s definition of a transformative intellectual as one who attempts

to insert teaching and learning directly into the political sphere by arguing that schooling represents both a struggle for meaning and a struggle over power relations. . . one whose intellectual practices are necessarily grounded in forms of moral and ethical discourse exhibiting a preferential concern for the suffering and struggles of the disadvantaged and oppressed (1986, in Capper, p. 9).

In these early articulations, we see the beginning of the major divergence between transformational and transformative leadership theories: the former focuses primarily on what happens within an organization while the latter is also concerned with the broader social and political sphere, recognizing that the inequities and struggles experienced in the wider society affect one’s ability both to perform and to succeed within an organizational context.

This distinction emerged gradually during the 1990s, with much writing still using the terms transformational and transformative interchangeably and synonymously. Sagor (1992), for example, wrote about three transformative leaders – principals who made a difference by using the three building blocks of “transformational leadership” (p. 13). Bates (1995) used the term transformative leader, but in a way more aligned with current transformational leadership, as he emphasized the work of leaders who reshaped and focused corporate culture carrying workers along with the vision (p. 11). Day *et al.* (2001) also perpetuated the confusion by continuing to use the terms interchangeably, alluding to the work of Leithwood and Jantzi as transformative leadership and then making reference to Burns’ (1978) transactional and transformational concepts (pp. 19–20). Sergiovanni (1990) again references Burns and states that “transformative leadership is first concerned with higher order psychological needs for esteem, autonomy, and self-actualization and, then, with moral questions of goodness, righteousness, duty, and obligation” (p. 23). A few pages later, Sergiovanni equates the concept he then calls transformational leadership with the notion of “value-added leadership” (p. 25). His appeal to psychology is important here as it soon helps to distinguish the development of the two strands of leadership. Transformative leadership, in the next decade, sheds most of its psychological trappings and focuses much more directly on sociological and cultural elements of organizations and the wider society in which they are embedded.

Quantz *et al.* (1991) outline many of the tenets of what has come to be known as transformative leadership. They argue that traditional theories of leadership are inadequate for democratic empowerment and that “only the concept of transformative leadership appears to provide an appropriate direction” (p. 96). Before clarifying their use of the term, they stipulate that the literature is still somewhat unclear about its meaning, and that the term requires critical reinterpretation (p. 96). Although they, too, fall at one point into the trap of using both terms interchangeably

(p. 97), their work is firmly grounded in the critical elements that distinguish transformative-leadership theory from formulations and characteristics of transformational leadership. They posit that schools are sites of cultural politics that serve both to reproduce and perpetuate the inequities inherent in gender, race, and class constructs and which “confirm and legitimate some cultures while disconfirming and delegitimizing others” (p. 98). They go on to argue that because organizations must be based on democratic authority, transformative educational leaders must learn to diminish “*undemocratic power relationships*” (p. 102, italics in original) and use their “power to transform present social relations” (p. 103). Transformative leadership, they assert, “requires a language of critique and possibility” (p. 105); a “transformative leader must introduce the mechanisms necessary for various groups to begin conversations around issues of emancipation and domination” (p. 112).

Transformative Leadership Today: A Theory of Critique and Possibility

Thus, despite confounding of the terms transformative and transformational into the present decade, it is clear that as early as the mid-1980s with the work of Aronowitz and Giroux (1985), Foster (1986), Quantz *et al.* (1991), and others, a theory of transformative leadership grounded in the twin concepts of critique and possibility was emerging. The increasing clarity and the developing body of literature associated with the concept of transformative leadership continue to emphasize and reinforce these two ideas. Weiner (2003) drawing, as have many others, on Freire’s work, emphasized both the individual and collective nature of transformative leadership. He wrote:

transformative leadership is an exercise of power and authority that begins with questions of justice, democracy, and the dialectic between individual accountability and social responsibility (p. 89).

The differences between the foregoing statement and Leithwood’s (2009, this volume) statement that transformational leadership has four dimensions (setting directions, developing people, redesigning the organization, and managing the instructional program) demonstrate how much the two theories have diverged in terms of central preoccupations and emphases. Although it would be wrong to conclude that either theory precludes either instructional excellence or social responsibility, there are clear distinctions between the ways in which leaders subscribing to each theory conceptualize, interpret, and approach these tasks. Weiner delineates the responsibilities of the transformative leader to instigate structural transformations, to reorganize the political space, to understand the relationship between

leaders and the led dialectically (and not hierarchically). He also calls for leaders to

confront more than just what is, and work toward creating an alternative political and social imagination that does not rest solely on the rule of capital or the hollow moralism of neoconservatives, but is rooted in radical democratic struggle (p. 97).

A fundamental task of the educational leader in this tradition is to ask questions about, for example, the purposes of schooling, about which ideas should be taught, and about who is successful. Critique lays the groundwork for the promise of schooling that is more inclusive, democratic, and equitable for more students. It is “antiracist, antisexist, antihomophobic, and responsive to class exploitation” (Weiner, p. 100).

Weiner (2003) makes the important additional point that transformative leaders always experience the challenge of having “one foot in the dominant structures of power and authority.” Where this not the case, they would not likely have attained the formal recognition as leaders that casts them as “willing subjects of dominant ideological and historical conditions” (p. 91). At the same time, transformative educational leaders work from within dominant social formations to exercise effective oppositional power, to resist courageously, to be activists and voices for change and transformation. They must be willing to

take risks, form strategic alliances, to learn and unlearn their power, and reach beyond a “fear of authority” toward a concrete vision of the work in which oppression, violence, and brutality are transformed by a commitment to equality, liberty, and democratic struggle (p. 102).

At the beginning of the twenty-first century, other educational researchers and theorists took up the task of articulating, and advocating for, transformative educational leadership (see Anderson, 2004; Dantley, 2003; Shields, 2003a; Tillman, 2005). For example, Shields (2003b) critiqued the typical silence of educational leaders that tends to pathologize differences, stating that

transformative educators and educational leaders must address issues of power, control, and inequity; they must adopt a set of guiding criteria [...] to act as benchmarks for the development of socially just education; and they must engage in dialogue, examine current practice, and create pedagogical conversations and communities that critically build on, and do not devalue, students’ lived experiences (p. 128).

Many of these theorists also cite the work of Astin and Astin (2000) who associate transformative leadership and societal change, saying:

We believe that the value ends of leadership should be to enhance equity, social justice, and the quality of life; to expand access and opportunity; to encourage respect for

difference and diversity; to strengthen democracy, civic life, and civic responsibility; and to promote cultural enrichment, creative expression, intellectual honesty, the advancement of knowledge, and personal freedom coupled with responsibility (p. 11).

Brown (2004) combines adult learning theory, transformative learning theory, and critical social theory in her development of a framework for the preparation of educational leaders that would emphasize transformative leadership. Once again, key elements of this framework include the need to address issues of power and privilege, dialog aimed at disequilibrium that results in meaningful change, and a call to activism.

The Current State of Transformative Leadership Theory: Problems and Promises

By the end of the first century of the twenty-first century, the theory of transformative leadership has been consistently articulated as a form of leadership grounded in an activist agenda, one that combines a rights-based theory that every individual is entitled to be treated with dignity, respect, and absolute regard with a social justice theory of ethics that takes these rights to a societal level. It emphasizes the socially constructed nature of society and the attendant outcome “that certain individuals occupy a position of greater power and that individuals with other characteristics may be associated with a higher likelihood of exclusion from decisions...” (Mertens, 2007: 87). Transformative leadership, therefore, recognizes the need to begin with critical reflection and analysis and to move through enlightened understanding to action – action to redress wrongs and to ensure that all members of the organization are provided with as level a playing field as possible – not only with respect to access, but also with regard to academic, social, and civic outcomes. In other words, it is not simply the task of the educational leader to ensure that all students succeed in tasks associated with learning the formal curriculum and demonstrating that learning on norm-referenced standardized tests; it is the essential work of the educational leader to create learning contexts or communities in which social, political, and cultural capital is enhanced in such a way as to provide equity of opportunity for students as they take their place as contributing members of society.

Not unexpectedly, the theory is not without its critics as well as its advocates. It is sometimes believed to be too idealistic, too demanding, and to place too much responsibility on the shoulders of educators and educational leaders for redressing global ills. Some argue that a focus on power, equity, and social justice can only occur at the expense of intellectual development and accountability.

These arguments are countered by those who posit that addressing issues of equity *is* the only way to transform education to achieve the success of all students – a goal that, although elusive, is at the heart of most current educational leadership theories. In addition, the tensions identified earlier by Weiner of needing to work within the system to effect transformation cannot be overestimated and can certainly not be discounted.

Perhaps the most salient criticism, and one that remains to be addressed in the coming years, is that although there is a considerable body of conceptual work, there is little empirical research related to transformative leadership: few studies have operationalized transformative leadership and examined its impact in real-life settings. (This is, of course, in opposition to the extensive empirical work already completed with respect to the sister theory of transformational leadership.) One exception is the nationally funded Te Kotahitanga project operating in New Zealand to effect transformation that leads to increased success on the part of Maori students (see Bishop *et al.*, 2003). There, a strong focus on leadership and power distribution, attitudes and positions, and on what might be termed critical democratic pedagogies is demonstrably overcoming the traditional achievement gaps between Maori and Pakeha (white European) New Zealanders. Other studies, in recent years, include those by Glanz (2007), Hoffman and Burrello (2004), Kose (2007), McLaughlin (1989), Shields (2009), and Theoharis (2007) – although there are still nuanced variations in the understanding and use of the theory. Much work, however, remains to be done to realize the potential of transformative leadership.

Concluding Thoughts

Unlike transformational leadership, which has the most potential to work well when the organization and the wider society in which it is embedded are synchronous, transformative leadership takes account of the ways in which the disparities in the outside world create inequities in the internal operations of educational organizations. Moreover, because these inequities are apparent almost everywhere with respect to schooling – both in developing and developed countries – many researchers and theorists advocate transformative leadership as a way forward. The dual focus on critique and promise emphasizes the need to do more than bemoan current failures. Transformative educational leaders need to act with courage and conviction to bring about the promise of education as apprenticeship that teaches students, as Greene (1988) states, to be “citizens of the free world – having the capacity to choose, the power to act to attain one’s purposes, and the ability to help transform a world lived in common with others.” Transformative educational

leadership not only works for the good of every individual in the school system; at its heart, it works for the common good of society as well.

See also: A Distributed Perspective On School Leadership and Management; Barriers to Participation in Adult Education; Curriculum and the Education of Cultural and Linguistic Minorities; Curriculum in Postcolonial Contexts; Educational Leadership and Social Capital; Educational Reform; Effective Leadership in Challenging Schools; Equity and Educational Effectiveness; Evaluation Governance and Planning; History of Educational Leadership/Management; Leaders for Productive Schools; Leadership: Authentic; Leadership: Democratic; Leadership in Diverse Cultures; Leading Diverse Schools; Leading/Managing Schools in Communities Made Poor; On School Management – International Council on School Effectiveness and Improvement, Commonwealth Council on Educational Administration and Management; Participation in Early Childhood Education and Care Programs: Equity, Diversity and Educational Disadvantage; Poverty, Effects of on Social and Emotional Development; Principals Role in Restructuring Schools; Public Policy and Inequality in Postsecondary Opportunity: Educational Statistics and the Failure of Education Reform; Reforming the American School System; Social and Cultural Capital in Education; Social and Cultural Influences on Teacher Education; System Leadership; The Big Challenge: Leadership for Inclusion; The Changing Role of University Presidents, Vice-Chancellors and Rectors; The Changing Roles of University Governing Boards and Councils; The Educational Leader's New Work; Transformational School Leadership; Understanding How Leadership Influences Student Learning; Justice and Care.

Bibliography

- Anderson, G. L. (2004). William Foster's legacy: Learning from the past and reconstructing the future. *Educational Administration Quarterly* 40(2), 240–258.
- Aronowitz, S. and Giroux, H. (1985). *Education under Siege: The Conservative, Liberal, and Radical Debate over Schooling*. South Hadley, MA: Bergin and Garvey.
- Astin, A. W. and Astin, H. S. (2000). *Leadership Reconsidered: Engaging Higher Education in Social Change*. Kellogg Foundation. <http://www.wkcf.org/Pubs/CCT/Leadership/Pub3368.PDF> (accessed August 2007).
- Bass, B. M. (1985). *Leadership and Performance Beyond Expectations*. New York: Free Press.
- Bass, B. M. and Avolio, B. J. (1994). *Improving Organizational Effectiveness Through Transformational Leadership*. Thousand Oaks, CA: Sage.
- Bates, R. (1995). A socially critical perspective on educational leadership. *Paper Presented at the Flinders University Conference on Educational Leadership*, Adelaide, South Australia, July.
- Bishop, R., Beryman, M., Tiakiwai, S., and Richardson, C. (2003). *Te Kotahitanga: The Experiences of Year 9 and 10 Māori Students in Mainstream Classrooms*. Wellington: Ministry of Education. <http://www.minedu.govt.nz/goto/tekotahitanga> (accessed August 2009).
- Brown, K. M. (2004). Leadership for social justice and equity: Weaving a transformative framework and pedagogy. *Educational Administration Quarterly* 40(1), 77–108.
- Burns, J. M. (1978). *Leadership*. New York: Harper and Row.
- Capper, C. A. (1989). Transformative leadership: Embracing student diversity in democratic schooling. ED 305 714. *Opinion Papers*.
- Dantley, M. (2003). Critical spirituality: Enhancing transformative leadership through critical theory and African American prophetic spirituality. *International Journal of Leadership in Education* 6(1), 3–17.
- Davis, S. H. (2006). Influencing transformative learning for leaders. *School Administrator* 63(8), 7.
- Day, C., Harris, A., and Hadfield, M. (2001). Grounding knowledge of schools in stakeholder realities: A multi-perspective study of effective school leaders. *School Leadership and Management* 21(1), 9–42.
- Duncan, M., Alperstein, M., Mayers, P., Olckers, L., and Gibbs, T. (2006). Not just another multi-professional course! Part 1. Rationale for a transformative curriculum. *Medical Teacher* 28(1), 59–63.
- Duncan, P. K. and Clayburn, C. (1997). Why haven't I heard from you? Evoking the voices of adult learners through transformative teaching. *Paper Presented at the Annual Meeting of the University Council for Educational Administration*, Orlando, November.
- Evans, S. D., Hanlin, C. E., and Prilleltensky, I. (2007). Blending ameliorative and transformative approaches in human service organizations: A case study. *Journal of Community Psychology* 35(3), 329–346.
- Foster, W. (1986). *Paradigms and Promises*. Buffalo, NY: Prometheus.
- Franz, N. (2002). TL2: Transformative learning and transformative leadership in successful cooperative extension staff partnerships. *Paper Presented at the ALE Conference*.
- Freire, P. (1970). *Pedagogy of the Oppressed*. New York: Herder and Herder.
- Freire, P. (1998). *Pedagogy of Freedom: Ethics, Democracy, and Civic Courage*. Lanham, MD: Rowan and Littlefield.
- Glanz, J. (2007). On vulnerability and transformative leadership: An imperative for leaders of supervision. *International Journal of Leadership in Education* 10(2), 115–135.
- Greene, M. (1988). *The Dialectic of Freedom*. New York: Teachers' College Press.
- Hoffman, L. P. and Burrello, L. C. (2004). A case study illustration of how a critical theorist and a consummate practitioner meet on common ground. *Educational Administration Quarterly* 40(2), 268–289.
- Keddie, A. (2006a). Pedagogies and critical reflection: Key understandings for transformative gender justice. *Gender and Education* 18(1), 99–114.
- Keddie, A. (2006b). Gender and schooling: Frameworks for transformative social justice. *Discourse: Studies in the Cultural Politics of Education* 27(3), 399–415.
- King, K. P. and Biro, S. C. (2000). A transformative learning perspective of continuing sexual identity development in the workplace. *New Directions for Adult and Continuing Education* 112, 17–27.
- Kose, B. W. (2007). Principal leadership for social justice: Uncovering the content of teacher professional development. *Journal of School Leadership* 17, 276–312.
- Leithwood, K. and Jantzi, D. (1990). Transformational leadership: How principals can help to reform school cultures. *School Effectiveness and School Improvement* 1(4), 249–280.
- Leithwood, K. and Jantzi, D. (1999). The relative effects of principal and teacher sources of leadership on student engagement with school. *Educational Administration Quarterly* 35(supplemental), 679–706.
- MacKinnon, D. (2000). Equity, leadership, and schooling. *Exceptionality Education Canada* 10(1/2), 5–21.
- McLaughlin, D. (1989). Power and the politics of knowledge: Transformative leadership and curriculum development for minority language learners. *Peabody Journal of Education* 66(3), 41–60.
- Mertens, D. M. (2007). Transformative considerations: Inclusion and social justice. *American Journal of Evaluation* 28(26), 86–90.
- Mezirow, J. (1991). *Transformative Dimensions of Adult Learning*. San Francisco, CA: Jossey-Bass.
- Mezirow, J. (1996). Contemporary paradigms of learning. *Adult Education Quarterly* 46(3), 58–172.
- Miettinen, R. (2006). Epistemology of transformative material activity: John Dewey's pragmatism and cultural-historical activity theory. *Journal for the Theory of Social Behaviour* 36(4), 390–408.
- Quantz, R. A., Rogers, J., and Dantley, M. (1991). Rethinking transformative leadership: Toward democratic reform of schools. *Journal of Education* 173(3), 96–118.

- Sagor, R. D. (1992). Three principals who make a difference. *Educational Leadership* **February**, 13–18.
- Sergiovanni, T. J. (1990). Adding value to leadership gets extraordinary results. *Educational Leadership* 23–27.
- Shields, C. M. (2003a). *Good Intentions Are Not Enough. Transformative Leadership for Communities of Difference*. Lanham, MD: Scarecrow.
- Shields, C. M. (2003b). Dialogic leadership for social justice: Overcoming pathologies of silence. *Educational Administrative Quarterly* **XI**(1), 111–134.
- Shields, C. M. (2009). *Courageous Leadership for Transforming Schools: Democratizing Practice*. Norwood, MA: Christopher-Gorgon.
- Starratt, R. J. (2004). The dialogue of scholarship. *Educational Administration Quarterly* **40**(2), 259–267.
- Sterling, D. R., Matkins, J. J., Frazier, W. M., and Logerwell, M. G. (2007). Science camp as a transformative experience for students, parents, and teachers in the urban setting. *School Science and Mathematics* **107**(4), 134–148.
- Stromquist, N. P. (2006). Gender, education and the possibility of transformative knowledge. *Compare* **36**(2), 145–161.
- Taylor, E. W. (2006). The challenge of teaching for change. *New Directions for Adult and Continuing Education* **112**, 91–95.
- Theoharis, G. (2007). Navigating rough waters: A synthesis of countervailing pressures against leading for social justice. *Journal of School Leadership* **17**, 4–27.
- Tillman, L. C. (2005). Mentoring new teachers: Implications for leadership practice in an urban school. *Educational Administration Quarterly* **41**(4), 609–629.
- Watkins, J. (2000). Leading via caring-healing: The fourfold way toward transformative leadership. *Nursing Administration Quarterly* **25**(1), 1–6.
- Weber, M. (1924). Legitimate authority and democracy. In Pugh, D. S. (ed.) *Organization Theory*, pp 3–15. London: Penguin.
- Weiner, E. J. (2003). Secretary Paulo Freire and the democratization of power: Toward a theory of transformative leadership. *Educational Philosophy and Theory* **35**(1), 89–106.

Further Reading

- Bennis, W. (1986). Transformative power and leadership. In Sergiovanni, T. J. and Corbally, J. E. (eds.) *Leadership and Organizational Culture*, pp 64–71. Chicago, IL: University of Illinois Press.
- Foster, W. (1989). The administrator as a transformative intellectual. *Peabody Journal of Education* **66**(3), 5–18.
- Foster, W. P. (2004). The decline of the local: A challenge to educational leadership. *Educational Administration Quarterly* **40**(2), 176–191.
- Nur-Awaleh, M. (2008). *Transformative leadership in Culturally Responsive Schools in an Urban Setting*. (2008). http://www.allacademic.com/meta/p_mla_apa_research_citation/2/7/5/5/4/p275546_index.html (accessed August 2009).

Relevant Website

- <http://www.cookman.edu> – Bethune-Cookman University, Earn your Master's in Transformative Leadership.

Strategic Leadership

B Davies and B J Davies, University of Hull, Hull, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

What do we understand by strategic leadership? Strategic leadership is not a new categorization or type of leadership; rather, it is best considered as the strategic element within the broader leadership paradigm. Initially, a definition of strategy can make use of five concepts. First, it is concerned with the idea of direction setting. To decide on the direction for the institution, it is necessary to understand its history and its current situation (Garratt, 1995: 2). Second, strategy, while very often associated with planning in traditional definitions (Fiddler, 1996), might better be thought of as a perspective, as a holistic way of looking at things. Third, strategy does not get involved in the detail day-to-day activities but is concerned with broad major dimensions of the organization. Fourth, a medium to longer term time framework is useful when considering strategy. A final useful concept is that strategy can be used as a template against which to set shorter term planning and activities (Davies, 2006, 2009).

Defining leadership presents a challenge owing to the expanding amount of literature in the field from which it is drawn. The forms of leadership are extensive and terms such as symbolic leadership, transformational leadership, learning-centered leadership, constructional-leader leadership, emotional leadership, ethical leadership, distributed leadership, invitational leadership, entrepreneurial leadership, and sustainable leadership are but a few. So where does one start? Bush and Glover (2003: 10) in their review of the leadership literature for the National College for School Leadership (NCSL) define leadership as "... a process of influence leading to the achievement of desired purposes. It involves inspiring and supporting others toward the achievement of a vision for the school which is based on clear personal and professional values." Strategic leadership can be seen to comprise elements of the traditional definitions of leadership but critically it is broad in scope, its impact is felt over long periods of time, and usually involves fundamental organizational change.

The Nature and Dimensions of Strategic Leadership

This section considers the nature and dimensions of strategic leadership by examining what strategic leaders do to promote a strategic approach. Davies (2003: 303) has

identified the problem in isolating essential characteristics of effective strategic leaders:

The difficulty in reviewing the literature, or interpreting the results of my current research, is that it is not always easy to distinguish the characteristics of 'good leadership' from those of 'strategic leadership'.

In an attempt to resolve this difficulty, it is possible to use the work of Boal and Hooijberg (2001: 516) who identify eight activities associated with strategic leadership as:

1. making strategic decisions;
2. creating and communicating a vision of the future;
3. developing key competences and capabilities;
4. developing organizational structures, processes, and controls;
5. managing multiple constituencies;
6. selecting and developing the next generation of leaders;
7. sustaining an effective organizational culture; and
8. infusing ethical value systems into the culture.

We have translated Boal and Hooijberg's eight points into the following categories, which in terms of leadership characteristics are concerned with:

- developing the strategic and organizational process(es) (1, 2, 4);
- leading and developing people (5, 6);
- developing the culture and value system (7, 8); and
- developing distinctive competences of the organization (3).

These elements are supported by other writers. Bolman and Deal (1991) suggest there are four leadership functions to serve four aspects of organization structural leadership, human resource leadership, political leadership, and symbolic leadership. There is an element of match between the two sets; for example, strategic and organizational processes could be defined as Bolman and Deal's structural leadership; similarly, the culture and value system could be represented as symbolic leadership, and developing people can be represented as human resource leadership. Bolman and Deal do not identify the importance of a strategic leader for developing core competences. Sergiovanni (1984) proposes a model of five leadership forces to explain how the head's leadership is related to excellent school performance: technical, human, educational, symbolic, and cultural. Unlike Bolman and Deal, Sergiovanni does not identify political leadership as an important element. Cheng

(2002: 56–57) based his five-dimension model of strategic leadership on a combination of the ideas of Bolman and Deal (1991) and Sergiovanni (1984). This is a useful synthesis when considering strategic leadership. He asserts that strategic leadership must have a concern for the development of each of these aspects as follows:

- developing positive relationships and enhancing staff commitment (human);
- developing clear goals, organizational structures, and holding people to account (structural leadership);
- building alliances and encouraging participation in decision making (political);
- inspiring and stimulating leaders to pursue organizational vision and building new values (cultural leadership); and
- directing the learning and teaching curriculum and professional improvement (educational leadership).

Encouraging participation in decision making could be considered as an aspect of the leader establishing a strategic process (developing the strategic and organizational process(es) from first bulleted list). Cheng highlights all the areas and gives a useful definition of ‘political’. It is important to add another point, a concern with developing networks to the original synthesis which can be summarized as, strategic leaders are concerned with:

- the strategic and organizational process;
- people;
- the culture and value system;
- distinctive competences of the organization; and
- networks.

Strategic Leaders Are Concerned with Developing Strategic and Organizational Processes

Cheng (2002: 64) asserts that “strategic leadership in an educational institution can be considered as leadership for initiating, developing and maintaining the strategic process”. What is needed then is a leader who will inspire and challenge, enacting the strategic process, not just talking about it. Hitt *et al.* (1998) argue that “organisational success depends on strategic flexibility”. This flexibility would enable a school to respond quickly to altered conditions and expectations and to take advantage of future opportunities. Boal and Hooijberg (2001: 518) pose the question “does strategic leadership matter?” Their answer is:

Strategic leadership does indeed matter. . . it seems to us the real question is not whether it matters but rather under what conditions, when, how and on what criteria.

They believe that when a leader makes a decision is just as important as what decision or action is taken. Bartunek and Necochea (2000) define the ability to take

the right action at a critical time as Kairos time. Boal and Hooijberg (2001: 528) suggest that strategic inflection points create a kairotic moment and it is during these moments that learning and change are possible if only the leader possesses the discernment to take notice and the wisdom to act.

Javidon (1991), quoted in Korak-Kakabadse and Kakabadse (1998: 10), suggests that “visioning depends on understanding existing realities (culture, history, formative context) and developing a clear sense of direction for the organisation”. As a leadership skill, it would seem that the leader’s own personal values and ideas are paramount in this process and the skill involves making it real for others. Leaders therefore need to understand themselves and the values they hold and be able to nurture quality communication. De Pree (1993: 99) classifies this lavish communication found in organizations with cultures which promote truth and do not suppress or limit the distribution of information. Making a vision real for others needs skills of conviction and passion, and it involves emotion. Boal and Hooijberg (2001: 516) state that “strategic leadership focuses on the creation of meaning and purpose for the organisation”. Strategic leadership therefore is concerned with the “development of the organisation as a whole which includes its changing aims and capabilities” (Selznick, 1984: 5).

Strategy can be defined as the art of making intelligent choices. Educational leaders need to make intelligent choices. Strategy should be a process of discovery, part of self-evaluation and self-improvement processes. Bennett (2000: 3) suggests that “vision used to be sufficient of itself, but it is no longer”. The ability to think big and imagine the impossible requires passion, in addition to vision. Establishing the processes may not be enough, strategic leaders may require a dedication beyond the normal in order to transform a culture or re-engineer the approach to learning. A personal enthusiasm or fervor may be required.

Strategic Leaders Are Concerned with Leading and Developing People

It is important to find a way to build a connection between thinking and action. The concept of a learning organization helps here: an organization of people who are attuned to changes and able to respond to them have valuable insights into how individuals and groups learn and how to convert this knowledge into organizational action. Pietersen (2002: 181) states that all learning organizations have developed a “culture of giving” which “fosters teamwork, experimentation, learning, and knowledge sharing”.

More recent leadership theories focus on transformational (Leithwood and Jantzi, 1999) and visionary leadership, which emphasize the interpersonal processes between leaders and followers. Boal and Hooijberg (2001: 526) suggest that research into transformational leaders stress upon factors such as intellectual stimulation and inspiration.

Cheng (2002: 53) found two recurring elements of leadership in various definitions: “first, leadership is related to the process of influencing others’ behaviour; and second, it is related to goal development and achievement”. This view is reflected in the previous discussion on the process of strategy; leaders need the skills to be able to influence people and their actions and they need to direct those actions through setting goals and creating meanings. This resonates with the perspective of transformational leadership, where a leader is proactive about the future view, shaping members’ beliefs, values, and attitudes while developing strategies for the future. Bass (1985) identified that transformational leaders, in educational settings, motivate people to do more than they are originally expected to do by raising their level of awareness and getting them to go beyond their own self-interest for the sake of the team or the organization. Strategic leadership, therefore, is about altering attitudes, values, and beliefs, all of which influence the culture of an organization.

Digman (1990) defines strategy as “the organisation’s pre-selected means or approach to achieving its goals or objectives, while coping with current and future external conditions”. He considers that strategy has a corporate dimension, that of achieving objectives, and as previously discussed, a functional operational dimension, but Digman also identifies that strategy has a future dimension. He contends that organizations have to consider why they exist and how they function in society. This dimension determines the relationships the organization has with those who have an interest in what the organization does and how it does it. This would form the framework within which the other types of strategies operate. The implication from this view is that many people are involved in strategic decision making at any one time. Digman refutes this since he believes that the way decisions are made means that decisions are taken at the corporate level. He believes that a decision is the “result of a stream or sequence of inputs and actions by a number of people.” Everyone shapes the decision but the managers or executives make the choice.

Hambrick (1998) argues that strategic leadership occurs in an environment embedded in ambiguity, complexity, and informational overload. It is important therefore for strategic leaders to recognize new information, analyze it, and apply it to new outcomes; in other words, leaders need the ability to learn. Boal and Hooijberg (2001: 517) call this absorptive capacity and argue that leaders “have a unique ability to change or reinforce existing action patterns” within the organization. Therefore, they should create an organizational context where learning can take place. They define the ability to change as adaptive capacity. Boal and Hooijberg (2001: 518) add a third key element to their essence of strategic leadership – that of managerial wisdom which they define as the combination of discernment and Kairos time. Discernment

is the ability to perceive variation in the environment and to develop an understanding of the social relationships and behavior of the participants. Kairos time is the critical moment (time) when key decisions can and should be made. Gardner (1985) and Sternberg (1985) talk about interpersonal intelligence or social intelligence. The key components, and therefore necessary skills, of this area are social awareness or empathy and social skill or conflict resolution. Sternberg (1985) usefully extends this definition suggesting that social intelligence is not just the ability to understand but also the ability to act on that understanding of others.

Strategic Leaders Are Concerned with Developing the Culture and Value System

Whittington (2001: 55) suggests that “being a good strategist is not necessarily enough. Leadership is about more than just fitting strategy to the . . . environment; it is about fitting yourself to the social environment”. As Dahrendorf (1959) put it, “to be successful means to be liked, and to be liked means, in many ways to be alike”. It is possible for the unliked leader to be successful in the short term, but long-term success needs the support of all, which is more likely to be given when the values are shared and those involved have like ideas, direction, and values. Bennett (2000: 1) supports this view: “to some extent the culture reflects the style and personality of the school’s leader”.

Boal and Hooijberg (2001: 528) suggest that “many of the new theories of leadership appear context free”. It would appear that they do not consider how the environment or organizational context influences the process.

In the competing values framework, Quinn (1980) argues for multiple measures of effectiveness at multiple levels of analysis. The second dimension is control, social actions focused on clarity of goals versus flexibility, social actions focused on being adaptive to people and the external environment. The model recognizes that leaders face competing demands of those with an interest in the organization. The two dimensions define four quadrants and eight leadership roles which highlight ways in which leaders can handle the competing requirements. The model highlights the necessity of a leader’s capacity to change. “Leaders who have a large repertoire of leadership roles at their disposal and know when to apply these roles are more likely to create effective change than leaders who have a small repertoire of roles and who apply them discriminately” (Boal and Hooijberg, 2001: 530).

Society has very high expectations of schools and is increasingly looking to them for moral leadership and social responsibility. Schools are seen to be the place where children are able to experience consistent moral role models (Davies and Brighouse, 2008). This has always been the case but in an increasingly diverse society, with diverse values, the expectations on school leaders are

increasing and this complexity must be reflected in the skills of the strategic leader.

Strategic Leaders Are Concerned with Developing Distinctive Competencies

Quinn (1980: 85) argues that the competing values framework reflects different perceptual biases that influence how we see social action. This idea has important relevance for schools since one of the dimensions it highlights is the internal versus the external focus. This double focus distinguishes between social actions, which satisfy internal effectiveness, criteria such as employee satisfaction, and social actions focused on satisfying external effectiveness. The external effectiveness is defined as market share and profitability, which does not easily relate to the school context but there is a tension between the internal effectiveness, as stated and the external accountability of judgments made about school effectiveness. Pietersen (2002: 9–10) identifies three strategic questions:

- What is the environment in which the organization must compete and win?
- What are the few things our organization must do outstandingly well to win and go on winning in this environment?
- How will we mobilize our organization to implement these things faster and better than our competitors?

If we translate these questions into school terms, we see that the leaders in a school need to understand the environment in which the school functions, what the school needs to focus on in order to succeed in that environment, and how leadership stimulates the participants to implement the decisions jointly arrived at. Education is inextricably linked to our understanding of intelligence and of learning. New forms of understanding are emerging and leaders must be able to devise strategies for developing learning. As Bennett (2000: 3) states “school leaders must develop a curriculum which ensures that cognitive and emotional development work in harmony”. The distinctive competency for leaders must be related to developing children’s learning.

Strategic Leaders Are Concerned with Developing Networks

Strategic leaders are not only interested in the internal functions of the organization but also the role of the organization in the community and the impact of the wider education field. These different interactions may require different behaviors. Bennett (2000: 2) suggests:

Schools have already developed far from the hierarchical model of local authority control. . .now strategic development is of less importance too. Even the typical three-year

plan, with twelve months firm and two years draft, cannot survive in an age of uncertainty. The importance of strategy will remain, but it is through relationships, and not plans that it will be exercised.

Therefore, the nature and the quality of the leader’s relationships could be critical for the school. These relationships will be both internal and external – the local community, partnerships in business, and the ability to secure funding in the latest local and national bidding rounds. Success is becoming dependent on establishing relationships and securing benefits from those relationships.

Is This All?

Boal and Hooijberg (2001: 529) suggest that “the three cornerstones of strategic leadership are the capacity to learn, the capacity to change and managerial wisdom.” Throughout this discussion, the personal qualities of leaders have been discussed, for example, the values they hold, the ability to inspire and stimulate, their social intelligence, the ability to be passionate. All of these qualities affect the way a leader learns and is able to change. Sanders (1998: 5) supports the view that strategic leaders need the ability to change and learn by asserting that mastering chaos, complexity, and change requires new ways of seeing and thinking. Whittington (2001: 43) suggests that “leaders need an enduring sense of purpose and a continuous sense of motivation”. Boal and Hooijberg (2001: 532) suggest that most leadership researchers agree that leaders need to have important interpersonal skills such as empathy, motivation, and communication. Bennett (2000: 3) expands the importance of personal values:

If moral leadership is to be exercised and pedagogy re-engineered with any degree of success, then future leaders will need a firm set of personal values. No doubt many will have their own lists, but integrity, social justice, humanity, respect, loyalty and a sharp distinction between right and wrong, will all need to be included. Strategic relationships will soon flounder unless such a value system is held with conviction and exercised on a regular consistent basis.

Although not specifically included in Bennett’s comprehensive list, social intelligence is important for strategic leadership because the process of decision making, solution implementations, and organizational improvement are rarely free of emotion. Social intelligence includes having a thorough understanding of the social context, and is defined by Gardner (1985: 9) as the ability “to notice and make distinctions among other individuals. . .in particular among their moods, temperaments, motivations and intentions”. Hence, a key component of social intelligence is the ability

to discern emotions both in self and in others. Gardner identifies this as intrapersonal intelligence and interpersonal intelligence (1993). The ability to build the involvement of others and to resolve conflicts will be increasingly vital in the context of developing strategic relationships and finding creative solutions. Bennett (2000: 4) also identifies the importance of strength and courage; “visionary projects, delivered with passion, will fail unless the leader has the ability to counter adversaries and remain confident until the conclusion has been reached”. These personal skills underpin the five dimensions of strategic leadership.

Davies *et al.* (2005) in a major national study in the UK researched how strategic leadership contributed to sustainable strategic success in schools. The schools in their research which achieved ongoing success in student achievement and organizational development have leaders who undertook certain activities and possessed certain characteristics. Their taxonomy is shown in **Table 1**.

This provides a framework for both individual leaders reflecting on their strategic-leadership skills and also a developmental framework for aspiring leaders.

Conclusion

We started this section by stating that strategic leadership needs to consider three factors: it is broad in scope; its impact is felt over long periods of time, and usually involves fundamental organizational change. The challenge for strategic leaders is to develop the five dimensions of strategic leadership in such a way to meet these three factors. However, it would be a mistake to think that operational short-term planning and management is in conflict with

longer term strategic leadership. A critical idea is that strategic leadership provides a template against which to set short-term actions and activities. Leaders and managers who are aware of the long-term direction of the organization should take short-term actions that not only deal with the immediate challenges but build capacity for longer term development. By doing this, strategic leaders can build a sustainable progress for their organizations that is rooted in the realities of the present but develops an effective vision for the future.

Bibliography

- Bartunek, J. M. and Necochea, R. (2000). Old insights and new times. *Journal of Management Inquiry* 9(2), 103–113.
- Bass, B. M. (1985). *Leadership and Performance Beyond Expectations*. New York: Free Press.
- Bennett, D. (2000). *Leadership Evidence Base: The School of the Future*. Nottingham: National College of School Leadership.
- Boal, K. B. and Hooijberg, R. (2001). Strategic leadership research: Moving on. *Leadership Quarterly* 11(4), 515–549.
- Bolman, L. G. and Deal, T. E. (1991). *Reframing Organisations*. San Francisco, CA: Jossey-Bass.
- Bush, T. and Glover, D. (2003). *School Leadership: Concepts and Evidence*. Nottingham: National College for School Leadership.
- Cheng, Y. C. (2002). Leadership and strategy. In Bush, T. and Bell, L. (eds.) *The Principles and Practice of Educational Management*, pp 51–69. London: Paul Chapman.
- Dahrendorf, R. (1959). *Class and Conflict in Industrial Society*. London: Routledge/Kegan Paul.
- Davies, B. (2003). Rethinking strategy and strategic leadership in schools. *Education Management and Administration* 31(3), 295–312.
- Davies, B. (2006). *Leading the Strategically Focused School*. London: Sage.
- Davies, B. (2009). *Essentials of School Leadership*, 2nd edn. London: Sage.
- Davies, B. and Brighouse, T. (2008). *Passionate Leadership*. London: Sage.
- Davies, B., Davies, B. J., and Ellison, L. (2005). *Success and Sustainability, Developing the Strategically Focused School*. Nottingham: National College for School Leadership.
- De Pree, M. (1993). *Leadership Jazz*. New York: Dell.
- Digman, L. A. (1990). *Strategic Management — Concepts, Decisions, Cases*, 2nd edn. New York: Irwin.
- Fiddler, B. (1996). *Strategic Planning for School Improvement*. London: Pitman.
- Gardner, H. (1985). *The Mind's New Science*. New York: Basic Books.
- Garratt, B. (1995). *Developing Strategic Thought*. London: McGraw-Hill.
- Hambrick, D. C. (1998). *Navigating Change*. Cambridge: Harvard Business School.
- Hitt, M. A., Keats, B. W., and De Marie, S. M. (1998). Navigating in the new competitive landscape. *Academy of Management Executive* 12, 22–41.
- Javidon, M. (1991). Leading a high commitment, high performance organisation. *Long Range Planning* 24(2), 28–36.
- Korak-Kakabadse, N. and Kakabadse, A. P. (1998). Vision, visionary leadership and the visioning process: An overview. In Kakabadse, A. P., Nortier, F., and Abramovici, N. (eds.) *Success in Sight*, pp 1–34. London: International Thompson Business Press.
- Leithwood, K. and Jantzi, D. (1999). Transformational school leadership effects: A replication. *School Effectiveness and School Improvement* 10(4), 451–479.
- Pietersen, W. (2002). *Reinventing Strategy Using Strategic Learning to Create and Sustain Breakthrough Performance*. New York: Wiley.
- Quinn, J. (1980). *Strategies for Change*. Homewood, IL: Irwin.

Table 1 Strategic leaders – their activities and characteristics

<i>What strategic leaders do</i>	<i>The characteristics strategic leaders possess</i>
They set the direction of the school.	They challenge and question – they have a dissatisfaction or restlessness with the present.
They translate strategy into action.	They prioritize their own strategic thinking and learning, and build new mental models to frame their own and others' understanding.
They align the people, the organization, and the strategy.	They display strategic wisdom based on a clear value system.
They determine effective strategic-intervention points.	They have powerful personal and professional networks.
They develop strategic capabilities within the school.	They have high-quality personal and interpersonal skills.

- Sanders, T. I. (1998). *Strategic Thinking and the New Science*. New York: Free Press.
- Selznick, P. (1984). *Leadership in Administration*. Berkeley, CA: University of California Press.
- Sergiovanni, T. J. (1984). Leadership and excellence in schooling. *Educational Leadership* **41**(5), 4–13.
- Sternberg, R. J. (1985). *Beyond IQ*. Cambridge, MA: Cambridge University Press.
- Whittington, R. (2001). *What Is Strategy and Does It Matter?* 2nd edn. London: Thomson Learning.

Teacher Leadership and Organizational Development

A Harris, University of Warwick, Coventry, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Leadership remains firmly centerstage in contemporary discussions about organizational change and development. It seems that no modern concept has been more powerfully received in the consciousness of those concerned with school reform and improvement than leadership. Effective leadership has long been identified with school improvement and transformation (Leithwood *et al.*, 2006; Fullan, *et al.*, 2007). Research has shown that leadership makes a difference to organizational and individual learning outcomes (Day *et al.*, 2007).

Current theorizing about leadership reinforces the view that leadership is not the preserve of one individual but is a “social influence process whereby intentional influence is exerted by one person (or group) over other people or groups to structure the activities and relationships in a group or organization” (Yukl, 1994). Such collective action resonates closely with the theory of distributed leadership which is generating much interest within the contemporary leadership field (Spillane *et al.*, 2001; Harris, 2007).

Work by Spillane *et al.* (2001) suggests that distributed leadership is a way of understanding leadership that focuses upon interaction and the exploration of complex social processes. It implies that the practice of leadership is one that is shared and realized within extended groupings and networks (Spillane *et al.*, 2001; Harris, 2004; Gronn, 2000). In this sense, leadership is best understood as “practice distributed over leaders, followers and their situation” (Spillane *et al.*, 2001: 13). This distributed view of leadership “incorporates the activities of multiple groups of individuals in a school guiding and mobilizing staff in the process of instructional change” (Spillane *et al.*, 2001:13).

Distributed leadership theory implies collaborative individualism where individuals work collaboratively to act upon and transform systems. Other researchers have similarly argued for a form of constructivist leadership (Lambert, 1998) where leadership is primarily about teachers learning together, and constructing meaning and knowledge collectively and collaboratively. Most recently, Crowther *et al.* (2009: 38) have argued for parallel leadership – a process whereby teacher leaders and their principals engage in collective action to build school capacity. It embodies mutual respect, shared purpose, and allowance for individual expression. Crowther *et al.* (2000) suggest a division of leadership responsibilities where principals assume primary responsibility for strategic leadership and

teachers assume primary responsibility for pedagogical or instructional leadership.

Implicit within current leadership theory is the central idea that leadership is something many people are able to exercise and that leadership is not the realm of certain people in certain parts of the organization (Ogawa and Bossert, 1995: 225). Teacher agency, instruction, and collaboration are implicit within the distributed span of leadership, even though this connection is rarely made (Harris and Muijs, 2004).

Teacher Leadership

The idea of teacher as leader has not only gained widespread popularity in recent years, but has also become gradually embedded in the language and discourse of organizational change and improvement. The ascendancy of teacher leadership has been prompted, in part, by new understandings about organizational development and by the new theories of leadership such as distributed leadership. Contemporary interest in teacher leadership also stems from the fact that its central tenet aligns with broader discussions of professionalism and by association, a professional model of change premised upon teacher enquiry, knowledge generation, and internal accountability (Elmore, 1990).

This article draws upon two major reviews of the teacher leadership literature to evaluate and assess the empirical strength of this knowledge base (York-Barr and Duke, 2005; Muijs and Harris, 2003). Initially, it considers different interpretations and definitions of the term teacher leadership and subsequently explores evidence relating to the impact and outcomes of teacher leadership.

Defining Teacher Leadership

It is clear from the international literature that there are overlapping and competing definitions of teacher leadership. Reviews of the literature (Muijs and Harris, 2003; York-Barr and Duke, 2004) reinforce that many conflicting and competing definitions of the term teacher leadership prevail and persist. As York-Barr and Duke (2005: 260) note, “in writing about teacher leadership, many authors readily assert its importance but usually fail to define it.” Similarly, in their review of the leadership literature Leithwood and Duke (1999: 45) state that it is important to be clear from the outset that what has been

learned about leadership in schools over the last century has not depended on any clear, agreed upon definition of the concept, as essential as this would seem at first glance. Inevitably, the lack of a precise definition has resulted in teacher leadership becoming a catch-all phrase for any form of teacher activity associated with change or development (York-Barr and Duke, 2005: 260).

In their review of the literature, Leithwood and Duke (1999) identified six categories of leadership: transformational, moral, instructional, participative, managerial, and contingency. Looking at each category, it appears that teacher leadership is most closely related to instructional and participative forms where leadership equates with the behaviors of teachers as they engage in activities directly affecting the growth of students (Leithwood and Duke, 1999: 47). In their work, Katzenmeyer and Moller (2001: 17) define teacher leaders as teachers, who are leaders within and beyond the classroom, who identify with and contribute to a community of teacher learners and leaders, and influence others toward improved educational practice. Boles and Troen, (1994: 11) characterize teacher leadership as a form of collective leadership in which teachers develop expertise by working collaboratively.

Muijs and Harris (2003) suggest that there are four discernable dimensions of the teacher-leadership role:

- The first dimension concerns the way in which teachers work with and across school boundaries and structures to establish social linkages within the community (Acker-Hocevar and Touchton, 1999). This brokering role remains a central responsibility for the teacher as leader as it ensures that links within and across schools are in place and that opportunities for meaningful development among teachers are maximized.
- A second dimension of the teacher leader role focuses upon participative leadership where teachers work collegially with other teachers to encourage the examination of instructional practices (Walsey, 1991).
- A third dimension of teacher leadership in school improvement is the mediating role. The literature suggests that teacher leaders are important sources of instructional expertise and information because they demonstrate high levels of instructional expertise (Snell and Swanson, 2000).
- The final and most important dimension of the teacher leadership-role is forging close relationships with individual teachers through which mutual learning takes place. The evidence shows that as leaders, teachers build trust and rapport with colleagues, establish solid relationships, and influence school culture through these relationships (Little, 1990, 2000).

In one of the most extensive studies on the work of teacher leaders, Lieberman *et al.* (2000) focused on what teachers actually did when they assumed leadership positions designed to provide assistance to other teachers.

This research found that the work of lead teachers was varied and largely specific to the individual context of the school. The authors concluded that restructuring school communities to incorporate leadership positions for teachers necessitated teacher leaders taking certain actions. These included placing a nonjudgmental value on providing assistance, modeling collegiality as a mode of work, enhancing teachers' self-esteem, using different approaches to assistance, making provisions for continuous learning and support for teachers at the school site, and encouraging others to provide leadership to their peers.

In summary, from the literature there are a number of important things to highlight about the definition of teacher leadership:

- First, teacher leadership is associated with the creation of collegial norms among teachers that evidence has shown can contribute to school effectiveness, improvement, and development.
- Second, teacher leadership equates with giving teachers opportunities to lead, which according to research has a positive influence upon the quality of relationships and teaching within the school.
- Third, at its most practical, teacher leadership means teachers working as instructional leaders influencing curriculum, teaching, and learning.
- Finally, teacher leadership is associated with re-culturing schools where leadership is the outcome of the dynamics of interpersonal relationships rather than just individual action.

There is an implicit assumption made in much of the literature about the positive impact of teacher leadership upon individual and organizational outcomes but the evidence of such effects, especially at the levels of classroom practice and student learning are sparse (York-Barr and Duke 2005: 282). The next section considers the evidence from the literature about the impact of teacher leadership upon organizational and individual outcomes.

The Impact of Teacher Leadership: The School Level

The findings from the literature about the impact of teacher leadership at the school level are rather variable and inconsistent. Some studies indicate strong school-level effects (Griffin, 1995) while others indicate the negligible influence of teacher leadership upon the school level. The greatest support for a relationship between teacher leadership and organizational change can be located within the school effectiveness, improvement, and reform literature. Here evidence suggests that generating teacher leadership, with its combination of increased collaboration and increased responsibility, has positive effects on transforming schools as organizations (Hargreaves, 1991; Little, 1990; Rosenholz, 1989; Little, 2000). In the

most effective schools, Pellicer *et al.* (1990) found that instructional leadership was a shared responsibility of teachers and principals. Other studies also report positive effects of teacher participation in decision making such as increased teacher motivation and decreases in teacher absenteeism (Rosenholz, 1989; Sickler, 1988).

Recent research has highlighted that an organization's ability to improve and sustain improvement largely depends upon its ability to foster and nurture professional learning communities or communities of practice (Stoll and Louis, 2007). Linked to the idea of communities of practice, school-improvement advocates have suggested that schools should operate as professional learning communities. Hargreaves (2003: 3) suggests that:

professional learning communities lead to strong and measurable improvements in students' learning. Instead of bringing about 'quick fixes' or superficial change, they create and support sustainable improvements that last over time because they build professional skill and the capacity to keep the school progressing.

This idea of professional learning communities embraces the notion of teacher leadership as it is assumed that teachers will be the catalysts for change and development within a professional learning community.

In summary, the literature confirms the positive effects of teacher collaboration upon organizational development and change. Where teacher leadership manifests itself in the development of trusting and collaborative relationships with colleagues, there is evidence that it does positively influence school culture and can generate the conditions where instructional and organizational development flourish.

The Impact of Teacher Leadership: The Teacher Level

The literature reveals that by far the most discernible and powerful effect of teacher leadership is on teacher leaders themselves. There is some evidence to support the assertion that as teachers lead, they grow in leadership skills and organizational perspectives (Ryan, 1999). Research findings also suggest that empowering teachers to take on leadership roles enhances teachers' self-esteem and work satisfaction, which in turn leads to higher levels of performance due to higher motivation, as well as possibly higher levels of retention in the profession (Katzenmeyer and Moller, 2001; Ovando, 1996).

As highlighted earlier, the nature and quality of leadership within schools has been found to be an important condition for maximizing school effectiveness and improvement. Yet, the relationship between teacher leadership and teacher effectiveness is generally implied in the literature rather than proved. The evidence shows that schools where achievement is higher tend to be more

confident in allowing teachers to take on leadership roles. As one US study showed, improving student outcomes appeared as a condition for, rather than a result of, teacher leadership (Dickerson, 1992).

Teacher leadership has been reported to have effects on teacher practices at the classroom level. Research by Ryan (1999) revealed a high level of perceived impact on instructional practices of colleagues. Smylie's (1994) review of re-designated teacher work and its effects on classroom practice drew two primary conclusions. First, that change in classroom practices was more likely to occur among those teachers whose work was redesigned that is, the teacher leaders, and second, changes in classroom practice were more likely to occur when initiatives were collective, as opposed to individual.

In summary, the findings from the available studies suggest there is evidence that teacher leadership is positively related to changes in teachers' classroom practice and their instructional effectiveness.

The Impact of Teacher Leadership: The Student Level

In their review of the literature, York-Barr and Duke (2005) found only five studies that directly examined the effects of teacher leadership on students. Research by Ryan (1999) studied three schools where it was found that the teacher leaders were perceived to be having a positive effect on students. In contrast, Louis and Marks (1998) did not discern a direct relationship between teacher empowerment and student learning but strongly supported the argument that empowerment did positively influence teachers' efforts to improve instruction. Taylor and Bogotch (1994) found no significant difference in terms of student attendance, achievement, or behavior between schools with high degrees of teacher participation in decision making as opposed to schools with low degrees of teacher participation in decision making. The other two studies conducted by Leithwood and Jantzi (1999; 2000) explored the effects of school and teacher leadership on students' engagement with school. This research found no statistically significant relationship between teacher leadership and student engagement. However it did conclude that teacher leadership outweighs principal leadership effects before taking into account the moderating effects of family educational culture.

Research by Silins and Mulford (2002) concluded that student outcomes are more likely to improve where leadership sources are distributed throughout the school community and where teachers are empowered to lead. Similarly, a study of 86 US middle schools, found that both teacher professionalism and collegial leadership were positively related to improved student outcomes (Hoy *et al.*, 1998). Other work by Lemlech and Hertzog, (1998) has suggested that encouraging teachers to take

on leadership roles positively affects self efficacy and behavior which subsequently influences student-learning outcomes. Recent research by Harris and Muijs, (2004) found an indirect relationship between teacher involvement in decision making and improved student outcomes.

In summary, while the teacher leadership literature confirms a link between teachers' collaborative practice and improved teacher behaviors, a direct link between teacher leadership and student outcomes is less strongly established.

Coda

The literature suggests that a number of ways in which teacher leadership can be developed and enhanced in schools.

- Time needs to be set aside for teachers to meet to plan and discuss issues such as curriculum matters, developing school-wide plans, leading study groups, etc.
- Rich and diverse continuous professional development is required for teacher leadership to flourish. The evidence would suggest that professional development for teacher leadership needs to focus not just on the development of teachers' skills and knowledge but also on aspects specific to their leadership role. Skills such as leading groups and workshops, collaborative work, mentoring, teaching adults, action research, collaborating with others, and writing bids need to be incorporated into professional development (and indeed initial teacher training) to help teachers adapt to the new roles involved (Katzenmeyer and Moller, 2001).
- Structured programs of collaboration or networking need to be set up to ensure that teacher leaders can fully develop their leadership potential (Clemson-Ingram and Fessler, 1997; Gehrke, 1991). Through collaborating with teachers in other schools, engaging in trialing new teaching approaches, disseminating their findings to colleagues, and engaging in action research, the potential for teacher leadership has been shown to be significantly enhanced (Darling-Hammond *et al.*, 1995).

The teacher leadership literature points toward the beneficial effects of teacher leadership; however, the empirical evidence upon which to base some of these claims remains limited. While the teacher leadership literature is vast with over hundreds of potential sources for consideration, including books, articles, chapters, and other media, there are acknowledged limitations with this particular body of literature (York-Barr and Duke, 2005; Muijs and Harris, 2003). However, despite these limitations, there is evidence to suggest that teacher leadership can have a positive impact on teachers' self-efficacy, instructional knowledge, and professional relationships. The evidence would also support the positive benefits to

other teachers and endorse a strong relationship between teacher leadership and organizational development.

See also: A Distributed Perspective On School Leadership and Management; Leadership in the Implementation of Innovations; Leadership: School Improvement; Organizational Learning in Schools; Professional Learning Community; Transformational School Leadership.

Bibliography

- Acker-Hovevar, M. and Touchton, D. (1999). A model of power as social relationships: Teachers as leaders describe the phenomena of effective agency in practice. *Paper Presented at the Annual Meeting of the American Educational Research Association*, Montreal, Quebec, Canada.
- Boles, K. and Troen, V. (1994). Teacher leadership in a professional development school. *Paper presented at the Annual Meeting of the American Educational Research Association*, New Orleans, LA.
- Clemson-Ingram, R. and Fessler, R. (1997). Innovative Programs for Teacher Leadership. *Action in Teacher Education* 19(3), 95–106.
- Crowther, F., Hann, L., McMaster, J., and Fergusson, M. (2000). Leadership for successful school revitalisation: Lessons from recent Australian research. *Paper Presenters at the Annual Meeting of the American Educational Research Association*. New Orleans, LA.
- Darling-Hammond, L., Bullmaster, M. L., and Cobb, V. L. (1995). Rethinking teacher leadership through professional development schools. *Elementary School Journal* 96(1), 87–106.
- Day, C., Sammons, P., Harris, A., *et al.* (2007). The impact of school leadership on pupil outcomes. *Interim Report to the Department for Children, Schools and Families*. London.
- Dickerson, C. (1992). *The lead teacher model: A restructured school. Paper Presented at the Annual Meeting of the North Carolina Association of Research in Education*, Chapel Hill, NC.
- Elmore, R. F. (1990). Introduction: On changing the structure of public schools. In *Restructuring Schools: The Next Generation of Educational Reform*, pp 1–28. San Francisco, CA: Jossey-Bass.
- Fullan, M., Hill, P., and Crevola, C. (2007). *Breakthrough*. Thousand Oaks, CA: Corwin.
- Gehrke, N. (1991). *Developing Teacher Leadership Skills*. ERIC. Retrieved ED330691, <http://www.eric.ed.gov> (accessed August 2009).
- Griffin, G. A. (1995). Influences of shared decision making on school and classroom activity: Conversations with five teachers. *Elementary School Journal* 96(1), 29–45.
- Hargreaves, A. (1991). Restructuring restructuring: Postmodernity and the prospects for educational change. *Paper Presented at the Annual Meeting of the American Educational Research Association*. Chicago, IL.
- Harris, A. (2004). Distributed leadership: Leading or misleading. *Educational Management and Administration* 32(1), 11–24.
- Harris, A. and Muijs, D. (2004). *Improving Schools through Teacher Leadership*. London: Open University Press.
- Hoy, W. K., Hannum, J., and Tschannen-Moran, M. (1998). Organizational climate and student achievement: A parsimonious and longitudinal view. *Journal of School Leadership* 8(4), 336–359.
- Katzenmeyer, M. and Moller, G. (2001). *Awakening the Sleeping Giant. Helping Teachers Develop as Leaders*, 2nd edn. Thousand Oaks, CA: CorwinPress.
- Lambert, L. (1998). *Building Leadership Capacity in Schools*. Alexandria, VA: ASCD.
- Lemlech, J. K. and Hertzog, H. (1998). Preparing teachers for leadership roles. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Diego, CA, April 1998.
- Leithwood, K. and Jantzi, D. (1999). The relative effects of principal and teacher sources of leadership on student engagement with school. *Educational Administration Quarterly* 35, 679–706.

- Leithwood, K., Day, C., Sammons, P., Harris, A., and Hopkins, D. (2006). *Seven Strong Claims about Successful Leadership*. London: DfES.
- Leithwood, K. and Jantzi, D. (2000). Distributed leadership and student engagement in school. *Paper Presented at the Annual Meeting of the American Educational Research Association*. San Diego, CA.
- Lieberman, A., Saxl, E. R., and Miles, M. B. (2000). Teacher leadership: Ideology and practice. In *The Jossey-Bass Reader on Educational Leadership*, pp 339–345. Chicago, IL: Jossey-Bass.
- Little, J. W. (1995). Contested ground: The basis of teacher leadership in two restructuring high schools. *Elementary School Journal* **96**(1), 47–63.
- Little, J. W. (2000). Assessing the prospects for teacher leadership. In *The Jossey-Bass Reader on Educational Leadership*. Chicago, IL: Jossey-Bass.
- Louis, K. S. and Marks, H. (1998). Does professional community affect the classroom? *American Journal of Education* **106**(4), 532–575.
- Muijs, D. and Harris, A. (2003). Teacher leadership: Improvement through empowerment. *Educational Management and Administration* **31**(4), 437–449.
- Ogawa, R. T. and Bossert, S. T. (1995). Leadership as an organizational quality. *Educational Administration Quarterly* **31**(2), 224–243.
- Ovando, M. (1994). Effects of teacher leadership on their teaching practices. *Paper Presented at the Annual Meeting of the University Council of Educational Administration*, Philadelphia, PA.
- Rosenholz, S. (1989). *Teachers' Workplace: The Social Organisation of Schools*. New York: Teachers College Press.
- Sickler, J. L. (1988). *Teachers in charge: Empowering the professionals*. *Phi Delta Kappan* **69**, 354–356; 375–376.
- Silins, H. and Mulford, B. (2002). Leadership and school results. In Kenneth, A. and Hallinger, P. (eds.) *Second International Handbook of Educational Leadership and Administration*, pp 561–612. Kluwer.
- Snell, J. and Swanson, J. (2000). The essential knowledge and skills of teacher leaders. *Paper Presented at the Annual Meeting of the American Educational Research Association*, New Orleans, LA.
- Smylie, M. A. (1995). New perspectives on teacher leadership. *Elementary School Journal* **96**(1), 3–7.
- Spillane, J., Halverson, R., and Diamond, J. (2001). Towards a theory of leadership practice: A distributed perspective Northwestern University, *Institute for Policy Research Working Article*.
- Taylor, D. L. and Bogotch, I. E. (1994). School-level effects of teachers' participation in decision making. *Educational Evaluation and Policy Analysis* **16**(3), 302–319.
- York Barr, J. and Duke, K. (2004). What do we know about teacher leadership? Findings from two decades of scholarship. *Review of Educational Research* **74**(3), 255–316.
- Yukl, G. (1990). An evaluative essay on current conceptions of effective leadership. *European Journal of Work and Organisational Psychology* **8**(1), 33–48.

Further Reading

- Lieberman, A. (1988). Teachers and principals: Turf, tension and new tasks. *Phi Delta Kappan* **69**, 648–653.
- Pechman, E. M. and King, J. A. (1993). *Obstacles to Restructuring: Experience of Six Middle-Grades Schools*. New York: National Center for Restructuring Education, Schools, and Teaching.
- Troen, V. and Boles, K. (1992). Leadership from the classroom: Women teachers as the key to school reform. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Francisco, CA.
- Wasley, P. A. (1991). *Teachers Who Lead: The Rhetoric of Reform and the Realities of Practice*. New York: Teachers College Press.

Relevant Website

<http://www.almaharris.co.uk> – Alma Harris, a leading educational writer and international researcher (the website of the author).

Understanding How Leadership Influences Student Learning

K Leithwood, University of Toronto, Toronto, ON, Canada

B Levin, University of Toronto, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This article describes a selected set of conceptual and methodological issues associated with developing a better understanding of the links between school leadership practices and student learning. We provide a general framework for thinking about these links and unravel some of the key challenges associated with each part of the framework.

A General Framework

Most efforts to understand the relationships between leaders' practices and student learning have assumed that leader effects are primarily indirect (e.g., Silins and Mulford, 2002a; Hallinger and Heck, 1999). Based on this assumption, one of the primary challenges for those trying to understand how leadership exercises its influence on students is to identify the most promising variables mediating and moderating leaders' effects. A second significant challenge is to uncover the nature of the relationships among these variables and between leaders and such variables.

Figure 1 is a framework for helping to understand the links between what leaders do and their influence on student learning. This figure indicates that leadership practices or behaviors have direct effects on potentially a wide range of variables which stand between or mediate the effects of leadership when those effects are determined to be student learning. **Figure 1** also includes a set of moderating variables. These are features of the organizational or wider context in which leaders' work that interact with leadership practices and/or mediating variables. These interactions potentially change the strength or nature of relationships between the leadership practices and mediating variables or the mediating variables and student learning.

Leadership Practices

In an extensive review of research on educational leadership up to 1999, Leithwood and Duke (1999) found evidence of five distinctly different types of leadership models reflected in that literature.

- Instructional leadership focuses on the behaviors of teachers as they engage in activities directly affecting the learning of pupils. The more fully developed models in this category (e.g., Hallinger, 2003) also include attention to broader sets of organizational variables, such as school culture or climate, thought to influence teachers' classroom practices.
- Transformational leadership focuses on the commitments and capacities of organizational members, as well as their willingness to engage in extra effort on behalf of their organizations. While the bulk of the evidence about this approach to leadership has been collected in non-school contexts (e.g., Avolio and Yammarino, 2002), educational researchers have recently begun to redress this imbalance (e.g., Nguni, 2005; Lunenburg, 2004).
- Moral leadership is concerned with the ethics and values of those exercising leadership. Specifically, it aims to clarify the nature of the values used by leaders in their decision making and how conflicts among values are best adjudicated (Begley and Johansson, 2003; Begley and Leonard, 1999). A strand within this approach to leadership specifically aims to promote democratic values and the empowerment of a large proportion of organizational members (e.g., Johansson, 2003; Starratt, 2003).
- Participative leadership shines a spotlight on group decision-making processes. Educational research inquiring about this approach builds on a strong foundation of research in other sectors dating back to seminal studies in the early 1930s (e.g., Mayo, 1933) about increases in organizational effectiveness associated with greater participation of employees in meaningful decisions about their work. The extensive body of research on teacher participation in decision making reasonably can be viewed as part of the body of evidence about this model of leadership (e.g., Conley, 1991). Rapidly growing literatures on both teacher leadership (Harris and Chapman, 2002; York-Barr and Duke, 2004) and distributed leadership (Gronn, 2002; Spillane *et al.*, 2000) are the most recent evolutions of this approach.
- Managerial and strategic leadership encompasses a range of tasks or functions found in the classical management literature (reviewed in Rost, 1991), including tasks such as coordination, planning, monitoring, and the distribution of resources. Educational literature from the United Kingdom reflects a far greater interest in this form of leadership than does the North American literature. Addressed much more extensively in the UK

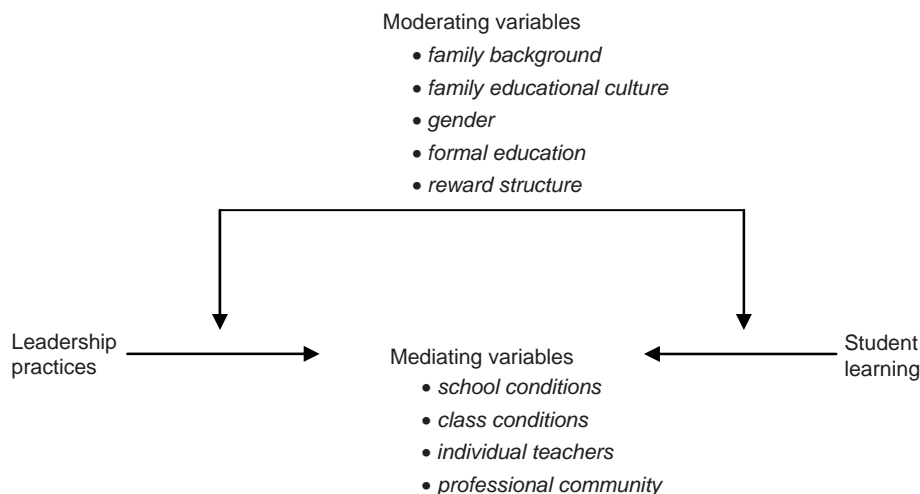


Figure 1 A framework for helping to understand how leadership influences student learning.

than in the North American literature as well is the entrepreneurial, creative, and change-oriented strategic leadership sometimes thought to be the exclusive purview of those occupying senior levels of the organizational hierarchy (Yukl and Lepsinger, 2004) (see, e.g., the special issue of *School Leadership and Management* (2004, vol. 24, no. 1) edited by Brent Davies).

- Contingent leadership emphasizes the need for leaders to be responsive to the unique demands of their organizations and the contexts in which those organizations function. While this approach is quite mature in both education and noneducation sectors (e.g., Blake and Mouton, 1964), its original conception was limited to a very small number of dimensions along which leadership styles could vary in response to context (primarily the initiation of structure and demonstrations of consideration for employees). Current leadership research continues to call for more sensitivity to the context in which leaders work and greater flexibility on the part of leaders across a much larger number of dimensions (Yukl and Lepsinger, 2004).

With the exception of instructional leadership, all of the approaches to leadership explored by the educational research community are also active areas of research in other sectors. However, academic leadership research (term used in reference to systematic, theoretically informed, empirical inquiry about leadership – as distinct from the highly popular genre of leadership literature which is autobiographical, anecdotal, and/or exclusively case based) in these other sectors reflects an additional range of approaches or models. For example, Dansereau *et al.* (1995) identify a total of 13 approaches to leadership nested within four categories they refer to as classical, contemporary, alternative, and new wave approaches.

Some education sector leadership models have been specified in detail and tested with instruments which are quite well developed. This is the case, for example, with

Hallinger's instructional leadership model (see Hallinger and Murphy, 1985), Leithwood's transformational school leadership model (Leithwood and Jantzi, 2000) and Marks and Printy's (2003) synthesis of both of these forms of leadership. Several versions of Bass' (1985) *Multifactor Leadership Questionnaire* have been used extensively to study the transformational leadership effects primarily in nonschool organizations, but in schools and districts, as well (e.g., Nguni, 2005). Many other instruments are available for measuring leadership, especially for doing so in noneducation organizations (Clark and Clark, 1990).

Without diminishing their many other contributions, one limitation of some of the most widely cited reviews of leadership effects on student learning is that they confound estimates of such effects by failing to distinguish among alternative approaches to, or models of, leadership (e.g., Hallinger and Heck, 1996b). Other reviews seem to infer comprehensive assessments of leadership effects while actually limiting themselves to the behaviors associated with a particular model of leadership (e.g., Witzier *et al.*, 2003). Furthermore, some original studies of leadership effects use secondary data sources originally created for other purposes, typically estimating the effects of potentially incoherent or incomplete models of leadership. Virtually all large-scale quantitative leadership effect studies in education restrict their attention to only part of what it is that leaders do.

These shortcomings in the actual measurement of leadership point to the importance of clearly specifying those leadership practices which are hypothesized to affect student outcomes. Failure to do this arises for both practical and conceptual reasons. Practically, available resources will often press researchers and evaluators to rely on existing evidence, evidence that is an imperfect match for their purposes. Conceptually, a major source of the problem is lack of agreement about the nature of leadership, as we discussed earlier.

Student Learning

Students' academic achievement, as it is typically measured, is just one of the several indicators of student learning. Others of a more long-term nature include graduation rates, drop-out rates, and engagement in school, for example. These two sets of outcomes are quite different. Achievement measures reflect pupils' skills and knowledge in a specific curriculum domain. Secondary school graduation rates, however, reflect not only specific curricular goals but also course selection decisions, course load, exam difficulty, and the like.

Achievement test scores are necessarily informed by pupils' entire previous school careers, as well as their personal lives. Indeed, there is a strong case to be made that the important outcomes of education are, in fact, the broader and longer-term measures such as participation in further or higher education, employment, and other measures of social participation. Many people care far more about these kinds of outcomes than they do about, for example, science test scores at age 15. Broader measures tend to present fewer data problems.

Nonetheless, the current preoccupation with student test scores, as the dependent measure of choice in inquiries about leadership effects, is not likely to go away anytime soon. So what are the challenges associated with this measure of student outcomes?

While purpose-built achievement measures could be used by researchers (although they would have their own limitations), in practice, levels of funding as well as government policies mean that most research studies end up using existing measures. These measures are typically part of national, state, or provincial student testing programs with three well-known limitations as estimates of leadership effects.

One of these limitations is the narrow focus of such testing programs. Most large-scale testing programs confine their focus to math and language achievement with occasional forays into science. Technical measurement challenges, lack of resources, and concerns about the amount of time for testing explain this typically narrow focus of large-scale testing programs. But this means that evidence of leaders' effects on student achievement using these sources is evidence of effects on pupils' literacy and numeracy. There is evidence, however, that leadership effects are of a different magnitude for even these two areas of achievement.

A second limitation of many large-scale testing programs is lack of reliability of their results at the school level. Most testing programs are designed to provide reliable results only for large groups of pupils. So as the number of pupils diminishes, as in the case of a single school or even a small district or region, few testing systems claim to even know how reliable are their results (e.g., Wolfe *et al.*, 2004). Researchers would do well to limit

analysis of achievement to data aggregated above the level of the individual school or leader.

Difficulties in estimating change are a third limitation of attempting to use the results of large-scale tests for research designed to assess leadership effects. Monitoring the extent to which a school improves the achievement of its pupils over time is a much better reflection of a school's (and leader's) effectiveness than is its annual mean achievement scores. Technically, however, arriving at a defensible estimate of such change is difficult. Simply attributing the difference between the mean achievement scores of this year's and last year's students on the province's literacy test to changes in a school's (and/or leader's) effectiveness overlooks a host of other possible explanations such as cohort differences, test differences, and differences in testing conditions.

Linn (2003) has demonstrated that these challenges to change scores become less severe as change is traced over 3 or 4 years. The lesson for those attempting to better understand leader effects is to rely on changes in student achievement over relatively long periods of time.

Variables Mediating Leadership Practices and Student Learning

We have already alluded to the commonly held assumption that the effects of leadership on student learning are largely indirect; **Figure 1** is based on this assumption. This is most obviously the case for leadership exercised by those in roles outside the classroom such as principals or head teachers. For the leadership of people in these roles to affect student learning, they must exercise some form of positive influence on the work of other colleagues such as teachers, as well as on the key conditions or characteristics of their organizations that, in turn, have a direct influence on students. These people and conditions are the mediating variables in **Figure 1**. Leaders potentially have a direct relationship or influence on these variables, and these variables, in turn, have a direct influence on student learning.

One recent review of literature summarized an extensive body of empirical evidence collected largely in North America about four categories of mediating variables which can be influenced by school leaders and which have a positive influence on student learning including school conditions (e.g., organizational culture and school structures), classroom conditions (e.g., teaching loads), individual teacher characteristics (e.g., pedagogical content knowledge) and teachers' professional community (Leithwood *et al.*, 2004). Silins and Mulford (2002b), among others, have also provided convincing evidence of the important mediating effects of organizational learning processes.

Variables Moderating Leadership Effects

The direct effects of leadership on mediating variables, as well as the indirect effects of leadership on pupils, are depressed, neutralized, or enhanced by some features of the situation or context in which leadership is exercised. For example, the same leadership practice may have quite different effects on teachers, depending on their gender, age, amount of experience, or levels of stress. So these are potential moderating variables.

Since moderating variables help explain how or why certain effects will hold, careful attention to moderating variables is a key part of understanding leadership effects on students, something that has been badly neglected in educational leadership research to date. Indeed, such inattention to moderating variables is one of the more plausible explanations for contradictory research findings and the general skepticism that often follows about the potential of research to provide clear guidance for policy and practice.

The important influence of moderators also provides an explanation for why an apparently effective leader in one school can become a struggling or ineffective leader when moved to another school. Differences in the new context react negatively to the leaders' habitual ways of doing business.

A final point about moderators concerns the basis on which a variable is assigned moderator status in a research study. The same variable might be defined moderator, mediator (or even dependent) variable status depending entirely on the theory or framework used to guide a leadership effects study. For example, teacher trust is often treated as a moderating variable in leader effects research. On the other hand, trust is also viewed as a dependent measure in leadership studies when researchers are curious about the forms of leader behaviors which promote its development (e.g., Kouzes and Posner, 1995). Trust is also conceived of as a mediating variable in studies concerned with the effects of leader behaviors on employees' acceptance of decisions (Tyler and Degoe, 1996). The theory-driven nature of moderator designation means that the examples of moderating variables provided here should be understood as an illustration of what is of theoretical interest to leadership researchers not as a category of unique variables.

The most frequently used moderating variables in studies of leadership effects on pupils are student and family background characteristics. Indicators of wealth (e.g., student eligibility for free or reduced lunch at schools) and/or socioeconomic status (e.g., parental occupation and minority status) are typically used to represent these variables. However, other moderating variables are evident in recent educational leadership effects research. For example, Leithwood and his colleagues used family educational culture as a moderating variable on the grounds that it is

the feature of student and family backgrounds that most directly influences the learning of pupils (Leithwood and Jantzi, 1999; Leithwood *et al.*, 1999). Marks and Printy (2003) incorporated into their study, as moderators, classroom compositional variables related to student gender and ethnicity.

Other examples of potential moderators can be found in recent reports of research carried out in both education and noneducation sectors. Our own scanning of such reports suggests at least five categories of promising moderators. The first category, pupils, includes a range of features associated with pupils' family background, family educational culture (OECD, 2004), and pupils' gender and ethnicity. The second category, teachers, includes gender, formal education, and tenure (age and experience have been included here). Additionally, part of this category are teacher ethnicity, beliefs and values, morale, trust, and the confidence they have in their leaders, as well as teachers' leadership prototypes. Two characteristics of leaders themselves are evident in the literature at present, leader gender and their level in the organizational hierarchy.

Five features of the organization have been conceptualized as moderators, including school size, what it is people are rewarded for, and opportunities for job enrichment. The difficulty of the tasks people are expected to accomplish, the interpersonal dynamics among people in the organization, and the availability and use of information in decision making are also identified as moderators in this category. Finally, in the category of organizational context, the nature of other stakeholders in the school or district and their relationships with the school, as well as the policy environment in which the organization finds itself, have been reported as moderators of the amount and nature of leaders' influence.

Conclusions

This article has described selected conceptual and methodological challenges associated with better understanding how and to what extent school leadership influences student learning. We offered suggestions about how each of these challenges might be addressed.

Although not discussed earlier in the article, there is an additional challenge which has been largely ignored to date and crosses the challenges we have identified to this point. It is the unit of analysis challenge. Most empirical leadership research in school contexts has treated the school as the unit of analysis. Even though it is rarely awarded serious attention, however, most of this research also has found greater within-than between-school variance in student achievement explained by leaders' influence.

These findings challenge the prevailing view that leaders have relatively uniform effects on the schools, as a

whole. There are two quite plausible explanations for the findings as well.

One explanation is that leaders treat their individual colleagues quite differently, for example, awarding much more discretion to those that they trust than those they do not. This explanation adheres closely to our actual experience in organizations and receives substantial support from a body of research, now well developed outside the education sector called the leader-member exchange theory (e.g., Erdogan and Liden, 2002; Graen and Uhl-Bien, 1998). A second explanation is that a largely common set of behaviors engaged in by leaders is interpreted quite differently by leaders' colleagues. This explanation also aligns closely with our actual life experiences (think of the different accounts provided by multiple witnesses of the same car accident). Furthermore, it has been extensively explored through information-processing approaches to leadership (e.g., Hall and Lord, 1998; Lord and Maher, 1993).

Neither leader-member exchange theory nor information-processing approaches to leadership, however, are reflected in the current corpus of evidence about educational leadership and its effects on students. Research reflecting both of these theoretical lenses would contribute substantially to our further understanding of the complex relationship between school leaders and student learning.

Bibliography

- Avolio, B. and Yammarino, F. (2002). Reflections, closing thoughts, and future directions. In Avolio, B. and Yammarino, B. (eds.) *Transformational and Charismatic Leadership: The Road Ahead*. Oxford: Elsevier.
- Bass, B. M. (1985). *Leadership and Performance beyond Expectations*. New York: Free Press.
- Begley, P. and Johansson, O. (eds.) (2003). *The Ethical Dimensions of School Leadership*. Dordrecht: Kluwer.
- Begley, P. and Leonard, P. E. (eds.) (1999). *The Values of Educational Administration*. London: Falmer.
- Blake, R. R. and Mouton, J. S. (1964). *The Managerial Grid*. Houston, TX: Gulf.
- Clark, K. E. and Clark, M. B. (eds.) (1990). *Measures of Leadership*. West Orange, NJ: Leadership Library of America.
- Conley, S. (1991). Review of research on teacher participation in school decision making. In Grant, G. (ed.) *Review of Research in Education*, pp 225–268. Washington, DC: American Educational Research Association.
- Dansereau, F., Yammarino, F., and Markham, S. (1995). Leadership: The multiple-level approaches. *Leadership Quarterly* 6(3), 251–263.
- Davies, B. (ed.) (2004). Strategy and strategic leadership in schools [Special Issue]. *School Leadership and Management*, 24(1).
- Erdogan, B. and Liden, R. C. (2002). Social exchanges in the workplace: A review of recent developments and future research directions in leader-member exchange theory. In Neider, L. L. and Schriesheim, C. (eds.) *Leadership*, pp 65–114. Greenwich, CT: Information Age Publishing.
- Graen, G. B. and Uhl-Bien, M. (1998). Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. In Dansereau, F. and Yammarino, F. (eds.) *Leadership: The Multi-Level Approaches*, vol. 24, pt. B, pp 103–134. Stamford, CT: JAI.
- Gronn, P. (2002). Distributed leadership. In Leithwood, K. and Hallinger, P. (eds.) *Second International Handbook of Educational Leadership and Administration*, pp 653–696. Dordrecht: Kluwer.
- Hall, R. J. and Lord, R. G. (1998). The role of within-individual cognitive structures in determining higher-level effects. In Dansereau, F. and Yammarino, F. (eds.) *Leadership: The Multi-Level Approaches*, vol. 24, part B, pp 159–184. Stamford, CT: JAI.
- Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education* 33(3), 329–351.
- Hallinger, P. and Heck, R. (1996). The principal's role in school effectiveness: An assessment of methodological progress, 1980–1995. In Leithwood, K. and Hallinger, P. (eds.) *International Handbook of Educational Leadership and Administration*, pp 723–783. Dordrecht: Kluwer.
- Hallinger, P. and Heck, R. (1996). Reassessing the principal's role in school effectiveness: A review of empirical research, 1980–1995. *Educational Administration Quarterly* 32(1), 5–44.
- Hallinger, P. and Heck, R. (1999). Next generation methods for the study of leadership and school improvement. In Murphy, J. and Louis, K. (eds.) *Handbook of Research on Educational Administration*, 2nd edn., pp 141–162. San Francisco, CA: Jossey-Bass.
- Hallinger, P. and Murphy, J. (1985). Assessing the instructional management behavior of principals. *Elementary School Journal* 86(2), 217–247.
- Harris, A. and Chapman, C. (2002). Democratic leadership for school improvement in challenging contexts. *Paper Presented at the International Congress of School Effectiveness and Improvement*. Copenhagen, Denmark.
- Johansson, O. (2003). School leadership as a democratic arena. In Begley, P. and Johansson, O. (eds.) *The Ethical Dimensions of School Leadership*, pp 201–219. Dordrecht: Kluwer.
- Kouzes, J. M. and Posner, B. Z. (1995). *The Leadership Challenge: How to Keep Getting Extraordinary Things Done in Organizations*, revised edn. San Francisco, CA: Jossey-Bass.
- Leithwood, K. and Duke, D. (1999). A century's quest to understand school leadership. In Murphy, J. and Louis, K. S. (eds.) *Handbook of Research on Educational Administration*, pp 45–72. San Francisco, CA: Jossey-Bass.
- Leithwood, K. and Jantzi, D. (1999). The relative effects of principal and teacher sources of leadership on student engagement with school. *Educational Administration Quarterly* 35(supplemental), 679–706.
- Leithwood, K. and Jantzi, D. (2000). *The Transformational School Leadership Survey*. Toronto, ON: OISE/University of Toronto.
- Leithwood, K., Jantzi, D., and Steinbach, R. (1999). *Changing Leadership for Changing Times*. Buckingham: Open University Press.
- Leithwood, K., Louis, K. S., Anderson, S., and Wahlstrom, K. (2004). *How Leadership Influences Student Learning: A Review of Research for the Learning from Leadership Project*. New York: The Wallace Foundation.
- Linn, R. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher* 32(7), 3–13.
- Lord, R. G. and Maher, K. J. (1993). *Leadership and Information Processing*. London: Routledge.
- Lunenburg, F. (2004). Transformational leadership: Factor structure of Bass and Avolio's MLQ in public school organizations. *Paper Presented at the Annual Meeting of the University Council for Educational Administration*. Kansas City, MO.
- Marks, H. and Printy, S. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Leadership Quarterly* 34(3), 370–397.
- Mayo, E. (1933). *The Human Problems of an Industrial Civilization*. Boston, MA: Harvard Business School.
- Nguni, S. (2005). *Transformational Leadership in Tanzanian Education: A Study of the Effects of Transformational Leadership on Teachers' Job Satisfaction, Organizational Commitment and Organizational Citizenship Behaviour in Tanzanian Primary and Secondary Schools*. Nijmegen: Radboud University.
- OECD (2004). *Learning for Tomorrow's World: First Results from PISA 2003*. Paris: OECD.

- Rost, J. C. (1991). *Leadership for the Twenty-First Century*. New York: Praeger.
- Silins, H. and Mulford, W. (2002a). Leadership and school results. In Leithwood, K. and Hallinger, P. (eds.) *Second International Handbook of Educational Leadership and Administration*, pp 561–612. Dordrecht: Kluwer.
- Silins, H. and Mulford, W. (2002b). Schools as learning organizations: The case for system, teacher and student learning. *Journal of Educational Administration* **40**, 425–446.
- Spillane, J. P., Halverson, R., and Diamond, J. (2000). Toward a theory of leadership practice: A distributed leadership perspective. *Paper Presented at the Annual Meeting of the American Educational Research Association*. New Orleans, LA.
- Starratt, R. J. (2003). Democratic leadership theory in late modernity: An oxymoron or ironic possibility? In Begley, P. and Johansson, O. (eds.) *The Ethical Dimensions of School Leadership*, pp 13–31. Dordrecht: Kluwer.
- Tyler, T. R. and DeGoey, P. (1996). Trust in organizational authorities: The influence of motive attributes on willingness to accept decisions. In Kramer, R. M. and Tyler, T. R. (eds.) *Trust in Organizations: Frontiers of Theory and Research*, pp 331–356. Thousand Oaks, CA: Sage.
- Witzier, B., Bosker, R., and Kruger, M. (2003). Educational leadership and pupil achievement: The elusive search for an association. *Educational Administration Quarterly* **34**(3), 398–425.
- Wolfe, R., Childs, R., and Elgie, S. (2004). *Final Report of the External Evaluation of the EQAO's Assessment Process*. Toronto, ON: OISE/University of Toronto.
- York-Barr, J. and Duke, K. (2004). What do we know about teacher leadership? Findings from two decades of scholarship. *Review of Educational Research* **74**(3), 255–316.
- Yukl, G. and Lepsinger, R. (2004). *Flexible Leadership: Creating Value by Balancing Multiple Challenges and Choices*. San Francisco, CA: Jossey-Bass.

LEADERSHIP AND MANAGEMENT – POLITICS AND GOVERNANCE

Contents

Equity and Educational Effectiveness

Leading/Managing Schools in Communities Made Poor

Principals Role in Restructuring Schools

School Reform and Restructuring: Self Managing School

System Leadership

Equity and Educational Effectiveness

P Sammons, University of Nottingham, Nottingham, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

CRS – Comprehensive School Reform.

Ecological fallacy – A widely recognized error in the interpretation of statistical data, whereby inferences about the nature of individuals are erroneously inferred from aggregate statistics collected for the group to which those individuals belong.

EPPE – Effective Provision of Pre-school Education Project.

NICHHD – National Institute of Child Health and Human Development.

OECD – Organization for Economic Cooperation and Development.

PIRLS – Progress in International Reading Literacy Study.

PISA – Programme for International Student Assessment – OECD.

SEN – Special educational needs.

SER – School effectiveness research.

SES – Socioeconomic status.

STAR – Student Teacher Achievement Ratio Study (Tennessee).

TIMSS – Trends in International Mathematics and Science Study.

UN – United Nations.

Development. (OECD), World Bank, etc.). Education is both affected by and influences the process of globalization in different ways in different contexts. In many societies the prime concern is to increase access to education and information technology, achieve the goal of universal primary education for all children and, in particular, improve the education prospects of girls and those living in poverty. The education of girls is seen to promote health goals for children, reduce population growth, and enhance economic prosperity in developing countries. In other contexts the prime concerns are to raise quality and standards, including increasing participation rates in further and higher education, and promoting lifelong learning.

Raising standards of achievement is seen as fundamental to economic performance and the promotion of democratic engagement in many countries. Education reform is given a high profile as governments attempt to modernize their education systems to face the challenges of the twenty-first century, and demand greater returns for their investment in education in terms of higher student achievement levels. International surveys of student achievement such as Progress in International Reading Literacy Study (also known as PIRLS), Trends in International Mathematics and Science Study (often shortened as TIMSS), and Programme for International Student Assessment (PISA) receive considerable media coverage with the creation of league tables of country results. They have become increasingly influential with governments concerned to boost both their nation's average attainment levels and reduce the variation in attainment within their country, particularly the equity gap in achievement between different groups of students. The political impact of low performance has been considerable, for example in both Denmark and Germany major reviews of their education system were conducted in response to poor

Increased awareness of the interdependence of societies, and the destabilizing impact of poverty and environmental degradation has led to a greater focus on promoting equity as a policy goal for many governments and transnational organizations (United Nations Organization (UNO), Organization for Economic Cooperation and

performance in PISA 2000 and, interestingly, both countries adopted a school effectiveness research (SER) framework to inform their reviews.

The contribution of SER to understanding variations in school performance, the role of school leadership, and its implications for promoting quality in education and raising standards are explored later in this article. First, attention is paid to define concepts and terms used in studying educational equity.

Defining Educational Disadvantage

Attempts to define equality and equity in education and the role of disadvantage draw on notions of social justice and social inclusion. Foci include:

- formal equality of access/provision;
- equality of circumstance;
- equality of participation; and
- equality of outcome.

A helpful discussion is provided by the UK government's *The Equalities Review* (2006).

In most education systems students from disadvantaged backgrounds (especially those from minority ethnic origins and those experiencing a range of social disadvantages such as low income, parents lacking qualifications, unemployed or in low socioeconomic status (SES) work, poor housing) are more likely to experience educational failure or underachievement. Nonetheless the size of the equity gap in achievement is wider in some systems than others, as is shown in international achievement studies and this is linked to broader structural features related to social processes and the characteristics of different education systems.

West and Pennell (2003) explore the concept of underachievement noting that some psychologists define this as the discrepancy between measured intelligence (IQ) and attainment in educational tests. They see this view as controversial since it proposes IQ as the main determinant of differences in achievement, and can be seen as circuitous since all IQ tests are, to some degree, tests of achievement themselves. An alternative approach is to use statistical measures of the relative differences in attainment between various social or ethnic groups as an indicator to identify relatively low-attaining groups.

Most studies in the political arithmetic tradition measure absolute differences in attainment between specific groups of students to identify the size of the equity gap. But this may not illuminate the factors that underlie such differences and their definition. Low attainment does not just relate to minority ethnic group background in a uniform way. There are significant variations in attainment between different minority ethnic groups, some outperforming majority students and others significantly

underperforming. Subgroups of students may show different profiles with black African students showing higher attainment levels than black Caribbean students, Chinese and Indian students typically showing high levels of attainment, and Pakistani or Bangladeshi students relatively low levels in the United Kingdom, for example. Within this, there can be within-group differences, the gender gap being larger for some ethnic groups. Other factors, particularly refugee status, time spent in the education system, whether first-, second-, or third-generation status are also important. Language is a powerful determinant with second-language status associated with lower attainment overall; differences exist within this category too though – bilingual students fully fluent in the majority language often show high levels of attainment whereas those not fully fluent often show lower attainment and reduced adjustment in school. The size of the equity gap differs between countries and the particular groups with low attainment vary in relation to economic, historical, and cultural factors that affect educational opportunities (tribal, religious, or caste systems often show a powerful association with achievement patterns especially in developing societies).

Factors associated with low attainment can be divided into broad categories such as individual characteristics (age, birth weight, gender); family SES characteristics (particularly family structure, parents' qualification levels, health, occupational level, whether in paid work, and income level); community and societal characteristics, both local and national (neighborhood context, cultural expectations, social structural divisions especially in relation to social class); and educational experiences related to preschool, school, and peer characteristics (including access, quantity, and quality of provision). In addition, the influence of family cultural capital, especially the powerful impact of the child's home learning environment in the early years, can be an important predictor of attainment.

Low educational attainment is linked to the concept of social exclusion where individuals or groups of people, often concentrated in specific areas or localities, experience a combination of linked problems such as unemployment, poor skills, low participation rates in further/higher education, poor housing, high crime, and bad health. A range of area-based intervention policies have been developed to address the geographical clustering of disadvantaged groups, and this often includes targeted interventions in education. In many developing countries a regional and/or a rural-urban divide in educational attainment levels and opportunities is evident, with rural poverty linked to the achievement gap; this leads to aid and regional development policies often with an education as well as an employment focus. In developed countries, a common problem involves the concentration of disadvantaged groups in inner city urban areas and this has also led to targeted educational interventions.

The strategy of positive discrimination was first introduced with Title 1 Program or war on poverty in the United States in 1965. The program intended to improve the basic skills of students from low-income groups and focused on schools with high concentrations of these students. Similar policies of education priority were followed in many western European countries including the Netherlands and England. More recent policy interventions in England with a strong area focus include Education Action Zones and Excellence in Cities involving schools, and Sure Start aimed at families with young children (Karsten, 2006).

The concept of multiple disadvantage focuses on the ways clusters of characteristics, each associated with an increased risk (likelihood) of low attainment, can have a cumulative if not a directly additive effect, while the intergenerational transmission of disadvantage is illustrated by the concept of the cycle of disadvantage, where poor opportunities in childhood lead to poor outcomes as adults including early child-bearing, lack of social mobility, and poverty thus perpetuating the divide. Multiple disadvantages can include a range of individual and family factors and also neighborhood characteristics such as living in an area with many other disadvantaged families. Low attainment is thus seen as a consequence of different combinations of disadvantage for particular groups (though this does not mean that all disadvantaged individuals have low attainment, as discussions of the ecological fallacy imply) (The ecological fallacy refers to the misinterpretation of statistical data, arising from drawing inferences about the nature of individuals from aggregate statistics collected for a group to which an individual belongs. For example, although typically there is a group difference in average attainment levels between high and low SES groups of students, some individual low SES students will show better attainment levels than some high SES students). Research indicates that multiply disadvantaged groups are significantly more likely than others to be identified as having some form of special educational need (SEN) at school and thus requiring learning support. In addition to strong links with low attainment, disadvantage (particularly multiple disadvantage) is associated with poorer educational outcomes on a range of other indicators including attendance, behavior, school exclusion, and early school-leaving.

The OECD (1995) defined those at risk as referring to children and youth who are in danger of failing at school or making a successful transition to work. Poverty is probably one of the greatest risk factors of poor life chances due to persistent associations with negative outcomes including school failure, teenage pregnancy, poor health, and violent crime. Cox (2000) argues that it is the multiplicative and interactive nature of risk factors that gives the concept of educational disadvantage its complexity.

The Role of Early Education

Neurological research indicates that extreme deprivation stunts normal brain development in very young children and supports the view that early experiences have a profound impact on later developmental outcomes. Longitudinal early intervention studies have pointed both to the damaging effects of early disadvantage on later school and adult life experiences and also to the benefits of high quality early childhood intervention programs for the most disadvantaged. (Barnett, 1995). In the 1960s, the High/Scope Perry Preschool Project targeted children from low-income African-American families and provided them with a comprehensive preschool program including education and health and family support services. Three- and four-year-old children attended the preschool program 5 days a week, two and one-half hours a day, for three-quarters of the year. Teachers also made home visits. The follow-up of the High/Scope intervention indicated lower incidence of crime, better jobs, and better social and emotional adjustment for those involved in comparison with an equally disadvantaged control group (Schweinhart *et al.*, 1993).

Sylva (2000) summarizes the benefits of good early childhood education as boosting cognitive skills that underpin academic learning personal attributes that give a child confidence, curiosity, and perseverance; and social commitment that binds the child to the community. Research by the National Institute of Child Health and Human Development (NICHD) (2002) in the United States has pointed to the benefits of high quality and longer duration of preschool for children's cognitive and several aspects of social development, while in England the Effective Provision of Pre-school Education Project (EPPE) longitudinal research indicates that preschool quality and duration are important for both improved cognitive and social behavioral outcomes, concluding that preschool is an effective intervention that helps to counteract the impact of educational disadvantage (Sylva *et al.*, 2004). In particular, good preschool experiences reduce the risk of later identification of SEN at primary school. Poor quality preschool offers few benefits by contrast, especially for disadvantaged children.

The Influence of Schools

There is increasing interest in the ways educational influences may compound existing inequalities related to student background, through unequal access to particular types of school, or differences in the quality of provision and the experience of teachers, for different groups of students. Policies of early selection and divisions into

academic and vocational schools are strongly associated with a larger equity gap in attainment, in international comparisons of PISA data. In addition, within school, differences can lead to unequal access to the curriculum through tracking, streaming, and setting arrangements resulting in concentrations of poor and minority students in lower tracks. A study of multiplying inequalities in the United States showed that unequal opportunities to learn science and maths contributed to unequal outcomes for disadvantaged groups (Oakes, 1990). School inspection has also indicated an association between poorer quality educational experiences and level of disadvantage of a school's student intake (Matthews and Sammons, 2004). High-poverty schools have more difficulty in recruiting and retaining teachers and in the United States such schools are found to employ significantly more teachers from less selective training institutions and with poorer grades (Wayne, 2002).

Other structural features are also relevant, for example, the STAR project in Tennessee pointed to benefits of smaller classes for younger age groups and particularly for disadvantaged (black and low-income) students (Pate-Bain *et al.*, 1992). Studies of summer regression in attainment indicate a sharp drop in performance in math and to a lesser extent in reading for low-income students over the summer holiday. The explanation given is that low-income students have less opportunity than others to pick up or maintain skills learnt at school. This feature is particularly marked in systems with long summer breaks, such as the United States. It has led to experiments such as summer schools, and proposals to have several short rather than one long holiday. Other developments such as full service schools (in the United States) or integrated community schools/extended schools in the United Kingdom seek to promote multi-agency working, linking health, social services, and education to support students, families, and communities and promote better educational outcomes.

Influential studies during the 1960s to early 1970s claimed that the particular school attended by a student had little influence on the educational outcome in comparison with factors such as IQ, race, and SES (Coleman *et al.*, 1966; Jencks *et al.*, 1972). The focus was thus on structural inequalities rather than on the way schools could contribute to either exacerbating or ameliorating inequities in outcomes. Subsequent research in a growing number of countries has pointed to the existence of significant school effects, while acknowledging the important influence of student background (Teddlie and Stringfield, 2000; Stevens, 2007).

SER seeks to disentangle the complex links between the student's dowry (the mix of abilities, prior attainments, and personal and family attributes) which any young person brings to school, from those of their educational experiences, and explore the way these jointly influence their later attainment, progress, and development. The main foci are the impact of social institutions

(including size of school effects), characteristics that promote better educational outcomes, the influence of context, the processes of institutional change, and the long-term impact of schooling on life chances (Sammons, 1999).

Early SER research in the United States incorporated explicit aims concerned with equity and excellence and focused on the achievement in basic skills (reading and numeracy) of poor/ethnic minority children in elementary schools. An effective school is defined as one in which the students progress further than might be expected from consideration of its intake. An effective school thus adds extra value to its students' outcomes, in comparison with other schools serving similar intakes. Effectiveness is thus a retrospective, relative concept that is both outcome- and time-specific.

A systematic meta-analysis concluded that net school effects (after control for intake) are larger for mathematics than language outcomes, and largest for studies based on composite measures of achievement. Effect sizes are generally greater in studies of developing countries where variations in quality and resources between individual institutions tend to be wide. On average, schools account for around 5–18% of the achievement differences between students after control for initial differences, but classroom level or teacher effects tend to be substantially larger than school effects. It concluded that "Schools matter most for underprivileged and/or initially low achieving students. Effective or ineffective schools are especially effective or ineffective for these students" (Scheerens and Bosker, 1997: 96).

Social and affective measure such as student attendance, attitudes, behavior, motivation, and self-esteem are also important in their own right but can act as intermediate outcomes, which affect, and can themselves be influenced by students' attainment and progress (Carneiro *et al.*, 2006; Cunha *et al.*, 2005; Feinstein, 2000). Relationships may be reciprocal. Improving a student's attainment and confidence as a learner can improve self-esteem, engagement, and attitudes to school and vice versa. Young students with low attainment are more at risk of developing poor attendance, and poor self-esteem and behavior as they grow older and move into secondary school, thus early intervention is vital, particularly for disadvantaged children.

The importance of school for emotional well-being is receiving attention. Students' perceptions or feelings of school connectedness have been shown to reduce adolescent emotional distress (Resnick *et al.*, 1997). Other research has identified relationships between students' sense of their school as a community and lower involvement in problem behaviors such as drug use and delinquent behavior and increased psychological resiliency (Battistich and Hom, 1997).

Three key questions relevant to the promotion of greater equity in education provision form the foci of much SER:

Table 1 The processes of effective schools

1. The processes of effective leadership	Being firm and purposeful Involving others in the process Exhibiting instructional leadership Frequent personal monitoring Selecting and replacing staff
2. The processes of effective teaching	Unity of purpose Consistency of practice Collegiality and collaboration
3. Developing and maintaining a pervasive focus on learning	Focusing on academics Maximizing school learning time
4. Producing a positive school culture	Creating a shared vision Creating an orderly environment Emphasizing positive reinforcement
5. Creating high and appropriate expectations for all	For students For staff
6. Emphasizing responsibilities and rights	Responsibilities Rights
7. Monitoring progress at all levels	At the school level At the classroom level At the individual level
8. Developing staff skills at the school site	Site-based Integrated with ongoing professional development
9. Involving parents in productive and appropriate ways	Buffering negative influences Encouraging productive interactions with parents

- Effective in promoting which outcomes?
- Effective for which student groups?
- Effective over what time period?

They provide a basis for monitoring at both the system level and for individual schools and can be used for school improvement planning and evaluation with the promotion of equity in mind (Sammons, 1999).

Characteristics of Effective Schools

The correlates of effectiveness identified by SER have been identified in a number of reviews (Teddle and Reynolds, 2000) and nine aspects are highlighted (see Table 1).

Characteristics of Ineffective Schools

Stoll and Fink (1996) review characteristics of ineffective schools and highlight:

- lack of vision;
- unfocused leadership;
- dysfunctional staff relationships; and
- ineffective classroom practices.

Ineffective classroom practices were characterized by:

- inconsistent approaches to the curriculum and teaching;
- generally lower expectations for students of low SES;

- an emphasis on supervising and communicating about routines;
- low levels of teacher–student interaction;
- low levels of student involvement in their work;
- student perceptions of their teachers as people who did not care, praise, provide help, or consider learning as important; and
- more frequent use of criticism and negative feedback.

The features of structured teaching have been identified as particularly relevant to promoting cognitive attainments in the basic skill areas especially for socioeconomically disadvantaged groups (Muijs and Reynolds, 2005).

The Role of School Leadership

School effectiveness research has drawn attention to the importance of school leadership as a key characteristic of effective schools (see Table 1) and leadership judged to be poor is a well-documented feature of ineffective schools according to inspection evidence in the United Kingdom (Matthews and Sammons, 2004). School improvement research has highlighted the principal's role in the turn-around of ineffective schools and its importance for schools in disadvantaged contexts. A review by Leithwood *et al.* (2006) highlighted seven strong claims about school leadership:

- school leadership is second only to classroom teaching as an influence on pupil learning;

- almost all successful leaders draw on the same repertoire of basic leadership practices; and
- school leaders improve teaching and learning indirectly and most powerfully through their influence on staff motivation, commitment, and working conditions.

The LOSLO study in Australia shows how leadership influences both organizational learning (OL) directly and, through this, teaching and learning and student outcomes (indirectly). It indicates that leadership needs to be distributed amongst a range of school personnel not just focused on the principal, and that four factors defined OL in schools:

- a trusting and collaborative climate;
- a shared and monitored mission;
- taking initiatives and risks; and
- ongoing, relevant professional development (Mulford *et al.*, 2004).

In a review for the OECD, Mulford (2003) draws attention to the increased importance attached to the role, recruitment, and development of school leaders for the improvement of education and evidence of their impact on teacher and school effectiveness and, through this, on student outcomes (see Mulford and Silins, 2010; Silins and Mulford, 2007).

While the challenges facing schools serving disadvantaged communities are greater, the characteristics of successful schools in such contexts are not radically different from those reported by SER as a whole. Research on high-achieving low-income secondary schools in Canada concludes that they improve educational outcomes by stressing clear expectations and through supportive structures and services, which motivate students and emphasize the importance of leadership (Henchey, 2001). A study of high-attaining Welsh primary schools in disadvantaged settings similarly points to important features of primary school culture and leadership (James *et al.*, 2006). In a review of improving schools in socioeconomically disadvantaged areas Muijs *et al.* (2004) argue that there are a number of strategies that support improvement in disadvantaged contexts including a focus on improving the quality of teaching and learning for all students as the core business of the school, effective distributed leadership, creating an information-rich environment, a positive school culture, learning environment, and a strong emphasis on continuous professional development.

Comprehensive School Reform (CSR) models designed to promote improvement have received considerable attention and investment in the United States. A meta-analysis shows that CSR is equally effective in relatively lower- and higher-poverty schools, and that achievement effect sizes increase over time with longer participation by schools. The strongest evidence of effectiveness was found for three models: Direct Instruction, School Development Program, and Success for All. The successful

expansion of CSR shows that research-based models of improvement can be brought to scale across many schools and varying contexts (Borman *et al.*, 2003).

Conclusions

Health, housing, income, and the home-learning environment remain powerful influences on students' educational outcomes and joined-up policies aimed at combating social exclusion are called for. Positive educational experiences in the early years seem to be particularly important in combating disadvantage. Effective schools that boost academic attainment and promote prosocial attitudes and behaviors, and a positive self-image, also have the potential to improve the life chances of disadvantaged students. SER indicates that school effects (positive or negative) are strongest for disadvantaged students and the knowledge base is particularly relevant to schools serving disadvantaged groups. School leadership is identified as a key factor associated with school effectiveness and improvement, especially important for the improvement of schools in disadvantaged contexts.

Hopkins and Levin (2000) argue that education reform in many systems could learn much from the SER tradition and evaluations of school improvement. They conclude that reform requires extra resources linked to clear plans for improvement based on the best available evidence and that policy alignment needs to be both vertical and horizontal, focused on supporting instructional goals and strategies and reducing the equity gap in achievement.

The Every Child Matters agenda in the United Kingdom and No Child Left Behind in the United States suggest a new policy commitment to promote greater equity and greater recognition of the need for additional resources and better strategies to enhance the life chances of vulnerable groups. However, controversy remains as to the extent that schools and educational interventions can be expected to compensate for persistent structural inequities in society. In England considerable success in promoting the improvement of weaker schools has followed intense policy priority and interventions, but although attainment standards have risen markedly for all groups of students the equity gap between disadvantaged and advantaged groups remains large (Sammons, 2008).

Bibliography

- Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *Future of Children* 5(3), 25–50.
- Battistich, V. and Hom, H. (1997). The relationship between students' sense of their school as a community and their involvement in problem behaviours. *American Journal of Public Health* 87(12), 1197–2001.

- Borman, G., Hewes, G., Overman, L., and Brown, S. (2003). Comprehensive school reform and achievement: A meta – analysis. *Review of Educational Research* **73**(2), 125–230.
- Carneiro, P., Crawford, C., and Goodman, A. (2006). *Which Skills Matter?* London: Centre for the Economics of Education, London School of Economics. <http://cee.lse.ac.uk/cee%20dps/ceedp59.pdf> (accessed September 2009).
- Coleman, J., Campbell, E., Hobson, C., et al. (1966). *Equality of Educational Opportunity*. Washington, DC: National Center for Educational Statistics/US Government Printing Office.
- Cox, T. (ed.) (2000). *Combatting Educational Disadvantage: Meeting the Needs of Vulnerable Children*. London: Falmer Press.
- Cunha, F., Heckman, J., Lochner, J., and Masterov, D. (2005). Interpreting the evidence of life cycle skill formation. *NBER Working Paper 11331*. London. <http://papers.nber.org/papers/w11331> (accessed September 2009).
- Feinstein, L. (2000). *The Relative Economic Importance of Academic, Psychological and Behavioural Attributes Developed in Childhood*. London: Centre for Economic Performance, London School of Economics and Political Science, University of London. <http://cep.lse.ac.uk/pubs/download/dp0443.pdf> (accessed September 2009).
- Henchey, N. (2001). *Schools that Make a Difference: Final Report Twelve Canadian Secondary Schools in Low Income Settings*. Kelowna, BC: Society for the Advancement of Excellence in Education.
- Hopkins, D. and Levin, B. (2000). Education reform and school improvement, education reform for the 21st century. *National Institute for Research Advancement Review* **7**(3), 21–26.
- James, C., Connolly, M., Dunning, G., and Elliott, T. (2006). *How Very Effective Primary Schools Work*. London: Paul Chapman.
- Jencks, C., Smith, M., Acland, H., et al. (1972). *Inequality: A Reassessment of the Effects of Family and Schooling in America*. New York: Basic Books.
- Karsten, S. (2006). Policies for disadvantaged children under scrutiny: The Dutch policy compared with policies in France, England, Flanders and the USA. *Comparative Education* **42**(2), 261–282.
- Leithwood, K., Day, C., Sammons, P., Harris, A., and Hopkins, D. (2006). *Seven Strong Claims about Successful School Leadership*. Nottingham: National College for School Leadership/DfES.
- Matthews, P. and Sammons, P. (2004). *Improvement through Inspection an Evaluation of the Impact of Ofsted's Work*. London: Ofsted/Institute of Education. HMI 2244. <http://www.ofsted.gov.uk/publications/index.cfm?fuseaction=pubs.summary&id=3696>.
- Muijs, D., Harris, A., Chapman, C., Stoll, L., and Russ, J. (2004). Improving schools in socioeconomically disadvantaged areas – a review of research evidence. *School Effectiveness and School Improvement* **15**(2), 149–175.
- Muijs, D. and Reynolds, D. (2005). *Effective Teaching: Evidence and Practice*. London: Sage.
- Mulford, B. (2003). *School leaders: Changing roles and impact on teacher and school effectiveness. Paper Commissioned by the Education and Training Policy Division*, OECD. Paris: OECD.
- Mulford, W., Silins, H., and Leithwood, K. (2004). *Educational Leadership for Organisational Learning and Improved Student Outcomes*. Dordrecht: Kluwer Academic Publishers.
- NICHHD (2002). Early child care and children's development prior to school entry: Results from the NICHHD study of early child care. *American Educational Research Journal* **39**(1), 133–164.
- Oakes, J. (1990). *Multiplying Inequalities. The Effects of Race, Social Class and Tracking on Opportunities to Learn Mathematics and Science*. Santa Monica, CA: Rand.
- OECD (1995). *Our Children at Risk*. Paris: OECD.
- Pate-Bain, H., Achilles, C. M., Boyd-Zaharias, J., and McKenna, B. (1992). Class size does make a difference. *Phi Delta Kappan* **74**(3), 253–256.
- Resnick, M. D., Bearman, P. S., Blum, R. W., et al. (1997). Protecting adolescents from harm findings from the National Longitudinal Study on Adolescent Health. *Journal of the American Medical Association* **278**, 823–832.
- Sammons, P. (1999). *School Effectiveness Coming of Age in the Twenty-First Century*. Lisse: Swets and Zeitlinger.
- Sammons, P. (2008). Zero tolerance of failure and new labour approaches to school improvement in England. *Oxford Review of Education* **34**(6), 651–664.
- Scheerens, J. and Bosker, R. (1997). *The Foundations of Educational Effectiveness*. Oxford: Pergamon.
- Schweinhart, L. J., Barnes, H. V., and Weikart, D. P. (1993). *Significant Benefits: The High/Scope Perry Preschool Study Through Age 27. Monographs of the High/Scope Educational Research Foundation, No. 10*. Ypsilanti, MI: High Scope Press.
- Silins, H. and Mulford, B. (2007). Leadership and school effectiveness and improvement. In Townsend, T. (ed.) *International Handbook of School Effectiveness and Improvement*, pp 635–658. Dordrecht: Springer.
- Stevens, P. (2007). Researching race/ethnicity and educational inequality in English secondary schools: A critical review of the research literature between 1980–2005. *Review of Educational Research* **77**(2), 147–185.
- Stoll, L. and Fink, D. (1996). *Changing Our Schools: Linking School Effectiveness and School Improvement*. Buckingham: Oxford University Press.
- Sylva, K. (2000). Early childhood education to ensure a 'fair start' for all. In Cox, T. (ed.) *Combatting Educational Disadvantage*, pp 121–135. London: Falmer Press.
- Sylva, K., Melhuish, E. C., Sammons, P., Siraj-Blatchford, I., and Taggart, B. (2004). *The Effective Provision of Pre-School Education (EPPE) Project: Technical Paper 12 – The Final Report: Effective Pre-School Education*. London: DfES/Institute of Education, University of London.
- Teddle, C. and Reynolds, D. (2000). *The International Handbook of School Effectiveness Research*. London: Falmer.
- The Equalities Review (2006). *The Equalities Review: Interim Report for Consultation*. London: The Cabinet Office.
- Wayne, A. J. (2002). Teacher inequality: New evidence on disparities in teachers' academic skills. *Education Policy Analysis Archive* **10**(30).
- West, A. and Pennell, H. (2003). *Underachievement in Schools*. London: Routledge Falmer.

Further Reading

- Equalities Review (2007). *Freedom and Fairness. The Final Report of the Equalities Review*. London: Equalities Review Cabinet Office. <http://archive.cabinetoffice.gov.uk/equalitiesreview/publications.html> (accessed September 2009).
- Mulford, B. and Silins, H. (2007). Organisational learning in schools. In Baker, E., McGaw, B., and Peterson, P. (eds.) *International Encyclopedia of Education*, 3rd edn. Oxford: Elsevier.
- Sammons, P. (2006). School effectiveness and equity: Making connections. *Keynote Paper Presented at the Annual Meeting of the International Congress for School Effectiveness and Improvement, Embracing Diversity: New Challenges for School Improvement in a Global Learning Society*, Fort Lauderdale, FL, 4th January, 2006. http://www.leadership.fau.edu/ICSEI2006/Papers/Sammons_keynote.pdf (accessed September 2009).
- Teddle, C. and Stringfield, S. (1993). *Schools Make a Difference: Lessons Learned from a 10 Year Study of School Effects*. New York: Teachers College Press.

Leading/Managing Schools in Communities Made Poor

P Thomson, Nottingham University, Nottingham, UK

© 2010 Elsevier Ltd. All rights reserved.

Across the world, compulsory schooling has historically benefited some children more than others. Despite the rise in the overall level of mass education, as evidenced in the continued elevation of the school-leaving age, the children who do not get their fair share of education are overwhelmingly from families and neighborhoods where there are low levels of both income and formal education (Anyon, 1997; Berliner, 2006; Burniaux *et al.*, 1998; Fincher and Saunders, 2001; Fitz *et al.*, 2006; Flaherty *et al.*, 2004; Foster *et al.*, 1996; Rainwater and Smeeding, 2003; Teese, 2000). Of course, education provides an avenue for social mobility for a minority of such children (Hills and Stewart, 2005; Marshall *et al.*, 1997); however, there is a persistent and stubborn nexus between social inequality and the unequal distribution of educational qualifications across the population.

The historical polarization of cities, states, and provinces (Murphy and Watson, 1994; Pacione, 1997), and waves of urban regeneration (Bryson and Winter, 1999; Lupton, 2003) and migration (Brah *et al.*, 1999), means that there are recognizable localities and neighborhood schools where there are concentrations of children living in poverty, and where the everyday norm is that of economic and social deprivation and exclusion (Gough *et al.*, 2006). However, poverty is not confined to inner cities. It is found in rural areas and regional towns (Elder and Conger, 2000; Kenway *et al.*, 2006), in indigenous communities where the cumulative effects of colonialism still take a large human toll (Rowse, 2002), and in outer urban areas, towns, and villages, where manufacturing industries have decamped to more profitable climes, leaving their workforce behind (Allen *et al.*, 2000; Fine and Weis, 1998). However, all of these locations, despite their differences, do share some common characteristics – as do their citizens and their schools. These commonalities – the patterns produced by social, economic, and cultural structures and changes – provide the basis for this article.

Poverty creates particular challenges for both positional leader/managers and functional leaders (Kugelmass, 2004) who are charged with reforming the schools that serve neighborhoods made poor. These are:

- the challenges of dealing with the everyday realities of lives in poverty;
- the challenges of the material realities of disadvantaged schools; and
- the challenges of making a difference.

This article addresses each in turn and subsequently considers some consequences for leader/managers.

The Challenges of Working with and for Poor Children and Families

Living in poverty has both material and relational effects on the people who endure it as their everyday experience (see graphic details in Charlesworth, 2000; Peel, 2003; Shirk *et al.*, 1999). We are accustomed to think of these effects as hunger and ill health, but Lister (2004) points out that unacceptable hardship also produces disrespect, humiliation, shame and stigma, assault on dignity and self esteem, othering, denial of human rights, diminished citizenship, lack of voice, and powerlessness. Wilkinson (2005) suggests that these experiences are the consequences of inequality, rather than poverty *per se*.

Children and young people who live with poverty/inequality cannot leave these feelings and experiences at the school gate. They are inevitably in school. Teachers and school leaders respond to the realities of lives in poverty in various ways, not all of them positive or productive. One of the most consistent research findings in schools serving families and communities made poor is that educators (including policy makers) mobilize a deficit discourse to describe poverty/inequality effects (Comber, 1998; Shields *et al.*, 2005; Valencia, 1997). Two of the most commonly held beliefs are that:

1. Poverty is due to laziness and “they” could get jobs if they wanted. The reality is that many families depend on below-poverty-line welfare payments (Saunders, 1996; UNICEF, 1996) and poorly paid semiskilled work (Ehrenreich, 2001; Newman, 1999; Toynbee, 2003). Such poverty is often persistent and hard to shift (Mishel *et al.*, 2005).
2. Parents do not value education; they do not have books at home and cannot/will not help their children with their homework.

Yet, studies (Connell *et al.*, 1982; Duncan and Brooks-Gunn, 1997; Ellsworth and Ames, 1998; Newman, 1999) consistently show that the vast majority of poor families understand the importance of education and schooling success and do try to help their children stay at school and do well.

Deficit views translate into educational solutions which are at best remedial, and which at worst simply reproduce the inequitable achievement gap (Connell, 1993). Children who seem to take more time to learn are slowed down further through the use of instructional methods which use small, linear steps to teach concrete functions (Anyon, 1980). At the same time, their more

affluent counterparts have engaging lessons which require the use of complex concepts and creative tasks. The decelerating pedagogies of poverty (Haberman, 1991) actively widen the achievement gap.

Leaders/managers who work from a deficit position do establish classroom reform, welfare, and counseling programs (see Payne, 1998), but fail to build connectivity (Deforge and Abouchaar, 2003) between school and home. Communication with families is characterized by disrespect and paternalism (Tatto *et al.*, 2001), and necessary transactions are enacted with an eye on efficiency (de Carvalho, 1997), rather than on building relationships. Parents/carers are helped, while compared unfavorably with middle-class families who are now highly skilled in maintaining the benefits they accrue from the achievement gap (Ball, 2003; Devine, 2004; McGrath and Kuriloff, 2004; Power *et al.*, 2002).

Research suggests that working-class children have word, number, image, and knowledge resources and practices equal, but different to those of their middle-class peers (e.g., Comber *et al.*, 2001; Dyson, 1993; Freebody *et al.*, 1995; Hill *et al.*, 1998; Moll, 1992). This body of work argues that parents are unable to do the kind of complementary education work on which success at school largely depends (Griffith and Smith, 2005) because the school does not recognize their specific resources and practices. What is required, therefore, is a not only a shift to inclusive school approaches to home-school relations and classroom pedagogies but also school and system tests which measure what all children know, rather than what some do not (Cairney, 2002; Comber *et al.*, 2001; Daspt and Weaver, 2000; Finn, 1999).

Leader/managers with a socially just approach set out to counter deficit views. They establish ethnographically oriented professional development, helping teachers to get to know their students and the complex, positive, and negative realities of their everyday lives (Frank, 1999; Gordon, 2002). They support teachers to find, value, build on, and incorporate family and community knowledges, understandings, ethical practices, and literacies into the school curriculum and pedagogies (Brooke, 2003; Gonzales *et al.*, 2005; Johnson *et al.*, 2005; Kamler and Comber, 2005; Romano and Glascock, 2002). They see the school as an important site for building social capital within the community and work with community leaders to mobilize community strengths and assets (Driscoll and Kerchner, 1999; Gold *et al.*, 2005; Kilpatrick *et al.*, 2002). Leader/managers committed to this approach allocate funding for specific community liaison staffing, offer school premises for community use and dedicated rooms for parent and community meetings, and allocate some of their own time to making connections with families and community leaders (Allen and Martin, 1992; Epstein, 1994; Pettit, 1980). However, this kind of community orientation is not easy to maintain at the same time as building an achievement culture, which ultimately

means that young people must physically and socially move away from the neighborhoods in which they grow up (Miron and Lauria, 2005).

Within the context of this broader curriculum and community development approach, the specific needs of some troubled and stressed lives can be attended to, and integrated health, welfare, and educational services can be tailor-made for specific children and groups (Thomson, 1999). However, care needs to be taken that interagency approaches do not reinforce welfarism, the domination of professionals over those that they aim to assist (see contrasting approaches in Anning *et al.*, 2006; Williams and Pritchard, 2006). Leader/managers may also find that they spend significant amounts of time in interagency partnership meetings out of their schools. This means that other staff must take up the leadership/management slack. While this can be good experience for middle managers, it can also lead to resentment about the invisible boss. Leader/managers must carefully mediate these external and internal pulls and pushes, and make the benefits of external connections and projects very clear and demonstrable to children.

The Challenges of Everyday Life in Disadvantaged Schools

Schools serving neighborhoods made poor are not on an equal footing with their more comfortable counterparts. Despite the unique nature of each school there are common patterns brought about by the conjunction of school mix (Thrupp, 1995), public policy agendas, and neighborhood effects (Thomson, 2000).

Family unemployment/underemployment/tenuous employment (the local working out of global economic shifts and macro- and microeconomic reform policies of the national and state governments) plays out in schools in a range of ways – many students have no money for educational expenses; some children are the only ones in their household who have to get up regularly everyday and frequently do not; some parents show acute levels of anxiety and depression arising from their lack of security – this is manifest in health problems and angry outbursts in schools. There are young people who firmly believe that they will not find any work at all, and they see little point in compliance to school rules or in application to curriculum. This produces significant behavior problems which are seldom adequately addressed by vocational programs (Ball and Lamb, 2001). There are also often regular and disillusioned unemployed visitors to schools, urging their younger counterparts to join them in the mall. Many students also have to work part time and these wages are often a necessary addition to the family income; however, casual work requirements of employers generally conflict with homework and regular attendance requirements.

The increase in the numbers of diverse families, many of whom are dependent on government income support, particularly those for lone parents and those designated as the working poor with young children, are manifest in disadvantaged schools. Significant numbers of children have to deal with the pressures of living in domestic situations which are not only unstable, but also strained from serious financial worries. Disadvantaged schools report regular incidences of “staying with grandma,” and children moving between family members. There is often significant family and student transience, and some children exhibit evidence of considerable insecurity and anxiety.

Low-income neighborhoods often become trapped in vicious downward spirals, in which those who are able to move out, do, and those who have no choice stay put. Small businesses close, and those who move into the area do so because it is all they can afford. Such areas often have a large quantum of housing funded from the public purse, some of which may be available for emergency shelter. Such neighborhoods also often become home to continuing intakes of refugees fleeing intolerable political situations. Schools in such localities face additional demands for translation and interpreting and support in a range of welfare- and immigration-related activities (Thomson *et al.*, 2004); in addition, there are sometimes significant shortfalls caused by the levels of resources available compared to the demand for services.

Such circumstances have a major impact on the schools that serve poor communities. They must do more with less (see Abu El-Haj, 2006; Kozol, 1991; Lupton, 2004; Riddell, 2003; Thomson, 2002). They are often financially worse off than schools that service more comfortable neighborhoods. They can raise little money from parents through voluntary contributions, and from fundraising and sponsorship. Their buildings are more likely to be aging and unable to cope with the demands of contemporary curricula. In addition, because they are located in neighborhoods where many experience the alienating and depowering effects of poverty/inequality, they are also more likely to experience ongoing vandalism which eats away cash which might otherwise be directed to major developments. There are other costs too. Working with children with stressed lives is tiring. Working with children who are more likely to be unhealthy might mean that teachers are more exposed to infectious diseases. Schools are likely to have higher staffing bills, resulting from illness and stress. A proportion of children who suffer poverty/inequality require a great deal of time and attention. This is time given by teachers, and by those in formal leadership positions, as well as those responsible for counseling and welfare. Making this time available inevitably means some redirection of instructional, curriculum development, and school reform time to discipline and welfare matters (Manicom, 1995). In order to maintain good social order and to meet the needs of particular

children, disadvantaged schools are likely to spend more on remedial and counseling support staff than that allowed for through targeted poverty funding.

Leaders/managers in disadvantaged schools also often have to manage a series of short-term policy programs intended to support change. While additional funds are welcome, multiple accountabilities for multiple programs with different time frames – each of which seeks to demonstrate that it is their particular intervention which produces change – are not. Often, such policy strategies are competitive, causing leader/managers in disadvantaged schools to spend disproportionate amounts of time on entrepreneurial activities as they attempt to garner more funds and assistance for their schools.

Schools in disadvantaged areas also have considerable difficulty with teacher turnover, which in turn results in: difficulties in pacing, sequencing, and consolidating children's learning; planning and sustaining reform; and building the trust of children and the families in their school. Senior manager/leaders must therefore be highly skilled in staff recruitment, selection, and retention. They must find and keep key staff, such as heads of department and team leaders, who do not possess rigid and reductive approaches to curriculum, pedagogy, and assessment. These key middle leaders have day-to-day responsibility for the ongoing work of adapting, developing, and trialing approaches to learning and teaching that do not simply reproduce unacceptable educational differences. They are essential partners in the collective achievement of educational/instructional reforms and school redesign (Gunter, 2005; Spillane, 2006) and must be able to cope with the inevitable ambiguities and tensions of such work (McInerney, 2005).

Finally, leader/managers must live with and plan for the reality that at least some of their most experienced teacher leaders will leave on a regular basis for more senior positions in other schools and in the wider system. As Hargreaves and Fink (2006: 48) puts it, such schools face an ironic change dynamic in which either the sustainability of innovation is sacrificed, as key leaders move out to carry the reforms to other parts of the system, or reform across the system is prevented by maintaining the presence of creative leaders within the single school. The role of disadvantaged schools in producing not only leaders, but also reforms for the wider system, is rarely acknowledged in policy and therefore not remunerated. It remains a necessary reality for leader/managers to anticipate and plan for in their staffing strategies in order to sustain change.

The Challenges of Making a Difference

Contemporary policies which frame the work of leaders/managers in disadvantaged schools focus heavily on the

ways in which education can assist people in poverty to climb out of the poverty trap and into new forms of knowledge and service work. The solution for people living in conditions of poverty seems to be that their schools must do better.

There is significant debate about how much this can be done. Studies on school effectiveness and school improvement have demonstrated that school effects make a relatively small difference in the distribution of achievement. Classroom effects are greater, but the income–parent education nexus is still a powerful indicator of likely educational achievement (Slee *et al.*, 1998; Stoll and Myers, 1998; Teddlie and Reynolds, 2000). Simply raising the level of achievement of children in poverty is also insufficient if the levels of achievement of the remainder of the population increase at a greater rate. The gap actually widens (see Power *et al.*, 2003).

Some educational researchers suggest that the answers to closing the gap might lie outside education in employment, regeneration and income support, housing, transport and health policies, and politics (Anyon, 1997, 2005; Berliner, 2006; Levin, 2003, 2006; Rothstein, 2002; Thomson, 2002). However, recognizing that there might be limitations to what can be done in schools does not mean that the task of improving students' learning and formal outcomes in disadvantaged schools should be abandoned (West and Pennell, 2003; Whitty and Mortimore, 1997). Rather, it is a matter, as Thrupp (1999) suggests, of being realistic about what can be achieved.

The task for leaders/managers in disadvantaged schools is to maintain a sense of efficacy while also being sanguine about what can be achieved. They must help the staff to maintain a collective and individual pragmatic optimism, an equilibrium which is neither wildly idealistic (translated into inappropriate demands on students and themselves) nor cynically negative (translated into lowered expectations and resistant divisive behaviors with colleagues) (Ainscow and West, 2006; Harris *et al.*, 2006). In England, and some parts of Canada and the USA, where there are narrow targets for learning outcomes on standardized tests and exams (see debate on targets – Fullan, 2006; Hargreaves and Fink, 2006) and media and policy naming and shaming of schools that allegedly fail (Fink, 1999; Hargreaves, 2004; Johnson, 1999; O'Connor, *et al.*, 1999), this is no easy matter.

Leader/managers can take some comfort in taking a modest approach to reform. Even small improvements are significant when translated into individual children's profiles and trajectories. They can be the difference between passing and failing, between getting and not getting the grade required to get into the vocational course of choice, and between getting into university and not. Measuring the achievement gap at the macro level can obscure the micro material benefits that can result when apparently paltry shifts are made.

Implications for Leader/Managers

Leaders/managers must have knowledges and practices that are up to the task of closing the gap – even if it is only a little.

Studies of unsuccessful heads (e.g., Baker, 2005) suggest the adverse impacts of: inconsistency in approach to students; poor communication; lack of attention to basic management and forward planning; refusal of the public symbolic acts of positional leadership (assemblies and so on); and unwillingness to face problems. By contrast, writings by and about successful leader/managers (Hampton and Jones, 2000; Monroe, 1997; Stubbs, 2003; Winkley, 2002) show a complex tangle of school, the everyday lives of children and their families, and broader social, political, cultural, and economic forces. Successful leaders demonstrate, in the face of adversity, a notable sense of agency and purpose which is profoundly ethical, and just (Scapp, 2006; Starratt, 2003).

There are a number of typologies of leadership practices for schools serving deprived communities (e.g., Harris and Chapman, 2002; Keys *et al.*, 2003; Leithwood and Steinbach, 2002; Ofsted, 2000), as well as various theorizations of leadership practices (e.g., Brighouse and Woods, 1999; Lingard *et al.*, 2003; National College for School Leadership, 2005; Ryan, 2003, 2005; Shields, 2003; Shields and Edwards, 2005). Headteachers in London (National College for School Leadership, 2005) examined and named their own key knowledges and practices as: enduring resilience, focus and simplicity; community engagement; open and connected leadership; leading learning; filtering, judging and acting; accountability and consistency; and purposeful influencing. They highlighted personal qualities, such as courage and conviction, as well as professional practices focused on organizational and classroom learning and keeping the bureaucracy on side (cf Haberman, 1999; Lyman and Villani, 2004).

Starting from the cultural diversity of students, Shields (2003) takes the view that leaders must develop cultural expertise that is both intellectual and lived. She argues for a headship that is as theoretically sophisticated as it is practically sensitive. She suggests that heads need to understand debates about justice and multiculturalism and to explicitly position themselves and their actions in the school in an ethicopolitical manner. She promotes the idea of grounded dialog and thinking of the school as a public sphere (Shields and Edwards, 2005). Ryan (2003, 2005) argues not for distributed leadership, but for anti-hierarchical schools in which the different contributions of people with different skills and perspectives are honored. He outlines levels of potential involvement, the different roles that students, parents, and teachers might take, and possible areas of participation. He suggests that questions of inclusion and diversity, again related to the populations of urban schools, are keys to successful reform.

In case this is all too serious, Peacock (2005), writing about the process of getting a school out of special measures, points to the importance of a sense of fun and some joy.

The teachers were exhausted. I gave them non teaching time and quality feedback so that they knew they had a voice and were being listened to. Whereever possible their requests for new resources were responded to and the school began to feel more loved.

The children were demotivated. Determined to show them that things could be different, we brought in drama specialists, artists, a teenage rock band, anyone we could think of to enliven the atmosphere, everyone needed to know that this school could be an exciting place to be. These visits gave the staff the opportunity to witness their children responding differently and provided rich CPD opportunities that were nothing to do with criticism of their teaching or providing targets for improvement. (Peacock, 2005: 6)

Peacock's example clearly shows how leading and managing must be thought and done together, and how an emphasis on the feelings and beings of the teachers and children are an important avenue to addressing teacher's deficit views of children, and the children's sense of alienation.

However, it is critical that these theorizations of leadership are not translated into the notion of the rescuing charismatic hero/heroine. Researchers in the US (Spillane *et al.*, 2003; Spillane and Zoltner-Sherer, 2004) and Australia (Gronn, 2003) have demonstrated something of what is actually involved in distributing leadership. Halverson (2003), for example, focuses on the role of policies, practices, and programs as the activities which do the work of initiating and sustaining distribution. Spillane *et al.* (2003) highlight structures such as meeting agendas and timetables which serve to construct and regulate activity through framing and supporting the ways in which people come together regularly to accomplish particular tasks. It is systems such as these which are potentially associated with what Leithwood and Reihl (2003) call building organizational structure and culture – and what many heads would simply call management. Money, time, meetings, information, policies, publications, rewards, and sanctions are important sites for a head teacher's change practices. Working on management issues is often a key activity for heads in schools that are judged to be in serious weaknesses or in special measures. National College for School Leadership (NCSL) associate Brown (2002) looked at the first 100 days of a head in a such a situation and noted how much of her time had to be spent simply making the school functional by addressing secretarial support, plant, and resourcing.

It is important to remember that the work that leaders/managers in disadvantaged schools find most rewarding is

educational reform and community development. The burdens of managing scarce resources and fielding demands for tracking and monitoring of multiple projects take too much of their time, and they welcome the opportunities they get for engagement with the real work of educational change. Finding ways to minimize administrative tasks through the use of staff other than teachers, or through redesigning the leadership/management systems of the school (Thomson and Blackmore, 2006), are critical to leader/managers being able to maintain a focus on the prime purpose for their work – that of making some difference.

Achieving this kind of agency is difficult. Current research issues document the increasing weight of managerial responsibilities on all leader/managers (Blackmore and Sachs, 2008; Hartley, 1997; Wright, 2001), suggesting that the profession is increasingly reluctant to take on the task of leadership/management in adverse circumstances. This is particularly true in schools that serve neighborhoods made poor, where there is evidence that they receive fewer applications for leadership vacancies than other schools (Barty *et al.*, 2005; Pounder and Merrill, 2001). Given the importance of leadership in school reform, this shortage of applicants bodes poorly for the prospects of closing the gap. The looming leadership/management supply issue is one which families and children at the bottom of the socioeconomic pecking order depend on us to resolve.

Bibliography

- Abu El-Haj, T. (2006). *Elusive Justice. Wrestling with Difference and Educational Equity in Everyday Practice*. New York: Routledge.
- Ainscow, M. and West, M. (eds.) (2006). *Improving Urban Schools. Leadership and Collaboration*. Buckingham: Open University Press.
- Allen, G. and Martin, I. (eds.) (1992). *Education and Community. The Politics of Practice*. London: Cassell.
- Allen, J., Massey, D., and Cochrane, A. (2000). *Rethinking the Region*. London: Routledge.
- Anning, A., Cottrell, D., Frost, N., Green, J., and Robinson, M. (2006). *Developing Multiprofessional Teamwork for Integrated Children's Services*. Buckingham: Open University Press.
- Anyon, J. (1980). Social class and the hidden curriculum of work. *Journal of Education* 162(1), 67–92.
- Anyon, J. (1997). *Ghetto Schooling. A Political Economy of Urban Educational Reform*. New York: Teachers College Press.
- Anyon, J. (2005). *Radical Possibilities. Public Policy, Urban Education and a New Social Movement*. New York: Routledge.
- Baker, B. (2005). *Transforming Schools. Illusion or Reality*. Stoke-on-Trent, Trentham.
- Ball, K. and Lamb, S. (2001). *Participation and Achievement in VET of Non-Completers of School. LSAY Report 20*. Melbourne, VIC: Australian Council for Educational Research.
- Ball, S. (2003). *Class Strategies and the Education Market. The Middle Classes and Social Advantage*. London: Routledge Falmer.
- Barty, K., Thomson, P., Blackmore, J., and Sachs, J. (2005). Unpacking the issues: Researching the shortage of school principals in two states of Australia. *Australian Educational Researcher* 32(3), 1–14.

- Berliner, D. (2006). Our impoverished view of educational research. *Teachers College Record* 108(6), 949–995.
- Blackmore, J. and Sachs, J. (2008). *Performing and Reforming Leaders: Gender, Educational Restructuring and Organisational Change*. New York: State University of New York Press.
- Brah, A., Hickman, M., and Mac An Ghaill, M. (eds.) (1999). *Global Futures. Migration, Environment and Globalisation*. London: St Martins Press.
- Brighouse, T. and Woods, D. (1999). *How to Improve Your School*. London: Routledge.
- Brooke, R. (ed.) (2003). *Rural Voices: Place-Conscious Education and the Teaching of Writing*. New York: National Writing Project and Teachers College Press.
- Brown, P. (2002). *The First 100 Days: An Enquiry into the First 100 Days of Headship in a Failing School*. Nottingham: National College for School Leadership.
- Bryson, L. and Winter, I. (1999). *Social Change, Suburban Lives. An Australian Newtown 1960s to 1990s*. Sydney, NSW: Australian Institute of Family Studies, Allen and Unwin.
- Burniaux, J. -M., Dang, T. -T., Fore, D., et al. (1998). *Income Distribution and Poverty in Selected OECD Countries. Economics Working Papers. No 189*. Paris: OECD. http://www.oecd.org/eco/working_papers (accessed May 2009).
- Cairney, T. (2002). Bridging home and school literacy: In search of transformative approaches to curriculum. *Early Child Development and Care* 172(2), 153–172.
- Charlesworth, S. (2000). *A Phenomenology of Working Class Experience*. Cambridge: Cambridge University Press.
- Comber, B. (1998). The problem of 'background' in researching the student subject. *The Australian Educational Researcher* 25(3), 1–21.
- Comber, B., Badger, L., Barnett, J., Nixon, H., and Pitt, J. (2001). *Socio-Economically Disadvantaged Students and the Development of Literacies in School. A Longitudinal Study* vols. 1–3. Adelaide, SA: Department of Education, Training and Employment (SA) and University of South Australia.
- Comber, B., Thomson, P., and Wells, M. (2001). Critical literacy finds a "place": Writing and social action in a low income Australian grade 2/3 classroom. *Elementary School Journal* 101(4), 451–464.
- Connell, R. (1993). *Schools and Social Justice*. Toronto, ON: Our Schools/Ourselves Foundation. Pluto Press.
- Connell, R., Ashenden, D., Kessler, S., and Dowsett, G. (1982). *Making the Difference. Schools, Families and Social Divisions*. Sydney, NSW: Allen and Unwin.
- Daspit, T. and Weaver, J. A. (eds.) (2000). *Popular Culture and Critical Pedagogy. Reading, Constructing, Connecting*. New York: Garland Publishing.
- de Carvalho, M. (1997). *Rethinking Family-School Relations: A Critique of Parental Involvement in Schooling*. Mahwah, NJ: Erlbaum.
- Deforge, C. and Abouchaar, A. (2003). *The Impact of Parental Involvement, Parental Support and Family Education on Pupil Achievement and Adjustment, A Literature Review. DfES Research Report 433*. Norwich: The Queen's Printer.
- Devine, F. (2004). *Class Practices. How Parents Help Their Children Get Good Jobs*. Cambridge: Cambridge University Press.
- Driscoll, M. and Kerchner, C. (1999). The implications of social capital for schools, communities and cities. In Murphy, J. and Louis, K. S. (eds.) *The Handbook of Research on Educational Administration. A Project of the American Educational Research Association*, pp 385–404. San Francisco, CA: Jossey-Bass.
- Duncan, G. J. and Brooks-Gunn, J. (eds.) (1997). *Consequences of Growing up Poor*. New York: Russell Sage Foundation.
- Dyson, A. H. (1993). *The Social Worlds of Children. Learning to Write in an Urban Primary School*. New York: Teachers College Press.
- Ehrenreich, B. (2001). *Nicked and Dimed. On (not) Getting by in America*. New York: Henry Holt.
- Elder, G. H. and Conger, R. (2000). *Children of the Land. Adversity and Success in Rural America*. Chicago, IL: University of Chicago Press.
- Ellsworth, J. and Ames, L. (eds.) (1998). *Critical Perspectives on Project Head Start. Revisioning Hope and Challenge*. Albany, NY: State University of New York Press.
- Epstein, J. (1994). Theory to practice: School and family partnerships lead to school improvement and success. In Fagnano, C. and Werber, A. (eds.) *School, Family and Community Interactions. A View from the Firing Lines*. Boulder, CO: Westview Press.
- Fincher, R. and Saunders, P. (eds.) (2001). *Creating Unequal Futures? Rethinking Poverty, Inequality and Disadvantage*. Sydney, NSW: Allen and Unwin.
- Fine, M. and Weis, L. (1998). *The Unknown City. The Lives of Poor and Working Class Young Adults*. Boston, MA: Beacon Press.
- Fink, D. (1999). Deadwood didn't kill itself. A pathology of failing schools. *Educational Management and Administration* 27(2), 131–141.
- Finn, P. (1999). *Literacy with an Attitude. Educating Working-Class Children in Their Own Self Interest*. New York: State University of New York Press.
- Fitz, J., Davies, B., and Evans, J. (2006). *Educational Policy and Social Reproduction. Class Inscription and Symbolic Control*. London: Routledge.
- Flaherty, J., Veit-Wilson, J., and Dorman, P. (2004). *Poverty: The Facts*, 5th edn. London: CPAG.
- Foster, P., Gomm, R., and Hammersley, R. (1996). *Constructing Educational Inequality*. London: Falmer Press.
- Frank, C. (1999). *Ethnographic Eyes. A Teacher's Guide to Classroom Observation*. Portsmouth, NH: Heineman.
- Freebody, P., Ludwig, C., and Gunn, S. (1995). *Everyday Literacy Practices in and out of Schools in Low Socio-Economic Urban Communities*. Canberra, ACT: Commonwealth of Australia, Curriculum Corporation.
- Fullan, M. (2006). *Turnaround Leadership*. San Francisco, CA: Jossey-Bass.
- Gold, E., Simon, E., and Brown, C. (2005). A new conception of parent engagement. Community organising for school reform. In English, F. (ed.) *The SAGE Handbook of Educational Leadership. Advances in Theory, Research and Practice*, pp 237–268. Thousand Oaks, CA: Sage.
- Gonzales, M., Moll, L., and Amanti, C. (2005). *Funds of Knowledge*. Mahwah, NJ: Erlbaum.
- Gordon, J. A. (2002). *Beyond the Classroom Walls. Ethnography as Pedagogy*. New York: RoutledgeFalmer.
- Gough, J., Eisenschitz, A., and McCulloch, A. (2006). *Spaces of Social Exclusion*. London: Routledge.
- Griffith, A. and Smith, D. (2005). *Mothering for Schooling*. New York: RoutledgeFalmer.
- Gronn, P. (2003). *The New Work of Educational Leaders. Changing Leadership Practice in an Era of School Reform*. London: Paul Chapman.
- Gunter, H. (2005). *Leading Teachers*. London: Continuum.
- Haberman, M. (1991). The pedagogy of poverty vs good teaching. *Phi Delta Kappan* 73(4), 290–294.
- Haberman, M. (1999). *Star Principals: Serving Children in Poverty*. Indianapolis, ID: Kappa Delta Pi.
- Halverson, R. (2003). Systems of practice: How leaders use artifacts to create professional communities in schools. *Educational Policy Analysis Archives* 11(37) <http://www.epaa.asu.edu/epaa.v11n37/> (accessed May 2009).
- Hampton, G. and Jones, J. (2000). *Transforming Northcote School. The Reality of School Improvement*. London: RoutledgeFalmer.
- Hargreaves, A. (2004). Distinction and disgust: The emotional politics of school failure. *International Journal of Leadership in Education* 7, 27–43.
- Hargreaves, A. and Fink, D. (2006). *Sustainable Leadership*. San Francisco, CA: Jossey Bass.
- Harris, A. and Chapman, C. (2002). *Effective Leadership in Schools Facing Challenging Circumstances*. Nottingham: National College for School Leadership. <http://www.ncsl.org.uk/research> (accessed May 2009).
- Harris, A., James, S., Gunraj, J., Clarke, P., and Harris, B. (2006). *Improving Schools in Exceptionally Challenging Circumstances. Tales from the Frontline*. London: Continuum.
- Hartley, D. (1997). The new managerialism in education: A mission impossible? *Cambridge Journal of Education* 27(1), 47–57.
- Hill, S., Comber, B., Loudon, B., Reid, J., and Rivalland, J. (1998). *100 Children Go to School: Connections and Disconnections in the*

- Literacy Experience Prior to School and in the First Year of School, vols. 1–3. Canberra, ACT: Department of Education, Employment, Training and Youth Affairs.
- Hills, J. and Stewart, K. (eds.) (2005). *A More Equal Society? New Labour, Poverty, Inequality and Exclusion*. Bristol: The Policy Press.
- Johnson, L., Finn, M., and Lewis, R. (eds.) (2005). *Urban Education with an Attitude*. New York: State University of New York Press.
- Johnson, M. (1999). *Failing School, Failing City*. Chipping Norton, UK: Jon Carpenter.
- Kamler, B. and Comber, B. (2005). *Turn around Pedagogies*. Sydney, NSW: PETA.
- Kenway, J., Kraack, A., and Hickey-Moody, A. (2006). *Masculinity beyond the Metropolis*. London: Palgrave.
- Keys, W., Sharp, C., Greene, K., and Grayson, H. (2003). *Successful Leadership of Schools in Urban and Challenging Contexts*. Nottingham: National College for School Leadership.
- Kilpatrick, S., Johns, S., Mulford, B., Falk, I., and Prescott, L. (2002). *More than an Education: Leadership for Rural School-Community Partnerships*. Canberra, ACT: Rural Industries Research and Development Corporation.
- Kozol, J. (1991). *Savage Inequalities. Children in America's Schools*. New York: Harper Perennial.
- Kugelmass, J. (2004). *The Inclusive School. Sustaining Equity and Standards*. New York: Teachers College Press.
- Leithwood, K. and Reihl, C. (2003). *What We Know about Successful School Leadership*. Nottingham: National College for School Leadership.
- Leithwood, K. and Steinbach, R. (2002). Successful leadership for especially challenging schools. In Davies, B. and West-Burnham, J. (eds.) *International Handbook of School Leadership*, pp 25–43. Dordrecht: Kluwer.
- Levin, B. (2003). *Better Education. The Role of Canadian Governments*. Canadian Education Association Whitworth Award presentation. <http://www.cea-ace.ca/res.cfm?subsection=whi&page=ble> (accessed May 2009).
- Levin, B. (2006). Schools in challenging circumstances. A reflection on what we know and what we need to know. *School Effectiveness and School Improvement* 17(4), 399–407.
- Lingard, B., Christie, P., Hayes, D., and Mills, M. (2003). *Leading Learning: Making Hope Practical in Schools*. Buckingham: Open University Press.
- Lister, R. (2004). *Poverty*. Cambridge: Polity.
- Lupton, R. (2003). *Poverty Street. The Dynamics of Neighbourhood Decline and Renewal*. Bristol: The Policy Press.
- Lupton, R. (2004). *Schools in Disadvantaged Areas: Recognising Context and Raising Performance*. CASE Paper 76. London: Centre for Analysis of Social Exclusion, London School of Economics.
- Lyman, L. and Villani, C. (2004). *Best Leadership Practices for High-Poverty Schools*. Lanham, MD: Scarecrow.
- Manicom, A. (1995). What's health got to do with it? Class, gender, and teacher's work. In Campbell, M. and Manicom, A. (eds.) *Knowledge, Experience, and Ruling Relations: Studies in the Social Organisation of Knowledge*, pp 135–148. Toronto, ON: University of Toronto Press.
- Marshall, G., Swift, A., and Roberts, S. (1997). *Against all Odds. Social Class and Social Justice in Industrial Societies*. Oxford: Clarendon Press.
- McGrath, D. J. and Kuriloff, P. J. (2004). They're going to tear the doors of this place: Upper-middle class parent school involvement and the educational opportunities of other people's children. *Educational Policy* 13(5), 603–629.
- McInerney, P. (2005). *Making Hope Practical: School Reform for Social Justice*. Brisbane, QLD: Postpressed.
- Miron, L. and Lauria, M. (2005). *Urban Schools: The New Social Spaces of Resistance*. New York: Peter Lang.
- Mishel, L., Bernstein, J., and Allegretto, S. (2005). *The State of Working America 2004/2005*. Ithaca, NY: Cornell University Press.
- Moll, L. (1992). Literacy research in community and classrooms: A sociocultural approach. In Beach, R., Green, J., Kamil, M., and Shanahan, T. (eds.) *Multidisciplinary Perspectives on Literacy Research*. Urbana, IL: National Council of Teachers of English.
- Monroe, L. (1997). *Nothing's Impossible. Leadership Lessons from Inside and Outside the Classroom*. New York: Public Affairs.
- Murphy, P. and Watson, S. (1994). Social polarisation and Australian cities. *International Journal of Urban and Regional Studies* 18(4), 573–590.
- National College for School Leadership (2005). *A Model of School Leadership in Challenging Urban Environments*. Nottingham: National College for School Leadership.
- Newman, K. (1999). *No Shame in My Game. The Working Poor in the Inner City*. New York: The Russell Sage Foundation/Alfred Knopf.
- O'Connor, M., Hales, E., Davies, J., and Tomlinson, S. (1999). *Hackney Downs. The School That Dared to Fight*. London: Cassell.
- Ofsted (2000). *Improving City Schools*. Norwich: The Queens Printer.
- Pacione, M. (ed.) (1997). *Britain's Cities. Geographies of Division in Urban Britain*. London: Routledge.
- Payne, R. (1998). *A Framework for Understanding Poverty*. Baytown, TX: AhaProcess.
- Peacock, A. (2005). Raising standards: What do we really want? *Forum* 47(2/3), 91–96.
- Peel, M. (2003). *The Lowest Rung. Voices of Australian Poverty*. Melbourne, VIC: Cambridge University Press.
- Pettit, D. (1980). *Opening up Schools*. Ringwood, VIC: Penguin.
- Pounder, D. and Merrill, R. (2001). Job desirability of the high school principalship: A job choice theory perspective. *Educational Administration Quarterly* 37(1), 27–57.
- Power, S., Edwards, T., Whitty, G., and Wigfall, V. (2002). *Education and the Middle Classes*. Buckingham: Open University Press.
- Power, S., Warren, S., Gillbourn, D., et al. (2003). *Education in Deprived Areas. Outcomes, Inputs and Processes*. London: Institute of Education, University of London.
- Rainwater, L. and Smeeding, T. (2003). *Poor Kids in a Rich Country. America's Children in Comparative Perspective*. New York: Russell Sage Foundation.
- Riddell, R. (2003). *Schools for Our Cities. Urban Learning in the 21st Century*. Stoke-on-Trent: Trentham.
- Romano, R. and Glascock, C. (2002). *Hungry Minds in Hard Times. Educating for Complexity for Students of Poverty*. New York: Peter Lang.
- Rothstein, R. (2002). *Out of Balance: Our Understanding of How Schools Affect Society and How Society Affects Schools*. Chicago, IL: Spencer Foundation.
- Rowse, T. (2002). *White Flour, White Power*. Melbourne, VIC: Cambridge University Press.
- Ryan, J. (2003). *Leading Diverse Schools*. Dordrecht: Kluwer.
- Ryan, J. (2005). *Inclusive Leadership*. San Francisco, CA: Jossey Bass.
- Saunders, P. (1996). Income and welfare: Poverty and deprivation in Australia. In *Year Book Australia*. ABS Catalogue No. 1301.0. Canberra, ACT: Australian Bureau of Statistics.
- Scapp, R. (2006). *Managing to Be Different. Educational Leadership as Critical Practice*. London: Routledge.
- Shields, C. (2003). *Good Intentions Are Not Enough. Transformative Leadership for Communities of Difference*. Lanham, MD: Scarecrow Press.
- Shields, C., Bishop, R., and Mazawi, A. E. (2005). *Pathologising Practices. The Impact of Deficit Thinking on Education*. New York: Peter Lang.
- Shields, C. and Edwards, M. (2005). *Dialogue Is Not Just Talk. A New Ground for Educational Leadership*. New York: Peter Lang.
- Shirk, M., Bennett, N., and Aber, J. L. (1999). *Lives on the Line. American Families and the Struggle to Make Ends Meet*. Boulder, CO: Westview Press.
- Slee, R., Weiner, G., and Tomlinson, S. (eds.) (1998). *School Effectiveness for Whom? Challenges to the School Effectiveness and School Improvement Movements*. London: Falmer.
- Spillane, J. (2006). *Distributed Leadership*. San Francisco, CA: Jossey Bass.
- Spillane, J., Halverson, R., and Diamond, J. (2003). Towards a theory of school leadership practice. Implications of a distributed perspective. *Journal of Curriculum Studies* 36(1), 3–34.
- Spillane, J. and Zoltner-Sherer, J. (2004). *A Distributed Perspective on School Leadership Practice as Stretched out over People and Place*. San Diego, CA: Paper presented at the American Educational Research Conference.

- Starratt, R. (2003). *Centering Educational Administration. Cultivating Meaning, Community, Responsibility*. Mahwah, NJ: Erlbaum.
- Stoll, L. and Myers, K. (eds.) (1998). *No Quick Fixes. Perspectives on Schools in Difficulty*. London: Falmer.
- Stubbs, M. (2003). *Ahead of the Class. How an Inspiring Headmistress Gave Children Back Their Future*. London: John Murray.
- Tatto, M. T., Rodriguez, A., Gonzalez-Lantz, D., et al. (2001). The challenges and tensions in reconstructing teacher-parents relations in the context of reform: A case study. *Teachers and Teaching. Theory and Practice* 7(3), 315–333.
- Teddlie, C. and Reynolds, D. (eds.) (2000). *The International Handbook of School Effectiveness*. London: Falmer.
- Teese, R. (2000). *Academic Success and Social Power*. Melbourne, VIC: Melbourne University Press.
- Thomson, P. (1999). Against the odds: Developing school programs that make a difference for students and families in communities placed 'at risk' *Children's Issues* 3(1), 7–13.
- Thomson, P. (2000). Like schools, educational disadvantage and 'thisness'. *Australian Educational Researcher* 27(3), 151–166.
- Thomson, P. (2002). *Schooling the Rustbelt Kids. Making the Difference in Changing Times*. Sydney, NSW: Allen and Unwin/Trentham.
- Thomson, P. and Blackmore, J. (2006). Beyond the power of one: Redesigning the work of school principals and schools. *Journal of Educational Change* 7(3), 161–177.
- Thomson, P., Ellison, L., Byrom, T., and Bulman, D. (2004). *An Investigation of Queries that School Offices Receive from Parents/Carers*. Department for Education and Skills, London. <http://www.dcsf.gov.uk/research/data/uploadfiles/RR575.pdf> (accessed May 2009).
- Thrupp, M. (1995). The school mix effect: The history of an enduring problem in education, research, policy and practice. *British Journal of Sociology of Education* 16(2), 183–203.
- Thrupp, M. (1999). *Schools Making a Difference: Let's Be Realistic! School Mix, School Effectiveness and the Social Limits of Reform*. Philadelphia, PA: Open University Press.
- Toynbee, P. (2003). *Hard Work. Life in Low-Pay Britain*. London: Bloomsbury.
- UNICEF (1996). Solo poverty. *Luxembourg Income Study. Working Paper 127*, <http://www.unicef.org/pon96/insolo.htm> (accessed May 2009).
- Valencia, R. (ed.) (1997). *The Evolution of Deficit Thinking. Educational Thought and Practice*. London: Falmer.
- West, A. and Pennell, H. (2003). *Underachievement in Schools*. London: Falmer Press.
- Whitty, G. and Mortimore, P. (1997). *Can School Improvement Overcome the Effects of Social Disadvantage?* London: Institute of Education.
- Wilkinson, R. (2005). *The Impact of Inequality. How to Make Sick Societies Healthier*. London: Routledge.
- Williams, R. and Pritchard, C. (2006). *Breaking the Cycle of Educational Alienation. A Multiprofessional Approach*. Buckingham: Open University Press.
- Winkley, D. (2002). *Handsworth Revolution. The Odyssey of a School*. London: Giles De La Mare.
- Wright, N. (2001). Leadership, 'bastard leadership' and managerialism. Confronting twin paradoxes in the Blair education project. *Educational Management and Administration* 29(3), 275–290.

Further Reading

- Anyon, J. (2005). *Radical Possibilities. Public Policy, Urban Education and a New Social Movement*. New York: Routledge.
- Connell, R. W. (1994). Poverty and education. *Harvard Educational Review* 64(2), 145–162.
- Kozol, J. (2000). *Ordinary Resurrections. Children in the Years of Hope*. New York: Crown Publishers.

Principals Role in Restructuring Schools

G M Crow, Indiana University, Bloomington, IN, USA

K D Peterson, University of Wisconsin-Madison, Madison, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

From Restructuring to New Structures

The school principal's role has undergone significant changes since the last edition of the international encyclopedia was published. Since our last description, the principal's role in restructuring schools and the complexity of both the role and the context in which this occurs have increased. The importance of the knowledge society, the globalization of society, and the increasing demographic diversity of schools have called for new, or at least changed, roles for school principals. Clearly, the specific nature of these changes varies with national and regional contexts; but internationally too, the roles of principals are changing (Cheung, 2000; Crow *et al.*, 2002; Earley and Weindling, 2004). In this article, current changes in the nature of school improvement and the resulting changes in the principal's roles, relationships, rules, and results are identified. The article ends with an identification of implications of these changes for research, policy, practice, and preparation.

Current Changes in the Nature of School Improvement

Although there are several changes in the nature of school improvement, four major movements have been identified: re-culturing, professional learning communities, distributed leadership, and social justice. These four movements are believed to be especially important for understanding how the principal's role has changed. These movements and other external pressures on principals, however, create tensions and conflicts within the role, which will also be identified.

Reculturing

The first major change in principals' roles is centered on the ways these school leaders shape the cultures of their schools. While many countries are increasingly pressing schools to become more tightly structured, more focused on outcomes and accountability, the school principal is being pressed to be more managerial and administrative. Nonetheless, researchers (Deal and Peterson, 1999; Fullan, 2001; Gurr *et al.*, 2006) point out that schools, like other organizations, must attend to the culture of the organization. Reinforcing and, where necessary, reculturing school

cultures are crucial roles of principals. Shaping the school culture means being able to read, assess, and reinforce key elements of the underlying norms and values, rituals and ceremonies of the school (Deal and Peterson 1999). Principals must be managers – structuring roles and responsibilities, implementing school improvement plans and so forth – but they also must be symbolic leaders reinforcing fundamental norms and values in their daily meetings, decisions, and interactions.

Professional Learning Community

The second major change in the nature of school improvement is the creation of professional learning communities. This strategy emerged from at least two ideas: teachers as professionals and organizational learning. The perspective of teachers as professionals focused on the unique knowledge and skills and the need for teachers to have the appropriate discretion to diagnose student needs and identify and implement responses to those needs. Rather than treating teachers as lone professionals isolated in their classrooms performing these diagnosis, identification, and implementation tasks, the literature on professional learning communities emphasizes the community of professionals who work together to promote student learning. This idea emerged from the organizational learning literature (Senge *et al.*, 2000) that emphasized the importance of the social rather than individual construction of knowledge and learning environments. This literature also acknowledges that student learning occurs throughout the school rather than only in the classroom. McLaughlin and Talbert (2006) define teacher learning communities as “teachers’ joint efforts to generate new knowledge of practice and their mutual support of each others’ professional growth” (p. 75). The characteristics of professional learning communities have been identified (Louis *et al.*, 1996; Scribner *et al.*, 2002; Smylie and Hart, 2000), as fostering collaboration, de-privatizing practice, and committing to shared values around student learning.

Principals have a special role in these professional learning communities. Crow *et al.* (2002) note that principals must nurture the intellectual capacity of teachers rather than view themselves as the only source of ideas, help teachers develop a common mission, and advocate for the schools in the face of external challenges (p. 198). Earley and Weindling (2004) identify the cultural role of

principals in learning communities that includes questioning the status quo, affirming the knowledge of all individuals in the organization, and distributing leadership. McLaughlin and Talbert (2006) found that principals who were unwilling to share control with teachers or who de-valued teacher contributions undermined school improvement efforts.

Distributed Leadership

A third major adaptation in the nature of school leadership is the attention to distributed leadership. Distributed leadership, different from the focus in the 1990s on the structural notion of shared leadership, focuses on the interactions and actions of the principal and staff that move school improvement forward (Spillane, 2006). The distributed leadership paradigm suggests that principals and teacher leaders should focus on the artifacts, routines, policies, and social interactions that move the school forward, improve practice, and enhance learning (Leithwood and Jantzi, 1998; Spillane, 2006). This model seems to suggest that school principals need a deeper understanding of the social, cognitive, and contextual factors that combine to produce outcomes. The focus in the school becomes more on the practice of leadership than the school administrator. School principals therefore need to learn how to spread leadership routines across tasks, processes, and interactions.

Social Justice

A fourth shift in the nature of school improvement is an emphasis on social justice, specifically ensuring equitable educational outcomes and closing the achievement gap (Gillborn, 2001; Ladson-Billings, 2006). Although equity is not a new concept in school improvement, government mandates and advocacy movements have raised the stakes. For example, the US Congress's No Child Left behind Act has, at the least, increased the rhetoric and the expectations that all students can and should learn. Although the Act and other similar governmental mandates are clearly contested in educational and political arenas, the resulting demands on principals are more intense than before.

In addition to government mandates, advocacy activities geared to raise the consciousness of policymakers, educators, and the public to reverse the oppressing and privileging tendencies within education, that disadvantage student groups based on race, ethnicity, gender, sexuality, language, or disability, have increased. These activities, both inside and outside educational organizations, have added a new element to the principal's role and require greater attention to how these student groups are impacted by school and societal forces (Skrla *et al.*, 2004).

Conflicts/Tensions in School Improvement

These four recent changes in the nature of school improvement are contested areas, routinely debated in various educational arenas. For school principals, they create at least three types of conflicts or tensions in advancing school improvement and student learning.

Accountability and Professionalism

A major tension that exists in the school reform literature in general and for the principal's role in supporting reform is the potential conflict between accountability and professionalism. Accountability is an international phenomenon (Leithwood and Earl, 2000), but is multifaceted and involves different contexts, such as government agency, local educational authority, school site council, and parent board/association (Marks and Nance, 2007). Even within, for example, the government context, accountability may involve different approaches, for example, market, decentralization, professionalization, and management (Leithwood, 2001). As Leithwood maintains, professionalism may be a type of accountability. Frequently, the tension and conflict are between other approaches, such as managerial or market accountability and professionalism. Major governmental initiatives in several countries around increasing student achievement for all students have intensified accountability by raising the stakes by connecting the increase in student test scores to student graduation and to school resources. In addition, some governmental initiatives have connected the increase in achievement scores to job security or performance pay for educators.

As noted previously, one initiative in new school reform agendas has been professional learning communities that emphasize a stronger, discretionary role for teachers and administrators. This professionalism argues for valuing the expertise and practical knowledge that teachers develop over time and the need for teachers to have the discretion to use this knowledge in responding to the individual and unique needs of their students. Professionalism places the decision-making role in the hands of educators rather than regulators or the market.

On the face of it, there are major inconsistencies and contradictions between some forms of accountability and professionalism. One seems to place decision-making authority in the hands of regulators and policymakers, while the other places it in the hands of educators. Emphasizing accountability seems to diminish professionalism. However, the tension is not necessarily this simple. As Hargreaves (2003) has argued, there is a difference between standardization and standards. A managerial accountability approach that emphasizes standardizing not only the

outcomes but also the processes of education clearly diminishes the importance of the educator's discretionary role. By contrast, accountability that emphasizes the need for research-based standards that can be used by educators to guide and evaluate their practices toward advancing learning for all students, does not necessarily contradict professionalism. The future of this apparent tension will depend on regulators, governing boards, and educators as they develop balances in standards that reflect and inform educator practices. However, the various accountability demands on schools "creates significant leadership dilemmas (Wildy and Loudon, 2000), and school leaders attempting to respond to their government's demands for change can be excused for feeling that they are being pulled in many different directions simultaneously. They *are* being pulled in many different directions simultaneously" (Leithwood, 2001: 228).

New Expectations for Tight or Loose Coupling

The structures of schools appear to be changing in terms of their structural coupling. For many decades, schools in the United States, for example, were allowed to be relatively loosely coupled, with considerable flexibility accorded to teachers and principals regarding curriculum, instruction, and the expectations for student achievement, for example, the types of students expected to learn (Weick, 1976). Recently, US policymakers at all levels have increased the outside expectations for: tighter design of curriculum (state standards), required instructional strategies (such as direct instruction), and specific targeted student achievement guidelines (No Child Left Behind Act). In the US, this change in the structural requirements of schools has put pressure on principals to become more administrative, bureaucratic, and evaluative. Interestingly, in some countries, for example, Taiwan, that had employed tightly aligned curriculum and assessment for decades, there is a movement to reduce governmental reigns, allowing schools to develop locally relevant curricula (Hughes and Stone, 1999). Principals in these countries must now engage in less of the administration of existing programs to more work on innovation and creation of emergent programs. Neither approach to restructuring, from loose to tight or from tight to loose is easy or simple and produces tensions for principals.

New Forms of Organization

Recent approaches to restructuring have involved new forms of schools, such as charter schools, middle college high schools, and schools-within-schools. These different forms are changing the roles of principals. Charter schools in countries such as the United States and Canada (O'Reilly and Bosetti, 2000) are designed to be free from the majority of curricular and assessment requirements

of the government. In these schools, principals often must be able to establish a coherent vision and set of administrative structures, actively recruit students, engage in marketing practices, and work with a governance board. But some researchers have noted the tendency for these schools to ignore fairness and equity considerations (Hausman, 2000). Middle college high schools were developed to serve disconnected youth who have shown some academic potential but have dropped out of school or are likely to. Located on college campuses, these small (100–140 students) high schools meet on an innovative schedule, offer regular and college classes, provide more personalization and counseling, and a chance to both graduate from high school and gain college credits (Grier and Peterson, 2007; Weschsler, 2001). Principals in these schools must seek new organizational structures, connect on an individual basis with disaffected youth, and find ways to foster a culture of support and personalization. Finally, the small-schools movement has encouraged the development of schools-within-schools. In these contexts, principals must become innovators and school design experts who are able to develop a unique mission and set of structures within the same building as another school. In these settings, principals must blend and balance educational leadership with political leadership as they negotiate use of space, facilities, and equipment.

Resulting Changes in the Principal's Role

In the previous edition of the international encyclopedia, new roles, relationships, rules, and results that principals confronted were identified. In this section, these areas are discussed but shifts are suggested that have resulted from the contemporary changes in the nature of school improvement and the conflicts/tensions in school improvement identified earlier in this article.

New Roles

Changes in the Political Role of Principal

The changes and tensions we identified earlier have created new political roles for the principal to facilitate school improvement and advance student learning. These involve both internal and external roles. Through facilitating professional learning community and distributed leadership, the principal is confronted with an increase in potential conflicts over values and resources. Although involving more people in the collaboration and decision-making process has significant benefits as the research on professional learning communities has demonstrated, costs of time, coordination, and resource allocation exist and many of these involve potential conflicts (Crow, 1998). In addition, the principal's political role in advocating for social justice

and the rights of all students to learn can confront the principal with resource and ideological conflicts.

The principal's political role also changes in external terms. The demands of accountability and the tension between different accountability contexts and approaches, for example, market and professionalism, (Leithwood, 2001; Marks and Nance, 2007) set up potential conflicts among groups such as district/LEA administrators, governmental regulators, parents, out-of-school social service experts, and teachers. These political conflicts can involve disagreements that require the principal's attention and expertise in resolving conflicts around allocating resources, defining achievement goals, and balancing achievement and social justice.

Changes in the Cultural Role of Principal

The restructuring of schools from loosely coupled (Weick, 1976) organizations to more tightly structured, standards-based educational settings actually accentuates and extends the culture-shaping role of principals (Deal and Peterson, 1999). With schools tightening their curriculum, increasing their external accountability requirements, and, at times, defining instructional approaches in the classroom, being able to assess and nurture a professional school culture is even more important. Principals are required oftentimes to re-culture existing school norms so they match the standards-based rules and responsibilities. In fact, using data for decision making and school improvement is a useful and important task, but one that is new to many school cultures. Therefore, principals and teacher leaders will need to find ways to reinforce, articulate, and support new cultural norms and traditions.

Changes in the Environmental Role of Principal

Many of the new forms of school organization as well as accountability demands have changed the environmental role of principals. Clearly, principals have for several years held a role beyond the school. However, as Crow *et al.* (2002) note, market competition and accountability among other forces have required the principal to become more entrepreneurial and to take on a larger societal role. These environmental roles involve reminding teachers and parents of external expectations, such as market or governmental interests, advocating for social justice in communities, and educating community, regulatory, and policymaking audiences of student and teacher needs and perspectives.

Changes in the Instructional Role of Principal

The combination of changes in the specificity and focus of student learning outcomes, standards-based curriculum reform, the availability of computer-managed student

achievement data, and intense accountability systems have substantially changed the instructional role of principals. Increasingly, principals are expected to be highly skilled users of computer-based statistical analyses of student achievement data. Principals and their teacher leaders now must know statistical fundamentals, be able to identify critical questions related to student achievement, and then analyze quantitative student performance data. Additionally, they need to be able to use those analyses to drive school improvement, planning, curriculum refinement, and instructional changes. These new roles require sometimes new and sometimes deepened skills in the use of data for decision making, computer skills, an understanding of statistical reasoning, and complex knowledge of the change process that is often not taught in preparation programs. Moreover, principals have the additional instructional role of removing barriers to teachers' use of data for instructional decision making (Ingram *et al.*, 2004). In some cases, this involves resolving conflicting policy demands and creating technical cultures that support norms of data use.

New Relationships

The new roles that have been identified for contemporary principals imply new relationships as well. The number and diversity of stakeholders to which the principal must relate have increased dramatically (Crow *et al.*, 2002). In addition to school and community stakeholders, principals must interact with other social service and healthcare professionals involved in students' well-being and with regulators and policymakers seeking to ensure student and educator productivity. In addition to the increasing number and diversity of stakeholders, the internal and external boundaries of schools are blurring. For example, students are increasingly taking college courses before they complete their high school graduation. Community groups see the school as a pivotal agency and resource to influence policy and implement programs. The school is increasingly an ideological battleground, with the principal confronting disputes, advocating student and family interests, and forming alliances.

New Rules

These new roles and relationships more frequently arise without rules or with new, fluid rules. Principals face these conflicts and tensions without the scripts common to school leadership practice in the past. The rapidity of change in a knowledge society, the demands for customized responses, and the contextualization of leadership require principals to think quickly, create unique responses, and be ready to adjust their responses to constantly changing circumstances.

(Crow *et al.*, 2002). In addition, the new rules of the principalship demand not heroic, sole leadership practice, but shared and distributed practice. As Murphy (2000) argues, this involves leadership from the middle, rather than the top.

New Results

The principal's role in restructuring for school improvement now exists in a context with a new set of expected results. Internationally, this varies with national context and purpose. In many settings, the importance of student achievement for all students is more pronounced than before. Principals and other educators are expected to ensure student learning not only for the most able or even the average, but for all students. This result, however, is being contested as educators, regulators, policymakers, and other interest groups confront the ideological and resource consequences (Ladson-Billings, 2006). Increasingly, educators, parents, and others are expanding the notion of student achievement to include affective components in which principals are expected to include civic, esthetic, and affective instructional goals as well as other sociopsychological effects such as student engagement in school (Leithwood and Earl, 2000). In other international contexts, the results take on a more governmental perspective in which schooling is a vehicle for nationalistic visions, norms, and strategies (Hughes and Stone, 1999; Tsai, 2002). As education is recognized as a powerful instrument for furthering governmental initiatives and interests, the principal's role takes on a different set of results and the principal becomes an agent of the state with responsibility to promote national objectives.

Implications

Internationally, the late 1990s and early twenty-first century has been a period of rapid and widespread change in educational policy and practice. These changes and the press for restructuring schools have necessitated the rethinking of principals' roles and practices. Clearly, given these changes there is considerable need for systematic, field-based studies of these changes on the daily work of principals, on the nature of leadership in these new contexts, and on the interaction of new structures and new behaviors. At the same time, there is a need for the reconsideration of preparation, licensure, and the professional development of school leaders. Preparation needs to be reconfigured to address the new realities of the work. Licensure, in settings where it is used, needs to be redefined to address the continued changes that will occur. Finally, professional development of principals needs to be deepened and expanded to reflect the wider array of

demands coming from the new relationships, new rules, and new results expected of school principals.

Bibliography

- Cheung, R. M. B. (2000). Securing a better future: A Hong Kong school principal's perception of leadership in times of change. In Dimmock, C. and Walker, A. (eds.) *Future School Administrators: Western and Asian Perspectives*, pp 225–248. Hong Kong: The Chinese University of Hong Kong.
- Crow, G. M. (1998). Implications for leadership in collaborative schools. In Pounder, D. G. (ed.) *Restructuring Schools for Collaboration: Promises and Pitfalls*, pp 135–153. Albany, NY: SUNY.
- Crow, G. M., Hausman, C. S., and Scribner, J. P. (2002). Reshaping the role of the school principal. In Murphy, J. (ed.) *The Educational Leadership Challenge: Redefining Leadership for the 21st Century*, (101st Yearbook of the National Society for the Study of Education), pp 134–161. Chicago, IL: National Society for the Study of Education/The University of Chicago Press.
- Deal, T. and Peterson, K. (1999). *Shaping School Culture: The Heart of Leadership*. San Francisco, CA: Jossey-Bass.
- Earley, P. and Weindling, D. (2004). *Understanding School Leadership*. London: Paul Chapman.
- Fullan, M. (2001). *Leading in a Culture of change*. San Francisco, CA: Jossey-Bass.
- Gillborn, D. (2001). "Raising standards" or rationing education? Racism and social justice in policy and practice. *Support for Learning* 16(3), 105–111.
- Grier, T. and Peterson, K. (2007). *Middle College High Schools: Secondary Schools for Disconnected Youth*. Paper presented at the American Educational Research Association, Chicago, IL.
- Gurr, D., Drysdale, L., and Mulford, B. (2006). Models of successful principal leadership. *School Leadership and Management* 26(4), 371–395.
- Hargreaves, A. (2003). *Teaching in the Knowledge Society: Education in the Age of Insecurity*. New York: Teachers College Press.
- Hausman, C. S. (2000). Principal role in magnet schools: Transformed or entrenched? *Journal of Educational Administration* 38(1), 25–46.
- Hughes, C. and Stone, R. (1999). Nation-building and curriculum reform in Hong Kong and Taiwan. *The China Quarterly* 60, 977–991.
- Ingram, D., Louis, K. S., and Schroeder, R. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record* 106(6), 1258–1287.
- Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher* 35(7), 3–12.
- Leithwood, K. (2001). School leadership in the context of accountability policies. *International Journal of Leadership in Education* 4(3), 217–235.
- Leithwood, K. and Earl, L. (2000). Educational accountability effects: An international perspective. *Peabody Journal of Education* 75(4), 1–18.
- Leithwood, K. and Jantzi, D. (1998). *Distributed Leadership and Student Engagement in School*. Paper presented at the American Educational Research Association, San Diego, CA.
- Louis, K. S., Kruse, S. D., and Marks, H. H. (1996). Schoolwide professional community. In Newman, F. M. A. (ed.) *Authentic Achievement: Restructuring Schools for Intellectual Quality*, pp 179–203. San Francisco, CA: Jossey-Bass.
- Marks, H. M. and Nance, J. P. (2007). Contexts of accountability under systemic reform: Implications for principal influence on instruction and supervision. *Educational Administration Quarterly* 43(1), 3–37.
- McLaughlin, M. W. and Talbert, J. E. (2006). *Building School-Based Teacher Learning Communities*. New York: Teachers College Press.
- Murphy, J. (2000). *The Quest for a Center: Notes on the State of the Profession of Educational Leadership*. Columbia, MO: University Council for Educational Administration.
- O'Reilly, R. R. and Bosetti, L. (2000). Charter schools: The search for community. *Peabody Journal of Education* 75(4), 19–36.
- Scribner, J. P., Hager, D., and Warne, T. R. (2002). The paradox of professional community: A tale of two high schools. *Educational Administration Quarterly* 38(1), 45–76.

- Senge, P., Cambron-McCabe, N., Lucas, T., et al. (2000). *Schools that Learn*. London: Nicholas Brearley.
- Sklra, L., Scheurich, J., Garcia, J., and Nolly, G. (2004). Equity audits: A practical leadership tool for developing equitable and excellent schools. *Educational Administration Quarterly* 40(1), 133–161.
- Smylie, M. A. and Hart, A. W. (2000). School leadership for teaching learning and change: A human and social capital development perspective. In Murphy, J. and Louis, K. S. (eds.) *Handbook of Research on Educational Administration*, pp 421–441. San Francisco, CA: Jossey-Bass.
- Spillane, J. (2006). *Distributed Leadership*. San Francisco, CA: Jossey-Bass.
- Tsai, C-T. (2002). The Chinese-ization and the nationalistic curriculum reform in Taiwan. *Journal of Educational Policy* 17(2), 229–243.
- Weick, K. (1976). Educational organizations as loosely-coupled systems. *Administrative Science Quarterly* 21, 1–16.
- Weschler, H. S. (2001). *Access to Success in the Urban High School: The Middle College Movement. Reflective History Series*. New York: Ford Foundation.
- Wildy, H. and Loudon, W. (2000). School restructuring and the dilemma of principals' work. *Educational Management and Administration* 28 (2), 173–184.
- Dimmock, C. Walker, A. (eds.) *Future School Administrators: Western and Asian Perspectives*. Hong Kong: The Chinese University of Hong Kong.
- Earley, P. and Weindling, D. (2004). *Understanding School Leadership*. London: Paul Chapman.
- Gorinski, R. and Shortland-Nuku, C. (2006). *Building Innovative Communities of Professional Learning: A Leadership Practice Challenge*. Paper presented at the Commonwealth Council for Educational Administration and Management, Nicosia, Cyprus.
- Gronn, P. (1999). *The Making of Educational Leaders*. London: Cassell.
- Leithwood, K. and Jantzi, D. (2000). The effects of transformational leadership on organizational conditions and student engagement with school. *Journal of Educational Administration* 38(2), 112–129.
- Matthews, L. J. and Crow, G. M. (2003). *Being and Becoming a Principal. Role Conceptions for Contemporary Principals and Assistant Principals*. Boston, MA: Allyn and Bacon.
- McLaughlin, M. W. and Talbert, J. E. (2001). *Professional Communities and the Work of High School Teaching*. Chicago, IL: The University of Chicago Press.
- Murphy, J. (ed.) (2002). *The Educational Leadership Challenge: Redefining Leadership for the 21st Century* (101st Yearbook of the National Society for the Study of Education). Chicago, IL: National Society for the Study of Education/The University of Chicago Press.
- Mulford, B., Silins, H., and Leithwood, K. (2004). *Leadership for Organisational Learning and Student Outcomes: A Problem-Based Learning Approach*. Dordrecht: Kluwer.
- Southworth, G. (2003). *Primary School Leadership in Context. Leading Small, Medium, and Large-Sized Primary Schools*. London: RoutledgeFalmer.

Further Reading

- Crow, G. M. (2006). Democracy and educational work in an age of complexity. *UCEA Review* 48(1), 1–5.
- Deal, T. and Peterson, K. (1999). *Shaping School Culture: The Heart of Leadership*. San Francisco, CA: Jossey-Bass.

School Reform and Restructuring: Self Managing School

B J Caldwell, Educational Transformations Pty Ltd, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Most systems of public education around the world have included decentralization of decision making in their plans for school reform or restructuring. There is a seeming paradox in that there has also been significant centralization. The Organization for Economic Cooperation and Development (OECD) explained it in these terms:

An important factor in educational policy is the division of responsibilities among national, regional and local authorities, as well as schools. Placing more decision-making authority at lower levels of the educational system has been a key aim in educational restructuring and systemic reform in many countries since the early 1980s. Yet, simultaneously, there have been frequent examples of strengthening the influence of central authorities in some areas. For example, a freeing of 'process' and financial regulations may be accompanied by an increase in the control of output from the centre, and by national curriculum frameworks (OECD, 2004: 34).

Decentralization of decisions to schools goes by several names including self-management, school-based management, or local management. Nomenclature varies from country to country and self-management is the term used in this article.

Definition

A self-managing school is a school in a system of education "to which there has been decentralized a significant amount of authority and responsibility to make decisions related to the allocation of resources within a centrally determined framework of goals, policies, standards and accountabilities. Resources are defined broadly to include knowledge, technology, power, materiel, people, time, assessment, information and finance" (Caldwell and Spinks, 1998: 4, 5).

It is important to distinguish between self-management and autonomy, independence or privatization of schools. The extent of decentralization in self-management is constrained by the requirement that self-managing schools operate in a centrally determined framework. Such schools continue, by and large, to be built, owned, operated, and funded by a public authority. There is, however, increasing interest in creating autonomous schools within systems of public education. Charter schools in Canada and the United States are examples; however, there are still constraints on their operation by the public authorities that

established them in the first place or that could close them for any reason. Majority of schools in the Netherlands are owned and operated by a nonpublic authority but receive their funds from government. It is unconstitutional and, therefore, illegal to differentially fund schools on the basis of ownership. This is an example of a long-standing system of self-managing schools that may be privately owned and operated. Normally, private, nonpublic, or independent schools are self-managing schools and this article is primarily concerned with those in systems of public education.

Trends

The trend to self-management is part of a general trend to decentralization in public education as summarized in the opening paragraph. What follows is drawn from surveys conducted by the OECD among its 30 members and 20 partner nations and published in an annual report on indicators of performance. The 2004 report (OECD, 2004) included information on trends in centralization and decentralization. It considered the locus and mode of decision making in four domains. Locus referred to which of six levels decisions were made: national, state, regional, municipal, local, or school. Mode referred to which of four ways decisions were made: full autonomy at the level concerned, consultation with other bodies at that level, independently but within a framework set by a higher authority, or other. The four domains were organization of instruction, personnel management, planning and structures, and resources.

The report compared patterns in 1998 and 2003 and found that "in 14 out of 19 countries decisions are taken at a more decentralized level in 2003 than in 1998." The following points summarize the major findings as far as decentralization to the school level is concerned:

- Decisions are more often taken at the school level in the Czech Republic, England, Hungary, New Zealand and the Slovak Republic and in particular in the Netherlands where all decisions are taken at the school level.
- Decisions on the organization of instruction are predominantly taken by schools in all OECD countries, while decisions on planning and structures are mostly the domain of centralized tiers of government. The picture is more mixed for decisions on personnel management and allocation and use of resources.

- Just less than half of decisions taken by schools are taken in full autonomy, about the same proportion as those taken within a framework set by a higher authority. Decisions taken by schools in consultation with others are relatively rare. Schools are less likely to make autonomous decisions related to planning and structures than related to other domains. (OECD, 2004: 21, 22)

Care should be taken in interpreting some of these patterns, as there are important differences within countries. This is particularly the case in Australia that is reported as being one of the most centralized. In Australia, like Canada and the United States, constitutional powers for making laws in relation to education lie with the states (provinces in Canada), with the Australian Government able to influence arrangements through its powers to make grants to the states. The statement that the country is highly centralized is a generalization that cannot be applied to all of the states. It does not apply to states that have shifted significant authority and responsibility to the school level in recent times. In Victoria, for example, 94% of the state's recurrent budget is decentralized to the school level for local decision making, albeit within a centrally determined framework, and this level exceeds that in England and New Zealand that are reported as being highly decentralized.

The aforementioned countries of Australia, Canada, England, New Zealand, and the United States, or states/provinces or districts therein, include some of the high-profile examples of self-managing schools. Apart from the general patterns among OECD members described above, there are also noteworthy examples in partner or observer nations. For example, decentralization including school-based management is a major feature of nationwide reforms in the Philippines and Thailand (partners) and in Israel (observer).

Driving Forces

Changes in patterns of decision making are partly explained by new directions in education policy. The OECD report cited above provided the broad context:

Changing social and economic conditions have given education an increasing central role in the success of individuals and nations. Human capital has long been identified as a key factor in combating unemployment and low pay, but there is now also robust evidence that it is associated with a wide range of non-economic benefits, including improvements in health and a greater sense of well-being . . . The benefits of education have driven increased participation in a widening range of learning activities – by people of all ages, from earliest childhood to advanced adulthood. As the demand for learning grows and become more diverse, the challenge

for governments is to ensure that the learning opportunities provided respond to real, dynamic needs in a cost-effective manner (OECD, 2004: 11).

While these factors explain much of the energy in educational reform, there are other reasons that account for the trend to self-management since the late 1960s. Many of the landmark social, political, and religious movements of the 1960s and early 1970s spawned much interest in empowerment. These were the years of social unrest, as seen in the student riots in Paris in 1968, protests against the Vietnam War, and Vatican II. By the end of the 1970s, there were important developments in school-based management in Canada and the United States. A review in 1977 identified four factors, working singly or in combination, representing demands for increased sensitivity to local needs and problems, reversal of the effects of size and centralization, accountability and professionalism, and a desire for participative management (Caldwell, 1977). One of the noteworthy reforms of this period was in the Edmonton Public School District in Alberta, Canada, known initially as school-based budgeting because the major part of the district's budget was decentralized to schools for local decision making. There is further mention of Edmonton later in this article because of its pioneering approach to needs-based formula funding of schools that continues as an exemplar.

These same forces were at work in Australia, which has traditionally been considered to have a highly centralized system of education. Reports of distinguished scholars were highly critical of the arrangement (Kandel, 1938; Butts, 1955). While there were precursors at the state level, the seminal event in shifting the balance of centralization and decentralization was the release of the report of the Interim Committee of the Australian Schools Commission (1973), generally known as the Karmel Report. Decentralization, or devolution as it was referred to at the time, was elevated to the status of a value that underpinned its recommendations. The seven values were devolution of responsibility, equality, diversity, public and private schooling, community involvement, special purposes of schools, and recurrent (lifelong) education. Most developments in self-managing schools, including the far-reaching reform in Victoria, cited earlier, can be traced to these reforms of the 1970s.

The local management of schools in England had its foundation in the community education movement several decades before the major thrust of the Education Reform Act of 1988. The notion of community empowerment ran through many of the developments in several counties about this time, being essentially part of a liberal democratic tradition. However, the tenor of debate changed profoundly with the passing of the 1988 Education Reform Act under the Thatcher Conservative Government. This was the era in which the government endeavored to weaken the power

of unions and wind back the influence of the state. Market reforms were in the ascendancy and self-managing schools were seen by many as being part of this movement, despite their foundations in an earlier period. Robust criticisms were mounted (Smyth, 1993).

The 1980s and the 1990s marked a struggle on the merits of the approach, especially in Australia, England, and New Zealand. In Victoria, for example, momentum in the early 1980s under a moderate left-of-center Labor Government stalled in the late 1980s with resistance by unions to proposals to decentralize most of the budget and significant authority at the school level in respect to the selection of staff. In the 1990s, a change in government to the right of center resulted in more than 90% of the state's budget for government schools being decentralized for local decision making. There were similar developments in New Zealand. However, much of the sting in the debate and its ideological overtones were removed when a return to left-of-center governments produced little change. In England, under the Blair New Labor Government elected in 1997, the self-management reforms of the previous Conservative Government were extended, so that more than 90% of the budget for schools has been decentralized. Similarly in Victoria, where the approach was extended when a left-of-center government was returned to power, and in New Zealand, although there has been some weakening of school authority in that country.

Issues

There are four major issues in the contemporary implementation and critique of self-managing schools: (1) the extent to which there has been an impact on learning; (2) the mechanism that is used to allocate funds from the center to schools; (3) whether the decentralization of funds to the local level has increased the opportunity for corruption in the use of public funds; and (4) the impact of self-management on the role of the principal and other school leaders.

Link to Learning

There was often a demand for evidence that self-management led in cause-and-effect fashion to improved student outcomes. It was sobering to note the consistent finding in early research that there appeared to be few, if any, direct links between local management, self-management or school-based management, and learning outcomes (Malen *et al.*, 1990; Summers and Johnson, 1996). Some researchers noted that such gains are unlikely to be achieved in the absence of purposeful links between capacities associated with school reform, in this instance, self-management, and what occurs in the classroom, in learning

and teaching and the support of learning and teaching (Bullock and Thomas, 1997; Cheng, 1996; Hanushek, 1996, 1997; Levačić, 1995; Smith *et al.*, 1996; OECD, 1994).

A review of research suggests that there have been three generations of studies and it is only in the third that evidence of the impact of decentralization on outcomes has emerged, provided that certain conditions are fulfilled. The first generation was in times when impact on learning was not a primary or even secondary purpose. The second generation was when such purposes may have been to the fore but the database was weak. The third, emerging in the late 1990s and gathering momentum in the early 2000s, coincides with a preeminent concern for learning outcomes and the development of a strong database (Caldwell and Spinks, 1998; Caldwell, 2002, 2003, 2005).

Woessmann undertook a comprehensive study of why students in some countries did better in the Trends in Mathematics and Science Study (TIMSS) and found a powerful connection between decentralization of decision making to the school level and student achievement (Woessmann, 2001). It is a connection that has been affirmed in subsequent results in the Programme for International Student Assessment (PISA). Schleicher identified decentralization as one of several policy levers for student achievement. He found that, in the best-performing countries:

- Decentralized decision-making is combined with devices to ensure a fair distribution of substantive educational opportunities
- The provision of standards and curricula at national/sub-national levels is combined with advanced evaluation systems
- Process-oriented assessments and/or centralized final examinations are complemented with individual reports and feedback mechanisms on student learning progress
- Schools and teachers have explicit strategies and approaches for teaching heterogeneous groups of learners
- Students are offered a variety of extra-curricular activities
- Schools offer differentiated support structures for students
- Institutional differentiation is introduced, if at all, at later stages
- Effective support systems are located at individual school level or in specialized support institutions
- Teacher training schemes are selective
- The training of pre-school personnel is closely integrated with the professional development of teachers
- Continuing professional development is a constitutive part of the system
- Special attention is paid to the professional development of school management personnel (Schleicher, 2004).

While a balance of centralization and decentralization is evident in the above, it is important to note that, even in contemporary times, there may be no impact on learning unless purposeful links are made to the student and classroom levels. There is a need to ensure an impact across all schools in a system. Fullan *et al.* (2006) demonstrated the limits to improvement under self-management by describing how gains in literacy have plateaued in England, and how decentralization of decision making in Chicago, Milwaukee, and Seattle has not led to large-scale improvement: “They contain glimpses of what will be required, but they fail to touch deeply day-to-day classroom instruction, and to touch it in a way that will get results for all” (Fullan *et al.*, 2006: 6). They proposed a system to lift the performance of schools to achieve a breakthrough, comprising three components: personalization, professional learning, and precision. “The glue that binds these three is moral purpose: education for all that raises the bar as it closes the gap” (p. 16).

Needs-Based Funding

A key issue is the determination of a funding mechanism to allocate resources from central sources to schools in systems of self-managing schools through mechanisms known variously as global budgets or student resource packages. Allocations typically include a per capita component, with weights that differ according to stage of schooling, and needs-based components that reflect student and school characteristics. Allocations for the per capita component generally reflect historical approaches, especially in respect to a class rather than student focus and assumptions about student–teacher ratios. Allocations that reflect school characteristics invariably take account of size and economies of scale; location, especially in remote or rural settings; and stage and specialization in schooling, where there are different resource requirements. Allocations that are more student focused typically take account of the socioeconomic status of the families or communities of students and the extent of special education needs, including disabilities and impairments. The pioneering system of self-managing schools in Edmonton, Alberta continues to provide a template. Good progress was made in the 1990s in several countries. Levačić and Ross (1999) provide a summary of approaches in Australia, Canada, England, New Zealand, United States, and Wales.

By 2007, attention was shifting in extensively decentralized systems of self-managing schools, especially in Australia and England, to how allocations from the system to the school level could take account of efforts to secure success for all students in all settings and to personalize the learning experience as far as possible. Critically important is how resources, once received, are best allocated at the school level. Student-focused planning models are emerging (Caldwell and Spinks, 2008).

Opportunities for Corruption

An important issue is whether the introduction of self-management with decentralization of funds to schools leads to corruption at the local level. This was the subject of a major study of the International Institute for Educational Planning (IIEP) of the United Nations Educational, Scientific and Cultural Organization (UNESCO) (Levačić and Downes, 2004). Case studies were provided of formula funding for schools under conditions of decentralization in Australia (Victoria), the United Kingdom (England), Poland (Kwidzyn and Swidnik), and Brazil (Rio Grande do Sul).

The reason for the study was stated in the following terms. “Given that the proportion of the national budget devoted to education is significant for both developed and developing countries, it is essential that public funds be directed effectively and used for the purposes for which they are allocated. The misuse of public funds is a serious matter both in terms of ethical and criminal implications of the abuser and in terms of the deprivation of funding inflicted on students” (Levačić and Downes, 2004: 15). The focus was on practice in systems of self-managing schools. Particular attention was given to transparency, the accurate collection of data, the avoidance of fraud, and the need for a range of auditing procedures at different levels. Few instances of fraud were uncovered in the study.

Recommendations were concerned with training, preparing manuals of financial procedures, removing opportunities for collusion, designing an agreed format for financial reporting across the system, local monitoring that is frequent and independent of the principal and administrative staff, the use of independent auditors, external checking of statistics that are used in determining allocations, and clarity in explanations of funding formulae so that they can be readily understood by all stakeholders. The report concluded that “formula funding for schools reduces the potential for corruption by increasing transparency as the amount each school should receive and the basis for this is public knowledge” (Levačić and Downes, 2004: 145).

Impact on the Workload of Principals

A final issue is concerned with the impact of self-management on the role of the principal and other leaders at the school level. A study was conducted in Victoria (Department of Education and Training, 2004) on the workload in government (public) schools and its impact on the health and well-being of the principal and assistant principals. On workload, the number of hours per week for principals in Victoria was similar to that for principals in England, as reported in a survey at about the same time, being about 60 h per week. In both places, this is well above the average of leaders and managers in other professional

fields in several European nations (about 45 h per week). The report contained evidence of a negative impact on the emotional and physical well-being of principals.

There is also evidence that fewer people are seeking appointment to principal in countries with self-managing schools. In England, for example, a typical school seeks a new principal once every 7 years, which means about 14% advertise each year. The number advertising in 2005 was 12%, with about one-third unable to make an appointment after the initial advertisement. Education Data Surveys (EDS) reported that re-advertisement reached record levels (Smithers, 2006). Despite the workload and declining numbers seeking to be principals in some countries, it is clear from the results of surveys over a decade that most serving principals in systems of self-managing schools would not wish to return to more centralized arrangements (Bullock and Thomas, 1997; Caldwell and Spinks, 1998; Department of Education and Training, 2004).

Issues related to workload and willingness to serve include the preparation and ongoing support of principals and other school leaders, and these are addressed in the final section.

Prognosis

Self-managing schools have been one manifestation of a general trend to decentralization in public education in many countries since the late 1960s, with momentum gathering as the twentieth century drew to a close. The practice was introduced for a range of reasons but much of the heat from often contentious debates about its efficacy was dissipated in the early years of the twenty-first century, as most governments and system authorities settled on the enhancement of learning as its primary purpose. The logic of the argument was relatively straightforward: each school contains a unique mix of student needs, interests, aptitudes, and aspirations and those at the school level are best placed to determine the particular mix of all of the resources available to the school to achieve optimal outcomes (Cheng, 1996; Caldwell, 2003). Early research was generally unable to confirm the logic, either because the design of the reform did not include a connection to learning, or because the database on student achievement was poorly constructed, thus thwarting any effort to determine the connection. Research at macro- and micro-levels tends to confirm the association but it requires purposeful efforts by a skilled profession to make it effective. The agenda for the next decade and beyond will include a high priority on these efforts, especially if emerging success in a growing number of schools is to be scaled up across an entire system.

Early efforts in self-management placed the focus on management, with particular attention being given to

planning and resource allocation. It is understandable that preparation and professional development programs for leaders and managers at the school level tended to focus on these. However, with heightened expectations for schools, especially in terms of success for all students in all settings, the focus will continue to shift to leadership and the building of professional capacity to achieve an alignment of curriculum, pedagogy, and resources with the mix of learning requirements at the school level. In this context, the student becomes the most important unit of organization rather than the classroom, school, or school system (Caldwell, 2006).

If practice through the first decade of the twenty-first century is a guide, leaders in self-managing schools will need to be adept at drawing on all of the resources of the community to meet expectations, and these include intellectual capital, social capital, spiritual capital (broadly defined), as well as financial capital (Caldwell and Spinks, 2007). It will be a more demanding role than ever before and it is likely that new roles will emerge, extending those which made their appearance at around this time in England, with an executive head becoming a person of influence across a number of schools. Remuneration levels will be high. Support will be drawn from a range of sources and the central, regional, and district offices in systems of public education will need to take on stronger support roles than in the past.

See also: Autonomy; Principals Role in Restructuring Schools; School Effectiveness in Developed Societies; School Finance Reform; School Finance: An Overview; Studies of School Improvement in Developing Countries; Transformational School Leadership.

Bibliography

- Bullock, A. and Thomas, H. (1997). *Schools at the Centre: A Study of Decentralization*. London: Routledge.
- Butts, R. F. (1955). *Assumptions Underlying Australian Education*. Melbourne: Australian Council for Educational Research.
- Caldwell, B. J. (1977). Decentralized school budgeting in Alberta: An analysis of objectives, adoption, operation and perceived outcomes in selected school systems. Unpublished Doctoral Dissertation, Department of Educational Administration, University of Alberta.
- Caldwell, B. J. (2002). Autonomy and self-management: Concepts and evidence. In Bush, T. and Bell, L. (eds.) *The Principles and Practice of Educational Management*, ch. 3, pp 24–40. London: Paul Chapman.
- Caldwell, B. J. (2003). A theory of learning in the self-managing school. In Volansky, A. and Friedman, I. A. (eds.) *School-Based Management: An International Perspective*, ch. 3, pp 93–116. Israel: Ministry of Education.
- Caldwell, B. J. (2005). *School-Based Management. No. 3 in the Education Policy Series of the International Academy of Education*. Paris: International Institute of Educational Planning (IIEP) (UNESCO).
- Caldwell, B. J. (2006). *Re-Imagining Educational Leadership*. Camberwell/London: ACER Press/Sage.
- Caldwell, B. J. and Spinks, J. M. (1998). *Beyond the Self-Managing School*. London: Falmer.

- Caldwell, B. J. and Spinks, J. M. (2008). *Raising the Stakes: From Improvement for Transformation in the Reform of Schools*. London: Routledge.
- Cheng, Y. C. (1996). *School Effectiveness and School-Based Management: A Mechanism for Development*. London: Falmer.
- Department of Education and Training (DET) (Victoria) (2004). *The Privilege and the Price*. Melbourne: Department of Education and Training.
- Fullan, M., Hill, P., and Crévoila, C. (2006). *Breakthrough*. Thousand Oaks, CA: Corwin Press.
- Hanushek, E. A. (1996). Outcomes, costs and effectiveness in schools. In Hanushek, E. A. and Jorgenson, D. W. (eds.) *Improving America's Schools: The Role of Incentives*, ch. 3, pp 29–52. Washington, DC: The Brookings Institution.
- Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis* 19(2), 141–164.
- Interim Committee of the Australian Schools Commission (1973). Karmel, P. (chair) *Schools in Australia*. Canberra: Australian Government Printing Service.
- Kandel, I. L. (1938). *Types of Administration*. Melbourne: Australian Council for Educational Research.
- Levačić, R. (1995). *Local Management of Schools: Analysis and Practice*. Buckingham: Open University Press.
- Levačić, R. and Downes, P. (2004). *Formula Funding of Schools, Decentralization and Corruption: A Comparative Analysis*. Paris: International Institute of Educational Planning (IIEP) (UNESCO).
- Levačić, R. and Ross, K. N. (eds.) (1999). *Needs-Based Resource Allocation in Education via Formula Funding of Schools*. Paris: International Institute for Educational Planning (IIEP) (UNESCO).
- Malen, B., Ogawa, R. T., and Kranz, J. (1990). What do we know about site-based management: A case study of the literature – a call for research. In Clune, W. and Witte, J. (eds.) *Choice and Control in American Education Volume 2: The Practice of Choice, Decentralization and School Restructuring*, pp 289–342. London: Falmer.
- OECD (2004). *Education at a Glance*. Paris: OECD.
- OECD, Directorate of Education, Employment, Labor and Social Affairs, Education Committee (1994). *Effectiveness of Schooling and of Educational Resource Management: Synthesis of Country Studies. Points 22 and 23*. Paris: OECD.
- Schleicher, A. (2004). I risultati dell'Italia nell'indagine OCSE "Education at a Glance" Paris: OECD. <http://www.oecd.org/dataoecd/33/33/33732967.ppt> (accessed July 2009).
- Smith, M. S., Scoll, B. W., and Link, J. (1996). Research-based school reform: The Clinton Administration's agenda. In Hanushek, E. A. and Jorgenson, D. W. (eds.) *Improving America's Schools: The Role of Incentives*, Paper 2, pp 9–27. Washington, DC: National Academy Press.
- Smithers, R. (2006). 'Headteacher vacancies expose schools crisis'. *Guardian*, 12 January.
- Smyth, J. (ed.) (1993). *A Socially-Critical Review of the Self-Managing School*. London: Falmer.
- Summers, A. A. and Johnson, A. W. (1996). The effects of school-based management plans. In Hanushek, E. A. and Jorgenson, D. W. (eds.) *Improving America's Schools: The Role of Incentives*, Paper 5, pp 75–96. Washington, DC: National Academy Press.
- Woessmann, L. (2001). Why students in some countries do better: International evidence on the importance of education policy. *Education Matters*, Summer, 67–74.

System Leadership

D Hopkins and R Higham, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

System leaders are those headteachers who are willing to shoulder system wide roles in order to support the improvement of other schools as well as their own. Such system leadership is a new and emerging practice that embraces a variety of responsibilities that are developing in England either locally or within discrete national networks or programs. When taken together these roles may have the potential to contribute to a more school-led, rather than government-prescribed, education system. This article elaborates this concept of system leadership and illustrates its potential power as a catalyst for systemic renewal in three ways. First, it provides an initial conceptualization of system leadership and raises a series of concerns about the way the concept is being interpreted. Second, by drawing on research in England, it proposes taxonomy of roles that system leaders are currently assuming. Third, and based on these analyses, a potential model for system leadership is suggested and the tensions involved in further developing the concept are explored.

Defining System Leadership

The concept of system leadership is one that has recently caught the educational imagination. Take for example, these quotations from two significant opinion makers. The first from the general secretary of the Association of School and College Leaders in England and the other from a leading educational commentator whose work has a global reach.

Dunford (2005) in a recent address to the National Conference of the Specialist Schools and Academies Trust (SSAT) argued that:

The greatest challenge on our leadership journey is how we can bring about system improvement. How can we contribute to the raising of standards, not only in our own school, but in others and colleges too? What types of leaders are needed for this task? What style of leadership is required if we are to achieve the sea-change in performance that is demanded of us?

In *Systems Thinkers in Action*, Fullan (2004) argued that:

a new kind of leadership is necessary to break through the status quo. Systematic forces, sometimes called inertia, have the upper hand in preventing system shifts. Therefore,

it will take powerful, proactive forces to change the existing system (to change context). This can be done directly and indirectly through systems thinking in action. These new theoreticians are leaders who work intensely in their own schools, or national agencies, and at the same time connect with and participate in the bigger picture. To change organizations and systems will require leaders to get experience in linking other parts of the system. These leaders in turn must help develop other leaders within similar characteristics.

These quotations share three implicit assumptions. The first is that if we are ever to achieve sustainable education change, it must be led by those close to the school; the second is that this must have a systemic focus; and the third is that system leadership is an emerging practice. (As we will see, the conceptual concerns of system theory for relationships, structures, and interdependencies Katz and Kahn (1976), Senge (1990), Campbell *et al.* (1994) underpin the contemporary work of system leaders in practice.) There is also an implicit tension between system leadership being a national policy or a professional movement. This dichotomy, as we will see, has profound implications for the prospect of sustainable educational reform.

In *Every School a Great School* (Hopkins, 2007) it was suggested that the five striking characteristics of system leaders, those distinguishing them from broader collaborative activity, are that they deploy their experience, knowledge, and skills to:

- actively lead improvements in other schools and measure their success in terms of student learning, achievement, and welfare, both to raise the bar and narrow the gap(s);
- commit staff in their own and other schools to the improvement of teaching and learning, engaging them deeply in the organization of teaching, learning, curriculum, and assessment so as to ensure learning is increasingly personalized for students;
- lead the development of schools as personal and professional learning communities, building relationships across and beyond each school to provide a range of learning experiences and professional development opportunities;
- lead work for equity and inclusion through acting on context and culture, not only in response to poverty but also to employ educational resources to help give communities a greater sense of aspiration and empowerment; and

- manage strategically the impact of the classroom, school, and system on one another, understanding that in order to change the larger system one has to both engage with it in a meaningful way and manage subsequent change at a school level.

It is a compelling proposition that such leadership holds significant potential to contribute to systemic educational improvement. This is an idea that has been positively advocated over the past 4 years by government ministers. David Miliband, member of the parliament (MP), for example, when minister of state for school standards in the England, saw that the development and deployment of cadre of system leaders could go a long way to responding to the key challenges he had identified for school leadership as part of a new relationship between schools and government (Miliband, 2004). This included raising productivity in education, effecting greater social justice, and ensuring sustainable improvement (Miliband, 2003).

Taken together, this combination of informed comment and progressive government policy suggests that the concept of system leadership is an idea whose time has come. One can summarize these aspirations in saying that:

- by being willing and able to work for the success and well-being of students in other schools as well as their own, system leaders can show how self-managed schools, emerging from an era of competition, might work together for greater social equity;
- by seeing collaboration with other schools and agencies as vital to delivering on aims that individual schools cannot achieve by themselves, system leaders can create alternative solutions to a range of educational challenges that have traditionally become the responsibility and preserve of the central apparatus of the state; and
- by starting from the problems and solution of schools themselves, rather than those identified by others, system leaders can show how professionals might take more control of educational reform and contribute directly to a broader renewal of the education system.

However, before we get too carried away with too much enthusiasm for the concept of system leadership, we need to admit that much of the well-intentioned advocacy is based on aspiration and a few early individual cases rather than systemic evidence. Above all, and perhaps most significantly, it is not actually clear what or how many system leadership roles are currently being undertaken. Possibly as a result of being a relatively new professional practice, there appears to have been no attempt to date to document how system leadership is being enacted. In an attempt to correct this situation, research is reported in this article (Hopkins and Higham, 2007) that provides an initial perspective on the nature of system leadership in England in mid-2006. Based on this analysis, a taxonomy

of system leadership roles are outlined; an initial conceptualization and model of what effective system leaders do is proposed; and finally the key emerging challenges for the future development of this practice are explored.

Taxonomy of Roles

The research cited above pointed to five distinct yet overlapping categories of system leadership and leads to the following taxonomy.

First, are those headteachers who develop and lead a successful educational improvement partnership between several schools. These are most usually focused on a set of specific themes that have clear outcomes and reach beyond the capacity of any one single institution. Examples include partnerships on curriculum design and specialisms, including sharing curricular innovation, 14–19 consortia, and behavior and hard-to-place students. While such partnerships often currently remain in what is commonly referred to as soft organizational collaboratives, some have moved to harder more formalized arrangements in the form of (co)federations (to develop stronger mechanisms for joint governance and accountability) or education-improvement partnerships (to formalize the devolution of defined delivery responsibilities and resources from their local authority). As a result of recent legislation, such groupings now have the possibility of forming independent school trusts.

Second, are headteachers who choose to lead and improve a school in extremely challenging circumstances. A dual objective of system leadership is to both raise the bar and close the gap(s) in systemic student achievement. As such a key task is to change contexts in our most challenging circumstances by choosing to lead and improve low-achieving schools and then sustain them as high valued-added institutions over a significant period of time.

Third, are those headteachers who partner another school facing difficulties and improve it. This includes both executive heads and leaders of more informal improvement arrangements who are differentiated from category 1 on the basis that these leaders work from a lead school into a low-achieving or underperforming school (or schools) that require intervention. Executive heads provide an example. They are responsible for two or more schools that have either entered into a federation or a local (often time-bound) agreement focused on a lead school working to improve a partner.

Fourth, are headteachers who act as community leaders to broker and shape partnerships or networks of wider relationships across local communities to support children's welfare and potential. In related research, our colleague Matthews (2006) conceives of four key dimensions to this work as organizing resources for learning from the community, widening learning experiences beyond the

school, drawing support for child and family welfare into the school or network, and providing for the lifelong learning needs of the community.

Fifth, are those head teachers who work as change agents or expert leaders. The focus is on providing practical knowledge and guidance as well as the transfer of best practice within a formalized school-improvement program. This is currently the most numerous category and includes national-leadership roles developed within centrally organized programs such as School Improvement Partners, Consultant Heads, and/or National Leaders of Education.

What Effective System Leaders Do

In establishing taxonomy of the five key areas of identified system-leadership activity, what seems increasingly important is how these leaders actually work – for, as we are keenly aware, their work often enters new territories that have novel challenges and no well-rehearsed solutions.

An important perspective on this is offered by Heifetz's (1994) concept of adaptive leadership. Heifetz's argument is that leaders increasingly require skills that move beyond traditional management solutions for technical problems to provide adaptive responses to challenges without easy answers. Technical problems, such as how to teach numeracy, and their solutions will of course remain vital. But system leaders will also need to work adaptively to lead people and organizations beyond restrictive boundaries, perceived wisdoms, and entrenched cultures where they exist as obstacles to improvement.

This theme also underpins Fullan's (2005) exposition of the role he believes school leaders will need to play as system thinkers in action if sustainable large-scale reform is to be achieved. This, Fullan argues, will necessarily involve adaptive challenges that "require more sophisticated leadership" (p. 53). For Fullan, examples of this new work include: leading and facilitating a revolution in pedagogy (p. 57); understanding and changing the culture of a school for the better (p. 57); relating to the broader community, in particular with parents; and integrating and coordinating the work of social service agencies into the school as a hub (p. 61). This will demand:

above all ... powerful strategies that enable people to question and alter certain values and beliefs as they create new forms of learning within and between schools, and across levels of the system. (Fullan, 2005: 60)

These demands are further illuminated in theory by Senge (1990) who argues that for organizations to excel, they have to become learning organizations in which "new and expansive patterns of [system] thinking are nurtured ... and where people are continually learning to see the whole together" (p. 3). To Senge, the key to

becoming a learning organization is for leaders to tap into people's commitment and capacity to learn at all levels, so that broader systemic interdependencies and how to change them effectively can be made clearer (Senge, 1990: 4).

There is a loud and clear read across here from system theory to the key areas of system-leadership activity we have already identified. There is also a sense of a shared, central skill set for system leaders that needs to be effective and reflects the established literature on educational leadership (for a comprehensive review see Leithwood *et al.*, 1999).

There is also however a real concern about the increasing tendency in the literature to distort the generic competencies of leaders through celebrating singular aspects of the role. Leithwood *et al.* (2004: 4) express this worry succinctly:

... we need to be skeptical about the "leadership by adjective" literature. Sometimes these adjectives have real meaning, but sometimes they mask the more important themes common to successful leadership, regardless of the style being advocated.

These are wise words and of course it has not escaped our notice that we ourselves are in danger of being hoisted by our own petard. In defense, we would claim that the concept of system leadership is embracing rather than esoteric. This claim is made on three grounds. First, the concept of system leadership as we have seen locates itself within the general literature on systems theory and thinking and as such is inclusive rather than exclusive. Second, as we shall see shortly, system leadership is a theory of action that embraces a catholic range of disciplines in order to exert its power (see, e.g., Elmore, 2004; Leithwood *et al.*, 2006). As hopefully is becoming clear, system leadership will only exert any influence to the extent that it focuses on teaching and learning (i.e., is instructional), shares its authority with others (i.e., is distributed) and so on. To reiterate, system leadership as a concept is integrative rather than exclusive.

This discussion emphasizes the focus on the key capabilities required by system leaders. Inevitably, this demands the development of hypotheses as we move inductively from our data. Here we have found it helpful to draw on Leithwood and Riehl's (2003) core leadership practices in proposing an initial outline of the capabilities system leaders bring to their role. This is set out in **Figure 1**.

Building on these key capabilities, and combining them with the range of identified roles, it is possible to offer a model of system-leadership practice that emerges inductively from the actions of a research sample of system leaders. This is set out in **Figure 2**.

The model exhibits a logic that flows from the inside-out. At the center, leaders driven by a moral purpose related to the enhancement of student learning, and seek to empower teachers and others to make schools a critical force for

Core practices	System leadership components
Setting direction	<p>Total commitment to enable every learner to reach his/her potential with a strategic vision that extends into the future and brings immediacy to the delivery of improvements for students.</p> <p>Ability to translate vision into whole school programs that extend the impact of pedagogic and curricular developments into other classrooms, departments, and schools.</p>
Managing teaching and learning	<p>Ensure every child is inspired and challenged through appropriate curriculum and a repertoire of teaching styles and skills that underpin personalized learning.</p> <p>Develop a high degree of clarity about and consistency of teaching quality to both create the regularities of practice that sustain improvement and to enable sharing of best practice and innovation.</p>
Developing people	<p>Enable students to become more active learners, develop thinking and learning skills, and take greater responsibility for their own learning. Involve parents and the communities to minimize the impact of challenging circumstances on expectations and achievement.</p> <p>Develop schools as professional learning communities, with relationships built and fostered across and beyond schools to provide a range of learning experiences and professional-development opportunities for staff.</p>
Developing the organization	<p>Create an evidence-based school, with decisions effectively informed by student data, with self-evaluation and external support used to seek out approaches to school improvement that are most appropriate to specific contextual needs.</p> <p>Extend an organization's vision of learning to involve networks of schools collaborating to build, for instance, curriculum diversity, professional support, extended and welfare services, and high expectations. In so doing, build a school's capacity to support wider system-leadership roles.</p>

Figure 1 A conceptualization of key system-leader capabilities.

improving communities. This is premised on the argument already made, that sustainable educational development requires educational leaders who are willing to shoulder broader leadership roles – those who care about and work for the success of other schools as well as their own.

It is also clear from our research that system leaders share a characteristic set of behaviors and skills. As illustrated in the second inner ring of **Figure 2**, these are of two types. First, system leaders engage in personal development usually informally through benchmarking themselves against their peers and developing their skill base in response to the context they find themselves working in. Second, all the system leaders we have studied have a strategic capability; they are able to translate their vision or moral purpose into operational principles that have tangible outcomes.

Taken together these two central circles of **Figure 2** reflect the core practice of setting directions as noted in **Figure 1**. As is denoted in the third ring of the model, the moral purpose, personal qualities, and strategic capacity of the system leader find focus in three domains of the school detailed above – managing the teaching and learning process, developing people, and developing the organization.

Finally, although there are a growing number of outstanding leaders that exemplify these qualities and determinations, they are not necessarily system leaders. A system leader not only needs these aspirations and capabilities but in addition, as seen in the outer ring of the model, works to change other contexts by engaging with the wider system in a meaningful way. We have included in the outer ring, the range of roles identified from the research that focuses on improving other schools, sharing curriculum innovations, empowering communities, and/or leading partnerships committed to enabling all schools to move forward. The model represents a powerful combination of practices that give us a glimpse of leadership in a new educational landscape. Realizing that landscape, however, may also require a bigger shift within the broader education system.

Developing System Leadership

It is clear that the practice of school-led system leadership will bring new challenges and roles for school leaders and

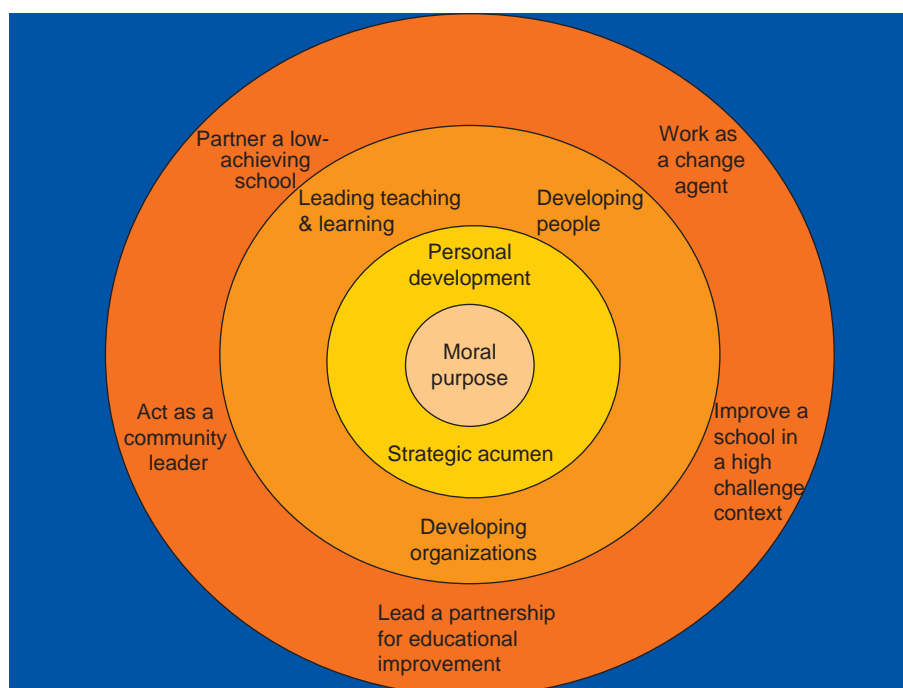


Figure 2 A model of system-leadership practice.

their schools. What is less clear is what this might mean for other system actors, most notably government. If school leaders are to take on wider responsibilities for system reform, how should a government work to develop such activity? Would a real and authentic commitment by government to do so inevitably mean a rebalancing of roles, agency, and control between schools and the government? In considering these issues, we now turn finally to explore some of the key systemic challenges for government.

As we have seen, fundamental elements of system leadership are the more authentic collaboration on and transfer of school-improvement intelligence and leadership best practice by experienced leaders. These are inherently professionally led, bottom-up solutions. Yet, they are also solutions to problems that have traditionally been the responsibility and preserve of the central apparatus of the state. This includes the deeply ingrained workings of the accountability, funding, and governance systems that place the unit of an individual school at their center. It also includes the location of agency and incentives, and the focus of support and professional development.

In considering these tensions, it is worth briefly returning to a distinction that was identified earlier between system leaders working in national programs and those working in locally organized often *ad-hoc* roles. The majority of school-level system leaders operate in national programs that have incentivized activity through organization, funding, and professional development that have in turn created new opportunities for headteachers. This is the enabling state at work. It is an important step toward rebalancing

agency by making it more possible for headteachers to lead technical and adaptive solutions in a widening professional domain of cross-school and system improvement.

Yet, within these opportunities, we may already be witnessing limitations of government-led activity. For, while new leadership roles emerge, the government's tendency to check and control does not seem to significantly diminish. This tendency is related to a focus on effectiveness and value for money. But it also seems to betray a government that has yet to develop sufficient trust in the profession. The result is less-than-intelligent accountability and at the extreme, a tick-box bureaucracy, rather than dynamic system reform.

A greater degree of freedom exists on the other side of the divide, in the roles we identified as being locally developed, often *ad hoc* and contextually responsive. In such activity, professionals not only deploy their experience and skill to lead improvements, they also define the terms on which such activity is undertaken and sustained. In this way, it is understandable why a number of our respondents conceived of these roles to be a more authentic form of system leadership. With no single framework or protocol, a range of models is developed in relation to specific needs (and times). Some centrally driven momentum is sacrificed, but on the principle that system leadership must inherently be a professionally led agenda. Furthermore, from this perspective, the role of an enabling state becomes focused on reducing barriers to collaboration and wider policy disincentives, with national agencies providing bespoke professional development to individual

system leaders. But there are limitations here too. Momentum seems to be significantly sapped by the time taken and energy used to develop and agree to the relatively procedural project and governance structures that underpin these roles. There will be concerns for rigor if effectiveness is not clearly defined or where an individual system leader does not have a wider perspective on emerging best practice.

There are of course variations to this bottom-up/top-down dialectic. For instance, strategic local-leadership partnerships already exist between headteachers and local authorities. In one such model, the authority retains legal responsibility for value for money while delegating decision making to a partnership of headteachers who bring coherence and accountability to local collaboration. A perspective on how these (and other) possibilities may inform current professional action and government influence will be dependent on a range of criteria. If, however, a shared criterion is to develop effective system leadership in a growing number of schools, then the following three suggestions for more short-term action may prove instructive:

1. *Incentivize rather than legislate.* The traditional response has been intervention and management from government, national agencies, or local authorities. The argument here is that this leadership now needs to come more from headteachers themselves or from agencies committed to working with them in authentic ways. It is clear that the more bureaucratic the response, the less likely it will work. A more lateral approach may be to create the conditions within the system to promote system leadership and collaborative activity through, for example, adjusting accountability requirements and funding for capacity building. With the right incentives in place, schools will naturally move toward these new ways of working and mold them to the context in which they operate and to the challenges they face.
2. *Place the agency close to the school.* There are now in England, at least three emerging change-agent roles within the system – National Leader of Education and School Improvement Partners – whose remit is specifically school improvement. The intention that must be maintained is that instead of creating a new bureaucracy, their brief is increasingly focused on facilitating relationships between schools to maximize the potential of purposive collaboration. This approach to school transformation is made increasingly possible by the highly sophisticated data now available on school and student performance. It enables groups of schools to identify (1) issues where they share both strengths and weaknesses, that is, their capacity for sharing and (2) common issues where they are likely to need some external input.
3. *Use school independence collaboratively to tackle inequality.* The underlying assumption here is that if state schools

are to become increasingly independent of local control, they must also be encouraged to work collaboratively. To tackle local inequalities, which have been entrenched during two decades of quasi-market reforms, a crucial condition is that all schools increasingly accept responsibility for the education of students within their geographic area. For example, this may enable the best of practice in the most successful federations to apply across other schools and to bring together a range of policy initiatives to give a real bite to local renewal.

A Final Word

This article has provided a conceptualization of system leadership and mapped the current system-leadership landscape based on data gathered in England in 2006. Although based on the English experience, the analysis may have a wider application. It has been demonstrated that system leadership can contribute decisively to a full range of government and local agendas by sharing of expertise, facilities and resources, innovation and creativity, leadership and management, and vocational education and skills support. In addition, a full range of children's-services agendas and constructive links between parents and schools, businesses, and further higher education providers and schools may be best served by such arrangements. The collective sharing of skills, expertise, and experience creates much richer and more sustainable opportunities for rigorous transformation than can ever be provided by isolated institutions.

We have also seen, however, that the notion of system leadership itself is not unproblematic. There is evidence to show that replicating best practice from one school to other is not easily achieved. It is clear that there are limits to headteachers working across the system without becoming distracted from sustaining improvement in their own schools. There are also significant contradictions within, and tensions between, government-led system leadership and demands for giving school leaders more agency to take the lead. Above all, there is a lack of evidence to demonstrate that school leaders engaged in system thinking and action can consistently provide solutions to complex systemic problems. Yet, while the whole idea of system-level reform is territory that is neither clearly charted nor uncontested, it is as we have seen, terrain that is beginning to be productively explored both in practice as well as theory by the new breed of educational-change adventurers.

Acknowledgment

Some of the research reported in this article was funded by a grant from the National College for School Leadership. Their support is gratefully acknowledged as is the

contribution of Elpida Ahtaridou who assisted with collection of the survey data.

Bibliography

- Campbell, D., Coldicott, T., and Kinsella, K. (1994). *Systemic Work with Organisations. A New Model for Managers and Change Agents*. London: Karnac Books.
- Dunford, J. (2005). Watering the plants: Leading schools and improving the system. *Address to the National Conference of the Specialist Schools and Academies Trust*.
- Fullan, M. (2004). *Systems Thinkers in Action: Moving beyond the Standards Plateau*. London/Nottingham: DfES Innovation Unit/NCSL.
- Fullan, M. (2005). *Leadership and Sustainability: System Thinkers in Action*. London: Sage.
- Heifetz, R. A. (1994). *Leadership without Easy Answers*. Harvard: Belknap Press.
- Hopkins, D. (2007). *Every School a Great School*. Buckingham: Open University Press.

- Katz, D. and Kahn, R. L. (1966). *The Social Psychology of Organizations*. New York: Wiley.
- Leithwood, K., Jantzi, D., and Steinbach, S. C. (1999). *Changing Leadership for Changing Times*. Buckingham: Open University Press.
- Leithwood, K. and Riehl, C. (2003). *What We Know about Successful School Leadership*. Nottingham: NCSL.
- Matthews, P. (2006). Aspects of system leadership: Community leadership. *Paper Presented to NCSL*.
- Miliband, D. (2003). *School Leadership: The Productivity Challenge. Address at the National College for School Leadership, October 2003*.
- Miliband, D. (2004). A new relationship with schools. *Address at the North of England Conference, January 2004*.
- Senge, P. (1990). *The Fifth Discipline*. New York: Doubleday.

Further Reading

- Higam, R. (2006). Sustaining improvement: System leadership in challenging circumstances. *Paper Presented to NCSL*.

LEADERSHIP AND MANAGEMENT – SCHOOL AND COMMUNITY

Contents

Educational Leadership and Social Capital

Networked Learning Communities School-to-School Collaboration as an Essential Component of a System Reform Strategy

Networks and Communities of Knowledge

School Community Relations

Social Capital, Educational Institutions and Leadership

Educational Leadership and Social Capital

J West-Burnham, St Mary's University College, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bonding – The means by which a community develops social capital through developing homogeneity and exclusivity.

Bridging – The means by which a community develops social capital by engaging with other communities through inclusive relationships.

Community – Any human social enterprise which obtains its identity through a geographical, social, religious, intellectual, or historical sense of shared identity.

Find and fix – A management strategy which is essentially reactive - waiting for things to go wrong and then repairing them.

Predict and prevent – A strategy which uses measurement to predict failure in a system so allowing for early intervention to prevent the failure occurring.

Social capital – The extent to which a community (qv) is effective in terms of the quality of networks, social relationships and interdependent, collaborative approaches to social life.

Trust – A pivotal component of social capital which can be best understood in terms of confidence, consistency and integrity in human relationships.

new bottles or re-branding to keep up with changing intellectual fashions. There is little doubt that the publication of Putnam's (2000) *Better Together* has led to widespread debate and interest. This article seeks to clarify social capital as an intellectual construct, to consider its implications for education, and then to review its potential impact on our current understanding of educational leadership.

The central thesis of this discussion is that social capital provides an integrated conceptual framework to amalgamate two areas which have historically been seen as distinct – the school as an organization and the school in its environment. For a generation educational policy, research and practice have focused on the school as an autonomous institution with models of school effectiveness and school improvement being descriptive and/or normative of the school *qua* school. This led in turn to a model of the school managing its boundaries with an emphasis on various permutations and applications of commercial models of marketing. This autonomy was reinforced in turn, noticeably in England and parts of the USA, by a model of accountability which was institutionally focused, performance related, market orientated, and made no concessions to social or environmental factors. Underpinning these elements was the widespread hegemony around the nature of professional autonomy which was extrapolated into institutional autonomy.

A range of trends have led to a questioning of the school as an autonomous and self-referential institution:

- the plateauing of performance with existing strategies demonstrably not working in certain systems;
- recognition of the correlation between social justice, child well-being, and high academic performance;

Introduction

To the cynic the current growth in interest in the concept of social capital might appear as a classic case of old wine in

- the stubborn nature of the tail of underachievement in some school systems and its failure to respond to existing policies and nostrums;
- increasing awareness of the high correlation between social factors and educational attainment and achievement;
- recognition of the benefits of networks and clusters in securing improved outcomes; and
- increasing concern about the viability and sustainability of existing patterns of school organization given demographic changes in the teaching profession.

The recognition of these trends has coincided with the growth in awareness and understanding of social capital as both an explanation of how societies and organizations work and a strategy to secure improvement and effectiveness in both.

In reviewing the implications of social capital for educational leadership this article focuses on the following themes:

- social capital as an intellectual construct,
- defining social capital,
- implications for systems and schools, and
- implications for educational leadership.

Social Capital as an Intellectual Construct

The provenance of the current usage of the concept of social capital can be attributed (with increasing confidence) to Bourdieu, Coleman, and Putnam respectively: through the work of these writers the concept has grown in clarity, specificity, availability, and applicability. In *Reproduction in Education, Culture and Society* (Bourdieu and Passeron, 1970), the concept of different types of capital is introduced as a way of describing and explaining power relationships in society and the dominance or subordination of certain groups via the interaction of class relationships and cultural norms.

Coleman's interest in social capital emerged from a concern to explain the factors influencing educational achievement – notably through an understanding of social inequality:

...social capital is the set of resources that inhere in family relations and in community social organization and that are useful for the cognitive or social development of a child or young person. (Coleman 1994: 300)

Coleman introduced the concept of social capital as being primarily concerned with the nature and quality of relationships, networks, and attitudes. His work is also significant because of its empirical basis.

Bowling Alone (Putnam, 2000) is the tipping point in terms of awareness of the concept of social capital and the widespread adoption of the concept as a means of

explaining the relative success or failure of societies. Putnam defines social capital as the:

...connections among individuals – social networks and the norms of reciprocity and trustworthiness that arise from them. In that sense social capital is closely related to what some have called “civic virtue.” The difference is that “social capital” calls attention to the fact that civic virtue is most powerful when embedded in a dense network of reciprocal social relations. (p 19)

For Putnam, social capital is about reciprocity, trustworthiness, mutual obligation, networks, and community. It is Putnam who makes the explicit link between social capital and community – talking of them as conceptual cousins. Putnam specifically reinforces Coleman's emphasis of the importance of social capital to education:

Child development is powerfully shaped by social capital. ...trust, networks, and norms of reciprocity within a child's family, school, peer group, and larger community have wide-ranging effects on the child's opportunities and choices and, hence, on his behavior and development. ...Social capital keeps bad things from happening to good kids. (Putnam 2000: 296)

...social capital is second only to poverty in the breadth and depth of its effects on children's lives.

(Putnam 2000: 297)

Since 2000, the concept of social capital has become increasingly available and understood as both a means of describing and explaining the nature of society and as the universal panacea – it seems most of the world's ills will be cured by increasing the level of available social capital.

In spite of this wide-ranging acceptance, there are problems with the concept of social capital. For example, in spite of Putnam's empirical base, is social capital anything more than a metaphor? It could be argued that it is as elusive as happiness or wellbeing – measuring the components does not describe the total human experience. Equally, there is a problem in equating human experiences with capital when used in an economic sense – social capital is simply not the same as economic capital – and the same is true of human or intellectual capital. In the final analysis, economic capital is money – which can be measured, transferred, and substituted. There are those in education who find introducing the notion of capital into any discussion of educational issues offensive. However, even in business there are significant changes in thinking taking place:

The new art and science of wkinomics is based on four powerful new ideas: openness, peering, sharing, and acting globally. These new principles are replacing some of the old tenets of business. (Tapscott and Williams, 2006: 20)

Thus, competition is replaced by collaboration, secrecy by sharing, hierarchy by lateral relationships, all in the

context of an international perspective and open communication. In many ways, Tapscott and Williams are describing a process in some businesses that is very similar to the one that is being advocated for schools in this article.

There is an increasingly robust empirical base to inform the discussions about social capital and this base is growing in terms of validity and reliability. Although the debate about social capital is growing in analytical integrity, much of the language that surrounds it remains normative and subject to a wide range of complex and contingent variables. It remains very difficult to develop a historical, geographical, or cultural comparative framework to fully explore and understand the dynamics of social capital. Thus, much of the debate about the nature of community tends to be normative rather than analytical.

Putnam's work is compelling because of its intuitive appeal – we want to live in communities. What he describes corresponds to the archetype of how we want to be. Yet, accepting the problems with capital, there are issues with social capital in terms of political power, ethnicity, gender, and, fundamentally, the changing nature of society. Is social capital a historical phenomenon, descriptive of a stage in the evolution of society which we are nostalgic about in the same way that we might mourn the loss of the comforts and certainties of childhood? Just as Christmas cards depict scenes and images that exist only in a utopian collective consciousness that owes more to reminiscence therapy than any view of reality, so also, the debate about social capital can be self-indulgent and idealized.

If social capital is only a metaphor, then it is a very useful and robust metaphor; however, its evidential base is sufficiently rigorous to allow us to use it to explore, analyze, explain, and act on the many social dimensions of education and schooling.

Defining Social Capital

The purpose of this section is not to problematize what social capital might be, but rather to see it as a theory-in-action and explore how it might inform the key debates about the nature of education and educational leadership. The scholarly writing on social capital tends to eschew detailed itemization of its components. An inductive reading of Putnam (2000) on social capital and Delanty (2003) on community produces the following criteria for social capital expressed through the concept of an effective community:

shared norms and values;
sophisticated networks and relationships;
volunteering and community action;
civic participation and engagement;
interdependence and reciprocity;
open communication;

trust;
shared rituals and symbols; and
common aspirations.

This list is not intended as a set of prescriptive criteria but rather as a means of identifying the significant variables that define a community. Their relative significance varies according to context. However some, particularly in an educational context, do seem to have relatively greater significance. For example, there is a very well-established consensus that any community or organization is successful to the extent that there is an alignment on norms and values. A shared ethical foundation expressed in clear values and consistent moral behavior seems fundamental to any human endeavor. There is robust evidence that points to successful schools having a clear sense of identity based on common values; the lack of such a sense of shared identity is the negative corollary that often explains failing schools.

Equally there is a growing recognition of the centrality of trust in social interactions. There is an interesting parallel between countries with high levels of trust and societies in which children have high levels of well being as shown in **Table 1**.

The work of Bryk and Schneider (2002) has demonstrated the influence and significance of trust at institutional level – they point to direct correlations between trust and academic achievement:

That is, even after controlling for differences among schools in various aspects of school context, student composition, and teacher background, we still find strong effects linking changes in relational trust to improvements in academic productivity. (pp 113–114)

As a social resource for school improvement, relational trust facilitates the development of beliefs, values, organizational routines, and individual behaviours that instrumentally affect students' engagement and learning.

(Bryk and Schneider 2002: 115)

Thus it may be that a focus on social capital provides a potent strategy to focus on the issues identified in the introduction of this article.

It is easy to see how the above criteria for an effective community could be applied to almost any successful human enterprise – a village, a suburb, a sports team, a

Table 1 Levels of social trust and child well-being

<i>Levels of social trust</i>	<i>Child well-being</i>
Denmark	Netherlands
Sweden	Sweden
Norway	Denmark
Iran	Finland
Finland	Spain

From Halpern, D. (2005). *Levels of Social Trust*. p 60; and UNICEF (2007). *Child Well Being*. p 4.

military unit, a criminal gang, and a terrorist group. It is these last two examples that point to one of the problems with social capital – it is morally neutral. The criteria could well describe a high performing school, but they could also describe the most anti-social street gang. Much of the writing about social capital assumes a liberal, democratic, and pluralist context – it is both a product of and a reinforcing agent for such hegemony. However, if taken out of our consensual context, then it could well be used to undermine all that is taken for granted in much of the literature. Even where there is an apparent moral consensus, it can often be very fragile (e.g., New Orleans in the aftermath of Hurricane Katrina in 2005). Perhaps the important point here is that it is not just a matter of having shared norms and values but resolutely reviewing why they are held as well as what they are.

Social capital is what creates a community, yet there is debate as to what constitutes a community – is a family a community? Are the supporters of Manchester United (found in the most remote parts of the world) a community? Does geographical proximity constitute a community? Are the contributors to this volume a community? It seems that it is not enough to be within shouting distance (Aristotle's definition of a civic society) to be a community, or to wear the same uniform, or not to meet together regularly. What makes the difference according to Putnam (2000) is the way in which communities cohere and interact.

Of all the dimensions along which forms of social capital vary, perhaps the most important is the distinction between bridging (or inclusive) and bonding (or exclusive). Some forms of social capital are, by choice or necessity, inward looking and tend to reinforce exclusive identities and homogeneous groups. Examples of bonding social capital include ethnic fraternal organizations, church-based women's groups, and fashionable country clubs. Other networks are outward looking and encompass people across diverse social cleavages. Examples of bridging social capital include the civil rights movement, many youth service groups, and ecumenical religious organizations. (p 22)

Bonding social capital constitutes a kind of sociological superglue, whereas bridging social capital provides sociological WD-40. (Putnam 2000: 23)

It is the extent to which there is bonding of all the criteria listed above that an effective community will function. According to the type of community, bonding will often have to be extended into bridging: a key proposition of the next section of this article is that schools have become very good at bonding but will need to engage in bridging to respond to the challenges identified in the introduction to this article. A contemplative order of nuns will only be concerned with bonding; a mendicant order of friars will live their lives by bonding and bridging. Bonding is

characterized by introversion and homogeneity, bridging by extroversion and heterogeneity – it is very difficult to reconcile such polarized imperatives; yet, the complexity of modern society may require leadership that integrates these apparently contradictory imperatives.

Diamond (2005) tells the tragic story of the medieval Scandinavian settlement in Greenland that starved to death over a winter because their cultural norms and social expectations prevented them from learning from their Inuit neighbors. Bonding made them a community, failure to bridge meant that they died within sight of help. The same issue may, in less-stark terms, be facing schools. The focus on social capital is essentially recognition of the basic interdependence of all people and all human endeavors. At its most basic, social capital is about recognizing common humanity as the basis of human activity, and education is no exception to this.

Implications for Systems and Schools

Perhaps the most significant issue raised by a focus on social capital is the debate around the variables that influence educational success, however defined. A focus on the significance of the community helps create a new perspective on the work of educational leaders:

At present, the tragedy of school change is that only about 30 per cent of the explanation for variations in school achievement appears to be attributable to factors in the school. . . Perhaps it is now time for leaders to lead their schools and exert their influence far beyond the school walls. . . (Moreno *et al.*, 2007: 5)

The movement from bonding to bridging represents a fundamental challenge to prevailing orthodoxies concerned with the nature of national education systems and the basic premise on which most schools operate. In both cases, the system and the school tend to operate in parallel with other services but not, in any way, integrated with them. The classic image is that of the silo – provision for children and young people can be seen as a number of silos (education, social services, health, housing, etc.) – each operating according to its own model of professionalism, bureaucratic systems, culture, priorities, and sense of purpose. Education systems work by bonding, so do schools. This is manifested in cultural and organizational territorial imperatives and models of professional autonomy which define the dominant models of most public sector organizations. This also explains much of the orthodoxy around school improvement, as the Audit Commission (2006) of the United Kingdom describe it:

Traditional school improvement activity has tended to concentrate on teaching and learning at individual school level. Critical though this is, by itself the approach is limited. . . (p 2)

They argue that the reason for this is:

Children's educational underachievement is linked with a wide range of deprivation factors: low parental qualifications, low family income, ill-health, family problems and wider community factors such as low aspirations and unemployment. (p 4)

This leads to their, somewhat radical, assertions that:

Because of this, improving schools and improving the prospects of the most disadvantaged pupils in schools is not a matter for schools alone. . . (p 7)

And:

School improvement and renewal are inseparable issue from neighbourhood improvement and renewal, particularly in the most disadvantaged areas. While schools are profoundly affected by their neighborhoods, the equally have a key role in promoting cohesion and building social capital. (p 7)

The movement to bridging, in order to respond to the issue of social capital, requires a paradigm shift of the most profound nature. In essence, this involves a movement from parallel silos to integrated services. The specific implications of this movement might be summarized as follows:

1. A reconsideration of core purpose: a movement away from education systems and schools being focused on academic attainment and outcomes to a more holistic and integrated approach to maximizing life chances.
2. The restructuring of services: the realignment of provision to reflect the focus on holistic approaches, that is, the subsuming of education and other services concerned with children and young people into one coherent and interdependent strategy. One implication of this would be all relevant data about every child will be available to all professionals.
3. Shared resources: schools, in particular becoming community resources rather than just collections of classrooms functioning for about one sixth of the year. This implies community access on equal terms (rather than on a grace and favor basis) and schools sharing accommodation with other public services, for example, social services, police, health promotion, and services for all age groups.
4. Integrated strategies: moving away from piecemeal national policies randomly applied to the development of specific cross-community initiatives focusing on issues such as intervention in the early years, healthy living initiatives, drug abuse programs, teenage pregnancy, etc. Further strategies might apply to the socially disadvantaged, the elderly, the unemployed, etc.
5. Changes in roles: recognizing that the education of a child or young person is more than just delivering the curriculum.

Fundamental to all of the above is a switch in culture which can best be summarized as a movement from find and fix to predict and prevent. Education systems and schools have long worked on a philosophy of cure rather than prevention: early years' provision compensates for inadequate family life; primary education compensates for the lack of early years' provision; secondary schools remediate the failings of primary schools; universities and employers bemoan the time they have to spend compensating for the inadequacies of secondary schools, and when all else fails, prisons provide basic education.

The data that we now have on the impact and significance of social capital in terms of life chances and educational success allows us to begin to formulate strategies based on identification of potential failure based on predictive evidence and intervention strategies to prevent or minimize such failure. However, the professional culture and prevailing policies in many countries are rooted in cure rather than prevention – witness the relative spending on health promotion and surgical intensive care and early years' provision and higher education. It is no coincidence that the professional silos of health and education give higher status to the consultant and professor than workers in health promotion and early years' workers.

A practical example of cautious and tentative moves in the direction being advocated above is the every-child-matters strategy, the related Children Act (2004) and the Children's Strategy in England. They do represent a significant realignment of the public services in terms of purpose, structure, and roles; but in terms of both principle and practice, they have to work alongside the prevailing culture of academic standards based accountability – which makes no concessions to the issues emerging from the impact of social capital on educational achievement. It is rather like two tectonic plates grinding away, generating a great deal of heat, but with the new continent yet to emerge.

The OECD PISA (2007) studies have demonstrated a high correlation between educational and academic success and countries with relatively lower levels of social disadvantage. Putnam (2000) makes an explicit and direct connection between states in the USA with high social capital and states with higher educational outcomes. It may be that the emphasis on school improvement has to be matched, or exceeded, by investment in social improvement.

Implications for Educational Leadership

Although beset by semantic differences, there is a clear consensus across the canon of educational leadership that the primary focus of educational leadership is the school *qua* school. Most models of effective leadership in education, accountability structures, and professional, contractual, and social expectations focus on the school as an organization. The dominant themes of most programs preparing school leaders are focused on internal effectiveness

with phrases such as boundary management or marketing being used to describe relationships with factors external to the school.

In the policy context, this has led to the focus on school improvement with an emphasis on enhancing the performance of the individual school by addressing its internal processes and structures. The concentration on parental choice as a mechanism for improving performance in some school systems has tended to reinforce institutional autonomy by introducing competition – for those who are able to exercise choice in the marketplace.

There are thus contradictory imperatives at work but the issues are very clear:

These new theoreticians are leaders who work intensely in their own schools, or national agencies, and at the same time connect with and participate in the bigger picture. To change organizations and systems will require leaders to get experience in linking to other parts of the system. These leaders in turn must develop other leaders with similar characteristics. (Fullan, 2004: 9)

We need to develop a new generation of leaders who come to the task with a different understanding of the job and with a different skill set than previously required. Being masters of space and place must yield to proficiency with connection, communication and collaboration. (Houston, 2004: 2–3)

In systems thinking, those at both local level and at the centre take into account each other's world, i.e. their world-view enlarges. Recall Senge's phrase – "a shift in mind from seeing ourselves as separate from the world to connected to the world"... (Fullan, 2004: 12)

However this is more than a realignment of roles and responsibilities; it is a fundamental reorientation of what it means to be an educational leader.

The hardest part of sustainable leadership is the part that provokes us to think beyond our own schools and ourselves. It is the part that calls us to serve the public good of all people's children within and beyond our community and not only the private interests of those who subscribe to our own institution. ...Sustainable leadership is socially just leadership, nothing simpler, nothing less.

(Hargreaves and Fink, 2006: 158).

In essence, introducing social capital and community into the complex equation that defines educational leadership requires a movement from bonding to bridging across the system, in the school and at a personal moral and emotional level. In practical terms, this means leaders developing the knowledge, behaviors, and qualities to bond and bridge in the following contexts:

- the school as a community;
- relationships with other schools;

- relationships with other agencies and aspect of provision for young people;
- links with employers and universities;
- engagement with international partners; and
- engagement with families and the immediate community that the school serves.

Of course, most schools will already be involved in many of these areas. The important point is the basis on which they are involved, that is, moving from bonding to bridging; from organizational effectiveness to community development; from marketing and liaison to building networks – all rooted in developing and nurturing communities rich in social capital.

This perspective has numerous implications for our understanding of educational leadership. It is not so much a matter of redefining leadership as realizing the core knowledge, behaviors, and qualities. The themes that would seem to be most significant might include:

1. moral leadership which is committed to education as social justice and an overarching belief in securing equity and excellence for all;
2. clarity of purpose which focuses on the education of the whole person – not just the academic achievement of the student;
3. the ability to build trust through consistency, credibility, and confidence expressed primarily in authentic relationships;
4. working through networks and building complex coalitions through consent and consensus rather than control;
5. Sharing leadership by moving from leadership as personal status to leadership as collective capacity across the wider community; and
6. The willingness to work as social entrepreneurs, building social capital in communities and acting as an agent for social change.

These themes, in turn, imply a movement away from positional power to moral purpose, from personal status to collective authority, from control to consent and, most significantly, from the institution to the system – from bonding to bridging.

Bibliography

- Audit Commission (2006). *More Than the Sum*. London Audit Commission.
- Bourdieu, P. and Passeron, J. -C. (1970). *Reproduction in Education, Culture and Society*. London: Sage.
- Bryk, A. and Schneider, B. (2002). *Trust in Schools*. New York: Russell Sage Foundation.
- Coleman, J. S. (1994). *Foundations of Social Theory*. Cambridge, MA: Belknap.
- Delanty, G. (2003). *Community*. Oxford: Routledge.
- Diamond, J. (2005). *Collapse*. London: Allen Lane.
- Fullan, M. (2004). *Systems Thinkers in Action: Moving beyond the Standards Plateau*. Nottingham: DfES Publications.

- Hargreaves, A. and Fink, D. (2006). *Sustainable Leadership*. San Francisco, CA: Wiley.
- Houston, P. D. (2004). *Advancing System Leadership*. <http://www.aasa.org/publications/execprespectdetail.cfm> (accessed May 2009).
- Moreno, M., Mulford, B., and Hargreaves, A. (2007). *Trusting Leadership: From Standards to Social Capital*. Nottingham: NCSL.
- PISA Results from (2000). *Student Engagement at School*. OECD. <http://www.oecd.org> (accessed May 2009).
- Putnam, R. (2000). *Bowling Alone*. New York: Simon and Schuster.
- Tapscott, D. and Williams, A. (2006). *Wikinomics: How mass collaboration changes everything*. London: Atlantic Books.
- Baron, S., Field, J., and Schuller, T. (eds.) (2000). *Social Capital: Critical Perspectives*. Oxford: Oxford University Press.
- Collarbone, P. and West-Burnham, J. (2008). *Systems Leadership*. London: Continuum.
- Field, J. (2003). *Social Capital*. Abingdon, PA: Routledge.
- Fullan, M. (2003). *The Moral Imperative of School Leadership*. Thousand Oaks, CA: Corwin Press.
- Halpern, D. (2005). *Social Capital*. Cambridge: Polity Press.
- Putnam, R. and Feldstein, L. M. (2003). *Better Together: Restoring the American Community*. New York: Simon and Schuster.
- Sergiovanni, T. J. (2005). *Strengthening the Heartbeat: Leading and Learning Together in Schools*. San Francisco, CA: Jossey-Bass.
- UNICEF (2007). *Child Poverty in Perspective: An Overview of Child Well-being in Rich Countries*. Innocenti Report Card 7.
- West-Burnham, J., Farrar, M., and Otero, G. (2007). *Schools and Communities*. London: Network Continuum Education.

Further Reading

Amit, V. (2002). *Realizing Community*. London: Routledge.

Networked Learning Communities School-to-School Collaboration as an Essential Component of a System Reform Strategy*

D Jackson, National College for School Leadership, Bedford, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Education action zones (EAZs) – A government program aimed at raising pupils' standards of achievement in areas of social and economic disadvantage by providing additional support to groups of schools and other partners to work collaboratively. The program focused on achievement in key curriculum areas: literacy, numeracy, information and communications technology, personal and social development, and professional development.

Excellence in cities (Eic) – A support program delivered locally by schools working in partnership with their local authority. It targeted deprived areas of the country seeking to transform urban secondary education by developing strategies and resources focused on teaching and learning, behavior and attendance, and leadership.

Innovation unit (IU) – The unit in England established by the government to foster practitioner-led innovation. Its role is to stimulate, incubate, and accelerate innovative practices on behalf of wider system learning and policy learning. Now in its second phase of development, most of IU's programs use the disciplines of the next-practice methodology that it has developed.

National College for School Leadership (NCSL) – The world's first national agency established in 2001, for the purpose of developing and supporting all aspects of school leadership. It seeks to grow and support current and future school leaders so that they can have a positive impact both within and beyond their schools.

Networked learning community (NLC) – A group of schools committed to interdependent working and learning on behalf of all the children they served, within the program. For the duration of the NLC program they became a new unit of educational meaning, a larger organizational form, and broader educational community. They also committed to making their learning and practice more widely

available, not just to the other schools in their NLC, but also to the wider system.

The 20th century was the century during which we built large institutions to do things for people. The 21st century is the century in which we help people to do things with and for one another

(Stephen Heppell).

Some Things We Know

There is a delightful saying in the organizational learning literature, attributed to Peter Senge, in which he notes that there are some organizations that remain steadfastly unable to learn what everyone in the organization knows.

Systems are like that, too. Some things we all seem to know, but the systems we work in are impervious to that collective knowledge. For example, it is widely accepted that the historical model of public-service change – centrally designed and delivered technorational strategies – will no longer work. The historical architecture of delivery, control, and accountability has had its time. We need a new way.

The English education system contains almost 24 000 individual school units, all relatively autonomous (one to another), yet accountable to central expectations through a national curriculum, national improvement strategies, key stage testing, centrally set target-and-inspection regimes. Local market accountabilities, through league tables and parental choice, follow on from these. Together, during the 1990s, these forces combined to foster system as a marketplace and to render schools more competitive than collaborative. Institutional success became a survival requirement, a stronger imperative than collective success.

Across the English-speaking world, the dominant school-improvement models utilize similar characteristics: schools designed on factory production principles; the profession layered and structured; the system tiered – a hierarchy of school, local authority (or school district), state, and national agency. Policy is mandated, practices are prescribed, and outcome targets specified. The logical route to improvement appears to be to strengthen delivery mechanisms and tighten accountabilities through targets, inspection, financial incentives, and consumer choice.

Such top-down-, outside-in-change approaches are seductive, because they appear to work well in the

* "An earlier version of this paper was published by the centre for Strategic Education in Melbourne in its Seminar Series as paper No. 156"

short-term – the system mobilizes itself around the targets, teachers teach to the tests – but then, as we have seen with England’s national strategies, improvements stall. Even more important, although this reform model raises general levels of attainment in the short term, it fails to close the gap in educational achievement between the most and least advantaged.

The paradigm is wrong for the times. Contemporary change needs are too rapid, knowledge is too ubiquitous, contexts of knowledge application are too diverse – and centrally coordinated strategies are unlikely either to be responsive enough, or to be sensitive to the unique challenges of these diverse circumstances. Such strategies do not stimulate or harness practitioner innovation and ownership. Worse, over time, they denude the system of its energy and creativity – its own innovative capacity.

In a sentence, not only are overcentralized reform strategies intrinsically unsustainable, they are also, over time, capacity denuding.

Much We Do Not Yet Know

There is an increasingly widespread view that a more collaborative, adaptive, and long-term problem-solving approach to educational change is the way to go. Sustainable improvement requires a shift to a different model: one which emphasizes capacity building; which spreads and uses leadership widely; which enables and encourages rapid knowledge transfer; which fosters and utilizes practitioner innovation and creativity; and which values system learning and builds for sustainability. The problem is, we just do not yet have the tools and know-how to orchestrate such a pervasive change.

Part of the resistance is that England’s policymakers are heavily influenced by history of reform – particularly the more permissive 1960s and 1970s – which would seem to tell us that random, unstructured, and unconnected innovation does not serve the system well either. Past experience (it says) suggests that unfettered innovation is unlikely to achieve the common purpose and connectivity required to bring coherence and alignment to system-improvement efforts.

Despite these reservations, there is a history in the UK, from different governments, of policies designed to stimulate collaboration, from technical and vocational education initiative (TVEI) more than 20 years ago, through education action zones (EAZs) and excellence in cities (EiCs), to more recent examples. Most have been heavily prescribed from the center, mandated, overly incentivized or resource dependent, and targeted at the most intractable problem areas – inner cities and schools perceived to be failing. They have tended not to be models designed to be either sustained or scaled. At best, the jury is out as to what we learn from these collaborative policy models,

but any sense that there was, as we entered the twenty-first century, a cumulative body of knowledge from them would be fanciful (Demos, 2000). The history of reform in this area has not encouraged the accumulation of a body of learning – what evaluations exist tend to be focused on outcomes rather than learning.

So, on the ground we do not yet have the practice knowledge or the ways of spreading it if we did. At the center, the system is not well served by its own research history from previous initiatives, nor, if we are honest, informed by best available public knowledge, particularly that from other sectors, where collaborative norms are more fully developed. Most tellingly, even if there were clearer understandings about the characteristics of disciplined collaboration, the current architecture of reform gets in the way. You cannot mandate collaborative practices through delivery and accountability levers.

The opposite question: “What is the role of policy and the machinery of state in promoting and supporting emergent, adaptive, locally owned collaborative arrangements?” is the theme addressed in this article.

This was the situation we faced in 2000. Both logic and evidence from practice were indicating that purposeful collaboration between schools could be more conducive to organizational learning than competition. We wanted to know what effective collaboration looked like, and also how it could be incentivized, mobilized and supported – the conditions that would optimize success.

Our belief was that networks of schools engaged in orchestrated and disciplined networked learning offered an alternative way of providing connection and alignment around the multiple concerns facing education. In fact, we argued that networked learning offered a highly effective method of adaptation and integration. We were confident that it would improve outcomes for young people.

These grounded beliefs lay at the heart of what England’s National College for School Leadership (NCSL) set out to achieve in the networked learning communities (NLCs) program (Jackson, 2003, 2005), a system-wide development and enquiry program.

The Networked Learning Communities Program

The NLC program was a system-wide development and research initiative involving 137 networks (1500 schools) that took place between 2002 and 2006. It was specifically designed to inform system learning and national policy, as well as to provide practice evidence about network design and implementation issues, network size and type, facilitation and leadership, formation processes and growth states, brokerage, and system support and incentivization. It sought to generate evidence about how and under what conditions networks can make a contribution to raising

student achievement and attainment, about the leadership practices that prove to hold most potential for collaborative learning and about the new relationships emerging between networks as a unit of engagement and their local authority and community partners.

There were six strands to the basic framework of the NLC design and, unlike traditional delivery approaches, it was designed from the inside out. The six levels of learning were:

- pupil learning (a classroom-learning focus);
- adult learning (professional learning communities);
- leadership learning (shared and distributed leadership opportunities);
- organizational learning (new organizational norms);
- school-to-school learning (networked learning); and
- network-to-network learning (lateral-system learning).

Each network additionally elected to engage other appropriate partners, usually one or more from a higher education institution (HEI) or local authority (LA) or community partner. There were also four further non-negotiables:

- moral purpose – a commitment to success for all children (If you do not care almost as much about the success of children in other schools as you do about those in your own, this would not be the program for you.);
- shared leadership (e.g., co-leadership);
- enquiry-based practice (evidence and data-driven learning); and
- adherence to a model of learning.

This model of learning provided a program-wide discipline and analytical template for what we called networked learning (Figure 1).

The use of this model of learning in support of collaborative joint work practices (networked learning) proved to be a very powerful discipline for networks of schools in the program. Participants across schools agreed a shared

purpose and content focus for their learning – something more aspirational than they could achieve alone.

Networked learning takes place when individuals from different schools in a network come together in groups to engage in rigorous and challenging joint work activity informed by the public knowledge base, using their own know-how and co-constructing knowledge together, as described in the three-fields learning model. In doing so, they are involved in four distinct learning processes:

1. Learning from one another (exchange): where groups capitalize on their individual differences and diversity through sharing their knowledge, experience, expertise, practices, and know-how.
2. Learning with one another (joint work): where individuals learn together, notice that they are learning together, co-construct learning, and make meaning together. Collaborative practitioner enquiry and collaborative learning about recent research are good examples of this activity.
3. Learning on behalf of: where learning between individuals from different schools is undertaken on behalf of other individuals within their school and network – or the wider system.
4. Meta-learning: where individuals are additionally learning about the processes of their own learning and the way that the system learns in ways that build sustainable capacity.

Such collaborative learning in networks has been shown to expand professional identity, from the school as a unit of community to the locality (the wider network) as a unit of educational community. Contextual relevance focuses activity on local change. Changes in teachers' knowledge and understanding and changes in school-level behavior and classroom practice are intermediate outcomes. The ultimate purpose of networked learning is to improve student learning, achievement, and attainment. Wider benefits also include improved confidence and self-esteem among participants, enhanced motivation, a richer sense of professional efficacy, broader professional identity (and liberation from context), and a capacity routinely to engage in wider professional partnerships.

These features begin to answer inevitable questions about cost benefit. There is no doubt that the early phases of collaborative activity are hard and typically not cost effective. The new kinesthetics of collaboration run against the grain of historical practices and we have to learn how to do it well. However, if viewed not as a short-term test of efficacy, but as a necessary phase of the learning process, the value is cumulative.

When practitioners within a network come together, it takes an effort of will and discretionary effort. They need no persuading that the engagement has to be made worthwhile in terms of changed practices and learning gains. Strikingly, the final evaluation of the program found a

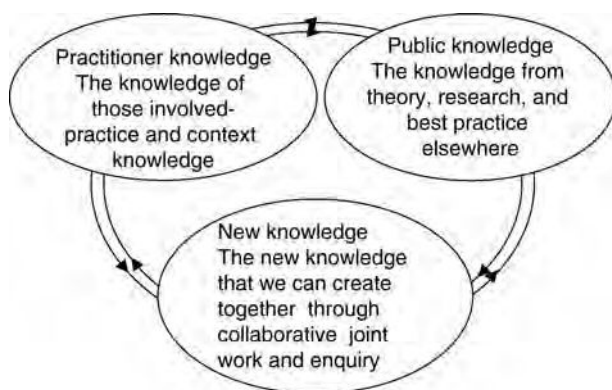


Figure 1 Three fields of knowledge.

direct correlation between levels of involvement and pupil achievement gains – the more staff involved in the network, the more gains for students.

Alternative Approaches to Incentivization and Implementation

Launched as a design framework at a series of regional continuing professional development (CPD) conferences in the autumn of 2001, the groundswell of support for the NLC program took those of us leading the program by surprise. It captured a moment. The profession was clearly weary of a climate of competition, outside-in change programs, normative improvement agendas, and externally generated accountability systems. There was a feeling that we needed to create space in the system for local creativity, lateral learning, and accountability, distributed leadership and the sharing of practice. Having achieved short-term improvements through the externally generated reform agenda, the next phase would take on different and less-predictable shapes. It would occupy different directions of travel, too, being less outside-in than inside-out; less top-down than lateral in direction; and less one (the center) to many than many (the network) to many (other networks).

Informing this, an explicit set of beliefs invited participation (NCSL, 2003). The logic sequence proved persuasive. The first was that normative notions of intelligence are out of date – that all children can be intelligent and powerful learners if schools can learn how to enable it. It is not any longer children that fail in the system so much as the system that fails young people. Sufficient is now known that every child can be a powerful learner. Second, schools, teachers, and leaders who aspire to continuous learning look beyond their own school – recycling existing internal knowledge is a poor foundation for growth. Collaboration (teachers and schools working together rather than competing negatively) is the route to learning and is professionally motivating. Third, the knowledge and the will exist but coherent models, a facilitative policy environment, and a commitment to learning at all levels of the system are required to make it happen. Finally, there is no profession with greater leadership potential than teaching to achieve these things.

The program therefore asked schools to form interdependent networks, to work with and for one another for the benefit of all children, the schools, and the communities they serve. It was formulated as a design intervention, and the design was theoretically and practically informed. The development team utilized both the literature base and practice base, visiting 14 locations around the world to build from best available knowledge. The prototype model (and the challenges of its implementation) was then further co-developed through a sequence of seminars

with practitioners, experts, potential participants, and consultants from both education and beyond.

In the end, what that design meant in practical terms on the ground was to be locally determined – not as an indulgence, but because we believed that they would know best. Networks were supported to co-construct their own responses: their local architecture, the focus of their collaboration, and the enquiry engagement relevant for their context. The aim was to develop involvement, engagement, and spread of leadership, not to prescribe structures.

In support of this, there were some uncharacteristic dimensions to the implementation components of the program. Documentation was minimal and intended to support local efforts rather than to meet central requirements. There was funding for the program, but networks had to match it in kind – showing that they were putting effort and resource into new ways of working and network-dedicated activity. We asked for finance templates (expected network spend), but the only audit was to declare variances – we were interested in how emergent learning shifted spending decisions and why. There was to be no external review for 2 years; instead, networks were expected to formulate their learning into artifacts such that knowledge could be shared with other networks. What review existed was self- and peer review. Networks were encouraged to visit one. Talking about what did not work was valued as highly as what did.

Another feature was that networks could not have one leader – co-leadership was the experimental model for the first year, and one which none of the 137 networks moved away from in subsequent years. These co-leaders operated as system leaders within the network.

Even the process of writing submissions was unusual and emphasized learning values. Regional submission-writing seminars were held during which representatives from potential networks helped one another to design proposals through peer critique and the exchange of ideas. (What is the best idea that you have had today? We would like you to give it away to everyone else in the room.) Formative feedback was given on written proposals – redrafting based on feedback encouraged. The final assessment phase (given that we received 150 proposals – more than 10 times what was expected) involved presentation to peer-review teams, the thinking being that ideas would be shared and disciplines learned from the assessment process.

Throughout the program, the central team operated as a surrogate intermediate system, but with an emphasis on their role as knowledge brokers. The networked learning group was tasked with three core goals:

1. the development of good networks;
2. learning about networked learning; and
3. enabling learning to inform the wider system.

As such, direct support for networks and the facilitation and brokerage of lateral learning and exchange were

significant early tasks. Increasingly, the synthesis of program knowledge and the design of tools, protocols, and publications to support both program and wider system learning were an increasing part of the work.

What Did We Learn?

Three sections from the 3-year external evaluation of the NLC program are illuminating to frame this section. The first, the one that matters most to most people, states that networks made a difference for young people. They worked in raising standards of achievement:

There is a connection between the participation in a network and improvement in pupil attainment. The study provides evidence that when networks of schools work together, there is an impact on pupil learning. The number of people in the school who are active in the network was positively correlated with pupil outcomes. . . . and the level of network attachment was related to change in pupil outcomes. . . . Network attachment was also correlated with intermediate outcomes of changes in thinking and practice in schools

(Earl *et al.*, 2006).

The second explicitly links collaborative and joint work with learning and change:

An important and necessary finding in this study was the emergence of a new factor for both schools and networks that we have called “rigorous and challenging joint work”. It is high order collaboration. It requires participants to suspend judgement, challenge their assumptions and intentionally seek out new information in the quest for ideas and practices that work. Rigorous and challenging joint work may be at the heart of the power of networks. Networks can provide the forum for colleagues to address genuinely new, and often difficult, ideas in a safe environment, away from the risk of censure or even retribution in their daily place of work

(Earl *et al.*, 2006).

The third links learning and leadership – both distributed and formal:

Trust relationships and mutual challenge are the things that make the links in networks; tapping explicit (public) knowledge and exposing tacit (private) knowledge provide the process; and leadership, both formal and distributed, can create the forums and provide the necessary support and capacity building opportunities to move the process forward

(Earl *et al.*, 2006).

If this is an overview of external evaluation findings, our work at NCSL also helped us to pick out a more finely grained set of learning points along the way. The first is a

set of verities. As stated earlier, the NLC program drew from best available knowledge worldwide. In accumulating that body of evidence, certain themes recurred which were subsequently further validated through the program’s work.

These verities are, chiefly, that collaboration relies on voluntarism and having a compelling reason to work together – one that accords with the strong moral purpose of the teaching profession. Networks also require good internal leadership. As for all organizational forms, leadership is critical, but it manifests itself differently when influence and facilitative support (rather than positional authority and hierarchical status) are key characteristics. In addition, networks benefit from external brokerage and critical friendship. Given these conditions, networks expand access to good ideas and enable greater sharing of professional knowledge.

Our ongoing process of learning and research throughout the program helped us to understand the challenges, the dynamics and the rewards of collaboration more deeply. Research (internal and external to the program), evaluation activities, network reviews, program-wide enquiries, sustained facilitator engagement, and the involvement of a number of respected international associates together yielded a huge body of evidence. This in turn has allowed us to expand the boundaries of what is known about successful collaboration.

A full account exists elsewhere (Jackson, 2005). In summary, we learned that new ways of working together emerge with difficulty and at high early transaction cost; that trust is an important issue for networks, but it is more an outcome of rigorous joint work than it is a precondition; that some of the motivational energy for participants stems from the acquisition of a sense of expanded professional identity; and that networked learning surfaces the reciprocity and generosity which is at the heart of a professionally collaborative culture.

Moreover, and importantly, networks spread leadership influence. They act both to distribute leadership and also liberate new types of leaders. These leaders understand their connection to the wider profession – learning on behalf of (mentioned above) was a major motivator for knowledge exchange and joint work activities. As such transfer of practice is far too naive a concept to describe the knowledge-exchange processes in networks. Collective problem solving, joint work projects, collaborative enquiry – in short, a range of innovation and knowledge-creation activities – better describe the learning processes. Perhaps unsurprisingly, this expanded leadership capacity accentuated rather than diminished the significance of formal leaders (school headteachers), whose capacity to enable and empower the participation of others, to gatekeep network outcomes within their own school – and to model commitment to networked learning – was directly correlated to success. The eternal

evaluation finding that gains for students in schools positively correlated with formal leader engagement with the network was no surprise.

Where successful, NLCs created new units of meaning, locality provision for young people – and new units of engagement for local authorities and other external partners.

Beyond this, there are some more tentative propositions that emerged from the NLC work and which bear the warranty of experience more than research.

Perhaps most importantly, we came across the well-established problem of causality – how can networks prove that they are making a difference greater than the sum of their parts and, when they do, how can they take the credit for it? We were able to show that pupils in the 1500 NLC schools made statistically significant attainment gains greater than those in schools not in NLCs, but proving causality is another matter!

We knew that voluntarism was vital, but we now believe it can be orchestrated with the help of intermediaries. Since the start of the program, some local authorities have implemented learning networks across all their schools, co-designing local solutions with groups of schools. Good system-brokerage skills are a vital part of this work and local authorities as system designers might be one of the differentiating skill sets for the future.

Successfully orchestrating participation also relies on a structural architecture built around shared challenges and goals for learning, and in which people are prepared to experiment with new communication channels that break the traditional, hierarchical mould. Logic tells us that ICT should play a vital role in that process, and yet there are very few examples of it being integral to the success of networks. As that suggests, networked learning is most likely to work when participants unlearn some of their old practices and embrace new ways of working. Those in positional roles enshrining institutional power have the most challenging unlearning agendas.

Finally, variability of commitment is another inherent and perhaps inevitable challenge for learning networks. It is, of course, for all organizations, but it is more visible in networks. Accommodating those variances while still retaining collective buy-in and commitment to core purposes is a key leadership skill for network leaders.

What Did We Learn about Implementing and Facilitating Networks?

The team that supported the NLC program – NCSL's networked learning group – was itself trying to explore what role intermediary bodies could play in catalyzing and supporting networks on the ground. Its performance was a potential model for organizations like local authorities in developing collaborative work.

There is no doubt that we did some things well. The program was both theoretically and practically informed (McCormick, 2004). We succeeded in generating a compelling design, underpinned by values, which motivated engagement. The co-design orientation, and subsequent co-development work, stimulated new relationships between system partners. We did not have the answers, so these were learning relationships. The model of learning was a seriously important shared discipline. The development team brokered lateral connections and stimulated peer-to-peer learning.

If improved outcomes for children are the critical determinant of success, then the external evaluation shows that this was achieved. If creating a body of knowledge is the yardstick, then it is fair to say that, as the program progressed, we wrestled – successfully on the whole – with the challenges of real-time learning now manifested in our tools and publications. The legacy of products and materials is quite probably of national and international significance.

We never, of course, pretended (or wanted to pretend) that we had the answers, or that everything was successful. This was a program committed to learning through doing, and we were open about our uncertainties and failures with the networks, just as we asked them to be open with one another. With the clarity of hindsight, we could have done a number of things more effectively.

In retrospect, the NLCs design came to be too close to a brand. In proposing NLC as an ideal prototype to support development, we unwittingly gave the impression that we were proposing this model as the single, or best way forward. Again, also with the wisdom of hindsight, our central infrastructure grew too large. Networked systems can be supported with a minimum of central capacity by utilizing leadership from the front line – and in so doing grounding decisions and thinking in the experience of practice. Our central infrastructure did not assist us to sustain the co-development orientation in the eyes of networks, either.

Two early decisions proved to be only partially wise. First, the networks received nominal funding for the first 3 of the 4 years. That may have been a mistake. We wanted to stimulate new capacities and ways of working and knew that some funding would help. So it did, but a proportion of networks saw the removal of funding as a barrier to sustainability. No funding could have been the way to go.

Second, we knew that baselining at the outset would be crucial to evaluating achievements. We wanted each network to take ownership of its own baselining strategy and we developed tools to scaffold that work. Looking back, that work was pretty variable, not least because the networks were at such a formative stage when asked to do it.

The last reflection links into the concluding points below. The network learning group was itself a part of a national agency, NCSL. In common with many innovators, we felt under siege early on, felt that we were challenging historical practices and were in turn being

challenged by those who were skeptical or threatened by the program. We could have integrated our work into the college more effectively – and as a result had less influence there than might have been possible.

That is an honest appraisal. The final section makes some equally honest concluding observations about barriers to taking collaborative practice to scale across entire public-service systems.

Conclusion

Systems can seem dumb. It is difficult for them to learn what is widely known, not least because the system's learning structures are themselves a part of the problem.

Enough is known. It is not primarily lack of knowledge that prevents collaborative working – either between schools or across services. The NLC program alone has created a huge store of evidence and learning and there are other rich sources of knowledge, both in this country and abroad. In a system configured for one-to-many rather than many-to-many dissemination, though, there are certainly problems making what knowledge we have widely available sufficiently quickly, at scale, to support emergent practices.

In addition, the issue of scale is also a part of the problem. In her internationally regarded study on scale Coburn (2003) identifies four dimensions in which we lack a resolute theory: depth, sustainability, spread, and shift in reform ownership. The third of these poses interesting challenges for the support of a collaborative system. About spread she says:

Rather than thinking of spread solely in terms of expanding outward to more and more schools and classrooms, this emphasis highlights the potential to spread reform-related norms within the district. At the district level, spread involves the ways in which reform norms and principles influence district policies, procedures, and professional development.

This is clearly a conundrum. We know from our own work that a network-based system requires brokerage and support. There is a crucial facilitation role for local authorities and a critical role in generating facilitative policy at a national level. To make the shift, both local and national agencies will have to internalize the challenging norms and principles of a more networked and collaborative system. But local and national government often appear irretrievably hierarchical and bureaucratic.

Returning for a moment to the epigraph at the start of this piece, Stephen Heppell referred to the organizations set up in the twentieth century to do things for people. As we seek to liberate collaborative arrangements, to help people to do things with and for one another, what could well stand in the way are those same organizations, hierarchically

structured, culturally siloed, relatively impermeable to learning, set up to provide answers, and to believe that they know best.

In addition to the external evaluation, as a part of the wider research program, we also commissioned a study on the role of local authorities in supporting school networks. Aubyn Howard in the report from this piece of work writes:

Networked learning is part of a system-wide shift towards a more collaborative and interconnected way of working – the system local authorities support is becoming more networked. Alongside this, LAs are also being asked to network internally to deliver Every Child Matters. However, many don't know either how to do this or how to reconfigure to support school-to-school networks – why would they? Moving into networked learning mode themselves (within services, between services or agencies and between LAs) such that they model practices and apply them in new forms of engagement with networks of schools, is a good way to start. This means embracing collaboration, networking and learning within the way they work. (They) need to engage in a deeper internal process of organisational and cultural change if their efforts to model networking and work differently in support of ECM are to be sustained

(Howard, 2006).

If this applies to local authorities, it is equally applicable to national agencies, to government, and to the structures of the civil service. The critical orientation for change is a learning orientation. This means engagement with front-line collaborative practices, not to provide answers, not to hold to account, or to bureaucratize, but to learn how better to accommodate risk, and to provide enablement and support. It is a role of system broker and shaper, and it is a very different orientation.

Bibliography

- Coburn, C. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher* 32(6), 3–12.
- Demos (2002). Learning the Lessons: How past policy initiatives can help practitioners in the networked learning communities programme. www.ncsl.org.uk/nlc (accessed July 2009).
- Earl, L., Katz, S., Elgie, S., Ben Jaafar, S., and Foster, L. (2006). How networked learning communities work. *Final Report of the three year External Evaluation of the Networked Learning Communities Programme*. <http://www.ncsl.org.uk/nlc> (accessed July 2009).
- Howard, A. (2006). The 'networked learning' change agenda for local authorities. <http://www.ncsl.org.uk/nlc> (accessed July 2009).
- Jackson, D. (2003). Networked Learning Communities: The role of knowledge-based networks in school and system improvement. *Paper Presented at the CERIOECD Forum Knowledge Management in Education and Learning*, March 2002. <http://www.ncsl.org.uk/nlc> (accessed July 2009).
- Jackson, D. (2005). Learning themes from the networked learning communities programme. *Paper Presented at the International*

Congress for School Effectiveness and Improvement Conference, Barcelona, 2nd–5th January 2005. <http://www.ncsl.org.uk/nlc> (accessed July 2009).

McCormick, R. (2002). *Theoretical Perspectives of Relevance to Networked Learning Communities*. <http://www.ncsl.org.uk/nlc> (accessed July 2009).

NCSL (2003). *Like No Other Initiative*. <http://www.ncsl.org.uk/nlc> (accessed July 2009).

Relevant Websites

<http://www.innovation-unit.co.uk> – Innovation Unit.

<http://www.ncsl.org.uk> – National College for School Leadership.

<http://www.ncsl.org.uk/nlc> – National College for School Leadership, Networked Learning Community (Getting started with network study visits, System leadership in action, Networked learning knowledge store).

Networks and Communities of Knowledge

T Schuller and H Theisens, Centre for Educational Research and Innovation, OECD, Paris, France

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The production and transmission of knowledge has always involved networks and communities. The band of young men surrounding Socrates, the academia of Plato, the medieval monasteries, the cathedral schools, the guilds, universities, and scientific disciplines are all examples of communities and networks of knowledge. However, while knowledge has always played an important role in human societies, the term knowledge society refers to the fact that knowledge has now moved to the economic and social core of our societies and that the speed of knowledge creation, accumulation, depreciation, and dissemination has accelerated (OECD, 2004). The communities and networks that produced and transmitted knowledge in past tended to be rather closed, homogeneous, and were either small (Socrates) or institutionalized (universities). With knowledge now central to society and economy (and both readily available and necessary for many people), the communities and networks that produce and transmit knowledge have become more heterogeneous and less institutionalized. The overall context, therefore, is one where it becomes increasingly important to understand the formal and informal mechanisms through which knowledge is created, transmitted, and validated.

Knowledge Production and Its Context

The rise of the knowledge society is the result of a number of different but interrelated trends. The whole process of knowledge production and use has strongly accelerated due to the steep decline in the cost of replicating and transferring information (OECD, 2004). Due to the ongoing mechanization of manual labor and the growth of the service sector, knowledge has become increasingly important as a production factor. These developments have been made possible and have also been driven by an increasingly educated workforce. Lately, the effects have been fortified by globalization: for many developed countries, knowledge-intensive production (and related innovations) is one of the major competitive advantages left, compared to developing countries with lower production costs. However, the balance of advantage here is changing fast as large countries such as China and India catch up.

There are several further features of this general trend:

- the growth of the service sector in all OECD economies and the weight of highly educated knowledge workers in this sector;
- the massification of higher education and the fast growth of academic research since World War II;
- the growing importance of knowledge-based innovation as an essential element for economic success (and the increasing spending on research and development (R&D) as a result), in the context of global competition;
- loss of a monopoly position of universities and academies when it comes to knowledge development and the transmission of high-level knowledge; this is now spread over a multitude of organizations (e.g., companies, think tanks, nongovernmental organizations (NGOs)); and
- the development of a knowledge infrastructure connecting these different knowledge organizations, especially via the Internet acting as a fast-growing repository of codified knowledge, to which search engines provide easy access.

As a result of these developments, different forms of knowledge production are coming into existence. Increasingly, knowledge workers are generating and transmitting knowledge outside the formal knowledge institutions like universities. Gibbons *et al.* (1994) describe these phenomena as new modes of knowledge production (they refer to mode 2 as opposite to mode 1). This new mode-2 knowledge production is marked by:

- an increase in the number of potential sites in which knowledge can be created (not only universities, but independent research centers, government agencies, laboratories, think tanks, consultancies, and in the interaction of all of these);
- the linking of sites together in a variety of ways – electronically, organizationally, socially, informally – through the functioning networks of communication;
- the simultaneous differentiation at the linking sites, of fields and areas of study into finer and finer specialties, but at the same time greater interdisciplinarity; and
- a strong focus on problem-solving as a central goal of knowledge production.

While mode-2 knowledge production refers mainly to new forms of knowledge production (in the terminology of mode 1: research), learning in this new situation is transformed too. With the advent of the knowledge

economy, learning becomes an integral part of economic processes. Communities and networks are important collective mechanisms through which such learning takes place.

New Knowledge Relationships

The combined impact of the importance of knowledge, the speed of change, and the fragmentation of the sites for knowledge production poses challenges for the production, accumulation, and transmission of knowledge. With knowledge workers dispersed through a great number of different organizations, with very different traditions and objectives, there is the question of whether the traditional mechanisms for knowledge reproduction are still operating. Recent literature has abundant descriptions of different responses to this challenge, and a plethora of related concepts.

Organizations have always been learning in the sense that they have always needed to respond to unforeseen changes. When new technologies were incorporated in the production process, unforeseen problems needed to be solved by developers, producers, and users; in other words, the organization learned by doing and using. However, this kind of learning often remains localized; its results do not spread beyond the boundaries of the firm and are only relevant and understandable in the context in which the learning took place. One reason for this is that while the cost of replicating information has been reduced to almost nothing, thanks to modern technology, reproducing knowledge remains a far more expensive process. Many cognitive capabilities are not easy to articulate explicitly or to transfer to others, as tacit knowledge (Polanyi, 1976). This is why knowledge reproduction has traditionally relied on some version of the master-apprentice system, or on interpersonal transaction among members of the same profession or community of practice. These means of reproducing knowledge were at the heart of many professions and traditions, but they fail to operate when social ties unravel, when contact is broken between older and younger generations, and when professional communities lose their capacity to act in stabilizing, preserving, and transmitting knowledge (OECD, 2004). In this situation, new forms of knowledge production, with results that allow for generalization (beyond the boundaries of closed and homogenous communities) may be needed.

These new patterns of knowledge production and learning are developing already; Lundvall (2006), Piore and Sabel (1984), and Granovetter (2000), for example, point to networks of small- and medium-sized firms that allow users and producers to interact and share what they have learned from producing and using each others' products. In several fields, from high-performance sports equipment to computer software, users are joining in the

design process, pooling their knowledge and experience in order to enable producers to meet their needs more closely (Hippel, 2006). These are the kinds of changing relationship which will characterize at least some future networks. In interacting and sharing their learning outcomes, beyond the boundaries of their own firms, producers and consumers need to formulate their knowledge in ways that makes it understandable outside its original context. This generalization of the learning outcomes then creates the possibility of sharing learning outcomes with many actors in a network.

Most of the new forms of knowledge production and learning identified in the literature share the use of the concepts of networks and communities. Networks can be loosely defined as a set of actors (persons or organizations) linked by ties or relations of a specified type. A tie or relation between actors has both strength and content. The content might include information, advice, or friendship, shared interests, or memberships and typically some level of trust. In this article, we are concerned with professional networks and particularly with the flow of information through the networks. The level of trust is important though because it lowers the transaction costs within the networks and facilitates the sharing of information (see section entitled 'Social capital and communities of knowledge').

The concepts of community and network are often used interchangeably in the literature. The distinction is a fluid one. Communities tend to be smaller, more homogenous, and with stronger ties in which people share a common set of values. Networks are larger, less homogenous, have weaker ties, and more plurality of values. This provides us already with a number of dimensions on which communities and networks can be classified, that is, size, strength of tie, and degree of homogeneity. In addition, we can examine the extent to which the community is defined by occupational or professional identity; the extent to which it is bureaucratically organized; its geographical dispersion; the degree to which technology defines the mode of communication within it; and the degree of its disciplinary boundedness or spread.

Other aspects of a slightly different kind refer to the values and orientations displayed: for example, whether the network has primarily a commercial orientation; its legal or official status; and the extent to which members of the community define their identities with reference to it.

There are certainly other dimensions that could be added. Network analysis is a methodology that is beginning to assume a new lease of life, as novel techniques allow a more sophisticated empirical mapping of their scale and nature. The essential point is that networks of different kinds will play a significant part in defining the extent to which knowledge is generated and transmitted.

Social Capital and Communities of Knowledge

One systematic way of looking at communities of knowledge is by using the social capital theory. At a societal level, this is primarily associated with the work of Robert Putnam (1993, 2000), but other literature in this area deals very closely with networks, notably that of Burt (1982) on structural holes. The relationship between social capital and education has been explored in social theory by Coleman (1998) and Bourdieu (1986), and in an international policy context by OECD, exploring the interaction between human and social capital (OECD, 2001). Social capital as a concept is often defined specifically in terms of networks, stressing the norm-laden nature of relationships within and between them. A common differentiation of types of social capital is into three basic forms (Woolcock, 1998):

- bonding social capital, which refers to relations within or between relatively homogenous groups;
- bridging social capital, which refers to relationships within or between relatively homogenous groups; and
- linking social capital, which refers to relationships between people or groups at different hierarchical levels.

These are not mutually exclusive, that is, a community can be high on both bonding and bridging forms. Analyses that explore the interaction between these different types of social capital can be very fruitful for determining the dynamics of knowledge creation and use (Schuller, 2006). Thus, networks that are strong on bonding but weak on bridging may be powerful creators of knowledge within quite well-defined frameworks, such as an established research group with high levels of commonality in the members' approach to research and understanding of the field. However, a community that also includes bridging social capital may be more open to an infusion of new ideas and to the intellectual challenges that come from exposure to different approaches, and benefit from this in the longer term. As for the application of knowledge, this may depend in part on how far linkages to higher levels of power exist; but often network structures may allow such hierarchies to be side-stepped. In short, there is no single ideal combination of these different forms of social capital in the construction of communities of knowledge.

One crucial function of social capital is its deployment of trust. Some level of trust is implicit in almost any concept of community (though maybe less so for networks, where the interdependence is weaker). The importance of trust is in its capacity to help members of the community of knowledge to assess and validate knowledge as it evolves: in other words, to scrutinize knowledge claims, for example, from other researchers, or from practitioners or policy-makers, and by drawing on the community's resources to

enable members to decide on the quality or level of knowledge. This function of critical validation is especially important in the face of the tidal waves of information and misinformation that slosh around the electronic world. The validation is not only a matter of truth determination, but also of selection, utility, and application. A community of knowledge draws on its members' collective experience and intelligence to decide which of the huge range of possible choices of knowledge is likely to be most favorable. It reduces the cost of information, both in searching and verification and raises the levels of reliability. These are valuable both financially and psychologically. Trust also promotes the sharing of ideas, because it diminishes fear of intellectual theft. Intellectual property is an increasingly recognized form of asset in knowledge societies, and this brings with it a growing complexity in how such an asset is to be managed and this includes issues related to ownership, distribution maintenance, and exploitation. One of the strengths of the open source movement is its ability to promote the accumulation of knowledge by providing a normative and legal framework within which people can share ideas, in the expectation that such sharing will result in improvements to the asset without it being appropriated for individual gain at the expense of a common good.

High Levels of Bonding: Communities of Practice

Communities of knowledge with high level of bonding social capital develop a common language, common approaches to problem solving, etc., which reduce the transaction costs and greatly stimulate the exchange and creation of knowledge. Foray (OECD, 2004) argues that there are three elements that constitute the basics of such a community. First, a significant number of a community's members combine to produce and reproduce new knowledge. Second, the community creates a public space for exchanging and circulating the knowledge. Third, new information and communication technologies are intensively used to codify and transmit new knowledge. He goes on to argue that communities characterized by these three components tend to be fundamentally geared to knowledge-driven production. The advantages of this are:

- the communities facilitate interaction, which increases the possibilities for learning and development of knowledge;
- in order to be able to exchange ideas, knowledge must be codified, which leads to greater storage and communication capacity and makes it possible to develop new cognitive approaches;
- quality control is assured because each member can reproduce, test, and criticize new knowledge;

- everyone has access to the knowledge produced and, therefore, the same items will not end up being reinvented; and
- learning productivity is made greater by the fact that an individual can learn to learn by reproducing the knowledge of others.

An example of communities with high levels of bonding is what Lave and Wenger (1991) have termed communities of practice. Knowledge production is about collective practices that may or may not be formalized. It involves learning that takes place within certain contexts. The links between knowledge production and learning is apparent on the surface, but not always recognized. A community of practice defines itself along three dimensions (Wenger, 1998):

- its content, that is, what it is about as a joint enterprise continually renegotiated by its members;
- its mode of functioning, that is, the types of engagement that bind members together into a social entity; and
- its capability, that is, the shared repertoire of routines, vocabulary styles which members develop over time.

The overlap with some of the dimensions already identified is evident, but the emphasis is different. Reference to communities of practice shifts attention to the ways in which people work together, and it places the accent on the close connection between knowledge and activity. In other words, the community is formed and continually reformed not only through the exchange of abstract knowledge, but also through the activities which constitute the ways in which members of the community relate to each other. Lave and Wenger's interest is in how educators form such communities; it therefore extends the range to include relationships between teachers and learners, although without placing them in two separate categories.

High Levels of Bridging: Open Source Communities

Networks of knowledge rich in bridging social capital are important, in that they bring together ideas, values, and approaches from potentially very different groups and individuals. The combination and recombination of these ideas is a great potential source of innovation. Granovetter (1973) referred to this as "the strength of weak ties." His illustration is the fact that people find jobs not through sources strongly tied to them (i.e., in their community) but through more peripheral contacts. This is so because it is those individuals further removed who have access to very different communities and can form a unique bridge to these other communities. While

networks may be less dense and ties less strong, this also makes for a greater inclusiveness, increasing the likelihood that ideas developed in very different communities are shared.

More recent technological developments have enabled the growth of new forms of communities of knowledge with potentially high bridging social capital. One outstanding example is the development of the open source movement and, following on that, of the emergence of open educational resources. The open source is a community-based approach to software developing based on voluntary participation and voluntary selection of tasks. Anyone can join an open source project, and anyone can leave at any time. The collaborative development process of open source software, strongly simplified, starts with a software developer trying to solve a problem he or she is facing. The developer needs a particular application that does not exist and decides to write it and invites others, who might have the same problem, to collaborate. The collaborators propose new features, modifications of the code or bug fixes that might or might not be accepted by the original developer. Small projects might have simple decision-making processes; larger ones develop hierarchies of gatekeepers, developed on an implicit principle of meritocracy.

Open educational resources (OERs) are digitalized learning materials placed on the Internet with free access. OERs are a much more recent phenomenon than the open source movement that inspired these. It is, therefore, uncertain whether its further growth will take a quite different path. However, in any case, OERs are beginning to have an impact on education. More and more educational institutions, including elite ones, are moving to create OERs, for reasons of prestige, position, or social responsibility. Developments are also taking place from the bottom up, as individuals from very different backgrounds join together over the net to pool their knowledge and experience in the form of learning materials (OECD, 2007) on a permanently evolving basis. The spread of OERs is potentially significant in terms of opening up access to educational materials worldwide; it also poses a number of issues in relation to sustainability, ownership, and intellectual property.

One of the most significant features of the OER trend is the way in which its potential to be open and free depends nevertheless on the existence of a clearly defined framework within which those terms are to be understood. Access to the materials does not bring with it access to teaching or to certification: the fact that the materials from all courses taught at MIT (and now many other universities) are available freely on the web does not mean that anyone can sign up for a course there. Nor does the fact that OER materials can be remixed, modified, and then retransmitted mean that anyone can

appropriate them. Membership of the OER community can take a number of forms, some quite ideologically connected to collective improvement, and some geared to organizational change or other motives. It is a good illustration of how loose and diverse networks can be.

System Level Networks: The Triple-Helix and Brokerage Institutions

On a system level, high bonding social capital is less probable because of size-related issues (and, therefore, to the social and spatial distance between individuals) and the diversity within many systems. Increasingly though, as a consequence of several developments, like the shift in knowledge production from mode 1 to mode 2, bridging social capital becomes an important aspect of successful research and innovation systems.

Etzkowitz and Leydesdorf (2000) introduced the concept of the triple-helix to refer to the new relationships between university, industry, and government. In their analysis, there is an increasing collaboration and overlap of mission between these three formerly separate spheres. This has many different aspects. For example, within universities there is increasing stress on collaboration with industry; this has resulted in developments like intermediary offices, spin-off firms, and science parks. At the same time, science-based companies can no longer afford to be isolated organizations. In an increasingly global and competitive environment, firms need to access sources of knowledge and technology outside their own organizational boundaries. Companies are increasingly working together with universities and even other firms and government laboratories as a potential source of useful knowledge and technology. These kinds of cooperative developments are often encouraged by government both at the regional and national levels. Of course, there are also conflicts of interest and issues about intellectual property and its relation to basic research.

All of this means that the linear model of innovation, a one-way flow from fundamental to applied research and to product development is replaced by new ideas and alternative models based upon interdisciplinarity and spiral feedback links between technology and science. The essence of the triple-helix model is that the interconnection between industry, government, and science has resulted in a layer of tri-lateral networks and hybrid organization that not only in itself is in permanent transition, but it also permanently challenges the three founding systems of this layer.

In the field of education, this process can be observed in the increasing interest in strengthening links between educational researchers, policymakers, and practitioners (OECD, 2007a). This reflects in part a concern in many

countries over the weakness of education as a knowledge-based service. Ironically, education, in spite of being the basic producer of most knowledge-generators, does not seem to draw systematically and extensively on a knowledge base.

One approach to tackling this has been to set up various agencies or initiatives designed to improve these linkages, sometimes referred to as brokerage agencies. There are practical and specific goals, such as to strengthen the quality of the research on which policy is based. The aim is often also to expand the capacity of the research community, making more explicit (less tacit) the kinds of rules which govern, or ought to govern, research and bring in a wider range of stakeholders. Arguably, this trend exemplifies a move to shift the education research community from bonding to bridging. The point here is that, there is a conscious attempt to extend and diversify the member of communities whose function it is to expand and strengthen the knowledge base of education as a service.

Conclusions

Important work needs to be done in making some of the research findings on networks and communities of knowledge relevant to educational policymakers and leaders. This would entail moving from describing and analyzing already existing networks and communities to thinking about how communities and networks can be created, successfully sustained, and steered in the right direction. Two major issues stand out here.

A first main issue is the relationship between traditional policy instruments, such as regulation, funding regimes, and inspection, on the one hand, and communities and networks on the other. Leadbeater (2005: 22) warns that the collaboration needed for effective networks “can be held back by regulation, inspection and funding regimes that encourage schools to think of themselves as autonomous, stand alone units.” Levin (NCSL, 2005b: 6) agrees with this point of view, pointing out that there “are inevitable tensions between the idea of learning networks, which are based on ideas of capacity building as a key to reform, and . . . reform through central policy mandate.”

A second important question is: what does leadership mean in the context of horizontal and fluid networks? Mulford (2007) points out the importance of leadership for the creation of both bridging and bonding social capital. Leadership in this context involves activities such as ensuring the participation of the right kind of people (courting) in networks and ensuring that, once trust is established, parameters are set for collaboration, resources are secured, and, at all times, a critical mass of enthusiasts participate fully in the network (aligning). In other words, leadership, of a very particular nature, is of key importance for effective networks and communities of learning.

However, the relation between formal leaders and communities and networks of knowledge is not problem-free. Traditionally, leaders derive their authority from their position in formalized hierarchies. Authors like Edith Rusch (2005) point out the inherent conflicts between fluid networks and the formal system. Structures are seen as malleable in networks but fixed and hierarchical in the system. Conflict is open and valued in networks, whereas it tends to be hidden and feared in the system. Communication is open and unbounded in networks, but controlled and closed in the system. Leadership tends to be fluid in networks, whereas it is hierarchical and assigned in the system. At least some of these tensions need to be resolved for leaders to be effective in networks and communities of knowledge.

It would be easy to oversimplify the place and functions of networks and communities of knowledge, just as it is easy to indulge in simplistic rhetoric about knowledge societies. Key needs are for further refinement of the typologies of the mechanism for the development and exchange of knowledge; further empirical application of models to explore the trajectories – including decline as well as growth – of different communities; and the impact on the quality, distribution, and utilization of knowledge in different social and occupational contexts.

Bibliography

- Bourdieu, P. (1986). The forms of capital. In Richardson, J. (ed.) *Handbook of Theory and Research for the Sociology of Education*, pp 241–248. New York: Greenwood.
- Burt, R. S. (1982). *Towards a Structural Theory of Action*. New York: Academic Press.
- Coleman, J. S. (1997). Social capital in the creation of human capital. In Halsey, E. H., Lauder, H., Brown, P., and Wells, A. S. (eds.) *Education, Culture, Economy and Society*, pp 80–95. Oxford: Oxford University Press.
- Etzkowitz, H. and Leydesdorf, L. (2000). The dynamics of innovation: from national systems and “Mode 2” to a triple helix of university–industry–government relations. *Research Policy* 29(2), 109–123.
- Gibbons, M., Limoges, C., Nowotny, H., et al. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: Sage.
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380.
- Granovetter, M., Castilla, E. J., Hwang, H., and Granovetter, E. (2000). Social networks in Silicon Valley. In Chong-Moon, L., et al. (eds.) *The Silicon Valley Edge*. Stanford: Stanford University Press.
- Hippel, E. von (2006). Democratizing innovation: The evolving phenomenon of user innovation. In Kahn, B. and Foray, D. (eds.) *Advancing Knowledge and the Knowledge Economy*, Cambridge, MA: MIT Press.
- Lave, J. and Wenger, E. (1991). *Situated Learning. Legitimate Peripheral Participation*. Cambridge: University of Cambridge Press.
- Leadbeater, C. (2005). The shape of things to come. London: DfES Innovation Unit. <http://www.standards.dfes.gov.uk/innovation-unit>. (accessed May 2009).
- Lundvall, B. (2006). Interactive learning, social capital and, economic performance. In Kahn, B. and Foray, D. (eds.) *Advancing Knowledge and the Knowledge Economy*, Cambridge, MA: MIT Press.
- Mulford, B. (2007). Building social capital in professional learning communities: Importance, challenges and a way forward. In Stoll, L. and Louis, K. S. (eds.) *Professional Learning Communities: Divergence, Depth and Dilemmas*, pp 166–180. London: Open University Press and McGraw Hill.
- NCSL (2005b). *Establishing a Network of Schools*. Cranfield: Network Learning Group, National College for School Leadership.
- OECD (2001). *The Well-Being of Nations*. Paris: Organisation for Economic Cooperation and Development.
- OECD (2004). *Innovation in the Knowledge Economy, Implications for Education and Learning*. Paris: Organisation for Economic Cooperation and Development.
- OECD (2007). *Giving Knowledge for Free: The Emergence of Open Educational Resources*. Paris: Organisation for Economic Cooperation and Development.
- OECD (2007a). *Evidence in Education: Linking Research and Policy*. Paris: Organisation for Economic Cooperation and Development.
- Piore, M. J and Sabel, C. (1984). *The Second Industrial Divide*. New York: Basic Books.
- Polanyi, M. (1967). *The Tacit Dimension*. New York: Anchor Books/Double Day and Company/Garden City.
- Putnam, R. D. (1993). *Making Democracy Work. Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.
- Putnam, R. D. (2000). *Bowling Alone. The Collapse and Revival of American Community*. New York: Simon and Schuster.
- Rusch, E. (2005). Institutional barriers to organisational learning in school systems: The power of silence. *Educational Administration Quarterly* 41(1), 83–120.
- Schuller, T. (2006). Social capital, networks, and communities of knowledge. In Kahn, B. and Foray, D. (eds.) *Advancing Knowledge and the Knowledge Economy*. Cambridge, MA: MIT Press.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge: Cambridge University Press.
- Woolcock, M. (1998). Social capital and economic development: Towards a theoretical synthesis and policy framework. *Theory and Society* 27, 151–208.

School Community Relations

R L Crowson and E B Goldring, Vanderbilt University, Nashville, TN, USA

© 2010 Elsevier Ltd. All rights reserved.

A movement to establish clear standards for school performance and use standardized tests of student achievement to assess that performance is at the heart of recent educational reform in the US. The central objective of these reforms is to improve the quality of schooling, particularly to improve the educational attainments of those students who have tended to be left behind by the schools and the larger society.

The standards movement in the US is regarded as a top-down initiative with much federal government involvement. This is rather unusual within a system of schooling long celebrated for local control. Despite the increased federal role in education, there has also been a resurgence of localism in the early twenty-first century. Particularly, there has been a resurgence of interest in the bottom-most unit in the educational hierarchy: the individual school and its surrounding community or neighborhood. Generally known as the new localism in education, this restored emphasis on the local school site, and its community relationship is by no means breaking away from the emphasis upon national standards and their attainment. A primary focus upon student learning and upon measurable evidence of school effectiveness in instruction remains a central objective. Explorations of how schools and communities can be linked and can interact successfully around both student learning and school improvement provide the driving questions in studies of the community relationship in the new localism paradigm.

As an approach to understanding and using the school–community relationship to improve learning, the new localism covers much territory. Four broad categories are central:

- a sense of place and a new pedagogy of place in education;
- the importance of community capacity and social capital in learning;
- communities and families as key instructional resources; and
- a new local politics of partnership and empowerment in education.

This article addresses each in turn and considers some actions for leaders in education.

A Sense of Place and a Pedagogy of Place

In the opinion of many scholars, the new localism starts with a concept of place (Driscoll, 2001; Smith, 2002;

Gruenewald, 2003). Interestingly, a longtime criticism of localism was that the schools and their clientele can all too easily be seriously constrained by the cultural myopia, the prejudices and narrowness, and the restricted range of norms and values that characterize their closely surrounding communities. From this perspective, it was a central mission of the educator to broaden horizons, introduce children to competing values and perspectives, and open minds to a much wider world than they tend to find in their own neighborhoods. A sense of place from this perspective was a limiting, not an enlarging, concept.

Perspectives have changed, and a rediscovery of the importance of place is emerging. Researchers and educators are recognizing that while the schools do need to widen horizons for their students, the schools are indeed embedded in their communities, and much that defines or characterizes the community is embedded in the schools. Family values, local dialects and language patterns, area lifestyles, neighbor-to-neighbor linkages, day-by-day patterns of activity, physical surroundings, the nature of the neighborhood economy, and neighborhood safety/security – all matter, and matter in ways that influence the work of the schools.

In an examination of the concept of place, Driscoll (2001) identified four central dimensions: (1) Place has a territorial dimension, a topography that is characterized by particular features and which can often be fixed in our long-term memories. (2) Places have boundaries, which provide roots and help us to mark the known from the strange and unfamiliar. (3) Places are endowed with social and cultural meanings, a socially constructed source of cultural socialization and behavioral learning for us. (4) Places have a temporal quality – helping us to organize the past, present, and future, and often help share experiences across generations.

What are the key implications of a sense of place for the schools? Again, Driscoll (2001) suggested a set of four implications: (1) Schools as spaces matter, signifying their own place in the community as centers of children's development and learning. (2) School boundaries matter, respecting the neighborhood as an integrative whole without establishing a fortress mentality that keeps other influences out. (3) History and culture matter, wherein the (often unique) cultural narrative of a community is also embraced by the school. (4) Roots matter, helping children learn who they are by virtue of where they are.

Smith (1998, 2002) and others (Gruenewald, 2003a, 2003b; Martin, 2001; Theobald and Curtiss, 2000) have

translated the concept of sense of place into suggestions for a curriculum of place-based education. Some early forays involved arming students with cameras to capture what they might personally see in their own neighborhoods. Other early efforts in the mid-1990s, as noted by Smith (2002: 589), involved students in in-depth explorations of their nearby natural environments, investigating the regional cultures in their part of the nation, some service learning and real-world problem solving vis-à-vis their communities, and explorations into work lives and occupations in their communities.

A central objective of place-based education, summarized by Smith (2002: 594), is to “resituate learning within the context of communities,” strengthening “children’s connections to others and to the regions in which they live.” This effort reconnects, rather than separates children from the world, serving both individuals and their communities.

Beyond the design of curricula, another example of place-based learning is supplied in Morris’ (2002) study of a single African American community in St. Louis and its longtime relationship with the neighborhood school, Farragut Elementary. In describing what he refers to as a communally bonded school, Morris (2002) identifies a tie between a low-income minority population and its academically strong elementary that transpired for many decades.

Among the key elements of place that specifically drew the school and community together, noted Morris (2002), were a number of actions. First, they cultivated intergenerational and cultural bonds, especially supported by the school’s tradition of serving as a site for historical and cultural celebrations for the community. This led to such an extremely low incidence of faculty turnover that children often received teachers who had taught their parents. Second, they reached out to families with many parent-invited functions and special events. There was a tradition by the school’s longtime principal to attend nearly every wedding, graduation, and funeral occurring in the community. Third, they emphasized personal and symbolic excellence, with teachers playing parent-like roles as they expected high-quality achievement as well individual-level responsibility. The school was positioned as a church-like substantial pillar of the black community.

The Importance of Community Capacity and Social Capital in Learning

There was a rapid expansion of efforts to coordinate children’s and family services throughout the United States from the late 1980s on. With roots in some of the Great Society interventions of the 1960s, many realized by the 1980s that several service needs of families and their children in poverty circumstances were not being met; these unmet needs could be interfering with

children’s opportunities for successful learning (see Kirst, 1989; Levy and Copple, 1989; Kagan and Neville, 1993).

An added problem was that the few available services to families were frequently so haphazardly delivered and so widely fragmented that little improvement in the lives and educational opportunities for children in poverty seemed likely. Following both state legislation (starting in New Jersey, Kentucky, and California) and major private foundation support, concerted efforts to provide coordinated services to families developed quickly across the US (see Smrekar and Mawhinney, 1999).

The services ranged widely from health clinics to family counseling, employment training and assistance, parenting classes, English-language learning, and housing/welfare assistance. As the services movement matured a bit, attention turned to the addition of programs of direct assistance to students via after-school tutoring, youth-development programs, summer schooling, pre-schooling, and day care (see Behrman, 1999; Jacobson, 2000). The services were sometimes school based, with facilities located on school grounds often in mobile units. In other places, they were school linked, that is, the services were held in other social-service agencies with efforts to bridge effectively between multiple service providers (e.g., churches, libraries, and boys and girls clubs) and local schools.

It was quickly discovered that the coordination of children’s services, particularly the coordination of the school-based variety, was extraordinarily difficult to achieve (Crowson and Boyd, 1993; Smylie *et al.*, 1994). Additionally, some major foundation funding was discontinued in the mid-1990s, amid disappointing results (White and Wehlage, 1995). Perhaps most significantly, a realization developed that the delivery of added professional services to families falls far short of the full scope of efforts needed to strengthen families and improve learning opportunities for children (Schorr, 1997; Driscoll and Kerchner, 1999; Boyd and Crowson, 2001).

By the late 1990s, a paradigm shift developed in the services movement that has continued on into the twenty-first century. First, interest began to focus on community development in place of service delivery. Key goals of community development include improving the economic, social, and cultural capital of the communities, along with the abilities of communities to use their capital to assist in the learning and development of the community’s children. Second, the movement has refocused interest around strengthening communities in order to strengthen the schools with a direct pedagogical emphasis upon evidence of improved learning.

An excellent exemplar of the shift from services to capacity for learning is found in the UK. In what British authorities prefer to identify as joined-up thinking, close to 25 education action zones (EAZs) were announced by the New Labour government in 1998, with a second round

of zones added a year later. The EAZ concept is built upon community development partnerships between an array of private, voluntary, and public agencies representing many societal sectors (e.g., business, health, welfare, recreation, church, and police; Power, 2001). There is typically close partnering with the schools and often, an identification of standards for EAZ success, such as achievement gains, improved school attendance, and reduced school leaving (Storey, 2005).

In the EAZs, there is also a deinstitutionalization of learning (Power, 2001: 24). The idea of deinstitutionalizing learning represents an effort on the part of each zone to take learning out of the sole responsibility of the classroom toward extended learning opportunities throughout the community. This is often achieved through partnerships with a local airport, engineering and construction firms, a local trucking company, or local retailers. The idea of creating extended learning systems for communities, which include, and also go far beyond the schools, has also been proposed in the US, in a 2007 report by a special task force funded by the Mott Foundation (Greifner, 2007).

An important issue in extended learning is how to pay for it. Increasingly, the twin notions of capacity and extended learning have led to an interest in extended funding. Picus *et al.* (2002) discovered that less than half of the money spent on children's education in Los Angeles and in many communities can be accounted for in the funding allocated to neighborhoods by the public schools. Many additional resources, with important learning consequences, flow out of the work of childcare servers, after-school offerings, nonprofits, youth services, city agencies, and, of course, family-to-family spending.

Picus *et al.* (2002) proposed an examination of the total public resources spent on children and families in specific geographical areas, and proposed further some financial rethinking around the idea of a children's budget for education in place of the traditional school-to-school budget. One major city that has come close to implementing this concept is Toronto, Canada – which produces an annual report card on children, tracing allocations of family services by census tract across the city, linking these data to trends in school achievement over time zone by zone.

The notion of capacity, in providing communities with opportunities to influence children's learning, is not limited to economic resources and investments. Sanyal (1994) draws upon the work of Hirschman (1971) in noting that the social construction of hope in a community can be a powerful developmental force. A community's capacity to monitor the institutions that serve it also represents an important strength, one that emerged as a key element in the empowerment of low-income communities in Shirley's (1997, 2001) research into Texas' Alliance schools. Another important element is the connective

capacity of communities (Yoshikawa *et al.*, 2006), wherein networks of families influence learning opportunities by helping one another, helping with elements such as neighborhood security, and passing on information to one another. Franklin and Savage (2004) documented the remarkable connectedness that often characterized African American communities and their drive toward educational advancement in the US, in predesegregation times (see also Anderson, 1988; Franklin, 1992; Walker, 1996).

While capacity is a term often used at the community level to describe the resources available for learning, the closely related and equally important concept of capital is typically used at the family level of analysis (Coleman, 1987, 1988). The financial capital available to a family is rather self-evidently important, for there can be wide variations in the resources that families, on their own, devote to their children's education. Human capital is not uncommonly regarded as the central measure of educational accomplishment for both individuals and the larger society. For individuals, human capital is often measured in terms of school attainment as a representation of earnings potential.

Cultural capital casts a wide net around learning-related patterns of child rearing, language use, family values and priorities, lifestyles, traditions, and concepts of right or wrong. Cultural differences, once commonly viewed as deficits by many educators, are now much better understood as frequent sources of cultural strengths. These strengths can be drawn upon in their own right and in their own way for improved learning (see Lareau, 1987, 1989; Trueba, 1999, 2002). Finally, social capital represents an important element of family access to resources by virtue of their connections, networks, and interactions with other members of the community (see Coleman, 1988; Putnam, 2000, 2002; Putnam and Feldstein, 2003). Social capital is used by parents in navigating their way through the structures and processes of schooling, and in their own involvement in school activities (Sheldon, 2002).

What is especially important in the school–community relationship is the ability of the school to activate the social capital of families in the surrounding community. Two terms used frequently in the consideration of activating and extending social capital are: bridging and bonding. Bridging attempts to bind across social distances; it is important, observe Putnam and Feldstein (2003: 279), “for reconciling democracy and diversity,” and for engaging successfully in the “tough task of fostering social ties that reach across social divisions.” Bonding social capital reflects the ties that link individuals or groups who do have much in common. Through bonding, community members reinforce shared identities, norms and values, traditions, likes and dislikes, and togetherness. The term can reflect the time parents and children share and interact together, including the time parents and children spend on learning activities (Coleman, 1988).

Communities and Families as Instructional Resources

Once held at arms length by school administrators, the realization today is that communities and families are vital parts of the learning equation. From a conceptual perspective, this realization owes much to advances in cognitive science, particularly with the development of a well-focused body of recent inquiry into how people learn (Bransford *et al.*, 2000).

Often labeled the new science of learning, the cognition theory understands that knowledge is best mastered when it is developmentally scaffolded onto preexisting knowledge, the lion's share of which learners bring to school from home and community contexts. The community context can both help and hinder in-school learning, but at the very least, argue Bransford (2000) and others (see, Driscoll and Goldring, 2006; Melaville *et al.*, 2003), the school should seek to understand the out-of-school context and should use that context effectively in its own instructional programming. A community-centered approach to instruction is urged, wherein both a deep understanding and a wide use of the nonschool environment are central to the school curriculum. Melaville *et al.* (2003: 27) note "effective learning occurs when schools, after school programs, and other organizations use the resources and challenges of the community as a living textbook for learning."

A community-centered approach to schooling proceeds from the ecological understanding that each community is a web of learning environments. Rather surprisingly, despite increasing interest in and attention to the community context, there has not yet developed an exceptionally strong knowledge base for educators to draw upon in using the community effectively as an instructional resource.

Traditionally, the focus in attempting to strengthen the school-community relationship has been upon the use of partnerships and collaborations (see Henry, 1996; Rigsby *et al.*, 1995). However, not every part of a community will or can partner. Yet, many nonpartnering elements in the community are still quite essential to learning. Thus, the renewed goal is to seek to engage the community instructionally, rather than to partner with it.

Four key areas of school-community engagement are generally found in today's learning environment: (1) programming to meet the needs of children, (2) the active engagement of parents in learning, (3) linkages with other community institutions in assisting the work of the schools, and (4) actions to help build a community's capacity for learning.

Schools throughout the US have increasingly been drawn, often through legislative mandates, into taking responsibility for attending to the special needs of children and youth. This has required deeper relations with an array of other social service agencies such as welfare,

housing, health, nutrition, and delinquency prevention and treatment. The current emphasis upon student achievement in educational reform provides an added incentive to help those with difficulties in learning. Thus, a much wider array of service relationships are now found helping families in areas such as financial and legal services, counseling, medical care, tutoring, English-language learning, and after-school programming (Springate and Stegelin, 1999).

As never before, parent involvement is accepted and encouraged in schools throughout much of the US. The usual array of homework activities has been broadened in many locales to include assistance to families with parenting skills, access to family support services, counseling assistance, legal services, and the like (Springate and Stegelin, 1999; Epstein *et al.*, 2002). Traditional roles for parents in the school have been to assist at ceremonial levels of involvement. However, over time, roles for parents in learning activities within the school have broadened considerably to visiting classrooms, tutoring, and performing aide roles (Sandell, 1998). Increasingly, the schools have also become more comfortable with advisory and governance roles for parents. Most recently, in some schools, parents are able to play a direct role in school evaluation or assessment (Springate and Stegelin, 1999).

Some scholars have expressed a concern that the overall level of civic engagement between families and community institutions, including the schools, is on the wane (see Skocpol and Fiorina, 1999). Other observers claim that the learning-related engagement of broad-based community institutions (from faith-based to business to community organizations, to health care, to the non-profits) remains strong and is, in fact, on the increase (Chadwick, 2004). Both observations may be true. Communities vary enormously in their sense of themselves as an engaged, education-minded collectivity. Chadwick (2004) notes, however, that an engaged community tends to demonstrate greater social capital, an added commitment to education, and improved student achievement.

Finally, an engaged community can increase its capacity for learning – with interorganizational connections, and broker organizations that help to negotiate relationships for improved learning. These communities can also develop specialized partnerships with specific organizations toward well-defined ends such as a coalition of police, youth organizations, neighborhood-watch volunteers, and business leaders to help reduce crime (Chaskin *et al.*, 2001).

A Local Politics of Partnership and Empowerment

The new localism in education is also reflected in altered political relationships at the community level, and in new

approaches to the empowerment of the school clientele. Two models that have surfaced most impressively in the twenty-first century are: (1) an alliances approach to community organizing for school improvement (see Shirley, 1997, 2001; Crowson, 2003) and (2) a preferences or choice model of client empowerment (see Berends *et al.*, 2007).

Much of the community organizing that characterized early efforts attempted to empower individual communities toward taking charge of their local institutions, including the schools. These efforts heavily focused upon organizational enemies, with a confrontational style that could often be quite successful in bringing inequities to light and in calling public attention to unmet needs. Much of this approach, however, was heavily based upon alienation, rather than cooperation. The result was often brief, without lasting cooperation. By the late twentieth century, this old-style activism began to be replaced by a much less confrontational style in community organizing. There was a movement toward a strategy of working with community institutions, rather than against them. Interestingly, many of the alliance initiatives in the US that used community organizing around improving the public schools came out of the early leadership of faith-based institutions.

Shirley (1997, 2001) described, in depth, some of the first community alliances that developed in low-income neighborhoods in the state of Texas. These alliances used a strategy that often started with small goals such as seeking an added school crossing guard at a busy corner, and subsequently continued on to greater and greater activity and engagement in school improvement. Shirley noted that by mid-2001, a network of community alliances in Texas had grown to nearly 150 schools, all based in some of the poorest communities in the state. The alliances approach to community organizing around school reform has subsequently spread to other states around the US. The public- and private-sector cooperation found in the alliances arena is quite similar to the EAZs effort in the UK.

Also widespread in the US is a strategy for school improvement through local empowerment that relies heavily on the choices and educational preferences of individual families. The capacity to vote with one's feet, that is, to exit a low-performing school to a private school or move to a higher-performing school district has long been available to families of means, those who can afford such a move. A rationale of the current choice movement in the US is that the same set of options should be available to lower-income families. The rationale continues, with a belief that the availability of choice creates competitive pressures upon the public schools to improve.

Choice options have become increasingly available across much of the US, in the form of a steadily growing array of magnet schools, charter schools, and, to a somewhat lesser extent, voucher-supported schools. The numbers, along with the terms and conditions of their

development, vary considerably from state to state. The federal education reform legislation in the US allows students attending under-performing schools to transfer to charter schools, with the money following the student to the charter school. The district is responsible for transportation as well. Kirst (2007) has observed that charter schools had spread to 40 states in the US between 1991 and 2005, but only 21 of the states have strong charter laws permitting and encouraging significant expansion and innovation.

The choice phenomenon has grown so rapidly in the US that many of the larger implications for communities remain to be fully investigated. In some early inquiry, Smrekar and Goldring (1999) found that individual families do not tend to choose to move toward a new and better school, but choose instead to move from a school with which they are dissatisfied. The same authors expressed a concern about a possible paradox of choice, in which individual preferences and one-family-at-a-time empowerment can result in a loss of connections in the geographic community (Smrekar and Goldring, 1999). Nevertheless, observe DeBray-Pelot *et al.* (2007), school choice advocacy remains extremely lively in the US, and currently transcends traditional (left-right) ideological positions.

By no means is the US alone in the use of choice- and market-oriented approaches to school reform and to local-level empowerment. West and Hind (1997) note that across much of the world, there has been a press toward greater reliance on the market in educational improvement (see also Glenn, 2007). In examining parental choices among state-maintained secondary schools in London, West and Hind (2007) found clear differences by type of secondary school in school composition across racial, ethnic, and income divisions. Glenn (2007), in examining choice programs adopted by Denmark, France, and Spain, notes that there are deep questions surrounding school choice in finding a balance point in empowerment between local autonomy and governmental accountability.

Summary

Centralizing tendencies, with emphases upon national standards and outcomes-based accountability, have provided a foundation for recent school reform, particularly in the US. At the same time, there has been a restoration of considerable interest in the basic unit of education: the individual school site and its surrounding community. Described as the new localism, this restored interest in the school-community relationship is rooted in the central objective of improved opportunities for student learning. Four broad categories that focus on improved learning in the school-community relationship and the new localism have emerged: (1) the role of place in

education, (2) the importance of local capacity and capital in learning, (3) the community and the family as instructional resources, and (4) an emerging politics of the school–community relationship around two very different models of alliances and choice.

In the opinion of many scholars, the new localism starts with a concept of place. Schools matter as places in their communities, and places matter as social and cultural settings for the schools. Represented in the concept of place are important developmental elements in learning – such as cultural socialization, the development of a sense of roots, shared experiences and memories, and socially constructed patterns of behavior and language. A pedagogy of place-based education has also emerged from the place concept, with the central objective of resituating learning for children within the context of their communities and strengthening the connections of children to other persons in their communities.

It has been thoroughly recognized for some time that some communities are less able than others to provide the elements of capacity and capital that facilitate successful learning in school. Efforts to compensate for weakened capacity characterized state and national assistance to schools and communities in the US from the 1960s on, and further characterized efforts to provide coordinated children's and family services (or full-service schooling) in many areas later in the century. A paradigm shift has occurred in twenty-first-century efforts to generate added community capacity and social capital. The shift is well represented in programs in the UK which engage actively in community development. These initiatives are often an array of public and private partnerships geared toward capacity development and the institutional extension of learning opportunities community-wide, beyond just the schools.

A new science of learning is a natural accompaniment to the new localism. It recognizes that learning in school environments builds, or is scaffolded, upon learning that comes initially from home and community contexts. This perspective also takes into account the realization that each community is a web of learning environments that includes, but is by no means limited to, the school. It furthermore recognizes that successful schools can develop out of and profit mightily from outreach efforts to bring about active engagement of parents, community institutions, local leadership, profit sectors, and the non-profits into sets of learning connections.

The new localism is also represented in community empowerment, but best represented in two quite different models at present. Alliances of community institutions and the schools, using community organizing toward shared school-improvement efforts, have gained increased attention and credibility, particularly across the US and England. Simultaneously, the empowerment of individual families and students through the use of choice has gained considerable interest across many nations of the world.

The consequences of choice for communities and the consequences of using market forces as the key vehicle to improve opportunities for learning remain to be fully understood; however, these topics are currently receiving considerable research attention along with a good bit of policy experimentation around the globe.

Bibliography

- Anderson, J. (1988). *The Education of Blacks in the South, 1860–1935*. Chapel Hill, NC: University of North Carolina Press.
- Behrman, R. E. (ed.) (1999). When school is out. *The Future of Children* 9(2), 4–20.
- Berends, M., Langevin, W., and Springer, M. G. (eds.) (2007). Policy, politics, and organization of school choice. *Special Issue. Peabody Journal of Education* 82(2–3).
- Boyd, W. L. and Crowson, R. L. (2001). Coordinated school-linked services: The U.S. experience. In Riddell, S. and Tett, L. (eds.) *Education, Social Justice and Inter-agency Working: Joined-up or Fractured Policy?* pp 58–69. London: Routledge.
- Bransford, J., Brown, A., and Cocking, R. (eds.) (2000). *How People Learn: Brain, Mind, Experience, and School. Commission on Behavioral and Social Sciences and Education, National Research Council*. Washington, DC: National Academy Press.
- Chadwick, K. G. (2004). *Improving Schools through Community Engagement*. Thousand Oaks, CA: Corwin Press.
- Chaskin, R. J., Brown, P., Venkatesh, S., and Vidal, A. (2001). *Building Community Capacity*. New York: Aldine de Gruyter.
- Coleman, J. S. (1987). Families and schools. *Educational Researcher* 16(6), 32–38.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology* 94, S95–S120.
- Community Partnerships: Your Handbook for Action*, 2nd edn. Thousand Oaks, CA: Corwin Press.
- Crowson, R. L. (2003). Empowerment models for interprofessional collaboration. In Brabeck, M. M., Walsh, M. E., and Latte, R. E. (eds.) *Meeting at the Hyphen: Schools–Universities–Communities–Professions in Collaboration for Student Achievement and Well Being*, pp 74–93. Chicago, IL: The University of Chicago Press.
- Crowson, R. L. and Boyd, W. L. (1993). Coordinated Services for children: Designing arks for storms and seas unknown. *American Journal of Education* 101(2), 140–179.
- DeBray-Pelot, E. H., Lubienski, C. A., and Scott, J. F. (2007). The institutional landscape of interest group politics and school choice. *Peabody Journal of Education* 82(2–3), 204–230.
- Driscoll, M. E. (2001). The sense of place and the neighborhood school. In Crowson, R. L. (ed.) *Community Development and School Reform*, pp 19–41. Oxford: Elsevier Science.
- Driscoll, M. and Goldring, E. (2006). How can educational leadership incorporate communities as contexts for student learning? In Firestone, W. and Riehl, C. (eds.) *A New Agenda for Research in Educational Leadership*, pp 61–80. New York: Teachers College Press.
- Driscoll, M. E. and Kerchner, C. T. (1999). The implications of social capital for schools, communities and cities. In Murphy, J. and Louis, K. S. (eds.) *Handbook of Research on Educational Administration*, 2nd edn., pp 385–404. San Francisco, CA: Jossey-Bass.
- Epstein, J. L., Sanders, M. G., Simon, B. S., et al. (2002). *School, Family, and for Children and Families: Understanding the Past to Shape the Future*. New Haven, CT: Yale University Press.
- Franklin, V. P. (1992). *Black Self-determination: A Cultural History of African-American Resistance*. Brooklyn, NY: Lawrence Hill.
- Franklin, V. P. and Savage, C. J. (eds.) (2001). *Cultural Capital and Black Education: African American Communities and the Funding of Black Schooling, 1865 to the Present*. Greenwich, CT: Information Age Publishing.
- Glenn, C. L. (2007). Common problems, different solutions. *Peabody Journal of Education* 82(2–3), 530–548.

- Greifner, L. (2007). Panel favors extended view of learning. *Education Week* 26(20), 1, 25.
- Gruenewald, D. (2003a). The best of both worlds: A critical pedagogy of place. *Educational Researcher* 32(4), 3–12.
- Gruenewald, D. (2003b). Foundations of place: a multidisciplinary framework for place-conscious education. *American Educational Research Journal* 40, 619–654.
- Henry, M. E. (1996). *Parent–School Collaboration*. Albany, NY: State University of New York Press.
- Hirschman, A. O. (1971). *A Bias for Hope: Essays on Development and Latin America*. New Haven, CT: Yale University Press.
- Jacobson, L. (2000). New event making case for better after-school options. *Education Week* 20(6), 6.
- Kagan, S. I. and Neville, P. R. (1993). *Integrating Services FOR Children and Families: Understanding the Past to Shape the Future*. New Haven, CT: Yale University Press.
- Kirst, M. (ed.) (1989). *The Conditions of Children in California*. Berkeley, CA: Policy Analysis for California Education (PACE).
- Kirst, M. W. (2007). Politics of charter schools: Competing national advocacy coalitions meet local politics. *Peabody Journal of Education* 82(2–3), 184–203.
- Lareau, A. (1987). Social class differences in family–school relationships: The importance of cultural capital. *Sociology of Education* 60(2), 73–85.
- Lareau, A. (1989). *Home Advantage: Social Class and Parental Intervention in Elementary Education*. London: Falmer Press.
- Levy, J. and Copple, C. (1989). *Joining Forces: A Report from the First Year*. Alexandria, VA: National Association of State Boards of Education.
- Martin, J. (2001). Learning to teach students in the community and environment. *Clearing Magazine* (Fall 110), 10–13.
- Melaville, A., Shah, B., and Blank, M. (2003). *Making the Difference: Research and Practice in Community Schools*. Washington, DC: Coalition for Community Schools, Institute for Educational Leadership.
- Morris, J. E. (2002). A community-bonded school for African American students, families, and a community. *Phi Delta Kappan* 84(3), 230–234.
- Picus, L. O., McCoskey, J., Robillard, E., Yao, J., and Marsenich, L. (2002). Using student level data to measure school finance adequacy: An exploratory analysis. In Roelke, C. and Rice, J. K. (eds.) *Fiscal Policy in Urban Education*, pp 181–201. Greenwich, CT: Information Age Publishing.
- Power, S. (2001). ‘Joined-up’ thinking: Interagency partnerships in education action zones. In Riddell, S. and Tett, L. (eds.) *Education, Social Justice and Inter-agency Working: Joined-up or Fractured Policy?* pp 14–28. London: Routledge.
- Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.
- Putnam, R. D. (ed.) (2002). *Democracies in Flux: The Evolution of Social Capital in Contemporary Society*. Oxford: Oxford University Press.
- Putnam, R. D. and Feldstein, L. M. (2003). *Better Together: Restoring the American Community*. New York: Simon & Schuster.
- Rigsby, L. C., Reynolds, M. C., and Wang, M. C. (eds.) (1995). *School–Community Connections*. San Francisco, CA: Jossey-Bass.
- Sandell, E. J. (1998). Parents in the schools. In Fuller, M. L. and Olsen, G. (eds.) *Home–school Relations: Working Successfully with Parents and Families*, pp 127–150. Boston, MA: Allyn and Bacon.
- Sanyal, B. (1994). Social construction of hope. In Rodwin, L. and Schon, D. A. (eds.) *Rethinking the Development Experience*, pp 131–144. Washington, DC: The Brookings Institution.
- Schorr, L. B. (1997). *Common Purposes: Strengthening Families and Neighborhoods to Rebuild America*. New York: Anchor Books.
- Sheldon, S. B. (2002). Parents’ social networks and beliefs as predictors of parent involvement. *The Elementary School Journal* 102(4), 302–318.
- Shirley, D. (1997). *Community Organizing for Urban School Reform*. Austin, TX: University of Texas Press.
- Shirley, D. (2001). Linking community organizing and school reform: A comprehensive analysis. In Crowson, R. L. (ed.) *Community Development and School Reform*, pp 139–169. Oxford: Elsevier Science.
- Skocpol, T. and Fiorina, M. P. (eds.) (1999). *Civic Engagement in American Democracy*. Washington, DC: Brookings Institution Press.
- Smith, G. (1998). Rooting children in place. *Encounter* 11(4), 13–24.
- Smith, G. (2002). Place-based education: learning to be where we are. *Phi Delta Kappan* 83(8), 584–594.
- Smrekar, C. and Goldring, E. (1999). *School Choice in Urban America: Magnet Schools and the Pursuit of Equity*. New York: Teachers College Press.
- Smrekar, C. E. and Mawhinney, H. B. (1999). Integrated services: Challenges in linking schools, families, and communities. In Murphy, J. and Louis, K. S. (eds.) *Handbook of Research on Educational Administration*, 2nd edn., pp 443–461. San Francisco, CA: Jossey-Bass.
- Smylie, M. A., Crowson, R. L., Chou, V., and Levin, R. (1994). The principal and community-school connections in Chicago’s radical reform. *Educational Administration Quarterly* 31(3), 342–364.
- Springate, K. W. and Stegelin, D. A. (1999). *Building School and Community Partnerships through Parent Involvement*. Upper Saddle River, NJ: Prentice-Hall.
- Storey, V. A. (2005). The eye of power. Partnership dynamics and their impact on an area basic initiative in England. Unpublished PhD Dissertation: Vanderbilt University, Nashville, TN.
- Theobald, P. and Curtiss, J. (2000). Communities as curricula. *Forum for Applied Research and Public Policy* 15(1), 106–111.
- Trueba, H. T. (1999). *Latinas Unidas: From Cultural Diversity to the Politics of Solidarity*. Lanham, MD: Rowman and Littlefield.
- Trueba, H. T. (2002). Multiple ethnic, racial, and cultural identities in action: From marginality to a new cultural capital in modern society. *Journal of Latinos and Education* 1(1), 7–28.
- Walker, V. S. (1996). *Their Highest Potential: An African American School Community in the Segregated South*. Chapel Hill, NC: University of North Carolina Press.
- West, A. and Hind, A. (2007). School choice in London, England: Characteristics of students in different types of secondary schools. *Peabody Journal of Education* 82(2–3), 498–529.
- White, J. and Wehlage, G. (1995). Community collaboration: if it is such a good idea, why is it so hard to do? *Educational Evaluation and Policy Analysis* 17(1), 23–38.
- Yoshikawa, H., Weisner, T., and Lowe, E. D. (eds.) (2006). *Making It Work: Low-wage Employment, Family Life and Child Development*. New York: Russell Sage Foundation.

Further Reading

- Sobel, C. F. (1994). Learning by monitoring. In Rodwin, L. and Schon, D. A. (eds.) *Rethinking the Development Experience*, pp 231–274. Washington, DC: The Brookings Institution.

Social Capital, Educational Institutions and Leadership

S Kilpatrick and S Johns, University of Tasmania, Launceston, TAS, Australia

B Mulford, University of Tasmania, Hobart, TAS, Australia

© 2010 Elsevier Ltd. All rights reserved.

The idea of social capital has enjoyed a remarkable rise to prominence. The concept now sits alongside other forms of capital (such as economic and human) as an accepted contributor to our individual, community, and national well-being. International bodies such as UNESCO, OECD, and World Bank have engaged in extensive conceptual, empirical, and policy-related work in the area (Isham *et al.*, 2002) and a number of websites are devoted entirely to the area. However, the application of the concept of social capital to educational institutions has been limited.

Yet, schools, universities, and other educational institutions are taking on a much broader role than just educating students. Schools and universities are enduring institutions in our communities. As a result of the changing nature of Western society and the weakening of social institutions including the family and church (Coleman, 1988; Driscoll and Kerchner, 1999) there is an expectation on the part of the public and governments that educational institutions will contribute to the greater good of the communities in which they are located. By building social capital, educational institutions can play a community development role that will create and promote better opportunities for students and other community members.

What Is Social Capital?

In a recent analysis of contemporary academic literature in the area, the World Bank (Grootaert *et al.*, 2004) found that social capital has been discussed in two related but different ways. The first approach was subjective or cognitive in nature and referred to the resources (such as information, ideas, support) that individuals were able to procure by virtue of their relationships with other people (Portes, 1998). The second approach was structural in nature and referred to the type and extent of one's involvement in various informal networks and formal civic organizations (Putnam, 2000).

However, social capital is defined most simply as the norms and networks that enable people to act collectively (Woolcock and Narayan, 2000). It is a set of resources that resides in the relationships among people that allows them to share their knowledge and skills, or human capital. Sharing of knowledge and skills is facilitated by structures, procedures, and mechanisms that bring people together in places and environments that encourage and

support interaction (Coleman, 1988). Social capital is simultaneously drawn on and reproduced (used and built) in collaborative action. The quality and nature of social capital depends on the qualitative dimensions of the social interactions in which it is produced, such as the extent and quality of trust, reciprocity and shared values, and norms (Falk and Kilpatrick, 2000). Social capital is associated with social cohesion and civic participation.

Field (2005) argues that social capital is an independent variable that explains some variations in learning. People's social relationships play a vital part in their capacity for learning. Social connections help generate trust between people, and thereby foster the exchange of information and ideas. People's networks are learning resources, which can give them greater access to and enhanced capabilities of using, information and skills. Schuller (2007) argues that human capital which is not linked to social capital is harder to acquire and its value is harder to realize.

Social capital is a useful lens for analyzing lifelong learning and its relationship to community development (Kilpatrick *et al.*, 2003). Community social infrastructure, in particular organization structures, rules and procedures, opportunities for meeting, and human infrastructure, influences the effectiveness and efficiency of networks and associated norms. Schools, universities, and other educational institutions are significant pieces of community social infrastructure, and their leadership and staff are part of a community's human infrastructure that can forge and maintain networks and share and influence norms.

The concept of social capital is multidisciplinary, drawing on sociology, economics, political science, and education. From the sociological perspective, Bourdieu (1983) identified cultural, symbolic, and social capital, all of which he regarded as disguised forms of economic capital. Coleman (1988) applied social capital in analyzing how the family, schools, and community generate human capital. Trust emerged as an important dimension of social capital in Putnam's (1993) work on the relationship between differing civic traditions and democratic effectiveness in Italy. Work by the OECD and others (Schuller, 2001) has found that social capital enhances the effectiveness of physical and human capital; it oils the operation of industries, organizations, and communities by facilitating access to networks. Trust smoothes negotiations and so reduces transaction costs.

There are three types of social capital networks or ties. Bonding networks are ties with like others, typically family, friends, or members of a tribe. Bridging networks connect acquaintances or others in similar positions in different groups or communities, that is, those outside the close social or work circle. Linking ties are between those in different social groupings, or levels of the power hierarchy (Woolcock, 1999). Strong bonding social capital can act to exclude groups or individuals if it is not balanced by sufficient bridging ties, and is referred to as the dark side of social capital (Schuller, 2001).

Building and Using Social Capital

Social capital is built through interactions within and outside communities or groups. Falk and Kilpatrick (2000) identified two kinds of resources which people bring to interactions that are intended to result in some action for mutual benefit: knowledge and identity resources. Knowledge resources are a knowledge of who, when, and where to go for advice or resources and knowledge of how to get things done. Knowledge resources relate to networks, and an understanding of procedures and how people work effectively together. Identity resources are being able and willing (committed) to act for the benefit of the group or community and its members. Identity resources include self-confidence, norms such as reciprocity, trust, and values

and visions that are shared between the parties in the interaction. Falk and Kilpatrick's model of building and using social capital is reproduced in **Figure 1**.

Opportunities for individuals and groups to interact are important, because they allow people to use their social capital for mutually beneficial actions, and also help build or strengthen social capital. However, the quality of the outcomes possible from interactions depends on the quality of the social capital resources that are used.

Social capital has been criticized as difficult to measure, which it has been argued has reduced its uptake and application in policy. In addition, a number of studies that have attempted to link schools to the development of community social capital have been criticized because of their largely qualitative nature (e.g., case studies), or lack of specific measurement indicators. There are, however, a number of well-accepted indicators:

- voluntary participation;
- trust;
- reciprocity;
- shared norms, values, attitudes, and vision;
- network extent and structure;
- acceptance of diversity; and
- identification with the community.

Educational institutions have two spheres of influence in terms of using and building social capital: (1) building social capital within the education institutional community

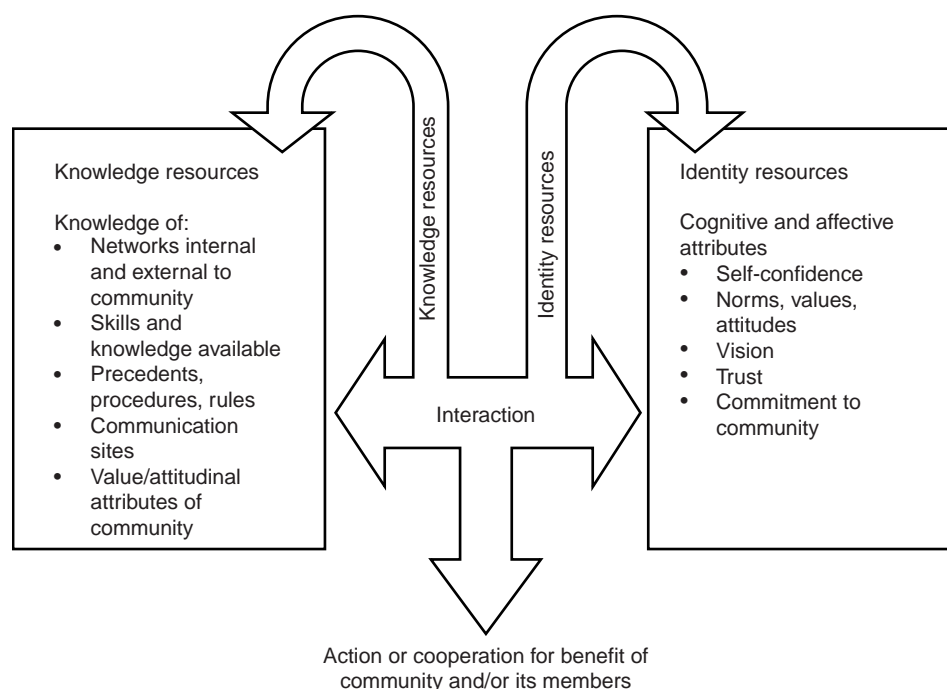


Figure 1 Building and using social capital. From Falk, I. and Kilpatrick, S. (2000). What is social capital? A study of interaction in a rural community. *Sociologia Ruralis* 40, 87–110.

(students, staff, parents) and (2) building social capital between institutional community and broader community. As the first is well documented in the research on learning organizations and professional learning communities, this article concentrates on the second sphere of influence, building social capital between institutional community and broader community.

Why Should Educational Institutions Build Community/Regional Social Capital?

The ability of leadership of the institution and community to draw on and build community social capital through collective learning, can deliver benefits to both. Educational institutions can supply the human infrastructure that makes collaborations and partnerships work. This includes enabling and supporting others to participate in the leadership process. The benefits for educational institutions and the community can be summarized as follows.

1. Benefits for educational institutions
 - Benefits for staff: responsibility for learners is shared within the community; drawing on and sharing knowledge and skills from a wider pool than just staff.
 - Benefits for students: increased access to employment and social networks, sense of identity and belonging within the community; building of inter-generational trust.
2. Benefits for community
 - Benefits for parents: social, health, well-being outcomes when schools are community hubs for delivery of services; development of social support networks.
 - Increased access to institutional facilities or shared institution–community facilities that may not otherwise be available locally; schools as hubs for lifelong learning for all.
 - Industry benefits from customized or targeted education and training provision for local needs.
 - Regional universities contribute to skilled regional workforces. They also bring a research capacity to their home locations that is rarely available elsewhere in the region. Research projects carried out locally can provide an opportunity for regional communities to examine their practices through a different lens. Research presents many opportunities for regional universities and their communities to learn together.

Regions where educational institutions, especially universities, business, and other organizations work and learn

together are more economically successful (Maskell, 2000). Social capital has been cited as the missing link in explaining the different economic success achieved by nations (and regions) with similar patterns of physical and human capital. There are similar relationships between social capital and social, cultural, and environmental benefits. It is no accident that much of the schools and social capital research focuses on either low socio-economic status or rural schools where the need to expand opportunities and build community is greatest. By virtue of their location within communities, and their place within a broader power hierarchy, educational institutions facilitate access to linking social capital for individuals and communities.

How to Build Social Capital between Education Community and Broader Community

Models

There are numerous examples of community engagement with educational institutions resulting in the building and use of social capital. These tend to fall into three broad groupings.

The first broad grouping is the educational institution as a community center or hub. Different models of educational institutions as community hubs exist, ranging from community use of educational facilities, to the delivery of health and welfare services through and with a school. Although specific terminology such as full-service schools, schools at the center, health-promoting schools and university–community engagement tend to come in and out of favor at various times, effective models demonstrate the critical role of educational institutions in building individual, family, and community social capital. Educational institutions promote community cohesiveness and identity through school/university-based events.

The second broad grouping is authentic learning opportunities with(in) the community. The move toward the increased use of learning opportunities within the community is reflected in community-based projects undertaken by school, technical and further education and university students, and vocational or technical education and training opportunities where students combine study at an educational institution with workplace experience. Evidence suggests that these activities have a range of social and economic benefits for individuals and communities such as increased youth retention in rural communities, the provision of lifelong learning opportunities for adults and more positive attitudes to education and learning within the community, and opportunities for employers to contribute to the community.

The third broad grouping revolves around the creation of new knowledge with the community. Institutions such as universities increase the research capacity of the community to connect to the global economy.

The Process

The development of social capital derives from a common purpose or vision between the educational institution and the community, and a process in which leadership is gradually shared between educational institution and community. Triggered by a need or opportunity, the process moves through an informal initiation stage that builds the foundation for a linkage or project to address the need or opportunity, towards a more formal development stage (see **Figure 2**).

Effective leadership for implementing institution–community partnerships goes further than involving or consulting with all stakeholders during the decision-making process. Rather, effective leadership for partnerships is a collective process during which educational institution and community go about developing and realizing shared visions. The stages of the leadership process for implementing school–community partnerships (**Figure 2**) illustrate how leadership gradually shifts from being the responsibility and domain of individuals early in the process, to a collective or group responsibility. That is, the vision of individuals at the trigger stage gradually shifts to a shared group (institution and community) vision during the initiation and development stages. The group develops a sense of ownership and common purpose in relation to the linkage

and, therefore, they have a vested interest in ensuring its maintenance and sustainability.

Criticism can be leveled at the project-based nature of many social capital building activities, and the associated fixed-term resourcing, leading to the question of whether such activities are sustainable in the long term or whether they create a culture of dependency. Schorr (1997) highlights the need for a systemic approach to the social capital building initiatives of educational institutions, to ensure their sustainability and replication.

A range of factors influences the extent to which educational institutions are able to develop the social capital of their communities. Those cited as having particular significance include the policy environment, resourcing issues including continuity of staffing, and leadership.

In many Western countries, there has been a move over the past decade or more, toward the development of government policy at all levels to support increased engagement between educational institutions and their communities. While such policy legitimizes community engagement activities by schools and other institutions, tension is created when there is a mismatch between policy and practice. Issues such as privacy legislation and public liability provide further challenges. Developing effective partnerships requires educational institutions and communities to learn together how to adapt, mould, and shape policy to support their vision, and to take risks when necessary.

Continuity of personnel is a key determinant of effective community engagement by educational institutions (Kilpatrick *et al.*, 2002). Rural and regional educational institutions and low socioeconomic schools, in particular,

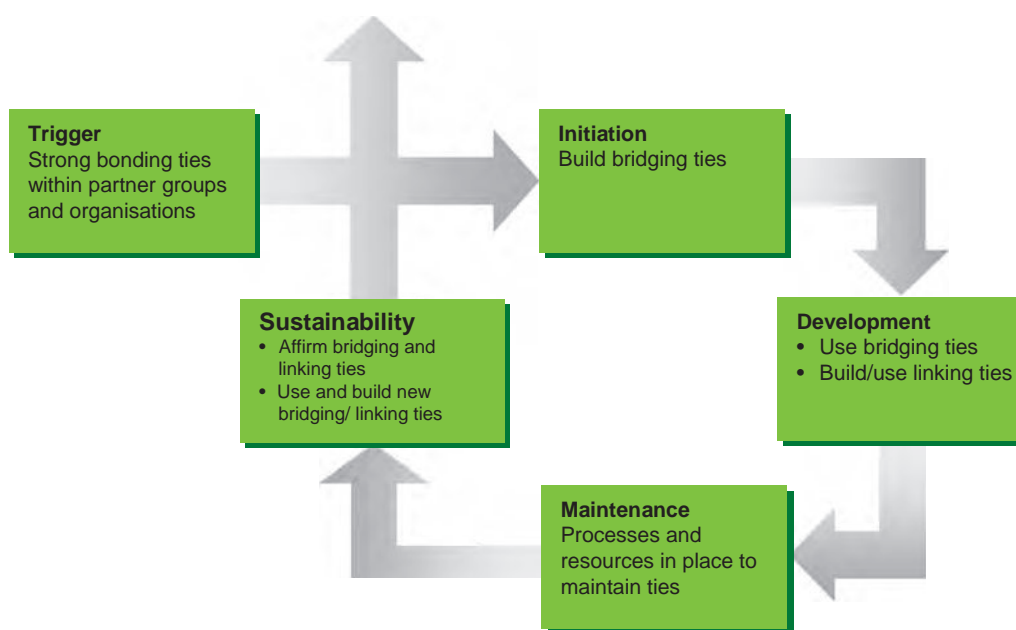


Figure 2 Building social capital through development of partnerships.

face challenges in terms of staff continuity, given their relatively high staff-turnover rates. Continuity of resources impacts on all stages of community engagement by educational institutions, but is vital as initiatives move from the development to the sustainability stage.

Leadership and Social Capital

There is increasing recognition by effective educational leaders of the value and legitimacy of community social capital outcomes, in addition to teaching and learning outcomes, and student outcomes (Mulford and Johns, 2004). Key players in unlocking and building social capital are formal educational leaders with a firm commitment to lifelong learning and a view of the community as one of a number of resources available to the school. Their roles include developing the leadership capacity of others within and external to the educational institution, and facilitating organizational learning.

In addition to the formal leaders, evidence (Kilpatrick *et al.*, 2002) suggests that effective community engagement is facilitated by boundary crossers, who speak the language of the educational institution and community, and are able to move freely between the two. Formal educational leaders recognize and nurture boundary crossers as integral to the process of community engagement.

Indicators of effective community engagement by educational institutions that use a social capital approach include the following:

- Educational leaders are committed to fostering increased integration between educational institution and community.
- Educational institution has in-depth knowledge of the community and resources available.
- Educational institution actively seeks opportunities to involve all sectors of the community, including boundary crossers, and those who would not normally have contact with the institution.
- Educational leaders display a leadership style which empowers others within the school and community and facilitates collective visioning.
- Educational institution and community have access to and utilize extensive internal and external networks.
- Educational institution and community share a vision for the future, centered on their youth and other learners.
- Educational institution and community are open to new ideas, willing to take risks, and willing to mould opportunities to match their vision.
- Educational institution and community value the skills of all in contributing to the learning of all.
- Leadership for engagement between educational institution and community is seen as the collective responsibility of the institution and community.

- Educational institution and community both view the institution as a learning center for the whole community, which brings together physical, human, and social capital resources.

Social Capital and Education in the Future

The concept of social capital is still as relevant as it was in the 1990s when it came to prominence. It underpins or is closely linked to concepts that have more recently found favor in the educational literature including learning communities, capacity building, and community engagement. The research evidence reviewed in these areas is clear in its strong support for social capital. The outcomes are impressive, not the least of which are improved student engagement, academic performance and later life chances, improved teaching and learning, reduced within-school variation, retention of teachers in the profession, and increased individual and community capacity to influence their own futures.

A way forward is to see the task of establishing learning communities as developmental, starting with the building of social capital (see Figure 3). A message is that those in educational institutions and their communities must learn how to lose time in order to gain time. Awareness of, and skill development in group and organizational processes is a first step in any effective change. Instead of others trying

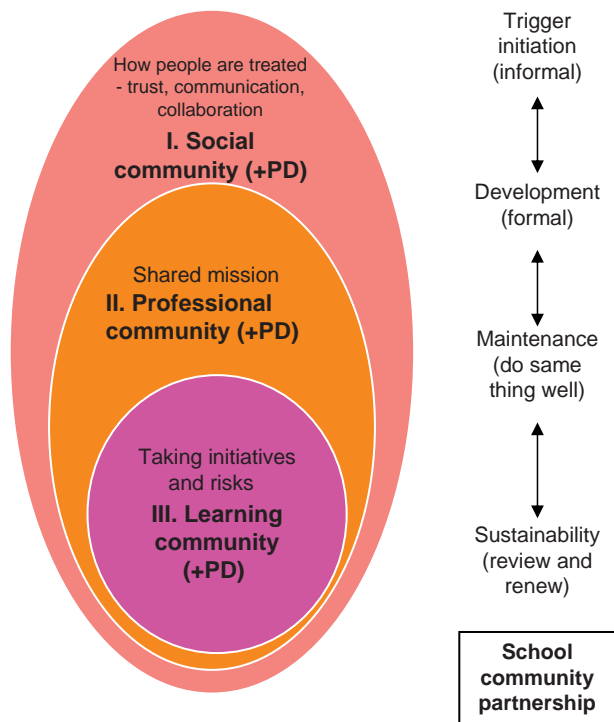


Figure 3 Social capital as the basis for learning communities.

to insert something into a school's or community's culture, the school and its community, and especially its leadership, should first be trying to help that culture develop an awareness of and a responsiveness to itself.

In brief, the position taken identifies three major, sequential and embedded elements in successful and ongoing school reform. It takes the two elements in the definition of social capital – groups, networks, norms, and trust – and for productive purposes, extends them to include a third element, learning. The first element in the sequence relates to the social community, how people are communicated with and treated. Success is more likely where people act rather than are always reacting, are empowered, involved in decision making through a transparent, facilitative, and supportive structure, and are trusted, respected, encouraged, and valued (see oval I in **Figure 3**). It is a waste of time moving to the second element until this social community is established. The second element concerns a professional community. A professional community involves shared norms and values including valuing differences and diversity, a focus on implementation and continuous enhancement of quality learning for all students, de-privatization of practice, collaboration, and critical reflective dialog, especially that based on performance data (oval II). But a professional community can be static, continuing to do the same or similar thing well. The final element relates to the presence of a capacity for change, learning, and innovation, in other words, a professional learning community (oval III).

Each element in this model, and each transition between the elements, can be facilitated by appropriate leadership and ongoing, optimistic, caring, nurturing training programs. Also, each element is a prerequisite for the other – as the ovals-within-ovals or eggs-within-eggs diagram implies, they are embedded within each other with only the emphasis changing. For example, when learning is occurring there is still a need to revisit the social community and the professional community, especially where there has been a change of personnel and/or a new governmental direction announced.

Using this analysis of social capital to understand the importance of, challenges to and developmental nature of communities, can assist in better translating the research into policy and practice. It can help us:

- understand better and be able to take action on the intricacies involved in moving an educational institution or community, or part thereof, from where it is now to becoming truly a place of ongoing excellence and equity without those in educational institutions and communities being bowled over by the demands for change that surround them;
- target appropriate interventions to ensure more effective progression through the stages. In targeting interventions, recognition will need to be given to the fact that it is a journey and that actions at one stage may be

inappropriate, or even counterproductive, at another stage; and,

- support the position that an educational institution and/or community will need to be evaluated differently depending on the stage it has reached (Mulford, 2007).

See also: Educational Leadership and Social Capital; Networked Learning Communities School-to-School Collaboration as an Essential Component of a System Reform Strategy; Networks and Communities of Knowledge; Organizational Learning in Schools; Professional Learning Community.

Bibliography

- Bourdieu, P. (1983). Economic capital, cultural capital, social capital. *Soziale-Welt* 2, 183–198.
- Coleman, J. (1988). Social capital and the creation of human capital. *American Sociological Review* 58, 1–15.
- Driscoll, M. and Kerchner, C. (1999). The implications of social capital for schools, communities, and cities: Educational administration as if a sense of place mattered. In Murphy, J. and Louis, K. S. (eds.) *Handbook of Research on Educational Administration*, 2nd edn, pp 385–404. San Francisco, CA: Jossey-Bass.
- Falk, I. and Kilpatrick, S. (2000). What is social capital? A study of interaction in a rural community. *Sociologia Ruralis* 40, 87–110.
- Field, J. (2005). *Social Capital and Lifelong Learning*. Bristol: The Policy Press.
- Grootaert, C., Narayan, D., Jones, V., and Woolcock, M. (2004). *Measuring Social Capital*, Working Paper No. 18. Washington, DC: World Bank.
- Isham, J., Kelly, T., and Ramaswamy, S. (2002). *Social Capital and Economic Development: Well-Being in Developing Countries*. Northampton, MA: Edward Elgar.
- Kilpatrick, S., Field, J., and Falk, I. (2003). Social capital: An analytical tool for exploring lifelong learning and community development. *British Education Research Journal* 29, 417–433.
- Kilpatrick, S., Johns, S., Mulford, B., Falk, I., and Prescott, L. (2002). *More than an Education: Leadership for Rural School-community Partnerships*. Barton, ACT: Rural Industries Research and Development Corporation.
- Maskell, P. (2000). Social capital, innovation and competitiveness. In Baron, S., Field, J., and Schuller, T. (eds.) *Social Capital: Critical Perspectives*, pp 111–123. Oxford: Oxford University Press.
- Mulford, B. (2007). Building social capital in professional learning communities: Importance, challenges and a way forward. In Stoll, L. and Louis, K. S. (eds.) *Professional Learning Communities: Divergence, Depth and Dilemmas*, pp 166–180. London: Open University Press/McGraw Hill.
- Mulford, B. and Johns, S. (2004). Successful school principalship. *Leading and Managing* 10, 45–76.
- Portes, A. (1998). Social capital: Its origins and applications in contemporary sociology. *Annual Review of Sociology* 24, 1–24.
- Putnam, R. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.
- Putnam, R. (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.
- Schorr, L. (1997). *Common Purpose: Strengthening Families and Neighborhoods to Rebuild America*. New York: Anchor/Doubleday.
- Schuller, T. (2001). The complementary roles of human and social capital. *Canadian Journal of Policy Research* 2, 18–24.
- Schuller, T. (2007). Reflection on the use of social capital. *Review of Social Economy* 65, 11–28.

- Woolcock, M. (1999). Social capital: The state of the notion. Paper presented at a multidisciplinary seminar on *Social Capital: Global and Local Perspectives*, Helsinki, 15 April.
- Woolcock, M. and Narayan, D. (2000). Social capital: Implications for development theory, research and policy. *World Bank Research Observer* **15**, 225–249.

Further Reading

- Baron, S., Field, J., and Schuller, T. (eds) *Social Capital: Critical Perspectives*. Oxford: Oxford University Press.
- Coleman, J. (1988). Social capital and the creation of human capital. *American Sociological Review* **58**, 1–15.
- Driscoll, M. and Kerchner, C. (1999). The implications of social capital for schools, communities, and cities: Educational administration as if a sense of place mattered. In Murphy, J. and Louis, K. S. (eds.) *Handbook of Research on Educational Administration*, 2nd edn., pp 385–404. San Francisco, CA: Jossey-Bass.
- Dryfoos, J. (1994). *Full-Service Schools: A Revolution in Health and Social Services for Children, Youth and Families*. San Francisco, CA: Jossey-Bass.
- Falk, I. and Kilpatrick, S. (2000). What is social capital? A study of interaction in a rural community. *Sociologia Ruralis* **40**, 87–110.
- Field, J. (2005). *Social Capital and Lifelong Learning*. Bristol: The Policy Press.
- Kilpatrick, S., Johns, S., Mulford, B., Falk, I., and Prescott, L. (2002). *More than an Education: Leadership for Rural School-community Partnerships*. Barton, ACT: Rural Industries Research and Development Corporation.
- Miller, B. (1995). The role of rural schools in community development: Policy issues and implications. *Journal of Research in Rural Education* **11**(3), 163–172.
- Mulford, B. and Johns, S. (2004). Successful school principalship. *Leading and Managing* **10**, 45–76.
- Schuller, T. (2001). The complementary roles of human and social capital. *Canadian Journal of Policy Research* **2**, 18–24.
- Woolcock, M. and Narayan, D. (2000). Social capital: Implications for development theory, research and policy. *World Bank Research Observer* **15**, 225–249.

Relevant Websites

- <http://www.ruraledu.org>– The Rural School and Community Trust.
- <http://www.ksg.harvard.edu>– The Saguaro Seminar: Civil engagement in America. Social capital reading list.
- <http://www.socialcapitalgateway.org>– The Social Capital Gateway: Resources for the study of social capital.
- <http://web.worldbank.org>– World Bank, Social Capital.

LEADERSHIP AND MANAGEMENT – SCHOOL EFFECTIVENESS AND IMPROVEMENT

Contents

Leadership and Technology
Leadership in the Implementation of Innovations
Leadership: School Improvement
Organizational Learning in Schools
Professional Learning Community
Transformational School Leadership

Leadership and Technology

C F Webber, University of Calgary, Calgary, AB, Canada

© 2010 Elsevier Ltd. All rights reserved.

Information and communication technology has transformed societies around the world. The introduction of radio in the first half of the twentieth century meant that news and information could spread worldwide within hours rather than days, weeks, or even months. The subsequent widespread adoption of television in the 1950s and 1960s opened windows on cultures and lifestyles that were dramatically different from those of viewers and challenged previously unexamined assumptions, beliefs, and values. As powerful as the influences of radio and television were, they were relatively minor compared to the combined effects of millions of portable computers and databases linked by Internet providers around the globe.

Computer technology has reshaped virtually every aspect of society, from business and industry to social services, entertainment, and education (Gaudelli, 2006; Krug, 2004). For citizens of all ages, portable computers that integrate the functions of a web browser, telephone, television, and radio are part of the most ordinary aspects of daily life such as checking train schedules, reading newspapers, watching favorite television shows, ordering groceries, and communicating with friends and family.

What follows is a description of current and possible relationships between technology and educational leadership. It is not a discussion that focuses on technology *per se* but, rather, on the potential of information and communication technology to build leadership capacity.

Reshaping Education

Within the educational world, information and communication technology is perceived as the most powerful change

agent in recent memory (Davies, 2002; see **Figure 1**). While educators learn to use hand-held computers to increase school security, access student information databases, document disciplinary procedures, and respond to e-mail from parents (Barrera and Warner, 2006; Brazell, 2005; Maxwell, 2006), they also facilitate the adoption of assistive technologies for students with learning and physical disabilities (Carnahan, 2006; Jeffs *et al.*, 2006; Quinn, 2003).

At the same time, expanding global access to information, ideas, and perspectives means that educators and learners in every content area and every part of the world must address challenges to disciplinary boundaries previously imposed by parameters such as culture, time, and space (Robertson and Webber, 2002). For instance, traditional classroom activities are now complemented by computer software that allows learners to access external people and information sources (Collins, 2001; Krumsvik, 2005). Social-networking software facilitates the formation of online communities that promote information sharing and collaboration among teachers and learners. Examples of social networking software include weblogs or blogs where the participants' comments to one another are archived and publicly accessible, wikis that allow website visitors to edit content, instant messaging and text messaging using computer software or telephones, and podcasting of multimedia files for recording and viewing at the convenience of recipients (Lamb and Johnson, 2006).

Social networking software brings together individuals with widely varying worldviews. They use different languages, may hold competing values, and live diverse lifestyles. At its best, social networking software may encourage the development of what Banks (2004) described as cosmopolitanism or global competency, and

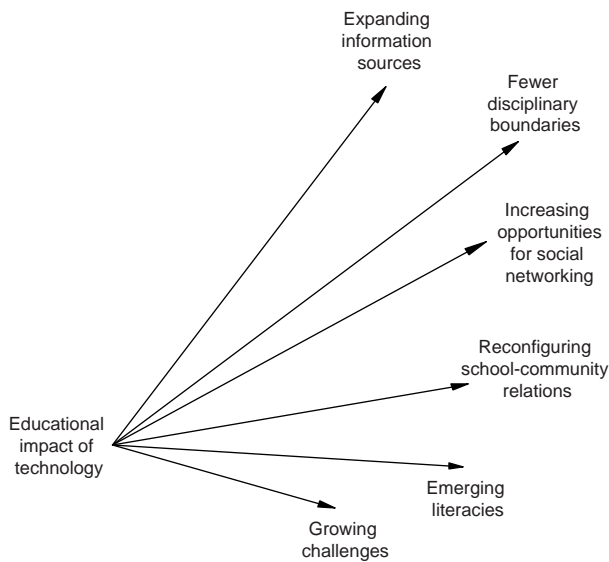


Figure 1 The educational impact of technology.

contribute to what Krumsvik (2005) called “a more open, authentic, and contextual learning environment” (p. 41). In fact, Elmborg (2006) cautioned that those who argue in favor of learning environments based on a single culture and single language risk being perceived as defensive and reactionary, and speculated that national identities dependent upon common cultural artifacts truly no longer exist.

Information and communication technology also has led to a reconsideration of how we define literacy. In the past, literacy was defined in relation to communicating within a particular community. However, social networking and information-gathering software have led to the need for multiple literacies that permit communication across cultural contexts (Elmborg, 2006).

The relations between schools and their communities have been changed dramatically by information and communication technology. Ideally, the use of technology in schools is shaped through public deliberations among educators, students, and community members, and also viewed from a critical perspective in order to optimize the benefits for learners. However, schools must also determine the pedagogical relevance of specific aspects of technology, become comfortable with the ambiguity and confusion generated within the community by constantly changing technology, and address the negative manifestations of technology, such as cyber bullying, which spill into the community (Webber and Mulford, 2007).

New software and widespread Internet access mean that educators must consider the benefits and challenges associated with several forms of instruction: traditional face-to-face instruction, hybrid or blended courses that integrate online and face-to-face teaching, online classes that are delivered entirely through the Internet, distance education that may not only use electronic tools but also

take the form of paper-based correspondence classes, and e-learning that incorporates instructional activities that are technology mediated in some way (Tallent-Runnels *et al.*, 2006). These overlapping instructional formats are complex and require different teaching skills and knowledge from those associated with traditional teaching.

Finally, it is important to note that a strong technology infrastructure does not guarantee positive outcomes for the members of educational communities (Krumsvik, 2005). In fact, the impact of technology on education has been viewed with some suspicion and skepticism (Gaudelli, 2006). Nonetheless, technology has dramatically altered how people live their lives, including how young people and adults learn.

The Need for Leadership

There is an unprecedented need for effective educational leadership due to, for example, widespread public demand for accountability, cost effectiveness, and educational improvement (Lepard, 2002). Such expectations are complex; however, at least there is a body of theoretical and empirical literature developed over many years that educational leaders can use. Nevertheless, educational leadership related to technology use is a field of study that is less mature and, therefore, it can be more difficult for school leaders to decide how technology use should be implemented (Geer, 2002; Sivan, 2000).

Nonetheless, promising efforts to provide educational technology guides for school leaders have emerged. The Technology Standards for School Administrators Collaborative (2001) has outlined six broad standards for school administrators that have been widely adopted throughout North America (e.g., Donlevy, 2004). The first standard is leadership that fosters a shared vision for technology use. A technology vision provides the base for short- and long-term planning, policy development, and resource allocation. Second, educational leaders should facilitate the integration of technologies that promote teaching and learning. Optimal use of technology should aim to improve student achievement, meet diverse learning needs, and promote higher-order thinking. Third, educational leaders should model appropriate use of technology that enhances communication and collaboration within school communities, nurtures effective professional development, and facilitates ongoing school improvement. Fourth, educational leaders should manage and support educational technology by ensuring system compatibility, allocating sufficient financial and human resources, and integrating plans and policies to maximize efficiency. The fifth standard relates to the assessment and evaluation of technology use in teaching and learning activities. For example, educational leaders should use technology to gather, interpret, and communicate data

that can be used to improve instruction, plan professional development, and inform staffing decisions. Last, educational leaders should possess a thorough knowledge of the social, legal, and ethical aspects of educational technology. That is, educational leaders should strive to ensure equitable access to technology for students and teachers, plan and enforce the responsible use of technology, promote technology use that is safe and ethical, and create policies that address copyright and intellectual property issues.

A rubric for linking educational technology with educational leadership proposed by Webber (2003a) extended the Technology Standards for School Administrators Collaborative (2001) standards by juxtaposing conservative and educative technology leadership. A conservative framework was described as based on standardization, predictability, and resistance to change. Within such a framework, educational leaders focus on technology *per se*, provide clear role definitions for learners and educators, attend to individual learning, react to emerging technologies, and pay much attention to resource acquisition and management. In contrast, leaders operating within an educative framework aim for technology use that is flexible, fragile, and a potentially high reward for learners. That is, educative leaders focus not on technology itself, but on its use as an instructional tool. They focus on vision building, problem solving, and innovation. The leaders seek engagement rather than compliance, offer ongoing professional development and cross-institutional and community boundaries, and make decisions that are transparent and sensitive to the needs of the learning community as well.

Webber (2003b) also described four leadership dimensions of learning networks supported by educational technology: professional, role, environmental, and emotional. The professional dimension was characterized by a focus on learning possibilities and role flexibility that allowed leadership to emerge from all sectors of a learning network. This dimension was also described as evidence based and future oriented. The second dimension outlined the role of leaders as challengers of standardization, intellectual explorers, entrepreneurs, politicians, and networkers. The third dimension focused on how leaders relate to their professional environments as they differentiate, for example, between seeking equity by treating everyone the same and positioning for equity by providing resources and opportunities strategically. Other environmental considerations include learning from and moving beyond the past and, also, providing opportunities for cross-role participation in learning initiatives so that students, teachers, parents, and community members can engage in overlapping activities that meet related learning needs. The fourth dimension relates to the emotions associated with active engagement in learning networks made possible by educational technology. Leaders must be prepared to deal with the passion and curiosity associated with learning

that fosters enhanced creativity, demands sensitivity and compassion, and promotes caring.

Other aspects of the leadership required by new technologies include the ability to serve as an effective change agent. While some school systems may present insurmountable challenges for educational leaders who seek to create new learning opportunities for learners through the use of technology (Hunter, 2001), leaders with the ability to facilitate successful change can achieve significant cultural and structural shifts in educational communities (Hartnell-Young, 2006). Successful change agents are able to master the tasks of acquiring and maintaining appropriate technological tools and of finding time and money to deliver meaningful professional development (Staples *et al.*, 2005). They avoid models of change that are primarily top-down and build ownership for change among key stakeholders (McGrail, 2006). Finally, they understand that effective change processes increase the likelihood that learning opportunities will be optimized for both students and teachers (Lai and Pratt, 2004).

Teacher Leadership

The widespread use of information and communication technology in schools has established a clear need for professional development among teachers (Copeland and Gray, 2002; Hunter, 2001). Just as literacy skills for students have been redefined because of technology, teachers must also develop literacy skills that permit them to navigate, assess, and utilize information sources effectively, which is to understand what Elmborg (2006) called a new grammatical system for communicating using new technology.

Technology skills are only part of what Johnson (2006) described as skills for the knowledge worker. However, new technology can facilitate the development of basic skills such as understanding visual imagery, numeracy, and effective communication, which form the framework around which higher-order thinking skills can be constructed. Discipline and profession-specific skills build upon basic skills to form a breadth of knowledge that spans the social and physical sciences, literature, and history. Basic technology skills also promote higher-order thinking skills that lead to independent learning and social responsibility, and foster conceptual skills such as the abilities to think holistically, synthesize information, and create meaning. It is essential that teachers access professional development that helps them to develop the skills of knowledge workers so they can create coherence across the curricula they teach, the technologies they use with students, and the larger mission of the school community (Staples *et al.*, 2005).

Creating educational coherence is a complex task in the information age because technology has changed teaching and learning from what most teachers were trained to do.

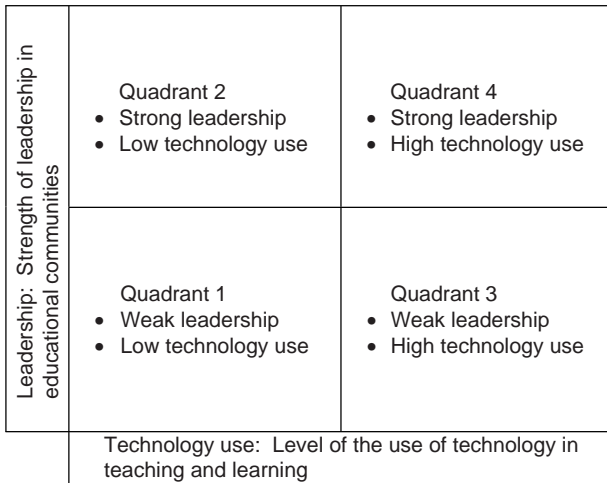


Figure 2 A framework for considering educational leadership and technology.

Teachers now are called upon to be educational entrepreneurs (Webber, 2005a) who are able to innovate, network, work in synchronous and asynchronous environments, and demonstrate concurrent local, national, and global cultural literacies. They are also expected to create educational organizations which serve as knowledge centers that integrate traditional and technology-mediated instruction.

Figure 2 portrays the relationship between educational leadership and the integration of technology in schools. In **Quadrant 1** schools, little use of technology is made by teachers, and educational leadership is ineffectual. Strong leadership exists in **Quadrant 2** learning environments; however, it is not focused on technology and teachers realize little of the potential of technology as a learning tool.

Teachers may use technology in **Quadrant 3** schools, but there is little academic coherence or shared understanding of a pedagogical rationale for utilizing it. In **Quadrant 4** schools, strong leaders facilitate the integration of appropriate technology into classroom practices so that teachers and support staff can optimize student learning.

Technology and Leadership Development

Successful leadership in the information age is largely dependent upon access to relevant professional development (see **Table 1**). In particular, it is essential that educational leaders, including principals and teachers, be able to access professional development that leads to the development of a strong sense of agency, enhanced altruism, critical analysis skills, and emotional engagement with learning (Robertson and Webber, 2002). Challenges for educators include a lack of time in their workday to participate in professional development initiatives (Hunter, 2001; Lai and Pratt, 2004). The challenge is greater for principals

Table 1 Desirable technology-mediated leadership development

Attributes	Evidence
Quality reflected in	Role-appropriate content Compelling e-learning format
Access that is	Timely Flexible
Opportunities to	Shape the learning experiences Engage Reflect Network locally, nationally, and internationally
Support infrastructure that	Invites Builds community Attends to stages of learning Reflects cross-role and cross-cultural understandings
Results in	Strong sense of professional agency Enhanced altruism Critical analysis skills Emotional engagement with learning

because the most professional development programs that address the use of educational technology are designed for teachers rather than for principals (Adsit, 2004).

Since many professional and academic organizations seek to address professional learning needs by using technology-mediated delivery formats, several frameworks for assessing e-learning programs have been proposed. For example, Killion (2002) suggested that educators seeking professional development consider several criteria, including their personal learning readiness, content quality, and flexibility of access. Other criteria include the form of e-learning, level of interactivity, technical support, and the esthetics of online interfaces.

Adsit (2004) suggested that educators considering technology-mediated professional development become knowledgeable about different levels of technology integration. That is, they should consider the suitability of low-tech/low-interactivity professional development that utilizes basic technology such as word processing, electronic handouts, videotapes, and audiotapes versus medium-tech/medium-interactivity programming that relies upon, for example, listserv and message boards. High-tech/high-interactivity professional development uses synchronous discussion groups, multimedia, interactive e-learning programs, and synchronous teleconferencing or videoconferencing. Adsit (2004) also urged educators to seek programs that promote reflection and reduce professional isolation.

Most leadership development programs are offered by universities, departments of education, and professional organizations. However, educators now have the option of participating in technology-mediated leadership development programs delivered literally from around the world

(Robertson and Webber, 2002). Therefore, leadership development programs should be assessed according to whether they promote active participation in national and international learning networks using highly interactive technologies. Another criterion suggested by Robertson and Webber (2002) is the degree to which programming facilitates cross-role dialog that expands the cultural literacy of participants to include a range of educational, community, and business settings. Delivery formats ought to include synchronous and asynchronous components that reflect an understanding of the time zones of participants and their wide range of professional obligations. Further, participants in leadership development programs should be able to engage with providers in shaping the design, delivery, and evaluation of programming so that it is relevant and useful. In other words, educational leaders need to experience leadership development activities in which they acquire the skills and knowledge needed to do similar work in their schools and communities.

Providers of technology-mediated leadership development programs should anticipate that participants require access to a support infrastructure that meets their information and administrative needs (Webber and Clark, 2006). That is, educational leaders need support in order to move beyond familiar leadership practices. They ought to feel that they are part of an inviting and supportive learning environment; therefore, community building is an essential aspect of leadership development. Providers should anticipate that participants in leadership development initiatives will move through a predictable pattern that begins with a focus on the managerial aspects of the initiative, such as how the program is organized and what is expected of participants. The second stage of participant development is likely to focus on the context of the leadership program. That is, educational leaders will want to know the forms of literacy and the skill sets that are required. They must also acquire the ability to critically analyze theoretical and empirical aspects of the program. The third stage is personal and it is characterized by self-reflection about who the learner is and the type of professional growth that is desired. Participants seek a deeper understanding of their personal strengths and challenges. In the fourth stage observed by Webber and Clark (2006), learners express concern about how others in the leadership development initiative are affected by their participation. They also seek to be considerate of the challenges that other participants face because of professional responsibilities and geographic locations.

The content of technology-mediated leadership development ought to include some of the components of educational leadership programs delivered in more traditional ways: leadership theory, change and influence in education, educational environments, and globalization (Webber, 2005b). Technology-mediated leadership development allows for a rich discussion of leadership theory,

for example, because participants can be from contexts with quite different ways of considering interpersonal communication, gender, diversity, ethics, and values in leadership and management. Similarly, the consideration of change theory will be enriched by the different organizational cultures and governance structures represented by participants. The environments in which participants work will have different political attributes and varying forms of accountability. Further, globalization will have disparate effects upon participants from venues where legal, cultural, and technological features vary. It is important that technology be kept in the background and that content and learning processes be the primary focus.

Technology-mediated leadership development is likely to contain cross-cultural components because of the ability of educational leaders to participate from many parts of the world (Webber and Robertson, 2004). Thus, the benefits of participation may include an increased ability to understand one's local educational context in relation to other settings. Other results can include heightened self-awareness, reduced professional isolation, and resistance to conformity. However, it is possible that participants may come to some naive conclusions about their own and other cultures. Another benefit of cross-cultural dialog may be a shift from ethnocentricity to cross-cultural understanding. As a result, participants may gain insight into the tension between tradition and innovation, recognize the limits of their personal cultural literacy, have more realistic understandings of the educational practices of other nations, and develop a degree of cultural humility. A corresponding caution is that some participants may come to believe incorrectly that they can remove cultural bias from their worldviews. Other benefits that may accrue to participants in technology-mediated, cross-cultural leadership development programming, according to Webber and Robertson (2004), may include the use of new technological learning tools, identification of previously unseen connections between schools and the larger community, heightened emotional engagement with work, and a higher level of altruism. These benefits may be accompanied by an unrealistic sense of power that, in the absence of an ongoing support infrastructure, may dissipate rapidly following the conclusion of the leadership development opportunity.

Social and Legal Challenges

As educational leaders engage with technology in their professional development and daily practices, they may be challenged by a set of social and legal issues that confront educators in many nations. For example, Berliner (2006) outlined the ongoing educational challenges posed by poverty in his country, the United States. While not writing specifically about educational technology, Berliner's cautions, nevertheless, are germane. Clearly, poverty

has a major impact on the access of learners to high-quality schools, including those where technology plays a major role in teaching and learning. When the barriers of poverty are combined with associated factors of race and language, the result is a serious challenge to the potential benefits of technology-enhanced learning environments.

Davies (2002) raised parallel cautions about the ability of some students, families, and schools in England to afford access to technology and the danger of a widening educational gap between rich and poor. Davies went so far as to ask if technology will unite or divide society because of disparities in access. Clearly, educational leaders have an obligation to seek wider access for students and colleagues in disadvantaged circumstances.

In addition to social justice issues, technology raises legal challenges. Quinn (2003) noted that technology changes so rapidly that legislators and policy makers are unable to keep pace. As a result, there is widespread uncertainty about how educational leaders ought to address challenges of freedom of speech and personal privacy, for example. School communities are also confronted with the need to create policies to address plagiarism, intellectual property rights, software licensing, and cyber bullying (Geer, 2002; Quinn, 2003; Shariff, 2004).

Conclusion

New technology has led educational leaders to reconsider major aspects of their practice, including the configuration of people involved in teaching and learning in school communities, the leadership principles that guide their work, the difficult questions that must be addressed, and

the possibilities for future directions in teaching and learning (Robertson and Webber, 2004).

Figure 3 depicts the influence of technology on professional practice and the learning community. School communities now require teachers with both technological literacy and content knowledge. It is not enough for teachers to have content knowledge and satisfactory classroom teaching skills. They must now possess the ability to select appropriate software, utilize relevant databases, and link learners from local venues and from afar in meaningful ways. They need to know enough about instructional design, multimedia, and web development to accomplish their work and support the work of their students besides knowing how to access technological support. In addition, teachers must organize, nurture, and promote technology use in their content areas. Therefore, educational leaders must know how to form teaching communities that can satisfy the academic and technological needs of learners.

While recruiting and retaining teachers who are academically and technologically knowledgeable, educational leaders need to be guided by a set of important understandings. For example, technology means that much of the work of school leaders is visible not only to the local community, but also to the world because so much of it is publicly accessible. Most schools in the Western world have websites that contain large amounts of information about members of the school community and district websites often summarize important data about the system. Provincial and state departments of education routinely display summaries of school-by-school achievement test results. Even satellite photographs of individual schools are publicly available worldwide in databases simply by typing in a school address. Thus, educational leaders need to be comfortable with the public nature of their work and use

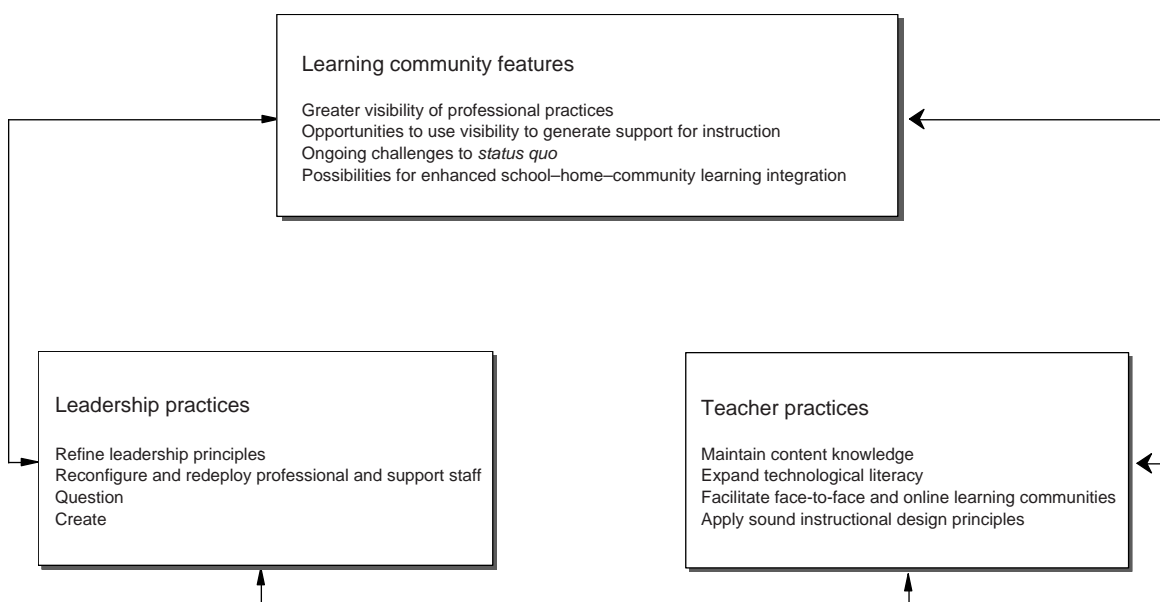


Figure 3 The influence of technology on professional practice and the learning community.

their visibility to promote support for instructional improvement. Similarly, educational leaders need to understand that technology should complement and not determine academic content and instructional practices, that online learning communities have the potential to be even stronger than traditional learning communities, and that technological infrastructures must be sustainable.

Educational leaders must have the capacity to recognize and address important questions raised by technology. For instance, leaders should strive to determine how face-to-face teaching can be balanced with e-learning to meet the needs of learners in their particular school communities. Leaders need to ask how commercial software enhances or limits learning success and if cross-cultural collaboration by students and teachers is mutually beneficial. When the *status quo* is challenged by technology, principals must ask if the *status quo* is beneficial or driven by unexamined beliefs and assumptions.

School leaders must pursue and be open to questions and possibilities opened by new information and communication technology. For instance, are new configurations of publicly and privately funded educational support appropriate? Further, are schools the most appropriate venue for all learners all the time or should technology-enabled combinations of school-, home-, and community-based schooling be considered?

Finally, leadership skills required in technologically supported educational environments include technological literacy, inventiveness, networking capacities, entrepreneurialism, and cosmopolitanism.

Bibliography

- Adsit, J. N. (2004). *Technology-Mediated Professional Development Programs for Teachers and School Leaders*. Washington, DC: American Association of Colleges for Teacher Education.
- Banks, J. A. (2004). Teaching for social justice, diversity, and citizenship in a global world. *Educational Forum* 68(4), 296–305.
- Barrera, J. M. and Warner, L. (2006). Involving families in school events. *Kappa Delta Pi* 42(2), 72–75.
- Berliner, D. C. (2006). Our impoverished view of educational research. *Teachers' College Record* 108(6), 949–995.
- Brazell, W. (2005). Handheld computers: A boon for principals. *Principal* 84(3), 48.
- Camahan, C. R. (2006). Photovoice: Engaging children with autism and their teachers. *Teaching Exceptional Children* 39(2), 44–50.
- Collins, J. (2001). Using the Internet as a distance learning tool and selected secondary school areas. *Journal of Research on Computing in Education* 33(4), 431–455.
- Copeland, L. L. and Gray, T. C. (2002). Developing Maryland's technology education leaders for the 21st century: Technology educational leadership project. *Journal of Industrial Teacher Education* 39(3), 104–121.
- Davies, B. (2002). Rethinking schools and school leadership for the 21st century: Changes and challenges. *International Journal of Educational Management* 16(4), 196–206.
- Donlevy, J. (2004). Preparing future educational leaders: Technology standards for school administrators. *International Journal of Instructional Media* 31(3), 213–217.
- Elmborg, J. (2006). Critical information literacy: Implications for instructional practice. *Journal of Academic Librarianship* 32(2), 192–199.
- Gaudelli, W. (2006). Convergence of technology and diversity: Experiences of two beginning teachers in web-based distance learning for global/multicultural education. *Teacher Education Quarterly* 33(1), 97–116.
- Geer, C. (2002). Technology training for school administrators: A real-world approach. *TechTrends* 46(6), 56–59.
- Hartnell-Young, E. (2006). Teachers' roles in professional learning and communities of practice supported by technology in schools. *Journal of Technology in Teacher Education* 14(3), 461–480.
- Hunter, B. (2001). Against the odds: Professional development and innovation under less-than-ideal conditions. *Journal of Technology and Teacher Education* 9(4), 473–496.
- Jeffs, T., Behrmann, M., and Bannan-Ritland, B. (2006). Assistive technology and literacy learning: Reflections of parents and children. *Journal of Special Education Technology* 21(1), 37–44.
- Johnson, D. (2006). Skills for the knowledge worker. *Teacher Librarian* 34(1), 8–13.
- Killion, J. (2002). Loading e-learning shopping cart. *Journal of Staff Development* 23(1), 12–16.
- Krug, D. H. (2004). Leadership and research: Reimagining electronic technologies for supporting learning through the visual arts. *Studies in Art Education* 46(1), 3–5.
- Krumsvik, R. (2005). ICT and community of practice. *Scandinavian Journal of Educational Research* 49(1), 27–50.
- Lai, K. W. and Pratt, K. (2004). Information and communication technology in secondary schools: The role of the computer coordinator. *British Journal of Educational Technology* 35(4), 461–475.
- Lamb, A. and Johnson, L. (2006). Want to be my "friend"? What you need to know about social technologies. *Teacher Librarian* 34(1), 55–57.
- Lepard, D. H. (2002). Using peers and technology to strengthen leadership. *Journal of Personnel Evaluation in Education* 16(1), 11–28.
- Maxwell, L. A. (2006). Web systems help schools screen visitors. *Education Week* 25(32), 5.
- McGrail, E. (2006). "It's a double-edged sword, this technology business": Secondary English teachers' perspectives on a schoolwide laptop technology initiative. *Teachers College Record* 108(6), 1055–1079.
- Quinn, D. M. (2003). Legal issues in educational technology: Implications for school leaders. *Educational Administration Quarterly* 39(2), 187–207.
- Robertson, J. M. and Webber, C. F. (2002). Boundary-breaking leadership: A must for tomorrow's learning communities. In Leithwood, K. and Hallinger, P. (eds.) *Second International Handbook of Educational Leadership and Administration*, pp 519–556. Dordrecht: Kluwer Academic Publishers.
- Robertson, J. M. and Webber, C. F. (2004). International leadership development through web-based learning. *International Electronic Journal for Leadership in Learning* 8, 12. <http://www.ucalgary.ca/iejll/vol8/Robertson>.
- Shariff, S. (2004). Keeping schools out of court: Legally defensible models of leadership. *Educational Forum* 68(3), 222–233.
- Sivan, Y. Y. (2000). Patterns of leadership in virtual professional communities. *Journal of Technology and Teacher Education* 8(1), 53–67.
- Staples, A., Pubach, M. C., and Himes, D. (2005). Rethinking the technology integration challenge: Cases from three urban elementary schools. *Journal of Research on Technology Education* 37(3), 285–311.
- Tallent-Runnels, M. K., Thomas, J. A., Ahern, T. C., Shaw, S. M., and Liu, X. (2006). Teaching courses online: A review of the research. *Review of Educational Research* 76(1), 93–135.
- Technology Standards for School Administrators Collaborative (2001). *Technology Standards for School Administrators*. Washington, DC: Author.
- Webber, C. F. (2003a). New technologies and educative leadership. *Journal of Educational Administration* 41(2), 119–123.

- Webber, C. F. (2003b). Technology-mediated leadership development networks: Expanding educative possibilities. *Journal of Educational Administration* **41**(2), 201–218.
- Webber, C. F. (2005a). Entrepreneurship and educational leadership development: A Canadian perspective. *Paper Presented to the Commonwealth Council for Educational Administration and Management Society Symposium Held during the Annual Meeting of the American Educational Research Association*. April 13, 2005, Montréal, Québec, Canada. .
- Webber, C. F. (2005b). Reshaping leadership development through distance doctoral education. *Educational Research and Perspectives* **32**(2), 142–164.
- Webber, C. F. and Clark, V. B. (2006). Exploring new dimensions of leadership development. *Leading and Managing* **12**(1), 31–47.
- Webber, C. F. and Mulford, B. (2007). School and community. In Burger, J., Webber, C.F., and Klinck, P. (eds.) *Intelligent Leadership: Constructs for Thinking Education Leaders*, pp 119–142. Dordrecht, The Netherlands: Springer.
- Webber, C. F. and Robertson, J. M. (2004). Internationalization and educators' understanding of issues in educational leadership. *Educational Forum* **68**(3), 264–275.
- McPherson, S., Wizer, D. R., and Pierrel, E. (2006). Technology academies: A professional development model for technology integration leaders. *Learning and Leading with Technology* **33**(5), 26–31.
- Pardini, P. (2002). Inside the wired district. *Journal of Staff Development* **23**(1), 18–22.
- Rakes, G. C., Fields, V. S., and Cox, K. E. (2006). The influence of teachers' technology use on instructional practices. *Journal of Research on Technology in Education* **38**(4), 409–424.
- Rowley, J., Dysard, G., and Arnold, J. (2005). Developing a new technology infusion program for preparing tomorrow's teachers. *Journal of Technology and Teacher Education* **13**(1), 105–123.
- Starratt, R. J. (2005). Responsible leadership. *Educational Forum* **69**(2), 124–133.
- Woelfel, K. D., Murray, K., and Hambright, A. (2004). Making sense of technology in educational leadership programs. *TechTrends* **48**(5), 29–33.
- Zirkel, P. A. (2006). Paralyzing fear? Avoiding distorted assessments of the effect of law on education. *Journal of Law and Education* **35**(4), 461–495.

Further Reading

- Armstrong, S. and Warlick, D. (2004). The new literacy: The 3Rs evolve into the 4Es. *Technology and Learning* **25**(2), 20–28.
- Fletcher, G. (2003). Making sense of NCLB's technology component. *T.H.E. Journal* **30**(7), 56.
- Lajeane, G. T. and Knezek, D. G. (2002). Standards for technology-supported learning environments. *State Education Standards* **3**(1), 14–20.

Relevant Websites

- <http://aera-cr.asu.edu> – American Educational Research Association SIG Communication of Research.
- <http://caret.iste.org> – Centre for Applied Research in Educational Technology.
- <http://www.iste.org> – International Society for Technology in Education.
- <http://pkp.sfu.ca> – Public Knowledge Project.
- <http://www.ucalgary.ca/iejll> – University of Calgary, International Electronic Journal for Leadership in Learning.

Leadership in the Implementation of Innovations

I E Bogotch, Florida Atlantic University, Boca Raton, FL, USA

T Townsend, University of Glasgow, Glasgow, UK

M Acker-Hocevar, Washington State University, Pullman, WA, USA

© 2010 Elsevier Ltd. All rights reserved.

A Brief History of Change Theory

The history of organizational change can be traced to both sides of the Atlantic Ocean. In fact, as far back as 1920, the Tavistock Institute of Human Relations in London used concepts taken from psychoanalytical theory in order to better understand group dynamics. While the professional literature of organizational change did not emerge as an interdisciplinary field of study until the 1960s, the social, political, psychological, and educational ideas of B. F. Skinner (i.e., behaviorism), John Dewey (i.e., experimentalism), and Kurt Lewin (i.e., social theories) could already be observed in educational research and practices. The latter's works influenced the next generation of US researchers, Rogers (1962, 1965), Miles (1964, 1965), Clark and Guba (1965), Haverlock and Benne (1965), and Bennis *et al.* (1969), among others – whose ideas on planned change preceded implementation theory (Berman, 1981).

Internationally, Walker's (1970) *Theory and Practice in Educational Administration* became a landmark argument for the development of a theory as well as improving the practice of educational administration in Australia, but also throughout the British Commonwealth. Other key international contributors to the development of an understanding of the role of the educational leader as an agent of change and innovation included Ray Bolam in the United Kingdom (Bolam and Pratt, 1976) whose work specifically addressed the issue of innovation in schools and Thomas B. Greenfield in Canada (Greenfield, 1974), who suggested that human organizations, rather than having a natural order, were characterized by complexity, uncertainty, and the vagaries of human behavior. In other words, the field of administration, almost a decade prior to Fullan's conclusion, recognized and accepted that all aspects of change were in part deliberate while at the same time emergent or resulting from human interactions (Greenfield, 1974). Organizations began to experiment with new research and development centers, sending workers for external training laboratories, as new academic fields of study from organizational sociology to leadership development emerged. Scholars such as Argyris (1982) and Argyris and Schon (1978) began to pose questions about whether organizations could learn from their actions. Thus began the systematic study of organizational and managerial changes and innovation grounded in theories of action and use.

The 1970s and 1980s saw a focus on empirical change study research evolving from US national/federal innovative efforts (see Rand Corporation Study research by Berman and McLaughlin, 1979; McLaughlin, 1990) and local school-based efforts that promoted new concepts and designs of professional development (i.e., Fullan and Pomfret, 1977; Fullan, 1982). During this same period, within Europe, Ray Bolam, Per Dalin, Michael Huberman, and others—all acknowledged by Fullan (1982) influenced the direction of research. It is interesting to note that in the 1980s and 1990s, when discussions of the changing nature of leadership were at their height, the most innovative activities that emerged were driven from the top down. During this period, the intervention of government into the knowledge and management issues surrounding schools led to perhaps the greatest period of change in school history. Within half a decade, the Thatcher government in England (local management of schools), the Lange government in New Zealand (tomorrow's schools), and the Kennett government in Victoria, Australia (schools of the future), had all instituted radical forms of self-managing schools (see Caldwell and Spinks, 1988, 1992, 1998), that rapidly became a model for a structural change in education that swept the world, with charter schools in the USA and various levels and forms of decentralization in many countries of the world. This was essentially a political activity, fueled by the school effectiveness research which strove to show that schools could make a difference, but was often misinterpreted by government bodies as only schools can make the difference (Reynolds, 1994). Although context issues have been accepted and are now being studied by those involved in the school effectiveness and school improvement research to the point where the definition of effectiveness might need further interpretation (see Townsend, 2007), governments remain attached to the simplistic views of the concept, which led to the school indicator research (Kochan, 2007), to guide the accountability systems in place. Fuhrman (2003) argued that accountability systems were school indicator systems with sanctions.

The rival tensions of top-down and bottom-up reform offered two completely different roles for school leaders. With respect to government intervention, the school leader became the implementer of decisions made by others, whereas shared decision making, school-based management, and self-managing schools all looked to leadership

to develop the capacity of others. These dual roles for the school leader involved implementing innovations dictated from above (e.g., new school structures, new funding mechanisms, new administrative arrangements, and new curriculum and assessment) to respond to the new accountability regime, versus implementing local innovations, with the teachers, students, parents, and local community that addressed the needs of the students in the schools they led. Sergiovanni (1989, p. 39) approached this duality as follows:

Because of the unpredictability of the world and the limits of human rationality, it makes sense to emphasize building capabilities of people first, encouraging them to develop ways and means for using their capabilities, than it does to develop plans and then seek the know-how and commitment to implement the plans.

Barber (2002, cited in Caldwell, 2006) labeled the above tensions as uninformed prescription in the 1980s which became “informed prescription” in the 2000s (pp. 50–51). Glennan *et al.* (2004) took a more holistic approach. They argued, “an important part of the principal’s role as school improvement leader is the active management of change from above (e.g., mandates from the district office) and below (requests or demands from teachers)” (p. 20). Their work suggests that leaders’ dual roles require seeking resources and support from district staff, while simultaneously fostering and monitoring teachers’ implementation of reform components within the classroom. Their research indicates a “strong, positive relationship between the quality of principals’ leadership and the extent to which schools implemented whole-school designs” (p. 20). These tensions across roles, rules, and relationships inside of educational institutions and policy debates suggested that educational leadership and educational innovation are to be considered as a unique context both socially and politically.

If so, then exercising educational leadership becomes more than a deliberate, cognitive act; rather, educational leadership is an act that requires courage, or the ability of the leader to act differently, if not disruptively (Schlechty, 2007). That is, “the kinds of innovations required [inside of educational institutions] will likely exceed the present social systems’ capacity to sustain them. These changes will necessarily be disruptive...” (p. 222). One way to begin to do this is to stimulate a particular kind of dialog (Brown *et al.*, 2005), one that opens up the possibility of seeing the world through the collective wisdom of all participants and shares what they already know. Dalin (1996), a scholar from Norway, asked educators pointedly whether they believed they had the capacity to create the secondary schools required to respond to the external changes that marked our entry into the twenty-first century. Simply put, Dalin inquired, “Can schools learn?” and “What will it take?” Furthermore, he challenged educators to examine how their deeply held assumptions about, “Who is the teacher? Who is the student? Where is

the school? And, what is the curriculum?” that may serve to hold them back or move them forward and into the future.

To get to the heart of the fundamental questions posed by Dalin, change theories which focus on a either a bottom-up or top-down approach may limit our leaders’ abilities to ask the type of provocative questions that enable them to find fresh solutions that lead to creative and innovative responses. Hargreaves and Fink (2005) suggested that educators must make deep structural changes to schooling that impact current practices. These deep changes require examining how we organize ourselves in schools for teaching, learning, and leading. If we continue to frame change from traditional approaches of top-down reform, and do not ask the tough questions, then, we continue to control and confine information, actions, and ideas tightly, within bureaucratic and hierarchical organizational structures that limit innovative responses. Likewise, bottom-up reform may, too, be more of a reaction to the hierarchical pressures from above than offering thoughtful responses and innovative responses; in other words, we may find ourselves in a no-win situation unless we challenge existing practices and power relationships (Schlechty, 2007; Snyder *et al.*, 2000).

Changes that emanate from outside the school system rest entirely on the shoulders of the leadership within schools to implement successfully, often without funding and the needed discretion to hire the right personnel. Frequently, mandated changes trump how leaders can spend their time to listen to the needs both inside their schools and outside their schools (Bogotch and Brooks, 1994). At odds, these traditional approaches to change have locked leaders into sidestepping how to respond to unplanned events – or emergence – where required actions which are not mandated take less precedence and can easily be ignored. These evolving events, which cannot be controlled or predicted, and may, therefore, be disregarded by educators, may furtheracerbate their being out of step with the realities of the world. Said differently, the institutional contexts of schooling become the educators’ total view of reality, shutting out other perspectives (Bogotch *et al.*, 1996).

Taking action based solely on political expediency rarely leads to innovation and deep change (Bogotch *et al.*, 1995). By operating within boundaries of shared values (Barth, 1990), conceptualizing organizational change as a collective pursuit, the paradigm shifts away from bureaucratic and control mechanisms to emergence and adaptation (Lincoln, 1985) that positions change theory not only within the new sciences, but also within the new meanings of educational theory (see Biesta, 2006; Blake *et al.*, 1998). Staying within the traditions of hierarchical and bureaucratic structures may well limit innovation into the future, reducing actions that serve only to mollify the powers that be. Innovation requires more response-ableness and a collective vision (Hargreaves and Fink, 2006).

Complex Systems' Frameworks

In contrast to contextual and institutional realities, Maguire *et al.* (2006), in their comprehensive review of the literature on complexity science and organizational studies, asserted:

Peters (1987) argues that organizations should be 'thriving on chaos'; and Nonaka (1988a: 72) maintains that that 'the self-renewal strategy of an organization lies in its ability to manage continuous dissolution and creation of organizational order' (p. 173).

Chaos and complexity theories, derived from these new sciences such as quantum physics and evolutionary biology, offer us a fresh lens to view change as adaptation and to view unplanned change – that is, emergence – as learning through actions taken and reflected upon. Importantly, relationships form primary building blocks to construct new ways of responding to these unanticipated changes (Wheatley, 1992). Leaders' work, therefore, calls for ways to build the needed connections with others that engenders trust in their abilities to work together to address issues that arise (Argyris and Schon, 1974).

Specifically, positioning leadership within a theory of chaos and complexity theories views leadership as neither top down or bottom up. Rather, the intersection of leadership, implementation, and innovations becomes grounded in the leader's ability to forge strong connections with others, to establish and promote listening systems, to engender trust that promotes meaningful dialog that leads to creative solutions. What is often left unresolved, however, is the dichotomy between forging common and shared understandings and collective agency and action within shared meanings versus the acceptance of "intrinsic but complementary differences" (Gould, 2003, p. 247).

Serious attention to all members of the set may well unify our mental lives by forging a consensus on values and results. However, such a consensus could only emerge from independent contributions, knitted together by serious and generous dialogue among truly different, and equally valid, ways of knowing, each responsible for a swatch on wisdom's quilt, with the swatches abutting and interfingering in gorgeously complex patterns of interaction (pp. 255–256).

The alternative to the binary top-down and bottom-up perspective was succinctly captured by Stephen Jay Gould (2003) who introduced the term consilience, literally meaning jumping together (pp. 247–260). Gould explained how distinct worlds can contribute to building new theories based on differences, not reductionism in policy and practice. The logical and the intuitive are allowed jump together as equally valid and necessary approaches to knowing and understanding and acting

upon what is actually going on. We can observe this line of research in Hargreaves and Fink (2004) who have attempted to describe their findings outside the binary opposites and conjunctions previously used by Fullan (1982). For example, to understand sustainability with respect to planning and preparing for leadership succession, the authors are mindful of maximizing a leader's ability to remain in schools so long as change is evident. Here, the authors recommend moving from the leader as an individual to distributive models of leadership so that continuity does not rest with any one individual or end with succession. In fact, all of their findings have that jumping together quality: shared vision, linking school and community leaders to ideas of social justice, obtaining both intrinsic rewards and extrinsic incentives, promoting diversity and variation, building strategic alliances, and establishing system supports.

These considerations might lead to where "the distinction between leaders and followers blurs," where although "school principals might still have overall responsibility for the ongoing work of the school, teachers, students, and parents all play leadership roles in terms of making decisions, acting on decisions made and taking responsibility" that requires the school to "change its focus from one where teachers teach, students learn and parents support, to one where everyone in the school contributes what they know (teach), and works towards common goals (support), one of which has to be the ongoing improvement of what the school does (learns)" (Townsend and Otero, 1999: 57).

Whether these principles of sustainability or the many choice options in and surrounding education, such as institutionalized schooling, homeschooling, deschooling (Illich, 1971), or notions of a learning organization (Senge, 1990), who built on and extended ideas of scholars such as Argyris and Schon (1978) and Emery and Trist (1965) (see also Silins and Mulford, in press) can result in complex relational and dialogic patterns of knowledge is both under-theorized and understudied. One futuristic research question to emerge from this line of thinking is whether national and local educational systems can empirically embrace the Wiki philosophy of editing and re-editing already existing content knowledge?

Although many studies substantiate how deep changes in a school's culture necessitated educators working collaboratively, what is often lost is how collaborations enable participants to co-construct and co-evolve their futures together – futures that impact the achievement patterns of students and permit educators to respond positively to changes in their distinct local and national environments. Looking at educational leadership and change through the lenses of chaos and complexity theories requires that we study different communities of practices (Wegner, 1998; Wegner *et al.*, 2002) that support formal and informal knowledge networks.

Re-Conceptualizing Educational Innovation

“Innovation is a noble pursuit” (Caldwell, 2006: 44) in the sense that educational reformers have tried to position public education as a fundamental human right. Unfortunately, education does not necessarily hold this position of public legitimacy in all societies of the world; moreover, educational innovations are not always linked to the primary purposes of education, but rather have been viewed in a limited way to the core technologies of teaching and learning. Thus, new school structures, new instructional methods and technologies, new funding mechanisms, new administrative arrangements, new curriculum and assessment, etc. connect to the practices of educators – but only contingently and on the margins. What is missing, therefore, for theorists such as Fullan and Goodlad (2004) among others, is the deliberateness of educational innovations as necessary for the purposes of public education as complex systems. What we mean here is that it is important to reposition the world of education as fundamental and also distinct from other institutionalized worlds, including the world of business.

The ethic of business is hard work, productivity, market competition, and profit (Drucker, 1985; Kouzes and Posner, 1991). Repeatedly, Drucker (1985) distinguishes a successful innovation from a bright idea. “I know of not one such ‘flash of genius’ that turned into an innovation. They all remained brilliant ideas” (p. 133). That is, innovations have to work in the present with people who can understand them. To the extent that the social system becomes relevant and aligned with the economic needs of people, innovation has to do with instrumental motivation and personnel issues in terms of keeping workers satisfied and focused on the job (Kouzes and Posner, 1991).

In contrast, innovations in education are inextricable not only from sociopolitical systems (House, 1981), present structures (Schlechty, 2007), and individual needs, but they are also relevant to the past, present, and future uses of knowledge, which is “something that hasn’t happened yet, but which [educators] are looking forward to” (Townsend, 2003: 17). Miles (1964) asserted that innovation was “*always* operant in relation to a given social system” (p. 14). Gittell and Hollander (1968) conceptually placed innovations at the center of education because it is “‘always present in some degree in all school systems’ (p. 3) even though its meaning and significance may be minimal” (cited in Bogotch *et al.*, 1995: 7). Thus, Bogotch *et al.* (1995) argued that “the concept of innovation is fundamental to social organizations such as schools. Innovations are not just a matter of structural rearrangements or technical advances, but rather are a new foci on social dynamics, management and people” (pp. 6–7). The argument is that educational innovations must be fundamental

to the many sociopolitical systems not just instrumentally used to crush the competition – as in business.

Beare (2006) recently advanced the argument beyond older systems’ thinking:

The leaders we admire are. . . frequently impatient with the merely functional, the technical, the tidy. They seem to be discomfited with what is, driven by a nagging discontent, and consumed by a compulsion to revise, remodel, and improve things. They have the explosive combination of insatiable restlessness about what is orthodox, conventional and acceptable, together with the ability to handle and invent ideas and to synthesize them in a new way which makes them coherent and persuasive for others. In doing so, they break open the boundaries (p. 8).

However, what makes Beare’s description somewhat problematic with respect to innovations, educational or otherwise, is that empirically, the attributes of innovations influence their adoption and acceptance (Rogers, 1962, 1995). Rogers identified four attributes of relative advantage of an innovation: compatibility, trialability, observability, and less complexity. The compatibility of the innovation refers to the degree it is consistent with existing values, past experiences, and the needs of the potential adopters. The complexity of the innovation is the degree to which it can be readily understood by most members of a social system. The innovation’s trialability refers to the degree to which it can be experienced with on a limited basis. The observability of an innovation refers to the degree to which its results are visible to others. In each of these attributes, the tensions of practicality, institutionalism, and change emerge – which are as relevant inside of rational bureaucracies as they are inside of complex systems. According to Diamond (2006), the perceived attributes of an innovation and its characteristics help explain different rates of adoption. The innovation’s *relative advantage* is the degree to which an innovation is perceived to be better than the idea it supersedes. What is then needed is a discussion of how educational leadership can respond, socially and institutionally, to these influences and constraints in order to create new ways of learning.

Beyond Societal and Institutional Barriers

Grounded in the ideas presented here, we can summarize the many sociopolitical challenges facing educational leaders in the twenty-first century:

1. To re-establish the international legitimacy and social capital of public education as a basic right (i.e., on principle) as well as reflecting societal goods (i.e., as different cultural and societal values).
2. To re-assert the capacity of educators to not only manage educational systems, but also renew and

co-construct complex systems with new patterns of knowledge diffusion and dissemination; when innovations are not institutionalized, the external mandates and educational leadership (see Cuban and Useem, 2003) work against genuine accountability, educational responsibility (Biesta, 2006), trust (Argyris and Schon, 1974), and innovation.

3. To develop theories of educational action and innovations that jump together the tensions between deliberate efforts and emergent unpredictable happenings; school development should be seen as ongoing changes that promote innovation through social networks and building new *Wiki*-like dialogs that challenge knowledge, positional hierarchical roles, and relationships.
4. To reconnect the purposes and processes of public education (not just schooling) to the needs and desires of citizens regardless of the different and disparate human conditions. That is, leadership and education are valid in all contexts.
5. To sustain leadership and innovations beyond a single individual, single location in time and space. Here, we note the limits of the leader's ability to control, plan, and predict human behaviors. Rather, leaders must rely upon systems of values and beliefs that underpin actions and create social networks of learning, for others to interact, and share information and power.

The lens of complex theories is thus necessary to go beyond societal and institutional barriers that have limited educational theory, the educators' capacity for learning and change, the educational systems' willingness to learn and transform with social, political, and economic realities. While the terms we chose to use here differ from the vocabulary of planned change, implementation theory, the barriers imposed externally onto education through government authorities continue to limit educational leaderships' ability to make changes for the right reasons – namely social capital and the public good.

The Progress of Educational Change: A New Integration

We have argued for the uniqueness of education institutionally and socially and therefore, we further assert that educational leadership and innovation demands a new theory. In *Re-imagining Educational Leadership*, Caldwell (2006), following Drucker (1999), sees the need to balance innovation with abandonment (p. 117). In other words, we must be cognizant of history as we think and act in the present. Therefore, without denying past successes, it is important to continuously start afresh, not unlike the newly hired CEO of Toys R Us who talks of transforming the traditional notion of toys to a world in which children (of whatever age!) play with gadgets. The distinction we

see here is not the traditional toys versus gadgets or toys and gadgets, but rather the jumping together of toys/gadgets. The point we are making is that we must in our educational theories and practices go beyond binary opposites (i.e., transitioning from one world view to another by abandoning the past), and going beyond compound conjunctions of two ideas without seeing the necessity for changes in both of the original ideas. We refer to this phenomenon in terms of wiki-like editing of existing knowledge. With respect to leadership and change, however, the argument suggests that we engage in mobilizing change despite the fact that existing models cannot provide us with definitive answers. Thus, we turn to the critical importance of intuition in real time by paying active attention to our surroundings for answers that cannot be derived from rational mindsets alone. (We are indebted to our Florida Atlantic University colleague, Mr. John Hardman for this insight.)

In **Table 1**, we have reconfigured Fullan's 1982 conclusions taken from *The Meaning of Educational Change*. Fullan, to whom we all owe much in terms of our ideas on change, divided his conclusions into two categories – transitions, that is, moving from x to y (column 1) and co-joining, that is, adding x to y (column 3). We have inserted in between our own two columns where we have re-conceptualized these categories in terms of Gould's notion of consilience, jumping together, or x/y (columns 2 and 4).

Thus, the transition from cognitive to social development no longer covers the range of possibilities in terms of teaching, learning, or leading. The developmental stages and grade and age-level configurations evolve culturally and behaviorally with demonstrated changes in the brain. The transitions of alignment of curriculum and processes as well as the privatism of learning – so necessary for routine matters and benchmarks, no longer spans the range of knowledge and skills needed by children, adults, and educators around the world today. The role of professional development in education still stands as one of Michael Fullan's major contributions to understanding education and change. We must now, however, take these notions further to collective learning and wisdom into developing social networks and communities of learning and practices that reflect diversity and differences. Such changes demand new skills and understandings that can make change possible across different settings. Finally, while we might take issue with the dichotomy of grandeur and incremental changes, we, too, see the need for making small deliberate changes inside of institutional routines, roles, and responsibilities that promote time, leadership, and meaning into the lives of educators, as teachers, leaders, and students. Only then, in the present moment, can we demonstrate to those who study schools and educational relationships that our practices in fact reflect human dignity and the capacity for us to learn, grow, and change.

Table 1 A reconceptualization of Fullan's categories using 'consilience'

<i>Fullan's binary opposites (transitions)</i>	<i>Jumping together</i>	<i>Fullan's compound conjunctions</i>	<i>Jumping together</i>
Cognitive vs. social development goals	The whole child/the whole person	Time and change	Knowing the past/acting in the present/unafraid of the future
Fidelity vs. variation	Human differences/uniqueness/human interactions/uniqueness	Leadership and change	Capacity building for self and other/substantive changes within a single generation
Privatism vs. professional development	Self/other/culture/social capital and the public good	Meaning and change	Everyday procedural justice in schools through asking, questioning, testing assumptions, experimenting on knowledge content, and critiquing educational practices
Specific vs. generic capacity for change	Student/teacher/administrator as learner/educator/leader	School and society	Combating injustices through education practices
Grandeur vs. incrementalism	Big bright idea/workable little idea		

Conclusion

Following futurists such as Toffler, Caldwell (2006) sees accelerated changes for the world's schools (p. 192). Consistent with our own thesis, Caldwell does not see this as the future of top-down and bottom-up logics, but rather, as a future of synergies, social networks, technology, and learning in real time. Many of the binaries that Fullan conceptualized in 1982 have been transformed; others have been reified in practice. We again ask, why? The answer for us cannot be found in the effectiveness or efficiency of linear models of diffusion or dissemination as rational systems or as scaling up of reforms. Rather, the challenges we face are located in ourselves as educators and our ability to develop the intellectual capacity to become educational leaders for ourselves and for others across societies.

In part what is given in teaching, in the initiation into a culture, is a gift that cannot be refused. What we come to know in this way precedes the possibility of our autonomy. Closer to us than we can readily acknowledge, it is familiar, occasioning a kind of love, as perhaps for things close to the heart (Blake *et al.*, 1998: 88).

The future of educational leadership implementation of innovations must remain largely unknown if it is to advance nations beyond their societal and institutional barriers. Not only must the results be different for students and citizens, but also must be the paths taken by educational leaders.

See also: Curriculum and Complex Systems Theory; Curriculum Planning and Systems Change; School Reform and Restructuring: Self Managing School; System Leadership.

Bibliography

- Argyris, C. (1982). *Reasoning, Learning, and Action: Individual and Organizational*. San Francisco, CA: Jossey-Bass.
- Argyris, C. and Schon, D. (1974). *Theory in Practice: Increasing Professional Effectiveness*. San Francisco, CA: Jossey-Bass.
- Argyris, C. and Schon, D. (1978). *Organizational Learning: A Theory of Action Perspective*. Reading, PA: Addison Wesley.
- Barth, R. (1990). *Improving Schools from within*. San Francisco, CA: Jossey-Bass.
- Beare, H. (2006). *Leadership for a New Millennium. ACEL Monograph Series, No. 38*. Melbourne, VIC: Australian Council for Educational Leadership.
- Bennis, W., Benne, K., and Chin, R. (1969/1985). *The Planning of Change*, 4th edn. New York: Holt, Rinehart, and Winston.
- Berman, P. (1981). Educational change: An implementation paradigm. In Lehming, R. and Kane, M. (eds.) *Improving Schools: Using What We Know*, pp 253–287. Beverly Hills, CA: Sage.
- Berman, P. and McLaughlin, M. (1979). *An Exploratory Study of School District Adaptation*. Santa Monica, CA: Rand Corporation.
- Biesta, G. (2006). *Beyond Learning: Democratic Education for a Human Future*. Brookline, MA: Paradigm.
- Blake, N., Smeyers, P., Smith, R., and Standish, P. (1998). *Thinking Again: Education after Postmodernism*. Westport, CT: Bergin and Garve.
- Bogotch, I. and Brooks, C. (1994). Linking school level innovations with an urban school district's central office. *Journal of School Leadership* 4(1), 12–27.
- Bogotch, I., Brooks, C., MacPhee, B., and Riedlinger, B. (1995). An urban district's knowledge and attitudes toward school-based innovation. *Urban Education* 30(1), 5–26.
- Bogotch, I. E., Gahr, J. A., and Rougelot, S. B., 3rd (1996). Practitioner knowledge of personnel functions: A comparative analysis. *Journal of Personal Evaluation in Education* 10(3), 227–246.
- Bolam, R. and Pratt, S. (1976). *Management in Education: The Management of Innovation in Schools*. Milton Keynes: Open University Press.
- Brown, J. and Isaacs, D. the World Café Community (2005). *The World Café: Shaping Our Futures through Conversations that Matter*. San Francisco, CA: Berrett-Koehler Publishers.
- Caldwell, B. (2006). *Re-Imagining Educational Leadership*. Melbourne, VIC: ACER Press.
- Caldwell, B. J. and Spinks, J. M. (1988). *The Self-Managing School*. London: Falmer Press.

- Caldwell, B. J. and Spinks, J. M. (1992). *Leading the Self-Managing School*. London: Falmer Press.
- Caldwell, B. J. and Spinks, J. M. (1998). *Beyond the Self-Managing School*. London: Falmer Press.
- Clark, D. and Guba, E. (1965). An examination of potential change roles in education. *Paper Presented at the Seminar on Innovation in Planning School Curricula*. Columbus, OH, USA.
- Cuban, L. and Useem, M. (2003). What happened in six cities? In Cuban, L. and Useem, M. (eds.) *Powerful Reforms with Shallow Roots: Improving America's Urban Schools*, pp 147–170. New York: Teacher's College Press.
- Dalin, P. (1996). Can school learn? Preparing for the 21st century. *NASSP Bulletin* **68**, 9–15.
- Diamond, D. (2006). *Attributes That Enable a Virtual High School to Go to Scale*. Unpublished dissertation proposal. Boca Raton, FL: Florida Atlantic University.
- Drucker, P. (1985). *Innovation and Entrepreneurship*. New York: Harper and Row.
- Drucker, P. F. (1999). *Leadership Challenges for the 21st Century*. Oxford: Butterworth Heinemann.
- Emery, F. E. and Trist, E. L. (1965). The causal texture of organizational environments. *Human Relations* **18**, 21–32.
- Fullan, M. (1982). *The Meaning of Educational Change*. New York: Teachers College Press.
- Fullan, M. and Pomfret, A. (1977). Research on curriculum and instruction implementation. *Review of Educational Research* **47**(1), 335–397.
- Fuhrman, S. (2003). *Redesigning Accountability Systems for Education (CPRE Policy Briefs)*. Philadelphia, PA: Consortium for Policy Research in Education (CPRE).
- Glennan, T., Bodilly, S., Galegher, J., and Kerr, K. (2004). *Expanding the Reach of Education Reforms Perspectives from Leaders in the Scale-Up of Educational Interventions*. Santa Monica, CA: Rand Corporation.
- Gittell, M. and Hollander, T. (1968). *Six Urban School Districts: Models for Analyzing Change*. New York: Praeger.
- Goodlad, J. (2004). *Romances with Schools*. New York: McGraw-Hill.
- Gould, S. J. (2003). *The Hedgehog, the Fox, and the Magister's Po*. New York: Harmony Books.
- Greenfield, T. B. (1974). The theory about organisation: A new perspective and its implications for schools. *Paper Presented at the Third International Intervisitation Programme in Bristol in 1974*.
- Hargreaves, A. and Fink, D. (2004). The seven principles of sustainable leadership. *Educational Leadership* **61**(7). Alexandria, VA: ASCD.
- Hargreaves, A. and Fink, D. (2005). *Sustainable Leadership*. San Francisco, CA: Jossey-Bass.
- Hargreaves, A. and Fink, D. (2006). Sustainable leadership and improvement. *Paper Presented at the International Congress on School Effectiveness and Improvement (ICSEI) Annual Conference*. Fort Lauderdale, FL, USA.
- Haverlock, R. and Benne, K. (1965). An exploratory study of knowledge utilization. *Paper Presented at the Seminar on Innovation in Planning School Curricula*. Columbus, OH, USA.
- House, E. (1981). Three perspectives on innovation. In Lehming, R. and Kane, M. (eds.) *Improving Schools: Using What We Know*, pp 253–287. Beverly Hills: Sage Publications.
- Illich, I. (1971). *Deschooling Society*. Harmondsworth: Penguin.
- Kochan, S. (2007). Evolution of school performance research in the USA: From school effectiveness to school accountability and back. In Townsend, T. (ed.) *The International Handbook of School Effectiveness and Improvement: Review, Reflection and Reframing*. Dordrecht: Springer.
- Kouzes, J. and Posner, B. (1991). *The Leadership Challenge*. San Francisco, CA: Jossey-Bass.
- Lincoln, Y. (ed.) (1985). *Organizational Theory and Inquiry*. Beverly Hills, CA: Sage.
- Maguire, S., McKelvey, B., Mirabeau, L., and Oztas, N. (2006). Complexity science and organizational studies. In Clegg, S., Hardy, C., Lawrence, T. B., and Nord, W. (eds.) *Handbook of Organization Studies*, 2nd edn., pp 165–214. London: Sage.
- McLaughlin, M. (1990). The rand change agent study revisited: Macro perspective's and micro realities. *Educational Researcher* **19**(9), 11–16.
- Miles, M. (1965). Some propositions in research utilization in education. *Paper Presented at the Seminar on Innovation in Planning School Curricula*. Columbus, OH, USA.
- Nonaka, I. (1988). Creating organizational order out of chaos: Self-renewal in Japanese firms. *California Management Review* **30**(3), 57–73.
- Peters, T. (1987). *Thriving on Chaos: Handbook for a Management Revolution*. New York: Alfred A. Knopf.
- Reynolds, D. (1994). School effectiveness and school improvement in the 1990s. In Ribbens, P. and Burrige, E. (eds.) *Improving Education: Promoting Quality in Schools*, pp 83–99. London: Cassell.
- Rogers, E. (1962). *Diffusion of Innovations*. New York: Free Press.
- Rogers, E. (1965). What are innovators like? In Carlson, R., et al. (eds.) *Change Processes in the Public Schools*, pp 55–64. Eugene, OR: University of Oregon.
- Schlechty, P. (2007). Understanding the normative system. In Fullan, M. (ed.) *The Jossey-Bass Reader on Educational Leadership*, 2nd edn., ch. 16, pp 221–236. San Francisco, CA: Jossey-Bass.
- Senge, P. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Doubleday/Currency.
- Sergiovanni, T. (1989). Value-driven schools: The amoeba theory. In Walberg, H. and Lane, J. (eds.) *Organizing for Learning toward the 21st Century*, pp 31–40. Reston, VA: NASSP.
- Silins, H. and Mulford, B. (in press). Leadership and school effectiveness and improvement. In Townsend, T. (ed.) *The International Handbook of School Effectiveness and Improvement: Review, Reflection and Reframing*. Dordrecht and New York: Springer.
- Snyder, K. J., Acker-Hocevar, M., and Snyder, K. M. (2000). *Living on the Edge of Chaos: Leading Schools into the Global Age*. Milwaukee, WI: Quality Press.
- Townsend, T. (2003). *Radical Intervention: Working with Those That Struggle to Learn*. Melbourne: Incorporated Association of Registered Teachers of Victoria.
- Townsend, T. (2007). School effectiveness and improvement in the 21st century: Reframing for the future. In Townsend, T. (ed.) *The International Handbook of School Effectiveness and Improvement: Review, Reflection and Reframing*, pp 933–962. Dordrecht: Springer.
- Townsend, T. and Otero, G. (1999). *The Global Classroom: Activities to Engage Students in Third Millennium Schools*. Melbourne, VIC: Hawker Brownlow Education.
- Walker, W. G. (1970). *Theory and Practice in Educational Administration*. St Lucia, QLD: University of Queensland Press.
- Wegner, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. New York: Cambridge University Press.
- Wegner, E., McDermott, R., and Snyder, W. (2002). *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Boston, MA: Harvard University Press.
- Wheatley, M. J. (1992). *Leadership and the New Science: Learning about Organization from an Orderly Universe*. San Francisco, CA: Berrett-Koehler.

Leadership: School Improvement

R H Heck, College of Education, Honolulu, HI, USA

P Hallinger, Asia Pacific Centre for Leadership and Change, Hong Kong Institute of Education, Hong Kong

© 2010 Elsevier Ltd. All rights reserved.

Overview of Leadership for School Improvement

School change and improvement have been well studied over the past several decades (Louis *et al.*, 1999). Over time, researchers' conceptualizations of school change evolved to incorporate a variety of different lenses for understanding the role of leaders in school improvement.

Conceptualizing School Improvement

Interest began with a focus on the diffusion and adoption of educational innovations from the 1930s to the 1960s and shifted toward studying planned change efforts during the 1960s and 1970s as national governments sought to implement ambitious policy mandates, curricula, and programs in schools. During the 1980s, researchers interested in change and school improvement centered their attention on studying the internal nature of effective, innovative, and successful schools. In the 1990s, the focus of researchers shifted away from school effectiveness and toward the role of leadership in initiating and sustaining longer-term site- and classroom-level changes and improvements. Finally, most recently, during an era dominated by concerns for accountability and improved student learning, studies of school improvement have become centrally concerned with the role of leadership as the key catalyst for change.

Today, there exists a strong expectation that school leaders can and should be held accountable for improvement of the school's academic performance and improvement (Hallinger and Heck, 1996, 2002; Heck and Hallinger, 2005; Leithwood, 1994; Leithwood *et al.*, 2004; Slegers *et al.*, 2002; Southworth, 2002). As Teddlie and Reynolds (2000) concluded in an review of international studies of school effectiveness, "we do not know of a study that has not shown that leadership is important within effective schools... Indeed, 'leadership' is now centrally synonymous with school effectiveness for many, including many operating within the school improvement paradigm also" (p. 141).

Despite these broad-stroke conclusions concerning the central role of leadership in school improvement, there remain significant challenges in interpreting this literature (Louis *et al.*, 1999). Indeed, we must note at the outset of this article that despite the frequency of its use, there is no commonly accepted definition of the term school improvement. It has been defined variously as the

successful implementation of a program, changes in teacher behavior, transformation of the school's culture, an alteration of a school structure, or an increase in student learning or school effectiveness (Clark *et al.*, 1984; Firestone and Corbett, 1988; Fullan, 1982; Heck and Hallinger, 1999; Leithwood, 1994; Louis *et al.*, 1999). School improvement studies, taken as a whole, have been important in providing information about the process of improvement even if they have been less successful in documenting how improvement processes affect student outcomes (Fullan, 1991; Hall and Hord, 1987; Louis, 1994; Reynolds and Teddlie, 2000).

In addition to acknowledging the diversity of ends that define our understanding of school improvement, there is also disagreement over the means that educators may choose to facilitate improvement (Louis *et al.*, 1999). School improvement processes have been described as complex, multidimensional, and dependent on the relationship between the school, its community, and its cultural context (Hallinger, 1998; Hallinger and Kantamara, 2001; Hallinger and Leithwood, 1998; Meyer and Rowan, 1977; Sarason, 1982). Conceptualizations of school improvement, therefore, must go beyond adopting an innovation, changing a structure, copying the practices of successful schools, or implementing new management systems such as site-based management (Barth, 1986, 1990; Crandall *et al.*, 1986; Cuban, 1984; Fullan, 1992; Sarason, 1982).

School improvement efforts are, however, a study in contrasts. The past several decades have seen school improvement efforts organized around the adoption of a curricular program, the implementation of teacher development programs, planned school improvement and school development programs, development of schools as learning organizations, and whole-school adoption of an organized set of teaching and learning practices. Despite this variety in approaches, a consistent theme emerging from research across this domain over a period of decades is the importance of leadership in facilitating improvement efforts (Berman and McLaughlin, 1978; Firestone and Corbett, 1988; Fullan, 1982, 2000, 2001; Fullan and Pomfret, 1977; Hall and Hord, 1987; Leighton, 1996; Leithwood, 1994; Leithwood *et al.*, 2004; Leithwood and Montgomery, 1982). School leaders play a central role in initiating internal changes in schools, providing direction and support, and sustaining those changes over time by linking the internal and external environments of the school (Firestone and Louis, 1999; Fullan, 2001; Meyer

and Rowan, 1977). Indeed, based upon experience and research of the past five decades, it is possible to conclude that sustained school improvement in the absence of evidence of leadership is a rarity.

Conceptualizing Leadership

For the purpose of examining leadership and school improvement, we define leadership as a process of influence by which leaders (i.e., school principals and others within the school) identify a direction for change, develop strategies for action (formally or informally), and facilitate the efforts of teachers and other stakeholders that lead to improvement of identified outcomes. Although improved student learning is often the outcome most salient to policymakers, a wide range of other outcomes have also been studied as outcomes of school improvement efforts. These include student engagement, attendance, and persistence; and changes in teacher behavior and successful implementation of curricular programs.

Up until the 1990s, most research conducted in this domain focused on the role of the school principal in facilitating change and improvement. More recently, however, scholars have sought to conceptualize leadership for school improvement. This has resulted in a focus on multiple roles and shared responsibility of various people in the school (Barth, 1999; Lambert, 1998, 2002; Marks and Printy, 2003) or as a characteristic of the organization (Ogawa and Bossert, 1995).

Frames for Understanding Leadership for School Improvement

A review of the empirical literature on school improvement suggests several different models have dominated theory and research on leadership for school improvement from the late 1970s through the present (e.g., Firestone and Corbett, 1988; House, 1981; Louis *et al.*, 1999; Van den berg and Vandenberghe, 1986; Van Velzen *et al.*, 1985). In the first *Handbook for Research in Educational Administration* (Boyan, 1988), Firestone and Corbett (1988) provided a review of the extant research on efforts directed at planned school change. Planned change focused on intentional systematic efforts to modify some aspect of the school's organization or practice. Consistent with House (1981), they identified three perspectives describing research on school change: rational-technical, political, and cultural.

Rational-Technical Perspectives on School Improvement

A technical-rational view dominates the literature on educational change and school improvement. There was

and continues to be a common perception – especially in the educational policy community – that school improvement derives primarily from externally imposed efforts of policymakers at higher levels of government (e.g., federal and state levels). The means used by policymakers consist of a combination of laws, policies, mandates, grants, training programs, and curriculum projects (e.g., Berman and McLaughlin, 1978). Within this model, leadership is synonymous with the school principal. As middle managers operating in a hierarchical system, they represent the staff members holding primary responsibility for implementing change efforts initiated at other governmental and organizational levels. As Firestone and Corbett (1988) summarized this view, principals were "...central to change because they are best placed structurally to establish, or at least affirm, norms for the whole school" (p. 337).

Consistent with rational-technical theories of organizations, this conceptual framework assumes that strong linkages should exist between the goals and intentions of policymakers and managers, the activities of staff, and student outcomes (Bolman and Deal, 1992; Scott, 2002). Underperforming schools have often been documented as evidencing unclear goals and a loose coupling between management and teaching and learning activities (Weick, 1976). From this perspective, school improvement consists of processes and activities designed to tighten the linkages between goals, activities, and outcomes (Murphy *et al.*, 1984, 1985).

In rational-technical models of planned change, the principal's leadership is without question the most critical lever available to policymakers interested in strengthening these linkages (Fullan, 1982, 2001; Murphy *et al.*, 1984, 1985). Principals exercise leadership for school improvement through goal setting, often incorporated into formal school-improvement programs and projects. These school-improvement processes themselves reflect a rational orientation toward organizations that entails planning, organizing, directing, coordinating, and controlling school-level activities designed to achieve specific goals and produce intended outcomes (Scott, 2002).

This rational-technical model of leadership for school improvement is perhaps best reflected in the effective schools literature which emerged during the 1980s. Findings from studies of instructionally effective schools for the urban poor were often loosely translated into prescriptions for the improvement of all schools (Brandt, 1982; Clark *et al.*, 1984; Edmonds, 1979; Levine, 1991; Mortimore, 1993; Rutter *et al.*, 1979; Stoll and Fink, 1996). A key finding from this body of related literature concerned the presence of strong principal instructional leadership in schools that were producing better-than-expected learning outcomes for traditionally underachieving student populations (Edmonds, 1979; Bossert *et al.*, 1982). This picture of strong instructional leadership contrasted sharply with prior research that found principals to

be largely disconnected from the technical core activities of teaching and learning in schools (Bossert *et al.*, 1982; Cuban, 1984). This led to prescriptions for school principals to become more actively engaged as instructional leaders. Cuban (1984) portrayed this picture of instructional leaders as hip-deep in matters of curriculum, instruction, and school improvement.

Notably, within this frame, school improvement was defined implicitly as oriented specifically toward the improvement of student learning outcomes. As the school-effectiveness studies were cross-sectional, however, this approach left open the question of whether and how instructional leaders could help schools increase their effectiveness over time through various improvement strategies (Barth, 1986; Cuban, 1984; Hallinger, 2003; Jackson, 2000; Slegers *et al.*, 2002).

A strong focus on instructional leadership continues to dominate the rational-technical approach into the present era, although with some adaptations. More recent views of instructional leadership, while accepting the importance of the principal in the process of school change, have highlighted constraints on the principal's time, educational expertise, and moral authority in assuming sole responsibility for leading school improvement (Barth, 1986, 1990; Blasé, 1993; Cuban, 1984; Hall and Hord, 1987). This recognition has led to a broadening of the instructional leadership construct to incorporate other key leaders in the school. The result has been a conceptualization of instructional leadership as a responsibility shared among by the principal with other school administrators and teachers (Barth, 1990; Hallinger, 2003, 2005; Lambert, 1998, 2002; Kleine-Kracht, 1993; Marks and Printy, 2004).

Political Perspectives on School Improvement

Yet, even as support for the importance of leadership as a key factor in school change emerged, other scholars called into question the underlying assumptions of the planned change model of school improvement (e.g., Barth, 1986; Berman and McLaughlin, 1978; Cuban, 1984, 1990; Fullan and Pomfret, 1977; Sarason, 1982; Weick, 1976). For example, the Rand studies during the 1970s highlighted the difficulty of actually implementing reforms to change educational processes across levels of educational systems. This research found that features of the local district and school site contexts moderated the implementation process to determine the outcomes of school-reform efforts (Berman and McLaughlin, 1978). As McLaughlin (1990) and Cuban (1990) later suggested, the nature, amount, and pace of change at the individual school is a product of local factors that are often out of the control of higher-level policymakers.

Political values and interests at the local level often conflict with the underlying purposes of reform mandates imposed from higher levels of the system. Schools are

characterized by multiple goals, diverse instructional strategies, and a relatively high degree of teacher autonomy (Barth, 1986, 1990; Cuban, 1990; Weick, 1976). This results in informal negotiation over the goals and strategies of school improvement. Based on these observations of school improvement in practice, researchers reconceptualized school improvement as a process of mutual adaptation, rather than as one of rational-technical imposition of change (Berman and McLaughlin, 1978).

Subsequent scholarship highlighted the role of micro-political processes within the school system as a key dimension influencing school improvement. This research recast the role of school leaders in school improvement from that of instructional or transformational leaders to that of negotiators, blending the needs and interests of various stakeholders (e.g., principals, teachers, parents, and the wider community) to determine goals and strategies and to allocate resources (Blasé, 1993; Bolman and Deal, 1992). The political frame challenged the dominant view of school improvement as a rational, top-down, goal-oriented process. It refocused attention on the role of leaders in interpreting local values endorsed by the school community and in working collaboratively with teachers (Blasé, 1993; Ogawa and Bossert, 1995).

We would also include critical and feminist perspectives on leadership within this framework on school improvement. These approaches again question the underlying assumption of school improvement as a goal-oriented process directed toward student achievement. Instead, these perspectives refocus attention on how well-intentioned school-improvement efforts can reinforce social inequities at the school level (Anderson, 1990). They examine the means through which dominant social values and shared meanings about unjust school processes (e.g., student grouping and access to curriculum) develop and seek to document the outcomes. They often involve a critique of existing social relationships and advancement toward desired ones (Keith, 1996).

One role of school administrators has been to legitimize their functions as allocators of social values, often consciously or unconsciously institutionalizing societal inequities. These perspectives reframe the role of leaders as questioning the underlying purposes of schooling and refocusing the efforts of schools into increasing social justice. Although these perspectives help us see school leadership in facilitating school improvement in more expansive and diverse ways, they have been less influential to date in terms of changing school-improvement practice.

Cultural Perspectives on School Improvement

Like the political frame, cultural-institutional approaches question the underlying assumptions of rational-technical models of school improvement (Bolman and Deal, 1992; Cuban, 1990; Deal, 1990; Sarason, 1990). Researchers

employing these alternative lenses again have remarked on the observed gap between intentions and results of large-scale reform efforts. Sarason (1982) characterized this perspective as “the culture of the school and the problem of change.” Numerous features of the local school culture – its decision-making structure, predominant values, past experience with change, administrative support, program focus, and other normative practices of the principal, teachers, and students – affect the implementation of improvement programs and activities.

From this perspective, improvement is defined in terms of reshaping normative structures within schools so that change can take root. Cultural perspectives on school change, view school improvement as a process of capacity building and growth rather than program implementation (Fullan, 2002, 2003; Stoll and Fink, 1996). In contrast to the rational–technical frame which has tended to emphasize instructional leadership, the dominant leadership perspective within the cultural frame has been transformational leadership (Leithwood, 1994; Leithwood and Jantzi, 1999, 2000).

The transformational model of school leadership (or models that focused on leadership as an organizational quality) was often associated with efforts to change the school’s culture as a means of changing its outcomes (e.g., Leithwood, 1994; Leithwood *et al.*, 1993; Mulford and Bishop, 2003; Pounder *et al.*, 1995; Rowan and Denk, 1984; Silins, 1994). The central distinction between the transformational and instructional leadership models is that instructional leadership focuses more on first-order changes in classroom practice, while transformational leadership targets second-order changes in the school’s culture (Hallinger, 2003, 2005).

Transformational leaders seek to build a shared vision for the school, foster staff collaboration and learning, develop norms that support continuous improvement, and increase involvement in decision making as means of implementing and sustaining school improvement (Barth, 1990; Fullan, 2001, 2002; Leighton, 1996; Leithwood *et al.*, 2004; Louis and Miles, 1990; Mulford and Bishop, 1997; Mulford and Silins, 2003; Stoll and Fink, 1996). Recognizing that change is driven from the bottom-up as well as from the top-down, there is again an emphasis on the distribution and sharing of leadership responsibilities within the school (Barth, 1990; Blasé, 1993; Fullan, 2002, 2003; Lambert, 2002; Stoll and Fink, 1996).

Studies of transformational school leadership have documented ways in which principals and teacher leaders develop conditions that support school improvement. Such studies have focused on a broader set of outcomes than just student learning, including staff development, collaborative culture, changes in teacher attitudes and behaviors, professional community, teacher commitment, and student engagement (e.g., Leithwood *et al.*, 1993; Leithwood, 1994; Leithwood and Jantzi, 1990, 1999; Marks and Printy, 1996;

Mulford and Bishop, 1997; Mulford and Silins, 2003; Silins, 1994; Silins and Mulford, 2002; Wiley, 2001). The results of these studies have generally confirmed the important role principals and school leaders play in facilitating the development and implementation of school cultures that support school improvement (Leithwood *et al.*, 2004).

In recent years, school improvement has also come to be viewed in a broader global perspective. Studies of educational change in different nations raised awareness of how the cultural contexts of different countries influence goals and strategies for school improvement. Although recognition of contextual differences represents a significant advance in and of itself, less progress has been made in defining how the role of leadership differs across cultures (Dimmock and Walker, 2000; Hallinger and Kantamara, 2000; Hallinger and Leithwood, 1998). This set of issues will undoubtedly engage the attention of the next generation of scholarship as research on school improvement expands into a truly global enterprise.

Challenges in Facilitating School Improvement

Reviews of previous research provide evidence that school leadership can influence school improvement as measured by various different outcomes (e.g., Hallinger and Heck, 1996, 2002; Heck and Hallinger, 2005; Leithwood, 1994; Leithwood *et al.*, 1982, 2004; Teddlie and Reynolds, 2000). Over time, a broader range of studies investigating both leadership and school improvement have emerged. These reflect increasingly diverse epistemological views which raise different research questions and employ different methods for conducting empirical research (Heck and Hallinger, 2005). At present, there is growing policy concern for identifying improvement strategies that are successful at improving levels of student learning over time (Slegers *et al.*, 2002) at the same time that a growing set of scholars question the underlying assumptions of the rational–technical approaches to school improvement (Day *et al.*, 2001). With respect to this more narrowly defined improvement objective, it is important to note limitations in the knowledge base concerning leadership for school leadership.

The positive contributions attributed to school leadership – specifically that of the principal – for improving student learning derive from two general types of studies. The first type is cross-sectional studies of school effectiveness. The second type consists of studies of school change and school improvement projects that observed leadership as a key factor in successful school improvement projects, but which were not explicitly designed to test this as a causal variable. The paucity of systematic longitudinal studies designed to test the contribution of leadership to school improvement leaves unanswered

important questions of causality as well as how school leaders actually facilitate efforts to improve student learning outcomes over time (Heck and Hallinger, 1999; Slegers *et al.*, 2002; Smylie and Hart, 1999).

What remains is to put together various pieces about features of improving schools. Bolstered by stronger theory about organizational change, scholars would inquire into how school processes can be altered in ways that result in improved outcomes. This research would also seek to describe how school leadership can facilitate improvement processes over time. More specifically, how and why do improvement strategies succeed (e.g., Hallinger and Heck, 2002; Louis *et al.*, 1999; Ouston, 1999)? Sustained inquiry about changing school processes and their impact on student learning over time will be required to fill in the gaps in our understanding of how leadership facilitates school improvement.

Our discussion of leadership and school improvement suggests several challenges for further research. First, the school effectiveness literature may have overemphasized the role of school leadership in influencing school improvement – especially as associated solely with school principals. A persisting criticism of this research has been a lack of theory in actually guiding schools toward achieving increased effectiveness (Day *et al.*, 2002; Ouston, 1999; Smylie, 1994). There is ample support for the notion that leadership is an organizational property (e.g., Ogawa and Bossert, 1995), with groups of individuals sharing responsibility for managing changes successfully (e.g., Berman and McLaughlin, 1978; Louis and Miles, 1990). This suggests broadening the investigation of leadership and its effects beyond the principalship (Hallinger and Heck, 2002; Leithwood and Jantzi, 1999, 2000).

Rather than focusing on generalized theories of school leadership and school improvement, scholars may be well advised to focus on building more domain-specific theories of how leadership affects various types of school outcomes not necessarily limited to student learning (Leithwood *et al.*, 1996). It remains to develop greater continuity in purposes and methods in order to find linkages and effects of school leadership on improvement outcomes. Studies in various settings define the goals of the research quite differently, and there is little consistency in terms of the types of contextual variables that may be important to consider (Hallinger and Heck, 1996; Teddlie and Reynolds, 2000). We noted that knowledge development seems to occur primarily in domains where there are externally driven demands for school accountability and increasing student learning (Heck and Hallinger, 2005).

Second, although various schools of scholarship would certainly disagree, we assert the need to place student learning at the center of research on school improvement (e.g., Heck and Hallinger, 2005; Leithwood *et al.*, 2004). This priority reflects a global concern for ensuring improvement in learning results for students. Unfortunately, to date

neither school effectiveness nor school improvement research has addressed what must be done to improve school outcomes (Ouston, 1999). This type of investigation will require improved use of theory about how these processes unfold in schools and greater use of longitudinal data collection. As Willms (1992) has noted, improvement in student learning is a more salient outcome than level of school outcomes measured at one point in time. Slegers *et al.* (2002) suggest that growth in student outcomes has not yet been applied often enough as a dependent variable in studies to be able to evaluate long-term effects.

Studying school improvement over time requires being able to link individual student data to teachers and (or) schools over several measurement occasions. This has made it more difficult to obtain data on both outcomes and changing school process variables that might affect school improvement since it often exceeds the current capacity of educational assessment systems (Heck, 2006). Besides the technical details of appropriate data collection, focusing on how leadership facilitates student learning will involve untangling the conceptual confusion about the types of contributions made by school leadership. Leithwood *et al.* (2004), for example, have made one recent attempt to propose a model of leadership and student learning that can serve as a framework for further empirical investigation.

Third, in our recent review of research on leadership and organizational mission (Hallinger and Heck, 2002), we noted that cognitive perspectives on leadership identify variables such as vision, mission, and goal setting as important in the improvement process. Vision and corresponding goal setting have been identified as important avenues of influence for school improvement (Hallinger, 1998; Hallinger and Heck, 2002; Leithwood, 1994; Leithwood *et al.*, 1998; Slegers *et al.*, 2002). This is closely associated with mediating teacher variables such as commitment to change, collaboration, and participation in decisions about school improvement (Hallinger and Heck, 2002). As Ouston (1999) argues, the school-improvement literature has noted the importance of variables such as school culture, goal setting, planning, and evaluation. Rarely, however, has any attempt been made to develop a dynamic model of school processes, although this is frequently mentioned as being desirable.

Finally, there is a need to determine the proper targets of improvement strategies. There is considerable difference of opinion about where efforts should be most appropriately placed. Changing teacher instructional behavior suggests an emphasis on classrooms, since teachers have a direct impact on student learning (Teddlie and Reynolds, 2000). The literature on large-scale reform efforts to affect outcomes, however, suggests that changing teaching practices is difficult to accomplish (Slegers *et al.*, 2002). At the school level, there is only mixed evidence that professional learning (especially because of the interacting effects of context) or site-based management have

measurable impacts on student outcomes (e.g., Smylie and Hart, 1999). Research does indicate, however, the role of leadership in terms of redefining norms and educational processes is significant (Hallinger and Heck, 2002; Leithwood *et al.*, 2004).

Other more recent proposals emphasize targeting the district and state and local educational levels simultaneously and interactively in order to sustain school improvement (Barber and Fullan, 2005; Fullan *et al.*, 2006). This approach to leadership for school improvement combines features of the rational–technical, political, and cultural perspectives on school improvement and has emerged out of a process of learning from putting research into practice in several large educational systems. As Fullan asserts, “It has become increasingly clear that leadership at all levels of the system is the key lever for reform, especially leaders who (1) focus on capacity building and (2) develop other leaders who can carry on” (Fullan, 2006: 33). This conclusion certainly finds support in the literature on leadership for school improvement. The implementation of tri-level school improvement efforts in large educational systems could provide the opportunities needed for the type of empirical investigation that we have suggested is necessary for furthering our understanding of the role of leadership in school improvement.

Bibliography

- Barber, M. and Fullan, M. (2005). Tri-level development: It's the system. *Education Week* **24**(25), 32; 34; 35.
- Barth, R. (1986). On sheep and goats and school reform. *Phi Delta Kappan* **68**(4), 293–296.
- Barth, R. (1990). *Improving Schools from within*. San Francisco, CA: Jossey-Bass.
- Barth, R. (1999). *The Teacher Leader*. Providence, RI: The Rhode Island Foundation.
- Berman, P. and McLaughlin, M. (1978). *Federal Programs Supporting Educational Change* vol. 3. Santa Monica, CA: Rand Corporation.
- Blasé, J. (1993). The micropolitics of effective school-based leadership: Teachers' perspectives. *Educational Administration Quarterly* **29**(2), 142–163.
- Bolman, L. and Deal, T. (1992). *Reframing Organizations*. San Francisco, CA: Jossey-Bass.
- Bossert, S., Dwyer, D., Rowan, B., and Lee, G. (1982). The instructional management role of the principal. *Educational Administration Quarterly* **18**(3), 34–64.
- Brandt, R. (1982). On school improvement: A conversation with Ronald Edmonds. *Educational Leadership* **40**, 13–15.
- Clark, D., Lotto, L., and Astuto, T. (1984). Effective schools and school improvement: A comparative analysis of two lines of inquiry. *Educational Administration Quarterly* **20**(3), 41–68.
- Crandall, D., Eisman, J., and Louis, K. S. (1986). Strategic planning issues that bear on the success of school improvement efforts. *Educational Administration Quarterly* **22**(3), 21–53.
- Cuban, L. (1984). Transforming the frog into a prince: Effective schools research, policy, and practice at the district level. *Harvard Educational Review* **54**(2), 129–151.
- Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher* **19**(1), 10–21.
- Day, C., Harris, A., and Hadfield, M. (2001). Challenging the orthodoxy of effective school leadership. *International Journal of Leadership in Education* **4**(1), 39–56.
- Deal, T. (1990). Reframing reform. *Educational Leadership* **47**(8), 6–12.
- Dimmock, C. and Walker, A. (2000). Introduction justifying a cross-cultural comparative approach to school leadership and management. *School Leadership and Management* **20**(2), 137–141.
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership* **37**, 15–24.
- Firestone, W. A. and Corbett, H. D. (1988). Planned organizational change. In Boyan, N. (ed.) *Handbook of Research on Educational Administration*, pp 321–340. New York: Longman.
- Fullan, M. (1982). *The Meaning of Educational Change*. New York: Teachers College Press.
- Fullan, M. (2001). *Leading in a Culture of Change*. San Francisco, CA: Jossey-Bass.
- Fullan, M. (2002). The change leader. *Educational Leadership* **59**(8), 16–20.
- Fullan, M. (2006). *The Development of Transformational Leaders for Educational Decentralization*. Toronto, ON: Michael Fullan.
- Fullan, M., Hill, P., and Crevola, C. (2006). *Breakthrough*. Thousand Oaks, CA: Corwin Press.
- Fullan, M. and Pomfret, A. (1977). Research on curriculum and instruction implementation. *Review of Educational Research* **47**(1), 335–397.
- Hall, G. and Hord, S. (1987). *Change in Schools*. Albany, NY: SUNY Press.
- Hallinger, P. (1998). Educational change in Southeast Asia: The challenge of creating learning systems. *Journal of Educational Administration* **36**(5), 492–509.
- Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education* **33**(3), 329–351.
- Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and Policy in Schools* **4**(3), 221–240.
- Hallinger, P. and Heck, R. H. (1996a). Reassessing the principal's role in school effectiveness: A review of the empirical research, 1980–1995. *Educational Administration Quarterly* **32**(X), XX.
- Hallinger, P. and Heck, R. H. (1996b). The principal's role in school effectiveness: An assessment of methodological progress, 1980–1995. In Leithwood, K., Chapman, J., Corson, D., Hallinger, P., and Hart, A. (eds.) *International Handbook of Educational Leadership and Administration*, pp 723–783. Dordrecht: Kluwer Academic Publishers.
- Hallinger, P. and Kantamara, P. (2000). Leading educational change in Thailand: Opening a window on leadership as a cultural process. *School Leadership and Management* **20**(1), 189–206.
- Hallinger, P. and Leithwood, K. (1998). Unseen forces: The impact of social culture on leadership. *Peabody Journal of Education* **73**(2), 126–151.
- Heck, R. H. (2002). Issues in the investigation of school leadership across cultures. In Walker, A. and Dimmock, C. (eds.) *School Leadership and Administration: Adopting a Cultural Perspective*, pp 78–100. New York: RoutledgeFalmer.
- Heck, R. H. (2006). Assessing school achievement progress: Comparing alternative approaches. *Educational Administration Quarterly* **42**(5), 1–33.
- Heck, R. H. and Hallinger, P. (2005). The study of educational leadership and management: Where does the field stand today? *Educational Management, Administration and Leadership* **33**(2), 229–244.
- Jackson, D. (2000). The school improvement journey: Perspectives on leadership. *School Leadership and Management* **20**(1), 61–78.
- Kleine-Kracht, P. (1993). Indirect instructional leadership: An administrator's choice. *Educational Administration Quarterly* **18**(4), 1–29.
- Lambert, L. (1998). *Building Leadership Capacity in Schools*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Lambert, L. (2002). A framework for shared leadership. *Educational Leadership* **59**(8), 37–40.
- Leighton, M. S. (1996). *The Role of Leadership in Sustaining School Reform: Voices from the Field*. Washington, DC: US Department of Education.
- Leithwood, K. (1994). Leadership for school restructuring. *Educational Administration Quarterly* **30**(4), 498–518.

- Leithwood, K. and Jantzi, D. (1999). The relative effects of principal and teachers sources of leadership on student engagement with school. *Educational Administration Quarterly* **35**, 679–706.
- Leithwood, K. and Jantzi, D. (2000). The effects of transformational leadership on organizational conditions and student engagement with the school. *Journal of Educational Administration* **38**(2), 112–129.
- Leithwood, K., Louis, K. S., Anderson, S., and Wahlstrom, K. (2004). Review of research: How leadership influences student learning. Wallace Foundation. <http://www.wallacefoundation.org/NR/rdonlyres/E3BCCFA5-A88B-45D3-8E27-B973732283C9/0/ReviewofResearchLearningFromLeadership.pdf>.
- Leithwood, K. and Montgomery, D. (1982). The role of the elementary principal in program improvement. *Review of Educational Research* **52**(3), 309–339.
- Levine, D. U. (1991). Creating effective schools: Findings and implications from research and practice. *Phi Delta Kappan* **72**(5), 389–393.
- Louis, K. S. (1994). Beyond “managed change”: Rethinking how schools improve. *School Effectiveness and School Improvement* **5**(1), 1–24.
- Louis, K. S. and Miles, M. B. (1990). *Improving the Urban High School*. New York: Teachers College Press.
- Louis, K. S., Toole, J., and Hargreaves, A. (1999). Rethinking school improvement. In Murphy, J. and Seashore Louis, K. (eds.) *Handbook of Research on Educational Administration*, 2nd edn., pp 251–276. San Francisco, CA: Jossey-Bass.
- Marks, H. and Printy, S. (2004). Principal leadership and school performance: An integration of transformation and instructional leadership. *Educational Administration Quarterly*.
- McLaughlin, M. (1990). The Rand change agent study revisited. *Educational Researcher* **5**, 11–16.
- Meyer, J. and Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology* **83**(2), 340–363.
- Mortimore, P. (1993). School effectiveness and the management of effective learning and teaching. *School Effectiveness and School Improvement* **4**, 290–310.
- Mulford, B. and Bishop, P. (1997). *Leadership in Organisational Learning and Student Outcomes (LOLSO) Project*. Launceston: University of Tasmania.
- Mulford, B. and Silins, H. (2003). Leadership for organisational learning and improved student outcomes – what do we know? *Cambridge Journal of Education* **33**(2), 175–195.
- Murphy, J., Hallinger, P., and Mesa, P. (1985a). School effectiveness: Checking progress and assumptions and developing a role for state and federal government. *Teachers College Record* **86**(4), 615–642.
- Murphy, J., Hallinger, P., and Mesa, R. P. (1985b). Strategies for tightening the coupling in schools. *NASSP Bulletin* **69**(478), 7–13.
- Ogawa, R. and Bossert, S. (1995). Leadership as an organizational quality. *Educational Administration Quarterly* **31**(2), 224–243.
- Ouston, J. (1999). School effectiveness and school improvement: Critique of a movement. In Bush, T., Bell, L., Bolam, R., Glatler, R., and Ribbins, P. (eds.) *Educational Management: Redefining Theory, Policy and Practice*. London: Paul Chapman.
- Reynolds, D., Teddlie, C., Hopkins, D., and Stringfield, S. (2000). *Linking School Effectiveness and School Improvement*. In Teddlie, C. and Reynolds, D. (eds.) *The International Handbook of School Effectiveness Research*, pp 206–231. London: Falmer.
- Rowan, B. and Denk, C. (1984). Management succession, school socioeconomic context and basic skills achievement. *American Educational Research Journal* **21**(3), 517–537.
- Rutter, M., Maugham, B., Mortimore, P., Ouston, J., and Smith, A. (1979). *Fifteen Thousand Hours: Secondary Schools and Their Effects on Children*. Cambridge, MA: Harvard University Press.
- Sarason, S. B. (1982). *Culture of the School and the Problem of Change*, 2nd edn. Boston, MA: Allyn and Bacon.
- Sarason, S. B. (1990). *The Predictable Failure of Educational Reform – Can We Change Course before It's Too Late?* San Francisco, CA: Jossey-Bass.
- Scott, W. R. (2002). *Organizations: Rational, Natural and Open Systems*, 5th edn. New York: Prentice-Hall.
- Silins, H. (1994). The relationship between transformational and transactional leadership and school improvement outcomes. *School Effectiveness and School Improvement* **5**(3), 272–298.
- Silins, H. and Mulford, B. (2002). Leadership, restructuring and organisational outcomes. In Leithwood, K., Hallinger, P., Furman-Brown, G., et al. (eds.) *Second International Handbook of Educational Leadership and Administration*. Dordrecht: Kluwer Academic Press.
- Sleegers, P., Geijsel, F., and van den Berg, R. (2002). Conditions fostering educational change. In Leithwood, K. and Hallinger, P. (eds.) *Second International Handbook of Educational Leadership and Administration*, pp 75–102. Dordrecht: Kluwer Academic Publishers.
- Smylie, M. A. (1994). Understanding school restructuring and improvement. *Educational Researcher* **23**(3), 39–40.
- Smylie, M. A. and Hart, A. (1999). School leadership for teacher learning and change: A human and social capital development perspective. In Murphy, J. and Louis, K. S. (eds.) *Handbook of Research on Educational Administration*, 2nd edn., pp 421–441. San Francisco, CA: Jossey-Bass.
- Southworth, G. (2002). Instructional leadership in schools: Reflections and empirical evidence. *School Leadership and Management* **22**(1), 73–92.
- Stoll, L. and Fink, D. (1996). *Changing Our Schools: Linking School Effectiveness and School Improvement*. Buckingham: Open University Press.
- Teddlie, C. and Reynolds, D. (2000). *The International Handbook of School Effectiveness Research*. London: Falmer.
- Van den Berg, R. and Vandenberghe, R. (1986). *Strategies for Large-Scale Change in Education: Strategies and Dilemmas*. Leuven: Acco.
- Van Velzen, W., Miles, M., Ekholm, M., Hameyer, U., and Robin, D. (1985). *Making School Improvement Work: A Conceptual Guide to Practice*. Leuven: Acco.
- Weick, K. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly* **21**, 1–19.
- Wiley, S. (2001). Contextual effects on student achievement: School leadership and professional community. *Journal of Educational Change* **2**(1), 1–33.
- Willms, D. (1992). *Monitoring School Performance*. London: Falmer.

Further Reading

- Caldwell, B. (1998). Strategic leadership, resource management and effective school reform. *Journal of Educational Administration* **36**(5), 445–461.
- Firestone, W. A. and Wilson, B. L. (1985). Using bureaucratic and cultural linkages to improve instruction: The principal's contribution. *Educational Administration Quarterly* **20**(2), 7–30.
- Geijsel, F., Sleegers, P., Leithwood, K., and Jantzi, D. (2002). Transformational leadership effects on teachers' commitment and effort toward school reform. *Paper Presented at the Annual Meeting of the American Educational Research Association*, New Orleans.
- Hallinger, P. and Heck, R. H. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement* **9**(2), 157–191.
- Hargreaves, A. (1992). Cultures of teaching: A focus for change. In Hargreaves, A. and Fullan, M. B. (eds.) *Understanding Teacher Development*, pp 216–236. New York: Teachers College Press.
- Louis, K. S. and Marks, H. (1998). Does professional community affect the classroom? *American Journal of Education* **106**(4), 532–575.
- Southworth, G. (1999). Headship, leadership, and school improvement. In Southworth, G. and Lincoln, P. (eds.) *Supporting Improving Primary Schools: The Role of Heads and LEAs in Raising Standards*, pp 69–86. London: Falmer.
- Vandenberghe, R. (2002). Teachers professional development as the core of school improvement. *International Journal of Educational Research* **37**, 653–659.

van de Grift, W. (1990). Educational leadership and academic achievement in elementary education. *School Effectiveness and School Improvement* **1**(3), 26–40.

Weatherly, R. and Lipsky, M. (1977). Street-level bureaucrats and institutional innovation: Implementing special-education reform. *Harvard Educational Review* **47**(2), 171–191.

Wiley, S. (1998). Contextual effects on student achievement: School leadership and professional community. *Paper Presented at the Annual Meeting of the American Sociological Association*.

Witziers, B., Bosker, R., and Kruger, M. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly* **39**, 398–425.

Organizational Learning in Schools

B Mulford, University of Tasmania, Hobart, TAS, Australia

H Silins, Flinders University of South Australia, Adelaide, SA, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In the 1990s, a veritable explosion of books and journal articles on organizational learning (OL) in business (e.g., Senge, 1990; Easterby-Smith *et al.*, 1999) contrasted with a small but increasing number of educational writers taking up the OL baton (e.g., Argyris, 1993; Miles, 1993; Watkins and Marsick, 1993; Fullan, 1993; Louis, 1994; Keating, 1995; Cousins, 1996; Leithwood and Louis, 1998; Mulford, 1998). These writers were followed by an even smaller number of educational researchers (e.g., Louis *et al.*, 1995; Mitchell, 1995; Hajnal *et al.*, 1998; Leithwood *et al.*, 1998; Marks *et al.*, 2000; Silins and Mulford, 2002a; Voulalas and Sharpe, 2005). Toward the end of the 1990s, Leithwood and Louis (1998: 7) pointed out that while the logical case for OL was compelling, “empirical support for the claim that increases in such learning will contribute to organizational effectiveness or productivity is embarrassingly slim” and that a “review of empirical research on organizational learning in schools alone would make a very quick read indeed.”

Why the growing interest in OL in schools? Almost 30 years ago, Argyris and Schön (1978: 320) argued that those intervening in organizations “have had to recognize that their main challenge is not to help an organization become more effective at the performance of a stable task in the light of stable purposes, but rather to help an organization restructure its purposes and redefine its task in the face of a changing environment.” In a similar vein, but more succinctly stated, Senge’s (1990) learning organization (LO) focuses on “continually expanding its capacity to create its future” (p. 14). Argyris and Schön (1978: 331) gave elementary importance to “an organization’s capacity for conscious transformation of its own theory of action, and to individuals’ ability to appreciate and transform the learning systems in which they live.” The better organizations are at organizational enquiry or learning, the more likely they “will be able to detect and correct errors, and to see when they are unable to detect and correct errors” and the more likely they “will be at being innovative or knowing the limits of their innovation” (Argyris, 1993: 1).

What Is OL?

Consistent with the reasons for the growth of interest in OL, Leithwood and Aitken (1995: 63) defined OL in

schools as “... a group of people pursuing common purposes (individual purposes as well) with a collective commitment to regularly weighing the value of those purposes, modifying them when that makes sense, and continuously developing more effective and efficient ways of accomplishing those purposes.” According to Fullan (1994: 5), in schools that demonstrate OL, two things are key, “learning is considered both ‘essential’ (important) and ‘integral’ (linked) to goals and activities.” Keating (1995: 20) pointed out that the area of OL had many descriptions and case studies, but little comparative or systematic research. He believed that “key elements noted by thoughtful observers can be summarized.” For Keating, these elements included: coordinated group effort toward commonly shared goals; active commitment to continuous improvement and to the diffusion of best practices throughout the organization; horizontal networks of information flow to help bring together expertise as well as links with the external world; and the ability to understand, analyze, and use the dynamic system within which the staff functions.

Mitchell (1995: 47) also believed that much of the work on OL was “theoretical in nature, based on literature reviews and anecdotal reports rather than on empirical evidence.” She noted that, in her reading of the OL literature, she was reminded of the story of the three blind men trying to describe an elephant from a different vantage point. Nevertheless, Mitchell (1995) concluded that there are points of convergence including the following set of indicators of school OL: developing a spirit of trust; developing common understandings; developing shared vision; sharing information openly and honestly; engaging in collaborative practices; engaging in professional learning and growth; using reflective self-analysis to raise awareness of assumptions and beliefs; examining current practices critically; understanding the inevitability of conflict; engaging in dialog in order to understand the frames of reference of others; raising sensitive issues for discussion; experimenting with new practices; changing personal frames of reference if warranted; managing differences of opinion through inquiry and problem solving; understanding systemic influences and relationships; and correcting disruptive power imbalances.

In Mitchell’s (1995) subsequent research in an urban Canadian elementary school, she found that a school operating effectively as an LO could be described using four processes, three basic assumptions, and three distinct

phases. To quote Mitchell, of the four processes, two cognitive and two affective:

The cognitive processes of reflection and conversation enabled the teachers to become aware of their practices and of those of their colleagues, to assess the desirability of those practices, and to discover new possibilities. The affective processes of affirmation [of each other as professionals] and invitation [into school deliberations] served to create positive working relationships by affirming the professional capabilities of individuals and by valuing the contributions of all staff members. (Mitchell, 1995: 223)

These four processes were founded upon three basic assumptions (Mitchell, 1995: 226): “that each individual was responsible for the welfare of the group and the success of the school; that diversity among individuals was recognized, honoured, and valued; and that psychological safety would be maintained in group deliberations.” Finally, these four processes moved through three phases (Mitchell, 1995: 226): “naming and framing to clarify positions and opinions; analysing and integrating especially when new ideas are opened up for possible experimentation; and applying and experimenting.” Indicators of OL synthesized from the literature by Mitchell (1995: 231) were found to emerge differentially at each of the three phases thus implying that “these behaviours and processes do not develop overnight, but rather build on one another over time”. This developmental nature of OL highlighted for Mitchell (1995: 243) “the importance of the ‘naming and framing’ and ‘analysing and integrating’ phases.” The comfort level with OL grew as teachers developed their own understandings about the concept, analyzed their own practices in the light of those understandings, and applied appropriate aspects of what they had learned.

In related work, Boyd and Hord clustered the indicators of schools as professional learning communities (PLCs) into five functional groupings (Boyd and Hord, 1994; Hord, 1997): shared values and vision; collective learning and application; supportive and shared leadership; supportive conditions; and shared personal practice (with a strong and unwavering focus on student learning). Subsequently, other researchers examined the conditions that foster the development of these five dimensions in schools that successfully improved student learning (Hord, 2004; Hipp and Huffman, 2003). Bolam *et al.* (2005) identified eight characteristics of schools that were PLCs: inclusive membership; mutual trust, respect, and support; openness, networks, and partnerships; shared values and vision; collective responsibility for pupils’ learning; reflective professional inquiry, collaboration focused on learning; and group as well as individual professional learning. A number of external and internal factors facilitated or inhibited these characteristics, including the internal school processes of resources and structures, promotion of professional learning, evaluating

and sustaining the PLC, and leadership. These characteristics and processes were found to be exhibited in varying degrees depending on where a school was placed on the sequence of starter, developer, or mature PLC as judged by the percentage of staff involved in key PLC activities.

From an extensive review of the noneducational and educational literature, including Senge’s (1990) seminal model of the five disciplines of OL (mental models, shared vision, personal mastery, team learning, and systems thinking), Silins *et al.* (2002) defined OL as schools that: employed processes of environmental scanning; developed shared goals; established collaborative teaching and learning environments; encouraged initiatives and risk taking; regularly reviewed all aspects related to and influencing the work of the school; recognized and reinforced good work; and provided opportunities for continuing professional development. This seven-factor structure, used to generate questionnaire items for a large research project on leadership for organizational and student learning in Australian high schools, was not supported empirically. Instead, a four-factor nested model of OL was supported comprising a trusting and collaborative environment, a shared and monitored mission, taking initiatives and risks, and ongoing, relevant professional development. Subsequent case study analysis suggested that the first three of these factors were sequential (Mulford *et al.*, 2004).

The emergence of consistent lists of identifying the characteristics of OL can be noted. In addition, some of the interrelationships among these characteristics are becoming clearer. For example, from a within-the-organization focus, the identifying characteristics tend to describe a journey rather than a destination. The first stage involves developing common understandings, honesty, and trust through dialog, sharing, and managing the inevitable conflict. These learning processes are then employed to link to the outside, examine current practice critically, and develop shared values as well as a vision for the school. The processes, the content (or identified changes), and shared values are employed to actually make the changes identified, thus taking initiatives and risks while continuously learning and improving. The stages of the journey are sequential and cyclical, lifting practice through renewed commitments and improved abilities (Silins and Mulford, 2002a).

In summary, OL in schools can be defined as a school staff’s collective commitment to a continuous journey pursuing, monitoring, and, where necessary, changing and improving shared processes and purposes.

Why Is OL Important?

Louis (1994: 20) argued that existing models of change management in education were inadequate and that the

OL model was more promising because the “image of change that emerges in the organizational learning paradigm has elements of both managed change (organizational learning is affected by structure and leadership) and anarchy (the emergence of alternative paradigms and the selection of a new paradigm is a chaotic, largely unpredictable process).” She noted that the concept had the “potential for helping to think about the problem of how schools change basic assumptions about ‘what it is we do here’ when demands for significant reforms are made” (Louis, 1994: 9). The conclusions of a 5-year longitudinal study of policy implementation in the Canadian province of British Columbia (Leithwood, *et al.*, 1995) confirmed that individual and collective learning processes (OL) best explained the variation in the productivity of schools. Collective learning processes were measured through interviews and included exchange of information through informal discussions among colleagues, experimentation with new practices, and joint planning and teaching. Newman and Wehlage’s (1995) influential research on successful school restructuring concurred, arguing the need for schools to become LOs by building organizational capacity.

The demands of fundamental and rapid change make sense of the shorthand statement that learning is change, as does the emphasis on active dialog among those with diverse points of view. To speak only with people sharing the same perspective runs the risk of those involved not being moved to thoughts of something better (Greene, 1995). The emphasis on the quality and character of human relationships and the necessity for individuals, groups, and organizations to construct meaning for themselves is appealing. This presumes the rejection of imposing one’s will on others for successful change. Louis (1994) argued that the principles of OL should shift discussion of educational problems away from blaming and finger-pointing to a consideration of systemic barriers to learning by those actually involved in change.

This emphasis on learning within a school might create an environment where the practice of the most effective teachers can be used to support and develop others’ work. Greater consistency in the range of opportunities that learners experience during their education is especially important given that variation in performance within schools has been found to be 4 times as great as variation in performance between schools (OECD, 2000). During 2003–04, 24 UK schools belonging to the National College for School Leadership’s Leadership Network (Connor, 2005) explored this issue of within-school variation. Researchers found schools applying a range of strategies to reduce this variation. Grouped into four themes, these were: the collection, analysis, interpretation, and use of data; the development of strategies that focus on teacher learning through, for example, the focused observation of specific aspects of practice; proposals for

curriculum reform, especially ensuring a relationship with the interests and preferences of learners; and the development of middle leaders, and facilitating learning from the innovative practice of others in the school.

What Are the Effects of OL?

Argyris (1993: 1) pointed out that in contexts such as those facing education today, organizations “have the most difficulty at learning when the problems are difficult, and embarrassing or threatening, in short, precisely when they need learning most.” Harassed and harried, caught on Deal’s (1990: 145) “high-speed carousel of change,” most schools and the systems in which they reside may not have the time, energy, or inclination for understanding, let alone developing, OL. Although the current research output on OL in schools is small, it is impressive and goes some way to countering Leithwood’s and Louis’ (1998) criticism that whether OL will contribute to organizational effectiveness or productivity is still to be empirically determined. Across a number of countries and school levels, OL has been clearly shown to be linked to school leadership, successful school improvement and innovation, school development, including collective teacher efficacy, and/or a range of student outcomes. A representative sample of this research is summarized in what follows.

A number of studies have linked OL and school leadership. A study of six Canadian schools by Leithwood *et al.* (1999) considered promising sites of OL, including four secondary schools, and found that school leadership practices had among the strongest direct and indirect influences. Those practices included identifying and articulating a vision, fostering the acceptance of group goals, structuring the school to enhance participation in decisions, providing intellectual stimulation, and conveying high expectations, that is, practices associated with transformational leadership. Such uniformly positive contributions of school leadership to OL have been confirmed by Sheppard and Brown (2000) and Silins and Mulford (2002a, 2002b).

Further studies link OL and successful school improvement initiatives and innovation. Hajnal *et al.* (1998) sought to identify successful school improvement initiatives and indicators of institutionalization in 93 Canadian schools. They concluded that leadership and OL were critical to school improvement. Leadership required attention to school and personal matters. School matters included a shared vision, resources, and an environment supportive of change. Personal matters included paying attention to individual differences, facilitating and empowering, encouraging collaboration, and building an environment of trust and care. OL required established practices that supported extensive collaboration, individual learning, and a clear

vision for the school. Sheppard and Brown (2000) collected data in two Canadian high schools with national reputations as innovative schools. Despite strong existing shared staff knowledge structures and the immediacy of teachers' work environments, the schools were found to gradually take on the domains of knowledge that Fullan (1995) argued were required for OL. Teachers and their administrators established partnerships with outside groups, which resulted in significant additional funding and new program initiatives, took ownership of the change process, and formalized collaborative structures. In time, the staff came to focus directly on teaching and learning, even though this was not their starting point. Eventually, the staff evolved into continuous learning communities, taking responsibility for their own professional development. Finally, the staff adopted programs targeted at the needs of all students.

Additional studies linked OL and school development, including teacher efficacy. Johnston's (1997) case studies of three Australian high schools with OL characteristics identified four key elements as instrumental in their development: an inclusive collaborative structure; effective communication channels; integrated and inclusive professional development programs; and learning-focused leadership. The latter was described as empowering (all within the school could be leaders) rather than heroic or hands on leadership. Ross *et al.* (2004: 177) found that "the strongest effects on collective teacher efficacy" were characteristics associated with OL, including "shared school goals, school-wide collaboration, fit of plans with school needs, and empowering school leadership." Studies have also documented a strong link between perceived teachers' collective efficacy and differences in student achievement. Bandura (1993) demonstrated that the effects of collective efficacy on student achievement were stronger than the direct link between socioeconomic status (SES) and student achievement. Goddard *et al.* (2004) found that even after controlling for students' prior achievement, race/ethnicity, SES, and gender, collective efficacy beliefs have stronger effects on student achievement than student race or SES. Goddard *et al.* (2004: 420) calculated, for example, that "a 1 SD increase in collective efficacy is associated with a gain of about 0.25 SD in terms of the number of students who pass high-stakes assessments in 12th grade." Goddard (2002) had also found that where teachers have the opportunity to influence important school decisions, they also tend to have stronger beliefs in the co-joint capability of their staff.

Finally, and most crucially, are studies linking leadership, OL, and student outcomes. Mawhinney *et al.* (2005: 5) examined how, "under pressures of accountability, districts in United States ... are undertaking research to support their development of strategic actions to foster organizational learning in schools." In a study of all teachers ($N=2448$) and administrators ($N=17$) employed

in 49 schools (31 elementary, nine middle, and eight high schools), Mawhinney *et al.* (2005) found collective efficacy and PLC (using a measure of OL) are positively related with the strongest relationships being with dimensions of shared values and vision, collective learning and application of that learning, and supportive conditions. While only two variables, the students' previous year's proficiency and collective efficacy beliefs, were significant predictors of student reading proficiency, "teachers who perceive their schools to be characterised by shared leadership, focused vision, collaborative work, shared observation, and supportive conditions also perceive their colleagues to be effective in bringing about student learning" (Mawhinney *et al.*, 2005: 20). Elementary, low-poverty, and white schools as well as female teachers perceived the highest levels of collective efficacy.

In the first study of its kind, Wheelan and Tilin (1999) examined the relationships between 292 teacher perceptions of staff group effectiveness and development and the actual levels of productivity in ten USA elementary, middle, and high schools. A survey measured the group development (from inclusion and dependency to conflict and counterdependency, trust and structure, and, finally, work and productivity) and data were gathered on student grades, standardized test scores, and the degree of parental involvement. Wheelan and Tilin (1999: 77) "found significant relationships between ... group development level and maths rank, reading rank and total achievement rank (a combination of maths and reading)." The staff in schools, classified as high in reading and total rank, has significantly lower scores on conflict and significantly higher scores on trust and structure as well as on work and productivity stages of group development. In addition, those high on trust and structure as well as on work and productivity also reported higher levels of parental involvement.

Marks *et al.* (2000) selected 24 predominantly urban, economically disadvantaged elementary, middle, and high schools in USA, where restructuring had been particularly thorough, in which to collect data using a battery of quantitative and qualitative instruments. Controlling for gender, race, ethnicity, SES, and student ability, the researchers examined the relationships among three dependent variables, authentic academic achievement (the sum of student scores in mathematics and social studies on analysis, disciplinary concepts, and elaborated written communication), National Assessment of Educational Progress (NAEP) achievement (a test of basic knowledge and skills in mathematics and reading/writing), and authentic pedagogy (a composite measure combining teachers' scores on observed classroom instruction and assessment tasks), and the independent variable, capacity for OL. Capacity for OL involved six dimensions: school structural conditions (school size, extent of decentralized governance, and the amount of

time teachers spend meeting with colleagues); teacher empowerment; shared commitment and collaborative activity; knowledge and skills (including school-oriented staff development, openness to innovation, and pedagogical content knowledge); supportive leadership; and feedback and accountability. Marks *et al.* (2000) found that 18–26% of the dependent variables were attributable to differences among schools. A school's capacity for OL was a strong predictor of success on the measures of authentic achievement, NAEP achievement, and pedagogical quality. They concluded that the capacity for OL supports the central teaching and learning activities of the school. However, they warn that the presence of only some of the dimensions of this capacity is not sufficient to activate OL. In the related area of PLCs, an almost 3-year research by Bolam *et al.*, including 393 surveys and 16 case study schools in England, also found

a positive, though weak, link between full expression of PLC characteristics and pupil outcomes – in particular value added performance... attendance, interest in learning and actual learning, as well as on the individual and collective professional learning, practice and morale of teaching and support staff. (Bolam *et al.*, 2005: iii)

Identified by reviews in the area (Bell *et al.*, 2002; Leithwood *et al.*, 2006) as one of the most thorough examinations of the relationship between leadership, OL, and student learning is the longitudinal Leadership for Organizational Learning and Student Outcomes (LOLSO) project. The LOLSO research on 96 Australian high schools involving over 5000 year-10 students and 3700 teachers (Silins and Mulford, 2002a, 2002b, 2004; Silins *et al.*, 2002; Mulford *et al.*, 2004) found that leadership that makes a difference has been found to be both position based (principal and transformational) and distributive (administrative team and teachers). However, both were indirectly related to student outcomes. OL, defined (not unlike collective teacher efficacy) by a trusting and collaborative environment, shared and monitored mission, supporting initiatives and risks, and providing appropriate professional development, was the important intervening variable between leadership and teacher work and, subsequently, student outcomes. That is, leadership contributed to OL, which in turn influenced what happened in the core business of the school – the teaching and learning. It influenced the way students perceived teachers organizing and conducting their instruction, and their educational interactions with, and expectations for, their students. Pupils' positive perceptions of teachers' work directly promoted their participation in school, academic self-concept, and engagement with school. Pupil participation was directly and pupil engagement indirectly (through retention) related to academic achievement. School size was negatively linked, while SES and, especially, student home educational environment, positively linked to these relationships.

Constraints

Effective OL involves those in a school taking responsibility for outcomes. Yet, how do those in schools share responsibility for outcomes/learnings/problems when several matters are not within a school's control? Levin (1994: 48) points out that “arguably, the most powerful educative (and miss-educative) forces in our society lie outside the school system” and that “educational reformers might direct their efforts with greater effect to the mass media, popular culture, and the emerging Information Superhighway [sic].” Hargreaves (1995: 12) goes further in suggesting that, “the dynamics of blame. . . the claim that we share responsibility for all our problems seems . . . pernicious. Schools and teachers have to deal with many mandates they do not control. Public schools are not bounded systems.” Voulalas and Sharpe (2005) conducted interviews with 22 Australian metropolitan principals identified by district superintendents as taking action to help their schools become LOs. They found that respondents lacked a clear understanding of LOs and that traditional hierarchical school structures and cultures, lack of implementation time, and difficulty in obtaining the support of staff and parents were major barriers to implementation.

How can we better handle the potential incompatibility between accountability and empowerment (Glickman, 1990) faced by schools? Teacher empowerment without accountability can lead to anarchy, whereas accountability without empowerment can lead to subservience. The OL in schools' literature suggests that to succeed in a rapidly changing and increasingly complex world, it is vital that schools grow, develop, adapt, and take charge of change so that they can control their own futures (Gitlin and Margonis, 1995; Stoll *et al.*, 2003). OL involves a different way of operating than many are used to. Accountability necessitates a movement away from external, extrinsic compliance to intrinsic, shared responsibility. Rusch (2005) reinforced this movement by identifying competing institutional scripts that act as barriers to OL in school systems. Structures in OL networks are malleable, not fixed or hierarchical as in traditional systems; conflict is open and valued rather than hidden and feared; communication is open and unbounded rather than controlled and closed; leadership is fluid rather than hierarchical or assigned; relationships are egalitarian rather than meritocratic; and knowledge is based on enquiry and valued for the learning that takes place rather than based on expertise and valued for the extent of knowledge. Gitlin and Margonis (1995) suggested that if schools are to become effective at OL, then at least one of the grammars of schooling, that is, how they use time, will need to be profoundly reviewed. Louis *et al.* (1995) and Lieberman (1995) made this same point, as did Mitchell (1995: 231) when she pointed out that the behaviors and processes required for effective OL “do not

develop overnight, but rather build on one another over time.” Specifically, Mitchell (1995: 243) found that her first phase of building effective OL, naming and framing, “is an active process, and teachers need time to talk about and to reflect on initiatives. We could be forgiven for wondering whether the fundamentals of education – teaching and learning – have been displaced for those inside and outside schools. Has displacement occurred because the improvement of interpersonal relationships, curriculum, and instruction call for delayed gratification, with the signs of progress not easily detected?”

Conclusion

Reporting on research in 200 business organizations, Redding and Catalanello (1994: xiii) stated that “the quality of strategic plans did not seem to have much to do with companies’ ability (or lack of ability) to change. Nor did the degree of effort devoted to the implementation of strategic plans . . . what seemed to matter most was the organization’s readiness for change.” These authors concluded, as have Beer *et al.* (1990), Drucker (1995), Kanter (1991), and Mintzberg (1994), that “fundamental organizational change most often results from journeys of learning – of setting out in a direction, of gaining new insights and making discoveries en route, of going back, adjusting old maps, developing revised plans, and taking new actions” (p. 7). This emphasis on learning, seeing journey’s of learning, learning one’s way forward, and learning as change, involving dialog and constructed meaning, is consistent with the reasons for the growing interest in, definitions of, research on, and continuing advocacy (e.g., Collinson *et al.*, 2006) of OL in schools.

OL deserves closer attention since it is argued that schools need to become LOs to fully engage staff and students as well as to meet the heightened multiple expectations now placed on them (Mulford, 2003). New types of relationships evolve between students, teachers, and leaders within OL schools based around a reasonably common set of characteristics that include a trusting and collaborative environment, a shared and monitored mission, supporting initiatives, risk taking, and ongoing, relevant professional development. These characteristics of OL have been found to be staged and built over time. For example, one message that emerges is that those in schools must also learn how to lose time in order to gain time. Awareness of, and skill development in, group and organizational processes is the first step. Instead of others trying to insert something into a school’s culture, the school, and especially its leadership, should first be trying to help that culture develop an awareness of, and a responsiveness to, itself. Success is more likely where people act rather than always react; are empowered; involved in decision making through a transparent, facilitative, and supportive

structure; and are trusted, respected, encouraged, and valued. Somech (2005) found that an interesting catch-22 situation developed when teachers simultaneously possess high levels of personal and team empowerment in that high team commitment can place constraints on a teacher’s individual autonomy and his or her personal empowerment. Somech (2005: 260) concluded that teachers “might achieve a ‘win-win’ situation if the norms of the profession of teaching do not translate autonomy into isolation and if collaboration is not perceived as violating teachers’ freedom.”

While collaboration does not necessarily suit or benefit everyone in a school (Johnson, 2003; O’Neill, 2000), it is a prerequisite for success (e.g., for the success of: high school English departments (Fuller *et al.*, 2005)); school-based management (Murphy and Beck, 1995); and integration of information and communication technology (ICT) (Robertson *et al.*, 2006). In addition, a trusting and collaborative environment (or naming and framing to clarify opinions or inclusion, dependency, conflict, counterdependency, and trust) is a necessary prerequisite, but not the end of the process. It is important to heed Dalin and Rust’s (1983: 91) warning that it “is fatal to assume that the school is learning only if it is undergoing change. More crucial is its ability to monitor its own system constantly, and accept as well as reject new practices and new products as they are found to be appropriate and inappropriate (that is, to be a learning organization).” The second stage concerns a professional community. A professional community involves a shared mission and shared norms and values including valuing differences and diversity, a focus on implementation, and continuous enhancement of learning for all students, deprivatization of practice, collaboration, and critical reflective dialog, especially that based on monitoring and performance data. However, a shared and monitored mission (or analysis and integration of structure, work, and productivity) is also not the end of the process.

The final stage relates to the presence of a capacity for change, and innovation. It means using the trusting and collaborative environment and a shared and monitored mission as a safe base from which to venture forth to take the initiative and risks. In brief, it is about the journeys of learning, learning one’s way forward, and learning as change. All stages are required for effective OL, with each stage and each transition between them able to be facilitated by appropriate leadership and ongoing, optimistic, caring, nurturing professional development.

OL offers a way for a school to make sense of paradox, to ride the seesaws of change (Handy, 1994) and increased diversity, to establish and maintain a sense of connectedness, direction, and continuity, and reduce variation in performance. OL offers the potential of stability for change, an opportunity for schools to move ahead without losing their roots (Peters, 1987). In other words, it is a change

strategy with the potential to address current change agendas. Essentially, it is a change strategy worthy of further development and analysis. Care will be needed, however, especially in cross-cultural situations. Our world is shrinking. Access to international information, institutions, and expertise increases the likelihood of policies and solutions to problems employed in one national context to be considered for use in other nations. Hallinger (1995) is one of a growing number who question this development. He wonders why, for example, “Western treatises on the nature of leadership . . . are often transferred across cultures with relatively little concern for cultural validity” (Hallinger, 1995: 4). It may be that, compared with the Western emphasis on separateness, things, individuality, the answer, and short time frame, Chinese cultures (Senge, *et al.*, 1994), for instance, with their greater appreciation of the interconnectedness of life, processes, community, interdependencies, and long time frame, or Scandinavian countries (Johansson, 2004), with their priority for democratic school leadership, would be more amenable to the concept of OL than the Australian and North American material that dominate this article.

See also: Professional Learning Community.

Bibliography

- Argyris, C. (1993). *Knowledge for Action: A Guide to Overcoming Barriers to Organizational Change*. San Francisco, CA: Jossey-Bass.
- Argyris, C. and Schön, D. (1978). *Organizational Learning. A Theory of Action Perspective*. Reading, MA: Addison-Wesley.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist* **28**(2), 117–148.
- Beer, M., Eisenstat, R., and Spector, B. (1990). *The Critical Path to Corporate Renewal*. Boston, MA: Harvard Business School Press.
- Bell, L., Bolam, R., and Cubillo, L. (2002). *A Systematic Review of the Impact of School Head Teachers and Principals on Student/Pupil Outcomes*. London: EPPI-Centre, Social Science Unit, Institute of Education, University of London.
- Bolam, R., McMahon, A., Stoll, L., Thomas, S., and Wallace, M. (2005). Creating and sustaining effective professional learning communities. *UK Department for Education and Skill Research Report RR637*. London: UK Department for Education and Skills.
- Boyd, V. and Hord, S. (1994). Principals and the new paradigm: Schools as learning communities. *Paper Presented to AERA*. New Orleans, LA, April.
- Collinson, V., Cook, T., and Conley, S. (2006). Organizational learning in schools and school systems: Improving learning, teaching and leading. *Theory into Practice* **25**(2), 107–116.
- Connor, C. (2005). *Within-School Variation*. Nottingham: NCSL.
- Cousins, B. (1996). Understanding organizational learning for school leadership and educational reform. In Leithwood, K., Chapman, J., Carson, D., Hallinger, P., and Hart, A. (eds.) *International Handbook of Educational Leadership and Administration*, pp 589–652. Dordrecht: Kluwer.
- Dalin, P. and Rust, V. (1983). *Can Schools Learn?* Windsor, UK: NFER-Nelson.
- Deal, T. (1990). Healing our schools: Restoring the heart. In Lieberman, A. (ed.) *Schools as Collaborative Cultures: Creating the Future Now*, pp 127–149. New York: Falmer Press.
- Drucker, P. (1995). *Managing in a Time of Great Change*. New York: Penguin.
- Easterby-Smith, M., Burgoyne, J., and Araujo, L. (eds.) (1999). *Organizational Learning and the Learning Organization*. London: Sage.
- Fullan, M. (1993). *Change Forces*. London: Falmer Press.
- Fullan, M. (1994). Teachers as leaders. *Report to the Joint Legislative Education Oversight Committee and the North Carolina State Board of Education*.
- Fullan, M. (1995). The school as a learning organization: Distant dreams. *Theory into Practice* **34**(4), 230–235.
- Fuller, A., Hodkinson, H., Hodkinson, P., and Unwin, L. (2005). Learning as peripheral participation in communities of practice: A reassessment of key concepts in workplace learning. *British Educational Research Journal* **31**(1), 49–68.
- Gitlin, A. and Margonis, F. (1995). The political aspect of reform: Teacher resistance as good sense. *American Journal of Education* **193**(4), 377–405.
- Glickman, C. (1990). Open accountability for the 90's: Between the pillars. *Educational Leadership* **47**(7), 38–42.
- Goddard, R. (2002). Collective efficacy and school organization: A multilevel analysis of teacher influence in schools. *Theory and Research in Educational Administration* **1**, 169–184.
- Goddard, R., Hoy, W., and Woolfolk Hoy, A. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher* **33**(3), 1–13.
- Greene, M. (1995). Recapturing the political dimension. In Ginsburg, M. (ed.) *The Politics of Educators' Work and Lives*, pp 1–10. New: Garland Publishing.
- Hajnal, V., Walker, K., and Sackney, L. (1998). Leadership, organizational learning and selected factors related to the institutionalisation of school improvement initiatives. *Alberta Journal of Educational Research* **44**(1), 70–89.
- Hallinger, P. (1995). Culture and leadership: Developing an international perspective on educational administration. *UCEA Review* **36**(2), 1, 4–5, and 10–13.
- Handy, C. (1994). *The Age of Paradox*. Boston, MA: Harvard Business School Press.
- Hargreaves, A. (1995). Renewal in the age of paradox. *Educational Leadership* **52**(7), 12–19.
- Hipp, K. and Huffman, J. (2003). Professional learning communities: Assessment, development, effects. *Paper Presented at ICSEI*. Sydney, NSW, January.
- Hord, S. (1997). *Professional Learning Communities: Communities of Continuous Inquiry and Improvement*. Austin, TX: Southwest Educational Development Laboratory.
- Hord, S. (2004). *Learning Together: Leading Together: Changing Schools through Professional Learning Communities*. New York: Teachers College Press.
- Johansson, O. (2004). A democratic, learning and communicative leadership? *Journal of Educational Administration* **42**(6), 697–707.
- Johnson, B. (2003). Teacher collaboration: Good for some, not so good for others. *Educational Studies* **29**(4), 337–350.
- Johnston, C. (1997). Leadership and learning organization in self-managing schools. Unpublished EdD thesis, University of Melbourne.
- Kanter, R. (1991). Change: Where to begin. *Harvard Business Review* **69**(4), 8–9.
- Keating, D. (1995). *The Learning Society in the Information Age*. Working Paper No. 2. Toronto, ON: The Canadian Institute for Advanced Research Program in Human Development.
- Leithwood, K. and Aitken, R. (1995). *Making Schools Smarter*. Thousand Oaks, CA: Corwin.
- Leithwood, K., Jantzi, D., and Steinbach, R. (1995). *An Organizational Learning Perspective on School Responses to Central Policy Initiatives*. Toronto, ON: OISE.
- Leithwood, K., Jantzi, D., and Steinbach, R. (1998). Leadership and other conditions which foster organizational learning in schools. In Leithwood, K. and Louis, K. S. (eds.) *Organizational Learning in Schools*, pp 67–90. The Netherlands: Swets and Zeitlinger.
- Leithwood, K. and Louis, K. S. (eds.) (1998). *Organizational Learning in Schools*. The Netherlands: Swets and Zeitlinger.

- Leithwood, K., Day, C., Sammons, P., Harris, A., and Hopkins, D. (2006). *Successful School Leadership: What It Is and How It Influences Pupil Learning. Report to the UK Department for Education and Skills*. Nottingham: Nottingham University.
- Levin, B. (1994). Schools coping with a changing world. *A paper presented at the American Educational Research Association Annual Conference*. New Orleans, April.
- Lieberman, A. (1995). *The Work of Restructuring Schools*. New York: Teachers College Press.
- Louis, K. (1994). Beyond managed change: Rethinking how schools improve. *School Effectiveness and Improvement* 5(1), 2–24.
- Louis, K., Kruse, S., and Bryk, A. S. (1995). *Professionalism and Community*. Thousand Oaks, CA: Corwin.
- Marks, H. M., Louis, K. S., and Printy, S. M. (2000). The capacity for organizational learning. In Leithwood, K. (ed.) *Understanding Schools as Intelligent Systems*, pp 239–265. Stamford, CT: JAI Press.
- Mawhinney, H., Hass, J., and Wood, C. (2005). School district design for improving instructional capacity: Assessing teachers' collective efficacy beliefs and perceptions of school professional learning community maturity. *A Paper Presented to AERA*. Montreal, April.
- Miles, M. (1993). 40 years of change in schools: Some personal reflections. *Educational Administration Quarterly* 29(2), 213–248.
- Mintzberg, H. (1994). *The Rise and Fall of Strategic Planning*. New York: Free Press.
- Mitchell, C. (1995). Teachers learning together: Organizational learning in an elementary school. Unpublished PhD thesis, University of Saskatchewan, Saskatoon.
- Mulford, B. (1998). Organizational learning and educational change. In Hargreaves, A., Lieberman, A., Fullan, M., and Hopkins, D. (eds.) *International Handbook of Organizational Change*, pp 616–641. Norwell, MA: Kluwer.
- Mulford, B. (2003). School leaders: Challenging roles and impact on teacher and school effectiveness. Commissioned Paper by the Education and Training Policy Division, OECD, for the Activity 'Attracting, Developing and Retaining Effective Teachers', Paris. <http://www.oecd.org/dataoecd/61/61/2635399.pdf> (accessed April 2009).
- Mulford, B., Silins, H., and Leithwood, K. (2004). *Leadership for Organizational Learning and Student Outcomes: A Problem-Based Learning Approach*. Dordrecht: Kluwer.
- Murphy, J. and Beck, L. (1995). *School-Based Management as School Reform: Taking Stock*. Thousand Oaks, CA: Corwin Press.
- Newman, F. and Wehlage, G. (1995). *Successful School Restructuring*. Alexandria, VA: Association for Supervision and Curriculum Development.
- OECD (2000). Programme for international student assessment (PISA). <http://www.pisa.oecd.org> (accessed April 2009).
- O'Neill, J. (2000). 'So that I can more or less get them to do things they really don't want to do'. Capturing the situational complexities of the secondary school head of department. *Journal of Educational Enquiry* 1(1), 13–34.
- Peters, T. (1987). *Thriving on Chaos*. London: Harper and Row.
- Redding, J. and Catalanello, R. (1994). *Strategic Readiness*. San Francisco, CA: Jossey-Bass.
- Robertson, M., Grady, N., Fluck, A., and Webb, I. (2006). Conversations toward effective implementation of information communication technologies in Australian schools. *Journal of Educational Administration* 44(1), 71–85.
- Ross, J., Hogaboam, A., and Gray, P. (2004). Prior student achievement, collaborative school processes and collective teacher efficacy. *Leadership and Policy in Schools* 3(3), 163–188.
- Rusch, E. (2005). Institutional barriers to organizational learning in school systems: The power of silence. *Educational Administration Quarterly* 41(1), 83–120.
- Senge, P. (1990). *The Fifth Discipline: The Art and Practice of Organizational Learning*. New York: Doubleday.
- Senge, P., Roberts, C., Ross, R., Smith, B., and Kleiner, A. (1994). *The Fifth Discipline Fieldbook: Strategies and Tools for Building a Learning Organization*. New York: Doubleday.
- Sheppard, B. and Brown, J. (2000). The transformation of secondary schools into learning organizations. In Leithwood, K. (ed.) *Understanding Schools as Intelligent Systems*, pp 293–314. Stamford, CT: JAI Press.
- Silins, H. and Mulford, B. (2002a). Leadership and school results. In Leithwood, K. and Hallinger, P. (eds.) *Second International Handbook of Educational Leadership and Administration*, pp 561–612. Norwell, MA: Kluwer.
- Silins, H. and Mulford, B. (2002b). Schools as learning organizations: The case for system, teacher and student learning. *Journal of Educational Administration* 40(5), 425–446.
- Silins, H. and Mulford, B. (2004). Schools as learning organizations: Effects on teacher leadership and student outcomes. *School Effectiveness and School Improvement* 15(3–4), 443–466.
- Silins, H., Mulford, B., and Zarins, C. (2002). Organizational learning and school change. *Educational Administration Quarterly* 38(5), 613–642.
- Somech, A. (2005). Teachers' personal and team empowerment and their relations to organizational outcomes: Contradictory or compatible constructs? *Educational Administration Quarterly* 41(2), 237–266.
- Stoll, L., Fink, D., and Earl, L. (2003). *It's about Learning: And It's about Time*. London: Falmer Press.
- Voulalas, Z. and Sharpe, F. (2005). Creating schools as learning communities: Obstacles and processes. *Journal of Educational Administration* 43(2), 187–208.
- Watkins, K. and Marsick, V. (1993). *Sculpting the Learning Organization: Lessons in the Art and Science of Systematic Change*. San Francisco, CA: Jossey Bass.
- Wheeler, S. and Tilin, F. (1999). The relationship between faculty group development and school productivity. *Small Group Research* 30(1), 59–81.

Further Reading

- Mulford, B. (2008). *The Leadership Challenge: Improving Learning in Schools*. Camberwell, Melbourne: Australian Council for Educational Research, Australian Education Review No. 53. www.acer.ed.au/research_reports/AER.html.
- Sackney, L., Walker, K., and Hajnal, V. (1995). Organizational learning, leadership, and selected factors relating to the institutionalization of school improvement initiatives. *A Paper for AERA*. San Francisco, April.
- Silins, H. and Mulford, B. (2007). Leadership and school effectiveness and improvement. In Townsend, T. (ed.) *International Handbook of School Effectiveness and Improvement*, pp 635–658. Dordrecht, The Netherlands: Springer.

Professional Learning Community

L Stoll, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Capacity – An individual's, group's, or organization's ability to engage in and sustain learning and improvement.

Co-construction of knowledge – Jointly creating new understanding and meaning.

Collective knowledge creation – Whereby a learning community interacts, engages in serious dialog and deliberates about information and data, interpreting it communally and distributing it among them.

Collective learning – In which a community interacts, engages in serious dialog and deliberates about all the information it has and data it collects, interpreting it communally and distributing it among themselves and collectively creating new knowledge.

Collegiality – Joint work promoting the kind of interdependence between colleagues that allows serious challenge and adjustment of practice.

Community of practice – A community focused on a topic of shared interest where participants gradually absorb and are absorbed in a 'culture of practice,' giving them exemplars, leading to shared meanings, a sense of belonging and increased understanding.

Deprivatization of practice – Opening up one's practice to the scrutiny of others.

Distributed leadership – Reciprocal leadership actions of a number of people; teachers and others involved in the practices of leadership.

Interdependence – Belief among members of a community that achieving better teaching practices would be considered impossible without collaboration.

Micropolitics – Internal politics and power issues within a community's relationships.

Reflective dialog – Conversations where tacit knowledge is converted into shared knowledge.

Reflective professional inquiry – Seeking new knowledge by frequently examining practice, through mutual observation and case analysis, joint planning and curriculum development, applying new ideas and information to problem solving and solutions that will address students' needs.

Social capital – Groups, networks, norms, and trust that are available to people to help them achieve their goals.

Despite shades of interpretation in different contexts, there is increasing international consensus that the term professional learning community broadly refers to an inclusive and mutually supportive group of people with a collaborative, reflective, and growth-oriented approach toward investigating and learning more about their practice in order to improve students' learning. The history of exploration is relatively recent with its main origins in North America, and largely focuses on schools and teachers within schools. The last few years, however, have seen greater attention in many countries from researchers, trying to gain greater nuanced and contextualized understanding of professional learning community, and policymakers who view its potential for the capacity building needed to implement educational reform. This has led to the development of various programs and initiatives promoting the development of professional learning communities. This article probes the meaning and purpose of professional learning community, membership, identified characteristics, levels of impact, and process and processes of development.

Meaning and Purpose

By examining the words making up the term, it is possible to see its theoretical underpinnings and its development.

Community

Core to the concept is the notion of community. The focus is not just on individual teachers learning, but of a sociocultural notion of learning within a collaborative community context – the notion of collective learning in a community of practice (Wenger, 1998). Features of communities – shared beliefs and understandings, interaction and participation, interdependence, concern for individual and minority views, and meaningful relationships through personal connections – are important to a professional learning community. The cultural norms of community are those of collaboration, interpersonal caring, and mutual support. Shared memory develops and is passed on to newcomers, and those involved view the group as a serious collective enterprise.

Professional

The word professional emphasizes that the community's work is underpinned by: a specialized and technical

knowledge base; a service ethic that orients members toward client needs; strong collective identity through professional commitment; and professional autonomy through collegial control over practice and professional standards (Talbert and McLaughlin, 1994). Changing notions of professionalism have also served to strengthen the link between the development of professionals and the development of the whole school in which they are located (Hargreaves, 1994), with increasing interest in understanding whole-school professional community in secondary schools.

Learning

Although described as a professional community of learners (Astuto *et al.*, 1993), much of the research in the United States still focuses on what has become known as professional community. Insertion, more recently, of the word learning has coincided with a shift in emphasis toward the objective of improvement – a learning community with a collective purpose of enhancing student learning. Here, for some investigators and proponents, there is a connection with other fields of study, including school improvement and social learning theory. Traditional forms of professional development emphasize opportunities for individuals to hone their knowledge and skills in and out of their school settings. Other means of collaborative learning in the workplace – including peer observation, coaching, and collaborative forms of action research – are now also being examined by some of those studying professional learning community. Collective learning in the context of professional learning communities also involves social learning, and working together toward a common understanding of concepts and practices. What is held in common does not supplant what teachers learn individually but supplements it and collaborative professional learning to support and spread learning throughout the entire community.

Nature of Membership

During the 1990s, those exploring professional community were largely concerned with schools and departments as mediating contexts for teaching, and concluded that these can differ strikingly from one another in the strength of their professional community. Examining professional community within high school departments highlighted how professional community can operate at different levels within a school, and parallels can be drawn with other work exploring differences in the culture of subject departments. Due to its smaller size, professional community within an elementary school is more usually concerned with the whole school.

A broadening of perspectives has begun to emerge of who might be counted as members of a professional learning community. Some would still argue that professional learning community should be interpreted as referring to

groups of teachers supported by leaders. Within this perspective, allowances are sometimes made for attention toward the involvement of support staff including counselors and teachers' aides, both of who have previously been seen as having a peripheral role to play in school development.

Another variation broadening the meaning of professional learning community is based on the argument that the increased complexity of a fast-changing world has brought new challenges for schooling that are too great for those in any one school to address alone. The idea of community here is extended to a network of schools, operating as a wider professional learning community, or learning network. These networks differ from those serving teachers with particular shared interests, such as subject networks, in that their focus is on whole-school change (Veugelers and O'Hair, 2005). In this sense, networks of principals, with a shared focus on specific school-wide change, would also be included within this category. Notably, it appears that more successful networked learning communities are ones that link and support professional learning communities within schools, where the school's professional learning community is the strong local locus for changes in practice but is enhanced by the strength of ideas and support from the learning network.

The third variation takes as its starting point the argument that schools exist within a wider social context. Other people with a valid stake in making a difference to students' learning and life chances therefore need to participate in professional learning communities. Expanding the concept of professional learning community in such a way means the inclusion of a wider range of members and drawing on broader knowledge bases than traditional teaching knowledge bases (Stoll and Louis, 2007). Some current investigations in this area include links between schools and their local communities, and connections within schools and school districts of staff from different agencies, including, for example, health and social welfare. Here, the issue of relationships and building social capital comes to the forefront. In particular, it highlights the necessity of bridging and linking social capital, where connections are made with people who bring different backgrounds and values and where power differentials may get in the way of equal relationships. For a professional learning community, this may mean challenging the status system of whose knowledge counts most. Work thus far in this area suggests that the difficulty of doing this cannot be underestimated.

Broadening perceptions of professional learning communities means rethinking notions of specific location for professional learning community. Professional learning communities can cross boundaries, both the social differentiations that develop between groups within the school, and the clearer borders that separate the school's members from those in the community and in other schools. As with any boundary crossing, expanding ideas about who belongs presents challenges to the existing culture.

As the notion of broadening perceptions of professional learning community is new and contested, the focus of the remainder of the article largely relates to the knowledge base about professional learning community within schools.

Professional Learning Community Characteristics

There is no universal consensus on the precise characteristics of professional learning communities, but analysis of the literature suggests that they share five key characteristics or features, which operate together. These are:

1. *Shared values and vision.* A shared vision, sense of purpose, and focus on all students' learning that is undeviating is critical to ensure that individual autonomy does not reduce teacher efficacy when teachers cannot count on colleagues to reinforce objectives. This shared value base is intended to provide a framework for shared, collective decision making.
2. *Collective responsibility.* Members of a professional learning community consistently take collective responsibility for student learning. Such collective responsibility appears to help sustain commitment, putting peer pressure and accountability on those who do not contribute equally, and eases isolation.
3. *Reflective professional inquiry.* In a professional learning community, members seek new knowledge by frequently examining their practice, through mutual observation and case analysis, joint planning, and curriculum development. They also engage in reflective dialog, conversations where tacit knowledge is converted into shared knowledge, and then regularly apply new ideas and information to problem solving and solutions that address students' needs. Deprivatization of practice (Louis *et al.*, 1995) lies at the heart of this.
4. *Collaboration.* Staff are involved in developmental activities that go beyond superficial exchanges of help or assistance and have consequences for several people: for example, joint review and feedback. Feelings of interdependence are central to such collaboration: achieving better teaching practices would be considered impossible without such collaboration. While micropolitics may exist, conflicts are managed more effectively, fitting with the notion that in communities, divergent views are expected.
5. *Group, as well as individual, learning is promoted.* All teachers are learners with their colleagues, and professional self-renewal is a communal feature. Collective learning is also evident, through collective knowledge creation (Louis, 1994), whereby the school learning community interacts, engages in serious dialog, and deliberates about information and data, interpreting them communally and distributing them among themselves.

Impact

Interest in the impact of a professional learning community is focused at a number of different levels. Historically, the main focus was on improving teachers' morale and practice, with further interest in the difference professional learning community might make for the whole organization. More recently, attention is also increasingly being turned toward the impact on students. Debate rages about whether professional community is and should be outcome oriented, rather than a significant goal in its own right. The argument on one side is that the value of community has to be unraveled from instrumental values of improving measurable student outcomes because community is concerned with the quality of life in schools. By contrast, others take a stand that all improvement and development efforts in schools must have an ultimate goal of improving outcomes for students.

Enhancing Teacher Morale

With issues of an aging teaching staff and teacher recruitment facing schools in many countries, this form of impact has an appeal. Job satisfaction is clearly important, and it appears that professional learning community can sometimes act as a buffer against the kind of issues causing teachers to leave the profession. Being able to feel part of a school-wide community may have greater resonance in today's world that is characterized by loss of traditional neighborhood community, changing family structures, greater mobility, and uncertainty about meaning and values.

Teacher Learning and Improvement in Practice

Changing what teachers do in the classroom has been one of the greatest challenges of school improvement. This is a major goal of professional learning community. Increasingly, international evidence suggests that where professional learning community is developed, teachers' knowledge base can be enhanced and there is a significant impact more generally on teachers' classroom work. It seems that in schools with a genuine sense of community, an increased sense of work efficacy leads to increased classroom motivation and work satisfaction, with subsequent increases in collective responsibility for student learning. The path between professional learning community and improvement in practice is, therefore, not necessarily direct. Rather, professional learning community appears to foster change in practice by creating an environment that supports innovation and experimentation, and this is what actually promotes teacher learning (Bryk *et al.*, 1999).

Sustained, collaborative continuing professional development is highlighted as one mechanism for promoting professional learning community. Studies on its effect on

teaching and learning conclude that collaborative professional development opportunities frequently have a positive impact on: teachers' confidence; their beliefs in their power to make a difference to students' learning; enthusiasm for collaborative working, despite initial anxiety about opening up their practice to the potential scrutiny of others; and commitment to changing practice and willingness to try new things. Several features of collaborative professional development, working in combination, appear to be linked to these positive outcomes. These include: use of external expertise linked to school-based activity; observation; feedback, frequently based on observation; emphasis on peer support rather than leadership by supervisors; scope for teacher participants to identify their own learning focus; processes to encourage, extend, and structure professional dialog; and processes for sustaining professional learning over time to enable teachers to embed the practices in their own classroom settings.

Organizational Capacity Building and Improvement

A recent major interest of policy makers is capacity building to ensure that educational reform is implemented, spread throughout the system and has a positive impact on students' outcomes. The level of the school is important here because it is the level at which performance is judged in many countries and jurisdictions. Improvements in individual teachers' practice does not change practice in whole schools. Understanding the means by which learning and development is dispersed throughout schools has, therefore, been seen as something of a holy grail. It is now generally agreed that professional learning communities may provide at least some of the answers because, where they are functioning at a high level, they appear to have the capacity for learning, inquiry, change, and innovation. It is generally thought that this is because they are a means to achieving long-term cultural change in an organization. Establishing lasting collaborative cultures that constantly focus on building the capacity for continuous learning and improvement are critical in a complex and fast-changing world. They are also particularly important in helping spread and sustain improvements in individual teachers' practice and morale.

Student Outcome Oriented

Some proponents are clear that the goal is not being a professional learning community, but to enhance students' learning. While teacher learning is considered important, the purpose of promoting staff learning is seen as maximization of the learning of students. Taken from this perspective, increases in staff morale and improved practice become intermediate outcomes. Analyses of student achievement in

schools have begun to indicate that the presence of professional community centered on student learning makes a significant difference to measurable student achievement (e.g., Bolam *et al.*, 2005). A lesser subject of exploration has been whether standardized achievement tests are the right measures of the difference made by professional learning community to students' learning or whether broader measures of student engagement with the school as a learning community might be more appropriate indicators. This partly depends on the focus of the professional learning community, and studies of smaller professional learning communities, in this case subject departments, have found specific links between a professional learning community orientation of the teachers and improved results in the specific subjects. The aggregate of extensive research in the school effectiveness tradition, however, has suggested that while intermediate variables like the professional relationship between staff and extent to which they work collaboratively are significant, they account for less variation in effectiveness than other within-school factors directly related to the teaching and learning process (Creemers, 1994).

Development of Professional Learning Community

Considerable effort has been and continues to be channeled toward gaining greater understanding of the specific processes involved in developing professional learning community, as well as the process of development over time. Approaches differ, with some carrying out sets of research projects over a period of years, and others evaluating the efforts of development projects based on earlier research. More is known about specific processes of professional learning community development than how the actual process of change occurs over time.

Processes of Professional Learning Community Development

Developing professional learning communities depends on working on a number of processes. While there is no precisely agreed set of processes, and some differentiation between researchers in what might be described as characteristics, features, or processes, there is concurrence about the broad nature of the kinds of processes involved.

Leadership of learning

There is general agreement that developing professional learning community depends on active support of leadership at different levels. In particular, principal commitment distributed leadership and coordination of professional learning appear significant.

Principal leadership

Through their leadership and management, principals create the conditions within which professional learning community either thrives or is diminished. The school principal has a major impact on the nature of school culture, and the kind of deep learning processes on which professional learning community depends are best supported and nurtured in a culture that values these types of processes and creates opportunities for them to occur. Promoting high expectations and teacher as well as student learning are critical facets, and there is also evidence that modeling learning can be important. Closely connected is an orientation toward, and promotion of a culture of, inquiry. This is particularly evident in more recent studies carried out during a period where access to, and expectation about use of, data are increasingly part of schools' reality.

Distributed leadership and leadership of professional learning

Given that collective responsibility is a key feature of professional learning community, and with the nature of demands on school leadership having intensified, it is perhaps not surprising that accomplishing workplace responsibility depends on reciprocal leadership actions of a number of people (Gronn, 2003). Exploration of this aspect of professional learning community is at a relatively early stage, with some suggestions that teachers' pedagogic leadership may work in parallel with principals' strategic leadership.

With the inclusion of the word learning in the expression, has come a greater focus on teachers' professional learning and, along with it, some attention to the leadership and coordination of professional learning activities. Here, what has increasingly been seen across a number of Organization for Economic Cooperation and Development (OECD) countries is a greater alignment between the focus on developing individual teachers' practice and a focus on managing the development of the whole school as a learning community as part of a more general approach to school change and improvement (Centre for Educational Research and Innovation (CERI), 2001).

Other social resources

Developing professional learning community is a human and emotional enterprise with all of the associated complexity of bringing about change, and there is agreement that making effective use of human and social resources is a key dimension. Working together productively depends on positive relationships and collegiality, joint work promoting the kind of interdependence between colleagues that allows serious challenge, and adjustment of practice (Little, 1990). Some commentators are quick to point out that there is no necessity for personal friendships, but a dynamic of dysfunctional relationships has been shown to

have a negative effect on a school. Without a climate of trust and respect from colleagues, teachers do not feel safe to take the risks associated with collaboration, open dialog and deprivatization of practice. Cautions are offered, however, that if trust provides a context for predictability, stability, assurance, and safety, it can inhibit innovative activity by keeping individuals satisfied with their current situation. Furthermore, it is possible to romanticize the notion of collegiality while, in reality, a political dimension of change means that the balance of power in relationships, both between teachers and between teachers and leaders can affect the nature of collaboration and the potential for real collegiality.

Structural resources

Schools are bounded by structures that shape their capacity to develop a professional learning community. In particular, the consensus appears to be that time, space, communication mechanisms, and interdependent roles are important. Teacher talk and exchange about and joint reflection on professional issues are critical elements of the collaborative activity necessary to develop professional learning community. The provision of time is required to facilitate this, as well as a physical structure where it is easier for colleagues to connect on a regular basis. Physical proximity has been found to facilitate opportunities for professional exchange. This is one reason why it is generally more challenging to develop professional learning community across a whole secondary school when teachers are based in subject departments that are often located some distance from the whole faculty lounge and share workrooms with colleagues in their subject department. Regular meetings with agendas focused on learning, and team teaching or collaborative lesson design, appear to be other structural mechanisms often cited as aids to development of professional learning community.

Promotion of learning

It could be argued that learning in professional learning communities can be assumed, but the evidence suggests that professional community, per se, can be a means by which weak practice is distributed and recycled among colleagues. Some models of professional learning community processes, therefore, specifically emphasize teacher-learning processes, some of which derive from sociocultural theories and others which are more aligned with learning and, in particular, social learning theories.

Professional development, work-based learning, and inquiry-based learning

Those who include the promotion of professional learning and development opportunities have argued that if the professional learning community is to be intellectually vigorous, members need a solid basis of expert knowledge and skills, strongly emphasizing the professionalization of

teachers' work through increasing expert knowledge. It is now generally agreed, however, that traditional forms of continuous professional development, consisting of short training courses, do little more than raise awareness of issues. By contrast, professional learning based on self-development, reflective practice, and work-based learning supported by peers is now widely seen as more effective.

With a broader definition of professionalism, and increased accountability, data analysis and use are now an increasingly important part of teachers' jobs in many countries and jurisdictions and are viewed in some models as forms of professional learning. As more data and evidence become available to schools, developing an inquiry orientation toward analysis and use of student and other data appear to take time, although where professional learning community is more developed, data collection and analysis appear to become part of an iterative process of reflection, and change.

Collective learning through knowledge creation and transfer

Learning within professional learning communities involves active deconstruction of knowledge through reflection and analysis, and its reconstruction through action in a particular context, as well as co-construction through collaborative learning with peers. Wenger (1998) proposes that when learning in communities of practice, participants gradually absorb and are absorbed in a culture of practice, giving them exemplars, leading to shared meanings, a sense of belonging, and increased understanding. Professional learning community is distinguished by an emphasis on collective learning. While effective teaching and learning depends on individual teachers, in addition, individual knowledge, skills, and attitudes are combined in a collective enterprise. Organizational learning theory also offers insights about these connections, suggesting that as the school community interacts, engages in serious dialog, and deliberates about all the information it has and data it collects, they interpret it communally distributing it among themselves and collectively creating new knowledge. Dialog also appears to be a key link, being seen as the process through which the gap between individual and organizational learning is bridged.

As the definition of professional learning communities begins to broaden, some models are also beginning to include processes whereby those within a specific organizational community interact and draw on colleagues in other parallel communities and other external agents, although this aspect is at an early stage of exploration.

Developmental Process over Time

A recent avenue of investigation is considering how a professional learning community may go through different phases of development. Models differ. Sequential

patterns of stages of organizational life cycles have also been examined, although there has been some question as to whether these can be applied to development and sustainability of learning communities where a key goal is continuous learning rather than implementing a specific change initiative. Others draw on the knowledge base about stages of educational change and levels of use of innovations, examining how the characteristics and processes involved appear to change over time as the professional learning community broadly moves through the phases. Inevitably, it is not a linear path and there is some suggestion that it is a more fluid, rather than fixed, entity, perennially evolving with accumulating collective experience (Bolam *et al.*, 2005), although little is still known about sustainability. The attempt of such research is to map how the professional learning community develops and grows, variously with a view to informing policy and/or helping practitioners and change facilitators identify and plan for the development of a professional learning community through the provision of research-based materials and guidelines. There is some suggestion that, where a broader perspective of professional learning community is taken, building of social capital is necessary before attending to other processes of professional learning community development (Mulford, 2007), but this remains to be tested. It may be, however, that a community has to be at a certain stage of readiness before it is able to embark on the collective learning that characterizes professional learning community.

Summary

Professional learning community, therefore, suggests that the focus is not just on individuals' learning but on professional learning within the context of a cohesive group, that focuses on collective knowledge and growth, and occurs within an ethic of interpersonal caring that permeates the life of the community. In essence, professional learning community implies that what colleagues do together outside of classrooms is critically important in supporting what they do inside to promote student learning, as well as to enhance their own and their school's development. It is a cultural way of working oriented toward and building capacity for continuous and sustainable learning. The general assumption by those exploring the area, as well as proponents with policy intentions, is that working toward professional learning community is a worthwhile endeavor, although not one without many challenges. There remains, however, concern that an overriding demand for evidence of student achievement outcomes will taint serious attempts to enhance teacher learning and growth within a supportive community and overshadow what might, otherwise, be a potentially profound mechanism for educational change.

See also: Contemporary Approaches to Teacher Professional Development; Social Capital, Educational Institutions and Leadership; Studies of School Improvement in Developing Countries.

Bibliography

- Astuto, T. A., Clark, D. L., Read, A. -M., McGree, K., and Fernandez, L. (1993). *Challenges to Dominant Assumptions Controlling Educational Reform*. Andover, MA: Regional Laboratory for the Educational Improvement of the Northeast and Islands.
- Bolam, R., McMahon, A., Stoll, L., et al. (2005). Creating and sustaining effective professional learning communities. *DfES Research Report RR637*. University of Bristol.
- Bryk, A., Camburn, E., and Louis, K. S. (1999). Professional community in Chicago elementary schools: Facilitating factors and organizational consequences. *Educational Administration Quarterly* **35** (Supplement), 751–781.
- Centre for Educational Research and Innovation (CERI) (2001). *New School Management Approaches*. Paris: OECD.
- Creemers, B. (1994). *The Effective Classroom*. London: Cassell.
- Gronn, P. (2003). Leadership: Who needs it? *School Leadership and Management* **23**(3), 267–290.
- Hargreaves, D. (1994). The new professionalism: The synthesis of professional and institutional development. *Teaching and Teacher Education* **10**(4), 423–438.
- Little, J. W. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teachers College Record* **91**(4), 509–536.
- Louis, K. S. (1994). Beyond 'managed change': Rethinking how schools improve. *School Effectiveness and School Improvement* **9**(1), 1–27.
- Louis, K. S. and Kruse, S. D. (1995). *Professionalism and Community: Perspectives on Reforming Urban Schools*. Thousand Oaks, CA: Corwin.
- Mulford, B. (2007). Building social capital in professional learning communities: Importance, challenges and a way forward. In Stoll, L. and Louis, K. S. (eds.) *Professional Learning Communities: Divergence, Detail and Dilemmas*, pp 166–180. Maidenhead: Open University Press/McGraw-Hill Education.
- Stoll, L. and Louis, K. S. (eds.) (2007). *Professional Learning Communities: Divergence, Detail and Dilemmas*. Maidenhead: Open University Press/McGraw-Hill Education.

- Talbert, J. E. and McLaughlin, M. (1994). Teacher professionalism in local school contexts. *American Journal of Education* **102**, 123–153.
- Veugelers, W. and O' Hair, M. J. (eds.) (2005). *Network Learning for Educational Change*. Maidenhead: Open University Press/McGraw-Hill Education.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. New York: Cambridge University Press.

Further Reading

- Bryk, A. and Schneider, B. (2002). *Trust in Schools*. New York: Sage.
- Hipp, K. K. (Guest ed.) (2006). *Special Issue: Leadership and Student Learning in Professional Learning Communities*. *Journal of School Leadership* **16**(5), 471–640.
- Hord, S. M. (1997). *Professional Learning Communities: Communities of Continuous Inquiry and Improvement*. Austin, TX: Southwest Educational Development Laboratory.
- King, M. B. and Newmann, F. M. (2001). Building school capacity through professional development: Conceptual and empirical considerations. *International Journal of Educational Management* **15**(2), 86–93.
- Little, J. W. (2002). Locating learning in teachers' communities of practice: Opening up problems of analysis in records of everyday work. *Teaching and Teacher Education* **18**(8), 917–946.
- Louis, K. S. and Marks, H. (1998). Does professional community affect the classroom? Teachers' work and student experience in restructured schools. *American Journal of Education* **106**(4), 532–575.
- McLaughlin, M. W. and Talbert, J. E. (2001). *Professional Communities and the Work of High School Teaching*. Chicago, IL: University of Chicago Press.
- Stoll, L., Bolam, R., McMahon, A., Thomas, S., and Wallace, M. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change* **7**(4), 221–258.
- Toole, J. C. and Louis, K. S. (2002). The role of professional learning communities in international education. In Leithwood, K. and Hallinger, P. (eds.) *Second International Handbook of Educational Leadership and Administration*, pp 245–280. Dordrecht: Kluwer.
- Westheimer, J. (1999). Communities and consequences: An inquiry into ideology and practice in teachers' professional work. *Educational Administration Quarterly* **35**(1), 71–105.
- Wiley, S. (2001). Contextual effects of student achievement: School leadership and professional community. *Journal of Educational Change* **2**(1), 1–33.

Transformational School Leadership*

K Leithwood, University of Toronto, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This article briefly describes the roots of transformational leadership and summarizes evidence of its effects. A school-specific formulation of this approach to leadership is then briefly outlined. The article concludes with a consideration of several common criticisms of this form of leadership.

The Roots of Transformational Leadership

The roots of transformational leadership are often attributed to James McGregor Burns' (1978) Pulitzer Prize-winning book entitled simply *Leadership*. In this book, Burns argued that transforming leadership "occurs when one or more persons *engage* with others in such a way that leaders and followers raise one another to higher levels of motivation and morality" (1978: 20). Such a conception of what effective leadership aims to accomplish diverged radically from the more common view of the day in which leaders engage in an exchange relationship with followers based on followers' individual, typically monetary, and otherwise extrinsic interests.

Following Burns's lead, Bernard Bass (1985) provided a more specific formulation of transformational leadership which, along with a survey-based measure, has enjoyed by far the greatest empirical attention in the ensuing several decades. Bass' initial formulation of transformational leadership consisted of five dimensions serving transformational purposes – charisma, inspirational leadership, individualized consideration, and intellectual stimulation. Bass' formulation also included several transactional dimensions reflecting the exchange relationships on which other leadership conceptions were based – contingent reward, management-by-exception, and *laissez-faire* leadership. Through these sets of practices, Bass claimed that transforming leaders:

convert followers to disciples; they develop followers into leaders. They elevate the concerns of followers on Maslow's (1954) needs hierarchy from needs for safety and security to needs for achievement and self-actualization, increase their awareness and consciousness of what is really important, and move them to go beyond their own self-interest for the

good of the larger entities to which they belong. The transforming leader provides followers with a cause around which they can rally (1985: 467).

Bass also argued that transformational leaders can be directive or participative, authoritarian, or democratic, depending on the context. In contrast to Burns' (1978) original view, Bass also claimed that transformational leadership augments rather than substitutes for transactional leadership. Finally, unlike many earlier theories of leadership which emphasized rational processes, transformational leadership theory emphasizes emotions and values, attributes importance to symbolic behavior, and conceptualizes the role of the leader as helping making events meaningful for followers (Yukl, 1989).

In addition to writing that refers explicitly to transformational leadership, writings about charismatic, visionary, cultural, and empowering concepts of leadership can be viewed as part of a transformational orientation, an orientation which assumes that the central focus of leadership ought to be the commitments and capacities of organizational members. Higher levels of personal commitment to organizational goals and greater capacities for accomplishing those goals are expected to result in extra effort and greater productivity.

Effects of Transformational Leadership

Evidence from studies in both school and nonschool contexts justifies continuing interest in this form of leadership. For example, meta-analyses of research in mostly nonschool contexts (Dumdum *et al.*, 2002; Lowe *et al.*, 1996) indicate that the overall effects of transformational leadership are typically positive and significant, although those effects are much stronger for subordinate perceptions of effectiveness than for objective measures of effectiveness, such as a company's financial performance.

A recent review of 32 studies of transformational leadership in school contexts (Leithwood and Jantzi, 2005) found positive and significant effects on teachers, for example, several forms of teacher commitment, teacher job satisfaction, changed classroom practices, collective teacher efficacy, and pedagogical or instructional quality. This form of leadership, the review indicated, also has a positive influence on elements of the school organization including its culture, planning, and strategies for change and organizational learning. While still limited in amount,

* Some sections of this paper are based on Leithwood, Day, Sammons, Harris, and Hopkins (2006).

evidence about transformational approaches to school leadership are reported to have promising effects on several important types of student outcomes (e.g., Leithwood and Jantzi, 1999; Silins and Mulford, 2004).

Transformational Leadership for Schools

This section provides an account of transformational leadership specifically designed for school organizations. The account draws on a substantial strand of work carried out by the author and his colleagues over a 15-year period aimed at developing a school-specific model of transformational leadership. By now, this model is extensively specified and evidence of its consequences reasonably well documented (e.g., Leithwood and Jantzi, 2005). One of the unique features of the model is its incorporation of leader behaviors associated with other approaches to leadership but which are both compatible with transformational intentions and demonstrably effective in both school and nonschool organizations. The model has four major dimensions – setting directions, developing people, redesigning the organization, and managing the instructional program – each of which includes three or four more specific sets of practices.

Setting Directions

This category of practices carries the bulk of the effort to motivate leaders' colleagues (Hallinger and Heck, 1998). It is about the establishment of moral purpose (Fullan, 2003; Hargreaves and Fink, 2006) as a basic stimulant for one's work. Most theories of motivation argue that people are motivated to accomplish personally important goals for themselves. For example, such goals are one of the four sources of motivation in Bandura's (1986) theory of human motivation.

Three more specific sets of practices are included in this category, all of which are aimed at bringing a focus to both the individual and collective work of staff in the school. Carried out skillfully, these practices are one of the main sources of motivation and inspiration for the work of the staff.

Building a shared vision

Building compelling visions of the organization's future is a fundamental task included in both transformational and charismatic leadership models. Bass' (1985) inspirational motivation is encompassed in this practice, a dimension that Podsakoff *et al.*, define as leadership behavior "aimed at identifying new opportunities for his or her unit. . . and developing, articulating, and inspiring others with his or her vision of the future" (1990: 112). Silins and Mulford (2002) found positive and significant effects of a shared

and monitored mission. Harris and Chapman's (2002) small-scale study of effective leadership in challenging schools in England found that the alignment of staffs' and heads' values and vision was a key to success. Locke (2002) argues that formulating a vision for the organization is one of eight core tasks for senior leaders and a key mechanism for achieving integration or alignment of activities within the organization.

Fostering the acceptance of group goals

While visions can be inspiring, action typically requires some agreement on the more immediate goals to be accomplished in order to move toward fulfilling the vision. Building on such theory, this set of practices aims not only at identifying important goals for the organization, but also doing so in such a way that individual members come to include the organization's goals among their own. Unless this happens, the organization's goals have no motivational effect. This set of practices includes leader relationship behaviors "aimed at promoting cooperation among [teachers] and getting them to work together toward a common goal" (Podsakoff *et al.*, 1990: 112).

In school settings, improvement-planning processes are among the more explicit contexts in which these behaviors are manifest. One of the 11 effective managerial behaviors included in Yukl's multiple linkage model, encompasses a portion of these practices. Planning and organizing include: "Determining long-range objectives and strategies. . . , identifying necessary steps to carry out a project or activity. . ." (1989: 130). This apparently rational planning process cannot be affected without attention to fostering acceptance of group goals.

High-performance expectations

This set of leadership practices is included as part of direction setting because it is closely aligned with goals. While high-performance expectations do not define the substance of organizational goals, they demonstrate the leader's values and, as Podsakoff explains, "the leader's expectations of excellence, quality, and/or high performance" (Podsakoff *et al.*, 1990: 112) in the achievement of those goals. Demonstrating such expectations is a central behavior in virtually all conceptions of transformational leadership.

Developing People

The three sets of practices in this category make a significant contribution to motivation. Their primary aim is capacity building; however, capacity building refers not only to the knowledge and skill staff need to accomplish organizational goals, but also to commitment and resilience, the dispositions needed to persist in applying

that knowledge and skill (Harris and Chapman, 2002). Individual teacher efficacy is arguably critical to these dispositions and it is a third source of motivation in Bandura's (1986) model. People are motivated by what they are good at. In addition, mastery experiences, according to Bandura, are the most powerful sources of efficacy. Therefore, building capacity which leads to a sense of mastery is highly motivational, as well.

Providing individualized support/consideration

Bass and Avolio include, as part of this dimension, "knowing your followers' needs and raising them to more mature levels...[sometimes through] the use of delegation to provide opportunities for each follower to self-actualize and to attain higher standards of moral development" (1994: 64). This set of behaviors, claims Podsakoff *et al.* (1990), should communicate the leader's respect for his or her colleagues and concerns about their personal feelings and needs (emotional understanding and support). This is a set of practices common to both the two-dimensional models of leadership (Ohio state, contingency theory, and situational leadership theory), which include task orientation and consideration for people. Encompassed by this set of practices are the supporting, and recognizing and rewarding managerial behaviors associated with Yukl's (1989) multiple linkage model, as well as Hallinger's (2003) model of instructional leadership. This set of leadership behaviors has attracted more leadership research outside of schools since the 1960s than any other.

Intellectual stimulation

Behaviors included in this dimension include encouraging colleagues to take intellectual risks, re-examine assumptions, look at their work from different perspectives, rethink how it can be performed (Avolio, 1994; Podsakoff *et al.*, 1990), and otherwise "induc[e]...employees to appreciate, dissect, ponder and discover what they would not otherwise discern..." (Lowe *et al.*, 1996: 415–416). Waters, Marzano, and McNulty (Marzano *et al.*, 2005; Waters *et al.*, 2003) include challenging the *status quo* among the practices contributing to leader effects on students.

The leader's role in professional development has been found to be especially important for leaders of schools in challenging circumstances (Gray, 2000; Harris and Chapman, 2002). However, there are many informal, as well as formal, ways in which such development occurs, reflecting current understandings of learning as constructed, social, and situated. All models of transformational leadership include this set of practices. A considerable amount of the educational literature assumes such practices on the part of school leaders, most notably the literature on instructional leadership which places school leaders at the center of instructional improvement efforts in their schools (e.g., Day *et al.*, 2000; Stein and Spillane, 2005).

Providing an appropriate model

This set of practices entails leading by example. These are practices also associated with models of authentic leadership (Avolio and Gardner, 2005) and include demonstrating transparent decision making, confidence, optimism, hope, resilience, and consistency between words and deeds. Locke (2002) claims that core values are established by modeling core values in one's own practices. Both Hallinger (2003) and Waters *et al.* (2003) note the contribution to leader effects of maintaining high visibility in the school, a visibility associated with high-quality interactions with both staff and students.

Also encompassed by this dimension is Bass' idealized influence, a partial replacement for his original charisma dimension. Avolio (1994) claims that leaders exercise idealized influence when they serve as role models with the appropriate behaviors and attitudes that are required to build trust and respect in followers. Such modeling on the part of leaders "sets an example for employees to follow that is consistent with the values the leader espouses" (Podsakoff *et al.*, 1990: 112).

Redesigning the Organization

There is little to be gained by increasing peoples' motivation and capacity if working conditions will not allow their effective application. Bandura (1986) theorizes that beliefs about the setting in which one is working is a fourth source of motivation. People are motivated when they believe the circumstances in which they find themselves are conducive to accomplishing the goals they hold to be personally important. The three more specific sets of practices included in this category are about establishing the conditions of work which will allow staff to make the most of their motivations and capacities.

Building collaborative cultures

A large body of evidence has accumulated since Little's (1982) early research which unambiguously supports the importance of collaborative cultures in schools as being central to school improvement, the development of professional learning communities, and the improvement of student learning (e.g., Louis and Kruse, 1998; Rosenholtz, 1989). Additional evidence clearly indicates that leaders are able to build more collaborative cultures and suggests practices that accomplish this goal (e.g., Waters *et al.*, 2003). For leaders of schools in challenging circumstances, creating more positive collaborative and achievement-oriented cultures is a key task (West *et al.*, 2005).

Connolly and James (2006) claim that the success of collaborative activity is determined by the capacities and motivations of collaborators together with opportunities for them to collaborate. Success also depends on prior conditions. For example, a history of working together successfully will sometimes build trust, thus

making further collaboration easier; whereas a history of unsuccessful attempts to collaborate will reduce trust. Trust is increasingly recognized as a key element in encouraging collaboration. People are more likely to trust those with whom they have established good relationships (Bryk and Schneider, 2002; Louis and Kruse, 1995). Participative leadership theory and leader–member exchange theory are concerned with the nature and quality of collaboration in organizations and how to manage it productively.

Leaders contribute to productive collaborative activity in their schools by being skilled conveners of that work. They nurture mutual respect and trust among those involved in collaborating, by being trustworthy themselves, ensure the shared determination of group processes and outcomes, help develop clarity about goals and roles for collaboration, encourage a willingness to compromise among collaborators, foster open and fluent communication among collaborators, and provide adequate and consistent resources in support of collaborative work (Connolly and James, 2006; Mattessich and Monsey, 1992).

Restructuring

This is a set of practices common to virtually all conceptions of management and leadership. Organizational culture and structure are two sides of the same coin. Developing and sustaining collaborative cultures depend on putting in place complementary structures, typically something requiring leadership initiative. Practices associated with such initiatives include creating common planning times for teachers and establishing team and group structures for problem solving (e.g., Hadfield, 2003). Hallinger and Heck (1998) identify this variable as a key mediator of leaders' effects on students. Restructuring also includes distributing leadership for selected tasks and increasing teacher involvement in decision making (Reeves, 2000).

Building productive relationships with families and communities

Shifting the attention of school staffs from an exclusively inside-the-school focus to one which embraces a meaningful role for parents and a close relationship with the larger community was identified during the 1990s as the biggest change in expectations for those in formal school leadership roles (e.g., Goldring and Rallis, 1993). More recently, Muijs *et al.*, (2004) have identified this core practice as important for improving schools in challenging circumstances. Attention to this focus has been encouraged by evidence of the contribution of family educational cultures to student achievement in schools (e.g., Coleman, 1966; Finn, 1989), the increase in public accountability of schools to their communities through the widespread implementation of school-based management

(Murphy and Beck, 1995), and the growing need for schools to actively manage public perceptions of their legitimacy (e.g., Mintrop, 2004).

Connecting the school to its wider environment

School leaders spend significant amounts of time in contact with people outside of their schools seeking information and advice, staying in tune with policy changes, and anticipating new pressures and trends likely to have an influence on their schools and the like. Meetings, informal conversations, phone calls, e-mail exchanges, and Internet searches are examples of opportunities for accomplishing these purposes. The extensive number of network learning projects facilitated by the National College of School Leadership in England provides especially powerful opportunities for connecting one's school to its wider educational environment (Jackson, 2002) as do those in other countries. Referring to it as networking, Yukl includes it in his multiple linkage model of leadership as one of the 11 critical managerial practices. He describes this practice as "Socializing informally, developing contacts with people who are a source of information and support, and maintaining contacts through periodic interaction, including visits, telephone calls, correspondence, and attendance at meetings and social events" (1994: 69).

Bringing in external support may also be a productive response to schools engaged in significant school-improvement projects (Reynolds *et al.*, 2001).

Managing the Instructional Program

Both Burns' (1978) and Bass' (1985) conceptions of transformational leadership include several different sets of transactional leadership behaviors built on exchange theory. These behaviors have proven to be among the most problematic features of transformational leadership theory as it has been subject to empirical tests. One element of Bass' formulation (contingent reward) has almost always behaved as a transformational practice and most of the remainder makes little or no contribution to important outcomes of leadership. For this reason, the author and his colleagues have replaced transactional leadership with managerial practices in their school-specific model of transformational leadership. When these managerial practices, four in total, have been included in research on school leadership effects, they have proven to be consequential (e.g., Leithwood and Jantzi, 1999) in creating stability and strengthening the organization's infrastructure.

Staffing the program

Although not touched on by Hallinger (2003) or Waters *et al.* (2003), this has proved to be a key function of leaders engaged in school improvement. Finding teachers with the interest and capacity to further the school's efforts is

the goal of this activity. Recruiting and retaining staff constitute a primary task of leading schools in challenging circumstances (Hargreaves and Fink, 2006).

Providing instructional support

This set of practices, included in both Hallinger's (2003) and Waters' *et al.* (2003) research on effective leadership, includes supervising and evaluating instruction, coordinating the curriculum, and providing resources in support of curriculum, instruction, and assessment activity. West *et al.* (2005) indicate that, for leaders of schools in challenging contexts, focusing on teaching and learning is essential. This includes controlling behavior, boosting self-esteem, and talking and listening to pupils. It also may include urging pupils and teachers to place a strong emphasis on pupil achievement. Such an academic climate makes significant contributions to achievement (De Maeyer *et al.*, 2006).

Monitoring school activity

Waters *et al.* analyze associated leadership effects on students with leader monitoring and evaluating functions, especially those focused on student progress. The purposeful use of data is reported by West *et al.* (2005) to be a central explanation for effective leadership in failing schools (see also Reynolds *et al.*, forthcoming). Hallinger's (2003) model includes a set of practices labeled monitoring student progress. Monitoring operations and environment is one of Yukl's (1989) 11 effective managerial practices. Furthermore, Gray (2000) reports that tracking student progress is a key task for leaders of schools in challenging circumstances.

Buffering staff from distractions to their work

A long line of research has reported the value to organizational effectiveness, of leaders who prevent staff from being pulled in directions incompatible with agreed-on goals (Copland, 2003; DiPaola and Tschannen-Moran, 2005; Dwyer, 1985). This buffering function acknowledges the open nature of schools and the constant bombardment of staff with expectations from parents, the media, special interest groups, and the government. Internal buffering is also helpful, especially buffering teachers from excessive pupil-disciplinary activity.

The four sets of leadership practices in this category provide the coordination for initiatives stimulated by the other core leadership practices. They help provide the stability which is so necessary for improvement to occur.

Conclusion

In his summary of transformational approaches to leadership, Northouse (2007) identifies five common criticisms of transformational approaches to leadership. First, some

have argued that this approach to leadership lacks conceptual clarity, meaning that its dimensions and associated practices overlap with other views of leadership. The school-specific model described in this article is especially guilty of this quality. However, this is a criticism only if the goal is to distinguish one model of leadership from another. If, as in this case, the goal is to develop a model of leadership that incorporates practices which reflect the best evidence about what works, then lack of distinguishing features is inevitable and certainly not a weakness.

A second criticism is that transformational approaches treat leadership as a personality trait rather than a set of practices that can be learned. This is a legitimate complaint of models that include a charismatic dimension. At least some forms of charisma depend on the possession of attractive personality traits. However, many models of transformational leadership no longer include charisma. The school-specific model described in this article includes only behaviors or practices that almost anyone with an interest in leadership could learn or improve on.

Some have also criticized transformational leadership as being elitist. This criticism seems to have arisen because much of the research literature in nonschool contexts has focused on the leadership of those at the top of the organizational hierarchy. The model outlined in this article includes practices that can be widely distributed throughout the organization; recent evidence suggests that they often are (Leithwood *et al.*, 2003; Leithwood *et al.*, 2007). This evidence is also a response to the criticism that transformational approaches reflect heroic images of leadership.

A final criticism is that this form of leadership has the potential for abuse. Early descriptions of transformational leadership sometimes cited, as examples, highly influential leaders who, nevertheless, pursued morally and socially unacceptable goals – Hitler, for example. Considerable effort has been made since those early writings to craft an account of transformational leadership in pursuit of morally and socially desirable ends, an authentic as opposed to inauthentic or pseudo-transformational form of leadership (Bass and Steidlmeier, 1999; Price, 2003). Generally, charisma-free models are less susceptible to this criticism.

Both the demonstrably positive effects of transformational leadership and the insubstantial nature of its apparent limitations argue for continuing to further develop and assess this leadership approach.

Bibliography

- Avolio, B. J. (1994). The alliance of total quality and the full range of leadership. In Bass, B. M. and Avolio, B. J. (eds.) *Improving Organizational Effectiveness through Transformational Leadership*, pp 121–145. Thousand Oaks, CA: Sage.

- Avolio, B. J. and Gardner, W. (2005). Authentic leadership development: Getting to the root of positive forms of leadership. *The Leadership Quarterly* **16**(3), 315–338.
- Bandura, A. (1986). *Social Foundations of Thought and Action*. Englewood Cliffs, NJ: Prentice Hall.
- Bass, B. M. (1985). *Leadership and Performance beyond Expectations*. New York: The Free Press.
- Bass, B. M. and Avolio, B. J. (1994). *Improving Organizational Effectiveness through Transformational Leadership*. Thousand Oaks, CA: Sage.
- Bryk, A. S. and Schneider, B. (2002). *Trust in Schools: A Core Resource for Improvement*. New York: Russell Sage Foundation.
- Burns, J. M. (1978). *Leadership*. New York: Harper and Row.
- Coleman, J. S. (1966). *Equality of Educational Opportunity*. Washington, DC: Government Printing Office.
- Connolly, M. and James, C. (2006). Collaboration for school improvement: A resource dependency and institutional framework of analysis. *Educational Management Administration and Leadership* **34**(1), 69–87.
- Copland, M. A. (2003). The Bay Area school reform collaborative: Building the capacity to lead. In Murphy, J. and Datnow, A. (eds.) *Leadership Lessons from Comprehensive School Reforms*, pp 159–183. Thousand Oaks, CA: Corwin Press.
- Day, C., Harris, A., Hadfield, M., Tolley, H., and Beresford, J. (eds.) (2000). *Leading Schools in Times of Change*. Buckingham: Open University Press.
- De Maeyer, S., Rymenans, R., Van Petegem, P., van der Bergh, H., and Rijlaarsdam, G. (2006). Educational leadership and pupil achievement: The choice of a valid conceptual model to test effects in school effectiveness research. Unpublished paper: University of Antwerp.
- DiPaola, M. F. and Tschannen-Moran, M. (2005). Bridging or buffering? The impact of schools' adaptive strategies on student achievement. *Journal of Educational Administration* **43**(1), 60–71.
- Dwyer, D. C. (1985). Contextual antecedents of instructional leadership. *The Urban Review* **17**(3), 166–188.
- Dumdum, U. R., Lowe, K. B., and Avolio, B. J. (2002). A meta-analysis of transformational and transactional leadership correlates of effectiveness and satisfaction: An update and extension. In Avolio, B. J. and Yammarino, F. J. (eds.) *Transformational and Charismatic Leadership: The Road Ahead*, pp 35–66. Oxford: Elsevier Science.
- Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research* **59**(2), 117–143.
- Fullan, M. (2003). *The Moral Imperative of School Leadership*. Thousand Oaks, CA: Corwin Press.
- Goldring, E. and Rallis, S. (1993). *Principals of Dynamic Schools*. Newbury Park, CA: Corwin.
- Gray, J. (2000). *Causation Concern but Improving: A Review of Schools' Experience*. London: Department for Education and Skills.
- Hadfield, M. (2003). Building capacity versus growing schools. In Harris, A., Day, C., Hopkins, D., et al. (eds.) *Effective Leadership for School Improvement*, pp 107–120. New York: RoutledgeFalmer.
- Hallinger, P. (2003). Leading educational change: Reflections on the practice of instructional and transformational leadership. *Cambridge Journal of Education* **33**(3), 329–351.
- Hallinger, P. and Heck, R. (1998). Exploring the principal's contribution to school effectiveness: 1980–1995. *School Effectiveness and School Improvement* **9**(2), 157–191.
- Hargreaves, A. and Fink, D. (2006). *Sustainable Leadership*. San Francisco, CA: Wiley.
- Harris, A. and Chapman, C. (2002). *Effective Leadership in Schools Facing Challenging Circumstances*. Nottingham: National College for School Leadership (NCSL).
- Jackson, D. (2002). *The Creation of Knowledge Networks: Collaborative Enquiry for School and System Improvement*. Paper presented at the CER/IOECD/DfES/QCA ESRC Forum Knowledge Management in Education and Learning, Oxford, UK.
- Leithwood, K. and Jantzi, D. (1999). The relative effects of principal and teacher sources of leadership on student engagement with school. *Educational Administration Quarterly* **35**(supplement), 679–706.
- Leithwood, K. and Jantzi, D. (2005). A review of transformational school leadership research: 1996–2005. *Leadership and Policy in Schools* **4**(3), 177–199.
- Leithwood, K., Mascal, B., Strauss, T., et al. (2007). Distributing leadership to make schools smarter: Taking the ego out of the system. *Leadership and Policy in Schools* **6**(1), 37–67.
- Little, J. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. *American Educational Research Journal* **19**, 325–340.
- Locke, E. A. (2002). The leaders as integrator: The case of Jack Welch at General Electric. In Neider, L. L. and Schriesheim, C. (eds.) *Leadership*, pp 1–22. Greenwich, CT: Information Age Publishing.
- Louis, K. and Kruse, S. (1995). *Professionalism and Community: Perspectives on Reforming Urban Schools*. Newbury Park, CA: Corwin.
- Louis, K. S. and Kruse, S. D. (1998). Creating community in reform: Images of organizational learning in inner-city schools. In Leithwood, K. and Louis, K. S. (eds.) *Organizational Learning in Schools*, pp 17–45. Lisse, The Netherlands: Swets and Zeitlinger.
- Lowe, K. B., Kroeck, K. G., and Sivasubramanian, N. (1996). Effectiveness correlates of transformational and transactional leadership: A meta-analytical review of the MLQ literature. *Leadership Quarterly* **7**(3), 385–425.
- Marzano, R. J., Waters, T., and McNulty, B. A. (2005). *School Leadership that Works: From Research to Results*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Mattessich, P. W. and Monsey, B. R. (1992). *Collaboration: What Makes it Work?* St. Paul, MN: Amherst.
- Mintrop, H. (2004). *Schools on Probation: How Accountability Works (and Doesn't Work)*. New York: Teachers College Press.
- Muijs, D., Harris, A., Chapman, C., Stoll, L., and Russ, J. (2004). Improving schools in socioeconomically disadvantaged areas – A review of research evidence. *School Effectiveness and School Improvement* **15**(2), 149–175.
- Murphy, J. and Beck, L. (1995). *School-based Management as School Reform*. Thousand Oaks, CA: Corwin.
- Northouse, P. (2007). *Leadership: Theory and Practice*, 4th edn. Thousand Oaks, CA: Sage.
- Podsakoff, P., MacKenzie, S., Moorman, R., and Fetter, R. (1990). Transformational leader behaviors and their effects on followers' trust in leader satisfaction and organizational citizenship behaviors. *Leadership Quarterly* **1**(2), 107–142.
- Reeves, J. (2000). Tracking the links between pupil attainment and development planning. *School Leadership and Management* **20**(3), 315–332.
- Reynolds, A. J., Stringfield, S., and Muijs, D. (forthcoming). Results for the high reliability schools project. Unpublished manuscript.
- Reynolds, D., Hopkins, D., Potter, D., and Chapman, C. (2001). *School Improvement for Schools Facing Challenging Circumstances: A Review of Research and Practice*. London: Department for Education and Skills.
- Rosenholtz, S. J. (1989). *Teachers' Workplace: The Social Organization of Schools*. New York: Longman.
- Silins, H. and Mulford, B. (2004). Schools as learning organizations – Effects on teacher leadership and student outcomes. *School Effectiveness and School Improvement* **15**(3–4), 443–466.
- Silins, H. and Mulford, W. (2002). Leadership and school results. In Leithwood, K. and Hallinger, P. (eds.) *Second International Handbook of Educational Leadership and Administration*, pp 561–612. Dordrecht: Kluwer.
- Stein, M. and Spillane, J. (2005). What can researchers on educational leadership learn from research on teaching: Building a bridge. In Firestone, W. and Riehl, C. (eds.) *A New Agenda for Research in Educational Leadership*, pp 28–45. New York: Teachers College Press.
- Waters, T., Marzano, R. J., and McNulty, B. (2003). *Balanced Leadership: What 30 Years of Research Tells Us about the Effect of Leadership on Pupil Achievement. A Working Paper*. Aurora, CO: Mid-continent Research for Education and Learning.
- West, M., Ainscow, M., and Stanford, J. (2005). Sustaining improvement in schools in challenging circumstances: A study of successful practice. *School Leadership and Management* **25**(1), 77–93.
- Yukl, G. (1989). *Leadership in Organizations*, 2nd edn. Englewood Cliffs, NJ: Prentice Hall.
- Yukl, G. (1994). *Leadership in Organizations*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.

Further Reading

Leithwood, K., Day, C., Sammons, P., Harris, A., and Hopkins, D. (2006). *Successful School Leadership: What It is and How It Influences Pupil Learning*. London: Department for

Education and Skills. <http://www.dfes.gov.uk/research/data/uploadfiles/RR800.pdf>(accessed May 2009).

Leithwood, K., Jantzi, D., and Dart, B. (1990). Transformational leadership: How principals can help reform school cultures. *School Effectiveness and School Improvement* **1**(4), 249–280.

LEARNING AND COGNITION – ISSUES AND CONCEPTS

Contents

Attention in Cognition and Early Learning

Concept Learning

Cooperative Learning

Intelligence

Learning from Multiple Information Sources

Memory

Metacognition

Personal Epistemology in Education: Concepts, Issues, and Implications

Piaget: Recent Work

The Adult Development of Cognition and Learning

Vygotsky and Recent Developments

Attention in Cognition and Early Learning

M Heimann, Linköping University, Linköping, Sweden; The Swedish Institute for Disability Research, Linköping, Sweden; The Norwegian Network for Infant Mental Health, Oslo, Norway

T Tjus and K Strid, University of Gothenburg, Gothenburg, Sweden

© 2010 Elsevier Ltd. All rights reserved.

“From a psychological view, attention includes changes from sleepiness to high alertness, from focused OR to a single object to unfocused awareness of the general scene, from responsiveness to external event to responses driven by the achievement of a particular goal.”

(Posner and Rothbart, 2007a: 16)

A child's ability to direct his or her attention drives awareness; only objects and events attended to will enter the child's mind. In this sense, attention processes are essential for learning and cognitive development. Attention influences and is, in turn, influenced by various brain systems, an interaction that creates priorities affecting both perception and action. Thus, alertness and the way attention regulates, or is regulated by brain functions become highly relevant for all domains of learning. Without attention it becomes difficult, if not impossible, to take in new and important information. We, adults and children, are active in selecting what to attend to and what to ignore – a process that becomes more and more voluntary during development.

The view of an infant as passive, incapable of communicating and remembering, and unaware about self and others has changed radically during the last 50 years. Typically, developing infants have a sensory system, which makes them capable of perceiving and interacting with the environment promoting communication, emotional

exchange, and signals for caretaking. In addition to this sensory system, they use their own body as a vehicle, not only for exploration of themselves, that is, building sensory motor schemas, but also for assembling information about other individuals. To learn from and understand other people and socially relate to them is crucial for a healthy development. Therefore infants' face processing is seen as an early indicator of this social attending. Newborns prefer face-like patterns suggesting an innate mechanism favored in brain processing. Furthermore, by 4 months infants look longer if there is a mutual eye gaze with the adult compared to averted eye gaze. Also, auditory perception in 6-month-old infants is linked to sensitive detection of human voices, such as discrimination of phoneme segments, both in native and foreign languages. However, at the age of 12 months this ability declines showing that early on our brain is prepared for a general language acquisition, but that exposure to one language with its specific prosodic and phonetic patterns reduces this capacity.

Attention in Infancy

All sensory systems are functional at or before birth and of these, the visual system is probably both the best-understood and least-mature system. It is actually a difficult

task for the newborn infant to respond to visual stimuli. When a newborn infant acts on complex visual information, as evident when the neonate imitates facial movements, the infant responds in spite of the fact of the immaturity of the visual system: Jerky eye movements make it hard to control vision and it is difficult for the immature system to capture fast-moving objects. In short, the infant's control of fixation is mostly not under voluntary control at birth and, during the first months of life, infants cannot resist being drawn to certain patterns. Visually, the newborn infant is attracted to high spatial frequencies (e.g., checkerboard pattern and edges), to slowly moving objects (moving stimuli are more attractive than static ones), and to face-like patterns. The attraction to edges provides necessary input for the visual cortex to develop (input helps to organize cell columns in visual cortex) while the attraction to face-like patterns guides the infant toward the social world.

All nonvisual sensory systems are also functional at birth but their role in early attentional processes is less well studied, although auditory attention probably is as important as vision for early development. Hearing is essential both for language learning and for making the social world interesting. The newborn infant recognizes human voices at birth (in fact, a fetus can learn to identify the mother's voice several weeks post-partum) and the melody of the mother tongue is identified within the first weeks. The human voice, especially the female voice, attracts the infants' attention both to the social world and to language.

Selective Attention

Selective attention develops rapidly over the first months of life. The newborn infant has less oculomotor control, less control over attentional shifts, and is attracted by salient details in perceptual displays. A neonate is not able to understand a partly occluded object as a unity and reacts only to direct visible information. It becomes a very difficult task if the neonate must identify a center-occluded object – the infant sees only the top and the bottom part of the object – in order to solve a task. This does not imply that a newborn child is not able to take in the relevant information, only that during the first months of life, perception and attention is more driven by environmental input (exogenous processes) coupled with biologically driven subcortical processes. Slowly, vision becomes more and more cortically controlled, acuity develops, the visual field increases, the eyes become better coordinated and inspection times decrease. A shift is usually seen around 2–3 months making the infant more visually competent when inspecting new objects, and, maybe most important, the child now becomes able to partake in prolonged face-to-face interactions with the caregiver.

The problem with perceptual completion, to see an object as a whole in spite of the fact that it is partly

occluded, also changes during the first months of life. Some capacity is observable already at 2 months but it is not until 4–5 months that the ability to solve perceptual-completion problems can be expected to be robust. Interestingly, children at 3 months seem to be in a transition phase. A recent study (Amos and Johnson, 2006) identified two groups of children, perceivers and non-perceivers: the children who were able to perceive unity (i.e., they solved the perception-completion task) used a more efficient strategy when solving a visual search task. These differences may stem from the possibility that the two groups reflect different stages in early brain development. Selective attention makes it possible for the child to become an active participant in his/her development and early differences in this ability might have an impact on later cognitive development.

The capacity to attend selectively is also related to disengagement, an ability that develops rapidly over the first months of life. As depicted in **Figure 1**, infants younger than 2 months tend to focus on one stimulus at a time. If a second stimulus is added to the visual display while the first stimulus is still visible, the infant is more or less unable to make a gaze shift. However, by 3–4 months, this is no longer a problem.

Selective looking and habituation have dominated research on infant attention. Generally, the length of looking declines during infancy, especially from 3 to 4 months and onward. Younger infants need longer inspection time (familiarization) than older infants; studies indicate a negative correlation between visual recognition and the inspection time needed. Selective attention has also been used to study auditory attention (e.g., the ability to discriminate between specific sound signals or between mother's voice and a stranger's voice).

Phases of Attention

Behavioral Phases

Attention in infancy is usually described as made up of three or four distinct phases representing different underlying processes (see **Figure 2**): The first phase, AR (AR), reflects the physiological readiness of the organism to react to any stimuli. Usually AR varies from deep sleep to active wakefulness through several intermediate phases. A characteristic of the young infant is that state changes occur often and rapidly, which affects the manner in which early attention is modulated. Without adequate AR, no attention can take place. During the second phase, usually called orienting (OR) or selective attention, the infant's attention is directed toward specific stimuli, that is, an interesting event is identified and the system becomes prepared for further inspection. Sustained attention (SA), phase three, reflects a phase of active information processing or encoding. SA is often described as a voluntary

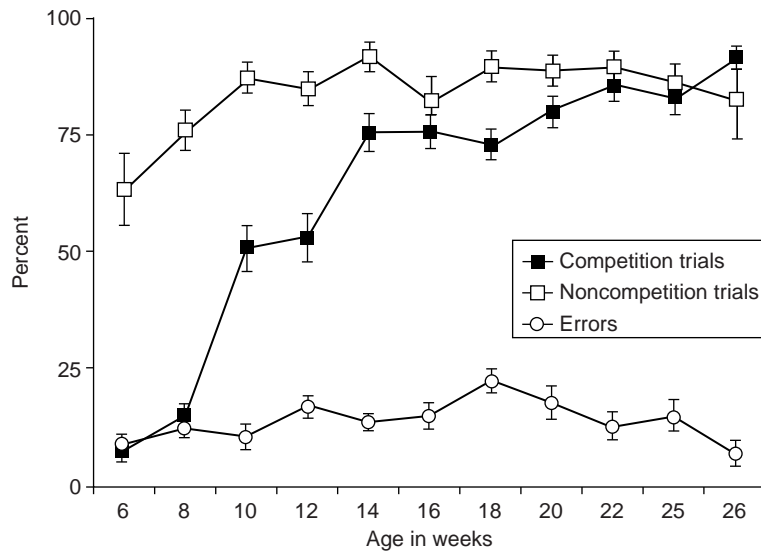


Figure 1

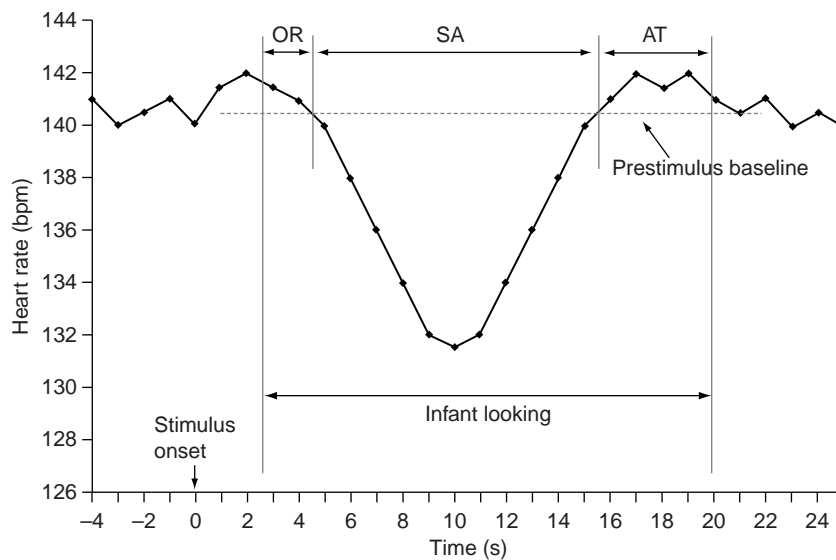


Figure 2

process and it has been linked to anterior brain systems. The final phase of attention, attention termination (AT), on the other hand, is linked to disengagement processes reflecting continuous looking after cessation of any active information processing. AT has been proposed to reflect the posterior brain system identified by Posner and colleagues and the disengagement processes have been observed to be a stronger prediction of recognition than sustained attention, at least for static stimuli.

Attention: Brain Networks

A recent attempt by Posner and Rothbart (2007b) to integrate behavioral studies of attention with current

findings within cognitive neuroscience identifies three neural networks that underlie central aspects of attention: orient, alert, and executive attention.

Alerting signifies that the child is in a state of readiness for reacting to new incoming stimuli. From a biological point of view, alerting has been linked to subcortical processes (thalamus) as well as to right parietal and right frontal areas. Norepinephrine is the main neurochemical modulator. . .

Orienting means that attention aligns with incoming sensory signals and that some information is selected. It is activated by posterior parts of the brain, mostly the parietal lobe (e.g., the superior parietal lobe) but also some frontal (frontal eye fields) and subcortical areas

(superior colliculus). The main neurochemical component is acetylcholine.

Executive attention indicates the ability to monitor responses, thoughts, and feelings, processes that not only involve mainly anterior brain areas (frontal cortex) but also some parts of the basal ganglia. It is hypothesized that executive attention plays an important role in the ability of the developing self to regulate positive and negative affect. Dopamine is the main neurochemical modulator.

The more posterior parts of the attentional system develop relatively early. Thus, the young infant is more able to alert and orient than to show executive attention, although it takes time for these systems to become fully mature as well. The first signs of the executive system can be noted toward the end of the first year, but it will take 8–10 years for it to reach adult-like levels.

All three networks – especially the executive network – are essential for the infant to develop an ability to display effortful control, that is, to self-regulate one's own behavior.

Aspects of Attention

Novelty

Novel objects and novel locations affect attention from birth although the parallel process of familiarity also exerts a strong influence on attention during the first months of life. Behaviorally, novelty preference can be reliably observed from approximately 4–6 months of age. Identification of new information leads to faster and more reliable OR, activates the alerting system, recruits available relevant brain resources (working memory and sustained attention), and prepares the system to encode a new file (transfer information to long-term memory).

The identification of a new object is governed by two processes. The first process to emerge is the ability to identify a new location. This ability is related to the control of eye movements and is, in some form, present at birth. It is observable through inhibition of return (IOR), which means that the visual system resists going back to a previous location; new locations are more attractive. The second process, the identification of a new object, is a slower-developing visual skill. It is usually established by 4–6 months and is related to the development of object recognition. This ability is observable through visual recognition memory.

Visual Recognition Memory

From 3 to 4 months and onward, the infant's ability to process new information has proved to be robust. This can be shown in paired-comparison experiments or tests when the child's preferential looking pattern is observed. Typically, a child prefers new information (a novel stimulus) in

comparison with familiar information, which makes it possible for the researcher to calculate a novelty preference score. This score provides an index of the child's visual recognition memory, because the child has to remember the familiar target in order to show a preference for the novel one. Measures of novelty preference/visual recognition memory in early infancy have been found to be a significant predictor of later intelligence quotient (IQ) as well as of nonverbal communication and language (e.g., Colombo, 1993; Bornstein and Sigman, 1986). As an example, it has been reported that visual recognition memory measured in infancy predicts receptive language at 3 years even when controlling for general IQ. These results suggest that attentional and memory capacities tapped by visual recognition memory as measured by novelty preference are important for later developing communication skills.

Visual recognition memory seems to tap into the very basic functions of our nervous system but exactly how these early attentional and memory processes affect later cognitive processes are still unexplained. One clue to an explanation has been proposed by Colombo and collaborators who found that, at 4 months, attention termination explained more of the observed variance in a novelty recognition task than sustained attention. The research group interprets their findings as supporting the "hypothesis that individual differences in the disengagement underlie the relation between look duration and cognitive performance in early to mid infancy" (Colombo *et al.*, 2001: 1605).

Face Processing

Human infants show a preference for faces right from birth, a tendency that seems to be especially strong for moving face-like stimuli (Johnson, 2005). This has been interpreted by some as indicating a pre-wired system molded by evolution. Such a system is described as being controlled by subcortical parts of the visual system functional at birth making the face a highly salient stimulus for the newborn. Thus, the child is endowed with an in-built perceptual attention grabber for faces. Although the propensity for faces among newborns is not debated, the mechanism is. In a series of studies, Simion and co-workers have provided support for an alternative view. They argue that the infant is not born with a pre-wired schematic configuration for faces but with a "domain-general bias towards configurations with more elements in the upper than lower half (i.e., top-heavy patterns)" (e.g., Cassia *et al.*, 2004: 379). The conflict between these two contrasting views is not resolved; we still do not know exactly how to describe the mechanism responsible for making newborn babies especially attracted to the human face.

Furthermore, observations also suggest that infants are better at processing female than male faces, a finding that probably stems from the fact that infants have much more experience in processing female faces. However, conclusive studies addressing this interpretation are still lacking. Finally, over the course of the first year, infants begin to process naturally looking faces in the same fashion as adults do; that is, they analyze faces in a holistic gestalt-like fashion. While 4-month-old children focus more on internal facial features, 10-month-old children use a holistic strategy. Children at 6 months seem to be in a transition phase using both strategies.

Attention and Learning

The Directed-Attention Model

Memory span of objects is around one item at 6 months and two to three items at 1 year. This means that infant learning and infants' complex social responses are carried out with a memory and attention system that is very limited compared to older children and adults. A model for how this can be achieved has been proposed by Reid and Striano (2007) who outlines a five-stage directed-attention model of infant social cognition. The model describes five perceptual stages that infants typically will master within the first year of life. The model is tentative but provides an ambitious attempt to integrate known perceptual and cognitive abilities and skills with the amazing social competence seen in the human infant. In brief, the proposed stages or phases are:

1. *Detection of socially relevant organisms.* Infants are born with a nervous system that directs them toward the social world. They are sensitive to distinctions between animate and inanimate objects, they prefer moving stimuli to static, biological motion to nonbiological motion, they are sensitive to the human voice, to rhythm, and they imitate (mimic) facial gestures.
2. *Identification of socially relevant organism.* The process to differentiate individual persons starts immediately at birth. The newborn infant rapidly learns to identify the mother's voice, smell, and face. More complex responses, such as imitation, are also used early on (from 6 weeks) to identify persons.
3. *Assessment of the locus of attention.* Once the socially relevant organism has been detected infants start to "attend towards characteristic that index the locus of attention of the observed organism" (Reid and Striano, 2007: 105). Eye and head movements provide important information to the infant in early social interactions. The infant's predisposition to enter into the social interactions with the caregiver is probably driven by intrinsic and biologically based motives to communicate with other humans.

4. *Detection of object-oriented attention.* Already at 4 months infants can use an adult's gaze to learn about objects and, by 8–9 months, objects become parts of highly motivating joint-attention encounters. Objects or an aspect of the environment that has been highlighted through joint attention with an adult will become more salient to the infant who will direct more of his or her attention to those objects/areas. The first sign of an emerging declarative memory is observed at 6 months through the child's ability to act on a memorized representation after a delay (deferred imitation).
5. *Inference of goals and/or prepare response.* Toward the end of the first year, infants begin to respond differently to accidental and intentional action. The child now starts to understand that people have goals that motivate their actions. The capacity of working memory increases, which makes it possible for the child to hold more than one piece of information online simultaneously.

Dyadic Attention

The interaction between a parent and the infant is characterized by rhythm, intimacy, and emotional exchange, the so-called proto-conversations. This interaction is encouraged by the mother through smiles and increased gaze in such a way that the interaction is prolonged and infants are active partners in this interaction, creating turn-taking sequences. Both mother and infant are sensitive to the contingency and quality of this interaction and have already created expectancies of specific patterns of communication from each other. This is evidenced by using still-face conditions where mother is either not responding to infant's communication or by using double-video-technique showing pre-recorded interaction creating unsynchronized turn-taking from both partners. Furthermore, infants very early learn to anticipate, not only contingent communicative patterns, but also contingent behavior from their mothers. When infants' distress is accompanied by parents soothing, the association between AR, parents' response, and subsequent relief is easily learned – soothing is anticipated by the infant already at 4 months of age.

This intrinsic communication with human beings could not be enhanced without the infant exploring the physical environment. With an increasing ability of motor control, the child will experiment with its own actions on objects such as mobiles and thereby develop an early understanding of agency according to physical cause and effect. With the growth of contingent perception, the infant is able to accompany this physical agency into detection of cause and effect/reciprocity with partners into further developed nonverbal turn-taking sequences – both vocal and with objects (like pushing a ball back and

forth). This means that infants are able to coordinate their attention with another person and understand the framework in which communication takes place between two persons.

Triadic Attention/Joint Attention

Nine-month-old infants begin to understand that actors are pursuing goals and they combine the awareness of outside objects, events, and persons in order to share and coordinate their attention or perception of goal activities, the so-called joint attention. By performing joint activities such as building a tower of bricks or rolling a ball back and forth, the infants understand the concept of sharing goals. This is accompanied by attention skills such as following the other persons' eye-gaze or pointing in order to direct others' attention to objects in the surrounding, or trying to modify a persons' behavior with gestures.

Attention skills that are initiated by the child (e.g., pointing) develop slightly later than behaviors that are responses to others (e.g., gaze-following). The capacity to direct another persons' gaze to objects by the infants' own interest develops between 9 and 12 months and infants can use the pointing gesture for different purposes, for both sharing attention and requesting.

It has been argued that declarative pointing, in contrast to imperative pointing, relies on the understanding of others as mental agents and is driven by a motivation to share attention and interest with other persons; it has therefore been suggested to be especially difficult. In typically developing infants, the motivation to share attention and interest is probably strong and declarative gestures are common, in contrast to children with social impairments like autism. Experiment with 12-month-olds has revealed that the social context is crucial for the amount of points the infant makes. Only when the adult was active in sharing the infants' attention to the event the infant pointed at, the infants' pointing increased. The interpretation of this result was that the infant did not only want to direct the adult's attention, but also wanted to share this attention.

It has been shown in several studies that the capacity for joint attention is an important precursor to later-developing language and cognitive skills. Together with early memory measures (e.g., visual recognition memory and deferred imitation) joint attention probably lays the ground for later-emerging social cognition including intentional understanding.

Understanding Intentions

Humans are special in their capacity to understand others' intentions. This is obvious when children pass false-belief tasks when they are about 4 years old. It can be argued that infants, by the time they begin to follow

and direct others attention, have acquired some understanding of others as intentional agents and that other people act on the basis of their own view of the world. From 6 months of age, infants follow another person's gaze to objects in the surrounding environment but they have been shown to pay more attention to where the head is turning, while older infants pay more attention to the eyes indicating understanding of the adult's intentions. When adults turn to an object with their eyes open or closed it was observed that infants by the age of 12 months looked at the target if the adult turned to it with open eyes but did not do so if the person turned to it with eyes closed.

Furthermore, studies have revealed that infants in their second year understand the intention behind an action and not only the action they actually have seen; when 18-month-old infants were observing an action that the adult failed to perform they did not imitate the failure, instead they performed the complete action. Another evidence for early intentional understanding is that 9-month-old infants show anger and distress toward an adult who is unwilling to give them a toy, but not to an adult who is unable to do it.

Compensatory Systems and Plasticity

Questions have been raised whether the sensory system can compensate for a deficit in one area, since it is striking how blind people can use auditory and tactile cues for orienting themselves and getting information. In one study, it was shown that spatial tuning of tactile attention is more accurate in early blind compared to sighted individuals when areas for Braille reading are stimulated, suggesting a compensatory scaffolding for individuals who have experienced visual deprivation from birth or early infancy. The plasticity of the brain making it possible to overcome deficits may be underlined by the fact that the sensory systems work together, especially, according to the dimensions of space, time, and intensity. This so-called redundant, amodal information is seen as a cornerstone of perceptual development, for example, when an adult takes a child's hands and uses them for clapping, auditory, visual, and tactile systems are involved, discerning both rhythm and rate to the child.

Conclusion

With increasing age, looking and attention come more and more under voluntary control. The child becomes more able to decide what to focus on; he or she will, by 4–5 years of age, be able to choose to attend to information even if it is boring. However, attention never becomes completely voluntary. Processes like novelty preference and habituation influence attention throughout life.

In the future, knowledge might be increased by detecting specific genes influencing attention, maybe even specific genes for each network or various attentional processes (see Posner *et al.*, 2007). This might help us gain a better understanding of how attention develops, how experience and biology co-act to create alertness and sustained attention, or the ability to initiate attention in another person – knowledge that will also be highly relevant for children with known disabilities such as ADHD or autism spectrum disorders.

See also: Cognition and Emotion; First Language Acquisition; Neuroscience Bases of Learning.

Bibliography

- Amos, D. and Johnson, S. P. (2006). Learning by selection: Visual search and object perception in young infants. *Developmental Psychology* **42**, 1236–1245.
- Bornstein, M. H. and Sigman, M. D. (1986). Continuity in mental development from infancy. *Child Development* **57**, 251–274.
- Cassia, V. M., Turati, C., and Simion, F. (2004). Can a nonspecific bias to more top-heavy pattern explain newborn's face preference? *Psychological Science* **15**, 379–383.
- Colombo, J. (1993). *Infant Cognition: Predicting Later Intellectual Functioning*. Newbury Park, CA: Sage.
- Colombo, J., Richman, W. A., Shaddy, J., Greenhoot, A. F., and Maikranz, J. (2001). Heart rate-defined phases of attention, look duration, and infant performance in the paired-comparison paradigm. *Child Development* **72**, 1605–1616.
- Johnson, M. H. (2005). *Developmental Cognitive Neuroscience*, 2nd edn. London: Blackwell.
- Posner, M. I. and Rothbart, M. K. (2007a). *Educating the Human Brain*. Washington, DC: American Psychological Association.
- Posner, M. I. and Rothbart, M. K. (2007b). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology* **58**, 1–23.
- Posner, M. I., Rothbart, M. K., and Sheese, B. E. (2007). Attention genes. *Developmental Science* **10**(1), 24–29.
- Reid, V. M. and Striano, T. (2007). The directed attention model of infant social cognition. *European Journal of Developmental Psychology* **4**, 100–110.
- Butcher, P. R., Kalverboer, A. F., and Geuze, R. H. (2000). Infants' shifts of gaze from central to peripheral stimulus: A longitudinal study of development between 6 and 26 weeks. *Infant Behavior and Development* **23**, 3–21.
- Colombo, J. (2002). Infant attention grows up: The emergence of a developmental cognitive neuroscience perspective. *Current Directions in Psychological Science* **11**, 196–200.
- Farroni, T., Csibra, G., Simion, F., and Johnson, M. H. (2002). Eye contact detection in humans from birth. *PNAS* **99**, 9602–9605.
- Forster, B., Eardley, A. F., and Eimer, E. (2007). Altered tactile spatial attention in the early blind. *Brain Research* **1131**, 149–154.
- Fridlund, A. J. (1997). The new ethology of human facial expression. In Russell, J. A. and Fernández-Dols, J. M. (eds.) *The Psychology of Facial Expression*, pp 103–127. Cambridge: Cambridge University Press.
- Gomes, H., Molholm, S., Christodoulou, C., Ritter, W., and Cowan, N. (2000). The development of auditory attention in children. *Frontiers of Bioscience* **5**, 108–120.
- Grossman, T. and Johnson, M. H. (2007). The development of the social brain in human infancy. *European Journal of Neuroscience* **25**, 909–919.
- Johnson, M. J. and Morton, J. (1991). *Biology and Cognitive Development: The Case of Face Recognition*. Oxford, UK: Blackwell.
- Johnson, P. J. (2004). Development of perceptual completion in infancy. *Psychological Science* **15**, 769–775.
- Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience* **5**, 831–843.
- Laucht, M., Becker, K., and Schmidt, M. H. (2006). Visual exploratory behaviour in infancy and novelty seeking in adolescence: Two developmentally specific phenotypes of DRD4. *Journal of Child Psychology and Psychiatry* **47**, 1143–1151.
- Moore, D. G., Goodwin, J. E., George, E., Axelsson, E. L., and Braddick, F. M. B. (2007). Infants perceive human point-light displays as solid forms. *Cognition* **104**, 377–396.
- Paterson, S. J., Heim, S., Friedman, J. T., Choudhury, N., and Benasich, A. A. (2006). Development of structure and function in the infant brain: Implications for cognition, language and social behaviour. *Neuroscience and Biobehavioral Reviews* **30**, 1087–1105.
- Ramsey-Rennels, J. L. and Langlois, J. H. (2006). Infant's differential processing of female and male faces. *Current Directions in Psychological Science* **15**, 59–62.
- Rose, S. A., Feldman, J. F., and Jankowski, J. J. (2004). Infant visual recognition memory. *Developmental Review* **24**, 74–100.
- Schwarzer, G., Zauner, N., and Jovanovic, B. (2007). Evidence of a shift from featural to configural face processing in infancy. *Developmental Science* **10**, 452–463.
- Smith, L., Fagan, J. F., and Ulvund, S. E. (2002). The relation of recognition memory in infancy and parental socioeconomic status to later intellectual competence. *Intelligence* **30**, 247–259.
- Striano, T., Chen, X., Cleveland, A., and Bradshaw, S. (2006). Joint attention social cues influence infant learning. *European Journal of Developmental Psychology* **3**, 289–299.
- Trevarthen, C. and Aitken, K. (2003). Regulation of brain development and age-related changes in infants' motives: The developmental function of regressive periods. In Heimann, M. (ed.) *Regression Periods in Human Infancy*, pp 107–184. Mahwah, NJ: Erlbaum.

Further Reading

- Bahrick, L. E., Lickliter, R., and Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science* **13**(3), 99–102.

Concept Learning

S J Unsworth, San Diego State University, San Diego, CA, USA
D L Medin, Northwestern University, Evanston, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

In what follows, we use concept to refer to a mental representation and category to refer to the set of entities or examples picked out by the concept. It is generally accepted that instances of a concept are organized into categories. Almost all theories about the structure of categories assume that, roughly speaking, similar things tend to belong to the same category and dissimilar things tend to be in different categories. For example, robins and sparrows both belong to the category bird and are more similar to each other than they are to squirrels or pumpkins. Similarity is a pretty vague term, but most commonly it is defined in terms of shared properties or attributes. Although alternative theories assume concepts are structured in terms of shared properties, theories differ greatly in their organizational principles.

Theories of Concept Representation and Learning

The Classical View

The classical view assumes that concepts have defining features that act like criteria or rules for determining category membership. For example, a triangle is a closed geometric form of three sides with the sum of the interior angles equaling 180°. Each of these properties is necessary for an entity to be a triangle, and together these properties are sufficient to define triangle.

A fair amount of research has examined people's knowledge about object categories such as bird, chair, and furniture and this evidence goes against the classical view. Not only do people fail to come up with defining features, but also they do not necessarily agree with each other (or even with themselves when asked at different times) on whether something is an example of a category. Philosophers and scientists also have worried about whether naturally occurring things such as plants and animals (so-called natural kinds) have defining features. The current consensus is that most natural concepts do not fit the classical view.

The Probabilistic View

The major alternative to the classical view is the probabilistic view which argues that concepts are organized around properties that are characteristic or typical of category members but crucially, they need not be true of all members. That is, the features are only probable. For

example, most people's concept of bird may include the properties of building nests, flying, and having hollow bones, even though not all birds have these properties (e.g., ostriches and penguins). The probabilistic view has major implications for how we think about categories. First, if categories are organized around characteristic properties, some members may have more of these properties than other members. In this sense, some members may be better examples or more typical of a concept than others. For example, it has been found that the more frequently a category member's properties appeared within a category, the higher was its rated typicality for that category. Robins were rated to be very typical birds and penguins rated as very atypical birds. A second implication is that category boundaries may be fuzzy. Nonmembers of a category may have almost as many characteristic properties of a category as do certain members. For example, whales have a lot of the characteristic properties of fish, and yet they are mammals. Third, learning about a category cannot be equated with determining what the defining features are because there may not be any.

Typicality: Central tendency versus ideality

Is typicality only based on central tendency? Although typicality effects are robust (and problematic for the classical view), other research shows that the underlying basis for typicality effects may vary with both the kind of category being studied and with the population being studied. While the internal structure of taxonomic categories is based primarily on the central tendency (or the average member) of a category, the internal structure of goal-derived categories, such as things to wear in the snow, is determined by some ideal (or the best possible member) associated with the category. The best example of snow clothing, a down jacket, was not the example that was most like other category members; instead, it was the example with the maximum value of the goal-related dimension of providing warmth.

One might think that ideals will only come into play when the category of interest lacks the natural similarity structure that characterizes common taxonomic categories, such as bird, fish, and tree. However, for tree experts (people who know a lot about trees, such as landscapers, parks workers, and taxonomists), the internal structure of the category tree is organized around the positive ideal

of height and the negative ideal of weediness. The best examples of tree are not trees of average height but trees of extraordinary height (and free of weedy characteristics like having weak limbs, growing where they are not wanted, and being susceptible to disease).

Indeed, research does suggest that people who have considerable knowledge in a domain tend to base typicality judgments on ideals and not on the number of typical features. For example, for Itzá Maya adults living in the rainforests of Guatemala, the best example of bird is the wild turkey which is culturally significant, prized for its meat, and strikingly beautiful. The fact that US tree experts based typicality on ideals suggests that it is not just that the Itzá have a different notion of what typicality means. It has also been found that Native American and European American fishermen's typicality judgments were based on ideals, although those ideals differed somewhat across groups.

Prototype versus exemplar theories

If categories are not represented in terms of definitions, what form do our mental representations take? One suggestion about how concepts are represented is known as the family resemblance principle. The general idea is that category members resemble each other in the way that family members do. A simple summary representation for such a family resemblance structure would be an example that possessed all the characteristic features of a category. The best example is referred to as the prototype.

In a prototype model of categorization, classifying a new example is done by comparing the new item to the prototype. If the candidate example is similar enough to the prototype for a category, it is classified as a member of that category. More detailed analyses, however, show problems with prototypes as mental representations. Prototype theory implies that the only information abstracted from categories is the central tendency. A prototype representation discards information concerning category size, the variability of the examples, and correlations among attributes, and people can use all three of these types of information.

An alternative approach, which is also consistent with the probabilistic view, assumes that much more information about specific examples is preserved. This approach appropriately falls under the general heading of exemplar theories. Exemplar models assume that people initially learn some examples of different concepts and then classify a new instance on the basis of how similar it is to the previously learned examples. The idea is that a new example reminds the person of similar old examples and that people assume that similar items will belong to the same category. For example, suppose someone is asked whether large birds are more or less likely to fly than small birds. He/she will probably answer "less likely," based on retrieving examples from memory and noting

that the only nonflying birds one can think of are large (e.g., penguin and ostrich).

Quite a few experiments have contrasted the predictions of exemplar and prototype models. In head-to-head competition, exemplar models have been considerably more successful than prototype models. Why should exemplar models fare better than prototype models? One of the main functions of classification is to allow one to make inferences and predictions on the basis of partial information. Relative to prototype models, exemplar models tend to be conservative about discarding information that facilitates predictions. For instance, sensitivity to correlations of properties within a category enables finer predictions: from noting that a bird is large, one can predict that it cannot sing. In short, exemplar models support predictions and inferences better than do prototype models.

More recent research has pointed to three major limitations of these simple forms of prototype and exemplar models:

1. they have narrowly focused on categorization and have paid little attention to how other conceptual functions, such as communication and inference, may affect concept representation and learning;
2. they view learning as a passive accumulation of statistical information rather than an active learning that may reflect particular learner goals; and
3. they pay little attention to how theoretical notions and causal reasoning organize learning.

With respect to the second point, we have just reviewed evidence from a number of populations, indicating that typicality is driven by ideals and that later learning builds on earlier learning. If category ideals tend to be learned first then they will have an important role in the development of categories, and modelers are beginning to shift to this more active view of learning. With respect to the role of theories, there is evidence that using (abstract) similarity relations may be likely to be a strategy of last resort, used only when more relevant information is unavailable. Let us examine the theory view in a bit more detail.

The Theory View

A number of researchers have argued that the organization of concepts is knowledge based (rather than similarity based) and driven by intuitive theories about the world. The idea that concepts might be knowledge based rather than similarity based suggests a natural way in which concepts may change – namely, through the addition of new knowledge and theoretical principles. There is also good evidence that these theories help determine which abstract and observable features learners pay attention to. We have a different set of categories for mental disorders now than we had 100 years ago, in part because our

knowledge base has become more refined. Often knowledge of diseases develops from information about patterns of symptoms to a specification of underlying causes. For example, the advanced stages of syphilis were treated as a mental disorder until the causes and consequences of this venereal disease were better understood. Recently, it has been shown that clinical psychologists organize their knowledge of mental disorders in terms of rich causal theories and that these theories (and not the atheoretical diagnostic manual they are supposed to use) guide their diagnostic classification and reasoning.

Domain Specificity

Several constraints have been hypothesized to mold concept formation in different domains, including the domains of biology, psychology, mathematics, and physics. The current consensus is that the potential for variation in conceptual knowledge across cultural communities is mediated by universal constraints on learning and the ways in which they interact with culture-specific experiences. Concepts are the building blocks of thought and one way to understand the flexibility of concept learning is to consider whether people in different cultures think differently. Usually, this question is tied up with the question of whether and how language influences thought and we will not give a separate treatment of this issue. Of course, if thought processes of two cultural groups were radically incommensurable, one would quickly realize that there were dramatic differences but feel at something of a loss to explain them. The fact that one part of learning a foreign language involves finding out what term or word is used in that language to refer to bird, or fish, or chair, or Tuesday, or mother suggests that comparable concepts and categories are in play. Nevertheless, culture affects learning and knowledge construction. Rather than provide a comprehensive catalog of the various principles constraining knowledge construction in each domain, we present a few detailed accounts of cultural research on concept formation, using for illustration cross-cultural conceptions of plants and animals (the domain of folk biology) and counting and calculation (the domain of folk mathematics).

Concept Learning in the Domain of Biology

The field of folk biology is blessed with many intriguing and important issues that lend themselves to an analysis in terms of culture and cognition. Biological concepts are believed to be processed and organized according to evolved cognitive structures that are functionally autonomous with respect to biological information, and for this reason are thought of as belonging to a separate domain of cognitive processing. Building on decades of work in

ethnobiology, research has shown that a few key principles guide the recognition and organization of biological information in similar ways across cultures, although important variation is produced by differences in expertise and other cultural factors.

First, there is marked cross-cultural agreement on the hierarchical classification of living things, such that plants and animals are grouped according to a ranked taxonomy with mutually exclusive groupings of entities at each level. For instance, across cultural groups, the highest level of taxonomic organization includes the most general categories, such as the folk kingdom rank (which includes groupings, such as plants and animals), and lower levels distinguish between increasingly greater degrees of specificity (e.g., life forms, such as tree or bird; generic species level, such as oak or blue jay). Furthermore, the generic species (in local settings the vast majority of genera are represented by a single species, so we use this term) level appears to be consistently privileged for inductive inference when generalizing properties across plants and animals (it is the most abstract level for which inductive confidence is strong and only minimal inductive advantage is gained at more subordinate levels). There is also cross-cultural agreement in the assumption that the appearance and behavior of every generic species is caused by an internal biological (and usually unspecified) essence that is inherited from the birth parents and is responsible for kindhood persistence in the face of physical and developmental transformation.

However, there is also considerable variability within these universal constraints in concept formation as a function of both experience with the natural world and cultural salience (two highly related factors). For instance, the basic level (the level at which they possess the greatest knowledge) for urban undergraduates is the life form (e.g., bird, fish, and tree), but for groups that have more direct experience with the natural environment and greater expertise, the basic level corresponds to the generic species level.

The remarkable cross-cultural agreement in the structure of folk biological organization is, at the same time, culturally variable. Correlations across groups of 0.70 appear quite strong but explain less than half the variance. Although some of these differences might be attributed to experience, other findings implicate cultural differences. For instance, when asked to sort biological kinds into categories, individuals from different communities vary not only in their taxonomic sorting but also in the degree to which they spontaneously sort along ecological dimensions. This difference is not as predictable on the basis of expertise alone. For example, Menominee Native American fisherman and European American fishermen, who both live in rural Wisconsin and have equivalent expertise about fish and fish habitats, differ in that Menominee fishermen are significantly more likely to sort in terms of ecological relationships.

Similar differences in ecological orientation have been found for children from these communities, such that Menominee children were more likely to reason about shared properties between living things using ecological relations, relative to rural European American children. In turn, rural European American children were more likely to employ ecological-based reasoning for shared properties than were urban children. In short, differences in ecological orientation reflect a confluence of experience-based and culturally based factors in folk biological thought.

Cultural differences in cognitive processing, concept representation, and behavior can be thought of as reflecting routines of practices or habits of the mind. Cultural groups establish practices over time, and the history of these practices may lead to regularities in the ways groups participate in the everyday activities within their communities. These practices may be associated, implicitly or explicitly, with different epistemologies that determine what sorts of things are presupposed, go without saying, and seem natural. For example, European Americans tend to conceive of nature as something external, to be cared for, and respected; in contrast, Native Americans are more likely to see themselves as part of nature. These sorts of presuppositions are likely to be embedded in curricula and school practices and represent a challenge to students from cultures and communities that do not share them.

Cultural practices are not immutable, static traits that are attached to participants (a view which can lead to overly deterministic views of cognition), but exist in tension with emergent goals, practices, and situationally specific affordances. Thus, one might design a biology curriculum for Native American students emphasizing ecological relationships, but then build on this base to suggest the value of other forms of organization (e.g., taxonomic). There is increasing evidence that taking advantage of the cultural practices that children bring to the classroom leads to better motivation, identification with learning, and academic performance.

Concept Learning in the Domain of Mathematics

Folk biological research has tended to compare different cultural groups and to identify robust similarities (and differences) in reasoning and representation. Studies of mathematical concepts have expanded on this strategy by using developmental comparisons and analyzing similarities between human and nonhuman species to identify universal or core principles. The domain of mathematics spans a wide variety of concepts, including numerosity, geometry, trigonometry, and so on.

We will limit our review to numerosity, counting, and calculation. A great deal of evidence suggests that for humans and other species there are evolved principles that assist in the representation of numerosity, and

that different principles can constrain representations in particular ways, depending on the set size of elements. Importantly, however, it has been proposed that the systems for large and small numerosity can interact for humans in ways not possible for nonhuman species. Number words and verbal counting may link together systems for small and large numerosities so that, through counting, distinctions can be made between large numerosities that differ in as little as one element.

The flexibility in concepts of numerosity afforded by natural language leads to questions about variability in representations of numerosity and counting as a function of language and other cultural inputs. Some innovative research has examined the different counting systems that have emerged in different cultural communities throughout the world. For instance, before contact with Western culture in 1940, the Oksapmin people in the West Sepik province of Papua New Guinea used a 27-body-part count system, beginning with the thumb on one hand and enumerating discrete points along the upper half of the body (including head and shoulders) and ending on the little finger on the other hand. Counting past 27 involves moving back along the same 27 points until the desired numerosity is reached. In addition, as individuals become more involved in the cash economy, this counting system becomes coopted for arithmetic calculations in addition to or in the place of enumeration and, in some cases, is even transformed to a base-10 system. Although cultural differences in counting systems are well established, not much work has examined the impact of these systems on the representation of numerosity.

Other research has examined the ways in which mathematical concepts, such as calculation processes and representations, are shaped by context-specific goals and culture-specific practices. For instance, grocery shoppers engage in mathematical calculations in response to specific shopping-related goals, and these calculations depend on the resources and environmental tools available to the shopper in the grocery store. Examples have been reported in which a shopper who, upon suspecting a price error for a block of cheese, sorted through a bin of cheese to find a block of similar weight and noted the difference in price that confirmed his suspicions. Had the bin of cheese not been available, the shopper would have had to mentally calculate the correct price based on listed-price-per-weight information.

An important issue is the relation between these sorts of out-of-school goal-related strategies and in-school mathematics learning. Community-specific goals can lead to a greater frequency and, therefore, greater proficiency for some calculations over others. For example, research has shown that 10–12-year-old children in Brazil with little or no education, who sold candy in urban streets, were highly likely to use ratio calculations during vending activities and were better at ratio comparisons than same-aged

children with formal education experience. Other work has revealed that African American middle school and high-school students vary in the extent to which they engage in mathematical calculations to evaluate basketball performance because of differences in the structure of the practice of basketball and level of commitment to basketball. High-school students were more likely to calculate formal statistics (such as average and percent) of their own and others' basketball performance, and these calculations were higher when ways of keeping and reporting basketball statistics were increasingly available to students. This work points out that the players' use and approach to mathematics during their everyday cultural practice may differ dramatically from the approach taken to school mathematics – the use of mathematics in a student's own cultural context is often more engaging. Related work with the children of sugarcane farmers found complementary tendencies to approach mathematical problem solving in different ways depending upon the value ascribed to the context or practice in play.

Conclusions

Concept learning is one of the most exciting and fundamental research areas within cognitive science because it concerns the very building blocks of thought. Early models which assumed that category learning consists of the accumulation of information about entities in the world have been superseded by approaches which stress that learning is in the service of goals, that it is guided by evolved, domain-specific constraints, and molded by cultural practices.

See also: Knowledge Domains and Domain Learning; Learning Science; Mathematics Learning; The Adult Development of Cognition and Learning.

Further Reading

- Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Erlbaum.
- Atran, S., Medin, D., and Ross, N. (2005). The cultural mind: Environmental decision making and cultural modeling within and across populations. *Psychological Review* **112**, 744–776.
- Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **11**, 629–654.
- Berlin, B. (1992). *Ethnobiological classification*. Princeton, NJ: Princeton University.
- Carey, S. (1985). *Conceptual Change in Childhood*. Cambridge, MA: Bradford Books.
- Gutierrez, K. D. and Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher* **32**, 19–25.
- Hauser, M. D. and Spelke, E. (2004). Evolutionary and developmental foundations of human knowledge. In Gazzaniga, M. (ed.) *The Cognitive Neurosciences, III*, pp 853–864. Cambridge: MIT Press.
- Keil, F. (1989). *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.
- Lave, J. (1988). *Cognition in Practice*. New York: Cambridge University Press.
- Medin, D. and Atran, S. (2004). The native mind: Biological categorization and reasoning in development and across cultures. *Psychological Review* **111**, 960–983.
- Murphy, G. L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.
- Nasir, N. (2000). "Points ain't everything": Emergent goals and average and percent understandings in the play of basketball among African-American students. *Anthropology and Education Quarterly* **31**, 283–305.
- NRC/National Research Council (2000). *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. Washington, DC: National Academies Press.
- NRC/National Research Council (2005). *How Students Learn: History, Math, and Science in the Classroom*. Washington, DC: National Academies Press.
- Rosch, E. and Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology* **7**, 573–605.
- Saxe, G. B. (1999). Cognition, development, and cultural practices. *New Directions for Child and Adolescent Development* **83**, 19–35.
- Saxe, G. (2006). The mathematics of child street vendors. *Child Development* **59**, 1415–1425.
- Smith, E. and Medin, D. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Cooperative Learning*

R E Slavin, Johns Hopkins University, Baltimore, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

Cooperative learning refers to a set of instructional strategies in which students work together in small groups to help each other learn academic content. Cooperative learning methods vary widely in their details: group sizes may be from two to several; group members may have individual roles or tasks, or they may all have the same task; and groups may be evaluated or rewarded based on group performance or the average of individual performances, or they may simply be asked to work together.

In one form or another, cooperative learning has been used and studied in every major subject, with students from preschool to college, and in all types of schools. It is used at some level by hundreds of thousands of teachers. One national survey in the 1990s found that 79% of elementary teachers and 62% of middle school teachers reported regular use of cooperative learning (Puma *et al.*, 1993). Antil *et al.* (1998) found that 93% of a sample of teachers reported using cooperative learning, with 81% reporting daily use.

There have been hundreds of studies of cooperative learning focusing on a wide variety of outcomes, including academic achievement in many subjects, second-language learning, attendance, behavior, intergroup relations, social cohesion, acceptance of classmates with handicaps, attitudes toward subjects, and more (see Slavin, 1995; Johnson and Johnson, 1998; Rohrbeck *et al.*, 2003).

This article focuses on research on achievement outcomes of cooperative learning in elementary and secondary schools, and on the evidence supporting various theories to account for effects of cooperative learning on achievement.

Theoretical Perspectives on Cooperative Learning

While there is a fair consensus among researchers about the positive effects of cooperative learning on student achievement, there remains a controversy about why and how cooperative learning methods affect achievement, and, most importantly, under what conditions cooperative learning has these effects. Different groups of researchers investigating cooperative learning effects on achievement begin with different assumptions and conclude by explaining the achievement effects of cooperative learning in terms that are substantially unrelated or contradictory. In

earlier work, Slavin (1995) identified motivationalist, social cohesion, cognitive developmental, and cognitive elaboration as the four major theoretical perspectives on the achievement effects of cooperative learning.

The motivationalist perspective presumes that task motivation is the single most impactful part of the learning process, asserting that the other processes, such as planning and helping, are driven by individuals' motivated self interest. Motivationalist-oriented scholars focus more on the reward or goal structure under which students operate, even going so far as to suggest that under some circumstances, interaction may not be necessary for the benefits of cooperative goal structures to manifest (Slavin, 1995). By contrast, the social cohesion perspective (also called social interdependence theory) suggests that the effects of cooperative learning are largely dependent on the cohesiveness of the group. This perspective holds that students help each other learn because they care about the group and its members and come to derive self-identity benefits from group membership (Johnson and Johnson, 1998). The two cognitive perspectives focus on the interactions among groups of students, holding that in themselves, these interactions lead to better learning and thus better achievement. Within the general cognitive heading, developmentalists attribute these effects to processes outlined by scholars such as Piaget and Vygotsky. Work from the cognitive elaboration perspective asserts that learners must engage in some manner of cognitive restructuring (elaboration) of new materials in order to learn them. Cooperative learning is said to facilitate that process. One reason for the continued lack of consensus among cooperative learning scholars is that each perspective tends to approach the topic without reference to the body of similar work from other perspectives and without attending to the larger picture.

This article offers a theoretical model of cooperative learning processes which intends to acknowledge the contributions of work from each of the major theoretical perspectives. It places them in a model that depicts the likely role each plays in cooperative learning. This work further explores conditions under which each may operate, and suggests research and development needed to advance cooperative learning scholarship so that educational practice may truly benefit from the lessons of 30 years of research.

The alternative perspectives on cooperative learning may be seen as complementary, not contradictory. For example, motivational theorists would not argue that the cognitive theories are unnecessary; instead, they would

* Portions of this paper are adapted from Slavin, 1995.

assert that motivation drives cognitive process, which in turn produces learning. They would argue that it is unlikely that over the long haul, students would engage in the type of elaborated explanations found by Webb (2008) to be essential to profit from cooperative activity, without a goal structure designed to enhance motivation. Similarly, social cohesion theorists might hold that the utility of extrinsic incentives must lie in their contribution to group cohesiveness, caring, and pro-social norms among group members, which could in turn affect cognitive processes.

A simple path model of cooperative learning processes, adapted from Slavin (1995), is diagramed in **Figure 1**, below. It depicts the main components of a group learning interaction, and represents the functional relationships among the major theoretical approaches to cooperative learning.

This diagram of the interdependent relationships among each of the components begins with a focus on group goals or incentives based on the individual learning of all group members. That is, the model assumes that the motivation to learn and to encourage and help others to learn activates cooperative behaviors that will result in learning. This would include both task motivation and motivation to interact in the group. In this model, motivation to succeed leads to learning directly, and also drives the behaviors and attitudes that lead to group cohesion, which in turn facilitates the types of group interactions – peer modeling, equilibration, and cognitive elaboration, which yield enhanced learning and academic achievement. The relationships are conceived to be reciprocal, such that as task motivation leads to the development of group cohesion, that development may reinforce and enhance task motivation. By the same token, the cognitive processes may become intrinsically rewarding and lead to increased task motivation and group cohesion.

Each aspect of the diagrammed model is well represented in the theoretical and empirical cooperative learning literature. All have well-established rationales and some supporting evidence. What follows is a review of the basic

theoretical orientation of each perspective, a description of the cooperative learning mode each prescribes, and a discussion of the empirical evidence supporting each.

Four Major Theoretical Perspectives on Cooperative Learning and Achievement

Motivational Perspectives

Motivational perspectives on cooperative learning presume that task motivation is the most important part of the process, believing that the other processes are driven by motivation. Therefore, these scholars primarily focus on the reward or goal structures under which students operate (see Slavin, 1995). From a motivationalist perspective (e.g., Johnson and Johnson, 1998; Slavin, 1983, 1995), cooperative incentive structures create a situation in which the only way group members can attain their own personal goals is if the group is successful. Therefore, to meet their personal goals, group members must both help their groupmates to do whatever enables the group to succeed, and, perhaps even more importantly, to encourage their groupmates to exert maximum efforts. In other words, rewarding groups based on group performance (or the sum of individual performances) creates an interpersonal reward structure in which group members will give or withhold social reinforcers (e.g., praise and encouragement) in response to groupmates' task-related efforts (see Slavin, 1983).

The motivationalist critique of traditional classroom organization holds that the competitive grading and informal reward system of the classroom creates peer norms opposing academic efforts (see Coleman, 1961). Since one student's success decreases the chances that others will succeed, students are likely to express norms that high achievement is for nerds or teachers' pets. However, by having students work together toward a common goal, they may be motivated to express norms favoring academic achievement, to reinforce one another for academic efforts.

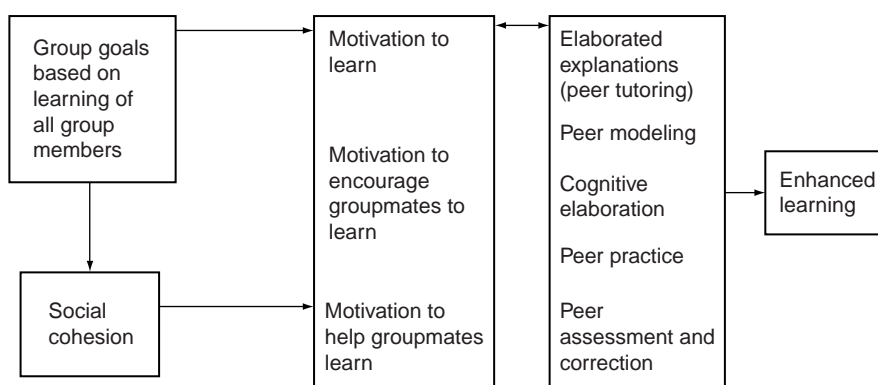


Figure 1 A model of achievement effects of cooperative learning. Adapted from Slavin, R. E. (1995). *Cooperative Learning: Theory, Research, and Practice*, 2nd edn. Boston, MA: Allyn and Bacon.

Not surprisingly, motivational theorists build group rewards into their cooperative learning methods. In methods developed at Johns Hopkins University (Slavin, 1994, 1995), students can earn certificates or other recognition if their average team scores on quizzes or other individual assignments exceed a preestablished criterion. Methods developed by Johnson *et al.* (1998) and his colleagues at the University of Minnesota often give students grades based on group performance, which is defined in several different ways. The theoretical rationale for these group rewards is that if students value the success of the group, they will encourage and help one another to achieve.

Empirical support for the motivational perspective

Considerable evidence from practical applications of cooperative learning in elementary and secondary schools supports the motivational position that group rewards are essential to the effectiveness of cooperative learning, with one critical qualification. Use of group goals or group rewards enhances the achievement outcomes of cooperative learning if and only if the group rewards are based on the individual learning of all group members (Slavin, 1995). Most often, this means that team scores are computed based on average scores on quizzes which all teammates take individually, without teammate help. For example, in Student Teams-Achievement Divisions, or STADs (Slavin, 1994), students work in mixed-ability teams to master material initially presented by the teacher. Following this, students take individual quizzes on the material, and the teams may earn certificates based on the degree to which team members have improved over their own past records. The only way the team can succeed is to ensure that all team members have learned, so the team members' activities focus on explaining concepts to one another, helping one another practice, and encouraging one another to achieve. In contrast, if group rewards are given based on a single group product (e.g., the team completes one worksheet or solves one problem), there is little incentive for group members to explain concepts to one another, and one or two group members may do all the work (see Slavin, 1995).

In assessing the empirical evidence supporting cooperative learning strategies, the greatest weight must be given to studies of longer duration. If well executed, these are bound to be more realistically generalizable to the day-to-day functioning of classroom practices. A review of 99 studies of cooperative learning in elementary and secondary schools that involved durations of at least 4 weeks compared achievement gains in cooperative learning and control groups. Of 64 studies of cooperative learning methods that provided group rewards based on the sum of group members' individual learning, 50 (78%) found significantly positive effects on

achievement, and none found negative effects (Slavin, 1995). The median effect size for the studies from which effect sizes could be computed was +0.32 (32% of a standard deviation separated cooperative learning and control treatments). In contrast, studies of methods that used group goals based on a single group product or provided no group rewards found few positive effects, with a median effect size of only +0.07. Comparisons of alternative treatments within the same studies found similar patterns; group goals based on the sum of individual learning performances were necessary to the instructional effectiveness of the cooperative learning models (e.g., Fantuzzo *et al.*, 1989, 1990).

Social Cohesion Perspective

A theoretical perspective somewhat related to the motivational viewpoint holds that the effects of cooperative learning on achievement are strongly mediated by the cohesiveness of the group. The quality of the group's interactions is thought to be largely determined by group cohesion. In essence, students will engage in the task and help one another learn because they identify with the group and want one another to succeed. This perspective is similar to the motivational perspective in that it emphasizes primarily motivational rather than cognitive explanations for the instructional effectiveness of cooperative learning. However, motivational theorists hold that students help their groupmates learn primarily because it is in their own interests to do so. Social cohesion theorists, in contrast, emphasize the idea that students help their groupmates learn because they care about the group. A hallmark of the social cohesion perspective is an emphasis on teambuilding activities in preparation for cooperative learning, and processing or group self-evaluation during and after group activities. Social cohesion theorists have historically tended to downplay or reject the group incentives and individual accountability held by motivationalist researchers to be essential. They emphasize, instead, that the effects of cooperative learning on students and student achievement depend substantially on the quality of the group's interaction (Battisch *et al.*, 1993). For example, Cohen (1994: 69, 70) stated "if the task is challenging and interesting, and if students are sufficiently prepared for skills in group process, students will experience the process of groupwork itself as highly rewarding . . . never grade or evaluate students on their individual contributions to the group product." Cohen's (1994) work, as well as that of Sharan and Sharan (1992) and Elliot Aronson (Aronson *et al.*, 1978) and his colleagues, may be described as social cohesiveness theories. Cohen, Aronson, and the Sharans all use forms of cooperative learning in which students take on individual roles within the group, which Slavin (1983) calls "task specialization" methods. In Aronson's Jigsaw method, students study material on one of four or five topics

distributed among the group members. They meet in expert groups to share information on their topics with members of other teams who had the same topic, and then take turns presenting their topics to the team. In the Sharans' Group Investigation method, groups take on topics within a unit studied by the class as a whole, and then further subdivide the topic into tasks within the group. The students investigate the topic together and ultimately present their findings to the class as a whole. Cohen's Finding Out/Descubrimiento program has students play different roles in discovery-oriented science activities.

One main purpose of the task specialization used in Jigsaw, Group Investigation, and Finding Out/Descubrimiento is to create interdependence among group members. In the Johnsons' methods, a somewhat similar form of interdependence is created by having students take on roles as checker, recorder, observer, and so on. The idea is that if students value their groupmates (as a result of teambuilding and other cohesiveness-building activities) and are dependent on one another, they are likely to encourage and help one another to succeed.

Empirical support for the social cohesion perspective

There is some evidence that the achievement effects of cooperative learning depend on social cohesion and the quality of group interactions (Battisch *et al.*, 1993). The achievement outcomes of cooperative learning methods that emphasize task specialization are less clear. Research on the original form of Jigsaw has not generally found positive effects of this method on student achievement (Slavin, 1995). One problem with this method is that students have limited exposure to material other than that which they studied themselves, so learning gains on their own topics may be offset by losses on their groupmates' topics. In contrast, there is evidence that when it is well implemented, Group Investigation can significantly increase student achievement (Sharan and Shachar, 1988). In studies of at least 4 weeks' duration, the Johnson *et al.* (1998) methods have not been found to increase achievement more than individualistic methods unless they incorporate group rewards (in this case, group grades) based on the average of group members' individual quiz scores (see Slavin, 1995). Studies of forms of Jigsaw that have added group rewards to the original model have found positive achievement outcomes (Mattingly and Van Sickle, 1991).

Research on practical classroom applications of methods based on social cohesion theories provides inconsistent support for the proposition that building cohesiveness among students through teambuilding alone (i.e., without group incentives) will enhance student achievement. In general, methods which emphasize teambuilding and group process, but do not provide specific group rewards based on the learning of all group members, are no more effective than

traditional instruction in increasing achievement (Slavin, 1995), although there is evidence that these methods can be effective if group rewards are added to them.

Cognitive Perspectives

The major alternative to the motivationalist and social cohesiveness perspectives on cooperative learning, both of which primarily focus on group norms and interpersonal influence, is the cognitive perspective. The cognitive perspective holds that interactions among students will in themselves increase student achievement for reasons which have to do with mental processing of information rather than with motivations. Cooperative methods developed by cognitive theorists involve neither the group goals that are the cornerstone of the motivationalist methods nor the emphasis on building group cohesiveness characteristic of the social cohesion methods. However, there are several quite different cognitive perspectives, as well as some which are similar in theoretical perspective, but have developed on largely parallel tracks. The two most notable of these are described in the following sections.

Developmental perspectives

One widely researched set of cognitive theories is the developmental perspective (e.g., Damon, 1984). The fundamental assumption of the developmental perspective on cooperative learning is that interaction among children around appropriate tasks increases their mastery of critical concepts. Vygotsky (1978: 86) defines the zone of proximal development as "... the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in *collaboration with more capable peers*" (emphasis added). In his view, collaborative activity among children promotes growth because children of similar ages are likely to be operating within one another's proximal zones of development, modeling in the collaborative group behaviors more advanced than those they could perform as individuals.

Similarly, Piaget (1926) held that social-arbitrary knowledge – language, values, rules, morality, and symbol systems – can only be learned in interactions with others. Peer interaction is also important in logical-mathematical thought in disequilibrating the child's egocentric conceptualizations and in providing feedback to the child about the validity of logical constructions.

There is a great deal of empirical support for the idea that peer interaction can help nonconservers become conservers. Many studies have shown that when conservers and nonconservers of about the same age work collaboratively on tasks requiring conservation, the nonconservers generally develop and maintain conservation concepts (see Bell *et al.*, 1985). From the developmental perspective, the effects of cooperative learning on student

achievement would be largely or entirely due to the use of cooperative tasks. In this view, opportunities for students to discuss, argue, and to present and hear one another's viewpoints constitute the critical elements of cooperative learning with respect to student achievement.

Empirical evidence for the developmental perspective

Despite considerable support from theoretical and laboratory research, there is little evidence, from classroom experiments conducted over meaningful time periods, that pure cooperative methods, which depend solely on interaction, do produce higher achievement. However, it is likely that the cognitive processes described by developmental theorists are important mediating variables, which can help explain the positive outcomes of effective cooperative learning methods (Slavin, 1995).

Cognitive elaboration perspectives

A cognitive perspective on cooperative learning quite different from the developmental viewpoint is one which might be called the cognitive elaboration perspective. Research in cognitive psychology has long held that if information is to be retained in memory and related to information already in memory, the learner must engage in some sort of cognitive restructuring, or elaboration of the material (Wittrock, 1986). One of the most effective means of elaboration is explaining the material to someone else. Research on peer tutoring has long found achievement benefits for the tutor as well as the tutee (Devin-Sheehan *et al.*, 1976). In this method, students take roles as recaller and listener. They read a section of text, and then the recaller summarizes the information while the listener corrects any errors, fills in any omitted material, and helps think of ways by which both students can remember the main ideas. The students switch roles on the next section.

Empirical evidence for the cognitive elaboration perspective

Donald Dansereau and his colleagues at Texas Christian University have found in a series of brief studies that college students working on structured cooperative scripts can learn technical material or procedures far better than can students working alone (O'Donnell, 1996). Dansereau and his colleagues found in a series of studies that while both the recaller and the listener learned more than did students working alone, the recaller learned more (O'Donnell and Dansereau, 1992). This mirrors both the peer-tutoring findings as well as those of Noreen Webb (2008), who discovered that the students who gained the most from cooperative activities were those who provided elaborated explanations to others. In this research as well as in Dansereau's, students who received elaborated explanations learned more than those who worked alone, but not as much as those who served as explainers. Studies

of reciprocal teaching, in which students learn to formulate questions for each other, have generally supported its positive effects on student achievement (Palincsar *et al.*, 1987; Rosenshine and Meister, 1994; O'Donnell, 2000).

Structuring Group Interactions

There is some evidence that carefully structuring the interactions among students in cooperative groups can be effective, even in the absence of group rewards. For example, Meloth and Deering (1992) compared students working in two cooperative conditions. In one group, students were taught specific reading comprehension strategies and given think sheets to remind them to use these strategies (e.g., prediction, summarization, and character mapping). In the other group, students earned team scores if their members improved each week on quizzes. A comparison of the two groups on a reading comprehension test found greater gains for the strategy group.

However, there is also evidence to suggest that a combination of group rewards and strategy training produces much better outcomes than either alone. Fantuzzo *et al.* (1992) directly made a comparison between rewards alone, strategy alone, and a combination, and found the combination to be by far the most effective. Further, the outcomes of dyadic learning methods, which use group rewards as well as strategy instruction, produced some of the largest positive effects of any cooperative methods, much larger than those found in studies that provided groups with structure but not rewards. As noted earlier, studies of scripted dyads also find that adding incentives adds to the effects of these strategies (O'Donnell, 1996). The consistent positive findings for Cooperative Integrated Reading and Composition (CIRC) (Stevens *et al.*, 1987), which uses both group rewards and strategy instruction, also argue for this combination.

Reconciling the Four Perspectives

The model shown above in **Figure 1** illustrates how group goals might operate in enhancing the learning outcomes of cooperative learning. Provision of group goals based on the individual learning of all group members might affect cognitive processes directly, by motivating students to engage in peer modeling, cognitive elaboration, and/or practice with one another. Group goals may also lead to group cohesiveness, increasing caring and concern among group members and making them feel responsible for one another's achievement, thereby motivating students to engage in cognitive processes which enhance learning. Finally, group goals may motivate students to take responsibility for one another independently of the teacher, thereby solving important classroom organization

problems and providing increased opportunities for cognitively appropriate learning activities. Scholars whose theoretical orientations deemphasize the utility of extrinsic rewards attempt to intervene directly on mechanisms identified as mediating variables in the model described earlier. For example, social cohesion theorists intervene directly on group cohesiveness by engaging in elaborate teambuilding and group processing training. Cognitive theorists would hold that the cognitive processes that are essential to any theory relating cooperative learning to achievement can be created directly, without the motivational or affective changes discussed by the motivationalist and social cohesion theorists.

From the perspective of the model diagrammed in **Figure 1**, starting with group goals and individual accountability permit students in cooperative learning groups to benefit from the full range of factors that are known to affect cooperative learning outcomes. While group goals and individual accountability may not always be absolutely necessary, to ignore them would be to ignore the tool with the most consistent evidence of positive effects on student achievement.

In summary, although cooperative learning has been studied in an extraordinary number of field experiments of high methodological quality, there is still much more to be done. Cooperative learning has the potential to become a primary format used by teachers to achieve both traditional and innovative goals. Research must continue to provide the practical, theoretical, and intellectual underpinnings to enable educators to achieve this potential. This article has advanced a cohesive model of the relationships among the important variables involved in the functioning of cooperative learning. It offers a framework for discussion and continued debate while calling for a move away from competitive attempts to explain this complex phenomenon toward a unified theoretical model which can guide future research efforts and inform education practice.

Acknowledgment

This work was supported by funding from the Institute of Education Sciences, US Department of Education (Grant No. R305A040082).

See also: Peer Interaction and Learning; Piaget: Recent Work; Vygotsky and Recent Developments.

Bibliography

- Antil, L. R., Jenkins, J. R., Wayne, S., and Vadasy, P. F. (1998). Cooperative learning: Prevalence, conceptualizations, and the relation between research and practice. *American Educational Research Journal* **35**(3), 419–454.
- Aronson, E., Blaney, N., Stephan, C., Sikes, J., and Snapp, M. (1978). *The Jigsaw Classroom*. Beverly Hills, CA: Sage.
- Battisch, V., Solomon, D., and Delucci, K. (1993). Interaction process and student outcomes in cooperative learning groups. *Elementary School Journal* **94**(1), 19–32.
- Bell, N., Grossen, M., and Perret-Clermont, A.-N. (1985). Socio-cognitive conflict and intellectual growth. In Berkowitz, M. (ed.) *Peer Conflict and Psychological Growth*, pp 88–112. San Francisco, CA: Jossey-Bass.
- Cohen, E. G. (1994). *Designing Groupwork: Strategies for the Heterogeneous Classroom*, 2nd edn, pp 69–70. New York: Teachers College Press.
- Coleman, J. (1961). *The Adolescent Society*. New York: Free Press.
- Damon, W. (1984). Peer education: The untapped potential. *Journal of Applied Developmental Psychology* **5**, 331–343.
- Devin-Sheehan, L., Feldman, R., and Allen, V. (1976). Research on children tutoring children: A critical review. *Review of Educational Research* **46**(3), 355–385.
- Fantuzzo, J., King, J., and Heller, L. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment: A component analysis. *Journal of Educational Psychology* **84**(3), 331–339.
- Fantuzzo, J. W., Polite, K., and Grayson, N. (1990). An evaluation of reciprocal peer tutoring across elementary school settings. *Journal of School Psychology* **28**, 309–323.
- Fantuzzo, J. W., Riggio, R. E., Connelly, S., and Dimeff, L. A. (1989). Effects of reciprocal peer tutoring on academic achievement and psychological adjustment: A component analysis. *Journal of Educational Psychology* **81**, 173–177.
- Johnson, D., Johnson, R., and Holubec, E. (1998). *Cooperation in the Classroom*. Boston: Allyn & Bacon.
- Johnson, D. W. and Johnson, R. T. (1998). *Learning Together and Alone: Cooperative, Competitive, and Individualistic Learning*, 5th edn. Boston, MA: Allyn and Bacon.
- Mattingly, R. M. and Van Sickle, R. L. (1991). Cooperative learning and achievement in social studies: Jigsaw II. *Social Education* **55**(6), 392–395.
- Meloth, M. S. and Deering, P. D. (1992). The effects of two cooperative conditions on peer group discussions, reading comprehension, and metacognition. *Contemporary Educational Psychology* **17**, 175–193.
- O'Donnell, A. M. (1996). The effects of explicit incentives on scripted and unscripted cooperation. *Journal of Educational Psychology* **88**, 74–86.
- O'Donnell, A. M. (2000). Interactive effects of prior knowledge and material format on cooperative teaching. *Journal of Experimental Education* **68**(2), 101–108.
- O'Donnell, A. M. and Dansereau, D. F. (1992). Scripted cooperation in student dyads: A method for analyzing and enhancing academic learning and performance. In Hertz-Lazarowitz, R. and Miller, N. (eds.) *Interaction in Cooperative Groups: The Theoretical Anatomy of Group Learning*, pp 120–144. New York: Cambridge University Press.
- Palincsar, A. S., Brown, A. L., and Martin, S. M. (1987). Peer interaction in reading comprehension instruction. *Educational Psychologist* **22**, 231–253.
- Piaget, J. (1926). *The Language and Thought of the Child*. New York: Harcourt Brace.
- Puma, M. J., Jones, C. C., Rock, D., and Fernandez, R. (1993). Prospects: The congressionally mandated study of educational growth and opportunity. *Interim Report*. Bethesda, MD: Abt Associates.
- Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., and Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology* **94**(20), 240–257.
- Rosenshine, B. and Meister, C. (1994). Reciprocal teaching: A review of research. *Review of Educational Research* **64**, 490–530.
- Sharan, S. and Shachar, C. (1988). *Language and Learning in the Cooperative Classroom*. New York: Springer.
- Sharan, Y. and Sharan, S. (1992). *Expanding Cooperative Learning through Group Investigation*. New York: Teachers College Press.
- Slavin, R. E. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin* **94**, 429–445.

- Slavin, R. E. (1994). *Using Student Team Learning*, 2nd edn. Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools.
- Slavin, R. E. (1995). *Cooperative Learning: Theory, Research, and Practice*, 2nd edn. Boston, MA: Allyn and Bacon.
- Stevens, R. J., Madden, N. A., Slavin, R. E., and Farnish, A. M. (1987). Cooperative integrated reading and composition: Two field experiments. *Reading Research Quarterly* **22**, 433–454.
- Vygotsky, L. S. (1978). *Mind in Society*, Cole, M., John-Steiner, V., Scribner, S., and Souberman, E. (eds.). Cambridge, MA: Harvard University Press.
- Webb, N. M. (2008). Learning in small groups. In Good, T. L. (ed.) *21st Century Education: A Reference Handbook*, pp 203–211. Los Angeles, CA: Sage.
- Wittrock, M. C. (1986). Students' thought processes. In Wittrock, M. C. (ed.) *Handbook of Research on Teaching*, 3rd edn, pp 297–314. New York: Macmillan.

Intelligence

R J Sternberg, Tufts University, Boston, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Crystallized intelligence – One of two major subfactors of general intelligence, it represents the accumulation of knowledge over the life span of the individual and may be measured by tests in areas such as vocabulary, general information, and achievement (cf. fluid intelligence).

Deviation intelligence quotients (IQs) – A means of determining intelligence test scores, based on deviations from an average score, calculated such that the normative equivalent for the median score is 100, about 68% of the scores are computed to fall between 85 and 115, and about 95% of the scores fall between 70 and 130; strictly speaking, they are not IQs because no quotient is involved (cf. mental age and ratio IQ).

Emotional intelligence – The ability to perceive and express emotion, assimilate emotion in thought, understand and reason with emotion, and regulate emotion in the self and others.

Fluid intelligence – One of two major subfactors of general intelligence, it represents the acquisition of new information or the grasping of new relationships and abstractions about known information (may be measured, e.g., by timed tests involving analogies, series completions, or inductive reasoning; cf. crystallized intelligence).

Heritability coefficient – The degree to which heredity contributes to individual differences in intelligence, expressed in terms of a number on a scale from 0 to 1, such that a coefficient of 0 means that heredity has no influence on variation among people, whereas a coefficient of 1 means that heredity is the only influence on such variation.

Intelligence – The ability to implement goal-directed adaptive behavior.

Mental age – A means of indicating a person's level of intelligence (generally in reference to a child), based on the individual's performance on tests of intelligence, by indicating the chronological age of persons who typically perform at the same level of intelligence as the test-taker (used less frequently today than in the past; cf. deviation IQs, intelligence, and ratio IQ).

Mental retardation – Low level of intelligence, usually reflected by both poor performance on tests of intelligence and poor adaptive competence

(the degree to which a person functions effectively within a normal situational context).

Ratio IQ – A means of indicating performance on intelligence tests, based on a quotient of mental age divided by chronological age, times 100 (cf. deviation IQs).

Reaction range – The broad limits within which a particular attribute (e.g., intelligence) may be expressed in various possible ways, given the inherited potential for its expression in the particular individual.

Theory of multiple intelligences – A theory suggesting that intelligence comprises eight distinct constructs – bodily–kinesthetic intelligence, interpersonal intelligence, intrapersonal intelligence, linguistic intelligence, mathematical–logical intelligence, musical intelligence, naturalist intelligence, and spatial intelligence – that function somewhat independently, but may interact to produce intelligent behavior.

Triarchic theory of successful intelligence – A theory of intelligence that asserts that intelligence comprises three aspects, which deal with the relationship of intelligence to the internal world, to experience, and to the external world.

Theories of Intelligence

Implicit Theories

How do psychologists or educators know about what intelligence is? One way they find out is by asking people. For example, they might ask experts. What do experts say?

In a 1921 symposium on the definition of intelligence, an American psychologist, Lewis M. Terman, emphasized the ability to think abstractly. Another American psychologist, Edward L. Thorndike, emphasized learning and the ability to give good responses to questions. In a similar 1986 symposium, however, psychologists generally agreed on the importance of adaptation to the environment as the key to understanding both what intelligence is and what it does. They also emphasized learning skills and understanding one's own cognitive processes. Therefore, adaptation, learning, and metacognition (i.e., understanding oneself) won the day.

In a set of studies published in 1981, Robert J. Sternberg and his colleagues asked laypeople in the United States what they thought intelligence was. Three factors emerged from their responses: practical problem solving, verbal ability, and social competence. The conceptions that arise depend, however, on who is asked.

Robert J. Sternberg and his colleagues have studied implicit theories in various cultures. In comparable studies done in Taiwan, four factors prevailed: cognitive skills, getting along with other people, understanding oneself, and knowing when to and when not to show that one is smart. In other studies conducted in Kenya, cognitive skills also were emphasized less than in the West. In contrast, obedience, respect, and understanding people were emphasized more.

Most theories of intelligence are explicit rather than implicit. They are elicited not by asking people what they mean by intelligence, but rather, by having people perform tasks believed to require intelligence. There are several different kinds of explicit theories.

Psychometric Theories

Psychometric theories have generally sought to understand the structure of intelligence: What form does it take, and what are its parts, if any? Such theories have generally been based on and tested by the use of data obtained from tests of mental abilities. These tests include assessments of vocabulary, numerical reasoning, analogical reasoning, and visualization of what forms would look like if they were rotated in space.

The first major psychometric theories were proposed by the British psychologist Charles E. Spearman. In a 1904 article, Spearman argued that just two kinds of factors underlie virtually all individual differences in test scores. Spearman called the first and more important kind of factor the general factor or *g*. It was said to pervade performance on all tasks requiring intelligence. The second kind of factor, according to Spearman, was specific to each test. However, what exactly is *g*? In 1927, Spearman proposed it might be something he labeled mental energy.

An American psychologist, L. L. Thurstone, instead suggested that seven factors, or primary mental abilities, underlie individual differences in mental test performance: verbal comprehension (knowledge of vocabulary and in reading); verbal fluency (writing and producing words in response to a prompt, such as words beginning with the letter d); number (solving simple arithmetical computation and reasoning problems); spatial visualization (mentally visualizing and manipulating objects); inductive reasoning (completing a number or letter series); memory (remembering people's names or faces); and perceptual speed (rapidly proofreading to discover typographical errors in a typed text).

Raymond B. Cattell and John B. Carroll, among others, suggested that abilities are hierarchical. At the top of the hierarchy is *g* or general ability. Below *g* in the hierarchy are successive levels of gradually narrowing abilities, ending with Spearman's specific abilities.

Cattell suggested that general ability can be divided into two basic abilities, fluid and crystallized. Fluid abilities are the reasoning and problem-solving skills measured by tests such as the analogies, classifications, and series completions. Crystallized abilities derive from fluid abilities and are viewed as their products, which include vocabulary, general information, and knowledge about specific fields. John L. Horn suggested that crystallized ability more or less increases over the life span, whereas fluid ability increases in the earlier years and decreases in the later ones.

John B. Carroll proposed a three-stratum model of intelligence. It is considered by some to be the most definitive psychometric model of intelligence because it is based upon reanalyses of hundreds of data sets. According to this model, general ability is at the top of a hierarchy of abilities. At the next lower stratum are various broad abilities (including learning and memory processes and the effortless production of many ideas). At the bottom of the hierarchy are many narrow, specific abilities, such as spelling ability and reasoning speed.

J. P. Guilford, an American psychologist, proposed a structure-of-intellect theory, which, in its earlier versions, postulated 120 abilities. For example, in an influential 1967 work, Guilford argued that abilities can be divided into five kinds of operations, four kinds of contents, and six kinds of products. These various facets of intelligence combine multiplicatively for a total of $5 \times 4 \times 6$, or 120 separate abilities. An example of such an ability would be cognition (operation) of semantic (content) relations (product), which would be involved in recognizing the relation between lawyer and client in the analogy problem, lawyer:client::doctor:?. In 1984, Guilford increased the number of abilities proposed by his theory, raising the total to 150.

In an address to the American Psychological Association in 1957, Lee J. Cronbach proposed that psychologists unite the two disciplines of scientific psychology – experimental and differential (the study of individual differences). His proposal led to cognitive theories of intelligence, which are derived and tested by experimental means.

Underlying most cognitive approaches to intelligence is the assumption that intelligence comprises a set of mental representations (e.g., propositions and images) of information and a set of mental processes that can operate on the representations. A more intelligent person is assumed to mentally represent information better and also to operate more quickly on these representations than a less intelligent person does.

In 1975, Earl B. Hunt, Clifford E. Lunneborg, and J. Lewis showed that a critical ability underlying verbal intelligence is that of rapidly retrieving lexical information, such as letter names, from memory. A few years later, Robert J. Sternberg identified key mental processes alleged to underlie many cognitive tasks, especially ones involving inductive reasoning. The processes included, among others, encoding stimuli, inferring relations between stimuli, and applying what one has learned.

Ian Deary and his colleagues have sought to understand intelligence through the study of inspection time. They have found that more intelligent individuals can discriminate the lengths of the lines with lesser stimulus duration (inspection) times than less intelligent individuals can.

Two leaders in the field of cognitive psychology, Allen Newell and Herbert A. Simon, used computers to model intelligence. The underlying idea is that computers can, in some sense, show a kind of intelligence similar to that shown by humans. In 1972, Newell and Simon proposed a general theory of problem solving, much of which was implemented on the computer. It involves heuristics such as means–ends analysis, whereby one seeks, at each step of problem solving, to reduce as much as possible the distance between the solution and where one is in the problem.

Marcel Just and Patricia Carpenter showed that complex intelligence test items, such as figural matrix problems involving reasoning with geometric shapes, could be solved by a computer program at a level of accuracy comparable to that of human test-takers.

The models described above are serial-processing models, whereby the computer takes steps in sequence. David E. Rumelhart and Jay L. McClelland, proposed what they call parallel distributed processing models of the mind. These models postulate that many types of information processing occur at once, rather than just one at a time.

Perhaps the dominant cognitive approach today is one that stresses the role of working memory in intelligence. Patrick Kyllonen, Randall Engle, and others have suggested that the main source of differences in performance on intellectual tasks is in people's differential working-memory capacities, that is, their ability to remember and manipulate recently presented information in their minds.

Cognitive-Contextual Theories

Cognitive-contextual theories deal with the way cognitive processes operate in various environmental contexts. Two of the major theories of this type have been proposed by Howard Gardner and Robert Sternberg.

In 1983, Gardner proposed a theory of what he called multiple intelligences. In the 1999 version of the theory, the multiple intelligences include, at minimum, linguistic, logical–mathematical, spatial, musical, bodily–kinesthetic, naturalist, interpersonal, and intrapersonal intelligences. Gardner has also speculated about the existence of an existential intelligence.

An alternative theory, also taking into account both cognition and context, is Sternberg's triarchic theory of successful intelligence. According to Sternberg, intelligence has three aspects. These aspects relate intelligence to what goes on internally within a person, what goes on in the external world, and to experience, which mediates between the internal and external worlds.

The first aspect is the set of cognitive processes and representations that form the core of all thought. Sternberg has distinguished three kinds of processes: those involved in deciding what to do and, later, in deciding how well it was done; those involved in doing what one has decided to do; and those involved in learning how to do it in the first place. The second aspect is the application of these processes to the external world. According to Sternberg, mental processes serve three functions in the everyday world: adaptation to existing environments, the shaping of existing environments into new ones, and the selection of new environments when old ones prove unsatisfactory. According to the theory, more successful intelligent persons are not just those who can execute many cognitive processes quickly or well. Greater intelligence is additionally reflected in knowing what one's strengths and weaknesses are and capitalizing upon strengths while remedying or compensating for weaknesses. Successfully intelligent persons, then, find a niche in which they can operate effectively.

The third aspect of Sternberg's triarchic theory is the integration of the internal and external worlds through experience. One measure of intelligence is the ability to cope with relatively novel situations. The abilities to cope with relative novelty and to automatize cognitive processing are seen as interrelated: The more a person is able to automatize the tasks of daily life, the more mental resources there are left to cope with novelty.

Peter Salovey and John Mayer proposed the construct of emotional intelligence. It was popularized by Daniel Goleman in a 1995 book. It is the ability to perceive accurately, appraise, and express emotion; the ability to access and/or generate feelings when they facilitate thought; the ability to understand emotion and emotional knowledge; and the ability to regulate emotions to promote emotional and intellectual growth. Several tests are now available to measure emotional intelligence. They generally show modest to moderate correlations with conventional tests of intelligence. They also predict various forms of real-world behavior.

Biological Theories

Biological theories are based in the neuropsychological functioning that produces, or at least is correlated with, intelligent behavior. Several biological approaches have been proposed.

One approach has been the investigation of the types of intellectual performance as related to the regions of the brain from which they originate. An American researcher in this area, Jerre Levy, found that the left hemisphere is superior in analytical functioning, of which a prime example is the use of language. The right hemisphere is superior in many forms of visual and spatial performance and tends to be more synthetic and holistic in its functioning than the left.

A second approach to research has involved the use of brain-wave recordings to study the relation between these waves and either performance on ability tests or in various kinds of cognitive tasks. Researchers, such as British researcher P. G. Caryl, have found a relationship between certain aspects of electroencephalogram (EEG) and event-related potential (ERP) waves and scores on standard psychometric tests of intelligence.

A third approach involves the measurement of blood flow in the brain, which is a fairly direct indicator of functional activity in brain tissue. In such studies, the amount and location of blood flow in the brain is monitored while subjects perform cognitive tasks. Using positron emission tomography (PET), Richard Haier discovered that people who perform better on conventional tests of intelligence often show less activation in relevant portions of the brain than those who do not perform as well. This pattern of results suggests that the better performers find the tasks easier and thus invoke less effort than the poorer performers do.

Development of Intelligence

There have been diverse approaches to studying the development of intelligence. Psychometric theorists, for instance, have sought to understand how intelligence develops in terms of changes in the factors of intelligence over time and changes in the amounts of the various abilities that children have. For example, the concept of mental age was popular during the first half of the twentieth century. A given mental age was believed to represent an average child's level of mental functioning for a given chronological age. Thus, an average 10-year-old would have a mental age of 10, but an above-average 8-year-old or a below-average 12-year-old might also have a mental age of 10 years. The concept of mental age has fallen into disfavor, however, and is used only rarely now. The concept does not work well over a chronological age of roughly 16. Moreover, its assumption of perfectly smooth continuous mental development is questionable.

The Theory of Jean Piaget

Jean Piaget, a Swiss psychologist, suggested that the child explores the world, observes regularities, and makes generalizations, much as a scientist does. Two fundamental cognitive processes are alleged to work in a somewhat reciprocal fashion. Assimilation involves incorporating new information into an already existing cognitive structure. Accommodation involves forming a new cognitive structure to incorporate new information. Cognitive development, according to Piaget, represents a dynamic equilibrium between these two processes of assimilation and accommodation.

Piaget also postulated that there are four major periods in intellectual development. The first, the sensorimotor period, extends from birth until roughly 2 years of age. During this period, a child learns how to modify reflexes to make them more adaptive, coordinate actions, retrieve hidden objects, and, eventually, to begin representing information mentally. During the second, preoperational period, from about 2 to 7 years of age, a child experiences the growth of language and mental imagery. He or she also learns to focus on single perceptual dimensions, such as color and shape. The third, concrete-operational period, is from about 7 to 12 years of age. It is the period during which a child develops an important set of skills that are referred to as conservation skills. The child will recognize that substances stay the same in amount, regardless of their form. Finally, children emerge into the fourth, formal-operational period, which begins at about age 12 and continues throughout life. The formal-operational child develops thinking skills in all logical combinations and learns to think with abstract concepts.

The Theory of Lev Vygotsky

Lev S. Vygotsky, a Soviet psychologist, suggested that intellectual development may be largely influenced by a child's interactions with others: a child sees others thinking and acting in certain ways and then internalizes and models what is seen. Vygotsky also proposed the notion of a zone of proximal development (ZPD), which is the range of ability between a person's observable level of ability and the person's latent capacity. This latent capacity is not directly observable, but may be detected by providing a context in which the latent capacity may be revealed and expressed. The ZPD is sometimes termed the zone of potential development.

Measuring Intelligence

Early Historical Background

The publication of Charles Darwin's *The Origin of Species* in 1859 had a profound effect on many lines

of scientific work. Darwin suggested that the capabilities of humans are, in some sense, continuous with those of lower animals. Hence, they can be understood through scientific investigation. One person who was strongly influenced by Darwin's thinking was his cousin, Sir Francis Galton. For 7 years – between 1884 and 1890 – Galton maintained a laboratory at the South Kensington Museum in London. For a small fee, visitors could have themselves measured on a variety of psychophysical tasks. These tasks included weight discrimination and sensitivity to musical pitch. Galton believed that these tests measured more than just psychophysical abilities. He believed that psychophysical abilities are the basis of intelligence and, hence, that his tasks were measures of intelligence. Galton's intelligence test, therefore, required a person to perform simple tasks, such as deciding which of two weights was heavier or showing how forcefully he could squeeze his hand. The Galtonian tradition was taken to the United States by the psychologist James McKeen Cattell.

The Intelligence Quotient Test

A more influential tradition of mental testing was developed in France by Alfred Binet and his collaborator, Theodore Simon. In 1904, the minister of public instruction in Paris named a commission to study or create tests that would insure that mentally retarded children received an adequate education. The minister was also concerned that certain children were being placed in classes for the retarded not because they were retarded, but because they had behavior problems, and teachers did not want them in their classrooms. Binet and Simon proposed that tests of intelligence should measure skills such as judgment, comprehension, and reasoning – the same kinds of skills measured on most intelligence tests today. Binet's early test was taken to the United States by a Stanford University psychologist, Lewis Terman, whose version came to be called the Stanford-Binet test. This test has been revised frequently and continues to be in use.

The Stanford-Binet test and others similar to it have traditionally yielded, at the very least, an overall score referred to as an intelligence quotient (IQ). In its most recent form, this test as well as the Wechsler Adult Intelligence Scale and the Wechsler Intelligence Scale for Children yield an overall IQ and other scores as well.

More recent tests of intelligence have expanded the range of abilities tested. For example, in 1997, J. P. Das and Jack Naglieri produced a test, the Cognitive Assessment System, based on a theory of intelligence first proposed by a Russian psychologist, Alexander Luria. The test measures planning abilities, attentional abilities, and simultaneous and successive processing abilities.

IQ was originally computed as the ratio of mental age to chronological (physical) age, multiplied by 100. It is

thus sometimes referred to as a ratio IQ. For example, if a child of age 8 had a mental age of 10 (i.e., performed on the test at the level of an average 10-year-old), the child was assigned an IQ of $(10/8) \times 100$, or 125. If the 8-year-old had a mental age of 6, the child's IQ would be $(6/8) \times 100$, or 75. A score of 100, whereby the mental age equals the chronological age, is average.

As discussed above, the concept of mental age has fallen into disrepute, and few tests continue to involve the computation of mental ages. Many tests still yield an IQ, but it is most often computed on the basis of statistical distributions. The scores are assigned on the basis of what percentage of people of a given group would be expected to have a certain IQ. Scores computed in this way are called deviation IQs.

The Distribution of IQ Scores

Intelligence test scores follow an approximately normal distribution, meaning that most people score near the middle of the distribution of scores. Scores drop off fairly rapidly in frequency in either direction from the center of the distribution. For example, on the IQ scale, about two-thirds of all scores fall between IQs of 85 and 115, and about 95% of scores fall between 70 and 130. Put another way, only one out of 20 scores differs from the average IQ (100) by more than 30 points.

It has been common to associate certain levels of IQ with labels. For example, at the upper end, the label gifted is sometimes assigned to people with IQs over a certain point, such as 130. In addition, at the lower end, mental retardation has been classified into different degrees depending upon IQ; therefore, for example, IQs of 70–84 have been classified as borderline retarded, 55–69 as mildly retarded, 40–54 as moderately retarded, 25–39 as severely retarded, and IQs below 25 as profoundly retarded.

Many psychologists now believe that IQ represents only a part of intelligence, and intelligence is only one factor in both mental retardation and giftedness. Most current definitions of mental retardation stress adaptive skills as well as IQ and also emphasize attributes such as creativity, motivation, and achievement in conceptions of giftedness.

The Heritability and Malleability of Intelligence

Historically, intelligence has been viewed as a more or less fixed trait. This view conceives intelligence as something people are born with. The function of development is thus to allow this genetic endowment to express itself. A number of investigators have suggested that intelligence is highly heritable, and that it is transmitted through

the genes. Heritability is here defined as the proportion of individual-differences variance that is genetically transmitted. Other investigators believe that intelligence is minimally heritable, if at all. Most authorities take an intermediate position. However, all agree that heritability operates within a reaction range, meaning that a given genotype for intelligence can result in a wide range of phenotypes, depending on how environmental factors interact with genetic ones.

Several methods are used to assess the heritability of intelligence. The most well known is perhaps the study of identical twins reared apart. For a variety of reasons, identical twins are occasionally separated at or near birth. If the twins are raised apart, and if it is assumed that when twins are separated, they are randomly distributed across environments (often a dubious assumption), then the twins would heritably have in common all of their genes, but none of their environment, except for chance environmental overlap. As a result, the correlation between their performances on tests of intelligence can provide an estimate of the proportion of variation in test scores due to heredity. Another method of computing the hereditary effect on intelligence involves comparing the relationship between intelligence test scores of identical twins and those of fraternal twins.

It appears that roughly half the variation in intelligence test scores is caused by hereditary influences. However, Robert Plomin and others have shown that the heritability of intelligence increases with age, suggesting that genetic factors become more important and environmental factors less important to individual differences in intelligence with increasing age. In adulthood, heritability may reach as high as 70% or more. The estimates are computed, for the most part, on the basis of intelligence test scores, so that the estimates are only for that part of intelligence measured by the tests. An interesting recent finding by Eric Turkheimer and his colleagues is that heritability appears to be substantially higher in higher social classes than in lower social classes.

Whatever the heritability factor of IQ, an entirely separate issue is whether intelligence can be increased. The work by a New Zealand researcher, James Flynn, has shown that, in the middle and later twentieth century, scores on intelligence tests rose rather steadily throughout the world. The precise reason for the increase is unknown, although speculations include better education, better nutrition, and advances in technology.

Despite the general increase in scores, average IQs continue to vary across both countries and different socioeconomic groups. For example, many researchers have found a positive correlation between socioeconomic status and IQ, although they disagree over the reason for the relationship. Most psychologists agree that differences in educational opportunities play an important role, and some investigators believe that there is a hereditary basis

for the difference as well. However, there is simply no broad consensus on the issue of why the differences exist. Again, the differences are based on IQ, not broadly defined intelligence.

No matter how heritable intelligence is, some aspects of it are still malleable. Heritability of a trait is a separate issue from its malleability. For example, height is not only highly heritable, but also highly modifiable by nutrition and other environmental factors. Thus, with intervention, even a highly heritable trait can be modified. There is a growing body of evidence suggesting that aspects of intelligence also can be modified. Intelligence, in the view of many psychologists, is not merely a fixed trait. A program of training in intellectual skills can increase some aspects of a person's level of intelligence. No training program – no environmental condition of any sort – is likely to make a genius of someone with low measured intelligence. However, some gains are possible, and a number of programs have been developed for increasing intellectual skills. A main trend for psychologists working in the intelligence field has been to combine testing and training functions in order to enable people to optimize their intelligence. Such work needs to take into account technological advances because the very existence of technology changes the nature of intelligence. Today, skills in computer literacy and database management, which were needed only rarely 50 years ago, have become important. Therefore, the nature of intelligence is, to some extent, a moving target, one that researchers in the field must constantly track.

See also: Cognition: Overview and Recent Trends; Piaget: Recent Work; Vygotsky and Recent Developments.

Further Reading

- Binet, A. and Simon, T. (1916). *The Development of Intelligence in Children*. Baltimore, MD: Williams and Wilkins.
- Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. New York: Cambridge University Press.
- Cattell, R. B. (1971). *Abilities: Their Structure, Growth, and Action*. Boston, MA: Houghton Mifflin.
- Ceci, S. J. (1996). *On Intelligence*. Cambridge, MA: Harvard University Press.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist* **12**, 671–684.
- Deary, I. J. (1999). Intelligence and visual and auditory information processing. In Ackerman, P. L., Kyllonen, P. C., and Roberts, R. D. (eds.) *Learning and Individual Differences: Process, Trait, and Content Determinants*, pp 111–130. Mahwah, NJ: Erlbaum.
- Deary, I. J. and Caryl, P. G. (1997). Neuroscience and human intelligence differences. *Trends in Neurosciences* **20**, 365–371.
- Engle, R. W., Kane, M. J., and Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In Miyake, A. and Shah, P. (eds.) *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, pp 102–134. Cambridge: Cambridge University Press.

- Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.
- Gardner (1999). *Intelligence Reframed: Multiple Intelligences for the 21st Century*. New York: Basic books.
- Guilford, J. P. (1950). Creativity. *American Psychologist* 5(9), 444–454.
- Haier, R. J., Siegel, B., Tang, C., Abel, L., and Buchsbaum, M. S. (1992). Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence* 16(3–4), 415–426.
- Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In Sternberg, R. J. (ed.) *The Encyclopedia of Human Intelligence*, vol. 1, pp 443–451. New York: Macmillan.
- Hunt, E. B., Lunneberg, C., and Lewis, J. (1975). What does it mean to be high verbal? *Cognitive Psychology* 7, 194–227.
- Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability*. Westport, CT: Praeger/Greenwood.
- Just, M. A. and Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99, 122–149.
- Kyllonen, P. C. (1996). Is working memory capacity Spearman's g? In Dennis, I. and Tapsfield, P. (eds.) *Human Abilities: Their Nature and Measurement*, pp 49–75. Mahwah, NJ: Erlbaum.
- Mayer, J. D. and Salovey, P. (1993). The intelligence of emotional intelligence. *Intelligence* 197, 433–442.
- Neisser, U. (ed.) (1998). *The Rising Curve*. Washington, DC: American Psychological Association.
- Newell, A. and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Piaget, J. (1972). *The Psychology of Intelligence*. Totowa, NJ: Littlefield-Adams.
- Plomin, R. (1997). Identifying genes for cognitive abilities and disabilities. In Sternberg, R. J. and Grigorenko, E. L. (eds.) *Intelligence, Heredity, and Environment*, pp 89–104. New York: Cambridge University Press.
- Preiss, D. D. and Sternberg, R. J. (2006). In everyday life, tool-free abilities do not exist. *Educational Technology* 46(2), 43–47.
- Spearman, C. (1904). 'General intelligence,' objectively determined and measured. *American Journal of Psychology* 15(2), 201–293.
- Spearman, C. (1927). *The Abilities of Man*. New York: Macmillan.
- Sternberg, R. J. (1990). *Metaphors of Mind: Conceptions of the Nature of Intelligence*. New York: Cambridge University Press.
- Sternberg, R. J. (1997). *Successful Intelligence*. New York: Plume.
- Sternberg, R. J. (ed.) (2000). *Handbook of Intelligence*. New York: Cambridge University Press.
- Sternberg, R. J. (2003). *Wisdom, Intelligence, and Creativity Synthesized*. New York: Cambridge University Press.
- Sternberg, R. J. (2004). Culture and intelligence. *American Psychologist* 59(5), 325–338.
- Sternberg, R. J., Conway, B. E., Ketron, J. L., and Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology* 41, 37–55.
- Sternberg, R. J., Grigorenko, E. L., and Kidd, K. K. (2005). Intelligence, race, and genetics. *American Psychologist* 60(1), 46–59.
- Terman, L. M. and Merrill, M. A. (1937). *Measuring Intelligence*. Boston, MA: Houghton Mifflin.
- Thurstone, L. L. (1938). *Primary Mental Abilities*. Chicago, IL: University of Chicago Press.
- Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., and Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science* 14, 623–628.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Learning from Multiple Information Sources

H I Strømso and I Bråten, University of Oslo, Oslo, Norway

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Recent research on reading and reading instruction has led to significant improvements in the understanding of word decoding and the teaching of primary-grade readers. This research is based on a single-text paradigm where students are invited to read texts as isolated pieces of information. Skills involved in single-text reading, such as decoding and inference making, are vital in all kinds of text-comprehension activities. However, when readers use multiple information sources, skilled reading involves even more complex processes. As readers in today's society are continually bombarded with information from innumerable, often conflicting, information sources, it is essential to understand how individuals handle challenging integrative reading tasks involving very complex cognitive processes.

In this article, multiple information sources refer to different text-based sources on the same topic. For instance, a student may read a newspaper article about immigration in Europe. Next day, the student reads another article about immigration and integration in a magazine, and later the same day a member of a right-wing party offers the student a flyer about the number of criminal immigrants in the country. The three different texts are focusing on different aspects of immigration, and some of the information may be closely related, even overlapping, whereas other parts of the texts may be partly conflicting, supplementary, or not related at all. The student may leave each text as is or, alternatively, take on the challenging task of trying to construct a coherent meaning from the three texts. Learning from multiple information sources is a common task in modern societies, not only in educational settings, but also at home and at most workplaces.

There is no unified theory regarding how students deal with manifold information sources. Several researchers prefer to use the term *intertextuality* when studying how people deal with different information sources. This term is traditionally associated with domains such as literary criticism, linguistics, and biblical research. In educational research, *intertextuality* has been linked to reading, writing, and literacy practices in classrooms (Shuart-Faris and Bloome, 2004). The focus is on the construction of meaning through texts, based on the assumption that texts are always related to other texts. Studies on *intertextuality* may include text materials such as transcripts of dialogs, student reports, textbooks, and several other forms of

written material. However, text may also be defined more broadly, as a product of textualizing, meaning that all sorts of phenomena can be made parts of a language system and thereby textualized (Bloome and Egan-Robertson, 1993). Generally speaking, this kind of research is interested in social and cultural processes in text-based meaning making.

Another line of research pays more attention to the interaction between readers and texts. This article primarily discusses work related to this last-mentioned approach, where multiple information sources mainly refer to different kinds of written nonfictional texts, both paper based and digital. Such texts may also include graphical representations, maps, and pictures, but this article does not discuss different kinds of multimedia sources. So far, much of the research on learning from multiple sources has been conducted within the domain of history, where the interpretation of different kinds of sources is an essential activity. Still, research on the reading of multiple information sources also spans other domains, for example, science, social science, language arts, and law. Most studies include adult participants, partly because the reading of multiple sources has been considered such a complex activity that younger students do not benefit from it, even though findings are not conclusive with respect to this issue.

Integrating Information from Multiple Sources

The reading of multiple information sources involves the same processes as single-text reading, but, in addition, the reader must integrate information from various sources and attend to specific features of each source. Part of the integration process seems to be more or less automatic, as when information in one source overlaps with information in another source (Kurby *et al.*, 2005). Other parts seem to require more strategic effort on the part of the reader. In an already classic study in this area, Wineburg (1991) found that expert historians demonstrated three strategies not shared by novices: *corroboration*, *sourcing*, and *contextualization*. *Corroboration* involves comparing details located in one source with those located in another to decide whether they are trustworthy or plausible. In this way, the reader may judge if the information is complete and in agreement with other sources. If not, the reader may question the credibility of the source. Wineburg (1991) found that experts used this strategy

frequently, whereas high school seniors tended to consider each source in isolation. Sourcing refers to how expert historians evaluate historical documents by first looking up a document's source information. In Wineburg's study, source information with respect to type of document, author, and date of publication, was included at the bottom of each document. Still, the experts focused on this information prior to reading the document, whereas the high school students generally paid little attention to source information. The third strategy, contextualization, refers to how experts use prior knowledge to reconstruct events described in historical documents in a broad context of time and space. Weinberg's seminal study demonstrated that experts benefit from having superior prior knowledge about text content. In addition, experts take advantage of more specialized knowledge about information sources and how to use them.

Later research has suggested that the use of strategies such as corroboration and sourcing is more a question of documentary expertise than of domain expertise. Rouet *et al.* (1997) found that graduate students familiar with different kinds of primary and secondary sources in their own content area (psychology) were able to benefit from this when working with documents in another content area (history). Thus, higher education may seem to promote skills in handling multiple sources, at least the kind of skills Wineburg (1991) identified. In contrast, students at lower levels do not spontaneously demonstrate such skills (Stahl *et al.*, 1996). However, there may be limitations to the kind of general documentary expertise noted by Rouet *et al.* (1997), as history and psychology differ less than for example history and physics. Readers' strategic processing of multiple sources probably adapts to disciplinary constraints, at least in part, but thus far, no comparative studies exist to confirm the impact of such constraints.

Still, results from several studies outside the domain of history indicate that at least adult college readers may actively integrate information from multiple sources and also benefit from intertextual reading. For example, Mannes (1994) studied undergraduate psychology students reading a text and one of two versions of an outline about the use of microbes in industry. One version of the outline represented the same perspective as the text, whereas the other version of the outline, although containing the same information, was written from a different perspective. Only the last version represented a distinctly different source in relation to the text. Results showed that students reading the text and the different-perspective outline integrated the two texts better than students reading the text and the same-perspective outline. The first group of students elaborated more on the content and produced more links between the two sources, thus generating a richer and more flexible representation of the content of the two documents.

Research within the domains of law (Strømsø *et al.*, 2003) and social science (Bråten and Strømsø, 2006b) confirms the importance of elaboration strategies when studying multiple information sources. That research also demonstrates the importance of monitoring strategies when students struggle to integrate the content of different sources and need to assess their understanding of a text in relation to other available sources. Moreover, use of such deep-level comprehension strategies may result in a better understanding of a topic when students read multiple sources than when they read a single text (e.g., a textbook chapter) with the same content.

In brief, the integration of multiple information sources may both require and generate active strategic reading. There is also evidence that such integration may result in a richer and more flexible mental representation of the text content than the reading of one single source.

Mental Representation of Multiple Information Sources

To construct an adequate mental representation of the content while reading a single text, a reader must select important information, generate coherence where necessary, and link information in the text to relevant prior knowledge. In one of the most influential theories of text comprehension, Kintsch (1988) argues that two central processes, construction and integration, produce a mental representation – or model – of the situation described in the text. When reading two or more related texts, a reader can produce two or more situation models which are completely separate from each other. Alternatively, a reader can construct a single highly integrated situation model based on the texts, or the result can fall somewhere in between the separate models and the integrated model. If there is much content overlap between texts, the likelihood of producing an integrated situation model increases, while contradictory texts, conversely, increase the likelihood of producing separate models. When one text explicitly refers to another text, this is also likely to promote the construction of an integrated situation model. In most situations where students read multiple sources about the same topic, the most adequate form of comprehension is neither a fully integrated representation of the texts, nor separate and isolated representations of each text. It is important, however, that students understand not only the content of each single source but also the relationships between the content of different sources and are able to judge the trustworthiness of the sources they study. This way of representing multiple sources has been called a documents' model (Perfetti *et al.*, 1999).

One component of a documents' model is the reader's representation of relationships between the content of multiple sources. In the introductory example with three

texts about immigration, the newspaper article presents descriptive data on immigration in different countries and describes trends during the last 20 years. The magazine contains a feature article about riots among young immigrants in Paris, and this article argues that lack of integration and unemployment are main explanations for public disturbance. In the right-wing flyer, it is argued that immigrants outnumber natives in statistics on crime in all European countries, and that this is due to genetic differences between Europeans and immigrants. Based on such information about the texts, it is possible to consider relationships between them. The newspaper article seems to be supplementary to the two other texts, provided it does not present any explanations for riots and crimes. The magazine article does not explicitly contradict the flyer, but the explanations it presents seem to be in conflict with the genetic explanation of the flyer. Thus, relationships between the three texts can be described as both supplementary and conflicting, and the reader needs to detect and remember the specific relationships between the texts to construct an adequate mental representation of the texts – a documents' model.

Different kinds of source information constitute a second component of a documents' model. In trying to understand the content of the texts and the relationships among them, the reader will benefit from being aware of properties of each text, such as who wrote it, possible motives for writing it, its cultural context, and genre. If, for example, the newspaper article was written by a journalist in a conservative newspaper usually arguing for severe restrictions on immigration, the reader might ask whether the descriptive data were selected to support a certain point of view. The reader's interpretation of the article might also be affected were the reader able to place it in a certain context, for example, in the context of a discussion about government expenses related to social security. Such a context would frame the article differently than the context of a discussion about the importance of immigration for economic growth. Finally, knowledge of genre could influence the reader's understanding of the content. In the example, it would be helpful to know the difference between a newspaper article and a research report. Likewise, knowledge about properties of a feature article and a flyer would probably affect the reader's construction of a mental representation of the texts, and also become part of such a representation. When reading multiple texts, source information forms an important part of the documents' model that good readers construct.

Thus, a good reader will probably construct a mental representation of the texts which includes the content of the texts, the relationships among the texts, and relevant source information tagged to the different information units. In this way, the reader not only constructs a representation of the central information in the texts, but also

keeps in mind where the different information units come from and how they are related.

Learning from Multiple Information Sources

Although textbooks still play an important role in most schools, as well as in higher education, students are continually exposed to multiple information sources about the same topic. Sometimes, a multitude of sources seem to benefit learning. From a pedagogical point of view, it is important to identify factors that may promote or constrain learning from multiple information sources. Such factors could be related to both individual-difference variables and to aspects of the task.

The Importance of Expertise

Given that the reading of multiple texts about a topic is a complex task, it is not surprising that expertise matters (Wineburg, 1991). First, good reading skills seem to be prerequisite for dealing effectively with multiple information sources (Britt *et al.*, 1999; VanSledright, 2002). Students who have difficulties with word decoding or reading comprehension need to concentrate their efforts on understanding each single text in isolation and they have less cognitive capacity to compare and integrate different sources. Still, there are indications that reading difficulties do not necessarily constrain students at higher levels, who may be able to compensate for such difficulties by employing deep-level strategies and activate relevant content knowledge (Strømsø *et al.*, 2003).

Second, experience with complex text materials and proficiency in the content area have an impact. Stahl *et al.* (1996) confronted high school students with multiple documents about an important incident taking place during the Vietnam War. It was found that students at this level did not benefit from reading more than two texts, most likely due to their limited knowledge about different kinds of historical documents. VanSledright and Kelly (1998) found that fifth graders reading multiple sources in history tended to accumulate as much information as possible without attempting to integrate information across sources or judge the trustworthiness of each source. Still, some students showed emerging awareness of the diverse views represented by different sources. These studies indicate that young students have the potential to develop competence in dealing with multiple sources but do not spontaneously come to grips with them. The dominance of the textbook as the single authoritative source in both primary and secondary education may explain why so many students struggle to comprehend and reason from manifold sources. Most students are

simply not trained to read and benefit from the reading of such complex text materials.

Personal Epistemology

Personal epistemology concerns the beliefs individuals hold about the nature of knowledge and the process of knowing. For example, some students may view scientific knowledge as a collection of proven facts transmitted from authorities, whereas others may believe scientific knowledge to be theoretical, problematic, and a result of ongoing inquiry. Lately, some studies have linked personal epistemology to the comprehension of multiple sources, with this body of research generally indicating that more naive epistemic beliefs (e.g., that knowledge is certain and simple) are related to poorer comprehension performance (for review, see Bråten, 2008).

For example, Jacobson and Spiro (1995) found that only students who preferred working with complex knowledge in multiple ways and valued active learner construction of knowledge were able to profit from the reading of multiple texts presented in a hypertext environment. Moreover, Rukavina and Daneman (1996) found that students who believed that knowledge is simple and consists of isolated facts had problems integrating information from two separate texts representing competing theories about scientific problems. Finally, Bråten and Strømsø (2006a) found that the reading of multiple texts facilitated deeper understanding only among students with more sophisticated epistemic beliefs.

Thus, research suggests that the understanding of multiple sources seems to require relatively sophisticated epistemic thinking, where knowledge is seen as constructed through both rational processes and the melding of information from different perspectives. Apparently, multiple sources constrain rather than promote comprehension for readers who hold the naive beliefs that knowledge consists of unchanging, isolated bits of information handed down by authority rather than tentative, interrelated concepts constructed by the reader.

The Reading Task

The nature of the reading task is also likely to affect how students understand multiple sources. A student looking for specific pieces of information does not read texts in the same way as a reader whose purpose is to construct a coherent and integrated understanding of the content of the texts. The knowledge of how reading tasks or purposes affect students' reading of multiple sources is still limited. However, some studies have demonstrated that students alter their text processing in accordance with differences in study-related reading tasks. For example, Strømsø *et al.* (2003) found that when law students reading self-selected study texts changed their understanding of the reading

task from reading to keep up with lectures to reviewing for the examination, they also adjusted their text processing to fit the new reading task. Results from other studies indicate that not only text processing but also level of understanding is affected by the task.

For example, Wiley and Voss (1999) had undergraduates read multiple documents about a historical event under the different task conditions of writing an argumentative essay, a narrative, a summary, or an explanation. They found that students who read in order to write arguments gained more integrated and deeper understandings of the topic than did students who read the same documents in the other three task contexts. The importance of the reading task for learning from multiple sources was also examined by Cerdan and Vidal-Abarca (2008), who showed that relatively broad questions may facilitate text integration and prevent readers from concentrating on isolated bits of information. Very narrow questions, on the other hand, may occasion that readers search for, find, memorize, and reproduce isolated bits of information, which may be sufficient for answering the narrow questions correctly but constrain rather than promote a deeper understanding of the texts' content.

Multiple Information Sources at the World Wide Web

The factors already described as important when learning from multiple information sources are also important when the sources are located at the Web. However, the myriad of informational Internet resources represents additional challenges to the reader. For instance, research on the readability of Web pages indicates that the texts located there are frequently quite difficult to read (Kim and Kamil, 2003). In addition, readers often have to read such texts framed by links to additional information, advertisements, or other potentially disturbing features. This article, however, focuses specifically on challenges regarding sourcing and integration.

First, students' sourcing skills are even more important as the majority of Web pages do not undergo the quality control most paper-based publications do. The judgment of texts' trustworthiness and quality is to a much greater extent left to the students. For instance, a primary school student looking for information about Martin Luther King at the Web may hit the apparently reliable website <http://www.martinlutherking.org/> containing plenty of information to include in an upcoming report. Only when comparing this information with other sources of information, the student may detect contradictions between sources and be able to evaluate the trustworthiness of each source. A competent reader includes source information in reflections on which source to trust. However, identifying the source may be difficult when reading

Web documents (Britt and Gabrys, 2000). In the above example, the attentive reader will notice that the document is hosted by Stormfront, and more research will reveal that this means Stormfront White Nationalist Community. Printed documents tend to carry very standard information about the author and the type of document, whereas Web documents in many cases do not.

Second, students face another challenge when they try to integrate information from different Web sources and construct a coherent understanding of a topic. This is related to the fact that Web texts are often linked to numerous other texts, and it is the student's own responsibility to choose what links to pursue. Links may be more or less relevant to the topic, and availability is often more important than relevance when links are embedded in a text. In such cases, links may interfere with students' efforts to construct a coherent representation of the content. On the other hand, links may also represent opportunities to gain a deeper understanding of a topic, provided the reader is able to compare and integrate information from the retrieved sources. In pursuing links across several texts, such integration may be quite demanding, however. Results from several studies indicate that students' prior knowledge about the topic is vital for their ability to cope with and benefit from using links in Web documents (DeStefano and LeFevre, 2007).

Educational Implications

The ability to integrate information from multiple information sources needs to be trained. Given the overflowing information in present-day knowledge societies, many students have received surprisingly little instruction in how to handle such sources. Consequentially, students ranging from primary to undergraduate level often strive with this task. However, results from several studies indicate that training may improve students' skills in handling various sources on the same topic.

For example, VanSledright (2002) attempted to teach fifth graders how to practice history by introducing both primary and secondary sources on a historical event and instructing the students to become historical detectives. They were explicitly taught that such detectives operated by identifying the nature of the sources and the different perspectives in the texts. They were also taught to check and corroborate evidence before drawing any conclusions. The results indicated that young students can benefit from being exposed to multiple texts, including primary sources, but that explicit training in dealing with such sources is necessary. Moreover, Goldman and Bloome (2004) studied academic discourse of literary analysis in seventh and eighth grade. At both levels, processes of construction and integration were stimulated and externalized in the classrooms. Students were instructed to

compare texts and look for both differences and similarities. When discussing the texts, students were instructed to back up explanations and arguments by always referring to the source. While these studies focus on efforts to expose young students to multiple sources and to externalize construction and integration processes in the classroom, other researchers have introduced computer-based training programs.

Britt and Aglinskas (2002) developed the Sourcer's Apprentice, a computer application teaching the skill of sourcing while reading multiple documents to research a historical controversy. In this approach, students are presented with a set of different documents about each controversy, varying in type from textbook excerpts to primary documents, which they read in a self-selected order. While reading each document, they fill in a note card with information about the author, document, and content. When the student thinks that he or she has studied the documents enough, the student answers a series of questions about the content and sources of the documents. Finally, the student writes an essay about the controversy, with only the note cards available during writing. Britt and Aglinskas (2002) showed that students who used the Sourcer's Apprentice in place of an integrated single-text presentation of the same content wrote essays on the controversy that were more integrated, cited more sources, and referenced more information from primary and secondary sources than did the comparison group. Later, Britt and colleagues supplied the Sourcer's Apprentice with a feedback mechanism that provides automatic and immediate feedback regarding skills in attending to and citing sources, as well as in integrating content from different sources, during the writing process (Britt *et al.*, 2004). The resulting Sourcer's Apprentice intelligent feedback mechanism has been found to help students write better essays after reading multiple documents.

Conclusion

The majority of research on text comprehension has thus far been restricted to a single-text perspective. However, being a proficient reader demands not only basic skills in decoding and the comprehension of single texts, but also skills in integrating and judging the trustworthiness of multiple sources. Results from several studies indicate that such skills partly depend on domain expertise, and it has been more or less assumed that students in primary school are not sufficiently mature and competent to handle such complex reading tasks. However, other studies indicate that students as early as fifth grade can be trained to cope with multiple information sources and also that they can benefit from reading them.

Further educational research on students' processing of multiple sources is important for two main reasons. First, students are confronted with multiple information

sources on a daily basis and therefore need skills to handle this situation. Second, the reading of multiple information sources can lead to richer and more flexible representations of text content, with such representations being less tied to any specific text and more accessible under a variety of circumstances. Research in this field has been dominated by studies on how students deal with different information sources in history, and there is a need for more studies on reading of multiple texts in other domains. As electronic learning environments become increasingly common, the study of learning from multiple texts should also be linked more closely to research on learning with multimedia technologies.

See also: Personal Epistemology in Education: Concepts, Issues, and Implications; Problem Solving and Human Expertise; Reading Comprehension: Reading For Learning.

Bibliography

- Bloome, D. and Egan-Robertson, A. (1993). The social construction of intertextuality in classroom reading and writing lessons. *Reading Research Quarterly* **28**, 305–333.
- Britt, M. A. and Aglinskias, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction* **20**, 485–522.
- Britt, M. A. and Gabrys, G. L. (2000). Teaching advanced literacy skills for the World Wide Web. In Wolfe, C. R. (ed.) *Learning and Teaching on the World Wide Web*, pp 73–90. San Diego, CA: Academic Press.
- Britt, M. A., Perfetti, C. A., Sandak, R., and Rouet, J. -F. (1999). Content integration and source separation in learning from multiple texts. In Goldman, S. R., Graesser, A. C., and van den Broek, P. (eds.) *Narrative, Comprehension, Causality, and Coherence: Essays in Honor of Tom Trabasso*, pp 209–233. Mahwah, NJ: Erlbaum.
- Britt, M. A., Wiemer-Hastings, P., Larson, A. A., and Perfetti, C. A. (2004). Using intelligent feedback to improve sourcing and integration in students' essays. *International Journal of Artificial Intelligence in Education* **14**, 359–374.
- Bråten, I. (2008). Personal epistemology, understanding of multiple texts, and learning within Internet technologies. In Khine, M.S. (ed.) *Knowing, Knowledge, and Beliefs: Epistemological Studies across Diverse Cultures*, pp 351–376. Amsterdam: Springer.
- Bråten, I. and Strømso, H. I. (2006a). Effects of personal epistemology on the understanding of multiple texts. *Reading Psychology* **27**, 457–484.
- Bråten, I. and Strømso, H. I. (2006b). Constructing meaning from multiple information sources as a function of personal epistemology: The role of text-processing strategies. *Information Design Journal* **14**, 56–67.
- Cerdan, R. and Vidal-Abarca, E. (2008) The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology* **100**, 209–222.
- DeStefano, D. and LeFevre, J. A. (2007) Cognitive load in hypertext reading: A review. *Computers in Human Behavior* **23**, 1616–1641.
- Goldman, S. R. and Bloome, D. M. (2004). Learning to construct and integrate. In Healy, A. F. (ed.) *Experimental Cognitive Psychology and Its Applications: Festschrift in Honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer*, pp 169–182. Washington, DC: American Psychological Association.
- Jacobson, M. J. and Spiro, R. J. (1995). Hypertext learning environments, cognitive flexibility, and the transfer of complex knowledge: An empirical investigation. *Journal of Educational Computing Research* **12**, 301–333.
- Kim, H. S. and Kamil, M. L. (2003). Electronic and multimedia documents. In Sweet, A. P. and Snow, C. E. (eds.) *Rethinking Reading Comprehension*, pp 166–175. New York: Guilford Press.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review* **95**, 163–182.
- Kurby, C. A., Britt, M. A., and Magliano, J. P. (2005). The role of top–down and bottom–up processes in between–text integration. *Reading Psychology* **26**, 335–362.
- Mannes, S. (1994). Strategic processing of text. *Journal of Educational Psychology* **86**, 577–588.
- Perfetti, C. A., Rouet, J. -F., and Britt, M. A. (1999). Toward a theory of documents representation. In van Oostendorp, H. and Goldman, S. R. (eds.) *The Construction of Mental Representations during Reading*, pp 99–122. Mahwah, NJ: Erlbaum.
- Rouet, J. -F., Favart, M., Britt, M. A., and Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction* **15**, 85–106.
- Rukavina, I. and Daneman, M. (1996). Integration and its effect on acquiring knowledge about competing scientific theories from text. *Journal of Educational Psychology* **88**, 272–287.
- Shuart-Faris, N. and Bloome, D. (eds.) (2004). *Uses of Intertextuality in Classroom and Educational Research*. Greenwich, CT: Information Age Publishing.
- Stahl, S. A., Hynd, C. R., Britton, B. K., McNish, M. M., and Bosquet, D. (1996). What happens when students read multiple source documents in history? *Reading Research Quarterly* **31**, 430–456.
- Strømso, H. I., Bråten, I., and Samuelstuen, M. S. (2003). Students' strategic use of multiple sources during expository text reading: A longitudinal think-aloud study. *Cognition and Instruction* **21**, 113–147.
- VanSledright, B. A. (2002). Fifth graders investigating history in the classroom: Results from a researcher–practitioner design experiment. *Elementary School Journal* **103**, 131–160.
- VanSledright, B. A. and Kelly, C. (1998). Reading American history: The influence of multiple sources on six fifth graders. *Elementary School Journal* **98**, 239–265.
- Wiley, J. and Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology* **91**, 301–311.
- Wineburg, S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology* **83**, 73–87.

Further Reading

- Goldman, S. R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In Shuart-Faris, N. and Bloome, D. (eds.) *Uses of Intertextuality in Classroom and Educational Research*, pp 317–351. Greenwich, CT: Information Age Publishing.
- Rouet, J. -F. (2006). *The Skills of Document Use: From Text Comprehension to Web-Based Learning*. Mahwah, NJ: Erlbaum.

Memory

S Magnussen and T Brennen, University of Oslo, Oslo, Norway

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Declarative memory – An essential property of declarative memory is that one is conscious of what one is remembering. Such memories are, by their very nature, verbal. For example, one can talk about one's experiences, and about new words that one has learned.

Episodic memory – Personal experiences and events are stored in episodic memory (see declarative memory).

False memory – A memory that appears real to the rememberer, but is not based on veridical experience is referred to as a false memory. They can be induced in the laboratory, but they are also prevalent in everyday situations, sometimes with drastic legal and interpersonal consequences.

Forgetting – It is the inability to recall an episode or a skill that was previously learned. In forgetting of declarative memory (*quod vide*), it is suspected that many failures to recall are in fact due to a failure to access a still-existent memory trace.

Memory systems – In contrast to previous schools of thought in psychology, where memory was considered to be unitary, in cognitive psychology one considers memory to be a multifaceted set of modules. There is a host of evidence to suggest that we possess many different types of memory.

Nondeclarative memory – We are not aware of everything that affects our behavior. These nonconscious effects of previous experience are evidence of nondeclarative memory. Also, our memory for motor skills, such as walking, typing, and skiing, is nondeclarative because motor skills are essentially nonverbal.

Perceptual representation system – Conceived of as the interface between perception and memory, it is a basic memory system storing perceptual information that enables us to perceive the world as meaningful.

Priming – Things that you have just seen, heard or thought about will facilitate thoughts about related concepts. This expression of memory is referred to as priming.

Semantic memory – Believed to be stored separately from personal experiences or episodes, semantic memory refers to a person's general knowledge of the world (see episodic memory).

Working memory – A coordinated set of short-term memory mechanisms that allow us to consciously manipulate what we are thinking about at any given moment.

Memory is a central part of the brain's attempt to make sense of experience, and to tell coherent stories about it. These tales are all we have of our past, so they are potent determinants of how we view ourselves and what we do. Yet our stories are built from many ingredients. Snippets of what actually happened, thoughts about what might have happened, and beliefs that guide us as we attempt to remember. Our memories are the powerful but fragile products of what we recall from the past, believe about the present, and imagine about the future (Schacter, 1996: 308).

A few years ago, a Norwegian newspaper (*Dagbladet, Magasinet*, 27 March, 2004) reported the story of Dodo, a young man of Asian origin who, in January 2003, woke up on the freezing ground in a small village in Switzerland with his well-equipped rucksack nearby, stuffed with expensive clothes and a money belt containing US\$5000, but no identity papers or tickets and with absolutely no personal memory. Dodo wandered around in Europe for some weeks, and, for reasons he could not explain, somehow managed to travel to Oslo, Norway. His memory loss of the time before he woke up in Switzerland is massive, he has no idea who he is, and he did not recognize his own face in the mirror. He had even lost his native language – he spoke heavily accented English but not any Asian language; the only thing he knew about himself was that he smoked Camel and liked pop music.

Dodo suffers from the condition of retrograde amnesia. That is, he has lost his memory in the sense people usually use the term memory, the recollection of private experiences and facts we have learned about the world. But Dodo remembered many things, he remembered what cigarettes were for, the workings of photographic equipment, he understood the value of money, was able to buy food, and mastered the skill of traveling by public transport. So he could make use many of the things he had learned. This type of selective memory impairment is a main argument cited by memory researchers in support of the idea that memory rather than being a single cognitive process or system, is a collective term for a family of neurocognitive systems that store information in different formats (Schacter *et al.*, 2000; Tulving, 2002).

Varieties of Human Memory

Modern taxonomies distinguish between several forms of memory, or memory systems. An important distinction is drawn between declarative memory, which refers to the conscious recollection of facts and personal experiences, and nondeclarative memory, which refers to behavior changes resulting from previous experiences that may or may not be accompanied by conscious recollection (see **Figure 1**).

The story of Dodo illustrates well the central role of declarative memory in human life. This young man had lost not only his personal past – his autobiographical memories – but also large parts of his general knowledge of the world and even his ability to speak his native language. Thus, the systems or forms of memory that we term episodic and semantic memory are heavily affected. Episodic memory is assigned a special role in human life: it is unique in the sense that memories are associated with a place and a time, and is, according to Tulving (2002), the only known example of a process where the arrow of time is turned back and the past can be re-experienced, a feat probably unique to humans. Without episodic memory, the mental representation that psychologists call the self – the organization of personal memories in a historical context – is lost. Obviously, this is what had happened to Dodo; however, his nondeclarative memory seemed largely intact.

In nondeclarative memory, the effects of previous experiences and exercises manifest themselves directly in behavior; the individual learning sessions may be vaguely remembered or be completely inaccessible to conscious recollection. Procedural memory is responsible for the maintenance of the cognitive and motor skills that we have acquired throughout life, from knowing how to eat with a knife and fork to the mastery of swimming, cycling, or driving a car, as well as the advanced skills of playing billiards or playing a saxophone. Conditioning is a basic memory system that in humans is particularly important in tying emotional reactions to external stimuli

or situations. For example, in phobic reactions, anxiety is triggered by the phobic stimuli but the person is typically not aware of the learning episodes in which the connection between the emotion and stimulus was established. Similarly, a piece of music, or a specific odor may evoke romantic feelings without an accompanying experienced-memory episode. Perceptual learning, or the perceptual representation system (PRS), enables us to perceive the world consisting of meaningful entities, because in order to produce a perceptual experience, online sensory signals must join stored representations, and this linking is part of the perceptual process itself. To see is to recognize, or to realize that you do not recognize. Tulving and Schacter (1990) also identified priming as an expression of non-declarative memory, coupled to the PRS. In priming, the person's performance on a specific task is facilitated (or inhibited) by the previous presentation of related information. For example, perceptual identification of a portrait as representing Hillary Clinton is speeded by the prior presentation of a picture of Bill Clinton, compared to the prior presentation of a picture of, say, the Pope.

An important feature of the memory systems concept is the idea that the various systems store information in different formats, and the information stored in one format is not directly translatable into other formats. This implies that the information stored in one system is not immediately accessible to other systems. However, assuming that the memory systems operate independently and in parallel, most experiences would be recorded and stored in parallel in different formats, and the memory performance assisted by several memory systems working in concert (Tulving, 2002).

Another basic distinction in memory theory is the division between a short-term memory mechanism that is the seat of consciousness and active processing, and is able to store limited quantities of information for a limited period of time, and a long-term memory mechanism that stores unlimited amounts of information for unlimited periods of time (Atkinson and Shiffrin, 1968). The short-term memory mechanism is currently associated with

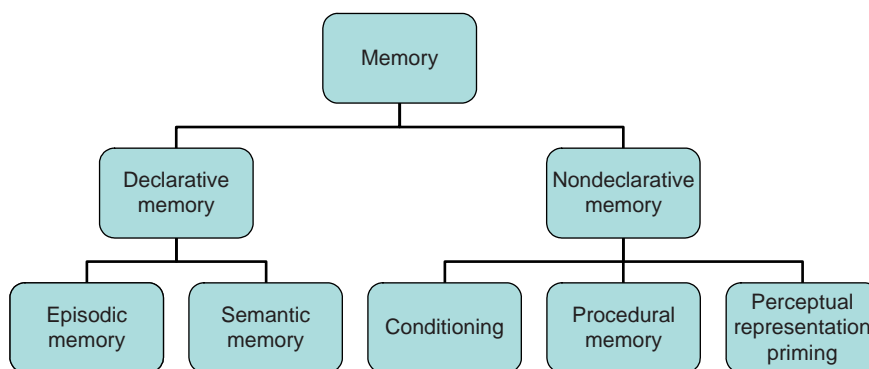


Figure 1 A common textbook taxonomy of human memory.

Baddeley's (1986, 2003) concept of working memory, a coordinated set of mechanisms that combine incoming information with information retrieved from long-term memory with the assistance of three support systems that actively manipulate verbal, visuo-spatial, and episodic information. The concept of working memory is closely associated with the concept of attention; only those pieces of information that are actively attended to, enter memory. Since the capacity of attention and working memory is limited, it follows that only part of the potentially available information confronting us at any time is registered, and much of the information that is registered is immediately forgotten. In a sense, most of the things we believe we have forgotten, never really entered a durable memory system.

In the context of education and everyday life, it is long-term memory that is the important aspect of memory.

Metaphors and Models of Memory – Strategies of Research

In the attempt to understand memory, scientists as well as laymen are forced to try to describe something that is not directly observable. We may be aware of some of the processes that take place during learning, and similarly of the processes that take place when we attempt to search for a piece of information in memory, but we have little access to the processes that mediate between encoding and retrieval. In order to make sense of the phenomena to which we have little access, both laymen and researchers tend to use metaphors and analogies, many of which are borrowed from the physical world (Magnussen *et al.*, 2007).

Koriat *et al.* (2000) pointed out that the study of memory following Ebbinghaus (1885) has been dominated by a storehouse metaphor, in which discrete items of information are stored and later retrieved, and which tends to evaluate memory in terms of the number of items that are retained (or lost). The storehouse metaphor invites an interest in the number of stored items. An effective store is one that contains many items, retains these items for long periods of time, and allows easy access to them. Memory then is evaluated in terms of its quantity – how much is retained, how much is lost.

Quantitative aspects of memory are frequently of great interest in everyday life, and research relevant to the classroom and academic achievement typically follows this strategy. However, an alternative approach to the study of memory, derived from the correspondence metaphor and that can be traced to the work of Bartlett (1932), evaluates memory performance in terms of the correspondence between the original event and the memory of it; the accuracy of the memory report. A high-quality memory is more faithful to the remembered event than an inferior memory, and contains fewer distortions and errors. Sometimes people remember events that

never happened. The assessment of memory correspondence should therefore start with the output – what the person reports – rather than with the input – what actually happened. An output-bound assessment reflects the accuracy of what is remembered – how much of what the person reports did in fact occur (Koriat *et al.*, 2000). In many real-life situations, such as in court, there is greater concern with the accuracy of the report than with the amount of information reported.

The Seven Sins of Memory

The modern study of the accuracy of episodic memory has convincingly disproved another popular metaphor, the idea of memory as a video recorder taping and replaying the original events. Episodic memory does not reproduce, it constructs, and the reconstruction of previous episodes is based on information from many sources with the assistance of many neural systems (Rubin, 2006). As some of the sources of information used in the construction of episodic memories are external to the original event, memory accuracy suffers. In a widely read book, Schacter (2001) identifies seven factors that he terms the “seven sins of memory,” which are at the origin of the tricks that episodic memory plays on us. The sins are: absent-mindedness; to be encoded into episodic memory, a piece of information or an event that must be attended to. When focal attention is diverted or attention is switched to auto-pilot, information may be missed, and the event is not properly remembered because the information was never registered in the first place. Transience refers to fact that memories fade. Some memories fade rapidly, others fade more slowly, depending on a number of factors, but no memory gets better over time. The general course of memory decay, first described by Ebbinghaus (1885), is the negatively accelerated function shown in **Figure 2(a)**, which represents the long-term forgetting of ordinary life events. However, as illustrated by **Figure 2(b)**, frequently studied or over-learned material shows very little decay, and some extraordinary life events, in particular dramatic and traumatic experiences, appear to be very resistant to fading and rather obey the sin of persistence. In fact, some memories we would like to forget, haunt us for the rest of our lives. And sometimes information that we know that is not forgotten – it is on the tip of the tongue – is unavailable to memory because of a failure of retrieval; this is the sin of blocking. Memories not only fade, they are subject to influences from several sources. The human mind is suggestible, our memories of events and facts are influenced by information from external sources, from what we read, and from what other people tell us. Memories may be biased in various ways. For example, our memories of the past depend on or are colored by the present, or as Schacter (2001) puts it, “the way we were depends upon the way we are.” The sin of misattribution refers to the confusions and

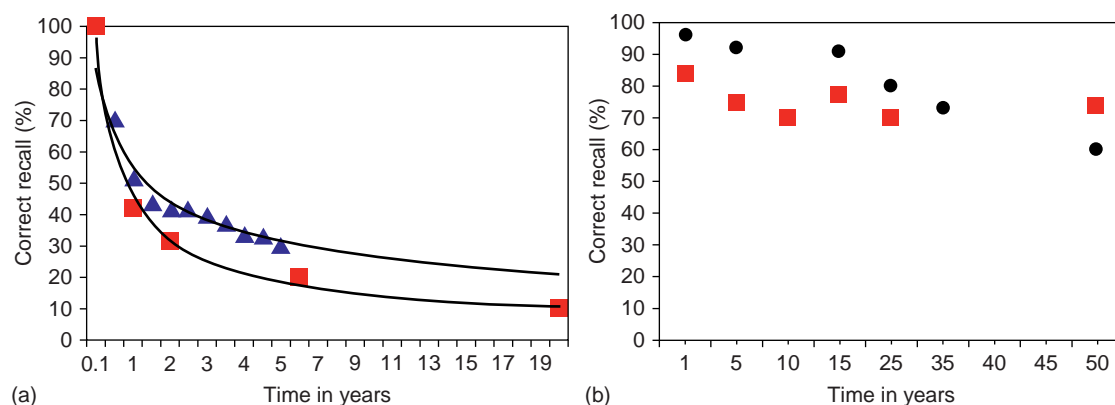


Figure 2 (a) Memory for personally experienced everyday events, (▲) and (■). (b) Very long-term memory for over-learned material, obtained by asking participants to match pictures of high-school classmates' faces to their names, (●), and testing participants' ability to translate unused Spanish vocabulary learned in school to English, (■). Despite the fact that we treasure our autobiographical memories and they play an integral role in our personal identity, we become incapable of bringing to mind the specifics of most of what we experience within a year or so of it happening. By contrast, multiple repeated exposures to something that is quite unimportant for one in later life and never practiced or recalled can allow access to memories decades later. This may be seen as paradoxical until one realizes that motivation to remember is not always as important as the amount of practice one has put in to learning something. (a) Adapted from Wagenaar, W. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology* **18**, 225–252 and White, R. (2002). Memory for events after twenty years. *Applied Cognitive Psychology* **16**, 603–612. (b) Adapted from Bahrick, H. P., Bahrick, P. O., and Wittlinger, R. P. (1975). Fifty years of memories for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General* **104**, 54–75 and Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General* **113**, 1–29.

mistakes we often make concerning the source of a particular piece of information, confusions of times, people, and places, and in the extreme case, the confusion of genuine memory and fantasy that gives rise to false memories.

Memory in Everyday Contexts

Ebbinghaus' (1885) pioneering studies investigated his own memory for nonsense syllables, and a century later, as cognitive psychology became the dominant school in psychology in the 1970s, the typical experiment on memory still clearly bore Ebbinghaus' imprint: university undergraduates sat in a darkened lab and learned long lists of words. A degree of discomfort with the potential lack of relevance of these experiments for understanding life outside the lab, led to a push for memory research to be more ecologically valid, and the study of applied memory has since grown into a very active research field. Drawing on the correspondence metaphor of memory, a key underlying question in this field has been, on how accurate is our memory, as it is used in everyday life. Another change has been in the types of material used: breaking out of the verbal learning tradition (learning lists of words), a concerted research effort over the past 20 years has, for instance, tried to shed light on our ability to remember more real-life stimuli, like people, and events we have experienced in our lives, a skill that Dodo apparently no longer possesses.

How do we remember people? A person represents an extremely complex and multifaceted biological stimulus:

while we most often recognize a person from their face, we can also judge who a person is from their name, their voice, their gait, besides many other potential routes. It turns out that the complexity of the stimulus is reflected in the diversity of memory representations we have for people. Not all types of person information are equally memorable either. In particular people's names pose a problem when it comes to recall. All readers will have experienced the tip-of-the-tongue phenomenon for a name, when one is unable to recall a name but feels that one might recall it at any time, and may be able to give partial information about the missing word. This phenomenon occurs particularly for people's names, and its frequency increases with age (Schwartz, 2002; Valentine *et al.*, 1996). On the other hand, our ability to distinguish between faces and to recognize that the one you are looking at now is one you have seen before is prodigious. This is perhaps not surprising from an evolutionary point of view: humans are social animals and there is a strong need to distinguish between family, friends, and enemies. On the other hand, more recent research in the applied memory tradition has cast doubt on our face-recognition skills in some practical situations. For instance, the use of credit cards with a picture of the holder's face is now widespread, but Kemp *et al.* (1997) have shown that, in fact, shop assistants are terrible at detecting that the face on the card does not match the person holding the card.

How much of our lives do we remember? Cognitive psychologists would agree with the sentiments of the novelist Milan Kundera who pointed out that in fact we

are only able to recall a vanishingly small proportion of the events in our lives. While we do of course remember many individual events from our lives, much of our memory appears to be based on generalized averages of similar events, so that while you may only be able to specifically recall a few of the times you have been to the Odeon cinema, you may know very well where you usually park, what sort of films you see, who you go with, and what snacks you buy. A systematic approach to answering the question was taken by Wagenaar (1986) who wrote down one incident from his life each day over a period of 4 years. When subsequently tested, he found that being cued by the who, what, or where of an event was more powerful a cue than when, and that single piece of information (who, what, or where) triggered recall of the incident only on about 50% of occasions for the first year and declined to 20% after 4 years. Multiple cues triggered over 60% of memories, even after 4 years, and even cases where he had apparently forgotten the incident completely could be recalled in cases when another person was central in the incident and could be asked to provide more cues. Thus, it is a case of the glass being half-full or half-empty: even events which were thought to be completely forgotten turned out to be retrievable, but on the other hand, normal cues failed to trigger a majority of memories. Results from Wagenaar's (1986) together with results from a similar study carried out across a 20-year interval by White (2002) are shown in **Figure 2(a)** and illustrate the dramatic decline of ordinary everyday memories. White (2002) tested his memory by having an assistant read event descriptions picked randomly, and he decided if he remembered the event vividly, vaguely, or not at all. At all testing times, most of the remembered events were vaguely remembered, and he encountered not a single case where an event forgotten at one time was remembered at a later time. Assuming the conditions of retrieval are the same at different testing points, memory does not improve over time.

False Autobiographical Memories

In the 1990s, the science of memory was the subject of a fierce debate in scientific journals, newspapers, and courts of law. A phenomenon that was first observed in the United States but that has since spread to other countries is that cases of alleged child-abuse were being brought to court by the victims, years and even decades later, because the victims had previously been unable to retrieve the memories. This was a powerful and deeply shocking claim that split psychology into those who thought of these memories as recovered, and those who thought of them as false, the so-called memory wars.

Since Bartlett's seminal research on how our memory for narrative introduces errors that fit in with the rest of our knowledge, memory research has known about the

malleability of our memories for events. Due to the memory wars, such memory illusions have been much studied over the past decade, showing that under certain circumstances, we can become convinced that, several years previously, one had been up in a hot air balloon, or had knocked over a punch bowl at a wedding when these events simply had not happened (Loftus, 2003). The key ingredients appear to be that one believes the authority telling one the information, and that one is given time to try to retrieve the information: over time, the false memories come to feel real to the rememberer. When it turns out that in many documented cases of recovered memories of child-abuse, the victim had been in therapy with a clinician who believed that symptoms of depression and anxiety in fact are the manifestation of early child-abuse, there is the very real possibility that the often grotesque and implausible accounts were genuinely believed by the rememberer but nonetheless entirely false, produced by the combination of an authority figure, the therapist, and several practice sessions in which to recover the memories. A number of documented cases show that false memories may be quite dramatic, as for example, remembering being witness to a murder that never happened (Goodman *et al.*, 2007). Modern research also fails to support the idea that early traumatic childhood experiences can be stored in a detailed fashion but repressed or blocked from consciousness, only later to be exhumed by therapeutic exercises. Rather, well-controlled studies show that such memories obey the sin of persistence (Goodman *et al.*, 2003).

Knowledge of the constructive nature of autobiographical memory is important for many professions, especially those dealing with psychiatric patients, children, and other groups with little power. The vulnerability of autobiographical or episodic memory is also the focus of another important area of applied memory research, the study of factors affecting the reliability of eyewitness testimony (Loftus, 2003; Wells *et al.*, 2006). In the interrogation room and in court, failures to remember are less damaging to the judicial process than are memory errors; absence of memory points nowhere, memory errors may point in the wrong direction.

Memory in the Classroom

One of the secrets to creating robust long-term memories, already shown by Ebbinghaus' experiments and known to centuries of teachers, is repetition or over-learning. Schoolchildren spend a lot of time in classrooms learning from teachers and books. The vast majority of children in the West spend upward of a decade of their lives in such a system. It is generally agreed that education is a good thing, but the exams that are used to test that the requisite knowledge has been acquired often take place within a few months of the teaching, which begs the question of the

fate of this knowledge in the long term. How much of the Spanish vocabulary that you learned in high school can you still recall? What about the fate of your university degree that you haven't used since your graduation exams? Seminal research by Harry Bahrick has shown that we are able to recall surprising amounts of knowledge even several decades after learning (see for example, Bahrick, 1984). Lots of material seems to become unavailable within the first year of learning. About 2 years after studying a topic, people are able to perform at around 50% of the level they were at when they stopped studying, but the surprising finding is that up to 50 years later, one performs at about the same level as one did 2 years after: very little forgetting appears to occur after an initial period. Similar results have been reported for mathematics, Spanish, and for cognitive psychology, so that this function appears to reflect general properties of the long-term semantic memory system, or permastore as Bahrick termed it (Figure 2(b)).

Another question is that of how best to study to ensure long-term learning. In the light of the prevailing orthodoxy in education, it might be surprising to some that a body of evidence has built up recently to suggest that the best way to study for a test is to test oneself. For instance, Roediger and Karpicke (2006) asked participants to read and learn a short essay. After all participants had read the passage through, one group was then given a test on what they had remembered, whereas the other group reread the passage. There was then a final memory test given, either 5 min, 2 days, or a week subsequent to the reading. After 5 min, those participants who had reread the passage performed better than the other group. However those tested after 2 days and those tested after a week had the opposite pattern: an extra opportunity to study the material led to worse memory compared to those who were tested also in the first phase (see Figure 3). The message from this study and many replications is clear: the very act of testing improves performance. It is to be hoped that educational systems will confront the challenge to incorporate the consequences of this finding into their classrooms: testing is not only a way of assessing students' performance, it also a way of facilitating their learning.

The Future for Memory Research

Memory research finds itself at the touching point between basic and applied research. On the one hand, it is becoming more integrated into neuroscience, with studies using cognitive tasks in conjunction with brain imaging methods, for instance. On the other hand, the emphasis on an understanding of memory in everyday contexts is strengthening. Up to now, the findings of memory research have been relatively slow to get out to practical situations. This is unfortunate because the focus

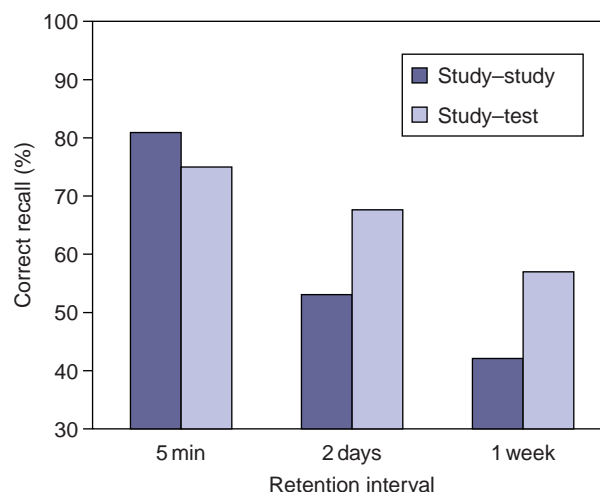


Figure 3 Having heard a story they were trying to learn, some participants were immediately tested on what they could recall, while others had a second chance to learn the story. Five minutes later, the second-chance group (the dark column) had better recall, but of potential importance to educationalists, the immediate-test group (the light column) did better on tests 2 and 7 days later. Adapted from Roediger, H. L. and Karpicke, J. D. (2006). Test-enhanced learning – taking memory tests improves long-term retention. *Psychological Science* 17, 249–255.

in recent decades on more applied sides of memory – how it is actually used in daily settings – has produced important results, which, at the risk of hyperbole, would appear important to be more widely known about. For all in the legal system, from the police to jury members, the studies showing that people can be induced to create false memories in themselves, should be required reading. Similarly, for the educational system, the generalized, positive gains obtained by the very act of testing someone's knowledge should become more widely known, and integrated into teaching systems.

See also: Cognition and Emotion; Cognition: Overview and Recent Trends; Problem Solving and Human Expertise; Remembering as Social Activity.

Bibliography

- Atkinson, R. C. and Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In Spence, K. W. and Spence, J. T. (eds.) *The Psychology of Learning and Motivation*, vol. 1, pp 89–195. New York: Academic Press.
- Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.
- Baddeley, A. D. (2003). Working memory. Looking back and looking forward. *Nature Neuroscience Reviews* 4, 829–839.
- Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General* 113, 1–29.
- Bartlett, F. (1932). *Remembering*. Cambridge: Cambridge University Press.

- Ebbinghaus, H. (1885). *Über das Gedächtnis*. Leipzig: Dunker.
- Goodman, G. S., Ghetti, S., Quas, J. A., et al. (2003). A prospective study of memory for child sexual abuse: New findings relevant to the repressed memory controversy. *Psychological Science* **14**, 113–118.
- Goodman, G. S., Magnussen, S., Andersson, J., et al. (2007). Memory illusions and false memories in real life. In Magnussen, S. and Helstrup, T. (eds.) *Everyday Memory*, pp 157–182. Hove: Psychology Press.
- Kemp, R., Towell, N., and Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology* **11**, 211–222.
- Koriat, A., Goldsmith, M., and Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology* **51**, 481–537.
- Loftus, E. (2003). Our changeable memories: Legal and practical implications. *Nature Neuroscience Reviews* **4**, 231–234.
- Magnussen, S., Endestad, T., Koriat, A., and Helstrup, P. (2007). What do people believe about memory, and how do they talk about memory? In Magnussen, S. and Helstrup, T. (eds.) *Everyday Memory*, pp 5–25. Hove: Psychology Press.
- Roediger, H. L. and Karpicke, J. D. (2006). Test-enhanced learning – taking memory tests improves long-term retention. *Psychological Science* **17**, 249–255.
- Rubin, D. C. (2006). The basic-systems model of episodic memory. *Perspectives on Psychological Science* **1**, 277–311.
- Schacter, D. L. (1996). *Searching for Memory: The Brain, the Mind, and the Past*. New York: Basic Books.
- Schacter, D. L. (2001). *The Seven Sins of Memory: How the Mind Forgets and Remembers*. Boston, MA: Houghton Mifflin.
- Schacter, D. L., Wagner, A. D., and Buckner, R. L. (2000). Memory systems of 1999. In Tulving, E. and Craik, F. I. M. (eds.) *The Oxford Handbook of Memory*, pp 627–643. Oxford: Oxford University Press.
- Schwartz, B. L. (2002). *Tip-of-the-Tongue States: Phenomenology, Mechanism and Lexical Retrieval*. London: Erlbaum.
- Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology* **53**, 1–25.
- Tulving, E. and Schacter, D. L. (1990). Priming and human memory systems. *Science* **247**, 301–306.
- Valentine, T., Brennen, T., and Brédart, S. (1996). *The Cognitive Psychology of Proper Names: The Importance of Being Ernest*. London: Routledge.
- Wagenaar, W. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology* **18**, 225–252.
- Wells, G. L., Memon, A., and Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest* **7**, 43–75.
- White, R. (2002). Memory for events after twenty years. *Applied Cognitive Psychology* **16**, 603–612.

Further Reading

- Bahrick, H. P., Bahrick, P. O., and Wittlinger, R. P. (1975). Fifty years of memories for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General* **104**, 54–75.
- Cohen, G. (2004). *Memory in the Real World*, 2nd edn. Hove: Psychology Press.
- Kundera, M. (2002). *Ignorance*. New York: Harper Collins.
- Magnussen, S. and Helstrup, T. (eds.) (2007). *Everyday Memory*. Hove: Psychology Press.
- McNally, R. J. (2003). *Remembering Trauma*. Cambridge, MA: Harvard University Press.
- Toglia, M. P., Read, J. D., Ross, D. F., and Lindsay, R. C. L. (eds.) (2007). *The Handbook of Eyewitness Psychology. Vol. I, Memory for Events*. London: Erlbaum.
- Tulving, E. and Craik, F. I. M. (eds.) (2000). *The Oxford Handbook of Memory*. Oxford: Oxford University Press.
- Squire, L. and Kandel, E. R. (1999). *Memory. From Mind to Molecules*. New York: Freeman.

Metacognition

L Baker, University of Maryland, Baltimore County, Baltimore, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cognitive monitoring – It refers to any activity aimed at evaluating or regulating one's own cognitions, including planning, checking, self-testing, assessing one's progress, and correcting one's errors.

Comprehension monitoring – It refers to metacognitive control in reading or listening, which involves deciding whether or not one understands (evaluation) and taking appropriate steps to correct comprehension problems that are detected (regulation).

Executive function – It is the cognitive system that controls and manages other cognitive processes, is associated with frontal lobe functioning, and includes processes typically regarded as metacognitive in nature, such as planning, monitoring, and error correction and detection.

Metacognition – It refers to thinking about thinking; knowledge about cognition and regulation of cognition; knowledge about person, task, and strategy variables that affect performance in a given domain; and cognitive monitoring and control.

Metacomprehension – It refers to understanding one's understanding, that is, realizing that one has understood.

Metamemory – It is knowledge about memory and control of memory, that is, appraisal of one's memory on a given task.

Other-regulation – It refers to regulation of the behaviors and cognitive activities of one individual by another, usually more knowledgeable; it involves Vygotsky's proposed sequence of other-regulation to self-regulation.

Reciprocal teaching – It is an approach to foster cognitive and metacognitive strategies in reading that involves peer collaboration; students are taught to use strategies of predicting, clarifying, summarizing, and questioning.

Self-regulation – It is the ability to control one's own behavior and cognitive activities; self-regulated learning is self-directed, intrinsically motivated, and under the deliberate, strategic control of the learner; it involves effective use of processes included within the regulatory component of metacognition, such as planning, monitoring, and evaluating.

Self-system – It is the cluster of motivational and affective processes that influence why we do what we do; it includes self-perceptions, attributions for performance outcomes, goals for engaging in an activity, and achievement motivation.

Definitional Issues

A question of ongoing interest and importance is how and when students develop knowledge and control of their cognitive processes. This higher-level cognition was given the label metacognition by the American developmental psychologist John Flavell in the mid-1970s. Metacognition is concerned with our ability to reflect on our own thinking, and in an academic context it includes knowledge about ourselves as learners, about aspects of the task, and about strategy use. Metacognition also involves self-regulation of our own cognitive efforts, including planning our actions, checking the outcomes of our efforts, evaluating our progress, remediating difficulties that arise, and testing and revising our strategies for learning.

The construct of metacognition has had wide appeal and wide applicability, stimulating a great deal of research across a broad spectrum of disciplines. Research today spans the subdisciplines of educational, developmental, cognitive, clinical, social, comparative, and cognitive psychology, as well as related fields such as neuroscience, linguistics, and second language learning, special education, and speech and communication disorders. Once developmental psychologists began to study metacognition in the 1970s, the construct quickly attracted the attention of educational researchers seeking an explanation for why some students fared better in school than others. The consistent finding for more than 30 years has been that students who are more successful in a domain exhibit higher levels of metacognitive knowledge about the domain and are more skilled at regulating their cognitive processes. Instructional interventions to promote metacognition quickly became popular and remain so today. The focus in this article is on metacognition in educational contexts.

Metacognition has been defined in different ways, and that of course affects how it has been studied. The term literally means thinking about thinking. John Flavell and Ann Brown initially defined metacognition as knowledge about cognition and regulation of cognition.

This two-component conceptualization of metacognition has been widely but not exclusively used since then by educational researchers. Some have restricted the term to the knowledge component, whereas others have restricted it to the control component. For example, researchers within the cognitive-experimental tradition use a definition proposed by Thomas Nelson that includes two components, monitoring and control. Two closely related constructs are associated with the control aspect of metacognition: self-regulation and executive functioning. The term self-regulation is often used by educational psychologists to refer to the use of skills included within the regulatory component of metacognition, such as planning, monitoring, and evaluating. (Developmental psychologists have a broader definition of self-regulation that encompasses impulse control.) Executive function is a term with origins in cognitive psychology and neuroscience. It includes processes typically regarded as metacognitive in nature, such as planning, monitoring, and correcting and detecting errors, and it is typically linked to frontal lobe functioning in the brain.

Whereas metacognition once was studied solely in terms of how it is related to cognition, it is now recognized that one cannot understand how and why people perform as they do on cognitive tasks without an examination of motivational and affective as well as metacognitive factors. Accordingly, many contemporary researchers examine the role of the self-system, which includes motivation, perceived competence, and attributional beliefs, in conjunction with metacognition and cognition.

Origins of Metacognition

Much of the research on metacognition in educational contexts is descriptive in nature, documenting developmental and individual differences in knowledge and control. Less often do researchers address basic questions about the mechanisms of metacognitive growth. Nevertheless, it is now widely agreed that the origins of metacognition lie at least in part in social interactions. Russian psychologist Lev Vygotsky proposed that children first learn how to engage in cognitive tasks through social interaction with more knowledgeable others, usually parents or teachers. The expert initially takes the responsibility of regulating the novice's activity by setting goals, planning, evaluating, and focusing attention on what is relevant. Gradually, the expert gives over more and more responsibility to the novice as the novice becomes capable of assuming it, until finally the novice internalizes the regulatory mechanisms and can perform without expert assistance. Many successful interventions aimed at fostering metacognition draw on Vygotsky's proposed sequence from other-regulation to self-regulation.

Whereas Vygotsky emphasized expert-novice interactions, another influential social perspective is that of Swiss psychologist Jean Piaget, who emphasized the importance of peer interactions. Piaget argued that peers challenge one another's thoughts and thus advance cognitive development. Discussion and collaboration help students to monitor their own understanding and build new strategic capabilities. Instructional interventions often supplement an emphasis on individual cognition with an emphasis on peer support for monitoring, reflection, and revision. These interventions, too, have met with success, providing validation for the perspective that metacognition has its origins in social interaction.

It is important not to overlook maturation as a further contributor to metacognitive growth. Adolescence is a period of major development in the prefrontal cortex, the portion of the brain involved in executive function. As noted previously, executive function includes metacognitive processes such as planning and monitoring. The prefrontal lobes are the last portions of the brain to develop, with maturation not complete until late adolescence or early adulthood. This late maturation helps explain why limitations in metacognitive control are still apparent in high school and college students.

Methods for Assessing Metacognition

Considerable controversy exists as to the best ways to measure metacognition. Methods vary in part depending on the theoretical orientation of the researcher. Methods also must be selected with developmental considerations in mind. Therefore, a variety of approaches have been used to study metacognition. Each approach has limitations, so it is advisable to use multiple measures that converge on the construct.

Verbal Reports

The most frequently used approach to assess both metacognitive knowledge and metacognitive control is to ask students directly about what they know or what they do while engaging in cognitive activities. Such self-reports have been collected in a variety of ways. For assessing metacognitive control, participants may be asked to think aloud about what they are doing and thinking as they solve a problem or read a text or they may be asked to provide written comments periodically throughout the session. Participants may alternatively be asked to complete checklists of strategies they used in a given task after its completion, or they may complete questionnaires or study-strategy inventories.

Whereas verbal reports are but one way for assessing metacognitive control, they are the primary basis for collecting information about metacognitive knowledge, either

through interviews or questionnaires. John Flavell's early study of metamemory that fueled the interest in metacognition used a structured interview format. Questionnaires were subsequently developed, often based on interview findings or think-aloud protocols. The major limitation of such approaches is that there is not necessarily a correspondence between what people say they do and what they actually do. Comparisons of general questionnaire responses with performance measures on a given task often yield rather low correlations. In addition, students, particularly children, often respond according to what they think they should say, rather than what they actually believe or do. Nevertheless, despite their limitations, there is a general consensus that verbal reports can be valid and reliable sources of information about cognitive processes when elicited and interpreted according to established guidelines.

Most questionnaires designed to assess metacognition are domain specific (e.g., reading or mathematics), but a few are intended to be more domain general. A domain-specific inventory might tap a student's understanding of variables that affect reading outcomes and of strategies that are effective for comprehending text. A domain-general inventory might assess an individual's knowledge about cognition (including declarative, procedural, and conditional knowledge) and regulation of cognition (including planning, monitoring, debugging, and evaluating learning). Relatively few well-validated and reliable instruments are available for research or practical use.

Online Processing Measures

Metacognitive control is frequently studied by asking students to engage in a task and collecting process measures as the task is being completed. For example, to assess metacognitive control during reading, a passage may be presented to the reader on a computer screen. Patterns of movement through the text are collected automatically, revealing whether the reader paused at a particular point, whether he or she looked back at previous text, whether he or she jumped ahead. These measures are often supplemented by asking readers to reflect on what they were thinking, or by giving them comprehension questions and then relating outcomes to the processing measures. Computer technology has similarly been used to capture processing during writing tasks and during mathematical problem solving. An advantage of these approaches is that they reveal what students do instead of what they say they do. However, the naturalness of the cognitive task is often disrupted by the specialized technology demands.

Judgments of Learning and Predictions of Performance

In judgments of learning (JOL) tasks, students are presented with to-be-learned material, such as a list of words

or a passage, and then they are given a test over the material. They are then asked to judge how well they learned the material or how well they answered the comprehension questions (often referred to as an index of metacomprehension). Judgments of learning are then examined in relation to actual performance.

A related approach, often referred to as knowledge monitoring, involves presenting students with material and asking them how well they think they would perform on a test. For example, they might be presented with a list of vocabulary words and be asked how many they would be able to define, or a set of math problems and asked how many they can solve. They are then asked to complete the task, and their performance is compared to their predictions.

These types of tasks are more likely to be used in research conducted by cognitive psychologists interested in basic processes than by educational researchers concerned with more applied issues. Nevertheless, the approach has its origins in seminal research on memory monitoring, and research on studying (applied memory) often includes assessments of test readiness.

Metacognition in Specific Academic Domains

Most of the research on metacognition is focused on a particular academic domain, in recognition that metacognition is largely domain specific. In other words, a student may have a great deal of metacognitive knowledge about reading, but that knowledge does not necessarily generalize to mathematics. This section addresses research findings on metacognition as it has been studied in reading, writing, mathematics, and science.

Metacognition in Reading

Research on metacognition in reading began in the 1970s, shortly after the publication of Flavell's seminal studies on metamemory. Pioneers in this area of research were Ann Brown and Scott Paris. Reading is the academic domain most frequently studied, and so it is given more attention in this article than the other content areas.

Studies of metacognitive knowledge

Much of the research on metacognitive knowledge about reading uses structured interview procedures adapted from Flavell's metamemory work. In a typical study, children are interviewed about their knowledge of person, task, and strategy variables involved in reading. The pattern that has been found consistently is that younger readers have little awareness that they must attempt to make sense of text. They focus on reading as a decoding process, rather

than as a meaning-getting process. Ability-related differences in knowledge about reading, like developmental differences, have been documented in countless studies, across age groups ranging from early childhood through later adulthood. Students' metacognitive knowledge about reading, whether assessed through interviews, questionnaires, or verbal reports, remains an active and important area of inquiry.

Studies of comprehension monitoring

Effective text comprehension requires an important metacognitive control component known as comprehension monitoring. Comprehension monitoring involves deciding whether or not we understand (evaluation) and taking appropriate steps to correct whatever comprehension problems we detect (regulation). Research has shown that students of all ages often are ineffective at monitoring their understanding of text. These difficulties are perhaps most apparent when students are asked to read information text, such as science and history textbooks. Failures to evaluate and regulate understanding reduce the likelihood of meaningful learning.

The majority of studies of comprehension monitoring over the years have used the error-detection paradigm. In this approach, errors or problems are deliberately introduced into texts, and various indices are used to determine whether readers notice the problems and attempt to resolve them. For example, readers may be asked to underline or report detected errors, or online processing measures may reveal longer pauses when readers encounter problematic text. Caution is needed in interpreting these studies because of students' propensity to believe texts are true and well structured and because of their reluctance to acknowledge comprehension difficulties. Moreover, some types of errors are more likely to be reported than others; for example, younger and less-skilled readers are more likely to evaluate their understanding using a word-level criterion than an internal-consistency criterion.

Studies aimed at fostering metacognitive skills

Intervention studies began to be implemented in the 1980s, providing solid evidence that metacognitive knowledge and control could be enhanced through direct instruction. Comprehensive classroom reading interventions incorporated metacognitively oriented instruction, with the goal not of enhancing metacognition *per se*, but rather of promoting students' reading comprehension by increasing their metacognition. Interventions that use some variant of reciprocal teaching, in which peers collaborate to learn and apply strategies of predicting, clarifying, summarizing, and questioning, have yielded substantial effects on reading comprehension. The research

base for the effectiveness of metacognitive intervention is now so strong that professional organizations and national panels recommend that metacognition be included in reading comprehension curricula.

Metacognition in Writing

Metacognition plays a role in the production of text as well as in the comprehension of text. Research on metacognition in writing has followed a similar course to that on reading. In the 1980s, researchers began to compare the metacognitive knowledge and control of more-skilled and less-skilled writers. Across all age groups, clear differences are apparent in students' conceptions of writing, their knowledge of the writing process, and their abilities to implement effective control strategies. For example, better writers focus more on the function of writing, whereas poorer writers focus more on form. When asked about their conceptions of writing, better writers discuss the qualities of good writing, such as having a clear beginning, middle, and end, whereas poorer writers discuss spelling all of the words correctly. Skilled writers have higher-order awareness of the writing process, such as awareness of the need for clarity, organization, and audience sensitivity, whereas less-skilled writers tend to focus on lower-order processes dealing with spelling, grammar, and punctuation.

Interest in metacognitive control of writing coincided with the publication of a cognitive process model of writing by Linda Flower and John Hayes in the early 1980s. The model includes three recursive processes: planning, translating (sentence generation), and revising, all of which are controlled by a monitor or executive component. Research has demonstrated consistent differences in how expert and novice writers handle the planning and revision processes. For example, experts are more likely to have global goals for their writing that take into account the communicative purpose of the task, and they revise at a global level. In contrast, novices seldom have overall plans for their writing and their revisions are typically made at a sentence-by-sentence level.

Also, as with reading, classroom interventions have been devised and implemented to increase students' metacognitive awareness and control of writing processes. Many of these interventions aim to reduce the cognitive processing demands on less-skilled writers by providing prompts, cues, and scaffolds. Marlene Scardamalia and Carl Bereiter developed an approach known as procedural facilitation that has proven effective with elementary school students and has been adapted up through the college level. Prompts are provided to remind students of the steps in the planning process, for example, or things that need to be taken into account when revising a paper. Often the translation demands are reduced so that

students who are overly focused on spelling, punctuation, and handwriting do not get bogged down with the lower-level components (e.g., a word processing system is used).

Metacognition in Mathematics

Research on metacognition in mathematics also began to appear in the 1980s. As with reading and writing, attention focused on both knowledge and control and on students of all ages. A paper by Alan Schoenfeld in 1987 was particularly influential in bringing metacognition to the attention of mathematics educators (What's all the fuss about metacognition?). Schoenfeld presented evidence from the college-level math course he taught that showed that many students did not reflect on the problem-solving strategies they used and that they frequently failed to connect the solutions they obtained with the real world.

Mathematical strategy knowledge includes basic knowledge of algorithms and heuristics, as well as metacognitive awareness of strategies to help in comprehending problem statements, organizing information or data, planning solution attempts, executing plans, and checking results. Even first-grade children have some specific strategy knowledge about math, but students seldom monitor their approaches to problem solving until much later. At first, problem solving is often taken one step at a time, and students show little understanding of the general principles of the problems. This is similar to the sentence-level focus of young writers who approach the task without an overall plan.

Just as students must monitor their comprehension of what they read, so too must they monitor their cognitive processes while doing math. Consider the different levels at which one may monitor the adequacy of a solution to a word problem. One may evaluate the results of arithmetic procedures carried out to obtain an answer; such monitoring could occur simply by checking the arithmetic involved (e.g., subtracting an addition result). One may evaluate whether the procedure one has chosen is correct. This requires some semantic analysis of the text, whether rereading the entire text or simply finding a key word. One may also evaluate the sensibleness of the problem itself. This involves looking at the relations expressed in the text and making a decision about whether those relations make sense. Research has shown difficulties in monitoring at all three levels, with younger and poorer math students more likely to focus on the low-level calculation standard of evaluation. These results are comparable to the reading and writing research, which shows novices more likely to evaluate low-level aspects of the task (e.g., decoding or spelling).

Metacognitive research in mathematics also includes a line of work focused on students' judgments of whether they will be able to solve problems correctly and judgments of which problems they succeeded in answering

correctly. When judgments or predictions are compared with actual performance, the general pattern of results is that younger and less-skilled students are less accurate in monitoring their abilities than their older and higher-achieving peers.

As with literacy, professional organizations now advocate that reflection and metacognition be central components of mathematics teaching. Classroom interventions have been devised and successfully implemented to help students plan, monitor, and evaluate their own thinking during mathematical problem solving. The National Council of Teachers of Mathematics in the United States emphasized the importance of writing as a way of helping students reflect: "writing in mathematics can also help students consolidate their thinking because it requires them to reflect on their work and clarify their thoughts about the ideas" (NCTM, 2000: 61). Research has shown the value of peer collaboration and discussion in enhancing mathematics performance, as students make their cognitive processes, assumptions, and strategies explicit. Consistent with a Piagetian constructivist perspective, the cognitive conflicts that may arise as students justify their approaches are effective in stimulating cognitive and metacognitive growth.

Metacognition in Science

The body of research on metacognition in science focuses primarily on the reading of scientific texts; the research results are comparable to those already discussed in the domain of reading. Scientific problem solving is clearly another academic endeavor where metacognitive knowledge and control are critical, but it is much less researched. The term metacognition may be absent from the following statement in the National Science Education Standards (NSES) in the United States, but there can be no doubt as to the central concern: "Engaging students in inquiry helps students develop an appreciation of 'how we know' what we know in science" (p. xx). The standards call for teachers to guide students to understand the purposes for their own learning and to develop their abilities to assess and reflect on their own scientific accomplishments. They also call for providing students with opportunities to apply standards of scientific practice to their own and others' scientific efforts.

Summary and Conclusions

It might be tempting to conclude from this brief review of metacognition within specific academic domains that metacognition is domain general. This conclusion may be drawn because the developmental and individual differences that have been revealed across domains are conceptually similar. Nevertheless, correlations in performance across domains

are often rather low. It is therefore important not to assume that metacognitive skills can and should be fostered in settings devoid of academic content.

It also might be tempting to conclude that the relation between metacognition and cognition is uni-directional, given the evidence that metacognitive training can promote greater achievement in a given domain. However, reciprocal causation is most likely; that is, improvements in metacognition contribute to improvements in cognition (reading, writing, problem solving, etc.), which in turn contribute to further improvements in metacognition.

One of the reasons why young students exhibit limited metacognitive awareness and control is because the basic processes needed to carry out the activity are not yet routinized or fluent. For example, beginning readers who must allocate all of their processing capacity to decoding have little cognitive capacity left to devote to meaning construction, let alone evaluate the adequacy of the meaning that was constructed. With repeated experience, readers learn to decode and to monitor their comprehension sufficiently well that they do not have to allocate attention to the processes; it is only when an obstacle is noted that attention is directed to the problem area. Automaticity is also critical to mathematical performance – students need to be sufficiently fluent with basic number facts so that simple addition and subtraction do not take up the cognitive resources needed for higher-level conceptual processes. Similarly, monitoring and evaluation of ongoing processes need to become automatic, so that they can proceed without deliberate attention. The same logic follows for writing. In other words, students should be able to proceed with a cognitive activity fluently and automatically, with metacognitive control processes reaching conscious awareness only at times when they must be deployed.

When students have knowledge and control of their own cognitive processes, learning is enhanced. This holds regardless of the domain of learning, whether reading, writing, science, or any other activity that involves thinking. Educators across disciplines now have the common goal of fostering metacognitively sophisticated learners, and the means to accomplish this goal appear to be similar: (1) Scaffolded instruction should take place embedded within a meaningful context. (2) Metacognition should not be promoted as an end in itself but rather as a means of promoting learning and achievement. (3) Students should be taught to apply a critical stance toward the information they encounter, overcoming the common tendency to accept at face value the accuracy and plausibility of information conveyed to them by perceived authorities. (4) Students should be given opportunities to work collaboratively with their peers on common problems and tasks and to articulate their cognitive and metacognitive processes.

Although the construct of metacognition is now widely known in the educational research community, empirical findings still look a great deal like they did when metacognition first became an area of inquiry. For example, studies conducted in the current decade reveal that 8–10-year-olds rely almost exclusively on word-level criteria for evaluating their understanding, replicating the findings of more than 20 years earlier. Similarly, young students' conception of a good reader as one who reads quickly without making any mistakes is the same as that identified almost 30 years earlier. These patterns are troubling because they illustrate how slowly advances in research knowledge are translated into changes in classroom practice that in turn bring about changes in student outcomes.

See also: Knowledge Domains and Domain Learning; Personal Epistemology in Education: Concepts, Issues, and Implications; Piaget: Recent Work; Problem Solving and Human Expertise; Self-Regulated Learning and Socio-Cognitive Theory; Vygotsky and Recent Developments.

Further Reading

- Baker, L. (2004). Reading comprehension and science inquiry: Metacognitive connections. In Saul, W. (ed.) *Crossing Borders in Literacy and Science Instruction: Perspectives on Theory and Practice*, pp 239–257. Newark, DE: International Reading Association.
- Baker, L. (2005). Developmental differences in metacognition: Implications for metacognitively-oriented reading instruction. In Israel, S. E., Block, C. C., Bauserman, K. L., and Kinnucan-Welsch, K. (eds.) *Metacognition in Literacy Learning: Theory, Assessment, Instruction, and Professional Development*, pp 61–79. Mahwah, NJ: Erlbaum.
- Baker, L. (2008). Metacognitive development in reading: Contributors and consequences. In Mokhtari, K. and Sheorey, R. (eds.) *Reading Strategies of First- and Second-Language Learners: See How They Read*, pp 25–42. Norwood, MA: Christopher Gordon.
- Baker, L. and Brown, A. L. (1984). Metacognitive skills and reading. In Pearson, P. D., Kamil, M., Barr, R., and Mosenthal, P. (eds.) *Handbook of Research in Reading*, vol. 1, pp 353–395. New York: Longman.
- Baker, L. and Cerro, L. (2000). Assessing metacognition in children and adults. In Schraw, G. and Impara, J. (eds.) *Issues in the Measurement of Metacognition*, pp 99–145. Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska.
- Carr, M. and Biddlecomb, B. (1998). Metacognition in mathematics from a constructivist perspective. In Hacker, D. J., Dunlosky, J., and Graesser, A. C. (eds.) *Metacognition in Educational Theory and Practice*, pp 69–91. Mahwah, NJ: Erlbaum.
- Chambres, P., Izaute, M., and Marescaux, P.-J. (eds.) (2002). *Metacognition: Process, Function and Use*. Boston, MA: Kluwer.
- Desoete, A. and Veenman, M. (eds.) (2006). *Metacognition in Mathematics Education*. Hauppauge, NY: Nova Science.
- Hacker, D. J., Dunlosky, J., and Graesser, A. C. (eds.) (1998). *Metacognition in Educational Theory and Practice*. Mahwah, NJ: Erlbaum.
- Hartman, H. J. (ed.) (2001). *Metacognition in Learning and Instruction: Theory, Research and Practice*. New York: Springer.
- Israel, S. E., Block, C. C., Bauserman, K. L., and Kinnucan-Welsch, K. (eds.) (2005). *Metacognition in Literacy Learning: Theory, Assessment, Instruction, and Professional Development*. New York: Erlbaum.

- McCormick, C. (2003). Metacognition and learning. In Reynolds, W. M., Miller, G. J., and Weiner, I. B. (eds.) *Handbook of Psychology: Volume 7, Educational Psychology*, pp 79–102. New York: Wiley.
- NCTM (2000). *Principles and Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Schraw, G. and Impara, J. (eds.) (2000). *Issues in the Measurement of Metacognition*. Lincoln, NE: Buros Institute of Mental Measurements.
- Sitko, B. M. (1998). Knowing how to write: Metacognition and writing instruction. In Hacker, D. J., Dunlosky, J., and Graesser, A. C. (eds.) *Metacognition in Educational Theory and Practice*, pp 93–115. Mahwah, NJ: Erlbaum.
- Van Haneghan, J. P. and Baker, L. (1989). Cognitive monitoring in mathematics. In McCormick, C. B., Miller, G., and Pressley, M. (eds.) *Cognitive Strategy Research: From Basic Research to Educational Applications*, pp 215–238. New York: Springer.
- Wong, B. Y. L. (1999). Metacognition in writing. In Gallimore, R., Bernheimer, L. P., MacMillan, D. L., Speece, D. L., and Vaughn, S. (eds.) *Developmental Perspectives on Children with High Incidence Disabilities*, pp 183–198. Mahwah, NJ: Erlbaum.

Personal Epistemology in Education: Concepts, Issues, and Implications

I Bråten, University of Oslo, Oslo, Norway

© 2010 Elsevier Ltd. All rights reserved.

What Is Personal Epistemology?

Epistemology, the study of knowledge and knowing, has absorbed philosophers from ancient times. As a branch of philosophy, epistemology concerns the nature, origins, and limitations of knowledge, as well as the justification of truth claims. Recently, educational researchers have become interested in the theories and conceptions of knowledge and knowing that students hold, with the term personal or folk epistemology used to distinguish the lay person's view about knowledge and knowing from the trained philosopher's view (Hofer and Pintrich, 2002; Kitchener, 2002). Thus, personal epistemology essentially refers to the theories or beliefs that students (and other individuals) hold about knowledge and the process of knowing (Hofer and Pintrich, 1997, 2002). Likewise, epistemic theories or beliefs refer to individuals' views about knowledge and knowing (i.e., the epistemic).

How Does Personal Epistemology Develop?

Most educational research on personal epistemology has been rooted in Perry's (1970) longitudinal interview studies at Harvard, resulting in a scheme describing the development of personal epistemology during the college years. Several researchers have continued Perry's (1970) effort to identify developmental stages or sequences in students' personal epistemology, mostly through the use of interviewing methodology (e.g., Baxter Magolda, 1992; King and Kitchener, 1994; Kuhn, 1991). This line of research has generally described a developmental trajectory beginning with a dualist view where knowledge is seen as either right or wrong and where it is possible to know what is right with certainty. This is followed by a period of multiplicity where multiple conflicting views are acknowledged and accepted as equally valid. Finally, a more evaluativistic perspective develops where individuals acknowledge that there is no absolutely certain knowledge but still believe that it is possible to evaluate competing knowledge claims and justify claims through the use of supporting evidence.

While educational researchers have been most concerned with development in personal epistemology during adolescence and early adulthood, some research has also

been conducted with children. Such research has linked the development of personal epistemology to children's theory of mind and also identified a predualistic stage of egocentric subjectivism where children believe that their own perspective is the only perspective (e.g., Burr and Hofer, 2002). Moreover, research conducted at younger ages has identified progression through the developmental stages observed in college students regardless of the age of the participants. Such repetition of previously passed stages, possibly several times from childhood to adulthood, suggests that the development of personal epistemology is recursive rather than linear, with recursion possibly occurring when students enter new educational contexts (e.g., traverse from elementary to secondary and from secondary to postsecondary education) (Muis *et al.*, 2006). However, other plausible explanations for the developmental-recursiveness puzzle also exist, for example, that development varies with the kind of facts considered (Hallett *et al.*, 2002).

What Are the Dimensions of Personal Epistemology?

Some researchers have been less interested in the development of personal epistemology than in its dimensionality. This approach was pioneered by Schommer (1990), who described personal epistemology as a system of more or less independent beliefs about certain knowledge (ranging from the belief that knowledge is absolute and unchanging to the belief that knowledge is tentative and evolving), simple knowledge (ranging from the belief that knowledge is best characterized as isolated bits and pieces to the belief that knowledge is best characterized as highly integrated concepts), omniscient authority (ranging from the belief that knowledge is handed down by authority to the belief that knowledge is derived from reason), quick learning (ranging from the belief that learning takes place quickly or not at all to the belief that learning is gradual), and fixed ability (ranging from the belief that ability to learn is given at birth to the view that ability to learn can be increased). Using a questionnaire to examine the described system, Schommer and associates identified factors corresponding to all the proposed dimensions except omniscient authority, even though this factor structure has not been consistently replicated by other

researchers (e.g., Bråten and Strømsø, 2005; Hofer, 2000; Qian and Alvermann, 1995).

Whereas the three first dimensions in Schommer's (1990) conceptualization fall under the cited definition of personal epistemology as theories or beliefs about the nature of knowledge (certain knowledge and simple knowledge) and knowing (omniscient authority), the two remaining dimensions are more controversial because they mainly concern beliefs about learning (quick learning) and intelligence (fixed ability). Hofer and Pintrich (1997) therefore placed quick learning and fixed ability outside the realm of personal epistemology and purified the construct by considering two dimensions concerning the nature of knowledge (what one believes knowledge is) and two dimensions concerning the nature or process of knowing (how one comes to know). In their conceptualization of personal epistemology, the dimensions certainty of knowledge and simplicity of knowledge, both concerning the nature of knowledge, correspond to the dimensions certain knowledge and simple knowledge as described by Schommer (1990). Within the area of nature of knowing, the dimension source of knowledge, in part paralleling omniscient authority in Schommer's belief system, ranges from the conception that knowledge originates outside the self and resides in external authority, from which it may be transmitted, to the conception that knowledge is actively constructed by the person in interaction with others. Finally, the dimension justification for knowing, also concerning the nature of knowing, refers to how individuals evaluate and justify knowledge claims. This dimension ranges from justification through observation and authority, or on the basis of what feels right, to the use of rules of inquiry and the evaluation and integration of multiple sources.

Even though the dimensionality proposed by Hofer and Pintrich (1997) is currently the most authoritative view on the dimensionality of personal epistemology, and as such widely recognized by researchers in the field, the conceptually derived dimensions have not been unequivocally empirically verified through factor analysis (Hofer, 2000). However, use of qualitative methodologies such as observations and interviews (Hofer, 2004a) or think-aloud protocols (Hofer, 2004b) indicates that all the four dimensions proposed by Hofer and Pintrich (1997) are represented in students' epistemic thinking.

Hofer and Pintrich (1997) conceptualized individuals' epistemic beliefs to be theory-like, that is, integrated and coherent rather than existing as more or less independent beliefs, as Schommer (1990) originally suggested. Regarding this issue, recent evidence suggests that the various dimensions of personal epistemology may exist and operate independently, meaning that a student may hold what could be characterized as a more sophisticated belief on one dimension (e.g., that knowledge is integrated) and, at the same time, hold what could be characterized as a more

naive belief on another dimension (e.g., that knowledge is unchanging) (Buehl and Alexander, 2005).

Is Personal Epistemology Domain General, Domain Specific, or Both?

Whereas initial educational research was based on the assumption that epistemic beliefs were independent of academic domains, implying, for example, that students would hold the same beliefs about knowledge and knowing in mathematics as they would in education, some studies conducted in the 1990s started to question this assumption. Those studies could be divided into between-subjects investigations, where students majoring in different domains were compared with respect to their epistemic beliefs, and within-subjects investigations, where the same students were asked about their epistemic beliefs in different domains. Buehl and Alexander (2001) found that most existing evidence supported the view that epistemic beliefs varied as a function of academic domains and, moreover, that such variation was related to domain structuredness (e.g., between mathematics as a well-structured domain and education as an ill-structured domain). However, Buehl and Alexander (2001) also argued that students' epistemic beliefs were not solely domain specific, that is, students could simultaneously hold both domain-specific and more domain-general or overarching epistemic beliefs.

A later review by Muis *et al.* (2006) confirmed that this is not an issue of either-or. Of the 19 studies that they reviewed, eight between-subjects investigations and 11 within-subjects investigations, 17 established evidence for domain specificity on one or more of the dimensions of personal epistemology, with six of those 17 studies also indicating some degree of domain generality. The review by Muis *et al.* (2006) also confirmed that epistemic belief similarities and differences across domains were related to whether the compared domains were similar or different with respect to structuredness, as well as whether they were similar or different on the hard-soft and the pure-applied dimensions. In general, students seem to view knowledge as more certain and integrated and more readily accept experts as sources of knowledge in well-structured or hard domains such as mathematics than in ill-structured or soft domains such as education.

The conclusion that personal epistemology includes levels of both domain generality and domain specificity was further supported by Buehl and Alexander (2005), who used cluster analysis to compare the student profiles that emerged from different dimensions of personal epistemology across the domains of mathematics and history. While the distinct epistemic belief profiles that emerged differed across the two domains, there was also some consistency in students' profile membership in mathematics and history, with this finding also consistent with a dual-level

conception of personal epistemology. Moreover, a useful distinction may concern domain and topic-specific epistemic beliefs. Just as domain knowledge and topic knowledge may form subcategories of formally acquired or schooled knowledge, with domain knowledge referring to the breadth of one's knowledge about a domain (e.g., psychology or history), and topic knowledge representing the depth of one's knowledge about particular contents or concepts within a domain (e.g., intelligence or World War II) (Alexander *et al.*, 1991), beliefs about knowledge and knowing in a domain may be distinguished from epistemic beliefs about topics within domains. As an example of such a topic-specific approach to personal epistemology, Trautwein and Lüdtke (2007) used questionnaire items to examine students' epistemic beliefs about specific scientific theories, for example, about biological theories concerning natural selection and extinction of the dinosaurs, respectively. It was found that epistemic beliefs differed considerably across theories. At the same time, a small but statistically significant association between topic-specific epistemic beliefs and more general epistemic beliefs about scientific knowledge was found. Likewise, Bråten (2008) described research on topic-specific epistemic beliefs concerning student views on knowledge about climate change and how one comes to know about climate.

There is a clear need to further examine how different levels of personal epistemology develop in interaction and how they operate together to promote or constrain various aspects of student motivation and learning. Possibly, personal epistemology at different levels of specificity has strongest impact on facets of academic learning at comparable levels of specificity (Schraw, 2001). Clearly, the consideration of various levels of personal epistemology also implies that measures of personal epistemology become tailored to particular levels.

The Role of Personal Epistemology in Student Motivation, Cognition, and Performance

Schommer's (1990) departure from the developmental paradigmatic approach to personal epistemology and her introduction of quantitative assessment in the form of a paper-and-pencil questionnaire initiated an important line of research on relations between personal epistemology and other academic constructs. In accordance with Schommer's (1990) initial findings, quite a few other studies have linked epistemic beliefs to students' text-based learning and comprehension (e.g., Buehl and Alexander, 2005), with this body of research generally indicating that beliefs that have traditionally been located at the naive ends of epistemic belief continuums (e.g., beliefs that knowledge is certain or simple) are related to poorer learning and comprehension. Other findings show that

more naive epistemic beliefs are negatively related to argumentative reasoning (e.g., Kuhn, 1991), conceptual change learning (e.g., Qian and Alvermann, 1995), and graded academic performance (e.g., Wood and Kardash, 2002). Such relations between epistemic beliefs and student learning, comprehension, and performance may well be mediated by the use of cognitive and metacognitive strategy use, as several researchers have documented that personal epistemology is related to students' strategic processing (for review, see Muis, 2007).

Bråten (2008) noted that research on personal epistemology and text-based learning and comprehension had almost exclusively focused on students' reading of one single text, and, moreover, that most research on personal epistemology and aspects of learning had been conducted in traditional print environments rather than in new technological environments. As personal epistemology may be particularly important when students work on complex learning tasks (Spiro *et al.*, 1996), these limitations are not trivial.

Regarding the reading of multiple texts, Rukavina and Daneman (1996) provided some early evidence that students holding more sophisticated epistemic beliefs about the complexity of knowledge were better equipped to integrate ideas across two texts presenting conflicting information on a topic. Later, Bråten and Strømsø (2006) provided new evidence that at least adult college readers are able to deal adequately with the challenge of integrating information from multiple, even conflicting, texts, provided that they hold relatively sophisticated beliefs about the nature of knowledge and knowing. Otherwise, even college readers may be better off when encountering the same content in an integrated textbook format.

Regarding the importance of personal epistemology when learning with hypermedia technology, Jacobson and Spiro (1995) provided preliminary evidence that students who believed in simple knowledge had problems handling the nonlinear and multidimensional nature of an ill-defined hypertext system. Bendixen and Hartley (2003) provided additional evidence to suggest that students' epistemic beliefs play an important role in hypermedia-learning environments. Moreover, Hofer (2004b) reported that when students thought aloud during online searching, those expressing naive epistemic beliefs were likely to pursue the searching task in a brief and perfunctory way, not seeing the need for additional sources or reflecting on the credibility and accuracy of the sources they located. Finally, Bråten and colleagues (e.g., Bråten *et al.*, 2005) found that students' epistemic beliefs predicted their Internet-based learning activities. In brief, the more naive beliefs students held about Internet-based knowledge and knowing, the more naive they seemed to be about the ease with which relevant Internet-based sources could be identified and used, also displaying an overreliance on the Internet as a communication tool and overestimating the value of

virtual exchanges at the expense of real-life encounters. Given the importance of being able to construct integrated meaning from multiple textual sources in today's knowledge society, with those sources more often than not located in complex computerized information systems, further research on the role played by personal epistemology in such endeavors is greatly needed.

Thus far, fewer studies have addressed relations between personal epistemology and academic motivation than between personal epistemology and academic cognition and performance. However, Bråten and Strømsø (2004, 2005) reported that naive epistemic beliefs were negatively related to adaptive motivational beliefs such as mastery goal orientation, self-efficacy, and interest. Conversely, Buehl and Alexander (2005) identified subgroups based on epistemic belief profiles and showed that clusters characterized by more sophisticated patterns of epistemic beliefs had higher levels on both expectancy and value components of academic motivation.

Although there is a fairly solid research base for asserting the importance of personal epistemology for student motivation, cognition, and performance, the complexity of the relationships among those constructs are not very well understood. As Schraw (2001) noted, structural models that specify the direct and indirect linkages among academic constructs, including epistemic beliefs, should be generated and then tested empirically through structural equation modeling to examine such complex relationships collectively.

How Can Personal Epistemology Be Assessed?

Perry (1970) and other later researchers primarily interested in the development of personal epistemology (e.g., Baxter Magolda, 1992; King and Kitchener, 1994) mainly conducted lengthy in-depth interviews to provide thick descriptions of how individuals' beliefs about knowledge and knowing change over time. For example, King and Kitchener (1994) have conducted structured interviews about ill-structured real-life problems or dilemmas, for example, about news or food additives, asking participants about their views on such problems and how they would justify their views. The interview tool used by these researchers is called the reflective judgment interview, with research providing evidence for the validity and reliability of this tool.

However, the possibility for doing larger-scale investigations of personal epistemology greatly increased with Schommer's (1990) introduction of a questionnaire allowing for group administration and statistical analyses of student scores. Schommer's epistemological beliefs questionnaire assesses personal epistemology at a domain-general level. Modifications of the questionnaire have

tried to improve its psychometric qualities. However, the dimensionality of personal epistemology has been somewhat different across the modifications, and the internal consistency reliabilities (Cronbach's Alphas) for the dimensions measured with those instruments have sometimes been smaller than required. Still, Schommer's questionnaire has been the most widely used quantitative assessment of personal epistemology.

Research on the domain-general versus domain-specificity issue eventually led to the construction of domain-specific personal epistemology questionnaires. First, a modified version of Schommer's domain-general questionnaire was developed where about every third item explicitly mentioned the domain (e.g., mathematics or science) that the student should keep in mind when completing the questionnaire (Schommer and Walker, 1995). However, Hofer (2000) constructed a measure more specifically devised to assess domain-specific epistemic beliefs, where each item on the questionnaire referred to a particular field or subject matter (psychology or science) as a frame of reference (e.g., in this field, knowledge is certain.). Likewise, Buehl *et al.* (2002) developed a self-report measure specifically devised to test for domain-specific epistemic beliefs, with the items focusing on either mathematics or history. In their more recent work, Buehl and Alexander (2005) have selected and combined items from the two above-mentioned domain-specific measures, the discipline-focused epistemological belief questionnaire (DFEBQ; Hofer, 2000) and the domain-specific beliefs questionnaire (DSBQ; Buehl *et al.*, 2002). Buehl and Alexander (2005) reported that the resulting domain-specific questionnaire captured the dimensions of certainty, simplicity, and source of knowledge in mathematics as well as in history.

As existing measures of personal epistemology primarily focused on conventional-print environments rather than new technological environments, Bråten *et al.* (2005) designed the Internet-specific epistemological questionnaire (ISEQ), a questionnaire specifically assessing beliefs about Internet-based knowledge (what one believes knowledge is like on the Internet) and Internet-based knowing (how one comes to know on the Internet), with this questionnaire especially suitable for studying relations between personal epistemology and learning with hypermedia or Internet technologies. As mentioned above, some researchers have also started to construct questionnaires that assess epistemic beliefs at a topic-specific level, for example, concerning specific theories within domains (Trautwein and Lüdtke, 2007) or specific scientific topics (Bråten *et al.*, in press; Stahl and Bromme, 2007).

Assessment of personal epistemology through survey methodology relies heavily on Likert-type rating scales where individuals express their degree of agreement with beliefs about knowledge and knowing. One issue with such scales is whether they can capture the complex and multifaceted nature of personal epistemology dimensions,

particularly the source of knowledge and justification of knowing dimensions (Hofer, 2004b). Another issue is whether they are suitable for measuring all the different epistemic positions identified within the developmental approach (i.e., dualistic, multiplistic, and evaluativistic positions) (Hofer, 2004b; see also, Muis *et al.*, 2006). Given such limitations, there has been a call for more qualitative, dynamic assessments of personal epistemology. One such approach, used by Hofer (2004b), involves having students think aloud during actual learning and knowledge construction and then analyzing the think-aloud protocols for instances of epistemic thinking (see also, Mason and Boldrin, 2008). Another possibility, also used by Hofer (2004a), is to use ongoing observations of classroom discourse combined with student interviews to examine how students' personal epistemology unfolds over time in the context of subject-matter instruction.

Finally, multi-method approaches combining quantitative and qualitative data sources may be especially valuable when assessing personal epistemology (e.g., Hofer, 2006; Schraw, 2001). For example, combining the use of questionnaires with in-depth interviewing may not only lead to a refinement of existing questionnaires (Hofer, 2006), but also allow for a triangulation of data that can give both researchers and educators a more complete picture of student epistemic beliefs.

Educational Implications

Schraw (2001) noted that one educational implication that might follow from the existing research on personal epistemology is that teachers should be helped to understand and change their own epistemic beliefs. According to Schraw (2001), teachers' beliefs affect their curricular and pedagogical decisions, and such decisions may, in turn, affect student epistemic beliefs (see also, Schraw and Olafson, 2003). This was partly confirmed in a qualitative study where Hofer (2004a) combined observations and interviews in two versions of introductory-level college chemistry. Hofer (2004a) observed dramatic differences in pedagogical approaches and implicit messages about knowledge and knowing in the two courses that she studied, also showing that students' existing epistemic beliefs were influenced by the instruction they experienced. Hammer and Elby (2002) suggested that innovative pedagogical approaches where class discussion is more typical than lectures, and where students are engaged in activities of design and construction to accomplish authentic tasks, are more likely than traditional pedagogical approaches to activate sets of epistemic resources that are productive for learning.

According to Muis *et al.* (2006), however, the dominant epistemologies of educational domains (e.g., of science or history education) often seem to reinforce the beliefs that

knowledge consists of right answers and unquestionable facts possessed by authorities and transmitted to students. In history, for example, students typically engage in the gathering of factual information about different topics, without much evaluation or questioning of the validity of that information. This concentration on accumulating historical facts may be reinforced by teacher beliefs in knowledge and knowing in history as the memorization and reproduction of factual information (cf., VanSledright, 2002). Instead, students need to be taught that there may be multiple opinions about historical events, with these opinions backed by varying evidence. Moreover, students need help to understand that interpretations of historical events can be justified by the amount of evidence that they account for, and that all available sources should be considered, not only a few select ones (Wolfe and Goldman, 2005).

Thus, given that teachers' epistemic beliefs in many instances seem to be less sophisticated than desirable, it is indeed an important task to try to promote belief change in teacher students as well as in more- and less-experienced teachers. Preliminary evidence suggests that this is not an impossible task. For example, Gill *et al.* (2004) showed that teachers may revise their existing epistemic beliefs through the reading of refutational text especially designed for this purpose.

Regarding the development of more sophisticated epistemic beliefs in students, Schraw (2001) suggested that schools should encourage an ongoing discussion and evaluation of such beliefs. In particular, Schraw argued that schools should try to promote critical-thinking skills and conceptual change among students, for example, by encouraging cooperative learning where students can discuss and evaluate their own epistemic beliefs. There is currently some research to underpin these suggestions, indicating that having students struggle to understand complex issues by reading texts presenting them with multiple perspectives on a topic, integrated with discussions of both text content and their current epistemic thinking, may bring about belief change. Thus, Valanides and Angeli (2005) observed that students who read a text presenting opposing views on a controversial topic and then discussed the text content, reflected on their thinking about the issue, and evaluated their thinking in light of principles for critical thinking, developed more sophisticated epistemic beliefs after the intervention. Accordingly, Kienhues *et al.* (2008) found that students who held naive epistemic beliefs concerning the scientific topic of genetics considered knowledge about this topic to be more complex and variable after reading a text focusing on the uncertainties of genetic fingerprinting. Bråten (2008) argued that the reading of multiple texts containing contrasting perspectives on a topic would be a good starting point for reflection on both content and epistemic beliefs in relation to that content, with such

reading and concomitant collective reflection presumably having the potential to foster the belief revision and conceptual change that many students seem to need.

Since personal epistemology seems to vary across domains, helping students become aware of their own epistemic beliefs, as well as develop more sophisticated beliefs, should probably take place within the frameworks of particular domains or even topics within domains. That is, an important part of instruction in a domain should concentrate on challenging students' existing beliefs about knowledge and knowing in the domain and, moreover, help them develop more sophisticated, expert-like beliefs about the nature of knowledge and the processes or methods of knowing within that domain. According to Hofer (2006), helping students develop from multiplicity to evaluativism is a particularly difficult instructional task in postmodern educational environments.

See also: Knowledge Domains and Domain Learning; Learning as Inquiry; Learning from Multiple Information Sources; Learning Strategies; Metacognition; Problem Solving and Human Expertise; Self-Regulated Learning and Socio-Cognitive Theory.

Bibliography

- Alexander, P. A., Schallert, D. L., and Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research* **61**, 315–343.
- Baxter Magolda, M. B. (1992). *Knowing and Reasoning in College: Gender-Related Patterns in Students' Intellectual Development*. San Francisco, CA: Jossey-Bass.
- Bendixen, L. D. and Hartley, K. (2003). Successful learning with hypermedia: The role of epistemological beliefs and metacognitive awareness. *Journal of Educational Computing Research* **28**, 15–30.
- Bråten, I. (2008). Personal epistemology, understanding of multiple texts, and learning within internet technologies. In Khine, M. S. (ed.) *Knowing, Knowledge, and Beliefs: Epistemological Studies across Diverse Cultures*, pp 351–376. New York: Springer.
- Bråten I., Gil, L., Strømso, H. I., and Vidal-Abarca, E. (in press). Personal epistemology across cultures: Exploring Norwegian and Spanish university students' epistemic beliefs about climate change. *Social Psychology of Education*.
- Bråten, I. and Strømso, H. I. (2004). Epistemological beliefs and implicit theories of intelligence as predictors of achievement goals. *Contemporary Educational Psychology* **29**, 371–388.
- Bråten, I. and Strømso, H. I. (2005). The relationship between epistemological beliefs, implicit theories of intelligence, and self-regulated learning among Norwegian post-secondary students. *British Journal of Educational Psychology* **75**, 539–565.
- Bråten, I. and Strømso, H. I. (2006). Effects of personal epistemology on the understanding of multiple texts. *Reading Psychology* **27**, 457–484.
- Bråten, I., Strømso, H. I., and Samuelstuen, M. S. (2005). The relationship between internet-specific epistemological beliefs and learning within internet technologies. *Journal of Educational Computing Research* **33**, 141–171.
- Buehl, M. M. and Alexander, P. A. (2001). Beliefs about academic knowledge. *Educational Psychology Review* **13**, 385–418.
- Buehl, M. M. and Alexander, P. A. (2005). Motivation and performance differences in students' domain-specific epistemological belief profiles. *American Educational Research Journal* **42**, 697–726.
- Buehl, M. M., Alexander, P. A., and Murphy, P. K. (2002). Beliefs about schooled knowledge: Domain specific or domain general? *Contemporary Educational Psychology* **27**, 415–449.
- Burr, J. E. and Hofer, B. K. (2002). Personal epistemology and theory of mind: Deciphering young children's beliefs about knowledge and knowing. *New Ideas in Psychology* **20**, 199–224.
- Gill, M. G., Ashton, P., and Algina, J. (2004). Authoritative schools: A test of a model to resolve the school effectiveness debate. *Contemporary Educational Psychology* **29**, 389–409.
- Hallett, D., Chandler, M. J., and Krettenauer, T. (2002). Disentangling the course of epistemic development: Parsing knowledge by epistemic content. *New Ideas in Psychology* **20**, 285–307.
- Hammer, D. and Elby, A. (2002). On the form of a personal epistemology. In Hofer, B. K. and Pintrich, P. R. (eds.) *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*, pp 169–190. Mahwah, NJ: Erlbaum.
- Hofer, B. K. (2000). Dimensionality and disciplinary differences in personal epistemology. *Contemporary Educational Psychology* **25**, 378–405.
- Hofer, B. K. (2004a). Exploring the dimensions of personal epistemology in differing classroom contexts: Student interpretations during the first year of college. *Contemporary Educational Psychology* **29**, 129–163.
- Hofer, B. K. (2004b). Epistemological understanding as a metacognitive process: Thinking aloud during online searching. *Educational Psychologist* **39**, 43–55.
- Hofer, B. K. (2006). Beliefs about knowledge and knowing: Integrating domain specificity and domain generality: A response to Muis, Bendixen, and Haerle (2006). *Educational Psychology Review* **18**, 67–76.
- Hofer, B. K. and Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research* **67**, 88–140.
- Hofer, B. K. and Pintrich, P. R. (eds.) (2002). *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*. Mahwah, NJ: Erlbaum.
- Jacobson, M. J. and Spiro, R. J. (1995). Hypertext learning environments, cognitive flexibility, and the transfer of complex knowledge: An empirical investigation. *Journal of Educational Computing Research* **12**, 301–333.
- Kienhues, D., Bromme, R., and Stahl, E. (2008). Changing epistemological beliefs: The unexpected impact of a short-term intervention. *British Journal of Educational Psychology* **78**, 545–565.
- King, P. M. and Kitchener, K. S. (1994). *Developing Reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults*. San Francisco, CA: Jossey-Bass.
- Kitchener, R. F. (2002). Folk epistemology: An introduction. *New Ideas in Psychology* **20**, 89–105.
- Kuhn, D. (1991). *The Skills of Argument*. Cambridge: Cambridge University Press.
- Mason, L. and Boldrin, A. (2008). Epistemic metacognition in the context of information searching on the web. In Khine, M. S. (ed.) *Knowing, Knowledge, and Beliefs: Epistemological Studies across Diverse Cultures*, pp 377–404. New York: Springer.
- Muis, K. R. (2007). The role of epistemic beliefs in self-regulated learning. *Educational Psychologist* **42**, 173–190.
- Muis, K. R., Bendixen, L. D., and Haerle, F. C. (2006). Domain-generality and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educational Psychology Review* **18**, 3–54.
- Perry, W. G. (1970). *Forms of Intellectual and Ethical Development in the College Years: A Scheme*. New York: Holt, Rinehart, and Winston.
- Qian, G. and Alvermann, D. (1995). Role of epistemological beliefs and learned helplessness in secondary school students' learning science concepts from text. *Journal of Educational Psychology* **87**, 282–292.
- Rukavina, I. and Daneman, M. (1996). Integration and its effect on acquiring knowledge about competing scientific theories from text. *Journal of Educational Psychology* **88**, 272–287.
- Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology* **82**, 498–504.
- Schommer, M. and Walker, K. (1995). Are epistemological beliefs similar across domains? *Journal of Educational Psychology* **87**, 424–432.

- Schraw, G. (2001). Current themes and future directions in epistemological research: A commentary. *Educational Psychology Review* **13**, 451–464.
- Schraw, G. and Olafson, L. (2003). Teachers' epistemological world views and educational practices. *Issues in Education* **8**, 99–148.
- Spiro, R. J., Feltovich, P. J., and Coulson, R. L. (1996). Two epistemic world-views: Prefigurative schemas and learning in complex domains. *Applied Cognitive Psychology* **10**, S51–S61.
- Stahl, E. and Bromme, R. (2007). The CAEB: An instrument for measuring connotative aspects of epistemological beliefs. *Learning and Instruction* **17**, 773–785.
- Trautwein, U. and Lüdtke, O. (2007). Predicting global and topic-specific certainty beliefs: Domain-specificity and the role of the academic environment. *British Journal of Educational Psychology* **77**, 907–934.
- Valanides, N. and Angeli, C. (2005). Effects of instruction on changes in epistemological beliefs. *Contemporary Educational Psychology* **30**, 314–330.
- VanSledright, B. (2002). *In Search of America's Past: Learning to Read History in Elementary School*. New York: Teachers College Press.
- Wolfe, M. B. W. and Goldman, S. R. (2005). Relations between adolescents' text processing and reasoning. *Cognition and Instruction* **23**, 467–502.
- Wood, P. K. and Kardash, C. (2002). Critical elements in the design and analysis of studies of epistemology. In Hofer, B. K. and Pintrich, P. R. (eds.) *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*, pp 231–260. Mahwah, NJ: Erlbaum.

Further Reading

- Khine, M. S. (ed.) (2008). *Knowing, Knowledge, and Beliefs: Epistemological Studies across Diverse Cultures*. New York: Springer.

Piaget: Recent Work

L Smith, Freelance Researcher, Lake District, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Constructivism – Problems about knowledge and reality are twofold. One problem is in metaphysics (sometimes called ontology): what is reality like? The other problem is in epistemology: how is reality known? These problems have led to three main answers – realism, nominalism, and constructivism. To see what is at issue, take a particular case about the nature and knowledge of number. The ontological problem is: what are numbers, really? One answer is realism (also called Platonism after its first sponsor): really, numbers are abstract objects with properties independent of any knower. A second answer is nominalism: really, numbers are numerals for manipulation within some mathematical language. A third answer is constructivism: really, numbers are constructions guided by norms and causes in human actions and operations. The epistemological problem follows on: how are numbers known? Realists identify a human faculty, such as intuition, by whose use numbers as abstract objects can be grasped. Nominalists identify some aspect of language that enables language users to manipulate numerals. Constructivists identify human capabilities that forge meaningful connections in learning and knowing. Standard objections also follow on. Realism seems to lead to an ontological slum of infinitely many possible worlds actually unknowable by our lights. Nominalism seems to lead to a relativism without objectivity. Constructivism seems to lead to an inherent subjectivity that precludes objectivity – if all knowledge is relative to its level of construction, by reference to which level is any knowledge objective? On the positive side, all three answers have current sponsors.

Frege's rule – The rule is usually referred to as the fallacy of psychologism and has its basis in the work of Gottlob Frege, the founder of modern logic. There is an ambiguity in the notion 'laws of thought'. If these laws are laws of logic, they are true independently of being acknowledged. If these laws are psychological laws, they are true in one context or culture, but not in others. Frege's rule is that conflating this distinction amounts to psychologism, and that psychologism is a fallacy. Although psychologism is widely accepted to be a fallacy, it does not entail a division of labor with no joint work in the formal and empirical sciences.

Nonpsychologism is an alternative under which complementary work is possible without lapsing into psychologism.

Hume's rule – The rule, so-called after its author, David Hume in his *Treatise of Human Nature*, states that the is-ought relation is not entailment: the same facts are compatible with contrary 'norms' about what ought to be or what ought to be done (*q.v.* norms). By contrast, any deductively valid argument is such that contrary conclusions could not be entailed by the same premises.

Kant on the formation of knowledge – Immanuel Kant in his *Critique of Pure Reason* identified two questions about concept formation and other forms of human rationality. Questions about the origins are factual (*quid facti*), whereas questions about constitution are normative (*quid iuri*) – the right concepts are required to make true as opposed to false judgments. For Kant, an adequate account should deal with both since the many forms of pseudo-rationality have an origin and yet are fallacious in leading to error.

Norms – The term 'norm', and cognate terms such as 'normativity' have a variety of uses in biology (pattern of phenotypic expression), mathematics (function in a vector space), sociology (normality), and statistics (average). These uses are all valid. However, they differ from a norm-laden use central in making right/wrong judgments. Norms in this norm-laden sense include rules, imperatives, directives, customs, obligations, and ideals. While moral norms are norms, not all norms are moral. Norms operate in mathematics, " $3 + 4 = 7$, so $3 \times 4 = 12$. Yes, that's right," and in human action and thought generally due to their implications for what has to be done and what has to be.

Normative facts – Although a norm is not a fact, not all facts are causal facts. Normative facts are facts occurring in an individual's commitment to a norm operative in action or thought irrespective of its explicit recognition. Recognition of a norm involves some degree of conscious knowledge. The operation of a norm is extrinsic when a person conforms to others' norms due to the sociocultural causality of parents, peers, or teachers. Extrinsic conformity in this sense is normative pressure. The operation of a norm is intrinsic when a person makes a commitment

to a norm on his or her own account, whether to retain the same norm, revise it, or to recast it. Intrinsic commitment is normative autonomy, a potentially powerful source of novelty and creativity.

Peirce's semiotics – Charles Sanders Peirce elaborated a distinctive semiotic theory, that is, theory of meaningful signs, according to which all meaning is triadic. The triad consists of a sign-vehicle, object, and interpretant. It is thereby incompatible with any dyadic theory of meaning, for example, meaning is a relation between a word and object. In a triadic theory, things other than words have meanings, for example, a metal arrow on a pole can depict wind direction, its object. The metal is a physical object in the causal world, and it depicts only in terms of an interpretant that is itself capable of standing in triadic relations of its own. For Peirce, a person serves as an interpretant. In Piaget's account, frameworks (schemes, structures, systems) are interpretants.

The Story So Far

Jean Piaget (1896–1980) made the seminal contribution to the study of human development in the twentieth century – *Time* included him in the 100 most important persons ranking 4th in psychology and 77th in science, generally. Despite this, Piaget's work has led to contrary interpretations and evaluations.

Under the standard interpretation of Piaget's work, human development is a sequence of four stages by analogy with a ladder or staircase with four steps. These stages are universal: the same stages fit all human development in the same order irrespective of differences in cultures and contexts. Yet, there is massive evidence to show two things. One is that heterogeneity – not the predicted homogeneity – is characteristic of human development. The other is that alternative accounts can explain these differences that are not readily explained in Piaget's program. Further, the advance from one stage to the next is due to a special mechanism – equilibration – or the change from one level of equilibrium to a better level. However, this mechanism is reckoned to be obscure, resisting clarification, and so stage-like advance is a mystery, apparently downsizing, even eliminating, the contribution of nature and nurture. A twofold conclusion is usually drawn that Piaget's work was historically important in identifying fundamental questions, and that its answers have now been superseded by better accounts (Scholnick *et al.*, 1999).

Running contrary to this is the constructivist interpretation of Piaget's work under which human development

is a hierarchical series of levels of construction with neither a first nor last level (*construction sans fin, régression sans fin*) (Smith, 2009a). The implication is that developmental stages are merely some of these indefinitely many levels. A principal feature of this interpretation is that Piaget's work is necessarily interdisciplinary in that its epistemology, psychology, and pedagogy are interdependent parts for systematic, not piecemeal, analysis. The conclusion typically drawn is that Piaget's work has addressed questions of human knowledge that are both fundamental and unresolved. Actually, Piaget's answers are often argued to be superior to available alternatives (Müller *et al.*, 2009).

Piaget was referring to the standard interpretation when he made an ironic remark about the recognition that his work had received: "I'm pleased by it, of course, but it's pretty catastrophic when I see how I'm understood" (Bringuier, 1980: 54). This remark motivates the stance taken in this article which is in three parts:

- Recent work in education.
- Piaget's constructivist program.
- Ways forward.

Recent Work in Education

Most educational research on Piaget's work is based on the standard interpretation under which Piagetian stages seem to be biological constraints that preclude instructional interventions, thereby making Piagetian pedagogy a nonstarter. Two adverse consequences follow. One is that a Piagetian child is viewed as a solitary knower for whom teaching is nonrelevant. The other is a preoccupation with the American question: can progress through Piagetian stages be improved? However, viewed under the constructivist interpretation, both consequences are flawed, and Piagetian pedagogy is promisingly positive.

Piaget's Pedagogy

Piaget (1998; commentary on this French text is available in Smith, 2009b) frequently noted that educational research required evidence about instructional interventions, a precondition he accepted he had not met in his own work. However, two qualifications should also be noticed. One is that others have remedied this omission leading to strikingly successful results – more on this below. The other is that Piaget's program did include principles for formative assessment, that is, teaching-based assistance with a view to making learning possible, easier, or better. Four principles elaborate key aspects of this assistance in Piaget's pedagogy:

1. *Children have to be taught to think.* "Each individual is led to think and re-think the system of collective notions"

(Piaget, 1995: 76). Note the use of ‘led’: any individual requires social assistance in being led to think with a view to any re-think in the sequel. For Piaget (1998: 163) “children have to be taught to think. . . To think is to search for oneself, that is, to criticize freely, and to demonstrate autonomously.” Learning without teaching has a tendency to lapse into the twin pitfalls of anomie and anarchy, alienation and willfulness. Neither amounts to individual autonomy which is quite another matter.

2. *Teaching creates contexts for learning in which learners use their judgment.* The role of the teacher is to design contexts that facilitate the formation of better learning (Piaget, 1998: 190–91). The difference between creating learning and creating a context for learning is an important difference. The teacher’s role is not to make learners think – that can amount to surface learning – but rather to create conditions in which learners make their own judgments about what they think in the light of standards of coherence and objectivity. Making the right judgments in human experience, generally, is an outstanding issue for us all.
3. *Good learning is collaborative.* Indeed, it “necessarily presupposes collaboration in work” (Piaget, 1998: 46). This principle is explicitly stated and its basis lies squarely in Piaget’s epistemology: “human knowledge is essentially collective, and social life constitutes an essential factor in the creation and growth of knowledge, both pre-scientific and scientific” (Piaget, 1995: 30). It is patently incompatible with the standard interpretation that a Piagetian child is a solitary knower. Making the right judgments is no easy matter, and access to others’ judgments is indispensable.
4. *Good learning requires self-government.* “The method of self-government consists in attributing to pupils a share in the responsibility for scholarly discipline” (Piaget, 1998: 167). Self-government is polymorphic in assuming different forms, including national and international collaborations based on freedom of thought and speech, institutional control with individual and collective responsibilities, personal development notably with regard to rules and directives, and self-discipline in academic work.

These principles are incompatible with the standard interpretation; they are, however, a direct consequence of the constructivist interpretation of Piaget’s program.

Recent Research on Assessment and Intervention

National assessment based on adaptations of classic Piagetian tasks has proved to be highly reliable in two respects. First, the evidence from a major British study in the 1970s (Shayer and Adey, 2002) has stood up well in

recent international studies (Müller *et al.*, 2009). Second, the same Piagetian tasks from the 1970s were used three decades later with different cohorts and different curricula, revealing a substantial decrease in intelligence, that is, anti-Flynn effects or measured decreases in intelligence over generations (Shayer, 2008). This novel finding is significant: there are no other standardized measures applicable to different cohorts other than psychometric tests that have revealed a progressive increase in intelligence over generations, that is, Flynn effects.

Scientific assessment requires the use of an objective method. Piaget’s method – a critical method – has a distinctive rationale for entering the child’s mind by teachers and investigators (Ginsburg, 1997). This method with its intra-individual function is indispensable on two counts. One is to monitor spontaneity, such as individual differences in children’s judgments in their own right, not as mere noise due to causal variance. The other is to check on the robustness of a particular individual’s reasoning through a train of thought. This method continues to be indispensable on a diverse range of non-Piagetian tasks. It complements quantitative methods with an inter-individual function directed on the frequency of prespecified responses in a population. There is no inherent incompatibility in the joint use of complementary methods. One reason is the need for dialog between investigators in the human sciences in view of the dual requirements of measurement and meaning (Dawson *et al.*, 2006). Another is based on educational practitioners’ beliefs about reconciling the heterogeneity of educational problems and the objectivity of science (Niaz, 2008).

A series of interventions were designed to improve children’s understanding under classroom conditions (Shayer, 2008; Shayer and Adey, 2002). Piaget’s account of children’s reasoning through serial levels was used as a measure of the population baseline. Each intervention had several parts, including Piagetian tasks adapted for classroom use; pedagogical focus on understanding; teacher training in their use; designated classes 1 h weekly or fortnightly over 1–2 school years; subject-specific delivery with adolescents, but nonspecific with children; collaborative learning with individual reflection; and a quasi-experimental design with the control of school differences. The first intervention covered learners in science in grades 7–8 (aged 11–13 years) over 2 years with an outcome measured in externally organized, national examinations. The level of improvement of youngsters in the experimental classrooms over the control classrooms was almost double for good passes in three core subjects (Science, Mathematics, and English). The second intervention was a replication that resulted in a comparable level of improvement. An ongoing intervention dealt with mathematics in grades 1–2 (aged 5–7 years). The success rate of children in control classes on pre- and post-test tasks was substantially lower than that of children in the experimental classes.

This research evidence shows that Piaget's program works well under classroom conditions. It is incompatible with the standard interpretation in view of its constructivist rationale.

Piaget's Constructivist Program

Piaget (2006) called his program genetic epistemology (*épistémologie génétique*). Today, such a name is liable to be misleading. His program is not about genetics and DNA. Developmental epistemology is an alternative name for this program whose central aim is to explain the formation of knowledge in terms of its human development. Four main notions in this program are now analyzed and illustrated: formation of knowledge, norms, actions, framework, and reasons.

Formation of Knowledge

Kant distinguished two questions about the formation of knowledge (Smith, 2009a):

1. What is its origin (*quid facti*)?
2. What is its constitution (*quid iuris*)?

Objective knowledge has an origin in the human mind, but so too do the many forms of pseudo-rationality from astrology to zoism. Establishing the origin of something leaves open its qualities – are these the right ones? An appellation on a wine label identifies where the grapes were grown, but not whether the wine in the bottle is actually at the right standard. Kant's point was that human minds have the right intellectual tools for the acquisition of properly constituted knowledge. These tools have to be rich enough to capture the necessity of some types of knowledge. His famous example was that $7 + 5 = 12$ is necessarily true, and so could not be otherwise. Kant's conclusion was that the capacity to make necessary judgments is fundamental to the human mind whose proper functioning requires *a priori* knowledge based on norms that are neither innate nor learned, but are instead independent of experience.

The insight behind Kant's distinction is integral to the constructivist interpretation of Piaget's program; it is lost in

the standard interpretation due to its sole focus on (1) the origin of knowledge, without regard for (2) its constitution by the knower. However, Piaget modified Kant's conclusion in two ways, notably about the formation of necessary knowledge. First, necessary knowledge is not innately present in the human mind. But nor is it straightforwardly acquired through learning – seven drops of water added to five drops can make one pool. Young children are prone to conflate number with other objects: 2×3 objects added to one box are less than 3×2 added to another (Piaget, 2001: 57). For Piaget, *a priori* knowledge is not fully formed at the lower levels of human development, even if its formation is completed later. Second, Kant's implied division of labor – facts in (1) for psychology, norms in (2) for epistemology – is replaced by a linking science directed jointly on both. This linking science is Piaget's developmental epistemology whose focus is on the relations between norms and facts without committing the fallacy of psychologism, that is, without reducing norms to facts.

Norms

There are many varieties of norms including rules, commands, directives, customs, moral principles, and ideals (Smith, 2006). The function of norms is to lay down what is right, legitimate, appropriate, or proper, notably in making the right judgments not only in mathematics and morality but in human experience, generally. Norms have a binding force, the 'has to' of obligation in action and of necessity in thought – see the examples in **Figure 1**.

This binding force is manifest in human judgments based on norms. Norms are not causes even though normative commitments may have causal consequences. An instructive example is this (Brandom, 2000). An iron bar in wet conditions responds by rusting and turning red. Shown a rusty iron bar, a parrot may say "That is red." Yet neither the bar nor the parrot recognizes the incompatibility "So it can't be green," or the implication "But it must be colored." Inferential capabilities require norms for recognizing these necessitations. Norms are not directly observable, but become manifest as normative commitments made in their use, unwittingly or consciously. A commitment made through the causality of normative pressure can always be reaffirmed or revised in

Luther: "has to" in action	Spinoza: "has to" in thought
For they have contradicted each other. I cannot and I will not recant anything.	1, 2, and 3, therefore 6.
Here I stand, I cannot do otherwise.	The proportion in these numbers had to be so, and could not be otherwise.

Figure 1 Adults' norms. From Smith, L. (2003). Children's reasoning by mathematical induction: Causal facts and normative facts. *International Journal of Educational Research* 39, 719–742.

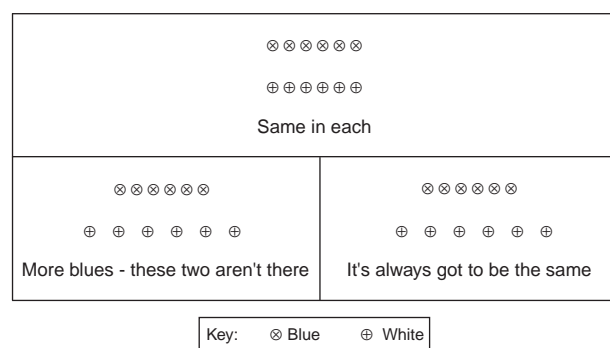


Figure 2 Children's norms. From Smith, L. (2002). *Reasoning by Mathematical Induction in Children's Arithmetic*. Oxford: Elsevier Pergamon Press.

an agent's autonomous commitments. For Piaget, a person is 'always norm laden' (*toujours 'normé'*) (Smith, 2006: 113). Normative capabilities are under development during childhood with an example in **Figure 2**. Two lines of blue or white buttons were presented in one-one correspondence with coincident endpoints, such that the children accepted that there is the same in each. After extending the white line, differences emerged. For some children, white outliers were disqualified from line-membership, and so were not counted – there were now “more Blues”. These children were using a personal norm – no doubt a pseudo-norm – to determine which buttons are still in the line, and so their judgments amount to normative misconception, not miscounting. For other children, necessary knowledge based on public norms was superbly on display about what has to be.

The use of a norm is a normative fact. Since facts are not norms, a normative fact is not the norm corresponding to it. And since norms are not causes, a normative fact is not a causal fact. However, any fact can be investigated empirically, including normative facts. Inferential reasons about entailments and implications, incompatibilities and contradictions, are paradigm examples of normative facts. That is why children's reasons for their judgments are always analyzed in Piaget's empirical studies. Piaget's constructivism deals both with norms in human development and with normative facts, that is, their actual use during childhood and the history of science.

Actions

An action is agent-initiated behavior and any living agent is always in action with objects in the world. Actions range from physical activities to acts of judgment, from sensorimotor activities to mental acts of thought and understanding, from knowing how to knowing that.

An action is an action in object selection (what is to be done with this object), not in object representation (which properties of the object are to be copied). For Piaget,

actions are not representations that copy or encode reality, and so his constructivism avoids the fatal objections that undermine the representational theories that are ubiquitous in psychology and education. Instead, a human action is always meaningful to its agent. An instructive parallel is with Peircean semiotics: a sign is meaningful because it stands in two relations, one to its object, the other to its interpretant. A metal arrow on a pole is a physical object in a causal world; it functions as a sign of wind direction only through some interpretative system. Lacking an interpretant, the arrow fails to comprehend the wind direction it is indicating. For Piaget, infants' actions have indicative, that is, indexical, meanings; symbolic and sign-based meanings emerge in later actions. This is because human agents are themselves interpretants and, for Piaget, an action framework is an interpretant.

Frameworks

The interpretant of agent's actions is a framework (*cadre*) or network of norms. Schemes and structure are types of framework. Every fact is registered through some framework. “Let us recall that, from the very beginning and even among our youngest subjects, a physical fact is recorded only within a logico-mathematical framework” (Piaget, 2001: 320). Since agents are always in action, frameworks are always in a state of Heraclitian flux during the agent's life, patterned through an agent's regulation. Changes due to regulation include: augmentation of new cases to existing norms; creating or annulling norms; recasting the relations between norms. Any change in a framework is recognized through the inferences made by their agent. These inferences are signifying, that is, meaningful implications, amounting to a logic that changes over time (Müller *et al.*, 2009).

Normative commitments are made both unwittingly and wittingly – see the examples in **Figure 2**. Either way, norms are continually being constructed and reconstructed in their use (*pratique*), that is, in the commitments made by a normative agent as advances to their conscious realization (*conscience*). These advances include conscious awareness and reflecting abstraction as aspects of equilibration (Müller *et al.*, 2009). A principal contribution to these advances is through inferential reasons.

Reasons

Piaget's early claim – “reason is a capacity born from action” – was recast in his final paper where the role of reasons is “to introduce new necessities into systems where they were merely implicit or remained unacknowledged” (2006: 8). Notice three things here. First, reasons may be trivial or profound and, either way, reasons may be learned or autonomous, that is, one's own reason due to human spontaneity. Following Kant, such is intellectual freedom (Piaget, 1998: 161). Second, reasons are never singletons

since the question “why?” is indefinitely iterative. Similarly, norms are never singletons, but always co-occur with other norms in networks. Third, reasons can embody inferences in being necessitating reasons, whether implicitly or explicitly. These inferential reasons are based on norms amenable to systematization in alternative logical models.

The plurality of norms and of reasons results in their complexity, and that is why the formation of knowledge takes time through serial levels of comprehension. Examples of children’s understanding of necessitating reasons are in **Table 1**.

An agent’s inferential reasons are norms in use (*pratique*) reflecting the norms in an operative framework. These reasons are gradually accessible both to their owner and to investigators (*conscience*). A science of the mind that ignores inferential reasons has no doubt given priority to measurement without due regard for meanings, notably the norm-laden meanings in the actions and acts of agents during their human development.

Piaget’s constructivism, then, looks like this. Human knowledge has its origin in action and its legitimation in action frameworks whose organization changes in their use through the life span. While factors in nature and nurture are always influential, there is a third factor operating interdependently with them. This third factor is equilibration, or the interplay between norms and their applications to specific cases by their self-regulating agents. Equilibration amounts to the formation of better levels of organization. Any level of organization currently in equilibrium always breaks down, that is, the balance between existing judgments is broken. No norm is a singleton with dilemmas arising when different norms point in contrary directions. No coherent network of norms can cover all possible cases, notably in rapidly changing modern worlds. Equilibration is the search – with success never written in – for a better level of equilibrium, whereby existing judgments are temporarily restored to balance when new judgments are coherently made. Inferential reasons are the accessible manifestations of epistemic advances, reasons that are accessible both to their agent and to investigators. Even the best scientific theories swim in a sea of anomalies; even sound systems of normative rules break down due to endless hard cases. Yet advances in the natural and human sciences does in fact take place. For Piaget, such advances are veritable human constructions. Central to Piaget’s developmental epistemology is the extent of the match between constructions in individuals and in science.

Ways Forward in the Future

In the twenty-first century, Piaget’s constructivist program embodies insights and evidence that merit and require attention, including reinterpretation and revision. Three ways forward are as follows:

Table 1 Varieties of necessitating reasons

^aNecessity denial

Switzerland has 26 cantons (counties) one of which is Vaud.

Florence was asked where she lived and her nationality.

JP What is your nationality?

FLO I’m from Vaud.

JP Are you also Swiss?

FLO You can’t be both at the same time, you have to choose.

^bPseudo necessity

The task was to make two collections of colored chips (towers) equal in number, taking chips 2 x 2 from one pile and 3 x 3 from another.

DOM (Initially denies that they ever could be equal) No, they’d have to be the same number (taken each time).

^cAd hoc necessity

Analogical reasoning requires two pairs of objects to be related in the same way, i.e., the relation between relations is identical.

Children were shown several drawings in a jumbled order. They were asked to place the drawings in pairs that go well together; then they were asked to bring together two pairs that go well together. Children who succeeded on the first comparison displaying a limited understanding of necessity did not always succeed on the second comparison.

CAN (Vacuum cleaner-socket) otherwise you can’t vacuum; (bird-feather) otherwise it can’t fly; and (dog-fur) otherwise he’ll be cold. (CAN does not succeed on any problem that requires relations among relations.)

^dNecessity

The task required the coordination of cardinal and ordinal numbers. Two boxes, one higher (H), the other lower (L), were connected by a tube and the child rolled a cardinal number of marbles down the tube. The question dealt with ordinal numbers. At the outset, H contained 5 marbles with 10 now in L.

JP Where is the 10th?

JOE In L.

JP How many are in H?

JOE Five.

JP How do you know?

JOE Because ten have been dropped and the sum is 15. (H = 5; L = 3)

JOE Three in L, five in H.

JP How did you do it?

JOE There are three of them in L, plus what is necessary to make eight. Therefore $8 - 3 = 5$

Sources

^aPiaget, J. (1995). *Sociological Studies*, ch. 7. London: Routledge.

^bPiaget, J. (2001). *Studies in Reflecting Abstraction*, ch. 2. Hove: Psychology Press.

^cPiaget, J. (2001). *Studies in Reflecting Abstraction*, ch. 2. Hove: Psychology Press.

^dPiaget, J. and Garcia, J. (1991). *Toward a Logic of Meanings*, ch. 4. Hillsdale, NJ: Erlbaum Associates.

1. In recent evaluations, Piaget’s account is regarded as distinctive and superior to alternatives. These evaluations provide pointers and directives about how to refine, recast, and re-evaluate explanations of the formation of coherent and objective knowledge in human development.
2. The normativity of inferential reasons has priority over causality in the explanation of human development.

Piaget's work always respected this order of priorities; the converse order is evident in most alternative accounts. A major and outstanding task is the identification of good specimens of the many forms of normativity, notably norm-laden meanings, norms in use, pseudo-norms, and novel norms.

3. Education consists in norm-laden exchanges between teachers and learners. These exchanges amount to paradigm cases of the interplay between norms and normative facts. Yet the normative aspects of assessment and intervention contexts have been given scant recognition in most accounts. Human agents have normative capabilities by whose use human creativity gains its expression. It is a fundamental error to confuse the regularities in nonliving things with the regulations of living agents with the potential to change the world.

See also: Personal Epistemology in Education: Concepts, Issues, and Implications.

Bibliography

- Brandom, R. (2000). *Articulating Reasons: An introduction to Inferentialism*. Cambridge, MA: Harvard University Press.
- Bringuier, J. -C. (1980). *Conversations with Jean Piaget*. Chicago, IL: University of Chicago Press.
- Dawson, T., Fischer, K., and Stein, Z. (2006). Reconsidering qualitative and quantitative research approaches: A cognitive developmental perspective. *New Ideas in Psychology* **24**, 229–239.
- Ginsburg, H. (1997). *Entering the Child's Mind*. Cambridge: Cambridge University Press.
- Müller, U., Carpendale, J., and Smith, L. (2009). *Cambridge Companion to Piaget*. New York: Cambridge University Press.
- Niaz, M. (2008). A rationale for mixed methods (integrative) research programmes in education. *Journal of Philosophy of Education* **42**, 287–305.
- Piaget, J. (1995). *Sociological Studies*. London: Routledge.
- Piaget, J. (1998). *De la pédagogie*. Paris: Odile Jacob.
- Piaget, J. (2001). *Studies in Reflecting Abstraction*. Hove: Psychology Press.
- Piaget, J. (2006). Reason. *New Ideas in Psychology* **24**, 1–29.
- Scholnick, E., Nelson, K., Gelman, S., and Miller, P. (1999). *Conceptual Development: Piaget's Legacy*. Mahwah, NJ: Erlbaum.
- Shayer, M. (2008). Intelligence for education: As described by Piaget and measured by psychometrics. *British Journal of Educational Psychology* **78**, 1–29.
- Shayer, M. and Adey, P. (2002). *Learning Intelligence*. Buckingham: Open University Press.
- Smith, L. (2006). Norms and normative facts in human development. In Smith, L. and Vonèche, J. (eds.) *Norms in Human Development*, pp 103–137. Cambridge: Cambridge University Press.
- Smith, L. (2009a). Piaget's developmental epistemology. In Müller, U., Carpendale, J., and Smith, L. (eds.) *Cambridge Companion to Piaget*, pp 64–93. New York: Cambridge University Press.
- Smith, L. (2009b). Piaget's pedagogy. In Müller, U., Carpendale, J., and Smith, L. (eds.) *Cambridge Companion to Piaget*, pp 324–343. New York: Cambridge University Press.
- ## Further Reading
- Adey, P. (2007). The CASE for a general factor in intelligence. In Roberts, M. J. (ed.) *Integrating the Mind: Domain General versus Domain Specific Processes in Higher Cognition*, pp 369–386. Hove: Psychology Press.
- Atkin, A. (2006). *Peirce's Theory of Signs*. <http://plato.stanford.edu/entries/peirce-semiotics> (accessed August 2009).
- Duckworth, E. (1996). *The Having of Wonderful Ideas: Essays on Teaching and Learning*, 2nd edn. New York: Teachers College Press.
- Feldman, D. H. (2004). Piaget's stages: The unfinished symphony of cognitive development. *New Ideas in Psychology* **22**, 175–274.
- Lourenço, O. and Machado, A. (1996). In defence of Piaget's theory: A reply to 10 common criticisms. *Psychological Review* **103**, 143–164.
- Neisser, U. (1998). *The Rising Curve: Long-Term Gains in IQ and Related Measures*. Washington, DC: American Psychological Association.
- Piaget, J. (1918). *Recherche*. Lausanne: La Concorde.
- Piaget, J. (1980). *Experiments in Contradiction*. Chicago, IL: University of Chicago Press.
- Piaget, J. (1985). *Equilibration of Cognitive Structures*. Chicago, IL: University of Chicago Press.
- Piaget, J. (1995). *Sociological Studies*. London: Routledge.
- Piaget, J. (1998). *De la pédagogie*. Paris: Odile Jacob.
- Piaget, J. (2001). *Studies in Reflecting Abstraction*. Hove: Psychology Press.
- Piaget, J. (2006). Reason. *New Ideas in Psychology* **24**, 1–29.
- Piaget, J. and Garcia, R. (1989). *Psychogenesis and the History of Science*. New York: Columbia University Press.
- Shayer, M., Ginsburg, D., and Coe, R. (2007). Thirty years on – a large anti-Flynn effect? The Piagetian test volume & heaviness norms 1975–2003. *British Journal of Educational Psychology* **77**, 25–41.
- Smith, L. (2002). *Reasoning by Mathematical Induction in Children's Arithmetic*. Oxford: Pergamon Press.
- Smith, L. (2006). Norms in human development: Introduction. In Smith, L. and Vonèche, J. (eds.) *Norms in Human Development*, pp 1–34. Cambridge: Cambridge University Press.
- Smith, L. (2009). Wittgenstein's rule-following paradox: How to resolve it with lessons for psychology. *New Ideas in Psychology* **27**, 228–242.
- Vonèche, J. (2006). The implicit normativity of developmental psychology. In Smith, S. and Vonèche, J. (eds.) *Norms in Human Development*, pp 35–56. Cambridge: Cambridge University Press.
- von Wright, G. H. (1963). *Norm and Action*. London: Routledge/Kegan Paul.
- Vuyk, R. (1981). *Overview and Critique of Piaget's Genetic Epistemology, 1965–1980* (2 vols.). London: Academic Press.

The Adult Development of Cognition and Learning

E A L Stine-Morrow and J M Parisi, University of Illinois at Urbana-Champaign, Champaign, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Cognition in adult development arises out of the dynamic interplay of (1) gains in knowledge-based systems, expertise, and skill and (2) losses in speed of processing, working memory (WM) capacity, and inhibitory control processes. These changes have important effects on the nature of learning and entail that strategies in instruction be developmentally sensitive throughout the life span. At the same time, education early in the life span shapes these lifelong trajectories, while education and engagement in work and leisure through adulthood play an important role in engendering and maintaining competencies. Our goal in this article is to examine this interaction between the adult development of cognition and education. In the following sections, we examine the nature of cognition through adulthood and consider implications for learning. We also discuss recent research suggesting lifelong plasticity in neural networks, so that education is intimately tied with successful aging.

Theories of Life-Span Development

We review learning and cognition in adulthood and aging against the backdrop of a long history in psychology and education in which adulthood has been viewed merely as the culmination of what was achieved in childhood followed by a biologically driven decline. To the contrary, contemporary theories and empirical findings suggest that there are dynamic changes in cognition, motivation, and regulatory heuristics throughout the life span.

Multidirectionality in Cognition and Intellectual Function

While there are a number of competing theories for mechanisms that underlie age-related changes in cognition, there is broad agreement that age effects on cognition can be characterized at a coarse level as the result of two competing forces. On the one hand, the senescence process drives a decline in mental mechanics, the capacity to control attention and perform basic mental computations with accuracy and speed. On the other hand, there is accumulating evidence that the brain has immense potential for plasticity into late life; therefore, experience-based articulation of knowledge systems, skill, and expertise offers potential for growth. These divergent trajectories have also been characterized in terms of fluid versus crystallized

abilities, derived from the factor structure of typical tests of intellectual ability. Fluid ability is manifested in tasks that require the encoding and transformation of information and detecting underlying relationships (e.g., block design, digit symbol, inductive reasoning, and spatial orientation). Crystallized ability, or pragmatics, depends on the acquisition of culture (e.g., vocabulary, semantic memory, and world knowledge).

Beyond that first cut, there are various theories characterizing the specific nature of declines in mechanics with age. The slowing hypothesis suggests that aging brings a systematic decrease in the speed with which mental operations can be performed. The WM hypothesis posits that aging is associated with a decline in the capacity to perform basic processing operations and store their products (as might be required, e.g., as one listens to a lecture and tries to construct an understanding of what is currently being discussed and integrate it with what has come before). According to the inhibition deficit hypothesis, aging brings a decreased effectiveness in suppressing irrelevant or no-longer-relevant information, thus reducing the functional capacity of WM. Another recent variant on the WM hypothesis is the effortfulness hypothesis, which identifies sensory loss as a critical contributor to age-related deficits in mental mechanics, not simply because they diminish the quality of information coming into the cognitive system, but also because central resources are strained by effort allocated to interpret the muddled input.

At the same time, aging may bring growth in a variety of capacities. Unlike declines, which come for free as a consequence of the senescence process, growth appears to arise out of dedicated engagement with particular types of experience. Initial conceptualizations of life-span growth focused on the normative development of capacities that could be scaled within a cultural context, for example, vocabulary and areas of common knowledge (e.g., why does bread rise? Who wrote *Huckleberry Finn*?), termed crystallized ability. In fact, vocabulary and some aspects of verbal ability are often found to increase or show stability through adulthood, with some evidence suggesting that such growth depends on habitual engagement with literacy-based activities. More recent conceptualizations acknowledge that normative growth of knowledge within certain cultural contexts early in the life span arises out of age-graded curricula in which there can be considerable commonality in the types of experiences to which children are exposed. However, relative to childhood, experiences in adulthood are more diverse and, rather than arising out of

common school curricula, are tied to occupational and leisure activities. The particularized knowledge and skill systems that develop during adulthood are key achievements during this period and are intimately tied to selective allocation of resources to particular activities. Skill systems can be maintained well into late life, with recent evidence highlighting an important role for deliberate practice.

These two dimensions of intellectual function interact in interesting ways. First, because knowledge growth and skill development require attentional control, expansion of crystallized ability, particularized knowledge, and expertise ultimately depend on fluid abilities to some extent. Consequently, early in the life span, fluid and crystallized abilities grow apace with each other. In midlife to early old age, however, these trajectories diverge; knowledge and skill can continue their upward trajectories, but at a shallower slope as it takes a greater investment in attention to achieve the same gains, relative to earlier in the life span. On the other hand, to the extent that an individual has already invested in creating knowledge structures and skill systems (regardless of age), subsequent growth in those areas can be accomplished more easily. In part, this is because of a greater efficiency in processing information when existing knowledge can support learning. Another factor that contributes to knowledge-driven learning is that knowledge can support attentional control (i.e., effective allocation of effort).

Selectivity

A hallmark of adult development is increased selectivity, the focus of effort on a subset of available options. While the earlier part of the life span is a period in which effort is allocated to hone a diverse repertoire of skills and to expand social networks, movement through adulthood brings a motivational shift toward selectivity. There are two developmental forces that drive this.

First, the decline in mental mechanics limits the resources available to promote growth; therefore, selective focus of effort will increase the likelihood that selected domains will thrive. This strategy of selective optimization is distinctively an adult skill, with evidence suggesting that well-being in later life is enhanced by mindful selection of domains to which effort will be allocated toward growth, with the acknowledgment that this will entail loss in unselected domains. Expertise and knowledge systems may enhance selectivity by enabling more efficient focus on the most relevant features of a situation.

Selectivity is also thought to be driven by a changing perspective on the temporal horizon across the life span. Time is perceived to be open ended in youth, with unlimited – and unknowable – possibilities for experience. This expansive temporal horizon engenders concerns with information acquisition, skill development, and enrichment of social networks as a strategy for preparing for an uncertain future. With movement through the life span, the temporal horizon comes into view. Time is not limitless

and choices have consequences that may squander that limited resource. This salience of the temporal horizon may increase attention to emotional concerns and decrease motivation toward purely cognitive goals. This theory has been used to explain, for example, why social networks typically decrease in size but become more emotionally satisfying. One implication of this socioemotional selectivity theory for learning is that cognition itself may be used most reliably in service of emotional goals; therefore, learning is expected to be relatively enhanced in later life if it is well integrated with the socioemotional system.

Self-Regulation

Another theme that emerges in adult developmental theory is the extent to which aging brings a change in the ability to engage processing resources for learning. While adult education textbooks often proclaim that adult learning is self-directed, the ability to self-initiate and self-direct learning may well depend on the level of existing knowledge and skill in the domain. In fact, one theory of cognitive aging holds that it becomes more difficult to self-initiate processing and that age-related difficulties in learning can be ameliorated to a large extent by the availability of environmental supports to guide processing.

Beliefs about one's capacity to accomplish cognitive and intellectual tasks appear to play a critical role in learning throughout the life span, but these may be particularly important in later adulthood and old age. Two related constructs that have received a lot of attention in the literature are self-efficacy and perceived control, each of which appears to be domain specific (e.g., one can conceive of being able to master health-related behaviors, but be undone by a cognitive task like filing a tax form). Self-efficacy is the confidence that one can execute the behavior or process necessary to achieve the desired outcome; perceived control is a multidimensional construct defining beliefs about the locus of control for achieving desired outcomes, with internal control reflecting beliefs of personal efficacy in achieving outcomes, and external control (chance or power others) reflecting beliefs that control of achievement rests with other sources. Relative to younger adults, older adults are sometimes found to have reduced levels of self-efficacy and may be more likely to believe that there are other people who will be able to control cognitive outcomes. To some extent, certain effects of aging are beyond our control; therefore, such beliefs may arise, in part, from an overgeneralization of a veridical perception. Another factor that may contribute to these beliefs is negative aging stereotypes, which can be internalized, and thus reduce effort to cognitive performance. In any case, such beliefs can become self-fulfilling prophecies. In fact, individual differences in self-efficacy and control beliefs can often account, in part, for age differences in cognitive performance, effects that are themselves sometimes mediated by strategy utilization.

Learning through Adulthood

Memory and Aging

Memory is among the first cognitive domains to be studied through adulthood and it remains a complex and vibrant literature. Memory is often conceptualized as involving three stages: encoding of information into a relatively durable trace, retention, and retrieval of the information into consciousness. Memory failure is among the most prevalent of aging stereotypes, and with aging, many adults complain about a difficulty with memory. In fact, age differences in memory performance can depend, to a large degree, on the task conditions, the materials, and the educational levels of the samples compared. Semantic memory (e.g., the meanings of words) and retrieval of well-learned information can show great resilience throughout the life span. By contrast, episodic memory, the ability to associate information with a particular learning context, may show pronounced age differences under some circumstances.

Deficits have been attributed, in part, to the demands that effective encoding places on mental mechanics and attentional control. Normatively, in an episodic memory task, older adults may be less likely to encode information in a way that is organized, elaborate, and distinctive. Critically, older learners have a more difficult time forming new arbitrary associations. Fairly modest interventions (e.g., instruction in organizational strategies) can improve memory performance. Interestingly, even though memory training can improve performance into very late life, training gains are typically greater among the young, suggesting that it is ultimately age-graded declines in mental mechanics that limit how well these strategies can be implemented.

Age changes in the effectiveness of episodic memory can also be attributed to retrieval. For example, age deficits in memory performance are often exaggerated in free recall relative to recognition or cued recall, a difference that presumably resides in the reduced demands to generate the information. In recognition, older adults are more likely to rely on familiarity rather than direct recollection.

Difficulty with name retrieval is one of the more common memory complaints. Aging does bring more tip-of-the-tongue experiences, in which an individual knows the word he/she wants to say but is unable to retrieve any of the phonological information. This is most likely to happen for relatively rare words or for infrequently encountered names. Happily, these are most typically resolved, even among older adults.

Learning from Text

Learning from text is often conceptualized as involving distinct processing systems that operate in concert to construct different facets of the language representation. At the surface level, individual lexical items (words) are encoded from the orthographic or acoustic signal and their meanings are activated. The semantic representation can be described

in terms of integrated ideas (or propositions) that establish relationships among concepts described by the text, a representation called the textbase. Knowledge plays a role in facilitating integration, enabling elaborative inference, and evoking a simulation of the situation suggested by the text. Consistent with the divergent age trajectories of mechanics and knowledge-based processes, age deficits are more likely for the resource-consuming aspects of language processing.

Understanding words

Vocabulary often shows an increase with age, particularly among those who are regularly exposed to text; therefore, word recognition and word-level comprehension appear to be highly resilient in reading. Visual-processing declines can impact reading rate, especially if the font is small or hard to decode. In speech processing, declines in auditory processing can make spoken word recognition more demanding; therefore, more acoustic information is needed to understand individual words. Such effects may not merely disrupt encoding of the surface form, but also tax WM resources that would otherwise be used to construct a representation of the text's meaning. For example, elders with normal and impaired hearing listening to a word list interrupted periodically to report the last word presented may show negligible differences; however, if asked to report the last three words, the hearing-impaired elders will show deficits. The explanation for such a provocative finding is that the hearing-impaired elders overcome a sensory loss at some attentional cost so as to exert a toll on semantic and elaborative processes that enhance memory. Presumably, the same mechanisms operate in ordinary language processing. At the same time, there is evidence that older adults can take differential advantage of context in the recognition of both spoken and written words, especially in noisy environments. One area of difficulty that older adults may have in word processing is in deriving the meaning of novel words from context, with research showing that older adults are likely to infer more generalized and imprecise meanings relative to the young, a difference that can be largely accounted for in terms of declines in mental mechanics.

Textbase processing

Older adults typically show poorer memory for the content expressed directly by the text. Processes used to construct the textbase (e.g., to instantiate and integrate concepts in the text, essentially an associative memory task) are among the most resource-consuming of those required in learning from text and are, hence, the most vulnerable to aging. When reading is self-paced, older adults require more time for effective propositional encoding (e.g., as indexed by effective reading time, the time allocated per idea unit recalled). In listening, when the pace is controlled by the speaker, older adults may have particular difficulty in understanding and retaining the information, especially as

informational density is increased or in noisy environments. Older adults appear to have no difficulty drawing anaphoric inference (i.e., correctly identifying the referent when the pronoun is used to refer to a noun that was introduced earlier) over short distances, but may find it difficult when the pronoun and referent are separated by intervening text. Thus, the general impression from this literature is that the semantic (textbase) representation is more fragmented and less distinctive as a function of aging.

There is an important exception to age declines in text memory that appears to derive from socioemotional selectivity. That is, memory for text may be very good if it is consonant with emotional goals or if the task is embedded in a social context. For example, it has been reported that emotional content in narratives (e.g., characters' emotional reactions) is well retained among older readers relative to emotion-neutral information. It has also been reported that age differences in narrative memory may be minimized if there is a social goal (e.g., tell a story to a child) as opposed to an information-acquisition goal (e.g., recall the text to an experimenter).

Situation model

Aside from deriving ideas directly from the text, learning from the text also involves elaboration on these ideas based on existing knowledge. Some theories focus on the perceptual quality of this level of representation, which gives rise to a perceptual simulation of the events described by the text. Therefore, for example, in narrative understanding, readers track goals and emotional reactions of characters as well as their movement through space and time. Behavioral methods to study this level of representation include probe recognition for objects in the narrative, as well as reading time, both of which show subtle effects of situation model processing. Readers are slower to verify the existence (in the narrative world) of objects that are spatially distant from the protagonist relative to those that are nearby. Readers also slow down when new characters are introduced, or when there is a spatial discontinuity (e.g., the locus of narrative events shifts from the village to the castle) or a temporal discontinuity (e.g., *The next day...*). When the text describes a goal to be achieved (e.g., Susan intends to buy her mother a purse for her birthday), the goal is activated in memory until it is achieved; therefore, concepts related to the goal are more quickly verified as long as the goal is open (e.g., purse will be more quickly verified if Susan can find the purse when she goes to the store relative to a condition in which she could not find one). To the extent that these paradigms have been used to explore adult age differences in situation model processing, there has been very little evidence of developmental differences in situation construction and updating; if anything, attentional allocation to situation construction may increase with aging. This is important because it suggests that the experiential aspects of reading and language understanding

(i.e., the phenomenal experience of entering the world described by the discourse) are resilient or even enhanced through adulthood. In addition, to the extent that perceptual simulation is required to understand expository or procedural text (e.g., how the heart works or how to put together a grill), the preservation of situation model processing may enable authentic learning from text – even if measures of explicit recall might suggest otherwise.

Older adults may particularly rely on the situation model to support textbase processing. For example, in ambiguous text (e.g., “The strength and flexibility of this equipment is remarkable. Not everyone is capable of using it even though most try at one point or another. . .”), older readers take differential advantage of titles (e.g., driving a car) that disambiguate the meaning to facilitate processing. Since the title renders the situation instantly transparent, both younger and older adults are more efficient in reading when it is available; however, older adults show this effect to a larger degree.

To the extent that the hallmark of situation model processing is an integration of textbase content with knowledge, one might expect that older adults would be particularly adept at inferential processing; however, this is not always the case. While older adults are more likely to draw elaborative inferences (e.g., in recall, to annotate their recollections with personal experiences or related information learned in another context), if inference is constrained so that it requires retrieval of textbase content, age deficits are the norm.

Discourse structures and context

Beyond sentence processing, different genres of text have characteristic forms. For example, narratives typically begin by introducing a setting and characters and proceed to describe a series of episodes in which goals or problems are introduced to be resolved, and so on. Expository texts have certain characteristic forms of argumentation (e.g., problem–solution and thesis–evidence). Older readers generally appear to track these larger discourse structures in the same way as the young. Adult readers have also been shown to benefit from explicit instruction in discourse forms to enhance memory and understanding.

Cognitive Reserve: Lifelong Effects of Education

Differential Developmental Trajectories as a Function of Education

A growing body of evidence suggests that there is a relationship between early educational experiences and cognitive development in adulthood. Numerous studies linking education and cognition have found that greater educational attainment is associated with higher levels of cognitive performance and lower risk for the development

of Alzheimer's disease. Although the mechanisms for education–cognition relationships remain unclear, four plausible explanations have been suggested, which may be operative individually or interactively. First, educational level may be a marker for innate levels of vitality or for capacities that are developed very early in the life span, such that those who are initially more able are also more likely to succeed in the educational system. Second, educational experiences early in the life span may expand neural networks, so as to create a lifelong cognitive reserve that enables relatively high levels of cognitive functioning even as the senescence process winnows neural connections later in life. Third, early educational experiences may be related to cognition through their association with socioeconomic status; therefore, it is socioeconomic advantage that enables lifelong conditions (e.g., nutrition, leisure that affords regular exercise, and medical care) that promote health and, thereby, successful cognitive aging. Finally, educational attainment may afford self-regulatory skill that promotes lifelong mental stimulation. Since educational experiences often lead to occupational, professional, and leisure experiences that provide intellectual challenge in domains in which one is invested, it may be that lifelong patterns of mental stimulation promote neural health, thereby engendering cognitive vitality. It is this latter possibility that is particularly exciting, and an important thrust of recent research in psychology, cognitive neuroscience, and education.

Cognitive and Neural Plasticity

In fact, there is growing evidence from animal models and human research that exposure to stimulating environments promotes neural growth and cognitive vitality. Studies that have administered cognitive training or practice sessions have revealed that the cognitive abilities of older adults show considerable plasticity. For example, the Advanced Cognitive Training for Independent and Vital Elderly (ACTIVE) trial is a randomized clinical trial to examine the effectiveness and durability of cognitive interventions on basic cognitive processes (memory, reasoning, and speed of processing). Results to date indicate that, in spite of the fact that each intervention has targeted a domain tapping into mental mechanics, training effects are ability specific (e.g., increased speed of processing does not transfer to better reasoning or memory). Interestingly, these effects have been shown to be durable up to 5 years. Such data not only provide evidence for the modifiability of cognitive abilities into late life, but also demonstrate that training effects may be highly selective.

On the other hand, certain conditions appear to enhance executive control (e.g., the ability to switch between two tasks) that may ultimately have the potential to affect a relatively wide array of activities. For example, aerobic exercise can increase executive function throughout

adulthood. Recent evidence suggests that language processing may impact this control function as well, with several recent demonstrations of enhanced executive function among fluent bilinguals who habitually manage two language systems.

There may be limits to plasticity very late in the life span, with some training studies showing reduced effects with increasing age. Such findings highlight the dynamic nature of change during adulthood, and also imply that cognitive vitality past the age of 85 or 90 years may depend on the cognitive reserve established up to that point.

Neuroimaging studies offer further insight into how the brain is shaped by learning and experience, in particular, showing effects that are specific to experience. For example, adults with long-term experience in navigation show enhanced neural development in brain regions thought to be responsible for spatial processing (posterior hippocampus). Experimental studies in which individuals are randomly assigned to receive training in a particular skill (e.g., juggling and videogames) show distinctive patterns of change in neural structure and function (e.g., among jugglers, bilateral expansion of mediotemporal and left posterior parietal areas, thought to be responsible for visual storage and processing).

At the same time, older adults may show compensatory patterns of resource allocation and neural recruitment. In a number of different task domains, neuroimaging data have provided evidence that older adults show reduced hemispheric asymmetry in activation patterns, demonstrating expanded recruitment of brain regions from both hemispheres. Older adults, especially those with relatively better performance, are likely to show greater activation of the prefrontal cortex, suggesting that successful cognitive performance may increasingly require executive attentional control with age. Interestingly, recent research also suggests that increased frontal recruitment with aging may be exaggerated among those with relatively high levels of education, providing some support for the self-regulatory account of education–cognition relationships.

Conclusion

This article has discussed the dynamic interplay between cognitive development and educational experiences throughout the life span, so that trajectories of adult cognitive development must shape educational practices and educational practices can promote cognitive vitality. Within this framework, education throughout the life span becomes a public health issue.

Sociologists have distinguished between age-segregated and age-integrated social structures. In age-segregated structures, permissible social roles are tightly tied to chronological age (e.g., education during youth and work during midlife). Within developed countries, advances in

medicine and health practices have stimulated a worldwide shift in demographics toward older populations, such that age segregation is no longer a tenable model. Life is simply too long now: effective work cannot rest on temporally removed education, and intellectual engagement is critical to vitality at every stage of the life span.

See also: Cognition and Emotion; Cognition: Overview and Recent Trends; Memory; Problem Solving and Human Expertise; Problem Solving and Reasoning; The Neuroscience of Aging and Cognition.

Further Reading

- Baltes, P. B. (1997). On the incomplete architecture of human ontogeny: Selection, optimization, and compensation as foundation of developmental theory. *American Psychologist* **52**, 366–380.
- Barnes, D. E., Tager, I. B., Satariano, W. A., and Yaffe, K. (2004). The relationship between literacy and cognition in well-educated elders. *Journal of Gerontology: Medical Sciences* **59A**, 390–395.
- Beier, M. E. and Ackerman, P. L. (2005). Age, ability, and the role of prior knowledge on the acquisition of new domain knowledge: Promising results in a real-world learning environment. *Psychology and Aging* **20**, 341–355.
- Bialystok, E., Craik, F. I. M., and Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia* **45**, 459–464.
- Carstensen, L. L., Mikels, J. A., and Mather, M. (2006). Aging and the intersection of cognition, motivation, and emotion. In Birren, J. E. and Schaie, K. W. (eds.) *Handbook of the Psychology of Aging*, 6th edn., pp 343–362. New York: Academic Press.
- Johnson, R. E. (2003). Aging and the remembering of text. *Developmental Review* **23**, 261–346.
- Kramer, A. F., Bherer, L., Colcombe, S. J., Dong, W., and Greenough, W. T. (2004). Environmental influences on cognitive and brain plasticity during aging. *Journal of Gerontology: Medical Sciences* **59A**, 940–957.
- Krampe, R. T. and Charness, N. (2006). Aging and expertise. In Ericsson, K. A., Charness, N., Feltovich, P. J., and Hoffman, R. R. (eds.) *The Cambridge Handbook of Expertise and Expert Performance*, pp 723–742. New York: Cambridge University Press.
- Lachman, M. E. (2004). Development in midlife. *Annual Review of Psychology* **55**, 305–331.
- Lachman, M. E. (2006). Perceived control over aging-related declines: Adaptive beliefs and behaviors. *Current Directions in Psychological Science* **15**, 282–286.
- Meyer, B. J. F. and Pollard, C. K. (2006). Applied learning and aging: A closer look at reading. In Birren, J. E. and Schaie, K. W. (eds.) *Handbook of the Psychology of Aging*, 6th edn., pp 233–260. New York: Elsevier.
- Riley, M. W. and Riley, J. W. Jr. (2000). Age integration: Conceptual and historical background. *Gerontologist* **40**, 266–270.
- Scarmeas, N. and Stern, Y. (2003). Cognitive reserve and lifestyle. *Journal of Clinical and Experimental Neuropsychology* **25**, 625–633.
- Stine-Morrow, E. A. L. (2007). The Dumbledore hypothesis of cognitive aging. *Current Directions in Psychological Science* **16**, 300–304.
- Stine-Morrow, E. A. L., Miller, L. M. S., and Hertzog, C. (2006). Aging and self-regulated language processing. *Psychological Bulletin* **132**, 582–606.
- Thornton, R. and Light, L. L. (2006). Language comprehension and production in normal aging. In Birren, J. E. and Schaie, K. W. (eds.) *Handbook of the Psychology of Aging*, 6th edn., pp 261–287. New York: Academic Press.

Vygotsky and Recent Developments

J V Wertsch, Washington University in St. Louis, St. Louis, MO, USA

© 2010 Elsevier Ltd. All rights reserved.

The study of learning and cognition has been heavily influenced by the ideas of Lev Semënovich Vygotsky. Vygotsky was born in Byelorussia in 1896 and spent the most important period of his professional life working in research and educational institutes in Moscow, where he died of tuberculosis in 1934. During his short life, Vygotsky addressed a remarkable range of topics in philosophy, psychology, semiotics, pedagogy, and literary analysis.

Vygotsky's works fill more than ten volumes, most written during the final decade of his life. His ideas have spawned a new round of research in the post-Soviet years, including intellectual biographies and analyses of how his ideas are tied to those of other major figures of his time such as Gustav Gustavovich Shpet.

Vygotsky's contributions to contemporary psychology can be spelled out in terms of three general themes that run throughout his writings. To one degree or another, these themes also characterize much of the writing of contemporary researchers influenced by Vygotsky, researchers who have developed ideas about cultural–historical and sociocultural approaches to psychology, and they characterize much of the writing on activity theory as well. These themes can be stated as follows:

- Cognition must be understood developmentally (i.e., genetically) in terms of its origins and subsequent development at individual and cultural levels of analysis.
- Cognition is mediated by semiotic mechanisms, especially natural language.
- Higher (i.e., uniquely human) mental processes such as problem solving, voluntary memory, and voluntary attention have their origins in social activity.

All of these themes reflect the general idea that mental functioning in the individual is fundamentally shaped by the social and historical context. Instead of focusing on how individual mental processes are made manifest in social expression, Vygotsky focused on how social expressions become manifest in individual mental processes.

Although such an approach emphasizes the social world, social interaction, and what Vygotsky termed intermental functioning, it does not amount to some form of social reductionism. Individual agency is still a central part of the picture. Like Mead, Vygotsky postulated that the key to understanding higher mental processes is how individuals participate in, and appropriate, social processes. In a nutshell, the approach asserts that developmental or genetic analysis must be employed to study how

processes that originate in social action shaped by semiotic mediation are transferred to the individual plane and shape higher mental processes.

Genetic Analysis

Vygotsky approached the early years of ontogenesis in terms of the interaction of two lines of development: the natural and the cultural, and he viewed the qualitative transformation of mental functioning that results from this interaction as key. He viewed any attempt to reduce all mental functions to amalgamations of lower order, biologically driven stimulus–response patterns as fundamentally misguided. The development of will or volition, for instance, is not to be seen as growing out of an iterative process of stimulus–response mechanisms, but as a qualitatively different type of phenomenon grounded in social processes. This was part of his effort to avoid the traps of the associationist and other reductionist approaches.

Like other major developmental theorists of his day, Vygotsky did not view developmental analysis as applying only to ontogenesis, and although much of his empirical study focused on children, he by no means assumed that developmental analysis could be equated with child psychology. Instead, it applies to several, qualitatively distinct genetic domains (Wertsch, 1985), namely phylogenesis, sociocultural history, ontogenesis, and microgenesis. As such, his notion of developmental psychology was much broader than what is often assumed in contemporary writings.

Vygotsky's approach to genetic analysis tends to interpret cultural differences in terms of developmental hierarchy, something that is no longer readily accepted by many cultural anthropologists. In their comparative studies of abstract reasoning in the 1930s, for example, Vygotsky and Luria interpreted differences between Uzbek and Russian performance as reflecting different stages in a grand developmental hierarchy. For them, this was what might be termed a cross-historical, rather than a cross-cultural study since they interpreted differences in the groups' performance as reflecting different stages in the evolution of a single general form of human civilization. The findings from these studies and the methods Luria used to generate them continue to provide inspiration for empirical research today. For example, a great deal of fruitful work over the last three decades on

cross-cultural comparisons of the psychological effects of literacy and schooling stems from the ideas of Vygotsky and Luria.

Contemporary modes of interpreting empirical findings, however, are quite different from what Vygotsky and Luria used. While accepting genetic analysis as a valuable technique in domains such as sociocultural history and ontogenesis, investigators today are likely to reject Vygotsky's assumption that cross-cultural differences can somehow be reduced to cross-historical differences. At least since the work of Franz Boas (1966) and Edward Sapir (1921), this assumption has been highly suspect in disciplines such as cultural anthropology in the US and Europe. Specifically, any tendency to view cultural differences in terms of historical evolution is likely to lead to charges of Eurocentricism since it is virtually always the case that the perspective used to do the comparing turns out to be at the top of the developmental hierarchy. Even with this caveat, however, Vygotsky's ideas have had a powerful impact on cross-cultural comparisons of cognition and other forms of mental functioning (Cole, 1996).

Mediated Nature of Human Mental Functioning

For Vygotsky and sociocultural approaches to cognition in general, the key to development is that social interactions are mediated by semiotic systems, most importantly language. His emphasis was on how forms of language use that shape human communication are appropriated by children in the development of cognitive processes on the individual plane. Vygotsky made increasingly strong claims toward the end of his career to the effect that an understanding of language and other cultural tools provides the foundation for the rest of his approach. Under the general heading of psychological tools, he included "language; various systems for counting; mnemonic techniques; algebraic symbol systems; works of art; writing; schemes, diagrams, maps, and mechanical drawings; [and] all sorts of conventional signs" (Vygotsky, 1981a: 137). Researchers such as John-Steiner (1991) have explored the role of symbol systems including drawing, gesture, music, or diagrams in the development of cognition.

An essential aspect of Vygotsky's treatment of mediational means is that its incorporation into human action (including mental functioning) does not simply make this action easier or more efficient in some quantitative sense. Instead, its incorporation typically results in a qualitative transformation. In his view "by being included in the process of behavior, the psychological tool [sign] alters the entire flow and structure of mental functions. It does this by determining the structure of a new instrumental act, just as a technical tool alters the process of a natural adaptation by determining the form of labor operations" (Vygotsky, 1981a: 137).

Vygotsky's emphasis on language as a cultural tool is evident in the writings of his followers as well. For example, Luria (1982) argued that just as mastery of a physical tool transforms human physical activity, mastery of the symbolic tool transforms human mental activity:

Language, in the course of social history, became the decisive instrument which helped humans transcend the boundaries of sensory experience, to assign symbols, and to formulate certain generalizations or categories. Thus, if humans had not possessed the capacity for labor and had not had language, they would not have developed abstract, "categorical" thinking. (p. 27)

Such studies by Vygotsky and his colleagues reveal that they tended to view language and other cultural tools as always working in favor of more advanced human functioning, as inevitably leading to more sophisticated performance. Other analysts have challenged Vygotsky's relatively uncritical stance toward mediation and have explored ways in which cultural tools constrain as well as facilitate action. These critiques suggest that language and other cultural tools may restrict, as well as enable activity often because they emerge or are privileged for reasons other than to facilitate the action in which they are eventually embedded.

The study of forces that give rise to cultural tools has not usually been the main focus of analyses of mediated action, but there are a few general points to make nonetheless. Perhaps the most interesting of these is that many of the cultural tools employed in mediated action were not designed for the role they have come to play. An illustration of this can be found in the keyboards used to type in English. Almost all users of such keyboards use the so-called QWERTY version, named after the fact that these letters are located at the upper left-hand portion of the array. Unless otherwise informed, most users of this keyboard assume that it was designed to facilitate their typing. In actuality, however, just the opposite is the case from today's perspective. The QWERTY keyboard was designed in an era of mechanical typewriters when the biggest impediment to efficient typing was having two or more keys jam together. As a result, the designers of the QWERTY keyboard specifically devised it to slow typists down.

With the appearance of electric typewriters and word processors, there is obviously no such need to slow typists down. Nevertheless, the vast majority of individuals who type in English continue to use the QWERTY keyboard, something that is made all the more striking by the fact that there is a readily available alternative keyboard design that is superior for most typists in terms of speed and accuracy. For example, the Dvorak keyboard is relatively easy to master, and most computer keyboards can easily be converted to its configuration.

The fact that the vast majority of individuals typing in English continue to use the QWERTY keyboard speaks

of the power of historical, economic, and other forces in shaping the cultural tools we employ. It also speaks of the tendency to use whatever psychological tools are handed to us in an uncritical way. This suggests that many of these tools may not be designed, or may not have evolved to facilitate the forms of mediated action in which they are currently employed. The particular case of the QWERTY keyboard is sometimes viewed as an isolated example of how technological and economic forces can go wrong. However, as authors such as Norman (1993) have argued, institutional, cultural, and historical forces often result in technology that is far from ideally designed from the perspective of the user, and this raises the question of whether similar issues might not be involved for all sorts of cultural tools.

The history of natural language presents an intriguing set of problems from this perspective. For the most part, language is not consciously planned or designed, a point that makes it somewhat different from the QWERTY keyboard example. However, many of the lessons of this illustration apply to language as well. For example, literacy and its impact on social and individual action raise several interesting questions. Literacy skills acquired in formal educational settings are associated with a specific set of cognitive skills, and the kind of language use required in formal literacy training is related to a willingness and ability to engage in tasks such as syllogistic reasoning.

However, it is generally accepted that literacy did not emerge in human history as part of an effort to facilitate skills such as those required in abstract reasoning tasks. Instead, literacy emerged in response to needs such as keeping records and conducting communication about commercial transactions. Furthermore, specific writing systems have often emerged when speakers of one language have borrowed the script used for another. Such facts serve to reinforce the claim that many cultural tools arise in response to forces that have little to do with the range of functions they are eventually required to serve.

In summary, cultural tools are often not simply neutral cognitive instruments. Instead, they may introduce historical and political dimensions into mental functioning and its socialization. Indeed, the distribution of psychological tools is often part of larger sociocultural debates and social differentiation. An example of this can be found in debates in the US surrounding the efficacy of Spanish or African-American vernacular as classroom instructional languages. These debates provide stark reminders that all mediational means are not equally valued in a society. Nor are they made equally available. This of course also applies to the distribution and use of tools such as computers in modern societies. The digital divide in the US means that computers may be widespread in many well-funded public and private schools, and all but absent in low-income urban or rural schools.

This suggests that cultural tools are implicated in the reproduction of social hierarchies. Many analysts of learning and cognition may consider these issues to be outside the boundaries of their area of inquiry, but to the extent that sociocultural psychology is concerned with the intersection of human mental functioning and the institutional, historical, and cultural contexts in which it occurs, it must take account of the social and political aspects of cultural tools. As scholars such as Duncan (1996) suggest, those working within a Vygotskian tradition must critically appraise the function of cultural tools and pattern of access to social spheres that they afford.

Social Origins of Individual Mental Functioning

In his approach to human mental functioning, Vygotsky outlined an account that began with action, namely mediated action (Zinchenko, 1985). Furthermore, he argued that the origins of this action are social, and in this connection he sought the developmental precursors of individual mental functioning in social processes. Perhaps the most general statement of this theme in Vygotsky's writings can be found in his general genetic law of cultural development:

Any function in children's development appears twice, or on two planes. First it appears on the social plane and then on the psychological plane. First it appears between people as an interpsychological category and then within the individual child as an intrapsychological category. . . .but it goes without saying that internalization transforms the process itself and changes its structure and function. Social relations or relations among people genetically [i.e., developmentally] underlie all higher functions and their relationships. (Vygotsky, 1981b: 163)

In this view, human mental functioning originates in inter-individual activities and only gradually develops into intramental processes. The very definition of mind is expanded such that its origins can be traced to activities between people, and the structural and functional organization of mind on the intermental plane provides the foundation for intramental functioning.

An essential part of Vygotsky's formulation of the intermental and intramental planes is that he viewed them as being inherently related. Indeed, the boundaries between social and individual functioning are quite permeable in his account, and his concern was with ongoing transformations between intermental and intramental processes rather than with any sharp distinctions that can be drawn. From this perspective, an element of sociality characterizes even the most private and internal forms of mental functioning:

[Higher mental functions'] composition, genetic structure, and means of action – in a word, their whole nature – is social. Even when we turn to [internal] mental processes, their nature remains quasi-social. In their own private sphere, human beings retain the functions of social interaction. (Vygotsky, 1981b: 164)

This statement does not assume that higher mental functioning in the individual is a direct and simple copy of socially organized processes; the point Vygotsky made in his formulation of the general genetic law of cultural development about transformations in internalization warns against any such view. Furthermore, it does not assume that nothing of interest goes on in the mind or brain of the individual when participating in intermental functioning. Instead, it simply posits a close connection, grounded in genetic transformations, between the specific strategies and processes of intermental and intramental functioning.

Vygotsky's general genetic law of cultural development underlies several aspects of his account of human mental functioning. For example, his research on what Piaget had called egocentric speech convinced him that the origins of children's problem solving and concept development lay not in interaction with the physical environment, but in their participation in social processes. By participating in social interaction, children appropriate certain linguistically mediated problem solving, thinking, and regulatory techniques first for external, social activity, then for individual cognitive activity as well. In Kozulin's (1990) words, "Development is therefore not an unfolding or maturation of pre-existing 'ideas'; on the contrary, it is the formation of such ideas – out of what originally was not an idea – in the course of socially meaningful activity" (p. 114).

The general genetic law of cultural development has received the most attention in the West in its incarnation as the zone of proximal development. The implications of this construct have been examined from perspectives such as general development and learning and psychoeducational assessment. In many cases, Vygotsky's comments about this zone are extracted from the more general context of his argument, and as a result it may be difficult to appreciate that it is just one way that he played out the implications of the general theme about the social origins of individual mental functioning. In fact, he developed the notion of the zone or proximal development fairly briefly on only a couple of occasions in his writings.

Vygotsky defined the zone of proximal development as distance between the performance level of an apprentice operating independently on the intramental plane and the level of intermental functioning involving an apprentice and an expert. It has provided the foundation for analyzing adult-child interaction and instruction; interaction and learning of children with disabilities; assessment; and other purposes.

Vygotsky's discussion of intermental processes has also played a role in the formulation of ideas about socially shared cognition, distributed cognition, and other topics. In several of these cases, the discussion does not posit a transition from the social to the individual plane that Vygotsky mentioned in the general genetic law of cultural development. Instead of speaking of social origins, with the assumption that the primary role of intermental functioning is to give rise to intramental functioning, investigators of socially shared cognition are often concerned with human cognitive activity that remains on the intermental plane. This is now widely recognized in studies of workplace activities, and it has taken on new importance in educational settings as well with the rise of interest in issues such as reciprocal teaching (Palincsar and Brown, 1984) and communities of learners (Lave and Wenger, 1991).

In analyzing these processes, investigators have raised questions about how to understand and assess intermental functioning in its own right, that is, independent of how it may give rise to intramental functioning. This brings with it some interesting new assumptions about how the expression cognitive development is to be used. In contrast to the usual assumptions grounded in methodological individualism (Lukes, 1977), the point is that intermental functioning itself may be examined from the perspective of development. From this perspective, it is appropriate to examine the development of cognition of a group and not just of the individuals in it.

Some dyads and larger groups such as institutions and even entire societies seem to function differently and perhaps at more advanced levels than others. Differences in how institutions think (Douglas, 1986) or societies remember (Connerton, 1989) have long been recognized by anthropologists, sociologists, and other scholars, but such expressions, let alone the conceptual framework behind them are quite alien to most studies of cognitive development.

This raises questions requiring conceptual frameworks that will be quite different from those we currently employ. What does it mean for a group – as a group – to develop cognitively? How can we formulate the processes involved such that they can be studied in some kind of a principled way? How would we go about assessing the relative levels of development of groups? Such issues need to be addressed without falling into the traps of strong versions of collective memory or cognition (Wertsch, 2002).

Returning to the issue of social origins of individual mental functioning as outlined by Vygotsky in his general genetic law of cultural development, it becomes crucial to consider how the transition from intermental to intramental functioning is envisioned. As suggested by Vygotsky, development should not to be understood as simply the internalizing for private uses of what were originally

social forms of behavior. In fact, Cazden (1988) warns against a “mechanical conception of the process of internalization whereby overt social interaction (speaking and listening) becomes transformed into covert mental processes (thinking)” (p. 108).

Instead, during learning activity, a transfer of competence – or the transfer of strategic responsibility – from expert to novice occurs. In the process, both the learner and the activity are transformed. In order for this transfer and transformation to take place, both the learner (novice) and the teacher (expert) must be active partners in the dialog surrounding a task.

This focus on active participation on the part of the tutor as well as the apprentice has been a major theme in the writings of Rogoff (1990) on guided participation. From this perspective, it is as essential to recognize and understand the contributions made by the learner, or apprentice, as it is to recognize those made by the teacher. This amounts to a corrective to what some view as sort of cultural transmission model inherent in Vygotsky’s view, a model in which the learner is taken to have little active role. Instead of being passive recipients of an input or a hypothesis-generating algorithm, children (or adult novices for that matter) are taken to be active participants in the co-construction of conversation and activity. Strategic responsibility for the task is gradually transferred to them, and through activity, they transfer strategies for organizing and monitoring problem solving from the intermental to the intramental plane. On the way, the practice itself undergoes qualitative changes.

Conclusion

This overview of Vygotsky’s model of cognitive development has focused on three basic themes that run throughout his writings. The first of these is the supposition that genetic or developmental analysis provides the foundation for understanding human mental functioning. For Vygotsky, genetic analysis was not simply one among many modes of inquiry – it was the most important and fundamental one. Furthermore, this vision of developmental analysis did not apply only to ontogenesis, but to other genetic domains as well.

A second theme that runs throughout Vygotsky’s writings concerns the mediated nature of human mental functioning. Instead of viewing cognition as a process that occurs within the skin, his approach posits that human mental functioning is typically distributed between active agents and cultural tools. His insights about cultural tools and the mediated action to which they give rise bring with them a range of conceptual implications that are still to be fully explored. Among other things, it leads us to introduce cultural and political questions into the study of cognition by asking where the cultural tools that shape

cognition come from and whether they are accessed in equal or unequal ways in the contemporary world.

The third theme in Vygotsky’s writings concerns the social origins of individual mental functioning. This constitutes a second sense, along with mediated action, in which Vygotsky viewed mind as extending beyond the skin and as being distributed. His claims about how higher mental processes appear first on the intermental, and then on the intramental planes of functioning underlie many other aspects of his thinking, including his claims about the zone of proximal development.

Although Vygotsky died in 1934, many of his ideas have come to have a powerful impact on discussions of cognitive development only over the past few decades in the West. This impact has grown drastically as contemporary researchers continue to employ his theoretical claims to formulate new empirical studies. There is every reason to expect this trend to continue as we focus on how cognitive development occurs in complex sociocultural settings.

See also: An Overview of Language and Literacy in Educational Settings; Apprenticeship Approach to Learning; Constructivism and Learning; Cultural–Historical Activity Theory; Learning in a Sociocultural Perspective.

Bibliography

- Boas, F. (1966). Introduction. In Boas, F. (ed.) *Handbook of American Indian Languages*. Lincoln: University of Nebraska Press.
- Cazden, C. (1988). *Classroom Discourse: The Language of Teaching and Learning*. Portsmouth, NH: Heinemann.
- Cole, M. (1996). *Cultural Psychology: A Once and Future Discipline*. Cambridge, MA: Harvard University Press.
- Connerton, P. (1989). *How Societies Remember*. Cambridge: Cambridge University Press.
- Douglas, M. (1986). *How Institutions Think*. Syracuse, NY: Syracuse University Press.
- Duncan, G. (1996). Space, place, and the problem of race: Black adolescent discourse as mediated action. *Journal of Negro Education* 65(2), 133–150.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Lukes, S. (1977). Methodological individualism reconsidered. In Lukes, S. (ed.) *Essays in Social Theory*, pp 177–186. New York: Columbia University Press.
- Palincsar, A. S. and Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction* 1, 117–175.
- Rogoff, B. (1990). *Apprenticeship in Thinking: Cognitive Development in Social Context*. Cambridge: Cambridge University Press.
- Salomon, G. (ed.) (1993). *Distributed Cognitions: Psychological and Educational Implications*. Cambridge: Cambridge University Press.
- Sapir, E. (1921). *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace.
- Vygotsky, L. S. (1981a). The instrumental method in psychology. In Wertsch, J. V. (ed.) *The Concept of Activity in Soviet Psychology*, pp 134–143. Armonk, NY: M.E. Sharpe.
- Vygotsky, L. S. (1981b). The genesis of higher mental functions. In Wertsch, J. V. (ed.) *The Concept of Activity in Soviet Psychology*, pp 144–188. Armonk, NY: M.E. Sharpe.
- Wertsch, J. V. (1985). *Culture, Communication and Cognition: Vygotskian Perspectives*. New York: Cambridge University Press.

Further Reading

- Brown, A. L. and Ferrara, R. A. (1985). Diagnosing zones of proximal development. In Wertsch, J. V. (ed.) *Culture, Communication, and Cognition: Vygotskian Perspectives*, pp 273–305. New York: Cambridge University Press.
- Cole, M. and Engeström, Y. (1993). A cultural–historical interpretation of distributed cognition. In Salomon, G. (ed.) *Distributed Cognitions: Psychological and Educational Considerations*, pp 1–46. Cambridge: Cambridge University Press.
- Daniels, H. (in press). Pedagogy. In Daniels, H., Cole, M., and Wertsch, J. V. (eds.) *The Cambridge Companion to Vygotsky*. Cambridge: Cambridge University Press.
- Edwards, A. (in press). An interesting resemblance: Vygotsky, mead, and american pragmatism. In Daniels, H., Cole, M., and Wertsch, J. V. (eds.) *The Cambridge Companion to Vygotsky*. Cambridge: Cambridge University Press.
- Engeström, Y. (in press). Putting Vygotsky to work: The change laboratory as an application of double stimulation. In Daniels, H., Cole, M., and Wertsch, J. V. (eds.) *The Cambridge Companion to Vygotsky*. Cambridge: Cambridge University Press.
- Gee, J. P. (1999). *An Introduction to Discourse Analysis: Theory and Method*. New York: Routledge.
- John-Steiner, V. (in press). Vygotsky on thinking and speech. In Daniels, H., Cole, M., and Wertsch, J. V. (eds.) *The Cambridge Companion to Vygotsky*. Cambridge: Cambridge University Press.
- Kozol, J. (1991). *Savage Inequalities: Children in America's Schools*. New York: Crown Publishers.
- Kozulin, A. (1990). *Vygotsky's Psychology: A Biography of Ideas*. Brighton, UK: Harvester Wheatsheaf.
- Kozulin, A. and Gindis, B. (in press). Sociocultural theory and education of children with special needs. In Daniels, H., Cole, M., and Wertsch, J. V. (eds.) *The Cambridge Companion to Vygotsky*. Cambridge: Cambridge University Press.
- Leont'ev, A. N. (1981). The problem of activity in psychology. In Wertsch, J. V. (ed.) *The Concept of Activity in Soviet Psychology*, pp 37–71. Armonk, NY: M.E. Sharpe.
- Linell, P. (1998). *Approaching Dialogue: Talk, Interaction, and Contexts in Dialogical Perspectives*. Amsterdam: John Benjamins.
- Luria, A. R. (1979). *The Making of Mind: A Personal Account of Soviet Psychology*. Cambridge, MA: Harvard University Press.
- Luria, A. R. (1982). *Language and Cognition*. New York: Wiley.
- Mead, G. H. (1955). In *Mind, Self & Society from the Stand-Point of a Social Behaviorist*. Morris, C. W. (ed.) Chicago, IL: University of Chicago Press.
- Meshcheryakov, B. G. (in press). Terminology in L.S. Vygotsky's writings. In Daniels, H., Cole, M., and Wertsch, J. V. (eds.) *The Cambridge Companion to Vygotsky*. Cambridge: Cambridge University Press.
- Middleton, D. and Edwards, D. (eds.) (1990). *Collective Remembering*. London: Sage.
- Minick, N. (1987). Introduction. In Vygotsky, L. S. (ed.) *Thinking and Speech*. New York: Plenum.
- Newman, F. and Holzman, L. (1993). *Lev Vygotsky: Revolutionary Scientist*. London: Routledge.
- Norman, D. A. (1988). *The Psychology of Everyday Things*. New York: Basic Books.
- Norman, D. A. (1993). *Things that Make Us Smart: Defending Human Attributes in the Age of the Machine*. Reading, MA: Addison-Wesley.
- Olson, D. R. (1994). *The World on Paper: The Conceptual and Cognitive Implications of Writing and Reading*. Cambridge: Cambridge University Press.
- Resnick, L. V., Levine, J. M., and Teasley, S. D. (eds.) (1991). *Perspectives on Socially Shared Cognition*. Washington, DC: American Psychological Association.
- Rogoff, B. (1981). Schooling and the development of cognitive skills. In Triandis, H. C. and Heron, A. (eds.) *Handbook of Cross-Cultural Psychology*, vol. 4, pp 233–294. Boston, MA: Allyn and Bacon.
- Rogoff, B. (1997). Evaluating development in the process of participation: Theory, methods, and practice building on each other. In Amsel, E. and Renninger, A. (eds.) *Change and Development*, pp 265–285. Hillsdale, NJ: Erlbaum.
- Rogoff, B. and Wertsch, J. V. (1984). Children's learning in the "zone of proximal development." In Rogoff, B. and Wertsch, J. V. (eds.) *New Directions for Child Development*, vol. 23, pp 19–30. San Francisco, CA: Jossey-Bass.
- Scollon, R. (1998). *Mediated Discourse as Social Interaction: A Study of News Discourse*. London: Longman.
- Scribner, S. and Cole, M. (1981). *The Psychology of Literacy*. Cambridge, MA: Harvard University Press.
- Shweder, R. A. (1990). Cultural psychology – What is it? In Stigler, J. W., Shweder, R. A., and Herdt, B. (eds.) *Cultural Psychology: Essays on Comparative Human Development*, pp 1–46. New York: Cambridge University Press.
- Stenberg, R. J. and Grigorenko, E. L. (2002). *Dynamic Testing: The Nature and Measurement of Learning Potential*. New York: Cambridge University Press.
- Taylor, C. (1985). *Human Agency and Language: Philosophical Papers I*. Cambridge: Cambridge University Press.
- van der Veer, R. (in press). Vygotsky in context: 1900–1935. In Daniels, H., Cole, M. and Wertsch, J. V. (eds.) *The Cambridge Companion to Vygotsky*. Cambridge: Cambridge University Press.
- van der Veer, R. and Valsiner, J. (1991). *Understanding Vygotsky: A Quest for Synthesis*. Oxford: Blackwell.
- Vygotskaya, G. L. and Lifanova, T. M. (1996). *Lev Seménovich Vygotskii: Zhizn', deyatel'nost', shtrikhi, i portrety (Lev Seménovich Vygotskii: Life, Activity, Traits, and Portraits)*. Moscow: Smysl.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. In Cole, M., John-Steiner, V., Scribner, S., and Souberman, E. (eds.). Cambridge, MA: Harvard University Press.
- Vygotsky, L. S. (1986). *Thought and Language*. Kozulin, A. (trans.) Cambridge, MA: MIT Press.
- Werner, H. (1948). *Comparative Psychology of Mental Development*. New York: International Universities Press.
- Wertsch, J. V. (1979). From social interaction to higher psychological processes: A clarification and application of Vygotsky's theory. *Human Development* 22, 1–22.
- Wertsch, J. V. (1991). *Voices of the Mind: A Sociocultural Approach to Mediated Action*. Cambridge, MA: Harvard University Press.
- Wertsch, J. V. (1998). *Mind as Action*. New York: Oxford University Press.
- Wertsch, J. V., del Rio, P., and Alvarez, A. (1995). Sociocultural studies: History, action, and mediation. In Wertsch, J. V., del Rio, P., and Alvarez, A. (eds.) *Sociocultural Studies of Mind*, pp 1–36. New York: Cambridge University Press.
- Wertsch, J. V., Tulviste, P., and Hagstrom, F. (1993). A sociocultural approach to agency. In Forman, E. A., Minick, N., and Stone, C. A. (eds.) *Contexts for Learning: Sociocultural Dynamics in Children's Development*, pp 336–356. New York: Oxford University Press.
- Zinchenko, V. P. (1985). Vygotsky's ideas about units for the analysis of mind. In Wertsch, J. V. (ed.) *Culture, Communication, and Cognition: Vygotskian Perspectives*, pp 94–118. New York: Cambridge University Press.
- Zinchenko, V. P. (in press). Thought and word: The approaches of L.S. Vygotsky and G.G. Shpet. In Daniels, H., Cole, M., and Wertsch, J. V. (eds.) *The Cambridge Companion to Vygotsky*. Cambridge: Cambridge University Press.

LEARNING AND COGNITION – ISSUES AND CONCEPTS – FOCUS ON COGNITION

Contents

Cognition and Emotion

Cognition: Overview and Recent Trends

Development of Creativity

Knowledge Domains and Domain Learning

Problem Solving and Human Expertise

Problem Solving and Reasoning

Reasoning, Moral and Social

Remembering as Social Activity

Cognition and Emotion

F Pons, University of Oslo, Oslo, Norway

M de Rosnay, University of Sydney, Sydney, NSW, Australia

F Cuisinier, University of Paris Ouest, Nanterre, France

© 2010 Elsevier Ltd. All rights reserved.

Introduction

What is the causal impact of emotion on cognition, and of cognition on emotion? Although this question is as classical as the one about, for example, the relation between nature and nurture, or consciousness and unconsciousness, it is not possible at this time to find a mapping of its possible answers within the fields of educational and psychological sciences. The investigation of the causal relation between cognition and emotion within these fields can be characterized as a disparate archipelago of research islands each somewhat isolated from the others and often covered or surrounded by an almost impenetrable fog. This article presents a representative map of this territory.

Why is it important to address the nature of the causal relation between cognition and emotion? The reasons are multiple. First, this question is interesting as a fundamental concern of human psychology that has attracted relatively little empirical attention. Of more than 200 000 psychological and educational publications (articles, book chapters, and books) including in their title the terms cognition or emotion (or equivalent terms such as intelligence, intellect, reasoning, mood, feeling, or affect), less than 4% include both terms (source PsycINFO and ERIC, May 2009). Moreover, of this 4%, almost all investigate cognition and emotion as the cause or the effect of a third phenomenon (e.g., the impact of gender on cognition and emotion or the impact of cognition and emotion on music). Few studies address the causal relation between cognition and emotion

and, when they do, it is typically in one direction. Second, a better understanding of the causal relation between cognition and emotion could improve our understanding of both phenomena via the analyses of a sometimes underestimated dimension of their nature; that is, the emotional nature of cognition and the cognitive nature of emotion. Third, within pragmatic or applied contexts, such as educational settings, these two phenomena are currently assessed independently; for instance, educational attainment and behavioral/emotional problems are known to be linked but are treated as separate domains. Improved understanding of the relation between children's cognitive and emotional development should contribute to the elaboration of integrated methods for assessment and treatment or intervention; not just for atypical children, for all those within an educational setting. Last but not least, most cognition and emotion researchers and professionals within psychology and education recognize that cognition and emotion are in some sort of relation. Yet, most of the time, such individuals are quite ignorant of each other's expertise or hold false beliefs about one another even when they work in the same institution. For these four reasons, among others, the simultaneous consideration of cognition and emotion, and the nature of their relation, has much to offer.

Why does the nature of the relation between cognition and emotion still remain so open today? There are at least two reasons: the oldness of the schism between these two domains of psychological and educational enquiry, and the relative youth of affective sciences (i.e., the scientific

study of emotions, moods, and feelings, and their constitutive elements). Indeed, the failure to provide integrated accounts of cognition and emotion as psychological phenomena has its roots in the history of philosophy, within which there is a classical distinction between reason and the passions, reason having been given pride of place in philosophical enquiry. Solomon explains the persistent failure to integrate these two domains as a function of the inferior role accorded to emotion, beginning with Plato, and the fact that emotion and cognition have been treated as though they are different kinds, thereby excluding the necessity of simultaneous consideration (Solomon (2000) in Lewis and Haviland-Jones (2000)). Pascal (1670/1998) eloquently captured this latter distinction when he said, “The heart has its reasons that the reason does not know.”

Contemporary cognitive science continues to be dominated by a preoccupation with cognition as a cold phenomenon. Predominantly, emotion is not considered or it is viewed as a confounding variable. When emotional and cognitive phenomena are considered together, emotion is often treated as a heating process, not directly integrated with cognition but causing spikes in temperature that interfere with cognitive processes.

The cognition–emotion dualism is to a great extent perpetuated methodologically within contemporary psychological and educational sciences. Indeed, in experimental sciences, you cannot study everything at the same time; it is very difficult to be simultaneously deep and broad. The fine-grained questions of experimental sciences need focused methods and are, in many cases, deeply connected to certain procedures such as reaction time or failure–success paradigms. By contrast, the study of emotion has relied to a great extent on subjective self-report at one end of a spectrum, and expressive behaviors (e.g., facial expressions) or physiological measures (e.g., blood volume pulse) at the other end of the spectrum. Such methodological disparities are lessening as a function of the rapid growth in neuroscientific methods (e.g., functional magnetic resonance imaging (fMRI), positron emission tomography (PET), second-generation electroencephalography (EEG)) but it remains unclear when we will be able to identify and assess both cognitive and emotional functions unfolding in real time: this is a goal for the future. In the meantime, cognitive scientists and emotion researchers very often talk a different talk and walk a different walk.

As indicated earlier, the affective sciences are still relatively young. It is also notable that they have to some extent define themselves against a backdrop of cognitivism – just as the cognitive scientists once defined themselves against behaviorism. Whereas the demise of behaviorism opened the way to studying cognitive processes within the black box, the affective revolution no longer considers persons merely as cognitive processors, more or less similar to Turing machines, and has allowed scientists to ask what

was happening in the heart of people: emotion has become an object of scientific enquiry to be described, explained, predicted, and even transformed. At one end of the methodological spectrum, private, conscious experiences such as feelings and moods and their self-report via introspection are again considered as legitimate scientific objects and methods. At the other end of the spectrum, neuroscientific methods raise the possibility that scientists will find some of the physical correlates of peoples’ subjective experiences, thus reinforcing their status as a legitimate object of scientific enquiry. Such methods may also undermine the dominant dualistic attitude in psychological and educational sciences regarding the mind–body relationship: they open a space where the study of emotion can take place (as the missing link) between the mind and the body.

The rest of the article is divided into four sections. We start with a short discussion about the meaning of emotion and cognition. Then we discuss the straightforward conceptions of the causal relationship between cognition and emotion: cognition as a cause of emotion, and emotion as a cause of cognition. In the conclusion, we speculate about the circular causal relationship between cognition and emotion and about some promising future research and intervention programs.

The Problem of Definition

Problems surrounding the meaning of cognition and emotion are at least as old as psychological and educational sciences. The question of how to define these terms is sometimes construed as a false problem with no existence, a real problem with no solution, or a problem with as many solutions as people trying to resolve it! It is not the function of the article to offer a definitive definition for each term but, rather, to consider how the two constructs have been related to one another. Nevertheless, it is important to set out the scope of these terms as they are used in the psychological and educational sciences. Indeed, many misunderstandings and conflicts centered on the relation between cognition and emotion are related to the absence of a common understanding of the phenomena in question and to the overgeneralization of some quite specific definitions.

Here, cognition refers to the different forms of knowledge (e.g., belief, thought, etc.) that we have and, critically, to the mental functions (e.g., systems, schemas, processes, etc.) making the acquisition, storage, retrieval, transformation, and use of this knowledge possible, for example, memory, attention, intelligence, language, mental imagery, and so on. This knowledge can be empirically or logically true or false, real or unreal, and more or less certain. It can be more or less simple or complex, sensory or symbolic, temporary or permanent, general or specific, conscious or unconscious, controllable or uncontrollable, and universal or idiosyncratic.

Importantly, it is traditionally held that knowledge is (usually implicitly) rule bound or structured and that the mental functions operate in accordance with such rules; for example, knowledge of language implies some form of syntax (that is consciously inaccessible to language users without formal education). Finally, such knowledge and mental functions are considered as interrelated states and processes of the mind (e.g., subjective experience of the color red), the body (e.g., neuronal correlates associate with the perception of the color red), and the culture (e.g., words to represent the color red: red, *rod*, *rojo*, rouge, etc.).

Emotion, on the other hand, is a class of feelings directed to objects, where objects include persons, things, and situations, both real and imagined. Emotions can be pleasant or unpleasant and more or less intense or moderate, but they can be differentiated from sensations and states of bodily arousal because of their inherent aboutness (i.e., intentionality) and their close connection with actions (Kenny, 1963). Emotions can be more or less basic (e.g., happiness, anger, fear, sadness, disgust, or surprise) or complex (e.g., guilt, shame, pride, jealousy, and mixed emotions), sensory or symbolic/abstract, temporary (e.g., sadness about a specific event) or enduring (e.g., an ongoing depressive mood), conscious or unconscious, controlled or uncontrolled, general or specific, and universal (e.g., fear of death) or idiosyncratic (Spanish Duende, Portuguese Saodade, etc.). Not all these aspects of emotion are uncontested, of course, but they fall under the common sense notion of emotion and the emotional lives of persons. Emotions are generally thought to be organized in that they are discrete (particularly in childhood) and they have their own natural history: certain situations and thoughts, as well as expressions and states of the body, are universally constitutive of specific emotions presumably because of our evolutionary origins (Darwin, 1872/1899; Ekman, 1999 in Dalglish and Power, 1999; Lazarus, 1991). Finally, emotions are interrelated states and processes of the mind (e.g., subjective experience of happiness about or of something), of the body (heart rate changes, respiration rate, muscle tension, pupil dilatation, etc.), and of the culture (e.g., the social and cultural norms related to the feeling of anger).

In the foregoing discussion, we have provided a sketch of emotion that is clearly cognitive in some sense: That is, the inherent aboutness of emotion entails that an emotion has cognitive content. For example, we are scared of the ferocious dog or happy about recent political events. The object of an emotion (e.g., the ferocious dog) and the beliefs on which an emotion is founded (e.g., that ferocious dogs may do us an injury) have a conceptual rather than a causal relation to the emotion (see Solomon, 2000 in Lewis and Haviland-Jones (2000) for a discussion). In the following sections, however, we take as a starting point the historical assumption that emotion and cognition may be treated as separable phenomena that

may stand in a cause or effect relations to one another – an assumption that may ultimately unravel.

Before we embark on this discussion, however, consider some research that suggests that emotional responses can occur before there is time for any cognitive processing (even unconscious). For example, LeDoux (1996) demonstrated that some emotional answers, taking place within a few milliseconds (e.g., fear of already known stimuli) result from an immediate mid- and lower-brain (i.e., the amygdala and the thalamus) response to a stimulus and can be produced without the involvement of the cortex, where cognitive processes are assumed to take place. It should be noted, however, that LeDoux also showed that learning a new emotional reaction to a stimulus involves the (sensory) cortex and therefore, it is reasoned, some cognitive processing, until this emotional reaction is completely automatic. Interestingly, to the best of our knowledge, no experimental study has tried to demonstrate that cognitive answers can be produced without being linked to emotion (e.g., without the emotional areas of the brain being involved), though this seems entirely plausible.

Cognition as a Cause of Emotion

Until 30 years ago, explanations of emotional phenomena had been preoccupied mainly with the body. Two hypotheses were then in competition: the James-Lange and the Cannon. In a nutshell, the former postulated that emotions resulted from subjective perceptions of bodily states (e.g., we are sad because we cry) and the latter postulated almost the opposite (e.g., we cry because we are sad). The debate about the relation between the subjective experiences of emotions and their body correlates is still very much alive (e.g., Cacioppo *et al.*, (2000) in Lewis and Haviland-Jones, 2000; Damasio, 1994; Ekman and Rosenberg, 2005).

With the affective revolution, however, attention shifted to the role of cognition – specifically, the role of cognitive appraisal or evaluation. Schachter and Singer (1962) were among the first to give an empirical demonstration of the existence of this kind of appraisal. They injected people with adrenaline. One group was told that the injection would have an impact on their heart beat (which was true) while a second that it would have no impact (which was untrue). Although, the two groups had the same bodily experience (increase of heart beat), only the second group of people reported feeling emotions. Moreover, when the second group was exposed to a happy person, they reported happiness, and anger when exposed to an angry person (these emotions were feigned by actors present in the room with the subject).

The cognitive-appraisal hypothesis has been the object of active debate (see Lazarus, 1982, 1984 in Lazarus, 1991; Zajonc, 1980 in Zajonc, 1984) and criticism (see Reisenzein (1983) for a review): is cognition (whether conscious or not)

a necessary (albeit nonsufficient) condition of emotion? Nevertheless, the basic premise that cognitive appraisal often has an influence (both as an antecedent and as a modulator) on emotional experience (whether that be the valence of the emotion experienced or the level of emotional arousal) found numerous empirical supports (e.g., Frijda, 1986; Roseman, 1984; Scherer, 1984; Smith, 1989), and underpins those cognitive-psychotherapy theories that aim to alter the cognitive-appraisal process in emotional disorders (e.g., Beck, 1976). Several dimensions of the cognitive appraisal of stimuli have been identified: novelty/familiarity, valence, goal/need significance, coping potential, and compatibility with personal and cultural norms (e.g., Kappas (2006) and Scherer *et al.*, (2001) for reviews). For example, a stimulus that is appraised as pleasant (valence) and acceptable (according to personal and cultural norms) could result in feeling happiness.

While there are many variants of cognitive appraisal, they are unified in stressing a very tight and, in most cases, subjectively instantaneous connection between an experienced emotion and the beliefs or thoughts that accompany it. There is, however, another sense in which our emotional experience is affected by our understanding of circumstances: in addition to knowing things about the situations in which we feel emotions, we also have knowledge of emotions (of their nature, causes, consequences, and possibilities of regulation) which has a tremendous potential to impact on our emotional experience by changing the nature of the relation we have to the emotion eliciting circumstances or the emotion itself (e.g., see for reviews, de Rosnay *et al.*, 2008; Haga *et al.*, 2008; Harris, 2006; Pons *et al.*, 2005). For example, between 4 and 5 years of age, children begin to understand the effect of memories on emotions. They realize that the intensity of anger decreases over time; looking at a picture of a lost loved one can reignite sadness; or thinking about a positive past event can cause joy. Development brings further emotion insights. From about 8 years of age, even sooner under certain conditions, children begin to understand how feelings can be regulated via the use of cognitive strategies such as the cognitive re-evaluation of the situation (e.g., “that’s not the end of the world”) and the re-orientation of attention (e.g., “let’s think about something positive”); thus children demonstrate conscious knowledge of the influence of thought process (cognition) on emotions. Consider as an example the well-documented association between depression and the understanding of the strategies to regulate feelings: children and adolescents who think that the use of strategies such as rumination and passivity are better to deal with negative emotions than strategies such as reevaluation of the situation or re-focusing show more depressive symptoms.

Furthermore, there is accumulating evidence that children’s knowledge about emotions has an impact on their emotional experiences and well-being in a social context.

This impact has been identified in both preschool and school children (see Pons *et al.*, 2005 for a review). For example, young children who better understand situation–emotion regularities (e.g., feeling happy when receiving a gift or sad when breaking a toy) are also the most popular with their day-care friends. At 5 years of age, children who are better able to recognize basic emotions are also the most popular with their classmates 1 or 2 years later. The relationship between emotion understanding and social functioning is also seen in middle childhood: 9-year-old children (particularly girls) with a good understanding of negative emotion-regulation strategies are considered, by their classmates and teachers, to be the most socially competent.

Surprisingly, there is relatively little research on the impact of school achievement on emotion. Some studies reveal a complex pattern of relations between positive and negative emotions, and school performance and self-evaluation of academic competencies. The link seems to be stronger between emotions and self-evaluation of academic competencies than between emotions and actual school performances (Gumora and Arsenio, 2002).

Emotion as a Cause of Cognition

It is widely held, in some form or another, that our emotions put us in some kind of meaningful relation to the world: A big ferocious dog is not just a fast-moving, big-toothed, furry creature; it is something to be feared because it is a potential threat to our well-being. How, then, does this meaningful relation between the person experiencing the emotion and the world influence our cognition? The answer to this question, within psychological and educational sciences, is twofold: emotional arousal is considered to be motivational, and emotional valence is considered to be a compass that guides our cognition.

At a general level, emotional arousal has an impact on cognition because it mobilizes the mind, as well as the body and the culture, to act and react (e.g., fear prepares the mind, the body, and the culture for fight or flight). This motivational function is the only one recognized by Piaget (1954/1981). He acknowledged that the intellect needs emotional arousal, which can speed up or slow down the functioning of the intellect and therefore its development. In 1908, Yerkes and Dodson postulated that an appropriate level of emotional arousal is needed to achieve an optimal level of cognitive performance. Too much anxiety (high arousal) or not enough interest (low arousal) might have a negative impact on the cognitive functions. They also postulated a functional relation between emotional arousal and task demands: the more difficult the task, the lower the level of emotional arousal required to reach the optimum level of cognitive performance.

Emotion not only makes cognition move but also orients its movement. Emotional arousal provides energy, whereas emotional valence gives the direction for cognition to move and therefore to develop. This is well illustrated by the behaviorist research tradition. Most classical and operant conditioning would not be possible without emotional valence: Emotional reactions (more or less positive or negative) to a stimulus (more or less conditioned or unconditioned) very often determine the behavioral response to it such as the appearance, disappearance, prioritization, and transformation of the (cognitive) behavior. This compass function of emotional valence for cognition is also recognized by Freud (1905/2002). The expressed or repressed drive coming from the Id not only gives its energy to the cognitive self but also orients (e.g., activation and inhibition) its functioning and development (in collaboration with the cognitive super-ego).

Numerous experimental studies have demonstrated the impact of emotional arousal and valence on cognition (e.g., memory, attention, and creativity). For example, when shown a list of positive and negative words, most people recall more positive words. If these people are clinically depressed, however, then they will tend to recall more negative words especially when these words have a clear negative valence (death, cancer, war, etc.). In a similar vein, typical people recall more elements of a story when the emotional valence of the story matches their own current mood, especially when the story is sad and they are in a sad mood while reading the story. It should be noted that when typical people are in a neutral mood, they remember more positive than negative elements of the story (e.g., Matt *et al.* (1992) for a review). It is as if, by default, when we are feeling nothing or at least at peace, our memory has an emotionally positive orientation, which may be adaptive. Other studies have also shown that the recall of information is facilitated when the mood of the person is the same at encoding and recall, irrespective of the emotional valence of this information. This effect is stronger when the information is autobiographical. For example, the recall of a list of neutral words is facilitated if the person is in a sad mood both when trying to learn (encoding) and to recall this list (see Eich and Forgas, (2003) for a review). Such findings may help to explain why the recall of a traumatic autobiographical event is facilitated when people are already in a negative mood – a vicious cycle recognized by many schools of psychotherapy (Beck, 1976).

Emotional arousal and valence also have a direct impact on attentional processes (Eysenck *et al.* (2007); MacLeod (2005) for reviews). People with mood disorders pay more attention to information (words, pictures, faces, etc.) with an emotional valence matching their mood. For example, depressed people have the tendency to pay more attention to their personal characteristics (internal locus of control) when explaining their failures (e.g., because who I am,

what I did, or what I said) and they pay more attention to their environment (external locus of control) to explain their success (e.g., because of luck, fate, or external events). In stark contrast to typical people, those who are clinically anxious (phobia, posttraumatic stress disorder (PTSD), etc.) direct their attention more toward threatening words (cancer, evil, death, etc.) than neutral words (house, picture, chair, etc.). When confronted with homonym words (e.g., batter-pancake vs. batter-assault), non-anxious people have the tendency to activate the neutral meaning of the word, whereas anxious people have the tendency to activate the negative meaning.

A person's level of emotional arousal and the valence of their emotional experience may also play an important role when constructing new information. Projective tests such as the Thematic Apperception Test (TAT) and the Rorschach are partially built on this assumption. It is argued, for example, that people's emotions determine, via a process of projective identification, their interpretation of reality, especially when this reality is ambiguous (ambiguous inkblots could be interpreted as a loving mother, a threatening father, a lost child, an unfaithful partner, etc.). Attachment theory (Bowlby, 1969/1997) presents us with an interesting illustration of this phenomenon. It is widely held that distinctive patterns of emotion co-regulation in infancy, so-called attachment styles (i.e., secure, ambivalent, avoidant, and disorganized), continue to inform the child's understanding of relationships much later on via enduring cognitive representations of such relationships (i.e., internal working models). Thus, if presented with a quasi-ambiguous picture of a mother and father leaving a child as they go on holiday for 2 weeks, children's understanding and construal of the events depicted are to a great extent determined by their emotional attachment style in infancy.

Of course, emotions may have an impact on other aspects of memory, attention, and creativity (e.g., research on flashbulb memories and autobiographical memory) and more generally on cognition (e.g., intelligence, language, perception, and moral reasoning). It should also be noted that most studies of the impact of emotion on cognition were conducted with adults (e.g., Brennen *et al.*, 2007; Overskeid, 2000). However, findings with children seem to confirm those obtained with adults.

Indeed, a substantial number of studies have shown that emotions such as anxiety (but also joy, pride, shame, and fear) have an impact on several aspects of learning at school: achievement, motivation, interest, goals, metacognition, etc. These studies showed that this impact may differ from one academic domain to another (i.e., mother tongue, mathematics, arts, etc.) and from one pupil to another (i.e., individual differences due to gender) (see Lafortune and Pons 2005 in Pons *et al.* (2005); and Schutz and Pekrun, (2007) for reviews). For example, induced positive mood in children (e.g., a compliment about

clothes) improves their performance in a block-design task, which is a classical measure of intelligence. Anxiety can prevent pupils from exercising all of their capacities and can, in some cases, prevent them from doing any mathematical reasoning altogether. Anxiety also influences the functioning of metacognition. Certain students feel that when mathematical explanations are given, a veil, even a wall, suddenly appears in front of them, stopping them from reaching the concentration level necessary for understanding what they are being shown. They are thus prevented from engaging in the metacognitive processes necessary to solve the problems. More generally, these studies showed that pupils' emotional competences (i.e., their capacity to experience, recognize, express, control the expression of, regulate the experience of, and understand emotions) have an impact on their school achievement.

Until quite recently, most of the studies on the impact of emotions at school had been almost exclusively focusing on pupils. Lately, a new line of research has emerged which suggests some interesting relations. Positive emotions experienced by teachers are related to the level of support from parents and colleagues, and to pupils' cooperative behaviors and learning achievement. Teachers' negative emotions (e.g., anger, frustration, or sadness) stem from the absence of support from colleagues and parents, and are related to pupils' disruptive behaviors. Teachers' emotions also vary according to their professional experience and the academic level of their teaching: Novice teachers report more anxiety than experienced teachers and primary school teachers report more sadness and helplessness than secondary school teachers (often related to their pupils' life difficulties). In sum, the emotional constellation of the teacher may significantly influence the educational environment of the classroom.

Conclusion

The aforementioned studies demonstrate how cognition and emotion can work together. Cognition and emotion may be thought of as two different languages, to represent and communicate about the world (ourselves, others, the physical world, etc.) that coexist within all typical individuals: Every person is emotionally and cognitively bilingual.

While illustrating impressive relations between cognition and emotion, many of the aforementioned studies actually make it difficult to know which is causally antecedent. The allocation of cause, in many cases, may depend on the moment you are taking the snapshot in the flow of the person's subjective experience (e.g., "I am anxious, which makes me focus my attention on anxious information, which in return reinforces my anxiety, which in return makes me focus on anxious information, etc.").

We may also speculate on the fact that, as a function of the situation (context, circumstances, etc.) and the individual (personality, level of development, etc.), either cognition or emotion may be dominating the individual's mental functioning and, further, that the absence of this circularity would be dysfunctional for the individual.

Until quite recently, pedagogy had focused its attention almost exclusively on the cognitive dimension of the mind, for pupils and teachers. There are at least two challenges for researchers and professionals in education and psychology. The first is the development of reliable and valid, but also realistic and interdisciplinary, methodologies to improve emotional competences within school settings. This applies to children and adolescents in relation to their school achievement, and also to teachers in relation to their teaching skills. Today, few if any of these methodologies have been properly developed, let alone evaluated for their reliability and validity. There is a need for interventions that make a discernable impact on pupils' and teachers' emotional competences and bring about stable long-term positive changes in pupils' school achievement and teachers' pedagogical competences. The second challenge is to introduce emotional competences (i.e., the abilities to experience, recognize, express, control the expression of, regulate the experience of, and understand emotions) in the standard curriculum of preschool and compulsory school, and in the training programs of the teachers. This implies strong political decision. Indeed, the introduction of emotion pedagogy at school is still often seen as irrelevant when there are no existing problems – the object and goal of the school institution being still often considered as cognitive and instructive rather than emotional and educative. This change in thinking about the importance of emotions at school also implies a reorganization of pupils' and teachers' curriculum, teachers already having much to teach and learn. Such reorganization can be quite problematic not only because it is a source of change (and that can cause some resistance) but also because it implies that some topics which are currently taught have to be reduced within the curriculum and may be even eliminated.

Summary

Three main results can be abstracted from the educational and psychological studies discussed in this article:

1. Some (old) emotional reactions can be elicited without the intervention of cognition (via a direct biological appraisal of the stimulus and without the involvement of the cortex). Surprisingly, the opposite (i.e., cognition without emotion) has not yet been so well documented; this absence perhaps being due to the fact this documentation is too trivial (or impossible).

Some researchers have been tempted by an overgeneralization of these no-causal findings. However, the fact that cognition and emotion are sometimes not causing one another (or that they are not the only cause of one another) does not imply that they are never causing one another.

2. Indeed, numerous studies have demonstrated that many emotional reactions and modulations are caused by the individual's cognitive appraisal of the stimuli coming from his/her body, mind, and environment. They have also demonstrated a robust relationship between children's cognitive understanding of emotions (including the possibility to regulate emotions) and their social and emotional well-being.
3. Numerous studies have equally demonstrated that emotional arousal and/or valence have an impact on memory (encoding, storage, and retrieval of existing information), attention (activation and inhibition of existing information), and creativity (elaboration of new information). They have also shown that the impact of emotional arousal on cognition is not linear and that the impact of positive and negative emotional valences is not symmetrical. Emotional arousal has to be neither too low nor too high to have an optimum impact on cognition (i.e., not an additional burden for cognitive executive functions) and the impact of negative emotions on cognition seems clearer than the impact of positive emotions.

While illustrating impressive relations between cognition and emotion, many of the studies discussed in this article actually make it difficult to know which is causally antecedent. The allocation of cause, in many cases, may depend on the moment you are taking the snapshot in the flow of the person's subjective experience (e.g., "I am depressed, which makes me focus my attention on depressive information, which in return reinforces my depression, which in return makes me focus on depression information"). We may also speculate on the fact that, as a function of the situation (context, circumstances, etc.) and the individual (personality, level of development, etc.), either cognition or emotion may be dominating the individual's mental functioning and, further, that the absence of this circularity would be dysfunctional for the individual.

Finally, although most of the studies reported here have been conducted in laboratories, an increasing number of studies seems to confirm, with some nuances, the mutual impact of cognition and emotion on pupils and teachers. In the future, it would be interesting to further investigate this mutual impact of cognition and emotion at school. Such investigations could have a positive influence on pupils' school achievement and on the quality of teachers' teaching. It would also test the validity of our understanding of the relation between cognition and emotion. Indeed, one thing is to demonstrate the logical coherence and the

empirical correspondence of our understanding; another is to demonstrate that our understanding can change pragmatically the world of practice for the best.

See also: Affect, Mood and Emotions; Anxiety; Assessment in Schools – Affective Domain; Cognition: Overview and Recent Trends; Coping with Stressful Situations: An Important Aspect of Self-Regulation; Educating Students with Emotional and Behavioral Disorders; Educational Measurement: Overview; Emotion in Educational Contexts; Empathy; Interest; Learning Through Play; Psychoanalysis and Education; Social and Emotional Outcomes of Learning; Volitional Control of Learning; Wellbeing.

Bibliography

- Beck, A. (1976). *Cognitive Therapy and the Emotional Disorders*. New York: Meridian.
- Bowlby, J. (1969/1997). *Attachment and Loss. Volume 1: Attachment*, 2nd edn. Sydney: Random House.
- Brennen, T., Dybdahl, R., and Kapidzic, A. (2007). Trauma-related and neutral false memories in war-induced posttraumatic stress disorder. *Consciousness and Cognition* **16**, 877–885.
- Dalgleish, T. and Power, M. (eds.) (1999). *Handbook of Cognition and Emotion*. Chichester: Wiley.
- Damasio, A. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Putnam.
- Darwin, C. (1872/1899). *The Expression of the Emotions in Man and Animals*. New York: D Appleton (Plain Label Books).
- de Rosnay, M., Harris, P., and Pons, F. (2008). Emotion understanding and developmental psychopathology in young children. In Sharp, C., Fonagy, P., and Goodyer, I. (eds.) *Social Cognition and Developmental Psychopathology*, pp 343–385. Oxford: Oxford University Press.
- Eich, E. and Forgas, J. (2003). Mood, cognition, and memory. In Healy, A. and Proctor, R. (eds.) *Handbook of Psychology: Experimental Psychology*, vol. 4, pp 61–83. Hoboken, NJ: Wiley.
- Ekman, P. and Rosenberg, E. (eds.) (2005). *What the Face Reveals? Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford: Oxford University Press.
- Eysenck, M., Derakshan, N., Santos, R., and Calvo, M. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion* **7**(2), 336–353.
- Freud, S. (1905/2002). *The Joke and Its Relation to the Unconscious*. London: Penguin Books.
- Frijda, N. (1986). *The Emotions*. Cambridge: Cambridge University Press.
- Gumora, G. and Arsenio, W. (2002). Emotionality, emotion regulation and school performance in middle school children. *Journal of School Psychology* **40**, 395–413.
- Harris, P. (2006). Social cognition. In Kuhn, D., Siegler, R., Damon, W., and Lerner, R. (eds.) *Handbook of Child Psychology: Vol 2. Cognition, Perception, and Language*, 6th edn., pp 811–858. Hoboken, NJ: Wiley.
- Haga, S., Kraft, P., and Corby, E. (2008). Coming to terms with emotion regulation: A cross-cultural exploration of antecedents and well-being outcomes. *Journal of Happiness Studies*.
- Kappas, A. (2006). Appraisal are direct, immediate, intuitive, and unwitting... and some are reflective... *Cognition and Emotion* **20**(7), 952–975.
- Kenny, A. (1963). *Action, Emotion and Will*. Thetford: Lowe and Brydone.
- Lazarus, R. S. (1991). *Emotion and Adaptation*. Oxford: Oxford University Press.

- LeDoux, J. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon and Schuster.
- Lewis, M. and Haviland-Jones, J. (eds.) (2000). *Handbook of Emotions*. New York: Guilford Press.
- MacLeod, C. (2005). The Stroop Task in clinical research. In Wenzel, A. and Rubin, D. (eds.) *Cognitive Methods and Their Application to Clinical Research*, pp 41–62. Washington, DC: American Psychological Association.
- Matt, G., Vasquez, C., and Campbell, W. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychological Review* **12**, 227–255.
- Overskeid, G. (2000). The slave of the passions: Experiencing problems and selecting solutions. *Review of General Psychology* **4**, 284–309.
- Pascal, B. (1670/1998). *Œuvres Complètes*. Paris: Gallimard.
- Piaget, J. (1954/1981). *Intelligence and Affectivity. Their Relationship during Child Development*. Palo Alto, CA: Annual reviews.
- Pons, F., Hancock, D., Lafortune, L., and Doudin, P. -A. (eds.) (2005). *Emotions in Learning*. Aalborg: Aalborg University Press.
- Reisenzein, R. (1983). The Schachter theory of emotion: Two decades later. *Psychological Bulletin* **94**(2), 239–264.
- Roseman, I. (1984). Cognitive determinants of emotions: A structural theory. In Shaver, P. (ed.) *Review of Personality and Social Psychology*, vol. 5, pp 11–36. Beverly Hills, CA: Sage.
- Schachter, S. and Singer, J. (1962). Cognitive, social and physiological determinants of emotional state. *Psychological Review* **69**, 379–399.
- Scherer, K. (1984). On the nature and function of emotion: A component process approach. In Scherer, K. and Ekman, P. (eds.) *Approaches to Emotion*, pp 293–318. Erlbaum: Hillsdale.
- Scherer, K., Schoor, A., and Johnstone, T. (eds.) (2001). *Appraisal Processes in Emotion: Theory, Methods, Research*. Oxford: Oxford University Press.
- Schutz, P. A. and Pekrun, R. (eds.) (2007). *Emotion in Education*. Burlington: Elsevier.
- Smith, C. (1989). Dimensions of appraisal and physiological response in emotion. *Journal of Personality and Social Psychology* **56**, 339–353.
- Zajonc, R. (1984). On the primacy of affect. *American Psychologist* **39**, 117–123.

Further Reading

- Braisby, N. and Gellatly, A. (eds.) (2005). *Cognitive Psychology*. Oxford: Oxford University Press.
- Davidson, R., Scherer, K., and Goldsmith, H. (eds.) (2003). *Handbook of Affective Sciences*. Oxford: Oxford University Press.
- Manstead, N., Frijda, N., and Fischer, A. (eds.) (2004). *Feelings and Emotions: The Amsterdam Symposium*. Cambridge: Cambridge University Press.
- Saarni, C. (1999). *The Development of Emotional Competence*. New York: Guilford Press.

Cognition: Overview and Recent Trends

J Wiley, University of Illinois at Chicago, Chicago, IL, USA

B D Jee, Northwestern University, Evanston, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

All processes of thought, conscious and unconscious, fall into the realm of cognition. These processes operate by manipulating information-laden mental representations, which are either retrieved from memory or constructed from sensory information. In this way, the mind can be understood as an information processor, continuously adding to its repertoire of mental representations as well as producing overt physical behaviors. The study of human cognition thus becomes the study of the information-processing characteristics of the mind: What is the nature of the representations? How are they manipulated? How much information can be active at once? Cognitive researchers are concerned with discovering such facts about cognition, with the overarching goal of explaining human behavior in its various forms.

The study of human cognition is relevant to many fields, and researchers from several disciplines have contributed to our understanding of the mind. These disciplines include psychology, philosophy, linguistics, artificial intelligence, anthropology, education, and neuroscience. In fact, many cognitive researchers are multidisciplinary, simultaneously working in a variety of fields, and interdisciplinary, integrating aspects of the different disciplines.

The growing field of cognitive science was born out of the perspective that cognitive research must span multiple disciplines. However, less than 50 years ago, the study of cognition was largely considered unscientific, at least among many psychologists in North America. To understand the current approach to the study of cognition, it is important to consider its origins, as well as its obstacles. It is not the goal of this article to provide a detailed historical review of inquiries into human cognition (for such a review, see Gardner, 1985). Rather, we focus on the relatively recent emergence of the information processing framework that permeates contemporary cognitive science, and consider various major areas of cognitive research.

The Cognitive Revolution

The modern approach to the study of human cognition was forged by a series of events in the mid-twentieth century, known as the cognitive revolution. While scholars continue to debate whether the cognitive revolution truly qualifies as a scientific revolution, the era certainly marks an important shift in psychological theory and methodology, especially for psychology in North America. In the

first half of the twentieth century, experimental psychology in North America was dominated by behaviorism, which generally disavows the use of mentalistic explanations of behavior. Mentalistic terms, such as belief and desire, were branded as superfluous and unscientific, and removed from accepted terminology. Behaviorism came in a variety of forms, including theoretical behaviorism and methodological behaviorism. The methodological behaviorist did not necessarily dismiss the existence of mental constructs, but would argue that psychological science should not include such unobservable entities. Theoretical behaviorism went further, holding that mentalistic constructs could be reduced to overt behaviors. Under this view, remembering a certain stimulus, like a sugar cube or a foul-tasting liquid, amounts to producing a specific learned behavior in response to it. Rather than building explanatory theories, behaviorism had the goal of describing relationships between reinforcements and observable behavior. Such descriptions were, according to behaviorism's founding father, Watson, as far as experimental psychology could and should progress.

The apparent scientific rigor of behaviorism played a large role in its rise to prominence in psychology. However, to many American psychologists of the mid-twentieth century, it became increasingly apparent that limiting psychology to a science of behavior would be inadequate for explaining learning and higher-order processes of problem solving, reasoning, and decision making. Evidence for this inadequacy came from various sources, including a central research area in behaviorism, animal learning. For example, studies finding that rats could quickly learn an association between ingesting salty water and illness or an association between a certain noise and a shock, but had difficulty learning other pairings of the same elements (salt with shock or noise with illness), seemed to defeat the notion that learning is purely a result of experiential factors. The behaviorist approach could also not explain instances of creative cognition where novel responses are generated.

It would be incorrect to assert that the decline of behaviorism in American psychology was immediate, or due to any single set of events (Miller, 2003). Indeed, behaviorism was widely debated and defended throughout the 1960s and into the 1970s in various areas of psychology. In addition, the shift from behaviorism to cognitivism that marks the cognitive revolution in American psychology was influenced by developments outside of

America. For example, the Swiss psychologist Jean Piaget argued that the organization of the human mind is shaped by biological factors, while Russian psychologist Lev Vygotsky stressed sociocultural influences. Besides advancing new theories of psychological development, these researchers provided examples that psychology could maintain its scientific rigor without the constraints of strict behaviorism.

The cognitive revolution was also fueled by developments outside of psychology, notably in the fields of linguistics and artificial intelligence (AI). The linguist Noam Chomsky's famous critique of B. F. Skinner's *Verbal Behavior* stands out to many as a critical event. Skinner, the leading behaviorist researcher of his era, attempted to provide an explanation of language strictly in terms of environmental input. Chomsky argued that behaviorist principles are inadequate to explain human verbal behavior, because language acquisition does not seem to develop as a function of the environmental conditions that are presented to the learner. Chomsky cited evidence of children's rapid language learning, and of people's ability to create and comprehend entirely novel expressions. Chomsky argued that the external conditions are too impoverished to support such learning, and posited universal, innate language abilities. While Chomsky's arguments did not settle the issue on language learning, they did, in the eyes of many, serve to defeat the purely behaviorist account.

Fueled by post-World War II interests in advancing science and technology, the field of AI flourished in the 1950s and 1960s, and played a crucial role in the rise of cognitivism. The birth of AI can be traced back to 1956, when an influential conference on AI was held at Dartmouth College (Miller, 2003). Attendees included John McCarthy, Marvin Minsky, Herbert Simon, and Allen Newell, each of whom became leading figures in AI and cognitive-science research for decades to come. AI labs around the United States began producing computer programs that were capable of carrying out tasks that were previously believed to require the highest human intellect, such as solving complex logic and math problems. The new synergy between computer science and psychology was apparent not only in the tasks that computers were programmed to perform, but also in how they carried these tasks out. For example, McCarthy's List Processing (LISP) possessed a unique memory organization for its time, which stored not only the contents of independent entries, but also their connectedness. These memory trees allowed the computer to store complex relational structures, a powerful new tool.

With advancing computer technology bringing promise of creating AI, the abstract notion of the computer served as a powerful metaphor for the human mind. Broadbent is widely credited with developing this computer metaphor into the information-processing model of the mind, which has become the dominant model for cognitive science.

The new idea was that the mind could be described in terms of input, internal representation, processing, and output. In this way, the mind was not only like a computer, it was a computer. Psychological theories could be legitimately instantiated in computer models, a view that more closely bound psychology to the field of AI.

The Birth of Cognitive Psychology

The emergence of new scientific fields outside psychology and the open consideration of alternative theoretical positions within psychology supported the birth of a new psychological field, one committed to the study of human cognition under the information-processing framework. The publication of Neisser's *Cognitive Psychology* in 1967 marked the emergence of this new field. Neisser defined cognition as all of the processes that transform, reduce, elaborate, store, recover, and use sensory input. For Neisser, the realm of cognitive psychology included how actions and experiences are affected by perceptions, memories, and beliefs. Like behaviorists, cognitive psychologists apply a strict scientific method to study the mind, but unlike behaviorists, they accept the existence of internal mental states within their science.

Armed with the information-processing framework, cognitive psychologists set out to explore many new questions about the mind. A primary concern was the mind's basic architecture: Is the mind comprised of distinct cognitive systems, and if so, how do these systems operate? Research on human memory was central in addressing these questions. Memory is not simply a storage system, but is thought to be actively engaged in all aspects of cognition. The memory system has three interrelated functions: encoding, storing, and retrieving information. Evidence from numerous studies suggested the existence of distinct, but interacting systems of memory, distinguished primarily by the duration that information is retained. First, there is sensory memory, which holds a large amount of incoming sensory information for only a few hundred milliseconds. There is thought to be a unique sensory store for each of the five senses, although visual sensory memory has received the most research attention. The next memory system is short-term memory, which receives input from sensory memory, and holds, by recent estimate (Cowan, 2001), about four chunks (i.e., meaningful units) of information for a few seconds. A more contemporary view, Baddeley's model of working memory (e.g., Baddeley, 2007), emphasizes not only short-term storage, but also how information is manipulated in immediate memory. The most recent incarnation of the working-memory model consists of three distinct short-term memory subsystems, commanded by a central executive. One subsystem processes and stores visuospatial information, a second deals with phonological information

(including speech), and a third integrates different modalities of information and integrates temporal information into an episodic representation. A final major memory system is long-term memory, characterized by its near-unlimited capacity and duration. Information from working memory may be stored in long-term memory for later retrieval; however, the storage of information depends on many factors, including how it is processed during encoding. Long-term memory can also input information into working memory, affecting the interpretation and elaboration of its contents, and making information accessible for use in higher-order processes such as comprehension, problem solving, and reasoning.

The field of cognitive psychology, realizing the potential of information-processing theory and building on basic assumptions about the architecture of memory, flourished in the latter quarter of the twentieth century. New areas of research emerged, exploring a range of cognitive phenomena, from the most basic processes, such as perception and pattern recognition, to more complex, higher-order processes, like comprehension and problem solving. Traditionally, cognitive researchers aim to determine specific processes that underlie human performance using experiments that isolate the particular demands of any given task. Participants are usually presented with fine-grained variations of stimuli or tasks, and performance is typically measured in terms of response accuracy or reaction time. This approach is essentially reductionist in nature, but it allows for causal determinations to be made through adherence to the scientific principles of quantification and strict experimental control. Beyond the basic controlled experiment, cognitive scientists have a number of other approaches in their methodological arsenal. Cognitive modeling, the simulation of cognition through computer programs, is a popular tool for instantiating theories and pitting them against one another. Technology in neuroscience, particularly functional magnetic resonance imaging (fMRI), has made it possible to examine the online brain activity of normal adults. Similar recordings can be obtained in terms of event-related potentials (ERPs), eye movements, skin conductance, and pupil dilation. Verbal and video protocols also provide a rich source of data, including gestures and think-aloud reports, which may give insights into the behaviors that people engage in during cognitive tasks.

Overview of Research in Cognition

Given that one of the main goals of education is to train and inform students through instruction, research in cognition has direct application, especially with respect to issues of learning. There are several areas of cognitive research, including perception and attention, language acquisition and reading, memory, comprehension, problem solving

and reasoning, and metacognition, which can all inform educational practice.

Perception and Attention

One of the most basic areas of research within cognitive science concerns perceptual and attentional processes, which serve as gatekeepers for stimuli and determine what is available for further cognitive processing. The extent to which information must be explicitly processed, how attention is drawn to or away from stimuli, and how often information needs to be attended to during learning, are important and well-studied topics (Pashler, 1998). Repeated exposure or practice is a central mechanism in skill and knowledge acquisition, and can result in automaticity and proceduralization of cognitive processes over time. These principles can inform our understanding of successful acquisition of spoken language, reading skills, and various problem-solving skills. A related literature explores individual differences in the ability to control one's attention, or the ability to focus on a goal in the face of interference. Research has found that these abilities affect the amount of information that people can consider in immediate memory, which in turn impacts their performance on many cognitive tasks. As such, the ability to control one's attention has also been referred to as working-memory capacity (Conway *et al.*, 2005).

Language Acquisition and Reading

First-language acquisition is perhaps the most impressive of all cognitive skills. Language develops universally in some form in all humans, and development typically proceeds in a remarkably rapid fashion. Present approaches emphasize both the biological predisposition to acquire language as well as the role of exposure to patterns in the environment (Gleitman and Newport, 1995). The differences between first-language acquisition occurring in infancy, and second-language acquisition that occurs afterward, are important issues, as are the questions of how bilinguals represent and process information in their second language (Kroll and Tokowicz, 2005). Additionally, recent research has highlighted several cognitive advantages among young bilingual students and aging bilingual adults, including superior executive functioning and facilitated acquisition of reading skills (Stock, 2001). A great deal of cognitive research has also explored reading behavior more generally. This research supports two important conclusions: that mastering the alphabetic principle (that written symbols correspond to sounds, or phonics) is essential to becoming proficient, and that methods that teach the alphabetic-decoding principles that underlie our written language are more effective than methods that do not, especially for children who are having difficulty learning to read (Rayner *et al.*, 2001).

Memory

The relevance of cognitive research on memory is perhaps obvious if a main goal of education is the acquisition of knowledge. Memory research concerns the mechanisms that enable the encoding, storage, and retrieval of new information. A long tradition of research has demonstrated more durable learning and memory result from spaced practice (Cepeda *et al.*, 2006), and when information is presented in a context that allows for imagery, elaboration, or integration with prior knowledge (Baddeley, 2007). Other highly relevant lines of investigation examine how memory tests affect subsequent memory for information, in some cases by introducing false memory traces through distractors in multiple-choice tests (Roediger and Karpicke, 2006). Testing may also cause patterns of facilitation for the tested information, and interference for nonretrieved information.

Comprehension and Conceptual Understanding

Perhaps the most common medium for subject-matter transmission is through expository text, with lectures and discussion also being common forms of instruction. In these cases, learning depends on text and discourse processing which is another major area of cognitive research. As information is processed, multiple levels of representation are constructed, and as Kintsch (1998) has suggested, it is especially the deepest level of representation, called the situation model, that represents understanding or comprehension of phenomena. This area of research has explored the contexts and individual differences that support construction of better-situation models during learning, including encouraging the activation or use of prior knowledge and encouraging active processing through tasks such as question-asking or self-explanation (Graesser *et al.*, 1997; McNamara, 2007). A special case of learning occurs when new information or evidence in some way conflicts with prior understandings. In these cases, conceptual change or belief change may be required, especially for learning in science (Chi, 2000; Chinn and Malhotra, 2002).

Problem Solving and Reasoning

A great deal of research on problem solving has been informed by using a contrastive approach – comparing better or more expert problem solvers to less-effective, or less-expert, problem solvers. This work has shown that experts use their experience to see the deep (more explanatory) structure of a problem, while novices often process problems at a superficial level. This line of research has in fact spurred the research on self-explanation cited above (Chi, 2000). In a similar vein, instruction that emphasizes procedural elements of math problem solving can promote shallow understanding, while conceptual or

mixed approaches can lead to better understanding as well as improvement in procedural skill (Siegler, 2003). A third important area of problem-solving research has explored the effectiveness of scaffolded practice, worked examples, and feedback (Atkinson *et al.*, 2000).

Ill-structured problems represent a special class, where solutions, goals, and assumptions are less constrained. But even here, expertise in a domain gives the solver ways to constrain the problem space (Voss *et al.*, 1983) and turn the problem into a well-structured one. Solving this type of problem, where no single agreed-upon answer exists, requires argumentation or informal reasoning skills. In such cases, the reasoner must consider the evidence and arguments and whether they provide support for conclusions. Such skills underlie disciplinary thinking in both the humanities and the sciences, as well as reasoning in most everyday contexts. Thus, the development of scientific reasoning and argumentation skills in students has been an area of much recent interest.

As opposed to these skills of informal reasoning, formal reasoning relies on the application of the rules of logic or mathematics, with the goal of determining the validity of syllogistic or propositional arguments. Formal-reasoning performance often improves with more familiar or meaningful contexts, as well as when arguments are phrased in more intuitive language. Several forms of reasoning have been found to improve with training in statistics, but generally, people tend to be poor at all types of reasoning without training. Substantial individual differences in reasoning abilities have been attributed to thinking dispositions (Stanovich *et al.*, 2003) which have been found to correlate with the ability to learn and engage in many higher-order cognitive tasks. This in turn suggests a critical role for instruction in reasoning that is largely absent from our current curricula.

Metacognition

Metacognition is the act of monitoring cognitive performance, which serves as input to self-regulation of cognitive behaviors such as studying. Much research on this topic has used prediction paradigms for memory and comprehension tasks. Metamemory paradigms typically consist of the task of learning word pairs, such as a word in a foreign language and its English translation. When students are asked to predict their ability to recall a translation, metamemory accuracy has been found to be quite good, especially when judgments are made at a delay (Dunlosky *et al.*, 2007). Predictive accuracy for comprehension tests following the reading of expository texts is typically much lower, although recently some contexts that promote accurate metacomprehension have been found. When readers are asked to self-explain while reading, or to generate summaries or keywords at a delay before making their judgments, metacomprehension accuracy

can be substantially improved (Thiede *et al.*, 2009). The ultimate value of supporting better predictive accuracy among students is that it then can help them to make effective decisions of what material to re-study as they attempt to learn material on their own.

Recent Trends

Transfer

The goal of instruction is ultimately to promote learning that transfers to new problems and situations, especially ones that are quite dissimilar from the initial learning context. When students are able to activate prior knowledge or learned-solution strategies and apply them by analogy in novel contexts, it is referred to as far transfer. However, a recurring theme in the cognitive literature is that students often fail to activate relevant prior knowledge as they process new information. Thus, little experimental evidence has supported the existence of far transfer (Barnett and Ceci, 2002). More often than not, people are unlikely to recall critical information in situations that are superficially dissimilar to the context of acquisition. Despite the elusiveness of findings of far transfer in cognitive laboratories, there is evidence from observational studies that people can and do apply analogies to solve novel problems in the real world. Scientists, for example, spontaneously construct analogies to help solve problems that they encounter in their research. One reason for the apparent void between laboratory and observational findings could be that experimental participants are not able to represent the analog at a level of abstraction that is sufficient to support transfer. The issue of whether initial instruction should be concrete and contextualized, versus abstract and symbolic, is currently receiving a great deal of attention, with some advantages being found for each mode of acquisition. When students are taught a principle through concrete examples that gradually become more abstract, they are better able to transfer the principle in some contexts (Goldstone *et al.*, 2008). However, in other contexts, starting with abstract or symbolic representations leads to better learning and transfer, as concrete representations may distract learners from recognizing the basic principles.

Spatial Thinking and Gesture

Spatial skills are thought to be important to many aspects of cognition, including thinking and problem solving in math and science domains. Individual differences in spatial ability have been well documented, and recent work is attempting to differentiate among distinct subtypes of spatial abilities (Hegarty and Waller, 2005). One important aspect of spatial thought is the ability to parse, mentally manipulate, and use symbolic representations, such as maps and diagrams. To the extent that spatial skills

affect learning and understanding, it is particularly encouraging that such skills do seem to improve with practice or input from the environment (Newcombe and Huttenlocher, 2000). Another important aspect of spatial cognition is the bodily gestures that people produce (Alibali, 2005). Gestures can be used to support spatial thinking and communication, but also may be used to express knowledge that is difficult to verbalize. Gestures may reveal a student's preparedness for learning a concept even when their verbal expressions suggest poor understanding. Interestingly, gestures seem to be used as much for the self as for others. Such findings suggest that cognitive processes may be tied to the body's interactions with the world, and an extreme version of this perspective, referred to as embodied cognition, is receiving a great deal of attention in the recent literature.

Culture and Cognition

Research in early science education and developmental psychology has demonstrated that most children need to overcome some of their intuitive ideas of the world around them in order to comprehend scientifically accurate explanation of phenomena. However, the vast majority of this work has involved children in mainstream, North American, urban, technologically dependent populations. Recent work by Medin and his colleagues has shown how intuitive concepts are shaped by society, and how children in specific cultures (such as Native American communities) differ in their understandings of biological concepts (Bang *et al.*, 2007). These differences are attributed to exposure to cultural values and beliefs transmitted through the discourse within the community. Interestingly, children in Native American communities were found to have more advanced understandings of living things and ecologies. Yet, disturbingly, they did less well than mainstream students in academic classroom performance. These results suggest conflicts between traditional science instruction and the ways of thinking instilled by different communities. Such findings highlight the need to recognize the differences in intuitive understandings that children from many different kinds of communities bring to the classroom, and represent one example of the recent research trend to better understand the interaction of culture and cognition.

Collaboration

A final trend in recent research explores the impact of social interaction on learning. The communication between the teacher and a class is essentially a social interaction, but because of the number of students, it is essentially one-sided. The teacher necessarily becomes the transmitter of information, and there is little role for either individualized discourse or interactive activity on the part of most students.

Thus, it is not surprising that researchers have attempted to increase interactive communication for all students through the creation of small group exercises, peer-collaboration activities, or intelligent tutoring environments. Interacting with others can be particularly motivating, but there are also a number of pitfalls that can make collaborative learning less successful than individual work. Intelligent tutors can provide students with timely feedback and expert knowledge, but if systems provide too much scaffolding or help, then the students may never actually engage with the content (Anderson *et al.*, 1995). Alternatively, working with peers can be quite motivating and effective as a supplement to classroom instruction, but guidance for students on how to collaborate is critical, otherwise they may engage in superficial discussions that do not reach a high level of discourse (O'Donnell and King, 1999). The goal of current investigations is determining which contexts are most effective and in what ways, for different subject matters, and different kinds of students (VanLehn *et al.*, 2007), as well as the discovery of the mechanisms that are responsible for successful collaborative learning and transfer.

See also: Cognition and Emotion; Concept Learning; Memory; Metacognition; Neuroscience Bases of Learning; Problem Solving and Human Expertise; Problem Solving and Reasoning; The Adult Development of Cognition and Learning.

Bibliography

- Aibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating and thinking about spatial information. *Spatial Cognition and Computation* 5, 307–331.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., and Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences* 4(2), 167–207.
- Atkinson, R. K., Derry, S. J., Renkl, A., and Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research* 70(2), 181–214.
- Baddeley, A. D. (2007). *Working Memory, Thought and Action*. Oxford, UK: Oxford University Press.
- Bang, M., Medin, D., and Atran, S. (2007). Cultural mosaics and mental models of nature. *Proceedings of the National Academy of Sciences* 104, 13868–13874.
- Barnett, S. M. and Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Journal of Experimental Psychology: General* 128, 612–637.
- Bialystok, E. (2001). *Bilingualism in Development: Language, Literacy, and Cognition*. New York: Cambridge University Press.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., and Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin* 132, 354–380.
- Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In Glaser, R. (ed.) *Advances in Instructional Psychology*, pp 161–238. Mahwah, NJ: Erlbaum.
- Chinn, C. A. and Malhotra, B. A. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology* 19, 327–343.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., *et al.* (2005). Working memory span tasks: A methodological review and users guide. *Psychonomic Bulletin and Review* 2(5), 769–786.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage. *Behavioral and Brain Sciences* 24(1), 87–185.
- Dunlosky, J., Serra, M. J., and Baker, J. M. C. (2007). Metamemory applied. In Durso, F. (ed.) *Handbook of Applied Cognition*, 2nd edn., pp 137–161. Chichester, UK: Wiley.
- Gardner, H. (1985). *The mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books.
- Gleitman, L. R. and Newport, E. L. (1995). The invention of language by children: Environmental and biological influences on the acquisition of language. In Gleitman, L. R. and Liberman, M. (eds.) *An Invitation to Cognitive Science, Vol. 1: Language*, 2nd edn. Cambridge: MIT Press.
- Goldstone, R. L., Landy, D., and Son, J. (2008). A well-grounded education: The role of perception in science and mathematics. In Glenberg, A., DeVega, M., and Graesser, A. (eds.) *Proceedings of the Garachico Workshop on Symbols, Embodiment and Meaning*, Universidad de La Laguna, Tenerife.
- Graesser, A. C., Millis, K., and Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology* 48, 163–189.
- Hegarty, M. and Waller, D. (2005). Individual differences in spatial abilities. In Shah, P. and Miyake, A. (eds.) *Handbook of Higher-Level Visuospatial Thinking*, pp 121–169. New York: Cambridge University Press.
- Kintsch, W. (1998). *Comprehension: A paradigm for Cognition*. New York: Cambridge University Press.
- Kroll, J. F. and Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In Kroll, J. F. and de Groot, A. M. B. (eds.) *Handbook of Bilingualism: Psycholinguistic Approaches*, pp 531–553. New York: Oxford University Press.
- McNamara, D. S. (ed.) (2007). *Reading Comprehension Strategies: Theories, Interventions, and Technologies*. Mahwah, NJ: Erlbaum.
- Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences* 7, 141–144.
- Newcombe, N. S. and Huttenlocher, J. (2000). *Making Space: The Development of Spatial Representation and Reasoning*. Cambridge, MA: MIT Press.
- O'Donnell, A. and King, A. (1999). *Cognitive Perspectives on Peer Learning*. Mahwah, NJ: Erlbaum.
- Pashler, H. (1998). *The Psychology of Attention*. Cambridge, MA: MIT Press.
- Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., and Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest* 2(2), 31–74.
- Roediger, H. L. and Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science* 17, 249–255.
- Siegler, R. S. (2003). Implications of cognitive science research for mathematics education. In Kilpatrick, J., Martin, W. G., and Schifter, D. E. (eds.) *A Research Companion to Principles and Standards for School Mathematics*, pp 119–233. Reston, VA: National Council of Teachers of Mathematics.
- Stanovich, K. E., Sá, W. C., and West, R. F. (2003). Individual differences in thinking, reasoning, and decision making. In Leighton, J. P. and Sternberg, R. J. (eds.) *The Nature of Reasoning*, pp 375–409. New York: Cambridge University Press.
- Thiede, K. W., Griffin, T. D., Wiley, J., and Redford, J. (2009). Metacognitive monitoring during and after reading. In Hacker, D. J., Dunlosky, J., and Graesser, A. C. (eds.) *Handbook of Metacognition in Education*, pp 85–106. New York: Routledge.
- VanLehn, K., Graesser, A. C., Jackson, G. T., *et al.* (2007). When are tutorial dialogues more effective than reading? *Cognitive Science* 31(1), 3–52.
- Voss, J. F., Greene, T. R., Post, T. A., and Penner, B. C. (1983). Problem-solving skill in social sciences. *Psychology of Learning and Motivation* 17, 165–213.

Development of Creativity

M A Runco, University of Georgia, Athens, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Creative performance – The actual use of creative talent.

Creative potential – The range of possible talent determined largely by genetic makeup.

Domain of performance – The different areas (e.g., art and mathematics) which seem to require different skills and seem to be associated with different backgrounds.

Family structure – The number of persons in the home, birth order, and age gap among siblings.

Fourth-grade slump – A drop in creative thinking experienced by approximately 50% of children at age 9.

Humane creativity – A domain wherein talents are used in a moral and socially beneficial manner.

Old-age style – The tendency of some artists to make changes to their work simply for the sake of change and personal originality.

Problem finding – This is required before problem solving and often determines the quality of a solution. A creative solution may depend on the identification and definition of a creative problem.

Remote model – An individual in books or the media whom children may imitate.

Creativity is an extremely valuable resource, both for individuals and for society as a whole. It is most accurate to distinguish creative potential from creative performance, especially in studies of development. This is because an individual who is putting his or her creative talents to work is actually performing or applying those talents. The individual may not be performing publically – many creative endeavors involve problem solving, adaptation, coping, or other everyday, mundane, or personal actions. Yet many other individuals, including most children, have notable creative potential yet are not using all of their talents. Studies of development are useful in part because they suggest how such potentials can actually be fulfilled and how individuals can make the transition to actual creative performance and achievement.

This is not to say that creative potentials are entirely fulfilled through systematic experiences. Indeed, creative talents exemplify the same interplay between nature and nurture that is apparent in most aspects of human development. In addition, there is a kind of bidirectionality in

that creative talent both results from maturation and experience and contributes to learning and adaptation. It is, put simply, both a cause and an effect of development.

Creativity behavior requires both originality and effectiveness. Each of these follows its own developmental trajectory. To chart these trajectories, it is useful to define originality and effectiveness more specifically. The former can be defined in terms of novelty, uniqueness, or unusualness, and the latter in terms of fit, appropriateness, and value. When creative behavior is directed toward a problem, and is thus a kind of problem solving, effectiveness can be defined in terms of the quality of the solution(s). Yet, not all creative behavior is problem solving. Some precedes solving problems and is involved in what has been called problem finding or problem discovery. The creative behavior of children is often entirely independent of problems and is simply original self-expression. Adults also sometimes express themselves creatively; and in fact there are strong connections between self-actualization, spontaneity, authenticity, and creativity. Yet another manifestation of creative talent, which is independent of problem solving, is being proactive. Problem solving is usually reactive in that the individual is responding to an obstacle or problem. Proactive creativity may entirely circumvent the problem. This, of course, is a very important form of creativity and may be especially useful, given some of the contemporary threats facing our society. Indeed, it would make a great deal of sense to focus developmental and educational efforts specifically on proactive creativity, and on humane creativity or creativity in the moral domain.

The concept of domains is vital for understanding the development of creativity. The background of an artist may differ dramatically from that of an engineer, scientist, leader, manager, or educator. Even within the arts there are differences among subdomains, with dancers differing from writers, for example, and poets differing from novelists.

The developmental trajectory of originality begins at what is for many individuals the peak or maximal level. Young children are often uninhibited and unconventional. In fact, they are unaware of what conventions are, and, of course, do not take them into account when emitting behavior. They are instead enormously spontaneous. This lack of inhibition, unconventionality, and spontaneity supports their originality. Adult originality and spontaneity are often stifled by inhibition and by a recognition of conventions (e.g., cultural norms, traditions, and expectations of others). Children also do not make the same assumptions as adults but instead tend to be much more

mindful of immediate experience – they have few, if any, routines. Adults, on the other hand, often rely on assumption and routine. This helps the adult to conserve resources and is often beneficial, but it is difficult, if not impossible, to be original if the person is merely following routine or making assumptions rather than mindfully attending to the situation at hand.

Note, however, that all these provide children with easy access to originality, but not necessarily to creative action. Again, creative behavior requires effectiveness as well as originality, and children are often ineffective, at least in the sense that they are egocentric and ignore conventions. Their originality is unchecked rather than discretionary, and for that reason it is often just originality and not creativity. As a matter of fact, many of the advantages just listed, giving children easy access to originality, make it very difficult to strike an optimal balance with effectiveness. Very often, then, the developmental trajectory of originality peaks early and drops with age and increased conventionality, while the trajectory for effectiveness has the nadir early but gradually increases as the individual learns to fit in and exercise discretion and appropriate conventional action. The balance of originality and effectiveness has been called postconventional, the idea being that the individual is aware of conventions but mindfully decided whether or not to conform to them.

This is not to say that the development of creativity is entirely smooth and continuous. Various stages have been proposed. The problem-finding talents related to creative performance above, for instance, have been assigned to a postformal stage of development, which implies that it would be apparent only late in adolescence or in early adulthood – if at all. There are also various developmental slumps. One common slump occurs at approximately age 9 and was labeled the fourth-grade slump by E. Paul Torrance in 1968. This has also been found in several other countries around the world and has been reported much more recently as well. It is, however, only apparent in approximately 50% of the children of that age. Another slump occurs late in life. This involves cognitive flexibility, which is very frequently associated with creative performance. It is apparent when an individual is able to consider various perspectives, generate different kinds of ideas and solutions, and avoid rigidity and functional fixedness. Apparently, many older adults have difficulty doing just that and rely more and more on routines. They become more rigid and less flexible, and their creative thinking is likely to suffer.

However, just as the fourth-grade slump is only experienced by approximately 50% of the children of that age, so also is rigidity common and flexibility lacking only in a segment of the population. Consider in this regard the creative breakthroughs that are sometimes discovered by individuals late in life. Piaget and Picasso, for instance, were productive and creative well into their 80s. Apparently some

individuals actively invest in their creative potentials throughout their lives. Among artists, this has been labeled the old-age style. Simplifying it a little, this is seen when the older artist intentionally changes his or her style or medium, even if this means leaving previous successes behind.

Nature and Nurture in Development

Individual differences, such as those implied by the fourth-grade slump and the old-age style, can be understood in terms of the interplay of nature and nurture. Some individuals, for example, probably inherit the genes which increase the likelihood of a slump. Great strides have recently been made toward a deeper understanding of the genetic basis of creativity and its development. There is, for example, one gene known as *DRD2* which is involved in the reception of the neurotransmitter dopamine and which is found in perhaps 20% of the population. It is correlated with certain tests of creative potential. Other indications of a heritability of creativity and the role of nature include the genealogies of some famous creative persons. This evidence is not entirely convincing, however, because while there are a number of examples of creative talent within multiple generations of one family (many of which were identified by Galton in his 1869 classic titled *Hereditary Genius*), there are also clear counter examples (e.g., William Shakespeare did not seem to have any outstandingly talented relatives).

Whatever genetic contribution does exist, it does not work in isolation. Behavioral scientists do contrast nature with nurture any longer; most now acknowledge that biology and experience work together. Thus, it is most appropriate to say that the development of creativity is a function of both nature and nurture. This can be further explained with the concept range of reaction whereby the genetic potential for creativity sets a range of possibilities, and nurture and experience determine where within that range and individuals perform. This idea of an interplay between nature and nurture is important because although not just anyone can be an Einstein or Mozart, everyone does have potentials which can be fulfilled. Hence, much of what we learn about the development of creativity applies very widely, if not universally.

Models, Motivation, and Values as Influences on Creative Development

The experiences which contribute to the development of creativity and the fulfillment of creative potential are quite diverse. Nurture may include, for example, formal and informal education. Much of the latter occurs within the context of the family. Indeed, a great deal of research

has suggested that early family experiences are critical for the development of creativity. Creative potentials are most likely to be fulfilled if a child has numerous opportunities to experience and practice creative behaviors, as well as models, including the parents, who demonstrate how to be creative and convey the idea that creativity is a valuable thing, and who provide appropriate feedback and reinforcement for creative behavior. The last of these is probably the most complicated because certain kinds of reinforcement may in fact undermine creativity. Creative behavior is frequently, though not always, tied to intrinsic motivation. This is the motivation that comes from within the individual. Frequently extrinsic factors, including rewards and incentives, inhibit intrinsic interests. This is known as overjustification. Hence, a child must be allowed to explore his or her own interests, and it is unwise to excessively reward creative behavior.

The motivational influences are also complicated because creative behavior is tied to originality. This means that it is often unexpected, and children may do something which is creative but very different from what the parents – or teachers – expect or even desire. Yet if it is original, it might be tied to creativity, and at times must be appreciated. This can be quite difficult. Creative children may require a certain amount of tolerance. They may do surprising, unexpected, and original things, and they may be unconventional and highly autonomous. Parents may target mature behaviors and socialize a child to what is conventional within that culture, but some of the time the child needs to experience the latitude for self-expression. The idea of postconventional behavior is very relevant here: parents and teachers may both attempt to support that balance of originality and socially appropriate behavior. Also relevant is that adults sometimes value creativity in the abstract but have difficulty in supporting the specific behaviors and traits which contribute to creative behavior (e.g., unconventionality, autonomy, nonconformity, and intrinsic motivation). The problems may be especially acute in a classroom, where the educator has not one or two children but several dozens, which makes it particularly difficult to support independence and originality.

Both parents and teachers may be creative models for children. Remote models may be found in books and entertainment media. These models are individuals who a child never actually meets but with whom the child does have some sort of experience. All models probably exert two different kinds of influence: children may learn specific creative behaviors by interacting with creative adults, but they may also internalize the values which are implied by the fact that the adult behaves in that fashion. The impact of values may be most clear in cross-cultural studies, for certain cultures emphasize harmony and collectivism, while others more clearly support creativity because their values emphasize individuality and originality.

Family Structure

Some of the clearest developmental influences on creative behavior are those which are determined by family structure. Family size, for instance, may be related to creative behavior, with large families possibly having an advantage. There are very clear data indicating that small families often have children who do well academically and in conventional areas of performance. Yet, academic performance and conventional or convergent thinking are distinct from the divergent and original thinking which is more typically associated with creative behavior, and divergent and original behavior seem to be better nurtured in large families.

The research on birth order is even more clear-cut. More often than not, firstborn children tend to develop a need for achievement in conventional areas, and, as a result, tend to do very well academically and in similar conventional areas. Second-born and many middle children, on the other hand, apparently attempt to obtain their individuality and develop strategies which ensure that they too can earn attention and obtain resources within the context of the family, even though they are younger than their older brother or sister. This situation has been compared to the evolutionary ecology which influences species diversification. If two species share an environment, they are the least likely to compete for resources if they are different from one another. Perhaps one is nocturnal and the other diurnal. If both are nocturnal, however, competition is likely. Something similar holds true in families. If two siblings are similar, they probably will compete, and often the older of the two will have an advantage because of his or her age and maturity. However, if a second child puts his or her effort into a unique area, that child will be seen as an individual and is much more likely to receive attention and other resources from the parents. One key point in this theory is that eldest children tend to learn the most directly from their parents and are, more often than not, the most conventional of the children within a family. This may in part reflect the fact that the eldest children were at one point only children, and, without any effort, received all of their parents' discretionary attention. When the second child was born, however, the child very likely had to put some effort into obtaining even a small portion of that same discretionary attention. There is a cute and helpless sibling demanding much of it! How can the eldest child regain parental attention? By achieving in conventional areas – hence the need for achievement. What direction does this leave the second or middle child? The unconventional area, which means that the child will probably put most effort into nontraditional and perhaps original areas. If the firstborn is academically astute, for instance, the second born may not put much effort into school but instead may perceive sports or some creative domain.

With this in mind, it is not surprising that Frank Sulloway found very clear-cut tendencies for revolutionaries to be middle children. He referred to the second child with a family as born to rebel. Such rebellion may take the form of creative behavior.

Final Comments

The best way to view the development of creativity is as a fulfillment of potentials. If potentials are fulfilled, actual manifest creative action and performance are likely. The individual and society will both benefit. Creativity has, for example, been tied to psychological and physical health, adaptability, problem solving, invention, entrepreneurship, and innovation.

There is reason to believe that we can be optimistic about creativity because in some ways the fulfillment of potential seems to depend on intentions and decisions. Think back on the idea of postconventionality, for example, and the discretion exercised when the individual recognizes conventions but ignores them in order to

express an original idea. Consider also the old-age style and artists' decisions to make changes, which are not necessary except in the sense that it keeps them out of ruts and routine. Many other examples of the importance of choice and decision making are available, but the point is that, although genetics set limits, each of us has a range of potentials and can fulfill those potentials and be as creative as possible by making the right choices as we grow older.

Further Reading

- Lindauer, M. S. (1992). Creativity in aging artists: Contributions from the humanities to the psychology of aging. *Creativity Research Journal* 5, 211–232.
- Runco, M. A. (2006). *Creativity: Theories and Themes: Research, Development, and Practice*. San Diego, CA: Academic Press.
- Runco, M. A. and Charles, R. (1997). Developmental trends in creativity. In Runco, M. A. (ed.) *Creativity Research Handbook*, vol. 1, pp 113–150. Cresskill, NJ: Hampton.
- Sulloway, F. (1996). *Born to Rebel*. New York: Pantheon.
- Torrance, E. P. (1968). The forth grade slump. *Gifted Child Quarterly* 18, 168–170.

Knowledge Domains and Domain Learning

L Maggioni and P A Alexander, University of Maryland, College Park, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Conditional knowledge – A state of knowledge that pertains to when and where knowledge (declarative or procedural) could or should be applied. For this reason, it has been described as “knowing when and where”.

Declarative knowledge – Involves factual information and is the state of knowledge referred to as “knowing what”.

Disciplinary knowledge – Represents the specialized knowledge associated with a foundational discipline (e.g., history or mathematics) and includes a specific taxonomy, vocabulary, concepts, theories, research methods, and standards of justification aligned with that discipline. It is knowledge of an academic discipline that is taught; a specialized field, or study, or particular branch of learning.

Discourse knowledge – A type of conceptual knowledge that is more particularly about language and its use.

Heuristics – Refer to general guidelines and practices that facilitate a systematic approach to a task and thus favor arriving at a solution (e.g., identifying the data of a problem, looking for related problems, and checking the results).

Knowledge domain – Refers to the part of the world investigated by a specific discipline. Such knowledge is characterized as the object (e.g., plants, numbers, or the past) of a specific body of knowledge (e.g., botany, mathematics, or history).

Metacognition – Typically defined as knowledge of knowledge; knowledge about one's cognition and the regulation of that cognition.

Misconception – An erroneous, sometimes naïve, idea or theory that can seriously hinder understanding.

Procedural knowledge – Described as “knowing how.” Procedural knowledge is, in effect, that state of knowledge encompassing certain processes, strategies, or routines.

Schemata – (pl. of schema) Knowledge structures that are interconnected with and embedded in, one another. The term schemata has been associated with the organization of conceptual knowledge or used to represent all one knows about the physical, social, or mental world.

Second-order knowledge – Refers to concepts regarding how historical knowledge is generated (i.e., the doing of history), such as evidence, accounts, change, and cause.

Skill – A procedure that has been routinized or habitualized, a mental habit that prompts specific actions in front of simple and highly familiar situation, usually operating at an unconscious level.

Strategy – A mental operation or technique, employed to solve problems or to enhance performance. Some strategies are called general, because they can be successfully applied across various domains, while others are called specific, because they target a specific domain or task within a domain.

Substantive knowledge or first-order knowledge – Refers to concepts that identify clusters of kinds of things in the world and, as such, they refer to what a discipline is about (e.g., market, migration, revolution). Such knowledge implies understanding of the rule that guides the formation of those clusters and being able to identify instances of that rule across time and space.

Although educational psychologists' interest in how knowledge develops within specific domains has fluctuated across time, formal education throughout the world is organized around particular fields of studies, and students soon come to characterize their experience of school in terms of subject matters. Thus, in education, domains seem to make a difference. The term knowledge domain refers to the part of the world investigated by a specific discipline. In other words, the domain can be characterized as the object (e.g., plants, numbers, or the past) of a specific body of knowledge (e.g., botany, mathematics, or history).

Herein, we offer a survey of the development of increasingly specialized disciplines and of the corresponding identification of ever-narrower knowledge domains. In particular, we consider the cultural trends that accompanied these changes and the major influences on the structure of formal education. Subsequently, we turn to consider the characteristics of learning and generating knowledge within specific domains. Specifically, we consider the findings of educational research in understanding the processes

of learning and teaching in history, mathematics, and science. We also include contributions regarding the processes of reading and writing, given their relevance in the school curriculum and the tendency in educational research to conceptualize these activities as domains of learning in their own right.

The Development of Disciplines

Besides being characterized by its systematic knowledge of a particular domain, each discipline is also distinguished by a specific way of thinking about associated phenomena. Thus, disciplinary knowledge includes a specific taxonomy, vocabulary, concepts, theories, research methods, and standards of justification. Histories of science, such as those offered by Fehl (1966), Hall and Hall (1988), and Libby (1917) illustrate that disciplines developed over the centuries as privileged pathways toward an understanding of almost any topic. The reciprocal influences between the prevalent cultural climate, disciplinary developments, and formal education are well exemplified by cultural histories of Western education, such as the one composed by Butts (1955).

These studies concur in affirming that organized bodies of knowledge arose in conjunction with the human needs of gaining understanding of the world, establishing some control on the physical environment, and organizing social life. In ancient Egypt, for example, efforts toward controlling and predicting the floods of the Nile fostered remarkable advances in geometry and astronomy. However, it was in the Greek cities of Asia Minor and, later, Athens that the investigation of philosophers into the nature and origin of the universe introduced the method of rational inquiry that so deeply influenced the development of knowledge in Western civilization.

In fact, Western thought owes much of its systematization of knowledge to the Greeks and to their reliance on critical reason in speculating about the origin and nature of the universe. The Greeks came to regard rationality as the human faculty allowing the acquisition of knowledge and truth; hence, the intellectual formation afforded by mathematics and philosophy was considered a privileged road to knowledge and learning and acquired centrality in education.

Over time, increased complexity in the knowledge and skills characterizing competence in specific areas fostered the development of formal instruction and the establishment of schools as separate institutions. The process of organization of knowledge was greatly favored by reliance on the written form, a practice adopted relatively soon by the liberal arts, but not by the practical arts. In the practical arts, informal apprenticeship remained the prevalent way of passing technical skills from one generation to the next. On the other hand, formal education tended to focus

on those bodies of knowledge that were systematically organized in written form.

The importance of accurate definitions in furthering thoughts and the power accorded in Athenian life to those who could speak effectively prompted the development of grammar and rhetoric. The need to think clearly encouraged the development of logic or dialectics. In addition, Greek philosophy began to organize into bodies of knowledge studies regarding the ultimate reality of things (metaphysics), the theory of knowledge (epistemology), human nature and human conduct (ethics, political science, economics, sociology, and psychology), the physical world (astronomy, geography, physics, mechanics, hydraulics, mineralogy, and botany), and the living world (zoology, physiology, and anatomy). The search for intellectual discipline and the problems investigated by the developing sciences favored the progress in the field of mathematics and the refinement and systematization of theories and concepts in arithmetic, geometry, and trigonometry. In addition, Greek inquiry extended to the human past, investigating the causes of historical events and thus laying the foundation of history. Finally, although the fine arts were not organized into systematic bodies of knowledge, the Greeks cultivated several forms of literary criticism, thus giving systematic organization to the theory of art and esthetics.

Roman culture widely drew from Greek thought, furthering the systematization of knowledge and its organization in forms suitable for teaching. By the end of the fourth century, the liberal arts had been circumscribed to the study of compendia (written in Latin) of those Greek works deemed suitable to the spiritual and intellectual development of the pupils. Specifically, the trivium, that is, the elementary liberal art curriculum, included grammar, rhetoric, and logic. The higher liberal arts, or quadrivium, mainly incorporated mathematical studies and comprehended arithmetic, geometry, astronomy, and music. The boundaries of these disciplines did not necessarily include the same content that they comprise in modern times. For example, grammar also involved the study of poetry and literature. In addition, the content reflected changes in the Roman intellectual climate. For instance, the purpose of rhetoric changed from preparation to active participation in the debates over public policies to the study of elegant language to be employed in public celebrations. At the same time, the Romans did not tend to value the sciences for their own sake, but applied the theoretical knowledge gained from the Greeks in geometry, astronomy, and natural philosophy to the solution of practical problems.

The disciplines included in the trivium and the quadrivium continued to constitute the backbone of knowledge during the Middle Ages. However, many changes were made in each of the seven disciplines in response to the varying needs and interests of the time. Latin increasingly

became the language of educated people; hence, grammar gained preponderance among the elementary liberal arts during the Early Middle Ages when the non-Latin people began to take part in the intellectual life of Europe. Nonetheless, Medieval Latin became increasingly different from classical Latin, reflecting the influence of the various languages spoken by the European peoples. Rhetoric lost most of its celebratory purpose and focused on the use of the written language for drawing up legal and feudal documents. Logic was increasingly identified with the rules of deductive thinking and, essentially, became distinguished from philosophy.

Arithmetic, geometry, and astronomy saw important developments during the Middle Ages. The cultural exchanges with the Arab world and the translation of the most important Greek, Arab, and Hindu works into Latin laid the foundations for the scientific and mathematical investigations of later centuries. Within the quadrivium, music retained its theoretical nature, even if its performance gained importance in medieval life. Although knowledge of the Greek tradition was praised by Christian philosophers for the contribution it gave to truth, and thus as an aid to the understanding of God and the soul, the intellectualism of pursuing knowledge for knowledge's sake was rejected. In the Benedictine tradition, where the monastery became the home of practical agriculturalists, as well as of religious, artists, and scholars, the alliance between learning and concrete reality foreshadowed the importance of factual knowledge and the relation between science and technology that came to characterize modern science.

Beginning in the twelfth century, teachers and students began to organize themselves into guilds for protections against the king, the bishop, or anyone else who tried to control them. Over time, the term *universitas* began to refer specifically to guilds of teachers (faculties) and students. Reflecting the articulation of knowledge of the time, the typical guilds of teachers were the faculties of liberal arts, law, medicine, and theology. Specific universities became famous for one specific faculty, thus deeply influencing the development of a particular discipline. In addition, the university system fostered an expansion of the liberal arts curriculum, adding the works of Aristotle on the physical sciences, ethics, politics, and metaphysics. The system also promoted the institutionalization of the educational curriculum (with its degrees, licenses to teach, exams, and titles) and, thus, the grouping of studies into separate faculties.

The Renaissance celebrated the return to the classics (Latin, Greeks, and Hebrew) and highly regarded rhetoric as a way to cultivate polite letters and expression. Logic fell in disrepute due to the humanist opposition to scholasticism. However, these changes had a deeper effect on the content of the trivium and quadrivium than on the disciplines taught in the universities, where medieval

philosophy conserved its predominance. In the sciences, humanists turned away from the deductive methods of argumentation and advocated the use of the inductive method (observations of facts and generalization).

The belief that the method of induction is the proper method to gain scientific knowledge is at the root of the classification of human knowledge provided by Francis Bacon in the seventeenth century. Here, the disciplines came to be mainly characterized as histories of nature, collections of descriptions regarding a vast array of natural phenomena. At the same time, a method common to all sciences began to take shape. The scientist should observe nature, collect facts, identify their common qualities, and express these similarities in general formulas. Empiricism highlighted the patient work of scientists in acquiring facts. However, it dismissed the guiding role of theory in deciding observations and experiments. Although the almost exclusive reliance on induction disappeared in later work on method by Descartes, the uniqueness of the scientific method to gain knowledge was not challenged. In fact, the certitude granted by mathematics was upheld as the goal of scientific knowledge, and the scientific method was considered applicable and appropriate to all fields of human thought.

Whereas in the physical sciences this new method of gaining knowledge proved compatible with the emergence of a unitary principle of explanation, that is, the mechanism, a close relation between abstract thought and scientific investigation failed to surface in the biological sciences. The complexities involved in studying living things and the philosophical impossibility of reconciling the existence of human soul with a completely mechanistic physiology supported the specificity of different bodies of knowledge.

The search for natural laws was extended to the study of society and political economy during the era of Enlightenment. The rules for scientific data gathering also began to be applied in the field of the social sciences, and historians began to identify progress as the fundamental law of history. This process was furthered by the role played by Darwinism in the biological science, where the process of change assumed a central explanatory role. From biology, the idea that change is an inherent part of natural and human development influenced the social sciences and their methodological approach, which increasingly tried to emulate the scientific approach of the natural sciences. In the nineteenth century, the rise of positivism in philosophy further promoted the assumption that reality obeys general and universal laws. The purpose of the disciplines became to discover by observation and experimentation relationships able to explain nature, the universe, human nature, and social institutions. Psychology, anthropology, and scientific medicine were all deeply influenced by this way of thinking.

In the nineteenth century, the effect on education of the increasingly important role attributed to the sciences in the overall knowledge landscape was delayed by the lingering humanist belief that a truly liberal education has to be strongly based on a deep acquaintance with the classics. Secondary instruction was particularly successful in protecting this view; in contrast, in the universities, and especially in Germany, where professors and students were recognized for their remarkable degree of independence and freedom, mathematics and science became the dominant studies during the second half of the nineteenth century.

In the American colleges and universities, the attempt to extend to all fields of knowledge the application of the scientific method and the rise of professional organizations of scholars and specialists in various fields encouraged the subdivision of traditional bodies of knowledge into specialized subjects. What, at the beginning of the century, was studied under the label of natural history was, by the end of the century, subdivided into the various biological and natural sciences, thus including botany, zoology, physiology, psychology, paleontology, ornithology, entomology, and anthropology. Natural philosophy was articulated into specialized physical sciences (astronomy, physics, chemistry, mineralogy, geology, meteorology, and physical geography); similarly, history, economics, political sciences, sociology, and anthropology acquired their own specificity within the field of moral philosophy.

The scientific discoveries of the twentieth century in astronomy and physics challenged the positivistic assumptions of a rigid and indestructible matter, obedient to rigid laws. More generally, these new insights questioned the close correspondence between what the universe is really like and the picture rendered by science at a certain point in time, and brought the debate about the epistemic status of scientific knowledge to the forefront. At the same time, and almost paradoxically, the process of disciplinary subdivision went even further as scholars tended to specialize in ever-narrower aspects of one discipline. However, as noted by Easton and Schelling (1991), real-world problems are rarely confined to a specific knowledge domain and the parceled understandings afforded by the increasing specialization do not easily reassemble into a unitary view of the issue at hand. Proposed solutions include interdisciplinary training and teamwork within specific research topics, although the point of departure for these attempts at integration tends to remain the specialized knowledge granted by the various disciplines.

On one hand, the overview of the development of disciplinary knowledge showed that a certain partition of the world in different domains reflects broad cultural and institutional trends. On the other hand, the nature of the object investigated makes specific methodological choices more effective and fruitful than others, thus characterizing

each discipline as a specific way of knowing. In the next section, we turn to consider the influence that these differences in thinking have on learning and teaching.

Domain Learning

Do teaching and learning differ across domains? As documented by Shulman and Quinlan (1996), educational psychologists' answer to this question has changed considerably across time. During the first two decades of the twentieth century, the answer has been mainly in the affirmative. In particular, Dewey (1902/1916) advocated the need to psychologize the subject matter of the studies by making explicit the research work that generated knowledge in a specific domain and by referring it to the present experience of the child. His experimental research was thus located within the naturalistic setting of the laboratory school.

In the next decades, following the lead of researchers such as Thorndike, educational psychologists' interest shifted toward the search for general theories of learning. Even when their research was nested within a particular subject matter, the general theoretical framework constrained the analysis of the disciplinary tasks. In addition, the controlled experiment in laboratory setting became the preferred methodological approach, further detaching psychological research from the educational setting.

Beginning in the 1980s, subject matters regained centrality. For instance, studies on classroom teaching increased understanding of the importance of pedagogical content knowledge. Further, comparisons between novices and experts uncovered the role played by heuristics typical of a certain discipline in the performance on domain-specific tasks. Moreover, the classroom, with its complexities, came to be considered a viable and preferred setting for studying these issues. Overall, researchers found that performance of domain-specific tasks calls for particular psychological processes. These findings spurred investigations of what pedagogical practices can best favor the development of these processes and thus improve students' learning.

In the ensuing sections, we consider some of the outcome of this research for the domains of reading, writing, history, mathematics, and science. Traditionally, these domains occupy a large portion of the educational curriculum, especially in the early years. Thus, it is not by chance that a large body of research about domain learning focused on these areas. We also mention a few controversial issues regarding the translation of research results into pedagogical practice. A general review of the work of educational psychologists in regard to learning within these specific domains may be found in the relevant chapters of the first and second editions of the *Handbook of Educational Psychology* (Alexander and Winne, 2006; Berliner and Calfee, 1996). Alexander (2006) offers an introductory overview of these topics.

Reading

Theories of learning have played a critical role in identifying what is meant by reading and writing, and, more generally, by literacy. Thus, definitions of reading span from the ability to decode (i.e., breaking the linguistic code) and encode (i.e., convert written signs fluently into meaning) to being well learned in a variety of topics as well as in a set of cultural practices (with more or less emphasis on the power structure engrained in them). The method chosen to study the process of reading and writing contributed to influence the definition of literacy also. For example, the exclusive focus on observable behavior characterizing research in the behaviorist tradition precluded the possibility to study understanding. Thus, these researchers mainly focused on handwriting, grammar, word recognition, and perception of print. Further, assuming that meaning is inherent in the text and the individual's role consists in uncovering it, understanding was mainly studied by observing vocabulary and recall. The pedagogical implications of this approach, still present in current curricula, include the suggestion to break the reading process into steps and to teach reading as a series of skills and subskills.

As well exemplified by Bruner (1990), the cognitive revolution focused on explaining mental processes; thus, understanding how meaning was generated became crucial. However, within the cognitive tradition, researchers conceived meaning in different terms. Specifically, the researchers working within the framework of information-processing theory assumed that meaning is transported from the author to the readers, while constructivists posited that meaning is constructed by the readers on the basis of information provided by the author.

This difference notwithstanding, cognitive psychologists fostered understanding of how individuals make sense of information conveyed by texts. By investigating the nature of readers' schemata and how information was organized in memory, researchers studied the role of background knowledge (Anderson and Pearson, 1984). Other studies focused on how different texts work, paying attention to their various structures (e.g., narrative or expository; Alexander and Kulikowich, 1994). Finally, research on the control that individuals maintain during meaning making has explored individuals' knowledge of their own cognitive processes (i.e., metacognition). In particular, researchers investigated individuals' knowledge of the tasks and goals required by reading or writing (i.e., declarative knowledge), of the strategies that allow one to pursue these goals (i.e., procedural knowledge), and individuals' awareness of how, when, and where to use a specific strategy (i.e., conditional knowledge; Garner, 1987). This work provided the background for research in strategy instruction and, limits notwithstanding, shifted the attention of educational psychologists from product to process.

Finally, cognitive psychologists fostered understanding of two basic processes of reading: phonemic awareness and automaticity. Phonemic awareness is the ability to think and manipulate sounds and plays a central role in the study of reading development. Automaticity regards the ability to process perceptual information necessary to the decoding of print with a minimum cognitive load, thus freeing attention for meaning-making purposes; within this process, accuracy and speed emerged as good predictors of comprehension (Stanovich, 1990). These findings suggest the soundness of a balanced reading program, in which mastering basic processes and meaningful engagement with text reinforce each other.

The social constructivist perspective highlighted the social dimension of learning. Thus, the definition of text was extended beyond the printed words to include conversation, media, and, more generally, social discourse. This perspective also implied a crucial epistemological shift since it defined knowledge as the consensus reached by the community of knowledgeable peers for the time being. Thus, according to this view, the external world may exist, but knowledge is not defined by a correspondence to it anymore (Bruffee, 1986). Such a view also suggested that thinking is an internalized version of conversation; thus, learning happens first on the social plane and, subsequently, is internalized by the individual (Vygotsky, 1934/1986). Pedagogically, this perspective advocated practices such as reciprocal teaching and whole language approach. This framework also drew attention on the influence of the context on the meaning-making process.

Writing

The importance of writing for communication, learning, and knowledge transmission can hardly be overstated. However, the scientific study of the process of writing is relatively recent (Graham, 2006). In the previous section, we considered the influence of different theoretical approaches on the study of literacy in general and described their pedagogical suggestions in terms of reading. In this section, we focus on the body of research that specifically addressed the writing process and on its pedagogical implications.

Mainly relying on analyses of think-aloud protocols collected from individuals of different ages and levels of expertise while composing texts, researchers in this domain emphasized that writing is a self-directed process. Often proceeding in a nonlinear fashion, this process requires individuals to simultaneously attend to several cognitive demands and, thus, entails a high level of effort. Although most of the research focused on cognitive processes, motivational and contextual factors were also found to influence writing (e.g., competence beliefs and environmental support).

Most theorists view writing as the interplay of three main components (Hayes and Flower, 1980). The first comprises factors external to the writer, such as topic, audience, and text produced so far. The second component regards the cognitive processes involved in writing and includes planning (e.g., setting goals, generating ideas, and organizing ideas into a plan), text production (e.g., translating plans into a written text), and reviewing (critically reading the text, determining how to address emerged problems, and editing). The final component is the writer's long-term memory, which includes knowledge of the topic, of the intended audience, and of rhetorical devices, as well as general plans to perform the writing task.

This body of research suggests that teaching writing should address all its components (Hillocks, 1986). In particular, direct teaching of strategies for planning and revising has proven particularly effective (Graham, 2005). At the same time, mastering of basic skills such as handwriting and spelling is also crucial to allow individuals to attend to the multiplicity of writing demands.

The learning environment can sustain students' development of writing strategies by offering appropriate scaffolding and peer interactions. Finally, increased awareness of the processes of reading and writing, of their differences as well as of their similarities, can foster the development of programs that facilitate development in both domains (Shanahan, 2005).

History

Although there is general agreement in identifying the domain of history with the past, psychological research has associated strikingly different processes to learning in history. In the first decades of the twentieth century, for example, learning history was viewed as the acquisition of temporal perspective and moral ideals (Hall, 1911), the maturation of chronological and causal thinking (Judd, 1915), and the development of historic sense (Bell, 1917).

Educational researchers, overall, agreed that learning history could not be reduced to answering factual questions; however, they also realized that this component of learning was the easiest to test. Behaviorism further restricted the study of domain-specific topics; in the case of history, research was limited to study how to apportion facts in order to facilitate memorization. It was only with the cognitive revolution that the attention of researchers focused on the psychological processes involved in learning history (Wineburg, 1996).

Aided by the use of qualitative methodologies, educational psychologists explored learners' background knowledge in an attempt to uncover beliefs and conceptions that may have fostered or hindered thinking in history (e.g., ideas about time and chronology, sparse information about historical people and events, and

beliefs about the nature of history). They also studied how historians and novices generated historical understanding while reading primary and secondary sources about specific events (Wineburg, 2001).

This body of research suggested that learning history entails developing familiarity with concepts and ideas that allow a description of the past (i.e., substantive knowledge) and also becoming acquainted with the strategies employed by historical inquirers to research and interpret the past (i.e., procedural knowledge). In particular, the development of historical knowledge includes not only being able to answer who, what, when, where, and how questions about the past (i.e., first-order knowledge), but also developing an understanding of concepts such as causation, change, historical significance, empathy, evidence, and account that allow historical investigators to interpret the past (i.e., second-order knowledge; Lee, 2005).

By comparison, procedural knowledge regards being able to use strategies such as assessing the status of sources, corroborating sources, contextualizing events, constructing evidence-based arguments, and writing accounts (VanSledright and Limón, 2006). The difference of these processes with the experience of memorization of information commonly associated to learning history is striking.

Although research in the past 20 years has increased understanding of the process of learning history and highlighted its specificity, the intense debate about the role of history in the school curriculum makes it difficult to translate these findings into pedagogical practice. In schools, the goal of developing historical thinking is often countered by the will to use history to build a collective, national identity. While the first purpose is well served by pedagogical approaches that foster a view of history as a critical inquiry into the past, the second is better pursued by the transmission of a specific narrative viewed as coinciding with the past. In addition, more research is needed to understand how to prepare teachers to be able to foster historical thinking in their students.

Mathematics

Historically, the characterization of mathematics evolved from a science of numbers and space to a science of patterns. More recently, the view of mathematics as a human activity defined within historical, cultural, and social contexts also emerged. In psychological research, specific aspects of mathematics learning were often used for studying general theories and pedagogical approaches. However, the specificity of thinking within this domain was usually ignored (De Corte *et al.*, 1996).

As surfaced in other domains considered herein, different psychological theories of cognitive development espoused distinctive views of mathematics, used a various array of research methodologies, focused on particular aspects of mathematics learning, and thus reached

different pedagogical conclusions. Behaviorism and connectionism, for example, favored drill and practice of well-defined information, skills, and associations.

The view of mathematics learning changed remarkably with the cognitive revolution, shifting the focus on the processes involved in thinking mathematically. In characterizing the domain, researchers highlighted the dual nature of mathematics. Rooted in the perception and description of the order of objects and events, once mathematics succeeds in modeling these structures through a process of symbolical representation, these representations become amenable to study in an abstract fashion, independent from their real-world roots. The pedagogical consequence of acknowledging this feature of the domain is a shift of focus from computation to modeling (i.e., the thinking of the mathematician; Davis, 1992).

A second characteristic of the domain regards the fact that, historically, mathematical concepts evolved hierarchically through a series of restructuring of previously developed concepts, definitions, and functions. The parallel with the Piagetian characterization of cognitive growth through a series of restructuring engendered by a situation of disequilibrium perhaps explains why mathematics was so often used by cognitive psychologists to study general cognitive development. Pedagogically, this feature of the domain supports practices that help the child relive the developmental process of the discipline (Freudenthal, 1991). This indication does not suggest that children need to reinvent the product of this development (e.g., the definition of rational numbers), rather that they have to tread the same cognitive path of people facing mathematical problems (e.g., abstracting and formalizing).

Finally, development in mathematics requires that individuals are comfortable with multiple modes of representation (e.g., verbal/syntactic, visual/spatial, and formal/symbolic) and fluent in translating from one system to the other. In particular, researchers theorized that mathematical thinking implies the interplay of external representations (e.g., language, symbols on a page, and objects), and internal mental processes, which include internal representations, affective processes, and executive functions (e.g., planning, and monitoring; Goldin, 1992). Pedagogically, this theory suggests that learning mathematics means fostering the development of internal representations well connected and consistent with conventional external representations; fluid movement across different systems of representation also needs to be promoted.

In schools, mathematics is often associated to doing mathematics rather than thinking mathematically. Thus, problem solving takes a central place in the curriculum and a large body of research in mathematics focused on investigating this process. From these studies, it emerged that problem solving involves the interplay of four factors:

domain-specific knowledge, heuristic methods, metacognitive knowledge, and affection (beliefs and emotion).

Domain-specific knowledge refers to definitions, formulas, symbols, algorithms, and concepts typical of the domain. Although this knowledge in experts is well organized and thus flexibly accessible, misconceptions and defective skills often hinder problem solving in novices. In addition, experts tend to categorize problems according to their mathematical structure, whereas novices focus on problem surface characteristics (Chi *et al.*, 1988; Confrey, 1990).

Heuristics are general guidelines that facilitate a systematic approach to the task and thus favor arriving at a solution (e.g., identifying the data of the problem, looking for related problems, and checking the results). However, knowledge of isolated heuristics is usually not very helpful. Research suggests that a more successful approach requires teaching heuristics concurrently with metacognitive skills and while exposing students to a variety of situations, so that learners may understand when and how to use a certain heuristic (Schoenfeld, 1992).

Metacognitive knowledge regards knowledge of one's own cognitive functioning and self-monitoring of these cognitive activities. Skilled problem solvers demonstrate high control of their actions, including planning and monitoring, and, if necessary, make the required corrections to a previously implemented strategy. Finally, affective components of problem solving include beliefs about the self and about mathematics, and emotions such as interest in the task (McLeod, 1990). Research on problem solving has shown that it is possible to teach students to plan and monitor more effectively (Schoenfeld, 1985). In addition, when the learning environment is structured in such a way as to favor reflective practice and talk among students about their thinking, mathematical learning is usually improved (Lampert, 1990).

Science

Reflecting a trend evidenced in other domains, the definition of what it means to learn science and of what psychological processes are central to its development has been influenced by theories of cognitive development. The meaning of learning science has increasingly broadened to include an understanding of substantive concepts and of the nature of science, logical reasoning, procedures used to develop scientific explanations, and metacognitive awareness (Linn *et al.*, 1996).

Comparisons between expert and novice thinking demonstrated that novices' knowledge is organized around concrete factors (e.g., formulas), while experts' knowledge tends to be hierarchically organized around abstract elements (Chi *et al.*, 1981). Beginning in the late 1970s, researchers increasingly distinguished between the

development of logical skills and that of scientific concepts (Pfundt and Duit, 1991). In addition, research extended beyond the acquisition of science information to investigate learners' ideas about the nature of science (Hestenes, 1992). The acknowledgment of the role of metacognition in learning paralleled this shift in focus, suggesting that monitoring one's own cognition is an important component in learning science, once one abandons the idea that science consists of universal truth contained in textbooks (White, 1988).

The interpretation of students' misconceptions also changed; previously conceptualized as instances signaling faulty reasoning on the students' part, these ideas began to be interpreted as alternative intuitions or framework to model the world. Researchers noted that students' misconceptions may originate from inaccurate implications drawn from accurate observations, with a focus on inessential characteristics of phenomena, or on the adoption of standards of evidence and views of science markedly different from those espoused by the scientific community (Kuhn *et al.*, 1988).

From these insights, two lines of research developed; the first focused on conceptual change and the second, on the restructuring of knowledge. Consistent to the Piagetian theory, conceptual change tended to be promoted by using cognitive conflict as a mechanism to foster scientific understanding. More recently, researchers studied what can motivate students to change their current understandings and found that, in addition to facing disconcerting evidence about their current concepts, students also need to confront clear alternatives (Strike and Posner, 1985). This new approach prompted teachers to focus on fostering and guiding the reasoning process of students while they investigate phenomena, rather than contradicting the conclusions they reach.

The investigation of experts' organization of scientific knowledge showed that scientists often use qualitative models (e.g., free-body diagrams) as an aid in problem solving. They also tend to chunk knowledge in patterns that can be used in a variety of situations, developing production rules that link a condition (e.g., the object in contact has different temperature) to a consequent action (e.g., the objects will tend toward equilibrium). In addition, experts tend to entertain several models of scientific phenomena and to choose which one to rely on according to the specific problem they face (Reif and Larkin, 1991).

Studies of novices found that students often entertain conflicting interpretations of scientific phenomena. Some researchers hypothesized that students' scientific knowledge tends to be fragile and fragmented (diSessa, 2002); other studies showed that students tend to contextualize their views, applying one set of ideas to interpret phenomena within the classroom context and another set to deal with the same instance out of class (Gilbert and

Boulter, 2000). In addition, novices' ability to generalize a model across a range of experiences is usually limited.

These findings supported a change in the goals set for science education, with the focus shifting from acquiring science information and concepts to developing the ability to scientifically reason about phenomena (Linn and Eylon, 2006). Pedagogically, this view suggests that students are exposed to a variety of explanatory models and provided with criteria for selecting among them according to the problem considered. Far from engaging students in inquiry or discovery activities for their own sake, this approach highlights the development of background knowledge, modeling and discussion of key processes and strategies, instructional guidance, feedback, use of evidence to test one's own ideas, and monitoring of one's progress (Alexander, 2006).

Concluding Thoughts

The roots of current disciplines and domains of study reach well back in history. These areas of knowledge and practice have not only reflected societies and cultures of their time, but have also influenced them, especially through formal educational systems. The differences in domains continue to shape the landscape of academic practice due to their inherent and socially constructed nature. Here, we have explored several of those features that define contemporary domains and their instantiation in educational practice.

See also: Concept Learning; Early Writing; Learning Science; Learning to Read; Mathematics Learning; Reading Comprehension: Reading For Learning; Writing: Advanced.

Bibliography

- Alexander, P. A. (2006). *Psychology in Learning and Instruction*. Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.
- Alexander, P. A. and Kulikowich, J. M. (1994). Learning from physics text: A synthesis of recent research. *Special Issue: Print-Based Language Arts and Science Learning. Journal of Research in Science Teaching* **31**, 895–911.
- Alexander, P. A. and Winne, P. H. (eds.) (2006). *Handbook of Educational Psychology*, 2nd edn. Mahwah, NJ: Erlbaum.
- Anderson, R. C. and Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In Pearson, P. D., Barr, R., Kamil, M., and Mosenthal, P. (eds.) *Handbook of Reading Research*, vol. 1, pp 255–293. New York: Longman.
- Bell, J. C. (1917). The historic sense. *Journal of Educational Psychology* **8**, 317–318.
- Berliner, D. C. and Calfee, R. C. (eds.) (1996). *Handbook of Educational Psychology*. New York: Simon and Schuster/Macmillan.
- Bruffee, K. A. (1986). Social construction, language, and the authority of knowledge: A bibliographic essay. *College English* **48**, 773–790.
- Bruner, J. (1990). *Acts of Meaning*. Cambridge, MA: Harvard University Press.

- Butts, F. R. (1955). *A Cultural History of Western Education: Its Social and Intellectual Foundations*, 2nd edn. New York: McGraw-Hill.
- Chi, M. T., Feltovich, P., and Smith, E. L. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science* 5(2), 121–152.
- Chi, M. T., Glaser, R., and Farr, M. J. (eds.) (1988). *The Nature of Expertise*. Hilldale, NJ: Erlbaum.
- Confrey, J. (1990). A review of the research on student conceptions in mathematics, science, and programming. In Cazden, C. B. (ed.) *Review of Research in Education*, vol. 16, pp 3–55. Washington, DC: American Educational Research Association.
- Davis, R. B. (1992). Reflections on where mathematics education now stands and on where it may be going. In Grouws, D. A. (ed.) *Handbook of Research on Mathematics Teaching and Learning*, pp 724–734. New York: Macmillan.
- De Corte, E., Greer, B., and Verschaffel, L. (1996). Mathematics teaching and learning. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 491–549. New York: Simon and Schuster/Macmillan.
- Dewey, J. (1902/1916). *The Child and the Curriculum*. Chicago, IL: University of Chicago Press.
- diSessa, A. A. (2002). Why “conceptual ecology” is a good idea. In Limón, M. and Mason, L. (eds.) *Reconsidering Conceptual Change: Issues in Theory and Practice*, pp 29–60. Dordrecht: Kluwer.
- Easton, D. and Schelling, C. S. (eds.) (1991). *Divided Knowledge: Across Disciplines, across Cultures*. Newbury Park, CA: Sage.
- Fehl, N. E. (1966). *Science and Culture*. Hong Kong: Chung Chi Publications.
- Freudenthal, H. (1991). *Revisiting Mathematics Education*. Dordrecht: Kluwer.
- Garner, R. (1987). *Metacognition and Reading Comprehension*. Norwood, NJ: Ablex.
- Gilbert, J. K. and Boulter, C. J. (2000). *Developing Models in Science Education*. Dordrecht: Kluwer.
- Goldin, G. A. (1992). On developing a unified model for the psychology of mathematical learning and problem solving. In Geeslin, W. and Graham, K. (eds.) *Proceedings of the Sixteenth Annual Meeting of the International Group for the Psychology of Mathematics Education*, vol. 3, pp 235–261. Durham, NC: University of New Hampshire Press.
- Graham, S. (2005). Strategy instruction and the teaching of writing: A meta-analysis. In MacArthur, C., Graham, S., and Fitzgerald, F. (eds.) *Handbook of Writing Research*, pp 187–207. New York: Guilford.
- Graham, S. (2006). Writing. In Alexander, P. A. and Winne, P. H. (eds.) *Handbook of Educational Psychology*, pp 457–478. Mahwah, NJ: Erlbaum.
- Hall, A. R. and Hall, B. M. (1988). *A Brief History of Science*. Ames, IA: Iowa State University Press.
- Hall, G. S. (1911). *Educational Problems*, vol. II. New York: Appleton.
- Hayes, J. and Flower, L. (1980). Identifying the organization of writing processes. In Gregg, L. and Steinberg, E. (eds.) *Cognitive Processes in Writing*, pp 3–30. Hillsdale, NJ: Erlbaum.
- Hestenes, D. (1992). Modeling games in the Newtonian world. *American Journal of Physics* 60(8), 732–748.
- Hillocks, G. (1986). *Research on Written Composition: New Directions for Teaching*. Urbana, IL: National Council of Teachers of English.
- Judd, C. H. (1915). *Psychology of High-School Subjects*. Boston, MA: Ginn.
- Kuhn, D., Amsel, E., O’Loughlin, M., and Schauble, L. (1988). *The Development of Scientific Thinking Skills, Developmental Psychology Series*. Orlando, FL: Academic Press.
- Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal* 27, 29–63.
- Lee, J. P. (2005). Putting principles into practice: Understanding history. In Donovan, M. S. and Bransford, J. D. (eds.) *How Students Learn: History in the Classroom*. Washington, DC: The National Academies Press.
- Libby, W. (1917). *An Introduction to the History of Science*. Boston, MA: Houghton Mifflin.
- Linn, M. C. and Eylon, B. (2006). Science education: Integrating views of learning and instruction. In Alexander, P. A. and Winne, P. H. (eds.) *Handbook of Educational Psychology*, pp 511–544. Mahwah, NJ: Erlbaum.
- Linn, M. C., Songer, N. B., and Eylon, B. (1996). Shifts and convergences in science learning and instruction. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 438–490. New York: Simon and Schuster/Macmillan.
- McLeod, D. B. (1990). Information-processing theories and mathematics learning: The role of affect. *International Journal of Educational Research* 14, 13–29.
- Pfundt, H. and Duit, R. (1991). *Students’ Alternative Frameworks*, 3rd edn. Kiel: Institut für die Pädagogik der Naturwissenschaften (Institute for Science Education at the University of Kiel).
- Reif, F. and Larkin, J. H. (1991). Cognition in scientific and everyday domains: Comparison and learning implications. *Journal of Research in Science Teaching* 28(9), 733–760.
- Schoenfeld, A. H. (1985). *Mathematical Problem Solving*. New York: Academic Press.
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In Grouws, D. A. (ed.) *Handbook of Research on Mathematics Teaching and Learning*, pp 334–370. New York: Macmillan.
- Shanahan, T. (2005). Relations among oral language, reading, and writing development. In MacArthur, C., Graham, S., and Fitzgerald, J. (eds.) *Handbook of Writing Research*, pp 171–185. New York: Guilford.
- Shulman, L. S. and Quinlan, K. M. (1996). The comparative psychology of school subjects. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 399–422. New York: Simon and Schuster/Macmillan.
- Stanovich, K. E. (1990). A call for an end to the paradigm wars in reading research. *Journal for Reading Behavior* 22, 221–232.
- Strike, K. A. and Posner, G. J. (1985). A conceptual change view of learning and understanding. In West, L. H. and Pines, A. L. (eds.) *Cognitive Structure and Conceptual Change*, pp 211–231. Orlando, FL: Academic Press.
- VanSledright, B. and Limón, M. (2006). Learning and teaching social studies: A review of cognitive research in history and geography. In Alexander, P. A. and Winne, P. H. (eds.) *Handbook of Educational Psychology*, pp 545–570. Mahwah, NJ: Erlbaum.
- Vygotsky, L. S. (1934/1986). *Thought and Language*. Kozulin, A. (trans.) Cambridge, MA: MIT Press.
- White, R. T. (1988). Metacognition. In Keeves, J. P. (ed.) *Educational Research, Methodology and Measurement: An International Handbook*, pp 70–75. Oxford: Pergamon.
- Wineburg, S. (1996). The psychology of learning and teaching history. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 423–437. New York: Simon and Schuster/Macmillan.
- Wineburg, S. (2001). *Historical Thinking and other Unnatural Acts: Charting the Future of Teaching the Past*. Philadelphia, PA: Temple University Press.

Further Reading

- Donald, J. (2002). *Learning to Think: Disciplinary Perspectives*. San Francisco, CA: Jossey-Bass.
- Elby, A. (2001). Helping physics students learn how to learn. *Physical Education Research: American Journal of Physiology Supplement* 69(7), 54–64.
- Gardner, H. (1999). *The Disciplined Mind: What All Students Should Understand*. New York: Simon and Schuster.
- Harris, R. W. (1960). *Science, Mind, and Method*. Oxford: Basil Blackwell.
- Hashweh, M. (1996). Effects of science teachers’ epistemological beliefs in teaching. *Journal of Research in Science Teaching* 33(1), 47–63.
- Hirst, R. H. (1973). Liberal education and the nature of knowledge. In Peters, R. S. (ed.) *Philosophy of Education*, pp 87–101. Oxford: Oxford University Press.

- Larkin, J., McDermott, J., Simon, D. P., and Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science* **208**(20), 1335–1342.
- Messer-Davidow, E., Shumway, D. R., and Sylvan, D. J. (eds.) (1993). *Knowledges: Historical and Critical Studies in Disciplinarity*. Charlottesville, VA: University Press of Virginia.
- Moll, L., Amanti, C., Neff, D., and Gonzales, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory into Practice* **31**, 132–141.
- Seixas, P. (2000). Schweigen! die Kinder! Or, does postmodern history have a place in the schools? In Stearns, P. N., Seixas, P., and Wineburg, S. (eds.) *Knowing, Teaching, and Learning History*, pp 19–37. New York: New York University Press.
- VanSledright, B. (2002). *In Search of America's Past: Learning to Read History in Elementary School*. New York: Teachers College Press.
- Venezky, R. L. (1984). The history of reading research. In Pearson, P. D., Barr, R., Kamil, M., and Mosenthal, P. (eds.) *Handbook of Reading Research*, vol. 1, pp 3–39. New York: Longman.
- Wilson, S. M. (1990). Mastodons, maps, and Michigan: Exploring uncharted territory while teaching elementary school social studies. *Report No. 24*. Elementary Subjects Center Series, Michigan State University, .
- Wineburg, S. and Wilson, S. (2001). Wrinkles in time and place: Using performance assessment to understand the knowledge of history teachers. In Wineburg, S. (ed.) *Historical Thinking and other Unnatural Acts: Charting the Future of Teaching the Past*, pp 173–214. Philadelphia, PA: Temple University Press.

Problem Solving and Human Expertise

T J Nokes and C D Schunn, University of Pittsburgh, Pittsburgh, PA, USA
M T H Chi, Arizona State University, Tempe, AZ, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Absolute methods – A research methodology that consists of an in-depth examination of high-level experts, usually in a task specific to their domain of expertise (e.g., playing a game of chess for chess experts).

Chunk – A unit of knowledge that is composed of several smaller units of information.

Declarative knowledge – This consists of descriptions about the world including facts, strategies, and principles, and is commonly referred to as knowing that.

Knowledge compilation – A cognitive mechanism hypothesized to interpret declarative knowledge into a set of specific procedural rules given a particular goal.

Procedural knowledge – This consists of information for how to perform particular actions to accomplish task goals, and is commonly referred to as knowing how.

Relative methods – A research methodology that consists of comparing more- to less-experienced participants, often in a neutral task not typically practiced in their domain of expertise (e.g., recalling chess positions for chess experts).

Schema – A hierarchical knowledge structure that includes prototypical information about the type of problem, including declarative knowledge of objects, facts, strategies, and constraints and may also include procedural operators for solving the problem.

Problem solving is a critical cognitive activity that permeates many aspects of our day-to-day lives. We solve problems at home, school, and work ranging from the simple – such as figuring out the tip on a bill – to the complex – such as planning the logistics of a family trip. Problems sometimes have clear goals and steps you can take (as in algebra problems), but sometimes have vague goals or ambiguity about what solution methods are possible. The latter are called ill-structured problems and are considered much more difficult than well-structured problems. Developing expertise in problem solving is critical to the success of a wide range of human activities, including pursuits in science, art, business, and politics. As our society becomes ever-more technologically diverse and

sophisticated, experts are sought in more and more specialized fields. Having a scientific explanation of expert performance is needed to understand its development and to facilitate its acquisition. Knowing what to teach influences the methods of teaching. Expertise research is an area that provides a basis for determining what needs to be taught. Our purpose in writing this article is to provide an integrative review of the psychological research on expert problem solving by taking a close look at what it is, how it is acquired, and the implications for education and instruction.

We structure the article around two interrelated themes. First, that expertise can be understood from an information-processing perspective by focusing on the role of knowledge, its content, and the cognitive processes that bring that knowledge to bear during problem solving. Second, that expert performance is acquired through deliberate practice (Ericsson *et al.*, 1993). This view that expertise can be decomposed into a set of knowledge structures that are learned has implications for how to structure learning environments in order to facilitate its acquisition. In the rest of this article, we explore these themes beginning with a brief review of the methods used to examine expertise, followed by a detailed analysis of how expertise impacts each stage of problem solving. We then review the research on its acquisition with a focus on the underlying cognitive processes. In the final section, we discuss current directions and implications for instruction.

Methods

Researchers have typically used one of two approaches to study expertise, what Chi (2006) has called absolute and relative methods. Absolute methods consist of an in-depth examination of high-level experts, usually in a task specific to their domain of expertise, such as playing a game of chess for chess experts. Defining the level of expertise occurs through established criteria for a particular domain. For some domains, there are written criteria (a rating or scoring system) to determine rank, such as in chess. In other domains, expertise is determined by a certain level of professional achievement, such as becoming a professional ballet dancer, physicist, or a commercial airline pilot. The absolute approach is aimed at providing an in-depth description of the knowledge and cognitive processes underlying

expert performance. This approach includes both observational studies as well as historical analyses of famous cases (e.g., James Clerk Maxwell; Nersessian, 1992). Relative methods involve comparing more- to less-experienced participants, often in a neutral task outside of their domain expertise, such as recalling chess positions for chess experts. The advantage of this approach is that it can uncover the structures and processes of performing the task, and not merely the ways that experts can excel. Both approaches have made extensive use of verbal protocols to obtain detailed data as to the thinking processes that accompany expert (and novice) performance (Ericsson, 2006).

These approaches have produced a wealth of findings on the nature of expertise (for general reviews see the section titled 'Further reading'). In the next section, we draw upon this literature to examine the impact of expertise on problem solving. We begin by describing a general theory of problem solving and then at each stage of the process describe the differences between expert versus novice performance and the explanations to account for those differences.

Expert Problem Solving: Major Findings

Most theories of human problem solving consist of some formulation of the following seven stages:

1. problem categorization,
2. construction of a mental representation of the problem,
3. search for the appropriate problem-solving operators (e.g., strategies or procedures),
4. retrieval and application of those operators to the problem,
5. evaluation of problem-solving progress and solution,
6. iterating stages 1–4 if not satisfied with progress/solution, and finally
7. storage of the solution (e.g., Newell and Simon, 1972).

These stages may not be strictly sequential, but may be iterative. In the following subsections, we describe the expertise findings relevant to each stage and discuss the theories proposed to account for them (see **Figure 1** for an illustration of the problem-solving stages and the impact of expertise on each one).

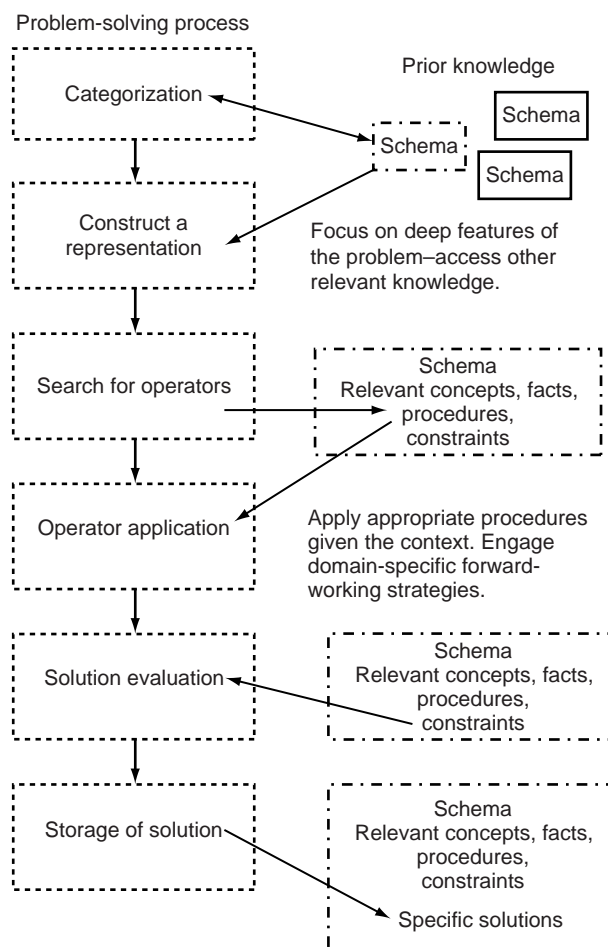


Figure 1 The impact of expertise on each stage of the problem-solving process.

Problem Categorization

The first stage involves the categorization of the problem. This stage is critical as it impacts all subsequent problem-solving processes, such as determining what knowledge to use and what strategies are relevant. For example, after a statistician categorizes a statistics problem as a permutations problem, she or he can proceed by retrieving and applying the appropriate formula to solve it. Much research has shown that experts' domain knowledge actually influences problem perception. When experts are presented a problem or task relevant to their domain of expertise, they see the problem in terms of prior meaningful patterns of information. For example, Chase and Simon (1973a, 1973b) found that expert chess players recalled more than novices on a memory task in which they were briefly presented a game scenario that they had to reconstruct. The experts recalled approximately four times as many pieces as the novices but only for scenarios that were from real games; when the scenario consisted of randomly placed pieces, experts performed at the same level as the novices. It was hypothesized that the experts' prior knowledge facilitated the recognition and recall of domain-relevant patterns, or chunks, of information from the scenarios (see **Figure 2** for an example of chunks in chess). These chunks provide experts useful ways to perceive and reason about large amounts of domain-relevant information.

Similar effects have been shown in research on medical expertise. For example, Lesgold *et al.* (1988) compared expert to novice physicians as they diagnosed X-ray films of the lungs. The physicians were asked to draw on the X-rays to identify the important features of their diagnosis. Both groups noticed abnormalities associated with a collapsed lung. However, experts were much more likely to identify the correct shape and size of the abnormality, whereas novices identified abnormalities that were approximately half the size of those identified by the experts. This work shows that experts and novices can perceive a problem very differently even when looking at the exact same stimulus. This finding that expert knowledge impacts problem perception has been found in a variety of tasks and domains including: architecture (Akin, 1980), mathematics (Silver, 1979), and naturalistic decision-making (NDM) tasks such as a fireman determining the safety of a room in a burning building (Klein, 1998).

A related effect is the finding that experts are more likely than novices to categorize problems at a deep level of abstraction (or function), whereas novices are more likely to categorize problems based on the surface features. For example, in the seminal work by Chi *et al.* (1981) experts and novices were asked to sort physics word problems based on their similarity. Experts sorted them according to their underlying physics principles, such as Newton's second

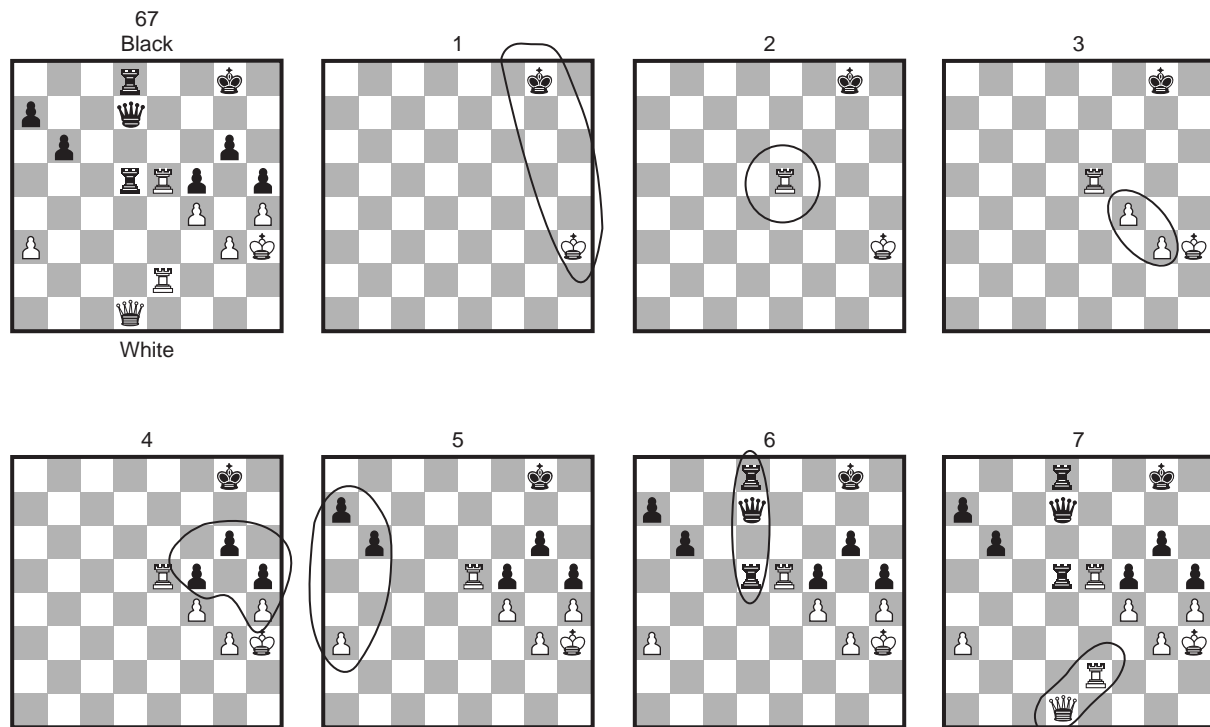


Figure 2 An example of a chess master's chunk-by-chunk recall of position 67 from Reinfeld (1945) with each new chunk circled. From Chase, W. G. and Simon, H. A. (1973). *The minds eye in chess*. In Chase, W. G. (ed.) *Visual Information Processing*, pp 215–281. New York: Academic Press. With permission from Elsevier.

law, whereas novices sorted them based on their surface features such as inclined planes or pulleys. Similar results have been shown in mathematics, where novices categorized algebra problems on the basis of the problem content (e.g., river problems), whereas more experienced students categorized them based on the underlying equation or principle (Silver, 1979). This effect has been found in a number of domains, including computer programming (Adelson, 1981), medicine (Groen and Patel, 1988), and engineering design (Moss, *et al.*, 2006) among others.

These results have typically been explained by the hypothesis that experts' problem schemas are organized differently than novices. Schemas are hierarchical knowledge structures that include prototypical information about the type of problem, including declarative knowledge of objects, facts, strategies, and constraints, and may also include the procedural operators for solving the problem (Marshall, 1995). Expert schemas are hypothesized to include many principle or structural features of the problem type, whereas novice schemas include few structural features and shallower, surface features. Schemas play a critical role in categorizing a problem. See **Figure 1** for the interactive role schemas play in both problem perception and construction of the problem representation. Next, we discuss how experts and novices construct a mental representation of the whole problem.

Construction of a Representation

After a problem has been categorized, the problem solver can begin to elaborate their mental representation that goes beyond the given information of the task environment. For some simple or well-practiced tasks (e.g., puzzle-type tasks, NDM tasks, and procedural skills), this step happens as rapidly as categorization but for other complex, multistep problems, such as those in physics or ill-structured tasks such as design tasks, constructing a mental representation is an iterative process that takes time to develop. Constructing a representation involves specifying the important features of the task such as the relevant objects, operators, and constraints. When experts are solving complex or ill-structured problems, they approach them qualitatively, first examining and elaborating the givens of the problem and then refining that representation. For example, Voss and Post (1988) examined how experts and novices in political science solved an open-ended problem on how to increase crop productivity in the Soviet Union. They showed that experts spent more time than novices in developing their representation of the problem by elaborating the history and causal factors underlying the problem.

These results are consistent with the schema account of expertise in which experts' schematic knowledge provides access to additional knowledge and strategies to

help elaborate and develop the initial problem representation. This process has been hypothesized to be highly interactive (Chi *et al.*, 1981). Based on the initial categorization, the activated schema can provide additional information, strategies, constraints, and expectations to further characterize and elaborate the problem representation that may in turn activate other relevant schemas. For very complex problems, this process may take several iterations. Consistent with the categorization results described earlier, McDermott and Larkin (1978) (see also Reimann and Chi, 1989) have proposed that physicists construct problem representations at different levels of abstraction including: literal, naive, scientific (qualitative), and algebraic (see **Table 1** for a description of each level). Differences in levels of representation have also been shown in medicine where experts represent text descriptions of patient cases with an abstract situation model, whereas novices represent them more at the text-based (or surface) level (Groen and Patel, 1988).

Not only does expert knowledge facilitate the development and elaboration of the problem representation, but research also shows these representations are very durable. Experts have been shown to quickly encode problems and are able to easily access that representation even after disruption, whereas novices often take much longer to encode and re-represent a problem (Ericsson and Kintsch, 1995). The durability of expert memory and encoding has been shown in a variety of domains, including bridge (Charness, 1979), medicine (Norman *et al.*, 1989), and computer programming (McKeithen *et al.*, 1981). To account for these findings, Ericsson and Kintsch postulate that experts have developed effective long-term working memories that use very specific cues in the task environment to reliably retrieve prior knowledge structures (chunks and schemas).

Table 1 Four different levels of abstraction in representing physics problems

Representation level	Description
Literal	Representations containing keywords from the text.
Naive	Representations containing literal objects and their spatial relationships, often accompanied by a sketch of the situation.
Scientific	Representations containing idealized objects (points, bodies) and physical concepts (forces, momenta).
Algebraic	Equations containing physical concepts and their relationships.

From Reimann, P. and Chi, M. T. H. (1989). Human expertise. In Gilhooly, K. J. (ed.) *Human and Machine Problem Solving*, pp 161–191. New York: Plenum, with permission from Springer.

Application of Problem-Solving Procedures

After a problem representation has been constructed, the problem solver can then access and apply the appropriate problem-solving strategies and procedures to solve it. Experts have been shown to have more reliable access than novices to domain-specific solution procedures for well-practiced problem types. For simple problems, they make decisions faster and more accurately than novices. Research has also shown that experts and novices use different types of strategies when solving simple problems. Experts are more likely to use forward-working strategies for well-practiced problems, whereas novices use backward-working strategies. Forward-working strategies consist of working toward the solution from the domain principles. For example, physics experts first identify the principles for the task and then apply the domain-specific strategies and procedures, working step-by-step toward the solution (Simon and Simon, 1978). In contrast, novices have been shown to use general problem-solving heuristics, such as means-ends analysis to work backward from the problem goal (e.g., a sought value in physics or math). However, strategy use for both experts and novices critically depends on the relationship between prior knowledge and the task. Experts may also use general problem-solving methods and backward-working strategies when solving very novel tasks in the domain (e.g., physicists in their own research).

Solution Evaluation and Storage

Solution evaluation is the process of assessing a problem solution. Research has shown that experts spend more time than novices evaluating their solutions to make sure they satisfy task constraints (Groen and Patel, 1988; Voss and Post, 1988). Experts are also more likely than novices to identify and correct errors. For example, historians given a problem outside their subdomain are more likely than novices to seek additional resources and information to revise their initial framing of the problem, whereas novices are more likely to proceed with their initial incorrect assumptions (Wineburg, 1998). This research suggests that experts have developed better meta-cognitive skills (i.e., reflective monitoring) than novices for domain-relevant tasks. These skills may be particularly useful when adapting their knowledge to novel tasks in the domain.

After a solution has been generated it can be stored for later use. Much research shows that prior knowledge has a large impact on what is learned. For example, it is easier for experts to acquire new knowledge in the domain than for novices. Baseball experts have better recall than novices after listening to the broadcast of a novel baseball game (Spilich *et al.*, 1979) and expert pilots recall more than novices after listening to new air traffic control messages (Morrow *et al.*, 2001). Experts' rich, well-organized knowledge structures

enable them to easily incorporate (assimilate) new information into their prior knowledge.

Summary

Theoretical accounts of expert-novice differences are primarily articulated in the representation and organization of expert knowledge. Not only do experts have more conceptual and procedural knowledge than novices, but their knowledge is also organized in ways that facilitate effective problem solving. They are able to quickly recognize large chunks of domain-relevant information, see the deep features of the problem, and effectively elaborate their initial problem representations. They can apply domain-specific strategies, efficiently monitor their problem-solving progress by refining and correcting solutions, and can learn new domain-relevant information easier than novices. In the next section, we briefly review the theoretical accounts of how this knowledge is acquired.

Acquisition of Expertise

Much research shows that a minimum of 10 years of daily deliberate practice is necessary to develop expertise in most domains (Ericsson *et al.*, 1993). Ericsson and colleagues refer to deliberate practice as

repeated experience in which the individual can attend to the critical aspects of the situation and incrementally improve her or his performance in response to knowledge of results, feedback, or both from a teacher (Ericsson *et al.*, 1993: 368).

This perspective emphasizes how the type and structure of practice is critical to the acquisition of expert performance. In contrast to this perspective is the view that expertise is due to some talent or innate ability. The talent perspective, originally proposed by Galton (1869), is the notion that psychological traits, like physical traits, are inherited and family lineage (i.e., genes) strongly influences the person who achieves expert performance. Most modern formulations of this perspective hypothesize that expertise is the result of a complex interaction between genetic dispositions and experience (e.g., Simonton, 1999). Given that the talent perspective has received limited empirical support (see Howe *et al.*, 1998 for a discussion and commentary) and much research shows that expert advantages are due to their domain knowledge (and not general reasoning or memory abilities), we focus on the cognitive learning processes that give rise to this knowledge.

Expert knowledge is composed of both declarative and procedural components. Declarative knowledge consists of descriptions about the world, including facts, strategies, and principles, and is commonly referred to as knowing that. Procedural knowledge consists of information for how to

perform particular actions to accomplish task goals, and is commonly referred to as knowing how. Different learning processes have been hypothesized to account for the acquisition of these two types of knowledge. Learning declarative knowledge has been hypothesized to occur through observation, comprehension processes for oral and written discourse, induction, analogy, inference, and self-explanation (see Chi and Ohlsson, 2005 for a recent review of the learning mechanisms that lead to the acquisition of complex declarative knowledge). The key point is that declarative knowledge can be acquired through a number of reflective cognitive processes. Learning environments (e.g., classroom instruction) can be structured to facilitate its acquisition by including and improving these processes.

The acquisition of procedural knowledge or skill is hypothesized to occur through the repeated practice of a particular task or problem (Anderson, 1982, 1987). Fitts (1964) has characterized skill acquisition into three stages of performance, including the cognitive, associative, and automatic stages. During the cognitive stage, a person applies declarative knowledge to solve a problem and performance is characterized as being slow, effortful, and error prone. In domains such as mathematics and physics, novices rely heavily on declarative knowledge from prior examples to solve new problems (e.g., VanLehn, 1998). Students often apply this knowledge by making an analogy between the current problem they are solving and a previous problem that was solved similarly or had similar content. In the associative stage, the skill is practiced and performance becomes faster, more accurate, and less susceptible to interference. In this stage, students rely less on examples and more on applying learned rules to solve the problem. In the automatic stage, the skill has become proceduralized and is characterized by the fast application of the knowledge (or rules) with little or no errors and requires minimal cognitive resources.

Research on skill acquisition has revealed a power-law relationship between the amount of practice and performance. Generally, it shows that performance improves most when first learning a task, followed by decreasing learning gains as practice continues until performance asymptotes. However, the pattern of learning is more specific than the fast-then-slow pattern: when plotted on a logarithmic scale, the power-law relationship is revealed as an exact straight line. This exact relationship has been shown to be a very general phenomenon and has been observed in a variety of activities from learning to roll cigars to learning to solve math problems (see Proctor and Dutta, 1995 for a review). See Figure 3 for a real-world example of the power-law relationship.

One mechanism hypothesized to account for procedural learning is knowledge compilation (Anderson, 1987). Knowledge compilation acts as a translation device that interprets, or compiles, declarative knowledge into a set of specific procedural rules given a particular goal. As those

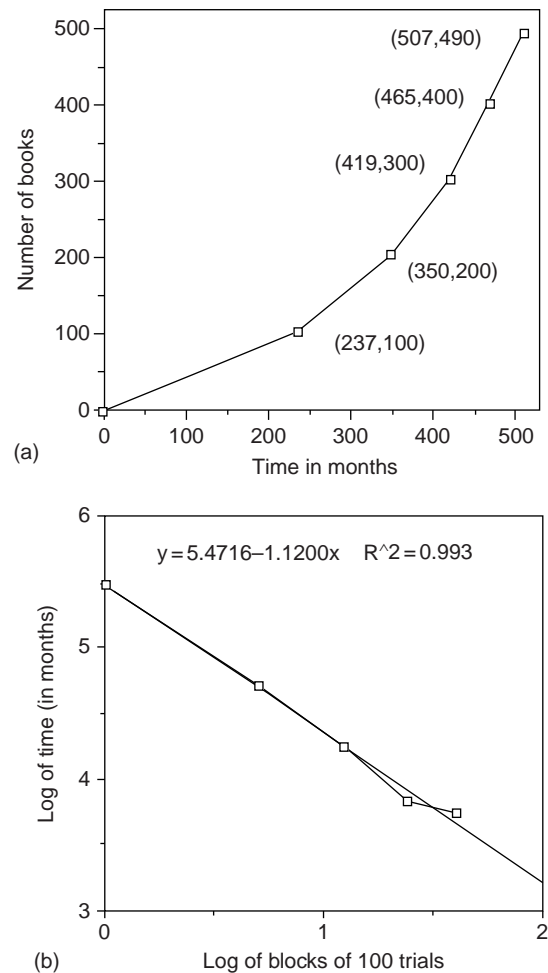


Figure 3 An illustration of the power-law relationship for the development of Professor Asimov's professional writing skills. (a) The number of books Professor Isaac Asimov wrote as a function of time in months. (b) The time to complete 100 books as a function of practice, plotted with logarithmic coordinates on both axes. From Ohlsson, S. (1992). The learning curve for writing books: Evidence of Professor Asimov. *Psychological Science* 3, 380–382. With permission from Wiley-Blackwell.

procedures (rules) get repeatedly applied they become concatenated or chunked together into more compact rules. This mechanism shows how cognitive processing changes from relying on the interpretation and retrieval of declarative knowledge to embedding that knowledge into a set of procedural rules that become more compact with use. The result is a context-specific representation of the skill that can be quickly and efficiently executed.

In sum, research has shown that the acquisition of expert performance requires extended deliberate practice in the domain. Expert knowledge is composed of both declarative and procedural knowledge and research on learning has shown that declarative knowledge can be acquired through multiple cognitive pathways, whereas procedural knowledge comes from the repeated practice of a task. This view suggests that the type and structure of

the learning environment are critical to the acquisition of expert performance. In the final section, we discuss two extensions to the traditional paradigm for research on expertise.

Current Directions

Current research extends the traditional paradigm in a number of ways. In this section, we focus on two: collaborative expertise and using expert–novice differences to determine targets of learning. In recent work, Schunn and colleagues (Tollinger *et al.*, 2006) had the unique opportunity to examine how over 50 NASA scientists worked together to plan the day-to-day operations of the two Mars rovers (Mars Explorer Rover Mission). The scientists' daily task was to analyze the data from the previous day and then come up with a plan for what experiments the rovers would conduct on the next day. They found that the amount of planning decreased across days and followed a learning curve similar to those typically observed for the acquisition of individual expertise, suggesting that expertise can also be acquired at the group level. Initial analyses suggest that the speedup in planning was due to both cognitive factors, such as individual knowledge chunking, plan reuse, and reducing task uncertainty, as well as social factors, such as coordinating information with others and the effect of leadership on the group.

In other recent work, Nokes *et al.* (2006) conducted a laboratory experiment on the effect of expertise on collaborative problem solving. They examined both expert and novice pilots' problem-solving performance when either working alone or with another participant of the same level of expertise. They found that experts working in pairs showed much larger collaborative benefits than novices working together, particularly for complex problem-solving tasks. Analysis of verbal protocols revealed that expert collaborative performance was supported by both domain knowledge (e.g., elaborating each other's contributions) and collaborative skill (e.g., acknowledging and restating the partner's contributions). The pilot and NASA scientist work extends the traditional paradigm and asks how expertise impacts cognitive and social processes at both the individual and the group level.

A second direction focuses on using expertise research to help identify targets of learning for novices. For example, Mestre and colleagues have used some of the classic findings in physics expertise (e.g., Chi *et al.*, 1981) to develop an instructional intervention to help students adopt similar strategies to that of the experts (Dufresne *et al.*, 1992; Mestre *et al.*, 1993). In one study, students were instructed to perform conceptual analyses vis-à-vis a computer interface that was based on the way experts strategize and solve problems, by first identifying the appropriate principles, justifying the use of those principles, and then

articulating the solution procedures. They found that this type of strategizing improved student's conceptual understanding and subsequent problem solving compared to control conditions where students used more traditional approaches to solve problems (e.g., textbook instruction). This research provides one example for how findings from the expertise literature can be used to help improve instructional techniques.

Conclusions

In this article, we reviewed the psychological research on expertise in human problem solving. We saw that expert knowledge impacts each stage of the problem-solving process from problem perception to solution storage. Expert knowledge is composed of both declarative and procedural knowledge and is organized into knowledge structures (e.g., chunks and schemas) that facilitate the categorization and construction of a mental representation of the problem, support the selection of appropriate strategies and procedures, provide constraints to evaluate problem-solving progress, and provide a framework to effectively store new information about the domain. These knowledge structures are acquired through deliberate practice, and learning environments can be designed to facilitate their acquisition. Future work should build upon this rich knowledge base to further advance theories of learning and instruction.

Acknowledgments

Preparation of this paper was supported by Grant SBE0354420 from the Pittsburgh Science of Learning Center and Grant R305B070085 from the Institute of Education Sciences.

See also: Concept Learning; Memory; Metacognition.

Bibliography

- Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. *Memory and Cognition* **9**, 422–433.
- Akin, O. (1980). *Models of Architectural Knowledge*. London: Pion.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review* **89**, 369–403.
- Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. *Psychological Review* **94**, 192–210.
- Charness, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology* **33**, 1–50.
- Chase, W. G. and Simon, H. A. (1973a). The minds eye in chess. In Chase, W. G. (ed.) *Visual Information Processing*, pp 215–281. New York: Academic Press.
- Chase, W. G. and Simon, H. A. (1973b). Perception in chess. *Cognitive Psychology* **4**, 55–81.

- Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In Ericsson, K. A., Charness, N., Feltovich, P. J., and Hoffman, R. R. (eds.) *The Cambridge Handbook of Expertise and Expert Performance*, pp 21–30. Cambridge, UK: Cambridge University Press.
- Chi, M. T. H., Feltovich, P. J., and Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science* **5**, 121–152.
- Chi, M. T. H. and Ohlsson, S. (2005). Complex declarative learning. In Holyoak, K. J. and Morrison, R. G. (eds.) *The Cambridge Handbook of Thinking and Reasoning*, pp 371–399. Cambridge: Cambridge University Press.
- Dufresne, R. J., Gerace, W. J., Hardiman, P. T., and Mestre, J. P. (1992). Constraining novices to perform expertlike analyses: Effects on schema acquisition. *Journal of the Learning Sciences* **2**, 307–331.
- Ericsson, K. A. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In Ericsson, K. A., Charness, N., Feltovich, P. J., and Hoffman, R. R. (eds.) *The Cambridge Handbook of Expertise and Expert Performance*, pp 223–241. Cambridge, UK: Cambridge University Press.
- Ericsson, K. A. and Kintsch, W. (1995). Long term working memory. *Psychological Review* **102**, 211–245.
- Ericsson, K. A., Krampe, R. T., and Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review* **100**, 363–406.
- Fitts, P. M. (1964). Perceptual-motor skill learning. In Melton, A. W. (ed.) *Categories of Human Learning*, pp 243–285. New York: Academic Press.
- Galton, F. (1869). *Hereditary Genius: An Inquiry into Its Laws and Consequences*. London: Julian Friedman Publishers.
- Groen, G. J. and Patel, V. L. (1988). The relationship between comprehension and reasoning in medical expertise. In Chi, M. T. H., Glaser, R., and Farr, M. J. (eds.) *The Nature of Expertise*, pp 287–310. Hillsdale, NJ: Erlbaum.
- Howe, M. J. A., Davidson, J. W., and Sloboda, J. A. (1998). Innate talents: Reality or myth? *Behavioral and Brain Sciences* **21**, 399–442.
- Klein, G. (1998). *Sources of Power: How People Make Decisions*. Cambridge, MA: MIT Press.
- Lesgold, A., Rubinson, J., Feltovich, P., et al. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In Chi, M. T. H., Glaser, R., and Farr, M. J. (eds.) *The Nature of Expertise*, pp 311–342. Hillsdale, NJ: Erlbaum.
- Marshall, S. (1995). *Schemas in Problem Solving*. New York: Cambridge University Press.
- McDermott, J. and Larkin, J. H. (1978). Re-representing textbook physics problems. In *Proceedings of the 2nd National Conference of the Canadian Society for Computational Studies of Intelligence*. Toronto: University of Toronto Press.
- McKeithen, K. B., Reitman, J. S., Reuter, H. H., and Hirtle, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology* **13**, 307–325.
- Mestre, J. P., Dufresne, R., Gerace, W. J., Hardiman, P. T., and Touger, J. S. (1993). Promoting skilled problem solving behavior among beginning physics students. *Journal of Research in Science Teaching* **30**, 303–317.
- Morrow, D. G., Menard, W. E., Stine-Morrow, E. A. L., Teller, T., and Bryant, D. (2001). The influence of expertise and task factors on age differences in pilot communication. *Psychology and Aging* **16**, 31–46.
- Moss, J., Kotovsky, K., and Cagan, J. (2006). The role of functionality in the mental representations of engineering students: Some differences in the early stages of expertise. *Cognitive Science* **30**, 65–93.
- Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In Giere, R. N. (ed.) *Cognitive Models of Science: Minnesota Studies in the Philosophy of Science*, vol. XV, pp 3–44. Minneapolis, MN: University of Minnesota Press.
- Newell, A. and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Nokes, T. J., Meade, M. L., Morrow, D. G., and Stine-Morrow, E. A. L. (2006). Investigating the effect of domain knowledge on collaborative problem solving. In Sun, R. and Miyake, N. (eds.) *Proceedings of the 28th Annual Cognitive Science Society*, pp 2572. Vancouver, CA: Cognitive Science Society.
- Norman, G. R., Brooks, L. R., and Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15**, 1166–1174.
- Proctor, R. W. and Dutta, A. (1995). *Skill Acquisition and Human Performance*. Thousand Oaks, CA: Sage.
- Reimann, P. and Chi, M. T. H. (1989). Human expertise. In Gilhooly, K. J. (ed.) *Human and Machine Problem Solving*, pp 161–191. New York: Plenum.
- Silver, E. A. (1979). Student perceptions of relatedness among verbal problems. *Journal of Research in Mathematics Education* **10**, 195–210.
- Simon, D. P. and Simon, H. A. (1978). Individual differences in solving physics problems. In Siegler, R. (ed.) *Thinking: What Develops?* Hillsdale, NJ: Erlbaum.
- Simonton, D. K. (1999). Talent and its development: An emergent and epigenetic model. *Psychological Review* **106**, 435–457.
- Spilich, G. J., Vesonder, G. T., Chiesi, H. L., and Voss, J. F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior* **14**, 506–522.
- Tollinger, I., Schunn, C. D., and Vera, A. H. (2006). What changes when a large team becomes more expert? Analyses of speedup in the Mars exploration rovers science planning process. *Paper Presented at the Human Computer Interaction Consortium*. Frasier, CO.
- VanLehn, K. (1998). Analogy events: How examples are used during problem solving. *Cognitive Science* **22**, 347–388.
- Voss, J. F. and Post, T. A. (1988). On the solving of ill-structured problems. In Chi, M. T. H., Glaser, R., and Farr, M. J. (eds.) *The Nature of Expertise*, pp 261–285. Hillsdale, NJ: Erlbaum.
- Wineburg, S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science* **22**, 319–346.

Further Reading

- Chi, M. T. H., Glaser, R., and Farr, M. J. (eds.) (1988). *The Nature of Expertise*. Hillsdale, NJ: Erlbaum.
- Ericsson, K. A. (ed.) (1996). *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games*. Mahwah, NJ: Erlbaum.
- Ericsson, K. A., Charness, N., Feltovich, P. J., and Hoffman, R. R. (eds.) (2006). *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge, UK: Cambridge University Press.
- Ericsson, K. A. and Smith, J. (1991). Prospects and limits of the empirical study of expertise: An introduction. In Ericsson, K. A. (ed.) *Toward a General Theory of Expertise*, pp 1–38. Cambridge, UK: Cambridge University Press.

Problem Solving and Reasoning

R E Mayer, University of California, Santa Barbara, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Problem solving and reasoning are frequently in the spotlight of educational reform. In addition to remembering information, students must be able to use what they have learned to solve new problems. A major goal of education is to help students become effective problem solvers, that is, people who can generate useful and original solutions when they are confronted with problems they have never seen before.

Definitions

A problem consists of a given state (i.e., a description of the current situation), a goal state (i.e., a description of the desired situation), and a set of operators (i.e., rules for moving from one state to another). A problem occurs when a situation is in one state, the problem solver wants it to be in another state, and there are obstacles preventing a smooth transition from one state to another. Duncker (1945: 1) defined a problem in the following way: “A problem arises when a living creature has a goal but does not know how this goal is to be reached.” Although Duncker’s definition is still valid, it must be updated to include the possibility of problem solving by machines, and so the term problem solver can refer to both living creatures and machines.

Problem solving occurs when a problem solver engages in cognitive activity aimed at overcoming a problem. Duncker (1945: 1) noted that “when one cannot go from the given situation to the desired situation simply by action, then there has to recourse to thinking” and “such thinking has the task of devising some action which may mediate between the existing and desired situations.” Similarly, Polya (1981: ix) defined problem solving as “finding a way out of a difficulty, a way around an obstacle.” In developing computer simulations of problem solving, Newell and Simon (1972) defined problem solving as a search for a path between the given and goal states of a problem. Mayer (1992) summarized three major elements in a definition of problem solving: (1) problem solving is cognitive because it occurs internally within the problem solver’s cognitive system, (2) problem solving is a process, because it involves manipulating or performing operations on the problem solver’s cognitive representations, and (3) problem solving is directed, because the problem solver is attempting to achieve some goal. In short, problem solving is directed, cognitive processing aimed at finding a way to achieve a goal.

What is the relation between problem solving and other high-level cognitive processes such as thinking and reasoning? Thinking can be broken down into two types – directed and nondirected thinking. Problem solving is a common and pervasive type of thinking, namely, directed thinking in which the thinker engages in cognitive processing aimed at achieving some goal. On the contrary, in nondirected thinking, the thinker engages in cognitive processing that is not aimed at achieving some goal, such as daydreaming or the abnormal thinking of autistic or schizophrenic people. In general, the terms problem solving and thinking can be used interchangeably, with the recognition that nondirected thinking is excluded.

Reasoning can be viewed as a type of problem solving, and is required in deductive reasoning tasks and inductive reasoning tasks. In deductive reasoning, the problem solver is given premises and must apply the rules of logic to derive a conclusion. For example, if you know all four-sided polygons are quadrilaterals and all squares are four-sided polygons, you may logically conclude, all squares are quadrilaterals. In inductive reasoning, the problem solver is given a series of instances or events or examples and must infer a rule. For example, after learning the Spanish words *la casa*, *el libro*, *la escuela*, *el perro*, *la muchacha*, and *el muchacho*, you may conclude that the article *la* goes with words ending in *a* and the *el* goes with words ending in *o*, a grammatical rule that is not without exceptions in Spanish.

Finally, creative thinking occurs when a problem solver generates ideas and critical thinking occurs when a problem solver evaluates them. Two important criteria in creative and critical thinking are that the ideas must be original and useful.

Types of Problems

An important distinction – based on the clarity of the problem statement – can be made between well-defined problems and ill-defined problems. A well-defined problem has a clear given state, a clear goal state, and a clear set of allowable operators. For example, finding the value of x in an algebraic equation such as $2x + 5 = 8$ is a well-defined problem because the given state is the equation, the goal state is a value for $x = \underline{\hspace{1cm}}$, and the operators are defined by the rules of algebra and arithmetic. In contrast, an ill-defined problem has a poorly specified given state, goal state, and/or operators. For example, choosing an

appropriate education for a career path is an ill-defined problem because the goal and allowable operators are not clearly specified. Most problems encountered in school are well-defined problems, whereas most crucial problems in everyday life are ill-defined.

Another important distinction – based on the knowledge of the problem solver – can be made between routine and nonroutine problems. Routine problems are identical or very similar to problems that the problem solver already knows how to solve, and therefore require what Wertheimer (1959) called reproductive thinking – reproducing responses that have been produced previously. For example, a routine problem for most high school students is “ $5 + 5 = \underline{\quad}$ ” or “The headquarters of the United Nations is located in the city of .” In the strictest sense, routine problems do not conform to the definition of problem, since they do not include an obstacle between the given and goal states. In contrast, nonroutine problems are different from any problems that the problem solver already knows how to solve and therefore require what Wertheimer (1959) called productive thinking – creating a novel solution. Examples for most high school students include writing a computer program to compute the mean and standard deviation of a sample, or working out why Spanish explorers waited several centuries before colonizing California. In school, students often work on routine problems called exercises; however, most important problems in everyday life are nonroutine.

A third distinction can be made between problems requiring convergent and divergent thinking. Convergent thinking problems have a single correct answer that can be determined by applying a procedure or retrieving a fact from memory. Examples include arithmetic computation problems and answering factual questions. Divergent thinking problems have many possible answers, and so the problem solver’s job is to create as many solutions as possible (Guilford, 1967). Classic examples include uses problems, such as “List all the possible uses of a brick,” and consequences problems, such as “List all the consequences of humans having six rather than five fingers.” Creativity can be measured in terms of the originality and usefulness of the answers, and divergent thinking skills, which underlie creativity, are taught in courses on creative thinking (Sternberg, 1999). Although divergent thinking is the hallmark of creativity, most school-based problems require convergent thinking.

Cognitive Processes and Types of Knowledge in Problem Solving

Problem solving can be divided into two major phases: problem representation and problem solution. Problem representation involves building a mental representation of the problem, and includes the cognitive process of

representing (i.e., building a situation model, that is, a mental representation of the situation described in the problem). Problem solution involves devising and carrying out a plan for solving the problem, and includes the cognitive processes of planning (i.e., devising a plan), executing (i.e., carrying out the plan), and monitoring (i.e., tracking the effectiveness of the plan).

The cognitive processes involved in problem representation and problem solution may interact, rather than occur in linear order. For example, a student may be given the following word problem: “Sarah has three marbles. David has two more marbles than Sarah. How many marbles does David have?” In representing the problem, the student must translate each sentence into an internal mental representation, such as “Sarah’s marbles = 3” and “David’s marbles = Sarah’s marbles + 2,” and mentally integrate them into a situation model, such as a spatial representation consisting of a bar for Sarah’s marbles (3 units high), a bar on top of it for the difference between Sarah’s and David’s marbles (2 units high), and a bar next to these for David’s marbles (indicating that Sarah’s marbles and the difference set of marbles are subsets of David’s marbles). Planning involves determining the operations to be performed, such as determining that 3 and 2 must be added together. Executing involves carrying out the operation(s), such as computing that 5 is the sum of 3 and 2. Monitoring involves detecting when a plan is not working, a step was not executed correctly, or an answer is questionable. Although school instruction tends to emphasize execution of basic skills, students’ major difficulties are in learning how to represent problems, devise plans, and monitor problem-solving processes.

Several types of knowledge are required for successful problem solving: facts, concepts, procedures, strategies, and beliefs (Anderson *et al.*, 2001; Mayer, 2008). Facts are elements of a factual knowledge about the world, such as, “There are 1000 milliliters in a liter.” Concepts are principles or models (which are elements of conceptual knowledge) such as, “In the number 567, 6 refers to the number of tens,” or categories or schemas (which are elements of schematic knowledge), such as knowing that “What is probability of flipping a fair coin three times and getting heads all three times?” is a joint probability problem. Procedures are step-by-step processes, such as knowledge of the procedure for long division used for 252 divided by 12. Strategies are general methods for problem solving, such as knowing how to break a problem into smaller parts. Beliefs are thoughts about one’s cognitive processing, such as believing, “I am good at solving statistics problems.” Facts and concepts are useful in the process of representing; procedures are useful for the process of executing; and strategies and beliefs are useful for planning and monitoring. Although instruction may tend to emphasize facts and procedures, all five kinds of knowledge are needed to support problem solving.

Rigidity in Thinking

A major obstacle to effective problem solving is rigidity in thinking. For example, in some problem-solving situations, the problem solver must use an object in a new way, such as using a brick as a doorstop or using a pencil as a bookmark. When a problem solver can only conceive of using an object in its most common function, then the problem cannot be solved. Duncker (1945) used the term *functional fixedness* to refer to a situation in which a problem solver cannot think of using an object in a new function that is required to solve the problem. Another example of rigidity occurs when a problem solver uses a well-learned procedure on a problem for which the procedure is inappropriate. For example, if a student solved a long series of arithmetic story problems that all contain the word 'more' and require adding the numbers together, the student may incorrectly carry out this same addition procedure for a new problem that actually requires subtracting the numbers from one another. Luchins (1942) used the term *einstellung* (or problem-solving set) to refer to this phenomenon. A goal of instruction in problem solving is to help students avoid rigid thinking.

Problem-Solving Transfer

Transfer is the effect of prior learning on new learning. When the new learning task is a problem to solve, we can use the term *problem-solving transfer* to refer to the effect of prior learning on solving a new problem (Mayer and Wittrock, 2006). Positive transfer occurs when previous learning helps you on a new task, whereas negative transfer occurs when previous learning hurts you on a new task. For example, if you have learned arithmetic, it should be easier for you to solve an arithmetic word problem – which would indicate positive transfer. If you learned to drive on the right side of the road in the United States, you may experience negative transfer in trying to learn to drive on the left side of the road in Australia. How does transfer work? This has been a central research question in psychology and education since Thorndike's (1931) (Thorndike and Woodworth, 1901) pioneering work in the early 1900s, and has generated three alternative explanations: general transfer, specific transfer, and mixed transfer. General transfer is the idea that learning task *A* can help you on task *B*, even if *A* and *B* have nothing specifically in common. For example, the doctrine of formal discipline (which is a classic theory of general transfer) posited that learning certain school subjects, such as Latin and geometry, would improve students' minds in general and thereby help them on unrelated tasks in the future. The doctrine of formal discipline was used to justify the establishment of Latin schools, in which the curriculum consisted of learning Latin, Greek,

geometry, and similar subjects. In one of the first experiments in the field of educational psychology, Thorndike (1931) (Thorndike and Woodworth, 1901) was able to show that students who learned Latin did not perform any better in learning bookkeeping than did students who had not learned Latin. Subsequent research (Singley and Anderson, 1989) also found little support for the idea of general transfer. In contrast, specific transfer is the idea that learning task *A* will help you in task *B* only to the degree that *A* and *B* have identical elements in common. Thus, learning Latin may help you learn Spanish because some of the verb conjugations are similar and many of the words are similar. The theory of specific transfer is problematic for educators because it suggests that students need to learn every specific piece of knowledge they will ever need.

Mixed transfer is a compromise between general and specific transfer that involves specific transfer of a general principle, that is, mixed transfer occurs when a learner abstracts a general principle from learning *A* and is able to apply it to solving a new problem *B*. The transfer is specific because both *A* and *B* can be solved by the same general principle but what is being transferred is a general principle rather than specific behaviors. For example, if students learn to make sense out of text passages by producing summaries, this general method can be transferred to the task of making sense out of a new text passage. Research on teaching of cognitive strategies shows that students can benefit from learning general strategies such as summarizing in reading comprehension (Pressley and Woloshyn, 1995).

Promoting positive transfer is a fundamental goal of education, that is, educators seek to help students learn in ways so that they will be able to use what they have learned to solve new problems. Mayer and Wittrock (2006) have described seven ways to promote problem-solving transfer: load-reducing methods, such as helping students build automaticity in basic skills; structure-based methods, such as using concrete models; schema-based methods, such as using advance organizers or pretraining; generative methods, such as encouraging learners to engage in elaboration or self-explanation; guided discovery methods, such as providing hints as someone solves a problem; modeling methods, such as providing worked examples; and teaching thinking skills, such as training people to use effective methods and strategies.

The Distinction between Productive and Reproductive Thinking

Why is it that some people invent clever solutions when confronted with a problem, whereas others do not? The Gestalt psychologist, Wertheimer (1959) attempted to answer this question by distinguishing between two kinds of thinking mentioned above, namely productive thinking and reproductive thinking. Productive thinking

involves producing a novel solution when confronted with a problem, whereas reproductive thinking occurs when problem solvers use solution procedures that they already know as a result of solving previous problems.

For example, Wertheimer (1959) described two ways of learning how to find the area of a parallelogram – learning by rote and learning by understanding. In learning by rote (or rote learning), the student is taught to measure the height, measure the base, and then multiply height times base. According to Wertheimer, students who learn by rote perform well on retention tests, such as finding the area of similar parallelograms, and poorly on transfer problems, such as finding the area of an unusually shaped parallelogram. In contrast, students who learn by understanding (or by meaningful learning) are encouraged to discover that the parallelogram can be converted into a rectangle by cutting the triangle off one end and moving it to the other end. Students who learn by understanding are expected to perform well on both retention and transfer tests. Thus, rote learning leads to reproductive thinking (as measured by retention tests) whereas meaningful learning leads to productive thinking (as measured by transfer tests).

The Nature of Insight

Insight is the cognitive process by which a problem solver suddenly moves from a state of not knowing how to solve a problem to a state of knowing how to solve a problem (Mayer, 1995). Insight plays a crucial role in creative thinking (Sternberg, 1999), in which a problem solver invents novel solutions to a problem. How does insight work? Gestalt psychologists and others have offered five somewhat interrelated explanations (Mayer, 1995): insight as completing a schema, insight as sudden visual reorganization, insight as reformulation of a problem, insight as removing mental blocks, and insight as finding a problem analog. Selz, working in the early 1900s in the Netherlands, produced psychology's first explanation of insight as completing a schema (Frijda and De Groot, 1982). For example, when given a problem such as "What is a co-ordinate of baseball?" a problem solver may say, "Let's see. Baseball is a sport. Another sport is football, so football is the answer." In this case, the problem solver is not following a chain of associations, but rather is trying to build a cognitive structure that has a superset (sport) linked to two subsets that are co-ordinates (baseball and one more), so coming up with an answer amounts to completing a schema. The idea that meaningful learning requires active construction by the learner is the fundamental theme of many current theories of learning (Bransford *et al.*, 1999).

Kohler, also working in the early 1900s, provided evidence that insight is a process of sudden visual reorganization in which the problem solver literally sees how all the parts of the problem fit together (Kohler, 1925).

For example, when an ape was put in an area that had stackable crates on the floor and bananas hanging overhead out of reach, the ape looked around and then in an apparent flash of insight, suddenly stacked the crates to form a sort of ladder leading to the bananas. This approach is consistent with current interest in using computer-assisted visualizations and concrete representations to help people understand how various systems work.

Duncker (1945) described insight as a reformulation of the problem, particularly a restatement of the givens or the goal in a new way. For example, in the tumor problem, you are asked to free a person of an inoperable stomach tumor by using "rays which destroy organic tissue at sufficient intensity" (p. 1). In order to solve the problem, the problem solver must restate the goal as, "lower the intensity of the rays as they pass through the healthy tissue," which leads to the solution of having many weak rays all converge on the tumor. This approach is consistent with the current idea that the most difficult aspect of problem solving is mentally representing the problem in a productive way.

Duncker (1945) also described insight as a process of removing mental blocks, that is, of being able to use an object in way that is different from its conventional use. For example, in the candle problem, the problem solver is given a box containing candles, a box containing tacks, and a box containing matches, and is asked to mount a lighted candle on a wall. The solution – involving using a box as the base, which is tacked into the wall – is much more difficult if the objects are in the boxes rather than next to them. According to Duncker, presenting the objects in the boxes creates functional fixedness – the tendency to be able to conceive of only one use of an object even though a problem solution requires using an object in a new way. Removing mental blocks is a key focus of current programs aimed at teaching thinking skills.

Finally, Wertheimer (1959) offered a fifth explanation of insight – finding a problem analog – in which the problem solver abstracts a general principle from one problem and applies it to a new one. Thinking by analogy is still an important theme in cognitive science, and is the basis for more current views of problem-solving transfer (Holyoak, 2005).

A sixth explanation – insight as nothing new – holds that solving insight problems is no different from solving other problems (Weisberg and Suls, 1973), although Metcalfe and Wiebe (1987) have shown that problem solvers use qualitatively different thinking processes for insight problems and routine problems.

Problem Space and Search Processes

Information-processing theories of problem solving focus on constructing a problem space and finding a path through the problem space (Newell and Simon, 1972;

Novick and Bassok, 2005). A problem space consists of a representation of the initial state, goal state, and all intervening states. For example, the problem space for solving the equation, $2X - 5 = X$, has this equation as the initial state, and $X = \underline{\hspace{1cm}}$ as the goal state. Two of the intervening states, directly after the initial state are $2X = X + 5$ and $2X - X - 5 = 0$, which were created by applying legal operators such as add 5 to both sides or subtract X from both sides. Similarly, other states are created by applying operators to these states, and so on.

Once a problem is represented as a problem space, the problem solver's task is to search for a path from the initial state to the goal state. Means-ends analysis is a search strategy in which the problem solver works on one goal at time; if that goal cannot be achieved directly, the problem solver sets a new goal of removing barriers, and so on. This search strategy is commonly used in computer simulations of problem solving and is consistent with the way that beginners solve problems.

Problem Solving in Realistic Situations

Although classic research focused mainly on solving artificial puzzles or formal syllogisms, cognitive science research on problem solving and reasoning has been shifting toward realistic situations including everyday problem solving, expert problem solving, and problem solving in subject areas (Ericsson *et al.*, 2006; Holyoak and Morrison, 2005).

Research on everyday thinking shows that people rarely use school-taught methods to solve problems encountered outside of school (Lave, 1988; Nunes *et al.*, 1993). For example, to determine the best buy in a supermarket – such as 90 cents for a 10-ounce can of peanuts or 45 cents for a 4-ounce can – the school-taught procedure is compute the unit cost of each item (i.e., 9 cents vs. 11.25 cents, respectively). However, Lave (1988) found that people almost never used the school-taught procedure; instead, they invented arithmetic procedures suited to the situation, such as the ratio strategy in which the problem solver notes that the larger one is a better buy because it costs twice as much and gives you more than twice as many ounces.

Research on expert problem solving compares differences in how novices and experts solve problems in domains such as physics, medical diagnosis, computer programming, and the game of chess. For example, when Larkin (1983) asked experts and novices to think aloud as they solved physics problems, she found that experts were more likely to focus on underlying physics concepts (such as forces and weights) whereas novices focused on surface characteristics (such as pulleys and ropes). Similarly, when Chi *et al.* (1981) asked experts and novices to sort physics problems into groups, experts sorted problems based on their underlying physics principle (such as conservation of energy), whereas novices sorted the problems based on

their surface characteristics (such as inclined planes and springs). Results of expert–novice studies suggest that experts represent and solve problems differently from novices, and so instruction can focus on helping novices think more like experts.

Another example of problem solving in realistic situations involves psychologies of subject matter, that is, research on problem solving in subject areas, such as reading, writing, mathematics, science, and history (Mayer, 2008). Instead of asking, how do people think in general, researchers ask, how do people think about testing a scientific theory, solving a mathematics word problem, or explaining why a historical event happened, or how do people think as they create an essay or make sense out of a printed passage. This approach suggests that instruction in subject matter areas should focus on helping students learn the cognitive processes and strategies required for successful problem solving.

See also: Cognition: Overview and Recent Trends; Knowledge Domains and Domain Learning.

Bibliography

- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., *et al.* (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Bransford, J. D., Brown, A. L., and Cocking, R. R. (1999). *How People Learn*. Washington, DC: National Academy Press.
- Chi, M. T. H., Feltovich, P. J., and Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science* 5, 121–152.
- Duncker, K. (1945). On problem solving. *Psychological Monographs* 58(270), 3.
- Ericsson, K. A., Feltovich, P. J., and Hoffman, R. R. (eds.) (2006). *The Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press.
- Frijda, N. H. and de Groot, A. D. (1982). *Otto Selz: His Contribution to Psychology*. Hague: Mouton.
- Guilford, J. P. (1967). *The Nature of Human Intelligence*. New York: McGraw-Hill.
- Holyoak, K. J. (2005). Analogy. In Holyoak, K. J. and Morrison, R. G. (eds.) *The Cambridge Handbook of Thinking and Reasoning*, pp 117–142. New York: Cambridge University Press.
- Holyoak, K. J. and Morrison, R. G. (eds.) (2005). *The Cambridge Handbook of Thinking and Reasoning*. New York: Cambridge University Press.
- Kohler, W. (1925). *The Mentality of Apes*. New York: Liveright.
- Larkin, J. H. (1983). The Role of Problem Representation in Physics. In Gentner, D. and Stevens, A. L. (eds.) *Mental Models*, pp 75–98. Mahwah, NJ: Erlbaum.
- Lave, J. (1988). *Cognition in Practice*. Cambridge, UK: Cambridge University Press.
- Luchins, A. (1942). Mechanization in problem solving. *Psychological Monographs* 54(248), 6.
- Mayer, R. E. (1992). *Thinking, Problem Solving, Cognition*, 2nd edn. New York: Freeman.
- Mayer, R. E. (1995). The search for insight: Grappling with Gestalt psychology's unanswered questions. In Sternberg, R. J. and Davidson, J. E. (eds.) *The Nature of Insight*, pp 3–32. Cambridge, MA: MIT Press.
- Mayer, R. E. (2008). *Learning and Instruction*. Upper Saddle River, NJ: Merrill Prentice Hall.

- Mayer, R. E. and Wittrock, M. C. (2006). Problem solving. In Alexander, P. A. and Winne, P. H. (eds.) *Handbook of Educational Psychology*, 2nd edn, pp 287–304. Mahwah, NJ: Erlbaum.
- Metcalfe, J. and Wiebe, D. (1987). Intuition in insight and non-insight problem solving. *Memory and Cognition* **15**, 238–246.
- Newell, A. and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Novick, L. R. and Bassok, M. (2005). Problem solving. In Holyoak, K. J. and Morrison, R. G. (eds.) *The Cambridge Handbook of Thinking and Reasoning*, pp 321–350. New York: Cambridge University Press.
- Nunes, T., Schliemann, A. D., and Carraher, D. W. (1993). *Street Mathematics and School Mathematics*. Cambridge, UK: Cambridge University Press.
- Polya, G. (1981). *Mathematical Discovery*. New York: Wiley.
- Pressley, M. and Woloshyn, V. (1995). *Cognitive Process Instruction*, 2nd edn. Cambridge, MA: Brookline Books.
- Singley, M. K. and Anderson, J. R. (1989). *The Transfer of Cognitive Skill*. Cambridge, MA: Harvard University Press.
- Sternberg, R. J. (1999). *Handbook of Creativity*. New York: Cambridge University Press.
- Thorndike, E. L. (1931). *Human Learning*. New York: Century.
- Thorndike, E. L. and Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review* **8**, 247–261.
- Weisberg, R. W. and Suls, J. (1973). An information processing model of Duncker's candle problem. *Cognitive Psychology* **4**, 255–276.
- Wertheimer, M. (1959). *Productive Thinking*. New York: Harper and Collins.

Reasoning, Moral and Social

N S Jampol, C B Richardson and M Killen, University of Maryland, College Park, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Behaviorism – The psychological theory stating that all internal cognitive function can be captured by observing individual behavior. According to this theory, no additional information about an individual's psyche can be gleaned from attempting to understand internal processes.

Heteronomy – Piagetian stage of moral development during which an individual is said to be adherent to the demands of an authority source without question or detraction.

Moral justifications – Reasoning that appeals to justice, others' welfare, and rights to the end of supporting a moral judgment.

Psychological justifications – Reasoning that appeals to autonomy, individuality, and personal choice to the end of supporting a psychological judgment.

Social cognitive domain theory – The proposition that social cognition and the perception of action can be parsed into three mutually exclusive but interactive domains of social knowledge: moral, conventional, and psychological.

Social-conventional justifications – Reasoning that appeals to values existing to promote the orderly function of groups such as customs, conventions, and etiquette.

Stage model – A theory that posits that an individual's development occurs in a linear and nonrecursive direction.

Transgression – An infraction. A potentially subjective belief that an action runs contrary to agreed upon way of operating in the world.

Introduction

Moral and social reasoning, which refer to how individuals evaluate social issues and events that occur throughout the life time, emerge in early childhood and evolve through adolescence. Moral reasons refer to those pertaining to justice, others' welfare, and rights. Social reasoning includes social-conventional reasoning about customs, conventions, and etiquette, as well as psychological reasoning about autonomy, individuality, and personal choice. The historical

background about the study of moral and social reasoning will be described followed by current findings on how these types of reasons develop, are acquired, and emerge in different cultures.

Historical Background

Jean Piaget (1932), a Swiss child developmental psychologist, first theorized about the origins of moral reasoning in the early part of the latter century. Piaget drew on deontological moral philosophy to define morality, which he posited referred to judgments that were universal, autonomous of authority, and distinct from cultural norms. Piaget did not examine children's performance on exams or tests, but rather developed an interview method to examine the forms of reasons that children used in their everyday interactions, such as the underlying principles associated with the rules and conventions of the game of marbles. This developmentally appropriate methodology represented a challenge to the prevailing measurement of behavior as an index of morality in the 1950s, with Piaget's focusing instead on intentions, motivations, and principles. Adapting the stage-model reasoning used to explain cognitive development (1932), Piaget postulated that there were three distinct stages of moral development, premoral, heteronomy (authority orientation), and autonomy (principles of justice). One of the unique aspects of Piaget's theory was the focus on the role of peer interaction in the development of moral judgment. Piaget demonstrated how children's peer interactions enable them to develop concepts of equality and fairness.

Lawrence Kohlberg (1969, 1971), a professor at Harvard University, extended the work of Piaget (1932) and critiqued the behaviorist conception of morality as learning rather than knowing (Kohlberg, 1971). Kohlberg characterized morality as a set of stages with universal moral principles as the highest level (1969). It was postulated that an individual's morality developed in a series of six stages moving from the personal, to the conventional, then finally to the moral. At this final stage of moral development, it was theorized that an individual would only act in accordance with universal morality, never heeding to competing commands from would-be sources of authority. Kohlberg hypothesized that moral reasoning developed much later than Piaget had first documented. Kohlberg developed a dilemma used throughout the

social sciences to measure reasoning, which involved a man's decision to steal a drug to save his wife's life (the Heinz dilemma task).

More recently, Elliot Turiel (1983, 1998), who worked with Kohlberg, expanded the study of moral reasoning by outlining three domains of knowledge: the moral, social-conventional, and psychological. Turiel provided evidence to show that children as young as 4–5 years of age were not premoral in the way that Piaget and Kohlberg suggested, and instead understood basic ideas about fairness, equality, and justice. At the same time that children recognize that certain transgressions such as hitting and not sharing with others are wrong due to the intrinsic negative consequences to others, they also begin to develop an understanding about conventions that serve to make groups function. Turiel's approach to social and moral reasoning was applied to children and adolescents in many cultures around the globe by many of his students and colleagues (see Smetana, 2006).

In the next sections, moral and social reasoning in childhood and adolescence are described, followed by how these forms of reasoning emerge in different cultures, as we note implications for education, and educational programs.

Preschool and Early Childhood

One of the first studies to assess moral transgressions in comparison to social conventional violations was undertaken by Nucci and Turiel (1978). Nucci and Turiel (1978) identified these differentiated domains within a preschool-aged sample. The researchers found that children and adults were quick to respond to moral transgressions, but only the adults reacted to the social conventional violations. Specifically, Nucci and Turiel (1978) discovered that moral transgressions were often addressed by peers when violated (as compared to social-conventional violations) and done so in the language that a child knows best: the language of another child. This study confirmed the basic distinction between domains of social knowledge, specifically assessing that conventional transgressions are bound up with authority dictates (i.e., a teacher's comments about the transgressions), while moral transgressions are generalizable and lack authority contingency (i.e., both teachers and peers responded to physical transgressions).

Further extending the empirical and theoretical framework of domain theory, Judith Smetana conducted a study using a preschool-aged sample to assess their understanding of the nature of morality. Smetana (1981) assessed seriousness, rule contingency, rule relativism, and amount of deserved punishment for ten different age-appropriate moral (hitting and not sharing toys) and social-conventional (child sitting in the wrong seat and putting a toy in the wrong place) violations. Across age and sex, all children considered moral violations to be more serious and more deserving

of punishment than social-conventional transgressions. Furthermore, children considered moral transgressions to be wrong regardless of the presence of a rule forbidding such an act, judging the inherent wrongness of a moral transgression to be what made the action wrong. Neither the presence of a rule, context, nor consensus determined the morality of a transgression. In comparison, social conventions were more likely to be contingent upon rules and relative contexts (place, time, etc.). This seminal work by Judith Smetana provided both confirmation and extension of social cognitive domain theory and an elaboration of Nucci and Turiel's earlier findings (Smetana, 2006).

A study by Killen and Smetana (1999) assessed children's evaluations of moral, social-conventional, personal, and mixed-domain events. Observing children's and teachers' reactions to a variety of rule violations, the researchers coded different behaviors and subsequent responses. When observing student/teacher interaction, the researchers found that teachers used the most didactic language when addressing moral and social-conventional violations. Conversely, teachers were not nearly as direct when addressing personal-domain and mixed-domain violations. Perhaps most significantly, children often vocalized personal-domain assertions ("I want juice, not milk"), thereby demanding autonomy when they believed their personal free choice was violated. However, not just children are aware of this domain of thought. When interviewed, teachers identified a personal domain for their students as well as understood the children's need for autonomy (the child can choose the type of toy with which they want to play).

Middle Childhood

As we might expect, children's social and moral reasoning do not simply plateau in early childhood. Rather, the developmental story of social and moral reasoning continues through the years. Conducting interviews with over 100 children, Tisak (1986) assessed children's conceptions of parental authority. Specifically, Tisak examined the boundaries of parental authority and found that children view authority as heteronymous and context dependant. Children were asked to consider the legitimacy of a parent's moral rule (do not steal toys), social-conventional rule (do not leave dishes on the table), and personal choice (you cannot play with a certain child). In addition, children were asked to evaluate the obligation to follow an established rule once it has been put in place. Rather than viewing rules and authority as homogeneous, children considered the domain-specific demands of each rule, differentiating moral rules (stealing) to be most legitimate, followed by social-conventional rules (family chores), and finally personal (friend choice) to be the least legitimate within the range of reasonable parental jurisdiction. It seems that children's

expectations of autonomy and evaluations of authority change dramatically as the child develops.

Considering the issue of authority, rules, and regulations, Laupa *et al.* (1995) examine children's concepts of authority (Laupa *et al.*, 1995) as a means of assessing the different components of authority evaluation, reasoning, and the subsequent act of obedience or dissent. Children understand what underlies authority jurisdiction, and, contrary to popular conceptions, they do not view authority as right because of punishment or strict obedience. Children judge the nature of the authority comments, whether the request is legitimate, and the basis for why the authority figure is making a command (Tisak, 1986).

Laupa's study addresses the source of the command and the attributes of the authority. The authority is assessed by age (or adult/child status), knowledge specific to the command, and social position (e.g., the authority's place in the hierarchy within an organization or system: a school teacher compared to a principal, etc.). Laupa *et al.* (1995) also study the relationship between dominant and subordinate individuals. This relationship between authority figures and subordinates is contingent on the specific context and content of each situation, as those sources of authority deemed legitimate in one situation may not be considered viable in another.

Accordingly, individuals critically consider the source of an individual's legitimacy when assessing the imperative to obey an order (Laupa, 1991). If the social context is the source of a body's authority, then the risk of punishment may influence rates of obedience (Laupa, 1991). Compliance to authority is not punishment-avoidant (Kohlberg, 1969) or heteronomous, as Piaget hypothesized (1932/1965), but based on the specific exigency of the specific command and its components.

Children do not just consider the immediate or tangible when reasoning about social and moral rights, but rather actively assess larger issues of freedom and civil rights. Charles Helwig (1995) found that individual rights to freedom of religion and speech were considered moral: natural, unalterable, and generalizable with children across a variety of ages, all rejecting the impediment of these rights by the government. In addition, Helwig found that despite the moral imperative to maintain freedom of religion and speech, children were aware of the complexities and intricacies of these rights. Specifically, children understood that these rights presented choices that younger individuals may not yet be fit to make and are accordingly denied such rights till they have reached a more experienced age. This understanding of the complexity of civil rights emerges with age. Children as young as 6 were aware of these civil rights and reason that such rights should be upheld without waiver. However, by age 8 children begin to evaluate these rights in light of greater contextual needs and issues as they attempt to coordinate the mutual demands of multidomain issues (e.g., "Is a racial slur protected as free speech?").

Adolescents

In another study by Nucci (2001), it was found that as children develop, a greater number of issues (such as those pertaining to the body and privacy) become considered matters of personal choice, existing outside of both moral and social-conventional domains (1991). Some issues of personal autonomy and privacy, which had never been contentious in the past, become all the more important as a child develops. Expectations of increased privacy and autonomy both mark the development of an individual from late childhood into adolescence. Specifically, the issue of drug use was studied by Nucci *et al.* (1991) insofar as how 9th and 12th graders reason about the use of controlled substances. This study revealed that across age groups and frequency of drug use, adolescents view the choice to use drugs as one of personal choice rather than a moral or social-conventional matter. However, those participants who were considered high users were more likely than their low-user counterparts to consider drug use an entirely personal choice. Conversely, participants who were considered low users also used prudential reasoning (concerns for harm to the body and physical dangers associated with drugs) in addition to asserting autonomy and identifying the choice to use drugs as a personal one (Nucci *et al.*, 1991).

Expected autonomy over conventional, prudential, and personal issues (as well as issues which require the coordination of these distinct domains) predictably increases through adolescence. However, Smetana and colleagues found that adolescents and parents reason differently about these three types of issues.

When interviewed, mothers thought prudential and social-conventional decisions were within the realm of reasonable and expected parental authority. Conversely, adolescents viewed prudential and social-conventional issues as a matter of joint authority, and not the sole decision of a parent. Multifaceted personal issues requiring the coordination of several domains were considered by adolescents to be best explored with the aid of a parent. Although adolescents are not content to merely act on a parent's opinion, it was shown that they do still seek out parental input as a valuable source of information when making a complex or significant choice. This choice to seek out additional information from parents seems to be a highly beneficial one for adolescents, as outcomes were found to be dramatically better for those participants who sought this additional parental perspective. Yet, a balance is required. As teenagers enter late adolescence, a lack of autonomy is also often marked by increased depression and a reduction in the adolescent's sense of self-worth. The work of Smetana would seem to argue for a careful balancing act in which parents and parenting matters a great deal to adolescents, yet, these emerging adults also require some degree of autonomy with respect to the increased number of responsibilities expected of an older teenager (Smetana *et al.*, 2004).

Culture

Up to this point, the studies described for social and moral reasoning have been empirically investigated in the United States. However, an entire body of cross-cultural research on consistencies and variations in moral reasoning has also been conducted.

Wainryb (1995) explored the moral reasoning of both Israeli Jewish and Israeli Druze populations. Concerned with both personal and justice-based decisions within authority-sanctioned or interpersonal contexts (whether or not it was alright to keep someone else's lost money if sanctioned by a parental authority), Wainryb assessed reasoning in both groups of participants. It was found that culture was not a good predictor for how individuals framed and reasoned about different moral and personal issues. Rather, the content and context-specific attributes of the vignette was the better predictor of participant outcome across the two distinct and very different cultures.

Considering the traditional social psychology dichotomization of Eastern and Western cultures, the findings of Kristin Neff and Charles Helwig are of particular import. Neff and Helwig (2002) examined conceptions of rights and individualism in China, India, the United States, and Canada. Importantly, prior research had almost exclusively associated individualistic cultures (the United States, Canada) with concerns for rights and justice and collectivist cultures (China, India) with tradition and authority. However, all four cultural groups expressed concerns for justice and rights as well as for tradition and authority simultaneously. Furthermore, commonly held conceptions and traditions were not blindly accepted by participants, but rather participants questioned the cultural traditional practices with a critical eye.

Ardila-Ray and Killen (2001) studied evaluations of loci of control, compliance, teacher legitimacy, and teacher methods of conflict resolution in moral, social-conventional, and personal domains in Colombia, which is considered a particularly collectivistic culture for the Western hemisphere. Developmentally, the expectation of personal autonomy for choice of activity, playmate, and adherence to social-convention increased with age. Older children expected more freedom for making such choices. In addition, students preferred more elaborated explanations from their teachers with age, as students were increasingly less satisfied with explanation-void punishment. These findings, in what has been traditionally considered a collectivistic culture, indicate that children critically evaluate their rights and the role of authority.

Similar to Colombian children's homogenous and critical evaluations of rights and authority, school-aged Chinese children critically evaluated social situations. Yau and Smetana (1996) examined the developmental demands of children for autonomy in Hong Kong. Consistent with a global model of emerging demands in the personal

domain, children, as they move from early adolescence (9th grade) into late adolescence (12th grade), require more autonomy. Accordingly, Chinese children's reports of warmth and closeness with family members were negatively correlated with rating of conflicts and controlling behaviors. This finding supports a developmental model of social reasoning with respect to autonomy that is not an exclusively Western phenomenon but rather an emerging need and concern throughout the world.

Nucci *et al.* (1996) reported that middle and lower socioeconomic status (SES) school-aged children in Brazil found that concerns for physical well-being were subject to parental authority and judgment. In addition, the personal domain was examined and an effect was found to have an interaction with SES where children in middle-class families expected more autonomy and attributed a greater number of choices to the personal domain than their lower SES peers. However, and, perhaps most interestingly, this SES-based difference for distinguishing the contents of the personal domain disappeared with age. Lower and middle SES children both expected comparable levels of personal freedom and autonomy by the time they reached age 16.

This cross-cultural research indicates a break from the traditional dichotomy of Eastern and Western cultures. Rather, researchers find that both Eastern (traditionally conceptualized as collectivistic) and Western cultures (traditionally conceptualized as individualistic) value traditions and values. Furthermore, as the literature of subversion will support, there is heterogeneity within these larger groups.

Complex issues that are not prototypically moral, social-conventional, or personal often demand the simultaneous coordination of multiple domains at once. Killen and colleagues have investigated exclusion in intergroup relation settings, where moral, social-conventional, and psychological issues bear on decisions about who to include and who to exclude. In a study with preschool-aged children, participants gave priority to fairness in the context of gender stereotypic reasons for exclusion (e.g., girls exclude a boy from doll playing) when the exclusion was straightforward, but relied on stereotypes when the exclusion involved multiple considerations (Killen *et al.*, 2001). Further, in a study with first, fourth, and seventh graders, children were asked about exclusion in straightforward and equal and unequal qualification situations (Killen and Stangor, 2001). Results showed that in the equal qualifications context, participants were more likely to choose the nonstereotypical child to join the group (e.g., choose the boy for ballet and the girl for baseball cards), but were less likely to include the nonstereotypical child if that child was not as qualified as another child (Killen and Stangor, 2001). Developmentally, seventh graders were most sensitive to context, justifying exclusion in the unequal condition more strongly than either the first or the fourth graders. Older children have more experience with groups and know that effective

functioning is important. Thus, their reasons for picking a more qualified child often pertain to considerations of group dynamics and group functioning.

Stereotypes and Social Reasoning

There is a paucity of research specifically addressing the interplay between stereotypes and social reasoning; however, it is important to address the role of stereotypes in social contexts involving exclusion, because stereotypes inform an individual's judgments on who to include, as well as their justifications for doing so. A stereotype that negatively informs a judgment could result in someone denying resources to the stereotyped individual. Conversely, a stereotype that positively informs judgments may result in inclusion of the stereotype-inconsistent person because of the knowledge that that person may not get many chances to do something stereotype inconsistent (i.e., Killen and Stangor, 2001).

In a related study, Horn (2003) analyzed high school participants' judgments regarding exclusion based on group membership (e.g., cheerleaders, jocks, preppies, etc.) when given ambiguous and nonambiguous information about the situation and the target's qualifications. Similar to the findings by Killen and Stangor (2001), Horn and colleagues found that participants judged exclusion in the ambiguous condition as more wrong than exclusion in the nonambiguous condition (Horn, 2003). Extending previous findings, they also found that participants used more group-related stereotypes when judging ambiguity. Specifically, ninth graders used more stereotypes than did 11th graders in the ambiguous condition (Horn, 2003). This research, along with the Killen and Stangor (2001) study, has been important in showing that in exclusion situations, different factors are taken into account, and also that group functioning is important in the eyes of adolescents. However, it is important to understand that stereotype-laden social reasoning is a product of an individual's belief and understanding of facts, rather than being a matter of simple bigotry.

Educational Implications

The place of morality in the classroom – and whether or not it is in fact a school's responsibility to offer moral education – has been debated for decades. Larry Nucci has examined the trend and controversy of moral education as it had emerged and shifted over the course of the 20th century. Nucci (2001) suggests that moral education is not so much about judgment as it is about reasoning. We now know that children begin to learn about the features that underlie and distinguish different domains of social knowledge from a much younger-than-expected age. Findings suggest that children are engaged in this task

by age 5, the same time that the child begins formal schooling. The according import of teaching students to critically assess and evaluate, rather than didactically instruct on what is or is not right is undeniable (Wainryb *et al.*, 2004). The most significant work before teachers charged with educating students is therefore the skill and art of critical thinking.

Conclusions

In summary, social and moral reasoning begin in childhood and reflect a central part of what it means to live in a culture and be a member of a society and community. The process of basing everyday decisions on such concepts as fairness, justice, equality, and respect for autonomy is universal. It is not unique to one culture or to one category of persons. Giving priority to moral and nonmoral social considerations varies by the context, and thus, in many situations, children are learning how to do this for the first time. Parents and educators can facilitate this process by helping children to determine the various salient issues in any specific decision-making context, particularly those involving social and moral reasoning.

See also: Piaget: Recent Work.

Bibliography

- Ardila-Rey, A. and Killen, M. (2001). Middle-class Colombian children's evaluations of moral, social-conventional and personal events in the classroom. *International Journal of Behavioral Development* **25**, 246–255.
- Helwig, C. C. (1995). Adolescents' and young adults' conceptions of civil liberties: Freedom of speech and religion. *Child Development* **66**, 152–166.
- Horn, S. (2003). Adolescents' reasoning about exclusion from social groups. *Developmental Psychology* **39**, 11–84.
- Killen, M., Pisacane, K., Lee-Kim, J., and Ardila-Rey, A. (2001). Fairness or stereotypes? Young children's priorities when evaluating group exclusion and inclusion. *Developmental Psychology* **37**, 587–596.
- Killen, M. and Smetana, J. G. (1999). Social interactions in preschool classrooms and the development of young children's conceptions of the personal. *Child Development* **70**, 486–501.
- Killen, M. and Stangor, C. (2001). Children's social reasoning about inclusion and exclusion in gender and race peer group contexts. *Child Development* **72**, 174–186.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In Goslin, D. A. (ed.) *Handbook of Socialization Theory and Research*, pp 347–480. Chicago: Rand McNally.
- Kohlberg, L. (1971). From is to ought (how to commit the naturalistic fallacy and get away with it in the study of moral development). In Mischel, T. (ed.) *Cognitive Development and Epistemology*, pp 151–235. New York: Academic Press.
- Laupa, M. (1991). Children's reasoning about three authority attributes: Adult status, knowledge, and social position. *Developmental Psychology* **27**, 321–329.
- Laupa, M., Turiel, E., and Cowan, P. (1995). Obedience to authority in children and adults. In Killen, M. and Hart, D. (eds.) *Morality in*

- Everyday Life: Developmental Perspectives*, pp 131–165. Cambridge, UK: Cambridge University Press.
- Neff, K. and Helwig, C. C. (2002). A constructivist approach to understanding the development of reasoning about rights and authority within cultural contexts. *Cognitive Development* **17**, 1429–1450.
- Nucci, L., Camino, C., and Sapiro, C. (1996). Social class effects on northeastern Brazilian children's conceptions of areas of personal choice and social regulation. *Child Development* **67**, 1223–1242.
- Nucci, L. P. (2001). *Education in the Moral Domain*. Cambridge, UK: Cambridge University Press.
- Nucci, L. P., Guerra, N., and Lee, J. (1991). Adolescent judgments of the personal, prudential, and normative aspects of drug usage. *Developmental Psychology* **27**, 841–848.
- Nucci, L. P. and Turiel, E. (1978). Social interactions in the context of moral and conventional transgressions. *Child Development* **49**, 400–407.
- Piaget, J. (1932). *The Moral Judgment of the Child*. Oxford, UK: Harcourt Brace.
- Smetana, J. G. (1981). Reasoning in the personal and moral domains: Adolescent and young adult women's decision-making regarding abortion. *Journal of Applied Developmental Psychology* **2**, 211–226.
- Smetana, J. G. (2006). Social domain theory: Consistencies and variations in children's moral and social judgments. In Killen, M. and Smetana, J. G. (eds.) *Handbook of Moral Development*, pp 119–154. Mahwah, NJ: Erlbaum.
- Smetana, J., Campione-Barr, N., and Daddis, C. (2004). Longitudinal development of family decision making: Defining healthy behavioral autonomy for middle-class African American adolescents. *Child Development* **75**, 1418–1434.
- Tisak, M. S. (1986). Children's conceptions of parental authority. *Child Development* **57**, 166–176.
- Turiel, E. (1983). *The Development of Social Knowledge: Morality and Convention*. Cambridge, UK: Cambridge University Press.
- Turiel, E. (1998). The development of morality. In Damon, W (series ed.) and Eisenberg, N. (vol. ed.) *Handbook of Child Psychology: Volume 3. Social, Emotional, and Personality Development*, pp 863–932. New York: Wiley.
- Wainryb, C. (1995). Reasoning about social conflicts in different cultures: Druze and Jewish children in Israel. *Child Development* **66**, 390–401.
- Wainryb, C., Shaw, L. A., Langley, M., Cottam, K., and Lewis, R. (2004). Children's thinking about diversity of belief in the early school years: Judgments of relativism, tolerance, and disagreeing persons. *Child Development* **75**, 687–703.
- Yau, J. and Smetana, J. G. (2003). Adolescent–parent conflict among Chinese adolescents in Hong Kong. *Child Development* **67**, 1262–1275.

Further Reading

- Arsenio, W. (2006). Happy victimization: Emotion dysregulation in the context of children's instrumental/proactive aggression. In Snyder, D., Simpson, J., and Hughes, J. (eds.) *Emotion Regulation in Families: Pathways to Dysfunction and Health*, pp 101–112. Washington, DC: American Psychological Association.
- Damon, W. (1977). *The social world of the child*. San Francisco, CA: Jossey-Bass.
- Kahn, P. (2001). *Human Relationship with Nature*. Cambridge, MA: MIT Press.
- Killen, M., Lee-Kim, J., McGlothlin, H., and Stangor, C. (2002). *How Children and Adolescents Evaluate Gender and Racial Exclusion. Monographs for the Society for Research in Child Development*. Serial no. 271, vol. 67, no. 4. Oxford, UK: Blackwell.
- Killen, M., Rutland, A., and Jampol, N. (2008). Social exclusion in childhood and adolescence. In Rubin, K. H., Bukowski, W., and Laursen, B. (eds.) *Handbook of Peer Relationships, Interactions, and Groups*, pp 249–266. New York: Guilford Press.
- Killen, M. and Smetana, J. G. (eds.) (2006). *Handbook of Moral Development*. Mahwah, NJ: Erlbaum.
- Nucci, L. P. and Killen, M. (1991). Social interactions in the preschool and development of moral and social concepts. In Scales, B., Almy, M., Nicolopoulou, A., and Ervin-Tripp, S. (eds.) *Play and Social Context of development in Early Education*, pp 219–233. New York: Teachers College Press.
- Smetana, J. G. (1995). Morality in context: Abstractions, applications, and ambiguities. In Vasta, R. (ed.) *Annals of Child Development*, vol. 10, pp 83–130. London: Jessica Kingsley.
- Turiel, E. (2002). *Culture and Morality*. Cambridge, UK: Cambridge University Press.
- Wainryb, C., Smetana, J., and Turiel, E. (2008). *Social Development, Social Inequality, and Social Justice*. Mahwah, NJ: LEA.

Remembering as Social Activity

D Middleton, Loughborough University, Loughborough, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Conversation analysis (CA) – The detailed analysis of how talk makes things happen in social life, business life, healthcare, education, leisure, politics, etc.

Discursive psychology – The study of the relations between psychological topics and everyday talk through the examination of the ways in which we use common-sense ideas about thoughts, emotions, attitudes, etc., as resources for accounting for everyday life.

Metacognitive formulation of remembering and forgetting – It refers to how we use what it is to remember and forget in warranting our claims to experience and accounts of events and our position on and within them.

Schema – A conceptualization of a principle for the organization of the content of experience.

Introduction

Memory is of critical importance in any consideration of learning. To become knowledgeable through any process of learning appears deeply bound up with what it is to remember and forget. Without memory we have no purchase on what we knew, know, and might come to know. So then, we might expect that all that is required is an understanding of how we consign to, contain within, and conscript the contents of memory.

Charles Dickens graphically illustrates rote learning and the centrality of memory as the container of knowledge in his cryptic parody of English Victorian education in *Hard Times*. Indeed, Gradgrind's teaching is emblematic for all that was lacking in schooling where verbatim rehearsal of multiplication tables, scientific formulae, grammar, poetry, and historical dates was the order of the day:

'Bitzer,' said Thomas Gradgrind, 'Your definition of a horse.' 'Quadruped. Graminivorous. Forty teeth, namely twenty-four grinders, four eye-teeth and twelve incisive. Sheds coat in the spring; in marshy countries sheds hoofs too. Hoofs hard, but requiring to be shod with iron. Age known by marks in mouth.' Thus (and much more) Bitzer.

'Now girl number twenty,' said Gradgrind. 'You know what a horse is' (p. 6).

For Bitzer to thrive in Gradgrind's world, he must enter a particular world defined by what Gradgrind marks as institutionally relevant. Bitzer's mnemonic performance is isomorphic with what it is to know. Equally, girl number 20 is presented with a model performance in Gradgrind's terms of what sort of practice frames and circumscribes what she should take to be relevant about a horse. We may pride ourselves that such iconoclastic domination of the tutorial process is no longer a property of contemporary educational practices. However, the key point here is, that notwithstanding the limitations of such a rote-approach learning, this fictional representation is accomplished in terms of the practice of memory; and its performance is part of what it is to display competent membership of the classroom setting within which Gradgrind, Bitzer, and all those other serried ranks of numbered pupils belonged.

It is fair to argue that until comparatively recently, psychologists have overwhelming thought of memory in terms of the storage of past experience. This container metaphor has dominated psychological research in the same way that talk of wax tablets and theatres was common for more ancient discussions on the topic. The problems that arise from treating mind as a vast warehouse wherein individual memories are cataloged and retrieved have been much rehearsed, not least by psychologists themselves. Such a view leads us to expect a kind of clarity and order to exist in human remembering that is simply absent from empirical evidence. Container metaphors may then ultimately be the wrong kind of metaphors in which to approach memory.

Work in psychology focusing on social practices of remembering and forgetting aims to displace the container view of memory in favor of an understanding of memory as a form of communicative action. Work in this area argues that psychological phenomena such as memory are better understood as accomplishments that occur in the course of communicative action, whether talk or text. Memory, on this account, is something that speakers/writers perform rather than simply possess in the course of routine interaction. These performances are informed by cultural understandings of what is to be counted as adequate and felicitous recall. The discursive study of memory entails a shift from viewing memory as some entity, whether in the mind or in the world. Instead the focus is on remembering and forgetting as social acts, as ways of accomplishing some activity in the present through invoking the past in an appropriate and resourceful manner.

This discursive approach also draws inspiration from Maurice Halbwachs' (1925/1992, 1950/1980) prioritization of language as the principle means in the collective organization of remembering and forgetting. If we treat remembering as a situated and contingent communicative act, three key issues emerge. First, since speakers must necessarily act toward and respond to one another in real time in order to sustain an interaction, we are led to attend to the ways in which remembering is sequentially organized. Second, given this sequential organization, we may analyze how inferences about the past are used by speakers when they co-opt one another into projects of remembering and forgetting. Third, that co-option into such projects involves the realization of memberships that are made relevant in the social occasioning and organization of remembering and forgetting.

These discursive issues – sequential organization, co-option, and membership – open up promising lines of empirical research into the social practices of memory. They also have relevance for understanding the memory practices in educational settings (see e.g., Edwards, 1987).

Interactive and Sequential Organization of Memory

Early exploration of discursive remembering focused upon the conversational devices we use in producing joint version of events and how we build a taken as given (default) continuity, as part of the sequential organization of talk concerning narrative sequence of events (Edwards and Middleton, 1986). In doing this, we share the burden of recall, comparing recollections, inserting out-of-sequence reminiscences, and commenting when we run into difficulty on what it is to remember and forget.

Such enquiry demonstrates the usefulness of conversational data in making visible the way we provide explicit formulations of inferential links and rationalizations concerning the past in our accounts of experience. In the course of such public thinking we often explicate such links and causal inferences. In addition, we justify and warrant versions against alternatives and the possibility of refutation. That is, our acts of remembering are always cast in an anticipated argumentative context. We are aware that we may be called upon, almost immediately, to justify, warrant, or produce supporting evidence for the claims we make about the past, and therefore build in to our claims rhetorical features that are designed to address possible objections (such as reporting mundane detail; using reported speech of others; and, for example, the manipulation of whether we are make a claim about the past from a first of third-person position – I vs. they).

These early findings, derived from conversational action within groups deliberately set up for the purposes

of studying joint recall, have been demonstrated in a wide variety of everyday contexts, including family, work, and leisure contexts where participants talk to each other in the social creation and justification of memories. However, the joint recall of events is a familiar and well-practiced activity that may occur across a wide range of settings, including classrooms (see e.g., Mercer, 2000). A range of linguistic resources are identifiable for remembering jointly in settings such as: overt requests for assistance that signal or invite ratification of claims concerning what may or may not have happened in the past; overt agreements (“yes that’s right”); ratification through the repetition of the previous speaker’s contribution; metacognitive formulations of the process of remembering itself and the specific problematic of remembering (“I’ve been trying to remember for quite a while”).

Metacognitive formulations are more than public expression of our personal understanding regarding the nature of our own mental processes (Flavell and Wellman, 1977). In fact, to understand these expressions solely in this way is to commit the same error as treating a schema as purely a mental representation. Metacognitive formulations arise in an occasioned manner, in particular discursive contexts, typically at points of difficulty – for instance, when speakers claim not to be able to remember something, or are suddenly reminded, or attempt to remember and try to fit a putative version of events with what another speaker is saying. On all these occasions, a public reflection on one’s own mental processes may strengthen or warrant a subsequent claim.

Instead of treating conversational action as a window upon mental processes, we can see that conversations act as significant environments in which thoughts are formulated, justified, and socialized according to how other speakers talk about mental processes. We do not need to know whether the claims people make about their mental processes – their memory – are actually correct or not in order to study how remembering and forgetting are socially accomplished (Middleton and Edwards, 1990). For example, consider the following two instances of talk analyzed by Charles Goodwin:

‘what was that guy’s name.=Blake?’

‘and what was the solution do you remember we discussing it here?’

Goodwin (1987) points out that displays of uncertainty in conversation are typically glossed in talk as remembering or forgetting. Orientations to uncertainty and forgetfulness can provide speakers with “resources for shaping their emerging interaction” (p. 116). In other words, saying something like “I don’t remember” is a good way for speakers to leave options open in the conversation. This is an instance of metacognitive construction – acquiring a conventional vocabulary and discourse for mental life designed to serve the social pragmatics of

conversation (Edwards, 1997). All of these linguistic devices contribute to the sequential unfolding of recollection, as each speaker's utterance dovetails with the last. A continuity by default is constructed, where each contribution builds upon and adds to the last.

Inference and Co-Option in Socially Organized Settings

Discursive analyses can also demonstrate that remembering is a fundamental aspect of the social ordering of the settings within which it occurs. Indeed, its occurrence appears to be functional to the continuing integrity not only of the psychological functioning of people in those settings but of the settings themselves. For example, we can examine conversations between parents and young children talking about family photographs. Such conversations frequently turn on establishing just what it is that people are looking at. However, the issue rarely stops there. In working out what is depicted in such snapshots, in jointly remembering, children are drawn into conversations that provide the basis for elaborating issues of identity, relationships, and emotional reactions to previous experiences.

In such conversations, past experiences are offered as a meaningful part of a continuing biography and development of personal identity. This is more than accessing the contents of memory. It is part of the ways in which what it is to remember and forget are used in communicative action. For example, parents may demonstrate something important both about remembering and about taking meaning from pictures – the role of inference: “it must have been a sunny day in that photograph(.) mustn't it?”. Photographs can be read for clues about the scene depicted. Sunny days are associated with pictures with sharp contrast and deep shadows, with sea and sand, blue skies, summer activities and clothing, and so on. None of those clues are explicit in what a parent says, but the implication is clear, that photographs in some way permit or afford inferences about the weather.

Parents can therefore demonstrate how remembering can be done by displaying how to read the photographs for clues to the contextual circumstances depicted. By so doing, children are taught to treat past events as memorable in terms of how they are affected and how they reacted to them as a family group. Inference serves as a basis for reconstruction. Children are therefore inducted into two sets of procedures. First, the use of artifacts such as photographs as inferential puzzles which may resource acts of remembering, and second, the resolution of such puzzles as a way of displaying a continuity between past and present. For instance, by establishing a continuity of personal experience over time, past experiences, and preferences, as they are inferred from the photographs, are

constructed as relevant to a discussion of the sort of persons children are at the time of speaking. This sort of discursive analysis also makes visible how we recruit or co-opt one another into social activities. Remembering can be achieved through talking to people. Children can learn how this is accomplished through the ways in which parents actively involve them in conversational projects concerning the past.

The collective accomplishment of remembering appears to be functional to the continuing and indeed developing lives of children and parents as members of family groups. As Bakhurst (1990) argues, in settings where remembering is mediated by cultural objects such as photographs, “we remember by constructing narratives which require the recall of past events for their intelligible completion” (p. 211). In other words, our narratives set up puzzles that are seemingly completed by the act of remembering, which as a consequence does a work of establishing continuities in identity. Such narrative puzzles then make up and sustain the very setting that entails what it is to be an accountable member of a grouping such as a family where the past is a direct topic of conversational concern.

Membership

This concern with continuity is illustrated in a final detailed example of discursive analysis taken from the conversations between a 4-year-old boy (Paul) and his mother as they looked through a family photograph album. Paul is constantly positioned as the next speaker, and as a consequence of the sequential organization of talk thereby co-opted into the project of remembering. How is the potential significance of Paul's experience achieved in terms of making his actions and supposed feelings (at the time of taking the photograph) relevant to family relationships, implied identities, and emotional significances? This involves the use of a series of contrasts between the interdependent experiences of a series of family members (including his father and sister).

From Middleton, 1997:

- Mother: oh look (.) there's when we went to the riding
[stables wasn't it?
Paul: [yeh (.) er er
Mother: you was trying to reach up and stroke that horse
Paul: where? (laughs)
Mother: would you like to do that again?
Paul: yeh
Mother: you don't look very happy though
Paul: because I thought I was going to fall off
Mother: you thought you was going to fall off did you?
(.) right misery (.) Daddy was holding on to you
so that you didn't (1) did it FEEL very bumpy?
Paul: yeh

- Mother: did it make your legs ache? [Paul laughs]
 Rebecca enjoyed it
- Paul: yeh
- Mother: she's a bit older wasn't she? (.) you were a little boy there

The mother's talk constantly appoints Paul as the next speaker in the project of remembering. In doing this, the mother provides candidate experiences in terms of Paul's actions and feelings, both physical and emotional, that would count as plausible interpretations of the events depicted and his position within them ("you was trying to reach up and stroke that horse"; "would you like to do that again?"; "you don't look very happy though"). The interactive appointment of Paul as accountable elicits an explanatory reminiscence from him ("because I thought I was going to fall off"). That is, Paul is required to produce a recollection in order to resolve the narrative puzzle that his mother has set up for him. In doing this, Paul provides an account occasioned in their talk for the display of unhappiness his mother has assigned to his actions in the photograph. What is more, the particular kind of account which Paul produces is of necessity one couched in terms of mental states, since only this kind of account will adequately address the narrative demands which have been interactionally established.

The mother responds by putting another question ("you thought you were going to fall off") and retakes the initiative of monitoring and displaying the form and content of the interaction in terms of candidate accounts of the actions, motives, and feelings of those depicted in the photo: of Paul as a ("right misery"); of his father's actions in the account offered of the events ("daddy was holding on to you"); for Paul's continued discomfort ("did it feel very bumpy"); and for his sisters engagement in the same event ("Rebecca enjoyed it"). In other words, the interactionally occasioned claims concerning his supposed experience of the events depicted in the snapshot are presented as relevant in terms of family relationships; relative and changing identities ("she's a bit older wasn't she"; "you were a little boy there"); and comparative affective reactions of family members ("Rebecca enjoyed it"). In summary, the experiences claimed on Paul's behalf are contrasted with those of other family members and with himself in the past. The negotiation of Paul's experience is then simultaneously a restating of what it means to be an accountable member of this particular family.

The sequential organization of such conversational remembering privileges the claims made in terms of their relevance beyond a pure statement of what is depicted in the photo – children, horses, adults, beaches, etc. We can see how the Mother's pursuit of contextual reminiscences co-opted Paul into justifying the claims being made on his behalf about the nature and consequence of his experience. For example, Paul is in some sense forced to accommodate

to the evaluation that he is a right misery because to do otherwise would involve challenging the mother's purported statement of fact that "daddy was holding on to you so that you didn't [fall]". When we use reminiscences in this way we interactively commit others to the relevance of our experience claims, or as in this case, the claims made by the mother on Paul's behalf. In this way, experience claims are not just an individual concern, they are presented as a concern of others and a concern in terms of the collectivities with which people are associated within conversational remembering, be they families, teams, classrooms, or any other educational setting. Accomplishing accountable memberships of a setting is no easy matter. The work of the conversation analyst Harvey Sacks (1992) is instructive on this issue. Sacks argues that it is not enough just to claim entitlements to the consequences of particular experiences. We must show them to be warranted in their conversational action. The point being that when we tell a story about events and describe their consequences for us, what we can say we felt is closely regulated in terms of the interactions of the moment. He makes the point that the experience and its consequences that we claim to have had in virtue of our participation in events must be the ones we are entitled to have. Sacks goes on to suggest and show that such entitlements are accomplished through the way we place ourselves as speakers in the events we narrate. This is of pertinence to the socially organized engagement of pupils and teachers in, and reproduction of, educational settings.

Summary

We can see how a discursive approach to psychology studies and how everyday versions of events (including persons, things, and states of affairs) are constructed and occasioned in talk and text. As descriptions of events can be indefinitely varied, discursive analysis examines how specific versions are produced and fitted to the occasions of their production (see Heritage, 1984). Overall, discursive analysis aims to reveal how versions are produced (how events are described, and so on), and what interactional business we accomplish by constructing our descriptions in one way rather than another (Potter and Wetherell, 1987). Any account of events reported is one of an indefinite number of possible reports. Often accounts are contrasted rhetorically, with alternative versions (Billig, 1996). Even the idea of versions being wrong can be treated through considering the way we orient to and address issues of accuracy, rather than deciding such matters in advance (as is typically done in experimental studies of memory). The factual and cognitive status of a report, as being offered from memory, or on the basis of direct experience, or as lies, error, or hearsay, are also studied as matters to be established by participants, rather

than as categories applied by analyst after the event. Psychological concerns are redefined within discursive psychology as those that may arise for participants (Edwards and Potter, 1992).

These issues go beyond the relevance of memory for educational settings as one of consigning, containing, and conscripting content in the service of understanding. It is possible to examine how our efforts as speakers at recollection are dependent on the contributions of others, or as premised upon a collectively held experience. Remembering therefore unfolds on a social and collective basis. Moreover, it is possible to identify various facets of the interactional organization of remembering, such as sequential organization, inference and co-option, and membership. These facets are all integral to the way speakers in educational setting communicate past experience as part of current pedagogical projects and activities.

See also: Learning in a Sociocultural Perspective; Memory; Situated View of Learning.

Bibliography

- Bakhurst, D. (1990). Social memory in Soviet thought. In Middleton, D. and Edwards, D. (eds.) *Collective Remembering*, pp 203–226. London: Sage.
- Billig, M. (1996). *Arguing and Thinking: A Rhetorical Approach to Social Psychology*, 2nd edn. Cambridge: Cambridge University Press.
- Edwards, D. (1987). Educational knowledge and collective memory. *Quarterly Newsletter of the LCHC* 9(1), 38–48.
- Edwards, D. (1997). *Discourse and Cognition*. London: Sage.
- Edwards, D. and Middleton, D. (1986). Joint remembering: Constructing an account of shared experience. *Discourse Processes* 9, 423–459.
- Edwards, D. and Potter, J. (1992). *Discursive Psychology*. London: Sage.
- Flavell, J. H. and Wellman, H. M. (1977). Meta-memory. In Kail, R. V. and Hagen, J. W. (eds.) *Memory in Cognitive Development*, pp 3–33. Hillsdale, NJ: Erlbaum.
- Goodwin, C. (1987). Forgetfulness as an interactive resource. *Social Psychology Quarterly* 50, 115–130.
- Halbwachs, M. (1950/1980). Dider, F. J., Jr., and Ditter, V. Y. (trans.) and Douglas, M. (intro.). *The Collective*. New York: Harper and Row.
- Halbwachs, M. (1925/1992). *On Collective Memory*, Coser, L. A. (trans. and Intro.). Chicago, IL: University of Chicago Press.
- Heritage, J. (1984). *Garfinkel and Ethnomethodology*. Cambridge: Polity Press.
- Mercer, N. (2000). *Words and Minds: How Do We Use Language to Think Together*. London: Routledge.
- Middleton, D. (1997). The social organisation of conversational remembering: Experience as individual and collective concerns. *Mind, Culture and Activity* 4(2), 71–85.
- Middleton, D. and Edwards, D. (1990). Conversational remembering: A social psychological approach. In Middleton, D. and Edwards, D. (eds.) *Collective Remembering*, pp 23–46. London: Sage.
- Potter, J. A. and Wetherell, M. (1987). *Discourse and Social Psychology: Beyond Attitudes and Behaviour*. London: Sage.
- Sacks, H. (1992). Lecture 4: Storyteller as 'witness'; Entitlement to experience. In Jefferson, G. (ed.) and Schegloff, E. A. (intro.) *Lectures on Conversation*, pp 242–248. Oxford: Blackwell.
- Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.
- Bellah, R. N., Madsen, R., Sullivan, W. M., Swidler, A., and Tipton, S. M. (1985). *Habits of the Heart: Individualism and Commitment in American Life*. Berkeley, CA: University of California Press.
- Draaisma, D. (2000). *Metaphors of Memory: A History of Ideas about the Mind*. Cambridge: Cambridge University Press.
- Edwards, D. and Mercer, N. (1987). *Common Knowledge*. London: Methuen/Routledge.
- Edwards, D. and Middleton, D. (1988). Conversational remembering and family relationships: How children learn to remember. *Journal of Social and Personal Relationships* 5, 3–25.
- Eisenberg, A. (1985). Learning to describe past experiences in conversation. *Discourse Processes* 8, 177–204.
- Fivush, R. (1994). Constructing narrative, emotion and self in parent-child conversations about the past. In Neisser, U. and Fivush, R. (eds.) *The Remembering Self: Accuracy and Construction in the Life Narrative*, pp 136–157. New York: Cambridge University press.
- Gibson, F. (1989). *Using Reminiscence*. London: Help the Aged.
- Harré, R. and Gillet, G. (1994). *The Discursive Mind*. Thousand Oaks, CA: Sage.
- Middleton, D. and Brown, S. D. (2005). *The Social Psychology of Experience: Studies in Remembering and Forgetting*. London: Sage.
- Middleton, D. and Edwards, D. (eds.) (1990). *Collective Remembering*. London: Sage.
- Miller, P. J., Potts, R., Fung, H., Hoogstra, L., and Mintz, J. (1990). Narrative practices and the social construction of self in childhood. *American Ethnologist* 17(2), 292–311.
- Norrick, N. R. (2000). *Conversational Narrative: Story Telling in Everyday Conversation*. Amsterdam: John Benjamins.
- Norrick, N. R. (2003). Remembering and forgetfulness in conversational narrative. *Discourse Processes* 36(1), 47–76.
- Orr, J. E. (1990). Sharing knowledge, celebrating identity: Community memory in a service culture. In Middleton, D. and Edwards, D. (eds.) *Collective Remembering*, pp 167–189. London: Sage.
- Schacter, D. L. (1996). *Searching for Memory: The Brain, the Mind, and the Past*. New York: Basic Books.
- Schrager, S. (1983). What is social in oral history? *International Journal of Oral History* 4, 76–98.
- Sutton, J. (2006). Remembering. In Robbins, P. and Aydede, M. (eds.) *Cambridge Handbook of Situated Cognition*, pp 217–235. Cambridge: Cambridge University Press.
- Wertsch, J. V. (2002). *Voices of Collective Remembering*. Cambridge: Cambridge University Press.

Further Reading

Relevant Website

<http://www.lboro.ac.uk/departments/ss/centres/darg/dargindex.htm>.

LEARNING AND COGNITION – ISSUES, CONCEPTS, TYPES

Contents

Computer-Supported Collaborative Learning: Basic Concepts, Multiple Perspectives, and Emerging Trends
Learning as Inquiry
Learning Through Narrative Socialization
Neuroscience Bases of Learning
The Neuroscience of Reading

Computer-Supported Collaborative Learning: Basic Concepts, Multiple Perspectives, and Emerging Trends

S R Ludvigsen and A I Mørch, University of Oslo, Oslo, Norway

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Computer-supported collaborative learning (CSCL) is the field concerned with how information and communication technology (ICT) might support learning in groups (co-located and distributed). It is also about understanding the actions and activities mediated by ICT. Its educational applications range from generic collaboration environments (e.g., forums) to tools for developing domain-specific knowledge. The research questions addressed by CSCL include how individuals learn with specific tools, how small groups interact and develop shared meanings over time, how institutions change and create new conditions for teaching and learning, and even how the opportunities for learning change as society adopts new models for education. Societies increasingly require new types of knowledge, new means of knowledge advancement and, consequently, new models of education.

To account for the multiple perspectives associated with CSCL without risking oversimplification, we adopt a sociocultural approach and present the main concepts and results. In particular, we make use of two overarching concepts – scaffolding and mediating artifact. Scaffolding is an instructional technique with which the teacher models the learning task, then gradually fades away and shifts responsibility to the students (Wood et al., 1976). In collaborative learning, students might also take on this role. By a technological scaffold, we mean features built into educational software that perform similar functions (e.g., guidance in virtual collaboration).

Related to scaffolding is the concept of mediation, proposed by Vygotsky (1986). This implies that technology for teaching and learning is first of all a mediating and

enhancing artifact. In the article we explore a special type of mediation, which is the relation between design-based research (DBR) and innovative CSCL tools (Collins *et al.*, 2004), on the one hand, and how this creates new opportunities for education on the other (Andriessen *et al.*, 2003).

CSCL emerged in response to skills that are important in a knowledge-based society. These are skills that were previously associated with deep learning of specialized knowledge, metacommunication, metacognition, and task reconceptualization (Järvelä and Salovaara, 2004). These skills are not easily taught through memorizing and fact finding using textbooks, which are the prevailing methods for learning basic skills. In fact finding, for example, the goal of the activity is most often invisible to students and the focus tends to be on tasks (Hewitt, 2001).

The shift in perspective concerning learning and cognition that CSCL provides is, in part, a result of the raised expectations in a knowledge-based society. The labor market of today demands: (1) specialized (domain-specific) skills and (2) an ability to work in teams (the capacity to integrate different types of knowledge and skills through collaboration). Specialized knowledge is important because the labor market is fragmented and interwoven in complex ways. The demands for skills in collaboration and knowledge integration (e.g., critically evaluate information resources found on the World Wide Web (WWW)) have come to the foreground over the past 10 years, making the teaching of communication, information seeking, and collaboration more important than ever.

CSCL is about teaching and learning the knowledge and skills required for participation in the knowledge-based society in concert with the basic skills they rely upon. The view we present here is consistent with those of

scholars who argue that the needs for specialized knowledge and collaboration skills must be met with a comprehensive approach (Järvelä and Salovaara, 2004; de Jong, 2006; Scardamalia and Bereiter, 2006).

CSCL is also a result of the widespread use of Web-based information systems and their acceptance by a broad group of actors at all levels in the education sector and by many workplaces (e.g., learning management systems and discussion forums). Two specific directions in educational research that have taken advantage of CSCL are ICT as mediators of accumulated knowledge (Paavola and Hakkarainen, 2005) and ICT as cultural tools (Wertsch, 1991; Ludvigsen and Mørch, 2003).

This article covers the major issues, research approaches, and questions concerning CSCL. We start by providing an overview of the field, including a presentation of key events and directions. Then we provide an overview of some important results grouped into two research approaches. Next, an overview of DBR and its implications for CSCL is provided. In closing, we identify the remaining open issues and point out some directions for further research.

Background

CSCL is a new and emerging field in the educational sciences (Stahl *et al.*, 2006). The term was first publicly used at an international workshop in 1989 in Maratea, Italy. The first international conference was organized in 1995 (Koschmann, 1996), and since then a biannual series of CSCL conferences has been arranged across Europe, North America, and Asia. In 2006, the *International Journal of Computer-Supported Collaborative Learning* (IJCSCL) brought-out its inaugural issue.

Broadly speaking, there are two main traditions within the learning sciences: cognitive psychology and the situated/sociocultural perspective. The former is based on the information-processing perspective (e.g., Anderson, 1993) and the latter on American pragmatism (e.g., Dewey, Mead, Garfinkel, and Schön,) and Soviet psychology (e.g., Vygotsky, Leontiev, Luria, and Wertsch). In CSCL studies, methods and techniques from both traditions are used and sometimes blended (e.g., interaction analysis). However, within each tradition there are unique interpretations of key concepts, methods, and empirical design.

Technical advances in computer science have contributed to CSCL in various ways. For example, researchers in computer-supported cooperative work (CSCW) have developed groupware systems that have been adopted for educational purposes (e.g., Stahl, 2006). Groupware provide shared spaces (Bannon and Bødker, 1997) on the WWW for storing and sharing information (messages, documents, pictures, and videos) and engaging the learners in social interaction (Girgensohn and Lee, 2002). When adopted in schools, they allow teachers and learners to interact online using a variety of communication and collaboration tools.

Examples of shared spaces are Basic Support for Collaborative Learning (BSCL) (Stahl, 2006), Future Learning Environment (FLE) (Muukkonen *et al.*, 1999), and Knowledge Forum (Scardamalia and Bereiter, 2006).

The relationship between computer support (CS) and collaborative learning (CL) in CSCL is complex as a result of the broad scope of the research questions, the multiplicity of approaches, and the plethora of educational technologies. This complexity needs to be unraveled in order to provide a more comprehensive account of the field. Although technology has had an enormous influence on CSCL, educational technologies should not be thought of as recipes for organizing teaching and learning. Similarly, the term CL does not imply that learning in small groups is better than individual learning. They are both needed and depend on each other. Findings indicate that detailed analyses of talk are necessary to understand how collaboration is carried out (Arnseth and Ludvigsen, 2006). It is not a question of an either/or situation, but rather how to identify specific situations that require mastery of new skills and design scaffolds for those situations. The common denominator is Vygotskian epistemology, which states that social interaction precedes learning and cognition at the level of the individual. To design for this requires CSCL tools and pedagogical models that foster social interaction and ease the transition from social interaction to learning and development. In this way, CSCL both represents a subfield in educational research and broadens the scope of educational research since it interacts with computer science and information systems.

Research Approaches in CSCL

We have grouped mainstream CSCL research into systemic and dialogical approaches (Arnseth and Ludvigsen, 2006). This distinction gives us the possibility to provide a more aggregated picture of what we know about research in CSCL.

Systemic Approach

The systemic approach concerns the generation of models of how specific features of technological tools afford or constrain collaboration, reasoning, knowledge representation, and inquiry (Dillenbourg, 1999) and to what extent these features will enhance students' capacities to solve problems in different domains (Arnseth and Ludvigsen, 2006). From a systemic approach, the analytic purpose is to identify interdependencies between quantifiable variables. The unit of analysis is the individually acting and thinking agent, and the two important cognitive processes are internalization (acquiring new or improving existing knowledge) and transfer (what is learned in one situation is applied to another similar situation). Models (often computer generated) of

how individuals construct, store, retrieve, and modify information serve as explanations of these phenomena (Anderson, 1993; Greeno, 2006).

Using a systemic approach in a traditional classroom setting in mathematics and reading, Lamon *et al.* (1996) demonstrated that students with CSCL tools performed better than students without such tools. The Jasper project (The Cognition and Technology Group at Vanderbilt, 1990) also reported similar results. The researchers in this large project studied the impact of educational technology on mathematics education in North America. They found that cooperative problem solving and discussion helped to engage the students in learning. In Europe, Fischer and colleagues studied how different types of scripts and tasks impact students' CL activities (Fischer and Mandl, 2005). Computational scripts were used to scaffold actions, and social scripts, such as role distribution, were used to organize turn taking. The authors found that scripts were useful for scaffolding learning and knowledge construction. However, these findings did not hold for all conditions (these conditions are further discussed in the last section of the article).

One of the most influential approaches within the systemic approach is knowledge building developed by Scardamalia and Bereiter (2006). Knowledge building is a model for distributed CL that is based on how professional scientists work to solve problems. The authors developed CSCL tools to support knowledge building (Computer-Supported Intentional Learning Environment (CSILE), Knowledge Forum). The latest version includes a Web-based shared space (WebCSILE). A further development of the Knowledge Forum is FLE (Muukkonen *et al.*, 1999).

The activities students engage in when involved in knowledge building can be formulated as a scientific-inquiry process, and many studies have been conducted using this approach (e.g., Hewitt, 2001). The phases of scientific inquiry include problem identification, proposing personal theories or hypotheses, experimentation, critical evaluation, data interpretation, scientific explanation, and summarizing. The studies in this area demonstrate that students who are engaged in knowledge building develop a deeper understanding of the domain under study. However, not all students benefit from it (Ludvigsen and Mørch, 2003). This may be related to the approach used, since a shortcoming in many of the studies is the timeframe adopted, which may range from a few hours to a few days. A consequence of this brevity is that conversational data is analyzed without taking into account the historical context of the interaction, which unfolds over time. The implication of such an approach is that the students' learning trajectories become less visible in the analysis (Crook, 1998; Rasmussen, 2005).

de Jong (2006) summarized recent research in scientific inquiry learning. He found that a number of students learn more effectively and develop deepening knowledge when supported with CSCL tools. The recommendation

de Jong proposes is that it is possible to design scaffolding mechanisms into CSCL environments that enhance students' learning. On the other hand, the results also show that most students have problems using predefined structures and processes adopted from professional science. An explanation for this difficulty is that the students do not have sufficient background knowledge to grasp the significance of the scientific inquiry process, and prefer instead to use everyday interpretations of scientific phenomena. Using a professional model of science to scaffold a learning environment provides a certain kind of insight, but it also generates new problems that are not easily resolved with a systemic approach.

In summary, the systemic approach gives useful guidelines for how we can build scaffolds for cognitive processes like hypothesis generation, data interpretation, and scientific explanation. However, this model-based approach to learning and cognition needs to be supplemented by a situational approach from a social and cultural perspective to provide a full account of CSCL.

Dialogic Approach

The dialogic approach is based on the idea that learning is a socially organized activity. The unit of analysis is a group of individuals interacting to accomplish a shared goal. Key concepts are mediation, artifacts and tools, and social practice, and mediation by tools to support learning is essential. It is through talk and interaction with significant others that we can understand how participants use tools and resources in learning and cognition. Externalization is seen as the main cognitive activity. The dialogic approach is influenced by research in situated learning and sociocultural perspectives (Greeno, 2006; Vygotsky 1986; Rommetveit, 1992; Valsiner and van der Veer, 2000; Wertsch, 1991). A basic premise is that both physical and abstract tools mediate human activities, and the main abstract tool is language (Vygotsky, 1986). The use of tools for learning is not only goal driven, but can also be seen in connection with how tools connect us with the past (predecessor artifacts) and with the future (unexplored potentials). The tools span both spatial and temporal dimensions.

In a study performed by Mercer and Wegerif (1999), students were exposed to a set of ground rules for communication. These rules included the use of arguments, disputes, clarifications, and explanations. The students and teachers were trained to talk together in specific ways in order to develop shared knowledge about a specific phenomenon of interest. In a series of interventions, the ground rules became a focus of the inquiry and previously implicit structures, like norms for participation, became explicit and transparent. This increased the probability for the kind of talk Mercer and Wegerif (1999) refer to as exploratory talk. Exploratory talk is characterized by the mutual

development of problems and ideas over time as a result of reflection and elaboration. They designed learning environments and new types of tasks to support these activities to promote more productive interactions in classrooms. It is further suggested that productive interaction needs to be understood not only as sequences of interaction, but as part of a broader context of institutional activities and sociocultural developments (Crook, 1998; Arnseth and Ludvigsen, 2006).

An example of a micro-study using the dialogic approach is the study of the effects of copy and paste on learning productivity. The cognitive effort involved in using copy and paste for text production and school presentations is low. As a result, many students use this technique uncritically. Some scholars have argued that it does not promote learning and should be discouraged (Hewitt, 2001; Kumpulainen and Wray, 2002). Using the dialogic approach, we can study this phenomenon in conjunction with how participants use the tools they have at their disposal to identify how the talk among the participants unfold as a result of tool mediation and emergent intermediate processes. Rasmussen (2005) found that students used copied texts as resources to deepen and broaden their understanding of the subject they studied.

Another finding concerning the use of the dialogic approach is that tasks are often open ended and cannot be taken for granted (Rasmussen, 2005). When studying how talk emerges in interaction as an analytic approach, the task needs to be constructed among the participants (Linell, 1998). The effect of this is that understanding the task becomes a learning activity in its own right, and this will stimulate the development of a higher-order skill (task conceptualization). When we assume that students working together share goals, task reconceptualization should be seen as an outcome of the activities rather than part of the premises for working together.

Suthers (2005) has identified intersubjective meaning making (Rommetveit, 1992) as one of the unique areas that CSCL is well equipped to support, and he suggests CSCL researchers undertake studies that attempt to understand how intersubjective meaning making impacts learning and how it can be mediated by technology affordances (Norman, 1999) embedded in CSCL tools. Suthers (2005) defines intersubjective meaning making as a joint composition of interpretations of a dynamically changing context. With this proposal, he provides a bold attempt to go beyond an information-sharing conception of CL. Technological affordances for exploratory learning and cooperative problem solving are proposed to support this process.

In summary, a dialogic approach to CSCL provides new analytic concepts to analyze how students and teachers interact in collaborative learning. The dialogic approach gives broader insights and explanations concerning the development of traditional skills, and pays particular attention

to skills such as those for communication, coordination, information seeking, information sharing, collaboration, negotiation, critiquing, and decision making, and how to design CSCL tools to support these activities.

Design-Based Research

Pedagogical Design

DBR has influenced research methodology in CSCL. DBR provides a solution for one of the dilemmas that confront researchers in the field – on the one hand, understanding how people learn, particularly within school settings, and on the other designing ways to ensure that learning will happen in a better manner in these settings (Brown, 1992; Collins *et al.*, 2004). The development of DBR has been on theoretical and methodological levels. On the methodology level, DBR suggests partnerships among researchers and educators with the goals of conducting rigorous and reflective inquiry, testing, and refining innovative learning environments, and defining new design principles based on previous research (Sandoval and Bell, 2004). On the theoretical level, design principles are the practical application of what we know about learning. As such, DBR does not provide direction for which research approach would be appropriate. Both the systemic and the dialogic approaches could be used.

Technology Design

The link between DBR and technology design is harder to establish. This is a result of the focus on theoretical and methodological issues (not well integrated with technological issues) and the difficulty involved in creating design principles that are practically useful for technology developers. There is an implied link between design principles and technology affordances in that the latter operationalize the former. However, this is a normative assumption upheld by some CSCL researchers, arguably strongest in the systemic tradition. Although many educational researchers agree that the basic principles of the sociocultural approach are important for the design of learning environments, the adoption of these principles have been hampered by a complex chain of elaborations before the principles can be used for developing specific tools. This is an important area for further work in CSCL (e.g., Suthers, 2005).

The basic idea of design principles in DBR is that we make use of what we know about previous research on learning when we design new learning environments. Although there is not an exact correspondence between the design principles proposed based on previous empirical studies and the design of an innovation for a new setting, the idea that the designers try to support the learning processes and anticipate its outcomes in specific directions

is likely to succeed over time. For example, the principle of deep learning can be found in many CSCL environments, in various forms (e.g., Linn *et al.*, 2004, Scardamalia and Bereiter, 2006). de Jong (2006) provides a recent overview.

The operationalization of design principles into technology affordances works best for principles that lend themselves to tool support, such as scaffolding (Wood *et al.*, 1976). This principle has been successfully incorporated into many CSCL tools, often in the form of automated feedback and/or guidance. Technology scaffolds takes advantage of regularities of:

- subject domains (Fischer *et al.*, 1991);
- knowledge types and scientific inquiry (Muukkonen *et al.*, 1999; de Jong, 2006);
- presence of others, group awareness, and social networks (Kreijns and Kirschner, 2004); and
- feedback and advice for online collaboration (Soller *et al.*, 2005; Mørch *et al.*, 2005).

One debate among technology developers concerns the degree to which computerized feedback should simulate or provide higher-level representations of user-interaction data before output (feedback or guidance) is generated. Soller *et al.* (2005) suggest three levels of feedback: mirroring (awareness), metacognitive tools, and guidance. This gradually increases the system's interpretation of the user data and consequently requires the users to be equally critically aware of the feedback generated.

Another debate explores to what extent automated feedback should be proactive, reactive, or requested (Mørch *et al.*, 2005). A system that provides sentence openers and step-by-step guidance is proactive. If it allows wrong actions to be taken before it gives hints and critique, it is reactive. If the system does not take any suggestive action on its own at all, but allows the learner to request guidance upon demand, it is requested (Mørch *et al.*, 2005). All three intervention strategies are important in CSCL environments, but not at the same time. The equation for balancing the three strategies depends on the complexity of the knowledge domain to be supported (e.g., the severity of making a wrong move vs. stimulating explorative learning) and the choice of research approach (systemic vs. dialogic).

Open issues and Directions for Further Work

In the 1990s, many people discussed how technology and the Internet would revolutionize schools and educational institutions. Now, after a 10-year period, these assumptions seem rather opaque and romantic. The CSCL research has, to a large degree, provided insight concerning the condition under which we can expect students to develop deep

knowledge using innovative technology support. A reasonable interpretation for the CSCL field across the different traditions we have surveyed in this article emphasizes that such capacity needs to be cultivated over a number of years, and it is dependent on the design of the learning environment, the social norms of the actors involved, and the institutional settings (Krange and Ludvigsen, 2008).

Two of the more general tendencies in complex CSCL environments are the following: first, that teachers and students need to engage deeply in specific problem-solving activities in order to learn concepts that are part of their actual knowledge development; second, that such a deep engagement often involves disagreement, identifying problems and conflicting ideas that need to be resolved (problematizing), and providing explanations, negotiations, etc. However, disagreement is not always a necessary condition. In exploratory talk, for example, reciprocal elaboration also serves as a means for engaging in deep learning. We need to conceptualize tensions, breakdowns, alignments, and elaborations as basic activities for learning to become productive for students over a long period of time.

Improving educational settings, with the scaffolding techniques for collaborative learning, is one of the aims of CSCL. This improvement includes pedagogical models and technological tools for problematizing tasks, hypothesis generation, elaboration, judgment about resources from the Internet, interpretation of data, evaluation of performance (metacognition), deliberate perspective shifts, etc. This is likely to be accomplished by scaffolding at the level of action and activities in classrooms, and supported by CSCL tools. The approaches developed by the CSCL community deal with micro-level phenomena and educational practices as seen from the teachers' and learners' points of view. In this way, CSCL has contributed to how schools can become better places for teaching and learning, and it is through the adoption and use of technology as a mediating artifact that it has achieved this status (Rasmussen, 2005; de Jong, 2006; Scardamalia and Bereiter, 2006). The historical tensions between old and new social practices create grounds for further development. It is the cognitive, social, historical, and institutional aspects, in combination, that must be taken into account for us to understand how we can improve the learning condition for students.

The systemic and the dialogical approaches to CSCL provide directions for how educational practices can and should be changed. However, an ongoing issue concerns how to constructively combine them without over simplification. Carefully designed (e.g., model-based) CSCL environments are important for improving our understanding of learning with ICT tools, but the analysis should be done from multiple perspectives, drawing on a broader set of student skills. Only by taking multiple perspectives as a starting point can we identify commonalities across

approaches that enrich our understanding of social interaction and its relationship with learning and cognition, as well as how to design new learning environments that enhance both productive learning and cognitive performance.

Acknowledgments

We want to thank InterMedia, The University of Oslo, and the Competence and Media Convergence (CMC) program at the university for financing our work with this article.

See also: Apprenticeship Approach to Learning; Classroom Discourse and Student Learning; Cultural-Historical Activity Theory; Learning in a Sociocultural Perspective.

Bibliography

- Anderson, J. R. (1993). *Rules of Mind*. Hillsdale, NJ: Erlbaum.
- Andriessen, J., Baker, M., and Suthers, D. (eds.) (2003). *Arguing to Learn: Confronting Cognitions in Computer-Supported Collaborative Learning Environments*. Dordrecht: Kluwer.
- Amseth, H. C. and Ludvigsen, S. (2006). Approaching institutional contexts: Systemic versus dialogical research in CSCL. *International Journal of Computer-Supported Collaborative Learning* 1(2), 167–185.
- Bannon, L. and Bødker, S. (1997). Constructing common information spaces. *Proceedings of the European Conference on Computer Supported Cooperative Work (ECSCW'97)*, pp 81–96. Dordrecht: Kluwer.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences* 2(2), 141–178.
- Collins, A., Joseph, D., and Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences* 13(1), 15–42.
- Crook, C. (1998). Children as computer users: The case of collaborative learning. *Computers and Education* 30(3 and 4), 237–247.
- de Jong, T. (2006). Scaffolds for scientific discovery learning. In Elen, J. and Clark, D. (eds.) *Handling Complexity in Learning Environments: Research and Theory*, pp 107–128. London: Elsevier Science.
- Dillenbourg, P. (ed.) (1999). *Collaborative Learning: Cognitive and Computational Approaches*. Amsterdam: Pergamon Press.
- Fischer, G., Lemke, A. C., Mastaglio, T., and Mørch, A. (1991). The role of critiquing in cooperative problem solving. *ACM Transactions on Information Systems* 9(2), 123–151.
- Fischer, F. and Mandl, H. (2005). Knowledge convergence in computer-supported collaborative learning: The role of external representation tools. *Journal of the Learning Sciences* 14(3), 405–441.
- Girgensohn, A. and Lee, A. (2002). Making web sites be places for social interaction. *Proceedings of the ACM 2002 Conference on Computer Supported Cooperative Work*, pp 136–145. New York: ACM Press.
- Greeno, J. G. (2006). Learning in activity. In Sawyer, R. K. (ed.) *The Cambridge Handbook of The Learning Science*, pp 79–96. Cambridge: Cambridge University Press.
- Hewitt, J. (2001). From focus on tasks to a focus on understanding: The cultural transformation of a Toronto classroom. In Koschmann, T., Hall, R., and Miyake, N. (eds.) *CSCL 2. Carrying Forward the Conversation*, pp 11–41. Mahwah, NJ: Erlbaum.
- Järvelä, S. and Salovaara, H. (2004). The interplay of motivational goals and cognitive strategies in a new pedagogical culture. A context oriented and qualitative approach. *European Psychologist* 9(4), 232–244.
- Koschmann, T. (1996). Paradigm shifts and instructional technology: An Introduction. In Koschmann, T. (ed.) *CSCL: Theory and Practice of an Emerging Paradigm*, pp 1–23. Mahwah, NJ: Lawrence Erlbaum.
- Krange, I. and Ludvigsen, S. (2008). What does it mean? Students' procedural and conceptual problem solving in CSCL environment designed within the field of science education. *International Journal of Computer supported collaborative learning* 3, 25–51.
- Kreijns, K. and Kirschner, P. A. (2004). Designing sociable CSCL environments. In Stribos, J. W., Kirschner, P. A., and Martens, R. L. (eds.) *What We Know about CSCL: And Implementing It in Higher Education*, pp 3–31. Boston, MA: Kluwer /Springer.
- Kumpulainen, K. and Wray, D. (eds.) (2002). *Classroom Interaction and Social Learning. From Theory to Practice*. London: Routledge/ Falmer.
- Lamon, M., Secules, T., Petrosino, A., et al. (1996). Schools for thought: Overview of the project and lessons learned from one of the sites. In Schauble, L. and Glaser, R. (eds.) *Innovations in Learning: New Environments for Education*, pp 243–288. Mahwah, NJ: Erlbaum.
- Linell, P. (1998). *Approaching Dialogue: Talk, Interaction and Contexts in Dialogical Perspectives*. Amsterdam: John Benjamins.
- Linn, M. C., Bell, P., and Davis, E. A. (2004). Specific design principles: Elaborating the scaffolded knowledge integration framework. In Linn, M., Davis, E. A., and Bell, P. (eds.) *Internet Environments for Science Education*, pp 315–341. Mahwah, NJ: Erlbaum.
- Ludvigsen, S. R. and Mørch, A. (2003). Categorisation in knowledge building: Task specific argumentation in a co-located CSCL environment. In Wasson, B., Ludvigsen, S., and Hoppe, U. (eds.) *Designing for Change in Networked Learning Environments. Proceedings of the International Conference on Computer Support for Collaborative Learning*, pp 67–76. Dordrecht: Kluwer.
- Mercer, N. and Wegerif, R. (1999). Is 'exploratory talk' productive talk? In Littleton, K. and Light, P. (eds.) *Learning with Computers. Analysing Productive Interaction*, pp 79–101. London: Routledge.
- Mørch, A., Jondahl, S., and Dolonen, J. (2005). Supporting conceptual awareness with pedagogical agents, information systems frontiers. *Special Issue. Computer Supported Collaborative Learning Requiring Immersive Presence* 7(1), 39–53.
- Muukkonen, H., Hakkarainen, K., and Lakkala, M. (1999). Collaborative technology for facilitating progressive inquiry: Future learning environment tools. In Hoadley, C. and Roschelle, J. (eds.) *Proceedings for: Computer support for Collaborative Learning. Designing New Media for a New Millennium: Collaborative Technology for Learning*, pp 406–415. Stanford University, CA: Erlbaum.
- Norman, D. A. (1999). Affordances, conventions and design. *Interactions* 6(3), 38–43.
- Paavola, S. and Hakkarainen, K. (2005). The knowledge creation metaphor: An emergent epistemological approach to learning. *Science and Education* 14, 537–557.
- Rasmussen, I. (2005). Project Work and ICT: A Study of Learning as Trajectories of Participation. PhD Thesis, InterMedia, University of Oslo, Norway.
- Rommetveit, R. (1992). Outlines of a dialogically based social-cognitive approach to human cognition and communication. In Wold, A. H. (ed.) *The Dialogical Alternative: Towards a Theory of Language and Mind*, pp 19–44. Oslo: Scandinavian University Press.
- Sandoval, W. A. and Bell, P. (2004). Design-based research methods for studying learning in context: Introduction. *Educational Psychologist* 39(4), 199–201.
- Scardamalia, M. and Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In Sawyer, R. K. (ed.) *The Cambridge Handbook of The Learning Science*, pp 97–118. Cambridge: Cambridge University Press.
- Soller, A., Martinez, A., Jermann, P., and Muehlenbrock, M. (2005). From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence in Education* 15, 261–290.
- Stahl, G. (2006). *Group Cognition: Computer Support for Building Collaborative Knowledge*. Cambridge, MA: MIT Press.
- Stahl, G., Koschmann, T., and Suthers, D. (2006). Computer-supported collaborative learning. In Sawyer, R. K. (ed.) *The Cambridge*

- Handbook of The Learning Science*, pp 409–425. Cambridge: Cambridge University Press.
- Suthers, D. D. (2005). Technology affordances for intersubjective learning: A thematic agenda for CSCL. In Koschmann, T., Suthers, D., and Chan, T. W. (eds.) *Proceedings of Conference on Computer Supported Collaborative Learning 2005*, pp 662–671. Mahwah, NJ: Erlbaum.
- The Cognition and Technology Group at Vanderbilt (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher* **19**(6), 2–10.
- Valsiner, J. and van der Veer, R. (2000). *The Social Mind: Construction of the Idea*. Cambridge, MA: Cambridge University Press.
- Vygotsky, L. (1986). *Thought and Language*. Cambridge, MA: MIT Press.
- Wertsch, J. V. (1991). *Voices of the Mind. A Sociocultural approach to Mediated Action*. Cambridge, MA: Harvard University Press.
- Wood, D., Bruner, J. S. and Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry* **17**, 89–100.

Further Reading

- Brown, J. S. (2000). *The Social Life of Information*. Boston, MA: Harvard Business School Press.

Learning as Inquiry

S R Goldman, J Radinsky, S Tozer and D Wink, University of Illinois, Chicago, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Learning as Inquiry

Inquiry as a purposeful approach to teaching and learning – sometimes called inquiry-based learning, or discovery learning, or inquiry-based teaching – is at least as old as Socrates. In the *Republic*, Socrates uses questioning techniques to prove to his students that anything worthy of the name knowledge had to be constructed by the learner through processes of questioning and answering. Teaching-as-telling could produce imitation, but not real knowledge. Even in arithmetic, he says, “No one uses it rightly; no one treats it as something that can truly lead to [knowledge of] reality.” (Plato in *Sterling and Scott*, 1985: 216, 523-a).

What is now called Socratic method was, in Plato’s telling, actually various approaches to using inquiry to help students learn. Chief among these were guided inquiry, in which the teacher already knew the answer but was posing questions so that the student might discover, construct, or even recover knowledge (as in the *Meno*), and shared inquiry, in which the teacher knew what questions to pose but was joining students in pursuit of an answer that even the teacher did not yet know. In either case, the teacher’s expertise lay in the method of finding answers, and his role was to help students learn that method, as well as the answers to which that inquiry method might lead. As Socrates said of the act of teaching, “[S]ome things are likely to provoke thought and some not” (*Republic*, 218, 524-d).

In short, inquiry learning foregrounds the question rather than the answer and places the locus of learning in the learner, not in the material that can be transmitted by the teacher. Although there are many approaches to inquiry learning, addressing all the core subjects taught in schools, these approaches share a defining feature: the learner’s engagement with the processes of answering one or more questions. The learner, the teacher, or the learning material can frame the question and the answer is developed from that point. This is a common design feature of inquiry-based curriculum materials for pre-K-12 schools. Ultimately, answers to questions do matter. However, in inquiry learning, the processes of inquiry are valued for their particular impact on the learner’s knowledge, skills, and dispositions toward learning itself. Educators and researchers find that these learning outcomes differ from those that arise when the learner focuses primarily on an answer to be memorized, a skill to be used algorithmically, or something otherwise learned apart from the process of arriving at that answer for him or herself.

The early history of public schools in the United States included little use of inquiry methods, Socratic or otherwise. Rather, rote memorization, recitation, and repetition were the primary approaches to teaching in rural and urban schools. In the twentieth century, progressive educators such as John Dewey began to explore how forms of inquiry learning could be implemented in public and private schools. “The rise of what is called new education and progressive schools,” Dewey wrote in *Experience and Education* “is of itself a product of discontent with traditional education. In effect it is a criticism of the latter” (Dewey, 1959: 4). In books such as *How We Think* (first published in 1910); *Democracy and Education* (first published in 1916); *Logic: The Theory of Inquiry* (first in 1938); and *Experience and Education* (also in 1938), Dewey so shaped contemporary thinking about the role of inquiry in teaching and learning that in 1992 Donald Schön titled an analysis of Dewey’s influence “The theory of inquiry: Dewey’s legacy to education’.

Like Plato, Dewey began with a notion of what it means to know something, rather than what it means to teach it or learn it. In contrast to Plato, for whom wisdom and virtue were the highest forms of knowing, Dewey focused on scientific knowing as the model for knowledge. For Dewey, science was an instructive model because it produced consensus about scientific answers only through shared agreement about the methods of inquiry needed to reach those answers. Dewey saw that the methods of science were essential for arriving at warranted knowledge, just as Plato saw that the methods of philosophical inquiry were indispensable for arriving at truth. In this way, Dewey wrote in *Experience and Education* of “scientific method as the pattern and ideal of intelligent exploration and exploitation of the potentialities inherent in experience” (1959: 108).

Dewey therefore saw the domain of science as particularly useful in thinking about teaching and learning in schools. To teach science as static subject matter, as opposed to teaching it through processes of inquiry, was not to teach science at all. To teach science meant that students needed to engage in the methods of inquiry essential to science, at whatever level their development would permit. Without using the term inquiry learning, in *Experience and Education*, Dewey described some of its particulars and challenges.

Adaptation of the method to individuals of various degrees of maturity is a problem for the educator, and the constant factors in the problem are the formation of

ideas, acting upon ideas, observation of the conditions which result, and organization of facts and ideas for future use. The ideas, the activities, the observations, or the organization are not the same for a 6-year-old person as they are for a 12- or 18-year-old, to say nothing of the adult scientist (Dewey, 1959: 112).

Dewey claimed that the experience of such learning differed markedly from the experience of more traditional learning approaches, and that the nature and quality of experience influenced the nature and quality of learning. Traditional learning approaches fostered rote learning and recitation of information as evidence of learning. Such shallow understanding would be less likely with inquiry-learning approaches. Students would not simply learn new subject-matter knowledge; they would also learn habits of mind – the disposition to form and pursue questions, the tendency to think critically and analytically, a reflective stance toward experience, and an interest in learning itself. Most importantly, he argued, these habits of mind are critical to future learning.

Collateral learning in the way of formation of enduring attitudes, of likes and dislikes, may be and often is more important than the spelling lesson or lesson in geography or history that is learned. These attitudes are fundamentally what count in the future. The most important attitude that can be formed is that of the desire to go on learning. (Dewey, 1959: 49)

Dewey and other progressive educators believed that such habits of mind were important to adult participation in a democratic society, which places a premium on individual as well as collective problem solving and decision making. In *Democracy and Education*, Dewey popularized the notion of the classroom as a community and as a laboratory for democracy, where groups of students could engage in shared problem solving and making decisions together under a teacher's guidance. *Democracy and Education* applied this thinking explicitly to a range of school subjects beyond science, demonstrating that different domains of knowledge afford different opportunities for students to engage in different kinds of learning – and that inquiry could be motivated and guided by student interests based on the prior experiences they brought to the classroom.

Despite the inquiry-learning influence of progressive-era educators, at no time in the twentieth century could it be said that inquiry teaching and learning were dominant in American public schools. Throughout the late twentieth century, studies of American classrooms typically documented modes of instruction that centered on teacher dissemination of information through presentation and books, rather than on students' engagement in inquiry methods, either guided or shared. However, research on learning over the past 40 years generated a renewed interest in inquiry learning in that it demonstrates that the learner is an active participant in the learning process. The active view of the learner contrasts

with the behaviorist tradition and its view of the learner as a passive recipient of knowledge, the view that pervades traditional classroom instruction.

Contemporary Classrooms and Inquiry Learning

Inquiry-focused education was simply not compatible with behaviorist traditions and instructional approaches based on the passive learner. Transmission from the teacher to the student was the dominant instructional philosophy, with metaphors of students as sponges or empty containers dominating education. However, findings from cognitive and social research conducted over the past 40 years painted a different picture of the learner and the role of disciplinary communities of practice. This research showed that learners were active participants in the acquisition and construction of knowledge and that learning was a social, interpersonal activity, not just an individual act. These characteristics were manifest in the functioning of communities of practice in the discipline. Furthermore, twenty-first-century society is a knowledge society with an increasing basis in technology. In a knowledge society, inquiry learning is an essential skill for successful and productive functioning. People need to be able to formulate appropriate questions and seek out their answers. Therefore, at the beginning of the twenty-first century, inquiry-based learning is increasingly evident in classroom instructional programs, albeit in many diverse forms.

Examples abound of both guided inquiry and shared inquiry, from unique teacher-invented learning exercises to commercial, mass-produced curricula that support teachers' use of inquiry learning in many different subject areas. The diversity of these reflect different strategies to include purposeful questioning by learners about a system, whether it is for learning about a discipline or learning to use knowledge obtained from a discipline. In addition, classrooms vary in the nature of the teachers, the students, the resources available, and the degree to which the school environment supports the independence that teachers need for inquiry teaching. This means that particular inquiry curricula are implemented in different ways in different classrooms and different student outcomes result. Implementation that is not completely faithful does not necessarily compromise the benefits of inquiry, although that is often the case.

Relationships between Professional Inquiry and Inquiry Learning in Schools

Many educators and scholars, usually following or responding to Dewey, have examined the nature of the relationship between professional communities of inquiry and their K-12 classroom counterparts. There are clearly

essential differences between these two communities: professional inquiry is conducted by an exclusionary community of people who have undergone extensive training and apprenticeship, while classroom communities are by definition inclusive, lacking in specialized training, and unfamiliar with many aspects of the domain. Professional inquiry involves well-established procedures for vetting claims, such as peer review, while classroom inquiry results in fragmentary claims that may lack complete information and lead to unwarranted arguments.

Beyond these differences in the participants and their backgrounds, disciplinary and classroom practices also differ in their epistemological assumptions. Scientists, historians, and other researchers investigate questions with the goal of creating new knowledge, challenging or clarifying existing assumptions. This contrasts sharply with the most common classroom goal of inquiry, that is, to help students acquire knowledge that is assumed to be already well established among more knowledgeable others outside the classroom community of inquirers. In this sense, the products of professional inquiry – understandings that are seen as tentative and subject to challenge within professional inquiry communities – are treated as something quite different by classroom communities, namely as established and unimpeachable facts.

Another difference lies in the source of authority in each community. Of necessity, a single authority (the teacher) exists in classrooms to serve as arbiter and facilitator of all inquiry activity. In contrast, complex hierarchies of authority exist within professional communities, with any member of the community vested with the authority to challenge any proposition made by any other. In this sense, classroom communities of inquiry may be seen as both more egalitarian, in that all students have nominally equal status, yet also more authoritarian, in that the teacher holds a position of ultimate authority which can only be challenged with his or her own consent.

In light of these differences, the appropriate relationship of inquiry activity in the classroom to professional inquiry is not self-evident. Some educators seek ways to simulate aspects of professional inquiry practices in classrooms and scaffold the knowledge and skills required to engage in them. Others attempt to simplify overly complex aspects of investigations, an effort that can have the unfortunate consequence of changing which cognitive processes are engaged by the inquiry process, sometimes turning inquiry into rote procedural execution. Still others seek to create ways for students to engage as peripheral participants in the actual work of scientific communities. In contrast, other approaches focus on the act of problematizing the conceptual content of the domain itself, rather than in the trappings of real-world investigations, seeing this activity as the most important relationship between professional communities of inquiry and the work done by students in classrooms. Finally,

there are those that question the relevance of the professional practices of scientists for the design of science instruction, instead focusing on the nature of inquiry as a knowledge-building practice that can be enacted in classrooms based on the experiences, lives, and cultures of the members of the classroom community itself.

Implementing Inquiry in Classrooms

The range of conceptions of inquiry that exist mean that, unsurprisingly, inquiry-based instruction takes on many forms in classrooms. It varies from single-classroom innovations designed by the teacher to large-scale curricula distributed by publishing houses or educational research-and-development institutions. Many of the established inquiry curricula share certain characteristic approaches to designing materials and learning experiences to scaffold classroom inquiry. At the same time, the particular characteristics of inquiry vary depending on the domain and, sometimes, the subdomains of inquiry (e.g., science, history, mathematics; biology, ecology, and physics) to reflect the epistemological orientation of the content area.

Inquiry Cycles

Notable among characteristic approaches to inquiry is the use of inquiry cycles that emphasize (to different degrees) the iterative relationship among research questions, investigation procedures, findings, and new or modified research questions. Such inquiry cycles typically move through a problematizing phase (engagement, posing of the challenging task), an investigation phase (e.g., guided inquiry within small groups), a feedback phase (e.g., analysis and reflection), and a revision phase based on the feedback. For example, the Inquiry Island curriculum and instructional approach includes six elements: ‘question, hypothesize, investigate, analyze, model, evaluate’, and returning to question. A similar circular model underlies the design of the inquiry-support software *Symphony*. It represents an investigation process as being constituted of five steps: develop problem, collect data, visualize data, model data, and review progress. A similar five-phase cycle for younger children consists of engage, prepare to investigate, investigate, prepare to report, and report. Important to the inquiry cycle is a driving question to necessitate collaborative investigation, communication of findings, and generation of work products that learners can talk about and use as a basis for feedback and revision.

Participation in these types of inquiry cycles in science supports students’ long-term engagement with their own learning. Each phase presents “different learning opportunities and teaching challenges” . . . and “different types of thinking and activity on the part of the students and the teacher; hence, each phase has a unique role to play in

supporting the development of scientific knowledge and ways of knowing” (Magnusson and Palincsar, 2005: 428). Within each phase, instruction emphasizes the metacognitive knowledge relevant to the work students will be doing, including the what, when, and why (i.e., the procedural and conditional knowledge) necessary for students’ developing expertise and ownership of strategies and concepts taught and applied appropriately to the content learning. Instruction itself takes the form of a gradual release of responsibility, in which teachers and students are co-participants in the learning process. As students gain more knowledge and control over strategies taught, they assume greater responsibility for the full learning cycle. The development of the metacognitive or reflective skills and habits needed to manage the complexity of these inquiry projects is often an explicit instructional goal.

Many inquiry-oriented interventions include the use of software or other highly structured materials to constrain and enable particular kinds of thinking and activity. This is in part because inquiry teaching often involves long-term projects with multiple phases and types of activity, requiring students to reason across many experiences to make sense of their investigation.

Some educators emphasize the value of more open-ended or student-driven inquiry processes as an alternative to sequential or otherwise highly structured investigations. In the domain of mathematics, applications of concepts to the real world are not the only way for students to engage in inquiry. Instead, students can engage in constructing solutions to problems about the conceptual issues of mathematics themselves and discover new relationships in mathematics. Given the opportunity to engage with such problems when learning mathematics, children’s inquiry replicates elements of argumentation and reasoning used by mathematicians, suggesting modes in which inquiry learning aligns to inquiry in a field. Conversely, it may be most important that students engage in discourse that draws out the unseen knowledge, assumptions, and experiences they bring into the classroom from their home cultures and lives into the classroom. The investigation of any science phenomenon through science talk that both values students’ understandings and challenges them to establish evidence for claims has the potential to become a focus of authentic scientific inquiry.

Although science and mathematics instruction are the most common home of inquiry curricula, other subject areas are increasingly providing examples of inquiry. For example, inquiry projects in history engage middle-school students in generating explanations for historical events, using historical documents to gather evidence to support or refute particular explanations, and critiquing one another’s interpretations. In one such investigation, students tried to establish the precise seat which Rosa Parks refused to surrender on a Montgomery bus, the incident that triggered the anti-segregation bus boycott of 1955.

Design projects in the domain of engineering emphasize inquiry and prioritize the work of creating an actual designed artifact as a key component of the inquiry process. These inquiry projects include not only design and artifact production but also phases of experimentation, testing, communication of results, and communal negotiation of the nature of the problem being addressed and the relative qualities of competing designs. This form of inquiry emphasizes knowledge as a building tool and brings out the usefulness of knowledge to solve problems.

Inquiry learning also can be found in the study of literature. Cultural modeling scaffolds high school students’ investigations of great American novels. This approach helps students develop inferences about layered meanings in the novel, supporting them with evidence drawn from the text. Knowledge building is an approach that provides computer tools for developing inferences about texts, giving and receiving feedback, and building connected knowledge across a community of inquirers. Finally, collaborative reasoning and book clubs engage youngsters of elementary school age in conducting inquiries into moral and social dilemmas often faced by the characters in literary works written for the young reader.

Challenges for Inquiry Learning in Classrooms

Research and development efforts over the past 15 years have brought to light several challenges that arise in conducting inquiry learning in the classroom, as well as approaches to dealing with the challenges. The four challenges discussed below echo the issue Dewey raised regarding the adaptation of the inquiry method so that it is implemented in classrooms in developmentally appropriate ways.

1. Providing sets of resources that enable students to understand the problem, including sources that provide information essential to solving the problem. Many informational text sources are too complex from a conceptual, vocabulary, and often grammatical perspective for students to be able to use them successfully. Some approaches provide students with strategies designed to assist them in understanding the key concepts and relations among the concepts. Examples of these are ‘Questioning the Author, Leslie’s Notebook’, text annotation, and concept webs or mapping. Other approaches find or adapt more complex sources so that the concepts are accessible to the target student group. Still others move away from traditional written texts as sources of information and use alternative media, such as video and dialog, to present the information. Embedded data and embedded teaching are examples. The issue of accessibility of text resources may be more problematic in some

disciplines than in others due to the importance placed on working with primary source documents.

2. Ensuring accessibility of the processes of thinking about the information and using it to address the problem to be solved. Students need to develop both the language for structuring acceptable arguments in a discipline and the processes for generating the content from sources. A number of scaffolding tools assist students in this. Some help students organize, sort, and categorize information, while others prompt for explanation, thinking, and reflection and some enable both.

3. Understanding and being able to use a discipline's representational formalisms. These include maps, various forms of data arrays (e.g., graphs and tables), equations and formulas, and dynamic and static physical models, for example, of chemical elements and their behavior. Efforts to introduce these formalisms to students include various technology-based modeling environments of dynamic graphs and chemistry phenomena. In addition, several groups of researchers make the issue of data representation a major focus of their work with school-age children.

4. Creating or providing relevant and meaningful problems that students care about. Such problems increase the likelihood that students will engage in intentional, conscious processes of learning complex ideas. A number of major development projects have produced inquiry materials rooted in significant societal, local, or personal needs and research indicates that these are engaging to students.

There are three challenges that relate to the classroom instructional context and the importance of the teacher in providing guidance to the inquiry process.

1. Teachers preparedness to guide inquiry-learning projects. Effective guidance of inquiry-learning activities requires that teachers understand the process and content of the inquiry. They must also understand how students are likely to engage with the inquiry and the problems and strategies that arise in moving students' thinking forward. Accordingly, many inquiry development projects include both development of curricula for students and teacher professional development. Some development efforts involve teachers as collaborators and, in so doing, create both professional development opportunities and curriculum materials.

2. Changing the norms of classroom discourse from the traditional pattern of teacher initiation, student response, and teacher evaluation (called I-R-E) to those of reasoned arguments in which claims are made and supported with evidence in ways appropriate to the discipline. A related change in classroom discourse patterns concerns the kinds of questions that teachers ask. Known-answer questions predominate in traditional classrooms: Teachers ask questions that they already know the answers to and students know that the teacher knows the answer. When teachers ask questions to which there are not known

answers, students have something genuine to contribute to the classroom discourse and become more engaged and energized. Such questions engage students in inquiry rather than test their knowledge. Making these changes in the patterns of classroom discourse is often challenging for teachers who typically have had little opportunity to experience this form of interaction themselves. Professional development institutes often model and provide opportunities for teachers to engage in more inquiry-oriented discourse.

3. Related to changing instructional interactions is the challenge of moving from the informal language of everyday life to more precise technical language of the discipline. As students engage in inquiry processes, their descriptive language needs to move toward the language used in the disciplines. For example, words such as conjecture, hypothesis, evidence, data, and conclusion take on specific meanings in the context of discipline-based inquiry. Yet, many of these words have less formalized meanings in the everyday world. Developing inquiry learning involves acquiring a good command of the ways in which members of the disciplinary community use these terms and expressions. Guiding the inquiry-learning process involves introducing and modeling the use of the language as much as the understanding of the content.

Although the challenges of doing inquiry in classrooms are substantial, there is an increasing body of evidence that instruction and materials can be designed to engage students with inquiry-learning problems over extended periods of time. This engagement produces evidence of student learning on assessments closely aligned with the inquiry-learning situations as well as on more generic achievement tests.

See also: An Overview of Research in Curriculum Inquiry; Classroom Discourse and Student Learning; Cognition: Overview and Recent Trends; Curriculum and Human Rights; Curriculum and Syllabus Design; Curriculum and Teacher Change; Curriculum Development in the Area of Reading; Curriculum Reform; Curriculum Use in the Classroom; Formative Assessment; Knowledge Domains and Domain Learning; Learning Science; Literature; Mathematics Learning; Mathematics; Pedagogical Content Knowledge; Problem Solving and Human Expertise; Problem Solving and Reasoning; Science; Situated View of Learning; Teacher Learning as Workplace Learning; Technology Supports for Acquiring Inquiry Skills; Technology Supports for Science Learning.

Bibliography

- Dewey, J. (1959). *Experience and Education*. New York: MacMillan.
 Magnusson, S. J. and Palincsar, A. S. (2005). Teaching to promote the development of scientific knowledge and reasoning about light at the elementary school level. In Bransford, J. D. and Donovan, S. (eds.)

How Students Learn, pp 421–474. Washington, DC: National Academy of Sciences.
 Plato (1985). *The Republic*, Sterling, R. W. and Scott, W. C. (trans). New York: W.W. Norton.

Further Reading

Barab, S., Zuiker, S., Warren, S., *et al.* (2007). Situationally embodied curriculum: Relating formalisms and contexts. *Science Education* 1–33. <http://www.interscience.wiley.com> (accessed June 2009).
 Cognition and Technology Group at Vanderbilt (1997). *The Jasper Project: Lessons in Curriculum, Instruction, Assessment, and Professional Development*. Mahwah, NJ: Erlbaum.
 Dewey, J. (1966). *Democracy and Education*. New York: Free Press.
 Goldman, S. R. (2005). Designing for scalable educational improvement. In Dede, C., Honan, J. P., and Peters, L. C. (eds.) *Scaling up Success: Lessons Learned from Technology-Based Educational Improvement*, pp 67–96. San Francisco, CA: Jossey-Bass.
 Kolodner, J., Camp, P. J., Crismond, D., *et al.* (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design into practice. *Journal of the Learning Sciences* 12, 495–547.
 Lee, C. D. (2001). Is October Brown Chinese? A cultural modeling activity system for underachieving students. *American Educational Research Journal* 38(1), 97–141.
 Linn, M. C., Davis, E. A., and Bell, P. (eds.) (2004). *Internet Environments for Science Education*. Mahwah, NJ: Erlbaum.
 Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., *et al.* (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. *Journal of Research in Science Teaching* 41, 1063–1080.
 Nemirovsky, R., Rosebery, A., Solomon, J., and Warren, B. (2004). *Everyday Matters in Science and Mathematics: Studies of Complex Classroom Events*. Mahwah, NJ: Erlbaum.
 Reiser, B. J., Tabak, I., Sandoval, W. A., *et al.* (2001). BGulLE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms.

In Carver, S. M. and Klahr, D. (eds.) *Cognition and Instruction: Twenty-Five Years of Progress*, pp 263–305. Mahwah, NJ: Erlbaum.
 Scardamalia, M., Bereiter, C., and Lamon, M. (1994). The CSILE Project: Trying to bring the classroom into World 3. In McGilly, K. (ed.) *Classroom Lessons: Integrating Cognitive Theory and Classroom Practice*, pp 201–228. Cambridge, MA: MIT Press.
 Songer, N. B., Lee, N., and McDonald, S. (2003). Research towards an expanded understanding of inquiry science beyond one idealized standard. *Science Education* 87, 490–516.
 Stevens, R., Herrenkohl, L. R., Bell, P., and Wineburg, S. (2005). The comparative understanding of school subjects: Past, present, and future research agenda. *Review of Educational Research* 75, 125–157.
 White, B. Y. and Fredericksen, J. R. (2005). A theoretical framework and approach for fostering metacognitive development. *Educational Psychologist* 4, 211–223.

Relevant Website

<http://www.thinklinklearning.com> – Adventures of Jasper Woodbury.
<http://www.biokids.umich.edu> – BioKids.
<http://www.ciera.org> – Book Club Plus.
<http://csr.ed.uiuc.edu> – Collaborative Reasoning.
<http://www.k12science.org> – Down the Drain.
<http://www.worldwatcher.northwestern.edu> – GEODE (Geographic Data in Education).
<http://www.hi-ce.org> – Investigating and Questioning Our World.
<http://www.knowledgeforum.com> – Knowledge Forum.
<http://www.perseus.tufts.edu> – Perseus Project.
<http://www.literacymatters.org> – Questioning.
<http://www.simcalc.umassd.edu> – SimCalc.
<http://www.globe.gov> – The Globe Program (Global Learning and Observations to Benefit the Environment).
<http://www.jasonproject.org> – The Jason Project.
<http://www.thinkertools.org> – Thinkertools.
<http://www.wise.berkeley.edu> – Web-based Inquiry Science Environment (WISE).

Learning Through Narrative Socialization

A Fasulo, University of Portsmouth, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Homesign – The sign language that spontaneously develop between deaf children and their hearing or nonsigning parents and caretakers.

Tellability – The property of a narrative that justify its emergence in the interactional context, and makes it relevant for its audience.

Narrative Socialization in the Life-Course

The capacity of the narrative to be an instrument of learning extends well beyond the stage of childhood socialization. Adolescents build their way into adulthood and gendered identities through gradual acquisition of narrative sophistication (Bamberg, 2004; Moissinac, 2007). Groups and institutions produce narratives as part of their daily functioning and these become fundamental socializing resources for newcomers and for the circulation of knowledge in general (Orr, 1990); societies reveal their major concerns and incipient changes in narrative fictions (Moretti, 1987). Anthropologists discuss narratives, especially those aimed to public representation, as reflexive devices through which community members reflect upon their social world and their role in it (Turner, 1974). Important changes in life – immigration (De Fina, 2003), conversion (Ayometzi, 2007), physical (Mattingly, 1998), or psychological treatments (Cain, 1991) – typically involve the emplotment (Mattingly, 1998) of ordinary actions and of one's life trajectory in a new but socially provided narrative form. Emotions have also been described as narrative plots that actors verbally and bodily enact when the situation calls for it (Sarbin, 2001).

The narrative is so central to human life that an evolutionary hypothesis wants it derived from mammals' chase play, as both are seen as interactional spaces in which to learn, in a neurophysiological condition aroused by entertainment adaptive behavioral models, and their variations (Steen and Owens, 2001; Hernadi, 2002). An independent motivation to narrate has been spontaneously produce narratives in homesign even when entirely deprived of language input, verbal or signed (Van Deusen-Phillips *et al.*, 2001). On the contrary, following Bruner and Feldman (1993), such an innate propensity to narrate is what is crucially missing in autism, so that autistic

children and adults are deprived of the basic communicative resources to share personal experience and access socio cultural knowledge. Narratives, in fact, not only contain ordered facts but also repertoires of interpretive procedures and lay theories about social and psychological functioning; narrative practice enables individual selves to develop along a path legible to comembers.

Narrative Discourse in Children

Early Narrative Stages

Antecedents to narratives have been identified in the formats, routine sequence of vocalizations and gestures which, in the story of parents and infants, evolve through variations and complications (Bruner, 1983). Such variations following Molinari (2002) are comparable to the complicating events that are the defining features of stories (Labov and Waletzky, 1967). Narrative activity has also been considered similar to symbolic play (Fasulo, 1997; Nicolopolou, 2002, Sawyer, 2002), for both reliance on scripted action and necessity of improvisation. Scarcity of research and theoretical efforts in this direction are also noted (Nicolopoulou, 2002): this might be due both to the more frequent focus of research on parent – child over peer interaction, the latter a setting in which fantasy narratives, play, and in fact the meshing of the two (Sawyer, 2002) are more likely to be encountered, and to a preference for personal or familial past events over other genres.

Genres

According to a growing number of studies on children's narratives, narrative discourse does not coincide with story telling solely, but includes a wider set of discourse types and genres. Heath and Branscombe (1986) and Ochs *et al.* (1992) enlarged the range of inquiry comprehending unproblematic reconstructions of events (respectively called recounts and reports), and ongoing description of action (event-casts and broadcasts). Meshing of narrative and nonnarrative genres have also been reported together with oscillations between different spatio-temporal locations (Ochs, 1994; Ochs and Capps, 2001). As a unifying feature and coding index for young children's narrative and protonarratives, Sperry and Sperry (1996) adopt the category of displacement, and are able to confront the longitudinal emergence and perseverance of a variety of past, future, and fictional genres (among the most striking

results, a higher percentage of fiction before 3–5 years, suggesting, in accordance with the play-narrative similarity, a common base in the work of imagination).

Such a broad perspective does not constitute an adjustment to immature language, but reflects an understanding of narrative as a natively hybrid genre, or metagenre, well established in literary theory (Bachtin, 1981). Genres, conceived of as linguistic means to accomplish culturally defined tasks (Eggins and Martin, 1997), are part of the interrelated set of practices children have to learn in order to act competently in the different provinces of their world (Wolf *et al.*, 1994).

Conarration

Narratives emerging naturally in interaction can show high degrees of conarration whatever the age of the participants (Goodwin, 1986; Lerner, 1992). As vehicles of evaluative perspectives (Labov and Waletzky, 1967), narratives can also easily become arenas where different views or interpretations are confronted.

Children's narratives at home exist basically in conarrated form: Parents and siblings can be responsible for eliciting narratives from children or introducing them, contributing substantial information, such as internal states and responses, and challenging the point, or argumentative line the narrative is trying to pull (Edwards and Middleton, 1988; Ochs *et al.*, 1992). Multiparty interactions more than dyadic ones represent complex environments in which, on the one hand, the floor is harder to gain and the audience more difficult to keep; on the other hand, family members' incursion into children narratives and the sheer listening in offer children precious occasions for narrative learning. In the talk of family reunited at the dinner table, children have accordingly been shown to narrate quantitatively less, compared with talk in dyadic settings, but to be involved in more complex participatory roles and to be more frequently faced with teasing and irony by family members (Aukrust, 2002). Also, the same data, reveal an overall higher proportion of narrative talk of all members in multiparty than dyadic interaction, thus exposing the child to more narrative discourse. As concerns the role played by children, children up to 6 years of age have been found to be more often the protagonists of stories told by others compared to their older siblings who figure more frequently as narrators; children in general appear to introduce far more less narratives than parents, although they may frequently become tellers of stories that parents ask them to narrate (Ochs and Taylor, 1992).

Participation formats also appear to reflect culturally valued aspects to narrative: Georgakopoulou's (2002) observations of middle-class families in Greece reveal a pattern of conarration inflected by the importance accorded in Greece to skilled narrative performance and

by the local practice of reiterating communal stories in support of group solidarity: Children would often prompt parents to tell stories from the family lore and also stories about the children themselves, while parents would intervene in the course of children's narration, assisting in the shaping of the stories' high point and seeing about the achievement of the best dramatic effects.

Such parental activity is bound to have long-term effects: coconstruction, via exposition to parental narrative styles, has been amply documented as affecting both quantity and nature of children's production, at least in the area of events' recall and autobiographical memory (Fivush, 2008; Fivush and Nelson, 2004; Reese, 2002).

Finally, participation may cross the borders of the present scene of the interlocutors to implicate absent actors: Goodwin (1982) illustrates a narrative genre through which preadolescent children instigate peers future confrontation by reporting words of an absent mate. These narratives sometimes developed into rehearsals of the future confrontational scene through imaginary dialogs. Practicing politics in the peer group, as well as sitting in at adult's meetings (Haviland 1986), foster learning on the subtleties of narrating about third persons.

As these studies demonstrate, narratives can be conceived of as chain-rings of interpersonal events, born to be repeated, spread across groups, and used like weapons or tokens of alliance. Inner politics of groups and communities, with their enmities and partisanship, have narrative as a resource for reaching remote audiences, gathering consensus, and keep relevant episodes within the active memory of the group.

Narrative Socialization to Cognition

Narrative Thinking

One of the most powerful effects of narrative relies in its fostering thinking through analogy, or *exempla*. From different arenas ranging from philosophy to psychology to history of science, analogical or narrative thinking has been acknowledged as a vital component in a variety of intellectual activities (Cazden and Hymes, 1980; Holton, 1992; Polkinghorne, 1988; Melandri, 2004). Even scientific paradigms (Kuhn, 1962) are sometimes described as normative models of a discipline built upon a core research enterprise that became the master narratives of the field. Concepts in the social and moral domains are often but a collection of stories about exemplar behaviors, positive or negative, and their expectable consequences (Feyerabend, 1999; Gergen, 1988).

Ontogenetically, as illustrated by the analyses of the child Emily's monologs (Nelson, 1989), narrative logic is mastered before piagetian logic and when the second appears in language is at the service of narrative rhetorics (Bruner, 1990) and narrative problem-solving

(Feldman, 1989). The joining and conflict between narrative versions, together with the pursuit of consensus, that characterize narrating within the family, has been likened to the process of scientific theory-building activity, where an initial storyline may lead to argumentation in the specialists' community and undergo several redrafting before achieving at least temporarily general acceptance (Ochs *et al.*, 1992).

Practicing narrative fosters abstract thinking also by exercising in decontextualized use of language, high-order coherence among interrelated actions of characters (Nicolopolou, 2002: 122), and the logic of part-whole relationships (Heath, 1983).

Narrative and Theories of the Mind

In the traditions of linguistic anthropology (Whorf, 1956; Cardona, 1985) sociology of knowledge (Berger and Luckmann, 1966) and discursive psychology (Edwards and Potter, 1992), language contributes in structuring the categories of experience and forming ideas about the functioning of the human mind. For example, the Western primacy of individual intentionality in the assessment of the meaning and moral weight of an action is not shared in many non-Western or Westernized countries (Duranti, 1993): such different perspectives on the relation between internal states and observable actions are enforced by a range of language practices and played out in narrative plots. Theories of minds are different across cultures although they may all rely on the assumption that others too have minds and socialization plays a key role in their formation. Children acquire lay psychological theories through immersion in the language of their community, but it is narratives that articulate more thoroughly observable events and behaviors with the inner world of thoughts and feelings (Bruner, 1986). The idea of others' minds and of their basic similarity to one's own is to a great extent owed to encounters with the represented minds of minds of narrative characters as well as to being a character in others' stories.

Narrative discourse embeds theories of cognition in conventional formulas and specific genres (e.g., the "do you remember when" or "I'll never forget..." initiations of, respectively, narratives of shared experience or extraordinary facts; Norrick, 2005). As, Reese (2002) points out, a theory upon others' mind is likely to be as much a product than a precursor of autobiographical memory, which is in turn deeply dependent on social factors.

Narrative Socialization and Sociocultural Groups

Although the studies reported above concerns children from very provenances, the discussion up to this point has

not taken into account the differences in narrative socialization on children of different gender, social class, or ethnic group. The following section presents an overview of research results focused on sociocultural differences and the production of culturally inflected identities.

Narrative Socialization and Gender

A large psycholinguistic study on 1300 elicited personal narratives of white children of working-class families has not revealed significant differences in the narratives of boys and girls, apart from girls resulting slightly more loquacious than their male counterparts (Peterson and McCabe, 1983). Differences have been found in subsequent studies in relation to narrative contents: analyzing narratives produced by preschools middle-class children in California and Massachusetts, Nicolopolou (1977a, b) finds that girls' characters and scenes were mainly drawn from family life or fairy tales themes, and their stories tended to depict harmonious relationships: boys, on the contrary, were more keen to represent "powerful and frightening characters" and include "conflict, movement, and disruption" (Nicolopolou, 2002: 127). In a different study on African-American children aged from 2 to 3.5, girls produced a lower number of narrative overall, and less fictional narrative than boys proportionally (Sperry, 1996). These studies, all set in the US, converge in showing a preference for boys to engage in imaginary worlds dominated by action and for girls to revolve on more tranquil and reality-based situations across different social origins and age of children. Mothers, in turn, have been found to be more elaborative, evaluative, and emotional when engaged in personal narratives with their daughters than with their sons (Fivush and Buckner, 2003), an observation matched by the presence in adult women of earlier and more complex memories (Pillemer, 1998).

Gender differences such as the one illustrated above do not ensure exclusively from caretakers' socialization, though: in her inspection of Grimm's fairy tales, Ortnor (1996) finds that the agency of female characters is either weaker than that of males – as when there are victims of misfortunes – or deflated in the course of the tale, when their initiatives end badly and it is left to male characters to resolve the complications and be rewarded for that. A very vast array of available narratives contribute, in synergy with other factors, to inform children's imaginative life along the preferential route that society designs for their gender.

Narrative Socialization and Social Class

In Bernstein's (1971) influential work on the relationship between social class, linguistic code, and school career, abilities to narrative appeared to vary with the children

familiar milieu. Middle-class children in UK, in an experiment he discusses, could more easily produce formally correct, decontextualized stories closer to the written code than working-class children; on the other hand, the stories of the latter were more rich and animated, and these narrators were more at ease with requests of invention and free elaboration. In school, though, only the former kind of stories were met by reward, coherently with the decontextualized approach to knowledge favored by educational institutions (Snow and Blum-Kulka, 2002). Bernstein discusses these results invoking socialization into societal roles and linguistic style of parental instructive practices. While groundbreaking at the time, Bernstein's work posits a dividing line between social classes which is problematic in light of the very different circumstances that could be subsumed under the middle- and working-class labels.

The issue of the internal differentiation of social classes is taken up by Miller and colleagues in a series of fieldwork studies on middle- and working-class children of different areas. Miller *et al.* (2005) discuss the results of a long-term project about young (2–3 years old) children's narrative in two working class communities in South Baltimore and Chicago, as well as middle-class community of another Chicago neighbourhood. The first community consisted of poorer, single-mother households which received public assistance or had low-income jobs. In the second one, children lived with both parents and the parents were regularly employed as blue-collar workers at jobs in which they had generally achieved higher positions with respect to parents from the first community. Researchers analyzed both adults' and children's narratives. Comparison between the two Chicago communities revealed that working-class mothers produced a higher rate of spontaneous narratives than middle-class mothers, and more frequently centered on negative events; their narratives, supported by a rich and dramatic performance, presented the image of a combative self struggling in the face of difficulties. Middle-class parents' narrative privileged instead a focus on inner experience and the use of emotion state terms. The same pattern held for the children of these mothers: both the South Baltimore and Chicago working-class children were more prolific narrators and recounted more negative events, although both the quantity and the dramatic intensity of negatively framed stories were higher in the poorer South Baltimore group. The preference of this group was also for colourful and action-based descriptions, in contrast with their middle-class counterparts' greater recourse to psychological language. In the working-class households, furthermore, children were often engaged in conarration, encouraged to recruit a public for the sharing of personal experience, and frequently contradicted, corrected, or otherwise pushed toward what was considered an adequate rendition of events, whereas middle-class mothers mitigated or withheld

correction even in spite of patent inaccuracies. This result is interpreted by the authors as indicating a more serious attitude toward narrative activity in working-class environments than in middle-class ones. However, as the authors remark, these data can also be read with a focus on the similarities among the groups, which largely overlap in engaging in conarration with the children, reminiscing both adults' and childrens' past and family events, and in short revealing that narrative is intended and used as a significant socializing tool.

Narrative Socialization and Cultural Communities

It has already been observed that the way narratives are told in the family reflects (and reinforces) local views on both what good stories look like and how they can serve societal values (Blum-Kulka, 1993; Georgakopoulou, 2002). Research on Italian families shows that children's narrative activity often takes place within disputes, and that family members' contributions supports the enhancement of children's rhetorical skills, encouraging them to present stories as persuasive evidence (Pontecorvo and Fasulo, 1997). These results are in line with Corsaro and Rizzo's (1988) observations regarding Italian preschool children routine engagement in dispute-like interactions.

Preferences for certain narrative topics over other may also vary across different communities. In a comparative study on Norwegian (Oslo) and US (Cambridge) families, Aukrust (2002) finds that in the former, which are richer in narrative altogether, there are also more frequent reports of children's school events, and a larger gallery of characters related to the children; more recurrent are also retellings of known events, although they are not simply replayed but enriched with new details. Differently, in the Cambridge corpus, the school genre enjoys a lower degree of tellability, often for the resistance of children to disclose details of their school life; when school narratives to appear, they tend to be about events still unknown to members. Overall, explanatory discourse genres are generally favored over narrative ones.

Focusing on the theme of transgression, Miller *et al.* (1997) compare Chicago and Taipei middle-class adults' and children's narratives, and find that Chinese parents typically recall children's transgression within a didactic narrative genre, whereas Chicagoan parents avoid to perpetrate the memory of such events or treat them in a humorous mode so to preserve a positive image of the child. The reverse holds for narrative about themselves or adults in general: in Taipei, parents never tell the children about transgression or faults committed in their young years, concentrating instead on instructive stories about their sacrifices and rewarded efforts on the contrary, Chicagoan parents may often disclose to their children transgressive episodes from their past (Miller *et al.*, 2001).

Interestingly, mothers from the two groups offered rather similar motives for such different practices, namely reducing distance between generations and conveying instructions for life. But whereas Taipei mothers thought that stories illustrating exemplar conducts would lead their children on the right path, Chicago mothers thought that acquaintance with parental misdeeds, beside being funny, would give children trust in the possibility of happy endings even in the face of initial mistakes.

Chinese mothers were therefore more didactic and serious in tone and emphasized children's immaturity and need to learn, whereas the American ones were more protective of the child image and less protective of their own. These results hint at different ideas of the construction of the "good life" (Freeman, Brockmeier, 2001), the first being steady and coherent and the second admitting false steps and diversions.

Following the results of the first comparative study by Miller *et al.* (1997) Van Deusen-Phillips *et al.* (2001) scanned their corpus of narratives from US (Philadelphia and Chicago) and Chinese (Taipei) deaf children for differences in the same direction. In line with the divergent ethics portrayed above, the authors found that Chinese children's narrative and protonarratives contained a much higher degree of evaluative comments than their American counterparts (100 vs. 5), and that these comments tended to be serious in the Chinese corpus and playful in the US one. As already reported, these children had not been exposed to language, narrative or otherwise: these findings suggest that essential cultural components of narrative can be socialized through embodied and other environmental resources.

Doing Research on Narrative Socialization

The present article has highlighted at several junctures how narrative is sensitive to multiple aspects of the situation of the telling. Hence, a crucial dimension for the study of children's narrative socialization is the methodology chosen for harvesting stories.

A multiparty setting has proven effective for the study of the range of participatory roles children can engage in, and of the types and quantity of other speakers' interventions in their stories (Ochs *et al.*, 1992), but the total number of narratives involving children in whatsoever role can be lower than in dyadic settings (Aukrust 2002). Anyhow, some ethnographers consider the multiparty situation a central site for children's socialization, in that it encourages observational learning and posits the need to sort out complex interpersonal demands. Researcher-elicited narratives can on the contrary provide robust narrative corpora, as proved by the seminal study by Peterson and McCabe (1983), but at the risk of scarce correspondence with *in vivo* story-telling characteristics:

Beals and Snow (2002) compared dinner-table to elicited narratives and found no correlation in the competence evidenced by the same children in the two settings.

School appears a setting wherein it is possible to construct a variety of narrative-relevant situations, circumventing uncontrolled contextual interpretations and also possibly creating, as Nicolopolou's research shows, highly motivating and socially meaningful tasks of narrative performance.

Interaction around written narratives has also proved an interesting site for the study of children understanding and production of narrative discourse (McArthur *et al.*, 2005). Perhaps counterintuitively, the written support does not impoverish narrative discourse but creates a steady common framework between child and adult for narrative elaboration and other types of reasoning. The paradigmatic and the narrative, as pointed out by Bruner (1990) can collaborate to enrich and expand the story realm. Also, joint reading is for children a gratifying experience and one they spontaneously engage in, whereas solitary consumption of texts seems to be more a normative prescription of adults that a spontaneous way of treating books (Sterponi, 2007).

A final comment concerns intercultural and intergroup comparisons. The study of children's narratives addressed across different communities is an invaluable window into the fundamentals of social and cultural groups, beside revealing the pedagogic ethnotheories of adult caretakers. A point by point comparison, though, as in all cross-cultural research, must be undertaken with some caution. Isolating children's stories from their growing soil and focusing on differences in skills, length, or other variables of production could, lead to interpretations of the resulting differences in terms of stereotypical or at best generic cultural traits, often accompanied by a negative portrayal of one of the groups (e.g., in terms of poor input, lack of pedagogical efforts etc.). The pioneering studies of Heath (1983), and the more recent research by Miller and her colleagues, are good illustrations of methodological choices that resist easy, dual minded comparisons in favour of dense and articulated accounts of the nature of differences and the wider circumstances of their emergence.

See also: An Overview of Language and Literacy in Educational Settings; Learning in a Sociocultural Perspective; Peer Interaction and Learning; Reasoning and Explanatory Talk: Learning in Everyday Settings.

Bibliography

- Aukrust, V. (2002). "What did you do in school today?" Speech genres and tellability in multiparty family mealtime conversations in two cultures. In Blum-Kulka, S. and Snow, C. (eds.) *Talking to adults*. Hillsdale NJ.: LEA.

- Ayometzi, Castillo, C. (2007). Stories as becoming: Identity through the telling of conversion. In Bamberg, M., De Fina, A., and Schiffrin, D. (eds.), *Selves and identities in narrative and discourse*. Amsterdam: Benjamins.
- Bakhtin, M. (1981). *The dialogical imagination. Four Essays*. Austin: University of Texas Press.
- Bamberg, M. (2004). "I know it may sound mean to say this, but we couldn't really care less about her anyway". Form and functions of "slut-bashing" in male identity constructions in 15-year-olds. *Human Development* 47, 331–353.
- Beals, D. E. and Snow, C. (2002). Deciding what to tell: Selecting and elaborating family topics in family interaction and children's elicited personal experience stories. In Blum-Kulka, S. and Snow, C. (eds.) *Talking to adults*. Hillsdale NJ.: LEA.
- Berger, P. and Luckmann, T. (1966). *The social construction of reality*. New York, Garden City: Doubleday and Co.
- Bernstein, B. (1971). *Class, Codes and Control*. London: Routledge & Kegan Paul.
- Blum-Kulka, S. (1993). You gotta know how to tell a story: Telling, tales and teller in American and Israeli narrative events at dinner. *Language in Society* 22, 361–402.
- Bruner, J. (1986). *Actual Minds, Possible Worlds*. Cambridge: Harvard University Press.
- Bruner, J. and Feldman, C. (1993). Theories of mind and the problem of autism. In Baron-Cohen, S., Tager-Flusberg, H., and Cohen, D. J. (eds.), *Understanding Other Minds: Perspectives from Autism*, pp 267–291. Oxford: Oxford University Press.
- Cain, C. (1991). Personal stories: Identity acquisition and self-understanding in Alcoholics Anonymous. *Ethos* 19, 210–253.
- Cardona, G. R. (1985). *I sei lati del mondo. Linguaggio ed esperienza [The six sides of the world. Language and experience]*. Laterza: Roma-Bari.
- Cazden, C. and Hymes, D. (1980). Narrative thinking and story-telling rights: A folklorist's clue to a critique of education.. In Hymes, D. (ed.) *Language in education: Ethnolinguistics Essays*. Washington DC: Center for Applied Linguistics.
- Corsaro, William, A. and Thomas, A. Rizzo (1988). Discussion and Friendship: Socialization Processes in the Peer Culture of Italian Nursery School Children. *American Sociological Review* 53, 879–894.
- De Fina, A. (2003). *Identity in Narrative. A study of Immigrant Discourse*. Amsterdam: John Benjamins.
- Duranti, A. (1993). Intentions, self and responsibility: An essay in Samoan ethnopragmatics. In Hill, J. H. and Irvine, J. T. (eds.) *Responsibility and Evidence in Oral Discourse*, pp 24–47. Cambridge: Cambridge University Press.
- Edwards, D. and Middleton, D. J. (1988). Conversational remembering and family relationships: How children learn to remember. *Journal of Social and Personal Relationships* 5, 3–25.
- Edwards, D. and Potter, J. (1992). *Discursive psychology*. London: Sage.
- Eggins, S. and Martin (1997). *Genres and registers of discourse*. In van Dijk, T. (ed.) *Discourse: A Multidisciplinary Introduction*. London: Sage.
- Fasulo, A. (1999). L'acquisizione della competenza conversazionale [Acquiring conversational competence]. In Galatolo, R. and Pallotti, G. (eds.) *La conversazione. Un'introduzione allo studio dell'interazione verbale, [Conversation. An introduction to the study of verbal interaction]*, pp 287–317. Milano: Cortina.
- Feldman, C. (1989). Monologues as problem-solving narrative. In Nelson, K. (ed.) *Narratives from the crib*. Cambridge: Harvard University Press.
- Feyerabend, P. (1999). *Conquest of abundance*. Chicago: The University of Chicago Press.
- Fivush, R. (2008). Remembering and reminiscing: How individual lives are constructed in family narratives. *Memory Studies* 1(1), 49–58.
- Fivush, R. and Buckner, J. P. (2003). Constructing gender and identity through autobiographical narratives. In Fivush, R. and Haden, C. A. (eds.) *Autobiographical memory and the construction of a narrative self: Developmental and cultural perspectives*. Mahwah, NJ: Erlbaum.
- Fivush, R. and Nelson, K. (2004). Culture and language in the emergence of autobiographical memory. *Psychological Science* 15(9), 573–577.
- Freeman, M. and Brockmeier, J. (2001). Narrative integrity: Autobiographical identity and the meaning of the "good life." In Brockmeier, J. and Carbaugh, D. (eds.) *Narrative and identity: Studies in autobiography, self, and culture*, pp 75–99. Amsterdam & Philadelphia: John Benjamins.
- Georgakopoulou, A. (2002). Greek children and familiar narratives in family contexts: En route to cultural performances. In Blum-Kulka, S. and Snow, C. (eds.) *Talking to adults*. Hillsdale NJ: LEA.
- Gergen, M. (1988). Narrative structures in social explanation. In Antaki, C. (ed.) (a cura di) *Analyzing lay explanation: A casebook of methods*. London: SAGE.
- Goodwin, C. (1986). Audience diversity, participation and interpretation. *Text* 6(3), 283–316.
- Goodwin M. H. (1982) "Instigating": Storytelling as a Social Process. *American Ethnologist*, 799–817.
- Haviland, J. B. (1986). 'Con Buenos Chiles' Talk, targets and teasing in *Zinacantan Text* 6(3), 249–282.
- Heath, S. B. (1983). *Ways with words. Language, life, and work in communities and classrooms*. New York: Cambridge University Press.
- Heath, S. B. and Branscombe, A. (1986). The book as narrative prop in language acquisition. In Schieffelin, B. B. and Gilmore, P. (eds.) *The Acquisition of Literacy: Ethnographic perspectives*, pp 16–34. Norwood, NJ: Ablex.
- Hernadi, P. (2002) Why is literature. A coevolutionary perspective on imaginative worldmaking.
- Holton, G. (1992). L'immaginazione nella scienza [Imagination in science]. In Preta, L. (ed.) *Immagini e metafore della scienza (Images and metaphors of science)*, pp 3–38. Bari-Roma: Laterza.
- Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University Chicago Press.
- Labov, W. and Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In Helm, J. (ed.) *Essays on the verbal and visual Arts*, pp 12–44. Seattle: University of Washington Press.
- Lerner, G. (1992). Assisted story-telling: Deploying Shared Knowledge as a Practical Matter. *Qualitative Sociology* 15(3), 247–271.
- Mattingly, C. (1998). *Healing dramas and clinical plots*. Cambridge, UK: Cambridge University Press.
- McArthur, D., Adamson, L., and Deckner, D. (2005). As stories become familiar: Mother-child conversations during shared reading. *Merrill-Palmer Quarterly* 51(4), 389–411.
- Melandri, E. (2004). *La linea e il circolo. Studio logico-filosofico sull'analogia* (The line and the circle. Logico-philosophical study upon analogy). Quodlibet: Macerata.
- Moissinac, L. (2007). "Mr. Lanoe hit on my mom": Re-establishment of believability in sequential narratives by adolescent boys. In Bamberg, M., DeFina, A., and Schiffrin, D. (eds.) *Selves and Identities in Narrative and Discourse*, pp 229–252. Amsterdam: John Benjamins.
- Moretti, F. (1987). *The Way of the World. The Bildungsroman in European Culture*. London: Verso.
- Miller, P., Cho, G., and Bracey, J. (2005). Working-class children's experience through the prism of personal story-telling. *Human Development* 48, 115–135.
- Miller, P., Sandel, T., Liang, C-H, and Fung, H. (2001). Narrating Transgression in Longwood: The discourses, meaning, and paradoxes of an American socializing practice. *Ethos* 29(2), 159–186.
- Miller, P., Wiley, A., Fung, H., and Liang, C-H. (1997). Personal story-telling as a medium of socialization in Chinese and American families. *Child development* 68, 557–568.
- Molinari, L. (2002). *Psicologia dello sviluppo sociale*. Bologna: Il Mulino.
- Nelson, K. (1989). *Narratives from the crib*. Cambridge: Harvard University Press.
- Nicolopoloulou, A. (2002). Peer group and narrative development. In Blum-Kulka, S. and Snow, C. (eds.) *Talking to adults*. Hillsdale NJ.: LEA.
- Nicolopoloulou, A. (1997a). Children and narratives: Toward an interpretive and sociocultural approach. In Bamberg, M. (ed.) *Narrative development: Six approaches*. Mahwah, NJ: Erlbaum.
- Nicolopoloulou, A. (1997b). Worldmaking and identity formation in children's narrative play. In Cox, B. D. and Lightfoot, C. (eds.) *Sociogenetic perspective on internalization*, pp 157–187. Mahwah, NJ: Erlbaum.
- Norrick, N. (2005). Interactional remembering in conversational narrative. *Journal of Pragmatics* 37, 1819–1844.

- Ochs, E. (1994). Stories that step into the future. In Biber, F. (ed.) (a cura di), *Perspectives on register: Situating register variation within sociolinguistics*, pp 106–135. Oxford: Oxford University Press.
- Ochs, E. and Capps, L. (2001). *Living Narrative*. Cambridge: Harvard University Press.
- Ochs, E. and Taylor, C. (1992). Family narrative as political activity. *Discourse and Society* 3, 43–57.
- Ochs, E., Taylor, C., Rudolph, D., and Smith, N. (1992). Storytelling as a theory-building activity. *Discourse Processes* 15, 37–72.
- Orr, J. (1990). Sharing knowledge, celebrating identity: War stories and community memory among service technicians. In Middleton, D. S. and Edwards, D. (eds.) *Collective remembering*, pp 169–189. London: Sage.
- Ortner, S. (1996). *Making Gender. The politics and erotics of culture*. Boston Ma: Beacon Press.
- Peterson, C. and McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at a child's narrative*. New York: Plenum.
- Pillemer, D. (1998). *Momentous events, vivid memories*. Cambridge, MA: Harvard University Press.
- Polkinghorne, D. E. (1988). *Narrative Knowing and the Human Sciences*. Albany, NY: State University of New York Press.
- Pontecorvo, C. and Fasulo, A. (1997). Learning to argue in family shared discourse. The reconstruction of past events. In Resnick, L., Säljö, R., Pontecorvo, C., and Burge, B. (eds.) *Discourse, tools and reasoning: Essays on situated cognition. NATO Series*. Berlin: Springer.
- Reese, E. (2002). Social factors in the development of autobiographical memory: The state of the art. *Social Development* 11(1), 2002.
- Sarbin, T. (2001). Embodiment and the narrative structure of emotional life. *Narrative Inquiry* 11(1), 217–225.
- Sawyer, K. (2002). Improvisation and narrative. *Narrative Inquiry* 12(2), 319–349.
- Snow, C. and Blum-Kulka, S. (2002). From home to school: School-age children talking with adults. In Blum-Kulka, S. and Snow, C. (eds.) *Talking to adults*. Hillsdale NJ.: LEA.
- Sperry, L. and Sperry, D. (1996). Early development of narrative skills. *Cognitive Development* 11, 443–465.
- Steen, F., Owens, S. (2001) Evolution's pedagogy. An adaptionist model of pretense and entertainment. www.cogweb.ucla.edu
- Sterponi, L. (2007). Clandestine interactional reading: Intertextuality and double-voicing under the desk. *Linguistics and Education* 18(1), 23.
- Turner, V. (1974). *Dramas, fields, and metaphors: Symbolic action in human society*. Ithaca: Cornell University Press.
- Van Deusen-Phillips, S., Goldin-Meadow, S., and Miller, P. (2001). Enacting stories, seeing worlds: Similarities and Differences in the cross-cultural narrative development of linguistically isolated deaf children. *Human Development* 44, 311–336.
- Whorf, B. (1956). *Language, thought, and reality*. Cambridge: M.I.T. Press.
- Wolf, D., Moreton, J., and Camp, L. (1994). Children's acquisition of different kinds of narrative discourse: Genres and lines of talk. In Sokolov, J. and Snow, C. (eds.) *Handbook of research in language development using CHILDES*. Hillsdale, NJ: Erlbaum.

Further Reading

- Bamberg, M. (1997). A constructivist approach to narrative development. In Bamberg, M. (ed.) *Narrative Development: Six Approaches*, pp 89–132. Mahwah, NJ: Erlbaum.
- Brockmeier, J., Wang, M., and Olson, D. R. (2002). *Literacy, Narrative and Culture*. Richmond: Routledge.
- Silliman, E. R. and Champion, T. (eds.) (2002). Narrative analysis in a multicultural society (special issue) *Linguistics and Education* 13(2), 143–150.
- Walkerdine, V. (1990). *Schoolgirl Fiction*. New York: Verso.
- Wang, Q. (2007). "Remember when you got the big, big bulldozer?" Mother-child reminiscing over time and across cultures. *Social Cognition* 25(4), 455–471.

Relevant Website

www.web.lemoyne.edu – Le Moyne College, Narrativepsych.com.

Neuroscience Bases of Learning

M H Immordino-Yang, University of Southern California, Los Angeles, CA, USA

K W Fischer, Harvard University Graduate School of Education, Cambridge, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Domain – A culturally constructed area of knowledge, such as language, math, music, or social interaction.

Neural network – A set of neurons that are structurally and functionally interconnected so that they activate in coherent patterns associated with mental functions.

Neuroimaging – A variety of research techniques, some invasive and some not, concerned with measuring and mapping the physiology and structure of the brain.

Neuromyth – A misguided, oversimplified, or incorrect tenet in education that concerns the brain or neuroscience.

Skill – An ability to behave or think in an organized way in a particular context.

Beyond Neuromyths: Mind, Brain, and Education Is a Cross-Disciplinary Field

All human behavior and learning, including feeling, thinking, creating, remembering, and deciding, originate in the brain. Rather than a hardwired biological system, the brain develops through an active, dynamic process in which a child's social, emotional, and cognitive experiences organize his or her brain over time, in accordance with biological constraints and principles. In the other direction, a child's particular neuropsychological strengths and weaknesses shape the way he or she perceives and interacts with the world. Like the weaving of an intricate and delicate web (Fischer and Bidell, 2006), physiological and cultural processes interact to produce learning and behavior in highly nuanced and complex patterns of human development.

People in the field of education often begin with a preconception that biology refers to traits that children are born with, that are fixed and unfold independent of experience, while children's social and cultural experiences, including schooling, are at the mercy of these biological predispositions, somehow riding on top of, but not influencing, biology. However, current research in neuroscience reinforces the notion that children's experiences shape their biology as much as biology shapes

children's development. The fields of neuroscience and more broadly biology are leading education toward analyzing the dynamic relationship between nurture and nature in development and schooling. A more nuanced understanding of how biology and experience interact is critically relevant to education. As neuroscientists learn about which aspects of experience are most likely to influence biology and vice versa, educators can develop increasingly tailored educational experiences, interventions, and assessments.

Due to this bi-directional relationship between a child's biological predispositions and social and cognitive experiences, the fields of neuroscience and education are coming increasingly into a research partnership. This relationship can be studied at many levels of analysis, from the workings of genes inside cells to the workings of communities inside cultures. However, in order for new information about the brain and learning to influence the design of learning environments, teachers and others involved in educational policy and design need to know about the newest principles about the brain and learning. Likewise, neuroscientists need to investigate phenomena that are relevant to real-world learning and development. To these ends, a new field has gradually taken shape over the last few years: mind, brain, and education (MBE). As a field, MBE encompasses educational neuroscience (a branch of neuroscience that deals with educationally relevant capacities in the brain), philosophy, linguistics, pedagogy, developmental psychology, and others.

In this interdisciplinary and applied climate, educators are in a particularly good position to help generate new questions and topics for research on learning and the brain, as they deal on a daily basis with the developmental issues and situations that affect real children and adults in their learning. For this reason, educators including teachers should have some familiarity with neuroscience and brain functioning, in order to become more informed consumers of educationally relevant findings as well as, ideally, contributors who help identify and shape new questions for neuroscience to pursue. For example, teachers can use information on the development of networks for numeric processing to design more effective curricula to teach math concepts, and educational assessments of students' math learning can help to shape new scientific questions about the development of math networks.

However, this does not mean that neuroscience is capable of contributing insights into all educational problems.

One of the challenges for the new field of MBE is for educators to learn about the applicability, implications, and limits of neuroscience research methods to various sorts of educational questions, and for neuroscientists at the same time to learn about the problems, issues, and processes of education, so that the two fields can collaborate as profitably as possible. For this to happen, educators and educational researchers need to know something about the tools, techniques, assumptions, and approaches that guide neuroscience research on learning, and need to develop a critical ability to consume and digest neuroscience findings and evaluate them for their potential applicability in the classroom. Toward this goal, teacher-training programs are beginning to incorporate information about the science of learning into their course offerings, and several new graduate programs in MBE have been launched at major universities in several countries in the last few years.

Before proceeding further, we felt the need to insert a strong cautionary note. As is typical during periods of rapid discovery, technological innovation, and theoretical advance, the field of MBE, as well as other related fields seeking to apply brain science to mainstream societal issues, are experiencing a lag between new technologies and findings on the one hand and the ability to interpret these findings on the other. In recent years, multiple examples of brain research misapplied have gone forward, including, for example, the overt labeling of elementary students as different categories of learners, from kinesthetic to auditory and beyond. Indeed, the scientific community agrees that much of what has been called brain-based education rests on very shaky ground. There is a proliferation of books written by nonscientists about the applications of neuroscience to learning, and while some of these books might present useful interpretations of neuroscience for educators, many of them suffer from a lack of basic understanding about the meaning and limitations of neuroscience research on learning and related processes. These books should be read with skepticism, as they often present models that are so oversimplified as to be misleading or even harmful or dangerous to children.

Overall, major changes in neuroscience research methods and theory are allowing better applicability of brain findings to educational issues and questions, and new insights into the processes that happen in schools. In this article, we focus on the prominent contribution of neuroimaging to the current view of learning as the construction of distributed neural networks that support skills, and how the development and recruitment of these neural networks is modulated and facilitated by domain-general processes in the brain, including emotion, attention, and mechanisms of social learning. We conclude with a call for further research that evaluates neuroscientific principles as they play out in classroom contexts.

New Neuroscience Methods Bring New Information and New Challenges for Interpretation

Educators' views of brain research have shifted in the past few years. While many educators continue to cling to so-called neuromyths, neuroscientists in the MBE field have been working to dispel these myths. In particular, the last decade has seen huge advances in *in vivo* neuroimaging technologies. Scientists are now able to study the workings of the human mind in healthy participants as they solve problems and perform other sorts of cognitive and emotional tasks in real time. Availability of these new-research technologies is pushing the field forward at an unprecedented pace; hardly a week goes by, it seems, without a picture of the brain appearing on the cover of a major magazine or in a major newspaper article.

To make sense of the new findings, it is critical that educators understand the logic and constraints in the neuroscience research underlying these articles. While neuroimaging techniques differ in their specifics, there are three main approaches. The first approach involves measuring and localizing changes in the flow of blood in the brain as subjects think in different ways, under the assumption that changes in regional blood flow are indicative of changes in neural activity. The second approach involves measuring the electrical activity of the brain, generated by the firing of networks of neurons (brain cells). The third approach involves measuring changes in the anatomy and structure of the brain. In conjunction or separately, these techniques can be used to study the neurological correlates of a wide variety of tasks, such as reading, math, or social processing, as well as developmental changes (for reviews, see Katzir and Pare-Blagoev, 2006; Thatcher *et al.*, in press).

While these recent advances in neuroimaging have had a profound effect on the field of neuroscience and its potential relevance to education, it is important to remember that new technological capabilities inevitably come with limitations. For example, in functional magnetic resonance imaging (fMRI), the changes in regional blood flow in the brain associated with a particular task of interest are not absolute, but either implicitly or explicitly calculated from comparisons between a target and a control task. The design of the two tasks and the differences between them are critical to the findings and interpretation. When one brain area is reported to light up (i.e., to become more active) for a particular task, this does not mean that the lighted brain area is the only area actively processing. Instead, this means that this particular area was relatively more active for this task than for the control task. Many other areas are certainly actively involved, but are equivalently active in the two conditions. In reality, a network of neural areas always supports the skill being tested. As educators are concerned with supporting the

development of coherent functional skills rather than isolated brain areas, it is essential that neuroimaging findings be correctly interpreted before any attempt can be made to apply them in the classroom.

Educational Skills are Supported by Specialized Neural Networks

Nonetheless, the advent of neuroimaging has precipitated major advances in neuroscientists' understanding of how the brain works. In the past, the neuroscientific localization tradition prevailed; that is, cognitive functions were mapped onto specific locations in the brain, as much as possible in one-to-one correspondence. However, neuroscientists now understand that learning involves the development of connections between networks of brain areas, spread across many regions of the brain. This means that while specific brain areas do carry out characteristic kinds of processing, skills for real-world and academic tasks are embodied in the networks they recruit, rather than in any one area of the brain. For example, there is no music, reading, or math area of the brain that is not also involved in processing many other skills and domains (culturally constructed areas of knowledge).

Instead of one brain area, learning involves actively constructing neural networks that functionally connect many brain areas. Due to the constructive nature of this process, different learners' networks may differ, in accordance with the person's neuropsychological strengths and predispositions, and with the cultural, physical, and social context in which the skills are built (Immordino-Yang, 2008). There are various routes to effective skill development, for example, in reading (Fischer *et al.*, 2007) or math (Singer, 2007). The job of education is to provide support for children with different neuropsychological profiles to develop effective, yet flexible skills. Children use whatever capacities they have to learn the most important skills in their lives, and although there is often a modal way of learning a specific skill, people can adapt their capacities to learn skills in diverse ways. For example, Knight and Fischer (1992) found that young children followed one of three pathways in learning to read words. In a related vein, in studying two high-functioning adolescent boys who had recovered from the surgical removal of half of their brain, Immordino-Yang (2007) found that each boy had compensated for weaknesses by transforming important neuropsychological skills into new ones that suited the boys' remaining strengths.

Neural Networks for Mathematics

One area that has seen much advance in the past few years is the study of neurological networks underlying processing

for mathematics and number representation. Overall, the findings suggest that networks for processing in math are built from networks for the representation of quantity that start in infancy – one for the approximate representation of numerosity (numeric quantity), and one for exact calculation using numbers (Dehaene *et al.*, 2004). These networks are further organized and differentiated with development and training in math concepts (Singer, 2007). For example, preschoolers go beyond innate number systems to build a mental number line, gradually adding one digit at a time (Le Corre *et al.*, 2006).

Interestingly, this math network shares many processing areas and features with language processing, including reading. Current research is exploring how math processing relates to other domains, such as spatial representation, as well as the development of math networks in atypically developing populations, such as children with learning disabilities.

Neural Networks for Reading

Another area of concentrated research interest is the study of reading development, both in typically developing and dyslexic children. Acquiring literacy skills impacts the functional organization of the brain, differentially recruiting networks for language, visual, and sound representation in both hemispheres, as well as increasing the amount of white-matter tissue connecting brain areas. Work on individual differences in the cognitive paths to reading has enriched the interpretation of the neurological research (e.g., Knight and Fischer, 1992), and helped to bridge the gap between the neuroscience findings and classroom practice (Katzir and Pare-Blagoev, 2006; Wolf and O'Brien, 2006). In dyslexic readers, progress is being made toward better understanding of the contributions of rapid phonological processing (Gaab *et al.*, 2007), orthographic processing (Bitan *et al.*, 2007), and visual processing to reading behaviors, as well as to thinking in other domains (Boets *et al.*, 2008). For example, the visual field of dyslexics may show more sensitivity in the periphery and less in the fovea compared to nondyslexics, leading to special talents in some dyslexics for diffuse-pattern recognition (Schneps *et al.*, 2007). Most recently, research looking at developmental differences in neurological networks for reading across cultures has begun to appear (e.g., Cao *et al.*, 2009), which ultimately may contribute to knowledge about how different kinds of reading experiences shape the brain.

The neural networks for learning reading and math have important implications for education, as the most effective lessons implicitly scaffold the development of brain systems responsible for the various component skills. For example, successful math curricula help students to connect skills for calculation with those for the representation of quantity, through scaffolding the

development of mental structures like the number line (Carey and Sarnecka, 2006; Griffin, 2004; Le Corre *et al.*, 2006). While different students will show different propensities for the component skills, all students will ultimately need to functionally connect the brain systems for quantity and calculation to be successful in math.

Domain-General and Emotion-Related Processes Enable Learning

The brain is a dynamic, plastic, experience-dependent, social, and affective organ. Due to this, the centuries-long debate over nature versus nurture is an unproductive and overly dichotomous approach to understanding the complexities of the dynamic interdependencies between biology and culture in development. New evidence highlights how humans are fundamentally social and symbolic beings (Herrmann *et al.*, 2007), and just as certain aspects of our biology, including our genetics and our brains, shape our social, emotional, and cognitive propensities, many aspects of our biology, including processes as fundamental as body growth, depend on adequate social, emotional, and cognitive nurturance. Learning is social, emotional, and shaped by culture!

For a stark example of this interdependence between biology, social interaction, and cognitive stimulation, in their work with Romanian orphans, Nelson *et al.* (2007) found that cognitive, social, and physical growth were delayed in institutionalized children, relative to their peers raised in foster or biological families. Although the institutionalized children's basic physical needs were met, the lack of high-quality social interaction and cognitive stimulation lead these children not to thrive.

Overall, while educators often focus on neural networks for domain-specific skills like reading and math, domain-general and emotion-related networks function as modulators and facilitators of memory and domain-specific learning. These networks include emotion, social processing, and attention.

Emotion and Social Processing

One cutting-edge area of research in neuroscience is the study of affective and social processing. All good teachers know that the way students feel, including their emotional states (e.g., stressed vs. relaxed, depressed vs. enthusiastic) and the state of their bodies (e.g., whether they are sick or well, whether they have slept enough, or whether they have eaten), are critical factors affecting learning. In addition, it is now becoming increasingly evident that emotion plays a fundamental role not only in background processes like motivation for learning, but in moment-to-moment problem solving and decision making as well (Adolphs and Damasio, 2000; Haidt, 2001). That is, emotion forms the

rudder that steers learners' thinking, in effect helping them to call up information and memories that are relevant to the topic or problem at hand. For example, as a student solves a math problem, she is emotionally evaluating whether each cognitive step is likely to bring her closer to a useful solution, or whether it seems to be leading her astray.

From a neurobiological perspective, emotional processing in the brain depends on somatosensory systems – the systems in the brain responsible for sensing the state of the viscera and body. These systems can reflect actual changes to the state of the body during emotions (i.e., increased heart rate during fearful states, or a feeling of having been kicked in the stomach when hearing bad news), or they can reflect simulated body states, conjuring how the viscera and body would feel, without actually imposing those physiological changes onto the body (see Figure 1 from Immordino-Yang and Damasio, 2007). Through regulating and inciting attention, motivation, and evaluation of simulated or actual outcomes, emotion serves to modulate the recruitment of neural networks for domain-specific skills, for example, for math or reading. In this way, cognition and emotion in the brain are two sides of the same coin, and most of the thought processes that educators care about, including memory, learning, and creativity among others, critically involve both cognitive and emotional aspects (Figure 1).

In addition, social processing in the brain is strongly interrelated with the processing of emotion. People's behavior is organized and influenced by cultural factors and the social context, which in turn reflect experience and learning. For example, many of the reasons the student above solves her math problem relate to the emotional aspects of her social relationships and cultural goals – the way her parents will feel about her behavior, or her desire to go to college. In turn, she feels the influences of these cultural constructs as emotional reactions that play out in her body and mind, and predispose her to think in particular ways.

But how does this student internalize or predict the emotional reactions of her parents? Interestingly, research over the past decade has revealed glimmers of the workings of a basic biological system for internalizing the actions, emotions, and goals of others, in order to learn from, empathize with, and influence others in social contexts (Immordino-Yang, 2008; Oberman *et al.*, 2007). Specifically, it appears that watching other people's actions and inferring their emotions and implicit goals recruits some of the same neural systems involved in planning and carrying out those actions in one's own self. This discovery was dubbed as mirroring by its discoverers (Gallese *et al.*, 1996; Umiltà *et al.*, 2001), and while neural systems for mirroring do not tell the whole story of the neurological system for social learning, current research suggests that they afford an important low-level mechanism on which social and cultural learning can build.

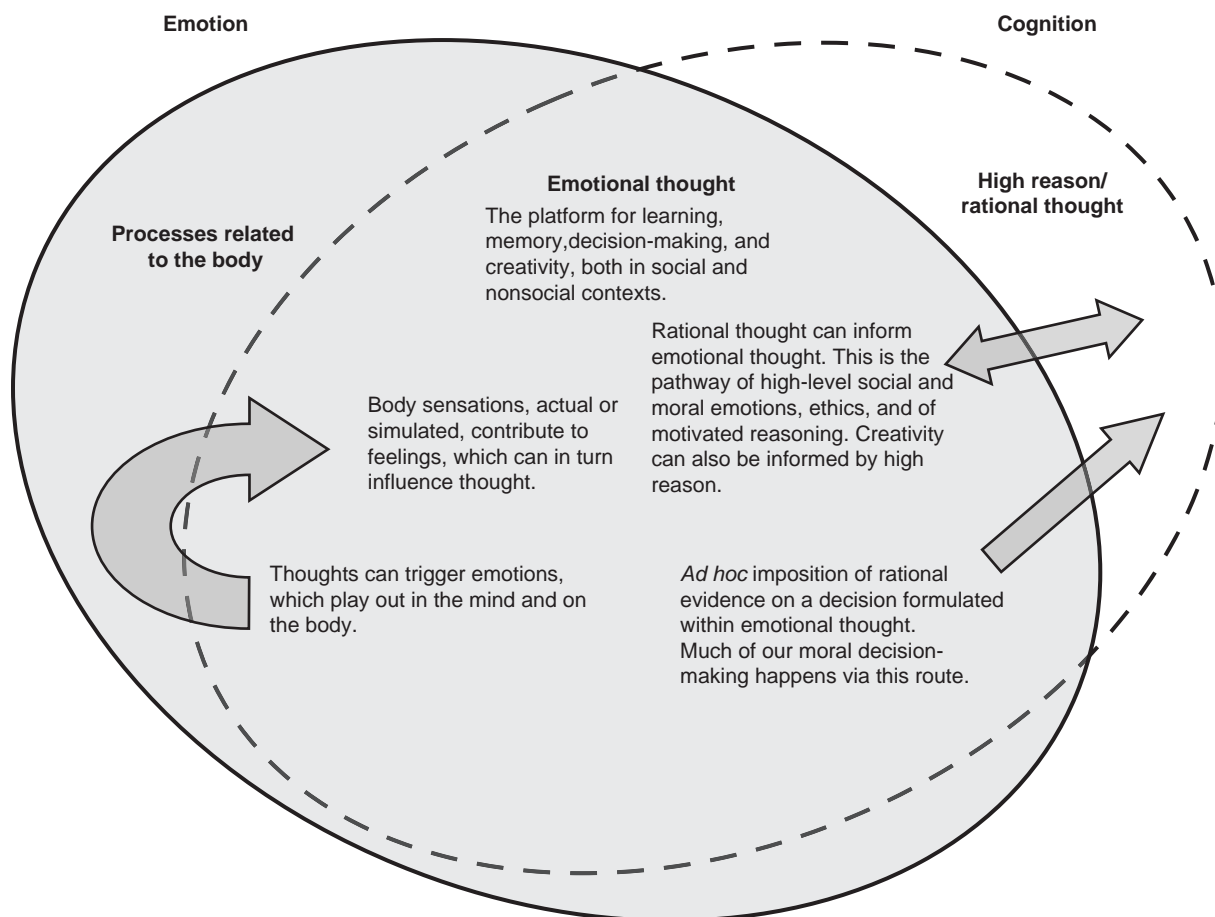


Figure 1 Emotion and cognition come together to produce the thought processes that educators care about, among them learning and memory. In the figure, the solid ellipse represents emotion and the dashed ellipse represents cognition. The extensive overlap between the two ellipses represents the domain of emotional thought. Note that emotional thought reflects a dynamic relationship between the brain and body. Reprinted from Immordino-Yang, M. H. and Damasio, A. R. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain and Education* 1(1), 3–10, with permission from Blackwell publishing.

Memory and Attention

To understand the current state of research on memory and attention, it is helpful to first discuss current views on how reality is constructed in the mind and brain, and the relationship of this process to perception. Work in various areas of neuroscience, for example, in vision or somato-sensory perception and location of the body in space, has shown that unlike the often predominant intuitive view, we humans do not construct reality directly from our perception of the environment, as if we were equipped with some sort of internal video camera. Instead, our prior learning, our neuropsychological predispositions, and the current context heavily influence the reality that we construct and experience. That is, reality is never perceived directly from the environment. Instead, we construct reality based on our own best guesses, interpretations, and expectations. For a trite but illustrative example, imagine why visual illusions work: our visual system

uses context and prior experience with the world to construct images that incorporate our best guesses about the color, form, movement, and identity of what is actually in front of our eyes.

Related to this, our memories do not reflect the objective replaying of an actual occurrence, but our iterative mental reconstruction of an event, fact, or procedure, for example, the skills to solve a math problem, or a student's conversation with her teacher about her test grade. This means that the iterative reconstruction or mental conjuring of a remembered event will be very similar to the neural processes for imagining an event that never happened, or for simulating possible outcomes of future events. Notably, each of these processes is organized by our emotions, and reflects the subjective meaningfulness and relevance of the remembered, imagined, or simulated thought, as well as the social, physical, biological, and developmental contexts in which the person is operating.

Given all these factors, it is no wonder that different teachers and learners perceive, experience, and remember lessons and educational contexts in different ways!

Another process that is related to the study of memory and emotion, and that is an important prerequisite for the recruitment of neural networks, is attention. The last decade marks theoretical and methodological advances in the study of attention and its relationship to the development of academic skills (Corbetta and Shulman, 2002). In particular, Posner and colleagues have distinguished three different attentional networks important for learning, including networks for alerting, orienting, and executive attention (for a review, see Posner and Rothbart, 2007). They have also shown that individual differences in attention networks can be related to genetic and environmental factors, and that training in these aspects of outwardly directed attention, that is, the ability to regulate one's focus on different aspects of the environmental context, can improve preschooler's academic abilities in various areas such as reading skills and social interaction at school (Berger *et al.*, 2007). Future work should investigate how attention monitoring can be taught in schools, as a way to increase the efficiency with which neural networks are built and recruited.

Back to the Big Picture: Mind, Brain, and Education are Becoming Usefully Connected

Over a decade ago, John Bruer cautioned educators that given the current state of knowledge, directly connecting brain science and education was premature – a bridge too far (Bruer, 1997). But, much has happened since then to narrow the chasm between these two sources of knowledge about development and learning. A new field has been established whose aim is to further knowledge about children's learning by bringing together methods and evidence from various fields, among them neuroscience, psychology, cognitive science, and education.

In this stimulating climate, it is important that new neuroscience advances be carefully examined in light of psychological, developmental, and pedagogical theory and research, to ensure that the field proceeds with caution as well as optimism toward educational innovation. In the past, techniques and ideas from so-called brain-based education have led to the formation of neuromyths – oversimplified, misunderstood, or misapplied notions whose integration into educational contexts is unjustified and, in some cases, detrimental or even dangerous. Instead, findings from neuroscience must be carefully implemented and evaluated, starting in educational microcosms such as research schools, where students and faculty partner with cognitive neuroscientists in the design and assessment of research.

In conclusion, it is an exciting time for the field of MBE, and for studying the neuroscientific bases of learning. In the end, learning happens primarily in the brain; studying the neuroscientific bases of learning can therefore provide educationally relevant insights that, with careful implementation and evaluation, may improve schools and other learning environments for the generations to come.

See also: Attention in Cognition and Early Learning; Knowledge Domains and Domain Learning; The Neuroscience of Aging and Cognition; The Neuroscience of Reading.

Bibliography

- Adolphs, R. and Damasio, A. R. (2000). The interaction of affect and cognition: A neurobiological perspective. In Forgas, J. P. (ed.) *Handbook of Affect and Social Cognition*, pp 27–49. Mahwah, NJ: Erlbaum.
- Berger, A., Kofman, O., Livneh, U., and Flenik, A. (2007). Multidisciplinary perspectives on attention and the development of self-regulation. *Progress in Neurobiology* **82**(5), 256–286.
- Bitan, T., Burman, D. D., Chou, T. L., *et al.* (2007). The interaction between orthographic and phonological information in children: An fMRI study. *Human Brain Mapping* **28**(9), 880–891.
- Boets, B., Wouters, J., Wieringen, A. V., Smedt, B. D., and Ghesquière, P. (2008). Modelling relations between sensory processing, speech perception, orthographic and phonological ability, and literacy achievement. *Brain and Language* **106**(1), 29–40.
- Bruer, J. (1997). Education and the brain: A bridge too far. *Educational Researcher* **26**(8), 4–16.
- Cao, F., Peng, D., Liu, L., *et al.* (2009). Developmental differences of neurocognitive networks for phonological and semantic processing in Chinese word reading. *Human Brain Mapping* **30**(3), 797–809.
- Carey, S. and Sarnecka, B. W. (2006). The development of human conceptual representations: A case study. In Munakata, Y. and Johnson, M. H. (eds.) *Processes of Change in Brain and Cognitive Development*, pp 473–498. Oxford: Oxford University Press.
- Corbetta, M. and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Neuroscience Reviews* **3**, 210–215.
- Dehaene, S., Molko, N., Cohen, L., and Wilson, A. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology* **14**, 218–224.
- Fischer, K. W., Bernstein, J. H., and Immordino-Yang, M. H. (eds.) (2007). *Mind, Brain and Education in Reading Disorders*. Cambridge: Cambridge University Press.
- Fischer, K. W. and Bidell, T. (2006). Dynamic development of action and thought. In Damon, W. and Lerner, R. (eds.) *Handbook of Child Psychology: Volume 1. Theoretical Models of Human Development*, 6th edn, pp 313–399. Hoboken, NJ: Wiley.
- Gaab, N., Gabrieli, J. D. E., Deutsch, G. K., Tallal, P., and Temple, E. (2007). Neural correlates of rapid auditory processing are disrupted in children with developmental dyslexia and ameliorated with training: An fMRI study. *Restorative Neurology and Neuroscience* **25**(3–4), 295–310.
- Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain* **119**, 593–609.
- Griffin, S. (2004). Building number sense with number worlds: A mathematics program for young children. *Early Childhood Research Quarterly* **19**(1), 173–180.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* **108**(4), 814–834.
- Herrmann, E., Call, J., Hernandez-Lloreda, M. V., Hare, B., and Tomasello, M. (2007). Humans have evolved specialized skills of

- social cognition: The cultural intelligence hypothesis. *Science* **317**(5843), 1360–1366.
- Immordino-Yang, M. H. (2007). A tale of two cases: Lessons for education from the study of two boys living with half their brains. *Mind, Brain and Education* **1**(2), 66–83.
- Immordino-Yang, M. H. (2008). The smoke around mirror neurons: Goals as sociocultural and emotional organizers of perception and action in learning. *Mind, Brain, and Education* **2**(2), 67–73.
- Immordino-Yang, M. H. and Damasio, A. R. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain and Education* **1**(1), 3–10.
- Katzir, T. and Pare-Blagoev, J. (2006). Applying cognitive neuroscience research to education: The case of literacy. *Educational Psychologist* **41**(1), 53–74.
- Knight, C. and Fischer, K. W. (1992). Learning to read words: Individual differences in developmental sequences. *Journal of Applied Developmental Psychology* **13**, 377–404.
- Le Corre, M., Van de Walle, G., Brannon, E. M., and Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of counting as a representation of the positive integers. *Cognitive Psychology* **52**, 130–169.
- Nelson, C. A., Zeanah, C. H., Fox, N. A., et al. (2007). Cognitive recovery in socially deprived young children: The Bucharest early intervention project. *Science* **318**, 1937–1940.
- Oberman, L. M., Pineda, J. A., and Ramachandran, V. S. (2007). The human mirror neuron system: A link between action observation and social skills. *Social, Cognitive and Affective Neuroscience* **2**(1), 62–66.
- Posner, M. I. and Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology* **58**, 1–23.
- Schneeps, M. H., Rose, T. L., and Fischer, K. W. (2007). Visual learning and the brain: Implications for dyslexia. *Mind, Brain, and Education* **1**(3), 128–139.
- Singer, F. M. (2007). Beyond conceptual change: Using representations to integrate domain-specific structural models in learning mathematics. *Mind, Brain, and Education* **1**(2), 84–97.
- Thatcher, R. W., North, D. M., and Biver, C. J. (2008). Development of cortical connections as measured by EEG coherence and phase delays. *Human Brain Mapping* **29**(12), 1400–1415.
- Umiltà, M. A., Kohler, E., Gallese, V., et al. (2001). I know what you are doing: A neurophysiological study. *Neuron* **31**(1), 155–165.
- Wolf, M. and O'Brien, B. (2006). From the Sumerians to images of the reading brain: Insights for reading theory and intervention. In Rosen, G. (ed.) *The Dyslexic Brain*, pp 5–10. Timonium, MD: York Press.

Further Reading

- Damasio, A. R. (2003). *Looking for Spinoza: Joy, Sorrow and the Feeling Brain*. New York: Harcourt.
- Goswami, U. (2006). Neuroscience and education: From research to practice? *Nature Reviews Neuroscience* **7**(5), 406–411.
- Jossey-Bass Publisher (2008). *The Jossey-Bass Reader on the Brain and Learning*. San Francisco, CA: Wiley.
- National Research Council (1999). *How People Learn: Brain, Mind, Experience, and School*, ch. 5. Washington, DC: National Academy Press.
- Organization for Economic Co-Operation and Development (OECD) (2007). *Understanding the Brain: The Birth of a Learning Science*. Paris: OECD.
- Powell, K. (2006). How does the teenage brain work? *Nature* **442**, 865–867.
- Ramachandran, V. (1998). *Phantoms of the Brain: Probing the Mysteries of the Human Mind*. New York: William Morrow.
- Rose, D. H. and Meyer, A. (eds.) (2006). *A Practical Reader in Universal Design for Learning*. Cambridge, MA: Harvard Education Press.

Relevant Websites

- <http://www.uknow.gse.harvard.edu> – Harvard Graduate School of Education, Usable Knowledge forum.
- <http://www.imbes.org> – International Mind, Brain and Education Society.
- <http://www.oecd.org> – Organization for Economic Cooperation and Development, Centre for Educational Research and Innovation (CERI): Brain and Learning.
- <http://faculty.washington.edu> – UW Faculty, Neuroscience for Kids.

The Neuroscience of Reading

U Goswami, University of Cambridge, Cambridge, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

EEG – Electroencephalography, a time sensitive imaging method based on direct measurement of electrical brain activity.

fMRI – Functional magnetic resonance imaging, a spatial imaging method based on blood flow.

MEG – Magnetoencephalography, a time sensitive imaging method based on the magnetic fields generated by electrical brain activity.

Introduction

The neural demands made by learning to read appear to be very similar in different languages. Current neuroimaging research suggests that the core neural systems involved are the same across languages, and that the atypical neural processing characteristic of children with developmental dyslexia is also highly similar across languages. Current neuroimaging technologies are able to provide information about the time processes in reading (electroencephalography (EEG) technology), information about the parts of the brain that are involved in reading (functional magnetic resonance imaging (fMRI) technology), and information about the temporal sequential activation of different neural regions (magnetic source imaging (MSI)). Connectivity analyses also provide important information on the white matter tracts (information highways in the brain) that develop to support successful word recognition.

The most prevalent neuroimaging technologies used with children work in different ways and provide different kinds of information. In EEG, sensitive electrodes are placed on the child's scalp and brain electrical activation is recorded directly. The sensitive electrodes can measure the low-voltage changes in networks of neurons caused by cells communicating with each other (an electrochemical process). EEG is very time sensitive, and is able to record changes in brain activity at the millisecond level. For example, EEG can pinpoint the exact time at which a child recognizes a word. However, EEG signals are difficult to localize, hence the neural regions responsible for word recognition are not revealed by this technique. A second prevalent measure – currently the most dominant measure in use with children – is fMRI. fMRI

measures changes in blood flow in the brain, through detecting the magnetic resonance signal generated by the protons of water molecules in neural cells. This generates a blood oxygenation level-dependent (BOLD) response. The BOLD response peaks over time, and relative activation in different neural regions will be recorded. Images are typically acquired over 0.5–several seconds, and so fMRI lacks the millisecond temporal resolution of EEG. However, fMRI offers very good spatial resolution, indicating where exactly in the brain reading-related neural activity takes place. Finally, MSI depends on a combination of magnetoencephalography (MEG) and MRI. MEG measures the magnetic fields generated by the electrical activity in the brain rather than the electrical activity itself (the latter is measured by EEG). These magnetic fields are tiny; they are 1 billion times smaller than the magnetic field generated by the electricity in a light bulb. By combining this information with MRI scans – which detect changes in blood flow – both the time course and the spatial localization of brain activity is possible. There are currently very few MSI studies of children's reading.

The Development of Reading

The history of behavioral research on reading acquisition in English reveals considerable debate about the extent to which visual processes versus phonological processes are critical for learning to read. For example, in the 1970s there was debate concerning the relative contribution of Phonician (code-breaking) versus Chinese (visual memorization) reading-acquisition strategies. Developmental dyslexia was assumed to be a primarily visual disorder, conceptualized as a congenital word blindness. More recently, behavioral researchers have proposed stage or phase models of reading acquisition in English, in which early logographic or visual memorization strategies are replaced by code-breaking or phonological recoding strategies, with a final phase of automatic orthographic word recognition. Dual-route models of reading acquisition were also popular. These models assumed that, developmentally, children could choose to learn to read by either whole-word-based strategies (a visual or direct route to meaning) or by using grapheme–phoneme assembly (a phonological or indirect route to meaning). It was also assumed that developmental dyslexia could either reflect visual-orthographic difficulties in developing a sight vocabulary (surface developmental dyslexia), or difficulties

in learning to recode words using letter-sound rules (phonological surface dyslexia). Most of these ideas are looking increasingly dated with the advent of brain imaging.

Brain imaging studies available at the time of writing suggest that reading begins primarily as a phonological process. In the earliest phases of reading acquisition, it is the neural structures for spoken language that are particularly active. As reading expertise develops, an area in the visual cortex originally named the visual word-form area (VWFA) becomes increasingly active (Cohen and Dehaene, 2004). Although described as processing visual forms, this area is not a logographic system that accesses meaning directly from a holistic pattern. Although close to the visual areas that are active during picture naming, the VWFA is also active during nonsense word reading (reading aloud nonlexical forms, such as *tegwump*). Nonsense words do not have visual word forms in the mental lexicon, hence the VWFA is thought to store orthography–phonology connections at different grain sizes. Orthographic sequences and their connections to sound appear to be stored at both the whole-word and sub-word level.

Many behavioral studies in developmental psychology have demonstrated the critical role of phonological awareness in learning to read (see Ziegler and Goswami, 2005, for a recent review). Phonological awareness is thought to develop from natural language-acquisition processes. Word learning is exponential between the ages of 1 and 6 years, and in order for the brain to represent each learned word as a distinct and unique sequence of sounds, the mental lexicon must incorporate knowledge of the sound elements that comprise a particular word, and the order in which these elements occur. Behavioral studies across languages have shown that a child's awareness of these sound elements at different linguistic levels (syllable, onset–rime, and phoneme) predicts the acquisition of reading (see Ziegler and Goswami, 2005, for a review). Further, training phonological awareness has positive effects on reading acquisition across languages, particularly when it is combined with training about how letters or letter sequences correspond to sounds in that language (e.g., Schneider *et al.*, 2000). Children with developmental dyslexia across languages appear to have specific problems in detecting and manipulating component sounds in words, called a phonological deficit (e.g., Snowling, 2000).

Neuroimaging Studies of Learning to Read

There are still relatively few neuroscience studies of children who are learning to read. Most neuroimaging studies of reading have been carried out with adults (see Price and McCrory, 2005, for a recent synthesis). Adult fMRI studies show a very consistent picture concerning the neural networks that underpin skilled reading. Skilled word recognition appears to depend on a left-lateralized

network of frontal, temporoparietal, and occipitotemporal regions – whatever language is being read. However, there is some additional recruitment of visuospatial areas for languages with nonalphabetic orthographies (e.g., left middle frontal gyrus for Chinese, see meta-analysis by Tan *et al.* (2005)). The frontal, temporoparietal, and occipitotemporal regions essentially comprise the language, cross-modal, and visual areas of the brain. At a very simple level, semantic and memory processing is thought to occur in temporal and frontal areas, auditory processing in temporal areas, visual processing in occipital areas, and cross-modal processing in parietal areas. Data on neural timing from EEG studies suggest that the brain has decided whether it is reading a real word or a nonsense word within 160–180 ms of presentation. This has been demonstrated both for children and adults across languages (e.g., Csepe and Szucs, 2003; Suaseng *et al.* 2004).

Studies of children using fMRI have tended to use a restricted range of tasks. These include asking participants to read single words and then comparing brain activation to a resting condition with the eyes closed; asking participants to pick out target visual features while reading print or false font (false font is made up of meaningless symbols matched to letters for visual features like the ascenders in the letters b, d, and k); making phonological judgments while reading words or nonsense words (e.g., “do these items rhyme?”: *leat* and *jete*), and making lexical decisions (e.g., pressing a button when a word is presented, and a different button when a nonsense word is presented). Nevertheless, the developmental studies that have been completed demonstrate a high degree of consistency between the neural networks recruited by novice and expert readers. For example, work by Turkeltaub and colleagues used fMRI and the false-font task to compare neural activation in English-speaking children and college students aged from 7 to 22 years (Turkeltaub *et al.*, 2003). This task was selected because 7-year-olds can perform it as well as adults, meaning that changes in reading-related neural activity are likely to reflect developmental differences rather than differences in task expertise. Turkeltaub and colleagues reported that adults activated the usual left hemisphere sites, including left posterior temporal and left inferior frontal cortex. They then restricted the analyses to children younger than 9 years of age. In this instance, the main area engaged was the left posterior superior temporal cortex. This region is traditionally considered the focus of phonological activity, and is thus thought to be active during grapheme–phoneme translation. As reading developed, activity in left temporal and frontal areas increased, while activity previously observed in right posterior areas declined. This pattern was interpreted as showing that reading-related activity in the brain becomes more left-lateralized with development.

In further analyses focusing only on the younger children, the researchers investigated the relationships

between three core phonological skills and word processing. The three core phonological skills are usually taken to be phonological awareness, phonological memory, and rapid automatized naming (RAN). Turkeltaub and colleagues calculated partial correlations between activated brain regions and each of these three measures while controlling for the effects of the other two measures. They reported that the three different measures correlated with three distinct patterns of brain activity. Brain activity during phonological awareness tasks appeared to depend on a network of areas in the left posterior superior temporal cortex and the inferior frontal gyrus. The level of the children's phonological skills modulated the amount of activity in this network. Activity in the inferior frontal gyrus increased with reading ability. This area is also a key phonological area (Broca's area) – important for the motor production of speech. Phonological short-term memory (digit span) appeared to depend on a different neural network, including left intraparietal sulcus – the dominant site of working memory in adults – the middle frontal gyri (bilaterally), and right superior temporal sulcus. RAN appeared to depend on a different, bilateral network including right posterior superior temporal, right middle temporal, and left ventral inferior frontal gyri. Other studies report increasing engagement of the VWFA as reading ability increases. This suggests that the VWFA is a kind of skill zone – with greater activation reflecting increasing expertise with orthography–phonology connections (see Pugh, 2006, for an overview).

A similar developmental picture emerges from an fMRI study of 119 typically developing readers in the age range 7–17 years by Shaywitz and colleagues (Shaywitz *et al.*, 2007). This study used a rhyme-decision task (e.g., “do these items rhyme?": leat and kete), and a visual line orientation task (e.g., “Do [V] and [V] match?"). Shaywitz and his colleagues reported that networks in both left and right superior and middle frontal regions were more active in younger readers. Activity in these networks declined as reading developed. In contrast, activity in the left anterior lateral occipitotemporal region increased as reading developed. This region includes the putative VWFA. Hence both Turkeltaub *et al.* (2003) and Shaywitz *et al.* (2007) report decreased right hemisphere involvement as reading develops. The difference in the behavioral tasks used (e.g., false font vs. rhyme judgment) may explain why somewhat different neural networks became less active.

Overall, therefore, current neuroimaging data appear to support a single route model of reading development. Neuroimaging data suggest that phonological recoding to sound rather than logographic recognition is the key early reading strategy, accompanied by an incremental process of developing orthographic–phonological connections at different grain sizes in the VWFA. This reflects the development of an orthographic lexicon containing both whole words and fragments of familiar words

such as orthographic rimes (Pugh, 2006). The VWFA is not a logographic or visual lexicon, able to support Chinese processing or a direct route from printed word to meaning postulated by dual-route theory.

Neuroimaging Studies of Developmental Dyslexia

The Networks Recruited for Reading

Neuroimaging studies of adult readers with developmental dyslexia suggest biological unity with regard to the affected neural networks. These studies report atypical activation in the three important neural sites for reading – namely, the left posterior temporal regions, the left inferior frontal regions, and the left occipitotemporal regions (such as the VWFA) (e.g., Paulesu *et al.* (2001), for a comparison of Italian, French, and English-speaking adult dyslexics). Further, neuroimaging studies carried out with children with developmental dyslexia report a similar pattern to adult studies (e.g., Shaywitz *et al.*, 2002, 2007; Simos *et al.*, 2000). However, when interpreting neural differences in developmental studies, it is important to ensure that behavioral performance in the tasks being used is equivalent. If children with dyslexia are worse in such tasks than control children, then any differences in neural activity could reflect differing levels of expertise rather than the neural differences that are core to being dyslexic.

In one of the first developmental studies, Shaywitz *et al.* (2002) studied 70 children with dyslexia, mean age of 13 years, and compared them to 74 11-year-old typically developing controls (although the controls were not matched for reading level). The children lay in the fMRI scanner while performing a variety of reading-related tasks – namely, letter identification (e.g., Are t and V the same letter?), single-letter rhyme (e.g., Do V and C rhyme?), nonword rhyming (e.g., Do 'leat' and 'jete' rhyme?), and reading for meaning (e.g., Are 'corn' and 'rice' in the same semantic category?). Brain activity in each condition was contrasted with activity in a baseline condition; the line-orientation task (e.g., do [V] and [V] match?). Shaywitz *et al.* (2002) reported that the children with developmental dyslexia showed underactivation in the core left temporoparietal networks, with older dyslexics showing overactivation in right inferior frontal gyrus. The children with developmental dyslexia also showed increased activation in right temporoparietal networks. However, some of the differences found in brain activation could have reflected differing levels of expertise with the tasks. For example, for the nonword-rhyming measure, the controls (79% correct) were significantly better at the task than the children with dyslexia (59% correct). In a subsequent study of an expanded sample, Shaywitz *et al.* (2007) used in-magnet nonword-reading

ability as a covariate to control for this problem. Shaywitz and colleagues compared 113 dyslexic children in the age range of 7–18 years to the 119 typically developing readers discussed above in the nonword rhyme and visual line-orientation tasks. Compared to the typically developing children, the dyslexic children showed no age-related increase in the activity of the VWFA. Instead, activity both in the left inferior frontal gyrus (speech articulation) and the left posterior medial occipitotemporal system increased, and reading did not become left-lateralized, with continued right hemisphere involvement.

Developmental Differences in the Time Course of Neural Activation

If basic word-recognition processes are delayed in developmental dyslexia, this will delay access to semantics and, therefore, affect reading comprehension. Similarly, cognitive processes such as grapheme–phoneme conversion might take longer in developmental dyslexia. EEG and MSI technologies can help us to study these questions.

One of very few longitudinal neuroimaging studies of children learning to read used MSI to gain information about developmental differences in the time course of neural activation. Simos and his colleagues studied 33 English-speaking children – 16 of whom were thought to be at high risk of developing dyslexia. The researchers compared brain activation in a letter-sound task (the child saw a letter and had to provide its sound) and a simple nonword-reading task (recoding nonwords like ‘lan’ to sound). Both tasks were administered at the end of kindergarten and again at the end of the first grade (see Simos *et al.*, 2005). In kindergarten, the high-risk group were significantly slower to show neural activity in response to both letters and nonwords in the occipitotemporal region (requiring, on average, 320 ms compared to 210 ms for those not at risk). The high-risk group also showed atypical activation in the left inferior frontal gyrus when performing the letter-sound task. For this task, the onset of activity actually increased developmentally – from 603 ms in kindergarten to 786 ms in the first grade. The typically developing readers did not show a processing-time increase.

When Simos and his colleagues compared the onset of activity of the three core neural networks for reading, they found that low-risk children showed early activity in the left occipitotemporal regions. This was followed by activity in the temporoparietal regions, predominantly in the left hemisphere, and then by bilateral activity in the inferior frontal regions. In contrast, high-risk children showed little differentiation in terms of the time course of activation between the occipitotemporal and temporoparietal regions. High-risk children who were also nonresponsive to a phonological remediation package being administered during the study ($N = 3$) were distinct in showing earlier onset of

activity in the inferior frontal gyrus compared to the temporoparietal regions. Simos and colleagues commented that the increased inferior frontal activation probably reflected the role of compensatory articulatory processes. This may indicate that children with phonological difficulties rely more heavily on networks for articulation when phonological processing is required.

The Neural Effects of Remediation

Neuroimaging studies of different types of remediation for reading difficulties are consistent in showing that when targeted phonology-based interventions are used with affected children, neural activity is normalized. This indicates that levels of activation in the left-lateralized network of reading areas typically improve following such interventions (e.g., Shaywitz and Shaywitz, 2005; Simos *et al.*, 2002; Temple *et al.*, 2003). Most such studies rely on fMRI, and hence it is difficult to be sure how extensive such normalization really is, given that the BOLD signal reaches a peak over several seconds. Simos and his research group, therefore, used MSI to explore neural activation in eight children with developmental dyslexia who had received 80 h of intensive training with a phonological remediation package (Simos *et al.*, 2002). MSI scans were taken during a nonword rhyme-matching task (e.g., ‘yoat’ ‘wote’), both before the intervention and following remediation. Prior to the intervention, the dyslexic children showed the usual hypoactivation of left temporoparietal regions. Following the intervention, all eight children showed a dramatic increase in the activation of left temporoparietal regions, predominantly in the left posterior superior temporal gyrus. As will be recalled, these networks support grapheme–phoneme recoding in typically developing readers (see Turkeltaub *et al.*, 2003). However, Simos and colleagues found that neural activity was delayed in the children with dyslexia relative to the controls even after this remediation. For example, the peak in left superior temporal gyrus activity occurred at 837 ms, on average, for the dyslexic children, and at 600 ms for the controls. Therefore, even with intensive remediation, children with dyslexia are slow to achieve the reading fluency shown by nondyslexic children.

Shaywitz and Shaywitz (2005) used retrospective examination of the large sample of children with developmental dyslexia reported in Shaywitz *et al.* (2002) to compare whether children at risk for reading difficulties showed different neurodevelopmental trajectories. They distinguished three groups within the sample – a group of persistently poor readers (PPR), who had met criteria for poor reading both in the second/third and in the ninth/tenth grades; a group of accuracy-improved poor readers (AIR), who had met criteria for poor reading in the second/third grades but who did not meet criteria in the ninth/tenth grades; and a control group of nonimpaired readers

(C), who had never met criteria for poor reading (the participants had been studied since age 5). Shaywitz and Shaywitz reported that both the PPR and the AIR groups showed hypoactivation in neural networks in left superior temporal and occipitotemporal regions. However, the groups were distinguished by their neural activity when reading real words. The AIR group still demonstrated underactivation in the usual left posterior areas for real-word reading, whereas the PPR group activated the left posterior regions to the same extent as controls.

Shaywitz and Shaywitz then carried out further analyses based on connectivity. Connectivity analyses examine the neural areas that are functionally connected to each other during reading. The connectivity analyses suggested that reading achievement depended on memory for the PPR group, and not on the normalized functioning of the left posterior regions. The unimpaired controls demonstrated functional connectivity between left hemisphere posterior and anterior reading systems, but the PPR group demonstrated functional connectivity between left hemisphere posterior regions and right prefrontal areas associated with working memory and memory retrieval. Prospective longitudinal studies comparing patterns of neural activation and connectivity in dyslexic children as high-frequency words become over-learned would clearly be very valuable.

Conclusions

Neuroscience studies using fMRI have provided a clear picture of the neural networks that underpin reading both in typically developing and dyslexic readers. It has been shown that the functional organization of the networks for reading is similar in typical development and in dyslexia. Rather than relying on different neural networks, dyslexics show hypoactivation of crucial parts of the network of areas involved in word recognition, and an atypical pattern of continuing right hemisphere involvement. These fMRI studies are essentially correlational studies. They do not answer research questions about what goes wrong in the dyslexic brain. Neuroimaging methods that provide data on the time course of neural processing, such as MEG (MSI) and EEG, are beginning to answer such causal questions. For example, studies using MSI demonstrated that neural activation is delayed in core components of the left-lateralized reading network, and suggested that core components of the reading network may be activated in a different order in developmental dyslexia. We need more longitudinal studies to find out whether there are different neurodevelopmental routes to word recognition for dyslexic children compared to typically developing children. Only longitudinal studies can illustrate the response of a dyslexic brain to being trained to learn to read.

It is believed that the most informative studies with respect to causation in reading will be longitudinal

prospective studies that use neuroscience techniques to study basic sensory processing in at-risk children. Such studies may give insight into the causes of the phonological deficit. In this regard, the most promising studies to date are those investigating basic auditory processing, using EEG and MEG methodologies. For example, a large-scale Finnish study (the Jyväskylä Longitudinal Study of Dyslexia or JLD; see Lyytinen *et al.*, 2004) has followed babies at familial risk for dyslexia since birth. EEG measures of auditory sensory processing (evoked-response potentials to speech and nonspeech cues) have been found to distinguish the at-risk babies from controls even during infancy. Hence EEG technologies may soon be able to offer robust neural markers of risk for reading difficulties, enabling much earlier intervention.

See also: Learning to Read; Neuroscience Bases of Learning; Reading Comprehension: Reading For Learning.

Bibliography

- Cohen, L. and Dehaene, S. (2004). Specialization within the ventral stream: The case for the visual word form area. *NeuroImage* **22**, 466–476.
- Csepe, V. and Szucs, D. (2003). Number word reading as a challenging task in dyslexia? An ERP study. *International Journal of Psychophysiology* **51**, 69–83.
- Lyytinen, H., Ahonen, T., Guttorm, T., *et al.* (2004). Early development of children at familial risk for dyslexia: Follow-up from birth to school age. *Dyslexia* **10**, 146–178.
- Paulesu, E., Démonet, J. F., Fazio, F., *et al.* (2001). Dyslexia: Cultural diversity and biological unity. *Science* **291**(5511), 2165–2167.
- Price, C. J. and McCrory, E. (2005). Functional brain imaging studies of skilled reading and developmental dyslexia. In Snowling, M. J. and Hulme, C. (eds.) *The Science of Reading: A Handbook*, pp 473–496. Oxford: Blackwell Publishing.
- Pugh, K. (2006). A neurocognitive overview of reading acquisition and dyslexia across languages. *Developmental Science* **9**, 448–450.
- Sauseng, P., Bergmann, J., and Wimmer, H. (2004). When does the brain register deviances from standard word spellings? An ERP study. *Cognitive Brain Research* **20**, 529–532.
- Schneider, W., Roth, E., and Ennemoser, M. (2000). Training phonological skills and letter knowledge in children at-risk for dyslexia: A comparison of three kindergarten intervention programs. *Journal of Educational Psychology* **92**, 284–295.
- Shaywitz, S. E. and Shaywitz, B. A. (2005). Dyslexia (specific reading disability). *Biological Psychiatry* **57**, 1301–1309.
- Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., *et al.* (2002). Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biological Psychiatry* **52**(2), 101–110.
- Shaywitz, B. A., Skudlarski, P., Holahan, J. M., *et al.* (2007). Age-related changes in reading systems of dyslexic children. *Annals of Neurology* **61**, 363–370.
- Simos, P. G., Breier, J. I., Fletcher, J. M., Bergman, E., and Papanicolaou, A. C. (2000). Cerebral mechanisms involved in word reading in dyslexic children: A magnetic source imaging approach. *Cerebral Cortex* **10**, 809–816.
- Simos, P. G., Fletcher, J. M., Bergman, E., *et al.* (2002). Dyslexia-specific brain activation profile becomes normal following successful remedial training. *Neurology* **58**, 1203–1213.
- Simos, P. G., Fletcher, J. M., Sarkari, S., *et al.* (2005). Early development of neurophysiological processes involved in normal reading and reading disability: A magnetic source imaging study. *Neuropsychology* **19**(6), 787–798.

- Snowling, M. J. (2000). *Dyslexia*. Oxford: Blackwell.
- Tan, L. H., Laird, A. R., Li, K., and Fox, P. T. (2005). Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: A meta-analysis. *Human Brain Mapping* **25**(1), 83–91.
- Temple, E., Deutsch, G. K., Poldrack, R. A., *et al.* (2003). Neural deficits in children with dyslexia ameliorated by behavioral remediation: Evidence from functional MRI. *Proceedings of the National Academy of Sciences* **100**, 2860–2865.
- Turkeltaub, P. E., Gareau, L., Flowers, D. L., Zeffiro, T. A., and Eden, G. F. (2003). Development of neural mechanisms for reading. *Nature Neuroscience* **6**(6), 767–773.
- Ziegler, J. C. and Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin* **131**(1), 3–29.

LEARNING AND COGNITION – ISSUES, CONCEPTS, TYPES – FOCUS ON LEARNING

Contents

Learning Strategies

Learning Through Play

Situated View of Learning

The Neuroscience of Aging and Cognition

Learning Strategies

C E Weinstein and J Jung, The University of Texas at Austin, Austin, TX, USA

T W Acee, Texas State University – San Marcos, San Marcos, TX, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Comprehension monitoring – This occurs when learners check their understanding of something they are trying to learn using some form of self-assessment.

Domain-dependent learning strategies – The learning strategies that are specific to a particular content area or type of academic task.

Domain-independent learning strategies – The learning strategies that are widely applicable across content areas and academic tasks.

Elaboration strategies – The deep-level strategies used to build bridges or meaningful connections between what the learner is trying to learn and their prior knowledge, experience, attitudes, and/or beliefs.

Learning strategies – Any behavioral, cognitive, metacognitive, motivational, or affective process or action that facilitates understanding, learning, and meaningful encoding into memory.

Organization strategies – The deep-level strategies used to meaningfully translate or transform information into a configuration or form that creates some sort of coherent scheme that makes the information more understandable and easier to encode and remember.

Rehearsal strategies – The surface-level repetition strategies used to memorize discrete information or hold information in working memory so that it can be processed further.

Self-regulation learning strategies – The strategies used by learners to control, manage, and

oversee cognitive, motivational, emotional, and environmental factors that influence learning.

Our desire to understand our world and learn what we need to know and do to survive and thrive has existed since the origin of mankind on the earth. However, it is only since the 1970s that the psychological study of the strategies we use to learn about our world began in earnest. In the broadest sense, a learning strategy is any behavioral, cognitive, metacognitive, motivational, or affective process or action that facilitates understanding, learning, and meaningful encoding into memory. The problem with such a broad definition is that it includes almost any psychological variable. Everything affects learning to at least some extent. For this reason, educational psychologists study variables that have been found to have the greatest direct or indirect impact on learning and encoding into memory, and that can be modified by some type of training or educational intervention. This second criterion excludes some variables such as personality traits because, although they may affect learning, they are not amenable to change by an educational intervention (however, we can teach students learning strategies that reduce the negative impact of some individual variables such as impulsivity or attention deficit disorders).

The major categories of learning strategies include: cognitive information acquisition and processing strategies, motivational strategies, self-regulation and monitoring strategies, affect regulation strategies, and behavioral strategies. Several of these strategies are discussed further after the presentation of a brief historical overview of this area. This is followed by a discussion of the ways in which models of

strategic and self-regulated learning incorporate learning strategies, how learning strategies are assessed, and types of instruction used to teach learning strategies.

Historical Overview

From the time humans first appeared on the earth, they were concerned about learning and memory. Some anthropologists have hypothesized that cave drawings were at least in part an attempt to learn and remember animal migrations and the seasons for different foods. They may also have been a way to help the young learn the knowledge, culture, and norms of life as it then existed. The Greeks are famous for the development of mnemonic devices, or mental memory aids, some of which are still used today (e.g., using the ABC song in the United States to learn the alphabet, associating a new item we are trying to remember with an existing image, or the first letter mnemonic HOMES (Huron, Ontario, Michigan, Erie, and Superior) to remember the five Great Lakes in America). In the 1800s and early 1900s many parlor entertainers at parties or gatherings would use mnemonics to perform prodigious feats of learning and memory. Today, advertising experts use mnemonic tunes and phrases to help people remember to buy their brand when they see it in a store.

For the first half of the twentieth century the study of mnemonics or any other cognitive learning strategy lay dormant as the overwhelming charge of behaviorism swept through psychology. Behaviorists, in their attempt to make psychology a science, rejected not only philosophical or introspective methods of research but also any theoretical or empirical work involving mental processing. The world of environmental manipulations, reinforcement, and extinction had little interest in the black box (mind).

It was not until the late 1960s and 1970s that the study of mental processes made its way back into psychological theories and research. Once again, mnemonics played a key role. Some of the earliest studies during this period focused on mnemonics but this time, once it was established that they were not just parlor tricks, the focus shifted to why and how mnemonics worked and what we could learn about human information processing from them. What cognitive processes were involved in the creation and use of mnemonics? What fundamental information did this give us about human information processing? In the 1970s, this work was greatly enriched by other early researchers in the area of what we now call cognitive educational psychology. For example, a major influence was the work of John Flavell and his colleagues who developed the concept of metacognition, or thinking about our thinking. Basically, metacognition is thinking that focuses on knowledge, self-reflection, and analyses of

how we think and learn. As a direct result of this early work examining cognition, variables that might influence cognition and the degree to which cognition might or might not be influenced through some form of educational intervention, the area of learning strategies developed as a field of study.

Types of Learning Strategies

There is general agreement about the basic types of learning strategies as originally codified by Weinstein and Mayer (1986) and expanded upon by continuing theoretical and research work in this area. The development of interactive models of strategic and self-regulated learning at all educational levels has, however, changed the ways we explore, study and teach students how to use learning strategies. This is discussed further later in this article.

Rehearsal Strategies

These are the most basic of all learning strategies and can be used most effectively in the first stages of building a knowledge base in a content domain. Rehearsal strategies are most useful for basic memorization, processing new information at a surface level, but are not generally useful for creating a deep and sophisticated understanding of the material. Researchers differentiate between passive rehearsal, which is not very effective for most learning tasks, and active rehearsal, which can be more beneficial. Passive rehearsal is based on early mental muscle models of learning which assumed that the mind is like a muscle – the more you practice, the more you will build up the strength of your memory. It involves learners repeating something over and over until they remember it (similar to the use of flash cards) or using other memory aids, such as mnemonics. While these methods may be useful for learning discrete bits of information (e.g., isolated facts and lists of information), they are not very effective as learning strategies for more complex content or learning tasks involving reasoning. In addition, these methods do not contribute efficiently to the integration of new information one is trying to learn with existing knowledge and expertise, a major goal for many learning tasks.

Active rehearsal, while using some of the same methods as passive rehearsal, differs in terms of the goals for using this strategy. In passive rehearsal, repetition is the end point of the process, while in active rehearsal, repetition is used as an enabling tool to hold onto information so that it can be further processed and encoded into more stable areas of memory. Highlighting important information in class notes or a textbook and then reviewing the highlighted material at a future time would be examples of using repetition to help learn the content by creating

additional opportunities to further process it. Even something like flash cards could be used for active repetition, if the goal was to continue thinking about and encoding the information on the cards.

Elaboration Strategies

Elaboration strategies are the largest and most diverse category of learning strategies. Fundamentally, elaboration involves building bridges or connections between what the learner is trying to learn and their prior knowledge, experience, attitudes, and beliefs. Building these bridges forces the learner to actively process the new information and it is this engagement that is believed to be the core cognitive mechanism involved in reaching learning goals. It is not just the elaborations that result from using these strategies but the process of creating those elaborations that facilitates meaningful encoding into memory. In addition, elaboration has been found to increase related variables such as task focus and concentration, task interest and enjoyment, motivation, and positive attitudes and emotions toward the learning content and task itself.

Elaboration learning strategies take many forms depending on the nature of the content, the learning task, and the learner's individual differences and learning goals. The most basic forms of elaboration involve paraphrasing and summarizing. Even though these are fundamentally a form of repetition, unlike rehearsal strategies, they are not simple verbatim recall which does not involve higher-level cognitive processes. To paraphrase or summarize what a student is trying to learn requires some degree of encoding and transformation of the targeted information so active cognitive engagement is required. More advanced and complex forms of elaboration include: using everyday experience to try to understand a new concept, applying what the student is learning to new and diverse tasks, trying to teach the material to someone else, perspective taking, visualization, using a problem-solving strategy in a new context, creating analogies, using compare-and-contrast methods to highlight the differences and similarities between two related concepts, and creating and responding to questions about the material being studied.

Research has shown that the active processing involved in using elaboration strategies is what is key to learning and that the specific elaboration strategy used for a learning task is not as important. However, it has been shown that it is important for students to learn a repertoire of diverse strategies so that they can both develop their preferences and have alternative strategies to fall back on if their preferences do not work for a particular learning task. This issue also relates to what strategies students should learn and in what context they should learn them. The instructional issues this raises are discussed further in later sections.

Organization Strategies

Organization strategies involve translating or transforming information into another configuration and creating some sort of scheme to provide structure to this new way of characterizing or representing the information. These strategies are used to organize information into meaningful categories, hierarchies, and sequential structures so that it can be visualized, analyzed, understood, and encoded meaningfully into memory. Similar to elaboration strategies, organization strategies facilitate deep processing of the information and meaningful encoding into memory. By organizing the new information the learner is actively engaged with the material and it is believed that this active engagement underlies much of the benefit of using organization strategies. In addition, the product that results can be used in the future to review and deepen understanding. Moreover, similar to elaboration, organization strategies have been found to increase related variables such as task focus and concentration, task interest and enjoyment, motivation, and positive attitudes and emotions toward the learning content and task itself.

Creating outlines, concept maps, and concept matrices are types of organizational strategies. Creating an outline involves organizing material into a hierarchical structure with a logical flow using an outline format (e.g., I. Thesis statement; A. Major point; 1. Supporting detail). Creating concept maps is another organizational strategy that is used to graphically represent relationships among and between concepts. For example, connecting concepts with arrowed lines and identifying those relationships with phrases such as: results in, contributes to, decreases, is a defining attribute of, or is a subcategory of is a common type of organization strategy. Creating concept matrices refers to graphically organizing information about related concepts into a matrix of rows and columns in order to learn and analyze those concepts (e.g., organizing the names of concepts in the first column, concept definitions in the second column, and examples of the concepts in the third column). Organizational strategies have been found to help students analyze, learn, and remember their course material at a deep level. Research suggests that the effectiveness of organization strategies derives not only from the active processing that is required to create the organizational structure but also from the product itself since it can be later used for review, as a study aid, or incorporated into a larger and more encompassing scheme in that knowledge domain.

Self-Regulation Learning Strategies

Self-regulation learning strategies are used by students to control, manage, and oversee cognitive, motivational,

emotional, and environmental factors that influence learning (both positively and negatively). Goal-setting/planning, implementing/monitoring, and evaluating both process and outcomes are types of self-regulation strategies. Goal-setting/planning refers to setting learning goals and planning for how one will reach those goals (e.g., choosing one rehearsal strategy, two elaboration strategies, and one organizational strategy that will be used to reach the learning objectives for an upcoming exam). Implementing/monitoring involves implementing these learning plans and monitoring the pursuit of the learning goals (e.g., implementing and monitoring a plan to summarize each section of a textbook). Evaluating the success of one's strategic approaches to achieve a learning goal is another type of self-regulation strategy (e.g., evaluating whether or not the cost in time of creating a concept map was worth the payoff in learning). Self-regulation learning strategies can help students fine-tune their strategic approaches to reaching their learning goals and develop more effective and efficient study routines. These strategies can also be used to oversee and manage the regulation of motivational, emotional, and environmental variables that influence learning.

Metacognitive Strategies – Comprehension Monitoring

An important component of self-regulation involves using metacognitive strategies for learning. Within this broad area, the most relevant type of metacognitive strategy for the purposes of this article is comprehension-monitoring strategies. Basically, comprehension monitoring involves checking our understanding of something we are trying to learn using some form of reviewing or self-testing. Comprehension strategies both support and contribute to meaningful learning. Without them, learning could be incomplete or errors might persist undetected. Reviewing and self-testing also contribute to knowledge consolidation and integration across topics. Using mental reviews, going over notes and course materials, thinking of potential questions to guide reading or help prepare for an exam, trying to use new information in novel ways, and trying to apply a principle or method are all important methods for checking understanding, consolidating new knowledge, and integrating related information (both from what is being learned and from what is already known). Although comprehension-monitoring strategies include many of the strategies discussed under elaboration, the purpose is different. When learning new content material, using a strategy like applying a principle is designed to enhance understanding and encoding into memory. When applying a principle such as a comprehension-monitoring strategy, the goal is to see if that understanding and accessible memory encoding has been established.

Need for a Learning Strategies Repertoire

Rehearsal, elaboration, and organization strategies can be used to help students actively process and learn new information. However, students differ over which learning strategies work best for them both within and across different types of learning tasks and contexts. For this reason, it is important that students learn and develop a repertoire of learning strategies both within and across all three of these categories so that they can mindfully develop their preferences and have a range of strategies to fall back on if their preferences do not work.

Domain-Independent Strategies Versus Domain-Dependent Strategies

In current theory, research, and practice, the applicability, or generalizability, of particular learning strategies to different learning content areas or tasks is still being debated. The general issue is whether it is best for students to learn domain (content or task)-specific strategies (e.g., strategies for solving a particular type of physics problem or learning a new vocabulary term in a foreign language) or more generalizable, or domain (content or task)-independent strategies that can be applied to many content areas (e.g., how to approach an unfamiliar textbook or using self-testing to check your understanding of what you are learning). In fact, if you think about it in terms of a generalization gradient, they are really just different points on the line. If the strategy has a narrower domain of applicability (i.e., it can only be used for a relatively small number of learning or performance activities), then it is domain dependent. If, on the other hand, it can be used in a wide variety of situations or content areas, then it has a wide domain of applicability and is domain independent. Like many controversies, it appears that it takes a bit of both to help students become self-regulating, strategic learners. Some strategies may be more effective and efficient for the content and tasks in one particular academic area, while others may be helpful for a wider variety of academic areas and tasks.

Learning Strategies in Models of Strategic and Self-Regulated Learning and College Readiness

The study of learning strategies has evolved from an isolated area of study into a critical component of most models of strategic and self-regulated learning. This integration of learning strategies into more complex and interactive models of academic learning is exemplified

by the work of Pintrich (2004), Weinstein *et al.* (2004), and Zimmerman and Schunk (2001). As an example of these theoretical models that guide much of the current research and instructional development designed to help students become more strategic and self-regulated learners, the most recent version of the Model of Strategic Learning (MSL) developed by Weinstein *et al.* (2006) is briefly discussed (see **Figure 1**).

Model of Strategic Learning

Similar to other recent models, the MSL is a multidimensional, interactive model where the focus is on the interactions among the components rather than the simple effects of one or two elements. Even many of the recent advances in statistical analyses are based on the need for analyzing interactive models. The core of the model (within the triangle) is the learner and all of the individual differences, self-system variables, and long-term goals learners bring to any learning event. In addition, most meaningful learning is goal-driven and the model will be

used in different ways by students with different learning tasks and goals.

The variables outside the rectangle (e.g., requirements of the task and instructor expectations) are important variables for completing a learning activity but they are not usually under the student's direct control. The three main components of the model are listed at the three points of the triangle: skill, will, and self-regulation. Skill focuses on declarative (knowing what to do), procedural (knowing how to do it), and conditional (knowing when to use the strategies) knowledge that students need to develop in order to become more strategic learners (e.g., not only knowing about different learning strategies but also knowing how to use them effectively and when it is most appropriate to use a particular strategy). Will focuses on attitudes, beliefs, and goals that help students thrive and persist when faced with roadblocks to learning (e.g., setting specific and challenging, yet realistic, learning goals and avoiding or coping with self-sabotaging beliefs and attitudes). The self-regulation component focuses on managing the learning process and one's own cognition, motivation, and emotion related to the task (e.g., time

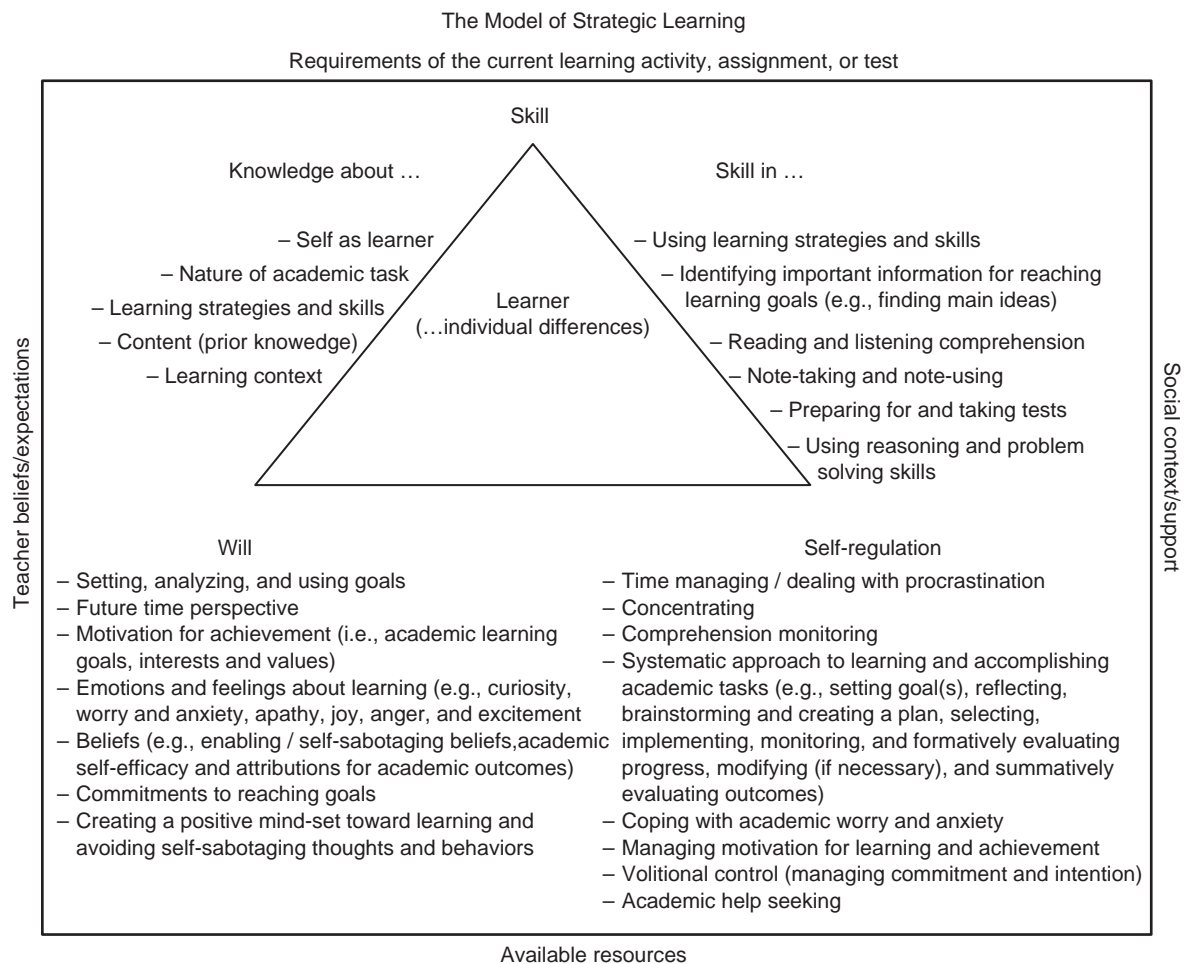


Figure 1 The Model of Strategic Learning.

management, comprehension monitoring, and coping with anxiety). For any given learning task, the student must take into account variables from each of the three components. Like the systems of the body, variables related to skill, will, and self-regulation need to function together in order to facilitate strategic learning. Which elements of each component will be most important to think about or use for completing a task will be largely determined by the learner's goals, prior knowledge, past experiences, and the nature of the task.

Assessment of Student Learning Strategies

Purposes for Assessment

The primary purpose of most learning strategy instruments is to identify students' knowledge and use of learning strategies in order to:

1. investigate correlational and causative relations with other variables, such as motivational goal types, class participation, or academic performance;
2. identify students' strengths and weaknesses in different areas of learning strategies to identify students who might need additional instruction in the areas where they scored low;
3. use as a pre-post and/or delayed outcome measure for interventions designed to teach learning strategies;
4. provide information to educators about individual needs of their students so they can target part of their instruction to helping their students learn or enhance their use of learning strategies; and
5. help developmental educators, student affairs counselors, and advisors who work with at-risk students to identify students who may be at risk of failure or dropping out from higher education settings because of their lack of knowledge and use of learning strategies so that they can be placed in learning strategies courses or other types of interventions.

Approaches and Instruments Used to Assess Students' Learning Strategies Knowledge and Use

There are a number of experimental and published instruments that assess students' knowledge and use of learning strategies. These instruments use some type of self-report method, usually involving some type of instrument or questionnaire. Some studies do have students describe or explain their strategies as they are using them, or just after completing a learning task, but these assessment methods still rely on self-report. While the limitations of self-report methods have been repeatedly

documented, it remains the best method for providing a window on the mind.

Some of the more commonly used measures include: The Learning and Study Strategies Inventory (2nd edition) by Weinstein *et al.* (2002), the Learning and Study Strategies Inventory – High School Version (2nd edition) by Weinstein and Palmer (1990), the Motivated Strategies for Learning Questionnaire by Pintrich *et al.* (1991), the Approaches and Study Skills Inventory for Students by Entwistle (1997), the Learning Process Questionnaire by Biggs (1987), the Survey of Study Habits and Attitudes, Form C by Brown and Holtzman (1984), and the Study Behavior Inventory by Bliss *et al.* (2000).

Teaching of Learning Strategies

Another controversy in the current literature focuses on the best way to teach students about learning strategies and how to use various learning strategies. This controversy also relates to the debate over the domain of applicability issue and whether we should teach content-dependent or content-independent strategies. Proponents on one side of the controversy believe that learning strategies should be taught in content courses, such as math, history, or biology, and not as a separate course. Proponents on the other side of the controversy believe that learning strategies should be taught as part of a course or training program in strategic or self-regulated learning. Recent research literature indicates that both groups are right – for students who are highly deficient in their knowledge and use of learning strategies, an adjunct course is best. For students who already have some level of skill, refining their knowledge and skills in content-dependent settings appears to be more helpful.

However, the most powerful instructional model appears to be a combination of the two. Using an adjunct course to help teach students general knowledge and skills in using strategies with a broad domain of applicability combined with what Weinstein has called the metacurriculum in content courses. Basically, a metacurriculum involves purposefully teaching learning strategies while also teaching course content. It can be as simple as paraphrasing a lesson and then teaching the class to do the same thing on their own to check their understanding, or as complex as teaching students how to develop, implement, monitor, and modify a test-preparation plan for an upcoming exam. For example, implementing a metacurriculum would involve not only teaching students what to learn in a history course but also how to learn and think like an historian.

Concluding Statement

It has often been said that the present belongs to those who have learned but the future belongs to those who are

learning. Increasing longevity and our increasingly complex and technologically sophisticated world requires that students be prepared to be lifelong learners throughout the different stages of their life span. Having an extensive repertoire of learning strategies is one step toward helping individuals become more effective and efficient learners.

See also: Knowledge Domains and Domain Learning; Metacognition; Personal Epistemology in Education: Concepts, Issues, and Implications; Self-Regulated Learning and Socio-Cognitive Theory.

Bibliography

- Biggs, J. B. (1987). *The Learning Process Questionnaire (LPQ): Manual*. Hawthorn: Australian Council for Educational Research.
- Bliss, L. B., Kerstiens, G., and Marvin, R. (2000). *The Study Behavior Inventory (Version 2.0)*. Torrance, CA: Androgogy Associates.
- Brown, W. F. and Holtzman, W. H. (1984). *Survey of Study Habits and Attitudes, Form C*. New York: Psychological Corporation.
- Entwistle, N. J. (1997). *The Approaches and Study Skills Inventory for Students (ASSIST)*. Edinburgh: University of Edinburgh Centre for Research on Learning and Instruction.
- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review* 16(4), 385–407.
- Pintrich, P., Smith, D., Garcia, T., and McKeachie, W. (1991). *Motivated Strategies for Learning Questionnaire*. The University of Michigan: National Center for Research to Improve Postsecondary Teaching and Learning.
- Weinstein, C. E. and Mayer, R. E. (1986). The teaching of learning strategies. In Wittrock, M. C. (ed.) *Handbook of Research on Teaching*, 3rd edn., pp 315–327. New York: Macmillan.
- Weinstein, C. E. and Palmer, D. R. (1990). *Learning and Study Strategies Inventory-High School Version*. Clearwater, FL: H & H Publishing.
- Weinstein, C. E., Schulte, A., and Palmer, D. R. (2002). *The Learning and Study Strategies Inventory*, 2nd edn. Clearwater, FL: H & H Publishing.
- Weinstein, C. E., Tomberlin, T. L., Julie, A. L., and Kim, J. (2004). Helping students to become strategic learners: The roles of assessment, teachers, instruction, and students. In Ee, J., Chang, A., and Tan, O. (eds.) *Thinking about Thinking: What Educators Need to Know*, pp 282–310. Singapore: McGraw-Hill.
- Zimmerman, B. J. and Schunk, D. H. (eds.) (2001). *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*, 2nd edn. Mahwah, NJ: Erlbaum.

Further Reading

- Biggs, J. and Tang, C. (2007). *Teaching for Quality Learning at University*, 3rd edn. Berkshire: Open University Press/McGraw-Hill.
- Boekaerts, M., Pintrich, P. R., and Zeidner, M. (eds.) (2000). *Handbook of Self-Regulation*. San Diego, CA: Academic Press.
- Boekaerts, M., Renninger, K. A., Sigel, I. E., Damon, W., and Lerner, R. M. (2006). *Self-Regulation and Effort Investment*. Wiley: Hoboken, NJ.
- Dembo, M. H. and Sell, H. (2007). *Motivation and Learning Strategies for College Success: A Self-Management Approach*, 2nd edn. Erlbaum: Mahwah, NJ.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist* 34(10), 906–911.
- Janac, K., Kipperman, D., and Linder, D. (1997). *Learning Strategies Matrix*. http://edweb.sdsu.edu/courses/ET650_online/MAPPS/Strats.html (accessed August 2009).
- Jones, B. F. and Idol, L. (1990). *Dimensions of Thinking and Cognitive Instruction*. Hillsdale, NJ: Erlbaum.
- McNamara, D. S. (2007). *Reading Comprehension Strategies: Theories, Interventions, and Technologies*. Mahwah, NJ: Erlbaum.
- New Horizons for Learning (2002). *Teaching and Learning Strategies*. http://www.newhorizons.org/strategies/front_strategies.html (accessed August 2009).
- Paris, S. G. and Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist* 36(2), 89–101.
- Schunk, D. H. and Zimmerman, B. J. (eds.) (2007). *Motivation and Self-Regulated Learning: Theory, Research and Application*. Mahwah, NJ: Erlbaum.
- Weinstein, C. E. and Hume, L. M. (1998). *Study Strategies for Lifelong Learning*. Washington, DC: American Psychological Association.
- Weinstein, C. E., Meyer, D. K., Van Mater Stone, G., and McKeachie, W. J. (2006). Teaching students how to learn. In McKeachie, W. J. and Svinicki, M. (eds.) *Teaching Tips: Strategies, Research, and Theory for College and University Teachers*, 12th edn, pp. 300–317. Lexington, MA: Houghton Mifflin.
- Wolters, C. A. (2003). Regulation of motivation: Evaluating an underemphasized aspect of self-regulated learning. *Educational Psychologist* 38(4), 189–205.

Relevant Websites

- <http://www.howtostudy.org> – How to Study.
- <http://www.muskingum.edu> – Muskingum, CAL Learning Strategies Database.
- <http://www.studygs.net/index.htm> – Studies Guides and Strategies.

Learning Through Play

K Aronsson, Stockholm University, Stockholm, Sweden

© 2010 Elsevier Ltd. All rights reserved.

Play Dichotomies

Historically, play has often been described as a type of mode that constitutes a natural way of being close to children, innocence, and nature. Play has been set up as a contrast to school, where pupils have been seen as unduly tamed citizens, colonized by methods that are alien to the child itself. Critical thinkers like Rousseau and Huizinga, and acclaimed authors like Goethe and Tolstoy have all foregrounded and celebrated the creative nature of play. Consequently, they have advocated pedagogies that offer leeway for the child's own free roaming, and playful explorations. For instance, on his estate in tsarist Russia, Tolstoy educated the children of poor peasants through Socratic dialogues, play, and creative writing.

Within the social sciences, play has often been studied in terms of instrumental goals. Following Sigmund Freud, Erik H. Erikson, and other psychodynamic thinkers, play has been advocated as the therapeutic method par excellence in work with clinical and psychiatric problems of children. For instance, play-world methods, which are widely used within child psychiatry, and in other clinical contexts, draw on the basic idea that children display and learn to handle envy, jealousy, and other difficult feelings through the use of play. Thereby, play becomes both a diagnostic and therapeutic method. Similarly, play has been advocated as an important method for learning. In both cases, diagnostic/therapeutic or instructional goals are set as the privileged or superordinate goals of play activities. Play is thus situated within a set of implicit value dichotomies that form a hierarchy, where play and exploration is generally cast as what is less valuable and that has to be upgraded through connections to what is serious:

Play	Work/school
Play	Game
Role play	Play with rules
Simplicity	Complexity
Nature	Culture
Exploration	Instruction

Play is thus not studied per se but subsumed under ulterior societal goals, such as diagnosis, therapy, or instruction. Moreover, such theorizing on play situates play in terms of a number of set more-or-less fixed dichotomies.

In contrast, exploratory models of play celebrate the inherent qualities of play. Play in itself is worthy of an

important role in child development as it is the child's natural or authentic mode of being. In fact, the human being should be seen as a playing creature, someone who has an inborn capacity and predisposition for play, what Huizinga has called *homo ludens*. Within Vygotskian theorizing and activity theory (e.g., Vygotsky, 1976, 1995), play has been theorized as a key mode of thinking in childhood. On a theoretical note, two of the most influential developmental theorists of the twentieth century, Jean Piaget and Lev Vygotsky, in fact discussed the role of exploration in learning, arguing that play forms an important aspect of learning.

Play, Semiotic Competence, and Mediation

Piaget (1945) claimed that cognition and learning always require adaptation. More specifically, adaptation necessarily involves both imitation (accommodation) and play and subjective thinking (assimilation). From the infant's first attempts at world making, learning and cognition draw both on imitation and other attempts to mimic others, and on assimilation, that is, play and subjective meaning. For instance, an infant may imitate words of its caregivers but endow them with idiosyncratic meaning. In their very first appropriations of words in their mother tongue, Swedish infants may, for instance, employ *pappa* (daddy) for *lampa* (lamp) as the two Swedish words are phonologically fairly similar, and both phenomena tend to be objects of interest within the young child's world making.

In much of his early work in the 1920s–1940s, Jean Piaget had a phenomenological focus, manifested in a series of ethnographic studies of children as philosophers. This phenomenological period involved extensive theorizing on play, on the basis of observations of his own children's (infants/toddlers) actions in nursery-room contexts (Piaget, 1945). He claimed that cognition involves a constant balancing (equilibration) between the known and the unknown. The child therefore repeats, experiments with, and recombines what is not yet quite mastered. The learning of a first language is a clear example of the ways in which children learn, not through explicit instruction, but through observations, exploration, and experimenting, which can, for instance, be seen in children's hyper-corrections and other self-corrections (for a review of contemporary work, see Rogoff *et al.*, 2003). Within

Piagetian theorizing, such exploratory work will characterize all the semiotic modes of young children: language, art, imitation (including pantomime), and role play as well as construction games (Piaget, 1945). A basic notion is that play is a form of intelligence that is parallel to and equivalent to language, imitation, and art in the child's early development. However, fantasy play or role play is primarily an aspect of the young child's early learning.

In contrast, Vygotsky (1976, 1995) locates both play and game activities at the heart of his theorizing on children's activities and development. In fact, he sees play (including games) as something that defines much of the child's imagination and combinatorial thinking. In contrast to Piaget, who did not discuss the role of fantasy play in later learning or modernists like Herbert Read (who discussed children's art work as something that is arrested through schools and schooling), Vygotsky (1995) claimed that the child's imagination continues to develop all though the school years. This is partly an effect of the fact that mediation is always an important aspect of learning. A given mode of mediation, for example, creative writing, may involve difficulties that the child does not encounter in verbal story telling (Vygotsky, 1995). The medium as such (e.g., speaking/writing/in contemporary society also new media) involves specific difficulties and challenges. In Vygotskian theorizing, one of the important roles of teachers and other masters is to build on children's play activities by introducing topics that appeal to their imagination, and by choosing types of mediation (drama, play, writing, art work, and new media activities) that appeal to children in relation to their stage of thinking and in relation to a specific area of learning. Today, it can be seen that children spontaneously engage in quite complex cognitive activities when engaging in, for instance, digital activities on the Internet. Play as such is one type of mediation, what Vygotsky would call an intellectual tool in children's development. This also means that play can be an intellectual resource all through life.

Piaget, Vygotsky, and Peer Learning

Drawing on his work on children's thinking around games of rule, Piaget (1932) presented a model of informal learning and social order, claiming that school-age children develop much of their moral thinking from social exchange with peers, and in particular from intellectual disputes with peers about games with rules (e.g., marbles). Peer relations are more symmetrical than adult-child dialogs, and children may therefore engage more wholeheartedly in critical disputes with peers, challenging the others' thinking. In contrast to their conflicts with adults, children who argue with peers thus tend to engage in more extended and more elaborate arguments. Thereby, they develop their perspective talking skills in ways that

they do not do with adult authorities. However, even though Piaget emphasized the role of peers for critical thinking, he did not document sibling or peer-play dialogs, whereas he did document adult-child-play interactions (Piaget, 1945), and adult-child dialogs on games with rules (Piaget, 1932).

In contrast, Vygotsky (1976) meant that all human behavior, including social life, includes rules. Therefore, it is not meaningful to see role play as an activity that has to be less complex than game activities and construction play. In Russian, there is but one word, *igra*, for both play and games. However, Vygotsky (1976) explicitly challenged a major divide between role play and play with rules, as well as any major divide between exploration and instruction. Instruction does not arrest imagination and free exploration. In fact, imagination increases as children appropriate more knowledge of the world (Vygotsky, 1995).

He theorized on the role of apprenticeship in children's learning, and on informal master-apprentice relations that form zones of proximal development, where children may surpass themselves, moving into a higher stage of mental processes, if assisted by a master. He claimed that, when playing, children would rise above themselves. "In play, a child is always above his average age, above his daily behaviour; in play it as though he were a head taller than himself" (Vygotsky, 1976: 552). Play thus constitutes a potential zone of proximal development, where children may assist each other in exploring the world. Moreover, play is about recollections or memory in action which means that children collaboratively advance their imaginative faculties, as they gradually appropriate more knowledge. This, of course, also means that collaborative play may profit from the accumulated prior experiences of all co-present peers.

Whereas Piaget foregrounded the developmental role of games with rules and peer disputes, Vygotsky emphasized the role of all kinds of play. Thereby, he also indirectly laid the ground for a focus on peers in children's development. Along with Piaget's analyses of peer disputes, his foregrounding of mediation, including both play and games, has prepared the ground for a novel focus on peer dialogs in children's learning. In contemporary language, socialization studies, and sociocultural theorizing, play and play communities have become central foci, and there is today a growing awareness of the role of peers for children's learning (Goodwin, 2006; Whalen, 1995).

Play and Intent Participation

Recently, Rogof *et al.* (2003) have reviewed various ways in which children acquire new knowledge without formal instruction but through keen observations and listening-in, what they have called intent participation. Such

informal apprenticeship would often take place in various horizontal types of collaboration, for example, between siblings or peers or with adults when children are allowed to watch and listen-in. In contrast to the lesson-style conversational practices that often characterize middle-class adults and their ways of socializing children, for example, at dinner time or in school contexts, such informal practices may thus involve keen observations, mimicking, and exploration, rather than instruction. These practices may involve various aspects of children's appropriations of cultural know-how: acquiring a first language, learning other tasks, and the appropriation of local norms for what is an accepted community member.

Play as Participation and Play Communities

Recently, a substantial body of work has been devoted to participation and its role in informal apprenticeships, and more specifically, to the ways in which people become fully competent members of communities of practice (Lave and Wenger, 1991). It is argued that apprenticeship is often linked to membership in communities of practice, where members advance from onlookers and peripheral participants to fully accepted members (Lave and Wenger, 1991). Lave and Wenger primarily discussed adult communities, such as alcoholics anonymous, where participants gradually move from peripheral to full participation. In the field of play, such emergent participation may, for instance, involve informal play communities, related to hip-hop culture, pogo-game activities (Sparrman and Aronsson, 2003), computer gaming, graffiti, skateboarding, or hopscotch (Goodwin, 2006), that is, various play and game activities, where children and youth in informal groups learn from each other within informal communities of practice.

Vygotsky (1995) described children's play products as crystallized imagination. In line with Vygotskian theorizing, recent research shows that such imagination is not located in a social vacuum. Imagination emerges within a given social order, and it is socially ordered. The value that is placed on, for instance, play materials tends to be strictly regulated within local communities of practice, where peers play together and in various ways assess each others' play performances and play materials. Children who, for instance, play with pogs (Sparrman and Aronsson, 2003) constantly assess each others' pogs and their trade values in ways that establish local hierarchies of what Bourdieu would call symbolic capital.

At a first glance, children's play worlds may seem totally free and accessible. Yet, one of the first matters to be learnt by young children is when and how to get access to play – to playmates, to peers' sustained attention, and to play materials. In his documentations of young toddlers'

play practices, Corsaro (1979) documented ways in which children acquired various access rituals for gaining the attention of desirable playmates. Such access rituals were apparently learnt through prior successful and unsuccessful attempts at achieving joint attention in play. What could be seen was that even very young preschoolers engaged in quite complex perspective taking when trying to secure access to ongoing play activities. In her work on mixed-age peers, Whalen (1995) has similarly documented that quite young preschoolers engage in complex tactics, calibrating their play directives in relation to peers' prior utterances, and that they delicately avoid alliances that might upset co-participants (and eventually lead to the arrest of play). It could be argued that children recurrently form what could be called play communities, where it is important for participants to be able to take part in local teasing, joking, and other play practices. Obviously, such communities are closely tied to participation, and to local rules for inclusion or exclusion in play activities.

On the basis of a broad review of prior work on play in different cultures and in contemporary society as well as in historical settings, Schwartzman (1978) has demonstrated that play negotiations take up a great proportion of children's playtime throughout much of children's play lives, and she therefore advocated an increasing attention to issues concerning access to play. The negotiations may concern who gets access to material resources as well as who gets to play what role or in what ways or with whom. Play negotiations – or what we could call play politics – thus concern both physical and virtual play resources. Play politics of course also concerns who gets to play, and even more importantly, who does not get to play (Garvey, 1993). In his work on preschoolers' play improvisations, Sawyer (1997) observed that young children spent more time than older ones on planning, negotiations, and other types of meta-communication about play. Moreover, he found that friendship had an effect that was similar to age and experience in that friends would go more seamlessly into play; needing less time for planning and coordination.

Threat-Tell, Teasing, Format Tyings, and Play Politics

As shown, jokes and other playful performances are often important discursive resources for securing participation in local play communities. Teasing, threats, and play politics have been documented in great detail by Marjorie Goodwin, who has conducted video ethnographies of children's play practices in school yards and neighborhood areas (e.g., Goodwin, 2006). She has analyzed the ways in which children spend much time, negotiating with each other about inclusion and exclusion and play rules,

as well as about material resources. Moreover, she has moved beyond gender stereotypes, showing that both boys and girls engage in insults, gossip, teasing, and other play practices that, at times, border on what is aggressive or violent. In fact, verbal insults, at times, take the form of what she has called highly complex format tying practices, that is, whole or partial repeats of prior utterances, where the participants build upon each others' contributions. Through creative repetitions of the prior speaker's insults, a child may thus build on peers' repertoires of invectives and insults, building new and more elaborate versions.

Such conversational work can already be seen in the play of preschoolers. Traditionally, Piagetian and other theories on play have generally positioned games and rule play as more complex forms of play than role play or so-called imaginative play. However, contemporary research, involving videorecordings of children's play worlds, has revealed a different picture. One of the pioneers within psychology is Garvey (1993). Like Goodwin, she has shown that recyclings, or what she called alternatives, were highly productive tactics in conflict talk among preschoolers' play practices. She also reported on ways in which role play may be quite complex, requiring substantial account work, that is, social explanations or reason giving, and other types of mitigations on the part of the participants. On the basis of analyses of the conversational work of preschool and school children's play dyads, Aronsson and Thorell (1999) have reported similar findings, showing in detail ways in which preschool children engage in quite elaborate perspective taking in order to solve play-conflict scenarios (about who is to use the only available play equipment, in this case a bike).

In their work on children's joking events and second-language learning in a Swedish first-grade-immersion-classroom setting, Čekaite and Aronsson (2004) have documented how refugee and immigrant children recurrently engaged in meta-pragmatic play, that is, play that comments on ways of saying things in that the participants joked about or enacted stylized performances of the ways in which something would be said by other role players, for instance, teachers. During informal work in the classroom, peers thus made funny role enactments or teasing each other for community members' mutual enjoyment.

Recently, work on teasing practices have also shown ways in which older school-age children create complex tactics, where they build upon prior insults in subtle ways, turning the weapons of their attackers against them as it were (Tholander and Aronsson, 2002).

Learning What?

What do children learn in their play dialogs? Apparently, much of a first language (or a second language in

immersion contexts) is learnt through playful repetitions, and increasingly more complex variations. Such informal learning involves all types of semiotic mediation, and both, what is said and how, that is, matters of style and so-called pragmatic awareness. Children are neither formally instructed in their first language, nor in how it should be used in terms of everyday diplomacy or politics, for instance, when and how to deploy mitigations and other politeness strategies. Yet, they obviously acquire quite broad repertoires of styles that can be deployed to upgrade or downgrade the intensity (and offensiveness) of potentially sensitive messages. Play dialogs are important sites for displaying and learning such strategies. If we do not assume that there is substantial learning through intent participation, it is difficult to explain the skills with which even preschool-age children eventually engage in intricate play dialogs, adopting the perspective of others, and predicting the ways in which conversational strategies may upgrade or downgrade potential challenges to the hearer. Role play is thus often as complex as play with rules in that children have to take into account both their co-participants' past moves, and their projected uptake. Simultaneously, all this has to be calibrated to local rules of comportment, local diplomatic protocols for what may be said in what way. Yet, this largely takes place without any explicit instruction. It means that play theories should by now move away from any rigid dichotomies between role play, on the one hand, and play with rules, on the other. Vygotsky (1976) similarly claimed that there is no major divide between work and play. In children's lives, play is serious business. Contemporary studies show that play dialogs involve important relational work in that children display and acquire conversational tactics for handling conflicts. This can also be seen both in children's extended play-access rituals, and in their sustained engagement in play activities.

See also: Learning in a Sociocultural Perspective; Learning Through Narrative Socialization; Peer Interaction and Learning; Piaget: Recent Work; Reasoning and Explanatory Talk: Learning in Everyday Settings; Situated View of Learning; Vygotsky and Recent Developments.

Bibliography

- Aronsson, K. and Thorell, M. (1999). Family politics in children's play directives. *Journal of Pragmatics* **31**, 25–47.
- Čekaite, A. and Aronsson, K. (2004). Repetition and joking in children's second language conversations. Playful recyclings in an immersion classroom. *Discourse Studies* **6**, 373–392.
- Corsaro, W. (1979). 'We're friends, right?' Children's use of access rituals in a nursery school. *Language in Society* **8**(3), 315–336.
- Garvey, C. (1993). Diversity in the conversational repertoire. The case of conflicts and social pretending. *Cognition and Instruction* **11**(3 & 4), 251–264.
- Goodwin, M. H. (2006). *The Hidden Life of Girls. Games of Stance, Status and Exclusion*. Oxford, UK: Blackwell.

- Lave, J. and Wenger, E. (1991). *Situated Learning. Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Piaget, J. (1932). *The Moral Judgment of the Child*. London: Routledge Kegan Paul.
- Piaget, J. (1945). *La formation du symbole chez l'enfant. Imitation, jeu et rêve, image et représentation*. ((1951); *Play, Dreams, and Imitation in Childhood*. New York: Norton.) Paris: Delachaux et Niesté.
- Rogoff, B., Paradise, R., Arauz, R. M., Correa-Chávez, M., and Angelillo, C. (2003). Firsthand learning through intent participation. *Annual Review of Psychology* **54**, 175–203.
- Sawyer, R. K. (1977). *Pretend Play as Improvisation*. Mahwah, NJ: Erlbaum.
- Schwartzman, H. (1978). *Transformations. The Anthropology of Children's Play*. New York: Plenum Press.
- Sparman, A. and Aronsson, K. (2003). Pog game practices, learning and ideology. Local markets and identity work. In Walford, G. (ed.) *Investigating Educational Policy through Ethnography*, pp 169–192. Amsterdam: Elsevier.
- Tholander, M. and Aronsson, K. (2002). Teasing as serious business. Collaborative staging and response work. *Text* **22**, 559–595.
- Vygotsky, L. (1976). Play and its role in the mental development of the child. In Bruner, J., Jolly, A., and Sylva, K. (eds.) *Play. Its Role in Development and Evolution*, pp 537–554. Harmondsworth: Penguin Books.
- Vygotsky, L. (1995). *Fantasi och kreativitet i barndomen ('Imagination and Creativity in Childhood')*. Göteborg: Daidalos. (Vobraženie i tvorčestvo v detskom vozraste; Moscow: Prosvesčenie.)
- Whalen, M. R. (1995). Working toward play. Complexity in children's fantasy activities. *Language in Society* **24**, 315–348.
- Aronsson, K. (1998). Identity-in-interaction and social choreography. *Research on Language and Social Interaction* **31**, 75–89.
- Aronsson, K. and Thorell, M. (2002). Voice and collusion in adult-child talk. Toward an architecture of intersubjectivity. In Blum-Kulka, S. and Snow, C. E. (eds.) *Talking to Adults*, pp 227–293. Mahwah, New Jersey: Erlbaum.
- Čekaite, A. and Aronsson, K. (2005). Language play, a collaborative resource in children's L2 learning. *Applied Linguistics* **26**(2), 169–191.
- Cromdal, J. and Aronsson, K. (2000). Footing in bilingual play. *Journal of Sociolinguistics* **4**, 435–457.
- Evaldsson, A. C. (2005). Staging insults and mobilizing categorizations in a multiethnic peer group. *Discourse and Society* **16**, 763–786.
- Evaldsson, A. C. and Corsaro, W. (1998). The ethnographic study of play and games in children's peer culture. An interpretative approach. *Childhood* **5**, 377–402.
- Garvey, C. and Kramer, T. (1989). The language of social pretend play. *Developmental Review* **9**(4), 364–382.
- Goodwin, M. H. (1990). *He-Said-She-Said. Talk as Social Organization among Black Children*. Bloomington: Indiana University Press.
- Huizinga, J. (2004). *Homo Ludens*. Stockholm: Natur och Kultur. (original in German, 1938).
- James, A. (1996). Learning to be friends. Methodological lessons from participant observations among English schoolchildren. *Childhood* **3**, 313–330.
- Maynard, D. (1985). How children start arguments. *Language in Society* **14**(1), 1–30.
- Mead, G. H. (1934). *Mind, Self and Society*. Chicago, IL: Chicago University Press.
- Rindstedt, C. and Aronsson, K. (2002). Growing up monolingual in a bilingual community. The Quichua revitalization paradox. *Language in Society* **31**, 721–742.
- Sheldon, A. (1996). You can be the baby brother, but you aren't born yet. Preschool girls' negotiations for power and access in pretend play. *Research on Language and Social Interaction* **29**(1), 57–80.

Further Reading

- Andersen, E. (1990). *Speaking with Style. The Sociolinguistic Skills of Children*. London: Routledge Kegan Paul.

Situated View of Learning

A Collins, Northwestern University, Evanston, IL, USA

J G Greeno, University of Pittsburgh, Pittsburgh, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cognitive apprenticeship – A method of teaching school subjects that embodies many of the practices of apprenticeship training, such as modeling, coaching, and careful observation of the learner's practice.

Community of practice – A group of people participating together to carry out different activities, such as garage bands, ham-radio operators, recovering alcoholics, and research scientists.

Goal-based scenario – Educational settings where learners are given real-world tasks and the support they need to carry out such tasks; they can be set either in computer-based environments or naturalistic environments.

Legitimate peripheral participation – Describes the relationship of apprenticeship learners who have access to the practices that they are expected to learn and to genuine participation in the activities and concerns of the group.

Situative – A perspective on cognition, learning, and education that analyzes processes of interaction, in which individuals participate, along with other material and informational systems.

Many practices of conventional schooling consider knowledge and skill as discrete structures of cognition that can be adequately transferred from teachers to students in classrooms and studied in laboratories. Knowing and thinking, in this view, are assumed to go on in individual minds isolated from the complexity of the world outside, from which abstract knowledge can be successfully distilled. However, a growing body of research that considers cognition and learning in activities outside of specialized learning environments is undermining the plausibility of these presuppositions (e.g., Brown *et al.*, 1989; Engeström, 2001; Greeno *et al.*, 1996; Hutchins, 1995a; Lave and Wenger, 1991; Nersessian *et al.*, 2003; Rogoff, 1990; Rogoff and Lave, 1984). This research supports the view that knowing and learning by individuals are inextricably situated in the physical and social contexts of their acquisition and use. It is a mistake to think that classrooms or laboratory experiments produce knowledge or follow principles of learning

that are somehow context free. Cognition and learning by individuals always occur in a context; the issue has to be what the context is, not whether there is one.

For most of the last half-century, active research programs have been studying structures and processes of social interaction, as well as cognitive processes of representing and transforming information. However, these research programs have been largely separate from each other. Situative research and theorizing attempts to unify the two perspectives of individual cognitive theory and the analysis of interactional structures and processes. The primary level of a situative analysis is an activity system, in which one or more individuals participate along with material and informational resources in the environment. Cognitive processes are understood as aspects of the practices of a community or group. Studies include analyses of perceiving (Goodwin, 1996), remembering (Hutchins, 1995b), reasoning and understanding (Greeno and van de Sande, 2007; Ochs *et al.*, 1996), and learning (Bowers *et al.*, 1999; Engeström, 2001; Engle, 2006; Stenning *et al.*, 2002). In these analyses, successful cognitive performances are considered as part of an interactive system, and analyses focus on how the multiple participants coordinate their contributions. Information structures, which the individual cognitive perspective attributes to individual minds, are attributed in the situative perspective to the interacting group as achievements of communication that enter the group's common ground (cf. Clark, 1996). Such analyses do not preclude also having analyses of the same events that focus on one or more individual participants, identifying their respective contributions to the interactions, and explaining these in terms of their individual capabilities, and with other participants and systems considered as the context (cf. Bowers *et al.*, 1999; Hatano and Inagaki, 2003).

If knowing is understood as successful situated participation, then many conventional assumptions must be questioned. In particular, a situative theory of knowing challenges the widely held belief that abstraction of knowledge from situations is the key to transferability. An examination of the role of situations in structuring knowledge suggests that abstraction and explication provide an inherently impoverished and often misleading view of knowing. Knowing by an individual is fundamentally a capability of the person to interact in the world. In this view, hypotheses or assessments of an individual's or

group's knowing are about their capabilities for interacting in situations. Hypotheses that represent knowledge only as abstract propositions do not capture the densely interwoven nature of knowing.

The situative perspective views knowing as distributed among people and their environments, including the objects, artifacts, tools, books, and the communities of which they are a part. Analyses of activity focus on processes of interaction of individuals with other people and with physical and technological systems. Several research traditions have contributed to the situative perspective. The best established of these is ethnography, including the study of cultural practices and patterns of social interactions, as well as discourse analysis and conversation analysis in activity theory, sociolinguistics, anthropology, and sociology. Another research tradition is ecological psychology, which studies behavior as physical interaction in which animals, including people, participate in physical and technological systems. A third research tradition is situation theory in logic and philosophy, which analyzes meaning and action as relational systems and is developing a reformulation of logic to support these relational analyses. Knowing in this perspective is both an attribute of groups that carry out cooperative activities and an attribute of individuals who participate in the groups. Learning by a group or individual involves becoming attuned to constraints and affordances of the material and social systems with which they interact. Discussions of motivation in this perspective often emphasize engagement of individuals with the functions and goals of the community, including interpersonal commitments and ways in which individuals' identities are enhanced or diminished by their participation.

Apprenticeship and Identity

When knowing is viewed as practices of communities and of the abilities of individuals to participate in those practices, then learning is the strengthening of those practices and participatory abilities. Systems in which individuals learn to participate in social practices are very common, and include apprenticeship and other forms of being initiated into the practices of a group. Lave and Wenger (1991) reviewed several studies of learning by newcomers to communities of practice and concluded that a crucial factor in the success of such a system is that learners must be afforded legitimate peripheral participation, which involves access to the practices that they are expected to learn and genuine participation in the activities and concerns of the group. Lave and Wenger characterized learning of practices as processes of participation in which beginners are relatively peripheral in the activities of a community, and as they become more experienced and

adept, they progress toward fuller participation. A crucial issue in the nature of learning is whether, and in what ways, the peripheral participation of beginners is legitimate. They described four cases of learning by newcomers and emphasized how learners' identities derive from being part of the community as they become more fully participating members in the community. They also noted that an apprenticeship relationship can be unproductive for learning, as in a case of meat cutters they cited, where the apprentices worked in a separate room and were isolated from the working community. For an environment of apprenticeship to be a productive environment of learning, learners need to have opportunities to observe and practice activities in order to progress toward fuller participation.

The degree to which people participate fully and are respected by other members of a community determines their sense of identity (Lave and Wenger, 1991; Wenger, 1998). The fully participative roles are those that most directly contribute to the collective activities and knowledge of the community. The motivation to participate more fully in a community of practice can provide a powerful incentive for learning. Smith (1988) argued that children learn to read and write if the people they admire read and write. That is, they will want to join the literacy club and will work hard to become members. Learning to read is part of becoming the kind of person they want to become. Identity is central to deep learning.

An important aspect of learners' identities is the way in which they are positioned in the participant structures (Phillips, 1972) of learning activities. An important distinction by Pickering (1995) involves different kinds of agency, called conceptual and disciplinary. Students who are positioned with a disciplinary agency only participate as receivers and reproducers of the established meanings and procedures of the discipline, and their learning is evaluated only by whether they can perform procedures and explanations correctly. Students who are positioned with conceptual agency are expected to question and adapt concepts and methods of the discipline. For example, they might construct understandings that utilize disciplinary concepts in novel ways or consider alternatives to standard definitions of concepts. As an example, research by Boaler (2002) compared learning of mathematics in two English secondary schools and found that students who learned primarily through investigations understood mathematics as a general resource for understanding and problem solving, whereas students whose learning was primarily mastery of set procedures understood mathematics as a set of rules to be followed.

Wenger (1998) argued that people participate in a variety of communities – at home, at work, at school, and in hobbies. In his view a community of practice is a group of people participating together to carry out different

activities, such as garage bands, ham-radio operators, recovering alcoholics, and research scientists.

For individuals, it means that learning is an issue of engaging in and contributing to the practices of their communities. For communities, it means that learning is an issue of refining their practice and ensuring new generations of members. For organizations, it means that learning is an issue of sustaining the interconnected communities of practice through which an organization knows what it knows and thus becomes effective and valuable as an organization (Wenger, 1998: 7, 8).

The view that learning occurs through participation is at the root of the practices of apprenticeship, where apprentices are guided and supervised by masters. In successful apprenticeship learning, masters teach by showing apprentices how to do a task (modeling), and then helping them as they try to do it on their own (coaching and fading). Lave and Wenger (1991) emphasized how an apprentice's identity derives from becoming part of the community of practitioners. The motive for becoming a fuller participant in a community of practice can provide a powerful motivation for learning. Of course, what is learned in apprenticeship may not generalize easily to other contexts. Collins *et al.* (1989) attempted to characterize how the modeling, coaching, and fading paradigm of apprenticeship might be applied to learning the cognitive subjects of school in an approach they called cognitive apprenticeship.

Educational Applications of the Situated View

A major goal of educational reform is to have students participate more actively and legitimately in learning communities, including participation in formulating and evaluating questions and problems, and constructing and evaluating hypotheses, evidence, arguments, and conclusions (Brown and Campione, 1996). Abilities for participating in these activities have to be learned, and the research literature on that kind of learning is sparse. Several projects have been focused on creating classroom practices of discussion and inquiry, and the investigators in those projects have discussed some aspects of the process of establishing norms and expectations by the students that support productive collaborative learning (Cohen, 1986; Lampert, 1990; Slavin, 1983).

In the view of learning as coming to participate more fully in a community of practice, transfer is often thought to be a problematic issue (e.g., Anderson *et al.*, 1996). Viewed in the situated perspective, transfer can occur when learning leads to better performance or learning of new practices within a community (e.g., for school communities this might mean working on new problems

or accomplishing new kinds of tasks) or outside the community (e.g., for school these might be work environments such as those studied by Beach (1995) and Saxe (1990)). Many of the resources and supports that occur within a community of practice do not carry over to a different community, and so the problem of transfer becomes one of marshalling the resources needed to be successful in a new environment. This requires sophisticated social and information-processing skills, which are the kinds of skills that businesses think they will need in the future.

In a view of transfer in the situative perspective proposed by Greeno *et al.* (1993), transfer depends on constraints and/or affordances that are invariant under the transformations that change the learning situation into the transfer situation. For transfer to occur, learners must become attuned to those invariants in their initial learning. One of the ways to be attuned is to have an abstract representation that can be applied in the new situation, but this is only one possible way for attunement to occur, and may not be the typical way to generalize many learned activities (Greeno, 1997).

Although the situative view insists that all cognition and learning are situated, learning designers who take a situative perspective generally attend to the activity settings in which learning is to occur. For example, in goal-based scenarios (Schank *et al.*, 1994, Nowakowski *et al.*, 1994) learners are given real-world tasks and the scaffolding they need to carry out such tasks. They can be set either in computer-based environments or naturalistic environments. In one computerized goal-based scenario, learners are asked to advise married couples as to whether their children are likely to have sickle-cell anemia, a genetically linked disease. In order to advise the couples, learners must use the facilities in the system to find out how different genetic combinations lead to the disease and run tests to determine the parents' genetic makeup. There are scaffolds in the system to support the learners, such as various recorded experts who offer advice. Other goal-based scenarios support learners in a wide variety of challenging tasks, such as putting together a news broadcast, solving an environmental problem, or developing a computer-reservation system. Goal-based scenarios make it possible to embed cognitive skills and knowledge in the kinds of contexts where they are to be used. Therefore, people learn not only the basic competencies they will need, but also when and how to apply these competencies.

Video and computer technology has enhanced the ability to create simulation environments where students are learning skills in context. A novel use of video technology is the Jasper series developed by the Cognition and Technology Group (1997) at Vanderbilt University to teach middle-school mathematics. In a series of 15–20-min videos, students are put into various problem-solving contexts: for example, deciding on a business plan for a school fair or a rescue plan for a wounded eagle. The problems are

quite difficult to solve and reflect the complex problem solving and planning that occurs in real life. Middle-school students work in groups for several days to solve each problem. Solving the problems develops a much richer understanding of the underlying mathematical concepts than the traditional school-mathematics problems.

Another novel use of technology is the curriculum developed by the middle-school mathematics through applications project (MMAP) at the Institute for Research on Learning (Goldman and Moschkowich, 1995; Greeno *et al.*, 1999). The leading activities in the MMAP curriculum are design problems, supported by software that provide computer-aided design environments in which students design floor plans of buildings, models of population growth and decline, lexicographic codes, or geographical analyses of environmental quality. Mathematical reasoning and problem solving involving topics such as proportional reasoning, linear and exponential functions, and geometrical properties of geographical space are required for successful progress in the design activities. Printed curriculum materials are provided to support teachers in organizing activities for students to encounter, recognize, and learn important mathematical concepts and methods.

These kinds of learning tasks are different from most school tasks, because the contexts of most school tasks lack characteristics of practices that occur outside of school. Imagine learning tennis by being told the rules and practicing the forehand, backhand, and serve without ever playing or seeing a tennis match. If tennis were taught that way, it would be hard to see the point of what you were learning. But in school, students are taught algebra and Shakespeare without being given any idea of how they might be useful in their lives. That is not how a coach would teach you to play tennis. A coach might first show you how to grip and swing the racket, but very soon you would be hitting the ball and playing games. A good coach would have you go back and forth between playing games and working on particular skills – combining global learning with focused local knowledge. The essential idea in the situative view of learning is to consider learning and cognition as participation in an activity system. This view supports designers' and educators' efforts to tightly couple a focus on accomplishing authentic tasks with a focus on the underlying competencies needed to carry out the tasks.

See also: Apprenticeship Approach to Learning; Learning in a Sociocultural Perspective.

Bibliography

- Anderson, J. R., Reder, L. M., and Simon, H. A. (1996). Situated learning and education. *Educational Researcher* 25(4), 5–11.
- Beach, K. (1995). Activity as a mediator of sociocultural change and individual development: The case of school-work transition in Nepal. *Mind, Culture, and Activity* 2, 285–302.
- Boaler, J. (2002). *Experiencing School Mathematics: Traditional and Reform Approaches to Teaching and Their Impact on Student Learning*. (revised and expanded edition). Mahwah, NJ: Erlbaum.
- Bowers, J., Cobb, P., and McClain, K. (1999). The evolution of mathematical practices: A case study. *Cognition and Instruction* 17, 25–64.
- Brown, A. L. and Campione, J. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In Schauble, L. and Glaser, R. (eds.) *Innovations in Learning: New Environments for Education*, pp 289–325. Mahwah, NJ: Erlbaum.
- Brown, J. S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* 18(1), 32–42.
- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Cognition and Technology Group (1997). *The Jasper Project: Lessons in Curriculum, Instruction, Assessment, and Professional Development*. Mahwah, NJ: Erlbaum.
- Cohen, E. G. (1986). *Designing Groupwork*. New York: Teachers College Press.
- Collins, A., Brown, J. S., and Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In Resnick, L. B. (ed.) *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, pp 453–494. Hillsdale, NJ: Erlbaum.
- Engeström, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work* 14, 133–156.
- Engle, R. A. (2006). Framing interactions to foster generative learning: A situative explanation of transfer in a community of learners classroom. *Journal of the Learning Sciences* 15, 451–498.
- Goldman, S. and Moschkowich, J. (1995). *Environments for Collaborating Mathematically: The Middle-School Mathematics through Applications Project*. CSCL '95 Proceedings. Hillsdale, NJ: Erlbaum.
- Goodwin, C. (1996). Transparent vision. In Ochs, E., Schegloff, E. A., and Thompson, S. A. (eds.) *Interaction and Grammar*, pp 370–404. Cambridge: Cambridge University Press.
- Greeno, J. G. (1997). On claims that answer the wrong questions. *Educational Researcher* 26(1), 5–17.
- Greeno, J. G., Collins, A., and Resnick, L. B. (1996). Cognition and learning. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, ch. 2, pp 15–46. New York: Macmillan.
- Greeno, J. G., McDermott, R., Cole, K., *et al.* (1999). Research, reform, and aims in education: Modes of action in search of each other. In Lagemann, E. and Shulman, L. (eds.) *Issues in Education Research: Problems and Possibilities*, pp 299–335. San Francisco, CA: Jossey-Bass.
- Greeno, J. G., Smith, D. R., and Moore, J. L. (1993). Transfer of situated learning. In Detterman, D. K. and Sternberg, R. J. (eds.) *Transfer on Trial: Intelligence, Cognition, and Instruction*, pp 99–167. Norwood, NJ: Ablex.
- Greeno, J. G. and van de Sande, C. (2007). Perspectival understanding of conceptions and conceptual growth in interaction. *Educational Psychologist* 42, 9–23.
- Hatano, G. and Inagaki, K. (2003). When is conceptual change intentional? A cognitive-sociocultural view. In Sinatra, G. M. and Pintrich, P. R. (eds.) *Intentional Conceptual Change*, pp 377–406. Mahwah, NJ: Erlbaum.
- Hutchins, E. (1995a). *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive Science* 19, 265–288.
- Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal* 17, 29–64.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Nersessian, N. J., Kurz-Milcke, E., Newstetter, W. C., and Davies, J. (2003). Research laboratories as evolving distributed cognitive systems. In Alterman, R. and Kirsh, D. (eds.) *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*, pp 857–862. Mahwah, NJ: Erlbaum.

- Nowakowski, A., Campbell, R., Monson, D., *et al.* (1994). Goal-based scenarios: A new approach to professional education. *Educational Technology* 34(9), 3–32.
- Ochs, E., Gonzales, P., and Jacoby, S. (1996). “When I come down I’m in the domain state”: Grammar and graphic representation in the interpretive activity of physicists. In Ochs, E., Schegloff, E. A., and Thompson, S. A. (eds.) *Interaction and Grammar*, pp 328–369. Cambridge: Cambridge University Press.
- Phillips, S. U. (1972). Participant structures and communicative competence: Warm Springs children in community and classroom. In Cazden, C. B., John, V. P., and Hymes, D. (eds.) *Functions of Language in the Classroom*, pp 370–394. New York: Teachers College Press.
- Pickering, A. (1995). *The Mangle of Practice*. Chicago, IL: University of Chicago Press.
- Rogoff, B. (1990). *Apprenticeship in Thinking: Cognitive Development in Social Context*. New York: Oxford University Press.
- Rogoff, B. and Lave, J. (1984). *Everyday Cognition: Its Development in Social Context*. Cambridge, MA: Harvard University Press.
- Saxe, G. (1990). *Culture and Cognitive Development: Studies in Mathematical Understanding*. Hillsdale, NJ: Erlbaum.
- Schank, R. C., Fano, A., Bell, B., and Jona, M. (1994). The design of goal-based scenarios. *Journal of the Learning Sciences* 3(4), 305–346.
- Slavin, R. E. (1983). *Cooperative Learning*. New York: Longman.
- Smith, F. (1988). *Joining the Literacy Club*. Portsmouth, NH: Heinemann.
- Stenning, K., Greeno, J. G., Hall, R., Sommerfeld, M., and Wiebe, M. (2002). Coordinating mathematical with biological multiplication: Conceptual learning as the development of heterogeneous reasoning systems. In Baker, M., Brna, P., Stenning, K., and Tiberghien, A. (eds.) *The Role of Communication in Learning to Model*, pp 3–48. Mahwah, NJ: Erlbaum.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. New York: Cambridge University Press.

Further Reading

- McLellan, H. (1996). *Situated Learning Perspectives*. Englewood Cliffs, NJ: Educational Technology Publications.

The Neuroscience of Aging and Cognition

W J Hoyer and B Bessette-Symons, Syracuse University, Syracuse, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cognition – A collection of largely integrated mental operations that include thinking, perceiving, remembering, and learning.

Cognitive neuroscience – The interdisciplinary field of study that aims to describe and explain cognitive behavior in terms of neural mechanisms and structures.

Development – The outcomes of the interplay between the accumulation of experiences and time-related biological processes that affect behavior and physiology throughout the life span of individuals.

Dopamine – A substance that is manufactured in the cells of the substantia nigra and that travels in pathways connected to the frontal cortex. Reductions in dopamine level have detrimental consequences for cognitive and psychomotor functions.

Epigenesis – The premise that development throughout life is shaped by an in-common genetic plan that gradually unfolds with age. This genetically programmed unfolding of development occurs in a particular cultural or environmental context and that context can modify the outcome of the genetic program.

Executive control processes – The cognitive processes that are controlled by the individual and that serve to coordinate and regulate multiple tasks and actions.

Functional magnetic resonance imaging (fMRI) – A noninvasive technique for assessing brain function. fMRI uses the magnetic qualities of water molecules to evaluate the changing distribution of oxygenated to deoxygenated blood, an indirect measure of neuronal activity.

Neural plasticity – The extent to which change and effective adaptation is possible within the individual.

Neurogenesis – The production of new neurons originating from stem cells and progenitor cells.

Introduction

As the average age of the population in developed nations continues to increase, in large part due to increased

longevity, there is growing interest in trying to understand the cognitive and neurobiological changes that accompany aging. As individuals are living longer, it is important to explicate the potentials as well as the limits of cognitive functioning in older individuals. One general observation is that there are large interindividual differences in the amount and kinds of cognitive deficits that can be observed. That is, some individuals show only minor deficits in particular forms of remembering as they grow older, and continue to function effectively in everyday life. But some individuals show substantial declines in cognitive functioning, not associated with dementia or disease. Inevitable age-related changes in neurobiological processes contribute directly to the effectiveness of cognitive functions. Certainly some portion of interindividual differences in cognitive effectiveness corresponds directly to within-person changes in particular brain structures and neurobiological processes. Scientific knowledge about the specific relations between brain aging and cognitive aging has increased tremendously in recent years, in part due to advances in noninvasive techniques for assessing cognition–brain relations, such as structural and functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). Recent advances in neuroimaging allow identification of the brain regions associated with age-related changes in selected cognitive functions. Improved technologies of neuroimaging, such as diffusion tensor imaging (DTI), reveal relatively detailed descriptions of the relations between neural activity and cognitive function. For example, DTI studies show that the greatest amount of age-related alterations in white matter occur in the prefrontal cortex and the anterior corpus callosum. The kinds of cognitive tasks that show the largest age-related deficits depend on exactly these brain regions.

It is important to point out that the aging of brain–cognition relations is a consequence of the interplay between multiple etiologies and moderators that involve not only neural mechanisms and structural changes in the brain, but also the influences of genetic factors, environmental factors, and health and lifestyle experiences of the individual. We begin this article with a brief summary of brain aging. Then we briefly discuss current knowledge about age-related changes in cognition in healthy adults. Next, we discuss the relations between cognitive aging and brain aging and recent advances in the understanding of age-related changes in brain–cognition relations.

Brain Aging

Brain aging reflects the interplay of a variety of mechanisms. Time-related processes, individually and interactively, alter neuroanatomical structures and the biochemistry of neural functions. Structurally, MRI data using voxel-based morphometry show age-related shrinkage in the volume of the ventricles of the brain, and show that the degree of shrinkage varies somewhat for different brain regions. The ventricles are interconnected cavities within the brain that contain cerebrospinal fluid. As the ventricular cavities are interconnected, changes in the ventricular system can be considered to be a fair index of the general health of the brain. As the brain shrinks in size with aging, the ventricles expand to take up more space.

Epigenesis refers to the premise that development throughout life is shaped by a genetic plan. Age-related changes in brain structures unfold to a large extent according to an epigenetic plan at the level of cell physiology. However, the rate of genetically-programmed aging depends on a host of within-brain factors and external factors. Neural integrity and neural functions are impaired by the accumulation of beta amyloid and by inflammations within the environment of neurons. Within-brain environmental factors affect the vitality of neurons and their physiology. For example, either acute or chronic exposure to toxins can accelerate inflammation in the central nervous system and the pace of epigenesis (Campbell, 2004).

Neurogenesis refers to the production of new neurons originating from stem cells and progenitor cells. Stem cells are unprogrammed cells that divide and can change into cells with specialized functions. A host of factors (including individually controlled actions such as exercising and maintaining a healthy diet) can stimulate or deter neurogenesis. The hippocampus, a brain region that is centrally involved in associative learning and memory, is known to produce new neurons throughout life (e.g., Abrous *et al.*, 2005; Kemperman, 2006). A steady production of new neurons supports effective hippocampal function, whereas deterred neurogenesis eventually results in serious memory loss. There is some evidence to suggest that cognitive status is associated with increased neurogenesis and the strengthening of particular pathways in the hippocampus in line with functional demands (Hertzog *et al.*, 2009).

Intraindividual changes in cognition depend on the interplay between neural mechanisms and brain structures, genetic influences, and environmental influences. For example, recent findings show that the catechol-O-methyltransferase (COMT) genotype affects age-related variability in decision time in younger and older adults (Li *et al.*, 2008). The COMT gene affects the degradation of dopamine in the prefrontal cortex. Dopamine is manufactured in the cells of the substantia nigra and travels in the pathways connected to the frontal cortex.

Dopaminergic systems normally decline with aging across various brain regions. Age-related reductions in dopamine are known to have detrimental consequences for cognitive and psychomotor functions including the planning and execution of actions (Bäckman *et al.*, 2006). Compared to older individuals with a genetic predisposition for efficient dopamine signaling in the prefrontal cortex, older individuals with a genetic predisposition for less-efficient dopamine signaling in the prefrontal cortex show greater variability in performance and an overall lower rate of information processing when carrying out a simple two-choice perceptual-decision task (Li *et al.*, 2008).

A relation between the cholesterol ester transfer protein (CEPT) genotype and cognitive function has also been demonstrated (Barzilai *et al.*, 2006). A variant of the CEPT genotype is associated with preserved cognitive function in very old adults. This variant is associated with lower CEPT levels and a favorable lipoprotein profile. These and other findings of associations between particular genetic variants and polymorphisms and differences in relative efficiency of neuromodulation and cognitive function point to the necessity of a multicausal view in building a comprehensive understanding of developmental changes in brain–cognition relations.

Cognitive Declines in Healthy Aging

One way to examine the degree to which aging affects memory and other cognitive functions is to ask individuals to report the frequency of their memory failures in everyday situations. Older adults' self-reports about memory losses are fairly accurate, but it is also true that older adults tend to exaggerate and worry about their memory failures more so than younger individuals. Discrepancies between self-reports and actual memory performance may have to do with mistaken ideas about the functions and organization of human memory. Human memory is more than a filing cabinet for storing precise recollections of events and experiences and their circumstances. In this article, we approach the topic of memory and learning from the perspective that there are multiple forms of memory. This view is useful for the purposes of understanding cognitive declines with aging, since not all forms of memory show decline to the same extent as individuals grow older, and there are interindividual differences in the extent to which different forms of memory decline. Further, recent findings in cognitive science suggest that different forms of memory recruit and depend on somewhat different neural mechanisms.

Short-Term Memory and Long-Term Memory

More than 100 years ago, in 1890, the psychologist and philosopher William James suggested a distinction

between the conscious awareness of immediate events (primary memory) and the retrieval of events that had left consciousness (secondary memory). Modern findings substantiate James's early speculations about the differences between primary or short-term memory and secondary or long-term memory. Developmental studies show that span of short-term memory is only slightly affected by aging, whereas there are large age-related deficits in the amount and accuracy of retrieval of details about events and experiences that have been out of consciousness (even for just a few minutes).

Cognitive neuroscience research has shown that the extent to which there are age-related declines in long-term memory depends on the brain regions that are recruited for retrieval and the kind of information that is being retrieved. That is, age-related deficits in long-term memory are larger in tasks that require individuals to literally declare and consciously retrieve past events. Only minor age deficits are usually observed when memory for past experience is assessed indirectly or by non-declarative measures.

One way to study nondeclarative memory is to assess performance on priming tasks. In a priming task, individuals are instructed to identify or make judgments about stimuli that were (or were not) presented previously. Priming is demonstrated if previous exposure to items improves performance. Consider the following example of a two-phase priming experiment. In phase 1, participants study a list of words such as motel. In phase 2, the participants are presented with three-letter stems for the words that were on the study list (e.g., mot–) and for properly matched words that did not appear at study, and participants are instructed to complete the word stems with the letters that first occur that spell a word. Priming or nondeclarative memory is demonstrated if more stems are completed correctly for items presented in phase than for nonstudied items, and usually there are only minor age-related differences in such tasks.

The extent to which there are age-related differences in long-term memory tasks also depends on whether or not conscious recollection is required for retrieval (i.e., explicit memory). In an explicit memory task, an individual is instructed to deliberately recollect particular events. Typical measures of memory, such as recognition, recall, and cued recall, assess explicit memory. For example, in a cued-recall task, in contrast to a priming task, participants might be shown a list of words in phase 1, and then in phase 2 asked to complete three-letter word stems with the words from the study list. Usually, there are large age-related differences in explicit memory, and small differences in tasks that do not require explicit recall of information (i.e., implicit memory). Recent findings also suggest that the age-performance differences on implicit and explicit memory tasks are due to the recruitment of different brain regions.

Episodic and Semantic Memory

The extent to which there are age-related differences in long-term memory tasks also depends on whether the task taps episodic memory or semantic memory. Tasks that require episodic memory require conscious recollection of specific details of previous events. In contrast, semantic memory refers to retrieval of general information or acquired knowledge without the specific details about when and where the information was acquired. Semantic memories are not accompanied by the kinds of source details that mark episodic memories. Aging negatively affects episodic, but not semantic, memory.

Recollection and Familiarity

The extent to which there are age-related differences in long-term memory tasks also depends on whether the task at hand taps recollection or familiarity. Recollection refers to the retrieval of specific details about particular events or experiences and their contexts. In contrast, memories based on familiarity depend on nonspecific feelings of recognition or oldness about the event. When older adults do not recollect the specific details of an experience, they can instead rely on their familiarity with similar experiences. Age-related deficits are greater on tasks that require recollection than on tasks that allow familiarity.

Executive Functions and Working Memory

Substantial age-related declines are found on cognitive tasks that require individuals to manage or coordinate multiple tasks (i.e., executive function) and on tasks that require individuals to simultaneously process and store information in immediate memory (i.e., working memory). Executive functions refer to higher-order cognitive actions that are controlled by the individual and that serve the planning and regulation of complex tasks.

General Effects of Aging on Cognition

Behavioral and neurophysiological findings suggesting differential aging of specific types of memory must be tempered by data analyses that show that a large portion of the observed age changes in various types of cognition can be accounted for by general or core deficits. Across-the-board slowing in the speed of carrying out any mental operation seems to account for much of the decline in a wide variety of cognitive tasks. Moreover, age-related sensory deficits and age-related deficits in the ability to inhibit or suppress irrelevant information form core ingredients in many kinds of cognitive tasks, and contribute substantially to observed age differences. In addition, the effect sizes for the age difference on tests of episodic memory are consistently about one standard deviation

across a wide range of measures of episodic memory. Overall effect sizes are consistently somewhat smaller for episodic recognition.

Aging of Cognition–Brain Relations

The findings reported above derive from comparisons of different age groups (i.e., cross-sectional studies) or descriptions taken from groups of individuals followed across the adult years. Data from cross-sectional and longitudinal studies are consistent in showing age-related declines in long-term memory (declarative memory and associative memory), in the encoding and retrieval of events and experiences (i.e., episodic memory and recollection), in executive functions and the ability to simultaneously maintain and manipulate information in memory (i.e., working memory), and in the speed at which information can be processed and retrieved (see Hoyer and Verhaeghen, 2006). In recent meta-analyses, it is reported that the mean weighted-effect sizes are between -0.6 and -1.0 standard deviations in cross-sectional studies of episodic memory (using measures of recall), working memory, and processing speed (Old and Naveh-Benjamin, 2008; Verhaeghen and Salthouse, 1997). In contrast, the mean weighted-effect sizes for age-related differences in memory for emotional content, short-term memory span, nondeclarative memory (e.g., perceptual priming and implicit memory), and retrieval of semantic knowledge are much smaller or negligible.

The observed differences in the patterns of the effects of aging on measures of memory reflect the distinctive effects of aging on the status of the particular neural systems that serve particular behavioral and cognitive functions. The well-known findings of age-related deficits in episodic memory in contrast to smaller age-related decrements in nondeclarative forms of memory (e.g., repetition priming) are associated with age-related changes in the medial temporal area, encompassing the hippocampus and surrounding cortical regions (perirhinal, entorhinal, and para-hippocampal cortices). The extent to which the hippocampus supports the formation, organization, and retrieval of richly detailed memories and their distinctive contexts diminishes with aging.

In terms of age-related deficits in recollection, fMRI studies show different patterns of activation for recollection and familiarity across several brain regions. In the hippocampus, it was found that recollection was directly related to activation whereas familiarity was inversely related to activation. In the prefrontal cortex, it was found that recollection was related to an anterior medial region, and that familiarity was related to the anterior and dorsolateral prefrontal cortex. In the lateral parietal cortex, it was found that recollection was related to the temporal region, and that familiarity was related to a more superior region.

In the medial parietal region, it was found that recollection was related to the posterior cingulate, and that familiarity was related to the precuneus.

In terms of executive functions, neuroimaging data show reliable patterns of activation in three regions of the lateral frontal cortex in relation to cognitive activities that involve the executive functions of task coordination and the updating and maintenance of information in memory. Recent findings suggest that individual differences in executive cognitive function are largely determined by genetic factors (Friedman *et al.*, 2008). Of course, the person's prior experience with related tasks, and the person's abilities to initiate and deliberately deploy strategies that bring about effective learning and memory serve to reduce the magnitude of observed age-related deficits in executive tasks.

Neurocognitive Plasticity

Cognitive aging is dynamic and reflects adaptation and compensation in response to cognitive losses. Perhaps the most striking headlines in the cognitive neuroscience of aging are about discoveries of continued neural plasticity, continued functional reorganization and compensation, and continued neurogenesis throughout life in healthy adults. Neural plasticity is essential to effective adaptation to changes in internal and external conditions. The evidence suggestive of continued cortical reorganization and neurocognitive plasticity in healthy older adults does not contradict the evidence documenting the deleterious consequences of aging on neurobehavioral functions. Evidence suggestive of continued neural plasticity informs the extent to which there is potential for older adults to function effectively in society. An appreciation of the multicausality and multidimensionality of aging mechanisms is central to the understanding of boundaries of late-life cognitive adaptation.

Conclusions

The magnitude of the overall age-related deficit observed for episodic memory and related forms of learning, memory, and cognition is about one standard deviation. Individual differences in the extent to which deficits are observed in samples of older adults are due to a mix of factors, including ability status, genetically based differences in predisposition to neural inefficiencies and disease, and individual actions (exercise, diet, cognitive enrichments, and interventions) that can help to sustain effective cognitive functioning. In a long-term prospective study of cognitive change in a group of healthy octogenarians, for example, individuals who showed cognitive decline over a 13-year period had worse memory at

entry into the study and carried a particular genetic variant, the APOE4 allele (Howieson *et al.*, 2003).

Developmental changes and individual changes in neural plasticity represent a limit as well as potential. Neural plasticity undoubtedly decreases with advancing age in adulthood, but the extent to which there is continued neural plasticity indicates the potential of individuals to adapt to insidious losses associated with normal aging and to relatively abrupt insults associated with disease. The extent to which there is neural plasticity depends on actions of the individual that stimulate compensation or restoration of function and on cognitive interventions and educational programs that can promote plasticity.

See also: Neuroscience Bases of Learning; The Neuroscience of Reading.

Bibliography

- Abrous, D. N., Koehl, M., and Le Moal, M. (2005). Adult neurogenesis: From precursors to network and physiology. *Physiological Reviews* **85**, 523–569.
- Bäckman, L., Nyberg, L., Lindenberger, U., Li, S.-C., and Farde, L. (2006). The correlative triad among aging, dopamine, and cognition: Current status and future prospects. *Neuroscience and Biobehavioral Reviews* **30**, 791–807.
- Barzilai, N., Atzmon, G., Derby, C. A., Bauman, J. M., and Lipton, R. B. (2006). A genotype of exceptional longevity is associated with preservation of cognitive function. *Neurology* **67**, 2170–2175.
- Campbell, A. (2004). Inflammation, neurodegenerative diseases, and environmental exposures. *Annals of the New York Academy of Sciences* **1035**, 117–132.
- Friedman, N. P., Miyake, A., Young, S. E., *et al.* (2008). Individual differences in executive function are almost entirely genetic in origin. *Journal of Experimental Psychology: General* **137**, 201–225.
- Howieson, D. B., Camicioli, R., Quinn, J., *et al.* (2003). Natural history of cognitive decline in the oldest old. *Neurology* **60**, 1489–1494.
- Hoyer, W. J. and Verhaeghen, P. (2006). Memory aging. In Birren, J. E. and Schaie, K. W. (eds.) *Handbook of the Psychology of Aging*, 6th edn., pp 209–232. San Diego, CA: Elsevier.
- Kemperman, G. (2006). Adult neurogenesis. In Baltes, P. B., Reuter-Lorenz, P., and Rösler, F. (eds.) *Lifespan Development and the Brain: The Perspective of Biocultural Co-Constructivism*, pp 82–110. New York: Cambridge University Press.
- Li, S.-C., Lindenberger, U., Chicherio, C., *et al.* (2008). Catechol-O-Methyltransferase genotype affects decision time variability and drift rate in young and older adults: Evidence for neuromodulation of processing robustness and information accumulation. *Poster Presented at the Cognitive Aging Conference*. Atlanta, GA.
- Old, S. R. and Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging* **23**, 14–118.
- Verhaeghen, P. and Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and non-linear age effects and structural models. *Psychological Bulletin* **122**, 231–249.

Further Reading

- Albert, M. S. and Killiany, R. J. (2001). Age-related cognitive changes and brain-behavior relationships. In Birren, J. E. and Schaie, K. W. (eds.) *Handbook of the Psychology of Aging*, 5th edn. San Diego, CA: Academic Press.
- Bartzokis, G. (2001). Brain still developing in middle age. *Archives of General Psychiatry* **58**, 461–465.
- Buckner, R. L. (2005). Three principles for cognitive aging research: Multiple causes and sequelae, variance in expression and response, and the need for integrative theory. In Cabeza, R., Nyberg, L., and Park, D. (eds.) *Cognitive Neuroscience of Aging*, pp 267–285. New York: Oxford University Press.
- Colcombe, S., Erickson, K. I., Raz, N., *et al.* (2003a). Aerobic fitness reduces brain tissue loss in aging humans. *Journal of Gerontology: Medical Sciences* **58A**, 176–180.
- Colcombe, S. J. and Kramer, A. F. (2003b). Fitness effects on the cognitive function of older adults: A meta-analytic study. *Psychological Science* **14**, 125–130.
- Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., and Alkire, M. T. (2005). Structural brain variation, age, and response time. *Cognitive, Affective, and Behavioral Neuroscience* **5**, 246–251.
- Hertzog, C., Kramer, A. F., Wilson, R. S., and Lindenberger, U. (2009). Enrichment effects on adult cognitive development. *Psychological Science in the Public Interest* **9**, 1–65.
- Jones, S., Nyberg, L., Sandblom, J., *et al.* (2006). Cognitive and neural plasticity in aging: General and task-specific limitations. *Neuroscience and Biobehavioral Reviews* **30**, 864–871.
- Li, S.-C., Brehmer, Y., Shing, Y. L., Werkle-Bergner, M., and Lindenberger, U. (2006). Neuromodulation of associative and organizational plasticity across the life span: Empirical evidence and neurocomputational modeling. *Neuroscience and Biobehavioral Reviews* **30**, 775–790.
- Li, S.-C., Lindenberger, U., Hommel, B., Aschersleben, G., and Prinz, W. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life-span. *Psychological Science* **15**, 155–163.
- Lindenberger, U., Li, S.-C., and Bäckman, L. (2006). Delineating brain-behavior mappings across the life-span: Substantive and methodological advances in developmental neuroscience. *Neuroscience and Biobehavioral Reviews* **30**, 713–717.
- Pascual-Leone, A., Amedi, A., Ffregni, F., and Merabet, L. B. (2005). The plastic human brain cortex. *Annual Review of Neuroscience* **28**, 377–401.
- Posner, M. I. and Rothbart, M. K. (2005). Influencing brain networks: Implications for education. *Trends in Cognitive Sciences* **9**, 99–103.
- Raz, N. (2005). The aging brain observed *in vivo*: Differential changes and their modifiers. In Cabeza, R., Nyberg, L., and Park, D. (eds.) *Cognitive Neuroscience of Aging*, pp 19–57. New York: Oxford University Press.
- Raz, N. and Rodrigue, K. M. (2006). Differential aging of the brain: Patterns, cognitive correlates, and modifiers. *Neuroscience and Biobehavioral Reviews* **30**, 730–748.
- Small, S. A., Tsai, W. Y., DeLaPaz, R., Mayeux, R., and Stern, Y. (2002). Imaging hippocampal function across the human life span: Is memory decline normal or not? *Annals of Neurology* **51**, 290–295.
- Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory* **82**, 171–177.
- Willis, S. L., Tennstedt, S. L., Marsiske, M., *et al.* (2006). Long-term effects of cognitive training on everyday functional outcomes in older adults. *Journal of the American Medical Association* **296**, 2805–2814.

LEARNING AND COGNITION – LANGUAGE, LITERACY AND SUBJECT-RELATED

Contents

An Overview of Language and Literacy in Educational Settings

Bilingualism and Learning

Classroom Discourse and Student Learning

Early Writing

First Language Acquisition

Foreign Language Learning

Learning Science

Learning to Read

Mathematics Learning

Music Learning

Reading Comprehension: Reading For Learning

Second Language Learning

Writing: Advanced

An Overview of Language and Literacy in Educational Settings

V G Aukrust, University of Oslo, Oslo, Norway

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Genre or register – It refers to the constellation of lexical and grammatical features that characterize particular uses of language.

Lexicon or vocabulary – It is the body of words used in a particular language or in a particular sphere of activity.

Morphology – It is the way in which words are formed and related to each other. Morphology is therefore the subsystem of the smallest units of meaning in a language – words and parts of words.

Narrative – It is a sequence of language or words that describe a sequence of events.

Phonology – It refers to the sound system of a language which involves the rules for combining sounds.

Pragmatics – It refers to the rules of language use, and the way members of a community use language to accomplish their intentions and achieve their goals. It includes systems of commonly held conventions and constraints in interpersonal and institutional relations and purposes (e.g., indirect forms of expression in which the meaning is inferred rather than stated explicitly in polite conversation).

Semantics – It refers to the ways in which language conveys meaning.

Syntax – It is the system of phrases and clauses that are used to create sentences.

Word recognition – It refers to how phonological and visual–orthographic information is combined for the identification of individual words in a text.

Language plays an essential part in education. It is a prerequisite for most of the learning that takes place in schools. Language is also a set of complex cognitive and linguistic skills that schools foster. The role of language in learning has attracted educational researchers from various disciplines, such as linguistics, psychology, sociology, and anthropology. Moreover, across the disciplines, perspectives on the role of language in learning have been underpinned by a number of different theoretical contributions. In examining the role of language in education, this article asks the following questions:

1. How is language learned? In responding to this question, emphasis is placed on two major perspectives on language learning: the cognitive and the sociocultural perspectives.

2. What is the role of language in literacy acquisition? Most researchers agree that language and literacy are integrated systems, and that language is the primary system. But, more precisely, how does research based on the cognitive and the sociocultural perspectives relate language to literacy?
3. What is the role of language in subject-matter learning? Although most researchers agree on the importance of language for literacy learning, the role of language in subject-matter learning that is not reading is less obvious and much less examined. This article examines the impact of academic language on subject-matter learning.

Most children around the world learn more than one language. Several languages may be spoken at home and the child will learn to use them and switch between them, or the child will learn to use one language at home and perhaps another language (the majority language) in the community and school. Many children will also learn one or several foreign languages in school. The multilingual learning situation of most children is not addressed here. A large body of research has addressed the early phases of language learning, emphasizing continuities and discontinuities in the transition from being a child communicating through natural (rather than conventional) vocalizations that are interpreted by the child's caretakers, to the child that acquires and uses conventional vocabulary and syntactical structures, and follows pragmatic rules. This article addresses first-language learning from an age when language has been acquired in at least a rudimentary form.

How Is Language Learned?

Approaches to the study of language rest within a broader set of approaches related to human learning. Recent reviews of learning theories often distinguish between three broad views on learning: the behaviorist/empiricist view, the cognitive/rationalist view, and the situative/pragmatist–sociohistoric view (Greeno *et al.*, 1996). These are distinct traditions in educational theories and practices and are presented here as behavioral, cognitive, and sociocultural perspectives. The behaviorist perspective has played an important role in language research in the past; the cognitive and sociocultural perspectives are contemporary comprehensive accounts of language learning and they will be the focus of this article. There are significant variations within each of these last two broad perspectives in terms of their theoretical positions and the questions they have asked, as well as points of contact and overlapping issues between them.

Language learning theories vary and can be distinguished in the following ways:

1. Theories that concentrate on how the task of language learning is conceptualized, and by how they define what it means to know a language. Does it primarily have to do with developing the conceptual basis for language use, or being able to participate fluently in everyday linguistic practices?
2. Theories that relate to the nature/nurture dimension. How children acquire language has long been contended: on one side, Chomskian linguistic generative structuralists articulate a nativist position in which basic presuppositions are claims about the poverty of the stimulus. This means that however rich the input is, children can never extract the underlying structural rules of grammar from it. On the other side are psychologists who have a functional approach focused on social inputs and child strategies. Most theories of language development acknowledge both an inborn and a learning component, but differ in the emphasis they give to these elements.
3. Theories that pay attention to the individual versus the community or culture in which the child grows up.
4. Theories that look at different aspects of language. Language is a complex system with subsystems such as phonology, morphology, semantics, lexicons, and pragmatics. Are phonological skills that support children's word processing basic in language learning, or should the language researcher turn to how children learn to express their intentions in language, acquire meaning, and learn the pragmatics of language use in their speech communities?
5. Theories that examine how the relationship between language and literacy is conceptualized.

The Behaviorist Perspective on Language Learning

There are many different hypotheses concerning language acquisition that fall under the general heading of behaviorism. All these approaches share a common focus on the observable aspects of language behavior. Learning is seen as the formation, strengthening, and adjustment of associations between stimuli and responses. Behaviorist perspectives on language learning following from Skinner explain language through two major types of conditioning – classical and operant – and through imitation and shaping (Bohannon and Warren-Leubecker, 2001). Language learning takes place when stimuli in the environment become associated with internal responses, which then become associated with overt behavior through the application of classical and operant conditioning, imitation, and shaping. Operant conditioning is the learning principle that explains productive speech. Simply put, behavior that is rewarded tends to be repeated by the learner. Behaviorists assume that caregivers reward children when their speech, or parts of it, most closely approximate adult speech. Children's word combinations

are assumed to be acquired in much the same way as single words, through shaping, imitation training, and rewards. The hypothesis is that smaller units must be mastered as a prerequisite for more complex units. Therefore analysis of complex tasks into learning hierarchies has been used in designing instruction sequences.

The behaviorist perspective is on the very end of the nature/nurture continuum, emphasizing the role of experience in language learning. Behaviorists focus on how the small units of language are acquired, and how more complex learning outcomes build on these. The behaviorists' view has been criticized for relying too much on simple principles to explain language development. The idea that smaller behavioral units must be mastered as a prerequisite for more complex units has also been questioned. Moreover, caretakers have not been found to consistently provide rewards for grammatically correct utterances or punishment for incorrect ones. Rather, they respond to the content of the utterance. In spite of its significance in the history of language-learning research, behavioral views play a minor role in contemporary educational research, compared to cognitive and sociocultural perspectives which we now turn to.

The Cognitive Perspective on Language Learning

The process of language acquisition by children remains one of the critical issues for contemporary cognitive theory. Within the cognitive perspective, learning is considered as change in conceptual and cognitive structures. Various theoretical contributions – connectionism, constructivism, and the sociocognitive approach – have contributed to the cognitive perspective on language. In this article, the sociocognitive contribution is included in the group of cognitive studies because of its commitment to study individuals, even though sociocognitive research has incorporated sociocultural ideas of the mediated mind, and its conceptual contributions such as intention reading, perspective taking, and communicative collaboration have impacted sociocultural views on language learning. The three theoretical examples of the cognitive perspective in this article share assumptions regarding the importance of the individual as the processor of input from the environment. They differ in their views on the relative role of innate conditions versus environmental input in language learning, and in the relative emphasis on individual versus social and cultural dimensions in explaining learning.

Connectionism

Connectionism emphasizes constraints on the processing of information and the experience-driven reorganization of lower-level behavior into more complex patterns.

Connectionists argue against the need to postulate abstract rule systems to explain emerging regularities in language use (cf. Chomsky and the generative structuralists) and emphasize the role of experience or input. The question is how a system like the brain, given streams of input from the structured world, comes to discover that structure. For example, when children start to produce the past tense of verbs they initially get them correct (e.g., she went), and then they begin to make errors (e.g., she goed). This has been cited as evidence for the theory that linguistic development proceeds in stages that reflect the children's acquisition of rules (first learn the word, e.g., went, then acquire the rule about the past tense, and finally discover that rules have exceptions – the two-route way to verb learning). A connectionist simulation of children acquiring knowledge of the use of the past tense showed that they used a mixture of correct and erroneous past tenses in a way that was not stage-like (the one-route way to verb learning).

Constructivist approaches

The constructivist model emphasizes the child's ability to actively organize and construct knowledge. Piaget's influence on conceptual learning as the basis for language learning has been considerable. Piaget focused on the conceptual structures that children develop and which their language presupposes and builds on. Generally, the constructivist approach following from Piaget stresses that language is only one of many complex cognitive skills that children acquire.

Sociocognitive approaches

Although the cognitive perspective on learning has traditionally focused its attention on the individual's learning processes, such as the development of cognitive structures, and has regarded social interaction as only one factor influencing learning, more attention has been paid recently to the social context in which learning occurs. Tomasello (1999) and Nelson (1996) have contributed to these recent developments within the cognitive perspective.

Tomasello emphasizes on children's sociocognitive skills that enable them to participate with others in joint attention, and learn from others, to understand the communicative intentions of others, to take the perspective of others in interpreting and using language, and to collaborate communicatively. Children apply sociocognitive skills to acquire and use language – skills of joint attention, intention reading, perspective taking, and communicative collaboration. Tomasello presents a two-way relationship between children's social cognition and their language. Their sociocognitive skills support their language acquisition, but engaging in linguistic communication leads them to create new sociocognitive skills such as taking the perspective of the other. Cultures shape the development of children,

especially through the intentional instruction of others. According to Tomasello, grammar, regardless of its language, is a perspective-taking device, learned in part by the human capacity to see experience from someone else's perspective.

Similarly, Nelson's sociocognitive theory emphasizes the relationships between cognitive development and language. Language is not separable from cognitive development, and cognitive development cannot be detached from language. Nelson takes as a starting point that the primary cognitive task of the child is to make sense of his or her situated place in the world in order to participate in its activities. She argues that, for the child, knowledge of the object world is embedded within knowledge of the sociocultural world, with the basic task for the child being to predict activities. She introduced the idea of participatory interactions, referring to children taking part in activities without fully understanding them. Through participation they learn their parts and acquire knowledge about them. The meaning of words and language forms such as narrative is acquired through participation in social discourse and it requires extensive practice. When internal language representations come into play, thinking in language emerges.

The Sociocultural Perspective on Language Learning

Sociocultural views of language learning emphasize the social nature of learning (also known as situative, pragmatist, sociohistoric views, and as social constructivism and cultural psychology). This article foregrounds three contributions that are related: the Vygotskian tradition, sociolinguistic research with an emphasis on language socialization, and ethnographic studies of discursive practices in homes and schools. These three sociocultural branches represent a family of related frameworks. They share the approach that the unit for studying language is the socially situated individual. In this group of approaches, language learning is based on children's repeated participation in everyday activities with other more-competent participants. Through extended engagement in activities with others who are more knowledgeable or expert, children transform the specific means for realizing these activities into individual skills.

The Vygotskian tradition

According to Wertsch (1985), there are three major lines of thought in Vygotskian sociocultural theory. First, to understand a psychological phenomenon, such as language or literacy learning, it is necessary to understand the origin of the phenomenon and the processes by which it is acquired. Vygotsky referred to it as genetic or developmental analysis. Central to this analysis is the conception that any competence appears on the scene twice, first

as an intermental capacity between the language-learning child and more competent others, and subsequently through a process of internalization as an intramental capacity in the child. Second, for Vygotsky, mind is social in nature and constituted through language-based social interactions with others. Third, human action is mediated by signs and tools – primarily psychological tools such as language. Researchers working with the Vygotskian idea of internalization have recently included the Bakhtinian notion of appropriation to emphasize the active making-of-one's own that is part of the process of learning language (Aukrust, 2001).

The child's developing capacities are seen as internalization or appropriation of the society's practices via the support of adults or more knowledgeable peers. Through guided participation, children appropriate cultural tools such as language, literacy practices, and social rituals. Children's participation and subsequent development are mediated by the symbolic tools and resources that constitute their learning environments. The metaphorical term scaffolding has become common for this guided assistance that supports the child's growing linguistic participation in everyday routines and formats. Scaffolded assistance in the zone of proximal development supports children's development into the practices deemed important in their culture. Studies of play (Bruner, 1983) and book reading (Snow and Goldfield, 1983) have shown that children's early language acquisition takes place within structured and repeated scaffolded interactions in familiar routines.

Sociolinguistic approaches emphasizing language socialization

Language socialization studies emphasize the complex social norms that govern the use of specific language constructions. These studies clearly define the implicit and explicit socialization practices involved in the child's developing language.

The term language socialization refers to the interactional processes through which a child develops the competence required for participation in the social life of a particular community, including routine cultural practices, such as language and literacy activities. Language socialization research builds on the view that acquiring a language is part of a much larger process of becoming a member of society. The language socialization paradigm is concerned with two phenomena: how children are socialized to use language, and how they are socialized through the use of language.

Ochs and Schieffelin (1995) argue that language socialization accounts for children's grammatical development in terms of the indexical meanings of grammatical forms. Children are viewed as tuned into certain lexical meanings of grammatical forms, which link those forms to, for

example, the social identities of interlocutors. They may use a form they do not hear often because it is indexically appropriate to use it. A language socialization approach relates children's use and understanding of linguistic forms to how information is linguistically presented within and across socially recognized situations.

Ethnographic studies

Ethnographic studies emphasize that language always takes place in a specific social and cultural setting. Acquisition of language and literacy is conceptualized as a gradual process of adopting local cultural practices. This group of studies, closely related to the language socialization approach, was developed originally as part of linguistic anthropology. This research demonstrated the cultural specificity of language and literacy socialization practices and their developmental consequences in the transition from home-based to school-based activities (cf. Heath's study below).

During the last two decades, ethnographic studies have examined the characteristics of language learning in the multiparty interaction of classrooms (for review, see Cazden, 2001; Wolf *et al.*, 2006). Many classroom activities are created through classroom discourse, and therefore its role is important in the creation of learning environments and thus of individual learning. This group of studies has been concerned with children's pragmatic learning – their learning of discourse participation. The most common participant structure in classrooms is described as initiative–response–feedback (IRF), initiative–response–evaluation (IRE), triadic dialog, and recitative conversation. In its characteristic form, this participant structure comprises three moves: initiative (I), usually in the form of a teacher's question which often relates to known information; response (R) where the student answers the question; and feedback (F) or evaluation (E), where the teacher follows up or comments on the student's response. A number of international studies have documented that classroom conversations displaying the IRF structure serve to control the students and make them passive within an asymmetrical balance of power: it is the teacher who controls the theme, gives the questions to the students, and the opportunities for them to answer, and it is the teacher who evaluates the answers. However, several recent studies have also emphasized the learning potential in the IRF sequence and have concluded that the relationship between the students' responses and the teachers' feedback is a decisive factor in establishing shared background knowledge. Educators have also explored more complex classroom discourse under terms such as reciprocal teaching (Brown *et al.*, 1996) and dialogic enquiry (Wells, 1999), which refer to discursive activity in classrooms that permits the co-construction of meaning between teachers and students.

Comparing the Cognitive and Sociocultural Perspectives

Cognitive and sociocultural perspectives differ in the ways they conceptualize language learning, in their relative emphasis on innate versus learned aspects of language learning, in the role of the individual versus the collective in language learning, and in which aspects of language they are focused.

How is the task of language learning conceptualized?

In the cognitive perspective, language is a system that must be acquired in terms of a processing system. For the majority of theorists working within the cognitive perspective, learning is a process of acquiring new representations that gradually take on a more decontextualized quality. This is the task the individual child faces when acquiring language, but there is considerable variation among researchers working within the cognitive perspective when it comes to how this task is conceptualized – from the constructivist tradition following Piaget, who addresses individual construction, to the culturally embedded approaches of Tomasello and Nelson.

For the sociocultural researcher, language is a system of cultural practices into which the child is gradually socialized. Language learning is not so much a process of incorporating information or individual construction, but a process of socialization in which the novice gradually participates more and engages in cultural practices. These two main perspectives therefore differ in their view of what it is to know a language (concepts, schemas, networks, versus participation in social activity). The sociocultural perspective tends to disregard acquisition of decontextualized knowledge and emphasizes increasingly competent and fluent participation in cultural routines.

What is the relative emphasis of innate versus acquired aspects of language learning?

Although the innate versus acquired dimension of language development has historically received much attention in language development research, neither the cognitive nor the sociocultural perspectives have a one-sided stance in this question. In the extent to which these issues are discussed in the cognitive and sociocultural research literature, the argument seems to be that young children are biologically predisposed to interact with their environments, but that the effects of language input, and the language structure and culture on children's developing language systems, suggest an enormous susceptibility on the part of the language learners to the effects of input.

Studies applying a cognitive perspective tend to expect that relatively little higher-level knowledge is innate. They tend to emphasize how lower-level learning mechanisms

lead to higher-level concepts and rule-governed behavior. The connectionists propose a combined nativist and input-sensitive language-learning system; similarly, the sociocognitivists emphasize that children are biologically prepared for culture (Tomasello), but that their participation in culture supports acquisition of new cognitive and linguistic skills. They start with the premise that language serves cognitive and social functions of human life, and claim that the structures of language emerge from its functions. Generally, socioculturalists do not speak about input as such, but focus on the culturally or institutionally embedded interactions in which children participate. The analytic focus is not on the child receiving input to process, but on ways of participating fluently and competently in significant cultural routines surrounding the child.

What is the role of the individual versus the social or collective in language learning?

The primary analytic level of the cognitive perspective is the individual processing information. The sociocultural perspective often distinguishes between three analytic levels: individual development, social interaction, and the cultural activities in which both take place. Sociocultural studies commonly view individual development as changed ways of participation in social interaction embedded in cultural activities. The individual level is therefore not the primary focus when socioculturalists explain individual change or learning. Although the sociocognitive approach is concerned with how culture enters the young mind and with how thinking becomes culturally mediated when it emerges in language, its focus is on the child constructing this knowledge.

As there is variation along the individual–collective continuum between the cognitive and the sociocultural perspectives, there is also variation within each of these perspectives regarding the conceptualization of the individual as active versus fairly passive and about the focus of the social interaction.

The cognitive perspective on learning language consists of a number of approaches that differ in the importance they attribute to the individual versus the social context. At one end of the continuum is Piagetian constructivism, which emphasizes the individual's mental processes in the construction of meaning. At the other end of the continuum is the sociocognitive approach, which pays more attention to the social context in which the individual is constructing knowledge. Piaget highlighted the child as an active constructor of knowledge and conceptual structures, whereas connectionism tends to view the child as a receiver of information. Also the sociocognitive approach emphasizes the actively constructing child, but unlike the Piagetian tradition, the research focus is on the child making use of culturally embedded meanings.

While the cognitive perspective pays attention to the individual and distinguishes between individual children and the cultures into which they are born, the sociocultural perspective turns primary attention to social/institutional/cultural opportunities and constraints for learning, and emphasizes the role of social experience and cultural tools in a child's developing language skills. Learning is a matter of how people transform through participation in the activities of their communities. Within the sociocultural perspective, studies differ when it comes to their specific focus on social/cultural analysis, whether transformed participation is studied within dyadic interaction (often characteristic of many Vygotskian-based language learning studies), within cultural activities (Heath, 1983; Ochs and Schieffelin, 1995), or power relationships (many ethnographic studies of classroom interaction).

Which aspects of language learning are focused on most?

The cognitive and sociocultural perspectives differ in their orientation to the aspects of language that are studied. The former perspective has a long tradition of interest in grammar, whereas the latter has primarily addressed social meanings, genres, and the mediating role of language. For example, cognitive research on word learning tends to focus on the number of semantic features the child knows, whereas socioculturalists emphasize the child's usage of a word across various contexts and performance rather than generalized knowledge. It is worth noting that sociocognitive theory has offered significant work on language-mediated narrative thinking and on children's acquisition of socially and culturally embedded word meanings, while research within the sociocultural perspective has examined children's functional learning of grammar, from the starting point that grammatical forms are linked to social meaning and social identities.

What is the Role of Language in Literacy Acquisition?

Literacy is not acquired in the same way as language and needs to be taught. There is consensus in the field of literacy research that language plays a major role in reading and writing, and that these activities are fundamentally linguistic activities. But what is the role of language in literacy? We now turn to how this relationship is conceptualized within the cognitive and sociocultural perspective. Various bidirectional relationships exist between oral language and literacy, but the impact of reading on language is not addressed here.

Language in Literacy: The Cognitive Perspective

Researchers who study literacy from cognitive and psycholinguistic orientations generally agree that literacy

draws upon multiple interrelated skills, including oral language, phonological awareness, knowledge of the graphic features of print, and an understanding of how sounds map into print. Cognitive research, particularly within psycholinguistics, has identified components of skills that need to be mastered in order to become a qualified reader. Within the cognitive perspective, literacy has been analyzed as a combination of skills to encode information from text into mental representations of letters and words, to recognize the words and activate representations of their meanings, to combine representations of words into phrases and sentences, and to form representations of the propositions they cover.

Early predictors of reading comprehension are word recognition and oral language skills. Two processes are involved in word recognition: a visual process and a phonological decoding process, which concerns the correspondences between printed letters and the sounds of the language, especially phonemes (the small sound units within spoken words). Research in alphabetic reading has developed the consensus that phonological decoding is a routine part of skilled word recognition. How the phonological and visual-orthographic information combines for the identification of individual words has been the focus of much research within the cognitive tradition, reflecting theoretical debates about how to conceptualize the cognitive mechanisms of word recognition. In addition to supporting word recognition, phonological processing supports cognitive processes such as comprehension and memory of the text. It is well documented that word recognition is an important source of individual variation in reading. In word recognition, accuracy and fluency are crucial for processing meaning. Fluency enables readers to access and apply relevant higher thinking skills for comprehension.

Moreover, reading comprehension is predicted by oral language which again is comprised of multiple skills. As noted, phonological skills are precursors to literacy. The importance of phonological awareness (the ability to reflect upon and manipulate the sound structure of spoken language) is well documented. Moreover, vocabulary has long been known to be predictive of later-reading comprehension. One of the most robust findings in the field of literacy research is actually the high correlation between vocabulary and reading comprehension. Also, variability in syntax skills has been found to be related to reading. Finally, knowledge of discourse types (expository text, narrative) predicts comprehension. Good comprehenders tend to have a more advanced understanding of various discourse types, which helps them interact with texts strategically and construct meaning successfully. Two types of extended discourse – explanations and narratives – have received much attention in literacy research, and children's exposure to such discourse has been linked to literacy-related outcomes. Over all and in sum, it is widely agreed that specific and multiple oral-language skills prepare children for comprehension.

Phonological processing and oral-language skills influence different aspects of reading acquisition at different times. The association between oral language – particularly vocabulary – and reading has been found to be greater for the later years, from fourth grade on. This does not mean that oral-language skills have no importance for early reading. Oral-language skills seem to have an indirect relationship with phonological processing as early as kindergarten and across the early years of schooling.

Language in Literacy: The Sociocultural Perspective

Within the sociocultural perspective, literacy practices are considered part of social practices, the most fundamental unit of analysis of human action within the field. Sociocultural researchers often adopt a broad view of literacy and have also proposed multiple literacies, such as digital literacy and science literacy.

Language is considered the most significant resource for creating and reproducing meaning and is an essential feature of social practices. Human knowledge is stored in discourse, systematically organized systems of language in which meaning is produced according to certain semiotic principles. Socioculturalists' emphasis on language and literacy recognizes the role of mediational means in understanding human collective and individual action.

Two important contributions to the role of language in literacy within the sociocultural perspective are considered: Heath's anthropological study of continuities and discontinuities in ways with words in home and school, and the new literacy studies (Gee, 1996).

Heath (1983) described how children in different communities in the USA learned their community's way of using language in everyday social practices. She suggested that each community had specific ways of socializing members. For example, Heath found that children from Trackton, a mostly African-American community, and Roadville, a mostly white community, exhibited very different story-telling behaviors. Heath documented how children learned to use language and how their use of language established their identity, roles, and relationships among family and friends. Following these children into their schools, she documented how their distinct ways of learning language affected their integration into academic life. Most critical to academic success were certain ritualized uses of language, such as the assignment of labels to objects and responses to questions about events already known to the questioner. She also described how children from some communities were encouraged to look at letters and books at home. The children developed a familiarity with texts that supported their learning in school. The study reports on continuities and discontinuities in literacy practices at home and school, and underlines the many relationships between language,

literacy, and academic learning, beneficial for children who experienced continuity in ways with words between home and school and detrimental for those who did not. The study finds support in more recent ethnographic studies, which have concluded that the failure of many children from low-income homes in reading results from a mismatch between home and community patterns of language and literacy use, and the uses of language and literacy typically expected in schools.

The new literacy studies took as a starting point that reading and writing only makes sense when studied in the context of social and cultural practices. This group of studies is closely related to work on situated cognition that has argued that knowledge does not reside solely in individuals' heads, but rather is distributed across the social practices (including language practices) and the various tools, technologies, and semiotic systems that a given community of practice uses in order to carry out its activities. Learning to read and write is to become a participant in literate social practices, and further learning is a matter of changing patterns of participation in literacy practices. The new literacy studies pays attention to different styles of oral language use and the various types of literacy that are tied to specific domains, such as science in classrooms, or to specific sociocultural groups. Cognition is studied in terms of how thinking is mediated by the physical and social environment, various semiotic systems, and different sorts of tools.

The Role of Language in Literacy – The Two Perspectives Compared

In the cognitive perspective, language plays an important role in reading comprehension. Various oral skills have been examined, and many of them, such as vocabulary, phonological processing, syntactic skills, and discourse knowledge, showed a relationship to reading outcomes. Though there is much developmental continuity between language and literacy, and though language is a strong predictor of later literacy, there are also discontinuities and differences, the most important being that literacy needs to be taught to be acquired.

Research based on the sociocultural perspective does not apply a skills approach to literacy. Literacy is not considered an individual accomplishment, but a social practice to which communities and schools socialize children. Texts and other cultural tools mediate the ways people communicate and think in a culture. The formation of literate identities and the acquisition of academic social languages are important achievements in literacy education. Moreover, the sociocultural perspective does not differentiate sharply between language, artifacts such as the alphabetic system and texts, and social practices. Artifacts are conceived as forms of human knowledge and as mediating resources that structure social practices as well as human thinking.

This article has culled out some differences in how researchers within the cognitive and sociocultural fields respond to questions of the role of language in literacy. When it comes to reading programs, there are many examples of researchers who draw from both the individual skills and the social practices perspective. An example is reciprocal teaching which is partly Piaget based and partly Vygotsky based, and is a cognitive apprenticeship model for teaching reading comprehension within communities of learners (Brown *et al.*, 1996). Teachers and students take turns leading a discussion that includes four types of strategic activities: predicting, questioning (making up a question on the main idea in the text), summarizing, and clarifying.

The Role of Language in Learning Subject Matter: Academic Language

While much research has examined relationships between language and literacy, less is known about the impact of language on learning subject matter in general, outside the mediating role of reading for learning. This final section turns to the role of academic language in children's discipline learning. Both the cognitive and the sociocultural perspectives have paid attention to academic language, the former with a focus on the particular language skills required to cope with textbooks and classroom discourse, the latter related to genres or registers as mediating tools for learning. While the term academic language represents a fairly recent attempt at identifying those qualities of language that support acquisition of content-area knowledge, a variety of terms have been used to capture the same language qualities, and these are reviewed briefly first:

1. *Decontextualized language.* Decontextualized language is talk that goes beyond the here and now, and relies on language to convey information about other times and places. The term was frequently used in child language research from a cognitive and psycholinguistic perspective in the 1980s and 1990s. Offering oral definitions is, for example, a discourse task in which children need to rely on their decontextualized language skills. The term has also been used to point out similarities in spoken and written language. Oral stories, in which the speaker does not presuppose shared knowledge with the interlocutors, take on some of the characteristics of written language: they have a relatively complete message, adequate explicit references, and an assumed audience that does not share much background information with the writer. These characteristics are all examples of decontextualized language skills. Advocates of the sociocultural perspective have argued that all language use is contextualized in social practices. They have also suggested applying the term recontextualized to messages that do not presuppose shared background knowledge.

2. *Extended discourse.* Extended discourse is the term suggested in child language research to cover much of the same meaning as decontextualized language. The term extended has two meanings. First, it emphasizes sustained use of language, such as in conversations, as opposed to simple directives – that is, talk that extends over several utterances or turns. Second, it refers to talk that extends the here and now. Narratives and explanations are examples of extended discourse that children participate in and that have been found to support their development of oral language and reading. Making convincing arguments and providing formal definitions are also examples of extended discourse.
3. *Cognitive Academic Language Proficiency (CALP).* The distinction between cognitive academic language proficiency (CALP) and basic interpersonal communicative skills (BICS) was proposed by Cummins (1979) to draw attention to the time required by immigrant children to acquire second-language conversational fluency (BICS) compared with grade-appropriate academic proficiency which was needed for school learning. Conversational fluency was often acquired within about 2 years of exposure to the second language, whereas academic aspects of the second language required a much lengthier process to be learned.
4. *Accountability talk.* The term accountability talk is used to point out various features of types of talk that support reading comprehension. These three features are:
 - *Accountability to accurate knowledge.* Participants make use of specific and accurate knowledge, recognize the kind of knowledge required to address a topic and appropriate evidence for claims and arguments, and show commitment to getting it right.
 - *Accountability to rigorous thinking.* Participants use strategies to present arguments and to challenge each other's reasoning, synthesize several sources of information, construct explanations, formulate hypotheses, and apply generally accepted standards of reasoning.
 - *Accountability to the learning community.* Participants make efforts to ensure that all participants understand the ideas and positions shared, elaborate and build on each other's ideas and work to clarify or expand propositions, and ask each other questions aimed at clarifying propositions (Cazden, 2001; Wolf *et al.*, 2006).

Academic Language in Cognitive versus Sociocultural Perspectives

Academic language in the cognitive/psycholinguistic perspective

Academic language has various typical features: the interpersonal stance is detached or authoritative, the information

load is dense, the organization of information involves explicit marking of text structures (metadiscourse), and at the lexical level academic language is diverse and precise, including terms that are discipline-specific as well as terms used academically across disciplines. Moreover, these linguistic features must be combined with additional cognitive achievements: genre mastery, mastery of argumentative strategies, and content-area knowledge (Snow and Uccelli, 2009). Academic language, like all linguistic communication, involves challenges at the level of self-representation, of representing a message, and of constructing discourse.

Academic language in the sociocultural perspective

Within this perspective, academic language is integrated with social identities and social practices and is at the core of school-based literacy (Gee, 2004). Academic language is part of social languages which again are defined as ways of using language so as to enact a particular socially situated identity and carry out a particular socially situated activity. School success is based on the ability to cope with academic language, which is different from everyday language in face-to-face conversation. Science has, for example, been conceptualized as a special-purpose language. To acquire an academic social language, students must be willing to see the acquisition of the academic social language as a gain (gaining traits, categorizations, etc.) and also accept certain losses (attitudes, interests, values left out by the speaker/writer).

What is academic language: Summary

Even in the early grades, children are expected to learn new information from content-area texts, so failure to understand the academic language of those texts can hinder their access to new information. Academic language is often defined as the language used in textbooks and in schools, that is, academic language is defined by the context in which it appears. In addition, academic language has special linguistic features, puts specific demands on self-representation (the cognitive/psycholinguistic perspective), and is integrated with the enactment of social identities (the sociocultural perspective).

Concluding Comments

The behaviorist perspective was historically a main line of development in the psychology of learning and played a significant role in the study of language learning in the 1950s and 1960s. The cognitive perspective became influential in research on learning and cognition in the 1970s, while recent developments like connectionism and socio-cognitive research continue to vitalize the study of language learning. The sociocultural perspective is the youngest of the three, but defines an area of educational

research with vital activity during the last decades. In their broad review of the field of learning and cognition, Greeno *et al.* (1996) discuss potential future developments and relationships between these theoretical perspectives. One possibility is that the different views will live side by side, complementing each other and analyzing processes of learning at different levels of aggregation. Another possibility is a somewhat more competitive relation among the perspectives, in which one perspective can provide a kind of synthesis of the others. One salient theoretical question for the study of how language is learned and how it impacts literacy and academic school learning will be the continuing clarification of relations among research perspectives.

See also: Apprenticeship Approach to Learning; Attention in Cognition and Early Learning; Bilingualism and Learning; Constructivism and Learning; First Language Acquisition; Learning in a Sociocultural Perspective; Learning Through Narrative Socialization; Learning to Read; Piaget: Recent Work; Reading Comprehension: Reading For Learning; Reasoning and Explanatory Talk: Learning in Everyday Settings; Second Language Learning; Vygotsky and Recent Developments.

Bibliography

- Aukrust, V. G. (2001). Agency and appropriation of voice: Cultural differences in parental ideas about young children's talk. *Human Development* **44**, 235–249.
- Bohannon, J. N. and Warren-Leubecker, A. (2001). Theoretical approaches to language acquisition. In Gleason, J. B. (ed.) *The Development of Language*, pp 254–314. Columbus, OH: Merrill.
- Brown, A., Metz, K., and Campione, J. (1996). Social interaction and individual understanding in a community of learners: The influence of Piaget and Vygotsky. In Tryphon, A. and Vonèche, J. (eds.) *Piaget–Vygotsky: The Social Genesis of Thought*, pp 145–171. Hove: Psychology Press.
- Bruner, J. S. (1983). *Child's Talk: Learning to Use Language*. New York: W. W. Norton.
- Cazden, C. (2001). *Classroom Discourse: The Language of Teaching and Learning*. Portsmouth: Heinemann.
- Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism* **19**, 121–129.
- Gee, J. P. (1996). *Social Linguistics and Literacies: Ideology in Discourses*. London: Taylor and Francis.
- Gee, J. P. (2004). Language in the science classroom: Academic social languages as the heart of school-based literacy. In Yerrick, R. (ed.) *Establishing Scientific Classroom Discourse Communities: Multiple Voices of Teaching and Learning Research*, pp 19–38. Mahwah, NJ: Erlbaum.
- Greeno, J. G., Collins, A., and Resnick, L. (1996). Cognition and learning. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 15–46. New York: Macmillan Library Reference USA.
- Heath, S. B. (1983). *Ways with Words: Language, Life, and Work in Communities and Classrooms*. Cambridge: Cambridge University Press.
- Nelson, K. (1996). *Language in Cognitive Development: Emergence of the Mediated Mind*. Cambridge: Cambridge University Press.
- Ochs, E. and Schieffelin, B. (1995). The impact of language socialization on grammatical development. In Fletcher, P. and MacWhinney, B. (eds.) *The Handbook of Child Language*, pp 73–94. Oxford: Blackwell.
- Snow, C. E. and Goldfield, B. A. (1983). Turn the page please: Situation-specific language acquisition. *Journal of Child Language* **10**, 551–569.
- Snow, C. E. and Uccelli, P. (2009). The challenge of academic language. In Olson, D. and Torrance, N. (eds.) *The Cambridge Handbook of Literacy*, pp 112–113. Cambridge: Cambridge University Press.
- Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.
- Wells, G. (1999). *Dialogic Inquiry. Towards a Sociocultural Practice and Theory of Education*. Cambridge: Cambridge University Press.
- Wertsch, J. (1985). *Vygotsky and the Social Formation of Mind*. Cambridge, MA: Harvard University Press.
- Wolf, M., Crosson, A. C., and Resnick, L. (2006). *Accountable Talk in Reading Comprehension Instruction*. Los Angeles, CA: Center for the Study of Evaluation, University of California.
- Bakhtin, M. (1986). *Speech Genres and other Late Essays*. Austin, TX: University of Texas Press.
- Chomsky, N. (1957). *Syntactic Structures*. Hague: Mouton.
- Dickinson, D. and McCabe, A. (2001). Bringing it all together: The multiple origins, skills, and environmental supports of early literacy. *Learning Disabilities Research and Practice* **16**, 186–202.
- Griffin, T., Hemphill, L., Camp, L., and Palmer, D. (2004). Oral discourse in the preschool years and later literacy skills. *First Language* **24**, 123–147.
- Piaget, J. (1959). *The Language and Thought of the Child*. London: Routledge/Kegan Paul.
- Skinner, B. F. (1957). *Verbal Behaviour*. Englewood Cliffs, NJ: Prentice-Hall.
- Snow, C. E., Burns, M. S., and Griffin, P. (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.
- Stone, C. A. (2004). Contemporary approaches to the study of language and literacy development. In Stone, C. A., Silliman, E. R., Ehren, B. J., and Apel, K. (eds.) *Handbook of Language and Literacy Development and Disorders*, pp 3–24. New York: Guilford Press.
- Street, B. (ed.) (2001). *Literacy and Development: Ethnographic Perspectives*. London: Routledge.
- Vygotsky, L. (1986). *Thought and Language*. Cambridge, MA: MIT Press.
- Wertsch, J. (1998). *Mind as Action*. New York: Oxford University Press.

Further Reading

Bilingualism and Learning

L S Verplaetse and E Schmitt, Southern Connecticut State University, New Haven, CT, USA

© 2010 Elsevier Ltd. All rights reserved.

Bilingualism is defined as a speaker's ability to use two languages for communication. Due to the complexity of its nature, the study of bilingualism relies on several fields within linguistics, including anthropology, psychology, neuroscience, and education. The study of bilingualism describes language behaviors of bilingual speakers, social and pragmatic patterns of using two languages, language development, and acquisition and loss, among many other issues. Researchers attempt to explain the effect that bilingualism has on human cognition, societal relationships, and education of bilingual children. This article discusses bilingualism and learning from three perspectives: types of bilingualism, bilingual processing, and bilingualism and academic learning.

Types of Bilingualism

One of the issues that has a direct effect on education is the distinction between individual bilingualism and societal bilingualism. The term societal bilingualism describes various situations when two or more languages are spoken in a society. Societal use of two languages may take different shapes. For example, during the colonization period, societies were formed where two monolingual groups spoke different languages to each other. An example of a different pattern of societal bilingualism are societies where two or more languages are used on a daily basis, as in many African countries and in parts of Asia. Yet another pattern of societal bilingualism is represented in the US, where minority groups must learn the main language of the community, while the socially dominant group remains monolingual (Hoffman, 1991). Finally, there are societies where two languages that may be closely related have different degrees of prestige. The high-prestige language is used in formal contexts, whereas the language of low prestige is spoken at home. Such a situation is referred to as diglossia (Ferguson, 1959) and is widespread in German Switzerland, Morocco, and Paraguay, to name just a few countries. Many factors account for societal bilingualism. Historically, reasons for societal bilingualism may include occupations, conquests, political marriages, colonization, neocolonization, migration, immigration, and internationalization.

As a result of these factors, the majority of the world's societies are bilingual. What is especially noteworthy is that about 70% of the world population is bilingual to some extent. Therefore, globally, bilingualism is much more of a norm than monolingualism.

Individual bilingualism is a term that describes an individual speaker who has command of two languages. Some of the reasons for individual bilingualism include immigration, migration, close contact with other linguistic groups, schooling, and growing up in a bilingual family. Depending on the reason for acquisition of a second language (SL), the level of the speaker's competence in two languages will vary. Only one of these languages is commonly used in the society where the speaker lives. An individual bilingual uses each language for different purposes and usually with different people. For example, in the United States, a Russian-English bilingual speaks Russian to his/her family members about household issues, but uses English at a car dealership or in the bank. Individual bilingualism may be viewed as a positive or negative attribute, depending on the society and its tradition. When the attitude toward bilingualism is positive, bilinguals take their ability to use the two languages as a fact of life. When the societal attitude toward individual bilingualism is negative, monolingual speakers may find it threatening and undesirable, which results in societal unacceptance of bilingual individuals. This attitude forces many young bilinguals to stop using their two languages in favor of the dominant societal language. Abrupt interruption in language maintenance leads to language loss and may result in semilingualism, a situation where speakers are unable to express themselves clearly in any language (Baker, 2006).

Researchers who study individual bilingualism attempt to explain how the two languages within one speaker are able to interact, to be activated and deactivated depending on the communicative needs of the individual, how these languages are represented in the brain, and what kinds of practices, attitudes, and behaviors such a speaker may experience that are different from a monolingual speaker.

Another important issue in the study of bilingualism is the age of acquisition of the SL. When a child grows up in a bilingual family and acquires two languages from birth, the acquisition is simultaneous and such bilinguals are referred to as early and simultaneous bilinguals. The term simultaneous bilingualism emphasizes that the SL is acquired through the same acquisitional mechanisms as the first language. This type of acquisition is qualitatively different from sequential acquisition where the SL is acquired after the first language is in place.

Under immigration, first-generation immigrants continue using their native language while acquiring the majority language of the country where they have settled.

Their children usually acquire the first language from their parents and the new language from people outside the home. In other words, first-generation immigrants are monolingual speakers who add another language to their repertoire and are expected to adapt to the norms of the new country. These bilinguals are referred to as sequential bilinguals; that is, they acquire their two languages in sequence, rather than simultaneously. Sequential bilinguals may be early bilinguals, if they acquire the SL during childhood, or late bilinguals, if they acquire the SL during the adult years.

One of the issues that arises for sequential early bilinguals is the maintenance of their native language. It is typical for young bilinguals to switch entirely to the majority language of their society. If such a switch occurs, it is likely that the first language will be lost and these speakers will become monolingual users of the SL.

Families that use more than one language sometimes make a decision to raise their children bilingually. In this case, they often use the one-parent-one-language strategy, where each parent consistently speaks only one language to the child.

The level of development and use of the two languages is another aspect of the study of bilingualism. Sometimes, bilinguals are equally proficient in both of their languages and are able to use them for similar purposes. This situation, described as balanced bilingualism, is difficult to achieve, especially when one of the languages is a minority language and is not used in the society at large.

A more common situation is unbalanced bilingualism – when the two languages are developed unequally in an individual because languages are used in different domains of life, for different purposes, and with different people. As unbalanced bilingualism constitutes a prevailing scenario, it is important to understand that the bilingual speaker is not a sum of two complete monolinguals in one (Grosjean, 1989). Rather, the language that the speaker is more fluent in is recognized as the dominant language. Importantly, the dominant language is not necessarily the first acquired language of the speaker. The balance between languages can change if the speaker's circumstances change. One of the languages may undergo attrition, fossilization, and forgetting by an individual or by the society (language death); on the other hand, one of the languages may be revitalized, maintained, or standardized.

It is noteworthy that the two languages may be present in the speech of an individual at the same time. If the conversation takes place between bilingual speakers who share the same languages, it is common that two languages will be used alternately within one sentence, across sentences, or even within one word. This type of alternation of languages is referred to as code switching. Code switching is often stigmatized as an ungrammatical, uneducated way of speaking. Despite the seemingly haphazard patterning of the two languages in code switching, one of the

languages controls the grammatical system of the utterance, while the other provides lexical information, that is, fills in the words. Therefore, code-switched production is a complex, rule-governed activity that involves activation and use of two languages simultaneously.

During code switching, both languages overtly appear in the production. However, when a bilingual's production in one language is underlyingly affected by the other language, only one language appears on the surface. The other language is used covertly.

Such covert incorporation of the dominant language is often referred to as convergence to the dominant language (Myers-Scotton, 2002). When converging, the speaker uses all the lexical and grammatical markers from one language, but parts of the structure come from another language in the speaker's repertoire. Such covert use of the SL is indicative of gradual language loss and may alert educators to the necessity to help the speaker maintain both languages.

One of the other aspects of bilingualism that may be of particular concern to educators is whether bilingualism of their students is additive or subtractive. In additive bilingualism, the new language adds to learners' skills and experiences without undermining their native language and existing knowledge. In subtractive bilingualism, learning of the new language takes place at the expense of the native language and the existing skills and knowledge. Researchers emphasize that successful learning of a new language takes place in additive situations due to high levels of motivation and feeling of self-worth that additive circumstances project (Cook, 2001). In monolingual societies, conditions are often created for subtractive bilingualism for those students whose primary language is other than English as it undermines the value of their native tongue. Subtractive bilingualism often results in anomie, which in a bilingual context refers to a bilingual's feelings of discontent and regret for losing ties with one group mixed with fearful anticipation of entering a relatively new group. Anomie can be avoided if educators and parents help young bilinguals to define their identities, which are different from those of monolinguals (Lambert, 1972).

Bilingual Processing

Researchers have established that bilinguals process language differently than monolinguals do as their two languages are in contact and they constantly make decisions as to what language to use. Linguists believe that both languages are continuously activated, but the degree of activation may differ for each language depending on the initial choice and intention of the speaker. Recent studies indicate that speakers have an option to switch to the SL

at any time during their utterance. This means that the degree of activation of both languages is quite high.

Depending on whether the speaker is an early or late bilingual and on the circumstances under which the SL was acquired, the speaker may rely on different types of knowledge when using the two languages. Explicit knowledge of the language refers to the speaker's familiarity with the facts about the language. For example, the speaker may be able to recite a grammar rule. Implicit knowledge refers to the computational operations that the speaker is never consciously aware of and that allow us to actually build and produce intended utterances. Importantly, explicit knowledge can never become implicit, as these are qualitatively different entities. However, it is important for educators to know that explicit knowledge may contribute to the faster development of implicit knowledge (Paradis, 2004). Moreover, studies in bilingual aphasia as well as magnetic resonance imaging (MRI) studies of the bilingual brain have shown that two languages of late bilinguals are localized in different parts of the brain, while early bilinguals seem to have one language center (Paradis, 2004).

Regardless of language location in the brain, activation is currently viewed as the key to language availability and maintenance. The activation threshold hypothesis proposes that in order for an item (vocabulary or grammar) to be activated, a certain amount of neural impulses need to reach neural substrate. The amount of neural impulses sufficient for activation is called the activation threshold. Each time an item is activated, the threshold becomes lower. If an item is not activated for a longer period of time, the threshold rises (Paradis, 2004). While the activation threshold operates on both monolingual and bilingual speakers, the activation is more complicated for bilinguals. In order for an item from one language to be activated, the competing item from another language must be inhibited. In other words, the activation threshold of a similar item in the unwanted language must be raised. Importantly, production of an item is more difficult than comprehension of the same item because the activation threshold needs to be lower for production than comprehension. That is why when bilinguals do not use one of their languages, their activation threshold may rise to the extent that they are no longer able to produce utterances in that language, but are still able to comprehend them (Köpke, 2002).

Some items require lower level of activation threshold than others. For example, lexical items require a higher level of activation threshold than do morphological features that are a more permanent part of the language. However, at the initial stages of acquisition of a new language, we observe that new lexical items of high frequency in the speaker's environment rather quickly achieve lower level of activation threshold, whereas grammatical markers are at high level of activation threshold. This often results in structural interference, that is, a negative effect of one language on the other.

Bilingualism and Academic Learning: The Education of Bilingual Children

Bilingual children and, in particular, school children who are in the process of becoming bilingual, pose particular challenges to school systems. Schools throughout the world deal with bilingualism in various ways. In some cases, schools' intentional goals are to produce bilingual children. In other cases, schools may approach the education of their language minority students (those students who do not speak fluently the language used in the school) as a problem that needs to be solved, because the language minority children may not have sufficient language proficiency to understand the language of the school instruction. In yet other cases, schools may intentionally or unintentionally ignore the linguistic plight of their language minority students and expect the students to either sink or swim in the mainstream academic setting.

Theoretical Constructs of Bilingual Education

Before examining the various types of school models of bilingual education, it is important to identify two theoretical constructs that help shape school choice in this area: the distinction between conversational language and academic language, and the interdependence between the first-language and SL literacy.

Language use varies depending upon setting. In simplistic terms, language use can be divided into two types: conversational language and academic language (Cummins, 1984). Conversational language use allows people to engage in basic, interpersonal conversation, such as visiting over coffee or on the school playground, ordering food in a restaurant, inviting someone over to the house, or narrating a story of what happened earlier in the day. Generally, conversational language assumes that: (1) both speakers (and hearers) have some sense of shared history or are speaking about a topic in the here and now; (2) both speakers can make use of gestures and visuals, and intonation; and (3) both speakers have the opportunity to negotiate meaning, that is, to signal if they do not understand and to seek clarification. As two speakers move away from any of these three features, the talk becomes less and less contextualized by the here and now, by visuals and gesture, and by the ability to negotiate meaning. When this happens, more of the burden of the talk is invested in the language, rather than the above-mentioned contextualizing devices. As the speakers move away from all three features (imagine this movement on a continuum), the language activity becomes less conversational and more academic.

Academic language (or literate language) is the second of the two language-usage types. Academic language is the language of school textbooks, expository writing,

classroom lectures, and some classroom discussions. This language type is distinctly different in its linguistic composition from conversational language, containing specialized lexical choices, more logical connectors, more clause-combining strategies, and more sophisticated use of tenses and aspect (Schleppegrell, 2004).

The distinction between conversational language and academic language is important for this reason. The time required to develop each language type varies significantly; conversational language skills take 1–2 years to develop, while academic language proficiency requires 5–9 years. (Hakuta *et al.*, 2000; Thomas and Collier, 2002). Consequently, language-learning students with strong conversational language skills may sound like they are fluent in their new target language, but may still be developing their academic language skills in this new language. As a result, teachers may mistakenly place students into mainstream instructional settings without the additional support required to help the students acquire the more refined academic/literacy related aspects of the new target language.

The second theoretical construct important to the education of bilingual children is the interdependent relationship between the literacy skills of the student's first language and the literacy skills of the SL. It has been determined that many of the literacy skills acquired in a student's first language are readily transferable to the student's SL. Therefore, time spent learning how to read in a child's first language is not time lost and, in fact, may be important time invested in that student's literacy development process (August and Shanahan, 2006; Cummins, 1991).

The educational implications of these two theoretical constructs are as follows:

1. Bilingual children, learning an additional language, will need sufficient time and academic support to develop the academic language skills of that target language.
2. Literacy in the first language, if at all possible, can be a most important step in that student's literacy development. It will not detract from the child's development of literacy in the SL, and it may provide a most important foundation.

Types of Bilingual Education

Schools throughout the globe offer a wide variety of education program types for bilingual students. These programs have a wide variety of names attached to them. In an attempt to solidify and simplify, we discuss six types:

1. bilingual immersion;
2. two-way/dual bilingual;
3. maintenance bilingual;

4. transitional bilingual;
5. sheltered content instruction; and
6. SL, such as English as a Second Language (ESL)

Bilingual immersion programs place students, all of whom share a similar language background, into a classroom setting in which students are immersed into a language other than their own. Successful examples of such programs include the Canadian Immersion Model, in which English-speaking children are immersed in French schooling for the first 2 years of school, or alternatively during the students' middle school years. Another example can be found in Hawaii, where native English-speaking children attend school in Hawaiian during the first 5 years of schooling. In each case, a small portion of schooling is conducted in the students' native language each day. It should be noted that immersion models are radically different from submersion models, in which students must simply sink or swim. In structured immersion models, the students speak a home language that has societal value and learn a target school language that holds a societal value or an historical value. Importantly, the value of their home/native language is not questioned, nor is there a danger of students losing their home language. In the submersion model, the language minority students receive no instruction or support in their home language. They are submersed into an instructional environment that is designed for the majority native-language speaking students. Therefore, the successful history of structured immersion programs cannot serve as justification for the use of submersion models.

Dual bilingual programs serve classrooms comprised of two groups of children, each of which speaks a different first language. The two languages represented by the students become the shared medium of instruction. Part of the school day or school week is conducted in one language; the other part of the day or week is in the other language. At any time, half of the students are working in their first language, while the other half of the students are working in a new, target language. The goal of these programs is to develop students who are bilingual in two languages, with conversational and academic language proficiency in both.

Maintenance bilingual programs serve classrooms comprised of all SL learners, students who speak a single, common minority language at home, and who need to acquire the societal majority language at school. In such programs, both languages are used. How the classroom teacher divides the class time between the two languages varies, but this is not done randomly. The program specifies if the use of the two languages is 50/50, or 60/40, or even 90/10 at the beginning of the student's time in school. Oftentimes, literacy is developed first in the home language, and later in the target language of the school. The goal of the program is to not only develop the target,

societal language, but also to maintain and develop the students' home language, and to develop the students' academic fluency in both languages.

Transitional bilingual programs serve a school population similar to that served by the maintenance bilingual program. However, the goal of the program is to develop the students' proficiency in the dominant, societal language. Therefore, the transitional program uses the students' first language only as a tool to bridge the students' academic development until they are able to fully perform academically in the target language of the school. In transitional programs, students often do not develop a strong literacy base in their first language, but rather, are taught literacy skills primarily through the target school language.

Sheltered content instruction is a relatively new method of teaching bilingual children in school systems. This method teaches course content in the language of the school, but in language environments that are highly structured. Sheltered instruction requires that teachers trained in specific content disciplines learn how to deliver instruction in these content areas in ways that make the content comprehensible to students who do not share the language of the course content. They do this by contextualizing the academic content through the use of visuals, shared history, and frequent negotiation of meaning (as described earlier in this article), and through the highly controlled use of the target, school language. The goal of sheltered instruction is to develop content and academic language proficiency in the new language.

SL programs (such as ESL) attempt to teach the target school language to bilingual students who speak a home language other than the school language. Students in one SL classroom may represent a large number of different home languages; therefore, the SL teacher must rely almost completely, if not entirely, on the school language. Bilingual students who need such help may be pulled out of their mainstream classes to attend SL classes; or, teachers trained in SL instruction may be pushed into the mainstream class, where that teacher works as a partner with the content teacher. SL classes can focus on conversational language or on the specific academic language of the content courses.

Efficacy of Bilingual Education

The decision of what type of bilingual educational program a school system will employ generates much public discussion, even political heat. In some cases, laws have been passed which strictly dictate how a school system will handle bilingual education. One recent study provides us an insight as to the efficacy of program types. Thomas and Collier (2002) compared dual, maintenance, transitional, and various ESL programs in major cities in the United States of America and found that only those

bilingual programs that develop both languages (dual and maintenance programs) ultimately resulted in bilingual students' ability to perform at minimally the 50th percentile or better as compared to native English-speaking students in these same school systems. While the Thomas and Collier study indicates that programs using only the target school language have not been successful for bilingual students, it should be noted that this study was conducted at a time when there was still too little ownership in mainstream education for the education of bilingual children. It was thought that the bilingual or ESL programs should bear the responsibility for the education of bilingual children until they were mainstreamed. Now that research has made it clear that it takes anywhere from 5 to 9 years to develop academic language, school systems are understanding that all teachers must bear the responsibility of the education of bilingual children. Hence, teachers are learning to employ sheltered strategies in their classrooms and to offer sheltered content courses to large groups of bilingual children. Educators remain hopeful that with increased training in sheltered and differentiated instruction for content teachers, future studies will reveal more positive findings for the academic performance of bilingual students.

Despite the findings of Thomas and Collier, the discussion of whether or not to use language minority children's first languages in their education remains a much heated, politically divisive debate in many countries. However, in general, the research is clear on this topic. A preponderance of evidence from various studies in various countries (Norway, England, Netherlands, Sweden, Australia, Mexico, and China) suggests that the use of students' first language in their educational experience does not hurt the child, and in many, if not most, cases students benefited, either in increased academic abilities, enhanced literacy skills, and/or even improved student behaviors (Krashen, 1999).

See also: An Overview of Language and Literacy in Educational Settings; First Language Acquisition; Foreign Language Learning; Second Language Learning.

Bibliography

- August, D. and Shanahan, T. (2006). *Developing Literacy in Second-Language Learners: A Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Erlbaum.
- Baker, C. (2006). *Foundations of Bilingual Education and Bilingualism*. Clevedon: Multilingual Matters.
- Cook, V. (2001). *Second Language Learning and Language Teaching*, 3rd edn. London: Arnold.
- Cummins, J. (1984). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. In Rivera, C. (ed.) *Language Proficiency and Academic Achievement*, pp 2–19. Clevedon: Multilingual Matters.
- Cummins, J. (1991). Interdependence of first- and second-language proficiency in bilingual children. In Bialystok, E. (ed.) *Language*

- Processing in Bilingual Children*, pp 70–89. New York: Cambridge University Press.
- Ferguson, C. A. (1959). Diglossia. *Word* **15**, 240–325.
- Grosjean, F. (1989). Neurolinguist, beware! The bilingual is not two monolinguals in one person. *Brain and Language* **36**, 3–15.
- Hakuta, K., Butler, Y. G., and Witt, D. (2000). How long does it take English learners to attain proficiency? *The University of California Linguistic Minority Research Institute Policy Report 2000-1*. <http://www.stanford.edu/~hakuta/www/research/publications.html> (accessed May 2009).
- Hoffman, C. (1991). *An Introduction to Bilingualism*. London: Longman.
- Köpke, B. (2002). Activation thresholds and non-pathological first language attrition. In Fabbro, F. (ed.) *Advances in the Neurolinguistics of Bilingualism. Essays in Honor of Michel Paradis*, pp 119–142. Undine: Forum.
- Krashen, S. (1999). *Condemned without a Trial: Bogus Arguments against Bilingual Education*. Portsmouth: Heinemann.
- Lambert, W. (1972). *Language, Psychology and Culture*. Stanford: Stanford University Press.
- Myers-Scotton, C. (2002). *Contact Linguistics. Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press.
- Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins.
- Schlepppegrell, M. J. (2004). *The Language of Schooling: A Functional Linguistics Perspective*. Mahwah, NJ: Erlbaum.
- Thomas, W. P. and Collier, V. P. (2002). *A National Study of School Effectiveness for Language Minority Students' Long-Term Academic Achievement*. Santa Cruz, CA: Center for Research on Education, Diversity, and Excellence.

Further Reading

- Baetens-Beardsmore, H. (1986). *Bilingualism: Basic Principles*. Clevedon: Multilingual Matters.
- Bialystok, E. (2001). *Bilingualism in Development: Language, Literacy, and Cognition*. Cambridge: Cambridge University Press.
- Brisk, M. E. (2006). *Bilingual Education: From Compensatory to Quality Schooling*. Mahwah, NJ: Erlbaum.
- Cloud, N., Genesee, F., and Hamayan, E. (2000). *Dual Language Instruction: A Handbook for Enriched Education*. Boston, MA: Heinle and Heinle.
- Clyne, M. (2003). *Dynamics of Immigrant Language Contact*. Cambridge: Cambridge University Press.
- Dewaele, J., Housen, A., and Wei, L. (2003). *Bilingualism: Beyond Basic Principles*. Clevedon: Multilingual Matters.
- Faltis, C. and Hudelson, S. (1998). *Bilingual Education in Elementary and Secondary Schools Communities*. Boston, MA: Allyn and Bacon.
- Walters, J. (2005). *Bilingualism: The Sociopragmatic-Psycholinguistic Interface*. Mahwah, NJ: Earlbaum.

Classroom Discourse and Student Learning

L Hemphill, Wheelock College, Boston, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Academic register – The language style most commonly used in academic contexts, typically incorporating features of written language.

Authentic questions – The questions that reflect genuine openness about the sought-after response.

Code-shifting – The alternation between two or more languages or language varieties within the same discourse.

I–R–E sequence – A three-part interactional sequence beginning with a teacher initiation followed by a student response and teacher evaluation.

Decontextualized language – The language that does not rely for its interpretation on the speaker and hearer's shared context; it is often characteristic of academic and written language.

Display question – A question that requires the respondent to verbally display information that is already known to the questioner.

Elicitation sequence – A chain of teacher questions and student responses in which the teacher attempts to elicit a sought-after response.

Nonstandard dialect – A community language variety such as African-American English vernacular that displays systematic differences with the standard form of the language.

Open-ended question – A question that does not elicit only a yes–no or phrasal response.

Speech event – A routinized communication situation involving particular participants, settings, move sequences, and communicative goals.

Definition/Scope

Classroom discourse research arose in the 1970s as a special focus within the new discipline of conversation analysis and early on identified some of the distinctive formal characteristics and social purposes of talk in schools. Research has generated descriptions of the forms of talk that are specific to academic contexts, characterizations of the rules that govern teacher–student talk, and accounts of the ways in which the development of language skills can be fostered or hindered in classrooms. As analytic tools have been applied to different kinds of

school settings, the field has contributed to better understanding of varied issues such as the socialization of academic language, gender roles in classrooms, and processes of second-language learning in school.

Differences with Everyday Conversation

As early research convincingly documented, teacher–student talk proceeds according to different rules and conventions than everyday conversation. These arise from the social purposes of much classroom talk, which center on the teacher's presentation of academic content and assessment of students' uptake of information. Largely absent from teacher–student communication are the phatic exchanges that are common in other forms of everyday talk: "How's it going?" Instead, teachers lecture, ask questions to check on student understanding, and provide brief evaluations of the adequacy of students' responses ("Good!" or "Anyone else?").

At the level of individual utterances, classroom talk often displays features of decontextualized language such as nominalizations (multiplication, thinking), and embedded clauses (such as that-clauses). Classroom discourse incorporates these features because teachers and students often talk about abstract concepts or entities that are physically absent from the context of conversation: carbon bonds, the Treaty of Paris, animal habitats, and the thoughts and feelings of fictional characters. In contrast to the topics of much everyday conversation outside of classrooms, the referents for classroom topics are often not visible or mutually accessible through other nonverbal channels, for example, through shared activity ("Could you hand me that?" "We can tackle it later."). The potentially supportive role of shared physical and social context is often diminished and therefore, in classroom discourse, interactants are typically obliged to use more explicit and verbally elaborated strategies for describing their conversational topics. Topical referents need to be first introduced through talk and then disambiguated and elaborated upon through verbal commentary. Repeated reference to the same topic is usually accomplished through definite reference in classroom talk (the theme of the story, the lesson, or the point) rather than through the deictic strategies more commonly used in thematic progression in everyday informal conversation (it, this, or that).

Classroom discourse is often motivated by the need to build up shared and elaborated understandings of

particular topics over time, for example, across the course of a unit or academic year; thus teachers and students use a variety of strategies to mark what is jointly understood and remembered about a topic.

Asymmetric Roles

A notable characteristic of teacher–pupil talk is an asymmetry of speaking roles: teachers lecture and ask questions and students answer the questions that are posed to them. Teachers dominate and monopolize the conversational floor in classrooms at every stage of the educational system, taking longer and more frequent turns than student participants. In contrast to other types of conversation in which children and young people can take more varied roles, in classroom discourse, teachers typically reserve the right to introduce new conversational topics and to engage in most forms of questioning. Teachers control access to the conversational floor through nominating respondents to their questions (“Hands, please!”) and often through limiting opportunities for side conversations or talk on student-initiated topics. In addition, different from other conversational contexts, many of the questions that teachers pose are close-ended, questions that require only minimal or fragmentary responses from students. The forms of teacher questions largely determine that syntax and degree of elaboration of student responses:

Teacher: Where do most cats live?

→ Student: Houses?

Teacher: They live in houses, right.

Many teacher questions take the form of what have been called display or known-answer questions. Unlike what have been termed authentic questions, display questions require students to present information that is already known to the teacher.

→ Teacher: So what’s the title of the story that we read?

Student: The Sea Serpent.

Student responses to known-answer questions are often quite brief; the point is to demonstrate recognition of the right answer rather than to present a unique or personal perspective on the topic. Responses to known-answer questions are often only interpretable within the context of the preceding teacher talk:

Teacher: Let me see your folder first. The book report cover. Okay, what do you have here? What’s that? (points)

→ Student: Title.

Teachers’ open-ended questions, on the other hand, require students to construct relatively autonomous turns that often present new information about the topic:

Teacher: Do you have a prediction about that?

Student: I think if we keep rolling the dice, we’ll keep getting numbers under 30. Because the dice numbers aren’t that big and if you multiply them, it will be like 20 and 16 and 10 and 8 like we have been getting.

Responses to open-ended question often elicit students’ most sophisticated language skills, requiring vocabulary that is specialized for particular subject areas and grammatical forms that include conditionals, subjunctives, and embedded clauses.

Three-Part Sequence

Much classroom interaction follows a three-part scheme, where the teacher first initiates a question–answer sequence, followed by a response from a student who either volunteers or is selected by the teacher to respond. A third element in what has been called an I–R–E structure is an optional teacher evaluation move, offering feedback to the student about the adequacy of his or her response. These three-part question–answer exchanges create the framework for most student contributions to talk within classrooms, typically alternating with longer stretches of teacher monolog or lecture.

Teacher: Do you know what that means, when they say water’s getting rough?

Student: That the water’s like getting wavier?

Teacher: Right.

Teachers use I–R–E exchanges for several social and pedagogical purposes: to assess student background knowledge before introducing a new topic or reintroducing a subject; to assess student uptake of new information, to break up monologic presentations of information; to mark transitions between topics; and to re-engage students whose attention may be flagging. Many teacher questions do not immediately elicit the desired response; thus classroom discourse often includes extended elicitation sequences in which the teacher repeats or reshapes the original question, seeking a closer approximation of a correct answer through hints, simplifications, and restatements:

Teacher: Look at the cover and tell me what you think the story will be about.

Student: It’s about animals.

→ Teacher: What kind?

Student: Cats. Mice.

The teacher evaluation moves that follow partial or incorrect answers (“What kind?”) provide cues to students that their expressed understanding of the topic is not yet fully adequate. A positive evaluation move by the teacher often serves to not only confirm the adequacy of an answer but can also mark the end of a particular topical

focus (we're clear about this aspect of the topic now) and can set up a shift in lesson focus:

Teacher: What's a landing? If you're in a boat, what's a landing?

Student: Oh you like go to the um to the um sand?

Teacher: You go to the where?

Student: To the sand, uh the boat.

Teacher: To the sand?

Student: Where the people are, where you play.

Teacher: You mean what, you go back to...?

Student: Where you were.

Teacher: Where you were. What's that called? What's that called, where you were?

Student: You go back to the shore?

→ Teacher: To the shore, very good. Good word. So it's a rough landing because, why? Why was it a rough landing? What was pushing them in? They thought it was the wind, but what was it?

Student: A sea serpent?

Teacher: The sea serpent was pushing them in and it made for a rough landing.

Effects on Opportunities for Student Participation

The combined effects of several characteristics of classroom discourse: teachers' monopoly over the introduction of conversation topics and access to the floor, teacher preferences for closed questions, and the use of I-R-E sequences, all may serve to limit opportunities for extended student participation. Student turns are short in many classroom contexts because student contributions function most often to ratify, rather than expand upon topics introduced by the teacher. Although the topics of classroom talk often call for decontextualized word choices and sentence patterns, very often student turns are grammatically incomplete because they respond to closed teacher questions that require only a yes/no or phrasal response (e.g., sodium, 1815, in Philadelphia, and sad). The goal of developing student thinking and mastery of content through talk is often subordinated to the more local goal of getting the right answer on the floor.

Socialization into Academic Registers

Classrooms present opportunities not only for learning academic content but also for learning how to talk about academic topics. Children begin school with discourse histories that reflect the norms and interactional practices of their families and communities. For some children, these include an expectation of close adult attention to child-initiated topics, considerable experience with holding the conversational floor, and experience of frequent encouragement from

adults to say more, to verbally elaborate on intentions, feelings, and beliefs. Not all cultures, however, view children as entitled to set the topics and take long turns in adult-child interaction. As a result, many children begin school with greater reticence in adult-dominated conversation and with less willingness to go beyond what is minimally required in adult-child exchanges. Children also start school with varying degrees of facility with the vocabulary and syntax of academic talk, including knowledge of verbs for talking about cognition (e.g., wonder, suppose, predict, and imagine), and understanding of academic vocabulary, for example, terms commonly used in primary schooling such as add, uppercase, government, and story character. These contrasts in children's preparation for engagement in academic talk often reflect differences in family social status, parental education levels, and cultural beliefs. Parents who have experienced success through the educational system seek, often unconsciously, to reproduce that success for their children through home-socialization strategies aimed at producing verbal responsivity and fluency within some of the adult-child questioning and responding routines that are characteristic of classrooms. Children from educated families enter school with experiences from family contexts like shared book reading and mealtime talk that position them to be highly responsive and confident in classroom discourse. Other children look for cues in classroom interaction on how to participate competently, often waiting for adult guidance and prompting.

Among the tasks that children face in learning how to engage in classroom talk in the early years (and that older students face at junctures in their education as well) include acquiring the vocabulary for particular academic subjects and learning to recognize the varied forms of participation that are required in response to different types of teacher questions: "Who can tell us something about penguins?" "How did you feel when the parents couldn't find Sylvester?" "Why do we need to regroup when we're adding 15 and 6?" To be successful participants, children need to know what content is relevant to particular academic contexts and the discourse shape this content should be presented in, for example, a request for a personal response to an event in a work of literature requires the expression of emotion state words (sad, worried) followed up by specific references to details in the literary work (his mom and dad walked right next to him and couldn't tell that he was there). Some teacher questions require fragmentary answers while others signal in subtle or more marked ways that more extended student responses are required.

Classroom Speech Events

Similar to talk outside of classrooms, classroom discourse is organized into speech events or subroutines that are set off from the flow of talk by movement into particular

settings: the book corner, the lab table, the rug, or by formal announcements, “It’s time for morning meeting.” Classroom speech events incorporate specific participants (the reading group members, lab partners, or the whole class) and specialized participant roles that children learn through participation: everyone will get a chance to speak, only one child speaks at a time, and a newsworthy contribution will merit an extended turn at speaking. Teachers use overt cues, “Let’s turn on our listening ears now,” and more implicit signals, for example, shifts in posture or eye gaze, to mark the initiation of specialized talk sequences. Classroom speech events typically follow a predictable sequence of moves, such as reciting names for attendance, doing a weather and calendar chart, and presenting personal news, the moves of a morning meeting speech event within many early childhood classrooms in the US. Moves within speech events also have differentiated rules governing access to speaking rights (going around the room one by one for attendance; raising hands to be selected to tell personal news) and moves have their own internal discourse structure, for example, the narrative forms required for telling personal news. The complexities of classroom speech events pose challenges to students at all stages of schooling; however, teacher talk provides support in acquiring the desired forms.

Modeling

One of the ways that teachers help children learn how to participate in classroom talk is through direct modeling of the desired response or response type. Teachers frequently answer their own questions in order to get some approximation of the sought-after response on the floor and to provide models for students of the kind of responses they are seeking. Teachers often effectively fill both slots, questioner and respondent, in the interactional sequence:

Teacher: What kind of a soup is it – how would you describe pea soup?

Student: Mmm. . . (shrugs)

Teacher: Hard to describe, huh? What does it look like?

Student: It’s like it’s a bunch of peas inside a pan with water.

→ Teacher: And the water’s called broth and broth is real, real thick. Have you ever had it? It’s a real thick – you know how some soups are almost watery?

Students: Yeah.

→ Teacher: Well this is real, real thick. And I want you to remember that cause that has something to do with why this story is called, “Pea Soup”.

Modeling is especially commonplace in classrooms for students who are second-language learners or speakers of nonstandard dialects. Teacher support through modeling

is also particularly common in classrooms where students are being introduced to new topics or initiated into new classroom speech events.

Scaffolding

Teachers use a wide range of scaffolding strategies to help students articulate the desired answers to the questions they pose. These include calling on another student (“Maybe Stephen can help.”), evaluating an answer as only partially correct (“Not quite.”), and providing concrete strategies for arriving at a better response (“Read to the end of the sentence.” “Look at the board.”). Scaffolding works towards the goal of having the student arrive at the correct answer and often forestalls a negative teacher evaluation.

Teacher: Okay. Right there, what does it mean, “. . .the helicopter descended?”

Student: Slowed down.

→ Teacher: Not quite.

Student: Stopped.

→ Teacher: Read to the end of the sentence.

Student: Oh, “came down”.

Recasts

Teachers also offer recasts or expansions of student responses that fall short of the desired answer. Frequently, these teacher moves repeat the student’s response, substituting more elaborate syntax, more precise vocabulary, or language specific to the academic subject:

Teacher: How would you get the answer?

Student: I would times it.

→ Teacher: So you times the two numbers, or multiply them to get the answer.

These moves juxtapose the learner’s response with a response that expresses the same content more formally, recasts are particularly effective in promoting acquisition of the vocabulary and syntax of academic language.

Learning Specialized Academic Registers

Teacher coaching and support are often needed as students begin to participate in talk within particular academic registers. The discourse of mathematics, for example, differs in many dimensions from the ways teachers and students talk about literature, even in the early years of schooling. Mathematics talk includes generalizations about the rules governing number relations and often presents hypothetical operations upon numbers. Talk about literature, on the other hand, incorporates elements of narration (e.g., descriptions of character actions and reactions) and values direct citation from the

literary text. Children learn that they are often called upon to relate a personal response or read a supporting passage during literature discussions, but virtually never are asked for these kinds of contributions during mathematics lessons.

In learning to talk in different ways during particular subject lessons, students begin to understand some of the ways of thinking that are characteristic of particular academic disciplines, for example, literary analysis' close attention to the text and mathematics' concern with the hypothetical and with logical proof and demonstration. These ways of thinking are embodied in language practices that are pervasive in classroom talk, even in the primary grades. The most successful teachers explicitly prompt students to produce discipline-specific talk and provide rich models and explanations of the language rules that operate within the disciplines.

Teacher Talk and Particular Populations

Classroom discourse research has been particularly helpful in elucidating language practices that both support and hinder the progress of particular groups of learners in schools. Second-language learners, speakers of nonstandard dialects, and girls, all face particular challenges in participating successfully in classroom talk.

Second-Language Learners

Schooling in a language of instruction that is not students' home language is common across the developing world and in communities with immigrant students in developed countries. In addition, in many countries, students from language minority communities are schooled in a national language such as English, Spanish, or Malay. Thus, many students must simultaneously acquire a new language in school and learn new academic vocabulary and discourse practices. Teachers whose students are still acquiring the language of instruction often simplify their talk and augment verbal exposition with gesture, demonstration, and illustration. Clarification sequences, where teachers use multiple interactional moves to clarify their meaning or their students' intended meaning, are commonplace in bilingual classrooms. Code-shifting and code-mixing, in which teachers and pupils alternate or mix languages within or across turns, are also common in second-language classrooms. Teachers and peers use code alternation both as a strategy for bridging or repairing communication gaps and as a social practice. Teachers who are themselves bilingual may present academic content in the national language but shift into a community language for greater clarity, for disciplinary exchanges ("¡Dámelo!"—Give it to me!), or to mark a boundary between whole class and private, off-the-record communication. Students

may read aloud in the official language but shift into the home language when offering commentary on the text. Such practices in second-language classrooms are natural extensions of the complexities of language use in multilingual communities and may in fact promote more rapid acquisition of the target language.

Speakers of Nonstandard Dialects

For students who are speakers of nonstandard dialects such as African-American vernacular English, classroom talk can be a productive context for acquiring a second, standard dialect or alternatively, a context for depreciation of community language norms. Although some teachers expect standard forms in all contexts of classroom communication, many teachers, particularly those working in language minority communities, hold differentiated expectations for the appropriateness of community dialects within different classroom speech contexts. These teacher expectations often mirror community beliefs about the appropriateness of dialect or standard language forms in particular contexts. For example, children may be corrected or prompted for use of the standard dialect when reading text aloud or in recitation contexts but may be allowed to use nonstandard forms in problem-solving groups and in private teacher–student exchanges. Classroom regulation of the use of nonstandard forms begins as early as preschool, through prompts like, "No street talk, Frankie," and accelerates as school-aged children acquire skills in two dialects and develop sensitivities to the social contexts that are intertwined with dialect variation.

Children's dialect patterns often extend beyond nonstandard word choices and syntactic constructions although these sentence-level differences with the standard dialect are especially likely to be targets of teacher corrections and recasts. Particularly complex and potentially problematic in classrooms are dialect patterns that affect discourse organization and participation roles, such as preferences for overlapping turns at talking, extended silences after teacher questions, or topical chaining in narrative discourse. When teachers and students do not share discourse styles, the task of acquiring skill within academic registers is particularly difficult. Teachers may interpret students' discourse patterns as signs of defiance, academic inadequacy, or failure to understand classroom-speaking rules. If there is little teacher awareness of differences in the discourse rules between the school's language and the language of many students, classroom discourse may be experienced as a context for failure, misunderstanding, or shaming.

Gender and Participation

The participation of girls and women in classrooms has been the focus of special investigation. In gender-mixed

classrooms, particularly at the secondary level, boys often volunteer to speak more readily than their female classmates and hold the floor longer when they are called on to contribute. Girls engage in more hedging in their contributions, diminishing their degree of commitment to the points they are presenting through verbal downgraders, such as “kinda,” “maybe,” “I think,” and through question intonation when they make statements. Male students have been observed engaging in more floor competition, overlapping a classmate’s turn before he or she is done speaking, and calling out answers even when the teacher is using a nomination strategy to select the next speaker turn. Although girls’ language development is often in advance of male classmates’, the rules of classroom engagement do not necessarily favor girls’ development. In fact, boys often occupy a middle position in the classroom social hierarchy, taking on some of the roles that are reserved for teachers when they produce longer floor-holding turns, carry on side conversations, and challenge classmates for control of the conversational floor.

Conclusion/Evaluation

Classroom talk serves as a critical context for the acquisition of more sophisticated language skills and for children’s socialization of academic identities. As early as preschool, children learn through contexts like classroom sharing time and morning meeting to participate in talk about physically and temporally absent topics: our family’s weekend events, yesterday’s weather, my pet. Different from many home contexts for extended discourse, the audience for such talk – peers and teachers – may lack first-hand knowledge of the entities that the child is talking about. Thus, classroom communication creates an essential motivation for learning to be clearer and more explicit than is typically required of children as they participate in everyday family talk.

The topics of classroom communication often draw upon children’s most advanced vocabulary and cognitive skills and require that children use newly acquired metacognitive verbs and complex syntax: conditional, subjunctive, and embedded clauses. Teachers (and sometimes peers) are often demanding as communicative partners, pushing for greater clarity and elaboration, and providing models for the use of academic vocabulary and sophisticated sentence structures. For children who speak other languages or dialects at home, classroom talk provides especially rich opportunities for language learning, although these opportunities may be blunted by failure to respect and include a role for the home language or dialect in school.

Although the field of classroom discourse began with critiques of the inauthentic character of much classroom talk and was accompanied by skepticism about classrooms

as environments for language learning, research of the past 15 years has demonstrated that many aspects of classroom talk support language learning and cognitive development. Teacher roles in modeling academic language and scaffolding effective contributions appear in the light of contemporary research to be much more positive than was suggested by earlier commentary. Teachers often use knowledge of the learner’s current stage of development to offer prompts that press for expression of the learner’s best understandings. Although the strategies teachers employ are inauthentic when compared to many adult roles in everyday conversational exchanges, teacher talk is often well adapted to the cognitive and interpersonal purposes of the classroom. Student-participation slots are often constrained and limited in classroom talk, but the simplification and orchestration of student contributions support an ability to focus on content and on precision of expression.

See also: An Overview of Language and Literacy in Educational Settings.

Further Reading

- Cazden, C. B. (2001). *Classroom Discourse: The Language of Learning and Teaching*, 2nd edn. Portsmouth, NH: Heinemann.
- Edwards, A. D. and Westgate, D. P. (1994). *Investigating Classroom Talk*. London: Falmer Press.
- Edwards, D. and Mercer, N. (1987). *Common Knowledge: The Development of Understanding in the Classroom*. London: Methuen.
- Lemke, J. L. (1990). *Talking Science: Language, Learning, and Values*. Norwood, NJ: Ablex.
- McHoul, A. (1990). The organization of repair in classrooms. *Language in Society* **19**, 349–377.
- Mehan, H. (1979). *Learning Lessons: Social Organization in a Classroom*. Cambridge, MA: Harvard University Press.
- Nystrand, M. (1997). *Opening Dialogue: Understanding the Dynamics of Language and Learning in the English Classroom*. New York, NY: Teachers College Press.
- O’Connor, M. C. and Michaels, S. (1993). Aligning academic task and participation status through revoicing: Analysis of a classroom discourse strategy. *Anthropology and Education Quarterly* **24**, 318–335.
- Sinclair, J. M. and Coulthard, R. M. (1975). *Towards an Analysis of Discourse: The English used by Teachers and Pupils*. London: Oxford.
- Snow, C. E. and Dickinson, D. K. (1991). Skills that aren’t basic in a new conception of literacy. In Jennings, E. M. and Purves, A. C. (eds.) *Literate Systems and Individual Lives: Perspectives on Literacy and Schooling*, pp 179–191. Albany, NY: State University of New York Press.

Relevant Websites

- <http://childes.psy.cmu.edu> – CHILDES: The Child Language Data Exchange System.
- <http://www.leaonline.com/loi/dp> – Discourse Processes, Lawrence Erlbaum Associates.
- <http://www.societyfortextanddiscourse.org> – Society for Text and Discourse.
- <http://talkbank.org> – TalkBank.

Early Writing

B E Hagtvet, University of Oslo, Oslo, Norway

© 2010 Elsevier Ltd. All rights reserved.

Writing as a Multimodal Phenomenon

An increasing awareness, during the 1970s, with regard to preschool children's literacy learning led to a growing body of research on early writing during the last part of the twentieth century. These studies complemented the work of pioneers Montessori (1971), Luria (1978), and Vygotsky (1978). They also marked a new era in literacy research – by highlighting how literacy emerges during preschool and early schooling, and by emphasizing children's early writing as an interdisciplinary field with great educational implications.

Writing is a social practice – an act of communication that uses script to compose messages. Similar to reading, writing basically consists of two skills: the ability to use a notational system (i.e., the alphabetic code) and the ability to handle word messages. Over the past decades, research on early writing has been prolific – covering the development of graphic script as well as the skills needed to write down verbal messages. There has also been a focus on linguistic and cognitive knowledge in the interface between oral and written modalities. For example, studies of the emerging notational system, of how early scribbling is transformed into letters via drawings and various graphic symbols, have shed light on the development of children's understanding of language as a system, and on the relationships between phonemes and graphemes. The research has also examined differences between drawing and writing and between signs and letters, as well as how older preschoolers and young school children tackle orthographic spelling and spelling rules. Furthermore, researchers have focused on the process of constructing verbal messages, on how children act, speak, and interact while transforming speech into written script. There have been studies of how children make use of pencils, computers, and toys in writing, and of how written language learning is culturally embedded with the script mediated by significant adults and peers in the child's environment.

Research on early writing has drawn attention to not only the understanding of emerging literacy, but also to the importance of the preschool years in literacy learning. In defining early literacy, there is no need to make a strong distinction about the point in time that children start formal writing instruction at school. Children – around the world – start school at different ages, and different school systems emphasize formal introduction to literacy differently. Thus, the term early writing is defined here more pragmatically as the period prior to a child mastering the conventional writing system with a certain consistency.

Disciplinary and Theoretical Roots

Prior to the 1970s, the idea of early writing was mainly associated with writing as a motor activity or with early spelling using alphabetic script. In some cases, it was also associated with the skills of young fluent readers, that is, children who taught themselves to read and write prior to formal schooling (Clark, 1976). Only in rare cases was early writing seen as an integral part of developmental theory – first and foremost by Luria and Vygotsky. The research was typically anecdotal and conducted in natural settings at home and in school.

In the 1970s, there was a major shift toward a broader concept of writing. The concept of emergent literacy was coined – reflecting an interest in the study of how literacy emerges during the preschool years. Writing was now more often studied ethnographically by focusing writing practices in different contexts: in nursery school, in collaborative play, or in informal situations in children's homes (Clay, 1975; Teale and Sulzby, 1986). In addition, children's creative spellings, that is, how children come to terms with the alphabetic and orthographic rules of the notational system by child-driven explorations, was focused upon (Read, 1986; Treiman, 1993).

Generally, the development of writing skills has been studied from four main perspectives (Tolchinsky, 2004):

1. The sociocultural approach originated in the early work of Vygotsky (1978) and Luria (1978). Focusing on written language as tools of linguistic and cognitive change in the development of the individual, their work contributed to moving the study of early writing to the scientific arena of higher-order psychological functions. In a literate society, a child is surrounded by print that is brought to the child by adults and more competent peers mediating the written language. These ideas of Vygotsky's about mediated learning have inspired much modern pedagogy, for example, that writing skills are learned while children collaborate and interact verbally with adults or more competent peers, and that tools (artifacts) such as words, pencils, lap tops, and symbolic toys are crucial mediating elements. Moreover, they have vitalized research on how children construct ideas about written language via writing practices, through collaborative writing, children's think alouds, and verbal dialogs during writing.
2. The psychogenetic approach dates back to the theoretical and methodological work of J. Piaget and was recreated in the field of early writing by the groundbreaking work of

E. Ferreiro and her collaborators (e.g., Ferreiro and Teberosky, 1983). With a constructivist point of departure, this strand of research focuses on how children achieve mastery of the script by exploring and experimenting with the script as an object. Children reconstruct the conventions of written language through processes of assimilation and accommodation (see this). They build hypotheses about connections between oral and written language that are not idiosyncratic but developmentally ordered. By making use of controlled tasks in combination with clinical interviews, researchers within this line of research have shed light on how children's growing knowledge of literacy may be described in terms of developmental stages and how each stage reflects a dominant or favored strategy of writing.

3. The emergent literacy approach typically involves ethnographic researchers studying how literacy emerges during the writing process in classrooms and families in various cultures and social contexts. A primary concern is how children's emergent graphic signs are influenced by their interaction with parents or peers, or by their talking to themselves during the writing process, and how drawing, talking, and writing interact during the writing process (e.g., Bissex, 1980; Clay, 1975; Dyson, 1989; Teale and Sulzby, 1986). Often influenced by sociocultural theory of learning, researchers within this line of research have highlighted phenomena such as children's self-regulatory speech during writing, scaffolding interactions, writing conferences, etc. (e.g., Graves, 1983). However, ethnographic researchers have expanded this paradigm by focusing more broadly on literacy events, that is, on learning to write as an event with specific purposes that are culturally defined by certain ways of relating to other participants and by expected text topics and structures, such as how children (and adults) collaborate in composing a letter or a story (Heath, 1983). Shedding light on cultural variations and on the importance to literacy of both family literacy and classroom practices, this strand of research has had great implications for educational practice, for example, in the ramifications of terminology such as classroom events.
4. Approaches focusing on invented spelling typically concentrate on the product and process of writing in a more restrictive sense than the above-mentioned approaches. Spelling is seen as one component of writing (Treiman and Cassar, 1996) and concerns the ability to write words in accordance with orthographic rules. In line with this narrow definition, the term invented spelling regards child-driven activities, referring to the children's own experimentation with meaningful links and with connections between spoken and written language (Read, 1986). Many will side with Richgels' (1995: 99) definition of invented spelling as children's ability to "sound units in words and associate letters with those units in a systematic though

non-conventional way before being taught to spell or read." However, the term is also used with broader meanings (Saada-Robert, 2004), blurring its distinction from terms such as invented or emergent writing. Invented spelling typically regards a later developmental period, when the child experiments with the conventional notational system, that is, the alphabetic system, while invented writing typically refers to a larger developmental period including also pre-alphabetic periods. Important issues relating to research on invented spelling in the last 20–30 years include: the role of phonology and morphology in early spelling, the development of spelling, including the roles of logos, children's names, and letter naming in children's creative attempts to break the alphabetic code. More recently, there have been comparative studies of how orthographic systems influence early spellings (Nunes and Hatamo, 2004).

Within each of these approaches, there are large variations, and the distinguishing criteria are more pragmatic than logical. The four approaches have inspired researchers theoretically and methodologically. In practice, research projects transcend theories and methodologies. For example, while the process of writing within a sociocultural paradigm is commonly studied by ethnographic methodology (observation, interviews, etc.), semi-structured, test-oriented methods associated with the microgenetic approach of the Piagetian constructivist paradigm are also used. While invented spelling is quite often studied by test-oriented procedures, for example, dictation of single words over time, it may also be studied ethnographically via spontaneous writing of informal notes.

Stages of Early-Writing Development

Despite differences associated with specific characteristics of scripts and languages across orthographies and cultures, certain universal periods or stages of writing skills may be revealed, for example, in Hebrew (Levin and Landsmann, 1987), Italian (Pontecorvo and Zuccheromaglio, 1989), Spanish (Ferreiro and Teberosky, 1983), English (Clay, 1975), German (Brügelmann, 1999; Valtin, 1997), French (Jaffré and David, 1998), Norwegian (Elsness, 2001; Hagtvet, 2003), and Swedish (Liberg, 1993). These developmental periods share similarities with Piagetian stages to the effect that the strategies used by children at each stage determine the characteristics of that particular stage. However, this does not imply that the individual child makes consistent use of one strategy at the time. Rather, a child typically makes use of multiple strategies cutting across different stages, so the borders between stages are variable. Nevertheless, at a specific point in time, there appears to be a domination in a child's preference for one specific strategy, even though it may be used inconsistently.

So, a developmental stage is a period when there is a domination of a specific strategy.

Stages of early writing have been described using various terminologies depending on the focus of study and classification criteria for a stage. The basis for the developmental description below is a combined set of criteria reflecting qualities associated with the process and the product of writing (Hagtvet, 1989, 2003).

Prephonetic writing

In prephonetic writing, the graphic signs do not refer to phonetic entities in the child's speech. They reflect a weak awareness about the linguistic segments that constitute sentences (words, syllables, and phonemes). Serving a multitude of functions, prephonetic writing has been operationalized differently, depending on the perspective taken by the researcher and on the focus of study, for example, pretend writing, preliterate spelling (Henderson and Templeton, 1986), visual writing (Valtin, 1997), and precommunicative writing (Ellis, 1997). It is often produced as an imitation of adult writing or in symbolic play (pretend writing). Written from memory, the graphic signs are visual renderings of reality (visual writing) – often with an unfocused intention of communicating (precommunicative). However, even young children typically intend to communicate when writing. Therefore, the term precommunicative does not necessarily capture the essence of this early stage.

Children's early drawings and scripts symbolize referents, and drawing is seen as a precursor of scripts. According to Vygotsky: "... A child's memory does not yield a simple depiction of representational images at this age. ... A major feature of this mode is a certain degree of abstraction, which any verbal representation necessarily entails. ... drawing is graphic speech that arises on the basis of verbal speech. This gives us reasons for regarding children's drawing as a preliminary stage in the development of written language" (Vygotsky, 1978: 112–113). A big leap in development takes place when children come to understand that graphic signs can represent objects or events, for example, big circles, much text, or large letters refer to big, many, or long objects (Ferreiro and Teberosky, 1983).

Three different types of scripts are typically seen during this prephonetic stage reflecting different sub-stages or subperiods:

Scribbling and symbols with no similarity with letters

This form has also been labeled figurative writing (Valtin, 1997) alluding to the symbolic and imaginative nature of early script. Its relation to the referent is initially imprecise reflecting a weak understanding of the relationships between segments of speech and graphic scripts. At an early age (typically before age 3), scribbling is not differentiated from drawing. It is performed

rather arbitrarily – similar to an imitation of the writing movement observed in adults and for a purpose that appears to be an accompaniment to speech. Somewhat later, the scribbling becomes loaded with meaning – as observed when children follow their text with eyes and fingertips while reading their newly written text (Sulzby, 1986). Slowly, the graphic signs take on a more advanced, symbolic content – representing a more conscious and verbalized interpretation of things and events in the real world (around age 4). This is illustrated in **Figure 1** Leonard, age 3.4, has been asked to draw a picture, to write what he has drawn, and to write his name. Leonard's reading of his texts indicates that he differentiates between his drawing ("many big snakes"), his writing of the word snake, and his name (Leo). Four months previously, he did not know the difference between drawing and writing, and when asked to write his name, he said he did not know how to.

Letters in rows that represent words

When letters are written more or less randomly on the page, children typically remember and experiment with the form of the letters, but without phonetic awareness.

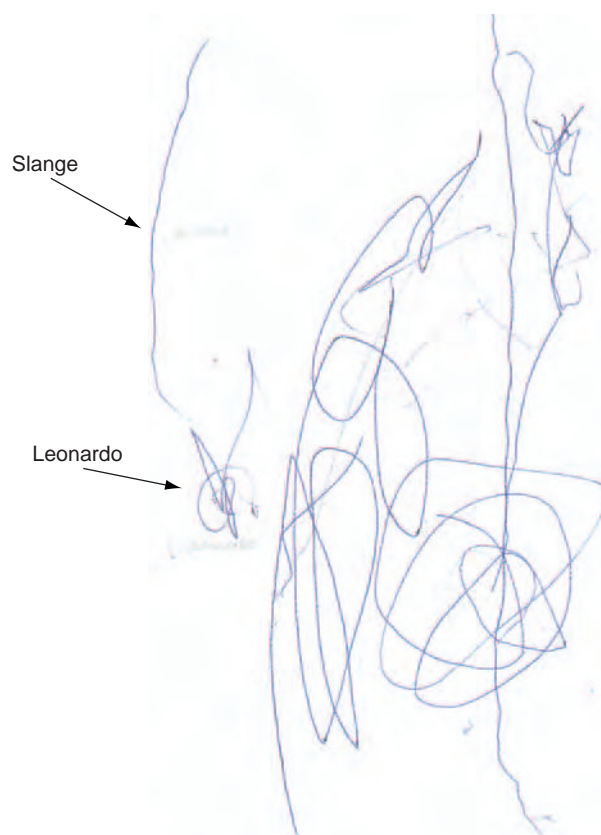


Figure 1 Leonard, age 3.4 draws and writes. From Hagtvet, B. E. (2003). *Språkstimulering. Tale og skrift i førskolealderen.* (Language stimulation. Speech and literacy before school age.) Oslo: Cappelen Akademisk.

Copying letters as pictures (sometimes from memory), this strategy has also been termed visual or logographic. **Figure 2** illustrates how letters may be used as pictures in a row when writing a Christmas card.

Global writing

When words are written as word pictures, the graphic scripts may be labeled global writing. Based on visual memory, this form of writing is still logographic. Own names and names of friends and family members are common among early words written by young children, but logos read on signs and posters (e.g., MILK, SHELL, and COKE) are also typical. This early writing of names commonly sets the stage for alphabetic exploration. Because these early names and word pictures are familiar and well-known in content – and, in most cases, emotionally important – they commonly act as catalysts for active and self-driven identification of phonemes in words. Slowly, children who regularly explore letter-sound relations in word pictures discover the alphabetic principle while writing.

Semiphonemic writing

Semiphonemic writing makes use of a phonetic strategy, but in a phonemically imprecise way: the child signals an awareness of the relationship between the phonological system and the script by means of graphic signs, pretend letters, or letters. An early phase of this developmental pattern is illustrated in **Figure 3** where Helene has written the sentence, *Helene liker is* (Helene likes ice cream). When reading the sentence, she underlines the syllabic pattern and rhythm of the sentence by means of voice punctuations and by tapping her finger tip on the relevant parts of the text revealing an awareness of the correspondence between word length and length of text: "He-le-ne li-ker is".

The discovery that different marks or signs in the text represent different aspects of speech (usually around

ages 3–4) is seen by Vygotsky as a shift in the development of written language from drawing things to drawing speech. From then on, spoken language disappears as the intermediate link between the world and the text, and written language slowly becomes as directly perceived as spoken language.

When letters are used in the semiphonemic stage, they are not copies of letters as pictures like they are in the prephonetic stage. Letters now represent sound patterns (sentences, words, onset, rhymes, and syllables) or language sounds (phonemes), and they now have a functional meaning to the child: they are used as building blocks in making word messages.

The writing of language sounds (phonemes) is a late phase in the semiphonemic stage. From then on, the writing becomes more and more phonemically oriented. First, the most prominent phoneme in the word is written down; each phoneme typically representing a larger unit of phonemes, for example, HS for house. The omission of vowels is a typical characteristic, as is the incorrect sequencing of phonemes or mirror imaging of words, for example SKRM for Markus.

Many children depend on letter naming during this early phonemic writing, for example, MSE (mess) or HLP (help) (Treiman and Cassar, 1997). However, this strategy may also be a barrier to proper phonemic writing by making the vowels less visible, which may explain why vowels are commonly omitted during this stage, for example, KM (come).

Phonemic writing

During the phonemic stage, children can identify phonemes in words, combine them with relevant graphemes, and write the graphemes in correct sequence. In early phonemic writing, children typically write as they say, tending to use this strategy systematically without



Figure 2 Experimenting with letters when writing a Christmas card, Christoffer, age 5.8. From Hagtvet, B. E. (2003). *Språkstimulering. Tale og skrift i førskolealderen*. (Language stimulation. Speech and literacy before school age.) Oslo: Cappelen Akademisk.



Figure 3 Syllabic writing, Helene, age 4.0.

orthographic information or spelling rules. This can yield charming examples such as RUDF? (Are you deaf?) (Bissex, 1980). They understand the alphabetic principle, but apply it orthodoxically phonemically . . .

Analyses of how early phonemic writing deviates from the adult norm have revealed that children's errors are quite logical. Children appear to actively construct a system of spelling patterns in accordance with the logic provided by their current level of skills. For this reason, it has been argued that spelling errors are windows into the child's written-language awareness, that is, to the systemic understanding of how the phonological structure of language matches the graphic structure of the script (Read, 1986).

Writing with conventional orthography

Increased experience with phonemic writing typically triggers children's interest in text and an acceleration of reading activities. This, in turn, generates an increased familiarity and awareness of the orthographic system and facilitates the inclusion of orthographic patterns in their writing (Treiman, 1994). The phonemic writing is then transformed into an orthographically driven process where orthographic representations are coordinated with phonological strategy. Jaffré and David (1998) see this developmental span from phonemic to orthographic spelling as two sub-phases, the first being a second logographic phase (in contrast to the logographic or visual phase described above). This is a phase where phonographic and morphological knowledge becomes integrated. Morphological knowledge is inherent in the orthographic system; it is based on rules and semantic insights. The second phase is the final orthographic phase.

The increasing awareness that the same sound can be spelled in one way when it represents one morpheme or has one grammatical meaning and in another way when it represents other ones, is demonstrated by children's unconventional or invented spellings (Read, 1986; Treiman, 1993). Children's spellings of regular past tense in English are good illustrations of this: the morphological rule being that past verbs end in -ed. In a longitudinal study of English children (Nunes *et al.*, 1997), children around age 6 ignored this ending and wrote past tense phonetically, for example, KIST for kissed. This was followed by an intermediate period when the children tended to avoid the -ed ending. Then, they used it too often, including non-verbs and irregular past verbs – in addition to some, but far from all, regular verbs. This phase is illustrated by the spellings of a boy aged 7.5 years: DRESD (dressed), FILLD (filled), SLEPED (slept), SOLED (sold), NECSED (next), and SOFED (soft). Following this period of overgeneralization, the children used the past tense with past verbs only, finally restricting its use to regular past tense verbs (Bryant and Nunes, 2004).

Individual Variations

Universality of stages

The extent to which these stages (prephonetic, semiphonemic, phonemic, and orthographic) are universal is debatable. Differences in language, culture, and school systems, no doubt, affect development. Literacy varies in importance and emphasis in different cultures (Heath, 1983). However, evidence suggests that despite cultural diversity, children in literate societies go through the same stages, but at a different pace (see above). Generally, there appear to be surprisingly many commonalities across cultures, languages, and orthographies.

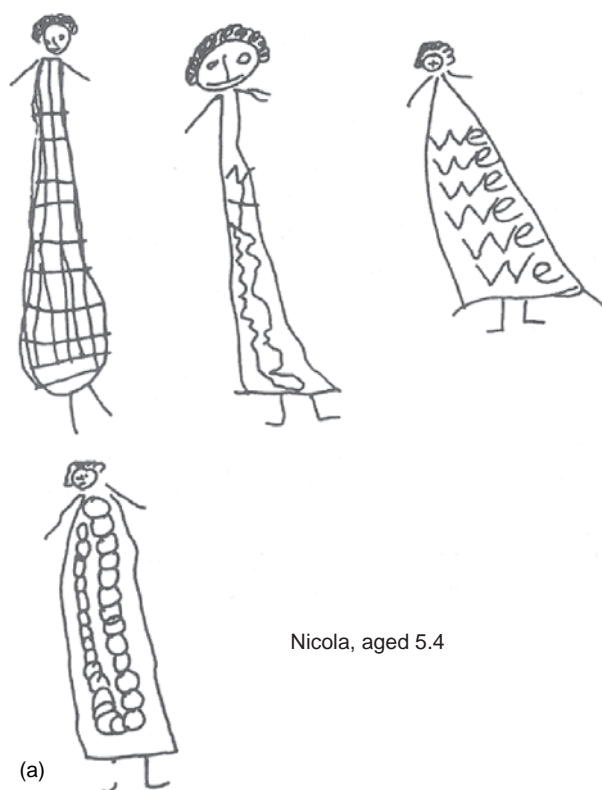
In addition, there are similarities in the onset of literacy development. It has been observed that the integration of scripts in children's drawings is suggestive of an increased interest in the written language on the part of the child. The universality of this observation, at least within the Western hemisphere, is illustrated in **Figure 4** where drawings by two girls age 5.4 are presented. Nicola grew up in New Zealand in the early 1970s while Benedicte grew up in Norway in the early part of this millennium. Nicola started school at age 5, while Benedicte started at age 7. Despite these differences of time and culture, their approach to early writing is surprisingly similar – in particular, in the style of their drawings and in their writing as an integral part of their drawing.

Variations within cultures

Despite general similarities, there are also noticeable differences – such as in the differences between boys and girls. Boys are often reported by their nursery school teachers to be less interested in paper-and-pencil activities than girls are. However, this may be an effect of the task more than biological differences. When encouraged to write on topics that appeal to them and allowed to make use of their own script, boys quite often produce more imaginative and detailed writing and drawings than girls do (Hagtvet, 2003). This is illustrated in **Figure 5** where Carl, age 5.3 has drawn a picture of the lives of sea animals and made up an explanatory text.

Girls appear generally less oriented toward details and events in the real world in their early writing. They are more inclined to tell stories about people, often inspired by fairy tales, such as tales about the prince and the princess, or a girl below a blue sky with a yellow sun as in **Figure 6**.

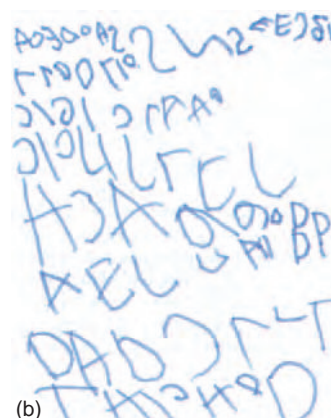
Carl and Melanie (Clare) approach the writing task differently, yet it is hard to rate one above the other. There are intragender differences as well. However, the gender differences in early writing are so striking that they invite more research attention – in particular, because boys tend to fall increasingly behind girls in literacy skills, according to recent international comparative studies (Programme for International Student Assessment (PISA)). Careful studies



Nicola, aged 5.4



Figure 4 Integrating scripts in drawings, Nicola, age 5.4 (New Zealand) and Benedicte (Norway), age 5.5. (a) Adapted from Clay, M. M. (1975). *What Did I Write? Beginning writing behaviour*. Heinemann: Auckland. (b) Adapted from Hagtvet, B. E. (2003). *Språkstimulering. Tale og skrift i førskolealderen*. (Language stimulation. Speech and literacy before school age.) Oslo: Cappelen Akademisk.



Carl, 5.3 år leser:
Der står haiene og der delfinene
og der står de derre bena
og så de derre bollene
også de der reunde ballene
også den gule ballen
også de små lekene
og regn og stener og pinner
og en til
kongler og pinner

(b)

Carl, 5.3 years reads while pointing:

It says the sharks and it says the dolphins
and it says those odd feet
and then those balls
and the yellow ball
and the small toys
and rain and stones and sticks
and one more
cones and pins

(c)

Figure 5 Carl, age 5.3: A boy's drawing and text. Text showing the child's reading of written text. From Hagtvet, B. E. (2003). *Språkstimulering. Tale og skrift i førskolealderen*. (Language stimulation. Speech and literacy before school age.) Oslo: Cappelen Akademisk.



Figure 6 Melanie, age 5.0: A girl's drawing and text. Text showing the child's reading of written text: "She was about to marry. She went out to pick flowers." From Hagtvét, B. E. (2003). *Språkstimulering. Tale og skrift i førskolealderen*. (Language stimulation. Speech and literacy before school age.) Oslo: Cappelen Akademisk.

of the importance of early writing for continued motivation for reading and writing appears to be a promising line of research.

Variations across languages

During the last 10 years, studies of literacy across languages have been a major focus of research (e.g., Joshi and Aaron, 2006; Perfetti *et al.*, 1997). In learning to spell orthographically, children are challenged differently by different orthographies. The closer the orthographic system depicts the pattern of pronunciation, the smoother is the road to mastery of conventional orthographic spelling (Wimmer and Landerl, 1997). For example, children learning to spell in languages with many written markers that have no corresponding pronunciation tend to be slow in conquering these silent orthographic patterns. This was demonstrated in French speaking children who typically learn to spell the silent endings in nouns, verbs, and adjectives rather slowly (Fayol *et al.*, 1999). On the other hand, even dyslexic Norwegian-speaking children making

use of a semi-structured orthography achieved mastery of regularly spelt words quite easily while irregularly spelled words were more difficult (Hagtvét and Lyster, 2003).

Summary

In conclusion, the pace and ways at which children come to terms with alphabetic and later orthographic writing reflects biological maturation as well as differences in languages, culture, and educational practice. When encouraged to write, children typically start writing once they know how to hold a pencil or use a computer, and literacy development is further supported by intervention means. Recent studies have shown that the quality of maternal mediation of children's writing in kindergarten affects children's literacy scores 2 years later (Aram and Levin, 2004). In addition, teaching of morphological rules to deaf children with spatially based sign language as mother tongue improved their orthographical spelling of

English (Burman and Pretzlik, 2004). These are examples underscoring the importance of active mediation to early writing. Mastery of orthographic spelling and skills in constructing word messages in different genres is typically achieved following several years of schooling, but for some children – for example, for dyslexic or deaf children – this is a continuous struggle.

Bibliography

- Aram, D. and Levin, I. (2004). The role of maternal mediation of writing to kindergartners in promoting literacy in school: A longitudinal perspective. *Reading and Writing: An Interdisciplinary Journal* **17**, 387–409.
- Bissex, G. I. (1980). *Gnys at Wrk*. Cambridge, MA: Harvard University Press.
- Brügelmann, H. (1999). From invention to convention: Children's different routes to literacy. In Nunes, T. (ed.) *Learning to Read: An Integrated View from Research to Practice*, pp 315–341. London: Kluwer Academic Press.
- Bryant, P. and Nunes, T. (2004). Morphology and spelling. In Nunes, T. and Bryant, B. (eds.) *Handbook of Children's Literacy*, pp 91–118. London: Kluwer Academic Publishers.
- Burman, D. and Pretzlik, U. (2004). Paths to literacy for deaf British Sign Language (BSL) users. In Nunes, T. and Bryant, B. (eds.) *Handbook of Children's Literacy*, pp 741–766. London: Kluwer Academic Publishers.
- Clark, M. M. (1976). *Young Fluent Readers*. London: Heinemann.
- Clay, M. M. (1975). *What Did I Write? Beginning Writing Behaviour*. Heinemann: Auckland.
- Dyson, A. H. (1989). *Multiple Worlds of Child Writers*. New York: Teachers' College Press.
- Ellis, N. (1997). Interactions in the development of reading and spelling: stages, strategies, and exchange of knowledge. In Perfetti, C. A., Rieben, L., and Fayol, M. (eds.) *Learning to Spell*, pp 271–294. Mahvali, NJ: Erlbaum.
- Elsness, T. F. (2001). *Stavestrategier hos barn i alderen 7–8 år*. (Spelling strategies in children aged 7–8 years.) Doctoral dissertation. Department of Educational Research: University of Oslo.
- Fayol, M., Thenevin, M.-G., Jarousse, J.-P., and Totereau, C. (1999). From learning to teaching to learning French written morphology. In Nunes, T. (ed.) *Learning to Read: An Integrated View from Research and Practice*, pp 43–63. London: Kluwer Academic Publishers.
- Frith, U. (1985). Beneath the surface of developmental dyslexia. In Patterson, K. E., Marshall, J. C., and Coltheart, M. (eds.) *Surface Dyslexia*, pp 301–330. London: Routledge and Kegan Paul.
- Graves, D. H. (1983). *Writing: Teachers & Children at Work*. London: Heinemann.
- Hagtvet, B. E. (1989). Emergent literacy in Norwegian six-year-olds. From pretend writing to phonemic awareness and invented writing. In Biglmaier, F. (ed.) *Reading at the Cross-Roads. Conference Proceedings, The 6th European Conference on Reading*, 31 July to 3 August, 1989, Berlin.
- Hagtvet, B. E. (2003). *Språkstimulering. Tale og skrift i førskolealderen*. Language stimulation. (Speech and literacy before school age.) Oslo: Cappelen Akademisk.
- Hagtvet, B. E. and Lyster, S. A. H. (2003). The spelling errors of Norwegian good and poor decoders: A developmental cross-linguistic perspective. In Goulandris, N. (ed.) *Dyslexia in Different Languages. Cross-Linguistic Comparisons*, pp 181–207. London: Whurr Publishers.
- Heath, S. B. (1983). *Ways with Words*. Cambridge: Cambridge University Press.
- Henderson, E. H. and Templeton, S. (1986). A developmental perspective of formal spelling instruction through alphabet, pattern and meaning. *Elementary School Journal* **86**(3), 305–316.
- Jaffré, J.-P. and David, J. (1998). Premières expériences en littérature. (First experiences in literacy). *Psychologie et Education* **33**, 47–61.
- Joshi, R. M. and Aaron, P. G. (2006). *Handbook of Orthography and Literacy*. Mahwah, NJ: Erlbaum.
- Levin, I. and Landsmann, L. T. (1987). Becoming literate: Referential and phonetic strategies in early reading and writing. *European Journal of Behavioural Development* **12**, 369–384.
- Liberg, C. (1993). *Hur barn lär sig läsa och skriva*. Lund: Liber Läromedel.
- Luria, A. R. (1978). The development of writing in the child. In Cole, M. (ed.) *The Selected Writings of A.R. Luria*, pp 145–194. White Plains, NY: M.E. Sharpe.
- Montessori, M. (1971). *The Montessori Elementary Material*. Cambridge, MA: Robert Bentley.
- Nunes, T., Bryant, P., and Bindman, M. (1997). Spelling and grammar – the nected move. In Perfetti, C. A., Rieben, L., and Fayol, M. (eds.) *Learning to Spell*, pp 151–170. Mahvali, NJ: Erlbaum.
- Nunes, T. and Hatamo, G. (2004). *Morphology, Reading and Spelling: Looking across Languages*. In Nunes, T. and Bryant, B. (eds.) *Handbook of Children's Literacy*, pp 651–672. London: Kluwer Academic Publishers.
- Perfetti, C. A., Rieben, L., and Fayol, M. (1997). *Learning to Spell*. Mahvali, NJ: Erlbaum.
- Pontecorvo, C. and Zuccheromaglio, C. (1989). From oral to written text: How to analyze children dictating stories. *Journal of Reading Behavior* **2**, 109–125.
- Read, C. (1986). *Children's Creative Spellings*. London: Routledge and Kegan Paul.
- Richgels, D. J. (1995). Invented spelling ability and printed word learning in kindergarten. *Reading Research Quarterly* **30**, 96–109.
- Saada-Robert, M. (2004). Early emergent literacy. In Nunes, T. and Bryant, B. (eds.) *Handbook of Children's Literacy*, pp 575–598. London: Kluwer Academic Publishers.
- Sulzby, E. (1986). Writing and reading: Signs of oral and written language organization in the young child. In Teale, W. H. and Sulzby, E. (eds.) *Emergent Literacy: Writing and Reading*, pp 50–89. Norwood, NJ: Ablex Publishing.
- Teale, W. H. and Sulzby, E. (1986). Introduction: Emergent literacy as a perspective for examining how young children become writers and readers. In Teale, W. H. and Sulzby, E. (eds.) *Emergent Literacy: Writing and Reading*, pp vii–xxv. Norwood, NJ: Ablex Publishing.
- Tolchinsky, L. (2004). Childhood conceptions of literacy. In Nunes, T. and Bryant, B. (eds.) *Handbook of Children's Literacy*, pp 11–29. London: Kluwer Academic Publishers.
- Treiman, R. (1993). *Beginning to Spell: A Study of First Grade Children*. New York: Oxford University Press.
- Treiman, R. (1994). Use of consonant letter names in beginning spelling. *Developmental Psychology* **30**, 567–580.
- Treiman, R. and Cassar, M. (1996). Effects of morphology on children's spelling of final consonant cluster. *Journal of Experimental Child Psychology* **63**, 141–170.
- Valtin, R. (1997). Strategies of spelling and reading of young children learning German orthography. In Leong, C. K. and Joshi, R. M. (eds.) *Cross-Language Studies of Learning to Read and Spell*, pp 175–193. London: Kluwer Academic Publishers.
- Vygotsky, L. S. (1978). The prehistory of written language. In Cole, M., John-Steiner, V., Scribner, S., and Soubberman, E. (eds.) *Mind in Society*, pp 105–119. Cambridge, MA: Harvard University Press.
- Wimmer, H. and Landerl, K. (1997). How learning to spell German differs from learning to spell English. In Perfetti, C. A., Rieben, L., and Fayol, M. (eds.) *Learning to Spell*, pp 81–96. Mahvali, NJ: Erlbaum.

Further Reading

- Farr, M. (1985). *Advances in Writing Research, Volume One: Children's Early Writing Development*. Norwood, NJ: Ablex Publishing.
- Ferreiro, E. and Teberosky, A. (1983). *Literacy before Schooling*. London: Heinemann Educational Books.

First Language Acquisition

N B Ratner, University of Maryland, College Park, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Affix – It is an element appended to a root word form, which may convey grammatical information (e.g., plural, past tense) or lexical information (e.g., un-, re-).

Dative sentence – It is a sentence form in which the normal sequence of direct object and indirect object has been inverted (e.g., I gave the present to the girl → I gave the girl the present).

Figurative language – It refers to the nonliteral use of language (e.g., a cold woman).

Finiteness – It is an attribute of verb forms that convey person, tense, number, and mood (e.g., he runs).

Formants – They are concentrations of acoustic energy caused by the resonant properties of the vocal tract that are particularly critical for distinguishing among vowels in a language.

Indirect language – They are linguistic forms that do not directly convey intended meaning, such as use of “it’s hot in here” to request that someone open a window.

Inflection – It refers to a grammatical affix.

Metalinguistic ability – It is the ability to use language to talk about the characteristics and use of language.

Morpheme/morphology – A morpheme is the smallest meaningful unit in a language, for example, words and affixes; morphology is the study of word structure.

Over-extension – It is the use of a word to refer to a larger category of referents than is appropriate, for example, calling all four-legged creatures doggie.

Passive sentence – It is a sentence form in which the order of subject and object are inverted, as in John(s) loves Mary(o) → Mary is loved by John.

Phonemic awareness – It refers to the ability to appreciate the fact that spoken words are composed of individual sounds (phonemes).

Phoneme/Phonology – A phoneme is a minimally contrastive sound in a language; phonology is the study of the sound system of a language.

Phonotactics – It refers to the permissible ordering of phonemes within a language.

Pragmatics – It is the study of language use in socially appropriate ways, which may vary with context, addressee, and other factors.

Pronominal reference – It refers to the use of pronouns to refer back to previously mentioned nouns or concepts (also called anaphora or anaphoric reference).

Segmental/supra-segmental – It is the distinction between information conveyed by sounds and words (segments) and prosodic characteristics such as intonation, volume, and stress that may extend over more than one segment.

Semantics – It is the study of meaning.

Specific language impairment – It is a developmental disorder in which comprehension and use of language is below age-expectations, in the absence of observable deficits in intellectual, hearing, or motor function.

Syntax – It is the study of sentence structure and grammatical morphology.

Vocabulary burst – It refers to a stage early in children’s vocabulary development during which words are learned at a very rapid pace.

The Nature of Human Language

Human languages are distinct from animal communication systems in a wide variety of ways. Among them are infinite creativity (the ability for speakers and hearers to produce and understand an infinite variety of utterances), their symbolic nature (the arbitrary relationships among words, utterances, and the concepts to which they refer), and hierarchical organization, which allows a number of levels of rules governing appropriate structure and use (see Fromkin *et al.*, 2007).

Within any language, there is a set of rules that governs appropriate use of sounds, words, grammar, and meaning. Moreover, competent language users must also master socially appropriate means of conveying and interpreting linguistic messages. Briefly, these subsystems of language knowledge consist of phonology, morphology, syntax, semantics, and pragmatics.

Phonological features of any given language specify its sound inventory (phonemes) as well as ways in which sounds may be legally combined to create well-formed words (phonotactics). The smallest units of language that convey meaning or grammatical distinctions are morphemes; for example, in English, a word such as *cats*

consists of one lexical morpheme, *cat*, which can stand alone, and one grammatical morpheme to signal the plural. Languages have large numbers of lexical (or open class) morphemes, and a much smaller and delimited number of grammatical morphemes (closed class), which may stand alone (such as *the* or *can*), or must be attached to lexical morphemes (such as the plural, past tense, possessive, etc.).

Mastery of syntax requires appropriate use of morphology as well as any rules governing the ordering of elements in sentences and their smaller constituents, such as noun and verb phrases. Some languages, such as Finnish, permit fairly free word order, while others are highly constrained. In addition, languages may differ in basic word order; for example, English tends to employ subject–verb–object as its canonical ordering, while the Philippine language Tagalog is primarily verb–subject–object.

Meaning in language can be conveyed by the meanings of individual words (as in knowing what the word *chair* refers to), as well as the order in which words are combined to reflect themes such as subject and object (most readers will readily appreciate that “John loves Mary” does not necessarily mean that “Mary loves John”). Finally, the meanings of sentences often go beyond the strict interpretation of their words and syntax. Pragmatic intent is obtained by evaluation of the sentence within a context to ascertain its function within conversation – whether one’s objective is to inform, warn, request action, etc., as might variously be the case in hearing someone say, “It’s late.” All of these rule systems must be adequately mastered in order for the child to function as a capable speaker–hearer of a language.

Techniques in Understanding Language Development

Facts about the typical milestones in language acquisition are obtained from a variety of sources. Historically, diary data and small, longitudinal studies offered the first insights into stages and strategies in child language development. This tradition has been continued with the broad use of open access data archives such as the Child Language Data Exchange System (CHILDES, <http://childes.psy.cmu.edu>) by developmental psycholinguists to investigate new questions about children’s behaviors using existing data from a broad variety of languages and populations.

Observational data are augmented by experimental or quasi-experimental studies in which production or comprehension of various language targets is elicited. More recently, the use of electrophysiological measures, such as event-related potentials (ERPs), eye tracking, and functional magnetic resonance imaging (fMRI), have

enabled researchers to explore the underlying substrates of infant and child linguistic performance (Karmiloff and Karmiloff-Smith, 2001).

Milestones in Speech and Language Development

Despite some degree of individual variation in timing and sequencing of the development of specific linguistic abilities, there is considerable uniformity in children’s language development. This predictability in sequencing permits educators and other professionals working with infants and children to apply developmental expectations to assessment and, if necessary, intervene with children not meeting typical pacing or sequencing of skill development (Oller *et al.*, 2006). Knowledge of typical developmental milestones also permits educators to construct appropriate curricular goals and strategies for all activities involving language and reading which build upon earlier linguistic skill development.

Pre-Linguistic Achievements

Language learning begins right from the womb. There is evidence that a fetus can be conditioned to recognize the rhythm and cadence of stories repeatedly read aloud in the last trimester of pregnancy; shortly after birth, use of non-nutritive sucking paradigms show that babies demonstrate a marked preference for the voices of their own mothers, and samples of conversational speech in their own language, rather than a foreign language or dialect.

In early infancy, babies show a capacity to distinguish linguistic contrasts used by all the world’s languages, rather than merely their own. For example, they can distinguish contrasts in voicing, place, and manner of consonants, as well as between vowels very similar in formant characteristics. However, by approximately 10 months of age, such discrimination ability wanes, as the infant begins to narrow contrasts to those used by his or her native language. Concurrently, their expressive babble begins to reflect segmental (sound) and supra-segmental (prosodic) characteristics of the ambient language or languages to which they are exposed.

Use of a number of laboratory paradigms has shown that the very young infant is able to begin to discriminate individual words within the conversational speech stream, as well as features within utterances that signal possible word boundaries, an important skill that will enable the child to begin to map early vocabulary and grammar. This ability to segment the speech signal is critical because even infant-directed speech typically takes the form of multiword sentences, with few words produced in isolation. Some milestones in segmentation skills occur relatively early. For example, typically developing infants

usually can discriminate their own names from similar-sounding foils, even in noise, by 4 months of age, while children later diagnosed with developmental language disorders may not show such ability. By 7–8 months of age, if an infant is familiarized with a novel word in isolation, she will listen longer to short spoken passages containing that word, and the inverse is also true. This discrimination is quite precise, and infants will not falsely respond to words that are similar, but differ by only one speech sound. Segmentation is not an all-or-none skill, and appears to rely on development of sensitivity to certain language-specific cues, such as the typical stress patterns within words and among words in sentences, as well as more general statistical regularities (such as co-occurrences of sounds or syllables). Word discovery progressively enables further and finer segmentation of elements in input, creating a circular relationship between segmentation and lexical development, with segmentation skill leading to enhanced word-learning, which in turn improves segmentation skill and future word learning.

Speech segmentation abilities may be quite delayed in children with identifiable syndromes known to be accompanied by significant cognitive and linguistic deficits, supporting the idea that segmentation may be a necessary precursor to normal language development. Infants who appear to be developing normally but fail to show typical segmentation abilities have significantly poorer language outcomes than infants who succeed in these tasks.

Production and Understanding Prior to First Words

Infant babble and conversational turn taking during vocal interactions with caretakers increasingly reflect the typically developing child's knowledge of both the shape of the ambient language as well as pragmatically appropriate eye gaze and gestures; these patterns appear to be mutually reinforcing, with more vocal children typically receiving a greater degree of conversational input from their caretakers. Failure for infants to develop appropriate eye-contact, reciprocal vocalization patterns, and gesture may be an early indicator of developmental delay, particularly autism spectrum disorder.

Typical early babble uses prosodic features of the language to convey pragmatic intent, such as requesting, notice, or displeasure, well before actual identifiable words are used for such functions; this pattern of vocalization is sometimes called jargon. By about 10 months of age, many infants will have developed proto-words, phonetically consistent forms that are regularly used to convey a particular message.

At about this same age, an important change occurs in infants' social cognition. They begin to intuit other people's thoughts, desires, and goals (sometimes called a theory of mind). They follow the direction of a point

(unlike animals, who will simply fixate on the hand doing the pointing), and will begin to follow others' gaze to an object or action of interest. Both behaviors signal the infant's ability to engage in joint attention. The pattern of responsiveness by both infants and adults is important to the rate of language development; inability of the infant to engage in joint attention is a marker of developmental delay, while the adult's ability to discern the infant's interest in objects or activities and comment on them contingently and responsively has been linked to more rapid achievement of early linguistic milestones.

The Nature of Early Word Production and Comprehension

Well before the production of first recognizable words of the language, usually at 12–15 months of age, infants use their segmentation skills to construct a fairly broad receptive vocabulary that can be assessed using parental report inventories. First words are usually not pronounced in an adult-like fashion, but are used consistently to convey intent. Across most languages that have been observed, first words are highly likely to be names of objects or people (a preference for nouns) that they actively interact with (rather than being merely the most frequently used word in a language), but individual children may show a preference for more conversational vocalizations, such as greetings, or terms used in play with adults. Semantic representation of first words is likely to be under-specified, leading to over-extension, in which words are applied to an overly broad category of referents (e.g., *doggie* for many additional types of four-legged, furry animals). However, word learning is governed by a number of principles that seem quite uniquely human, such as a tendency to associate labels with the entirety of a reference, rather than one of its component features, and the presumption that new words refer to novel, rather than known, referents.

Infants and toddlers comprehend more words than they tend to produce, and show good evidence of understanding general rules of syntax (such as preferred word order of the language, signaling the difference between "Big Bird pushes Ernie" and "Ernie pushes Big Bird") in laboratory tasks well before they produce multi-word utterances themselves. By 2 years of age, the typically developing child will have achieved at least a 50-word expressive vocabulary (the average is approximately 200) and begin to combine words into two-word utterances. Children who have failed to reach this stage by 24 months are significantly at risk for a variety of communicative disorders, although a substantial proportion of late talkers do appear to recover and later function within the normal range.

Early two- and three-word utterances are likely to lack appropriate grammatical affixes in a language such as

English, in which they are relatively sparse. Acquisition of the relatively closed set of grammatical morphemes in English has been tracked extensively, beginning with Roger Brown and his colleagues at Harvard in the early 1960s, and it follows a relatively consistent order and trajectory (see **Table 1**).

Classic studies, such as that conducted by Jean Berko in the late 1950s demonstrate that the learning of these morphemes represents abstraction and generalization of rules which can be extended to novel exemplars (see **Figure 1**) and even mis-applied to exceptional forms

Table 1 Typical order of acquisition for 14 common English morphemes

Morpheme	Example
Present progressive	Baby crying
Prepositions in/on	Cookie on table; Mommy in car
Regular plural	Cats, dogs, dishes
Past irregular (learned as lexical items)	Early examples: came, fell, broke, sat, went
Possessive	Daddy's car
Uncontractible copula	Eve is girl
Articles (a, the)	See the kitty, Give a cookie
Past regular	Hugged, kissed, patted, cried
Third person regular	Boy runs, doggie barks
Third person irregular	He does/has it
Uncontractible auxiliary	I am going.
Contractible copula	What's that
Contractible auxiliary	Puppy's eating

Mastery may span from 1 year and 6 months to 4 years in typically developing children.
Adapted from Brown (1973).

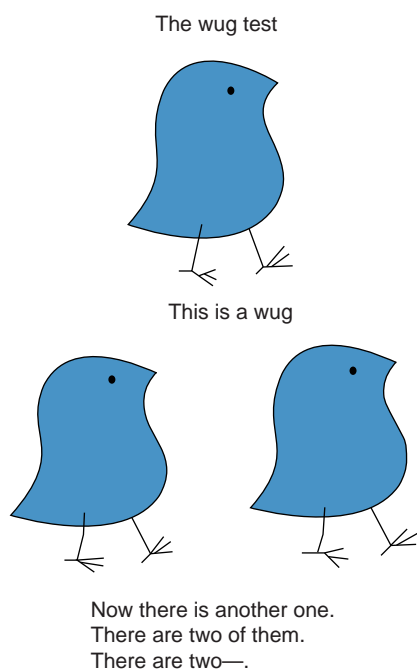


Figure 1 The Wug test.

(e.g., when a child says *foots* or *wented*), rather than rote memorization of items the child has overheard. Among the most difficult morphemes are those conveying tense and finiteness, and these continue to be problematic for children later diagnosed with specific language impairment. Children learning highly inflected languages will show earlier productive use of grammatical elements, even during the single-word stage.

While it may take many months for the child to compile the initial 50-word vocabulary, this achievement is usually followed by what some have termed a vocabulary burst, in which hundreds of words are acquired within a very short space of time. Most children entering kindergarten have vocabularies upward of 14 000 words, which may represent both unique roots as well as morphologically complex forms such as *un+comfort+abl+y*.

Combining Words into Utterances

In English, mastery of basic sentence forms tends to proceed from those types of utterances requiring few verbal auxiliaries, such as the imperative, to those that require manipulation of, or changes and additions to the typical components in declarative sentences, as in negative utterances (*He is/isn't going, I like/don't like spinach*), and finally, those that additionally require re-ordering or permutation of elements, as in questions (*Is he going? Doesn't he like ice cream?*). Mastery of *wh*-questions (e.g., *Where is he going?*) requires additional skills, including further understanding of the specific meanings conveyed by question terms such as *what*, *when*, *where*, and *why*, with the latter requiring perhaps the greatest level of sophistication.

Progressively longer and more complex utterances are formed by conjoining, using conjunctions, and embedding. Conjoining is typically first observed in object clauses, then subject clauses ("He likes cookies and milk vs. Daddy and Mommy are coming home"). The conceptually simpler conjunctions such as *and* are mastered before those requiring higher levels of understanding of causality (*because*), time (*before/after*), or conditionality (*if*). Prior to full understanding of such terms, the child is likely to respond to clauses linked using later-learned conjunctions using an order-of-mention strategy, in which the first clause is presumed to occur before the second.

In a language such as English, which shows a strong preference for subject–verb–object ordering, sentences deviating from this order are likely to lead to misinterpretation, or reliance on conceptual plausibility. Thus, inaccurate comprehension of the subjects of verbs in passives ("Big Bird (object) is pushed by Ernie (subject)"), direct objects (DO) and indirect objects (IO) in datives ("Mary showed the baby(IO) the kitty(DO)" vs. the more canonically ordered "Mary showed the kitty (DO) to the baby (IO)"), center-embedded relative clauses ("The man who lives next to my

sister is a doctor”), and other exceptional constructions (e.g., “The boy is easy to see”), is quite likely, at ages up to 8 years or so, depending upon situational context and other cues provided by the specific words in the sentence.

Later Acquisitions

As the child nears kindergarten entry and progresses through the early school years, refinements occur in phonological, lexical, and grammatical skills. Residual problems in adult-like pronunciation of late-acquired sounds (in English, e.g., sounds such as /r/ and /l/, consonant clusters containing these sounds, and /s/) subside before age 8 years. In lexical development, the ability to comprehend relative terms (those with no fixed meaning, that require perspective, such as *big* or *daughter*), and those which convey abstract concepts (such as *honest*) emerges, as does the awareness that words may be ambiguous (having more than one meaning). This last skill enables elementary school-aged children’s growing appreciation of jokes and riddles. The ability to decompose and assemble multi-morphemic forms (e.g., *anti+dis+establish+ment+arianism* or *mono+the+ism*) will continue across the school years and is critical to continued development of linguistic skills required of skilled readers. Skilled readers will also need to expand vocabulary mappings to include figurative and metaphorical uses of language common to advanced literary forms.

Other skills required in the academic setting include the ability to make sound–symbol associations (phonemic awareness) in the child’s language and orthography necessary for the decoding of the written word. Fluency in word decoding will need to be achieved in order for the child to retrieve the meanings of utterances before information fades from the short-term memory store.

Written language skills will require mastery of more complex language forms than those typically used in conversation, and the student’s writing will increasingly demonstrate frequent use of compound and complex sentences. In addition, success in writing for academic purposes will depend upon the child’s ability to comprehend complex text, including passages with long-distance pronominal references (also known as anaphoric reference, as in “The European colonization of the Americas in the fifteenth and sixteenth centuries resulted in progressive decimation of native populations, as *it* introduced diseases to which they had no natural resistance. . .”). Students are also expected to demonstrate metalinguistic knowledge of terminology for the forms they already understand and produce (e.g., notions such as subject–verb agreement). Finally, both oral and written assignments will require them to demonstrate facility with different genres of language use, such as narratives and formal expository text. Most classroom writing assignments will also require the child to comprehend text more linguistically sophisticated

than that he or she conventionally produces, and to then paraphrase that information into a unique, but accurate recast of the original wording found in source materials.

While even the youngest child has control over a variety of pragmatic functions, including the ability to request, argue, explain, engage in appropriate turn taking, etc., socially appropriate use of language continues to develop over the school years. Acceptable ways to convey politeness or deference will grow from relatively simple strategies (such as the use of *please*) to the use and comprehension of indirect forms (as in knowing that the phrase, “Yum, that looks good” is an appropriate way to convey a desire to receive a bite of the coveted food item).

Cross-Linguistic Variation in Profiles of First-Language Acquisition

It is both theoretically and practically relevant to understand that patterns of acquisition in one language do not necessarily map directly to others. The order in which aspects of grammar, in particular, are learned in a specific language may be impacted by a number of factors. These may include the pervasiveness of a linguistic concept (e.g., whether or not verbs are invariably marked for person and number as in Spanish, contrasted with relatively sparse marking in English). Other considerations may include conceptual simplicity of a grammatical form (whether a language distinguishes between singular and plural only, or requires some marking of physical number), and whether or not a grammatical rule is relatively uniform or is characterized by numerous exceptions. The study of different trajectories seen in languages which differ typologically can do much to identify which aspects of acquisition appear dependent upon language-specific features, generalized cognitive development and ability, or input characteristics, and which appear relatively uniform regardless of the language being learned. When relatively invariant patterns or strategies emerge across language differing widely in their rule systems, they provide evidence of innate predispositions that govern the process of language acquisition (see section titled “Theories of language acquisition”).

Practically speaking, knowledge of the specific stages and strategies in acquisition of individual languages enables the important process of assessing the adequacy of children’s progress in language learning, and the identification and remediation of delayed or disordered language development. It is already abundantly clear that even within a single language, dialectal variation may compromise the sensitivity or appropriateness of language-assessment instruments. Additionally, because concepts and rules are acquired at different points and in differing orders across languages, a simple translation of a test written in

English, for example, regardless of its solid psychometric properties when used with English-speaking children, will be inappropriate in assessing the development of a child learning Spanish or Urdu.

Individual Variation

While the many generalizable patterns of language learning within a specific language community enable us to have appropriate expectations of children of a given age, individual variation does exist in how children approach the task of language acquisition. Some children appear to be more attracted to learning the names of things early in lexical development, while some appear more attuned to the social and conversational uses of language. Some children appear to map the prosodic characteristics of adult language when producing early utterances, while others appear to pay more attention to the segmental properties of target words. Some children appear to make relatively rapid progress in acquiring an initial vocabulary, while some appear to accumulate first words more slowly. However, precocious language learners do not appear to keep their verbal advantage over the life span. In contrast, a significant number of slower language learners do continue to experience long-term delay (see section titled ‘Developmental language disorders’).

Theories of Language Acquisition

Given the rapidity with which a child learns his or her first language, it is not surprising that multiple theories of language acquisition have been advanced and debated. The traditional contrast has been between those researchers and theorists who attribute the major impetus in language development to innate or nativist predispositions and those who reserve a relatively larger role for social-interactionist factors, such as the nature, quality, and quantity of verbal interaction the child receives. The first approach is most classically associated with Noam Chomsky and Steven Pinker, while the second is often attributed to Jerome Bruner and Catherine Snow. In most respects, the contrast between the positions is artificial. There clearly appear to be features of language learning, particularly in the domains of syntax and the course of phonological development, that are constrained by innate strategies and are not very easily accounted for by aspects of the nature of input addressed to the child, nor generalized cognitive strategies. In contrast, aspects of a language such as its vocabulary and pragmatic conventions can only be learned through exposure, although some aspects of lexical learning appear constrained by apparently innate strategies.

Additional theories of language learning often appeal to more generalized cognitive strategies for the discovery and abstraction of linguistic rules and regularities. Among them are connectionist models of language acquisition, which liken the process to that seen in learning by computing networks. Researchers most highly identified with what might be termed an information-processing approach include Brian MacWhinney and the late Elizabeth Bates.

Developmental Language Disorders

Estimates vary, but between 6% and 10% of preschool and school-aged children demonstrate problems in language acquisition. Some delayed or deviant skills may be attributable to primary problems in intellectual development, hearing impairment, or autism spectrum disorder, but others appear relatively unique to language, thus leading to a diagnosis of specific language impairment (SLI). SLI has significant ramifications for later educational achievement, as it shows high co-morbidity with dyslexia, but also impacts later-reading/writing skills development and use that are reliant upon linguistic knowledge. Chief among these are the abilities to make phoneme–grapheme correspondences, decode complex multi-morphemic words, as well as to comprehend and use complex syntax. Weaknesses in decoding ambiguity, following pronominal reference and resolving metaphorical and figurative usages of language are common.

Early identification of language-learning disorders and appropriate intervention are crucial during the preschool and early elementary school years because of impacts on later school achievement. SLI is highly heritable and children at familial risk should be carefully monitored.

See also: An Overview of Language and Literacy in Educational Settings; Attention in Cognition and Early Learning; Classroom Discourse and Student Learning; Early Writing; Learning to Read; Reasoning and Explanatory Talk; Learning in Everyday Settings.

Bibliography

- Brown, R. (1973). *A First Language*. Cambridge, MA: Harvard University Press.
- Fromkin, V., Rodman, R., and Hyams, N. (2007). *An Introduction to Language*, 8th edn. Boston, MA: Thomson Wadsworth.
- Karmiloff, K. and Karmiloff-Smith, A. (2001). *Pathways to Language*. Cambridge, MA: Harvard University Press.
- Oller, J. W., Oller, S., and Badon, L. (2006). *Milestones: Normal Speech and Language Development across the Lifespan*. San Diego, CA: Plural.

Further Reading

- Berko, J. (1958). The child's learning of English morphology. *Word* **14**, 150–177.
- Gleason, J. B. and Ratner, N. B. (2008). *Language Development*, 7th edn. Boston, MA: Allyn and Bacon.
- Golinkoff, R. and Hirsh-Pasek, K. (1999). *How Babies Talk: The Magic and Mystery of Language Acquisition*. New York: Dutton/Penguin.
- Hoff, E. (2005). *Language Development*, 4th edn. Boston, MA: Thomson Wadsworth.
- Hulit, L. and Howard, M. (2006). *Born to Talk: An Introduction to Speech and Language Development*, 4th edn. Boston, MA: Allyn and Bacon.
- Jusczyk, P. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

- Nippold, M. (1998). *Later Language Development: The School-Age and Adolescent Years*, 2nd edn. Austin, TX: Pro-Ed.
- Paul, R. (2006). *Child Language Disorders: From Infancy through Adolescence*, 3rd edn. St. Louis, MO: Mosby/Elsevier.
- Traxler, M. and Gernsbacher, M. (2006). *Handbook of Psycholinguistics*. Amsterdam: Academic.

Relevant Website

- <http://www.childes.psy.cmu.edu> – The CHILDES (child language data exchange) Project.

Foreign Language Learning

C Kinginger, Pennsylvania State University, University Park, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Current Challenges in Research on Foreign-Language Learning

In the latter decades of the twentieth century, the dominant paradigm for both foreign- and second-language acquisition (SLA) research, derived from psycholinguistics, has been the input–interaction–output (IIO) perspective focusing on development of second-language grammars via mechanisms considered analogous to those of computer-information processing. Within this paradigm, language acquisition is viewed as a learner-internal cognitive process by which language input in the environment is entered into a transitional system (or interlanguage) governed by universal processing constraints. Input may be tailored for the individual acquirer, particularly in the context of conversational interaction, but the process itself is impervious to influence from the social environment, including the classroom and the efforts of teachers. Evidence for acquisition is provided by the extent to which the system's output approximates native speaking norms. Computational models of SLA emphasize the autonomy of the human individual while also validating communicative interaction as a source of learning. Research in this paradigm has yielded valuable insights on the sources of systematicity and the difficulty in the development of grammatical competence. However, due to their exclusive focus on psycholinguistic processes, computational models deliberately set aside explanation of relationships between the learner and the social context.

Many researchers continue to find inspiration and to perform productive research within the IIO perspective. However, in response to calls for more socially grounded theories, research on foreign-language learning has undergone considerable expansion in the variety and scope of its theoretical perspectives over the course of the past decade. Researchers eager to redress a perceived imbalance favoring the cognitive over the social aspects of language learning have introduced a range of descriptive or explanatory approaches from sociology (conversation analysis), anthropology (language socialization), literary study (narrative analysis), corpus linguistics, and sociocultural/historical psychology. Rather than a uniform monoculture, the field now resembles the flower garden predicted in Lantolf's (1996) metaphor, where diverse and quite distinct approaches grow – and compete for space – alongside one another. It is hardly surprising, in such an environment, that the more influential writers in the field argue in favor of an ecological approach where language learning is conceptualized as a relational human activity, “co-constructed

between humans and their environment, contingent upon the position in space and history, and a site of struggle for the control of social power and cultural memory” (Kramsch, 2002: 5).

Regardless of the precise form it takes, the need for conceptual expansion is highlighted by changes in the practice of foreign-language education within its broader societal context. Rapidly accelerating globalization, including the spread of communication technologies, the dominance of English as a *lingua mundi*, and the relative ease of travel, has exerted considerable influence on the field of foreign-language learning. As learners now typically enjoy ready and immediate access to foreign-language-mediated texts and interaction via the Internet, the focus of classroom pedagogy has shifted from enhancement or importation of authenticity to furthering students' skill in intercultural dialog. In order to meet this challenge, models of language competence beyond the grammatical must be taken seriously.

Meanwhile, the sociopolitical aspects of foreign language learning have come into clearer focus. Demand for English-language instruction is such that, in many countries, foreign-language learning has become synonymous with the study of English for international communication. Because English as a foreign language is plagued by questions about the cultural allegiance and colonial heritage of its speakers, the norms imposed by native-speaker models are also subject to critical scrutiny. In other words, the desired developmental end point implied within mainstream psycholinguistic approaches is no longer universally accepted. In parallel with the rise of English-language instruction, the promotion of foreign languages other than English, particularly among the educated populace of predominantly Anglophone countries, has become a matter of urgent concern. Teachers of foreign languages struggle to prove the relevance of their subject matter without allowing it to be cast in a strictly utilitarian light, and thus find support for their endeavors in broadly conceived models of language ability.

Defining Foreign-Language Ability

In the past several decades, researchers and language educators have devoted considerable attention to the definition of foreign-language ability. New perspectives from second and foreign-language acquisition research, language teaching, and corpus linguistics have challenged

prevailing views. Since the early 1980s, the dominant model of foreign-language ability has been based on the notion of communicative competence. This construct originated in a debate between the linguist Noam Chomsky and the anthropologist Dell Hymes during the mid-1960s. Chomsky defined linguistic competence in terms of an underlying, innate and universal grammar, or set of principles capable of generating the structural properties of any human language. For Hymes, this view of language was too restrictive in that it did not take into consideration knowledge of the social aspects of language use as shared within a given speech community. When such knowledge is absent, according to Hymes, utterances may be grammatically correct but socially inappropriate. In coining the term communicative competence, Hymes (1972) conveyed the notion that knowledge of language, and of variation within languages, is closely tied to the social conditions of use in particular contexts, including features of the setting, characteristics of the participants, and norms for interpretation.

In the early 1980s, under pressure to generate more realistic and practical approaches to language teaching, theorists began to elaborate models of communicative competence intended for use by the language-education profession. The first comprehensive model was presented to the profession by Canale and Swain (1980) and revised by Canale (1983). This model included four interrelated components that have remained at the core of the construct through subsequent revisions (see **Table 1**).

A similar model, under development by van Ek (1986) and the Council of Europe, added to these four components a provision for sociocultural competence, or knowledge of the social components of situations where communication is likely to occur, and for social competence, a category encompassing the learner's motivation, attitude, and stance toward second-language communications. In subsequent years, the model was revised, most notably by Celce-Murcia *et al.* (1995) to place discourse competence squarely at the core, with the other components seen as support for the practice of socially situated language use.

Models of communicative competence have traditionally attempted to portray the abilities of expert language

users within their own speech communities, that is, native speakers of their own first language. As noted above, the idealized native speaker has come under critical scrutiny due to the sociopolitical climate of foreign-language teaching in the late twentieth and early twenty-first centuries. In the 1990s, a number of foreign and SLA researchers also began to question the rational basis for proposing that first-language users should be held up as models for the emulation of second-language speakers. Such views imply a deficit model of the second-language user who can rarely achieve more than a native-like competence in circumscribed contexts. Moreover, multilinguals, it is argued, possess capabilities different from, and in some ways greater than those of monolinguals. For example, they exhibit greater metalinguistic awareness, and more divergent thinking, but may require more time than monolinguals do when performing language-related cognitive tasks. These and other observations led Cook (1999) to conclude that users of more than one language possess multicompetence wherein the development of second-language abilities interacts with first-language abilities to produce unique and complex communicative repertoires. The model for the language learner should therefore not be a native speaker of that language but instead a person possessing relevant multilingualism.

In the field of foreign-language teaching, researchers have also critiqued the notion that language learners should be interpreted as aspiring native speakers. Rather, they should be construed as developing intercultural communicative competence, or the knowledge and abilities required for communication in situations where the language in question is either the dominant medium of interaction or is the code shared among speakers of divergent primary languages. In Byram's (1989) model, intercultural communicative competence consists of: (1) a cognitive dimension, including knowledge of conventions for communicative activity in one's own and the other group; (2) an affective dimension including empathy and understanding of others' perspectives; and (3) a behavioral dimension including abilities such as the tolerance for diverse communicative styles and the ability to initiate interactions and form interpersonal relationships.

A further challenge to the idealized native-speaker model and the view of language competence as consisting of an underlying sentence-level grammar has come not from second- or foreign-language theorists but rather from the robustly empirical investigation of language in use via corpus linguistics. Based on computerized collections of spoken and written texts, this research aims to reveal regularities and patterns in the documented language use of real speakers and writers, thus demonstrating, for example, that expert users are in command not only of individual words but also of the ways in which these words conventionally form collocations in discourse. Corpus-based analysis also reveals that the grammar of spoken

Table 1 Components of communicative competence

<i>Grammatical competence</i> : knowledge of and ability to use the forms of the language (lexical items and rules of syntax, morphology, and phonology)
<i>Sociolinguistic competence</i> : ability to use the language appropriately in a variety of social settings
<i>Discourse competence</i> : ability to interpret and create spoken, written, or multimedia second-language texts that are cohesive (internally well structured) and coherent (interpretable and appropriate within their contexts)
<i>Strategic competence</i> : ability to compensate for lacunae in any of the other areas

language use should be described in terms of probabilistic statements based on empirical observation of discourse rather than as resulting from the application of deterministic rules (McCarthy, 1998).

Taken together, efforts to revise and refine the field's core definitional construct correspond to a perceived need for realistic, rationally defined, and empirically based goals for language learners to inform both research and instruction. Over time, a shift has occurred, moving the field away from a view in which the abstract, sentence-level grammatical competence of native speakers is taken as the desired developmental end point of foreign-language learning. In its place is a perspective suggesting that the field's centerstage should be occupied by the discourse of intercultural communication and the competence of multilinguals as revealed by empirical observation in the settings where foreign languages are used.

Investigating Foreign-Language Learning

In this climate of conceptual change, recent investigation of foreign-language learning has been characterized by a social turn (Block, 2003) favoring attempts to align the priorities outlined above with the foci of research. One consequence of this development is a new level of attention to aspects of communicative ability beyond the grammatical, particularly the characteristics of foreign-language pragmatics, or the study of how learners develop the ability to understand and perform action in a second language (Kasper and Rose, 2002: 5). Another effect of the social turn has been the importation of research methodologies from adjacent fields emphasizing language in use or grounded in usage-based linguistics. Increasingly, for example, in-depth description of learner-foreign-language use is informed by the rigorous empiricism of conversation analysis, and the growth of language ability is traced longitudinally through the collection and analysis of developmental-learner corpora. More broadly, the field has become increasingly responsive to the sociopolitical, cultural, and historical aspects of foreign-language learning, including the agency and identity of learners. Ethnographies and case studies attempt to illuminate relationships between language learning, the cultures of learners and host communities, personal history, and the qualities of learning environments. Inquiry based on learner and teacher narratives has pointed out the significance of sociopolitical and historical contexts in the shaping of motives and opportunities to learn.

Theories of mind or of language development incorporating a significant social dimension have begun to exert a considerable impact on the nature of foreign language research. From anthropology, researchers borrow the concept of language socialization (Ochs, 2002), focusing on the longitudinal interactive processes by

which novices acquire the competence needed for participation in the social life of a community or classroom, including routine cultural practices such as language and literacy activities, and local preferences for action, thought, and emotion. Others have emphasized the potential of Vygotskian sociocultural theory (Lantolf and Thorne, 2006), a theory of mind, in which language and other semiotic systems play a major role, whose core notion is that there exists an inherent, dynamic, and dialectical connection between human action and the cultural, historical, and institutional contexts in which it occurs. In arguing for a sociocognitive approach, Atkinson (2002: 534) adopts a radically integrative stance, suggesting that neither language acquisition nor language use – nor even cognized linguistic knowledge – can be properly understood without taking into account their fundamental integration into a socially mediated world.

For Kramsch (2002), the new relevance assigned to the social in second- and foreign-language studies suggests a need to address the relationship between approaches grounded in theories of language acquisition and those whose core metaphor represents language socialization. It has often been noted that the disciplinary roots and epistemological stance associated with these two approaches to research are seen as fundamentally incompatible, with each field conceptualizing the social context in its own way. For Kramsch, this conflict is best expressed in terms of underlying conceptual metaphors, with SLA researchers preferring the learner as computer, as noted above, and language socialization researchers, the learner as apprentice. Thus, these two fields of endeavor conceptualize the role of participation in learning in very different ways.

When researchers view the social context as an environment for acquisition of language, the relationship between participation and language learning is relatively simple: more is better. It is assumed, not unreasonably, that the more the social context exposes learners to relevant input, and the more the interactive setting highlights that input is effective in ways, the more we may expect the learner to acquire. When language acquisition does not result, or is not equally distributed among participants, it is further assumed that input processing has been hindered or blocked. Perhaps the process is slowed by factors related to the psychology of individuals who, for example, display motivational deficit, or employ a learning style inappropriate for the setting, or, alternatively, there has been insufficient provision of input via constraints on the learner's presence in communicative settings.

Approaches based on the concept of language socialization, emerging as they do from the field of linguistic anthropology, have at their core the notion that language learning is part of the much larger process of becoming a person in society. Here, the learner becomes an apprentice within communities of practice rather than an input-processor. "As novices learn more from expert members

how to use the language accurately and appropriately, they enact social relationships and other sociocultural phenomena that will make them into expert members” (Kramsch, 2002: 2). From the perspective of language socialization, language study is an opportunity for apprenticeship into and through language (Ochs, 2002: 106) within the socially organized practices of the classroom or host community. Language socialization research therefore strives to understand the dialectical relationship between individual development and broader sociocultural contexts, seeking a maximally holistic perspective while simultaneously attending to the micro-level details of language use. Normally, researchers adopt a longitudinal approach emphasizing developmental processes and choose research sites believed to be places of transformation, where change in the skills and capacities of novice participants is assumed to be observable over time. In foreign-language research, these sites are typically classrooms or other instructional settings such as language-immersion programs, but they may also include the homes, schools, and workplaces frequented by learners during their sojourns abroad.

In recent years, many researchers in this tradition have been influenced by “. . . the poststructuralist realization that learning is a non-linear, relational human activity, co-constructed between humans and their environment, contingent upon their position in space and history, and a site of struggle for the control of power and cultural memory” (Kramsch, 2002: 5). Language-socialization researchers increasingly emphasize the role of language as symbolic capital and the subtle mechanisms by which power is circulated, and relations of power reproduced, in the discourse of socialization processes. Agendas of social critique have effectively changed the view that language socialization is inevitable, desirable, and equally accessible to all. Rather, many researchers are sensitive to the dynamic and negotiated nature of this process, observing that ready access to language socialization may not be available to all comers, but may be constrained by interpretations of identity and societal norms on the part of all participants. Participants in discourse are positioned by their interlocutors in ways related to local ideologies and resources for the performance of identity, such as gender, nationality, and foreignness. Furthermore, language learners are endowed with agency, and may accept, accommodate, resist, or reject the practices they encounter. A great deal depends upon the ways in which students position themselves in interaction: how they envision their success as users of their foreign language and the extent which they can imagine their fuller participation in the communities whose languages they study in positive light.

Within the field of applied linguistics more broadly drawn, the relationship between acquisition and socialization research has been uneasy, and there has traditionally been little dialog between them. This lack of cross-disciplinary communication is due in large part to

differences in focus and in criteria for success in each case. Language acquisition researchers seek to illuminate the linguistic processes underway as individuals achieve mastery of grammatical and communicative competence. Language-socialization researchers traditionally examine the nature of sociocultural phenomena in play as individuals become assimilated members of communities. Ultimately, however, the distinction between language acquisition and language socialization is not clear cut. When the goal of SLA processes is expressed in terms of communicative competence (including pragmatic, sociolinguistic, or discourse competence), socialization clearly has a part to play. To the extent that development of expert membership requires precise linguistic expression, it is difficult to separate language acquisition from socialization.

Conclusion

Due to the recent history of debate over the province of foreign-language-acquisition research, and in equal part, due to widely perceived need for inclusive models of foreign-language ability, a key issue in foreign-language learning research has become the integration of the cognitive and the social. There is widespread appreciation of the need to overcome dichotomous views of mind and society, implied by the Cartesian heritage of Western research in the social sciences, in order to further the investigation of developing communicative competence, multicompetence, or intercultural communicative ability. Block (2003: 138), for example, suggests that the future of language acquisition research lies in approaches combining attention to acquisition and socialization as researchers “attempt to reconstruct detailed life stories of learners hand-in-hand with an interest in linguistic development over time.” Kramsch and Lam (2003: 144) argue for an ecology of language learning, focused on the meaning-making environment and attempting to capture the interconnectedness of psychological, social, and environmental processes. It is not yet clear as to how, or indeed if, this goal will be reached, or which among a broad range of candidate approaches will prevail in gathering the allegiance of researchers. In the meantime, while many continue to work within the IIO framework, examining foreign-language acquisition as a learner-internal process of cognitive restructuring based on received input, other investigators continue to explore alternative, often integrative or socially oriented approaches.

Considerable promise for the future of the field may be observed in several recent developments. First, a new level of tolerance for methodological ecumenicalism has emerged as researchers committed to language education agree to set aside debates over competing epistemologies in favor of returning to their core mission – to inform and assist the foreign-language teaching profession. The field

has thereby become more inclusive of efforts to elaborate and test the applicability of many socially grounded approaches to language learning, including language socialization, sociocultural theory, and sociocognitive theory. Along with this expansion at the level of theory has come a growth in the field's accepted repertoire of analytic techniques as well as the inclusion of new data sources both quantitative (e.g., developmental corpora) and qualitative (e.g., learner narratives and case histories). Second, there appears to be a new level of dedication to the grounding of conclusions in observation of authentic learner-language use. The empirical rigor of foreign-language learning research has been greatly enhanced by the application of usage-based approaches to linguistics and of conversation analytic methodologies, in combination with the development of new technologies such as first language and learner corpora or multimedia recording and transcription techniques. Finally, the social turn in language-acquisition research has been accompanied by an increase in the scope of inquiry as witnessed by renewed emphasis on the ethics of foreign-language teaching and learning. Attention to the sociopolitical contexts in which foreign-language learning takes place has added a dimension of social activism and critical reflection to the discourse of the field. Taken together, these observations suggest that the field of foreign-language-learning research is a dynamic enterprise entering the twenty-first century with a multifaceted and complex yet compelling agenda.

See also: An Overview of Language and Literacy in Educational Settings; Bilingualism and Learning; First Language Acquisition; Second Language Learning.

Bibliography

- Atkinson, D. (2002). Toward a sociocognitive approach to second language acquisition. *Modern Language Journal* 86(4), 525–545.
 Block, D. (2003). *The Social Turn in Second Language Acquisition*. Washington, DC: Georgetown University Press.

- Byram, M. (1989). *Cultural Studies in Foreign Language Education*. Clevedon, UK: Multilingual Matters.
 Canale, M. (1983). From communicative competence to communicative language pedagogy. In Richards, J. and Schmidt, R. W. (eds.) *Language and Communication*, pp 2–27. New York: Longman.
 Canale, M. and Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1(1), 1–47.
 Celce-Murcia, C., Dornyei, Z., and Thurell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics* 6(2), 5–35.
 Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly* 33(2), 185–209.
 Hymes, D. (1972). On communicative competence. In Gumperz, J. J. and Hymes, D. (eds.) *Directions in Sociolinguistics: The Ethnography of Communication*, pp 35–71. New York: Holt, Rinehart, and Winston.
 Kasper, G. and Rose, K. (2002). Pragmatic development in a second language. *Language Learning* 52(Supplement 1), 1–352.
 Kramsch, C. (ed.) (2002). *Language Acquisition and Language Socialization: Ecological Perspectives*. New York: Continuum.
 Kramsch, C. and Lam (2003). The ecology of an SLA community in a computer-mediated environment. In Leather, J. and van Dam, J. (eds.) *Ecology of Language Acquisition*, pp 141–158. Dordrecht: Kluwer.
 Lantolf, J. P. (1996). Second language theory building: Letting all the flowers bloom! *Language Learning* 46, 713–749.
 Lantolf, J. P. and Thorne, S. L. (2006). *Sociocultural Theory and the Genesis of Second Language Development*. London: Oxford University Press.
 McCarthy, M. J. (1998). *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
 Ochs, E. (2002). Becoming a speaker of culture. In Kramsch, C. (ed.) *Language Acquisition and Language Socialization: Ecological Perspectives*, pp 99–120. New York: Continuum.
 van Ek, J. A. (1986). *Objectives for Foreign Language Learning. Scope I*. Strasbourg: Council of Europe.

Further Reading

- Block, D. (2003). *The Social Turn in Second Language Acquisition*. Washington, DC: Georgetown University Press.
 Kramsch, C. (2002). *Language Acquisition and Language Socialization: Ecological Perspectives*. New York: Continuum.
 Lantolf, J. P. and Thorne, S. L. (2006). *Sociocultural Theory and the Genesis of Second Language Development*. London: Oxford University Press.

Learning Science

J Leach and P Scott, University of Leeds, Leeds, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Acquisition – Perspectives which portray learning as the acquisition of knowledge by individuals through various mechanisms.

Cognitive – Approaches to theorizing learning focusing on thinking by individuals.

Conceptual change – Perspectives which address how individuals' conceptual understanding changes during learning, or on descriptions of individuals' conceptual understanding before, and after, learning.

Development – Perspectives which focus on the maturation of individuals' intellectual abilities.

Participation – Perspectives which portray learning in terms of learners' growing ability to participate in various social practices.

Science concepts – The substantive ideas of science (force, evolution, etc.), as opposed to methodological and procedural ideas. For the purpose of this piece we do not differentiate these substantive ideas into facts, models, generalizations, theories, etc. Our primary focus is the physical and life sciences.

Social constructivist – Perspectives which focus upon how individuals make personal sense of meaning-making through social interactions.

Sociocultural – Perspectives which focus upon meaning-making through social interactions.

Introduction

In this article we present major perspectives that have influenced thinking on science learning. A range of perspectives on science learning can be found in the literature. We have therefore found it helpful to identify two key features which we use as organizing dimensions in developing and presenting the article. The first dimension is taken from the influential paper by Anna Sfard (1998) in which she proposed two key metaphors for learning: the acquisition metaphor and the participation metaphor.

According to Sfard (1998), human learning has been conceived of as an acquisition of something “since the dawn of civilization” and, in recent decades, “the idea of learning as gaining possession over some commodity has persisted in a wide spectrum of frameworks, from moderate

to radical constructivism and then to interactionism and sociocultural theories” (p. 6). Gaining possession implies that something is stored or held somewhere. Sfard makes clear that it is concepts which are learned and then stored in the learner's head: “Since the time of Piaget and Vygotsky, the growth of knowledge in the process of learning has been analyzed in terms of concept development. Concepts are to be understood as basic units of knowledge that can be accumulated, gradually refined, and combined to form ever richer cognitive structures” (Sfard, 1998: 5).

By way of contrast, Sfard sees the participation metaphor as offering a fundamentally different perspective on learning, in which “the learner should be viewed as a person interested in participation in certain kinds of activities rather than in accumulating private possessions” (Sfard, 1998: 6). According to this perspective, “learning a subject is now conceived of as a process of becoming a member of a certain community” (Sfard, 1998: 6).

From the outset it is important to recognize that the acquisition–participation dimension is not a continuum. The two metaphors offer fundamentally different perspectives on learning, or as Sfard (1998: 7) stated, “the acquisition/participation division is ontological in nature and draws on two radically different approaches to the fundamental question, ‘What is this thing called learning?’” Majority of approaches to conceptualizing science learning which we review here relate to the acquisition perspective.

The second dimension that we take into account is the distinction between individual and social perspectives on learning. This takes us from a starting point where the main focus is on the individual learner, moving toward approaches where increased account is taken of various social aspects of the learning process and of knowledge itself.

We begin this article with approaches to conceptualizing science concept learning which belong to the acquisition perspective and then move on to those which relate to participation.

Science Concept Learning as Acquisition: Cognitive Approaches

Following the ideas set out earlier, we first consider those approaches which see science learning as involving a process of acquisition and which focus on the individual in providing an account of that learning.

Perspectives on student concepts and conceptual learning in science have been heavily influenced by the seminal

work of the Swiss genetic epistemologist Jean Piaget. This influence was particularly dominant during the 1960s and 1970s, as can be confirmed by looking through the citations of Piaget in papers published in the main science education journals of that period. Piaget described an interactive learning process whereby an individual makes sense of the world through cognitive schemes, which are themselves modified as a result of the individual's actions on objects in the world. This model is summarized in the phrase '*L'intelligence organise le monde en s'organisant elle-même*' (intelligence organizes the world by organizing itself) (Piaget, 1937). Piaget emphasized the significance of the child's social environment on knowledge development, claiming that:

Society is the supreme unit, and the individual can only achieve his inventions and intellectual constructions insofar as he is the seat of collective interactions that are naturally dependent, in level and value, on society as a whole
(Piaget, 1971: 368).

Nonetheless, in most of Piaget's writing – and writing addressing the significance of Piagetian theorizing for science education – knowledge is portrayed as schemata in the individual's head, with little prominence being placed upon wider social aspects. The proposed mechanism for changes in intellectual organization as a result of interactions with the world (termed adaptation) involves the processes of assimilation and accommodation. Assimilation is the process by which an individual interprets particular sensory information and in so doing includes that information in their existing cognitive structure. Accommodation is the process by which cognitive structure adapts in order to make sense of specific information. Assimilation and accommodation cannot be dissociated: whenever an individual interacts with sensory information, both assimilation and accommodation take place.

Although primarily interested in development as a result of maturation rather than learning as a result of instruction, Piaget's empirical work addressed the development of children's knowledge about various aspects of the natural world including life, time, and mass, weight, and volume. Drawing upon this body of empirical work, an account of conceptual change based upon the development of content-independent logical structures was proposed (Piaget and Inhelder, 1956). Characteristic stages in the development of logical thinking were set out, based upon students' abilities to perform tasks involving skills such as conservation and seriation (the serial ordering of items). The concrete operational stage, for example, runs between approximately 2 and 12 years and is characterized by the development and coordination of conceptual schemes including conservation, classification, and seriation. Children at the concrete operational stage are not, however, capable of performing operations at a purely symbolic level; that competence is characteristic of the formal operational stage.

Piaget's work has influenced perspectives on student conceptions and conceptual learning in several ways. His account of how individuals come to know can be seen in much writing about students' conceptions, conceptual change, and personal constructivism through references to assimilation and accommodation. Piaget's methods for probing an individual's understanding, which involve an interviewer asking children questions without attempting to lead their responses, have also been drawn upon in research on students' alternative conceptions. Furthermore, Piagetian stage theory has been drawn upon to inform science curriculum design and sequencing (Science Curriculum Improvement Study, USA; Cognitive Acceleration through Science Education, UK).

Various criticisms of the use of Piagetian theory in science education have been advanced. Driver (1978), Donaldson (1978), and Carey (1985) have questioned the empirical basis on which claims for characteristic stages in logicomathematical thinking are founded. Specific criticisms include the following:

- Tasks requiring identical logicomathematical reasoning are made easier or more difficult by the degree of familiarity with the task's context (Donaldson, 1978).
- Tasks characteristic of a given stage can be performed by much younger children (Driver, 1978).
- The analysis used in Piagetian research is designed to validate existing theory rather than account for children's reasoning (Driver, 1978; Carey, 1985).

Perhaps the most significant break from the Piagetian account of conceptual learning in science can be traced back to the developmental psychology of David Ausubel (1968). Ausubel argued that the most significant influence on the learners' conceptual development is their existing conceptual knowledge in the target domain. During the early 1970s a small number of empirical studies were conducted which accounted for students' science learning in terms of domain-specific factors, rather than explaining learning in terms of global logicomathematical reasoning skills (e.g., Viennot, 1979).

An empirical research program subsequently developed (Novak, 1978), focusing upon the content of students' domain-specific reasoning (or students' alternative conceptions: Driver and Easley, 1978) about natural phenomena and involving researchers from around the world. Two particularly influential books in the development of research on students' alternative conceptions were *The Pupil as Scientist* by Rosalind Driver (1983) and *Learning in Science: The Implications of Children's Science* edited by Osborne and Freyberg (1985). The number of publications in this field of science education research moved into thousands. Helga Pfundt and Reinders Duit of the Leibniz Institute for Science Education (IPN) in Kiel, Germany, developed a comprehensive bibliography, *Student' Alternative Frameworks and Science Education* (Pfundt and Duit, 2007). All the

evidence suggests that there are strong commonalities in the alternative conceptions of students from different cultures and, furthermore, these ideas about the natural world have a profound influence on what is learnt as a result of science teaching, and some ideas are extremely resistant to change (Driver, 1989).

Learning as Conceptual Change

Recognition that prior knowledge influences learning (Ausubel, 1968), together with Piagetian ideas of accommodation and assimilation, and work from the philosophy of science all underpinned the seminal paper by Posner *et al.* (1982) on conceptual change in science learning. In their paper, the conditions needed for a major change in thinking within a scientific field (such as the shift from Earth-at-center to Sun-at-center models of the solar system) were considered to be analogous to the conditions that are needed to bring about accommodation or conceptual change in individual learners. Posner *et al.* (1982) identified four conditions which must be fulfilled before such an accommodation can occur. These conditions are that a learner must first be dissatisfied with existing ideas and then that the new ideas must be seen as intelligible, plausible, and fruitful. Empirical evidence from students' learning about the special theory of relativity was used to illustrate and exemplify this model of conceptual change learning. Though taking the view that learning is a rational activity, Posner *et al.* recognized that such accommodations might take considerable time, involving "much fumbling about, many false starts and mistakes, and frequent reversals of direction" (p. 223). The conditions of intelligibility, plausibility, and fruitfulness contribute to the status of an idea. During conceptual change the status of different ideas within a person's conceptual ecology (the range of ideas they hold) changes (Hewson and Lemberger, 2000). In addition, Scott *et al.* (1992) outlined two broad types of approach to conceptual change teaching. The first of these is based upon promoting cognitive conflict and follows from the model proposed by Posner *et al.* (1982), while in the second the learner's existing ideas are built upon and extended.

A significant point of confusion in this whole area of work concerns the different meanings which are attached to the term conceptual change. Sometimes conceptual change refers to the process of learning, while at other times to the products. Furthermore, conceptual change sometimes refers to situations where one concept (seen as a unit of knowledge) is exchanged for another; or where a concept is modified in some way, for example, by differentiation into two; sometimes where the relationship between concepts changes; or at other times where new concepts are added without loss of the original ideas. The interest in student misconceptions, or alternative

conceptions, in the 1980s led to a focus on conceptual change as revolutionary, with new ideas replacing the original ones (through a process of exchange) rather than evolutionary and gradual, with the possibility of several views existing simultaneously (through a process of addition) and used in different contexts (see, e.g., Sinatra, 2002).

What Changes During Conceptual Change?

The model of conceptual change by Posner *et al.* focused on the conditions under which radical accommodations occur. Alongside this, the focus of much work in developmental and cognitive psychology has been on what changes, exploring the performance of learners at different ages and attempting to explain this in terms of the ways in which concepts are mentally represented and related and the cognitive processes by which they are acquired and change.

One of the early proponents of domain-specific approaches, Susan Carey, proposed two forms of knowledge restructuring in learning – one similar to that demonstrated in the shift from novice to expert and one analogous to that of theory change in science. In the first, weak restructuring, the relations between concepts are changed. In the second, strong restructuring, the concepts themselves change (Carey, 1985) and this is regarded as being difficult to achieve. Considerable attention has been given to these latter situations where radical restructuring is needed, particularly in the context of learning physics concepts. The idea that learning occurs as discrete concepts are formed and then linked into more complex conceptual structures has largely given way to a view that concepts are part of larger relational structures from the start (Vosniadou, 1994). Following the seminal work of Keil (1979), ontological categorization is also seen as being of fundamental importance in learning science concepts (Chi, 1992). diSessa and Sherin (1998) pointed out some difficulties with the standard model of conceptual change. They argued that the notion of concept needs to be replaced by more carefully defined theoretical constructs within a knowledge system, which allow us to understand how that system functions. Focusing on the cognitive processes by which we gain information from the world, they proposed entities such as coordination classes and phenomenological primitives, or p-prims. Coordination classes include cognitive strategies such as selecting and integrating information and are "systematically connected ways of getting information from the world" (p. 1171).

Although Posner *et al.* (1982) noted that motivational and affective variables were not unimportant in the learning process, the model of conceptual change they

proposed was based on a view of learning as a rational activity. Pintrich *et al.* (1993) in their critique of cold conceptual change models proposed that the conditions of dissatisfaction with existing conceptions and the intelligibility, plausibility, and fruitfulness of the new, whilst necessary, are not sufficient to support conceptual change. Cognitive, motivational, and classroom contextual factors must also be taken into account as the individual student in the classroom is subject to influences from the broader social setting.

Cognitive Approaches: Summary and Implications

The following fundamental insights about science concept learning are common to the majority of cognitive perspectives:

- Individuals' beliefs about the natural world are constructed, rather than received.
- There are strong commonalities in how individuals appear to think about the natural world.
- A person's existing ideas about a given subject greatly influence the subsequent learning about that subject.

In addition, some have argued that there are more general aspects of reasoning, such as Piaget's logicomathematical reasoning skills.

Science Concept Learning as Acquisition: Sociocultural and Social Constructivist Perspectives

At this point in the article, we take a significant step in moving from approaches to characterizing science concept learning which focuses on the individual, while recognizing the influence of the social context, to those which take the social context as being an integral part of the learning process.

A fundamental theoretical reference point for sociocultural and social constructivist perspectives on learning is provided by Lev Semenovich Vygotsky (Vygotsky, 1987). Central to Vygotsky's views is the idea that learning involves a passage from social contexts to individual understanding (Vygotsky, 1978). Thus, we first meet new ideas (new to us, at least) in social situations where those ideas are rehearsed between people, drawing on a range of modes of communication, such as talk, gesture, writing, visual images, and action. Vygotsky referred to these interactions as existing on the social plane. The social plane may be constituted by a teacher working with a class of students in school; it may involve a parent explaining something to a child. As ideas are explored during the social event, each participant is able to reflect on, and

make individual sense of, what is being communicated. The words, gestures, and images used in the social exchanges provide the very tools needed for individual thinking. Thus, there is a transition from social to individual planes, whereby the social tools for communication become internalized and provide the means for individual thinking. It is no coincidence that Vygotsky's seminal book is titled *Thought and Language* (Vygotsky, 1962).

The social origins of learning are thus a fundamental and integral part of Vygotsky's account and it is the job of the teacher to make scientific knowledge available on the social plane of the classroom, supporting students as they try to make sense of it. Vygotsky brought the activities of teaching and learning together through his concept of the zone of proximal development or ZPD (Vygotsky, 1978). The ZPD provides a measure of the difference between what the student can achieve working alone and with assistance. The key point here is that the student's learning is conceived of as being directly connected to, and dependent upon, the supporting activity of the teacher on the social plane.

In addition to drawing attention to the social origins of learning, Vygotsky also emphasized the role of the individual in the learning process. The process of internalization, as envisaged by Vygotsky, does not involve the simple transfer of ways of talking and thinking from social to personal planes. There must always be a step of personal sense-making. Leontiev, one of Vygotsky's contemporaries, made the point in stating that, "the process of internalisation is not the transferral of an external activity to a pre-existing 'internal plane of consciousness': it is the process in which this plane is formed" (Leontiev, 1981: 57). That is, individual learners must make sense of the talk which surrounds them on the social plane, relating that talk in a dialogic way to their existing ideas and ways of thinking.

In this respect Vygotskian theory shares common ground with the constructivist perspectives outlined earlier, which emphasize that learners cannot be passive recipients of knowledge. It is perhaps with this point in mind that those contemporary approaches to conceptualizing science learning, which draw on Vygotskian sociocultural theory, are often referred to as social constructivist perspectives.

Social Constructivist Views of Learning Science

Vygotskian theory has been directly drawn upon by a number of researchers in developing an account of science learning (see, e.g., Driver *et al.*, 1994; Hodson and Hodson, 1998; Scott, 1998; Leach and Scott, 2002; Leach and Scott, 2003; Mortimer and Scott, 2003; Wells, 1999).

Hodson and Hodson, for example, outlined a social constructivist perspective on teaching and learning science, which was "based on the Vygotskian notion of enculturation" (1998: 33). They argued that this perspective

provides an alternative to personal constructivist accounts of learning, which they claimed often imply “that students who construct their own understanding of the world are building *scientific* understanding” (1998: 34, emphasis as in original).

Central to social constructivist views is the fundamental epistemological tenet that areas of knowledge such as science are developed within specific social communities. Thus, Driver *et al.* (1994) stated that, “. . . if knowledge construction is seen solely as an individual process, then this is similar to what has traditionally been identified as discovery learning. If, however, learners are to be given access to the knowledge systems of science, the process of knowledge construction must go beyond personal empirical enquiry. Learners need to be given access not only to physical experiences but also to the concepts and models of conventional science” (1994: 7). The implications of this point are fundamental. The understanding of an individual, acquired, on the one hand, through the individual’s interactions with the material world and, on the other, through introduction to the concepts and models of conventional science, are ontologically different. The concepts and models of conventional science embody practices, conventions, and modes of expression that are socially and institutionally agreed. Because scientific knowledge is the product of the scientific community, it cannot be learnt through interactions with the material world alone. Such differences between empiricist interpretations of personal constructivism and social constructivist accounts of learning are discussed by Leach and Scott (2003).

Following the ideas set out in the preceding sections, social constructivist accounts of learning can be deemed to be social in nature on two counts. First, in the sense of specifying the social origins of learning through the interactions of the social plane and, secondly, in recognizing the social context of the scientific community for the development of scientific knowledge.

The sociocultural view of learning offers an interesting perspective on the origins and status of alternative conceptions or misconceptions. From the sociocultural point of view an alternative conception such as the idea of a plant drawing its food from the soil, is representative of an everyday way of talking and thinking about plants. This is the way in which ordinary folks talk about such things and in this respect there is a very real sense in which the scientific point of view (based on the concept of photosynthesis) offers the alternative perspective. Viewed in this way, it is hardly surprising that the alternative conceptions or misconceptions identified by the science education community are robust and difficult to change. These are not the ephemeral outcomes of the solitary musings of a child trying to make sense of the natural world around them, but the tools of an everyday language which continuously acts to define, and reinforce, our ways of talking and thinking.

Social Constructivist Perspectives: Summary and Implications

The following insights about science concept learning are common to social constructivist perspectives:

- Learning scientific knowledge involves a passage from social to personal planes.
- The process of learning is consequent upon individual sense-making by the learner.
- Learning is mediated by various semiotic resources, with language being the most important.
- Learning science involves learning the social language of the scientific community which must be introduced to the learner by a teacher or some other knowledgeable figure.

Where do these distinctive aspects of the social constructivist perspective take us which are different from the interests and outcomes of the cognitive viewpoint? The most obvious development has been the increased attention, during the late 1980s and 1990s, to the role of the teacher and the ways in which teachers guide the discourse of the classroom to support the introduction of scientific knowledge and scientific ways of explaining (Edwards and Mercer, 1987; Ogborn *et al.*, 1996; Van Zee and Minstrell, 1997; Mortimer and Scott, 2003). Through this kind of work, we have a much better grasp of the ways in which teachers make scientific knowledge available on the social plane of the classroom.

Science Concept Learning as Participation

In this final section we take the step from approaches to conceptualizing science learning which are based on acquisition to those which entail some form of participation.

Situated Cognition

The metaphor of learning as participation has largely arisen through a perspective on learning known as situated cognition (see, e.g., Brown *et al.*, 1989; Lave and Wenger, 1991).

The pioneering work in this field focused on the use of mathematics in the workplace and in day-to-day life. For example, Scribner (1984) analyzed the arithmetical practices of people as they worked in a dairy factory, while Lave (1988) focused on the use of arithmetic in everyday shopping. These studies, and others, have identified forms of mathematics which are radically different from those taught in school. The skilled users of these everyday forms of arithmetic vary their problem-solving approaches depending on the specific situation and problems, which

appear to be structurally identical but are solved using different strategies. In this sense, the strategies are seen to be directly linked to context and thereby situated in nature.

According to the situated cognition perspective, learning is seen as a process of enculturation, or participation in socially organized practices, through which specialized skills are developed by the learner as they engage in an apprenticeship in thinking (Rogoff, 1990), or in legitimate peripheral participation (Lave and Wenger, 1991). According to Collins *et al.* (1989), the key components of the apprenticeship process include modeling, coaching, scaffolding, fading, and encouraging learners to reflect on their own problem-solving strategies. This apprenticeship leads to the learner becoming involved in the authentic practices of a 'community of practice' (Lave and Wenger, 1991). Seely-Brown *et al.* (1989) argued that, "Unfortunately, students are too often asked to use the tools of a discipline without being able to adopt its culture. To learn to use tools as practitioners use them, a student, like an apprentice, must enter that community and its culture." Roth (1995a: 29) suggested that authentic practices involve activities, "which have a large degree of resemblance with the activities in which core members of a community actually engage."

In the context of education, situated cognition perspectives have received a lot of attention, particularly in North America and particularly in relation to mathematics education. According to Cobb and Bowers (1999), "A situated perspective on the mathematics classroom sees individual students as participating in and contributing to the development of the mathematical practices established by the classroom community."

Situated perspectives on learning have also been drawn upon as part of a theoretical justification for inquiry-based approaches to science teaching and learning (see, e.g., Roth, 1995b). Roth (1995a: 29) suggested that "situated learning emphasizes learning through the engagement in authentic activities." He explained his use of the term authentic by suggesting that in classrooms focusing on scientific activities, the students would "(1) learn in contexts constituted in part by ill-defined problems; (2) experience uncertainties, ambiguities, and the social nature of scientific work and knowledge; (3) engage in learning (curriculum) which is predicated on, and driven by, their current knowledge state; (4) experience themselves as part of communities of inquiry in which knowledge, practices, resources, and discourse are shared; and (5) participate in classroom communities, in which they can draw on the expertise of more knowledgeable others" (Roth, 1995a: 29; see also: Wells, 1999).

Drawing explicitly upon these ideas, science instruction has been planned and implemented as the enculturation of students into practices such as field ecology (e.g., Roth and Bowen, 1995), environmental activism

(e.g., Roth and Désautels, 2002), and basic scientific research (e.g., Ryder *et al.*, 1999). Although the practices described in these studies can be argued to be authentic in the sense that they refer to situations in which science is actually used, it is more difficult to argue that they are closely related to the everyday experience of most science learners. Furthermore, the authors' analyses of teaching focus more upon students' learning about various practices that involve science (the use of instrumentation and specific technical procedures, the construction of arguments, the social relationships of various communities) than upon the development of conceptual understanding by students.

Participative Approaches: Summary and Implications

The following insights about learning are common to the participative approaches outlined earlier:

- Learning is seen as a process of developing participation in the practices of a particular community.
- The learner takes on the role of apprentice, while the teacher is seen as an expert participant.
- That which is to be learned involves some aspect of practice or discourse.

Perhaps the biggest question to be raised in relation to the participative approaches concerns the issue of subject matter and the very aims of science education. For example, what does it mean to suggest that learning science should involve participation in the practices of a scientific community? What does it mean to suggest that students should engage in the authentic practices of science? To what extent is it possible to reconfigure the science classroom as a seat of authentic scientific practices? Is it reasonable to expect that the teacher can act as an expert practitioner within this scientific community of the classroom? What would be the aims of such an approach to science education? What would be learned?

See also: Cognition: Overview and Recent Trends; Concept Learning; Constructivism and Learning; Learning in a Sociocultural Perspective; Piaget: Recent Work; Vygotsky and Recent Developments.

Bibliography

- Ausubel, D. P. (1968). *Educational Psychology: A Cognitive View*. New York: Holt, Rinehart and Winston.
- Brown, J. S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* 18(1), 32–42.
- Carey, S. (1985). *Conceptual Change in Childhood*. Cambridge, MA: MIT Press.
- Chi, M. T. H. (1992). Conceptual change within and across ontological categories: Examples from learning and discovery in science. In

- Gieryn, R. (ed.) *Cognitive Models of Science: Minnesota Studies in the Philosophy of Science*, pp 129–186. Minneapolis, MN: University of Minnesota Press.
- Cobb, P. and Bowers, J. (1999). Cognitive and situated learning perspectives in theory and practice. *Educational Researcher* **28**(2), 4–15.
- Collins, A., Brown, J. S., and Newman, S. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In Resnick, L. (ed.) *Cognition and Instruction: Issues and Agendas*, pp 453–494. Hillsdale, NJ: Erlbaum.
- diSessa, A. and Sherin, B. (1998). What changes in conceptual change? *International Journal of Science Education* **20**(10), 1155–1191.
- Donaldson, M. (1978). *Children's Minds*. London: Helm.
- Driver, R. (1978). When is a stage not a stage? A critique of Piaget's theory of cognitive development and its application to science education. *Educational Research* **21**(1), 54–61.
- Driver, R. (1983). *The Pupil as Scientist?* Milton Keynes: Open University Press.
- Driver, R. (1989). Students' conceptions and the learning of science. *International Journal of Science Education* **11**(5), 481–490.
- Driver, R., Asoko, H., Leach, J., Mortimer, E., and Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher* **23**(7), 5–12.
- Driver, R. and Easley, J. (1978). Pupils and paradigms: A review of literature related to concept development in adolescent science students. *Studies in Science Education* **5**, 3–12.
- Edwards, D. and Mercer, N. (1987). *Common Knowledge; the Development of Understanding in the Classroom*. London: Methuen.
- Hewson, P. and Lemberger, J. (2000). Status as the hallmark of conceptual learning. In Millar, R., Leach, J., and Osborne, J. (eds.) *Improving Science Education: The Contribution of Research*, pp 110–125. Buckingham: Open University Press.
- Hodson, D. and Hodson, J. (1998). From constructivism to social constructivism: A Vygotskian perspective on teaching and learning science. *School Science Review* **79**, 33–41.
- Keil, F. (1979). *Semantic and Conceptual Development: An Ontological Perspective*. Cambridge, MA: Harvard University Press.
- Lave, J. (1988). *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life*. Cambridge: Cambridge University Press.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. New York: Cambridge University Press.
- Leach, J. and Scott, P. (2002). Designing and evaluating science teaching sequences: An approach drawing upon the concept of learning demand and a social constructivist perspective on learning. *Studies in Science Education* **38**, 115–142.
- Leach, J. and Scott, P. (2003). Learning science in the classroom: Drawing on individual and social perspectives. *Science and Education* **12**(1), 91–113.
- Leontiev, A. N. (1981). The problem of activity in psychology. In Wertsch, J. V. (ed.) *The Concept of Activity in Soviet Psychology*, pp 37–71. Armonk, NY: Sharpe.
- Mortimer, E. and Scott, P. (2003). *Meaning Making in Secondary Science Classrooms*. Milton Keynes: Open University Press.
- Novak, J. (1978). An alternative to Piagetian psychology for science and mathematics education. *Studies in Science Education* **5**, 1–30.
- Ogborn, J., Kress, G., Martins, I., and McGillicuddy, K. (1996). *Explaining Science in the Classroom*. Buckingham: Open University Press.
- Osborne, R. and Freyberg, P. (eds.) (1985). *Learning in Science: The Implications of Children's Science*. Auckland: Heinemann.
- Pfundt, H. and Duit, R. (2007). Bibliography: Students' Alternative Frameworks and Science Education. <http://www.ipn.uni-kiel.de/aktuell/stcse/stcse.html> (accessed August 2009).
- Piaget, J. (1937). *La construction du réel chez l'enfant*. Neuchâtel: Felachaux et Niestlé.
- Piaget, J. (1971). *Biology and Knowledge*. Edinburgh: Edinburgh University Press.
- Piaget, J. and Inhelder, B. (1956). *The Child's Conception of Space*. London: Routledge/Kegan Paul.
- Pintrich, P. R., Marx, R. W., and Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research* **63**, 167–199.
- Posner, G. J., Strike, K. A., Hewson, P. W., and Gerzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education* **66**(2), 211–227.
- Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. Oxford, England: Oxford University Press.
- Roth, W. -M. (1995a). *Authentic School Science. Knowing and Learning in Open-Inquiry Science Laboratories*. Dordrecht: Kluwer Academic Publishers.
- Roth, W. -M. (1995b). Teacher questioning in an open-inquiry learning environment: Interactions of context, content and student responses. *Journal of Research in Science Teaching* **33**(7), 709–736.
- Roth, W. -M. and Bowen, G. M. (1995). Knowing and interacting: A study of culture, practices and resources in a grade 8 open-inquiry science classroom guided by a cognitive apprenticeship metaphor. *Cognition and Instruction* **13**, 73–128.
- Roth, W. -M. and Désautels, J. (2002). *Science Education as/for Sociopolitical Action*. New York: Counterpoints.
- Ryder, J., Leach, J., and Driver, R. (1999). Undergraduate science students' images of the nature of science. *Journal of Research in Science Teaching* **36**(2), 201–220.
- Scott, P. (1998). Teacher talk and meaning making in science classrooms: A Vygotskian analysis and review. *Studies in Science Education* **32**, 45–80.
- Scott, P. H., Asoko, H. M., and Driver, R. H. (1992). Teaching for conceptual change: A review of strategies. In Duit, R., Goldberg, F., and Niedderer, H. (eds.) *Research in Physics Learning: Theoretical Issues and Empirical Studies*, pp 310–329. Kiel: IPN, University of Kiel.
- Scribner, S. (1984). Studying working intelligence. In Rogoff, B. and Lave, J. (eds.) *Everyday Cognition: Its Development in Social Context*, pp 9–40. Cambridge, MA: Harvard University Press.
- Seely-Brown, J., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* **18**(1), 32–42.
- Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher* **27**(2), 4–13.
- Sinatra, G. M. (2002). Motivational, social and contextual aspects of conceptual change: A commentary. In Limon, M. and Mason, L. (eds.) *Reconsidering Conceptual Change: Issues in Theory and Practice*, pp 187–197. Dordrecht: Kluwer Academic Publishers.
- Van Zee, E. H. and Minstrell, J. (1997). Reflective discourse: Developing shared understandings in a physics classroom. *International Journal of Science Education* **19**(2), 209–228.
- Viennot, L. (1979). Spontaneous reasoning in elementary dynamics. *European Journal of Science Education* **1**(2), 205–221.
- Vosniadou, S. (1994). Capturing and modelling the process of conceptual change. *Learning and Instruction* **4**, 45–69.
- Vygotsky, L. S. (1962). *Thought and Language*. Cambridge, MA: MIT Press.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. S. (1987). *The Collected Works of L.S. Vygotsky*, Vol. 1, Rieber, R. W. and Carton, A. S. (eds.) and Minick, N. (trans.). New York: Plenum.
- Wells, G. (1999). *Dialogic Inquiry. Toward a Sociocultural Practice and Theory of Education*. Cambridge: Cambridge University Press.

Learning to Read

M S Burns and J K Kidd, George Mason University, Fairfax, VA, USA

© 2010 Elsevier Ltd. All rights reserved.

Essential Underpinnings of Reading

All beginning readers depend on language underpinnings that begin developing in the preschool years. During formal reading instruction, children continue to develop oral-language capabilities needed to approach and understand written language. As preschool children's oral language develops, their expressive oral language increases and their listening comprehension becomes more sophisticated. As this occurs, their ability to remember and use information increases and they learn how to listen differently when hearing expository text, storybooks, poems, and nonsense rhymes. They become aware of the sounds of the language and the phonemes and morphemes that comprise words. They learn the purpose of written language and the forms it takes. At a basic level, they learn how print works (concepts of print). A final but major part of an early literacy foundation is children's developing motivation to read. These essential skills and knowledge continue to develop during instruction in learning to read and as students continue to build reading competence (see **Figure 1**). Below, we elaborate on these underlying skills and knowledge.

Language and Listening Comprehension

Oral language begins in infancy and continues throughout life. It is the largest developmental domain relevant to reading. Oral-language skills form the foundation for the transition from understandings of spoken language to written language. During the preschool years, children develop their receptive language, which enables them to understand, remember, and use what they hear, as well as their expressive language, which gives them the ability to communicate their own needs and thoughts.

At the word level, young children develop speech discrimination in the languages they hear on a consistent basis. They hear words and also the separation between words, developing a sense of what a word is. They compare and contrast words, understanding that some begin with the same sounds and some end with the same sounds, developing phonological awareness. Through manipulation of this word-sound system, they begin to understand that words (speech) are made up of a sequence of sounds (phonemes) and are combined in different ways for different words. Given this opportunity and familiarity with many words, children eventually develop a mental model that enables them to break the code (i.e., understand

sound-letter correspondences). Children's model for learning and understanding new words is also rooted in morphological development. During the preschool years, children learn many aspects of morphology, for example, how to form past tense and possessives. A morpheme is the smallest unit of language that carries meaning, for example, the word play has one morpheme, that is, play, and the past tense of play, played, has two morphemes play and ed. Children develop phonological awareness and later morphological awareness, metalinguistic understandings. Words are connected and syntax developed.

Central to language development is semantic development and vocabulary. Young children learn sentential semantics, how phrases and sentences are ordered to obtain meaning. Lexical semantics and vocabulary flourish around 2–3 years of age when children acquire the naming insight, realizing that words are names for things. Developing vocabulary includes not only learning new words but the interrelationships between words. Young children play with word meaning as exemplified in the childhood joke, "Why did the girl throw the butter out the window? To see a butterfly." As children build upon these skills and gain experience with storytelling, they develop a sense of narrative and hone their listening-comprehensions skills.

Forms and Uses of Written Language

During the preschool period, children begin to learn that written language comes in different forms and is used for different purposes. They learn that words are arranged on paper in different ways (e.g., the food words on a menu vs. those on a grocery list vs. those in a cookbook). They simultaneously learn that the food words in these three different forms have different functions. They learn that a storybook (i.e., fiction) is different from a book in which they are learning new information (i.e., expository text). They learn the features of a narrative, how in a narrative a sequence of events relates to the central theme of a story. They learn that a written story has specific features such as quotes in parentheses along with an indication of who is saying the information in quotes (e.g., "Let's take the bicycles," said Mary).

Knowledge of the Written Symbol System and of Print Concepts

Before they can read, children as young as 3 years know we can read certain letter strings, like BOOK, but not

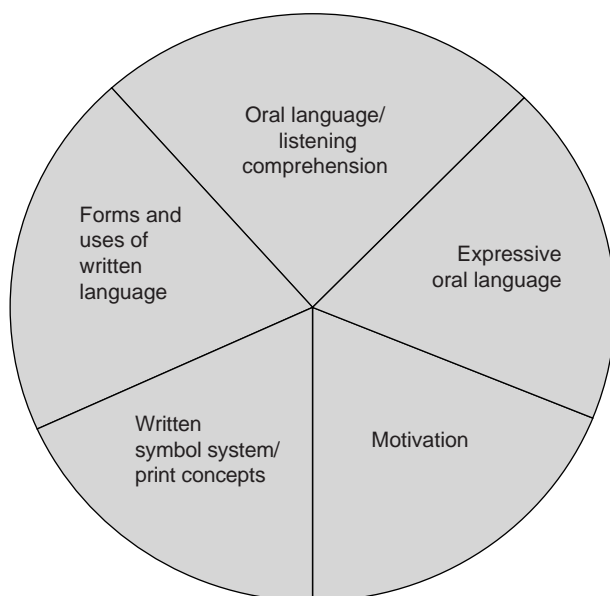


Figure 1 Essential underpinnings of reading.

TTTT. Young children also reliably classify BOOK as a word and 8965 as a number. They learn that it is the print in books that is read, and that it works in a certain way (e.g., in English it goes from top to bottom, left to right on the page). They learn knowledge of punctuation and letter knowledge. They develop the understanding that print can have meaning independent of immediate context.

Motivation: Becoming Enthusiastic About Reading and Writing

Literacy as a source of enjoyment takes place in the preschool years as children are read to in interactions that are positive and warm. These experiences are enhanced when the interactions address children's interests, take into account their prior knowledge, honor cultural continuity, and provide multilingual support.

Instruction That Develops Essential Underpinnings of Reading

Oral language, forms and uses of written language, knowledge of the written symbol system, concepts of print, and becoming enthusiastic about reading and writing are acquired before formal instruction in learning to read. These essential skills and knowledge continue to develop during instruction in learning to read and as students continue to build reading competence.

How do children learn the knowledge and skills during the preschool period? The social context is particularly important as young children communicate and participate in social and cultural activities, including experiences such

as storytelling, singing, and reading story books. Young children explore both the reading and writing systems when they, for example, pretend to read and write or act out stories that have been read to them. Peer play is essential. Sociodramatic play is especially important for supporting children's language. It is a time in which children use symbols for objects and participate in shared schemes with other children as they develop their play narratives. Adults enter interactions with children and teach and nudge by sharing their language, knowledge, and skills. The variety and richness of adult words (including rare words) used in those conversations influences children's use of vocabulary. Positive, warm exchanges have the most positive outcomes on children's language and literacy. Children develop their knowledge of the world and urge adults to provide new learning. These shared interactions between adults and children are sustained over time, developing depth of understanding of language and literacy.

Beginning of Formal Reading Instruction

Formal reading instruction builds upon these understandings of language and literacy that emerge as children interact with others and encounter language and literacy models in their world. When learning to read, children build upon what they know about language and text to simultaneously identify printed words and construct meaning. They draw upon their prior experiences, cultural knowledge, and developing vocabularies to make sense of the text. As they develop their abilities to decode and comprehend text, they become fluent readers who read with accuracy, prosody, and speed.

Identifying Printed Words

Identifying printed words requires knowledge of the relationship between letters and sounds and a developing repertoire of words that are recognized at sight. The development of letter-sound correspondences builds upon children's phonological sensitivity and their increasing awareness of phonemes. According to Ehri, children progress through five phases of alphabetic development. In the prealphabetic phase, children have no awareness that there is a relationship between letters and sounds. As children move to the partial-alphabetic phase, they gain an awareness of the relationship between letters and sounds, but focus their attention on the most prominent parts of words (e.g., initial letters). In the fully alphabetic phase, children know the sounds associated with the letters and blend the sounds together to pronounce the word. By the consolidated alphabetic phase, children recognize the whole word instantly and may use multiletter chunks to arrive at

the word. When reaching the automatic alphabetic phase, children recognize most words by sight and are able to apply a variety of strategies to decode unfamiliar words.

Movement through these phases requires increasing knowledge about letters and their associated sounds. Children gain this knowledge in a variety of ways. A focus on phonemic awareness combined with the development of letter–sound correspondences has been found to be beneficial in initial instruction. In this approach, children are taught to segment, blend, delete, and substitute phonemes as well as match and isolate sounds and match words. As phonemic awareness is developed, the sounds associated with particular letters are introduced, reinforced, and applied. Application of children’s knowledge of letter–sound correspondences can be seen in their written as well as oral representations of letter sounds. As children write to communicate their stories and ideas, writing the letters of the sounds heard reinforces the relationship between sounds and letters and provides insight into the development of letter–sound correspondences.

As children develop an awareness of letter–sound correspondences, a synthetic phonics approach that systematically teaches letters and their corresponding sounds can be helpful for some. This approach focuses on teaching the sounds associated with letters and then teaching children how to blend the sounds together to decode the word. However, many children use sequential decoding rather than synthetic phonics as they figure out unfamiliar words. Sequential decoding involves looking at all letters in an unknown word and then decoding the word by associating sounds with some but not all of the letters of the word. Children who decode in this way use what they know about letters and sounds as well as picture and context cues to predict the word. For example, children reading the sentence, “Stan pet the dog,” might use their knowledge of the sounds associated with the consonants in pet and dog along with the picture of a boy petting a dog and the context of the story to help decode both unfamiliar words. Caution has to be taken to ensure that students do not develop an overreliance on pictures for decoding as the overriding goal in decoding instruction is learning the letter–sound correspondences.

For many children who have some understanding of letter–sound correspondences, an analogy approach to learning to decode words is effective. This method of instruction focuses children’s attention on using patterns children know to decode unfamiliar words. This approach requires that children delete the initial consonants of a syllable, the onset, to segment the rime, the vowel and consonants that follow the vowel. For example in ‘dog,’ the ‘d’ is the onset and ‘og’ is the rime, and in ‘Stan,’ ‘St’ is the onset and ‘an’ is the rime. Once children recognize a familiar rime, they can decode unfamiliar words by substituting the onset in the known word with the onset in the familiar word. Therefore, children who can read dog, can

use the familiar pattern ‘og’ to read an unknown word fog by substituting the ‘d’ for an ‘f.’ This use of analogy is often more efficient and effective than sounding out each letter in a word and blending them together because the vowel sounds in these patterns are fairly constant. For example, the ‘a’ in ‘an’ remains stable whether used in Stan, can, or answer. Whereas, learning sounds for ‘a’ is variable depending on the letters that follow it (e.g., cat, cay, car, and care).

The reliance on patterns to decode words continues as children encounter multisyllabic words. However, the use of onset and rime is not nearly as prevalent as the use of morphemes in decoding longer words. Recognizing morphemes helps children decode words using familiar patterns, such as ‘jump’ and ‘ing’ to arrive at ‘jumping.’ As morphemes are the smallest unit of meaning, the use of morphemic patterns is important because they not only help children decode unknown words, but these patterns also help them construct meaning. For example, children who can decode bed, know what a bed is, and understand that adding an ‘s’ to form beds makes the word plural recognize that two or more beds are being referenced. Likewise, children who are familiar with bed and room comprehend what a bedroom is through their ability to decode and understand both morphemes.

As children’s decoding skills develop, so does their ability to recognize words automatically. Developing a sight-word vocabulary involves recognizing familiar words by linking printed words with words stored in memory to remember their meanings and how to pronounce them. As children are exposed to words, connections are made between the graphemes in the spelling of the words and their associated phonemes. The word spellings are a visual representation of the words stored in memory and serve to activate the pronunciation and meaning of these words as children encounter them in print. By developing a repertoire of sight words, children are able to expend less energy on the decoding of text and can focus on constructing meaning as they read.

The goal of decoding instruction is to foster automatic word recognition and enhance reading comprehension to produce fluent readers. Effective instruction builds upon the essential underpinnings developed before formal reading instruction, recognizes the influence of motivation on children’s willingness to interact with printed text, and develops decoding and comprehension skills and strategies simultaneously (see **Figure 2**). Children’s knowledge of the written symbol system and print concepts supports their ability to navigate the text (e.g., directionality; meaning of punctuation; knowledge of letters, words, and sentences) and helps them understand that identifying printed words is a process of not only being able to pronounce the written word, but also of being able to associate meaning. This requires children to use what they know about oral language, including their knowledge of phonemes,

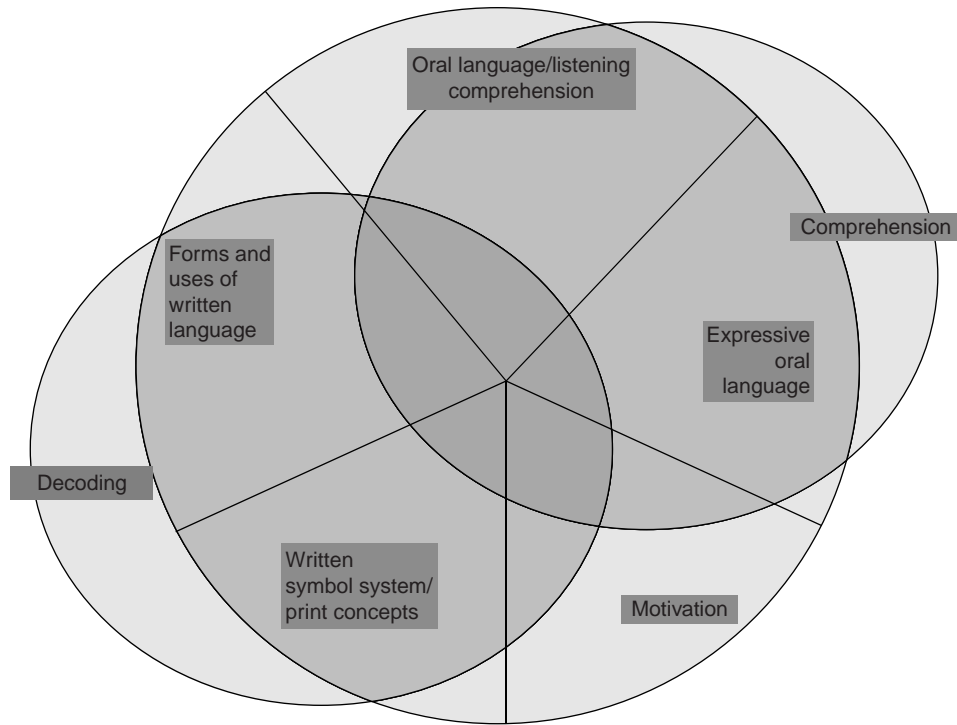


Figure 2 Formal reading instruction.

morphemes, syntax, and semantics, to read the word. For example, when encountering ‘read’ in print, children use sound–symbol correspondence, grammar, and the context to know how to pronounce and interpret the word. ‘Read’ used, pronounced, and understood in the past tense (e.g., she read the book yesterday) differs from ‘read’ used, pronounced, and understood in the future tense (e.g., she will read the book tomorrow). In essence, children use their knowledge of oral and written language to identify printed words to comprehend text while at the same time they rely on their comprehension of text to assist in their identification of the printed words.

Reading Comprehension

Reading comprehension occurs when readers construct meaning as they interact with the written word in an exchange of ideas between themselves and the message in the text. To comprehend text, readers rely upon their ability to decode text fluently, draw upon their extensive vocabularies, and employ comprehension skills and strategies to assist their understanding of the printed materials. This requires them to draw upon their prior knowledge and experiences to make connections between their existing knowledge and the information presented in the text. This interactive process between readers and text enables readers to understand, remember, and use information read and

is influenced by readers’ purpose for reading, motivation, and social context.

When learning to read, developing fluent word recognition is important for enhancing comprehension of text. Both decoding and comprehension demand memory which has a limited capacity. Effort devoted to decoding words detracts from the ability to expend resources on understanding text. Therefore, developing automatized decoding is vital to increasing comprehension. However, a focus on word recognition should not be at the expense of an emphasis on developing vocabulary and comprehension skills and strategies. All areas need to be developed simultaneously as children learn to read.

Vocabulary and concept development is also essential to developing comprehension in children learning to read. Knowledge of words and their meanings contributes to clearer understandings of the text. When vocabulary is unknown, children rely on clues from the text, which may result in an incorrect understanding of what was read. To build a rich vocabulary that enhances comprehension, children learning to read need to be immersed in language-rich experiences, both in natural as well as instructional contexts. Exposure to a wide range of vocabulary is needed, including contextualized oral language and the decontextualized language in oral storytelling and written text. These encounters with vocabulary that begin before formal instruction (see **Figure 2**) and continue over time enable children to develop deeper understandings of words and the meanings they convey.

The ability to identify printed words and attach meaning to the words enables children to construct literal interpretations of the text. To comprehend at a deeper level, children learn to activate their prior knowledge or schema to make inferences about what they read. Children reading about a trip to the park will use information based on their own experiences to help them understand. They will rely on what they know about parks to infer information that is not explicitly stated. When the connections they make are relevant and accurate, comprehension is fostered. However, if the connections they make are not relevant or are inaccurate or past experiences with and prior knowledge about the topic are minimal or nonexistent, drawing upon existing schema can limit comprehension. Therefore, children also learn to rely on the connections they make within the text to make inferences. This is especially helpful when they have limited knowledge about or experience with the topic. For example, children reading about an activity that is not familiar might infer that the activity is fun if the text refers to the characters' laughter as they engage in the activity.

As children learn to read, they develop a repertoire of strategies that promote active engagement before, during, and after reading. Prior to reading a text, children learn to establish a goal for reading and activate their prior knowledge. Children use their prior knowledge and the existing text to make predictions as they read. Active reading involves making predictions, confirming or modifying the predictions, and making new predictions. As children make predictions and confirm or modify those predictions, they are engaged in constructing meaning. However, when children hold on to incorrect predictions, comprehension is hindered. For example, children reading a book about the South Pole might originally make the prediction that the South Pole has a warm climate because they associate south with warm. As they read about the ice and the cold temperature, children who have learned to use the information from the text to modify their predictions will realize that the South Pole is cold and will use this information while continuing to read. Whereas, children who maintain their original prediction will not understand information that is reliant on knowing that the South Pole has a cold climate.

While reading, children also learn to monitor their reading. They become aware of the text structure, the relevance of what they are reading, and their own understanding or lack of understanding of the text. They learn to moderate their pace, reread when necessary, and use different reading strategies to enhance their comprehension. Their increasing repertoire of strategies may include generating questions, visualizing the text, and paraphrasing what they read. They learn to identify important information and organize details around the main ideas. The development of their metacognitive awareness enables children to regulate their reading and become more effective readers.

Reading Fluency

As with the previous processes addressed in this piece on learning to read, fluency is dependent upon and coordinated with all the previous mentioned aspects of reading as well as decoding and comprehension (see **Figure 3**) and is emphasized in instruction at all reading levels using various levels of text difficulty. Prosody, using emphasis and variations in intonation, pausing, etc., is a feature of reading that is observed in fluent readers. To achieve prosody, readers employ their knowledge of syntactic information and processing as they use meaningful phrasing and punctuation in the text being read. As mentioned in the section on identifying printed words, morphology and phonology play important roles in achieving decoding automaticity. Prior knowledge, vocabulary, and strategic processing are central in comprehension, as depicted in **Figure 2**. These processes overlap with each other and with the essential underpinnings of language and literacy. Listening comprehension and motivation explain variance in fluency.

Specific instruction to achieve fluency includes a number of forms of repeated reading, both assisted by teachers and peers as well as technology, to be effective. Silent reading with no accounting for comprehension is an additional practice that has been used to increase the amount of absolute reading but results are mixed for use of this method. It seems that the bottom line for fluency is that absolute amount of reading is the most important factor. Caution must be taken though to make sure that absolute amount of reading is defined as reading to get meaning from the text. Word calling without understanding, even if done with accuracy and expression, does not count. Fluency instruction focusing on decoding speech and accuracy and prosody can have such an impact, that is, word calling. The reader might not be reading accurately and comprehending during this time.

Levels of Reading

When children are learning to read, the children's levels of reading and the difficulty of the text must be taken into account. Children's decoding and comprehension of text and their ability to read fluently are affected not only by their reading development, but also by the text they are reading. The difficulty or ease in reading a text is influenced by textual features, such as the size and placement of the text; the presence or lack of pictures; difficulty of the words; the complexity of the sentence structure; and the children's familiarity with the content of the text. Recognizing that texts will pose different levels of challenge for children as they learn to read, it is important to match texts with children in ways that foster reading development. To facilitate this process, texts are generally

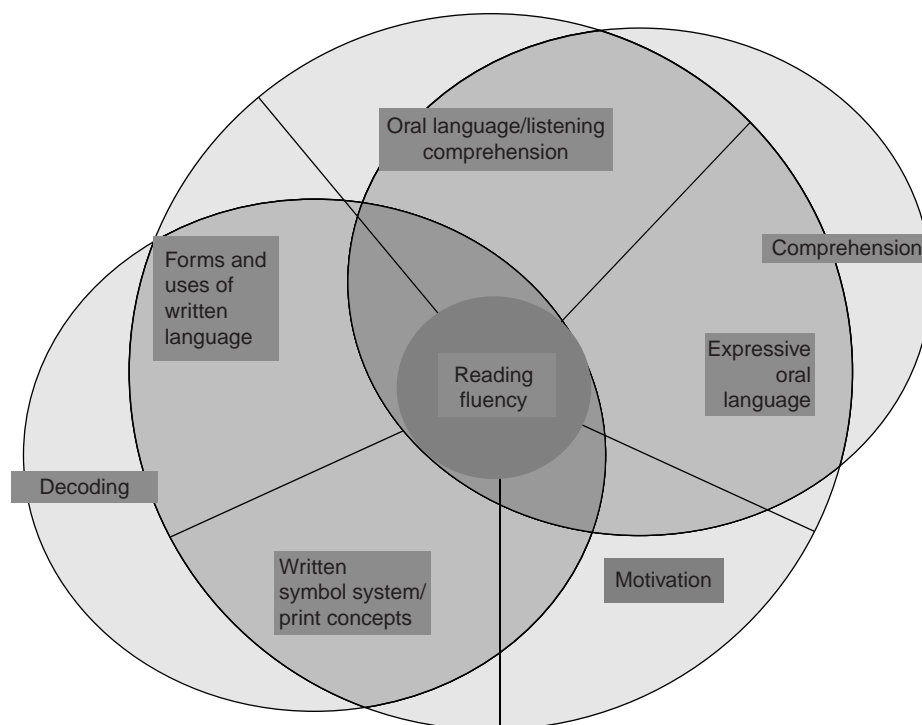


Figure 3 The fluent reader.

classified as being at children's independent, instructional, borderline instructional, or frustration level (see **Table 1**).

Texts at children's independent level are ones children can read fluently without assistance. Reading texts on their independent level provides children with an opportunity to practice and apply the reading strategies and skills they already possess and improves their decoding, comprehension, and fluency. Children's reading of instructional-level texts promotes reading development when support or scaffolding is provided by a more adept reader. The use of instructional-level texts during reading instruction enables teachers to instruct within the children's zone of proximal development and optimizes the potential effects of the instruction provided. This is also true when reading materials at the borderline instructional level are read during instruction. However, texts at this level require a high level of teacher support. This is often the level used when children are working with a reading specialist or tutor who can provide more intensive instruction and scaffolding. Texts at children's frustration level require such a high level of support that children typically do not benefit from instruction. With these levels in mind, teachers can enhance reading development by providing children with opportunities to read a variety of texts on their independent and instructional levels with appropriate levels of support and scaffolding.

Table 1 Children's levels of reading in relationship to text difficulty

<i>Level</i>	<i>Word recognition</i>	<i>Comprehension</i>
Independent	99–100%	90–100%
Instructional	95–98%	75–89%
Borderline instructional	90–94%	50–74%
Frustration	Below 90%	Below 50%

Summary

To learn to read, learners need opportunities to learn and integrate numerous underpinnings of reading, identify printed words, develop comprehension, and increase reading fluency. Underpinnings include oral expressive language, listening comprehension, forms and uses of written language, knowledge of the written symbol system, concepts of print, and becoming enthusiastic about reading and writing, motivation. These essential skills and knowledge continue to develop during formal instruction in learning to read and as students continue to build reading competence. To identify printed words, children continue to develop their understanding of the alphabetic system and apply a variety of strategies to decode unfamiliar words. To comprehend, children construct meaning as they interact with the text and draw upon their prior knowledge and experiences. Reading fluency occurs

when children's decoding and comprehension enables them to read with speed, accuracy, and prosody. A focus on any one aspect of learning to read should not be at the expense of an emphasis on other aspects. To the extent possible, different aspects should be integrated into instruction and the relationship among the processes should be the focus of instruction.

See also: An Overview of Language and Literacy in Educational Settings; Reading Comprehension: Reading For Learning.

Further Reading

Cunningham, P. M. and Cunningham, J. W. (2002). What we know about how to teach phonics. In Farstrup, A. E. and Samuels, S. J. (eds.) *What Research Has to Say about Reading Instruction*, pp 87–109.

Ehri, L. C. and McCormick, S. (1998). Phases of word learning: Implications for instruction with delayed and disabled readers.

Reading and Writing Quarterly: Overcoming Learning Difficulties **14**(2), 135–163.

Juel, C. and Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly* **35**, 458–492.

Nagy, W. E. and Scott, J. A. (2000). Vocabulary processes. In Kamil, M. L., Mosenthal, P. B., Pearson, P. D., and Barr, R. (eds.) *Handbook of Reading Research: Vol. III*, pp 269–284. Mahwah, NJ: Erlbaum.

Pressley, M. (2000). What should comprehension instruction be the instruction of? In Kamil, M. L., Mosenthal, P. B., Pearson, P. D., and Barr, R. (eds.) *Handbook of Reading Research: Vol. III*, pp 545–561. Mahwah, NJ: Erlbaum.

Snow, C. E., Burns, M. S., and Griffin, P. (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.

Snow, C. E., Griffin, P., and Burns, M. S. (eds.) (2005). *Knowledge to Support the Teaching of Reading: Preparing Teachers for a Changing World*. Indianapolis, IN: Jossey-Bass.

Snow, C. E. and Páez, M. (2004). The head start classroom as an oral language environment: What should the performance standards be? In Zigler, E. and Styfco, S. (eds.) *The Head Start Debates (Friendly and Otherwise)*, pp 215–244. Baltimore, MD: Brookes Publishing.

Tolchinsky, L. (2003). *The Cradle of Culture and What Children Know about Writing and Numbers Before Being Taught*. Mahwah, NJ: Erlbaum.

Mathematics Learning

L Verschaffel, Katholieke Universiteit Leuven, Leuven, Belgium

B Greer, Portland State University, Portland, OR, USA

E De Corte, Katholieke Universiteit Leuven, Leuven, Belgium

© 2010 Elsevier Ltd. All rights reserved.

Introduction

From some decades now, research in mathematics education has been emerging as a field of study in its own right. In building theoretical frameworks specific to its domain, while still productively drawing on cognitive and instructional psychology, this field is also dedicated to the mission of examining and improving the institutions and practices of school mathematics education. Accordingly, there is increasing interest in other social science disciplines and recognition of the cultural embedding of mathematics (education). After a long struggle for its identity, today one can say that mathematics education has become a full-fledged, interdisciplinary field of research and analysis, aiming at a better understanding of the processes underlying the acquisition and development of mathematical knowledge, skills, beliefs, and attitudes applied to the design of valuable tools and powerful environments for teaching/learning mathematics.

This article presents a review of important recent themes and developments in research on the learning and teaching of mathematical knowledge and thinking. As a framework we use a model for the design of powerful environments for learning and teaching mathematics that is structured according to four interrelated components, namely, competence, learning, intervention, and assessment (CLIA-model) (De Corte *et al.*, 2004). Space allows only a few specific examples, and like the research in the field, these are concentrated around primary and secondary school students' learning of numbers and arithmetic. For comprehensive reviews we refer to the following handbooks and handbook chapters (this list is itself testimony to the vitality of the field): Bishop *et al.* (2003); Bishop *et al.* (1996); De Corte *et al.* (1996); Ginsburg *et al.* (1998); Grouws (1992); Gutierrez and Boero (2006); and Lester (2007).

Components of Mathematical Competence

In recent times, discourse about what competence in mathematics means has been enriched by introduction of the notion of "mathematical disposition." For instance, in *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics, 1989), it is stated:

Learning mathematics extends beyond learning concepts, procedures, and their applications. It also includes developing a disposition toward mathematics and seeing mathematics as a powerful way for looking at situations. Disposition refers not simply to attitudes but to a tendency to think and to act in positive ways. Students' mathematical dispositions are manifested in the way they approach tasks – whether with confidence, willingness to explore alternatives, perseverance, and interest – and in their tendency to reflect on their own thinking. (Curriculum and Evaluation Standards for School Mathematics, 1989: 233).

This enhanced view is in agreement with the characterization of "mathematical proficiency" in the report of the National Research Council (2001a: 5), *Adding it up*, wherein proficiency is defined in terms of five interwoven strands, namely:

- conceptual understanding, "comprehension of mathematical concepts, operations, and relations"
- procedural fluency, "skill in carrying out procedures flexibly, accurately, efficiently, and appropriately"
- strategic competence, "the ability to formulate, represent, and solve mathematical problems"
- adaptive reasoning, "the capacity for logical thought, reflection, explanation, and justification"
- productive disposition, "a habitual inclination to see mathematics as sensible, useful, and worthwhile, coupled with a belief in diligence and one's efficacy".

The above descriptions are also consistent with more general models of competence which stress the acquisition of (1) a well-organized and flexibly accessible domain-specific knowledge, (2) heuristic methods, (3) meta-knowledge, (4) self-regulatory skills, and (5) affective components involving beliefs about, attitudes toward, and emotions about the domain and about thinking, learning, and teaching in this domain (De Corte *et al.*, 1996).

Using still another construct that was introduced in the literature more than two decades ago by Hatano and Inagaki (1986), one can say that what students have to acquire in mathematics education is "adaptive expertise" as opposed to "routine expertise." Hatano (2003: xi) describes adaptive expertise as "the ability to apply meaningfully learned procedures flexibly and creatively"; in contrast, routine expertise involves "simply being able to

complete school mathematics exercises quickly and accurately without (much) understanding” (Hatano 2003: xi). Like the other characterizations of (mathematical) proficiency or competence mentioned earlier, Hatano’s concept emphasizes that gaining expertise is accompanied by changes in interest, values, and identity besides changes in the purely cognitive realm.

A major point of agreement is expressed thus in the report of the National Research Council (2001a: 133): “Learning is not an all-or-none phenomenon, and as it proceeds, each strand of mathematical proficiency should be developed in synchrony with the others. That development takes time.” Consequently, from the very beginning of mathematics education, attention needs to be paid to the parallel and integrated acquisition of the different components of competence.

The handbook publications cited earlier contain numerous illustrations of the importance and interdependency of the different strands of proficiency in domains such as single-digit computation, multidigit arithmetic, estimation, word problem solving, etc. A prime example is the prominent position accorded to the notion of ‘number sense’ from the earliest grades of school, as mentioned in recent reform documents worldwide. Number sense refers to “a person’s general understanding of number and operations. It refers to a person’s general understanding of number and operations along with the ability and inclination to use this understanding in flexible ways to make mathematical judgements and to develop useful strategies for handling numbers and operations. It reflects an inclination and an ability to use numbers and quantitative methods as a means of communicating, processing and interpreting information. It results in, and reciprocally derives from, an expectation that numbers are useful and that mathematics has a certain regularity” (McIntosh *et al.*, 1992: 3).

Mathematics Learning

Research over the past decades has resulted in the view that mathematics learning is the active and cumulative construction – situated in a given community of learners but also within a broader sociocultural context – of mathematical meaning, understanding, and skills (De Corte *et al.*, 1996; Verschaffel *et al.*, 2007).

The view that learning is an active and constructive process has become common ground among mathematics educators, and there is substantial empirical evidence supporting it (see e.g., National Research Council, 2001a). What is essential in the constructivist view on learning is the mindful and effortful involvement of learners in the processes of knowledge and skill acquisition in interaction with the environment (physical and social) and building on their prior knowledge. This involvement is well illustrated in the idiosyncratic but accurate calculation procedures

that are invented by children in and out of school contexts (Ginsburg *et al.*, 1998; Nunes, 1992). However, the constructive nature of learning is also evidenced in a negative way in the misconceptions (such as the idea that ‘multiplication always makes bigger and division always makes smaller’) and defective procedures (such as, in multidigit subtraction, subtracting in each column the smaller digit from the larger) that many learners acquire (De Corte *et al.*, 1996; Verschaffel *et al.*, 2007).

The cumulative feature of learning, which is closely linked to its constructive nature, refers to the pivotal role of prior knowledge, again in both positive and negative ways. Research on conceptual change has yielded ample evidence that, as in scientific reasoning, students develop an initial framework theory about number, organized on the basis of certain core principles or presuppositions, based on their (out-of-school) experience with natural numbers, and that these framework theories facilitate some kinds of learning but inhibit others. For example, an early understanding of natural number and its properties strongly supports children’s further learning in the domain of natural numbers, while at the same time it stands in the way of their understanding of properties and operations of numbers beyond the natural numbers, such as rational numbers, negative numbers, and real numbers (Verschaffel *et al.*, 2007).

The idea that learning and cognition are situated activities emerged strongly in the late 1980s in reaction to the then dominant cognitivist view of learning and thinking as highly individual and purely mentalistic processes occurring in the brain, thus resulting in encapsulated mental representations (Brown *et al.*, 1989). The cognitivist view is in line with Sfard’s (1998) acquisition metaphor of learning focused on individual enrichment through acquiring knowledge, skills, etc. By contrast, the situated perspective converges with the participation metaphor that Sfard identifies, stressing that learning is enacted essentially in interaction with social and cultural contexts and artifacts, and especially through participation in cultural activities and contexts (Schoenfeld, 2006; Sfard, 1998; Verschaffel *et al.*, 2007). Such a socioculturally situated conception of learning and cognition is nowadays widely shared in the mathematics education community and encapsulated in the perspective of ethnomathematics (D’Ambrosio, 2006). The outcomes of a large series of ethnomathematical studies of the informal calculation procedures, problem-solving strategies, and learning mechanisms of particular groups of children and adults involved in everyday cultural practices such as business, tailoring, weaving, carpentry, grocery-packing, cooking, etc. has largely contributed to the popularity of the situated approach (for summaries see De Corte *et al.*, 1996; Nunes, 1992; Schoenfeld, 2006).

Although the situated nature of learning has been documented especially well within out-of-school contexts, it is obvious that this situatedness applies to school

learning also. For instance, the numerous examples of students' "suspension of sense-making" when doing school word problems (leading to computationally based solutions to absurd problems like "There are 26 sheep and 10 goats on a ship. How old is the captain?") can be considered as another line of evidence for the importance of the situatedness of mathematical thinking and learning (Verschaffel *et al.*, 2000).

Of special importance from an educational perspective is that the situativity view implies the importance of, and thereby reinforces the case for, interaction and collaboration in learning. Mathematics educators have broadly embraced the view that learning is not a solo and internal activity; rather, learning efforts are distributed over the individual student, his/her partners in the learning environment, and the (technological) resources and tools available. For instance, Yackel and Cobb (1996) consider social interaction essential for mathematics learning, with individual knowledge construction occurring throughout processes of interaction, negotiation, and cooperation. Although the situated approach has served an important role in de-emphasizing individual internal processes and stressing the sociocultural and collaborative aspects that were previously neglected, it also led to some misguided claims about mathematics learning and inappropriate educational suggestions (e.g., that learning is always grounded in concrete situations, knowledge has become less important, knowledge does not transfer between tasks, stimulating abstraction is of little value, etc.) (Vosniadou and Vamvakoussi, 2006). Accordingly, several researchers have warned against totally abandoning the individual acquisition perspective on (mathematics) learning and have made a plea for a proper balance between the individual and social (Sfard, 1998; Vosniadou and Vamvakoussi, 2006). According to Schoenfeld (2006), how to put things together – how to see everything connected to an individual (both in the sense of knowledge, identity, etc., and in terms of that person's relationship to various communities), and the communities to which the individual belongs, as a coherent whole – is one of the great challenges for the future.

The preceding discussion shows that recent research provides substantial evidence supporting the view that productive mathematics learning is a constructive and situated process of knowledge-building and skill-acquisition, involving ample opportunities for interaction, negotiation, and collaboration. Consequently, it will not come as a surprise that this integrated new perspective on learning mathematics has stimulated researchers and practitioners to take these basic characteristics of learning as major guidelines for the design of new curricula, textbooks, learning environments, and assessment instruments that aim at fostering in students the acquisition of a mathematical disposition. It should be acknowledged that such an approach to mathematics teaching and assessment was pioneered earlier by several (groups of) mathematics

educators, in particular Freudenthal (1983), who, with his colleagues, developed and implemented realistic mathematics education in The Netherlands long before these new ideas came into vogue.

Instruction

How can powerful mathematics learning environments be designed for inducing in students the intended learning activities and processes, and by so doing, effectively fostering in them the progressive development of a mathematical disposition?

Mainly over the past 15 years, this challenge has been addressed through intervention studies, under the banner of such terms as constructional research (Becker and Selter, 1996), developmental research (Gravemeijer, 1994), and design experiments (Cobb *et al.*, 2003). Becker and Selter (1996) define constructional research "as research that is connected with suggestions on how teaching ought to be or could be, to put it slightly more moderately. . . [it is] concentrating on the development of theoretically founded and empirically tested practical suggestions for teaching" (p. 525). According to Cobb *et al.* (2003: 9) "design experiments entail both 'engineering' particular forms of learning and systematically studying those forms of learning within the context defined by the means of supporting them. This designed context is subject to test and revision, and the successive iterations that result play a role similar to that of systematic variation in experiments." Note that, besides contributing to the innovation and improvement of classroom practices, this type of research is intended to advance theory-building about learning from instruction (Cobb *et al.*, 2003; De Corte *et al.*, 2004).

It is possible to identify many common features or family resemblances of the environments that have been designed, implemented, and evaluated in such studies, including: (1) the active and constructive view on mathematics learning; (2) the recognition of the crucial role of students' prior knowledge for future learning; (3) the orientation towards the integrated mastery of procedural skill, conceptual understanding and problem solving; (4) a reluctance to impose a single representational model or solution strategy on all learners and the pursuit of adaptive expertise (Hatano, 2003); (5) the importance of social interaction and collaboration in doing and learning mathematics, and more generally, of conceiving learning mathematics as a community activity; (6) efforts to embed mathematics learning in authentic and meaningful contexts; (7) the focus on basic principles or, on big ideas that underlie numerous concepts and procedures across topics; (8) the belief that the above instructional design principles do not hold only for teaching mathematics to mathematically able and/or higher socioeconomic status (SES) children, but also to low achievers and/or low SES learners; and

(9) substantial attention to the teacher's pivotal role in the children's learning process. These common features collectively imply a radical departure from a traditional skills approach to mathematics instruction based on the view that mathematics learning is a highly individual activity consisting mainly in absorbing and memorizing a fixed body of decontextualized and fragmented knowledge and procedural skills transmitted by the teacher (Baroody, 2003; Verschaffel *et al.*, 2007). Importantly, these features can also be differentiated from a *laissez-faire* approach to mathematics education, which was observed in the initial period of the implementation of the reform in countries such as the United States and the Netherlands. They generally fit with what Baroody (2003) called the investigative approach, that is, a blend of a conceptual and a problem-solving approach, aiming at the mastery of basic skills, conceptual learning, and mathematical thinking and characterized by both meaningful and inquiry-based instruction and purposeful learning and practice.

Assessment

A final component of the CLIA-model, namely assessment, is concerned with the design, construction, and use of instruments for determining how powerful learning environments are in facilitating in students the acquisition of (the different aspects of) a mathematical disposition. Assessments of mathematics learning can either be internal or external. Internal assessments are organized by the teacher in the classroom, formally or more informally. By contrast, external, usually large-scale, assessments are organized at the district, state, national, or even international level using standardized tests or surveys (De Corte *et al.*, 1996; National Research Council, 2001b). As argued by the National Research Council (2001b), assessments serve different purposes including assisting learning and teaching, measuring achievement of individual pupils, and evaluating school programs.

Particularly, but by no means exclusively, in the United States, external assessment of what students have learnt and especially whether they attain a certain achievement or proficiency level – high-stakes testing – is a hotly debated topic, especially in relation to mathematics learning and teaching. (The intensity of emotion generated by this topic, as described in Schoenfeld (2006), is particularly striking.) The massive use of standardized tests in education has always been more customary in the United States as compared, for instance, to Europe. But the 2001 No Child Left Behind Act and the related quest for accountability have both increased this practice and intensified the debates about the effectiveness and desirability of high-stakes testing. Especially since the beginning of the 1990s, traditional tests have been increasingly criticized (see e.g., De Corte *et al.*, 1996; Schoenfeld, 2006). Analyses

of widely used standardized tests show that there is a mismatch between the new vision of mathematical competence as described earlier and the content and style of those tests. Partly due to the excessive use of multiple-choice item format, the tests focus on the assessment of memorized facts, rote knowledge, and lower-level procedural skills (routine expertise, to use Hatano's (2003) phrase). On the other hand, they do not sufficiently yield relevant and useful information on pupils' abilities in problem-solving, modeling complex situations, communicating mathematical ideas, other higher-order components of mathematical activity and a mathematical disposition, and, generally, adaptive expertise (Hatano, 2003). A related criticism points to the one-sided orientation of the tests toward the products of pupils' mathematics work and the neglect of the underlying processes (De Corte *et al.*, 1996). An important consequence of this state-of-the-art is that assessment often has a negative impact on the implemented curriculum, the classroom climate, and instructional practices. Apart from the previous intrinsic criticisms on traditional standardized achievement tests, a major issue of debate is the accountability over their use as high-stakes tests, that is, their mandatory administration for collecting data on the attainment of students as a basis for highly consequential decisions about students (e.g., graduation), teachers (e.g., financial rewards), and schools and school districts (e.g., accreditation). According to the No Child Left Behind Act, this accountability use should result in the progressive acquisition, by all students, of a proficiency level in reading and mathematics. However, there are serious doubts whether current testing programs really improve learning and instruction (Amrein and Berliner, 2002). Moreover, there are many reports of unintended but unfavorable consequences, such as increased dropout rates, negative impact on minority and special education children, cheating on examinations by teachers and students, teachers leaving the profession, etc. Nowadays, most researchers in (mathematics) education believe that for large-scale assessments to indeed serve to improve student learning, it will be necessary to move away from the rationale, the constraints, and the practices of current high-stakes testing programs (Amrein and Berliner, 2002; National Research Council, 2001b).

Notwithstanding the relevance and importance of large-scale, external assessments, they need to be supplemented by internal classroom testing, whereby teachers continuously monitor students' progress in understanding and mastering knowledge and skills as a basis for guiding and supporting further learning and, if needed, for providing timely remediation. Such formative assessments also provide students themselves with informative feedback as a basis for monitoring and regulating their own learning (see e.g., National Research Council, 2001b). Whereas external assessments are useful and important for the large-scale monitoring of trends in mathematics education,

classroom assessment affords ongoing day-to-day basis information to improve student learning, taking into account the strengths and weaknesses of the class as a group as well as those of the individual pupils. Convincing examples of the usefulness of identifying students' reasoning underlying their responses as a valuable source for the regulation of learning are given by Wiliam (2007).

The preceding discussion shows that using assessment to assist instruction requires that assessment and instruction should be integrated as envisioned by the National Research Council (2001b): "Assessment should be an integral part of teaching. It is the mechanism whereby teachers can learn how students think about mathematics as well as what students are able to accomplish" (p. 69). For gathering data about pupils' performance and progress, teachers can use a variety of techniques: informal questions, seatwork and homework tasks, clinical interviews, portfolios, and more formal instruments such as teacher-made classroom tests, learning potential tests, and progress maps.

General Conclusions

Using the CLIA framework as an organizing device, this article has presented a selective review of research on development, learning, instruction, and assessment relating to mathematics that offers promise for innovation in, and improvement of, mathematics classroom practices in line with the new international perspectives on the goals and nature of mathematics education as manifested in reform documents worldwide.

With respect to each of the four interconnected components our empirically based knowledge has substantially advanced over the past decades, enabling a progressively better understanding concerning the components that constitute a mathematical disposition, the nature of the learning and developmental processes that should be induced in students to facilitate the acquisition of competence, the characteristics of learning environments that are powerful in initiating and evoking those processes, and finally, the kind of assessment instruments that are appropriate to help monitor and support learning and teaching.

An important question to ask here is whether this expanding knowledge base is relevant and useful in view of bridging the long-standing gap between theory/research and practice, and thus can contribute to the actual improvement of mathematics education practices. The available studies wherein reform-based instructional programs have been implemented and evaluated warrant some optimism. Indeed, the increasing number of "success stories" (e.g., see Becker and Selter, 1996; Verschaffel *et al.*, 2007) are building up increasingly to a critical mass of results, showing that under certain conditions, carefully designed, research-based learning environments can yield learning

outcomes in students, including weaker children and children from SES minority groups, that are in accordance with the current view of the goal of mathematics education as the acquisition of a mathematical disposition, including measures of conceptual understanding and problem solving, while showing either no significant differences or an advantage to the new curricula on measures of computational skills. However, while Schoenfeld (2006: 494) also concludes that "the 'score sheet' is uniformly in favour of the standards-based curricula," at the same time he acknowledges that at present "the evidence base is embarrassingly weak" (p. 494).

Optimism based on the available research is bolstered by the observation that inquiry-based ideas are gradually taking root in the mathematics education community, not just as manifested in the reform documents worldwide, and (subsequently) in curricula and textbooks, but also in the writings and practices of knowledgeable educational professionals (although, again for the United States, Schoenfeld's (2006) short-term view of the situation is much more pessimistic).

However, this general optimism in the long term is tempered by two major challenging problems for future research and development. The first issue relates to upscaling the new perspective on learning and teaching mathematics, and the design principles for learning environments that derive from it. The second (related) problem concerns the sustainability of innovative learning environments. Solving both problems has a serious price ticket and is largely a matter for educational policy and innovation and in preservice and in-service teacher professional development (see also: Schoenfeld, 2006; Verschaffel *et al.*, 2007; issues that are addressed elsewhere in this encyclopedia). As with the debate over assessment, and the perennial search for a proper balance and integration of routine expertise and adaptive expertise, it is increasingly being recognized within the field that these are fundamentally political issues (Schoenfeld, 2006) and that the endeavor of research in mathematics education is itself situated in social and political contexts (Valero and Zevenbergen, 2004).

See also: Constructivism and Learning; Learning as Inquiry; Learning Outside of School; Situated View of Learning.

Bibliography

- Amrein, A. L. and Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Educational Policy Analysis Archives* **10**(18). <http://epaa.asu.edu/epaa/v10n18> (accessed August 2009).
- Baroody, A. J. (2003). The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge.

- In Baroody, A. J. and Dowker, A. (eds.) *The Development of Arithmetic Concepts and Skills*, pp 1–34. Mahwah, NJ: Erlbaum.
- Becker, J. P. and Selter, C. (1996). Elementary school practices. In Bishop, A., Clements, K., Keitel, C., Kilpatrick, J., and Laborde, C. (eds.) *International Handbook of Mathematics Education*, pp 511–564. Dordrecht: Kluwer.
- Bishop, A., Clements, K., Keitel, C., and Laborde, C. (eds.) (1996). *International Handbook of Mathematics Education*. Dordrecht: Kluwer.
- Bishop, A. J., Clements, M. A., Keitel, C., Kilpatrick, J., and Leung, F. S. K. (eds.) (2003). *Second International Handbook of Mathematics Education*. Dordrecht: Kluwer.
- Brown, J. S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* **18**(1), 32–42.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., and Schauble, L. (2003). Design experiments in educational research. *Educational Researcher* **32**(1), 9–13.
- D'Ambrosio, U. (2006). *Ethnomathematics: Link between Traditions and Modernity*. Rotterdam/Taipei: Sense Publishers.
- De Corte, E., Greer, B., and Verschaffel, L. (1996). Mathematics teaching and learning. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 491–549. New York: Macmillan.
- De Corte, E., Verschaffel, L., and Masui, C. (2004). The CLIA-model: A framework for designing powerful learning environments for thinking and problem solving. *European Journal of Psychology of Education* **19**, 365–384.
- Freudenthal, H. (1983). *Didactical Phenomenology of Mathematical Structures*. Dordrecht: Reidel.
- Ginsburg, H. P., Klein, A., and Starkey, P. (1998). The development of children's mathematical thinking: Connecting research with practice. In Sigel, I. E. and Renninger, K. A. (eds.) *Handbook of Child Psychology: Volume 4. Child Psychology in Practice*, 5th edn., pp 401–476. New York: Wiley.
- Gravemeijer, K. (1994). *Developing Realistic Mathematics Education*. Utrecht: Freudenthal Institute, University of Utrecht.
- Grouws, D. A. (ed.) (1992). *Handbook of Research on Mathematics Teaching and Learning*. New York: Macmillan.
- Gutierrez, A. and Boero, P. (eds.) (2006). *Handbook of Research on the Psychology of Mathematics Education. Past, Present, and Future*. Rotterdam: Sense Publishers.
- Hatano, G. (2003). Foreword. In Baroody, A. J. and Dowker, A. (eds.) *The Development of Arithmetic Concepts and Skills*, pp xi–xiii. Mahwah, NJ: Erlbaum.
- Hatano, G. and Inagaki, K. (1986). Two courses of expertise. In Stevenson, H., Azumam, H., and Hakuta, K. (eds.) *Child Development and Education in Japan*, pp 262–272. New York: Freeman.
- Lester, F. K. Jr. (2007). *Second Handbook of Research on Mathematics Teaching and Learning*. Greenwich, CT: Information Age Publishing.
- McIntosh, A., Reys, B. J., and Reys, R. E. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics* **12**(3), 2–8.
- National Council of Teachers of Mathematics (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- National Research Council (2001a). *Adding It up: Helping Children Learn Mathematics*. Washington, DC: National Academy Press.
- National Research Council (2001b). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.
- Nunes, T. (1992). Ethnomathematics and everyday cognition. In Grouws, D. A. (ed.) *Handbook of Research on Mathematics Teaching and Learning*, pp 557–574. New York: Macmillan.
- Schoenfeld, A. H. (2006). Mathematics teaching and learning. In Alexander, P. A. and Winne, P. H. (eds.) *Handbook of Educational Psychology*, 2nd edn. pp 479–510. New York: Macmillan.
- Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher* **27**(2), 4–13.
- Valero, P. and Zevenbergen, R. (eds.) (2004). *Researching the Socio-Political Dimensions of Mathematics Education: Issues of Power in Theory and Methodology*. Dordrecht: Kluwer.
- Verschaffel, L., Greer, B., and De Corte, E. (2000). *Making Sense of Word Problems*. Lisse: Swets and Zeitlinger.
- Verschaffel, L., Greer, B., and De Corte, E. (2007). Whole number concepts and operations. In Lester, F. K. (ed.) *Second Handbook of Research on Mathematics Teaching and Learning*, pp 557–628. Greenwich, CT: Information Age Publishing.
- Vosniadou, S. and Vamvakoussi, X. (2006). Examining mathematics learning from a conceptual change point of view. In Verschaffel, L., Dochy, F., Boekaerts, M., and Vosniadou, S. (eds.) *Instructional Psychology: Past, Present and Future Trends*, pp 55–72. Oxford: Elsevier Ltd.
- William, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In Lester, F. K. (ed.) *Second Handbook of Research on Mathematics Teaching and Learning*, pp 1051–1098. Greenwich, CT: Information Age Publishing.
- Yackel, E. and Cobb, P. (1996). Sociomathematical norms, argumentations, and autonomy in mathematics. *Journal for Research in Mathematics Education* **27**, 458–477.

Relevant Websites

- <http://www.mathunion.org/ICMI> – The International Commission on Mathematical Instruction (ICMI).
- <http://igpme.org> – The International Group for the Psychology of Mathematics Education (PME).
- <http://www.nctm.org> – National Council of Teachers of Mathematics (NCTM).

Music Learning

S Hallam, Institute of Education, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Learning in Music

Learning is a natural process for human beings. We are pre-programmed to learn. In our everyday lives, during our interactions with others and with the environment, we are constantly engaged in learning. This may be deliberate and intentional or incidental and without conscious awareness. Whether what we have learned is retained in the long term depends on the extent of our ongoing engagement with it and how important it is to us. Learning in music is no exception to these general principles.

Music is a universal trait of humankind. We all have the capacity to learn in music, although whether everyone inherits the same level of potential continues to be hotly debated. While we cannot completely rule out the possibility that there may be some innate functional and structural brain differences which may predispose some individuals to become musicians, there is ever-increasing evidence that musical training itself leads to changes in brain function and structure (Schlaug, 2003). This suggests that whatever genetic inheritance an individual may have is greatly enhanced by a musically enriched environment.

Musical ability tests were first developed in the late nineteenth century to select children to play an instrument in formal educational settings, the underlying assumption being that musical ability was innate or at least determined at an early age. Resources were limited and the best way to ensure that resources were well used was to select children on the basis of their musical potential. The tests focused on aural abilities, usually pitch and rhythm, although some later tests attempted to assess musicality. We now know that these tests assess what has already been learned and may therefore discriminate against those who have experienced an impoverished musical environment prior to testing. They also neglect to assess many of the skills that are required for developing musical expertise.

Musical Skills and Endstates

A range of diverse and complex skills are involved in actively making music and in training to be a professional within the modern music industries as a composer, performer, producer, technician, critic, manager, retailer, or publicist. **Table 1** sets out the wide range of skills that may be involved in actively engaging in creating and making music.

Music Learning in Infants and Young Children

The implicit musical knowledge that adults acquire over time is built on structures present in infancy. These provide the basis for perceptual learning and enculturation, the process by which the child develops internal representations of the music of its culture. This process is mediated by the learning environment and begins when the human auditory system is functional, 3–4 months before birth. In early infancy, parents and the family provide musical stimulation including the use of motherese, nonverbal communication that takes place between parents and infants. These pre-verbal quasi-musical interactions constitute the beginnings of musical competences. Infants are also especially sensitive to lullabies. These have smooth pitch contours, a strong tonal center, repetitive rhythms, and a distinct vocal style, and when adults sing them to infants they tend to sing at a higher pitch than normal, with a slower tempo, and exaggerated emotion, a style of singing that is found in childcare contexts across cultures. Newborns and infants listen more attentively to singing which is in this infant-directed style.

From birth, the infant has very well-developed systems for processing music. Infants are predisposed to attend to melodic contour, rhythmic patterning, and consonant sounds and are similar to adults in their sensitivity to the pitch and rhythmic grouping of sounds. However, the complex skills required for understanding and analyzing music within any particular culture take time to develop and depend on the type and extent of exposure to music. The scale patterns of particular cultures are not innately known and have to be learned (Trehub and Trainer, 1990). In Western cultures, 1-year-old infants respond differently to diatonic and nondiatonic patterns before by 3 or 4 years, display the ability to respond to the degree of tonality of the stimuli. This process is likely to be aided by the tonal construction of nursery songs that usually emphasize the pitch content of basic scales. Depending on their musical environment, by the age of 5 years, children can generally organize songs around stable tonal keys but do not have a stable tonal scale system that can be used to transpose melodies. This develops later (Lamont and Cross, 1994). Children in their early years also begin to recognize musical depictions of emotion and respond to them. The specific age at which this occurs is likely to depend on the extent of the child's exposure to music within and across genres

Table 1 Musical skills*Aural skills are required for developing:*

- Rhythmic accuracy and a sense of pulse
- Good intonation
- The facility to know how music will sound without having to play it
- Improvisational skills

Cognitive skills are required in the processes of:

- Reading music
- Transposition
- Understanding keys
- Understanding harmony
- Understanding the structure of music
- The memorization of music
- Composing
- Understanding different musical styles and their cultural and historic contexts
- Evaluating and appraising

Technical skills are required for developing:

- Instrument-specific skills
- Technical agility
- Articulation
- Expressive tone quality

Musicianship skills are concerned with:

- Being able to play expressively
- Being able to project sound
- Developing control
- Conveying meaning

Performance skills include:

- Being able to communicate with an audience
- Communicating with other performers
- Being able to coordinate a group
- Presenting to an audience

*Learning skills are related to: being able to learn, monitor, and evaluate progress independently**Life skills include:*

- Social skills (being able to work with other musicians, promoters, the public)
- Planning and organizational skills (planning practice schedules, programs, travel arrangements)
- Time management (being punctual, meeting deadlines)

and their general cognitive and emotional development. A small number of children develop absolute pitch, the ability to identify or produce isolated pitches in the absence of a reference pitch. We may all be predisposed to develop this, but require an appropriate musical environment that provides exposure to specific pitches and a means of labeling them when young.

From 9 months, infants begin to make spontaneous babbling or singing sounds and, after 18 months, they begin to spontaneously generate music that has systematic form, uses discrete pitch levels, repeats rhythmic and melodic contours, and is recognizable as song. However, they lack a stable pitch framework and before in any single song, a very limited set of phrase contours are used. By the age of 5 years, individual contours and intervals are produced accurately and children can produce recognizable songs with stable tonality. They also improvise and invent songs.

Children are exposed to music in the home through the radio, recordings, or the television. The nature of that music will influence their musical enculturation. Globalization has meant that worldwide, most children are exposed to Western popular music. From infancy, children acquire implicit knowledge about the music they hear, its structures, rhythms, and tonality. The greater the exposure to music, the more fully and speedily this knowledge is acquired. It is not necessary for children to focus on listening to music for this knowledge to develop. It is learned without conscious awareness.

Listening, Responding to, and Appraising Music

When we listen to music, we process an enormous amount of information rapidly and often without conscious awareness. The ease with which we do this depends on our prior musical experiences and the culturally determined tonal scheme to which we have become accustomed. This knowledge is not always available to conscious thought and is applied automatically whenever we listen to music.

Human beings respond to music in a variety of different ways. Responses can be physiological, motor, intellectual, esthetic, emotional, or in relation to mood or arousal. In education, the focus has been on the intellect and the esthetic. However, other responses may be more powerful. Individuals can have very strong emotional experiences to music (see Gabrielsson, 2001 for a review), although music can also play an important role in helping overcome powerful emotions.

Listening plays a central role in the appraisal and evaluation of active music making, guiding future planning and activity. The process of appraising involves listening and making comparisons (implicit or explicit) with already-acquired internal representations of music. These may be specific, that is, of the same piece of music; generalized, but still pertaining to particular types of music; or generalized, relating to particular features of performance, for example, intonation, tone quality. Without comparison, with some already-held conception, appraisal cannot occur. Even quite young children are able to articulate ideas about their listening processes and as they get older they more frequently predict, compare, evaluate, express preferences, reflect, recognize, and make judgments about music. Musically trained children demonstrate similar skills to adults suggesting that the ability to think critically while listening to music is related to experience with music rather than age.

Acquiring Skills for Active Music Making

Learning to play an instrument, developing vocally, improvising, or composing, all require the acquisition of a wide range of skills. Three main stages have been

identified in skill development. In the initial cognitive-verbal-motor-stage, learning is consciously controlled, the learner understands what is required, and carries it out while consciously providing self-instruction. Learning is supported when the learner has a clear mental representation of both the process and the goal of learning and feedback is available, either directly from the environment or from an observer, usually a teacher. In the associative stage, the learner begins to put together a sequence of responses that becomes more fluent over time. Errors are detected and eliminated and feedback from others or self-monitoring continues to be important. In the final autonomous stage, the skill become automated, is carried out without conscious effort, and continues to develop each time it is used, becoming more fluent and quicker. As automaticity develops, the component processes become unavailable for conscious inspection and the learner has difficulty in explaining his or her actions to others (Fitts and Posner, 1967). A similar set of processes occurs in relation to the acquisition of knowledge. Initially, learners are unable to discern the difference between accurate/inaccurate and relevant/tangential knowledge. At the second stage, a foundational body of relatively cohesive, structured domain knowledge has been acquired which can be applied adopting surface and deep approaches to learning. To move to the third stage, a synergy among elements of knowledge is required. To attain high levels of coherence and automaticity requires increasing levels of personal motivation.

The acquisition of knowledge and skills is inextricably intertwined. Knowledge-based mental representations of what the music should sound like are required to check for errors; monitor progress toward performance, improvisational, or compositional goals; to select possible strategies for continued working; and to evaluate the final outcome. Due to the complexity of learning in music, as one set of skills becomes increasingly automated, others will be at earlier stages. The quality of the processing during these various stages is important in determining learning outcomes and is supported by having appropriate representations of what is to be attained and receiving appropriate feedback directly (from the sounds produced) or from observers. Knowledge of high-quality representations can be acquired through listening, imitation, analysis, and participating in making music with others who are more expert.

As the individual's expertise increases and skills become automated, s/he can more easily take a holistic view of the task and perceive large meaningful patterns. S/he is likely to spend longer analyzing the problem and represent it at a deeper level, while the speed of processing and memory related to the area of expertise increases. Monitoring of progress and evaluation of outcomes improves because of the high level of expert knowledge. Those with high levels of expertise know why they make

errors, why they fail to comprehend, when they need to check their solutions, and what they need to do next. They are better than novices in judging the difficulty of a problem and selecting appropriate strategies to solve it.

For high levels of musical expertise to develop, the individual needs to be actively engaged with music over a long, sustained period of time, in the case of professional musicians, many years. This may be in formal or informal learning contexts and involves systematic practice or playful, creative activity. Transfer of learning takes place in relation to some musical activities, for instance, in sight-reading, technical instrumental skills, and knowledge of a particular genre; but transfer between genres, for instance, classical and jazz, and music based on different tonalities is limited, and different types of skills are required for performance, improvisation, and composition. To perform well, particularly in formal contexts, in addition to being well prepared musically and technically, requires an understanding of how to utilize stage fright to enhance performance rather than allow it to disrupt it, while improvisation and composition require creative skills.

Practice is essential for all musical learning but is more effective when it is not merely repetitious but is systematic with careful monitoring of progress toward goals and the adoption of alternative strategies when progress is not being made. Learners benefit from having knowledge of a wide range of learning and problem-solving strategies, being aware of their own strengths and weaknesses, and having high-quality models to emulate. Mental rehearsal, where the learner imagines playing or performing, has been demonstrated to be effective as it stimulates the same processes in the brain as physically undertaking the activity. This requires no physical resources and therefore it enables practice to be undertaken at any time.

Metacognitive strategies concerned with the planning, monitoring, and evaluation of learning are crucial to all aspects of engagement with music and can be considered at the level of a particular task or in relation to the more global concerns of the learner to maintain or improve the standard of their work. In both cases, the knowledge of personal strengths and weaknesses, the nature of the task to be completed, possible strategies, and the nature of the learning outcome are important. There are considerable differences between beginners, novices, and experts in their knowledge and deployment of different learning and self-regulating strategies, as well as individual differences among musicians and novices at the same level of competence.

Managing self-motivation, and the working and learning environment are all-important in facilitating successful learning. High-achieving students maintain a balance between formal or required practice tasks and informal, creative, or motivating activities such as playing a favorite piece or improvising. The most creative student

composers set personal goals, negotiate the relationship between global and specific parameters, experiment and constructively criticize their work, and while demonstrating an optimistic outlook and internal locus of control, are more critical of their work than their peers (Priest, 2006).

To be successful, learning and creative problem solving must be thoughtfully approached and the strategies adopted tailored to the needs of the particular task. Learning may be exploratory, deliberate, playful, mental, physical, and expressive or performance oriented depending on the task and the stage reached in bringing it to fruition. Those wishing to become more effective learners and creators should:

1. develop metacognitive skills (identify the nature of the task and what will be most effective in tackling it, identify strengths and weaknesses, and monitor and evaluate progress);
2. identify clear goals and work toward them;
3. listen extensively to all kinds of music, comparing and critically evaluating it and its performance;
4. search out and act upon feedback from others;
5. think about the task when not actively engaged with it – engage in mental rehearsal;
6. invest time in analytic work;
7. plan;
8. develop strategies for managing motivation;
9. ensure that the learning or creative environment is conducive;
10. take advantage of opportunities to play, improvise, compose, and perform with others (this develops a wide range of skills); and
11. acknowledge the relationship between time engaged with music and achievement and set out to invest the time necessary.

Time invested in engagement with music and the quality of that engagement impact on the level of musical expertise attained. Learning how to learn and create leads to the time spent being more productive, challenging, and enjoyable.

Learning or Creating in a Musical Community

Participating in group musical activities (communities of practice), formal and informal, has many benefits. As members of musical groups, learners go through a progressive sequence of intentional stages through which they progress as auditors, competent listeners, inceptors, beginners, and competent players. Interacting musically with others develops a wide range of musical skills. Group music making or composition provides opportunities for collective involvement and decision making; support for the less confident; opportunities for playing and practicing together; exchange of ideas; opportunities for the less skilled to learn from the more experienced;

increased opportunities for experimentation; more choice of instrumentation; and shared opportunities for the rejecting and selecting of ideas. Other benefits can include increased confidence, the development of negotiation skills, the ability to defer to other people, respect for the work of others, the ability to give constructive criticism, and the ability to work freely within a disciplined environment. Informal learning based on combinations of trial and error, repetition, watching and taking advice from other players, reading, listening, and emulating develops metacognitive musical skills and can encompass organizational and practical skills, linguistic training, and the formation of personal identity.

Motivation

To attain high levels of skill in any aspect of music requires that a considerable amount of time be invested in active engagement with it. This requires motivation.

There are complex relationships between motivation, achievement, and time spent engaged with learning. While intrinsic motivation may be generated by interesting tasks in the short term, that interest must be internalized and become part of the individual's identity for motivation to be sustained over long periods of time. Internal motives include the desire for achievement, curiosity, and self-actualization, the need for personal fulfillment met by the emotion-inducing quality of music, satisfaction of a positive social response in performance settings, exploration of aggressive drives through the exploitation of the motor skills entailed in musical performance, and the satisfaction of some voyeuristic and exhibitionistic desires.

The particular musical activity engaged in is important for sustaining motivation. Choice of musical instrument should reflect interests and enthusiasms but it may be limited by convenience, availability, parents' views, the influence of the school or instrumental music provider, or friends. Gender may be a factor in that girls generally prefer small, high-pitched instruments, whereas boys choose large, low-pitched ones. Parents also tend to have stereotypical views of which instruments are appropriate. Those children violating these stereotypes may be subject to harassment.

To motivate learners, the early stages of active engagement with music need to be playful and intrinsically rewarding. The influence of early teachers, who are viewed as warm and sympathetic, seems to be particularly important in this respect. Relatively uncritical encouragement in the early stages of engagement with music encourages the development of a positive musical identity. Once this is established, later teachers provide high-status role models with whom the young musician can identify and emulate (Sosniak, 1985).

Sustained parental support is an important factor in the attainment of high levels of musical expertise. Parental enrichment of the musical environment from birth seems to stimulate the onset of early singing that in turn contributes to later success. (Howe *et al.*, 1995). Having musical parents or a musical home environment influences participation and success in music. Families can play a role in the identification of talent and parents of high achievers tend to have very high expectations and are demanding in those expectations. The child seems to internalize these and achievement becomes a need in itself. The parents of high achievers also tend to be more involved in initial practice, attend lessons with their children, and receive and act upon feedback from the teacher. Practical help in taking pupils to concerts and providing financial support is critical. However, there are exceptions where children have achieved professional status with little practical support from the family. Some highly effective individuals in music have histories marked by severe frustration, deprivation, and traumatic experiences. Personal determination enabled them to overcome these difficulties. The most highly accomplished young musicians appear to be self-motivated almost to the point of obsession, unable to separate their developing self-perception from that of being a musician. (Kemp, 1996) This enables them to overcome obstacles and bullying or loss of popularity as a result of their involvement. This high level of commitment may depend on the experience of intense, esthetic, emotional reactions to music, initially occurring in early childhood.

The development of a high level of commitment to music is also influenced by societal factors. Music is not valued equally in all cultures. In some places and at some times, the environment is very conducive to musical activity and there are considerable rewards for being musical. Educational institutions also vary in the level of support they give to musical activities. In some schools, music is highly valued and supported by senior managers and class teachers. An enthusiastic class teacher can modify pupils' attitudes toward music and counteract negative social influences.

In adolescence, the peer group is very powerful and can bring negative pressure to bear in relation to engagement with some types of music. To withstand this pressure, musical identities need to be well developed.

Those who drop out from specialist music tuition perceive that they are less musically able and receive less family encouragement. They tend to pursue other leisure activities as they view continuing to play as demanding too great a time cost for the relatively small rewards it offered. There are complex relationships between prior knowledge, motivation, effort, and perceived efficacy. When a child begins to learn an instrument, prior musical knowledge affects ease of learning and the time needed to achieve mastery of a task. While

undertaking additional practice may compensate for lack of prior knowledge, this has a time cost and requires perseverance. If a task proves challenging, the effort required to complete it may be perceived as too great and the individual may give up. Difficulties may be attributed to a lack of musical ability leading to a loss of self-esteem, loss of motivation, less practice, and a downward spiral leading to the termination of lessons.

Endnote

We are all born with the capability to make and appreciate music. The extent to which the individual develops musical skills depends on opportunity, the selection of an appropriate musical activity which s/he finds engaging, personal motivation and commitment, the adoption of effective learning and problem-solving strategies, and a supportive learning environment where music is valued.

See also: Culture in Motivation Research: A Challenging and Enriching Contribution; Metacognition.

Bibliography

- Fitts, P. M. and Posner, M. I. (1967). *Human Performance*. Belmont, CA: Brooks Cole.
- Gabrielsson, A. (2001). Emotions in strong experiences with music. In Juslin, P. N. and Sloboda, J. A. (eds.) *Music and Emotion: Theory and Research*, pp 431–452. Oxford: Oxford University Press.
- Howe, M. J. A., Davidson, J. W., Moore, D. M., and Sloboda, J. A. (1995). Are there early childhood signs of musical ability? *Psychology of Music* **23**, 162–176.
- Kemp, A. E. (1996). *The Musical Temperament: Psychology and Personality of Musicians*. Oxford: Oxford University Press.
- Lamont, A. and Cross, I. (1994). Children's cognitive representations of musical pitch. *Music Perception* **12**(1), 27–55.
- Priest, T. (2006). Self-evaluation, creativity and musical achievement. *Psychology of Music* **34**(1), 47–62.
- Schlaug, G. (2003). The brain of musicians. In Peretz, I. and Zatorre, R. (eds.) *The Cognitive Neuroscience of Music*, pp 366–381. Oxford: Oxford University Press.
- Sosniak, L. A. (1985). Learning to be a concert pianist. In Bloom, B. S. (ed.) *Developing Talent in Young People*, pp 19–67. New York: Ballantine.
- Trehub, S. E. and Trainor, L. J. (1990). Rules for listening in infancy. In Enns, J. (ed.) *The Development of Attention: Research and Theory*, pp 87–119. Amsterdam: Elsevier.

Further Reading

- Barry, N. and Hallam, S. (2002). Practising. In Parncutt, R. and McPherson, G. (eds.) *Science and Psychology of Music Performance*, pp 151–166. Oxford: Oxford University Press.
- Colwell, R. and Richardson, C. (eds.) (2002). *The New Handbook of Research on Music Teaching and Learning*. Oxford: Oxford University Press.
- Hallam, S. (1998). *Instrumental Teaching: A Practical Guide to Better Teaching and Learning*. Oxford: Heinemann.

Hallam, S. (2006). *Music Psychology in Education*. London: Institute of Education, University of London.

MacDonald, R., Hargreaves, D., and Miell, D. (2002). *Musical Identities*. Oxford: Oxford University Press.

McPherson, G. (ed.) (2006). *The Child as Musician: A Handbook of Musical Development*. Oxford: Oxford University Press.

Peretz, I. and Zatorre, R. (eds.) (2003). *The Cognitive Neuroscience of Music*. Oxford: Oxford University Press.

Reading Comprehension: Reading for Learning

C E Snow, Harvard Graduate School of Education, Cambridge, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Defining Reading Comprehension

The challenge of understanding reading comprehension derives, in part, from the difficulty of defining its borders. Comprehension was defined by the Research and Development (RAND) Reading Study Group (RRSG, 2002) as “the process of simultaneously constructing and extracting meaning through interaction and engagement with print.” This definition was intended to signal the importance of a number of key features of comprehension: the accurate decoding of print, a process of meaning construction through which inferences and information not available from the print are incorporated into the meaning representation, and active, motivated engagement from the reader. This definition works well for prototypical cases: the 10-year-old laughing while reading a joke book, the 15-year-old engrossed in a science fiction novel, and the 25-year-old being guided by a manual to install and run a new piece of software. The processes that occur during these prototypical comprehension events have been the subject of considerable research (see RRSG, 2002 for more detailed information about those processes), which has made clear that the success of any reading comprehension event is determined by variation on three dimensions: the text, the reader, and the task, all defined within a sociocultural context. The RRSG characterized successful comprehension as what occurs when the demands of the text, the challenges of the task, and the skills and proclivities of the reader are all well aligned, as exemplified by the prototypical cases listed above. Any pair of these dimensions can be the site of a mismatch that causes comprehension to fail and, as is described below, each introduces some ambiguity about where real reading comprehension begins and ends.

Texts

Consider a candidate text that might be found in a first grade reader:

Alex and Ali ran to the swings and jumped on.

What constitutes comprehension for this text? At a minimum, a mental representation of two individuals moving quickly toward and using some playground equipment should be conjured up, but is the inference that Alex and Ali are probably children part of the comprehension process or does that go beyond basic comprehension? Is it required that the comprehender assign genders to Alex

and Ali, or that gender assignment be postponed, recognizing that Alex could be short for either Alexandra or Alexander, that Ali could be a boy's name or a nickname for Alison? If Ali is provisionally classified as a boy, is it part of comprehension processing to infer that he comes from a Muslim family, or is that an inference that goes well beyond basic comprehension? If the reader has, for example, just arrived from China and has never encountered these first names before, has that reader fulfilled expectations with the inference that these are animate creatures – perhaps as likely to be cats as children? Must the reader infer that Alex and Ali actually started swinging, or does that go beyond comprehension into the realm of prediction? Does an inference that Alex and Ali were enjoying themselves belong to the realm of comprehending this sentence or comprehending the world? In other words, what is a sufficiently elaborated representation of this simple sentence to qualify as comprehension?

The dilemmas posed by considering different levels of processing of this brief text are, of course, greatly expanded if we consider the comprehension of longer and more complex texts, from paragraphs to newspaper reports or scientific articles to entire novels, let alone trying to establish what constitutes comprehension when reading an array of texts – reports of a political speech in right-wing versus left-wing newspapers, or scientific articles reporting conflicting results, or the entire oeuvre of a novelist – in conjunction with one another. At some point between the simple sentence above and the several volumes of *Remembrance of Things Past*, the definition of comprehension shape-shifts from a simple representation of an event to deep understanding of a worldview, but fixing the boundary between those activities is not easy.

Readers

Considering students at different points in development also dictates emphasis on different aspects and levels of comprehension, whether one is motivated to design instruction, select assessments, or investigate the underlying comprehension processes. For example, researchers and practitioners focused on reading to learn for students in secondary grades must take into account the overwhelmingly important contribution to successful comprehension of students' access to relevant background knowledge. Thus, in science, social studies, and math classes, there is often considerable emphasis on ensuring that students know something about a topic (using

lectures, videos, diagrams, hands-on demonstrations, or other nonliterate means) prior and as a support to their reading a text about that topic. On the other hand, researchers and practitioners more interested in early reading instruction and/or in remediation for struggling readers tend to emphasize issues related to reading and understanding the words in the text because that is where beginning readers encounter comprehension challenges, and it is often (though not always) the reason struggling readers do not comprehend well. In between these extremes of teaching beginning and struggling readers and teaching reading for learning, there is instructional emphasis on what might be thought of as simple comprehension – comprehension by students who have mastered word reading, reading texts which only make limited demands on background knowledge, but which do require (1) building and continually revising/expanding a text representation while reading, (2) making some inferences about connections among sentences and about connections to real world situations, and (3) perhaps some comprehension monitoring and comprehension repair mechanisms.

These differences related to developmental stage are also reflected in comprehension assessments, which for younger readers typically include items testing literal comprehension or basic inferencing, while items for older readers may require inferences that go farther beyond the text or draw more deeply on background knowledge (Snow, 2003). In other words, the definition of successful comprehension must be made conditional on at least the age and stage of development of the reader as well as the level and complexity of the text being read.

Task

A further complicating factor in defining successful reading comprehension has to do with the task being undertaken. There are important cultural, educational, and individual differences in the conceptualization of comprehension. In some literary and religious traditions, for example, literal memory for text is valued above interpretation of the text, whereas in others, attention to the actual words of the original text is much less important than coming to a justifiable interpretation of it, making connections to it, and even perhaps critiquing it.

Stark differences in task can be observed within cultures across disciplinary boundaries as well. For example, a science textbook is meant to be read for information, and comprehension can be said to occur when the reader expands and/or revises his/her understanding of some phenomenon by reading the information in the book; all too often, of course, the science textbook reader simply remembers the new information long enough to pass a test on it, without actually revising his/her enduring understanding. Therefore, the question that then arises is

whether this is a failure of comprehension or a failure of science learning.

In contrast, though successful comprehension of a novel read in a language arts or literature class does require learning the basics of characters, setting, and plot, just acquiring that information is not considered successful learning unless some appreciation is also engendered of the mood, the characters' and author's perspectives, the theme, the author's goal in writing the book, and other such ineffable features. One might well, in the course of reading some literary works, incidentally pick up information about scientific or historical or interpersonal topics treated in the book, and that would signal comprehension in one sense, but a literary reading would demand much more from the reader. Therefore, in literature classes, the question arises whether the dutiful student who can write an accurate plot summary of a novel, but fails to recognize, for example, that the narrator has taken an ironic stance or that the plot is a modern reenactment of the *Odyssey*, has failed at reading comprehension or at literary analysis.

Integrating Information about Reader, Text, and Task

Predicting comprehension success requires calculating information about the reader's stage of development, the complexity of the text being read, and the task being engaged in (see RRSg (2002) for further elaboration of each of these three aspects of any comprehension experience; **Figure 1**, reproduced from the RAND report, is a visual representation of this model). Successful comprehension occurs when these three dimensions are well

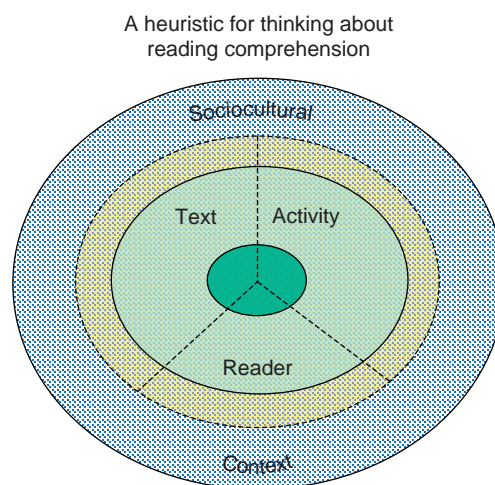


Figure 1 A heuristic for thinking about reading comprehensions devised by the Research and Development (RAND) Reading Study Group (2002) to depict the interaction of text, reader, and activity (or task) on reading comprehension.

aligned. For each of these dimensions of comprehension, though, there are simple cases and more marginal, gray areas where comprehension shades into learning or interpreting or functioning disciplinarily. The vast differences in what we would call successful comprehension across different levels of reader skill, text challenge, and task definition pose a challenge in summarizing what we know about reading comprehension, and in integrating or even providing a road map to the extensive research literature on comprehension development, assessment, instruction, and intervention.

A Taxonomy for Comprehension

For the purposes of this overview, we argue that identifying exactly where the boundary between reading comprehension and some other activity occurs is, to some extent, the individual's prerogative; even experts in the field achieve better agreement on identifying prototypical comprehension events than on placement of the boundaries. Furthermore, while the difficulty of deciding when a reading activity incorporates too many additional demands to be considered real comprehension may be obvious, there are also difficulties in deciding where comprehension begins at the bottom end, considering young readers and simple texts. What about the Jewish American or Muslim Turkish child who learns to read a sacred text, following the print faithfully while accurately pronouncing words that neither understands? It may seem obvious that this does not count as comprehension; however, what if that child has been told what the text is about, or has even been given a careful and complete translation of it? If the child is thinking of the translation while reading, is that comprehension? What if the child can understand a few of the words in the text, but has no understanding of the grammar of the written language? What if the child understands that verse, but none of the others in the book? When does reciting stop and real reading comprehension begin?

Reading comprehension might be thought of, then, as located on the radius of a set of concentric circles (see **Figure 2**). In the center circle are the basic reading processes that must be in place in order to access the text and form a mental representation of it: accurate word recognition, fluent access to word meaning, recognition of syntactic cues to sentence meaning, and short-term phonological memory. Variations of skill on these dimensions are clearly related to reading comprehension success – the reader who misidentifies words, who does not know the meaning of words in the text, who cannot parse the syntax of utterances, and who forgets the first sentence in a paragraph while reading the second will have difficulty comprehending (RRSG, 2002; Vellutino, 2003).

The second circle can be thought of as core comprehension processes – the ability to construct a mental representation of the ideas presented textually (Kintsch, 1998; Kintsch and Kintsch, 2005). Core comprehension requires text memory, making text-based inferences (e.g., tracing anaphors back to referents, keeping track of the order of events, and understanding implicit causal links), and making text-world links (e.g., bringing information about real dogs to bear in understanding what is strange and funny about a talking dog). Much early comprehension instruction focuses on helping learners activate relevant background knowledge before confronting text, on the theory that even children who have the required knowledge may not automatically access it while reading or integrate it with new information in the text. Another aspect of comprehension instruction for younger readers is a focus on self-monitoring, to ensure that the process of reading remains focused on building mental representations, and not just on reading the words.

The third circle comprises more elaborated comprehension processes, the processes involved in going beyond creating an unadorned text representation to a deeper understanding of the text. Many of the comprehension strategies that are recommended as part of comprehension instruction, for example, visualization, noting questions that arise while reading, and making text-to-text connections, are focused on these somewhat more elaborated comprehension processes. These processes also shade into ones that might be identified and taught as part of inquiry learning, such as figuring out how claims in one text relate to claims in another text, identifying the point of view a text presents, critiquing the argument in a text, and so on. In other words, rather than inquiry being a process applied to real-world phenomena, it is taken as a process to be applied to text itself. This is the theory underlying approaches to comprehension instruction such approaches to comprehension instruction as reciprocal teaching (Palincsar, 2003), questioning the author (Beck and McKeown, 2002), and reading apprenticeship (Shoenbach *et al.*, 1999).

An outer circle comprises highly elaborated comprehension processes that overlap with disciplinary studies or deep learning from text. Whereas ordinary readers might be expected to engage in moderately elaborated comprehension for purposes of understanding murder mysteries, psychological novels, columnists' political opinion pieces, or popular science articles, highly elaborated comprehension processes can only be expected of readers operating within domains where they have developed deep background knowledge and have had disciplinary training in how to read. These would encompass the processes involved, for example, in reading for purposes of literary criticism, historiography, constructing an intellectual history, or producing a parody.

The representation of these four kinds of reading in **Figure 2** as concentric circles with clear boundaries

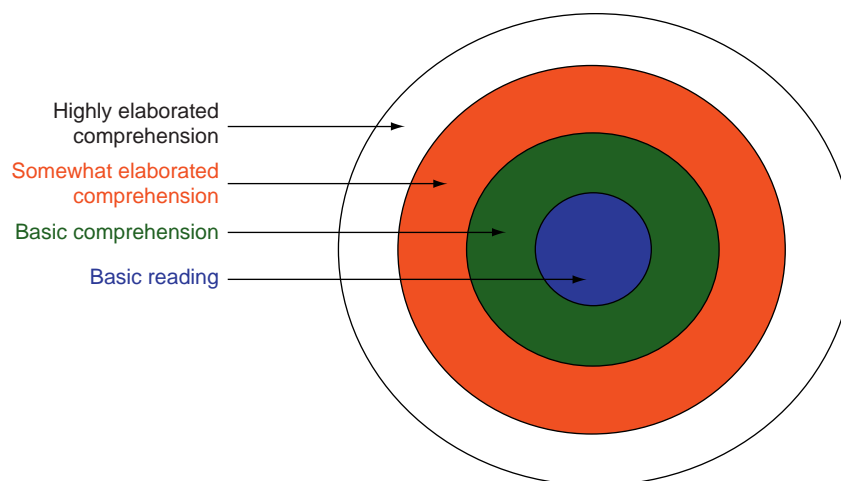


Figure 2 Basic reading processes, basic comprehension processes, and elaborated and highly elaborated comprehension processes represented as concentric circles, with reading comprehension located somewhere on a radius depending on the reader, text, and task.

between them should be viewed cautiously. First, there is no strong basis for placing a particular reading comprehension event on either side of the boundaries between central, elaborated, and highly elaborated comprehension processes. Second, this depiction is not meant to license an approach to reading instruction that starts in the middle and moves slowly outward; meaning construction, new learning, and interpretation should be part of the earliest literacy instruction, though these activities may be engaged in while reading texts aloud to children who are still mastering the code. Furthermore, the degree to which more sophisticated and elaborated comprehension might be expected of a lay literate versus a disciplinary literate depends, to a large extent, on the decisions a society makes about educational goals. Nonetheless, it may be useful in categorizing research, analyzing comprehension assessments, and understanding the challenges facing teachers of reading and of content areas to at least stipulate that reading comprehension is quite different when it occurs during code-focused reading as compared to reading for new learning and intellectual development.

Theories of Reading Comprehension

A few theories of reading comprehension have been particularly useful in guiding research and informing instruction. The simple view of reading (Gough and Tunmer, 1986) conceptualizes comprehension as the product of two capacities: the capacity to decode and the capacity to understand spoken language. The simple view claims, then, that comprehension is limited not only by speed and accuracy of word reading, but also by oral comprehension ability, and that if either of these abilities

is zero, then comprehension does not occur. Children following a normal developmental trajectory are subject to comprehension limitations stemming from constraints on word reading during the early years of schooling, and stemming from the limits on oral language skills thereafter. Under this view, it is clear that building oral language skills (vocabulary, comprehension of complex syntax, and comprehension of extended discourse forms) constitutes a key contribution to reading comprehension.

The simple view underemphasizes, though, the role of background knowledge and of motivation. The theory formulated by Kintsch introduces background knowledge by articulating how the textbase (the product of core comprehension processes) interacts with the mental model (the meaning representation constructed from the textbase and world knowledge; Kintsch, 1998; Kintsch and Kintsch, 2005). Kintsch (1998) also notes the importance of attending to the genre and the rules of reader–writer communication within the genre. Key in understanding the textbase and its links to the mental model, the genre, and the larger communicative act are various signals at the sentence level (e.g., after, same, and but) and the larger discourse level (e.g., headers and lists) of how the bits of information in the text are meant to be related to and integrated with one another (see also Graesser *et al.*, 2003; RRSg, 2002).

The role of motivation is emphasized in the work of Guthrie (2003), who points out that background knowledge is likely to be richer in areas of personal interest, and that readers are more likely to persist in wrestling with text if (1) they are interested in the topic and (2) they experience self-efficacy as readers. Reader self-efficacy grows with comprehension skill, which in turn supports reading engagement, which in turn further builds comprehension skills and background knowledge.

Instruction in Reading Comprehension

Instruction in reading comprehension is much less emphasized than instruction in basic reading skills or instruction focused on content areas without attention to the challenges of reading in these areas. Given the importance of background knowledge and vocabulary to successful comprehension, young children should have access to oral language-focused instruction, in which comprehension is modeled and vocabulary and background knowledge are taught by reading aloud from both fiction and nonfiction books.

The most frequent form of comprehension-focused instruction involves teaching comprehension strategies (National Reading Panel, 2000). While strategies such as monitoring one's own comprehension, stopping to note questions that one has, and visualizing can be supportive, it is important that instruction in these strategies focus on when to use them and why they can be helpful in creating meaning representations. A focus on content teaching creates a context for introducing comprehension strategies as targeted learning tools, as happens in Guthrie's program called Concept-oriented Reading Instruction (2002) and in Reading Apprenticeship (Shoenbach *et al.*, 1999), rather than teaching them as all-purpose comprehension aides.

Intervention with Struggling Comprehenders

Providing intervention to help struggling comprehenders before they fall far behind is a key responsibility in light of the overwhelming evidence that poor comprehension is associated with reduced opportunities to learn vocabulary and general knowledge (Stanovich, 1986), difficulties in learning across academic areas (RRSG, 2002), and ultimately, frustration with schooling and a higher likelihood of failure to graduate from high school or to achieve access to higher education. A challenge in providing comprehension intervention is that poor comprehension can be a product of a breakdown in any of a wide variety of reader skills (word reading accuracy, fluency, vocabulary, background knowledge, text memory, deployment of appropriate strategies, and engagement in reading), and effective intervention requires identifying the challenge and responding to it. For adolescent learners struggling with comprehension because of difficulties with word reading or fluency, it is often difficult to access instructional materials that offer minimal textual challenge, but are engaging and of appropriate cognitive level.

Deshler *et al.* (2007) provide an extensive review of interventions for struggling comprehenders, indexed by target of the intervention as well as developmental level

and type of learner (e.g., vocabulary focus for intermediate second-language learners). Unfortunately, very few of the programs they review, many of which have solid theoretical foundations, have been extensively evaluated or analyzed to determine under which circumstances and for which subgroups of learners they are most useful.

Summary

Reading comprehension is a complex topic. Predicting success in comprehension requires knowing about the reader, about the text being read, about the task being undertaken, and about the sociocultural context in which the reading is occurring. Since reading comprehension shades into learning, constructing a worldview, and discipline-specific literacy practices, it is difficult to establish firm boundaries around comprehension; nonetheless, it is clear that more attention to comprehension is needed across the grades. In preschool and primary grades, opportunities for building vocabulary and background knowledge and practicing oral comprehension should be provided while children are learning to decode. In later grades, students need explicit instruction in how texts are constructed and how language cues signal meaning at sentential and discourse levels, as well as practice and support in wrestling with content-rich texts for well-defined and engaging purposes.

See also: An Overview of Language and Literacy in Educational Settings; First Language Acquisition; Learning as Inquiry; Learning to Read.

Bibliography

- Beck, I. L. and McKeown, M. G. (2002). Questioning the author: Making sense of social studies. *Educational Leadership* 60(3), 44–47.
- Deshler, D. D., Palincsar, A. S., Biancarosa, G., and Nair, M. (2007). *Informed Choices for Struggling Adolescent Readers: A Research-Based Guide to Instructional Programs and Practice*. New York: International Reading Association.
- Gough, P. B. and Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education* 7, 6–10.
- Graesser, A. C., McNamara, D. S., and Louwerse, M. M. (2002). What do readers need to learn in order to process coherence relations in narrative and expository text? In Sweet, A. P. and Snow, C. E. (eds.) *Rethinking Reading Comprehension*, pp 82–98. New York: Guilford.
- Guthrie, J. (2003). Concept-oriented reading instruction: Practices of teaching reading for understanding. In Sweet, A. P. and Snow, C. E. (eds.) *Rethinking Reading Comprehension*, pp 115–140. New York: Guilford.
- Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. New York: Cambridge University Press.
- Kintsch, W. and Kintsch, E. (2005). Comprehension. In Paris, S. G. and Stahl, S. A. (eds.) *Children's Reading Comprehension and Assessment*, pp 71–92. Mahwah, NJ: Erlbaum.
- National Reading Panel (2000). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction. Reports of the*

- Subgroups*, NIH Publication No. 00-4754, Washington, DC: National Institute of Child Health and Human Development.
- Palincsar, A. M. (2003). Collaborative approaches to comprehension instruction. In Sweet, A. P. and Snow, C. E. (eds.) *Rethinking Reading Comprehension*, pp 99–114. New York: Guilford.
- RAND Reading Study Group (RRSG) (2002). Toward an R&D program in reading comprehension. Santa Monica, CA: RAND. http://www.rand.org/pubs/monograph_reports/2005/MR1465.pdf (accessed May 2009).
- Schoenbach, R., Greenleaf, C., Cziko, C., and Hurwitz, L. (1999). *Reading for Understanding: A Guide to Improving Reading in Middle and High School*. San Francisco, CA: Jossey-Bass.
- Snow, C. E. (2003). Assessment of reading comprehension: Researchers and practitioners helping themselves and each other. In Sweet, A. P. and Snow, C. (eds.) *Rethinking Reading Comprehension*, pp 192–206. New York: Guilford.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly* **21**(4), 360–406.
- Vellutino, F. R. (2003). Individual differences as sources of variability in reading comprehension. In Sweet, A. P. and Snow, C. E. (eds.) *Rethinking Reading Comprehension*, pp 51–81. New York: Guilford.
- Carnegie Corporation of New York, Washington, DC: Alliance for Excellent Education.
- Brown, A. L., Armbruster, B. B., and Baker, L. (1986). The role of metacognition in reading and studying. In Orasanu, J. (ed.) *Reading Comprehension: From Research to Practice*, pp 49–76. Hillsdale, NJ: Erlbaum.
- Clark, A. M., Anderson, R. C., Archodidou, A., et al. (2003). Collaborative reasoning: Expanding ways for children to talk and think in the classroom. *Educational Psychology Review* **15**, 181–198.
- Fry, E. (2002). Readability versus leveling. *Reading Teacher* **56**(3), 286–291.
- Moje, E. B., Dillon, D. R., and O'Brien, D. (2000). Reexamining roles of learner, text and context in secondary literacy. *Journal of Educational Research* **93**(3), 165–180.
- Perfetti, C. A., Landi, N., and Oakhill, J. (2005). The acquisition of reading comprehension skills. In Snowling, M. J. and Hulme, C. (eds.) *The Science of Reading: A Handbook*, pp 227–247. Malden, MA: Blackwell.
- Rasinski, T. V., Padak, M. D., McKeon, C. A., et al. (2005). Is reading fluency a key for successful high school reading? *Journal of Adolescent and Adult Literacy* **49**(1), 22–27.
- Sweet, A. P. and Snow, C. E. (eds.) (2003). *Rethinking Reading Comprehension*. New York: Guilford.

Further Reading

- Alexander, P. A. and Jetton, T. L. (2000). Learning from text: A multidimensional and developmental perspective. In Barr, R., Kamil, M., Mosenthal, P., and Pearson, P. D. (eds.) *Handbook of Reading Research*, vol. 3, pp 285–310. New York: Longman.
- Bailey, A. (ed.) (2007). *The Language Demands of School: Putting Academic Language to the Test*. New Haven, CT: Yale University Press.
- Beck, I. L. and McKeown, M. G. (2001). Text talk: Capturing the benefits of read aloud experiences for young children. *Reading Teacher* **55**(1), 10–20.
- Biancarosa, G. and Snow, C. (2004). Reading next: A vision for action and research in middle and high school literacy. *Report to the*

Relevant Websites

- <http://www.all4ed.org> – Alliance for Excellent Education.
- <http://www.carnegie.org> – Carnegie Corporation of New York.
- <http://www.ciera.org> – Center for Intervention in Early Reading Achievement.
- <http://www.excelgov.org> – Coalition for Evidence-Based Policy.
- <http://www.w-w-c.org> – Department of Education, What Works?
- <http://www.reading.org> – International Reading Association.
- <http://www.rand.org> – RAND Reports.
- <http://www.sedl.org> – Southwest Educational Development Laboratory (SEDL).

Second Language Learning

J S Arnfast and J N Jørgensen, University of Copenhagen, Copenhagen S, Denmark
A Holmen, Aarhus University, Copenhagen NV, Denmark

© 2010 Elsevier Ltd. All rights reserved.

Theories of Second Language Acquisition

Contrastive Linguistics

The first major paradigm of language-acquisition theory was the behaviorist–structuralist school which was informed by the contrastive hypothesis. This was built on the observation that learners with common mother tongues developed similar difficulties with specific foreign languages. The general principle of the hypothesis was that difficulties in second- or foreign-language learning were caused by structural differences between the mother tongue of the learner and the language to be acquired. The hypothesis had two degrees (Wardhaugh, 1970). The strong version claimed to be able to predict the difficulties of learners on the basis of a contrastive analysis, that is, a systematic comparison of the structures of the two languages involved. The original contrastive hypothesis was closely related to structuralist linguistics and its emphasis on *langue* (Lado, 1957). The view on learning was that of behaviorist psychology. Language was considered a set of habits, and second-language learning was a task which primarily involved a change of habits. The mother tongue of the learner would, in its capacity as a set of habits, interfere with the learner's new language (Weinreich, 1953). Contrastive analysis of entire languages were planned and carried out. Depending on the structural salience of the features involved, contrastivists also found that they were able to determine different degrees of difficulty. James (1980) suggests that contrastive analysis can be used by curriculum planners and teachers in their preparation, but not necessarily as a wholesale explanation of difficulties. Based on studies of learners' actual difficulties, however, a criticism of the contrastive hypothesis developed. It was found that difficulties predicted by contrastive analysis sometimes never appeared, and in other cases, difficulties arose which contrastive analysis did not explain. Furthermore, so-called errors came to be seen as necessary steps in the acquisition of a language (Corder, 1967). Many error analyses were carried out in order to shed light on real-life difficulties of learners. In some cases, the characteristics of the target language (L2) appeared to be more important than the mother tongue (Hyltenstam, 1978). The focus soon shifted to the study of the L2 performance of the learner as a language in its own right, a so-called interlanguage (*interlingua* or approximative system, Richards, 1974). An interlanguage has a grammar which follows universal principles just like any other grammar; it is just more variable.

This understanding of language differed from traditional structuralist linguistics. Interlanguage studies further led to theorizing over the patterns of acquisition of second-language learners. This theorizing was likewise far removed from the behaviorist view of learning, it concentrated on cognitive aspects of language acquisition.

A Cognitive Approach to Second Language Acquisition

The cognitive approach to language acquisition has its starting point in the human brain's capacity for processing and organizing information. The central notion that (second) language learning is skill learning opens the field to psychological accounts of the processes leading to language acquisition. The first major contributions (e.g., McLaughlin, 1987) to the upcoming field were to a certain extent based on the L1 (first language)-acquisition studies of Slobin (1985). Slobin argued that the mind is programmed to process information which is perceptually salient, and to organize this information according to specific operating principles eventually leading to language learning. This explains cross-linguistic similarities in first-language acquisition. In later studies Slobin (1996) points to a close connection between language typology and language processing. A process similar to Slobin's earlier studies was believed to work for second-language learning (Andersen, 1990), and different models were proposed to account for the information processing necessary for the automatization of second-language learners' linguistic knowledge and fluent speech. Both McLaughlin's (1987) information-processing model and Anderson's (1985) adaptive control of thought are related to this line of research. Like L1-acquisition, second-language acquisition (SLA) is considered to follow a concise developmental order by which certain parts or features of the language are acquired before others (e.g., Givón, 1985). These thoughts are further elaborated in Pienemann's (1998) processability theory, which claims that language acquisition follows similar patterns even across language families. In cognitively oriented language-acquisition studies, the role of memory is taken into consideration in more detail. Such studies also drew on cognitive psychology in order to investigate the limits of perception, that is, how many different items the brain is capable of processing. Short-term memory is able to process only a limited amount of new information at a time, and by ways we can still only hypothesize about, information is

stored in the long-term memory after being processed repeatedly and thereby automatized. However, when information is stored in long-term memory, it is considered difficult to retrieve it for further processing, for instance, if the structure of the linguistic information is incomplete (i.e., incomplete morphology, syntax, or pronunciation). Krashen (1985) attempted to formulate a comprehensive theory for SLA, consisting of five major hypotheses. His comprehensive theory has been criticized for being difficult to test empirically. Nevertheless, the discussions it has raised have been fruitful for the further development of the field. One of the outcomes of Krashen's comprehensive theory was the extension of the input hypothesis proposed by Long (1985), in the field generally known as the interaction hypothesis. Through studies from the 1980s onward, Long pointed to the interesting differences in conversations carried out among native speakers compared to conversations involving both a native speaker and a learner. Conversations and communicative difficulties and how they were managed were studied through conversation analysis. They showed a significant use of conversational strategies (e.g., repetitions, repairs, requests for clarification, and comprehension checks) by both parties to ensure that the meaning and content of the conversation were not lost due to misunderstandings. During language acquisition, through negotiation of meaning, the learner focuses on content as well as form (Long and Robinson, 1998). The idea that language is learned through interaction, and thus tied to a social practice, is part of a wider pedagogical movement which shifts the focus from teaching to learning. The learner is considered an active part of a learning process. Schmidt and Frota (1986) point to the importance of consciousness and noticing as central phenomena in the language-acquisition process. Noticing specific aspects of language is closely linked to cognitive development as well. Through the 1980s, this interactionist approach was further developed. The notion of communicative strategies which accounted for the psycholinguistic aspects of interaction from comprehension to production was widely discussed, e.g., Kellerman and Bialystok (1997).

Motivation

Two major paradigms dominated the studies of motivation in language learning from the beginning of the 1970s. One was the view on motivation considered in terms of instrumentality and integrativeness (Gardner and Lambert, 1972). According to this view the motivational drive to learn a second language is the prospect of achieving specific goals (e.g., job or education) by learning the language (instrumental motivation), or the drive stems from a desire to adapt the culture and habits associated with the language (integrative motivation). The other paradigm distinguished between intrinsic versus extrinsic motivation (Deci and Ryan, 1985). This approach is more

sociopsychological. Motivation is considered the outcome of individual regulation of behavior. Extrinsic motivation is connected with the demands and expectations formulated by persons (e.g., parents or teachers) in the individual's surroundings under the promise of reward or the threat of punishment. As the individual becomes more and more socialized into a certain behavior, the motivation to fulfill specific expectations would come from within (intrinsic motivation). In both paradigms, the relation between the motivation types is complex, as one type may be dominant in some situations, but not in others.

Dörnyei and Ottó (1998) express the idea that motivation could be subject to change during the learning process. They introduce time as a factor in their model. The learner regularly adjusts her or his goals and strategies according to an ongoing evaluation. Noels *et al.* (2000) point to a strong correlation between instrumental motivation and external regulation (parallel to the extrinsic-intrinsic motivation paradigm). Travels, friendship, and knowledge are all involved, and important. Norton (2000) points to the complex relationship between identity, language learning, and setting, including socioeducational and socioeconomic factors as relevant to the development of the individual's motivation to learn a new language. Norton's introduction of the concept of social investment reflects the complex conditions that govern motivation.

Social Theory in SLA

With the sociolinguistic turn in SLA research, focus shifted from seeing language learning as the gradual development of a language system in individuals to a perspective on language learners as language users in real-life environments. Dominant cognitive theories were replaced by or supplemented with sociocultural theories focusing on actual language use. One characteristic aspect of learners' second-language use is the high level of variability in their interlanguage at any given time, that is, their use of different versions of the same language construction (e.g., marker of past tense in verbs). In the 1980s and early 1990s, results from a number of descriptive case studies on variability were analyzed in the SLA literature to discuss the theoretical concept of acquisition. Among others, Tarone (1988) and Ellis (1994) discussed how accurate the use of a linguistic form must be in order to say that the learner has acquired the basic rule. Another consequence of the rise of sociocultural theories in SLA research was a focus on the model of communication evoked in different approaches. Whereas the majority of cognitive studies as well as early sociolinguistic studies were based on a transmission model of communication, sociocultural studies assumed an interactionist perspective. Firth and Wagner (1997) argue that a dialogic view on language is a fundamental concept in modern SLA research, thus criticizing the traditional view on language as monologic in

nature. A third and partly related aspect of sociocultural theories is the focus on the social environment of language learning and on the interplay between language learners and their community. Based on Hymes (1972), an ethnographic approach to language learning implies seeing learning as socially and collaboratively constructed in interaction in different contexts or speech events. Different speech events provide learners with different opportunities for interaction, affordances (van Lier, 2000), as they draw on different communicative patterns and are governed by social hierarchies and power relations. In instructed second-language learning, there is typically an orientation toward target norms in the social context and in the dialog between student and teacher. However, recent studies show that other less-institutionalized kinds of interaction also develop as participants negotiate their social identity through individual style or group patterns (e.g., through new rituals, Rampton, 1999).

This has brought about discussions of the role of normativity in second-language learning, similar to the debate on socialization in child-language studies (e.g., Ochs, 1996). Insofar as language learning consists of picking up conversational routines or adjusting to specific norms of language use, both learner and child may be seen as apprentices moving from the periphery toward the center of a community through interaction with experts (Lave and Wenger, 1991). Although normativity is an important characteristic of instructed second-language learning, views on language learning as a uniform process of approaching predefined target norms have come under attack from poststructuralists who see learning as “a non-linear and relational human activity, co-constructed between humans and their environment, contingent upon their position in space and history, and a site of struggle for the control of social power and cultural memory” (Kramsch, 2002: 5). According to this sociocultural or ecological position, the sociolinguistics of second-language learning is not restricted to background variables which explain the speed and rate of success in the process of language learning. It is placed at the core of theories on the dynamics of language learning. Language is seen as part of the complex and dynamic processes of social membership, culture, and identity. Language learning is not only a matter of individual capability, but also of opportunities for interaction at the microsocial level and of power relations and linguistic norms at the macrosocial level. The role of social interaction is to provide supporting structures or scaffolding for individual competencies and thus to mediate language learning through communicative patterns and other semiotic tools (van Lier, 2000; Lantolff, 2000). At the macrosocial level, language proficiency is developed in social settings shaped by social structures and orders of discourse. Language development interacts with identity issues, such as the manifestation of agency in a negotiation of competing

subject positions in conflicting discourse communities (Canagarajah, 2004). Due to the increasing complexity of modern multilingual and multicultural societies, both learners and native speakers belong to several speech communities with a range of membership roles. Consequently, the idea of the native speaker is criticized for being a political rather than a theoretical concept. Kramsch (1998) replaces it with the concept of intercultural speaker and Dabelsteen and Jørgensen (2004) with the concept of the languager.

Interlanguage

Selinker (1972) introduced the term interlanguage to refer to the language produced by a learner, seen as a unique linguistic system different from the learner's L1 and from the target language, but using elements from both. The term describes learner language as systematic, and yet as dynamic and transitional. The gradually growing complexity of learner language is seen as a continuum of interlocking systems.

Since the 1980s, the term interlanguage has come to be used with different meanings: On the one hand, it refers to the mental grammar which the learner constructs while developing a second or foreign language. The learner's competence at any given time is hypothesized to be the guiding principle behind instances of language use. On the other hand, the term refers to the learner's actual language use in a social situation. The understanding is that learner language is a systematic variety of language (Færch *et al.*, 1984). According to both uses, deviations in learner language use are not seen as imperfect learning, but rather as signs of learning through hypothesis testing or as linguistic expressions of social identity.

Acquisition and Variation

SLA leads to variation in the L2 production of learners. Interlanguage is a highly variable phenomenon which changes with a range of factors, including oscillating precision, in reaching the target-language norms. In addition to this, SLA leads to variation which is sociolinguistically related in the way L1 variation is. Around 2000 minority groups of young second-language speakers of majority languages in the European cities developed styles which were marked as minority-related, but nevertheless also used by young majority members (Kotsinas, 2000). It remains to be seen whether these phenomena develop into ethnolects of the majority languages, or perhaps socially stigmatized urban varieties.

Variation in the production of learners may also involve loans and code switching. Contrary to conventional wisdom in traditional textbooks on language learning and language teaching, switching may be productive for the understanding between learners, or between learners and

native speakers (Arnfast and Jørgensen, 2003). The learner may achieve on several levels by code switching away from the L2 in an otherwise L2 conversation. Switching can be a strategy to maintain fluency, and therefore obtain at the level of social relations. It may also represent a learning strategy to expand the learner's knowledge of L2 – or the culture in which the L2 is being learned. In learner groups, perhaps, particularly school classes, the L2 may come to represent specific values (or stereotypes) which can be exploited in the everyday interaction among the learners (Rampton, 1999) in code switches into the L2 during otherwise L1 interaction. An intricate and complex pattern of code choice by young Londoners is described by Rampton (1995) as crossing which involves both English, Punjabi, Creole English, and other varieties. Crossing is characterized by the fact that the speakers do not necessarily know very much of the languages from which they borrow in their interactions. Similarly, Turkish-Danish students switch between a range of different languages, several of which have been taught to them, in what has been termed languaging (Dabelsteen and Jørgensen, 2004). Languaging is the use of linguistic items and features – regardless of where they belong – by human beings in order to achieve communicative goals.

Code switching and crossing are (as yet) controversial phenomena which are often discouraged by language instructors. There could be a future development in language-acquisition research which attempts to demonstrate to what extent learners can achieve by employing all the linguistic skills they have.

Age and SLA

The relationship between child SLA and adult SLA is a crucial problem in acquisition theory and empirical studies. Behaviorism formulated the critical-age hypothesis which states that language learning changes profoundly with the lateralization of the brain in adolescence, and the individual's ability to acquire new language deteriorates, and the acquisition of the mother tongue must be well under way before the age of 13, or the acquisition will never be complete (Lenneberg, 1967). This has also been applied to acquisition of a second language. The theory has been criticized, for one, because experiments have shown that adults acquire some aspects of language faster than children. On the other hand, the end result apparently is more native-like, the earlier a continuous acquisition has begun. A crucial review article is by Long (1990). Long concludes that the concept of a critical period is not irrelevant. However, it is not one specific period, but a continuum of maturation which affect different aspects of language at different times. He relates the maturational constraints to a physiological process, myelination, an

ongoing process which renders the nervous system less flexible, and which runs parallel with the automatization of linguistic features. The effect of age on language learning is a richly discussed issue (e.g., Singleton and Lengyel, 1995). Evidence for the critical-age hypothesis is found in studies which conclude that adults fail to acquire second-language features which are different from their L1 (e.g., Schachter, 1996). Evidence against the hypothesis is presented and reviewed by Bialystok (1997). Abrahamsson and Hyltenstam (2004) find evidence that neither young learners nor adult learners achieve complete native-like command of a second language, and like Long, they cite physiological reasons. Burgo (2006) reaches a similar conclusion. A different perspective of age-related differences has its focus on the identity work performed by learners when investing personal resources in acquiring a second language. The reactions of particularly adult learners range from rejection of, and isolation from, the second language and the cultures associated with it to completely embracing the language and cultures (Norton, 2000). Identities ascribed to learners, especially minority learners, by society at large may also affect the outcome of the learning process (McKay and Wong, 1996). Expectations as well as identities affect young learners and adult learners differently, and may affect motivations differently.

Second-Language Learning and Foreign-Language Learning

The distinction between foreign- and second-language learning is traditionally considered to be a matter of learning environment and the possibility to encounter, outside a formal learning setting, the language being learnt. The distinction between the two becomes clearer, when the who, how, what, and why are compared.

Foreign-language learning is generally directed to students, usually children or adolescents in primary and secondary school, who are taught the primary *lingua franca* of their country as a part of their general education. At a later stage within the educational system, foreign-language learning is either a goal in itself (e.g., Scandinavian philology in Asia) or a means to qualify other educational goals (e.g., studying German in order to read philosophers in the original language or to specialize in business school). In most countries, the number and selection of foreign languages offered in the educational system is determined by historical, political, cultural, and sociolinguistic factors. Outside the educational system, foreign-language learning is most often associated with adult education on a leisure-time basis, and rarely with special instruction for instrumental reasons. The second-language learner, on the other hand, is typically considered to be an immigrant (or in some

cases the children of immigrants), learning the language that plays an institutional and social role in the community (Ellis, 1994). Learning a second language may take place in formal settings (e.g., language courses) in the target country, or in the informal settings of the workplace, contact with authorities or social institutions, or other specific communities of practice. In this way, second-language learning takes place in a much more complex setting and under much more complex conditions than foreign-language learning. Since the 1990s, the distinction between formal and informal second-language learning tends to be regarded as a question of sociolinguistic setting. In this view, it is important to consider the interaction between the two settings, with an overlap between the two.

With the ongoing globalization of communication, it is no longer possible to regard foreign-language learning as an activity isolated from the community where the language in question is spoken as the primary means of communication. School classes communicate with other classes in twin towns, students seek out communities of interest on the Internet, and possibilities of traveling are expanding, just to mention a few. This promotes intrinsic motivation to improve specific linguistic and communicative skills in order to participate in social activity – real or virtual.

Second-Language Learning for School

In schools, language is not only taught, but also used in specific ways to transmit and reproduce relevant knowledge in relevant ways. This develops the specialized academic registers of school. At the same time, language is used to socialize children into (what counts as) competent citizens. Thus, language is both a means and an end to school activities. Language is also an instrument which links children's everyday experiences and cultural and family background with relevant school categories and learning practices, through recontextualization of their knowledge (Bernstein, 1990). Schools make use of a communicative code which draws on the social and thematic priorities that are given in the local context, and which overlaps only little with the functions of everyday language. For younger children, schools tend to link their language use with bodily activities, visual and sensory support, etc. to provide a physical basis for new concepts. For older children, the development of new conceptual categories is predominantly based on linguistic cues and explanations. Other kinds of problem solving, for example, math activities, also become gradually more dependent on the child's mastery of the relevant semiotic systems. The role of language is crucial in understanding why schooling in general is intellectually and socially difficult for children who do not master the communicative code of the school, and why the so-called fourth-year

slump (e.g., Thomas and Collier, 2001) marks the onset of a growing gap between high- and low-performing students. As a consequence, language plays an important role in counteracting children's problems with school.

For minority students, who attend school in their second language, the situation is often particularly difficult. It takes several years to achieve high proficiency in a new language across several domains (Cummins, 2001), and especially to develop the academic language suited for educational purposes. The learning situation of the minority children is often hampered by lack of response to their special linguistic needs. Many school systems are misled by a monolingual, common-sense view on language learning, which claims that it is a limited process shared by all learners regardless of their age, social, or linguistic background. According to this view, learning will eventually lead to the same outcome, whether it takes place before or after school starts. As a consequence, there is a political pressure toward preschool language learning in several industrialized countries. The logic behind this is the idea that an earlier introduction of the majority language leads to better results. This idea has been termed the fallacy of early start by Skutnabb-Kangas (2000). The idea draws on a simplistic, quantitative view on language learning. It is in line with another common fallacy – that so-called natural acquisition of language automatically develops into school language. According to this, children pick up everyday language through interaction in the playground and gradually transform this into the academic language of schools. Crawford (2000) among others has cautioned against the so-called sink-or-swim-method in which minority children are left to cope on their own with the intellectually and linguistically demanding learning of curriculum without instructional support. In this situation, efficient language learning may be hampered by gate keeping and silencing (Santa Ana, 2004). This may lead to social marginalization and – at a very practical level – to reduced opportunity of using the target language in situations that could have promoted learning.

The quantitative perspective is challenged by a view on language proficiency as a complex and dynamic combination of skills, knowledge, and reflexivity, relevant for both majority and minority children. In order to change the situation of educational underachievement in general, the role and norms of language must be made explicit in the classroom. Discourses must be established which may form the bridge between students' everyday language and prior knowledge on the one hand and their development of the academic school registers on the other (Gibbons, 2006). For minority children, this must include strategies to incorporate the actual multilingualism of the students' environment as part of their learning potential (Thomas and Collier, 2001).

See also: An Overview of Language and Literacy in Educational Settings; Bilingualism and Learning; First Language Acquisition; Foreign Language Learning.

Bibliography

- Abrahamsson, N. and Hyltenstam, K. (2004). Mognadsbegränsningar och den kritiska perioden för andraspråksinläring. In Hyltenstam, K. and Lindberg, I. (eds.) *Svenska som andraspråk*, pp 221–258. Lund: Studentlitteratur.
- Andersen, R. (1990). Models, processes, principles and strategies: Second language acquisition inside and outside the classroom. In Van Patten, B. and Lee, J. (eds.) *Second Language Acquisition – Foreign Language Learning*, pp 45–68. Clevedon, UK: Multilingual Matters.
- Anderson, J. (1985). *Cognitive Psychology and its Implications*, 2nd edn. New York: Freeman.
- Arnfast, J. S. and Jørgensen, J. N. (2003). Code-switching as a communication, learning, and social negotiation strategy in first-year learners of Danish. *International Journal of Applied Linguistics* 13(1), 23–53.
- Bernstein, B. (1990). *The Structuring of Pedagogic Discourse*. London: Routledge.
- Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second Language Research* 13, 116–137.
- Burgo, C. (2006). Maturation constraints in SLA. LL Journal 1(1), 12–25, <http://ojs.gc.cuny.edu/index.php/lljournal/article/viewFile/burgo/64> (accessed August 2009).
- Canagarajah, S. (2004). Multilingual writers and the struggle for voice in academic discourse. In Pavlenko, A. and Blackledge, A. (eds.) *Negotiation of Identities in Multilingual Contexts*, pp 266–289. Clevedon, UK: Multilingual Matters.
- Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics* 4, 161–170.
- Crawford, J. (2000). *At War with Diversity. US Language Policy in an Age of Anxiety*. Clevedon, Avon: Multilingual Matters.
- Cummins, J. (2001). *Language, Power and Pedagogy. Bilingual Children in the Crossfire*. Clevedon, UK: Multilingual Matters.
- Dabelsteen, C. and Jørgensen, J. N. (eds.) (2004). *Language and Language Practices. Copenhagen Studies in Bilingualism* 36. Copenhagen: University of Copenhagen.
- Deci, E. and Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Press.
- Dörnyei, Z. and Ottó, I. (1998). Motivation in action: A process model of L2 motivation. *Working Papers in Applied Linguistics* (Thames Valley University) 4, 43–69.
- Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Firth, A. and Wagner, J. (1997). On discourse, communication and (some) fundamental concepts in SLA research. *Modern Language Journal* 81(3), 286–300.
- Færch, C., Haastrup, K., and Phillipson, R. (1984). *Learner Language and Language Learning*. Copenhagen: Gyldendal.
- Gardner, R. and Lambert, W. (1972). *Attitudes and Motivation in Second Language Learning*. Rowley, MA: Newbury House.
- Gibbons, P. (2006). *Bridging Discourses in the ESL Classroom. Students, Teachers and Researchers*. London: Continuum.
- Givón, T. (1985). Function, structure and language acquisition. In Slobin, D. (ed.) *The Crosslinguistic Study of Language Acquisition*, 2, pp 1005–1027. New York: Erlbaum.
- Hyltenstam, K. (1978). *Progress in Immigrant Swedish Syntax. A Variability Analysis*. Lund: University of Lund.
- Hymes, D. (1972). Models of the interaction of language and social life. In Gumperz, J. and Hymes, D. (eds.) *Directions in Sociolinguistics*, pp 35–71. New York: Holt, Rinehart and Winston.
- James, C. (1980). *Contrastive Analysis*. London: Longman.
- Kellerman, E. and Bialystok, E. (1997). On psychological plausibility in the study of communication strategies. In Kasper, G. and Kellerman, E. (eds.) *Communication Strategies: Psycholinguistic and Sociolinguistic Perspectives*, pp 31–48. London: Longman.
- Kotsinas, U.-B. (2000). *Kontakt, variation och förändring – studier i Stockholmspråk*. Stockholm: Almqvist and Wiksell.
- Kramsch, C. (1998). The privilege of the intercultural speaker. In Byram, M. and Fleming (eds.) *Language Learning in Intercultural Perspective: Approaches through Drama and Ethnography*, pp 16–33. Cambridge: Cambridge University Press.
- Kramsch, C. (ed.) (2002). How can we tell the dancer from the dance? In *Language Acquisition and Language Socialization*, pp 1–30. London: Continuum.
- Krashen, S. (1985). *The Input Hypothesis: Issues and Implications*. Harlow: Longman.
- Lado, R. (1957). *Linguistics across Cultures*. Ann Arbor, MI: University of Michigan Press.
- Lantoloff, J. P. (ed.) (2000). *Sociocultural Theory and Second Language Learning*. Oxford: Oxford University Press.
- Lave, J. and Wenger, E. (1991). *Situated Learning. Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: Wiley.
- Long, M. H. (1985). Input and second language acquisition theory. In Gass, S. and Madden, C. (eds.) *Input and Second Language Acquisition*, pp 377–393. Rowley, MA: Newbury House.
- Long, M. (1990). Maturation constraints on language development. *Studies in Second Language Acquisition* 12(3), 251–285.
- Long, M. H. and Robinson, P. (1998). Focus on form: Theory, research and practice. In Doughty, C. J. and Williams, J. (eds.) *Focus on form in Second Language Acquisition*, pp 15–41. Cambridge: Cambridge University Press.
- McKay, S. L. and Wong, S.-L. C. (1996). Multiple discourses, multiple identities: Investment and agency in second-language learning among Chinese adolescent immigrant students. *Harvard Educational Review* 66(3), 577–608.
- McLaughlin, B. (1987). *Theories of Second Language Learning*. London: Edward Arnold.
- Noels, K. A., Pelletier, L. G., Clément, R., and Vallerand, R. J. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning* 50, 199–218.
- Norton, B. (2000). *Identity and Language Learning. Gender, Ethnicity and Educational Change*. London: Longman.
- Ochs, E. (1996). Linguistic resources for socializing humanity. In Gumperz, J. and Levinson, S. C. (eds.) *Rethinking Linguistic Relativity*, pp 407–437. Cambridge: Cambridge University Press.
- Pienemann, M. (1998). *Language Processing and Second Language Development: Processability Theory*. Amsterdam: John Benjamins.
- Rampton, B. (1995). *Crossing. Language and Ethnicity among Adolescents*. London: Longman.
- Rampton, B. (1999). Deutsch in inner London and the animation of an instructed foreign language. *Journal of Sociolinguistics* 3(4), 480–504.
- Richards, J. C. (ed.) (1974). *Error Analysis. Perspectives on Second Language Acquisition*. London: Longman.
- Santa Ana, O. (ed.) (2004). *Tongue-Tied. The Lives of Multilingual Children in Public Education*. Lanham: Rowman and Littlefield.
- Schachter, J. (1996). Maturation and the issue of universal grammar in second language acquisition. In Ritchie, W. and Bhatia, T. (eds.) *Handbook of Second Language Acquisition*, pp 159–193. San Diego, CA: Academic Press.
- Schmidt, R. and Frota, S. (1986). Developing basic conversational ability in a second language: A case-study of an adult learner. In Day, R. (ed.) *Talking to Learn: Conversation in Second Language Acquisition*, pp 237–326. Rowley, MA: Newbury House.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics* 10, 209–231.
- Singleton, D. and Lengyel, Z. (eds.) (1995). *The Age Factor in Second Language Acquisition*. Clevedon, Avon: Multilingual Matters.
- Skutnabb-Kangas, T. (2000). *Linguistic Genocide in Education – or Worldwide Diversity and Human Rights*. Mahwah, NJ: Erlbaum.
- Slobin, D. (ed.) (1985). *The Crosslinguistic Study of Language Acquisition*, vol. 1–2. Hillsdale, NJ: Erlbaum.
- Slobin, D. I. (1996). From 'thought and language' to 'thinking for speaking'. In Gumperz, J. and Levinson, S. C. (eds.) *Rethinking*

- Linguistic Relativity*, pp 70–96. Cambridge: Cambridge University Press.
- Tarone, E. (1988). *Variation in Interlanguage*. London: Edward Arnold.
- Thomas, W. and Collier, V. (2001). National study of school effectiveness for language minority students long-term academic achievement. http://www.crede.org/research/llaa/1.1_final.html (accessed August 2009).
- Van Lier, L. (2000). From input to affordance: Social-interactive learning from an ecological perspective. In Lantolff, J. (ed.) *Socio-Cultural Theory and Second Language Learning*, pp 245–259. Oxford: Oxford University Press.
- Vygotsky, L. (1978). *Mind in Society. The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly* 4(2), 123–130.
- Weinreich, U. (1953). *Languages in Contact. Findings and Problems*. The Hague: Mouton.

Further Reading

- Byram, M. (1997). *Teaching and Assessing Intercultural Communicative Competence*. Clevedon, UK: Multilingual Matters.
- Dörnyei, Z., Csizér, K., and Németh, N. (2006). *Motivation, Language Attitudes, and Globalisation: A Hungarian Perspective*. Clevedon, UK: Multilingual Matters.

Writing: Advanced

R Andrews, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The history of teaching-of-writing approaches at advanced levels over the last 50 years or so can be characterized into two broad phases: the pre-computer phase, and the phase informed by reciprocal co-evolution with information and communication technologies.

The 1950s and 1960s saw an approach to writing which set high store by literary quality within a limited range of genres derived from nineteenth-century rhetorical categories. There was little emphasis on the processes of writing; more on the finished product and on a distanced, academic command of language types. Exercises deriving from the Renaissance and Elizabethan grammar school practice of *progymnasmata* (rhetorical models and imitation) were designed to build up competence. This period marked the high point in the teaching of formal grammar in the hope of improving writing quality and accuracy. It is against this background that much of the development in approaches to the teaching of writing at advanced levels can be gauged.

This article divides the main phases in the teaching and learning of writing at advanced levels into the decades from the 1960s to the 1980s; and from the 1980s to the present. Between these two phases, it looks at a number of different transitions that are critical to writing pedagogy.

1960s to 1980s

From the 1960s to the early 1980s there was more emphasis on expressiveness and the emergence of a personal voice in writing than in the previous decades. The foregrounding of the imaginative and creative (in a literary sense) continued, with a move away from tried and tested school genres toward experimentation. There was a celebration of self within societal contexts, best characterized by *Growth through English* (Dixon, 1967), the book emerging from the Dartmouth seminar of the International Federation for the Teaching of English in 1966. Two other key works in the late 1960s were Moffett's (1968/1987) *Teaching the Universe of Discourse*, and *Language, the Learner and the School* by Barnes *et al.* (1969/1989). The former charted the range of discourses that a student could expect to cover in formal education, with a clear indication that narrative modes and genres must come first, followed by expository and argumentative types of writing. This work,

therefore, sat very much in the rhetorical tradition, but marked a step forward from the conventional theorizing and practice of the 1950s, where theory was ignored or taken for granted and practice fell into well-trodden ways. As Moffett states (1987, vii), in reflecting on the first edition of his book "I wanted to recast into the psychological terms of human growth those familiar but opaque academic elements such as rhetoric, logic, grammar, and literary technique. . ." Perhaps the most salient aspects of Moffett's hypothesis is that language development takes a course from drama through narrative and exposition to argumentation. Although such a series – with the famous sentence that "whereas adults differentiate their thought into specialized kinds of discourse such as narrative, generalization and theory, children must for a long time make narrative do for all" (my italics) (1987: 49) – seems to suggest a Piagetian fixed sequence, Moffett suggests that the strands of discourse run through young people's development concurrently. The basis in dialog and drama still feels contemporary in the light of Vygotsky's theory of cognitive development and, for example, Alexander's (2006) notions of dialogic teaching.

Language, the Learner and the School was more seminal than Moffett's work, in that it took a fresh empirical look at practice in classrooms and found it wanting in a number of respects. First, speech in the classroom was seen to be teacher dominated, with little space for pupils or students to try out ideas or expound at length. Second, although there was a higher proportion of open questions in English as a secondary school subject than in other subjects, there were fewer questions that required reasoning than in mathematics or science lessons. Third, questions asked by teachers tended to seek factual rather than reasoned answers, thus downplaying the dialogic, negotiated potential of the curriculum for learning and knowledge building. The investigations recorded in the book suggest that a closer relationship between active pupils' talk and their writing would enhance performance in both modes of communication; nevertheless, there was not only a gap between teacher and pupil discourses, but also one between pupil talk and pupil writing. However, Barnes (1989: 72) notes that "though this paper focuses on spoken language, much of what I have said applies equally to written language." In many ways, the connection between talk and writing was made, bringing an expressive, thinking-through-writing, dialogic (and potentially dialectical) character to learning to write at advanced level. Such a connection is also made

in Britton (1967) and in the work of Dixon (op. cit.). The complex interdependency of speech and writing is also reflected, for example, in the work on Ong (1982/1988).

The period also marked the rise of narrative as a primary act of mind (Hardy, 1977) or even human paradigm (Fisher, 1989). Rosen (1985) continued his interest in narrative speech and writing with publications in the 1980s celebrating the power and ubiquity of narrative, not just for expressive writing in autobiography and storytelling, but also in transactional, nonfictional kinds of writing. The interest in the oddly negative nonfiction category was captured in Pavel's (1986) *Fictional Worlds*, which brilliantly charted the position of fiction in relation to other forms of writing, reading, and cultural engagement. In many ways, this period marked the beginning of the breakaway from the personal growth model of English which emphasized creative and imaginative fictional and poetic expression to a broader conception of the subject which celebrated a wider range of genres and text types, not without their own kinds of creativity, expression, and imagination (as had already been presaged by Moffett, 1968/1987).

An interest in genres and their place in the teaching of advanced writing increased in the second half of the 1980s, following the publication of Miller's (1984) article, 'Genre as social action'. The notion of genres and their importance to writing curricula and pedagogies derived from film studies and from sociological theory. Essentially, Miller's conception of genre was influential because it saw genres as constituted in social action rather than in text types. The spectrum of possibilities between the social, contextual view of genre and its more narrow textual (formal properties of texts) version has defined the debate and practices in the field, not least because it is at advanced stages in learning to write that the social and textual differentiation between kinds of writing manifests itself. On the one hand, the social, contextual view is represented in writing by Kress (1982/1994), Freedman and Medway (1994); the more linguistically focused view finds expression in the Sydney School of Australian Halliday-derived genre theory and practice, as evidenced in the work of Martin (1991), Christie (1987) and others. Excellent surveys of the field and of the contribution of genre theory to the development of writing abilities are found in Reid (1987) and Cope and Kalantzis (1993). The influence of such theory on curriculum design is felt in many countries.

A greater understanding of writing processes in expert writers that can be modeled in novice writers emerged in the 1980s to early 2000s, deriving from the work of psycholinguistic and discourse modeling by Graves (1983) and Scardamalia and Bereiter (1987). The understanding and mapping of process was reified into a formal system for teaching writing (Calkins, 1986), and creating an abreaction

(e.g. Harwayne, 2001) among those who believed it to be fossilized. The emphasis of this movement to celebrate and teach process was on drafting, editing, and peer conferencing (aided by use of word processing), aiming all the time to capture the voice in writing. Sometimes this movement was linked to or ran alongside a widening of the range of written (and spoken) forms – a development that in itself was reacting against the preponderance of a single, unified narrative/expressive mode and by taking such a wider perspective was aiming to prepare young writers for the world. Such widening of the range of discourse manifested itself in the later versions of the National Curriculum (UK) and in the Australian (specifically, the Sydney School, e.g., Christie, 2002) celebration (and calibration) of a range of genres in writing.

Transitions

There are at least three types of transition in writing at advanced levels that need to be mentioned:

1. one is concerned with the transition from literary or inward-looking textual composition to real-world discourses;
2. another is between imaginative, informational, and an often-neglected argumentative mode of writing; and
3. a third concerns the transition from primary/elementary schooling to secondary/high school, and again from school to college.

The first transition that gets too little attention is that between the classroom and the real world. Although Britton *et al.* (1975) made the distinction between poetic and transactional writing – the latter doing the work of the world and thus giving status to it – much of the writing that was practiced under that heading was informational and/or factual rather than making a difference in the world. In other words, it was inert rather than transactional. Indeed, writing curricula is often divided into creative and imaginative types of writing on the one hand, and informational and instructional writing on the other, suggesting not only that these are the two main types of writing, but also that they are somehow opposed. There are two exceptions to such reductive categorization: writing that made a difference in the world, and argumentative writing; both of which, from time to time, are temporarily given higher profile in writing curricula.

Writing that makes a difference in the world was the subject of an under-published report by Brown *et al.* (1990) which looked to develop English for the UK's Technical and Vocational Education Initiative (TVEI). The key change made to writing practices in this project was that the contexts and audiences for the writing reached beyond the classroom. For example, in one subproject, pupils were

commissioned by the school librarian to survey and make recommendations to her on the journals and magazines to which the library subscribed. A 65-page report, compiled and written over 6 weeks by a group of pupils, was delivered and made a significant difference not only to the library subscriptions, but also to the number of pupils reading in the library. The composition of the report was complex and highly motivating; the accuracy was enhanced by the need for public presentation.

Work on the place of argumentative writing in the curriculum was undertaken by Andrews (1995) and Mitchell (2000), suggesting that at secondary school and at further and higher education levels, the emphasis on the formal essay was not only in need of better teaching, but could be reinvigorated and diversified by exploration of the dialogic and dialectic nature of argumentation as well. To see the basis of learning to write at advanced levels as being in dialog suggests also the close and complex connection between writing and speech. Ong's exploration of orality and of oral cultures on the one hand, and Bakhtin's notion of a dialogic imagination (Bakhtin, 1982) on the other (composed in the 1930s, but published in the West in the 1970s and early 1980s) bring a new dynamic to the learning of and teaching of writing at advanced levels. Essentially, Ong's study of the shift from oral to chirographic (writing) cultures is both historical and synchronic. He sees writing cultures returning to a second orality in which dialogic forms are preferred, and short interchanges in writing (akin to speech) become prevalent. Bakhtin's notion that all writing is created in response to some existing discourse is helpful in that it makes writing appear to be more of a cultural exchange, and more like speaking, than is usually assumed.

The transitions from primary/elementary schooling to secondary/high schools, and/or through the middle school phase constitute a particular concern for the learning and teaching of writing.

Tabor (2004) sets out the particular problems in the teaching of writing in the transition from age 11 to 12 in the English system. In a doctoral study based on the progression of four pupils, he suggests that despite efforts made to ease transition by schools and by curriculum design and implementation (e.g., the National Literacy Strategy in England, followed by the Primary and Secondary National Strategies – with their particular focus on literacy development for 11–14-year-olds), there remains a dip in performance in writing at the start of secondary schooling. Such a dip would not be of concern if the progress of pupils was recovered, but some students do not make a recovery. The problem is part of a larger issue of progression in writing at advanced levels. Tabor (2004: 18–21) provides a very useful table on conceptualizing progression in writing, drawing on many of the key works mentioned in the present account and adding others. What is particularly interesting is that few, if any, of the theories or models of writing account for progression.

1990s to the Present

The period from the mid-1990s to the present has seen a tension between the functions of writing in wider society and those in schooling and assessment. The place of writing within multimodal communication, and especially recognition of its relation with the visual (still and moving images) in popular and indeed all culture(s) has begun to be explored, inspired by writers such as Lanham (1993) or Mitchell (1986). Lanham suggested that the “turn to the visual” in the 1990s, supposedly inspired by the visual/verbal interface of computer screens, was not new but was a phenomenon that had been a preoccupation for cultural analysts since at least the medieval period (through illuminated manuscripts and textual commentaries on iconic images). Mitchell took the idea further, exploring the spectrum of means of representation from image through icon to (written) text, linking the range of media to narrative. Although both writers operate at a high level of theory, their work has been instrumental in helping us to understand the relationship between word and image, and thus the particular characteristic and functions of writing, in late twentieth-/early twenty-first-century practice both within and without the classroom.

Writing processes moved from an understanding and practice of drafting and editing to design issues (Kress, 1995). A move away from the notion of a single personal voice to a multiplicity of voices was registered; and, indeed, through the improved quality of sound on computers (not quite up to general speech recognition and instant translation-into-writing quality). The advent of the mobile learner (see Sharples *et al.*, 2007), accessing written, visual, and audio material anytime, at anyplace, brought about an extension of academic and social space, especially for 11–16-year-olds. Writing began to be conceived as text-box filling, but not always briefly; there was much scope for extended writing, both of an in-depth and functional nature. The need for keyboarding skills became apparent as writing increasingly took place through word processing. One of the key papers of this period was by the New London Group (1996/2000): ‘A pedagogy of multiliteracies: Designing social futures’. This paper put the teaching and learning of writing within a multimodal conception of literacy; and furthermore, within changing social and economic patterns in personal, working lives and with regard to citizenship. It thus would see writing as embedded within school cultures (what Sheeran and Barnes (1991) call school writing), but as critically needing to reflect wider social practices and patterns or representation in order to stay relevant and engaging for advanced writers. The pedagogic problem of engineering such a productive relationship is seen to be a question of design: designs of meaning and the available designs of the grammars of various semiotic systems, providing the basis for the act of designing and redesigning.

In school practice, learning to write is often a matter of taking material from one genre (used in the social process sense) and redesigning it for another. The pedagogy of multiliteracies approach extends this conception in two ways: into the visual and other semiotic codes in hybrid combinations with the verbal spoken and written codes (often foregrounded or highlighted by information and communication technologies); and by looking for more interaction between the practices of school life and social/political practices beyond the classroom.

The major challenge posed by the place of writing in a universe of discourse, and the pedagogies associated with it, is shaping a writing curriculum for 11–19-year-olds that would:

1. Recognize the place of writing within multimodality.
2. Reengage and motivate disaffected or unengaged young people by:
 - a. bringing the genres of schooling closer to the genres of the wider social world, and
 - b. giving writing a range of real purposes.
3. At the same time, use the power of writing to explore depth in thought, reflection, and feeling.
4. Recognize the place of creativity and imagination in nonliterary forms of writing, as well as in literary forms;
5. Recognize and exploit the fact that writing and reading are reciprocal – investigate the similarities and differences, strengths and weaknesses of speaking and writing in different contexts and for different functions – and thus reestablish the link between speaking and writing.

The last point needs further exploration. Speaking and writing are primarily skills of language production, whereas listening and reading are primarily skills of reception.

Speaking and writing can be demanding in that they require expression, articulation, framing, and shaping. To explicate these terms briefly: expression requires motivation to speak or write and the intellectual and motor facility to do so; articulation requires clarity of intention and thought, or at least a move toward such clarity; framing requires selections from the repertoire of socially embedded and generated speech genres and text types (sometimes these genres are hybrid or newly created); and shaping requires the manipulation of language within the those frames of reference, often at the point of utterance. Expression and articulation are part and parcel of the current curriculum but both need to be reemphasized; framing derives from sociological and discourse theory (see, e.g., Tannen, 1993) as a way of making sense of the demands of meaning-making at text level; shaping derives from Britton's notion of shaping at the point of utterance (Britton, 1980), that is, giving credit to the fact that much oral and written communication is not preplanned, but is shaped at it happens.

Expression is important because it engages the self or personae and releases what may be felt and/or thought. It affords channels of communication and creates contact with others.

Articulation aims to make such communication clear. In speech terms, articulation is associated perhaps most readily with surface features such as clear enunciation of utterances; more importantly, the notion of articulation (joining) is about logical or a-logical connections between ideas, thoughts, feelings, and language, in speech and/or writing. Andrews *et al.* (2006b), in a systematic review of research on the teaching of argumentative writing at KS2 and 3, draw attention to the need for cognitive as well as linguistic work in improving writing in this mode. The findings of this report are mirrored in a recent report by the US-based Alliance for Excellent Education – see Graham and Perin (2006) – which, based on a meta-analysis of research studies, concludes there are 11 strategies for improving writing in middle and high schools, including writing strategies, summarizing, collaborative writing, specific product goals (audiences), word processing, sentence combining (cf. grammar review by Andrews *et al.*, 2006a), prewriting (planning), inquiry activities (research), and a process writing approach.

With framing and shaping, the emphasis needs to move from a focus on the end products – the frames (pedagogic scaffolds, genres, text types, and forms) and shapes that language uses and that need to be learned – to the act of framing and shaping that is at the heart of composition (literally, putting things together). Such a move will entail thinking more deeply about the early stages of composition: how ideas are formed; how they are framed; how inspirational ideas are supported by a climate for learning and development; how choices are made, early on, about the medium or media in which it is best to convey the message; how drafting and editing can be improved by critical dialog and reflection at the deeper levels of composition (structure, voice, position, and tone); how momentum and interest can be sustained; how speaking, reading, and listening can contribute to the composing process in writing; how issues of design, balance, and elegance (when is a piece finished?) can be taught and learned; and how a community of learners (speakers/writers/makers) can support such committed and high-quality composing.

One possible reason for the fact that writing performance lags behind reading for the most part is that when listening or reading, the material is given. The intellectual load on the audience or reader might be said to be lighter than when composing in speech or writing, although that load will vary with content and substance in each case. It is generally accepted that writing is the most difficult, if not the most complex, of the four language skills, requiring solitary, creative, thoughtful, accurate, and focused compositional energy; as well as a higher degree of reflective thinking and (usually) personal engagement.

Much has been done, for example, to introduce pupils to nonfiction text types, and to improve their control of stylistic features associated with them. Where practice is weaker is in generating the motivation and purpose to write; without such direction, pupils know how to write but not why they write, how to start or how to engage an audience, and how to generate and marshal ideas. Producing writing of such quality, along with other text types, is one of the keys to overall improvement in English.

It worth reciting how speaking and writing can support each other at advanced levels. First, speaking can be an important rehearsal for writing. Ideas can be discussed in pairs, small groups, in whole-class discussion, and in larger forums, then distilled, translated, and developed in writing. Such writing can be dialogic as well as monologic. Dialogic writing includes planning for Socratic dialog (question-and-answer format), colloquia, play scripts, and other dual- and multi-voiced text types. Monologic writing includes the more conventional forms such as essay, story, letter, and report, where translation from the multiple voices of speech to the single authorial voice of the writer can be more difficult.

Second, writing can be a rehearsal for speech. Individual and/or joint composition in writing can prefigure delivery in speech, as in the making of a speech, the production of an oral narrative, the composition of a persuasive case, or the scripting of a (radio) play or advertisement. Speech as a product in these cases is more than mere performance: it is part of a dialog that invites response in spoken, written, and other formats. It is in such transformation between different means of communication and different genres within those means that the day-to-day practice of English in classrooms takes place.

The problem of insufficient space for sustained speaking and writing in school curricula is compounded by assessment practices.

What is clear is that speaking and writing are central to learning in formal education because they afford the learner the ability to reflect, think, compose, and rearrange as well as respond spontaneously (particularly in the case of speech). Furthermore, as Meek (1983) proves, such emphasis on the productive language skills can be the key to improvements and even breakthroughs for weaker learners not only in speaking and writing themselves, but also in reading and listening as a result of increased motivation, commitment, and investment in making meaning in language; and increased awareness and exploitation of the reciprocity between writing/reading, and speaking/listening.

A further phase of development in writing practices, yet to arrive, will probably see advances in speech-recognition technology that might or might not obviate the need for keyboards or writing implements. The emphasis on composing written text (expression, articulation, framing, and shaping) will shift toward oral composition, while not

abandoning writing. There will be a renewed dynamic relationship between speaking and writing, with each finding their roles in a new economy of communication.

Conclusion

Writing pedagogy is in need of reform to keep up with developments in the digital age and with multimodal perspectives. In particular, there has been little research on writing development at advanced levels. There is no doubt that writing will continue to be an important mode and medium of communication in the twenty-first century. What is a matter of concern is that teachers of writing are not necessarily equipped themselves as writers at advanced levels in a wide range of genres; they thus find the teaching of writing more difficult than that of reading. This article has indicated the range of skills and capabilities required to be a competent and inspiring teacher of writing.

See also: Early Writing.

Bibliography

- Alexander, R. (2006). *Towards Dialogic Teaching*, 3rd edn. Thirsk: Dialogos.
- Andrews, R. (1995). *Teaching and Learning Argument*. London: Cassell.
- Andrews, R., Torgerson, C., Beverton, S., et al. (2006a). The effect of grammar teaching on writing development. *British Educational Research Journal* 32(1), 39–55.
- Andrews, R., Torgerson, C., Low, G., McGuinn, N., and Robinson, A. (2006b). Teaching argumentative non-fiction writing to 7–14 year olds: A systematic review of the evidence of successful practice. Report. In *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London. <http://eppi.ioe.ac.uk/reel> (accessed August 2009).
- Bakhtin, M. M., Holquist, M., and Liapunov, V. (1982). *The Dialogic Imagination: Four Essays*. Austin, TX: University of Texas Press.
- Barnes, D., Britton, J., and Rosen, H. (1969/1989). *Language, the Learner and the School*. Harmondsworth/Portsmouth, NH: Heinemann-Boynton/Cook/Pelican.
- Britton, J. (1967). *Talking & Writing: A Handbook for Teachers*. London: Methuen.
- Britton, J. (1980). Shaping at the point of utterance. In Freedman, A. and Pringle, I. (eds.) *Reinventing the Rhetorical Tradition*, pp 61–66. Las Vegas: Long and Silverman.
- Britton, J., Burgess, T., Martin, N., McLeod, A., and Rosen, H. (1975). *The Development of Writing Abilities (11–16)*. London: Macmillan.
- Brown, J., Clarke, S., Medway, P., Stibbs, A., and Andrews, R. (1990). *Developing English for TVEI*. Leeds: University of Leeds, School of Education.
- Calkins, L. M. (1986). *The Art of Teaching Writing*. Portsmouth, NH: Heinemann.
- Christie, F. (1987). Young children's writing: From spoken to written genre. *Language in Education* 1.1, 3–13.
- Christie, F. (2002). *Classroom Discourse Analysis: A Functional Perspective*. London: Continuum.
- Cope, B. and Kalantzis, M. (1993). *The Powers of Literacy: A Genre Approach to Teaching Writing*. London: Falmer.
- Dixon, J. (1967). *Growth through English*. Oxford: Oxford University Press for the National Association for the Teaching of English.

- Fisher, W. (1989). *Human Communication as Narration*. Columbia, SC: University of South Carolina Press.
- Freedman, A. and Medway, P. (eds.) (1994). *Learning and Teaching Genre*. Portsmouth, NH: Heinemann.
- Graham, S. and Perin, D. (2006). *Writing Next: Effective Strategies to Improve Writing of Adolescents in Middle and High Schools*. Washington, DC: Alliance for Excellent Education.
- Graves, D. (1983). *Writing: Children and Teachers at Work*. Portsmouth, NH: Heinemann.
- Hardy, B. (1977). *The Cool Web: The Pattern of Children's Reading*. London: The Bodley Head.
- Harwayne, S. (2001). *Writing through Childhood: Rethinking Process and Product*. Portsmouth, NH: Heinemann.
- Kress, G. (1982/1994). *Learning to Write*. London: Routledge.
- Kress, G. (1995). *Writing the Future: English and the Making of a Culture of Innovation*. Sheffield: National Association for the Teaching of English.
- Lanham, R. (1993). *The Electronic Word: Democracy, Technology and the Arts*. Chicago, IL: University of Chicago Press.
- Martin, J. R. (1991). *English Text: System and Structure*. Amsterdam: John Benjamins.
- Meek, M. (1983). *Achieving Literacy: Longitudinal Studies of Adolescents Learning to Read*. London: Routledge.
- Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech* 70, 151–167.
- Mitchell, S. and Andrews, R. (eds.) (2000). *Learning to Argue in Higher Education*. Portsmouth, NH: Boynton/Cook.
- Mitchell, W. J. T. (1986). *Iconology: Image, Text, Ideology*. Chicago, IL: University of Chicago Press.
- Moffett, J. (1968/1987). *Teaching the Universe of Discourse*. Portsmouth, NH: Heinemann-Boynton/Cook.
- New London Group (1996/2000). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review* 66, 60–92. (Revised in Cope, B. and Kalantzis, M. (2000), pp 9–37.)
- Ong, W. J. (1982/1988). *Orality and Literacy: The Technologizing of the Word*. *New Accents*, Hawkes, T. (ed.). New York: Methuen.
- Pavel, T. G. (1986). *Fictional Worlds*. Cambridge, MA: Harvard University Press.
- Reid, I. (ed.) (1987). *The Place of Genre in Learning: Current Debates*. Geelong: Deakin University Press, Centre for Studies in Literary Education.
- Rosen, H. (1985). *Stories and Meaning*. Sheffield: National Association for the Teaching of English.
- Scardamalia, M. and Bereiter, C. (1987). *The Psychology of Written Composition*. Hillsdale, NJ: Erlbaum.
- Sharples, M., Taylor, J., and Vavoula, G. (2007). A theory of learning for the mobile age. In Andrews, R. and Haythornthwaite, C. (eds.) *The Sage Handbook of E-Learning Research*, pp 221–247. London: Sage.
- Sheeran, Y. and Barnes, D. (1991). *School Writing: Discovering the Ground Rules*. Milton Keynes: Open University Press.
- Tabor, D. (2004). *Young Writers at Transition*. London: RoutledgeFalmer.

- Tannen, D. (ed.) (1993). *Framing in Discourse*. New York: Oxford University Press.

Further Reading

- Andrews, R. (ed.) (2004). *The Impact of ICT on Literacy Education*. London: RoutledgeFalmer.
- Beaufort, A. (1999). *Writing in the Real World: Making the Transition from School to Work*. New York: Teachers College Press.
- Cope, B. and Kalantzis, M. (2000). *Multiliteracies: Literacy Learning and the Design of Social Futures*. London: Routledge.
- Hillocks, G. (1995). *Teaching Writing as Reflective Practice*. New York: Teachers College Press.
- Jones, R. and Wyse, D. (2000). *Teaching English, Language and Literacy*. London: Routledge.
- Massey, A. J., Elliott, G. L., and Johnson, N. K. (2005). *Variations in Aspects of Writing in 16+ English Examinations between 1980 and 2004: Vocabulary, Spelling, Punctuation, Sentence Structure, Non-Standard English*. Cambridge: Assessment Directorate, Cambridge Local Examinations Syndicate.
- Medway, P. (1980). *Finding a Language: Autonomy and Learning in School*. London: Chameleon.
- Myhill, D. and Fisher, R. (2005). *Informing Practice in English: A Review of Recent Research in Literacy and the Teaching of English*. London: Office for Standards in Education (Ofsted).
- Russell, D. R. (2002). *Writing in the Academic Disciplines*. Carbondale, IL: Southern Illinois University Press.
- Sharples, M. (1999). *How We Write – Writing as Creative Design*. London: Routledge.
- Vygotsky, L. (1986). *Thought and Language*. Cambridge, MA: MIT Press.
- Wilkinson, A. (1986). *The Quality of Writing*. Milton Keynes: Open University Press.

Relevant Websites

- <http://www.caribou.bc.ca/disciplines>
- <http://www.ifte.net> – International Federation for the Teaching of English.
- <http://www.jslw.org> – Journal of Second Language Writing.
- <http://www.nate.org.uk> – National Association for the Teaching of English.
- <http://www.literacytrust.org.uk> – National Literacy trust.
- <http://www.writingproject.org> – National Writing Project.
- <http://www.ncte.org/prog/writing>
- <http://www.thinkingwriting.qmul.ac.uk> – Thinking Writing.

LEARNING AND COGNITION – OTHER ISSUES

Contents

Gender and Learning

Parent–Child Relationships in Early Learning

Peer Interaction and Learning

Gender and Learning

L Yates and J McLeod, The University of Melbourne, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Hidden curriculum – The knowledge and understandings that students learn in school that are not part of the official syllabus curriculum, and not intentional. For example, if textbooks consistently show men in positions of authority and women in subordinate roles (or if principals are always men and infant teachers always women) this is seen as shaping girls' and boys' expectations of their pathways in and beyond school.

Social construction/social constructionist – A description of how identities are shaped and formed by social factors, such as language, roles, traditions, expectations, and the like. It is contrasted with views that see identity as solely the result of fixed biological or natural impulses and orientations.

Voice, women's voice – In educational research, voice refers to the expression of someone's identity, as communicated in many ways, not only in speech. Women's voice conveys a sense of women's distinct qualities, orientations, and differences, such as preferring humanistic rather than abstract orientations to problems.

One of the most prominent developments in educational thinking in the last three to four decades has been the rise of gender as a central issue for learning and curriculum, and for education more broadly. This reflects a world where equality between women and men is a formal expectation for education systems and where there is a good deal of monitoring and research on patterns of educational participation and achievement by gender. The attention to gender in research and policy on learning also reflects some significant changes that began in the second-half of the twentieth century, with researchers moving away from simply accepting sex differences as an explanation for different patterns to focusing in much more detail on the

ways teaching practices and school arrangements interact with the characteristics and experiences girls and boys bring to learning. In this work, learning (and teaching) is understood as shaping gender as well as the converse.

A large array of different kinds of research and policy-making has been carried out on gender and learning. It can now be found in most disciplines (or foundations) that study education, in many areas of school curriculum subject-specific study, and in many different kinds of research: longitudinal, qualitative, quasi-experimental, and action-research projects as well as in large-scale data monitoring. Research on gender and school learning also encompasses a broad range of topics, including attitudes to subject areas and subject choice; the effect of the portrayal of males and females in text books; the influence of teacher and classroom language; effects of male and female teachers and role models and of different teaching methods and class arrangements; the impact of peer culture; cognitive style; assessment methods and their effects; and patterns of achievement in cohort or end-of-school examinations as well as in curriculum areas. The last four decades have seen some significant shifts of interest across the field of gender and learning: moving from debunking deficit assumptions of sex differences in the 1970s, through a broad concern with girl-friendly and inclusive approaches in the 1980s, to a new kind of concern with boys' learning in the 1990s.

The article begins with a snapshot of shifts in thinking about gender and learning and a brief comment on terms and concepts. This provides background to a mapping of key themes and dilemmas addressed in research on gender and learning, particularly following the influence of feminist and equal-opportunity reforms in many education systems from the 1970s onward.

Gender Compared to Sex Differences

The term gender itself is used in a variety of ways, and has been the subject of much debate. In general, the term gender

signals an attention to the socially formed nature of male and female identities, relationships, and orientations rather than seeing these as simply inherent or biological differences. This contrasts with earlier uses of the term sex to denote differences between males and females, which often assumed such differences to be a fixed, empirical, and innate phenomenon. From the late 1970s, gender began to replace sex as the preferred descriptor of male/female identity differences, and this is seen in a shift from a focus on sex differences in learning to a focus on gender and learning.

In the sex-differences mode, researchers tended to assume that the category of sex was stable, and research was concerned to investigate its effects. In the gender-research approach, research was looking to investigate how patterns of differentiation were produced, and, in the case of learning, how practices could be changed to produce more equal outcomes. With this latter approach, learning is viewed as an outcome of the social experiences in and out of schools to which young people are exposed. Gender was differentiated from biological sex, and was used to describe differences in what students liked and did not like, what skills and knowledge they found easy and what they found difficult. If girls did not progress as well in science, for example, the explanation might be found in their preschool play with dolls rather than mechanical objects; similarly, boys' weaker literacy skills were attributed to less-early-years involvement in group and language-based play. Or it might be the result of in-school social influences, such as particular types of teaching and assessment practices that are discussed below.

Initially, in many public and policy contexts, gender was regarded as a term denoting a rather radical perspective, a code for feminist approach. Sex remained the preferred, more familiar term. However, this view of the terms has shifted markedly in the last three decades in the English-speaking world. Many official sources now label their formal policy and data categories as gender rather than sex. Simultaneously, as research on gender has continued to develop, much of it no longer takes such a strong social-constructionist position about the desires and orientations girls and boys develop in the preschool years and beyond. A range of different research traditions, from brain researchers to cultural studies practitioners, now understand human differences and developments as arising out of interactions between biological and social. On the one hand, biology and drives are seen as producing some predispositions in individuals and, on the other, social, educational, and familial environments are seen as driving how and whether these dispositions are developed (including biologically) or turned into behaviors, preferences, and outcomes. Individuals are not seen as a *tabula rasa* upon which social messages are neatly inscribed. Nevertheless, the social environment is seen as a key formative shaper of individual identities and learning outcomes because of the way it interacts with and develops inherent biological character-

istics. The implication for school learning remains one that emphasizes social arrangements, in and out of school. To see gender as an important element in how learning takes place and in what learning occurs, requires attention to experiences and values developed and conveyed in families and the broader society, as well as attention to what is said and conveyed in a classroom.

In summary, since the late 1970s, there has been an explosion of research on the social construction of gender and an accompanying recognition of the importance of gender in relation to learning. This, in turn, has contributed to changes in conceptions of learning that have expanded beyond models or measures of individual cognition to include a strong focus on the sociocultural processes and contexts shaping learning. This research on learning also shows the influence of ideas coming from the new sociology of education, particularly notions such as the hidden curriculum, which drew attention to the powerful knowledge and know-how that was not part of the official syllabus and classroom program but which was a vital part of students' overall learning.

An additional point to emphasize here is that research on gender and learning is not only concerned with how gender influences learning, but also with how school learning influences gender. This is both in terms of individuals' developing sense of who they are, and what potential they have (or lack), and in terms of broader patterns of outcomes for men and women in the world. A further point is that although the definition of terms offered here explains some broad differences in how sex-differences research and gender-and-learning research have been oriented, the use of these terms is not invariable or mutually exclusive. A lot of sex-differences research still continues, and large-scale national and international testing, such as the Organization for Economic Cooperation and Development (OECD)'s Programme of International Student Assessment (PISA), continues to routinely monitor differences by sex or gender, as well as by the intersection of such differences with other social variables, such as level of parental education, country of birth, and numerous proxies for socioeconomic status.

Shifts in Focus of Work on Gender and Learning

Gender and learning became an issue for education as much because of social developments and new expectations regarding the roles of women and girls outside education research fields, as from developments within these fields. For example, the tendency for girls not to continue with mathematics and science and to do less well in some subjects had been known for some time, but to some extent was taken simply as natural. It was when different expectations about women's social outcomes

were raised from the 1970s onward that researchers began to give serious attention to whether such outcomes were, in part at least, the result of how schools and teaching were organized, and that these arrangements and practices could be changed.

Initially, in its translation into school policies and practices, gender was predominantly addressed as an issue that affected girls (Yates, 1998), and in reality much discussion and policy was directed to girls rather than to gender or gender relations. The new problem was understood to be why are girls not doing as well as boys? The challenge was to expose and critique what had been taken for granted as the normal state of affairs, and to show that differences in learning expectations, opportunities, and outcomes that systematically disadvantaged one gender group were unacceptable.

Research on gender and learning initially took two main forms: exposing flaws in previous research that had seemed to prove the innate inferiority of girls compared with boys, and identifying practices in schools where girls might be given different opportunities and messages than boys (AAUW, 1992). Before the 1970s, for example, it was common for the higher achievement and retention of boys in mathematics compared with girls to be attributed to innate differences, in particular to differences in spatial cognition, with boys on average judged to be naturally more adept at this than girls. A wave of influential feminist-based research in the following decade systematically reexamined this earlier work, exposing flaws in its design and arguments and encouraging greater participation of girls in traditionally male learning domains (e.g., Kelly, 1981; Fennema and Leder, 1990; Friedman, 1995). Research on the different ways schools conveyed implicitly different messages to girls about their potential included work on language and images in textbooks and work on teachers' different kinds of interactions with male and female students, as well as peer interactions.

One example of such research investigated how children responded to teacher feedback and found this shaped the learning and sense of competence of both girls and boys. Wood (1998) reported that while boys often received negative feedback from mathematics teachers, this was usually directed at their lack of attentiveness rather than their mathematical ability. When boys received positive feedback, it was likely to be in terms of their intellectual competence. In contrast, while girls tended not to receive a lot of negative feedback, any such responses were likely to be directed to their lack of ability. Thus "for boys criticism is common . . . and it does not reflect upon their competence. . . . Because negative feedback is rare for girls, when a girl does receive it, it forms a notable event and is likely to convey the idea, to herself and her peers, that she is not very good at the subject" (Wood, 1998: 287).

Another influential body of work examined the hidden or implicit but powerful expectations teachers held about

what kind of students were the right kind of learners. Valerie Walkerdine's studies of early childhood and primary-school classrooms showed some of the complex and subtle ways in which teachers' beliefs about what good learning looked like were gendered, with boys typically regarded as potentially more disruptive in the classroom but also more risk taking and independent as learners, and girls as more cooperative and diligent but less likely to be exceptional or brilliant students (Walkerdine, 1989). Walkerdine argued that conceptions of learning and cognitive development were themselves deeply gendered and that the ideal rational learner was implicitly masculine.

In other work, a wave of interest in women's ways of knowing or women's voice sought to identify specific qualities in women and girls' learning and to devise pedagogies and forms of curriculum that would promote the realization of these qualities. The work of Carol Gilligan was particularly influential. Gilligan's (1982) revisionist account of Kohlberg's framework of moral reasoning argued that hierarchies of intellectual development had been founded on male styles of reasoning, particularly valuing abstraction and judgment. She laid out ways in which connectedness, or reasoning that included attention to human effects, can be seen not simply as a lower stage of the first hierarchy, but as a different hierarchy altogether with similar stages of complexity. The well-known American study, *Women's Ways of Knowing* (Belenky *et al.*, 1986), influenced by Gilligan's work, argued that women had been silenced by traditional educational processes and that they were more likely to find their own voice and develop as learners when the learning is connected. This occurs when learning is connected to first-hand experience, values question posing, learning in context, and relative, context-specific truth. Such an orientation is contrasted to traditional masculine ways of learning that value objectivity, abstraction, and general principles. Considerable debate and further research followed from these attempts to identify female and male learning style, particularly in terms of whether these paid sufficient attention to differences within the category of girls, especially in relation to race and social-class differences.

By the mid-1990s, a new focus of gender research and policymaking began to emerge – the issue of boys and their learning preferences and needs (Weaver-Hightower, 2003). In part, this was motivated by concern about boys who were dropping out of school, and in part by a perception that there had been a lot of reform attention given to girls but little to boys. Some of this work took a similar form to the earlier gender-based research on girls, focusing on the disadvantaging effects of teaching methods. In the case of boys, targets included the growing need for higher levels of language skills in all subjects, and assessment practices – essay-based compared with short-answer testing – and how these interacted with patterns of achievement for boys (Epstein *et al.*, 1998). Other research tended more toward

a sex-differences approach, in emphasizing the strength of different innate drives, and questioning how well schools dealt with these. Within the boy turn in research on gender, there are some sharp disagreements about directions for boys and recommended school practices. But all the research broadly falls within a focus on gender as defined earlier in this article, because it all pays some attention to the interaction between social arrangements inside and outside schools, and the ways these produce orientations and outcomes in relation to schooling for a particular group.

The research on boys and learning is often crudely divided into pro- and anti-feminist work. Pro-feminist work rejects the idea that boys rather than girls are disadvantaged by school practices, and sees the need for attention to both. It sees its approach as building on the earlier work on girls and gender, and often focuses on the way gender (learned ways of being masculine) disadvantages boys' learning and needs to be changed. Anti-feminist approaches tend to argue that boys have replaced girls as the new disadvantaged group, and to see boys' gendered preferences as a given with which schools should work, rather than as socially constructed or sometimes detrimental. Weaver-Hightower (2003) provides a good overview of the different lines of work in this area.

Gender-Equity Reforms and Reviews

During the 1990s, after about two decades of deliberate gender-equity reform, many national governments and international agencies commissioned major studies of gender and school participation and performance (e.g., Arnot *et al.*, 1999). While there were national differences, one overwhelming finding from much of this overview research was that girls' participation and retention had increased, and that in many subject areas the average girl was performing at a higher level than the average boy. At the same time, this research drew attention to the question of who actually was the average girl or boy, and what factors other than gender were influencing school learning and educational outcomes. For example, although there were some average differences in outcomes between girls as a group and boys as a group, these were usually smaller differences than differences within the category of girls or within the category of boys, particularly in relation to socioeconomic status or social class. As noted above, the OECD's PISA continues to produce its own comparative data on these issues, showing some of the relative effects of different variables on school achievement.

The various national reviews of the effects of large-scale policy attention to gender in schooling have shown that changes in school teaching and assessment practices do produce different patterns of outcomes in relation to gender. That is, they have confirmed the understanding that changes in education practices – school teaching, learning, and assessment – are able to influence patterns

of outcomes by gender. This, in turn, raises some questions that are at issue in contemporary debates about boys. Should the mandate of schooling be to produce identical patterns of outcomes for girls and boys (e.g., regardless of how hard they work)? Should teachers and teaching approaches proceed on the assumption of gender-based differences in learning preferences and styles? Should there be common or differentiated expectations of learning? How far should forms of assessment be modified to produce gender-neutral outcomes?

For example, it is generally agreed that on average boys and girls develop at different rates in the early and primary years (for discussion of this issue see MacNaughton, 2003), but the degree to which school-starting age and national age-based or year-based testing should take account of that continues to be in some dispute in different policy settings. Similarly, a range of studies have found that boys are overrepresented in both remedial and gifted streams, where these exist, and in named learning difficulties categories, such as attention-deficit disorder, but there is debate on the extent to which social practices and assumptions in relation to gender produce such results and need to be changed.

Single Sex and Coeducation and Learning

The heightened interest in learning and achievement patterns by gender, and the recognition that school and classroom arrangements are important, has produced a continuing interest in the impact of single sex compared with coeducational arrangements. A very large number of studies have been conducted on this topic and it continues to prompt considerable scholarly, policy, and community attention (Salomone, 2003). The topic itself is an interesting case of the multifaceted way in which gender and learning come together, since whether or not a school or a classroom is girls-only or boys-only is only one of a number of important elements of the gender and learning environment, which includes the school culture overall, curriculum choices and supports, the expectations of teachers, and the expectations of students themselves and their parents. It is not surprising, then, that empirical data comparing the results of single sex and coeducation have often not been consistent between different studies or over time. However, they do show small overall indications that, controlling for other aspects of school-intake difference, single-sex groupings may have a slightly beneficial effect on girls' achievement and self-esteem and on their likelihood of choosing masculine subjects; and, further, that coeducational groupings may have a slightly positive effect on boys' achievement.

However, overall patterns here are of less significance than how particular arrangements may affect specific groups of girls or boys, or particular purposes. Some single-sex

arrangements have been put in place almost as a form of action research. These include a range of experimentation with single-sex classes to encourage greater participation in certain subject areas, such as single-sex classes for girls in for mathematics in the middle years; as well as attempts to establish small alternative single-sex schools for students who are at risk, especially boys.

Gender in Interaction with Other Factors

Much research has taken place on gendered preferences and achievement in different school subjects. Such research often attends to the ways in which particular subjects themselves have gendered identities (e.g., mathematics as male, English as female) that produce different patterns of take-up and achievement. But gender is not a single or simple phenomenon, and patterns of take-up are strongly influenced by socioeconomic and ethnic backgrounds of students as well as by expectations and practices of different kinds of schools. How gender intersects with social-class background leads to well-documented differences in patterns of achievement. Based on achievement data, it would appear that the higher their socioeconomic status, the fewer disadvantages girls experience in mathematics and science and the fewer disadvantages boys experience in English. Gender relativities in learning outcomes are thus weakest – but not absent – among those with the greatest material and cultural advantages (Teese and Polesel, 2003).

Current and Emerging Issues

The kinds of monitoring and comparisons involving gender are now well established in most national systems (at least those covered by the OECD) and are likely to continue. If obvious disparities are evident, these generate a new wave of research attention to that subject area, or to parts of the education system. In recent years, for example, new technologies and particularly the more widespread use of computers have been a significant issue for learning in schools and for work opportunities beyond schools. This has produced a strong focus on gendered take-up and preferences within this work, often mirroring the kinds of work that took place earlier in relation to mathematics and science, including attention to role models, the way the subject is portrayed culturally, and in schools, peer attitudes, and assessment practices.

Within broader policy directions, the increasing emphasis on international benchmarking, school effectiveness, and more intense testing commonly includes gender as a basic category for comparison, yet often works to downplay the range of the social and identity factors that students bring into classrooms. Its emphasis tends to be on effective techniques for teaching or

managing school systems, rather than social interactions within schooling and the classroom.

In terms of a global perspective, a lot of work on gender has focused on issues of access to schooling or to levels and programs within schooling systems. This research has received new impetus with the UN Millennium Development Goals, which included a target of eliminating gender disparity in primary and secondary education, preferably by 2005, and to all levels of education no later than 2015. The goals and targets here give continued attention to the ongoing strong interaction between schooling and social expectations and practices around gender. Two areas that are subject to renewed attention as part of this work are also areas of heightened concern in developed countries: bullying and sexual harassment as a hidden component of school learning; and the legitimacy or illegitimacy of promoting gender-differentiated expectations via the curriculum, including religious framing of education.

A focus on gender and learning has produced some ongoing attention to differences in what students bring to schooling as part of the learning environment, and some changing expectations of the responsibilities of schooling to acknowledge, support, or attempt to modify these. It has also produced a continuing attention to patterns of subject choice, retention, and success and failure in schooling as something that are not simply a given or produced by what students bring to the task, but in part at least an effect of what schools and teachers do, an effect of how learning environments operate for different kinds of students. The framing of research on gender and learning also reflects political differences in the questions and assumptions researchers bring to education and its relation to broader social forms, as well as changing social expectations about standards and outcomes for education for both women and men in the twenty-first century.

Bibliography

- AAUW (American Association of University Women) (1992). *How Schools Short Change Girls*. New York: Marlowe.
- Arnot, M., David, M., and Weiner, G. (1999). *Closing the Gender Gap: Postwar Education and Social Change*. Cambridge: Polity Press.
- Belenky, M. F., Clinchy, B. M., Goldberger, N. R., and Tarule, J. M. (1986). *Women's Ways of Knowing: The Development of Self, Voice and Mind*. New York: Basic Books.
- Epstein, D., Elwood, J., Hey, V., and Maw, J. (eds.) (1998). *Failing Boys? Issues in Gender and Achievement*. Buckingham, PA: Open University Press.
- Fennema, E. and Leder, G. (eds.) (1990). *Mathematics and Gender*. New York: Teachers College Press.
- Friedman, L. (1995). The space factor in mathematics: Gender differences. *Review of Education Research* 65(1), 22–50.
- Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women's Development*. Cambridge: Harvard University Press.
- Kelly, A. (ed.) (1981). *The Missing Half: Girls and Science Education*. Manchester: Manchester University Press.
- MacNaughton, G. (2003). *Shaping Early Childhood: Learners, Curriculum and Contexts*. Maidenhead: Open University Press.

- Salomone, R. (2003). *Same, Different, Equal: Rethinking Single-Sex Schooling*. New Haven, CT: Yale University Press.
- Teese, R. and Polesel, J. (2003). *Undemocratic Schooling: Equity and Quality in Mass Secondary Education in Australia*. Melbourne: Melbourne University Press.
- Walkerdine, V. and Girls and Mathematics Unit Institute of Education (1989). *Counting Girls Out*. London: Virago.
- Weaver-Hightower, M. (2003). The "boy turn" in research on gender and education. *Review of Education Research* 73(4), 471–498.
- Wood, D. (1998). *How Children Think and Learn*, 2nd edn. Oxford: Blackwell.
- Yates, L. (1998). Constructing and deconstructing 'girls' as a category of concern. In MacKinnon, A., Elgqvist-Saltzman, A., and Prentice, A. (eds.) *Education into the 21st Century: Dangerous Terrain for Women*, pp 155–167. London: Falmer.
- Berge, B.-M. and Ve, H. (2000). *Action Research for Gender Equity*. Maidenhead: Open University Press.
- Connell, R. W. (2002). *Gender*. Cambridge: Polity Press.
- Erskine, S. W. (ed.) (1999). *Gender Issues in International Educations: Beyond Policy and Practice*. New York: Falmer.
- Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women's Development*. Cambridge: Harvard University Press.
- Mael, F. A. (1998). Single-sex and coeducational schooling: Relationships to socioemotional and academic development. *Review of Educational Research* 68(2), 101–129.
- OECD (Organisation for Economic Co-Operation and Development) (2008). *PISA Database 2006*. <http://pisa2006.acer.edu.au/interactive.php> (accessed August 2009).
- United Nations (2000). *Millennium Goals: Goal 3, Promote Gender Equality and Empower Women*. <http://www.un.org/millenniumgoals/gender.shtml> (accessed August 2009).
- Walkerdine, V. (1988). *The Mastery of Reason: Cognitive Development and the Production of Rationality*. London: Routledge.

Further Reading

- Aikman, S. and Unterhalter, E. (eds.) (2005). *Beyond Access: Transforming Policy and Practice for Gender Equality in Education*. Oxford: Oxfam Publishers.

Parent–Child Relationships in Early Learning

C P Edwards, S M Sheridan and L L Knoche, University of Nebraska–Lincoln, Lincoln, NE, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Autonomy – The developmental achievement of the capacity to take care of one's own basic needs (e.g., feeding and dressing oneself, or toileting oneself); functioning in an age-appropriate way in one's usual settings (e.g., sitting at the dinner table or playing with peers at preschool); and exhibiting emotional self-regulation (e.g., controlling anger and distress or using cultural-display rules in expressing one's needs and feelings). Autonomy is always interpreted within a context defined by age, culture, and gender rules and involves interdependence as well as independence.

Curriculum of the home – The interactions and environment for child learning provided by parents and other family members. Parents provide for their young child's general learning both at home and outside the home in many direct and indirect ways, as apparent in their daily routines, strategies for teaching, disciplining, and scheduling children's lives. Some of this behavior and provision by parents is conscious and intentional, but much of it takes place without conscious planning and intent and is part of the implicit culture.

Nurturance – The behavior intended to promote the health, learning, or development of another person, through offering attention, approval, instrumental help, food, material goods, permission to proceed, or other kinds of emotional or physical resources.

Parent engagement – The nurturing behavior by children's primary caregivers (usually mothers or fathers, but sometimes guardians) intended to promote their survival and pass on skills important to their present and future success in their society. Parental engagement can be expressed in many styles and may focus on physical, emotional, and/or cognitive aspects of children's learning, development, and well-being.

Parental behavior during a child's first 5 years of life is critical for the development of important social and cognitive outcomes in children. The child's first relationships, it is now clear, are critical for the establishment of competences – cognitive, social–emotional, and self-regulatory skills – that set the stage for lifelong adaptation and functioning. The interactions and experiences that

children have in the home and family setting provide a framework for how the child will interpret his or her world and give meaning to culturally framed events. Even the degree to which children are prepared to benefit from later schooling is predicated in part on what transpires before they enter the school door. This article reviews some of the key findings about the importance of parent–child relationships in early learning. The term, parent, will refer to the primary parenting figure in a child's life; it may refer to the child's guardian or even to a small number of attachment figures who closely share parenting duties.

Parental behavior consists of an almost infinite variety of specific actions that unfold over time as the child develops, but in our work we have found it useful to summarize three key dimensions of parental behavior that we call parent engagement (Sheridan *et al.*, in press). These three dimensions appear to facilitate child learning and develop in conceptually distinct and practically important ways that are described. The dimensions of parent engagement include: (1) warmth and sensitivity, (2) support for a child's emerging autonomy, and (3) active participation in learning. All three dimensions influence the developmental pathways, including neural capacities, leading to social–emotional, cognitive, and communicative competence (Shonkoff and Phillips, 2000).

Parental Warmth and Sensitivity

The first dimension of parental engagement clusters around warm and sensitive responsiveness to the child's needs and cues. It includes all those behaviors variously described in the child development literature as loving nurturance, warmth and sensitivity, responsive contingency to children's cues, and emotional availability toward the child. Beginning in the neonatal period, parental responsiveness can be seen in adults imitating and highlighting infant behavior, pausing to give the infant an opportunity to respond, respecting the infant's needs for an occasional break from communication, responding enthusiastically and appropriately to the infant's interests, following the infant's attentional focus, and letting the infant initiate interactions. As children grow older, parental warmth and empathy have been identified as global qualities that lead children to interact more smoothly with their parents and to form a strong identification

with parental values that extends outside the home to cooperation with other adults and peers as well.

Children's very survival and development depend on parental warmth and sensitivity because children are inherently relationship-seeking beings. From the beginning of life, children seek to engage and interact with the people around them. When comfortable and fed, they direct their attention and interest outward toward others who seem friendly, exciting, or loving. They reach out to get responses from these people and to send them signals of distress or pleasure as they try to help manage the pace, flow, and intensity of interaction. They actively strive to participate in the life around them. Without intimate, nurturing responses from others, children become too upset and exhausted to accept food and comfort. They cannot make sense of sensory stimulation and understand the world, connect to it, or care about it. Warm and sensitive parents create the framework for this vital interaction in the process of meeting their infant's basic needs.

Ample evidence exists that this first dimension of loving care is positively related to the all-critical development of the child's first attachments and close, secure relationships with a few significant others. Warm and sensitive caregiving that includes encouragement and support, lays the foundation for secure behavior and exploration such as through extended play episodes and pretend play (Ainsworth *et al.*, 1972). Much of the evidence emanates from research conducted within the attachment paradigm. Securely attached children tend to engage in more spontaneous reading activities and perform better on emergent literacy measures than insecurely attached children. In preschool, observers describe securely attached children as more curious, self-directed, sensitive to others' needs, and eager to learn than children who were insecurely attached as infants. Children with less-secure relationships with their caregivers tend to have lower levels of behavioral and emotional control, less adaptive levels of autonomy, and to experience difficulties approaching learning tasks (e.g., Sroufe, 1983).

Parental interactions that include displays of affection, physical proximity, contingent positive reinforcement, and sensitivity have repeatedly related to children's cognitive growth over time. Specifically, research has identified that positive, early relationships between children and caregivers contribute to neural connections that facilitate children's long-term developmental success (National Scientific Council on the Developing Child, 2004). Children in more highly connected parent-child relationships tend to display more positive socioemotional outcomes, such as stronger prosocial orientations, more numerous and higher-quality friendships, and higher levels of peer acceptance in kindergarten (Clark and Ladd, 2000). Through connected interaction with parents, children appear to develop an empathic socioemotional orientation that serves as a foundation for interpreting social situations and responding more prosocially to age-mates and teachers.

Clearly, young children benefit in the short and long term from nurturant caregiving that is emotionally warm, available, and responsive. Yet, there are many styles in which this caregiving can be delivered. Nurturance can be demonstrated in many ways all of which seem to promote infant health and well-being. No single cultural group or set of parents uses all of the available techniques, but instead each selects out some of them to make the customary approach. Parents and communities often use styles that emphasize either a physical, social, or cognitive style of expressing warmth and sensitivity. For example, certain kinds of parents may emphasize a physical style of nurturance, for example, focusing on the child's desires for food, holding, and responsive touch (by day or night) (Edwards and Whiting, 2004; Whiting, 1994; Whiting and Whiting, 1975). Through provision of food, holding, and other primary care oriented to the child's survival, these parents communicate to their children that they love them and are devoted to them. Through gentle touch, physical games, or use of massage, they communicate their nurturing feelings and tell their child that they wish her to feel ease and comfort throughout her body. In contrast, other kinds of parents may take greatest pleasure in a social style of nurturing by singing to the child, grooming their child's hair, dressing the child up, taking her on visits, and teaching her social words and gestures. Indeed, in many cultures, adults take great delight in the social forms of nurturance and communicate their affection through beautifying their child and teaching the child the rudiments of good manners. Finally, a third kind of parents may emphasize a cognitive style of expressing warmth and sensitivity by responding to the child's developing interests and preferences, offering them objects to look at and manipulate, and following their eyes to see what they are looking at, in order to label those things and expand on the child's exclamations and words. These parents often are verbal in their interaction with even the youngest children, and they treat their babies as conversational partners and intelligent beings who wonder about how things work and what causes things to happen. Of course, all three styles can be combined.

In today's postindustrial societies, it is the third style, focused on cognitively stimulating interactions, which seems to lead to the optimal outcomes for children's school readiness and academic success. Warm interactions of the mother provide the foundation for compliance and internalized controls in young children; and limit setting and discipline may be less effective in the absence of positive, warm relationships. The expressions of positive affect and emotional availability are also associated with improved short-term cognitive performance and long-term effects of positive academic performance. The emotional, social, and behavioral competence of young children predicts their academic performance in first grade over and above their

cognitive skills and family backgrounds, whereas the absence of a secure attachment with a caregiver or multiple caregivers leaves a child at a distinctive disadvantage. Qualities of parental engagement have been linked to a number of adaptive characteristics in preschool children, such as good work habits, frustration tolerance, fewer behavior problems, and better social skills.

Parental Support for Autonomy

The second dimension of parental behavior clusters around parental guidance and support for autonomy. It includes all those behaviors variously described in the child development literature as discipline, positive guidance, and support for the development of independence, self-reliance, and self-regulation. Children cannot remain infantile forever and must learn to do things for themselves so they can get along without constant supervision. They must individuate from their attachment figures and develop a certain initiative in relation to their surroundings. This dimension of parent behavior begins at the child's birth but becomes particularly salient (and conscious to parents) during the toddler years when children begin to be resistant and to want to do things independently. Parents promote autonomy by helping children to care for their own needs in the areas of eating, dressing, and personal hygiene. They support their capacity to function maturely in the home, neighborhood, or school classroom by teaching them to regulate the expression of their needs and emotions, respond cooperatively and compliantly to adult authority and direction, resist temptations to misbehave and violate rules, and find ways to tolerate frustration and stay on task.

Parental support for children's autonomy has been associated with the development of many positive cognitive and social outcomes for young children. Parents teach and model skills that help children to recognize and express feelings in culturally appropriate ways so that they are not rejected by others for crying, screaming, or expressing anger in ways that are too violent and uncontrolled. By supporting their child's independence and inviting children to participate in decision making, parents foster self-regulatory skills and intrinsic motivation in children that serves them well in any situation, but especially in school and work-related settings. By providing developmentally sensitive support for problem solving, they promote children's ability to learn from others and work cooperatively on home or school tasks. Parent-child interactions that are attentive but nondirective provide children some guidance, but they also allow children the freedom to be expressive, initiating, and self-directed. Interactions that are monitored and responses that are matched to children's developmental abilities and interests can foster continued interest in a current activity,

comfortable exploration of its potential dimensions, and mastery motivation.

Research indicates that by promoting autonomy, parents promote desired outcomes such as effective communication with peers, self-regulation, adaptive levels of social assertiveness, and self-directedness in social and play interactions at preschool, and increased levels of cognitive competence in young children. By supporting their child's independence and inviting children to participate in decision making, parents foster self-regulatory skills and intrinsic motivation to accomplish tasks set by adults, peers, or themselves. Children of parents who support autonomy have shown higher scores on standardized tests of literacy and numeracy skills, as well as more appropriate social assertiveness and self-directedness. In contrast, parents who undermine autonomy through greater frequency of controlling, hostile behaviors have children with more disruptive behavior problems in early childhood. Children of parents who provide inconsistent guidelines, are harsh or coercive, disengaged, and/or are unable to appropriately monitor child behavior are likely to display more aggression and antisocial behavior. In studies of child-mother interaction, differences in parenting discipline account for a substantial portion of the variance in behavior problems in childhood.

As with warmth and sensitivity, there are many styles by which parents can promote their children's autonomy. Some families encourage motoric or physical autonomy by allowing their child lots of opportunity for active movement. They might allow their child to explore independently in a carefully childproofed home or yard, or take the child outside for regular vigorous exercise. Still others might encourage physical autonomy by being exceptionally patient as the child struggles to climb stairs, use a fork, put on shoes, wash their hands, pour cereal and milk, or put things away. Another dimension of autonomy is social. Some parents may emphasize social autonomy because it helps the child to function in a social group without constant parental intervention. Parents orient their babies toward social autonomy, for instance, when they help them learn to remain patient and pleasant during a long family meal or to control their jealousy of a smaller baby visiting their house. Socially oriented families might also encourage their child to enter a playgroup of same-aged peers, to freely share their toys, or to accept another adult's care to the extent of being able to join the fun of the family outing to the park or swimming pool. Still other families put a premium on cognitive styles of autonomy, and they demonstrate this by encouraging early mastery of language skills, so that the child can use words to express his needs ('up,' 'milk,' 'do car'). They may put special energy into helping the children solve their own problems and make independent use of a nice playroom with its rich store of books, creative art materials, and constructive toys.

Cultural values about autonomy influence the manner in which parents and other family members evaluate and set limits with their children. For instance, there are many ways that parents can express praise and approval for what a child is doing. Some parents and cultural groups tend to use applause and hurrahs to encourage small children to show off and do little performances. In other cultures, parents do not want their children to seek attention or be boastful and proud; so instead of giving overt praise, they comment to another adult how well the child is doing, give the child another responsibility that indicates his success with the first one, or wordlessly display the child's lovely picture to share it in the family. The child notices what the parent is doing and feels a quiet pride that does not make him the center of attention.

Parents' values about appropriate autonomy also influence what they see as too indulgent and as spoiling a child. All cultures have some areas in which they expect early attainment of autonomy and mature behavior, and other areas in which they are relatively lax and indulgent. When people are looking at families from other cultures, they tend to notice those areas of childrearing where the other culture is either much stricter or more indulgent than their own. However, they are unaware of what aspects of their own culture others tend to find either overly indulgent, or overly strict.

Parental Participation in Learning

The third dimension of parental behavior clusters around promoting and participating in children's learning. It includes all those behaviors variously described in the child development literature as teaching, scaffolding, facilitating, and promoting language and learning. This dimension of parent behavior begins at the child's birth but becomes particularly salient (and conscious to the parent) during the toddler or preschool years. Parents provide for their young child's general learning both at home and outside the home in many direct and indirect ways, and this is sometimes called the curriculum of the home. For example, they promote their children's learning by interacting with them in an attentive and interested way and by providing them ample opportunities to gain new information and encouraging or permitting them to solve their own problems. Parents have many opportunities throughout the day to engage in responsive language and learning interactions with their children and allow the child to learn through complex and constructive play, asking questions, shared book-reading, or involvement in household tasks, and open-ended exploration.

Ample evidence exists for the importance of parents' participation and engagement in their children's early learning. Well before starting school, children interact with materials that are important for the emergence of

literacy. By interacting with all the forms of environmental print, children gain valuable information about print, signs, and books; sounds, letters, words, and sentences; and when they practice the de-contextualized language associated with songs, rhymes, and stories they are enjoyably prepared for later, more formal literacy instruction in school. Parents play a critical role in influencing early-language learning by commenting, mimicking, and expanding on their children's play schemes and speech. The richness of the literacy environment strongly predicts children's language and academic outcomes. Specific interactions during shared storybook-reading, such as labeling pictures, pointing out words and letters, and relating the story to a child's own life are important for learning early literacy rules/conventions and children's later school success (Wood, 1980). Parents also influence children's learning by modeling and supporting simple verbal and written productions such as the alphabet song, nursery rhymes, or how to write their name on a greeting card. Young children who experience reading and writing as pleasurable events are generally more successful later in school. Mothers who use comments and open-ended questions during conversations and shared reading activities, rather than predominantly verbal directives and closed questions, usually have children who develop more advanced vocabulary and language skills (Hart and Risley, 1995). Parents who frequently engage in responsive language and literacy interactions with their children, and who provide a home environment rich in opportunities for learning through shared book-reading, constructive play, and exploration, have children who display higher language and cognitive skills in the preschool and primary years.

Additionally, parents directly and indirectly provide natural learning environments for young children by determining their everyday activities (i.e., mealtimes, interactions with siblings, or outdoor or indoor play) in the settings and community locations frequented by young children of a similar age, culture, and geographic region. Children interact with parents in routine daily activities (e.g., dialing the phone, reading the mail, writing a grocery list), and thereby learn and practice a variety of skills that will serve them well once they start school. The degree to which parents engage their children in these activities (e.g., by providing the opportunity to watch, imitate, practice, or ask questions) influences the amount of information children take from these interactions, and has been found to be associated with optimal developmental changes. Informal activities, such as eating a meal or getting dressed or playing in the park or backyard, are also important natural learning opportunities for children if parents use them to engage with children in positive ways and use language and problem-solving strategies to highlight the experiences and express high, realistic expectations for achievement, and become involved with their child's explorations. Parents' ability to provide such

a curriculum at home has been related to early-childhood language outcomes and literary success and positive academic outcomes.

Finally, parents play an important role in arranging for young children's out-of-home learning opportunities, by how they select a quality childcare or preschool experience for their children and how they engage fully with the children and staff. Assuring language- and literacy-learning opportunities at childcare or preschool or kindergarten involves parents' participation and collaboration with teachers and education professionals. Parents' regular participation in school activities, such as parent-teacher conferences, as well as involvement in class activities, observation visits, and take-home activities (songs, books, etc.) have been linked to young children's later academic success. However, successful home-school collaboration is the responsibility of both parents and education professionals, and effective communication between parents and schools is critical to the successful bridging of home and school learning opportunities for children (Christenson and Sheridan, 2001). A strong connection between schools and families assists children in developing the skills needed to be successful socially and emotionally, as well as academically, and has been shown to be a significant factor in children's overall achievement. When parents are involved in their children's schooling, children show improvements in many adaptive outcomes, including pro-social behavior, self-esteem, perseverance and mastery motivation, and participation in learning activities.

As with warmth and sensitivity, and promoting autonomy, there are many styles by which parents can participate in their children's learning. Parents in different cultural communities have distinctive beliefs about what they believe children should learn, and in a general way, these parental beliefs match the demands of the cultural context. For instance, a physical style of participating is promoted in cultural environments that contain strong physical dangers for young children (drowning, falling, getting burned, getting run over, or getting lost). By using a physical style, parents promote the acquisition of gross and fine motor physical skills that help children learn to move safely and efficiently through the world and to begin to handle and manipulate necessary tools and implements, whether they be spoons, knives, digging tools, or pens and pencils.

The social realm of language and learning is also important to most families, as has been described. Parents promote their child's incorporation of social skills and knowledge by letting them participate in household work and including them in the joyful celebrations and rituals that are most meaningful to the family. They can achieve similar effects by incorporating their youngest children in events that give the whole family pleasure, such as sports events. For infants and toddlers, sitting with the family on the sidelines through long games can be either barren and boring (when they are primarily

pacified with food and drink) or instead rich in learning and literacy experience, when family members take time to draw them out in extended conversation, teach them meaningful routines (e.g., the rudiments of the game), and show them all the numbers, letters, and words on the score boards, food containers, programs, and uniforms.

Today, however, the pre-academic or cognitive side of parental participation in learning has become at least as important as the social and physical sides because of its connection to readiness for school success. Indeed, explicitly symbolic learning that promotes emergent learning in the domains of literacy, math, science, and creative arts, reaches all the way down into the infant and toddler years. Parents set the stage for their babies' later school readiness treating them as conversational partners (echoing and expanding their vocalizations and utterances, e.g., when they say, "You want more milk in your bottle?" after their child says, "Bot-tle"). Likewise, they expand their children's future command of language by modeling and encouraging the pleasure of using words, whether in naming, describing, explaining, rhyming, joking, telling a story, singing, counting, comparing, or computing. Parents also support an early love of language and learning by introducing their children to the cultural arts (by providing drawing and listening materials, or taking the child to a puppet show, library, public garden, swimming pool, or park). Finally, they cultivate a pleasure in reading and future literacy by reading stories to them from infancy and providing a rich supply of books, literacy tools, and imaginative play materials in their home. The pre-academic methods of fostering school readiness are very desirable as part of the curriculum of the home, but they are not everything. Any and all of the parenting styles of promoting language and learning – physical, social, and cognitive – have their own merit and are positive supports to young children's present and future socioemotional and intellectual growth and development.

Acknowledgment

The work was supported by a grant awarded to Dr. Susan Sheridan and Dr. Carolyn Pope Edwards by the Department of Health and Human Services (DHHS) – National Institute of Child Health and Human Development (NICHD), Administration for Children and Families (ACF) and Office of the Assistant Secretary for Planning and Evaluation (ASPE); and the Department of Education (ED) – Office of Special Education and Rehabilitative Services (Grant #1R01H00436135). The opinions expressed herein are those of the investigators and do not reflect the funding agencies. Special appreciation is extended to the families and teachers who participated including the Head Start agency administration for their willingness to cooperate and learn with us throughout the project.

See also: Attention in Cognition and Early Learning; Cognition and Emotion.

Bibliography

- Ainsworth, M. D., Bell, S. M., and Stayton, D. J. (1972). Individual differences in the development of some attachment behaviors. *Merrill-Palmer Quarterly* **18**, 123–143.
- Christenson, S. L. and Sheridan, S. M. (2001). *Schools and Families: Creating Essential Connections for Learning*. New York: Guilford Press.
- Clark, K. E. and Ladd, G. W. (2000). Connectedness and autonomy support in parent-child relationships: Links to children's socioemotional orientation and peer relationships. *Developmental Psychology* **36**, 485–498.
- Edwards, C. P. and Whiting, B. B. (eds.) (2004). *Ngecha: A Kenyan Village in a Time of Rapid Social Change*. Lincoln, NE: University of Nebraska Press.
- Hart, B. and Risley, T. (1995). *Meaningful Differences in the Everyday Experiences of American Children*. Baltimore, MD: Brookes Publishing.
- National Scientific Council on the Developing Child (2004). *Young Children Develop in an Environment of Relationships*. Working Paper No. 1. Retrieved (November 4, 2009) from http://developingchild.harvard.edu/library/reports_and_working_papers/wp1/.
- Sheridan, S. M., Marvin, C., Knoche, L., and Edwards, C. P. (2008). *Getting ready: Promoting school readiness through a relationship-based partnership model*. *Early Childhood Services* **2**(3), 149–172.
- Shonkoff, J. P. and Phillips, D. (2000). *From Neurons to Neighborhoods*. Washington, DC: National Academy Press.
- Sroufe, L. A. (1983). Infant-caregiver attachment and patterns of adaptation in preschool: The roots of maladaptation and competence. In Permuter, M. (ed.) *Educating All Students in the Mainstream of Regular Education*, pp 41–84. Baltimore, MD: Paul Brookes.
- Whiting, B. B. and Whiting, J. W. M. (1975). *Children of Six Cultures: A Psychocultural Analysis*. Cambridge, MA: Harvard University Press.
- Whiting, J. (1994). Environmental constraints on infant care. In Chasdi, E. (ed.) *Culture and Human Development: The Selected Papers of John Whiting*, pp 107–134. New York: Cambridge University Press.
- Wood, D. (1980). Teaching the young child: Some relationships between social interaction, language and thought. In Olson, D. R. (ed.) *The Social Function of Language and Thought*, pp 87–99. New York: Norton.

Further Reading

- Edwards, C. P. and Liu, W. (2002). Parenting toddlers. In Bornstein, M. L. (ed.) *Handbook of Parenting, Second Edition: Volume 1. Children and Parenting*, pp 45–72. Hillsdale, NJ: Erlbaum.
- Gielen, U. P. and Roopnarine, J. L. (eds.) (2004). *Childhood and Adolescence: Cross-Cultural Perspectives and Applications*. Westport, CT: Praeger.
- Harkness, S. and Super, C. (eds.) (1996). *Parents' Cultural Belief Systems: Their Origins, Expressions, and Consequences*. New York: Guilford Press.
- Keller, H. (2007). *Cultures of Infancy*. Mahwah, NJ: Erlbaum.
- Landry, S. (2002). *Pathways to Competence: Encouraging Healthy Social and Emotional Development in Young Children*. Baltimore, MD: Paul H. Brookes.
- LeVine, R. A. and New, R. S. (eds.) (2008). *Anthropology and Child Development: A Cross-Cultural Reader*. Malden, MA: Blackwell.
- Lynch, E. W. and Hanson, M. J. (2004). *Developing Cross-Cultural Competence: A Guide for Working with Children and Families*, 3rd edn. Baltimore, MD: Paul H. Brookes.
- Raikes, H. and Edwards, C. P. (2009). *Extending the Dance: Attachment and Relationship in Infant and Toddler Caregiving*. Baltimore, MD: Paul H. Brookes.
- Ramey, C. T. and Ramey, S. L. (1999). *Right from Birth: Building Your Child's Foundation for Life, Birth to 18 Months*. New York: Goddard Press.
- Rogoff, B. (2003). *The Cultural Nature of Human Development*. New York: Oxford University.
- Whiting, B. B. and Edwards, C. P. (1988). *Children of Different Worlds: The Formation of Socialbehavior*. Cambridge MA: Harvard University Press.

Peer Interaction and Learning

S Blum-Kulka and S Dvir-Gvirsman, Hebrew University, Jerusalem, Israel

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In the broadest sense, situated peer learning can unfold in any human interaction between peers, children, or adults, when different viewpoints emerge against a backdrop of social equality and active participation. In the adult world, it may involve interaction between a lay person and a professional (learning new software on the computer), friends or acquaintances (exchange of recipes), and be either planned or unplanned (college students tutoring in math vs. casual conversation between friends). In a parallel fashion, peer learning in the world of children can take many forms, some institutionalized, as in group work in classrooms and labs, and others in natural settings, as in interaction between siblings in the family or between friends in preschool or school. From the educational perspective adopted here, the discussion of research on peer learning encompasses both the sites of children's natural habitus, as in ethnographic studies of socialization, and those of the more formal settings of schools, as in educational-psychology studies of peer collaboration. Included in this spectrum are children and young people of all ages, from kindergarteners to young adults.

Research on peer learning encompasses a wide variety of interests, including language, social skills, many subject matters at school, and more. Peer learning has been studied both experimentally and through participant observation of naturally occurring interaction; it has been of interest to scholars coming from at least four research traditions: (1) studies in cognitive psychology, focused on cognitive development, that have generated an abundance of experimental studies of the relative benefits of peer learning compared to child-adult learning (for a review see Garton, 2004; O'Donnell and King, 1999); (2) the educational psychology tradition, within which child collaborative learning has been widely adopted for its effectiveness in promoting subject-matter achievement, in math and science (e.g., Schwarz *et al.*, 2000); (3) the cross-cultural psychology approach to socialization, based on which researchers have studied the role of siblings in patterns of child-rearing in different communities around the world (Rogoff, 2003); and (4) the linguistic ethnography tradition, with its focus on the ways in which peer talk functions in children's acquisition of discourse skills and the co-construction of peer culture (e.g., Goodwin, 1990).

The methods used in these studies follow two major methodological groups or some combination thereof. The first group, mainly representing the cognitive approach in

the psychological and educational traditions, is often carried out in classroom settings during the academic year, and is based on classic pretest-posttest experimental design, with clear hypotheses and measurable outcomes (Calhoon and Fuchs, 2003). The second group, representing the sociocultural and language socialization approach, consists of studies in which the basic data consist of transcripts of natural talk among children; data analysis tools draw from both discourse-analytical and quantitative traditions, and the focus of analysis is on process rather than outcome (see Kyratzis, 2004 for review).

Studies of peer learning also vary in terms of the context of the situation focused on, age of participants, and domain studied. Structured contexts, such as classrooms, have been studied much more extensively (for review see Fuchs and Fuchs, 2005) than nonstructured ones, such as peer interaction among siblings in the family or talk among young peers in preschool (for a review see Blum-Kulka and Snow, 2004). Age of participants studied varies from young preschoolers to school-age children (for review see Ginsburg-Block *et al.*, 2006), high school (Schwarz *et al.*, 2000) and college students. Interests shift with age: in studies of young children, the focus is often on language-linked domains, such as second-language literacy and discursive skills (Blum-Kulka and Snow, 2004; Cekaita and Aronsson, 2004), whereas with school age the interest shifts to the learning of subject matter, such as math and science (King *et al.*, 1998), history, and social skills (Ginsburg-Block *et al.*, 2006). The overall conclusion from all the studies is that peer learning is a highly effective mode of learning.

Theoretical Underpinnings

The effectiveness of peer learning can be grounded in at least four theoretical perspectives on learning. From a Piagetian perspective, the learning process at large is triggered by states of cognitive conflict between what is understood and what is encountered in active interaction with and manipulation of the environment. Peer interaction promotes such cognitive conflict by exposing discrepancies between different participants' level of knowledge, resulting in a state of disequilibrium. Through dialog and discussion among individuals of equal status – as in peer interaction between children unhindered by the status and cognitive superiority of adults – a higher level of understanding evolves, cognitive change occurs,

and equilibrium is restored. This is regarded as an internal process, which then manifests itself in behavior (an inside-out theory Garton, 2004). Studies grounded in a Piagetian constructivist framework have largely supported this view by showing experimentally that working with a peer leads to greater cognitive benefit than working alone (Garton, 2004).

From a Vygotskian perspective, social activity at large is important as the primary site in which concepts are formed, before they become mental functions for the individual. Cognitive development depends on active social interaction, including reasoning and explanation, with a more-competent partner. The interaction unfolds presumably within the novice's zone of proximal development, namely with one of the participants entering the interaction with an initial gap in knowledge that can be closed in interaction with the more expert partner. The expert is considered as having responsibility for providing the appropriate support or guidance (scaffolding) to fit the novice's zone of proximal development. In his view, a child may learn more when interacting with a peer at a similar age than when interacting with an adult, due to the fact that they are both operating within one another's proximal zones of development, and can change roles as novice and expert, providing each other with appropriate scaffolding on different occasions. As a result of the process, a shared, intersubjective understanding is co-constructed from the participants' differing viewpoint which in turn is internalized by all participants (an outside-in theory Garton, 2004). The Vygotskian framework has been evoked to explain gains in peer learning in both experimental (e.g., Fawcett and Garton, 2005) and natural settings (e.g., Nicolopoulou and Richner, 2004). In line with the zone of proximal development (ZPD) idea, it has been shown that small differences in cognitive level between children were more effective than large differences (Kuhn, 1972).

More recent theories of situated learning (Lave and Wenger, 1991) take the Vygotskian argument further, claiming that all learning is quintessentially social. In this view, learning is a process of participation in communities of practice, participation that is at first legitimately peripheral but increases gradually in engagement and complexity. Legitimate peripheral participation (LPP) theory alters the locus of learning. It is not the individual mind that acquires mastery, since learning is a process that presumably takes place not in an individual mind, but in a participation framework. The shift in focus toward the socially constituted nature of learning resonates in cognitive theory that views thinking as a socially distributed process, with knowledge emerging from the language-mediated collaborative activity of several participants (Resnick *et al.*, 1991).

A fourth tradition, coming from social psychology, locates the locus of learning in the motivational factors explaining the interaction between the individuals in the

group (Slavin, 1996). Here, the interest in peer learning is restricted to certain forms of cooperative learning in school, and not to peer learning at large. The central process this theory emphasizes is the structure of the reward or goals under which the student operates (Slavin, 1996). The reward is given to the group as a whole, making the individual interest and the group interest as one. This kind of reward structure motivates the group members, not only to maximize their abilities, but also to help others in the group to do so. The most effective structure is attained when the group is rewarded based on the individual learning of its members. This procedure is intended to make group members more accountable, and to that end avoid social loafing. A related conceptualization emphasizes the importance of social cohesion, claiming that the reason the group's interest has been served by its members is not motivational, but rather emotional, building on the members' wish to help each other (Sharan, 1994). Though there is no one-to-one correspondence between theory and research tradition, in general terms, the cognitive approach tends to be associated with Piaget, while the sociocultural approach is affiliated with Vygotskian and situated-learning theories.

Peer Learning in the Cognitive Tradition: Empirical Evidence

In educational psychology, peer learning has been referred to by various terms, such as collaborative learning, cooperative learning, peer-assisted learning, and peer tutoring. It is beyond the scope of this article to enlist the conceptual differences between all these terms, and we shall focus instead on two often-cited definitions. Collaborative learning, in its radical version, "involves learners working together in small groups to develop their own answer through interaction and reaching consensus, not necessarily a known answer" (Ehrlich, 2002). Monitoring the groups or correcting "wrong" impressions is not the role of the trainer since there is no authority on what the answer should be.

Cooperative learning, on the other hand, is geared toward solving problems with right solutions. It has been defined as "classroom techniques in which students work on learning activities in small groups and receive rewards or recognition based on their group performance" (Slavin, 1980: 315). Three distinct types have been identified – information learning, formal learning, and study teams – varying in terms of the length of the task required to be done together, a parameter that affects the degree of involvement and relationship within the learning group (see Nastasi and Clements, 1991 for review).

The majority of the empirical studies conducted in education and psychology used a strict pre-post experimental

model, mostly in the field (though some were held in laboratories), with a very similar structure:

1. Children who are a representative sample of the population (that can be different from one study to the other in age, learning abilities, and so on) were randomly assigned to the control or experimental group; sometimes, this process is dependent on the child's scores in school.
2. A measure of the specific educational field, usually by standardized tests.
3. Applying the intervention that varied from one study to the other in a given length of time, the level of structuring in the assignment the dyads/groups receive, differences in children's academic level, and so on. The condition in the control group is also varied: in some studies, the children in the control group carry on with the original educational program of the school; in others, there is a different intervention, such as studying in small groups with close supervision/guidance of the teacher, which has been compared to peer learning.
4. Testing the children again in the same standardized test.
5. Testing to see a significant statistical improvement following the intervention, in comparison to the control group.

This approach, which uses a very specific outcome and a very specific definition for the effect of peer learning, has advantages as well as disadvantages. On the one hand, it is easy to establish the child's gains from peer learning and to identify improvement and the necessary conditions for it. On the other hand, we may be missing data that will help explain the ways in which peer learning works. Research in this tradition has become increasingly sensitive to the role of verbal communication during peer learning, either by including a +/-talk condition in the experimental design (Fawcett and Garton, 2005), or more frequently, by assessing the quality of discourse during peer interaction (see Nastasi and Clements, 1991 for review).

The effectiveness of peer learning has been demonstrated in many curriculum areas as well as in social and communication skills. In the child's early years, the main focus in the literature revolves around reading and writing abilities. Studies conducted as early as in kindergarten found that peer-assisted learning strategies (PALSs) have a beneficial effect on reading vocabulary and comprehension, decoding skills, and writing (Fuchs and Fuchs, 2005). These findings were significant over and above the impact of the child's socioeconomic background and his initial level (high vs. low).

The effects of peer learning on mathematical skills were assessed for both elementary and high school students. Overall, peer learning was found to have a

positive effect on the child's mathematical abilities. Studies conducted in elementary school showed that peer learning helped students generate conceptual mathematical explanations and helped high-achieving students' performances on complex mathematical tasks (Fuchs and Fuchs, 2005) and improved students' grades in both curriculum-specific tests and standardized tests (Slavin, 1980).

Studies on elementary schools have also shown that interventions that focus on academics can also improve social skills. PALS encourages making new friends among children who are less popular (Fuchs and Fuchs, 2005), promoting the social acceptance of children with handicaps (see Nastasi and Clements, 1991 for review) or intellectual disabilities (Yarrow and Topping, 2001), and improving social and self-concept. Those, in turn, are correlated with high academic scores (Ginsburg-Block *et al.*, 2006).

The number of studies conducted on high school students is significantly lower than the studies done in elementary school. Obviously, there is not much work on reading and writing. The main focus in high school is mathematical skills and social skills. Peer learning increases mathematical reasoning and understanding. The positive effect of peer learning for mathematics has been attributed to the peers' inquiry and raising of questions (King *et al.*, 1998), discursive practices which require the student to explain and defend his opinion (Schwarz *et al.*, 2000).

Peer learning also improves the adolescent attitude toward school, the dropout rate, and their self-concept (Raswal *et al.*, 1995). More recent studies are trying to assess the interaction between the learning method and the student's basic needs and personality (Hänze and Berger, 2007).

Peer learning has also been convincingly shown to be beneficial for special education students. In this case, studied extensively, peer learning proved to be beneficial from kindergarten to high school. It improved the reading abilities of the child, mathematical abilities (Calhoon and Fuchs, 2003), and social skills (Yarrow and Topping, 2001). These findings are valid for children with emotional and behavior disorders as well (Ryan *et al.*, 2004).

Peer Learning in Sociocultural Tradition: Empirical Evidence

One of the basic tenets of sociocultural theory is that children are active agents in their own socialization, learning language and culture through participation in linguistically marked meaningful events. The perspective on learning in this approach is conceptualized as becoming competent participants in communities of practice, as entering new worlds of discourse, rather than as just acquiring new skills and concepts. From this viewpoint, peer interaction is the essential condition for peer learning. Peer (and sibling) interaction is viewed as crucial for the

development of cognitive and linguistic skills, including pragmatic skills. As in all ethnographic and cultural approaches to development, the focus of the observation is on the process of the interaction in its potential contributions in the wider sociocultural context, rather than on measurable gains. Arguably, the outcomes of children's interactions need to be viewed in terms of their contributions to children's co-construction of their social identities and peer culture, as well as in terms of their cognitive and linguistic gains. Two major contextual features of peer interaction can be considered conducive to learning in this site: (1) its collaborative, multiparty, symmetrical participation structure and (2) its shared worlds of childhood culture. Peer interaction unfolds unhindered by the inherent asymmetry of adult-child interactions and hence presumably can promote a free and reciprocal exchange of ideas; the specifics of peer-group participant structure allows for reciprocity in expert-novice configurations with shared trajectories of ZPDs, inviting momentary teaching-learning opportunities in different domains, ranging from the use of computers to language learning (Blum-Kulka and Snow, 2004; Rogoff, 2003).

Evidence in this approach for the essential role peer interaction plays in socialization comes from three domains of study: (1) cross-cultural, ethnographic studies of socialization across cultures; (2) second-language learning studies; and (3) linguistic ethnography studies.

For many centuries and even in recent times and in diverse cultures, children have been acting as the major agents of socialization. In these societies, the care of toddlers is entrusted to their older siblings, as is the case for Polynesian children, Mayan children in Mexico, or African-American children in Louisiana, offering them ample opportunities to learn varied aspects of social behavior – from how to help with household chores to how to manage conflicts and compromises. Simultaneously, older siblings are young children's primary source of linguistic input, and have also been shown to engage in direct language teaching, as in the case of African-American children in Louisiana who taught their younger siblings to recite the alphabet, play word games, and name colors and numbers; the African-American girls in Piedmont who engaged younger children in word play, counting and naming body parts; and the American-Israeli bilingual teenagers who engaged their younger siblings in bilingual word play (Blum-Kulka and Snow, 2004; Heath, 1983; Rogoff, 2003).

Peer interaction is essential for second-language learning. For immigrant children, in many countries whose home language is different from school language, school experience is the major resource for learning the target language. Depending on the type and quality of educational support given, interaction with native peers may prove crucial in the process. Once immigrant children have acquired the rudiments of communication in the second language, their native peers willingly engage them in

talk and play, enhancing indirect language learning, and also act as explicit language teachers, correcting pronunciation and teaching vocabulary (e.g., Tabors *et al.*, 2000). From a theoretical viewpoint, the case of immigrant children can be considered a prime example for learning through legitimate peripheral participation; from an initial silent position at the periphery of their peers' participation structure, they move with time (and confidence in second language) to the center of this structure, aided in the process by their native peers. Second-language learners in immersion classrooms with limited second-language proficiency recurrently employ language play in conversations with peers, a speech genre which generates extended repair sequences serving as informal language lessons (Cekaita and Aronsson, 2004). Studies that show scaffolded help provided by peers include Japanese and American immersion preschool and kindergarten classes, British bilingual classrooms, and English as second language (ESL) classrooms (Kasper and Rose, 2002).

Linguistic ethnography studies of children's peer talk reveal that deep involvement in shared worlds of childhood culture creates interest and motivation, facilitating peer interaction and presenting rich opportunities for the mutual learning of a wide range of cognitive, social, and discursive skills, such as: the intersubjective skills of following collaboratively constructed complicated plots of pretend play (Sawyer, 1997); negotiating power asymmetries and gender identities in same-sex and cross-sex groups (Goodwin, 1990); practicing and refining their narrative skills, and developing child peer culture through narration (Maybin, 2006; Nicolopoulou and Richner, 2004); fostering conversational skills (Hamo and Blum-Kulka, 2007); and developing cultural styles of negotiation (Corsaro and Rizzo, 1990).

Summary

Peer learning emerges from this article both as an educational framework for a variety of educational practices (tutoring, collaborative learning, cooperative learning, etc.), as well as an interactional framework fostering learning in many forms of life. Theoretical interest in the field has a long and well-established history, going back to the two pillars of developmental psychology, Piaget and Vygotsky. The educational domains where peer learning has been shown to be effective include school subjects like mathematics, history, and literature, as well as social skills and self-confidence. Yet, research shows that the degree of success achievable with peer learning can depend on a host of contextual features, including age, gender composition, level of group heterogeneity, degree of task structuring, quality of interaction, and type of reward (Ginsburg-Block *et al.*, 2006; Jacob, 1999; Slavin, 1980). The gains and affordances of PL in natural settings are manifest in sibling care in traditional societies, second-language acquisition by

young children, and child-to-child talk in various contexts. Traditionally, in the cognitive tradition, gains are evident mainly by measurable outcomes; in the sociocultural tradition, gains are evident by close scrutiny of the interactional processes. Yet, research on peer learning has also seen a growing cross-over from sociocultural to cognitive approaches in fields traditionally associated with the latter, like mathematics education; embracing discursive approaches to research on math education entails replacing the language of personal acquisition by talk about learning as becoming participants in a collective activity, accompanied by detailed descriptions of the learning process backed by transcripts of student interactions (Kieran *et al.*, 2003). Arguably, the common denominator in all cases considered is communication: peer learning, whether planned or unplanned, institutionalized or natural, unfolds mostly through talk. It follows that the quality of talk is an essential component of its potential success. Indeed, several educational studies have shown that engagement in verbal pre-planning, active debate, explicit negotiation, listening to each other's explanations and arguments, and reflection on their logical consistency while working on tasks with peers is positively associated with individual cognitive gains on posttests (e.g., Barbieri and Light, 1992). The role of verbal active exchange of ideas as a key (if not the key) element of effective peer learning has been recognized by at least some researchers in the cognitive tradition, and has led to a rising emphasis on transcripts as a major source of data, and to some application of discourse-analytical methods to pinpoint the discursive strategies fostering learning (e.g., Fawcett and Garton, 2005; Jacob, 1999; Asterhan and Schwarz, 2009). Attention to language and discursive practices as primary to the inquiry of issues of socialization and development has been essential to most research in the sociocultural tradition; for some, like in the study of argumentation, the nature of the argumentative dialog is taken to be the very essence of the learning process (Pontecorvo and Girardet, 1993); for others, peer talk is conceived of as providing affordances for learning in a broad spectrum of domains and skills, leading to cultural, social, linguistic, and cognitive gains.

See *also*: Attention in Cognition and Early Learning; Cognition and Emotion.

Bibliography

- Asterhan, C. S. C. and Schwarz, B.B. (2009). Indications from protocol analyses of peer-to-peer Dialog. *Cognitive Science* **33**(3), 374–400.
- Barbieri, M. S. and Light, P. (1992). Instruction, gender, and performance on a computer-based problem solving task. *Learning and Instruction* **2**(3), 199–213.
- Blum-Kulka, S. and Snow, C. (2004). Introduction: The potential of peer talk. *Discourse Studies* **6**(3), 291–307.
- Calhoon, M. B. and Fuchs, L. S. (2003). The effects of peer-assisted learning strategies and curriculum-based measure

- on the mathematic performance of the secondary students with disabilities. *Remedial and Special Education* **24**(4), 235–245.
- Cekaita, A. and Aronsson, K. (2004). Repetition and joking in children's second language conversations: Playful recycling in immersion classrooms. *Discourse Studies* **6**(3), 373–393.
- Corsaro, B. and Rizzo, T. (1990). Disputes in the peer culture of American and Italian nursery-school children. In Grimshaw, A. (ed.) *Conflict Talk*, pp 21–67. Cambridge: Cambridge University Press.
- Ehrlich, D. (2002). *HRD 408: Instructional Design II*. <http://www.neiu.edu/~dbehrlic/hrd408> (accessed Jun. 2009).
- Fawcett, L. M. and Garton, A. F. (2005). The effect of peer collaboration on children's problem-solving ability. *British Journal of Educational Psychology* **75**, 157–169.
- Fuchs, D. and Fuchs, L. S. (2005). Peer-assisted learning strategies: Promoting word recognition, fluency, and reading comprehension in young children. *Journal Special Education* **39**(1), 34–44.
- Garton, A. F. (2004). *Exploring Cognitive Development: The Child as Problem Solver*. Oxford: Blackwell Publishing.
- Ginsburg-Block, M. D., Rohrbach, C. A., and Fantuzzo, J. W. (2006). A meta-analytic review of social, self-concept, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology* **98**(4), 732–749.
- Goodwin, M. H. (1990). *He Said She Said: Talk as Social Organization among Black Children*. Bloomington, IN: Indiana University Press.
- Hänze, M. and Berger, R. (2007). Cooperative learning, motivational effects, and student characteristics: An experimental study comparing cooperative learning and direct instruction in 12th grade physics classes. *Learning and Instruction* **17**, 29–41.
- Hamo, M. and Blum-Kulka, S. (2007). Apprenticeship in conversation and culture: Emerging sociability in preschool peer talk. In Valsiner, J. and Rosa, A. (eds.) *The Cambridge Handbook of Social-Cultural Psychology*, pp 421–444. Cambridge: Cambridge University Press.
- Heath, S. B. (1983). *Ways with Words: Language, Life and Work in Communities and Classrooms*. Cambridge: Cambridge University Press.
- Jacob, E. (1999). *Cooperative Learning in Context: An Educational Innovation in Everyday*. Albany, NY: State University of New York Press.
- Kasper, G. and Rose, K. A. (2002). *Pragmatic Development in a Second Language*. Oxford: Blackwell Publishing.
- Kieran, C., Forman, E., and Stard, A. (eds.) (2003). *Learning Discourse: Discursive Approaches to Research in Mathematics Education*. Dordrecht: Kluwer Academic Press.
- King, A., Staffieri, A., and Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology* **90**(1), 134–152.
- Kuhn, D. (1972). Mechanism of change in the development of cognitive structures. *Child Development* **43**, 833–844.
- Kyratzis, G. (2004). Talk and interaction among children and the co-construction of peer groups and peer culture. *Annual Review of Anthropology* **33**, 625–649.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Maybin, J. (2006). *Children's Voices: Talk, Knowledge and Identity*. Houndmills: Palgrave Macmillan.
- Nastasi, B. K. and Clements, D. H. (1991). Research on cooperative learning: Implications for practice. *School Psychology Review* **20**(1), 110–132.
- Nicolopoulou, A. and Richner, E. S. (2004). When your powers combine, I am Captain Planet: The developmental significance of individual and group authored stories by preschoolers. *Discourse Studies* **6**(3), 329–347.
- O'Donnell, A. M. and King, A. (eds.) (1999). *Cognitive Perspectives on Peer Learning*. Mahwah, NJ: Erlbaum.
- Pontecorvo, C. and Girardet, H. (1993). Arguing and reasoning in understanding historical topics. *Cognition and Instruction* **11**, 365–395.
- Raswal, G. M., Mims, A. A., Croce, R., *et al.* (1995). Effect of collaborative peer tutoring on urban seventh grades. *Journal of Educational Research* **88**(5), 275–279.

- Resnick, L., Levine, J., and Teasley, S. (1991). *Perspectives on Socially Shared Cognition*. Washington, DC: American Psychological Association.
- Rogoff, B. (2003). *The Cultural Nature of Human Development*. Oxford: Oxford University Press.
- Ryan, J. B., Reid, R., and Hepstein, M. L. (2004). Peer-mediated intervention studies on academic achievement for students with EBD: A review. *Remedial and Special Education* **25**(6), 330–341.
- Sawyer, K. (1997). *Pretend Play as Improvisation: Conversation in the Preschool Classroom*. Mahwah, NJ: Erlbaum.
- Schwarz, B. B., Neuman, Y., and Biezuner, S. (2000). Two wrongs may make a right . . . If they argue together! *Cognition and Instruction* **18**(4), 461–494.
- Sharan, S. (ed.) (1994). *Handbook of Cooperative Learning Methods*. Westport, CT: Greenwood Press.
- Slavin, R. E. (1980). Cooperative learning. *Review of Educational Research* **50**(2), 315–342.
- Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology* **21**, 43–69.
- Tabors, P., Aceves, C., Bartolom, L., Páez, M., and Wolf, A. (2000). Language development of linguistically diverse children in Head Start classrooms: Three ethnographic portraits. *NHSA Dialog* **3**(3), 409–440.
- Yarrow, F. and Topping, K. J. (2001). Collaborative writing: The effect of metacognitive prompting and structured peer interaction. *British Journal of Educational Psychology* **71**(2), 261–282.
- Blum-Kulka, S. (2005). Modes of meaning making in young children's conversational narratives. In Thornborrow, J. and Coates, J. (eds.) *The Sociolinguistics of Narrative*, pp 149–171. Antwerpen: John Benjamins.
- Blum-Kulka, S. and Snow, C. E. (eds.) (2004). *Special Issue: Peer Talk and Pragmatic Development*. *Discourse Studies* **6**(3), 291–307.
- Corsaro, W. A. (1985). *Friendship and Peer Culture in the Early Years*. Norwood, NJ: Ablex.
- Goodwin, M. H. (2006). *The Hidden Life of Girls: Games of Stance, Status Building and Power Asymmetries in Girls' Interaction*. Oxford: Blackwell Publishing.
- Gross-Davis, B. (1993). *Tools for Teaching*. San Francisco, CA: Jossey-Bass Publishers.
- Handel, G. (1988). *Childhood Socialization*. New York: Aldine de Gruyter.
- Hoyle, S. M. and Temple, A. C. (eds.) (1998). *Kids Talk*. New York: Oxford University Press.
- Johnson, D. W., Johnson, R. T., and Smith, K. A. (1991). *Cooperative Learning: Increasing College Faculty Instructional Productivity*. Washington, DC: School of Education and Human Development, George Washington University.
- Kanagy, R. (1999). Interactional routines as a mechanism for L2 acquisition and socialization in an immersion context. *Journal of Pragmatics* **31**, 1467–1492.
- McTear, M. (1985). *Children's Conversation*. Oxford: Blackwell.
- Riese, H. (2007). Learning outcomes in peer learning practices. A meta-synthesis of qualitative studies. *Paper Presented at the 12th EARLI Conference, Budapest*.
- Rogoff, B. (1990). *Apprenticeship in Thinking: Cognitive Development in Social Context*. New York: Oxford University Press.
- Sfard, A. and Kieran, C. (2001). Cognition as communication: Rethinking learning-by-talking through multi-faceted analysis of students' mathematical interactions. *Mind, Culture, and Activity* **8**(1), 42–76.
- Topping, K. J. (2005). Trends in peer learning. *Educational Psychology* **25**(6), 631–645.

Further Reading

- Aukrust, V. (2001). Talk-focused talk in preschools – culturally formed socialization for talk? *First Language* **21**, 57–82.

LEARNING AND COGNITION – PLACES OF LEARNING

Contents

Learning in a Cross-Cultural Perspective

Learning Outside of School

Reasoning and Explanatory Talk: Learning in Everyday Settings

Learning in a Cross-Cultural Perspective

L-F Zhang, The University of Hong Kong, Hong Kong

R J Sternberg, Tufts University, Boston, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Capital – In sociology, the word “capital” is a general term for four specific forms of capital: cultural, economic, human, and social capital. *Cultural capital* refers to the knowledge and experience that individuals gain chiefly through participating in cultural events. *Economic capital* essentially refers to wealth. *Human capital* refers to the knowledge and skills that individuals gain through education and experience. *Social capital* refers to the tangible benefits and resources that individuals build up because of their inclusion in social networks.

Field dependence/independence – This concept is also known as psychological differentiation and perceptual style. Individuals who are more *field-dependent* are less able to view objects separate from the overall environment. They are likely to be influenced by the prevailing context. Individuals who are more *field-independent* are good at identifying objects that have surroundings that might obscure their view. They tend to see things as discrete from their backgrounds.

Intellectual styles – This term is the general one for concepts such as cognitive styles, learning styles, teaching styles, and thinking styles. All these styles address people’s preferred ways of processing information.

Learning approaches – Ways of learning a task that are contingent upon the student’s intentions. Three common learning approaches are: *surface*, which involves meeting the minimum requirements by reproducing what is taught; *deep*, which involves a true understanding of what is learned; and *achieving*, which involves using a strategy that would maximize one’s academic grades.

Self-construal – A specific dimension of self-definition. Broadly speaking, there are two types of self-construal: interdependent and independent. Individuals with an *interdependent self-construal* have a tendency to view themselves in terms of their close relationships with others. Individuals with an *independent self-construal* are likely to perceive themselves as being autonomous and separated from others.

Thinking styles – This refers to people’s preferred ways of using the abilities that they have.

For nearly a century, researchers have sought alternative ways to account for learning gaps among various cultural groups. As a result of increasing globalization, understanding learning in a cross-cultural perspective has become more important than ever before. There is no such thing as a context-and/or meaning-free environment (LeVine *et al.*, 1988). We learn and make sense of our learning experiences within cultural contexts. Indeed, a substantial amount of literature concerning cross-cultural learning is converging on one conclusion – culture has a great impact on student learning. However, the major findings in this research have created several paradoxes. What could have happened? What are some of the major explanations regarding differential learning across cultures? What can scholars do to move the field forward?

This article aims at addressing the above questions in three parts. First, we describe important works concerning the learning process, focusing on intellectual styles. Second, we introduce the major works on learning outcome, centering on academic achievement. Finally, we draw some conclusions and suggest future research directions.

Intellectual Styles as Learning Processes

What are Intellectual Styles?

Intellectual styles, an encompassing term for such constructs as cognitive styles, learning styles, and thinking styles, refer to people's preferred ways of processing information and dealing with tasks (Zhang and Sternberg, 2006). The field of intellectual styles has a long history, dating back to Allport's (1937) study of personality types. One of the major reasons that theorists and researchers have maintained their interest in styles is that styles play an important role in the process of learning. They also account for individual differences in learning outcomes. Research suggests that intellectual styles are malleable to some degree, depending on many factors. One is culture. Sternberg (1997) proposed that culture is the foremost variable that helps shape people's styles and argued that some cultures are likely to be more rewarding of particular styles than of others.

Intellectual Styles and Culture

The impact of culture on intellectual styles is indisputable. However, very little concerted effort has been made to systematically investigate cross-cultural differences (or similarities) in intellectual styles. Most existing research examining intellectual styles in relation to culture has focused on testing the relationships between intellectual styles and academic achievement in different cultures, rather than on making direct comparisons of intellectual styles among students from different cultural settings. Direct comparisons may promote stereotypical views about certain cultural groups.

The studies documented in the literature have identified a certain degree of universality of intellectual styles across cultures as well as style differences based on culture (e.g., Berry, 1966). They have also presented challenges to some of the stereotypical views about particular cultural groups. These challenges pertain to claims that: (1) Black and Hispanic (including Mexicans, Puerto Ricans, Cubans, and people of other Hispanic origin) students have a higher preference for the surface approach to learning (see Watkins and Mboya, 1997); (2) Asian students are rote learners; (3) Black and Hispanic students are more field dependent than are Anglo-American students; (4) students from group-oriented societies tend to be more field dependent than students from individualistically oriented societies (see Bagley and Mallick, 1998); and (5) students from group-oriented cultures are more norm-conforming in their thinking styles (see Sternberg, 1997). The challenges to the first two claims (1 and 2) are the result of research based on Biggs' (1978) construct of learning approaches. The challenges to the next two claims (3 and 4) are

derived from research grounded in Witkin's (1962) concept of field-dependence/independence. The challenge to the final claim (5) is manifested through research based on Sternberg's (1997) theory of thinking styles.

Learning approaches

One group of cross-cultural studies of students' intellectual styles has been based on Biggs' (1978) three approaches to learning: surface, which involves meeting the minimum requirements by reproducing what is taught; deep, which involves a real understanding of what is learned; and achieving, which involves using a strategy that will maximize one's grades.

However, with regard to the first two claims above (1 and 2), recent studies have found no evidence to support them. Comparative studies on the learning approaches of Black (and Hispanic) and White students in several cultures (e.g., Watkins and Mboya, 1997) have failed to reveal any significant differences in learning approaches. Similarly, the results from research comparing the learning approaches of Asian students with those of Caucasian students are at odds with the stereotypical view that Asian students are rote learners. Traditionally, Asian students' higher academic achievement has been attributed to the value they place on hard work mainly characterized by rote learning, among other factors. However, such a view has not been supported. On the contrary, there is evidence indicating that White American students show a greater tendency to rely on the surface approach to learning than do their Asian peers.

Field dependence/independence

Field dependence/independence (FDI), alternatively known as psychological differentiation and perceptual style (Witkin, 1962), refers to the extent to which people are dependent on versus independent of the organization of the surrounding perceptual field. Field-independent individuals are thought to be better at cognitive restructuring because of their propensity for being free from external referents. Field-dependent individuals are considered to be more socially oriented as a result of their higher levels of sensitivity to external referents. Based on the definition of field dependence and of field independence, one would expect that students from individualistically oriented societies in which people enjoy social looseness (Pelto, 1968) would be more field independent and that students from group-oriented cultures would be more field dependent. Indeed, this hypothesis has been supported by numerous empirical studies.

At the same time, there is also evidence at variance with this hypothesis. For example, Buriel's (1978) comparative study found no difference in the level of field dependence/independence between Anglo-American and Mexican-American school children. In a review of the literature,

Bagley and Mallick (1998) concluded that, compared with their Western counterparts (in this case, students from Canada, the UK, and the US), students from group-oriented cultures (in this case, Chinese and Japanese cultures) were more field independent.

Thinking styles

A relatively recent theory of intellectual styles, Sternberg's (1997) theory of mental self-government, states that there are 13 different ways in which people can use their abilities (i.e., thinking styles). Zhang and Sternberg (2005) classified these styles into three types. Type I thinking styles tend to be more creativity generating and denote higher levels of cognitive complexity (e.g., the global style – the tendency to focus on the holistic picture; the liberal style – the propensity for adopting a new approach to tasks). Type II thinking styles denote a norm-favoring tendency and tend to require lower levels of cognitive complexity (e.g., the local style – the preference for focusing on details; the executive style – the inclination to implement tasks with given guidelines). Type III styles may manifest the characteristics of styles from either Type I or Type II groups, depending on the stylistic demands of a specific task (e.g., internal style – preferring to work on one's own; external style – preferring to work in groups).

One might predict that the internal style and Type I styles would be more encouraged in individualistically oriented cultures and that the external and Type II styles would be more beneficial in group – or collectively oriented cultures. However, this appears to be true in some cases, but not in others. For example, Type I styles have been found to be related to higher academic achievement among American students (e.g., Grigorenko and Sternberg, 1997), whereas Type II styles have been strongly associated with better achievement in collectively oriented cultures such as Hong Kong and the Philippines. However, contrary to the prediction, better academic achievement has been found to be related to Type I styles among university students in Korea, a highly collectively oriented society, whereas better achievement has been associated with Type II styles in Spain, a more individualistic society. Finally, higher academic achievement has been found to be related to the internal style in all cultures investigated so far, be they individualistically oriented or group oriented.

Academic Achievement as Learning Outcome

The two major bodies of literature on academic achievement as learning outcomes concern Asian and African students.

Research Centered on Asian Students

In the earliest and most comprehensive study of cross-cultural learning among 625 000 students from six ethnic groups (Coleman, 1966), Asian Americans emerged as the only ethnic minority group that performed as well as or outperformed Caucasian students on many indicators of intellectual value, aspiration, and achievement. The superior achievement of Asian students in the international arena has been maintained to this day (Trends in International Mathematics and Science Study, 2008).

This finding is puzzling for several reasons. Traditional educational and psychological theories assume that better academic performance should be associated with higher levels of self-efficacy and with a stronger sense of individuality; yet, the Asian students who outperform their Western peers do so despite their expressed lower levels of self-efficacy (e.g., Eaton and Dembo, 1997). Likewise, although psychologists generally maintain that higher academic achievement tends to be positively correlated with attributions to ability, not effort, Asian students tend to attribute their academic success to effort. Anthropologists and sociologists believe that school performance has a great deal to do with wealth and socioeconomic status. However, Southeast Asian refugee students thrive academically despite their low socioeconomic status. Such paradoxes require alternative explanations of Asian students' high academic performance.

Consider three explanations of the paradoxes. The first is from the perspective of cultural values; the second uses Markus and Kitayama's (1991) concept of self-construal; and the third is based on Sue and Okazaki's (1990) notion of relative functionalism.

Cultural values

Asian students' superior achievement is believed to be largely the outcome of the strong influence of Confucian philosophy, which emphasizes authoritarian moralism, collectivism, effort, endurance, filial piety, modesty, passion for learning, and self-improvement. Many scholars have argued that, as a result of such traditional values, students of Asian origin tend to be more respectful of their elders. Asian parents have been/are known for having higher expectations of and making greater sacrifices for their children's learning. In return, Asian students feel a strong obligation to repay their parents' devotion and sacrifice with good academic performance, which is normally perceived to be one of the best forms of filial piety. Empirical evidence also suggests that Asian people often show greater passion for learning than Westerners, who usually put more emphasis on fun, self-esteem, and social development (e.g., Chao, 1996).

Cultural values have also been used to address the paradox that Asian students outperform their Western peers despite attributing their superior performance to

effort rather than to ability (as would be expected by attribution theories), and while reporting lower levels of self-efficacy. Eaton and Dembo (1997) have suggested that modesty is deeply rooted in Confucianism. When Asian students do well academically, instead of boasting how naturally talented they are, they express their need to work harder and do better.

Modest people may tend to be more aware of their inadequacies and, as a result, feel the need for self-improvement through being engaged in their work more diligently. On a similar note, other researchers have proposed that Asian students may score lower on tests of self-efficacy because their Western counterparts overestimate their own abilities by evaluating themselves based on their underachieving peers or on their parents' lower expectations. Western students' higher levels of self-efficacy could also be attributed to the Western view that students need to feel good about themselves and about their achievements. For example, in both the United Kingdom and the United States, it is commonly accepted that having positive self-regard is critical for students to strive to achieve in all aspects of life.

Markus and Kitayama's concept of self-construal

Recently, several authors have borrowed Markus and Kitayama's (1991) concept of self-construal to resolve the paradox of high Asian achievement. According to Markus and Kitayama, the way in which achievement motivation drives goal-directed behavior is culture specific. Westerners tend to perceive goals as an expression of individual wishes. They consider themselves as independent selves, in contrast to Asians, who tend to view goals as an expression of their family or other communal objectives. Asians are more likely to consider themselves as interdependent selves. Moreover, Westerners tend to perceive individual action as the means of achieving personal goals, whereas Asians are likely to perceive the individual as being reflective of past and future generations. Such a strong sense of responsibility, not only for one's own success, but also for glorifying one's own family and community, has been found to be an important factor in Asian students' motivation to work harder and to find ways to fit their interests with the expectations of others.

Sue and Okazaki's relative functionalism

Sue and Okazaki (1990) have argued that in explaining the differential achievement between students of immigrant cultures and students in the host cultures, one should take into account the economic and social status of the immigrant groups in the host cultures, as well as the immigrant groups' traditional cultural values. The authors put forward their theory of relative functionalism to accommodate both cultural and social factors. They believe that immigrants make great efforts to take advantage of the opportunities (such as for education) not available in

their motherlands, with the ultimate goal of upward social mobility via education. The authors have further argued that education can be the principal vehicle for upward social mobility when other means are impossible, which is often true for the Southeast Asian refugees in Western cultures. Consequently, superior academic achievement is a principal route to achieving higher social status.

Research Centered on Black and/or Hispanic Students

Most research on the academic achievement of minority students has used the cross-race comparative paradigm. Typically, Blacks and/or Hispanics are compared with their Caucasian counterparts and with their Asian peers studying in industrial countries. Such studies have usually obtained largely the same finding – Black and/or Hispanic students' academic performance lagged behind that of their Caucasian and Asian peers. Black and Hispanic students underperform on many indicators of academic achievement (DeBlassie and DeBlassie, 1996). In early research (especially between the 1960s and the early 1970s), scholars turned their attention to the notion of compensatory education, viewing minority students' academic failure as the result of some kind of culturally or racially related deficit, most noticeably, poor home preparation for school experience (Deutsch, 1964). The essence of this viewpoint is best illustrated by Crow *et al.* (1966), who listed 60 alleged psycho-educational deficits of the culturally disadvantaged child.

Yet, this culturally/racially related deficit model cannot explain the academic success of some of the minority students and has given rise to several competing explanatory models. Below, we first discuss an explanatory model that is oriented toward racial/cultural differences – that of social class, wealth, and capital. We then discuss two alternative models that take within-culture differences into account, specifically, (1) types of minorities and (2) school structure.

Social class, wealth, and capital

Social class and wealth affect school achievement (e.g., Adler, 1968; Davis, 1948). Sociologists have integrated the concepts of social class and wealth with Bourdieu's theory of capital in explaining the achievement gap between Blacks/Hispanics and Whites. According to Bourdieu, there are four forms of capital: cultural, economic, human, and social. Although varying in their degrees of liquidity (or, the levels of ease by which they can be converted), the four forms of capital are convertible. Economic capital (i.e., wealth) is the most convertible and the most important for school achievement. Wealth, as economic capital, can enable parents to provide children with many important educational resources, such as books, computers, and access to school that is of higher quality.

Wealth can also be transformed into cultural capital – broadly defined as “instruments for the appropriation of symbolic wealth socially designated as worthy of being sought and possessed” (Bourdieu, 1977: 488). Cultural capital can be represented by activities such as attendance at art performances or exhibitions, symphony concerts, and other types of live shows. Activities such as these that require great financial expenses can enhance school achievement (e.g., Hattie *et al.*, 1997).

Wealth can also affect human capital, which in turn, affects children’s school achievement. Wealthy parents may be more cognizant of the role of self-confidence in school achievement than are less wealthy parents. Such awareness may help wealthy parents to pay more attention to adopting strategies that boost their children’s self-confidence.

Finally, wealth can also be transformed to social capital, which in turn affects academic achievement. Social capital is defined as the concrete benefits and resources that people build up by virtue of their inclusion in a social structure/network. Students from wealthy families have the opportunities to interact with people of greater social capital, who serve as role models for success. Seeing these role models can be very conducive to the positive development of learning motivation, and thus to the enhancement of academic achievement.

As Blacks/Hispanics are substantially less wealthy, on average, than are Whites, wealth, as manifested in different forms of capital, can be used to explain at least a portion of the racial-achievement gap. However, using social class, wealth, and capital to account for the racial-achievement gap can also be perceived as using a race-related-deficit approach, for it stresses the disadvantages associated with lack of wealth and capital commonly experienced by Blacks and Hispanics. Furthermore, it cannot explain the success of some Black and Hispanic students.

Ogbu’s two types of minorities

Ogbu (1978) differentiated between two types of minorities in accordance with their initial terms of incorporation and treatment by the majority in the host culture: voluntary minorities and involuntary minorities. Voluntary minorities are those who choose to go to the host country because of perceived opportunities for economic advancement and political freedom. Involuntary minorities refer to those who were either originally brought to the host country in the form of human labor or were subdued through colonization.

A major difference between the two minority groups is what Ogbu called cultural frame of reference. Voluntary minorities tend to hold a positive frame of reference, interpreting various obstacles they are confronted with in their host countries as temporary problems that can be solved

through hard work and education. In contrast, involuntary minorities tend to hold an oppositional cultural frame of reference, distrusting dominants and their institutions and perceiving them as the gatekeepers obstructing them from channels to success and social mobility.

Ogbu suggested that children from voluntary minority groups tend to achieve better than those from the involuntary groups because the parents of the voluntary minority children tend to adopt schooling strategies that are conducive to the social adjustment and academic success of their children. By contrast, the oppositional cultural frame of reference held by the involuntary minorities has promoted an anti-educational orientation that is counterproductive to academic success.

Ogbu’s model accounts for some of the differences in achievement across minority groups. However, the model has been criticized on the grounds that it does not explain why some students fail whereas other students succeed even though they are from the same minority group, be it voluntary or involuntary.

Institutional practice

Explaining the achievement gap between Blacks/Hispanics and Whites from the perspective of institutional practice is a relatively recent endeavor in response to Ogbu’s (1978) argument that involuntary minorities tend to take an oppositional stance toward school and that Black/Hispanic high achievers tended to act White. According to this alternative view, many institutional practices may affect the academic engagement and achievement of Black/Hispanic minority students (e.g., Flores-Gonzalez, 2005). We highlight three such factors.

The first is known as the subtractive schooling process, in which minority culture is devalued by school teachers and administrators. Under such a school system, teachers and administrators select and reward students who possess mainstream cultural attributes and deprive minority students of important cultural and social resources, “leaving them progressively vulnerable to academic failure” (Valenzuela, 1999: 3).

The second school institutional practice that is thought to be responsible for Black/Hispanic students’ low academic achievement is what is known as cultural tracking (see Flores-Gonzalez, 2005; Valenzuela, 1999). Cultural tracking is an institutional practice that puts students into particular programs based on their dominant-culture proficiency within schools. The result of such a practice is that students are divided into two main groups – those students who are believed to be equipped with dominant culture skills and those who are not. As minority culture is often devalued, the majority of minority students are sorted into the lower tracks. On the other hand, because White students have the advantage of possessing dominant culture skills, they (along with a very small number of minority students who demonstrate dominant-culture

proficiency) are placed into more advanced programs. Not only does such an institutional practice shape opportunities for academic success, it also places students into different peer groups on the basis of their socially defined racial backgrounds (e.g., Lucas, 1999).

A final institutional practice believed to underlie the low achievement of Black/Hispanic students is what has been identified as Eurocentric curricula and pedagogy. Although the call for culturally responsive curricula and pedagogy was made more than two decades ago (e.g., Boykin, 1983), the curricula used in industrial countries (in particular, the United States where the minority groups being considered reside) are designed to address issues in the predominant White culture. The pedagogy used often does not take into account minority students' ways of learning. Such an institutional practice tends to discourage minority students from being active in their learning because they see little connection between their educational experiences in predominantly White schools and their own cultural values.

Conclusions and Future Directions

Convergent empirical evidence on students' learning processes and outcomes enables us to draw several conclusions. First, culture does have a significant impact on student learning. However, the influence of culture on student learning is rarely direct. Rather, the relationship between culture and student learning is often mediated by other factors of a sociological or psychological nature. Existing studies on both learning processes and learning outcomes cast doubt on some of the stereotypical views about particular cultural groups. To achieve a better understanding of learning in a cross-cultural perspective, further research is needed regarding both learning process (as represented by intellectual styles) and learning outcome (as represented by academic achievement).

On Intellectual Styles

Research on intellectual styles may help resolve some of the issues arising in this article. First, cross-cultural research on intellectual styles should be conducted among a wider range of populations. Second, we need to know more how culture interacts with intellectual styles. Third, we need better instruments to assess styles – ones that can be used in a variety of cultures.

On Academic Achievement

We also need to have a better understanding of academic achievement. First, cross-cultural research on academic achievement should involve more countries.

At present, many countries in the world have been left out in existing cross-national studies. Furthermore, the bulk of the immigrant literature is about comparing the academic achievement of students from the so-called underdeveloped countries with that of students from Western industrial countries. Practically nothing is known about the differential (or similar) achievement among students across underdeveloped nations. Second, more attention should be given to the learning gaps between students from different regions or different ethnic groups within countries. Third, we need to examine more closely the impact of the interplay of culture, social class, family and school environment, psychological factors, and other student-developmental dimensions (e.g., cognitive development and psychosocial development) on student learning. Fourth, research comparing students' academic achievement should go beyond using results from tests of basic skills. Other intellectual skills such as higher-order cognitive skills, critical thinking, and creative problem solving should be examined. Finally, cross-cultural research on academic achievement should go beyond studying school achievement. Progress is being made, but there is a long way to go. Given the costs to nations of slow progress, there is little time to lose.

See also: Achievement Goal Theory: Definitions, Correlates, and Unresolved Questions.

Bibliography

- Adler, M. (1968). Intelligence testing of the culturally disadvantaged: Some pitfalls. *Journal of Negro Education* 37(4), 364–369.
- Allport, G. W. (1937). *Structure et développement de la personnalité*. Neuchâtel: Delachaux-Niestlé.
- Bagley, C. and Mallick, K. (1998). Field independence, cultural context and academic achievement: A commentary. *British Journal of Educational Psychology* 68, 581–587.
- Berry, J. W. (1966). Temne and Eskimo perceptual skills. *Journal of International Psychology* 1, 207–299.
- Biggs, J. B. (1978). Individual and group differences in study processes. *British Journal of Educational Psychology* 48, 266–279.
- Bourdieu, P. (1977). Cultural reproduction and social reproduction. In Karabel, J. and Halsey, A. H. (eds.) *Power and Ideology in Education*, pp 487–511. New York: Oxford University Press.
- Boykin, W. (1983). On task performance and Afro-American children. In Spencer, U. R. (ed.) *Achievement and Achievement Motives*, pp 324–371. Boston, MA: Freeman.
- Buriel, R. (1978). Relationship of the three field-dependence measures to the reading and math achievement of Anglo American and Mexican American children. *Journal of Educational Psychology* 70(2), 167–174.
- Chao, R. K. (1996). Chinese and European American mothers' beliefs about the role of parenting in children's school success. *Journal of Cross-Cultural Psychology* 27, 403–422.
- Coleman, J. S. (1966). *Equality of Educational Opportunity*. Washington, DC: U.S. Government Printing Office, US Department of Health, Education and Welfare, Office of Education.
- Crow, L. D., Murray, W. I., and Smythe, H. H. (1966). *Educating the Culturally Disadvantaged Child*. New York: D. McKay.
- Davis, A. (1948). *Social-Class Influences upon Learning*. Cambridge, MA: Harvard University Press.

- DeBlassie, A. M. and DeBlassie, R. R. (1996). Education of Hispanic youth: A cultural lag. *Adolescence* **31**(121), 205–216.
- Deutsch, M. P. (1964). The disadvantaged child and the learning process. In Reissman, F., Cohen, J., and Pearl, A. (eds.) *Mental Health of the Poor*, pp 172–187. New York: The Free Press.
- Eaton, M. J. and Dembo, M. H. (1997). Differences in the motivational beliefs of Asian American and non-Asian students. *Journal of Educational Psychology* **89**(3), 433–440.
- Flores-Gonzalez, N. (2005). Popularity versus respect: School structure, peer groups and Latino academic achievement. *International Journal of Qualitative Studies in Education* **18**(5), 625–642.
- Grigorenko, E. L. and Sternberg, R. J. (1997). Styles of thinking, abilities, and academic performance. *Exceptional Children* **63**(3), 295–312.
- Hattie, J., Marsh, H. W., Neill, J. T., and Richards, G. E. (1997). Adventure education and outward bound: Out-of-class experiences that make a lasting difference. *Review of Educational Research* **67**(1), 43–87.
- LeVine, R., Miller, P., and West, M. (eds.) (1988). *Parental Behavior in Diverse Societies: Volume 40. New Directions for Child Development*. San Francisco, CA: Jossey-Bass.
- Lucas, S. (1999). *Tracking Inequality: Stratification and Mobility in American Schools*. New York: Teachers College Press.
- Markus, H. R. and Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion and motivation. *Psychological Review* **98**, 224–253.
- Ogbu, J. U. (1978). *Minority Education and Caste: The American System in Cross-Cultural Perspective*. New York: Academic Press.
- Pelto, P. J. (1968). The difference between “tight” and “loose” societies. *Transaction April*, 37–40.
- Sternberg, R. J. (1997). *Thinking Styles*. New York: Cambridge University Press.
- Sue, S. and Okazaki, S. (1990). Asian–American educational achievements. *American Psychologist* **45**, 913–920.
- Trends in International Mathematics and Science Study (2008). The International Association for the Evaluation of Educational Achievement, Amsterdam.
- Valenzuela, A. (1999). *Subtractive Schooling: US–Mexican Youth and the Politics of Caring*. Albany, NY: SUNY Press.
- Watkins, D. and Mboya, M. (1997). Assessing the learning processes of Black South African students. *Journal of Psychology* **131**, 623–640.
- Witkin, H. A. (1962). *Psychological Differentiation; Studies of Development*. New York: Wiley.
- Zhang, L. F. and Sternberg, R. J. (2005). A threefold model of intellectual styles. *Educational Psychology Review* **17**(1), 1–53.
- Zhang, L. F. and Sternberg, R. J. (2006). *The Nature of Intellectual Styles*. Mahwah, NJ: Erlbaum.

Further Reading

- Lew, J. (2006). *Asian Americans in Class: Charting the Achievement Gap among Korean American Youth*. New York: Teachers College Press, Columbia University.
- Mandara, J. (2006). The impact of family functioning on African American males' academic achievement: A review and clarification of the empirical literature. *Teachers College Record* **108**(2), 206–223.
- Zhang, L. F. and Sternberg, R. J. (eds.) (2009). *Perspectives on the Nature of Intellectual Styles*. New York: Springer.

Learning Outside of School

T Nunes, University of Oxford, Oxford, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Learning is the acquisition of knowledge. Learning may be about new ideas, symbols, ways of performing (more automatically, more flexibly), or the use of something known in a new way (learning transfer). Learning happens everywhere but it is easy not to notice this because we are used to linking teaching and learning, so we tend to look for learning where there is teaching. This leads to a focus on learning in schools, colleges, and universities. In this article, the focus is on learning outside school, where much learning takes place without teaching.

This analysis is organized by a series of contrasts, which aim at making visible what is often invisible, because what we learn outside school is seen as common sense, everyday knowledge, social knowledge, professional knowledge – that is, knowledge that is taken for granted, which is part of the general culture or part of the culture of a group. “Like fish in water, we fail to ‘see’ culture because it is the medium within which we exist” (Cole, 1996). By contrasting learning in and out of school, a picture of the products and the processes of learning outside school are sketched. However, before we work our way through these contrasts, it must be said that looking at learning outside school is not a criticism of school learning. Schools have specific aims and functions, and one of these is to make learning an explicitly pursued venture, where the knowledge obtained can be examined critically and communicated; thus its refinement and transmission are, if not assured, made more likely, so that future generations may inherit it and build on it. Informal knowledge and learning are, by comparison, more difficult to share and their future is uncertain.

Formal and Informal Learning and Knowledge

Eraut (2000) proposed a definition of formal learning on the basis of these characteristics: (1) a prescribed learning framework; (2) an organized learning event or package; (3) the presence of a designated teacher or trainer; (4) the award of a qualification or credit; and (5) the external specification of outcomes. Not all of these are necessary to define formal learning; any one may suffice, but they tend to go together. Outside school, learning happens without explicit learning outcomes, a curriculum, or a designated teacher. The contrast between formal and informal

learning also applies to knowledge: formal knowledge receives recognition and informal knowledge, though often necessary for learning formal knowledge, does not. In fact, informal knowledge is taken for granted, considered obvious, and only identified as an achievement through research.

An example from mathematics learning can help clarify the difference between informal and formal knowledge. Outside school, children are used to sharing different sorts of things: chocolate bars, sweets, turns in games, etc. They learn that these are not shared in the same way. For example, turns in games are not a matter of equal shares: one child’s turn ends when a mistake is made, so you skip rope until you step on it, you play the video-game until the game is over. Turns in games are earned and regulated by the rules of the game. It is rather different when children share a chocolate bar or sweets: the assumption is that the shares will be equal, even if perfect equivalence of the shares is not attained. Children also learn through sharing that, the more people sharing, the smaller each one’s share will be. This informal knowledge of fractional quantities requires an insight into the inverse relation between the divisor (i.e., the number by which something is divided) and the quotient (i.e., the result of the division). Few 5-year-olds realize this, but the majority of 7-year-olds knows it. This important insight into mathematical relations can be used to give meaning to the learning of fractions in school. However, this idea is neither taught nor receives recognition in school – it is taken for granted and not perceived as knowledge of fractions. Children are only credited with knowledge of fractions if the symbols used to represent fractions are involved in the questions that they are asked. The formal equivalent of this insight would be to know, for example, which fraction is bigger, $1/3$ or $1/5$. If the children can answer this, they are credited with some knowledge of fractions.

What is the difference between these two forms of knowledge, if they involve the same insight into the inverse relation between the divisor and the quotient? The comparison between $1/3$ and $1/5$ is a comparison of two written symbols; mastery of written symbols and their manipulations are recognized school-learning outcomes. The oral forms one-third and one-fifth are readings of these written forms, and as such they also obtain recognition in school. In contrast, when one talks about one chocolate shared among three children, and one chocolate shared among five children, this is a description of a situation, and its understanding is not given the same

status as comparing two symbols that have a general meaning. Schools include in their curriculum the teaching of fractions through formal representations that can be used to designate quantities; they typically do not aim at teaching children about sharing chocolates or sweets – children should learn this at home or in preschool.

This example illustrates how some forms of knowledge are seen as aims for formal teaching and learning, whereas others are not. Two aspects of formal knowledge in this case merit further consideration – the emphasis on written versus oral forms of knowledge and the emphasis on general versus specific and situated knowledge. Each of these contrasts is discussed in turn in the context of mathematics knowledge.

Written and Oral Practices

The three R's – reading, writing, and 'rithmetic – are standard aims of Western schooling. The emphasis on written practices and their transmission is a mark of schooling; people are divided into literate and non-literate for comparative descriptions of the level of education across countries. There is no similar distinction in mathematics but the emphasis on mastering the use of written symbols is so strong in school that knowledge necessary to learn these symbols can be taken for granted and not recognized as a form of knowledge.

Analyses of written and oral arithmetic practices emerged as the interest in contrasting European and African thinking grew in anthropology and psychology. Gay and Cole (1967), Zaslavsky (1973), Reed and Lave (1981), and Lancy (1983) described mathematical thinking learned without schooling, in the market, playing games, or in the course of learning a trade. Perhaps the most systematic description of mathematics learned outside school was carried out in Brazil. The Brazilian school curriculum includes teaching children to write and read numbers and to calculate using algorithms applied to written symbols. The focus on the use of written symbols and their manipulation renders children's knowledge of mathematics developed outside school invisible. Therefore, many children fail arithmetic assessments in school when they are perfectly able to perform the same calculations outside school.

A powerful demonstration of the difference between mathematics learned in and out of school comes from a study where five young vendors working in the informal economy, selling produce at street markets and corners, were asked to solve the same set of problems as part of their commercial activities and, later, in school-like situations (Nunes *et al.*, 1993). The children were very successful when they solved problems in the course of their work in the market: 98% of their answers were correct. In contrast, when they were presented with the same numbers in the

form of computation exercises, only 37% of their answers were correct. The main difference between what the children did in these two situations was that they solved the problems in the market using oral numbers – they spoke to the customers as they solved the problems – whereas they tried to solve the operations by writing them down and using the computation routines that they had been taught in school. **Figure 1(a)** shows a young vendor in his work setting, a beach in Brazil. **Table 1** presents a contrast between how one of these young people solved the same



Figure 1 (a) A young vendor selling ice-cream on the beach. (b) A young man selling hats.

Table 1 A child's answers to 35×4

M, a boy aged 12 years, who had 5 years of schooling, was presented with the question 35×4 when selling coconuts and later asked to solve the same sum.

In the informal setting:

Customer: I'm going to take four coconuts (each coconut costs 35 Cruzeiros, the Brazilian currency at the time). How much is that?

M: Three will be 105, plus 30, that's 135 . . . one coconut is 35 . . . that is . . . 140 (. . . indicates pauses where the intonation suggests that the sentence is still being continued).

Solving the computation exercise 35×4 :

C writes down 35×4 in the traditional arrangement, with 35 on one line and $\times 4$ under it, and calculates out loud: Four times five is 20, carry the two (writes 0 as a partial answer). Two plus three is five, times four is 20 (writes 20 to the left of the zero). 200.

From Nunes, T., Schliemann, A. D., and Carraher, D. W. (1993). *Street Mathematics and School Mathematics*. New York: Cambridge University Press, p 24.

arithmetic problem presented as part of a sale and later, as a computation exercise.

Brazilian young people and adults who had different occupations – fishermen, carpenters, farmers, and foremen in construction sites – and either no, or at most primary, schooling can solve many different types of problem without recourse to written symbols. Their learning outside school enables them to solve problems with any of the four basic arithmetic operations and different contents: finding out the unit price from the price for a lot, how much wood is required for building a piece of furniture, how much shrimp a fisherman has to fish if a customer wants a certain amount of shelled shrimp, how many liters of wine are needed to fill a certain number of bottles, or the size of a wall from its representation on a scale drawing. Calculating with negative numbers, simple proportions, averages, area, and volume are within the domains of what can be learned outside school. What is learned depends on the trade but the type of reasoning is not trade specific. For example, fishermen know roughly how much shelled shrimp they can get from a certain amount of untreated shrimp. They are also able to apply this proportional reasoning to agricultural products: if they are told how much cassava flour you can get from a certain amount of cassava, they can use this information to say how much cassava flour you can get from other amounts of cassava. Learning arithmetic on the job, outside school, does not only lead to job-specific knowledge, but also leads to a more general understanding of how to operate with quantities.

General versus Specific Learning

Schools are transitional places. Students are not being prepared to be students, but to go on to do something

else. Neither they nor their teachers know what they will do after they leave school. The solution to this difficult task of preparing students for doing something not yet specified is to try to teach them general knowledge, which could be used in many different occupations and settings. The expectation is that they will be able to apply this general knowledge to specific situations in which it is required. However, the application of general knowledge to specific settings is not a trivial matter. Spatial learning illustrates well the issues of general and specific knowledge. General spatial knowledge is about how to organize space: left and right, north and south, east and west can be combined with measures of space and knowledge of shapes to support spatial learning. This knowledge can be used when we arrive in a new town, but it is not the same as knowing the town. We get to know the town by looking at maps, going around, recognizing landmarks, and ordering them so that we know what is farther north than what. This type of spatial learning involves the acquisition of specific knowledge, and also more general learning: when we use a map to get to know a town, we also learn about using maps in a more general way.

General spatial knowledge can be learned in school; getting to know a town requires learning outside school. When we focus on getting to know a town, we use general spatial knowledge embedded in the knowledge of the town. We may not say “going north” but say “going towards the Bronx”; this latter way of speaking often reduces the number of steps in our thinking about spatial relations. There is often a tension between the use of general knowledge, which is focused on general representations, and the use of specific knowledge, where these representations are interpreted through the meanings that they have in the situation where we learn. This contrast is explored in the next section.

Focus on Representations versus Focus on Meaning

Concept learning involves putting meanings together with their representations, which are often conventional. Learning vocabulary, for example, which starts many years before school, is about learning words – a verbal representation – and their meanings. Although representations and meanings are closely related, learning outside school influences the way we speak: we speak about the meanings rather than about the general aspects of knowledge.

A very simple example, which we encountered in our study of mathematics learned outside school, is that of a 9-year-old boy, R, solving a problem in a simulated shop situation. He used written arithmetic to solve the problem, as he would have used in school, but it was his understanding of the meaning in the situation that showed him that something had gone wrong with his calculations.

Table 2 presents an extract of his interaction with the interviewer.

A school-like argument, which a teacher could present to R, would be: if you take something away from 200, the answer cannot be 200 or greater than 200. This argument, formulated in a general way, contrasts with R's argument, which has the same sense but is presented embedded in the meaning that 200 has in this situation. He argues that he should not hand back to the customer the 200 note that the customer just gave him.

The focus on meaning when learning happens on the job was also observed by Noss *et al.* (1999), who asked nurses to read blood pressure charts and say what average blood pressure meant. Students usually learn in school a concept of average as a measure of central tendency, used to describe a population. Nurses look at blood pressure charts to see whether changes in the blood pressure of a patient are significant enough to call for medication. This gives them a different concept of average, as they add to the mathematical concept the meaning normal for the patient, which is not part of the mathematical concept. When they look at a blood pressure chart, they do not calculate the mean but consider the variation of the measurements, estimate the average to be around the most frequent value at the center of these variations, and then make a judgment about the need for medication. The meaning of average is placed in the context of their activities. When asked about the variations that are observed in the chart for a child, they think about what the child could have been doing at the time. In a statistics class, these could be treated as random variations or

measurement error. A quote from an interview illustrates this focus on the meaning of average interpreted in situational terms.

She was... it was settled here, the early hours of the morning... I don't know, what with the activities of the day probably getting on top of her here, so she's probably a bit more agitated there, so you could allow for the increase [in blood pressure], or whatever. But probably in the morning she's more settled... Yeah, I'd like to think that she was asleep then, so she had a settled blood pressure (Noss *et al.*, 1999: 37).

Noss *et al.* (2007) further illustrate this attribution of meaning in another situation where graph reading was learned on the job. Jim, a shift leader in a large factory that makes plastic sheets, had learned, through 31 years of experience on the job, to read the complex graphs that recorded variations in tension and temperature of the materials over time and at different places in the production process. When a problem occurred in the production process, he would take the graphs and examine what was happening just before the problem occurred. Using this information, he was able to identify the location of the problem, and maintenance engineers would then be called in to check the item he had identified as the likely cause. This was not seen by the chemical engineer, who was responsible for the overall production, as a simple accomplishment: Jim had to be able to identify readings in a graph as normal, thus attributing to them a meaning that was learned in the situation. She described briefly what he had done in one instance when he thought that one straight line on the graph (which contained many different lines, including some straight ones) looked too good to be true.

By examining the "revs [revolutions] per minute" historical data and seeing that it wasn't fluctuating as Jim expected but was a constant value, he determined something was wrong with it. He alerted the maintenance engineers who found the motor on the controller had been fitted to run backward after some work done on it the previous day and so the signal it was sending for revs per minute was false. Jim doesn't know anything about motor control – he just knew that the historical data looked "wrong" (Noss *et al.*, 2007: 377).

Thus the process of learning on the job, outside school, involves going beyond the information given in the representations, by attributing to them specific meanings taken from the situation, which the representations would not convey by themselves. R's argument that it is not sensible to return to a customer the note that he just received in payment for a merchandise, the nurses reading into the blood chart that the child's blood pressure was settled while she was asleep and went up when the activities started in the ward, and Jim's identification of a specific

Table 2 Using meaning to assess the results of a computation

Interviewer (I): Now I'm buying this, and it costs 35 Cruzeiros. I'm going to pay with a 200 note. How much will you give me back?
R: (writes 200–35 in the traditional alignment, and says): Five, to get to zero, nothing. Three, to get to zero, nothing. Two, take away nothing, two.
I: Is it right?
R: No. So you buy something from me, and it costs 35, you pay with a 200-cruzeiro note and I give it back to you?
I: Do it again, then.
R: (writes down 200–35 in the same way) Five, take away nothing, five. Three, take away zero, three. Two, take away nothing, two. Wrong again.
I: Why is it wrong again?
R: Now you buy something and it costs 35. You give me 200 and I give you 200, and 35 on top?
I: Do you know what the result is?
R: If it were to cost 30, then I'd give you one seventy.
I: But it is 35. Are you giving me a discount?
R: 165.

From Nunes, T., Schliemann, A. D., and Carraher, D. W. (1993). *Street Mathematics and School Mathematics*. New York: Cambridge University Press, p 46.

straight line on the graph as too good to be true, all exemplify this process of going beyond the information given in the symbols by imbuing them with situation-specific meaning.

Thought versus Action

Sylvia Scribner, one of the most distinguished researchers of learning outside school, argued that models of the mind based on the metaphor of the mind as a computer fail to capture those significant aspects of learning that are based on action. In a series of works about thinking in action (Tobach *et al.*, 1997), she reasoned that learning, memory, and thinking in everyday life are not separate from what we do but part of what we do. Therefore, what we do is also part of our learning, memory, and thinking. In order to develop this argument, she drew on Luria's (1973) concept of functional systems as the basis for higher mental functions. Luria's example of how memory works helps us understand the basic notion that there is no opposition between thinking and action, or mental and manual work. When we want to remember something, we might repeat what we want to remember many times to ourselves, and be able to recall it later. However, our memory function would be greatly restricted if this were all we could do. Fortunately, we can also write down what we need to remember, and be so organized about where we write what down that we know where to look for the information when we need it. Therefore, Luria argued, memory is not an activity restricted to the brain: it can be accomplished through functional systems that maintain the task of remembering constant but accomplish it through a variety of means, including the incorporation of external objects into our memory system.

Learning outside school leads people to incorporate into their thinking systems the tools and the characteristics of the objects that are used in the environment. Scribner (in Tobach *et al.*, 1997: 338–366) analyzed how product assemblers in a dairy factory in the US incorporated the characteristics of packing into their thinking as they pre-packaged items for an order. Products were packed as different units: a case equals 4 gallons, 8 half-gallons, 16 quarts, 32 pints, or 48 half-pints. When a product assembler is putting together an order, he might have a load-out form that shows 1 – 6, that is, 1 case minus 6 units. If this order called for something that is packaged in quarts and the assembler filled it literally, he would start from a full case (which has 16 quarts) and remove six from it. However, this is not necessarily how assemblers work. If there is a half-empty case close by, the assembler is much more likely to add 2 units to it, translating the order into half-case plus 2, a solution that saves effort and time. Scribner compared product assemblers' performance with that of students,

who would be able to use the same reasoning but had not learned about assembling on the job and had limited experience of assembling products through their participation in Scribner's study. Product assemblers were about 3 times more likely to use effort-saving strategies than students, when these did not coincide with the literal solution. Learning on the job granted product assemblers greater flexibility in treating the units in cases as units of thinking: in the same way that we think easily using the decades of our number system, they had learned to use the units of packing as thinking units. Just looking at a half-full case of quarts allowed them to know that there were 8 units there, and move smoothly to the action of adding 2 units.

Scribner concluded from her different studies of thinking in action that higher-order solutions such as those illustrated above had a major effort-saving impact on the manual execution of the work: in one estimate, she concluded that literal solutions with one item assembled at a time would involve walking roughly 4 miles to complete the jobs; this contrasted with the actual solutions, which involved a little more than 2 miles. Assemblers had not learned this effort-saving from lingering over the order forms or engaging in discussions with each other. Thus, Scribner concluded, they must have learned by representing quantities and space in a way that synthesized the different forms of organizing and regulating their own actions; they learned by thinking in action.

The incorporation of cultural and cognitive tools into the learner's thinking system is not a defining characteristic of out-of-school learning. Wertsch (1998) provides a convincing example of this incorporation in his discussion of how we solve multi-digit multiplications. If we are asked to say what is 343 times 862 without using a calculator, we would write down these numbers in vertical alignment and carry out the algorithm that we learned in school. This algorithm is a cultural and cognitive tool, which we incorporate into our thinking, and without which we would not be able to solve the problem. It differs across countries slightly – for example, in some countries, when we multiply 343 by the 6 in 862, we write the product one row to the left. We are, in fact, multiplying 343 by 60, not by 6, and this is simplified by acting as if the multiplication is by 6, thereby making the multiplication routine more similar across the units, tens, and hundreds. Learners can carry out the procedure without full awareness of why a place is skipped in the process – they are, in this case, relying on a tool and downloading part of the difficulties of the process onto the paper-and-pencil record (to use Hatano's, 1997, expression).

Learning in and out of school is a process that allows people to participate in socially organized activities, so incorporating the tools that are used in these activities into our thinking is a common feature of learning in and out of school.

Table 3 Some of the idealized characteristics of informal and formal education

<i>Informal education</i>	<i>Formal education</i>
1. Embedded in daily-life activities	1. Set apart from the context of everyday life
2. Learner is responsible for obtaining knowledge and skill	2. Teacher is responsible for imparting knowledge and skill
3. Personal; relatives are appropriate as teachers	3. Impersonal; teachers should not be relatives
4. Little or no explicit pedagogy or curriculum	4. Explicit pedagogy and curriculum
5. Learning by observation and imitation	5. Learning by verbal interchange, questioning
6. Motivated by social contribution of novices and their participation (in real activities)	6. Less-strong social motivation

Some of the idealized characteristics of informal and formal education (adapted from Greenfield & Lave, 1982, p. 183) from the book *CULTURAL PERSPECTIVES ON CHILD DEVELOPMENT* ed. by D. A. Wagner and H. W. Stevenson. Copyright 1982 by W. H. Freeman. Reprinted by permission of Henry Holt and Company, LLC.

The Places and Processes of Learning

There are similarities between learning in and out of school. However, it is important not to forget the differences, which are related to how roles are defined for the teacher and the learner in and out of school, what is expected of each, and what is the focus of their attention as education takes place. Greenfield and Lave (1982) summarized, some years ago, the processes by which schooling and informal education seek to promote learning; the characteristics relevant to this analysis are presented in **Table 3**.

Their analysis can be complemented by a more detailed description of the implicit pedagogy that emerges from analyses of learning outside school. Rogoff (2003) analyzed how novice girl scouts become experts in selling cookies and how mothers guide their children's learning in a variety of settings outside school. She termed *guided participation* the implicit pedagogy used by adults outside school, a term that indicates that novices do not take full responsibility from the outset in learning situations, but still have some responsibility for the activities they engage in. Girl scouts are able to participate in selling cookies, even as novices, through the use of forms that they fill in, which work as tools that organize the activity, and through their interactions with peers, as they work together.

Rogoff distinguished different forms of guided participation when mothers have the role of tutors. In middle-class US families, adults organize child-centered learning events, and structure the children's learning by directing their attention, promoting their motivation, and maintaining

their engagement. European and US middle-class mothers involve their children in events that are similar to school lessons in the family. Tizard and Hughes (2002) observed European mothers playing games that are aimed to prepare children for what they will learn in school (e.g., Mother: "I spy, with my little eye, something beginning with S"; child has to identify something in the room whose name begins with the letter S) and lesson-like activities in the home, where the mother writes something and the child copies it. Heath (1983) also observed, in the US, how mothers read stories to their children at bedtime and prepare them for adopting the concept of a story and for comprehension questions that are similar to those asked of children in school. In contrast, mothers from traditional societies, where tradition is a value, expect children to learn through watching and manage their own attention, motivation, and involvement. Rogoff cites examples from Japanese, Polynesian, and Mayan communities, where children are expected to learn through intent participation, being able to observe and listen in ongoing processes that were not set up for their learning. Guided participation then takes the form of encouragement of observation and responsive assistance, for example, rather than through school-like lessons enacted outside school.

Conclusion

In summary, learning outside school is learning in action, by doing something with an aim that is different from learning itself, by participation in an activity where the learner has some responsibility. Thinking in the course of these activities incorporates the representational tools and objects that are part of these activities, and does not single out, as more valuable, written over oral representations, or general over situated arguments and explanations. When using representational tools, such as number systems and graphs, people go beyond the information given in the representational tools, adding meanings from their learning in the situations where the tools are embedded. Outside school, learning is often guided, monitored, but not tested or given official recognition. There is no explicit curriculum, no designated teacher, no explicit learning outcome, and no accreditation – all of which make the learning that happens outside school invisible to some extent. The sample of studies presented here should help make out-of-school learning more visible, less taken for granted.

See also: An Overview of Language and Literacy in Educational Settings; Learning in a Sociocultural Perspective; Mathematics Learning; Reasoning and Explanatory Talk: Learning in Everyday Settings; Situated View of Learning; Vygotsky and Recent Developments.

Bibliography

- Cole, M. (1996). *Cultural Psychology: A Once and Future Discipline*. Cambridge, MA: Belknap Press of Harvard University Press.
- Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology* **70**, 113–136.
- Gay, J. and Cole, M. (1967). *The New Mathematics and an Old Culture. A Study of Learning among the Kpelle of Liberia*. New York: Holt, Rinehart and Winston.
- Greenfield, P. and Lave, J. (1982). Cognitive aspects of informal education. In Wagner, D. A. and Stevenson, H. W. (eds.) *Cultural Perspectives on Child Development*, pp 181–207. San Francisco, CA: Freeman.
- Hatano, G. (1997). Learning arithmetic with an abacus. In Nunes, T. and Bryant, P. (eds.) *Learning and Teaching Mathematics. An International Perspective*, pp 209–232. Hove: Psychology Press.
- Heath, S. B. (1983). *Ways with Words: Language, Life and Work in Communities and Classrooms*. Cambridge: Cambridge University Press.
- Lancy, D. F. (1983). *Cross-Cultural Studies in Cognition and Mathematics*. New York: Academic Press.
- Luria, A. R. (1973). *The Working Brain. An Introduction to Neuropsychology*. Harmondsworth: Penguin.
- Noss, R., Bakker, A., Hoyles, C., and Kent, P. (2007). Situating graphs as workplace knowledge. *Educational Studies in Mathematics* **65**, 367–384.
- Noss, R., Pozzi, S., and Hoyles, C. (1999). Touching epistemologies: Meaning of average and variation in nursing practice. *Educational Studies in Mathematics* **40**, 25–51.
- Nunes, T., Schliemann, A. D., and Carraher, D. W. (1993). *Street Mathematics and School Mathematics*. New York: Cambridge University Press.
- Reed, H. J. and Lave, J. (1981). Arithmetic as a tool for investigating relations between culture and cognition. In Casson, R. W. (ed.) *Language, Culture and Cognition: Anthropological Perspectives*, pp 437–455. New York: Macmillan.
- Rogoff, B. (2003). *The Cultural Nature of Human Development*. New York: Oxford University Press.
- Tizard, B. and Hughes, M. (2002). *Young Children Learning*, 2nd edn. Oxford: Blackwell.
- Tobach, E., Falmagne, R. J., Parlee, M. B., Martin, L. M. W., and Kapelman, A. S. (eds.) (1997). *Mind and Social Practice. Selected Writings of Sylvia Scribner*. Cambridge: Cambridge University Press.
- Wertsch, J. V. (1998). *Mind as Action*. New York: Oxford University Press.
- Zaslavsky, C. (1973). *Africa Counts: Number and Pattern in African Culture*. Westport, CT: Lawrence Hill.

Further Reading

- Greenfield, P. M. (2004). *Weaving Generations Together. Evolving Creativity in the Maya of Chiapas*. Santa Fe, NM: School of American Research Press.
- Lave, J. (1988). *Cognition in Practice*. Cambridge: Cambridge University Press.
- Resnick, L. B., Levine, J. M., and Teasley, S. D. (eds.) (1991). *Perspectives on Socially Shared Cognition*, pp 349–364. Washington, DC: American Psychological Association.
- Rogoff, B. and Lave, J. (1984). *Everyday Cognition*. Cambridge, MA: Harvard University Press.
- Saxe, G. B. (1991). *Culture and Cognitive Development: Studies in Mathematical Understanding*. Hillsdale, NJ: Erlbaum.

Reasoning and Explanatory Talk: Learning in Everyday Settings

D E Beals, The University of Tulsa, Tulsa, OK, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Explanatory talk with, and by, children offers a unique opportunity to examine both cognitive and linguist development and outcomes. Children may learn the relationship between objects, concepts, and/or events delineated in an explanation. Conversely, they must understand the relationship in order to explain it to someone else.

Historically, there have been two basic approaches to children's explanation. The first is based in the philosophy of science – linking scientific explanation with children's reasoning. The second is based in sociolinguistic and sociocultural theories – focusing on the interaction between children and more knowledgeable others that allows the child to participate in the language of reasoning. Although the two approaches have two somewhat distinct research and theoretical literatures, each perspective benefits the other by linking what a child learns and knows to what linguistic, social, and cultural contexts the child is exposed to and participates in. The first approach focuses on what is going on inside the child's mind (cognition), while the second focuses on what is going on in the world outside the child's mind (interaction). While these two approaches examine children's talk from different perspectives, there is much crossover. It is almost a cliché to say that a child's speech is a window to its mind. The window glass, however, is not transparent, but translucent – allowing researchers to infer what a child thinks from the forms and shadows of the internal workings of the mind visible from the outside. This article focuses more specifically on the sociolinguist approach; however, the cognitive approach provides a strong foundation – theoretically and historically – for the framework of the interactionist approach.

Definition of Explanation

The philosophical literature on why we seek and give explanations and what kinds of things need explanation is a long one. In the last century, much work has been limited to a logical-positivist concentration on explanation as mathematical proof – allowing for deductive-nomological thought (Hempel, 1965). According to Hempel and others, an explanation of a scientific phenomenon is a deduction from natural laws the nature of consequences or effects; major views of explanations in the last century have concentrated on the ability to predict future events based on

these deductions from laws of nature. There is a strong flavor of the language of logic and rationality in these definitions, focusing entirely on scientific explanation.

Keil (2006), however – in review of the literature – took issue with the narrow range of Hempel's explanation. He outlines several purposes for explanation, including: (1) predicting events in the future based on laws and consequences (Hempel's definition); (2) diagnosing system failures and repair of the system based on these; (3) placing of blame *post facto* since a specific prediction in advance is unlikely (Keil's example was with regard to the car crash that killed Princess Diana); and (4) justifying a specific action as result from reasonable or beneficent intentions. While the first two generally follow more traditional definitions and scientific explanation of reasoning, the last two are less about logic and more about human intention, which is not always rational. Keil refers to these as everyday explanations. He sees the social nature of these everyday explanations, claiming that “explanations view people less as autodidacts and more social interacting agents” (p. 230). The work of several developmentalists have examined children's ability to comprehend and produce scientific explanations and even argue for the similarities between the children as developing explainers, scientists, and psychologists.

Explanation as Discourse Genre

Bruner (1985) posited the existence of two major modes of thinking: narrative – which deals with intentions and actions and their consequences – and paradigmatic – which uses formal logical and scientific description and explanation. The narrative mode attempts to emphasize the particular and locate it in time and space, the paradigmatic mode attempts to “transcend the particular by higher and higher reaching of abstraction” (p. 13). These two modes are played out in different forms of discourse genres such as stories and plans for narrative thought and arguments and explanations for paradigmatic thought.

Bruner assumed that narrative is the earliest mode and discourse genre accessible to young children, with the paradigmatic thought and discourse emerging later. In fact, he argued that narrative thought is natural, untrained, and even primal. The ubiquity of story-telling and storybook-reading discourse in interaction with children is a reflection of such an assumption (Pappas, 1993). If the assumption is correct, it is important to examine the

emergence of paradigmatic discourse genres, such as explanation and argumentation.

Considerable attention has been paid to the development and display of narratives by numerous researchers, for example, Nelson and her colleagues (Nelson, 1986) – who examined a variety of narratives and charted the linguistic resources children use in narratives; and by Slobin and collaborators (Berman and Slobin, 1994) – who examined narrative production across cultures. In comparison, relatively little research has focused on the development of explanatory discourse, with the exception of the naturalistic, descriptive work of Barbieri (1991), Landolfi (1989), and Beals (1993). Additional developmental work on explanation has focused entirely on explanations at the sentence level.

Narratives and explanations share a number of features that might lead one to expect more or less parallel development. Both are structured forms of extended discourse around a specific topic that require formulation of a general goal and control over inter-utterance cohesion markers. Both require the participants to talk about topics outside of the immediate present. Both forms occur with relative frequency in everyday conversations and are characterized by collaborative multiparty participation structures. Exposure to both discourse forms relates to early indicators of school success, such as vocabulary, skill in rendering formal definitions, and comprehension. Further, research has shown that even young children such as kindergartners demonstrate no more difficulty remembering and retelling information in explanatory forms than in narrative forms found in children's books (Pappas, 1993).

There does not always appear to be a clear distinction between narrative and explanation in the literature; some scholars have argued that explanation is not a separate genre but a primary function of narrative. Kemper's (1984) work on narrative structure claimed that causal (and intentional) relationships are generally found in stories people tell. According to Kemper, stories lay out events in terms of intentions and evaluations of the storyteller through the use of temporal and causal connections. Ochs *et al.* (1992) studied the use of narrative as a form of problem solving at mealtimes in which family members co-narrate stories, and in the process, use and polish the social, cognitive, and linguistic skills that underlie scientific and other scholarly discourse. Ochs and co-workers claimed that the collaboration around narrative supports theory building among speakers, not unlike the work of scientists. Even more extreme views exist; some consider all forms of discourse to be some kind of narrative.

There are a number of crucial interrelated features to extended discourse. First, extended discourse requires paying attention to the needs of the audience. Bakhtin refers to the addressivity of utterances: speakers are always responding to utterances previously spoken and

in anticipation of the potential responses of their listeners. Discourse is then constructed by speakers from what they assume the point of view of the listener to be. Another audience requires that speakers take into consideration how much knowledge they have in common with their listeners in order to gauge how much detail and explicitness to provide in the course of discussing a topic. Listeners in conversation are able to provide feedback to the speaker by asking clarifying questions or by facial expressions, and the speaker monitors how effectively his talk or text is in meeting the audience's needs both in terms of the amount of information needed and in terms of preparing a response to the content of the interlocutor's utterance(s). This allows the speaker or writer to recognize when he is not connecting well with the listener or reader, and to revise his discourse in order to get his point across.

Another aspect of extended discourse is its structure. Discourse comes in a variety of genres, each with its own purposes, demands, and affordances. In written discourse, there is frequently a common structure that aids readers in comprehending the text by allowing them to predict and recognize relationships within and across sentences. Much work has been done identifying structures of various types of children's narrative genres and others have studied explanatory genres. While written discourse is more likely to demonstrate the usual structure of a genre, conversational discourse is less predictable in form and sequence, but generally the kinds of information found in the written forms are found in the conversational genres as well.

Children's Explanations

Talk between adults and children can take many forms, but one type that has been shown to have both cognitive and linguistic benefits is a type of talk that Sigel (1981) calls parental distancing behavior and Snow calls extended discourse (Snow and Kurland, 1996). In this sort of talk, the parent talks in ways that allows the child to distance herself from the immediate present, to represent mentally some other place and/or time. This sort of talk places not only cognitive demands on the child in order to comprehend and produce it, but it also makes linguistic demands on the speaker to be more explicit about people, places, objects, events, and the relations among them. As the topic is not necessarily present in the immediate social and physical context, the speakers must make assumptions about what the listeners know and what they are thinking about at the moment of conversation. Much research has demonstrated the relationships between children's exposure to such talk and their cognitive and linguistic skills.

Underlying a child's ability to explain something is its understanding of the logical relations involved.

These abilities, at least in the case of causal relations, have been shown to emerge early in language development. While the work of Piaget (1926, 1930) suggested that children do not have true understanding of causality and causal connectives (such as *because* and *so*) until age 7 or 8, other researchers have demonstrated that some understanding of cause and its linguistic expression emerge prior to age 3. Observational studies of children's spontaneous production of causal explanations in conversation with familiar adults provide some evidence that children as young as 2.5 years begin to use their first expressions of causal relations and use causal connectives correctly.

In his studies of children's conceptions of causality, Piaget (1928) suggested three types of causal relations: (1) physical or empirical (between two events); (2) psychological (between two psychological activities, mental states, and/or intentions; and (3) logical (between two ideas or judgments, or between a judgment and its proof or logical antecedent). Piaget's hypothesis was that – because of their egocentric thought – preoperational (ages 2–7) would fail to distinguish between these three types and would tend to overgeneralize the psychological relations (e.g., “The door slammed because it was angry”). In his analysis, children come to understand psychological relations first, followed by physical relations, then logical relations. His research confirmed this hypothesis. Later research on very young children (prior to age 3) confirmed that the most frequent explanations express psychological causality.

Following Piaget, Donaldson (1986) referred to these categories of explanations as the content of explanations. Each content category is based on the type of rules or laws that are invoked in the explanation. The following sentences illustrate these content categories (p. 3):

The window broke because a ball hit it. (Physical)
 Mary hit John because he pulled her hair. (Psychological)
 Half nine is not four because four and four make eight.
 (Logical)

Donaldson distinguished these content categories from modes of explanations. Modes correspond to the types of questions that explanations answer. The four modes of explanation are (1) empirical (answering the question “what caused it?”); (2) intentional (“for what purpose?”); (3) deductive (“how do you know it?”); and (4) procedural (“how do you do it?”). Donaldson emphasizes that content categories and modes of explanation are not totally independent. A specific content type constrains the mode for linguistic expression of that content type to some extent. Psychological content may be asserted in an empirical, intentional, or deductive mode of explanation, indicating a broad range of possibilities for expression. However, logical content is expressed exclusively by a deductive explanation mode, although physical and

psychological relations may be presupposed within the explanation.

Donaldson performed a series of studies on children's abilities to produce explanations of the three types of content. Utilizing tasks that focused on each of the three content categories, she elicited causal statements from children at different ages. She found that children as young as age 3 produced well-formed and appropriate explanations of psychological, physical, and logical phenomena. This contradicts Piaget's (1928) findings that children are unable to produce appropriate explanations until the age of 7 or 8, and once they can, they tend to psychologize, or give psychological explanations of physical phenomena. She attributes these contradictory findings to the fact that children in her studies were asked to explain events and emotions familiar to them, while Piaget asked children to explain for esoteric phenomena. Donaldson's results coincide with the earlier work that indicates, in spontaneous speech, that preschool children produce causal explanations for all kinds of things. Donaldson also performed a series of studies aimed at eliciting each mode of explanation (except procedural, which she chose to exclude from her research). The results of each study indicate that children, by age 5, have considerable skill in producing empirical, psychological, and deductive explanations.

Explanation as Social Interaction

Explanation – like other types of talk – is a social activity. Much of the research on children's production of explanations has focused on explanations of causality, exploring children's use of sentential connectives such as ‘*because*’ and ‘*so*’. The decision to focus on a causal definition of explanation was, most likely, a methodological one; this definition is easily operationalized by a search for the use (or implication) of causal terms and connectives in children's talk. There is another line of research that draws sociolinguistic and sociocultural theories – using discourse-analysis methodology to examine children's participation in everyday explanatory talk. This line generally differs from the more cognitively focused research by defining explanatory talk as a conversational exchange in which one speaker recognizes that another person needs some information, and as a result, offers the information. This information is often causal or intentional links are a major category, there remain other kinds of explanations that express a variety of logical relations, such as definitions and procedures. The focus on connectives, such as ‘*because*’ and ‘*so*’, implies that the sentence is the unit of analysis. However, social interactionists emphasize the fact that explanations occur in the give and take between interlocutors – in longer segments of discourse, often over several speaker turns. These studies examine the talk of the child and those around her in the conversation. Children hear explanations, give

explanations, and jointly construct explanations with parents, siblings, and friends.

Minimally, an explanation occurs when one person makes some connection (between events, actions, ideas, or judgments) clear to another. An explanation, however, is rarely a simple utterance. It is often prefaced by a request for information, interrupted by requests for clarification or expansion, elaborated by additional information from other persons, and closed by an acknowledgement of understanding. A more sociolinguistic approach to the study of the development of explanatory talk emphasizes examination of the immediate physical and social context of the conversations for evidence of the occurrence of an explanation. Such an approach has been utilized by Barbieri (1991) and Landolfi (1989).

Barbieri, arguing that explanations are defined not by content only, but also by context, studied explanations that arise in naturalistic settings. Analysis of children's conversations in play with other children revealed another category of mode of explanation. While Donaldson's modes of explanation posed 'why' and 'how' questions, Barbieri *et al.* found, in addition, that children produced 'what' explanations. These explanations answer such questions as "what is the type and/or the characteristics of an object?" and "what does a human action or verbal expression mean?" She found that six out of nine of her 3-year-old subjects produced such explanations spontaneously in conversation.

While Barbieri's dependence on context to reveal the presence of an explanation is intuitively appealing, operationalizing such a definition is not a simple matter. She suggested two sets of conditions necessary for the existence of an explanation in conversation. These include: (1) the preparatory conditions: joint attention to what is to be explained, and the belief on the speaker's part that such an explanation is called for; and (2) the essential conditions: the explanation is comprehensible to the hearer, and there must be something to be explained and some new information about it. These two conditions highlight the interactional nature of explanation, and provide a broader move toward an operational definition for explanations in children's conversations.

Based on her work on discourse in adult second-language classrooms, Landolfi (1989) distinguished between the synchronic and diachronic aspects of explanation. The synchronic aspect refers to "explanatory activities as they occur within social interactions (i.e., inter-subjective development)" (p. 137), while the diachronic aspect refers to "the history of synchronic explanation as internalized by each individual (i.e., intra-subjective development)" (p. 137). Only synchronic explanations are observable, while diachronic explanations are inferred.

Landolfi posited that synchronic explanations are jointly constructed by interlocutors, and that they unfold in a cycle (see Figure 1), which begins with a trigger

(a direct or indirect question or an observable fact or event) that creates a state of dissonance in a person's universe of knowledge. This causes the person to seek other sources of knowledge (e.g., another person, a book, the Internet) that re-establishes a state of assonance. The new knowledge that brings about assonance rearranges the prior universe of knowledge, producing a new universe of knowledge. The cycle then begins again, resulting in a diachronic spiral (Figure 2) – a repeating cycle bringing interlocutors into deeper and shared understanding of the universe of knowledge. In this way, an explanation is co-constructed by the person seeking new knowledge and the source of new knowledge. Although Landolfi's model grows out of research with adults in a structured learning environment, it provides a reasonable framework for identifying explanations and observing their joint construction with children in formal and informal conversational settings.

This social interactionist approach to the study of children's participation in explanatory talk has yielded

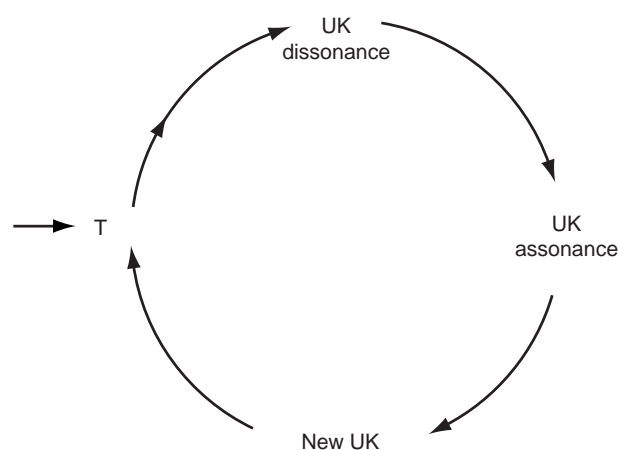


Figure 1 Synchronic cycle. T, trigger; UK, universe of knowledge.

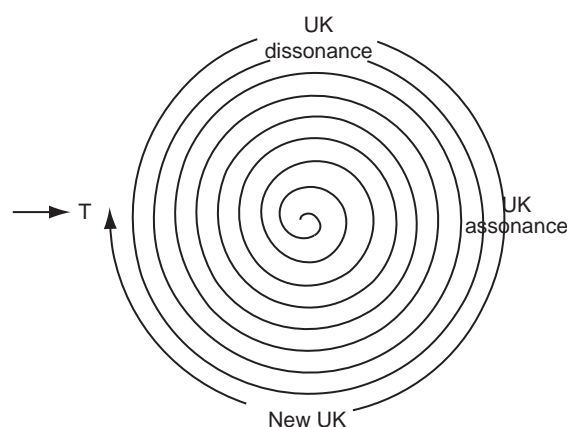


Figure 2 Diachronic cycle. T, trigger; UK, universe of knowledge.

much fruit. Following the work of Donaldson, Barbieri, and Landolfi, several researchers have analyzed everyday conversational settings, such as mealtimes and play time, in order to examine how children used explanations in spontaneous talk with adults and with other children. Drawing on data from a large longitudinal study, Beals (1993) examined spontaneously emerging explanatory talk in 160 transcripts of mealtime conversations of children in the study at age 3, 4, and 5. On average, 13–14 explanations took place in a mealtime (average length of recordings was about 20 min), accounting for about 15% of the all talk in the conversation. Beals found that approximately two-thirds of all explanations were about the intentions of a person in doing or saying something. Causal explanations were present but relatively infrequent.

This line of research has been further expanded to cultural comparative work. In a study comparing the use of explanatory talk during mealtimes in American and Norwegian families, Aukrust and Snow (1998) adapted Beals' categories of explanation. The findings indicated that American families tended to engage in more explanatory talk than did Norwegian families. Americans talked more of intentions and causes, while Norwegian families focused on social norms and violations of them. Aukrust and Snow interpreted these findings as reflecting the relatively individualistic norms of Americans and collectivistic norms of Norwegians.

Summary

The two approaches to the study of children's exposure to and engagement in explanatory have generated rich theoretical and research bodies, many of which merge into outlines of a portrait of children as explainers and seekers of explanations. The cognitive approach points to children's early understanding of causality and intentionality, based on their naive biological, physical, and psychological theories. Children as young as age 30 months are able to comprehend and participate in talk concerning explainable phenomena – much sooner than thought by Piaget – and they are able to do this with increasing frequency and facility throughout their preschool years. The developing mind of the child – able to comprehend and predict basis scientific and psychological phenomena – engages with the social world around it and expresses it in co-constructed discourse in a wide range of explanatory topics and forms.

Bibliography

- Aukrust, V. G. and Snow, C. E. (1998). Narratives and explanations during mealtime conversations in Norway and the U.S. *Language in Society* **27**, 221–246.
- Barbieri, M. S. (1991). The origin of explanations. *Contributi di Psicologia* **4**, 153–176.
- Beals, D. E. (1993). Explanations in low-income families' mealtime conversations. *Applied Psycholinguistics* **14**(4), 489–513.
- Berman, R. and Slobin, D. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale, NJ: Erlbaum.
- Bruner, J. (1985). *Actual Minds, Possible Worlds*. Cambridge, MA: Harvard University Press.
- Donaldson, M. (1986). *Children's Explanations: A Psycholinguistic Study*. London: Cambridge University Press.
- Hempel, C. G. (1965). *Aspects of Scientific Explanation*. New York: Free Press.
- Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology* **57**, 227–254.
- Kemper, S. (1984). The development of narrative skills: Explanations and entertainments. In Kuczaj, S. (ed.) *Discourse Development: Progress in Cognitive Development Research*, pp 99–124. New York: Springer.
- Landolfi, L. (1989). Explanation as social activity. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition* **11**(4), 136–141.
- Nelson, K. (1986). *Event Knowledge: Structure and Function in Development*. Hillsdale, NJ: Erlbaum.
- Ochs, E., Taylor, C., Rudolph, D., and Smith, R. (1992). Storytelling as a theory-building activity. *Discourse Processes* **15**, 37–72.
- Pappas, C. C. (1993). Is narrative "primary"? Some insights from kindergartners' pretend readings of stories and information books. *Journal of Reading Behavior* **25**(1), 97–129.
- Piaget, J. (1926). *The Language and Thought of the Child*. New York: Routledge.
- Piaget, J. (1928). *Judgement and Reasoning in the Child*. London: Routledge and Kegan Paul.
- Piaget, J. (1930). *The Child's Conception of Causality*. Totowa, NJ: Littlefield, Adams.
- Sigel, I. E. (1981). Social experience in the development of representational thought: Distancing theory. In Sigel, I. E., Brodzinsky, D. M., and Golinkoff, R. M. (eds.) *New Directions in Piagetian Theory and Practice*, pp 203–218. Hillsdale, NJ: Erlbaum.
- Snow, C. E. and Kurland, B. F. (1996). Sticking to the point: Talk about magnets as a context for engaging in scientific discourse. In Hicks, D. (ed.) *Child Discourse and Social Learning: An Interdisciplinary Perspective*, pp 189–220. New York: Cambridge University Press.

Further Reading

- Antaki, C. (1981). *The Psychology of Ordinary Explanations of Social Behaviour*. New York: Academic Press.
- Barbieri, M. S., Colavita, F., and Scheuer, N. (1990). The beginning of the explaining capacity. In Ramsden, G. C. and Snow, C. (eds.) *Children's Language*, vol. 7, pp 245–272. Hillsdale, NJ: Erlbaum.
- Keil, F. C. (2002). *Explanation and Cognition*. Cambridge, MA: Bradford Books/MIT Press.

Relevant Website

<http://plato.stanford.edu> – Scientific Explanation, Stanford Encyclopedia of Philosophy.

LEARNING AND COGNITION – THEORETICAL PERSPECTIVES – LEARNING

Contents

Apprenticeship Approach to Learning
Connectionism and Learning
Constructivism and Learning
Cultural–Historical Activity Theory
Learning in a Sociocultural Perspective
Self-Regulated Learning and Socio-Cognitive Theory

Apprenticeship Approach to Learning

K Nielsen, University of Aarhus, Aarhus, Denmark

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Apprenticeship – Training in an art, trade, or craft under a legal agreement defining the relationship between master and learner and the duration and conditions of their relationship. There are different connotations of the Anglo-Saxon apprenticeship and the German–Scandinavian *mesterlære* (master learning). In the Anglo-Saxon apprenticeship, the learning of the apprentice is stressed. Prevalence of the master's role for learning is implied in the Scandinavian–German term for apprenticeship – *mesterlære*.

Apprenticeship in the guilds – The formal European apprenticeship of the guilds involves a written contract between master and apprentice, stating obligations for both parts, with the term of apprenticeship often ending after a 4-year period with a journeyman's test and an official accreditation as a journeyman of the trade.

Community of practice – A term especially introduced in Lave and Wenger's *Situated Learning. Legitimate Peripheral Participation* (1991) meaning an activity system about which participants share understandings concerning what they are doing and what that means in their lives and for their communities.

Scaffolding – The concept of scaffolding is a metaphor from building constructions used to illustrate the interaction between tutor and child/pupil. The notion of scaffolding is influenced by Vygotsky's concept of zone of proximal development which is understood as the distance between the

actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers.

Scholastic approach to education – An approach to education characterized by a narrow concern for book learning and formal rules excluding knowledge or experience of practical matters.

Learning through apprenticeship was, for centuries, the central way of introducing youth to the skills, knowledge, and values of a craft and a profession. In this and the last century, apprenticeship has been replaced by schooling in many areas. Whereas learning in schools has been extensively researched, apprenticeship learning has enjoyed less attention from educational researchers. In the last few decades, a new interest in apprenticeship learning has emerged: in anthropological studies of learning as social practice, in philosophical analyses of skill learning and tacit knowledge, and in educational experiments with cognitive apprenticeship in the school curriculum.

In this article, some key issues are outlined from the renewed focus on apprenticeship. First, some definitions, main aspects, and varieties of apprenticeship are outlined. Second, different theoretical approaches to apprenticeship are emphasized – anthropological, philosophical, and psychological. Two dynamic perspectives on the dynamics of the learning process are presented: a person-centered versus a decentered approach. Different conceptions of the

nature of apprenticeship learning are described such as learning in a community of practice, learning through participation in practice, bodily learning and imitation, and learning through assessment of practice. Finally, some reasons for the current interest in apprenticeship are discussed.

Delineation of Apprenticeship

Today, the term apprenticeship is used in several different ways. This calls for some distinctions concerning definition, main aspects, and varieties of apprenticeship to be mentioned.

Definition

Apprenticeship is defined in *Encyclopedia Britannica* (2007) as: “training in an art, trade, or craft under a legal agreement that defines the duration and conditions of the relationship between master and apprentice”. Apprenticeship originates from the French phrase *a prendre* which means to learn or to grasp mentally (Webster, 1968). There are different connotations of the Anglo-Saxon apprenticeship and the German–Scandinavian *mesterlere* (master learning). In the Anglo-Saxon apprenticeship, the learning of the apprentice is stressed. The prominence of the master’s role for learning is implied in the Scandinavian–German term for apprenticeship – *mesterlere*.

Varieties of Apprenticeship

Apprenticeship occurs in many forms and fields. The formal European apprenticeship of the guilds involves a written contract between master and apprentice, stating obligations for both parties and terms of apprenticeship, which often end after a 4-year period with a journeyman’s test and an official accreditation as a journeyman of the trade. Within vocational education in parts of Northern Europe, the classical apprenticeship model has gradually been replaced by a dual system whereby the apprentices move between intermittent periods of practice at a workplace and attending lessons at technical schools (Christensen, 1995; Bailey, 1993; Hamilton, 1990). Other apprenticeships may be more informal, whereby the apprentices, so to speak, grow up with a trade. In Lave and Wenger’s (1991) development of a theory of situated learning, they draw upon five cases of apprenticeship: Yucatec midwives, Liberian tailors, naval quartermasters, meat cutters, and sober alcoholics. Of these, the apprenticeships of the meat cutters were formally defined in writing, the tailors in oral agreements, whereas the midwives grew up with the family trade.

While apprenticeship is commonly associated with the crafts, it takes place within a wide spectrum of fields. Within the field of classical music, apprenticeship learning is often used to describe the way the students learn at the

Academy of Music (Nielsen, 1999). Artists have long learned through apprenticeships with a master as, for example, in the studio of Rembrandt. Consequently, today, art historians have difficulties determining whether a painting is by the master or an apprentice of his. It is equally striking today how artists’ biographies emphasize the masters they have studied with; for example, see *Stocktakings of an Apprenticeship* by Boulez (1991). The role apprenticeship plays in learning advanced intellectual pursuits is documented by the Nobel laureates in the sciences (Kvale, 1997).

Apprenticeship is found in most cultures (Coy, 1989), as witnessed in the Asian spiritual traditions as well as in martial arts. The master–disciple relationship is essential for the transmission of truth in major wisdom traditions, such as Buddhism. While enlightenment and freedom cannot be obtained without relying on a master, it is the truth of the teaching which is all important; Buddha thus warned, “Rely on the message of the teacher, not on his personality” (Rinpoche, 1992).

Theoretical Approaches

Apprenticeship has been explored within different disciplines and theoretical positions. Three approaches – an anthropological, a philosophical, and a psychological–educational – are outlined shortly.

Anthropological Approaches to Apprenticeship

Inspired by fieldwork in third world countries, anthropologists have reconceptualized learning from the perspective of social practice. One background for this move was the increasing tendency in the Western industrial world to perceive learning as a mechanical consequence of teaching, as an individualistic endeavor to pass the next examination, or to fulfill the need of self-realization. By studying apprenticeship, anthropologists understood learning as a basic social phenomenon, initiating both change and continuity in a social practice (Coy, 1989; Lave and Wenger, 1991; Singleton, 1998). Learning was an aspect of cooperation between different participants in making a production work. This cooperation was embedded in everyday practice, and not isolated in educational institutions. From one perspective of anthropology, the real master in apprenticeship is the ongoing social and bodily activities, in particular organized practices, which in turn create a particular social habitus (Bourdieu, 1992).

Philosophical Approaches to Apprenticeship

One of the consequences of the efforts to create artificial intelligence has been an expansion of our understanding of

human competences and practical rationality. In the efforts to simulate different kinds of practical competences in a computer, the limitations of desituated and disembodied simulations of practice became apparent. Philosophers have turned to apprenticeship as a way to learn tacit and bodily competences (Polanyi, 1958). Not all competences could be formulated into explicit rules and transmitted verbally; some could only be transmitted through practical action. Only by observing the experienced master and by using the skills in a practical situation could the novice acquire these kinds of practical competences (Dreyfus and Dreyfus, 1986, 1999; Wackerhausen, 1996).

Psychological and Educational Approaches to Apprenticeship

A psychological approach to apprenticeship has been inspired by Bandura's theory of social learning, emphasizing identification and imitation as central means for learning (Jespersen, 1997). Cognitive apprenticeship is a term designated to a synthesis of school and apprenticeship and is based on insights gained from studies of apprenticeship in school curriculums. Through modeling and scaffolding, the teacher offers the students an understanding of the strategies behind solving a problem (Collins *et al.*, 1991). Gardner (1993) discusses in *Multiple Intelligences* apprenticeship and a scholastic approach to art education, and also depicts the use of apprentice-like situations in the classroom. Apprenticeship has also been used to designate the psychological and social development of the child. The interaction between parents and children metaphorically resembles the interaction between master and apprentices, and when the child learns cognitive structures from participation and observation of the parents' everyday activities (Rogoff, 1990).

The Dynamics of Apprenticeship Learning

Within the various theories of apprenticeship learning, different perspectives on the dynamics of the learning process are found. Two approaches – a person-centered and a decentered approach – are outlined.

A Person-Centered Approach to Apprenticeship Learning

With the absence of formal teaching in apprenticeship, the question remains of what advances the learning process – the master or the structures of the community in practice? In a person-centered approach to apprenticeship learning, the point of focus is centered on the relationship between master and apprentice (Dreyfus and Dreyfus, 1999; Polanyi, 1958). The master is the

practitioner, who, in practice, reflects and makes the thinking of the profession visible to the apprentice (Schön, 1987). The master may serve as a role model (Collins *et al.*, 1991), making the tasks to be learned visible, and also by serving as a source of identification. Ziehe and Stubenrauch (1983) have, in a discussion of learning at a school for clowns, pointed to a transition over time in the master–apprentice relation, moving from a personal to a structural identification. It is no longer the personality of the master, but his relation to the topic of the work and learning which is decisive. Learning is not merely an imitation of the master, but also based on the relation of the master to the topic of learning.

A Decentered Understanding of Apprenticeship Learning

In a decentered understanding of apprenticeship, the decisive factor is the participation of the learner in a community of practice. Lave and Wenger (1991) emphasize, in contrast to a strongly asymmetrical master apprenticeship relation, learning through legitimate peripheral participation in communities of practice. This leads to a focus on the intricate structuring of a community's learning resources: "To take a decentred view of master–apprentice relations leads to an understanding that mastery resides not in the master but in the organization of the community of practice of which the master is a part. . ." (p. 94). What motivates the apprentice is to become a fully legitimate member of the community of practice at the workplace.

From a decentered perspective on apprenticeship, suggestions for what is important for the learning process include: open structures of the community of practice, which facilitate observation and sharing of ways of solving problems; a gradual move to more central and important tasks in practice; access to observe older apprentices or journeymen's way of handling the production process; and the way the learning process is organized in relation to the production process and access to narratives of the trade (Lave and Wenger, 1991).

Central Features of Learning in Apprenticeship

The kinds of learning found in studies of apprenticeship introduce a shift in attention from the attributes of individual participants to the aspects of the environment, which make learning possible. This contextualization of the learning process implies a movement from individuals who learn in isolation to communities of practice where learning occurs.

The following is an elaboration of the specific features of learning processes that take place in apprenticeship such as learning in a community of practice, learning

through participation in practice, bodily learning and imitation, and learning through assessment of practice (see Nielsen and Kvale (2006) for an elaboration). Some rudimentary comments on how apprenticeship learning differs from learning in a school community are also offered.

Learning in a Community of Practice

Apprenticeship learning is often related to learning at a workplace. To learn in practice is of crucial importance if the workshop, shop, company, enterprise, or department wants to survive. Learning in apprenticeship entails becoming part of the culture of the workplace. The apprentice has to learn the rhythm and the jargon at the workplace, and show he or she is capable of participating in the various interhuman relations in order to gain access to the more professional tasks.

For the community of practice, it is essential that the workplace as a whole functions; the learning processes of the apprentices are secondary. In a school community, in contrast, it is essential that the individual student learns the knowledge of the subject, as it is outlined in the school curriculum. In a workplace, many different people from whom he/she can learn ideally support an apprentice's learning process. The apprentice is situated in a multi-generational environment and encounters a variety of different learning resources in the job situation (Akre and Ludvigsen, 1997). An important part of apprenticeship learning is for the apprentice to find the person who masters the specific part of the learning process which the apprentice is to learn (Hansen, 2003). In a school, the students are divided into classes, in which everyone is at the same level, and thus, some of the learning potential that exists in a manifold multigenerational environment is discarded. In other words, the school and the workplace as communities of practice are organized differently. Learning has often been comprehended as a hierarchical one-way process in keeping with an educational assumption that learning is a consequence of teaching. In descriptions of workplace activities, learning is situated in a multidimensional network of social and material resources (Nielsen and Kvale, 2006).

Learning through Participation in Practice

The notion of learning through participation in practice is used in a broad sense as a process in which the student changes his/her participation from just taking part to becoming a responsible participant in a community of practice (Elmholdt, 2003). If we look at the learning that takes place in practical situations, it may rarely be the result of direct teaching. Learning in apprenticeship is incorporated in daily activities, whereby one hardly notices that it takes place. At this point learning is indirect; it is a side effect when completing existing assignments. An

important element is that when learning happens through participation of practice, the individual apprentice is thrown into practical assignments (Aakrog, 2003). The apprentices are first thrown into different assignments, after which they develop an understanding for the context and the meaning of the assignments. In traditional schools, students are seldom thrown into different practical assignments right away. They are usually taught the principles and theoretical background first, whereby they gain an understanding of what they need to do in order to complete a given assignment.

Another aspect of learning through participation of practice is that the apprentices themselves have to be active and search for learning occasions in order to utilize existing learning resources in the workplace (Becker, 1972). Nielsen and Kvale (2006) present a number of examples of how apprentices make journeymen or masters aware of the fact that they need more demanding work assignments.

One way to conceptualize learning through participation in practice is to focus on responsibility (Hansen, 2003; Nielsen and Kvale, 2003). The concept of responsibility is brought to light in order to emphasize the relational and social dimensions of the learning process. By relating learning to responsibility, the learning process is seen as mutual interaction and an active dialog between individual and environment. In other words, responsibility is seen as being crucial to what one does in relation to others. To have learned something means to be capable of doing something responsibly. A substantial part of learning in apprenticeship is related to the fact that the apprentice is held responsible for a part of the production, which also means that he/she has to consider himself or herself as part of a larger unit. This is expressed in how the apprentice is not only responsible for own learning, but also faces responsibility in relation to others within the company. To participate responsibly also refers to a competence dimension, in which the individual as part of the learning process gradually improves in order to live up to and complete his/her assignments. The learning task is completed when the student or apprentice is properly capable of fulfilling the demands of the craft, which are required in order to fulfill an assignment, and the apprentice is capable of assuming the assigned responsibility.

In the apprentices' descriptions of their work, an assessment that leads to increased responsibility plays an important role. The different ways in which the apprentices participate in the work processes can be illustrated as a ladder of increasing responsibility (Wilbrandt, 2002). The apprentices move from the lower steps up in a continuous and mostly tacit assessment process. **Figure 1** illustrates the apprentices' increased responsibility on this ladder of tasks.

Figure 1 contains a description of workplace evaluation as a ladder of tasks with increasing responsibility,

Tasks	Steps
Work independently on assignments outside the workshop and managing the work of other apprentices.	7
Work alone (without journeymen) on complex tasks; the result of the work is unknown.	6
Work alone (without journeymen) on easy routine tasks; the apprentice knows the procedure and expected result.	5
Perform day-to-day tasks of the trade, as instructed by journeymen.	4
Observe journeymen work on complex work tasks, which display the skills of the craft.	3
Perform simple minor tasks; the apprentices take care of activities, which require no particular skills belonging to the trade.	2
Perform simple apprenticeship task (e.g., sweeping the floor, cleaning, and stocktaking), which are not specific to the trade	1

Figure 1 The apprentices' ladder of increasing responsibility.

starting from simple routine tasks (step 1), such as sweeping the floor, to more demanding tasks under the supervision of a journeyman, to performing the tasks alone, and finally, with full responsibility for the tasks and directing the work of the new apprentices and performing tasks outside (step 7). This implicit curriculum embedded in social practice has been termed the learning curriculum (see Lave and Wenger, 1991). The assigned tasks and the applied tools indicate which level of responsibility the apprentice has obtained in the workplace community (Wilbrandt, 2002).

Bodily Learning and Imitation

Learning processes in a workplace can be rather invisible. The execution of an assignment may appear as a daily routine, without being seen as learning. Imitation of other participants in a community of practice, and also identification with more experienced agents of the subject, takes place unintentionally (Jespersen, 1997). Learning through bodily action and by use of tools is incorporated in daily contact with the surroundings, and learning may take place without a deliberate plan for learning. It may not be until new apprentices enter the workplace that the person in question experiences that he/she has learned something.

In apprenticeship, the body is the learning subject. The body is rather absent in educational theory because learning is primarily understood as an intellectual activity. In schools, the body has simply been deported to practical subjects such as woodwork, sports, home economics, and art, which are traditionally low-status subjects within

the educational area. The pedagogical practice in many schools is often based on the deskbound and listening student, whereas in the workplace, the active, moving body is learning; the apprentice moves from area to area for different assignments in his or her learning process. The physical routines in a workplace are resources, and among the participants, it is unnecessary everyday to negotiate everything over and over again. In many cases, practitioners rely on bodily feelings and intuition instead of on calculated analyses of a given problem (Nielsen, 2006).

Imitations and routine work play a central role when addressing issues of learning in workplaces. Generally, today, educational researchers accentuate innovative learning, and ignore how innovation is rooted in familiar routines and traditions. Performing odd jobs and routine assignments may be viewed as a way of gaining access to a workplace culture, in which the apprentices learn to accept a shared responsibility for the workplace and become accepted members of the work community. In addition, performing routine assignments provides the new apprentices with an opportunity to obtain a general understanding of the various work processes (Museaus, 2003). Needless to say, performing only routine assignments throughout the whole period of workplace training is of little value to the apprentices.

In most workplaces, tools and equipment are dominant parts of the organization of the place. Screwdrivers, computers, hammers, briefcases, adjustable spanners, etc. hang on the wall, lie on a table, or lie on the floor. On an immediate level of experience, these tools address us in the daily performance of working assignments. The organization of the workplace itself has meaning for the

learning process. The order in which the tools are placed at the workplace, for example, in a garage, enables the mechanic, with a single look, to find the appropriate tool with the right dimensions needed in order to complete the task at hand. The tools and their positions in the room mean that they are a part of the work process, with which the apprentice does not have to be deliberately concerned.

Learning through Assessment of Practice

The survival of a workplace depends on the quality of its products. Therefore, ideally speaking, anyone who participates in the process of production is liable for the quality of the product. The apprentice's work is frequently checked to assess whether it fulfills the expected professional standards (Kvale, 2007). If the quality is unsatisfactory or mistakes have been made, it may have consequences for the entire community of practice. In other words, in a community of practice, the shared responsibility for the production also involves a responsibility for others' learning. Praise, recognition, or positive feedback make the apprentices grow through their own self-knowledge, whereas criticism, triviality, or negative feedback are experienced as hurtful. In both cases, emotional involvement plays an important role in the learning process.

The assessment aims of the individual apprentice's work pertain to its value for the work community. This means that actions are assessed in terms of their usefulness, providing the assessment with a crucial role in the process of learning. The apprentice is responsible for making a product; therefore, through the product, his or her process of learning is assessed. Is the table straight? Are the tiles even? Does the bread rise? These types of evaluations are made instantly as an integrated part of the work process in which it is decided whether a carpenter's, a tile maker's, or gardener's skills are sufficiently consolidated, when considering what he/she has produced. Professional assessments are part of the social structure in a workplace as the apprentice often spends substantial parts of his/her training with a journeyman. He or she functions both as a model to the apprentice and as the one who constantly keeps an eye on the quality of the apprentice's production. In a safe workplace, the apprentices themselves can freely test their skills without criticism. In the course of their training period, the apprentices gradually become capable of assessing their own and other workers' professional skills, whereby the professional, independent assessment becomes dominant in their working lives. Our actions have consequences for others in a community of practice. They acknowledge or invalidate the value of our acts. Older colleagues may assess the product with a brief comment, possibly just with a nod or a shake of the head. A significant type of assessment may take place with hardly any verbal comments on the

task performed, by assigning the apprentice to tasks that demand greater responsibility. The recognition for a well-done job is directly met by letting the apprentice participate in more complex and significant parts of the work process (see **Figure 1**)

Conclusion

A few conclusions on the present status of apprenticeship may be ventured. Today, there is a growing interest in apprenticeship learning. While the traditional forms of apprenticeship in the crafts have decreased, and more or less integrated with schooling, apprenticeship learning still exists in a variety of areas, in particular in the manual professions. Parallel to the modern advance of schooling, apprenticeship has existed in a variety of areas as an important type of learning. With the current anthropological and philosophical focus on apprenticeship, skill acquisition, and tacit knowledge, the many varieties of apprenticeship learning are coming to attention. Here, empirical investigations may clarify some of the contested claims of the merits and drawbacks of apprenticeship learning. A key issue remains the relevance of the traditional apprenticeship learning to the qualifications required by current forms of production.

Theoretically, the focus upon an institutionalized alternative type of learning as apprenticeship may serve to put school learning into perspective and question some of the assumptions taken for granted. Thus, leaving the presupposition of much educational research that learning depends upon teaching may open for investigating other potential sources of learning in the learner's environment. For educational practice, the question becomes how to design institutions that may advance apprenticeship types of learning today.

See also: Cultural–Historical Activity Theory.

Bibliography

- Aakrog, V. (2003). Læring i detailhandel. In Nielsen, K. and Kvale, S. (eds.) *Praktikkens læringslandskab – At lære gennem arbejde* (A Landscape of Learning. Learning through Work), pp 43–55. Copenhagen: Akademisk Forlag.
- Akre, V. and Ludvigsen, S. (1997). Learning medical practice. *Nordisk Pedagogik. Journal of Nordic Educational Research* 17(3), 152–159.
- Bailey, T. (1993). Can youth apprenticeship thrive in the United States? *Educational Researcher* 22(3), 4–10.
- Becker, H. S. (1972). A school is a lousy place to learn anything in. *American Behavioral Scientist* 16, 85–105.
- Boulez, P. (1991). *Stocktakings from an Apprenticeship*. Oxford: Clarendon Press.
- Bourdieu, P. (1992). *Outline of a Theory of Practice*. Cambridge: Cambridge University Press.

- Christensen, F. (1995). *Hånden, ånden og den faglige stolthed* (*The Hand, the Spirit and Professional Pride*). Uddannelse, April, 1995.
- Collins, A., Brown, J. S., and Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator* **15**, 6–11; 38–46.
- Coy, M. W. (1989). *Apprenticeship*. New York: State University of New York.
- Dreyfus, H. and Dreyfus, S. (1986). *Mind over Machine*. New York: Free Press.
- Dreyfus, H. and Dreyfus, S. (1999). Mesterlære og eksperter læring (Apprenticeship and experts' learning). In Nielsen, K. and Kvale, S. (eds.) *Mesterlære – Læring som Social Praksis* (*Apprenticeship – Learning as Social Practice*). Copenhagen: Hans Reitzels Forlag.
- Elmholdt, C. (2003). Metaphors for learning: Cognitive acquisition versus social participation. *Scandinavian Journal of Educational Research* **47**(2), 115–132.
- Encyclopaedia Britannica (2007). Apprenticeship. *Encyclopaedia Britannica Online*. <http://search.eb.com/eb/article-9008087> (accessed May 2009).
- Gardner, H. (1983). *Multiple Intelligences*. New York: Basic Books.
- Hamilton, S. F. (1990). *Apprenticeship for Adulthood: Preparing Youth for the Future*. New York: Free Press.
- Hansen, R. (2003). Læring gennem øget ansvar i en sportsbutik (Learning and responsibility in a sportsshop). In Nielsen, K. and Kvale, S. (eds.) *Praktikkens Læringslandskab – At lære gennem arbejde* (*A Landscape of Learning. Learning through Work*), pp 55–69. Copenhagen: Akademisk Forlag.
- Jespersen, E. (1997). Modelling in sporting apprenticeship: The role of the body itself is attracting attention. *Nordisk Pedagogik. Journal of Nordic Educational Research* **17**(3), 178–185.
- Kvale, S. (1997). Research apprenticeship. *Nordisk Pedagogik. Journal of Nordic Educational Research* **17**, 186–194.
- Kvale, S. (2007). A workplace perspective on school assessment. In Havnes, A. and McDowell, L. (eds.) *Balancing Dilemmas in Assessment and Learning in Contemporary Education*, pp 197–208. London: Routledge.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Situated Peripheral Participation*. Cambridge: Cambridge University Press.
- Museaus, P. (2003). Forefaldende arbejde – uddannelse eller udnyttelse (Occurring work – education or repetition). In Nielsen, K. and Kvale, S. (eds.) *Praktikkens læringslandskab – At lære gennem arbejde* (*A Landscape of Learning. Learning through Work*), pp 16–38. Copenhagen: Akademisk Forlag.
- Nielsen, K. (1999). Musical apprenticeship. Learning at the academy of music as socially situated. *Psykologisk Skriftserie* **24**(2) Aarhus: Psykologisk Institut/Aarhus Universitet.
- Nielsen, K. (2006). Learning to do things with things: Apprenticeship learning in bakeries as economy and social practice. In Costall, A. and Dreier, O. (eds.) *Doing Things with Things*, pp 209–224. Aldershot: Ashgate.
- Nielsen, K. and Kvale, S. (eds.) (2003). *Praktikkens læringslandskab – At lære gennem arbejde* (*The Practical Landscape of Learning. Learning through Work*). Copenhagen: Akademisk Forlag.
- Nielsen, K. and Kvale, S. (2006). The workplace – A landscape of learning. In Antonacopoulou, E., Jarvis, P., Andersen, V., Elkjaer, B., and Høyrup, S. (eds.) *Learning, Working and Living*, pp 119–135. Basingstoke: Palgrave Macmillan.
- Polanyi, M. (1958). *Personal Knowledge*. London: Routledge and Kegan.
- Rinpoche, S. (1992). *The Tibetan Book of Living and Dying*. San Francisco, CA: Harper.
- Rogoff, B. (1990). *Apprenticeship in Thinking*. New York: Oxford University Press.
- Schön, D. A. (1987). *Educating the Reflective Practitioner – Toward a New Design for Teaching and Learning in the Professions*. London: Jossey-Bass.
- Singleton, J. (1998). *Learning in Likely Places. Varieties of Apprenticeship in Japan*. Cambridge: Cambridge University Press.
- Wackerhausen, S. (1996). Apprenticeship, knowledge and representation. *Nordisk Pedagogik. Journal of Nordic Educational Research* **16**, 194–202.
- Webster (1968). *Webster's Third New International Dictionary*. Springfield, MA: Webster.
- Wilbrandt, J. (2002). *Vekseldannelse i håndværksuddannelser. Lærlinges oplæring, faglighed og identitet mellem skole og virksomhed* (*Alternating Training in the Vocational Education. The Learning of the Apprentice. Professional Competences and Identity between School and Workplace*). Uddannelsesstyrelsens temahæfteserie, No. 14.
- Ziehe, T. and Stubenrauch, H. (1983). *Ny ungdom og usædvanlige læreprocesser* (*New Youth and Unusual Learning Processes*). Copenhagen: Politisk Revy.

Further Reading

- Ainley, P. and Rainbird, H. (1999). *Apprenticeship: Towards a New Paradigm of Learning*. London: Kogan.
- Coy, M. W. (1989). *Apprenticeship*. New York: State University of New York.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Situated Peripheral Participation*. Cambridge: Cambridge University Press.
- Nielsen, K. and Kvale, S. (1997). Current issues of apprenticeship. *Nordisk Pedagogik. Journal of Nordic Educational Research* **17**, 130–139.
- Nielsen, K. and Kvale, S. (eds.) (1997). *Special Issue: Apprenticeship-Learning as Social Practice*. *Nordisk Pedagogik. Journal of Nordic Educational Research* **17**(3).
- Nielsen, K. and Kvale, S. (1999). *Mesterlære – læring som social praksis* (*Apprenticeship – Learning as Social Practice*). Copenhagen: Hans Reitzels Forlag.

Connectionism and Learning

T R Shultz, McGill University, Montreal, QC, Canada

© 2010 Elsevier Ltd. All rights reserved.

Glossary

1-of-n coding – A method of activation coding such that only one of a group of units is active.

Algorithm – Explicit instructions for solving a problem.

Base 10 – Having any of ten values, for example, 0–9.

Binary – Having either of two values, for example, 0 or 1.

Composite number – An integer greater than 1 having more than two divisors; thus, it is not a prime number.

Connection weight – A link between units in a neural network, represented as a real number. It is roughly analogous to a brain synapse.

Connectionism – A style of computing inspired by how brains are thought to function, using networks of simple interconnected units.

Constructivism – The idea that knowledge is constructed by the learner, often by building on top of existing knowledge.

Context – Incidental cues that accompany a learning experience.

Course coding – A method of activation coding such that one unit is very active and its neighboring units are a bit active.

Disequilibrium – In Piaget's theory, it refers to a mismatch between mental representation and environmental outcome, indicating a misunderstanding.

Divisor – An integer that divides equally (i.e., with no remainder) into another integer.

Dyslexia – Impaired reading.

Error – A measure of how far off a network is from learning its training patterns, computed as the sum of squared differences between actual outputs and target outputs, summed over all training patterns.

Error function – A hypothetical function relating network error to the size of a connection weight, usually conceptualized as a parabola with arms opening upward.

Excitation – Increasing the activity of a unit.

Generalize – To perform correctly on problems that have not been trained. It is used to distinguish understanding from mere memorization.

Gradient descent – A learning method that changes network weights in proportion to the slope of the error function in order to reduce network error.

Hidden unit – A unit that mediates between input and output units, thus hiding from the environment and enabling the learning of nonlinear relations.

Hz – The abbreviation for hertz: unit of frequency in cycles per second.

Inhibition – Dampening the activity of a unit.

Input – A vector of activations supplied to a network's input units. The pattern of these activations describes a problem to the network.

Integer – A whole number.

Integer part – The part of a real number preceding the decimal.

KBCC – Knowledge-based cascade correlation, a learning algorithm that builds a network topology by recruiting previously learned networks and single hidden units as needed.

Learning rate – A proportion that multiplies the error slope and, thus, controls the size of weight changes and the speed of learning. A programmer sets this proportion at the start of a simulation so that a network's weights change enough to support quick learning, but not so much that error oscillates across an error minimum.

Modeling – Building and running a computer program to simulate scientific phenomena.

Multiplicand – A number to be multiplied by another.

Net input – The net amount of input entering a unit, computed as the sum of products of sending-unit activations and respective connection weights.

Nonlinear – A kind of function that cannot be represented as a simple linear combination of its inputs, for example, distinguishing the points inside a circle from those outside the circle. The degree of nonlinearity typically makes learning more difficult.

Output – A vector of activations computed by the output units of a network. The pattern of these activations describes the network's response to a particular input pattern.

Prime number – An integer greater than 1 having only two divisors, 1 and itself.

Real number – A decimal number, not having an imaginary part.

Simulate – To reproduce natural phenomena, usually by presenting a computer program with the similar inputs to those found in nature. The program (or model) usually abstracts and simplifies so as not to be overwhelmed by natural complexity.

Slope – The rate at which one variable changes as a function of changes in another variable.

Source network – In KBCC, it refers to a previously learned network that may be recruited to aid the current learning task.

Target network – In KBCC, it refers to a network learning the current problem.

Target vector – The activation values supplied by a teacher or by the environment, indicating correct output values for a particular input pattern.

Threshold – The region in which a unit is highly sensitive to changes in net input.

Topology – The way a network is laid out in terms of its organization and connectivity.

Unit – The basic processing element of an artificial neural network. A unit has a fluctuating level of activity, represented as a real number, corresponding to the average firing rate of a neuron.

Vector – An ordered list of numbers.

Introduction

Connectionism is a style of computing that partly mimics the properties and functions of brains. Incorporating ideas from computer science, artificial intelligence, mathematics, physics, neuroscience, and psychology, connectionists build working computational models of learning and other psychological processes. A few connectionist projects have modeled educational phenomena such as reading and arithmetic, and these applications are highlighted here. After reviewing the basics of modeling and connectionism and focusing on some models of particular relevance to education, the broader implications of connectionism for education are discussed.

Importance of Modeling

The value of building computer models of psychological processes may not be immediately apparent. A general justification is that modeling was consistently useful in a variety of scientific disciplines. A good scientific model typically:

- implements a theory in a precise, concrete, easy-to-manipulate way;
- covers (or generates) a variety of interesting phenomena;
- helps to explain the phenomena that it covers;
- links various observations together, making them easier to understand;
- predicts new phenomena that can be tested; and
- can be improved when evidence reveals new facts or contradicts model predictions.

Since the renaissance of connectionism in the mid-1980s, all of these advantages have been realized to some degree. Indeed, of all the available techniques for psychological modeling, none has produced as many valuable insights into learning as connectionism has.

Modern Connectionism

The central idea of connectionism is that mental processes can be modeled by interconnected networks of simple units. These artificial neural networks can largely be understood in terms of activity and connectivity. As noted in the first row of **Table 1**, neural networks are composed of two elements: units and connection weights. Units are analogous to brain neurons, while weights are analogous to synapses, the connection points between neurons. Unit activity corresponds to the average firing rate in neurons, that is, how often they send an electrical impulse from their cell body down their axon to other neurons. Connection weights correspond to the ability of neurons to excite or inhibit the activity of other neurons. Both units and weights can be implemented in neural networks as vectors of real numbers. In psychological terms, unit activity corresponds to active memory, a focus on particular ideas, while weights encode long-term memory for how ideas are related. Unit activity changes over seconds, whereas changes to connection weights can occur rather quickly or over a period of years, enabling networks to learn.

An artificial neural network is thus a set of units and weights organized in a particular topology. Most neural networks are programmer designed and static, meaning that their topology remains constant, but others build their own network topology automatically during learning.

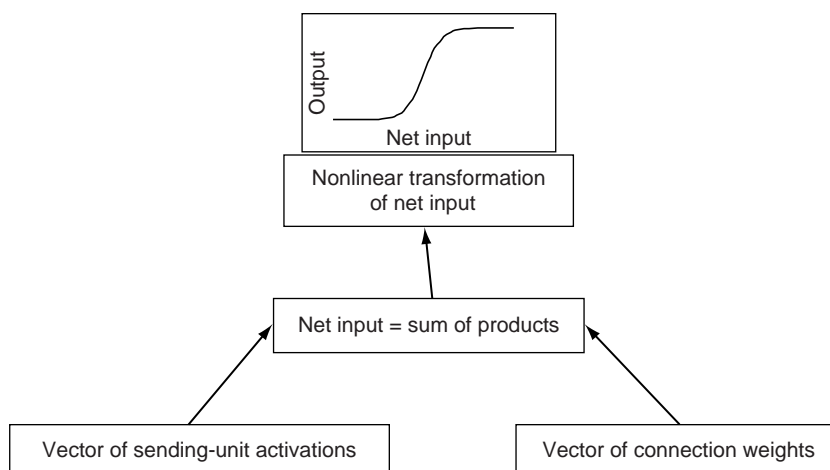
As shown in **Figure 1**, each unit in a network runs a simple program in which it computes a weighted sum of inputs coming from other units and then outputs a number, which is a nonlinear function of the weighted sum of inputs. This output is then sent to other units running this same simple program. Units are active at various levels and communicate their level of activity to other units in a way that modulates the activity of those other units.

As **Figure 1** shows, a receiving unit's level of activity changes in two steps. In the first step, a vector of sending-unit activations is multiplied by a vector of connection weights. Some of these products might be positive (with an excitatory influence on the receiving unit), while others are negative (having an inhibitory effect). The sum of these products constitutes the net input to the receiving unit.

In a second step, this net input is then passed through a nonlinear activation function like that shown in the top of **Figure 1**. Like real neurons, this function has a floor

Table 1 Activity and connectivity in artificial neural networks

<i>Feature</i>	<i>Elements</i>	<i>Brain analogy</i>	<i>Behavior</i>	<i>Implementation</i>	<i>Memory type</i>	<i>Time scale of change</i>
Activity	Units	Neurons	Firing rate	Real numbers	Active	Seconds
Connectivity	Connection weights	Synapses	Excitation, inhibition	Real numbers	Long term	Seconds to years

**Figure 1** Update of unit activation in a neural network.

(corresponding to no activity), a ceiling (corresponding to the maximum firing rate of about 300 Hz in real neurons), and a continuous but sharp transition between them (a threshold). As net input increases below the threshold, there is little change in activity of the receiving unit. However, as net input further increases across the threshold, there is a sharp increase in the activity of the receiving unit. Well beyond the threshold, if net input continues to increase, activity of the receiving unit would again fail to change much because it would be nearing its maximal level. Nonlinear activation functions enable the learning of complex, nonlinear target functions, which are common in the real world. They also prevent networks from overheating through continual increases in activity.

The primary way that neural networks learn is by adjusting their connection weights to reduce error. In supervised learning, connection weights are typically trained by being presented with examples – pairs of input values and target output values. Since there are often multiple inputs and multiple outputs, the example values (represented by real numbers) are presented in the form of vector pairs. In each pair, one vector holds input values and the other holds target output values. By processing such examples, a network gradually learns to produce the correct output vector in response to a particular input vector. The vector pairs used in such training are called the training set; those used in testing

generalization to untrained examples are known as the test set. Error at the output units is computed as the sum of squared differences between actual outputs and target outputs.

Imagine that the error contributed by a single connection weight is some unknown parabolic function of the value of that connection weight as shown in **Figure 2**. This makes sense on the assumption that there is some optimal value for each connection weight such that either increasing or decreasing the value of the weight from this optimum would increase error. If the precise shape of this error function was known, learning would be easy – simply adjust each connection weight to the value that minimizes error. Unfortunately, these error functions are unknown to the learner, and are unlikely to assume the simple, smooth parabola drawn in **Figure 2**. The neural network learner may know the size of the discrepancy between actual and target responses in the training set, but not the exact shape of the error function.

However, this meager information can be used to compute the slope of the error function at each connection-weight value that has been experienced. The slope (or gradient) is the first derivative of a function, evaluated at a particular point. Slope is a measure of how rapidly error changes as a function of changes in a connection weight. Two such hypothetical slopes are shown in **Figure 2** for two different weight values (w_1 and w_2). If the slope at a

given weight value is known, then the direction in which the connection weight should change is also known. If the slope is negative (as at w_2), then the weight should increase; however, if the slope is positive (as at w_1), then the weight should decrease.

However, how much should a connection weight change? If a weight is changed too much, the floor of the error valley could be missed, perhaps creating an oscillation in weight adjustments that never settles near the minimum error. One could make only tiny steps of weight change to avoid passing over the minimum error, but this slows learning. A better solution is to make the amount of weight change proportional to slope. With a steep slope, as at w_1 , a large change in weight is suggested, whereas with a shallow slope, as at w_2 , weight change should be smaller because the error minimum may be close. This technique of using information on the slope of the error function is known as gradient descent because it attempts to slide downhill on the error surface to reach the point of minimum error in weight space. Thus, weight change is a negative proportion (known as learning rate) of the derivative of error with respect to weight.

Relations to the Old Connectionism

Modern connectionism differs from the older connectionism of Thorndike. Thorndike was a founder of behaviorism, emphasizing the idea that organisms learn associations between stimuli and responses. Responses become habitual

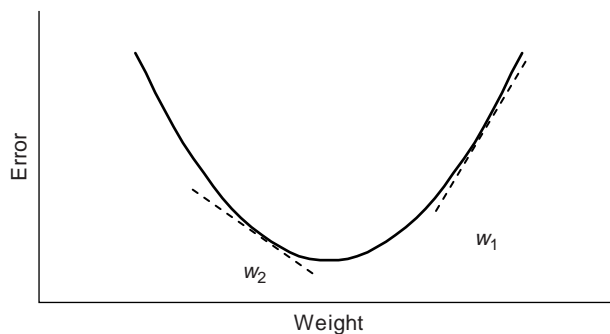


Figure 2 A hypothetical function relating network error to the value of a network weight. Slopes of the error function at two weight values (w_1 and w_2) are drawn with dashed lines.

by being rewarded and thus having their associations to stimuli strengthened. As in other versions of behaviorism, learning was explained without recourse to internal states. As in this article, there were applications to educational issues such as reading and arithmetic.

Despite minor historical influence on, and superficial similarities with, modern connectionism, the differences are more profound than is sometimes recognized. Some important differences between old and new versions of connectionism are listed in **Table 2**, most of which are important for applications to education. Whereas behaviorists wrote about single associations between stimuli and responses, modern connectionists deal with large, multilevel, massively parallel networks. Moreover, some modern networks are designed with recurrent connectivity which allows for sequential processing and complex network dynamics, akin to the interesting behavior of differential equations. Following from these differences, the knowledge representations in behaviorist schemes were entirely local. In contrast, modern networks can as well employ distributed schemes in which each unit represents many different ideas and each idea is represented by many different units. Such distributed representations are more efficient and robust and account for many interesting psychological phenomena, along with being more biologically realistic. Behaviorism disavowed interest in mental states, emphasizing a direct association between stimulus and response. Modern connectionists instead invest considerable energy into determining what their networks know at various points in learning. Such knowledge-representation analyses often play an essential role in explaining psychological phenomena.

Thorndike's law of effect emphasized that habit formation was controlled by rewards. Contemporary connectionist models have clearly demonstrated the difficulty of learning from evaluative reward signals that tell the organism that it is doing well or badly, but fail to indicate what is needed for improvement. In contrast, supervised neural networks learn from fully specified target vectors that indicate the correct response to particular inputs, making learning faster and more accurate. All of the foregoing provide modern networks with vastly more learning power than habits. A habit can handle only a simple linear relation between a stimulus and a response, whereas there are now proofs that a network with a single

Table 2 Some key differences between old and new connectionism

Characteristic	Topology	Knowledge representation	Emphasis	Feedback signal	Learning power	Method
Old connectionism	Single association	Local	Behavior	Evaluative	Linear	Vague speculation
Modern connectionism	Massive parallelism	Distributed	Knowledge representation	Informational	Nonlinear	Explicit modeling

layer of hidden units can learn any continuous function to any degree of accuracy if this layer has enough hidden units. There are also proofs that any function can be learned by a network with two hidden layers, provided there are enough hidden units in each layer. Finally, most of the theoretical work in behaviorism was vague and speculative; however, the new connectionism is aided by working computational models that clearly indicate data coverage and predictions.

Models of Reading

Connectionism has been more concerned with establishing a theoretical basis for understanding learning than with developing applications in the field of education. So far, the area of education best modeled by connectionism is that of reading. Reading is an unnatural act for humans that typically requires several years of education. A debate about whether it is better to teach reading by the rules of letter-to-sound correspondence or by learning to visually recognize whole words stretches back to the 1960s. There are hundreds of such phonic rules, but because letter-to-sound correspondence is only somewhat regular, they are not that useful to learn. There is some regularity in reading aloud, as in the pronunciations of the words *lint*, *mint*, and *bint*, but plenty of exceptions, as in *pint*.

A theoretical framework for an influential series of connectionist simulations of reading is pictured in **Figure 3**. The rectangles represent groups of network units encoding information on spelling, sound, or meaning, or hidden units that effect nonlinear transformations between these three encodings. The bidirectional arrows indicate connection weights in both directions. By adjusting these connection weights, the system might learn to transform written words into meanings or pronunciations, meanings into pronunciations or written words, and pronunciations into meanings or written words. So far, research has concentrated on the mapping from written words to pronunciations, that is, reading aloud. These models learned to pronounce the

words they were trained on, such as *gate* and *save*, and generalized successfully to novel words such as *rave*.

Such connectionist models also captured a number of psychological phenomena such as frequency, similarity, and consistency effects. The frequency effect is that common words are read more quickly than rare words. The model correctly predicted that frequency effects are smaller for words with many similar neighbors (e.g., *save*) than for more isolated words such as *sieve*. The consistency effect is that words with regular neighbors are read more quickly than words with irregular neighbors. For example, the word *gave* has a regular pronunciation, but irregular neighbors such as *have*. So words like *gave* take longer to read aloud than words such as *must*, which do not have irregular neighbors. Consistency effects are larger for lower-frequency words and less-skilled readers.

Since all words share the same set of connection weights and neural learning tries to reduce as much error as possible, frequent words, words with high similarity to other words, and words with regular neighbors are read quickly and accurately. Frequency, similarity, and consistency can compensate for each other in that words at a disadvantage in one respect might benefit from another factor. For example, an infrequent word might be read quickly and accurately because of its similarity to other words.

A curiosity of these models is that the network has no explicit representation of lexical entries (words). Not only is it common for language researchers to assume and use a lexicon, but the frequency effect in reading is also customarily explained by storing frequencies within a lexicon. Remarkably, networks exhibit the frequency effect without a lexicon, which shows that a lexicon is not required for this effect.

The recent focus on the mapping from spelling to meaning has enabled connectionist models to address the issue of how to teach reading. In contrast to previous models emphasizing a conflict between visual and phonetic routes, these networks reveal collaboration between the two routes. Early in training, networks relied more on the spelling–sound route (see **Figure 3**); however, with more training, the spelling–meaning–sound route increased in importance, mimicking a psychological progression. As with humans, skilled reading of words involved convergent contributions of both routes from print to sound.

Dyslexia can be simulated in these models by damaging either the network or its training. Reducing the number of hidden units would be analogous to a child with fewer cognitive resources. Limiting the amount of training would correspond to inadequate educational opportunity. Ignoring training in the spelling–sound route would correspond to teaching without phonics. Letting each letter string activate more spelling units could implement a visual impairment. In all of these impaired cases, network learning focuses on the largest current

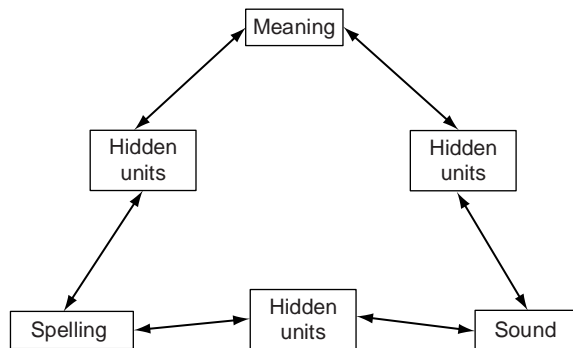


Figure 3 Theoretical framework for Seidenberg's connectionist models of reading.

source of error, namely that contributed by regular words, thus sacrificing exception words.

Other computational models using symbolic rules have relatively more difficulty accounting for these psychological phenomena. The main implication of connectionist models for reading instruction is to provide students with plenty of examples (not rules) of printed words and their meanings and pronunciations.

Models of Mathematics

Mathematics is another example of unnatural skills taught over several years of formal instruction that have been modeled with neural networks. The two examples considered here are learning of the single-digit multiplication table and prime-number detection.

Multiplication

Learning the 0–9 multiplication table requires 5–6 years of schooling and continues to generate errors even in adults. The most widely studied problem is the production task in which two single-digit multiplicands are presented and the

person is asked to provide their product. Several regularities are evident in the psychological literature:

- Computational procedures such as repeated addition ($m \times n =$ adding m , n times) are gradually replaced by recall.
- Reaction time increases with the size of the multiplicands, except that the 5's table and tie problems (e.g., 4×4 , 7×7) are quicker than expected.
- Adults under mild time pressure make about 8% errors.
- Errors are usually close to the correct product, and often substitute a close multiplicand (see Figure 6 for a breakdown of error types).
- There is a strong positive correlation between errors and reaction time across problems.

Building on the successes and limitations of several earlier models, Dallaway designed a feed-forward network learning with multiplicands 2–9 that captured the last four of these psychological phenomena. The topology of his network is shown in Figure 4, and the nature of his course coding of multiplicands is shown in Figure 5. Target output vectors were designed by turning on one product unit and leaving the others off. The percentage of error types plotted in Figure 6 for adults and Dallaway's

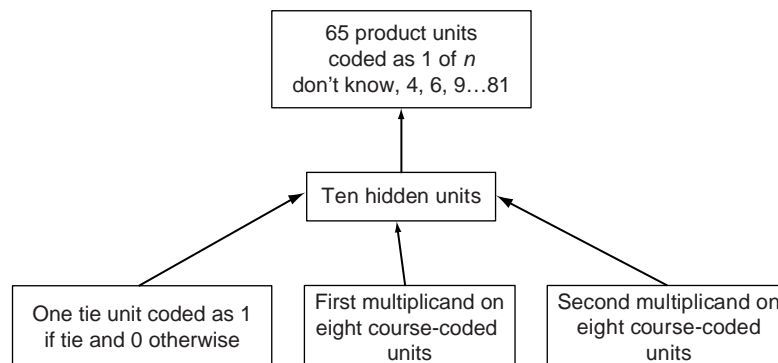


Figure 4 Dallaway's (1994) network for learning the 2–9 multiplication table.

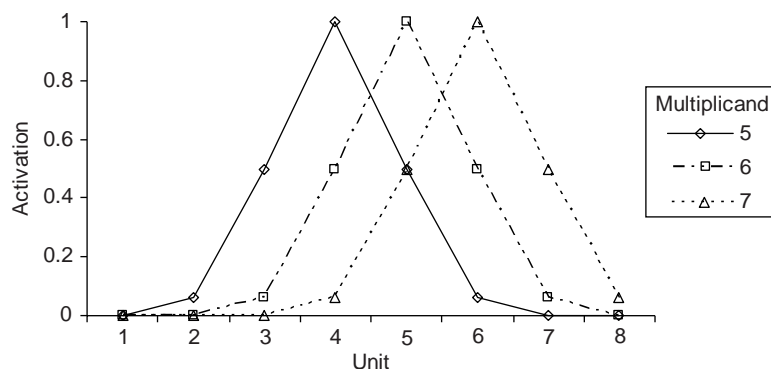


Figure 5 Course coding of the input multiplicands 5, 6, and 7 in Dallaway's (1994) network. Activation representing a given integer is maximal for the unit corresponding to that integer itself and drops off for neighboring values less than or greater than the integer. Only integers from 2 to 9 were included in this simulation; therefore, integer n centers on unit $n-1$.

model indicate a good fit, although the overall error rate was higher for the networks at 14%.

As shown in **Table 3**, operand errors involve changing one of the operands, close-operand errors involve changing to a close operand, and frequent-product errors involve selecting a frequently occurring product. Table errors involve choosing a less-frequent product that is in the multiplication table and does not share any multiplicands with the problem being tested. Operation errors involve adding instead of multiplying. Networks reacted about as quickly to multiplication by 6 as 5, complicating explanation of the 5's speedup. The fit to human data was not as good when the 0 and 1 tables were also trained. Some psychologists believe that multiplication by 0 and 1 is rule governed, rather than being based on connectionist pattern matching; however, it seems possible that the greater regularity of 0 and 1 multiplications just makes them seem rule governed. The fit to human reaction time data worsened when the training sample was not biased in favor of smaller multiplicands.

Applications of this and related models to education remain tentative, but likely would be similar to those made for reading, namely to include many examples of

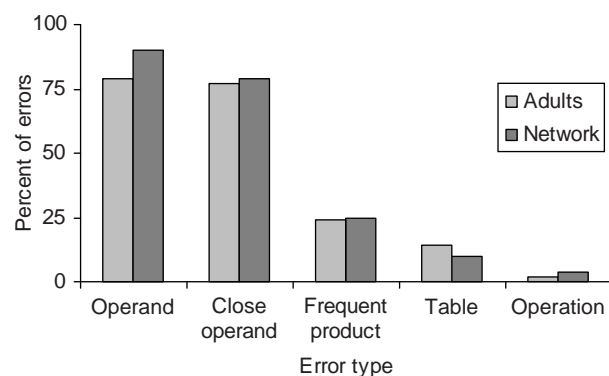


Figure 6 Percent of multiplication errors of different kinds in adults (Campbell and Graham, 1985) and neural networks (Dallaway, 1994). The values do not sum to 100 because error types are not exclusive.

Table 3 Multiplication table with three error types highlighted^a

x	2	3	4	5	6	7	8	9
2	4	6	8	10	12	14	16	18
3	6	9	12	15	18	21	24	27
4	8	12	16	20	24	28	32	36
5	10	15	20	25	30	35	40	45
6	12	18	24	30	36	42	48	54
7	14	21	28	35	42	49	56	63
8	16	24	32	40	48	56	64	72
9	18	27	36	45	54	63	72	81

^aLight-gray-shaded boxes indicate close operand errors for 5×4 ; dark-gray-shaded boxes indicate operand errors for 8×9 ; and frequent-product errors are given in italics.

correct multiplication that go just beyond the student's current ability. The role of addition in learning and understanding multiplication should be explored in future modeling because of its evident role in children's learning and its possible role in several multiplication errors.

Primality

Prime-number detection is a more advanced mathematical skill that has also been modeled with connectionist methods. The primality of an integer n can be determined by seeing if n is divisible by any integers between 2 and the integer part of \sqrt{n} . It is efficient to test in this order because the smaller the prime divisor, the more composites it detects in a fixed range of integers.

A connectionist system called knowledge-based cascade-correlation (KBCC) learned this algorithm from examples by recruiting previously learned knowledge of divisibility. KBCC is based on a simpler connectionist algorithm called cascade-correlation (CC) that learns from examples by recruiting single hidden units. CC and KBCC have simulated a large number of phenomena in learning and cognitive development. KBCC has the added advantage that it can recruit its existing network knowledge as well as single hidden units. Both CC and KBCC are constructive learners that build their new learning on top of existing knowledge.

For primality, the pool of source knowledge contained networks that had previously learned whether an integer could be divided by each of a range of divisors, for example, a divide-by-2 network, a divide-by-3 network, etc., up to a divisor of 20 (see **Figure 7**). These source networks were trained on integers from 2 to 360. Twenty KBCC target networks trained on 306 randomly selected integers from 21 to 360 only recruited source networks involving prime divisors below the square root of the largest number they were trained on (360). Moreover, they recruited these sources in order from small to large, and avoided recruiting single hidden units, source networks with composite divisors, any divisors greater than square root of

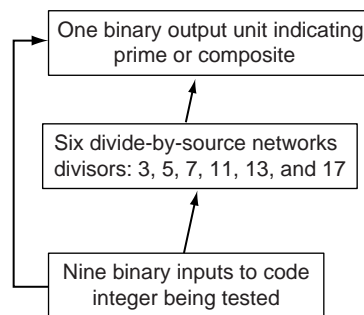


Figure 7 Topology of a KBCC network that learns prime-number detection (Egri and Shultz, 2006).

360 even if prime, and divisor networks with randomized connection weights.

KBCC target networks never recruited a divide-by-2 source network; however, this was because they, instead, used the least significant digit of n , binary coded, to directly determine if n was odd or even. As with people who use the 1's digit of base-10 numbers to check for divisibility by 5 or 10, this is a shortcut to dividing by 2.

The KBCC target networks learned to classify their training integers 3 times faster than did knowledge-free networks, with fewer recruits on fewer network layers, and generalized almost perfectly to novel test integers. Networks without knowledge of divisibility did not generalize better than chance guessing.

As predicted by this simulation, university students testing the primality of integers also mainly used prime divisors below \sqrt{n} and ordered their divisors from small to large. The recommendation for education is not only to use examples, but also to structure curricula so that later learning can build on existing knowledge.

Educational Relevance

As connectionist modeling of learning continues, potential applications to education could become more apparent. Some implications of connectionist models of reading and mathematics were already noted, namely the use of examples with fully specified feedback on what to do with these examples. Teaching with examples is compatible with the long-recognized notion of learning by doing. Since students may vary considerably in their skills, this can be challenging in a classroom, but could be accomplished with materials that vary enough in difficulty to continuously provide at least some examples just beyond the ability of each student.

Accompanying examples with full feedback is more informative than the evaluative feedback provided by rewards or equally vague cues to disequilibrium which were characteristic of the educational recommendations of classical behaviorist theory and Piagetian theory, respectively. Computational results make it clear that having to convert information about being wrong into a useful target vector makes learning both slower and more difficult.

Another educational suggestion stemming from connectionism is that repetition and patience are often required. This stems, in part, from the fact that connectionist learning can be quite slow. Some insight into this slowness has been gained from studying variation in learning-rate parameters. When learning rate is too high, networks may oscillate across error minima. To settle near such minima, it is often necessary to take small steps in connection-weight adjustments. This could also be true of brain networks, and if so, teachers should not rush through difficult material.

Of course, methods for increasing both the speed and accuracy of network learning are under active investigation. Networks learn faster and more accurately if they can recruit relevant existing knowledge. In the case of prime-number detection, successful generalization to untrained integers requires recruiting existing knowledge. This suggests curricula designed to ensure that lessons are optimally ordered. Network simulations might identify the most beneficial lesson sequences.

Another implication of neural modeling is that context is important and that it limits generalization. Connectionist learning algorithms naturally exhibit context effects whenever contextual cues aid learning. However, such contextual effects ensure that generalization may not be as universal as desired. Similar to network researchers, teachers might try to decontextualize learning by varying contextual cues in teaching. Again, exploratory network simulations could help.

Many of these educational suggestions emanating from connectionist research (practice, feedback, prior knowledge, and structured lessons) at first appear more consistent with teacher-centered, rather than child-centered, approaches to education. This seems paradoxical given that constructive connectionist approaches (CC and KBCC) are consistent with a Piagetian approach to knowledge acquisition that is the psychological basis for child-centered education.

Whereas teacher-centered education focuses on structured lesson plans, extensive practice, and feedback, child-centered education emphasizes curiosity, problem solving, and discovery learning. Although these approaches are often portrayed as being opposed, constructivist connectionist modeling suggests a possible rapprochement, with computational demonstrations that effective learning incorporates both methods. In connectionist learning, knowledge representations are constructed and abstracted, rather than memorized. Moreover, this learning is particularly effective when experience is well structured, building more complex ideas on top of earlier simpler ideas, and well practiced, with detailed feedback.

See also: Constructivism and Learning; Neuroscience Bases of Learning; The Neuroscience of Reading.

Bibliography

- Campbell, J. I. D. and Graham, D. J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology* **39**, 338–366.
- Chall, J. S. (2000). *The Academic Achievement Challenge: What Really Works in the Classroom?* New York: Guilford.
- Dallaway, R. (1994). *Cognitive Science Research Papers 306: Dynamics of Arithmetic: A Connectionist View of Arithmetic Skills*. Brighton: University of Sussex.
- Egri, L. and Shultz, T. R. (2006). A compositional neural-network solution to prime-number testing. In *Proceedings of the*

- Twenty-Eighth Annual Conference of the Cognitive Science Society*, pp 1263–1268. Mahwah, NJ: Erlbaum.
- Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition* **28**, 3–71.
- Foorman, B. R. (1994). The relevance of a connectionist model of reading for “the great debate” *Educational Psychology Review* **6**, 25–47.
- Hertz, J., Krogh, A., and Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*. Reading, MA: Addison-Wesley.
- Lemaire, P. and Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children’s learning of multiplication. *Journal of Experimental Psychology: General* **124**, 83–97.
- Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science* **14**, 238–242.
- Shultz, T. R. (2003). *Computational Developmental Psychology*. Cambridge, MA: MIT Press.
- Shultz, T. R. and Rivest, F. (2001). Knowledge-based cascade-correlation: Using knowledge to speed learning. *Connection Science* **13**, 1–30.
- Sun, R. (ed.) (2008). *Cambridge Handbook of Computational Cognitive Modeling*. Cambridge: Cambridge University Press.
- Thorndike, E. (1913). *Educational Psychology: The Psychology of Learning*. New York: Teachers College Press.
- Thorndike, E. (1932). *The Fundamentals of Learning*. New York: Teachers College Press.
- Elman, J. L., Bates, E. A., Johnson, M. H., and et, al (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.
- Harm, M. W. and Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review* **106**, 491–528.
- Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*, 2nd edn. Upper Saddle River, NJ: Prentice-Hall.
- Plaut, D. C. (1999). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science* **23**, 543–568.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review* **103**, 56–115.
- Rumelhart, D. E., Hinton, G. E., and McClelland, J. L. (1986). A general framework for parallel distributed processing. In Rumelhart, D. E. and McClelland, J. L. (eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, pp 45–76. Cambridge, MA: MIT Press.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In Rumelhart, D. E. and McClelland, J. L. (eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, pp 318–362. Cambridge, MA: MIT Press.

Further Reading

- Anderson, J. A. (1995). *An Introduction to Neural Networks*. Cambridge, MA: MIT Press.
- Arbib, M. A. (ed.) (1995). *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA: MIT Press.
- Dayan, P. and Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: MIT Press.

Relevant Website

- <http://www.psych.mcgill.ca/perpg/fac/shultz/personal/default.htm> – LNSC Cascade-correlation Simulator Applet: Allows users to run constructive and static neural networks on canned and designed learning problems, without programming; Tutorial on cascade-correlation; Tutorial on knowledge-based cascade-correlation.

Constructivism and Learning

S Sjøberg, University of Oslo, Oslo, Norway

© 2010 Elsevier Ltd. All rights reserved.

The Many Faces of Constructivism

The term constructivism is used in different fields and with many different meanings. If one looks it up in an encyclopedia, one gets somewhat disturbing results. In Encyclopaedia Britannica (2008, DVD version) constructivism is discussed in 29 articles. Most of these, including the article with the heading Constructivism, are related to a Russian art movement, and none of them are related to the constructivism we talk about in education. Similarly, Encarta (2008, DVD version) gives 30 entries, none of which are related to education.

Searches on the Internet provide entries in millions, even when limiting the searches by adding education or learning. One also notes that most of the educational entries are related to the teaching and learning of science, technology, and mathematics, although many also relate to other areas. Many hits relate to informal teaching and learning as found in science centers.

The level of precision of the term constructivism has led some critics (e.g., Matthews, 1994) to consider the term to be empty of meaning and its use purely ideological. It seems, according to Matthews, to be used to distinguish the good guys (constructivists) from the bad guys (traditionalists). Many critics say that the label constructivist teaching is used by many authors as more or less synonymous to any teaching that is somewhat child-centered, caring, inclusive, or based on enquiry, discovery, or any kind of active involvement from the learners. The literature abounds with lists of aspects that characterize constructivist classrooms, teachers, curricula and assessment. Most of these articles and books have a low precision on the definition of the term but they all seem to associate the term with something unquestionably positive.

Based on such observations, many critics argue that constructivism as a meaningful concept has lost its power. Some call constructivism a new orthodoxy, a fad and a fashion, a movement (Erickson, 2001), or even a religion with different sects (Phillips, 1995).

However, there is, of course, also serious theoretical writing and research that strongly opposes such characterizations. Many academics claim that there is a strong theoretical underpinning of constructivism. But they also disagree with one another about the epistemological and theoretical status of constructivism.

One should also note that even within the field of education, there are several varieties over the theme of

constructivism. Many scholars use qualifiers when they refer to constructivism. Hence, we find individual and cognitive constructivism (often with reference to Jean Piaget) and social constructivism (often with reference to Lev Vygotsky). Some use the term simple, mild, or even naive constructivism with reference mainly to some interpretations of Piaget, and with a contrast to radical constructivism, used by von Glasersfeld (1984). Other widely used versions include contextual constructivism (Cobern, 1993), sociotransformative constructivism (Rodriguez, 1998), and sociocultural constructivism (Tobin, 1998). The list can be made longer, and many of the above qualifiers are used in new and inventive combinations. It is beyond the scope of this article to go in detail on differences and similarities behind this flourishing terminology. The point is simply to warn the reader about the possibility for misunderstandings, as well as for real and false disagreements.

Constructivism: The Construction of What?

In this article, we discuss constructivism as a theory of learning, hence, we debate on how people construct meaning and knowledge. It is important to distinguish this from epistemology of scientific knowledge, that is, the growth, development, and status of scientific knowledge about the world.

We may ask what is constructed? Is it:

1. our individual knowledge about the world? (e.g., children construct their own knowledge)
2. the shared and accepted scientific knowledge about the world as it exists in established science? (e.g., scientific knowledge is socially constructed) or
3. the world itself? (e.g., the world is socially constructed)

The first of these questions is a problem of psychology and educational or learning theory, while the latter two are part of philosophy and epistemology. Question no. 2 is also addressed by the sociology of knowledge and also science.

Analytically, it is important to keep these questions apart. One may, for instance, be a strong supporter of constructivist learning theories while at the same time reject the two other stances, in particular the last and most extreme one. This latter kind of constructivism is criticized for being a subjectivist and relativist postmodern attack on

the rationality of science (Boghossian, 2006), a stance that certainly runs against any suggestions from, for instance, Piaget and Vygotsky to be discussed later in this article.

Much confusion and disagreements occur because one does not keep the fundamental differences between the nature of these constructivist claims in mind. A recent book title is called *Teaching Constructivist Science* (Bentley *et al.*, 2007). While the book is about constructivist methods of teaching, the title may suggest that the authors claim that science itself is constructed.

Constructivism and Learning – Core Ideas

Within the large family of constructivist-learning theorists, there are some ideas that more or less all subscribe to. Hence, they may be seen to be a mild version of constructivist claims. Here are some of these core ideas, partly based on the analysis of Taber (2006).

1. Knowledge is actively constructed by the learner, not passively received from the outside. Learning is something done by the learner, not something that is imposed on him.
2. Learners come to the learning situation (in science, etc.) with existing ideas about many phenomena. Some of these ideas are *ad hoc* and unstable; others are more deeply rooted and well developed.
3. Learners have their own individual ideas about the world, but there are also many similarities and common patterns in their ideas. Some of these ideas are socially and culturally accepted and shared and are often part of the language, supported by metaphors, etc. They also often function well as tools to understand many phenomena.
4. These ideas are often at odds with accepted scientific ideas and some of them may be persistent and hard to change.
5. Knowledge is represented in the brain as conceptual structures and it is possible to model and describe these in some detail.
6. Teachers have to take the learner's existing ideas seriously if they want to change or challenge these.
7. Although knowledge in one sense is personal and individual, the learners construct their knowledge through their interaction with the physical world, collaboratively in social settings and in a cultural and linguistic environment. (The relative stress on such factors account for the different versions of constructivism earlier alluded to.)

The Many Phases of Constructivism in Education: A Brief History

The following description is based on how these ideas have emerged, with emphasis on the field of science education.

Piaget and Constructivism

Many of the core ideas of constructivism are indeed very old, but it may be fruitful to present a historical account based on the influence of the Swiss psychologist and epistemologist Jean Piaget (1896–1980). Few intellectuals have been so widely used and misused, or understood and misunderstood as Piaget. A main reason for this is that his research and perspectives do not fit well into the established boundaries of academic disciplines.

A brief version of his life story exemplifies this: at the age of 21 he had a PhD in biology, and his specialty was how organisms adapted to their environment. He was also deeply interested in philosophy. For some years he worked in psychoanalysis with CG Jung, and later in testing of intelligence in the behaviorist tradition of Stanford-Binet and Cyril Burt. Piaget was, simultaneously or in succession, professor in several fields: in psychology, in sociology, in the history of science, and in genetic and experimental psychology. Moreover, he used mathematical logic and group theory to express his ideas. He even contributed to the development of this field. Most of his examples of the intellectual development came from observations and clinical interviews with children when they manipulated physical, technical, and chemical objects in experiments that were rather similar to traditional textbook examples (pendulum, balance, etc.).

Piaget's biological background and language, his use of examples from physics and technology as well as his wish to formulate general theories in a mathematical way may explain why Piagetian theory has a strong appeal to science and mathematics educators. He had the greatest admiration for physics as a discipline and considered it to be the most mature and developed of all scientific fields. He also corresponded and met with Albert Einstein, who was fascinated with Piaget's ideas.

As one can understand, Jean Piaget covered several academic disciplines. It is, however, a paradox that the field where he has had most influence in is education. He was not at all an educator, and actually wrote very little on teaching and pedagogy.

Piaget was academically active for more than half a century, and wrote more than 50 books and several hundred articles. Although he remained faithful to his initial research agenda, his theories developed during this period. When one tries to understand (or to critique) Piaget's theories, one must know which age of Piaget one has in mind. The early Piaget is very different from the older, and in many ways he was his own best critic. Some of his books were rather badly translated from his original in French to English, and then translated from English to other languages. Such factors may also have added to the confusion over his ideas. One might say that Piaget's strong interdisciplinary orientation was his strength, but it was also his problem because it made his thinking

difficult to access for people with a more typical (and narrow) academic background.

However, Piaget's basic research problem remained the same all the time. His problem was epistemological and philosophical: What is the nature of knowledge? How does it grow and develop? Piaget's epistemological research agenda is well reflected in the name of the institution that he established in Geneva in 1955: Centre International d'Epistémologie génétique (International Center for Genetic Epistemology.) Here, of course, genetic refers to the genesis and development of knowledge, and not to biological heredity.

The nature of knowledge should, according to Piaget, be studied empirically where it is actually constructed and develops. This can be done either through the historical development of knowledge, as it is found in well-established sciences (in particular, physics and mathematics), or may be studied in the growth and development of an individual. We may therefore say that Piaget's study of the development of children was in effect only to get empirical access to his epistemological research question: the growth of knowledge and the development of logical thinking.

Piaget's earlier writings, from the 1920s on, were on children's conceptions of the world. This was, in fact, also the title of one of his earlier books (Piaget, 1929). Already at that time, he had started to use the term constructivism. A very influential book had the title *La Construction du réel chez l'enfant* (*The Construction of Reality in the Child*). The French original came out in 1937, with the Introduction written as early as 1925. His other writings from the same period (late 1920s to 1940) included detailed studies on how children developed (or constructed) ideas about time, space, causality, logic, numbers, movement and velocity, conservation, and a long list of various natural phenomena.

As we can see, Piaget gave meaning to the term constructivism long before it was used by academics in other fields, like the psychologist George Kelly (Kelly, 1955) and sociologists Berger and Luckman (Berger and Luckmann, 1967).

He remained faithful to this constructivist perspective throughout his long active period. In his last publication, written the year that he died, some 55 years after his first use of constructivism, he summarizes his lifelong program: his task had been to

establish what we have called a *constructivist* theory of knowledge and, at the same time, refute the empiricist and nativist theories. The essential problem of a theory of knowledge is: How is new knowledge constructed? Is it, as empiricism contends, always derived from observing reality, or is it preformed in the human mind, and thus innate? Even our earlier work, I believe, clearly showed the insufficiencies of both the empiricist and preformist theories. (Piaget, 1980: 3)

In this statement, Piaget locates his own constructivist epistemology. He rejects the empiricist and behaviorist stance that knowledge derives directly from sense experiences. He also rejects the rationalist or preformist view that knowledge is innate and develops more or less biologically as we grow and mature. His whole life program consists of showing how both these views are insufficient.

Piaget developed his theory of knowledge based on ideas derived from biology like the process of adaptation, consisting of assimilation and accommodation. Other Piagetian concepts like self-regulation also indicate his basic belief that the development of intelligence and thinking should also be understood as the individual's biological adaptation to the external world.

As noted, Piaget was not very interested in education, let alone in teaching (Solomon, 1994). Nevertheless, his theories have probably been more widely used (and misused) in education than in other fields. As also noted, the main uses have been in science and mathematics education because of reasons explained earlier. But even in these areas, his theories have been used in a wide and often contradictory ways. Let us briefly look at this development.

Piagetian Constructivism Emerges

Piaget's theories were discovered by science educators in the early 1970s. At that time, most attention was given to his stage theory. A recurring theme in Piaget's writing was the description of stages of intellectual development that each individual had to pass. Names like the psychomotor, the intuitive, and the concrete operational and the formal operational stage soon became part of educational terminology.

He asserted, for instance, that an individual at the concrete operational level was able to understand and perform certain logical operations, while the spectrum was much wider at the formal operational stage. At this stage, the individual could reason in a way that is close to ways commonly used in science, like thinking with hypotheses and abstract models, understanding proportionality, and control of variables etc. These stages are described in Piaget's publications (Piaget and Inhelder, 1958). On the basis of these descriptions, educators developed written tests that they used to classify the learners by their Piagetian level. They also developed teaching material that was supposed to match the level of abstraction of the learner. This so-called matching model became very influential for a long period. Numerous studies based on the Piagetian stage theory were published in the journals for science and mathematics education, and large curriculum development projects were carried out. In the US, the Science Curriculum Improvement Study (SCIS) project (SCIS, 1974) was based on Piagetian stage theory. This program was also exported to several other countries and had some success. It is still in use some places.

The British Science 5/13 project had a similar foundation (Science 5/13, 1972).

Some of the strongest supporters of the stage theory claimed that this theory and its practical applications had turned science education into a real science. “Towards a science of science education” was the ambitious title of a book by two of the British pioneers in the field (Shayer and Adey, 1981). This book presented experiences from a large research and development program based on Piagetian stage thinking. They also provided evidence that the development from one stage to the next could be accelerated or enhanced by certain learning experiences and tasks (Shayer and Adey, 1992).

However, although this theoretical stance and the corresponding curriculum projects gave inspiration and positive impetus to science education, the stress on Piagetian stage theory gradually lost supporters and waned in popularity. In short, the approach did not hold what it promised. The theoretical underpinning was also under attack. Most important, however, is probably new and more promising developments.

The new turn in the thinking in science education may be exemplified by a seminal article written jointly by the American Jack Easley and his former PhD student Rosalind Driver from the UK (Driver and Easley, 1978). In hindsight, we may say that this article is a starting point for what we now label the constructivist tradition in science education.

In their article, Driver and Easley reviewed research on children’s ideas and perceptions regarding natural phenomena, and also provided a language and theoretical perspective to talk about the findings and their educational aspects. The article was strongly influenced by Piagetian theory and triggered a new interest for the actual contents of children’s ideas. While the Piagetian stage theory was centered around the more formal and logical aspects of children’s thinking, we now got a new interest for what concrete ideas that children bring. They also anchored their interpretations to Piaget’s constructivist epistemology as well as to his early studies of children’s ideas.

In the following years, the research agenda in science (and mathematics) education changed gradually. There was a growing concern about the ideas that children develop about the phenomena in the physical world (as well as other phenomena), and the term constructivism started to occur frequently. Rosalind Driver was a key person in the development of this constructivist movement. Alone, and in cooperation with colleagues from several other countries, she published academic articles (e.g., Driver and Oldham, 1986) as well as books meant for the classroom teacher (e.g., Driver, 1983). These (and similar) books are probably the main reason why the ideas behind constructivism soon became highly influential, not only in research, but also in teacher-training and in science classrooms.

Constructivism and Children’s Ideas

Although it had been known for a long time that pupils have their own ideas, often at odds with those of established science, about natural and other phenomena, the emergence of constructivism in the late 1970s and early 1980s triggered a virtual flood of such studies. Children’s ideas for all thinkable phenomena (like force and movement, electricity, heat, etc.) were studied in great detail. This also implied a rediscovery of the early Piaget (from the late 1920s and some decades to follow), where he had published many books on these themes. Also Piaget’s research methodology, the clinical interview, got a revival, but was also supplemented by many other research methods, qualitative as well as quantitative.

The many names given to children’s (and indeed adults’) ideas reveal that different researchers understood and interpreted their findings from very different theoretical perspectives. The literature reported on children’s ideas, their alternative paradigms, mini-theories, conceptions, misconceptions, etc. The variety in terminology reflects ambiguity and some confusion over the status of what one observed. This, of course, also has implications on how to act on such observations. If you face a misconception, you can simply try to correct the misunderstanding. If, on the other hand, the child’s explanation has the character of being a paradigm, an alternative way of seeing the world, well integrated with other strongly held beliefs, then the educational task is indeed very different!

However, the status of the ideas that children express may indeed differ from one type of phenomenon to another, and this is an empirical question that may be clarified by research. In some areas, one may find that the expressed ideas are very loose, often *ad hoc*, and even invented in the interview setting. Piaget himself wrote about these challenges in the 1920s. For other types of phenomena, children’s explanations are more deeply rooted, well integrated, and systematically used to understand a wide class of experiences. In this case, they may be very resistant to change, and sometimes one should think twice before embarking on such a project.

The research on children’s ideas, often under the theoretical umbrella of constructivism and/or conceptual change proliferated from the early 1980s. Reinders Duit at IPN, the German Institute for Science Education in Kiel, maintains a bibliography that, according to the Introduction is “an attempt to document constructivist research in science education”. The 2009 version contains 8400 research articles, a clear indication that this is by now a dominant perspective in science education research.

Constructivism: Widening the Perspective

The influence from Piaget is evident in nearly all the written reference to constructivism, especially in

the early phases. But Piaget's own perspective was, as noted, mostly on the general aspects of the development of knowledge *per se*. He was not so much interested in education, let alone teaching or conditions for good and effective learning. Constructivism has, as noted, developed from such a Piagetian perspective, and has drawn on other theorists who put more stress on social and cultural conditions for learning. This may explain why we now have so many varieties of constructivism.

A main contributor to this development has been from a contemporary of Jean Piaget, the Russian Lev Vygotsky (1896–1934) and his students, mainly Luria and Leontiev. Vygotsky's work remained virtually unknown in the West until its rediscovery in the 1960s, when the translation of *Thought and Language* from 1934 was published in English in 1962 (revised edition in 1986). It was not until the end of the 1970s that his works started to get attention, and his collected works were available in English only in the late 1990s.

Many books have been written where Piaget and Vygotsky are seen as more or less as opposites. It may be more productive to note some fundamental similarities. They may both clearly be seen as constructivists. Some of the differences between the two can be explained by the fact that they had rather different research agendas. While Piaget was interested in epistemology and knowledge *per se*, Vygotsky was more interested in understanding the social and cultural conditions for human learning. Hence, his writings may be closer to the concerns of educators.

With his stress on the social and collaborative nature of learning, Vygotsky is often considered to be the father of social constructivism, while Piaget is often classified as a father of personal (or cognitive) constructivism. It may be interesting to note that Piaget and Vygotsky corresponded, and that Piaget acknowledged to have been inspired by Vygotsky's ideas.

Constructivism: Paradigm or Research Program?

We have noted that constructivism is a dominant perspective in fields like science education. How profound is this influence? Some authors write about constructivism as a paradigm in a Kuhnian sense (Kuhn, 1970). Others find it more proper to use the terminology of the philosopher Lakatos (1970). They therefore argue that constructivism represent a so-called research program.

For this author, the term paradigm in the Kuhnian sense seems to be too strong. A paradigm implies that the idea is so strong that it completely dominates an entire field or discipline, and that other ideas practically do not exist (like the paradigm of evolution in biology, or of atoms or quanta in physics). Therefore, the term research

program seems more suitable. Lakatos (1970) has developed these ideas in detail, but here are some basic ideas. A research program is a set of ideas that provide a platform of common assumptions and ideas about certain phenomena. This core of commitment is shared among those working in this research program. (Core ideas of constructivism have been suggested earlier in this article.).

Several research programs may coexist within the same field over time. A research program provides direction and structure of a field of research. It contains a positive heuristic that may generate new knowledge and ideas for further research. When a research program develops, grows, and adds new knowledge, it may be called progressive. When it ceases to do so, it degenerates and may fade away, particularly if other research programs are seen to be more promising for new research.

Many authors have used the concept of research programs on constructivism. Driver and Easley (1978) used the term in their seminal article in 1978. Joan Solomon wrote about the rise and fall of constructivism as a paradigm as well as a research program in 1994. In spite of this, the influence of constructivist ideas and studies continued to grow in the following years. Erickson revisited the theme and strengthened the claim of constructivism as a still progressive research program in 2001, and these ideas are supported in a recent review article of the whole field of teaching and learning science (Taber, 2006).

The core set of ideas of constructivism suggested earlier in this article seem to be widely accepted by many. This is, in itself, a sign of their current dominance. But a set of principles for learning does not directly translate into a set of recommendations for good teaching. One cannot logically deduce a scientifically based pedagogy from a theory of learning.

Many books claim to present constructivist teaching, often in the form of lists of concrete recommendations. Many authors also give detailed prescriptions regarding constructivist curricula and constructivist assessments. It is beyond the scope of this article to enter into an analysis and critique of these widely differing recommendations, but can only advise the reader to exercise some care regarding such assertions.

See also: Curriculum and Constructivism; Learning Science; Mathematics Learning; Piaget: Recent Work; Vygotsky and Recent Developments.

Bibliography

- Bentley, M. L., Ebert, S. E., and Ebert, C. (2007). *Teaching Constructivist Science*. Thousand Oaks, CA: Sage.
- Berger, P. and Luckmann, T. (1967). *The Social Construction of Reality*. New York: Doubleday.

- Boghossian, P. (2006). *Fear of Knowledge: Against Relativism and Constructivism*. Oxford: Oxford University Press.
- Coburn, W. (1993). Contextual constructivism. In Tobin, K. (ed.) *The Practice of Constructivism in Science Education*, pp 51–69. Washington: DC: AAAS.
- Driver, R. (1983). *The Pupil as Scientist?* Buckingham: Open University Press.
- Driver, R. and Easley, J. (1978). Pupils and paradigms: A review of literature related to concept development in adolescent science students. *Studies in Science Education* 5, 61–84.
- Driver, R. and Oldham, V. (1986). A constructivist approach to curriculum development in science. *Studies in Science Education* 13, 105–122.
- Erickson, G. (2001). Research programmes and the student science learning literature. In Millar, R., Leach, J., and Osborne, J. (eds.) *Improving Science Education – the Contribution of Research*, pp 271–292. Buckingham: Open University Press.
- Kelly, G. A. (1955). *The Psychology of Personal Constructs*. New York: W.W. Norton.
- Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. Chicago, IL: Chicago University Press.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In Lakatos, I. and Musgrave, A. (eds.) *Criticism and the Growth of Knowledge*, pp 91–196. London: Cambridge University Press.
- Matthews, M. (1994). Discontent with constructivism. *Studies in Science Education* 24, 165–172.
- Phillips, D. (1995). The good, the bad, and the ugly: The many faces of constructivism. *Educational Researcher* 24(7), 5–12.
- Piaget, J. (1929). *The Child's Conception of the World*. London: Paul Trench and Trubner.
- Piaget, J. (1937). *La construction du réel chez l'enfant* (In English, 1955: *The Construction of Reality in the Child*). Paris: Delachaux et Niestlé.
- Piaget, J. (1980). *Cahier de la foundation archives*. Jean Piaget: CIEG.
- Piaget, J. and Inhelder, B. (1958). *The Growth of Logical Thinking, from Childhood to Adolescence*. London: Routledge and Kegan Paul.
- Rodriguez, A. J. (1998). Strategies for counterresistance: Toward sociotransformative constructivism and learning to teach science for diversity and for understanding. *Journal of Research in Science Teaching* 35, 589–622.
- Science 5/13 (1972). *With Objectives in Mind*. London: MacDonald Educational.
- SCIS (1974). *Teachers' Handbook*. Berkeley, CA: Lawrence Hall of Science.
- Shayer, M. and Adey, P. (1981). *Towards a Science of Science Teaching. Cognitive Development and Cognitive Demand*. London: Heinemann Educational Books.
- Shayer, M. and Adey, P. (1992). Accelerating the development of formal thinking in middle and high school students 3: Testing the permanency of effects. *Journal of Research in Science Teaching* 29, 1101–1115.
- Solomon, J. (1994). The rise and fall of constructivism. *Studies in Science Education* 23, 1–19.
- Taber, K. S. (2006). Beyond constructivism: The progressive research programme into learning science. *Studies in Science Education* 42, 125–184.
- Tobin, K. (1998). Sociocultural perspectives on the teaching and learning of science. In Larochelle, M., Bednarz, N., and Garrison, J. (eds.) *Constructivism and Education*, pp 195–212. Cambridge: Cambridge University Press.
- von Glasersfeld, E. (1984). An introduction to radical constructivism. In Watzlawick, P. (ed.) *The Invented Reality*, pp 17–40. London: W. W. Naughton.

Further Reading

- Driver, R., Asoko, H., Leach, J., Mortimer, E. F., and Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher* 23(7), 5–12.
- Duit, R. (2009). *Students' and Teachers' Conceptions and Science Education*. <http://www.ipn.uni-kiel.de/aktuell/stcse/stcse.html> (accessed September 2009).
- Duit, R. and Treagust, D. (1998). Learning in science – from behaviourism towards social constructivism and beyond. In Fraser, B. and Tobin, K. (eds.) *International Handbook of Science Education*, pp 3–26. Dordrecht: Kluwer Academic Publishers.
- Fensham, P., Gunstone, R., and White, R. (eds.) (1994). *The Content of Science: A Constructivist Approach to its Teaching and Learning*. London: Falmer Press.
- Larochelle, M., Bednarz, N., and Garrison, J. (eds.) *Constructivism and Education*. Cambridge: Cambridge University Press.
- Matthews, M. (ed.) (1998). *Constructivism in Science Education. A Philosophical Examination*. London: Kluwer Academic Publishers.
- Matthews, M. (1992). Constructivism and empiricism: An incomplete divorce. *Research in Science Education* 22, 299–307.
- Piaget, J. (1983). *Psychogenesis and the History of Science* (with Rolando Garcia). New York: Columbia University Press.
- Steffe, L. and Gale, J. (eds.) *Constructivism in Education*. Hillsdale, NJ: Erlbaum.
- Tobin, K. (ed.) (1993). *The Practice of Constructivism in Science Education*. Washington, DC: AAAS Press.

Relevant Website

- <http://folk.uio.no> – Svein Sjøberg, Universitetet i Oslo.
- <http://folk.uio.no/sveinsj/>.

Cultural–Historical Activity Theory

A Havnes, University of Bergen, Bergen, Norway

© 2010 Elsevier Ltd. All rights reserved.

The cultural–historical activity theory (CHAT) has its origin in the work of Lev Semenovich Vygotsky (1896–1934) and has since been developed, first by his colleagues and students in USSR, and later in the Western world. In USSR, Vygotsky's ideas had been suppressed during the Stalin period, but revitalized in the post-World War II period. Until the publishing of *Mind in Society* in 1978 (edited by Cole, John-Steiner, Scribner, and Souberman), little was known about the so-called Soviet school of psychology in the West. At this time, around 1980, Western scholars had become receptive to the ideas expressed in *Mind in Society*. *Thought and Language* had been translated into English in 1962, but had not attracted much interest until then.

Vygotsky described his approach as cultural–historical. Alexei N. Leontiev (1904–1979) developed the notion of activity. His approach has been termed the activity theory. Later, CHAT had been further developed by Western researchers, on the basis of the conceptualizations of both Vygotsky and Leontiev. Vygotsky's notion of culture, history, and mediation, and his emphasis on speech and thought and self-directed action has been revitalized. As a consequence of these reconceptualizations, the term CHAT is now commonly used.

While Vygotsky, Leontiev, and their colleagues were immersed in psychology – particularly educational psychology – the interest in activity theory has expanded and is now frequently applied in, for instance, educational research, organizational studies, linguistics, and human–computer interaction. This article classifies three main categorizations of CHAT: (1) the initial formulation by Lev Vygotsky in the early history of USSR where the basic principles of the theory were formulated; (2) the further development and reconceptualization of Vygotsky's ideas after his death, where the concept of activity was developed (Leontiev, 1978); and (3) the more recent development of CHAT in the West, particularly in the context of Northern European and American educational and psychological research.

Because of these reconceptualizations and the wide use of CHAT, it is hard to regard it as a strict and conceptually consistent theory. The relationship between Vygotsky's cultural–historical approach and Leontiev's activity theory is problematic, similar to the relationship between Leontiev's activity theory and the recent application of CHAT in diverse research contexts today. It is not easy to state exactly where the boundaries of CHAT go.

CHAT – A Theory in Development

The developmental stages described above are often referred to as three generations of CHAT (Engeström, 1996; Daniels, 2001), which imply that the history of CHAT represents an expansion of the research platform that Vygotsky developed. The expansion has taken place under different social, historical, and political conditions. It would be inconsistent to present CHAT as a thing in itself, detached from its context. The reconceptualizations can, therefore, also be understood in terms of the recontextualizations, as well as development of the internal logic of the theoretical approach. For instance, Kazunin (2005: 112) explains that Leontiev worked under conditions where “Soviet psychologists were expected to derive psychological categories directly from the works of Marx, Engels and Lenin.” While Vygotsky relied on theories and methods from, for instance, psychoanalysis and Gestalt theory and reframed Western European and American psychological and philosophical research, Leontiev and his colleagues relied on Marxist (political) theory in a more direct sense.

In the West, the social, cultural, and political conditions have been different, and Vygotsky's ideas have been revitalized. Researchers in a wide range of fields have come to see the potential of CHAT to improve their research, but generally without adopting the political agenda associated with Marxism. The theory has been a conceptual tool not only for transforming research agendas and methodologies, but also for making changes in institutions and local communities to improve the conditions for children's learning and development – some of which has a political undertone of supporting underprivileged children and families. Parallel to the growing interest in CHAT, the theory has been developed to accommodate the challenges of theoretical and empirical research under the social and political conditions in these researchers' societies. The history of CHAT illustrates that psychological and educational research develops in relation to the society in which it takes place. In addition, CHAT research has developed in a two-way process where the potentials of the theory open up for new ways of doing research at the same time as its theoretical basis is developed and expanded.

The interest in Vygotsky and CHAT has also merged with a broader interest in the social learning theory. These ideas are now often associated with a spectrum of more or

less overlapping approaches, for instance, sociocultural research, cultural psychology, situated learning, and dialogism. While Vygotsky developed his approach in a context where behaviorist stimulus–response psychology and Pavlov’s reflexology dominated – in opposition to the dominant theorizing of his time – the much-later introduction of his ideas in the West followed in the wake of the so-called cognitive revolution and largely in opposition to the individualist and mentalist perspective that has dominated psychological and educational research over the recent decades.

Lev Vygotsky, the Founder – A Mind in Society

In January 1924, the 28-year-old Vygotsky, from the small province town Gomel in western Russia, and unknown in the community of researchers at the conference, gave a presentation on ‘Methods of reflexological and psychological investigations’ at the Second All-Russian Psychoneurological Congress. He contested the dominant reflexology and behaviorism of the time by emphasizing human consciousness as a mediating factor in human functioning. A main point in his presentation was that by ignoring the problem of consciousness the psychology of that time had deprived “itself of access to the study of some rather complex problems of human behavior. It is forced to restrict itself to explaining no more than the most elementary connections between a living being and the world” (Vygotsky, 1925). He also criticized the deduction of human psychological functioning from animal studies: deducting from animal psychology to human psychology is not possible. The director of the Psychological Institute in Moscow appreciated his approach and invited him to take part in restructuring the institution. From this time on, Vygotsky was devoted to developing an alternative approach to psychology in opposition to the two dominant approaches of the time: objective behaviorism and reflexology and subjective–empirical introspection.

Vygotsky established a school of psychology that was founded on two main agendas: one basically theoretical, that is, to develop a psychology in the philosophical tradition of Marxism and another practical, that is, to find concrete ways of dealing with some of the massive problems in his society, he was particularly concerned about problems related to education, special education, illiteracy, and services to homeless (or so-called difficult) children and those who were mentally retarded. Both of these agendas can be seen in the perspective of the social, political, and scientific circumstances in the society where he had his academic life; postrevolution USSR. This was a period of massive social, economic, and political changes in society accompanied with the rethinking of dominant scientific theories and paradigms. These agendas – the

theoretical and practical – not only drove Vygotsky’s psychological research, but also merged with the mandate associated with restructuring psychological research in his contemporary society in line with the new social and political situation in his society. In a wider, international perspective, this was a period of heated scientific debate as well as optimism and progressivism in the progress of scientific knowing. In this respect, Vygotsky’s work represented a particular contribution to psychology at a particular time and place in society, locally and internationally. He was a mind in society, and a remarkable one at that. When he died from tuberculosis, 37-year-old Vygotsky had written about 180 papers and established a number of research collectives and research centers throughout the country.

The Early Phase – Basic CHAT Ideas

Levels of Development

As discussed, Vygotsky’s work was mainly in psychology; however, he was influenced by his background in literature and semiotics, his interests in European and American psychology (e.g., German Gestalt theory), and in new developments by contemporary biologists and evolution theorists in the USSR. Though he opposed applying findings from animal research to human psychology, and emphasized consciousness and higher mental functions, he had a great interest in animal studies, for instance, Köhler’s studies of chimpanzees and their tool use. Evolution was, in fact, one of the three lines of development that Vygotsky emphasized along with sociocultural development (societies and institutions) and ontogenesis (individuals). A basic assumption in Vygotsky’s approach is that there is no single factor or set of explanatory principles that alone can account for these different lines of development. Instead, there are multiple forces of development, each with its own set of explanatory principles.

Higher Mental Functions

In the article ‘The genesis of higher mental functions’, Vygotsky made a distinction between three stages in the development of behavior. “Instinct, or the innate, inherited fund of behavioral modes, forms the first. The second consists of [...] the stage of training or the stage of conditioned reflexes [...]. Finally, still higher, we have the third stage, the stage of intellect or intellectual responses that fulfill the function of adaptation to new conditions” (Vygotsky, 1981a: 154). Each of these three stages requires a specific set of explanatory principles. The intellectual response is of a different class than the mechanical formation of habits taking place through trial and error. The main problem of psychology is to understand the transformation of elementary processes, starting

with the newborn baby's reflexes, into higher mental functions. Vygotsky attributed two main components to the higher mental functions. The first was the making and use of tools and signs as auxiliary means. Tools provide the ability to change the object of an activity as they are externally oriented toward the mastering of nature, while the signs are internally oriented and aimed at mastering oneself. Second, higher mental functions are socially rooted and historically developed activities. The combination of tools and signs constitutes the unique feature of human psychology; it is the basis for the qualitative leap from animal to human psychology. Human practice is fundamentally mediated by tools and signs that have been created and made accessible through history and culture. Consequently, mental functioning is integral to complex social structures.

The Cultural and Historical in CHAT

The starting point for CHAT, then, is that human psychological functioning is embedded in culture and history – hence the term cultural–historical.

In his discussion of the problem of method in psychology, Vygotsky elaborated on the principles of the genetic or developmental approach. He emphasizes process analysis as opposed to object analysis: it is in movement that a body shows what it is. “By a developmental study of a problem, I mean the disclosure of its genesis, its causal dynamic basis” (Vygotsky, 1978: 62). It implies explaining a phenomenon by investigating its origin and its development instead of its appearance. We need to focus on how the higher forms of mental functioning are established. It implies analyzing individual development and learning in the contexts of social contexts and social practices. Individual development is interpreted as a trajectory of the individual's inclusion (or exclusion) in cultural contexts, and the context is interpreted in the perspective of its history and its interrelatedness with other contexts.

The cultural embeddedness implies that human action and interaction cannot be understood without including the social and cultural context in the analysis. It implies that the unit of analysis is person in context. Vygotsky's colleague, Luria, expressed it this way:

In order to explain the highly complex forms of human consciousness one must go beyond the human organism. One must seek the origins of conscious activity [...] in the social and historical forms of human existence.
(Luria, 1981: 25, quoted from Wertsch, 1998: 8).

Vygotsky formulated what he called the general law of cultural development, which states that the development of higher psychological functions implies mastery of processes that in their origin are social – they are external and social before they become internal. Any function in the child's cultural development appears twice, or on two

planes: first on the social plane and then on the psychological plane. First, it appears in social interaction and subsequently as an individual achievement. From this general law, Vygotsky formed the well-known theses that learning emerges as a consequence of the zone of proximal development:

It is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers.
(Vygotsky, 1978: 86)

Tool-Mediated Action

As discussed, CHAT takes as a starting point that human practice is mediated by tools or signs, which Vygotsky (1981b: 136–137) described as “devices for mastering mental processes [...]”. They are directed toward the mastery or control of behavioral processes – someone else's or one's own – just as technical means are directed toward the control of processes of nature” (Vygotsky, 1978: 55). They are partly externally oriented (changing an external object) and partly internally oriented (mastery of oneself). A variety of terms is used more or less synonymously: tools, signs, mediational means, and recently (cultural) artifacts is commonly used. Vygotsky particularly emphasized the meaning of psychological tools (signs) like language, models, and theories, as well as material tools like hammers, pens, and computers. By using psychological tools, the possibilities for behavior are immeasurably extended. Bateson's example of how the blind man extends his tactile space by using the stick illustrates the point that tools extend the capacity of the user. Wertsch (1998) emphasizes the irreducible relationship between agent and mediational means by defining the unit of analysis as the individual operating with mediational means.

Coming back to the cultural–historical aspect of CHAT, artifacts are seen as cultural tools often offered to the user by someone; mediational means are also ways of conveying the culturally given. Through the use of artifacts, we are linked to culture. Artifacts represent and constitute culture. When we learn to the use of artifacts, we are, in fact, introduced to the activities of prior generations. “Cultural mediation implies a mode of developmental change in which the activities of prior generations are cumulated in the present as the specifically human part of the environment” (Cole, 1996: 145). This means that signs and tools manifest cultural practices and, through their use, human practice and mental representations are linked to these cultural practices.

The first generation or conceptualization of CHAT is illustrated in **Figure 1**. The left part of the **Figure 1** illustrates Vygotsky's initial illustration which reframed

the stimulus–response approach: the simple stimulus–response process is replaced by a complex, mediated act. The relationship between a stimulus and a response goes through the use of mediational means. The right part illustrates the more recent conceptualization. It highlights the triangular interdependent relationship between (1) artifacts, tools, or mediational means; (2) the object or target of activity; and (3) the subject as tool operating and object oriented. The focus of Vygotsky’s conceptualization of CHAT is the genesis of higher mental functions in the perspective of ontogenesis, but with the intention to solve the individualism/social dualism: individual action and development emerge from social and cultural practice; they are an integral part of social practice.

The Notion of Activity in Chat

A. N. Leontiev, who worked with Vygotsky from 1924 to 1930 before being appointed to a position in Kharkov, Ukraine, developed the concept of activity. What does the term activity in CHAT mean?

First, activity includes, but is not identical with, the actions of individual agents. Leontiev differentiated between activity, action, and operation. To illustrate the distinction between activity and action, Leontiev used the example of the role of the beater in primeval collective hunt. The task of the beater was to frighten the animals and send them toward other hunters that were hiding in the ambush. His individual task ended with that, and the rest was completed by the (other) hunters. However, the meaning of his task can only be understood in the context of the collective task and actions, including the beater and the other hunters. “We can say, for example, that the beater’s activity is the hunt, and the frightening of the game his action” (Leontiev, 1981: 210). The activity is constituted by the joint and coordinated actions of more than one actor, often a wide range of persons with diverse roles and expertise. The frightening of the animal did not in itself lead to the satisfaction of the beater’s need for food, but it was needed both for the beater and the other hunters to be able to satisfy their needs. The object of the

coordinated activity was the hunt (which provided meat, skin, etc.) in which object the motive of the activity resided. At the action level, the beater’s intention was to drive it in the direction of the other hunters whose action was to kill the animal.

The second point, then, is that while the actions relate to the intentions of the actors, the activity is driven by a collective motive. To discover the true motive of the participants in the hunt, we need to address the complex coordinated and joint activity of actors with different roles, tasks, and goals. Taking schooling as another, more familiar example of an activity, the coordinated actions of many individuals constitute the activity: in particular, the actions of a number of different subject teachers and the students’ actions in and out of the classroom. “Consequently, it is the activity of the others that provides an objective basis for the specific structure of individual activity” (Leontiev, 1978: 281). It is the activity that attributes the overall meaning to the series of diverse actions.

The third point is that, in an activity, actors also perform a series of more or less automatic operations. Teachers respond immediately to situations or conditions that occur in the classroom, and students write without consciously reflecting on how to hold the pencil or having any present intentions about how to form letters. Such operations represent the third level of an activity. **Table 1** illustrates these different levels of an activity.

Leontiev did not emphasize semiotic mediation – Vygotsky’s point that, through signs and psychological tools, the subjects can master not only the world, but also oneself and others’ relationships to the world. Instead, he emphasized the direct relationship of psychological functioning with reality. Thus, “the subject assumed in [Leontiev’s] categories is the social–historical subject rather than the psychological–individual” (Kazulin, 2005: 117).

Activity System

Engeström (1987) took Leontiev’s approach further, and included the role of semiotic mediation and the mediated act in his reinterpretation of CHAT. His conceptualization of CHAT is frequently referred to in scientific publications (Roth and Lee, 2007). Following Leontiev, he regards activity as the most adequate object of study, but introduced the notion of an activity system. An activity system is a unit of analysis that includes both the active, participatory role of individuals and the impact of the

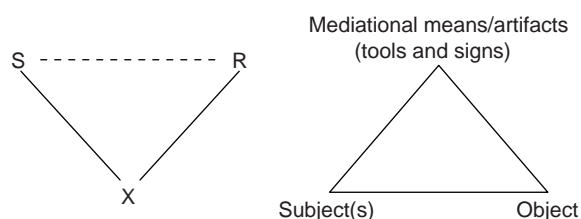


Figure 1 The mediate act and its recent reconceptualization. Modified from <http://www.edu.helsinki.fi/activity/pages/chatand.dwr/chat/>, with permission from the Centre for Activity Theory and Development Work Research, the University of Helsinki.

Table 1 Levels of activity

Level	Driven by	Agency level
Activity	Motive/object	Community
Action	Intention/goal	Individual/group/team, etc.
Operation	Conditions	Routine – potentially dehumanized (machine)

wider social system in which the agents operate. It incorporates the core mediating components of action and activity (Figure 2). The model integrates Leontiev's focus on object-related activity and collective object-driven motive with Vygotsky's emphasis on material and psychological tools and the mediated act.

As participants, we are engaged in social systems (communities) where certain rules apply and where the action and interaction between participants are regulated to some extent (division of labor). The upper part of the triangle is, thus, just the tip of the iceberg. Actions are anchored in social conditions and requirements that are not so clearly visible. The model depicts what we mean when we say that actions are culturally, institutionally, and historically situated and opens up paths for empirical research based on a contextual approach to learning. First, the core component of the activity system is the object of activity, which includes the outcome the system is directed toward. For instance, in studies of health systems, the object can be depicted as the sick patient and the outcome of the activity as the same person healthy again. Second, the model implies that the improvement of the activity presupposes changes in the diverse components of the activity system. Which component(s) that needs to be addressed to initiate change is an empirical question; it

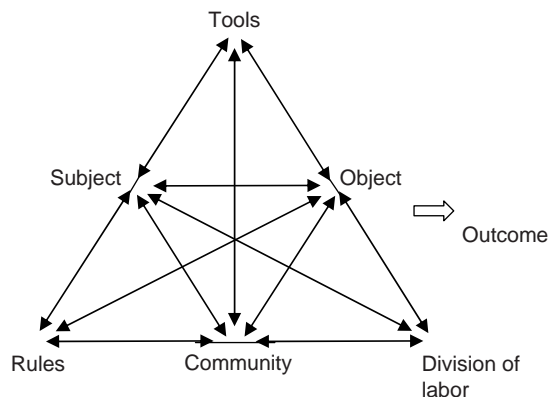


Figure 2 The triangular activity system. Modified from Engeström, Y. (1987). *Learning by Expanding*. Helsinki: Orienta-Konsultit.

will vary from one activity system to another. The model conceptualizes the genesis of social practice, the development of activity – social transformation rather than individual development, and it has led to the development of a strand of research called developmental work research.

Third Generation of CHAT

The reconceptualization of Leontiev's version of CHAT in the Western multicultural, democratic societies took place under other social and political conditions than the ones in USSR. CHAT has appealed to researchers who question the way institutions and societal practices are organized. However, the theory entered a multicultural and multivoiced social and academic context where diversity, dialog, and confrontation between various perspectives and viewpoints were prevalent. Second, while Leontiev, from his focus on psychological processes, stated that “the human individual's activity is a system in a system of social relations”, Engeström made the point that an activity system is also a system within a system of other activity systems. The principle that an individual is not an independent and isolated unit of analysis was applied on activities. To understand the meaning of an activity, we need to take into consideration its interrelatedness with other activities. It means, for instance, that the activity of schooling should be analyzed from the perspective of how it interrelates with, for instance, play, work, family life, or kinship relations out of school. Consequently, the minimal unit of analysis should include the system studied and one or more activity systems with which it interacts. In their interaction, a potentially shared object might occur and, thereby, a shared motive for both collective activity and individual actions (Figure 3). In healthcare, for instance, the focus might be on how primary and secondary health care institutions interact in the treatment of a sick patient.

In the third generalization of CHAT, where transformation of activities and organizational development are in focus, the potential tensions between activities and

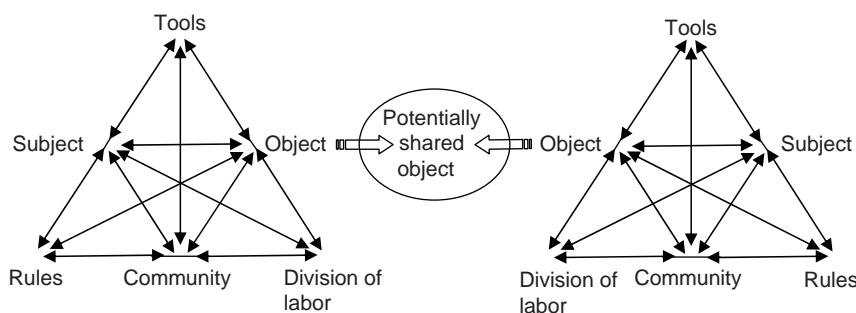


Figure 3 Third generation of CHAT. Modified from Engeström, Y. (1996). Developmental work research as educational research. *Nordic Educational Research Journal* 16, 131–143.

activity systems emerge as objects of study. It can be questioned if the emphasis on collective activity, activity systems, and social transformation has led to undercommunication of individual's practices and development. Stetsenko (2005: 72), for instance, suggests that the major challenge today "is to conceptualize psychological processes avoiding the extremes of reducing them either to a separate individual mental realm [...] or, alternatively, to the essentially sociological realm of collective discourses and practices."

Summing Up

As mentioned, CHAT is an approach that has been, and still is, in a process of development. Its development is not a linear process in the meaning of one continuous movement toward new stages where the next stage or conceptualization takes over from the previous. First, new conceptualizations did not follow directly from the previous. For instance, Leontiev's activity theory diverted, to some extent, from Vygotsky's emphasis on sign-mediated action and the mediating role of language. Second, new developments have also implied reinterpretations of core elements of the earlier versions of the theory. Thus, Vygotsky's initial ideas about, for instance, the social origin of individual development, the zone of proximal development, and mediational means are still under development and subject to reinterpretation. Wertsch (1998) has elaborated on and expanded Vygotsky's notion of mediational means, and Cole (1996), on the role of culture on mental functioning. Leontiev's notion of object-oriented activity has been developed by, for instance, Engeström (1987) and, in a recent special issue of the journal *Mind, Culture and Activity*, the interpretation of the object of activity has been reconsidered by Kaptelinin (2005) and Stetsenko (2005), among others.

It should be stressed that the philosophical and theoretical underpinnings of CHAT did not start with Vygotsky. For instance, Valsiner and van der Veer (2000) have elaborated on the influence on Vygotsky by his contemporaries in USSR, US, and Europe. From the very beginning, CHAT has itself been developed and applied in various cultural-historical phases and conditions. To understand the development of CHAT, it is needed to address both its external relationship to social, cultural, and political conditions and the internal dynamics of the theory – its potentials as a research platform and its inner tensions and dilemmas.

In applying CHAT as a platform for research, it is essential to understand that it is not straightforward and consistent. The three conceptualizations that have structured this presentation of the theory illustrate not

only how it can be applied in studies of both individual development and organizational development, but also how these developmental lines interact.

See also: Apprenticeship Approach to Learning; Computer-Supported Collaborative Learning: Basic Concepts, Multiple Perspectives, and Emerging Trends; Constructivism and Learning; Learning in a Sociocultural Perspective; Vygotsky and Recent Developments.

Bibliography

- Cole, M. (1996). *Cultural Psychology. A Once and Future Discipline*. Cambridge, MA: Harvard University Press.
- Daniels, H. (2001). *Vygotsky and Pedagogy*. New York: RoutledgeFalmer.
- Engeström, Y. (1987). *Learning by Expanding*. Helsinki: Orienta-Konsultit.
- Engeström, Y. (1996). Developmental work research as educational research. *Nordic Educational Research Journal* **16**, 131–143.
- Kaptelinin, V. (2005). The object of activity: Making sense of the sense-maker. *Mind, Culture, and Activity* **12**, 4–18.
- Kazulin, A. (2005). The concept of activity in Soviet psychology: Vygotsky, his disciples and critics. In Daniels, H. (ed.) *An Introduction to Vygotsky*, pp 101–123. New York: Routledge.
- Leontiev, A. N. (1978). *Activity, Consciousness, and Personality*. Englewood Cliffs, NJ: Prentice-Hall.
- Leontiev, A. N. (1981). *Problems of the Development of the Mind*. Moscow: Progress.
- Luria, A. R. (1981). Wertsch, J. V. (ed.) *Language and Cognition*. New York: Wiley Interscience.
- Roth, W.-M. and Lee, Y.-J. (2007). "Vygotsky's neglected legacy": Cultural-historical activity theory. *Review of Educational Research* **77**, 186–232.
- Stetsenko, A. (2005). Activity as object-related: Resolving the dichotomy of individual and collective planes of activity. *Mind, Culture and Activity* **12**, 70–88.
- Valsiner, J. and Van der Veer, R. (2000). *The Social Mind. Construction of the idea*. Cambridge: Cambridge University Press.
- Vygotsky, L. S. (1925). Consciousness as a problem in the psychology of behavior. <http://www.marxists.org/archive/vygotsky> (accessed May 2009).
- Vygotsky, L. S. (1978). *Mind in Society. The Development of Higher Psychological Processes*. Cambridge: Cambridge University Press.
- Vygotsky, L. S. (1981a). The genesis of higher mental functions. In Wertsch, J. V. (ed.) *The Concept of Activity in Soviet Psychology*, pp 144–188. Armonk, NY: Sharpe.
- Vygotsky, L. S. (1981). The instrumental method in psychology. In Wertsch, J. V. (ed.) *The Concept of Activity in Soviet Psychology*, pp 134–143. Armonk, NY: Sharpe.
- Wertsch, J. V. (1998). *Mind as Action*. Oxford: Oxford University Press.

Further Reading

- Bateson, G. (1972). Form, substance and difference. In Bateson, G. (ed.) *Steps to an Ecology*, pp 448–466. New York: Ballantine Books.
- Daniels, H., Cole, M., and Wertsch, J. V. (eds.) (2007). *The Cambridge Companion to Vygotsky*. Cambridge: Cambridge University Press.

- Engeström, Y., Miettinen, R., and Punamake, R. J. (eds.) (1999). *Perspectives on Activity Theory*. Cambridge: Cambridge University Press.
- Leontiev, A. N. (1981). The problem of activity in psychology. In Wertsch, J. V. (ed.) *The Concept of Activity in Soviet Psychology*, pp 37–71. Armonk, NY: Sharpe.
- Moll, L. C. (ed.) (1990). *Vygotsky and Education*. Cambridge: Cambridge University.
- Wertsch, J. V. (1981). *The Concept of Activity in Soviet Psychology*. Armonk, NY: Sharpe.
- Wertsch, J. V. (1985). *Culture, Communication, and Cognition: Vygotskian Perspectives*, pp 35–65. Cambridge: Cambridge University Press.

Wertsch, J. V. (1985). *Vygotsky and the Social Formation of Mind*. Cambridge, MA: Harvard University Press.

Relevant Websites

<http://www.edu.helsinki.fi> – Faculty of Behavioural Sciences.

<http://www.iscar.org> – International Society for Cultural and Activity Research (ISCAR).

<http://www.marxists.org> – Marxists Internet Archive.

Learning in a Sociocultural Perspective

R Säljö, Göteborg University, Göteborg, Sweden

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Issues of learning and development were at the heart of the attempts by Vygotsky (1896–1934) to take psychology out of what he described as its ‘crisis.’ He developed his ideas by dialoguing with the theoretical perspectives presented by many of the distinguished scholars who were active during the rich and dynamic period of psychology of the early twentieth century: Piaget, Janet, Thorndike, Pavlov, Freud, Bühler, and others who we now conceive as Grand Theorists. He was also eager to challenge the basic ideas of the influential schools of psychology of the time, such as behaviorism, pragmatism, and Gestalt psychology. Vygotsky carefully and critically scrutinized the conceptions of learning and development, underpinning these diverse traditions, and pointed to where they, in his opinion, were flawed. Some of these traditions he saw as idealistic and speculative rather than scientific; others he accused of being too reductionistic and failing to address the fundamental problems of consciousness and how the human mind develops and is transformed through social experience. (Informative accounts of the development of Vygotsky’s ideas have been written by Bruner, 1985; Kozulin, 1986; and Leontiev, 1997).

One way to understand some of Vygotsky’s revolutionary ideas on learning and development is to listen to his critique of the dominant tradition of his day in the Soviet Union: pavlovian reflexology. Reflexology, like behaviorism, represents a view of learning which from a philosophical perspective can be described as associationism. The core idea is that learning can be reduced to understanding the associations between stimuli and responses, the so-called $S \rightarrow R$ connection. This link between a stimulus (the presence of food) and a response (salivation by the dog in Pavlov’s famous experiments) was conceived as the atom of learning. Through the principle of conditioning, which implies that a new stimulus (in Pavlov’s case, a bell) is introduced to elicit a response, learning takes place. What was previously a natural response is now a conditioned, that is, learned response. In this tradition, more advanced behaviors can be understood as chains of increasingly complex connections between stimuli and responses. Vygotsky viewed this experimental approach to the study of psychological phenomena, although scientific, as insufficient for understanding how human beings learn and develop. Indeed, the adherence to this perspective on human psychological functioning was a significant part of the ‘crisis.’ This approach is only relevant for explaining what he referred to as the ‘lower psychological functions,’ that is, those

elementary processes which are largely biological and given to us by nature, as it were. A genuinely psychological understanding of phenomena such as learning and development must grapple with specifically human modes of thinking and acting, where language and other cultural tools play a decisive role. These are the “higher psychological functions” (Vygotsky, 1978), which are social, historical, and cultural in their origin and nature.

The Sociocultural Approach

Vygotsky formulated his ideas during a short and hectic period of about ten years. His style of writing and the richness and originality of his ideas led the philosopher Stephen Toulmin to refer to him as the Mozart of psychology in a review of *Mind in Society* when it was published in 1978. Together with his colleagues A R Luria (1902–77) and A N Leontiev (1904–79) he formed the famous troika that was to have a profound influence on psychology and allied disciplines. The history of the sociocultural tradition is also fascinating. The research and the ideas were dormant during the dark ages of political oppression in the East and the intellectual hegemony, even monotony, of behaviorism in the West. The interest resurfaced in a major way in the 1970s and 1980s, thanks to the efforts of Luria and Leontiev and scholars in the West such as Jerome Bruner, Michael Cole, and Jim Wertsch. Today, the influence of the legacy of Vygotsky is stronger than ever and his writings form a rich and inspiring source for developing our understanding of human learning.

One important clue to understanding the pillars on which Vygotsky tried to build a new, and in his view more productive, approach to the study of learning and development can be found in the terms which have been used by his followers to characterize his approach. Some authors refer to it as a sociocultural approach, while others use the terms cultural–historical or sociohistorical. Cole (1996) uses Vygotsky’s ideas as a foundation for what he refers to as cultural psychology. All these expressions communicate something significant about Vygotsky’s original intellectual project: he wanted to study the relationships between the sociocultural and sociohistorical development of human activities and societies on the one hand, and, on the other, what these transformations imply for learning and development of individuals during ontogeny (i.e., their life span). In other words, he wanted to understand how a biological being becomes a sociocultural being equipped

with language and a range of cultural and intellectual skills relevant for social life – skills that have no counterpart in any other species.

Mediation and Cultural Tools

One of Vygotsky's core ideas is that human beings, unlike animals, learn and develop by using cultural tools. Such tools, "language, different forms of numeration and counting, mnemotechnic techniques, algebraic symbolism, works of art, writing, schemes, diagrams, maps" (Vygotsky, 1997: 85) are the products of the development of practices in society over time. They are sociocultural also, in the sense that the nature of these tools may vary between societies. For instance, in some languages one uses some kind of phonetic alphabet for writing, others have syllabic and/or logographic writing systems. These writing systems are different cultural tools, each with their own particularities when it comes to how reading and writing are carried out. In a historical perspective most societies have not used writing at all. Growing up in the latter kind of environment implies that many human activities have to be done differently from what we are used to. For example, one cannot make notes to remember and there can be no such thing as a book of law as an element of legal practices.

Cultural tools develop through an evolutionary process, that is, they have a sociogenesis (de Graaf and Meier, 1994). This idea of the sociogenesis of tools, and of psychological processes, is central to a sociocultural perspective. Vygotsky made a distinction between intellectual/mental (although he used the term psychological) tools and technical or physical tools. Mental tools are symbolic systems used for activities such as writing, counting, measuring, and so on. Examples of physical tools, that is, artifacts, would be rulers, computers, hammers, compasses, etc. (Säljö, 1999). However, and as Cole (1996) argues very convincingly, the distinction between these two kinds of tools is difficult, maybe even pointless, to uphold; a mini-calculator or an abacus is an artifact but both these artifacts, although in different manners, incorporate intellectual tools such as number systems and rules for performing calculations. This ability of externalizing intellectual tools into physical tools is another example of a unique human talent of considerable significance for learning.

At the level of individual action, cultural tools serve as instruments of thinking and acting, to use Vygotsky's own expression. They enter into the flow of human actions and transform the manners in which we reason, communicate, and carry out all kinds of activities. From a psychological point of view, it is one thing to carry out multiplications of four-digit numbers with two decimals as mental arithmetic, another to do it with paper and pencil, and yet another to do it with a mini-calculator (Säljö *et al.*, 2006).

The manners in which we perform such operations differ radically in these situations. What is impossible to do as mental arithmetic becomes almost trivial when there is a calculator present.

In the language of sociocultural theory, cultural tools mediate the world for us in social practices. This notion of mediation is one of the most central concepts in the sociocultural approach. By learning to use cultural tools, we appropriate (Wertsch, 1998) portions of the accumulated experiences and knowing of our society. When learning how to make calculations using decimals or learning how to use a compass, we take over some of the insights and practical knowing that have emerged in our society. These tools are not the inventions of the individual; rather, we come to know and master them because they are used in the society in which we are socialized – and we use many such tools without knowing where they came from or how they emerged. Thus, all readers of this text are familiar with intellectual tools such as percent, decimals, and the = sign, but we may have no idea of their sociogenesis. We take them over as they are, and in this sense the world is pre-interpreted for us. When tools change, such as when we begin to use calculators for complex multiplications, remediation takes place.

For Vygotsky, language is the prime mechanism for mediating what happens in the world, it is "the tool of tools" as he puts it. The relationship between language and thought was one of his most prominent interests, and there are many reasons for this focus on language. One reason why language is so significant for learning and development, both at the collective and the individual level, is that it simultaneously serves as the link between people in interactive settings, and as a tool for thinking. This is how our thinking is socialized or, to put it even stronger, humanized. One of the most quoted passages from Vygotsky's work is the following, where he makes this point:

Every function in the child's cultural development appears twice: first, on the social level, and later, on the individual level; first *between people (interpsychological)*, and then *inside the child (intrapsychological)*. This applies equally to voluntary attention, to logical memory, and to the formation of ideas. All the higher functions originate as actual relationships between human individuals. (Vygotsky, 1978: 57; italics in original)

This formulation indicates how Vygotsky attempted to solve the problem of creating a link between the inner world (thinking) and the outside world (interaction with others). Through language, cultural tools are appropriated (or in Vygotsky's terms: internalized) by individuals in social interaction. Thus, in a sociocultural perspective, thinking is seen as a kind of inner speech carried out by means of cultural tools that we first encountered in communicative practices (Vygotsky, 1986). To think is to engage in internal dialogs with semiotic tools.

Language is also the prime mechanism of semiotic mediation, that is, thinking by means of signs. Sign-mediated communication is another uniquely human phenomenon which, for instance, behaviorist interpretations of learning are unable to deal with. Learning as conceived in behaviorism implies that the organism (be it a human being or an animal) responds directly to stimuli in the external world, that is, in an unmediated manner. Sign-mediated communication instead implies that people think in a roundabout way (Vygotsky, 1994: 61) by means of cultural tools. We, metaphorically speaking, engage with the world via cultural tools when we communicate and perform physical activities.

In passing it should be pointed out that Vygotsky in no way denied the significance of the biological basis of human cognitive and psychological functioning. In fact, in many of his writings, he explores various features of animal cognition, in particular the thinking of primates, and he had a strong interest in such comparative research. But he was adamant in his insistence on the fact that human thinking and learning cannot be reduced to biological processes. Humans are unique in the sense that they think, communicate, and work through cultural tools, and it is at this level that we have to study learning and development if our interest is in human psychological functioning. The sociocultural development is “not a simple continuation” of the biological one. On the contrary, when people begin to appropriate cultural tools through communication with those around them, the “*nature of development itself changes, from biological to sociohistorical*” (Vygotsky, 1986: 94; italics in original). In this sense, “animals are incapable of learning in the human sense of the term,” since “*human learning presupposes a specific social nature and a process by which children grow into the intellectual life of those around them*” (Vygotsky, 1978: 88; italics in original). The evolutionary psychologist Donald (2001) expresses this idea of the relationship between biology and culture even more strongly when he argues that the human brain is a biological structure which has been ‘hi-jacked’ by culture for cultural and interactional purposes.

Learning and Development: the Zone of Proximal Development

Another idea which is central to the sociocultural tradition, and which differs from other perspectives, is the focus on the change and dynamics of psychological processes. In fact, Vygotsky argued that when studying human beings, one always studies change. This idea he developed as a critique of the research of, among others, Piaget and his stage theory and of the representatives of traditional intelligence testing such as Binet. An assumption of such theories is that “learning trails behind development” (Vygotsky, 1978: 80). This implies that “[d]evelopment

or maturation is viewed as a precondition of learning but never as a result of it” (Vygotsky, 1978). From the point of view of educating children this implies that development is seen as a more fundamental process; children can only learn it when they are at the required stage of maturation. Instruction, thus, should be adapted to the developmental level of the child.

For Vygotsky, and from a sociocultural perspective, the opposite assumption, that is, that learning is constitutive of development, is more productive. It is by appropriating cultural tools that children develop and become familiar with the accumulated knowing and skills of their community. When children begin to appropriate the basics of addition and subtraction, they become familiar with specific cultural tools and “this provides the basis for the subsequent development of a variety of highly complex internal processes in children’s thinking” (Vygotsky, 1978: 90). Expressed differently, through learning the development of the child is set in motion in a specific direction. However, learning is not identical with development; rather, it is a necessary prerequisite for the child to develop “culturally organized, specifically human psychological functions” (Vygotsky, 1978).

These ideas of the dynamics of human thinking, and that learning contributes to development, are incorporated into the famous concept of the zone of proximal development (ZPD). Throughout his short life, Vygotsky had a strong interest in education and wanted to offer an alternative way of thinking about pedagogical practices to the ones offered by, for instance, stage theories and behaviorism. Instead of seeing maturation as a necessary prerequisite for learning, one should consider the manner in which children (or adults) appropriate cultural tools. This is done gradually and through the support the learner receives in social interaction by more expert partners. Thus, for the individual there is a ‘zone’ in which his or her familiarity with how to use a cultural tool is still at an early stage. Vygotsky (1978: 86) defined ZPD as:

the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers. (italics in original)

What this idea implies at a very concrete level is that when people are allowed to cooperate with more capable peers, their performance is usually much better than when they work alone. When a child is struggling with understanding the basics of addition and subtraction, support by a teacher or a parent will assist the child in understanding how to proceed when dealing with specific problems. These hints and suggestions may be indirect and subtle but when the adult and the child share a certain interpretation of the task, the child can make productive use of the skills of the more expert person. More capable

partners will thus ‘scaffold’ (Wood *et al.*, 1976) the activities of less experienced members of an activity.

The ZPD can be conceived as a developmental path in which the child’s appropriation of a cultural tool is still partial, and where external support will be necessary to accomplish a task such as tying a knot (Nilholm and Säljö, 1996) or solving a puzzle (Wertsch, 1985). But it can also be seen as the zone in which a child is particularly sensitive to instruction from a more knowledgeable partner. It is within this zone that the child is able to profit from the assistance provided by the adult in order to develop a more independent understanding of the cultural tool and how to use it in a specific setting.

The ZPD has served as an inspiration for important developmental work in education in many contexts. The attempts to develop reciprocal teaching and guided cooperative learning by the late Ann Brown and her colleagues (Brown, 1993) is one example of this. In reciprocal teaching, the role of the teacher is “to scaffold the involvement of learners in the discussion by providing the explanation, modeling, support, and feedback that will – in time – enable full participation of students” in activities such as understanding texts and other cultural tools. The idea of the ZPD has also been used as an underlying notion for the view of learning as a process of increasing participation in social practices emphasized by scholars such as Rogoff (1990). In much of this work there is also a critique of the rather one-sided view of the adult–child relationship that Vygotsky presents, where the child is always the dependent part and the adult is in control of the progress of the activity. Several of these scholars, for instance, Rogoff in her analyses of guided participation, instead emphasize the joint efforts by all parties of maintaining mutual understanding and shared perspectives on the world when engaging in conversations. This critique also draws inspiration from the more dialogical interpretations of learning and social interaction which have emerged during recent years when Vygotskian ideas have been combined with the dialogical perspective on language and interaction that is at the heart of Bakhtin’s approach to human communication (Wertsch, 1991).

Language, Learning, and the Formation of Mind

The most well-known publication by Vygotsky is his book *Thought and Language*. It was originally published in the final year of his life and has appeared in a number of English translations. The first English version, a highly abbreviated one, appeared in 1962 and more extended versions were published in 1986 and 1997 (the latter one with the title *Thinking and Speech*). In this volume, which undoubtedly is one of the most influential books in the fields of learning and development, Vygotsky explores fundamental problems

of the relationships between thinking, communication, and the formation of mind. His ambition is to scrutinize the role that language and communication play in human ontogeny, and he touches on several significant issues that pertain to how the mind is shaped by social experience.

Vygotsky’s emphasis on the role of linguistically mediated communication is rooted in his conviction of the role that semiotic mediation plays in the formation of human thinking and social interaction. It is through semiotic mediation that we develop the abilities characteristic of higher psychological functioning. We learn conceptual knowledge, how to remember, how to monitor our own activities and put them under conscious control, and a range of other sociocultural skills. Much of what we learn would in modern psychological parlance be described as meta-cognitive or meta-communicative skills; we learn to structure our own activities and the world through language (Rommetveit, 1985). The emphasis on these kinds of skills reflects Vygotsky’s conviction that as human beings we not only live in the world; rather, we are also able to reflect on it and develop knowledge about what we do (cf. Cole, 1996: 120). Language is the most significant tool in such processes.

From the educational point of view, one particularly interesting analysis concerns that of the distinction between the acquisition of spontaneous concepts and scientific concepts, respectively. We appropriate spontaneous concepts (brother, sister, and family) in everyday interaction, that is, we learn them from below, metaphorically speaking. Other concepts (one example that Vygotsky uses is Archimedes’ law) we have to learn through some form of explicit instruction or guidance. Somebody has to explain to us what the term and the concept mean, and appropriation then proceeds in a top-down fashion; we descend from the abstract to the concrete (cf. Kozulin, 1998: 48). Thus, understanding them implies learning to contextualize a particular class of phenomena by means of an abstract principle. Supporting this specific type of learning is the responsibility of schooling and the teacher. This is one of the reasons why schooling is so important, and why the specific learning practices of schooling are different from what we find in other settings. Thus, Vygotsky objected to Piaget’s view that children, when engaged in self-directed learning and experimentation, will come to master scientific concepts. Such learning, Vygotsky argued, presupposes guidance by someone who masters the relevant discourses. A significant role for schooling is to introduce children to those parts of the accumulated knowledge of our society that is not available through everyday practices.

Vygotsky and the Development of Education

In the debates on learning and development Vygotsky’s sociocultural approach has become central. The theory

has become part of many introductory books in psychology and education, and has found its way into discussions about pedagogy. An interesting dimension of the legacy of Vygotsky concerns his impact on education. Already in the Soviet period there were many who attempted to develop pedagogical practices that were rooted in Vygotsky's thinking. The success of these experiments is not convincing. In fact, many of these attempts now strike us as highly problematic. The notion of a scientific approach to education often resulted in a kind of scienticism with what appears to have been rather sterile, even authoritarian, teaching patterns dominated by adults. Children were, to some extent, treated as objects of teaching rather than active participants and contributors to their own learning. In recent developments in educational research, however, Vygotskian ideas have become important for approaches to teaching and learning that emphasize the situated nature of knowing and the necessity of actively engaging learners in the process of appropriating knowledge and skills in cooperation with adults and in various kinds of learning communities. The work by Brown and Rogoff and their colleagues mentioned earlier are but two examples of many contemporary attempts to take Vygotskian ideas into educational practices. This is also an important reminder of a point that John Dewey repeatedly made: theoretical principles are not enough for creating successful educational practices. This move from theoretical insights to pedagogical practices *per se* requires extensive work. But for anyone struggling with these important issues, the sociocultural approach is a rich and promising companion.

See also: An Overview of Language and Literacy in Educational Settings; Apprenticeship Approach to Learning; Constructivism and Learning; Cultural-Historical Activity Theory; Peer Interaction and Learning; Vygotsky and Recent Developments.

Bibliography

- Brown, A. L. (1993). First-grade dialogues for knowledge acquisition and use. In Forman, E. A., Minick, N., and Stone, C. A. (eds.) *Contexts for Learning. Sociocultural Dynamics in Children's Development*, pp 32–57. New York: Oxford University Press.
- Bruner, J. S. (1985). Vygotsky: A historical and conceptual perspective. In Wertsch, J. V. (ed.) *Culture, Communication and Cognition: Vygotskian Perspectives*, pp 21–34. Cambridge, MA: Cambridge University Press.
- Cole, M. (1996). *Cultural Psychology: A Once and Future Discipline*. Cambridge, MA: The Belknap Press.
- de Graaf, W. and Meier, R. (eds.) (1994). *Sociogenesis Reexamined*. New York: Springer.
- Donald, M. (2001). *A Mind so Rare. The Evolution of Human Consciousness*. New York: Norton.

- Kozulin, A. (1986). Vygotsky in context. In Vygotsky, L. S. (ed.) *Thought and Language*, pp XI–LVI. Cambridge, MA: The MIT-Press.
- Kozulin, A. (1998). *Psychological Tools. A Sociocultural Approach to Education*. Cambridge, MA: Harvard University Press.
- Leontiev, A. N. (1997). On Vygotsky's creative development. In Rieber, R. and Wollock, J. (eds.) *The Collected Works of L.S. Vygotsky: Volume 3. Problems of the Theory and History of Psychology*, pp 9–32. London: Plenum Press.
- Nilholm, C. and Säljö, R. (1996). Co-action and situation definitions. An empirical study of problem solving in mother-child interaction. *Learning and Instruction* 6(4), 325–344.
- Rogoff, B. (1990). *Apprenticeship in Thinking: Cognitive Development in Social Context*. New York: Oxford University Press.
- Rommetveit, R. (1985). Language acquisition as increasing linguistic structuring of experience and symbolic behaviour control. In Wertsch, J. (ed.) *Culture, Communication, and Cognition: Vygotskian Perspectives*, pp 183–204. Cambridge: Cambridge University Press.
- Säljö, R. (1999). Learning as the use of tools: A sociocultural perspective on the human-technology link. In Littleton, K. and Light, P. (eds.) *Learning and Computers: Analysing Productive Interactions*, pp 144–166. London: Routledge.
- Säljö, R., Eklund, A.-C., and Mäkitalo, Å. (2006). Reasoning with mental tools and physical artifacts in everyday problem solving. In Verschaffel, L., Dochy, F., Boekaerts, M., and Vosniadou, S. (eds.) *Instructional Psychology: Past, Present and Future Trends*, pp 73–90. Oxford: Pergamon.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. S. (1986). *Thought and Language*. Kozulin, A. (trans.) Cambridge, MA: MIT-Press.
- Vygotsky, L. S. (1994). The problem of the cultural development of the child. In van der, Veer, R. and Valsiner, J. (eds.) *The Vygotsky Reader*, pp 57–72. Oxford: Blackwell.
- Vygotsky, L. S. (1997). The instrumental method in psychology. In Rieber, R. and Wollock, J. (eds.) *The Collected Works of L.S. Vygotsky. Volume 3: Problems of the Theory and History of Psychology*, pp 85–89. London: Plenum Press.
- Wertsch, J. V. (1985). *Vygotsky and the Social Formation of Mind*. Cambridge, MA: Harvard University Press.
- Wertsch, J. V. (1991). *Voices of the Mind: A Sociocultural Approach to Mediated Action*. Cambridge, MA: Harvard University Press.
- Wertsch, J. V. (1998). *Mind as Action*. New York: Oxford University Press.
- Wood, D., Bruner, J. S., and Ross, G. (1976). The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry* 17, 89–100.

Further Reading

- Daniels, H. (2008). *Vygotsky and Research*. New York: Routledge.
- Daniels, H., Cole, M., and Wertsch, J. (eds.) (2007). *The Cambridge Companion to Vygotsky*. New York: Cambridge University Press.
- Kozulin, A., Gindis, B., Ageyev, V. S., and Miller, S. M. (eds.) (2003). *Vygotsky's Educational Theory in Cultural Context*. Cambridge: Cambridge University Press.
- Mercer, N. (2000). *Minds and Words: How we Use Language to Think Together*. London: Routledge.
- Valsiner, J. (2000). *Culture and Human Development: An Introduction*. Thousand Oaks, CA: Sage.
- Wells, G. (1999). *Dialogic Inquiry. Towards a Sociocultural Practice and Theory of Education*. Cambridge: Cambridge University Press.

Self-Regulated Learning and Socio-Cognitive Theory

P H Winne, Simon Fraser University, Burnaby, BC, Canada

A F Hadwin, University of Victoria, Victoria, BC, Canada

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Agency – The human capacity to choose and act on choices.

Coregulation – The temporary sharing or distributing of self-regulated learning between a learner and a more capable other (peer or teacher) to help the learner advance toward more effective self-regulated practice.

Enactive learning – Learning by doing or practicing.

Metacognition – The thought about cognition and the products of cognition usually involving monitoring and choosing among tactics and strategies.

Scaffolding – The dynamic interaction between a learner and an other in which the kind and degree of support is carefully matched to the goal of learning and the learner's current level of mastery relative to that goal.

Self-regulated learning – The intentional and strategic adaptation of learning activities to change cognition, motivation, and behavior outcomes.

Sociocognitive theory – A theory that acknowledges the joint roles of cognitive factors, self-beliefs, and environmental factors in human learning; an extension of social-learning theory.

Sociocultural theory – A theory that emphasizes social interactions and participation as central in enculturating learners into the practices of their learning communities.

Strategy – A carefully selected set of tactics coordinated to achieve a learning goal.

Tactic – A basic technique for learning that is initiated under specific conditions and produces a well-defined result.

Vicarious learning – Learning by observing others.

made to learning processes and strategies; to conditions, like motivation or factors in the learning environment, that affect learning activities or to learning outcomes and products.

For example, after writing a paragraph, I typically pause to read it. I check it against at least two standards; whether it is (1) clear and (2) presents content I intended it to. If the paragraph falls short of these standards, I adjust the words or phrasing (products), then reexamine the paragraph to judge whether these adjustments are sufficient to meet the standards I set.

I often write without making a written outline. However, when the topic is particularly complex or relatively new to me, past experience in producing less-than-satisfactory products leads me to adjust my writing process. In these cases, I begin by creating an outline, then writing to fill in its sections. I also add a third standard to be used when pausing and reviewing my writing, namely, examining the match between what I have written and the outline.

These examples of regulating my writing are instances of self-regulation because I, rather than some external person or source of feedback (e.g., a computer beep) initiated and performed the regulating activity. I determined the standards to use in monitoring whether adjustments were necessary, and I decided whether the product and processes met those standards. If instead, my coauthor read my draft and made changes to it, that is regulation by an external source or other-regulation. If my coauthor modeled the development of a concept map as a method for improving my writing, that adjustment to my typical writing process also is other-regulation. If my coauthor asked me questions that prompted me to reflect on and adjust my writing process or product, this would be considered coregulation because rather than doing the regulation or changes for me, she encouraged me to metacognitively reflect upon and regulate my own writing.

In sum, regulation occurs when learners adjust products they create or methods they use to create products. Adjustments are made in relation to standards that characterize ideal or sufficient products and processes. Regulation is metacognitive because the monitoring examines prior cognitive activities that were methods by which the person engaged in a task. When regulation is initiated and managed by one's self, that is self-regulation. When people regulate processes they use to acquire new knowledge – to learn – they engage in SRL.

What Is Self-Regulated Learning?

Self-regulated learning (SRL) refers to intentionally and strategically adapting learning activities to achieve goals of learning. Learners self-regulate by applying learning tactics they predict will be successful. They monitor how well tactics achieve goals and, when differences exceed a threshold, they make adjustments. Adjustments can be

Models of SRL

The number and scope of models of SRL make it impossible to address all of them here. We select models that are prevalent in the research literature and provide context for considering sociocognitive aspects of SRL.

Boekaert's Model of Adaptable Learning

Boekaerts' model of adaptable learning focuses on students in classrooms. A basic premise is that learners strive to balance two priorities: (1) extending knowledge and skills to enhance their personal resources – a mastery mode of learning – and (2) preserving what one knows and believes to avoid injuring the self – a coping or well-being mode of learning. Corresponding to these two modes, the learner draws on constellations of: (1) learning strategies that build resources and (2) coping strategies that protect resources. Thus, Boekaerts' model, like others, blends cognitive and motivational elements. Success in this depends on the learners' capacity to appraise the overall situation and, on that basis, control their approach to learning as well as elements in their task environment.

When learners participate in learning, they draw on three main sources of information. The first is their perception of the task environment. This is an amalgam of the task they are assigned, the teacher's (or textbook's) instructions, and features of the physical and social context in which the task is situated. The second source of information is a spectrum of knowledge and skills specific to the domain of the task that comprise an action plan. The plan includes: prior declarative and procedural knowledge, tactics and strategies they have successfully used before, plus metacognitive knowledge. The third source of information concerns the student's hierarchy of goals, motivations, and values.

Boekaerts' model emphasizes learners' goal setting and ways they apply control to deal with successes and failures. Productive self-regulating learners plan for how they can receive feedback about goals. When goals are not met, learners seek to balance goals relating to mastery and those of coping in ways that adapt to maintain a satisfactory balance and sense of self.

Winne and Hadwin's Model of Recursive Self-Regulation

Winne and Hadwin developed their model to describe occasions when learners study, as in doing homework or preparing for an oral presentation. They postulate that SRL unfolds over four flexibly ordered and recursive phases.

In phase one, learners identify what they perceive to be conditions that define the assigned task. Conditions fall

into two main categories. Task conditions are features of the assigned task such as objectives the teacher (or textbook) set, time available, whether peers are involved, social structure (e.g., cooperative vs. competitive, or whether responsibility is individual or shared), resources available to start and then support work on the task, and guidance or scaffolding. Cognitive conditions are internal to the learner. They include the scope and relevance of prior knowledge, motivational orientation, epistemological beliefs, known study tactics, and other qualities that make the learner a unique individual.

In phase two of this model, learners construct a perception of what the task is and, on that basis, set their own goals. In phase three, learners begin engaging with the task, taking steps to reach goals. In phase four, large-scale changes to preceding phases may be made, including changing metacognitive knowledge, to enhance success in the present and for future tasks.

Within each phase, Winne and Hadwin hypothesize that learners engage in metacognitive monitoring. For example, in phase one, learners may reexamine external resources or what they know about the task to revise their description of the assignment. When metacognitive monitoring in later phases reveals gaps of the kinds the learner perceives to be too great, learners retreat to a prior phase to make adjustments – this is the recursive property of their model.

Zimmerman's Social-Cognitive Model

Zimmerman's model extends Bandura's social cognitive theory. It also is a phased model. In the forethought phase, learners undertake two main activities. One couples an analysis of the task the learner perceives to be assigned to planning that identifies what the learner forecasts are optimal techniques for succeeding at the task. The other facet of the forethought phase is where learners survey their self-motivational beliefs, specifically: self-efficacy, expectations they have for the outcome(s) of the task, their intrinsic interest in the task and its value to them, and goals they seek to achieve.

In the second phase, learners engage in the task and simultaneously apply control processes to stay on track toward goals. One key component is applying self-control, for which there are four main elements: (1) self-instruction, where learners overtly or covertly describe what they are to do; (2) imaging, where learners mentally picture their activities; (3) attention focusing, where learners change their external and cognitive environment to screen out distractions; and (4) applying strategies that decompose a complex task into parts that can be managed. The second key component, self-observation, sets the stage for applying control. Learners record mentally or materially (e.g., in a personal log) what they do and how well it works, and

engage in cycles of self-experimenting to explore whether variations in their approach will lead to greater success.

In the third and final phase, learners self-reflect. This has two components: self-judgment, where learners determine the extent to which goals are achieved; and self-evaluation, where learners compare their achievements to standards. Standards have four forms: mastery, one's prior performance, normative expectations, and in collaborative task environments, whether one's expectations were successfully fulfilled.

Commonalities across Models of SRL

Task Environment

Whenever SRL occurs, it occurs in the context of a task the learner is pursuing. Tasks may be assigned by teachers, textbooks, or peers in a collaborative work group; or, tasks may be self-generated, as when a learner pursues knowledge and skills related to a hobby. Tasks are multifaceted and there is great variety among typologies for characterizing tasks. Common facets include time available, resources at hand, whether standards are explicit or implicit, and characteristics of feedback the learner can access (e.g., timing, topic, and guidance).

Agency

To have capacity to self-regulate learning, learners must be able to make choices and, within relatively wide limits, to act on their choices. Without such latitude of behavior, learning is other-regulated rather than self-regulated. Agency is the common term used to describe this capacity to choose and act on choices. Sociocognitive theory emphasizes the role of individual agency on human learning. Bandura proposed that human agency involves four core features including: intentionality, forethought, self-reactiveness, and self-reflectiveness. These features of agency are cornerstones in most models of SRL and account for the perception that self-regulation is a sociocognitive process. As such, every model of SRL rests on the assumption that learners can exercise agency although it is recognized that, for practical and other reasons, agency is not boundless. For example, expressions of agency are strongly prescribed by social conventions.

Goals

Making adjustments to products of learning or processes used in learning might sometimes be trial and error. Notwithstanding a universal assumption tacit in some models of SRL is that learners are purposeful and oriented to particular goals. Goals define standards for judging the adequacy or sufficiency of products and processes. Formally, a standard is a description of an observable

feature of a product or process that can be measured at least nominally, that is, present or absent. Goals also can have features that are measured at finer distinctions, such as ranks and intervals.

Goals have two characteristics. One is purely informational. In this sense, standards that constitute goals are just descriptions of products or processes. The other characteristic of goals is what they mean to people. In this sense, goals matter because they have *qualia*—feelings, emotions, motivation, and values. Some models of SRL give different prominence to the informational features and the *qualia* of goals.

Monitoring

Goals play roles in SRL across time. First, goals are information learners consider as they shape their approach to tasks, including whether the task is accepted as presented, accepted but modified, or rejected. Second, learners may consider goals in the midst of engaging in a task. Some tasks have explicit subgoals. Third, learners can examine products and processes used in learning at point of completion. In all three instances, standards inherent in the goals are compared to attributes of the products and processes at each point along the timeline. Monitoring is the cognitive process responsible for making this comparison and generating a description of the match between goals and events. In this sense, all models of SRL are metacognitive models because this monitoring is inherently metacognitive.

Memory and Reasoning

New situations are never exactly the same as prior experiences. Thus, learners interpret each new situation in relation to what they remember about previous situations they deem similar. In this way, all SRL is integrally dependent on what a learner can remember about previous situations, as well as what that knowledge permits them to perceive about new situations. In matching history to the present, learners interpolate or interpret.

When learners adjust a product or process, models of SRL assume learners do this with an intention that making that particular adjustment, as opposed to some other adjustment, will align the product or process with standards defined in goals. Having this intention entails that learners reason about the likelihood the adjustment will have intended effects.

Sociocognitive Accounts of SRL

When is Cognition Social?

From a sociocognitive perspective, learners are agents who can choose how they will behave. Bandura's triadic theory describes learners' choices as shaped by reciprocal

interactions among: (1) past behaviors; (2) personal variables, such as interest in a task and beliefs about whether they have efficacy to complete a task; and (3) environmental variables, such as instructional supports they can draw on, material resources, and task contexts. The reciprocity involving environment, particularly peers and models, in learning represents cognition as fundamentally social. As learners work on tasks, they generate new information that updates any or all three factors. In addition to reciprocal influences between personal beliefs and behavior, social and environmental factors mediate learning. Environmental factors include things such as social modeling, feedback, and instructional conditions. Each can influence on personal variables and behavior before, during, or after a task. Each can, in turn, be shaped or regulated by behaviors and personal variables. For example, my coauthor, who initially thought he would easily complete his writing task, might revise that belief after several unsuccessful attempts to generate an outline. Personal beliefs about the writing task might be influenced by the fact that the deadline is approaching. Collectively, these beliefs about his person and the environment might lead him to adopt a different strategy for writing, such as abandoning the preliminary outline, writing an introductory section, and using that introduction as a vehicle for collaboratively generating an outline with his coauthor.

Sociocognitive theory recognizes that learners can learn in two ways. They can learn by doing, called enactive learning; and by observing, called vicarious learning. Both kinds of learning are social because the activities of learning are embedded in a task environment that is reciprocally influenced by: (1) others directly (e.g., modeling and feedback), (2) memories of others' behaviors (e.g., attention to what is observed), and (3) tools others created (e.g., instructions and examples).

Sociocognitive Learning and SRL

Vicarious learning is central to sociocognitive accounts of SRL because regulation is described as a developmental trajectory moving from observation, through emulation of others usually involving guided practice, to self-control, and finally to self-regulation. Observation is fundamentally social because it involves attending to actions of a proficient model, participating in guided practice, and learning in relation to instrumental feedback. Through these social processes, students develop competence with the task, content, and context; as well as their capabilities to self-regulate learning.

From this perspective, SRL is a developmental process assisted by observing models, trying out the behavior, and receiving feedback from others. There is a tacit assumption in most sociocognitive descriptions of SRL that social supports are prominent while SRL proficiency develops

and become less prominent afterward. In fact, however, environmental factors such as instructional sequence and classroom context continually exert a reciprocal influence on self-regulatory behaviors and beliefs.

The three models of SRL introduced earlier share two features with sociocognitive perspectives of learning. First, they emphasize the role of individual agency. Second, they acknowledge the role of task environment in shaping self-regulatory activity.

Alternatives to Sociocognitive Theories of SRL

More recent perspectives of learning reveal increased interest in explaining the role of social and contextual influences on SRL.

Sociocultural influences on SRL

Sociocognitive perspectives of SRL emphasize self-regulation as developing within the individual and assisted by external modeling and feedback. In contrast, sociocultural perspectives of SRL emphasize it as a fundamentally social process wherein students learn to internalize language, signs, and activities existing first in the sociocultural practices of their communities. This shift in perspective changes the emphasis from self- to coregulation.

From a sociocultural perspective, SRL is a stage occurring as children are socialized into speech patterns and practices. Coregulation is the temporary sharing or distributing of self-regulatory processes and thinking between a learner and more capable other (peer or teacher), while the learner transitions toward self-regulatory practice.

Three basic concepts characterize coregulatory aspects of self-regulation. First, rather than focusing on the individual learner, the focus is on the relationships among individuals, objects, and settings. Second, regulating learning involves coordinating and negotiating social contexts as well as self and social expectations and goals. Third, instructional supports afford opportunities for learners to experiment with and learn motivation as well as strategies and self-evaluations central to SRL.

Informed by Vygotsky's notion of internalization, regulation is seen as a social process because it appears first on the intrapsychological plane and then later becomes part of a child's understanding, appearing on the interpsychological plane. Borrowing from partners in learning and joint problem solving are considered core coregulatory processes in the social exchange between learners and more capable others. From this perspective, the mark of SRL is when the activity and practice appears in a learner's own performance, and when those activities are internalized and automated.

Coregulation

From the sociocultural perspective, coregulation is a central transitional process in a learner's development of more

productive SRL. Coregulation is an interactive process whereby ownership of self-regulatory activity and thinking is first shared among participants, then gradually taken up or appropriated by the individual learner. During coregulation, student and teacher regulate together, sharing thinking and decision making and developing a shared or intersubjective task environment where each brings expertise and control to the task. Slowly, as knowledge and control are transferred to the learner, self-regulation emerges and students begin to develop realistic self-evaluations. Hence, this process is referred to as emergent interaction. Finally, the student independently engages behaviors, actions, and thinking associated with SRL.

In contrast to a sociocognitive perspective of SRL that emphasizes self-regulation developing within the individual in reciprocal interaction with external modeling and feedback, coregulation emphasizes a shifting in who shapes the regulation. Through scaffolding, cognitive demands arising from engaging in a task are eased by sharing the demands of metacognitively monitoring, evaluating, and regulating task processes. For example, rather than a mother modeling her own steps and thinking while tying a shoe lace, she takes on the regulation of her child's tying shoe laces by metacognitively monitoring and evaluating with the child. The mother might ask questions like: "What do you know about how to connect those two laces?" "How do you know when you have completed the first step properly?" "What do you need to do now?" In this way, the child focuses on task enactment while the mother supports metacognitive engagement and regulatory control of learning.

Research on SRL and Sociocognitive Models

Several core social components appear in theoretical and empirical accounts of SRL. We provide a snapshot of findings.

Feedback and SRL

Feedback provided by others is a social tool that exerts powerful effects. An important distinction about feedback is whether it is targeted at tactics and strategies, called process feedback, or at products, called knowledge of results. The distinction is somewhat artificial in the case of SRL because, when a learner is attempting to regulate processes, these are products of metacognition that are being affected by feedback describing how processes were carried out.

The small amount of research on how feedback affects SRL and products generated in tasks is quite consistent. When learners face challenging tasks for which their current tactics and strategies are not sufficient for success, performance in the long run can be improved by

providing feedback about processes in the short run. What occurs in this case is that, while the learner is shaping up tactics and strategies, there is little success in meeting the goals of the task *per se*. Delaying gratification, in the sense of postponing progress toward task goals, often frustrate learners and their teachers but overcoming this pays off in the end.

Modeling and SRL

Modeling is considered an important antecedent of self-regulation. When students observe models, they acquire knowledge and strategies for successfully completing a task. Under productive conditions of modeling, they also develop beliefs regarding their efficacy for that task. Models can display mastery of the task or methods for coping with the challenges of a task not yet mastered. Research shows both kinds of models are beneficial under different circumstances. When children have experienced failure or challenges, coping models enhance self-efficacy and skills better than mastery models. Observing multiple peer models is more effective. Modeling has been shown to improve self-regulation of academic skills in domains such as mathematics and writing, motor skills, strategy knowledge, and metacognitive strategies.

Models are effective when observers perceive them as competent, regardless of model age. Children are more influenced by models they perceive as similar in ability to themselves.

Scaffolding and SRL

Scaffolding is a dynamic interaction between the learner and another. The other provides support that is carefully calibrated to the goal of the task and the learner's current level of mastery relative to that goal. Through scaffolding, learners can achieve something that would not be possible by themselves. With effective scaffolding, the learner gradually assumes full regulatory control of cognition, metacognition, motivation, and behavior.

Research has yet to systematically examine the role of diagnosis, calibration, and fading on specific subprocesses of cognition, motivation, metacognition, and behavior associated with SRL. The picture that is beginning to emerge suggests that scaffolding takes time. It involves forms of coregulatory dialog. The teacher requests information, confirms the learner's results or interpretation, restates results, and models thinking or requests judgments about learning or performance. In coregulatory dialog, students request information, confirm interpretations of the teacher's activities, elaborate on features of the task and goals, and request evaluations of performance and learning. Furthermore, when scaffolding is adapted, timely and targets specific aspects of SRL such as planning,

monitoring, and using strategies, students experience shifts in their mental models, use more strategies, and regulate their learning by planning, activating prior knowledge, and monitoring their progress toward goals.

See also: Metacognition.

Further Reading

- Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology* **52**, 1–26.
- Boekaerts, M. (2006). Self-regulation and effort investment. In Sigel, E. and Renninger, K. A. (vol. eds.) *Handbook of Child Psychology: Volume 4. Child Psychology in Practice*, pp 345–377. Hoboken, NJ: Wiley.
- Boekaerts, M., Pintrich, P. R., and Zeidner, M. (2000). *Handbook of Self-Regulation*. Academic PressSan Diego, CA.
- Butler, D. L. and Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research* **65**, 245–281.
- Corno, L. and Mandinach, E. B. (2004). What we have learned about student engagement in the past twenty years. In McInerney, D. M. and Van Etten, S. (eds.) *Big Theories Revisited: Volume 4. Research on Sociocultural Influences on Motivation and Learning*, pp 299–328. Greenwich, CT: Information Age.
- Gallimore, R. and Tharpe, R. (1990). Teaching mind in society: Teaching, schooling, and literate discourse. In Moll, L. C. (ed.) *Vygotsky and Education: Instructional Implications and Applications of Sociocultural Psychology*, pp 175–205. New York: Cambridge University Press.
- McCaslin, M. and Hickey, D. T. (2001). Self-regulated learning and achievement: A Vygotskian view. In Zimmerman, B. J. and Schunk, D. H. (eds.) *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*, 2nd edn., pp 227–252. Mahwah, NJ: Erlbaum.
- McGivern, J. E., Levin, J. R., Ghatala, E. S., and Pressley, M. (1986). Can selection of an effective memory strategy be induced vicariously. *Contemporary Educational Psychology* **11**, 170–186.
- Puustinen, M. and Pulkkinen, L. (2001). Models of self-regulated learning. *Scandinavian Journal of Educational Research* **45**, 269–286.
- Rosenthal, T. L. and Zimmernan, B. J. (1978). *Social Learning and Cognition*. Academic PressNew York.
- Schunk, D. H. (1987). Peer models and children's behavioural change. *Review of Educational Research* **57**, 149–174.
- Schunk, D. H. (2001). Social cognitive theory and self-regulated learning. In Zimmerman, B. J. and Schunk, D. H. (eds.) *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*, 2nd edn., pp 125–151. Mahwah, NJ: Erlbaum.
- Schunk, D. H. and Zimmerman, B. J. (1997). Social origins of self-regulatory competence. *Educational Psychologist* **32**, 195–208.
- Winne, P. H. and Hadwin, A. F. (1998). Studying as self-regulated learning. In Hacker, D. J., Dunlosky, J., and Graesser, A. C. (eds.) *Metacognition in Educational Theory and Practice*, pp 277–304. Mahwah, NJ: Erlbaum.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In Boekaerts, M., Pintrich, P., and Zeidner, M. (eds.) *Handbook of Self-Regulation*, pp 13–39. San Diego, CA: Academic Press.
- Zimmerman, B. J. (2004). Socio-cultural influence and students' development of academic self-regulation: A socio-cognitive perspective. In McInerney, D. M. and Van Etten, S. (eds.) *Big Theories Revisited: Volume 4. Research on Sociocultural Influences on Motivation and Learning*, pp 139–164. Greenwich, CT: Information Age.
- Zimmerman, B. J. and Schunk, D. H. (2001). *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*, 2nd edn. Mahwah, NJ: Erlbaum.

N

NATIONAL SYSTEMS OF EDUCATION

Contents

Australia

Austria

Bangladesh

Barbados

Bulgaria

England

Estonia

Fiji

Finland

France

Germany

Ghana

Greece

Iceland

Indonesia

Israel

Jamaica

Latvia

Lebanon

Lithuania

Malawi

Malta

Mongolia

Namibia

New Zealand

Niger

Norway

Oman

Pakistan

Poland

Qatar

Russia

Rwanda

Scotland

Singapore

Slovenia

Solomon Islands

South Africa

Spain

Sri Lanka

Suriname

Taiwan

Tanzania
Thailand
The Gambia
The Philippines
Uruguay
USA
Zanzibar

Australia

C N Power, University of Queensland, Brisbane, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

General Background

Australia is a large island continent. With an area of 7.7 million square kilometers and a population of 21 million, it is one of the most sparsely populated countries on earth. The continent's first inhabitants arrived about 46 000 years ago. Most Australians are of European descent, but large-scale migration since World War II has produced a culturally diverse society.

Australia is a constitutional democracy based on a federal division of powers. Its basic political, legal, and administrative structure derives mainly from Britain. British colonization of the country began in 1788 and by the 1850s, six colonies had been established across the country. In 1901, the Commonwealth of Australia was established as a federal system, the former colonies becoming states. Retaining its constitutional ties with the United Kingdom, Australia is a member of the Commonwealth. However, it now enjoys substantial independence.

The first schools in Australia were private, but by the 1870s, each colony had established a public education system. Under the 1901 federal constitution, the six states retained responsibility for education. Australia also has two territories with similar education powers to the states. While there is no single national system as such, education and training operate within frameworks set by the Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) and the Ministerial Council on Vocational Education and Training (MCVET).

The population of Australia grew rapidly in the 1950s and 1960s when substantial numbers arrived from southern and northern Europe. Since then, population growth rate has slowed. Two-thirds of Australians live in the major capital cities. Almost 25% of Australians were born outside the country, and one in three is either a first- or a second-generation migrant. In the past decade or so, Asia and the Middle East have been prominent sources of immigration and overseas students.

Despite the diversity of languages and cultures, English continues to dominate all activities in daily life, including education. Given the number of immigrants coming from non-English-speaking countries, extensive provision has been made for the teaching of English as a second language, and most children study a language other than English as part of their general education.

For many years, the Australian economy has been dominated by agriculture and mining. These industries remain important, but now employ relatively few people. In 2006, there were 10.7 million people aged 15 years or over in the workforce, 70% working full time. Unemployment is low (under 5%) and concentrated among those with the least education. Over the years, there has been a significant upward movement in the occupational and educational structure of the labor force. More than half of the population aged 15–64 has at least one nonschool qualification, over 36% have completed year 12, and about 20% have a university degree.

With average yearly earnings at A\$28 400, Australia is a relatively wealthy country. Strong economic growth since the beginning of the twenty-first century has led to a skills shortage and the continuing growth and diversification of post-compulsory education. There has also been a strong push for reforms aimed at equipping workers with the competencies seen to be essential to improving productivity and competitiveness in the global knowledge society.

Goals of the Education System

For most of the twentieth century, there has been a broad agreement about the basic goals of the education system and the responsibilities of government. These are generally set out in legislation and are supplemented by more detailed policy documents. The goals of:

- the school sector are reflected in the national goals for schooling, and focus on developing the capacities and

talents of all young people so that they have the necessary knowledge, understanding, skills, and values for a productive and rewarding life;

- the vocational education and training (VET) sector focus on giving industry a highly skilled workforce to support strong performance in the global economy; and
- the higher education sector include a focus on advancing and applying knowledge and understanding to benefit the Australian economy and society.

National goals for schooling have been set by MCEETYA. In addition, it was agreed in 1999 that the curriculum in the compulsory years should encompass the arts, English, health and physical education, languages other than English, mathematics, science, technology, studies of society and environment, and most recently, history. Schools are required to develop computer, vocational, and enterprise skills, and an understanding of Aboriginal and Torres Strait Islander cultures and the value of cultural diversity.

High-quality outcomes for postschool education are critical for both individual welfare and national development. The focus in VET is on supporting the growing and changing Australian economy. Goals are set jointly with Australian industry and employers, and aim at ensuring the availability of a highly skilled and mobile workforce. Higher education institutions have more flexibility in defining their mission and managing themselves. But at the same time, government is pushing for a stronger focus on contributing to economic growth and social development, and on increasing enrolments in science, mathematics, and engineering.

Structure and Operation of the Education System

Figure 1 outlines the formal education system. Key statistics on enrolments and resources in the major education sectors are provided in Table 1.

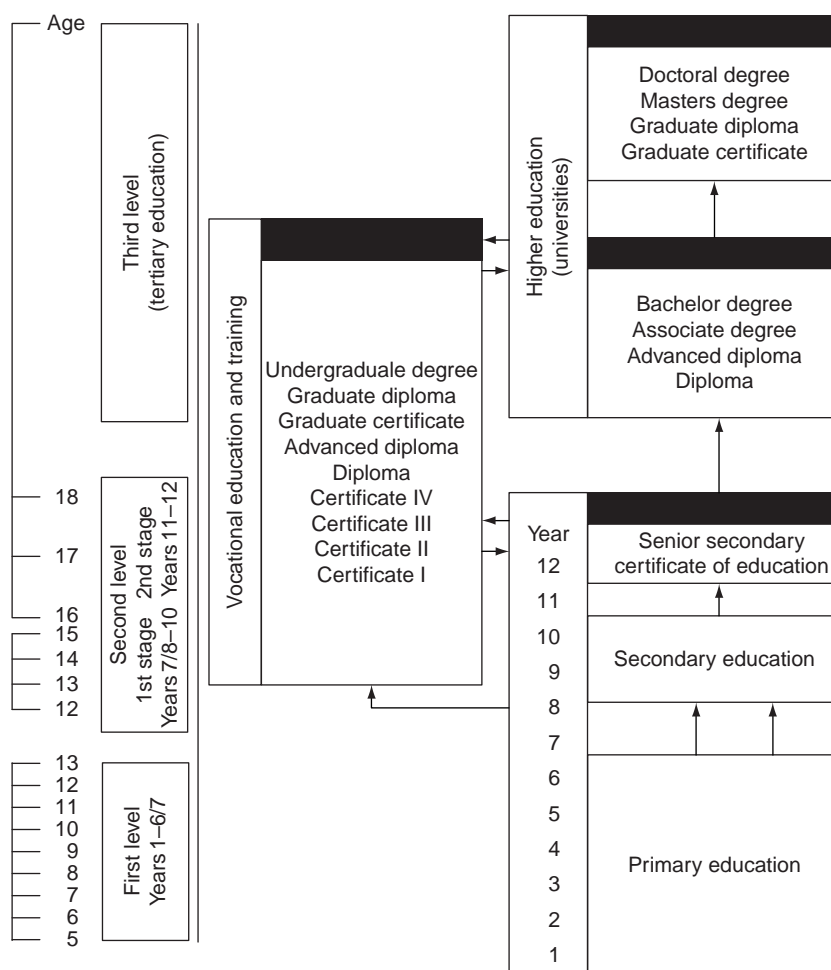


Figure 1 Outline of the Australian education and training system. Providers deliver qualifications in more than one sector. Schools, for example, are delivering certificates 1–2, universities are delivering certificates 2–4, and VET providers are delivering undergraduate degrees, graduate certificates, and graduate diplomas (higher education qualifications in some jurisdictions, but in others also VET), all subject to meeting relevant quality-assurance requirements. Adapted from National Offices of Overseas Skills Recognition (2000).

Table 1 Educational providers, students, and teaching staff, 2005–06

<i>Provider</i>	<i>Institutions</i>	<i>Students '000</i>	<i>Teaching staff '000</i>
Schools	9612	3375	239
Government	6902	2257	158
Catholic	1703	680	45
Independent	1007	441	36
Primary	6558 ^a	1937	121
Secondary	1478	1441	118
Males		1720	81
Females		1659	190
VET	1950 ^b	1638 ^c	27
Males		848	14
Females		790	13
Higher education	74 ^d	957	91
Males	.	432	44
Females	.	513	47

^aThese figures are for 2006 and do not include combined primary–secondary schools (1181 schools) and special schools (395 schools).

^bNote VET institutions include both government and private providers in 2005.

^cOf the 1.6 million people participating in VET in 2005, 1.2 million were in government-funded programs.

^dThe higher education sector includes 37 public universities, private, and other self-accrediting institutions, and higher education providers accredited by state and territory authorities. Adapted from Australian Bureau of Statistics (2007). *Australian Yearbook*. Canberra: Australian Bureau of Statistics Australia Department of Education, Science and Training (DEST) *Annual Reports 2005, 2006*. Canberra: Australian Government Publishing Service.

Preschool and Early Childhood Care

Many Australian children attend an early childhood center prior to formal education, this sector being increasingly privatized. In such centers, the focus is generally on structured play and social development. In 2006, over 217 000 4-year-olds (62% of the age group) were attending a public or private full-time preschool, kindergarten, or preparatory-year program. Subsidies and tax rebates are provided to support families of children in private childcare and preschool programs.

Primary (Elementary) and Secondary Education

Schooling in Australia is compulsory for children from the ages of 6–15 (16 in South Australia and Tasmania). In 2006, there were 9612 schools, of which 72% were government schools and 28% were nongovernment schools. Over 70% of schools were primary, 16% secondary, and the rest were combined primary/secondary schools. There were 3.4 million full-time school students and over 265 000 teachers. The number of students in nongovernment schools is now over 1.1 million, making up one-third of all students, up

from 26% in 1986. The number of indigenous students rose from 42 000 in 1986 to over 140 000 in 2006 (the total indigenous population being about 500 000).

Primary education lasts either for 6 or 7 years, depending on the state or territory concerned. Children who are 5 years old at the beginning of the year normally enter at that stage. The scattered rural population has necessitated a relatively large, but declining, number of small primary schools. The primary school day normally contains about 5 h of tuition, and comprises around 200 school days per year. Overall, the average number of students per full-time teacher equivalent (FTE) in primary schooling was 16.0.

Secondary education is normally for either 5 or 6 years depending on the duration of primary education in the state. The secondary year operates for around 200 days, with about 5.5 h of teaching per day. Twenty years ago, less than 50% of all students stayed in school to year 12, but now three-quarters of all students remain in school to year 12. Completion rates are significantly higher in independent schools and for girls than boys. The average number of students per FTE in secondary schools in 2006 was 12.2.

Each state or territory has established a statutory authority responsible for curriculum development and assessment. Some systems (e.g., Queensland) use moderated school-based assessments, while others combine school-based assessment with external examinations at year-12 level. Given increased retention rates, the curriculum at upper secondary level has been increasingly vocationalized. There are now over 210 000 secondary students enrolled in VET in school programs, most of which are industry-specific courses accredited within the Australian Qualifications Framework.

Other Schooling Arrangements

During the compulsory years, the goal in Australia is for inclusive schooling. But not all students of school age can attend a regular school. Thus, extensive use is made of a variety of delivery mechanisms (e.g., School of the Air, distance and open learning), and the children of some indigenous groups in remote areas receive schooling in homeland learning centers. An estimated 2% of children in Australia have disabilities requiring special education. Wherever possible, children with disabilities are integrated in regular schools, but some do attend special schools (397 in 2005) provided by government or nongovernment authorities.

Technical and Further Education

In recent years, considerable efforts have been made to effectively engage and retain young people in learning and to provide them with opportunities to develop the competencies expected of them in the workplace and to

improve their employability. All governments are seeking to develop smoother, more diverse, and flexible vocational and educational pathways.

In Australia, around 2000 registered training organizations operating in almost 10 000 locations deliver VET programs to 1.6 million students. The participation rate in VET programs exceeds 8% of 15–64-year-olds, and 34% of adults hold a VET qualification. Currently, over 25% of VET units completed are in the management and commerce, 18% in engineering technologies, 10% in health, and the rest in fields like society and culture, food and hospitality, architecture and building, and agriculture. In 2006, there were 389 000 apprentices and trainees, males dominating construction, automotive, electrical trades, the only trade with a majority of females being hairdressing.

Technical and Further Education (TAFE) colleges administered by state and territory governments provide a wide variety of technical and vocationally relevant courses, apprenticeship partnerships, and adult and continuing-education programs. MCVET assures harmonization of the national system of standards, performance assessment, and accreditation with goals agreed in a Commonwealth-State Agreement for Skilling Australia's Workforce. One of the guiding principles for the training scheme is that industry and business need to drive training priorities and delivery, and industry-skills councils have been created for this purpose.

Higher Education

The Australian higher education system is made up of 39 universities (37 public and two private) and other higher education providers recognized under Australian, state, or territory law. All universities are autonomous, self-accrediting institutions responsible for their own governance, under guidelines set by the Department of Education, Science and Training (DEST). There is now also an Australian branch of an overseas university and there are four other self-accrediting higher education institutions. The higher education sector includes over 150 providers that are accredited by state and territory authorities. The non-self-accrediting providers form a diverse group that ranges from theological colleges to those offering courses in areas such as business, information technology, and health.

In 2006, total higher education enrolments were over 843 000, the equivalent full-time student load being almost 430 000. About one-quarter of enrolments are in postgraduate programs. Over 70% of students are studying full time, and most (83%) are internal students, although external students continue to be an important feature of the higher education in Australia. About 40% of year-12 graduates enter higher education within a year or two of successfully meeting entry requirements. But there

are now many pathways for entry to higher education, and in most institutions, mature-age students outnumber those entering directly from school.

Despite rapid expansion in the systems in the past 20 years, students are still drawn disproportionately from urban areas and professional and managerial backgrounds. About 18% of students come from rural or isolated communities, one in seven from a socioeconomically disadvantaged background, and in 2005, there were 8370 indigenous students.

The major shift in the composition of the student body has been created by the enormous growth in the numbers of overseas students that began in 1986 when public institutions were first permitted to enrol overseas fee-paying students. There were some 345 000 overseas students enrolled in 2005, the proportion rising from 8% in 1996 to around 25%. Many Australian higher education providers have established joint-venture agreements with their counterparts in Asia, and about one in four overseas students enrolled in Australian programs is studying offshore.

Adult and Nonformal Education

Adult participation in formal education has increased significantly with the recognition of the need in the knowledge society of the twenty-first century for all to continue to learn through life. Enrolments in the non-formal (i.e., nonaward) sector are difficult to estimate, but almost 600 000 persons were enrolled for study not leading to a qualification in 2005. Programs provided in the nonprofit adult and community-education sector include vocationally oriented courses, secondary school completion, basic skills, personal development, and leisure activities.

The Teaching Profession

In 2005, there were 156 000 FTE teaching staff in government schools (67% of all teachers) and 79 200 employed in nongovernment schools. As a whole, school teaching is a highly feminized profession. Many people with teaching qualifications leave the profession or are not employed as teachers. In VET, there were 14 800 full-time and 7600 part-time teachers. Most (63%) full-time VET teachers are male, and most (65%) part-time teachers are female. There were about 39 000 academics and 51 500 nonacademic staff working in higher education institutions, 53% higher education staff being female. At all levels, teaching is an aging profession.

The training of teachers and specialist-educational personnel is undertaken by universities, and is normally 3 years for primary and 4 years for secondary school teachers. Australia is experiencing a shortage of teachers in rural and remote areas and in mathematics, science,

technology, and languages other than English. To cope, educational institutions at all levels are making increased use of part-time, contract teachers. Education authorities, teachers' unions, and professional organizations (such as the Australian College of Educators) are exploring ways to recruit more teachers into the profession and to improve retention. At the national level, the National Institute for Quality Teaching and School Leadership (now Teaching Australia) and the Carrick Institute for Learning and Teaching in Higher Education have been established to support efforts to make teaching more extrinsically attractive and to establish professional standards.

Administrative and Supervisory Structure

Government schools in Australia operate under the direct responsibility of the relevant state or territory minister, while nongovernment schools are established and operate under conditions determined by government registration authorities. State education departments recruit and employ teachers for government schools; supply buildings, equipment, and materials; and provide some discretionary funding for use by schools. Responsibility for the day-to-day administration of schools has been increasingly devolved to regional offices and the schools themselves.

Private education and training have always been an important part of the Australian education system. Most nongovernment schools have a religious affiliation. Each state has a substantial system of Catholic schools, and over 20% of all school students are enrolled in a Catholic school. Around 900 schools are independent, that is, they do not belong to a system. These schools vary from established, prestigious schools to relatively new, religious, and alternative-education schools. Aside from having to meet prescribed standards for registration and funding purposes, private schools are largely free from government direction, and governed by school boards.

In the tertiary sector, decision making, regulation, funding, and governance are shared among the Australian government, state, and territory governments and the institutions themselves. TAFE colleges are operated by state governments and governed by boards. Higher education institutions are autonomous bodies established under state legislation, each with their own independent governing body. Since 1974, the primary responsibility for higher education has rested with the Australian government, with DEST administering government funding, policy, and programs. While higher education institutions are autonomous, they are accountable to the Australian government through an institutional assessment framework, and their self-assessments are audited by the Australian Universities Quality Agency.

Educational Finance

Total expenditure on education has two components: public and private. For the financial year 2004–05, total government education and training operating expenses were A\$47 billion (5.3% of gross domestic product (GDP)). Private household expenditure on education was A\$18.2 billion (2% of GDP).

Government expenditure on primary and secondary education was A\$26 billion or almost 56% of total government expenditure on education. Government expenditure on preschool, special, and other education was A\$2 billion; VET A\$4.6 billion; and higher education A\$13.9 billion. Nongovernment schools receive the highest proportion (about 70%) of Australian government direct funding, while the states and territories provide over 90% of the funds for government schools.

Income sources for nongovernment schools include Australian government grants (53%), state/territory grants (19%), and private income – predominantly fees (28%). VET providers in receipt of public funds receive 48% of their recurrent funding from state and territory governments, 33% from the Australian government, and the remainder comes mainly from fee-for-service activities (11%) and student fees (4%).

Government funding for higher education comes mainly from the Australian government. It contributes toward the cost of an agreed number of government-supported places, and provides grants for research and other purposes. In 2005, the operating revenue of government-supported institutions amounted to A\$13.9 billion, of which 42% came from government grants. Competition for research grants administered by the Australian Research Council and the National Health and Medical Research Council is highly competitive. DEST also administers government formula-driven research and development.

In addition to government funding, institutions receive research grants from the private sector and revenue from student fees. Most Australian students contribute to the cost of their education through the Higher Education Contribution Scheme (HECS). In 2005, 15% of the operating revenue was raised from HECS, and other fees and charges accounted for a further 23% of operating revenue. In 2007, the fees for full-time HECS students range from about A\$4000 to A\$8400 (for fields such as medicine, law, and dentistry). Charges for full fee-paying students depend on the faculty and institution, and range between A\$14 500 per year for arts, and A\$38 200 for medicine. The revenue from fees paid by overseas students totaled A\$2.1 billion. In the 2007 budget, the Australian government announced funding increases for higher education of about 8% and an increased number of scholarships for overseas students.

Performance Monitoring, Evaluation, and Research

The performance of the education system is monitored by the national, state, or territory authority responsible, using data collected by agencies like the Australian Bureau of Statistics. Most educational research is undertaken by Australian Council for Educational Research, the National Center for Vocational Education Research, and universities. Partnerships are being forged linking university researchers, policymakers, and industry via government initiatives such as the cooperative research centers and collaborative university ventures such as the recently established Eidos Institute. Professional organizations like the Australian Association for Research in Education also play an active role in support of research in their field of expertise.

Increasingly, overall system performance is judged using international and national performance indicators. For example, overall education-participation rates in 2005 were 96% for 15-year-olds, 78% for 17-year-olds, 58% for 19-year-olds and 25% for 24-year-olds. School retention rates from years 7/8 to year 12 were 69% for males and 80.6% for females. Young people who are neither studying nor in the labor force are of particular concern, and are deemed to be at risk: 14% of 15–19-year-olds and 22% of 20–24-year-olds fell into this category.

The Australian Qualifications Framework indicates the level of knowledge, skills, and competencies associated with qualifications issued by accredited education and training authorities and include year-12 certificates, VET and higher education degrees, diplomas, certificates, and awards. In 2005, 51% of the 15–64-year-old age group held at least one nonschool qualification as their highest qualification, and half had completed year 12. The percentage of Australians aged 25–34 with a tertiary qualification (36% in 2002) was above the Organization for Economic Cooperation and Development (OECD) average (29%), but below that of countries like Canada, Japan, and Korea.

In addition to participation rates, performance indicators include learning outcomes. The National Assessment Program includes regular assessments of reading, writing, and numeracy performance at years 3, 5, 7, and 9, and of areas such as scientific literacy, information technology, civics, and citizenship. In 2005, the proportion not reaching national benchmarks in numeracy was 6% for year 3, rising to 9% for year 5, and 18% for year 7. At the international level, Australia consistently ranks among the highest performing countries in the OECD Program for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS).

The performance framework for VET revolves around objectives established under the national strategy for 2004–10. For example, most VET students (84%) and employers

(79%) are satisfied with the training provided. Higher education performance is assessed by institutions using international and national indicators and benchmarks of the quality of teaching, research outputs, graduate employment, and other outcomes set out in the national institutional-assessment framework. These assessments are audited by the Australian University Quality Assurance Agency.

Major Changes and Issues since the 1990s

The major changes and issues facing the Australian education system since the 1990s stem from the processes of globalization and the dominance of market economic theories in shaping policy. Education is big business, contributing an estimated 4.6% to Australia's GDP, and accounting for 7% of its workforce. International education is now Australia's fourth-largest export.

The application of market principles generally leads to policies designed to deregulate the education market. The Australian government has encouraged greater choice and diversity at all levels of education, and promoted cost sharing between participants in the education system and the society as a whole. The ultimate goal has been to increase competitiveness by improving the quality of education while containing or reducing its cost to government.

In most OECD countries, public investment in education actually increased between 1995 and 2002, regardless of changes in private spending. Australia is the main exception. The shift toward private expenditure has been accompanied by a fall in the level of public expenditure in real terms (especially in higher education). Critics of higher education policies of the national government claim that Australia is at risk of falling behind as the global knowledge economy gathers momentum. Efforts to increase Australia's share of the international education market have also led to criticism from abroad.

Successive national governments have encouraged state governments to adopt common approaches on issues such as school-starting age, teachers' registration and salaries, curriculum and assessment reform, quality assurance, and standard setting. A consensus has been reached on many issues such as the measures needed to improve early childhood education, literacy and numeracy outcomes, retention and participation levels, the attractiveness and quality of teaching, and education for indigenous children. Reforms at the state and territory level have focused on priority areas agreed at MCEETYA meetings, notably those set out in the National Goals for Schooling and Skilling Australia's Workforce. The Australian government has used its funding mechanisms to realign education in accordance with its national human capital reform agenda, which includes privatization and vocationalization of

education; standards and accountability mechanisms; partnerships with industry; science, engineering, and information technology; and history, civics, and citizenship education.

National intervention in areas for which the states are responsible has not always been welcomed, and, so on occasion, national initiatives (such as performance-based remuneration, and national year-12 curricula and examinations) have met with strong resistance from state and territory governments. However, the financial power of the Australian government and global pressures for reform mean that a more cohesive national approach to education and training seems inevitable. In 2008, the Australian Government established a National Curriculum Board to develop a nationwide K-12 curriculum in English, mathematics, science, and history.

While it is accepted that national and state governments must assume their responsibility to ensure that education systems and institutions meet their goals and are accountable, how this is to be done and by whom remains an issue. Those who believe that critical thinking, creative independence, and innovation are vital if we are to meet future challenges, argue that education and research in Australia are overregulated and underfunded. It is also likely that in an unequal, conflicted, and environmentally threatened world, there will be a renewed emphasis on education for sustainable development, and on what United Nations Educational Scientific and Cultural Organization (UNESCO)'s International Commission on Education for the 21st Century insisted are the key issues for the future – learning to live together and learning to be.

Further Reading

- Australian Bureau of Statistics (2007). *Australian Yearbook*. Canberra: Australian Bureau of Statistics.
 Australian Council for Educational Research (2007). *Year 12 Curriculum Content and Achievement Levels*. Melbourne: ACER.

- Australia Department of Education, Science and Training (2007). *Annual Report*. Canberra: Australian Government Publishing Service.
 Australia Department of Education, Science and Training (2005). *Skillings Australia: New Directions for Vocational Education and Training*. Canberra: Australian Government Publishing Service.
 Australian Government Productivity Commission (2007). *Report on Government Services, Part B-Education*. Canberra: Australian Government Publishing Service.
 Australian Vice Chancellors Committee (2001). *Our Universities: Our Future*. Canberra: AVCC.
 Hughes, P. (2007). *Opening Doors to the Future*. Canberra: Australian College of Educators.
 Marginson, S. (1997). *Education Australia: Governance, Economy and Citizen since 1960*. Cambridge: Cambridge University Press.
 Ministerial Council on Education (Employment and Training) (1999). *Adelaide Declaration on National Goals for Schooling in the Twenty-First Century*. Adelaide: MCEETYA.
 OECD (2005). *Education at a Glance: OECD Indicators 2005*. Paris: OECD.
 UNESCO (1996). *Learning: The treasure within. Report of the International Commission on Education for the 21st Century*. Paris: UNESCO.

Relevant Websites

- <http://www.annualreport.dest.gov.au> – Annual Reports, Department of Education, Employment and Workplace Relations, Australian Government.
<http://www.abs.gov.au> – Australian Bureau of Statistics.
<http://www.austcolled.com.au> – Australian College of Educators.
<http://www.acer.edu.au> – Australian Council for Educational Research.
<http://www.apo.org.au> – Australian Policy Online. <http://www.auqa.edu.au> Australian Universities Quality Agency.
<http://www.curriculum.edu.au> – Curriculum Corporation.
<http://www.dest.gov.au> – Department of Education, Employment and Workplace Relations, Australian Government.
<http://www.eidos.org.au> – Eidos Institute.
<http://www.edna.edu.au> – Education Network Australia.
<http://www.mceetya.edu.au>.
<http://www.ncver.edu.au> – National Centre for Vocational Education Research (NCVER).
<http://www.pc.gov.au/gsp/reports> – Report on Government Services, Report on Government Services, Productivity Commission, Australian Government.
<http://www.teachingaustralia.edu.au> – Teaching Australia.

Description of the Educational System

Education Population and Language of Instruction

In the academic year 2004/05, approximately 57% of the population under the age of 30 were in education, and there were 848 000 young people of compulsory education age. The language of instruction is predominantly German but, in regions with a linguistic minority, instruction at primary schools is divided between German and Slovenian, respectively, and Hungarian or Croatian in bilingual institutions.

Administrative Control and Extent of Public-Sector Funded Education

In the academic year 2004/05, nearly 88% of students attended state-funded educational establishments. Private institutions accounted for the remaining 12%.

Responsibilities for legislation and its implementation are divided between the federation (*Bund*) and the *Länder* (where it is executed by the *Länder* parliaments and the *Ämter der Landesregierungen*).

In specific matters enumerated in the Constitution, the federation sets the framework, while detailed legislation is implemented by the *Länder* parliaments (*Landtage*).

The Federation has an overwhelming responsibility for the educational system, including virtually all areas of school organization, the organization of school instruction, private schools, as well as the remuneration and retirement law governing educational staff.

As regards administration, the Federal Ministry of Education, Science and Culture is responsible for primary, secondary, and nonuniversity tertiary education, as well as for *Fachhochschulen* and universities. The Ministry of Economic Affairs and Labour has responsibility for in-company apprenticeship training. The *Länder* are mainly responsible for the provision of public-sector compulsory education. They support the local communities in establishing and maintaining these schools via school-construction funds that are administered by the *Länder*. The *Länder* have sole responsibility for crèches and kindergartens. Schools enjoy some autonomy in budgetary management and – up to a point – are free to adapt the curriculum to local needs. Province school inspectors in each of the nine Austrian *Länder* (assisted by district school inspectors for compulsory schools, and subject

inspectors for the intermediate and upper secondary levels) are responsible for inspections.

The 2002 *Universitätsgesetz* (University Act) transformed universities from federal institutions to public law entities which are no longer under federal administration. Federal funding consists of 3-year global budgets based on public-service agreements; universities are free to tap other sources of funding by virtue of their full legal capacity. Territorial entities and private legal entities may operate *Fachhochschulen*. At present, *Fachhochschule* study programs are mainly funded by the federal government in line with authorized study places.

Preprimary Education

Up to the age of 3, children may attend a crèche (*Krippe*), if available (Table 1). They may attend a kindergarten, when they are in the age range between 3 and 6. The training programs for kindergarten teachers have been aligned to take account of the growing number of children younger than 3 years who attend kindergarten. The provision of these facilities and decisions with regard to fees are the responsibility of the *Länder* or private institutions (providers). In 2005/06, 66.3% of all 3-year olds, 89.8% of all 4-year olds, and 91.9% of all 5-year olds attended pre-compulsory education (Figure 1).

Compulsory Education

Phases

Education is compulsory for 9 years and starts at the age of 6 (Table 2).

Sonderschule (special school; 6–15 years of age is described in section 7).

Admission criteria

Primary school pupils must have attained 6 years of age by 1 September in the year of admission. Children who have not yet attained the age of compulsory schooling and will be aged six by 1 March of the next following calendar year may be admitted earlier, if they are sufficiently mature and have the required social skills for attending school.

Transfer to lower secondary schooling requires successful completion of grade 4 of primary school. Admission to *Allgemein bildende höhere Schule* depends on achievement or is conditional on an admission test. With the exception of education in private schools, compulsory schooling is free of charge.

Length of the school day/week/year

The school year comprises 180 days (5 days a week) and lasts – according to a staggered schedule depending on the *Land* – from the first or second Monday in September to the Saturday between 28 June and 4 July, or between 5 July and 11 July.

Schools open 5 or 6 days a week.

The number of lessons per week ranges from 20 (primary school) to 35 (in secondary higher technical and vocational colleges). A lesson lasts 50 min. The minimum number of hours taught each year is around 630 for children of age 7, 750 for those aged 10, 870 at lower secondary, and 960 at the upper secondary level.

Class size/student grouping

The maximum class size is 30, while the minimum is 10 in primary school and 20 in secondary general school (*Hauptschule*) and the prevocational school. Pupils are grouped by age, but there is some grouping by ability in

German, mathematics, and modern foreign language at secondary level (secondary general school and prevocational school). At pre-vocational schools, pupils may be grouped in these subjects also by interest (or subject areas). As a rule, primary school pupils are taught by one teacher in all subjects except for religion and crafts; in secondary education, instruction is provided by subject teachers.

Curricular control and content

On the basis of proposals drafted by curricular task forces, the Ministry for Education, Science and Culture establishes a curricular framework in a consultation process that includes district and province educational bodies and organizations of teacher representatives.

The Ministry also approves textbooks. Schools are free to make their choice and enjoy some freedom to adapt the curriculum to the local context. Primary school compulsory subjects include religion, early science, German, reading, writing, mathematics, music, arts, textile/technical work, exercise and sports, modern foreign language, and road safety (without grading).

As of the school year 2003/04 onward, a modern foreign language has been taught on a mandatory basis at all primary schools – starting in grade 1. At the lower

Table 1

Krippe	0–3 years
Kindergarten	3–6 years

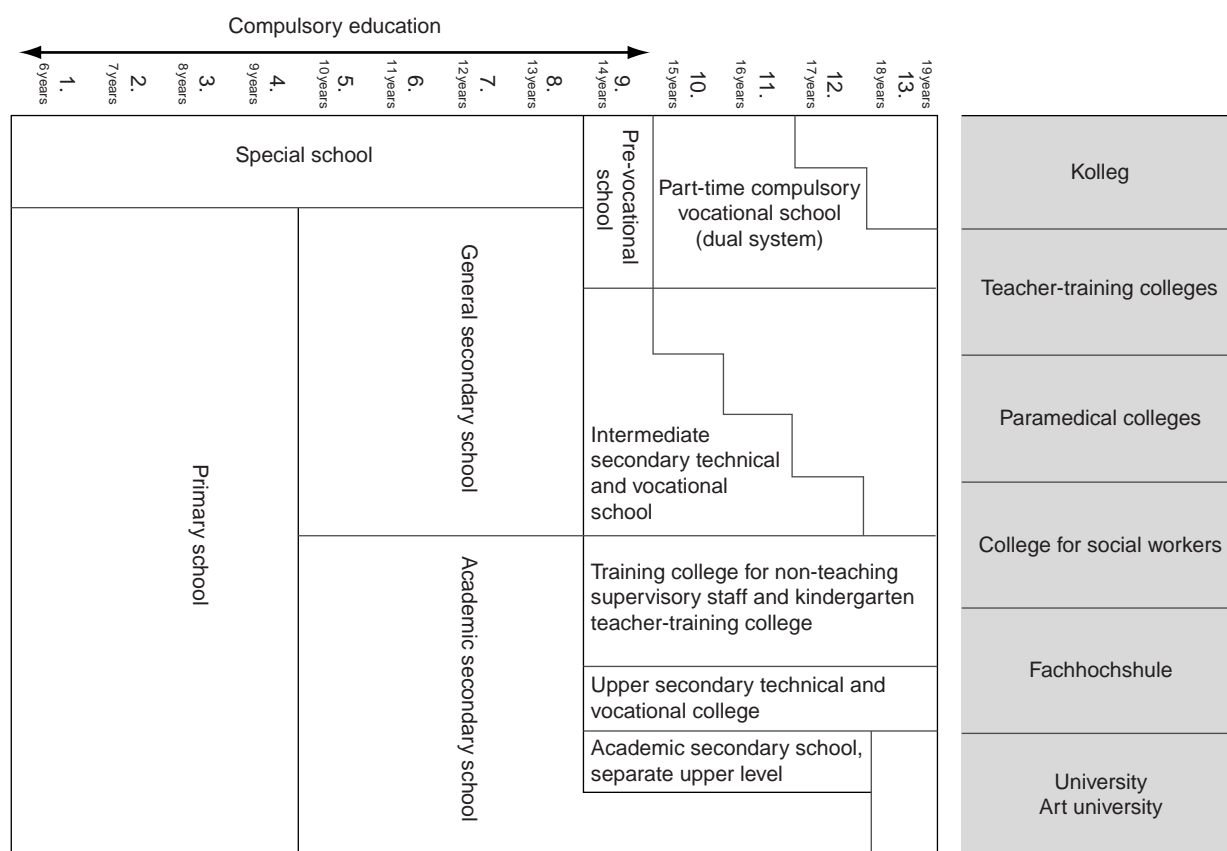


Figure 1

Table 2

(Volksschule or Grundschule) – primary education	6–10 years of age
Lower secondary education	10–14 years of age
Volksschule (upper primary level)	
Hauptschule (secondary general school)	
Allgemein bildende höhere Schule (secondary academic school – lower level)	
Grade 9	From 14 years of age
Polytechnische Schule (pre-vocational school)	
Or any other upper secondary school	

secondary level, pupils continue these subjects, in addition to science and technical subjects.

A special committee composed of teachers and parents (and – at secondary, intermediate, and higher level – students also) exercises an advisory and decision-making function in defined matters that are laid down in law.

Assessment, progression, and qualifications

There is no formal external testing during compulsory education. Teacher-generated assessment is based on classroom participation and oral, written, practical, and graphical work. Primary school pupils have to sit for written exams (school tests) in German and mathematics in grade 4.

As a rule, progression to the next higher form depends on achievement in all subjects (with exceptions at primary school level or unsatisfactory achievement in only one subject).

Pupils receive reports at the end of each term and at the end of the academic year.

In German and mathematics, pupils may attend instruction in the next lower or next higher form – if this is more appropriate for their individual learning situation.

Upper Secondary Education and Postsecondary Level

Types of education

*Although students start courses at these schools at the age of 14, compulsory education ends when they are 15 (Table 3).

Admission criteria

Following grade 8, students may opt for one of the many different courses in either the general or technical and vocational education (TVE) sector offered at secondary level II. Transfer requires successful completion of grade 8 and is also based on achievement.

Candidates failing to meet the achievement criteria must undergo admission testing.

Pupils without a leaving certificate following grade 8 may attend a pre-vocational school. The aim is to convey the best-possible support and skills for transfer

Table 3

<i>Upper secondary education</i>	<i>Age</i>
Allgemein bildende höhere Schule (secondary academic school – upper level)	14*–18
Polytechnische Schule (pre-vocational school)	14*–15
Berufsbildende mittlere Schule (technical and vocational school)	14+*
Berufsbildende höhere Schule (technical and vocational college)	14*–19
Bildungsanstalten für Kindergartenpädagogik/ Sozialpädagogik (schools for the training of kindergarten teachers and nonteaching education staff)	14*–19
Berufsbildende Pflichtschule (part-time vocational school)	15+
Post-secondary education Kolleg	18–20

to vocational training later on (dual system of training) or a higher-level school. Transferring back from secondary higher general or technical and vocational school to a pre-vocational school is possible only until 31 December of the current school year.

For some training routes at secondary higher technical and vocational schools or colleges, or colleges for kindergarten teachers and nonteaching education staff, candidates must prove their physical or artistic aptitudes as well as social and communication skills.

Following completion of 9 years of compulsory schooling, pupils may enter into an apprenticeship contract with an employer, minors require parental consent. Apprentices – at the same time – become pupils of a compulsory part-time vocational school.

Curricular control and content

Curricula at this level of education depend on the course chosen, although German, mathematics, and a modern foreign language are part of the general mandatory syllabus.

At pre-vocational schools, the emphasis lies on vocational orientation and basic vocational training in specialist areas according to widely scoped occupational fields such as technical-commercial occupations, commerce-clerical occupations, as well as service – industry and tourism.

TVE schools and colleges offer a wide variety of courses and special emphases. Consultation with the social partners is important, as curricula also include technical theory and practical instruction.

Teachers are free to select methods. The curricula emphasize student-centered learning, project work, cross-curricular activities, and the opportunity to acquire key skills.

In addition, in this area, textbooks and other materials are selected by the school.

Assessment, progression, and qualifications

Students who pass the examination at the end of an Allgemein bildende höhere Schule receive a matriculation certificate (*Reifeprüfungszeugnis*).

Apprentices receive a certificate on completion of their part-time course at school.

They may sit for an end-of-apprenticeship examination.

The *Lehrabschlussprüfungszeugnis* gives access to certain trades, foreman or master craftsman courses, professional diploma examinations, and courses for 'working adults', as well as to the *Berufsreifeprüfung* and the *Studienberechtigungsprüfung* (access to higher education).

Students on courses lasting 1–2 years receive pre-vocational training certificates.

Students at 3–4-year TVE schools take a leaving examination (written/practical and oral). The certificate entitles them to exercise certain regulated occupations or trades and provides access to upgrading or to the *Berufsreifeprüfung* examination.

Students of 5-year higher-level TVE courses may take the *Reife- und Diplomprüfung* – which includes written/practical and oral parts – and leads to double qualifications comprising general higher education access and professional qualifications for immediate entry into employment at middle-management level (EU diploma-level recognition).

Students of post-secondary courses (e.g., *Kollegs*) take a *Diplomprüfung* which is evidence of the acquisition of professional qualifications.

Higher Education Structure, Admission Criteria, Qualifications Awarded

Types of institutions

Higher education is offered at:

- Public universities including the arts universities;
- *Fachhochschulen* (institutes of technical/vocational higher education);
- the *Krems University of Continuing Education* (Danube University Krems);
- Private universities (following accreditation);
- *Pädagogische Akademien and Berufspädagogische Akademien* (general and vocational teacher-training colleges; as of October 2007 *Pädagogische Hochschulen*, Universities of Education);
- *Akademien für gebobene technisch-medizinische Dienste und Hebammenakademien* (colleges for higher-level medico-technical professions and midwifery colleges);
- Moreover, there are a number of educational institutions which offer university-type courses.

Access

Applicants to university (and to the *Akademien*) must have the *Reifeprüfungszeugnis* (or a *Reife- und Diplomprüfungszeugnis*) from a secondary higher school, or, alternatively, a *Berufsreifeprüfungszeugnis* or *Studienberechtigungsprüfungszeugnis* for the particular course of study. Depending on the course chosen, they may have to appear for supplementary examinations.

In 2005, there was a major change in the access requirements. Admission testing was introduced for those study programs which are affected by eight *numerus clausus* – study programs in Germany. This measure will last until 2007 and allows the restriction of access to university studies, mainly for medical study programs.

At the teacher-training colleges for vocational schools, the *Reifeprüfung* may be replaced by a *Meisterprüfung* (master-craftsman examination; the *Reifeprüfung* is, generally, required to access the future Universities of Education).

All applicants to universities of art and music have to pass an entrance examination. Candidates for a *Fachhochschule* – which, normally, requires an entrance exam – may either have a *Reifeprüfungszeugnis* or equivalent, or relevant vocational qualifications, as well as additional examinations, as the case may be.

The general tuition fees for university programs amount to € 363.36 per term for students from Austria and other EU – and EEA – countries, and to € 727.27 for other foreign students (exemption and reimbursement regulations apply to students from developing countries or to students in mobility programs). An amount of € 363.36 may be collected for *Fachhochschule* study programs.

Qualifications

Students at teacher-training study programs at *Pädagogische und Berufspädagogische Akademien* graduate as '*Diplompädagogin/Diplompädagoge*.' The teacher-training study programs at the future universities of education will lead to graduation as bachelor of education.

Students at colleges for higher-level medico-technical professions graduate as follows:

'*Physiotherapeutin/Physiotherapeut*'
 '*Biomedizinische Analytikerin/Biomedizinischer Analytiker*'
 '*Radiologietechnologin/Radiologietechnologe*'
 '*Diätologin/Diätologe*'
 '*Ergotherapeutin/Ergotherapeut*'
 '*Logopädin/Logopäde*'
 '*Orthoptistin/Orthoptist*'

Studies at midwifery colleges lead to graduation as a midwife.

As of the academic year 2006/07, training for higher-level medico-technical professions and midwives is also offered at *Fachhochschulen* in the form of bachelor-study programs.

University studies end with the following degrees:

Regular programs

- Diploma (after 8–12 terms)
- Bachelor (6–8 terms)
- Master (2–4 terms after the bachelor)
- Doctor (4 terms following the master or diploma degree)

- University programs ending with an international master degree, for example, master of business administration (MBA);
- University programs ending with the designation of *akademischer* . . . (at least 60 ECTS).

Fachhochschule programs end with a diploma degree – since 2002 – also with a bachelor and master degree similar to the classification for the universities. *Fachhochschule* graduates are allowed to enter doctoral programs at universities.

Special Needs

Pupils with special educational needs may be educated either in special schools or in inclusive settings in primary and lower secondary schools. Parents have the right to choose the kind of schooling they prefer for their child. Special curricula and/or adapted mainstream curricula are applied in response to pupils' individual needs. During the school year 2002/03, over 50% of all pupils with special educational needs attended integration classes.

Compulsory education

To be admitted to a special school, a pupil's special educational needs must be determined by way of official decree (Table 4).

Class size/student grouping

The maximum class size is as follows:

- special school for blind, deaf, and severely disabled pupils – 8;
- special school for visually impaired, hearing impaired pupils and remedial schools – 10;
- others – 15.

Curricular control and content

Pupils are taught according to the curricula of primary school or secondary general school or of the lower cycle of secondary academic school, if they are, generally, able to attain the objectives of instruction without being overtaxed. In all other cases, the curriculum of a special

school – geared to their disability – is to be applied. There are special curricula for general special schools (for performance-impaired pupils or pupils with learning difficulties), for the special schools for blind children, the special schools for deaf children, and the special schools for severely disabled children.

Assessment, progression, and qualifications

Pupils with special educational needs at general schools are entitled to transfer to the next higher form if this provides better opportunities for the pupil's overall development; the class conference decides on such transfers.

In the absence of pertaining statutory provisions, inclusive settings during the last (9th) year of compulsory schooling are, at present, provided in pilot projects only.

Specific programs for the transition period between school and the labor market are funded by the Ministry of Social Security, Generations and Consumer Protection and implemented in cooperation with schools.

Vocational training options for disabled or disadvantaged young persons

The Vocational Training Act was amended in the summer of 2003, adopting the integrative vocational training model. This amendment created a basis for young people with disabilities or at a disadvantage for better employment opportunities. Ever since, the number of young persons in integrative vocational training has risen steadily.

In addition, the Apprenticeship without Barriers campaign was launched to inform the public, companies, as well as the young people affected and their parents with regard to the opportunities of integrative vocational training. This media campaign, moreover, aims at raising awareness and understanding of the public at large for encouraging even more companies to accept apprentices for integrative vocational training.

Teachers

Kindergarten teachers complete 5 years of training from the age of 14 onward or a 2-year post-secondary course.

Primary and lower secondary school teachers must have completed a 3-year course at a teacher-training college (as of October 2007 *Pädagogische Hochschulen*, Universities of Education).

Teachers at an *Allgemein bildende höhere Schule* must complete a university study program of at least 4.5 years leading to a degree.

Depending on their subject area, teachers for the TVE sector are trained either at universities or at vocational teacher-training colleges (as of 2007, at universities of education). A relevant professional background and experience is required for teaching technical theory and practice at TVE schools and colleges.

Table 4

<i>Sonderschule</i> (special school): General special school consists of levels I and II and of an upper cycle.	6–15 years of age
Specialized branches for physically disabled children, for children with linguistic disabilities, for hearing impaired, deaf, visually impaired, blind, severely disabled, and maladjusted children.	
Pupils may attend special school during the entire period of compulsory schooling	

Primary school teachers do not specialize, whereas lower and upper secondary school teachers specialize in two subjects (at TVE schools and colleges, they may also specialize in a group of subjects or a subject area).

Teachers are employed by the federation or the *Länder* (civil servants).

Teachers at colleges for higher-level medico-technical professions and midwifery colleges are trained in specialist courses (following general training).

Current Reforms and Priorities

Reforms in Primary, Lower, and Upper Secondary Education

Afternoon childcare

Nearly 80 000 pupils have enrolled for afternoon childcare as per the autumn of 2006. This is more than 10% of a total of 760 000 pupils for whom afternoon childcare arrangements are available.

Every child in need of afternoon care should have a right to attend. Providing afternoon childcare to meet the given needs is an important task to be jointly fulfilled by the federation, the *Länder*, and the local authorities.

Nearly 10 h of afternoon childcare are allocated for 15 children each by the federation to the *Länder* at compulsory school level. Afternoon-childcare groups may be set up across classes, years, or different schools.

In early 2006, the schools informed the parents regarding afternoon childcare programs and identified existing needs. In the school year 2006/07, afternoon childcare will be offered wherever needed, while respecting parental freedom of choice.

Five-day week

As of the autumn of 2006, Saturdays at schools catering to the 6–14 age range will be free. Pupils should not work on more days of the week than their parents and will now have more time to spend together with their families.

Early language learning

Allocation of 1964 teaching posts as of the school year 2006/07 has made for a swift integration in primary school. The existing 1600 posts will be stepped up by an additional 330. Children at primary school who do not know German can henceforth be taught up to 11 h per week in small groups.

Learning German even earlier is better still. This is why the federal government sponsors language-learning courses at municipal and local nursery schools and has invested a total of € 640 000.

In recent months, 8000 children benefited from language-learning courses that are to ensure a good start of their school careers.

Expanding support for specially gifted pupils

Dispensation provisions (premature admission to primary school) and the possibility to skip a class (even at the primary/secondary transition) will take better account of the needs of highly gifted pupils. Reflecting their specific talents and skills, children should be able to terminate their school careers earlier.

Upgrading of class heads

Class heads have an important role to play in terms of furthering and assisting talented as well as underachieving pupils. This includes the early detection of strengths and weaknesses to make sure that children and adolescents can be supported in the best-possible manner through accompanying educational measures. They are also responsible for effectively implementing the early-warning system in cooperation with the school partners (parents, pupils, etc.), ensuring the target- and needs-driven use of remedial instruction, coordinating the educational and recreational activities during afternoon childcare, assuring high-quality teaching, and participating in evaluation. As of autumn 2006, class heads at compulsory school level will receive an additional € 70 per month as a performance-linked remuneration for these activities.

Schools now starting on Monday in the first school week

Instruction guarantee: No more than 2.5% of all classes may be canceled.

Monday and Tuesday in the first school week are days of familiarization. In order to ensure that classes start on the Wednesday of that week in line with the timetable, examination resits may be held on Thursday and Friday of the last week of the main holidays as of autumn 2006/07. The final dates will be determined by the school-partner forums.

Blocking of teaching units

According to an amendment to the School Periods Act (*Schulzeitgesetz*), teaching units may now be blocked. Classes can, therefore, be organized more flexibly and modern teaching methods – such as project work or interdisciplinary instruction – can be used.

Higher Education Reforms

University reform

The 1999 amendment to the *Universitäts-Studiengesetz* (University Education Act) introduced the three-tier system of universities studies, excluding teacher training, human medicine, and dental medicine studies – which are only offered as diploma programs. In the mid-range, diploma studies will, generally, be replaced by bachelor and master study programs. In the winter term of 2006,

211 bachelor and 277 master study programs existed alongside 179 diploma study programs.

The 2002 *Universitätsgesetz* (Universities Act) gave full legal status to the universities as of 1 January 2004. Newly recruited university staff is now employed under private law contracts. Universities are funded via a global budget which is set up for 3-year periods based on public service agreements which define the services to be rendered by the Federal Ministry and the university by way of negotiation. The first public service agreements between the universities and the Federal Ministry of Education,

Science and Culture will be concluded in the autumn of 2006. The first period for public service agreements is 2007–09. Following that time, every university must submit an annual performance report on the state of implementation of the public service agreement as well as an intellectual capital report. Budget reporting is done via the financial statements based on a commercial system of accounting. Twenty percent of all university budget allocations are formula-linked. This formula covers indicators on teaching, research and development, on the development of the arts, as well as on social objectives.

Bangladesh

R Z Dewan, Director (Research and Documentation) (Retired), National Academy for Educational Management (NAEM), Ministry of Education, Government of Bangladesh, Dhaka, Bangladesh

© 2010 Elsevier Ltd. All rights reserved.

Glossary

General stream of education – This is the mainstream at all stages of education: primary, secondary, and higher. The curriculum of general education at all levels contain arts, science, and business subjects forming substreams at different stages.

Madrasah – Islamic religious schools.

Madrasah education – Refers to the religious stream of education at all the three stages of education: primary, secondary, and higher education.

Stream – Streams are different courses of study at a particular stage of education.

Structure – Structure of education refers to the organization of the education system in terms of successive cycles of education along with their duration.

Upazila – Subdistrict.

Zila – District.

General Background

Introduction

Bangladesh emerged as a sovereign country in the world map in 1971. It lies in the northeastern part of South Asia. Having an area of 147 570 km² and inhabited by 139 million people, the country has the highest population density in the world. Bangladesh has a per capita gross domestic product (GDP) of US\$456 and the GDP growth rate is 5.38%. The adult literacy rate is 62.6% and the average life expectancy at birth is 65 years. Categorized as a low-income country (World Bank, 2006), the country ranks 137 in the list of 177 countries of the world in terms of Human Development Index (UNDP, 2006).

The territory constituting Bangladesh was a part of the Indian subcontinent which was under the Muslim rule from 1201 to 1757. In 1757, the British defeated the last Muslim ruler of Bengal and established British rule which lasted up to 1947, over the whole subcontinent. The British rule ended in 1947 through the partition of the subcontinent into India and Pakistan. Pakistan comprised two parts – East and West. The present Bangladesh was the

eastern part of Pakistan and was known as East Pakistan. Pakistan existed for 24 years. In 1971, East Pakistan won freedom through a 9-month War of Liberation.

The country is bounded on three sides by India and Myanmar and by the Bay of Bengal on the South. It is mostly constituted of plain fertile land composed of alluvial soil having some hilly regions in some parts of the country. The major rivers of the country, along with their 230 tributaries, form a river network which, covering the whole country, flow down to the Bay of Bengal. In the south-western part of the country along the coast, stands the Sunderbans – the world's largest mangrove forest. The total forest area covers about 14% of the land area. Bangladesh has a subtropical monsoon climate.

Poverty remains the critical issue to resolve. The national poverty rate based on the cost-of-basic-needs (CBN) method (using upper poverty line) is 40.0. The health situation has improved over the years in terms of safe drinking water users, and child (1–4 years) and maternal mortality rates. Over the period 1995–2005, the road network has enhanced by 34%. Bangladesh has recently been connected with high-speed submarine cable consortium.

Bangladesh has, according to the 2001 census, a population of 130.0 million. The population projected for 2006 stands at 139 million. The population density is 928 persons per square kilometer. The population growth rate is 1.48%. Bangladesh is a Muslim-majority country, having 89.6% Muslims, 9.2% Hindus, 0.7% Buddhists, 0.3% Christians, and 0.2% other ethnic minority groups. The sex ratio of the population is 105.4 males per 100 females. Twenty-five percent of the population are urban dwellers, while 75% live in villages. The total civilian labor force of the country is estimated at 44.3 million, of which 34.5 million are males and the rest females.

Bangladesh has a parliamentary form of government, with the prime minister as the head of government, while the president is the head of the state. The prime minister is assisted by a council of ministers. The National Parliament of the country is a uni-cameral legislature consisting of 345 members.

The country has six administrative divisions each of which comprises a number of districts (*zila*). Each district consists of a number of subdistricts called *upazila*, which is the lowest administrative unit. There is a local government covering both urban and rural areas. The local government is entrusted to bodies elected by the people.

Bangladesh is a linguistically, religiously, and culturally diverse country. The Bangla-speaking group accounts for 98% of the population. The remaining 2% consists of about 45 ethnic minority groups speaking their own languages and have their own religions.

These diversities in religion and languages have made the country own a checkered culture.

Bangladesh is an agrarian country with 52% of the labor force engaged in agriculture. Agriculture and fisheries contribute about 21% to the GDP. Although the economy is agro-based, industrialization is taking place at a slow pace. The industrial sector is contributing about 17% to the GDP. Other sectors having notable contribution to the GDP are construction, trade, and services sectors. Coal deposits are there awaiting exploration. The country has a considerable deposit of natural gas. Other mineral resources include hard rock, limestone, and silica sand.

Legal Frame Under Which the Education System Operates

Article 15 and 17 of the Constitution contains the national directives regarding education. Article 15(a) recognizes education as the basic need of the people and also describes fulfillment of this need as a state responsibility. The nation's vision regarding education has been articulated in Article 17 of the Constitution. Article 17 reads as follows:

The state shall adopt effective measures for the purpose of

- (a) establishing a uniform, mass-oriented and universal system of education and extending free and compulsory education to all children to such stage as may be determined by law.
- (b) relating education to the needs of society and producing properly trained and motivated citizens to serve those needs.
- (c) removing illiteracy within such time as may be determined by law.

Besides, Bangladesh is a signatory to World Declaration on Education For All and the Dakar Framework for Action adopted in Jomtien, Thailand in 1990 and at Dakar, Senegal in 2000, respectively. In the context of these constitutional and international obligations, acts, laws, rules, and regulations have been framed for different levels of education.

Background of the Education System

In Muslim India, there were two levels of education: primary and higher. However, the levels were not very

structured or organized. The primary-level institutions were of two types based on languages: Persian and indigenous (mainly Sanskrit) language. Higher education institutions also had two major types: *madrasas* for Muslim students and *tolls* for Hindu students.

Many schools were situated in the houses of people or private lands and were funded by the head of the family, community people, rich men of the locality, and, in some cases, the local rulers. Teaching was mainly a mission. The curriculum included literature, logic, religion, grammar, mathematics, astrology, and medicine. A learner generally needed 10–12 years to complete higher education.

During the early years of the British rule, this system of education continued. The first document relating to education, popularly known as Wood's dispatch, was published in 1854 by Mr. Charles Wood, an officer of the East India Company. This document included recommendations regarding primary to higher education levels.

Immediately after the publication of Wood's dispatch, in 1855, a few universities at different places of India were established for facilitating higher education. However, primary and secondary education were not given due importance. Thereafter, an Education Commission, under the leadership of William Hunter, was constituted. The commission's report was published in 1882. On the basis of this report, primary and secondary education were organized. The British rule in India ended with the division of the subcontinents into two sovereign nations – Pakistan and India. The education system in Pakistan was a continuation of the British system. Bangladesh inherited this system.

Goals/Objectives of Education System

In order to fulfill the constitutional obligations, the government of the country constituted education commissions/committees and task forces at different times. Educational goals set in these documents are reflected in the development plans and the curricula. Summarized, these goals are as follows:

- to accomplish learners' overall growth and development and to enable their creativity to flourish;
- to introduce work-oriented, value, and need-based education;
- to ensure equality in terms of gender and educational opportunities at all levels of education;
- to extend the foundation for science and technology education and creating a knowledge society; and
- to create conscientious, patriotic, and productive citizens capable of contributing to the overall development of the country.

Structure and Operation of the Education System

The Structure of Education

Education in Bangladesh has three major stages: primary, secondary, and higher education. Primary education is a 5-year cycle, while secondary education is a 7-year cycle having three substages: 3 years of junior secondary, 2 years of secondary, and 2 years of higher secondary. Higher secondary is followed by higher education in different streams (general, *madrasah*, technology, and medicine), which require 3/4/5 years to obtain a first degree and another 2/1/1 years to obtain a master's degree.

In the general stream, after the higher secondary, a learner needs 3 years to get a first degree with pass and 4 years for a first degree with honors. The pass and honors degree holders need 2 years and 1 year, respectively to get the master's degree. For a first degree in engineering and business, 4 years of schooling is necessary, while for a first degree in medicine and architecture, the requirement is 5 years.

The entry age for primary is 6 years. The corresponding age groups for primary, secondary, and tertiary education cycles are 6–10, 11–17, and 18–22.

There is a preprimary system run by many primary schools, English medium schools, kindergartens, and nursery schools mainly located in urban areas. Mosque-based institutions, such as *maktabs* also offer preprimary along with religious teaching. This stage is prior to primary and extends from 1–3 years. This preprimary stage is not recognized by the government (Figure 1). However, government is planning to include the preprimary stage into the formal education structure.

Stream

Primary education has two major streams – general and religious – while secondary education has three major streams: general, technical–vocational, and religious. Higher education has four streams: general (inclusive of pure and applied science, arts, law, business, and social science), religious (*madrasah* education), medicine, and technology (engineering, agriculture, textile, and leather) education. Teacher education forms a separate stream.

Another stream of education is running parallel to the general stream at the primary through secondary levels. This is the English medium schools located in city areas and offering the English system of 'O' (ordinary) and 'A' (advanced) levels. Mostly, city-dwelling and children of well-off people are enrolled in these schools. The learners of this system constitute a negligible portion of the total strength.

Institutions and Courses of Study

Primary

On the basis of ownership and management, primary-level institutions may be classed under four major heads:

1. government primary schools (GPSs),
2. registered nongovernment primary schools (RNGPSs),
3. nonregistered nongovernment primary schools (NRNGPSs), and
4. others.

The last category includes experimental schools, community schools, kindergartens, nongovernmental organization (NGO)-run schools, *Ebtedayee madrasahs*, and satellite schools. The total number of institutions, teachers, and

Grades	Cycles and corresponding ages	Cycles and corresponding ages
17	Master's degree 21–22 years	Master's degree 22 years
16		Ist degree (honours)/
15	Ist degree (pass) 18–20 years	business/engineering/
14		technology
13		18–21 years
12	Higher sec. (general, TVET, <i>madrasah</i>)	
11	16–17 years	
10	Secondary (general, TVET, <i>madrasah</i>)	
9	14–15 years	
8	Junior secondary 11–13 years	
7		
6		
5	Primary 6–10 years	
4		
3		
2		
1		

Figure 1 Structure of education (cycles, duration, and age). Reproduced from Jalaluddin, *et al.* (1992). *Secondary Education in Bangladesh: A sub-sector Study*. Dhaka: DSHE.

pupils in the primary education subsector are presented in **Table 1**.

Female teachers and students account for 36% and 50% of the total numbers, respectively. The average pupil–teacher ratio stands at 47:1. It is the highest in government primary schools and is almost double of that in the other primary-level institutions in which it is 29:1. Government primary schools account for 47% of the total strength and form the mainstream.

Secondary: General stream

The total number of educational institutions at the secondary level (general stream) is 20313. This number comprises:

1. junior secondary schools, all of which are non-government;
2. government-owned and managed secondary schools;
3. nongovernment secondary schools; and
4. government and nongovernment higher secondary schools.

Statistics related to the number of institutions, teachers, and students in these institutions is presented in **Table 2**.

A vast majority of the secondary-level institutions – 98% – are nongovernment.

Secondary: Technical and vocational stream

Technical vocational education and training (TVET) offers certificate and diploma courses to the learners. The certificate courses prepare skilled workers in different

vocations and trades spread over 1–2 years after completion of the junior secondary level in the general education stream. The diploma courses offer diploma in different branches of engineering. This is a 3-year course, and the minimum requirement for admission into these courses, is secondary-level (grades 6–10) completion certificate, that is, the Secondary School Certificate (SSC).

There is a variety of institutes in the area of TVET. The total number of institutions in 2005 stands at 2722 and that of the trainers/instructors thereof, at 17140. The enrollment comes to 241336. The pupil–teacher ratio is 14:1. Females account for 19% and 26% in the trainer and student categories, respectively.

Secondary: Religious stream (madrasah education)

For the secondary-level *madrasah* education, there are Dakhil and Alim *madrasas*. The major focus of *madrasah* education is on the Holy Quran and Al-Hadith. Courses in Bangla, English, arithmetic, science, and social science, as prescribed in the general education stream, are also taught. The total number of secondary-level *madrasas* is 8000. Teachers and enrollment are 123757 and 2786838, respectively. Females account for 9% and 48% in the teacher and pupil categories, respectively. The pupil–teacher ratio is 22:1.

Higher education

Higher education is imparted by two sets of institutions: colleges (all streams; public and private) offering first and

Table 1 Number of primary schools, teachers, and enrollment

Sl. No.	Type of schools	No. of schools	No. of teachers	Enrollment	Pupil:teacher
1.	Government	37 672	162 084	948 3891	58:1
2.	Nongovernment registered	19 682	76 566	3 572 686	47:1
3.	Nongovernment nonregistered	946	3456	158 059	46:1
4.	Others (including Ebtedayee madrasah)	22 101	102 683	3 011 022	29:1
	<i>Total</i>	80 401	344 789 (36.25%)	16 225 658 (50.13)	47:1

Figures in parenthesis indicate the percentage of females.

From the Baseline Report of the 2nd Primary Education Development Programme (2006), Phase-2 (PEDP 2). DPE, Ministry of Primary and Mass Education, Dhaka, Bangladesh.

Table 2 Number of secondary (general stream) schools, teachers, and enrollment (2005)

Sl. no.	Type of school	No. of schools	No. of teachers	Enrollment	Pupil–teacher ratio
1.	Junior secondary	4322	36 122(20%)	910 914 (58%)	25:1
2.	Govt. secondary	317	7452 (31%)	221 887 (47%)	30:1
3.	Nongovernment secondary	13861	194 584 (19%)	6 265 751 (52%)	32:1
4.	Higher secondary (both govt. and nongovernment)	1813	35 408 (20%)	261 229 (48%)	7:1
	<i>Total</i>	20 313	273 566(20%)	7 659 781 (52%)	28:1

Figures in parenthesis indicate the percentage of females.

From Bangladesh Bureau of Educational Information and Statistics (BANBEIS) (2006). *Bangladesh Educational Statistics, 2006*, Dhaka.

master's degree courses and universities (general and specialized; public and private). Colleges offering general education are affiliated to the National University (NU) and colleges for other streams are affiliated to the respective specialized universities. There are general universities for the general stream and specialized universities for studying medicine, agriculture, and technology. For higher-level *madrasah* education, there are Fazil and Kamil *madrasahs*. The universities also offer MPhil and PhD courses in selected subjects. Besides, there is an Open University which offers secondary to higher education level courses (mainly general stream) through distance mode. **Table 3** shows the number of institutions offering first and master's degrees in general, *madrasah*, technology, and medical streams:

Nonformal Education

The Ministry of Primary and Mass Education (MOPME) executes the nonformal education (NFE) policies through the Bureau of Non-Formal Education (BNFE) at the national level. The BNFE is headed by a director general. The organizational structure provides for 64 offices in 64 districts of the country. The NFE programs are disseminated through continuing education centers (CECs) and learning centers.

NFE covers disadvantaged groups, including children and youth with physical and mental disabilities, ethnic minorities, people living in ecologically difficult locations, street and working children, and people living in especially difficult circumstances.

Initially, the NFE programs were designed to impart literacy to illiterates of the 11–45 age group. At present, the employment-oriented Post-Literacy and Continuing Education (PLCE) project is being implemented, which covers 1.6 million neo-literates with a 9-month course. The other project run by this Bureau is the Basic Education For Hard To Reach Urban Working Children (BEH-TRUWC) which targets 0.2 million urban working children aged 10–14 years with 40 months of basic education program.

Teacher Education

There are 54 primary teacher training institutes popularly known as PTIs offering Certificate in Education (C-in-Ed). These are all public institutions and spread all over the country. For secondary school teachers, there are 99 (14 public and 85 private) teacher training colleges (TTCs) offering Bachelor of Education (BEd) degree. The Institute of Education and Research (IER) of the University of Dhaka also offers BEd courses. The Open University, through the distance mode, imparts BEd course to secondary teachers. There are training institutes for technical–vocational, *madrasah*, and physical education teachers and also for higher secondary teachers. The total number of teacher-training institutes, trainers, and enrollment (trainees) for all-level teachers are 188, 2123, and 36265, respectively, for the year 2005. The trainee–trainer ratio is 17:1. Besides, there are two national academies – National Academy for Primary Education (NAPE) and National Academy for Educational Management (NAEM) – for imparting educational management training to teachers and administrators of primary and secondary subsectors, respectively.

Although it is a precondition for entry into the teaching profession to have the requisite pedagogical training, a large proportion of teachers in the system are still untrained. The relevant training for a primary teacher is C-in-Ed, while that for a secondary teacher is BEd.

Operation of the System

The net enrollment ratio at the primary is 87%. The repetition rate at this level is 11.2% and as a percentage of the relevant age group, the completion rate is 73%. The survival rate to grade 5 is 53.9% and the years input per graduate is 8.2. The pupil–teacher ratio for all types of schools is 47:1. Taken individually, it is the highest in the mainstream government primary schools which is 58:1. Only 12% of the GPSs and 9% of the RNGPSs run single shift, while the rest run double shift. The contact hours for single shift and double shift schools are 900 h and 600 h, respectively, per academic year. The percentage of trained

Table 3 Number of institutions, teachers, and students at the higher education level (2005)

No.	Institutions by stream and level	No. of institutions	No. of teachers	Enrollment
1.	First-degree (pass and honors) college (general stream)	1248	48 069	719 524
2.	Master's degree college (general stream)	89	6924	386 493
3.	Medical, engineering, and other (dental, law, textile, leather) technology colleges/institutes	127	3460	42 507
4.	Madrasah (higher education level: Fazil and Kamil madrasahs)	1214	28 292	669 256
5.	University (public and private; general and specialized universities)	74	10 339	207 577
	<i>Total</i>	2752	97 084	2 025 857

From Bangladesh Bureau of Educational Information and Statistics (BANBEIS) (2006). *Bangladesh Educational Statistics*, 2006.

teachers at the primary is 73% and the transition rate from primary to secondary is 83.3%.

With the 58:1 pupil–teacher ratio in the mainstream primary schools, contact hours of 600 per annum in most of the schools, and 27% of teachers being untrained, the primary-level institutions are striving hard to ensure quality.

In the secondary (general), the net enrollment ratio is 48%. The dropout and completion rates in the secondary cycle are 80% and 20%, respectively. In the general, technical, and madrasah streams of secondary education, the pupil–teacher ratios are: 28:1, 14:1, and 22:1, respectively. The contact hours at this level, on an average, is 5 h a day. The number of working days is 228. With a very high dropout ratio and almost half of the teaching force being untrained, and even trained teachers mostly following the teacher-centered method of teaching, ensuring quality poses a problem. This is reflected in the success rates at the SSC and Higher Secondary Certificate (HSC) results which are 59% and 64%, respectively, in 2006.

At the primary level, there is no general public examination. There is a system of continuous pupil assessment (CPA) for grades 1–3. At the end of grades 3 and 4, there are annual examinations for promotion to the next grade. At the end of the cycle, there is the school-leaving examination conducted by the respective schools.

At the junior secondary level also, there is no public examination. The most important public examination is held at the end of grade 10. This is the SSC examination. For internal evaluation, pupils of each grade from 6 through 10 have to take three term examinations.

Those who complete the secondary level, have to take a 2-year higher secondary course and sit for a public examination titled HSC examination. The SSC and HSC are organized, conducted, and controlled by the Board of Intermediate and Secondary Education (BISE). For Madrasah and Technical Education, there is the Madrasah Education Board and Technical Education Board.

At the tertiary level, student evaluation is the responsibility of the universities.

Teacher Management

Recruitment of government primary school teachers is carried out through a competitive written examination conducted by a central committee headed by the Director General of the Directorate of Primary Education (DPE). For nongovernment schools, it is the School Managing Committee (SMC) that appoints teachers in accordance with the nongovernment Teacher Recruitment Rules. The minimum qualifications required for the primary teacher is the SSC for females and HSC for males, with a preservice training titled C-in-Ed.

Teachers of government and nongovernment secondary schools, colleges, *madrasahs*, and other institutions are recruited through competitive examinations. Government teachers are recruited by the Directorate of Secondary and Higher Education (DSHE) and nongovernment teachers by Nongovernment Teacher Registration and Certification Authority (NTRCA), respectively. After recruitment by the NTRCA, the Appointment Board constituted at the institutional level concerned and headed by the chairman of the SMC, selects and appoints teachers in nongovernment institutions. The eligibility requirements for a secondary teacher are a first degree from a university and a preservice training/degree, that is, BEd.

The pay scale for both government and nongovernment teachers are the same, but there are differences in salary-payment practices. The government pays the salary, including the allowances of government teachers, but for nongovernment teachers, the government pays only the basic pay, and not the allowances. The allowances have to be collected from other sources, including tuition fees from pupils.

The career prospects for teachers at both the government and nongovernment primary and secondary levels, are not bright. In course of time, assistant teachers become senior teachers and their possibility of becoming headteachers is bleak, as the positions of headteachers are few in relation to the number of teachers.

Teachers in government institutions are evaluated annually through the criteria set in a format called the Annual Confidential Report (ACR). Teachers in nongovernment institutions are evaluated informally. The evaluation has a direct link with their promotion.

Teachers at all levels of government institutions are transferable on completion of 3 years of service in one station. Teachers in nongovernment institutions are not transferable.

Administration and Supervisory Structure

The education system of Bangladesh functions under two separate ministries: MOPME, responsible for primary and mass education, and the Ministry of Education (MoE), responsible for secondary and higher education.

Administration

Administrative structure: Primary

The MOPME is the policymaking body for primary and mass education. The ministry is headed by an adviser with the status of a state minister, under the direct supervision of the prime minister. The adviser is assisted by the Secretariat headed by a secretary of Primary and Mass Education.

The principal executing body for the Secretariat is the DPE. The directorate is headed by a director general who

has six directors heading six units. These units are: training, finance, operations, monitoring and evaluation, planning and development, and administration. A number of deputy directors (DDs) and assistant directors (ADs) assist the directors of the six units. This is the national level. The directorate has structures at the zone, district, and subdistrict levels.

At the zonal level, the Zonal Primary Education Office is headed by the zonal DDs who are assisted by the ADs. Zonal DDs are under the direct control of the director of administration located at the national level.

At the district level, there is the district primary education officer (DPEO) and assistant district primary education officer (ADPEO). The subdistrict structure is manned by *upazilla* (subdistrict) education officers (UEOs). Each UEO in the 481 *upazillas* of the country is assisted by one assistant *upazilla* education officer (AUEO).

Alongside the DPE, there is a Compulsory Primary Education Implementation and Monitoring Unit (CPEIMU) for implementing and monitoring the programs and activities undertaken for executing the provisions of the Compulsory Primary Education Act enacted in 1990 and for ensuring education for all (EFA).

Administrative structure: Secondary and higher education (imparted by colleges)

The (MOE) is the policymaking body and also the apex body for educational administration, management, and planning for secondary and higher education. The MOE is headed by a minister who is assisted by a state minister. The ministry executes its functions through the secretariat which is headed by the education secretary.

The secretariat is assisted by two administrative directorates: Directorate of Secondary and Higher Education (DSHE) and Directorate of Technical Education (DTE) which have the key roles in the administration and management of postprimary education subsectors for the general and technical education streams. The government policy and planning for general secondary and higher education is implemented by the DSHE, while those for the technical education (secondary level) is implemented by the DTE.

The DSHE is headed by a director general who is assisted by 4 directors at the center, who head the four units of the directorate: (1) planning, (2) secondary education, (3) college education and administration, and (4) training. The directorate has substructures at the zonal, district, and subdistrict levels. The zonal offices are headed by DDs who are assisted by ADs and school inspectors. The district offices are headed by district education officers (DEOs) who are assisted by the respective assistant district education officers (ADEOs). At the subdistrict level, there are *upazila* (subdistrict) secondary education officers (USEOs) and assistant *upazila* secondary education officers (AUSEOs).

Likewise, the DTE is headed by a director general who has the requisite support offices at the central and lower levels.

There are some staff organizations located at the center which extend support to the administrative functions of DSHE. The National Curriculum and Textbook Board (NCTB) develops curricula for the primary and secondary (general streams) subsectors. There are six education boards called the Board of Intermediate and Secondary Education (BISE), one Madrasah Education Board (MEB), and one Technical Education Board (TEB). These boards are responsible for the accreditation of nongovernment secondary and higher secondary education institutions (general, *madrasah*, and technical-vocational streams) and the administration of public examination at the secondary- and higher-secondary levels (SSC and HSC examinations). MEB and TEB are also charged with the responsibilities of developing the curricula for Madrasah and technical-vocational streams of education, respectively. For administering public examinations of higher education imparted by colleges, there is the National University. The Directorate of Inspection and Audit (DIA) is engaged in inspecting the schools to ensure mainly the financial accountability of the schools.

The Bangladesh Bureau of Educational Information and Statistics (BANBEIS) is responsible for collecting data and furnishing information. Infrastructure support to the postprimary education subsectors is provided by the Education Engineering Directorate (EED). There are two apex bodies for providing professional support to MOE and MOPME. The support is provided through management training to education managers and administrators and research support to the two ministries for policy and decision making. These are (i) National Academy for Primary Education (NAPE) and (ii) National Academy for Educational Management (NAEM). The former is for primary education and the latter for secondary and higher education subsectors.

The above structure applies to both government and nongovernment institutions. One important feature of the administration and management of nongovernment primary, secondary, and higher-level educational institutions is the school-based management through the school managing committees (SMCs). For the college-level institutions, which impart higher education, these bodies are called the governing bodies. These are elected bodies having representatives from the community and parents. These bodies are responsible for all in-school management issues in the nongovernment schools, except academic matters. The heads of schools and colleges are the member secretaries of these administrative bodies.

Educational planning of the country is mainly a top-down process. The Planning Commission, an organ of the Ministry of Planning (MOP), is the principal authority for formulating the national-level plans of the country. The education wing of the Planning Commission formulates

the education plan. The planning cell in the MOE and MOPME, and the planning units of the DPE, DSHE, and DTE provide inputs to the education wing for planning primary, secondary, and higher levels of education.

Administrative structure: Higher education imparted by universities

There are 74 universities in the country. Of these, 21 are public universities and the rest 53 are private. These include general, medical, agricultural, engineering, and technology universities. Universities are autonomous bodies having their own set of rules and principles to govern themselves. However, there is the University Grants Commission (UGC), a supervisory body which provides operational regulations and allocates fund to the public universities.

Teacher Supervision and Support

Primary

At the primary level, the inspectorate comprises mainly the UEOs and AUEOs who work at the field level. The main functions of the UEOs and AUEOs are to visit schools which aim at both teacher supervision and school inspection. There are supervision guidelines, reporting formats, and also provision for follow-up actions. There is a regular monitoring system of the supervision staff.

The AUEOs have to organize subcluster (one subcluster comprising 15–20 schools) training for in-service teachers every month. The other in-service training methods that are imparted to the teachers include subject-based training, training on teaching methods, and management training for headteachers. Supply of teaching aids, learning materials, and teacher guides are the other forms of teacher support.

Secondary

Supervision of secondary schools and teachers are particularly the job of regional, district, and subdistrict offices of DSHE. Each of the nine DD's office is manned, so far as the supervisory staff is concerned, by two inspectors of school (ISs) and two assistant inspectors of schools (AISs). At the district level, the DDs are supported by DEOs who form the pivot of academic inspection and supervision activities. Each DEO is assisted by an ADEO. The USEOs and AUSEOs are the school supervisors at the subdistrict level. Besides, there is a body of academic supervisors (ASs) who have school supervision as their major responsibility.

The DIA and the BISE have responsibilities of inspecting and supervising schools for various purposes but not specifically for the purpose of academic supervision.

Despite the prevalence of a body of academic supervisors, academic supervision remains a weak area of secondary-education management mainly due to the high school-supervisor ratio. If the DEOs, who form the main body of academic supervisors, are considered, then the ratio is 314:1.

Financing of Education

As a percentage of gross national product (GNP), expenditure on education during the years 2000–04 was 2.2% which compares favorably with an average allocation of 1.8% during 1993–94.

The allocation to education in the revenue and development budgets of the government accounts for 15.03% and 12.8%, respectively, of the total allocations in 2005–06 (BANBEIS, 2006). The primary and mass education subsector is given the highest allocation of 61.51% in the development budget for education for the year 2005–06, followed by secondary and higher (general stream) education – 27.7%. In the revenue budget for the same year, the share of primary and mass education stands at 32.7%, while that for secondary and higher education stands at 52.5%. The rest goes to technical, university, *madrassa* education, teacher training, administrative and subsidiary services, etc.

The development budget for education is financed from two major sources – government and aid, grant, loan from development partners. The 2005–06 development budget has been shared at a 60:40 proportion by the Government of Bangladesh (GOB) and development partners. The other sources of financing education include parents and the community. NGOs also finance education development but mostly for primary and adult-literacy programs.

Nongovernment institutions at the primary and secondary levels account for 53% and 98%, respectively. The government pays only the basic salary of teaching and nonteaching staff of nongovernment institutions. The rest of the cost is borne by the private sector. In nongovernment institutions that account for 53% and 98% of the primary- and secondary-level institutions, respectively, the major share of the costs of primary and secondary education is borne by the private sector.

Performance Monitoring, Evaluation, and Research

The performance of each level of education is monitored by the administrative agencies concerned. The monitoring and evaluation of primary education is done by the monitoring and evaluation unit of the DPE. As there is no specific unit or cell for monitoring and evaluation in the DSHE, this responsibility is discharged by the secondary and college and administration units of the directorate. The performance indicators that are used are: literacy rate, participation rate (gross and net enrolment rates), and indicators of internal efficiency such as repetition, dropout, survival rates, the success rates, and the like.

Educational development programs and projects are undertaken on the basis of feasibility studies. Research studies prior to educational reforms are also practiced. These studies are conducted by different national agencies

under the MOE and development partners in collaboration with national agencies. The linkage between education research and education reforms and decision making is not yet well established. There is no central agency to undertake education research to extend research support to educational policy and decision making on a regular basis. However, public research institutions, such as the Institute of Education and Research (IER) of the University of Dhaka, the National Academy for Educational Management (NAEM) working under the MOE, and the Bangladesh Institute of Development Studies (BIDS), are organizations and agencies that are doing the needful in this respect at present.

Major Changes, Challenges, and Issues since 1990

Changes

The educational sector has undergone remarkable changes since 1990. In 1990, the Compulsory Primary Education Act was enacted. Children's Stipend Programme was introduced for retaining poor children in schools. This yielded a notable improvement in enrollment. Over a period of 10 years (1990–2000), enrollment enhanced by 46%. During the period 1990–2005, girls' participation enhanced from 45% to 50%.

For ensuring better administration and management of primary and mass education, a separate ministry titled Ministry of Primary and Mass Education, was established.

Another change that deserves mention in the primary education arena is the change in curriculum. From a traditional one, the curriculum was changed to a competency-based one in 1992. The curriculum containing 50 terminal competencies to be achieved by learners, was implemented phase-wise during the period 1992–97.

In order to enhance access and gender parity in secondary education, financial support in the form of stipends for girls through a number of female secondary stipend projects (FSSPs), was introduced in 1994, and this is still continuing. This resulted in enormous growth in girls' participation in secondary education, enhancing gender parity.

Because of very low girls' participation in technical education, the government has established six new polytechnic institutes for girls to support a wider coverage of girls.

In the secondary-education subsector, changes occurred in learner assessment. Previously, learners were assessed on the knowledge aspect of education only. With the introduction of school-based assessment (SBA), where 30% marks have been earmarked for assessing learners' behavioral aspects, the psychomotor and affective domains have come under the purview of assessment. The other reform in learner assessment is the introduction of the grade point average (GPA) system in 2001 in place of the absolute marking system. This has ensured more effective assessment of learners' merit and achievements.

A very major intervention in the quality arena is the enactment of the Nongovernment Teacher Registration and Certification Authority Act in 2005, aimed at recruiting and appointing quality teachers for the nongovernment educational institutions of all postprimary subsectors which constitute 98% of the total strength. The other objective is to bring them under the fold of registration which was never done in the past. Now, nongovernment teachers are needed to be registered with and certified by the authority for entry into the teaching profession.

In the higher education arena, the enactment of the Private University Act in 1992 had been a great step forward toward privatizing higher education and enhancing access to higher education. At present, 53 out of the 74 universities are privately owned.

Challenges and Issues

Till today, the major challenge is ensuring the participation of all in the education system. Despite remarkable achievements in enrollment and gender parity at both the primary and secondary levels, 13% of the primary-, 52% of the secondary-school-age children are still out of school.

Apart from this, the prime concern is quality. The internal efficiency is low as demonstrated by a high dropout and low completion rates both at the primary and secondary levels. Years input per student at primary is 8.2 years. In the secondary cycle, the dropout rate is 80%, the completion rate is 20%, and the repetition rate is 10%. Fifty percent of children at the primary level fail to achieve more than 50% of basic competencies. Improving this situation is the major challenge to the government and the education actors.

Low instructional time, ineffective teaching–learning, characterized by no lesson plan, and a weak learner–assessment system, are the major issues at the primary- and secondary-education subsectors.

The success rate at the public examination at the secondary- and higher-secondary levels (SSC and HSC) ranged between 42–53% and 39–59%, respectively, during the period 2000–05. All these point to the quality issue. The relevance of the curriculum is questionable as it is oriented toward higher education and examination, and not to practical skills. The assessment system has drawbacks. The common criticism against the system is that it encourages rote learning and not the understanding of the learners.

Improving the quality of secondary teachers is a challenge. Only 53% of the teachers have the requisite professional qualifications.

Enhancing the participation of women as both teachers and students still remains a challenge despite the fact that gender parity has been achieved at the pupil level at both the primary and secondary subsectors. Female teachers account for 36%, 20%, 19%, and 9% in the primary, secondary (general), secondary (technical), and secondary (*madrasah*) streams.

In the TVET arena, the major challenges are to establish linkage to the job market and to improve the effectiveness of training support and delivery.

The challenges facing higher education are improving the external efficiency, making higher education need-responsive, and establishing linkages to national and global markets.

Bibliography

Bangladesh Bureau of Educational Information and Statistics (BANBEIS) (2006). *Bangladesh Educational Statistics, 2006*, Dhaka.
 UNDP (2006). *The Human Development Report*. Basingstoke: Palgrave Macmillan.
 World Bank (2006). *World Development Indicators, 2006*.

Further Reading

Bangladesh Bureau of Educational Information and Statistics (BANBEIS) (1992). *Education System of Bangladesh*. Dhaka.

Bangladesh Bureau of Statistics, Planning Division, Ministry of Planning (2001). *Population Census 2001*.
 Bangladesh Bureau of Statistics, Planning Division, Ministry of Planning (2005). *Bangladesh Statistical Yearbook, 2005*.
 Bureau of Non Formal Education, *NonFormal Education (NFE) Policy (2006)*, Ministry of Primary and Mass Education (MoPME), Government of Bangladesh.
 Dewan, R. Z., Huque, M. A., and Hoque, M. M. (2004). *School Principals: Core Actors in Educational Improvement*. Paris: IIEP, UNESCO.
 Directorate of Primary Education, Ministry of Primary and Mass Education, Bangladesh (2006). *Baseline Report of Primary Education Development Project Phase-2, (PEDP 2)*, Dhaka.
 Jalaluddin, A. K. (1992). *Secondary Education in Bangladesh: A Sub-Sector Study*. DSHE: Dhaka.
 Ministry of Finance, Government of Bangladesh (2006). *Bangladesh Economic Review, 2006*, Dhaka.
 Ministry of Labour and Manpower, Government of Bangladesh (2003). *Labour Force Survey 2003*, Dhaka.
 Muqtadir, M. A., Mozammel, M. M., and Dewan, R. Z. (1998). *National Teacher Supervision and Support Services for Basic Education in Bangladesh: Supervision and Support Services in Asia*. vol. II, Paris: IIEP, UNESCO.
 Planning Commission, Ministry of Planning, Government of Bangladesh (2005). *National Strategy for Accelerated Poverty Reduction: Unlocking the Potential*. Dhaka: BG Press.

Barbados

R Rudder, Ministry of Education, St. Michael, Barbados

© 2010 Elsevier Ltd. All rights reserved.

General Background

Barbados is a small, island country of 166 square miles with a population of over 270 000 people. Most Barbadians are of African descent.

The first inhabitants of Barbados were Amerindians, who arrived here from Venezuela. The island was taken over by the Spanish in 1492. Spain, however, left Barbados in favor of the larger Caribbean islands. This left the island open when the British came to colonize it. The first English ship came to the island on 14 May 1625, under the command of Captain John Powell. The island was, therefore, claimed on behalf of King James I. On 17 February 1627, Captain Henry Powell landed with a party of 80 settlers and ten slaves to occupy and settle on the island. This expedition landed in Holetown, formerly known as Jamestown. The colonists established a House of Assembly in 1639. It was the third ever parliamentary democracy in the world.

The earliest attempts at formal education were introduced in approximately 1686, when a charity school was established for the education of poor white children. With the further efforts of private individuals, more charity schools were established for poor white children; however, it was only in 1818 that the first school for colored boys was established. In 1827, an elementary school for colored girls was established. By 1844, there were 215 schools – with a total enrolment of 7452 students – operated by churches, as well as private individuals throughout the island.

The first Education Act was passed, in 1850, and it provided for the establishment of an Education Committee with a part-time school inspector as its executive officer, who was also simultaneously, the head of the Central School for Boys. The committee was given responsibility for the educational policies and the administering of the annual grant of £3000, to be distributed to the schools in proportion to the average attendance of pupils and the amount raised in school fees.

In 1858, a second Education Bill was passed. The Inspector became a full-time officer and the subjects of the curriculum were determined. The pupil–teacher system was introduced and payment by results was initiated. Elementary education developed rapidly during this period and government grants were increased, annually, to £9200 by 1874.

The first serious attempt to organize higher education came in the early 1870s. During 1874–75, two important events occurred in education. The first was the affiliation

of Codrington College to Durham University and the establishment of a Classical Facility. The second was the appointment of a Commission – headed by Bishop Mitchinson – to study the system of education in the island and make recommendations for its improvement.

The Report of the Mitchinson Commission was completed in 1876. It found favor with the authorities who embodied the recommendations in the 1890 Education Act, which remained the principal Education Act until 1983. This Act resulted in many far-reaching consequences for the educational system. An Education Board – consisting of four members of the House of Assembly, three of the Legislative council, and two others, all of whom were appointed by the governor – replaced the former Committee. An inspector and an assistant inspector served as its officers.

The Board had general control of elementary education and could require the vestries (local parish authorities) to provide and maintain elementary school buildings.

Under the 1878 Act, two schools were designated First Grade Schools with governing bodies and received annual grants. Provision was also made for Second Grade Schools, (between First and Elementary) to be established. By the end of the nineteenth century, there were three First Grade Schools and five Second Grade Schools. There were 169 elementary schools receiving financial support from government with an enrolment of 24 145 students: 134 were Anglican, 19 Wesleyan, and 16 Moravian. School fees were charged at all schools.

Although education was provided from public funds, there was no distinction between government Elementary Schools and Church Schools. The Education Board with the Bishop as president made the rules and regulations for elementary schools. Teachers were appointed by the local Board of Managers and approved by the Education Board. Buildings were maintained by the vestries.

Many private schools were in existence at the time, but there was “no inspection or notice taken of them.” The number of children being educated in these schools – at that time – was in excess of 10 000.

During the first 25 years of the twentieth century, the educational system came in for severe criticism. It was reported that the Board lacked a definite policy and had lost control through decentralization. Elementary school buildings were unsuitable for educational purposes and grossly overcrowded. Elementary education was described as removed from the needs of the children, which rendered them unemployable to a large degree. Education in the

First Grade Schools was described as being severely academic and having little contact with reality. The Second Grade Schools were also described as unsatisfactory. Without adequate facilities, they tried to imitate the First Grade Schools and, as a result, they provided neither culture nor technical skill.

The Marriott–Mayhew Commission was appointed, in 1932, “to make a comprehensive investigation into the educational services of the colony.” The report – published in 1933 – drew attention to the need for the appointment of a Director of Education, and the necessity of improving the facilities for teacher-training, at the Rawle Training Institute attached to Codrington College. The Commission also advocated a reorganization of the schools and a review of the curriculum to bring education into closer touch with the needs of the students and the times. An important recommendation was the introduction of the modern Secondary School. Very few of the recommendations could be implemented during the years following the report because of social unrest in the region and the outbreak of World War II.

In 1943, the first recommendations of the Marriott–Mayhew Report came into force. A Director of Education was appointed and the Education Board became an advisory body. Several changes were immediately put into effect in the teaching service. Teachers in elementary schools became civil servants; the regulations for the award of certificates of competency were revised in relation to a new salary scale; secondary teachers’ salaries were also improved but they were not granted civil servant status.

The Rawle Training Institute, opened in 1912, was closed in 1945, with plans to open a new institution nearer to Bridgetown. Other changes taking place were the introduction of age-grouping in elementary schools, strengthening of the administration and inspectorate by creating the posts of assistant director, inspector, and assistant inspector of schools and three specialist posts for Handicraft, Domestic Science, and Infant Methods. A new Housecraft Centre was opened for training women in domestic science, in 1945.

Another milestone in educational history was reached with the creation of a separate Ministry of Education in 1958. The Minister’s responsibility for policy and development of education was more clearly recognized. The establishment for the teaching staff in primary and secondary modern schools was increased and provisions made for more acting staff.

The most significant step in the field of education was made in 1962, when school fees were abolished in all government secondary schools. This meant that all persons of ability would have free access to secondary education. In 1962, the Department of Education was integrated with the Ministry of Education and the Director of Education was re-designated the Chief Education Officer.

In 1963, a school Meals Pilot Scheme was introduced in three primary schools. During the same year, the College of Arts and Science at the University of the West Indies was opened. Erdiston College reintroduced the 2-year program while the 1-year course was continued for experienced, unqualified teachers.

A new Education Act was proclaimed in 1983. Its main objective was the democratization of education and the regularizing of the management and operation of all secondary schools. Under this Act, all teachers have become members of the civil service.

The Act is very comprehensive and covers all aspects of education from school attendance, registration of private institutions, duties of teachers, to management of primary and composite schools, and provision of special education.

Administration of Ministry of Education

The Ministry of Education was established in 1954, under the portfolio of the Premier. In 1958, a separate ministry was established with its own staff of administrative and technical officers to assist the minister in the execution of government policy.

The Ministry is divided into two main sections – technical and administrative. The Chief Education Officer, who heads the technical staff, is the Chief Professional Advisor while the Permanent Secretary is the Chief Administrative Officer with responsibility for finance. This dual system functions rather effectively, but can be a bit intricate where roles are not clearly defined.

The administrative section has responsibility for personnel, financial, and general administrative matters while the technical section has responsibility for specific educational matters.

For administrative purposes, the two main sections were subdivided into four divisions namely:

1. The Schools Division
2. The Planning and Development Division
3. The Personnel Management and Services Division
4. The Finance Division

In 1997, there was a restructuring of the Administrative and Technical sections to meet the new reform process initiated by the White Paper on Education Reform (1995).

The divisions have now been renamed:

1. Schools Supervision and Management Division
2. Planning, Research and Development Division
3. Human resources Division
4. Education – Management Information System Division
5. Finance Section
6. Student Services Division
7. Curriculum Division

8. Testing and Measurement Division
9. International Relations Division
10. Higher Education Award Division
11. Office Management
12. Parent Volunteer Support Division

The Technical Division

The Technical Division is headed by the Chief Education Officer who has two Deputy Chiefs to assist her/him. This Division has responsibility for schools, educational planning, and development.

The Schools Division

The schools division is made up of two subunits and is supervised by a deputy chief education officer. Each section in the schools division has administrative personnel to assist the technical teams in carrying out their duties. These administrative persons include administrative officers 1 and 11, senior clerk, and clerical officers.

The subunits are (1) School Supervision and Management, and (2) Student Services.

The schools supervision and management unit

This embraces the primary and secondary schools and is, at present, managed by a senior education officer. The main functions of this subunit include the supervision of public and private assisted schools, registration, selection of teachers for appointments (at primary level), and advising schools on educational and administrative matters. The public primary and private assisted secondary schools are divided into five districts – each district being supervised by an Education Officer.

The main goals of the Ministry are:

- To ensure high-quality, free education for all from preschool to tertiary level
- To provide a wide range of higher education and training opportunities to enable those best able to avail themselves of the facilities to develop the professional, technical, and other skills they can use to further their careers and contribute to the development of the Barbadian economy and society.
- To promote gender equality, which will produce citizens who are well-balanced, disciplined, industrious, creative, self-reliant persons who can think critically and function effectively in a modern society.

Early-Childhood Education

The Government of Barbados is committed to the achievement of the United Nations' Declaration on Education for All – Expansion of Early Childhood Care and the CARICOM Plan of Action to address early-childhood

issues. To this end, the Ministry has embarked on a program to facilitate the expansion of early-childhood education that would ensure universal educational access for children between the ages of 3 and 5. This program was started in the academic year 2005/06 and 825 new students have been enrolled since the commencement of the Nursery Expansion project. By September 2007, 6560 3–5-year olds or 91.15% of the cohort were enrolled in pre-primary education. There are, at present, seven public nursery schools and many private nursery schools.

The curriculum at this stage focuses heavily on the oral language development of the students. The Early Childhood Education (ECE) program seeks to assist these young children in developing oral competence in standard English before they are called upon to demonstrate literacy in it. In addition, the program – at this level – seeks to develop the natural curiosity of children and pave the way for them to be critical and creative thinkers from an early age.

As specified by the Ministry of Education, any ECE program should strive to enhance the following characteristics of the students:

- Ability to talk freely
- Innate curiosity and desire to learn
- Playful activity
- Ability to imitate and interact with each other, as well as adults
- Ability to use concrete and multisensory material and tools
- Ease in simultaneously acquiring knowledge and skills in multiple areas
- Capacity to learn at different rates
- Ability to distinguish right from wrong
- Need for success to build a sense of security and self-confidence

Observation is a critical assessment procedure employed by early-childhood teachers. Teachers observe the children throughout the school day, paying special attention to their progress in terms of physical growth, social growth, as well as the evolving of their literacy, mathematical, and practical abilities.

Primary Education

In Barbados, primary education is provided for children between the ages of 3+ and 11. At present, there are 71 public primary schools and 20 private primary schools; together, they cater to approximately 28 000 students.

Modes of Assessment at the Primary Level

With the introduction of flexible transfer in 1996, children as young as 9+ can appear for the Barbados Secondary

School Entrance Examination (BSSEE) to gain admission to secondary schools if they are ready. Conversely, those who reach age 11+ and have not fully covered the primary syllabus can postpone appearing for the BSSEE for 1 year.

As a means of ensuring that pupils reap optimum benefit from their primary education, the Ministry of Education has introduced a number of innovations in the area of testing and assessment. One such innovation is the Criterion – Referenced Test (CRT), which is administered at the end of Infants B (~6-years old) and Class 2 (~8-years old). This test provides teachers with information on the strengths and weaknesses of students.

Secondary Education

Secondary Education is provided for students between the ages of 10+ and 16+ at 22 public secondary schools. Four of these schools – Harrison College, Queen's College, Combermere, and Lodge – provide sixth-form education. In addition to the public secondary schools, there are eight assisted private schools that offer secondary education.

Admission to secondary schools is based on the BSSEE. Students take the examinations of the Caribbean Examination Council (CXC) at the basic and/or general proficiency levels between the ages of 15 and 17 years. Most students write four or more subjects and, those whose performance is deemed to be excellent, may write the Advanced Level Examination of the Caribbean Examinations Council, the Caribbean Advanced Proficiency Examination (CAPE) 2 years later at 17–19 years of age. Excellent performance in this examination may result in the award of scholarships or exhibitions to pursue studies at colleges or universities, local or overseas.

Students at most of the newer secondary schools still appear for the Barbados Secondary School Leaving Certificate, Stage 1.

Tertiary Education

Education, at the tertiary level, is provided at vocational and technical colleges as well as at the university. Admission to these institutions varies from the possession of a Barbados Secondary School Certificate Stage I to a Caribbean Examinations Council Certificate to G.C.E A Level qualifications. Possession of these qualifications represent completion of secondary education. The institutions for tertiary education are the Samuel Jackman Prescod Polytechnic, Erdiston Teachers' College, the Barbados Community College, and the University of the West Indies.

University of the West Indies

Qualified Barbadian students are eligible to enter the University of the West Indies – a regional institution

with campuses at Mona in Jamaica, St. Augustine in Trinidad, and Cave Hill in Barbados. The University of the West Indies has faculties of medicine, science, arts and general studies, social sciences, agriculture, engineering, and education and law.

The campus at Cave Hill was established in 1965; however, classes were started at a temporary site near the Bridgetown Harbour, in 1963. Courses offered at the Cave Hill Campus lead to the bachelor's degree in arts and general studies, law, natural sciences, theology, social sciences, and education. In addition, a number of certificate, diploma, and graduate-degree programs are offered.

Erdiston Teachers' College

Erdiston Teachers' College is a nonresidential institution providing a basic 2-year teacher-education program for nongraduate teachers of primary and secondary schools. In addition to general teacher-education courses, the college cooperates with the Samuel Jackman Prescod Polytechnic and the Barbados Community College to provide certification in technical and vocational areas such as business education, industrial arts, and home economics. These courses are mainly for teachers at the secondary level. The college also offers a number of post-basic and continuing-education courses designed to meet the special needs of teachers and the public. Among these are 1-year in-service part-time courses in the teaching of reading and remedial education along with 1-year full-time courses in early-childhood education and physical education.

Barbados Community College

The Barbados Community College – established in January 1969 – is an institution designed to improve the facilities available to the community for training in a wide range of skills at the technical para-professional, middle-management, and preuniversity levels.

The college offers courses in the divisions of liberal arts, fine arts, health sciences, sciences, commerce, hospitality studies, technology, general and continuing education, the language centre, and the departments of computer studies and physical education. In addition, the college offers a wide range of evening and summer courses as part of its service to the community. The courses of studies are of 2 years' duration and lead to the associate degree in arts or applied arts, and sciences or applied sciences.

Nursing education was transferred from the Tercentenary School of Nursing of the Ministry of Health to the Division of Health Sciences of the Barbados Community College – with effect from April 1986.

Samuel Jackman Prescod Polytechnic

The Samuel Jackman Prescod Polytechnic – an institution offering a wide range of trade courses – was officially

opened in January 1970. Craft-level training – which had been offered at the Barbados Technical Institute, since 1953 – was reorganized and expanded in scope within this new institution. Its objectives are to develop trade skills and occupational competencies up to the level of skilled craftsmen, as well as to prepare students for entry into the Division of Technology of the Barbados Community College.

The Housecraft Centre – formerly located at Bay Street – has been incorporated into this institution as the Division of Human Ecology.

The types of training offered include:

1. training for students about to enter the labor market;
2. training for persons who wish to improve their skills; and
3. re-training for those desirous of changing their occupations.

The courses of training are offered either on a full-time, part-time, or evening basis and may lead to either local certification or qualifications of the City and Guilds of London Institute or CXC.

Government-Aided Programs in Schools

There are several programs designed specifically to assist the disadvantaged students so as to ensure that they can take advantage of the educational system. Some of these programs are described below.

Textbook Loan Scheme

The TextBook Loan scheme was established by the government to ensure that all secondary students have easy access to the texts required for the study of programs offered in schools. Each student – on payment of an annual rental fee – is issued with the required textbooks. Provision has been made for the waiver of the rental fee when this is justifiable.

The monies collected as rental fees are used for the management of the scheme. Government provides the funds, annually, to replenish the stock.

School Uniform Grants

A scheme for the provision of uniforms for school children was introduced in 1980. The scheme provides a one time grant to each qualified student entering a secondary school, the senior department of a composite school, or a senior school for the first time.

School Lunch Programme

In 1963, a Schools Lunch Pilot Programme was introduced under which children attending primary schools were provided with a cooked lunch every day. The facility was

gradually expanded until all primary, composite, and senior schools became involved in the program. In addition, two private schools are benefiting from this program.

The Teaching Profession

In 2006, there were 2912 teachers in government schools from the pre-primary to the secondary levels. Of this number, 2082 were female and 830 were male; these figures clearly show that the teaching profession is highly feminized. The minimum qualification for a teacher is certification from the CXC at the general level in at least five subjects – including English language and mathematics. Teachers' salaries are determined based on qualifications. The salaries range from \$22 551.60 per annum to \$59 575.00 per annum for teachers and, for principals, from \$79 608.72 per annum to \$98 505.24 per annum.

Education Sector Enhancement Project

The central goal of the Education Sector Enhancement Programme (ESEP) is to comprehensively reform the entire primary and secondary school system in Barbados so that its output will reflect an increase in the number of young persons who can significantly contribute to the sustainability of the social, cultural, and economic development of Barbados. The implementation of the reform initiative is being done through the execution of civil works activities, teacher training, curriculum reform, and the introduction and integration of technology into the school environment.

The Programme Coordinating Unit manages and coordinates all matters pertaining to the planning, implementation, and monitoring of the program in conjunction with other ministry sections, departments, and institutions. The Education Project Implementation Unit has special responsibility for the management and execution of civil works, project accounts, and some aspects of procurement in the program. In addition, the Curriculum Section is responsible for curriculum reform, while the Erdiston Teachers' Training College is responsible for teacher training under the human resource development component of the program.

Functions

The functions of the abovementioned units involved in the implementation of the program include:

- Liaising with the program's international funding institutions (IFIs) – the Caribbean Development Bank (CDB) and the Inter-American Development Bank (IDB).
- Negotiating and preparing contracts relevant to the procurement of goods and services for the program, and preparing relevant Cabinet Papers.
- Preparing budgets and reports for internal and external users/purposes.

- Coordinating and monitoring all civil works and activities related to the rehabilitation of schools in the program.
- Coordinating the procurement of technological equipment and software, technical assistance, and general professional services for the ESEP.
- Coordinating teacher training relevant to the program as well as the training of the ministry's technical and administrative personnel in program-related areas.
- Coordinating the preparation and implementation of revised curricula for the primary and secondary schools.
- Communicating and working with the schools to ensure the effective implementation and management of the reform program at the school level.
- Coordinating the evaluation of the program.
- Preparing and disseminating information on the program – including publications and multimedia presentations – for education stakeholders and for the general public.

Education Financing

The Government of Barbados is committed to the development of an educational system, which will enable all persons to reach their fullest potential. Education – at all

Table 1 Government expenditure on education

<i>Year</i>	<i>Education as a % of total government expenditure</i>
1933–34	10%
1940–41	13%
1950–51	16%
1960–61	12%
1970–71	20%
1979–80	19%
1982–83	18.40%
1997–98	16.39%
2000–01	18.55%
2004–05	16.98%
2007–08	16.68%

levels – is free for all Barbadian Nationals. In the pre-independence era, the expenditure on education fluctuated between 10% and 16% of the total government expenditure, and with independence came an increase in expenditure on education.

The education budget averages between 16% and 20% of the total annual budget of the country. This reflects the government's policy of using education as an instrument of social change. **Table 1** reflects the percentage of government expenditure on education from the 1930s to the present day.

Bulgaria

A Chavdarova and D Vesselinov, Sofia University "St. Kliment Ohridski", Sofia, Bulgaria

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Asparuh or Isperih – was ruler of a Bulgarian tribe in the second half of the seventh century and is credited with the establishment of the First Bulgarian Empire in 680/681. He is the most famous Bulgarian ruler. The accuracy of the Turkic title *khan*, commonly applied to him and his successors, is a subject of some dispute.

Old Great Bulgaria or Great Bulgaria – was a term used by Byzantine historians to refer to Onoguria during the reign of the Bulgarian ruler Kubrat (father of khan Asparuh) in the seventh century north of the Caucasus mountains in the steppe between the Dniester and Lower Volga.

The Bulgars – (also *Bolgars*, *Bulghars*, *Huno-Bulgars* or *Proto-Bulgarians*) were originally semi-nomadic people, probably of Turkic descent, originating in Central Asia, who from the second century onwards settled in different parts of Europe.

General Background

The Republic of Bulgaria is a country in Southeast Europe, a European Union (EU) member state (01 January 2007) and a member of the North Atlantic Treaty Organization (NATO) (02 April 2004). Up to the year 2007, the population of Bulgaria was 7 679 290. According to the 2001 census, 83.9% of the people in the country are Bulgarian, 9.4% are Turkish, and 4.7%, Roma. The official language in the country is Bulgarian, spoken by 84.5% of the people, Turkish is spoken by 9.6%, and Roma is used by 4.1%. Most Bulgarians are Orthodox (82.6%), 12.2% Muslim, 1.7% Roman Catholics, 0.8% Jewish, and the remaining 1.6% are Protestants or other.

The first Bulgarian state existed on the Balkan Peninsula, and its adjacent parts, of Southeast Europe from 680 until 1018. It was founded by Khan Asparuh, who, after the dissolution of Great Bulgaria, brought a part of the proto-Bulgarian tribes, subjected the local Slavs, and entered into a peace treaty with Byzantium in 681. In 864, Bulgaria converted to Orthodox Christianity and, in 866, the Slavic alphabet and writing were introduced.

The second Bulgarian state (1186–1393) was founded by the brothers Asen and Peter after the period of the Byzantine rule (1018–1186). In the twelfth century, following the military success of Tsar Kaloyan over the crusaders, the country grew stronger and during the rule of Ivan Asen II, there was extension of territory, economic, and cultural development, as well as reestablishment of the Bulgarian patriarchy. In 1396, Bulgaria fell under the Ottoman rule which continued for almost five centuries.

After the Russian Turkish war for liberation (1877–78) the San Stefano peace treaty was signed in 1878 and it reestablished the Bulgarian state and marked the beginning of the third Bulgarian state. The Berlin Treaty, signed in 1878, partitioned Bulgaria and established an independent Bulgarian Principality and an autonomous region Eastern Rumelia. The unification of Bulgaria was achieved in 1885. In 1908, following the proclamation of the independence of the state, the kingdom of Bulgaria was established.

After World War II, Bulgaria fell into the Russian sphere of influence and in 1946 it became a people's republic. The Bulgarian Communist Party (BCP) headed the country and its governing role was provided for in the Constitution. In 1898, following the resignation of Todor Zhivkov as the secretary general of the BCP and as the Chairperson of the State Council (head of state), the transition of Bulgaria from socialism to market economy began.

The Bulgarian education developed during three periods: capitalistic (1878–1944), socialist (1944–1989), and transition (after 1989), wherein each one of them is characterized by specifics which are conditioned by the particular political, social, economic, ideological, and cultural conditions. During the first period of its establishment, the educational system had a three-tier structure: primary school with a 4-year tuition course; junior high school with three grades; and high school with five grades of tuition, separated into three profiles – classical, semiclassical, and real. After 1944, the high school period was reduced by 1 year, but the structure was retained – primary school, junior high school, and high school. A new three-stage educational model was affirmed in 1979. According to this model, the general education is received in a unified school and it lasts for 10 years (grades 1–10); after that there are two 1-year grades (11 and 12) in which the students receive broad profile professional training in special educational professional complexes. After the democratic changes in 1989, the educational system also changed.

Legislative Foundations

Public Education

Public education act (1991)

The Public Education Act (1991) being the first democratic law in the recent Bulgarian history, includes the following basic aspects:

- All limitations and privileges are abolished.
- School education begins at the age of 6 or 7 by choice of the parents or the guardians.
- The parents are granted the right to choose the district and the school according to their preferences and opportunities.
- There is permission to open private kindergartens and schools both by Bulgarian and foreign individuals.
- State educational requirements are introduced.
- Simultaneously with state subsidy, the schools and kindergartens may provide their own funds.

After passing the Act, numerous changes have been made, wherein the most significant ones were between 1998 and 2002.

An act for amendment and supplementation of the public education act (1998)

This act introduces significant changes with reference to more than 30 articles of the previous act. The most significant ones are:

- School education begins at the age of 7 for pupils turning 7 in the year of starting the first grade.
- Bulgarian is regulated as the official educational language.
- Two degrees of school education are introduced – primary and secondary (7)8 + (5)4.

An act for amendment and supplementation of the public education act (2002)

This act introduces the following significant changes:

- A compulsory preschool education is introduced which may be conducted both at school or kindergarten.
- The number of the state matriculation exams is reduced to two, wherein one of them is obligatorily Bulgarian language and literature and the other is chosen by the pupil.
- The number of textbooks to be approved by the Minister of Education and Science is limited and they may not be more than three for a school subject for each school year.

An act for amendment and supplementation of the public education act (2006)

The amendments in the act are minor and they are related to the pending accession of Bulgaria to the EU. In this respect, the following provisions are made:

- Pupils, whose parents are citizens of EU member states and work in Bulgaria, may join the educational system.
- The admission of pupils after grade 7 in the state and municipal schools is done by test-format exams.

An act for amendment and supplementation of the public education act (2008)

The amendments refer to the state matriculation exams, which were introduced for the first time in 2008, as well as to the regulation of the so-called protected schools (if such a school is closed, this will prevent access to education) and central schools (a school which is located in the closest inhabited area in the territory of the municipality, where pupils from the areas without a school are educated). In particular, this means:

1. With reference to state matriculation exams:
 - State matriculation exams are assessed by a national commission, which includes school teachers and university professors.
 - The pupils who have not passed the matriculation exams receive only a certificate for completed education degree.
2. With reference to the protected schools:
 - They may not be closed and they receive additional subsidy.
 - The list of schools is being updated every year.
3. With reference to secondary schools:
 - The list of secondary schools is accepted by an act of the Council of Ministers following a proposal of the respective municipal councils.
 - Funds are provided for additional financing of the education of traveling students.

Educational degree, general educational minimum, and study curriculum act (1999)

The 1999 Act is a separate law which regulates the study-curriculum structuring and the preparation of the study content. School subjects in the general-educational preparation are grouped into eight cultural and educational spheres: (1) Bulgarian language and literature; (2) foreign languages; (3) mathematics, informatics, and information technologies; (4) social studies, civil education, and religion; (5) natural sciences and ecology; (6) arts; (7) manner of life and technologies; and (8) physical education and sport. The Act establishes the portion of the compulsory preparation (CP) and compulsory optional preparation (COP), where CP represents 90% of the compulsory classes for primary school, 80–90% for junior high school, and 45–80% for high school. The freely chosen preparation (FCP) for all grades is four classes per week.

In 2002, 2004, and 2006, the Act underwent some amendments which refer to the matriculation exams and to the ways for receiving secondary education.

Professional education and training act (1999)

The Act provides for the functioning of professional training and the implementation of the entire professional training system in Bulgaria. It complies with the European trends in this sphere even though many of them are difficult for implementation under the Bulgarian economic and financial conditions. This necessitated a number of amendments in the course of its passing. The most significant aspects of the law are:

1. introduction of four stages of professional qualification;
2. introduction of six programs (categories) for professional training;
3. establishment of state educational requirements for professional education and training; and
4. establishment of new structures for management and financing of the professional training system.

Higher Education**The higher education act (1995)**

After the political, economic, and social changes, which were implemented in Bulgaria in 1989, the higher education is regulated by the Academic Autonomy of the Higher Education Institutions Act, which was passed as early as 1990, whereby higher education institutions were granted academic freedom with reference to tuition, structure, and organization of the study process as well as with reference to scientific and research activity.

After the passing of this Act, there were a number of amendments and supplementations, referring to both the structure and organization of the study process and to financing, academic autonomy, professors, and regulation of the functions of the state in the management of the higher education institutions.

The key aspects of the Higher Education Act are as follows:

- Higher education shall be received at higher education institutions which have been accredited by the National Agency for Evaluation and Accreditation, as well as established under the terms and conditions of the Act.
- The higher education institutions are both state owned and private.
- Higher education institutions shall be universities, the specialized higher education institutions and the independent colleges, which award only a bachelor's degree.

Unlike the Public Education Act which also includes rules for application and which refers to all schools in the Republic of Bulgaria, there are no unified rules in the sphere of higher education and each higher education institution develops its own rules which comply with the requirements under the Act.

After the accession of Bulgaria to the EU, there were a number of legislative changes with a view to harmonize the

higher educational system, the focus being on student and professor mobility, introduction of a system for transferring credits, and an European diploma attachment being added to the diploma on completion of an educational qualification.

Academic degrees and titles act (1972)

The Academic Degrees and Titles Act was passed in 1972 and a number of amendments and supplements thereto were passed, particularly after 1990 – 1992, 1995, 1996, and 2000. This Act regulates the procedure for awarding academic degrees and titles, which is unified and applies to all who wish to obtain them.

The academic degrees in Bulgaria are two:

- educational and academic degree PhD and
- academic degree – DSc.

The academic titles are: assistant, senior assistant, chief assistant, associate professor, and full professor.

The academic degrees PhD and DSc, as well as the academic degrees associate professor and professor are awarded by special academic councils and approved by academic commissions. The final say is of the Supreme Attestation Commission, which is appointed by the Prime Minister. Thus, the academic degrees and titles have state value.

This makes the procedures hard and slow – it takes about a year, which hinders the academic growth of the specialists considerably. In recent years, there have been proposals for amendment of this Act but up till now (2008) these have not been implemented.

Promotion of scientific research act (2003)

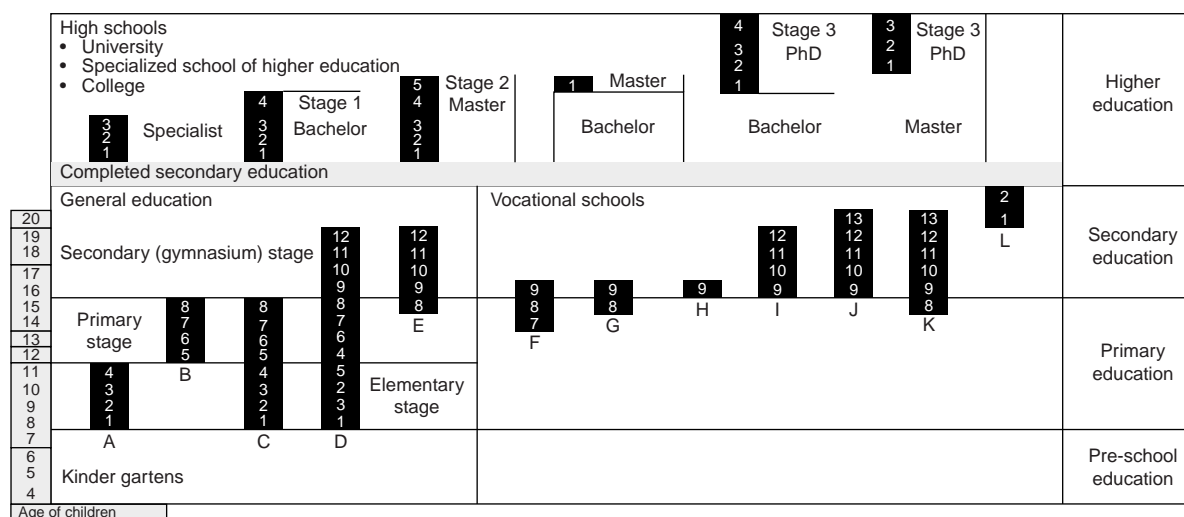
Currently, this Act has more of a theoretical scope rather than a specific practical value. It regulates the principles and mechanisms for the state policy for promotion of scientific research. There is an underlying national strategy for stimulation of the scientific research in the Act, in view of its significance for the economy, the social processes, and the human resources.

Structure of the Educational System

The educational system in Bulgaria includes the following degrees (see **Figure 1**).

Preschool Education

It is designated for children between the ages of 3 and 7 and it is implemented in kindergartens, which may be municipal and private. After amendments in the Public Education Act (2002) CP for children was introduced 1 year prior to starting first grade.



- A – Elementary (basic) school
 - B – Junior school
 - C – Primary school
 - D – Secondary school (general education)
 - E – Specialized secondary school
 - F,G,H,I – Vocational training school
 - J,K – Vocational secondary school
 - L – Vocational college
- 1 - 13 - Grades
1 - 5 - Duration of the education (years)

Figure 1 Structure of the educational system in Bulgaria.

School Education

In the Republic of Bulgaria, school education is compulsory until the age of 16.

According to the degree of education, the school education includes:

1. Primary education – with duration of 8 years (grades 1–8), which is subdivided into two stages:
 - elementary stage of the primary education (elementary education) – (grades 1–4), and
 - junior high school stage of the primary education (junior high school education) – (grades 5–8).
2. Secondary education – with duration of 4 or 5 years (grades 9–12), which is implemented into just one stage – high school. The profile (language, mathematics and sciences, and mathematics) and the professional high schools admit pupils after grade 7 (grades 7–12).

According to the contents of the preparation which are received in the two stages, the schools are:

- a. general educational schools;
 - elementary school (grades 1–4),
 - junior high school (grades 5–8),
 - primary school (grades 1–8),
 - high school (grades 9–12),
 - profile high school (grades 8–12),
 - secondary general education school (grades 1–12), and
 - secondary general education school (grades 5–12).
- b. Vocational schools;

- vocational high school (grades 7/8/9–12/12/13),
 - vocational trainings school (grades 9–12),
 - vocational college – 2 years after completion of secondary education.
- c. sports schools (grades 5/7–12);
 - d. art schools (grades 5/7–12);
 - e. special schools – for children with special educational needs – (grades 1–8/12); and
 - f. culture schools – two national schools (grades 8–12).

School education may be obtained in private schools as well, which have the same structure as the state and municipal schools.

There are also theological schools which are opened by the respective religious institutions in the country. Education there is implemented after completion of primary education and it may be equated to secular education if the state requirements for the respective stage have been observed.

Higher Education

Higher education is implemented at:

1. Universities – having the following degrees:
 - bachelor – 4 years,
 - master – 1–2 years, and
 - PhD – 3–4 years.
2. Special higher education institutions – having the same degrees.
3. Independent colleges:
 - specialist – 3 years, and
 - bachelor – 4 years

Education and Qualification of the Teachers

The education of teachers in Bulgaria is implemented according to the regulation for unified state requirements for acquisition of professional qualification as teacher (1997). According to this regulation the professional qualification teacher is obtained in higher education institutions which comply with the requirements of the Higher Education Act. This may be implemented simultaneously with the preparation for acquisition of the respective education qualification degree, as well as after graduation. The training includes a unified educational minimum for theoretical and practical preparation during full-time studies and 40% of it during part-time studies. The theoretical preparation is implemented by obligatory, optional, and elective study subjects. The minimum number of obligatory study disciplines and the minimum number of their classes are:

1. pedagogy (theory of bringing-up and didactics) – 60 h;
2. psychology (general, age, and pedagogical) – 45 h;
3. audio-visual and information technologies in tuition – 15 h; and
4. teaching methods – 60 h.

The optional study disciplines with the number of hours not less than 15 h for each one is distributed depending on the nature of the major into two groups, where the students obligatorily choose one or two study disciplines out of each of the stated groups:

- first group – pedagogical, psychological, and methodological; and
- second group – interdisciplinary and applied experimental disciplines, related to the professional pedagogical realization of the teachers.

The practical preparation is implemented under the following forms of training with a minimum number of hours as follows:

1. lesson observation – 30 h;
2. continuous pedagogical practice – 45 h; and
3. pre-graduation pedagogical practice – 75 h.

The training for obtaining a professional qualification as teacher ends with an integrated practical-applied state exam, comprising delivering and defending a lesson prepared by the student. The state exam is passed before a state-exam panel whose members are appointed by order of the head of the higher education institution.

The conditions for improvement of the qualification of the pedagogical professionals in the system of the public education and the order to obtain pedagogical professional qualification degrees is regulated by Order No. 5 (1997 and supplemented in 1999). The training for

improvement of the qualification of the pedagogical professionals is implemented within the structure of the higher education by Bulgarian higher education institutions, established under order, stipulated by the Higher Education Act and by the Information and Teacher Qualification Improvement Departments with the higher education institutions. In 2008 they were three in number.

Main forms of improvement of qualification are:

- a complex course – duration of attendance at least 3 weeks with a training program comprising current problems according to the functions of the position of the people trained;
- a thematic course – duration of up to 2 weeks with a training program comprising theoretical and practical preparation on a specific professional pedagogical issue;
- an instructive course – duration up to 1 week with a training program comprising the preparation of new study content, specific professional functions, etc.;
- professional pedagogical specialization;
- specialization in a specific scientific field;
- training for acquisition and improvement of professional pedagogical skills; and
- seminar, practice, problem group, conference, etc.

The improvement of the qualification is implemented at three levels:

- school – the head of the school or kindergarten organizes or coordinates it;
- regional – by the regional inspectorates of the Ministry of Education and Science (MES); and
- national – by the MES.

Depending on the level of their professional competence during their professional realization, the pedagogical professionals may obtain the following professional qualification degrees, which are awarded by the specialized institutes for improvement of qualification of the teachers with the universities in Sofia, Varna, and Stara Zagora:

1. fifth professional qualification degree – teachers should have at least four subsequent years of pedagogical experience and pass an oral exam according to a questionnaire with at least 4.50;
2. fourth professional qualification degree – teachers should have at least fifth professional qualification degree and should pass a written exam with at least 4.50 on an issue of the specific professional field of the candidate;
3. third professional qualification degree – teachers should have fourth professional qualification degree and should complete 1-year professional pedagogical specialization with an average score of at least 4.50;
4. second professional qualification degree – teachers should have third professional qualification degree and they should defend a written paper with at least

- 4.50, related to analysis of the results of a diagnostic procedure applied by the candidates; and
5. first professional qualification degree – teachers should have second professional qualification degree and they should defend successfully a written paper of research and innovative nature, related to the pedagogical practice of the candidate.

Management of the Educational System

Management of Public Education

The management in preschool education is implemented at four levels: national, regional, municipal, and internal.

The management in school education is implemented at three levels: national, regional, and internal.

The activities at the specified levels are implemented by different institutions, as follows:

1. *National level – the MES.*

The MES forms the national policy and coordinates the overall work for the functioning and development of the educational system by developing strategies, priorities and programs, state educational standards, and programs for qualification and specialization of the pedagogical professionals; it proposes, prepares, and confirms agreements for international cooperation. The stated activities are implemented by different departments which report to the minister of science and education.

2. *Regional level – regional education inspectorate (REI).*

The REIs are 28 in number and correspond to the administrative division of the country. The REIs are managed by heads who take the position after they win a competition and the operational activities are implemented by experts, where all of them have the status of civil servants. The structure of the REI is separated into general administration with administrative legal, financial and economic, and information services departments, and specialized administration, which makes inspections and performs organizational methodological activity.

3. *Municipal level – education department with the municipal administration.*

The activity of the education department comprises securing the obligatory school education, maintenance of the necessary equipment of the schools and kindergartens, providing funds for support, repairs, furnishing, facilities for recreation, canteen feeding, sports facilities, boarding houses, and transport for pupils and teachers.

4. *Internal level – headteacher and pedagogical council.*

The headteachers of the state and municipal kindergartens and schools are appointed after a competition, organized by the MES and by the regional inspectorate

respectively. The headteacher organizes, manages, and takes charge of the entire activity of the school or the kindergarten. He or she disposes budget and nonbudget funds on the principle of delegated budgets. The headteacher appoints the entire school staff and controls its activity, organizes the admission of children and pupils, keeps and maintains all school records, and represents the school before organizations and persons.

The pedagogical council is a special collective body for management of the kindergarten or school, which is convened by the headteacher at least once in 2 months. The pedagogical council accepts the strategy for the development of the educational institution, frames rules for the activity of the school and ensures their implementation, discusses and takes decisions about the education results.

The school board is an independent volunteer association for assistance in the development and the material support of the kindergarten or school. It is established by the initiative of the headteacher, the parents, teachers, and public figures.

Management of the Higher Education Institution

Higher education institutions are managed on two levels – national and internal.

1. *National level – the National Assembly, the Council of Ministers, the Ministry of Education and Science and the National Evaluation and Accreditation Agency.*

The National Assembly pursuant to decision, opens, restructures, renames, and closes the higher education institutions and determines the state budget subsidy for each higher education institution annually.

The Council of Ministers has broader functions related to opening, restructuring, and closing faculties, institutes, branches, and colleges in the state higher education institutions; it affirms the classification of the higher education spheres and the professional streams; the state requirements for educational qualification degrees; the number of students and postgraduates admitted annually; the list of the regulated professions, etc.

The MES is the state body for implementation of the national policy in the field of higher education; it maintains a register of the higher education institutions, organizes the acknowledgment and legalization of the diplomas of the people who have graduated abroad, and exercises control over the higher education institutions for compliance with the Higher Education Act.

2. *Internal management – management authorities of the higher education institutions are the general meeting, the academic council and the rector.*

The general meeting of higher education institutions is held once in 4 years during which it elects the rector,

the academic council, a chairperson, a deputy chairperson, and members of the control council by secret vote; it also accepts and amends the rules for the activities of the higher education institutions, as well as the annual report of the rector.

The academic council consists of 25–45 members and includes representatives of the academic staff, students, postgraduate students, and employees on a quota principle. It is within the powers of the academic council to elect the deputy rectors, to take decisions for opening, closing, and restructuring of chairs and departments, to approve the study curricula for the individual majors, to approve and control the implementation of the budget of the higher education institution, to approve a system for evaluation and maintenance of the quality of education and the quality of the academic staff.

The rector is elected for a 4-year term and represents the higher education institution; he or she enters into or terminates the employment agreements on all levels, takes final decisions on matters with reference to admission, deregistration, and movement of students, postgraduate students and specializing postgraduate students, and prepares reports about the activity of the higher education institution.

Financing of the Educational System

Financing of the Public Education System

The financing of the state and municipal kindergartens and schools is provided by the state budget, by the municipal budget, by sources under other laws and acts of the council of ministers, and by private sources. The financing is implemented in a manner provided for under the Public Education Act and the amendments and supplements thereto, as follows:

- The MES in coordination with the Ministry of Finance sets the annual subsidy for each child or pupil in the state and municipal kindergartens and schools.
- After approval of the funds in the state budget by the National Assembly and after approval of the municipal budgets, the MES relocates budget funds for investments, scientific research subsidizing, and for qualification of the teachers.
- The funds for support and salaries are stipulated separately in the municipal budgets and the MES budget. These funds are provided to the state and municipal kindergartens and schools and on the basis thereof the headteachers form their budgets.
- The MES controls the spending of funds for support and salaries both for state and municipal schools.
- The mayors control the manner for spending the funds provided to the municipal kindergartens and schools.
- The municipal budgets, including the one for education, are formed from state-budget proceeds and from their own proceeds from local charges and taxes. Therefore, the more their own proceeds they obtain, the fewer subsidies from the budget.
- The kindergartens and the schools are entitled to provide their own funds from rents, fees, donations, inheritances, produce realization, qualification and creative activity, and educational and other services.
- The ministries and the municipalities may not reduce the subsidy at the expense of the private proceeds, gathered from the abovementioned sources and activities.
- The construction, maintenance, repairs, reconstruction, and modernization of the material and technical equipment of the kindergartens and the schools are implemented with funds from the state budget or from that of the municipalities or with funds from other sources.

The delegated budgets were introduced in January 2007, which may provide greater effectiveness in the utilization of the funds, as well as differentiated payment for teachers. The principle that money follows the pupil was introduced and this applies to private schools as well.

Financing of the Higher Education System

The state higher education institutions are financed by the state budget on an annual subsidy, through financial aid from the municipalities, from donations, wills, sponsorships, inheritances, and with their own proceeds from different activities – research and scientific, arts and creative activity, healing and sports activity, and from proceeds from application and tuition fees, postgraduate qualification, administrative services, etc.

Under the Higher Education Act, financing is implemented according to the following principles:

- The state subsidy provides funds for: tuition support, scientific or creative activity, publishing study and scientific works, accommodation expenses, and for capital investments.
- The tuition support is estimated on the basis of: differentiated norms into professional streams for a single student, number of students and postgraduate students, and the results from the accreditation of the higher education institution.
- The funds for scientific studies are stipulated by norms for the tuition support.
- The council of ministers stipulates the rates for the remuneration of the people working in the higher education institution.
- The students and postgraduate students are entitled to apply for grants from the state budget.
- The students, the postgraduate students, and the specializing postgraduate students pay tuition fees, which are stipulated annually by the council of ministers.

Control and Assessment of the Educational System

School Education

In 2003, the MES published an order which determines the state educational requirement for the assessment system. Under this normative document, the assessment is done at these junctures: in the process of school education; upon completion of secondary education; upon completion of professional education; and upon completion of professional training. The assessment is both internal and external. In internal assessment, the teacher is the assessor, and in external assessment the assessors may be: the school panel, appointed by the headteacher of the school; officers or commissions, appointed by the minister of education and science, the head of the REI, or by the directors of the control and assessment units servicing the public education.

In 2005, the center for control and assessment of the quality of school education was established by the MES. It is in charge of the development of national tests, and currently there are such tests for grades 4, 5, 7, and 12. Some of them are the basis for a future external assessment at the end of the study stages – those for grades 4, 7, and 12. There is an ongoing process of standardization of school-education assessment and when the methodology of the tests becomes clear, there will be a system for independent assessment of the knowledge and skills of the students from the entire year at the end of each educational stage. Then the tests will be done by a real structure with nationwide coverage, which is external both to the ministry and the educational network; however, this has not been implemented yet. This year (2008) the tests for grade 5 were approved, next year there will be such a test for grade 6, and may be for some of the junior high school grades, and this should be the skeletal framework for the future annual internal assessment based on national criteria. This should remain a task of the school and its purpose it to standardize the outgoing assessment at the end of the year, which the school is obliged to do now as well.

Higher Education

A special state authority for assessment and accreditation of the quality of the activities of the higher education institutions is the National Agency for Assessment and Accreditation with the council of ministers. Its activities are related to:

- approval of criteria for assessment and accreditation in compliance with the state requirements;
- approval of procedures for assessment and accreditation and the documentation related to them;
- evaluation of projects for opening and restructuring of higher education institutions, basic units, branches, professional streams, and majors;

- approval of criteria and procedures for post-accreditation observation of the higher education institutions, the basic units, branches, professional streams, and majors;
- evaluation of the activity and the quality of the education in the higher education institutions on the basis of which accreditation is awarded or refused; and
- setup and maintenance of an informational system with data about the accredited higher education institutions, basic units, branches, professional streams, and majors.

Statistic Data

General Educational System

Tables 1–3 show the number of general educational schools, the number of students in the general educational schools, and the number of teachers in the general educational schools, respectively.

Special schools

Table 4 shows the number of special schools.

Professional educational system

Tables 5 and 6 show the number of professional schools, and the number of students in the professional schools, respectively.

Table 1 Number of general educational schools

Elementary schools (grades 1–4)	273
Primary schools (grades 1–8)	1760
Junior high schools (grades 5–8)	20
High schools (grades 8/9–12)	168
Secondary general educational schools (grades 1–12 and 5–12)	410
Total number	2631

Table 2 Number of students in the general educational schools

Grades 1–4	268 123
Grades 5–8	286 960
Grades 9–12	170 462
Total number	725 545

Table 3 Number of teachers in the general educational schools

Grades 1–4	16 886
Grades 5–8	25 647
Grades 9–12	15 472
Total number	58 005

Private schools

Tables 7–10 show the number of private schools, the number of students in the private schools, the number of private professional schools, and the number of students at the private professional schools, respectively.

Higher Education

Tables 11–14 show the number of state universities and higher education institutions, the number of private universities and higher education institutions, the number of

students in the universities, higher education institutions, and independent colleges, and the number of teachers in the universities, higher education institutions, and independent colleges, respectively.

Table 4 Special schools

Number of special schools	126
Number of students at the special schools	11 875
Number of teachers at the special schools	2333

Table 5 Number of professional schools

Arts schools	20
Professional high schools	453
Professional colleges	17
Professional schools	5
Total number	495

Table 6 Number of students in the professional schools

Arts schools	3984
Professional high schools	19 1649
Professional colleges	2810
Professional schools	2502
Total number	200 945

Table 7 Number of private schools

Elementary schools (grades 1–4)	4
Primary schools (grades 1–8)	22
Junior high schools (grades 5–8)	3
High schools (grades 8/9–12)	26
Secondary general educational schools (grades 1–12 and grade 5–12)	4
Total number	59

Table 8 Number of students in the private schools

Elementary schools (grades 1–4)	81
Primary schools (grades 1–8)	1912
Junior high schools (grades 5–8)	57
High schools (grades 8/9–12)	3712
Secondary general educational schools (grades 1–12 and 5–12)	282
Total number	6044

Table 9 Number of private professional schools

Arts schools	1
Professional high schools	25
Professional colleges	17
Total number	43

Table 10 Number of students at the private professional schools

Arts schools	31
Professional high schools	1509
Professional colleges	2328
Total number	3868

Table 11 Number of state universities and higher education institutions

State universities	24
State special higher education institutions	12
State independent colleges	1
Total number	37

Table 12 Number of private universities and higher education institutions

Private universities	4
Private special higher education institutions	2
Private independent colleges	8
Total number	14

Table 13 Number of students in the universities, higher education institutions, and independent colleges

Students at the state universities and higher education institutions	221 000
Students at the independent colleges	24 000
Total number	245 000

Table 14 Number of teachers in the universities, higher education institutions, and independent colleges

Number of teachers at the state universities, higher education institutions, and independent colleges	20 925
Number of teachers at the private universities, higher education institutions, and independent colleges	3008
Total number	23 933

Bibliography

Academic Degrees and Titles Act (1972). *APIS* Vol. 4(2), 260.
 An Act for Amendment and Supplementation of the Public Education Act (1998). *APIS* Vol. 4(1), 10.
 An Act for Amendment and Supplementation of the Public Education Act (2002). *APIS* Vol. 4(1), 10.
 An Act for Amendment and Supplementation of the Public Education Act (2006). *APIS* Vol. 4(1), 10.
 An Act for Amendment and Supplementation of the Public Education Act (2008). *APIS* Vol. 4(1), 10.
 Educational Degree, General Educational Minimum and Study Curriculum Act (1999). *APIS* Vol. 4(1), 5.
 Public Education Act (1991). *APIS* Vol. 4(1), 10.
 Professional Education and Training Act (1999). *APIS* Vol. 4(1), 1.
 Promotion of Scientific Research Act (2003). *APIS* Vol. 4(2), 251.
 Regulation for Unified State Requirements for Acquisition of Professional Qualification "teacher" (1997). *State gazette* 34, 1997.
 The Higher Education Act (1995). *APIS* Vol. 4(2), 270.

Further Reading

Antova, M., Atanasov, V., and Drakeva, R. (2000). Education for human rights, peace, stability, and democracy status, problems,

perspectives. *Country Reports on Education, Bulgaria*. http://www-gewi.kfunigraz.ac.at/csbsc/country_reports/Education_Bulgaria.htm (accessed July 2009).

Relevant websites

<http://www.globalgateway.org.uk> – Education Guide – Bulgaria, globalgateway.
<http://www.eurydice.org> – Eurybase: Education systems in Europe, EACEA.
<http://www.nationmaster.com> – NationMaster, Bulgarian Education Statistics.
<http://www.euroeducation.net> – Structure of Education System in Bulgaria.

England

C Whatford, UK National Commission for UNESCO, London, UK

© 2010 Elsevier Ltd. All rights reserved.

General

Education in England is compulsory for all children aged between 5 and 16 years of age. The government provides free education to all children up to the age of 19 via state primary and secondary schools and further-education colleges. Some parents, however, prefer to pay to send their children to private independent schools: these may be residential schools or pupils may live at home and attend on a daily basis. There are around 2300 independent schools in England and around 7% of the pupil population attend independent schools. Furthermore, although most parents educate their children at school, under UK law, parents are allowed to educate their children from home if they choose to do so.

The first 6 years of education are known as primary education and the remaining 5 years as secondary education. Between these ages, the subjects that children study and the way in which they study them are set out by the government according to the National Curriculum.

The National Curriculum has been established to allow all children the opportunity to learn the skills needed throughout their lives. The curriculum consists of the core subjects of English, Maths, and Science together with the foundation subjects of design and technology, information and communication technologies (ICT), history, geography, modern foreign languages, art and design, music, physical education (PE), and citizenship. It is flexible and allows teachers the freedom to set suitable learning challenges for all pupils, taking into account their needs, abilities, and interests.

The responsibility for the education service in England lies with:-

- The Department for Children, Schools and Families (DCSF), which brings together all aspects of policy affecting children and young people. The government's aim is to ensure that every child gets the best possible start in life, receiving the ongoing support and protection that they – and their families – need to allow them to fulfill their potential.
- The Department for Innovation, Universities and Skills (DIUS) which is responsible for investing in science, research, and innovation. It also takes responsibility for ensuring that UK has the skilled workforce it needs to compete in a global economy.

School Governors

All local authority (LA)-maintained schools in England must have a school-governing body. Set up by law as corporate bodies, responsibility lies with the whole governing body rather than individual members. Each school has an Instrument of government which specifies the constitution of its governing body.

Governing bodies are made up of key stakeholders:-

- *Parents.* Parent governors are elected by other parents at the school;
- *Staff, including the headteacher.* Staff, that is, teaching and other staff, governors, that is, members of the community;
- *The LA.* The LA appoints one or more local education authority (LEA) governors; and
- *Foundation governors.* For some schools, the foundation or relevant religious body appoints foundation governors.

The main responsibilities of the governing body are:

- Setting strategic direction, objectives, targets, and policies;
- Reviewing progress against the budget, plans, and targets;
- Approving the school budget;
- Acting as a critical friend to the headteacher by providing support and challenge; and
- Appointing a headteacher.

In practice, the governing body delegates much of the responsibility for the day-to-day management of the school to the headteacher, while retaining its strategic role in developing and monitoring the implementation of policy at the school.

Inspection of Schools and Higher Education Institutions

The Office for Standards in Education (Ofsted) was first set up on 1 September 1992. The new Ofsted, which came into being on 1 April 2007, is a nonministerial government department which brings together four formerly separate inspectorates. Ofsted inspects education and training for learners of all ages. Its main aim is to raise aspirations and

contribute to the long-term achievement of ambitious standards and better life chances for service users.

The Quality Assurance Agency for Higher Education (QAA) was set up in 1997 to provide an integrated quality-assurance service for UK higher education institutions. The agency is an independent body, funded by subscriptions from universities and colleges of HE, and through contracts with higher education funding bodies. Its board of directors comprises 14 members: four are appointed by the representative bodies of the heads of higher education institutions; four are appointed by the funding bodies in HE, and six are independent directors who have wide practical experience of industry, commerce, finance, or the practice of a profession.

Preprimary Education

For children aged 3 months to 3 years, provision is largely in the private and voluntary sectors, and parents pay fees. For children aged 3–5 years, places are in state-maintained nursery schools and classes in primary schools, and in voluntary and private settings.

The government has, in recent years, set targets to expand and develop publicly funded early-years education and childcare in cooperation with the private and voluntary sectors. Free part-time nursery provision is now available for all 3- and 4-year-olds. It is a statutory requirement that children at this stage of education work toward early learning goals, which cover six key curricular areas.

Primary Education

In England and Wales, the primary school normally has 7-year groups and comprises two key stages. Infant schools are for children aged 5–7, junior schools are for those aged between 7 and 11. Some primary schools have a nursery attached for children aged between 3 and 4.

Key stage 1 provides for children aged between 5 and 7 and key stage 2 provides for children aged 7–11. At the end of key stage 1, all pupils have to take classroom-based tests in English and Mathematics. The tests are marked by the class teacher, but set and audited by an outside agency.

At the end of key stage 2 (age 11), pupils take written tests in English, Mathematics, and Science. These tests are set and marked by an external agency which is appointed by the Qualifications and Curriculum Authority. National tests at the end of key stage 2 are not used to assess the ability or aptitude of pupils for the purpose of selection for secondary schools. Selective secondary schools (which exist in some areas of England only, and are commonly

known as grammar schools), set their own tests for this purpose, and these are usually administered earlier in the school year than the key stage 2 tests.

Secondary Education

Most secondary schools, which are maintained schools, are mainly nonselective and accept pupils regardless of their ability. These are known as comprehensive schools. Parents have an element of choice in choosing a secondary school and it is not uncommon, especially in towns and cities, for students to travel some distance to a school of their choice.

There has been a recent drive to designate comprehensive schools as specializing in particular areas, for example, technology, modern foreign languages, performing arts, physical education, or sport, etc. The government achieved its target to increase the number of specialist schools to 2000 by the year 2006, over a year early. The current target for the program is that 95% of all schools will be specialist by 2008.

The School Standards and Framework Act 1998 allows the admissions authority for a school with a specialism in certain prescribed subjects to give priority to up to 10% of pupils on the basis of aptitude in those subjects. This does not apply only to schools in the specialist schools program; any admissions authority may decide to give priority in this way.

There are still some schools which select all their pupils by ability; these are grammar schools.

In England, the first 3 years of secondary education, catering to pupils aged 11–14 years, are known as key stage 3 and the remaining 2 years of compulsory education, for pupils aged 14–16, as key stage 4.

Pupils are organized into year groups with a senior teacher as head of year.

Depending on the policy of the school, within the year group, pupils may be divided into classes or tutor groups with a designated teacher who has organizational and pastoral-care responsibilities. Tutor groups in some schools may include pupils from different year groups. The organization of teaching groups is also determined by the school. In secondary schools, teachers usually teach one or more specialist subjects.

Pupils may be grouped by their general ability (a practice known as streaming), taught in mixed-ability groups or, more commonly, grouped according to their ability in a particular subject (a practice known as setting). Most schools use setting for some subjects only, such as mathematics and languages, and teach other subjects in mixed-ability groups. Other schools use a combination of these teaching groups/methods. Pupils are placed in a class according to their age.

The first formal academic qualifications, which are taken by school children, are the General Certificate of Secondary Education (GCSE). Generally, these are taken by students at the end of their final year. Students are taught on average eight GCSE subjects comprising a mix of compulsory subjects, plus others which they have chosen to study. GCSEs are available in more than 40 subjects in total, including humanities (such as English literature and history), sciences (including biology and physics), social sciences (including health and social care and economics), technology (including ICT and engineering), creative subjects (including art and music), and many modern and ancient languages. GCSEs are awarded in a series of grades from A* the highest to G the lowest.

The National Curriculum and its key stages do not apply to pupils in postcompulsory schooling. The final 2 years of (postcompulsory) full-time secondary education for those aged 16–18+ (in years 12 and 13 of school education) are usually known as the sixth form. Pupils are taught by specialist subject teachers, normally in smaller groups than younger pupils. There are no regulations covering class sizes.

14–19 Education and Skills Reform

The White Paper 14–19 Education and Skills outlined plans to meet the government's ambition to ensure that post-16 participation rates rise from the current 76% to 90% by 2015. These reforms will give all young people the opportunity to choose a mix of learning which motivates, interests, and challenges them, and which gives them the knowledge, skills, and attitude they need to succeed in education, work, and life. The aim is to develop an education system where all young people have the opportunities to learn in ways that motivate and stretch them – a system where, through their own hard work and that of their teachers and tutors, young people are able to qualify themselves for success in life.

As part of its proposals to improve secondary, further, and HE, the government is introducing 14 specialized diplomas covering all sectors of the economy. From 2008, the first five diplomas will be ready for teaching.

Private Education

In general, private education is that which is provided in institutions which are largely privately funded, receiving most of their income from tuition fees, usually paid by parents. In some cases, private schools also receive some funding from donations and grants received from benefactors.

Six hundred and fifteen thousand children attend some 2500 independent schools in the United Kingdom. In

England, this represents about 7% of the school age population; it is lower in Wales.

Over 80% of these children are in 1283 schools regulated by membership of an association represented by the Independent Schools Council (ISC).

At present, under transitional arrangements, some existing pupils may have part or all of their fees (depending on parental income) paid by the government through the Assisted Places Scheme. This scheme was originally set up in 1981 under the provisions of the Education Act 1980, but was brought to an end by the Education (Schools) Act 1997.

The Music and Ballet Scheme (MBS) was also set up in 1981, to provide government grant aid with fees at seven independent specialist schools in England for children with exceptional talent in either music or dance. In 2002, the scheme was renamed the Music and Dance Scheme and another school added. The Aided Pupil Scheme, which operates within the scheme, is similar to the Assisted Places Scheme, but it also provides additional funds to cover the costs of specialist tuition and boarding fees. The grant is means tested. There are over 800 children participating in the scheme, which remains unaffected by the Education (Schools) Act 1997.

There is private provision at all levels of education. Private schools are often known as independent schools. Some long-established secondary private schools are known as public schools. Apart from city technology colleges and academies, independent schools receive no public funds (except for the financial advantages, e.g., tax relief, conferred by their charitable status). They are financed through fees, donations, and endowments.

Sixth Forms and Sixth-Form Colleges

The legal school-leaving age is 16, although many schools have provision for their pupils to stay on until they are 18. In these cases, students will most usually take 2-year courses of AS- and A-level.

A-levels are awarded at five grades, from A (the highest) to E (the lowest), and each A-level scores between 40 and 120 points on the Universities and Colleges Admissions Service (UCAS) tariff. The UCAS is the single clearinghouse for applications for admission to full-time first degree, foundation degree, and first-diploma courses at universities and other higher education institutions in the United Kingdom. A new UCAS tariff has been developed to provide a point-score system for expressing entry requirements to higher education. The new system was introduced in 2002 as the first cohort of applicants with the new post-16 qualifications (introduced in September 2000; see 5.17.2. and 7.14.) entered HE. It makes provision for a wide range of qualifications, including Scottish qualifications and the new Welsh Baccalaureate. However, use of the tariff is not obligatory.

A student who scores three grade As or more has an unlimited choice of where he/she can progress after their A-levels. The lower the score, the more limited the choice. Realistically, a student will find it difficult to be accepted for a degree course with a UCAS score of less than 180 points (equivalent to three A levels at lower grades or two at medium grades). The average score achieved by students on degree courses is around 240 points.

GCE A-level and GCE AS qualification passes are graded on a scale of A to E. The grade U denotes a fail. Pupils have a right of appeal if they are not satisfied with the grade they receive. Appeals (from the pupil's school or examination center) must be addressed in the first instance to the appropriate awarding body. Grades can be lowered, raised, or confirmed as a result of an appeal. The Examinations and Appeals Board (EAB) hears appeals which have failed to be resolved by an awarding body's own procedures.

Pupils can have access to their marked examination papers, on request. Requests can be made by schools, colleges, or examination centers or by students via their school, college, or examination center. A charge is made for this service. About 75% of UK students remain in full-time education beyond the age of 16.

Further Education

After the age of 16, students are free to remain in full-time education, to leave education and enter the workplace, or to undertake a mixture of part-time education and part-time work. Education undertaken at this stage is known as further education. It includes courses that expand on the knowledge learned at school, and courses that offer a second chance or new direction to students who have not obtained the results needed to progress to the next level (generally those who did not achieve five GCSEs at grades A* to C or equivalent). Courses may be 1 or 2 years long.

The 75% of 16-year-olds who choose to enter full-time further education undertake their studies either in the sixth-form college or at a college of further education. They choose from a wide range of education and training opportunities and, from this stage onward, most will specialize in a general direction, such as science, arts, or humanities, or train for a specific career or industry sector such as health and social care, engineering, or tourism. Education falls into two broad categories, academic and vocational.

Higher Education

After successfully completing further-education qualifications, many students progress to HE. Students usually need to be at least 18 years of age to begin a higher education course. HE includes academic courses, such

as diplomas of HE (Dip HE), bachelor's degrees, and, at the postgraduate level, master's degrees, masters of business administration (MBAs), and PhDs. Vocational higher education is covered by Higher National Certificates (HNCs) and Diplomas (HNDs) and Foundation Degrees, which are offered by HE colleges and universities, and some further-education colleges.

Students who opt for HE will usually begin their course at the age of 18 or 19. Courses can be anything from 1 to 4 years long and students may choose to stay on as postgraduates into their 20s. Adult education is increasing in popularity and many colleges and universities now have a growing number of mature students who are aged over 21. They have usually worked for a few years and now want to improve their career prospects.

Most students progress smoothly from compulsory school education to further education to HE. A government commitment in the White Paper *The Future of HE* was to increase participation in HE toward 50% of those aged 18–30 by 2010. Alongside this, the government wants to widen participation so that more people from nontraditional backgrounds have the opportunity to participate in HE. They are doing this through a number of initiatives focusing on raising the aspirations and attainment levels among groups currently underrepresented in HE; by improving the financial packages available to students and by improving progression routes into HE.

We are doing this through a number of initiatives focusing on raising the aspiration and attainment levels among groups currently underrepresented in HE; by improving the financial packages available to students and by improving progression routes into HE. Entry requirements for degree courses vary, depending on the reputation of the institution and the popularity of the course. Approximately, 180 UCAS points from A levels or the equivalent is the minimum acceptable for a course.

Bachelor's Degrees

Honors degrees form the largest group of HE qualifications. Typical courses last for 3 years (if taken full-time) and lead to a bachelor's degree with honors, having a title such as bachelor of arts (BA(Hons)) or Bachelor of Science (BSc(Hons)). Moreover, at this level are short courses and professional conversion courses, based largely on undergraduate material, and taken usually by those who are already graduates in another discipline, leading to graduate certificates or graduate diplomas.

Honors degrees are normally classified into first, second, and third class. Second-class degrees are further divided into two divisions, upper and lower, also known as 2i and 2ii. Students who do not achieve a high enough standard for an honors degree may be awarded a pass (or ordinary) degree.

Master's Degrees

Master's degrees are awarded after the completion of taught courses or programs of research, or a combination of both. Longer, research-based programs often lead to the degree of MPhil. Most master's courses last at least 1 year (if taken full time), and are taken by persons with honors degrees (or equivalent). Some master's degrees – in science and engineering – are awarded after extended undergraduate programs that last, typically, a year longer than honors degree programs. At this level, there are also advanced short courses, often forming parts of continuing professional development programs, leading to postgraduate certificates and postgraduate diplomas.

Doctorate Level

Doctorates are awarded for the creation and interpretation of knowledge, which extends the forefront of a discipline, usually through original research. The titles PhD and DPhil are commonly used for doctorates awarded on the basis of original research. Doctoral programs that may include a research component, but which have a substantial taught element, lead usually to awards that include the name of the discipline in their title (e.g., EdD for doctor of education). A doctorate normally requires the equivalent of 3 years full-time study.

Professional Qualifications

Some institutions offer professionally accredited courses in subjects such as engineering, accountancy, teacher training, librarianship/information science, and medical studies. The Chartered Institute of Library and Information Professionals (CILIP), for example, is a professional body which accredits degree courses in librarianship and information science offered by universities.

Qualifications specific to a profession and required for its practice, which may be taken at a university or after initial studies at a university are completed, are more often obtained through the successful completion of examinations set or accredited by professional bodies such as the Chartered Institute of Public Finance and Accountancy and the Inns of Court School of Law.

Finance

Undergraduate students apply for financial support to their LA, or Education and Library Board in Northern Ireland, which then carries out an income assessment. A notice is sent to the student advising how much support will be available in terms of grants and loans, and how much, if anything, the student or his/her family are expected to contribute. Some funds are administered by individual institutions.

Loans

Students entering HE can apply for an income-contingent loan which is repayable through the tax system through their LA. Loans are available to full-time students up to the age of 54. Part-time higher education students on low incomes become entitled to avail of student loans to manage the costs of their course.

Seventy-five percent of the maximum loan is available to all eligible students regardless of any other income they have. The remaining 25% depends on the student's income and that of his/her family. This is assessed by the student's LA.

The government's policy is that loans should be available to students on favorable conditions, which require borrowers to repay, in real terms, broadly the same amount as that borrowed. Thus, interest rates are indexed to inflation rates and adjusted each year in line with the retail price index. Students make payments at the rate of 9% of their income above a certain threshold.

Teacher Training

In England and Wales, teachers employed in maintained schools, including nursery schools, must have qualified teacher status (QTS) or be otherwise licensed or authorized to teach by the Secretary of State for Education and Skills, the Training and Development Agency for Schools, or the National Assembly for Wales.

Teachers of classes of pupils with hearing or visual impairments must, in addition, obtain a recognized specialist qualification within 3 years of appointment.

Persons training to teach in the postcompulsory education sector can work in further education, tertiary colleges, or sixth-form colleges. The sector also includes agricultural and horticultural colleges, and there are also opportunities to work in adult and community education and in private training organizations. Maintained schools may also have postcompulsory age students (16+) in sixth-form classes; regulations concerning teacher training and qualifications for those teaching in these schools are covered by those for the school sector.

HE institutions are autonomous corporations. Each determines its own requirements concerning staff and there is no single legislative provision in this area. Increasingly, institutions provide training for their teaching staff, especially those who are new to the profession. The admission requirements for teaching staff, curriculum content, and method, and the evaluation and certification of any professional development offered to staff, vary from one institution to another. However, national professional standards for teaching in HE are currently under development.

Estonia

M Keränen and J Kõrgesaar, Ministry of Education and Research, Tartu, Estonia

© 2010 Elsevier Ltd. All rights reserved.

General Information

The Republic of Estonia is located in northern Europe. Estonia is the northernmost of the three Baltic countries, situated on the eastern shore of the Baltic Sea and is bordered by the Gulf of Finland in the north, Russia in the east, and Latvia in the south. The official language in Estonia is Estonian which is closely related to Finnish. Along with Finnish, English, Russian, and German are also widely spoken and understood. Estonia has a population of 1 342 000 inhabitants spread over 45 000 km² making it quite a sparsely populated country in Europe. The capital of Estonia is Tallinn. Estonia is a member state of the European Union (EU) and the North Atlantic Treaty Organization (NATO).

The Estonian education system started with cathedral and monastery schools established in the thirteenth and fourteenth centuries. In 1583, the first Estonian Gymnasium was opened for a short period in Tartu by the Jesuits, followed by the gymnasiums established by King Gustavus II Adolphus of Sweden in 1630 in Tartu, and in 1631 in Tallinn. Tartu University, the first university of Estonia was founded in 1632. The Lutheran Church had an impact on Estonian culture and education from the sixteenth century onward, competing with pietist movement since the late eighteenth and, with the impact of freemasonry, in the beginning of the nineteenth century. The first peasants' schools were established in the seventeenth century, along with the 2-year college for training native teachers for peasants' schools from 1684. According to the 1881 population census, about 90% of Estonians were literate. Since the thirteenth century, Estonia has been dominated by Danes, Germans, Poles, Swedes, and Russians. Estonia was established as a modern nation-state on 24 February 1918. Since then, Estonian has been the dominant language of instruction in schools and universities. Before and after World War II, Estonia was occupied by the Soviet Union. Independence was regained in 1991 and the new constitution came into effect from 1992.

Prior to 1940, 90% of the population was Estonian. Between 1944 and 1998, intensive migration from different areas of the Soviet Union changed the structure of the population. According to population census in 2000, 68.6% of the population are ethnic Estonians (922 000) and 25.6% are ethnic Russians (345 000). This ethnic structure remains reflected in a language-based semidivided education. There are quite a few Russian-language or mixed-language

kindergartens, general education and vocational schools, and higher education institutions functioning in Russian-dominated or mixed-population areas along with prevailing Estonian-language and some English- or Finnish-language schools.

In a geopolitical sense, Estonia's change from Russia's west into Europe's east was formally adopted with the turn of millennium as Estonia became a member state of the EU and the NATO in 2004. Until 2008, Estonia's economic development has been very intensive. Estonian gross domestic product (GDP) per capita was US\$702 in 1992 and US\$20 300 in 2007.

Politics and Goals of the Education System

Estonia has done remarkably well in two major international comparative studies: Trends in International Mathematics and Science Study (TIMSS) 2003 (assessing eighth-grade students) and Programme for International Student Assessment (PISA) 2006 (assessing 15-year-olds).

Between 1940 and 1991, despite adoption of the over-politicized and formally unified Soviet educational structure and curricula, the Estonian educational system was able to maintain instruction in the Estonian language on all levels of education, to develop a relatively widely differentiated secondary education, even to secure 1 additional year of secondary education.

The political renaissance at the end of the 1980s also stimulated Estonians' sense of national identity in educational matters. This was a big moment, since, in Estonia, education has traditionally been highly valued. Estonians connected their hopes with education. A depoliticized form of child-centered humanistic education was highly valued by politicians.

The Republic of Estonia Education Act from 1992 provides legal basis for the formation, functioning, and development of the education system. This act was followed by the laws on adult education (1993), on universities (1995), on vocational education institutions, on private schools, on applied higher education (all three in 1998), and preschool education (1999).

The long-term goal of the Estonian education policy has been to raise the living standard of citizens. Education has a major role in further improving the economic competitiveness of Estonia. By the year 2007, there were policy documents for managing all areas and levels of

education, research, language, and youth. In order to increase the coherence and coordination between the fields, in 2006, the Ministry of Education and Research prepared and approved the development plan *Smart and Active People 2007–10*. The goals of all the relevant strategies have been summarized as follows:

- everyone has equal opportunities for lifelong learning following their abilities and interests;
- the quality of teaching and instruction is high;
- Estonian continues to be developing cultural language, while linguistic diversity is highly valued;
- youth work supports the development of the personality of young people; and
- research and development are of high quality and sustainable.

The Estonian education system expects to be positioned to benefit from international cooperation. Exchanging experience across national borders is a vital condition on decision-making processes. Estonia participates in a large number of international cooperation activities in EU, Organization for Economic Cooperation and Development (OECD), United Nations Educational Scientific and Cultural Organization (UNESCO), Council of Europe, and Nordic Council of Ministers.

The Formal System of Education

There is a 9-year compulsory basic school, which replaced the compulsory secondary school in 1988. There are 3–4-year professional higher education institutions introduced along with universities. There are private, municipal, and public educational institutions. The formal structure of education is shown in **Figure 1**. Education has the following levels: preschool education, basic education (first level of education), secondary education (second level of education), and higher education (third level of education).

Primary, Secondary, and Tertiary Education

Children who turn 7 years of age by first of October of the current year are obligated to attend school. The compulsory schooling obligation applies to children until they acquire basic education or turn 17 years of age. Basic education can be acquired in primary schools (grades 1–6), basic schools (grades 1–9), or predominantly in comprehensive schools with grades 1–12.

Basic school is divided into three stages of study:

- stage I: grades 1–3;
- stage II: grades 4–6; and
- stage III: grades 7–9.

In the 2007–08 academic year, there were 121 405 pupils acquiring basic education. It can be acquired on the basis of three national curricula: the national basic school and upper secondary school curriculum, the simplified national basic school curriculum, and the national curriculum for students with moderate and severe learning disabilities.

There are two main options after the basic school: upper-secondary general school (grades 10–12) and vocational school. The certificate confirming secondary education gives students the right to continue their education either in universities or in other tertiary educational institutions.

Attendance of state schools and general-education schools of local governments is free of charge. There were approximately 155 000 pupils in general-education schools in the 2007–08 academic year, 12 606 of them were in the first grade. There were 589 general-education schools (primary schools, basic schools, secondary schools or upper-secondary schools, and adult upper-secondary schools) in the 2007–08 academic year. The number dropped by 12 when compared to 2006–07. Local governments have established catchment areas for all municipal schools. Schools must admit all children who live in their catchment area obligated to attend school. There are also schools which recruit their students from all over the town area or even all over country. There are only a limited number of seats and so candidates have to pass entrance tests to get into such schools.

The provision of basic education on all levels of education is carried out on the basis of common curricula irrespective of the language of instruction, that is, any language may be the language of instruction. In the upper-secondary school stage, from the year 2011, 60% of the curriculum should be taught in Estonian in Russian-language schools. Any other language may become the language of instruction in the upper-secondary stage of municipal schools and in specified classes of municipal school, by the permission by the Government of the Republic.

Schools elaborate their curricula on the basis of the national curricula.

General and upper secondary education are provided pursuant to the national curriculum which is used by each school as the basis of their own curriculum. There are 35 study weeks in daytime study. One study week must include at least 32 lessons. Pupils in upper-secondary schools can acquire extensive knowledge in certain fields of study within the framework of elective subjects (in arts, science, natural sciences, etc.) or learn a profession taught in vocational schools. At the end of the 3-year study period, pupils take five graduation examinations, of which at least three are state examinations. In the 2007–08 academic year, 33 666 were studying in upper-secondary schools. Approximately 6300 pupils are acquiring basic and general secondary education in the form of evening study and

Age study period

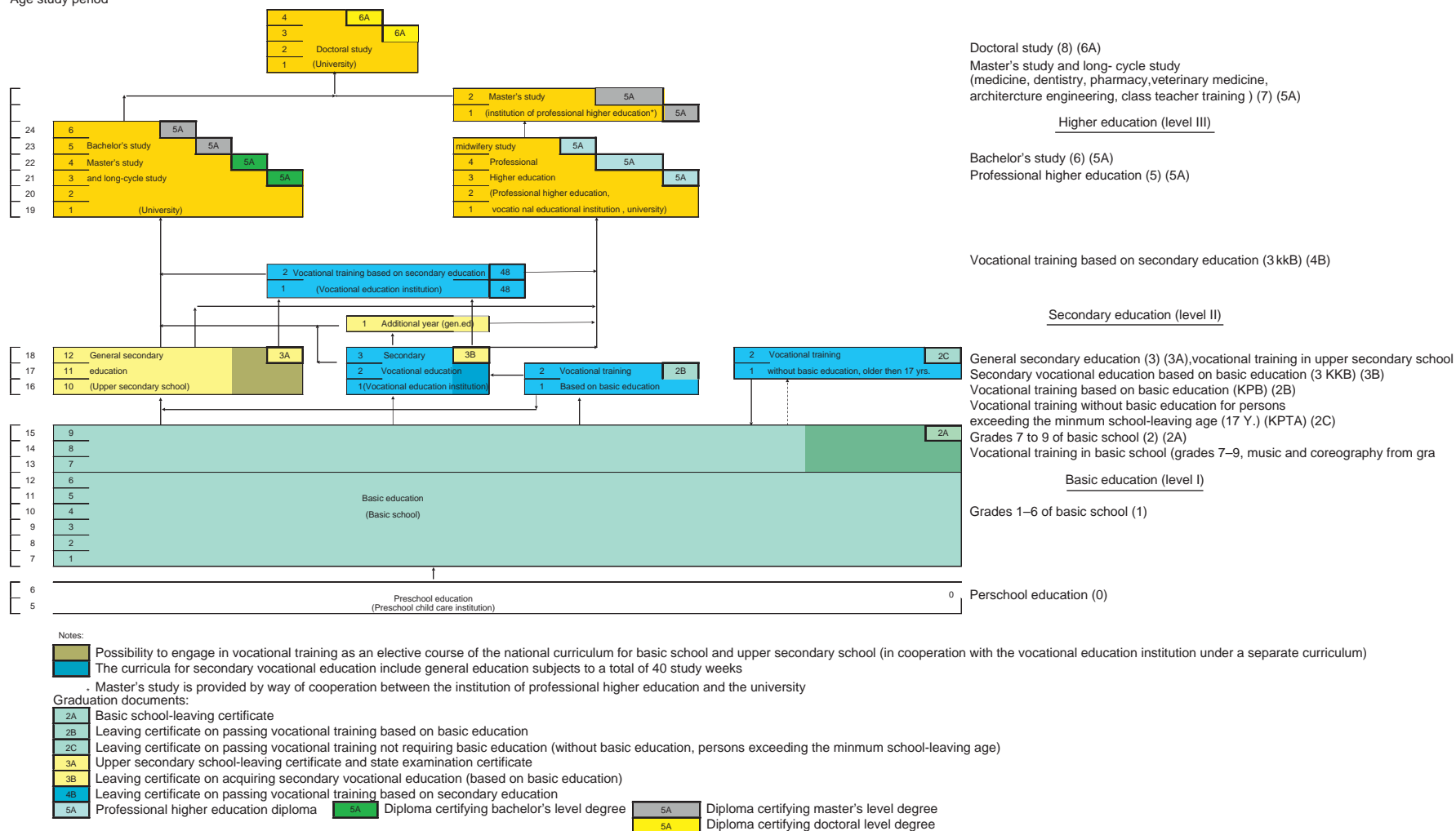


Figure 1 Structure of formal education and graduation documents in Estonia in 2006.

distance learning, and 5000 of them are acquiring general secondary education.

In vocational education, providing specialist knowledge and skills, one can acquire secondary vocational education after graduating from basic school or upper secondary school. There is an option to take up vocational studies without the requirements of basic education, in which case only professional skills are targeted. In the 2007–08 academic year, there were 47 different vocational training institutions enrolling 27 000 students.

Acquisition of general secondary education and secondary vocational education is free, while higher education can be either free of charge or tuition-fee based.

Estonia has a two-tier system of higher education. Universities are institutions of research, development, education, and culture; one of their duties is to offer research-based doctoral studies. Institutions of professional higher education prepare motivated specialists with good professional skills on the first level of higher education, meeting the needs of the labor market. Thirty-four educational institutions offer higher education in Estonia, as of 30 June 2008. There were six universities as legal persons in public law, four private universities, 10 state higher schools, 11 private higher schools. In addition three vocational education institutions provide applied higher education in a limited number of programs.

Students can choose between two types of curricula at the first level of higher education:

- theory-based bachelor-study curricula where professional skills are developed on the basis of theoretical principles; and
- practice-based professional higher education curricula where theoretical knowledge is developed primarily on the basis of practical needs. The curriculum contains at least 30% practical work.

The standard period of bachelor and master's study is at least 5 years and the study load specified in the curriculum is 200 credit points (300 credit European credit transfer system).

At the beginning of academic year 2007–08, 68 168 students were acquiring higher education. The number of students studying according to higher education curricula has increased 2.7 times in the 2007–08 academic year, when compared to 1994–95, when the number of students was 25 483.

Forty-six percent of the students enrolled obtain higher education at state-commissioned and 54% at nonstate-commissioned study places in the academic year 2007–08.

Sixty-two percent of students are women. The fields of study of national priority and with increasing employment are engineering, production and processing, information technology sciences, biosciences and environmental protection (environmental technology), physical natural sciences, and healthcare.

Preschool Education

According to the Education Act and the national curriculum of preschool child-care institutions, preschool education is acquired either in a preschool child-care institution or at home and thus, the acquisition of preschool education is the responsibility of the child's parents or guardians. Parents of children who go to child-care institutions or who stay at home have the right to receive advice about educating and raising their children from the teacher of the child-care institution in their region.

There are four types of preschool child care institutions – day nurseries (for children 1–3 years of age), nursery schools (for children 1–7 years of age), special nursery schools, and nursery primary schools. Most child-care institutions have the first three groups and such institutions are mainly located in rural areas. Child-care institutions with up to 12 groups dominate in cities. Local governments must provide all their resident children between 1 and 7 years of age the opportunity to attend child-care institutions in their catchment areas, if this is requested by the parents. Still, the need for nursery school places is very high in certain regions such as the capital city of Tallinn and its surroundings and university town Tartu.

Rural municipality or city governments provide children with physical, speech, sensory, or mental disabilities, and children who need special assistance or special care, the opportunity to grow and develop in the adaptation groups of the child-care institutions in their places of residence.

Rural municipality or city governments create special groups or establish special nursery schools if it is impossible for local child-care institutions to create an adaptation group. Children are admitted to adaptation or special groups on the basis of a written application of a parent and the decision of the counseling committee.

Preschool child-care institutions or schools have opened preparatory groups for children who do not attend nursery schools, and children can participate in these groups free of cost.

Special Education

Children with special needs have the right to learn in schools of their residence like all other children.

Pupils with special educational needs are pupils whose outstanding talent, learning or behavioral difficulties, health problems, disabilities or long-term absence from studies creates the need to make changes or adaptations in the content of studies, the study processes, or the learning environment (study aids, classrooms, language of communication, including alternative communications, specially trained teachers, support staff, etc.), or in the work plan prepared by the teacher for work with the relevant class.

The task of schools is to involve pupils in the study process and adapt the learning environment in such a manner that every pupil would be able to learn according to their abilities and to develop their abilities to the maximum.

Disabled children or children who need special care have the right to study in the nearest school that complies with requirements if the school of their residence does not have the possibilities and conditions for accommodating children with special needs.

Pupils in Estonian schools can study on the basis of different curricula. A simplified national basic school curriculum, which is used to teach children with slight mental retardation, exists in addition to the national basic school and upper-secondary school curricula. There is also the national curriculum for students with moderate and severe learning disabilities.

Adult and Nonformal Education

Adult education is divided into formal education, work-related training, and popular adult education.

Flexible study opportunities have been created for adult students to participate in formal education: distance learning and evening courses, learning as an external student, or in part-time study. It is possible to acquire basic and secondary education, to study in a vocational educational institution or higher education institution.

Popular adult-education institutions and several other training institutions offer nonformal training. Popular training courses, which primarily focus on the hobbies and interests of people, offer the opportunity to develop one's creativity, talents, and social skills.

There are 7% of learners in lifelong learning among people aged 25–64 in Estonia in 2007 according to the data of the Statistical Office. The goal set by the EU is to achieve a level where at least 12.5% of the working-age population participate in lifelong learning by 2010. Estonia plans to reach 10% by 2008.

Administrative and Supervisory Structure and Operation

The parliament establishes the principles, budget, and general structure of the educational system. The government is responsible for the establishment and maintenance of the state programs of education and regulates the governance of universities and other higher education institutions. The Ministry of Education and Research is responsible for the development and implementation of state educational programs and standards. It also grants licenses to private educational institutions.

Local municipalities establish and finance local municipal educational institutions. Local self-governments form

the structural units to administer preschool, primary, secondary, and vocational educational institutions. Determining of the national curricula for public educational institutions is in the jurisdiction of the government or the Ministry of Education and Research. Each school designs its own curricula on the basis of the national curricula.

Educational Finance

In recent years, 14–15% of the total public expenditure in Estonia has been invested in education each year, and in 2007 this amount was over 11 billion kroons. Expenditures on general education comprise 56% of the total public expenditure on education (in the second-half of the 1990s, expenditure on general education comprised 62–64% of the total public expenditure on education). The share of pupils on the general-education level has also decreased in the same period. The state pays for ongoing teacher training, some reconstruction and building costs, nationwide student fairs, hobby schools, and adult education. Local governments pay for the salaries of preschool teachers. Instruction at school is free. At basic school, the teaching materials are free. Meals are subsidized.

Vocation education expenditures now comprise 12% of total expenditures. Higher education expenditures amounted to 18% in 2007. The number of students has more than doubled when compared to the middle of the 1990s, but this has mainly occurred on account of students studying in nonstate-commissioned student places. Thirteen percent of the public expenditures on education are classified under other education costs – this covers costs that cannot be tied to any concrete level of education, including education projects that combine several levels of education, the expenditures in the areas of language and youth, and preschool education.

Education expenditures comprised approximately 4.7% of GDP in Estonia in 2007.

The EU has been supporting the areas of education, research, and development in Estonia through the European Social Fund (ESF) and the European Regional Development Fund (ERDF) since 2004.

Approximately 1.1 billion kroons of assistance was directed in the area of responsibility of the Ministry of Education and Research from 2002 to 2006. The largest share of the total costs was directed at vocational education.

Supplying Personnel for the Education System

The number of preschool teachers was 7111 in the beginning of 2007. The average age of teachers is 45 and the share of male teachers is 0.4%. The number of teachers who worked in general-education schools in the academic

year 2007–08 was 15 039; 1.073 of these worked in state schools and adult upper-secondary schools. The average age of teachers is slightly over 45. Fourteen percent of all teachers are men; 2214 teachers work in vocational educational institutions; 34% of them are men and 66% are women. The average age of vocational teachers has increased to 47.9 years. Approximately 2500 teachers work in hobby schools.

Teachers' standard of professional competence was approved in 2005. Teacher training is provided in universities and their colleges. The duration of the training program for class and subject teachers is 5 years, and basic school teachers acquire the competence for teaching several subjects within the framework of this training. Initial training of preschool teachers is primarily based on professional higher education and bachelor study curricula.

Since 2004, all teacher-training graduates need to undergo an induction year where they are supported by their colleagues/mentors and where they can also take part in the support program offered by institutions of higher education.

Teachers must also undergo at least 160 h of in-service training in 5 years. Professional in-service teacher training takes place in the form of independent work or in state and municipal institutions, private schools with training licenses, or by legal persons or sole traders governed by private or public law if their activities comply with the taught subject or specialty.

Curriculum Development and Teaching Methodology

Curricula, having been developed during the Soviet occupation, had to follow the general principles of the Soviet school ideology. Despite the strict demand to unify the curricula, Estonia was able to maintain some of its educational traditions. For example, instruction was carried out in the native language and a number of textbooks were written by Estonian authors. These are the potential factors, why the results of Estonian students in 2006 PISA comparative study on education turned out to be very good (fifth on natural sciences). The other reason may lie in the fact, that Congress of Estonian Teachers held on 1987 started the necessary process of restructuring the system. The curriculum was depoliticized, many syllabi modernized, factual material diminished, the time allocated to optional subjects and to foreign languages was increased, and new subjects were introduced. In 1996, the Estonian government approved the new national curriculum for basic and secondary education. The overarching vision was forward-looking, but not all subject syllabi managed to embrace changes. The revised state curriculum was adopted in 2002. It provides more freedom for schools in

choosing the direction of studies, makes it possible to consider the students' interest and abilities, coordinates the development of the system of external evaluation, decreases the study load, leading to the implementation of knowledge into practice, and emphasizes the importance of the integration of different subjects and the importance of competencies, as the basis of lifelong learning.

English, French, German, and Russian are taught as primary and secondary foreign languages, the choice of which is up to the school (taking into account the school's possibilities and students wishes). Students have to study at least two foreign languages, the choice of which is up to them. In grade 3 the first foreign language is introduced and in grade 6, the second foreign language.

The curriculum is nationwide for the core subjects but allows choices for the school and students in some subjects. According to the 2007–08 timetable, the number of classroom hours per week was 20 h in grade 1, 23 h in grade 2, 25 h in grades 3 and 4, 30 h in grades 6 and 7, 32 h in grade 8, and 34 h in grade 9. The length of the school year is 175 school days.

The System of Examination, Promotion, and Certification

The evaluation of students' achievement, the transfer of students from grade to grade, and graduation from school are administrated by the National Examination and Qualification Centre. The students' achievement is numerically evaluated on a five-point scale. In basic schools, verbal evaluation is also used. The reports given at the end of each grade at basic and secondary school levels monitor the student's progress. The first formal certificate is received at the end of basic school (grade 9). The next certificate is received at the end of secondary school.

The state exam in the mother tongue in the form of an essay is obligatory for all students graduating from secondary school; in schools where instruction is in a language other than Estonian, an exam in Estonian as a second language is also obligatory.

The mathematics state exam is obligatory to all. The rest of the state exams can be chosen from Russian language and literature; English; German; French; Russian; biology, chemistry, and physics; and geography, history, and civics.

The evaluation commission for state exams evaluate the state exams on a 100-point scale.

Since 2002, the passing minimum for graduation has been 20 points. Secondary education certificate does not confer any automatic right of admission to tertiary educational institutions or universities. The applicant must pass the examinations of the higher education institution in compliance with their admission requirements.

Educational Assessment, Evaluation, and Research

Estonian Education Information System (EHIS) is a web-based national register, which provides information about schools. All the information regarding the Estonian education system is entered into this online state register. System includes data about educational institutions, students, teachers, study modules, as well as education licenses and educational certificates. It was launched in 2005 and now there are also subregisters for kindergartens, juvenile committees, and hobby schools. System is obligatory for education institutions. Every school-leader nominates 1–2 responsible school administrators to update school data in the register. All the governmental statistics and financing decisions are made according to register data. This register is a tool not only for schools, but also for local governments and other state institutions.

In 2006, the compulsory internal assessment of preprimary child-care institutions, general-education schools, and vocational-educational institutions was introduced and altered the system of state supervision. The changes served to support the development of internal quality-assurance system in educational institutions. This is in line with the general-educational political trends toward promoting the independence and self-management of educational institutions. State consulting will be introduced to ensure the effective implementation of internal assessment.

The Research and Development Organisation Act provides the structure and functioning of Estonian research and development (R&D) system. The Government of the Republic establishes national R&D plans, submits them to the Parliament, obtains the approval for the programs, ensures the cooperation between the ministries, and enacts legislation. Research and Development Council provides consultation to the Government of the Republic on the matters of R&D. The Ministry of Education and Research implements national research policy, organizes the financing and evaluation of the activities of R&D institutions, and coordinates international research cooperation on the national level. The Ministry is also responsible for the planning, coordination, execution, and monitoring of the research policy related to the activities of universities and research institutes.

Estonia has been cooperating with the (OECD) since the beginning of the 1990s. In autumn 2006, Estonia joined the International Conference on Intelligent Engineering Systems (INES) program. On 16 May 2007, Estonia received an invitation to accession negotiations with the organization.

Current Topics

The impact of a drastic decrease in the number of children for the education system, labor market, and also for

the state is discussed by the education community and general public. There were approximately 155 000 pupils in the daytime form of study in general-education schools in the 2007–08 academic year: in 2004–05 there were 184 038 pupils. Closing municipal educational institutions particularly in peripheral areas has made many parents decide to send their children at an earlier age to schools at larger centers. Such a decrease in the number of students according to school stages has a direct influence on the school network, state-commissioned teacher training, the need for in-service training and retraining of teachers, the whole higher education system and also on the labor market.

Developing efficient financing models, which would ensure the quality and accessibility of education, is an urgent issue at all levels of education. Much progress has been made with renewing the financing system of general education-schools, which was approved by the government coalition in 2007. The principles agreed upon foresee replacing the per-capita financing model with a per-class model. The aim of the proposal is to retain small rural schools and to reform school network in order to guarantee basic schools and upper-secondary schools the implementation of the agreed availability and quality criteria and to offer assuredness to parents and teachers as regards the development of local schools. Financing of vocational training has been a national priority in recent years. Bearing in mind the relative stagnation of vocational training and the notable support of employers' central organizations, the goal related to vocational training is reaching the price of a study place, 1.5 times higher than the price of a study place in general-education schools in 2008. With the support of European structural funds, investments in the vocational training system infrastructure have also been increased in the recent years, the study support system has been implemented, and the government has approved the new curriculum-group coefficients, which, compared to earlier, take better account of the price of studies in different professional spheres. The financing-system reforms are also prepared in higher education where, according to the approved higher education strategy, the 3-year result-based contracts will be implemented latest by the year 2009.

Increasing dropout rates in basic school and the decreasing educational level of younger age groups sets challenges for decision makers. The ratio of early school-leavers in the 18–24-year age group was 14.3% in 2007 (the respective ratio being 14.2% in 2000). Public debates focus on the need for strengthening social support structures more than in previous times and for this purpose, a boarding school program has been launched, and eligibility for free school lunch has been broadened to basic school students and to students acquiring vocational education after basic school. The curriculum development process is of principal importance in this context.

In the academic year 2007–08, a transfer of municipal upper-secondary schools with Russian as the language of instruction to partial subject study in Estonian has begun. The goal of the transfer is to ensure that students whose mother tongue is other than Estonian get equal opportunities to attain higher education and scope in the labor market. The goal is to achieve a situation where at the upper-secondary level, at least 60% of compulsory courses are taught in Estonian.

In the context of a changing labor-market environment, it is necessary to bring enterprises' needs and the structure of higher education curricula quickly into compliance. At all levels of education, raising young people's interest in natural and exact sciences and technology is being debated. As measures to amend the situation, the following have been proposed: supporting respective hobby groups already at basic-school level, making digital study aids available to schools, and offering various e-study materials. Launching of special scholarship programs, enhancing the role of employers in rendering respective professions more popular, and better organization of practical training are also deemed important.

Further Reading

Eurybase – The Information Database on Education Systems in Europe.
The Education System in Estonia (2007/2008).

- Mikk, J., Veisson, M., and Luik, P. (eds.) (2008). *Reforms and Innovations in Estonian Education*. Peter Lang: Frankfurt.
- Ministry of Education and Research (2006). *Smart and Active People 2007–2010 (development plan)*. Tartu: Ministry of Education and Research.
- Ministry of Education and Research (2008). *Education and Research in Estonia*. Tartu: Ministry of Education and Research.
- Ministry of Education and Research (2008). *National Report of Estonia by the Ministry of Education and Research (for UNESCO)*. Tartu: Ministry of Education and Research.
- Ministry of Education and Research (2008). *PISA 2006, Estonian Results*. Tartu: Ministry of Education and Research.
- OECD (2007). *Reviews of Tertiary Education – Estonia*. Paris: OECD.

Relevant Websites

- <http://www.archimedes.ee> – Archimedes Foundation.
- <http://www.artun.ee> – Estonian Academy of Arts.
- <http://www.akadeemia.ee> – Estonian Academy of Sciences.
- <http://www.hm.ee> – Estonian Ministry of Education and Research.
- <http://www.emu.ee> – Estonian University of Life Sciences.
- <http://eacea.ec.europa.eu> – Eurybase: The Information Database on Education Systems in Europe.
- <http://www.innove.ee> – INNOVE (Foundation for Lifelong Learning Development).
- <http://www.oecd.org> – OECD Thematic Review of Tertiary Education.
- <http://www.smartestonia.ee> – SmartEstonia.ee.
- <http://www.stat.ee> – Statistical Office of Estonia.
- <http://www.tlu.ee> – Tallinn University.
- <http://www.ut.ee> – Tartu University.
- <http://www.ttu.ee> – Tallinn University of Technology.
- <http://www.ekk.edu.ee> – The National Examinations and Qualification Centre.
- <http://www.estonica.ee> – Web-Encyclopedia Estonica.

Fiji

I Senibulu, Ministry of Education, Suva, Fiji

© 2010 Elsevier Ltd. All rights reserved.

Background

Fiji lies in the heart of the Pacific Ocean; its exclusive economic zone (EEZ) covering about 1.3 million square kilometers, contains nearly 330 islands of which a third are inhabited. Fiji was ceded to Great Britain, in 1874, and remained a crown colony until it achieved independence, in 1970. When Europeans first arrived in the country, they began to claim large tracts of land in order to engage in plantation agriculture, initially growing cotton and later sugarcane. However, the British colonial government moved to prohibit any further alienation of native land by foreigners. As a result, at least 83% of the country's land still belongs to the Fijian population, being communally – rather than individually – owned.

With the need for manual labor in the sugar industry in its early years, the first group of Indian indentured workers was brought into the country, in 1879, under the then-Governor Sir Arthur Gordon, and their recruitment under indenture continued until 1916 – by which time, ~60 000 had arrived in Fiji. On completion of 10 years of contractual labor, there was an option to return to India at the expense of the colonial government. Approximately, two-thirds of the workers decided to make Fiji their home.

Political

A former British colony, it became a republic, in 1987, following two *coup d'etats*. Fiji was – until the December 2006 military coup – a sovereign democratic state. Fiji's Parliament prior to this, largely followed the procedures and customs of the British where there was an elected House of Representatives and a nominated senate. The senate complemented the work of the House of Representatives through the scrutiny and revision of Bills coming to it.

Fiji's Pacific interests focus on membership of the South Pacific Forum where it was a founding member – on bilateral relations – membership of the Melanesian Spearhead Group, and those with other regional agencies. Fiji has always strongly opposed Pacific nuclear tests and the use of driftnet fishing. While supporting the promotion of Small Island States (SIDS) sustainable development, it has continued to draw attention to the special environment threats relating to the small islands of the Pacific.

On the international front, the increased effectiveness of the roles of Fiji's 11 missions abroad will continue to be reinforced, to align their roles more toward trade promotion and securing market access for its export products.

The government has continued to strengthen the embassies and high commissions in Asia to maximize benefits – particularly in enhancing trade and economic relations, especially with China and India. Since 1978, Fiji has played a prominent role, and will continue to maintain a high priority in international peacekeeping by the United Nations (UN).

Early Schools

It is well documented that, by 1874 – when Fiji became a crown colony – there was a fairly well-established network of small village schools, most of them under the control of the Methodist church offering a 4-year program in which Fijian teachers taught basic numeracy and literacy using materials printed in the dominant *Bauan* (a Fijian) dialect. There were also a few centrally located Roman Catholic schools run by Marist missionaries offering a more academic education with European teachers and English as the medium of instruction. During the first decades of colonial rule, the British administration's approach to education for the indigenous people was one of nonintervention, although they did establish and support schools for European children in Suva and Levuka. Fijian aspirations for education were such that, by 1900, there was a school in almost every village and general acceptance of school attendance as a norm of Fijian childhood.

Population

Fiji is becoming increasingly urbanized as internal migration to urban areas continues. Extension of urban boundaries has contributed to this trend. In 1996, nearly 46% of the population were living in urban areas – up from 39% in 1986. At least 41% of Fijians and Rotumans now live in urban areas. Figures for the year 2006 showed that the capital Suva is the most populated with 1 67 975 persons, followed by Lautoka with 43 274 and Nadi at 30 884.

Language and Culture

While English is the official language, Fijian and Hindi are taught as part of the curriculum. Fijians and Indo-Fijians have their own dialects while the latter, generally, speak a distinctive Fiji–Hindi dialect; people of other nationalities who have made Fiji their home, include Europeans, Chinese, and other Pacific islanders.

Economy

Across 4 years, up to the end of 2006, Fiji experienced strong and continuous economic growth and, for 2004, it was estimated at 4.1%. Tourism then, continued to be the main foreign-exchange earner, followed by garments, sugar, fish, gold, and mineral water. Economic growth, for 2005, was projected at 1.4% and, for 2006, 0.8%. A significant proportion of Fiji's population continues to be involved in the natural resource sector. The sugar industry – on which ~200 000 people are dependent – has been the principal export for the past 100 years; it continues to occupy a dominant position in the economy.

The distribution of economic power in the country comparing the size of the major sectors shows that the primary-resource-based sector – encompassing agriculture, fisheries, forestry, and mining – has declined in its share of gross domestic product (GDP) since 1986, from 25% to 18%. It has been surpassed by wholesale and retail trade, restaurants, and the hotel sector. Economic growth, for 2005, was around 1.4% and, for 2006, 0.8%.

Goals of the Education System

The core function of the ministry is to ensure that all students – from kindergarten, compulsory school age group to Form 7, and including vocational education students – have access to quality education. The ministry is also charged with the responsibility of ensuring that standards in education are met and maintained, and that the human, physical, and financial resources allocated to education are appropriately directed and expended.

The ministry had set its plan for 2006–14 to target resources and support to allow all students to achieve full benefit from educational services. For social justice and equity, all Fiji students must have the opportunity to achieve their full personal best, and the opportunity to make a positive contribution to society.

Structure/Operation of Education System

Over the last decade, the government's efforts in partnership with the community have focused on increasing access to quality education for all. Significant progress has been made in making education accessible, especially to the rural and urban disadvantaged.

Nearly one-quarter of the entire population is of school-going age and, although education is not compulsory at any stage, over 98% of children between 6 and 14 years of age are attending school. In 1996, ~70% of the population had already received some secondary education (74% Fijians and 66% Indians), and nearly all children 6–8 years of age were enrolled in schools. In the same year, the literacy rate was 92.9% and was higher

among Fijians (92.9%) than among Indians (89%). The rate was higher among males (94%) than among females (91%).

In the year 2007, Fiji had a population of 8 35 230 of which 36 190 children were 4–5 years old (of preschool age) – 1 35 724 children were of primary school age (between 6 and 13 years of age), and 78 196 in the secondary school age group (between 14 and 18 years of age).

Fiji is guided by the goals of the South Pacific Forum's Basic Education Plan which, in turn, re-affirms the support for the global initiative Education for All. All major examinations have been localized and are conducted by the Ministry of Education. The government is planning to set up a semi-autonomous Fiji Schools Examinations & Assessment Board in the Ministry of Education to enhance transparency and improve on the quality of assessment and certification procedures.

The Education Summit document produced, in 2006, for the 10-year-period 2006–15 contains the platform for education reform in Fiji and looks at six themes that have been chosen as the most encompassing key educational aspects, with the child at the center of the whole education process; it is also the key theme. Included, as the other five themes, are the curriculum, the school, technical and vocational education and training (TVET), human resources, and the community.

The recently established Teacher Registration Board will monitor teacher quality and professionalism, and the Ministry of Education will closely monitor the development of teachers to improve discipline and performance. Recently, the teacher–pupil ratio has improved slightly, declining from 1.31, in 1992, to 1.27, in 2003, for primary education, and declining from 1.20, in 1992, to 1.17, in 2003, for secondary education; the teacher–pupil ratio in both primary and secondary education is higher for urban than rural schools.

Early-Childhood Education

There are 899 preschools in Fiji and all of them are non-governmental. Most of the early kindergartens were established in the late 1940s to early 1950s by qualified expatriate wives, initially for their own expatriate children. While many of the Fiji communities have shown initiative in setting up and maintaining some form of early-childhood education at the village or community level, there needs to be a consistent effort to ensure the sustainability of such programs. It is evident that preschool centers exist more in the urban than in the rural areas. Some centers offer awareness and support programs for parents and families.

Primary Education

In 2007, there were 720 primary schools of which, only two were government run; the others were run by a committee

of private schools. Primary schools offer education for classes 1–6 or 1–8. In the 6th year, students appear for the Fiji Intermediate Examination. A new test – the Fiji Islands Literacy & Numeracy Assessment (FILNA) – is, at present, offered to more than 50% of schools in Fiji at classes 4 and 6 in place of the Fiji Intermediate Examination.

Over 160 years of development have ensured that primary schooling is well established throughout Fiji. It is well documented that, by 1874, when Fiji became a British colony, there was a fairly well-established network of small village schools – most of them under the Methodist church offering a 4-year program in which Fijian teachers taught basic numeracy and literacy using materials printed in the dominant *Bauan* dialect. There were also a few centrally located Roman Catholic schools run by Marist missionaries, offering a more academic education with European teachers and English as the medium of instruction. During the first decades of colonial rule, the British administration's approach to education for the indigenous people was one of nonintervention, although they did establish and support schools for European children in Suva (capital) and Levuka. The introduction of the indentured labor from India to Fiji, between 1879 and 1916, saw – at the turn of the century – the establishment by the missions of some schools for Indian students, with more being provided by Indian religious organizations, laying the foundation for strong Indo-Fijian community provision of – and involvement in – the education of young people.

In the decades following 1969, education initiatives in the primary sector responded to the concerns about quality and relevance. Much curriculum revision was undertaken, in the 1970s, with the objectives of improving quality and standardizing the primary system, as well as making the system relevant to an independent Fiji. The recommendation to reduce the primary school years from 8–6 – thus moving from an 8-year primary/4-year secondary structure to a 6-year primary/6-year secondary one – was implemented in the early 1970s; although not made compulsory, many schools restructured accordingly.

Only two of the 731 primary schools are state owned, at present, with most run and managed by school committees or religious and cultural groups.

Secondary Education

Up to 2000, all areas of curriculum have been modified to better fit the local context, and all external examinations had, by then, been localized. There, then, were 169 secondary schools as compared to 57 in 1969; 12 of these were government run; and 55 were boarding schools. A total of 69 230 students attended secondary school in 2007, of which 6409 were boarders. Secondary schools have Forms 1–6 (or 7) or Forms 3–6 (or 7), junior secondary schools have Forms 1–4. In 2007, there were 11 junior secondary schools. At the end of Form 4, students sit for the Fiji Junior Certificate

(FJC) examination which assesses 2 years of work in Forms 3 and 4. At the end of Form 6, students sit for the Fiji School Leaving Certificate (FSLC Examination – which assesses work on Forms 5 and 6. Following the successful completion of Form 6, students can either continue in secondary school in Form 7 or pursue Foundation level at a university. At the end of Form 7, students sit for the Fiji Seventh Form Examination.

The Ministry's National Substance Abuse Advisory Council focuses on raising awareness on the effects of misuse and abuse of substances. Value education inculcates, in students, stronger feelings of national consciousness and respect for others as well as an appreciation of Fiji's rich multicultural heritage.

Other Schooling Arrangements

There are 17 special schools which look after the special needs of children. All these schools are nongovernmental and are run by the communities.

The 1996 census recorded nearly 12 000 people as being disabled, some suffering from more than one disability, having important implications for care, service, administration, and policy concerning the disabled. Special education delivery commenced in Fiji in the early 1960s. As awareness of disability grew, a handful of nongovernmental organizations were formed. Special schools were established, in the 1970s and 1980s, to take care of children with a variety of disabilities. There are, at present, 17 special education schools catering to over 1200 disabled children at various centers; there is a rehabilitation center for older students. All the schools are managed by societies and they operate in partnership with the government.

In the existing special education schools, there is provision for teaching and learning in the areas of early intervention from birth to 8 years. Services are also provided for those beyond these years right up to vocational training. Primary education is provided for children with visual, hearing, intellectual, and physical impairment in these special schools. The integration of students into regular primary and secondary schools and tertiary institutions is one of the more recent initiatives in the effort to embrace an inclusive educational practice.

At present, the approved establishment of civil servant positions is inadequate to meet the demands placed on the special education school, in terms of the needs of children who require greater teacher-time than children in the mainstream classes.

The Curriculum Development Unit continues in its efforts to provide a well-balanced, effective, and applicable curriculum in the formal education system. One of its major tasks, to date, is the revision of the whole school curriculum under the National Curriculum Framework initiative.

Technical and Further Education

Technical and vocational education is offered in secondary schools in the form of technical/vocational subjects. It is part of the secondary curriculum. Some secondary schools also offer vocational programs. Vocational education is also offered in some primary schools as enterprise education.

The institutions outside the formal school system, including the University of the South Pacific, University of Fiji, Fiji School of Medicine, Fiji College of Agriculture, the Fiji Institute of Technology, and the Training and Productivity Authority of Fiji, have continued to progress with growth in infrastructure developments, student enrollment numbers, diversification of programs, and a marked increase in the introduction of new delivery modes. With continued varied requirements in technical and further education, the government keeps abreast and ensures the provision of academic and practical courses to develop and equip Fijian youth in a rapidly changing society; scholarships are based on merit.

Higher Education

The cabinet recently endorsed the establishment of an independent advisory body, to be called the Higher Education Advisory Commission, whose role will be to regulate the higher education sector in Fiji. The present lack of regulation in higher education is a concern because there is no control over standards and quality, as these may severely compromise its national and international standing. The Commission will manage the accreditation and approval processes. Higher education is offered at technical institutes offering certificate, diploma, and degree programs. There are two universities – the University of the South Pacific, jointly owned by the Pacific Country Forum members, and the University of Fiji, established in 2004 and based in Lautoka. There are also two main vocational training centers.

Adult and Nonformal Education

Nonformal education (NFE) has existed in Fiji, in various forms, for almost as long as its formal counterpart. NFE has included programs in areas such as primary healthcare, literacy, vocational skills, farming, recreation, and income-generating projects. While the provision of NFE has been – and still is – the domain of nongovernmental organizations (NGOs), government ministries have played a part in NFE programs, such as through the Women's Interest Office and extension services of agriculture and fisheries. While NFE is for all communities, it is, often, associated with the rural population, as it is often located in rural areas and frequently caters to out-of-school youths; the target group is, however, broad and inclusive.

The Teaching Profession

Fiji recognizes the centrality of the teacher's role in improving the quality of education offered in schools in developing countries such as Fiji. Up to 2000, remuneration for teachers absorbed approximately 90% and 85% of the budget for primary and secondary schools, respectively. While the core function in primary education in Fiji is to provide, facilitate, and promote excellence in the teaching, learning, administration, and management of primary education, secondary education's core function is to provide optimal staffing resources for all secondary schools, and to promote a committed and competent workforce that will enable the ministry to maximize the use of its budgeted resources. The government operates two of the five teacher-training institutions.

Administrative and Supervisory Structure

There are four Education Divisions and each division is managed by a Divisional Education Officer; a division consists of one or more Education Districts. There are nine education districts, each administered by a District Senior Education Officer.

Educational Finance

The education budget for 2007 amounted to \$F320 678 400 – an increase of about 1% compared to that for the previous year. The funding of education in Fiji is a joint enterprise among the government, NGOs, and local communities. The initial funding of education was largely undertaken by NGOs, particularly religious organizations. The colonial government did not provide any financial assistance for the education of the community until about the beginning of the twentieth century, although it established schools in Levuka and Suva for European children. Since independence, the government has substantially increased public funding for education in all levels of the system. Schools are allowed to charge levies for various school activities such as admission and administration and administration, hiring of books, purchase of stationery and school equipment, extracurricular activities, and parent-teacher levy. School committees manage and control the finances of their schools.

The funding of education at the primary and secondary levels represents a more active partnership among the government, NGOs, and local communities. A formula for government funding of aided schools, based on school rolls, is being used at present. The government provides grants to assist in the construction and repair costs of facilities. The ministry meets the cost of teachers' salaries according to a formula in all registered schools, with exception in the case of a few schools.

The government continues to support fee-free primary education for classes 1–8 and tuition-free secondary assistance for Forms 1–6. Several new policy initiatives were introduced, since 2002, and these include the payment of premiums required by the Native Land Trust Board for new leases to nongovernment-owned schools on native leased land. Boarding school assistance is being extended to all rural boarding schools as well as secondary boarding schools. The normal fees for the three public examinations in secondary schools have all been abolished.

Performance Monitoring, Evaluation, and Research

The performance review and standards monitoring policy for schools was recently approved by the government and was formulated with assistance from AusAID; Phase 1 implementation has begun in 2008.

Standards monitoring in schools is an important element of the process of continuous improvement and critical to achieving the purpose of the Ministry. This policy is to provide a framework that will enhance and facilitate continuous improvement of student achievement through quality teaching and learning. This is enhanced and supported by the recognition of the rights of the child and a child-centered approach to growth and development.

Major Changes and Issues Since 1970

As Fiji approached independence in 1970, there was a clear need to assess the education system and plan for its future. A Royal Commission was established in 1968 and its Sherlock Report (1969) made 81 recommendations in three major areas: the quality and number of teachers, disparities between geographical and racial groups, and the curriculum. While not all of its recommendations were implemented, it was significant as the first substantive and critical review of the entire education system.

In 1970, Fiji had a very uneven situation of educational provision. As most schools were owned and run by communities and civil society, the poor rural communities had schools that were, often, small and ill-equipped. Communities with more resources were able to put in more for their schools. There was an overwhelming move, in the 1990s, for more technical and vocational subjects to be included, rather than maintaining the academic focus. The distinguishing partnership between the state and the wider community evolved out of historical circumstances and social demand – not conscious choice. The school inspection system was terminated, in the 1970s, and the Ministry's Curriculum Development Unit staff visits schools in an advisory capacity.

The government introduced fee-free education in 1973 and, by 1981, the first 8 years of school were tuition free and this was subsequently extended to 10 years.

From 2000, tuition-free assistance had been extended to Form-5 students and children from low-income families could claim fees remission in the remaining secondary school years. In 1997, the education ministry decided to revert back to the 8-year primary cycle with the change being linked to the compulsory education initiative as well as access and equity. As primary schools are much more widespread, children's chances of completing at least 8 years of schooling are optimized. The political crisis of mid-2000 had a profound effect on the education system.

International Programs

Fiji is part of the global community and, by virtue of this link, it has international responsibilities. Fiji has made commitments to international bodies such as the UN, AusAID, and EU. At the regional level, the country holds close ties with organizations, such as the Pacific Islands Forum, through bilateral and multilateral agreements. Fiji is also part of the EU/NZAID Pacific Regional Initiatives for the Delivery of Basic Education (PRIDE). It has commitments with United Nations Educational, Scientific and Cultural Organization (UNESCO)'s Education for All (EFA) and the Fiji Education Sector Programme (AusAID and EU). Of importance also is the government's commitment to the UN Millennium Declaration and the achievement of the eight Millennium Development Goals (MDGs). Fiji's new education plan features the commitment by government to the Decade of Education for Sustainable Development (DESD) 2005–14.

Fiji and Papua New Guinea (PNG) are the two Pacific countries among nine other African, Caribbean and Pacific (ACP) countries where the International Labour Organization, in cooperation with the EU, has launched a major project aimed at tackling child labor through education. The overall goal of the partnership is to accelerate the fight against poverty and work toward the achievement of the MDGs.

The Future

Fiji's education system is confronted by new and complex issues. Educating the Fijian child of the twenty-first century becomes a more challenging task than ever before. Social problems and tensions in schools placed added pressure on the system as a whole – challenging the curriculum, the teachers, the community, the school environment, culture, and ethos. Another serious issue is the continued rise in unemployment and school dropouts, challenging the relevance of the curriculum and programs in schools. As a result, there is a demand for relevance from all segments of society and the nation as a whole.

In conclusion, there will always be the need for the education system to continue to prioritize the child, and that the educational programs should be structured to

accommodate the needs of the child. Greater attention must continue to be directed to inculcating values and aesthetics. For a peaceful and prosperous Fiji, the country will need a generation of citizens who have been educated in a holistic manner.

Ministry of Education Annual Report (2006).
Ministry of Education Annual Report (2007).
Monitoring standards in schools.
Report of the Fiji Islands Education Commission/Panel; Learning Together: Directions for Education in the Fiji Islands.

Further Reading

Building a strategic direction for education in Fiji; 2006–2015.
Effective implementation of inclusive education policy in Fiji & Rotuma.
Fiji Today; 2005–2006.

Finland

J E Tuovinen, University of the Sunshine Coast, Maroochydore DC, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Licenciate – A 2-year post-master's, predoctoral degree which may be taken as complete set of studies on its own, or may be extended to lead to a doctoral qualification.

Lukio – An upper secondary school.

Matriculation – The national final school examination and accreditation process.

Polytechnic – A postschool professional training institution.

Sámi – An ethnic minority group, distributed throughout the northern regions of Finland, Sweden, Norway, and Russia.

General Background – National Context of the Education System

Finland is a country of 5.3 million people living in an area of 338 000 square kilometers between Sweden and Russia. In a land of severe winters and mild summers, the Finns and the Sámi people have carved a living from a land of dense forests, multitudinous lakes, and islands. Finland has a relatively low average population density of 17.4 people/km². The south of the country is more densely populated, with densities up to 218 people/km² (Uusimaa), whereas in northern Lapland, the population drops to 2.0 people/km². Finland's population mostly consists of Finnish-(91.24%) and Swedish-speaking (5.46%) Finns, with the Sámi-(0.03%) and Russian-speaking (0.85%) groups being the largest minorities (2007 figures).

Finland is a democracy, with a president and a national parliament. She achieved her independence from Russia in 1917 and, although she suffered badly in wars both before and during World War II, she has maintained her independence and democratic government since then. Her economy has been transformed from an agricultural base to a leading knowledge-based economy.

Both adults and school children value education very highly. Thus, strong community support and high expectations for education are evident in Finland. This was already demonstrated in 1852, while Finland was an autonomous region of Russia, by the establishment of the first professorial chair in education at the University of Helsinki. The community support has enabled Finland to reform its

education system, most recently in three phases from 1960s to late 1980s, during the recession in the early 1990s and in a continuing phase since then.

In the early 1960s, the educational reform concentrated on changing the two-track system of school education to a comprehensive common schooling. In the two-track system, the students at the end of grade 4 in primary (elementary) school elected to continue to either an academic stream consisting of eight further grades leading to matriculation and university entrance or a civic stream of three to five grades, which led to employment or vocational schools. The reform goal was for Finland to become internationally and economically competitive by producing a better-educated population and improving educational equity.

All students now participate in a common compulsory comprehensive schooling for the first 9 years. After this period, they can choose between upper general schooling with nongraded, modular curriculum, typically completed in 3 years (84% complete it in 3 years, less than 1% in 2 years, and 16% take 4 years to complete it), or a 3-year basic vocational school, or not to continue formal school education any further. Although it is possible to combine the vocational and general tracks, very few do so.

At the end of the upper general schooling, the students may sit for their matriculation examinations, the only high-stakes national testing conducted in Finnish schools. Based on these results, they can then take up higher education studies. After the education reform in the 1960s through the 1990s, some vocational secondary school students can also enter universities. A few places are reserved for them. However, they usually enter the polytechnics if they need further education to handle the increasingly knowledge-intensive trades and professions.

Aims of the Education System

The Finnish education system aims to promote quality, efficiency, equity, and internationalism. It is intended as a means to achieve and improve the competitiveness of the Finnish society, while maintaining its strong welfare and equity focus. Finland's constitution and laws guarantee equal opportunity for every resident of Finland to get comprehensive school education irrespective of their financial standing or location in the country. Thus, education is free throughout the country, from primary to higher education.

Structure and Operation of the System

Characteristics and Basic Statistics for Levels and Types of Education and Training

The Finnish education system consists of 1 year of an optional free preprimary education, which students can enter at the age of 6 years. It is followed by a compulsory 9-year common basic education phase for students aged 7–16 years. The common 9-year stage of schooling does not include streaming or other forms of separation of students to different ability or gender groups. This is followed by two streams of further school education, one general and the other vocational, with a possibility of combined general and vocation track, leading to university or vocational, polytechnic studies. The intention of the Finnish education system is for students to be able to continue their studies as far as they wish, irrespective of the pathways they choose along the way. Therefore, even though the upper general secondary school is the main pathway to the university, students can also get to the university through the vocational education institutes.

The 20 universities and 30 polytechnics are located throughout the country, allowing easy access to free post-school education. Traditionally, university education has provided a pathway to ever more sophisticated levels via

bachelor's degrees, master's degrees, licenciate, and doctoral level studies. The master's degree, not the bachelor's, is the typical initial university exit point for students in Finland. The polytechnics award polytechnic bachelor's degrees and polytechnic master's degrees. The structure of the program is shown in **Figure 1**.

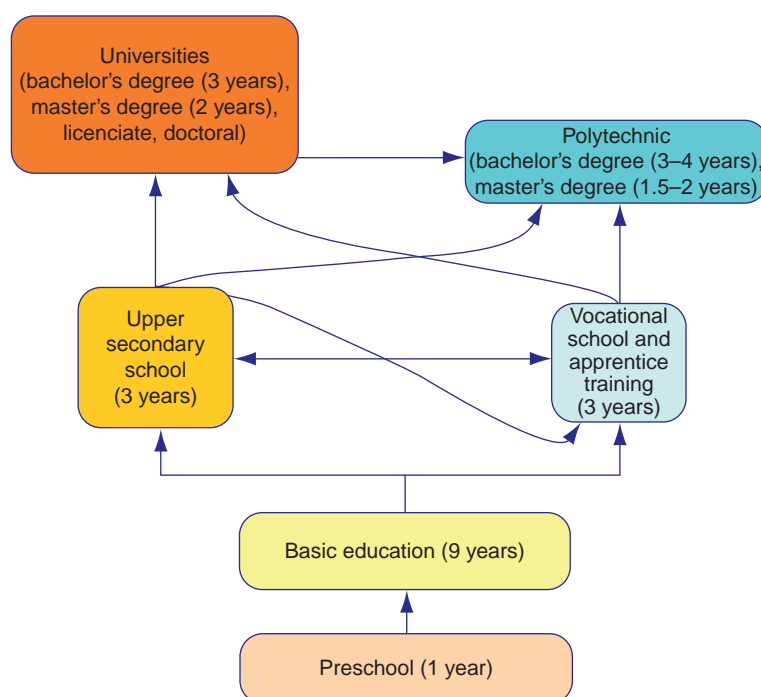
The numbers of schools, polytechnics, universities, and students are given in **Table 1**.

The school class sizes are smaller than the Organisation for Economic Co-operation and Development (OECD) average.

Preschool and Early Childhood Care

The first kindergartens were established in the nineteenth century, funded by donations, for example, the profits from the state-owned alcohol company. In 1927, a state subsidy for kindergartens was introduced. In 1973, state subsidies were granted to all forms of day care, including day-care centers, provision of family day care, and play activities.

The objective of preschool education is to support children's growth toward civilized human beings who are ethically responsible members of society and to provide



Note: Specialist vocational qualifications, further vocational qualifications, and contributing work experience are not shown for clarity.

Figure 1 Schematic structure of the Finnish education system. Aadapted from the Education System Chart on the Finnish National Board of Education Website.

Table 1 Educational Institutions and Students^a

<i>Type of educational institution</i>	<i>Number</i>	<i>Students</i>
Comprehensive schools	3067	547 500
Comprehensive school level special education schools	159	8300
Upper secondary general schools	406	122 600
Comprehensive and upper secondary general level schools	37	24 400
Vocational institutes	161	164 900
Special needs vocational institutes	12	4900
Specialized vocational institutes	37	28 900
Vocational adult education centers	32	42 700
Fire, police, and security service institutes	2	1400
Military vocational institutes	15	–
Polytechnics	30	143 300
Universities	20	176 300
Military academies	1	–
Music schools and colleges	89	62 600
Sports institutes	14	1600
Folk high schools	89	14 500
Adult education centers	234	506 900
Study circle centers	11	53 800
Summer universities	20	32 500
Other educational institutions	7	600
Total	4443	1 937 700

^a2007 figures, based on Statistics Finland information. Note that 49% of the pupils in the comprehensive school were girls and 51% were boys.

them with the basic knowledge and skills necessary in life. Preschool education aims to improve children's learning capabilities and to even out the individual differences in children's readiness to start school, to prevent growth in undesirable differences. The preschool curriculum is based on language and interaction, mathematics, ethics and philosophy, environmental and natural studies, health, physical and motor development, and art and culture.

Approximately 98–99% of 6-year-old children participate in preschool education (2007–08 school year estimate). It consists of at least 700 h per year, that is, about 18 h per week. However, there are no detailed national specifications of the school year, the number of school days, the length of school days, or the duration of periods of preschool teaching; instead, the practical arrangements of preschool education are decided by each local authority. The education providers adopt a curriculum for the education, prepared separately for instruction in the Finnish, Swedish, and Sámi languages and, where necessary, for instruction in other languages. Based on the curriculum, the education providers need to devise their own annual educational programs.

The students are grouped on the basis of achieving the curriculum objectives. During the preschool education, the necessary schoolbooks, learning materials, work equipment, and materials are free for children. Children are also provided free daily meals.

Basic Education

The students normally study their 9-year basic education in comprehensive government schools in their localities. There are also a few private schools. The regular comprehensive schools (about 3226 catering for 555 800 students in 2007) are funded by the local councils, with subsidies from the national government. The basic education is free to all students. The school sizes vary from less than ten students to schools with over 900 students. During the first 6 years, they are taught by classroom teachers for most of their subjects; however, in the last 3 years, they are generally taught by subject specialist teachers. Nevertheless, schools are increasingly deploying the subject specialist and generalist teachers across the 9 years of the basic education in order to create better continuity across the primary–secondary divide. Similarly, primary and secondary school phases are often integrated at the same physical sites and co-located buildings. On satisfactory completion of basic education, the students are entitled to enter the upper secondary school. Almost all (99.7%) of children in Finland complete their basic education.

Upper Secondary School

There are about 443 upper secondary schools, either established as separate schools or co-located with basic schools but functioning as distinct operating entities, with about 147 000 students. The typical length of upper secondary schooling (*lukio*) is 3 years, but may vary from 2 to 4 years. The age range of students is typically 16–19 years. Some of these schools may specialize in music, physical education, fine arts, languages, or natural science. The matriculation certificate entitles a student to enter university. The main aims of upper secondary education are to assist in students' moral development, in order that they become balanced and educated individuals and members of society, and to provide them with skills and knowledge needed in further studies, employment, recreational activities, and overall personal development. This stage of education is intended to develop them into life-long learners who will undertake self-development throughout their lives.

The studies in the upper secondary school are divided into courses of about 38 lessons each. The school year is usually divided into five or six stages, where the timetable focuses on particular subjects in each of these stages. The students' subject choices determine their progress and their study groupings. There are no year classes and the classes are nongraded. The progress throughout this stage of schooling is individually determined. Overall, the upper secondary syllabus consists of a minimum of 75 courses,

which are composed of compulsory, specialized, and applied courses approved by the Finnish National Board of Education.

Vocational Schools and Apprenticeship Training

The vocational schools or institutes provide 3-year studies after basic education. There are 161 vocational institutes with about 164 900 students (2007 figures). These studies authorized by the Ministry of Education and based on the national core curriculum from the Finnish National Board of Education lead to either higher education in polytechnics or universities or directly to employment. The key aims are vocational competence and entrepreneurial capability, as well as meeting the same general aims as the general upper secondary schools.

Students can select from 52 qualifications and 116 study programs. The core subjects in all courses are mother tongue, second national language, a foreign language, mathematics, physics, chemistry, social studies, entrepreneurship, workplace studies (a minimum of 6 months of workplace training outside the school is included in all vocational programs), physical and health education, and arts and cultural studies.

These qualifications may be obtained through school-based education or an apprenticeship training where an employment contract is set up between the student and an employer and confirmed by an education provider. The qualifications can also be gained by competency testing.

Polytechnic Education

There are 30 polytechnics with 143 300 students (2007 figures). They provide education in humanities and education; cultural studies; natural sciences; tourism; catering and domestic services; social services; health and sports; technology, communication and transport; social sciences; and business and administration. The students usually enter polytechnics after completing either a general or vocational secondary school, or with relevant work experience, for which they gain credit. They can complete their degrees in 3.5–4.5 years of full-time study. The polytechnic degrees have a professional emphasis and include a 6-month job placement.

University Education

The 20 universities cater to 176 300 students (2007 figures). Ten are multidisciplinary universities, three are technical universities, three specialize in economics and business administration, and four are art academies.

Approximately one-third of the possible national age group studies at a university. The universities provide courses in 20 fields of study, with the largest student numbers being in science and technology, humanities, and natural sciences.

The first level of study is a 3-year bachelor's degree followed by 2-year master's programs. Beyond the master's degree, further studies are provided at the licentiate and doctoral levels. The licentiate degree is a 2-year predoctoral degree which may be taken as complete set of studies on its own, or may be extended to lead to a doctoral qualification by about 2 more years of study which includes independent research culminating in a thesis incorporating a formal public defense. Licentiate degrees are accepted as sufficient qualification to practice medicine, veterinary medicine, and dentistry; however, many continue their studies to the doctoral level.

Adult and Community Education

In addition to the university and polytechnic education providers, various forms of adult education are provided in Finland by about 750 other institutions. These institutions are operated by the government, local authorities, collaborative municipal groups, private associations, foundations, and companies. Some programs are free, whereas others charge fees to cover the costs not wholly borne by the national government or other levels of public administration. Liberal adult education is provided in 258 adult education centers, 91 folk high schools, 11 study centers, 20 summer universities, and 14 physical education centers. General adult education is provided in 54 upper secondary schools for adults and vocational training at the upper secondary level is provided in 220 initial vocational education provider sites, 54 specialized vocational institutions, and 45 vocational adult education centers. The participants in these programs may study for qualifications or partial qualifications and train to prepare for competency-based examinations. A significant portion of the adult education is intended to upgrade existing competencies.

Teaching Profession

The Finnish teachers' high quality is based on thorough teacher preparation. This is a distinctive feature of the Finnish education system. A master's degree is required for permanent employment. They participate in school management, curriculum design and interpretation, resource materials selection, etc., in addition to classroom teaching. Studies in these areas during their extensive preservice education period prepare them for these wide-ranging responsibilities.

The Finnish school students receive special assistance with learning difficulties. Thorough teacher preparation through explicit teacher-education programs develops the teachers' capabilities to assist struggling students. The teacher-education programs have integral components helping them to identify and deal with learning problems. For example, in the basic preparatory program to become a primary classroom teacher, significant studies on dealing with learning difficulties are included. These studies are based on the premise that 15–20% of the children will experience developmentally significant learning problems during their schooling and, in fact, 21% of the Finnish basic school students accessed learning support services and 6.7% of the students were assigned to special needs education programs in 2003–04.

The high esteem of education in Finland means that teachers have high community regard. Thus, teaching as a profession attracts the best students in Finland as it is as highly valued as other high-status professions, such as medicine and information and communications technology. It is quite difficult to enter a teacher-education course in Finland. For example, only the top 10–15 % of the applicants are typically selected for primary school teacher-education programs. They are highly motivated and multitalented, with successful progression in academic disciplines as well as in art, music, and sport. The interview-based selection of students into university teacher-education courses takes into account more than their academic performance. This creates a highly capable teaching force that is strongly committed to their profession and finds a high degree of satisfaction and community support. However, as in many other countries, Finland finds it most difficult to attract the best students into mathematics and science teacher-education programs.

The teachers identify with the upper social strata in the community and are mostly politically conservative. For example, in the Finnish teacher-education politics, very few teachers hold radical or extreme Left positions. Instead, they have consistently strived for professionalism, seeking university-level teacher education for primary teachers since 1890. This has been achieved in stages; for example, in 1930s, the first University College of Education was formed at Jyväskylä and, since then, others were developed. All teacher education became a university-level provision as part of the national 1973–79 *Teacher Education Reform*.

Administrative and Supervisory Structures

The Finnish education system is governed by national acts of parliament and consequent regulations. Among other things, the regulations cover the core subjects taught to everyone and the number of hours of instruction. The core subjects comprise mother tongue and literature,

in either Finnish or Swedish, the second official language, one foreign language, environmental studies, health education, religion or ethics, history, social studies, mathematics, physics, chemistry, biology, geography, physical education, music, art and crafts, and home economics. Guidance counseling is also stipulated and some optional subjects may also be provided based on local decisions. The general curriculum is mandated at the national level; however, the details of its implementation are left to the municipalities, schools, and teachers. The national curriculum also defines cross-curricular themes, which integrate the various facets of the school programs. The core curriculum also specifies how special needs education, student welfare, student evaluation, and school–home collaboration should be conducted, but the implementation is decentralized.

Educational Financing

The primary and secondary schools in Finland are predominantly owned and operated by the municipalities. However, they receive substantial support from the national government in lumpsums not tied to particular funding requirements, allowing the schools and the municipalities to allocate their education budgets as they wish.

Education in Finland is free for the students and includes many forms of assistance. During the compulsory education phase, the textbooks and school transport and other facilities are also free, but not in the upper secondary or vocational schools. Students with learning problems can access a number of legally mandated support services. For example, schools have counselors (assisted by school nurses, psychologists, and doctors) to help in diagnosing learning problems; remedial teachers to work with individuals or small groups who are having difficulties; or class teachers who are paid for extra hours of work outside the normal class times with students who are having learning problems. The publicly funded school lunches at schools constitute an important factor in Finnish school resourcing, ensuring that, irrespective of the home environments, all students are fed relatively well at school with good cooked food.

Special needs assistants are employed to help the comprehensive school teachers if their students need extra care, for example, when special education students are integrated into general classes. Many of the school assistants proceed to teacher training programs after personal experience in schools.

Similarly, dental and health services are provided free to Finnish students. Thus, good physical health and dental basis for education is not out of the reach of any students because of cost, as the physical impediments to education receive early and continual attention and help. However, these service costs are causing community concern and some communities find it difficult to supply even the minimum standard provisions.

The cost of educational institutions in Finland as a share of the gross domestic product (GDP) is 6.0% (2002 figures). The cost of education as a share of total public funding is 12.7%. The expenditure per student in the lower basic school is US\$5090, in the upper level of the comprehensive school is US\$8200, in upper secondary school US\$6460, and in higher education US\$7330 (excluding research and development costs).

Performance Monitoring

Student testing in Finnish schools consists almost wholly of school-based assessments, which are mainly diagnostic in nature and lead to interventions to assist students with learning as described above. The main form of external testing is the matriculation examination at the end of the upper general education (*lukio*). About half of each age group participates in these examinations.

The student assessment in Finland is intended to guide and encourage student learning and help to develop students' self-assessment capabilities. The students' schoolwork, their total learning process, and conduct are assessed. The students are provided end-of-year reports and they may receive intermediate reports throughout the year. In the first 7 years of the basic compulsory school, the reports may be verbal or numerical, but later the assessments must be expressed in numerical terms. However, they may be accompanied by verbal assessment comments. Elective subjects have greater potential of assessment variability.

The scales used in the reports range from 4 to 10; 4 = fail, 5 = adequate, 6 = moderate, 7 = satisfactory, 8 = good, 9 = very good, and 10 = excellent knowledge and skills. The subject assessments are provided by the subject teachers and conduct and schoolwork assessment is jointly marked by all the teachers of a given student. The national curriculum provides criteria for the good (8) grades for all subjects. These are the fundamental tools used by the teachers to determine student grades.

The secondary school students may sit for the final national matriculation examinations in up to three consecutive stages. These examinations are held twice a year, in spring and autumn. Everyone must sit for a minimum of four examinations, depending on the students' pattern of study, but students may choose to attempt more examinations. Every student must sit for the mother tongue examination and a selection of three tests selected from the second national language, one foreign language, mathematics, sciences, and humanities.

Evaluation and research

In Finland, the education and training providers primarily evaluate their own activities. Evaluation is conducted

to provide the basis for educational policy decisions and information and performance management. The evaluation of education at local, regional, and national levels is coordinated by the Finnish Education Evaluation Council. School evaluations are primarily conducted through a regular sampling program; however, schools not scheduled to be included in a sample may also ask to be included. Finland also participates in international educational evaluations, for example, in the Programme for International Student Assessment (PISA) of international comparisons conducted by the OECD, and International Association in the Evaluation of Educational Achievement (IEA) studies, for example, in civic education (International Civic and Citizenship Study (ICCS)) and mathematics and science (Trends in International Mathematics and Science Study (TIMSS)). University and polytechnic evaluation is carried out by the Finnish Higher Education Evaluation Council. General, vocational, and adult education evaluation is the responsibility of the Educational Evaluation Council.

Educational research in Finland is mainly carried out by universities and, to a much lesser extent, by the polytechnics, or as part of national priority programs, through agencies such as the National Strategy for Education, Training and Research in the Information Society. Specialized research centers in educational research have been established at some universities; for example, at the University of Jyväskylä, there is the Institute for Educational Research, which has conducted a number of the PISA studies in Finland, and at the University of Helsinki, there is the Centre for Educational Assessment, which has conducted numerous school evaluations as well as the 2006 PISA study in Finland. The national surveys of educational outcomes in various subjects are mostly conducted by the National Board of Education.

Major Changes and Issues since the 1990s

The reform of the 1960s, discussed earlier, continued throughout the 1970s, 1980s, and the 1990s. The 1990s phase of the process coincided with a severe recession in Finland in the early 1990s, resulting in huge unemployment. However, Finland persisted in the reform of education and the recession gave added impetus toward reducing educational expenditure, reducing centralized control, and increasing regional educational autonomy. Thus, a number of functions specified centrally for the whole country were handed over to the municipalities who owned the schools and operated them. Instead of the previous systems of centrally specified detailed syllabuses, inspectorial visits, and selection of textbooks, the new system was based on trust that the local municipality in collaboration with school principals, teachers, and

the community could best decide on how to implement a much briefer form of centrally mandated syllabus and ensure the funds provided by the municipality and the National Education Department were used to the best effect.

This has led to an education system with nationally agreed direction, but implemented at the local level, with significant amount of professional discretion vested in teachers, schools, and local communities. The results have been very pleasing, with excellent student academic performance demonstrated in rigorous international comparisons and a remarkable level of equity in school performance across the country.

Further Reading

- Aho, E., Pitkänen, K., and Sahlberg, P. (2006). *Policy Development and Reform Principles of Basic and Secondary Education in Finland Since 1968*. Washington, DC: The World Bank.
- Centre for International Mobility (2003). *Higher Education in Finland. An Introduction*. Helsinki: CIMO.
- Dill, D. D., Mitra, S. K., Jensen, H. S., et al. (2006). *PhD Training and the Knowledge-Based Society*. Tampere, Finland: Finnish Higher Education Evaluation Council.
- Finnish Education Evaluation Council (2005). *New Directions in Educational Evaluation. Evaluation Programme 2004–2007*. Jyväskylä, Finland: The Finnish Education Evaluation Council.
- Jakku-Sihvonen, R. (ed.) (2002). *Evaluation and Outcomes in Finland – Main Results in 1995–2002*. Helsinki: National Board of Education.
- Jakku-Sihvonen, R. and Niemi, H. (eds.) (2006). *Research-Based Teacher Education in Finland – Reflections by Finnish Teacher Educators*. Turku: Finnish Educational Research Association.
- Kinnunen, E. and Halmevu, T. (eds.) (2004). *Student Assessment Guide for Vocational Education and Training*. Helsinki: Ministry of Education.
- Kyrö, M. and Nyyssölä, K. (2006). Attitudes towards education in Finland and other Nordic Countries. *European Journal of Education* 41(1), 59–70.
- Laukkanen, R. (2008). Finnish strategy for high-level education for all. In Soguel, N. C. and Jaccard, P. (eds.) *Governance and Performance of Education Systems*, pp 305–324. Netherlands: Springer.
- Linnakylä, P. (2004). Finland. In Döbert, H., Klieme, E., and Stroka, W. (eds.) *Conditions of School Performance in Seven Countries*, pp 150–218. Münster: Waxman.
- Linnakylä, P. and Arffman, I. (eds.) (2007). *Finnish Reading Literacy. When Quality and Equity Meet*. Jyväskylä, Finland: Institute of Educational Research.
- Ministry of Education (2004). *Education and Research 2003–2008. Development Plan*. Helsinki: Ministry of Education.
- National Board of Education (2003). *Vocational Qualifications in the School-Based System in Finland*. Helsinki: Ministry of Education.
- Pehkonen, E., Ahtee, M., and Lavonen, J. (eds.) (2007). *How Finns Learn Mathematics and Science*. Rotterdam/Taipei: Sense Publishers.
- Sahlberg, P. (2007). Education policies for raising student learning: The Finnish approach. *Journal of Education Policy* 22(2), 147–171.
- Simola, H. (2005). The Finnish miracle of PISA: Historical and sociological remarks on teaching and teacher education. *Comparative Education* 41(4), 455–470.
- Tuovinen, J. E. (2008). Learning the craft of teaching and learning from world's best practice. The case of Finland. In McInerney, D. M. and Liem, A. D. (eds.) *Teaching and Learning: International Best Practice*, pp 51–77. Charlotte, NC: Information Age Publishing.
- Välijärvi, J., Kupari, P., Linnakylä, P., et al. (2007). *The Finnish Success in PISA – and Some Reasons Behind It 2*. University of Jyväskylä, Institute for Educational Research.
- Westbury, I., Hansen, S.-E., Kansanen, P., and Björkvist, O. (2005). Teacher education for research-based practice in expanded roles: Finland's experience. *Scandinavian Journal of Educational Research* 49(5), 475–485.

Relevant Websites

- <http://www.eurydice.org> – Eurydice is the information network on education in Europe.
- <http://www.edu.fi> – Information in English about the activities of the National Board of Education is gathered on the NBE website.
- <http://www.isep.org> – International student exchange programs.
- <http://www.minedu.fi> – Ministry of education of Finland.
- <https://www.data.opi.fi> – National Board of Education, Research Database KOTU.
- <http://www.opi.fi> – National Core Curriculum for Basic Education 2004.
- <http://www.euroeducation.net> – The European education directory.
- <http://www.edev.fi> – The Finnish Educational Evaluation Council.
- <http://www.finheec.fi> – The Finnish Higher Education Evaluation Council (FINHEEC).
- <http://www.opi.fi> – The Finnish National Board of Education website.
- <http://www.helsinki.fi> – University of Helsinki.

France

M Sellier, Paris 75013, France

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Adaptation and integration school (AIS) – Since the 1970s, a policy to help special-needs children to adapt and to integrate into regular schools.

Baccalauréat – A diploma (equivalent A-levels) obtained at the end of secondary school and required for access to higher education.

Decentralization – It transfers powers and responsibilities of the State to local authorities or institutions, run by elected boards and managers.

Priority education zones (ZEPs) – Created in 1981, ZEPs are designed to provide additional means for the education of the most disadvantaged pupils by giving more for those who have less. Means are concentrated in social and economically disadvantaged areas.

Qualification – All learned competences (knowledge, know-how, and social skills) that people can mobilize in a job.

Recteur – A chief education officer appointed by decree by the President of the Republic, on proposal of the Minister of National Education. Responsible for primary, secondary, and higher education, he or she represents the Minister of National Education in an academy (there are 30 for 26 regions).

School life of the education community – All actions and reflection indissolubly linked to school as a complex system within an environment.

Secularism – Pupils and families have freedom of conscience: School as an institution must respect all political, philosophical, and religious choices although it recognizes none.

Technical education – It prepares pupils for different technical baccalauréats in secondary schools for general and technical education (LEGT) in the second cycle.

Vocational education – It prepares pupils for diplomas that enable immediate professional insertion. Today, students can opt for vocational high schools (LP) in the second cycle or for centers for apprenticeship (CFA). Both combine teaching vocational programs with training in enterprise.

French Context

France is a European country with its capital in Paris. The French Republic (official name) comprises metropolitan France as well as various overseas territories. Its total surface area is 675 000 km² (metropolitan France, 551 500 km²). In January 2008, the population was 64 473 000 (metropolitan France, 61 876 000), that is, 1% of the world's population. France's demographic growth is one of the fastest in Europe, combining a birth rate higher than the European average with a positive migratory balance.

The unification of France was achieved by its kings who, from Hugues Capet (987–996) to Francis I (1515–1547), gradually established their authority over the whole country. French, the language of the king, became the official language of the kingdom in 1539. The French Revolution of 1789 put an end to the Ancien Régime and reorganized French society on a new basis, which remains that of the Republic today. France has been a member of the European Union since its inception (Treaty of Rome, 1957).

Among the economic powers of the world, France ranked fifth or sixth (depending on the calculation method) in 2008. Although it is basically capitalist, the country's economy has been characterized by strong State intervention since Colbert (a Minister of Louis XIV in the seventeenth century). Two major problems plague the French economy today: its unemployment rate, which is one of the highest in Europe (7.5% in May 2008), particularly among women, people over 50 and the youth; and the size of the national debt and budget deficit, which put France in a difficult situation regarding the requirements of the Stability and Growth Pact adopted in 1997 by the Member States of the Euro zone.

In terms of the 1958 Constitution (which founded the Fifth Republic), France is an indivisible, secular, democratic and social Republic. Although it is a parliamentary democracy, the Constitution grants far-reaching powers to the President of the Republic, who is elected by direct universal suffrage.

The French Revolution laid the foundations for the edification of the Nation. One of its objectives was to ensure that education, until then provided largely by the Roman Catholic Church, was entrusted to the State and based on principles grounded in the Declaration of Human Rights and Rights of Citizens adopted in 1789, that is, public education, secularism, freedom, and equality. Secularism,

based on the separation of Church and State (the Law of 1905), stipulates that the State must respect all religions, although it recognizes none. A school quarrel has nevertheless raged for nearly two centuries between State education and private schools, most of them religious. The public education system is characterized today by its pluralism. It includes schools operating under contract, most of them Roman Catholic, which provide education for approximately 17% of all pupils – a rate that has remained stable since 1959.

Metropolitan France is divided into three tiers of local government authorities, namely, 36 568 communes, 96 departments (instituted by the Revolution), and 22 regions (since 1982). The State, which traditionally concentrated far-reaching powers at central level, has since 1980 introduced measures that have increased decentralization to local authorities, in particular in the field of education. In the past, several key events marked the intervention of the State in this respect: the organization of secondary education, by Napoleon; that of primary education teaching for boys (Guizot Law, 1833); and the institution, under the Third Republic (1870–1940), of free and compulsory schooling for boys and girls aged 6–13 years (Jules Ferry Laws, 1881–82). The system was completed under the Fourth (1946–58) and Fifth Republics.

Objectives of the Education System

The objectives of the French education system are set out in Section 1, Paragraph 1, Article L111-1 of the Education Code, which guarantees every citizen the right to education to enable them to develop their personality, to improve their level of schooling and further education, to participate fully in social and professional life, and to exercise their rights and responsibilities as citizens.

The 23 April 2005 framework law for the future of schooling added the following amendment to this first paragraph: “in addition to imparting knowledge, the main mission of schools, as established by the Nation, is to transmit the values of the Republic to pupils.” The public school system must allow access to all children, without exception, and strive to ensure their success.

The 11 February 2005 law on the equality of rights and opportunities for disabled persons reaffirms the right of special-needs children to education. The education system is now legally obliged to accept disabled children at the school closest to their home, which is their reference school. The parents of these children must be involved in the development of a personalized schooling project for their child.

The obligation to enrol children in school applies likewise to new immigrants who have been in the country

for less than a year (i.e., 3.9% of the total school population). These pupils are entitled to specific measures to facilitate their integration.

The role of schools is to contribute to equal opportunities for pupils to acquire knowledge, skills, and qualifications, on which their access to the labor market, especially in France, will depend. France is strongly attached to the principle of equal opportunities and the offer of the same quality of teaching for everyone throughout the country. It is only since the 1980s that priority education zones (PEZs) have been created to provide additional means for the education of the most disadvantaged pupils.

The Law of 23 April 2005, in the attached report, assigned quantitative objectives to the education system, to continue the democratization policy, and to meet France’s European commitments. “This is why the Nation has set the objectives for the education system to guarantee that, when they leave school, 100% of pupils have obtained a qualification, and to ensure that 80% of pupils in every age-group (i.e., born in the same year) reach the *baccalauréat* [equivalent A-levels] level.” The law also set the objective of ensuring that 50% of pupils in every age-group obtain a higher education degree or diploma.

Presentation of the Education System

In 2006–07, education was provided for 14 972 849 young people (pupils and students), in State and private schools, accounting for 23.5% of the total population in France. Schooling is compulsory from ages 6 to 16 years.

Pupils in Primary Education

In 2006, French primary schools provided education for 6 644 100 pupils, including 111 100 disabled pupils.

Nursery or pre-elementary education precedes the years of compulsory schooling (which begins at age of 6 years). In 2006, nursery schools, of which 87.6% were in the public sector, accommodated 2 578 445 children. Characterized by their strong and original identity, nursery schools fulfill an educational mission of which France is very proud. They provide education to all 3-year-old children and 23.4% of 2-year-old children.

Elementary schools, beginning at age 6, provide education to all French children (4 016 900 in 2006) for 5 years, from the preparatory course (CP) to the end of the intermediate course (CM2).

Special-needs children in the special schools/classes (AIS) set up to help them to adapt and to integrate into regular schools account for 3.5% of enrolments annually and total 48 700 pupils.

Pupils in Secondary Education

In 2006, a total of 5 418 100 pupils attended secondary school in France, 79.3% of them in public schools. Up to 44 300 were special-needs children who received individual or collective, part-time or full-time schooling. There were 201 100 students of foreign nationality in secondary schooling (no figures are available for primary schooling), that is, 3.7% of the total. Some 155 300 students, 2.9% of the total, were enrolled in schools falling under the Ministry of Agriculture and Fisheries.

Secondary schools in the first cycle

In the French system, the first cycle of secondary schooling (called *collège*, equivalent to junior high school) lasts for 4 years, theoretically from ages 11 to 15 years. The principle of secondary schooling for all was enshrined in a law passed in 1975 and has been applied since then. At the end of these 4 years, students sit for a national examination, *diplôme national du brevet* (DNB). In 2006, 57% of students (with or without the DNB) furthered their education in high schools for general and technical education, *lycées d'enseignement général et technologique* (LEGT) and 26% went on to vocational high schools, *lycées professionnels* (LP). Six percent of those who failed repeated the year and 11% left the system of schooling delivered under the Ministry of National Education (some continued with sandwich courses or training delivered by institutions falling under other ministries, while others exited the educational system altogether).

Secondary schools for general and technical education (LEGT) in the second cycle

The LEGT covers the last 3 years of schooling (*seconde*, *première*, and *terminal*). The first year of this cycle is decisive as students choose the main subjects that they are going to carry through to the end. Slightly under two-thirds of all students opt for the general stream, and just over one-third for the technology stream. At the end of the cycle, they sit for the national general or technical *baccalauréat* examination.

Vocational high schools (LPs) in the second cycle

The LPs offer 48 different specialities in which teaching at school is combined with training in enterprise. They prepare pupils for diplomas that qualify them for immediate professional insertion: *certificat d'aptitude professionnelle* (CAP); and *brevet d'études professionnelles* (BEP) or what is known as the *bac pro* (*baccalauréat professionnel*). Of the 26% of students who opt for an LP, 22% take up a BEP and 4% a CAP. The dropout rate at the end of the first year is high: about 16.5% in CAP and about 12.5% in BEP either drop out of schooling completely, or enter apprenticeship, or else move on to another training system.

At the end of the second year, 41% of students doing the BEP and 10% of those enrolled for a CAP choose

to continue for a further two years and to prepare a vocational *baccalauréat* (*bac pro*, created in 1985). Theoretically, this qualification is obtained 4 years after the DNB (although obtaining it in 3 years is currently under consideration).

Apprentices

Apprentices (385 859 in 2006) are young people in the 16–25 age group preparing for a vocational or technical diploma under an employment contract involving training in a company and vocational education programs conducted in a center for apprenticeship, a *centre de formation d'apprentis* (CFA). Most of these students (59%) follow a low level of training (leading mainly to the CAP and the BEP).

Students in Higher Education

In 2006, 2 254 000 students (of which 11.7% were foreign students) were registered in higher education institutions in France, among them 79.2% of the *baccalauréat* graduates from the previous year continuing their studies at university. Students registered in the universities that do not apply selection criteria at entry (62%) can do a 3-year degree called a license (equivalent of a bachelor's degree), a further 2 years to obtain a master's degree and, in a further 3 years, a PhD, following the European pattern (B/M/D).

Ten percent of students do a 2-year post-*baccalauréat* higher technical diploma, a *brevet de technicien supérieur* (BTS) delivered at high schools. The 76 160 students selected for preparatory classes for the prestigious *grandes écoles*, the *classes préparatoires aux grandes écoles* (CPGE), spend 2 or even 3 years on this preparation, as entry to these schools is highly competitive (e.g., the *Ecole Normale Supérieure* or the *Ecole Polytechnique*). This is a French particularity. Engineering degrees account for about 5% of all university degrees, which is considered insufficient to meet the country's needs.

Adults in Continuous Education

Despite being mandatory for firms, local authorities, and the State, since the Law of 16 November 1971, continuous and ongoing education concern only a limited number of people. In 2005, they accounted for 446 100 trainees in the National Education system.

Teaching Staff and Other Educational Personnel

In January 2007, the Ministry of National Education paid the salaries of 1 065 327 employees in the public education sector and another 144 501 in the private education sector under contract with the government. Four out of five

employees were teachers, that is, 982 678. In the public sector, 97.7% of teachers are civil servants with permanent posts, and their average age is 41.8.

Primary school teachers and all secondary education teachers are trained at university in *instituts universitaires de formation des maîtres* (IUFM). New recruits, who must have a license, study for 1 year to prepare for a competitive examination for entry into teaching. In the second year, if they are successful, they follow 1 year of practical training as paid trainee teachers. In 2006, teachers in the first stage of training numbered 13 404 (less than 15% of applicants in 2005), while those in the second stage of training numbered 9584 (less than 12% of applicants). Teaching as a profession seems to be suffering from a malaise and attention is being given to this as a general problem.

In 2007, 227 150 individuals had administrative, technical, or educational positions in the education system.

Administrative Organization and Management

For a long time, the political and administrative organization of France was characterized by the pre-eminence of central State authority, the heritage of an absolute monarchy. Until the end of the 1970s, the administration of education was characterized by a pyramidal system where all the decisions were made at the top by the Ministry of National Education. Corresponding to this administrative organization was a centralized trade-union structure.

The decentralization laws of 1982, 1983, and 1985 profoundly changed this configuration. A significant share of the State's responsibilities was transferred to local authorities: Communes (primary schools), departments (*collèges*), or regions (*lycées*), depending on the level of the educational institution concerned. Although autonomous since 1968, the universities have remained under national authority.

Decentralized powers concern the construction, maintenance, and functioning of educational institutions. As provided for by the Law of 13 August 2004, enrolment, catering, accommodation, general maintenance, and personnel management have fallen within the province of the departments – for *collèges* – and the regions – for *lycées* – since 1 January 2005.

In practice, local authorities have not allowed themselves to be confined to a strict management role. They are involved in pedagogical choices and the school life of the educational community and have, in particular, multiplied initiatives to support pupils in difficulty.

In parallel, many management decisions were transferred from national level to the *recteurs d'académie* (chief education officers, of which there are 30, for 26 regions in France), to the *inspecteurs d'académie* (directors of

departmental services, of which there are 100 in France), and to the principals of secondary schools (11 410 schools, public and private).

The State has nevertheless maintained important prerogatives: responsibility for national programs, the awarding of diplomas and degrees (including the *baccalauréat*), recruitment, remuneration, the management of most staff, including teaching/lecturing staff, the allocation of funds, and the control and evaluation of educational policies.

Overall, higher education remains within the exclusive domain of the State but receives substantial grants from local authorities. Primary schools (of which there are 55 667) have been a service delivered and managed by the communes since the nineteenth century (although the teachers are employed by the State). In secondary education, the State and local authorities have shared responsibilities and must work together.

To steer changes in the education system, which was traditionally regulated according to bureaucratic procedures, France is endeavoring to move toward a form of management-by-results and a new system of educational governance.

The country has an excellent system of statistics and of quantitative and qualitative evaluation, as well as tools for forecasting and reflection. In 1987, it set up a department of evaluation, forecasting, and performance, the *Direction de l'évaluation, de la prospective et de la performance* (DEPP) within the Ministry of Education. The government has entrusted the mission of evaluating all educational institutions, as well as the education system as a whole, to two general inspectorates: 159 IGEN, a general inspectorate of national education, and the 93 IGAENR, a general inspectorate of the administration of national education and research. A significant number of inspectors of primary and secondary schools (3237) are responsible for the quality of teaching. For higher education and research, an evaluation agency, the *Agence d'évaluation de la recherche et de l'enseignement supérieur* (AERES) was created in 2007 to replace the national evaluation committee, the *Comité national d'évaluation* (CNE). In 2005, the higher council for schools, the *Haut Conseil de l'Ecole* (HCE), replaced the higher council for the evaluation of schools, the *Haut Conseil de l'évaluation de l'école* (HCEE) that had been created in 2000. Every year, the HCE delivers a public report on the results of the education system to the President of the Republic. This report is submitted to Parliament.

A new evaluation culture has thus developed which, paradoxically, seems to be used far more at the level of the system as a whole than at that of individual schools.

With the *Loi organique relative aux lois de finances* (LOLF) the organic law relating to finance laws, passed by the Parliament on 1 August 2001, France obtained a new financial constitution. Since 1 January 2006, the Parliament has been able to discuss policy and budget

priorities and objectives concerning education, to which the government is then committed. On this basis, it votes on the required means for meeting those objectives and evaluation results. France is currently experimenting with this new mode of political steering of education.

The Cost of Education

In 2006, France devoted 121.4 billion euros to its education system, that is, 6.8% of its gross domestic product (GDP) and the equivalent of 1920 euros per capita, or 7160 euros per pupil or student. Excluding adult continuous education, this level of investment situates France above the average of the Organisation for Economic Co-operation and Development (OECD) countries (6.1% against an OECD average of 5.8% in 2004).

The share of expenditure on education in the national budget was appreciably increased in the early 1990s, reaching 7.6% in 1993 against 6.4% in 1980. Since then, levels of expenditure on education have continued to increase, but less quickly than the growth of national wealth. This growth in expenditure is explained less by the increase in the number of pupils and students than by the rising cost per pupil.

The State is the main source of funding of public education (63% in 2006, of which 55.9% was funded by the Ministry of National Education). The Ministry's budget is primarily used for salaries. Local authorities now foot 20.6% of the total bill (against 14% in 1980), while households contribute 8% and companies 6.4%.

At the time of writing, in 2008, due to financial constraints, a policy of not replacing one out of two teachers leaving on retirement is under consideration.

Results of the Education System

The French education system made important progress, from both a quantitative and qualitative point of view, during the three decades up to the mid-1990s. Since 1994, the rate of schooling in nursery schools of children from the age of 3 has been 100%, compared with 36% in 1960. During the 1960s and 1970s, access to secondary education was generalized. From the mid-1980s, pupils who graduated from *collège* massively opted for the second cycle to do a general, technical, or vocational *baccalauréat* (the latter was created in 1985), and then to go on to higher education.

The progress of a generation currently engaged in the French education system or which has just left it can be summarized as follows: almost all young people today complete the first cycle of secondary school (81.1% pass the DNB), 70% reach the level of the *baccalauréat*; 64% actually obtain their *baccalauréat*, 50% go on to higher education, and a little more than 40% obtain a higher education degree.

The objective announced in the mid-1980s and included in the Law of 2005, to ensure that 80% of a generation reached their final year of *lycée* –that is, *baccalauréat* or equivalent – may not have been achieved, but it has led to a spectacular rise (over 30 points) in the rate of access to the *baccalauréat* level in just one decade, from under 40% in 1985 to 70% in 1995. School has thus made it possible for young generations to reach much higher levels of education than those of their elders. At the end of the 1990s, slightly over 40% obtained a higher education degree, against 30% at the end of the 1960s, and 20% at the end of the 1950s.

Seventeen percent of young people (approximately 130 000 per generation) leave secondary school without any diploma. This proportion was close to 50% for the generation now aged 55–64. Out of that 17%, close to 6% leave the education–training system without a professional qualification (i.e., without having reached the level of the CAP or the BEP, and without having been accepted in an LEGT).

The rapid development of schooling has definitely supported the democratization of the education system, but all national evaluations show that school results depend strongly on the social origin of the pupils at all levels of schooling. Difficulties and inequalities persist, especially in the level of basic skills acquired, and these have serious consequences for young people's schooling and job opportunities. This is why the Law of 23 April 2005 now requires every young person at the end of the first cycle of secondary school (i.e., *collège*) to acquire a common basic level of knowledge and skills.

France still needs to make an effort to achieve the goals set by the European governments at the Summit of Lisbon in 2000 with respect to the generalization of the second cycle of secondary education, the reduction of dropouts, the development of adult education, the improvement of basic reading skills, and the increase in the number of graduates in science and technology.

The Major Issues Debated since 1995

The development of comparative studies at international level (OECD, EUROSTAT, and UNESCO), and, in particular, the Programme for International Student Assessment (PISA), have made it possible for France to continue its reflection on education.

Since 1995, several OECD indicators have shown no further improvements for France (e.g., the rate of access to the *baccalauréat* level continues not to exceed 70% of a generation). The education system appears to have reached a threshold that it is unable to exceed.

Whereas half of the pupils in France obtain excellent results when compared with students internationally, the other half's scores are among the worst of the developed countries. This suggests that the school system works

successfully for the pupils who adapt to it, but does not manage to meet the needs of pupils in difficulty, especially the needs of the 130 000 young people who leave school every year in France without any diploma and are very likely to find themselves unemployed.

After *collège*, 40% of pupils are directed (in certain cases, against their will) to a short period of vocational training to obtain a certificate or diploma. This is a very high percentage by comparison with the other countries, and explains the insufficient number of graduates in higher education.

Among the students who continue their studies, only 20% of a generation obtain a higher education degree equal to or higher than the license (*baccalauréat* + 3 years). This is not enough to meet France's needs.

This low rate is alarming because France seems to be one of the Western countries where the link between the qualifications obtained at school and at work is the strongest, not only at entry into the workforce but throughout life. It is difficult to compensate for the low level of initial training later in life because adult ongoing education is developing very slowly.

Although the school system has effectively been democratized over the past 20 years, it still reproduces social elites rather than guaranteeing equal opportunities for all. These results are not only contrary to the goal of equity at which the education system aims; they also seem to be counter-productive as regards the needs of the economy.

The 50% of pupils who obtain good results in the PISA assessments, tend to be those who have made it without having to repeat a year, whereas the other half consists of those pupils who obtain weak results even after repeating at least 1 or 2 years – as if repeating compounded their difficulties. Although requiring pupils to repeat a year is a practice that is decreasing, it remains deeply entrenched in the French education system which, in this respect, holds the worst record of European countries (up to 38% of pupils have repeated at least once at the end of the first cycle of secondary school). These extra years are a heavy social and financial burden for the country.

The pattern of boys' and girls' schooling tends to differ: more girls than boys obtain their *baccalauréat* and enter higher education. However, girls are overrepresented in the humanities and training for the service sector, to the detriment of the scientific and technological fields.

Of all these key issues open to debate, the most important and unquestionably the most difficult to solve, remains the evaluation of teaching and nonteaching staff. There is all-round agreement that good management of human resources (the term itself is controversial) appears necessary, but the need to establish a link between pupils' success and teachers' teaching practices, or even the management approach of a school principal, has been formulated only recently.

Because of the contrasting results obtained by the education system and its historical role in the construction of the Republic, the school system in France is a subject of

constant and impassioned debates, which makes any reform difficult to implement.

Further Reading

- Association française des administrateurs de l'éducation (AFAE) (2008). *Le système éducatif français et son administration coédité avec le CRDP de Poitou-Charentes*, p 373.
- Champy, P., Etévé, C., Forquin, J.-C., and Robert, A. D. (2005). *Dictionnaire encyclopédique de l'éducation et de la formation*, 3^e édition. Retz, p 1102.
- Crozier, M. (1963). *Le phénomène bureaucratique*, thèse de doctorat de lettres, Université de Paris, Paris, Le Seuil, p 402.
- Dubet, F. (2004). *L'école des chances: qu'est-ce qu'une école juste?* Le Seuil, La République des Idées, p 96.
- Duru-Bellat, M. (2002). *Les inégalités sociales à l'école, Genèse et mythes*. Paris, PUF, p 256.
- Forestier, C., Thelot, C., and Emin, J.-C. (2007). *Que vaut l'enseignement en France?* Paris, Stock, p 293.
- Inspection générale de l'Education nationale, Inspection générale de l'administration de l'Education nationale et de la Recherche, (2005), (2006), (2007). *Rapports annuels des inspections générales*, La Documentation française.
- Ministère de l'Education nationale – Direction de l'évaluation, de la prospective et de la performance, *L'état de l'Ecole de la maternelle à l'enseignement supérieur*, n° 17, octobre 2007.
- Ministère de l'Education nationale – Direction de l'évaluation, de la prospective et de la performance, *L'état de l'Enseignement supérieur et de la Recherche*, n° 1, novembre 2007.
- Ministère de l'Education nationale – Direction de l'évaluation, de la prospective et de la performance, *Géographie de l'Ecole*, édition n° 9, mai 2005.
- Ministère de l'Education nationale – Direction de l'évaluation, de la prospective et de la performance, *Repères et références statistiques sur les enseignements, la formation et la recherche (RERS 2007)*, septembre 2007.
- Mons, N. (2007). *Les nouvelles politiques éducatives, la France fait-elle les bons choix?* Paris, PUF, p 202.
- Sellier, M. (2005). *Dictionnaire encyclopédique de l'éducation et de la formation*, 3^eédition. Retz, pp 237–240.
- Sellier, M. (2005). *Dictionnaire encyclopédique de l'éducation et de la formation*, 3^eédition. Retz, pp 1038–1040.
- Sellier, M. (2006). coordination du numéro, *L'école au féminin. Administration et éducation, revue de l'AFAE*. n°2.
- Sellier, M. (2007). coordination du numéro, *L'élitisme républicain en question. Administration et éducation, revue de l'AFAE*. n°4.
- Toulemonde, B. (2006). *Le système éducatif* (La documentation française, coll. Les notices 2^eédition, pp 192.
- Van Zanten, A. (2004). *Les politiques d'éducation*, Paris, PUF, p 126.

Relevant Websites

- <http://www.afae.fr>
<http://www.afae.systemeduc.fr>
<http://eduscol.education.fr>
<http://www.cafepedagogique.net>
<http://cisad.adc.education.fr/hcee/index.html>
<http://www.hce.education.fr>
<http://www.education.gouv.fr>
<http://www.esen.education.fr>
<http://www.eurydice.org>
<http://ieeps.org>
<http://www.infocentre.education.fr>
<http://www.ladocumentationfrancaise/rapports/index.shtml>
<http://www.oecd.org>
<http://www.cpu.fr>
<http://www.premier-ministre.gouv.fr>

Germany

A Hüfner and K Hüfner, Berlin, Germany

© 2010 Elsevier Ltd. All rights reserved.

General Background

Since the unification of the two German states on 3 October 1990 which marked the end of over 40 years of division, the new Federal Republic of Germany covers a total of 357 thousand square kilometers. It lies at the heart of the European Union (EU) and shares its borders with nine neighboring countries. Except Switzerland, all of them are also members of the EU. With a population of over 82 million, Germany is one of the most densely populated countries in Europe. In 2005, 6.8 million were of foreign nationality, of which the Turks represented with 26% the largest group. It continues to be a major challenge for the educational system to integrate the children of migrant workers and to offer them adequate opportunities for study, training, and work.

Germany is a democratic and social federal state made up of 16 Länder (states) since 1990. The political, economic, social, and cultural integration process of the former two states takes longer than originally anticipated; the necessary changes from a socialist planned economy of the former German Democratic Republic toward a new type of social market economy are still underway. For the people living in the five new Länder of the eastern part of Germany, these transition processes from security without freedom toward freedom without security implied big problems of adjustment toward unknown market-oriented demands of flexibility. In the areas of culture, education, and science, the Unification Treaty (*Einigungsvertrag*) of 31 August 1990 contained basic provisions in order to establish a common and comparable basic structure in the school system and a common, though differentiated, higher education and research system in the 16 Länder of the new Federal Republic of Germany, primarily according to the model of the former Federal Republic of Germany.

Germany is not rich in natural resources. Therefore, the country is heavily dependent on imported goods (40% of gross national product (GNP) in 2006) and the export of finished products and services (45% of GNP in 2006). GNP per capita reached US\$35 000. Two percent of the 39 million employed work in agriculture, 26% in industry, and 72% in the service sector. Between 2005 and 2007, the unemployment rate went down from over 10.0% to 8.4% remaining high by Organization for Economic Cooperation and Development (OECD) standards. Until 2007, Germany enjoyed economic recovery after a long phase of stagnation since the end of 2004.

In its most recent economic survey of 2008, OECD has included a special chapter on educational policy stressing that improving education outcomes is important for the country's long-term economic performance and social cohesion. This analysis must also be seen in the light of the Lisbon goal of 2000 when the EU heads of state declared their ambition to make the EU "the most competitive and dynamic knowledge-based economy in the world by 2010, capable of sustainable economic growth, with more and better jobs and greater cohesion."

Institutions at the National Level

According to the Basic Law (*Grundgesetz*), the cultural sovereignty (*Kulturbobkeit*) implies the predominant responsibility of each of the Länder for education, science, and culture. Each Land is responsible for its educational and cultural policy. Educational and cultural legislation is primarily, the administration of education and culture almost entirely the responsibility of the 16 Länder. Therefore, one could also argue that Germany consists of 16 educational systems which can be characterized as highly differentiated systems (cf. **Figure 1** and especially the annotations as published by the Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany).

However, within the Federal Government (*Bundesregierung*), it is the Federal Ministry of Education and Research (BMBF) that is responsible for policy, legislation, and coordination with regard to out-of-school vocational training and continuing education, financial assistance for students, admission to higher education institutions and the degrees they confer. Furthermore, the BMBF shares with the Länder, joint tasks (*Zusammenwirken*) according to Article 91 b of the Basic Law. In order to coordinate cooperation among the Länder and with the Bund, the Länder established in 1948 the Standing Conference of the Ministers of Education and Cultural Affairs (*Ständige Konferenz der Kultusminister der Länder; Kultusministerkonferenz; KMK*).

The General Structure of the Education System

Although German is the normal language of instruction and training at general education and vocational schools as well as higher education institutions, some exception must be mentioned. Those include bilingual schools and

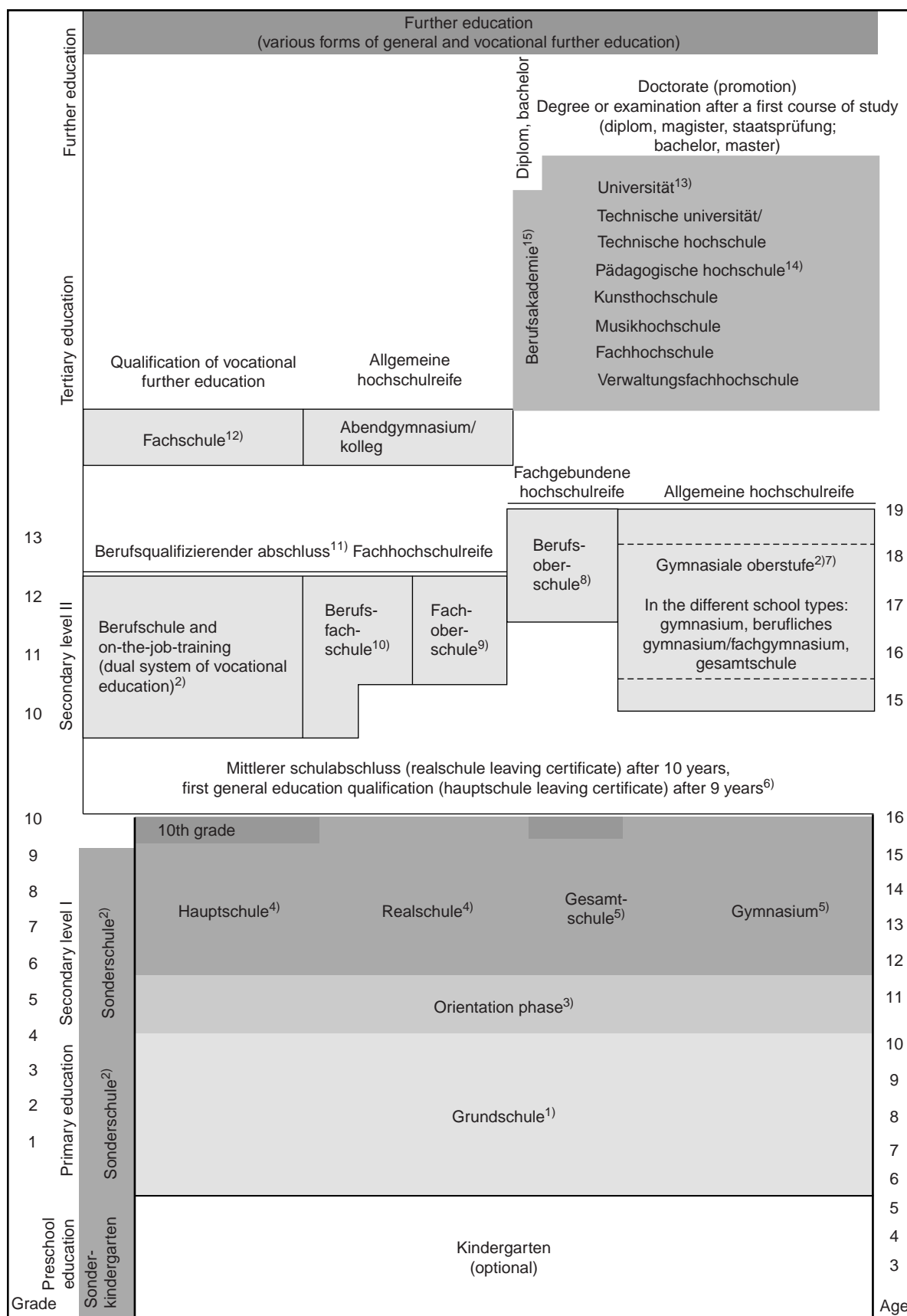


Figure 1 (Continued)

Annotations

Diagram of the basic structure of the education system. The distribution of the school population in grade 8 as per 2005 taken as a national average is as follows: *Hauptschule* 22.5%, *Realschule* 25.8%, *Gymnasium* 30.9%, *integrierte Gesamtschule* 8.5%, types of school with several courses of education 6.3%, special schools 5.3%.

The ability of pupils to transfer between school types and the recognition of school-leaving qualifications is basically guaranteed if the preconditions agreed between the Länder are fulfilled. The duration of full-time compulsory education (compulsory general education) is 9 years (10 years in four of the Länder) and the subsequent period of part-time compulsory education (compulsory vocational education) is 3 years.

1. In some Länder special types of transition from pre-school to primary education (*Vorklassen*, *Schulkindergärten*) exist. In Berlin and Brandenburg the primary school comprises six grades.
2. The disabled attend special forms of general-education and vocational school types (partially integrated with non-handicapped pupils) depending on the type of disability in question. Designation of schools varies according to the law of each Land (*Sonderschule/Schule für Behinderte/Förderschule/Förderzentrum*).
3. Irrespective of school type, grades 5 and 6 constitute a phase of particular promotion, supervision and orientation with regard to the pupil's future educational path and its particular direction. In some Länder, the orientation stage (*Orientierungsstufe* or *Förderstufe*) is organised as a separate school type.
4. The *Hauptschule* and *Realschule* courses of education are also offered at schools with several courses of education, for which the names differ from one Land to another. The *Mittelschule* (Sachsen), *Regelschule* (Thüringen), *Sekundarschule* (Bremen, Sachsen-Anhalt), *Erweiterte Realschule* (Saarland), *Integrierte Haupt-und Realschule* (Hamburg), *Oberschule* (Brandenburg), *Verbundene* or *Zusammengefasste Haupt-und Realschule* (Berlin, Hessen, Mecklenburg-Vorpommern, Niedersachsen) and *Regionale Schule* (Mecklenburg-Vorpommern, Rheinland-Pfalz), *Regionalschule* (Schleswig-Holstein), *Gemeinschaftsschule* (Schleswig-Holstein) as well as comprehensive school (*Gesamtschulen*) fall under this category.
5. The *Gymnasium* course of education is also offered at comprehensive schools (*Gesamtschule*). In the cooperative comprehensive schools, the 3 courses of education (*Hauptschule*, *Realschule* and *Gymnasium*) are brought under one educational and organisational umbrella; these form an educational and organisational whole at the integrated *Gesamtschule*. The provision of comprehensive schools (*Gesamtschulen*) varies in accordance with the respective educational laws of the Länder.
6. The general education qualifications that may be obtained after grades 9 and 10 carry particular designations in some Länder. These certificates can also be obtained in evening classes and at vocational schools.
7. Admission to the *gymnasiale Oberstufe* requires a formal entrance qualification which can be obtained after grade 9 or 10. At present, in the majority of Länder the *Allgemeine Hochschulreife* can be obtained after the successful completion of 13 consecutive school years (9 years at the *Gymnasium*). Yet in almost all Länder the gradual conversion to 8 years at the *Gymnasium* is currently under way, where the *Allgemeine Hochschulreife* can be obtained after a 12-year course of education.
8. The *Berufsoberschule* has so far only existed in a few Länder and offers school-leavers with the *Mittlerer Schulabschluss* who have completed vocational training or 5 years' working experience the opportunity to obtain the *Fachgebundene Hochschulreife*. Pupils can obtain the *Allgemeine Hochschulreife* by proving their proficiency in a second foreign language.
9. The *Fachoberschule* is a school type lasting for two years (grades 11 and 12) which admits pupils who have completed the *Mittlerer Schulabschluss* and qualifies them to study at a *Fachhochschule*. Pupils who have successfully completed the *Mittlerer Schulabschluss* and have been through initial vocational training can also enter the *Fachoberschule* directly in grade 12.

Figure 1 (Continued)

10. *Berufsfachschulen* are full-time vocational schools differing in terms of entrance requirements, duration and leaving certificates. Basic vocational training can be obtained during one- or two-year courses at *Berufsfachschulen* and a vocational qualification is available at the end of two- or three-year courses. Under certain conditions the *Fachhochschulreife* can be acquired on completion of a course lasting a minimum of 2 years.
11. Extension courses are offered to enable pupils to acquire qualifications equivalent to the *Hauptschule* and *Realschule* leaving certificates.
12. *Fachschulen* cater for vocational continuing education (1–3 year duration) and as a rule require the completion of relevant vocational training in a recognised occupation and subsequent employment. In addition, the *Fachhochschulreife* can be acquired under certain conditions.
13. Including institutions of higher education offering courses in particular disciplines at university level (e.g. theology, philosophy, medicine, administrative sciences, sport).
14. *Pädagogische Hochschulen* (only in Baden-Württemberg) offer training courses for teachers at various types of schools. In specific cases, study courses leading to professions in the area of education and pedagogy outside the school sector are offered as well.
15. The *Berufsakademie* is a tertiary sector institution in some Länder offering academic training at a *Studienakademie* (study institution) combined with practical in-company professional training in keeping with the principle of the dual system.

As at July 2007

Source: Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany. Documentation and Education Information Service.

Figure 1 Basic structure of the educational system in the Federal Republic of Germany.

classes as well as instruction and extra classes in the mother tongue for pupils whose mother tongue is not German. Furthermore, the Länder concerned apply the European Charter of Regional Minority Languages of the Council of Europe to those speaking Danish, Frisian, Sorbian, Romany, and Low German. In the higher education institutions, increasingly English is also used as a language of instruction and as working language. This is also in line with the Bologna Process leading to a European Higher Education Area.

School attendance is compulsory for children between the ages of 5–6 and 18 years. Depending on the Land, full-time schooling is mandated for either 9 or 10 years; it may be followed by part-time attendance at a vocational school, complemented by an apprenticeship (dual system). All compulsory schooling is free of charge.

In principle, the education system is divided into the following levels:

- pre-primary education,
- primary education,
- secondary education, and
- tertiary education.

Pre-Primary Education

Pre-primary education is provided by institutions catering for children from a few months to 5–6 years when they

usually start school (*kindergarten* being the traditional form for children from 3 to 5–6 years). Traditionally, those institutions are not regarded as part of the educational system; they are run by the nonpublic and public child and youth welfare services, for example, by churches, welfare associations, local authorities (*Kommunen*), and parents' associations. In most Länder, the legal and administrative competence lies with the ministries of social affairs.

All children having reached the age of 3 years have the legal right to be admitted to a kindergarten until school entry. The main responsibility is with the *Kommunen* which are obliged to provide places in day-care centers to all children from 3 years until they start school. Parents must pay fees for the attendance of pre-school institutions. The amount depends, *inter alia*, upon parents' income and family size. Parents can apply for full or partial reimbursement by the local youth welfare office if they cannot afford to bear the costs. Whereas in 1960 only 25% of the age group attended pre-school institutions, its share increased to 87% by 2006.

Primary Education

Primary education is compulsory for all. Deferment of school attendance is only possible in exceptional cases. In the majority of the Länder, those children must attend *Schulkindergarten* or preschool classes (*Vorklassen*) in order

to reach the necessary level of individual development. Earlier, compulsory schooling started on 1 August for all children having reached their sixth birthday. Nowadays, a general tendency can be observed to set the statutory qualifying date for starting school on 31 December. Berlin started in 2004, Bavaria and North-Rhine-Westfalia followed. This implies that compulsory education starts between the age of 5 and 6 years.

For the first 2 years of primary education (*Grundschule*), school work is not graded; instead, teachers provide evaluations of the pupils' strengths and weaknesses. Foreign language instruction often begins as early as third or fourth grade. By fifth grade, all students are learning a foreign language, in most cases English. Other subjects are German, mathematics, science, art, music, and physical education. The *Grundschule* lasts for 6 years in the Länder of Berlin and Brandenburg and for 4 years in all the other Länder.

Secondary Education

Upon completion of the *Grundschule* at which all children attend mixed-ability classes, the secondary school system is characterized primarily by three school types which divide the students of age 10–12 into various educational paths with different leaving certificates and qualifications: *Hauptschule*, *Realschule*, and *Gymnasium*. Despite all attempts to increase the horizontal mobility, this differentiated system at the secondary level is still the characteristic feature of German education. Apart from these three types, all Länder have *Gesamtschulen* (comprehensive schools), although highly differing in number (in 2003, e.g., two institutions in Bavaria and 217 in North-Rhine-Westfalia). Several Länder introduced types of schools in which the traditional courses offered by the *Hauptschule* and the *Realschule* are combined under one organizational umbrella. Names, however, differ from Land to Land (cf. **Figure 1** and the annotations).

Secondary education breaks down in lower secondary level (*Sekundarschule I*) from grades 5 to 9/10 and upper secondary level (*Sekundarstufe II*) from grades 10 to 12/13. Transition from *Grundschule* to one of the school branches at secondary level depends on legislation in the Land concerned. The decision is either taken by the parents

or the school or the school supervisory authority, based on an assessment made by the *Grundschule*.

Whereas *Hauptschulen* provide its students with a basic general education preparing them for an apprenticeship afterward, *Realschulen* offer a more extensive general education. The *Hauptschule* normally covers grades 5–9/10, the standard *Realschule* grades 5–10. A *Realschule*-leaving certificate qualifies for a transfer to a school that provides a vocational or higher education entrance qualification.

The *Gymnasium* provides an intensified general education; it comprises both the lower and the upper secondary level and covers grades 5 or 7–12/13; its leaving certificate (*Abitur*) provides a general higher education entrance qualification. In almost all the Länder, the reduction from 9 to 8 years at the *Gymnasium* is currently underway. As mentioned above, in addition to the three school branches, there are *Gesamtschulen* in the majority of Länder. Two types can be distinguished. Whereas the cooperative one combines the *Hauptschule*, *Realschule*, and *Gymnasium* in one organizational unit and offers instruction in classes designed for the achievement of various certificates, the integrated one forms one organizational and educational unit, differentiating classes in some of the subjects according to the level of proficiency.

Table 1 shows how the students in grade 8 were distributed over time. Whereas the share of students at the *Hauptschule* decreased from 72% in 1960 to 24% in 2005, a clear increase can be observed in the attendance at *Gymnasium* and *Realschule* moving from 17% and 11% to 33% and 27%. Since the *Hauptschule* tends to become a kind of left-overs school (*Restschule*) with a disproportionately high percentage of students from migrant families, presently alternative solutions are discussed or even started ranging from an integration (cf. **Table 1**) to its full abolition.

The vocational training system has enjoyed a high reputation based on its success in producing the necessary skills through the dual system of education, a combination of part-time vocational schooling and part-time on-the-job training, which is open to all lower secondary school leavers. It is organized for some 340 professions. In all economic fields, training regulations (the workplace element) are nationally coordinated and framework curricula are established by the Länder (the school element).

Table 1 Distribution among students in grade 8 according to school branch, 1960–2005 (in percent)

Year	Gymnasium	Realschule	Hauptschule	Gesamtschule	Integrated Haupt-und Realschule
1960	17	11	72	–	–
1970	23	21	56	–	–
1980	27	28	41	4	–
1990	30	29	34	7	–
2000	31	26	23	10	10
2005	33	27	24	10	7

From Bundesministerium für Bildung und Forschung (2008). *Grund-und Strukturdaten 2007/2008*, p. 25. Bonn, Berlin: BMBF. p. 25.

Students pass a final examination before an examination board composed of representatives of the chambers concerned. Concomitantly, they receive a second leaving certificate from the vocational school (*Berufsschule*) if they have achieved the necessary results.

Higher Education

The tertiary sector encompasses higher education institutions which include universities and equivalent institutions, for example, for teacher training (*Pädagogische Hochschulen*), theology, art, music, sport, and universities of applied sciences (*Fachhochschulen*) (cf. Table 3). In 2003, Germany had a total of 366 institutions of higher education of which 99 were private institutions (17 universities, 16 institutions for theology, 7 for art and 59 *Fachhochschulen*).

In 2007, 1 979 043 students were registered in German higher education institutions. 246 369 (= 12.4%) of them were foreign students. 57 933 of them were enrolled as students of foreign nationality who gained their higher education entrance qualification at a German school (*Bildungsinländer*). China headed the ranking of countries of origin, followed by Bulgaria, Poland, and Russia. Turkey which ranked first 10 years ago reached rank 5. As a host country for foreign students, Germany ranks third after the United States and the United Kingdom.

In 2006, students had an average monthly income of 770 Euros, 27% of them had less than 600 Euros per month. In a breakdown, 52% of the monthly income came from the parents, whereas 24% were earned. Since 2005, the Länder can impose study fees on students. Some Länder made use of this option for the first time in the winter-semester of 2006/2007 by imposing up to 500 Euros per semester. No matter under what conditions, the imposition of fees is highly controversial among the political parties in Germany.

The internationalization of German higher education shows different features. The Bologna Process is supposed to lead toward a European Higher Education Area, thus enhancing its competitiveness and attractiveness to other parts of the world. In this context, a new graduation system of bachelor's and master's degrees has been adopted for implementation in all higher education institutions since 1998. In addition, a European credit transfer system (ECTS) has been introduced facilitating an increase of student mobility.

In Germany, the speed of the reform process in higher education is accelerating, which indicates that competition is becoming an integral feature both among higher education institutions and among the Länder. In general, a move from detailed input-oriented Länder control mechanisms to highly flexible, output-oriented institutional budgets can be observed. Performance indicators will be introduced, also influencing the amount of salary supplements for professors.

Ongoing Debates about Future Developments

Currently, education receives highest political priority in Germany. Although without legal competences in education, the Federal Chancellor, Angela Merkel, proclaimed "education for all" as the message for the twenty-first century, thereby referring to "welfare for all" as postulated by the father of the concept of social market economy, Ludwig Erhard, in the 1950s. In the following, some of the major issues in education are summarized.

Pre-School Education

In all Länder, the development of educational concepts and the preparation and implementation of plans can be observed which deal with the pre-school sector of day-care centers for children between 3 and 5–6 years, thereby attaching particular importance to improving the linguistic competence of children with migrant background. In about half of the Länder, those measures also include their parents. In order to organize the school entrance phase in a more flexible way, cooperative links between pre-school institutions and primary schools had to be improved. In 2004, the KMK and the Conference of Ministers of Youth adopted a framework for early education in the pre-school sector and a recommendation to strengthen and further develop the overall relationship between education, upbringing, and supervision. Furthermore, the Day-Care Expansion Act of 2004, which came into force at the beginning of 2005, provides for an expansion of the day-care for children under the age of 3 years which should be extended by 2010 in such a way that it meets the actual needs of parents and their children.

Primary Education

Since the 1990s, teaching became more pupil oriented, enhancing their self-initiatives. Furthermore, all Länder have introduced foreign language teaching mainly for grades 3 and 4, in some Länder even for grades 1 and 2. In October 2004, the KMK adopted binding educational standards for the subjects German and mathematics. These determine the competences and knowledge students should have attained by grade 4.

Secondary Education

Germany is participating in several international comparative studies of student achievement. Among them is the OECD project entitled Programme for International Student Assessment (PISA) which provides Germany with indicators concerning the knowledge, skills, and abilities of 15-year-old students in reading, mathematics, and

sciences. The survey covers three thematically overlapping cycles (PISA 2000, PISA 2003, PISA 2006). After the publication of the results of PISA 2000 which focused on reading literacy, the KMK named seven areas in which the Länder and the KMK will undertake measures in order to

- improve linguistic competence as early as pre-school education,
- strengthen the link between the pre-school and primary school sector with the aim of an early school entry,
- improve reading literary and basic understanding of concepts in mathematics and sciences in primary education,
- support more efficiently educationally disadvantaged children, especially those from families of migrant workers (cf. **Table 2**),
- develop and assure the quality of teaching based on binding educational standards and result-oriented evaluation,
- improve the performance of teachers, and
- expand the provision of all-day activities and care aiming to increase educational opportunities.

According to the 2006 OECD, PISA study education achievement of 15-year-olds in Germany has been above the OECD average in science ranking eighth position among OECD countries. Their achievement has been average in reading and mathematics. Students from families of migrant workers had made lower scores in the three subjects than native students. Given the relatively large share of foreign students (cf. **Table 2**), their performance had a large negative impact on Germany's ranking.

Higher Education

Modernization and internationalization are the key efforts in Germany. This implies, *inter alia*, a higher degree of deregulation in the Länder, thus increasing the institutional autonomy and creating financial incentives via the use of performance indicators. The responsible authorities intend to enhance the international competitiveness of higher education institutions. One example is the Excellence Initiative of the Bund and the Länder for the

promotion of science and research in German institutions of higher education funding three areas, namely research schools for the promotion of scientific talent, excellence clusters for the promotion of leading science, and future concepts for top-class research at universities (total amount until 2011: 1.9 billion Euros).

Conclusions

In sum, it can be concluded that the issue of quality assurance through educational standards and educational reporting received the necessary attention after the PISA shock. As a consequence, the KMK had adopted the necessary measures step by step in 2003 and 2004. As a result of this and for the first time, quality development in the general education of all Länder can be checked against jointly agreed criteria. In order to guarantee a central comparative review of the achievement of educational standards, the KMK founded in 2004 the Institute for Quality Development in Education (IQB) which was set up at the Humboldt University Berlin. The comprehensive strategy for educational monitoring of the KMK is indeed a completely new approach in educational planning, thus replacing input-oriented approaches by output-oriented strategies. One step in this direction is the first joint educational report of June 2006, of the Bund and the Länder on education in Germany, which has been drawn up by an independent scientific consortium. This educational report will be published every 2 years and present a set of core indicators, informing the German public about education in the course of life, that is, from early education up to continuing education taking into account the necessity of lifelong learning.

In its second report of 2008 the group observed, *inter alia*, that

- the female students are increasingly successful at all levels of the German education system – a success story which is not yet reflected on the labor market;
- the demand for higher education remains relatively low, although the proportion of students qualifying for higher education increased;
- the socioeconomic background of parents has a stronger impact on student participation in secondary and higher education than in other OECD countries.

Over the last years, a broad spectrum of reforms related to all levels of the education system has been started in Germany. Whether the reform processes will actually lead to the desired results remains an open question.

Educational Finance

The Bund, the Länder, and Kommunen spent altogether 86.7 billion Euros in 2005 (in 1995: 75.9 billion Euros), among

Table 2 Share of foreign students according to educational level, 2000 and 2005 (in percent)

	2000	2005
Pre-primary level	15.2	14.0
Primary level	11.5	11.6
Secondary level	8.5	9.5
Secondary level II and post-secondary non-tertiary level)	7.2	6.2

From Bundesministerium für Bildung und Forschung (2008). *Grund-und Strukturdaten 2007/2008.*, p. 57. Bonn, Berlin: BMBF.

Table 3 Institutions of higher education, 1995, 2000, and 2003

	1995	2000	2003
Universities	89	97	100
Teacher training	6	6	6
Theology	17	16	16
Art	46	49	52
General <i>Fachhochschulen</i>	138	155	163
Public administration <i>Fachhochschulen</i>	30	29	29
Total	326	352	366

From Bundesministerium für Bildung und Forschung (2008). *Grund-und Strukturdaten 2007/2008*, p. 149. Bonn, Berlin: BMBF.

them 50.2 billion Euros for general and vocational education and 18.4 billion Euros for higher education (9.2 billion Euros for higher education research are not included).

Compared to the majority of OECD countries, overall educational expenditure in Germany increased more slowly in recent years than total public expenditure. Therefore, its proportion decreased from 9.9% to 9.7% between 2000 and 2005, whereas the OECD average increased from 12.8% to 13.2%.

In 2005, Germany spent 5.1% of GDP at all levels of the education system (OECD average: 6.1% of GDP). Looking at the expenditure per student at different educational levels, Germany remained below the OECD average in primary education, but above the OECD average in higher education and, due to the dual system, also in secondary education. The OECD clearly postulates that additional investments in education will be important. However, efficiency and productivity in education must also be improved.

See also: An Overview of Vocational Education and Training; Higher Education: An Overview; Progress Monitoring.

Further Reading

- Authoring Group Educational Reporting (ed.) (2008). *Education in Germany 2008. An Indicator-Based Report Including an Analysis of Transitions Subsequent to Lower Secondary Education. Summary of Important Results*. Bielefeld: W. Bertelsmann.
- Bundesministerium für Bildung und Forschung (2008). *Grund-und Strukturdaten 2007/2008*. Bonn, Berlin: BMBF.
- Cortina, K. S., Baumert, J., Lechinsky, A., Mayer, K. U., and Trommer, L. (eds.) (2008). *Das Bildungswesen in der Bundesrepublik Deutschland*. Reinbek/Hamburg: Rowohlt.

Deutscher Akademischer Austauschdienst (DAAD)/German Academic Exchange Service (ed.) (2008). *Wissenschaft weltoffen. Daten und Fakten zur Internationalität von Studium und Forschung. Facts and Figures on the International Nature of Studies and Research in Germany*. Bielefeld: Bertelsmann.

Federal Ministry of Education and Research (2007). *Economic and Social Conditions of Student Life in the Federal Republic of Germany in 2006. 18th social survey of the Deutsches Studentenwerk conducted by HIS Hochschul-Informationen-System – Selected Results* – Bonn, Berlin: BMBF.

Hüfner, K., Landfried, K. (guest eds.) (2003). German Higher Education: A System under Reform. *Higher Education in Europe*, vol. 28, No. 2, pp 141–198.

Hüfner, K., Naumann, J., Köhler, H., and Pfeffer, G. (1986). *Hochkonjunktur und Flaute. Bildungspolitik in der Bundesrepublik Deutschland 1967–1980*. Stuttgart: Klett-Cotta.

Kehm, B. (1999). *Higher Education in Germany. Developments, Problems, and Perspectives*. Wittenberg and Bucharest: UNESCO-CEPES.

Max Planck Institute for Human Development and Education (1983). *Between Elite and Mass Education*. Albany, NY: State University of New York Press.

OECD (1972). *Reviews of National Policies for Education: Germany*. Paris: OECD.

OECD (2007). *PISA 6: Science Competencies for Tomorrow's World. Volume 1 Analysis*. Paris: OECD.

OECD (2008). *Education at a Glance. OECD Indicators*. Paris: OECD.

OECD Economic Surveys (2008). Germany. Vol. 2008, Issue 7. OECD Publishing.

Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (2008). *The Education System in the Federal Republic of Germany 2006*. Bonn: KMK.

Relevant Websites

- www.bmbf.de/gus – Bundesministerium für Bildung und Forschung. Portal for Fundamental and Structural Data on Education.
- www.hrk-bologna.de – Bologna Centre of the German Rectors' Conference.
- www.che.de – Centre for Higher Education Development.
- www.wissenschaft.weltoffen.de – Data Report on the International Nature of Studies and Research in Germany.
- www.bildungsbericht.de – Education in Germany 2008: An Indicator-based Report Including an Analysis of Transitions Subsequent to Lower Secondary Education.
- www.enqa.eu – European Association for Quality Assurance in Higher Education.
- www.ec.europa.eu – European Commission.
- www.eurydice.org – Eurydice, the Information Network on Education in Europe.
- www.daad.de – German Academic Exchange Service.
- www.bildungsserver.de – German Education Server.
- www.kmk.org – Kultusministerkonferenz.
- www.oecd.org – Organization for Economic Cooperation and Development.
- www.destatis.de – Statistisches Bundesamt Deutschland.
- www.wissenschaftsrat.de – German Council of Science and Humanities.

Ghana

R Palmer, Network for Policy Research, Review and Advice on Education and Training, Amman, Jordan

© 2010 Elsevier Ltd. All rights reserved.

Glossary

COTVET – Council for Technical and Vocational Education and Training.

ESP – Education Strategic Plan.

FCUBE – Free Compulsory Universal Basic Education.

JSS/JHS – Junior Secondary School (JSS), called Junior High School (JHS) since September 2007.

MoESS – Ministry of Education, Science and Sports.

MoMYE – Ministry of Manpower, Youth and Employment.

NVTI – National Vocational Training Institute.

SSS/SHS – Senior Secondary School (JSS), called Senior High School (JHS) since September 2007.

STEP – Skills Training and Entrepreneurship Promotion Program.

TVET – Technical and Vocational Education and Training.

VTI – Vocational Training Institute.

General Background

Ghana is situated in West Africa on the Gulf of Guinea. It achieved independence from Britain in 1957 and has been a self-governing republic since then.

Between independence and 2000, Ghana has seen almost 20 years of military rule, about 12 years of restricted emerging democratic practice, and some 10 years of democracy. The election of the New Patriotic Party (NPP) in December 2000 ushered in a new political context for Ghana: for the first time, a sitting government was ousted from office by means of the ballot box. Ghana has a parliamentary form of government with the president serving 4-year terms. The elections in December 2004 saw the NPP peacefully retain power. The last elections were in December 2008. Ghana is now considered an island of peace and stability in the West Africa Sub-region.

Early formal schooling in the Gold Coast (Ghana) became, for a time, a subsidiary function of the large merchant companies who operated schools usually attached to castle chaplaincies which preceded the main mission activity in the Gold Coast by nearly a 100 years. It was the Portuguese who established the first school at Elmina Castle (in the mid-sixteenth century). In 1821, provision was

made from the British Crown funds to set up a series of government schools, though these were later absorbed by missions.

It was not until the late-nineteenth century, with the rise of European missionary societies, that the development of formal educational facilities really took off. Two of the most important missions were the Wesleyan Methodist Missionary Society and the Basel Mission Society. Outside of these early formal schools, skills formation and replication among the early occupational specialists – for example, the goldsmiths, weavers, and wood-carvers – took place through traditional apprenticeship learning. While the literature makes little mention of this type of training during the colonial period, it probably remained the largest provider of skills training in the Gold Coast – as it did in Ghana in 2009.

While World War I increased the opportunities for Africans in the labor market, the expansion of primary schooling during this period soon led to concerns about unemployed school leavers. Technical and vocational education and training (TVET) as a solution to this problem of rising unemployment became popular among policy-makers. In the 1920s, Governor Guggisberg vehemently promoted school-based TVET as a means of re-orientating youth toward manual work thereby reducing unemployment. However, the unpopularity of school-based TVET led to school curricula becoming increasingly more academic during the 1930s until the period of self-governance in 1951. The development of Achimota secondary school during the 1920s is an interesting case in point. The initial intention was to create a secondary school with a strong agricultural core as part of the curriculum but, as a result of pressure from youth, parents, educated Africans, and the press, by 1951, Achimota had become a very academically orientated institution.

By 1951 there were only 23 technical institutions with some 3300 pupils – less than 1% of pupils in all types of school. Unemployment has remained a major problem in the post-1951 period. During the 1950s–1970s the government of Ghana attempted to tackle this with work-creation schemes like the Brigades and by setting up technical institutes and national vocational training institutes (NVTIs). Since the 1960s, there have been repeated attempts to reform the education system to make it more orientated toward work. The 1987 attempt to vocationalize the junior secondary level has been officially acknowledged as a failure. In September 2007, a new education reform came into effect (see **Case study 1**).

Case study 1 Ghana's 2007 Education Reform

In 2004, the Government of Ghana (GoG) came out with a *White Paper on Educational Reforms* which outlined the reforms for the education and training system which commenced in September 2007. This reform came about due to the persistent high levels of youth un- and underemployment and the belief that something must have gone wrong with the education and training system to have caused, or at least not solved, the youth employment problem. The previous education reforms of 1987 – which were intended to equip the youth with directly employable skills for the world of work through a vocationalized junior secondary school system – have been deemed a failure by many in the GoG. Moreover, the new reforms are meant to put in place a second-cycle system that caters in a better manner for the majority of youth (c.60%) who complete basic education and do not continue to formal senior-secondary education. The major elements in the reform include:

1. Universal and compulsory basic education extended to include 2 years of preschool teaching and extending basic education from 9 to 11 years.
2. At the primary level, fewer subjects are to be taught so that grounding in the basic skills (literacy, numeracy, problem-solving skills, and creative arts) would be improved.
3. The junior secondary school (JSS) system, that includes – in theory, but frequently not in practice – an element of pre-vocational and pre-technical skills training, has been discontinued. In place of the JSS will be the renamed Junior High School (JHS), with a more general, comprehensive curriculum. The government intends that the JHS should not be like the present JSS “which served as a terminal programme for most pupils. . . [but] should become the entry stage” (p. 5) to further postbasic education and training in the new diversified system of senior high schools or in a new system of structured apprenticeship.
4. Senior secondary schools will be renamed Senior High Schools, extended from 3 to 4 years, and diversified into four streams: vocational, technical, agricultural, and general education. It is the decision of the government “that the Senior High School system should be organised both as terminal education for entry into the world of work, and as a preparatory stage for entry into tertiary education” (p. 8).
5. The government intends to partner with the private sector to promote more formalized apprenticeship training programs, with government assuming full responsibility (p. 9) for the first year of the program. Related to the new education reform is the new TVET policy which is meant to complement the White Paper and to reform skills training in Ghana and to dramatically expand the formal TVET sector.
6. By 2015, all levels of education are to be staffed by professionally trained teachers.

Factors critical to the successful achievement of the proposed education reforms include: training sufficient teachers to staff all levels of education and effective decentralization of responsibility for provision and management of primary and second cycle schools to the districts. District assemblies would be responsible for providing equipment and infrastructure to all basic schools.

From GoG (Government of Ghana) (2007). *Preliminary Education Sector Performance Report June 2007*. Accra: Ministry of Education, Science and Sports; Palmer, R. (2005). *Beyond the basics: Post-basic education and training and poverty reduction in Ghana. Post-Basic Education and Training Working Paper Series No. 4*, Centre of African Studies, University of Edinburgh; and page references refer to GoG (2004b). *White Paper on the Report of the Education Reform Review Committee*. Accra: GoG.

Ghana is a predominantly rural, low-income country with a population of about 22.1 million (in 2005) which is expected to grow on an average of 2.1% a year between 2000 and 2015. Ghana had a gross national income (GNI) per capita of US\$450. The structure of Ghana's economy, with an overreliance on primary products (e.g., agriculture, timber, and gold) and a large informal economy, has changed little since independence in 1957. Less than a third of Ghanaians live below the national poverty line (poverty has fallen from 52% in 1991/92 to 28% in 2005/6 – Ghana Living Standards Survey 2005/6). However, income poverty remains widespread, particularly in the three northern regions, among nonexport food-crop farmers and those in the informal economy.

According to the Ghana Statistical Service (GSS, 2000), unemployment in Ghana stood at 6.7% in 2000 though it was more acute in urban than rural areas (11.9% vs. 4.3%). However unemployment figures in Ghana should be treated with a high degree of caution. Not only do they vary widely according to how unemployment is measured, but the huge number of working poor in low-end informal micro-enterprises means that there is a great deal of disguised unemployment or underemployment (Palmer, 2007). The 1998/99 Ghana Living Standards Survey estimated that total informal employment in Ghana was 86.3% (74.1% in urban areas and 91.9% in rural areas). Most new jobs are created in the informal economy, with formal sector employment growth largely stagnant.

Goals of the Education System

In May 2003, the Ministry of Education (MoESS) published the *Education Strategic Plan (ESP) 2003–2015* (GoG, 2003). The ESP was informed by many documents including the Education For All (EFA) Goals, the millennium development goals (MDGs), the President's Committee on the Review of Education (GoG, 2002), the Ghana Poverty Reduction Strategy (GPRS), and many reports of committees on educational reforms. The ESP is the main policy document for the education sector and sets out the government's mission statement and policy goals. (A revised ESP (2010–2020) now replaces the 2003–2015 version.)

The mission of the Ministry of Education is to provide relevant education to all Ghanaians at all levels to enable them to acquire skills that will assist them to develop their potential, to be productive, to facilitate poverty reduction, and to promote socioeconomic growth and national development.

These policy goals are set out to:

1. increase access to and participation in education and training;
2. improve quality of teaching and learning for enhanced pupil/student achievement;

3. extend and improve technical and vocational education and training;
4. promote good health and environmental sanitation in schools and institutions of higher learning;
5. strengthen and improve educational planning and management;
6. promote and extend the provision of science and technology education and training;
7. improve the quality of academic and research programs;
8. promote and extend preschool education;
9. identify and promote education programs that will assist in the prevention and management of human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS); and
10. provide girls with equal opportunities to access the full cycle of education.

Ghana's *TVET Policy Framework* (GoG, 2004a) lays out the specific policy goals of the TVET sector.

Structure and Operation of the System

Note that in September 2007, a new education reform began to be implemented (see **Case study 1**).

Key statistics on enrolments and resources in the major education sectors are provided in **Table 1**.

Preschool and Early-Childhood Care

In the 2006/7 year, the total number of kindergarten schools was 13 750 representing an increase of 2% when compared to the 2005/6 figures of 11 931 (GoG, 2007). Approximately one-third are private institutions (UNESCO, 2007). Enrolment increased from 687 643 to 1 142 784 between 2003/4 and 2006/7.

Primary Schools

The 1992 constitution reiterated the earlier educational intention, as spelt out in the Education Act of 1961, by

Table 1 Key statistics on enrolments and resources in the major education sectors (2006/7 unless otherwise stated)^a

	No. of institutions	No. of students	GER	GPI(%)	Pupil teacher ratio	% of trained teachers	% of total education spending (allocated 2007)
Kindergarten/ preschool	13 750	1 142 784	89	99.1	34:1	36	4.3
Public	10 008	n/a					
Private	3 742	n/a					
Primary	16 410	3 473 229	94	95.6	34:1	62	38
Public	12 880	n/a					
Private	3530	n/a					
Junior	9054	1 170 801	77	91.2	18:1	77	18.5
Secondary							
Public	7122						
Private	1932						
Senior	605	492 120					9
Secondary							
Public	478						
Private	127						
TVET							2.4
Public	n/a	18 005					
Private	n/a						
Teacher training (TTC)	n/a	26 025					3.4
Tertiary	n/a						
Public							20
Universities	5	84 078					
Polytechnics	n/a	29 047					
Private universities	12	n/a					
Nonformal							0.6

^aNote that these statistics refer only to institutions under the Ministry of Education, Science and Sports.

n/a data not available.

From GoG (Government of Ghana) (2007). *Preliminary Education Sector Performance Report June 2007*. Accra: Ministry of Education, Science and Sports.

introducing a program of free compulsory universal basic education (FCUBE) (launched in September 1995), making schooling in Ghana compulsory and free for children from the ages of 6 to 15. In practice, however, before the introduction of capitation grants in 2005, FCUBE was not free as students were still charged levies for sports, cultural activities, practicals, and so on. FCUBE is still not universal or compulsory in practice. There are still about 990 000 children of primary age out of school (2005) (UNESCO, 2007), and, while basic education (primary and junior secondary school (JSS)) is technically compulsory, there is no real mechanism in place to enforce the compulsory element of FCUBE.

Primary schooling in Ghana lasts 6 years. In the 2006/7 year, the total number of both public and private primary schools was 16 410, representing an increase of 7% when related to the 2005/6 figures of 15 307 (GoG, 2007). Approximately one-fifth are private institutions (GoG, 2007; UNESCO, 2007). Between 2003/4 and 2006/7 enrolments grew by 17.4% (2 957 491–3 473 229). The year 2006/7 saw large improvements in the net enrolment ratios, indicating that the number of appropriately aged children for primary has increased. The primary completion rate is 86%.

Junior Secondary Schools

The transition rate from primary to lower-secondary school is approximately 80%. Under the new education reform (**Case study 1**), the 3-year junior high schools will have a general curriculum. In 2006/7 there were 9054 schools (7122 public and 1932 private) against the 2005/6 figure of 7696 schools. Between 2003/4 and 2006/7 enrolments grew by 20% (984 111 to 1 170 801) (GoG, 2007). The lower secondary completion rate is 65%. The basic education certificate examination (BECE) is currently the only way of testing learning outcomes.

Senior Secondary Schools

The senior secondary school (SSS) level of education starts to become more exclusive, with about 36% of lower-secondary graduates gaining access to this level. The current course duration of the senior high school (SHS) is 4 years (increased from 3 years under the new education reform – see **Case study 1**).

Ghana's Poverty Reduction Strategy Paper (GoG, 2005) plans to create one good, or model, senior secondary in each district. The ESP further notes the intention of the government to expand and improve postbasic education, by rehabilitating 10% of existing senior secondary schools by 2015 and providing at least 75% of the districts with a model SHS by 2015.

In the 2006/7 year, there were 478 public senior secondary schools and 127 approved private secondary

schools. Between 2005/6 and 2006/7 enrolment increased by 28% from 384 455 to 492 120. The upper secondary completion rate of 34% indicates that there are many dropouts at this level of schooling. On completion, students receive an SSS certificate.

The policy-related issues in the secondary education subsector are: equitable distribution of well-equipped secondary schools; well-trained and motivated teachers; and inability to absorb the majority of pupils who complete the compulsory basic-education program (GoG, 2005).

Technical and Vocational Education and Training

In Ghana, TVET takes place under no less than nine different ministries as well as in the private sector. TVET is delivered in schools, in institution-based, vocational-skill centers, institutes, and colleges and in the informal economy in traditional apprenticeship.

Most formal public TVET is delivered by the MoESS and the Ministry of Manpower, Youth, and Employment (MoMYE). Formal TVET in Ghana occurs at three levels. Polytechnics, the highest-level training institutions beneath universities, are located in regional capitals ($n = 10$) and are followed by the technical training institutes ($n = 23$), both under the MoESS. Public sector vocational training institutes (VTIs) under the MoMYE number about 111 (national vocational training institutes – NVTIs (30), integrated community centers for employable skills – ICCES (59), and other VTIs (22)). Other ministries have vocationally related institutes, for example: the Ministry of Lands, Environment and Forestry has 15 regional technology transfer units; and the Ministry of Local Government and Rural Development has about 22 VTIs (Palmer, 2007).

In addition to public-skills provision, there is a huge range of private for-profit and nonprofit provision. There are over 450 private training institutions, many of them church-based VTIs with some 10 000 trainees. Some of these, for example, the Catholic Don Bosco training institutes in Tema/Ashiaman and Sunyani, are probably unequalled in quality anywhere in Ghana, and they have a history of training children from poorer families. International nongovernmental organization (NGO) provision is usually concerned with meeting the training needs of poorer children, while much private for-profit provision is beyond the means of poorer parents.

Much of Ghana's public (and, to a lesser degree, private) formal TVET system is characterized by low-quality, obsolete equipment, outdated curricula, and a lack of relationship with industry and employment. Moreover, most public and private VTIs are largely out of reach of the poorest individuals. The formal TVET sector in Ghana, as in other developing countries, is largely orientated toward formal rather than informal employment.

The numbers that formal public TVET reach are low. In 2006/7, enrolment targets set for institutions under the MoESS were not achieved due to lack of interest in the selection of technical vocational causes by lower secondary graduates and their parent/guardians; enrolments were only 18 005 (Table 1) for MoESS TVET institutions.

Altogether there are approximately 50 000 students in different formal TVET institutions (public and private) nationwide at all these levels – in all ministries. Entry requirements for most of the institutes in the TVET system is a (BECE) or middle school leaving certificate (MSLC), and the training period is usually 3 years. Some, like the Integrated Community Centre for Employable Skills (ICCES), do not have any specific entry requirements and cater to both school dropouts and JSS leavers.

Another type of skills-delivery mechanism has been the short-duration modality. In the mid-1990s, the government and the World Bank implemented the Vocational Skills and Informal Sector Support Project which focused on skills upgrading for mastercraftsmen and traditional apprentices. More recently (2003/5) the Skills Training and Entrepreneurship Programme (STEP) offered 3–12 months of training to some 27 000 youth with the intention that the STEP graduates would immediately enter self-employment. The STEP was a highly politicized program that largely had disappointing outcomes (Palmer, 2007).

Traditional apprenticeship training remains the largest provider of skills and is responsible for some 80–90% of all basic skills training in Ghana, compared to 5–10% from public training institutions and 10–15% from NGO for-profit and nonprofit providers. The general features of the traditional apprenticeship training modality in Ghana are well known and similar to those described in other countries in West Africa (e.g., see Johanson and Adams, 2004; Palmer, 2007). Ambitious plans are being put in place in 2007/8 to move toward a more regulated, or formalized, informal apprenticeship system (Palmer, 2009).

Formal TVET is mostly undertaken by males. For example, only 19% of all trainees in formal TVET institutions under the MoESS are female. In traditional apprenticeships, training is highly segregated along gender lines.

Higher Education

Both the GPRS and the ESP call for an improvement in existing tertiary education. By 2015, the objective is to have tertiary education available for at least 15% of the postsecondary cohort. Between 2001/2 and 2006/7, Ghana's tertiary education system saw an increase in enrolments of approximately 66%, from about 67 000 to approximately 113 000. This growing demand for

tertiary-level education has led to a growth in private tertiary institutions, predominantly in Accra.

Private institutions charge fees while government provides support to the institutions in other ways such as access to Teaching and Learning Innovation Fund (TALIF) and mass transport. Since the establishment of the Ghana Education Trust Fund in 2002, students pursuing accredited programs are eligible for the same student loans as their colleagues in public institutions.

At the tertiary level, the critical policy issues relate to: inadequate physical infrastructure to absorb the growing number of young adults who seek admission to tertiary institutions; and insufficient numbers of qualified and well-motivated academic staff (GoG, 2005).

Adult and Nonformal Education

The Non-Formal Education Department (NFED) is primarily responsible for the development and delivery of education programs for out-of-school youth and adults. At present, its program focus is on the coordination and implementation of the National Functional Literacy Project (NFLP), which aims to increase the basic literacy skills for adults and youth, nationally, with particular focus on females and the rural poor. The key ESP target for the nonformal education (NFE) sector is to achieve a decrease in the national rate of adult illiteracy and to improve the gender parity of literate adults. By 2004, 696 177 learners had been recruited into the NFLP. In 2005, 300 000 new learners were recruited. Over 700 000 inputs are distributed to learners – including pencils, exercise books, primers, supplementary readers, etc., with incentive packages provided for 7500 facilitators annually. The target for adult literacy was 64.5% by 2007 (GoG, 2007).

Teaching Profession

The teacher-education division has the responsibility for the training of preservice teachers and the provision of ongoing inservice training (INSET) for serving teachers. The preservice teacher-education program is of 3 years duration leading to a diploma in basic education. In October 2004, the upgrading of teacher training college (TTC) programs to diploma status commenced. Institutional upgrading of the TTCs is in process to ensure the attainment of tertiary status by 2010 (GoG, 2007).

Administrative and Supervisory Structures

Public schools and technical institutes operate under the direct responsibility of the MoESS, while other ministries (principally MoMYE) are responsible for VTIs.

The MoESS is embarking on decentralization to institutionalize the district educational directorate as the primary management unit directing and reporting on education-service delivery.

A growing number of private universities have emerged in Ghana, in part due to the inability of public universities to satisfy the growing social demand for higher education, but also as a result of the increasing demand for employment-oriented courses. The Education Act of 1960 empowered the Minister of Education to approve the establishment of private tertiary institutions, close institutions, and make regulations regarding the conduct of the affairs of institutions. Following the 1987 education reform, the National Accreditation Board now performs these functions. The work of the National Accreditation Board among others will ensure quality assurance in the private institutions to meet acceptable standards.

Ghana has recently set up a Council for TVET (COTVET) which will coordinate and oversee all aspects of TVET, across ministries and the broad spectrum of formal, informal, and nonformal education and training. In November 2007, the board of the new COTVET was inaugurated, over a year after the COTVET Bill was passed in Parliament (July 2006). Three committees are to be set up under COTVET which will undertake the functions of the council:

1. National TVET Qualifications Committee – the main regulatory authority for the National Qualifications Framework.
2. Training Quality Assurance Committee – ensures training providers and qualification awarding agencies maintain satisfactory standards in the delivery of training and the award of qualifications.
3. Industry Training Advisory Committee – to develop national occupational standards or knowledge, skills, and work-performance standards for the definition and issuance of qualifications.

While much is expected of COTVET, there are still concerns that it will not differ greatly from its predecessor – the National Coordinating Committee on Technical and Vocational Education and Training (NACVET) (Case study 2).

Under the education reform, several new structures are being put in place to enhance the effective implementation of the reform (e.g., COTVET).

Educational Financing

In 2006, actual expenditure on education was 6% of gross domestic product (GDP) (same as the target suggested by the United Nations Educational Scientific and Cultural organization (UNESCO) and the African Union).

Case study 2 NACVET and COTVET in Ghana: A new era of coordination or new wine in old wineskins?

In 1990, an attempt was made to resolve the interdepartmental rivalries and bureaucratic constraints that affected the management of training; the National Coordinating Committee on Technical and Vocational Education and Training (NACVET) was set up to coordinate a national skills development system (in particular between providers of skills from the ministries of education and employment). However, even a decade after it had been set up, NACVET had failed to develop a national skills-development policy, continued to suffer from capacity problems and ongoing tensions between the two parent Ministries, and remained inhibited by a lack of any legislative backing.

It is hoped that NACVET's successor, the Council for Technical and Vocational Education and Training (COTVET), that does have legal backing, will be more effective. Parliament passed on 27 July 2006 a Bill to establish the COTVET, which has been almost 10 years in the making. Its remit is to formulate national policies for skills development across the broad spectrum of formal, informal, and nonformal education. With the COTVET in place, the expectation is that a comprehensive demand-driven system can be established and financed. However, there remain some serious challenges to the successful functioning and implementation of COTVET; not least those related to cooperation between, and the multiple (often conflicting) agendas, interests, and politics within each Ministry. Alongside the COTVET is a new TVET Policy Framework which aims to reform the system so that it will be demand driven and competency based. However, the financing mechanisms are yet to be resolved.

From Palmer, R. (2007). *Skills Development, the Enabling Environment and Informal Micro-Enterprise in Ghana*. Doctoral Thesis, Edinburgh: Centre of African Studies, University of Edinburgh.

There are serious questions to be asked about how the expanded vision of second-cycle education and training – as envisaged in the ESP and the new education reform – will be funded. There is concern among some of the donors that the rapid expansion of postbasic levels of education in Ghana risk undermining the push for the education MDG of universal primary education (UPE) by 2015 (cf. Palmer, 2005).

Looking at the total resource envelope for education in Ghana (Table 2), we see that resources to primary education, as a percentage of total resources, has decreased from 40% in 2003 to 27.6% in 2006. However, there has been an actual increase in resources to the primary level over the same time frame. The overall percentage of total resources for the JSS level has decreased from 22% in 2003 to 16.8% in 2006. At the SSS level, resource allocation 2003/6 remained similar in 2003 and 2006 at around 15% (in 2004 and 2005 this figure was about 20%). The most significant increases in resources, in both percentage and actual terms, can be seen at the tertiary levels where resource allocation has increased from 14% to 22.5% between 2003 and 2006.

The budget allocation for 2007 is indicative of policy shifts (Table 2); primary education funding is set to

Table 2 Expenditure by level of education as a % of total resource envelope 2003/6

	2003		2004		2005		2006				2007	
Level	(Expenditure)		(Expenditure)		(Expenditure)		(Allocation)		(Expenditure)		(Allocation)	
	¢ (million)	%	¢ (million)	%	¢ (million)	%	¢ (million)	%	¢ (million)	%	¢ (million)	%
Preschool	98 862	2%	231 761.64	4.00%	250 299	3.40%	360 143	4.10%	371 448	3.90%	483 320	4.30%
Primary	1 635 339	40%	1 830 916.96	31.60%	2 201 159	29.90%	2 696 678	30.70%	2 626 272	27.60%	4 220 284	37.90%
JSS	910 353	22%	927 046.56	16.00%	1 310 389	17.80%	1 300 027	14.80%	1 599 216	16.80%	2 062 239	18.50%
SSS	630 244	15%	1 153 014.16	19.90%	1 531 241	20.80%	1 862 201	21.20%	1 503 828	15.80%	1 017 244	9.10%
TVET	45 312	1%	63 734.45	1.10%	88 341	1.20%	61 488	0.70%	85 999	0.90%	266 213	2.40%
SPED	16 477	0%	23 176.16	0.40%	29 447	0.40%	105 408	1.20%	38 356	0.40%	59 663	0.50%
NFED	37 073	1%	92 704.66	1.60%	139 873	1.90%	96 624	1.10%	67 369	0.70%	68 419	0.60%
Teacher education	164 770	4%	214 379.52	3.70%	287 108	3.90%	351 359	4.00%	331 190	3.50%	381 888	3.40%
Tertiary	572 575	14%	1 216 749	21.00%	1 442 900	19.60%	1 827 066	20.80%	2 145 645	22.50%	2 242 499	20.10%
Management and subvented	4119	0%	28 970.21	0.50%	73 617	1.00%	105 408	1.20%	734 384	7.70%	333 421	3.00%
HIV-AIDS	4119	0%	11 588.08	0.20%	7362	0.10%	17 568	0.20%	24 743	0.30%		0.00%
Total	4 119 243	100%	5 794 041	100.00%	7 361 737	100%	8 783 969	100%	9 528 450	100.00%	11 135 189	100%

From GoG (Government of Ghana) (2007). *Preliminary Education Sector Performance Report June 2007*. Accra: Ministry of Education, Science and Sports.

increase from 28% to 38% in 2006/7; lower secondary (JSS) expenditure is set to remain constant; upper secondary (SSS) expenditure is set to decline from 16% to 9%; TVET (under MoESS) will increase from 1% to 2.4%; and tertiary budget allocation is set to remain around 20%.

The MoESS receives a substantial amount of funding from sources other than Government of Ghana (GoG). In 2006/7, approximately 34% of the total resource envelope for the MoESS is provided by nonstatutory and external sources (Development Partners, Ghana Education Trust fund, Heavily Indebted Poor Countries Initiative, District Assembly Common Fund). In addition, the monetary contribution made by NGOs, community-based organizations (CBOs), etc. to the education sector (e.g., the building of schools) is not accounted for in the total resource envelope at all.

The proposals set out in the 2004 White Paper on Educational Reforms (GoG, 2004b) including increasing the duration of SSS from 3 to 4 years and diversifying this level have huge cost implications. These proposals worry development partners in Ghana who see them as potentially taking money away from basic, and especially primary, education (Palmer, 2005). Nonetheless the GoG appears to be very keen to push for this home-grown education strategy in spite of donor pressure to shift their thinking. But given financial constraints, it is not clear how the senior secondary level can be expanded significantly and quality ensured at this level, as well as quality improvements at the basic level achieved so that it does not deteriorate further. This leads to the question: What financing mechanisms are most appropriate for funding postbasic education and training expansion?

Performance Monitoring, Evaluation, and Research

The Planning, Budgeting, Monitoring and Evaluation (PBME) department of the MoESS is responsible for performance monitoring and evaluation. The National Education Sector Annual Review (NESAR), instituted with the implementation of the ESP 2003–15, provides the opportunity for all sector stakeholders to participate in the review of sector performance annually. The process, from its inception, has taken place at the national level with representatives from regional and district offices participating. The establishment of the NESAR has led to tremendous improvement in education delivery in the country. This is a collaborative approach to ensure the pooling together of resources and the harmonization of programs and activities for the realization of the goals and objectives of the education sector. Through the involvement of all stakeholders in the review, the NESAR enhances accountability and transparency within the sector; a key to the successful implementation of the

ESP, where all stakeholders in the sector work together under the overall lead of the MoESS. At the NESAR 2006, the recommendation was made for the conduct of Regional Education Sector Annual Reviews (RESAR) as a prelude to the organization of NESAR. The regional reviews in all the ten regions took place in the first week of May 2007 (GoG, 2007).

The majority of all externally funded initiatives in Ghana's education and training system have built in monitoring and evaluation components. However, public formal TVET in Ghana has very weak monitoring and evaluation dimensions; in many cases, evaluation is completely absent.

Much educational research is financed externally. For example, the UK Department for International Development is currently financing three research programs examining education access, quality, and outcomes in Ghana.

Major Changes and Issues since the 1990s

The 1992 constitution reiterated the earlier educational intention, as spelt out in the Education Act of 1961, by introducing the FCUBE.

With regard to the MDGs, Ghana is judged to be on track to achieve MDG 2 (UPE) by 2015 (and expects to achieve this by 2010), and progress toward UPE received a boost in 2005 with the introduction of capitation grants and the associated removal of all fees and levies on basic education (effective since the start of the 2005/6 academic year); this caused, for example, enrolment of 6–11-year-olds in primary schools to jump by 400 000 in 2005/6, an increase of 14% on the previous year. This increase in enrolments could, by 2015 result in a mini-surge of JSS graduates entering the labor market and also in those seeking postbasic education and training. While capitation grants have undoubtedly increased enrolments, questions still remain concerning the extent to which quality can be improved, or even maintained, given the large increase in student numbers. Indeed, a concern for government now is the need to train more teachers and provide more schooling infrastructure in order to maintain quality. Even with the capitation grants, there are still likely to be large numbers of children and youth out of school.

Since 2002, the NPP government has followed a dual-track education policymaking process. In parallel to the ESP, an education strategy document was drawn up, which had a focus on basic education; the government in 2002 commissioned a panel of academics and other educationalists to examine the education system in Ghana. Some of the main recommendations of this commission (GoG, 2002) were adopted by government in the 2004 *White Paper on The Report of The Education Reform Review Committee* (GoG, 2004b). The White Paper, among other things, proposes to expand postbasic education and to

vocationalize the secondary level (see **Case study 1** on the education reforms). This turn of events, toward post-basic education and training, is obviously one met with some concern, particularly among donors who are focused on the MDGs, like Department for International Development (DFID). DFID Ghana is well aware of this apparent move away from basic education toward postbasic education, and has voiced its concerns to the government and to other donors. A 2005 DFID Ghana report on the education sector notes that the new White Paper is strikingly different in its priorities.

Since 2001, there has been a renewed government focus on skills development and its relationship with combating unemployment. This renewed focus, driven by democratic, poverty-reduction, and economic imperatives, has been reflected in a number of government documents; for example, the 2003/5 GPRS I, the 2006/9 GPRS II, and the 2004 Draft TVET Policy Framework for Ghana. TVET, delivered through public and private schools, VTIs, and traditional apprenticeships, continues to be seen as an important link to work. Between 2003 and 2005 the GoG funded short-duration skills training through the STEP – a direct response to government concerns regarding unemployment. The 2004 education White Paper and the 2004 TVET policy framework have set a challenging agenda for skills reform in Ghana.

Bibliography

- GoG (2002). Meeting the challenges of education in the twenty first century. *Report of the President's Committee on the Review of Education Reforms in Ghana*. Accra: GoG.
- GoG (2003). *Education Strategic Plan 2003 to 2015: Volume 1. Policies, Targets and Strategies*. Accra: MoE.
- GoG (2004a). *Draft TVET Policy Framework for Ghana*. Developed for Ministry of Education Youth and Sports, January 2004, Accra: MoEYS.
- GoG (2004b). *White Paper on the Report of the Education Reform Review Committee*. Accra: GoG.
- GoG (2005). Growth and poverty reduction strategy (GPRS II). The coordinated programme for the economic and social development of Ghana (2006–2009). *Final Draft, September 2005*. Accra: National Development Planning Commission.
- GoG (Government of Ghana) (2007). *Preliminary Education Sector Performance Report June 2007*. Accra: Ministry of Education, Science and Sports.
- GSS (Ghana Statistical Service) (2000). *Ghana Living Standards Survey: Report of the Fourth Round (GLSS 4)*. Accra: Ghana Statistical Service.
- Johanson, R. and Adams, A. (2004). *Skills Development in Sub-Saharan Africa*. Washington, DC: World Bank.
- Palmer, R. (2005). Beyond the basics: Post-basic education and training and poverty reduction in Ghana. *Post-Basic Education and Training Working Paper Series No. 4*. Centre of African Studies, University of Edinburgh. http://www.cas.ed.ac.uk/research_projects/beyond_basics (accessed August 2009).
- Palmer, R. (2007). *Skills Development, the Enabling Environment and Informal Micro-Enterprise in Ghana*. Doctoral Thesis, Edinburgh: Centre of African Studies, University of Edinburgh.
- Palmer, R. (2009). Formalising the Informal: Ghana's New Apprenticeship System. *Journal of Vocational Education and Training* 61(1): 67–83.
- UNESCO (2007). Education for all by 2015. Will we make it? *EFA Global Monitoring Report 2008*. Paris: UNESCO.
- DFID (2007). *Ghana Fact Sheet (October 2007)*. London: DFID.
- Palmer, R. (2006). Beyond the basics: The need for a balanced education and training system in developing countries. *Journal of Education in International Development* 2(1). <http://www.equip123.net/JEID/articles/2/BeyondBasics.pdf> (accessed August 2009).
- Palmer, R. (2007). Education, Training and Labour Market Outcomes in Ghana: A Review of the Evidence. *Improving the Outcomes of Education for Pro-Poor Development, RECoup Working Paper 9*. Paper for a DFID funded research project. Edinburgh: University of Edinburgh.
- Palmer, R. (2007). Skills for work: From skills development to decent livelihoods in Ghana's rural informal economy. *Special Issue: Beyond Basic Education and towards an Expanded Vision of Education for Poverty Reduction and Growth? International Journal of Educational Development* 27(4).
- Palmer, R. (2009). Skills Development, Employment and Sustained Growth in Ghana: Sustainability Challenges. *International Journal of Educational Development* 29(2), 133–139.
- Palmer, R., Wedgwood, R., Hayman, R., King, K., and Thin, N. (2007). Educating out of poverty? *A Synthesis Report on Ghana, India, Kenya, Rwanda, Tanzania and South Africa. Researching the Issues Series, No. 70*. DFID: London. <http://www.dfid.gov.uk/documents/publications/educating-out-poverty-70.pdf> (accessed August 2009).

Further Reading

Relevant Website

www.norrag.org – NORRAG: Network for Policy Research, Review and Advice on Education and Training.

Greece

K Toura and A Karagiannis, Greek Ministry of National Education and Religious Affairs, Athens, Greece

© 2010 Elsevier Ltd. All rights reserved.

Glossary

AEI – Universities
APOLITIRIO – School-leaving certificate obtained when a pupil successfully completes primary or secondary education
DEMOTICO SCHOLEIO – Primary School
DEPPS – Cross-Thematic Curriculum Framework for Compulsory Education
DOATAP – National Academic Recognition and Information Center
ECTS – European Credit Transfer and Accumulation System
EPAGGELMATIKES SCHOLES – Upper Secondary Vocational Training Schools
EPAGGELMATIKO LYKEIO – Upper Secondary Vocational School
ESEEK – National System of Vocational Education and Training
GENIKO LYKEIO – General Upper Secondary School
GSAE – General Secretariat of Lifelong Learning
GYMNASIO – Lower Secondary School
HOU – Hellenic Open University
ICT – Information and Communication Technology
IDEKE – Institute of Continuing Adult Education
IEK – Vocational Training Institutes
IHU – International Hellenic University
KEE – Education Research Centre of Greece
LLL – Lifelong Learning
LYKEIO – Upper Secondary School
NIPIAGOGEO – Kindergarten
NQAA – National Quality Assurance Agency
OEEK – Organisation for Vocational Education and Training
OLOIMERA NIPIAGOGEOIA – All-day Kindergartens
OLOIMERA SCHOLEIA – All-day Schools
PANEPISTIMIA – Universities
PEK – Regional Training Centers
PI – Pedagogical Institute
SCHOLEIO – School
SCHOLES – Training Schools
SDE – Second Chance Schools
TEI – Technological Education Institutes
YPEPTH – Ministry of National Education and Religious Affairs

General Background

Greece is located in the south-eastern edge of Europe and covers an area of 131 957 km². According to the data from the last census in 2001 provided by the National Statistical Service of Greece, its population is now 10 934 080. Today, the majority of the population lives in large cities; more specifically, in Athens, Thessaloniki, or Patras.

An interesting aspect of the demographic composition of Greece during the recent years is the large immigration influx mainly from the Balkan and Eastern European countries, which started in the beginning of the 1990s. According to the 2001 census, immigrants comprise 761 813 of the total number of the country's inhabitants.

Greece is a presidential parliamentary republic. The Greek Constitution was last revised in 2001. According to it, the legislative function is primarily exercised by the Parliament and the President of the Republic, while the executive function is exercised by the President of the Republic and the government which must also enjoy the Parliament's confidence.

In accordance with the Greek Constitution, the prevailing religion in Greece is that of the Eastern Orthodox Church of Christ. The official language of the country is Greek. It is used throughout the Greek territory and taught at all levels of education.

Education policy in the Greek Republic is mainly implemented by the Ministry of National Education and Religious Affairs (YPEPTH), which is assisted in its work by several supervised bodies with specific goals, jurisdictions, and responsibilities.

Goals of the Education System

According to the Greek Constitution, education is one of the basic missions of the state and every Greek citizen is entitled to it free of charge at all levels. It aims at the moral, intellectual, physical, and professional education of the Greeks, the development of national and religious awareness, as well as the development of free and responsible citizens. It should be noted that foreign and repatriated students (migrants/refugees) enjoy the right to free education exactly as natives do. Every child living in Greece is entitled to education regardless of their parents'/guardians' legal status in the country. Therefore, minor aliens who

reside on Greek territory are entitled to a 10-year minimum compulsory education, just like their Greek peers.

With regard to primary and secondary education, it aims at contributing to the complete, harmonious, and balanced development of the intellectual, psychological, and physical potential of the pupils, so that regardless of their gender or origin they may become integral personalities and live in harmony.

The national priorities with regard to educational policies among others include:

- ensuring high-quality education accessible to all;
- combating educational exclusion;
- catering for every child's specific educational needs, while respecting the child's diversity;
- developing the pupil's abilities;
- providing general education;
- making pupils aware of human-rights-related issues;
- fostering the principles of democratic citizenship;
- raising the pupils' environmental awareness;
- preparing them for the use of information and communication technology (ICT);
- connecting education with the labor market;
- strengthening vocational education and training; and
- promoting adult education and lifelong learning (LLL).

Structure and Operation of the Education System

Education is provided at three levels:

1. Primary education, which includes kindergarten (*nipia-gogeo*) and primary school (*demotiko scholeio*);
2. Secondary education, which includes lower-secondary school (*gymnasio*), general upper-secondary school (*geniko lykeio*), upper-secondary vocational school (*epaggelmatiko lykeio*), and upper-secondary vocational training schools (*epaggelmatikes scholes*);
3. Higher education, which includes universities (*panepistimia*, AEs) and technological education institutes (*technologika ekpaideftika idrymata*, TEIs), the school of fine arts as well as the Hellenic Open University (HOU) and the International University of Greece.

According to the Greek legislation, the 10-year pre-school, primary, and lower-secondary education are compulsory. Preschool education lasts for a year and primary education (*demotiko scholeio*) lasts for 6 years. Pupils who have successfully attended primary school can register in lower-secondary school (*gymnasio*), which lasts for 3 years.

Attendance in the upper-secondary schools is not compulsory and the duration of studies in it is from 2 to 3 years, depending on the type of school.

The training provided by the vocational training institutes (*institutouta epaggelmatikis katartisis*, IEK) is formal and falls into the category of post-secondary nontertiary education.

In state primary and secondary schools, attendance is free of charge and textbooks are distributed free of charge by the state. On successful completion of primary and lower or upper-secondary education, students receive a certificate, the so-called *Apolytirio*.

Preschool Education

Preschool education, which is part of primary education, is provided in public or private kindergarten (*nipiagogeio*). Children can attend it from the age of 4 and it lasts for 2 years. The second year is compulsory for young children who have completed their fifth year of age by 31st December of the year of enrolment. Most kindergartens (*nipiagogeia*) are run by the state. Attendance in state kindergartens is provided free of charge. All-day kindergartens (*oloimera nipiagogeia*) operate on an extended timetable engaging children in creative activities for at least 8 h per day.

Primary Education

Primary education is provided in public or private primary schools. Attendance in primary school (*demotiko scholeio*) is compulsory and lasts for 6 years. It includes six grades and children are admitted at the age of 6. Public education and schoolbooks are provided free of charge. All-day schools (*oloimera scholeia*) operate in parallel with the ordinary primary schools and have an extended timetable and an enriched curriculum. A serious effort is being made for all primary schools to become all-day schools.

Public primary education schools branch out into experimental, intercultural, and special education ones, apart from the ordinary and all-day schools mentioned above.

Secondary Education

Secondary education in Greece is provided at two levels: (1) compulsory (lower-secondary education) and (2) non-compulsory (upper-secondary education).

Compulsory secondary education is provided in lower-secondary schools (*gymnasia*). Attendance lasts for 3 years and the age cohort of pupils is 12–15 years. Upper-secondary education involves two types of schools: general upper-secondary school (*geniko lykeio*) and upper-secondary vocational school (*epaggelmatiko lykeio*). Attendance in both types of secondary schools lasts for 3 years. There are also upper-secondary vocational training schools (*epaggelmatikes scholes*) that operate during the day and attendance there lasts for 2 years.

Furthermore, alternative structures of schools provided by the state include experimental, intercultural, minority, ecclesiastic, arts, sports, and musical lower-secondary schools and upper-secondary schools, along with special education ones.

In state secondary education schools, attendance is free of charge and textbooks are distributed free of charge.

Postsecondary Nontertiary Education/ Vocational Training

The Ministry of National Education and Religious Affairs supervises formal vocational education and training through the Organisation for Vocational Education and Training (OEEK, a legal entity under public law). This organization is independent in administrative and financial terms.

Its objectives are:

- the realization of the goals of the National System of Vocational Education and Training (ESEK);
- the organization and operation of public IEK; and
- the supervision and control of private IEK.

The work of the OEEK mainly focuses on:

- providing any type of vocational training, initial or continuing;
- putting forward proposals to the National Ministry of Education and Religious Affairs for policymaking and programming of vocational education and training;
- drawing up the curricula for both public and private IEK;
- determining the vocational rights at all levels of vocational education and training, in cooperation with the competent ministries and other social partners;
- recognizing the certificates or diplomas that are granted by other public bodies for vocational education and training, certifying the equivalence of certificates or diplomas obtained abroad, and providing information concerning the recognition of the professional rights and of the certificates/diplomas, as well as concerning other terms of access to legislatively guaranteed vocations.

The IEK fall under postsecondary nontertiary education. They provide formal initial vocational education and training as well as the possibility of obtaining a certificate or diploma of vocational training equivalent to that of upper-secondary vocational schools, which equips IEK graduates with professional rights, thus enabling them to access the labor market in both the public and private sector. These institutes accept both lower- and upper-secondary school graduates depending on the specializations they offer.

Today, there are 114 public IEK, of which 100 are independent from an administrative point of view, while 14 operate as branches of independent ones. Alongside the public IEK, private IEK were established for the first time in 1993. There are currently 52 private IEK, whose programs of study and specializations are the same as those of the state IEK.

Tertiary Education

According to the Greek Constitution, tertiary education falls exclusively under the jurisdiction of the state and is

provided free of charge at the undergraduate level. Tertiary education institutions are fully self-governed legal entities of public law and are supervised by the Ministry of National Education and Religious Affairs. Admission of students to tertiary education institutions depends on their performance in the pan-Hellenic university entrance exams which are held at the end of the upper-secondary school on the basis of a *numerus clausus* system.

Tertiary education consists of:

1. the university sector which includes universities, technical universities, and the higher school of fine arts; and
2. the technological sector which includes technological education institutions.

The HOU provides distance-learning courses (at undergraduate and postgraduate level).

The International Hellenic University constitutes an independent and fully self-governed higher education institution. It is a legal entity of public law supervised by the state. It runs study programs at undergraduate and postgraduate level taught in English to foreigners interested in studying in Greece.

In terms of funding, tertiary education institutions are mainly funded by the state. According to the Organization for Economic Cooperation and Development (OECD) data, Greece is one of the countries with the highest public funding of tertiary education institutions in Europe, reaching over 98%.

One of the most important recent reforms that has contributed greatly toward the autonomy and the qualitative improvement of tertiary education institutions is the drawing up and submission to the ministerial services of a 4-year academic development program by all tertiary education institutions. In this way, they will be able to define their mid-term and long-term goals and plan their strategy on the basis of their profile, size, particular needs, and special characteristics. Therefore, planning and decision making can become more flexible and effective.

The 4-year academic-development programs are related to issues such as the particular aims of each academic unit; the designation, the programming, and the measures for the development and support of the educational and research activities and other offered services; the infrastructure and equipment development; the optimal investment of higher education institutions' property and other financial sources; and the number of new students that can be admitted every year which will be agreed upon for the first time in collaboration with the Ministry.

Adult Education

The General Secretariat of Life Long Learning, supervised by the Greek Ministry of National Education and Religious Affairs, undertakes the planning, coordination,

and implementation of LLL at a national level. Under its competence fall the:

- prefectural committees of adult education;
- vocational training centers;
- the center of distance lifelong education and training for adults; and
- education in correctional institutions.

The GSAE is assisted by the Institute of Continuing Adult Education (IDEKE) that serves the purpose of providing technological and scientific support to the programs of the General Secretariat and of realizing activities concerning LLL. Some of the most characteristic lifelong education structures running with the support of the IDEKE are:

- adult education centers,
- parent schools, and
- second-chance schools.

Second-chance schools (*scholeia defteris efkerias* (SDE)) were established in 1997 and offer formal education services. They are attended by adults who have not completed compulsory education and who, after successful attendance, are granted a certificate of study equivalent to the certificate of study of primary or lower-secondary school.

Access to formal adult education can also be gained through attendance of evening lower- and upper-secondary schools and evening upper-secondary vocational schools. These schools belong to secondary education both legislatively and administratively. Any working person interested, can take these classes irrespective of their age. By law, there is a special quota allowing graduates of evening schools – of all types – to be admitted to tertiary education.

Recent reforms in LLL enable higher education institutions to establish lifelong learning institutes and offer lifelong learning programs to graduates and employees.

Finally, the HOU contributes considerably in the field of adult education as well through its undergraduate and postgraduate distance-learning courses.

The Teaching Profession

Teachers of both primary and secondary education complete at least 4-year university studies and they follow preservice training upon appointment. All teachers in the public sector are civil servants. In 2006–07 there were 11 846 teachers in kindergartens, 70 827 in primary schools, and 80 250 in secondary schools. As for tertiary education, 15 761 and 13 102 teaching staff were employed in universities and technological institutions respectively.

The training of teachers runs on both a preservice and in-service basis, and usually takes place at the regional training centers (PEK). The training of teachers is implemented by the Organisation of Teacher Training (O.EPEK)

and supervised by the Ministry of National Education and Religious Affairs in cooperation mainly with the Pedagogical Institute (PI) as well as the Ministry.

The PI conducts a significant number of teacher-training programs. In particular, 8000 newly appointed teachers were trained during 2005–07. The training programs usually concern the new school textbooks, the cross-curricular approach, the teaching of second foreign languages in primary education, modern teaching approaches in the use of ICT in education, classroom management, school administration, as well as European integration issues.

Administrative and Supervisory Structure

The general organization and administration of the education system is as follows:

- Ministry of National Education and Religious Affairs;
- Primary and Secondary Education Regional Directorates;
- Primary and Secondary Education Directorates and Offices at Prefectural Regional Level; and
- school units.

The Ministry of Education and Religious Affairs is responsible for the administration of all schools in the country. The administration and supervision is carried out through the central and regional services as well as through advisory and expert councils that have been created and function in the Ministry's central and regional services.

Tertiary education institutions are self-administered legal entities under public law and the Ministry supervises and monitors the legality of their actions and decisions through the Ministry's central services.

Educational Finance

For the years 2007–13, education is financed by:

- national resources through the regular budget of the state as well as the public investments budget. In Greece, public expenditure accounts for 3.9% of the GDP, with the larger share thereof being allocated to tertiary education (approximately 1.26% of the gross domestic product (GDP) for 2004, i.e., higher than the average European rate for the same period);
- the European Social Fund (ESF), which co-finances the operational program for education and LLL; and
- the European Regional Development Fund (ERDF), which finances all regional operational programs for the construction and maintenance of buildings, for equipment at all levels of education, and also the programs for initial vocational training and LLL.

Performance Monitoring, Evaluation, and Research

Law 2986/2002 lays out particular provisions with regard to teacher evaluation. The relevant measures/actions are implemented transversally and horizontally. The aim is to diffuse a climate of continuous and transparent accountability, enhance the quality of education services, and sustain the constructive efforts of all those involved in the educational process, whether teaching practitioners or administrators.

Teacher self-evaluation is supplemented by a process of assessment/evaluation conducted by the school head and the school counselors. The evaluation and training department of the PI is directly involved in the entire process.

At the institutional level, the PI is entrusted with the comprehensive evaluation of school units and teaching practitioners on a grand scale whereas the Centre for Educational Research (KEE) is responsible for the development and implementation of indicators and benchmarks monitoring performance and learning outcomes at both the school unit and the systemic levels. It also collects and processes reports prepared by the regional support and educational planning centers along with the school unit self-evaluation reports.

With regard to tertiary education, during the period 2004–07, new laws came into effect whose main priority was the establishment of the National Quality Assurance Agency (NQAA). Among them, the Law on Quality Assurance established the European Credit Transfer and Accumulation System (ECTS) as well as the obligatory issuing of the diploma supplement automatically and free of charge by universities.

Consequently, the work and research in higher education institutions is subject to continuous evaluation with the objective of assuring and improving the quality of research and studies. The evaluation of higher education institutions is carried out in two stages: the first stage involves the evaluation performed by the academic units of the higher education units themselves in connection with their profile, objectives, and mission (internal evaluation). After completion of the first stage the second stage follows, which consists of the evaluation of the work performed by a committee of independent experts that also takes into consideration the results of the self-evaluation report of the academic unit (external evaluation).

Major Changes and Issues since the 1990s

After the mid-1990s, important reform measures concerning the education system were implemented. Among others they included:

- establishment of all-day schools and second-chance schools;
 - foundation of the regional counseling and orientation centers;
 - changes in the admission procedure for higher education;
 - upgradation of TEIs by incorporating them into the higher education area;
 - systematization of LLL policies and structures; and
 - significant action in the field of intercultural education to address the needs of groups with special social, cultural, or religious characteristics.
- At the beginning of the twenty-first century, in the context of a broader reform effort, the most important changes that have taken place are the following:
- increase in the duration of compulsory education from 9 to 10 years by including preschool education;
 - introduction of the cross-thematic curriculum framework for compulsory education (DEPPS) and corresponding new textbooks in primary education;
 - introduction of the teaching of a second foreign language in primary school;
 - creating a new institutional framework for the operation of higher education institutions securing the combination of autonomy and accountability (Law 3549/2007);
 - setting up of the quality assurance agency and putting in place the internal/external evaluation procedures to align the domestic higher education landscape with European and compatible international patterns (Law 3374/2005); within the same framework, the stipulation of the diploma supplement and the implementation of the academic credit transfer and accumulation system have further contributed to student mobility and aligned the Greek higher education system with European Union (EU) standards;
 - enhancement of higher technological education through, among others, the improvement of the academic components of the TEIs and the consolidation of their graduates' professional rights; setting up of upgraded and more flexible upper-secondary education vocational schools and training schools (Law 3475/2006);
 - transparency and efficiency with regard to the procedures involved in the recognition of academic titles through the setting up of the *Diepistimonikos Organismos Anagnorisis Titlon Akadimaikon kai Pliroforisis* (DOATAP)/Hellenic NARIC (Hellenic National Academic Recognition and Information Centre) (Law 3328/2005);
 - major policy priority in Greece given to modernization and integration of all LLL platforms into a comprehensive institutional and functional system to cut across both education/training and employment policies, which has oriented the national LLL strategy toward not only the qualitative enhancement of the provided education, but also the increase of employability, and the diffusion of flexible re-skilling training setups; also, social inclusion

and the promotion of citizenship competences have been at the core of the new education and employment outlook on LLL; the new law on the systemization of LLL (Law 3369/2005) constitutes the legal backbone of the entire matrix as it establishes a comprehensive and holistic web of agencies, platforms, and sub-networks operating on the basis of supplementarity and synergy;

- the establishment of the International Hellenic University (IHU) (Law 3391/2005) along with the adoption of the ECTS and diploma supplement and the introduction of postgraduate programs and degrees upheld jointly with foreign academic institutions (Law 3404/2005) which have further opened up Greek higher education to the European and international setups;
- upgrading of the relevant institutional framework on research and technology (Law 3653/2008); and
- enforcing a new law for inclusive education to ensure equal opportunities and equitable access for people with disabilities or special educational needs (Law 3699/2008).

Further Reading

- Center for Educational Research (2005). *Imprint of the Educational System of School-Level Units*. In Koulaidis, B. (ed.). Athens: Center for Educational Research.
- Education Research Centre of Greece and Greek Ministry of Education and Religious Affairs. (2004). *A Report on Education and Training in Greece*. In Koulaidis, V. (ed.). Athens: Education Research Centre of Greece and Greek Ministry of Education and Religious Affairs.
- European Commission, Directorate – General for Education and Culture (2006). Eurydice, The Information database on Education Systems in Europe. *The Education System in Greece, 2005/2006*. European Commission.
- European Commission, Directorate – General for Education and Culture (2008). *National Summary Sheets on Education Systems in Europe and Ongoing Reforms*. Greece, 2008 edn.
- Eurydice European Unit (2007a). *European Glossary on Education, Vol. 5: Decision-Making, Advisory, Operational and Regulatory Bodies in Higher Education*. European Commission: Directorate – General for Education and Culture.
- Eurydice European Unit (2007b). *Special Report: The Structure of Higher Education in Europe, 2006–2007: National Trends in the Bologna Process*. European Commission, DG Education and Culture.
- Eurydice European Unit (2008a). *Levels of Autonomy and Responsibilities of Teachers in Europe*. European Commission, Directorate – General for Education and Culture.
- Eurydice European Unit (2008b). *Higher Education Governance in Europe: Policies, Structures, Funding and Academic Staff*. European Commission, Directorate – General for Education and Culture.
- Eurydice European Unit and Eurostat (2005). *Key Data on Education in Europe 2005*. European Commission, Directorate – General for Education and Culture.

- Eurydice European Unit and Eurostat (2007). *Key Data on Higher Education in Europe, 2007 Edition*. European Commission, Directorate – General for Education and Culture.
- General Secretariat of Communication and General Secretariat of Information (2007). *About Greece*. Athens: General Secretariat of Communication and General Secretariat of Information.
- General Secretariat of National Statistical Service of Greece. *Education Statistics* <http://www.statistics.gr> (accessed July 2009).
- Hellenic Republic, Ministry of Economy and Finance (2007). National Strategic Reference Framework 2007–2013: Sectoral Operational Programmes. *Operational Programme for Education and Lifelong Learning*.
- Kyriazis, A. and Asderaki, F. (2008). *Higher Education in Greece*. Bucharest: UNESCO/CEPES.
- Ministry of Education (2007). *Our Work on Education. Activities of Ministry Upgrading the Quality of Public Education. March 2004 to May 2007*. Athens: Ministry of Education.
- Ministry of National Education and Religious Affairs (2007). *Hellenic Higher Education Institutions. International Co-operations*. Athens: Directorate of International Relations in Education.
- Ministry of National Education and Religious Affairs and Managing Authority of Operational Programme: Education and Initial Vocational Training (2000). *Operational Programme for Education and Initial Vocational Training 2000–2006*.
- OECD (2008). *Education at a Glance 2008*. OECD Indicators.
- Pedagogical Institute. (2004). *A Cross-Thematic Curriculum Framework for Compulsory Education. Diathematikon programma*. Athens: Pedagogical Institute.

Relevant Websites

- <http://www.greeklanguage.gr> – Centre for the Greek Language.
- <http://www.kee.gr> – Education Research Centre of Greece.
- <http://www.eurydice.org> – European Eurydice Unit.
- <http://www.gsae.edu.gr> – General Secretariat of Lifelong Learning.
- <http://www.statistics.gr> – General Secretariat of National Statistical Service of Greece.
- <http://www.ypepth.gr> – Greek Ministry of National Education and Religious Affairs.
- <http://www.sch.gr> – Greek School Network.
- <http://www.gunet.gr> – Greek Universities Network Portal.
- <http://www.eap.gr> – Hellenic Open University.
- <http://www.ihu.edu.gr> – International Hellenic University.
- <http://www.doatap.gr> – National Academic Recognition Information Centre.
- <http://www.ekep.gr> – National Centre for Vocational Orientation.
- <http://www.ein.gr> – National Youth Foundation.
- <http://www.epeaek.gr> – Operational Program of Education and Initial Vocational Training.
- <http://www.oEEK.gr> – Organization for Vocational Education and Training.
- <http://www.oepek.gr> – Organization of Teacher Training.
- <http://www.pi-schools.gr> – Pedagogical Institute.
- <http://www.osk.gr> – School Buildings Organization.
- <http://www.iky.gr> – State Scholarships Foundation.
- <http://www.isocrates.gr> – Special Secretariat for Greek Studies Abroad and Intercultural Education.

Iceland

M Hardardottir and M Gunnlaugsdóttir, Ministry of Education, Science and Culture, Reykjavik, Iceland

© 2010 Elsevier Ltd. All rights reserved.

General Background

Iceland is an island in the North Atlantic Ocean, bordering the Temperate and Arctic Zones. Its extreme northern point touches the Arctic Circle. Iceland's total area is approximately 103 000 km², of which only about 23% is covered by vegetation. As of 1 December 2007, the population of Iceland was 312 872, with a population density of approximately 3.0 inhabitants per square kilometer.

Iceland was settled from Norway and by Norse settlers from the British Isles in the ninth and tenth centuries. From 1262 to 1264, allegiance was sworn to the Norwegian Crown. In 1383, the Danish and Norwegian Crowns were united, and Iceland was thereafter part of the Danish–Norwegian kingdom. In 1918, Iceland was proclaimed a sovereign state in personal union with Denmark.

Since 1944, Iceland has been a republic with a written constitution and a parliamentary form of government. The president is head of state, elected for a term of 4 years by direct vote of the electorate. Parliament is composed of 63 members, elected for 4 years by proportional representation. The government holds executive power and is headed by the prime minister. Iceland is a representative democracy. The current constitution came into effect on 17 June 1944, when Iceland achieved its independence from Denmark. According to the Constitution of Iceland, the Parliament and the President jointly exercise legislative power. Judges exercise judicial power.

Local government is exercised by 79 municipalities.

Iceland was almost exclusively an agricultural society until the late nineteenth century, when fisheries gradually came to replace animal husbandry as the mainstay of the economy. Apart from fish, Iceland's most significant resource is its vast natural-energy potential, comprising hydro- and geothermal power.

Unemployment has been low in recent years. The unemployment rate was 2.3% in 2007. Gross domestic product (GDP) per capita was USD 39 986 (current *purchasing power parities*; PPPs) in 2006. Economic growth in 2006 was 4.2%.

The twentieth century saw great demographic changes in Iceland, as regards both the number of inhabitants and their patterns of residence. Until the twentieth century the population of Iceland was almost entirely rural. Urbanization was slow, and in 1850 the inhabitants of Reykjavík, the largest urban center, constituted only 1.94% of the total population. By 2005, however, approximately 92.8% of the population lived in communities of

over 200 inhabitants (60 localities), with around 62.1% of the country's inhabitants living in the capital area (including Reykjavík proper and the surrounding communities). This development mirrors the economic changes which took place during the last century, as increasing industrialization, especially in the fisheries and the fishing industry, led to the growth of urban communities, with a corresponding decrease in the importance of agriculture.

In 2006, about 23% of the total population was aged 0–15 years, 67% was 16–66 years, and 10% was 67 years and over. It is estimated that by 2030 the population will have reached 342 000.

Iceland has until recent years been a relatively homogeneous society. However, since 1990 and especially after 2000, there has been a considerable increase in the number of immigrants.

The Evangelical Lutheran Church is the official state church, with 82% membership in 2006.

Icelandic is the native tongue of Iceland and is its only official spoken and written language. It belongs, along with Norwegian and Faeroese, to the West Scandinavian branch of the North Germanic family of languages. Morphologically it has remained the most conservative of the Scandinavian languages, retaining, for example, three genders and a full system of case endings for nouns and adjectives.

From the middle of the seventeenth century, the law required that the clergy provide children with religious education, and during the 1790s it became one of their tasks to ensure that parents taught their children to read. In 1880, legislation was enacted on the instruction of children in reading, writing, and arithmetic and the Christian doctrine of the Lutheran Church. Parents were responsible for this instruction, while all supervision was placed in the hands of the ministers of the state church.

During the nineteenth century, there was a general movement toward public education and an interest in educational affairs. The development of a modern school system at the close of the nineteenth and the beginning of the twentieth century can reasonably be seen as one of the fruits of this movement.

In 1907, the first law on formal schooling for children aged 10 to 14 years was passed. This legislation marked the beginning of state involvement in educational affairs in Iceland, and education, free of charge, became the legal right of every child in the country from 10 to 14 years of age.

School levels in Iceland are four: pre-primary, compulsory (single structure – primary and lower-secondary education), upper secondary, and higher education.

Education in Iceland has traditionally been organized within the public sector, and there are relatively few private institutions in the school system. Almost all private schools receive public funding.

Goals of the Education System

A fundamental principle of the Icelandic education system is that everyone should have equal opportunities to acquire an education, irrespective of gender, economic status, residential location, religion, possible handicap, and cultural and social background.

Basic goals of the education system are set out in legislation for each school level and are supplemented in curriculum guidelines for pre-primary, compulsory, and upper-secondary education and in other policy documents.

In pre-primary schools children receive education and support for their all-round development, thus preparing them for compulsory school and life itself. At the pre-primary school level, the nucleus of educational work is play.

The main purpose of compulsory schooling is to prepare pupils for life and work in a continuously developing, democratic society. The organization of the school, as well as its work is thus to be guided by tolerance, Christian values, and democratic cooperation.

The primary aims of upper-secondary education are to prepare pupils for life and work in a democratic society by offering them suitable opportunities to learn and develop individually, and prepare them for employment through specialized studies leading to professional qualifications or further study.

The main purpose of adult education is to encourage equality of opportunity among adults without regard to location, age, gender, occupation, or previous education.

The role of higher education institutions is contributing to the creation and dissemination of knowledge and skills to students, as well as to society in general. The higher education institutions shall aim at strengthening the infrastructure of Icelandic society and its position in an international context.

Structure and Operation of the Education System

Figure 1 outlines the formal education system. Key statistics on number of schools, students, and teaching staff are provided in Table 1

Pre-primary School and Early-Childhood Care

The present Pre-primary School Act was passed in 2008. The Act's first article defines pre-primary schools as the first level of the educational system, and under the Act, pre-primary schools are to provide education for children

who have not reached the age at which compulsory education begins, that is, 1 September of the year in which the child turns six. The current national curriculum guidelines are based on a child-centered ideology, where the individual development and needs of each child is to be the focal point. All handicapped children have the same right as other children to attend pre-primary schools, and in many cases are given priority status with regard to admission. The program for handicapped children is the same as for other children, but adapted to their ability.

Pre-primary school education is to be provided by staff who have professional training in working with children at this level.

Compulsory Education (Primary and Lower Secondary)

Under the Compulsory School Act of 2008, local municipalities are responsible for the operation of schools at that school level.

The local municipalities pay for instruction (general teaching, substitute teaching, special education, and the teaching of children in hospitals), administration, and specialist services, as well as establishing and running schools at the compulsory level. The Act makes it the duty of parents to ensure that their children register for and attend school. The Compulsory School Act also specifies the duty of the state and local municipalities to ensure that mandated instruction be given.

Compulsory schools operate for 9 months a year. The number of school days is 180.

The state monitors compliance with educational law and regulations; it is also responsible for the publication of educational materials. The state is in charge of coordinated school surveys, national coordinated examinations in core subjects in years 4 and 7, and national coordinated examinations in three subjects in year 10. Furthermore, the state is responsible for evaluating individual schools and the educational work that is carried out.

Upper-Secondary Education

A new Upper Secondary School Act came into force in 2008.

Students may enter upper-secondary schools at the end of compulsory schooling in the year they turn 16. Upper-secondary education is not compulsory; anyone who has completed compulsory education has the right to enter a course of studies in an upper-secondary school. Majority of study programs are 4-year programs. In recent years, 93–94% of students who completed compulsory education have entered upper-secondary education directly thereafter; the dropout rate during

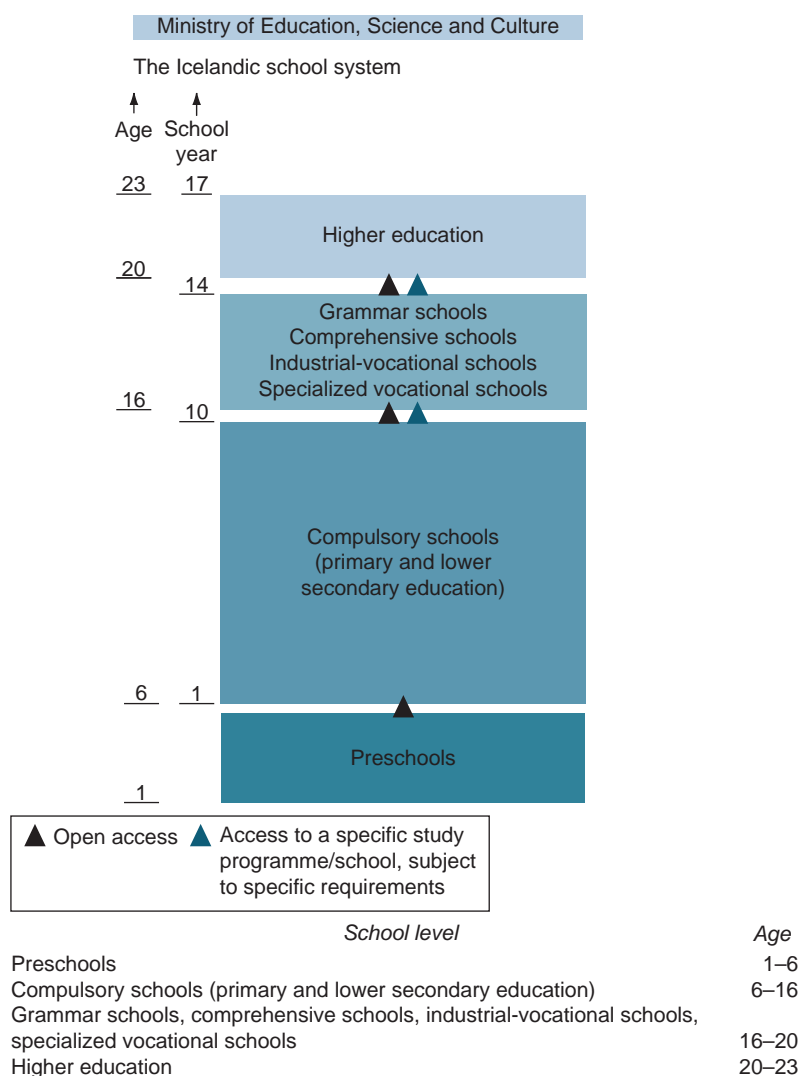


Figure 1 Diagram of the Icelandic school system. From Ministry of Education, Science, and Culture.

Table 1 Educational providers, students, and teaching staff 2007

Provider	Institutions ^a	Students			Teaching staff		
		Total	Male	Female	Total	Male	Female
Pre-primary school	267	17.216	8.778	8.438	4,459 ^b	134	4.325
Compulsory school	173	43.875	22.412	21.463	4.969	1.074	3.895
Upper-secondary school	36	24.459	11.440	13.019	1.880	889	991
Higher education	8	16.738	6.305	10.433	2,039 ^c	1.060	979

^a31 pre-primary schools, 7 compulsory schools, 7 upper-secondary schools, and 3 higher education institutions were private with government support in 2006.

^bTotal number of all involved in teaching. Number of pre-primary school teachers and staff with other educational training: 1792.

^cTotal number of all involved in teaching. Number of permanent teaching staff: 694.

From: Statistics Iceland and Ministry of Education, Science and Culture.

upper-secondary schooling has, however, been considerable. Completion rates are significantly higher for women than for men.

The school year, which lasts for 9 months, is divided into autumn and spring terms. Students generally attend 32–40 lessons per week, with each lesson lasting 40 min.

The main types of upper-secondary schools are as follows:

- Grammar schools (high schools) which offer 4-year academic programs of study which conclude with matriculation examinations.

- Industrial-vocational schools, which offer theoretical and practical programs of study in certified and some noncertified trades.
- Comprehensive schools, which provide academic programs comparable to those of the grammar/high schools and vocational programs similar to those offered by the industrial-vocational schools, as well as other specialized vocational-training programs.

Upper-secondary schools either have a traditional class structure, or operate according to a unit-credit system. The unit-credit system is the most common form of upper-secondary education.

The main types of programs of study are the following: (1) academic programs leading to matriculation, (2) vocational programs, (3) arts programs, and (4) a general program of study.

General academic education is primarily organized as a 3–4-year course leading to a matriculation examination. Vocational education is divided between the school and the workplace or takes place exclusively in the school. A general short program of study is intended for pupils who need further preparation for academic or vocational studies or for those pupils who are undecided as to what to do after compulsory education. In addition there are also fine arts programs of study. The length of vocational courses varies, but most are 3- or 4-year courses. Pupils in vocational programs of study can choose between training for the certified trades or vocational training in other areas, for example, in the field of agriculture, travel industry, fisheries, food production industry, health, or commerce.

Higher Education

The Icelandic higher education system is made up of seven university and nonuniversity institutions of which three are private. The Higher Education Act, enacted in 2006, establishes the general framework for the activities of higher education institutions under the auspices of the Ministry of Education, Science and Culture. The role of each public higher education institution is further defined in a separate act of parliament on its activities. The charters of private institutions define their engagement in research, internal organization, etc.

The Higher Education Act fully implements the Bologna Process into Icelandic higher education. The Act emphasizes quality work and quality assurance in higher education. All institutions that are currently operating are to apply for accreditation in each field of study before July 2008.

Legislation on the Science and Technology Policy Council was enacted in 2003 with the purpose of strengthening science, research education, and technical development in the country. All higher education institutions have in recent years placed increasing emphasis on research activities.

The foundation of the University of Iceland in 1911 marks the beginning of the modern Icelandic system of higher education. This first national university was established by merging three professional schools founded during the previous century – a school of theology, a school of medicine, and a law school – and adding a new faculty of humanities. Before that time Icelandic students had mainly traveled to Denmark for higher education.

In the last three decades the higher education system has grown more diverse. New higher education institutions have been established and some secondary institutions have been upgraded to the higher education level.

In recent years there has been a significant rise in the number of students in higher education, followed by an increase in the availability of higher education programs of various lengths.

Private institutions receive state support and can also charge tuition fees. Public institutions are only authorized to charge registration fees.

Adult and Continuing Education

Adult and continuing education is an extensive sphere in Icelandic society, provided by various parties such as public authorities, private institutions, companies, and organizations. The educational opportunities provided by public authorities are open to all, although with certain restrictions in some cases. They are intended to serve everyone according to their needs, especially young school leavers and adults who want to improve their basic education, general knowledge, or professional capabilities, or who want to cultivate their hobbies.

The Teaching Profession

The training of teachers and specialist educational personnel for the pre-primary, primary, and secondary levels is undertaken by universities. According to a legislation of 2008, training for pre-primary, primary, lower, and upper secondary teachers is to be 5 years.

In 2006, 78% of compulsory school teachers (primary and lower secondary) and 53% of upper-secondary school teachers (general and vocational education) were women. In pre-primary school education 97% of teachers were women. Many people with teaching qualifications are not employed as teachers. As there is a shortage of qualified teachers, especially in some regions outside the capital area and in mathematics and science, some of the teaching staff is unqualified. In 2006, 16% of compulsory school teachers and 24% of upper-secondary school teachers were unqualified. At the higher education level 34% of permanent educational staff in 2006 were women.

A variety of in-service training courses for teachers are offered annually.

Administrative and Supervisory Structure

The Icelandic parliament is legally and politically responsible for the educational system. It determines its basic objectives and administrative framework. All education comes under the jurisdiction of the Ministry of Education, Science and Culture.

The educational system has to a large extent been decentralized, with regard to both responsibilities and decision making. This reflects a general trend in Icelandic society. Local municipalities are responsible for the operation of pre-primary schools and compulsory education (primary and lower-secondary schools). On the other hand, the state runs upper-secondary schools and institutions at the higher education level.

The Ministry of Education, Science and Culture issues national curriculum guidelines for compulsory and upper-secondary education. These national curriculum guidelines are intended both to provide the detailed objectives necessary to implement the law and offer direction as to how they should be carried out in practice. The Ministry furthermore issues national curriculum guidelines for pre-primary schools that specify the aims that pre-schools are to follow and describe the basic means and attitudes that apply in the education of young children.

The National Centre for Educational Materials (under the auspices of the Ministry of Education, Science and Culture) develops and publishes educational materials for compulsory schools and distributes them to schools free of charge.

The Educational Testing Institute is an independent institution funded by the state through the Ministry of Education, Science and Culture. The institution is responsible for organizing, setting, and grading the national coordinated examinations and for undertaking comparative analysis of the educational system through participation in international surveys such as Organization for Economic Cooperation and Development Programme for International Student Assessment (OECD PISA).

Educational Finances

Since the late 1990s, the government has considerably raised expenditure on education. The OECD figures in *Education at a Glance 2007* indicate that Iceland's overall spending on education as a proportion of GDP was 8.0% in 2004; Iceland's expenditure was then the highest of any OECD nation. In that year the average figure for OECD nations was 5.8%.

Spending on education accounted for 19.94 % of total public spending in 2006. Almost half, or 9.54%, went to the compulsory educational level, 3.42% to upper-secondary

education (including vocational training), and 4.05% to higher education.

The Icelandic Student Loan Fund offers loans to students in higher education and some vocational students to cover costs incurred by studies, as well as the cost of living. In order to receive loans, students must complete at least 75% of full-time studies according to the program of the educational institution, approved by the board of the Fund. Repayment of loans begins 2 years after completion of studies.

Performance Monitoring, Evaluation, and Research

The Ministry of Education, Science and Culture is responsible for the evaluation and supervision of educational institutions and of the entire educational system in Iceland. The monitoring of the educational system consists mainly of external evaluations of schools and institutions or special aspects of the school system; national coordinated examinations in years 4, 7, and 10 of compulsory education; and by the use of data collected by Statistics Iceland.

Laws on pre-primary, compulsory, upper-secondary, and higher education levels place strong emphasis on regular internal evaluation by individual schools/institutions. Every compulsory and upper-secondary school is expected to implement methods to evaluate its work, including its teaching and administrative methods, internal communications, and contacts with parties outside the school. The internal-evaluation methods of compulsory and upper-secondary schools are subject to regular external evaluations by the Ministry of Education, Science and Culture.

Schools/institutions at all educational levels may be subjected to an external evaluation organized by the Ministry of Education, Science and Culture. This evaluation is conducted by external evaluators. The purpose of external evaluations of schools/institutions is to obtain an overall picture of each school's or institution's activities or of specific aspects at any given time. Attention is directed toward various features of the school or institution's internal activities, such as administration, teaching, development work, cooperation, and communications within the school/institution, study achievements, and the connection between the school or institution and society.

Every 3 years, the Minister of Education, Science and Culture is to deliver comprehensive reports to parliament on compulsory and upper-secondary education, based on systematically gathered information, national and international research, and evaluation.

Research is mostly carried out by higher education institutions in the country. Iceland has in recent years also participated in international research such as the

OECD PISA and Progress in International Reading Literacy Study (PIRLS). In addition, Iceland participates on a regular basis in OECD work on developing student-achievement indicators. The Icelandic educational system has been regularly reviewed by OECD experts. Iceland also participated in the OECD thematic review of tertiary education. The report: Thematic Review of Tertiary Education, Iceland Country Note, was published by OECD in April 2006.

Major Changes and Issues since the 1990s

Great changes have taken place in education in Iceland since 1990, at the same time as society has undergone considerable change due to the influence of globalization and a free-market economy. Greater demands are made on the educational system in terms of quality and efficiency.

As the economy has become more diversified and internationalized the demand for a highly educated workforce has increased significantly. Rising numbers of students over the period, and increased services, have necessitated an increase in the number of people involved in education; they represented 5.7% of the workforce in 1991, which had risen to 7.5% by 2006.

Participation in international collaboration in education has increased greatly since 1990, not least due to Iceland's membership of the European Economic Area since 1994. Iceland thus became a full member of European Union collaborative programs in the field of education and science. Participation in Nordic collaboration in education and science has also increased, for example, the Nordplus programs.

In the 1990s legislation was enacted for all four educational levels. The Pre-Primary Schools Act was passed in 1994, the Compulsory Schools Act in 1995, the Upper-Secondary Schools Act in 1996, and the Higher Education Act in 1997. A new Higher Education Act was passed in 2006. The legislation entailed *inter alia* increased autonomy and responsibility for educational institutions, and greater emphasis on evaluation and monitoring of educational activities. Under the Compulsory Schools Act of 1995, the operation of the schools was transferred to local government.

Current curricula for the compulsory school level date from 2007, and for the pre-primary and upper-secondary level from 1999. The duration and structure of upper-secondary studies leading to matriculation examinations have been under review by the Ministry of Education, Science and Culture.

Considerable changes have taken place in higher education over the period. The range of study programs has increased greatly, not least at the postgraduate level. The

number of higher education institutions has increased, *inter alia* by educational institutions previously at upper-secondary level being upgraded to higher education level. Seven institutes of higher education now operate in Iceland; three are privately run, but with state funding. Their funding is based upon the same allocation rules as for state-run higher education institutes.

At all educational levels except the compulsory level, proportional enrolment has increased greatly. From 1997 to 2006 the number of students in higher education in Iceland doubled, that is, from 8372 to 16 738. In addition, 2308 Icelandic students were studying abroad in 2006. Women constituted 62% of the student body in Icelandic higher education institutions in 2006 and were 67% of graduates in the academic year 2005–06.

The proportion of 16-year olds enrolled in the first year of upper-secondary education rose over the same period from 83% to 93%. In 2006, about 95% of children aged 3–5 attended pre-primary school.

Research and development expenditure was 3% of GDP in 2006, up from about 1% in 1990.

See also: An Overview of Educational Assessment; Assessment in Schools Related To Literacy: Reading; Post-Structuralism and Education.

Further Reading

- Educational System in Iceland (2002). Ministry of Education, Science and Culture. Reykjavík.
- Eurydice (2007). National Dossier on the Icelandic Educational System 2006. <http://www.eurydice.org> (accessed July 2009).
- Landshagir (2007). *Statistical Yearbook of Iceland*. Reykjavík: Hagstofa Íslands.
- Ministry of Education, Science and Culture in Iceland (2005). Ministry of Education, Science and Culture. *Publication No. 18*. Reykjavík.
- OECD (2005). Thematic review of tertiary education. *National Background Report for Iceland*. Ministry of Education, Science and Culture. Paris: OECD.
- OECD (2006). Adapting the Icelandic Education System to a Changing Environment. *Economic Department Working Papers No. 516*. Paris: OECD Publishing.
- OECD (2006). Thematic Review of Tertiary Education. *Iceland Country Note*. Paris: OECD.
- OECD (2007). *Education at a Glance 2007. OECD Indicators 2007*. Paris: OECD.

Relevant Websites

- <http://www.althingi.is> – Alþingi.
- <http://www.eurydice.org> – Eurydice.
- <http://www.hagstofa.is> – Hagstofa Íslands (Statistic Iceland).
- <http://www.menntamalaraduneyti.is> – Menntamálaráðuneyti (Ministry of Education, Science and Culture).
- <http://www.namsmat.is> – Námsmatstofnun.
- <http://www.oecd.com> – oecd.com.
- <http://www.hi.is> – School of Education, University of Iceland.

Indonesia

A C Alwasilah and Furqon, Indonesia University of Education, Bandung, West Java, Indonesia

© 2010 Elsevier Ltd. All rights reserved.

Indonesia has an area of 1.9 million sq km and is inhabited by more than 230 million people; thus making it the world's fourth most populous country and the world's largest archipelago country consisting of more than 17 000 islands. The country consists of urban areas (95 449 616 sq km) and rural areas (123 020 384 2 km). More than 60% of the population (and more than 500 ethnic groups) live in the rural area. The country includes 33 provinces, 440 districts or municipalities, 5641 sub-districts, and 71 555 villages. The country's official religions include Islam (191 303 211), Christian (13 242 316), Catholic (7 846 779), Hindu (4 022 993), and Buddhist (2 789 701). The per capita income (or gross domestic product) is Rp 11 193 855.8 (around US\$1120). **Table 1** summarizes the school age-group population and the target population of students.

Until 2004, the average number of years of schooling of 15-year-olds or above was 7.2 years, while their literacy rate was 90.4%. Early-childhood education programs were very limited and disproportional. Out of 28.2 million 0–6-year-olds, only 7.2 million (25.3%) received education. Out of 8.14 million 5–6-year-olds, only 2.63 million (32.36%) went to kindergarten. School participation of 7–12-year-olds and 13–15-year-olds was 96.8% and 83.5%, respectively. Around 76% of the families suggested that economic factors were the major reasons for this matter, where 67% had no budget and 8.7% of school-age youth had to earn a living. In 2004–05, there were around 330 000 students in 2870 junior secondary schools. They were students who could not afford to continue their studies at regular secondary schools. There were 1.5 million students with special needs, but only 4% of these received an appropriate education. In tertiary education, the college participation of 19–24-year-olds was relatively low at 14.6%. **Figure 1** depicts the ladder of national education.

The National Education System

The national education in Indonesia evolved from its earlier form during the three-and-a-half centuries of Dutch occupation and the 2-year Japanese occupation. The nation formally started its national system in 1945 immediately after independence. Elitist policymaking by the Dutch resulted in resentments leading to independent movements among indigenous people. This was even more noticeable during the Japanese occupation when Indonesian was encouraged to be used as the medium of instruction. As in other parts of the globe, education in

Indonesia developed over the times to meet the challenges of contemporary life. National education is based on *Pancasila* (five pillars of the state's philosophy, i.e., belief in God, humanism, nationalism, democracy, and social justice) and the 1945 Constitution of the Republic of Indonesia.

Referring to Law No. 20/2003 on national education system, the vision and missions of national education are described as follows. National education has a vision of realizing an Indonesian society that observes peace and democracy; possesses expertise and competitiveness; and achieves advancement and prosperity in the unity of the Republic of Indonesia supported by citizens who are healthy, independent, faithful, and pious to God, the Almighty; possess morals and noble character; love the country, observe laws, and conserve the environment; master science and technology; and possess work ethos and discipline. To accomplish the vision, the Ministry of National Education (MONE) has the following missions:

1. to strive for the expansion and even distribution of opportunities for quality education for all Indonesian citizens;
2. to assist and facilitate the development of children's potentials, from early childhood throughout life, in order to bring into being a learning society;
3. to improve quality of educational inputs and processes to optimize the building of moral character;
4. to enhance the professionalism and accountability of educational institutions as centers for acculturation of sciences, skills, experiences, attitudes, and values based on national and global standards; and
5. to empower community participation in the provision of education based on the principles of autonomy in the context of the unity of the Republic of Indonesia.

The current education system is a result of legal reform, beginning with the amendment of the 1945 Constitution. The amendment stipulates that education is not only the right of Indonesian citizens but it is now part of human rights. All Indonesian citizens have to get primary education and the government has to pay for it. The amendment also mandates at least 20% of the state budget at the national as well as regional levels to be allocated for education. The amendment was then elaborated in the following laws, among others: Law No. 20/2003 on educational system, Law No. 14/2005 on teachers and lecturers, and Law No. 43/2007 on library. During the 2005–07 period alone, MONE has accomplished nine major policies which are as follows:

1. *Massive education funding.* The government spent a total amount of Rp 11 532 484 429 000 00 for school-operating grant, special grant for students, and quality management grant.
2. *Improvement of qualification, competencies, as well as certification of teachers and educational personnel.* As mandated by Law No. 14/2005, teachers as professionals should have S1 or D4 qualifications, while university lecturers should have S2 or S3 qualifications. It suggests that there are around 1.75 million teachers yet to achieve S1/D4 qualifications and more than 150 000 university

lecturers yet to achieve S2/S3 degrees; there are around 2.5 million teachers and 300 000 university lecturers to be certified in the next 10 years. In 2007, MONE allocated Rp 579 billion to improve the qualification of 81 800 teachers and 8540 lecturers. Moreover, 147 217 teachers were already certified.

3. *E-learning and E-administration.* Information technology (IT) has been used to support the three pillars of national education, namely (1) better access to education, (2) better quality of education, and (3) improved governance, accountability, and public image. IT is relevant for Indonesia with a population of 230 million, with more than 50 million of formal education participants occupying more than 17 000 islands.
4. *New facilities.* To facilitate the implementation of the Education Act, the government built new schools, state universities, libraries, and laboratories.
5. *Reconstructing educational facilities.* On reconstructing facilities, the central government and regional government shared the funding in a ratio of 50:30:20% of the fund was to be paid by the central government, provincial government, and city or district administration.
6. *Textbook provision.* The textbook policy has changed drastically as follows: (1) there is no more monopoly by MONE on textbook writing, provision, publishing,

Table 1 Group age population in thousands

Group age and population		Target group of education		
No	(2007)			
1	7-12 years	248 357	Elementary	285 330
2	13-15 years	129 341	Secondary	112 381
3	16-18 years	127 251	Senior secondary	72 793
4	19-24 years	253 181	Tertiary	39 400
5	15-older	1 558 166		
	Total population	2 216 543		

Source: Ministry of National Education (2007).

Official school age	Level of education	Academic education		Professional education			
		MORA	MONE	MORA/MONE			
26	Higher education	Islamic doctorate program (S3)	Doctorate program (S3)	Second professional program (SPII)			
25		Islamic master program (S2)	Master program (S2)	First professional program (SPI)			
24		Islamic master program (S1)	Graduate degree program (S1)	Diploma 4 program	Diploma 3 program	Diploma 2 program	Diploma 1 program
23							
22	Secondary education	Islamic GSSS	General SSS	Vocational SSS	Islamic vocational SSS		
21							
20	Basic education	Islamic GJSS	General JSS				
19							
18		Islamic primary school	Primary school				
17							
16	Early childhood	Islamic kindergarten	Kindergarten				
15							
14							
13							
12							
11							
10							
9							
8							
7							
6							
5							
4							
3							
2							
1							
0							

Figure 1 Education in Indonesia. GJSS: General Junior Secondary School; GSSS: General Senior Secondary School; JSS: Junior Secondary School; SSS: Senior Secondary School; MORA: Ministry of Religious Affairs; MONE: Ministry of National Education. Source: Ministry of National Education (2006: 15).

and distributing; (2) textbooks are selected by school committees and are to be used for at least three consecutive years; (3) students are recommended to buy the books from the bookstores, while the teachers are not allowed to get involved in the textbook business; (4) education units have to provide libraries with textbooks as many as necessary to help the students who cannot afford to buy any; (5) MONE and Ministry of Religious Affairs (MORA) are encouraged to buy the copyrights of selected textbooks so that they can be mass-produced for students at relatively low costs; and (6) MONE, MORA, and provincial government are encouraged to provide the prospective textbook shop-owner with subsidy and capital.

7. *Improvement of quality, relevance, and competitiveness.* MONE has adopted a comprehensive strategy to improve the quality, relevance, and competitiveness of education through the enactment of Law No.19/2005 on the national standards of education. The law mandates the government to form the National Board of Education Standards, whose function is to set the eight national standards on the following: (1) content, (2) competency, (3) educators and educational personnel, (4) educational facilities, (5) evaluation, (6) educational process, (7) management, and (8) funding. In addition, the government promotes school-based curriculum, where every educational unit has autonomy to develop its own curriculum.
8. *Improving the governance, accountability, and public image.* The government has taken several steps, among others as follows: (1) canceling laws or regulations that contradict other laws and regulations, (2) enacting new laws or regulations to enhance the three pillars of education, and (3) restructuring the MONE organization.
9. *Intensifying nonformal and informal education.* To reach the hitherto unreached sectors of the population, the government has launched some programs among others as follows: (1) nonformal early-childhood education programs, (2) literacy campaign, (3) package A, B, and C, (4) life-skill education, (5) community literacy circle, and (6) gender mainstreaming. In 2007, nonformal early-childhood education programs served 9 542 776 out of 28 426 500 children, while the formal early-childhood education programs served 5 193 298 children. In 2004, the number of illiterates aged 15 years or older was 15 million (10.2%), and in 2007, the number significantly decreased to 11 million (7.2%).

Educational Streaming

Indonesian education consists of formal education, nonformal education, and informal education. Level of education consists of basic education, secondary education,

and higher education, while types of education include general education, vocational education, and special education. The streams, levels, and types of education can take the form of an educational unit organized by the central government, local government, and/or community. All these are managed by MONE through four directorate generals as follows: Directorate General for the Management of Primary and Secondary Education, Directorate General for Higher Education, Directorate General for Nonformal and Informal Education, and Directorate General for Quality Improvement of Teacher and Educational Personnel.

Basic education is the foundation for the secondary education. Basic education takes the form of primary schools, *Sekolah Dasar* (SD) as well as *Madrasah Ibtidaiyah* (MI), or other schools of the same level, and junior secondary school, that is, *Sekolah Menengah Pertama* (SMP) as well as *Madrasah Tsanawiyah* (MTs).

1. Secondary school is the continuation of basic education. Secondary education comprises general secondary education and vocational secondary education. Secondary education takes the form of senior general secondary schools, *Sekolah Menengah Atas* (SMA), *Madrasah Aliyah* (MA), senior vocational school, *Sekolah Menengah Kejuruan* (SMK), and *Madrasah Aliyah Kejuruan* (MAK).
2. Higher education is the level of education after secondary education consisting of diplomas (1–3-year program), bachelor's (*sarjana*), master's, and specialized postgraduate programs, and doctorate programs imparted by a higher education institution. Higher education institutions are provided in a flexible system, and take the form of academies, polytechnic (*sekolah tinggi*) institutes, or universities. Higher education institutions can also run academic, professional, and/or vocational and technical programs.
3. Nonformal education is designed for community members who need education service that functions as a replacement, complement, and/or supplement to formal education within the framework of supporting lifelong education. It is aimed at developing learners' potential emphasis on the acquisition of knowledge and functional skills, and developing personality and professional attitudes. Nonformal education comprises life-skills education, early-childhood education, youth education, women-empowerment education, literacy education, vocational training and internship, equivalency programs, and other kinds of education aimed at developing learner's ability.
4. Informal education can be in the form of self-learning, provided by families and communities. The outcomes of informal education shall be recognized as being equal to the outcomes of formal education and nonformal education after successfully passing an assessment according to national education standards.

5. Early-childhood education is organized prior to basic education. It is provided through formal education, non-formal education, and/or informal education. Early-childhood education provided through formal education can take the form of kindergarten, or *Taman Kanak-kanak* (TK), as well as *Bustanul Athfal*, or *Raudhatul Athfal* (RA). Early-childhood education provided through nonformal education can take the form of playgroups, that is, *Kelompok Bermain* (KB), childcare centers, or *Taman Penitipan Anak* (TPA), or other similar forms of nonformal education. Early-childhood education provided through informal education can take the form of family education or social education.
6. In-service education is professional education provided by government departments or nongovernment institutes. In-service education functions to enhance the ability and skills to carry out the duties for government officials, and is provided through formal education and nonformal education.
7. Religious education is provided by government and/or by any group of people belonging to the same religion in accordance with the law. Religious education functions to prepare learners to become community members who understand and practice religious values and/or acquire expertise in religious studies. Religious education can be conducted through formal education, nonformal education, or informal education. Religious education can take the form of *diniyah* education, *pesantren*, *pasraman*, *pabbaja samanera*, and other similar forms of education.
8. Distance education can be organized in all streams, levels, and types of education. It provides educational services to any group of people in the community who cannot attend face-to-face courses or regular classes. Distance education is organized in various forms, models, and coverage by learning facilities and services, and an assessment system which ensures that the quality of graduates is in accordance with national education standards.
9. Special education is provided for learners who have difficulties in following the learning process because of physical, emotional, mental, or social deficiencies. It is also for those with proven intelligence and those who are especially gifted. Education with special services is provided for learners in the remote and less-developed areas, isolated areas, and/or for learners who are victims of natural disasters who suffer from social deficiencies, and those who are economically disadvantaged.

National Curriculum Guidelines

The national center for curriculum development has the task of developing curriculum with reference to the national standard of education. At all levels of education, the curriculum is developed based on the diversification

principle relevant to education units, local potential, and students. The curriculum is developed within the framework of maintaining the unified state of Indonesia, based on the following principles:

- strengthened faith and obedience to God Almighty;
- improved character;
- increased potential, intellect, and interest of students;
- diversity of local potential and environment;
- local and national development demands;
- employment demands;
- development in science, technology, and arts;
- religion;
- global developments; and
- national unity and national values.

The curriculum of elementary and secondary education must include the following: religious education, civic education, language, mathematics, science, social studies, arts and culture, physical education and sports, vocational skills, and local contents. The curriculum of higher education should include religious education, civic education, and language.

Teacher Education

According to Law No. 20/2003 on national education system, teachers are professionals who are responsible for carrying out the following tasks: planning and implementing the learning process, as well as evaluating its results. Teachers must have the minimum qualification and be certified to teach at relevant levels of educations; be healthy physically and mentally; and be capable of implementing the goal of national education. Teacher promotion is based on educational background, experiences, competencies, and achievements in the field of education. In formal education settings, teaching at the level of early-childhood education, elementary education, and secondary education must be done by graduates of accredited universities. Teacher certification is offered by selected universities that have relevant educational programs. The central and local governments must provide support for training teachers at all levels in their respective jurisdictions. **Table 2** summarizes the number of teachers at all school levels.

Table 2 The number of teachers in 2005/2006

No. of schools private/ public	No. of teachers	No. of students
144 262	1 346 846	25 982 590
23 853	616 364	8 073 389
15 342	469 360	5 729 347
187 457	2 432 570	39 785 326

Source: Indonesia Educational Statistics in Brief 2005/2006.

Since independence, teacher education has been carried out by public or private institutes of teacher training or faculties of education. In the 1990s, those institutes changed into universities with the wider mandate mission of carrying out education as well as noneducation programs. Throughout the country, there are 12 public teacher training universities and a number of private ones. The existing teacher training has been carried out using the concurrent mode rather than consecutive mode of training. Since the first semester, the prospective teachers have to take subject-matter courses as well as pedagogical courses. However, Law No. 14/2005 on teachers and lecturers implicitly states that a graduate with a bachelor's degree from any university can be admitted into the certification program and later become a certified teacher. In other words, any graduate of a noneducation university can apply for certification to be a certified teacher. Many suspect that in the long run, the existing concurrent program could be legally terminated, in favor of consecutive mode of training as is common in other countries.

Higher Education

Up to 2007, there are 82 state universities and 2800 private universities, offering 14 000 study programs at various levels, including diploma, S1 (BA), S2 (master's), and S3 (doctorate). The participation rate at universities under MONE, MORE, and open university has reached 17.25%, which is far from ideal. There are 231 757 lecturers teaching at 2811 universities, excluding those under non-MONE such as MORA, and others. In 2008, there are around 2.6 million students at universities under MONE, so quantitatively the lecturer–student ratio of 1:11 is fairly good. In general, lecturers at public universities are more qualified than their counterparts at private universities. There are 42 069 lecturers with S2/3 qualifications, that is 62.61% of the total of 67 192 lecturers at

MONE. **Table 3** depicts academic qualification of lecturers in MONE. As mandated by Higher Education Long Term Strategy (HELTS) IV 2003–2010 that higher education developments are aimed at the following: (1) creating competitiveness among universities, (2) developing autonomy, and (3) establishing healthy organizations.

The government encourages all public universities to be autonomous. There are already seven public universities that have enjoyed this autonomy, namely UI in Jakarta, ITB in Bandung, IPB in Bogor, UGM in Yogyakarta, USU in Medan, UPI in Bandung, and UNAIR in Surabaya. Among the public universities some are already included among the top 500 universities in the world and they are also among the best 100 universities in Asia. In 2007, according to the *Times Higher Education Supplement*, six universities were short listed, namely UGM (218 study programs), ITB (100 study programs), UI (202 study programs), UNDIP (115 study programs), UNAIR (96 study programs), and IPB (80 study programs), while UT (47 study programs) is accredited by OECD.

Funding

A comparative study on educational funding in the 1990s showed that the budget for education allocated at the central government as well as provincial government is relatively low. In 2007, for example, the budget for education was around US\$4.63 billion. This figure does not include the education budget allocated at the provincial levels for 33 provinces and 440 districts or cities. With this allocation, MONE is the department that has the highest budget. In general, the budget is allocated for social assistance (63.47%), buying goods (16.01%), paying salaries (15.22%), and investment (5.29%).

In response to the call for reform in 1997, the People's Consultative Council has amended the 1945 Constitution. Chapter 31, Article 4 of the Constitution mandates that the government should allocate 20% of the state budget

Table 3 Number of lecturers and their qualifications in MONE

No	Levels of education	Gender		Statistics	
		Male	Female	Total	Percentage (%)
1	S-3	5962	1635	7597	11.72
2	S-2	19 371	9998	29 369	45.30
3	S-1	16 390	9709	26 099	40.26
4	Sp-1	612	239	851	1.31
5	Sp-2	231	32	263	0.41
6	D-4	232	52	284	0.44
7	D-3	68	6	74	0.11
8	D-1	2		2	0.00
9	Profession	133	161	294	0.45
	Total	43 001	21 832	64 833	100
	Percentage	66.33%	33.67%	100%	

Source: Ministry of National Education (2008).

for education, so that in 2009 the allocation for education will be around US\$22.8 billion. This figure includes the budget for education administered by agencies outside the MONE, yet does not include the education budget at regional levels. To ensure quality education at the school level, Law No. 20/2003 mandates standardization of education budgeting. To realize it, the government has issued Regulation No. 9/2005 on national standards of education, which sets the standards for financing schools and other educational units in terms of investment, operation, and personal budgets.

Major Changes and Issues since the 1990s

One of education milestones in Indonesia is the enactment of a new law on education known as Education Act No. 2 on the national educational system that was officially launched on 27 March 1989. Some of the major issues in the law include the following:

1. Consistent with the centralistic management of administration system as a whole, national education management is the responsibility of the MONE. The school learning activities are based on the national curriculum. Recently, there have been the Curriculum 1975, Curriculum 1984, and Curriculum 1994, all of which were based on ministerial decisions and were implemented at the elementary and secondary education levels nationwide. The law mandates the implementation of 9-year compulsory education, namely 6-year elementary and 3-year secondary education. However, up to now (2008) this has not been fully implemented. Due to economic reasons, many elementary school-agers are not admitted into school or drop out from school.
2. There were four major issues of education in Indonesia, that is, access to education, quality of education, relevance of graduates to the needs of community, and effectiveness and efficiency of educational management. Access to education is made possible through the 9-year compulsory education programs and the establishment of new schools in various remote areas to accommodate all school-agers for elementary and secondary education. The issue of quality education is concerned with attempts to help all students achieve a specified standard of competency. Hence, the society will expect that graduates of education levels have the minimum of a specified set of competencies. The issue of relevance of education is concerned with the relevance of quality and competence of school graduates with the needs of society and industry. The link-and-match policy was once introduced to ensure the relevance of graduates with the community's needs. The issue of effectiveness and efficiency in managing educational resources from the lowest level of educational units up to the national level was facilitated by programs to propose solutions to the three problems mentioned above.
3. Since the 1980s, EBTANAS (National Final Learning Assessment) has been implemented at SD/MI, SMP/MTs., SMA/MA, and SMK. EBTANAS was a solution to the perceived quality of national education which decreased since the school examination replaced the national examination in the early 1970s. Some people proposed to revive the national examination to increase the quality of education, while some others wanted to continue with the school examination. The current EBTANAS was a compromise to combine the scores of the national examination (called EBTANAS) and those of the school examination (called EBTA). When the 9-year compulsory education policy was mandated, EBTANAS was no more relevant for SD/MI.
4. In response to the issue of relevance of education with the community's needs, the Minister of Education launched the link-and-match policy. The school was recommended to collaborate with the industry in designing the curriculum and learning programs to ensure that the school graduates have skills more relevant with the employment demands.
5. There were at least two factors that should be noticed during the implementation of the Education Act No. 2 1989 that significantly affect educational development in Indonesia: national consensus on the implementation of decentralization and democratization, and the introduction of a new curriculum known as competence-based curriculum. Under Law No. 22 Year 1999 on district autonomy, management of education with few exceptions is decentralized to district level. The central government focuses, among others, on the establishment of regulations needed for the implementation of the new law, and the development, monitoring, and evaluation of national standards in education. Meanwhile, MONE through the Center for Curriculum Development developed a new national curriculum for primary and secondary school known to be a competence-based curriculum.
6. With respect to those issues and national political consensus on decentralization and democratization through Law No. 74/2000, education reform was badly needed. In response to the need for education reform, MONE established a commission for education reform or KRP (*Komisi Reformasi Pendidikan*). It is worth noting that education reform is a part of nation reform as a whole. The commission was to review and revise Education Act No. 2 of 1989 with respect to the issues above and district autonomy under coordination of the head of the office of research and development of national education. The commission came up with a new law of education known as that of Education Act No. 20 of 2003.

The Major Issues in Education Law No. 20/2003

The economic crisis that hit the country in 1997 caused total reform movements in education. There have been major changes in education policies ever since, among others, as follows:

1. Consistent with the law on regional autonomy, Law No. 20/2003 mandates decentralization of government. Education management was shared by the district or city and the central government with the focus on (1) developing, monitoring, and evaluating the national standard of education and its implementation, and (2) proposing or drafting regulations for the implementation of Law No. 20/2003.
2. Early-childhood education is specifically mandated in several articles. Also mandated in the law is special education. Earlier managed by a section in the directorate of kindergarten, elementary education, and secondary education, special education is now managed by a directorate. Distance learning, which was formally mandated to *Universitas Terbuka* or open university is now explicitly stated in the law.
3. National standards of education (SNP) are now explicitly stated in the law. Chapter 35 lists eight standards, namely on content process, competency process, graduate's competency, education personnel, facilities, management, funding, and evaluation. The government has formed an independent board of standardization and quality assurance responsible for implementing and monitoring the standards.
4. As realization of Law No. 20/2003 on the national education system, some government regulations have been enacted, such as Regulation No. 19 Year 2005 on national standard of education. The board of national standard of education was formed and mandated to assist the minister to develop and monitor the implementation of the national standards, that is, on content process, competency process, graduate's competencies, educator and educational personnel, facilities, management, funding, and evaluation; and to make recommendations to the government. BSNP is also tasked with administering the national assessment and evaluating the elementary and secondary textbooks in terms of content, presentation, language, and graphics.
5. A new directorate general was formed to manage teachers and educational personnel called directorate

general for quality improvement of teacher and educational personnel improvement. Consistent with government commitment to improvement, supported by PGRI (Indonesian Teachers Union) and Association of LPTK (Teacher Training University), Law No. 14/2005 on teachers and lecturers was enacted. According to the law, any teacher has to be certified. Practically all working teachers have to be certified, and those certified teachers will get extra payment.

6. The government has set a policy on increasing the number of vocational high schools so that by 2015, the ratio of vocational high schools to general high schools will be 60:40. The policy was motivated by demands for middle labor force required to transform Indonesia from an agricultural country to an industrial one.
7. The government reiterates the commitment to education characterized by access to quality education, cheap education, and access to education for all. To implement this, the government has launched inexpensive book programs. The public are encouraged to write elementary and secondary school textbooks. The books are evaluated by the Board of National Standard of Education (BSNP) and quality textbooks will be mass-produced for school children to access manually as well as electronically.

Further Reading

- Jalal, F. and Bachrudin, M. (eds.) (2001). *Education Reform in the Context of Regional Autonomy: The Case of Indonesia*. Indonesia: Ministry of National Education and National Development Planning Agency.
- Ministry for Coordinating People's Welfare (2008). *Pendidikan di Indonesia: Masalah dan Solusi*.
- Ministry of National Education (2006). *Educational Indicators in Indonesia 2005/2006*.
- Ministry of National Education (2006). *Indonesia Educational Statistics in Brief 2005/2006*.
- Ministry of National Education (2006). *Report on Achievements of Department of National Education, 2006*.
- Ministry of National Education (2006). *Fifty Years of Development of Education in Indonesia, 1996*.
- Ministry of National Education (2006). *Strategic Planning of National Education for 2005–2009*.
- Ministry of National Education (2007). *Pembangunan Pendidikan Nasional 2005–2007*.
- Ministry of National Education (2008). *Laporan Portofolio Perguruan Tinggi di Indonesia Tahun 2007*.

Background

As of late 2006, Israel's population was at 7.1 million – 16.5% Muslim, 2% Christian, 1.6% Druze, and slightly over 80% Jewish. The founding of the State of Israel in 1948 as a Jewish national home – following the destruction, on the European continent, of a third of the Jewish people during World War II – triggered waves of Jewish immigration to Israel from countries around the world – a phenomenon that continues to this day. Israel is a rich and unusual demographic mosaic, in which diverse cultures intermingle and the streets resound with a medley of languages besides the official Hebrew and Arabic. Jews from 79 countries, representing about 39 different languages (CBS, statistical Abstract of Israel, 2009: 237) have made their home here. In addition to the formidable task of building a nation, Israel has had to cope, through the years, with a tense geopolitical situation in the Middle East. These facts elucidate the complex reality with which the Israeli education system is faced.

Complex though it may be, Israeli society is perceived as a creative and vibrant one in all spheres of life. For example, during the 5-year period between 1999 and 2003, Israel ranked third in the number of scientific papers *per capita*, a world leader in applied science, and ranked second worldwide in astrophysics, third in computer science, and fourth in molecular biology (Czapski and Ilan, 2004). All seven of the country's universities were included in the list of the world's 500 leading research institutions in 2005. The local thirst for culture has led to the creation of a wide and growing variety of cultural institutions. Indeed, recent years have witnessed an unprecedented growth of cultural activity, reflected in a threefold increase in cultural consumption on the part of Israeli children and adolescents over the last two decades.

On the economic level, since the 1980s, Israel has undergone the transition from a planned to a liberal economy. As of 2007, Israel's gross domestic product (GDP) is US\$26 500 *per capita*; economic growth rate was 5% in 2007, nearly double that of the Organization for Economic Cooperation and Development (OECD) countries. Annual inflation is about 2%, the labor-force participation rate is 55.2%, and the unemployment rate is 8%. During the 1990s, the information and communication technology (ICT) sector was the main force driving Israel's economic growth. To give some perspective, in 2004, the sector's share of Israel's total exports was 20%, compared with 13% for the same year in the OECD countries.

Despite Israel's achievements in the areas of economic growth, immigrant absorption, reach of cultural life, science, and research, social disparities persist. A quarter of Israel's children live below the poverty line; the disparity in per-pupil spending on education between the highest and lowest decile has held steady for the past decade at 11-fold (Israel Shekel Exchange Rates (NIS) 126 per month for the lowest decile vs. NIS 1355 for the highest decile in 2005).

Against this background, it becomes easier to understand the complex circumstances in which the Israeli education system operates: absorbing new immigrants whose native language is not Hebrew, profound social and economic disparities, and, above all, the drive to meet international standards – as well as a demanding public's expectations – for scholastic performance. The Israeli education system based on the foundations laid during a 50-year prestate period, of all walks of life, was administered by the Vaad Leumi, or Jewish National Council, mainly during the British Mandate in Palestine. This infrastructure was influenced by the numerous initiatives, organizations, and ideological tracks that were active in Palestine from the end of the nineteenth century onward and from which a growing variety of educational enterprises emerged. These enterprises included vocational education networks as well as boarding school systems (youth village) that rescued children via youth *aliyah* frameworks prior to the European catastrophe besides assisting young people in immigrating to Israel after the establishment of the state. At that time, these youth villages served a third of Israel's children. Alongside the youth village system, a joint school system developed, that encompassed over 400 kibbutzim and which in itself absorbed numerous immigrant children.

While the language of instruction for Jewish pupils is Hebrew, the language of instruction for Arab pupils is Arabic. Israel's Arab and Jewish education systems do not operate on an equal footing. Over the years, various infrastructural and service disparities have been identified between the Jewish and Arab, as discussed below. These disparities have been a persistent source of frustration as well as of action.

In this article, we look at the various levels of Israel's education system, from early-childhood education through higher education, with particular attention to the system's internal contradictions, complexities, and weaknesses, as well as to the major challenges faced at each successive level.

Early-Childhood Education

Israel's Compulsory Education Law calls for compulsory schooling from age 5 through 18. However, when the law was passed in 1949, a year after the founding of the state, compulsory education was defined as 8 years of schooling only, from age 5 to 13. Israeli legislators at the time were divided into two camps – those who attached importance to the age at which schooling was to end preferring that it continue until age 15, and who, due to limited resources, proposed that schooling begin at age 7, and those who attached special importance to preprimary education, insisting that compulsory schooling begin at age 5. The State of Israel was founded in May, 1948, while the Compulsory Education Law was enacted in September, 1949. The latter approach viewed early-childhood education as a preparatory stage crucial to ensuring young children's readiness for primary school and, in particular, to narrowing educational disparities. This approach has found practical expression, particularly since the 1980s, in intensified governmental efforts to encourage the enrolment of children in preschools from age 3 – efforts that have taken the form of state subsidies for early-childhood education in Israel's disadvantaged population sectors. As a result of this policy, 37% of Israeli children, aged 2 and above, 75% aged 3 and above, 86% aged 4 and above, and 95% aged 5 and above were enrolled in preschools in 2006, with a total preschool enrolment of 400 000 between ages 2 and 5.

From a pedagogical point of view, the Israeli early-childhood education system has gone through two distinct stages. During the first stage, the focus was on education for positive social habits, arts and crafts, music, holidays and festivals, hygiene, and linguistic usage (Michaelovich, 1999), while until the late 1980s, the preprimary education system came to be perceived as a framework for preparing children in accordance with the requirements of older pupils in the system. This trend led to the development of a more demanding curriculum that taught basic scholastic skills and strove to ensure a smooth continuum between kindergarten and primary school. What occurred was, in essence, a complete overhaul of the kindergarten curriculum, which was now oriented toward teaching arithmetic, science, art, life skills, and the historical-cultural heritage. The transition from a relaxed pedagogy, free of external policy pressure to a policy with more defined standards and prerequisites generated considerable controversy within Israel's education system that has continued to resonate into the twenty-first century. Despite the pedagogical differences of opinion, Israel's early-childhood education system is generally thought to possess many positive features and has earned a high degree of professional and public esteem.

Primary Education

Israel's primary education system, in which some 800 000 pupils currently participate at 2200 educational institutions, developed in three main stages.

The first stage, which spanned the first two decades of Israel's existence, was characterized by a melting-pot policy. Half of the country's pupils came from immigrant families; in some localities, immigrants formed the decisive majority. In order to promote the development of a homogeneous society and to blur differences and disparities between immigrants and veteran Israelis and between Jews of Western and of Middle Eastern ethnic background, a uniform curriculum was adopted for the entire pupil population. The main goal was to do away with the various educational streams that had existed prior to the founding of the state and to absorb and unite the immigrants under a single uniform curriculum – to ensure the development of one unified people within one state, by means of a single curriculum (Tsabar-Ben Yehoshua and Zilberstein, 1999). Schools and teachers were regarded as partners in the ideological mission to forge a cohesive society. This aspiration to uniformity cast the teaching profession in a conservative mold and impaired its ability to address the tremendous differences that actually existed between different pupil groups (Dror, 2004).

The second period began in the late 1960s, when an approach gradually emerged that called for emphasizing differences in pupil needs over uniformity in the provision of educational services. This approach engendered the policy of a child-centered approach. It led Israel's primary education system to stress upon individually tailored instruction, to encourage independent learning, and to foster the acquisition of inquiry-based learning; it promoted active learning, enrichment of the learning environment, the offering of electives, schedule flexibility and a multiyear curricular structure, democratization of the school social framework, involvement in the community, an emphasis on identifying individual pupil needs, and a diminished reliance on standardized assessments (Bentoitz, 1960; Eden, 1971; Tsameret, 1977; Shtall, 1991; Tsabar-Ben Yehoshua, 1990). Additionally, a trend toward differential allocation of learning resources took hold, in which priority was given to disadvantaged populations (Tsabar-Ben Yehoshua, 1988; Zilberstein, 1984; Tsabar-Ben Yehoshua and Zilberstein, 1999). This in turn created the conditions for large-scale adoption of a school autonomy philosophy – to a point where a third of Israel's schools became self-managing (Volansky, 2003; Dror, 2006; Gibton and Goldring, 2002; Gibton *et al.*, 2000).

The third stage in the development of Israel's primary education system is characterized by contradictions in the realm of pedagogical policy. On the one hand, there was a teacher-empowerment policy and a trend toward trans-

ferring authority to the schools – an understanding of the new information age and accessibility to knowledge and its concomitant recasting of the teacher as facilitator guiding the pupil to knowledge sources, rather than serving as the sole source of knowledge in the classroom. This trend harmonized with the growing tendency toward encouraging pupils to actively build their own individual knowledge maps and toward employing the computer for instructional purposes. At the same time, contradictory pedagogical approaches began to garner support, approaches that called for increased standardization of teaching, learning, and testing methodologies. These approaches drew strength from Israel's medium-to-low rankings on international scales of scholastic performance, and from growing criticism of instructional methods that gave pupils excessive degrees of creative and scholastic latitude. This is reflected in the growing use of strictly defined and rigid instruction frameworks for the entire pupil population, as well as in the development of a culture of high-stakes testing.

Israel's primary schools have, thus, undergone three major changes in pedagogical approach – changes that have, necessarily, affected their organizational and administrative structures. Schools have transitioned from a climate of extreme centralization to one of pedagogical autonomy and self-management, while a current trend toward standardization is restricting the freedom of the teacher and the educational institution and transferring significant pedagogical latitude from the schools back to the central planning authorities.

Secondary Education

At the end of primary education, pupils move on to the secondary education system, divided into junior and senior high schools. In 2007, 610 000 pupils were enrolled in 1588 schools comprising grades 7–12.

Nearly half (47%) of these schools are operated by not-for-profit organizations, while the other half are run by local authorities (38%) or the Israeli government (15%) (Facts and Figures, 2004).

Until 1979, when secondary education became open to the majority of Israel's pupils, a selective system was in place. Graduates of the primary education system were divided into three groups: those headed for academically oriented study in grammar schools, those headed for vocational training at technology schools, and those who entered the workforce directly. A series of policy changes opened the gates of secondary education (through grade 12) to a growing percentage of young people. These measures included: the decision at the end of the 1960s to enact an education reform lengthening the period of compulsory education to age 16 and the period of free schooling to age 18; the establishment of junior high

schools with upgraded curricula; the elimination, in 1973, of screening examination for secondary education; the reformed matriculation examination system, instituted in 1979, which provided for a more varied curriculum reflecting pupil needs, multiple intelligence and differences; and the trend toward differential budgetary allocation as a means of strengthening disadvantaged populations. All of these things contributed to a profound change in the structure of secondary education in Israel (Iram and Schmida, 1998). Over the course of the last three decades, this change has been reflected in three main indices: a greater number of pupils from the middle-to-low range of the socioeconomic scale completing the 12th grade; a greater number of pupils finishing high school with a matriculation certificate; and a greater number of pupils gaining admission to institutions of higher education (Drori, 1999).

These measures, taken both individually and together, led to changes in Israel's educational structure, particularly in terms of offering educational opportunity to disadvantaged populations for whom the gates of learning had previously been closed (Shmueli, 1998; Volansky, 2005). These changes may largely be credited to the increased scope given to schools for defining study disciplines with a high degree of relevance to their pupils and communities, without renouncing the duty to teach and test the entire Israeli pupil population in six or seven core subjects. A wide variety of unique and innovative curricula were developed that had never before been on the education system's agenda: animal care, cosmetology and beauty, music and art, sports, and culture – all in addition to the required subjects comprising Hebrew/Arabic, English, mathematics, the sciences, history, Bible, geography, and civics. This was an era of exceptional creativity and innovation in the education system, one that lives on in memory as a golden age.

This age came to an end during the 1990s, when the education system was forced to yield to growing pressure from Israeli society. The first to oppose the aforementioned educational approach were the universities (Volansky, 1999, 2005), which argued the impossibility of ensuring qualitative matriculation standards in a situation where study disciplines were developed by the schools and not by the state. The second group to oppose the innovative educational approach was the Mizrahi Democratic Rainbow Coalition, a movement involved in guaranteeing the rights of Jews of Middle Eastern (*Mizrachi*) origin. One of the movement's claims was that the education system's vaunted flexibility and variety had actually created a situation in which *Mizrachi* pupils were being channeled toward less-prestigious courses of study, mainly in the technology schools, and were, thereby, being barred from acquiring higher education – in contrast to native Israeli pupils and to those of European/American origin. A third, back-to-basics group called, quite simply, for the elimination of noncore disciplines and for a re-embrace of the basics,

that is, of six or seven uniform subjects for the entire Israeli pupil population. By the early years of the twenty-first century, these pressures had gathered momentum in generating change in the education system (Yogev, 2007). The change consisted primarily of a transition to an academically oriented study, consistent with the back-to-basics approach. This approach has been a source of discord since it began to gain currency during mid-1990s. On the one hand, the overall percentage of Israeli pupils eligible for the matriculation certificate has risen to about 50%. This is, indeed, an impressive improvement, considering that in 1990 only 35% of Israeli pupils were eligible for the matriculation certificate. On the other hand, this contraction of the curricular structure has left the other half of the secondary school population behind. These pupils have been deprived of the kind of flexible educational option that they need in order to complete 12 years of compulsory schooling in an environment that engages them, challenges them, brings their talents and abilities to the fore, and secures them a matriculation certificate so that they can go on to tertiary education.

Brand-naming schools as successful or outstanding and using league tables to rank them have intensified pressures with regard to pupil screening. The slashing of education budgets during the past decade has also resulted in the inability of schools to implement special curricula which by their very nature are more expensive. This development has caused schools to limit their study offerings and to focus overwhelmingly on curricula that require little more than a blackboard and chalk. School networks that developed over decades as frameworks offering a broad variety of curricula spanning the range of academic, technological, and practical subjects, have been forced over the last 10 years to confine the bulk of their instructional activity to scientific/engineering subjects geared mainly for the more academically oriented segment of the pupil population. In effect, school principals have been obliged to intensify their pupil-screening activity as a way of boosting the prestige of their schools, and to adopt curricula that are, in practice, suitable for only the upper 50% of the pupil population. The ramifications of this policy may be seen in Table 1.

Table 1 shows both the strengths and weaknesses of Israeli secondary education. The larger the cohort, the greater the number of pupils who complete 12 years of schooling; but in absolute terms, the number of pupils completing grade 12 without a matriculation certificate adequate for tertiary education has also grown, reaching 54 000 in 2006. This illuminates the significant challenge facing the Israeli education system – that of adapting the curricular structure to make it relevant and meaningful for the lower 50% of its enrollees who, despite the system's success in ensuring that they complete grade 12, nevertheless remain bereft of a ladder of opportunity leading to higher education.

Table 1 Percentage of those completing 12 years of schooling and percentage of those eligible for the matriculation certificate, per cohort, since the founding of the state

Year	Size of cohort ^a	% of cohort members enrolled in grade 12	% of cohort members eligible for the matriculation certificate at the end of grade 12
1949 ^b	12 000	6.7	1.6
1960	25 000	27.0	14.2
1970	52 900	39.7	20.0
1980	65 500	53.4	21.3
1990	85 000	72.0	34.7
2000	103 000	86.0	43.6
2006	107 000	92.0	49.2

^aCohort of 17-year-olds.

^bJewish population only.

Source: State of Israel. *Statistical Yearbooks*, 1986–2007. Jerusalem: Central Bureau of Statistics.

It may be assumed that the coming decade will bring with it a search for balance between the two trends described above, that of instructional standardization and uniformity, and that of variety in educational offerings – variety that encompasses a broader range of intelligences and talents than that currently addressed by the system.

Higher Education

Legal authority for Israel's system of higher education, which in 2007 comprised a quarter of a million students in 65 educational institutions, lies with the Council for Higher Education. A higher education planning and budgeting committee operates within the Council's framework. These bodies, which are responsible for formulating higher education policy, initiated a reform in 1993, which led, over the course of the subsequent decade, to a substantial restructuring of higher educational opportunity in Israel.

For most of the State of Israel's existence, the higher education system was a monolithic structure founded mainly on seven research universities. Starting in the mid-1990s, the system underwent a profound transformation, expanding to include new colleges and evolving into a pluralistic structure that includes 65 institutions offering a broad range of study options. This increase in the number of institutions was made possible in part by the trend toward legitimizing degree conferral in disciplines not normally associated with the research but on professional studies. Five main reasons may be identified for this large-scale opening of the gates of academia. The first of these is the disparity that existed around 1990 between the percentage of those eligible for the matriculation certificate (35%) and those admitted for first-year studies at Israel's institutions of higher learning (20%). A particularly low

rate of admittance to higher education was found among *Mizrachi* Jews and in the Arab sector. The second factor was the growing pressure exerted by professional postsecondary institutions for accreditation to grant academic degrees. These were institutions offering courses of study in tourism services, communications, insurance, optometry, dental hygiene, laboratory technology, architecture and design, and various technological disciplines – as well as teacher-training colleges, only a minority of which awarded academic degrees prior to the reform.

The third factor was the rigorous screening policies that prevailed in the more desirable and prestigious disciplines, such as law, psychology, medicine, computer science, and economics. As the planning echelon of that period felt that the economy's needs in these fields was limited, the number of student places was likewise limited in conformity with projected needs. The fourth factor that led to the development of academic colleges was the rapid increase in the number of students at the universities – 8% – 4 times higher than the annual growth rate of the population (to ensure their stabilization at current levels through 2000). Growing public expenditures on higher education were the fifth factor that came into play. The national expenditure per full-time student (including tertiary and higher education) in Israel was high at the time (US\$11 100) compared with other countries. All of these factors culminated in the decision to develop academic colleges whose cost of maintenance as teaching facilities would be a third lower than that of expanding the universities. A decision of the Council for Higher Education was supported by the Israeli government decision and by a legal amendment passed by the Knesset.

Thus, starting in the mid-1990s and over the course of a decade, Israel's higher education system doubled in size and absorbed nearly one out of every two (42%) members of the high-school graduate cohort, compared with one out of every five at the beginning of the 1990s, mainly due to the development of the academic colleges. **Table 2** illustrates the change in access to higher education:

The rapid growth of the higher education system also had a swift impact on the quality of teaching and learning within the system. Governmental funding in institutional budgets, particularly those of the universities, was cut by 20%; the senior faculty/student ratio, which in the mid-1990s was 1:16, had reached 1:25 by 2006; branches of foreign institutions in Israel were opened; nine new private institutions were established; and there was an increase in the number of public colleges and an overall rise in the number of institutions to 65 in 2007. The structural changes have produced an effect opposite to that anticipated – intensified competition, which led to lowering academic quality rather than enhancing it. Criticism of academic quality and frequent budgetary crises, particularly in the universities, gained momentum and led to the creation of a governmental commission that, in 2007, formulated

Table 2 Percentage of students admitted for first-year academic study, per age cohort^a (1949–2004)

% enrolled in institutions of higher education, per cohort	% eligible for matriculation certificate, per cohort	Size of cohort	Year
1.6	6.7	11 902 ^b	1949
11.6	14.2	25 032	1960
17.8	20.0	52 900	1970
21.3	21.3	65 500	1980
22.9	34.7	85 000	1990
36.2	42.0	106 300	2000
42.1	49.2	113 000 ^c	2006

^aDoes not include students enrolled in the Open University.

^bJews only.

^cMedian age cohort 20–24.

Source: Data of the State of Israel. *Statistical Yearbooks*, 1986–2007. Jerusalem: Central Bureau of Statistics and Council of Higher Education and Planning and Budgeting Committee Yearbooks, 1981–2007, Jerusalem.

recommendations intended to improve the state of affairs. These recommendations included returning the budget to its 2001 level; raising student tuition by 50%; expanding the support system for students from the lower rungs of the socioeconomic ladder; merging existing colleges; recruiting new faculty staff to enhance instructional and research quality; defining the universities as the exclusive venues of research and the colleges as institutions geared solely toward academic teaching; and having an external quality control system.

The commission's recommendations, which will be coming up for government approval in early 2008, are controversial. While the universities regard the recommendations as a potential lifeline, two other factions oppose them: the students, who are unwilling to pay higher tuition, and the colleges, which are unwilling to see research defined as an area of activity in which their instructional personnel have no part. All of the stakeholders are gearing up for the coming battle.

Special Issues

The *Haredi* (Ultra-Orthodox) Sector

Until the early 1990s, the percentage of Israeli pupils studying in the *Haredi* sector was 6%. This figure began to increase in the early 1990s, and by 2006, about a quarter (25.6%) of Israel's Jewish primary education population was *Haredi*. Main features of education in the *Haredi*-school sector include the nonstate status but a public funding of most of its expenses; a focus on religious studies with minimal attention to secular subjects; a sense of alienation from the state; and a tendency to shirk civic duties. This situation has two main consequences: first, many young people who graduate from

the *Haredi* education system find it difficult to gain admittance to institutions of higher education, due to an inadequate education background in secular subjects; and second, this population has obstacles taking its place in the labor force. In a study that was conducted with unemployed *Haredim*, the latter identified their lack of background in English, mathematics, and computer studies as factors preventing them from finding work (Naon *et al.*, 2006). Their limited potential for labor-force participation partly explains why a high percentage of *Haredim* live below the poverty line, and why a high percentage of Israeli children living in poverty belong to this growing sector.

Despite awareness on the part of the *Haredi* spiritual leadership of the community's severe economic plight, and despite increasing calls from within the population, particularly women, to permit secular studies for the purpose of ensuring family livelihoods, the formula has yet to be found that will enable hundreds of thousands of *Haredi* pupils to study secular subjects at the level and scope needed in order to earn a living. To a great degree, the issue is in the hands of the *Haredi* leadership, but it is not their province exclusively. The state has means available to it for generating a new kind of dialog, one capable of striking a new balance between the *Haredi* sector's curricular structure and the budgetary support that the state provides to *Haredi* educational institutions.

The Arab Sector

The educational disparities that exist between Israel's Jewish and Arab populations date back to the prestate period. Jewish organizations were established during the nineteenth and early twentieth centuries to promote Hebrew education in Palestine (Elboim-Dror, 1986). Differences in strength of leadership and budgeting policies led to disparities as early as the British Mandate period, and their effects were felt during the early years of Israel's existence. For example, in 1935 only "20% of Muslim children aged 5–15, versus 80% of Jewish and Christian children," were enrolled in educational institutions (Elboim-Dror, 1986: 36). During the years of Israel's existence, the gaps have narrowed continually. For instance, in 2006, the percentage of those enrolled in grade 12 in the Jewish sector was 93%, compared with 83% in the Arab sector.

Another issue relevant to Jewish–Arab educational gaps is that of the strength of the local authorities. The Arab local authorities' limited economic capabilities make it difficult to collect local rate and to establish complementary educational services, although the government pays teacher salaries and funds most educational services.

The 1980s witnessed a turning point in governmental policy toward the Arab sector. The government initiated 5-year plans to eliminate the deep disparities between the Jewish and Arab sectors. The plans had an

affirmative-action orientation that called for increased investment in Arab education. Although discrepancies between planning and implementation still exist, there has been a gradual trend toward closing the gaps on all measures of inequality. The percentage of school-aged Arabs enrolled in study frameworks has grown faster than that of school-aged Jews; there has been a significant increase in the percentage of Arab pupils eligible for the matriculation certificate; the percentage of Arabs studying in institutions of higher education has grown, although it is still low relative to the Arab presence in the overall population; during the 1990s, priority was given to school construction in the Arab sector; and, no less importantly, in 2003 a new formula was determined for the allocation of teaching hours, giving precedence to the Arab sector over the Jewish one. These corrective measures reached their height in 2007, when the Education Ministry made a politically charged and publicly contentious decision to allow the events of 1948 and their implications for Palestinian society to be taught in accordance with the Israeli Arab community's traditional narrative. The sensitivity surrounding this authorization is rooted in the fact that, while Israel's achievement of statehood – the establishment of a Jewish national home in Eretz, Israel, constituted a tremendous achievement for the Jewish people, for the region's Arab inhabitants, it occasioned a collective trauma, the *Nakba*, inasmuch as many of them left for neighboring countries in the wake of Israel's declaration of independence.

Another issue is that of the development of an Arab educational leadership. From the time of Israel's founding until 1987, education in the Arab sector was administered by Jewish position holders. This had consequences for the development of a senior Arab leadership in the educational sphere, and for the degree to which the sector's educational needs were actually addressed. The Ministry of Education's administrative dominance and its exertion of pedagogical authority were felt in nearly all areas of the education system. This reality led, unintentionally, to a stifling of any original thought that might have encouraged the development of a professional leadership capable of adapting pedagogical programs to Arab-sector needs. This dependence, though it did in some cases generate educational initiatives, nevertheless compromised the Arab-sector leadership's ability to come up with creative programs.

In contrast to this stifling effect of the centralized educational leadership, recent years have seen the emergence of an intelligentsia that has left its mark at every point of the educational map to which it has traveled: inspectors, school principals, professional online forums, or graduate students in education. These individuals have demonstrated professional awareness and a higher degree of administrative ability than anything seen heretofore in the Arab educational sphere. A critical perspective;

educational innovation; needs-assessment capability; a broad educational background; professional leadership ability – all of these attributes have been increasingly evident since the late 1990s in encounters with the new cadre of Arab educators. The paradox of this leadership lies in the fact that it is still trapped in a dependent relationship, with the ideas coming from the Jewish sector. This sense of dependence is still having a paralytic effect and is rendering largely useless the vision, original thinking, talents, pedagogical initiative, and courage of the new intelligentsia. At present, midway through the first decade of the twenty-first century, Israeli Arabs' potential for social integration at middle-class status depends more on the sector's professional leadership and its determination to drive educational initiatives aimed at eliminating disparities, than it does on the establishment itself. Recent years have, unquestionably, witnessed a turning point: philanthropic initiatives to raise funds for education institutes; the establishment of organizations to improve school performance; initiatives on the part of numerous local authorities to increase the percentage of pupils eligible for the matriculation certificate; and an impressive rise in the percentage of Israeli Arabs studying in institutions of higher education (Volansky, 2005, 2007). More than at any time in the past, the Arab sector seems to be aware that the way to upgrade its status in Israeli society is to focus on quality education. The Arab leadership has a decisive role to play in ensuring the continued development of this trend.

The Teacher Crisis

The key to the current teacher crisis in Israel is the teaching profession's relative undesirability, to which several factors have contributed over the years. First, various occupational alternatives to teaching have emerged that are considerably more attractive to young people. These alternative sectors include finance, insurance, tourism, recreational/leisure services, ICT, complementary medicine, etc. Another factor is the harsh public criticism to which the education system is subjected, and the prevailing negative image of the teacher. Articles by key figures in Israeli society that portray the Israeli teacher as a failure, and crude insults to the profession on the part of public leaders have caused this image to become ingrained in the public mind and have deterred young people from joining the teaching profession. A third factor is the expanded sphere of individual, parental, and pupil rights which, albeit unintentionally, has undermined the teacher's authority. Fourth, teacher salaries have eroded to the point where teachers are relying on national-insurance income supports in order to achieve a mandatory minimum income – or taking on additional jobs beyond their professional teaching positions. A fifth factor is the trend that developed during the first decade of the twenty-first century toward a shorter

school day (14% on average), which indirectly contributed to a heavier teacher burden. The final factor is the intensifying public critiques with Israel's performance on international tests.

All of these factors have caused demand for teaching as an academic study discipline in Israel's teacher-training colleges and universities to decline. In 2006, the discipline-demand breakdown in Israel's 27 academic teacher-training institutions is shown in **Table 3**.

Conclusion

Israel's education system is faced with several challenges. The first and most visible of these challenges is the teaching profession's lack of demand. Cumulative damage to the image of the teacher; changes in the Israeli economy's employment structure, including the rise of higher status and better paid occupational sectors; the growing complexity of the teacher's job; and the higher median age of the teacher population – all of these things have led to a situation in which only profound change will draw young people with cultural/educational capital and personal charisma to the Israeli education system's 45 000 classrooms. A second challenge relates to the clear choice of a pedagogic strategy. Over its years of existence, the Israeli education system's standards have traditionally revolved around the matriculation exams administered at the end of grade 12. However, the permeation of high-stakes testing at all educational levels, including the early-childhood level, has created a situation rife with contradictions. Those who support this approach, view it as a major, if not the sole, tool for improving scholastic performance throughout the country. Critics see it as harmful to knowledge-building processes and cultivation of mind by development of thinking skills, and consider it to be a return to mechanistic and formal teaching methods. Those who oppose the standardization process see in the development of ICT a golden opportunity to create challenging new teaching methods relevant to the pupils' world, and they feel that the education system must under no circumstances waste its chance to incorporate ICTs into the teaching and learning process. Hence, today's great challenge is that of choosing a clear pedagogic strategy capable of resolving the existing contradiction.

A third challenge is also connected to the pedagogical sphere – the need to enable the lower 50% of the pupil population to complete 12 years of schooling, based on curricula that will function as a ladder to advancement and give these pupils the requisite background for tertiary education. In order for this to happen, there has to be choice and variety and flexibility in the curricular structure, as in the ideology of variety that came to prominence in the late 1970s – an ideology at odds with the current trend toward a monolithic, academically oriented system

Table 3 First-year student enrolment in primary education programs at teacher training colleges – 2006–2007

Subject	No. of students enrollment (state education)	No. of students enrolment (state religious education)	No. of students enrolment (Arab sector)
History	0	0	0
Judaism	13	22	0
Bible	7	119	0
Talmud	0	20	0
Hebrew language	4	0	0
Hebrew literature	4	0	1
Geography	0	0	0
Israel studies	0	0	1
English	45	3	35
Arabic	0	0	50
Science	45	4	0
Mathematics	87	50	33

Source: Training Teacher Division, Ministry of Education, 2007 (unpublished). The low student enrolments in teacher-training programs explains, at least in part, the current teacher crisis in Israel.

founded on a back-to-basics approach that favors the upper 50% of the population. Unless a different attitude is taken toward curricular structure at the secondary level, it will be difficult to ensure equal opportunity for tertiary education to tens of thousands of high-school graduates.

A fourth challenge lies in the *Haredi* sector. The continual growth of this population, which now accounts for a quarter of all pupils in the Jewish sector and which focuses primarily on religious studies at the expense of secular subjects, is responsible for the plight of the *Haredim* themselves, who are largely unable to integrate into the labor force. This in turn dooms the *Haredim* to economic distress, with most of the population's children living below the poverty line; it also keeps the state from increasing the percentage of its citizens who participate in the labor market (the Israeli figure is 10% lower than that of most developed countries). Any effort in this area will entail new definitions for the relationship between the state and the *Haredi* community, definitions capable of encouraging and supporting secular studies within this sector, if only in certain and minimum core disciplines essential for their integration into the labor market.

A fifth challenge is that of formulating an education budget capable of addressing the system's various other challenges. The budget cuts that Israel's education system has absorbed over the last few years have led to larger class sizes, fewer weekly class hours, a heavier burden on the teacher, a loss of pedagogical flexibility in the schools, a cessation of the trend toward school autonomy

and self-management, and fewer resources for advancing disadvantaged populations, including the Arab sector. These are the challenges faced by the Israeli education system in the middle of the first decade of the twenty-first century.

Acknowledgment

The article was supported by The Israel National Commission for UNESCO.

Bibliography

- Bentoitz, Y. (1960). *Education in the State of Israel*. Tel Aviv: Chachik.
- Czapski, G. and Ilan, Y. (2004). *International Status of Israeli Research: A Comparative Analysis Using Scientometric Indices*. Haifa: Neeman Institute, Technion.
- Dror, Y. (2004). The education system as an agent for Jewish patriotism in Israel: From pioneering Zionism to balanced Israeliness. In Ben-Amos, A. and Bar-Tal, D. (eds.) *Patriotism: Homeland Love*, pp 137–173. Tel Aviv: Deyonon and Hakibbutz Hameuchad.
- Dror, Y. (2006). Past reforms in the Israeli education system: What history can teach us about the Dovrat report. In Inbar, D. (ed.) *Toward Educational Revolution? Bnei Brak: Hakibbutz Hameuchad*.
- Drori, Y. (1999). A historical overview of 50 years of education in Israel: Periods and dilemmas. In Peled, E. (ed.) *Fifty Years of Israeli Education*, pp 39–66. Tel Aviv: Ministry of Defense.
- Eden, S. (ed.) (1971). *On New Curricula*. Tel Aviv: Ma'alot.
- Elboim-Dror, R. (1986). *Hebrew Education in Palestine*. Jerusalem: Yad Ben Zvi.
- Facts and Figures (2004). Ministry of Education, Culture and Sport. 43.
- Gibton, D. and Goldring, E. (2002). The role of legislation in education decentralization: The case of Israel and the United Kingdom. *Peabody Journal of Education* 76(3&4), 81–101.
- Gibton, D., Tsabar, N., and Goldring, E. B. (2000). How principals of autonomous schools in Israel view implementation of decentralization and restructuring policy: Risks, rights and wrongs. *Educational Evaluation and Policy Analysis* 22(2), 193–210.
- Iram, Y. and Schmida, M. (1998). *The Educational System of Israel*. Westport, CO: Greenwood Press.
- Michaelovich, R. (1999). Pre-primary education in Israel: The development and application of varying approaches. In Peled, E. (ed.) *Fifty Years of Israeli Education*, pp 80–859. Tel Aviv: Ministry of Defense.
- Naon, D., King, J., and Wolde-Tsadick, A. (2006). *Populations Not Fully Participating in the Labor Market: Extent, Characteristics and Programs to Promote their Employment and Occupational Mobility*, pp 36. Jerusalem: Meyers-JDC-Brookdale Institute.
- Shmueli, A. (1998). The Israeli education system. In Yavin, H. (ed.) *Fifty Years of Israeli History*, pp 462–470. Jerusalem: Zahir.
- Shtall, A. (1991). *Preparing Learning Materials for Heterogeneous Class*. Jerusalem: School for Senior Teachers.
- Tsabar-Ben Yehoshua, N. (1988). *Metamorphosis of a Curriculum*. Tel Aviv: Yachdav.
- Tsabar-Ben Yehoshua, N. (1990). Planning school curricula: The concept, and the risks and dangers inherent in its implementation. In Friedman, Y. (ed.) *Autonomy in Education: Conceptual Frameworks and Implementation Processes*, pp 130–144. Jerusalem: Szold.
- Tsabar-Ben Yehoshua, N. and Zilberstein, M. (1999). *From Uniformity to Diversity: Curriculum Development*, pp 194. Tel Aviv: Ministry of Defense.
- Tsabar-Ben Yehoshua, N. and Zilberstein, M. (1999). From unity to diversity: The development of curricular planning. In Peled, E. (ed.) *Fifty Years of Israeli Education*, pp 193–203. Tel Aviv: Ministry of Defense.
- Tsameret, T. (1997). *Across a Narrow Bridge*. Sde Boker: Ben Gurion Heritage Center and Ben Gurion University.

- Volansky, A. (1999). The dialectic between center and periphery in the education system of Israel. In Peled, E. (ed.) *The Jubilee Book on the Israeli Education System*, pp 283–300. Tel Aviv: Defense Ministry.
- Volansky, A. (2003). From experiment to educational policy: The transition to school-based management in Israeli schools. In Volansky, A. and Friedman, A. I. (eds.) *School-Based Management – An International Perspective*, pp 217–232. Jerusalem: MOE.
- Volansky, A. (2005). How to shatter the glass ceiling: Issues in education policy. In Rosenthal, R. (ed.) *An Israeli Agenda for Social Change*, pp 328–348. Jerusalem: Keter.
- Volansky, A. (2005). *Academia in a Changing Environment – Higher Education Policy of Israel 1952–2004*. Bnei-Brak: Hakibbutz Hameuchad and Neeman Institute.
- Volansky, A. (2007). School autonomy for school effectiveness and improvement: The case of Israel. In Townsend, T. (ed.) *International Handbook on School Effectiveness and Improvement*. Dordrecht: Springer.
- Yogev, A. (2007). Ideology, pedagogy and education policy in Israel. In Aviram, U., Gal, J., and Katan, J. (eds.) *Formulating Social Policy in Israel: Trends and Issues*, pp 131–148. Jerusalem: The Taub Center for Social Policy Studies.
- Zilberstein, M. (1984). The place of the teacher in educational planning in Israel. *Iyunim BeChinuch* 40, 131–150.

Further Reading

- Cohen, A. (1999). Fifty Years of primary education. In Peled, E. (ed.) *Fifty Years of Israeli Education*, pp 881–893. Tel Aviv: Ministry of Defense.
- Shohamy, E. (2006). *Language Policy: Hidden Agendas and New Approaches*. London: Routledge.

Relevant Websites

- <http://ed.sjtu> – Academic Ranking of World Universities.
- <http://www1.cbs.gov.il> – Central Bureau of Statistics.
- <http://www.cbs.gov.il> – Central Bureau of Statistics.
- <http://www.che.org.il> – Council for higher education.
- <http://www.haaretz.co.il> – Haaretz.co.
- <http://www.mof.gov.il> – Ministry of Finance.
- <http://cms.education.gov.il> – The state of Israel: Ministry of Education: Ministry of Structure.

Jamaica

E Hannam, Jamaican National Commission for UNESCO, Kingston, Jamaica

© 2010 Elsevier Ltd. All rights reserved.

Jamaican Education System

Jamaica is an island nation in the Caribbean. It is part of the Greater Antilles, with a population of approximately 3 million. It was a colony of Great Britain and became independent in 1962. It is the most popular Anglophone country in the Caribbean/North American region, after the United States and Canada.

The Jamaican education system for the masses was established after the emancipation of the slaves. During the period of slavery, there were few schools to educate the local population as the plantation owners sent their children back to England to pursue education. After emancipation, the Indian Commission granted a sum of money to establish elementary schools. The churches also played an important role in the early beginnings of the public school system.

Presently, education in Jamaica is administered primarily by the Government through the Ministry of Education. The involvement of churches and trusts has, however, continued. Hence, there are two types of provisions – public (Government) and private (churches, trusts, and individuals).

The education system consists of four levels:

- Early childhood
- Primary
- Secondary
- Tertiary

Early Childhood

This level caters to children between the ages of 3 and 5 years. There are both public and private schools at this level. There are two types of private schools, namely, basic schools, which are individual- or community-managed and preparatory schools, which are managed by churches or individuals. There are 136 preparatory schools.

The public schools are called infant schools and are fully funded by the State. There are 29 of these schools.

Primary Schools

These schools cater to children between the ages of 6 and 12 years. Primary schools are funded by the State. There are also privately funded schools (preparatory) at this level. Children who attend primary school are mostly

from the middle and lower SES classes of the society while the majority of those who attend preparatory schools are from the upper middle and upper SES groups.

There are 892 public institutions which offer primary level education (grades 1–6).

Promotion from primary to secondary schools is through the Grade Six Achievement Test (GSAT), which is administered by the Ministry of Education annually. Children in grade 6 of primary and preparatory schools sit the test and are placed in secondary schools based on parents choice, performance, and geographic location. Children with the highest performance are normally placed in what are regarded as the “better schools.”

High Schools (Secondary)

High schools cater to students between ages 12–18 years (grades 7–13). Up till recently, secondary education was provided in two types of secondary institutions, high schools and new secondary schools. The high schools were selective schools, meaning that students entered the schools through a selective examination at the end of the primary level. The new secondary schools were nonselective. Students entered these schools from neighboring primary schools (feeder schools), without having to sit a selective examination.

Currently, all schools offering grades 7–11 education are called high schools. Students enter these schools through the GSAT which, as mentioned before, is taken by children in grade 6 of primary and preparatory schools.

There are 155 public high schools with a total enrolment of 236 410 (2005/6), whereas there are 85 private high schools with an enrolment of 9456. Among the public high schools are two agricultural schools and 14 technical high schools. The majority of students at the end of grade 11 sit the Caribbean Secondary Examinations Certificate (CSEC) Examination administered by the Caribbean Examinations Council (CXC). Students at the end of grades 12 and 13 sit the Caribbean Advance Proficiency Examination (CAPE) which matriculates them into the regional and extra-regional universities.

Tertiary

Tertiary-level education is offered in teacher education colleges, community colleges, and universities. There are

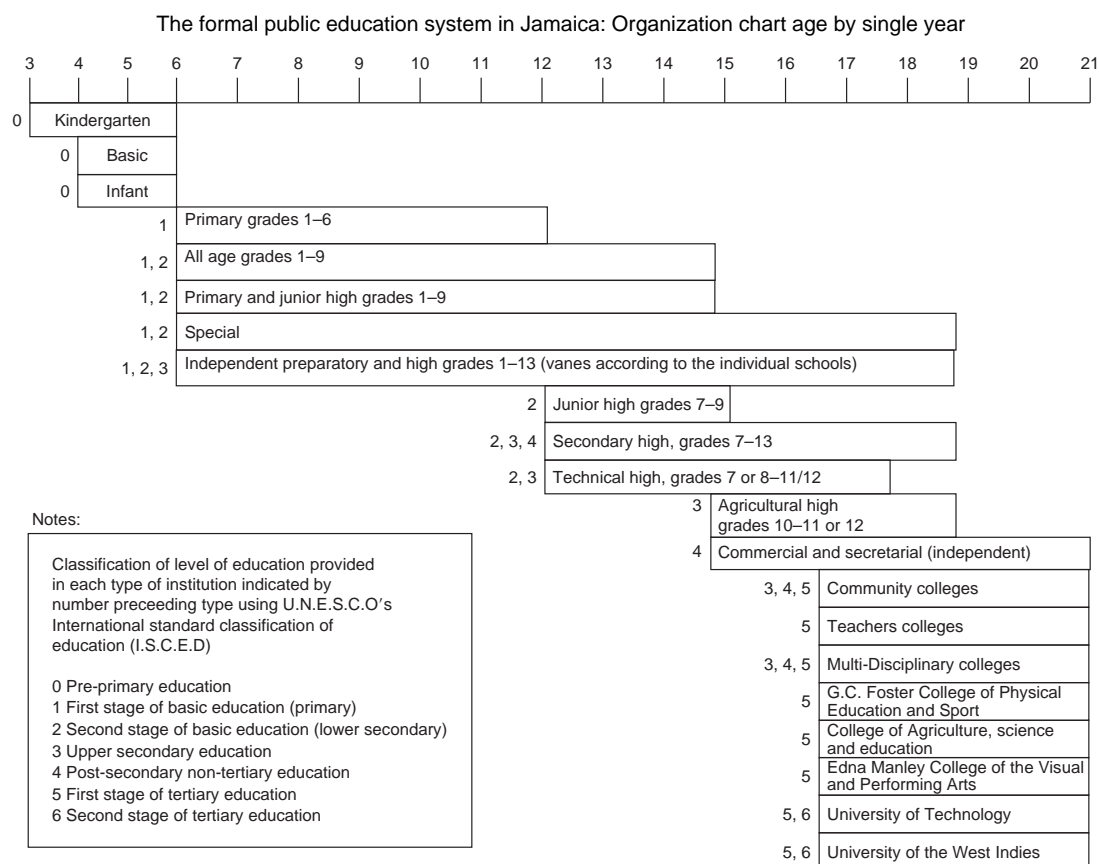


Figure 1 The formal public education system in Jamaica Organization chart age by single year.

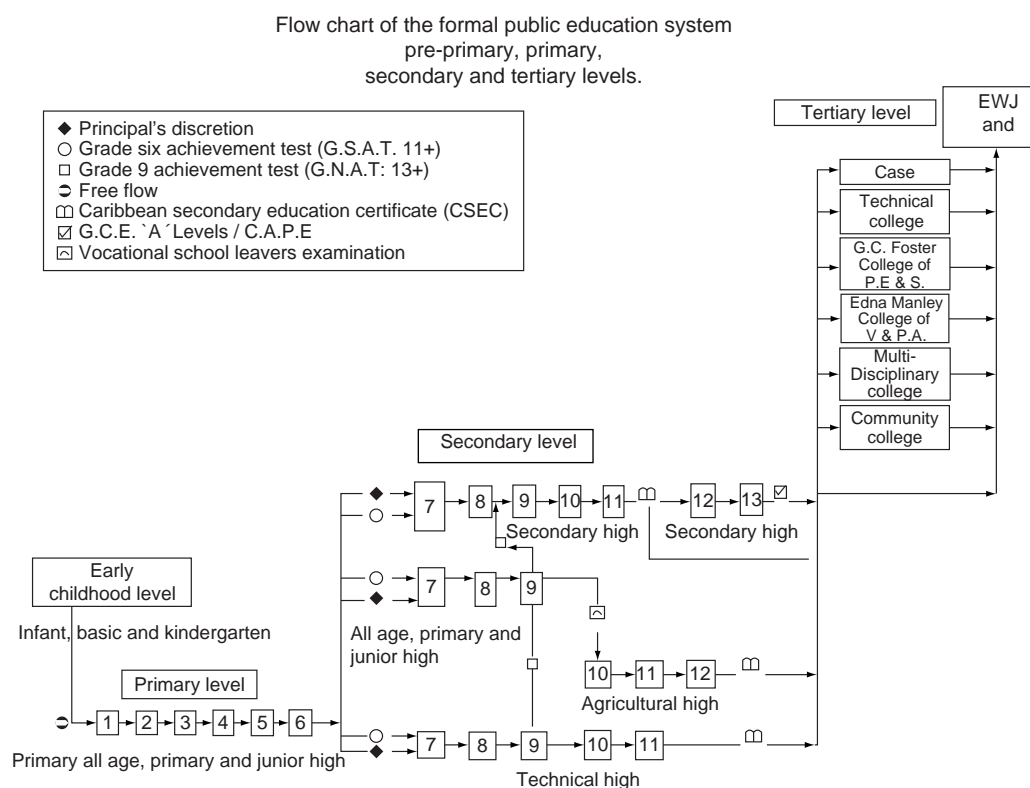


Figure 2 Flow chart of the formal public education system are primary, primary secondary, and tertiary levels.

five teacher education colleges, four community colleges, three multidisciplinary colleges, and two universities. The above mentioned are all public institutions. There are two local private universities. A number of overseas universities offer tertiary-level programs on the island.

There are 59 137 (2005/6) students enrolled in institutions offering tertiary-level education.

Latvia

A Joma, State Education Development Agency, Riga, Latvia

© 2010 A Joma. Published by Elsevier Ltd. All rights reserved.

General Background

Latvia is a small country in north-eastern Europe, on the coast of the Baltic Sea. The landscape of the country is marked by lowland plains, rolling hills, and thousands of rivers and lakes. Most of the territory of 64 589 km² or 24 937 mi² is less than 100 m above sea level. Latvia has a moderate oceanic climate with average summer temperatures of +20°C and winter temperatures of −5°C, but extremes of +30°C and −25°C or more are not unusual. Latvia is not richly endowed with natural resources, so its future depends in large measure on its intellectual capacity, which is regarded as the nation's greatest resource.

The name Latvia originates from the ancient Latgalians, one of the four eastern Baltic tribes that formed the ethnic core of the Latvian people (c. AD 8th to 12th century). The region subsequently came under the control of Germans, Poles, Swedes, and finally, Russians. Latvia is a relatively new national state, founded on 18 November 1918. Latvia was annexed by the Soviet Union in 1940 and regained its independence only in 1991. Currently, the population of almost 2.3 million inhabitants (in 2006) comprises Latvians, 57.7%; Russians, 29.6%; Byelorussians, 4.1%; Ukrainians, 2.7%; Poles, 2.5%; Lithuanians, 1.4%; and others, 2%. As of 2004, Latvia is a member country of the European Union. It is a parliamentary democracy. The state language is Latvian, which belongs to the Indo-European family of languages.

Since regaining independence, Latvia has made considerable progress in the stabilization and restructuring of its economy from centrally planned to free market. Reforms undertaken in the 1990s have strengthened the private sector. High domestic demand promotes development of services, especially trade and construction. The service sector, which contributed only 32% to the total value added in 1990, had more than doubled to 71% by 2002. In 2007, the total gross domestic product (GDP) for Latvia was slightly under 11.3 million LVL or 4932 LVL per capita (up more than 10% from 2006). Unemployment was at 6.8%.

In 1211, Christian missionaries from Germany founded the first school in the territory of modern-day Latvia. From the thirteenth to the sixteenth century, German was almost exclusively the language of instruction. Schools first started providing Latvian language education in the mid- to late sixteenth century. The purpose was to spread literacy among the Latvian serfs, to promote knowledge of religious literature, including the Bible – translated into Latvian by

Ernst Gluk in 1689. The first textbooks in Latvian were alphabet books, the oldest of which were published in 1683.

From 18 November 1918, with the foundation of an independent Latvian state, the guaranteed right to obtain all forms of education in Latvia was established. From the end of 1919, compulsory free primary education was made compulsory by law for children aged 8–14 (pre-school from age 7). The state also promoted the establishment of schools for ethnic minorities, with teaching imparted in their native language. All types of schools were maintained by municipal or state authorities, and there were private teaching institutions as well. The University of Latvia was established in 1919. Overall, in the inter-war period (1918–40), a modern, unified system of education developed in Latvia, consisting of primary education, secondary education, specialist secondary education, vocational education, and higher education. In general terms, this structure of education is still retained today. Currently, overall instruction is given in Latvian, which is the recognized official language. Latvia also has state-financed ethnic minority schools or classes at the primary through to the upper secondary level where courses are partially presented in Belarussian, Estonian, Hebrew, Lithuanian, Polish, Roma, Russian, and Ukrainian. State-funded higher education programs are taught in Latvian.

Goals of the Education System

According to the Education Law (1998; last amended in January 2007), the National Education Standards provide the strategic goals and main tasks of compulsory curricula, syllabus, basic principles, and procedures for the assessment of education acquired by a pupil.

The long-range conceptual framework document approved by the Parliament in 2005 *The Latvian Growth Model: People Take First Place* proscribes a person-centered model for growth in Latvia. An educated and knowledge-based society becomes the key to internal and external national security. Thus the main goals of the Latvian education and training system are to support the building of a knowledge-based, democratic, and socially integrated society, in order to increase the competitiveness of the Latvian population and economy, and simultaneously preserve and develop the cultural values typical of Latvia.

An analysis of performance in the education sector during the previous planning period (2002–05) indicated a shortage of preschool education providers, unsatisfactory

quality of basic education, low prestige of the teaching profession, poor fit between the education and training profiles offered by vocational and higher education and labor market needs, as well as barriers to access. In order to address these problems, as well as taking into account the new long-term model for national growth and the national development plan, a new basic policy framework for education was approved in September of 2006 wherein particular attention is devoted to the quality of pre- and primary school education, the modernization of vocational training, and the competitiveness of higher education. Emphasis is placed on modernizing the equipment and facilities for acquiring practical skills needed in the labor market. Effective measures for ensuring support to children with special needs and those facing social risk include capacity building for education managers and teachers, as well as increasing cooperation between schools, parents, and society at large.

Structure and Operation of the Education System

Preschool and Early-Childhood Care

While preprimary education was optional until the 2001–02 school year, it became mandatory on 1st of September 2002 for 5- and 6-year-old children. The objective of preschool education is to foster the overall development of children and to prepare them to enter the primary stage of basic general education. It is provided by various preprimary-education institutions, kindergartens (nurseries), or at special preprimary classes in general-education institutions. Preschool education is considered the first stage of general education and all children should complete it by the age of 7. In special cases where the child suffers from certain health or psychological problems, preschool education can be continued, based on either parental request or doctors' recommendations, until the age of 8. The entire teaching process is organized through various targeted games.

Basic (Primary and Lower Secondary) and Upper Secondary Education

Children in Latvia are obliged by law to attend school from the year they turn 7 till the age of 16. In some special cases, the acquisition of basic education may continue till the age of 18. In 2006, there were 974 general education day schools. Around 5.4% of schools were primary (grades 1–4), 48.6% were basic schools (grades 1–9), and the rest were combined basic/upper secondary (grades 1–12) schools. The number of school-age children has remained fairly constant at about 226 000 annually, with around 26 760 teachers. In the next 6 years, the number of school children will decrease significantly (even up to

50% of the current total) due to low birth rates toward the end of the 1990s.

The basic school day comprises of 5 lessons in grades 1–3 and up to 8 lessons in grade 9. The school year is 34 weeks per year in first grade and 37 weeks in ninth grade. The compulsory curriculum includes four subject areas: Introduction to technologies and science; languages; art; and man and society. The curriculum for comprehensive education is defined by 20 subject standards. In 2006, the average number of students per full-time teacher in basic education was ten. After finishing basic education, most young people continue their education at upper secondary level in general upper secondary schools, while about a fourth attend technical and further-education schools.

General upper secondary education normally covers 3 years (grade 10–12). The school year comprises 35 weeks in grades 10 and 11 and 38 weeks in grade 12 with not more than 36 lessons per week and 8 lessons per day. The compulsory curriculum of the 3-year general upper secondary schools is determined by the national standard in the four standardized educational profiles: (1) comprehensive education, without intensive teaching of any particular subject; (2) the humanities and social sciences program; (3) the mathematics, natural science, and technical-science program; and (4) the vocational program, where the general-education curriculum places emphasis on subjects in these particular fields.

Educational standards and regulations regarding assessment of learning achievement for compulsory school and general upper secondary school are drafted by the Centre for Curriculum Development and Examination, which is a national administrative body reporting directly to the Minister of Education and Science, and are approved by the Cabinet of Ministers. Learning achievements of school children are assessed through exams organized both locally at school level and centrally at the national level.

Other Schooling Arrangements

Children may attend vocation-oriented education in arts and music or other areas of personal interest in addition to attending regular school. The so-called little art and music schools are generally founded and maintained by municipal governments.

This type of education is voluntary and provides for a person's individual educational needs and wishes. This education does not lead to a professional qualification.

The number of students enrolled in special-interest education programs has been increasing annually. In the 2001–02 school year, 240 273 students were enrolled, but in 2005–06, 293 527 or 92% of the total number of students had enrolled in general, vocational, and special-education schools.

Special education schools or special education classes within general education schools provide education for children with special needs that corresponds to each individual's physical and mental condition. The structure of special education is very similar to that of the mainstream education, providing opportunities for persons with special needs to attain general knowledge and skills with a strong emphasis on their applicability, thus facilitating social inclusion. There are 64 special education schools in Latvia. Almost all of these schools provide dormitory housing for their pupils.

Vocational and Further Education

In 2006–07, there were 92 nationally accredited vocational-education schools in Latvia with a total enrolment of 40 439 students. The schools employed 4824 teachers, 70% of whom were women. The majority of vocational schools in Latvia provide 2, 3, and 4-year vocational education and training programs at upper secondary level, and only some programs are designed for basic vocational and training purposes.

The national standard of vocational education and the occupational standards determine the curriculum content of vocational education programs. Practical training on the job comprises at least 20–25% of these programs. The theoretical part provides the knowledge needed in modern society: a command of languages, mathematics, natural sciences, and social studies, and good communication skills. In addition, vocational training also underlines the following skills: entrepreneurship, global environmental awareness, internationalization, and citizenship. Sample curricula are drawn up by the Vocational Education Administration, an institution under the authority of the Ministry of Education and Science. Both occupational standards and sample curricula are developed in cooperation with professional associations representing different economic sectors. Based on the relevant standards and sample programs, schools develop their own training programs in accordance with the needs of the local labor market.

Basic vocational and training programs are intended to provide training in simple trades for pupils who have not finished compulsory 9-year basic education by the age of 16. By successfully completing this educational program, the pupil obtains a basic education certificate and may go on to upper secondary education. The number of students enrolled in these programs is not large – only about 2.3% of the total. The aim of this form of education is to prevent marginalization of pupils facing social risk.

Vocational education programs lasting 2 or 3 years provide the theoretical and practical knowledge required by a skilled worker. Even on completion of such a vocational

education program, the student has not obtained a full upper secondary education. For this, an additional year of schooling is required. The proportion of students enrolled in these vocational programs is 18% of the total.

Upper secondary vocational education programs covering 4 years provide much wider theoretical knowledge and practical skills. This is a full secondary education program, so the school leaver is entitled to enter higher education. Approximately 70% of students in vocational education choose upper secondary programs.

In 2006, there were 10 464 graduates of vocational-education programs. Of these, 43% had qualifications in engineering (including processing and technologies), 24.7% in the services, 14% in commerce, 5% in information technologies, 5% in humanities and arts, 4% in healthcare, 1.8% in agriculture, and 2.5% in others.

Higher Education

Latvia has two types of higher education programs – academic and professional. Such a division of higher education allows the student to choose either research or professional activity in the future. Professional higher education is divided into two levels. First-level professional higher education programs cover 2–3 years after upper-secondary education and are provided by colleges. Second-level professional higher education programs cover at least 4 years after upper secondary education or 2 years after acquiring a bachelor's degree (3-year studies). These programs, as well as higher academic education programs are offered at universities or nonuniversity-type higher education institutions.

The right to enter a higher education institution is enjoyed by all those who have completed general upper secondary education or graduated from a 4-year vocational program.

In Latvia, there are both state-financed and private higher education institutions. In 2007, of the 60 accredited higher education institutions, 34 were university/nonuniversity-type institutions (19 of them government-funded and the rest funded by other legal entities or private individuals) and 26 colleges (18, state and eight private) (see **Table 1**).

Of the 43 860 students who enrolled in 2007, 22% enrolled in postgraduate programs, and 1982 were doctoral students. Of the first-year students, 13% had completed upper secondary vocational education, while 87% had completed general upper secondary education. There is no distinct disproportion between city and country dwellers. The total number attending state and private higher education institutions and colleges was 129 497. Almost 63% of students are doing full-time courses. There were 4926 full-time equivalent academic staff members, of whom 23% were professors. In Latvia, there are 568 students per 10 000 inhabitants, which is the second largest student

Table 1 Educational providers, students, and teaching staff, 2006–07 academic year

<i>Provider</i>	<i>Institutions</i>	<i>Students</i>	<i>Pedagogical staff</i>
Total schools	1008	279 872	33 937
Day schools	974	266 111	30 138
Evening schools	34	13 761	1058
Special schools	64	9376	2741
Elementary (grades 1–4)	53	<i>Day/evening/special</i> 72 669/1/3021	1031
Primary (grades 5–9)	474	128 511/2156/5184	10 239
Secondary (grades 10–12)	383	55 555/11 591/11 171	18 868
Males	-	139 732	4995
Females	-	140 140	28 978
VET	92	40439	4824
Males		23 662	1411
Females		16 777	3413
Higher Education	60	129 497 ^a	4926
State budget	37 ^b	30 172	
Private/own funds	23 ^c	99 325	

^a82 771 higher education students, and of them, 63.9% are females.

^bOf these, 19 are university-type and other higher education institutions, and 18 are colleges.

^cOf these, 15 are university-type and other higher education institutions, and 8 are colleges.

Adapted from Latvia, Central Statistical Bureau (2007). *Izglītības ikgadējie dati (Annual Data on Education)*. Central Statistical Bureau.

proportion in the world. Foreign students comprised only 1.1% of the total enrolment in 2007. Programs for foreign students are mainly offered in English or Russian.

Adult and Nonformal Education

Adult education in Latvia is supported by the principle of lifelong education. The Education Law (1998) specifically concerns the development of adult formal and nonformal-education programs. Arrangements for the provision of adult formal education are set by the Education Law, Vocational Education Law, Higher Education institutions Law, and other statutes and regulations.

Different types of further education and training are offered to persons after graduating from general upper secondary or 2–3-year vocational education and training programs. These programs are focused on mastering professional skills and knowledge in line with the requirements of the respective qualification level. The training process and assessment of achievements are organized in a manner similar to that in vocational secondary education and training programs. In 2006–07, 290 279 persons were involved in further education; of these, 24.7% received support from their employers to cover training costs.

Labor market training for the unemployed accounts for the largest number of participants in the adult-education sector. Employment authorities provide a wide range of labor market training mainly targeted at unemployed adults. The courses are free of charge for persons who qualify for official unemployment status.

Adults can also follow a complete program of basic education or upper secondary education at so-called evening schools.

The Teaching Profession

Teachers in Latvia are trained at five higher education institutions. Two training routes can be taken. The most common is a professional bachelor's degree program covering 4 years, which provides a teaching qualification for a specific level of education (preschool, primary, and secondary) and, for secondary school teachers, a specific subject area. Preschool and primary school teachers are qualified to teach all subjects. The second route requires two stages – a bachelor's degree (3 years) in education sciences, plus an additional 2 years of study in a second-level professional program of studies to qualify as a teacher in a specific level of education and/or subject area. Vocational school teachers generally have a professional diploma in a vocational area with an additional qualification in vocational teaching.

Currently, the teaching profession is not attractive due to low social prestige and relatively low salary levels. Of the total number of teachers in the general education sector (grades 1–12) 88% are women (70% in the vocational-education sector). Additionally, 9.09% of teachers currently employed are at pension age. In 2000–01, 19% of teachers were 30 years old or younger; in 2004–05, the proportion is only 15%. However, considering the decreasing numbers of school-age children in the upcoming 6 years, this should not create a drastic shortage of teachers.

Administrative and Supervisory Structure

The educational system is administered on three levels – national, municipal, and institutional. The parliament,

cabinet of ministers, and ministry of education and science are the main decision-making bodies at the national level. The ministry as the leading public administration institution in the field of education and science implements a single national policy and development strategy in education, develops education, science, sport, youth and state-language policies, organizes and coordinates their implementation, in addition to developing draft-regulatory legislation regulating the sector, and draft-policy planning documents.

Every municipality supervises the preschool, basic, and general upper secondary schools located in its administrative territory and participates in funding the maintenance expenses of those educational institutions. A board of education is established by each municipality to perform its education-related functions.

Vocational education and training schools are mainly maintained and supervised by the Ministry of Education and Science in strong cooperation with branch ministries and social partners. Only art and music schools are placed under the authority of the Ministry of Culture. Private and municipal schools must be registered with the National Registrar of Education and comply with government education standards.

In the tertiary sector, decision-making, regulation, funding, and governance are shared between the government and the institutions themselves. Higher education institutions are autonomous bodies established under national legislation, each with its own independent governing body (senate). The primary institution responsible for higher education is the Ministry of Education and Science, which administers government funding, policy, and programs. While higher education institutions are autonomous, they are accountable to the government via accreditation.

Educational Finance

The national education budget in 2007 was slightly over 172 million LVL, up by 27.5% in absolute figures. The proportion of the education budget within the national budget was 4.89 % (4.79% in 2006).

Public primary and secondary general education in Latvia is free and is financed from the municipal budget. At primary and secondary schools, the state pays teachers' wages, while the local authority finances the maintenance of the school building and grounds and covers other expenses connected with teaching. In 2007, the government subsidy to general education schools was almost 5.36 million LVL, or 3.11% of government expenditure on education.

Vocational schools in Latvia usually belong to the state or the local authority, which accordingly finances both teaching and maintenance of the school. Government

spending on vocational education in 2007 amounted to slightly over 56 million LVL, or approximately 32.6% of the education budget.

Private educational establishments at primary–secondary level may set their own tuition fees. Accredited private lower and upper secondary schools receive a subsidy from the national government budget to cover teachers' salaries and related social-security taxes in order to fund the provision of accredited lower and upper secondary education programs. Accredited private vocational education schools receive a national government subsidy based on a training agreement between a government ministry and the training institution for the training of a fixed number of persons in a specific vocational area.

The Latvian state finances all the higher education institutions belonging to the state. In 2007, government expenditure on higher education was slightly over 111 million LVL, or 64.3% of the education budget. The Higher Education Council allocates a certain number of state-financed study places in each field of studies. In 2007, there were 30 172 (23.3%) students funded by the state budget. Those students who pass their entrance examinations at state higher education institutions, but whose marks are not sufficiently high to grant them state-supported education, can take up studies as fee-paying students. Similarly, fees are charged at all private higher education institutions.

Adult education can be financed from the government or local government budget, employers' resources, learners' resources, donations and grants, as well as from other sources. There are different institutions on national and local levels involved in the organization and provision of adult education. The tuition fee regulations for continuing or inservice education and training are established by training providers and covered by individuals or employers. Further training or retraining of persons who are legally classified as unemployed is financed from the state budget.

Performance Monitoring, Evaluation, and Research

The laws on education, general education, vocational education, and higher education prescribe the monitoring and assessment of each respective subsector of the education system by relevant national and local government authorities. Every education institution, regardless of whether it is funded by national or local government or private concerns, must be registered with the National Registrar of Education. In order to begin providing education services, a license must be obtained. The right to deliver nationally recognized diplomas and certificates is only conferred on education institutions accredited by the respective national institutions (the State Agency for Quality Assessment in General Education, the

Vocational Education Administration, and the Higher Education Quality Evaluation Centre).

Statistics collected by the Ministry of Education and Science comprise 182 data tables on general education. This makes it possible to compare performance levels in Latvia with those of other countries in Europe, using the same indicators. The assessment indicators developed by European Centre for the Development of Vocational Training (CEDEFOP) are applied in evaluating vocational education in Latvia. Preparation of annual statistical reports and policy-planning documents on higher education is based on internationally comparable indicators.

The Institute of Education Research at the University of Latvia Faculty of Education and Psychology has participated in 10 major international comparative research projects as a member of the International Association for Educational Assessment since 1992, as well as in the Organization for Economic Cooperation and Development (OECD) Programme for International Student Assessment (PISA) in 2000–2003–2006. A comparison of the PISA results over the three assessment cycles indicates that while there was a significant improvement in reading skills from 2000 to 2003 with a decrease in the percentage of students at or below level 1, from 30.6% to 23.7%, the average score in reading has dropped in 2006 (from 490 points to 470). Math skills, assessed in 2003 and 2006, show a slight improvement – from 482 to 484 points on average. The very small proportion of students performing at top levels (0.3% in science, 1.1% in math, and 4.5% in reading) draws Latvia into a comparative level below OECD and European averages.

Major Changes and Issues Since the 1990s

The break up of the Soviet Union and the restoration of independence in 1990 has generated a considerable transformation in Latvia's system of education. The enrolment ratio in universities has doubled. The Latvian language has been restored to its former prestige and found new utility. The process of modernizing syllabi and teaching standards has begun. Textbook content has been revised. All of these changes have been informed and guided by new legislation and new administrative decrees, and supported by public investment which, in real terms, was 16% higher in the year 2000 than it had been 10 years earlier. Recent national budget endowment in education has also increased.

Local governments have become active in the improvement process of educational institutions through development and implementation of regional, county, and community-level education policy. The facilities and equipment at education institutions have been modernized, information technologies are widely used, and

teaching methods and forms have been enriched. The National Tri-partite Cooperation Council has established a subcommittee on education and employment which approves professional standards and is involved in the accreditation of vocational and higher education programs. The Ministry of Education and Science has established a fund for distributing student loans.

Changes have occurred that will irreversibly impact the development and improvement of the quality of education. Based on these criteria, among others, it can be said generally that education in Latvia has undergone a successful transition from a system effective for a planned economy to a system effective for a market economy and a democratic society.

See also: Characteristics of Adult Learning; Educational Measurement: Overview; Financing of Adult and Lifelong Learning; Germany; National Assessments; Overview of Lifelong Learning Policies and Systems.

Further Reading

- Latvia, Academic Information Centre (2003). Education in Latvia's transition: The challenge of management. *Report on Education in Latvia 2001/2002 Executive Summary*. Riga: Academic Information Centre.
- Latvia, Cabinet of Ministers (2006). *Izglītības attīstības pamatnostādnes 2007–2013 (Policy Framework for the Development of Education 2007–2013)*. Cabinet of Ministers.
- Latvia, Cabinet of Ministers (2007). *Mūžizglītības politikas pamatnostādnes 2007–2013. gadam (Policy Framework for Lifelong Learning 2007–2013)*. Cabinet of Ministers.
- Latvia, Central Statistical Bureau (2007). *Izglītības ikgadējie dati (Annual Data on Education)*. Central Statistical Bureau.
- Latvia, Ministry of Education and Science Higher Education Department (2008). *Pārskats par Latvijas augstāko izglītību 2007. gadā (skaitļi, fakti, tendences) (Report on Higher Education in Latvia 2007 (Figures, Facts, Trends))*. Ministry of Education and Science.
- Latvia, Vocational Education Development Agency (2006). *Education and Career Guidance in Latvia*.
- OECD (2007). PISA 2006 results. *Executive Summary*. Paris: OECD.
- Republic of Latvia (1999). *Vispārējās izglītības likums (Law on General Education)*.

Relevant Websites

- <http://www.csb.lv> – Central Statistical Bureau, Latvia.
- <http://www.isec.gov.lv> – Curriculum Development and Examination Centre, Latvia.
- <http://www.fitforeurope.info> – Fit For Europe: Education, Study and Career Information on Europe.
- <http://www.enic-naric.net> – Gateway to Recognition of Academic and Professional Qualifications.
- <http://www.aiknc.lv> – Higher Education Quality Evaluation Centre, Latvia.
- <http://www.eurydice.org> – Information Network in Education in Europe.
- <http://www.li.lv> – Latvian Institute.
- <http://www.izm.gov.lv> – Ministry of Education and Science of the Republic of Latvia.
- <http://www.viaa.gov.lv> – State Education Development Agency, Latvia.
- <http://www.izmpia.gov.lv> – Vocational Education Administration, Latvia.

Lebanon

N A Mneimneh, Ministry of Education and Higher Education, Beirut, Lebanon

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Gross enrolment rate (GER) – Calculated by dividing the total number of students enrolled in a level or stage of education by the wider population of that age group corresponding to that level or stage.

Leadership development program – A professional development program designed for school principals and training-center directors to improve leadership, motivational, and management skills, together with improved understanding of international trends in competency-based contextualized learning.

Medium-term expenditure framework (MTEF) – A multiyear budget that also brings in key performance indicators and an integrated process for monitoring expenditures against budget and performance against targets.

Net enrolment rate (NER) – Estimated by calculating the ratio, dividing the number of pupils attending a certain school cycle (e.g., primary and secondary school, respectively), at normal ages (e.g., 6–11 for primary, 12–18 for secondary) by the number of children belonging to the age group.

Paris III International Conference for Support to Lebanon – The conference held in the French capital on 25 January 2007, in support of Lebanon in which 36 countries and seven regional and international institutions participated and whose pledges amounted to \$7.6 billion.

Performance-based budgeting (PBB) – The budgeting process that focuses on program outcomes and the measure of the program's ability to meet short-term and/or long-term objectives. The PBB systems incorporate three components: planning, budgeting, and performance measurements.

Student–teacher ratios (STRs) – Calculated by dividing the total number of students in a school or type of school by the total number of teachers teaching them.

Subsidized free schools – The private tuition-free schools, which are usually subsidized by the government on a per capita basis. Subsidized free schools cover only the elementary (cycles 1 and 2) level.

Trends in International Mathematics and Science Study (TIMSS) – The study carried out by the International Association for the Evaluation of

Education Achievement (IEA) and used to provide indicators of the degree of learned concepts in math and science for the fourth and eighth grade students.

General Background

The Republic of Lebanon, a parliamentary democratic republic based on respect for public liberties, especially the freedom of opinion and belief, and respect for social justice and equality of rights and duties among all citizens without discrimination, is located in the Middle East. Lebanon became an independent country in 1943 after being occupied by France since 1920. The country's constitution, fashioned after that of the Third Republic in France, was adopted on 23 May 1926. After a 25-year civil war extending from 1975 till 1990, the Ta'if Accord, also called the National Accord Convention, was put in place in October 1989 to restore peace and national reconciliation in Lebanon.

Lebanon is a small country with an area of 10 452 km² and an estimated population of 3 925 502 (the population estimates are as of July 2007, according to the Central Administration for Statistics (CAS) Factbook). The country's population is mainly composed of the 15–64 years age group which constitutes 66.7% of the total population (2 616 889), followed by the 0–14 years age group that accounts for 26.2% of the total population (1 029 439), and finally, the 65 years and over age group that represents the remaining 7.1% of the total population (279 174). Lebanon's population is almost equally divided between females and males as the former constitute 51.4% and the latter account for the remaining 48.6%. Lebanon's per capita income is estimated at \$5700, placing the country in the upper-middle-income economies category along with 40 other countries. (The classification is adopted from the World Bank's country groups classification. The 41 upper-middle-income economies – one among which is Lebanon – are: American Samoa, Argentina, Belize, Botswana, Brazil, Bulgaria, Chile, Costa Rica, Croatia, Dominica, Equatorial Guinea, Gabon, Grenada, Hungary, Kazakhstan, Latvia, Lebanon, Libya, Lithuania, Malaysia, Mauritius, Mayotte, Mexico, Montenegro, Northern Mariana Islands, Oman, Palau, Panama, Poland, Romania, Russian Federation, Serbia, Seychelles, Slovak Republic, South Africa, St. Kitts and Nevis, St. Lucia,

Table 1 Value added by sector 2004

Sector	2004 value (LBP bn)	2004 value (\$bn)	% of GDP
Agriculture and livestock	1697	1.13	5.24%
Energy and water supply	83	0.06	0.26%
Industry	3782	2.51	11.69%
Construction	2407	1.60	7.44%
Transportation and communications	2384	1.58	7.37%
Market services	10 629	7.05	32.85%
Trade	7763	5.15	23.99%
Government	3611	2.40	11.16%
Total GDP	32 357	21.46	100.00%

From *Economic Accounts of Lebanon, 2004*.

St. Vincent and the Grenadines, Turkey, Uruguay, Venezuela, and the Republic of Bolivia (RB)).

The country's economy is free and ensures private initiative and the right to private property. The economy is services oriented as this sector accounted for 32.85% (40.22% when the transportation and communications sector is added to the services sector) of the country's 2004 gross domestic product (GDP; see **Table 1**). The unemployment rate was estimated at around 20% in 2006.

Lebanon is characterized by its cultural openness and its highly qualified human resources. Although Arabic is the official language, the country's constitution derives from that of the Third French Republic. As such, French might be considered the second language in Lebanon. However, English has never been absent. In recent years, the English language popularity as the first foreign language and the second language after Arabic in schools has increased. This can be seen by the fact that the number of students enrolled in schools where English is the first foreign language and the second language after Arabic increased continuously from 33% in the academic year 2000–01 to 36.4% in the academic year 2005–06. In addition, the number of schools that teach English as the first foreign language and as the second language after Arabic has also increased from 19.6% in the academic year 2000–01 to 22.6% in the academic year 2005–06 (see **Tables 2 and 3**).

Goals of the Education Sector

The education sector in Lebanon is governed by

the fundamental principles cherished in the Lebanese Constitution and the National Accord Convention as well as the laws and regulations governing the educational matters which emphasize freedom of and right to education and ensure the accessibility and equality in

Table 2 Distribution of students (%) according to first foreign language

Academic year	French	English
2000–01	67.0%	33.0%
2001–02	66.4%	33.6%
2002–03	65.7%	34.3%
2003–04	64.8%	35.2%
2004–05	64.1%	35.9%
2005–06	63.6%	36.4%

From *ECRD Statistical Bulletins*.

Table 3 Distribution of schools (%) according to first foreign language

Academic year	French	English	Both English and French
2000–01	59.4%	19.6%	21.0%
2001–02	59.0%	20.2%	20.8%
2002–03	58.3%	20.7%	21.1%
2003–04	56.9%	21.6%	21.5%
2004–05	56.5%	21.7%	21.8%
2005–06	55.8%	21.6%	22.6%

From *ECRD Statistical Bulletins*.

opportunities and requirements of education to all. (The National Educational Strategy, 2007: p. 1.)

The education sector goals stress on the availability of education in Lebanon on the basis of equal opportunity as well as on the providing of good quality education that contributes to the building of an information society and that promotes social integration and economic development. The Ministry of Education and Higher Education (MEHE) clearly stated these objectives in the National Educational Strategy, which was sent to the Council of Ministers for endorsement in September 2007.

The National Educational Strategy (The Technical Committee assigned for setting the National Educational Strategy was established under Decree no. 82-3-2004 dated 11 February 2004.) – hereon, the Strategy – is derived from International and Arab educational principles that adopt a humanistic approach to education and that build on the

comprehensive development of the individual; the reinforcement of respect for human beings and their basic freedom; the development of the ability to actively participate in a free society; the development of a sense of responsibility in a spirit of understanding, peace, and friendship; and the commitment of wide social partnership among those concerned with education in order to ensure meeting the human need for education and for building a knowledge society.

The National Educational Strategy states that the mission of MEHE is to:

1. Take a leading role in producing a national educational curriculum which would serve as the reference framework for all its policies, organizational structures, and programs, projects, and activities.
2. Take a leading role in transforming the Lebanese society into an information society and in preparing scientists, experts, and qualified staff in all domains of life.
3. Take a pivotal role in achieving social progress and economic development.
4. Shoulder the responsibility of preparing a thinking, productive citizen and educate her/him in national belongingness and social integration in a society where justice, freedom, democracy, and peace prevail.

Structure and Operation of the Education Sector

Lebanon's educational system is composed of general education (GE), vocational and technical education (VTE), and higher education (HE). The GE system is public, private, or subsidized free (the subsidized free education is only for cycles 1 and 2) with 1399 public, 1025 private, and 364 subsidized free schools, for a total of 2788 schools for the academic year 2005–06. The VTE system, accessible only for students who have completed their cycle-2 elementary education, is either public (83 institutions) or private (362 institutions), while the HE comprises one single public university, the Lebanese University, and

39 private institutions for the academic year 2006–07 which include universities, university institutes, technological institutes, and theological institutes.

General Education

The GE system comprises the pre-elementary kindergarten, the elementary (cycles 1 and 2), the intermediate (cycle 3), and the secondary (cycle 4). With each cycle consisting of 3 years, the total school years (excluding the pre-elementary kindergarten) amount to 12 years (see **Tables 4** and **5**).

Pre-elementary kindergarten

Net enrolment rate (NER) in the pre-elementary kindergarten for children aged 3–5 reached 77.7% for the year 2004. (The NER is calculated for the year 2004, as the latest National Survey of Household Living Conditions, which includes the population distribution by age group, is for the year 2004; see **Table 6**.) When considered for every age, NER is at 26.6%, 72.2%, and 79.6% for children of ages 3, 4, and 5 years, respectively. The drastic percentage drop for the 3-year age group is attributed to the absence of clear policies to incorporate 3-year-old children in the educational ladder and the absence of a nursery class in most public preschools.

There is a huge discrepancy between the student enrolment in private and public pre-elementary kindergarten which is explicit in the fact that for the academic year 2005–06, 63.34% of total number of students were enrolled in the private pre-elementary kindergarten

Table 4 Distribution of students in GE according to cycle and class 2005–06

Cycle	Grade	Public	Subsidized free	Private	Total
Pre-elementary	Preschool	939	2515	17 355	20 809
	KG1	14 532	9456	37 174	61 162
	KGII	16 592	10 344	39 441	66 377
Elementary cycle 1	First	19 780	15 982	36 356	72 118
	Second	21 984	16 673	36 558	75 215
	Third	23 121	15 934	34 454	73 509
Elementary cycle 2	Fourth	32 097	16 006	34 841	82 944
	Fifth	25 747	15 143	32 809	73 699
	Sixth	24 550	13 201	32 357	70 108
Intermediate cycle 3	Seventh	32 795	0	40 639	73 434
	Eighth	26 393	0	38 160	64 553
	Ninth	22 015	0	34 233	56 248
Secondary	First	26 394	0	20 828	47 222
	Second scientific	10 855	0	15 457	26 312
	Second humanities	7862	0	3757	11619
	Third humanities	2214	0	1405	3619
	Third sociology and economics	9372	0	6307	15 679
	Third general sciences	1930	0	2799	4729
	Third life sciences	5479	0	6479	11 958
Total		324 651	115 254	471 409	911 314

From *ECRD Statistical Bulletins*.

Table 5 Distribution of students in GE according to situation upon registration 2005–06

<i>Cycle</i>	<i>Grade</i>	<i>Promoted</i>	<i>1-time repeater</i>	<i>2-times repeater</i>	<i>From home</i>	<i>Total</i>
Pre-elementary	Preschool	31	115	0	20 663	20 809
	KG1	18 871	733	4	41 554	61 162
Elementary cycle 1	KGII	58 080	1280	24	6993	66 377
	First	64 871	4113	17	3117	72 118
	Second	70 289	4909	17	0	75 215
Elementary cycle 2	Third	68 770	4728	11	0	73 509
	Fourth	69 235	13 639	70	0	82 944
	Fifth	65 527	8158	14	0	73 699
Intermediate cycle 3	Sixth	62 938	7157	13	0	70 108
	Seventh	60 688	12 714	32	0	73 434
	Eighth	56 860	7679	14	0	64 553
Secondary	Ninth	52 543	3697	8	0	56 248
	First	43 229	3980	13	0	47 222
	Second scientific	26 042	269	1	0	26 312
	Second humanities	11 506	113	0	0	11 619
	Third humanities	3484	133	2	0	3619
	Third sociology and economics	14 837	831	11	0	15 679
	Third general sciences	4625	101	3	0	4729
	Third life sciences	11 548	394	16	0	11 958
Total		763 974	74 743	270	72 327	911 314

From *ECRD Statistical Bulletins*.

Table 6 Distribution of residents according to age

<i>Age</i>	<i>Males % Females</i>	<i>% of total</i>
0–4	299 242	8.0
5–9	337 582	9.0
10–14	386 040	10.3
15–19	373 027	9.9
20–24	372 744	9.9
25–29	298 373	7.9
30–34	275 920	7.3
35–39	256 710	6.8
40–44	249 793	6.7
45–49	193 968	5.2
50–54	165 746	4.4
55–59	137 583	3.7
60–64	126 052	3.4
65–69	112 916	3.0
70–74	83 809	2.2
75–79	49 365	1.3
80–84	26 818	0.7
85 and above	9343	0.2
Total residents	3 755 031	100

From *The National Survey of Households Living Conditions 2004*, UNDP, MoSA.

cycle, while 21.61% and 15.04% were enrolled in the public and the free subsidized pre-elementary kindergarten cycle, respectively (see **Table 7**).

Elementary cycles 1 and 2 and intermediate cycle 3

The NER in the elementary level reached 97% in 2004 (National Educational Strategy, 2007: 6.), while the gross

enrolment rate (GER) for students aged 5–9 and 10–14 reached 98.4% and 95.2%, respectively (The National Survey of Household Living Conditions, p. 149). As for the intermediate level (cycle 3), the average NER drops to 68.5%, clearly indicating that the dropout or the repetition rate at the end of the elementary level is very high. The average NER at the intermediate level concerns the GE enrolment only and excludes a very meager number of students enrolled in the VTE at the intermediate level, a number that would not affect the average NER. At the end of the intermediate level (cycle 3), students sit for official exams; and those who succeed become holders of the Brevet certificate.

In Lebanon, basic education extends from age 6 (grade 1 of the elementary level) to age 15 (grade 9 of the intermediate level); however, Law no. 686 dated 16 March 1998, states that compulsory education stops at the age of 12 (grade 6 of the elementary level).

Secondary (cycle 4)

The secondary cycle-4 level is of 3 years, with the first as general; the second, sciences or humanities; and the third, sociology and economics, general sciences, or life sciences. The percentage of 15–19-year olds enrolled in school is 70.8% (The National Survey of Household Living Conditions, p. 149); however, not all of the students in this age group are enrolled in secondary education, as 31.3% of them are in the intermediate cycle and 16.5% of them are in higher education. As such, only 52.3% of them are enrolled in the secondary level. When one considers the 15–17 age group, the typical age of students in both general

Table 7 Distribution of currently studying residents according to current educational level and age

Age	Preschool	Elementary	Intermediate	Secondary	University	Special Edu.	No response	Total
0–4	47 530	0	0	0	0	0	37	47 567
5–9	48 384	284 127	0	0	0	387	112	333 010
10–14	0	173 892	184 450	8416	0	600	211	367 569
15–19	0	1952	60 758	150 666	50 971	719	0	265 066
20–24	0	0	970	19 347	106 939	346	0	127 602
25–29	0	0	0	803	19 413	0	0	20 216
30–34	0	0	83	0	4493	76	0	4652
35–39	0	0	0	0	964	0	0	964
40–44	0	0	0	0	556	0	0	556
45–49	0	0	0	0	261	0	0	261
50–54	0	0	0	0	316	0	0	316
55–59	0	0	0	0	0	0	0	0
60–64	0	0	0	0	58	0	0	58
Total	95 914	459 971	246 261	179 232	183 971	2128	360	1 167 837

From *The National Survey of Households Living Conditions 2004*, UNDP, MoSA.

and vocational and technical secondary schools, the NER drops to 50% (The National Educational Strategy, 2007: 8). At the end of the secondary level (cycle 4) students sit for official exams; and those who succeed become holders of a high-school (terminal) diploma, each according to his specialization in one of the three above-mentioned categories.

It is worth noting that the education system in general and the public schools in particular do not cater to integrating students with special needs, whether with disabilities or gifted students. As such, and despite the issuance of Decree no: 11853 dated 11 February 2004 – that states the need for establishing a combined committee between MEHE and local nongovernmental and governmental organizations aimed at setting methodologies that ensure the integration of students with disabilities and gifted students – there has not been issued, to date, any laws that take into consideration the needs of these students and ways to integrate them in schools.

Vocational and Technical Education

The VTE in Lebanon has been reorganized within new curricula under the Council of Ministers Decree dated 16 August 2000, as the importance of VTE has been highlighted to suit the country's construction and development needs. The VTE in Lebanon includes three levels. The first level of vocational preparation consists of two phases culminating in the issuance of a *Certificat d'Aptitude Professionnelle* (CAP) followed by a *Brevet Professionnel* (BP). The second level is a 3-year vocational and technical secondary education for those who have completed their BP or the intermediate cycle; this level branches into two tracks: the Technical Baccalaureate (BT) and Technical Secondary Diploma-Dual System (SD). The third and last level is the vocational and technical higher education divided into the 3-year medium framework which leads to the

Table 8 Number of VTE specializations at every level for the academic year 2005–06

Level	Specializations
CAP	8
BP	19
SD	7
BT	27
TS	27
LT	17
LET	3
3-month diplomas	21
1-year diplomas	33

BP – Brevet Professionnel; BT – Technical Baccalaureate; CAP – Certificat d'Aptitude Professionnelle; LET – Licence d'Enseignement Technique; LT – Licence Technique; SD – Technical Secondary Diploma-Dual System; TS – Techniciens Superior.

From *ECRD Statistical Bulletin*, 2005–06.

Techniciens Superior (TS) degree, and the 2-year higher framework following the TS which branches into the Licence Technique (LT) and the Licence d'Enseignement Technique (LET; see **Table 8**).

The total enrolment in the 83 public VTE schools and higher education institutes amounted to 35 090 students for the academic year 2005–06, while the enrolment in the 362 private VTE schools and higher education institutes amounted to 41 270 for the same academic year. Similarly, enrolment in the 3-months and 1-year VTE diplomas amounted to 22 435, for a combined total of 98 795 students for the academic year 2005–06, or some 10.84% of the total 911 314 students enrolled in the GE sector (see **Table 9**).

The VTE subsector suffers from high repetition, retardation, and dropout rates. In addition, the VTE as the GE lacks the proper integration systems and procedures for students with special needs. The VTE, in particular,

Table 9 Number of students in VTE at every level for the academic year 2005–06

<i>Level</i>	<i>Public VTE institutions</i>	<i>Private VTE institutions</i>	<i>Total</i>
CAP	200	925	1125
BP	3123	4477	7600
SD	709		709
BT	16 904	23 860	40 764
TS	11 138	11 716	22 854
LT	2290	292	2582
LET	726		726
3-month diplomas		11 220	11 220
1-year diplomas		11 215	11 215
Total	35 090	63 705	98 795

From *ECRD Statistical Bulletin, 2005–06*.

is characterized by the weak link between the subsector's specializations and the labor needs and the absence of career guidance.

Higher Education

The HE in Lebanon dates back to 1866 when the American University of Beirut (AUB) was established followed by Saint Joseph University (USJ) in 1875. They were the first Anglophone and first Francophone universities to open in Lebanon, respectively. It was not until 1951 that the national public university, Lebanese University (LU), was established. Both the private and the public higher education sectors fall under the jurisdiction of MEHE.

For 2007, the private higher education sector encompassed 39 higher education institutions, including 26 universities, seven university institutions, two technology institutions, and four theological university institutions. The LU, the only public university, is composed of 13 faculties and four institutions comprising 46 branches. Branches of the faculties of LU were established in the different Lebanese regions after the war broke out in 1977. Lately, LU has created three doctoral schools (Decree No. 74, March 2007): Doctoral School of Science and Technology; Doctoral School of Law and Political, Administrative and Economic Sciences; and Doctoral School of Literature and Human and Social Sciences.

The average number of students enrolled in higher education is 4143 per 100 000 habitants, one of the highest in the Arab region. In addition, the average enrolment rate in education for the 20–24 years age group is 29.7%, with about 1/4 of those enrolled in preuniversity education. The number of students enrolled in higher education in 2006–07 was 146 961 students of whom 70 627 (48.1%) were enrolled in the LU, 79 256 (53.9%) were females, and 131 070 (89.2%) were Lebanese.

Academic mobility within HE in Lebanon is highly limited. There is no comprehensive system for HE where definitions of curricula, degrees, and courses are agreed upon in such a way that would allow for academic mobility among institutions of HE. Universities in Lebanon are characterized by cultural diversity where some universities are Francophone while others are either Anglophone or universities whose main language of instruction is Arabic. The LU, with its faculties and branches, is the most diversified and heterogeneous institution with respect to these systems.

In the past few years, the universities that follow the French system started to be influenced by the changes resulting from the European Union Licence Maîtrise Doctorat (LMD) system that was applied in Europe in the beginning of 2004–05. Adoption of this new system might lead to increasing the opportunities for academic mobility within higher education institutions.

Regarding private higher education, there are few universities whose quality has been recognized as a result of meeting international, European, or American standards. Among those, some universities sought accreditation, an external evaluation scheme to assess their quality.

In 2004, the LU conducted a self-evaluation report which pinpointed several quality problems. These problems are attributed to the university's not adopting a reform project, not availing the needed resources, and losing its autonomy in the last decades. The report also presents clear recommendations regarding the needs of the university.

The Teaching Profession

The total teacher workforce in Lebanon in GE reached 87 459 teachers in the academic year 2005–06, of which 40 550 worked in the public sector, 6202 in the subsidized free schools, and 40 707 in private schools (see **Table 10**). The workforce is divided into three categories: within the cadre, contractual, and a very small percentage (2.3% of the total) that are in neither of the categories but rather volunteer teachers (mainly religious clerks). For the academic year 2005–06, the average student–teacher ratio (STR) in GE was 9, 19, and 12 in the public, subsidized free, and private schools, respectively.

The appointment of teachers in Lebanese public GE schools suffers serious problems at all levels, among which is the high percentage of teachers who are above 50 years of age and that account for 30.3% of all public school teachers, compared to 11.4% and 14.8% in the subsidized free and the private schools (see **Table 11**). At the pre-school kindergarten level, the percentage of teachers who have university degrees or special diplomas is less than 30.6% of all preschool teachers (The National Educational Strategy, 2007: 17). Laxity in academic conditions

to appoint teachers at the public elementary education level is evident by the high percentage of teachers whose education has stopped at the baccalaureate level or even before (see **Tables 12** and **13**) and the choosing of teachers with university degrees on a discretionary basis and without having any educational preparation. For the secondary public education, the practice of recruiting

Table 10 Distribution of teachers by category in GE for the academic year 2005–06

Category	Public	Subsidized free	Private	Total
Cadre	28 348	4668	23 017	56 033
Contractual	11 135	1326	16 984	29 445
Volunteers	1067	208	706	1981
Total	40 550	6202	40 707	87 459

From *ECRD Statistical Bulletin*, 2005–06.

Table 11 Distribution of teachers by age in GE for the academic year 2005–06

Age	Public	Subsidized free	Private	Total
< or = 20	25	83	414	522
21–25	2638	1084	5728	9450
26–30	6258	1115	6463	13 836
31–35	5037	935	6339	12 311
36–40	3831	1037	6678	11 546
41–45	3937	763	5494	10 194
46–50	6532	475	3547	10 554
51–55	6190	351	2685	9226
56–60	4492	196	1932	6620
61 and above	1610	163	1427	3200
Total	40 550	6202	40 707	87 459

From *ECRD Statistical Bulletin*, 2005–06.

teaching staff also suffers from the contracting of individuals with university degrees to teaching positions without them having educational preparation and the appointment of full-time teachers after offering them only short professional preparation session. As for training, there is a large-scale project which started in 2004–05 to offer continuous training for public elementary and secondary teachers.

The VTE subsector teacher-and-administrative workforce amounts to 16 183 teachers, of which 9141 are in the public VTE schools. Most of the VTE teachers are hired on a contractual basis without consistent hiring guidelines and standards. In addition, there are very limited training programs for VTE teachers in the use of modern technical and educational tools.

As for Lebanon's higher education, the total teacher and administrative workforce in this subsector amounts to 13 770 teachers, among which 4390 teach in the LU, the only public university in the country.

The teaching force at the LU and other private universities and university institutions suffers from the high percentage of hourly contractuels and the rise in the average age of full-time tenured faculty, mainly in the LU.

Administrative and Supervisory Structure

The current organization of MEHE dates back to 1959. Despite the attempts in the 1990s to split MEHE into three ministries (Ministry of Vocational and Technical Education, Ministry of Culture and Higher Education, and Ministry of National Education, Youth and Sports), currently there is only the Ministry of Culture and the

Table 12 Distribution of GE teachers by degree and number of schools they teach in for the academic year 2005–06

Degree	1 School	2 Schools	3 Schools or more	Total
PhD	393	93	14	500
Engineering	357	40	4	401
Higher Education	2193	478	76	2747
Kafaa	1365	492	78	1935
Bachelor's degree	26 348	3423	528	30 299
University diploma	760	123	18	901
Primary teacher training	8280	273	15	8568
Baccalaureate part II or equivalent	22 159	539	90	22 788
Diploma in intermediate edu.	190	18	0	208
Baccalaureate part I	1356	19	4	1379
Higher-primary edu. or equivalent	3149	32	10	3191
TS	1538	53	5	1596
BT	3803	36	7	3846
Other VTE degrees	744	108	13	865
Other education degrees	316	46	9	371
Not specified	200	1	0	201
Total	73 151	5774	871	79 796 ^a

^aThe difference between the total 87 459 and 79 796 teachers is 7663 teachers who teach in more than one school and that should be counted only once.

From *ECRD Statistical Bulletin*, 2005–06.

Table 13 Distribution of public GE teachers by degree and number of schools they teach in for the academic year 2005–06

<i>Degree</i>	<i>1 School</i>	<i>2 Schools</i>	<i>3 Schools or more</i>	<i>Total</i>
PhD	200	20	0	220
Engineering	67	9	1	77
Higher education	973	147	13	1133
Kafaa	1588	81	3	1672
Bachelor's degree	12 958	1389	128	14 475
University diploma	312	61	8	381
Primary teacher training	8235	75	0	8310
Baccalaureate part II or equivalent	8256	141	44	8441
Diploma in intermediate edu.	189	4	0	193
Baccalaureate part I	322	6	2	330
Higher-primary edu. or equivalent	1425	6	8	1439
TS	234	11	3	248
BT	404	5	2	411
Other VTE degrees	525	39	0	564
Other education degrees	105	32	7	144
Not specified	5	0	0	5
Total	35 798	2026	219	38 043 ^a

^aThe difference between the total 40 550, and 38 043 teachers is 2507 teachers who teach in more than one school and that should be counted only once.

From *ECRD Statistical Bulletin, 2005–06*.

Ministry of Youth and Sports that still operate along with MEHE.

The MEHE is composed of the Common Management Services Directorate, the General Directorate of Education, the General Directorate of Higher Education, and the General Directorate of VTE. The Educational Center for Research and Development (ECRD), an autonomous entity in charge of curriculum development, education reform, preparation of didactic material, data collection, and evaluation, and the Lebanese University are independent institutions attached to MEHE.

Despite the fact that MEHE is expected to carry out its mandate by looking after the public interest in both the public and the private educational sector, its role in the private sector has been limited to

conducting national examinations in general education, vocational and technical education, and higher technical education. It has also included keeping track of the names of students enrolled in private schools, offering financial support for public sector employees to help them send their children to private schools, and offering financial support to subsidized private schools. It is worth mentioning that the Ministry's monitoring of subsidized private schools required by law has been absent during the last few years (The National Educational Strategy, 2007: 52).

Educational Financing

The financing of the education sector in Lebanon is complex given the diversity of the financial resources and the intricacy of the procedures and programs used

to cover the costs of education in the country. Public funds flow to both public and private institutions. A distinction considered very important is that between public funds provided by public education authorities and those provided by agencies whose primary responsibilities lie outside education. The former includes the MEHE and other education departments and municipalities. The latter are of two kinds: (1) ministries or entities with noneducational responsibilities but who provide scholarship allowances to children of civil servants (such as the Civil Servants Cooperatives; ministries of defense, interior, justice; public institutions; mutual funds . . .), and (2) public institutions whose functions cut across educational responsibilities in terms of provision of schools and educational establishments such as the Council for Development and Reconstruction (CDR) and Council of the South (COS).

The MEHE has the highest budget allocation among social ministries (Ministry of Public Health (MPH), Ministry of Culture (MoC), and Ministry of Social Affairs (MoSA). Current expenditures on education within MEHE are mainly concentrated on wages and salaries (see **Tables 14–16**).

As for private spending on education, households spend on tuition fees as well as school supplies, textbooks, and transportation. Public funding from, for example, the Cooperative of Civil Servants, allocated to households for spending on education is simply a transfer from the government and cannot be considered private. However, households also receive funds for education from commercial, industrial, and other sectors (banking, insurance, commerce, and construction).

The mean estimated total expenditure per student in the public and private sector are very similar at the

Table 14 Budget of MEHE by directorate (current LBP billions)

<i>General Directorates</i>	<i>2004</i>		<i>2005</i>		<i>2006</i>		<i>2007</i>	
	<i>Total</i>	<i>% of MEHE's budget</i>	<i>Total</i>	<i>% of MEHE's budget</i>	<i>Total</i>	<i>% of MEHE's budget</i>	<i>Total</i>	<i>% of MEHE's budget</i>
Common Administrative Directorate	13.98	1.62	11.02	1.26	10.84	1.17	11.66	1.25
General Directorate of Education	612.62	70.89	591.84	67.43	592.90	64.24	600.41	64.47
General Directorate of Higher Education	156.55	18.11	174.26	19.85	174.74	18.93	172.60	18.53
General Directorate of Vocation and Technical Education	53.02	6.14	76.20	8.68	75.55	8.19	76.93	8.26

From Ministry of Finance.
LBP – Lebanese pound.

Table 15 Budget of MEHE by section (current LBP billions)

<i>Year</i>	<i>2002</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>
Section 1: Operations	727.8	704.7	834.5	853.3	854.0	861.6
Section 2A: Short-term investments	3.9	54.0	1.6	2.73	2.6	1.5
Section 2B: Long-term investments	61.1	52.0	28.0	21.69	66.3	68.3
Total MEHE budget	792.7	810.7	864.1	877.7	922.9	931.3

From Ministry of Finance.

Table 16 Budget of MEHE and other social ministries (current LBP billions)

<i>Year</i>	<i>2002</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>
Total budget of the MEHE	791.93	809.94	864.16	877.73	922.86	931.33
Share of MEHE in public budget, debt servicing included (%)	8.44	9.42	9.19	8.78	8.24	7.87
Total budget of MPH	289.53	285.13	344.61	360.27	346.94	361.63
Share of MPH in public budget, debt servicing included (%)	3.08	3.32	3.67	3.60	3.10	3.05
Total budget of the MoC	18.80	16.13	20.44	17.48	18.75	18.60
Share of MoC in public budget, debt servicing included (%)	0.20	0.19	0.22	0.17	0.17	0.16
Total budget of the MoSA	106.18	100.11	108.93	87.30	96.92	102.20
Share of MoSA in public budget, debt servicing included (%)	1.13	1.16	1.16	0.87	0.87	0.86

From Ministry of Finance.

primary and intermediate levels (1618 thousand LL per student in the public sector and 1680 thousand LL (the current LP/US\$ exchange rate is fixed at LP1507.5/\$) in the private sector). There are greater differences in estimated total expenditure per student between the sectors at the secondary level with an estimated total expenditure per student in the private sector of 2581 thousand LL and 1823 thousand LL per student in the public sector.

Performance Monitoring, Evaluation, and Research

Lebanon participated in the Trends in International Mathematics and Science Study (TIMSS) 2003.

(The TIMSS, carried out by the International Association for the Evaluation of Education Achievement (IEA) is used to provide indicators of the degree of learned concepts in math and science for the fourth and the eighth grade. Lebanon participated only for the eighth grade.) The results showed that Lebanon ranked 31st among 45 countries in mathematics with a score of 433 compared to an international average of 466 and 41st in science with a score of 393 compared to an international average of 473. Compared to the 11 Arab countries that participated in the study, Lebanon came first in mathematics and last in science. This low rank in science achievement may be attributed to the fact that science is taught in Arabic for most of the fourth- and eighth-grade students, while the TIMSS administered the questions in English (see **Table 17**).

Table 17 Mathematics and science achievement of eight grade students in Middle East North Africa – TIMSS 2003

<i>Country</i>	<i>Mathematics average score</i>	<i>Science average score</i>
Lebanon	433	393
Jordan	424	475
Iran	411	453
Tunisia	410	404
Egypt	406	421
Bahrain	401	438
Palestine	390	435
Morocco	387	396
Syria	358	411
Saudi Arabia	332	398
Yemen	278	250
International average	466	473

From Highlights from The TIMSS 2003, International Association for the Evaluation of Education Achievement (IEA), December 2004.

On another performance-monitoring front, in its 2008 budget circular, Ministry of Finance (MoF) decided to introduce the notion of a medium-term expenditure framework (MTEF) and request line ministries, one among which is MEHE, to prepare forward estimates (FEs) for their financial needs for 2009 through 2011. In addition, MEHE has been chosen as a pilot ministry to prepare performance-based budgeting (PBB) for 2008–09 and present it to MoF by March 2010 as a shadow budget along with the classical budget. As such, MEHE has already set key performance indicators (KPIs) that depict baseline and target levels (see **Table 18** and **Appendix 2**).

Major Changes and Issues Since the 1990s

Pre-Paris III Reforms

The reform efforts starting in the 1990s and culminating in the International Conference for Support to Lebanon, Paris III, held on 25 January 2007 (36 countries and seven regional and international institutions participated in this donor conference and pledges amounted to \$7.6 billion), recognized the importance of education reform, among other social reforms, as a cornerstone for development. Consequently, several steps were and are currently taken to restore education quality in public institutions, the information for which appears in The Needs Assessment Report prepared by the Helsinki Consulting Group for MEHE through the Education Development Project (EDP) in 2005.

In 1993, the government launched an interministerial commission to set general policy guidelines for the sector. In 1994, the ECRD issued a 10-year Plan for Educational

Reform in Lebanon, which aimed at restructuring the education system.

In May 1996, the government convened the First National Conference on Education, setting out its priorities and using this opportunity to familiarize both internal and external stakeholders and development partners. The same year, Parliament passed a law making elementary education free and compulsory until the age of 12. A second law extending compulsory education until the age of 15 is under preparation.

In 1997, a project was launched with assistance from the United Nations Development Project (UNDP) to support basic education development. The project, which is executed by the United Nations Educational, Scientific and Cultural Organization (UNESCO), emphasizes strengthening national capacities in education planning and management (education management information system (EMIS), school mapping, training, monitoring and evaluation, etc.) and aims at ensuring quality basic education accessible for all.

In 1998, the law on compulsory primary education was enacted, following advocacy efforts by the United Nations International Children's Emergency Fund (United Nations Children's Fund, UNICEF) and several organizations involved in child rights, in coordination with the Parliamentary Committee for the Rights of the Child, thus providing the basis for ensuring primary education for all.

Since the year 2000, MEHE has been implementing the EDP, a \$57 million loan project funded by the World Bank. The EDP's project development objective (PDO) is to enhance access, and build the knowledge base, analytical tools, and institutional capacity within the MEHE, necessary to initiate reform of the general and vocational education systems.

Paris III Reforms and Beyond

The MEHE presented a 3-year reform program at the Paris III conference held on 25 January 2007. The Education Sector Reform Action Plan for the years 2007–09 spans 7 reform programs: (1) consolidating policy, planning, and resource allocation; (2) achieving Universal Basic Education (UBE) for ages 6–15; (3) improving the efficiency, effectiveness, and competence level of the teaching workforce; (4) enhancing the quality of education; (5) strategic management of educational facilities; (6) rationalizing VTE; and (7) quality assurance in HE (see **Table 18** and **Appendices 1** and **2**).

The National Education Strategy document was sent to the Council of Ministers in September 2007 for endorsement. The strategy aims at providing: (1) education on the basis of equal opportunity, (2) quality education that contributes to building a knowledge society, (3) education that contributes to social integration,

Table 18 MEHE's reform programs and their key monitoring indicators

<i>Program</i>	<i>Monitoring indicators</i>
Consolidate policy, planning, and resource allocation	<ul style="list-style-type: none"> • Submission of a shadow budget under MTEF for MEHE by March 2008 • Enhancement in the organization capacity of MEHE through decreasing process duration time by 50%
Achieve Universal Basic Education (UBE) for ages 6–15	<ul style="list-style-type: none"> • Increase in NER for grades 1–6 by 1% from 97% to 98%, and for grades 7–9 by 4% from 68.5% to 72.5% • Increase classroom occupancy for grades 1–6 by 1 from 17 to 18, and for grades 7–9 by 2 from 19 to 21 • Increase the percentage of public basic education students with free textbooks in cycle 1 to 100%
Improve efficiency, effectiveness, and competence level of the teaching workforce	<ul style="list-style-type: none"> • Decrease the number of teachers without a baccalaureate degree and above age 58 and/or who are physically and/or psychologically disabled by 2% (from 9.6% to 7.6%) by 2009 • Increase the number of public school principals who have undergone professional development programs to improve competence to 50% of all public school principals (by 300, from 450 to 750) by 2009 • Increase by 10 500 the number of teachers who have undergone professional development
Enhance quality of education	<ul style="list-style-type: none"> • Decrease in repetition rate by 2% (from 21% to 19% for grades 1–6) and 2% (from 23% to 21% for grades 7–9) by 2010 • Increase of the NER for KGII by 3% from 79.6% to 82.6% in 2009 • Enhance efficiency of examination process through 100% automation of official exams
Enhance the utilization of educational facilities	<ul style="list-style-type: none"> • Decrease the number of unviable schools by 69% from 130 to 40 through consolidation and decrease resource centers by 48% from 33 to 17 • Decrease in the percentage of schools with less than 100 students by 10% from 25% to 15%
Rationalize vocational and technical education	<ul style="list-style-type: none"> • Decrease the number of nonviable VTCs through consolidation by 8% from 20% to 12%, close 10 out of the total 17 unviable VTE schools • Consolidation and reduction of specializations based on the labor-market needs and aligned with the investment priorities in VTE • Redefinition of the program framework in compliance with the labor-market needs
Develop quality assurance in higher education	<ul style="list-style-type: none"> • Increase in the accountability within the Lebanese University to assess the performance of teachers and administration • Enhancement in the autonomy of the Lebanese University • Implementation of national structures for quality assurance in higher education

Appendix 1 Action plan for the Education Sector Reform (2007–09)

<i>Sector reform objectives</i>	<i>Short-term actions 2007</i>	<i>Medium-term actions 2008</i>	<i>Longer-term actions 2009</i>	<i>Responsible party</i>	<i>Monitoring indicators</i>
1. Consolidate Policy, Planning And Resource Allocation MEHE organizational development	Complete the review of MEHE's structure, and propose new restructured organization in line with the requirements to implement the reform strategy	Issue amendments to MEHE laws and articles for fundamental restructuring of MEHE in accordance with the requirements of strategic development plan Develop process maps, procedures, communication, and HR management systems for new MEHE structure	Monitor implementation of new MEHE structure	EDP MEHE	MEHE has a new organizational structure Amendments to laws and articles have been issued and endorsed

Continued

Appendix 1 Continued

<i>Sector reform objectives</i>	<i>Short-term actions 2007</i>	<i>Medium-term actions 2008</i>	<i>Longer-term actions 2009</i>	<i>Responsible party</i>	<i>Monitoring indicators</i>
Policy, planning, and information management	<p>Integrate policy, planning, and information management functions at MEHE</p> <p>Nominate staff to fill key functions during the organizational restructuring transition period</p> <p>Develop and support with training, systems, and technology a core financial planning team within MEHE</p> <p>Develop financial scenario software tool and train a planning team on its use</p> <p>Perform information audit to identify relevant information and information needs</p> <p>Initiate the development of an EMIS</p>	<p>Develop and implement MEHE staff development program</p> <p>Undertake policy option analysis as required for strategy implementation</p> <p>Lead implementation of education sector strategy</p> <p>Initiate the development of PBB for the sector</p> <p>Develop procedures and methods for assessing effectiveness of strategies in meeting Government of Lebanon goals and objectives</p> <p>Implement EMIS at MEHE managed by IMU</p> <p>Implement a school information system that is operational at school level and connected through a national education network to central ministry database</p>	<p>Undertake policy option analysis as required for strategy implementation</p> <p>Lead implementation of education sector strategy</p> <p>Disseminate and discuss the model PBB within MEHE and other Government of Lebanon entities (MoF, CDR, Parliament, etc.</p> <p>Continue implementation of EMIS, and SIS</p>	<p>MEHE MEHE EDP IMU MEHE</p>	<p>MEHE personnel have undergone a staff development program</p> <p>Mandate for establishment of policy, planning, and information management unit</p> <p>Core team (national and regional) in place and tools and systems developed</p> <p>Model PBB for 2008–09 developed and disseminated</p> <p>EMIS is operational</p>
2. Achieve Universal Basic Education (UBE) For Ages 6–15					<p>2006: NER grades 1–6 = 97%</p> <p>2006: NER grades 7–9 = 68.5%</p> <p>2009: NER grades 1–6 = 99%</p> <p>2009: NER grades 7–9 = 77%</p>
Ensure equitable access for all students	Update the nation-wide survey of currently owned and rented physical infrastructure (buildings, equipment, and facilities) to provide data for upcoming geographic information system (GIS) for purposes of decision making	Utilize the school mapping capability to ensure placement for projected levels of enrolment	Initiate school construction/replacement program	MEHE	Laws ratified and decrees enacted

Continued

Appendix 1 Continued

<i>Sector reform objectives</i>	<i>Short-term actions 2007</i>	<i>Medium-term actions 2008</i>	<i>Longer-term actions 2009</i>	<i>Responsible party</i>	<i>Monitoring indicators</i>
		Formulate a medium-term plan for school construction, replacement, or rental to ensure universal and equitable access		Council of Ministers	
		Develop a teacher allocation plan to ensure efficient staffing of all schools		Parliament	
Enact compulsory education to age 15	Prepare draft-relevant laws and regulatory decrees to enact compulsory basic education to age 15	Undertake education infrastructure needs assessment arising from implementation of UBE Approve/ratify relevant laws	Implement UBE	MEHE	2006: average grades 1–6 classroom occupancy = 17 2006: average grades 7–9 classroom occupancy = 19
		Develop an implementation plan to match and meet the physical needs of UBE		CDR	2009: average grades 1–6 classroom occupancy = 18 2009: average grades 7–9 classroom occupancy = 22
Eliminate registration fees at third cycle (grades 7–9)	Prepare draft laws	Approve/ratify relevant laws Implement UBE	Implement UBE	MEHE Council of Ministers Parliament	Laws ratified
Engage participation of local society in the financing of school fund	Undertake consultations on feasibility and modalities of local society participation	Prepare relevant decrees, mechanisms, proposals Approve/ratify relevant laws	Implement	MEHE Min. of Interior	Local society is participating in the financing of parent's fund
Provide free textbooks to public-school students through revolving textbooks program	Establish mechanism and program for revolving textbook program Initiate phased implementation	Implement	Implement	MEHE ECDR	All public-school students receive free textbooks Procedures and mechanisms for revolving textbooks are established and enacted 2009: 100% of public-school students with textbooks
Provide free school transportation for cycle-3 public students arising from their reallocation	Set up possible scenarios for implementing the scheme of issuing transportation vouchers (organization, administration, responsibilities, financing)	Implement	Implement	MEHE	100% of cycle-3 public students with transportation vouchers in accordance with agreed criteria

Continued

Appendix 1 Continued

<i>Sector reform objectives</i>	<i>Short-term actions 2007</i>	<i>Medium-term actions 2008</i>	<i>Longer-term actions 2009</i>	<i>Responsible party</i>	<i>Monitoring indicators</i>
	Implement measures for consolidated schools (where need exists)			Ministry of Transport	
3. Improve Efficiency, Effectiveness And Competence Level of the Teaching Workforce					2006: student/teacher = 9 2009: student/teacher = 10
Leadership development	Continue implementation of leadership development program for school principals	Institutionalize leadership development program within MEHE and LU Continue implementation of leadership development program for school principals	Continue implementation of leadership development program for school principals	EDP LU	2006: Number of school principals who have undergone school leadership program – 124 2009: Number of school principals who have undergone school leadership program – 650 2006: Number of master trainers – 20 2009: Number of master trainers – 40
Enhance the professional competence of the teaching workforce	Develop evaluation tools for continuous instructors development program	Enhance the program for teacher training consistent with needs identified from the evaluation of continuous instructors development program	Continue implementation of teacher-training program	EDP	TOT program has been redeployed and trainers are undergoing training based on new strategy
	Expand and implement the master trainers program	Continue implementation of teacher-training program		ECRD	2006: Number of master trainers – 224
	Expand second-language learning opportunities	Continue expanding second-language learning opportunities			2009: Number of master trainers – 300
Enhance the efficiency of teacher workforce utilization	Assess the official workload of teachers and develop a plan for optimal utilization and deployment Ensure deployment of teachers with relevant language competence	Apply the relevant laws and regulatory decrees to nominate teachers in all cycles based on university degrees and possession of teaching diploma Implement the optimal utilization through reallocation	Continue implementing the optimal utilization through reallocation	MEHE	Number of teachers trained – 20 000 2006: Total # of permanent teachers – 28 786 (at schools) Number of full-time equivalent teachers – 25 356 2009: Student-teacher ratio: Total # of permanent teachers: to be determined Number of full-time equivalent teachers: to be determined

Continued

Appendix 1 Continued

<i>Sector reform objectives</i>	<i>Short-term actions 2007</i>	<i>Medium-term actions 2008</i>	<i>Longer-term actions 2009</i>	<i>Responsible party</i>	<i>Monitoring indicators</i>
Early retirement program	Prepare a framework and criteria for early retirement scheme Offer separation packages for 1000 teachers identified as having physical and/or psychological disabilities	Offer separation packages for 1884 primary teachers who are above age 58 and who do not hold a baccalaureate degree Implement an early retirement program	Continue implementing early retirement program	MEHE	2006: 12% of teachers are above 58 2006: 9.6% of teachers do not hold a baccalaureate degree 2009: 6% of teachers are above 58 Teachers who are not physically or psychologically fit to teach have been given early retirement packages Reduction of teachers who are sick 2% of teachers do not hold a baccalaureate degree 2006: Repetition rates (1–6) = 21% Repetition rates (7–9) = 23% 2009: Repetition rates reduced by 10% to 19% (grades 1–6) 21% (grades 7–9) New curriculum developed
Curriculum review and development	Review and update curriculum development process, procedures, and responsibilities	Initiate curriculum content enhancement based on analysis of learning achievement	Implement curriculum content enhancement	ECRD	
Enhance the quality and content of the student textbooks and teachers guide	Initiate a review of the process for textbook and teacher guide standards and content	Subject to review of outcome	Subject to review of outcome	MEHE Civil society	
Learning achievement	Continue implementation of assessment of academic achievement standards Develop examination question bank Develop examination generation system Complete examination management system	Continue implementation of assessment of academic achievement standard Implement usage of examination question bank system Implement usage of examination generation and management systems	Continue implementation of assessment of academic achievement standard Implement usage of examination question bank system Implement usage of examination generation and management systems	EDP ECRD Department of examinations	A question bank system has been developed for official exams The official exams are being generated and managed by an information system

Continued

Appendix 1 Continued

<i>Sector reform objectives</i>	<i>Short-term actions 2007</i>	<i>Medium-term actions 2008</i>	<i>Longer-term actions 2009</i>	<i>Responsible party</i>	<i>Monitoring indicators</i>
Expand access to public preschool education in poor, rural areas	Undertake needs assessment, costing, program design, and furniture requirements to provide 3 years of preschool education to all children	Initiate provision of school infrastructure Purchase furniture and equipment Initiate implementation of preschool expansion for KGII level (age 5)	Initiate implementation of preschool expansion for KGII and KGI levels (ages 3–4)	MEHE CDR ECDR	2006: NER KGII = 79.6% 2009: NER KGII = 84%
Introduce Dropout Prevention Program targeting pedagogical reasons behind dropout and repetition problem	Formulate a process for identification of students at risk Expand psychosocial support programs	Develop a program of remedial instructions (learning materials, instructional methods) Implement Dropout Prevention Program	Implement Dropout Prevention Program in schools Evaluate and expand	MEHE ECDR	Analysis of the determinants of student achievement completed and disseminated -28 Learning Material to meet the needs of students at risk from grade 1 to 9 have been developed TOT program has been provided to 96 trainers from 6 resources centers across provinces 20 000 teachers have been provided training on the developed learning materials and are implementing the program Reduction of investment years required to produce 9th-grade completer from 22.7 to 19.6 18% increase in the number of students who complete grade 9 within 11 years 2006: 25% of public schools have less than 100 students 50% of public schools have average student/classroom less than 15 2009: 15% of public schools have less than 100 students
5. Strategic Management of Educational Facilities (Enhance the Utilization of Physical Infrastructure)					

Continued

Appendix 1 Continued

<i>Sector reform objectives</i>	<i>Short-term actions 2007</i>	<i>Medium-term actions 2008</i>	<i>Longer-term actions 2009</i>	<i>Responsible party</i>	<i>Monitoring indicators</i>
Assess and upgrade the physical requirements of schools	Launch development of an educational GIS in order to optimize the utilization and investment in physical infrastructure	Activate Gis Undertake a gap analysis on infrastructure/ maintenance and investment requirements Establish procedures for continuous assessment and monitoring of school infrastructure together with responsibilities for reporting and updating of GIS	Identify and provide required infrastructure in compliance with MEHE norms for classroom and school materials and equipment Initiate upgrading in accordance with identified requirements and MEHE standards Establish procedures to ensure follow-up surveys to assess the annual needs of schools	MEHE CDR COS ECD EDP	40% of public schools have average student/ classroom less than 15 Physical needs at the school level have been identified Mechanisms have been established to ensure optimal utilization of physical infrastructure
Consolidate small GE and VTE schools	Establish criteria for school consolidation (# of students, student/ teacher ratios, constructed area, distance, etc.) Identify priority nonviable and receiving schools that could be consolidated Develop a consolidation plan for priority nonviable and receiving schools Issue mandate to consolidate priority non viable and receiving schools Implement consolidation for 30 priority nonviable receiving schools	Continue implementation of school consolidation plan	Continue implementation of school consolidation plan	MEHE	2007: Nonviable schools have been identified 2009: Nonviable schools have been consolidated Improvement in student-teacher ratios from 9 to 9.5
Reallocate students in cycle 3	Establish criteria for reallocation of cycle-3 students (# of students, student/teacher ratios, constructed area, distance, etc.)	Develop a student reallocation plan (identification of receiving schools and students) Implementation of student reallocation plan Issue decrees to reallocate cycle-3 students	Continue implementation of student reallocation plan	MEHE	2006: Cycle 3 Student/ classroom = 19 2009: Cycle 3 Student/ classroom = 23

Continued

Appendix 1 Continued

<i>Sector reform objectives</i>	<i>Short-term actions 2007</i>	<i>Medium-term actions 2008</i>	<i>Longer-term actions 2009</i>	<i>Responsible party</i>	<i>Monitoring indicators</i>
6. Rationalize Vocational and Technical Education					
Reexamine the role and mission of VTE in Lebanon	Examine the viability of CAP/BP programs Conduct a rate of return assessment to the various specializations	Implement consolidation of specialization	Conclude implementation of consolidation of specialization	MEHE DGVTE	2006: VTE policy vacuum 2008: ERR of VTE completed 2009: Specific policy objectives for VTE established
Establish employer community relationship	Develop strategies and assign responsibilities for the maintenance of relationships with priority economic and employment sectors	Establish VTE program advisory committees Engage employers in conducting a VTE program framework review	Implement VTE programs in collaboration with employers	DGVTE MEHE	2006: No mechanisms in place for employer input into VTE 2009: 6 program advisory committees established with employer inputs
Restructure VTE program framework, delivery modules, and assessment	Collaborate with employer community in redefining VTE program framework, credential levels, and occupational demands	Close VTE programs and/or VTE schools in accordance with current or projected occupational demands	Initiate curriculum upgrade in competency-based format Redevelop learning assessment tools to measure occupational competency and employability skills	DGVTE MEHE	2006: VTE system output inconsistent with labor-market demand 2009: VTE program framework realigned
Realign VTE workshop and equipment requirement	Update equipment inventory developed through VTE technical assistance	Assess equipment requirements against revised program frameworks	Initiate equipment procurement and/or maintenance as required	DGVTE MEHE	2006: Equipment inventory in place 2009: Equipment inventory update procurement plan in place
VTE instructor development	Update VTE instructor inventory and competency levels developed through VTE technical assistance	Establish VTE instructor competency requirements Formulate instructor occupational and pedagogical upgrading training program	Implement VTE instructor upgrading program	DGVTE MEHE	2006: Instructor cadre occupation competency profile misaligned with programme requirements 2009: VTE staff development programme defined
7. Quality Assurance in Higher Education					<i>A national council for quality assurance is established and started functioning on both public and private higher education institutions</i>
Review and update the LU law	Prepare and issue a new and contemporary law for LU	Approve/ratify the relevant law Take measures to ensure the administrative, academic, and financial independence of the LU and activating accountability mechanisms	implement		2006:

Continued

Appendix 1 Continued

<i>Sector reform objectives</i>	<i>Short-term actions 2007</i>	<i>Medium-term actions 2008</i>	<i>Longer-term actions 2009</i>	<i>Responsible party</i>	<i>Monitoring indicators</i>
Develop standards for institutional quality assurance	Draft standards compatible with international patterns in quality assurance in higher education and with local requisites	Undertake a national consultation on the drafted standards and approve them	Implement	MEHE Council of Ministers Parliament	2006: No standards for institutional quality assurance 2009: Standards for institutional quality assurance
Establish national structures for quality assurance	Prepare relevant laws and regulatory decrees to create a national council for quality assurance and define its structure and mechanisms	Approval/ratification of relevant law and decrees	Implement	MEHE	2006: No national structures for quality assurance 2009: Structures for quality assurance established

Appendix 2 Initiatives and their respective KPIs in terms of baseline, current value, and target value

<i>Programs</i>	<i>KPI</i>	<i>Baseline (July 2007)</i>	<i>Current value</i>	<i>Target (closest date)</i>
Consolidate policy, planning, and resource allocation	Percentage of educational policies and plans based on EMIS output	0%	0%	100% (Dec 2008)
	Submission of a shadow PBB for MEHE	PBB	PBB	
	Percent decrease in number of days and processes required to complete a sample of 10 tasks	0% decrease in total number of days	0% decrease in total number of days	50% decrease in total number of days (Dec 2008)
Achieve Universal Basic Education (UBE) for ages 6–15	Percent increase in NER for grades 1–6	97%	97%	98% (Dec 2009)
	Percent increase in NER for grades 7–9	68.5%	68.5%	72.5% (Dec 2009)
	Percent increase in class occupancy for grades 1–6 (cycles 1 and 2)	0%	0%	5.88% from 17 to 18 (Dec 2009)
	Percent increase in class occupancy for grades 7–9 (cycle 3)	0%	0%	10.53% from 19 to 21 (Dec 2009)
	Percent increase in public students with free textbooks in cycles 1 and 2	0%	0%	100% (Dec 2007)
Improve efficiency, effectiveness, and competence level of the teaching workforce	Percentage of teachers without a bacc. and above age 58: who are physically/psychologically disabled	9.6%	9.6%	6.9% (Dec 2008)
	Percentage of PS Principals who have undergone professional dev. prog. to improve competence	10%	10%	30% (Dec 2007)
	Number of teachers who have undergone professional development	16 500 teachers	16 500 teachers	20 000 teachers (Dec 2007)
Enhance quality of education	Percent decrease in repetition rate for grades 1–6	21%	21%	19% (Dec 2010)
	Average student repetition rate in grades 7–9	23%	23%	21% (Dec 2010)
	Percent increase in NER for children aged 3–5 enrolled in public KGs	21.61%	21.61%	29.69% (Dec 2010)
	Percentage automation of public examination process and exams	0%	0%	100% (Dec 2008)
Enhance the utilization of educational facilities	Percent decrease in number of nonviable schools or ecoles normales	0% (nonviable schools) and 0% (ecoles normales)	0% (nonviable schools) and 0% (ecoles normales)	23.1% (nonviable schools) (Dec 2007) and 48.5% (ecoles normales) (Dec 2009)

Rationalize VTE	Identification of number of VTE specializations that comply most with labor-market needs	Zero VTE specializations identified	Zero VTE specializations identified	15 VTE specializations identified (Dec 2009)
Develop quality assurance in higher education	Submission of standards compatible with international patterns in quality assurance in higher education and with local requisites	No unified set of standards submitted	No unified set of standards submitted	A unified set of standards submitted (Dec 2009)

(4) education that contributes to economic development, and (5) governance of education.

Throughout the year 2007, a number of initiatives have been launched to complete the review of MEHE's structure, and to propose new restructured organization in line with the requirements to implement the reform strategy; integrate policy, planning, and information management functions at MEHE; nominate staff to fill key functions during the organizational restructuring transition period; develop and support with training, systems, and technology a core financial planning team within MEHE; initiate the development of an EMIS; develop financial scenario software tool and train a planning team on its use; develop a PBB model for MEHE's budget; perform information audit to identify relevant information and information needs; and prepare and issue a new and contemporary law for LU.

Further Reading

Educational Center for Research and Development (ECRD). Statistical Bulletins 2000–2006.
 Education Sector Reform Strategy Action Plan 2007–2009.
 Lebanese Association for Educational Studies (2006). National Educational Strategy in Lebanon.
 Lebanon Constitution, International Constitutional Law (ICL) (1990). *International Constitutional Law (ICL)*. <http://www.servat.unibe.ch/law/icl/home.html> (accessed September 2009).
 Ministry of Finance, Lebanon. Budget Laws and Proposals 2002–2007. Ministry of Finance, Lebanon.
 Ministry of Education and Higher Education. The National Work Plan of education for all. 2004–2015.
 Ministry of Finance, Lebanon (2006). *International Conference for Support to Lebanon–Paris III. First Progress Report*. Ministry of Finance, Lebanon.

Ministry of Finance, Lebanon (2006). *Key Fiscal, Debt, Monetary, Trade and Market Reforms*. Ministry of Finance: Lebanon.
 Minister of Finance, Lebanon (2007). *Key Reforms @ Finance for 2007 and Developments in 2006*. Minister of Finance: Lebanon.
 Nakhle, N. (2007). *Performance Based Budgeting: Lebanon's MEHE Next Challenge*. Ministry of Education and Higher Education. Education Development Project (EDP) World Bank Loan # 7010-LE.
 Republic of Lebanon Ministry of Finance (2004). *Reform Program @ The Ministry of Finance*. Ministry of Finance: Lebanon.
 Survey Report and Analysis of Results conducted by the Helsinki Consulting Group for MEHE through the EDP in 2006.
 The International Association for the Evaluation of Education Achievement (IEA) (2004). *Highlights from the Trends in International Mathematics and Science Study (TIMSS) 2003*. The International Association for the Evaluation of Education Achievement. (IEA), Washington, DC.
 The Needs Assessment Report prepared by the Helsinki Consulting Group for MEHE through the EDP in 2005.

Relevant Websites

<http://www.crdp.org> – Center for Research and Development.
<https://www.cia.gov> – Central Intelligence Agency: The World Fact Book.
<http://www.higher-edu.gov.lb> – Directorate General of Higher Education (DGHE), Republic of Lebanon Ministry of Education and Higher Education.
<http://www.laes.org> – Lebanese Association for Educational Studies.
<http://www.opportunities.com.lb> – Opportunities in Lebanon.
<http://www.schoolnet.edu.lb> – Schoolnet Lebanon Project 2010.
<https://www.aub.edu.lb> – The American University of Beirut.
<http://portal.unesco.org> – UNESCO Office Beirut and Regional Bureau for Education, Seiko Sugita: UNESCO SHS.
<http://www.ul.edu.lb> – Université Libanaise.
<https://www.usj.edu.lb> – Université Saint-Joseph.

Lithuania

J Navickaitė, Ministry of Education and Science, Vilnius, Lithuania

L Paurienė, R Zablackė, A Tijūnelytė and V Kožemiakina, Education Development Centre, Vilnius, Lithuania

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Compulsory education – Compulsory and state-guaranteed provision of primary and basic education programs available to individuals under 16 years, citizens of the Lithuanian Republic, and foreigners holding the right of permanent or temporary residence in the Republic of Lithuania.

Education monitoring – Continuous analysis, assessment, and forecasting of the conditions and change of education.

Education provider – School, independent teacher, or any other education provider (institution, enterprise, or an organization involved in education as secondary activity) granted the right to provide instruction services in accordance with the order established by the law on education of the Republic of Lithuania.

Formal education – Instruction provided to teach programs approved and registered in the order regulated by legal acts and granting graduates primary, basic, secondary, postsecondary, or higher education and (or) qualification.

Learner of special needs – One whose ability to learn and participate in social life is limited by congenital or acquired disability.

Nonformal education – Diversity of educational courses providing for different educational needs of professional development and additional competences.

Nonstate-funded school – A school founded by a legal entity of the Republic of Lithuania (except Lithuanian Seimas (Parliament), the Government, the Ministry of Education and Science or other ministries, governmental agencies, agencies at the ministries, county administration office, and municipality council) or a natural person; a school founded by a legal entity or natural person jointly with legal entities or natural persons of other states, or a school founded by a legal entity or natural person of another state.

School-maintenance fund – Funds related to the education process indirectly (dedicated for the running of school as economic entity and exploitation

of its economic assets, providing a lift to pupils and other economic needs).

State-run school – School founded by the Lithuanian Seimas, the Government, the Ministry of Education and Science or other ministries, governmental institutions, institutions at the ministries, county administration office, or municipality council and a school jointly funded by a state institution with legal entities or natural persons of the Republic of Lithuania.

Universally available education – State-guaranteed education available to all citizens of the Republic of Lithuania and foreigners holding the right of permanent or temporary residence in the Republic of Lithuania.

Political, Social, and Economic Context

The Lithuanian State is a democratic republic. Seimas (Parliament), the president of the republic, government, and court are bodies of state power. The boundaries of power vested with each of them are established by the constitution. Questions critical in the life of the state and the nation are decided by referendum. The territory of the Lithuanian state is uniform and further indivisible into state units. Lithuanian is the state language.

The total area under the state of Lithuania is 65 300 sq. km with a population of 3 371 000 (estimate of the fourth quarter of 2007) and a population density of 51.6 people per sq. km. The Lithuanian capital is Vilnius, with a population of 553 600.

Short History

The name of Lithuania is first mentioned in the *Annales Quedlinburgenses*. The State of Lithuania emerged in the thirteenth century. On 6 July 1253, Mindaugas, the Grand Duke of Lithuania, was crowned king for the first time to become the only Lithuanian king. In 1323, Lithuanian Grand Duke Gediminas founded Vilnius, the capital of Lithuania. In 1387, Lithuania embraced Christianity.

The first school was opened in 1397 at Vilnius Cathedral. In 1547, Martynas Mažvydas (Martynus Mosvidius) published the first Lithuanian book, *Catechismus*. Vilnius University was founded in 1579. On 3 May 1791, the Commonwealth of Poland and Lithuania accepted the first written constitution, the first one in the history of Europe. In 1793, an educational commission was established in charge of the education system in the Grand Duchy of Lithuania. It was among the first secular education ministries in Europe. The first Lithuanian language newspaper, *Aušra* (*The Dawn*) was printed in 1883. The Independence Act of Lithuania was declared on 16 February 1918. In 1940, Lithuania was occupied by the Soviet Union. In 1988, a Lithuanian rebirth movement emerged leading to the declaration of the reestablishment of the independence of Lithuania. On 25 October 1992, the constitution of the Republic of Lithuania came into power.

From 29 March 2004, Lithuania is a part of the North Atlantic Treaty Organization (NATO). From 1 May 2004, Lithuania is a member of the European Union (EU).

State Governance

Lithuania has a parliamentary republic form of governance. The head of the state is the president of the republic elected by direct suffrage for a term of 5 years. The president oversees foreign and national security policy. Valdas Adamkus was elected president on 12 July 2004.

The Seimas is the highest body of state power. MPs are elected for 4-year terms. All citizens have the right to vote from the age of 18. From the age of 25, citizens are eligible candidates to the Seimas. The sessions of the Seimas are conducted by the chairperson or the deputy chairperson. The functions of the Seimas are established by the constitution, the statute of the Seimas sets down the structure and working procedures of the Lithuanian Parliament.

The Government of the Republic of Lithuania is composed of the prime minister and 13 ministers. The prime minister is appointed and dismissed by the president with the approval of the Seimas. Ministers are appointed and dismissed by the president on the advice of the prime minister. At the national government level, the responsibility for education is vested with the Ministry of Education and Science.

Religion

There is no state religion in Lithuania. Based on the 2007 census, 79% of the Lithuanian population were Roman Catholics.

National Minorities

Based on nationality, Lithuanian population represents the following composition: Lithuanians – 84.6%; Polish – 6.3%;

Russians – 5.1%; Belorussians – 1.1%; Ukrainians – 0.6%; Roma – 0.3%; Latvians, Tatars, German, and Jewish – 0.1%.

In Lithuania, national minorities enjoy the opportunity to be recognized as such, to teach children their native language and history of their nation, to foster their culture, and practice their religion. During the school year 2006–07, 8% of day-shift learners at general-education schools received instruction in a language other than the state language. Schools of Russian, Polish, and Byelorussian national minorities operate in Lithuania, as well as other mixed schools offering instruction in several languages. There are 46 Sunday schools across Lithuania.

Demographic Situation

The current trend in Lithuania is a decrease in the population. The number 3 403 300 in the early 2006 represents a decrease of 22 000 from early 2005.

Since 1990, the rate of birth in Lithuania has been decreasing and has dropped twice over 15 years. Summary birth rate indicator (average number of births given by a female) was 1.27 in 2005.

In 2006, *migration saldo* (negative net migration) was – 8782 or –2.6 per 1000 of the population.

Economic Situation

The economic development in Lithuania during 2006 was extremely rapid as compared with that in 2005. The gross domestic product (GDP) grew by 7.5% compared with 2005. In 2006, the GDP created per individual increased by 15.8% compared with 2005.

The number of unemployed decreased over 2006 as also did the proportion of unemployed among citizens of working age. Over the last 5 years, the level of unemployment dropped from 13.8% (in 2002) to 5.6% (2006).

The income of the Lithuanian population has grown in recent years. Average gross monthly salary paid to individuals employed with the national economy (except individual enterprises) was LT 1826.3 in the second quarter of 2007, an increase of 20.2% from the second quarter of 2006. It was LT 1847.3 in the state sector, increasing by 16.5% over the same time span, and LT 1814.0 in the private sector, showing an increase of 22.7% from the previous year.

The Development of Education Policy

The first document of significance for education strategy, *The Conception of National School* (1988), was created prior to the reestablishment of independence. Its key idea was to create opportunities for the ex-USSR republics to create their independent education systems.

The Law on Education of the Republic of Lithuania introduced in 1991 defined key goals for the education

system. In 1992, the Lithuanian government approved the *General Concept of Education in Lithuania*, a central conceptual document that informs the reform of education in the country. The concept sets out that in order to achieve its education goals a permanent system based on differentiated and integrated education has to be shaped.

In 2003, the Lithuanian Seimas passed a new edition of the Law on Education of the Republic of Lithuania, which created opportunities to further develop and improve the national education system. Rather than regulating institutional makeup of the system, it introduces a flexible and qualitative process of teaching. The law is learner focused.

On 4 July 2003, the Lithuanian Seimas approved a central political document for the development of education in the Republic of Lithuania, *Provisions of the National Education Strategy 2003–12*. This document sets forth key guidelines and indicators for national education.

Strategy and Goals of Education

The 2003 approved *Provisions of the National Education Strategy 2003–12* identifies three priorities for the development of education:

1. to create an efficient and cohesive education system based on responsible management, targeted financing, and rational use of resources;
2. to develop a continuous, universally available, and socially just system of education providing for lifelong learning; and
3. to guarantee quality education capable of meeting needs of individuals functioning in open society and market economy.

The *Provisions of the National Education Strategy 2003–12* formulates goals for the development of the Lithuanian education system and identifies means to achieve these goals. The provisions define the main qualitative and quantitative indicators to judge by and assess the development of Lithuanian education over the period 2003–12.

General Structure of Education and Levels of Education

Compulsory education in Lithuania starts at the age of 7 and lasts until the learner reaches the age of 16. Parents may opt for their child to start schooling at 6, if he or she shows the necessary maturity.

In Lithuania, boys and girls learn together. For scope of education, numbers of schools, learners, and teachers see **Figures 1 and 2** and also **Tables 1–4**.

Preschool Education

Preschool education is provided to children between 1 and 5 (or 6) years of age. Preschool and preprimary education are not compulsory. The child may be educated at home until the age of 7; the child or his/her family is entitled to receive educational support.

Preprimary Education

Preprimary education in Lithuania was first introduced in 2000. Currently, preprimary education is universal, yet not compulsory. Preprimary education is provided to children who turn 6 that particular calendar year. Preprimary education may be deemed compulsory for individual children should children's rights' protection institutions decide so.

Children of ages 5 and 6, not yet receiving education at educational institutions, are given education support in the family.

Primary Education

Children start primary education in form 1, the year when they reach 7 years. Learners spend 4 years in primary school, from the age of 6–7 to 10–11. After completion of the primary-education curriculum, learners acquire primary education. Learning achievements are not assessed in the form of marks during the stage of primary education. In forms 1–4, children are taught by one teacher providing

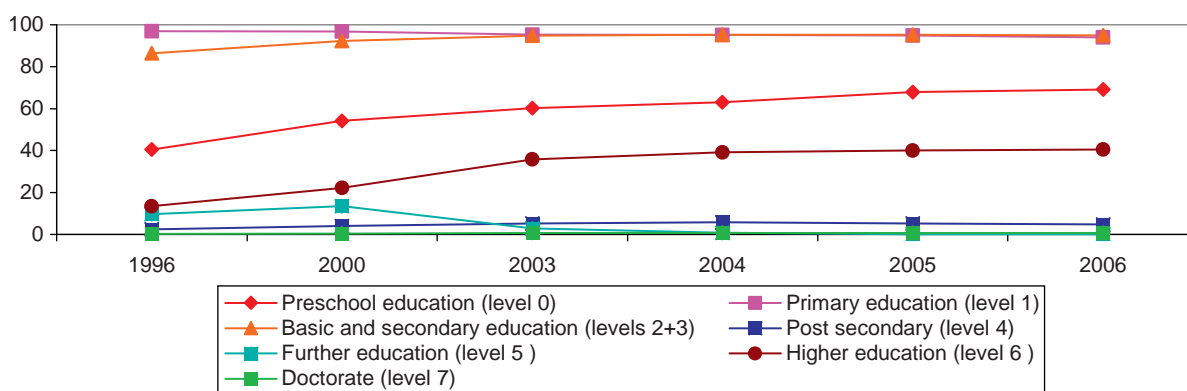


Figure 1 Scope of education, percent (net). From Statistics Department at the Government of the Republic of Lithuania.

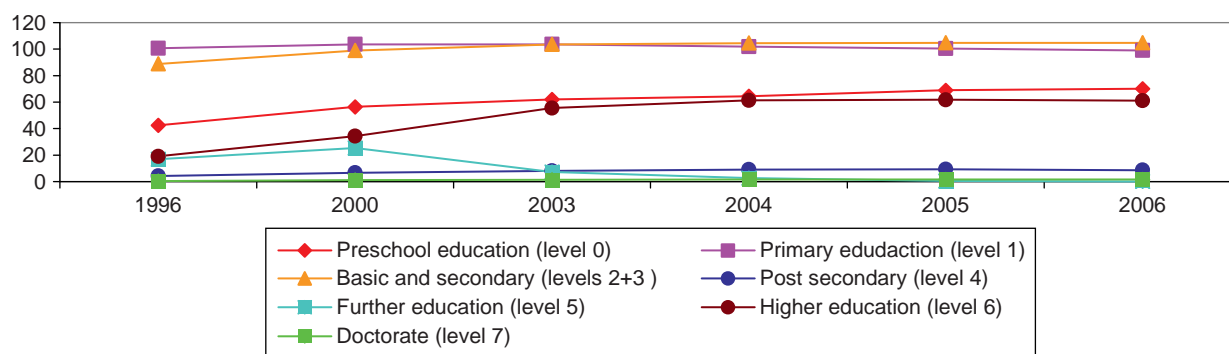


Figure 2 Scope of education, percent (gross). From Statistics Department at the Government of the Republic of Lithuania.

Table 1 Number of schools 2000–06

	2000	2001	2002	2003	2004	2005	2006
Total number of schools	3178	3085	2983	2735	2410	2315	2284
Preschool-education schools	714	699	686	672	655	656	652
General-education schools	2354	2270	2172	1932	1634	1534	1502
Vocational schools	84	81	82	83	73	76	80
Colleges	7	16	24	27	27	28	28
Universities	19	19	19	21	21	21	22

From Statistics Department at the Government of the Republic of Lithuania.

Table 2 Number of learners 2000–06 (thousands)

	2000	2001	2002	2003	2004	2005	2006
Total number of children, pupils, and students	840.0	854.5	875.4	887.7	888.5	872.5	850.1
Preschool-education schools	90.1	89.8	90.9	89.5	88.4	90.0	90.6
General-education schools	603.8	602.4	594.3	583.1	563.1	538.5	514.6
Vocational schools	47.0	45.0	44.4	44.4	46.3	46.3	45.4
Colleges	3.5	10.4	26.2	40.5	52.2	55.9	56.3
Universities	95.6	106.9	119.6	130.2	138.5	141.8	143.2

From Statistics Department at the Government of the Republic of Lithuania.

Table 3 Number of teachers 2000–06 (thousands)

	2000	2001	2002	2003	2004	2005	2006
Total number of teachers	78.7	77.6	77.9	79.0	77.6	74.7	74.0
Preschool-education schools	12.5	12.3	12.1	12.0	11.9	12.1	12.2
General-education schools	51.9	50.9	50.2	49.5	48.0	44.6	43.9
Vocational schools	4.9	4.7	4.7	4.7	4.9	4.9	4.7
Colleges	0.4	1.1	2.0	3.5	3.6	3.7	3.6
Universities	9.0	8.6	8.9	9.3	9.2	9.4	9.6

From Statistics Department at the Government of the Republic of Lithuania.

instruction in all subjects. In forms 1–4, the maximum allowed number of learners in a class is 24.

Primary school year (forms 1–4) lasts 170 school days.

Schools providing primary education are required to comply with the requirements set forth by the state

general-education curricula and standards for education, yet they can also apply alternative ways of organizing provision of primary education. Alternative education schools of Waldorf, M. Montessori, Suzuki, and others operate in Lithuania.

Table 4 Adult participation in education in 2006 (all forms: formal, nonformal, and informal)

	<i>Total</i>	<i>Men</i>	<i>Women</i>	<i>Towns</i>	<i>Countryside</i>
Total number of residents between 25 and 64 years	1795.1	851.8	943.3	1248.5	546.6
Participated, number	987.1	439.5	547.6	757.7	229.4
Did not participate, number	808.0	412.3	395.7	490.8	317.2
Participated, percent	55.0	51.6	58.1	60.7	42.0
Did not participate, percent	45.0	48.4	41.9	39.3	58.0

From Statistics Department at the Government of the Republic of Lithuania.

Basic Education

Basic education is provided to learners who have acquired primary education. After completion of the basic-education curriculum, learners acquire basic education.

School year for learners in form 5 is 170 school days; in forms 6–10, it lasts 195 school days.

Learning achievements are assessed according to a ten-point scale. Schools are free to select other scales, but need to establish a procedure to convert their results into the ten-point scale.

The learner graduating from basic education receives basic education certificate.

Secondary Education

Secondary education curricula are taught by day-general-education schools, vocational schools (their day-time sections), adult-education schools (evening sections), and adult education centers.

Secondary education is acquired after completion of the general-education curriculum and passing of maturity examinations. Currently, graduates may select to take school and/or state maturity examinations. School maturity examination features tasks prepared on the basis of the program (requirements) for school maturity examinations. Such examinations are taken at the school or a local center. The assessment is criteria based and recorded on a ten-point scale. State maturity examinations involve tasks based on the state maturity examinations program (requirements) and are taken at the centers of state maturity examinations. The results are assessed in a centralized manner. Normative assessment is used and results are ranked on a 1–100-point scale. Graduates from secondary school pass the following examinations: credit-tests in the Lithuanian language (state language), native language, and one foreign language; they take compulsory examination in the Lithuanian language and no fewer than three selected maturity examinations.

Vocational Education and Training

Vocational education and training may be initial and continued. Initial vocational education is formal and is

designed to acquire the first qualification. It is available to learners who have basic or secondary education. Learners who have only basic education can receive vocational training combined with secondary education. Initial vocational education can be provided even to learners who do not have basic education but are younger than 14 years.

Continued vocational education is provided to individuals of different ages who already have a qualification. It is designed to improve the existing qualification or provide a new one. The continued vocational education comprises both, formal and nonformal vocational training.

Vocational education and training encompasses formal and nonformal vocational education programs. The formal vocational education program has to meet the requirements in the standards describing the given qualification. Such formal vocational education programs can be provided together with general education programs. Providers of formal vocational education are vocational-education institutions and other institutions licensed by the Ministry of Education and Science. Vocational training can also be organized on the job in the form of apprenticeship, in companies or enterprises.

Individuals receive a qualification upon acquiring the competencies required for it as established by the standards describing the given qualification. The assessment of acquired competencies is performed by institutions (social partners or other legal entities) accredited for that purpose by the qualifications service.

Higher Education

Studies of higher education are universal and are provided to individuals who have acquired secondary education and are capable of studying independently.

The curricula of higher education studies are taught by higher education institutions: universities and colleges. Students may study at different higher education institutions with intermissions and complete these programs in separate modules.

Upon completion of higher education study programs, alumni receive higher education and (or) qualification. Scientific research degrees are granted in accordance with the requirements formulated by the Higher Education Law.

Special Education

Special education is provided as part of all-compulsory and universal-education programs. These programs are modified and adjusted, and also separate specialized programs are created to be provided alongside, with additional support in order to meet learners' special needs.

Special education is provided by all schools teaching compulsory and universal education programs, other education providers, and, in separate cases, by specialized schools.

Formal education programs, when modified for the purpose of special education, can be covered in longer than originally established periods. Learners may take such a program with intermissions and in separate modules.

When individuals with special needs take education programs meeting state education standards, they can acquire education and (or) qualification. In separate cases, qualification may be acquired without standard general education.

Nonformal children-education programs are offered by music, art, sports, and other schools, individual teachers, and other education providers. Music, art, sports, and other schools teaching programs of nonformal education or their modules fall under the same category as the schools teaching formal education programs. Individuals seeking higher education in the fields of music, visual and performing arts have to complete, alongside with secondary education, a specified program approved by the Education and Science Minister, if such a requirement is included in the admission requirements of the given higher education institution.

Adult Education

Adult education is provided to all individuals seeking it, but who are not younger than 18.

There are over 60 adult-education centers across the country operating in each municipality, and over 400 private nonformal adult-education providers.

Competencies acquired through informal education may be recognized as part of a formal education program or qualification in accordance with the order established by the Government of the Republic of Lithuania, its authorized institution, or higher education institution.

Nonformal adult education is regulated by the Law on Nonformal Adult Education.

Education Availability and Quality

Each citizen of the Republic of Lithuania, or a foreigner holding the right of permanent or temporary residence, has the right to learn and acquire education and qualification. Instruction of basic, secondary, and vocation

programs provided by state-run and municipality-funded schools is free.

The Right to Select Educational Programs

Learners are given the opportunity to select educational programs, modifications of these or modules, study programs in different subjects, in line with individual abilities.

Employers must create conditions to learn for working individuals. Vocational and higher education schools provide working learners with the opportunities of learning in different ways, including distance education.

Children in social exclusion, those from poor families, deprived of family and care, and refugees' children who do not attend school are given the opportunity of alternative education in youth schools. Individuals under temporary deprivation or restriction of liberty are provided with the opportunity to learn at the place of rehabilitation or penalization and acquire primary, basic, or secondary education qualification; opportunities of informal education are also available. Instruction according to the modules of general-education programs and non-formal-education programs is available to conscripts during their service period; they can also undertake informal education.

Learners with special needs, based on the decision of parents (guardians and caretakers) have options available to take a full- or partial-integration program at the general-education school, to learn at the school teaching special-education programs, or opt for school sanatorium. Education at medical facilities or at home, also arrangements of independent learning and taking of examinations are available to learners unable to attend general-education schools because of illness or disability. Individuals with special needs have the right of special privileges introduced into admission procedure by vocational or higher education schools when entering these institutions.

Assessment of Learners

Learner achievements are assessed by teachers, education providers, school founders, the Ministry of Education and Science, and authorized institutions. Institutions authorized by the Minister of Education and Science organize examinations and order the creation of other tools to assess learner achievements.

The procedure of examinations, tasks, and the forms of assessment can be modified to assess the achievements of learners with special needs.

Individual learning achievements of general education are assessed by the national examination center. The center organizes and conducts examinations and credit tests; performs assessment of education achievements and literacy, tests of knowledge, abilities, interim and final assessments of educational periods; and performs generalization and analysis of examinations, tests, and education.

Education Quality

The education provider is responsible for the quality of education. The state guarantees for the quality of formal education, and partially so, of nonformal education.

The Minister of Education and Science approves preschool and general/nontraditional/vocational/nonuniversity study programs of general education, establishes the criteria for preschool and nonformal children-education programs, approves general-education plans for general-education schools and standards for primary/basic/secondary education and vocational training; the ministry sets regulations for study fields, teacher training, and maturity examinations programs.

Teachers and Pedagogical Staff

Teacher Education and Training

Teacher education and training represent the first-level university or nonuniversity studies (including teaching practice at school), pedagogical placement, and qualification examination.

Individuals with secondary education are eligible candidates to take university and nonuniversity teacher-training programs. There is no age limit for the applicants. Applicants have to take additional tests to study some of the specialties.

Both parallel and consecutive models are applied to train teachers. Parallel model represents a method when all components of teacher-training curriculum content are studied in parallel. According to consecutive model, academic studies are followed by professional studies, for example, disciplines in education science, methodology of teaching different subjects or integrated courses, and school practice. Teacher training model in Lithuania includes common curriculum content components such as: studies of educational sciences, studies of academic subjects and some disciplines related to school educational programs, methodology of individual subjects or integrated study courses, and pedagogical/school practice.

Teachers working at school, yet having no professional qualification or adequate qualification in the subject they teach, can acquire it by full-time studies, evening studies, or studies by correspondence (distance studies). They acquire knowledge provided by specialized professional studies and write a graduation thesis.

Teacher Professional Development and Retraining

Teacher retraining for a different qualification is an integral part of teacher training. Such retraining studies are organized by university and nonuniversity schools and retraining institutions hold a license for that purpose from

the Ministry of Education and Science. The duration and format of retraining depend on the program, the nature of specialty, and the potential of institution offering retraining services; it also depends on the level of the individual learner and its duration is no longer than 3 years.

Teacher Professional Development

There are four teacher qualification categories in Lithuania: teacher, senior teacher, teacher-methodologist, and expert teacher.

These are the following types of institutions undertaking teacher and school leadership professional-development functions: teacher methodological centers, municipality teacher-education centers, county education centers with regional computer-training centers, regional computer-training centers, and professional-development institutions at higher education schools (with regional computer-training centers).

The following institutions are indirectly involved in teacher professional development: school methodological centers, childhood pedagogy centers, centers of supplementary education under the Ministry of Education and Science, and nongovernmental organizations.

The services provided by professional-development institutions are free. Part of the professional-development events' expenses are covered from the state or municipal budget. Financing for professional development is also earmarked from the school budget; its absolute amount depends on the number of learners.

Work Hours and Workload of Teachers and Pedagogical Staff

The government has established that teachers and psychologists of general education, professional, post-secondary schools, higher education schools, and supplementary-education institutions, and teachers of healthcare and social-care institutions are entitled to 56 calendar-day holidays per year. Teachers of higher education institutions are also entitled to holidays of the same duration.

The teachers' contact hours are defined and limited. The maximum allowed workload (contact hours) of a teacher is 34 h per week. The time teacher spends at school is not limited.

Each academic year, higher education institutions establish the required number of contact hours per position. State-run colleges establish teacher monthly payroll salaries based on 700–740 contact hours per academic year.

School Administration and Management

The founder of the school assigns and dismisses the school head (except higher education institutions). Heads of state-run and municipality-funded schools are hired through

public competition. Heads of nonstate schools are assigned, elected, and dismissed as per the order established by the labor laws.

The rector of the state-run university is elected by school's senate for a term no longer than 5 years, not exceeding two tenures in a run. The procedure is public competition and secret ballot held in the order established by the statute of each higher education school. Eligible candidates are established scholars or artists holding the title of professor. Rectors of nonstate universities are elected or appointed in the order established by the statute of the school.

Education Management

The purpose of education management is to guarantee the quality of the implementation of state education policy with managerial tools: monitoring, planning, distribution, and supervision of authority and responsibility.

Education Monitoring

Education monitoring is performed in accordance with the provisions established by the description of the order of state education monitoring and the list of indicators of state education monitoring. Data and information for the purpose of education monitoring are presented by the education management information system (EMIS) research into the state of education, the national learner achievement, and other types of international research.

The priorities of Lithuanian education, long-term goals, directions in the changes of curricula content, and priorities for financing are established by the State Education Strategy 2003–12. Drafted by the Ministry of Education and Science, the state strategic-education plans are approved by the Lithuanian Government. In order to implement these strategic plans, annual education-activity programs are run on the national, country, municipal, and school level.

School Self-Governance

School self-governance depends on school's specific goals of education, concrete programs run by school, and the school's ethos. The diversity of self-governance bodies, their competence, and principles of formation are established by school regulations (statute). School council is the highest self-governance body representing learners, teachers, parents (foster-parents), guardians, or caretakers and local community. Vocation school council also represents social partners. Teacher council (except higher education institutions and nonformal education schools) is a permanently operating school self-governance body dedicated to solving teachers' professional questions and general

issues of education process. It includes school leadership, all teachers on staff, medical personnel, psychologists, social, social pedagogs, librarians, and other individuals directly involved in education process. Schools may also have other types of self-governance bodies (students, parents (foster-parents), guardians, and caretakers).

Financing of Education

Investments into Development of Education

On the state level, investments into the development of education are made based on approved programs. One-year or long-term investment programs in the field of education are drafted by the Ministry of Education and Science in accordance with the State Education Strategy 2003–12 and the Lithuanian Republic Government Program. At the municipality level, investments into development of education are made based on programs and investment projects included into the budget of municipality.

Financing of Education Programs and Schools

Education is financed from the state budget, municipality budget assignments, and other resources. Preschool and formal education programs run by state, municipal, and nonstate-funded schools (except higher education schools) are financed from state and municipal budgets on the principle of expenditure per pupil. Municipal and nonstate-funded schools providing preschool and general education receive financing for education from the special, targeted donations to municipal budgets from the state budget for concrete year. State-run schools receive financing as state budget assignments. Funds for school maintenance are assigned by the founder in the order it has established. Healthcare at school is covered from the budget of compulsory health insurance fund, state and municipal budgets, and other resources.

Nonstate-funded, traditional religious community, or society schools, when they run programs of formal education, can be financed as per the order established by the Lithuanian Government and receive the same education and school-maintenance funds as do state or municipality schools of the same type.

Programs of nonformal children education are financed by the founder of the school, learners (their representatives), and supporters. The founders also finance vocational, non-formal, adult-education schools and institutions providing support to learners, teachers, and schools.

Each state higher education institution receives annual financing from the state budget, assigned by the Seimas. Financing for state higher education schools includes: state budget funds (earmarked for studies, development of scholarly research and artistic creation, administration and maintenance, maintenance of cultural properties,

coverage of students' tuition fees, students' stipends, implementation of state investment and other programs, and the development of international exchange), other state funds, revenues generated from scholarly work, economic activity, and provided services, recourses received from international and other funds and other organizations, tuition fees paid by students receiving just partial state financing, tuition fees by students who study at their own cost, as well as other legally generated recourses. The founder establishes the structure and allocation of assets of nonstate-funded higher education schools.

Payment for Work

Salary paid to state (except higher education schools) and municipal school head, deputy head for education, and section supervisor responsible for the organization of education depends on individual's education, years of pedagogical work, category of qualification, and complexity of work. The procedure of payment for work is established by the Government.

The head of state higher education school is paid based on the effectiveness of performance and workload; teachers and scholarly staff are paid based on responsibility, scholarly qualifications, involvement in the projects of international scientific and technological-development programs or commissions from economic subjects, and performance of urgent, key, or complex tasks.

Salaries of nonstate schools are established in the order set by the law.

Material Support

Learners taking an initial vocational training program can receive a stipend and be provided with material support.

Students at state higher education schools taking basic (bachelor) university-level programs, integrated and second-stage (master) programs, and doctoral or licentiate students can receive academic (based on academic results) or social stipends. Stipends are paid from the state budget assignments. Students of higher education schools can receive state loans to pay tuition fee, accommodation and subsistence loans, and toward partial studies based on international treaties and agreements.

Learners taking nonformal education programs can also receive material support.

Payment for Education

State and municipality schools teach preschool, primary, basic, secondary, and vocational education (initial qualification) programs for free. Payment for education, teaching, or studies is established by the founder of nonstate school.

Payment for nonformal education (except preprimary) provided by school is established by the founder; it can be reduced with regard to learner's ability and social conditions of parents (foster-parents), guardians, or caretakers.

High-achieving students of state higher education schools are guaranteed free higher education: no lesser than 30% of students following each study program (except those who study at their own cost), selected each semester based on the best academic results (entrance results for the first semester) get their tuition fee paid from the state budget. The number of students whose tuition fee is paid from the state budget is established by the government. If full tuition fee for studies at the state higher education school is paid by the students who take the same or lower-level study program that they have already taken at the state higher education school, they need to pay this fee for only half of completed program's credits. It is also paid by individuals, who simultaneously take two or more programs of the same level, and at least one of these study programs is fully or partially paid by the state; this applies to foreigners also, unless international agreements of the Lithuanian Republic establish otherwise. It also applies to individuals who agree to pay the full fee and get enrolled into evening studies or extra-mural studies.

Students taking nonconsecutive studies or repeating separate subjects from the program of nonconsecutive studies provided by the state higher education school also pay tuition fee proportionate to the scope of subject taken.

Evaluation and Research

Currently, evaluation of educational institutions performs the following functions: formative (when evaluation is done for the sake of improvement), summative (for the sake of accounting for work performed), and administrative (for the sake of improving administrative function).

The founder constantly observes the activity of general-education schools and performs its overall evaluation every 7 years. The evaluation is based on the analysis of self-evaluation, observation of lessons (instruction periods), analysis of learners work, and interviews with learners, teachers, leaders, parents (guardians), and other individuals related with the activity of the organization, or a survey of these analyses of documentation, facts, and concrete indicators. Such an evaluation is performed by the National School Assessment Agency.

The coordinating function for the self-evaluation and the assessment of study quality at scientific and study institutions is performed by the Lithuanian Centre for Quality Assessment in Higher Education. Currently, institutional assessment is performed only in colleges. It is going to be performed in universities only in the future; therefore universities currently prepare self-evaluation only for external evaluation of study programs, performed

on a periodical basis. Lithuanian Centre for Quality Assessment in Higher Education has prepared a methodology for the summary of self-evaluation of study programs and methodological recommendations for colleges on summarizing self-evaluation.

Research in Education

Since 2001, the Ministry of Education and Science orders annual researches on the condition of education, including different levels of education, educational processes, problems arising in teacher training, questions of educational support, and other activities of current significance in the field of education. The results of the research are used to perform analysis of the education system in planning the activity of the education system and decision making with the goal of specifying priorities in the development of education and science.

Annually, national research of learners' achievements is conducted and includes achievements of pupils in forms 4, 6, 8, and 10. The research covers the following fields of education: Lithuanian as native language, mathematics, natural sciences, and social sciences.

Lithuania takes part in international research: International Association for the Evaluation of Educational Achievement Trends in International Mathematics and Science Study (IEA TIMSS), Progress in International Reading Literacy Study (IEA PIRLS), International Civic and Citizenship Education Study (IEA ICCS), IEA Second Information Technology in Education Studies (SITES), Organization for Economic Cooperation and Development Programme for International Student Assessment OECD PISA, OECD Teaching and Learning International Survey (TALIS), and European School Survey Project on Alcohol and Other Drugs (ESPAD).

Scholarly research in Lithuania is carried out by higher education institutions, agencies under the auspices of ministries, scholarly institutions, and institutes.

Further Reading

Classification of Lithuanian Education (1999). Vilnius: Ministry of Education and Science of the Republic of Lithuania.
 Concept of Lithuanian Education (1992). Vilnius: Ministry of Education and Science of the Republic of Lithuania.

Decree of the Seimas of the Republic of Lithuania No. IX-1700 of 4 July, 2003. On the Approval of Provisions of the State Education Strategy 2003–2012 (State News, 2003, No 71–3216).

Education 2006 (2007). Vilnius: Statistics Department at the Government of the Republic of Lithuania.

Law of the Seimas of the Republic of Lithuania No I-1489 of 13 March 2000, Law on Education of the Republic of Lithuania (State News, No. 23–593).

Law of the Seimas of the Republic of Lithuania No. VIII-1586 of 21 March 2000, Law on Higher Education of the Republic of Lithuania (State News, 2000, No. 27–715).

Law of the Seimas of the Republic of Lithuania No. VIII-450 of 14 October 1997, Law on Vocational Education and Training of the Republic of Lithuania (State News, 1997, No 98-2478).

Lithuanian Education in Figures 2006 (2006). Vilnius: Education Supply Centre.

Lithuanian Statistical Yearbook 2006 (2007). Vilnius: Statistics Department at the Government of the Republic of Lithuania.

Modernization of Education and Teaching System: In Pursue of Joint EU Goals Set for 2010 (2005). Vilnius: Education Supply Centre.

The Implementation of the State Education Strategy 2003–2012.

Materials for discussions (2007). Vilnius: Education Supply Centre.

Relevant Websites

<http://www.aikos.smm.lt> – Aikos.

<http://www.ipc.lt> – ITC.

<http://www.svis.smm.lt> – ITC Švietimo valdymo informacinė sistema (ITC Education Management Information System).

<http://lma.lt> – Lietuvos Mokslo Akademija.

<http://www.vmsfondas.lt> – Lithuanian State Science and Studies Foundation.

<http://www.egzaminai.lt> – NEC – Nacionalinis egzaminų centras.

<http://www.pprc.lt> – Pedagogų profesinės raidos centras (PPRC).

<http://www.smpf.lt> – ŠMPF.

<http://www.sppc.lt> – Specialiosios pedagogikos ir psichologijos centras ((SPPC) Special Education and Psychology Center).

<http://www.stat.gov.lt> – Statistikos Departamentas.

<http://www.skvc.lt> – Studijų kokybės vertinimo centro ((SKVC) Study Quality Evaluation Center).

<http://www.smm.lt> – Švietimo ir mokslo ministerija (Ministry of Education and Science).

<http://www.mokslas.lt> – Švietimo ir mokslo ministerija (Ministry of Education and Science).

<http://www.pedagogika.lt> – Švietimo plėtros centras (ŠPC) – Education Development Center (SPC).

<http://www.tpa.lt> – Tarptautinio mokslo ir technologijų plėtros programų agentūra (International Science and Technology Development Program Agency).

Malawi

F R Mkandawire and D Mulera, Malawi National Commission for UNESCO, Lilongwe, Malawi

© 2010 Elsevier Ltd. All rights reserved.

General Background

Malawi is found on the southern part of the East African Rift valley lying between latitude 9 and 17° E and longitude 33 and 36° S of the equator. The country has an area of 118 484 km² of which 94 276 km² is land mass.

Malawi's total population was estimated to be around 13 million in 2007 based on data from the Population and Housing census conducted in 1998. The most recent population and housing census was conducted in June 2008. Based on the 1998 population and housing census projections, Malawi's population density is 105 persons per square kilometer. The population growth rate is 2.2% per annum. The percentage of the population under 18 years is 50.5. Malawi's socioeconomic indicators are poor and among the lowest in the world. Its gross national income (GNI) per capita was averaged less than US\$200 over the past decade. According to a recent household survey, 52% of the population in Malawi are classified as poor while 22% are classified as ultra-poor, that is, living in extreme poverty (National Statistical Office, 2005). The incidence of poverty is severe in rural areas with an estimated 56% of the rural population classified as poor compared to 24% of the urban population. Inequality levels are quite high with 20% of the richest consuming 56% of the wealth and 20% of the poorest consuming only 5% (National Statistical Office, 2005). Infant mortality stands at 76 deaths per 1000 live births, child mortality at 133 deaths per 1000 live births, and maternal mortality at 984 deaths per 1000 live births and are among the highest in the world (NSO and ORC Macro, 2005). Malawi is also one of the countries worst hit by the acquired immune deficiency syndrome (AIDS) epidemic with an estimated 14.1% of the sexually active population (i.e., 15–49 years) infected with human immunodeficiency virus (HIV; NSO and ORC Macro, 2005). As a consequence of the AIDS epidemic, life expectancy for the average Malawian decreased from 45 years in 1995 to 39 years in 2003. However, the recent Human Development report puts Malawi's life expectancy at 44.5 years (UNDP, 2008). This is attributed to the improved nutritional status due to the bumper harvest of the staple food, maize, that Malawi has experienced from 2006 and improved campaign on the fight against HIV and AIDS (UNAIDS, 2008).

Since the implementation of structural adjustment programs from the early 1980s, Malawi's economic performance was erratic averaging 1% between 1980 and 1994. In 1994, Malawi went through a political transition from a

one-party dictatorship to a multiparty government. At that time, Malawi had an inflation rate of 70% and a budget deficit of 28% of gross domestic product (GDP). Good progress was, however, made in the immediate post-1994 years. Fiscal deficits fell from 28% to 8% in 1997. Inflation fell to 9% in 1997. Growth in GDP averaged over 3% in the period 1992–97. However, during the second multiparty term, between 1999 and 2004, of the then ruling party there was a decline in economic performance. Malawi's official currency, the Kwacha, experienced rapid devaluation from MK68 in December 2000 to MK107 to one USD by January 2004. This was attributed to, among others, withdrawal of external budgetary support from donors and high domestic borrowing by government from the banks. Poor rain pattern coupled with poor policy decisions regarding the sale of the strategic grain reserves led to serious shortage of the staple food, maize, triggering high inflation rates.

The country's economy is currently growing at 8.5% which is above the Southern African Development Community (SADC) average rate of 5.5% and the world's average of 5.4%. The inflation rate decreased from 17.1% in February 2006 to 14.9% in July 2006 and to 7.9% in July 2007. The depreciation of the Kwacha has been contained to less than 10% for a period of 2 years. There has been improvement in managing public finances and domestic debt repayment has improved which has seen interest rates being reduced from 30% in May 2004 to the current 15%. The percentage of people that lives below the poverty line has moved down from 60% to 52%.

Education Policies and Reforms

The legal framework of education is based on the 1962 Education Act, which is currently under review to take into account the current sociocultural and economic environment. Malawi's Constitution defines the nation's educational objectives and regulates the sharing of responsibilities for education among the three key players – the state, religious groups, and the private sector. The religious groups control a large proportion of the primary schools, around 60%.

The Constitution of the Republic of Malawi recognizes the right to education for all persons under chapter IV, Section 25.

The broad policy on education is to develop an efficient and high-quality system of education of a type and size appropriate both to the available resources and to the political, social, and economic aspirations of the nation.

Owing to the limited employment market, the emphasis of education is to equip students with the skills and desire for self-employment and entrepreneurship rather than conventional wage employment (MOEST, 2001). The vision of the education sector in Malawi, as outlined in the National Education Sector Plan (MOEST, 2008), is to be a catalyst for socioeconomic development, industrial growth, and an instrument for empowering the poor, the weak, and voiceless. Its mission is to provide quality and relevant education to the Malawian nation, enabling people to acquire relevant knowledge, skills, expertise, and competencies to perform effectively as citizens, work-force, and as leaders of Malawi, thereby reducing poverty. In implementing the vision and mission, Malawi has made the improvement of quality, equity, relevance, access, and efficiency in basic education as the main strategic priority. Complementing this main priority, at secondary level, the focus is to double the enrolment over the 10-year plan period while at the same time focusing on upgrading quality and on retention of girls. Access to technical and vocational education will expand and the colleges will be rehabilitated. At higher education level, enrolment in existing public universities will be doubled with significant increase in efficiency to minimize costs to the budget, complemented by expansion of private tertiary education. HIV and AIDS mitigation is considered a cross-cutting intervention at all levels/subsectors. The prioritization of the sector's interventions and strategies is premised on the fact that although there has been significant progress in terms of enrolment of school-age children (an increase of 60%) since the introduction of the free primary education (FPE) following the government's commitment in 1994 to the Millennium Development Goals (MDGs) for education and the Dakar, 2000 Education For All (EFA) initiative, both of which prioritize primary school-age children, the quality of primary education is acknowledged to be unsatisfactory and completion rates are very low (approximately 25%). The secondary and higher education sub-sectors (including technical and vocational training) also face some challenges related to quality, equitable access and relevance, and problems related to resource management.

Structure and Operation of the Education System

The Ministry of Education, Science and Technology (MOEST) is responsible for the provision of primary, secondary, and tertiary education. The Ministry of Women and Child Development (MOWCD) is responsible for early childhood development which encompasses pre-primary education, and adult literacy and education, while the Ministry of Youth Development and Sports (MOYSD) is responsible for aspects of nonformal education for out-of-school youths. While the government is the main provider of education at all levels, the private sector is currently

active in the provision of pre-primary, primary, secondary, and tertiary education.

The formal education system under MOEST is divided into three main levels – primary, secondary, and tertiary – and follows an 8–4–4 structure: 8 years of primary, 4 years of secondary, and up to 4 years of tertiary education.

Scope of Basic Education

The Malawi EFA Plan of Action and the NESP recognize the scope of basic education to be from preschool through years of primary school to secondary education as adopted by member states of the SADC in the SADC Education Protocol. This includes all those who may have dropped out of primary school and are pursuing vocational or skills training, and those who enter education at an advanced stage.

Early Childhood Care and Education

Currently, there are 4005 Community-Based Child Care (CBCC) centers and 1940 preschool centers in Malawi that provide early childhood development (ECD) services. A total of 6240 early learning centers have been opened and are operating across the country. A total of 615 478 children, orphans, and other vulnerable children of ages 1–6 targeted through either the preschools or the CBCC program have been reached representing a national coverage of 26.76% of children below 6 years – 2006. A total of 14 127 caregivers were registered within communities with 5335 targeted through a 2-week basic training in ECD. The concentration trend is now changing, majority of CBCCs are concentrated in rural and peri-urban areas with the aim of providing psychosocial support and care to orphans and vulnerable children, whereas most preschools are in urban and semi-urban areas, and the majority of caregivers in both situations are not trained.

So far, early learning programs cater for only 26.7% of the targeted population. This implies that almost 70% of children aged 0–5 do not have access to early learning services.

Primary (Elementary) and Secondary Education

Primary school enrolments which had been expanding, albeit gradually since independence, increased extremely rapidly in 1994 following the introduction of FPE when an additional 13 million children enrolled in school (from 1.9 million to 3.2 million), representing an overall increase of 51% and nearly 60% increase in enrolments in the first grade alone. Prior to the introduction of FPE, 50% of the

school-age children in Malawi were not attending school and the majority of these were girls. In 1990, girls comprised less than half the enrolment – 45%. The proportion of girls steadily increased over the years with remarkable progress seen in the period when girls' education became a focus of government's policy from 1991 with the launching of the Girls Attainment in Basic Literacy and Education (GABLE) program funded by United States Agency for International Development (USAID). During this period, the proportion of girls increased to 48% of the primary school enrolments in 1993–94 only to decrease by one percentage point in 1994 when FPE was introduced. The decrease was due to the fact that many more older boys took advantage of the FPE and enrolled in school as opposed to older girls. Girls' enrolments picked up again after FPE. By 2005, nearly 50% of the primary school enrolments were girls and by 2006 there were more girls enrolled in primary schools, which is reflective of the proportion by sex of the school-age population.

Secondary Education

Secondary education begins after the 8-year primary education cycle, and consists of a junior and senior cycle. The junior cycle consists of the first 2 years of secondary education after which students sit for the Junior Certificate of Education (JCE). The senior secondary education also consists of 2 years, after which students sit for the Malawi School Certificate of Education (MSCE). The MSCE opens a variety of opportunities in terms of employment and further education and training for those who pass this examination. The school census survey of 2005 indicated that there are 978 secondary schools, which enrol 183 854 students (about 17% of the age cohort), with girls comprising 43% of the secondary school enrolment. The secondary school education is offered mainly through four types of schools, namely: conventional secondary schools (CSSs), community day secondary schools (CDSSs), grant-aided secondary schools, and private secondary schools. CSSs are government-supported schools and are the most privileged with regard to educational infrastructure and quality of teaching staff. There are 95 CSSs enrolling 41 412 students, 41% of whom are girls. The CSSs enrol 23% of the secondary school population.

CDSSs, previously known as distance education centers (DECs), are also government schools established with community assistance and they are the most disadvantaged. There are 553 CDSSs with an enrolment of 84 351 students, 40% of whom are girls. The CDSS enrol 46% of the students at the secondary level, which is 66% of the students in the government-run schools. The third group of schools is grant-aided schools. These schools are owned

by nonprofit mostly Church organizations but are also supported by the government and are of similar quality as CSSs. There are 55 grant-aided schools enrolling 15 382 students, 49% of whom are girls. The students in the grant-aided schools make about 8% of the secondary school enrolment. The fourth category is that of private schools managed by individual entrepreneurs with no subsidy from government. Majority of these private schools are comparable to CDSSs from the viewpoint of the available teaching/learning resources.

There are 259 registered private secondary schools enrolling about 39 887 students, of whom 47% are girls. Students in the registered private schools make up 22% of secondary school enrolment. No accurate statistics are available on the number of students attending unregistered private secondary schools.

Access and Equity in Secondary Education

Expanding access to primary education has traditionally been a primary concern of the government of Malawi since 1994. This emphasis has resulted, however, in a lower attention to secondary education. However, the increased enrolment and retention at the primary level inevitably increases demand for the next levels. Accordingly, the government policy gives increased importance to raising the enrolment in secondary education from 17% in 2005 with 43% girls to 30% by 2012, with 50% being girls, as stipulated in the PIF. This would include an increased role for the private sector and community involvement in the provision of secondary education.

The government policy also makes provision for the promotion and strengthening of secondary education programs involving CDSSs, which provide secondary education to the neediest segment of the secondary school age group, and continuing education for the unqualified teachers who staff these schools. Furthermore, although the proportion of girls in secondary education has increased in recent years and some policies and actions to promote girl's education in secondary have been put in place, there is a need to examine more deeply the issues related to further expanding access for girls and improving their retention in the schools.

Quality of Secondary Education

In addition to the problem of low access, the quality of secondary education provided is poor, especially in CDSSs. Examination results taken as one of the standard measurements of quality reveal that the overall students' performance is poor. In 2006, only 29.98% of the students who sat for the MSCE passed the examination. This low quality of the secondary education provided is largely attributed to the inadequate levels of qualified teachers, insufficient teaching and learning materials, shortage of

teaching facilities, and deficient school management and supervision.

There are an estimated 8975 teachers (1723 or 19% female) in the secondary school system. About 72% of these teachers are not qualified for secondary school teaching, adversely affecting the quality of education provided. In most CSSs, the curriculum is delivered by qualified teachers, but in the CDSSs, the great majority (86%) of the teachers is academically and professionally underqualified. The conversion of distance education centers to CDSS in 1998 increased the need for secondary teachers as students are taught face to face. Some primary teachers have been reassigned to teach at CDSSs. These teachers are considered untrained or unqualified to teach at secondary schools until they complete a secondary-teacher-training program. The challenge for the Ministry is increasing the number of qualified teachers graduating from secondary teacher programs as well as upgrading the qualifications of the underqualified teachers. In addition, strategies for addressing these challenges need to be reviewed.

The inadequate provision of teaching and learning materials also negatively impacts on the education provided. In line with government policy of supplying schools with an initial stock of books, textbooks for form IV students were provided in 2004 to secondary schools that did not have the books. Textbooks for form III students were supplied in 2002 to some schools. However, most CDSSs are critically short of or lacking textbooks for forms I, II, and III, which make up over 50% of the secondary school enrolment in government-run secondary schools. Once an initial stock of textbooks is provided, availability of textbooks is sustained through the schools Textbook Revolving Fund (TRF). Schools retain MWK 250 of the school fees per student each year toward the TRF, which each school maintains to partially finance their school needs.

Similarly, the lack of sufficient number of classrooms, science-teaching facilities, and libraries, principally at the CDSSs, undermines the quality of teaching. In general, the condition of classrooms and science teaching and learning facilities (if at all they exist) in CDSSs does not provide conducive environment for teaching and learning to take place and most schools lack libraries and facilities for teaching science.

Technical and Vocational Education

In the face of serious unemployment and especially rapidly increasing numbers of unemployed youth, the government of Malawi initiated the reformation in 1994 of its technical education and vocational training policies (MOEST, 1994). Emphasis was to be placed on the preparation of the trainees for employment in the informal sector, especially by incorporating entrepreneurship

development. In 1998, the government launched the Technical Entrepreneurial Vocational Education Policy. The main objective of the new policy was to enhance the level of productivity in both the formal and informal sectors so as to contribute effectively toward poverty reduction. The new training system was to be very broad-based and no longer laying emphasis on formal educational background as a requirement for training.

As part of these changes, the Technical Entrepreneurial, Vocational Education and Training Act (TEVET Act, 1999) was enacted and the Technical, Entrepreneurial and Vocational Education and Entrepreneurship Training Authority (TEVETA) was created. TEVETA is an autonomous body that aims to regulate and coordinate the provision of technical education, vocational and entrepreneurship training in Malawi.

There are four different kinds of training providers for technical and vocational education (TVE): (1) technical colleges which fall under the Directorate of Technical Vocational Education and Training (DTVET), (2) community-based training centers, (3) church-based training centers, and (4) private commercial training providers, and there is the traditional apprenticeship as well.

The DTVET currently operates seven technical colleges offering formal training in some 22 trades in the fields of building/construction, engineering, horticulture, and printing.

Distance and Continuing Education

The Malawi College of Distance Education (MCDE) was instituted in 1965 under the name of Malawi Correspondence College which in 1987 was changed to MCDE. The idea of establishing MCDE was to meet the increasing social demand for secondary education.

Ever since secondary education was introduced in 1940, only a small fraction of primary school leavers has been able to continue with secondary education. With the attainment of independence, there was a great need for education in rural areas where the vast majority of the people lives. The shortage of resources, both in manpower and financial terms, caused the government to find alternative methods of providing education. MCDE was therefore established to increase accessibility to secondary education; that is, distance education. As described under secondary education, the MCDE centers were turned into CDSSs.

MCDE was also introduced to cater to secondary education needs of those who could not fit into conventional secondary education through its night secondary schools, for example, adults who had missed the normal opportunity to complete certain levels of education. Currently, MCDE remains with its night secondary schools,

correspondence students without face-to-face tutorial assistance and teacher-upgrading programs.

Special Education

This represents the specialized educational services delivery offered to children and adults who are unable to cope with the regular school/class organization and methods. The MOEST has developed a draft policy on special needs education and has plans to develop a national institute for special needs education to replace the only dedicated institute to special needs education since independence which has been turned into a private university by its proprietor, the Roman Catholic Church. The aim is to provide adequate education for all categories of children and adults who require special education services as well as to provide a diversified and appropriate curriculum for each category of disability.

Nonformal Education and Adult Literacy

Adult literacy and nonformal education programs in Malawi are implemented by several ministries. The main ministry that implements adult literacy programs is the Ministry of Women and Child Development. Other ministries include the Ministry of Youth Development and Sports (for nonformal education) and the Ministry of Education which is implementing a complementary basic education program targeting children who drop out of school and the overaged. Other ministries that are involved in one way or another in the implementation of adult literacy and nonformal education programs are the Ministry of Agriculture and Food Security, Ministry of Labour and Social Development, and the Ministry of Health. There are a number of nongovernmental organizations (NGOs) that are also actively involved in the implementation of adult Literacy and nonformal education programs.

The country's concerted efforts to fight illiteracy stretches back to the colonial period. However, the major initiative in the country's overall strategy to combat illiteracy was undertaken between 1981 and 1985. During this period, a functional adult literacy program was launched on a pilot basis, and in this program literacy was perceived as a method of providing illiterate adults aged 15 and above learning opportunities which they missed out in their youth. The idea was that this would enable them access information on health, nutrition, and agriculture and thus improve the standard of living. The pilot phase culminated in the launch of the National Adult Literacy Programme (NALP) in 1986, which is coordinated by the National Center for Adult Literacy and Education (NACLAE).

A recent development in the adult literacy education sector in the country was the launch of the Sustainable

Social and Economic Empowerment Programme (SSEEP) in 2003. The overall objective of the SSEEP, piloted in 12 districts across the country, is to promote the Regenerated Freirean Literacy through the Empowering Community Techniques (REFLECT) approach, which has been championed and popularized by several NGOs notably Action Aid. This approach encourages and enables participants to critically assess their lives, take control of their futures, enhance their literacy skills, generate a written vocabulary, which is relevant to their own community or situation, recognize and build upon their knowledge, and mobilize for individual and collective actions.

It is estimated that there are over 8000 adult literacy classes with an average of 25 learners per class operating across the country annually since the turn of the 1990s. This estimate includes literacy programs provided by other service providers besides the NALP such as NGOs, Faith-Based Organizations (FBOs), Community-Based Organizations (CBOs), and even the private sector. Classes run entirely by the NALP are estimated at 4400 countrywide annually. Some notable nonstate actors involved in the provision of adult literacy programs include Action Aid, World Vision, Oxfam, and the Presbyterian Church of Central Africa (CCAP). However, despite the apparent impressive coverage and a series of interventions stretching back to the colonial era, the problem of illiteracy remains intractable in the country.

The overall illiteracy rate is currently estimated at 42%, but when disaggregated in terms of gender more women than men are illiterate. The illiteracy rates for women and men are estimated at 76% and 23%, respectively, representing an illiteracy ratio of 3:1. The paradox, however, in Malawi is that illiteracy has been decreasing over the years but only in percentage terms. It has been steadily worsening in absolute terms. It is estimated that illiteracy stood at about 87.6% in 1966, 77.9% in 1977, 55% in 1987, and at about 40% following the 1998 Population and Housing Census (NSO, 1998). The number of illiterate adults was projected at 1 976 218 in 1977, 3 848 414 in 1987, and at about 4.6 million during the 1998 Population and Housing Census. The statistics for 2005 indicated that illiteracy in the country had dropped from 90% in 1964 to 37.2%.

The goal of the NALP at its launch was to reach out to at least 2 million of the 4.6 million illiterate adults by the year 2000, but by 2005, the program had only reached about 730 000 learners. These statistics clearly underlie the magnitude of failure of the set of interventions and programs that have hitherto been carried out to fight illiteracy in the country.

The Teaching Profession

There has been an improvement in teacher training and recruitment since 2000. According to Education Management Information System (EMIS), 2007, the teacher-pupil

ratio is 1:78, but the qualified teacher–pupil ratio is 1:88. This is an improvement from the teacher–pupil ratio of 1:118 in 2000 and qualified teacher–pupil ratio of 1:138 in the same year. The country is, however, short of 34 203 and 60 203 primary school teachers to achieve teacher–pupil ratio of 1:60 and 1:40, respectively by 2015. Annually, the country needs to recruit 5200 teachers on average and train 5244 teachers. The government has since 2007 introduced a new teacher training program (the 1+1 Teacher Training Programme) replacing the previous Malawi In-service Integrated Teacher Education Programme (MIITEP) in order to address teacher supply and teacher quality. In terms of gender, the country is short of at least 10 000 female primary school teachers now and needs to recruit around 30 000 more teachers by 2015 to close the gender gap between male and female teachers who are currently 26 893 and 16 304, respectively. There is a need to train and recruit between 10 875 and 15 500 specialist teachers to cater to children with special needs and an expanding special needs sector.

Primary school teachers are trained in four public teacher training colleges. Secondary school teachers are trained at public institutions which include University of Malawi (Chancellor College), Mzuzu University, and Domasi College of Education. Some private universities and colleges have recently been opened up which are also training teachers, such as the University of Livingstonia, African Bible College, the Catholic University, and DAPP Teachers Training College. Admission to these institutions is based on individual academic qualifications.

There is generally a critical shortage of teachers' houses, especially in rural areas which has been cited in some studies as a contributing factor to the inadequate numbers of teachers in the rural areas as majority of teachers, especially female teachers, migrate to urban areas.

Higher Education

There are two public universities in Malawi. The first is the University of Malawi (UNIMA) established in October 1964. The University of Malawi operates five constituent colleges: Chancellor College (for general degrees, education, public administration, and law), the Polytechnic (accountancy, engineering, and journalism), Bunda College (agriculture), College of Medicine, and Kamuzu College of Nursing. The second public university, Mzuzu University, opened its doors in 1998. Currently, there are plans for a third national university which will focus on science and technology.

In addition, in recent years, a number of private universities have been established by churches.

It is clear however, that Malawi has under-invested in the education sector overall, and, more particularly, in tertiary education. The university does not have sufficient

capacity to satisfy the needs of a rapidly growing population of students thirsty for a university education. The growth in absolute student enrolment has been quite modest – growing from a student intake of 90 in 1965 to 5000 in 2004–05 academic year.

Administrative and Supervisory Structure

The MOEST is managed through central, divisional (6), and education district (34) structures. Primary schools are further organized through a zonal system (315 zones) with each zone served by a teachers' development center (TDC). A similar cluster system has been developed for the secondary subsector, with approximately 74 clusters.

Community involvement and partnership among stakeholders are becoming accepted practices in many communities in Malawi, but the bulk of the partnership appears largely limited to efforts of parent–teacher associations (PTAs). Other groups and organizations in the communities are yet to identify fully with the school programs.

Educational Finance

The national budget allocation to MOEST has hovered around 20% in recent years. It should, however, be noted that this does not include education programs implemented by other ministries such as early childhood development and adult literacy, and out-of-school youth programs which are under the Ministry of Women and Child Development and the Ministry of Youth, Sports and Culture, respectively.

The major constraints and challenges of financing are:

- *Mobilizing sufficient funding for education subsectors.* The government should devise a means of mobilizing sufficient funds for the education sector at all levels. This means that all the stakeholders in the education industry should live up to their responsibility of making available sufficient funds for the running of the educational sector in Malawi. These stakeholders include the state, local government/assemblies, the private sector, members of the National Assembly, the civil society, and development partners.
- *Reducing hidden costs to parents of the less privileged.* Although primary education is free, there are some hidden costs that poor households encounter which outweigh the opportunity costs of sending children to school. Such hidden costs include school uniforms/clothes, and learning materials, for example pens/pencils, meals/snacks. A reduction in the hidden costs will relieve the heavy burden on parents of low-income groups. The government has a policy on noncompulsory uniform and has also introduced school feeding programs in some schools.
- *Expanding and strengthening partnerships.* Financing education appears to have become a heavy burden on

government and thus requires participation from other stakeholders. Hence, there is need for partnerships among stakeholders in the education sector, not only with international development partners, but also with private sector, nongovernment agencies, as well as other philanthropic organizations and individuals.

- *Prioritizing needs for optimizing scarce resources.* The current challenge in the education sector is the need for prioritization of needs because the available resources appear grossly inadequate. Currently, there are many compelling needs as far as costing and funding of the education subsector in the country is concerned. These include classroom construction, renovation, furniture provision, teachers' salaries, capacity building, special programs such as HIV and AIDS, problem of enrolment, and enhancing the quality of education delivery at all levels. Given these circumstances, efforts should be made to put these needs in order of their importance and treated as such to ensure optimum use of scarce resources. In this regard, it is required that school management determines which area of expenditure should come before the other. Resource allocation in the various subsectors of the education should also be based on simulation modeling, whereby allocation to each subsector is specified. This should be done in such a way that when all the subsectors are put together it would produce the best result in terms of costing and financing of education in Malawi.

Bibliography

- MOEST (2008). *National Education Sector Plan 2008–2017, A Statement*. Lilongwe.
- National Statistical Office (2005). *Integrated Household Survey 2004–2005. Volume 1. Household Socio-Economic Characteristics*. Zomba: National Statistical Office (NSO).
- NSO and ORC Macro (2005). *Malawi Demographic and Health Survey 2004*. Calverton, MD: NSO and ORC Macro.
- UNAIDS (2007). *Uniting the World against AIDS: Malawi Situation Analysis*. Lilongwe.
- UNDP (2008). *United Nations Development Framework (UNDAF) 2008–2011*. Lilongwe.
- MOE (2001). *Policy and Investment Framework for the Education Sector*. Lilongwe.
- UNAIDS (2006). *2006 Report on the Global AIDS Epidemic*. Geneva.
- UNAIDS Inter-Agency Task Team (IATT) on Education (2005). *Report on the Education Sector Global HIV/AIDS Readiness Survey 2004*. Geneva: UNAIDS.
- UNDP (2006). *United Nations Human Settlement Programme*. <http://www.unhabitat.org> (accessed September 2009).
- UNDP (2007). *The UN Business Plan 2008–2009: Unified UN Country Strategy – Malawi*. Lilongwe.
- UNESCO (2000). *Dakar Framework for Action, Education for All: Meeting Our Collective Commitments*. Paris: UNESCO.
- UNESCO (2005a). *Approved Programme and Budget 2006–2007 (C/5)*. Paris: UNESCO.
- UNESCO (2005b). *2006 EFA Global Monitoring Report – Literacy for Life*. Paris: UNESCO.
- UNESCO (2006). *Global Action Plan to Achieve the EFA Goals*. (Draft Edition of 5 June 2006). Paris: UNESCO.
- UNESCO (2007). *2008 EFA Global Monitoring Report*. Paris: UNESCO.
- World Bank Group; Department for International Development (UK) (2005). *Country Partnership Strategy for the Federal Republic of Malawi 2005–2009*. The document of The World Bank. (Report No. 32412–NG).
- World Bank Group; Department for International Development (UK) (2005). *Millennium Development Goals – Malawi*. Lilongwe.
- World Bank Group; Department for International Development (UK) (2007). *Country Assistance Strategy (CAS) 2007–2010*. Lilongwe. <http://go.worldbank.org/GZOBGT43> (accessed September 2009).

Further Reading

Malta

C Farrugia, University Ombudsman, Valletta, Malta

© 2010 Elsevier Ltd. All rights reserved.

Malta is a small island state, the smallest in the European Union (EU). A total landmass of 316 km² and a population of 405 000 render Malta one of the most densely populated countries (1282 km⁻²). The Islands have an aging population currently averaging 38.5 years, a birth rate of 2.6 per mother, while longevity is 80 years for females, 77 for males. The literacy rate is 92.8% with the majority of illiterates falling in the 60+ age group.

The Maltese Islands lie at the crossroads of the Mediterranean: 100 km south of Italy, 280 km off the North African coast, and almost equidistant from Spain and Greece. Their extensive, sheltered, and deep natural harbors have attracted the dominant seafaring powers in the region. Romans and Carthaginians fought over Malta in the Punic Wars. So did the Byzantines and the Arabs in the Middle Ages, the Christian Knights of St. John and the Muslim Turks in the sixteenth century, the British and the French in the nineteenth century, and the Allies versus the Axis during World War II. Malta's best-kept secret is the origin of the numerous Neolithic Temples dating from 5500 BC and acknowledged as the oldest freestanding buildings in the world. The vast majority of Maltese are Roman Catholics, their predecessors having been converted to Christianity by St. Paul in AD 60. The national language is Maltese, the only Semitic language written in Roman script, while English is the second official language. Many Maltese also speak Italian.

Malta is a parliamentary democracy, gaining independence from Britain in 1964 and becoming a republic in 1974. The Islands joined the EU in 2004, but remain an active member of the British Commonwealth. The Nationalist Party (Christian Democrats) has been in power since 1987, except for the period 1996–98 when the Labour Party (Social Democrats) enjoyed a parliamentary majority. In tune with its small-island-state characteristics, social, cultural, financial, environmental, and religious issues in Malta tend to become highly contentious, often divisive, and invariably politicized, even though the two main political parties espouse similar national policies.

Local economists disagree whether Malta prospers most in time of peace or in time of war. Economic data show, however, that the national economy has flourished since independence when the Maltese embarked on an industrialization drive to divert reliance on British military expenditure to locally generated incomes. Up to the 1990s manufacturing enterprises concentrated on textiles and light industries especially the manufacture of electronic components, while shipbuilding and ship repairing became

a financial hemorrhage. Tourism and construction have evolved as the mainstays of the Maltese economy. Currently, financial services, international banking, e-commerce, and information and communication technology (ICT) make strong inroads as new major employers and important foreign currency earners. In mid-2007, Malta's gross domestic product (GDP) stood at €1504.6 million in real terms, gross national income (GNI) at market prices was €1332.3 million, with unemployment at 7.12%.

History of Education

The first mention of a public school emerges *c.* 1460. The school for boys was co-financed by the Cathedral and the Municipality of Mdina, then the capital city. Girls were tutored in private homes.

On arrival in Malta in 1530, the Knights of St. John set up a school for their young recruits. Members of a naval power, they were taught mathematics, navigation, cartography, and architecture; and in line with their primary vocation as knights hospitalers; medicine and surgery. In 1592, the Knights commissioned the Jesuits to open the Collegium Melitense, which became a public university in 1769, the precursor of the present University of Malta.

The education provided at the time, as in the early years of British rule, catered to the elite. The British launched public-funded education for the masses following the 1838 Royal Commission Report, which castigated the colonial administration in Malta for neglecting to educate the common people. The criticism served its purpose and by 1900 every town and most villages had their own primary school administered by the government's Education Department. Concurrently, the University's Lyceum provided secondary education for boys. In the mid-1920s secondary education for some 5% of the age cohort of both sexes became the responsibility of the Education Department. During the same period, private education evolved through church schools and private colleges. Educational policy in Malta followed closely the UK's, a trend that persists today.

Compulsory school attendance became law in 1924, but could not be enforced until the end of World War II. An extensive schools' repair and building program rendered full-day free primary and secondary schooling possible in the mid-1950s. Free University education was introduced in 1971. In 1974, schooling became compulsory from age 5–16.

Today, two-thirds of the Islands' student population attends state schools; the rest go to church or independent schools (see Table 1). The number of students in the nonstate sector shot up in the mid-1970s when a well-intentioned but badly planned and rashly executed switch from grammar- to comprehensive-type education in the state secondary sector, prompted many parents to transfer their children to church schools. State education still suffers from the shockwaves. Several social engineering devices were tried to stem the flow. First, was the forced abolition of church school fees to render them financially nonviable. This measure had the direct opposite of the intended result since it made the enrolment into church schools less costly, even when parents made voluntary donations. Second, was the establishment of state-run Junior Lyceums, identical to the old grammar schools. The second move, together with the demand for entry into church schools, which far exceeded the supply, led to two highly competitive eleven-plus-type secondary-school entrance examinations. Preparation for these examinations, including extensive private tuition, thwarted the aims of primary education.

All state education from kindergarten to university is free; currently the State heavily subsidizes church schools, while independent schools (very much a minority) charge fees with parents receiving tax rebates. Of the

€2 498 463 000 estimates for 2008, the government devoted 10.88% to education.

In 1978, student stipends were introduced to spur youngsters to tertiary education with the argument that human capital is Malta's main natural resource and it behooves the state to nurture and develop it. Local pundits doubt whether Malta's educational budget affords to devote €21 million, or 7.7% to this extraordinary expenditure, and yet no party in government dares remove it.

Structure of the Educational System

The Minister of Education is responsible also for Culture, Youth and Sports. Her remit reflects the government's determination to transform the Islands' into a hi-tech servicing community with education as the foundation for future economic and social development. Figure 1 illustrates the responsibilities of the education directorates.

A major structural reorganization in the delivery and monitoring of the educational services occurred in 2007 when an Amendment to the Education Act transformed the Education Division into two new directorates. The Directorate for Educational Services includes schools as the service providers; it also acts as the resources provider supplying schools with their human and material

Table 1 Malta: Education Statistics, 2005

	<i>Institutions</i>	<i>Students</i>	<i>Male</i>	<i>Female</i>	<i>Teachers</i>	<i>Male</i>	<i>Female</i>
Totals	251	88 828	45 541	43 287	8119	2543	5576
Government	122	63 649	32 679	30 970	5943	2089	3854
Kinder		5253	2680	2573			
	78 ^a				2259	283	1976
Primary		18 435	10 006	8429			
Special	6 ^b	251	151	100	148	17	131
Secondary	32	18 251	8909	9342	2351	961	1390
Postsecondary	3	4965	2040	2925	211	130	81
Postsecondary Vocational	2	6964	4720	2244	329	210	119
University	1	9530	4173	5357	645	488	157
Church	83	16 928	8075	8853	1517	337	1180
Kinder	34	1444	636	808			
Primary	25	7260	2997	4263	Teachers in church schools teach across levels		
Secondary	22	7576	4107	3469			
Postsecondary	2	648	335	313			
Independent	46	8251	4787	3464	659	117	542
Kindergarten	24	1854	976	878			
Primary	13	3901	2398	1503	Teachers in independent schools teach across levels		
Secondary	8	2377	1341	1036			
Postsecondary	1	119	72	47			

Part-time evening: 17 299 students follow courses in 125 government and private institutions.

^aKindergartens and primary schools are often in the same building and fall under the same administration.

^bAs many as possible students with special needs are integrated in mainstream schools.

Compiled from National Statistics Office "Official Statistics of Malta 2005 - Education Statistics", Valletta, Malta, 2007.

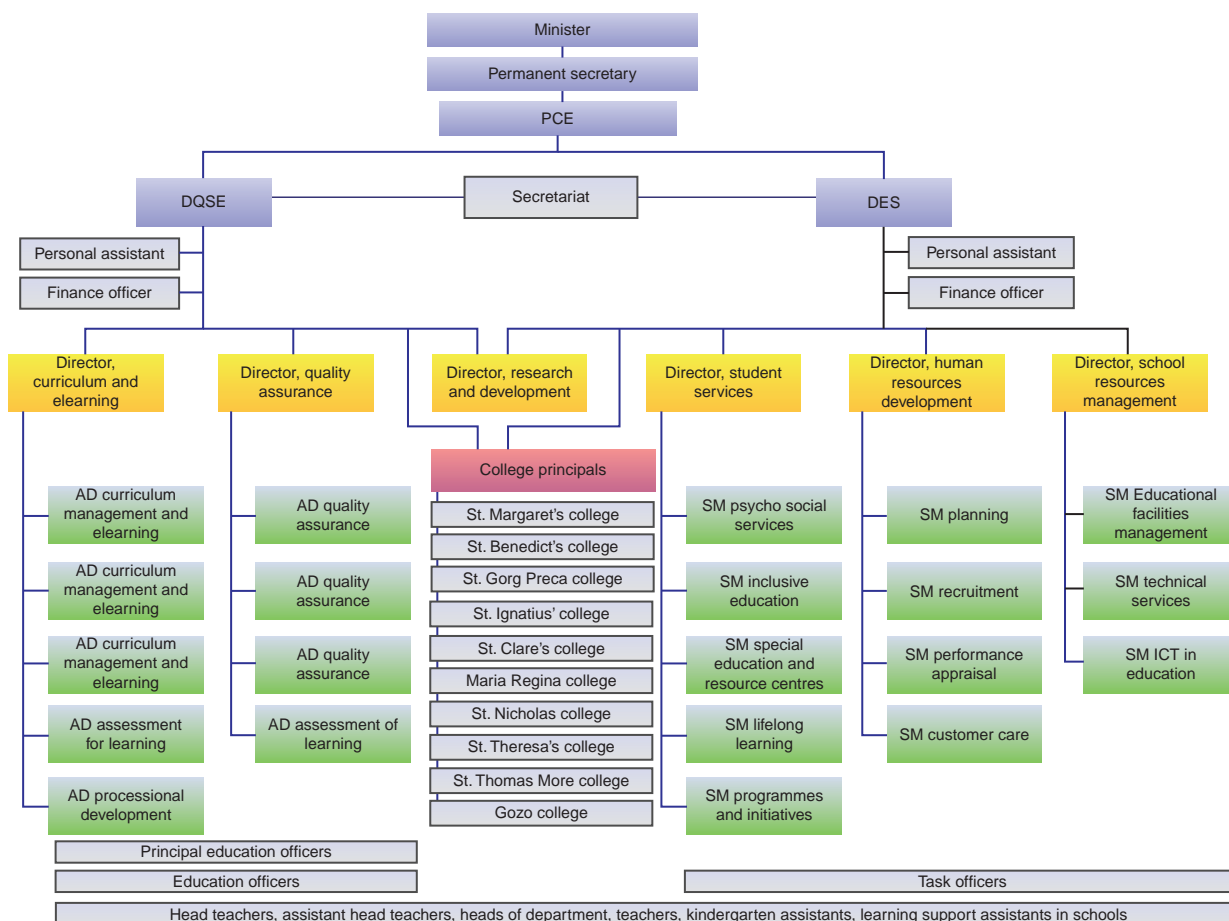


Figure 1 The responsibilities of the education directorates. From: Ministry of Education, Culture, Youth and Sports (2008). *Malta: A Guide to Education and Vocational Training*. Floriana, Malta: Government of Malta.

requirements. The Directorate for Quality and Standards in Education monitors the delivery of educational services to ensure best practice. Each directorate is headed by a director general answerable to the permanent secretary and the minister. An important provision in the Amendment allows both Directorates much greater freedom of operation than the old Division enjoyed.

Education Paths

Figure 2 represents the educational paths that Maltese students can follow.

Day Care Centers and Nurseries

The state does not offer services at this level; they are provided by church organizations and private entities. The government does ensure that such services, as well as kindergartens, comply with its guidelines.

Kindergartens

Parents send their children to either free-of-charge state kindergartens or fee-paying ones run by church and independent schools. All three sectors provide good services based on structured play and social development, and the parents' choice depends more on convenience, or the belief that what one pays for is bound to be superior to what is free, than on an impartial evaluation of pedagogical worth. The quality should improve further in the near future when early childhood education teachers will have to hold a Bachelor's degree in the area.

Almost 93% of 3- and 4-year olds attend kindergartens. As a result, grandparents play a crucial role in the upbringing of kindergarten- and primary-aged children where both parents work. Primary schools and kindergartens close at 2.30 p.m. with many grandparents minding the children until the parents return home. Problems arise when elderly grandparents babysit two or three children, or conversely when younger, working grandparents cannot provide the service. It is only a matter of time before the state,

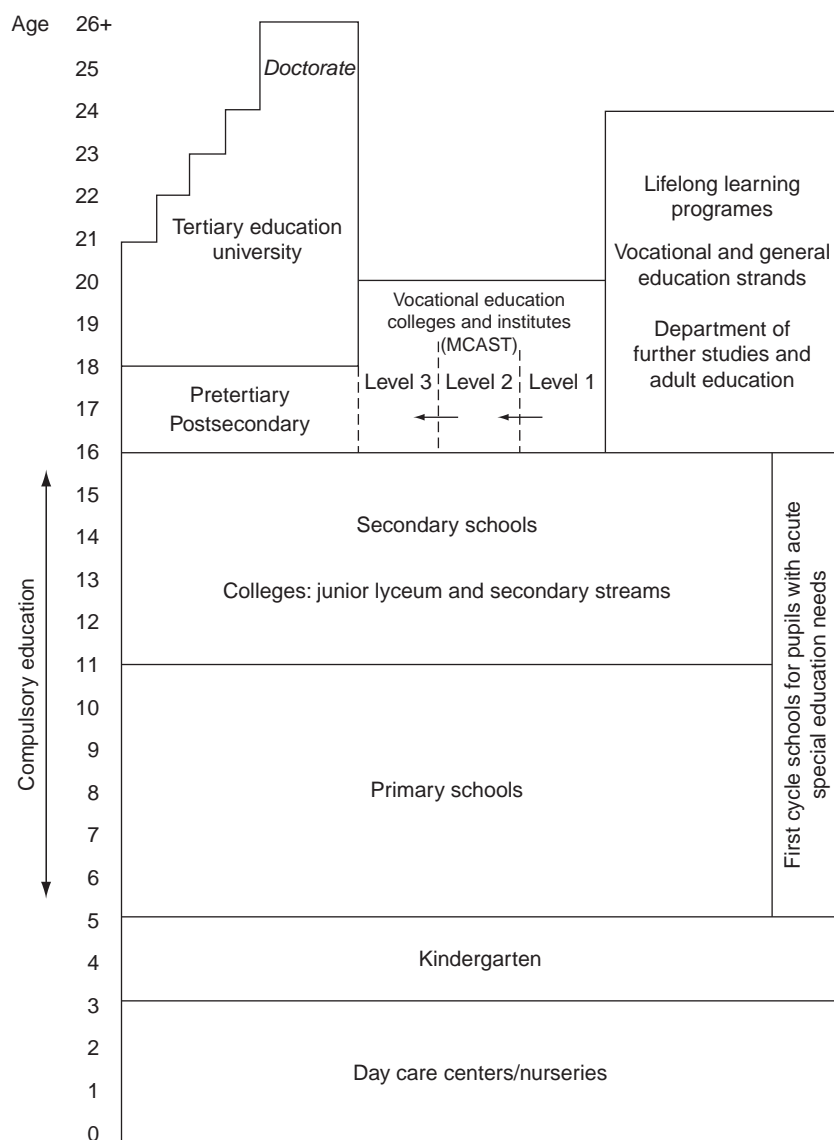


Figure 2 The Maltese mainstream education system. From: Ministry of Education, Culture, Youth and Sports (2008). *Malta: A Guide to Education and Vocational Training*. Floriana, Malta: Government of Malta.

in addition to campaigns to increase the female workforce, will have to offer extended after-school services.

special-needs students in primary and secondary schools, and retains this extremely expensive commitment as an integral part of its inclusive-society policies.

Special Education

The Maltese education system provides robust provisions to integrate the maximum number of students with special needs into the mainstream of the educational process. As a result, only a small fraction of students attend special-education institutions and that too, only when the children are at risk in normal schools. An ambitious support system provides one-to-one facilitators to students needing help to integrate with their intellectually or physically stronger classmates in the mainstream. The state provides the facilitator service free of charge to all

Primary Schools

Children aged 5–11 attend primary schools. The scholastic year starts in mid-September and ends in June, comprising 957 hours of schooling. Primary state and independent schools are coeducational, as are some primary church schools. All state and church secondary schools are single-sex institutions.

All schools in Malta are expected to conform to the National Minimum Curriculum established in 1989, revised

in 1999, renewed into the National Curriculum Framework (NCF) in 2006, periodic revisions since then. The NCF is a far-reaching work in which local experts in the theory and practice of education expound the goals and quality of education Maltese students deserve.

State education officers try to ensure that schools follow the NCF. In the case of primary schools they face a major obstacle. The first 3 years of state primary schooling are generally tranquil for pupils, their parents, and the teachers. The learning is experiential with emphasis on literacy in Maltese and English, numeracy, general knowledge, physical and social development, with prominence given to self-expression through various media. The learning is broad and the pace moderate. Pupils move from one level to the next on the basis of continuous assessment. The situation changes drastically in year 4 when the race toward the secondary school entrance examinations begins. Admission to the state's grammar-type Junior Lyceums and the church's colleges is highly selective, greatly sought after, and consequently exceedingly competitive.

Students wishing to attend independent schools pay for the privilege. Those seeking a place in prestigious state or church secondary schools endure the ordeal of preparing for the 11-plus-type Junior Lyceum Entrance Examination (JLEE) in the case of state schools, or the Common Entrance Examination (CEE) for entry into Church schools. Practically all the children finishing primary school sit for both, although efforts are under way to half the ordeal by amalgamating the two into one national examination. Unfortunately, these examinations condition all pedagogical procedures in the last 3 years of primary school. Pupils are subjected to one continuous swotting exercise toward successful results: the higher the marks, the better the chances of entry into the desired school. In the process the aims of the NCF are seriously undermined.

Pupils attending those church and independent schools that offer both primary and secondary education are more fortunate: they simply proceed from one level to the next without the need to sit for examinations. The demand for places in these church schools is so great that pupils are admitted by a public draw.

Secondary Schools

Successful students in either the JLEE or the CEE proceed to a state Junior Lyceum or a church secondary school as the case may be. The rest either opt for a fee-paying independent school, or attend a tuition-free state area secondary school. Parents and students alike label the latter failure schools. This label is unjustified, because a high number of students will have missed entry into the more sought-after schools by a few percentage points in one of five subjects. With guidance, such students have proceeded to

Matriculation and university studies in tandem with their Junior Lyceum and church schools' counterparts.

Most educators and some parents condemn the transfer of pupils from primary to secondary schools through such a competitive and stress-generating procedure. However, even if parents complain about the enormous tension that the JLEE and the CEE generate, the majority favor this differentiation and selection process. They regard the entrance examinations as the fairest and best chance their children have of attending the better secondary schools. Community pressure has prevented efforts to abolish the examinations.

Another attempt to lessen the impact of the JLEE is currently being tried in state schools, and its potential success is heightened by the fact that the model already exists in several church and independent schools. Students in these schools are admitted at kindergarten or primary levels and proceed to the secondary level of the same or an associated school without having to swot and undergo the trauma of the eleven-plus. This author researched the procedures and their outcome, and proposed the adoption of this system in state schools. The Minister and his advisors saw the potential of the recommendations and expanded them significantly to establish ten state colleges and their networks.

Each college incorporates a Junior Lyceum, an area secondary school, and several feeder primary schools. Students from the Junior Lyceum and the area secondary school merge and are set in classes according to their abilities in the various school subjects. The pupils' progress from the feeder Primary schools is evaluated on a continuous assessment package which, so far, includes a modified form of the JLEE. When the JLEE is sufficiently transformed, primary school pupils will be able to proceed to their college's secondary level without differentiation. The college concept goes beyond the academic sphere since the network's administration, resources, and finances are organized into one management structure regulated by the college principal and the heads of schools within the network. They are allowed a high degree of autonomy, which in the past, only nonstate schools enjoyed.

The college networks are still at their initial stage. Should they succeed in building their ethos and academic reputation sufficiently to defuse the negative impact of the JLEE, or abolish it completely, the sound educational aims of the NCF can come to fruition. Senior state education officials hope that church schools will join the state college networks. This will not occur before church schools feel convinced that they will retain their autonomy, an expectation that in many ways counteracts the principles of networks.

The secondary school program lasts 5 years, at the end of which students sit for the Secondary Education Certificate (SEC) enabling them to proceed to postsecondary education.

Postsecondary/Sixth-Form Colleges

In 2007, five institutions provided a 2-year postsecondary course for 6200 students, 3200 of them attending the University's Junior College. The course leads to the Malta Matriculation Certificate (Matsec) and entry to university. During these 2 years, students replicate their stress-ridden experiences prior to secondary school entry, this time in efforts to join their preferred university course. Although the *numerus clausus*, or restricted entry number, into certain courses no longer applies, most University of Malta courses have entry requirements far more demanding than their counterparts in renowned institutions worldwide.

The government aims to increase the number of students attending postsecondary and tertiary education to the higher percentages prevalent in the EU and provides students with monetary incentives to continue studying beyond secondary education. As a result, all postsecondary and undergraduate education in Malta is tuition free. In addition, students are paid monthly stipends in order not to become a financial burden on their families for the duration of their studies. Furthermore, the state has launched schemes of postgraduate scholarships, mainly in science and technology areas, tenable in Malta and overseas.

Economic commentators claim that such state expenditure is unsustainable. Yet, politicians on both sides of the House of Representatives argue that the expenditure is a justified investment in Malta's human capital. Enrolment figures bear out this argument since the number of university students shot up from 1447 in 1987 to 10 000, 20 years later.

Malta College of Arts, Science and Technology

Malta College of Arts, Science and Technology (MCAST) was revived in 2001 to provide pretertiary vocational education. The institution is an amalgamation of pre-existing vocational schools, renamed institutes, brought together to offer coordinated, harmonized, and improved educational services. Its institutes are: agribusiness, art and design, building and construction engineering, business and commerce, community services, electronic and electronics engineering, information and communication technology, maritime, mechanical engineering, and the MCAST Gozo Centre. The Institute of Tourism Studies, with 900 students is expected to join the MCAST structure in the near future.

The institutes provide tuition to 6000 full-time equivalent students, with ICT and business and commerce

being the most popular. The academic standing of courses varies from sixth-form equivalents to almost first-degree level. The government is investing heavily in the institution and plans that in due course MCAST will become Malta's second university. This should come about through the initiatives of the National Council for Tertiary Education which coordinates the work of MCAST and the University in line with the Islands' social and economic aspirations.

The University of Malta

Malta's university prides itself as the oldest in the British Commonwealth outside the UK. From a Jesuit College in 1592 with the authority to confer degrees, it became a public university in 1769. Arts (including mathematics and science), law, medicine and surgery, and theology formed the foundation faculties; science, dentistry, and architecture and civil engineering were established in the mid-twentieth century; education, engineering, and economics, management, and accountancy were upgraded from polytechnic departments to the University of Malta (UoM) faculties in the late 1970s. The Faculty of Information and Communication Technology is a 2007 creation. In 1999, the University opened its Link Campus in Rome, Italy.

The University also houses 27 institutes and centers, with the Institute of Health Care being the largest. Institutes and centers complement the faculties' work mainly through cross-subject, interfaculty teaching and research. The University's entities, with the Senate's approval, offer certificate, diploma, bachelor, master, and doctoral-level courses, which follow European Education Area directives and the Bologna Agreement of study-units and academic-credits structures. In October 2007, the University had 10 000 full-time equivalent students, of which 750 were foreigners.

A recent phenomenon has been the opening in Malta of numerous branches of EU, mainly British, universities with students enrolling on a fee-paying, part-time basis. Their chief attraction lies in offering subjects which either are unavailable at the UoM or whose entry requirements are not as demanding.

Justifiably proud of its history and traditions, the UoM occasionally finds it hard to shed past practices. Most course content, teaching, and research are at the cutting edge, but administrative and managerial procedures lag behind. In 2006, Council, the University's highest governing body, appointed a new rector with the primary remit to restructure and upgrade the institution's administrative and financial management. Currently, the University strives to enhance quality assurance, minimize bureaucracy, and maximize the efficiency of its human and material resources without sacrificing academic standards.

It endeavors to become less dependent on public funding by generating income through specialized services. The UoM aspires to become administratively and financially autonomous from the government to the same high degree that it enjoys in academic freedom. One augurs that in the process the university will not mutate into a super-efficient bureaucracy driving a soulless institution.

Adult and Continuing Education

The Maltese educational system provides services to encourage and support learning beyond the normal schooling provisions. It offers courses in self-improvement, hobbies and leisure, traditional arts and crafts, as well as in drama and music. Some 600 senior citizens attend the University of the Third Age (U3A) each year. A particularly successful venture is the Employment Training Corporation (ETC) which, besides serving as the national employment agency, runs worker retraining courses, sometimes on its own, more often in collaboration with the University, MCAST, and local industry.

Teacher Education

State-employed teachers began to receive rudimentary forms of teacher training in the nineteenth century. Proper teacher education began in 1946 with the opening of two residential colleges, one for males, the other for females. The initial 1-year full-time course was eventually extended to 2-, 3-, and 4-year durations. Both colleges were closely linked to, and externally examined by, UK colleges associated with the Institute of Education at the University of London. By the mid-1970s, the colleges were reorganized first as an amalgamated Malta College of Education, and later as the Department of Educational Studies at the local polytechnic with sufficient academic clout to offer a bachelor of arts (BA) in Education.

In 1978, the Department evolved into the Faculty of Education at the University of Malta, with this author as its founder Dean. It now offers qualifications ranging from undergraduate diplomas to PhDs with external examiners from European universities forming part of its assessment procedures. The Faculty's contribution to education in Malta has been substantial: the vast majority of teachers hold a bachelor of education (BEd) (Hons) or a Postgraduate Certificate in Education (PGCE) with specializations in either early childhood, or primary or secondary education. Heads of schools must hold the Faculty's Postgraduate Diploma in Education (Administration and Management). The Faculty also offers degree courses in psychology and youth studies. The academic staff has developed a strong research and publications profile

with engagements in national and international projects. Observers argue, with some justification, that in its quest for academic respectability, the Faculty has neglected somewhat the practical pedagogical formation of its student-teachers. The current Dean, the first female Dean at UoM, plans to redress the balance.

The Education Act of 1988 recognized teachers as professionals placing them on the same salary scale as medical doctors and lawyers in public service. In reality, however, they have yet to gain the social status and prestige of traditional professions. The 2006 Amendment to the Education Act established a Teachers' Council with the aim of promoting teachers' work, monitoring their output, and in the process issuing teacher's warrants in lieu of the Ministry. The Malta Union of Teachers (MUT) ardently supports the Council and has a strong representation on it. The MUT, established in 1919 as Malta's first trade union, is powerful and influences government policy in education, mostly with positive results.

Conclusion

The demands made by the extensive education services in Malta constitute a heavy burden on the Islands' limited human and financial resources. In spite of this, the state is committed to provide them, even when the community asks for more. All the indicators show that the state's commitment to and investment in educational development are bearing fruit. One example: the rundown of the British military base in the 1950s and 1960s caused widespread unemployment and hardships with mass emigration being the only solution. In contrast, Malta is managing remarkably well the current transformation from a manufacturing, labor-intensive economy to a striving service and e-commerce one.

Improvements are still necessary. The educational authorities, having overcome the problem of quantity, should now strive to improve the quality, especially in the primary and secondary sectors. At these levels too much attention is devoted to examinations and certification, and not enough to experiential, hands-on, problem-solving, independent learning. Even at the pretertiary and tertiary level, people still need to appreciate the fact that due to globalization the Maltese economy is changing radically. Too many parents, and school career counselors, still push students toward the traditional professions, ignoring the far wider, and more rewarding, career opportunities in such new areas as financial services, special-interest tourism, and ICT. Many Maltese have still to grasp the fact that the right measure of a liberal education combined with new and refined forms of vocational training can improve their quality of life by relying more on their brains, less on their brawn.

Further Reading

- Clews, S. (ed.) (2006). *Malta Yearbook 2006*. Cospicua: De La Salle Bros. Publication.
- Malta, Ministry of Education (2000). *National Minimum Curriculum – Creating a Future Together*. Floriana: MEC.
- Malta, Ministry of Education, Ministerial Committee on Inclusive Education (2000). *Inclusive Education: Policy Regarding Students with Disability*. Floriana: MEC.
- Malta, Ministry of Education, National Curriculum Council (2002). *Guidelines for the Implementation of the National Curriculum on Inclusive Education*. Floriana: MEC.
- Malta, Ministry of Education, Youth and Employment, Chalmers, R. (2004). *State Higher Education Funding*. Floriana: MEYE.
- Malta, Ministry of Education, Youth and Employment, Clark, M. (2005). *School Attendance Improvement*. Floriana: MEYE.
- Malta, Ministry of Education, Youth and Employment, Debono, M. (2007). *A Career Guidance Policy and Strategy for Compulsory Schooling in Malta*. Floriana: MEYE.
- Malta, Ministry of Education, Youth and Employment, Grima, G. (2005). *MATSEC: Strengthening a National Examinations System*. Floriana: MEYE.
- Malta, Ministry of Education, Youth and Employment, National Commission for Higher Education and Secretariat for Higher Education (2005). *The Future of Higher Education in Malta*. Floriana: MEYE.
- Malta, Ministry of Education, Youth and Employment, Policy Unit (2005). *For All Children to Succeed: A New Network Organization for Quality Education in Malta*. Floriana: MEYE.
- Malta, Ministry of Education, Youth and Employment, Sollars, V. (2006). *Early Childhood Education and Care: A National Policy*. Floriana: MEYE.
- Malta, Ministry of Education, Youth and Employment, Spiteri, L. (2005). *Inclusive and Special Education Review*. Floriana: MEYE.
- Malta, Ministry of Education, Youth and Employment, Zarha, J. F. X. (2007). *A Quality Assurance Framework for Further and Higher Education in Malta*. Valletta: NCfME.
- Malta, National Statistics Office (2007). *Official Statistics of Malta 2005-Education Statistics*. Valletta: Malta Government Press.
- University of Malta (2007). *University of Malta Calendar 2007–2008*. Msida: UOM Communications Office.
- University of Malta, Farrugia, C. (2001). *Strategic Plan 2001–51*. Msida: UOM Communications Office.

Relevant Websites

- <http://www.gov.mt> – Government of Malta portal.
- <http://www.maltadiocese.org.mt> – Malta Church Schools.
- <http://www.mut.org.mt> – Malta Union of Teachers.
- <http://www.education.edu.mt> – Ministry of Education, Youth and Employment.
- <http://www.um.edu.mt> – University of Malta.

Mongolia

B Yembuu, National Centre for Non Formal and Distance education, Ulaanbaatar, Mongolia

© 2010 Elsevier Ltd. All rights reserved.

General Background

Located between Russia and China, Mongolia is the world's second-largest landlocked and least densely populated country in the world. It spans a territory of 1565000 km², which is larger than the overall combined territory of Great Britain, France, Germany, and Italy. Mongolia's terrain is characterized by vast semidesert, grassy steppe, and mountains. Arable land constitutes less than 1% of the total area. The country is also well known for its extremely cold and dry continental climate.

In 2007, with a population of 2.6 million, the annual growth rate was about 1.4%. As of 2007, the median age is 23.7 years, and life expectancy is 64.52 years; about 33% and 75% of its people are under 14 and 35, respectively. Sixty-three percent of the population lives in cities, the rest live a seminomadic lifestyle.

Mongolia is a homogeneous nation populated mainly by Mongols, who consist of Khalkha and other groups, all distinguished primarily by dialects of the Mongol language. However, a small proportion (up to 5%) is represented by Kazakh and Tuvas. The official language is Mongolian.

The country is divided administratively into 21 *aimags* (provinces), including the capital city, Ulaanbaatar, which has independent administrative status. Further local subdivisions include *soums* (districts), and *bags* ((subdistricts). As of 2007, 1 million people or 38% of its population live in the capital city, which is growing rapidly.

The country uses a unicameral parliamentary system consisting of 76 seats and the president has a symbolic role.

The Human Development Report 2007 ranks Mongolia as 114th out of 177 countries, with its Human Development Index (HDI) at 0.700, and categorizes it as a Medium Human Development Country.

By the time the Mongolian nation came to the dramatic attention of the world in the early thirteenth century, it already had a long and trustworthy history. After declaration of the People's Government in 1921, the country followed Soviet leadership for nearly seven decades. In the period since 1991, Mongolia has been in the midst of transition, from a totalitarian regime with a centralized command economy, to a market-oriented economy. A new constitution was adopted in 1992.

Economic growth reached 8.4% in 2006, bringing Mongolia's gross domestic product (GDP) per capita to \$1037. The animal-husbandry sector continues to be the main source of income for 40% of the labor force, it

produces 21.7% of GDP, and grew by 14.4% in 2006; crop production increased by 63.4% the same year. Mongolia's export is heavily dependent on a few commodities (minerals, cashmere, and textiles) that constitute 80% of total exports and international market prices of which are highly volatile and unpredictable.

The second major contributor to the economy is the mining sector; it continues to benefit from high international gold and copper prices. The tertiary sector is dominated by wholesale and retail trade, as well as transport and communication.

Mongolia's relatively good education system during the socialist period included free and widespread access, good quality, impressive levels of attainment, a policy on nondiscrimination, and measures to deliver education services to the nomadic population through boarding schools. Mongolia claims one of the highest literacy rates (97.8%) in the world (the percentage of male literacy was 98.7% and female literacy was 99.2%). By the mid-twentieth century, literacy rates throughout Mongolia increased drastically with the shift from traditional Mongolian script to Cyrillic.

The national education system consists of a complex set of successive education programs, including formal schooling and a nonformal education (NFE) sector. Mongolia's Education Development Index (EDI) is 0.952 as of 2007. A recent Human Development Report (HDR) revealed that 80% of medical doctors, 70% of lawyers, and 73% of teachers in Mongolia are women.

Goals of the Education System

The Mongolian Constitution (1992) guarantees its people general education free of charge. The fundamental assumptions stated in the Education Law reflect the principle of equality in education: every citizen has equal rights to his or her education regardless of race, ethnicity, nationality, sex, religion, disability, social status, and economic condition. Furthermore, "education shall be humanistic and democratic, universally available and continuing." These provisions state that Mongolia must gear toward a public-education system grounded upon equality in educational opportunities.

The Education Law and sub-sector-specific laws, including the Pre-school Education Law, the Primary and Secondary Education Law, Vocational Education Law, and Higher Education Law, have been enacted. The recent amendments

(2006) to the Education Law of Mongolia provide a legal framework for school structural reform (from 10- to 12-year system); upgrading teachers' social status through the rationalized pay and incentive system and better protection of the rights of children.

The amendments require education settings to be free of discrimination and any kind of emotional, physical punishment and economic pressure on students; education curricula and contents to be gender sensitive. Furthermore, it obliges local governments and education authorities to be responsible for the provision of education for all children in their respective areas, and for state education budget allocations to be sensitive to geographic location and particular specifics of schools, and specific learning needs of children with disabilities. Primary and secondary education delivery through a nonformal, equivalency training program is to receive funding from the state.

As per the Education Sector Master Plan (2006–15), the objectives of the Mongolian education system are: to increase the preschool gross enrolment up to 99%, reduce disparities to provide quality education for students and also provide support to them to enjoy rights to study,

provide continuous educational services in conformity with study needs and the lives of students, and improve accessibility to nonformal and adult educational services.

The goal of these schools is to provide appropriate education for pupils with special abilities in certain fields. The strategic goals of preschool education are directed at improving both accessibility and quality – particularly for poor, rural/remote children, ethnic minority, and children with disabilities, as well as policy regulation and coordination on early childhood care and development (ECCD). Improving efforts of family and community involvement were noted as significant factors in accomplishing these goals, but specific ways to resolve or improve and how to apply family and community involvement, is not yet clear in government policy.

Structure and Operation of the Education System

The Mongolian education system consists of preschool education, primary education, secondary education, technical and vocational education and training, and higher education and a broad range of NFE activities devoted to the various target groups of the population.

The new structural reform of the education system moved from a 4-4-2 system to a 5-4-2 and then to a 6-3-3 in line with international practices. Beginning with the 2004–05 school year, schools began to serve 7-year olds, and from the 2008–09 school year, the school entry age is being reduced to 6.

Preschool and Early Childhood Care

Preschool education is not compulsory in Mongolia. As of 2007–08, there are 768 kindergartens (666 of them are public), with 130758 children nationwide (see **Table 1**). Only 54.1% of preschool-aged children are enrolled in kindergarten (was 30.0% in 2000). Of these, 41.8% are enrolled in formal kindergarten and another 12.3% in nonformal or alternative forms of preschool education delivery (shift classes, *ger* and mobile kindergartens, as well as mobile teachers). Herders' children count for only 15% of the total preschool population. Such disparities cause unequal development opportunities for youngsters and consequent effects on basic educational attainment at later stages. Private-sector participation is invited and the private-service coverage is planned to reach 10% of the total preschool children.

Primary (Elementary) and Secondary Education

General education in Mongolia is 12 years (6 years, primary; 3 years, lower secondary; and 3 years, upper secondary) which has started from the academic year 2008–09. Basic education (6+3) is compulsory and provided by the state free

Table 1 Educational providers, students, and teaching staff, 2007–08

<i>Provider</i>	<i>Institutions</i>	<i>Students</i>	<i>Teaching staff</i>
<i>Preschool education</i>	768	130 758	3446
Public	666		
Private	102		
<i>Schools</i>	754	537 546	23 897
Government	597		
Private	152		
Foreign investment	5		
Primary		239 663	7694
Lower secondary		212 243	10 873
Upper secondary		85 640	5330
Males		264 275	4435
Females		273 271	19 462
VET	56	29 906	1451 ^a
Government	23	14 629	
Private	10	1643	
State colleges	9	7861	
Private colleges	4	2581	
State higher education institutions	8	2809	
Secondary school based	2	383	
Males			906
Females			545
<i>Higher education</i>	165	142 411	6818
Private	117	48 552	
Government	48	93 478	
International branches	6	381	
Males		56 228	2913
Females		86 183	3905

^aOnly for independent VET.

Source: *Statistical Year Book of MECS*, 2008.

of charge. Schools for the primary, lower-secondary, and upper-secondary levels, generally do not exist separately and are combined in one school campus. There are only 79 schools offering just primary education in Mongolia (mostly in remote rural areas).

In the 2007–08 scholastic year, in all, 537546 students were studying at 754 schools (public 597, private 152, and international 5) nationwide. The gross enrollment rate of children aged between 7 and 15 years is 95.1%. The dropout rate is claimed at below 2%, which reached its highest point, 8.8%, in 1992–93.

Lower-secondary education is the final stage of compulsory schooling and lasts 3 years (ages 12–15), followed by 3 years of upper-secondary education (ages 16–18). Graduates from grades 9 through 12 are eligible to enter technical and vocational training schools. After graduating, the upper-secondary school (not compulsory) students can enter higher education institutes. In the rural areas, children of nomads are often sent to the *soum* (county) or *aimag* (province) centers to attend boarding schools; they return home only during breaks and summer vacation. The primary school day normally contains about 5 h of tuition, and comprises around 200 school days per year.

Other Schooling Arrangements

Students with special needs: there exist countrywide five schools for students with severe physical or mental disabilities, all located in Ulaanbaatar. Officially, students with special needs are supposed to be integrated into regular schools. According to the National Policy on Inclusive Education approved by the Government in 2006, the goodwill of individual schools is required to systematically integrate students with special needs into schools. The more common forms of schooling for children who have special needs are home schooling by parents and, if private financial resources are available, private tutoring by teachers.

There also exist numerous classes at lower-and/or upper-secondary levels for students who are especially gifted. Majority are classes within a regular school, but in greater agglomerations, there also exist separate schools for gifted students. They provide additional courses in particular subjects for a fee that is retrieved from parents. The most popular subjects are the natural sciences. High-stakes entry exams are used to ensure selective admission.

Technical and Further Education

The vocational education and skills training has been one of the most neglected areas of education in Mongolia. After the initial transition shock in the first half of the 1990s, the number of students in technical and vocational education and training (TVET) sharply declined from 31000 in 1989 to 8000 in 1995. The total number of

students enrolled in technical and vocational schools was 29 906 by year 2008.

The TEVT is provided by professional training and production centers and private-sector participation is increasing. In addition, some branches of colleges and universities also provide TEVT. These schools offer secondary vocational education programs to train skilled workers and technicians.

Higher Education

Higher education in Mongolia is provided by universities, colleges, and institutes. Colleges offer mainly undergraduate programs, while universities focus more on research and graduate study. The higher education sector has expanded rapidly since 1990 mainly due to new private-sector players in the market. As of the 2008–09 academic year, there are 165 higher education institutions, including 48 state- and 117 private-run institutions as well as 6 branches of foreign institutions, and 74% of all higher education institutions are based in the capital city and only 49% are accredited. There are 142 411 students in higher educational institutions, of which 48 552 students study in private institutions. Public institutions of higher education are nonprofit organizations, while private institutions may be either nonprofit or for-profit.

Formerly, higher education institutions used to mainly offer undergraduate programs, leading to the award of a higher education diploma with the title of specialist. In 1992, a more Western model, consisting of BA, MA, and PhD degrees, was introduced. A student must complete a minimum of 120 credit hours to earn a bachelor's degree which requires 3–5 years of full-time study. Professional degrees in dentistry, pharmacy, and veterinary medicine require 5 years, and degrees in medicine are conferred after 6 years. The master's degree requires a total of at least 150 credit hours, which is awarded after 11/2–2 years of study beyond the bachelor's degree. The Doctor of Philosophy requires several years of advanced study beyond the master's degree, in addition to a dissertation and public defense. A total of not less than 210 credit hours is needed for doing a PhD. The doctor of science degree (similar to the German habilitation doctorate) is awarded as an advanced degree, requiring 21/2–3 years of study beyond the PhD.

Some colleges (nonuniversity higher education) offer vocational training, which lasts between 2 and 4 years, and leads to a higher education diploma. Women currently account for more than 63% of university students, with 65% of them earning master's degrees.

Adult and Nonformal Education

The NFE sector caters to two groups. The primary target group consists of those with no or limited formal schooling, that is, school-aged children and youth who are out of

school, as well as adults who never enrolled in or dropped out of schools through the Literacy and Equivalency Programme. The second target group is more broadly defined and includes literate adolescents and adults who take courses in life skills, (e.g., health, market economy, legal education, ecology, and small-business skills), vocational skills, creative skills, postliteracy programs for adults on life skills, income-generating activities, etc. The establishment of the National Non-formal and Distance Education Centre is a major milestone, a nationwide structure to deliver NFE with designated staff. The NFE learning centers – Enlightenment – play a crucial role in providing NFE service. As of 2008, 492 NFE facilitators and teachers worked in 352 NFE learning centers across the country, at least one in each *soum* or district.

The Teaching Profession

The advantage of the education system at all levels is that most of the teachers are professionally prepared in universities, higher schools, and colleges. Professionals who graduate from universities and colleges with a teacher's diploma have the rights to teach. In primary and secondary schools, 98.5% of teachers are professionals with the relevant educational qualification. There are 23897 teachers (7694 for primary, 10873 for lower secondary, and 5330 for upper secondary) working in all schools and 1451 teachers in TEVT. On average, the pupil–teacher ratio (PTR) in primary education in rural areas is 31:1 compared to 34:1 in the capital. However, there are rural schools with PTR as low as 11:1 and urban schools with PTR as high as 77:1 due to the ongoing trends of in-country migration. Woman teachers comprise 99.8% of staff in primary education and 81.3% in secondary education.

Primary school teachers are trained at teacher training colleges. Secondary school teachers are mostly trained at the State University of Education. Graduating students are awarded a Bachelor's degree in education. Also, some private institutions, such as the Gurvan-Erdene Teacher Training Institute, train school teachers.

Administrative and Supervisory Structure

The central education authority in Mongolia is the Ministry of Education, Culture and Science (MECS) which is the central administrative body that formulates the national educational policy and sets the standards for each level of education. The MECS also administers general provisions on education matters, teacher training, curriculum development, and state examination procedures, and is responsible for the accreditation of higher education institutions. The administrative fields of the ministry include not only preschool, primary, secondary,

vocational, higher education, and educational research, but also cultural and scientific affairs, and NFE.

In addition, general provisions on educational matters, such as scheduling of the school year, preparation and publication of textbooks for general secondary education, and state examination procedures, are administered by the ministry. Furthermore, the ministry ensures the smooth implementation of the elaborate contents and standards, and is responsible for providing licenses to establish higher educational institutions, and setting general typical provisions for teaching and research.

Education and culture departments within the local governments at the provincial level serve as the local educational authority. These departments are responsible for the administration and management of education services relating to formal education and NFE, and in-service teacher training. There are some state educational/professional institutions, including the Higher Education Quality Accreditation Centre, Education Assessment Centre, State Treasury Fund for Student Loans and Grants, and Institute of Education, which operate with some degree of autonomy.

Educational Finance

The education sector has always been the priority sector in national development strategy and resource allocation even in the most difficult years of the transition period. The Education Law stipulates that at least 20% of the government budget is to be allocated to education, and since 2000, the actual allocation ranged between 18% and 20%. In 2007, the education budget increased by 29.1% (8.3% of GDP). The Law permits education delivery through both for-profit and not-for-profit institutions. However, more than 70% of the expenditure goes for teachers' salary and heating costs. High spending does not contribute enough to quality.

As of 2007, shares of the subsectors of the education budget were as follows: early childhood care and education (ECCE), 16.8%; primary and general secondary education, 48.3%; VTE, 4.7%; higher education, 14.4%; and other areas, including teacher training, higher education student grants and loans, NFE, and adult education, were at 15.9%.

Schools in Mongolia plan their budget based on a funding formula that distinguishes between variable and fixed costs. Variable spending is proportional to projected student enrollment and covers teacher remuneration items – salaries, supplements, bonuses, insurances, taxes, and pension plans – and other expenses such as stationery, books, periodicals, postage, and communication costs. Fixed costs cover items, such as heating, water, electricity, and sewage, and are estimated from past expenses.

Based on estimations of student numbers and inflation rates, MECS and the Ministry of Finance propose an annual per-student budget every fiscal year, which is called the

normative means that apply to the variable cost component of the education budget. Normative means differ regionally and the government approves them every year. Higher education was fully subsidized by the state until 1993, when fees for students were introduced for the first time. Financing of public higher education institutions comprises three main parts: state budget, tuition fee, and other sources of funding. However, the government continues to provide financial assistance in the form of grants and loans to students from low-income families and to those who demonstrate outstanding achievement.

According to this policy, students from low-income families, disabled families, families of herdsmen with less than 200 heads of livestock, and students from partially orphan families, have access to loans; orphans and disabled students are eligible for tuition-fee grants; and students who have outstanding achievements to show, are exempted from tuition-fee payment.

Performance Monitoring, Evaluation, and Research

Promotion to each educational level is through a system of exams. At the end of primary, lower-secondary, and upper-secondary education, students are required to take state examinations given to pupils who are on the verge of completing each level. Primary education examinations are given in Mongolian language and mathematics. Admission to both university and nonuniversity programs require the Gerchilgee Diploma, awarded at the end of secondary school. Students must also take a competitive entrance examination administered by all institutions of higher education. The examination is held once a year, usually at the end of June and early July.

In 1998, the National Council for Higher Education Accreditation (NCHEA) was established, and since then, all higher education institutions are required to undergo accreditation. Only those institutions that have successfully gone through the accreditation process are eligible to receive government financial support. Likewise, only students enrolled in accredited institutions are eligible for government grants and loans. The private institutions are approved by the MECS, which also sets standards for private higher education. Students enrolled in accredited private higher education institutions are eligible for various aid programs granted by the government.

Major Changes and Issues since the 1990s

With a very strong education system until 1989, Mongolia's educational achievements were seriously undermined following an economic and financial crisis in the 1990s that resulted after the political and economic transition from the

socialist system to a free-market economy. The education system saw a severe erosion in both quality and quantity with falling enrollments.

Innovations in the educational sector, which have been implemented according to a package of laws on education, created a legal basis to adjust the education system and content to those of highly developed countries, enhanced the quality of education at all school levels, upgraded equipment and facilities, and raised social guarantees for students.

The school curricula and contents have been revised and the number of students at all levels of education has also increased. Competence-oriented new Education Standards started from 2004. The Government decided to extend the school system to a 12-year system in 2008, which will align Mongolia's education system with international standards.

Hence, the higher education sector has been transformed from a single, state-run multipurpose university into a decentralized group of specialized universities, all having the freedom to appoint their own instructors, set their admission policies, and confer degrees.

However, there is an imbalance between supply and market demand. Every year, approximately, 14 000 graduates from tertiary institutions join the labor market. Oversupply of specialists, such as medical doctors, economists, business managers, lawyers, accountants, and teachers, appears to be a serious problem, especially when it is combined with inadequate skills of graduates who have demanding specializations.

Many students go abroad to pursue higher education studies, especially to developed countries such as United States, Germany, Japan, and South Korea, instead of the former Soviet Union and Eastern Europe before the 1990s. There has also been a recent shift away from the study of Russian to include more English, Japanese, and Chinese.

Although there were some efforts and progress in promoting student-centered teaching/learning, the practical application is uneven and still limited. The use of information and communication technology (ICT) in education and computerization is increasing.

The key issues in the current Mongolian education system are:

1. poor quality and relevance of the current curriculum for primary and secondary education,
2. inadequate teaching qualifications and skills,
3. lack of consistent education standards,
4. weak learning and teaching environments in schools, and
5. supply-driven irrelevant vocational education for the youth.

To conclude, the Mongolian education sector is at a very important, indeed critical stage of its development with its medium-term development strategy to achieve

the Education for All (EFA) goals and Millennium Development Goal (MDG) education targets of the country. The main challenges ahead include significant improvements in the access, quality, and relevance of education, as well as improving the effectiveness and efficiency of the education system through the successful application of rights-based strategic planning, results-based management, and delivery of education, taking best advantage of the government's genuine commitment to education and the continued support of donors and development partners.

Further Reading

Education for All Global Monitoring Report (2008). Mongolia country case study. *Open Society Forum*. Ulaanbaatar: Mongolia.

Education Statistics (2007). *MECS Yearbook*. Ulaanbaatar: Mongolia.

Non Formal Education Sector Analyses. National Centre for Non Formal and Distance Education. Ulaanbaatar: Mongolia.

UNESCO National Education Support Strategy (*UNESS*) (2008). Ulaanbaatar: Mongolia.

UNDP (2008). *Mongolian Human Development Report*. Ulaanbaatar: Mongolia.

<http://mirror.undp.org/Mongolia/publications/UNDP-Brochure-Eng-FINAL.pdf> (accessed July 2009).

<http://www.welcome2mongolia.com/2007/10/primary-and-secondary-education/>.

Relevant Websites

<http://www.cmuc.edu.mn> – Consortium of Mongolian Universities and Colleges.

<http://www.country-studies.com> – Country Studies, Mongolia Education.

<http://education.stateuniversity.com> – Education Encyclopaedia, StateUniversity.com, Mongolia.

<http://www.britannica.com> – Encyclopaedia Britannica, Mongolia.

<http://aduuchin.tripod.com> – God Bless the Mongolia, Mongolian Education.

<http://www.unesco.org> – International Association of Universities, Mongolia's System of Education.

<http://www.ibe.unesco.org> – International Bureau of Education (UNESCO).

<http://www.wes.org> – International Grade Conversion Guide for Higher Education, World Education Services.

<http://www.policy.hu> – IPF (International Policy Fellowships), Financing Public Higher Education Institutions in Mongolia.

<http://www.mcc.gov> – MCC (Millennium Challenge Corporation), Mongolia.

<http://www.mecs.pmis.gov.mn> – Ministry of Education, Culture and Science.

<http://www.mtu.edu.mn> – Mongolian University of Science and Technology.

<http://www.mongolinternet.com> – Mongol Internet, Mongolian Education, and Universities.

<http://www.accmon.mn> – National Council for Higher Education Accreditation – NCHEA.

<http://nso.mn> – National Statistical Office.

<http://www.num.edu.mn> – National University of Mongolia.

<http://www.wes.org> – WENR (World Education Services): Education in Mongolia.

<http://www.wes.org> – WES Grade Conversion Guide, Mongolia.

Namibia

J Ellis, Ministry of Education, Windhoek, Republic of Namibia

© 2010 Elsevier Ltd. All rights reserved.

General Background

The political and educational history of Namibia has made education a national preoccupation. Education policy has been intertwined with colonial rule, first by Germany from 1885 to 1915, and then by South Africa, until independence in 1990. Since both colonial powers sought to limit the education of the indigenous black population, striving for education became part of the national liberation struggle, going back to the beginning of the twentieth century when missions were the main providers of education.

African cultures, of course, include traditional forms of education, some of which still exist today, but these are not the subject of this article, for want of space.

Namibia is a vast and arid southern African country of some 825 000 km². Currently, the total population is about 2 million, with settlement being denser in the northern regions because of rainfall patterns. The people are culturally and linguistically diverse, although the Ovambo groups form about 50% of the population, while less than 10% are of European extraction. The education system currently uses 13 languages, and English is the official language, for want of a better choice.

There are perennial rivers only on the southern border with South Africa and the northern borders with Angola, Zambia, and Botswana, the latter lying to the east of Namibia. The westerly Namib Desert, on the Atlantic coast, gave the country its name, partly as this ancient and inhospitable desert kept colonizers at bay until the nineteenth century. With the Namib Desert in the west and the Kalahari Desert in the east, Namibia has been called the land between two deserts.

Although short of water, the country has rich mineral resources, including diamonds, uranium, copper, silver, zinc, gold, and others, which make Namibia a lower-middle-income country, albeit with perhaps the highest inequality in the world. Namibia has significant fisheries, although there was overexploitation under South African rule, also exports red meat, and tourism is of growing economic importance. About 80% of manufactured goods are still imported from South Africa.

The independence constitution of 1990, in Article 20, guarantees the right to education, compulsory and free primary education, and the right to establish private schools at own cost. This has been reinforced through

legislation, notably the Education Act of 2000, and other legislations dealing with open and distance learning, higher education, and vocational education and training (VET). Government expenditure on education has varied between 20% and 25% of the national budget since independence.

Goals of the System

The goals of the Namibian education system were defined in the influential 1993 policy document *Toward Education for All* as access, equity, quality (including lifelong learning), democracy, and using resources well.

Structure and Operation of the System

This section also includes the characteristics and statistics for levels and types of education and training.

Namibia has a ten-level National Qualifications Framework administered by the Namibia Qualifications Authority (NQA) (see diagram).

The national system for schools covers 12 years. Children must be 6 years old to enter grade 1. The school year is from January to December and the three terms provide for 195–200 school days.

The lower primary phase consists of grades 1–4, upper primary grades 5–7, junior secondary grades 8–10, and senior secondary grades 11 and 12. In the first 3 years of schooling, local languages are the medium of instruction. English is a subject from grade 1 and becomes the medium of instruction from grade 4 onwards.

Soon after independence, Namibia decided to jettison the South African curriculum. The University of Cambridge provided the school leaving examination at grade 12. By 2006 this had been localized, with Cambridge still certifying the standard. A national examination is also conducted at junior secondary (grade 10) level.

About 45% of grade-10 learners in the formal school system cannot currently be accommodated in grade-11 classes. However, the Namibian College of Open Learning (NAMCOL) caters to most of those wishing to improve their grade-10 marks in some subjects (and thus gain readmission to the formal stream) or pursue grade-12 subject examinations.

Grade 10 also provides the usual entry point for vocational education, which is now being reorganized with

heavy involvement of the private sector, around occupations and competencies. Access to vocational education without grade 10 is also possible through a system of Community Skills Development Centres (COSDECs). Specialized vocational training has been taken out of schools and replaced by generic subjects such as entrepreneurship and design and technology.

In 2008, the government introduced a 1-year pre-primary year in 100 disadvantaged communities. This will be gradually expanded in future years. The 2001 Census suggested that about one-third of preschool children were attending early childhood centers organized privately or at community level.

Participants in adult literacy programs are able to complete a certificate equivalent to primary education and to continue their studies through COSDECs or NAMCOL. Annually about 25 000 adult learners enrol in the National Literacy Programme, while NAMCOL attracts 28 000 learners, making it the largest educational institution in the country. About 3000 trainees make use of vocational training opportunities.

At tertiary level, Namibia has the University of Namibia (UNAM), the Polytechnic of Namibia, and four colleges of education. Many students also study at a distance, particularly with the University of South Africa (UNISA). Namibia probably has about 30 000 students at tertiary level.

The total number of learners in formal schools was 570 623 in 2007, of whom 50.8% were female. There were 20 333 teachers, of whom 62.1% were female. There were 1661 schools, of which 99 were private. The average class group size was 30.9 learners (30.7 at primary and 32.0 at secondary). The learner–teacher ratio, however, was 28.1 on average (29.9 in primary and 24.6 in secondary). Enrolments in grade 1 were 68 861 learners (33 202 females), grade 5 had an enrolment of 64 552 (31 603 females), grade 8 had 50 575 (26 875 females), and grade 11 had 16 977 (9177 females). The average repetition rate in 2007 was 17.9%, a problem which has slowly been getting worse over the years. The survival rate to grade 5 in 2006 was 94% (95% for females), giving a good indicator of the number entering school who will have acquired stable literacy skills. The survival rate was 81% (82% for females) and 39% (40% for females) for grades 8 and 11, respectively.

In 2007, there were 19 290 classrooms in the system, of which 16 518 (85.6%) were permanent structures, 1171 prefabricated, 1420 traditional, and 181 hired. (There were 16 970 classrooms in 2001.) However, only 76% of schools had adequate toilets, while 80% had their own water supply.

Teaching Profession

One of the main changes in education since independence has been the attention paid to teacher education. In 1993

(3 years after independence), only 12.6% (13.8% for females) of teachers were qualified to teach at primary level, and 40.7% (47% for females) were qualified to teach at secondary level. Overall, the percentage of qualified primary teachers grew to 41.1% in 2001 and to 71.0% in 2007, while at secondary level it grew to 73.2% in 2001 and to 90.3% in 2007.

The current minimum requirement for appointment as a teacher is the Basic Education Teacher's Diploma (BETD), a 3-year qualification taught at the four colleges of education. It has also been provided at a distance by both the National Institute for Educational Development (NIED) and UNAM, although the distance mode is being phased out as demand is declining due to the high level of qualified teachers that has been reached and the limited number of annual vacancies. Equivalent qualifications are also accepted, subject to the assessment of the NQA, and many take this route through South African institutions and their local (private) partners.

Teachers at senior secondary level should have a 4-year degree in education, provided by UNAM, or another degree and a postgraduate diploma in education.

In 2007, 76.6% of teachers had more than 2 years of tertiary education, and most were between 20 and 49 years of age. Of the 13 675 primary teachers, 69.6% had more than 2 years of tertiary education. At secondary level, 91.2% of teachers had more than 2 years of tertiary education; 5.6% of teachers with more than 2 years of tertiary education transferred to another school in 2007. Those without such training were more inclined to stay put. Between 2001 and 2007, the number of teachers has grown on average by 1.9% per year, while the number of teachers with at least 3 years of tertiary education has grown by an annual average of 3.2%.

In 2006, the National Standards for the Teaching Profession were developed under the NQA. They are currently being used to reform preservice and in-service teacher education, and may in future result in the introduction of system of teacher licensing. The Namibia National Teachers Union (NANTU) has accepted the national standards but expresses its opposition to teacher licensing. Namibian teachers are regarded as public servants, but a committee of the Public Service Commission addresses the particular interests of the teaching profession.

One of the big controversies in Namibian education is about why this improvement in teacher education has not resulted in a dramatic improvement in the quality of results achieved by learners. While considerable efforts have been made to inculcate a learner-centered approach to teaching, doubts have been expressed about the subject knowledge of teachers. Preparations are currently being made for programs to ensure that all teachers achieve proficiency in English and mathematics.

Attention is now being given to the development of standards for school managers and the training to go with it.

Although teachers are reportedly not held in high regard as was once the case (when few other professions were accessible to black Namibians), attrition rates do not seem to be alarmingly high, at about 6% per annum. As part of a comprehensive strategy on human immunodeficiency virus (HIV) and acquired immune deficiency syndrome (AIDS), the Ministry of Education is currently implementing a workplace wellness program for all staff to limit the depredations of the pandemic. Members of the public service medical aid scheme have access to antiretroviral (ARV) treatment when needed.

Administrative and Supervisory Structures

At independence, Namibia inherited a number of separate and unequal education departments, for whites, coloreds (or people of mixed race), and various African ethnic groups, each with its own structures and even different curricula. Fortunately, the incoming government had the authority and determination to rapidly establish one integrated system. Drawing on its experience of setting up schools in exile, the South West Africa People's Organization (SWAPO) government changed secondary education first. The National Institute for Educational Development was created to work on curricula.

Initially, seven regional education offices were established, and these were later increased to 13 in accordance with the administrative regions that were determined for the country in the meantime. In 2008, the education system entered a decentralized administration with most functions, staff, and funds being decentralized to the regional councils. The Head Office will (in general terms) remain responsible for policy, curricula, standards, and monitoring and evaluation.

At the subregional level, Namibian schools are organized into inspection circuits and at the next level clusters of five to eight schools.

A national inspectorate administers the National Standards for Schools. These standards (introduced in 2005) have indicators, which have been described in some detail and four levels determined for each. Each school does an internal self-evaluation each year, and the national inspectorate arranges for a limited number of schools (about 30 at present) to be evaluated and rated by a national evaluation team. A school's evaluation in terms of the national standards should result in the drafting of a strategic plan for the development of the whole school.

A subject advisory service provides support for lower primary education and for the various subjects. Its deployment is complicated by the large number of subjects on the curriculum and limited human resources.

In a country as big as Namibia, communication and transport are a constant headache, not to say expense.

Namibia has had several ministerial structures for education since independence. A separate ministry for higher education was instituted in 1995(?) but in 2005(?) the Ministry of Education was again put in place.

Educational Financing

Almost all funds for education come from the central government. In 2008–09, the budget for the Ministry of Education is N\$4.78 billion (about US\$620 million) out of a total budget of N\$22.46 billion (about US\$2.9 billion.) Education thus currently makes up about 21% of total government expenditure. About 60% of all education expenditure is taken up by personnel costs.

Although no fees are charged by government for primary education, learners do pay examination fees at secondary level, and subsidized fees are also charged by NAMCOL and public higher education bodies. The law allows school boards to run a school development fund with voluntary contributions from parents. Although contributions to the school development fund are not compulsory, and are theoretically limited, informal pressures are often brought to bear on parents and learners to ensure payment. At a rough estimate, it seems that private contributions make up about 10% of expenditure on education, and are one of the complicating factors in achieving greater equity.

International aid to the Ministry of Education in 2008–09 is estimated at about N\$100 million.

Performance Monitoring, Evaluation, and Research

The performance of the Namibian education system is currently being measured against targets in terms of the following key performance indicators:

1. number of children from disadvantaged communities who enter primary education having successfully completed 1 year of public pre-primary education;
2. enrolment in grade 11;
3. percentage of learners achieving D or better in mathematics, science, and English in grade 10;
4. percentage of learners achieving D or better in mathematics, science, and English in grade 12;
5. national average Southern and Eastern African Consortium on the Monitoring of Education Quality (SACMEQ) test score (in reading and mathematics);
6. percentage of learners receiving life-skills education;
7. total enrolment in vocational education;
8. average annual completion rates for vocational trainees;
9. employment rates of VET graduates within 1 year of graduation;
10. number of diploma, degree, and postgraduate graduates in key human resource categories; and
11. adult literacy rate.

Particular mention should be made here of the SACMEQ. This regional mechanism for testing primary school learners came as a nasty surprise to Namibia when Namibian learners were found to be at the bottom of the regional league table. Much debate has followed about what should be done. Besides continuing its participation in SACMEQ, Namibia has decided to set up additional national diagnostic tests at grades 5 and 7 within the next few years. It seems that while much attention has been given to secondary education, not enough effort has been put into primary education, where language is the main complicating factor for both learners and teachers.

Research is one of the main functions of NIED. Because of national and international interest in Namibian education, a fair amount of research has been done. It is, however, beyond the scope of this article to deal with this topic.

Major Changes and Issues Since the 1990s

Public concern about the quality of education, reinforced by the finding of SACMEQ and other studies, led to the establishment of the Presidential Commission on Education, Culture and Training in 1999. The Commission consulted widely with the Namibian public and with Namibian educators. They concluded as follows:

The first nine years of education in Namibia have been the years of access. The Commission believes that the next decade should be the decade of equity. . . . This will also be a strong influence on the quality and efficiency of the system. . . . The Commission believe that Namibia should aspire to being "A Learning Nation". . . . Because of globalization Namibia must compete in the knowledge-based production of goods and services, or face increasing marginalization.

Apparently concerned about the slow pace of improvement in education, and doubting that the education system could make its expected contribution to the equitable and knowledge-based society foreseen in Namibia's Vision 2030, the government in 2002 invited the World Bank to undertake a study of the Namibian education system. This was a controversial move as Namibia has no economic need to borrow from the World Bank.

The World Bank study concluded that:

. . . at present, the (education) system is too weak to effectively play its expected role. Key weaknesses pertain to poor quality or ineffectiveness; low efficiency; inequalities; low economic relevance; and low capacity for knowledge creation and innovation.

Further technical assistance from the World Bank, and extensive participation by Namibian educationalists, gave

rise to the Education and Training Sector Improvement Programme (ETSIP), first as a 15-year strategic plan (2005–20) and subsequently as an implementation plan for the first phase (2006–11).

The first phase of ETSIP focuses on strengthening the immediate supply of middle- to high-level skilled labor to meet labor-market demands and support overall national development goals. Second, ETSIP will strengthen the quality, effectiveness, and efficiency of the general education and training system. Third, it will strengthen the knowledge creation and innovation system. Fourth, it will improve the effectiveness, quality, efficiency, and development relevance of the tertiary education and training system. Fifth, it will strengthen the policy, legal, and institutional frameworks to support equitable access to high-quality and responsive adult learning.

ETSIP implementation plans have been developed in greater detail than any previous plans in Namibian education and have attracted widespread support from the government, the Namibian private sector, and international development partners. In fact, ETSIP consists of nine subprograms dealing with all aspects of education from early childhood development to tertiary education, and with cross-cutting programs for information and communication technologies (ICTs), HIV and AIDS, and capacity development. Taking its cue from the Presidential Commission of 1999, ETSIP bears the subtitle, 'Planning for a learning nation'.

Further Reading

- Ellis, J. (1984). *Education, Repression & Liberation: Namibia* (Also published in German with an introduction by Goldschmidt, D. and Melber, H. by World University Service Deutsches Komitee.) London: World University Service UK and Catholic Institute for International Relations.
- Ellis, J. (2004). Making space for adult education in independent Namibia. *Convergence* 37(3), 105–113.
- Government of the Republic of Namibia (1999). *Report of the Presidential Commission on Education, Culture and Training*. Windhoek: Gamsberg Macmillan.
- Marope, M. T. (2005). *Namibia Human Capital and Knowledge Development for Economic Growth with Equity*. Washington, DC: World Bank.
- Ministry of Education (2007). *Education Management Information Statistics*. Windhoek.
- Ministry of Education and Culture (1993). *Toward Education for All, A Development Brief for Education, Culture and Training*. Windhoek: Gamsberg Macmillan.

Relevant Websites

- <http://www.dnea.gov.na> – Directorate of National Examinations and Assessment.
- <http://www.etsip.na> – Education and Training Sector Improvement Programme (ETSIP).
- <http://www.hamu-nam.net> – HIV/AIDS Management Unit, Republic of Namibia.

<http://www.mec.gov.na> – Ministry of Education, Republic of Namibia.
<http://www.namcol.com.na> – Namibian College of Open Learning (NAMCOL).
<http://www.namqa.org> – Namibia Qualifications Authority.
<http://www.nied.edu.na> – National Institute for Education Development.
<http://www.nln.gov.na> – National Library of Namibia.

<http://www.nta.com.na>.
<http://www.polytechnic.edu.na> – Polytechnic of Namibia.
<http://www.sacmeq.org> – Southern and Eastern Africa Consortium for Monitoring Educational Quality.
<http://www.tech.na> – TECH/NA!, Namibia's ICTs in Education Initiative.
<http://www.unam.na> – University of Namibia (UNAM).

New Zealand

R Baker and C Wylie, New Zealand Council for Educational Research, Wellington, New Zealand

© 2010 Elsevier Ltd. All rights reserved.

Glossary

ERO – Education Review Office, the government department that inspects every early childhood education and school on a 3-yearly cycle.

Kōhanga Reo – Māori-medium early childhood education centers, supporting language and culture as well as early education.

Māori – The indigenous population of New Zealand.

NCEA – The National Certificate of Educational Achievement is the secondary qualification. It has three levels, with the first level usually taken in Year 11.

NZQA – The New Zealand Qualifications Authority is the government department that sets and reviews standards related to qualifications, administers the national secondary school examinations related to the NCEA, and approves courses.

Pacific – Students and people of Pacific Island origin.

Pakeha-European – New Zealander of European ethnic origin.

Te Whāriki – The early childhood education curriculum.

Treaty of Waitangi – The founding document of New Zealand.

Background

New Zealand is a small, mountainous country located in the South Pacific Ocean, 2000 km southeast of Australia. Its total land area of 269 000 m² makes it three-quarters the size of Japan. The population of just over 4 million people live mostly in the coastal cities located on the two main islands. While New Zealand has three official languages, English, Māori, and sign language, English predominates and few people are bilingual.

New Zealand is a democracy with constitutional ties to the United Kingdom. Its founding document is the Treaty of Waitangi, an agreement signed in 1840 by Māori (the indigenous people of New Zealand) and the British Crown. From the British perspective, it was the means by which they gained the right to govern and establish laws and in return the Crown was to guarantee and actively protect Māori tribal authority over their lands,

fisheries, forests, villages, treasures, and culture and extend to them the status and rights of British citizens. The significance of the Treaty of Waitangi has been actively debated in New Zealand over the past 30 years and in recent years the government has sought to address its earlier failure to keep to its treaty commitments.

Once a colony of Britain, by the middle of the twentieth century, New Zealand's agricultural products had made it a relatively wealthy country. However, changes, such as the establishment of the European Union (EU), led New Zealand to look at new ways to generate wealth. More recently, the aim has been to transform the national economy from a basic reliance on primary production to more added value and knowledge-based production. Consequently, a priority for government is investing in educational programs designed to build the kind of skills and knowledge that will be required for a knowledge economy.

Until recently, the country's economic indicators were positive with the labor-force-participation rates rising steadily through the 1990s, matched by a fall in the unemployment rates. There has been strong economic growth in recent years and while the Organization for Economic Cooperation and Development (OECD) economic survey of New Zealand 2007 rates the country as having one of the most flexible and resilient economies in the OECD, it highlights concerns with productivity, the need to maximize improvements in living standards, and to actively manage spending on health and superannuation as the population ages.

The other major change in recent years is New Zealand's ever-increasing ethnic diversity. Of the overall student population in 2004, 59% were identified as European, 21% as Māori, 8% as Pacific, and over 8% as Asian. Statistics New Zealand estimates that by 2021, the population of New Zealanders under the age of 15 will consist of 51% Pakeha-European, 23% Māori, 14% Pacific, and 12% Asian.

Education in New Zealand

The education system is based on several guiding principles including: access to quality early childhood services; legally free publicly provided primary and secondary education; equitable and affordable access to tertiary education; and quality-assured and portable education qualifications. The provision of flexible pathways for study is also an important feature: for example, students are not

streamed or channeled through particular types of school. Māori-medium education also supports the revitalization of the Māori language and culture. There are linked national curricula for early childhood education (ECE) and compulsory schooling which provide frameworks rather than close prescription. Assessment tools that allow comparison with national norms are widely used in schools, and will be used to assess students against national standards in literacy and numeracy from 2010.

The overall aim of education policy is to build a world-leading education system that equips all New Zealanders with the knowledge, skills, and values to be successful citizens in the twenty-first century.

A Devolved System

As part of wide-ranging economic- and public-sector reforms in the 1980s, New Zealand education moved from a relatively centralized structure in relation to resources and staffing, to one in which individual schools and tertiary institutions have considerable responsibility for their own governance and management, working within the framework of guidelines, requirements, and funding arrangements set by central government and administered through its agencies. The Ministry of Education

(MOE), established under the Education Act 1989, carries out functions such as providing policy advice to the Government; allocating funding and resources to schools and ECE providers; and monitoring the effectiveness of the education system as a whole. It does not have a direct supervisory relationship with individual schools, though it provides some support for schools deemed to be at risk. Local authorities have no role in education in New Zealand, nor are there any school districts. New Zealand has the most devolved system of school self-management in the Western world, with each school having its own board of trustees, largely parents elected by parents, in a governance role, and accountable through reporting of annual progress on self-set targets.

Separate education agencies have national responsibilities for quality and qualifications. The Education Review Office (ERO) provides 3-yearly reviews of early childhood centers and schools. The New Zealand Qualifications Authority (NZQA) sets and regularly reviews standards as they relate to qualifications, administers national school-based examinations, and approves courses. The New Zealand Teachers Council registers teachers, renews their practicing certificates, and approves teacher-education programs that lead to registration. All state and private schools and kindergartens employ only teachers with a current practicing certificate, or a limited authority

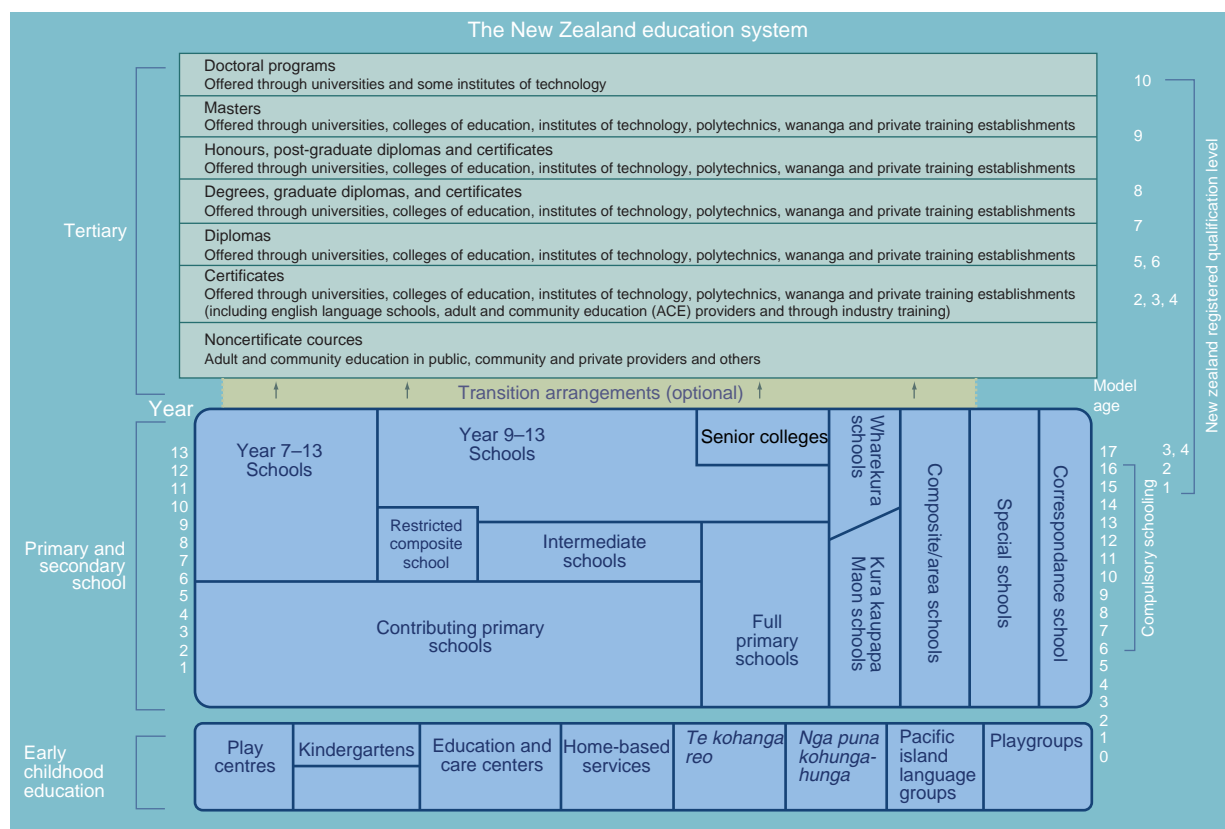


Figure 1 New Zealand education system overview. Courtesy of the International Division, Ministry of Education.

to teach. Registration is optional for teachers in other early childhood centers and in tertiary institutions.

While the Ministry has a strategic leadership role, and monitoring and evaluation roles with respect to the tertiary sector, it is the responsibility of the Tertiary Education Commission (TEC) to oversee implementation of the Tertiary Education Strategy and associated set of priorities. The TEC is responsible for funding all postcompulsory education and training offered by a diverse range of institutions (see **Figure 1**).

Early Childhood Education

New Zealand was the second country in the world after Iceland to reorganize ECE services by integrating responsibility for all of those services within the education system with childcare services moving from the social welfare to the education portfolio in 1986. ECE enrolments rose markedly over the 1990s, and in 2006, 94.5% of new school entrants had had some ECE experience. Since July 2007, 3- and 4-year-olds are entitled to up to 20 h free ECE a week.

ECE services are not state-owned, provided, or managed. They receive per-child government funding, and there are also family-income-related subsidies for some children. They operate within a framework of regulations, 3-yearly review, and an innovative bicultural ECE curriculum, *Te Whāriki*. The Ministry of Education supports ECE services to make the most of this curriculum by providing resources including assessment, planning, and self-review with associated professional development from contracted providers. Most ECE services are teacher-led; some are led by parents, often working alongside a teacher. The latter include playcentres, and *nga kōhanga reo*, providing Māori-medium ECE.

Primary (Elementary) and Secondary Education

As on 1 July 2006 760 761 students were enrolled in primary and secondary school, at which attendance is compulsory between the ages of 6 and 16. Students generally start school on their fifth birthday. Primary schools cater for students from age 5 (Year 1) to Year 8. Secondary schools usually provide for students from Year 9 until the end of Year 13. There are 2584 schools. Approximately 350 are secondary schools, and 46 are schools for students with special educational needs (though most such students attend regular schools). There are 105 private schools. Within the state school system, 10% are integrated schools which represent a range of special characteristics, including religious affiliation. There are also 68 *kura kaupapa Māori* (Māori-medium schools), and a

national correspondence school for students who cannot access a local school due to their remote location or other reasons, or who cannot access a particular subject at their local school. Students and their families can choose any state school, but a quarter of state schools have enrolment schemes, giving first priority to students within a set geographical zone. State schools ask for set donations, and though these are voluntary, they can be sizeable in some schools in high-income communities; and state-integrated schools charge fees to cover property costs that can also be substantial. Private schools receive some government funding.

The average primary school has just 216 students, reflecting the large number of small schools we have in rural communities. There are 49 241 teachers employed in state schools. In the primary schools, the average teacher:student classroom ratio is 1:23.9, and in secondary schools, 1:17.7.

Curriculum and Assessment

New Zealand has an outcome-focused national curriculum that sets the direction for student learning rather than prescribing specific content. In a high-trust/low-stakes accountability climate, schools are charged with developing a curriculum specific to the needs of their students, based on this framework. The curriculum specifies eight broad levels of achievement from Year 1–13. This is intended to bring about a seamless alignment between the school curriculum and *Te Whāriki*, across the years of school, and on into the various tertiary-learning pathways. The key competencies in **Figure 2** (thinking; using language symbols and texts; managing self; relating to others; participating and contributing) are woven into all eight learning areas and form a key curriculum link both to ECE and the tertiary sector.

Assessment up until the final 3 years of school also follows the high-trust/low-stakes policy model. There is an emphasis on formative assessment, and on ensuring all students are making progress relevant to their needs. Schools set their own performance targets and demonstrate their progress toward meeting these. Two main sets of nationally referenced tools are available to help them do this. Both sets of tools focus on literacy and numeracy, as emphasized in the curriculum. A range of other tools are also used at the schools' own discretion and a national monitoring program (NEMP) provides a process for sampling national progress across the wider curriculum, on a 3-year cycle. Internationally, New Zealand participates in the Programme for International Student Assessment (PISA), Progress in International Reading Literacy Study (PIRLS), Trends in International Mathematics and Science Study (TIMSS), and International Adult Literacy Survey (IALS).

The National Certificate of Educational Achievement (NCEA) is the official school-leaving qualification. This is

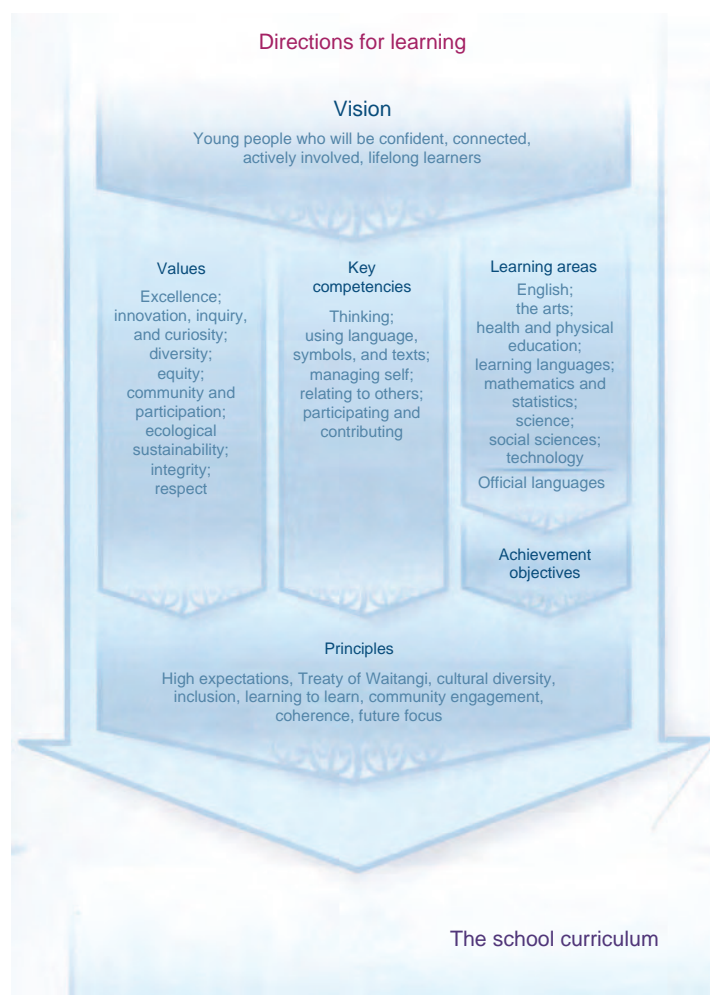


Figure 2 The New Zealand curriculum (2007). Published by Learning Media Ltd. Copyright © Crown 2007. All rights reserved.

awarded at three levels, broadly corresponding to the final 3 years of secondary school (years 11–13) and designed to integrate with other level 1–3 qualifications offered by tertiary providers in a seamless National Qualifications Framework (NQF) administered by NZQA. NCEA awards are earned by aggregating credits from achievement standards (which specify learning directly aligned with the school curriculum) and unit standards which are often more vocational/applied in nature and may also be offered by tertiary providers. In theory, both types of standards have parity of esteem. Credits from them can be combined in any balance to reach the total needed for an award. The modular nature of the standards provides for flexibility in curriculum design. Students wanting to enter university need to ensure their level-3 NCEA credits are aggregated in the correct proportions from an approved subjects list, and that they demonstrate at least a basic level of competence in literacy and numeracy. The most able students may also sit competitive scholarship examinations in individual subjects in their final school year. These provide financial rewards and also

help determine direct entry into elite university courses where limited places are offered. Open entry at the first year of university is more usual, although this is changing.

Tertiary Education

The term tertiary education in New Zealand is used to describe all aspects of postschool education and training. It includes over 900 institutions providing the full range of education from adult literacy and second chance education through to certificates, diplomas, bachelors, masters, and PhDs. It also covers industry training, apprenticeships and adult and community education. The diversity of provision is reflected in the many types of organizations offering tertiary education. The public tertiary education institutions include 8 universities, 20 institutes of technology and polytechnics, and 3 wananga (Māori tertiary education institutions). There are also 46 industry-training organizations, and approximately 895 private training establishments, which include private English-language

schools, registered by the NZQA. In 2006, 448 000 students were enrolled in tertiary education. In 2006, 58% of school-leavers went from school to tertiary education, 30% to certificate-level courses, 23% to degree-level courses, and 5% to diploma-level courses.

While the proportion of those participating in tertiary-level courses has increased markedly in recent years, the growth has been in certificate-level courses, not degree-level courses. There has also been growth in part-time enrolments, with students combining work and study.

Quality within the tertiary system is assured by a requirement that all organizations must be able to demonstrate that they have the policies and procedures relevant for their respective programs. These are then reviewed through processes of registration, approval, accreditation, and monitoring and audit.

Teaching Profession and Teacher Education

Most schools can attract qualified teachers, but in recent years, schools have expressed some concern over the shortage of experienced teachers. There are shortages in low-income and rural areas, and in some subjects in secondary schools. With an increased emphasis on professional development, often occurring during school hours, some schools are also experiencing difficulties finding experienced relief teachers. There are longstanding concerns about the workloads faced by school principals and middle managers, and signs of difficulty in filling these roles in a sizeable minority of schools.

Most primary student teachers undertake a 3-year undergraduate teaching degree, most early childhood teachers undertake a 3-year diploma or degree program, and most secondary teachers undertake a 1-year graduate diploma, following a subject-specific undergraduate degree.

Registration is compulsory for all primary and secondary teachers. Half of all ECE teachers are currently registered, and there are government plans to increase this proportion. Provisional registration requires graduation from an approved preservice program, plus a satisfactory police check, and verification that the graduate meets the teacher's council criteria for being of good character and fit to be a teacher.

All schools employing provisionally registered teachers are required to provide them with a 2-year advice-and-guidance program that enables them to meet the satisfactory teacher dimensions for fully registered teachers. Each school employing a provisionally registered teacher receives a 0.2 salary allowance for the first year and 0.1 for their second year of provisional registration.

Early childhood services employing provisionally registered teachers are entitled to a grant of \$3700 for each of the 2 years to support their induction. The recent

move to increase the registration of early childhood teachers has created issues for the sector in providing appropriate supervision.

Educational Finance

Government spending on all three educational levels has gradually increased, both as a percentage of gross domestic product (GDP) (5.6% in 2006–2007) and a percentage of total government expenses (18.6% in 2006–2007). Our spending is close to the OECD average once the country's relative wealth and demographic structure are taken into account. As New Zealand has a relatively young population, the proportion of government expenses used for education is at the top end of the OECD countries.

Thirteen percent of school funding comes from non-government sources such as parent donations, fundraising, community trusts, and investment. The funding that schools receive for their running costs has risen, but is under pressure with rising costs due to information and communication technology (ICT), the costs of administrative staff due to school self-management, increasing legal compliance costs, and rising expectations of what schools should offer. Teaching-staff capacity is separately calculated in relation to the number of students on the roll, with schools making their own hiring and firing decisions, but with the government meeting the actual costs of the staff appointed. Most resourcing for schools is on a per-student basis, including some additional funding for schools serving low-income populations, in rural areas, or providing Māori-medium education.

The rationale for government funding of tertiary education relies on two principles. The first is that there are both public and private benefits from tertiary education and so costs are shared. Course costs have increased since the 1990s. Student loans to finance tertiary education began in 1991; by 2007, they had been used by 21.7% of the population aged 15 and over. The second funding principle is that there are some areas of more strategic value to society, and these are more heavily subsidized. In practice, institutions are funded through an integrated funding framework. This has three components: funding for domestic students, funding for research, and targeted funding. The latter is designed to promote the overall tertiary education strategy and comprises supplementary grants, for example, to support the participation and achievement of Māori and Pacific students, and funds to support innovation, e-learning, and links with business.

Performance Monitoring, Evaluation and Research

In 2006, the Ministry of Education began an annual report, *State of Education in New Zealand*, which gives

an overview of the overall performance of the education system, providing useful information on trends over time in key indicators and comparisons with other countries. An annual report on the school sector has been made to Parliament every year since 1994. The Ministry of Education also contracts research organizations and universities to undertake evaluations of specific initiatives and policies. The ERO provides a number of evaluative reports each year giving a national picture drawn from analysis of its individual school reports of aspects of particular interest, for example, quality of teaching in a specific curriculum area, quality of school governance. Research into aspects of education is carried out by the New Zealand Council for Educational Research, which receives some annual government funding in order to ensure that there is a sustainable national independent body that can carry out and disseminate educational research, by the universities, and a number of consultant organizations and individuals.

Major Changes and Issues since the 1990s

Compulsory Schooling

During the 1990s, educational institutions were largely left to their own devices, with government agencies taking responsibility for policy, funding, and review, at a distance. This hands-off decade left a legacy of both distrust and sometimes overconfidence in school internal knowledge and skills (in a country that has prided itself on its do-it-yourself approach). But government agencies have taken more of a leadership role to improve the capability and capacity of educators and educational institutions, as it became clearer that there had been no overall improvement in student outcomes over the 1990s. The original school review model, for example, was modified to include advice – if schools ask for it. The ideology of the self-managing school, however, remains, making many aspects of schools sharing information about their operations and progress a sensitive area. Longstanding issues about principal workloads, stress levels, the focus on administration at the cost of leading learning (NZ principals spend more of their time on administration than do principals in most other countries), difficulties in attracting experienced principals to small, rural schools, or those in difficulty, the low level of required preparation for the principalship (a quarter of primary principals have come to their position from being a class teacher with no school management experience) and reports of inequitable or poor appointments made by school boards remain problematic in a system which takes self-managing schools literally, and which does not provide the kinds of connective relationships and processes that exist in the public-school systems of most other countries.

While there was a market element in New Zealand's shift to self-managing schools, it was not accompanied by the kind of performance-management system introduced in England, which made a similar shift at around the same time. Although the media compiles league tables of secondary schools based on their performance in the NCEA, the government does not want league tables for primary schools, when national standards are introduced in 2011. The levers that are being used to improve quality in education, and raise student achievement are centered around developing capability through well-anchored professional development, teaching resources, and assessments. These include national literacy and numeracy programs that have resulted in gains for student achievement, particularly those with initially low achievement. Other levers include funding for schools to work in clusters, for example, on using ICT for learning, and for innovations linked to the new curriculum. Programs for new principals and refresher courses for existing principals have been introduced in recent years, with increasingly high take-up.

The student-achievement issues faced by New Zealand are not new, but there is less acceptance now of the gap between Māori and Pacific student achievement levels and others, of the gap between students from poor homes and others, and the level of students leaving school without a meaningful qualification. While the country is generally in the second-highest set of performers on the international assessments we take part in, with a high proportion scoring at the highest levels, we also have more students performing at low levels than other countries with similar average scores.

Around 90% of students are estimated to stay at school until age 16, but only 71% to age 17. Around 25% leave with no qualification, and 40% leave school without the level-2 NCEA qualification that would provide access to many career pathways. In part, this is because New Zealand has had high employment, and it has been possible for young people without qualifications to find work.

As a consequence, the Ministry of Education has focused on three key areas that research and experience show can make the greatest different to raising student achievement and reducing disparity. These are effective teaching for all students; family and community engagement in education; and developing quality providers. The work program within these three areas includes a major investment in increasing the information and research base so that new policies and the associated development of new practices in teaching, preparation of teachers and principals, and professional development, are well supported by evidence. This work program includes a series of best-evidence syntheses that distil underlying principles and contextual effects from a synthesis of research that analyses the impact of practice on student outcomes. The most recent of these include the teaching of mathematics,

professional development, and school leadership. These best-evidence syntheses are developed through a collaborative knowledge-building approach across policy, research, and practice, with a commitment to knowledge utilization that has led to embedding opportunities for dialog for all key stakeholders within the knowledge-building processes. The national work program also includes research and development focused on researchers working with professional developers, educators, and resource providers on trials of new approaches.

At the same time that the government is implementing policies that support the continuous improvement of the current education system, there is a developing debate about what learning for the knowledge age might actually mean and how schools might better provide such opportunities.

Noncompulsory – ECE and Tertiary

Participation in ECE has increased since the 1990s, largely due to increased government funding. However, concerns remain about uneven quality. Compared to the 1990s, there is less reliance on competition and consumer choice to ensure that children receive good quality ECE, with some movement toward more regulation around gross indicators of quality (e.g., ensuring ECE staff have ECE qualifications), and as with the school sector, providing more resources and professional development to improve teaching capability.

In 2002, *Pathways to the Future – Ngā Huarahi Arataki*, the first long-term strategic plan for any education sector in New Zealand, was published after extensive consultation with the sector. Three goals provide the framework for the plan: increasing participation in quality ECE; improving the quality of ECE services; and promoting collaborative relationships.

During the 1990s, there was a huge expansion in the tertiary sector as it moved from an elite, largely university-based system, to a much more diverse and accessible system. There was an emphasis on market-like competition, student choice, and on the private returns to tertiary education. Associated with this was the requirement for students to pay more for their tertiary education and a student loan system was instigated. Since 2000, while the general direction was maintained, there has been more governmental steering in an attempt to align tertiary education more closely with the social and economic needs of the country.

During 2006, the government embarked on a program of sustainable reforms to the tertiary education system. Under the new funding system, the TEC will fund providers on the basis on investment plans agreed over a 3-year period, against the priorities outlined in the tertiary education strategy. This strategy has six themes: strengthening of the system capability and quality; contributing to achieving the Māori-development aspirations,

raising foundation skills to allow participation in the knowledge society; developing the skills needed for a knowledge society; educating for Pacific people's development and success; and strengthening the research knowledge creation and uptake function.

See also: Classroom Assessment in Policy Context (New Zealand); National Assessments; School Reform and Restructuring: Self Managing School; The Education of Indigenous Students.

Further Reading

- Alton-Lee, A. (2007). The Iterative Best Evidence Synthesis Programme, New Zealand. In *Evidence in Education: Linking Research and Policy*, pp 71–80. Paris: OECD.
- Bishop, R., Berryman, M., Cavanaga, T., and Teddy, L. (2007). *Te Kōtahitanga Phase 3 Whānau Tangata: Establishing a Culturally Responsive Pedagogy of Relations in Mainstream Secondary School Classrooms*. Wellington: Ministry of Education.
- Fiske, E. B. and Ladd, H. F. (2000). *When Schools Compete: A Cautionary Tale*. Washington, DC: Brookings Institution Press.
- Gilbert, J. (2005). *Catching the Knowledge Wave*. Wellington: NZCER Press.
- Goedegebuure, L., Santiago, P., Fitznor, L., Stensaker, B., and van der Steen, M. (2007). *Thematic Review of Tertiary Education: New Zealand Country Note*. Paris: OECD Directorate for Education, Education and Training Policy Division.
- Ministry of Education (2002). *Pathways to the Future: Nga Huarahi Arataki*. Wellington: Author.
- Ministry of Education (2006). *Educate. Ministry of Education Statement of Intent 2006–2011*. Wellington: Ministry of Education.
- Ministry of Education (2007a). *The New Zealand Curriculum for English-Medium Teaching and Learning in Years 1–13*. Wellington: Crown/Learning Media.
- Ministry of Education (2007b). *Statement of Strategic Intent 2007–2012*. Wellington: Author.
- Ministry of Education (2007c). *State of Education in New Zealand 2007*. Wellington: Author.
- Ministry of Education (2007d). *Tertiary Education Strategy 2007–12*. Wellington: Author.
- Ministry of Education (2008). *Ka Hikitia – Managing for Success: Māori Education Strategy 2008–2012*. Wellington: Crown/Ministry of Education.
- OECD (2007). *Improving School Leadership: New Zealand Background Report*. Paris: OECD.
- Timperley, H., Wilson, A., Barrar, H., and Fung, I. (2007). *Teacher Professional Learning and Development. Best Evidence Synthesis Iteration*. Wellington: Ministry of Education.
- Wylie, C. (2007). *School Governance in New Zealand: How Is It Working?* Wellington: New Zealand Council for Educational Research.

Relevant Websites

- <http://www.educationcounts.govt.nz> – Education Counts.
- <http://www.ero.govt.nz> – Education Review Office – Te Tari Arotaki Mātauranga.
- <http://www.minedu.govt.nz> – Ministry of Education.
- <http://www.nzcer.org.nz> – New Zealand Council for Educational Research.
- <http://www.nzqa.govt.nz> – New Zealand Qualifications Authority.
- <http://www.teacherscouncil.govt.nz> – New Zealand Teachers Council.
- <http://www.tlri.org.nz> – Teaching and Learning Research Initiative (TLRI).
- <http://www.tec.govt.nz> – Tertiary Education Commission.

Niger

A A I Daouda, University Abdou Moumouni de Niamey, Niamey, Republic of Niger

© 2010 Elsevier Ltd. All rights reserved.

Context

Niger is a desert country (three-quarters of its surface), its area being 1 267 000 km². In the area bordering Benin and Nigeria are concentrated three-quarters of the population, but conditions are difficult because of insufficient and irregular precipitation and poor fertility of the soil. Almost 84% of inhabitants live in rural areas. The population of Niger is estimated at 13.475 million (2007), women representing 50.1% of the population. Niger has high birth rates (7.1 children per woman), infant mortality rate (14.8%), and maternal death rate (648 out of 100 000). According to the UN projections, the population will reach 27 million in 2025. The age structure of the population shows a clear predominance of young people; indeed, half the population is less than 15 years, while only 3.2% of the total population is aged over 65 years. The extent to which basic needs are met is set out in **Table 1**.

Niger is a country dominated by the primary sector (namely, agriculture) occupying more than 80% of the population. This sector grew at a rate of 2.2% from 1995 to 2000, and its share of GDP reached 42.8% between 2002 and 2006. Given the climatic risks, the economy is frequently unstable. The share of the secondary industry remains less important, and fell to 13.2% in the period 2002–06. This situation is explained by the prevalence of an unaccounted largely rural sector (70.8% of the GDP in 1990; 74.3% in 2000). Such a structure results in very limited revenues from taxes and affects the capacity of the state to finance its operating costs. According to government's strategy or the development and reduction of poverty (SDRP), "the evolution of the economy of Niger over the decade 1990–2000 was characterized by a weak growth with a real rate annual average of 1.9%. The principal factors related to this poor economic performance are the unfavorable climatic conditions, an inappropriate economic policy characterized by delays in the reforms, an unfavorable external environment, even hostile, and a social and political climate that is often unstable. In terms of structure, the GDP was marked by a preponderance of the tertiary sector coming from trade. The contribution of this component to the formation of the GDP was on average 42.2% during the decade 1990–2000."

The demographic growth rate (3.3%) until 2003 was higher than the annual rate of growth of schooling (3% in 2001, compared to 2% in the preceding years), that is, the supply of education did not match demographic growth. From 2003, with the implementation of the PDDE, access

to schooling made important gains; however, with demographic growth rate reaching 5.1% in 2006, the problems facing education remain enormous. In general, the social deficits with respect to the right of access to basic services (e.g., education, health, information, credit, human rights and protection, land) and other development needs have an extremely harmful effect on the most vulnerable groups, in particular women and children. In the field of education, the indicators place Niger in a very low position globally with a great disparity between boys and girls, and also between urban and rural areas. The education of girls is an essential requirement for development in Niger. Low levels of education and literacy are among a number of the important factors that explain the very low level of the index of human development (IDH) in Niger.

While French is the official language of Niger, in fact, ten languages are spoken, the principal ones being Haoussa (53%), Zarma (21%), and Peulh (10.5%). Niger is a country strongly Islamized (95% of the population, against 4% Christianity, and 1% other religions).

Structure and the Management of Education

At the time of independence, in 1960, Niger was the African country with the lowest provision for education. Educational policy was determined by the need for a rapid expansion of an inherited school system. There were neither better alternatives, nor the time to develop and try out new systems of education. The policy of manpower is certainly called into question during debates on the reform of education: for example, (1974) the meeting of the national commission on the reform of teaching; (1975) the national seminar of reflexion on the reform of teaching; (1979) the seminar general on education; (1982) the national debate on the school native of Niger; (1987) the setting-up of the commission of reflexion on the training of the trainers; (1987) the seminar on the actualization of the programs of the lesson of first and second degree; (1992) states general of education; (1998) law of orientation of the education system; (2001) letter of educational policy; and (2002) decennial program of development of education. However, in practice, the implementation of reforms put forward at national forums met many obstacles, and so the education system evolved in a context characterized by the absence of planning and management according to immediate needs. The first law

Table 1 Basic needs indicators

Access to health services	65% in 2005 vs. 47.6% in 2001
Access to safe drinking water	68.7% in 2005 vs. 43% in 2000
Access to sanitation	78.4% of households have no toilet
Rate of schooling of the children	59% in 2007
Adult literacy rate	28.7% in 2005

for the orientation of the education system was promulgated only in 1998. The present education system of Niger is managed by three government departments:

- the Ministry of National Education (MEN);
- the Ministry for Educations Secondary and Higher of Research and Technology (MESSRT); and
- the Ministry for the Vocational Training and Technical (MFPT).

The formal sector of education comprises of:

- basic education (pre-school, basic cycle 1, and basic cycle 2);
- post-basic education; and
- higher education.

The nonformal education sector comprises of:

- elimination of illiteracy and
- Koranic teaching.

The government departments cover different systems: MEN – preschool education, primary education or basic cycle 1, and elimination of illiteracy; MESSRT – basic education cycle 2, secondary teaching, and higher education; MFPT – the establishments and centers providing vocational training and technical training at secondary level (basis 2). The three departments have an organizational structure similar to that at the regional level, which is under regional and local control:

- in each area, the educational zone is placed under the supervision of a regional director responsible for the coordination of all educational actions and policy;
- each under-area is subdivided into inspectorates according to the type of teaching;
- each inspectorate manages the establishments belonging to its entity at the level of secondary basic education 1, the inspections are organized in teaching sectors directed by education advisers; and
- educational establishments are managed by directors in the administrative and educational plan.

The formal education system in Niger is composed of:

- Preschool education is conducted in the nursery class and kindergarten for children from 3 to 4 years. The duration of this cycle is 3 years.

- Basic cycle 1 (primary education teaching) accommodates 7-year-old children. It includes traditional schools, free-Arabic schools, experimental schools, and specialized schools; it has a cycle of 6 years of studies sanctioned by the CFEPD or the BOLETUS free-Arabic.
- Basic cycle 2 (first cycle of the secondary education) accommodates, by way of an entrance examination in sixth year for 4 years, sanctioned by the BEPC or the BEPC french-Arabic.
- Secondary education (second cycle of the secondary education) that constitutes the second degree of teaching is composed of a general teaching and vocational teaching. The access to average teaching is open to the holders of the patent of studies of the first cycle of secondary education (BEPC). This cycle lasts 3 years and is sanctioned by obtaining the baccalaureat.
- Higher education is the third level of education and includes the whole of the post-school sector.

Educational Policy

The primary goal of the educational policy of Niger is the construction of a system of education able to better develop human resources for a harmonious social, economic, and cultural development of the country. Article 13 of the education law stipulates that “education must be complete. It aims at the development of the intellectual abilities, physical and morals, the improvements in education for a social and professional integration and the full exercise of the citizenship.” The government, in dialog with all the partners of the school and the support of technical and financial partners (PTF), has undertaken the development of the Decennial Programme of Development of Education (PDDE) in Niger over the period 2003–13. This program must henceforth be used as a comprehensive framework of reference to all the interventions in the sector. After the adoption of the law on 1 June 1998, the government confirms by this program, its firm political will and engagement to make education one of its priorities of first order. This decennial approach conforms to the orientations of the Decade of Education of the OAU, with the Special Initiative of the United Nations for Africa, with Tally of action of Dakar on Education for All (EPT) and with recommendations of the summit of the heads of states of the six least-provided-for countries (in terms of education) (Bamako), also translated the national concerns many times affirmed in the world conferences and summits. In fact, it is a question of putting in harmony the ambition of universal schooling and that of the development of a country economically strong, socially structured and stable, and culturally open. The required ultimate goal is the achievement of the fundamental educational needs for all layers of the population in order to enable them to ensure the control of

their destiny by effective development and utilization of human resources.

Strategic Objectives of the PDDE

The primary strategic objectives of PDDE are the following:

- to reinforce and develop the institutional capacities of the MEB1/A;
- to develop a partnership, active and diversified in favor of the system;
- to especially accelerate schooling in rural areas, particularly for girls;
- to promote nonformal education, particularly for young people and women;
- to reform and adapt the provision of education to reduce the disparities between areas, and between boys and girls;
- to improve quality of education by the restoration of the curricula; and
- to reinforce and develop vocational training and technical education.

Operational Objectives of the PDDE

The primary operational objectives of PDDE include:

- to change the rate of pre-schooling from 1% in 2002 to 5% by 2013;
- to raise the primary rate of schooling from 41.7% in 2002 to 91% by 2013;
- to carry the rough rate of schooling of the girls to the basic cycle 1 from 33.3% in 2002 to 91% by 2013;
- to change the primary rate of schooling in rural zones from 38% in 2002 to 90% by 2013;
- to carry the rate of schooling to the basic cycle 2 from 13% in 2001 to 23% by 2013;
- to change the proportion of manpower of the rural CEG from 19% in 2001 to 40% of the total staff complements of the basic cycle 2 by 2013;
- to increase the share of private from 14% in 2001 to 20% of manpower of the colleges by 2013;
- to increase the rate of elimination of illiteracy from 19.9% in 2000 to 38% by 2013;
- to create 50 centers of training under development community (CFDC);
- to carry the rate of survival to the basic cycle 1 from 57% in 2002 to 93% by 2013;
- to reduce the rate of redoubling to CM2 from 36% in 2002 to 10% by 2013 with an intermediate rate of 15% by 2006;
- to reduce the rate of average redoubling of the 6 to 3 from 20% in 2001 to 7% by 2013;

- to reduce the rate of redoubling in 3 from 27% in 2001 to 10% by 2013;
- to carry the rate of survival to the college from 52% in 2002 to 76% by 2013; and
- to carry the rate of success in elimination of illiteracy from 42% in 2000 to 80% by 2013.

Concerning post-primary education, the PDDE, in theory, will implement the reform program as it has with the primary education. Currently, after nearly 5 years of implementation of the PDDE at the level of basic education, significant progress has been made even if much remains to be done. With regard to preschool education, the number of establishments has strongly increased (15.7% on average annually between 2003 and 2005 as a whole and approximately 12% in the rural zones), but the total rate gross of preschool hardly exceeds 1.4% (2005–06). For basic education 1, the gross enrolment rate of primary schooling grew from 41.7% in 2002 to 54% in 2006 (against an objective of 57%). However, disparities according to place of residence remain high. The rate of completion improved from 26% in 2002 to 40% in 2006 (against an initial objective of 46%).

For secondary school education 2 (CEG, first cycle THESE, colleges offering free–Arabic learning), the gross rate of enrolment grew from 13% in 2002 to 17% in 2006. However, there was a high drop-out rate for girls. For secondary education, the gross enrolment rate is 4.1% (girls 2.7%, boys 5.9%). Drop-outs are more frequent among girls, particularly in rural areas. Concerning higher education, the rate of access to the university was only 73 students per 100 000 inhabitants in 2004, lower than the average for sub-Saharan Africa (230 students per 100 000 inhabitants). Technical and vocational training (EFPT) accommodated, in 2005, nearly 2.5% of the pupils in secondary education; of these more than half were in private establishments. Regarding the elimination of illiteracy, remarkable progress was recorded in recent years, even for adults (28.7% in 2005 against 19.9% in 2000). Lastly, teaching in the Koranic schools takes place within ten schools with a total staff complement of 5243, of which 2.948 are female.

Management

With the development and the adoption of the law of June 1998, Niger was equipped, not only with a legal framework for its education system, but also of a new model of decentralized management for its management, the aim of which is to ensure the participation of communities and local actors in financing and managing educational structures. The decentralization policy adopted by the ministry of education follows the law that stipulates that financing and managing publicly owned establishments will be ensured by the state, the communities, the

Table 2 Principal macroeconomic indicators, 2002–06

<i>Key indicators</i>	<i>2002</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>	<i>Average period</i>
Real GDP/head (in thousands of FCFA)	130.8	129.6	125.2	139	141.8	133.3
Growth rate of the real GDP (in %)	5.2	3.2	−0.8	7.2	4.8	3.9
Growth rate of the real GDP/head	1.9	−0.1	−4.1	3.9	1.5	0.6
Unaccounted sector in the nominal GDP (%)	69.9	71.8	69.8	70.4	70.7	70.5
Rate of investment (%)	16.1	16.3	14.6	23.1	21.8	18.4
Rate of interior saving (%)	8.2	7.6	3.7	10.2	10.4	8.0
Final consumption (variation in %)	6.1	4.2	3.1	−0.2	3.7	3.4
Rate of inflation (in %)	2.6	−1.6	0.2	7.8	0.1	1.8
National debt VAN (as % of exports)	221.7	217.6	428	134.6	46	209.6

Sources: INS/MEF; DGEP/MEF.

Table 3 Types of education, official age of entry, and duration of cycles

<i>Order of teaching</i>	<i>Official age of entry in the cycle in years</i>	<i>Duration of the cycle in years</i>
Preschool	4	3
Basic cycle 1	7	6
Basic cycle 2	13	4
Secondary	17	3
Superior	20	4

families, and all the concerned persons or entities. The burden-sharing is as follows:

- State: infrastructure, equipment, handbooks, and school stationery, training the trainers, hired labor charges, logistics, and purses.
- Communities: infrastructure, equipment, supplies, hired labor charges, maintenance, electricity, water, telephone.
- Families: supplies, maintenance of the schools.
- Partners in development: infrastructure, equipment, training the trainers, research, logistics.
- Company: formation, research, and training.
- Other NGOs: gifts and legacy.

Implementation of this law required the creation of several bodies: national, regional, and local. At the national level, the National Council of Education (CNE) is the body that determines the general orientation of education. The Regional Council of Education (CRE) and various Regional Councils of Education (CSRE) follow the same mission as the CNE, and are in charge of the follow-up of the educational policy. At the local level, the Board of Management of the School Establishments (COGES) is the management body that determines the day-to-day running of the school establishments.

Educational Sector Analysis

Tables 4–6 give data charting the progress of education since 2000.

1. Primary education: The access rate TBA (65% in 2007) remains below 100% in spite of a remarkable evolution during the last 5 years. There has been an annual increase of 13% in the number of new pupils entering schooling since 2003 following the implementation of the PDDE. Like the TBA, the gross enrolment ratio shows a marked improvement (37% in 2001 and 57% in 2007).

Taking into account the level of poverty of the populations entirely dependent on the state, the percentage of children from deprived families in private schools remains very low.

2. Secondary: The weakness of the TBS at the secondary level testifies to the weak development of this sub-sector and is explained by a very strong downturn due mainly to the policy of automatic promotion in primary education, and also to other factors such as the early marriage of girls especially in rural areas and the precarious living conditions of pupils obliged to leave their villages and go to urban centers to study. However, the continuing development of primary education as part of the implementation of PDDE involves new challenges with secondary education in Niger.
3. Superior: Like other education sub-sectors, higher education is underdeveloped in Niger. The country has only one national university and one Islamic university managed by the Organization of the Islamic Conference (OCI).

As for the other levels, the rate of access to higher education in Niger is lower than that in other countries of sub-Saharan Africa with similar levels of income. In 2005–06, this rate was estimated at 122 students for 100 000 inhabitants. At the University Abdou Moumouni (UAM), between 2001–02 and 2003–04, the enrolments dropped from 6833 to 6612 students before rising to 8710 in 2005–06. It is estimated that these enrolments are two to three times the capacity for absorption in the labor market of Niger, given its current economic base.

The UAM accommodates most of Niger's higher education student population. Within the UAM, the

Table 4 Evolution of gross rates of admission to schooling

	2000-01	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07
Provided education for manpower	158.623	204.069	215.946	242 047	248 910	279 210	279 353
TBA	40%	50%	51%	55%	55%	59%	65%
TBS	37%	42%	45%	50%	52%	54%	57%

Table 5 Pupils in public and private education (2006-07)

<i>Public + private</i>			<i>Only deprived</i>			<i>Percentage in private schools</i>		
<i>M</i>	<i>F</i>	<i>T</i>	<i>M</i>	<i>F</i>	<i>T</i>	<i>M</i>	<i>F</i>	<i>T</i>
722 653	512 412	1235 065	26 109	21 495	47 604	3.6%	4.2%	3.85%

Source: Statistical directory of the MEN, 2006-07.

Table 6 Evolution gross rate of schooling and enrolment of secondary pupils 2001-07

	2001-02	2002-03	2003-04	2004-05	2005-06	2006-07
Enrolment	109 293	123 930	215 946	155 151	177 033	210 626
TBS	10%	11.2%	13.7%	14.9%	17.1%	17.0%

distribution of students between various faculties is unbalanced. In 2005-06, the arts and social science (FLSH) and economics and legal studies (FSEJ) alone accommodated two-thirds of the registered students accounting for 35% and 32% of enrolments. Since 2001-02, enrolments in these faculties fell by 6%. This unequal distribution reflects an imbalance in admissions to the baccalaureat.

4. Technical and vocational education and training: Technical and vocational training represents the poor relation of the education system of Niger. It accommodates only 2.5% of students. However, the recent adoption of a national policy on technical and vocational training and the PPDE will undoubtedly allow it in future years to develop in ways adapted to national needs.

Disparities

1. Primary education: Regional disparities in term of gross enrolment rates declined over the period 2000-06. Indeed, dispersion around the national rate measured by the standard deviation of the index of disparity moved from 0.62 to 0.34 over this period.
2. Secondary level:

Qualitative Aspects

Success in examinations at the end of the year

1. Primary education level: One of the indicators of the quality of primary education is the rate of success in examinations at the end of the yearly cycle.

The rate of success in the CFEPD shows an improvement since 2001. However, one notes a drastic fall in 2005-06.

2. Secondary level: There is a downward trend in the rate of success in the BEP. This tendency was accentuated in 2006 with the reform in the organization of the BEPC that prohibits any fraud. The fall in the rate of success in the baccalaureat particularly since 2001 was due to the reforms introduced into the organization of the baccalaureat. With the BEPC, as with the VAT, the rate of success was affected by the problems of quality involved in the nearly exclusive recruitment of teachers of the new type (ASCN, EC). In 2006, the fall is explained exclusively by the anti-defraud provisions taken by the government.
3. Higher level: The gender disparities are marked even in higher education: sociocultural and economic obstacles mean a large loss of girls at school, making it rare for them to reach the higher level (source: statistical directory 2005-06, DEP/MESSRT, January 2007).

Factors influencing quality of education: teachers

The improvement of schooling observed in the implementation of the PDDE was achieved, thanks to a massive recruitment of teachers with numbers almost doubling between 2000 (15 668) and 2006 (31 131). However, the recruitment of full-time teaching staff has slowed while that of under-paid contractual teachers has grown from 17% of educational personnel in 2001 to almost 80% of the primary teaching service today, creating problems of professionalization and quality of teaching. In addition,

Table 7 Higher education in places other than Niger: Establishments and enrolments 2005–06

	<i>Publicly owned establishments</i>			<i>Private establishments</i>		<i>Total</i>
	<i>UAM</i>	<i>UIS</i>	<i>Large schools</i>	<i>University</i>	<i>Institutes^b</i>	
Number of establishments	1	1	6	1	26	35
Number of students	8710 ^a	800	1150 ^c	NB	2498	13 158

^aIncluding ENS.^bs/tutelle MESSRT.^cENAM+EMIG alone.

Source: Directory statistics 2005–06, DEP/MESSRT, January 2007; study on the development of the ESRS in Niger, J-C Lanzallavi, January 2007.

Table 8 Evolution of enrolments by faculty with the UAM, 2001–06

<i>Faculty</i>	<i>2001–02</i>	<i>2002–03</i>	<i>2003–04</i>	<i>2004–05</i>	<i>2005–06</i>
Sciences (FS)	943 14%	666 10%	558 8%	618 8%	719 8%
ENS	138 2%	184 3%	167 3%	219 3%	254 3%
Agronomy (F)	286 4%	299 5%	354 5%	324 4%	359 4%
Letters and social sciences (FLSH)	2648 39%	2624 41%	2609 39%	2971 40%	3086 35%
Sciences and health (FSS)	1153 17%	1093 17%	1088 16%	1227 17%	1523 17%
Economic and legal sciences (FSEJ)	1665 24%	1613 25%	1836 28%	2009 27%	2769 32%
Total UAM	6833	6479	6612	7368	8710

Source: Statistical directory 2005–06, DEP/MESSRT, January 2007.

Table 9 Technical vocational training, 2005–06

	<i>Public</i>	<i>Nongovernment</i>	<i>Total</i>
Number establishments ^a	12	41	53
Number of pupils	2 431	2 758	5 189

^aEstablishments concerned with the MFPT, including mixed establishments offering a higher formation.

Source: DEFPT/MFPT.

the small proportion of full-time professional teachers exacerbates the disparities already noted in access to basic education.

1. Secondary level: The weak development of the sub-sector of secondary education is compounded by small number of secondary teachers: 4267 for all cycles for 211 374 pupils. One finds here still the problem of the under-representation of women, and the fact that most secondary teachers are not civil servants (nearly 85%). Quality is problematic, given that teachers who are not civil servants rarely profit from any professional training.
2. Higher level: The number of permanent academic teaching-research staff has also grown only slowly from 245 in 1997–98 to 293 in 2005–06. On the other

hand, more and more recourse to non-permanent personnel was evident: these accounted for 47% of the teaching at the UAM in 2005–06.

Education and Economic and Sociocultural Development: External Effectiveness

There is neither an effective link between education and the labor market, nor is there an employment policy. With the support of the ILO, in May 2004 a study of the employment policy was undertaken, but as yet, the results of the study are not available. Indeed, the National Agency for the Promotion of Employment (ANPE) is the public utility of the modern job created in 1996 in place of the service of the labor. One of the missions of this institution is the follow-up of the labor market; it also provides information on the employment picture. ANPE has set up and carries out programs of promotion in various categories of the labor market. These include a program of insertion of young graduates in small to medium enterprises, a program of reconversion of the unemployed, a program of support of initiatives for independent employment, and a program of support for creation of rural jobs. Moreover, a steering committee was created in May 2003

Table 10 Evolution of gross enrolment ratios and regional disparities, 2000–07

<i>DREN</i>	<i>TBS %</i>				<i>Disparity index</i>			
	<i>2000–01</i>		<i>2006–07</i>		<i>2000–01</i>		<i>2006–07</i>	
	<i>Total</i>	<i>Girls</i>	<i>Total</i>	<i>Girls</i>	<i>Total</i>	<i>Girls</i>	<i>Total</i>	<i>Girls</i>
Agadez	44	38	73.6	68.6	1.2	1.3	1.3	1.4
Diffa	32	29	38.5	35.9	0.9	1.0	1.0	0.8
Dosso	42	31	63.0	50.7	1.1	1.0	1.1	1.1
Maradi	34	24	55.4	42.1	0.9	0.8	1.0	0.9
Niamey	98	96	101.3	99.0	2.6	3.2	1.8	2.1
Tahoua	32	21	52.3	37.2	0.9	0.7	0.9	0.8
Tillabéri	34	29	53.5	47.9	0.9	1.0	0.9	1.0
Zinder	26	20	48.6	42.0	0.7	0.7	0.9	0.9
Together	37	30	57.1	47.4	1.0	1.0	1.0	1.0

Table 11 Gross enrolment rates and disparities, 2005–06

	<i>Base 2 (%)</i>			<i>Means (%)</i>		
	<i>Boys</i>	<i>Girls</i>	<i>Together</i>	<i>Boys</i>	<i>Girls</i>	<i>Together</i>
Agadez	31.3	22.2	26.5	11.6	4.4	7.7
Diffa	11.4	8.0	9.6	3.9	2.2	3.0
Dosso	22.7	11.8	17.1	3.5	1.4	2.4
Maradi	23.9	8.7	15.7	4.9	1.4	2.9
Niamey	60.7	49.1	54.5	26.4	18.3	22.2
Tahoua	17.4	7.3	12.2	3.0	0.9	1.8
Tillabéri	13.5	8.3	10.7	1.9	0.9	1.4
Zinder	17.9	8.9	12.8	4.8	1.6	2.9
Niger	22.3	12.5	17.1	5.9	2.7	4.1

Source: Directory statistics 2005–06, DEP/MESSRT, January 2007.

Table 12 Evolution of the rate of success to the CFEPD of 2000–06

<i>Candidates</i>	<i>2000–01</i>	<i>2001–02</i>	<i>2002–03</i>	<i>2003–04</i>	<i>2004–05</i>	<i>2005–06</i>
Presented	86 688	76 565	72 221	85 948	100 161	111 603
Admitted	29 102	36 168	38 799	51 616	65 575	475 22
%	34%	47%	54%	60%	65%	43%

for the follow-up and evaluation of programs of employment. This committee is composed of the representatives of ANPE, the trade unions of employers, workers and the unemployed. However, it should be noted that the implementation of these programs is always not effective. Under these conditions, prospects for employment will improve only if the effectiveness of implementation improves.

It is noted that distribution in various sectors has not evolved systematically. However, for banks and insurance, transport and communication, and social services, there was a consistent increase in the number of employees from 1999 to 2003.

Table 13 Evolution of the rate of success to the BEPC and to the VAT, 1998–2006

<i>Year</i>	<i>% success BEPC</i>	<i>% success VAT</i>
1998/99	28	54
1999/00	32	43
2000/01	32	33
2001/02	31	24
2002/03	48	28
2004/05	55.4	39.2
2005/06	11.9	32.5

Source: DEP/MESSRT & ratio n°32649-, the World Bank, AFTH June 2, 2005.

Table 14 Disparities kind among faculty of the UAM, 2005–06

<i>Faculty</i>	<i>Enrolment</i>	<i>Gender</i>		<i>% Total girls</i>
		<i>Females</i>	<i>%</i>	
Sciences (FS)	719	63	9%	3%
Agronomy (F)	359	58	16%	3%
Letters and social sciences (FLSH)	3086	781	25%	41%
Sciences and health (FSS)	1523	458	30%	24%
ENS	254	36	14%	2%
Economic and legal sciences (FSEJ)	2769	487	18%	26%
Total UAM	8710	1883	22%	100%

Source: Statistical directory 2005-06, DEP/MESSRT, January 2007.

Table 15 Evolution of teaching service at the primary level, 2000–06

<i>Teachers</i>	<i>2000–01</i>	<i>2001–02</i>	<i>2002–03</i>	<i>2003–04</i>	<i>2004–05</i>	<i>2005–06</i>
Number	15 668	18 441	20 553	22 427	24 091	28 163
% women	33	34	35	35	38	40
% Contractual	17	44	52	58	63	71

Education and Sociocultural Development

In Niger, access to education at the primary level is primarily the responsibility of the state. The educational sector reflects the deep disparity that exists between the possibilities provided by the education system and the progression of young people in education and the labor market. With indicators among lowest in the world, the challenges facing schooling in Niger are made more complex with a great disparity between boys and girls. The quality of education is also a concern because the state lacks the capacity to attract, train, and employ qualified teachers; the unsuitability of much of the curriculum to the local context; the lack of textbooks and teaching materials; and deficiencies in infrastructure and school facilities. Enormous disparities exist between the cities and rural areas, provision for girls, and the distribution of educational services. The fundamental problem is poverty: Niger lacks the financial and material resources needed by the sector and it needs workable strategies to remedy the deficiencies. The lingering problems of illiteracy and lack of education for all are among the key factors that explain the very low level of IDH in Niger. However, the national authorities are conscious of the strong correlation that exists between the level of poverty in Niger and the low level of education of the population. Therefore, great efforts are being made by the government with support from partners in the development of programs that aim at increasing levels of schooling, elimination of illiteracy, and the vocational training in general, and in particular for girls and women.

As for the use of national languages in primary education, this has been a 27-year-old experiment undertaken in

Table 16 Situation of secondary teachers, 2006–07

<i>Teachers</i>	<i>Totals</i>	<i>Civil servant</i>	<i>Not civil servant</i>
Number	4267	655	3612
%	100	15.35	84.65
% women	20.70	27.80	19.40

42 schools. An evaluation was carried out in 1999 by the Ministry for National Education in collaboration with the GTZ. This evaluation showed that the experimental schools are more effective and better adapted to the socio-cultural context compared to the traditional schools and that, in certain localities, it encourages more of the population to be interested in school life. It showed, moreover, that the use of mother tongue-like languages for teaching facilitated learning by pupils. The training in mother tongue during the first years of the schooling does not act to the detriment of the acquisition of French – in fact, it is just the opposite, a good initiation in reading and writing in mother tongue facilitates the training in French.

Nonformal Teaching

Nonformal education covers the elimination of illiteracy and the training of adults, the community centers of formation under development (CFDC), and the renovated Koranic schools.

Of all these components, only the elimination of illiteracy has permanent structures that facilitate initial and continuous training. Thus, the programs of functional elimination of illiteracy are summarized as:

Table 17 Student and faculty numbers at UAM, 2005–06

	<i>Number</i>				
<i>Faculty</i>	<i>Students</i>	<i>Permanent faculty</i>	<i>Student/staff ratio</i>	<i>% of students</i>	<i>% of permanent staff</i>
Sciences (FS)	719	78	9	8%	28%
Agronomy (F)	359	26	14	4%	9%
Letters and social sciences (FLSH)	3086	71	43	35%	26%
Sciences and health (FSS)	1523	40	38	17%	14%
ENS	254	28	9	3%	10%
Economic and legal sciences (FSEJ)	2769	34	81	32%	12%
Total UAM	8710	277	31	100%	100%

Source: Statistical directory 2005–06, DEP/MESSRT, January 2007.

Table 18 Distribution of employees by branch of industry from 1999–2003

<i>Branch of activity</i>	<i>1999</i>	<i>2000</i>	<i>2001</i>	<i>2002</i>	<i>2003</i>
Agriculture, fishing, pisciculture	1412	1323	1623	1438	1150
Mining	2371	2373	2157	2285	2505
Manufacturing industry	1313	2574	2546	3069	3921
Electricity, gas, and water	2700	2862	2583	3789	3536
Public works and building	3679	3214	4284	5232	4177
Services: hotel, bar, restaurant	4077	5021	4621	5904	6664
Transport and communication	4504	4899	5243	5338	6847
Bank, Insurance	776	1518	1594	2198	2341
Social services	6259	6766	6489	7707	7679
Total	27 091	30 550	31 140	36 960	38 910

Table 19 Evolution of public education budgets (of billion F CFA currents)

<i>Voted budgets</i>	<i>2002</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>
Budget general	240.2	407.0	407.9	450.2	456.7	498.4
Budget MEN	23.6	35.0	48.4	56.0	59.3	69.0
% Budget MEN	9.8	8.6	11.9	12.4	13.0	13.8

Source: Ministry of the Economy and Finances and DRFM/MEN.

Table 20 Evolution of the recurring expenses between 1990 and 2002[illegible]

- learning to read, write, and count in the national languages;
- topics on health, for example, AIDS prevention, hygiene, pregnancy, and nutrition; and
- environment: for example, bush fires, desertification, agriculture.

The program for neo-literates involves support for the village press and libraries. The village press treats topics relating to agricultural and animal production, hygiene and improvement of living conditions of the rural populations, and the organization and management of community activities. As for the village library, it aims at encouraging adults to read, to realize the need to teach reading and writing, and to popularize technical topics. It also takes part in informing the rural population on its rights and duties and in explaining clearly to them the process of holding elections and the outcomes of democracy. All in all, the village library complements the rural press.

Financing of Education

Public expenditure

Since the implementation of the PDDE in 2003, the undersector of basic education has profited from important funds allocated by the financial partners.

In total, the expenditure of education consists of the wages of the personnel, which account for approximately more than 75% of the total expenditure.

Other sources of financing

Apart from public financing (state, parents, communities), education native to Niger receives other sources of financing from donor communities and development banks as technical and financial partners.

Conclusion

The education system is probably one of the least developed ones in the world. However, the difficult context and the deficiencies in the performance of the system have not prevented a few years of relative dynamism of the sector. This dynamism resulted from the engagement of the national authorities, the educational administration, educators, and other partners of the school. Since 2003, the interventions of the PTF through the PDDE also impacted the changes. However, at all levels of education, the development of the education system of Niger is characterized by significant disparities between the zones, the areas, and the sexes. The cover of the education system is higher in urban than in rural zones with regard to basic teaching I as compared to basic teaching II. The gender inequalities are more pronounced in basic teaching II. The improvements in the education system are especially represented by important increases in the number of educational personnel recorded in basic cycle I since the beginnings of the

implementation of the PDDE in 2003. Concerning teachers, one notes that, owing to lack of means, the state is recruiting very few full-time teaching staff and is forced to use contractual and the so-called national civic service staff with secondary education; the latter might not have received any teacher training. With the passing of years, these new types of teachers have become the majority, and the established minority of professionally trained teachers is concentrated in the urban zones. At the secondary level particularly, in spite of the crucial lack of classroom teachers, a great number of teachers easily give up chalk for an administrative office. Lastly, it is necessary to note that even the best education system will be functional only if the pupils remain in good health, and if the combined effects of education and health contribute to economic growth and reduction of poverty – two key objectives for all countries.

Further Reading

- Issa Daouda, A. A. (2005). *Optimal Management of Secondary Teacher in Niger*. Paris: UNESCO-IIEP.
- Republic of Niger, Cabinet of the Prime Minister (2002). *Document of Strategy of Reduction of Poverty: DRSP 1^{ère} Phase Prepared by the Government of Niger*. Niamey: Permanent Secretariat of the DSRP.
- Republic of Niger, Cabinet of the Prime Minister (2003). *Strategy of reduction of poverty: Report/ratio of implementation*. Niamey.
- Republic of Niger, Cabinet of the Prime Minister (2007). *Strategy of Development and Reduction of Poverty: SDRP 2nd Prepared Phase the Government of Niger*. Niamey: Permanent Secretariat of the DSRP.
- Republic of Niger, Ministry for Basic Education (2004). *Memorandum of the first joint review of the government of the Republic of Niger and the technical and financial partners. (PTF) of the sector of education, PDDE, Niamey*.
- Republic of Niger, Ministry for Basic Education. Ministry for Education – Secondary and Higher, Research and Technology (2004). *Policies of recruitment and teaching staff management in the sectors of education, Niamey*.
- Republic of Niger, Ministry for National Education (1998). *Law of the 1^{er} bearing June 1998 orientation of the education system native of Niger*. Niamey.
- Republic of Niger, Ministry for National Education (2007). *Tally of Expenditure with Medium Term 2006–2008*. Niamey: CDMT/MEN.
- Republic of Niger, Ministry for Education – Secondary and Higher, Research and Technology (2004). *Report/Ratio of the Committee Charged to Reflect on the Reorganization of the DIES and the Conditions of Access Has the Higher Teacher Training School*. Niamey: University Abdou Moumouni/ENS.
- Republic of Niger, Ministry for Education – Secondary and Higher, Research and Technology (2007). *Document of Project of the Primary PDDE Post*. Niamey: MESS/RT/MFP/T.
- Republic of Niger, Ministry for Education – Secondary and Higher, Research and Technology (2006–2007). *Directory Statistics*. Niamey: MESSRT/DEP.
- Republic of Niger, Ministry for National Education (2006–2007). *Directory Statistics*. Niamey: MEN/DEP.
- World Bank (2000). *Constraints and Spaces of Freedom for the Development in Quantity and Quality of Education in Niger*. Washington, DC: World Bank.

Relevant Website

<http://www.ins.ne> – Institut National de la Statistique.

Norway

H Farstad, Ministry of Education and Research, Oslo, Norway

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Sami population – Aboriginal minority population living in the Norway, Sweden, Finland, and Russia.

State education loan fund – Public Financial institution that provides grants and soft loans to students in higher education and upper secondary education and training.

Stortinget – National assembly of Norway.

Tertiary vocational education – Post-secondary vocational courses between 6 months and 2 years, not recognized as higher education.

Training company – Company or public institution that has accepted responsibility and been formally approved for the training of apprentices, as a regular part of national upper secondary vocational training.

Udir – Directorate of Education and Training. Under the Ministry of Education and Research.

General Background

Norway is located in the far north of Europe, with large areas north of the Polar Circle. It consists of the mainland and the islands of Svalbard and Jan Mayen, with a total area of 387 000 km² and a population of 4.7 million. Population density is 15 km⁻² in mainland Norway. Up to 31% of the population are in the age group 0–24 and 20% are 60 years and older. The annual population increase is 0.5–0.6%.

Norway is a unitary state, monarchy, and parliamentary democracy, independent since 1905, after several hundred years under Danish and Swedish rule. Norway is a member of North Atlantic Treaty Organization (NATO) and European Free Trade Association (EFTA). In a referendum in 1994, 52% of the population decided against full European Union (EU) membership. However, through the European Economic Area (EEA) Agreement, Norway is a member of the European Single Market and participates in several EU programs and institutional arrangements, for instance, in the area of education and training.

Norway has three political and administrative levels. Apart from the national level, there are 19 counties (regions) and 431 municipalities. Each of these units has a locally elected decision-making assembly and an executive body appointed by the relevant assembly. The national assembly (Stortinget) decides national political

principles and goals, as well as budgets and legal frameworks for activities under each ministry. Counties and municipalities have specific responsibilities within this framework. Local autonomy is a strong principle. Both counties and municipalities report directly to the national level.

Geographic, topographic, and climatic conditions as well as the availability of major natural resources – oil and gas, fish, forests, and waterfalls – strongly influence the economy and the production and population structure. The topography and climate limit the importance of agriculture. Small- and medium-sized enterprises (SMEs) constitute more than 99% of all companies. As much as 80% of them have less than five employees. SMEs employ nearly 70% of the total labor force. Only about 1 000 enterprises have 100 employees or more.

Despite the dependency on natural resources, Norway is a modern industrial nation. Approximately 21% of employment is in manufacturing and construction, whereas 75% is in tertiary sectors. Mechanization and automation characterize most industries, ensured by a high level of investments. Deliveries of goods and services to the oil sector are substantial. Energy-intensive metals and chemicals production, as well as shipbuilding, are large export-oriented industries. Fish farming has boomed over the last 20 years. The economy is open with an extensive foreign trade and thus vulnerable to fluctuations in international markets. In recent years, inflation has been at 2–3% and the unemployment rate fluctuates between 2% and 4%.

The general education level is high. More than 95% of young people enter upper secondary education and training, and more than 35% attend higher education. Female students are in the majority in upper secondary and higher education. Public spending in education is 6.6% of gross domestic product (GDP) and 16% of total public expenditures. The use of information and communication technology (ICT) is part of everyday family life and work for the vast majority of the population.

The immigrant population comprises nearly 380 000 persons (8% of the total population), of which 75% originate from non-Western countries. Immigrants reside in all 431 municipalities. Up to 48% of non-Western immigrants live in the Oslo area and constitute about 22% of the total population in the capital city. The level of education in the immigrant population varies greatly, according to the background of the country. Among persons born in Norway of two foreign-born parents, enrolment in higher education is higher than the country

average for the age group 25–29. Unemployment is higher than the country average in all immigrant groups.

Education and training are viewed as a central means to achieve national social, economic, and regional policy goals. Hence, the provision of those is considered a national, public responsibility and the education and training policy is shaped in the interface between cultural, economic, and social distribution policies. Vocational education and training (VET) is strongly emphasized in mainstream education, as well as in a broader Lifelong Learning (LLL) perspective.

Principles and Goals of the Education System

Norwegian education is rights based. The overall goal of the Norwegian education system is to give each individual the opportunity and necessary support to identify and develop his/her talents and interests to the end of becoming established as a satisfied, active member of society. Similarly, the education system is expected to provide the various sectors of society with qualified, critical, and constructive individuals who will support and contribute to the maintenance and further development of the welfare society.

Equality and freedom of choice are general political principles, which lie at the heart of Norwegian education and vocational training policy. All residents are to be ensured equal rights of access to quality education, irrespective of gender and social, geographical, and cultural background. Accordingly:

- all education and training in the public domain, including higher education, is supplied free of charge, where costs are covered by public budgets;
- every young person completing compulsory education is entitled by law to 3 years of upper secondary general education or VET;
- the supply of education and training should be of high quality and broad enough to allow for a range of choices, irrespective of geographical location and social factors; and
- state grants and soft loans are available to all students in need of financial support.

Structure and Operation of the Education System

Preschool education is gradually becoming part of the mainstream education system; however, attendance is not compulsory. Kindergartens cater to children aged 0–5 years. In 2006, almost 93% of all children aged 3–5 years had a place in kindergarten.

Mainstream education in Norway is 7 + 3 + 3 years – primary, lower secondary, and upper secondary education (see **Figure 1**). It has a centralized framework with national legislation, curricula, financing, and quality control. However, the delivery of education and training is decentralized to a large number of public and private providers, including more than 35 000 approved training companies (with apprentices). Vast majority of students are enrolled in public education institutions. Both public and private providers of national education and training receive financial support from the government. They are subjected to national quality control and are obliged to organize examination and assessment according to national guidelines, involving external experts.

In 2006, nationally recognized education and training programs were delivered by 3131 primary and lower secondary schools, 453 upper secondary schools, and 70 tertiary education institutions (**Table 1**).

Preschool and Early Childhood Care

Kindergartens cater to children aged 0–5 years. The national Framework Plan for the Content and Tasks of Kindergartens states that they should work in a goal-oriented manner toward children's development and learning, and stimulate children's linguistic and social competences. Besides being quality pedagogical institutions that provide all children a foundation for personal development and regular schooling, they provide care for children while their parents are at work or study. Hence, the kindergartens are also considered a means to promote gender equality in the adult population.

In 2006, about 80% of all children aged 1–5 years had a place in kindergarten. It is the aim of the government to obtain full coverage and introduce a legal right to attendance from 2009. Lack of qualified personnel is a serious obstacle to achieving this goal, as national law states that every kindergarten and each of its department must have staff with a specialized education at the bachelor's level. Approximately 30% of the 70 000 staff in Norwegian kindergartens are trained preschool teachers.

The municipalities are responsible for the establishment and operation of the preschool provision and for ensuring its quality. Nearly 50% of the kindergartens are privately owned and operated with public support, based on formal approval by the municipalities. Financing is shared between the public and the parents; the latter cover 22–30% of the total costs.

Primary and Lower Secondary Education

Compulsory schooling in Norway is for 10 years and children start school at the age of 6. Primary and lower secondary education are founded on the principle of a

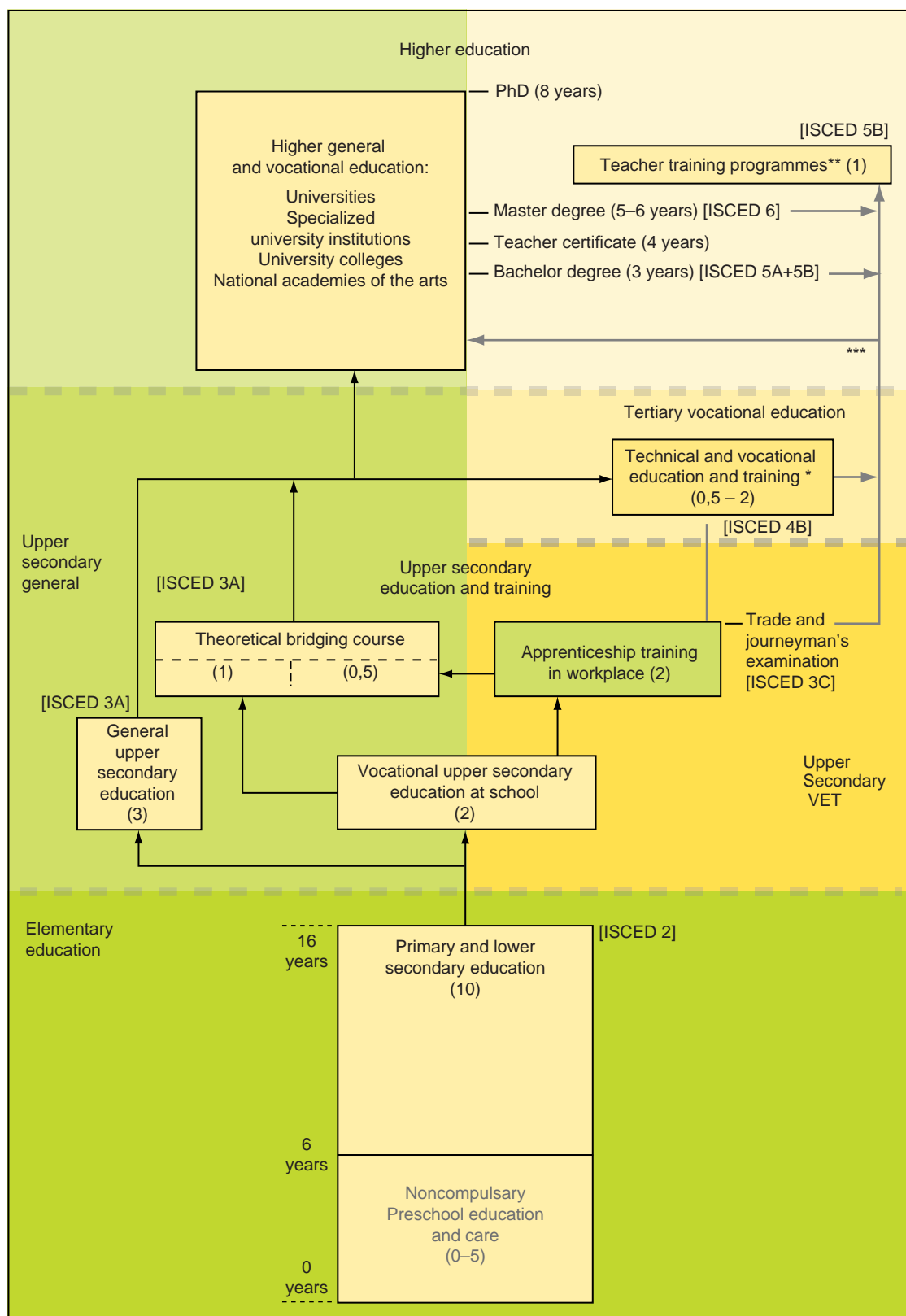


Figure 1 Education and training in Norway – 2008. *Technical school, training of civil servants, health personnel, etc.
 For candidates who decide to work as teachers after completing specialized academic or vocational training. *Candidates above 25 years that do not meet ordinary entrance requirements may obtain right of admission based on assessment of former practice and training. () Duration of training, years. From Ministry of Education and Research (Kunnskapsdepartementet).

Table 1 Key statistics on enrolment and resources in education, 2006

	<i>Institutions</i>	<i>Students (female)</i>	<i>Teaching staff (female)</i>
<i>Preschool (0–5 years)</i>	6436	235 000	69 700 (63 400)
Of these:			
Private	3535	107 700	
<i>Primary and lower secondary</i>	3131	619 000	51 400 (36 000)
Of these:			
Private	146	13 600	
Adult immigrants		30 400 (15 900)	
Adult Norwegians		10 700 (5 400)	
<i>Upper secondary</i>	453	223 500	20 700 (10 700)
Of these:			
Private	77	11 500	
<i>Vocational programs</i>		117 500	
Apprenticeship		34 300 (9 300)	
Adult Norwegians		67 400 (45 800)	
<i>Tertiary vocational^a</i>		10 000	
Of these:			
Private ^a		6000	
<i>Higher education</i>	70	211 600 (127 571)	16 800
Of these:			
Private	32	29 300 (17 500)	

^aReliable statistics missing due to change of system. Numbers are rough estimates for 2007

Data from Statistics Norway web pages, public web portal for the education sector and Statistics Norway publication: *Utdanning 2007*.

unified school system that provides equal and adapted education for all on the basis of a single national curriculum. An adapted curriculum is being used in the Sami population, and most Sami children have the opportunity to receive instruction in their mother tongue.

The subject curricula lay down a common learning content for all pupils, which increases in scope throughout the school and is greatest at the lower secondary stage. There is, however, room for adaptation to local conditions and to the particular needs of individual pupils. Immigrant students receive mother tongue language training.

Teacher education in Norway is currently 4 years at a university college, but an extension of 1 year is being discussed, thereby turning it into a 5-year master program. Teachers are employed by the municipalities, which own and run the schools at this level. Only those with a teacher diploma can be granted a permanent teacher contract.

Upper Secondary Education and Training

Young people, who have completed compulsory education, are legally entitled to 3 years of upper secondary education leading to admission to higher education, vocational qualifications, or to basic skills. All counties have a follow-up service for dropouts and young persons, who have not applied for, or not accepted, a school or training place despite their statutory right. The purpose and task of the service is to assist the youth in finding appropriate education, work, or to establish a tailor-made combination of education and work for them.

Table 2 Available vocational study programs and number of course alternatives at the various levels in Norwegian upper secondary education and training

<i>Study program</i>	<i>Level</i>		
	<i>VG 1</i>	<i>VG 2</i>	<i>Apprenticeship (VG 3)</i>
Building and construction	1	9	22
Design, arts, and crafts	1	24	46
Electricity and electronics	1	5	20
Health and social care	1	6	9
Media and communication	1	1	3
Utilization of natural resources	1	7	10
Restaurant and food processing	1	2	12
Service and transport	1	4	8
Technical and industrial production	1	16	56
Total	9	74	186

From Ministry of Education and Research

From 1976, VET, including apprenticeship, is an integrated part of upper secondary education and, from 1998, is regulated by the same act as general education. Entrants to upper secondary level can choose between 12 alternative study programs. Three of them prepare primarily for further academic studies, whereas the other nine equip the students with specialized vocational knowledge and skills (Table 2).

In general, upper secondary VET includes 2 years of school-based education with practical training in school

workshops and short work placements in industry, followed by 2 years of formalized apprenticeship training and productive work in an enterprise or public institution. National curricula in upper secondary VET cover the school-based as well as the apprenticeship part of the training programs. Students in all VET programs must also follow some classes in general subjects. Those who complete vocational training can add a theoretical bridging course and meet entrance requirements for higher education.

Public-private partnerships (PPPs) constitute the backbone of upper secondary VET. Employers' organizations and trade unions as well as individual companies and public institutions are major contributors to policy formation and operation of the system at national and county levels. The extensive collaboration is firmly institutionalized and takes place within formal frameworks at national and county levels.

Postsecondary VET

Tertiary vocational education is based on completed upper secondary general or vocational education and training, or equivalent knowledge and skills obtained through informal and nonformal learning.

Tertiary vocational education comprises a wide variety of vocational courses lasting from 6 months to 2 years. Around 3000 out of the some 10 000 students (in 2007) attend a 2-year training program for a technician. These students generally have completed upper secondary VET and possess a trade certificate. Technical training is in most cases publicly financed and delivered by the counties. All institutions and study programs under this arrangement must be approved by the government agency, National Agency for Quality Assurance in Education (NOKUT) (see below).

The program of continuing vocational education and training (CVET) provides training as a master craftsman for holders of a trade certificate and several years of relevant work experience for those who wish to set up their own business or hold a managerial position in a craft enterprise.

Higher Education

In 2007, higher education and professional training were delivered by 38 public and 32 private institutions. All the seven universities are public, whereas there are both public and private specialized university institutions and university colleges.

The public universities and colleges are owned and financed directly by the Ministry of Education and Research (MOER), but they have a high degree of professional and budgetary autonomy. In general, the recognized institutions have a degree structure, grading

system, and quality assurance system aligned with other European countries, based on the levels of bachelor, master, and PhD. Only a few university colleges have been approved for the delivery of PhD education, and some education programs still differ from the main degree structure.

About 12.5% of students in higher education attend private institutions. Most of them receive financial support from MOER, provided they meet professional requirements and deliver study programs that comply with national standards. Students in private institutions normally have to pay a tuition fee, but may receive financial support from the Norwegian State Educational Loan Fund that covers most of these expenses.

Higher education normally builds on the successful completion of 3 years of upper secondary general education, but enrolment is also granted on the basis of a combination of work experience and exams in six general subjects from upper secondary general education. Since 2001, access can also be granted for those older than 25 years on the basis of documented informal and nonformal competence.

Adult Education and Continuing Education and Training

All adults have a statutory right to primary (from 2002) and secondary (from 2000) education and training. The municipalities and counties, respectively, are obliged to organize appropriate schooling, free of charge. More than 50 000 persons participate annually in adult education provided by municipal and county authorities.

Second-chance opportunities for dropouts, who wish to come back and complete their education, have been in operation for decades. In general, one does not have to attend organized courses to be allowed to sit for the exam in individual subjects or full study programs. Candidates who meet the intake requirements for a study can register and take the examination as a private candidate. Examinations for private candidates are organized regularly.

Around one-quarter of the adult population participates in work-related updating as well as further education and training, annually. This reflects the need for a highly qualified workforce having gained experience in the industry in the knowledge-based economy. Those with highest education tend to attend CVET more often than those with less formal education, resulting in an increasing educational gap. This is a major challenge to the politicians and to the social partners, who cooperate closely with the government on ensuring that relevant LLL opportunities are available.

Training for the unemployed and for immigrants is organized and financed by public authorities. The volume of training provided varies according to labor market fluctuations and influx of immigrants. A variety of other CVET courses are organized by public, private, and semiprivate

providers. Universities and colleges annually provide further education and training for 80–100 000 individuals in their own program areas, delivered as full-time or part-time studies. Training, according to national curricula, on secondary and tertiary levels, as well as work- and personally-oriented courses are provided by 21 recognized adult education associations, 13 distance education institutions, and 77 folk high schools. In 2007, some 550 000 students attended courses organized by these structures.

Administrative Structure

Whereas legislative power lies with the Norwegian Parliament (Stortinget), the MOER has overall responsibility for national policy development and administration of mainstream education and vocational training at all levels.

Individual municipalities are responsible for the pre-school provisions, and for regular and adult primary and lower secondary education. The counties are responsible for public upper secondary general education and VET for all groups.

The Directorate of Education and Training (Udir) is a national agency responsible for development of curricula, national examinations, and quality control in compulsory school and upper secondary education and training. The National Agency for Quality Assurance in Education (NOKUT) is responsible for approval and quality control of tertiary vocational and higher education institutions and programs.

A county governor represents the government in each county. The county governor's education office controls

delivery and results of compulsory and upper secondary education and VET, including adult education (Figure 2).

In upper secondary VET, social partner organizations and the industry hold the majority of seats in all the most influential bodies at national and county level, listed in Table 3.

At the local level, training offices owned by local enterprises and institutions recruit and supervise training enterprises during apprenticeship on behalf of the county authorities. They often provide training for company instructors and vocational theory classes to apprentices.

Financing of Education

There is no political controversy over the current policy of free education delivery to all pupils and students at all levels in public training institutions. The government finances compulsory and upper secondary education and training through block grants to the municipalities and counties. MOER, in general, covers 80–90% of total budgets of public universities and university colleges, calculated for each institution according to a multi-element formula. The remaining part of their budgets is covered by contracted research and provision of tailor-made training to private enterprises and public institutions.

The municipalities are responsible for providing the establishment and operation of the preschool provision and for ensuring the quality. Nearly 50% of the kindergartens are privately owned and operated with direct financial support from MOER, based on formal approval by

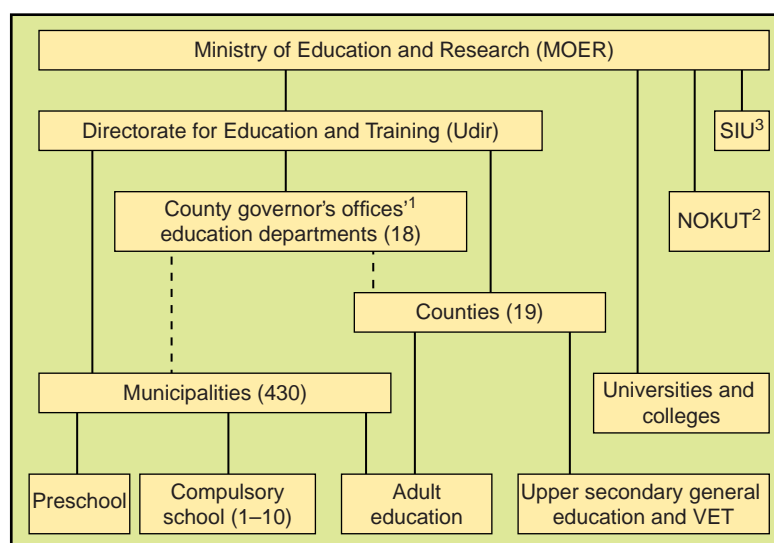


Figure 2 Administration of education in Norway. (1) The county governor is the representative of the national government at county level. (2) The Norwegian Agency for Quality Assurance in Education. (3) The Norwegian Centre for International Cooperation in Higher Education. From *Ministry of Education and Research* (Kunnskapsdepartementet).

Table 3 Overview of bodies in upper secondary VET in which external partners hold the majority of seats and hence execute their influence

<i>Level</i>	<i>Body</i>	<i>Role, responsibilities</i>
National	SRY ^a	Develops national framework for and structure of public VET, including school-based VET, recommends structure of vocational training councils and organizational representation in these, develops national procedures for examination and certification.
	Vocational training councils ^b	Define qualification levels of training in trades under their authority; give input to development of national curricula (VET); and appoint members of SRY and appeals boards.
	Appeals boards	Evaluate and take the final decision when candidates who fail the trade and journeyman's examination appeal the decision of the county-based examination board.
County	Vocational training committees	Recommend regional provision structure and volume of VET. Appoint examination boards, supervise and control apprenticeship training and trade examinations, and issue trade certificates. Main responsibility for development and implementation of quality control system in VET.
	Examination boards	Organize and implement the practical part of the trade and journeyman's examinations and evaluate the results.

^aThe National Council for Vocational Education and Training (Samarbeidsrådet for yrkesopplæring – SRY).

^bThere are nine vocational training councils, one for each of the nine vocational program areas.

the municipalities. Financing is shared between the public and the parents; the latter cover 22–30 % of the total costs.

Approved private education and training institutions that provide public programs receive financial support at the level of 75–90% of costs from the government and cover the remaining part of their budgets by tuition fees paid by the students. Training companies and training offices also receive some financial support from the education authorities, as they accept responsibility for training under public education programs.

A basic principle is that education and training that go beyond mainstream education should be financed by the beneficiaries, that is, the enterprises, the social partners, and the individuals. However, since it is in the interest of the nation to have a well-educated population and a competitive production sector, there is a system of public co-financing by means of targeted subsidies and tax deduction on training investments.

In order to ensure equal access regardless of family economy, students in upper secondary and tertiary education and training, including apprentices, may receive personal grants and subsidized loans from the State Education Loan Fund. This also applies to adult employees who take a leave for further education and training.

Internationalization is a major political objective in upper secondary and tertiary education. All students in tertiary education are encouraged to take half a year or more of their studies in another country. At least 15 000 Norwegians are studying abroad at all times, financially supported by the State Education Loan Fund.

Education and training delivered by formally recognized adult education associations, distance education institutions, and folk high schools is subsidized by the Ministry, but these nonprofit organizations cover most of their budgets by participant fees.

Training for employment under the labor market measures, as well as education of legal immigrants for integration purposes, is fully financed by the government.

Monitoring, Evaluation and Education Research

The legal framework for assessment, validation, and accreditation of learning, as well as for relevant quality-control mechanisms, varies according to the type and level of training and the formal status of training and its appurtenant certificate.

In general, public education authorities are concerned only with training and validation arrangements that are formally recognized under the national education system. Privately organized training that is not aiming for formally recognized certification is not regulated or monitored by the education authorities. However, like any other commercial transaction, privately organized training deliveries are covered by general market regulations.

Individual public and private institutions that have been approved for delivery of recognized training programs are obliged to have a documented system for and to conduct internal quality control of performance and outcomes. Municipalities, counties, and owners of private institutions are responsible for ensuring that such systems are developed and are being used. The national agencies, Udir and NOKUT, conduct random quality control of the training providers.

In every county, there is a system for assessment of prior (nonformal and informal) learning outcomes in relation to upper secondary education and training, whereby immigrants as well as Norwegians may have their vocational qualifications assessed and formally documented with reference to the national program curricula.

Immigrants with documented higher education qualifications from countries and institutions that are not covered by formal agreements on mutual approval of education credits may have their papers assessed by NOKUT.

Education research is conducted by a number of research institutes and higher education and research institutions all over the country, often at their own initiative. MOER and Udir frequently commission studies on selected topics and the ministry employ a number of analysts who are working continuously to identify and further investigate interesting or alarming events and trends in the education sector. Observations are reported back to the decision makers for policy adjustment purposes. Applied and formative research is often implemented in preparation of education reforms and during reform implementation, respectively, in order to establish the situation and reform effects on the ground and guide policy adjustments.

Major Changes and Reforms Since the 1990s

In a rapidly changing world, education reform and policy adjustments are necessarily a continuous exercise. Hence, all parts of the education system have been subjected to adjustments and reforms of varying significance: preschool/kindergarten (2004–2006), compulsory school (1997, 2006), upper secondary education and VET (1994, 2006), tertiary VET (2000, 2003), higher education (1994, 2004), and adult education (1999–2002). (Changes and reforms in terms of legal adjustments with implications for rights and obligations, major changes in content and structure, as well as the introduction of new institutions and re-distribution of responsibilities.)

The reforms have been based on broad political consensus and have targeted both structure and content of training at all levels with the intention to mutually adapt the various parts to each other and create a more streamlined system. A common core curriculum developed in the early 1990s (MOER, 2007a) constitutes the political foundation, providing an overarching framework of principles and goals in education and training, emphasizing both individual rights and needs of society. Major objectives of the reforms have been to:

- further strengthen the general access to the various parts of the education system for all groups, promote equality, and reduce obstacles;
- improve quality and relevance in all parts of the system to meet the needs of the individuals, the labor market, and society in general;
- enhance flexibility in the system in order to allow local adaptations and rapid responses to changing needs relating to developments in technology, markets, and society;

- reduce obstacles to students who wish to move horizontally between academic and VET training paths, to the end of optimal development of human resources;
- ensure the availability of relevant provisions of lifelong and lifewide education and training for adults, as well as various support arrangements to meet the needs of adults choose to educate themselves further.
- increase the efficiency and cost efficiency in all parts of education, at all levels; and
- promote internationalization in education and training.

To these ends, MOER has:

- decentralized responsibilities, but at the same time, strengthened central-level administration and quality control;
- adjusted the administrative and regulatory systems, as well as the financial arrangements for students and providers of education and training according to the strategic intentions of outreach, quality, relevance, control, and flexibility;
- strengthened the emphasis on a broad and general initial education including some core skills in preparation for further specialization and LLL;
- updated the curricula in all types and at all levels of education and training; and
- aligned the Norwegian credit and degree system in higher education with the system of most European countries within the framework of the so-called Bologna Process.

In 2004, the Norwegian government introduced a political strategy with the aim to promote entrepreneurship in education, thereby stimulating innovations and establishment of new production. Three ministries are behind the strategy: MOER, Ministry of Trade and Industry, and Ministry of Local Government and Regional Development. The major social partner organizations and the industry were actively involved in planning the strategy and have formed a separate administrative structure for the implementation, which is based on joint public–private financing (MOER, 2006a).

Major Challenges and Education Policy Issues as of 2008

Enrolment figures show that there are no major structural obstacles to school attainment at compulsory and upper secondary levels in Norway. Inclusive education and follow-up arrangements for groups and individuals with specific needs of pedagogical adaptation and financing are in place. However, performance indicators show that there is still a potential for quality improvements, particularly related to the provision of core skills and the theoretical foundation for further learning:

- Norwegian scores in international tests, for example, PIRLS and Pisa, fall short of the political ambitions; and

- nearly 30% of students who enter upper secondary education drop out or fail to obtain a completed general or vocational certification during a period of 5 years.

Many factors influence these results. The general opinion is, however, that more emphasis must be put on equipping the children with basic literacy and numeracy skills in the early stages of education. More resources should be available for early intervention in cases where children need particular support. Recent research shows that the major explanation for the high dropout rate during the school-based part of VET is insufficient theoretical results in compulsory school and low motivation for further theoretical studies.

In response to the recent findings, MOER is now proposing to give each child a legal right to preschool provision and has increased the financial support to the development of new kindergartens nationwide. School curricula have been changed with the aim to give more attention to basic literacy and numeracy skills at the lowest levels of education, and the training of preschool and ordinary teachers is being strengthened accordingly.

Further improvements in quality and relevance are a permanent issue on the agenda in all parts of the education system. It is a major challenge to further develop appropriate mechanisms to the end of foreseeing short-term and long-term qualification requirements of the labor market. The need to strengthen the services related to counseling and vocational guidance in an increasingly complex world of education and work opportunities, is another issue that receives increasing attention. In the LLL perspective, there is a need for a new and effective strategy on how to motivate and recruit adults with weak educational background to further training, including relevant financial and other support arrangements. Last but not least, an important task is to further align the Norwegian education and training system with systems in other countries in order to further reduce obstacles to international mobility.

Bibliography

MOER (2006a). *See the Opportunities and Make them Work! Strategy for Entrepreneurship in Education and Training*. Oslo. <http://www.regjeringen.no/Upload/KD/Vedlegg/Grunnskole/Strategiplaner/STRATEGI%20ENG.pdf> (accessed May 2009).

MOER (2007a). *Core Curriculum for Primary, Secondary and Adult Education in Norway*. Oslo. http://www.udir.no/upload/larerplaner/generell_del/Core_Curriculum_English.pdf (accessed May 2009).

Further Reading

Byrhagen, K. N., Falch, T., and Strøm, B. (2006). *Frafall i videregående opplæring: Betydningen av grunnskolekarakterer, studieretninger og fylke. (Dropout in Upper Secondary Education and Training: The Significance of Results from Compulsory School, Study Programmes and County)*. Trondheim. Senter for Økonomisk Forskning AS. SØF report 08/06. http://www.regjeringen.no/upload/kilde/kd/rap/2006/0031/ddd/pdfv/301003-sof_frafall_endelig_med_isbnnr.pdf (accessed May 2009).

CEDEFOP (2005). *Overview of Norwegian VET, all levels*. Thessaloniki. http://www.trainingvillage.gr/etv/Information_resources/NationalVet/Thematic/criteria_reply.asp (accessed May 2009).

Farstad, H. (2007). Public-private cooperation in vocational education and training: framework, implementation and experiences. The Norwegian case. *Report prepared for the Biennale 5th-9th May 2008 in Maputo on Post-Primary Education, organized by ADEA—the Association for Development of Education in Africa*, Oslo: MOER.

Markussen, E., et al. (2008). *Bortvalg og Kompetanse. (Dropout and Competencies)*. Report 13/2008. Oslo. NIFU STEP. <http://nifu.pdc.no/publ/index.php?sid=132878&t=R> (accessed May 2009).

MOER (2004). *The Development of Education 2000–2004*. Oslo. National report of Norway. http://www.regjeringen.no/nb/dep/kd/dok/rapporter_planer/rapporter/2004/The-Development-of-Education-20002004-National-Report-of-Norway.html?id=106213 (accessed May 2009).

MOER (2005–2007). *The Knowledge Promotion Reform*. Several documents, Oslo. <http://www.regjeringen.no/en/dep/kd/Selected-topics/andre/Knowledge-Promotion.html?id=1411> (accessed May 2009).

MOER (2006). *Directives to the Education Act*. Oslo. http://www.lovdato.no/cgi-wift/wiftldes?doc=/usr/www/lovdato/for/sf/kd/kd-20060623-0724.html&emne=oppl%C6ringslov*& (accessed May 2009).

MOER (2007). *Education Act, English Translation*. <http://www.ub.uio.no/ujur/ulovdata/lov-19980717-061-eng.pdf> (accessed May 2009).

Wærness, Jarl Inge and Lindvig, Yngve (2005). *Analyse av Lærlingsinspektorene 2003–2004. (Analysis of the Apprentice Inspectors 2003–2004)*. Oslo. LÆRINGSslaben NIFU-STEP. Rapport 1/2005. http://udir.no/upload/Rapporter/Analyse_av_larlingsinspektorene2003-2004.pdf (accessed May 2009).

Relevant Websites

<http://www.regjeringen.no> – MOER, Norway.

<http://www.oecd.org> – OECD (2007). Education at a Glance. Paris.

<http://utdanning.no> – Portal for education information (only in Norwegian).

<http://www.ssb.no> – Statistics Norway: Latest National Statistics in English by Theme.

<http://www.udir.no> – UDIR.

General Background

Oman is located on the southeastern edge of the Arabian Peninsula, with a total land area of 309,500 km². It shares its borders with Saudi Arabia, the United Arab Emirates, and Yemen and overlooks the Arabian Gulf, the Gulf of Oman, and the Arabian Sea. As a gateway between the Indian Ocean, East Africa, and the Arabian Gulf, Oman's location has always been strategically important. It covers 309,500 km² with a coastline extending 1700 km. Coastline plains represent about 30% of the total land area, mountain ranges occupy another 15%, and the remainder of the country is mainly sand and gravel desert. Although the mountains only cover about one-sixth of Oman's total area, they have always dominated the country at their feet, and determined the pattern of settlement there. The Sultanate of Oman is divided into eight administrative regions, which are further subdivided into 59 districts or *wilayats*. Each *wilayat* is governed by a mayor, *wali*, who is responsible to the Ministry of Interiors.

Oman's history begins in the early third or late fourth millennium BC, with the rise of a society that had cultural and trade ties to ancient Mesopotamia. Between the fourth and seventh century AD the area was dominated by successive Persian empires. During the first century AD Arab tribes began to migrate into Oman. When the new religion of Islam spread throughout the region in AD 630, Persian rule ended and Oman's Islamic-Arab character was firmly established. In 751, *Ibadhi* Muslims, a moderate branch of the *Kharijites*, established an Imamate in Oman. (The term *Ibadhiyah* applies to a Muslim group, which was considered by most writers as a moderate branch of the *Kharijite* movement. *Al-Ibadhiyah* is one of the earliest Islamic sects, the foundation of which goes back to the first half of the first century H. The adherents of this school still form a number of independent communities holding fast to its teachings. The largest of these today is in the Sultanate of Oman, where *Ibadhis* form the majority and *Al-Ibadhiyah* is the state religion. There are other minorities in Zanzibar on the east coast of Africa, in Jabal Nafusah and Zuwarah in Libya, on Jerba Island in Tunisia, and in Wad Mzab in Algeria.)

Contact with the Western world began when Portugal seized Muscat and other coastal strongholds in the early sixteenth century. Portuguese power waned after 1624, when a strong line of imams asserted itself, and Muscat was recaptured in 1650. The Imamate then flourished again under the *Ya'aribah* dynasty, which extended Omani rule or

influence to both sides of the Arabian Gulf and the East African coast.

Omani society is highly structured, both politically and religiously. Politically, it is a long-standing monarchy; the Sultan is head of state, the highest and final authority, and the supreme commander of the armed forces. There have been some democratizing trends, however, created by Royal Decree in 1997, the basic statute of the state established the state Council. The Council is a bicameral legislature with the *Majlis al Shura* (Consultative Council), whose members are elected by Omani citizens, and the *Majlis al Dawla* (State Council), whose members are appointed by the Sultan. (State Council (*Majlis al Dawla*) is an appointed body of 58 members, picked by the Sultan, established in 1997 and Consultative Council (*Majlis al Shura*) is an elected body of 83 members, established in 1991.)

Islam, the state religion, exerts profound influence over most aspects of Omani society in that the *Quran* prescribes politics, economics, law and justice, and social behavior, as well as theology. Most Omanis are *Ibadhi* with minorities adhering to *Sunni* or *Shia* Islam (Karabenick and Moosa, 2005). Oman's legal system is founded upon the Islamic *Shari'a* traditions of the *Ibadhi* school. Civil, criminal, and commercial cases are handled by the courts of first instance; matters of personal status and family law fall under the jurisdiction of *Shari'a* courts. The White Book which was enacted in 1996 is the most important piece of legislation. Although the Basic Law states that the judiciary is independent, it remains subordinate to the Sultan and the Ministry of Justice. In general, the legal system is based on both English common law and Islamic law. Religious conflicts have not been prominent in Oman since tolerance to both Muslims and non-Muslims is emphasized.

Oman's population was 2,331,391 as of the 2003 Census with a quarter of this number being expatriates: the percentage of females is 49.1% of the total national population. **Table 1** shows the distribution of Omani population by educational status and gender. However, the number has been increasing since that year. It is estimated that the total number extends to 3 million.

Education has expanded dramatically since 1970, when only three primary schools existed and few girls received any schooling. Education is provided free to all Omanis but is still not mandatory. About three-fourths of Oman's adult population is literate; there has been a substantial increase in the number of literate women. Karabenick and Moosa (2005) stated that comparative enrollments of boys

and girls indicate progress in reducing the gender gap with only a slight difference in the gross net enrolment ratios (Table 2). Thus, education is one area in which there is greater equity. Ninety-two percent of all girls who are eligible attend schools (grades 1–12). Girls also represent half of the students attending public schools, approximately half of the students attending Sultan Qaboos University (SQU), and they generally receive half of the government grants to study abroad.

As of 2007 statistics, the number of public schools is 1053, enrolling 563,602 students, employing 39,993 teachers and 4521 administrators. In addition, the number of private schools is 170, enrolling 32,134 students (Ministry of Education, 2007).

The Seventh Five-Year Development Plan (2006–2010), is a continuation of the previous 5-year plans but distinct from the others in that it witnesses the implementation of the development of education beyond basic education representatives in grades 11 and 12 along with the expansion of the application of basic education, in addition to the continuous development and improvement of the quality of educational services in light of the outcome of educational and performance indicators achieved by the Sixth Five-Year Development Plan (2000–2005). One can recognize three stages in the development of education in Oman:

Table 1 Percentage distribution of Omani population by educational status and gender

<i>Educational status</i>	<i>Male %</i>	<i>Female %</i>	<i>Average %</i>
Illiterate	11.81	23.82	17.77
Can read and write	21.17	18.46	19.83
Primary stage of education	23.66	18.10	20.90
Preparatory stage of basic education	19.14	15.78	17.47
Secondary stage	17.36	18.39	17.87
Postsecondary nontertiary	2.77	2.29	2.53
First university degree	3.50	2.73	3.12
Master degree	0.38	0.10	0.24
PhD degree	0.07	0.02	0.04
Not stated	0.14	0.31	0.22

From Sultanate of Oman, Ministry of Information (2007). *Oman 2006–2007*. Oman: Ministry of Information.

- Stage one emphasized the rapid quantitative development of education.
- Stage two started in the early 1980s, when the Ministry of Education (MoE) initiated serious efforts to improve the quality of education.
- Stage three began in 1995, after the Conference on Oman's Economic Future, Vision 2020, when a number of reforms were introduced in order to cope with the educational requirements of the future.

Goals of the Education System

The major aim of education and its philosophy in the Sultanate is to develop the current capabilities of the Omani people. Education is one means through which this aim can be realized. It provides the individual with the basic knowledge that enables her/him to objectively and advantageously employ scientific knowledge and skills that are acquired throughout life. Education also stimulates innovation and helps to develop the linguistic skills necessary for oral and written communication that mirrors the Islamic religion and its high ideals. Furthermore, it seeks to inculcate in Omanis the precepts of Islam to the extent that they are reflected in attitude, word, deed, and lifestyle. It also takes interest in the psychological and mental well-being of the individual, encouraging healthy practices and attitudes to work, health, and hygiene. The general aims of education before 1995 are as follows (Al-Shanfari, 1991: 90–114):

1. To create an Omani generation that is capable of fulfilling its responsibility within the national and patriotic framework.
2. To keep up with the progressive changes of modern life.
3. To develop knowledge in keeping with the noble principles derived from Islam.
4. To develop the physical and emotional aspects of the individual.
5. To acquire an appropriate awareness and skills of the nature of common professions and careers in the individual's environment.
6. To provide Omanis with an awareness of their rights, duties, and the commitment that they should have toward their nation.

Table 2 Enrollment of Omani students by gender in general education 2005–2006

<i>Grade level</i>	<i>Gross enrolment rate</i>			<i>Net enrolment rate</i>		
	<i>Males</i>	<i>Females</i>	<i>Total</i>	<i>Males</i>	<i>Females</i>	<i>Total</i>
Grades 1–6	100.0	99.53	99.83	89.26	90.15	89.69
Grades 7–9	99.85	93.49	96.72	73.98	75.65	74.80
Grades 10–12	85.56	81.99	83.78	62.42	63.60	62.98

From Ministry of Education (2006). Education for all in the Sultanate of Oman 1970–2005. A contribution to the celebration on UNESCO's 60 year anniversary. Muscat: Oman.

7. To develop the students' ability to maximize the use of their leisure time.
8. To provide every member of society with educational opportunities.
9. To inculcate love, pride, and sacrifice for the homeland and the need to protect it and maintain its independence.

The general aims were reviewed in 1995 as a result of changes going on at the national and international levels, and reforms were introduced in the system. The policies and aims were explicit in the National Five-Year Plans (1996–2000), which highlighted the strategies and aims of developing human resources. The plan emphasized the following goals:

- Strengthening and improving science, mathematics, and English-language programs.
- Assessing the quality of the education system and schools' input, processes, and output.
- Introducing computer laboratories at all levels of schooling.
- Using computers in all subjects, especially in English, mathematics, and science.
- Providing learning resource centers (LRCs) in all schools; LRCs would also be designed for use by members of the wider community.
- Developing the inspectorate system and improving processes for the recruitment of staff.
- Application of information technology to the needs of the Ministry.
- Improving the teaching profession through:
 1. Ensuring that all 11,000 existing Omani teachers with 2 years diploma attain a 4-year degree by 2015.
 2. Establishing a pool of substitute teachers in each region to cover longer absences.

Omanization and Manpower Development

The issue of Omanization came into being with the publication of Oman's third National Development Plan (1980–1985). Omanization, according to public agreement, is a process by which jobs are taken up by efficient and competent Omani nationals to fulfill organizational needs at the governmental level.

Al-Lamki and Suleiman (1992) state that in fact all three facets of Omanization – those impacting upon the Omani government, the private sector, and the Omani citizen – resonate with improved education and administration systems. For the individual Omani, Omanization represents new potential for personal economic advancement. Within the government, it has been seen as a countermeasure to the unpopular importation of foreign labor, so necessary to the development of the Sultanate's infrastructure and even

more so, potentially harmful to the delicate fabric of the culture and national security of the Sultanate. Al-Lamki (1992) argues that there have been few instances where the implementation of Omanization in the private sector has been the result of a long-term systematic strategy for the recruitment, retention, and advancement of Omani nationals into the management structure of companies.

While in the private sector, Omanization may result in a slight (and reluctant) increase in Omani employees, the public sector has to abide by certain restrictions regarding the recruitment of expatriate staff. These are embedded in the Civil Service Regulations, which state the following (The Fourth Five-Year Development Plan, 1995):

1. An expatriate official will fill no position in the civil service if a qualified Omani citizen is available for the post.
2. The positions needed by different ministries will be clearly specified once a year.
3. More efforts should be made to achieve self-sufficiency, and at the same time to continue recruiting expatriates when Omanis are not available.
4. Personnel departments should be managed by Omanis in the private sector establishments to ensure that there is an appropriate climate for the placement of the national labor force.
5. Compensation is to be given to the private sector for all expenses incurred in training Omanis to replace expatriate labor and to fill new jobs which require a particular standard of training and skills.

The Ministry of Higher Education contributes toward the promotion of the policy and objectives by increasing the enrolment in colleges and universities, as well as, by ensuring that Omanization remains an important criterion for employment at all levels of educational institutions. Manpower training is an important part of the enhancement process and the government gives top priority to it. The volume of education and training and the resulting number of those who pass through the system depends on the financing of programs. The level of the State's contribution to education affects the number of possible institutions that can be established, the current budget, and therefore the amount of manpower available for employment.

Structure and Operation of the Education System

The educational system in Oman is centralized and controlled by the government, with the exception of prebasic education and private schools. The MoE, in cooperation with other ministries and bodies such as The Sultanate *Diwan*, the Ministry of Social Development, the Ministry of Manpower, the Ministry of Defense and Police, is responsible for education, its expansion, and supervision.

The MoE is the principal executive authority for the general education system which includes preschool education, basic education, and postbasic education. At present there are 11 educational directorates and departments within the Ministry which is distributed throughout all regions of Oman. However, the organizational structure has gone through many changes to meet the expanded educational needs in the growing regions. At present, the Ministry undertakes certain duties. It draws up educational policy within the general policy framework of the State in accordance with its principles and national aims. The Ministry has to keep pace with the needs of production and services. It makes plans and establishes programs and means of implementing policy. It also has to respond to community needs and keep abreast with modern scientific progress. Tertiary education is the portfolio of the Ministry of Higher Education, while vocational training is the charge of the Ministry of Manpower. In addition, there are other government and private institutions that offer specialized education and training in various fields.

Preschool and Early Childhood Care

This education includes kindergartens and nurseries. Most of them come under the jurisdiction of the Department of Private Education within the MoE. The number of children attending these institutions has been increasing since 1970, especially when Omani women began to participate in the labor market. Women's associations, other ministries, and the private sector contribute to develop this kind of education. It is increasingly privatized and parents pay fees to register their children in these institutions which serve about 2.5% of the total age range (3–6 years). The kindergarten curriculum is designed to promote self-learning and enhance emotional, imaginative, and esthetic development through a

range of activities with due consideration being given to the objectives and aims of education as detailed by the MoE. As a result of an increase in the number of kindergartens throughout the Sultanate of Oman, a need arose to expand supervision to include regions outside Muscat. In the academic year 2003–2004, kindergarten supervisors were trained and deployed throughout the various educational regions in the Sultanate of Oman.

Basic Education

In the academic year 1998–1999, MoE introduced basic education. The time span allotted to the process is 10 years. It is divided into two cycles: Cycle One includes grades 1–4 and Cycle Two includes grades 5–10. The MoE pointed out that the distinguishing characteristics of the new system endeavors to encourage students themselves to investigate phenomena and find answers to questions rather than teachers merely giving and explaining the answers to students. The reform plan has adopted the concept of basic education with these aims (Ministry of Education, 2001):

- Integration between theory and application, thought and work, education and life.
- Comprehensive development of the individual's whole personality.
- Acquisition of self-learning skills in the context of a lifelong education.
- Inculcating the values and practices necessary for mastery and excellence in learning.
- Meeting the need of human development in the context of comprehensive social development.

Basic education was phased in gradually in its first year in 1998, starting with 17 schools. By 2004, it had been phased in throughout the country. In general, the basic education program consists of many subjects that aim to

Table 3 Subjects and classes per week for basic education, grades 1–10

Subjects	Grades									
	1	2	3	4	5	6	7	8	9	10
Islamic studies	6	6	6	5	5	5	4	4	4	4
Arabic language	12	12	10	7	7	7	7	7	7	7
English language	5	5	5	5	5	5	5	5	5	5
Mathematics	7	7	7	7	7	7	7	7	7	7
Science	3	3	3	5	5	5	7	7	7	7
Social studies	0	0	2	3	4	4	4	4	4	4
Life skills	1	1	1	1	1	1	1	1	1	1
Information technology	1	1	1	2	2	2	2	2	2	2
Physical education	2	2	2	2	2	2	1	1	1	1
Arts	2	2	2	2	1	1	1	1	1	1
Music	1	1	1	1	1	1	1	1	1	1
Total	40	40	40	40	40	40	40	40	40	40

From Ministry of Education (2006). Education for all in the Sultanate of Oman 1970–2005. A contribution to the celebration on UNESCO's 60 year anniversary. Muscat: Oman. <http://www.moe.gov.om/portal/SiteBuilder/Sites/EPS/English/home.aspx> (accessed July 2009).

satisfy the stated objectives – these subjects are stated in **Table 3**.

Emphasis within the curriculum should be given to the basic social arts of communicating, recording, computing, and measuring, which may be translated as reading, writing, and speaking. The principles of health instruction and inculcation of good health practices are basic phases in the school program. Creative and expressive skills and activities should be part of a child's school experience. A good education should provide for the widening of the space horizon and the boarding of time perspective. Geography and history serve to acquaint learners with the world lying beyond their own immediate experience.

Postbasic Education

Postbasic education system is 2 years of schooling followed by basic education. The new system has been introduced since 2007–2008 with the following goals of education in mind:

1. Promoting loyalty to the homeland and to His Majesty the Sultan.
2. Affirmation of belonging to one's own community and to the Gulf Arab and Islamic world, and to developing awareness of global trends.
3. Reaffirming faith in the principles of the Islamic religion and creed, and the consolidation of spiritual values in the minds of learners – making it a standard of conduct.
4. Having pride in the Arabic language while developing adequate skills in a language of international communication.
5. Developing thinking skills of all kinds and positive trends toward work.
6. Using self-developed skills effectively and continuously learning.
7. Contributing positive trends in the environment.
8. Interacting with others and participating in community activities.

The aim of the school programs is to prepare students for life after school, whether this is for higher education, for further education, or for entry into the labor market. Personal choice is regarded as important in determining individual curricula at grades 11–12. A diverse range of courses relevant to the varying abilities, interests, and aspirations of students are offered. The curricular model emphasizes the learning of essential skills such as communicational skills, mathematical skills, nano skills information, problem-solving skills, and personal and social skills. The program in the postbasic education includes: research methods, life skills, geography and modern techniques, science and technology, environmental science, computer skills, communication skills, and business visual design. The essential skills are delivered by

integrating them into core-subjects-based courses – subjects taken by all students.

Career Guidance

Within the academic plan for grades 11 and 12, where a student has the freedom to choose from a diverse range of subjects, the role of career guidance, and career advisors becomes paramount. They help students to discover their abilities and how to channel them to converge with the goals and ambitions of their future careers. Students as well as their parents receive assistance in identifying and weighing the various employment and/or further education and training opportunities that will be made available to them. The National Center for Career Guidance is also established. It provides the following services: (1) a database of professions and jobs and their academic standards commensurate with the readiness and ability of each student; (2) providing information on the different functions and nature of the work in order to help a student choose his/her future career; (3) adequate information on higher education programs and conditions of enrollment; and (4) organizing of workshops and training programs for career advisors working at schools.

Other Schooling Arrangements

Special education

There are special educational institutions that cater to students with special needs, such as the disabled, deaf, and blind. Most of these schools are located in Muscat. They are coeducational and offer programs for both males and females. They prepare them for any kind of work that may help them and their society and are under the patronage of the MoE and the Oman Women's Association. According to the Ministry's statistic (2006–2007), the institutions are currently serving 770 students.

Private schools

The government has encouraged the growth of private schools and welcomes the contribution made by the private sector to the educational system. In the academic year 2006–2007, there were 170 private schools and kindergartens in the Sultanate providing preschool, general education (grades 1–12) for 17,978 boys and 14,156 girls, as well as about 8452 children between the ages of 3 and 6 years. Private school owners must be of Omani nationality. They follow the same curriculum as government schools and students sit for the same examinations. The Sultan's School at Seeb is an example of an educational institution that operates a demanding bilingual program with specialization in science in higher grades. Emphasis is placed on preparation for citizenship and leadership and the school offers a very wide range of extracurricular activities (Ministry of Information, 2008).

International schools

According to ministerial decree issued on 4 August 2006, international schools are administered by the International Board of Directors elected by the parents of students, and the Council is the body responsible for mediating all administrative matters concerning the school to the MoE. Students are admitted to schools based on admission requirements established in the states of residents (?). With the approval of the Council, schools are allowed to organize cultural celebrations, recreational programs, or other appropriate money-generating activities in order to implement specific projects and/or to provide more educational services to students. The school provides the Ministry with a copy of the annual audited accounts.

Technical and Further Education

The policy of higher education as identified in Vision 2020 is to expand the institutions and specialization of higher education in accordance with skilled labor-market requirements.

In an attempt to implement the policy to further technical education, the State has set out the following measures (Sultanate of Oman, Ministry of Development, 1997):

1. Provide advanced technical education for a period of 2 or 3 years to allow for rapid technological development, as well as to cater to labor-market requirements, that is, technicians in computer, mechanical, and electrical engineering, and building construction.
2. Provide university education and postgraduate studies with emphasis on expansion and diversification, according to market needs.

There are also some private technical and vocational institutes which offer 1-year courses after secondary level or some courses which emphasize focused and specific skills for particular industrial, commercial, and service agencies.

Higher Education

The Ministry of Development (1997) has stated that Vision 2020 for Oman's economy determines the aims of the higher education system as follows:

- To develop a high level of educational system which is flexible and upgradable.
- To increase the enrolment in Higher Education from 9% to 40% in 2020.
- To increase the enrolment of women from the age group 20–40 to 40% in order to provide equal opportunities for males and females in higher education.

- To give special consideration to students from families receiving social welfare benefits and to those from limited income categories.

Oman is attempting to attract more private investment in the country's higher education sector through incentives such as land grants and a matching financial grant for capital investment for private universities and colleges. The government also supports private higher education through the payment of fees for certain qualifying students.

The higher education sector in the Sultanate of Oman consists of a variety of institutions. These can be divided into two basic types: public institutions which are managed and fully funded by the government; and private institutions, which are managed and funded by private companies and individuals and supported by the government through the provision of facilities and the granting of subsidies.

At present, there are 14 privately owned colleges of higher education which operate under license from the Ministry of Higher Education and are at different stages of development. Three of these are designated university colleges: Caledonian College of Engineering, *Majan* College, and *Sur* University College. Most of them are affiliated to universities in the UK, USA, Australia, Jordan, Germany, and India. The existing provision for higher education is still insufficient, which means that a considerable number of Omanis go abroad for their postsecondary education.

Serving a rapidly developing economy and society, Oman's system of higher education has grown very fast over the past two decades since the establishment of Oman's premier government university – SQU in 1986 – with a current enrolment of 13,500 students. The total number of students in higher education in Oman has exceeded 50,000. The system of private higher education grew from one college in 1995 to the present 23 higher education institutions (HEIs), with an approximate number of 17,000 entrants each year.

Some international universities offer degrees through local HEIs (e.g., the University of Hull MBA), but to date, the only international university approved to operate 'independently' in the Sultanate is the Arab Open University, which is affiliated with the British Open University. Over time, the Sultanate may attract a number of prestigious universities to establish 'branch campuses' in Oman, as is already the case in some other Gulf countries. SQU (something, a sentence is missing here) with (Sohar University, University of Dhofar, University of Nizwa, and the Oman-German University of Technology). There are at present six Ministry of Higher Education (MoHE) Colleges of Applied Sciences (CAS) enrolling approximately 7000 students, with a seventh CAS soon to be added. In addition, there are six higher colleges of technology under the jurisdiction of the Ministry of Manpower with an enrolment of 6000 and 16 institutions of health science with an intake of 707 students.

Despite further development of the private higher education sector, there will still be a great need to expand provision so that justifiable expectations for further/higher education can be met. The future development of private universities may absorb more than 2000 school leavers each year. With over 50,000 school leavers seeking further or higher education or jobs, this will impact Oman's development programs.

However, the programs and specializations for these educational establishments tend to reflect a particular vision of the labor-market and market requirements in the next decades. Courses in computer sciences, engineering, commerce, and business are available at most private colleges. Most award diplomas and a few are now able to award degrees.

Adult and Nonformal Education

Literacy has been a key objective of the Government of Oman since the beginning of the Renaissance in 1970. The Government has proceeded to spread adult literacy along with the spread of education among the young. The illiteracy rate is 9.1% in the age group 15–29 years, according to statistics of the census of population and housing conducted in 2003. **Table 4** shows the illiteracy rate of two age groups. Despite the difficulties of reaching people living in desert and mountainous areas, tackling regional disparities remains one of the MoE's major commitments.

Teaching Profession

Teachers represent about 25% of the total labor force in the public sector and 3.8% of the total labor force in the private sector. There are 29 teachers for every 1000 Omanis between the ages of 15 and 64 years. This percentage is higher than that of developed countries where it is estimated that there are 24 teachers for every 1000 people in the same age category (Donn and Issan, 2007).

In light of teacher training, The MoHE was responsible for setting up six teacher training colleges in 1995 to replace the 2-year diploma-awarding intermediate colleges, but in 2005 they were phased out and transformed

to industrial colleges, all except one. At present, College of Education at SQU is the main one that is in charge of qualifying teachers as well as educational administrators.

Various alternatives for retraining have been developed. Currently, Leeds University in England is involved in a 5-year English-language program with the MoE and has been enrolling 46 applicants a year since 2000–2001 on courses at Leeds University at OR5m. The MoE is considering an affiliation with the University of Melbourne to provide information technology, science, and mathematics courses. SQU is involved in a 10-year educational administration program and has commenced enrolling 90 students per year since 2003–2004. Al Rustaq College of Education has also started its programs to upgrade their qualifications in Arabic, Islamic sciences, and social sciences.

That being the case, it has been argued that there is a clear need for continued training of teachers in Oman.

Administrative and Supervisory Structure

Education finance

The government, through the Ministry of Finance, decides on the level of spending for education every year. On behalf of the government, MoE has the responsibility for carrying out a range of functions, including strategic planning, purchasing educational services, managing the education sector assets owned by the government, and providing a range of education related services. MoE General Directorate of Financial Affairs is the body responsible for the allocation of funds both within the ministry and the educational regions. The Ministry receives no income from students since government schooling is offered free of charge for all for the first 12 years. A large share of the annual budget is allocated to education every year and even when the country suffered a significant reduction in the gross domestic product (GDP) in 1999, the government increased the budget for the education sector as seen in **Table 5**.

Performance monitoring, evaluation, and research

As part of the overall reform process, the Ministry encourages a shift in the culture of its schools. The aim is for schools to work toward the continuous improvement of the whole school. Schools should be self-critical: identifying their strengths and areas for development and taking responsibility for their development. With these objectives in mind, the Ministry has introduced the School Performance Evaluation Project in 2002–2003. The aim of the project is to establish national criteria and to introduce a comprehensive system of school self-evaluation which is complemented by an occasional thorough external evaluation. The Ministry is intended to phase in the new arrangements to all schools over a 7-year period.

The aims of assessment and evaluation systems should allow students to seek out assistance to help them address their weaknesses, gain confidence from their strengths,

Table 4 Illiteracy by age group 1993 and 2003

Age Group	1993 Census			2003 Census		
	Male (%)	Female (5)	Total (%)	Male (%)	Female (5)	Total (%)
15–29	4.5	21.0	12.8	1.5	4.2	2.8
30–49	38.8	82.2	60.5	13.5	49.2	31.8

Source: Sultanate of Oman, Ministry of National Economy, Census Department (2003). *Selected Data and Indicators of General Census: 1993–2003*. Oman: Ministry of National Economy.

Table 5 MoE budget as proportion of gross domestic product (GPD) and state budget – 1999–2005 (in Omani rials).^a

Financial year	GDP	State budget	Education budget	Edu. budget as % of GDP
1999	2.012.000.000	.000.0001.214	239.702.069	11.9
1999	1.525.000.000	.000.0001.133	248.277.476	16.28
2000	2.091.000.000	.000.0001.321	278.019.151	13.3
2001	.000.0002.495	.000.0001.400	292.422.273	11.7
2002	.000.0002.490	.000.0001.479	313.793.094	12.6
2003	.000.0002.600	.000.0001.525	343.986.652	13.2
2004	2.925.000.000	1.587.000.000	389.345.522	13.3
2005	3.140.000.000	1.735.000.000	448.983.429	14.3

^aone OR = three US\$.

From Ministry of Education, General Directorate of Financial Affairs (2005). A Financial Report 1999–2005. Oman: Ministry of Education (unpublished data).

and provide them with opportunities for further challenges and demands. To help achieve these aims, there has also been a move away from the expectation that students should merely rote learn and memorize information. Instead, an increased emphasis has been placed on the testing of higher thinking skills. A wider range of assessment instruments have now been introduced, with continuous classroom assessment being given added prominence in the assessment and evaluation system. Arising out of ordinary classroom activities, teachers are now expected to use a variety of continuous assessment instruments such as short written or oral tests, quizzes, performance assessment tasks, projects, and student self-assessments.

Major Changes and Issues since the 1990s

The vast influx of foreigners into this formerly secluded region emphasized the need to catch up. Oman's government realized that there was a need to provide a well-educated, resourceful local workforce for the future, displacing the continual need for expatriates to undertake even basic maintenance of state utilities. International exposure and cross-cultural interaction is essential if Oman is to attain high standards and remain relevant in today's world. However, Oman witnessed a spectacular educational renaissance in the last three decades of the twentieth century. Spreading education in Oman was an uphill task in view of sparse population over a vast geographical area with difficult terrain. Oman managed to reduce illiteracy ratio from 41% in 1993 to 24.4% in 2000. However, this ratio has been decreasing during the last 8 years of the third millennium. Oman made good progress in merging females in various levels and therefore bridging a sustainable part of the gender gap in the field of education, an indication of good progress in the field of human development. Public spending on education as a percentage of total government expenditure increased from 9.5% in 1991 to 14% in 2005. This put Oman midway in comparison to other Gulf Cooperation Council (GCC)

countries, yet illiteracy and gender issues are challenges to be tackled at all levels and kinds of education.

Bibliography

- Al-Lamki, A. and Suleiman, M. (1992). *Higher Education and Under-Employment in Oman*. Sultanate of Oman: Sultan Qaboos University.
- Al-Shanfari, A. M. S. (1991). *The Development of Government Policy in Education and Vocational Training in the Sultanate of Oman 1970–1990*. PhD Thesis, University of Wales.
- Donn, G. and Issan, S. (2007). Higher education in transition: Gender and change in the Sultanate of Oman. *Scottish Educational Review* 39(2), 173–185.
- Karabenick, S. A. and Moosa, S. (2005). Culture and personal epistemology: US and middle eastern students' beliefs about scientific knowledge and knowing. *Social Psychology of Education* 8(4), 375–393.
- Ministry of Education (2007). *Educational Statistics Year Book. Issue Thirty Seven*. Muscat: Mazoon Publishing Company.
- Ministry of Information (1997). *The White Book – The Basic Law of the Sultanate Of Oman (24 Jumada al Akhira 1417 AH)*. Royal Decree No. 101/96. Muscat: Sultanate of Oman.
- Ministry of National Economy (2003). *Oman – Human Development Report*. Oman: Ministry of National Economy.
- Sultanate of Oman, Ministry of National Economy (1991). *The Fourth Five Years Development Plan*. Development Council General Secretariat, RPM LTD. Oman: Ministry of Information.
- Sultanate of Oman, Ministry of National Economy (1996). *The Fifth Five Years Developments Plan, Development: 1996–2000*. Development Council General Secretariat, RPM LTD. Oman: Ministry of National Economy.
- Sultanate of Oman, Ministry of Education (2001). *Basic Education in The Sultanate of Oman – The Theoretical Framework*. Muscat: Oman Printers & stationers.
- Sultanate of Oman, Ministry of Information (2008). *Annual Book 2007–2008*. Oman: Ministry of Information.

Further Reading

- Al Barwani, T. (2007). Women, Education and the Redefinition of Empowerment and Change in a Traditional Society: The Case of Oman (unpublished Paper).
- Al Farsi, F. N. J. (1994). *Omanisation and Staff Development of Academic Staff in Sultan Qaboos University*. PhD Thesis, University of Exeter.
- Al-Lamki, S. M. (1999). Paradigm shift: A perspective on Omani women in management in the Sultanate of Oman. *Advancing Women in Leadership* 5, 1–18.

- Al-Lamki, S. M. (2002). Higher education in the Sultanate of Oman: The challenge of access, equity and privatization. *Journal of Higher Education Policy and Management* 24(1), 75–86.
- Al-Lamky, A. M. S. (1992). Higher Education and Underdevelopment of Oman: Perceptions of University Graduates in the Context of Dependent Development 1970–1990. The George Washington University.
- Al-Talib, K. (1990). Modern Oman: Economic Diversification and Omaniisation of the Labor Force (MSc. University Of Nevada, 1990).
- CS1>UNDP (2008). Human Development Report 2007/2008-Country Fact – Sheet – Senegal, Oman The Human Development Index – Going Beyond Income, www.drstats.undp.org/countries/country_fact_sheets/cty_fs_SEN.html-30k (accessed July 2009).
- Ministry of Education (2006). Education for all in the Sultanate of Oman 1970–2005 A contribution to the celebration on UNESCO's 60 year anniversary. Muscat: Oman. <http://www.moe.gov.om/portal/SiteBuilder/Sites/EPS/English/home.aspx> (accessed July 2009).
- Ministry of Education and UNICEF Muscat (1999) EFA: Year 2000 Assessment Country Report.
- Pridham, B. R. (ed.) (1987). *Oman: Economic, Social and Strategic Development*. London: Croom Helm For The Center Of Arab Gulf Studies, University Of Exeter.
- Rassekh, S. (2004). *Case Study on Gender-Sensitive Educational Policy and Practice in the Sultanate of Oman*. Oman: Ministry of Education.
- Rassekh, S. (2004). *Education as a Motor for Development: Recent Education Reforms in Oman with Particular Reference to the Status of Women and Girls*. Geneva: International Bureau of Education, UNESCO.
- Sultanate of Oman, Ministry of National Economy (2003). Oman – *Human Development Report*. Oman: Ministry of National Economy.
- Sultanate of Oman, Sultan Qaboos University (2005). Academic Subject Review of the B. Ed in Teaching English Programme Self – Evaluation Document, December 2005.
- Sultanate of Oman, Sultan Qaboos University, College of Engineering (2006). An Institutional Profile – Appendix 2. (unpublished Report).
- The Ministry of National Economy (1995). *The Fifth Five-Year Development Plan (1995–2000)*. Development Council General Secretariat, RPM LTD. Oman: Ministry of National Economy.
- The Ministry of National Economy (2001). *The Sixth Five-Year Development Plan (2001–2005)*. Oman: Ministry of National Economy.
- The Ministry of National Economy (2006). *The Seventh Five-Year Development Plan (2006–2010)*. Oman: Ministry of National Economy.
- UNESCO (2005). *EFA Global Monitoring Report: Education for All the Quality Imperative*. Paris: UNESCO.
- UNICEF/Government of Oman (2005). *Situation Analysis of Children and Women in the Sultanate of Oman*. Muscat: Oman.

Relevant Websites

- <http://www.abegs.org> – Arab Bureau of Education for the Gulf States.
- <http://www.ita.gov.om> – Information Technology Authority (ITA) Portal.
- <http://www.kom.om> – Knowledge Oasis Muscat.
- <http://www.mocs.gov.om> – Ministry of Civil Services.
- <http://www.moe.gov.om> – Ministry of Education, Sultanate of Oman.
- <http://www.mohe.gov.om> – Ministry Of Higher Education.
- <http://www.omanet.om> – Ministry of Information, Sultanate of Oman.
- <http://www.manpower.gov.om> – Ministry of Manpower, Sultanate of Oman.
- <http://www.mone.gov.om> – Ministry of National Economy.
- <http://www.oac.gov.om> – Oman Accreditation Council.
- <http://www.onc.gov.om> – Oman National Commission for Education Culture and Science.
- <http://www.omannews.gov.om> – Oman News Agency.
- <http://www.omanscouts.gov.om> – Oman Scouts and Guides.
- <http://82.178.29.32> – Sultanate of Oman educational portal.
- <http://www.squ.edu.om> – Sultan Qaboos University.
- <http://www.rgotc.cjb.net> – The Royal Guard of Oman Technical College.

Pakistan

N Mahmood and R Malik, University of the Punjab, Lahore, Pakistan

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Assistant education officer (AEO) – AEO is responsible for day to day administration of schools in the geographical territory of tehsil (an administrative division under the district).

Dars-e-Nizami – Curriculum based mostly on religious teachings offered by *Deeni Madaris*. This includes subjects on applied and natural sciences.

Darul Aloom – Literal meaning is the abode of knowledge. Offers a 2-year certificate course equivalent to upper-secondary level, after 10 years of education offered by *Deeni Madaris*.

Deeni Madaris – (Singular *madrassah*) Offers 10 years of education (equivalent to secondary level) based mostly on religious teachings.

Directorate of Staff Development (DSD) – A provincial-level institution responsible for continuous professional development of public school teachers.

District education officer (DEO) – Head of the Education Department at the district level. Responsible for planning, budgeting, and coordinating all educational activities in the district.

Executive district officer (EDO) – Head of offices evolved at district level through local government ordinance.

Fiqh – Islamic jurisprudence based on various Islamic schools of thoughts.

Hadees – Sayings/texts compiling the life and traditions of the Holy Prophet (peace be upon Him (PBUH)).

Higher Education Commission (HEC) – Established in 2001 by the Government of Pakistan with a mission to raise the quality of education in the higher educational institutions and further, aimed at capacity building of faculty, developing research culture in the universities, and providing technical and financial assistance.

Iftah – Study of giving legal options (*fatwa*).

Katchi class – Also known as early childhood education. It is equivalent to nursery education.

Local Government Ordinance, 2001 – Also known as Devolution Plan. System of governance decentralizing the government's administrative and financial authority from provinces to the district and

union councils, ensuring maximum participation of the civil society.

National Education Assessment System (NEAS) – The program aimed at developing national capacity for monitoring the learning achievements of elementary-level students in order to improve the equality of services (curriculum, textual material, teachers' delivery, policy formulation, etc.) in the education sector. The basic objective of NEAS is to establish a system of student assessment in the Ministry of Education and to develop national capacity for conducting assessments periodically to monitor student achievement.

National Vocational and Technical Education Commission (NAVTEC) – The Commission is mandated to facilitate, regulate, and provide policy direction for technical education and vocational training to meet national and international demands for skilled manpower. NAVTEC's energies are also dedicated to develop national occupational skills standards, curricula, and trade testing certification systems for all sectors where technical education and vocational training are imparted.

Parent teacher association (PTA) – It is an association established by parents and teachers for resource generation for the uplift of the schools. It is based on equal partnership with the government.

Provincial Institute for Teacher Education (PITE) – A provincial-level institution responsible for imparting pre-service training, development of innovative learning material for in-service training, training of master trainers to work as trainers at local levels, and impart training through distance education for in-service teachers.

Tafseer – In-depth exegesis of the Quran.

Technical Education and Vocational Training Authority (TEVTA) – Aimed at enhancing global competitiveness, through a quality and productive workforce by developing demand-driven, standardized, dynamic, and integrated technical education and vocational-training service.

Village Education Committees/School Management Committees (SMCs) – Representatives from school administration and political leadership that works for the welfare of schools.

General Background

The roots of the Pakistani nation can be traced back both to the great Muslim civilization which flourished between the seventh and tenth centuries in Asia and Africa and to the Indus Valley civilization, which prevailed 5000 years ago. The excavations of Harappa, Mohenjo-Daro, and Soan Valley have provided evidence of the highly organized civilizations which existed around the Indus River. The Indus Valley civilization is one of the oldest civilizations in the world and it achieved a high level of sophistication at an early stage (3000 BC) and flourished for many centuries. The ruins at Mohenjo-Daro (Sindh, Pakistan) revealed that people living there had a well-organized social and economic structure. It was a very well planned city with proper sanitation, general civic facilities, metric system of measurement, and many other aspects indicating that the people of the Indus Valley civilization were well ahead of other civilizations which existed at that time, both in technological crafts and the sociolegal system. The people of that time sailed the Arabian Sea and the Persian Gulf in wooden boats and traded in gems, animal hides, and fabrics. In the fourth century BC, Alexander and his armies passed through this region, and in the sixth century BC, the Persian Empire was extended to the Indus Valley. The Indo-Aryans, who predominate in contemporary Pakistan, arrived between about 1500 and 1200 BC. Later on, Islam was introduced in this region during the seventh century. Art, literature, music, and architecture flourished during this period. This region, therefore, has been a melting pot for diverse races and cultures.

Demographic Characteristics

Pakistan obtained independence in 1947 comprising two regions: West Pakistan in the Indus River Basin and East Pakistan, located more than 1000 miles away in the Ganges River Delta. East Pakistan, now Bangladesh, broke away from West Pakistan in 1971.

Pakistan, constituted on 796 095 km², has four provinces – Balochistan, North Western Frontier Province (NWFP), Punjab, and Sindh. In addition to these provinces, Pakistan also has three areas which include Federal Administered Tribal Areas (FATA), Federal Administered Northern Areas (FANA), and Azad and Jammu Kashmir (AJK). The territory of Pakistan encompasses portions of the Himalayas, Hindu Kush, and Karakoram mountain ranges, and is home to the world's highest mountains. Pakistan shares borders with Afghanistan (2430 km), China (523 km), India (2912 km), and Iran (909 km). Its coastline totals 1064 km along the Arabian Sea.

Pakistan is the world's sixth largest country. Its population is estimated at 160 million; of this, 50% or 77 million is below the age of 18 years. Pakistan in other

words, has one of the youngest populations in the world. The rapid population growth is a heavy burden on the country's limited resources. A majority of the population (65.1%) resides in rural areas largely depending on agriculture. In spite of a significant decline in the level of fertility in recent years, Pakistan's population is still growing at the rate of close to 2% a year. The life expectancy for the year 2006–07 was estimated at 64 years for males and 66 years for females. The crude death rate (CDR) was 7.8 (per 1000) in 1999–2000, declining to 6.5 (per 1000) by the year 2006–07.

Pakistan's economy entered the fourth consecutive year of above 7.0% annual growth rate in 2007, and its poverty headcount fell from one-third to less than one-fourth of the population. In the last fiscal budget (2007–08), a sum of \$8.3 billion was allocated for social-sector development with special emphasis on poverty reduction. The average per capita income was \$680 (US), and the labor-force participation rates were 36.7% for males and 9.3% for females in 2006–07. Pakistan is predominantly an agricultural country with major exports in cotton, yarn, and cotton fabrics.

Despite these developments, however, the economic and social development indicators of Pakistan are still among the lowest not only in the Asia-Pacific region, but also in the world. Pakistan stands at the 136th position in terms of the Human Development Index and is at the 126th position out of 128 countries in the Global Gender Gap Index. Pakistan has scored 42 points and is placed among the ten least equitable countries of the world.

Goals of the Education System

Providing of education is the constitutional responsibility of the state. Based on government policies and target plans, the following are the main goals of education:

- To achieve universal primary and secondary education, following the targets set in Millennium Development Goals (MDGs) and Education for All (EFA). The aim is to eliminate gender disparities in primary and secondary education by 2015, with a focus on ensuring girls have full and equal access to and achievement in basic education of good quality.
- To expand and improve comprehensive early childhood care and education, especially for the most vulnerable disadvantaged children.
- To strengthen technical education at secondary and higher-secondary levels, establish polytechnics at district and vocational institutions at the *tehsil* level, and to introduce more diversified courses in technical and vocational education.
- To follow the gender-based policy in education specially focusing on providing free education to females

up to secondary level (grade 10), improving nutritional status of school-going girls through free school meals and regular health checkup, giving stipends and scholarships to needy students, and providing free uniforms and school bags.

- To translate the dream of a knowledge-based economy into reality by providing (good-quality, merit-oriented, equitable, and efficient higher education) which is the most crucial instruments; to focus on faculty development, developing industrial linkages, and conducting need-based research.

Education Administration

After the introduction of Local Government Ordinance in 2001, the education system has been devaluated. The Federal Ministry of Education is a coordinating body whereas the provincial education departments serve as a bridge between the federal and district governments. The district governments are fully empowered to prepare their own educational development and fiscal plans. However, the formulation of national educational guidelines for curricula, examination, evaluation, standard of education, learning/competency level, and teacher training is the responsibility of the federal government. The provincial education departments, headed by their respective provincial education ministers work in line with the Federal Ministry for the implementation of national policies and plans. In the district, the head of the Education Department is the executive district officer – education (EDO, education). Next in the hierarchy is the district education officer (DEO) followed by the assistant education officers (AEOs). The AEOs at the district level are responsible for providing academic guidance, as well as supervision of the schools. At the grass-roots level, that is, the union council level, village education committees/school management committees (SMCs), and parent teacher associations (PTAs) have been set up to coordinate with schools for their smooth functioning.

Structure and Operation of the Education System

Figure 1 shows the formal structure of education. The statistics on enrolments and resources in the education sector are provided in **Table 1** and **Figure 2**.

Currently, there are 122 349 primary, 38 449 middle, and 25 090 secondary schools in the country. Almost 37% of the total students enrolled are in the primary level, 20% in the middle, while 28% are in the secondary level. The overall teacher–student ratio at the primary level is 1:31, 1:21 at the middle level, and 1:23 at the secondary level in the country but it may vary by urban/rural and

public/private scenarios. At the primary level in rural areas, for example, there are 34 students per teacher while in urban areas the ratio is 1:24.

There are 12 153 *Deeni Madaris* in the country. Around 55% are located in the rural areas and around 97% are privately managed.

Literacy

The literacy levels in Pakistan have improved over time although at a moderate pace. The overall literacy rate (10 years and above) was 45% in 2001, which has increased to 54% in 2006–07, indicating a 9.0 percentage-point increase over a period of only 5 years. Male literacy rate (10 years and above) increased from 58% in 2001 to 65% in 2006–07 while it increased from 32% to 42% for females during the same period. The rate of increase in literacy for females is faster as compared to the males.

Multiple reasons for low literacy include social taboos, poverty, child labor, illiteracy among parents/families, poor infrastructure and educational facilities, constraints on girls' education, high dropout rate, etc.

Early Childhood Education

Until recently, early childhood education had not been formally recognized in public schools in Pakistan. The traditional *katchi* class, to some extent, serves the purpose of early childhood education. In the private sector, however, focus on early childhood education has been there for a long time. There are regular Montessori and kindergarten private schools. The Government of Pakistan, recognizing the importance of early childhood education, has recently included it as a component in Education Sector Reforms and provided funding to the provincial and district governments for the purpose. The gross enrolment rate (GER) for early childhood education rose from 36% in 2001–02 to 91% in 2006–07, for children aged 3–4 years.

Primary Education

Primary education is free and compulsory in Pakistan. The universalization of primary education (grades 1–5) has been recognized by the founding father of Pakistan, Muhammad Ali Jinnah, and Universal Primary Education (UPE) was established as a goal at the first National Education Conference in 1948, soon after independence. Over the years, the goal of UPE has been the of all the educational policies prioritization and plans, but despite this concentration, however, there are still many children between 5 and 9 years of age who are not attending school. Out of the total enrolment of students from primary to post-secondary level, almost 37% are enrolled at only primary level of education. During recent years, there has been considerable progress in the GER which reached

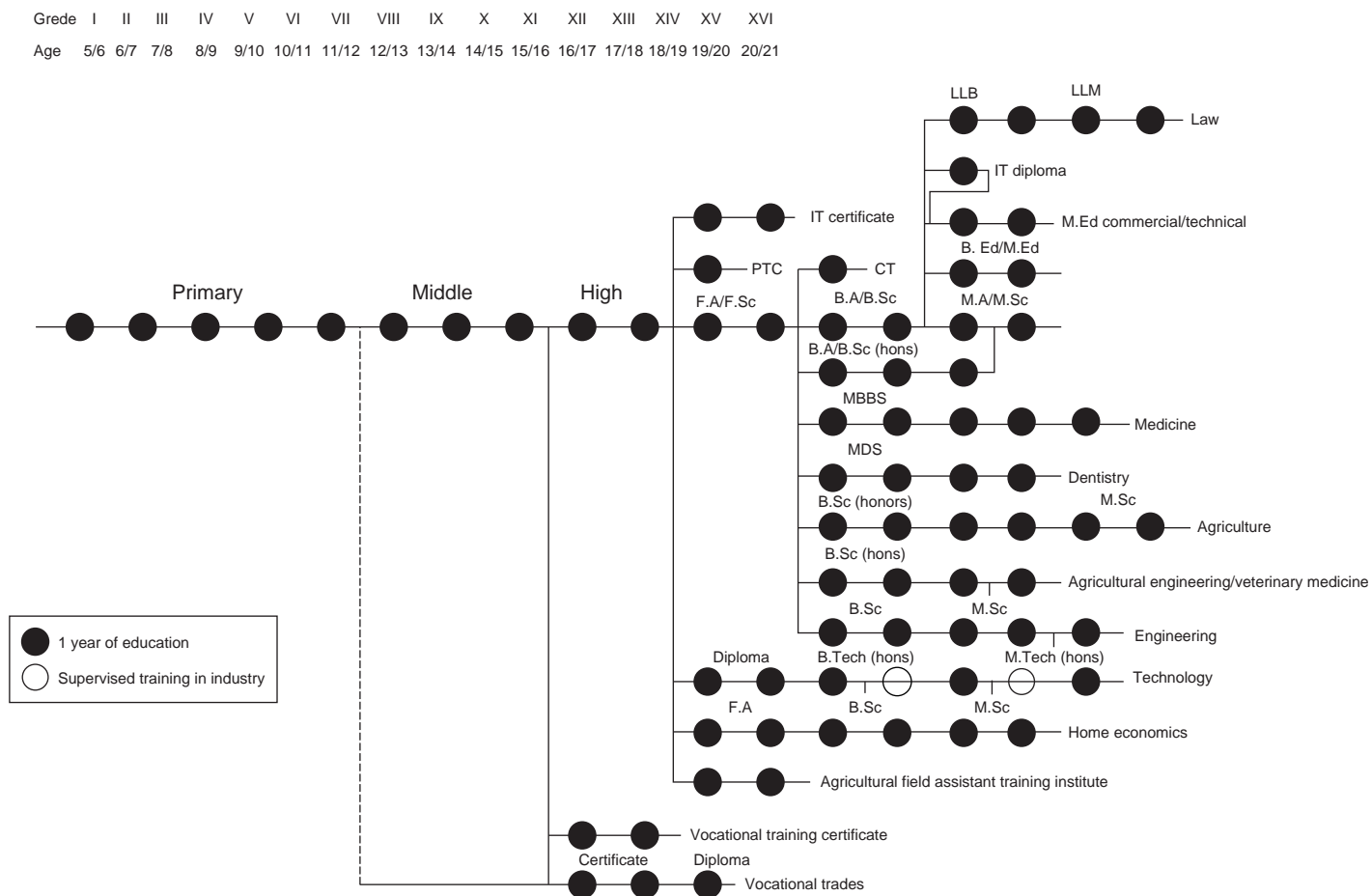
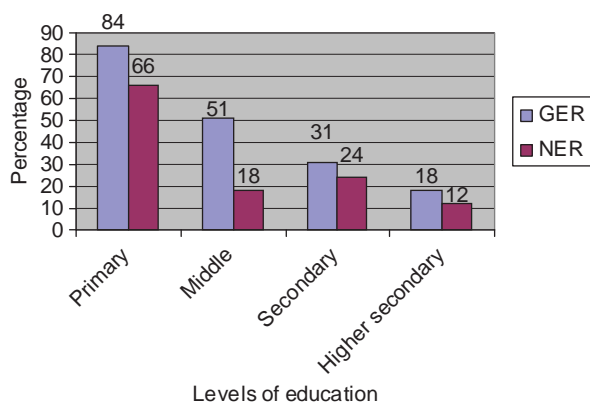


Figure 1 Structure of Education System in Pakistan.

Table 1 Number of institutions, enrolments, and teachers by level of education

Levels of education	Institutions	Enrolment	Teachers
Pre-primary	1081	61 364	3807
Mosque school	14 123	802 904	22 603
Primary	122 349	12 433 240	399 517
Middle	38 449	6 652 870	313 797
Secondary	25 090	9 473 525	418 376
Higher secondary	1882	1 158 589	59 097
Universities and degree-awarding institutions	99	419 189	11 434
Technical/professional	1324	361 534	30 334
Vocational/polytechnics	3059	238 687	15 339
Madrassahs	12 153	1 549 242	58 391
Others	8232	452 752	24 107
Total	227 841	33 603 896	1 356 802

From: Education Census, 2006.

**Figure 2** Gross and net enrolment rates by level of education, Pakistan 2007.

84% in 2006–07 up from 71% in 2001–02. Similarly, the net enrolment rate (NER) also rose from 57% in 2001–02 to 66% in 2006–07. The dropout rate at the primary level has also been controlled through the government's recent education reforms. It fell from 43% in 2001–02 to 28% in 2006–07. Despite progress, however, access rates remain low and Pakistan still faces the risk of defaulting on EFA and MDGs 2015 targets.

Middle-Level Education

The gross and net enrolment rates at the middle level (grades 6–8) have also shown improvement in recent years. The GER for the middle level was 51% in 2006–07 compared to 49% in 2005–06. Female and male GERs have increased slightly over the period, from 42% to 44% and 55% to 57%, respectively, between 2005 and 2007. There are, however, wide urban–rural disparities,

with gross middle enrolment in urban areas at 69%, and 43% in rural areas. The net enrolments are much lower than GERs. The GERs and NERs have been calculated using an age range of 10–12 years and because of large numbers of overage children enrolled at the middle level, the NER stands out as low as 18% in 2006–07.

Secondary and Higher-Secondary Education

Secondary and higher-secondary education is for 2 years each. The statistical figures show improvement in the participation at the secondary-school level in the recent past. The GER and NER had increased to 31% and 24% in 2006–07 in comparison to 31% and 24% in 2001–02, respectively. Enrolment in higher secondary education, which was 2.6% for 2001–02, rose to 3.7% in 2005–06. Secondary education is the first stage where students have to make a choice of subjects according to their interest and capabilities. The diversity offered to students is very limited and the students have to choose one of the two areas of specialization offered at the secondary level, that is, humanities and science. The humanities group consists of general science, civics, and mathematics; while the science-group students study chemistry, physics, mathematics, and biology. The students have to study compulsory subjects in addition to the area of specialization they choose. As a continuation of the area selected in secondary education, in higher-secondary school, the students take premedical, pre-engineering, or specialization in humanities corresponding to their area of specialization in secondary school. The science students are allowed to take any of the above-mentioned areas of specialization while students in humanities group can only study specializations in humanities in higher-secondary school.

Higher Education

Higher education is mainly provided by the universities and postgraduate colleges. There are more than 100 public and private universities in Pakistan. The enrolment in higher education is 0.23 million with males 0.15 million and females 0.08 million, respectively. The gross participation rate is around 4%. Higher education has received a big boost by the allocation of financial resources and improvement in the quality of education in recent years. The universities are being upgraded through a vigorous program of faculty development, scholarships and stipends to needy students, curriculum revision, equipping laboratories and libraries, connectivity to pooled and shared resources, and emphasis on research through the Higher Education Commission. Annual budget allocation has been raised 10 times from \$15 million to \$150 million, while for science and technology it has increased from \$3 million to \$100 million in 2006–07.

Technical and Vocational Education

In recent years, considerable efforts have been made to integrate skill development with the general stream of education. National Vocational and Technical Education Commission (NAVTEC) has been established to manage national policy and planning, curriculum development, standardization of technical education, training of trainees, national accreditation of private polytechnics and vocational-training institutes, and developing strong linkages with the industries. The policy thus formulated is implemented through Technical Education and Vocational Training Authority (TEVTA) in the provinces. At present, there are 1324 technical/professional colleges and 3059 vocational/polytechnic institutes with an enrolment of 361 534 and 238 687 students, respectively, in the country. Around 46 000 teachers are attached to these institutions. Emerging technologies like telecommunication, computer electronics, automation, petroleum, garments, food preservation, printing and graphics, textile, mining, sugar technology, tourism and hotel management, livestock and dairy sector, construction sector, women's participation in economy, etc., are given due importance in courses offered at the technical and vocational institutes.

Adult and Nonformal Education

The nonformal education system includes nonformal basic education (NFBE) schools and adult literacy centers. There are 10 000 NFBE schools with an enrolment of more than 300 000 students. The adult literacy centers cater to the needs of 15+ age-group people. There are 70 000 adult literacy centers having an enrolment of 1.7 million. These centers mainly focus on reading, writing, and numeracy skills.

There are approximately 56 million illiterate adults against the 66 million literate population. Financial allocations in the past for adult literacy have remained severely inadequate. At present, 2 billion Pakistani rupees (approximately \$29 million) per annum are allocated for literacy and nonformal education programs, whereas to meet the EFA targets about 15 billion Pakistani rupees per year are required.

Private Sector in Education

The private sector's role has been expanding in the country. Between 1983 and 2000, the number of private primary and secondary schools increased almost tenfold from 3300 to 32 000, much faster than the population of school-aged children. At present, private educational institutions enrol 3% of students studying basic education (pre-primary through higher secondary). In urban areas, private schools account for more students (51%) than the public sector (49%). However, the situation is reversed in rural areas, where over 80% of students are attending

public schools. Private education has the maximum share at the pre-primary level, where 42% of pre-primary students are enrolled. At the primary stage, 32%; secondary, 30%; higher secondary, 18%; technical/vocational, 52%; vocational/polytechnics, 57%; NFBE, 61%; and *Deeni Madaris*, 97% are enrolled.

A private sector is growing substantially in recent years and the number of private universities has risen up to 57 as compared to 67 in the public sector.

Religious Education

The *madrassab* system has been a part of the education system in Pakistan for long and runs parallel to the general stream of education discussed before. The *madrassab* has remained a popular system of education for the poor masses of the country because it offers free education along with free boarding and lodging facility to students. The age of students in *madaris* typically runs from 5 through 16 years. For study beyond 10 years offered by *madaris*, for grades 11 and 12, one would attend a *darul uloom* (literally, the abode of knowledge). The *darul uloom*, is equivalent to the upper-secondary school in the general stream of education. For study beyond *darul uloom*, one would attend a *jamia*, equivalent to college or university.

Estimates of *madaris'* enrolment range from about half a million to more than 2 million. There are about 4 775 000 enrolled *madrassab* students which is about 4.6% of the total enrolment in the country.

There are five boards (*wiqafhas*) running *madaris* according to their respective schools of Islamic thought: Ahl-i-Hadith, Bareilvi, Deobandi, Jamaat-i-Islami, and Shia. The boards approve the curriculum, conduct examinations, and issue graduation certificates and diplomas. There are around 10 000 *madaris* registered with these five boards. Roughly 70% are Deobandi, 16% are Bareilvi, 5% are Jamaat-i-Islami, 4% Ahl-i-Hadith, and 3% Shia.

Recently the government of Pakistan introduced *madaris* reforms. The Pakistan Madrassah Education Board Ordinance in 2001 established the Pakistan Madrassah Education Board. The Board has the responsibility of establishing new, exemplary *deeni madaris* and *darul ulooms*. The model *Deeni Madarssab* is meant to be semiautonomous, providing an exemplary model to existing *madaris* about how to modernize and train a new generation of liberal-minded *ulema* (religious scholars). In another effort to reform *madaris*, the government has issued an ordinance referred to as Madaris Regulation and Control Ordinance (2002). It demands that all *deeni madaris* and *darul ulooms* get registered with the government and submit regular financial statements. The *deeni madaris* and *darul ulooms* registered with the board would receive scholarships for their students, while the others that do not comply would be closed.

The *madaris* teach either in full or in part a specialized curriculum called *Dars-e-Nizami*. It covers about

20 subjects mainly divided into two categories: *al-uloom an naqliya* (transmitted sciences) and *al-uloom-al aqliya* (rational sciences). Most of the subjects included in the curriculum are religious in nature. Others include medicine, mathematics, astronomy, history, philosophy, poetry, and polemics. Some of the *madaris* also offer postgraduate specialization courses in *tafseer*, *badees*, *iftab*, and *fiqb*. The students can take these courses after completing *Dars-e-Nizami* and it takes at least 2 years to complete. The *madaris* offer degrees/certificates of different durations equivalent to the mainstream education, but the Government of Pakistan accepts only the highest degree (*Shahadatul Alamiya*) (MA in Arabic and Islamic Studies (Institute of Policy Studies, 2002)) offered by *madaris* after 16 years of education.

Almost all the *deeni madaris* (97%) imparting religious/ Islamic education are run by the private sector since the independence of Pakistan in 1947. A *madrassah* generally enrolls students for grades 1 through 10. Most of the *madaris* provide free boarding and lodging facilities to the students.

Curriculum

The Curriculum Wing established at the Federal Ministry of Education works closely with the provisional curriculum bureaus and textbook boards in all provinces for the development of curriculum. A National Curriculum Review Committee, represented by national and provincial experts, gives approval for the textbooks to be published by the provincial textbook boards which develop textbooks up to grade 12. The textbook boards are encouraging use of multiple textbooks as part of the national policy in recent years. The public schools are bound to follow the national curriculum. However, the private schools are free to choose curriculum in line with their aims and objectives. A reasonable number of private schools are affiliated with examination boards/bodies outside Pakistan (mainly British and American) and thus follow their curriculum to prepare students for their examination as external candidates.

Medium of Instruction

The medium of instruction in public schools (grades 1–12) is Urdu. However, at the primary level, medium of instruction could also be the provincial/local language which is different for all four provinces. In contrast to this, majority of the private schools follow English as the medium of instruction.

Teaching Profession

In Pakistan, 270 teacher education/training institutes exist of which 227 are run by the government sector and 53 are operated by the private sector. They offer pre-service and

in-service teachers' training programs. There are two kinds of programs for elementary school teachers (grades 1–8): the Primary Teaching Certificate (PTC), and the Certificate of Teaching (CT). The PTC program is meant for teachers who teach in grades 1–5. The duration of this program is 1 academic year (48 weeks). Admission to the PTC program requires that the applicant should possess Secondary School Certificate (10 years schooling). The CT program prepares teachers to teach all subjects up to grade 8, including English. The duration of this program, too, is 1 academic year. The condition for admission to the CT program is a Fellow of Arts (FA) or Fellow of Science (FSc) certificate, that is, 12 years of schooling.

The institutions preparing secondary-school teachers are known as colleges of education, affiliated with universities and those offering advanced education leading to an MA in Education, MEd, and PhD degrees are either University of Education or institutes of education and research which are part of a general university or departments of education in universities (see Table 2).

At the time of independence, there was only one college of education for men and one for women in the area now called Pakistan. There was no Department of Education or institute offering master's degrees in education. The first such institute, the Institute of Education and Research (IER) was established in 1960 at the University of the Punjab, Lahore, in collaboration with Indiana University (USA), following the recommendation by the National Commission on Education. Other IERs or departments of education were subsequently established. In 2002 the first University of Education was also established in Punjab. At present there is one University of Education, three institutes of education and research, and several colleges of teacher education attached to these universities.

Two types of programs are offered for the preparation of secondary-school teachers, requiring differences in the years of general education and professional training: (1) 1-year Bachelor of Education (BEd) program (14+1 model); and (2) 3-year Bachelor of Science Education (BSEd) program (12+3 model). For the 1-year program,

Table 2 Percentage of teachers (public) with professional qualifications

Professional qualification	Total	Female
PTC	37.3	41.5
CT	12.3	12.8
BEd	17.4	18.0
MEd	5.1	4.0
Other trained	11.1	10.1
Untrained	1.1	1.0
No information	15.6	12.3

PTC – Primary Teaching Certificate; CT – Certificate of Teaching; BEd – Bachelor of Education; MEd – Master of Education.
From: Shami *et al.*, National education 2005.

the minimum qualification required for admission is a BA/ BSc degree. In the 12+3 model, the minimum qualification required for admission is the FSc. The academic and professional qualifications required for the recruitment of teachers at various levels are shown in **Table 3**.

Each province has different departments for the training of in-service teachers. In Punjab, Directorate of Staff Development (DSD), Provincial Institute for Teacher Education (PITE), and Government Colleges for Teacher Education (GCET) are responsible for professional development of teachers. Bureau of Curriculum (BoC) and PITE are given the responsibility of in-service teachers' professional development in Sind and Balochistan while Directorate of Curriculum and Teacher Education (DCTE) and PITE are providing in-service teacher training in NWFP. The number of teachers in each province by type of school is given in **Table 4**.

Accreditation Council for Teacher Education has been established. The council is currently working on developing the national standard for teacher certification and accreditation.

Table 3 Academic and professional qualifications of teachers at different levels

<i>Level of teaching</i>	<i>Type of examination</i>	<i>Academic qualification</i>	<i>Professional preparation</i>
Primary	Grade 5	Matric (secondary)	PTC
Middle	Grade 8	FA/FSc (higher secondary)	CT
Secondary (High)	Matric	BA/BSc (except Punjab) MA/MSc (only Punjab)	BEd/BSEd
Intermediate	FA/FSc	MA/MSc	—
Degree colleges	BA/BSc	MA/MSc	—
University	MA/MSc	MA/MSc/PhD	—

Examination System

A system of automatic promotion up to grade 5 has recently been introduced in public schools. However, at the end of the fifth year of the primary stage, an examination is held by the District Education Department/Examination Commission, which has to be passed to get admission in the next grade (grade 6). Annual examinations are conducted by the respective schools for grades 6 and 7. At the end of middle school (grade 8), external examinations are taken by the respective district education departments/examination commissions. There are 18 boards of intermediate and secondary education (BISEs) in the country including one Federal Board of Intermediate and Secondary Education (FBISE) and one private-sector BISE, that is, Agha Khan University Examination Board (AKU-EB). These boards are responsible for holding examinations for Secondary School Certificate (after 10 years of education) and Higher Secondary School Certificate (at the completion of 12 years of education). Each such board has its defined geographical jurisdiction and all students enrolled in schools/higher-secondary schools in those areas have to appear for examinations conducted by their respective boards except FBISE and AKU-EB. The territorial jurisdiction of FBISE includes schools from all provinces of Pakistan and Pakistani schools in other countries while AKU-EB is allowed to register schools from all areas of Pakistan. The concept of an examination board in private sector is very recent and is meant to create competition among BISEs in government and private sectors for the improvement of the quality of examination in the country. The certificates issued by these boards are considered equivalent for all purposes. Inter Board Committee of Chairmen (IBCC) is responsible for determining the equivalence of certificates offered by these boards. The universities are responsible for conducting examinations at the territory level mainly through semester system.

Table 4 Number of teachers in each province by type of school

<i>Province/district</i>	<i>Teachers</i>				
	<i>Overall</i>	<i>Primary schools</i>	<i>Middle schools</i>	<i>High schools</i>	<i>Higher-secondary schools</i>
Pakistan	629 674	314 712	109 327	162 430	23 849
Punjab	298 665	126 964	67 042	85 406	10 338
Sindh	139 959	95 331	9 492	28 704	5 255
NWFP	97 173	57 292	12 199	18 992	5 310
Balochistan	40 551	16 150	8 573	12 107	—
AJK	23 594	4 334	6 810	9 550	978
FANA	4 942	2 232	1 038	1 672	—
FATA	18 442	10 547	3 548	3 941	165

AJK – Azad and Jammu Kashmir; FANA – Federal Administered Northern Areas; FATA – Federal Administered Tribal Areas; NWFP – North Western Frontier Province.

From: Shami *et al.*, Pakistan education 2005; Kardar (2005).

Pakistan is striving hard to participate in international comparative studies/programs for assessment of student achievement to get information about the comparative performance of Pakistani students. In this connection National Education Assessment System (NEAS) was established as a quality assurance innovative program to develop national capacity for monitoring the learning achievements of elementary-level students in order to improve the quality of services (curriculum, textual material, teachers' delivery, policy formulation, etc.) in the education sector. The NEAS has set up provincial/area educational assessment centers at provincial headquarters – FATA, FANA, and AJK.

Financing of Education

The major share of funding for education comes from the public sector, which spent 2.5% of the gross domestic product (GDP) in 2006–07 on education. Around 0.5% funds are contributed by the private sector, putting the combined resources at around 3% of GDP in 2006–07. The total public-sector budgetary expenditure on education has almost doubled in the last 6 years. It has increased from 1.8% in 2000–01 to 2.5% in 2006–07. However, the public expenditure on education is relatively low as compared to other developing countries in the region. In the allocation of budgetary resources, the primary sector has almost half of the education budget, that is, 44%. The secondary sector has 24% and 13% goes to the tertiary sector, while the rest is distributed among the other sectors. Expenditures on primary and secondary education, therefore, are about five times more than the expenditure on the tertiary sector.

Challenges and Way Forward

Pakistan is facing numerous challenges as many developing countries in the world are also doing. Achieving universalization of primary education, overcoming illiteracy, providing access to quality education, controlling student dropout, bridging the gender gap, addressing urban–rural disparities and active participation of private sector, and improving the quality of teachers are some very basic challenges the government is trying to cope with.

The major challenge facing Pakistan is the mass illiteracy in the country, particularly among females. The government being sensitive to the issue has increased the educational budget and proposes to increase it further in the year 2009–10. An additional 10 000 NFBE schools are planned for the next 4 years to lessen the shadows of darkness. Universalization of primary education and meeting the targets of MDGs and EFA is another top priority of the government. After the reinforcement of Compulsory Primary Education Ordinance all possible facilitation is being provided by the government to achieve

the targets by the year 2015. Pakistan is being supported by international partners like United Nations Educational, Scientific and Cultural Organization (UNESCO), United Nations Children's Fund (UNICEF), World Bank, AusAid, and United States Agency for International Development (USAID) in its effort to achieve the MDG by 2015. The issue of gender parity is also addressed as crosscutting objectives in all such efforts.

Access to education is the basic right provided in the constitution of the country to each individual in the country. Despite being high on the agenda of the government, the goal of universal access has not yet been achieved. The country is making a major effort to expand access to education. The issue is being approached through setting up more primary schools in the public sector and supporting nongovernmental organizations (NGOs)/the private sector to invest in this area. At the same time upgrading of existing primary schools to middle level is also being pursued aggressively.

Gender disparity in education is another major challenge which needs to be addressed, particularly in the rural areas of Pakistan. In urban areas, the public sector has almost closed the gender gap where the proportion of female-to-male enrolment was 1:49. In the private sector, the proportion of female-to-male enrolment was 19:81. This gap in enrolment is expected to close with expansion in private-sector facilities. In the case of enrolment in rural areas, the gender gap in both the public- and private-sector institutions is quite pronounced. In both cases, the proportion of female-to-male enrolment is 62%. With policy emphasis on female enrolment and launching of free textbooks and scholarship schemes, the gender gap should close in the rural areas as well. Both the federal and the provincial governments recognize the need for eliminating gender disparity. Policy reforms are focusing on increasing girls' enrolment and their retention.

As the government alone cannot accomplish the gigantic task of attaining the goal of sustainable quality education, there is greater need to facilitate the private sector by financial, administrative, and management empowerment through public–private partnership. It is expected that efficient private-sector leadership patronized by the public sector will result in better quality education and less dropouts.

Implementation and sustainability of educational goals and allocation of funds is now the responsibility of districts, therefore, increased concentration is being given to improving quality by hiring new teachers, increasing their accountability, improving the learning environment in schools, and cutting nondevelopmental educational expenditures.

Higher Education Commission established in 2002 has been working hard to raise the quality of higher education in accordance with the international standards and developing research culture in universities. There has

been a substantial increase in budgetary allocations to universities and several new disciplines and specializations have been introduced in recent years.

Relevant Websites

<http://www.balochistan.gov.pk> – Department of Education, Literacy and Non-Formal Education, Government of Balochistan, Pakistan.
<http://www.nwfp.gov.pk> – Department of Elementary and Secondary Education, North West Frontier Province, Pakistan.
<http://dsd.punjab.gov.pk/> – Directorate of Staff Development, Government of Punjab, Lahore.
<http://portal.punjab.gov.pk> – Education Department, Government of Punjab, Pakistan.
<http://www.sindh.gov.pk> – Education and Literacy Department, Government of Sind, Pakistan.

<http://www.hec.gov.pk> – Higher Education Commission, Pakistan.
<http://www.hec.gov.pk> – Higher Education Institution in Pakistan and their ranking.
<http://www.ibcc.edu.pk> – Inter Board Committee of Chairmen.
<http://www.sdpi.org> – Links to NGOs working in the area of Education in Pakistan.
<http://www.moe.gov.pk> – Ministry of Education, Islamabad.
<http://www.nef.org.pk> – National Education Foundation, Islamabad.
<http://www.nlp.gov.pk> – National Library of Pakistan.
<http://www.pcst.org.pk> – Pakistan Council for Science and Technology.
<http://www.psf.gov.pk> – Pakistan Science Foundation.
<http://www.usaid.gov> – Partnership for Education-USAID programs in Pakistan.
<http://www.un.org.pk> – Strengthening Teacher Education in Pakistan, a project of USAID and UNESCO.
<http://www.un.org.pk/unesco> – UNESCO Pakistan.
<http://www.worldbank.org.pk> – World Bank Projects-Pakistan Webpage.

Poland

A Janowski, Academy of Humanities and Economics in Lodz, Poland

© 2010 Elsevier Ltd. All rights reserved.

General Background

The official name of the Poland is Rzeczpospolita Polska, *rzeczpospolita* being the polonized version of the Latin *res publica*. Poland is a parliamentary republic and Warsaw is the country's capital. The beginning of the Polish state is associated with the dynasty of Piast at the end of the tenth century. In 966, Mieszko I, a member of this dynasty, was baptized and brought Christianity to the country. The Piasts united the lands which roughly resembled the present territory of Poland. During the following centuries, the Polish Commonwealth (Poland united with Lithuania) expanded eastward and ruled over vast territories situated between the Baltic and Black Seas. In the seventeenth and eighteenth centuries, neighboring Russia, Prussia, and Austria were increasing their political power. This led to the partitions of Poland during the last decades of the eighteenth century. Efforts to regain independence in the nineteenth century were fruitless. Poland emerged again as an independent state after World War I in 1918. In September 1939, the invasion of Poland by Germany started World War II. In 1945, Poland, occupied by Soviet troops, fell under communist rule. Authoritarian rule led to a growing opposition which culminated in the formation of the Solidarity trade union in 1980. In 1981, martial law was imposed and for many years Solidarity became an underground movement. In 1989, after the so-called round table negotiations, the first free parliamentary elections took place. A dynamic period of political and economic transformation began.

The functioning of the Polish State is defined by the Constitution of 2 April 1997. The legislative power is exercised by the Parliament (Sejm) and the Senate (Senat). The Sejm is the lower chamber of the parliament and consists of 460 members elected according to a proportional representation system for a 4-year term. The Sejm supervises the government. The Senat, the upper chamber, consists of 100 members. The executive power is exercised by the president and the council of ministers. The president, who is the head of the State and protects its sovereignty, is elected by the nation for a 5-year term. These political parties in the first-half of 2008 – since the general election that took place on 21 October 2007 – are the ruling coalition: *Platforma Obywatelska* (Citizens Platform) and *Polskie Stronnictwo Ludowe* (the Polish Peasant Party); and the opposition consists of: *Prawo i Sprawiedliwość* (Law and Justice) and *Sojusz Lewicy Demokratycznej* (the Alliance of Democratic Left).

Poland covers 312 685 square kilometres and has a population of 38.18 million. Sixty-one and a half percent of Poles live in towns and urban areas and 38.5% inhabit rural areas. From the year 2002, we have seen a natural decrease in the population. In the year 2005, there were 9.7 deaths to every 9.6 live births per 1000 inhabitants. In 2006, the birth ratio equalled +0.1%, the number of young people under 29 years of age accounted for 39.9% of the population, and the number of children of compulsory school age accounted for 13.0% of the population. The official language is Polish. Poland is very homogeneous as regards nationality; it is estimated that no more than 1 500 000 people are members of national minorities: German, Byelorussian, Ukrainian, Slovak, Lithuanian, Romany, as well as the ethnic minority of Kashubian. The situation of nationalities is unclear because it is estimated that *c.* 1 000 000 Poles left Poland after 2004 (for an undefined period or for ever?) and moved to Western Europe. Approximately the same number of persons arrived from the countries of the former Soviet Union and Vietnam and they now work in Poland's shadow economy. Poland is predominantly Roman Catholic (about 35.8 million baptized). The Catholic Church also includes the Uniate Church (Greek-Catholic) with a congregation of *c.* 82 000. Other religions and denominations are represented by relatively small communities. Among them the biggest are: Orthodox, 510 000; Protestant and related, 140 000; Jehovah witnesses, 126 000; and Old Catholic, 50 000.

After the collapse of the socialist form of enterprise, Poland is witnessing, in the present decade, the steady growth of gross domestic product (GDP), which in percentage terms was +5.4 in 2004, +3.2 in 2005, and +5.8 in 2006. At the end of 2006, the rate of unemployment was 14.9% and average annual inflation 2.1%. In 1991, Poland became a member of the Council of Europe, in 1996 a member of the Organisation for Economic Co-operation and Development (OECD), and in 1999 of the North Atlantic Treaty Organisation (NATO). On 1 May 2004 Poland became a member of the European Union (EU).

Goals of the Education System

The basic principles of the Polish education system are included in the Education System Act of 7 September 1991. Education is defined as part of the common welfare of the whole of society. It should be guided by the principles

contained in the Constitution. The ideas of unfinished school reform from 1999 create the basis for the future development. Two ideas have become target areas for reform of the system of education:

- creating equal educational opportunities for all children and teenagers and
- improving the quality of education, preparing the graduates both for adult life and for permanent self-development.

Structures and Operation of the Education System

The different levels of the education system in Poland are depicted in Figure 1.

Preschool and Early Childhood Care

Pre-primary education is regarded as the first level of the school system. It concerns children from 3 to 6 years of age.

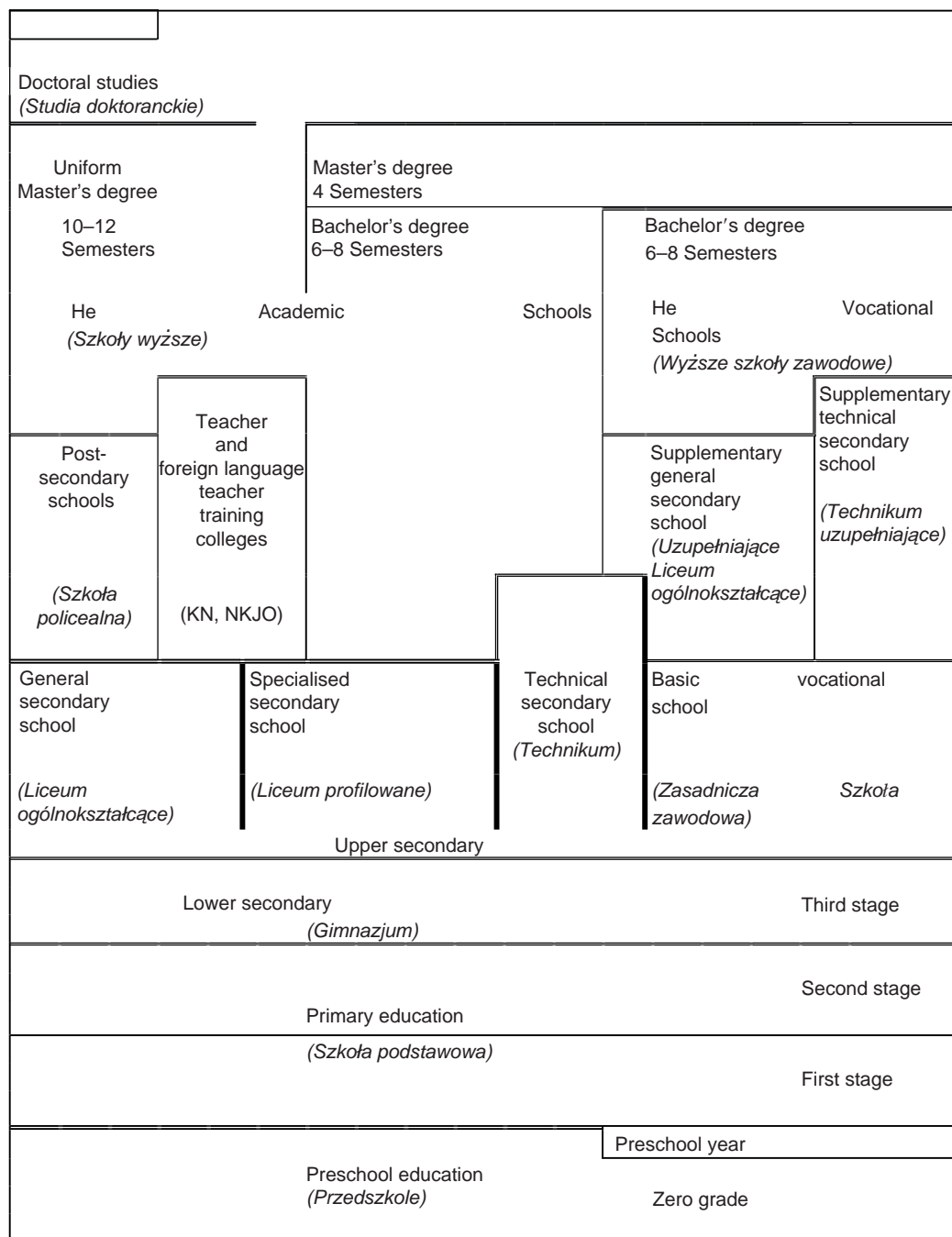


Figure 1 Structure of the education system in Poland.

As of 2004/05, 6-year-old children are obliged to complete a year of preparation for primary education called 0 grade. The classes are attached either to kindergartens or to primary schools.

Primary (Elementary) and Secondary Education

There are three phases of compulsory education: 0 grade, primary school, and lower secondary school.

1. Klasa zerowa (0 grade) is for 6-year-old children.
2. *Szkoła podstawowa* (6 years of primary school) for children 7–13 years of age. This school is divided into two stages:
 - stage 1 – integrated teaching, 7–10 years of age and
 - stage 2 – teaching based on separate subjects, 10–13 years of age.
3. *Gimnazjum* (3 years of lower secondary school) for children 13–16 years of age, and teaching is based on separate subjects.

Post-Compulsory Education (Upper Secondary and Post-Secondary Level)

There are three types of schools in upper secondary and post-secondary general education:

- *Liceum ogólnokształcące* (general upper secondary school) 16–19 years of age;
- *Liceum profilowane* (specialized upper secondary school) 16–19 years of age; and
- *Uzupełniające liceum ogólnokształcące* (supplementary general secondary school) 18/19–20/21 years of age.

Other Schooling Arrangements

Schools can be of two types: public (state) schools, which offer free education within the framework of the core curricula, and nonpublic. The latter can be civic (social), church, or private schools. The schools are called social or civic when they are owned by associations or foundations and they are nonprofit-oriented. All these schools may have their own curricula, which are approved by the minister of education. They are financed by fees received from parents. Most nonpublic schools have small numbers of pupils and small classes. They may be distinguished from the public schools by their personalized teaching programs, by a wider range of curriculum choice, and by a higher standard of foreign language teaching. In 2006/2007 school year, nonpublic school pupils formed 2.0–4.3% of the total number of pupils attending school.

Technical and Further Education

Post-secondary schools (*szkoły policealne*) are included as part of secondary education in the Polish classification and assigned to level 4 in the ISCED. Post-secondary

schools admit first of all graduates of general upper secondary schools. Students in these schools are trained as nurses, accountants, administrative personnel for enterprises and hotels, computer specialists, or librarians. Among upper secondary and post-secondary schools we have:

- *Technikum* (technical secondary school) 16–20 years of age;
- *Zasadnicza szkoła zawodowa* (basic vocational school) 16–18/19 years of age;
- *Technikum uzupełniające* (supplementary technical secondary school) 18/19–21/22 years of age; and
- *Szkoła policealna* (post-secondary school) 19–21 years of age (very rarely 20).

Higher Education

There are state and non-state (private) higher education institutions. The final examination certificate at the end of upper secondary school is required by all institutions for admission to higher education. The first-cycle studies last from 3 to 4 years and finish with the vocational qualification diploma and the title of *licencjat* (equivalent to BA) or *inżynier* (engineer) which gives access to the job market or to extended higher education (the second-cycle studies) to obtain the title of *magister* (MA) or its equivalent. After obtaining an MA diploma, one can apply for doctoral studies – the third cycle of studies.

Adult and Nonformal Education

Adult education aims at the acquisition and extension of general knowledge, upgrading of vocational skills and qualifications needed for a given occupation. Education, training, and in-service training can be provided on daily, evening, extra-mural, distance-learning, out-of-school basis, or in any other system combining the above forms. These tasks are carried out by public and nonpublic adult schools, centers for continuing education, practical training centers, and other institutions run by, among others, associations or foundations. In 2003, external exams were introduced.

All public schools for adults are organized and run by territorial self-government units. The *powiat* (district) authorities are responsible for running of post-gymnasium schools including continuing and practical education centers. The *gmina* authorities are in charge of primary and lower secondary schools for adults (apart from special schools). Nonpublic schools for adults are organized and run by individual persons or by bodies such as associations, social and religious organizations, etc. In the school year 2006/07, in 3666 schools for adults (all types), there were nearly 273.9 thousand students enrolled. Curricula in

schools for adults are usually the same as in youth schools, but adapted by teachers to the needs and requirements of adult education.

The Teaching Profession

The legal act defining the professional status and conditions of service of teachers employed in the school education sector is the Act of 26 January 1982 – The Teachers' Charter (with further amendments). According to the Teachers' Charter, the post of teacher may be taken by a person who has completed either a higher education course with appropriate pedagogical preparation or a course of study in a teacher training college. Any person graduating from these institutions is recognized as a qualified teacher within the specialization he/she has completed. The teacher can obtain the following professional grades: trainee teacher, contract teacher, appointed teacher, and chartered teacher.

The preschool teachers are mostly female – women make up 99.3% of teachers at this level. Teachers with a master's degree constitute 53.6% of preschool teaching staff. In grades 1–3 of primary school, one teacher teaches all subjects (integrated teaching), while in grades 4–6 each subject is taught by a different teacher. To date, teachers have been trained to teach one subject. The newly introduced teacher training standards envisage training teachers as specialists in two subjects. Acquisition of computing skills and a good command of one foreign language is also becoming obligatory.

Teacher training consists of training in subject matter and pedagogical training – teaching methods, psychology, and pedagogy. Teachers working in 6-year primary schools ought to have one of the following qualifications: university education (the title of *magister*), or 3-year teacher training college completed (the title of *licencjat*). Graduates with *licencjat* may complement their education with 2-year university study courses and obtain a master's degree (*magister*). The average salaries for teachers are calculated on the basis of the average salary of the trainee teacher, fixed at a level of 82% of the average state sector salary. This is a fixed mechanism for an annual revaluation and raise in teachers' salaries, identical to that for other parts of the public sector. The average salaries for teachers classified in the remaining categories are calculated as the following percentages of the average salary of the trainee teacher: contract teacher 125%, appointed teacher 175%, and chartered teacher 225%. There is a different division of powers concerning pay regulations. The powers of the minister of national education are limited to fixing the minimum rates of basic pay, whereas the rates for bonuses or allowances and the rules for granting these are determined by the body responsible for the management of a given school.

The working time of the teacher may not exceed 40 h per week. This workload includes the minimum teaching load which for the basic group of teachers (in all types of schools) is 18 h per week (45-min lessons). At the teacher's request, his/her weekly workload may now comprise up to 27 h per week, with any increase in the workload implying a proportionally higher salary. Teachers employed in the 3-year gymnasia ought to have – at a minimum – the qualifications reflected in the title of *licencjat*. Three-year gymnasium teachers are employed according to the same rules as those which apply to 6-year primary school teachers. Upper secondary school teachers ought to have completed a university education (*magister* degree) or equivalent.

Administrative and Supervisory Structure

There are two separate ministries in Poland: Ministry of National Education and Ministry of Science and Higher Education. The main role in initiating and exercising control over policy toward pre-primary, primary, and secondary educational institutions is played by the minister of national education. The Parliament is responsible for the final version of legal acts that determine the orientation of educational policy and the amount of money earmarked for education. Teachers' unions have a considerable role in shaping educational policy. In the last 20 years, we have also witnessed the emergence of a large number of non-government organizations (NGOs) involved in the problems of education, for example, The Centre for Civic Education or the foundation Education for Democracy. They play a very important and stimulating role.

The majority of vocational schools are the responsibility of the Ministry of National Education; only artistic schools, as well as correctional institutions are under the supervision of the Ministry of Culture and the Ministry of Justice, respectively. The national educational policy is developed and carried out centrally, while the administration of education and the running of schools, preschool institutions, and other educational establishments are decentralized. The minister of national education coordinates and carries out the state education policy, and is partly responsible for supervising the work of education superintendents (*kuratoria*). The minister determines the timetables, core curricula, conditions, and procedures for the approval of textbooks. He/she is responsible for the rules for assessing and promoting pupils and for conducting tests and examinations.

The regional level in Poland is the level of voivodship. The number of voivodships is 16. The education superintendent (*kurator*) is the chief educational body at regional level. The education superintendent implements the policy of the minister of education. On behalf of the voivode, the superintendent is responsible for pedagogical supervision

over public and nonpublic schools. At the regional level, self-governing voivodships are responsible for running certain types of educational institutions: teacher training colleges, teacher in-service training centers, pedagogical libraries, schools, and institutions of importance for a given region. The district level in Poland is the level of *powiat* (an intermediate administrative unit between the voivodship and the commune). There are 379 districts. The districts exercise administrative control over upper secondary general and vocational schools, as well as over post-secondary schools and public schools for children with SEN. They are also responsible for the management of art schools, lifelong education centers, psychological and pedagogical guidance centers, and out-of-school education centers.

There are about 500 centers for guidance and counseling in Poland. They are intended mostly for primary and lower secondary school pupils, with the aim of providing psychological assistance.

The local level in Poland is the level of the commune (*gmina*). There are 2478 communes, most of which rural ones. Communes exercise administrative control over preschool institutions, primary schools, and lower secondary schools called a *gimnazjum*.

School heads are recruited on the basis of an open competition and employed by the school governing body for 5 years. The school head decides on the employment of teachers. Some large schools employ a school pedagogical counselor or a psychologist.

Educational Finance

In 2004, public spending on the whole system of education (with higher education included) was estimated at 5.4% of GDP. In 2005, public spending on primary and secondary education (without higher education but with post-secondary schools included) was equal to 4.1% of GDP. All educational tasks carried out by the three levels of local government are financed within the framework of a general subsidy from the State Budget. In the year 2000, a uniform system of allocation of funds, using the algorithmic formula, was adopted. This formula is based on the real number of pupils, adequately increased by the system of weightings (taking into account specific conditions, integration of SEN pupils, sports schools). The local government unit, as a body running the school, is responsible for the designing of a budget program (i.e., a plan of expenditure) for all schools and educational institutions in its respective area.

Performance Monitoring, Evaluation and Research

Pedagogical supervision over the school is exercised by regional education authority (*kurator*-superintendent),

while organizational, administrative, and financial supervision is carried out by the school governing bodies. There are no formal principles for evaluating or monitoring pre-primary school children. The only exception is the so-called balance-sheet of the child aged 6 years which is connected with the recruitment to primary schools. In grades 1–3 of primary school, pupils are promoted automatically if their achievements are assessed positively. Starting with grade 4, pupils are assessed separately in each subject. The evaluation depends on the teacher. Only final marks per semester and at the end of the school year have to be approved by the teachers' council in each school.

An external standardized test upon completion of the primary school (grade 6) was introduced in 2002. These tests are comparable on the national scale. The primary school-leaving certificate is required for admission to lower secondary schools. Internal evaluation at *gimnazjum* level is the same as that in primary education. At the end of the third year of the *gimnazjum*, an external standardized examination is introduced. This examination checks abilities, skills, and knowledge. It is compulsory for all pupils. The results are indicated on the gymnasium-leaving certificate. They are comparable on a national scale. The number of points indicated on the leaving certificate decides about the pupils' admission to an upper secondary school. In the upper secondary general schools, the principles of internal evaluation are the same as those at lower secondary level. On completion of the 3-year general upper secondary school, pupils are awarded a school-leaving certificate. It gives access to the baccalauréat (*matura*) examination. At the end of upper secondary education, pupils may sit for baccalauréat, a national examination, which is compulsory only for those intending to receive the *matura* certificate and to gain access to higher education. At the end of upper secondary technical education (*technikum*), pupils may also sit for (*matura*) examination. In 2004, the new external vocational examination began to replace the old vocational examinations organized by schools.

The measurement of educational achievements and the assessment of school performance are carried out by the Central Examination Commission. This commission is totally responsible for the *matura* examination, as well as for all external evaluation. External support for schools and teachers is mainly provided by the National In-Service Teacher Training Centre, by regional centers, and by educational advisors. Research on educational problems is carried out by the Institute for Educational Research and a number of departments of education within universities.

Major Changes and Issues since 1990

As the first noncommunist government had already taken over the responsibility for education in 1989, it was necessary to:

- introduce new ideas into the teaching of humanities, particularly history;
- draw up and implement a civic education program; and
- increase the scope of foreign language teaching.

The demands for changes were at the beginning of the transformation based on the documents prepared earlier for the round table negotiations by *Zespół Oświaty Niezależnej* (Group for Independent Education) which was secretly active between 1982 and 1989. During those first years (1989–91) it was natural for new educational leaders to define their task in terms of negating whatever the communist authorities had been promoting. For decades, cut off from educational debates going on in the West, it was difficult to imagine what new perspectives for education may emerge. Between 1990 and 1999, there occurred many events which, in modernizing education, prepared the ground for more radical changes. Among them, the following should be considered in more detail:

1. The appearance of educational NGOs.
2. The Parliament's passing of the Act of 12 September 1990 on schools of higher education, which enabled private colleges and universities to be established. Between 1990 and 2005, the number of college and university students increased almost fivefold.
3. Intensive actions aimed at the development of foreign language teaching. According to our estimation, from 1989 there was a need for c. 25 000 teachers of English and the demand for teachers of German and French amounted to about 8000 each. Therefore, in the first half of 1990 a program was prepared for the new type of school – the Foreign Language Teacher Training College. During 1990 and 1991, 55 colleges opened all over Poland, mainly in towns in which there did not exist institutions of tertiary education.
4. The possibility of using other countries' educational experience – consent was given for the creation of Waldorff or Montessori-type schools or schools within the international baccalaureate system.
5. The undertaking of intensive actions aimed at introducing computers and the Internet into schools.
6. A large increase in the number of students applying to secondary schools awarding the baccalaureate (*matura*) diploma and, resulting from this, a decrease in the number of students wanting to study at secondary vocational schools (which did not award the diploma).
7. The abolition of the state's monopoly – that existed before 1989 – in creating the curricula and handbooks. Many institutions and economic entities became empowered to create and publish school books and teaching materials.
8. A gradual handing over to local, district, and regional authorities of the power to run schools and educational institutions and start of the long – and

far-from-finished – process of those authorities learning how to develop and shape local educational policies.

9. The appearance of the possibility for schools to take into account the needs of the region as well as needs connected with the pupils belonging to a specific ethnic group (e.g., teaching the Kashubian language has only been possible since 1990).
10. Until 1998/99, there existed an 8-year primary school, after which students could continue their education in a 4-year general secondary school, technical secondary school, general vocational school, or a 3-year basic vocational school. In 1999, a reform was introduced. Since 1999/2000, the present-day system – described earlier – exists.
11. In 2008, the Minister of Education main assumptions are aimed at finalizing the reform begun 10 years ago. The following activities are foreseen:
 - the dissemination of education of small children, including the compulsory education of all 5-year-olds (gradually, between 2009 and 2011); it will be obligatory for 5-year-olds to undergo 1 year of pre-school preparation;
 - lowering to 6 years, the starting age for children going into the 6-grade elementary school; the above change is connected with the necessity to verify and modify the existing educational programs; and
 - carrying out a curriculum reform of education on the lower and upper secondary schools.
12. Up to now, Poland has no long-term educational policy, that is, a policy which would foresee the state of education in 10–15 years time and point out how this state is to be achieved. In the first-half of 2008, we witness the attempts to create such a strategy.

Further Reading

- Eurydice (2008a). *Structures of Education, Vocational Training and Adult Education Systems in Europe*. Eurydice Unit, Poland.
- Eurydice (2008b). The system of education in Poland, p 64. Eurydice Unit, Poland.
- Główne zamierzenia (2008). MEN (Main intentions of the MONE) internet MONE 17 01 2008.
- Janowski, A. (1992). Polish education: Changes and prospects. *Oxford Studies in Comparative Education* 2, 41–55.
- Janowski, A. (1995). Teachers and multicultural education in Poland. In Gagliardi, R. (ed.) *Teacher Training and Multiculturalism – National Studies*, pp 171–192. Geneva: IBE-UNESCO.
- Janowski, A. (1996). Preface to the book *Building Civic Education for Democracy in Poland*. Washington, DC: National Council for the Social Studies.
- Janowski, A. (1999). Bullying at school. The country report: Poland. In Catalano, R., Junger-Tas, J., Morita, Y., et al. (eds.) *The Nature of School Bullying: A Cross National Perspective*, pp 264–275. London: Routledge.
- Janowski, A. (1999). Case study on information utilization in Poland. In Rokicka, W. (ed.) *Educational Documentation, Research and*

Decision Making: National Case Studies, pp 229–250. Geneva: IBE-UNESCO.

- Janowski, A. (1999). The specific nature and objectives of civic education in Poland: Some reflections. In Torney-Purta, J. (ed.) *Civic Education Across Countries: Twenty four National Case Studies from the IEA Civic Education Project*, pp 463–482. Amsterdam: IEA.
- Janowski, A. and Stanowski, K. (2003). Foundation education for democracy, Poland. In Dunn, S., Fritzsche, K. P., and Morgan, V. (eds.) *Tolerance Matters – International Educational Approaches*, pp 138–158. Gutersloh: Bertelsmann.
- Janowski, A. (2004). Transformacja oświaty w latach 1989–2004 (Transformation of education in the years 1989–2004). *Polska transformacja ustrojowa*, pp 190–213. Warsaw: Fundacja Innowacja.
- Janowski, A. (2007). Educational restructuring and change: Post-communist educational transformation in Poland. *Orbis scholae* 1(2), 80–108. Charles University in Prague.

Relevant Websites

- <http://www.bip.gov.pl> – BPI (Bulletin of Public Information).
- <http://www.codn.edu.pl> – Centralny Ośrodek Doskonalenia Nauczycieli (Central Service Teacher Training).
- <http://www.cke.edu.pl> – Centralna Komisja Egzaminacyjna (Central Examination Commission).
- <http://www.stat.gov.pl> – Główny Urząd Statystyczny (Central Statistical Office).
- <http://www.gus.pl> – Gus.pl.
- <http://www.men.gov.pl> – Ministerstwo Edukacji Narodowej (Ministry of National Education).
- <http://www.reformaprogramowa.men.gov.pl> – Reforma Programowa – Ministerstwo Edukacji Narodowej (Reform Program – Ministry of National Education).
- <http://www.scholaris.edu.pl> – Scholaris.

Qatar

S R Salman, Consultant, Doha, Qatar

© 2010 Elsevier Ltd. All rights reserved.

General Background

Qatar is a peninsula with an area of 11 580 km², which has borders – on the south – with the Saudi Arabia. The head of the State of Qatar is Emir Sheikh Hamad Bin Khalifa Al Thani. The right to rule is passed on within the Al Thani Family. The constitution was ratified in 2003. In 1998, the Emir issued a decree setting an elected municipal council.

The national language is Arabic, and the culture is predominantly Arabic Islamic. The Arabic language is dominant language of instruction in government schools, while English is used in teaching in independent schools, private schools, and in all universities.

In 1997, oil and gas accounted for more than 64.3% of (gross domestic product, GDP). However, oil revenue dropped to 56.2%, in 2003, while the natural gas percentage share to the GDP rose from 10.6% in 1997 to 34%, in 2003. The GDP rose from 29.6 billion QRs in 1995 to 232.4 billion QRs in 2007. The *per capita* income rose from \$14 500, in 1995, to \$47 753, in 2007. The rise in GDP enabled Qatar to increase its expenditure on health and education.

Formal education began in Qatar, in 1952, with the establishment of a primary school – for boys – called Qatar Primary School. In 1956, there were 45 teachers, 1000 students, and 14 schools and kindergartens.

Population and Labor Force

The population in Qatar is composed of Qatari citizens, and the expatriate community which originates mainly from the Asian and other Arab countries. The population increased rapidly in size as a result of both natural growth and expatriate immigration. The population increased from 200 000 in 1975 to 744 029 in 2004, and to an estimated 1.4 million, in 2008 – Qatar citizens representing 14% of the population and noncitizens, 86%. The average annual increase in population between 1997 and 2004 was 5.3%. The economically active population (15+) was 280 122, in 1997, and has increased to 535 807, in 2006, of which 11.5% were Qataris (Table 1).

Responsibility for Education and Training

Formal education in the State of Qatar started in 1952. Since then, comprehensive educational policies have been

developed and supervised by the Ministry of Education (MoE). Such policies have been adherent to the heritage of the Islamic nation and its moderate character and constant endeavors to benefit from modern and technological achievements and new educational experiments.

Until 2002, the MoE was mainly responsible for governmental schools and for the certification of private schools and institutions. Qatar started its reform of educational policy by the creation of Supreme Education Council (SEC), which was established by Emiri decree #37 in November 2002. In 2003, the Qatar leadership released an initiative to reform education – under the slogan “Education for a New Era” – aimed at providing the best education for Qataris to prepare them to meet the needs of economic and social development.

Goals of the Education System

The basic goals of the education in Qatar can be summarized in Qatar Vision, which states, “The state desire to build a modern, world class education system that provides students with a first rate education comparable to that offered anywhere in the world.” The report indicates the characteristics of this system as: “it will encourage critical thinking, creativity and freedom of thought, and will promote social cohesion.”

The goals of the educational system can be summarized as follows:

- Reform the education system to increase the capacity of the student to acquire and use knowledge to benefit the society, to adopt a critical and innovative way of thinking, to increase student understanding for global changes, and build the necessary skills and ethical Islamic values for a productive and rewarding life.
- Vocational education and training (VET) to focus on providing students with the competence and skill to serve the expanding industry sector and to link to the labor market – giving industry a highly skilled workforce to support strong performance in an ever changing economy.
- Higher education sector goals are focused on developing efficient, quality institutions that can provide students with knowledge and enable them to contribute to the economy.

National Goals for Schooling have been set by SEC and its two bodies – the Education Institute and the

Table 1 Economically active population (15+) by occupations and gender

2004				1997				Occupation
F%	M%	%	No.	F%	M%	%	No.	
1.5	2.9	2.6	11 514	1.2	2.7	2.5	6986	Legislators and senior posts
23.5	10.1	12.1	52 948	22.4	7.6	9.6	26 890	Professionals
9.0	7.2	7.5	32 777	5.3	4.7	4.8	13 532	Technical and assistant professionals
11.9	7.2	7.9	34 567	10.5	12.5	12.2	34 157	Clerks
7.9	9.6	9.3	40 938	5.6	8.6	8.2	23 068	Service workers and sales
0.0	1.2	1.0	4288	0.0	2.2	1.9	5336	Skilled workers in agriculture and hunting
0.4	27.1	23.2	101 434	0.1	24.6	21.3	59 719	Craftsmen
0.5	13.5	11.6	50 640	0.0	14.6	12.6	35 292	Tool and machine operators
45.3	21.2	24.8	108 455	54.8	22.3	26.7	74 654	Ordinary workers
00.0	0.0	0.0	0	0.1	0.2	0.2	488	Unclassified
100.0	100.0	100.0	437 561	100.0	100.0	100.0	280 122	Total

From General Secretariat of the Planning Council: Population Census 1997 and 2004.

Source: <http://www.qsa.gov.qa/Eng/publication/Annabs2007.htm>.

Evaluation Institute. The third body of the SEC – that is, the Higher Education Institute (HEI) – is responsible for supporting high-quality outcomes for postsecondary education, so it provides scholarship to best-performing students and sponsors them to the best and reputable universities, whether these are based in Qatar or otherwise.

As for VET, the State of Qatar recognizes the need to create a system that will certify and accredit VET institutions. In 2005, the Planning Council (at present, called the General Secretary for Development Planning (GSDP)) sponsored the Labor Market Strategy (LMS) that yielded a number of initiatives, which were coordinated by different ministries and agencies. This strategy aims to support the Qatarization process, build human capacity, and serve the fast-growing economic and social development. One of the programs that originated from this strategy is the National Qualification Framework for VET (NQFVET). This project will recommend a NQFVET, set of standards, and quality assurance (QA) system for the VET providers.

Structure and Operation of the Education System

The structure of the education system in Qatar is different from other countries. At present, there is no centralized body responsible for all education providers. The whole educational system is evolving and, possibly, in the near future, it will witness the development of a unified educational strategy:

The main policymaking agencies and their responsibilities are described below:

1. *Ministry of Education*. Founded in the mid-1951 and was called *Wizarat Al Maarif*. The ministry houses four sections: administration and financial affairs, education affairs, cultural affairs, and planning and curricula. It is fully responsible for the governmental schools in terms

of financing, evaluation and staff member recruitment, and educating older people.

2. *Supreme Education Council*. Until 2002, the MoE was mainly responsible for governmental schools and for the certification of private schools and institutes. The SEC was established in November 2002; it directs the nation's educational policy. It plays an integral role in the development and implementation of the education reform effort through its three institutions: The Education Institute – oversees and supports the independent schools; The Evaluation Institute – develops and conducts testing of students, monitors student learning, and evaluates school performance; The Higher Education Institute (HEI) – advises individuals with regard to career options and opportunities for higher education in Qatar and abroad, and administers scholarships and grants.

Population and Education

Population and their percentage distribution by education status in March 2004:

1. The population was 744 029 with the percentage of illiterate standing at 2.6% Qataris and 12.0% non-Qataris.
2. The percentage of population with highest qualification being completion of primary to secondary schooling was 45.4%, percentage of population with a bachelor, diploma, higher diploma, MSc, and PhD degree = 10.2% in 1986, and it increased to 19.8%, by 2004 – with almost all the change of 14.7% being due to the increase in number of university graduates.

Preschool and Early-Childhood Care

In 2006, there were 15 548 children in preschool care in the age range of 4–7 years, of which 6079 were 4 years old,

5935 were 5 years old, 916 were 6 years old, and 92 were aged 7. The total number of children attending kindergarten, in 2004 were 12 877, of whom 6030 were females and 6847 males. There were 93 preschools of which 42 were Arabic.

Primary (Elementary) and Secondary Education

Schooling in Qatar is compulsory for children. In 2006, there were 278 schools in Qatar (Table 2) of which 49.6% are from the MoE, 18.7% independent, and 31.6% are private schools. Over 48.5% of schools were primary, 29.1% preparatory, and 22.3% secondary. There were 112 498 full-time school students and 12 768 full-time teachers. The number of students in private schools is, at present, over 45 446, comprising 40.4% of all students. Teachers in private schools are 26.3%. The number of indigenous students rose from 42 000 in 1986 to over 140 000 in 2006.

According to 2006 statistics, Qatari students represent 59% in the MoE schools, 74% in independent schools, and 49% in private Arabic schools.

Primary education (1–6 years) lasts 6 years. The primary-school day, normally, contains about 5 h of tuition, and comprises two semesters each of 17 weeks. Overall, the average number of students per full-time equivalent (FTE) in primary schooling was 10.7.

Preparatory schools comprise grades 7–9, and last 3 years with 5 days per week. The average number of students per FTE in primary schooling was 10.4.

Secondary education is, normally, for grades 10–12 and is for 3 years with classes 5 days a week and two semesters each lasting 17 weeks. The average number of students per FTE in secondary schools, in 2006, was 8.4.

Other Schooling Arrangements

To decrease the illiteracy rate, the government educates students in night-classes and in all intervals. The number of such students decreased from 2871 in 2002, to 1845 in 2004.

An estimated 0.8% of children, in Qatar, have disabilities requiring special education. There are five institutes and a federation which deal with special needs at all ages. The Shafallah Center deals with special needs arising from disorders such as Down Syndrome and autism. Al Noor Institute deals with the blind. There are other institutes such as the mentally retarded school, Acoustic school, and Qatar Sport Federations, Rumeila and Oshwar.

Other schools which give special education, of all forms, starting from primary and up to secondary schooling are the Qatar Leadership Academy, and Qatar Academy Learning Center.

Technical and Further Education

Some international accredited private institutions offer short courses with certificates in information technology (IT), English as a Second Language (ELS), accounting,

Table 2 Educational providers, students, and teaching staff (2006–07)

<i>Provider</i>	<i>Institution</i>	<i>Students</i>	<i>Male (%)</i>	<i>Female (%)</i>	<i>Teaching staff</i>	<i>Male (%)</i>	<i>Female (%)</i>
Schools	278	112 498	50	50	12 768	31	69
Independent	52	29 019	54	46	2657	32	68
Moe schools	138	48 834	44	56	6747	29	71
Private arabic	37	10 801	65	35	877	42	58
International and community	51	23 844			2487		
Primary	178	70 730	50	50	7670	18	82
Preparatory	107	46 973	49	51	4846	48	52
Secondary	82	39 682	48	52	4104	46	54
VET							
North Atlantic ^a	1	1859	1236	623	492	269	223
HE ^a	9	9091	2553	6538	1109	680	429
University of Qatar ^a	1	8184	2198	5986	604	381	223
Texas A & M ^a	1	178	108	70	136	79	57
QVC ^a	1	190	1	189	NA	NA	NA
Carnegie Mellon ^a	1	118	42	76	91	52	39
Weill Cornell ^a	1	148	71	77	246	152	94
Georgetown ^a	1	0					
Northwestern ^a	1	0					
CHN ^a	1	273	133	140	32	16	16
Calgary university ^a	1	0					

^aPlanning Council, 2006–2007 statistics.

Source: Schools and Schooling in Qatar, 2006–2007 Education Institute, SEC, Doha, Qatar (http://www.english.education.gov.qa/section/sec/evaluation_institute/seoi/_statistics; <http://www.qsa.gov.qa/Eng/publication/Annabs2007.htm>).

and English. The total number of graduated students is estimated at 5835 per year.

Qatar Aeronautical College (QAC) was established in 1975, and is, at present, the Gulf region's leading provider of training for the aviation industry. It offers full-time, approved courses for pilots, aircraft maintenance engineers, air traffic controllers, meteorologists, and flight dispatchers, as well as short courses in a wide variety of aviation-related disciplines.

College of the North Atlantic, Qatar (QCNA) is a college run and administered by Canadians. Programs include business studies, engineering, and IT. Part-time courses are also available. For the year 2006–07, there were 1859 students registered, of which 623 were female and 1236 male.

Higher Education

Qatar has a number of postsecondary institutions. These include the Qatar University (QU), the Seven United States Universities located at the Education City, the University of Calgary, which is a Canadian University, and Stenden University, which is a Dutch University.

The QU is the only national university to have been established in 1977. It offers six areas of specialization: science, engineering, Sharia and Law, administration and economics, education, and business. The number of graduates in 2000–01 was 1445 (82% Qataris) and decreased to 1332 in 2005–06 (79% Qataris), with graduates from science and engineering representing ~42% of the total.

Since 1995, the Qatar Foundation (QF) established the Education City (EC), which accommodates some of the best world universities. In 1998, Virginia Commonwealth University School for Arts (VCUQ) started offering a bachelor of arts (BA) in design. It accommodates 190 students (2006–07 statistics with 189 being female and one, male). The Weill Cornell Medical College in Qatar (WCMC-Q) educates young people in this region to be physicians, biomedical researchers, and medical educators; it accommodates 148 students (2006–07, 77 female and 71 male). The Texas A&M University at Qatar (TAMUQ) opened in 2003. It offers undergraduate degrees in chemical, electrical, mechanical, and petroleum engineering, and accommodates 178 students (2006–07, 108 males and 70 females). Carnegie Mellon University in Qatar offers undergraduate degrees in two of its most highly ranked programs – business and computer science – and it accommodates 118 students (2007–08, 76 female, 42 male). Georgetown University, established in 2005, offers programs leading to a bachelor's degree in foreign service. The Faculty of Islamic Studies is an international center for Islamic thinking and dialog and is committed to enhancing research into the Islamic culture. The Northwestern University opened in 2008, and offers degree in

journalism and communication and it is the newest institution to open a campus at the EC.

In addition, 151 students study for a year in the Academic Bridge, which prepares students for entrance into the QF universities.

The Stenden University is a Dutch-administered university that offers diploma, bachelor, and master of arts (MA) degrees in hospitality, leisure and tourism, business, retail, and service management. In 2006–07, there were 273 students, of whom 140 were female and 133 male. The Calgary University Faculty of Nursing opened in May 2007 – this Canadian university offers a bachelor of nursing (BN) program. The total student enrolment was 273, of which Qatari students represent 71%.

The total number of students graduating in 2005–06, from all HEIs, was 2239 – among them, 69% were females and 64.8% were Qataris.

Part of the reform policy is the creation of the HEIs. It aims to enhance the quality of higher education, and thus, it has created several scholarship programs, namely:

- *Emiri scholarship.* Awarded to Qatari students with high grades from secondary schools. In 2005–06, there were 114 recipients of this scholarship – only six of them studied abroad and the rest in EC.
- *National scholarship.* Intended for highly qualified students who are accepted in universities selected to be on the national list (250 universities highly ranked, but with lower ranking than the Emiri list, which has 50 universities). A total of 204 students received this scholarship for the year 2005–06.
- *Employee scholarship.* Offered in partnership with the public and private sectors on condition that the students study in HEI-approved universities. The number was 119.
- *Pre-College (Academic Bridge in EC).* The number was 127.
- *VET Diploma.* Usually, students who do not qualify to Emiri or National HEI-approved universities can study in QCNA. The total number of students enrolled was 90.
- *Diploma, Qatar Aeronautical College.* The number was 187.
- *Pre-college grant.* This is for students who are accepted in Emiri and National Scholarship but need a pre-college additional academic preparation. The number of students enrolled in the United States, Canadian, and Australian institutions was 87, in Europe 256, and in the Middle East 261.

Adult and Nonformal Education

Qatar is also home to a very large number (around 90) of nongovernmental providers and others derived from credible overseas institutions. There are many where the quality is, as yet, quite unknown. A recent study identified a large number of organizations that provide training in specific skills highly valued by the labor market, such as

information and communication technologies (ICT) and English. There are no statistics for the number of students engaged in VET and nonformal institutions. Their certificate is only recognized by their employer and not by the MoE.

The Teaching Profession

In 2006–07 (**Table 2**), there were 6747 FTE teaching staff in MoE schools (52%), 2657 in independent schools (20.8%), and 3364 employed in private schools. As a whole, school teaching is a highly feminized profession, for example, in all schools, 69% of the staff are females.

For the higher education sector, there are 1119 full time teachers (FTT) (**Table 2**) 33.8% females.

The training for all primary and secondary schools and specialist educational personnel is undertaken by universities, especially the University of Qatar. In addition, the College North Atlantic provides a teaching diploma which is recognized by the MoE, and it has 492 (FTE).

Administrative and Supervisory Structure

Government schools in Qatar operate under the direct responsibility of the MoE. The independent schools come under the SEC. Private schools have their own administrations but they are established and operated under conditions determined by MoE registration authorities. MoE and SEC recruit and employ teachers for governmental and independent schools, supply buildings, equipment, and materials and provide funding for use by schools.

Private education has always been an important part of the Qatari education system. There are different private schools in Qatar. The first type is Arab schools and the second is the international schools that are administered and run by the foreign embassies in Qatar. The third type is the foreign schools – such as the British, French, Indian, and Iranian schools.

In the higher education sector, decision making, regulation, funding, and governance depend on the statutes of the university. For example, QU is funded by the government and the students. QF universities are governed by the board of the foundation and each university has its own board. Budget is allocated by the QF, together with students' fees. The Stenden University is a private university and it is governed by its own board.

Educational Finance

The total expenditure on education has two components: public and private. For the financial year 2003–04, total

government education and training operating expenses were QR 2 455 342 billion (2.85% of the GDP).

The source of income for government schools is from the government and from students' fees. Income sources for nongovernmental schools are mainly from students' fees and from fees for services. The governmental VET institutes are funded by the government and from fees for services. Private VET institutes are founded by students' fees and fee-for-service fees.

Performance Monitoring, Evaluation, and Research

The general education quality indicators are monitored by governmental agencies, such as the Planning Council (GSDP), MoE, and SEC. At present, all independent schools and their performance and programs are evaluated by the Evaluation Institute. All those evaluation and assessments programs are based on data gathered by different agencies such as SEC, MoE, and Qatar Bureau of Statistics. Educational research is carried out either by the MoE or SEC.

Research programs at QU and QF universities are funded by government grants and private companies. Two years ago, the QF created priority-research initiative and developed an ambitious program. The Emir of Qatar allocated 2.8% of the GDP for research. The QF stated that "Research is an essential component in QF strategy to make EC a world-class center for innovative education. QF Research, Science and Technology Division, with EC branch campuses and Research Centers, is aiming at pursuing cutting-edge research and development." The QF mission is "to organize Research around core platforms of Medicine, Biotechnology, Information and Communication Technologies, Environmental Sciences, Energy, Molecular Sciences and Nanotechnology. Qatar Science & Technology Park QSTP will be an incubator for private companies to partner with government agencies and academic institutions to develop research for commercial applications."

Increasingly, overall system performance is judged using international and national performance indicators. Drop-outs from schools range between 1.3% and 2.1% for females in primary schools compared to 1.3–8.7% for males. For preparatory years, the percentage ranges between 1.9% and 2.6% for females and 4.7–5.5% for males. For secondary schools, the percentage increases for both genders; so, for the females, it ranges from 3.7% to 7.3% and 8.6–13.0% for males. For secondary religious, trade, and industrial schools (only for males), it ranges from 14.3% to 20.8%.

The Evaluation Institute carries out student assessment using four elements: a national test of student's achievement, a process for regular administration, a process for monitoring of these tests, and a process for producing

reports on student's achievement at the national, school, and individual level. Student assessment is done according to a system known as Qatar Student Assessment System (QSAS). The SEC developed new curriculum standards for Qatar in Arabic, English, mathematics, and science. Students in night and illiteracy schools totaled 3271 in 1998, and 3041 in 2004.

All universities at the QF are accredited. In the case of professional schools at the QU, the College of Engineering received the Accreditation Board for Engineering and Technology (ABET), in 2005, for four of its programs; most colleges and departments at QU are, at present, working toward applicable accreditation in their fields. The College of Business and Economics is, at present, nearing completion of its accreditation by the Association to Advance Collegiate Schools of Business AACSB.

Major Changes and Issues Since 2000

Qatar society and government are facing great challenges due to increasing revenues from oil and gas, globalization, and internationalization. Qatar is willing to diversify its economy and education and to build a knowledge-based economy. Qatar is aware of the challenges and it answers them by publishing the document Qatar National Vision which, approved recently, states

The State of Qatar desires to build a modern, world class education system that provides students with a first rate education comparable to that offered anywhere in the world. This system will allow citizens to develop to their full potential and prepare them for international success, while helping them to understand and respect the values, traditions and heritage of Qatari society. It will encourage critical thinking, creativity and freedom of thought, and will promote social cohesion. To further support this pursuit of human development, Qatar aspires to be seen as an active centre in the fields of scientific research and intellectual activity.

Qatar witnessed a number of major changes in education and research. Those changes are outlined below:

1. The SEC was created in 2002, with three bodies. SEC increasing responsibilities are evidenced by the increasing number of independent schools, use of the international approaches to assess the performance of students, administration, and various departments.
2. The creation of QF and its universities, QSTP, the different research centers, the Sidra Hospital, and the Science Research Division. All these changes, combined with increasing expenditure on research, will increase the number of quality graduates, will involve the industry, universities, and research centers in performing research and development activities that will help

serve the society and enable Qatar to tackle a number of issues related to the development of a knowledge-based economy.

3. Qatar has also witnessed great development in many sectors such as construction, services, and hydrocarbon industries that require an increasing number of professionals and knowledge workers. This can only be achieved through providing a quality education, from VET to citizenship education.
4. Qatar is working to link education output to the labor market through programs adopted by the Ministry of Labour and Social Affairs while the GSDP is working with different agencies to create conditions for Qatar to enter the knowledge-based economy.

Further Reading

- Brewer, D. J., Augustine, C. H., Zellman, G. L., et al. (2007). *Education for a New Era*. Santa Monica, CA: Rand-Qatar.
- General Secretariat of the Planning Council (2007). *Population Census 1997 and 2004*. Doha, Qatar.
- Planning Council, State of Qatar (2004). *A Census, Doha*. <http://www.qsa.gov.qa/eng/index.htm>.
- Qatar Foundation (2007). Student Service, financial aid: Student employment program. <http://www.qf.edu.qa/output/page3.asp>.
- Qatar University (2006). *The Reform Project*. Doha, Qatar. <http://www.qu.edu.qa/theuniversity/reformproject/index.php>.
- Stasz, C., Eide, E. R., Martorell, F., et al. (2007a). *Post Secondary Education: Employer Demand, Student Choice, & Options for Policy*. Doha: Rand Qatar Policy Institute.
- Stasz, C., Eide, E. R., Martorell, F., et al. (2007b). *Post-Secondary Education in Qatar: Employer Demands, Students Choice, and Options for Policy*. http://rand.org/pubs/monographs/2007/RAND_MG644.pdf (accessed September 2009).
- Supreme Education Council (2007). *School & schooling in Qatar, 2006–2007: A statistical overview of aspects of schools and schooling in Qatar*. Evaluation Institute, SEC. <http://www.sec.gov.qa/EVI/SchoolingReport/English-2006-07.pdf> (accessed September 2009).
- UNDP (2006). *Human Development Report, 2006*. Doha: UNDP.
- World Bank (2005). *A Labour Market Strategy for the State of Qatar, Main Report vol. 1*. World Bank Study Prepared for the Planning Council, Doha, Qatar.

Relevant Websites

- <http://www.qu.edu.qa> – Accreditation, Qatar University.
- <http://www.ucalgary-qatar.ca> – Education, University of Calgary – Qatar.
- <http://www.education.gov.qa> – Evaluation Institute, Supreme Education Council, SEC.
- <http://www.sec.gov.qa/> – Ministry of Education Site MoE.
- <http://www.qatar.northwestern.edu> – Programs, Northwestern University in Qatar.
- <http://www.qatarization.com> – Qatarization.com, Qatar Petroleum Training Program Site.
- <http://www.qf.org.qa> – Qatar Foundation Site.
- <http://www.stenden.com/en/stenden/Locations/qatar/Pages/default.aspx>.
- <http://66.195.251.235/sciencefair/public/main.asp>.

Russia

V M Filippov and G F Tkach, Peoples' Friendship University of Russia, Moscow, Russian Federation

© 2010 Elsevier Ltd. All rights reserved.

General Information

At 17 075 400 km² (6 592 800 mi²), Russia is, by far, the largest country in the world – covering more than an eighth of the Earth's land area and extending across the whole of northern Asia and 40% of Europe.

With 142 013 000 people (as of 1 November 2007), it is the ninth largest by population with four-fifth of the population living in the European part of Russia. The majority of the population is urban – 73%. The population is multinational – besides Russians (79.83%), there are people from over 180 nationalities. The state language is Russian. The Russian Federation (RF) is a democratic federate state with the republican form of governing; it consists of 86 subjects with equal federal rights-republics, *krais* (territories), *oblasts* (provinces), and federal cities. Local self-government functions within the bounds of given authority.

In 2005, 74 million Russian citizens were economically active; of these, 68.6 million were actually involved in the country's economy. The number of the unemployed totaled 5.6 million people (7% of economically active population). The majority of people were involved in trade and service sector (20.7%), processing industry (17.8%), agriculture (9.5%), and education (9.3%). Russia's economy comprises almost all branches of industry and all forms and types of production and manufacturing. This fact accounts for existing or potential demand for employees of all vocations and levels, which justifies high requirements to the universality and scale of the national education system. On the other hand, fast changes of the vocational and qualification structure of demands for working force in the last 15 years call for high flexibility and adaptation of the educational system, which presents some difficulties considering its scale.

Aims and Tasks of the Educational System

The main strategic task of the development of the Russian education is specified in the Federal Target Program of educational development in 2006–10. The program envisages ensuring conditions for meeting the demands of citizens, society, and labor market in quality education by creating new institutional mechanisms of regulation in the sphere of education, renewing structure and content, developing fundamental and practical character of educational programs, and forming a system of continuing

education. To accomplish this aim, the following strategic tasks should be realized:

- improving content and technologies of education;
- developing the system of ensuring quality of educational services; and
- improving economic mechanisms in the educational sphere.

Talking about concrete indicators of educational activity, the abovementioned tasks envisage the following results by 2010:

- Elaboration and implementation of new standards of general education in 60% of educational disciplines.
- Introduction of third-generation standards in all fields and specialties of higher professional education.
- A 1.5 times increase – in comparison with 2005 – in the number of students trained with the use of information technologies.
- Increase in the number of entrants in institutions of secondary and higher professional education who are enrolled subsequent to the results of the Unified State Examination (USE) from 40% to 90%.
- Increase in the ratio of higher education institutions (HEIs), realizing two-level professional education programs from 15% to 70%.
- A 1.3 times increase in the number of adult population mastering supplementary education programs.

Structure and Functioning of the System of Education

The general scheme of the Russian education system is presented in **Figure 1**. Integral qualitative indicators for main sectors are shown in **Table 1**.

General Characteristics of the System of Education

The system of education of the RF ensures realization of programs of general and professional education of various levels and fields. Key parameters of the educational content and organization of the study process – at all levels of education – are regulated by federal state educational standards, which are elaborated and adopted no less than once every 10 years in the order stipulated by the government.

The standards are a combination of requirements compulsory for all main educational programs in all educational institutions (EIs) with state accreditation. These requirements define: structure, realization terms, and learning outcomes of educational programs. On the basis of these standards, federal bodies of education management elaborate model educational programs for all levels and fields of training – which serve as a basis for study plans and programs realized in various EIs.

Preschool Education

The system of preschool education comprises daycare centers for babies from 12 to 36 months and kindergartens for children from 3 to 7 years. In 2005, 4 630 000 children – that is, 57.3% of all children of the corresponding age

group – attended 46 500 preschool EIs. There is one tutor for approximately ten children. In rural areas, there are 40% of preschool facilities, with 20% of children attending them. Most preschool general education institutions are public (13.6%) or municipal (83.7%) and are funded from corresponding budgets. A recent trend is the appearance of private and semi-private kindergartens (2.2% of the total number). Alongside with educational functions, kindergartens function as social, tutorial, and health-improving facilities. Kindergarten elementary educational program serves as a basis for primary school.

Primary and Secondary General Education

General education comprises three levels corresponding to three levels of educational programs: primary general,

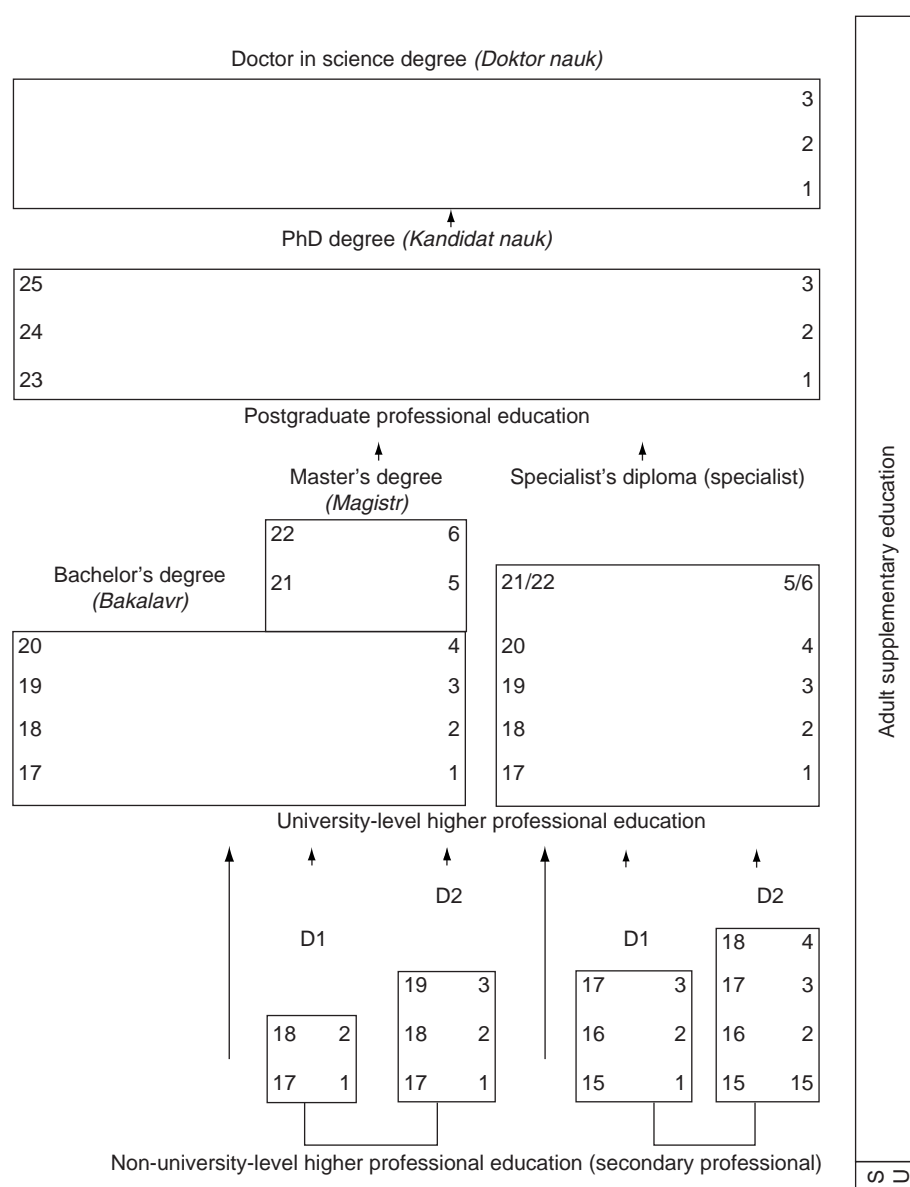


Figure 1 Continued

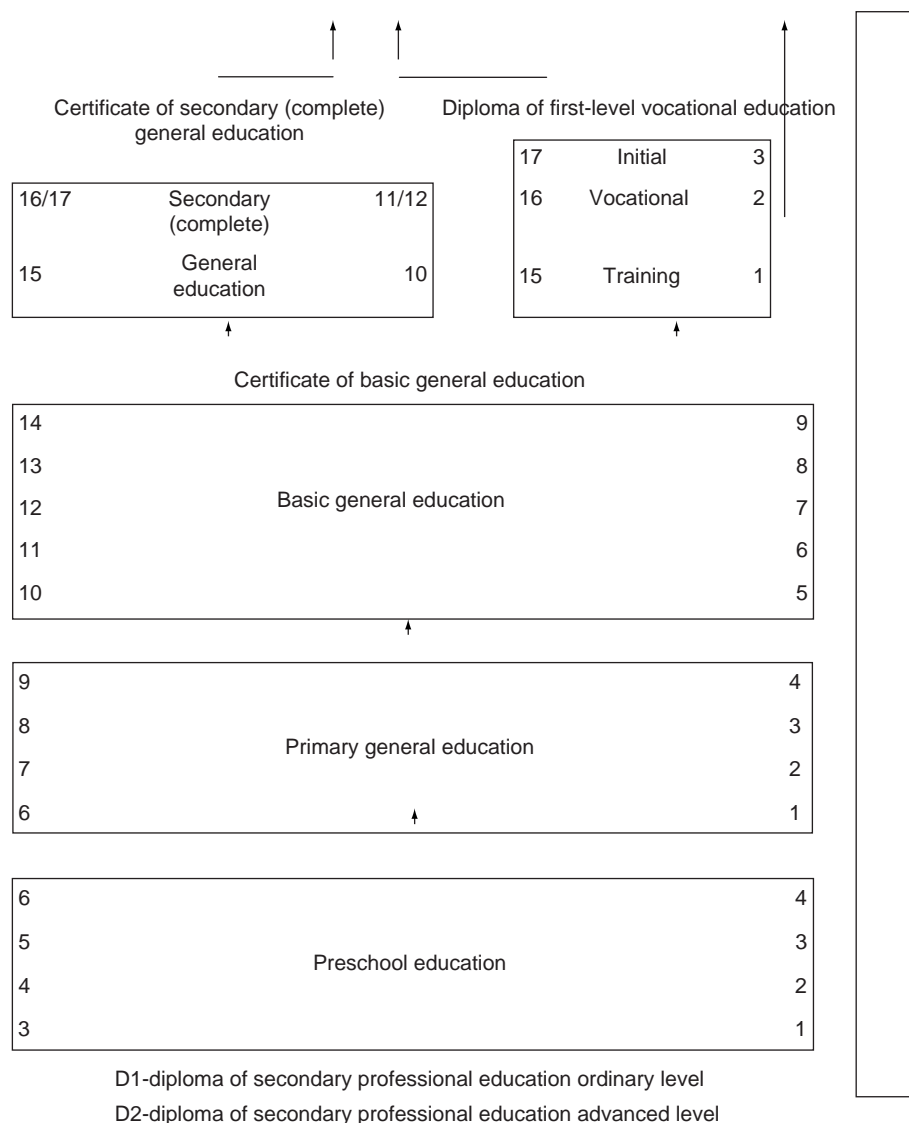


Figure 1 Education system of the Russian Federation.

basic general, and secondary (complete) general – the duration of each is 4, 5, and 2 years, correspondingly. These successive programs lasting, all in all, 11 years (12 in some schools with certain specializations) are usually realized in the same EI, that is, secondary general education schools, lyceums, and gymnasias. Basic general (9 years) education is compulsory. In 2007, a law on transition to compulsory complete (11 years) secondary education – at schools and institutions of initial vocational or secondary professional education – was adopted.

In 2005, 15.6 million pupils studied at 63 200 general education institutions (35% of pupils – primary general, 48% – basic general, and 17% – secondary complete general education). There is one teacher for, approximately, 17 pupils in primary school and 14 pupils in high school. On average, there are 18–19 pupils in one class (23 pupils

in urban areas and 13 in rural areas). Repeater ratio is less than 1%.

Children with disabilities and teenagers with deviant behavior attend specialized corrective EIs (70% of which are boarding schools) and specialized classes at general education schools. In 2005, 236 000 pupils studied at 1946 specialized corrective schools, and 189 000 pupils in specialized corrective classes at regular schools. All correctional schools and classes are either public or municipal.

Children start school at the age of 6. At primary school, they have traditional – for this stage (1st–4th forms) – subjects: the Russian language, mathematics, regional ethnography, the world around us, practical arts, music, arts, and physical training. From the 2nd form, a foreign language (in most schools, English) is introduced. In the 1st form,

Table 1 Education institutions, students, and teaching staff (2006)

<i>Level and type of education</i>	<i>Institutions</i>		<i>Students, thousand</i>		<i>Teaching staff, thousand</i>	
	<i>State</i>	<i>Nonstate</i>	<i>State</i>	<i>Nonstate</i>	<i>State</i>	<i>Nonstate</i>
General education	60 771	726	15 113	72	1575	19
Initial vocational	3025	–	1429	–	152	–
Secondary professional (nonuniversity-level higher professional)	2688	217	2473	118	140	10.4
University-level higher professional	655	413	5985	1079	322	65.2

the workload is 20 compulsory academic hours, 30 in the 4th form; children study 5 days a week.

The basic general education commences from the 5th form (10-year-old children). At the first stage of secondary education (5th–9th forms), mainly, attention is paid to fundamentals of natural sciences, humanities, and social disciplines with teachers in each subject. Study plans and programs at this stage are basically standard. Lyceums and gymnasias introduce elements of early specialization. National autonomies introduce additional hours of the national language.

In the final stage of school education (10th–11th forms), pupils complete the compulsory complex of main general education disciplines at the level of secondary (complete) general education. At this stage, education becomes more differentiated. Region wise, it means realization of the region component – which takes into consideration specific characteristics and demands of each region. Differentiation implies that each school can introduce certain disciplines for advanced studying within the alternative part of the study plan; pupils also take up some elective disciplines. The workload increases from 30 to 36 academic hours a week; pupils study 6 days a week. Russia is, at present, shifting to profession-oriented education in the 10th–11th forms, while simultaneously obtaining complete secondary education.

First and second stages of secondary general education culminate in state final attestation examinations. Final examinations are taken in 5–6 disciplines from the state standard list compulsory for all types of schools. After passing examinations in the 9th form, a pupil obtains a certificate of basic general education – which ensures the right to continue studying in the 10th form or to enter initial vocational or secondary professional EIs. The first two options are available without any entrance examinations or contest, while, at secondary professional EIs, there is a contest. After passing examinations in the 11th form, school-leavers of all Russian schools get a certificate of secondary (complete) general education, which entitles them to enter university and non-university level HEIs, irrespective of the chosen field. Starting from 2009, following long experiments, a USE is to be introduced. It combines state final attestation at schools and entrance exams at HEIs. Schools consider exam results as final attestation, while HEIs score these results as entrance.

Initial Vocational Training and Nonuniversity-Level Higher Education

Initial vocational training aims to prepare qualified workers and employees for all the spheres of economy – carried out mostly on the basis of basic general education. In 2005 – at 3392 professional–technical colleges, professional lyceums, and centers for continuing professional education – 1.5 million students mastered 1–3 years' programs of initial vocational training, 95% of them were full-time students. The majority (80%) were enrolled in 3 years' programs providing vocational training as well as secondary (complete) general education.

Nonuniversity-level higher education is aimed at preparing specialists on the basis of basic general, secondary complete general education, and initial vocational training. There are two types of nonuniversity-level HEIs: technical colleges realizing programs of the basic level (3 years for graduates of basic general school and 2 years for graduates of secondary complete general school), and colleges offering advanced programs (4 years for graduates of basic general school and 3 years for graduates of secondary complete general school). In 2005, 2.6 million students studied at 2905 nonuniversity-level HEIs (among them, 1.3 million students at 1250 colleges). Approximately 58% of students study at nonuniversity-level HEIs on the basis of secondary complete general education and 42% on the basis of basic general education.

At present, nearly 43% of nonuniversity-level HEIs come under the jurisdiction of federal bodies of the department of education; 49% of the subjects of the RF, 7.4% are nonstate. 118 000 students (4.5% of total number) study at 217 nonstate nonuniversity-level HEIs (among them 133 colleges).

All programs of nonuniversity-level higher education are compiled so that, in all cases, vocational training is based on secondary (complete) general education which is always realized at the initial or previous stage. Nonuniversity-level HEIs, thus, belong to the category of post-secondary EIs. Final state attestation of graduates envisages a final examination in a certain discipline, or a final interdisciplinary examination, and/or defense of a graduation-qualification paper. Graduates get a secondary

professional education diploma and are conferred a corresponding qualification in a certain specialty.

University-Level Higher and Postgraduate Professional Education

University-level higher professional education is realized on the basis of secondary (complete) general and secondary professional (nonuniversity-level higher) education. Entrants are admitted based on their performance in competitive examinations, which – starting from 2009 – will be USEs. For some lines of study (arts, sports, etc.) additional specialized tests and examinations are envisaged. Certain HEIs may also get the right to carry out such tests and examinations.

The HEIs are of the following types: universities, academies, and institutes. The category depends on a number of criteria – mostly on the range of lines of study and levels of training, as well as fundamentality and scale of research carried out by the HEI.

In 2005, more than 7 million Russian students studied at 1068 HEIs. Among them, nearly 6 million students studied at 655 public and municipal HEIs (among them: 332 universities, 172 academies, and 151 institutes). A total of 587 HEIs enjoy the federal status, 48 are under the jurisdiction of the subjects of the RF, and 13 of municipal authorities. Nearly 1.1 million students (15% of total number) studied at 413 nonstate HEIs (40% of all HEIs).

The modern structure of higher education includes study programs of higher professional education of two cycles – program leading to a master's degree and program leading to a specialist's diploma (engineer, economist, teacher, medical doctor, etc.). The master's program is based on the bachelor's degree, lasts a minimum of 6 years, includes practical training, and presupposes further research and/or teaching activity of the graduate. Holders of the bachelor's degrees are enrolled in the master's program based on performance in competitive examinations. The specialist's program means a 5–6-year program, including a complex of general scientific, general professional, and specialized disciplines, as well as training practice for further corresponding professional activity of the graduate. All programs of higher education culminate in a final state attestation including final examinations chosen by the HEI, and defense of the final attestation paper (project). The first level of higher education leads to the bachelor's diploma and, the second level, to the master's or specialist's diploma. The master's and specialist's diplomas give the same professional and academic rights to their holders, including the right to enrol in postgraduate programs.

The main aim of postgraduate professional education (doctoral studies), which is realized mainly in the framework of postgraduate studies and doctorate, is the training of highly qualified research and teaching staff.

Holders of the master's or specialist's degree can enrol in postgraduate programs based upon their performance

in competitive examinations. Postgraduate studies last a maximum of 3 years (full-time) and 4 years (extramural). Those with the PhD degree can enrol in the doctoral program based on their performance in competitive examinations. Doctoral studies last a maximum of 3 years. Postgraduate and doctoral programs culminate in public defense of the dissertation at the meeting of the Dissertation council, – which is entitled to confer corresponding degrees. Based on the results of the defense, a PhD or a doctorate degree is conferred. In 2005, 123 000 postgraduate and 4282 doctoral students trained in 640 HEIs.

Supplementary Education

Various types of supplementary education are available at all levels of general and professional education as well as outside of formalized education – at different establishments and institutions. Supposedly, every third pupil and student and every tenth job holder is engaged in supplementary education. Among various types of supplementary education, the best structured is supplementary professional education – including further training, traineeship, and professional re-training. EIs of supplementary professional education include academies, further training institutes, courses (schools, centers, etc.) of further training, and employment service-training centers. Every year, over 1.5 million people undergo supplementary professional training. Of these, nearly 75 000 are employees of professional EIs (17% – teachers of initial vocational training institutions; 26% – teachers of nonuniversity-level HEIs; 47% – university-level-HEIs teaching staff; and 10% – teaching staff of supplementary professional education institutions).

Teaching Staff

EIs are staffed according to their charters – with different types of EIs having different structure and composition. In preschool EIs, the main categories of teaching staff are – head of the kindergarten, tutors, and teachers. In initial vocational training institutions – directors and their deputies, foremen, tutors, counselors, and psychologists, as well as teachers of general educational and specialized disciplines. In nonuniversity-level HEIs – directors and deputies, teachers, and foremen. Main categories of university-level HEIs teaching staff are rectors, deans of schools, heads of departments, professors, associate professors, lecturers, and assistant lecturers.

In 2005, the total number of the teaching staff of EIs of all levels was 5.8 million people. Among them, in preschool and primary general education – 30%, secondary general (basic and complete) – 54%, higher professional – 15%, and adult learning – 1%. Nearly 80% of the teaching staff are women (86% in general education institutions). In initial vocational training, there are 66% women, nonuniversity-level higher professional – 76%,

university-level higher professional – 53%. University-level HEIs have the following structure of the teaching staff: managerial staff (rectors, vice-rectors, deans, and heads of departments) – 11.3%, professors at departments – 9.5%, associate professors – 37.5%, and senior lecturers, lecturers, and assistant lecturers – 41.7%.

Educational qualification and work experience of the teaching staff, on the whole, meet the requirements. Thus, 60% of primary school teachers have higher and 36% have secondary teacher-training education. At the nonuniversity-level higher education level, these figures are, respectively, 93% and 4%. At university-level HEIs, 48% of all the teaching staff possess PhD degrees and 12% have doctorate degrees.

Education Authorities

Authority in education between federal (central), regional, and local bodies of power are allocated according to the RF law, dated 10 July 1992, #3266-1 “On education” and the Federal law dated, 22 August 1996, #125-FZ “On higher and postgraduate professional education.” The legislative body adopts federal laws on education, approves the federal budget, stipulates education levels, and approves federal programs of education development.

The RF government stipulates the procedure of elaboration and approval of federal state-education standards, approves model regulations on state and municipal EIs, establishes the licensing, attestation, and state accreditation procedure for EIs, and acts as a founder of federal state HEIs.

Federal (central) managerial bodies in education are, at present, the Ministry of Education and Science, Federal Education and Science Supervision Service, and Federal Education Agency (both under the jurisdiction of the Ministry of Education and Science). The main function of the Ministry is working out state policy and adopting normative legal acts in the sphere of education. The Federal service and the Federal agency control and provide organizational, financial, material, and technical support to the educational system at the federal level.

Regions (subjects of the Federation) are responsible for: pursuing federal education policy in the region; elaboration and realization of regional programs of education development with due consideration of regional peculiarities; forming and managing state bodies of education management; setting up EIs of general, initial vocational, and secondary professional education; forming regional budgets with regard to educational expenses and adopting regional normative acts on education funding; organizing training, and re-training and refresher courses for teaching staff.

Local self-government bodies are responsible for realization of the right of citizens to compulsory basic general education. Their competence extends to: planning, organization, and control of activity of local bodies of education

management and EIs with regard to implementation of state educational policy; forming local educational budgets and adopting local normative acts with regard to education funding; setting up and reorganizing municipal EIs; appointing heads of municipal EIs; and constructing municipal EIs buildings and developing neighboring territories.

Each EI is independent in implementing the educational process, in choosing and appointing staff – in research, financial, and economic activity within the competence given to them by the RF legislation standard provisions on corresponding EIs and the charter. Nonstate EIs may be set up in the organization and legal forms stipulated by the RF Civil Code for noncommercial organizations. In their activity, such organizations shall be regulated by the education legislation and the charter elaborated independently and approved by the founder.

Education Funding

General education of all levels and initial vocational training are free of charge. State and municipal EIs provide programs of nonuniversity-level higher professional, university-level higher professional, and postgraduate professional education free of charge upon performance in competitive examinations – if the citizen obtains education of this level for the first time.

Funding of EIs is done on the basis of state and local standard norms established per student in each type and category of EI and level of educational program. The funding scheme is determined by standard regulations on EIs of corresponding types and kinds. Norms of funding of nonstate EIs cannot be lower than the ones for state and municipal EIs on the same territory. Nonstate EIs develop the right to state and/or municipal funding the moment they are granted state accreditation, provided they implement main educational programs. Irrespective of their organization and legal form, EIs have the right to attract additional funds rendering additional paid educational and other services stipulated by the charter in accordance with the RF legislation.

Educational expenses from the consolidated budget of the RF in 2005 totaled 801.8 billion rubles, which is 11.8% of total expenses from the RF consolidated budget, or 3.7% of gross national product (GNP). The federal budget accounted for 20% expenses; consolidated budgets of the subjects of the RF accounted for 80% of expenses. The distribution of expenses was as follows: preschool education – 16%, general (primary and secondary) – 51%, initial vocational – 6.4%, nonuniversity-level higher professional – 6.4%, university-level higher professional – 19.2%, and re-training and refresher – 1%.

Education Control and Quality Assessment, Scientific Research

Federal education and science supervision service is the main body authorized to control and assess the activity of EIs with a view to their congruence with educational legislation, federal state educational standards, and federal state requirements. Control is implemented through mechanisms of licensing, attestation, and accreditation.

Every EI is subject to licensing. It determines compliance of the educational process in the EI with state and local requirements with a view to observing material and technical, sanitary, construction and other rules and norms, adequacy of classroom equipment, education qualifications of the teaching staff, and staffing level.

State accreditation confirms the right of the EI to confer to its graduates a state-standard document of the corresponding educational level. State accreditation is granted following a favorable attestation report. Attestation is carried out no less than once every 5 years and is aimed at determining compliance of the content, level, and quality of preparation with federal state educational standards. The state accreditation certificate confirms the status of the EI for another term and determines a list of lines of study and specialties in which the EI can confer corresponding qualifications and state-standard diplomas. EIs can obtain public accreditation at various Russian, foreign, and international educational, scientific, and industrial associations and organizations.

One of the most important attestation parameters of complex assessment of HEIs is evaluation of scale and results of scientific-research activity and its contingency with the educational process. This parameter is crucial for obtaining university status. Participation in research and achieving results is one of the grounds for teachers' re-election and further career growth. Research is funded independently from funding educational activity.

In the overall structure of institutions and organizations doing research in the sector of higher education, 75% are universities and 20% are scientific research institutes and centers. Over 30 000 employees participate in research and elaborations; half of them enjoy the researcher status. Among them, approximately 40% and 30% work in the field of natural and engineering sciences. Social sciences account for 10%, medical and humanitarian – 5% each of the total number of this category of employees.

Principal Changes and Results Since the 1990s

In the first-half of the 1990s, the Russian educational system faced a crisis due to the break up of the Union of Soviet Socialist Republics (USSR) and consequent change of the socioeconomic system. Qualified teaching staff started leaving their jobs, certain sectors of educational activity saw stagnation and reorganization.

As a result, nearly all quantitative and qualitative indicators characterizing education development declined dramatically. However, there were some positive changes during this period – adoption of the 1992 and 1996 framework laws regulating functioning of education in new conditions and partial introduction of multilevel structure of higher education in 1992–93.

In the second-half of the 1990s, the state started taking decisive actions to improve the situation. First of all, it concerned funding education and increasing teachers' salaries. EIs and the educational community gradually started to adjust to new conditions. In the following years, the consolidation process gained momentum. This allowed elaborating and adopting the RF National education doctrine (2000), and the Concept of Modernization of Russian Education till 2010 (2001). The documents were approved correspondingly by the State Council and the government. The modernization concept served as a basis for short-term and mid-term programs of educational development, which formulated concrete tasks and indicated mechanisms of their realization as applied to different educational sectors and types of educational activity.

These programs envisage projects of modernizing structure and content of general and professional education, education quality-assessment system, EIs' material and technical base development, improving mechanisms of funding education, and ubiquitous introduction of information technologies. Further modernization of the Russian education is considered a key factor of socioeconomic development of the country, and envisages concentration of effort in the following directions:

- bringing education content, teaching technologies, and education quality-assessment methods in balance with demands of the modern society;
- elaborating managerial mechanisms adequate to educational system-development tasks;
- creating economic mechanisms ensuring investment attractiveness of the educational system; and
- ensuring effective participation in education internationalization processes.

The education policy of Russia – reflecting nationwide interests in the sphere of education and presenting them to the world community – also takes into consideration general tendencies of world development, which cause the necessity of significant changes in the educational system. There is sufficient evidence that notwithstanding remaining acute problems, the Russian education has, on the whole, overcome negative consequences of the 1990s' crisis and entered the phase of intense growth. In recent years, the attention of the society and government to education has grown considerably. Since 2005, educational development has become one of five priority national projects, which envisages, in particular, considerable

additional funding targeted, first of all, to the most needy sectors of education activity, and second, to intensive formation of education innovation potential.

Further Reading

Education in the Russian Federation (2006). *Statistical Yearbook*, pp 528. Moscow: State University, Higher School of Economics.
Ministry of Education and Science of the Russian Federation (2007). *Collection of Normative Documents for EIs of the Russian Federation Implementing Programs of General Education*, pp 443. Moscow: "Drofa" Publishing House.

Relevant Websites

<http://www.ed.gov.ru> – Federal Education Agency, US Department of Education.
<http://www.obrazov.gov.ru> – Federal Education and Science Supervision Service.
<http://www.mon.gov.ru> – Ministry of Education and Science of the Russian Federation.
<http://www.russianenics.ru> – National Information Centre for Academic Recognition and Mobility.

Rwanda

E Bahizi, Rwandan National Commission for UNESCO, Kigali, Republic of Rwanda

© 2010 Elsevier Ltd. All rights reserved.

General Background

The Republic of Rwanda is a relatively small, landlocked country in the Great Lakes region of eastern and central Africa, bordered by Uganda, Brundi, the Democratic Republic of Congo, and Tanzania. With an area of 26 338 square kilometers and a population of 8.1 million (2002 census), it is one of the most densely populated countries in Africa. A green and hilly country, Rwanda whose capital is Kigali enjoys a relatively mild climate, with the mean daily temperature being 24 °C and most of the country being located on the fertile eastern plateau. Rwanda is essentially an agricultural country, 83% of the people living in rural areas and over 88% of the working population engaged in agriculture, 3.3% in services, 3% unskilled workers, 2.6% in business, 1.6% professionals, and 0.8% in manufacturing. With a per capita income of \$ 290, Rwanda is ranked 161st on the human development index (HDI).

The population growth rate in 1998 was estimated to be 2.8%, birth rate at 3.9%, and the infant mortality rate was 107 deaths/1000 live births. Life expectancy at birth for the total population was 41.9 years.

In 1899, Rwanda became a German colony under the Germany East Africa. The Germans ruled indirectly through the King (*Mwami*) and his chiefs. After World War I, in 1919, Rwanda became a mandated territory of the League of Nations under the administration of Belgium. In 1946, after the World War II, the country became a Belgium Trustee on behalf of the United Nations. Rwanda became independent in 1962. The post-independence period has been characterized by political conflicts that affected social and economic growth.

In 1994, within a period of less than 3 months at least a million people were brutally massacred; thousands of women and young girls were raped, maimed – both physically and psychologically.

Education System

Rwanda's education system consists of 6 years of primary education, 3 years of junior secondary, and 3 years of senior secondary. After upper secondary, one can proceed to the university or go in for other tertiary education. University education takes 4–6 years depending on the course of study. Primary education is free and compulsory. A landmark development in the country is the introduction

of 9 years of free basic education. This includes 6 years of primary and 3 years of junior secondary. In implementing this, the Government of Rwanda seeks to ensure that all children leaving school will have acquired reasonable academic knowledge, technical and life skills to survive. Higher education is mainly provided by universities and specialized institutes.

A number of institutions provide formal education, ranging from preschools to universities. The civil service reforms within the ministry of education (MINEDUC) have led to organizational changes which are currently being implemented. The departments and units have been reorganized for increased efficiency and effectiveness. In terms of education delivery, MINEDUC leads policy formulation and is responsible for the setting of norms and standards, and for planning, monitoring, and evaluation at national level; districts administer the personnel and monitor performance.

The Government of Rwanda is also implementing major reforms in the areas of decentralization and public service. With decentralization and public service reform currently being implemented, responsibilities for program and plan implementation and monitoring at the district levels lie with the district education offices. It is anticipated that as school management is strengthened, schools themselves will make more decisions. MINEDUC also works with civil and faith-based partners and the private sector to ensure education provision for all.

National Goals and Policies

Rwanda's Vision 2020 aims at the development of a knowledge-based and technology-led economy. It seeks to eradicate poverty and to move Rwanda to the list of middle income countries (MICs) by the year 2020. The people of Rwanda are its greatest asset as it has extremely limited natural resources. Education is therefore at the heart of the Vision's strategy. In today's globally competitive environment, an educated population is a prerequisite for a country to take on technological challenges and development.

Mission statement for education

The global goal of the Government of Rwanda is to reduce poverty and in turn improve the well-being of its population. Within this context, the aim of education is to

combat ignorance and illiteracy and to provide human resources useful for the socioeconomic development of Rwanda through the education system.

Goals for the education sector

Goals for the education sector seek to build on the strengths of Rwanda while acknowledging the challenges to realize the mission statement for the education sector:

1. to educate a free citizen who is liberated from all kinds of discrimination, including gender-based discrimination, exclusion, and favoritism;
2. to contribute to the promotion of a culture of peace and to emphasize Rwandese and universal values of justice, peace, tolerance, respect for human rights, gender equality, solidarity, and democracy;
3. to dispense a holistic moral, intellectual, social, physical, and professional education through the promotion of individual competencies and aptitudes in the service of national reconstruction and the sustainable development of the country;
4. to promote science and technology with special attention to information communication and technology (ICT);
5. to develop in the Rwandese citizen an autonomy of thought, patriotic spirit, a sense of civic pride, a love of work well done, and global awareness;
6. to transform the Rwandese population into human capital for development through acquisition of development skills;
7. to eliminate all the causes and obstacles which can lead to disparity in education be it by gender, disability, geographical, or social group;
8. to educate a free citizen who is liberated from all kinds of discrimination, including gender-based discrimination, exclusion, and favoritism;
9. to contribute to the promotion of peace of a culture emphasise Rwandese and Universal values of justice, peace, tolerance, respect for human rights, gender equality, solidarity and democracy.
10. to dispense a holistic moral, intellectual, social, physical, and professional education through the promotion of individual competencies and aptitudes in the service of national reconstruction and the sustainable development of the country;
11. to promote science and technology with special attention to ICT;
12. to develop in the Rwandese citizen an autonomy of thought, patriotic spirit, a sense of civic pride, a love of work well done, and global awareness;
13. to transform the Rwandese population into human capital for development through acquisition of development skills; and
14. to eliminate all the causes and obstacles which can lead to disparity in education be it by gender, disability, geographical, or social group.

Specific objectives for the education sector

Specific objectives for the education sector are:

1. to ensure that education is available and accessible to all Rwandese people;
2. to improve the quality and relevance of education;
3. to promote the teaching of science and technology with a special focus on ICT;
4. to promote trilingualism in the country;
5. to promote an integral, comprehensive education orientated toward the respect of human rights and adapted to the present situation of the country;
6. to inculcate in children and sensitize them to the importance of environment, hygiene, and health and protection against HIV/AIDS;
7. to improve the capacity for planning, management, and administration of education; and
8. to promote research as a mobilizing factor for national development and harmonize the research agenda.

Challenges

The Government of Rwanda recognizes the challenges faced in the country which can be summed up as:

1. developing approaches to deal effectively with the legacy of the genocide, notably the education of orphans and child-heads-of-family and the provision of feeding and boarding programs for such children;
2. addressing the shortage of teachers, both qualitative and quantitative, at all levels and insufficiency of qualified personnel at central and provincial administration levels;
3. improving the status of the teacher and providing incentives for the job, given the salary and conditions of service which do not motivate;
4. rehabilitating destroyed or defective infrastructure in some parts of the country (including the replacement of furniture, equipment, and educational materials destroyed during the war and genocide);
5. increasing the recurrent budget for education of which the majority is absorbed by paying salaries, and managing changing donor support as Rwanda moves out of emergency into development;
6. monitoring the system with an inspection service that is still in the early stages of development and building its capacity to meet the norms of planning;
7. completing modernization process for educational legislation;
8. providing adequate supplies of textbooks and relevant educational material in schools;
9. improving internal efficiency (rate of failure, repetition, drop out, and insufficiently developed systems to recoup those who are excluded);
10. improving external efficiency (poor performance of leavers);

11. rehabilitation and strengthening of education in science and technology;
12. increasing the number and quality of researchers in all domains and increasing the very limited allocation of funding to research;
13. identifying appropriate educational strategies and measures to assist in addressing the HIV/AIDS pandemic; and
14. halving the rate of illiteracy.

Guiding principles

The education sector strategic plan (ESSP) is guided by the following six principles that derive from Vision 2020, the poverty reduction strategic paper (PRSP), and the education sector policy:

1. Education will be considered holistically as a sector and so a whole sector, or Sector Wide Approach (SWAp), will be developed to be used to assist the planning and management of the system.
2. A medium-term expenditure framework (MTEF) will be used as a tool to ensure that educational proposals are set within the national fiscal planning and management process in the short and medium term, with close monitoring and regular evaluation.
3. The government affirms the importance of partnerships between government, parents, communities, donors, the private sector, NGOs, faith-based organizations (FBOs), and civil society. There will be regular participative consultations, negotiations, and meetings coordinated by the government. A horizontal coordination between different actors will be established, and, through decentralization processes, there will be effective vertical links between central government, local government, and grass roots groups.
4. There is a need to balance access, quality, and relevance with a special emphasis on a curriculum that is outcome oriented and offers the skills and values necessary for development.
5. There shall be gender consideration especially in learning achievement for girls and access to education for women, especially in rural areas.
6. ICT in education shall be considered as the heart of the education system.

International Compliance

Education sector policies comply with important international goals and aspirations. The ESSP therefore builds upon the government's clear commitment in its PRSP to the key role that education can play to improve social and economic well-being and reduce poverty. This is consistent with the United Nation's 2000 Millennium Declaration and the millennium development goals (MDGs), particularly

those that underline the importance of universal primary education (UPE) and the removal of gender disparities.

In terms of poverty reduction and human resource development, education, particularly basic education, is of central importance. Rwanda subscribes to the education-for-all (EFA) principles and process. MINEDUC has developed an EFA plan of action that puts into effect the six goals arising from the World Education Forum in Dakar, April 2000.

Structure and Operation of the Education System

Preschool and early childhood care

Preschool education is financed by parents while the ministry of education provides policy and coordination services. The government offers technical support and plans to give incentives to the private sector to provide such services. Less than 10% of the eligible children are enrolled in pre-schools. In 2001, there were 257 nursery schools of which only two were government owned. These preschools cater to 18 399 children out of about 2 million preschool-aged children. This is just under 1%. These schools are managed by 327 teachers, the majority of whom have received only in-service training.

In the future, the government will play an active part in the financing of this subsector. In addition, the feasibility of using existing primary and other structures to provide such services will be examined, and further data will be obtained to determine the type of service desired by the public and the potential public benefits.

Primary education

The MINEDUC Planning Directorate has conducted school censuses annually since 1997 and substantial data are now available (**Table 1**) as a basis for planning future provision of primary education.

Eleven teacher training colleges (TTCs) offer both initial and in-service teacher training and help to improve the quality of teaching staff by upgrading existing under-qualified teachers and giving new recruits an intensive basic training. These TTCs have the capacity for approximately 2500 students in total each year. There are also several private or government-subsidized schools with special sections to train primary school teachers.

Secondary education

Currently, transition from primary to secondary schools is approximately 37% including all public and private schools, but only 25% for the public schools. There is a higher proportion of boys (almost 60%) (**Table 2**) accessing public and government-assisted secondary schools, and a higher proportion of girls (again, almost 60%)

Table 1 Primary education indicators (1998–2005)

1998	2005	
		1.0. Students
1 270 733	1 857 841	Number of students
635 765	912 207	Number of boys
634 968	945 634	Number of girls
970 161	1 265 205	Students of school age
-	137.3%	Gross enrolment rate
-	93.5%	Net enrolment rate
21.0%		Transition rate to secondary
56.2%		Promotion rate
32.1%		Repetition rate
11.7%		Drop-out rate
		2.0. Teachers
22 435	29 033	Number of head teachers and teachers
	26 944	Number of head teachers
10 002	12 330	Number of males
12 433	14 614	Number of females
10 463	25 255	Number of qualified teachers
		3.0. Teacher/student
57	69.0	Teacher student ratio
121	73.6	Qualified teacher student ratio
		4.0. Schools
1940	2295	Schools
	29 748	Classroom
	36 175	Number of streams

Table 2 Secondary education indicators (1998–2005)

1998	2005	
		1.0. Students
105 292	218 517	Students public and private
51 811	115 350	Number of boys
53 481	103 167	Number of girls
.	642 990	Boys of school age
.	670 200	Girls of school age
.	16.6	Gross enrolment rate (%)
.	9.0	Net enrolment rate (%)
.	8.7	Repetition rate (%)
		2.0. Teachers
4679	7610	Number of teachers public and private
73.6	78.7	% of male
23.3	21.3	% of female
1544	3940	Number of qualified teachers
33.0	52.2	% of qualified teachers
		3.0. Teacher/Student
21.1	29.6	Teacher student ratio (public)
24.8	27.6	Teacher student ratio (private)
		4.0. Schools
322	553	Schools(public and private)
167	337	Public schools
155	216	Private schools

attending private secondary schools. This is partly due to girls' underperformance at the end of primary school and not achieving the necessary level to enter public secondary schools, and partly due to a preference among many

families for girls to be attending boarding schools, most of which are private. Girls' performance at secondary level is also inferior to that of boys. Consequently, the education gap between girls and boys widens with implications for inequality in career opportunities.

Technical and vocational education

Vocational training

Vocational training is offered in youth training centers (YTCs). These centers cater to about 2000 young people out of a total target population of approximately 1 million. There are 22 public and 16 private vocational centers.

The curriculum in YTCs is tailored to the needs of young people with adequate basic education, so those who have dropped out of primary school very early are unlikely to be able to access these YTCs. The proportion of girls is very low and is concentrated in the traditional domains such as tailoring and catering. Only 5.8% of women benefit from the apprenticeship as against 9.1% of men, and only 2.6% of women benefit from short training. There is an acute shortage of equipment and qualified teachers in most YTCs. Fees are also payable, making access difficult for the poor.

Technical and professional/vocational education

Technical and professional/vocational education are provided at the second cycle of secondary education by the MINEDUC. Historically, Rwanda has had a shortage of technicians, and where technical education has existed, it has often been of poor quality. This situation was exacerbated by the events of 1994, where some of the few technicians that existed were killed in the genocide or fled the country.

At present, there are technical schools (ETOs). Most technical schools lack adequate equipment and teachers are not fully qualified. Technical education in agricultural fields is neglected and the participation of girls is still very limited due to the legacy of colonial education policies and gender stereotypes. Professional/vocational education focuses mainly on traditional skills such as secretarial skills, agro-veterinary, nursing, and teaching, and caters mainly to girls. This reinforces gender stereotypes in careers and professional opportunities.

The ministry of education is in the process of developing a policy for formal vocational training subsector and also revising the curriculum for vocational training centers. It also intends to improve the existing vocational training centers. Each district is encouraged to establish at least one vocational training center.

There exists a technical education policy and strategy. The intention of the ministry is to strengthen the existing ETOs and to upgrade some ETOs to providing advanced diplomas at the level of A1.

Higher Education

Higher education is offered at University of Rwanda and at public and private higher education institutions. However, higher education enrolments are still lower than average south of the Sahara (about 1%). There is a relatively low participation of girls in tertiary education (Table 3), particularly in science, technology, and related fields and in public institutions. Private sector participation is picking up, with women's participation encouraging at 49%, while in public institutions it remains at 26%. The issue of equal access and performance for women in higher education institutions, whether public or private, is being addressed through higher education subsector policy.

Currently, these higher education Institutes are relying on expatriate lecturers. The Government of Rwanda is committed to a staff development program by sending a number of recent young graduates overseas so that they can complete their training in masters or PhD level and replace foreign lecturers/professors when they return to Rwanda, and it also intends to initiate postgraduate programs at masters and PhD level so as to produce its own university professors and highly qualified executive personnel.

A cost-sharing scheme has recently been adopted for higher education, with students receiving loans from the government for their studies, and then having to pay this money back when they start employment after graduation. Thus, enrolment in higher education can be expanded, as students now bear more of the costs. This will reduce the burden on the already constrained government resources when loans begin to be repaid.

The Government of Rwanda (GoR) policy is to establish a coherent and comprehensive strategic framework

that sets out the governance, function, role, form, and shape of the higher education system. This framework is detailed in the higher education law, the higher education policy, and the higher education subsector plan.

A key component of the framework is the 2005 Higher Education Law. This law defines the operating environment for all higher education institutions (HEIs), both public and private. It specifies the roles, responsibilities, and duties of all HEIs. The law mandates the establishment of two semi-autonomous organizations. Together, these new organizations will oversee the development of higher education according to the government's stated aims and priorities. These are outlined in the higher education policy and subsector plan. Each institution will have a specific remit and set of responsibilities within the overall framework.

The National Council of Higher Education (NCHE) will support the development of a national strategic planning framework. It will establish a national quality system and a national qualification framework. It will also have the power to accredit and approve the operations of all higher education providers. The student financing agency for Rwanda (SFAR) will manage all students financing, such as loans, grants, bursaries, and scholarships.

The intended outcome of these initiatives and reforms are to allow maximum autonomy to individual institutions for their own micro-management, while at the same time ensuring that the system as a whole is responsive to national priorities and flexible to changing conditions.

An important premise for growth in higher education is the ability to attract private sector investment. A further area for consideration concerns the training of graduates in Rwanda. This will be done by organizing postgraduate studies within the country and also by sending students abroad to obtain training in those professional fields which cannot be accessed in Rwanda. This strategy will increase the number of nationals working in professional and technical areas, reduce the number of expatriate lecturers, and thereby lower the costs of higher education provision.

Adult and nonformal education

The GoR has set a target of having 80% of the population literate by 2010. The recent synthesis of the 2002 census indicates that some 60% of the population aged 15 years and above can read and write in at least any one language, and that men have a higher literacy rate than women. The highest rates of literacy are within the 15–34 year age range.

A draft policy for youth and adult literacy is being developed by MINEDUC. It will seek to be appropriate to Rwanda's context in terms of cultural links and practices (e.g., using the Umuganda tradition, ensuring peace and reconciliation, etc.). The proposed policy distinguishes between learning to read and write (Kinyarwanda,

Table 3 Higher education indicators

1996	2001	2006	
			1.0. Students
3948	12 802	37 159	Total public and private
3948	8723	16 193	Students(public)
0	4079	209 66	Students(private)
2942	8493	21 648	Total boys
74.5%	66.3%	58.3%	% of boys
1006	4309	15 465	Total girls
			2.0. Teachers
160	812		Total public and private
160	585		Teachers (public)
0	227		Teachers (private)
146	497		Numbers of males (public)
0	221		Number of males (private)
14	88		Females (public)
			3.0. Institutions
1	10	18	Total public and private
1	6	6	Public institutions
0	4	12	Private institutions

French, and English), and reading and writing to learn. The latter is important if people are to access information easily (life skills, healthcare, etc.).

The revised organizational structure at MINEDUC encourages the linkage between formal and nonformal education (NFE) and places the catch up and adult literacy program within the same NFE unit. This allows any literacy program to be not only a major goal for primary education but also for out-of-school education.

The challenge is to go further by having an intersectoral approach with strong partnerships with faith-based and civil society organizations. This would promote diversified strategies within an over-arching national program for vulnerable groups (rural communities, out-of-school youths, orphans, and vulnerable children and women).

The long-term strategy demands that adult literacy centers be upgraded and that there will be increased production and distribution of diversified literacy material available to adolescents, and younger and older adults. Consideration is also being given to financial incentives and forms of recognition for literacy trainers. Alternative ways of reaching nonliterate persons are important, such as through the media (e.g., radio, television for wide public information), and through the use of school facilities as community resources. An expanded nation-wide literacy program demands adequate funding by the government, sufficient trainers, and materials, as well as a rolled-out national plan within the decentralized context.

New Policy Initiatives

Education in Rwanda is in a dynamic and fast transitional state, and new policies and programs are either under consideration or are being introduced. These incoming policies and programs will affect the strategies and activities during the next years and will change the structure of the education system, besides having important implications for the financing of the education system (Table 4).

Nine-year basic education

This is one of the key emerging priorities. Achieving 9-year basic education for all is one of the main goals of the GoR. It will aim to raise the general level of knowledge and skills in the population, which has the potential to reduce poverty.

The implementation of the 9-year basic education program will demand increased numbers of trained teachers, more learning materials, and more classrooms. The challenge will be to develop effective implementation strategies for a large-scale investment in improved quality, expanded capacity, and managed growth. A major challenge will also be to access the required financial resources.

The following considerations and proposals have been put forward to manage the expansion to the 9-year basic education program:

Table 4 Educational finance

<i>Input indicator</i>	<i>2004</i>	<i>2005</i>	<i>2006 baseline</i>
Education spending as % of GDP ^a	5.0%	5.2%	5.2%
Education spending as % GoR budget ^b	14.6%	13.9%	16.5
Government spending on education as % of total public expenditure ^c	23%	24%	25%
Recurrent expenditure on primary education as % of total recurrent expenditure on education	41%	43%	45%
Recurrent expenditure on Tronc Commun education as % of total recurrent expenditure on education			14%
Recurrent expenditure on upper secondary education as % of total recurrent expenditure on education			6%
Recurrent expenditure on secondary education as % of total recurrent expenditure on education	23%	20%	18%
Recurrent expenditure on higher education as % of total recurrent expenditure on education	30%	29%	29%
Recurrent expenditure per primary pupil (RWF) ^d	9472	10 712	13 653
Recurrent expenditure per Tronc Commun pupil (RWF)	73 884	73 467	70 207
Recurrent expenditure per upper secondary pupil (RWF)	63 150	75 584	86 435
Average expenditure per higher education student	795 418	828 040	958 047
Ratio of higher education to primary education unit cost	86	77	70
STR (IRST) expenditure as % of GDP	0.07%	0.07%	0.08%

^aGDP figures are projections on 5% growth. The base 2004 figure is taken from MASABO (2005) report, MINECOFIN.

^bEducation spending is Government spending on primary, secondary, higher, informal, science, research, technology, and institutions support. This includes the recurrent and development budget of MINEDUC and the provinces.

^cThis recurrent education spending as proportion of the Government budget minus interests payments, debt repayments, arrears, and capital expenditure.

^dThe figures on recurrent expenditure per primary, tronc commun (lower secondary) and upper secondary pupil do not include expenditure on science and technology.

RWF, Rwanda Francs; STR, Science Technology and Research; IRST, Institut de Recherche Scientifique et Technologique.

1. Teacher supply and teacher training: The major strategy is to establish four national colleges of education (each eventually enrolling up to 500 trainees) to train the extra Tronc Commun teachers that will be needed. A new teacher training curriculum will have to be developed in line with the training approach chosen and the needs of new teachers at primary and Tronc Commun level.
2. Curriculum, learning materials, and assessment: These will have to be comprehensively revised within a conceptual framework that provides continuity of learning from grades 1 to 9. Priority will be on the development of a new national core curriculum. This will require decisions concerning which core subjects should be prioritized at primary and secondary levels and what the desired learning outcomes should be. A national assessment strategy linked to curriculum outcomes will have to be developed.
3. School infrastructure development: The increased numbers of pupils will demand extra classrooms and schools. The proposed program will assess the level of underutilization in schools and how much growth can be absorbed through better use of existing structures. However, additional classrooms will need to be built which implies a substantial building program and capital investment.
4. School management and efficiency: Expanded access requires reduction in costs per student if the resource envelop is not to be exceeded. The 9-year basic plan promotes reductions in subsidized boarding and aims to have no more than 10% subsidized boarding at Tronc Commun by 2015 and 20% at upper secondary. Increased efficiency will also be achieved through improved teacher deployment, reduced nonteaching administrative staff, and active and purposeful school improvement plans (SIPs). Management training for education personnel will therefore be required. Effective school monitoring and evaluation systems to track progress and identify potential problem areas will also need to be in place.

The GoR is committed to implementing the 9-year basic education program. Institutional and human resource capacity to implement the program will be critical to its successful implementation and to ensure that expanded access does not detract from the quality of the education in the classroom.

Science and technology in education

This is another key policy with emerging priority along with 9-year basic education for all. The science and technology policy within education is now integrated into the financial framework. Promoting science and technological education is an essential strategy to achieve the human development objectives set out in Vision 2020 and the

economic development and poverty reduction strategy (EDPRS). A national science, technology, scientific research, and innovation policy was approved in July 2005. It forms the basis for an operational integration of the science and technology education program. The policy identifies underdeveloped skills in all areas of science and technology and the lack of resources in teaching institutions to support the teaching of quality science and technology education. The policy proposes an enabling legal and policy framework and its strategy envisages strong partnership with the private sector and the ICT sector. The key objectives of capacity building in science and technology are to achieve sufficient number of students, who are trained to a high level so that graduates from vocational schools, technical schools, and higher technical, professional, and technological institutions meet the development needs of Rwanda. MINE-DUC and the national council for science, technology, and innovation will have coordinating roles and will work with the national council for higher education (NCHE). The development of an ICT curriculum for primary and secondary schools is in progress.

Post-basic education and training

An evolving coherent framework for post-basic education and training (PBET) is under consideration, which includes the re-conceptualization of the present structure of the education system as basic education will soon be P1–P6 + TC1–TC3. This will have consequent implications for the rationalization of the upper secondary curriculum and examinations and for the school building programs.

A major area of focus will be on the coverage or inclusiveness of PBET with regard to professional and technical and vocational training, as well as between formal and nonformal technical education. A coherent framework, backed up by adequate data, is to be developed for management and planning purposes to ensure that incoming policies, projects, and programs will be aligned with a coherent approach to the subsector as a whole. The PBET policy/strategy development will contribute to Rwanda's technical and education policy and strategy and to its national skills development.

Open, distance, and e-learning

There is a draft policy for open, distance, and e-learning (Ode-L) which promotes the expanded use of distance learning in all subsectors (primary, secondary, and tertiary). At present, the main distance education program, which has been developed in country, is for under-qualified secondary teachers (at Kigali Institute of Education (KIE)). The Kigali Institute of Science and Technology (KIST) and the National University of Rwanda (NUR) also have a number of programs offered through the African Virtual University (AVU).

The main challenges to be faced in the deployment of distance and e-learning in Rwanda will be that of developing a vision of the way in which blended learning solutions can be used to develop educational provision in Rwanda, and of developing realistic media and technology choices and access to technology. Building up capacity in distance and e-learning methods will thus be important. It will also be necessary to guarantee the quality and consistency of standards, and any open and distance learning strategy will have to design appropriate qualifications and credit structures that will embrace both face-to-face and distance learning.

The challenges and risks are many including the need for a strong change in the management process in existing institutions to re-gear to a distance learning approach – for example, logistical management (i.e., development, production and delivery schedules, and operations). The need for an effective quality assurance system to ensure the development of appropriate, high-quality materials produced in-country and for strong student support systems (tuition, help desks, etc.) will be important. A major program design consideration will be how to ensure that students acquire the necessary practical experience which is required for some courses.

Girls education

An education policy for girls is to be developed; a draft consultancy report is to be discussed at the JSR 2006. Recommendations have been made to establish a national task force for the coordination of girls education in Rwanda and to include girls education into the MTEF. A communication strategy for girls education is also being proposed so that the various partners (students, teachers, administrators, parents, the private sector and partners, and donors) can address particular issues and harmonize their efforts. MINEDUC further intends to coordinate a girls education movement to sensitize students at all levels in the formal and nonformal system, as well as those not in school.

Key areas of proposed intervention will be to ensure that girls have access to adequate sanitation in schools and that girls should be protected so that they can study in an environment free from abuse. There are also plans to scale up the number of Tuseme clubs in secondary schools so as to empower girls by helping them analyze issues affecting their education and come up with relevant solutions.

The quality of girls education is also a major area of focus in terms of improving the retention rate of girls in post primary and tertiary education, and especially for girls from poor households. Increased consideration of gender issues in education through training programs for all teachers is required if increased number of girls are to enter into maths, science, and technology subject areas.

Cross-cutting issues

HIV/AIDS, and hard-to-reach children are interlinked cross-cutting issues and they correlate strongly to poverty alleviation. They are major determinants of strategic choices – not only for education but for all social services.

The Government of Rwanda has been a strong advocate in ensuring that hard-to-reach children access education through flexible pilot complementary or catch-up programs. A major strategy has been to elaborate what should constitute a minimum learning package to achieve gender best practices, and the integration of life skills and HIV/AIDS education.

Special education needs

There is no policy for special education needs (SEN) but a study has been completed and will help produce a policy and long-term strategy, which will require incorporation into the annual operation plans of all subsectors given that some 10% of all students suffer from some form of disability. Most of the current SEN activities are run by NGOs and churches as there are only five educational centers for children with hearing, visual, physical, and mental disabilities and only one is at secondary school level. Inevitably, SEN activities are restricted to a small number of students with disabilities.

Peace and reconciliation

Education at all levels (primary, secondary, and tertiary) is an important means of addressing issues of peace and reconciliation in the context of Rwanda post-1994. National and individual values emphasizing peace, harmony, and reconciliation will infuse the revision of the history and civic education primary curricula for which the national curriculum development center (NCDC) is seeking resources. The guidance and counseling curriculum is also to be undertaken and will explore innovative ways of counseling for trauma healing of children at primary and secondary levels and to ensure resources in schools.

Private sector involvement

Under MINEDUC's sector-wide approach to planning in the new ESSP, it is recognized that the nongovernment sector must play a crucial role if the dual education aims are to be achieved. That is, all children will have access to 9 years of basic education. Management of the nongovernment sector's expansion and the quality of its provision in a fiscally sustainable manner is a critically important policy issue especially when the proposed route is one whereby the GoR intends to mix government finance with private delivery and private

ownership of a considerable proportion of the post-primary education service.

The Teaching Profession

There are 11 primary school TTCs. Initial and in-service training of secondary teachers is concentrated in two centers: one is at the National University of Rwanda (NUR), the other is KIE. NUR's annual output is approximately 60 teachers. KIE produces at least 300 new teachers for the Tronc Commun (general lower secondary) since the end of 2003. In 2001, a distance-training program (DTP) started at KIE to upgrade underqualified secondary teachers. This new program has ten satellite distance training centers around the country. The first 500 unqualified secondary teachers following the distance-training program graduated in 2006. The objective is, in the long term, to have every teacher qualified. According to administrative data for 2001/2, only 52% of the 6368 teachers in the secondary system were qualified. Of these teachers, approximately 1200 are women of whom only about 25% are qualified. Teacher/pupil ratios at secondary level are approximately 1:24.

The objective of the teacher development and management (TD&M) policy, linked to 9-year basic education, is to strengthen accountability for improving teacher quality in all basic education schools in Rwanda. It forms part of MINEDUC's contribution to Vision 2020 and to reducing poverty. It endorses measures to improve teachers' working conditions and status and affirms that teachers at all levels will be trained in sufficient numbers. The TD&M policy aims to promote a range of training approaches and it supports learner-centeredness. An important component will be to provide incentives to teachers veering them toward continuous professional development.

The policy on teacher development envisages teachers progressing through three professional stages – newly qualified, probation of 3 years, and post-probation with emphasis on continuous development. Core teaching values and competence profiles will be structured according to areas of competencies and will lead to the establishment of a coherent teacher education curriculum and the introduction of a range of new qualifications framework. School-based training, at primary and secondary levels, supported by TTCs and colleges of education (CoE) links, and on-the-job mentoring will become increasingly important. Teacher motivation at all levels is an important consideration.

The problem of teacher demand and supply as key policies like the 9-year basic education are implemented is given special attention. Two strategies are planned:

- First is to upgrade some TTCs to diploma-offering (A1) CoE to supplement KIE in training teachers for Tronc Commun.

- Second is all institutions to embark on internal efficiency mechanisms to ensure optimal production of teachers.

Two important strategies are emerging in teacher management and motivation. These are establishment of a task force to lead the process of establishment of the teachers' service commission. The second one is the establishment of the teachers' cooperative.

Monitoring Education Sector Performances

The process of accountability, to which the GoR is committed, dictates that there should be a joint review of the education sector (JRES) on a regular basis (annually for the foreseeable future). The JRES will review the education sector performance which is led by MINEDUC. It is conducted in concert with all internal and external development partners. In terms of the monitoring of education performance, the purpose of the JRES is to ensure that there are effective returns on the investments being made in the education sector and that the intended beneficiaries, that is, students, children, parents, and all other stakeholders, are all indeed benefited.

The first JRES in April–May 2003 signified the start of the review process. The performance of the sector is evaluated against commissioned reports, inspection reports, sector indicators, target completion, and visits to educational establishments. Working in tandem with their development partners, the performance review will inform future policy, future work programs, and the ongoing development of the ESSP itself (Table 5).

The JRES will be arranged to complement the annual progress report and the budgetary cycle so that informed decisions can be taken. Likewise, the review will take place prior to the revision of the annual operation work plans (AOWPs.) The review will cover whole sector performance and will, of necessity, cover all aspects of annual educational development, including projects and sub-programs.

The ESSP will serve to guide the formulation of the AOWPs and to provide a framework for the monitoring and assessment of achievements. Its purpose is to elaborate GoR's vision through policy objectives. These objectives are realized through outlined key activities within broad budget indications. The process is shown in Figure 1. The strategies and key activities must therefore be realistic and the indicators, standardized monitoring and assessment of relief and transition (SMART) if they are to be useful. Evidence of success includes qualitative indicators that require the application of professional judgment to assess the extent to which they have been achieved. The combination of measurable and qualitative indicators will make monitoring and evaluating progress a relatively straightforward task.

Table 5 Some performance indicators for Rwanda

<i>Input indicators</i>	<i>2004</i>	<i>2005</i>	<i>2006 baseline</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>2010</i>
Education spending as % of GDP ^a	5.0%	5.2%	5.2%	5.4%	5.6%	5.9%	6.1%
Education spending as % of GoR budget ^b	14.6%	13.9%	16.5%	17.0%	17.5%	18.0%	18.5%
Government spending on education as % of total public expenditure ^c	23%	24%	25%	25%	25%	25%	25%
Recurrent expenditure on primary education as % of total recurrent expenditure on education	41%	43%	45%	46%	46%	49%	46%
Recurrent expenditure on secondary education as % of total recurrent expenditure on education	23%	20%	18%	21%	20%	21%	22%
Recurrent expenditure on higher education as % of total recurrent expenditure on education	30%	29%	29%	26%	22%	20%	20%
Recurrent exp per primary pupil (RWF) ^d	9,472	10,712	13,653	15,770	18,230	21,100	24,165
Recurrent exp per Tronc Commun pupil (RWF)	73,884	73,467	70,207	71,306	72,406	73,518	74,281
Recurrent exp per upper secondary pupil (RWF)	63,150	75,584	86,435	94,392	102,874	108,845	121,344
Average exp per higher education student (RWF)	795,418	828,040	958,047	976,082	978,980	977,548	973,843
Ratio of higher education to primary education unit costs	86	77	70	62	54	46	40
STR (IRST) expenditure as % of GDP	0.07%	0.07%	0.08%	0.09%	0.10%	0.11%	0.12%
<i>Output indicators (primary)</i>	<i>2004 baseline</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>2010</i>
Pupil: teacher ratio ^e (not including head teachers)	66.9	69.0	63.5	61.8	60.2	58.7	55.6
Male qualified (%)	87%	94%					
Female qualified (%)	89%	94%					
Number of years schooling for primary graduate	11.5					6.6	
Class size ^f	51	51	51	51	51	51	50
Non-salary as % of recurrent spending	31.1%		38.1%	41.5%	45.0%	45.0%	45.0%
<i>Output indicators (Tronc Commun)</i>	<i>2004 baseline</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>2010</i>
Pupil-teacher ratio	30		30	30	31	31	31
Non-salary as % of recurrent spending			65%	64%	62%	60%	59%
% Pupils boarding	58%		41%	35%	30%	25%	21%
<i>Output indicators (upper secondary)</i>	<i>2004 baseline</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>2010</i>
Pupil-teacher ratio	26	26	26	27	27	26	26
Number of teachers (including non-public)	2885	3113	3278	3348	3415	3487	3615
Number of classrooms (public and L.S only)	896	940	975	1030	1093	1164	1244
Pupil-classroom ratio	40	40	40	40	40	40	40
Non-salary as % of recurrent spending			60%	60%	60%	60%	60%
% Pupils boarding	75%	72%	69%	66%	63%	60%	55%
<i>Output indicators (higher)</i>	<i>2004 baseline</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>2010</i>
Number of academic staff with Masters qualification							
Number of academic staff with PhD qualification							
Student/lecturer ratio							
% of ex-pat lecturers							
<i>Outcome indicators (primary)</i>	<i>2004 baseline</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>2010</i>
Gross enrolment rate (%) ^g	129%	136%	140%	141%	142%	142%	137%
Gross entry rate (%)	195%	206%	163%	149%	137%	125%	120%
Net enrolment rate (%)	92%	92%	94%	94%	95%	96%	97%
Primary completion rate (World Bank method)	42%		65%	75%	85%	109%	122%
Primary completion rate (UNESCO method)	51%		74%	84%	94%	119%	132%
Average repetition rate	19%		15%	13%	11%	10%	9%
Average drop out rate	14%		12%	10%	8%	5%	5%
Transition to Tronc Commun ^h	60%		47%	46%	48%	42%	43%

Continued

Table 5 Continued

<i>Outcome indicators (primary)</i>		<i>2004 baseline</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>2010</i>
Pupil-qualified teacher ratio		75.8	73.6					
Learning achievement in core subjects (MLA and or/national assessment test scores)								
<i>Outcome indicators (Tronc Commun)</i>	<i>2004 baseline</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>2010</i>	
Gross enrolment rate	20%	23%	26%	28%	30%	33%	36%	
Net enrolment rate								
Completion rate	14%	16%	21%	23%	25%	27%	30%	
Average repetition rate	9%	9%	8%	8%	7%	6%	6%	
Average drop-out rate								
Transition rate to upper secondary		91%	83%	67%	63%	60%	57%	
Pupil-qualified teacher ratio								
Exam performance								
<i>Outcome indicators (upper secondary)</i>	<i>2004 baseline</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>2010</i>	
Gross enrolment rate	11%	12%	13%	13%	14%	14%	15%	
Net enrolment rate								
Completion rate	8%	10%	11%	12%	12%	13%	13%	
Average repetition rate	5%	5%	5%	5%	5%	5%	4%	
Average drop-out rate								
Pupil-qualified teacher ratio								
Exam performance/learning achievement								
<i>Outcome indicators (higher)</i>	<i>2004 baseline</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>2010</i>	
Number of student places provided								
Gross enrolment rate								
Net enrolment rate								
Average repetition rate								
Average drop-out rate								
Completion rate								
Student-qualified lecturer ratio								
<i>Outcome indicators (STR)</i>	<i>2004 baseline</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>2010</i>	
Number of science students – upper secondary								
Number of science students – higher education								
Number of engineers qualified?								
<i>IMPACT indicators</i>	<i>2004 baseline</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>	<i>2008</i>	<i>2009</i>	<i>2010</i>	
Youth literacy rate (15–24-year olds)	84%							
Average literacy (15 and above)	60%							

^aGDP figures are projections on 5% growth. The base 2004 figure is taken from MASABO (2005) report, MINECOFIN.

^bEducation spending is Government spending on primary, secondary, higher, informal, science, research, technology, and institutions support. This includes the recurrent and development budget of MINEDUC and the provinces.

^cThis recurrent education spending as proportion of the Government budget minus interests payments, debt repayments, arrears, and capital expenditure.

^dThe figures on recurrent expenditure per primary, tronc commun (lower secondary) and upper secondary pupil do not include expenditure on science and technology.

^eThe target is that Pupil-Teacher Ratio is to fall, however, it has been rising since 2001. Head teachers are not included in Pupil-Teacher ratio (PTR).

^fClass size is smaller than PTR and Pupil Classroom Ratio because of double shifting.

^gGross enrolment rate (GER), Net enrolment rate (NER) and completion rate vary slightly from previously published statistics because of updated base population projections.

^hincludes transition to private tronc communs schools. Although this is projected to fall, tronc commun enrolment is still rising significantly.

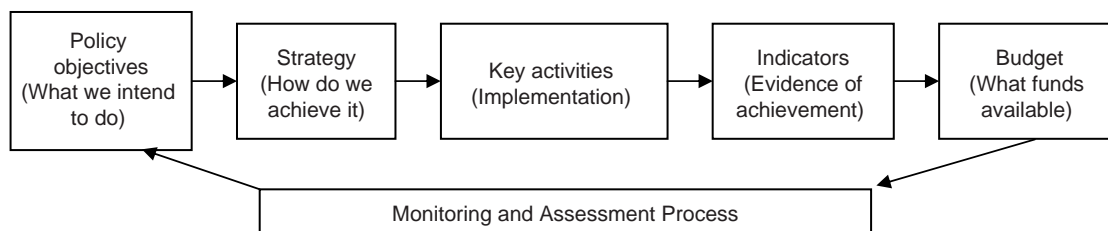


Figure 1 Policy and strategies – process.

Further Reading

Rwanda (2000). *Démographique de Santé (EDS)*. Geneva: UNAIDS.

Rwanda (2001). MINEDUC administration data. Kigali: Ministry of Education.

Rwanda (2002). *Education Sector Policy*. Kigali: Ministry of Education.

Rwanda (2002). *3rd General Census of Population and Habitat (RGPH), the Integral Survey on Households Living Conditions (EICV)*. Kigali: Government of Rwanda.

Rwanda (2002). *Synthesis of the 2002 Census*. Kigali: Government of Rwanda.

Rwanda (2003). *Country Status Report*. Kigali: CSR, MINEDUC/World Bank.

Rwanda (2003). *Millennium Development Goals: Status Report*. UN MDGs.

Rwanda (2003). *WEANDA Politique Nationale de Santé*. Kigali: Government of Rwanda.

Rwanda (2005). *Education spending is government spending on primary, secondary, higher, informal, STR and institutions support. MASABO Report, MINECOFIN*. Kigali: Ministry of Finance and Planning.

UNESCO (2006). *Fact Book on Education for All*. Nairobi: UNESCO Nairobi Office.

Scotland

D Raffe, University of Edinburgh, Edinburgh, Scotland

© 2010 Elsevier Ltd. All rights reserved.

Glossary

CLD – (Community Learning and Development). Diverse sector of postschool learning including community-based nonformal learning opportunities.

Curriculum for Excellence – Reform of 3–18 curriculum.

HEI – Higher education institution.

HMle – Her Majesty's Inspectorate of Education.

HNC/D (Higher National Certificate/Diploma) – Short-cycle higher education qualifications.

National Course – Subject-specific courses available at several levels leading to National Qualifications awarded by the SQA, delivered mainly in schools.

P1–P7 – Six stages (years) of primary school.

S1–S6 – Six stages (years) of secondary school.

SCQF – Scottish Credit and Qualifications Framework.

SQA – Scottish Qualifications Authority.

Standard grade – Subject-specific qualification awarded at end of compulsory schooling, to be phased out from 2014.

General Background

Scotland is a country, of five million inhabitants, which occupies the northern third of Great Britain. It is less densely populated than the rest of Britain but a large proportion of its population lives in the central belt which includes the two largest cities of Glasgow and Edinburgh.

Scotland was an independent country before 1707, when it joined with England and Wales to form Great Britain, under a single parliament. In 1801, it became part of the United Kingdom (UK) of Great Britain and Ireland (now Northern Ireland). The Scottish education system was already comparatively developed in 1707 and it preserved its distinct identity within the Union. Its historical links with the establishment of a national protestant church have stamped it with features which persist to this day, including institutional uniformity, centralized state control, inclusiveness, and an academic ethos.

In 1872, the Scotch Education Department was created to administer Scottish schools, and in 1885 this was

incorporated into the Scottish Office, a territorial department of the UK government responsible for certain domestic policy areas as they related to Scotland. The Scottish Office took responsibility for the vocational higher education institutions (central institutions) and further-education colleges which were established in the following decades. In the 1980s and 1990s it took over further policy responsibilities including universities and training. By 1999, education and training policy in Scotland was almost wholly the responsibility of the Scottish Office which, although part of the UK government, exercised significant autonomy. In 1999, the Scottish Parliament was created with legislative powers over education. The Scottish Office was relabeled the Scottish Executive (and subsequently, in 2007, the Scottish Government), led by Scottish ministers accountable to the parliament. The first two administrations (1999–2007) were formed by coalitions of the Scottish Labour Party and the Scottish Liberal Democrats; after the 2007 election, the Scottish National Party formed a government with a minority of seats in the Parliament.

The main languages are English, Scots (whose status as a separate language is contested), and Gaelic, together with languages spoken by various immigrant communities. Gaelic, once the majority language of Scotland, is now spoken by only 1% of its population. Scotland has traditionally been a country of emigration, with a net outflow of around 40,000 people per year in the mid-1960s, but this pattern has reversed and in the mid-2000s, there was a net inflow of more than 20,000 per year. These include migrants from new member states of the European Union, who have changed the ethnic composition of schools in many areas.

The last decades of the twentieth century saw a huge decline in employment in the coal, shipbuilding, and steel industries. The Scottish economy is now primarily based on services; financial services, tourism, health, and education are important sources of employment. Fishing and oil extraction are important in the regional economy of the north-east. The UK labor market is flexible, with weak regulation and weak occupational markets compared with other European countries. The Scottish unemployment rate stood at 4.9%, slightly below the UK average, in the 3 months to November 2007, but this figure masks significant regional variations, with unemployment black spots in the old industrial areas of the west.

Scotland is a rich, developed country but it includes areas of poverty and social deprivation. An index of multiple deprivation applied to data zones (small areas) in

the 2001 census showed that more than half of the zones in Glasgow belonged to the 15% most deprived zones nationally. Inverclyde, Dundee, West Dunbartonshire, and North Lanarkshire also had disproportionately high numbers of deprived zones.

Goals of the Education System

The general goals of education are not defined in legislation. The Standards in Scotland's Schools, etc. Act (2000) lays down the duty of (local) education authorities to provide education "directed to the development of the personality, talents, and mental and physical ability of the child or young person," but it does not address broader questions of national purpose. The Act empowered governments to define national priorities in education – in 2001, the government identified broad priorities under five headings: achievement and attainment, framework for learning (the skills of teachers, self-discipline of learners, and school environments), inclusion and equality, values and citizenship, and learning for life. Commentators typically describe the Scottish educational tradition as inclusive, meritocratic, academic, collectivist, and conservative; education tends to be more highly valued than elsewhere in the UK. However, the inquiry into the purposes of Scottish education by the Scottish Parliament's education committee noted that questions of philosophy and purpose had been ignored in debate about Scottish educational policy for several decades. It found widespread support for a broad view of educational purposes including the promotion of positive values and active citizenship. There was wide agreement that a merely utilitarian education is insufficient. In 2004, a government-appointed review group defined the purposes of the 3–18 curriculum as to enable all young people to be successful learners, confident individuals, responsible citizens, and effective contributors to society and at work. These four capacities became the guiding purposes of a large curriculum-reform program, A Curriculum for Excellence.

These statements refer primarily to school education. Another Scottish Parliament committee report, on lifelong learning, identified objectives under four headings: economic development, social justice, citizenship, and quality.

Structure and Operation of the Education System

Preschool Education

Since 2002, free, part-time preschool education has been available to all 3- and 4-year-olds whose parents wish it. In 2007, some 95% of 3-year-olds and 99% of 4-year-olds were registered with local authorities for preschool education,

but these figures double-count those who attended two centers. They attend a variety of centers including nursery classes in primary schools, stand-alone public nurseries, and voluntary and private providers, many of which receive public support. Centers aim to provide a variety of experiences affording opportunities for active learning, including purposeful play.

Primary and Secondary School

Schooling is compulsory from ages 5 to 16. Children usually enter primary school at (or soon after) the beginning of the school session in August, and they move up through 7 years of primary school (P1–P7) and up to 6 years of secondary (S1–S6). They remain with the same year group; grade repeating is rare. In 2006, there were 733,000 pupils in Scottish schools; pupil numbers were declining and were projected to fall by a further 12% by 2016.

About 4% of pupils attend private schools. This percentage is higher among secondary than primary pupils, and highest in Edinburgh where more than 20% of secondary pupils attend private schools. The remainder attend public schools run by the 32 elected local authorities. In 2006, there were 2184 primary schools, 381 secondary schools, and 190 special schools within the public sector. Schools serve defined catchment areas. Children are entitled to a place in their local school, but parents may choose another school if places are available. In 2005–2006 some 30,000 placement requests were received, of which 84% were granted. Most requests were for pupils entering primary school (22% of the P1 cohort) or transferring to secondary (14% of the S1 cohort).

Scottish schools are institutionally uniform. All publicly funded schools are comprehensive, coeducational (with one exception), community based, and nonselective. Slightly more than a fifth of pupils attend denominational schools, nearly all Roman Catholic, which have been fully integrated within the public system since 1918. Scottish secondary schools are less socially segregated than most other European systems and school effects vary little from school to school.

There is no statutory curriculum and local authorities have formal responsibility for the school curriculum. However, they all support national guidelines for the 5–14 curriculum, which define outcomes and recommend approximate time allocations across five broad areas: language, mathematics, environmental studies (science, technology, geography, history, and modern studies), expressive arts, and religious and moral education. The guidelines specify six levels of performance (A–F). Each level is expected to be achieved by a majority of pupils by a given school year, but teaching methods allow pupils to work through the levels at different rates. From 2010 the 5–14 framework is being replaced by A Curriculum for

Excellence, which specifies less-detailed outcomes and experiences in eight curriculum areas at five levels from 3 to 15 years.

The secondary curriculum has traditionally followed a 2 + 2 + 2 pattern, with a common curriculum in S1–S2 followed by courses in S3–S4 leading to subject-based standard-grade qualifications and up to two postcompulsory years in S5–S6. A Curriculum for Excellence proposes to replace this with a 3 + 3 pattern, with a new senior phase starting in S4. Standard grades are to be replaced with new qualifications, but the phasing of courses and qualifications across the 3 years of the senior phase is uncertain and may vary across schools. Opportunities for subject choice increase during the secondary years. The school curriculum is predominantly general but a growing number of vocational options are available from S3 onward. Most pupils are eligible to leave school after S4 or the first term of S5. In 2006, 67% of the year group stayed on beyond the first term of S5 and 44% stayed on to S6. S5 students typically take up to five single-subject national courses, available at different levels with entry determined by achievement in the corresponding standard grade; in S6 they may take further national courses in additional subjects or in the same subjects at a higher level.

Education authorities have the responsibility to identify, and provide for, pupils who require additional support to benefit from their education, for example, due to family circumstances, health, disability, or emotional and behavioral factors. In 2006, 36,148 children in publicly funded schools had additional support needs. Current policies favor meeting their needs within mainstream schools, where appropriate, but about one-fifth were in special schools.

Colleges

The 43 colleges of further education are the largest providers of lifelong learning. Much of their provision is vocational, but they also offer general and academic courses and more specialized opportunities such as literacy and numeracy courses and provision for learners with disabilities. The colleges have a tradition of access and responsiveness. Their programs vary widely in level, duration, and mode of study. There were 351,435 students enrolled on college courses in 2004–2005. A majority were aged over 25, but colleges also recruit young people. Nearly one in four school leavers entered a full-time college course in 2006; others entered apprenticeships or training which included part-time college study. Colleges increasingly provide vocational courses for school pupils aged 14–16, and each school is expected to have a partnership with at least one college. Colleges are also important providers of higher education, primarily

through subdegree courses such as higher national certificate and diploma (HNC/D) courses.

Higher Education

At the end of 2007, there were 20 higher education institutions (HEIs): 15 universities (including the Open University in Scotland), four specialist colleges, and the University of the Highlands and Islands (UHI) Millennium Institute, which offers higher education across north-western Scotland through local colleges and other providers. Universities are informally distinguished on the basis of their age: four ancient universities were founded before 1600, four old universities were created before 1992, and six new universities created since 1992 are former government-controlled central institutions. HEIs in these three categories tend to vary with respect to the age and social origins of students, the balance of full- and part-time study, the numbers of postgraduate students, and the emphasis on research.

Higher education expanded massively in the late twentieth century. The proportion of young Scots entering higher education (including subdegree courses) for the first time surpassed 50% in 2000, although it fell back to 47% in 2005. Participation by older students has also grown. In 2005, 21% of first-degree students were aged 25 or over and 32% were aged between 21 and 24; 60% were female. Of the 285,000 higher education students in Scottish institutions in 2005, 72% were from Scotland and another 10% from the rest of the UK. The remainder came from other countries, led by China, the US, India, and Malaysia. The number of students leaving Scotland to study is not known, but within the UK, the inward flows are much larger.

Other Education and Training

The largest publicly funded work-based learning program is modern apprenticeship (MA). In March 2007, there were 28,626 modern apprentices in training, of whom 73% were male and 65% were aged under 20. Other programs include skill seekers (lower level training for 16–18-year-olds, to be absorbed within the MA program), Get Ready for Work (for 16–18-year-olds with additional needs), Training for Work (for the long-term unemployed aged 25-plus), and New Deal (a UK program for specific groups including long-term unemployed 18–24-year-olds).

The term community learning and development (CLD) refers to informal learning and social development work with individuals and groups within their communities. This diverse sector includes youth work, adult learning in the community (including adult classes, literacy and numeracy classes, parenting education, and family learning), and

building community capacity. It is provided by local authorities and partner agencies in the public and voluntary sectors, and coordinated by local partnerships.

Qualifications

The Scottish Qualifications Authority (SQA) is the main awarding body for nonuniversity qualifications. These include National Qualifications awarded in schools and (nonadvanced) college education, HNCs and HNDs, and occupational Scottish Vocational Qualifications (SVQs) mainly gained in work-based training. The SQA also accredits SVQs awarded by other bodies. All SQA qualifications are unitized (except standard grades) and placed in the Scottish Credit and Qualifications Framework (SCQF), a nonregulatory framework for all qualifications in Scotland. Qualifications within the SCQF, and their component units, are rated in terms of level and credit volume. The SCQF aims to promote access and to encourage transparency and understanding of the system. It underpins articulation and credit transfer between sectors; for example, a relevant HND may allow progression to the third year of a 4-year degree course.

The Teaching Profession

In 2006, there were 51,658 full-time equivalent (FTE) teachers in publicly funded schools, and a further 3300 in private schools. Three-quarters were female. The pupil-teacher ratio was 16.3 in public primary schools and 12.0 in public secondary schools. A further 20,000 FTE staff worked in various support roles, including some 5700 classroom assistants mostly in primary schools. All school teachers must hold a teaching qualification and be registered with the general teaching council for Scotland. Full registration is achieved after completing a university course (either a 4-year education or combined degree, or a 1-year postgraduate course) and a period of probation. A new career structure and conditions of service for school teachers were introduced after 2001 following an agreement between the government and teacher unions, *A Teaching Profession for the 21st Century*.

In 2005–2006 there were 13,157 payroll teaching staff in colleges, of whom 4771 were full time. About one-half were female. Most full-time staff hold the teaching qualification (further education), or an equivalent qualification, taken after entry to the profession; under current proposals, all new permanent full-time and part-time lecturers are expected to gain it within 3 years and 5 years, respectively, of taking up post. In 2006–2007, there were 12,645 full time and 3370 part-time academic staff in HEIs, including those funded from external grants and contracts. There are no general qualification requirements for teaching in HEIs.

Administrative and Supervisory Structures

Legislative powers over education and training were devolved to the new Scottish Parliament in 1999. In 2007, the Scottish Government was organized into five main directorates-general; the directorate-general for education comprises three directorates, responsible respectively for schools, for lifelong learning, and for children, young people, and social care. Certain functions are delegated to nondepartmental public bodies, including the SQA, the Scottish Funding Council (SFC) which funds colleges and HEIs, and Learning and Teaching Scotland which is responsible for school-curriculum development and support.

Publicly funded schools are administered by the 32 local authorities, often through a department which also covers children's or community services. Since the 1980s, significant powers have been devolved to schools, but local authorities retain an important role. School parent councils were introduced in 2007 to represent parents, replacing a more formal system of school boards.

Most colleges are self-governing institutions, led by a board which includes community and industry representatives. Universities are private bodies; like colleges, they have charitable status and are governed by a court or board of governors with substantial lay membership. In the four ancient universities, the court is chaired by a rector elected by students and staff.

From 1990 to 2008, Scottish Enterprise and Highlands and Islands Enterprise, nondepartmental public bodies responsible for economic development in their respective areas of operation, were responsible for MAs and other government-supported training. As of 2008, Scottish Enterprise's training role, along with other skills-related responsibilities, was being transferred to a new body, Skills Development Scotland.

Educational Financing

Public preschool and school education is funded by local authorities out of annual grants from the Scottish Government and local property-based taxes. Headteachers have substantial budgetary autonomy over day-to-day spending. In 2006–2007, gross revenue expenditure was £1683 million (£4403 per pupil) on primary education, £1912 million (£6120 per pupil) on secondary education, and £464 million on special education in special and mainstream schools. About £2.9 billion of expenditure was devolved to headteachers. Total capital expenditure not funded from revenue was £385 million. Further capital spending was supported by arrangements which combine private and public financing.

Most public funding for colleges and HEIs is provided through annual grants from the SFC, following guidance from the government. In 2006–2007 grants to colleges and HEIs totaled £660 million and £1116 million, respectively. A large part of colleges' funding is based on a weighted measure of student numbers. Colleges also receive fees from some students and other customers such as employers. HEIs receive separate allocations for teaching and research. In 2005–2006, 41% of HEIs' income was from SFC grants; a further 21% was from tuition fees and education grants and contracts, 19% from research grants and contracts, and 19% from other sources including accommodation and other services.

School education and full-time, nonadvanced college courses are free. Colleges may provide means-tested bursaries to support students' maintenance costs, travel, and study expenses. In 2006–2007, 38,760 young people received an education-maintenance allowance of up to £30 per week, available to 16–18-year-olds from low-income families in full-time education. Older students on low incomes may receive an individual learning account of up to £200 to support part-time study.

Maintenance grants for UK higher education students were replaced by a system of loans during the 1980s and 1990s. In 1998, tuition fees were introduced for the first time, initially set at £1000 per annum (later increased to £3000 elsewhere in the UK). In the early years of the Scottish Parliament, upfront fees in Scotland were replaced by a one-off contribution of £2000 to a fund to support bursaries, collected through the tax system after graduation. In 2007, the new government introduced legislation to abolish this contribution.

Performance Monitoring, Evaluation, and Research

Her Majesty's Inspectorate of Education (HMIe), a Scottish Government agency, is responsible for evaluating quality and supporting improvement in schools, colleges, initial teacher education, and CLD. It also evaluates children's services and the education function of local authorities. There is a strong emphasis on institutional self-evaluation based on a framework published by HMIe, *How Good is our School?* Quality-improvement officers in each local authority challenge and support schools' improvement activities. Standard performance data are prepared for each school.

Performance in selected school subjects is monitored at a national level by the Scottish Survey of Achievement which samples pupils in selected year groups. Scotland participates as an 'adjudicated region' in the Programme for International Student Assessment (PISA) and in other international surveys. It has consistently performed above the OECD average in PISA with a relatively small proportion of 15-year-olds performing at the lowest levels.

Quality assurance in HEIs is carried out by the Quality Assurance Agency for Higher Education, a UK body with a

separate Scottish office. Since 2003, it has been primarily based on a process of enhancement-led institutional review, which focuses on institutions' self-evaluations and their own policies and practices for quality management. Performance indicators for HEIs, collected on a UK-wide basis, include the age and social background of students, retention, and graduate employment. The SFC collects performance data from colleges including volume of study, age, and social background (based on area) of students; retention and achievement; student satisfaction; and staff qualifications.

Most educational research is carried out by universities and private research organizations. Sources of funding include the Scottish Government and agencies, UK research councils, private foundations, and the European Commission. The Scottish Council for Research in Education, one of the world's first national educational research councils founded in 1928, is now a research center within Glasgow University. Other prominent university centers include the Centre for Educational Sociology (Edinburgh) and the Centre for Research in Lifelong Learning (Glasgow Caledonian and Stirling).

Major Changes and Issues Since the 1990s

The most significant recent change is the establishment of the Scottish Parliament in 1999. Its immediate effect was to place an existing tier of government (the former Scottish Office) under a Scottish legislature, but its longer term effects include further institutional and policy divergence from other parts of the UK and changes in the style of policy-making and administration within Scotland. The first years of the new parliament were an occasion for strategic reappraisal, with a national debate in 2002–2003, major parliamentary enquiries including lifelong learning strategy (2002) and the purposes of education (2003), and a range of strategic government policy reviews including the 3–18 curriculum, lifelong learning, higher education, and the college sector.

In 2006, a government-backed strategic audit identified the education system, along with a strong and well-qualified labor market, as national strengths. It also identified challenges including poor economic growth, poverty, and inequality, young people not in learning or employment, an aging population, a poor health record, and high levels of violent crime. In 2007, the new SNP government made sustainable economic growth its main policy goal and identified five strategic priorities, one of which (for a smarter Scotland) was to expand opportunities for Scots to succeed from nurture through to lifelong learning ensuring higher and more widely shared achievements. Its skills strategy, *Skills for Scotland*, emphasized the importance of skills for economic development.

In 2007, an OECD review of Scottish schooling concluded that Scotland is a well-schooled nation by international standards but identified current challenges

including an achievement gap between children from different socioeconomic backgrounds and low participation and attainment in postcompulsory education. While social inequalities are similar to those in many other OECD countries, the proportion of 16–19-year-olds not in education, employment, or training (12.4% in 2006) is comparatively high. A 2006 report, *More Choices More Chances*, outlined a strategy for reducing this figure, with particular emphasis on improving pupil engagement during compulsory schooling.

The OECD review attributed social inequalities to the uniformly academic ethos and culture of Scottish schools which, it argued, failed to engage learners from disadvantaged backgrounds. It endorsed the inclusive vision of *A Curriculum for Excellence*, the 3–18 curriculum reform, and its goal of a more diverse, personalized curriculum delivered through a wider range of teaching and learning approaches. *A Curriculum for Excellence* is probably the most significant school reform of the early twenty-first century, and it complements other initiatives, for example, to promote formative assessment or to foster school improvement.

Successive governments have promoted increased vocational and prevocational opportunities for young people, but the respective roles of schools and colleges in providing these opportunities is a subject of debate. Following a 2007 review, the Scottish Government identified four strategic purposes for colleges: to provide vocational education and training related to employment needs; to provide positive and clear routes into employment or into HEIs; to develop knowledge and skills to support the confidence of learners; and to encourage employers' demand and utilization of skills.

Current issues for higher education include meeting the country's economic development needs, widening access, and addressing the resource implications of the recent expansion. The policy divergence within the UK – with tuition fees removed in Scotland but increased in England – is seen to threaten the capacity of Scottish HEIs to compete within the UK and internationally.

Finally, a significant recent change is the incremental development of unified arrangements for coordinating a diverse learning system. The best expression of this is the SCQF, which builds on a series of qualification reforms since the 1980s. The SCQF has helped to showcase Scottish education internationally; it has influenced qualifications reform in other countries and it has informed the development of the European Qualifications Framework.

Further Reading

Bryce, T. and Humes, W. (eds.) (2003). *Scottish Education: Second Edition Post-Devolution*. Edinburgh: Edinburgh University Press.
 Eurydice (2007). *The Education System in Scotland: 2006/07*. Brussels: European Commission.

Gallacher, J. (2006). Widening access or differentiation and stratification in higher education in Scotland. *Higher Education Quarterly* 60, 349–369.
 Her Majesty's Inspectorate of Education (2007). *How Good is Our School? The Journey to Excellence, Part 3*. Livingston: HMIE.
 Hutchison, C. (2005). The journey so far: Assessment for learning in Scotland. *Curriculum Journal* 16, 225–248.
 Menter, I., Brisard, E., and Smith, I. (2006). *Convergence or Divergence? Initial Teacher Education in Scotland and England*. Edinburgh: Dunedin Academic Press.
 Munn, P., Stead, J., McLeod, G., et al. (2004). Schools for the 21st century: The national debate on education in Scotland. *Research Papers in Education* 19, 433–452.
 OECD (2007). *Quality and Equity of Schooling in Scotland*. Paris: OECD.
 Paterson, L. (2002). Scotland. In Gearon, L. (ed.) *Education in the United Kingdom*, pp 29–39. London: David Fulton.
 Paterson, L. (2003). *Scottish Education in the Twentieth Century*. Edinburgh: Edinburgh University Press.
 Raffae, D. (2004). How distinctive is Scottish education? Five perspectives on distinctiveness. *Scottish Affairs* 49, 50–72.
 Raffae, D. (2007). Making haste slowly: The evolution of a unified qualifications framework in Scotland. *European Journal of Education* 42, 485–502.
 Riddell, S. (2006). *Special Educational Needs: Providing Additional Support*, 2nd edn. Edinburgh: Dunedin Academic Press.
 Scottish Executive (2001). *A Teaching Profession for the 21st Century. Agreement Reached Following Recommendations Made in the McCrone Report*. Edinburgh: Scottish Executive.
 Scottish Executive (2004). *A curriculum for excellence: Report of the Curriculum Review Group*. Edinburgh: Scottish Executive.
 Scottish Executive (2004). *Ambitious, Excellent Schools: Our Agenda for Action*. Edinburgh: Scottish Executive.
 Scottish Executive (2006). *More Choices, More Chances: A Strategy to Reduce the Proportion of Young People not in Education, Employment or Training in Scotland*. Edinburgh: Scottish Executive.
 Scottish Government (2007). *Review of Scotland's Colleges. Transforming Lives, Transforming Scotland: An Overview by the Review's Core Group*. Edinburgh: Scottish Government.
 Scottish Government (2007). *Skills for Scotland*. Edinburgh: Scottish Government.
 Scottish Parliament (2000). *Standards in Scotland's Schools etc. Act 2000*. asp6. Edinburgh: The Stationery Office.
 Scottish Parliament (2002). *Enterprise and lifelong learning committee: 9th Report 2002. Final Report on Lifelong Learning*. SP Paper 679. Edinburgh: The Stationery Office.
 Scottish Parliament (2003). *Education culture and sport committee: 6th Report 2003, Inquiry into the Purposes of Scottish Education. Vol. 1-Report*. SP Paper 815. Edinburgh: The Stationery Office.
 Simpson, M. (2005). *Assessment*. Edinburgh: Dunedin Academic Press.

Relevant Websites

<http://www.ces.ed.ac.uk> – Centre for Educational Sociology.
<http://www.crl.gcal.ac.uk> – Centre for Research in Lifelong Learning.
<http://www.acurriculumforexcellencescotland.gov.uk> – Curriculum for Excellence.
<http://www.eurydice.org> – Eurydice network (Publishes Information on European Education Systems).
<http://www.hmie.gov.uk> – Her Majesty's Inspectorate of Education.
<http://www.LTScotland.com> – Learning and Teaching Scotland.
<http://www.scotland.gov.uk> – The Scottish Government.
<http://www.scottish.parliament.uk> – The Scottish Parliament.
<http://www.scqf.org.uk> – The Scottish Credit and Qualifications Framework.
<http://www.scre.ac.uk> – The Scottish Council for Research in Education.
<http://www.sfc.ac.uk> – Scottish Funding Council.
<http://www.sfeu.ac.uk> – Scottish Further Education Unit.
<http://www.sqa.org.uk> – Scottish Qualifications Authority.

Singapore

P T Ng, Nanyang Technological University, Singapore

© 2010 Elsevier Ltd. All rights reserved.

General Background

Singapore is a small island country at the southern tip of the Malay Peninsula. With an area of 660 Km² and a current total population of about 4.5 million, of which 3.6 million people are residents (citizens and permanent residents), she is one of the most densely populated countries in the world.

The modern history of Singapore began in 1819 when Englishman Sir Thomas Stamford Raffles established a port on the island and put the island under British colonial rule. Gaining importance through the India–China trade and the entrepot trade in Southeast Asia, Singapore rapidly became a major port city and a trade center. This growth came to a halt during World War II when Singapore was conquered and occupied by the Japanese from 1942 to 1945. After the surrender of the Japanese, Singapore reverted to British control, but with an increased level of self-governance.

From 1955 to 1963, Singapore gained internal self-governance from the British, though the latter still held on to defense and foreign affairs. In 1963, Singapore broke completely away from the British by formally merging with Malaya, Sarawak, and Sabah to form the Federation of Malaysia. However, the merger was far from smooth. There was much social unrest, ethnic tension, and many political disputes between Singapore and Malaysia. This resulted in the separation of Singapore from Malaysia on 9 August 1965. On that day, Singapore became a sovereign independent nation.

Facing severe challenges such as unemployment, housing crisis, and social tensions, Singapore embarked on a modernization program that focused on industrialization, developing public housing, and investing heavily in public education. The efforts paid off and Singapore's economy grew rapidly. Singapore's industrial structure developed from a basic production industry during the 1960s to knowledge-based industries and high-value services in 2000 and beyond. There has also been a strong push for reforms aimed at equipping workers with the competencies seen to be essential to improving productivity and competitiveness in the global knowledge society. By the 1990s, the country has become one of the world's most prosperous nations, with a highly developed free-market economy, strong international trading links, and one of the highest per capita gross domestic products (GDPs) in the world. In 2006, total GDP stood at S\$210 billion, per capital indigenous gross national income (GNI) at

S\$45 billion, and total foreign reserves at S\$209 billion. In the same year, the labor force was 2.6 million people, with 1.9 million residents (73%). The unemployment rate was a low 2.7%.

The residents in Singapore comprised 75% Chinese, 14% Malays, 9% Indians, and 2% of others in 2006. The early residents (prior to 1965) were mostly immigrants. Due to the diversity of ethnic languages and cultures, and the historical roots of Singapore as a British colony, English became the main language in many official and business activities, including education.

Goals of the Education System

The Singapore education system has always been an important vehicle for social and economic development. Since independence, while education policies have changed to reflect the emphasis of the times, two basic goals have not changed: education develops the individual and educates the citizen.

In terms of developing the individual, education in Singapore aims to develop the child intellectually, physically, morally, socially, and esthetically. There are compulsory components in the school curriculum to address each of these facets. In terms of educating the citizen, education in Singapore aims to teach the child to identify Singapore as home – the place to live in, improve, and defend. A good citizen fulfills his obligations and responsibilities to the nation.

One critical education policy to support the achievement of these two goals is bilingualism. The government adopted English and three local languages (Chinese, Malay, and Tamil) as official languages in 1956 and started compulsory bilingual education at primary level in 1966 and at secondary level in 1969. The bilingual policy requires students to learn two languages: English and a mother tongue, which may be Chinese, Malay, or Tamil. English is the working language in Singapore and the medium of instruction in schools. The policy is based on the premise that English is the language of international business, science, and technology, and a good command of the language will facilitate global business. Moreover, it is a vehicle of communication among the different races in the country and helps to maintain social harmony. But it is also important for students to learn their mother tongue to give them access to their cultural heritage and help them retain their cultural identity and roots.

In 1998, the Ministry of Education (MOE) developed The Desired Outcomes of Education, a statement which summarized the characteristics that the education system was supposed to develop in a student at different levels of their education (primary, secondary, postsecondary, and tertiary, and potential leaders). Briefly, the students who have gone through the education system have to be creative, entrepreneurial, and have a lifelong habit of learning, be able to think global and yet remain rooted to Singapore. They have to be morally upright, culturally rooted, and yet understand and respect differences. They have to be responsible to the family, community, and country. They have to believe in the principles of multiculturalism and meritocracy, appreciate the national constraints, and yet be able to see the opportunities.

Structure and Operation of the Education System

In 2006, there were 1200 preschools, 354 schools (primary, secondary, and preuniversity), five polytechnics, three colleges of the Institute of Technical Education (ITE) and three local government-funded universities. Out of the 354 schools, there were 172 primary schools, 155 secondary schools, 13 mixed-level schools, and 14 preuniversities. The total enrolment in the education institutions (primary, secondary, postsecondary, and tertiary levels) was 689 000 in 2006. The main educational pathways available to a Singapore student are shown in Figure 1.

Preschool Education

Many Singaporean children (age 3–6) attend a kindergarten or a child-care center prior to formal education in the

primary school. At this stage, the focus is generally on structured play and social development.

Kindergartens in Singapore are private schools run by organizations such as community foundations, religious bodies, social enterprises, and business companies. However, kindergartens have to be registered with the MOE. Child-care centers also offer similar preschool programs to children of the same age group but they are licensed by the Ministry of Community Development, Youth and Sports (MCYS).

In 2006, there were about 1200 kindergartens and childcare centers offering preschool education. Of these, 40% were kindergartens and 60% were child-care centers. An estimated 95% of children in Singapore received preschool education.

Primary Education

In Singapore, the Compulsory Education Act, implemented in 2003, ensures that children (except certain categories of children, e.g., those with special needs) receive a minimum period of education up to Primary 6. According to the act, a child of compulsory school age is one who is above the age of 6 years and who has not yet attained the age of 15 years. Parents have the responsibility for sending children to school and ensuring that they attend school.

In 2006, there were 172 primary schools (47% of all schools in Singapore from primary to preuniversity level) and a total of 285 000 primary students (54% of all students from primary to preuniversity level). Primary education consists of a 4-year foundation stage (Primary 1–4) and a 2-year orientation stage (Primary 5–6). At this stage, education aims to give students a good grasp of the English language, mother tongue, and mathematics. Students are also taught science, art and craft, music, health education,

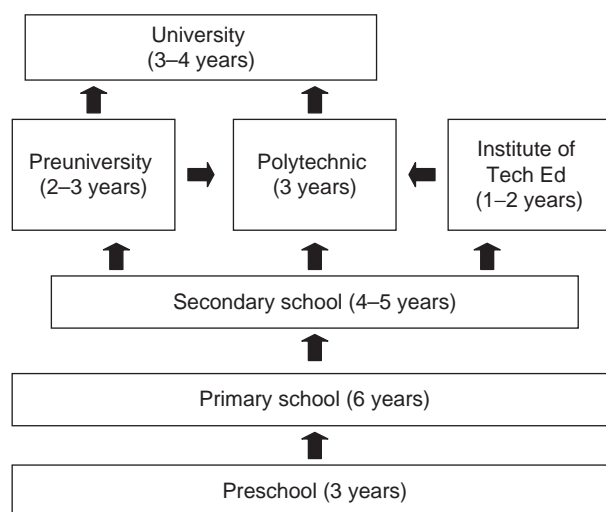


Figure 1 The main education pathways available to a student in Singapore.

physical education, social studies, and civics and moral education. They are also involved in co-curricula activities (CCAs) and Community Involvement Programme (CIP). The student to teacher ratio was 23.5 in 2005, an improvement compared to 25.1 in 2001.

At the end of Primary 6, students sit for the Primary School Leaving Examination (PSLE), except for those admitted to secondary schools under the direct school admission (DSA) scheme. Under the DSA scheme, some secondary schools in Singapore are allowed to admit a certain percentage of their students based on their own selection criteria. This percentage ranges from 5% to 50%, depending on school type and program. After the release of the PSLE results, students make six choices of secondary schools in rank order and the MOE posts them to schools based on merit (PSLE results) and choice.

Secondary Education

During the Secondary-1 posting exercise, the MOE places students in the special, express, normal (academic), or normal (technical) course according to their PSLE results. Each course has a different curricular emphasis designed to match the students' learning abilities. Special and express courses are 4-year programs leading to the General Certificate of Education (GCE) O-level examination. Special course students offer mother tongue at a higher level. Normal (academic) and normal (technical) courses are 4-year programs leading to the GCE N-level examination, in which a fifth year leading to the GCE O-level examination is available to normal (academic) students who perform well in their GCE N levels. Students are allowed to move from one course to another based on their performance and the assessment of their principal and teachers. Other than academic subjects, the students are also involved in CCA and CIP. The student to teacher ratio was 18.5 in 2005, an improvement compared to 19.6 in 2001.

In 2006, there were 168 secondary schools (include mixed-level schools offering secondary education; 47% of all schools in Singapore from primary to preuniversity level) and a total of 213 000 secondary students (41% of all students from primary to preuniversity level).

At the end of their secondary education, students may progress further in their education along a few pathways. For students who are academically inclined and have the necessary GCE O-level qualifications, they may choose to go to a preuniversity leading to the GCE A-level examination (2–3-year program). For students who wish to pursue a more practice-oriented curriculum and have the necessary GCE O-level qualifications, they may choose to go to a polytechnic (3-year program). Polytechnic graduates with good results may pursue tertiary education at the universities. For students with GCE O or N-level certificates and who wish to pursue a more vocationally oriented training, they may choose to go to a college of the ITE, which offers

1–2-year vocational courses. ITE students with good results may proceed to the polytechnics for diploma programs and subsequently progress to the universities, again upon good results.

Preuniversity

Preuniversity education prepares students for the GCE A-level examination at the end of the 2-year junior college or 3-year centralized institute course. The A-level curriculum equips students with life skills, knowledge skills, and knowledge in content-based subject disciplines, and is a preparation for university education. In 2006, there were 14 preuniversity institutions and a total of 31 000 preuniversity students (25% of all students receiving postsecondary education in preuniversity, ITE, and polytechnics).

Institute of Technical Education

Established in 1992, the ITE (formally the Vocational and Industrial Training Board), has established itself as a postsecondary technical institution of excellence. The primary role of the ITE is to ensure that its graduates have the technical knowledge and skills that are relevant to industry. It is also the national authority for standard setting and certification of skills. In 2006, there were three ITE colleges and a total of 24 000 ITE students (19% of all students receiving postsecondary education in preuniversity, ITE, and polytechnics).

Polytechnic

The mission of the polytechnics is to develop middle-level professionals to support the technological and economic development of Singapore. Polytechnic graduates are much sought after by industry because they are perceived as practice-oriented and knowledgeable professionals. The polytechnics are also a significant provider of continuing education and post-employment professional-development programs. In 2006, there were five polytechnics and a total of 68 000 polytechnic students (55% of all students receiving postsecondary education in preuniversity, ITE, and polytechnics).

University

In Singapore, the government aims to provide a diversity of university education for local and international students through local government-funded autonomous universities and private universities, including world-class universities (WCUs).

There are currently three local government-funded autonomous universities, namely the National University of Singapore (NUS), established formally in 1980; the

Nanyang Technological University (NTU), formally granted university status in 1991; and the Singapore Management University (SMU), established in 2000. Each university has its different strengths and emphases.

Private universities can be foreign or local in origin. They operate according to market forces and are at liberty to establish their own international to local student ratio and school fees. One such local private university is the Singapore Institute of Management (SIM) University (UniSIM) established in 2005 with an aim to provide university education and satisfy skills-upgrading needs of the working professional and adult learners. Other private universities include WCUs, which refer to top universities in the world that set up centers of excellence in Singapore to spearhead world-class research and development, transfer knowledge to the industry, and position Singapore as a premier educational hub. Examples of such WCUs that have set up centers in Singapore include *Institut Européen d'Administration des Affaires* (INSEAD), University of Chicago Graduate School of Business, University of Pennsylvania's Wharton School, Massachusetts Institute of Technology, and Johns Hopkins University.

In 2006, there was a total enrolment of 63 000 full-time and part-time undergraduate and postgraduate students in the three autonomous universities. Fifty-five percent of the enrolment was male. The ratio of full-time to part-time students was 3:1. The biggest portion of students enrolled in engineering sciences (37.3%), followed by business and administration (13.6%). About 25% of the enrolment was in a postgraduate program.

Other Schooling Arrangements

Children with disabilities are educated at special-education schools, which cater to distinct disability groups of children who are unable to benefit from mainstream schooling. In 2006, there were 21 such schools run by voluntary welfare organizations (VWOs) receiving funding from the MOE and the National Council of Social Service (NCSS).

There are also a few specialized independent schools such as the Singapore Sports School, the School of the Arts, Singapore, and the NUS High School of Mathematics and Science and private schools such as LaSalle College of the Arts and the Nanyang Academy of Fine Arts.

Adult Continuing Education

Increasingly, adult workers in Singapore continued with their academic education while they are on the job. In 2006, under the continuing academic education programs, some 25 000 places were taken up and 17 000 workers (67%) completed their program.

The Teaching Profession

In 2006, there were 27 000 teaching staff members in the Singapore education service (approximately 46% primary, 47% secondary, and 7% preuniversity). Generally, the education profession (primary to preuniversity) is highly feminized with a female to male ratio of about 3 to 1 in 2006. It is also a quality profession with a graduate to nongraduate ratio of about 5 to 2 in the same year.

The training of teachers and school leaders is undertaken by the National Institute of Education (NIE), an autonomous institute of the Nanyang Technological University. Teacher training at the NIE is normally 4 years for a bachelor degree cum postgraduate diploma in education, and 1 year for the postgraduate diploma in education for degree holders in other disciplines. The NIE also conducts in-service courses for teachers, leadership courses for school leaders, and courses at the master and doctoral levels. The MOE also provides round-the-year training courses for the teachers through its training and development division.

In 2006, there were approximately 1500 ITE, 3700 polytechnic, and 3900 university teachers. These institutions have their own recruitment and internal-development mechanisms for their staff.

Administrative and Supervisory Structure

The government is the principal provider of education at primary, secondary, and tertiary levels. However, there are two general ways in which the government relates with the education institutions: direct supervisory mode and autonomous mode.

The MOE exercises direct supervision over the primary schools, secondary schools, and preuniversities (except a few independent schools which are autonomous entities). That means that the principals of the schools report directly to the MOE. The local government-funded universities, NIE, polytechnics, ITE colleges, and a few independent schools operate in the autonomous mode.

The three local government-funded universities are autonomous entities, based on the recommendations by the University Autonomy, Governance and Funding (UAGF) Steering Committee in 2005. The universities exercise control over their internal governance, fees structure, admission requirements, staff remuneration, and fund utilization out of a block budget for recurrent expenditure within a broad strategic framework set by the government. The university councils are empowered to ensure that strategic plans are aligned to the university's goals, and oversee the internal quality-assurance systems. The NIE, polytechnics, ITE colleges, and a few independent schools are similarly autonomous institutions.

For nonformal education, the private sector plays the complementary role of conducting supplementary

education classes in areas such as business, computers, foreign languages, fine arts, and extra tuition. All private schools have to be registered with the MOE. But the government does not oversee their governance or operations. They are expected to operate within the general business laws of the land.

Educational Finance

Total government expenditure on education was S\$7.0 billion in 2006 (about 4% of GDP). Ninety-one percent of this expenditure went to recurrent expenses while the rest was spent on development. In terms of recurrent expenditure, the approximate cost in 2006 was:

- \$4270 for a primary school student;
- \$6186 for a secondary school student;
- \$10 001 for a preuniversity student;
- \$10 237 for an institute of education student;
- \$11 578 for a polytechnic student; and
- \$18 281 for a university student.

The breakdown of the recurrent government expenditure by the various types of education institutions as a percentage of the total recurrent expenditure in 2006 was as follows:

- primary 20%;
- secondary and preuniversity 28%;
- ITE 4%;
- polytechnic 11%;
- university 26%; and
- others 11%.

The Singapore government is proactive in its approach to education funding. For example, in 1993, the Edusave Endowment Fund was established to give every child between 6 and 16 years old, through the school, a grant to meet specified educational expenses. The money could be used for additional classes in schools, extra courses such as music and information and technology (IT) education, educational tours, and textbooks. The government contributed an initial capital sum of S\$1 billion to the fund, which reached the target of S\$5 billion in August 1997. In 1996, the Edusave scheme was widened to provide needy students with merit bursaries. The Edusave scheme helped to provide every child with resources for education because it was given to every child regardless of academic ability and every school regardless of academic standing.

Quality Assurance and Research

For the school sector (primary, secondary, and preuniversity), the main quality-assurance mechanism is the school

excellence model (SEM), implemented in 2000. Adapted from the various quality models used by business organizations, the SEM, which is a self-appraisal model, stresses the importance of self-assessment and school initiatives as a primary mechanism to drive school improvement. An external team from the MOE validates the self-appraisal results using the same criteria approximately once in 5 years. Linked to the SEM is the master plan of awards for schools, which recognizes schools for their achievement in the various SEM categories.

For the university sector, quality assurance is achieved through an accountability framework, comprising the quality-assurance framework for universities (QAFU) and the policy and performance agreements between the MOE and each university. The other autonomous institutions (such as polytechnics and ITE colleges) have their own quality-assurance frameworks agreed with the government.

Most educational research is undertaken by the MOE and the NIE. In particular, the Centre for Research in Pedagogy and Practice (CRPP), established in 2002 as part of the NIE, is the largest educational research center in the Asia Pacific. The center, which brings together researchers, educators, and administrators to research and develops innovative ways of teaching and learning, as well as informs policymaking, is funded by the MOE.

Performance of the education system is also measured by international performance indicators. Internationally, Singapore measures well in terms of achievement and resource input. For example, Singapore consistently ranks among the highest-performing countries in the Trends in International Mathematics and Science Study (TIMSS). Singapore was first in mathematics and science for grade 4 (Primary 4) and grade 8 (Secondary 2) in TIMSS 2003. In TIMSS 1999, Singapore was ranked first in mathematics and second in science at grade 8 (Secondary 2). In TIMSS 1995, Singapore ranked first at grades 7 and 8 (Secondary 1 and 2) in both mathematics and science. At grades 3 and 4 (Primary 3 and 4), she ranked second and first respectively for mathematics, and ranked seventh in science. In terms of resource input, according to the index of high availability of school resources for mathematics/science instruction by TIMSS, Singapore's 88% and 92% for grade-8 mathematics and science instruction compare very favorably with the international average of 26% and 26%, respectively.

Major Changes and Issues since the 1990s

From independence to the 1980s, Singapore has developed a robust education system through centrally driven policies. But such a centralized approach has also achieved a level of conformity to standards and reliance on external appraisal that in the 1990s, there was a need to

improve the innovation capacity of the schools to provide more quality educational pathways. The success formula that was good for the 1980s was obsolete in the twenty-first-century globalized knowledge economy.

A key milestone in this effort was the launch of the Thinking Schools, Learning Nation (TSLN) vision in 1997, after a strategic review of education, motivated by a preoccupation with the future. It guided subsequent initiatives in the education system until today. Thinking Schools is a vision of a school system that can develop creative thinking skills, lifelong learning passion, and nationalistic commitment in the young. Learning Nation is a vision of learning as a national culture, where creativity and innovation flourish at every level of society. TSLN prepares the nation to face a globalized knowledge economy by responding to the twin challenges of equipping its people for innovation and rooting its people in the country while encouraging them to go global.

Several major educational initiatives followed in the footsteps of TSLN. In particular, the education system adopted an ability-driven education paradigm (ADE) to replace the old efficiency paradigm in an attempt to tailor education to the ability and potential of the child. Creativity and entrepreneurship were emphasized. Syllabi, examinations, and university-admission criteria were changed to encourage thinking out of the box and risk taking. The MOE implemented a comprehensive strategy for developing an ICT-based teaching and learning environment in every school. National Education (NE), launched in 1997, was a significant initiative to develop national cohesion, the instinct for survival, and confidence in the future by fostering a sense of identity, pride, and self-respect as Singaporeans. So, while Singaporeans were encouraged to think global in their business, they were also encouraged to stay local in their commitment and loyalty.

Going into the 2000s and beyond, the government looks set to continue the process of emphasizing innovation and loosening the educational structure to encourage diversity. One sure sign of this phase of innovation and diversity was the shifting of the focus of TSLN to Innovation and Enterprise (I&E) in 2004, which aimed to develop an innovative and enterprising spirit in the young. The latest MOE initiative was innovation and diversity in the curricula and pedagogical aspects of education. Launched in 2005, Teach Less Learn More (TLLM) aimed to enhance the quality of teaching and learning through diverse and innovative methods tailored to the students' needs.

One of the current and major challenges for the Singapore education system is to find the breakthrough from quantity to quality. This is a great balancing act. Singapore has a robust, coherent, and systematic education system from primary to tertiary education, a system that is equally criticized and admired by many Western educators and policymakers. But she needs to address the problems of her students' lack of creativity, focus on memorization over application, and a disconnection between school learning and real-life situations. The challenge is to find the right balance so that the baby will not be thrown out with the bath water.

To encourage diversity and innovation, the government has progressively decentralized its power to the education institutions. However, the government still carries a great responsibility for achieving national outcomes and providing high value for public money. The functioning of the education institutions must be aligned to the goals of social and economic development in Singapore. Thus, there is a need to maintain strategic control so as to ensure accountability and standards. This is again a balancing act – giving power and yet maintaining control.

Whether Singapore can maintain its success and transform its education system from quantity to quality remains to be seen. But the Singapore government has proven itself to be a no-nonsense school master and a master of the balancing act. If history can be a guide to telling the future, there is reason for optimism.

Further Reading

- Department of Statistics (2007). *Yearbook of Statistics Singapore 2007*. Singapore: Department of Statistics, Ministry of Trade and Industry.
- Ministry of Education (2006). *Education in Singapore*. Singapore: Ministry of Education.
- Ministry of Education (2006). *Nurturing Every Child: Flexibility and Diversity in Singapore Schools*. Singapore: Ministry of Education.
- Ministry of Education (2007). *Many Pathways, One Mission: Fifty Years of Singapore Education*. Singapore: Ministry of Education.
- Tan, E. T. J. and Ng, P. T. (eds.) (2008). *Thinking Schools, Learning Nation: Contemporary Issues and Challenges*. Singapore: Prentice Hall.

Relevant Websites

- <http://www.moe.gov.sg> – Ministry of Education, Singapore.
- <http://www.nie.edu.sg> – National Institute of Education, Singapore.
- <http://www.singstat.gov.sg> – Statistics Singapore.

Slovenia

T Plevnik and A B Lakota, Ministry of Education and Sport, Ljubljana, Slovenia

© 2010 Elsevier Ltd. All rights reserved.

General Background

Slovenia is a small country with a population of 2 million, geographically positioned in the heart of Central Europe among neighboring states Austria, Italy, Hungary, and Croatia. Slovenia has a mixture of sub-Mediterranean climate on the coast, continental climate with mild to hot summers and cold winters on the plateaus and in the valleys to the east.

More than half of the territory of 20 273 square kilometers is forested, with hilly plains spreading across the central and eastern regions, the Alpine mountains, valleys, and numerous rivers. The future of Slovenia is conceived as the balance between the economic development and social well-being of its people while conserving the environment, protecting and preserving natural habitats of wild animals and plant species.

Currently, the population of over 2 million inhabitants (in 2007) comprises 81% Slovenians, 0.5% members of the official minorities (Italian and Hungarian), and 18.5 % of other nationalities, mainly of the former Yugoslavia. Slovenian is the state-wide official language and within designated ethnically mixed areas, Hungarian or Italian is used as the second official language. In the last census of 2002, 57.8% of the population declared themselves as Roman Catholics, 2.4% as Muslims, and 3.3% as believers of other religions.

The ancestors of the present-day Slovenians settled in the area in the sixth century BD. In the seventh century, Western Slavic tribes formed an alliance with the Slavic Duchy of Carantania (the center of today's Austrian Carinthia), which fell under the rule of the Frankish Empire in the middle of the eighth century. In the fourteenth century, most of Slovenia came under the power of the Habsburg dynasty and later became part of the Austro-Hungarian-Empire. In 1918, after the collapse of the Austro-Hungarian Empire, Slovenia joined the Kingdom of Serbs, Croats, and Slovenes, renamed the Kingdom of Yugoslavia in 1929. After World War II, Slovenia became a republic in the Socialist Federal Republic of Yugoslavia, and 50 years later, on 25 June 1991, it declared independence. Slovenia joined NATO on 29 March 2004 and the European Union (EU) on 1 May 2004.

Since independence in 1991, Slovenia's economic development has been very successful, making it one of the most prosperous countries in transition. Economic growth has been stable, reaching an average of about 4% per year, enhanced by internalization and the rapid

growth of private consumption and investments. The country is increasingly becoming a service economy; the service sector makes up 63.4% of the economy, the industrial sector 28.2%, and agriculture 2.5%. The services on the rise include telecommunications and business and tourism, with these potentially becoming the leading sectors of the Slovenian economy in the next few years. In 2006, Slovenia reached 88% of the average GDP of the EU per capita. The rate of employment has exceeded the European average (in 2007, 69%, women 62%). The average monthly gross wage in March 2008 was EUR 1352.87.

The history of education in Slovenia reflects the swift and profound changes that occurred in response to numerous political shifts and rapid economic development, affecting education to change from one extreme to another. Some disputable questions arise with each reform and are deeply rooted in Slovenian history. Such questions are: sensitivity to the national language and the fear of being overpowered by a foreign language, private schools and the processes of globalization, and harmonization of the national education system within the rules of the transnational community.

Goals of the Education System

After independence in 1991, Slovenia was faced with fundamental changes in all areas: political, economic, and social. The major education reform (1993–96) took place in the area of total ownership transformation, with the introduction of a market economy, parliamentary democracy, human rights, and increasing globalization. The transition from a centrally administrated system to an open market economy challenged previous systems of education and training. Following studies of developed Western education systems, the new Slovenian model was based on a broadly accepted concept presented in the White Paper on Education (1996) and an entirely new national legislation. For Slovenia, this was of utmost importance since in its history there was little opportunity to develop its own education system.

Independent Slovenia has thus devised and created its own national education system from preschool to university. At the same time, Slovenia has also been involved in the process of joining the EU and, as such, has been obliged to adopt the guiding principles, values, and global goals of the EU as well as those of the wider international community. Slovenia has strived to find certain practical

solutions for these global goals in a way that would be more compatible with its national traditions so that:

- Some management aspects in the autonomy of pre-schools, schools, and tertiary institutions as carried over from the rights given to self-managed communities by the former socialist regime have been maintained and new responsibilities have been added relating to the planning, performance, and evaluation of teachers' work and the professional development of teaching staff.
- An ideological neutrality of public schools has been established preventing the prevalence of any one single ideology; this also means that the constitutional principle of the state being separated from church is enforced in practice.
- An egalitarian understanding of a just school system has been promoted, postponing differentiation among children until they reach the age of 15 years, yet recognizing and respecting the differences among them and providing them with choices and assistance in learning.
- Besides the objectives set out by curricula focused on competence-oriented knowledge, there is also a strong emphasis on the acquisition of systemic theoretical knowledge; development of creativity and individual talents; arts; languages, and literature. History and geography are the backbone of general education in humanities and hold a special place in the general part of all curricula; the goal-oriented curricula are accompanied by catalogs of contents that must be taught.
- To facilitate the needs of rapid economic development, there is a highly flexible system of vocational education and training (VET) in place which is expected to adjust to the current and future needs of the labor market; to allow for a permanent dialog with social partners providing them with an opportunity for their input in the development of goals, course content, and types of qualifications and to establish multiple paths and options for initial vocational preparation.
- Until recently, a relatively isolated higher education system has officially joined the European higher education system; special attention is now given to the recommendation that education is key to the welfare of the individual in the knowledge society, as well as to the economic growth and competitiveness in the world. In this respect, entrepreneurship and privatization in higher education are encouraged.

The current government perceives education as the link between economic development and social welfare. In its strategic documents, lifelong learning is closely related to the economic advancements and employability of the individual thus also ensuring social inclusion. For the time being, the government remains almost entirely responsible for the education and social welfare of its people, although it promotes the possibility of privatizing

educational services. The wider community still recognizes education as a public benefit and in general opposes the spreading of privatization into areas which are, by definition, attributes of a social state, education being one of them. Students and traditional public universities oppose the subordination of higher education to the entrepreneurial interests of capital.

Structure and Operation of the Education System

Figure 1 outlines the formal education system. Key statistics are provided in **Table 1** about educational providers, students and teaching staff, 2007/08.

Preschool Education

After independence in 1991, preschool education became an integral part of the educational system and was transferred from the sector of family and social affairs to the jurisdiction of the ministry of education. The extensive systemic changes in 1996 introduced parental choice and better conditions for children and teaching staff. While the former program was based on group and routine activities, the new one strongly emphasizes individuality and the right of the child to use creative expression and play. Much attention is paid to the equal opportunities of children to progress, including children with special needs, children from ethnic minorities, immigrant groups, and low-income families.

In spite of demographic decline, the number of children in kindergartens has been increasing every year. In the academic year of 2007/08, there were 68% of children in the age group 1–5 years enrolled in preschool. A third of all parents of 1-year-olds opt for kindergarten, while at the age of 5 years, as many as 90% of children are enrolled. Private educational enterprises have started to appear only since 1996 and the percentage of children attending private kindergartens is still very low (1.7%). Regardless of public or private preschool provision, parents pay between 10% and 80% of costs dependent on their income. Families are supported additionally by income tax relief, child allowances, and lower payments for the second child.

Basic and Upper Secondary Education

Education is compulsory for pupils between 6 and 15 years of age. It is offered in basic schools and is organized as a single structure (without distinction between primary and lower secondary levels). Pupils receive the same education divided into three cycles lasting 3 years each without

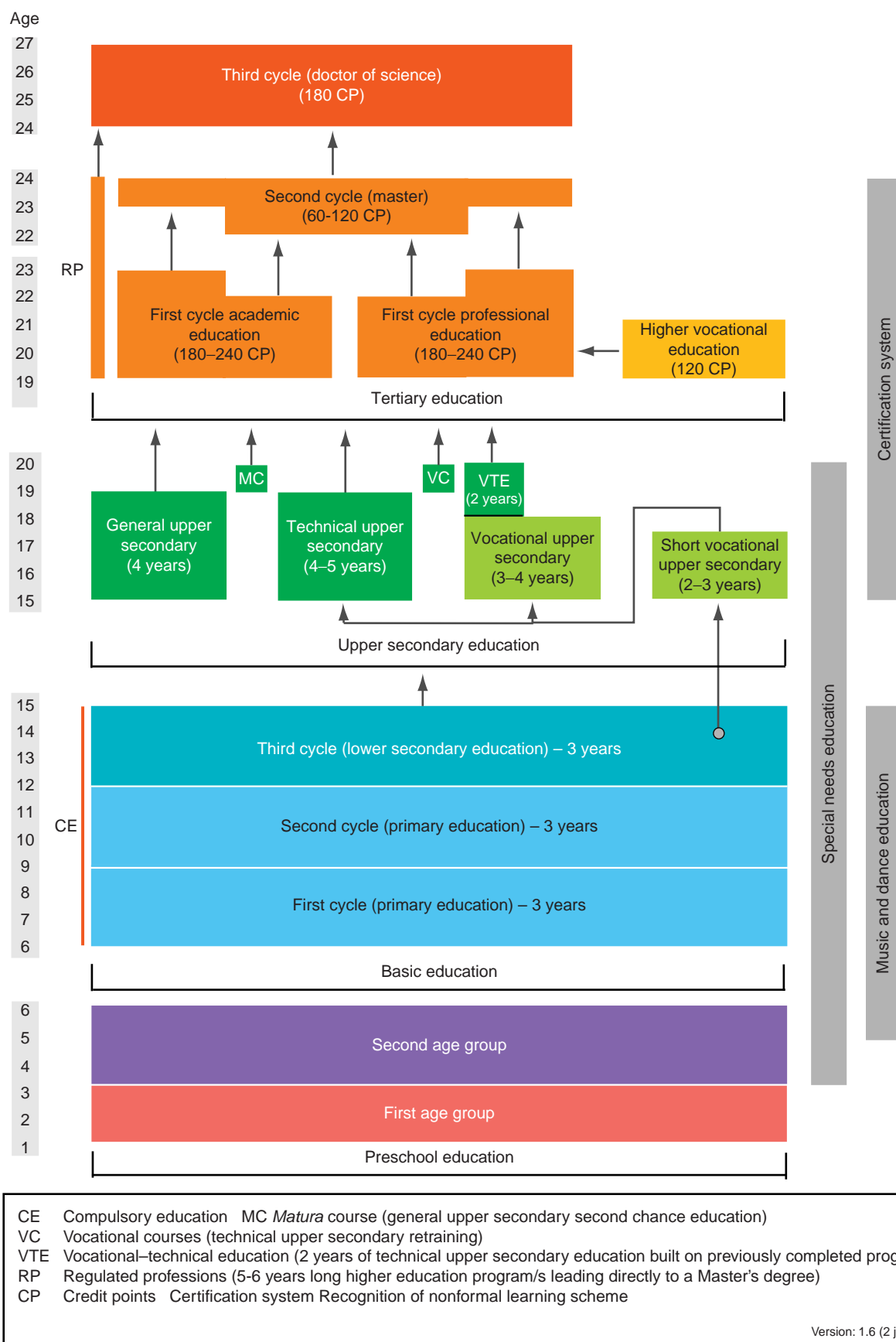


Figure 1 Structure of the education system in Slovenia.

Table 1 Educational providers, students, and teaching staff, 2006–07

Provider	Institutions		Students		Teaching staff (FTE)	
	Public	Private	Females	Males	Females	Males
Preschool institutions ^a	774	19	27 787	30 340	7096	88
Basic schools ^b	854	1	80 493	85 608	12 024	1641
Of which schools offering adapted curriculum	59	-	583	1041	394	80
Upper secondary schools ^c	136	6	46 953	49 357	4710	2510
-General programs			22 079	15 771	2004	773
-Technical programs			18 136	20 989	1919	1025
-Vocational programs			5560	12 004	780	708
-The second chance courses			1178	593	7	4
Higher education ^d	26	39	67 588	48 356	1212	2441
-Technical colleges	23	27	7925	7906	207	225
-Universities	3	1	56 425	38 372	960	2122
-Single institutions	-	11	3238	2078	45	94

Adapted from Rapid Reports and First Releases, Statistical Office of the Republic of Slovenia; 2007.

^aIn the school year 2006/07, nearly two-thirds of children aged between 1 and 6 years were included in preschool institutions (86% at the age of 5). Children with special needs (759) were included. The figure on professional staff includes assistant educators. Data on the number of preschool institutions include single institutions, branches, and kindergarten units within basic schools.

^bThe figures include basic schools and branches (796) and basic schools or units (59) which provide compulsory school curriculum adapted to the special learning needs of pupils (1624 pupils).

^cMany upper secondary schools do not differ in the type of education program they provide. Of the teaching staff in the second chance courses, only those teachers who work solely in these types of short courses are shown (1 year).

^dIn the academic year 2006/07, 48.2% of the population aged 19–23 were enrolled in higher education (115 944 students).

a change of school. Schools are mainly public and co-educational. Religious education and rituals are not allowed as the Constitution separates state affairs from religious communities. At the end of compulsory education, pupils have the choice to advance to general, technical, or vocational upper secondary education.

Today, almost 100% of 6-year-olds start compulsory education and more than 98% of pupils successfully complete it. Students who drop out are rare, yet they still have the opportunity to continue studies in short vocational courses. A fall in the birth rate since 2003 is reflected in the number of school children, which is 50% lower in 2007 than it was before 1991. A further decline of 17% is expected in the next 10 years. The average number of pupils per full-time equivalent (FTE) teacher is 12. The average number of pupils per class is 19.

Upon completion of compulsory education at the age of 15 years, students normally continue their education at the upper secondary level, split into 4-year general type programs (*gimnazije*) or technical programs and 2–4-year vocational type programs. Traditionally, many upper secondary schools provide all three types of programs. Coursework consists of 45-min lessons, 36 lessons per week (32–36 lessons), organized by subject, together with theory and practical work, compulsory elective components, excursions, seminar work, project work, and individual work. Teaching time is spread over 5 days per week and 38–42 weeks per year. Practical training in vocational or technical education may be provided in schools or in the

form of intensive practical workplace training with an employer.

Statistics show that after completing compulsory schooling, around 40% students enrol in general, 43% in technical, and 17% in vocational education. The state-wide examinations (the *Matura* and technical *Matura*) allow students to enter professionally or academically oriented higher education studies. Drop-out is less than 10% and almost 80% of all upper secondary graduates proceed to higher education, the majority of them being technicians. In 2006, 75% of the population aged between 15 and 29 were studying full-time, that is high above the European average (65%).

Much of the organization, financing, and programming of pre-university education in Slovenia is centralized. The most important decisions (e.g., the content and structure of programs, staffing requirements, and salaries) are taken at the national level. However, schools have a certain level of autonomy in the implementation of the core national curricula, choice of teaching methods, staffing and employment matters, and admission procedures.

Other Schooling Arrangements

Education for children with special needs in Slovenia has traditionally been incorporated into the general education system and integration with their peers has been a planned part of pedagogical treatment for the last 30 years.

A significant proportion of these children, assisted by special mobile pedagogical services, have been integrated in regular schools. There are still 0.9% pupils of compulsory school age who are unable to achieve the same outcome standards as required in mainstream education; they are included in adapted curricula with lower attainment targets. Only a small number of pupils are included in special schools. The education for children with special needs is provided only as a public service and is entirely the government's responsibility.

Music education, which was first introduced in the beginning of nineteenth century, holds a special place in Slovenia. Each education authority has strengthened its position in the public education network by arguing that losing it would impoverish the Slovenian culture and threaten Slovenian national identity. Almost 15% of school-age children attend basic school as well as music or ballet school. Basic schools recognize music education as an elective part of school requirements. Music school programs are public and music schools are almost entirely financed by the government and local councils.

Vocational and Further Education

Initial vocational education in Slovenia has been traditionally school based. Schools providing vocational education usually offer a whole range of educational programs: vocational, technical, and also general. The teaching staff is mostly the same for all three types of education. The responsibilities in this area are rather diffused among ministries, social partners, and councils of experts. In recent years, councils have agreed to broaden their standards, specifying the knowledge and skills that are common to a related set of occupations or functions in industry. The rather limited labor market in Slovenia requires flexible skilled workers with transferable general competences that can quickly adapt to changes in a job structure. One-quarter of the curriculum contents is left to the autonomy of schools and employers. Curricula are being constantly updated and new fields of study have emerged.

To provide a faster and more flexible response to labor-market needs which cannot be satisfied by formal vocational schooling, the so-called certification system for the assessment and award of national vocational qualifications has been developed (2003). It enables individuals to obtain a formal recognition or certification for the skills they have acquired informally. The system is administered jointly by the chambers and their vocational field commissions, trade unions, and the Ministry of Labour.

Continuing education and training also includes preparatory courses for obtaining technical qualification of master craftsman and *Matura* certificate; specialization programs; training and retraining programs; new forms

of project learning; foreign language; and leisure time courses. In 2006, 301 790 students actively participated in 357 registered training organizations. The participation rate in center for employment training (CET) and adult education of 25–64-year-olds has grown enormously: from 7.3% in 2000 to 15.0% in 2006.

Higher Education

Tertiary education in Slovenia is divided into a vocational college sector and a higher education sector with universities. The mission of vocational colleges is to meet the needs of national and local economies, with due regard to specific occupational profiles and middle management teams. Public colleges (25) offer technical subjects, whereas private colleges (34) are more likely to offer commercial and administrative courses. They have to meet the quality standards determined by the national accreditation bodies.

Higher education institutions are autonomous and self-governed institutions accredited by the National Council for Higher Education. This sector is represented by four universities and 19 other private higher education institutions. In 2008, the International Euro-Mediterranean University was founded. In the last two decades, the number of undergraduates has markedly increased and now stands at over 60% enrolments in each year group. In total, there are 89 593 students enrolled in university courses, 71% are full-time students, and more than half of them are female. Of all full-time students, 70% are enrolled in the largest university in the capital city. Foreign students represent less than 1% of all enrolments. The lecturer-per-student ratio is fairly unfavorable (22.2 in 2005). Statistical data show that study efficacy is less than satisfactory. Approximately 57% of students advance from their first year of study to the second one. Students take an average of 7 years to complete a degree even though actual study requirements, on average, do not extend beyond 5 years.

Adult and Nonformal Education

Adult education is a wide ranging and diverse sector that has expanded most rapidly in recent years. It includes all types and stages of formal education as well as nonformal learning opportunities. Private initiative has made significant progress, especially in the area of nonformal education offered by foreign language schools, information and communication technology (ICT) centers, and schools for financial management. In addition, professional associations and societies, municipal resource centers, churches, open universities, the Third Age University, and others also play a considerable role in the education and training

of adults. In 2006, 235 608 adult students participated in non-certified courses, 81% of them wishing to improve their occupational prospects.

The Teaching Profession

In Slovenia, teachers hold the status of public servants. Preschool teachers must have a higher education or university degree (at least 3 years of study). All other teachers and educators must hold a university diploma equivalent to a masters degree and a teaching qualification which can be acquired concurrently or consecutively.

In 2006, there were 26 415 school teachers in Slovenia, or rather 2.9% of the active population. Despite the demographic decline in recent years, the number of teachers in schools has continued to grow slightly. For years there have been no shortages of teachers as each year there are more graduates in education than there are vacancies. The teaching profession in Slovenia is attractive as earnings are solid and employment is relatively safe. There is a prevalence of females with the percentage of women being in inverse proportion to the level of school where they teach: 88% at the basic education level, 65% in upper secondary schools and 33.5% in higher education. Part-time employment is much rarer than in other occupations. There is compulsory training for head teachers provided by the Head Teachers School.

In 2006 there were nearly 9000 teachers and other professional staff employed at higher education institutions (5 786 FTE) and at vocational colleges (532.5 FTE).

Administrative and Supervisory Structure

Public education is managed and supervised by the government, while to some extent, especially in preschool and basic school, local councils also hold some founding responsibilities. Kindergartens and schools have the status of legal entities and, as such, hold many managerial, financial, and educational responsibilities. The government manages work relations and teachers salaries, while staffing and candidate selection is left to schools. However, the government controls the scope, work loads, and teacher qualifications by regulating the salary system and giving its approval to the classification of working positions.

Private education is regulated by the same legislation as in the public sector. Private kindergartens and schools (religious and alternative education schools) can run their own programs or carry out the public program. The government can fund up to 85% of the costs of public schools. Their operation is monitored through a system of program accreditations, funding, and inspection authorities.

Public universities in Slovenia are guaranteed autonomy by the Constitution. This is respected in terms of staff management, research, and pedagogical process, and also to

a great degree in financial matters. The ownership has been transferred from the state to universities. However, limitations set by legislation must be respected, for example, in cases of personnel dismissals, selection of students, and program accreditations. Public institutions are governed by boards with representatives from the state administration and industry. The same legislation and basic rules are applied in the private state-maintained sector, which is co-financed to a large degree by the government. Private institutions are governed by their own bodies; however, they must respect the national legislation in all respects. The independent sector in Slovenia is not significant.

Educational Finance

Funding of education is primarily the government's responsibility, with the exception of some kindergarten and basic school costs where local councils are also an important source of finance.

The percentage of expenditure for formal education in the GDP in Slovenia is relatively high; in 2006 it was 5.83% of GDP, and the total expenditure for educational institutions was 6.17% of GDP. In 2006, the highest percentage of all funds (54%) was budgeted for compulsory and pre-school education, 25% for upper secondary education, and 22% for higher education. The highest percentage of all public expenditure for formal education covered the immediate costs of educational institutions (91%), while 9% went for financial support to students and families.

Within the funding structure allocated for educational institutions, 86% were public funds, whereas 13% came from private sources.

Kindergarten attendance attracts a fee, while full-time education in schools and tertiary institutions is free. Kindergarten fees depend on the parents' socioeconomic situation (0–80% of full cost). Pupils buy or borrow their textbooks and pay for their meals and transport, but with considerable subsidies from local councils or government. The government provides university students with meals and accommodation at subsidized prices.

Performance Monitoring, Evaluation, and Research

Comparable education indicators show that Slovenia is rapidly approaching the standards of a knowledge society. The percentage of population aged between 25 and 64 holding a university degree reached 21.5% in 2006 which is just over 7% more than in 1995. The percentage of people having completed secondary education is also increasing and reached 31% in 2005. The percentage of premature school-leavers is relatively low, except in vocational education. The number of university graduates had been on the increase until 2006. Looking at the

number of years spent in school, Slovenia is close to the OECD (2004) average with 11.9 years.

The government monitors education using data, analyses, projects, and research carried out by the National Statistical Office, institutes, and agencies for development of education, education research institutes, and departments of education at universities. The government has developed a so-called strategic contracted research system, through which researchers are invited to undertake specific research work. Evaluation processes in pre-higher education are coordinated by an authorized expert body, the Council for Evaluation in Education. Evaluation includes self-evaluation, prescribed by law (2007) for all institutions, and other external forms and traditional methods, such as results analysis at the end of each assessment period, school inspections, and external assessment at the end of compulsory schooling and at the end of the 4-year upper secondary school.

Upon completion of the 9-year basic school, as well as completion of general and technical upper secondary school, an external knowledge assessment is compulsory and serves as a measure for the assessment of teaching standards. *Matura* results are used in selection procedures for enrolment in higher education. Slovenia is included in international student assessments, such as PISA, which is another assessment measure of the education system. PISA 2006 results show that Slovenian students' results are above the OECD average.

In higher education, quality assurance of the teaching process is monitored by institutions themselves and discussed by the Quality Assessment Commission (established in 1996). Universities have developed their own systems of quality control since 1999 when they first started addressing the question of quality. The council on higher education (CHE) is responsible for accrediting study programs and institutions and debating the findings from self-evaluation and external evaluation reports. The quality of teaching is annually evaluated by students using commonly accepted questionnaires. The evaluation results are made available for the public online. The results are pivotal when it comes to applications for new projects or for academic promotion.

Self-evaluation reports show high drop-out numbers and excessively long durations of study. Nevertheless, the University of Ljubljana was placed among the top 5% of universities in the world on the 2007 Times ranking table.

Major Changes and Issues since the 1990s

Education in Slovenia has always been very closely related to building and maintaining Slovenian national identity. This was precisely the reason why public education was introduced very early. Up until the 1990s, education was

developed within the guidelines defined by the former Yugoslav republics, one of which was Slovenia. In the beginning of the 1990s, Slovenia became an independent country; hence, one of the key challenges of this period was to develop a new education system in the newly independent country. The development of education system concepts was an extremely demanding task due to the role education had always played in Slovenian history. During this period, Slovenia began joining international associations and networks on a large scale, which became an important influence for the development of its own education policies. In the mid-1990s, Slovenian legislation for the first time defined the position of private and public education and made provisions for different, modern evaluation approaches.

Education policies in recent years have paid special attention to the core questions highlighted by common EU institutions. They include: how to achieve a balance between equity and efficiency, how to provide high quality education for all, how to provide access to education for all, and how to devise evidence-based policies. The Slovenian education policy is facing challenges similar to those faced by other EU member states, and the methods of solving these questions are often subjected to heated debates both in professional circles and in the general public. For decades, education policy was concerned primarily with the acquisition of formal knowledge in educational institutions; however, in recent years it has tried to systemically include and develop all elements of life-long education and has, for this purpose, devised a system of recognition for different types of knowledge and processes of skill acquisition.

Slovenia is one of the countries whose population is quickly aging. At the end of 2006, the average age of the population was 41 and in the same year, for every 100 inhabitants aged under 15, there were 114 inhabitants aged over 65. Projections up to 2050 show that the number of people aged over 65 will have doubled by then and a 33% decrease in active population is also anticipated, which places Slovenia at the top of EU rankings.

An aging population, migration, climate change, and energy restrictions are issues which necessitate a revision of educational concepts in Slovenia. Developmental documents adopted by the government in recent years have already introduced some new aspects and highlights, such as the development of social competences in young generations, development of intergenerational pedagogy, development of methodology to forecast future needs for knowledge and skills, and the development of flexible education forms and methods in adult education.

Further Reading

Eurydice Slovenia (2007). *Structure of Education, Vocational Training and Adult Education Systems in Slovenia*. <http://www.eurydice.org/>

- portal/page/portal/Eurydice/PubContents?pubid=041EN&country=null (accessed May 2009).
- Government Office for Growth (2008). *Resolution on National Development Projects for the Period 2007–2023; Effective use of Knowledge for Economic Development and Quality Jobs*. Ljubljana: Government PR and Media Office.
- Government Office for Macro-economic Analysis and Development (2007). *Poročilo o razvoju 2007 (Development Report 2007)*. Ljubljana: GOMAD.
- Government of the Republic of Slovenia (2005). *Strategija razvoja Slovenije 2007–2013 (Strategy of Development of Slovenia 2007–2013)*. Government Information Office.
- Medveš, Zdenko (2008). *Globalizacija in slovensko šolstvo (Globalisation and Slovene School system), v Sadobni pedagogiki 1/2008 (in Contemporary Pedagogy 1/2008)*. Ljubljana: ZDPDS.
- OECD. (2007). *Education at a Glance: OECD Indicators 2007*. Paris: OECD Publishing.
- Slovenia. Ministry of Education and Sport (2007). *Strategija vseživljenjskega učenja (Strategy of Lifelong Learning)*. Cabinet of the Minister of Education.
- Slovenia. Ministry of Education and Sport (2008). *Ključni dosežki na področju vzgoje in izobraževanja v Sloveniji v obdobju 2004–2008 (Key Achievements in Education in Slovenia 2004–2008)*. Ljubljana: Ministry of Education and Sport (2008).
- Slovenia. Ministry of Education and Sport; Office for Development of Education (2008). *Structure of the Education System in Slovenia 2007–2008*. ODE.

- Statistical Office of the Republic of Slovenia (2008). *Rapid Reports*.
- Student Organisation of Slovenia (2008). *Strategija zaposlovanja mladih; Analiza stanja (Strategy for Employment of Youth; State of the Art Analysis)*. Ljubljana: Student Organisation of Slovenia.
- University of Ljubljana. (2008). *Self-evaluation Report (2007)*.

Relevant Websites

- <http://www.arnes.si> Academic and Research Network of Slovenia.
- <http://www.ukom.gov.si> – Government Communication Office, Press releases.
- <http://www.uvi.si> – Government Communication Office.
- <http://www.mss.gov.si> – Ministry of Education and Sport.
- <http://www.mvzt.gov.si> – Ministry of Higher Education, Science and Technology.
- <http://www.stat.si> – Statistical Office of the Republic of Slovenia.
- <http://www.studentska-org.si> Student Organization of Slovenia.
- www.gov.si – The Republic of Slovenia.
- www.uni-lj.si – University of Ljubljana.
- <http://www.slovenijajutri.gov.si> – Web site in support of the implementation of Slovenia's Development Strategy and reforms of the work from May 2006 to December 2008.

Solomon Islands

D P Malasa, Ministry of Education and Human Resources Development, Honiara, Solomon Islands

© 2010 Elsevier Ltd. All rights reserved.

General Background

The Solomon Islands form an archipelago in the Southwest Pacific about 1900 kilometers northeast of Australia. With terrain ranging from ruggedly mountainous islands to low-lying coral atolls, the country stretches in a 1450-km chain southeast from Papua New Guinea across the Coral Sea to Vanuatu. It is made up of six large and 992 small islands with a total land area of 27 990 km² and an ocean area within an exclusive economic zone of 1.34 km². The main islands of Choiseul, New Georgia, Santa Isabel, Guadalcanal, Malaita, and Makira have rain-forested mountain ranges of mainly volcanic origin, deep narrow valleys, and coastal belts lined with coconut palms and ringed by reefs. The smaller islands are atolls and raised coral reefs and lagoons. These features impact substantially on logistics and communication and can affect the effective delivery of services such as education throughout the country.

The Solomon Islands comprise diverse cultures, languages, and customs. With a total population of about 496 000, the annual growth rate is estimated to be 2.8%. About 120 vernacular languages are spoken. Despite these many languages, the official language of instruction in schools, administration and commerce is English and the *lingua franca* is Pijin.

Only 57% of 5–14-year-olds attended school, although this percentage has increased in recent years with the change in the government policy to provide free primary education in line with the millennium development goals.

Other key characteristics of the Solomon Islands indicate that approximately 85% of Solomon Islanders live in rural areas, which are still largely subsistence, although they are both increasingly influenced by and also have a desire to join the cash economy. The remainder live in urban centers such as Honiara, Auki, Gizo, and Kirakira, or work in primary industries such as plantations and logging.

The Solomon Islands gained independence from Britain on 7 July 1978 and is a parliamentary democracy within the Commonwealth, with a unicameral parliament and a ministerial system of government. The British monarch is the head of state and is represented by a governor general, chosen by the Parliament for a 5-year term. The national parliament has 50 members, elected for 4-year terms. However, Parliament may be dissolved by majority vote of its members before the completion of its term. Parliamentary representation is based on single-member

constituencies. Suffrage is universal for citizens over age 18. The prime minister, elected by parliament, chooses the other members of the cabinet. Each ministry is headed by a cabinet member, who is assisted by a permanent secretary, who is contracted to direct the staff of the ministry.

For local government, the country is divided into 10 administrative areas, of which nine are provinces administered by elected provincial assemblies, and the tenth is the capital city of Honiara, administered by the Honiara City Council.

Most people reside in small, widely dispersed settlements along the coasts. Sixty percent live in localities with fewer than 200 persons with their distinct sociocultural settings, and only 10% live in urban areas. The capital city of Honiara, situated on Guadalcanal, the largest island, has over 30 000 inhabitants. The other principal towns are Gizo, Auki, and Kirakira.

Most Solomon Islanders are Christian, although about 5% of the population maintain traditional beliefs. The recognition of bonds of kinship, with important obligations extending beyond the immediate family group and local and clan loyalties far outweigh regional or national affiliations. This is considered one of the factors that fueled the recent ethnic tension. The social structure of most communities is generally egalitarian, emphasizing acquired rather than inherited status, and a strong attachment of the people to the land. Most Solomon Islanders maintain this traditional social structure and find their roots in village life. Land generally is still held on a family, clan, or tribal basis and may be handed down from mother or father according to local custom.

Formally organized schooling in the Solomon Islands was introduced by the Christian Missionaries in 1870s and later in 1950s by the British colonial administrators. The early church mission schools were aimed at Christianizing and civilizing the islanders. In contrast, early protectorate government schools were political in nature and were aimed at training elite which would take over the affairs of the country from Britain.

Until 1998, when world prices for tropical timber fell steeply, timber was Solomon Islands main export product, and, in recent years its forests were unsustainably over-exploited. Other important cash crops and exports include copra, cocoa, and palm oil. Since 1998, a gold mine has been developed at Gold Ridge on Guadalcanal, with its first gold export expected in early 2009. Minerals exploration in other provinces continued. Exploitation of Solomon

Islands' rich fisheries offers the best prospect for further export and domestic economic expansion. Tourism, particularly ecotourism and diving, is a potential service industry for Solomon Islands. However, growth in that industry is hampered by lack of infrastructure and transportation limitations.

With its per capita gross domestic product (GDP) of US\$340, Solomon Islands is ranked as a least-developed nation, with more than 75% of its labor force engaged in subsistence farming and fishing.

The recent conflict has devastated the national economy, and traumatized many who were affected by it. The nation and people have thus been urged to look critically at the situation and plan for the future. While the underlying social structures and values may have contributed to the conflict, they have also enabled families and communities to cope with its impact. Many people believe that the values and attitudes promoted through the Westernization of Solomon Islands society, and in particular by the education system, have been a root cause of the conflict. Although this is debatable, there is general acceptance that the current education system has increased tensions within communities, especially between younger people and their more conservative and traditional elders, by its promotion of and focus on economic advancement. The education system is seen by many as being unconnected and antagonistic to the social and cultural values on which Solomon Island communities and society are based. Such views pose great challenges to the government and educational leaders throughout the country.

The current education system in Solomon Islands is administered under the Education Act of 1978. Although in dire need of a review to accommodate the changing needs of the nation and its learners, the Act provides the legal basis on which much administration of the country's education system was decentralized to the education boards of the nine provincial governments and the Honiara City Council. This was necessary because of the geographic isolation and cultural diversity of the country and by the associated issues of communication and transportation. This has enabled most of the administrative problems to be dealt with at local level, and made the schools more responsive to the needs of their immediate communities.

The present education system is responsible for ensuring the operation and development of schools and training institutions across Solomon Islands. It manages over 600 primary schools with a student enrolment of 99 310, 140 secondary schools with a student enrolment of 25 435, and a teaching establishment of over 5609. The vision of the education system is that all Solomon Islanders will develop as individuals and possess knowledge, skills, and attitudes needed to earn a living and to live in harmony with others and their environment. It envisages a united and progressive society in which all can live in peace and harmony with fair and equitable opportunities for a

better life. It envisions an education and training system responsive to its clients and efficiently managed by its stakeholders and clients.

Goals and Strategies of the Education System

The three strategic goals proposed for the National Education Action Plan 2007–09 are:

- to provide equitable access to quality basic education for all children in the Solomon Islands;
- to provide access to community, technical, vocational, and tertiary education that will meet individual, regional, and national needs for a knowledgeable, skilled, competent, and complete people; and
- to manage resources in an efficient, effective, and transparent manner.

The overarching goal is to provide universal access to basic education for all children by 2015, and improved access to technical and vocational education and training. In the Solomon Islands, basic education refers to the first 10 years of formal education from the preparatory year, through standards 1–6 in the primary school, and including forms 1–3 in the junior secondary school. It may also include similar programs offered elsewhere at similar levels for out-of-school youth and adults.

To achieve these, the Solomon Islands Government has undertaken to place priority on refocusing education-sector expenditure on providing services at primary and junior secondary schools to achieve universal basic education by 2015.

Six immediate key strategies have therefore been developed as a focus for the period 2007–09. These include:

- to develop and implement a program of human resource development capacity building;
- to improve the implementation of grants to support school operations;
- to develop an improved school infrastructure;
- to deliver effective teacher training and development;
- to support a program of curriculum development; and
- to improve tertiary education.

Annual work plans and associated budgets have been developed by the Ministry of Education and Human Resources Development for each of 2007, 2008, and 2009 to give effect to these strategies.

Structure and Operation of the Solomon Islands Education System

Figure 1 summarizes the structure and operation of the Solomon Islands education system.

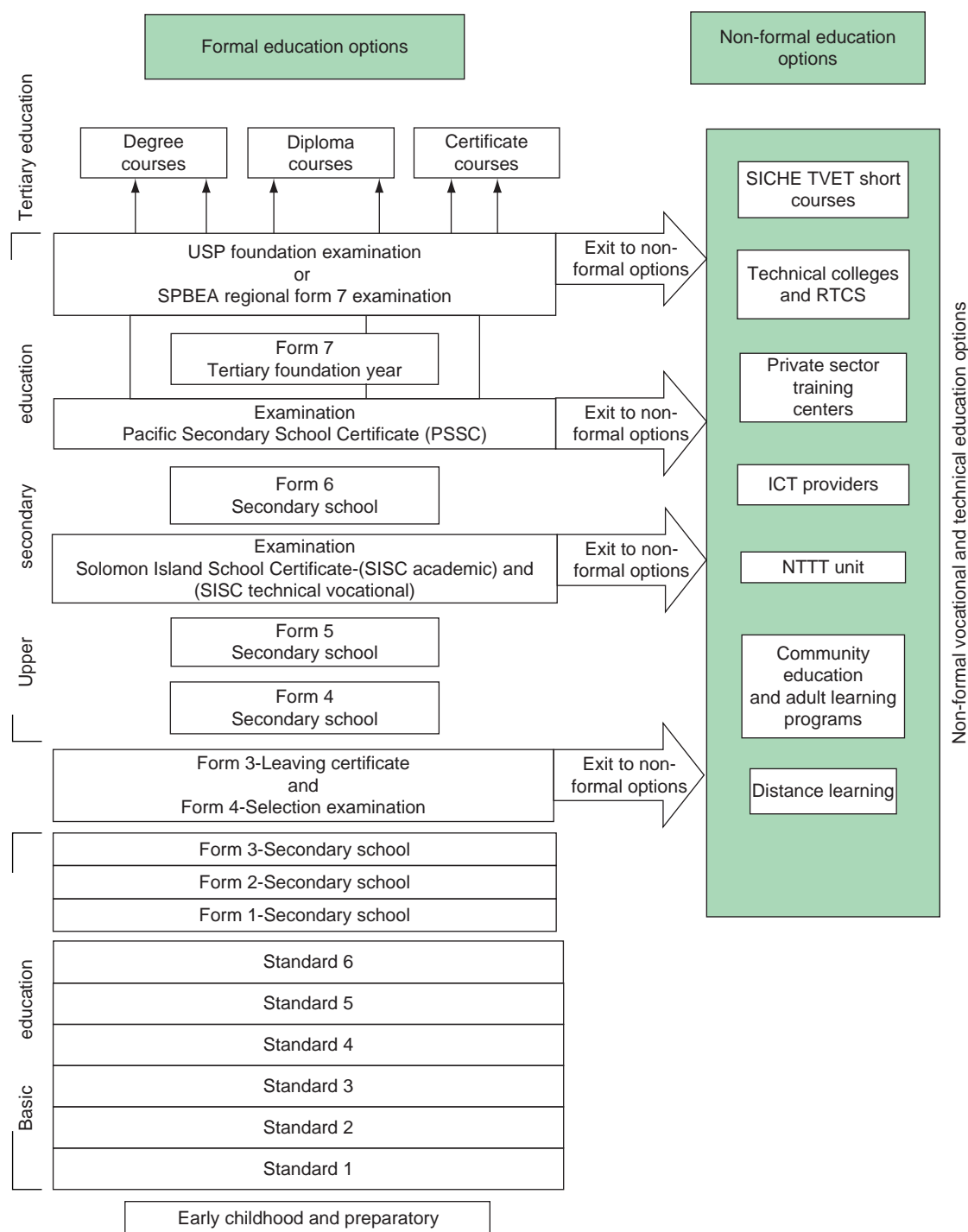


Figure 1 Structure of the Solomon Islands education system.

Early Childhood Education

Until recently, early childhood education (ECE) in the Solomon Islands has not received focused attention from the national or provincial governments. It is viewed as a parental rather than a state responsibility and has remained the domain of the churches and community groups, with little attention or support from the national

or provincial governments. Most ECE programs are therefore supported by the voluntary sector, and are not fully funded by central government. While ECE centers do not receive full financial assistance, they are subsidized to some extent. Salaries of ECE teachers trained at the Solomon Islands College of Higher Education to teacher's certificate level are paid by the government, and the Ministry of

Education and Human Resources Development (MEHRD) employs officers who monitor the provision of ECE.

No formal ECE curriculum has been approved for the Solomon Islands by the government. In general, the ECE curriculum in the Solomon Islands is not structured. Children at this age are encouraged to interact in an environment where they can explore, learn, and discover through play. Communities are encouraged to equip the centers with educational resources and equipment that will enhance children's learning and development. While there are some learning materials available to support ECE programs, supply of these materials is limited.

Thus, learning activities in these early childhood centers are focused on establishing the social values and attitudes and achievement of children in the formal education environment.

In 2006, the total ECE enrolment number was 11 228. This represents 27.5% of the estimated population of the 3–5 age group in the Solomon Islands. ECE therefore is not currently accessible to all children in the Solomon Islands. There is also inequality of access in some constituencies. The difficulty faced by some parents is that access to ECE by very young children is difficult if the ECE center is not close to the home, since young children cannot travel long distances. There is easier access to ECE in the urban centers where enrolment numbers are larger and where it is therefore more efficient to operate a center because the numbers attending make it financially viable. Access is more difficult in remote rural areas.

Primary Education

Primary education covers a formal span of 7 years. The preparatory year prior to standard 1 is regarded as part of primary education. In addition there are 6 years following the preparatory year, starting at standard 1 and ending at standard 6 with a terminal examination. This span makes a total of 7 years primary education in all. These 7 years constitute the basic-education program of the Solomon Islands. A child is expected to commence at the age of 6 or 7 and continue for 7 years.

The overall purpose of primary education is to develop children's literacy and numeracy skills (reading, writing, speaking, listening, and computational skills) and other skills and understandings that prepare young people to take part in society. These other skills and understandings include skills in science, social sciences, community studies, agriculture, art, music, and physical education.

Generally, primary education appears to be available to the majority of children in the Solomon Islands, although 6% of the primary-school-age population do not attend school.

It is worth noting that geographical and settlement patterns in the Solomon Islands greatly affect the question of access. The islands are rugged and divided by huge

stretches of sea and people live in small rural or coastal settlements. Many pupils have to walk or paddle long distances to attend school. Transport systems too are not very good, especially in rural areas. Besides, the weather is intolerable at times. These are unavoidable factors, which hinder access to primary education in Solomon Islands. However, the government and other authorities are doing all they can to minimize these factors, which have caused negative impact on universal access to primary education.

Secondary Education

There are three categories of secondary school in the Solomon Islands. National Secondary Schools (NSS) which were the original secondary schools operated by the government and the churches, and which enrol students from across the country; provincial secondary schools (PSS), which were initiated by the government but are run by the provinces, and whose student enrolment should be restricted to the province only; and community high schools (CHS), which started as primary schools, but which have had a secondary section added on. The CHS are built and managed by communities, and are assisted by church or provincial authorities.

Table 1 sets out the number of secondary schools in the Solomon Islands from 2003 to 2006.

Access to secondary schools is a major issue in the Solomon Islands. The lack of sufficient secondary school places, as a result of a lack of trained teachers and a lack of adequate facilities, is a major barrier to access. The difficulty with transportation also makes access difficult for a number of secondary school students, especially those living in isolated communities.

There is a total enrolment of 25 435 in the Solomon Islands secondary schools. This is an increase of 8.5% from 2005.

Control of the junior secondary curriculum is centralized under the Ministry of Education. Curriculum review of all secondary subjects is currently undertaken by the curriculum development center. Syllabuses, teachers' handbooks, and textbooks for students are developed by curriculum panels.

The current curriculum used by secondary schools is heavily theory based, and thus does not prepare children to go back to their villages to apply the skills they have learnt. The adoption of the policy on technical and vocational education and training (TVET) is intended to

Table 1 Number of secondary schools by type

<i>School type</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>
NSS	9	9	9	9
PSS	17	16	16	16
CHS	104	108	114	117
Total	130	133	139	142

allow junior secondary schools to be able to use TVET-modularized courses relevant to their communities in the near future.

The key questions with respect to the junior secondary curriculum are whether it remains relevant and up-to-date, and whether supporting learning materials and resources (including trained teachers) are adequately supplied to schools. The rapid expansion of CHS is throwing into sharp focus the adequacy of the training and specialist subject knowledge of the teachers who are allocated secondary classes in these junior secondary schools.

Technical and Vocational Education and Training

The development of this alternative pathway in the education system of the Solomon Islands will provide avenues for those young people who do not gain access to a place in the formal secondary school system. There is also a general need to provide adults with a range of vocational skills that are in demand in the workforce, so that they may make a contribution to the economy of the country.

TVET is available in the Solomon Islands through rural training centers (RTCs) controlled by the churches. There are no rural training centers which are administered by national and provincial governments. The TVET policy encourages establishment of more rural training centers and emphasizes the inclusion of practical subjects in the formal education system.

Access to TVET in the Solomon Islands is a significant issue. There are two main target audiences for whom access to TVET is essential: early school leavers, and adults who need return education and training to develop skills needed for participation in the workforce.

At the senior secondary level, it is planned that TVET should be developed as an alternative pathway for school leavers. In the Solomon Islands, there is a need to target those school leavers who leave the formal education system at the end of form 3 and at the end of form 5.

In 2006 the total enrolment in TVET was 2228. There are clearly issues relating to equitable access to TVET on the basis of gender. Currently, enrolments in TVET courses are dominated by males, at the expense of females. An estimated 28% of total enrolments are female, thereby constituting only slightly more than a quarter of total enrolments.

There are few opportunities for the physically disadvantaged to access TVET. In addressing access issues, both gender equity and equity for the disadvantaged will need to be considered.

Tertiary Education

Tertiary (or post-secondary) education in the Solomon Islands is provided by the Solomon Islands College of Higher Education (SICHE), and other overseas universities and

colleges, including the regional University of the South Pacific (USP) in Fiji, and its Solomon Islands campus in Honiara. The focus of SICHE is on middle-level skills and specialized training that leads to a formal qualification such as certificate and diploma, USP and other regional universities and colleges on the other hand offer higher-level tertiary education and training at undergraduate and postgraduate levels.

Most of the Solomon Islanders studying for a degree enrol with the University of the South Pacific, or attend universities and colleges in Papua New Guinea. A significant number also attend universities and colleges in Australia and New Zealand.

In 2006, over 800 students were studying in universities and colleges abroad, supported by the Solomon Islands Government as well as foreign governments (mostly Australia, New Zealand, and the Taiwan-ROC) Scholarship Awards.

Full-time student enrolment at SICHE averages 1800 each year. In addition to these opportunities, programs for part-time students are also offered by some of the schools of the college, on a day release and evening-class basis. The academic year consists of two semesters of 15 weeks each.

Other tertiary-education providers in the Solomon Islands include several church or private-funded education providers that range from theological colleges and rural training centers that focus specifically on religious and TVET. These other providers are effectively delivering tertiary (post-secondary) education, although often at a relatively low level of achievement and accreditation.

Access to tertiary-education opportunities in the Solomon Islands needs to improve significantly. Current participation levels are relatively low. Increasing participation in higher-level tertiary education courses will be necessary to lift the level of productivity and the performance of the workforce.

Adult and Nonformal Education

Many of the adult and working-age population in the Solomon Islands need to develop skills that may lead to employment, or skills that may assist in establishing viable local businesses or self-employment. The scope for developing and increasing access to adult and nonformal education to offer further education and training for the adult population is therefore extensive. Such informal and nonaward training opportunities are provided by existing tertiary (SICHE, USP) and rural training centers, schools, and community-based training centers.

Although enrolments in the nonformal sector are on *ad hoc* basis and are therefore difficult to quantify, such informal training offer opportunities for skills training for most adult and out-of-school population. However, there is still the issue of gender equity that needs to be considered.

The Teaching Profession

The total number of ECE teachers in the Solomon Islands in 2005 was 739 (652 women and 87 men). Over 80% of these teachers were unqualified, although many of these unqualified teachers may have had some field-based training. Teachers who have completed field-based training are technically classified as untrained and need to go on and complete the ECE teacher-training program at SICHE in order to qualify as a trained ECE teacher. The key need is to provide training to upgrade the skills of the ECE teachers who are unqualified. **Table 2** sets out further details.

The teacher pupil ratio for ECE in 2005 was 1:15. This ratio is as per the recommended teacher-pupil guidelines, and suggests that teacher supply is adequate. However, there is certainly an overall shortage of well-trained and qualified ECE teachers throughout the country.

There were 3964 primary school teachers in 2005 in the Solomon Islands education system, although about 33% of these teachers (approximately 1300) are categorized as unqualified. This situation has serious implications for the quality of teaching delivered to some students, and directly affects the standard of student achievement.

On the basis that the teacher:pupil ratio was 1:25.3, there appeared to be sufficient teachers in the system to manage the number of enrolments in 2005. Teacher supply does not appear to be an issue at present, but may become an issue in future as primary school rolls expand.

What does need to be monitored carefully and improved, however, is the quality of the primary school teaching staff. A key issue is the number of unqualified primary school teachers in the Solomon Islands.

Table 3 provides details of primary school teachers in the Solomon Islands in 2005 by province, gender, and sets out the percentage of teachers who are qualified.

Table 3 shows that approximately one-third of primary teachers in the country in 2005 were unqualified, and strategies need to be developed to address how the untrained teachers will be up-skilled. This issue requires

immediate attention if the quality of education in the Solomon Islands primary schools is to be improved.

The current supply of trained secondary teachers is not sufficient to meet the projected requirements of the junior secondary schools. In the junior secondary schools subject specialization is a requirement for teachers, unlike teachers at the primary level who teach all subjects of the curriculum. The curriculum reform being initiated will also require additional specialized teachers, since the introduction of 11 subjects in total (including more practical subjects) into the curriculum. Just over 20% of secondary teachers were estimated to be unqualified in 2005 (see **Table 4**). This percentage is too high.

It is evident from informed professional opinion that there is a shortage of adequately trained and qualified teachers at the junior secondary level. It will be necessary to provide training for the unqualified teachers in the junior secondary schools, and to lift the overall qualifications of teachers in the secondary teaching service.

In TVET, there are 108 trained and 82 untrained teachers. The key teacher-supply issue in this sector is finding an adequate supply of trained and competent

Table 3 Total primary school teachers by gender and percentage qualified

<i>Province</i>	<i>Female</i>	<i>Male</i>	<i>Total</i>	<i>% Qualified</i>
Central	65	135	200	79.0%
Choiseul	120	95	215	62.3%
Guadalcanal	252	403	655	64.6%
Honiara	228	113	341	75.1%
Isabel	68	113	181	92.8%
Makira and Ulawa	108	255	363	66.1%
Malaita	383	803	1186	59.8%
Rennell and Bellona	15	50	65	64.6%
Temotu	74	127	201	76.6%
Western	282	275	557	68.9%
Total	1595	2369	3964	67.3%

Table 4 Numbers of secondary teachers by province, and percentage qualified

<i>Province</i>	<i>Number of secondary teachers</i>			<i>% Qualified</i>
	<i>Male</i>	<i>Female</i>	<i>Total</i>	
Central	31	7	38	76.3%
Choiseul	15	11	26	61.5%
Guadalcanal	100	34	134	91%
Honiara	109	51	160	75.6%
Isabel	59	5	64	84.4%
Makira and Ulawa	51	18	69	76.8%
Malaita	188	58	246	77.6%
Rennell and Bellona	6	1	7	85.7%
Temotu	26	8	34	61.8%
Western	113	47	160	83.1%
Total	698	240	938	79.5%

Table 2 Total ECE teachers, by gender and percentage qualified, 2005

<i>Province</i>	<i>Female</i>	<i>Male</i>	<i>Total</i>	<i>% Qualified</i>
Central	16	6	22	13.6%
Choiseul	54	1	55	16.4%
Guadalcanal	81	9	90	6.7%
Honiara	104	10	114	51.8%
Isabel	98	9	107	6.5%
Makira and Ulawa	76	12	88	13.6%
Malaita	86	26	112	11.6%
Rennell and Bellona	18	2	20	0%
Temotu	26	7	33	30.3%
Western	93	5	98	16.3%
Total	652	87	739	18.3%

TVET tutors or instructors with high-level skills in the particular specialty in which training is desired.

Increasing the supply of trained TVET tutors is essential in order to increase the number of trainees who can enrol in TVET courses.

Table 5 sets out available 2004 data about the trained and untrained TVET teachers by province.

To summarize, the two key teacher-supply issues that need to be addressed are increasing the supply of qualified TVET instructors (i.e., the number of trained tutors), and providing training to up-skill those tutors who do not have appropriate training or qualifications.

Administration and Supervisory Structure

Schools in the Solomon Islands are operated and administered by their respective education authorities. An education authority is any person or organization the minister of education may approve to establish or maintain a school or schools within Solomon Islands. These include the provincial education boards, churches, and private individuals. There are a total of 24 approved education authorities in the Solomon Islands. A school may only be established or operated by an education authority.

Government schools are under the responsibility of the relevant education boards of their provincial education authority, while the churches and private schools are under the responsibility of their respective education boards.

Church and private-run schools form an important component of the Solomon Islands school system. Although most church-run schools have a religious affiliation, they are required to meet prescribed standards for registration, funding purposes, and to follow the prescribed national school curriculum.

Table 5 Total establishment: trained and untrained TVET teachers

<i>Province</i>	<i>Total trained</i>	<i>Total untrained</i>	<i>Total</i>
Central	-	-	-
Choiseul	4	5	9
Guadalcanal	33	33	66
Honiara	-	-	-
Isabel	3	3	6
Makira and Ulawa	14	20	34
Malaita	17	16	33
Rennell and Bellona	6	-	6
Temotu	2	4	6
Western	29	1	30
Total	108	82	190

From TVET 2004 Survey Report.

In the tertiary sector, higher education institutions in the country are autonomous bodies established under their respective legislation and independent governing body. However, primary responsibility for funding, policy, and programs remains with the Solomon Islands Ministry of Education and Human Resources Development.

Educational Finance

Although schools in the Solomon Islands are administered by their responsible education authorities, the national government provides over 70% of their operating costs. The rest are funded by parents, communities, and other private sources. In 2005, the total expenditure on education in the Solomon Islands was US\$184 million. It accounts for 7% of the country's GDP and 27% of its total recurrent expenditure.

However, 40% of this is disproportionately used to fund the tertiary sector (mostly on overseas scholarships), which comprise barely 2% of the total student population.

Government funding for early childhood, primary, secondary, and TVET is in the form of an annual grant and teacher salaries.

For tertiary and higher education, most of the government funding is in the form of grant contribution and scholarship fees, especially for students studying in overseas institutions. Most of the students studying at SICHE pay for their own fees.

Performance Monitoring, Evaluation, and Research

The performances of the national education system are monitored using data collected by the Solomon Islands Education Management Information System (SIEMIS).

Table 6 shows the 2006 detailed enrolment data of school-age children.

The above participation rates have shown that quite a number of 13–19-year-olds (30.7%) are neither at school nor engaged in employment activities. This is a threat to the social stability and security of the nation. The last ethnic tension was started by disengaged and disillusioned youths.

There are also some concerns about apparent attrition (particularly among girls) in the later years of primary schooling. There are various factors that may contribute to a significant attrition rate at the primary school level. These include geographical factors such as how children get to schools; the locations of schools far from the catchments areas, with the consequence that travel time is longer; relocation of villages to smaller remote communities; and the frequent movement of families. All these factors sometimes make it difficult for children to be

Table 6 Detail enrolment data by age, 2006

		<i>Official Age Range</i>	<i>Total Enrolment</i>	<i>Enrolment of Official Age Range</i>	<i>Population</i>
Early Childhood	M	3–5	5,720	–	–
	F		5,508	–	–
			11,228	–	–
Primary	M	6–12	52,543	5,635	49,936
	F		46,973	5,395	46,619
			99,516	11,030	96,555
Secondary	M	13–19	14,637	849	43,498
	F		10,974	742	39,888
			25,611	1,591	83,386

From Digest of Education Statistics (2006).

placed in the appropriate class. The cultural and social influence of parents and local communities are also sometimes hindrances to access and equity.

Gender equity does not appear to be a major problem, although there is a slight gender imbalance in favor of boys. Of the total enrolments in 2005, 53% were boys and 47% were girls. The reason for this imbalance may be attributable to some slight attrition of girls toward the end of primary schooling. The breakdown of enrolment figures by gender, however, is generally consistent with the distribution by gender in the primary school target-age groups, as reflected in the census estimates for the 2005 primary school-age population (52.2% male and 47.8% female).

Information was not available to determine if there are equity issues with respect to access to primary education for the disabled (e.g., children who are physically disabled, visually impaired, hearing impaired, intellectually disabled, or who have other types of disability).

The National Examination and Standards Unit (NESU) within the Ministry of Education is responsible for management, administration, and monitoring of national examinations, standards, selection procedures, and the progression of students from the exit points in the education system. It also sets national baseline data for literacy and numeracy to monitor educational standards through the administration of the Solomon Islands Standardized Test of Achievement (SISTA) and other support-assessment tools. In addition, NESU also administers three national examinations; Solomon Islands Secondary Entrance (SISE) at year 6, Solomon Islands for Three (SIF3) examinations at year 9 and the Solomon Islands School Certificate (SISC) at year 11, and Pacific Senior School Certificate (PSSC) at year 12 although the PSSC is directly under South Pacific Board for Educational Assessment (SPBEA) qualification framework.

The current national assessments heavily dependent on external examination for selection purposes are at year 6, 9, and 11, and also for selection into form 7 (year 13). The SISTA monitors students' literacy and numeracy levels at the primary. In addition, assessment for learning

tools are being developed to help teachers to address specific problem areas in the teaching and learning based on the key curriculum-learning objectives at the primary level.

The assessment for the TVET is done by individual institution based on the benchmarks set by the national trade and testing scheme. Higher education performance is also assessed by individual institutions using benchmarks and other performance outcomes set out by their respective boards of studies.

Major Changes and Issues since the 1990s

The major changes and issues facing the Solomon Islands education system since the 1990s are to maintain the delivery of education services by providing equitable access to quality education (ECE to tertiary) for all children throughout the country and to manage resources in an efficient and transparent manner.

At the same time, these changes are also linked to the global international goals (millennium development goal 2 and Education for All (EFA) goals) for education and to the broader regional context in the Pacific. In particular, the emphasis on achieving access to universal basic education for all children of Solomon Islands. The target date set by the Solomon Islands education strategic framework (2007–15) is therefore aligned with and is consistent with the target date of 2015 by the millennium development goals and the United Nations Educational Scientific and Cultural Organization (UNESCO)-sponsored Asia and Pacific Regional Framework for Action: EFA.

The national strategic goals are also linked to the regional goals for education adopted by the Pacific Education Forum Basic Education Action Plan and focus on the Pacific Islands Forum, and on the regional initiatives sponsored by the Pacific Regional Initiative for the Development of (basic) Education (PRIDE). The PRIDE initiative, which is consistent with the strategic direction adopted by the Solomon Islands education system, is also

based on the concept of regionalism – that is, countries working together for their joint and individual benefit. Regionalism under the Pacific Plan does not limit national sovereignty. It is not intended to replace national programs, only to support and complement them. This Pacific regional approach is supported because it adds value to the national efforts in the education sector here in the Solomon Islands.

Another initiative since the 1990s was the move to have greater input by the key stakeholders in planning of the education system. This has led to the development of Provincial Education Action Plans (PEAP) for the ten provinces (including Honiara city) that started in 2004. The outcomes, recommendations, and agreements that emerged from the ten PEAPs were synthesized and incorporated into the National Education Action Plan (NEAP) 2007–09, a national planning document that reflects the general needs of all provinces and needs specific to individual provinces.

It has been widely acknowledged that education of our children is a responsibility that can be shared with other education and training providers. Understanding its obligations to ensure that quality education and training is provided to its people, the MEHRD is in the process of establishing policies to guide and regulate other education and training providers to develop and deliver quality education. Other education and training providers include church and private education authorities running schools and tertiary institutions.

One of the key priorities identified by the MEHRD is to incorporate programs and activities that are aimed at achieving gender equity in access to basic education. Of equal priority is the need to identify the needs of students in the 10–19-year-old age groups who have been pushed out or dropped out of formal education, and to design programs to enable them to re-enter the system or to progress along alternate education and training paths.

In addition the MEHRD aims to conduct field studies to identify the scope of problems of people with disabilities and identify the economic constraints to achieving access to education and design programs to overcome these.

There are several issues concerning the current national external examinations, such as too many selection/exit points. While schools appreciate the current selective

examinations to select the bests, many believe that this has influenced and dominated the way teachers are teaching. In addition, critics argue that the current school curriculum and assessment system does not evaluate or report students' progress in achieving desired learning outcomes, and does not promote the adoption of teaching practices that support continued learning through the schooling period. There has been continuous evaluation of the links between the school-curriculum objectives and assessment criteria and procedures and the support for children's learning provided by current system-wide assessment methods. Another emerging issue relates to the relevance of the school curriculum to the Solomon Islands sociocultural context.

Further Reading

- Bray, M. (1991). (ed.). *Ministries of Education in small states: case studies of organization and management*. London: Commonwealth Secretariat.
- Education Act (1978). The statutes of Solomon Islands, No. 26.
- The MEHRD *Annual Report* for 2004
- The MEHRD *Annual Report* for 2005
- The MEHRD 2005 Statistical Digest
- Ministry of Education (2003). *Education Sector Plan 2000–2010*. Honiara, Solomon Islands.
- Ministry of Education (2004). *Education Strategic Plan 2004–2006*. Honiara, Solomon Islands.
- Ministry of Education (2005). *Education Corporate Plan 2006–2008*. Honiara, Solomon Islands.
- Ministry of Education (2005). *2004 Annual Report*. Honiara, Solomon Islands.
- National Education Action Plan (NEAP) 2007–2009*
- The Millennium Development Goals
- Pacific Education Forum Basic Education Action Plan
- TVET 2004 Survey Report
- UNESCO (2000). Asia and Pacific regional framework for action: Education for all. *Guiding Principles, Specific Goals and Targets for 2015. Adopted by the Asia-Pacific Conference on EFA 2000 Assessment*. Bangkok, Thailand, 17–20 January 2000.

Relevant Websites

- <http://www.pipofastaem.gov.sb>.
- <http://www.pmc.gov.sb> – Office of Prime Minister and Cabinet.
- <http://www.unesco.org> – United Nations Educational, Scientific and Cultural Organization.

South Africa

P Kgobe and J Pampallis, Centre for Education Policy Development, Johannesburg, South Africa

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The current education system in South Africa can only be understood within its historical context and the ongoing transformation efforts since the end of apartheid in 1994. The racial ideology of apartheid recognized four main population groups based on race – African, colored (of mixed descent), Indian, and white – and resulted in differential policies toward those groups, including in the field of education. The post-1994 era has seen major transformation efforts aimed at building a democratic system of education, improving access to education, improving the quality of education for all, and building a nonracial and nonsexist society. The transformation agenda has encompassed the entire education system, including early childhood development (ECD), basic education, further education and training, higher education, and adult education.

Background

In 2007, South Africa had a population of 47.9 million people, of which 51% were female. The population is very diverse, including indigenous peoples as well as substantial minorities of European and Asian origin. The four racial groupings identified by the apartheid system, though no longer a basis for differential treatment by the state, are still generally recognized. Africans constituted 79.3% of the total South African population in 2005, whites 9.3%, colored 8.8%, and Indians 2.5%. Many languages are spoken in the country and the constitution recognizes 11 of them as official languages. English and, to a lesser extent, Afrikaans tend to dominate in government and commercial life. Most of the nine indigenous official languages tend to be regionally based, although in the industrial and commercial heartland of Johannesburg and the rest of Gauteng province, there is a strong presence of all official languages.

Following the country's first nonracial elections in 1994, South Africa became a constitutional democracy. There are nine provinces, which enjoy substantial powers over matters which include, *inter alia*, agriculture, education, the environment, health, language policy, tourism, transport, and welfare. The constitution includes a bill of rights which guarantees the fundamental rights of individuals and groups.

South Africa's economic development owes much to the discovery of diamonds and gold in the latter half of

the nineteenth century. This led to significant infrastructural development and industrialization. The three main branches of the economy – agriculture, mining, and manufacturing – depended strongly on cheap black labor provided by indigenous Africans, whose traditional economy was destroyed as their land was systematically confiscated. For most of the twentieth century, the country enjoyed rapid economic growth. However, growth slowed in the last quarter of the century, partly as a result of labor constraints resulting from apartheid's restrictions on the internal markets for both labor and goods, and partly because of international economic sanctions imposed in protest against apartheid.

When the first democratically elected government took power in 1994, the South African economy was not showing any significant levels of growth. This changed by the late 1990s. From 1996 to 2002, the average real annual growth in the gross domestic product (GDP) was 2.8%. From 2003 to 2007, it rose to an average of over 5% per year.

There are high levels of unemployment in the country, with as much as 26% of the population considered unemployed in March 2007, using the official definition of unemployment which excludes those who are not actively seeking work. Using a broader definition, the unemployment rate is closer to 40%.

Education

Post-Apartheid Reform

Apartheid education was shaped by racial disparities in relation to access, resource allocation, and quality of provision. Education provision for each racial group was undertaken by a separate department, and led to different standards of provision. At the time that the democratic government took office, white education received more than 4 times the amount of money per learner than was allocated for black education. The colored and Indian populations had a level of provision between that of whites and Africans. This situation impacted on the quality of education provided for the various groups. The poor quality of black education is evident in rates of progression and completion of education: only 51% of African children reached grade 8 in 12 years of schooling. In contrast, 96% of white learners reached this grade in 8 years. Enrolments in higher education also reflected

the legacy of apartheid, with fewer blacks enrolling in higher education institutions, especially in fields such as engineering, mathematics, and science.

Through the bill of rights, all South Africans have the right to a basic education, including adult basic education, and the right to receive an education in any of the 11 official languages where it is reasonably practicable. The constitution splits control over schooling between the national and provincial levels of government. The national department of education is responsible for establishing national norms and standards, and nine provincial departments are responsible for administering the school system and establishing provincial policy within nationally set parameters. Higher education is the responsibility of the national government, but the higher education institutions enjoy a high level of autonomy.

When it was formed in 1994, the Government of National Unity tackled the reorganization of the entire education system. Within the schooling system, the first area of education-policy development that it addressed was the organization, governance, and funding of schools. After a lengthy process of consultation and negotiation, the main policies were set out in the South African Schools Act of 1996. The act established two legally recognized categories of schools – public and independent schools – and decentralized many powers to elected school-governing bodies. A new curriculum (known as curriculum 2005) was introduced in 1997. Due to its perceived complexity, it was later reviewed and a revised national curriculum statement (RNCS) was adopted in 2002. The further education and training (FET), higher education, ECD, and adult basic education and training (ABET) sectors also underwent major reforms.

An important component of the government's post-apartheid education-reform strategy was the establishment of the national qualifications framework (NQF). The South African Qualifications Authority (SAQA) is a statutory umbrella body which is responsible for coordinating this framework. The NQF aims to accommodate all recognized qualifications in South Africa – whether they have been obtained through formal education or through workplace training. Its task includes providing equivalence between different types of qualifications, and assuring their quality. The NQF has become understood more broadly as a means to achieve an integrated approach to education and training which recognizes the value of both.

South Africa spends almost 5.5% of GDP on education, one of the highest expenditure rates in the world.

Enrolments across the System

In 2005, there were 13 936 737 students in all sectors of the education system. Of these, 85% were in public schools and 2.3% in independent schools. Public higher education

institutions accounted for 5% of total enrolments, public FET colleges 2.7%, ABET centers 2%, ECD sites 2%, and special schools 0.6%. There were 34 162 educational institutions in South Africa, of which 26 592 were ordinary schools. There were 437 330 educators and lecturers in the education system. Policy is to include children with disabilities as much as possible in mainstream schools.

Early Childhood Development

Under apartheid, there was a rigid distinction between education and care. Child care was provided for in terms of the Child Care Act of 1983, which made provision for the registration of children's homes and places of care. According to the 1996 Interim Policy for Early Childhood, only 9–11% of all South African children from birth to 6 years had access to ECD facilities at the end of the apartheid period. One in three white infants and children received ECD services, compared with one in eight Indian and colored children and one in 16 African children. Twice as many urban children were receiving ECD services compared to their rural counterparts.

Current policy promotes an integrated vision of ECD and views it as the provision of physical, emotional, social, spiritual, and moral development for children between birth and 9 years. Provision is therefore a responsibility of various sectors, including a number of departments within the government (education, social development, and health) and the broader community. In the schooling system, a reception year (grade R) is currently being introduced. It will be part of an integrated 4-year foundation phase program (grades R, 1, 2, and 3) to be located in schools.

The introduction of grade R in schools has improved access to ECD for 5-year-old children. **Table 1** presents data on grade-R enrolment from 1999 to 2005.

In addition to grade-R classes in schools, there are many private providers which offer ECD for children from as young as 2 years, but their enrolments are currently unconfirmed. Although the number of both public and private ECD sites has been increasing, in May 2007

Table 1 Number of children in grade R

	Male	Female	Total
1999	77 718	78 574	156 292
2001	120 449	121 076	241 525
2003	157 532	157 855	315 387
2005	202 590	202 607	405 197

From Department of Education (2005). *Teachers for the Future: Meeting Teacher Shortages to Achieve Education for all*. Pretoria: Department of Education; Department of Education (2006). *Education Statistics in South Africa at a Glance in 2005*. Pretoria: Department of Education; and Department of Education (2006). *The National Policy Framework for Teacher Education and Development in South Africa*. Pretoria: Department of Education.

the Minister of Education identified the sub-sector as “one of our poor performance areas.”

Primary and Secondary Schooling

Compulsory schooling (also known as the general education band) is divided into three phases for curriculum purposes: foundation phase (grades 1–3), intermediate phase (grades 4–6), and senior phase (grades 7–9). The further-education band consists of grades 10–12. Grade 12 culminates in the senior certificate, commonly referred to as matric or matriculation. Most primary schools go from grade 1 to grade 7, while secondary schools normally encompass grade 8 to grade 12. The reason for this division, which breaks the intermediate phase between primary and secondary schools, is historical in that most schools were built before the new curriculum was introduced. There are no plans yet to change this state of affairs.

The South African Schools Act of 1996 made 9 years of education compulsory for all children between the ages of 7 and 15. Since then, enrolment in the schooling system has grown significantly, especially for the African population. Although schooling is still not universal, well over 90% of children of appropriate age are enrolled in the compulsory phase of the system.

Table 2 shows the gross enrolment ratio (GER) in public ordinary and independent schools by province in 1997, 2001, and 2003. The GER measures the number of students as a proportion of the total population of appropriate age. Whereas the table shows a GER of 111 nationally in 1997, this figure had declined to 94% in 2003. The decline is probably a result of new regulations, which have limited the number of times a child may repeat a grade to once in a 3-year phase; they have also prevented schools from enrolling children younger than 6 years of age. This

has resulted in a decrease in the enrolment of overage and underage learners in the system.

From time of the South African Schools Act, schools have been allowed to charge school fees if the majority of parents agree. It soon became clear that even schools in poor communities were charging fees under pressure from school principals and governing bodies who argued that they did not get enough state funding. This caused widespread distress in poor communities, and in 2006 government began to designate schools in poorer communities as no-fee schools. By 2009, 60% of schools will be so designated.

Adult Education and Training

In South Africa, basic education (i.e., up to the equivalent of grade 9) is a constitutional right. This includes adult basic education. In 2007, just over 10% of the population aged 20 and above had had no schooling; a decrease from 18% recorded in the 2001 census and 19% in 1996 census. Adult illiteracy in South Africa has a race and gender bias, as shown in **Table 3**.

There are a number of providers of ABET in South Africa, including various government departments, non-governmental organizations (NGOs), and employers.

In addition, Cabinet approved approximately ZAR6 billion for a mass literacy campaign to be undertaken from 2008. The project seeks to eradicate illiteracy by 2012, by reaching approximately 4.7 million people who are considered functionally illiterate. This is in addition to the various preexisting literacy programs.

Further Education and Training Colleges

FET colleges provide vocational education and training in six broad fields. These are:

Table 2 Gross enrolment ratio (GER) (percentage) in ordinary public and independent primary and secondary schools by province, 1997, 2001, and 2003

Province	1997			2001			2003		
	Prim	Sec	Total	Prim	Sec	Total	Prim	Sec	Total
Eastern Cape	144	85	120	122	68	100	118	72	99
Free State	125	98	113	107	82	97	99	81	92
Gauteng	115	88	104	102	77	92	100	81	92
Kwazulu-Natal	137	92	119	107	78	95	104	80	94
Limpopo	111	105	108	103	85	99	99	90	95
Mpumalanga	124	89	110	113	84	101	101	87	95
North West	99	63	85	103	80	93	98	82	92
Northern Cape	114	90	104	109	70	93	111	79	98
Western Cape	111	77	97	109	71	89	97	71	86
National	125	90	111	108	77	96	104	80	94

From Shindler, J. (2008). Public schooling. In Kraak, A. and Press, K. (eds.) *Human Resource Development Review, 2008: Education, Employment and Skills in South Africa*. Cape Town: HSRC Press.

Table 3 Percentage of people in the population aged 20 and above who reported they had no schooling by race and gender, 1996, 2001, and 2007

	1996 (%)	2001 (%)	2007 (%)
<i>By Race</i>			
African	24.2	22.3	12.8
Colored	10.1	8.3	5.3
Asian	6.5	5.3	3.6
White	1.2	1.4	0.6
<i>By Gender</i>			
Male	17.1	15.5	8.4
Female	20.9	20.0	12.1

From Statistics South Africa (2007b). *The Community Survey*. (2007). Pretoria: Stats SA.

- art and music,
- business studies,
- educare and social services,
- engineering studies,
- general education, and
- utility studies.

FET colleges offer programs at levels equivalent to the three highest grades of senior secondary school and beyond. While it is possible for students to proceed from FET colleges to higher education, this currently seldom happens. Efforts are underway to increase such mobility.

FET is an important sector, especially in terms of skills development. In South Africa, it has been linked to major skills-development initiatives such as the national skills strategy and the joint initiative for priority skills acquisition (JIPSA). A major program is under way to recapitalize the FET colleges. The aim is to ensure that these colleges have the necessary infrastructure, equipment, materials, administrative systems, teaching staff, and information technology for effective teaching.

In addition to the public FET colleges, education at this level is also made available by private providers, NGOs, community organizations, and employers. The public colleges were established by merging technical colleges, manpower and skills centers, and some former community colleges. There are 50 public FET colleges in South Africa, with more than 200 college campus sites and satellite sites. These enrolled 377 584 learners in 2005 and employed 6407 educators.

The private FET sector has grown rapidly in the post-1994 period. While government has been making efforts to register and regulate private FET colleges, there appears to be no certainty yet on the size of the private FET sector. However, some officials estimate that it is substantially larger than the public sector in the number of enrolments.

Table 4 Percentage of students by degree level, 2005

<i>Degree</i>	<i>% of students</i>
Undergraduate degrees (3 years)	59.8%
Professional undergraduate degrees (4 years)	22%
Postgraduate degrees below masters	8.4%
Masters and doctoral degrees	7.3%
Occasional courses (various levels)	2.6%

Compiled from information on the Department of Education website and the higher education management information system (HEMIS) for State-subsidized Universities and Technikons.

Higher Education

There are currently 22 public higher education institutions in South Africa – 11 universities, five universities of technology (formerly known as technikons), and six comprehensive universities (a combination of universities and technikons). This new institutional landscape emerged out of a restructuring process, which included a consolidation of the 36 higher education institutions that existed previously.

Enrolment in public higher education institutions stood at 741 380 in 2006. This represented an increase from the 270 874 recorded in 1985. The majority of students in public higher education institutions were enrolled for undergraduate degrees, as shown in **Table 4**.

Of the total number of students, 29% were registered in science, engineering and technology programs, 30% in business and commerce, 13% in education, and 28% in other humanities and social sciences. In terms of racial group, 61% of registered students were African, 25% were white, 7% were colored, and 7% were Indian. In 2005, 120 000 students completed their studies in higher education institutions, an increase from 77 000 earlier in 1998. The increase in higher education enrolments has been facilitated by the establishment of a national student financial aid scheme which provides assistance to poorer students, mainly in the form of loans which are repaid after the recipient has left university and found a job.

The racial profile of academic staff in higher education institutions has not changed as much as that of students. The proportion of white academic staff in permanent posts as a whole declined from 87% to 61% between 1993 and 2006, while black staff increased from 13% to 39% (with 25% being African).

Private higher education institutions are permitted in terms of the Higher Education Act of 1997, but are required to register with the department of education. The purpose of registration is to ensure that institutions offer an acceptable quality of education, that they have the capacity to deliver what they promise, and that students will receive qualifications aligned to the NQF. The

register of private higher education institutions in 2008 records 78 registered and eight provisionally registered. None of these currently qualify to call themselves universities. They range from fairly large corporate and overseas-based institutions, offering a limited range bachelor's and even master's degrees, to fairly small specialist colleges offering various courses including, *inter alia*, business studies, advertising, theology, media studies, information technology, and tourism.

Reliable data on enrolment, staffing, and output in the private higher education sector is not currently available, but they clearly cater for tens of thousands of students.

The Teaching Profession

There were 385 860 teachers in South Africa in 2006. Of these, 5% (19 407) were in independent schools. Of those in public institutions, 173 850 were in primary schools, 111 865 in secondary schools, and 53 988 in combined, intermediate, or middle schools. Others were in special schools (i.e., schools for the disabled), ECD, and ABET centers. Approximately 7% of teachers in the public school system are employed privately by school governing bodies, mainly in the more affluent areas where school fees provide the resources for this.

In 1994, 36% of teachers were considered underqualified or unqualified (the definition of qualified is 3 years of professional, postschool training). This figure declined to 18% in 2001, and stood at 8.3% in 2004. There are moves to raise the standard for qualified teachers by one additional year of professional training. **Table 5** shows teacher qualification by gender and race in 2004.

Notwithstanding the improved qualification profile of the teaching force, indications are that the majority of teachers have not yet been sufficiently equipped to meet the education needs of a growing democracy in a twenty-first-century global environment. A research report commissioned by the president's education initiative in 1999 concluded that the most critical challenge

for teacher education in South Africa was the limited conceptual knowledge of many teachers. This includes a poor grasp of their subjects as evidenced by a range of factual errors made in content and concepts during lessons. Similar concerns are expressed in the findings of other research.

The teaching profession faces a shortage of qualified educators, especially in the fields of mathematics and science. Estimates vary, but indications are that teacher training institutions are not producing sufficient teachers to meet the demands of the schooling system. This situation is further exacerbated by the impact of the HIV and AIDS pandemic. A recent study conducted for the education labor relations council estimated that 22% of educators in South Africa were living with HIV. The highest prevalent rates were among teachers aged 24–35. Although the attrition rate for teachers was between 5% and 5.5% (i.e., equivalent to 17 000–20 000 teachers in 2004), teacher-training institutions were producing at best 9000 graduates annually, of whom 3000 were already practicing teachers.

Major plans are underway to address these challenges, and to increase the number of people who choose teaching as a career. This includes bursary schemes, efforts to improve the quality of teacher education, and mitigate the impact of HIV and AIDS on teachers.

In order to maintain high levels of professionalism in the teaching profession, a South African council for educators was established by Act of Parliament in 2000. All educators must be registered with the council, which is responsible for ensuring the maintenance of professional ethics and values in the teaching profession.

Performance Monitoring and Evaluation

A number of performance monitoring and evaluation mechanisms have been put in place. The South African Qualifications Authority (SAQA) Act of 1995 provides for

Table 5 Teacher qualifications by gender and race, 2005

REQV ^a	Total (%)	Female (%)	Male (%)	African (%)	Colored (%)	Indian (%)	White (%)
14+	53	53	52	47	55	89	88
13	38	38	39	42	35	11	12
12	6	6	6	7	7	0	0
11	2	2	1	2	3	0	0
10	1	1	2	2	0	0	0
	100	100	100	100	100	100	100

^aRequired education qualification value (REQV) refers to the qualification level of teachers. Levels 10–16 refer to the number of postmatriculation years teachers have successfully completed to achieve their qualification. For example, a teacher at REQV 13 has completed the (school leaving) senior certificate plus 3 years of additional education.

From Department of Education (2006). *The National Policy Framework for Teacher Education and Development in South Africa*. Pretoria: Department of Education.

the establishment of education and training quality assurance bodies (ETQAs) across the various sectors of the education system. The main purpose of the ETQAs is to monitor and audit achievements in terms of national standards.

Two independent quality-assurance bodies have been established by statute. The first is Umalusi – the council for quality assurance in general and further education, which aims to enhance and assure education quality in public and private schools, further education institutions, and adult education providers. The second is the higher education quality committee (HEQC), a permanent committee of the council on higher education, which has a similar function in the higher education sector. In late 2007, the government took the decision to establish another similar body, the quality council for trades and occupations (QCTO), with responsibility for occupation-specific training programs in workplaces and institutions outside of registered universities and colleges.

Twenty-three sector education and training authorities (SETAs) have been established with responsibility for promotion and coordination of skills development within their respective economic sectors. This includes a quality-assurance role. They are accredited by SAQA to quality assure qualifications in their areas of primary focus. So, for example, the quality assurance of programs aligned to bricklaying or carpentry qualifications is delegated to the Construction SETA, which must ensure that all accredited training meets approved standards.

Within the system of school administration, the department of education has created the integrated quality management system (IQMS). This system serves four purposes – to identify the specific needs of educators, schools, and district offices with a view to supporting and developing them; to promote accountability; to monitor the overall effectiveness of institutions; and to evaluate educator performance. The IQMS uses three strategies, namely developmental appraisal, performance measurement, and whole school evaluation, all of which are aimed at enhancing and monitoring the performance of the education system as a whole.

There are also systemic evaluation studies which focus primarily on assessing the achievement of learners at the various transitional stages of the system – namely, grades 3, 6, and 9. The purpose of the systemic evaluations is to assess the effectiveness of the entire system and the extent to which the vision and goals of the education system are being achieved. Two systemic evaluations have been conducted to date (among grade-3 learners in 2003 and among grade-6 learners in 2004/2005). Both have given rise to major concerns regarding the teaching of literacy and numeracy in primary schools, as pupil performance has been unacceptably low. As a result, the department of education has developed special programs to tackle the problems identified.

Conclusion

Many structural changes have reshaped the way education is organized and managed in South Africa and have increased access to education at all levels. While the system will continue to expand, the focus for the next decade will be on quality improvement. The government has already put in place a number of initiatives in this regard, including various programs for quality assurance, increased emphasis on teaching literacy (both in primary schools and in adult classes), improving mathematics teaching, enhancing the relevance and quality of further-education colleges, improving the quantity and quality of teaching materials and educational equipment, upgrading the infrastructure of educational institutions, and improving both preservice and in-service teacher education. South African education remains very much a system in transition, and systemic reform will continue to dominate the educational agenda for many years to come.

Further Reading

- Chisholm, L. (ed.) (2004). *Changing Class: Education and Social Change in Post-Apartheid South Africa*. Cape Town: HSRC Press.
- Department of Education (2004). *Education Statistics in South Africa at a Glance in 1999*. Pretoria: Department of Education.
- Department of Education (2004). *Education Statistics in South Africa at a Glance in 2001*. Pretoria: Department of Education.
- Department of Education (2005). *Education Statistics in South Africa at a Glance in 2003*. Pretoria: Department of Education.
- Department of Education (2005). *Teachers for the Future: Meeting Teacher Shortages to Achieve Education for all*. Pretoria: Department of Education.
- Department of Education (2006). *Education Statistics in South Africa at a Glance in 2005*. Pretoria: Department of Education.
- Department of Education (2006). *The National Policy Framework for Teacher Education and Development in South Africa*. Pretoria: Department of Education.
- Department of Education (2007). *Annual Report. 2006/2007*. Pretoria: Department of Education.
- Fiske, E. B. and Ladd, H. F. (2004). *Elusive Equity: Education Reform in Post-Apartheid South Africa*. Washington, DC: Brookings Institution Press.
- Fleisch, B. (2002). *Managing Educational Change: The State and School Reform in South Africa*. Sandown, South Africa: Heinemann.
- Fleisch, B. (2007). *Primary Education in Crisis: Why South African Schoolchildren Underachieve in Reading and Mathematics*. Cape Town: Juta.
- Government Communication and Information System (GCIS) (2007). *South Africa Yearbook 2006/07*. Pretoria: GCIS.
- Kraak, A. and Press, K. (eds.) (2008). *Human Resource Development Review 2008: Education, Employment and Skills in South Africa*. Cape Town: HSRC Press.
- Morrow, W. (2007). *Learning to Teach in South Africa*. Cape Town: HSRC Press.
- Motala, E. and Pampallis, J. (eds.) (2001). *Education and Equity: The Impact of State Policies on South African Education*. Johannesburg: Heinemann.
- Nelson Mandela Foundation (2005). *Emerging Voices: A Report on Education in Rural Communities*. Cape Town: HSRC Press.
- Sayed, Y. and Jansen, J. (eds.) (2001). *Implementing Education Policies: The South African Experience*. Cape Town: University of Cape Town Press.

- Shindler, J. (2008). Public schooling. In Kraak, A. and Press, K. (eds.) *Human Resource Development Review, 2008: Education, Employment and Skills in South Africa*. Cape Town: HSRC Press.
- South African Institute of Race Relations (SAIRR) (2007). *South Africa survey, 2005/2005*. Johannesburg: SAIRR.
- Statistics South Africa (2007a). *Mid-Year Population Estimates 2007*. Pretoria: StatsSA.
- Statistics South Africa (2007b). *The Community Survey. (2007)*. Pretoria: Stats SA.
- Statistics South Africa (2008). *Unemployment and Employment in South Africa*. Pretoria: Statistics South Africa.
- Taylor, N. and Vinjevoold, P. (1999). *Getting Learning Right*. Johannesburg: Joint Education Trust.

Relevant Websites

- <http://www.che.ac.za> – Council on Higher Education.
- <http://www.education.gov.za> – Department of Education.
- <http://www.gcis.gov.za> – Government Communication and Information System.
- <http://www.saqsa.org.za> – South African Qualifications Authority.
- <http://www.statssa.gov.za> – Statistics South Africa.
- <http://www.thutong.org.za> – Thutong South African Education Portal.
- <http://www.umalusi.org.za> – Umalusi, Council for Quality Assurance in General and Further Education and Training.

Spain

A Tiana, Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Autonomous communities – Each of the 17 regions forming Spain. They have their own parliament and government and hold wide responsibilities in several fields, education among them.

Bachillerato – Academic upper-secondary level, noncompulsory, composed of two grades (grade 11 and 12).

Basic education – Educational period aimed at preparing all citizens for adult life, composed of ten grades in two levels (primary and *Educación secundaria obligatoria* (ESO)), compulsory and free of charge.

Childhood education – Educational level for children from birth to 6 years, equivalent to pre-primary or preschool education in other systems.

Conference of Ministers of Education – Body formed by the Minister of Education of Spain and the regional ministers and having the function of discussing all national regulations and establishing ways for cooperation in the field of education.

Early school leaving – Proportion of young people from 18 to 24 years having finished compulsory schooling at most and not continuing further studies.

Educación secundaria obligatoria (ESO) – Comprehensive lower-secondary level composed of four grades (7–10).

High Inspectorate – Body which assures that educational regulation of Spanish autonomous regions does agree with national legislation and respects central-government responsibilities.

National Agency for Quality Evaluation and Accreditation (ANECA) – National institution for the evaluation of quality and accreditation of university studies and staff.

National Professional Qualifications Catalog – National repertoire including all qualifications identified in the labor market with the necessary competencies for a professional exercise.

Popular universities – Educational institutions for adults, usually promoted by municipalities, offering formal and nonformal programs and cultural and leisure activities.

School council – Decision making and controlling body at school level, assuring the participation of

teachers, parents, and students in school governance.

Universities of the experience – University extension programs for adults over 55 years without academic requisites.

General Background

Spain is situated southwest of Europe, having an area of 505 km², most of it in the Iberian Peninsula. Its population grew from 18 594 405 inhabitants in 1900 to 40 709 455 in 2001, reaching 45 million in 2008, due in part, to the important immigration received since the late 1990s. Of the current Spanish inhabitants 8.4% are of foreign origin, mostly from Latin America, Europe, and North Africa.

Spanish economy has been traditionally based on extensive agriculture, which employed more than half of the active population until 1940. From 1960 on there were important changes in all economic sectors, producing as a result, a very significant growth of employment in industry and services and a parallel decline in agriculture. In 2001, employment in agriculture represented 4.7% of the active population, compared to 30.7% in industry and 64.6% in services. So, Spain has evolved into a typical postindustrial economy. According to Organization for Economic Cooperation and Development (OECD) data, gross domestic product (GDP) per capita in equivalent USD was 26 018 in 2004.

In the last two decades, an increasing number of women have joined the workforce, producing as a side effect a growing demand of educational services for early childhood. As the trend is likely to continue in coming years, it will be of particular influence for the development of the education system.

Spain is a diverse country, by geography, ethnics, culture, and languages. Spanish, a language spread worldwide, is the official language of Spain, but there are also some regional languages, which are co-official in their own territories, like Catalanian, Basque, or Galician. The existence of different official languages creates an obligation to offer education in all of them according to different regional regulations.

Spain is a constitutional monarchy according to the 1978 Constitution, which was the result of a broad national agreement reached after the end of the authoritarian Franco's regime. The long history of wars and conflicts in the nineteenth and twentieth centuries was closed with

the transition to democracy – a process started in 1975 and ended with the approval of the new Constitution in 1978.

The Constitution acknowledges the diversity of Spain, setting the basis for what is usually known as a ‘State of Autonomies.’ According to this system, Spain is organized into 17 autonomous communities, with their own regional parliaments and governments. Even if Spain is not a proper federal state, the autonomous regions have wide responsibilities in some fields, education among them.

The building of the national education system started with the 1812 Constitution, which obliged all towns of the Monarchy to have primary schools. The liberal governments of the first half of the nineteenth century approved the seminal rules for the national system and in 1857 the first general law for education was passed, introducing compulsory education for all Spanish children from 6 to 9 years.

The development of the education system was very slow and timid. For a long time, schools were under the responsibility of the municipalities. The national Ministry of Public Instruction was created in 1900 and extended its responsibilities into the field of public education. Nevertheless, the conflicts of the first part of the twentieth century and the negative effects they had in education were evident until the last quarter of the century. As a significant example, universal education from 6 to 14 years was only accomplished in 1984–85 and from 6 to 16 years, 10 years later.

The backwardness of the educational development in Spain was overcome in the last three decades of the twentieth century. The 1970 Act (*Ley General de Educación*, LGE) was an important step forward, reinforced by the 1990 Act. Since the late 1970s, the Spanish education system has developed very quickly, even if current indicators are not yet as positive as wished. For example, in 2005, only 49% of the adult population (25–64 years) attained at least upper-secondary education and 28%, tertiary education. Among the young adults (25–34 years) those figures increased up to 64% and 40%, for upper-secondary and tertiary education, respectively, thereby indicating the progress made in this period.

Goals of the Education System

Since the 1970 Act, a great emphasis was put on extending the right to education for all. Therefore, an immediate objective of this time was the quantitative extension of the education system, finally succeeding in making education universal for compulsory ages by mid-1980s and enlarging the proportion of young people over 16 years at school.

The main goals of education were set in the 1978 Constitution, where the following principles are stated:

- Right to education: all citizens have the right to receive education according to their will and needs, and that right must be guaranteed by public authorities.

- Freedom of teaching: all citizens are allowed to create educational institutions, which may be supported by the public authorities under certain conditions; parents may express their preferences in choosing a school for their children; teachers have academic freedom.
- Development of all capacities and talents: education is aimed at developing all aspects of human personality respecting the principles of democratic life and fundamental rights.
- Participation: all sectors will participate in the general programming of educational provision; teachers, parents, and students will participate in the controlling bodies of public and publicly funded schools (school councils).
- Religious and moral freedom: parents are entitled to ask for an education for their children which respects their own religious and moral values.
- University autonomy: universities have the right to develop an autonomous life.

These goals have been developed by several laws: *Ley de Reforma Universitaria* (LRU, 1983); *Ley reguladora del Derecho a la Educación* (LODE, 1985); *Ley de Ordenación General del Sistema Educativo* (LOGSE, 1990); *Ley Orgánica de Universidades* (LOU, 2001 and 2007); *Ley Orgánica de las Cualificaciones y de la Formación Profesional* (LOCFP, 2002); *Ley Orgánica de Calidad de la Educación* (LOCE, 2002); and *Ley Orgánica de Educación* (LOE, 2006).

Once basic education (grades 1–10) was made universal, the main goal has been to offer a high-quality education and training for all. Offering a common education in the context of attention to diversity is one of the outstanding principles concerning basic education in the Spanish legislation.

A common education implies: (1) introducing a comprehensive system from grade 1 to 10, without streaming students and leading to a single diploma of lower-secondary education; (2) defining a national core curriculum, covering 55–65% of the study time and further developed by each autonomous region; and (3) defining a set of eight key competencies to be developed in basic schooling.

Attention to diversity implies: (1) paying attention to personal differences among students; (2) introducing curricular and organizational measures to address different educational needs; and (3) offering optional subjects or ways adapted to personal projects and expectations.

After the period of basic schooling the main aim is to offer equal opportunities, giving every citizen the possibility to continue their studies regardless of their socio-economic situation. This implies the existence of a solid national grants system. In close connection, another aim of post-compulsory education is to offer high-quality education and training for coping with new social and economic demands. It also implies stressing the aim of making lifelong learning real in a national context not always sensitive to it.

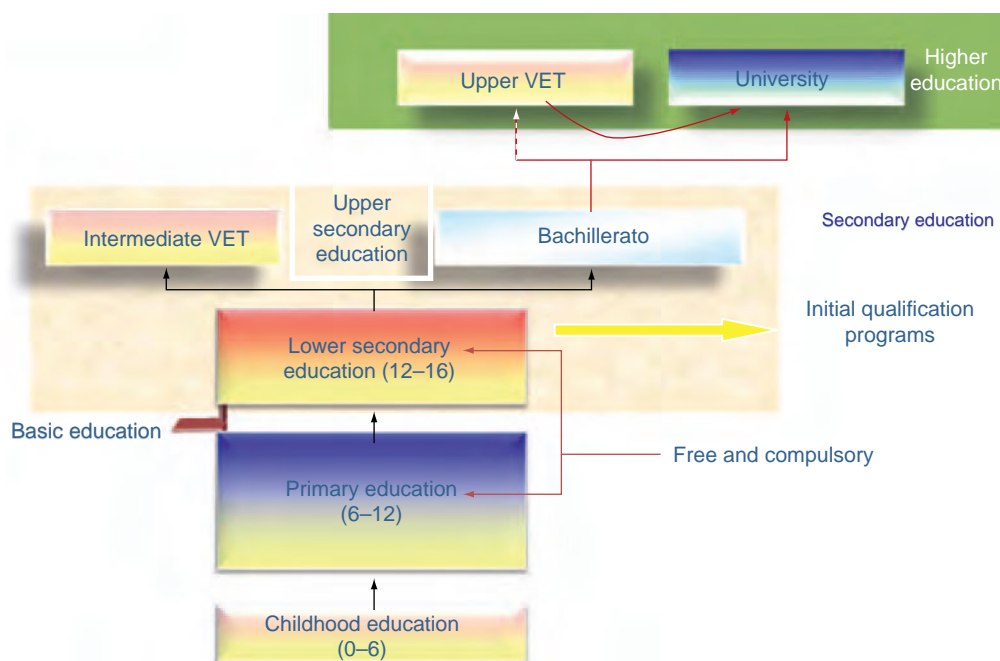


Figure 1 Structure of the formal education system in Spain.

Structure and Operation of the Education System

Figure 1 outlines the structure of the formal education system. Key statistics and indicators can be seen in **Tables 1** and **2**.

Childhood Education

Spanish law puts special emphasis on the educational purposes of childhood attention and care. This level is composed of two cycles: 0–3 and 3–6 years. Both are offered in children's schools, but the second one is also usually offered in mixed primary and children's schools. Contents and organization of the first cycle lies under the responsibility of the regions, but there is a national core curriculum for the second cycle.

In 2006, 96% of 3-year-olds were at school, schooling for 4- and 5-year-olds being almost universal. The rate of children from 3 to 6 years at school rose very quickly from the mid-1990s and now the main goal is enlarging educational provision for children younger than 3 years. A proportion of 16.6% children below 3 were at school in 2006. About 1 620 000 children attended this level in 2008.

Basic Education (Primary and Lower Secondary)

Basic education is compulsory and free of charge for all children from 6 to 16 years. It covers primary education and compulsory secondary education (*Educación*

Secundaria Obligatoria, ESO). The structure of these two levels (set in 1990 and revised in 2006) is the same for the entire country and there is a national core curriculum, even if autonomous regions are responsible for the organization and operation of schools.

Primary education consists of six grades, organized in three cycles. Normal school week covers 5 days with 5 h of tuition, usually with a break for lunch. Most of the primary teachers are nonspecialist, but some are specialized in physical education, music and arts, and foreign languages. Pupils not reaching the objectives and assessment criteria can repeat 1 year along the level, usually at the end of a cycle. In 2008, about 2 603 000 pupils attended this level. Pupils attend primary schools even in very small towns, with special provisions for rural areas (e.g., non-graded schools and associations of rural schools).

Compulsory secondary education consists of four grades. Grades 7–9 are common and grade 10 has a counseling purpose, introducing some optional subjects with different orientations. A normal school week covers 5 days with 6 h of tuition, with or without a break for lunch. Students not reaching the objectives and assessment criteria can repeat up to 2 years. Repetition is frequent in this level (25.2% of students had repeated at least once in 2003). Special programs are offered for students with difficulties. Schooling is possible in this level until 18 years. At the end of the level, students can get a lower-secondary diploma needed to continue studies at post-compulsory level. Secondary teachers are specialists in different subjects, but in some cases they may teach two or more subjects. In 2008, about 1 827 000 students attended this level.

Table 1 Students, teaching staff, and educational institutions, 2005–06

	<i>Institutions</i>	<i>Students</i>	<i>Teaching staff</i>
<i>Nonuniversity</i>	25 575	8 021 220	635 294
Public schools	18 667	5 687 938	470 469
Private schools	6908	2 333 282	164 825
Childhood ^a and primary	14 012	3 970 912	279 275
Secondary and VET	4482	2 983 961	280 855
Specialized	1439	657 425	29 626
Other	5642 ^b	408 922 ^c	45 538 ^d
Males		3 951 500	196 025
Females		4 069 720	398 343
<i>University</i>	71 ^e	1 510 072	109 838
Public	48	1 367 769	93 995
Private	23	142 303	15 843
Males		692 941	69 967
Females		817 131	39 871

^aOnly centers authorized by educational administrations.

^bSpecial education, formal adult education, and centers offering several levels are included.

^cSpecial education and formal adult education are included.

^dSpecial education, formal adult education, and teachers working in several levels are included.

^eNumber of universities (total number of university centers: 998 – 864 public and 134 private).

Adapted from Ministerio de Educación y Ciencia (2007b). *Estadística de las enseñanzas no universitarias. Resultados detallados. Curso 2005–06*. Madrid: Ministerio de Educación y Ciencia and Ministerio de Educación y Ciencia (2007c). *Las cifras de la educación en España. Estadísticas e indicadores. Edición 2008*. Madrid: Ministerio de Educación y Ciencia. (www.mepsyd.es) VET – vocational education and training.

Private schools offering these ten grades can receive public funds if they meet some legal requirements concerning admission policies, parents' participation, and no compulsory fees. They usually offer both levels in the same school, sometimes together with lower and upper levels. Public schools usually offer both levels in different institutions – primary schools and secondary institutes – even if some new models are being developed. A total of 23 678 institutions offered childhood, primary, and secondary education in 2008, of which 16 725 were public and 6953, private. In all, 33% of primary and lower-secondary students attended private schools in 2006 (less than 5% attended nonsubsidized ones).

Wherever possible, children with disabilities or special needs are integrated in normal schools and classrooms. Nevertheless, in 2008, a total of 29 555 children attended special schools.

Upper-Secondary Education

After basic education, students getting the lower-secondary diploma (in 2006 only about 70% of the age cohort) have two

Table 2 Selected educational indicators

<i>Indicator</i>	<i>Year</i>	<i>Data</i>
School expectancy at 6 years	2005–06	14.5 years
Schooling at 3 years	2005–06	96%
Schooling at 18 years	2005–06	62.3%
Early school leaving ^a	2006	29.9%
Lower-secondary graduation ^b	2004–05	70.4%
Low skills in science ^c	2006	19.6%
Upper-secondary education ^d	2006	61.6%
Postsecondary education ^e	2004–05	49.9%
Continuous training ^f	2006	10.9%
Public expenditure on education ^g	2007	4.36%

^aProportion of young people 18–24 having ended lower-secondary education at most and not continuing further studies.

^bRate of students getting the lower-secondary diploma compared to young population of the corresponding age group.

^cProportion of 15-year-old students in levels 1 or less in the science scale in Programme for International Student Assessment (PISA) 2006.

^dProportion of young people 20–24 with an upper-secondary diploma or certificate.

^eRate of students getting a post-secondary degree or certificate compared to young population of the corresponding age group.

^fProportion of adult people 25–64 having participated in a training activity during the month before survey.

^gPublic expenditure on education as a percentage of GDP

Adapted from Ministerio de Educación y Ciencia (2007). *Las cifras de la educación en España. Estadísticas e indicadores. Edición 2008* Madrid: Ministerio de Educación y Ciencia and *Informe 2007: Objetivos educativos y Puntos de referencia 2010* (www.mepsyd.es); OECD: *Education at a Glance* (2007). OECD Indicators. Paris: OECD.

different ways to continue their studies: upper-secondary education (*Bachillerato*) or intermediate-vocational training.

Bachillerato consists of two grades, organized by specialties: science and technology, social sciences and humanities, and arts. Students take some common courses plus a number of optional subjects, depending on the specialty. The 2006 law gave it a more flexible structure, aimed at preparing students for the new model of higher education. Students passing all subjects get the *Bachillerato* diploma, allowing them to go on to higher education. Students wishing to go to university must also pass an external examination, adapted to the different specialties, whose marks serve as the basis for choosing university degrees and institutions. In 2008, a total of about 625 000 students were at this level, 25% of them in private schools.

Vocational Training

Vocational training is offered in Spain in different ways. The education system offers formal intermediate- and upper-vocational training programs, usually full time for young students. Labor administrations offer both nonformal training courses for employment and continuous training activities for employees in cooperation with

trade unions and business organizations. Different types of formal and nonformal vocational training courses are offered in integrated centers, recently created.

Formal intermediate-vocational training is offered for students holding the lower-secondary diploma or passing a specific entrance examination. It is a part of the post-compulsory secondary education and usually lasts 2 years. It includes education in a school and a training period in a workplace.

Formal upper-vocational training is offered for students holding the *Bachillerato* diploma or passing a specific entrance examination. It is a part of the higher education system and usually lasts 2 years. It also combines tuition in an educational institution and a practical training period.

Both formal and nonformal vocational training are adapted to a National Professional Qualifications Catalog, allowing the organization of modular courses and the addition of successive periods of training. The NQS is also the basis for the recognition and accreditation of competencies acquired through working and social experience.

Vocational training has been traditionally underdeveloped in Spain. This is why the number of vocational education and training (VET) students is lower than those involved in academic ways. In 2005, 45% of the young population got a *Bachillerato* diploma, but only 25% got an intermediate-VET certificate, and 32.7% got a university degree, but 17.2% got an upper-VET certificate. The proportion of students in VET programs is currently growing due to the excellent rates of employment of those professionals. In 2008, there were 235 110 students in intermediate-VET programs and 208 830 students in upper ones.

University Education

Spanish universities have been offering two types of basic official degrees: first-cycle degree (*Diplomado*), lasting 3 years, and second-cycle degree (*Licenciado*), lasting 4 or 5 years. Moreover, they offer PhD studies. Master has not been until very recently an official degree, but universities did offer a number of them with professional focus and content. From 2009 on, the implementation of the Bologna scheme will change significantly the structure of university studies, moving to a different system of three degrees: graduate, master, and PhD. It will mean an important change for Spanish universities, in convergence with European ones.

In 2008, Spain had 75 universities – 50 public and 25 private. Public ones are economically dependent on the autonomous regions, but have organizational and academic autonomy. Governing boards of public universities are composed by academic authorities, teaching and services staff, and students, and the rector is elected by university members. They are accountable for the quality of the studies they offer to a national or regional agency for evaluation and accreditation.

In 2008, a total of 1 381 749 students were following official, non-PhD studies at university level, 557 511 in first-cycle degrees and 824 238 in second-cycle degrees. Percentage of young population getting a university degree is currently higher than the OECD average.

Other Formal Studies

The Spanish education system also includes some specialized studies: arts and design, drama, music and dance, languages, and sports. They are offered, whether at basic, intermediate, or upper level, in public or private schools. In 2007, a total of 695 098 students followed these courses. The most important ones are languages (381 367 students) and music (259 132 students).

Provision of education for adult population is important in Spain. Providers of this kind of educational activities are very diverse, offering from formal, second-chance education to nonformal education, as well as cultural and leisure activities. Adults involved in formal education were 366 670 in 2007, 62.1% of them being women. A great majority (almost 80%) of persons involved in nonformal adult education are also women. Adult centers, popular universities, the so-called universities of the experience, and other similar initiatives are being boosted by different institutions and administrations.

Distance teaching is offered in Spain, both at university and non-university level and mixed programs are being developed by traditional institutions.

The Teaching Profession

In 2008, a total of 455 130 teachers worked in public schools and 168 844 in private ones. University teaching staff was of 107 905, all categories included.

Since 2009, pre-primary and primary teachers need to have a specific first-cycle university degree, to be transformed into a 4-year graduate degree, including practical training. Secondary and VET teachers need to have a *Licenciado* or a graduate degree in any disciplinary field, plus a 1-year master degree of pedagogical and didactic orientation.

Teachers in the public system have the status of civil servants, after passing a rigorous concourse. Private institutions select their teachers by themselves. Teaching is a feminized profession, women being about two-thirds of non-university teachers.

Between 1998 and 2008, the number of teachers has increased by 24.2%, while students have decreased by 4%. In consequence, the student-teacher ratio in 2005 was 14.3 in primary education, 13.4 in all secondary education, and 11.6 in all tertiary education, significantly lower than OECD averages (16.7, 14.2, and 15.8, respectively).

In the years to come an important change in the teaching force will be produced. It is expected to incorporate about 200 000 new teachers, due to retirement and the need to create new schools. Educational administrations are thus giving high priority to the development of the new model for teacher training. No shortage of teachers is expected.

Administrative and Supervisory Structure

Spanish regions have important responsibilities in the operation of the education system. Central government keeps some crucial responsibilities to assure the existence of one single system, although some significant regional differences do exist.

Central government is responsible for setting the conditions for delivering official diplomas and certificates, establishing the basic rules for the structure and organization of the education system, defining the national core curriculum for all educational levels, and assuring equal opportunities for all students in Spain. One important responsibility lies with the High Inspectorate, a body which assures that educational regulation of autonomous regions agrees with national legislation and respects central-government responsibilities.

Autonomous regions are responsible for all the organization and operation of schools and educational services. For instance, they have autonomy for setting the model of school organization, selecting and training teachers, developing school curricula and pedagogical methods and tools, assessing the quality of teaching, and approving subsidies for private schools. Some of these responsibilities should be developed under a basic national framework or regulation. According to this distribution of roles, regions have their own inspectorate services to supervise schools.

There are some national bodies aimed at fostering cooperation among central and regional administrations. The most important ones are the Conference of Ministers of Education and the General Conference for University Policy, but there are others like the General Council for Vocational Training or the Higher Council for Arts Education. Central and regional governments also develop some cooperation programs, jointly defined and financed, in fields of special interest for educational development and improvement.

Educational Finance

The quantitative expansion of the education system in Spain has implied a higher expenditure in all levels. In absolute terms, the educational public expenditure has more than doubled from 1997 (22 785.5 million €) to 2007 (45 617.6 million €). Nevertheless, the important economic growth in this period has resulted in a decrease in relative expenditure to GDP (4.52% in 1997, 4.36% in 2007). Private expenditure in education has also grown, but in a

smaller proportion, from 5528 million € in 1997 (1.10% of GDP) to 8810 million € in 2007 (0.84% of GDP).

As the number of students has decreased by almost 300 000 in this decade, expenditure per student has also increased. In 2004, expenditure per student in public institutions related to GDP per capita was higher than the European average for both primary and secondary education (22.6% and 30.3% in Spain, 20.9% and 26.5% in the 27 member states of the European Union (EU-27)), but lower for tertiary education (37.4% in Spain, 39.7% in EU-27).

As for different education levels, in 2005, 32.9% of public expenditure in education was devoted to pre-primary, primary, and special education; 30.6% to secondary education (including *Bachillerato* and VET); 19.6% to university; and 7.2% to other educational activities. The national grant system received 3.1% of public expenditure. Grants do receive big attention as one of the outstanding systems for assuring equal opportunities. Central government contributed in 2008 with 1482.26 million €.

The distribution of public expenditure in education reflects the distribution of responsibilities among central and regional governments. In 2005, only 4.6% of the total public expenditure came from the central government, while some of the biggest regions, like Andalusia or Catalonia, contributed with 16.7% or 14.7%.

Students pay no fee in public schools or in subsidized private ones. University studies are highly subsidized in public universities, students covering only about 20% of the real cost. Annual fees cost between 500 € and 800 € in public universities and can be higher than 6000 € in private ones.

Performance Monitoring, Evaluation, and Research

Performance monitoring is a shared responsibility between the Ministry of Education and the regions. Each of the 17 regions is responsible for the monitoring and evaluation of the education system and schools in its own territory. Some of them have created regional centers or institutes for evaluation and have developed school-evaluation programs. There is a national statistical plan which allows the collection of national data on education and training on a regular basis and the contribution to international statistics.

The Evaluation Institute, belonging to the Ministry of Education but governed in cooperation with the regions, sets the frame for the general evaluation of the education system and the collection of national indicators. National evaluation programs set in place in the 1990s are now developing toward a new model based on the assessment of key competencies.

The Evaluation Institute also coordinates Spanish participation in international evaluations. Most of the Spanish

regions are participating in PISA with comparable samples, allowing comparison with other countries and regions.

Universities and their teaching staff are evaluated by the National Agency for Quality Evaluation and Accreditation (ANECA) or the corresponding agencies created by some regions.

Educational research is mainly undertaken by universities, with the participation of some public or private centers or foundations. The Center for Educational Research and Documentation, belonging to the Ministry of Education, plays a significant role in fostering research and disseminating its outcomes.

Major Changes and Issues since the 1990s

In the last few years three outstanding issues have been raised: the impact of immigration on schools, the need to reduce school failure, and the necessary European convergence.

The important immigration movement since the late 1990s has provoked a number of effects in the education system and in schools. On the one hand, it has changed the trend of a decreasing student population, raising it again and demanding more resources. On the other hand, it has raised new challenges for school inclusion and multicultural education, in the perspective of future social cohesion. From 1997 to 2007 the number of students of foreign origin has multiplied 9 times, reaching a proportion of 8.4% in 2008 (in some regions is higher than 12%). For a country traditionally of emigrants, this means a big change.

School failure and early school leaving are two important problems in Spain. Almost 30% of students don't get the lower-education diploma in due time (some of them do it later through adult education), and in 2006, 29.9% of young people from 18 to 24 years had left school after lower secondary without continuing post-compulsory studies. Early school leaving implies that the Spanish population holding a post-compulsory diploma is lower than the OECD average. The Conference of Regional Ministers has taken some decisions and put in practice a set of cooperation programs to reduce these rates by 2010 and afterward.

The EU has defined a set of educational objectives for 2010. Spain is committed to them, but still below the benchmarks for early school leaving, upper-secondary education attainment, and skills level of the 15-year-old population. A national strategy has been adopted for improving those results thereby approaching the European objectives in the coming years.

Further Reading

- Alvarez Lázaro, P. (2001). *Cien años de educación en España. En torno a la creación del Ministerio de Instrucción Pública y Bellas Artes*. Madrid: Ministerio de Educación, Cultura y Deportes and Fundación BBVA.
- Bonal, X., Rambla, X., Calderón, E., and Pros, N. (2005). *La descentralización educativa en España. Una mirada comparativa a los sistemas escolares de las Comunidades Autónomas*. Barcelona: Fundació Carles Pi i Sunyer.
- European Commission (2007). *Progress Towards the Lisbon Objectives in Education and Training. Indicators and Benchmarks*. Brussels: European Commission.
- Gimeno Sacristán, J. and Carbonell Sebarroja, J. (eds.) (2004). *El sistema educativo. Una Mirada crítica*. Barcelona: Praxis.
- Instituto de Evaluación (2006). *Sistema estatal de indicadores de la educación. Prioritarios 2006*. Madrid: Ministerio de Educación y Ciencia – Instituto de Evaluación.
- Instituto de Evaluación (2007). *PISA 2006. Informe español*. Madrid: Ministerio de Educación y Ciencia – Instituto de Evaluación.
- Ministerio de Educación y Ciencia (2007a). *Informe 2006: Objetivos educativos y Puntos de referencia 2010*. Madrid: Ministerio de Educación y Ciencia.
- Ministerio de Educación y Ciencia (2007b). *Estadística de las enseñanzas no universitarias. Resultados detallados. Curso 2005-2006*. Madrid: Ministerio de Educación y Ciencia.
- Ministerio de Educación y Ciencia (2007c). *Las cifras de la educación en España. Estadísticas e indicadores. Edición 2008*. Madrid: Ministerio de Educación y Ciencia.
- Ministerio de Educación y Ciencia (2007d). *Estadística del gasto público en educación. Año 2005*. Madrid: Ministerio de Educación y Ciencia.
- Ministerio de Educación, Cultura y Deportes (2001). *La educación en la España del siglo XX. Primer centenario de la creación del Ministerio de Instrucción Pública y Bellas Artes*. Madrid: Ministerio de Educación, Cultura y Deportes.
- OECD (2007). *Education at a Glance 2007. OECD Indicators*. Paris: OECD.
- Prats, J. and Raventós, F. (2005). *Los sistemas educativos europeos. ¿Crisis o transformación?* Barcelona: Fundación La Caixa
- Rey, R., de Puelles Benítez, M., Tiana Ferrer, A., et al. (2004). *Análisis y situación de las Comunidades Autónomas*. Madrid: Santillana.
- Viñao, A. (2004). *Escuela para todos. Educación y modernidad en la España del siglo XX*. Madrid: Marcial Pons.

Relevant Websites

- <http://www.csic.es> – Consejo Superior de Investigaciones Científicas.
- <http://www.educaweb.com> – educaweb.
- <http://www.ine.es> – Instituto Nacional de Estadística.
- <http://www.institutodeevaluacion.mec.es> – Instituto de Evaluación-Ministerio de Educación.
- <http://www.micinn.es> – Ministerio de Ciencia e Innovación.
- <http://www.mtas.es> – Ministerio de Trabajo e Inmigración.
- <http://www.eurydice.org> – The information network on education in Europe.
- <http://www.oei.es> – Organización de Estados Iberoamericanos.
- <http://www.mepsyd.es> – Portada-Ministerio de Educación.
- <http://www.inem.es> – Servicio Público de Empleo Estatal.
- <http://www.universia.es> – Universia España.

Sri Lanka

B M J Balasooriya, M M Wehella and S U Wijeratne, Ministry of Education, Battaramulla, Sri Lanka

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Free education policy and education subsidies –

Providing education free for all students from grade 1 to first degree (at university) level. Additionally, it is implementing a number of education subsidies programs such as: providing free textbooks for students in grades 1–11, free school uniforms for all students, mid-day meal for primary students and students in disadvantaged areas, public transport subsidies, scholarships and bursaries, special provisions for children with special educational needs (SEN) etc., and ensuring free education policy.

Generic skills – Examples are creativity, communication, problem solving, and decision-making.

Higher-order capital assets (HOCA) – Learning equipment, technology, books, reading materials, library, and computers.

Higher-order inputs – All materials, equipment and instruments used to improve the quality of education.

Higher-order learning spaces (HOLS) – Learning spaces for higher-order transferable skills of students: activity rooms, computer centers, libraries, laboratories, auditorium, gymnasium, and vocational/technical complex.

Higher-order processes (HOPs) – All services to improve students' higher-order skills.

Sri Lanka at a Glance

Background

Sri Lanka is an island in the Indian Ocean, situated to the south of the Indian subcontinent with a land area of 65 000 km², lying a few degrees north of the equator. It has a tropical climate. With nature's gifts of rain and fertile soil, the principal means of livelihood of the people from the dawn of history has been agriculture.

The proximity to the Indian subcontinent created close affinities between Sri Lanka and India. The original settlers were immigrants from India spreading principal religions (Buddhism and Hinduism) in the country. The ancient culture, language, art, architecture, technology, and social customs were of Indian origin. Political developments in

India, too, had their repercussions. The island was subjected to frequent invasion from South India.

However, Sri Lanka developed a culture of its own, with a distinct identity. Buddhism, which originated in India, remains the principal religion in Sri Lanka. Similarly, the language and other cultural aspects have developed their own distinguishing characteristics. The strategic location on the main sea routes resulted in travelers visiting this country in the past. The Greek and Arab merchant sailors described the island as Taprobane, Serendib, Ceilao, and finally Ceylon. The present name, Sri Lanka, is used with the promulgation of the Republican constitution in 1972. The European explorers conquered the island in the sixteenth century and ruled for four-and-a-half centuries.

Sri Lanka gained independence from the British in 1948. After independence, the country was governed by a democratically elected, Westminster type of government. In 1972, it became a Republic, and, in 1978, an executive presidency model was adopted. At the center, there is the Legislature, the Parliament, and the Executive President, but power is substantially devolved to elected provincial councils, and the local government system assures people's participation in managing their local affairs.

People, Society, and Economy

The population of Sri Lanka in 2006 was 19.5 million and the annual growth rate was 1.2%. It is estimated that by the year 2020, the population will stabilize at 22 million. The decline in the rate of population growth will enable the authorities to divert more resources for quality improvement in the education sector.

Sri Lanka is a multiethnic, multireligious society, with 74% Sinhalese, 12.6% Sri Lankan Tamils, 5.5% Indian Tamil, 7% Moors, and 0.8% others. It is predominantly an agricultural country; 72% of the population lives in rural areas, 21% in urban areas, and 6% in the plantation sector.

According to the United Nations Organization (UNO) Human Development Index (HDI), Sri Lanka enjoys a high quality of life in comparison to countries having a similar or a higher level of per capita income. Some of the components taken into account in the compilation of the HDI, such as literacy at 92%, life expectancy of 74 years, and infant mortality at 13 per 1000, are impressive achievements in social development. These are the results of the social welfare measures pursued by successive

governments after independence, to which free education has made a remarkable contribution.

On gender issues, the country has followed an enlightened policy. The right to vote, irrespective of any restrictions, granted with universal suffrage in 1931 and the provision of educational facilities for women have established gender equity in Sri Lankan society. Women outnumber men in enrolment in secondary education and in professions such as teaching.

Sri Lanka is a developing country with a per capita income of US\$ 1030 per person. The economy has been expanding at a real gross domestic product (GDP) growth rate of 7.4% in 2007. According to the World Bank, the economic progress of the country has been the result of market-oriented development policies that enable private sector participation in the industrial and service sectors.

Historical Evolution of Education

Ancient Period

The ancient period is divided into three eras: pre-eras of the arrival of Vijaya; period from Vijaya to King Devanampiyatissa; and period from the Arhanth Mahinda to year AD 1500. As the early history (543 BC to AD 1500) of Sri Lanka reveals, close affinity and cultural relationship with India from the Vedic times have had their impact on Sri Lanka. The introduction of Buddhism to Sri Lanka by Arhanth Mahinda during the reign of King Devanampiyatissa laid a firm foundation for Buddhist civilization that led to the establishment of Pirivenas – monastic institutions that were intended for the education of clergy but also enrolled lay students.

The Buddhist temples in villages became the institutions for primary education, while the Pirivena, where the Bhikkus received their teaching in the Dhamma, became the institution for secondary education, and the Mahavihara turned out to be the center for tertiary education. During the Polonnaruwa period (AD 1073–1215), more Pirivenas were started and the curriculum included Sinhala, other languages, and art. Nonformal education (NFE) provided an education of a high standard in subjects like handicraft, fine arts, and engineering.

Pre-independence Era

Portuguese period (AD 1505–1658)

When the Portuguese invaded Sri Lanka in 1505, the country was ruled by the kings of Kotte, Kandy, and Jaffna for about 150 years. The Portuguese introduced an education system in the provinces and started many schools, the main aim of which was conversion of local population to the Catholic religion. The curriculum consisted of Catholic religion, reading, writing, arithmetic, and languages, and the teachers were mostly Catholic priests.

Dutch period (AD 1658–1796)

The Dutch conquered the maritime provinces of Sri Lanka. The aims of their education system were to convert local people to Christianity and to suppress the spread of Catholic religion; teach reading and writing to new converts; and train local people to assist them in administration. Under the Dutch with their system of free compulsory education, large numbers of girls attended school. During the Portuguese and Dutch occupation, the state of affairs was not favorable for Buddhist education, which largely disappeared.

British period (1796–1948)

The British had their influence in the maritime provinces since 1796, before they conquered the Kandyan Kingdom. The period from 1805 to 1824 was known as the missionary period. Various Christian missionary societies started schools in all provinces of the country, and the large number of Christian schools led Buddhist, Hindu, and Muslim organizations to establish schools in premises donated by the community. As many different organizations were involved in education during this period, there was no proper supervision or administration. In 1841, a school commission was established but was replaced in 1869 by the Department of Public Instruction responsible for the supervision of schools. The curriculum was secularized, and several higher education colleges were established.

In 1931, Dr. C.W.W. Kannangara became the country's first minister of education. His efforts cleared the way for the establishment of a system that would ensure equal opportunities for all children. In 1942, a special committee recommended providing free education from the kindergarten to the university; using mother tongue as the medium of instruction in the primary schools; teaching English in all schools from standard III; and introducing a curriculum for the child which would develop its "head, heart, and hands."

Postindependence Era (1948–present)

Since independence, the government has given the highest priority to education. Within a period of less than 60 years, the number of schools increased by over 50%, the number of students increased by more than 300%, and the number of teachers increased by more than 400%. The literacy rate has grown correspondingly and by mid-1980s, over 90% of the population was literate.

The difficulties of educational planning in the past caused by the system of dual control and lack of coordination were finally removed by the nationalization of schools in 1960. The National Education Commission (NEC), in 1961, recommended a system of school zones, and the establishment of regional education departments and district education offices.

In 1970, there were 2.7 million students in 9928 government schools. Due to the increasing demand, government of

Sri Lanka (GoSL) faced the problem of improving the quality of education and the mismatch between education and the labor market. Therefore, GoSL implemented a range of reforms in 1972, and the 1981 White Paper on education recommended a new subject (life skills) in place of the two pre-vocational subjects; a dual system of evaluation; and a school cluster system. The 1997 education reforms covered major policy areas such as extending educational opportunities; improving the quality of education; developing practical and technical skills; education and training of teachers; and management and resource provision.

Present Education System and Administrative Structures

Sri Lanka faces all the political, economic, and social problems common to all developing countries. Despite these problems, experienced during the past 60 years since independence in 1948, the literacy rate increased to 91.6% in 2006. The government has accepted that education is a right of each citizen, and the Sri Lankan

constitution ensures everyone the right to universal and equal access to education at all levels.

This principle of providing equal opportunities has been a prominent characteristic of the policies of all governments that came to power since independence. Changing the medium of instruction from English to the mother tongue (Sinhala/Tamil) and the provision of financial aid to less-privileged students are other measures along regulations enforcing compulsory attendance for all children. However, Sri Lanka has not attained equity in education. Economic backwardness, climatic conditions in certain regions, the cultural traditions of certain ethnic groups, and problems in the employment and deployment of teachers in the school system have caused difficulties in achieving education-for-all targets.

Organizational Structure and Operation of the Education System

The education sector is organized into four major stages (see **Figure 1**). Sri Lanka has a 13-year schooling system that provides general education where age of admission to

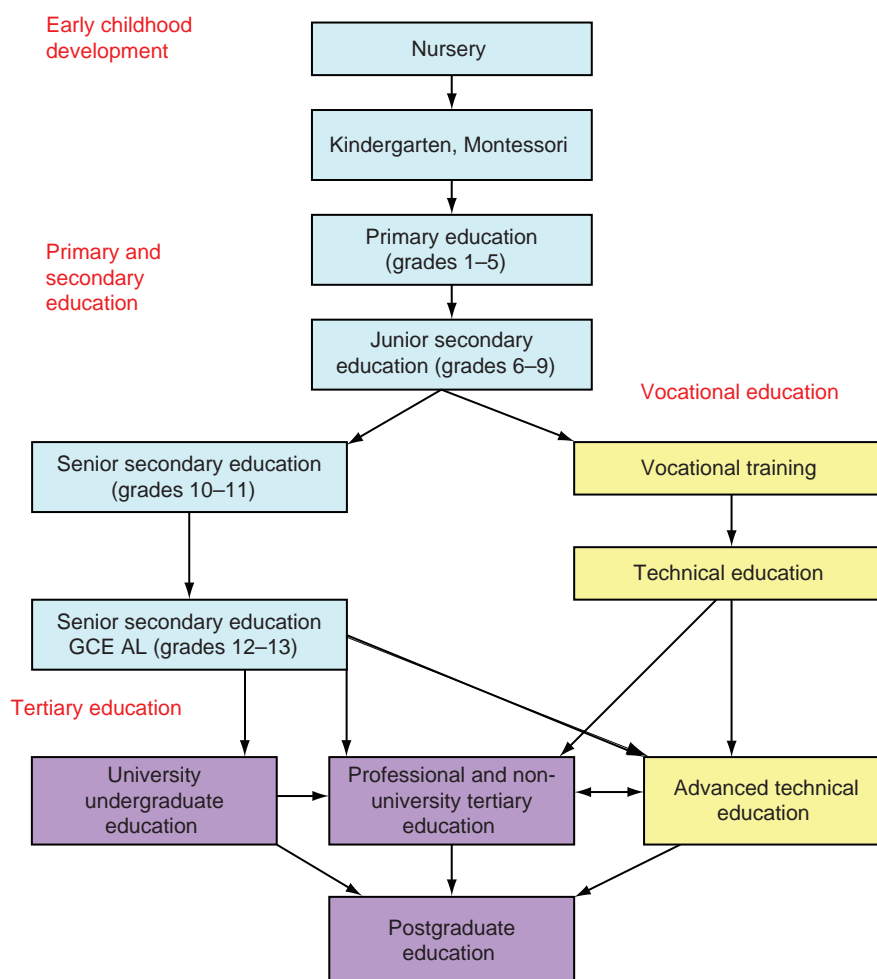


Figure 1 Organizational structure: National system of education in Sri Lanka.

grade 1 is 5 years. The education system is divided into three main levels as primary, secondary, and tertiary. The system has a compulsory phase from grades 1 to 9. The public sector accounts for 93% of schools and 95% of student enrolment.

Preschool/early childhood care

The earliest stage, catering to children aged 3–5, is that of early childhood development. It is not included in the formal public education system as virtually all preschools are in the private sector.

Primary and secondary education

The second stage includes formal primary, junior secondary, and senior secondary education. The primary school from grades 1–5 enrolls about 1.6 million students (50% are female). There are 3034 exclusively primary schools. Except for a few 1AB schools, all other schools offer primary education. The student–teacher ratio in primary grades is 1:26. The primary school curriculum is an activity-based curriculum and there is a reasonable level of equity in achievement levels. Annually, around 300 000 students enrol in grade 1. Net enrolment ratio has been 98% and the net completion ratio at grade 5 is 95%. The completion ratio at the end of compulsory basic education, at grade 9, is 83%.

The secondary school includes grades 6–13; the grades 6–11 form the junior secondary school; and grades 12–13 are the senior secondary school. The grades 6–13 have about 2.3 million students (50% are female). At the end of grade 11, students sit for the General Certificate of Education Ordinary Level (GCE OL) examination to qualify for senior secondary schooling.

The junior secondary school curriculum from grades 1–9 is oriented to provide all basic skills before one leaves school. Grades 10 and 11, leading to GCE OL examination, have core and optional subjects. The senior secondary school curriculum is streamlined to offer specializations in science, arts, and commerce streams. At the end of grade 13, a student sits for the General Certificate of Education (Advanced Level) (GCE AL) examination.

Tertiary education: Vocational and university education

The tertiary education system consists of the public and private universities, technical colleges, and tertiary-level institutions. Examination performance at GCE AL serves as the basis for admission to the public universities. Only about 34% who complete grade 13 get access to tertiary-level courses. There are 15 universities, including the Open University of Sri Lanka (OUSL) and the Sri Lanka Institute of Advanced Technological Education, 17 national colleges of education (NCoEs), 31 teacher training colleges (TTCs), and many other public and private sector higher education institutions (HEIs) providing tertiary education. Sri Lanka has a widely spread apprenticeship training system with three national-level institutes, five regional centers, and 52 vocational centers that absorb only about 10 000 trainees.

Public School System

In 2006, the total number of government schools was 9714, educating nearly 3.84 million students, and these schools are classified as:

- 1AB – classes up to grade 13, with GCE AL science, arts, and commerce streams;
- 1C – classes up to grade 13, and with GCE AL arts, and/or commerce streams;
- Type 2 – classes up to grade 11; and
- Type 3 – classes up to grade 5/8.

In addition, there is another category of government schools called national schools, mainly type 1AB. These schools are directly administered and financed by the Ministry of Education (MoE), whereas the administration and finance of the other schools are functions of the Provincial Education Authorities (PEAs). The total numbers of schools, teachers, and pupil enrolments in 2006 are given in **Table 1**.

Thirty-three percent of schools in the country are type 2, and 40% are type 3 schools. Types 2 and 3 schools are predominantly found in the rural areas where multi-grade teaching takes place with few teachers. Out of the

Table 1 Numbers of schools, teachers, and pupils by types of schools, 2006

<i>Type of school</i>	<i>No. of schools</i>	<i>No. of teachers</i>	<i>No. of pupils</i>	<i>Student–teacher Ratio (STR)</i>
National (1AB or 1C)	327	31 088	681 510	22
1AB	367	24 872	541 080	22
1C	1 820	62 474	1 220 456	20
Type 2	4 229	67 157	1 056 650	16
Type 3	2 971	19 054	337 852	18
Total	9 714	204 645	3 837 548	19

From Ministry of Education (2006a). *School Census–2006*. Battaramulla: Ministry of Education.

9714 schools, 4906 are small schools (less than 200 students). Most of the types 1AB, 1C, and type 2 schools include primary sections. However, schools in remote areas lack human and physical resources. Further, some schools are situated in tea and rubber estates where 6.3% of the population live and the level of human poverty is the highest.

Semi- and Nongovernment Educational Institutions

In addition to the government schools, there are three other types of educational institutions: pirivenas, private schools, and international schools (see **Table 2**).

Pirivena education

Pirivenas are functioning under the legal provision of the Pirivena Act, No. 64 of 1979. State-funded Pirivenas provide formal education from the primary level to higher education for Buddhist monks and lay male students. They are run by the head of each institution under the supervision of MoE. Every Pirivena is entitled to a per-pupil subsidiary grant provided by the government. From 2006, that the government has provided capital grants to improve their infrastructure facilities and provided learning resources enabling them to increase access and participation in education.

Private schools

Private schools are run by the management board of each school, and charge pupils for education. They fall under three categories (see **Table 3**).

The first category does not receive any funds from the government, while the second and third categories receive free textbooks, free school uniforms, and funds for the purchase of learning materials. Additionally, the third category is provided with salaries for its academic staff.

The government does not provide capital budgets for any of these categories. They are directed to follow the national curriculum.

Special education schools

GoSL does not provide capital budget for these schools, but the salary bill for teachers is paid. In addition to the pupils in these schools, 58 626 pupils (1.5% of enrolled students) were on the rolls in the mainstream of normal schools.

International schools

International schools offer English-medium education and charge fees for tuition. These schools have been established as companies and are registered with the registrar of companies and the Board of Investment in Sri Lanka. Around 200 international schools with a student enrolment of about 70 000, which prepare students to sit overseas examinations, were functioning in the whole country in 2007. GoSL does not provide any financial assistance to these schools. In 1999, the government has made the legal provision for pupils in international schools to take public examinations, but the autonomy of these schools keeps them outside the purview of MoE.

Nonformal Education

NFE initiatives of the government could be traced to the early 1970s. Then, NFE offered vocational training courses in some schools after school hours, adult education programs, weekend English courses for adults, and literacy courses for children of school-going age who were out of schools. However, the NFE program did not expand, given the relatively small number of school dropouts (lower than other countries in the region).

One of the main focuses of the 1997 reforms was to provide education for all children of the school-going age. The intention was to provide facilities for every child in the age group of 5–14 years to obtain education in schools or through suitable NFE programs. Approximately 8500

Table 2 Semi- and nongovernment educational institutions, 2006

<i>Institution/schools</i>	<i>No. of institutions</i>	<i>No. of teachers</i>	<i>No. of students</i>
Pirivenas	653	5 575	54 899
Private schools	80	4 009	95 092
Assisted special education schools (catering for disabled and retarded children)	24	459	2 924
International schools	200	NA	70 000

From Ministry of Education (2006a). *School Census–2006*. Battaramulla: Ministry of Education.

Table 3 Categories of private schools, 2006

<i>Category of school</i>	<i>No. of schools</i>	<i>No. of students</i>	<i>No. of teachers</i>
Recognized and certified	10	8 008	431
Unaided fee-levying private	15	21 938	1126
Aided non-fee-levying private	35	65 146	2452

From Ministry of Education (2006a). *School Census–2006*. Battaramulla: Ministry of Education.

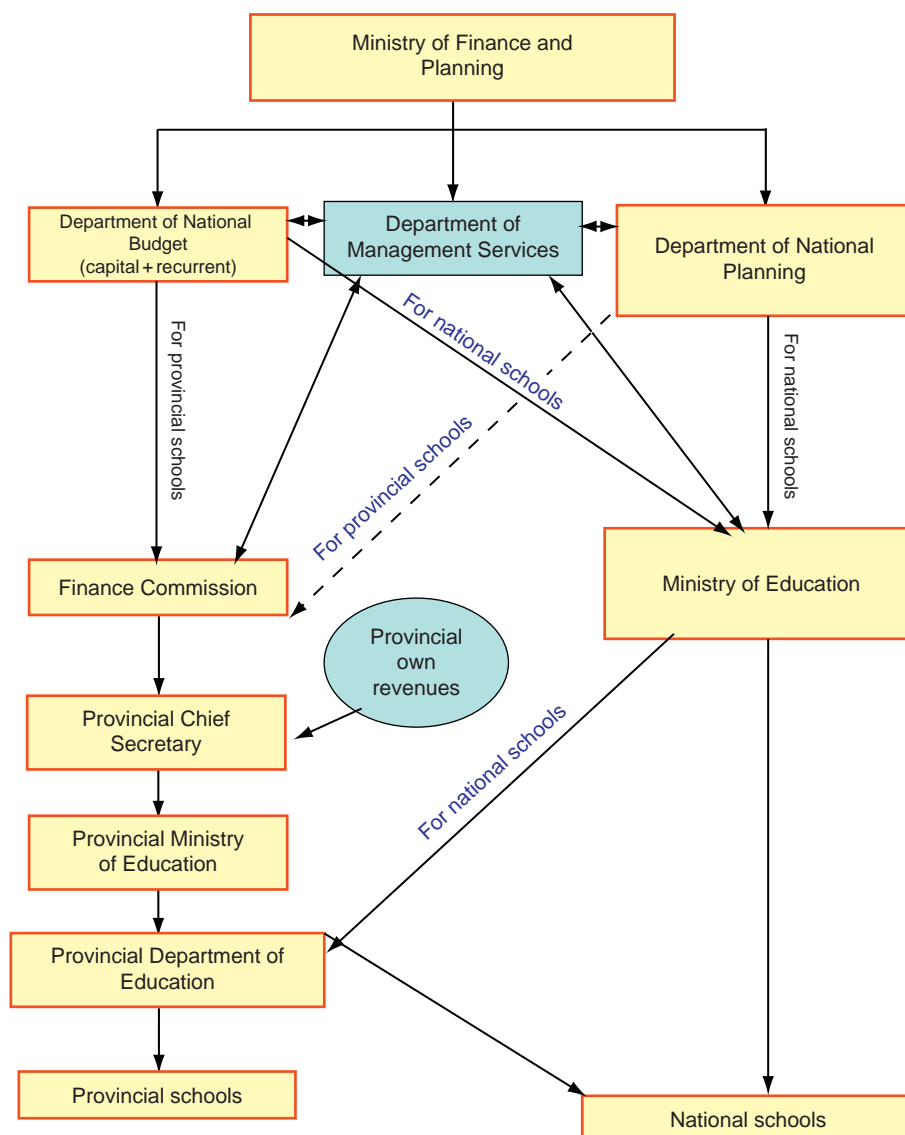


Figure 2 Education financial allocation flow in Sri Lanka, 2007.

school attendance committees were established for monitoring the attendance of children. NFE programs include career guidance programs; literacy classes; income generating programs; and programs for street children. Community learning centers also conduct NFE programs, and the NIE has introduced an open school to provide opportunities for persons who obtain basic literacy and pre-vocational training to receive recognized certificates.

Teacher Education

At present, 17 NCoEs provide preservice training to all teachers to be recruited. As many as 8692 students were trained as cadet teachers by 504 teacher educators. Continuing education will be provided through a network of 100 teacher centers (TCs). In addition, there are ten TTCs with the 1922 teacher trainees and 156 academic staff

operating in the system. Recently, teacher educators have promoted school-based teacher development programs.

Education Management and Administrative Structures

The governance framework for education is somewhat complex and combines elements of deconcentration, delegation, and devolution of functions and power. The central government is responsible for national education policy at all levels. However, PEAs play an important role in the administration of the school system.

Current policy thinking is to further devolve education management to empower frontline service providers (i.e., principals, section-heads, and teachers), and involve local communities. The present decentralized management

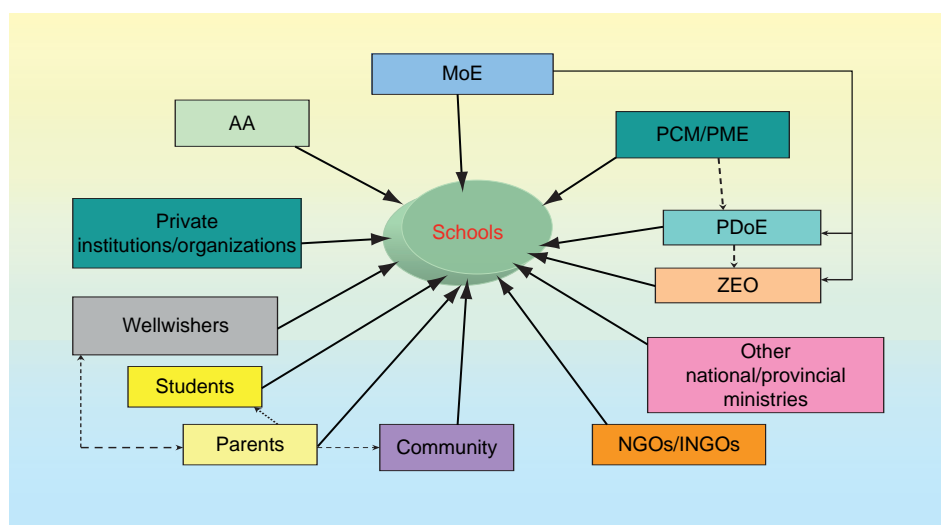


Figure 3 Pattern of provision of school financing and resources: Sri Lanka. Balasooriya, B. M. J. (2004). *An Evaluation of the Impact of School-Based Resource Management and Formula Funding of Schools on the Efficiency and Equity of Resource Allocation in Sri Lanka*. (Unpublished), London: Institute of Education, University of London.

structure consists of two interlinked layers: national and provincial (see **Figures 4 and 5**).

National level

The Ministry of Education is the national policymaking body with regard to general education in Sri Lanka. It is vested with legislative and executive authority to: formulate and ensure the implementation of national policy on education, and management of schools and education institutes; manage national schools, and control the functioning of private schools; planning, monitoring, evaluation, supervision, and research and development; donor coordination; review and coordinate policy implementation; set standards; provide incentives and subsidies; and exercise other powers relevant to education.

The National Education Commission, established in 1991, is mandated to identify and recommend policies in education for declaration by the President of Sri Lanka.

The National Institute of Education was established in 1986 and is responsible for providing leadership for the development of general education and curricula for general and teacher education, providing professional development, and spearheading change through research and innovation.

The Department of Examinations, established in 1968, conducts all types of national evaluations, and public and other local and foreign examinations.

The Department of Education Publications (DEP) was established as a separate budgetary entity for the production and publication of textbooks for use in schools. The main objective of the DEP when it was first established was the production and distribution of textbooks in national languages. This became necessary due to the

need for textbooks in Sinhala and Tamil following the change in the medium of instructions in the school system from English. Since 1980, DEP has published free textbooks for grades 1–11 in all schools for all subjects.

Provincial level

Provincial councils play an important role within the school system. The provincial education structure consists of: Provincial Ministry of Education (PME), Provincial Department of Education (PDoE), Zonal Education Offices (ZEO), Divisional Education Offices (DEO), and National and Provincial Schools. PEAs develop education plans and budgets, and deploy education administrators, principals, and teachers at the provincial level.

Provincial Ministry of Education is responsible for the implementation of national education policies, and formulation and implementation of provincial education policies in line with the national education policies.

Provincial Department of Education is responsible for the management and administration of all education programs of the province while being accountable and responsible for both the MoE and the PME.

Zonal Education Offices are responsible for administrative work of the schools and teachers in the zone, and the quality improvement of education. At present, 93 ZEOs are functioning island-wide.

Divisional Education Offices are responsible of supervision, monitoring of schools, and divisional administration. At present, 295 DEOs are functioning in the system.

National and provincial schools

Although the national schools are directly administered and financed by the MoE, many administrative powers are

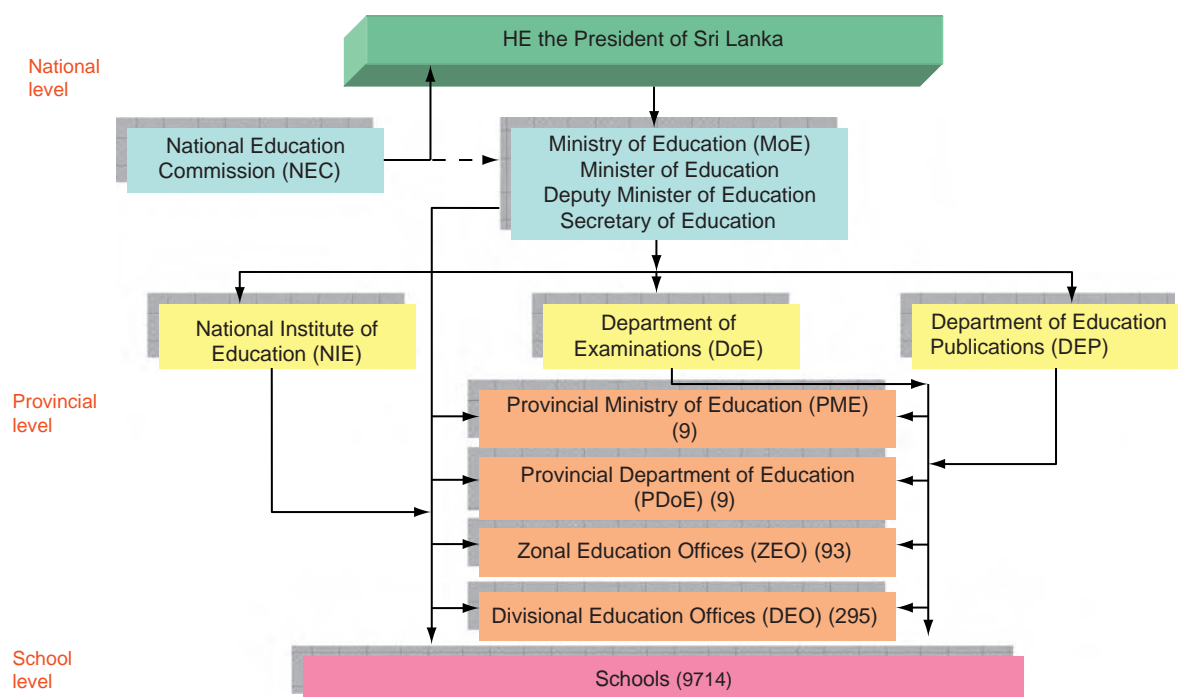


Figure 4 Administrative structure of education system: Sri Lanka.

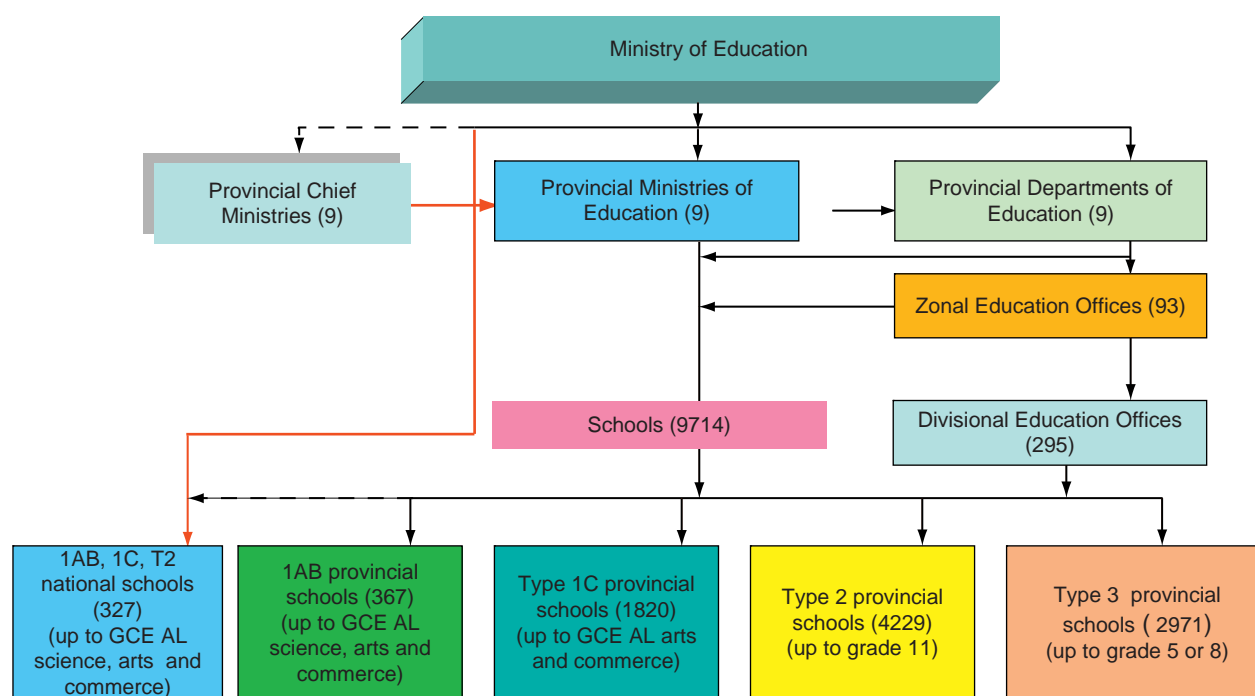


Figure 5 Implementation structure of education system: Sri Lanka.

being devolved to the provincial level. School principals act as heads of management and accounting officers for the schools and are accountable to the government through PEAs.

Education Financing

There are two main sources of expenditure on education: first, government expenditure and second, private/household

expenditure. Public expenditure on education for the period 2001–08 is shown in **Table 4**.

As **Table 4** shows, the government's expenditure on education has increased annually, although as a percentage of GDP, it has varied. In spite of its impressive achievements, by international standards Sri Lanka has spent small amounts on education. World Bank (1996) estimates the international average as 5% of GDP and 20% of total public expenditure. Asian averages range from 4% to 8% of GDP and 11–18% of total expenditure. Though the educational expenditure in Sri Lanka has increased with the expansion of the school system, it has remained between 6% and 10% of overall public spending since 1978. This may however, reflect both positive and negative features in the provision of education. First, it may reflect low unit costs and second, it may reflect less than desirable quality in the sector. Consequently, for further development of education financing, the Presidential Task Force on General Education in 1997 recommended that while current expenditure on education was around 2.5% of GDP, the allocation should be increased within the next few years to above 4.5% of GDP. PEAs have the main responsibility for providing education for children in schools, although there are many national schools located within the provinces which receive funding directly from the central government. The education sector has a complicated governance framework in relation to financial powers and authorities. The allocation flow for education in Sri Lanka is shown in **Figure 2**.

Assessing expenditure on primary and secondary education separately has been difficult due to the way that the school system has been organized. In 2000, recurrent public education expenditure was approximately 38% on primary education, 53% on secondary education and 9% on higher education.

Table 4 Expenditure on education, 2001–08

Year	Education expenditure SLRs m	Education exp. as a percentage of total government exp.	Education exp. as a percentage of GDP
2001	28 286.00	7.30	2.00
2002	33 575.38	9.20	2.40
2003	35 311.23	9.40	2.20
2004	38 680.05	8.90	2.10
2005	53 293.91	8.80	2.20
2006	60 876.58	8.00	2.10
2007	74 526.22	5.51	2.09
(Est.)			
2008	112 170.22	7.40	2.69
(Est.)			

From Central Bank of Sri Lanka (various years). *Central Bank of Sri Lanka Annual Report – Various Years*. Colombo: Central Bank Printing Press and Ministry of Finance and Planning (various years). *Budget Estimates – Various years*. Colombo: Department of Government Printing.

Household Expenditure on Education

The second-largest financial contribution to education is made by individuals, but these expenditures have not been computed at school, zonal, provincial, or national levels. These expenditures fall into two types. The first is expenditure contributed directly to school budgets. The second is expenditure incurred by the students and parents that does not necessarily go to the school budget. Sri Lanka has achieved high levels of school participation with low costs compared to other Asian countries. This depends on the household contributions. In Sri Lanka, both poor and rich parents aspire to obtain good quality education for their children. Their contributions have not been computed, and little attention has been paid to them when decisions are taken on school financing or education policies in general.

Output-Oriented Education Budgetary Approach

Currently, the resource mechanism is moving from an input-oriented approach toward an output/outcome-oriented approach; the budgeting approach followed medium-term budgetary framework (MTBF) for resourcing schools and other educational agencies. Within the MTBF, performance-based budgeting (PBB) and formula-based funding mechanisms have been implemented, ensuring equity and efficiency of resource allocation.

School Revenues

The entire financial requirements of schools are born by the government. Schools have been given permission to collect a very small amount as school facilities fees and for the school development society. Most of the schools have their own alumni associations (AA). Although statistics are not available, the total amount of these funds is not negligible. School revenue flow is given in **Figure 3**.

New trends: Education Sector Development Framework and Program (ESDFP)

Sri Lanka now faces the challenges of expanding equitable access to basic and secondary education and establishing a high-quality education system. The country has passed regulations on compulsory education for all children aged 6–14 years, but participation in basic education is only 81%, and needs to be increased.

The achievement tests show substantial shortfalls in languages and mathematics. Therefore, the government set priority to improve learning achievement levels and

skills, to revise and update curriculum, develop teacher development framework (TDF) and modernize testing mechanisms. Issues remain in relation to equitable and efficient distribution of resources, service delivery, governance, and capacity building.

To face these challenges, the government developed ESDFP as a way forward. Initially, ESDFP was developed for the period of 2006–10 within the multiyear planning and is organized following four themes and result-based monitoring (see Table 5).

Theme 1: Increase Equitable Access to Basic and Secondary Education

The objectives of this theme are to ensure that all students complete the compulsory basic education cycle, and have access to secondary education (grades 10–13). Sri Lanka has performed well in attracting and retaining students in the primary cycle (grades 1–5), with a net enrolment rate of 96% and a net completion rate of 95%. With better implementation, Sri Lanka will be able to reduce the number of out-of-school children.

Theme 2: Improve the Quality of Basic and Secondary Education

The objectives are to improve the quality of basic and secondary education by supporting initiatives to improve learning outcomes and orient the education system to the world of work. Current policy objectives are to promote the development of generic skills; to reform the curriculum; to improve the content, quality, and delivery of textbooks and examinations; and to strengthen professional development.

Theme 3: Enhance the Economic Efficiency and Equity of Resource Allocation

Investment in education in the past has been based on a single year budgeting framework, with resources allocated to various expenditure heads on a historical basis. Several policy initiatives have been introduced to improve the efficiency and equity of resource allocation. The government has agreed to meet financial performance targets in actual spending from 2006 to 2010, and to allocate additional funds for accelerated learning campaigns in disadvantaged provinces/regions.

Theme 4: Strengthening Education Governance and Service Delivery

This theme is implemented under the three key outcomes set out in Table 5.

Result-Based Monitoring and Evaluation

ESDFP has a strong focus on the monitoring and evaluation of agreed results/outcomes. Outcome indicators and targets for 2006–10 have been developed for each outcome indicator. It is noteworthy that all external and internal assistance can be accommodated within this policy framework.

Tertiary Education

The tertiary education system is a three-tier system, which consists of: public universities, public advanced technical institutions, and private postsecondary institutions.

Public University Education

Beginning with a single university in 1942, the university system has now grown to 14 conventional universities and one open university operating under the primary legislative enactment, the Universities Act. The University Grants Commission (UGC) is responsible for planning and coordination of university education, allocation of public funds to HEIs, maintenance of academic standards, and regulation of admission to the HEIs. Present student enrolment of conventional universities is about 40 000 students and the OUSL has an enrolment of 20 000 students. This represents a participation rate in university education of less than 3% of the age cohort.

Postgraduate and Higher Education Institutions

Six postgraduate institutes and seven state-owned and 53 private HIEs are functioning. Up to 50 000 students are enrolled in these HIEs. They spend over SLRs 5.00 billion per annum.

Technical and Vocational Education

The technical education and vocational training (TEVT) sector is currently made up of an extensive system of public, private, and NGO sector training providers. Public sector TEVT providers used to function under different ministries up to the mid-1990s.

A considerable number of Sri Lankan students study abroad. The annual outflow of students is approximately 1000 and annual foreign exchange outflow is SLRs 2 billion.

Conclusion

The historical initiatives show that planned interventions are not new to the education system. However, some interventions had only partial success. The Sri Lankan

Table 5 Main education development programs and outcome indicators (ESDFP)

<i>Theme</i>	<i>Key operation (KO)</i>	<i>Project/key activities</i>	<i>Outcome indicator</i>
Increasing equitable access to basic and secondary education.	KO-1: Provide equal educational opportunities ensuring free education policy.	-Mid-day meal program. -Free textbooks and school uniforms. -Public transport subsidies. -Grade 5 scholarship program.	Number and percentage of children aged 6–14 years enrolled in formal schooling, special education, and NFE programs.
	KO-2: Expanding the network of secondary schools to relax constraints on geographically equitable access to good quality secondary education.	-Isuru colleges. -Model primary school development project. -Project on development of selected primary and secondary schools in urban areas.	
	KO-3: Development of special education programs for children with SEN.	-Special education programs. -Enrol out-of-school children.	
	KO4: Development of NFE programs for out-of-school adolescents.	-NFE programs. -Implementation of compulsory education act	
Improving quality of basic and secondary education.	KO-5: Curriculum reforms at primary and secondary levels and facilitate curriculum implementation.	-Curriculum reforms and implementation. -Provision of essential and HOCA, and inputs. -Provision of financial allocation for HOP.	Learning competencies over the basic and secondary education cycles clearly specified and communicated to schools.
	KO-6: Establishment of a TDF, which would include school-based/on-site teacher development systems.	-Acquisition of basic and HOLS. -School-based/on-site TDF.	
	KO-7: Examination reform to promote acquisition of higher-order transferable skills.	-Capacity building of teacher educators.	Schools staffed with trained and certified teachers, and supported by a high-quality teacher development system.
		-Establish item bank.	
		-Develop examination guidelines.	
	KO-8: Production and timely delivery of textbooks.	-Research and development. -Implement HRD programs.	Examination item bank, guidelines specified and communicated to schools and stakeholders.
		-Improve quality and content of textbooks.	
		-Develop timely delivery mechanism of textbooks.	
	KO-9: Promoting values, ethics, civic consciousness, and social cohesion in schools.	-Maintain and publish standards of textbooks.	Quantity, timeliness of delivery and quality of textbooks available to children.
		-Peace education and social harmony.	
	KO-10: Forging synergy between curriculum reforms, school-based TDF, and examination reform.	-Establish mechanism to improve forging synergy between NIE, Teacher Education Section, DoE, and DEP.	

Continued

Table 5 Continued

<i>Theme</i>	<i>Key operation (KO)</i>	<i>Project/key activities</i>	<i>Outcome indicator</i>
Enhancing economic efficiency and equity of resource allocation.	KO-11: Development of an over-arching education sector development plan.	-Develop/revise medium-term education sector development plans.	
	KO-12: Establishment of MTBF.	-Strengthen national- and provincial-level planning capacities and EMIS. -Establishment of MTBF.	Recurrent education budget prioritized to support HOP and inputs. Capital education budget prioritized to support the development of HOLS and HOCA. Resource allocated for replacement of the education capital stocks, on the basis of funding formulae.
Strengthening education governance and service delivery.	KO-13: Establishment of PEQETS.	-Establishment of PEQETS.	
	KO-14: Program for school improvements (PSI).	-Awareness programs and principal training programs.	PSI pilot tested and evaluated, and, if successful, established countrywide after incorporating lesson from the pilot program.
	KO-15: Human resource development strategy for the education system. KO-16: Organizational analysis and capacity building.	-Develop plan and the policy on HRD. -Organizational review. -Revise organizational structure.	A long-term HRD plan produced and which forms the basis for staff development in the MoE, and among the PEAs. Organizational capacity review of the MoE, provincial, zonal and divisional education authorities conducted and organizational activities completed satisfactorily.
Results-based monitoring and evaluation.	KO-17: Results monitoring and evaluation.	-Implement result-based monitoring mechanism. -Policy analysis, research, surveys, and studies.	

From Ministry of Education (2006b). *Education Sector Development Framework and Programme (2006–2010): Policies and Expected Results*. Battaramulla: Ministry of Education. and Ministry of Education (2007). *Education Sector Development Framework and Programme*. Atul-Kotte: Graphics Systems.

education system has achieved success in providing wide-spread access to formal education enabling the country to attain a high level of human development relative to a low-income economy. Up to the early 1990s, Sri Lanka enjoyed the highest basic social development outcomes relative to per capita income among virtually all developing countries. This achievement was the result of strategic policy decisions over several successive governments to invest resources in education, health, and other social services. Sri Lankan policymakers who designed the basic framework of the education system in the 1930s and 1940s were far ahead of their time in perceiving human capital as a promising investment.

Currently, a number of policy initiatives have taken place to improve learning outcomes of students in line with the education quality development principles. It should be noted that the focus of present ESDFP is outcome oriented, and the reform program follows medium/long-term planning and budgeting horizons.

Further Reading

- Balasooriya, B. M. J. (2004). *An Evaluation of the Impact of School-Based Resource Management and Formula Funding of Schools on the Efficiency and Equity of Resource Allocation in Sri Lanka*. (unpublished), London: Institute of Education, University of London.
- Central Bank of Sri Lanka (2007). *Sri Lanka Socio-Economic Data 2007*, vol. XXX. Colombo: Central Bank Printing Press.
- Central Bank of Sri Lanka (various years). *Central Bank of Sri Lanka Annual Report – Various Years*. Colombo: Central Bank Printing Press.
- Department of Examinations (2003). *Examinations for Grade 5 pupils for the Selections of Four Popular Schools and Bursaries: Rules and Regulations*. Battaramulla: Department of Examinations Press.
- Government of Sri Lanka (1978). *The Constitution of the Democratic Socialist Republic of Sri Lanka*. Colombo: Department of Government Printing.
- Government of Sri Lanka (1979). *Pririvas Education Act, Number 64 of 1979*. Colombo: Department of Government Printing.
- Government of Sri Lanka (1985). *National Institute of Education Act, No. 28 of 1985*. Colombo: Department of Government Printing.
- Government of Sri Lanka (1997). *Compulsory Education Act No. 1003/5 and Dated November 25, 1997*. Colombo: Department of Government Printing.
- Lewin, K. M. and Mallawarachchi, S. (2001). Financing secondary schools in Sri Lanka: high participation and low costs. In Lewin, K. M. and Caillods, F. (eds.) *Financing Secondary Education in Developing Countries*. Paris: IIEP/UNESCO.
- Ministry of Education (1981). (in collaboration with the Ministry of Higher Education and the Ministry of Youth Affairs and Employment). *Education Proposals for Reforms: White paper for Education*. Colombo: State Printing Corporation.
- Ministry of Education (2005). *School Census–2005*. Battaramulla: Ministry of Education.
- Ministry of Education (2006). *Education Sector Development Framework and Programme (2006–2010): Policies and Expected Results*. Battaramulla: Ministry of Education.
- Ministry of Education (2006a). *School Census–2006*. Battaramulla: Ministry of Education.
- Ministry of Education (2006b). *Education Sector Development Framework and Programme (2006–2010): Policies and Expected Results*. Battaramulla: Ministry of Education.
- Ministry of Education (2007). *Education Sector Development Framework and Programme*. Atul-Kotte: Graphics Systems.
- Ministry of Finance and Planning (2000). *Budget Estimates – 2000*. Colombo: Department of Government Printing.
- Ministry of Finance and Planning (2001). *Budget Estimates – 2001*. Colombo: Department of Government Printing.
- Ministry of Finance and Planning (various years). *Budget Estimates – Various years*. Colombo: Department of Government Printing.
- National Education Commission (2003). *Envisioning Education for Human Development: Proposals for a National Policy Framework on General Education in Sri Lanka*. Nugegoda: Piyasiri Printing Systems.
- National Education Research and Evaluation Centre (2004). *National Assessment of Achievement of Grade Four Pupils in Sri Lanka: Achievement after Four Years of Schooling – National Report*. Colombo: NREC.
- National Institute of Education (1989a). *Historical Foundation of Education – Part 1*. The module of Postgraduate Certificate Course, Sinhala Version: 1.0 edn. Institute of Teacher Education, National Institute of Education, Maharagama. Colombo: Gunaratne Offset Limited.
- National Institute of Education (1989). *Kannangara Philosophy and Education Development*. Sinhala: 2nd edn. Published by the Research Unit of the National Institute of Education. Colombo: State Printing Cooperation.
- The Presidential Task Force on General Education (1997). *General Education Reforms*. Colombo: The Pacific Press Pvt. Ltd.
- UNESCO (2000). *Sri Lanka Country Document*. <http://www.ibe.unesco.org/International/Databank> (accessed May, 2009).
- United Nations' Organization (1948). *Universal Declaration of Human Rights, Article No. 26*. New York: UNO.
- Wehella, M. M. (2001). *Extending Educational Opportunities: A Study of the Causes and Effects of the Implementation of School-restructuring Programme in Sri Lanka*. (Unpublished), Institute of Education, University of London.
- World Bank (1996). *Sri Lanka Teacher Education and Teacher Deployment Project Implementation Papers*. Washington, DC: The World Bank.
- World Bank (2005). *Treasures of the Education System in Sri Lanka: Restoring Performance, Expanding Opportunities and Enhancing Prospects*, The World Bank, Human Development Unit, South Asia Region, The World Bank Colombo Office. Colombo: Ari Investment Limited.

Relevant Websites

- <http://www.centralbanklanka.org> – Central Bank of Sri Lanka.
- <http://www.statistics.gov.lk> – Department of Census and Statistics.
- <http://www.edupub.gov.lk> – Department of Education Publications.
- <http://www.doenets.lk> – Department of Examinations.
- <http://www.gov.lk> – Government of Sri Lanka.
- <http://www.moe.gov.lk> – Ministry of Education.
- <http://www.mohe.gov.lk> – Ministry of Higher Education.
- <http://www.most.gov.lk> – Ministry of Science and Technology.
- <http://www.nipunatha.gov.lk> – Ministry of Vocational Education.
- <http://www.nie.lk> – National Institute of Education.
- <http://www.ugc.ac.lk> – University Grants Commission.

Suriname

H R Ori, Ministry of Education and Community Development, Paramaribo, Republic of Suriname

© 2010 Elsevier Ltd. All rights reserved.

General Background

Suriname is situated in the north-east coast of South America, just north of the Amazon Delta. Its borders are the Atlantic Ocean in the north, Guyana in the west, French Guyana in the east, and Brazil on the south. The size of Suriname is approximately 163,820 km² and almost 90% of Suriname's surface is covered by tropical rainforest. The capital of Suriname is Paramaribo. This is the largest city in Suriname and is also the major import/export port as well as the major economic center. It is located on the west bank of the Suriname River and has about 250 000 inhabitants. The city is famous for its historical wooden architecture. The historical inner city has been placed on the United Nations Educational Scientific and Cultural Organization (UNESCO) World Heritage List as of July 2002.

The surface of Suriname is mainly covered by dense tropical rainforests, and you could find mainly lowland and marshes in the coastal area. The average temperature is 30°C and the humidity is approximately 90%. The annual average rainfall is above 2000 mm, and the south-west district has an average rainfall of 3000 mm annually.

The multi-ethnic population with their heartwarming hospitality is a token of the diversity of cultural traditions that exist side by side in harmony. The population is formed by several ethnic groups including Amerindians, Creoles, Maroons, East Indians, Javanese, Chinese, etc. Dutch is the official language as Suriname was a former colony of the Netherlands. The common spoken language is Sranan Tongo (*lingua franca*). Most of the people in Suriname can speak English as well.

Suriname has many natural resources that have to be developed and an enormous, particularly unexploited rainforest with many attractions for ecotourism. The world is nowadays discovering Suriname as an ecotourism destination, as well as a cruise-tourism destination. Suriname has many natural resources such as aluminum, gold, crude oil, etc. Most of the gold-mining projects in Suriname are outsourced to Canadian companies. Recently, more and more Chinese companies are investing in the wood industry of Suriname as well. The major countries to which natural resources are exported to are USA and the Netherlands. Approximately 18% of the population of Suriname is in the farming industry. Farmlands are located in the coastal area and in the district of Nickerie (one of the 10 districts in the country). The major farm products include rice, coconut, banana, tropical fruits, flowers, etc. Apart from this, there is also a huge market for seafood resources from

Suriname, including shark fins, fish maws, and shrimps. Most of them are exported to Hong Kong and China where these are very popular. One could also discover a lot of species that are only found in the Amazon River region. There are also fish and shrimp farms.

Goals of the Education System

In the twenty-first century, there has been a broad agreement about the basic goals of the education system and the responsibilities of government. These are generally set out in legislation and are supplemented by more detailed policy documents. In accordance with the Constitution and the Ministry of Education and Community Development (MOECD) policy document, the following mission has been formulated:

to create and maintain conditions, facilities and means of formal and informal education and community development for the benefit of every citizen and resident of Suriname for the acquisition of knowledge, skills, standards and values aimed at participating efficiently in our democratic society and the modern world, and participating freely in the multi-cultural life with a strongly developed environment awareness and participating optimally in socio-economic life of our country (Policy document 2000–05, MOECD).

The mission is based on the principles of quality, equality and equity, continuity, diversity, and also focuses on actors, in order to accomplish its objectives. The objectives at the macro-level have been set out as follows:

- a more effective and efficient educational system;
- a just educational system offering equal opportunities to all;
- an internationally competitive working population; and
- an educational system offering services that meet the quality standards set.

The government aims to create a seamless, high-quality, world-class education system in Suriname which will contribute to a highly skilled, talented, and knowledgeable workforce and the development of a people characterized by excellence in innovation. It is within this context that the MOECD is working diligently, along with other key stakeholders, in restructuring and developing the educational system in Suriname.

The national educational plan builds on much prior work that has been carried out in Suriname as well as in

different international for a in which Suriname has been involved. So, in addition to the national education plan benefiting from the Suriname documents referred to above, it has also benefited from international educational strategies such as those put forward in the Educational for All (EFA) campaign and the Caribbean Task Force which emerged from the Caribbean Development Strategy and is planned till the year 2020. Five strategies are outlined in the national education plan:

1. reducing the knowledge gap between and within countries;
2. turning the school and the classroom into the focus of the educational system;
3. undoing inequalities in the education system;
4. improving educational finance and management; and
5. strengthening regional cooperation.

The ministry has developed this national educational plan for the period 2005–10 which identifies several strategic priorities: focus on schools, change/reform the ministry, and involve the community. This plan establishes the framework within which educational goal setting, policy development, strategic planning, research, program development and implementation, establishment and management of quality standards, and monitoring and evaluation are conducted by the ministry's central administration.

The Aims and Purposes of Education at Each Level

The MOECD is in the stage of transforming the basic education system for 4–15-year-old children in the basic education improvement project (BEIP). The goal of the project is to contribute to the development of human capital in Suriname and its objective is the improvement of the quality and internal efficiency of basic education by means of updating the educational contents and processes, providing inputs to schools, and supporting institutional reforms to strengthen the MOECD. Therefore early childhood development (ECD) is combined with preschooling and primary junior secondary schooling into a new 11-year basic education cycle model. The underlying goals of early childhood care and education are to provide young children with opportunities to succeed in all areas of development utilizing the constructivist learning theory. Primary education serves to ensure the transmission of culture and values, deepen cognitive and social development, and prepare students with literacy and numeracy skills to successfully pursue a quality secondary-level program.

The core objective is to improve the quality and internal efficiency of basic education by introducing (1) updated curricula and educational processes, (2) the provision of inputs for schools, and (3) the establishment of several institutional amendments aimed at strengthening the

MOECD and hence build management capacity at the school level in Suriname. The BEIP includes four key components, with some components comprising several major subcomponents. They can be summarized as follows:

- support the establishment of a revised basic educational cycle and quality improvement;
- renovation of schools and MOECD infrastructure;
- strengthening of management capacities at school level; and
- modernization and strengthening of MOECD and the school system.

Structure and Operation of the Education System

General

The MOECD provides the general-education system for children from age 4 years and older. Children are legally bound to participate in education up to the age of 15 years. The Suriname's educational system includes (pre)primary, junior secondary, senior secondary, and tertiary schooling. Schooling for young children (ages 4 and 5) is integrated with primary education, and is optimal until age 6. Roughly half (48%) of all schools in the country are operated by religious bodies that are mostly financed and regulated by the MOECD; also, they follow the same national curriculum as those in the public schools (51%) and other private education (1%). The education system uses tracking based on pupil performance, which results in a primary school-leaver's certificate based on the sixth-grade final exam. For students who pass this exam, their entry into one of the six streams of junior secondary education is granted, based on the results.

Suriname's (pre)primary schools have relatively high enrolment rates of 90%, with approximately 90 000 children enrolled in 320 (pre)primary schools. Approximately 35% of the schools are located in the capital city Paramaribo and 55% are situated in surrounding (semirural) districts and 10% are found in the interior.

The ministry provides four levels of education: pre-primary, primary, secondary, and tertiary levels of education. All education is included within these four levels.

Preprimary and Primary Education

Nursery education is intended to offer preschool children of ages 4 and 5, basic education and teach them basic knowledge and basic skills. This is necessary with a view to their further development and to prepare them for primary education (GLO). Except for a few private schools, all nursery schools are accommodated at GLO-schools. Although most preschoolers in the age group 4–5 years go to school, compulsory attendance does not apply to this

group and there is no law on nursery education. This will be changed in the new law on the 11-year basis education law. The compulsory education age will be changed from 7–12 years to 4–15 years. Suriname is now in the stage of implementing the integral ECD approach (age 0–8) with special attention being paid to early childhood stimulation.

Children must have reached the age of 6 prior to 1 January of each calendar year in order to be admitted to primary education. Primary education is characterized by a very high degree of inefficiency. The percentage of repeaters is approximately 25% and also the drop-out percentage is rather high (8%). Only 50% of the pupils who start school, leave with a certificate.

The causes of this degree of inefficiency may be attributed to, among other things, the shortage of educational tools, obsolete curricula, deficient refresher and supplementary training of teachers, and high percentage of teacher absenteeism. At the end of the sixth year, pupils take the national primary test. On the basis of the test results and the school marks for the sixth year, pupils are admitted to one of the various types of junior secondary education (VOJ). At the moment, no diagnostic test or formative tests are taken prior to enrolling in secondary education. The interest and the capacity of the pupils should be and will be taken into account in the new 11-year basic cycle education program.

Education in the Interior

Education in the interior demands special attention. The development of the educational sector in the interior has always been hindered by the fact that the village communities are scattered over large geographical distances and moreover, are often difficult to reach and are accessible only by dugout or airplane. The percentage of repeaters and drop-outs in the interior is about 50%. Due to their isolated conditions, education in the interior is confronted with some specific problems: shortage of qualified teachers, expensive transport for teachers, deficient physical infrastructure, shortage of educational tools and adjusted curricula, insufficient guidance and inspection from city, and language problems. Many qualified teachers do not wish to work in the interior because of the geographic isolation, the lack of possibilities for further study, and the incidence of malaria disease.

Special Education

Special education targets children with a handicap and special educational needs. This type of education is provided in various schools and institutions at primary-education levels. This type of education is strongly differentiated according to the type of handicap and is intended for children with: visual, auditory, communication, physical, and/or mental

handicap; serious behavioral disturbance or psychiatric problems; and learning and behavior difficulties.

At primary level (GLO), dependent on the nature and intensity of the handicap, this type of education is provided in/at several school: classes for special education at primary schools and schools for children with learning difficulties, multiple disabilities, motor disabilities, visual and auditory disability.

Over 1800 pupils are involved in some form of special education. Almost all schools and institutes for special education fall under the management of private foundations and institutions, while the teachers are subsidized by the government. A few institutes provide secondary education.

Special education is characterized by numerous problems. Some of these are: the absence of adapting the international development (Salamanca Declaration, 1994), the lack of implementing new instruments (diagnostic systems, assessment aid, research centers), insufficiently qualified teachers and assistants, and insufficient opportunities to move back to regular education.

MOECD needs to develop a guiding philosophy on inclusive education which will address the provision of support and services to all learners, regardless of their physical, intellectual, social, or emotional conditions, and regardless of the ethnic group, religion, socioeconomic status, or gender. All children have a right to become all they are capable of becoming. The government successfully achieved universal primary-education capacity since the mid-1970s. The ministry is in the starting process of developing a policy on inclusive education which will address the provision of support and services to all learners regardless of their physical, intellectual, social or emotional conditions.

Junior Secondary Education

There are two levels of secondary education: junior secondary (VOJ) and senior secondary education (VOS). The junior secondary education consists of several streams: junior secondary general education (MULO, 4 years), junior vocation-oriented education (LBGO, 4 years), junior technical education (LTO, 4 years), junior domestic science and technical education (LHNO, 3–4 years), elementary technical education (ETS, 3 years), elementary vocational education (EBO, 3 years), and advanced special education (VSO, variable duration) (See **Figures 1 and 2**).

Pupils (12 years) are admitted to the different types of schools mainly on the basis of their primary-education test results, as well as a certain priority of performance and prestige. In MULO especially, there is a greater demand in view of the possibilities it gives to move up to various forms of senior secondary education. Of the pupils who participated in the primary education test in the past 10 years, 46.9% was referred to MULO, 30.3% to LTO, 14.4% to

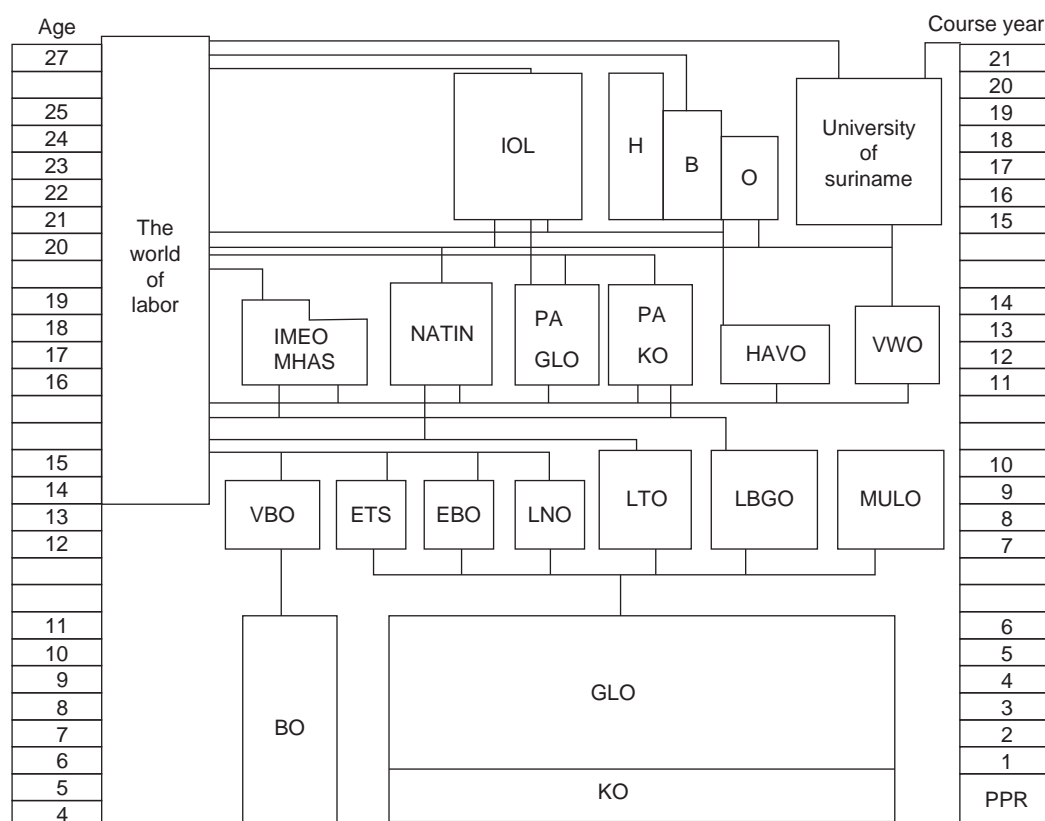


Figure 1 The current structure of the Suriname educational system. BO/SO, special education; EBO, simple vocational education; ETO, elementary vocational education; GLO, primary education; HAVO, senior general secondary education; HBO, higher professional education; IMEAO, Institute for Intermediate Business Education; IOL, Institute for Advanced Teacher Training; KO, nursery education; LBGO, junior vocation-oriented education; LNO, junior domestic science and technical education; LTO, junior technical education; MHAS, secondary evening commercial school; MULO, junior secondary general education; NATIN, Institute for Middle-Level Natural Resources and Engineering Studies; PA, pedagogical academy; VBO/VSO, advanced special education; VWO, preuniversity education.

LBGO, and 7.4% to the remaining school types. Over 30% had to repeat the sixth grade.

This level of education is confronted with the following problems: unqualified teachers, obsolete curricula, and a shortage of educational tools, high percentages of repeaters and drop-outs, low percentages of repeaters, and few possibilities of moving up horizontally.

Senior Secondary Education

This level comprises the following school types: university preparatory education (VWO, 3 years), senior general secondary education (HAVO, 2 years, night HAVO, 3 years), institute for natural resources and engineering studies (NATIN, 4 years), school for intermediate business education (IMEAO, 3 years), teacher-training college (4 years), secondary commercial night school (MHS, 2–3 years) and secondary technical night school (AMTO, 4 years) (see Figures 1 and 2).

For admission to VWO, HAVO, and NATIN, one must have passed the relevant entrance examination.

Participation in the exam is possible from the third or fourth year of MULO. For admission to the teacher-training colleges (SPI, CPI) and IMEAO, MULO diploma is a requisite. This school type is also confronted with several problems: lack of classrooms and buildings, lack of qualified teachers, high percentage of repeaters and drop-outs, etc. There is an increase in the demand for secondary education, especially for NATIN and IMEAO.

Adult Education

Within the framework of their nonformal education and development activities, many nongovernmental organizations (NGOs) are engaged in providing adult education. The duration of the course is approximately 1 year and classes of two periods are given on 2 days a week. The size of the classes varies from 12 to 20 participants, while the lessons are provided at two levels: completely illiterate and for people who have had 2–3 years of primary education.



Figure 2 The future structure of the educational system. BA, courses at bachelor's level; SO, special education; HAVO, senior general secondary education; HBO, higher professional education; IMEAO, Institute for Intermediate Business Education; IOL, Institute for Advanced Teacher Training; LNO, junior domestic science and technical education; LTO, junior technical education; MA, courses at master's level; MBO, upper secondary vocational education; PPR, pre-primary education; TR CEN, training center; VSO, secondary special education; VWO, preuniversity education.

Second-Chance Education

In the context of second-chance education, the courses offered by the foundation for labor mobilization and development (SAO) may be considered as supplementary education, just as the programs offered by the foundation toward a new alternative (TANA).

Higher Education

Higher education comprises all types of courses at postsecondary level. Such courses are provided by: the Anton de Kom University of Suriname (ADEKUS), the institutions for higher professional education (HBO); the Institute for Advanced Teacher Training (IOL), the Polytechnic College (PTC), the Teacher Training Vocational Education (LOBO), the Youth Dental Care (JTV), The Academy for Higher Education in Arts and Culture (AHKCO), and the Central Training for Nurses and Related Professions (COVAB).

JTV and COVAB fall under the Ministry of Health, the other courses fall under the MINOV. Furthermore, there are a few private institutions, including ones that offer courses in tourism/hotel and catering sector, and a part-time night course in business economics provided by the Hogeschool Inholland from the Netherlands.

The ADEKUS comprises three faculties and the Institute of the Graduate School (IGSR) in the academic year 2008–09. The IGRS offers several masters of science courses and leads the PhD trajects. The faculty of social sciences has six disciplines, 14 specializations, and 2400 students. The medical faculty offers two disciplines and 425 students and the technological faculty has six disciplines, 15 specializations, and 650 students. The university also has the *Schakelinstituut* (transition course) where students who do not meet the regular admission requirements relating to previous education, are prepared for admission to university studies.

The university also comprises of nine research centers: INTEC, IMWO, CELOS, IDPM, MWI, UCC, the National Herbarium, the National Zoological Collection, CMO, IKR, and IIR. The ADEKUS replaced the old 5-year doctoral studies, which are now dismantled, by a BA and MA (bachelors and masters) structure with the 3-year bachelor of science studies followed by the 2-year masters studies.

Suriname, like other countries, is facing numerous challenges in higher education. Adequate management and financing, organization of education, quality assessment of education and research, and establishment of national and international partnerships are some examples of these challenges. With the Flemish universities in Belgium, a sustainable relation was established for a duration of 10 years for

strengthening education, research, and services. The focus is set on six projects in transforming the ADEKUS Suriname from a primarily BSc-oriented education university to an accredited MSc-oriented research and education university.

Higher education is of great importance to the development of the country and is continuously under internal and external pressure. Internal because of the strong demand for highly trained staff (manpower requirement) and the great supply of students (social demand) involving considerable training costs: External, because of the need to satisfy international and regional-quality requirements despite the small scale and limited means.

In higher education, several issues need attention: legislation concerning higher education, alternative financing models, increase in the number of students, incompetent teaching staff and education reform (pedagogical competency of teachers, teaching large groups, use of new didactic methods, increase use of information and communication technology (ICT), laboratory courses, retraining, in-service training, etc.).

Administrative and Supervisory Structure

Education Administration

Government schools in Suriname operate under direct responsibility of the relevant MOECD, while nongovernment schools operate under authorities. The MOECD is in the stage of strengthening the core responsibilities and supporting services and processes of education. The changes will occur at departmental and school levels both. Much more focus will be given to the following key functions: strategy development, management of financial and material resources, human resources management, management of educational tools and means of development, quality control, information, communication, and responsibility.

Educational Finance

Suriname has to spend much more effort and money to develop the education system. There is a shortage of highly qualified staff and deficiencies in institutional and incentive structures in the education sector. Schools lack autonomy and financial allocation mechanisms are completely unrelated to performance.

The government in Suriname is financing almost financing the education sector. Which means 5–6% of the gross domestic product (GDP) is spent on the educational sector, which implies that annually between 20% and 25% of the current government budget is spent on it. The education budget for the period 1996–2000 averaged US\$37 million on an annual basis with a downward trend. In the last few years, this budget showed an upward trend once again. In 2002, it was US\$56 million and in 2003,

(US\$63 million). Roughly 80% hereof is spent on primary and junior secondary education (VOJ), 7% on senior secondary (VOS), and 13% on tertiary education (IOL, ADEKUS). (See **Tables 1** and **2**.) Government funding for higher education comes mainly from the Suriname government. For the financial year 2009, total government education and training operating expenses are projected to be 5.3% of GDP.

Assessment of the Performance of Schools

In Suriname, the assessment of the performance of schools is a matter of growing concern. This has to do with the national efforts to raise the standards of individual schools, both objective external assessments and advice from professionals or peers, who know the school, are equally important. Performance indicators can provide important evidence of a school's achievement but there should be a clear attempt to identify information which is genuinely related to the quality of education a school offers its students. Evaluation systems simply making schools accountable, whether to parents, the community, or to others, are unlikely, on their own, to lead improvements in performance. But to require accountability of a school's stewardship of its resources, mission, and educational responsibilities is highly desirable in the interest of transparency and democracy in education. Care should be taken to build on the expertise and professionalism of teachers, and to provide programs of staff development which enable the staff of an institution to develop and extend their professional activities into new forms of autonomy.

Major Changes and Issues for the Coming Years

Managing Information, Monitoring, and Evaluation

A factor to be considered is whether Suriname has the necessary human, technical, and infrastructure resources for a successful implementation of education reforms. The use of educational change at the level suggested here and the technological level required, demands a

Table 1 Education expenditure as % GDP 1990–2008, human development index (HDI) 2008

	Suriname	Guyana	Barbados	Jamaica
1990	6.4%	3.4%	7.8%	-
1990–2000	3.4%	3.0%	7.3%	4.5%
2000–02	5–6%	8.4%	7.6%	6.1%
2002–08	5–6%	9.4%	9.6%	8.1%
HDI rank (2008)	89	107	30	98

Table 2 Education statistics (pre) primary, primary and VOJ-level schools, teachers, students, and teacher–student ratio

	<i>KO</i>	<i>GLO</i>	<i>MULO (VOJ)</i>	<i>LBGO (VOJ)</i>	<i>Other (VOJ)</i>
Schools	281	300	51	31	19
Teachers	717	3.230	1.088	535	787
Students	17.219	62.167	17.652	6.285	7.293
Student–teacher ratio	24	19			
Teacher sex	M (0), F(717)	M (224), F (2.083)			
Students sex	M (8.763) F (8.456)	M (32.044) F (30.123)			

culture to support it. This culture cannot be present only at the higher levels; it has to be all over the system. Not only the coordinators and the technical-support personnel, but also the data generators and even the users have to be part of this handling and distribution of information. The fact that there are computers and the need of information is felt, is not enough to guarantee a good system or an effective performance. The personnel using it need to believe in its value and to know how to use it to their own advantage. The information needed is to understand what is necessary to know. If the country is dealing with information to be the framework of policy, information to support the development of policy, it is necessary to develop a clear understanding of what is to be achieved. Education needs to have a comprehensive and shared vision of what is desired. Only in that framework will it be possible to define what is the information needed to show progress.

An education management information system (EMIS) has been developed for MOECD. It is a system for the collection, integration, processing, maintenance, and dissemination of data and information to support decision making, policy analysis and formulation, planning, monitoring, and management at all levels of the education system.

Many times the systems of education provide a significant amount of raw data. Lots of information is produced describing what is happening, sometimes in great detail. But if it is only that, a description, it might not provide much help to establish whether there is progress or not. Actually, some might think that it signifies a lot of effort, time, and resources just for the sake of getting some information that will only be valuable for researchers, who will, most of the time, use it only to be critical of the system. Other times, it goes beyond the raw data and the records to provide some analysis. Nevertheless, it is still felt that the results are not valuable enough.

These analyses are done out of context providing some comparisons or even looking for causes and explanations of what are individual phenomena sometimes taken out of context or in an individual fashion. In these cases, the information provided does not and cannot provide a comprehensive understanding of the situation.

What is lacking in most of these cases is a framework that can provide the basis to interpret the information that would lead to policy decisions. In simple terms, what makes the information relevant is the framework in which it is analyzed and understood. If the educational system lacks a vision and a sense of purpose, the information will never be understood to its fullest. Only when there is a coherent strategic view is it possible to have information that makes sense and that can guide policy. Once the goals are defined and the road clarified, then the indicators can be defined and the numbers can become meaningful.

But direction and a strategic view, although necessary are still not sufficient. Once the road has been defined, the next step is to travel it. Once the strategy has been determined, the next step is to have an achievable and sustainable plan.

How many times are the proposed goals and objectives reached? How many times are they really expected to be reached? Experience says that not many. On most occasions, because the action plans are unrealistic in the context in which they are to be implemented.

ICT Promotion in Education

The MOECD has to promote more ICT in education. A policy on ICT in education should be designed and disseminated which will provide the framework for the integration of ICT into the operation and management of the business of the ministry. Additional ICT initiatives include the establishment of fully equipped computer laboratories at primary schools; networking of secondary schools and maintenance of ICT infrastructure; provision of instructional ICT training to educators; and provision of tools for educators to develop instructional software and collaborative classroom systems.

Decentralization of Education

The MOECD will restructure and decentralize as key strategy for transforming the education system. This strategy aims to improve learning, teaching, and management in the education system, as well as organizational

performance and effectiveness. In addition, the establishment of the national system of local school boards to assist school will be taken into consideration at primary and secondary level.

Curricular Policies

The MOECD has to identify the essential learning outcomes which will help define standards of attainment for all primary and secondary school students (esthetic expression, citizenship, communication, personal development, problem solving and technological outcomes, and prepare students for further study or for entry into the world of work). The national curricula should also provide all students with the maximum opportunity to develop their potential and should therefore reflect and support the Suriname national ideals: be flexible and responsive to the developmental needs, life experiences, and unique abilities of each individual, provide learners with the resources to construct knowledge that is relevant to their needs and interest, and equip all learners with the knowledge and skills to attain a good quality of life.

Further Reading

IDB (2004). Basic Education Improvement Project: Loan contract no. 1521/OC-SU between the Republic of Suriname and the Inter-American Development Bank, IDB, Lima, March 2004.

Inter-American Development Bank (1998). Suriname Education Sector Study, February 1998.

Mercer, M. (2006). *Education Planning Consultancy, Final Report*. July 2006.

MOECD (2004). Education Sector Plan 2004 – 2008. Ministry of Education and Community Development: November 2004.

MOECD (2004). Policy document 2000 – 2005, Ministry of Education and Community Development: November 2004.

MOECD (2004). Suriname Education Plan, Ministry of Education and Community Development: November 2004.

MOECD (2005). Policy document 2005 – 2010, Ministry of Education and Community Development: November 2005.

Riddell, A. (2003). *Road Map toward a Sector Wide Approach in Education in Suriname: Report of UNICEF Consultancy*, March 2003.

Suriname Bureau of Statistics (2004). *Statistical Yearbook 2004*. ABS, Paramaribo, November 2005.

Relevant Websites

<http://www.emis-sr.org> – Education Management Information System Suriname.

<http://www.plos.sr.org> – Ministerie van Planning en Ontwikkelingssamenwerking.

Taiwan

C-M Lee, National Taiwan Normal University, Taiwan, Republic of China

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cram schools – Cram schools (called “bu-shi-ban” in Taiwan) are specialized private institutions that offer classes after normal school time or on weekends to meet customized goals, most commonly to pass the entrance examinations of high schools and colleges or universities. Cram schools offer instruction in areas such as mathematics, natural sciences, English, and certain kinds of talented competencies, which are necessary for entering a higher level of schooling.

General Background

Taiwan, which includes a series of islands with a total area of nearly 36 179 km² and a population of 22.8 million, is located in eastern Asia and separated from the People's Republic of China (PRC) by the Taiwan Strait. From the 1500s to the 1660s, Taiwan was subject to the influence of a series of Portuguese, Spanish, Dutch, and English powers. Han Chinese settlements were established during the Ching Dynasty from 1683 to 1895; Japanese colonization from 1895 to 1945 had a great impact on Taiwan's culture as well. When the Chinese Communist Party established the PRC on the Chinese mainland in 1949, the Republic of China (ROC) government, led by the Kuomintang (KMT; the Nationalist Party), had already relocated to Taiwan and proceeded to rule the country until 2000.

Ethnically, the people of Taiwan consist of two groups: an indigenous minority of Austronesian origin comprising 14 aboriginal tribes, and a Han majority including the Southern Min, Hakka, and Mainlanders. Ethnic Han came in two large waves from China, one from the 1650s onward and the other around 1949. Beginning in the early 1990s up until 2007, around 38 000 people from Southeast Asia and China married Taiwanese and immigrated to Taiwan. Over the past five decades, the KMT government made Mandarin the official language, completely suppressing other dialects. Today, Taiwanese (also known as Minnanese) as well as Hakka and aboriginal languages are promoted and taught in elementary and junior high schools.

The KMT government lifted martial law in 1987 and launched a series of political reforms, which strengthened

the process of democratization. In the 1996 general election, Taiwanese elected their own president directly for the first time. In the elections in 2000, five decades of KMT rule came to an end and the opposition party, the Democratic Progressive Party (DPP), took power in a peaceful transition. In the 2004 election, the DPP won again. But in 2008, following the presidential election, the KMT has regained power. Taiwan has a dynamic capitalist economy with increasing liberalization by the government in investment and foreign trade. As a result, Taiwan has now transformed into a knowledge-based and high-tech country with a per capita annual GNP reaching US \$16 471 in 2006.

Owing to Taiwan's rapid political and economic development in the last 20 years, the government has been able to take a more active role in facilitating human rights and providing children, the aged, the handicapped, and underprivileged groups with better care and protection. In addition, the government has implemented the National Health Insurance (NHI) program with the objective of offering universal medical care to all citizens and residents. Article 13 of the constitution of Taiwan guarantees freedom of religion as a right of all people. Polytheistic and syncretic, Taiwan's religious and spiritual environments are dominated by ancestor worship, Taoism, Buddhism, and Confucianism but have never excluded the addition and development of other indigenous and foreign religions.

Goals of the Educational System

Taiwanese society has transformed from being authoritarian to democratic step-by-step. Accordingly, its educational system is moving from uniformity to diversity, from authoritarian centralization to deregulation and pluralism. From 1949 to the 1970s, the ultimate goal of Taiwan's educational system was to fulfill and reinforce the ideals enumerated by Dr. Sun Yat-Sen, the founder of ROC, and stated in his Three Principles of the People: nationalism, democracy, and the people's livelihood. Education was seen as a means of indoctrinating political ideology and Chinese traditional culture into the populace. Since the 1990s, educational reform has become a part of several social movements and introduced to meet the needs of a greatly changing and emerging democratic society as well as to make manifest the intrinsic value of education.

The sustained goals of the educational system stated in writing are the five educational goals in the National Education Act, promulgated in 1979 and amended later in 2004. According to Article 1, "The purpose of education is to cultivate the citizens with a comprehensive and balanced development in moral, intellectual, physical, social and aesthetic dimensions." Moreover, the Educational Fundamental Act, promulgated in 1999 and amended in 2006, indicates in Article 2 multiple aims for educational goals:

The purposes of education are to cultivate modern citizens with sense of national identity and international perspectives. The methods are to foster the development of wholesome personality, democratic literacy, ideas of rule of law, and humanities virtues, patriotic education, native soil care and information capability. Strengthening people's physical health as well as their abilities to think, judge and create is critical, too. In addition, it is indispensable to enhance respect for basic human rights, protection of ecosystems and natural environment, and understandings of and concerns for different countries, ethnic groups, sexes, religions and cultures. (Educational Fundamental Act, 1999, 2006)

The Taiwan Ministry of Education (MOE) announced the educational goals of New Campus, Sustainable Campus, Healthy Campus, and Friendly Campus which were enforced in 1999. In 2004, the MOE declared four major tracks for current educational policies in order to enhance students' capabilities in all fields with creativities and global vision, namely, cultivating modern citizens, affirming Taiwanese identity, developing global views, and improving social concern.

Structure and Operation of the Educational System

The Taiwanese government has laid a strong emphasis on the expansion and improvement of education over the past four decades. Since 1968, compulsory education has been extended from 6 years to 9 years (ages 6–15). Furthermore, a wide range of other educational options is available to citizens of all ages. The entire educational process covers 2 years for preschool education from age 4, 6 years for elementary school, 3 years for junior high school, 3 years for senior secondary education, 4–7 years for college or university, 2–4 years for a graduate school program, and 2–7 years for a doctoral degree program. There are also technological/vocational educational programs, special education programs, supplementary education, and adult/continuing education programs. **Figure 1** shows the completion times of the entire process of the educational system. The key statistics of Taiwan's major education levels are in **Table 1**.

Preschool and Early Childhood Education

In Taiwan, there are two subsystems of preschool and early childhood education. The first subsystem is the kindergarten system, which is the responsibility of the MOE in the central government and Bureaus/Departments of Education (BOEs/DOEs) in local governments. According to the Kindergarten Act, amended in 2003, the main purpose of Taiwan's kindergartens are to improve health education, life education, and ethical education. In 2006, around 200 000 students, aged 4 to 6, were attending 3329 kindergartens in 1507 public and 1822 private schools. The second sub-system is childhood care and nursery school system, which is under the purview of the Ministry of Interior (MOI). Around 300 000 children, aged 2 to 6, were attending 5526 centers and schools in 1404 public and 4122 private schools. Children 5 years and above attending both kindergartens and nursery schools reached 93.9% of that age demographic.

The government has provided subsidies to support Taiwan's remote district, ethnic minority, and lower-income families of children in private childcare and preschool programs since 2004. A new policy of preschool and early childhood care has been implemented since 2005 and is expected to integrate the above two subsystems by the end of 2009. The new policy has two phases for children up to 6 years of age. For toddlers up to age 2 there will be babycare centers and for children aged 3–6 kindergartens. In addition, the government is planning to make education compulsory for children of ages 5 and above.

Primary and Secondary Education

Taiwan's primary and secondary education is mainly composed of elementary school (six years), junior high school (three years), and senior (vocational) high school (three years). Elementary and junior high schooling are compulsory for children from the ages of 6 to 15. In February 2007, the MOE announced a 12-year basic education plan at primary and secondary education level in the near future and has been revising the Senior High School Law for that purpose. In 2006, there were 2651 elementary schools (98.6% public and only 1.4% private) and 736 junior high schools (98% public and 2% private). The student–teacher ratio in elementary schools was 17.86 and that in junior high schools was 15.7. However, the average class size for elementary school was 29.3 and for junior high school 35.6 in 2005. The enrolment rate in elementary school was 99.02%, compared with that age demographic. The enrolment rate of junior high school students who graduated from elementary school was 99.65%. The school day of elementary school and junior high school is normally of about 7 class hours of tuition and comprises around 200 school days per year.

In the past decade, Taiwan's MOE has undertaken a revolutionary educational reform program that has had a

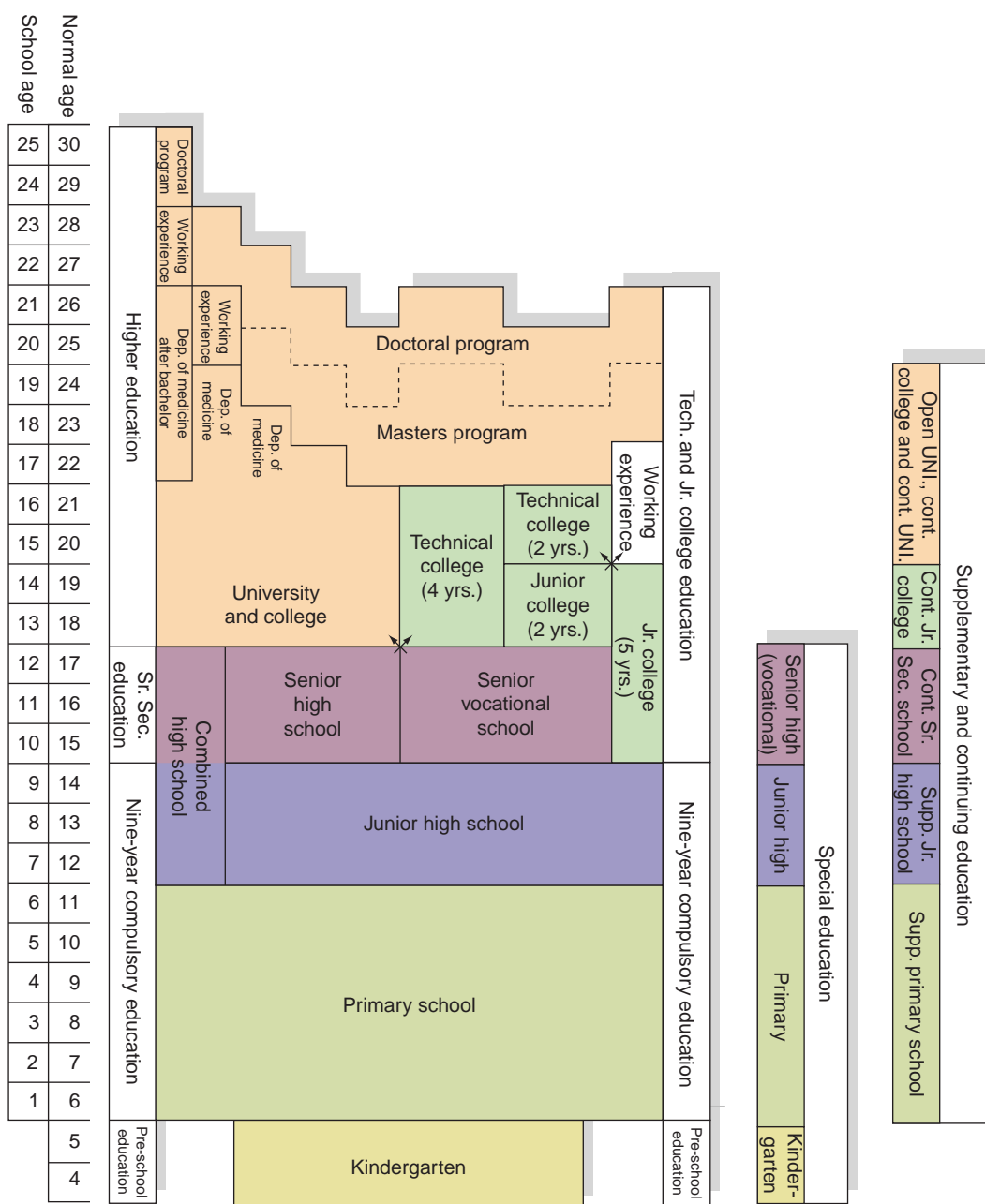


Figure 1 Taiwan's current educational system. Adapted from Taiwan Ministry of Education (2007). *2007 Education in Taiwan*. Taipei: Ministry of Education, p. 9.

huge impact on elementary and junior high schools. The MOE promulgated the Guidelines of Grade 1–9 Curriculum of Elementary and Junior High School Education in 1998 and implemented it fully by August 2004. The curriculum guideline encompasses seven major learning areas supplanting previous individual subjects and contains several important issues integrated into the overall curricula. The learning areas are language arts, health and physical education, social studies, arts and humanities, mathematics, science and technology, and integrative activities. The important issues are information

education, environmental education, gender education, human rights education, career development education, and home economics education. The guidelines emphasize ten goals as follows:

1. to enhance self-understanding and explore individual potential;
2. to develop creativity and the ability to appreciate beauty and bring one's own talents into full play;
3. to promote abilities related to career planning and lifelong learning;

Table 1 Key statistics in Taiwan's major educational levels 2007

<i>Educational levels</i>	<i>Schools</i>	<i>Students</i>	<i>Teachers</i>
Schools ^a	8254	5 287 226	273 971
public school	5870	3 635 373	208 485
private school	2384	1 651 853	65 486
Primary education	2651	1 798 436	100 692
Secondary education			
junior high school	736	952 642	49 749
senior high school	318	419 140	34 581
vocational high school	156	335 554	16 168
Higher education	163	1 313 993	50 388
Special education	24	6588	1736

^aThe total number of schools include preschool, primary, secondary, higher (tertiary), special education, and supplementary and continuing education. Cited from the Department of Statistics, Ministry of Education (Taiwan) http://www.edu.tw/EDU_WEB/EDU_MGT/STATISTICS/EDU7220001/ebooks/edusta/edusta.htm (in Chinese).

4. to cultivate knowledge and skills related to expression, communication, and sharing;
5. to learn how to respect others, care for the community, and facilitate teamwork;
6. to have further cultural learning and global understanding;
7. to strengthen knowledge and skills related to planning and organizing;
8. to acquire the ability to utilize technology and information;
9. to encourage the attitude of active learning and studying; and
10. to develop abilities related to independent thinking and problem-solving.

At the end of grade 9, students should attend the Multiple Admission Process to Senior High Schools and Vocational High Schools, which is based upon their scores of Basic Achievement Test and in conformity with the Grade 1–9 curriculum guidelines. In 2006, 145 308 students entered senior high schools, 119 533 students went to vocational high schools, and around 71 800 students attended other tracks of schools such as various supplementary schools or five-year junior colleges that cover part of the students' high school education and certain practical technical programs.

Regarding Taiwan's senior secondary education, there were around 318 senior high schools (56% public and 44% private) and 156 vocational high schools (59% public and 41% private) in 2006. The average number of students per full-time teacher in senior high schools was 19.3 and 18.38 in vocational high schools. The enrollment rate of senior (vocational) high school students who graduated from junior high school was 96.23%. The Curriculum Guidelines of Senior High Schools and the Curriculum Guidelines of Vocational High Schools (both promulgated by the MOE

on a trial basis in 2006, to be formalized in 2009 or 2010) are there to provide national standards of educational goals and the content for teaching all the courses of senior secondary education.

Technological and Vocational Education

The Technological and Vocational Education (TVE) system has several different levels in Taiwan: junior colleges (2-year and 5-year systems), colleges and universities of technology (2-year and 4-year systems), and the vocational high schools mentioned earlier. In 2006, there were 29 universities of technology, 46 colleges of technology, and 16 junior colleges in Taiwan. The total number of students was 1 001 098 including 331 604 vocational high schools students, 180 886 junior colleges students, and 465 840 undergraduates and 22 768 graduate students of colleges/universities of technology.

In 2001, the Taiwan MOE provided funds to establish the Testing Center for Technological and Vocational Education and the Committee of Recruitment Policy Advancement for Technological Colleges and Universities. The main purpose of those two institutions is to meet the needs of over 70% vocational high school students expecting to go on to further education and around 25% of senior high school students qualified to study in colleges/universities of technology. Their chief undertakings are to improve the research and development on the TVE joint college entrance examination and integrate diversified entrance schemes, simplify the admission process, enhance test quality, and allow the TVE colleges autonomy in recruiting students. Furthermore, the MOE has established six Centers for Regional Industry-Academia Collaboration and 40 technical research centers affiliated to various universities to foster more school–industry collaborations.

Higher (Tertiary) Education

In the past 10 years, Taiwan's higher education policies, among other factors, have led to a substantial increase in the number of colleges and universities. More than 80% of the senior high school graduates and 70% of the vocational high school graduates are able to enter institutions of higher education. The higher education system, including the technological and vocational colleges and universities, is made up of 97 universities (42 public and 55 private), 50 colleges (10 public and 40 private), and 16 junior colleges (3 public and 13 private). In 2006, the total higher education enrolment was around 1 313 993 (including 29 839 PhD program students), 163 583 masters program students, 966 591 undergraduate students, and 153 978 junior college students. The enrolment rate in that age demographic in Taiwan was 82.02% (79.14% male and 85.08% female).

Besides the track of technological and vocational education system, the MOE adopted the multiple admission channels for entrance to colleges and universities for senior high school students in 2002. One channel is based on the requirement of the score of General Scholastic Ability Test and refers to recommendation (school recommendation or individual student's application) and screening. The second channel is via the Subjects Competence Tests, in which three to six subjects would be designed by individual departments of the colleges and universities, and the schools will then assign students according to their test scores and applications by the Joint Board, College Recruitment Community (JBCRC). The College Entrance Examination Center (CEEC) conducts the General Scholastic Ability Test and Subjects Competence Tests nationwide.

Special Education

Special education in Taiwan incorporates that for the talented and gifted (including general intelligence, academic aptitude, creativity, and leadership) and the disadvantaged students (including mental retardation, health impairment, speech disorders, learning disabilities, severe emotional disturbances, multiple impairments, autism, developmental delays, and other significant disorders). The Special Education Act of 1997, first announced in 1984, showed that the trend of special education has shifted from separation to mainstream, integration, and overall inclusion.

In Taiwan there are special education units (the Special Education Task Force Unit) in MOE and several Special Education Units in local DOEs/BOEs, which are in charge of supervising the instruction, counselling, assessment, and placement for special education. There are 25 special education schools (23 public and two private) providing a comprehensive learning environment to cover each category of disadvantaged students. Besides, there are also various types of placement in regular education institutions such as special classes, resource programs, and regular classes receiving special education in elementary, junior, or senior high schools. At the college and university level, there are several resource classes for visually and hearing impaired students.

Supplementary and Continuing Education

Supplementary education has decreased and transformed its functions because a great majority of Taiwan's children have been able to receive formal education over the past four decades. In 2006, there were only 318 supplementary schools for primary education and 470 for secondary education. These schools provide basic education and work-skills for out-of-school youth. In addition,

supplementary schools emphasize lifelong learning and provide educational programs to enhance adult literacy, and in the last 10 years programs to enhance the educational services for foreign spouses as well.

Cram schools (called *bu-shi-ban* in Taiwan) are specialized private institutions that offer classes after normal school time or on weekends to meet customized goals, most commonly to pass the entrance examinations of high schools and colleges or universities. Cram schools offer instruction in areas such as mathematics, natural sciences, English, and certain kinds of talented competencies which are necessary for entering a higher level of schooling. Taiwan's cram schools have increased from 3478 in 1998 to 16 982 in 2007; however, the quality of cram schools varies considerably.

In Taiwan, an assortment of learning modes for adult and continuing education has greatly increased its depth and breadth. These modes contain libraries, museums, art education centers, science education centers, extension division of colleges/universities for in-service and continuing education, four national social education centers, two open universities, 276 social work stations, and more than 30 community universities, a program strongly advocated by the educational reform movement in the private sector and aiming to liberalize the dissemination of knowledge and advance the formation of civil society.

The Teaching Profession

In 2006, there were 19 037 full-time kindergarten teachers, 150 441 elementary and junior high school teachers, and 50 749 senior high school teachers in Taiwan. Among them, more than 85% are instructional faculty in public schools, more than 85% are younger than 50 years, and around 70% are female. Under the Teacher Education Act, first issued in 1994 and last revised in 2003, around 75 institutions including Normal Universities, University of Education, and colleges/universities with teacher education departments or centers, are eligible to provide the initial training for student teachers from kindergarten to secondary education level. The initial training generally includes a four-year program and a half-year of teaching practicum in order to obtain the certificate. Starting from 2005 the students have to take the six-month practicum and then pass a qualification examination held by the MOE for the affirmed teacher certification. As of today, more than 55 000 people with teaching qualifications are waiting for teaching jobs.

Regarding Taiwan's higher education, there were 10 393 full-time junior college teachers, 11 225 college teachers and 37 790 university teachers. Among them, about 32% are instructional faculty in public institutions, more than 73% are younger than 50 years, around 34%

are female, and 76.6% are teachers with a PhD degree in public schools and 51.7% in private schools. However, like many other countries, there is no preservice training program for higher education faculty in Taiwan.

Administrative and Supervisory Structure

The administrative and supervisory structure in Taiwan consists of two levels: the MOE at the central government level and the DOEs/BOEs at the county and municipal government level. The MOE mainly takes charge of nationwide educational and cultural affairs, and supervises around 24 affiliated institutions for social/civic education, national colleges/universities, and around 170 affiliated elementary and high schools. The DOEs/BOEs administrate elementary and high schools and social education units or organizations within their jurisdictions. Private schools of all levels in Taiwan are subject to the supervision of the MOE and DOEs/BOEs, but their personnel and financial policies still have to follow the regulations of the Private School Act, last amended in 2007. In addition, according to the University Act, last amended in 2007, colleges/universities are guaranteed academic freedom and autonomy as specified within related laws and regulations. The MOE is gradually empowering the colleges/universities and improving higher education quality through financial subsidies, grants, and school evaluation.

Educational Finance

The Educational Budget Allocation and Management Act, passed in 2000 and executed in 2002, stipulates that Taiwan government must guarantee a stable educational budget and establish a committee to regulate the use of the educational budget at different levels of schools. In 2005 the educational expenses reached NT\$ 676 billion (6.07% of GNP, 18.41% of total governmental budget), of which public education took up 72.46% (4.39% of GNP) and private education 27.54% (1.68% of GNP). The distribution of total educational expenditure at all levels of schools was 3.23% for kindergarten education, 40.26% for compulsory education, 15.56% for senior secondary education, 0.66% for special education, and 40.29% for higher education.

Over the last 10 years, revenue sources for Taiwan's higher education have undergone a significant transformation. Government funding for public colleges/universities have diminished from two-thirds to one-half of each public school's budget and less than one-third for private school budget. In addition to standard government funding, colleges/universities may receive special grants for research or instructional excellence from the government. The remainder of the school budget comes from the

private sector, school service, and revenue from student fees. In 2006, the average tuition fees for full-time college/university students was US\$1793 for public schools and US\$ 3265 for private schools. The fees in public higher education schools was 12.74% and that in private ones was 23.21% as related to per capita income of US\$ 14 070. The MOE supports the financially disadvantaged students with full scholarships, grants, and financial loans totaling nearly NT\$ 10 billion per year.

Performance Monitoring, Evaluation, and Research

The MOE and DOEs/BOEs are responsible for monitoring and supervising performance of each sector of the educational system. In addition, the government has to collect data through agencies like the Department of Statistics of MOE and to conduct research on the outcomes of the educational system. Scholars of teacher-training and education-related colleges/universities undertake most educational researches. Professional institutions, such as the Center for Educational Research of National Taiwan Normal University, have been playing an active role in educational research. The National Academy for Educational Research Preparatory Office, established in 2000, focuses on developing its functions on educational research, consultation, and in-service training of teachers and principals. In addition, the Higher Education Evaluation and Accreditation Council of Taiwan, founded in 2005, is responsible for university evaluations.

Several educational statistics and indicators reveal Taiwan's educational achievement. Overall education participation rates in 2006 were nearly 100% for the 6–12-year-olds, 99.83% for the 12–15-year-olds, and 82.02% for the 15–18-year-olds. In 2005, the national illiteracy rate for those over 15 years old was 2.7%. The number of elementary and junior high school dropouts was 1785 (around 0.06% of the students of the same age group). The percentage of the population that have acquired at least senior secondary education was 64.2% for the 25–64-year-olds and 87% for the 25–34-year-olds, and those with a higher education diploma was 30.7% for the 25–64-year-olds and 47.9% for the 25–34-year-olds.

At the international level, Taiwan consistently ranks among the top ten highest performing countries in the Trends in International Mathematics and Science Study (e.g., both grade 4 and grade 8 won the second prize in science and the fourth in mathematics in 2003) and International Olympiad for high school students. In 2007, Taiwan got two gold and two silver medals in chemistry; two gold, one silver, and one bronze medals in information technology; one gold, two silver, and two bronze

medals in physics; two gold and two bronze medals in biology; and two gold, three silver, and one bronze medals in mathematics. In addition, Taiwan's research papers of higher education presented in Science Citation Index ranked 18th and 11th in Engineer Information in 2005.

Major Changes and Issues Since the 1990s

Taiwan has gone through continual educational reforms beginning in the 1990s. Many educational officials and nongovernmental committees on educational reform have published numerous white papers and educational reports. Several education-related laws and policies have been promulgated, amended, and implemented. According to the MOE, the goals of educational reforms are to enhance the quality of education by establishing a flexible school system, reduce the pressures of entering higher schooling, shorten the gap between the educational resources of rural and urban areas, balance the resources allocated to public and private schools, promote lifelong education, build a learning-oriented society, and reinforce international cultural and educational cooperation. In general, Taiwan's educational reforms are oriented towards diversification, liberalization, and globalization. However, the effectiveness of Taiwan's educational reforms still needs to be carefully reviewed.

Taiwan still has to resolve numerous issues at different educational levels. In preschool and early childhood education, the main issues are how to integrate kindergartens and childhood care centers effectively, and the possibility of extending the compulsory education to include 5- to 6-year-olds. In primary and secondary education, the essential issues are how to plan the 12-year mandatory education for primary and secondary education in order to reduce students' academic pressure in entering high schools, how to enhance the educational equity for minorities, and how to revise and continuously monitor the curricula guidelines through school performance. As for the teaching profession, some critical issues are how to encourage the professional development and how to resolve the oversupply of preservice teachers. In technological and vocational education, the issues are how to emphasize the licenses of technical expertise and how to strengthen the connection between schools and industries. Finally, several major issues in higher education are of concern. Those issues include how to improve the globalization and competitiveness of Taiwan's colleges/universities, how to evaluate those colleges/universities properly, and how to improve the quality, and achieve successful transformation of the higher education system.

Further Reading

- Chien, M. F. and Lee, C. -M. (eds.) (2001). *Modern Educational Indicators: International Comparison Viewpoints (in Chinese)*. Taipei: Pro-Ed.
- Lee, C. -M. (2004). Changes and challenges for moral education in Taiwan. *Journal of Moral Education* 33(4), 575–595.
- Taiwan Ministry of Education (2005). *2005 Education in Taiwan*. Taipei: Ministry of Education.
- Taiwan Ministry of Education (2006). *2006 Education in Taiwan*. Taipei: Ministry of Education.
- Taiwan Ministry of Education (2007). *2007 Education in Taiwan*. Taipei: Ministry of Education.
- Taiwan Ministry of Education (2007). *Yearbook of Teacher Education Statistics of Republic of China (Taiwan)* (in Chinese). Taipei: Ministry of Education.

Relevant Websites

- <http://bsb.edu.tw> – All cities and counties of Short-Term Bu-shi-ban (Taiwan).
- <http://www.ceec.edu.tw> – College Entrance Examination Center (Taiwan) (in Chinese).
- <http://www.bctest.ntnu.edu.tw> – Committee of the Basic Competence Test for Junior High School Students (Taiwan) (in Chinese).
- <http://www.moi.gov.tw> – Department of Statistics, Ministry of Interior (Taiwan).
- <http://www.edu.tw> – Department of Statistics, Ministry of Education (Taiwan) (in Chinese).
- <http://www.edu.tw> – Department of Statistics, Ministry of Education (Taiwan). 2007 Educational Statistics of Taiwan (in Chinese).
- <http://www.ece.moe.edu.tw> – Early Childhood Education (in Chinese).
- <http://en.wikipedia.org> – Education in Taiwan – From Wikipedia, the Free Encyclopedia.
- <http://www.heeact.org.tw> – Higher Education Evaluation and Accreditation Council of Taiwan (HEEACT).
- <http://www.ibo-info.org> – International Biology Olympiad (IBO 2007).
- <http://www.hsin.hr> – 19th International Olympiad in Informatics (IOI 2007), Croatia.
- <http://www.icho39.chem.msu.ru> – 39th International Chemistry Olympiad (IChO 2007).
- <http://www.imo2007.edu.vn> – 48th International Mathematical Olympiad (IMO 2007).
- <http://www.ipho2007.ir> – International Physics Olympiad (IPhO 2007).
- <http://www.jbcrc.edu.tw> – Joint Board, College Recruitment Community (JBCRC).
- <http://law.moj.gov.tw> – Laws and Regulations Database of Taiwan.
- <http://english.moe.gov.tw> – Ministry of Education, Republic of China (Taiwan).
- <http://apcu.taiconet.com.tw> – National Association for the Promotion of Community University (in Chinese).
- <http://www.spc.ntnu.edu.tw> – National Special Education Information Network.
- <http://www.gio.gov.tw> – Taiwan Government Information Office, Republic of China.
- <http://www.president.gov.tw> – Taiwan Office of the President, Republic of China.
- <http://www.tcte.edu.tw> – Testing Center for Technological and Vocational Education (TCTE) (Taiwan).
- <http://nces.ed.gov> – Trends in International Mathematics and Science Study, National Center for Education Statistics (NCES).
- <http://education.yahoo.com> – World Fact Book: Taiwan.

General Background and History

The United Republic of Tanzania includes a large area of mainland East Africa and the islands of Zanzibar (Unguja, Pemba) and Mafia. It covers an area of 0.9 million square kilometers and has a population of around 39.4 million. The vast majority (80%) of the workforce are involved in agriculture, which accounts for around 40% of Tanzania's GDP. Since 2001, Tanzania has experienced a growth rate in GDP of around 6%. In spite of this economic growth, Tanzania remains one of the poorest economies in the world with a GDP per capita of US\$744 in 2005. By international standards, the majority of the population would be considered to be living in poverty, surviving on under a dollar a day.

The United Republic of Tanzania was formed by the union of Tanganyika and Zanzibar in 1964. Tanganyika became a German colony in 1885. After World War I, it was handed over to the British as a mandated territory under the League of Nations and, in 1961, it gained its independence. Zanzibar was ruled by the Portuguese in the sixteenth and seventeenth centuries and by Oman in the eighteenth and nineteenth centuries. In 1890, it became a British protectorate. It gained its independence in 1963.

Prior to German colonization, the only schools in Tanganyika were a limited number of small mission schools and Koranic schools restricted to the coastal areas. Some tribes had highly organized initiation-training periods, but otherwise there was no formal education. The Germans set up a limited number of schools to provide for the manpower needs of the administration. The first government school was opened in Tanga in 1892.

World War I shattered the education system that had been set up by the missionaries and colonial German powers. The missionary societies soon resumed their activities, but it was not until 1925 that the British developed a clear education policy for the territory. The overall policy was one of education for adaptation. It supported a broad-based, relevant education system with the curriculum adapted to rural African life. The education system was racially segregated into separate systems for Africans, Asians, and Europeans. The curriculum for African schools stressed agriculture, vocational skills, and hygiene.

During the interwar period, the priority was mass education, giving basic education to as many as possible, rather than focusing resources on the higher levels of education. Due to this policy, the colonial government

did not start developing secondary education until relatively late. In the last few years of colonial rule, the colonial administration paid greater attention to secondary education. This led to a proportionately dramatic increase in the number of secondary school places, but there were still only 700 African pupils in form 4 at the eve of Independence.

Immediately after Independence, the education system was desegregated and progressively indigenized through the replacement of colonial staff, curricula, and examinations. Education policy focused on manpower development, as it had in the final years of colonial rule, and secondary education expanded rapidly. Education was seen as a tool for bringing about rapid modernization. The curriculum was academic with science and technology strongly promoted.

Higher education became available in 1961, with the opening of the University College of Tanganyika as a college of the University of London. The college later became a constituent college of the University of East Africa and then, in 1970 it became an independent national university and was established as the University of Dar es Salaam.

In 1967, education policy in Tanzania was dramatically reoriented according to *Education for Self-Reliance* (ESR), a philosophical treatise on education by Julius Nyerere, the country's first president. Nyerere presented a bottom-up model of development based on a predominantly rural economy and a socialist ideology. ESR returned the focus of educational development to the provision of basic, relevant education for all. Agricultural activities became a central part of school life. Expansion of secondary education was tightly controlled and all non-government-run schools, with the exception of seminaries, were nationalized.

The Musoma Resolution of 1974 set the goal of achieving universal primary education (UPE) by 1978. The nation made an enormous effort to train new teachers and build new schools. Primary enrolment soared from 1.2 million in 1974 to 3.5 million in 1981, while secondary enrolment stagnated. However, many of the school structures built by communities were of poor quality and soon fell into disrepair. Teachers were recruited straight from primary schools and were trained by distance training.

The world economic recession of the 1980s and structural adjustment policies imposed by the International Monetary Fund led to a huge reduction in the level of public spending on education, compounding the decline

in quality from the push for UPE. The expanded primary cohorts reached the end of primary schooling only to find that the vast majority could not progress to secondary school. This reduced the incentive for parents to enrol their children in primary schools and absolute primary enrolments fell from 1983 to 1986. Although total enrolments rose slowly thereafter, the rise did not keep up with population growth and the net enrolment rate fell to below 60% by the end of the twentieth century.

While the state education system declined throughout the 1990s, liberalization enabled the private sector secondary education system to flourish and, by 1999, there were almost as many private secondary schools as state schools. More recently, liberalization and reform within the higher education system has led to a boom in the private university sector.

International donors have had a major influence on educational development in Tanzania throughout its history. In the 1970s, vocational biases were introduced into secondary schools with strong support from the World Bank. Since then the World Bank's thinking on the effectiveness of diversified secondary education has changed and more recent support to secondary education was only granted once the government agreed to remove vocational biases from secondary schools.

The new millennium saw major changes in the approaches to international development by multilateral and bilateral donors, both in terms of priorities and the mode of delivery. Poverty reduction and attainment of the millennium development goals, including universal, free, and compulsory primary education and gender equity in education, have become the priority of international development agencies. Tanzania has responded to these changes by producing a poverty-reduction strategy paper and by increasing the amount it spends on primary education. International aid, debt relief, and government reform with a focus on poverty reduction have enabled a major reversal in the decline in the state education system, and enrolments at all levels of education have increased dramatically since 2000.

Goals of the Education System

The goals of education and training in Tanzania are set out in the *Education and Training Policy* (United Republic of Tanzania, 1995). This gives the aims and objectives for the sector as a whole and specific aims for each sub-sector. Promotion of national unity through the acquisition and appreciation of Tanzanian culture, customs, and traditions is one of the overall goals.

Nyerere's vision of education for self-reliance saw education as a means to enable the country to develop through the efforts of its own people, without support or intervention from outside. It saw the primary goal of

education as preparing all children for life as self-reliant citizens in a primarily rural agricultural economy. The current *Education and Training Policy* (United Republic of Tanzania, 1995) has maintained some of Nyerere's language of education for self-reliance but the interpretation is very different, with a much greater emphasis on development of scientific and technological capacity. Whereas Nyerere argued that primary education should focus on preparing children to enter work on completion of the first 7 years of schooling, the current education and training policy focuses more on the role of primary education in preparing children for further education and training. Preparation of the child for the world of work is given as an aim of primary education, but it comes at the bottom of the list.

Tanzania's development vision, Vision 2025, sees education as a tool for national development. It stresses the importance of expanding tertiary education in order to develop a "critical mass of high quality human resources required to effectively respond and master the development challenges at all levels" (United Republic of Tanzania, 1999). At the same time, it stresses the importance of scientific literacy for all, stating that the education system "must instill a science and technology culture from its lowest levels, giving a high standard of education to all children between the age of 6 to 15." Attainment of computer literacy is also becoming a more prominent goal and in 2007, the Ministry of Education and Vocational Training produced its *Information and Communication Technology Policy for Basic Education* (United Republic of Tanzania, 2007b).

Since 2000, education has become more explicitly linked with poverty reduction in Tanzanian policy. In the *Education and Training Sector Development Programme Document* (United Republic of Tanzania, 2001) the goal of poverty reduction is given as a rationale for allocating the majority of educational spending to the primary level. While the language of poverty reduction is relatively new in Tanzania, the idea of prioritizing basic education for all over post-primary academic education for some was at the heart of Nyerere's *Education for Self-Reliance*.

Structure of the Education System

The official entry age into formal education in Tanzania is 5 years. Mainstream education on the mainland involves 2 years pre-primary, 7 years primary, 4 years lower secondary, 2 years upper secondary, and 3–4 years university degree. On Zanzibar, lower secondary school is split into two cycles, one of 3 years and a second cycle of 2 years. Primary school leavers can also enter vocational education training colleges, although increasing competition for the more popular classes means that primary school leavers often miss out to secondary leavers. Secondary graduates can enter a range of non-university-based training courses, most of which lead to a certificate or diploma.

Table 1

Level	NER (%)	GER (%)	GPI
Pre-primary	33	35	1.05
Primary	97	114	0.99
Lower secondary (1–4)	21	30	0.92
Upper secondary (5–6)	1	4	0.59

Table 1 shows the net enrolment ratios (NERs) and gross enrolment ratios (GERs) for the various levels of education. It also gives the gender parity index (GPI), calculated by dividing the number of females enrolled by the number of males.

In 2007, around one-third of children aged 5 and 6 attended pre-primary schools. Almost all (97%) children aged between 7 and 13 were enrolled in primary school. As illustrated by the higher GER, around 10% of children in primary schools were older than 13 years. In 2006, around two-thirds of primary school leavers were able to enrol in form 1 of secondary school and the secondary NER in 2007 was 21%. Recent expansion of secondary education has meant that many children over the official entry age of 14 years who missed out on selection for secondary school in previous years are now enrolling. In 2007, around 80% of form 1 pupils were 15 years or older. Because of this, the GER is much higher than the NER. Similarly, most pupils in upper secondary are above the official age (18–19 years).

At pre-primary schools, there are more girls than boys, but the enrolment levels become steadily more biased toward males going up the system. At the primary level, there is almost gender parity. At lower secondary level, boys outnumber girls by nearly 10% and at the upper secondary level, there are over twice as many males as females. At higher education levels, around one-third of the students are female.

Public examinations take place at the end of the fourth and seventh years at primary school, and at the end of the second, fourth, and sixth years at secondary schools. The key national qualifications from schools are the Primary School Leaving Examination (PSLE), the Certificate of Secondary Education (CSE or O' level), and the Advanced Certificate of Secondary Education (ACSE or A' level).

Pre-Primary

Formal pre-primary education for children aged 5–6 has expanded significantly since the turn of the millennium. Preschools provide 2 years of early education taught mainly in Kiswahili, following a national curriculum. The vast majority of children in registered pre-primary schools are in government schools (97.5%). Pupil–teacher ratios (PTRs) in government pre-primary schools have improved from 55 pupils per teacher in 2004 to 45 in 2007.

PTRs in non-governmental pre-primary schools are considerably lower (average PTR of 14 in 2007).

Primary

Primary education is free and compulsory. In all but one region, enrolment of primary school age children is over 90%. The vast majority of primary school students (99%) are in government schools. The medium of instruction is Kiswahili, and English is taught as a subject. Private primary schools are mainly English medium and tend to serve elites. The curriculum includes Kiswahili, English, mathematics, social studies, science, and vocational studies. The PTR is around 50, although it is as high as 70 in some regions.

Secondary

Within the government system, there are a small number of national secondary schools, most of which are boarding. Places are allocated based on students' success in the PSLE. Most government schools are community schools, built by the local community and staffed and run by the government. Nineteen percent of secondary students are in non-governmental schools. In 2006, there were 679 non-governmental schools. Some are run by faith-based organizations, others are run for profit. There are around 100 religious seminaries (56 Christian and 42 Muslim) that provide secondary education alongside preparation for religious service. Seminaries have tended to have the highest academic performance in national examinations.

The secondary curriculum for the first 2 years includes basic mathematics, English, Kiswahili, history, geography, civics, biology, and physics with chemistry. In the third and fourth years, students can choose to specialize in either arts or sciences. Religious education (Christian or Islamic) is taught by local religious leaders. French, further mathematics, and information and computer studies are also taught at some schools. At upper secondary level, students can choose a combination of three subjects to study.

The medium of secondary and higher education and training is English. Most secondary students and some teachers struggle with comprehension in English and it is seen as a major barrier to effective learning. In practice, Kiswahili is used extensively in most secondary schools. This has prompted many linguists and educationalists to lobby for a change to Kiswahili medium at secondary level. However, English medium education remains the popular choice because it is seen as vital for students to learn English.

Higher Education

In 2006, there were 30 universities in Tanzania, 11 of which were public and 19 private. The total student

population was 52 657 students, of whom 33.5% were female. The Open University of Tanzania (13 000 students) and the University of Dar es Salaam (12 000 students) together account for around 50% of all of the students in higher education in Tanzania. The other public universities are mainly specialist institutions. These include Sokoine University of Agriculture, Muhimbili University College of Health Sciences, and the University College of Lands and Architectural Studies. There are many other smaller higher education institutes. In 2006, these accommodated 16 372 students, 14.6% of whom were female.

Technical and Vocational Education

There is a wide variety of state-run vocational training institutes, including small post-primary colleges and folk development colleges offering 2-year courses in manual trades. There are also national vocational training colleges in the urban centers offering courses in areas such as secretarial skills, computing, catering, and tourism. Much of the vocational training in rural areas is provided by faith-based organizations. They tend to deliver 2-year courses in manual trades such as tailoring, masonry, and carpentry. They have relatively low fees and train mainly primary leavers. Private centers run for profit are mostly located in urban areas and offer shorter, more expensive courses in computing and commercial skills. Vocational courses are accredited through a national trade test.

The Teaching Profession

Teacher training is carried out through a network of colleges and the faculties of education at universities. Potential primary school teachers must study for the certificate of education. This is a 2-year course with 1 year in college and 1 year in school. Trainees must have passed their CSE, and they specialize in arts or sciences. The certificate qualifies teachers as grade-A teachers. Teachers currently in service who have not completed secondary school are classified as grade-B or -C teachers, depending on the level of training they have received. All grade-B/C teachers are expected to upgrade through a Kiswahili medium distance-learning teacher training program.

The diploma in education qualifies teachers to teach at lower secondary. It is a 2-year residential course and candidates must have successfully completed upper secondary education. University faculties of education deliver 3-year degree courses in education. Graduates from other degree programs can also qualify as teachers through a 1-year postgraduate diploma in education. A growing number of teachers with diplomas are studying part-time for education degrees through the Open University of Tanzania.

The official minimum qualifications for teachers at the various levels are:

- Pre-primary – pre-primary certificate of education
- Primary – CSE and certificate of education
- Lower secondary – diploma of education
- Upper secondary – degree in education or degree with PGDE

In practice, only a minority of pre-primary teachers are qualified (15%). The majority of primary school teachers are qualified at grade A and there are a small number of diploma holders working in primary schools. In 2007, around a quarter of the teaching force were still graded at B/C level.

According to the *Education and Training Policy*, teachers of forms 3 and 4 at secondary schools should be degree holders. In practice, there are very few degree holders teaching at O' level and most secondary school teachers who teach forms 1–4 are diploma holders. A growing proportion of lower secondary teachers are licensee teachers – these are ACSE holders who are given crash training. This involves 3 weeks of college-based training followed by in-service training through self-study.

Administrative and Supervisory Structures

Formal education from pre-primary to upper secondary school, adult education, and vocational training are the responsibility of the Ministry of Education and Vocational Training (MoEVT). Higher education and research are the responsibility of the Ministry of Higher Education, Science and Technology (MoHEST).

Curriculum development is carried out by the Tanzania Institute of Education (TIE). Earlier, TIE was the main producer of school textbooks, but in recent years it has relinquished this role, opening up textbook production to the commercial market. The National Examinations Council of Tanzania (NECTA) is responsible for national examinations at primary and secondary schools (PSLE, CSE, ACSE), and for qualifications from teacher training colleges (certificate and diploma).

Earlier, vocational education and training used to come under a number of ministries; however, since 2005 it has been administered together with formal education through the MoEVT. The Vocational Education and Training Authority (VETA) runs the national vocational training centers and is responsible for the administration of the trade tests.

Zanzibar remains semi-autonomous and has its own Ministry of Education, Culture and Sports. This ministry is responsible for pre-primary, primary, secondary, adult, and higher education.

Educational Financing

State Funding

In 2005/06, primary education and supporting services received around 64.5% of all spending on education. Twelve and a half percent went to secondary education, 22% to higher and tertiary education, and 1% to teacher training.

Primary education is officially fully financed by the state, with school fees being abolished in 2001. In practice, some schools raise additional funds through a system of parental contributions. State secondary schools charge school fees of around US\$20 per year but the main costs are met by the state. The government provides free higher education for a restricted number of candidates. Individuals can also fund themselves as private candidates at state or private universities.

VETA raises funds to run its centers through a levy on businesses. Non-government providers of vocational and technical education raise their funds through user fees and donations. They pay a fee to VETA for accreditation.

Donor Support

Throughout the 1990s, most of the donor funding was through the project approach and did not contribute to recurrent costs such as teachers' salaries. While having a major influence on educational development, it only constituted around 6% of all government spending on education. Donor support to education is now done largely through basket funding, which goes toward both recurrent and development costs. From 2002 to 2006 this was targeted at the primary sub-sector. The support has become proportionately much more significant. In the financial year of 2002/03, foreign support to education was 219 billion TSh out of a total of around 520 billion TSh, constituting over 40% of spending on education (United Republic of Tanzania, MoF 2003).

Performance Monitoring, Evaluation and Research

There are eight school inspectorate zones. According to policy, each school should be inspected once a year; in practice schools are inspected once every 2 years. National and regional education statistics for pre-primary to upper secondary education are published every year by MoEVT in the *Basic Education Statistics for Tanzania (BEST)* series. MoEVT and MoHEST make education statistics available on their websites, and progress in the education system is reported in annual budget speeches by the ministers. School performance in the national examinations is published each year by NECTA.

For the primary education development plan, which had major donor support, annual joint reviews were carried out by the government of Tanzania together with representatives from the donor community. There were also monitoring reports published by the President's Office – Regional Administration and Local Government (PORALG).

All research in Tanzania, including research into education, comes under the supervision of the Commission for Science and Technology (COSTECH). The commission has the power to award research grants. In practice, most academic research into the education system in Tanzania is funded by, and often conducted by, non-Tanzanians. The staff of the Faculty of Education in the University of Dar es Salaam are heavily involved in research, but most work on projects commissioned by or in partnership with international organizations. There is, therefore, a tendency for research to reflect the concerns of the international development community.

There are a number of independent research organizations that commission and conduct research into education, including The Economic and Social Research Foundation (ESRF), Research on Poverty Alleviation (REPOA), and HakiElimu, a Tanzanian non-governmental organization committed to publishing accessible research and analysis of education development within Tanzania.

Major Changes and Issues Since the 1990s

At the end of the twentieth century, the education system in Tanzania was in a state of disrepair. The first global monitoring report of Education For All (EFA) (UNESCO, 2002) grouped Tanzania in the lowest quadrant of countries, those that had a net enrolment ratio of below 70% (the figure given for Tanzania being 47%) and where enrolment rates had fallen since 1990. This group of countries was classed as at serious risk of not achieving the goal of universal primary education by 2015.

Since the year 2000, the number of children enrolled in primary schools has almost doubled, increasing from 4.4 million in 2000 to 8.3 million in 2007. Net enrolment rates have increased to above 95%. This rapid growth is a result of the decision to drop primary school fees in 2001 and the initiation of the primary education development plan (PEDP) in July 2002. PEDP received huge budgetary and technical support from the international donor community and also increased government spending on primary education. Besides mainstream education, the plan included programs, such as complementary basic education in Tanzania (COBET) and the integrated community based adult education (ICBAE) for youth and adults who have missed out on primary education.

In terms of quality, PEDP aimed at reducing PTR to 40:1, providing teachers with in-service training and providing a US\$10 grant per student to be spent by the schools on teaching–learning materials. These changes took longer to take effect than the quantitative changes but, by 2005, capitation grants were reaching schools and the output from teacher training programs had increased. There has been an apparent leap in quality as measured by the results of the PSLE, with over 70% passing in 2006 as compared with 22 % in 2000.

In 2004, the Secondary Education Development Plan (SEDP) was launched with financial support (in the form of a grant and a loan) from the World Bank. Government spending on secondary education was increased considerably and community mobilization, encouraged by matched government funding, led to the construction of hundreds of new secondary schools. In the first 3 years of SEDP, form 1 enrolment increased by a factor of 3 from 147 490 (2004) to 448 448 (2007), and the net enrolment ratio at lower secondary increased from 8.4% (2004) to 30.5% (2007). To accommodate for the huge influx of new students, upper secondary school leavers have been trained as teachers through a crash teacher training program. Aside from the World Bank, international support for SEDP has been much more limited than for PEDP. There is, therefore, less funding available for improving quality and it is yet to be seen what the full impact of the rapid expansion will be on the quality of secondary education.

Bibliography

- UNESCO (2002). *Education for All Global Monitoring Report*. Paris: UNESCO.
- United Republic of Tanzania (1999). *Development Vision 2025*. <http://www.tanzania.go.tz/vision.htm> (accessed May 2009).
- United Republic of Tanzania (2000). *Poverty Reduction Strategy Paper*. Dar es Salaam, October 2000.
- United Republic of Tanzania (2001). The education sector development programme document. Dar es Salaam.

- United Republic of Tanzania, M. o. E. C. (2003). Primary education development plan 2002–2006. Dar es Salaam: Ministry of Education and Culture.
- United Republic of Tanzania, M. o. E. C. (2004). Secondary education development plan 2004–2009. Final document. Dar es Salaam, Ministry of Education and Culture.
- United Republic of Tanzania, M. o. E. C. (1995). *Education and Training Policy*. Dar es Salaam: Government Printers.
- United Republic of Tanzania, M. o. E. V. T. (2007a). *Basic Statistics in Education 2007*. Dar es Salaam: Ministry of Education and Vocational Training.
- United Republic of Tanzania, M. o. E. V. T. (2007b). *Information and Communication Technology (ICT) Policy for Basic Education*. Dar es Salaam: Ministry of Education and Vocational Training.
- United Republic of Tanzania, M. o. F. (2003). Budget Execution Report: Budget for fiscal year 2002/2003. Dar es Salaam: Ministry of Finance.

Further Reading

- Galabawa, J. C. J., Senkoro, F. E. M. K., and Lwaitama, A. F. L. (eds.) (2000). *The Quality of Education in Tanzania*. Dar es Salaam: University of Dar es Salaam Faculty of Education.
- Nyerere, J. K. (1967). Education for self reliance. Arusha: United Republic of Tanzania.
- Roy-Campbell, Z. M. and Qorro, M. (1997). *Language Crisis in Tanzania: The Myth of English versus Education*. Dar es salaam: Mkuki na Nyota.
- Wedgwood, R. (2005). *Post-Basic Education and Poverty in Tanzania (Post-Basic Education and Training Working Paper Series No. 1)*. Edinburgh: Centre of African Studies, University of Edinburgh. <http://www.cas.ed.ac.uk/research/projects.html> (accessed May 2009).

Relevant Websites

- <http://www.esrf.or.tz> – Economic and Social Research Foundation Tanzania.
- <http://www.hakielimu.org> – People Making Difference in Education and Democracy.
- <http://www.repoa.or.tz> – Research on Poverty Alleviation.
- <http://www.costech.or.tz> – Tanzania Commission for Science and Technology.
- <http://www.tanedu.org> – Tanzania Education and Information Services Trust.
- <http://www.tie.go.tz> – Tanzania Institute of Education.
- <http://www.necta.go.tz> – The National Examinations Council of Tanzania.
- <http://www.moe.go.tz> – The United Republic of Tanzania, Ministry of Education and Vocational Training.
- <http://www.msthe.go.tz> – MOHEST.

Thailand

C Rukpollmuang, Chulalongkorn University, Bangkok, Thailand

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Basic education – The education provided before the level of higher education.

Lifelong education – The education resulting from the integration of formal, nonformal, and informal education so as to enhance ability for continuous lifelong development of quality of life.

General Background

The kingdom of Thailand was formerly known as Siam until 11 May 1949 when there was an official proclamation declared that the country would henceforth be known as Thailand, meaning land of the free, since the kingdom has never felt the harness of direct European colonialism.

The country is situated in the heart of the Southeast Asian mainland, covering an area of 513 115 square kilometers. According to the National Statistical Office, the total population was 64.86 million in 2006 and was estimated at 65 068 149 in July 2007. Thailand's population growth rate is on a decline, from 0.91% in 2006 to 0.68% in 2007. It is also reported that approximately 25% of the population is under the age of 15 years and population below the poverty line is 10.4%.

Since the political reform of the absolute monarchy in 1932, Thailand has had about 18 constitutions and charters. However, it is noteworthy that regardless of just a constitutional monarchy, as far as the Thai public is concerned, the authority of the much beloved king contributes a great deal to the nation unity and stability.

The official national language is Thai. Buddhism, the national religion, is the professed faith of 95% of the population. Islam (3.8%), Christianity (0.5%), Hinduism (0.1%), and other religions (0.6%) are embraced by the rest of the population. There is absolute religious freedom. The king of Thailand, under the constitution and in practice, is a patron of all major religions.

The Thai public administration is structured into central and local administration. According to the Department of Local Administration, Ministry of Interior, as of 30 September 2007, there are 75 provinces, 1276 municipalities and sanitary districts, as well as 6500 provincial administrative organizations (PAOs). Moreover, there are

two special forms, which are the Bangkok Metropolitan Administration and the city of Pattaya.

Traditionally, the kingdom economy has been solely based on agriculture; however, in the 1980s, Thailand emerged as one of Asia's leading tiger economies. However, after the economic crisis in the summer of 1997, primarily due to a renowned Orient syndrome, the country has revised its development goals. The present 10th National Economic and Social Development Plan (2007–11) has placed His Majesty King Bhumipol Adunyadej's philosophy on sufficiency economy as its foundation, and aims at three main objectives: people-centered development; balancing between economic, social, and natural resources and environment capitals; and leading to the green and happiness society.

The country's real gross domestic product (GDP) growth in 2007 was 4.7%, driven mostly by a strong export sector. The Bank of Thailand, the Thai Finance Ministry, and the Thai National Economic and Social Development Board have estimated that real GDP growth in 2008 will be between 4.5% and 5.5%.

Similar to other Asian countries such as China and India, the Thais highly value education. The literacy rate in Thailand is quite high. According to the national 2000 census, 92.6% of the population aged 15 years and above (94.9% male and 90.5% female) can read and write. In recent years, there has been an increased emphasis on education. The development of the kingdom's human resources is its highest priority.

The modern school system was initiated by His Majesty King Chulalongkorn or King Rama V. The first public school for commoners was set up in 1884, followed by the establishment of Ministry of Education in 1892, and a reorganization of the school system with the proclamation of early form of the national education plan in 1898. During the reign of King Rama VI, a law on compulsory primary education was issued in 1921, requiring every 7-year-old child to receive free primary education until the age of 14 years; in addition, the first university, Chulalongkorn University, was established in 1917.

With drastic social changes in the age of globalization, the prime concern of the government is on educational reform that will prepare and strengthen Thai people to cope with such changes, move through the crisis, and contribute to the social and economic development of the country. It is necessary that human development focuses on human worth, potential, morality, and ethics as the main aspects in moving toward a knowledge-based, and a peaceful and caring, society.

The first successful attempt of comprehensive educational reform was the inclusion of various provisions relating to education in the 1997 Constitution, followed by the promulgation of the 1999 National Education Act and Amendments (Second National Education Act BE 2545 (2000) and the preparation of the National Education Plan (2002–2016)) as mandated by Section 33 of the 1999 National Education Act.

Framework and Goals of the Education System

“Education Builds the Nation, Empowers the Individual, and Generates Employment” has been the principle which the Thai government adopts as a means of educational reform movement since the 8th National Economic and Social Development Plan (1997–2001) until the present 10th Plan (2007–11). Toward this end, education policy is integrated with the overall policy on social development and is called “Policy on quality human and societal development: Building a lifelong learning society.” This policy also emphasizes improvement of the country’s human resources in terms of knowledge, morality and ethics, as well as readiness to respond to the measures required to ensure the nation’s development and competitiveness.

The main goals of educational reform reflected in the principles and challenging guidelines of the 1997 Constitution, the present 2007 Constitution, and the 1999 National Education Act and Amendments (Second National Education Act 2000) are better equity and quality in education, and increasing the quality of life of Thai citizens.

Constitution of the Kingdom of Thailand

The Constitution BE 2540 (1997) has greatly increased the rights of Thai people to political participation and to voice public opinion on major problems. Relating to education, Section 81 of the constitution required that the state:

Provide education to attain knowledge and morality; issue laws relating to national education; improve education so as to be attuned to economic and social change; create and strengthen knowledge and inculcate sound awareness of politics and a democratic system of government under a constitutional monarchy; promote research in various disciplines; accelerate the application of science and technology for national development; promote the teaching profession; and encourage the revival of local wisdom, art and culture of the nation.

Improving education to be in harmony with economic and social change means that the government is committed to initiate educational reform whenever it is

necessary to keep up with the pace of change. Besides, it is provided in the constitution, for the first time, that all Thai people will have an equal right to receive basic education, for at least 12 years, of good quality and free of charge (Section 43). The constitution ensures that all people will have both the right and duty to receive education and training (Sections 30 and 69) as well as academic freedom (Section 42). It also includes the right to receive care and education for children, youth, women, the elderly, the underprivileged, and the handicapped as provided in Sections 53, 55, and 80. These provisions will protect the right to education of all Thai people, thereby moving forward toward a knowledge-based economy and society.

In providing education, maximum public benefit in national communication resources (Section 40) and the conservation and restoration of local wisdom (Section 46) will be taken into consideration. The role of the private sector in the provision of education at all levels is also emphasized (Section 43). It ensures the right of local organizations to participate in the provision of education which will promote the decentralization of educational management (Section 289). Furthermore, participation of local people and communities in educational provision will be enhanced, which will make education both relevant to the needs of the people and responsive to changing environments, demands, and opportunities at a local level.

These challenging guidelines mandated by the 1997 Constitution have been assured in Thailand’s 2007 Constitution, which has just been promulgated in August. Education rights and liberties are assured in this latest constitution. It emphasizes that a person shall enjoy an equal right to receive basic education for the duration of not less than 12 years which shall be provided by the state thoroughly, up to the quality, and without charge. Indigent, disabled, and handicapped persons are eligible for assistance and support from the state to enjoy equal education opportunities. Moreover, education and vocational training provided by private organizations, educational alternatives, self-education, and education for life shall be protected and promoted by the state (Section 48). As for academic freedom, it states that education, training, learning, teaching, researching, and disseminating such research according to academic principles shall be protected, provided that it is not contrary to his/her civic duties or good morals (Section 49).

The 1999 National Education Act

In order to meet the requirements of Section 81 of the 1997 Constitution, the first National Education Act was promulgated in August BE 2542 (1999), followed by its amendments in BE 2545 (2000). The act serves as the fundamental law for the administration and provision of education and training. It includes nine chapters prescribing the objectives

and principles; educational rights and duties; educational system; national education guidelines; educational administration and management; teachers, faculty staff, and educational personnel; resources and investment for education; and technologies for education.

The framework of education reform included (1) learning reform toward learner-oriented education; (2) administrative reform aiming at adjusting administrative structure to promote decentralized and school-based management; reorganizing systems for teachers, faculty staff, and educational personnel to upgrade the teaching profession; and increasing efficiency in the utilization of resources and investment for educational purposes; and (3) legal reform by preparing and/or amending a range of legislations and regulations according to the act.

The ultimate goals of education are to foster full development of the Thai people in all aspects: physical and mental health, intellect, knowledge, morality, integrity, and a desirable way of life so as to live in harmony with other people (Section 6). Various measures are now undertaken to make sure that morality development be inculcated in the young. In her 2005 lecture on “Education is the golden key,” Her Royal Highness Princess Maha Chakri Sirindhorn also emphasized that:

One important part of education is to help a person realize the importance of virtues and righteousness, so that you will not commit wrongdoings. . . . The expectation of good education is intellectual development and mental exercise.

As for the provision of education, according to the National Education Act, it shall be based on three principles, namely, participation by all segments of society, continuous development of the bodies of knowledge and learning processes, and lifelong education for all (Section 8).

The National Education Plan (2002–16)

As mandated by Section 33 of the 1999 National Education Act, a 15-year National Education Plan was prepared by the Office of the Educational Council to serve as a framework for formulating the development plans pertaining to basic education, vocational education, higher education, religion, and art and culture as well as to provide guidelines for formulating operational plans at the levels of educational service areas and educational institutions. This long-term reform plan brought together the relevant provisions of the constitution, the National Education Act, and the government policy. It centers on the integration of all aspects of the quality of life, including comprehensive and balanced human development, and the forging of a society of morality, wisdom, and learning. Authorized for subsequent implementation by the Council of Ministers on 17 June 2002, the plan stipulates three objectives: all-round and balanced human development; building a society of morality, wisdom, and learning; and development of social environment (see Figure 1).

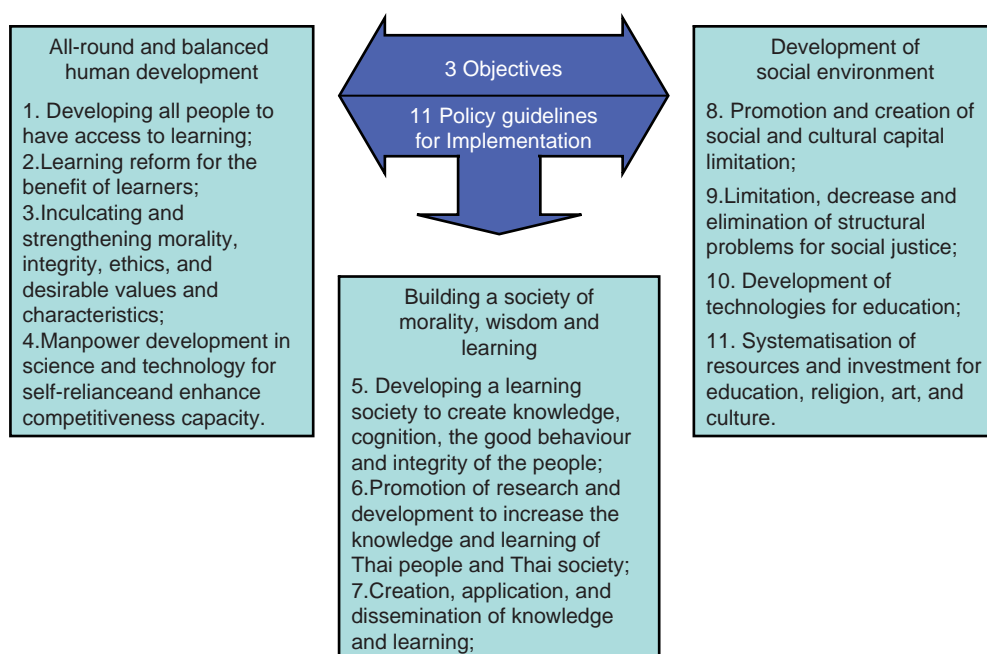


Figure 1 Objectives and policy guidelines of the National Education Plan. From Office of the Education Council, Ministry of Education (2004). *Education in Thailand 2004*. Bangkok: Amarin Printing and Publishing. With kind permission from Amarin Printing and Publishing.

Structure and Operation of the Education System

The 1999 National Education Act classifies education approaches into formal, nonformal, and informal. Formal education, comprising basic and higher education, is primarily provided in the school, while nonformal approaches are for all those outside schools and informal activities are available for self-learning through various sources. However, all types of education can be provided by educational institutions as well as learning centers organized by individuals, families, communities, community or private groups, local administration organizations, professional bodies, religious institutions, welfare institutions; and other social institutions.

The school system at basic education level is 6:3:3, covering preprimary education, 6 years of primary (level 1 or grade 1–3), 3 years of lower secondary (level 2 or grade 4–6), and 3 years of upper secondary education (level 4 or grades 10–12). It is provided by early childhood development institutions, schools, and learning centers (Figure 2).

The 12 years of free basic education (grades 1–12) were made available to students throughout the country for the first time in 2002. The government later increased that number to 14 years to include 2 years of preprimary education. On 25 March 2008, the Council of Ministers approved additional budget to provide 14 years of free

basic education. This measure is expected to begin in May or the first semester of the academic year 2008.

In the academic year 2007, as of 31 December 2007, the total number of formal educational institutions at both basic and higher educational levels are 38 509 (1524 in Bangkok and 36 985 in provincial and local areas). In addition, it was reported that, in that year, there were 14 467 833 students and 683 538 teachers and faculty members (Table 1).

Preprimary Education and Early Childhood Care

Harmonious physical, intellectual, emotional, and social development of children in the 0–5 years age group is highly recognized by the government. Preprimary education programs are provided by early childhood development institutions, day-care centers, nursery schools, or kindergarten and learning centers. However, most children in the range 3–5 years of age receive preprimary education through noncompulsory basic education services. In 2004, Office of the Education Council (OEC), Ministry of Education, reported that only 17.5% of the 3–5 years age group were under the full-time care of their families, while the majority, or 82.5%, participated in child development centers, kindergartens, and preschool classes. In the 0–3 years age group, 97.7% were cared for by their families, with the remainder, or 2.3 %, attending nurseries.

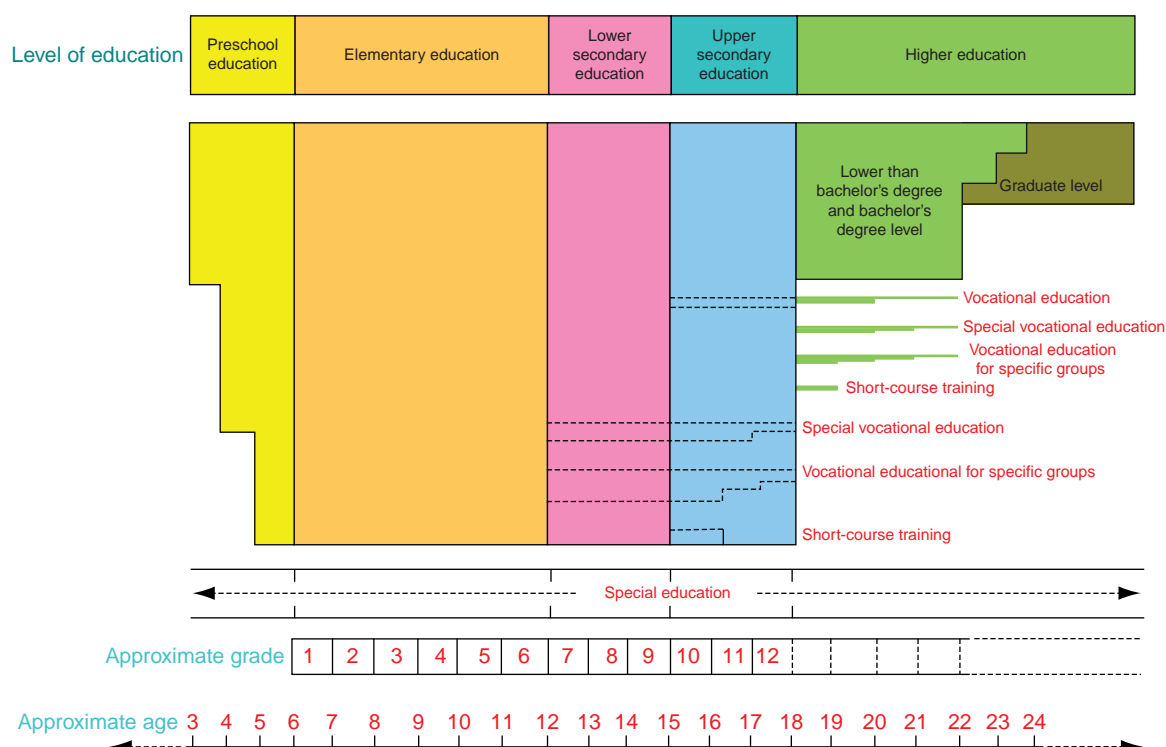


Figure 2 Present education system. From Office of the Education Council, Ministry of Education (2006). *Education in Thailand 2005/2006*. Bangkok: Amarin Printing and Publishing. With kind permission from Amarin Printing and Publishing.

Table 1 Number of hits in Google Scholar in March 2009 for various types of diversity in teacher education since 2004

Jurisdiction	Institutions			Teachers			Students		
	Total	Bangkok metropolis	Other provinces	Total	Bangkok metropolis	Other provinces	Total	Male	Female
Grand total	38 509	1524	36 985	683 538	88 157	595 381	14 467 833	7 188 319	7 279 514
Ministry of Education	36 694	1059	35 635	634 981	71 464	563 517	13 529 562	6 686 701	6 842 861
1. Office of the Permanent Secretary (Office of the Private Education Commission)	3887	836	3051	125 803	31 643	94 160	2 405 305	1 180 050	1 225 255
2. Office of the Basic Education Commission	32 256	154	32 102	419 911	13 088	406 823	8 346 591	4 166 498	4 180 093
3. Office of Vocational Education Commission	404	21	383	27 822	1606	26 216	687 630	412 275	275 355
4. Office of Higher Education Commission	146	48	98	61 330	25 127	36 203	2 089 320	927 457	1 161 863
5. Organization under the supervision of the Ministry of Education (Mahidol Wittayanusorn School)	1	-	1	115	-	115	716	421	295
Other organizations	1815	465	1350	48 557	16 693	31 864	938 271	501 618	436 653
1. Ministry of Interior (Dept. of Local Administration)	691	-	691	23 639	-	23 639	464 295	236 465	227 830
2. Ministry of Social Development and Human Security (Dept. of Social Development and Welfare)	3	1	2	46	-	46	478	243	235
3. Bangkok Metropolitan Administration	437	437	-	14 920	14 920	-	342 520	176 700	165 820
3.1 Department of Education	435	435	-	14 609	14 609	-	341 555	176 507	165 048
3.2 Department of Health	2	2	-	311	311	-	965	193	772
4. Ministry of Public Health (Praborommarajchanok Institute)	37	2	35	2135	-	2135	19 678	2432	17 246
5. Ministry of Transport	2	1	1	136	31	105	2811	1945	866
5.1 The Merchant Marine Training Center	1	-	1	105	-	105	1003	1003	-
5.2 The Civil Aviation Training Center	1	1	-	31	31	-	1808	942	866
6. Ministry of Defense (The Armed Forces Academies Preparatory School)	13	10	3	2129	1580	549	7762	6929	833
7. Ministry of Culture (Bunditpatanasilpa Institute)	16	3	13	951	162	789	10 337	3704	6633
8. Ministry of Tourism and Sports (Sport schools)	28	1	27	-	-	-	12 490	9167	3323
9. The Bureau of National Buddhism	396	10	386	2782	-	2782	48 220	48 220	-
10. Royal Thai Police	192	-	192	1819	-	1819	29 680	15 813	13 867
10.1 Border Patrol Police General Headquarters	191	-	191	1720	-	1720	28 718	14 851	13 867
10.2 Police Cadet Academy	1	-	1	99	-	99	962	962	-

From Office of the Permanent Secretary, Ministry of Education (2007). *Education Statistics: Academic Year 2007*. Bangkok: Ministry of Education. With kind permission from Ministry of Education.

According to the Ministry of Education, 1 754 328 pre-school children enrolled in both public and private institutions in academic year 2007. Among these, 1 591 410 were in institutions under the auspices of Ministry of Education and the remaining 162 918 were in centers of other government organizations such as Ministry of Interior, Ministry of Social Development and Human Security, the Bangkok Metropolitan Administration, and the Border Patrol Police.

Primary and Secondary Education

The compulsory education requirement covers 6 years of primary and 3 years of lower secondary education. Children are expected to be enrolled in basic education institutions from age 7 through the age of 16, except for those who have already completed grade 9.

The Basic Education Curriculum BE 2544 (AD 2001) consists of 12-year core curricula prescribed by the Ministry of Education. The substance of the curriculum is assembled into eight subject groups of learning processes: (1) Thai language; (2) mathematics; (3) science; (4) social studies, religion, and culture; (5) health and physical education; (6) art; (7) career and technology; and (8) foreign languages. Educational institutions are required to develop school-based curricula relating to the needs of community and society, and include local wisdom and desirable attributes for members of family, community, society, and the nation. The proportion of core curriculum to local content developed by the institution should be approximately 70:30, flexibly applied in compliance with the nature of each subject.

For several years until 2005, the percentage of students in primary education institutions had exceeded 100% because of the underage and overage student population added to the percentage of children in the 6–11 years age group. In the academic year 2007, the total number of primary school students was 5 561 768. The majority of students (4 001 470) studied in government schools under the auspices of the Office of the Basic Education Commission (OBEC), Ministry of Education. Another 1 001 969 students were in private schools; 15 034 enrolled in university demonstration schools; and 543 295 were in primary schools organized by other government agencies.

At the lower secondary education level, the percentage of students in classes compared to the population aged 11–14 years increased from 2.34 million, or 82.8%, in 2000 to 2.63 million, or 96.7%, in 2006. In the academic year 2007, the total number of lower secondary students was 2 790 190. The majority, that is, 2 634 163 students, studied in schools under the auspices of the OBEC.

The proportion of students enrolled in upper secondary classes per population aged 15–17 years during the academic years 2000–06 also increased from 57.3% in 2000 to 65.8% in 2006. According to the Ministry of Education, it

is reported that, in the academic year 2007, the total number of students in general education program was 1 174 014 and vocational education program was 773 895.

The number of children aged 3–17 years enrolled in basic education during the coming academic years is expected to increase because of the extension of free basic education from 12 to 14 years and the extension of compulsory education from 6 to 9 years. It is also very likely that students who have completed compulsory education will opt to continue on to 3 years of upper secondary instruction, and to pursue higher education at an increasing rate because the Income-Contingent Loan Fund will be available to pay for their tuition, and hence provide them with greater access.

Basic Education of Children with Special Educational Needs

The OBEC, Ministry of Education, has formulated a strategic plan for gifted and talented students. Suitable curricula, appropriate means to accelerate growth and development, a dynamic and vibrant environment, and well-trained mentors are approaches being implemented to nurture gifted children with talents for science and mathematics, languages, sports, music, computing, visual and performing arts, and many other fields. Incentives for teachers and schools, advice to parents, as well as the establishment of the National Centre for the Gifted and Talented are expected to further stimulate the development, promotion, and support of children with special talents.

Most of the socially and/or culturally disadvantaged students study in regular schools, while others study in special institutions, such as those established by the Border Patrol Police and the welfare schools supervised by the OBEC.

Nine types of disability are recognized: (1) hearing impairment; (2) mental impairment; (3) visual impairment; (4) physical or health-related impairment; (5) learning disabilities; (6) autism; (7) emotional and behavioral disorders; (8) speech and language disorders; and (9) multiple disabilities. These students receive formal education provided by the Bureau of Special Education Administration, the Department of Social Development and Public Welfare, as well as by some university demonstration schools, municipal schools, and private foundations and some hospitals which organize classes for disabled children suffering from chronic conditions. With an emphasis on individualized education programs, and commensurate with their potential, disabled students receive education from various types of schools: (1) special schools, (2) special centers, and (3) inclusive schools.

Education for Ecclesiastics

General education is also provided to novices and monks in general ecclesiastic schools in various Buddhist temples.

They offer lower and upper secondary education curricula equivalent to those provided by the OBEC. Apart from general subjects, the courses include learning units related to religious practice, Buddhist doctrine, and the Pali language. There are also two Buddhist universities in Bangkok, with various campuses elsewhere, offering courses at undergraduate and graduate levels. In the academic year 2006, there were 10 455 institutions and a total number of 1 303 605 students in Dharma and Pali sections.

Technical and Vocational Education

Thailand recognizes the important role of technical and vocational education and training as a critical tool for producing manpower with the necessary skills required for employment and/or entrepreneurship as well as for poverty alleviation. This type of education and training is provided in public and private educational institutions, enterprises, or through programs organized through cooperation between educational institutions and enterprises.

Several activities, organized by the Office of Vocational Education Commission and the Office of Higher Education Commission under the Ministry of Education, the Ministry of Labour, as well as the business sector, have been undertaken to improve the image of vocational education institutions, including improvement of the curriculum and increased collaboration between educational institutions and the production and service sectors. These include the improvement of curriculum and instruction, establishment of Thai Vocational Qualification, validation of experience, research and innovation, and career development.

Formal technical and vocational education and training is conducted at three levels: upper secondary, leading to the lower certificate of vocational education; post-secondary, leading to a diploma or the associate's degree in vocational education; and university level, leading to a degree. Under the supervision of the Office of the Vocational Education Commission, technical and vocational education is provided in 12 types of colleges: (1) technical; (2) vocational; (3) agricultural and technology; (4) commercial; (5) industrial and ship-building technology; (6) fishery; (7) administration and tourism; (8) polytechnic; (9) industrial and community; (10) automotive industry; (11) arts and crafts; and (12) the Golden Jubilee Royal Goldsmith College. At present, nine major fields of study are offered: trade and industry, agriculture, home economics, fishery, commerce and business administration, tourism, arts and crafts, textiles, and information technology and communication.

Higher education

The Office of the Higher Education Commission, Ministry of Education, currently supervises 88 limited admission

public universities, which include 38 public universities/institutes, autonomous universities, and open universities, 41 Rajabhat universities or former teachers colleges, nine Rajamangala universities of technology, 68 private universities and colleges, and 17 community colleges. Other ministries and agencies also organize 94 specialized institutions. The University Council of each institution is authorized to approve the offering of additional programs so long as the curriculum is in line with the criteria specified by the Board of Higher Education Committee. The Office of the Higher Education Commission acknowledges and reviews the council-approved curriculum.

The demand for higher education increases constantly. Student enrolment at diploma, bachelor, and graduate degree levels in higher education institutions, including that in open universities, rose continuously from 1 797 840 in 2000 to 2 993 499 in 2004. Compared to the population aged 18–21 years, the percentage of students in higher education institutions was 68.05% in the academic year 2006.

Systematic and continual improvement of higher education is essential in order to meet the rising demand of students and enable institutions to produce graduates of high caliber in accordance with the requirements of the country's social and economic development needs, and to serve as centers of excellence for innovation, research, and development. The First 15-Year Plan for Higher Education (1992–2007) identified five key areas of competencies, namely, pursuit of excellence, equality in educational opportunity, administrative efficiency, internationalization of higher education, and flexible management and promoting the role of the private sector. Consequently, the main goal of the Second 15-Year Plan (2008–2022) is the system of high quality through mechanisms and measures of good governance, financing instrument, higher education standards, and university networking; the foundation to this is university academic freedom, diversity, and unity of the system.

The admission system was reviewed and the newly modified university admission system can be summed up as (1) the direct university admission system, (2) the central university admission system, and (3) special programs and the quota system. Beginning academic year 2006, students completing grade 12 are obliged to take national educational tests of two types: the Ordinary National Educational Test (O-NET) and the Advanced National Educational Test (A-NET) administered by the National Institute of Educational Testing Service.

Continuous effort has been made to encourage public higher education institutions to transform to autonomous universities so that they can generate innovation, cost effectiveness, accountability, performance-based assessment, good governance, and long-term social and economic development. In addition, a pilot project of the Thailand Cyber University was initiated in early 2005.

Research and development in universities is strongly promoted in response to the government policy to enhance Thailand's competitiveness as well as to generate new products and innovations for commercial purposes.

Adult and Nonformal Education

Nonformal education services such as fundamental education for literacy, general education, technical and vocational education and training, family-based early childhood development programs, and activities for the improvement of quality of life are provided by both public and private bodies. Target groups of the Office of the Non-formal Education Commission primarily are those outside the school system, that is, the early childhood population, school-age population who have missed formal schooling, and the over-school-age population; however, services have now been expanded to include specific target groups such as the hill tribes, prison inmates, the disabled, the aged, farmers, local leaders, slum dwellers, and those having no opportunity to further their studies in formal schooling after compulsory education. In the academic year 2006, the enrolment number of all non-formal programs was 1 840 333 and the graduate number was 852 007.

Lifelong learning sources such as community learning centers and village reading centers are promoted, in accordance with Section 25 of the National Education Act, so that they will be in sufficient number and with efficient functioning. Several ministries are involved in providing informal education through information dissemination, educational activities, or academic and professional programs for different target groups relating to the responsibilities of each organization. It is noteworthy that nonformal and informal education services are provided to promote and support lifelong learning activities based on active participation from all segments of society.

The Teaching Profession

Chapter 7 of the National Education Act mandated that a system for teachers and educational personnel, including production and further refinement of personnel, shall be promoted so that teaching will be further enhanced and become a highly respected profession. Accordingly, the preservice curriculum was extended from 4 to 5 years, with the first 4 years dedicated to coursework and a final year devoted to teaching practice at an approved school. Besides a bachelor's degree in education, a professional teaching license is required of in-service teachers and school directors. College graduates with a bachelor's degree in fields other than education must complete a

1-year graduate certificate program in education before being certified.

The Teachers' Council has established a set of standards and a code of ethics for educational professionals, in line with the 2003 Teachers and Education Personnel Council Act, which specifies that professional standards be comprised of professional knowledge and experience, performance, and conduct. The National Institute for Development of Teachers and Educational Personnel was established as an independent organization for teacher development.

Several measures have been undertaken to attract qualified individuals to the teaching profession and retain them. Special salary scales have been introduced and a strategic plan has been formulated to solve the problem of teacher indebtedness. Contributions of outstanding teachers are recognized by organizations such as the Teachers' Council of Thailand, the Association of Science Teachers, and the OEC. OEC has selected outstanding teachers under three categories: national teachers, master teachers, and teachers of Thai wisdom.

In the academic year 2007, the total number of teachers and faculty members was 683 538. Among these, 634 981 are employed in educational institutions under the auspices of the Ministry of Education and another 48 557 are employed by other ministries.

Administrative and Supervisory Structure

Educational administration and management in Thailand is conducted by the state, local administration organizations, and the private sector. The Ministry of Education is responsible for promoting and overseeing all levels and types of education under the administration of the state. Overseeing administration, management, and monitoring the quality and standards of private educational institutions are also responsibilities of the state. However, local education administration is under the supervision of the Ministry of Interior. In addition, other ministries undertake management of education in specialized fields or for specific purposes (**Figure 3**).

Under the principle of decentralization, the main functions of the Ministry of Education at the central level are formulating policies, plans, and standards; support of resources; and monitoring and evaluation. There are five main bodies under the Ministry: the Office of the Permanent Secretary; the OEC; the OBEC; the Office of the Vocational Education Commission; and the Office of the Higher Education Commission. The ministry decentralizes powers in educational administration and management regarding academic matters, budget, personnel and general affairs administration directly to educational service areas and educational institutions in the areas.

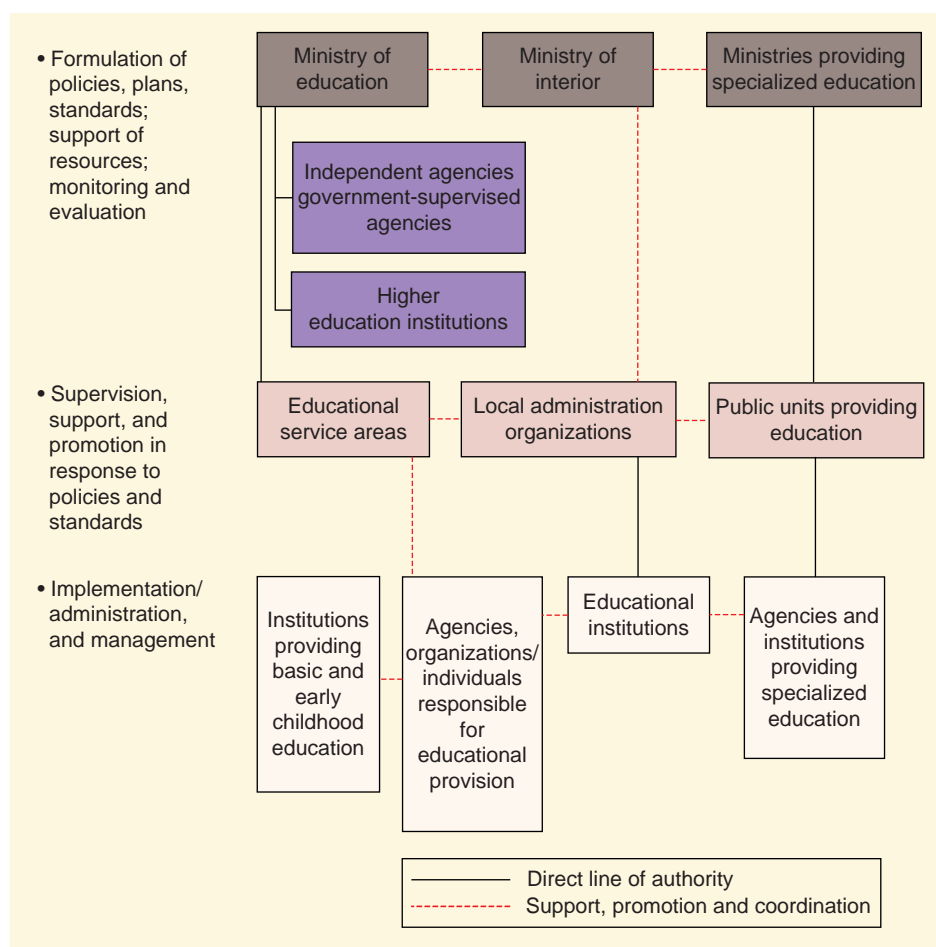


Figure 3 Educational administration and management structure. From Office of the Education Council, Ministry of Education (2006). *Education in Thailand 2005/2006*. Bangkok: Amarin Printing and Publishing. With kind permission from Amarin Printing and Publishing.

The OBEC has established 175 educational service areas with 172 areas in the provinces and the remaining three in Bangkok. At present, each educational service area comprises an area committee for education, with its office responsible for approximately 200 educational institutions and a student population of 300 000–500 000. Educational service areas are responsible for supervision, support, and promotion in response to policies and standards.

Administration and management of basic educational institutions are supervised and supported by a board consisting of representatives of parents, teachers, community groups, local administration organizations, alumni, and academicians. State universities are moving toward transformation to state-supervised institutions that function as legal entities which will enable each institution to develop its own administration and management system with greater flexibility and academic freedom under the supervision of the institutional council empowered by its own act.

Local administration organizations can provide education services at any or all levels commensurate with their

readiness, suitability, and the requirements of the local area. The Ministry of Education prescribes criteria and procedures for assessing readiness to provide education services, and assists in enhancing their capability in line with the policies and required standards. Additionally, the ministry advises on the budgetary allocations provided by local administration organizations.

Educational Finance

The current reform proposal encourages greater support for education from local resources. However, the major source of local funding for education still comes from the central government.

The Thai government has, for the last decade, allotted a generous proportion of the national budget to the education sector. The share of education as a proportion of the GDP has remained fairly constant, declining only slightly from 3.9% in 1997 to 3.7% in 2006. Over the past decade, the education share of the budget was largest

in the year 2000 at 25.7% and declined to 21.7% (Bht. 295 million) in 2006. The largest proportion of educational funding in 2006 (43.6%) has been allocated to preprimary and primary education. Secondary education and higher education received 26.6% and 15.3%, respectively, while only 1.3% of the total educational budget was allocated to nonformal education.

In 2004, two laws were enacted to increase private sector participation in educational provision, one exempting income tax and value-added tax on imported learning materials, and the other granting property tax exemptions.

Educational Standards and Quality Assurance

In order to ensure improvement of educational quality and standards at all levels, the 1999 National Education Act mandates, for the first time, that there shall be a system of quality assurance. Educational institutions follow guidelines for internal quality assurance standards developed by their supervising agencies. The Office for National Education Standards and Quality Assessment (ONESQA), established as a public organization in November 2000, employs the Amicable Assessment Model, to oversee external quality assessment at least once every 5 years.

Basic, vocational, and higher education institutions are assessed following standards relating to educational achievement (output/outcome); input/processes; and efficiency in administration and leadership. Moreover, on 26 October 2004, the Council of Ministers approved national education standards which will also serve as the basis for setting assessment standards of internal and external quality assurance mechanisms.

The national education standards comprised of three categories: (1) desirable characteristics of the Thai people, as both citizens of the country and members of the world community; (2) guidelines for educational provision; and (3) guidelines for creating a learning society/knowledge society.

Indicators for desirable characteristics consist of sound physical and mental health; required knowledge and skills sufficient for leading a meaningful life and social development; skills in learning and self-adjustment; social skills; and righteousness, public-mindedness, and consciousness of their citizenship of Thailand and the world.

The guidelines for educational provision consist of three indicators: development of a diversified curricula and ambiance enabling learners to develop themselves in line with their natural inclinations and to the best of their potential; systematic and effective development of administrators, teachers, faculty staff, and education personnel; and practice of school-based management.

The three indicators of guidelines for creating a learning society/knowledge society are provision of academic services and establishment of cooperation between educational institutions and community so as to transform educational institutions into a learning society/knowledge society; research and study, promotion of and support for learning sources and mechanisms; and generation and management of knowledge for the benefit of all levels and components of the society.

Concluding Remarks

The government has been supporting educational reform with the aim to prepare and strengthen Thai people to cope with rapid changes in the globalization age and to develop a knowledge-based society as a prerequisite for a knowledge-based economy. The first 1999 National Education Act was promulgated as a legal framework for education. Based on the principle of education for all and all for education, the major aspiration in the reform is to provide quality education and equal access to lifelong education and training. Participation from all stakeholders and support from government policy on education are, inevitably, the key success factors in pushing forward educational reform.

Further Reading

- Ministry of Education (2002). *Basic Education Curriculum B.E. 2544 (A.D. 2001)*. Bangkok: Department of Curriculum and Instructional Development.
- Office of the Commission on Higher Education, Ministry of Education (2007). *Framework of the Second 15-Year Long Range Plan on Higher Education (2008–2022)*. Bangkok: Office of the Commission on Higher Education.
- Office of the Education Council, Ministry of Education (2003). *Strategies and Roadmap for Higher Education Reform in Thailand*. Bangkok: Office of the Education Council.
- Office of the Education Council, Ministry of Education (2004). *Education in Thailand 2004*. Bangkok: Amarin Printing and Publishing.
- Office of the Education Council, Ministry of Education (2006). *Education in Thailand 2005/2006*. Bangkok: Amarin Printing and Publishing.
- Office of the Education Council, Ministry of Education (2006). H.R.H. Princess Maha Chakri Sirindhorn, Lecture on "Education is the Golden Key", the American International Club of Geneva, December 14, 2005.
- Office of the Education Council, Ministry of Education (2007). *Framework of Educational Development in the Tenth National Economic and Social Development Plan (B.E. 2550–2554) in Line with the National Education Plan (B.E. 2545–2559)*. Bangkok: Office of the Education Council.
- Office of the National Education Commission, Office of the Prime Minister (2003). *National Education Act B.E. 2542. (1999) and Amendments (Second National Education Act B.E. 2545 (2002))*. Bangkok: Pimdeekampim Co. Ltd.
- Office of the Permanent Secretary, Ministry of Education (2007). *Education Statistics: Academic Year 2007*. Bangkok: Ministry of Education.

Relevant Websites

<http://www.obec.go.th/> – Office of the Basic Education Commission.
<http://www.onec.go.th/> – Office of the Educational Council.
<http://www.mua.go.th/> – Office of the Higher Education Commission.
<http://www.nesdb.go.th/> – Office of the National Economic and Social Board.

<http://www.onesqa.or.th/> – Office for National Education Standards and Quality Assessment.
<http://www.thaigov.go.th/> – Royal Thai Government.
<http://www.ksp.or.th/> – The Teacher's Council of Thailand.
<http://www.moe.go.th/> – The Ministry of Education in Thailand.
<http://www.vec.go.th/> – Vocational Education Commission.

The Gambia

Y A-M Jobe, Gambia National Commission for UNESCO, Banjul, Gambia

© 2010 Elsevier Ltd. All rights reserved.

Background Information

The Gambia was a former British colony used as a port of transiting slaves from the West African subregion to Europe and other continents during the transatlantic slave trade. It acceded to independence in 1965 and to the status of Republic in 1970.

Physical and Climatic Characteristics

The Gambia is a small Sudano-Sahelian country in West Africa, lying between latitudes 13.0° and 13.7° North and longitudes 13.7° and 16.0° West. It consists of a narrow strip of land some 400 km long, and from 30 km wide in the west to 10 km wide in the east, on either side of the River Gambia. It is bordered on three sides by Senegal and on the west by the Atlantic Ocean. The country has a subtropical climate, typified by a long dry season (November to June) and a short wet season (July to October). Rainfall distribution is irregular and ranges from 800 mm a year inland to 2200 mm in the coastal areas. Average temperatures range from 18°C to 30°C during the dry season and 23°C to 33°C during the wet season.

Demographic Characteristics

The Gambia has a human population of some 1.4 million people (Gambia Bureau of Statistics (GBoS) 2003 census report), with an annual growth rate of 2.7% and a density of 127 km⁻² – one of the highest in Africa. Compared to the 4.2% growth rate in the 1993 census, this shows a decline of 1.5 points over the decade, albeit still higher than the 2.5% average for sub-Saharan Africa.

In addition, the 2003 census data shows that 42% of the population is under the age of 15 years, and the economic dependency ratio is 1.22, meaning every 100 persons working takes care of another 122 persons for their livelihood. Similarly, the youth (15–24 years) and adults (15 years and over) constitute 22% and 57% of the population, respectively (GBoS, 2003).

The population distribution in the Gambia is highly uneven with much of the population concentrated around the urban areas, about 50.4% compared to 49.6% in rural areas (GBoS, 2003). This is evident in the density of 2867 and 4272 persons per square kilometer in urban areas such as Banjul and Kanifing, and 45 and 77 in rural areas such as Mansakonko and Kerewan administrative regions, respectively (GBoS, 2003). At the national

level, population density averaged about 127 persons per square kilometer in 2003, indicating an increase of 31 persons per square kilometer from a 1993 density of 97 persons per square kilometer. The implications of these high population densities, especially in the urban area, calls for both expansion and diversification of educational facilities and programs to cater to the growing demand for educational services in the country.

Economic Characters

The Gambia has a very narrow economy with no important mineral or other natural resources and has a limited agricultural base. About 47% of the population depends on crops and livestock for their livelihood (GBoS, 2003). Small-scale manufacturing activity features the processing of groundnuts, fish, and hides. Agriculture accounts for about 27% of the gross domestic product (GDP) while industry and services account for 10.7% and 62.9% respectively (GBoS, 2007).

According to the 1993 census, approximately 51% of the population was categorized as economically active. Only about 32% of the population was classified as employed, although a substantial part of the remainder of the population was occupied in subsistence agriculture. The age structure of the population suggests that there is a high dependency ratio: those in employment support a large number of young children. Of those employed, some 82% are engaged in agriculture, fishing, and related activities; 8% in industry and related activities; and the remainder in the service sector. Eighty-six percent are employed in the informal sector and 14% in the formal sector of the economy, with 77% of the latter working in the Greater Banjul Area.

Sociocultural Characteristics

The Gambia is a multicultural society with a variety of ethnic and linguistic groups composing the nation, namely *Mandinka/Jabanka* (35.9%), *Wolof* (14.5%), *Fula/Tuklor* (21.9%), *Sarabule* (8.2%), *Fola/Karoninka* (11.4%), *Serere* (3.1%), *Manjago* (2.0%), *Bambara* (1.1%), *Creole/Aku* (0.5%), others (1.4%) and ethnicity not stated (0.2%). The main religions are Islam and Christianity, with minority following traditional African religions. Muslims constitute about 90% of the population, Christians about 8%, while other traditional believers account for the rest (GBoS, 2003). Despite the ethnic and religious diversities, people of different

ethnic or religious groups have always lived together happily in a social framework characterized by mutual respect, trust, and tolerance toward one another since independence. This social setup has greatly enhanced the delivery of educational service throughout the country, in that ethnic groups live together in many settlements and communities.

Goals of the Education System

The National Education Policy (NEP) 2004–15 and the Education Sector Strategic Plan (ESSP) 2006–15, from which this medium-term plan has been developed, is tied to The Gambia's Vision 2020; that is:

to transform the Gambia into a financial center, a tourist paradise, a trading, export-oriented, agricultural and manufacturing nation, thriving on free market policies and a vibrant private sector, sustained by a well-educated, trained, skilled, healthy, self-reliant and enterprising population, and guaranteeing a well-balanced eco-system and a decent standard of living for one and all, under a system of government based on the consent of the citizenry.

Thus, the NEP is grounded in the overall goal of ensuring that by 2015, universal access to relevant and high-quality education has been achieved. Another guiding precept is the mission statement of the Department of State for Basic and Secondary Education (DoSBSE), which seeks to:

- provide access to relevant and high-quality basic education for all,
- provide high-quality education services,
- ensure gender equity in education,
- provide relevant life skills, and
- promote the principle of lifelong learning.

To situate education within the national context, the major goals pursued by the basic and secondary education sector are closely linked to the Poverty Reduction Strategy Paper (PRSP); hence, education provision in the Gambia aims at reducing poverty. It is understandable, therefore, that the sector is guided by the following motto: 'A Provision of Responsive, Relevant and Quality Education for All Gambians for Poverty Reduction.'

While the legal basis for education-service delivery responds to upholding the right of everybody to quality basic education, as contained in the 1997 Constitution of the Gambia, the Education for All (EFA) goals and the education-related Millennium Development Goals (MDGs) continue to mark such service delivery. Therefore, these instruments have been fed into the current NEP (2004–15), which clearly articulates the basic aims and objectives of education.

In pursuance of the education sector's goals, aims, and objectives, several key areas have been prioritized. These

include access; quality education; teacher education, deployment, and utilization; technical and vocational education and training (tvvet); tertiary and higher education and sector management.

Structure of the Education System

The government's strategy is to develop the industrial and infrastructural subsector comprising manufacture, transport and communications, and energy. For the education sector, the strategy implies an emphasis on basic education (grades 1–9), increased access and quality; expansion of senior secondary education (grades 10–12) to ensure up to 50% transition from grade 9 to 10; skills training; the development of scientific and technological competencies; improved acquisition of numeracy and literacy skills; and the establishment of the University of the Gambia initially for a limited number of students in some critical disciplines to fill the dearth–capacity gaps in the labor force.

The DoSBSE in the Gambia has the sole responsibility for the provision of the full range of educational services. The formal system consists of 6 years of lower basic, 3 years of upper basic, and 3 years of senior secondary schooling followed by 3 years of continuing education at tertiary institutions or 4 years at the university level. The first 9 years of education constitute the basic education cycle, which is mainly provided by government while the senior secondary, technical and vocational, and tertiary and university education are funded largely through the grant-in aid arrangement and the private sector.

Basic Education

The Gambia's NEP (2004–15) provides for a unified basic education system, covering years 1–9, through an automatic transition with no transition examination at the end of the lower-basic cycle. Thus, there is a rapid expansion taking place at the upper-basic level, and a policy of integrating basic education facilities through the creation of basic-cycle schools and stand-alone upper-basic schools. (These are schools that combine within the same structures the lower and upper-basic components (grade 1–9) and administered by a single headteacher.)

The overall basic-education program consists of early childhood development, basic education (grades 1–9) and adult and nonformal education.

Early Childhood Care and Development

Under the expanded vision of basic education, preschool education is provided to children between the ages of 3 and 6. Until 1995, the number of preschool centers registered was 125, found mainly in the capital city, Banjul,

and the immediate surroundings. However, by 1998, this number had significantly increased to 265.

According to The Gambia's EFA 2000 assessment, early childhood care and development (ECCD) is entirely provided by the private sector, while the role of government continues to be one of coordination and supervision of service delivery. From the same study too, it was reported that only 46% of urban children and less than 17% of rural children have access to ECCD. This is mainly due to the high fees and parental inability to pay the fees charged in preschools.

The impact of the expanded vision of basic education is expected to raise the admission rate of 39.4% of 3-year-olds, which represented 40 124 children in 2006, to 45% by 2011. The enrolment output of this is expected to be 91 076.

Lower-Basic Education

At the level of the lower basic, enrolments for the period 2001/02 to 2006/07 increased from 157 544 to 220 423 registering a gross enrolment ratio (GER) increase from 82% to 92%, taking into account the Madrassa enrolment, which formed 10% and 15% of the total enrolment in 2001/02 and 2006/07, respectively. (This includes only the officially recognized Madrassas.) During the same period, the GER for boys showed a decrease from 85% in 2001/02 to 82% in 2004/05 but a slight increase to 92% in 2006/07, while the GER for girls registered an appreciable increase from 80% to 95%.

Upper-Basic School

In the area of upper-basic education, there has been rapid expansion between 2002 and 2006 with enrolment increasing from 42 094 to 68 024 translating into a GER growing from 43% to 65%. This growth in enrolment represents an annual growth rate of 15%, which exceeds the target of 12.7%. However, the period under review witnessed a drop in the GER from 65% to 61%, with the enrolment for girls dropping slightly from 32 597 in 2005/2006 to 32 047 in 2006/07, while that of boys increased from 33 433 to 34 442.

Madrassa Education

The Madrassa (Islamic Arabic Schools) is a formal education-delivery institution with Arabic as a medium of instruction and a strong Islamic orientation in content and practice. There are about 149 registered Madrassas, double the number initially planned. Between 2004 and 2006/07, Madrassas accounted for 65% of the enrolment increase in primary schools, and the Madrassas now account for an estimated 16% of lower-basic-school enrolment, up from 10% a decade earlier, and the majority of this increase is in Madrassas that have synchronized with the national curriculum.

Adult and Nonformal Education

According to the census report, 2003, cited in The Gambia CONFINTEA VI Report 2008 (GBoS, 2003), the literacy rate in the Gambia for the population aged 10 years and over is estimated at 52.1%, 40% for female, and 64% for males.

On the literacy levels for the official age cohorts for adult literacy in the Gambia, the 15–24-year cohort captures 63%, while 15-year-and-above cohort is estimated at 42.5%, with 30.6% for women and 55.1% for men (The Gambia report for CONFINTEA VI, 2008). Although these literacy rates have fallen below the average for sub-Saharan Africa (77%), there have been appreciable levels of improvements from the 1998 levels of the 15–24-age categories for men and women, which stood at 48.5% and 25%, respectively.

As many as 13 450 women and out-of-school youths in 309 communities are literate in the local languages (*Mandinka, Pular, Wolof, Fula, and Sarabulleb*) and 21 828 women and youths trained in various skills.

Secondary Education

In the Gambia, secondary education continues to be a major problem for the education sector as most of the schools at this level are privately operated, thereby, rendering access quite difficult for children from poor families, particularly those in the peri-urban areas. Notwithstanding the fact that majority of the secondary schools are located in the urban areas (regions 1 and 2), provincial children are not left out and have experienced an upsurge of secondary schools of late, which have resulted in an increase in enrolment of children into secondary schools in these regions.

Second, although not as bad as it used to be in the past, girls still continue to lag behind at this level in terms of numbers (boys, 18 661; girls, 14 330). Tuition fees and other cost (uniform, textbooks, transportation, private tuition/classes, etc.) of secondary education, including considerable opportunity costs, represent severe barriers to access for children from poor families.

The quality of secondary education, as reflected by performance of students at the West African Senior Secondary Certificate Examination (WASSCE), leaves a lot to be desired. In 2007, 90.93%, 66.65%, and 63.82%, of candidates failed mathematics, English, and science, respectively. On average, more than half of the candidates failed all the courses. Thus, the desire to achieve postbasic education (secondary education) will continue to be an important activity for the sector. These daunting problems warranted the sector to have the following targets and objectives set in the 2008–11 education medium-term plan.

Technical Vocational Education and Training

Expenditure on vocational and technical education has remained constant in real terms, at about 3% of the recurrent budget. This represents a sharp fall since the 7% in 1990, largely explained by the absence of a comprehensive policy on vocational education and technical training for the sector.

TVET in the Gambia has a variety of forms and a range of purposes, by far the most important of which is to mobilize human resources in response to manpower needs and employment opportunities in the Gambia.

Institutionally, the main providers of TVET are the Gambia Technical Training Institute (GTTI) and a network of other institutions, which are categorized under technical institutes, vocational and skills training, with the GTTI being the lead technical institution for technician training.

During the policy period (2004–15), there are plans to institutionalize accredited postbasic education programs such as apprenticeship or on/off-the-job vocational and technical training.

Since government recognizes the importance of TVET provision in relation to sustainable postbasic and secondary education and training systems, which builds on the skills base of our people and encourages the long-term development of the country, the National Training Authority (NTA) has been set up to offer the opportunity and the platform to launch a major initiative in the expansion of TVET provision in the Gambia. Through this instrumentality, efforts are being made to design and develop a variety of national diploma and higher national diploma programs to meet the existing shortages of trained manpower in the country.

The NTA, which is private sector led, will be issuing National Vocational Qualification (NVQ) in association with educational and training establishments and employers. Awarding bodies will be established to develop nationally accredited examinations for skills areas, trades, and professions, which are based on international best practice.

The NTA's role as a link representing the transition from secondary to higher levels of education through a variety of programs and courses that could be the building blocks for higher professional-degree programs is also being strengthened and developed.

Tertiary and Higher Education

In the Gambia, the tertiary-education program refers to all postsecondary programs, including universities, colleges, teacher-training institutions, professional institutes, and other postsecondary technical training. During the 2004–15 policy period, Government of The Gambia

(GOTG) will ensure that the tertiary education subsector provides a flexible and dynamic system of education and training that will address the demands of access and equity on the one hand, and the need for quality and standards (excellence), on the other.

There are, at present, four main public tertiary institutions in the Gambia that together provide a variety of programs at the postsecondary level: the Gambia College (GC) in education, agriculture, nursing, and public health; the Gambia Technical Training Institute (GTTI) in technical and vocational education; and the Management Development Institute (MDI) in the area of management studies provide trained and skilled human resources at the middle levels of services and industry. The programs provided by these three institutions are below the bachelor's degree level. The fourth tertiary institution, the University of The Gambia (UTG), provides programs leading to only the bachelor's degree in the basic sciences, agriculture, humanities, and medicine. Together, these four institutions serve only about 5000 students.

Each of the four institutions is autonomous with its own governing council/board and academic structures. Although all four institutions rely on government support, students also contribute toward their own education through tuition fees.

There were significant increases in enrolment at the GC and the UTG between 2005 and 2007, representing 69.8% and 61.3%, respectively. During the same period, the MDI and GTTI experienced significant drops in enrolment by 34.3% and 13%, respectively. These were due to the high rate of competition for program options with newly established institutions offering either cheaper or better programs of the same kind.

Teacher Education

Although tertiary education no longer falls under the purview of this Department of State, teacher education and training still continues to be the responsibility and interest of the sector. Thus, their training mechanism and supply remain to influence our drive toward attaining access and quality education as we endeavor to achieve the MDGs and EFA targets.

The Gambia Training College significantly enhanced its intake of Primary Teachers' Certificate (PTC) and Higher Teachers' Certificate (HTC) students over the past 5 years, resulting in an increase of 1024 qualified PTC teachers and 1034 HTC in the system.

Improved Sector Management

The Department of State for Education, which has always been responsible for the management and coordination of

education programs and services in the Gambia has, with effect from the year 2007, been divided into two departments of state namely, the DoSBSE and the Department of State for Higher Education, Research, Science and Technology (DoSHERST).

Under the new arrangement, the DoSBSE will continue to be responsible for policy development, management, and coordination in the basic and secondary education subsector. In spite of the separation, DoSBSE is still the largest employer of civil servants, numbering 7689, which represents 39.8% of the civil service. (July 2007 IFMIS payroll and quoted in the study 'The Gambia Improving Civil Service Performance,' February 2008. It must also be noted that the figure does not include staff in subvented institutions.) In an attempt to improve the management of the sector, six directorates have been established at headquarters and six directorates at regional level, each headed by a director. While the headquarter directorates monitor and evaluate policy implementation across the system, the regional directorates are responsible for program implementation at the regional level.

At another level of monitoring, two separate structures, namely, the senior management team (SMT) and the coordination committee, have been established to formulate policy and monitor policy implementation, respectively, on the ground. Both committees comprise all the directors and are chaired by the permanent secretary with the secretary of state serving as a member of the SMT only. In addition to the directors and the two deputy permanent secretaries, the coordinating committee is composed of key stakeholders in the education sector. These stakeholders include, among others, nongovernmental organizations operating in the education sector, the Gambia Teachers' Union, the West African Examinations Council (WAEC), and the GC, responsible for teacher training.

Education Expenditure

The Gambia has an average per capita income of US\$ 420 (2007). It ranks 155 out of the 177 countries in the 2007 human development index. The poverty head count is 58%, with higher pockets of poverty in rural communities. The economy has a small internal market and lacks diversification.

Total Recurrent Expenditure

Government expenditure on education during the late 1990s to date reflects an increased public allocation to education of 39% in current prices (19% in real terms). The 2001 planned expenditure is an increase of 50% in current and 23% in constant terms from 1998.

This translates to 7% per annum, substantially more than the planned target of 4.3% annual growth.

Education's share of total government expenditure (excluding debt service) rose from 24% in 1998 to 25% in 2000 and 2001. As a proportion of overall recurrent spending, it varied between 14% and nearly 17%. Compared with the GDP, recurrent education spending rose steadily, from 3.2% of GDP in 1998 to 4.1% in 2001.

A large proportion of investment spending is financed through project grants or loans by external donors plus government local fund (GLF). Total Gambian public sector spending on education (current and capital) grew from 3.6% of GDP to 4.3% over the period.

The shares of expenditure allocated to the different levels of education continue to reflect the policy priorities of the sector. Allocation to basic education increased from 56.8% in 1998 to 64.1% in 2001. This is as a result of the accelerated expansion at the basic education level, particularly in upper basic education, compared to senior secondary education whose share dropped from 14.5% in 1998 to 11.0% in 2001.

Given that the education recurrent budget grew faster than expected, the growth targets set for each level of education, with the exception of technical/vocational education and adult education, were far exceeded. In particular, expenditure on basic education increased at a real annual rate of 11%, reflecting the rapid increase in enrolment. Secondary spending actually fell in real terms by 2.3% a year. Real spending on the quality and relevance component (including spending on preservice and in-service teacher training) grew by 7% a year.

While most sub-Saharan countries spend well over 90% of their recurrent budgets on salaries, leaving very little for maintenance and books, expenditures on these items at the basic-education level (including grant in aid), in the Gambia, continues to be maintained below 80% on average. Operation and maintenance receives 4.2%, teaching and learning supplies receives about 2% of the recurrent budget, while approximately 3% is allocated to examinations, and about 15% to grant in aid.

Recurrent Expenditures

The education policy 2004–15 calls for private participation in the delivery of educational services. The current national expenditure for education is projected at GMD 615 966 000. Government commitment to the sector continues to be unwaveringly strong. However, due to economic constraints and external pressures, budgetary allocations to the sector had suffered significant drops in proportionate share over the past years. The Department of State for Finance and Economic Affairs (DOSFEA) has expressed its commitment to restoring the sector's share to a minimum of 20% by the end of the medium term.

This has already been demonstrated in the allocations to the sector in the 2009 round of bilateral budget discussions, in which allocations to the education subsector increased from 14.4% (11.3% for basic and secondary education) in 2008 to 19.1% (16.2% for basic and secondary education) in 2009. This has been possible sequel to receipt of government attainment of heavily indebted poor countries (HIPC) completion point and subsequent receipt of the multilateral debt relief initiative (MDRI) support. There is, therefore, a strong basis that this increase will be sustained over the medium term.

It would be observed that government is funding 17.3% (14.3% and 3.0% for investments and recurrent costs, respectively) while the donor partners are financing 31.4% and 10.3%, respectively. The catalytic funding is projected at 15.4% of the overall program cost (11.5% and 3.8% of investment and recurrent, respectively). The funding gap thereof stands at 34.9% of investments and 6.2% of recurrent expenses, an overall funding gap of 41.1% of the program cost.

Sector Monitoring and Evaluation

In order to ensure that an monitoring and evaluation (M&E) system is fully functional, the DoSBSE will ensure that monitoring and tracking reports are produced and circulated on time to all concerned parties for possible actions. It is expected that follow-ups to the reports, amendments, or reevaluations of the situation would be considered while targets, objectives, and timeframes would be redefined to address the realities on the ground.

The M&E system for the education system has the purpose of ensuring compliance of its educational objectives and goals by means of a continuous monitoring of ongoing activities, intervention strategies, and the products to be obtained in its six components. The system will function with a set of indicators defined in numerical values showing the real progress expected by the education action plan. The impact indicators will be measured annually.

To this end, annual data collection will be jointly conducted by the Planning, Policy Analysis, Research and Budgeting Directorate (PPARBD) and Human Resource Directorate (HRD). As regards quality targets, the Standard and Quality Assurance Directorate (SQAD) will ensure that the Learning Achievement Target (LAT) is conducted annually followed by the community participatory approaches to assess school performances. HRD will continue the instituted staff-appraisal schemes to inform the training and promotion prospects of teachers.

In order to provide guidance to the implementation process for the strategies outlined above, the bimonthly SMT and coordinating committee meetings (CCMs) will be maintained and strengthened to focus more on outcomes of and results for interventions.

The institutionalized arrangements of annual sector and donor reviews will be maintained and conducted regularly.

Since civil society has much experience and a crucial role to play in identifying barriers to EFA goals, and developing policies and strategies to remove them, participation in monitoring and evaluation will be attracted from civil-society organizations. This is essential in order to foster the development of accountable, comprehensive, and flexible educational-management frameworks.

Major Changes and Issues Since 1990s

Based on both normative and summative evaluations of various interventions conducted with support from the EFAI/Fast Track Initiative (FTI) and other bilateral and multilateral donors, there is evidence to indicate that significant successes have been registered within the sector. These include expanded access to education across all levels of the school system, particularly by girls in the rural area, where, in some instances, interventions led to the unintended decline of boys' enrolment. Whereas, for instance, the GER in 1998 was 85%, it rose to 92% in 2007. Concurrently, the GER at the upper-basic level, which stood at 29% in 1998, has risen by over 30 percentage points to 65%. These achievements have ultimately contributed to the attainment of gender parity at the basic level (i.e., from grades 1 to 9). Similarly, the GER at the senior-secondary level has doubled from 15% in 1998 to 32% in 2007.

Generally, resources are more equitably distributed and better utilized for program implementation. The reviews and updates of the curriculum have also resulted in a more relevant, and therefore, more responsive curriculum to stakeholder needs. (This is in accordance with the study commissioned by the Association for the Development of Education in Africa (ADEA) for the 2008 Biennale.) The acute shortage of instructional materials for effective learning that was lamented in the past has been reversed through substantial investments in various items, including core textbooks and teachers' guides, supplementary learning materials, especially books written by Gambian writers. Added to these, the sector has registered the following achievements:

- The expanded vision of basic education, which comprises early childhood education, adult, and nonformal education, and formal schooling of 9 years (i.e., from grades 1 to 9) is now being implemented in full and stands out as a successful model in Africa.
- The partnership between the DoSBSE and the Madrassa proprietors, through the General Secretariat for Islamic/Arabic Education, continues to enhance access, while improving the quality of education provided.

The teaching of English language in the Madrassas that are being granted aided by the DoSBSE and the harmonization of the various syllabuses of the Madrassas and their synchronization with the curriculum of the conventional schools have contributed to this success.

- There is a more equitable distribution of qualified teachers, as shown by the impact reports from the regional directorates on the hardship allowances currently paid to teachers.
- The textbook rental scheme has been abolished, and the student–textbook ratio is 1:1 for the core subjects at the lower-basic level; that is, from grades 1 to 6.
- While the transition from the lower-basic (primary) level to the upper-basic level increased from 72% in 1998 to 88%, in 2005 as a result of phasing out the Primary School Leaving Certificate Examination, the transition rate from the basic level (grade 9) to the senior-secondary level (grade 10) has exceeded the 50% target. (This is as a result of children repeating and dropping out at the end of grade 6.)
- The PTC extension program for unqualified teachers, being piloted in region 5, promisingly stands to contribute to addressing teacher requirements in the school system, particularly with regard to the training of female teachers (Gender Education Unit Evaluation Report, 2007).
- Interest and public involvement in education is broadening and deepening, as evidenced by participation in the weekly televised (Education Forum), and radio, programs as well as other discussions on education. The DOSBSE continues to hold monthly *Bantabas* (town hall meetings) to discuss pertinent issues with stakeholders. Early childhood development (ECD) is no longer considered a luxury but an integral part of the basic education program.
- As evidenced by the 2006 and 2007 results of the WASSCE and The Gambia Basic Education Certificate Examination (GABECE) respectively, the performance

of students is steadily improving (analyses of examination results by the planning directorate, DoSBSE, 2006 and 2007).

- Through the cluster monitors, a more effective and decentralized monitoring mechanism has been established with effect from 2007.
- The process has begun for instituting a performance-management system (PMS). Thus, operational unit and individual plans have now been developed, leading to the profiling of the sector personnel for appraisal purposes.

Bibliography

- Department of State for Basic Education (2003). EFA Proposal for the Fast Track Initiative.
- Department of State for Basic and Secondary Education (2004). National Education Plan 2004–2015.
- Department of State for Basic and Secondary Education (2008). Education Sector 2008–2011 Medium Term Plan.
- Department of State for Basic and Secondary Education (2008). Simulation Model 2008.
- Department of State for Education (1997). Education Master Plan 1997–2006.
- Department of State for Education (2000). Gambia's EFA 2000 Assessment.
- Department of State for Education (2004). Education Policy 2004–2015.
- Department of State for Education (2006). Education Sector Strategic Plan 2006–2015.
- Gambia Bureau of Statistics (2003). National Census Report 2003.
- Gambia Bureau of Statistics (2007). National Census Report 2007.
- Republic of The Gambia (1994). Vision 200.
- Republic of The Gambia (1997). Constitution of The Gambia.
- Republic of The Gambia (2007). PRSP II Annual Progress Report.
- The Gambia (2008). National Report for CONFITEA VI 2008.

Relevant Website

<http://www.edugambia.gm> – Ministry of Education, Gambia.

The Philippines

E A P Valenzuela, Southeast Asian Ministers of Education Innovation and Technology, Quezon City, The Philippines

© 2010 Elsevier Ltd. All rights reserved.

Country Profile

The Philippines is an archipelago comprising 7107 islands and stretches from China in the North to the Indonesian archipelago in the south. Its northernmost island is located south of Taiwan (Chinese Taipei) and southernmost islands are just 24 km away from the coast of Borneo. It is divided into three main island groups: Luzon, Visayas, and Mindanao, which are geographically divided into 17 regions and 71 provinces.

The Philippines had a population of about 88.5 million as of 2007 and Metro Manila had a population of almost 11 million, the largest in the whole archipelago (NSB, 2007). The Philippines had a predominantly Christian population, 85% of which is Catholic, about 10% is Muslim, and the rest account for other religions.

The national language of the Philippines is Filipino. For the purposes of communication and instruction, however, its official languages are Filipino and, unless otherwise provided, English (Philippine Constitution, 1987). It has 171 regional dialects and about ten main languages, including Tagalog, Cebuano, Hiligaynon, Tausog, Mindanaoan, Pangasinense, Waray, Iloko, Kapampangan, and Bicol.

The climate in the Philippines is generally tropical and varies according to its topography. The southwest monsoon lasts from June to November, which brings sultry wet weather. The northwest monsoon, on the other hand, occurs in December. The average temperature is 80°F/26.6°C–90°F/32°C.

Brief Historical Account of the Philippine Educational System

The Philippine educational system shows lasting influences of the Spanish, American, and Japanese regimes. After nearly four centuries of Spanish rule, an educational system covering the elementary to collegiate levels was established through the promulgation of the Educational Decree of 1863. The decree made possible the establishment of complete secondary and collegiate levels of instructions, the provision for government supervision and control of these schools, and the establishment of teacher training institutions (Estioko, 1994).

When the Americans occupied Manila in 1898, they immediately reopened the schools. The American regime was marked by an extensive public educational system that used English as medium of instruction. The regime

attempted to unify the archipelago, to spread the ideals of democracy, and to facilitate colonial rule. During this period, the Philippine educational system, particularly the public school system, became unified and well organized, enabling every citizen to enjoy the benefits that education provides. This was also the time when the educational system was overhauled, when grade 7 was eliminated and when the double-session plan was introduced to accommodate more students (Zwanopoe, 1972).

During the brief Japanese occupation (1941–44), the Japanese redirected the political and cultural orientation of education from Western to Japanese values rooted on love for neighbor. Vocational education and the use of Niponggo as medium of instruction were also introduced. The Japanese also stressed the importance of physical education and music, particularly singing Japanese songs (San Mateo and Tangco, 1997).

After the Philippines was granted independence on 4 July 1946, the government focused on rehabilitating the educational system and on conserving the Filipino heritage. The Philippine educational system also stressed the importance of attaining socioeconomic development, nonformal education, and vocational training. The system was revolutionized to bridge the gap between manpower development and the country's industrial needs.

Today, the Philippine educational system adopts a more quantifiable measure of performance, promotes values of good citizenship, English instruction, global education, and strong technical–vocational education (TVE). It aims to address multiple and diverse challenges and support a system of capacity building and institutional streamlining (Figure 2).

Goals of the Philippine Educational System

The 1987 Philippine Constitution serves as the major legal basis of the Philippine educational system, to wit:

The state shall protect and promote the right of all citizens to quality education at all levels and shall take appropriate steps to make education accessible to all.

The Philippine Education Act of 1982 or Batas Pambansa Blg. 232 provided the impetus for the establishment and maintenance of an integrated system of education that endeavors to achieve and strengthen national unity and consciousness and to preserve, develop, and promote desirable cultural, moral, and spiritual values in a changing world.

Governance of Education

The Congressional Committee on Education (EDCOM) restructured Philippine educational system. Congress passed laws to support trifocalization of education and refocused the educational governance as follows:

1. *Basic education.* The Department of Education (DepED) was mandated by the Basic Education Act of 2001 or Republic Act (RA) No. 9155 to cover elementary, secondary, and nonformal or alternative learning system (ALS). The DepED was vested with the authority to raise standards of basic education and enhance the efficiency of the system. It comprises three bureaus – the Bureau of Elementary Education, the Bureau of Secondary Education, and the Bureau of ALS. Each bureau is headed by a director who manages and supervises its operation and implements its projects to fulfill its functions.
To ensure the implementation of the DepEd's policies throughout the country, it has 17 regional offices, one in each region, in addition to a central office in Metro Manila. A region typically covers four to eight provinces; each DepEd regional office has a secondary education division unit responsible for implementing policies in and for supervising secondary schools.
2. *Technical education.* Midlevel skills education and training falls under the purview of the Technical Education and Skills Development Authority (TESDA), which was established through the enactment of RA No. 7796, also known as the Technical Education and Skills Development Act of 1994. TVE provides the country with skills training and development for a particular occupation or for midlevel skills training by adopting a competence-based approach.
3. *Higher education.* The Commission on Higher Education (CHED) was created under Republic Act 7722. It is mandated to oversee higher education programs in the country. It endeavors to inculcate values and attitudes for nation building, to develop professions and technologies appropriate for national development, to attain high standards of excellence and efficiency, and to strengthen academic freedom. The CHED is headed by a chairman ranked as department secretary and four commissioners ranked as undersecretaries. It has three major central offices – the Office of Programs and Standards; the Office of Policy, Planning, Research, and Information; and the Office of Student Affairs. It also has 17 regional offices that serve as extensions of its regional offices.

Philippine Educational Ladder

The Philippine educational system can be categorized into formal and nonformal. Formal education is defined

as a sequential progression of academic schooling in three levels – elementary, secondary, and tertiary. Education is compulsory from age 7 to 12 years, covering the six elementary grades. The number of years of formal schooling in the Philippines is shown in **Figure 1**.

The Philippine educational ladder has a 6 + 4 + 4 structure, that is, 6 years of elementary education, 4 years of secondary education, and, typically, 4 years to gain a bachelor's degree, which is considered one of the shortest cycles in the world. A bachelor's degree in the Philippines is obtained after 4 years, which may be extended up to another 4 years for courses such as engineering, medicine, and law. Postgraduate education is, on the other hand, usually completed in 2 years.

The academic school year (SY) in the Philippines starts in June and ends in March, covering a period of 40 weeks. An academic year (AY) usually comprises two semesters, after which students may opt to take optional summer classes.

English was the official language of instruction from 1935 to 1987 in the Philippines. However, the 1987 Philippine Constitution prescribed that both Filipino and English be used as official languages of communication and instruction. English continues to be widely used from the higher primary grades onward, owing to the dearth of materials and resources written in Filipino as well as Filipino-speaking teachers. This is especially true in technical–vocational fields.

Salient Features of the Educational System

Early-Childhood Education

Early-childhood education is governed by the Early Education Act or RA No. 6972, which establishes a daycare center in every *barangay* and institutes therein an overall children's development and protection program. This act mandates that the state shall defend the right of children to assistance, including proper care and nutrition, and to the provision of special protection against all forms of neglect, abuse, exploitation, and other conditions prejudicial to their development. There is a 6-month preschool program for prospective grade 1 pupils in fifth and sixth class municipalities and urban poor areas. However, there are elite preschools that also offer 4-year preschool programs, including nursery, kinders 1 and 2, and preparatory. These are optional, however, for children aged 3–6 years.

Elementary Education

Elementary education is the first stage in formal education. It is free and compulsory, comprising 6 years (grades 1–6) for children aged 6–11. Some elite private elementary

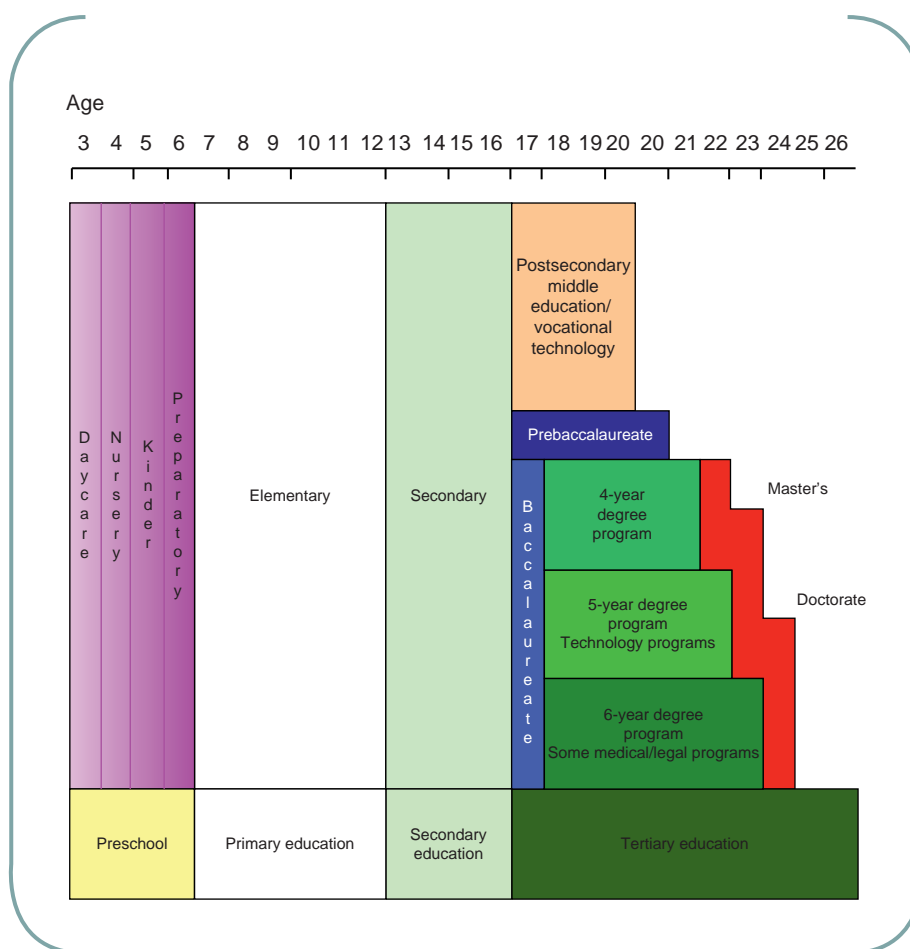


Figure 1 The Philippine education system.

schools also offer an extended elementary level, grade 7, which is optional. The basic education policy of the country prescribes a set of specific competencies in language arts (English and Filipino); mathematics; science; and social studies; or *Makabayan*, which is generally oriented toward civic development; music; arts; physical education; home economics; and livelihood education. The DepEd also prescribes a cumulative rating system where the passing grade is 75%. The students receive a certificate of graduation after completing 6 years of elementary education.

Secondary Education

Secondary education in the Philippines provides general education for students aged 12–15. It is defined as a 4-year comprehensive level, which prepares students for higher education or for the world of work. There are two kinds of secondary education – general and TVE. There are also special secondary schools in the country such as science, cultural, and sports high schools.

A general high school offers academic secondary education while a technical–vocational high school offers three specialized courses – fishery, arts and trade, and agriculture. Secondary school students are rated four times a year with a passing grade of 75%. If a student fails to get a final rating of 75% or higher in a particular subject, he/she repeats it the next year, but is, nevertheless, promoted to the next year. A certificate is issued to technical–vocational secondary school graduates. Philippine secondary schools are dominated by private schools, which require passing an entrance examination prior to admission. All science high schools require passing a competitive examination while admission to general public secondary schools is automatic for elementary education graduates.

The following subjects are taught in general education secondary schools– communication arts (English and Filipino); social studies, including history, economics, geography, and sociology; mathematics; science and technology; technical and livelihood education; computer education; youth development training (YDT); and elective subjects. Technical–vocational high schools, on the

other hand, follow a different curriculum. The first 2 years in such schools are similar to those in a general high school. However, in the third and fourth years, technical–vocational students specialize in any of the following fields – arts and trade, fishery, or agriculture.

Science high schools offer a highly specialized curriculum, which includes advanced science and mathematics courses. The Philippine Science High School (PSHS) falls under the jurisdiction of the Department of Science and Technology (DOST) while other regional science high schools fall under the supervision of the DepEd.

Every secondary education student takes a National Career Aptitude Examination (NCAE) to determine what collegiate courses he/she can take. The NCAE is administered by the DepEd every last week of August. A student who completes secondary education is awarded a diploma and a report card or Form 138, which provides a listing of all the classes he/she took and the corresponding grades he/she earned.

ALS or Nonformal Education

The DepEd boasts of a well-organized ALS to cater to out-of-school youth (OSY) and adult learners with a well-designed accreditation and equivalency (A&E) program developed by the Bureau of ALS, formerly known as the Bureau of Nonformal Education.

The nonformal education A&E system provides an alternative means of learning and certification for basically literate Filipinos and foreigners aged 15 and above who are unable to avail of formal education or who dropped out of a formal elementary or secondary school. It has four components – the NFE A&E curriculum framework, NFE A&E learning materials, the NFE A&E learning support delivery system, and NFE equivalency testing in the elementary and secondary levels. The system uses a range of innovative strategies designed to break

down traditional learning barriers – time, accessibility, and resources. It also uses a range of delivery modes within a framework of flexibility, providing learners with more control over what to learn, how to learn, when to learn, where to learn, and when to enter and exit the system (SEAMEO INNOTECH, 2007).

Higher Education

The Education Commission Report (EDCOM) recommended the creation of the CHED in 1994. The CHED is an independent government agency that is tasked to oversee higher education institutes (HEIs) in the country. The CHED currently oversees more than 1943 HEIs in the country, 451 of which are state universities and colleges (SUCs), local universities and colleges (LUCs), and other government-funded schools while 1589 are privately owned (**Table 1**). At present, there are 2 438 855 higher education students enrolled in these colleges and universities.

The CHED MIS (2007) reported the total enrolment in HEIs as 2 483 645 as of SY06–07. The bulk of enrollees are in the baccalaureate stage (2 134 856). There are 76 048 masters and 8939 doctoral students. The projected enrolment, however, shows a declining trend.

The Commission identifies centers of excellence (COEs); monitors the quality of higher educational programs; regulates the establishment or the closure of private HEIs, their program offerings, curricula, and teaching forces; enforces quality assurance and accreditation; and sets the ceiling for tuition-fee increases. Generally, private higher education institutions (PHEIs) abide by the regulations of the CHED. COEs are granted autonomy or deregulated status. The CHED also encourages HEIs to seek accreditation to receive benefits such as fewer regulations and other grants and benefits such as entitlement to student financial assistance and other awards.

Table 1 Distribution of HEIs by type (SY06–07)

<i>Sector</i>	<i>Institutional type</i>	<i>With/without satellite campuses</i>	<i>Total</i>
Public	SUCs	Main	111
		Satellite	260
	LUCs		65
	CHED-supervised institutions (CSIs)		1
	Other government-funded schools		9
Total public	Special schools		5
		Without satellite campuses	191
		With satellite campuses	451
Private	Sectarian		358
	Nonsectarian		1134
Total private			1492
Grand total		Without satellite campuses	1683
		With satellite campuses	1943

From CHED-management information system (MIS), December 2007.

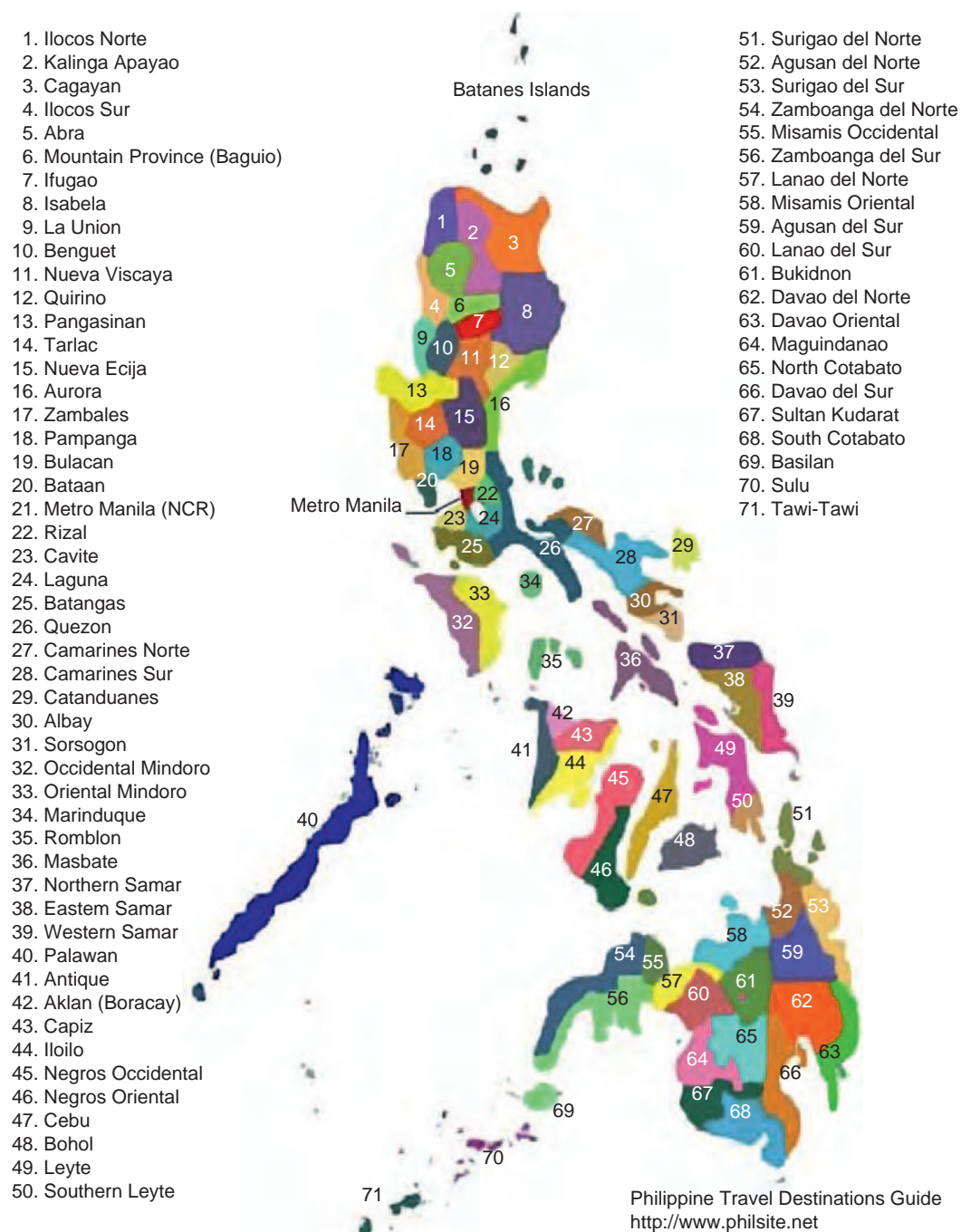


Figure 2 Regional map of Philippines.

State Universities and Colleges (SUCs)

SUCs are funded by the national government and are governed by their own charters. The funds for the local government universities and colleges (LGUCs) come from local governments. As such, their policies and programs are governed by elected local government officials. SUCs' charters ensure their autonomy and academic

freedom. They are empowered to develop their own curricula, to introduce competitive institutional programs, and to award their own degrees. The CHED ensures that SUCs adhere to the Higher Education Modernization Act or RA No. 8292, which provides for the uniform composition of the governing boards of SUCs with the CHED as chairman.

Private HEIs

Private HEIs are owned and administered by private individuals, groups, or organizations. They are classified as either sectarian or nonsectarian. Sectarian schools are usually nonstock, nonprofit educational institutions owned and operated by religious orders (Catholic and Christian schools) while nonsectarian schools are operated by private corporations, which are not affiliated with any religious organization. Majority of nonsectarian schools are stock corporations and only a few are nonstock, nonprofit corporations while a number are registered as foundations.

Private HEIs experience some degree of freedom only when their programs are recognized by CHED and when they have attained accreditation from the Federation of Accrediting Agencies of the Philippines (FAAP). Recognition of academic programs is awarded to private HEIs who comply with the minimum requirements prescribed by the CHED. Around 80 private HEIs are recognized as autonomous and have attained deregulated status according to the commission.

Admission Requirements for HEIs

Entrance to universities and other HEIs depends on possessing a high school diploma and a report card (Form 138) and, in some cases, on the results of the NCAE or, in many PUCs, the results of their own entrance examinations.

Higher education degrees

Bachelor's degree programs are usually for 4 years in length. With the short cycle of basic education, during the first 2 years, the students usually take general education courses (63 credits). In the remaining 2 years, they take professional or major courses. Some bachelor's degrees, however, require more than 4 years of study. Some of the 5-year programs are teachers' education, engineering, and architecture (226 credits). Dental and veterinary medicine, on the other hand, require 6 years of study.

Degrees such as bachelor of laws (LLB or LIB) also require 4 years of study, following a bachelor's degree. The newest program, Juris Doctor (JD), requires an additional eight units (two to four) of coursework and a thesis beyond the requirements for LLB. JD students are expected to complete an internship lasting not more than 12 months, the completion of which combined with completion of an LLB/JD degree, qualifies a student to take the bar examinations administered by the Supreme Court.

A master's of art or science degree in the Philippines usually requires 2 years of full-time study and a thesis or comprehensive examination. Admission to a master's degree requires a bachelor's degree in an appropriate discipline with an average grade equal to or better than 2.00, 85%, or B.

A doctorate in philosophy (PhD) or education (EdD) and in management (DM) requires a substantial amount of coursework. The main requirement for a PhD is a dissertation, which makes up a quarter or fifth of the degree's total credits. PhD, EdD, or DM programs usually require 2 or 3 years of full-time study beyond a master's degree. Admission to a doctoral program requires passing an entrance examination and a master's degree with an average grade of 2.00 or B.

Rating System in HEIs

The system of grading in Philippine colleges and universities is weighted according to a system of instructional units based on the number of class hours where one unit of instruction equals 1 h of lectures or 3 h of lab work per week. Four-year bachelor's degree programs in arts and sciences require a minimum of 120 units for graduation, although the minimum in many schools is likely to be 140–160 units. Some programs may require as many as 185 units.

Quality Assurance and Accreditation System

The Philippine quality assurance system is based on the concept of voluntary accreditation modeled on the regional accreditation system practiced in the United States. Four accreditation associations recognized by the DepEd and CHED were organized to encourage private institutions to raise their programs' levels above the minimal standard. These accreditation bodies are the Association of Christian Schools, Colleges and Universities Accrediting Agency Incorporated (ACSCU-AAI); the Philippine Accrediting Association of Schools, Colleges, and Universities (PAASCU); the Philippine Association of Colleges and Universities Commission on Accreditation (PACUCOA); and the Accrediting Agency of Chartered Colleges and Universities of the Philippines (AACCUP).

The FAAP coordinates and certifies the activities of individual accreditation agencies. To date, there are 221 HEIs with 832 programs in various stages of the accreditation process (Pijano, 2003). A voluntary accreditation system in the technical and vocational sector is currently being implemented by the Technical and Vocational Education Accrediting Agency of the Philippines (TVEAAP).

In the higher education system, there are four levels of accreditation. As defined by the CHED, Level I gives applicant status to schools that have undergone a preliminary survey and have been certified by the FAAP as capable of acquiring accredited status within 2 years. Institutions with programs accredited at Level II receive full administrative deregulation and partial curricular autonomy, including priority in funding assistance and subsidies for faculty development. Programs with Level III

accreditation status are granted full curricular deregulation, including the privilege to offer distance education programs. Level IV institutions are eligible for grants and subsidies from the Higher Education Development Fund and are granted full autonomy from government supervision and control. This level of accreditation is reserved for academic programs considered comparable in quality to those of internationally renowned universities. To date, there is only one institution in the country whose programs have been granted Level IV status – De La Salle University (Pijano, 2007).

Teacher Education

Admission to a teacher education program requires passing secondary education and a high NCAE score. The most common courses are bachelor's degrees in elementary education (BEEd), in secondary education (BSEd), and in another area with the addition of at least 18 units in professional education. All these programs require 4 years of study. Over the years, the teacher education sector has been vigorously and relentlessly exerting efforts to improve its existing curriculum, teacher educators' conditions, and purview. Toward this end, several reform strategies are being advocated, implemented, and revisited. The DepEd, the CHED, and the Teacher Education Council (TEC) now implement a singular competency-based development model – the national competency-based teacher standards, (NCBTS), which defines all aspects and phases of effective teacher development.

Educational Financing

Public education is generally supported by the Philippine government. It is annually appropriated by the House of Congress. Aside from this, local government units (LGUs) also offer a special education fund (SEF) as provided for under the Local Government Code of 1990. Local school boards determine the annual budgetary requirements for operating and maintaining public schools. The main source of SEFs is real property and other taxes collected by the government. The national appropriation for education covers payment of teachers and other school personnel's salaries; purchasing school materials, equipment, and textbooks; and funding for school officials' travels and other expenses. Funds for educational infrastructural projects are taken from the Department of Public Works and Highways (DPWH)'s budget. In higher education, SUCs are supported by government appropriations and matriculation fees paid by students. They also get a substantial amount of funding from income-generating projects and donations. Private educational institutions, on the other hand, do not receive government allocations or support. They mostly derive funds from tuition and other fees collected from their students.

Research in higher education

CHED prioritizes research and development activities of higher education institutions. It issued its National Higher Education Research Agenda and identified Zonal Research Centers. CHED supports incentives to higher education institutions in order to publish Refereed Journals. It also provides grant to those who conduct quality research outputs, studies, and professional papers for possible publications in high impact journal.

Major Changes and Issues since the 1990s

Trifocalization of Education

The Congressional Commission on Education (EDCOM) in 1990 recommended a major reform program that trifocalized the educational system. It created the Legal Education Board to improve the quality of law schools. It also institutionalized the dual training system that allows students to pursue technical vocational courses while pursuing their studies. It also established the CHED and the TESDA.

Ladderization Programs

The Philippine government espouses seamless education. This initiative was launched through Executive Order (EO) No. 358 (2005) entitled Institutionalizing a Ladderized Interface between TVE and Training and Higher Education. EO No. 358 stipulates a unified and articulation mechanism between the TESDA and CHED. It promotes credit transfer and recognition from technical–vocational training to a degree program or vice versa. It also establishes an enhanced equivalency system and a modularized program approach to promote accessible educational programs for all.

Quality Assurance

In the Southeast Asian region, there is a growing trend to move toward a modicum of standards. This development resulted in the identification of COEs and centers of development (CODs). This provides incentives to educational institutions, which consistently exhibit excellent qualities in instruction, research, and extension. A COE is given financial assistance to further improve the quality of its programs and activities in a particular field of study while a COD is further developed to become a COE. Given certain challenges, there is a need to develop more COEs and CODs that will emit quality assurance.

International Education and Mutual Recognition Programs

Globalization has brought about various perspectives and challenges, especially in recognizing and comparing

degrees. The country issued strict policies to regulate cross-border education (CBE) providers through a transnational education (TNA) policy. One concern is that cross-border education cannot guarantee quality of educational degrees, studies, and diplomas. As of now, the country lacked the monitoring and evaluation system governing CBE providers.

Language-in-Education Policies

The present regime supports English as a medium of instruction in response to globalization and booming business-process outsourcing trend in the country. This directive poses challenge to many educators especially for those serving the marginalized and minority groups who learn faster with the mother tongue.

Broadening Access to Education

The country is high in implementing education for all pursuant to the UN millennium development goals. It has instituted policies to provide access to education for Muslim learners and other disadvantaged groups. It has developed a Madrasah curriculum in the basic education sector and expanded the open high school program. There are also new scholarship programs enacted to enable the poor Filipino children to complete their basic education.

Bibliography

- Estioko, L. (SVD) (1994). *History of Education: A Filipino Perspective*. Manila: Logos Publication.
- Executive Order (EO) No. 358 (2005). *Institutionalizing a Ladderized Interface between TVE and Training and Higher Education*. Pijano, C. Philippine Association of Colleges and Universities.
- Pijano, C. (2003). Philippine Association of Colleges and Universities.
- Pijano, C. (2007). Paper presented at the 9th UNESCO Regional Convention on the Recognition of Higher Education Qualifications in Asia and the Pacific. Macau, China.
- San Mateo, R. and Tangco, M. G. (1997). *Foundations of Education II: Historical, Philosophical, and Legal Foundations of Education*. pp 83–85.
- Tullao, T. S. (2000). *Asian Cultural Response to a Global Environment: A Paper Presented in a Conference on Economic Crisis in Southeast Asia and Korea*.

- Valenzuela, E. A. P. (2007). *Situational Analysis of Education for International Understanding: A Paper Presented at the Regional Conference on Peace Education*. Seoul, South Korea.
- Zwanopool, P. (1972). *Tertiary Education in the Philippines, 16-11-1972: A Systems Analysis*. pp 565–566.

Further Reading

- FR CHED (2001). *Philippine Higher Education Guide*. Manila.
- CHED (2001). *Philippine Higher Education in the 21st Century: Strategies for Excellence and Equity*. Manila.
- CHED (2003). *Towards Rationalizing Philippine Higher Education*. Manila.
- CHED (2007). *ICT Education in the Philippines. Paper presented at the E-Global Leaders Meeting*. Seoul, South Korea.
- Commission on Higher Education. Republic of the Philippines (2003). *CHED Memorandum Order No. 6 s. 2003. Policies and Guidelines on Transnational Education*. Pasig City: Commission on Higher Education.
- Congressional Commission on Education (1993). *Making Education Work*. Quezon City: Congressional Oversight Committee.
- Department for Education (2007). *Briefing Materials for the Budget Hearing*. Pasig City: Department of Education.
- DepED (2006). *Literacy Coordinating Council Annual Report*. Manila.
- Johansson, R. (1999). *Higher Education in the Philippines: Technical Background Paper No. 3. Philippine Education for the 21st Century*. Asian Development Bank.
- Pijano, C. (2008). *Accreditation in the Philippines. A Paper Presented at the Basic Education Sector Reform Agenda Conference*.
- The 1987 Constitution of the Republic of the Philippines (1999). Manila: National Bookstore.
- Tullao, T. S. (2003). *Education and Globalization: Philippine APEC Study Network Discussion Paper*. Makati City, Philippines.
- Valenzuela, E. A. P. (2007). *Philippine Higher Education: A Paper Presented in the Regional Conference on the Recognition of Degrees, Studies, and Certificates in Higher Education*. Eighth Regional Convention on the Recognition of Degrees, Studies and Diplomas in Asia and the Pacific, Kunming, China.
- Zaide, G. (2000). *History of the Filipino People*. Manila: National Bookstore.

Relevant Websites

- <http://www.Nscb.gov.ph> – National Statistical Coordination Board-Philippines' policy making and coordinating body on statistical matters.
- <http://www.deped.gov.ph> – Republic of Philippines, Department of Education.
- <http://www.tesda.gov.ph> – Technical education and skill development authority.
- <http://www.ched.gov.ph> – The Commission on Higher Education is the governing body covering both public and private higher education institutions as well as degree-granting programs in all tertiary educational institutions in the Philippines.

Uruguay

B Guinovart and C Troncoso, National Administration for Public Education (ANEP), Montevideo, Uruguay

© 2010 Elsevier Ltd. All rights reserved.

General Background

Uruguay is located to the southeast of South America, its total area is 176 220 km² and it has 3.2 million inhabitants (52% women and 48% men). It is a Spanish-speaking country with a democratic government, organized in three different powers, administratively divided into 19 departments (**Figure 1**).

Of the population, 41% reside in the capital city (Montevideo) and 92% in urban areas. It has a low population growth rate. In the next 20 years, there will be a rise in the proportion of the older layers of society, with consequent impact on public expenditure, due to the huge burden of paying retirement benefits (**Figure 2**).

Uruguay is one among the countries with a high human development. It holds the third place in South America in terms of gross domestic product (GDP) per capita. Regarding the distribution of income, Uruguay is among the most equitable countries. Despite the fact that in the last 2 years poverty was mitigated, 18.3% of the homes and half the population under 13 years of age is poor (with income below US\$160).

The GDP in 2006 amounted to US\$19 308 million. Agro-industrial production is very important, and service areas, such as financial, communications, transportation, and software services, apart from the significant increase of forestry exploitation since the 1990s, have become increasingly relevant. According to official projections, for the 5-year term 2005–09, the GDP is expected to grow at a yearly rate of approximately 5% in average.

Unemployment has decreased in the last 3 years, reaching one-digit levels (9.4%).

In the last few years, inflation has reached one-digit figures (about 6%).

The median household income is approximately US \$800. The recovery of the same, which started in 2004 after a significant fall between 2001 and 2003, is basically due to the recovery of salaries and retirements.

José Pedro Varela established in the late nineteenth century – quite early when compared to the other countries in the region – the basic principles of public education, as being cost free, lay, and compulsory. These principles were also closely linked to the consolidation of the national state with a strong capacity for social integration.

At present, the public system of education has 85–87% of students registered at all levels (960 000 students approximately), with the exception of preschool, where the percentage is lower: 67%.

Goals of the Education System

Within the framework of the goals for the millennium, **Table 1** shows the existing gaps in the case of Uruguay and the main goals that are yet to be achieved.

General objectives

The general objectives mentioned in the 2005–09 strategic plans are:

- Contributing to sustainable human development by promoting the democratization of the educational system.
- Aiming at scientific, technological, and artistic development.
- Improving the quality of education at all levels.
- Diversifying educational offers, increasing their presence outside Montevideo.
- Dignifying the work conditions and improving the existing buildings to enhance production function of the education-related staff.
- Decentralizing management at all levels.
- Fostering management modernization processes which allow for an efficient support of the transformations.

Specific Objectives

The specific objectives mentioned in the 2005–09 strategic plans are:

- Introducing studies in recent history, sexual education, and human rights at primary and middle levels.
- Strengthening the pedagogical bond.
- In primary education: globalizing the educational coverage for 4- and 5-year-olds as well as ensuring the completion of primary education level; broadening the coverage by means of full-time schools (FTSs); developing a primary attention school model for critical sociocultural environments; and achieving effective results in the practice of physical education.
- In middle school: globalizing the coverage of the basic cycle; improving the levels of coverage of middle high school; significantly reducing dropout and fall-behind levels; and reducing excessive rotation of teachers.
- In teacher education: strengthening graduate education; fostering postgraduate education; and giving it a university-level status.
- In technical-professional education: developing and consolidating middle-school and professional and

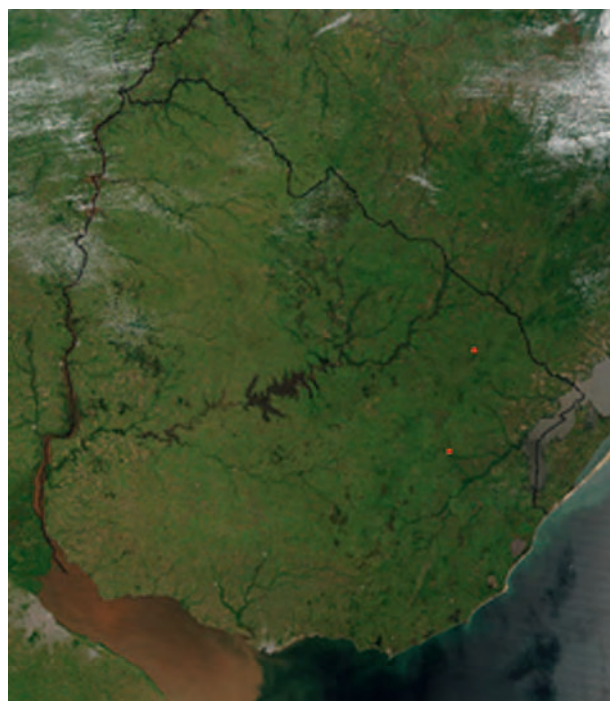


Figure 1

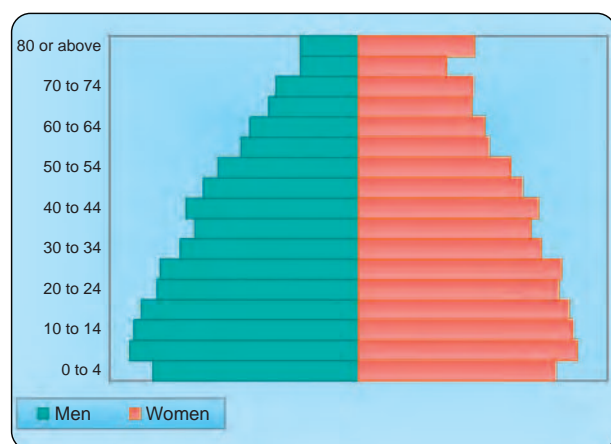


Figure 2 Distribution of the population according to the gender and the age range. From own elaboration based on Census Step 1 2004.

technological tertiary education within the framework of the country's productive and social development.

- In the education of young people and adults: raising the educational level of those members of society who have not completed primary studies.
- In tertiary education: generalizing advanced education establishing links with productive sectors; fostering creation of a tertiary education system; broadening and strengthening the offer of postgraduate and masters courses; playing an active role in the national innovation systems; and creating a national assessment and accreditation agency.

Table 1

Diagnosis for Uruguay (November 2003–March 2004)	<ul style="list-style-type: none"> • Six out of 10 children do not receive preschool education. • The highest dropout levels in Uruguay are at the middle-school level. • Half of the young people population are not able to finish middle high-school. • Twenty percent of the young people between the ages of 17 and 24 neither study nor work.
Uruguay's goals	<ul style="list-style-type: none"> • Globalization of the preschool education. • Globalization of the middle basic cycle education. • Extension of the middle high-school education.
Challenges	<ul style="list-style-type: none"> • Reduction of the inequities in primary school. • Teenage dropouts who do not enter the labor market. • Teacher professionalization.

Structure of the Education System: Regulations, General Functioning, and Main Institutionalities

The regulations in force set forward 10 years of compulsory education (11 years as of 2009): 1 year of preschool (5 years until now and 4 and 5 years as of 2009), 6 years of primary education, and 3 years of middle basic cycle.

Public education is free at all levels.

Figure 3 summarizes the education cycles and the type of administration.

The main institutions related to education are: the Ministry for Education and Culture (MEC), which is a part of the Executive Power; the National Administration for Public Education (ANEP), the organization in charge of planning, managing, and administrating preschool, primary, middle school, technical, and teacher public education as well as controlling private education at those same levels; the University of the Republic (UDELAR), the organization in charge of public university education and private education institutions at all levels; and the Commission for the Coordination of Education (CCE), a consultant organization.

The MEC is responsible for coordinating education nationwide. Unlike other countries, the design and application of educational policies are not its direct aims. It is also responsible for supervising private institutions at early childhood, tertiary, and university levels.

The centers for education and production (CECAPs) are also part of the MEC, providing 780 young school dropouts with basic skills. The Industrial Design Center (CDI) is in process of integration into the UDELAR.

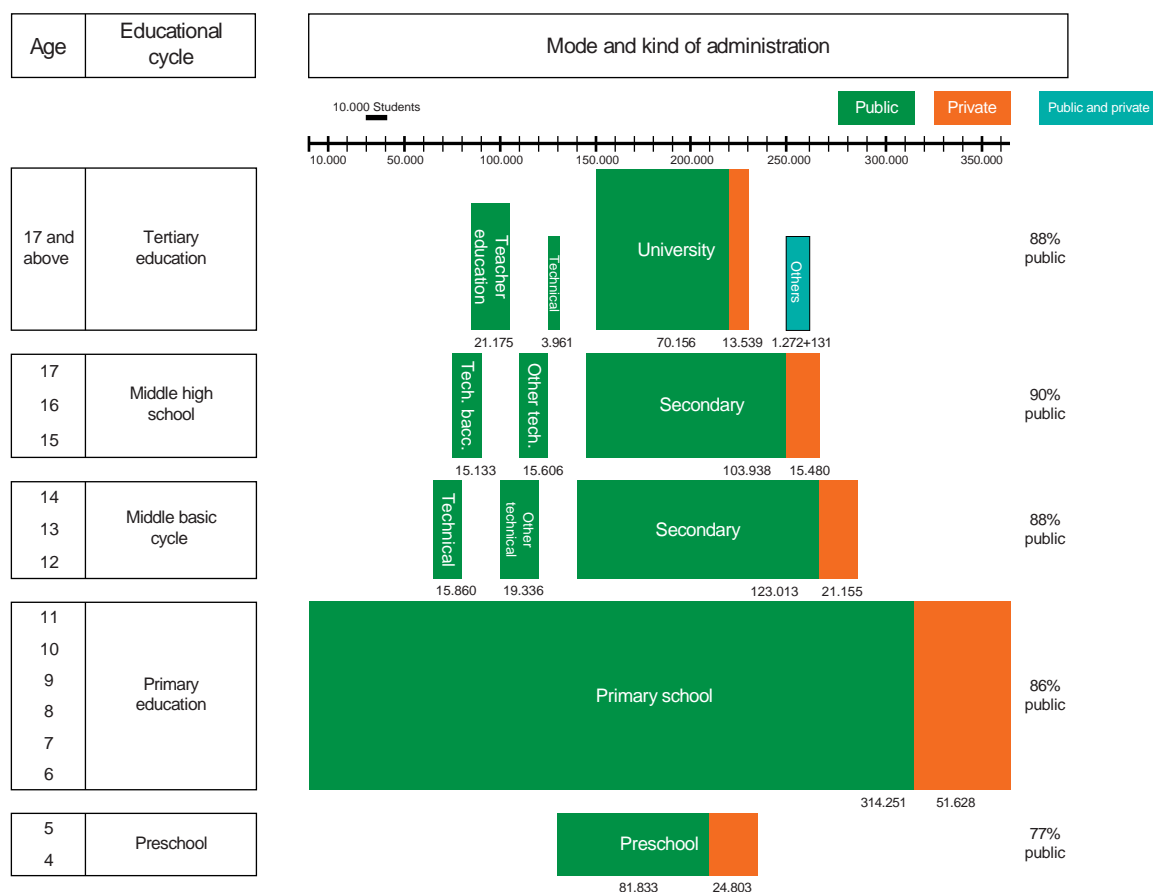


Figure 3

ANEP, which consists of 85% of the total number of students at all levels, is directed by five members of the Central Directorate Council (CODICEN). This agency is responsible for setting up the general orientation for primary, secondary, and technical-professional education plan of studies and curricula, as well as approving the plan of studies.

Within CODICEN, there is the Directorate for Teacher Education and Further Training (DFPD), and the program which is offered by them for young people and adults.

Preschool and primary as well as secondary and technical educations are directed by three independent councils, each one of which is integrated by three members: the Primary Education Council (CEP), the Secondary Education Council (CES), and the Technical-Professional Education Council (CETP). The decentralized councils above are in charge of providing education at the corresponding levels and supervising the approved private institutes at the mentioned levels.

The UDELAR – which represents 64.7% of the total number of tertiary-level students, 83% of the university students, and consists of 14 schools, three institutes, two university centers, four graduate services outside Montevideo, and one hospital – is an autonomous body that is

co-governed. Its responsibilities are higher public education, fostering scientific research and artistic activities and contributing to the study of general interest problems.

University education and university institutions were open to the private sector in 1985. Those institutions which offered teaching, research, and extension activities in three or more independent areas of knowledge were granted the status of university', whereas those which developed these activities in one or more related areas or in two independent ones, were considered university institutions.

At present, there are four private universities (Catholic University, ORT University, University of Montevideo, and University of the Enterprise) and seven private university institutions.

The private institutions are represented by two different associations: the Uruguayan Association for Catholic Education and the Association of Private Education Institutions. There is also a Council of the Presidents of Uruguay's Private Universities.

The Teacher's Meetings (ATD) held at primary, middle, and teacher public education schools are very important in the Uruguayan education system, as are the national teachers' meetings of each independent council.

This institutional order promotes and organizes teacher participation, giving teachers the rights of initiative and consultation regarding technical-pedagogical issues in their corresponding area of education.

Teacher and employee unions as well as student unions at the university level have existed for a very long time and constitute powerful groups within the educational system.

Parents and friends commissions established by public schools form the link between the community and the education centers.

Performance of the System

Population Literacy and Schooling

- The literacy rate is 97.8% (2006).
- Eighty-three percent of the population over the age of 15 have completed at least the primary education level (Figure 4).
- The population between 20 and 25 years of age have an average of 10.5 years of schooling, with significant differences depending on their levels of income (Figure 5).

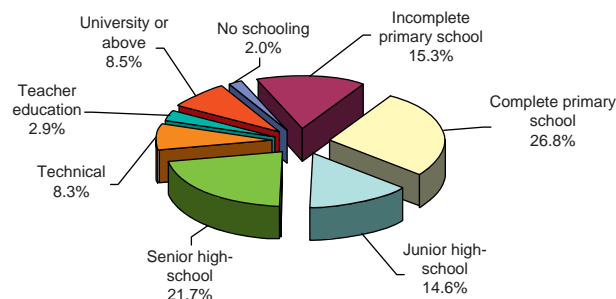


Figure 4 Educational level of the population aged 15 and above – year 2006.

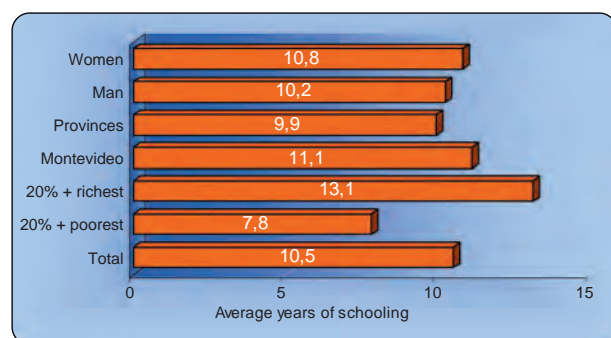


Figure 5 Years of schooling of the population between the ages of 20 and 25 – year 2005. From own elaboration based on ECH-INE.

Enrolment

The evolution of enrolment shows the following:

- There have been increases at the following levels: early childhood, private preschool, private primary school, public rural middle school, teacher education, and public and private university.
- After consistent increases until 2004, there were decreases at the following levels: public early childhood, public primary school, public middle school, public technical-professional middle-school education, teacher education, and tertiary technical-professional education (Table 2).

Schooling ages 3–17

The coverage rates in preschool show an upward growth curve.

At the primary education level, schooling is nearly global; 95% of the school-aged population attends school.

At the secondary education level, 70% of the population between 12 and 14 years of age attends an educational center, whereas the coverage for 15–17-year-olds is between 40% and 50% (Figures 6 and 7).

The differences in the coverage rates in relation to the level of income are significant at the preschool level and from the age of 14 (Figures 8 and 9).

Repeating and Dropping Out in Public Primary and Secondary Education

Repeating rates in public primary schools (8%) have shown a downward trend in the recent years, especially in full-time schools. The gap between schools belonging to different sociocultural contexts was narrowed (Figures 10 and 11).

Dropout rates in public primary education have decreased in different socioeconomic contexts, going from a 1.4% average in 1992 to 1.0% in 2005. Schools in unfavorable and very unfavorable contexts showed a higher dropout rate (Figure 12).

In public secondary school, repeating and dropout rates increased in the first 4 years (Figure 13).

Young People between the Ages of 15 and 17

Fifty-eight percent of the young people between the ages of 15 and 17 attend an educational center and 21% are over-age students. Of the 21% who are out of the educational system, 27% work, 23% are looking for a job, and 50% neither work nor look for a job.

Table 2

	Enrolment 2000–06						
Educational levels and institutions	2000	2001	2002	2003	2004	2005	2006
Early childhood			38 096	30 345	35 413	28 069	37 416
Public preschool (CEP)	84 984	87 155	86 906	85 268	84 612	83 546	81 833
Private preschool	20 802	18 642	19 317	17 991	20 603	22 025	24 803
Public primary school (CEP)	309 416	313 134	316 832	320 025	319 903	317 665	314 251
Private primary school	50 343	45 097	48 172	45 470	47 525	48 175	51 628
Public secondary education (CES)	208 587	219 324	233 956	242 319	236 538	228 946	225 133
Public rural middle school	641	1 037	1 498	1 771	1 869	1 797	1 818
Private secondary education	35 899	35 507	35 249	34 281	35 192	36 309	36 635
Public technical-professional middle school (CETP)	59 433	59 013	61 112	65 265	65 551	62 405	65 935
Public teacher education (DFPD)	13 166	14 788	16 610	19 298	20 968	20 435	21 715
Public technical professional tertiary school	283	2 314	2 564	3 514	3 671	4 024	3 961
Other public tertiary education centers	1 901	1 777	1 964	1 970	2 097	790 ^a	1 272
UDELAR	66 502	70 156	70 156 ^a	70 156 ^a	70 156 ^a	78 300 ^b	80 000 ^b
Private universities	8 077	9 716	9 497	10 354	9 637	13 465	14 997
Private university institutes	401	386	456	466	599	1 398	1 649
Other private tertiary institutes (incomplete data)	2 350	2 883	473	1 244	445	305	131
Total	862 785	880 929	942 858	949 737	954 779	946 864	963 177

^aIncomplete data.^bEstimate data.

Data from 2006 MEC Yearbook. For UDELAR 2000–04: MEC Yearbook for 2005 and 2006: Estimate based upon projections made by UDELAR.

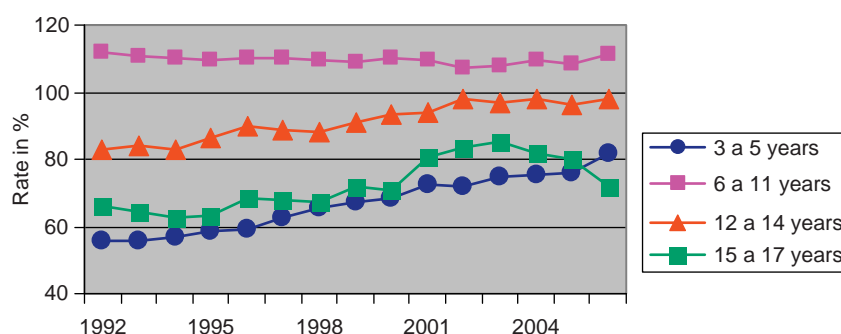


Figure 6 Gross schooling rate by education cycle. From Research, Evaluation and Statistic Department of Central Directorate Council based on ECH-INE.

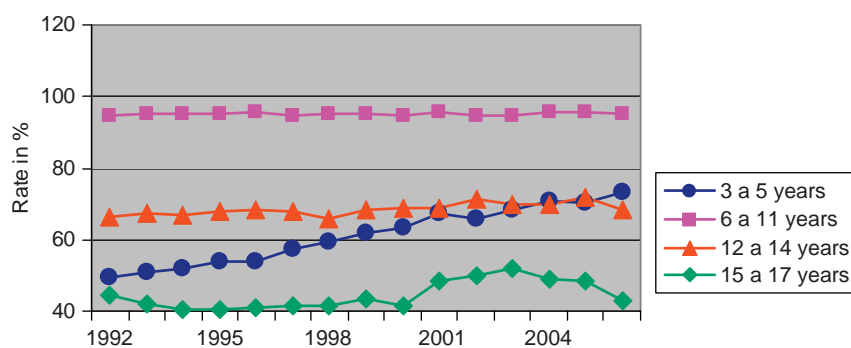


Figure 7 Net schooling rate by education cycle. From Research, Evaluation and Statistic Department of Central Directorate Council based on ECH-INE.

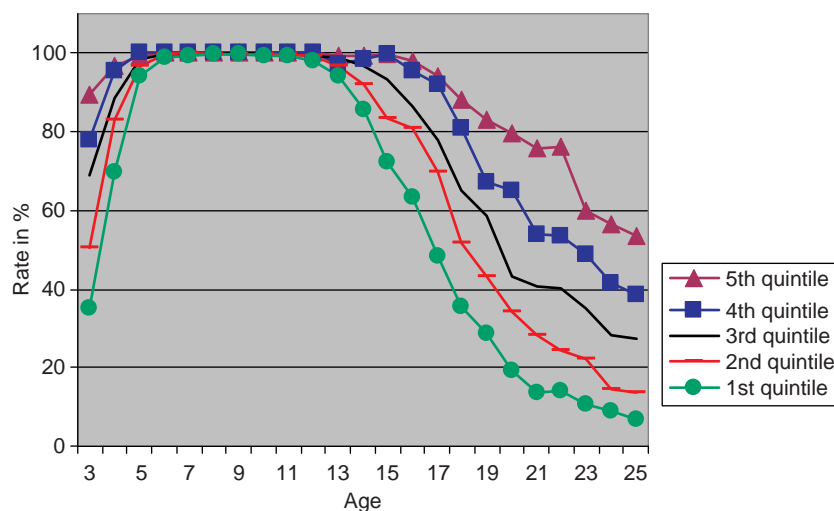


Figure 8 Attendance to an education center per income quintile – year 2006. From Research, Evaluation and Statistic Department of Central Directorate Council based on INE-ECH.

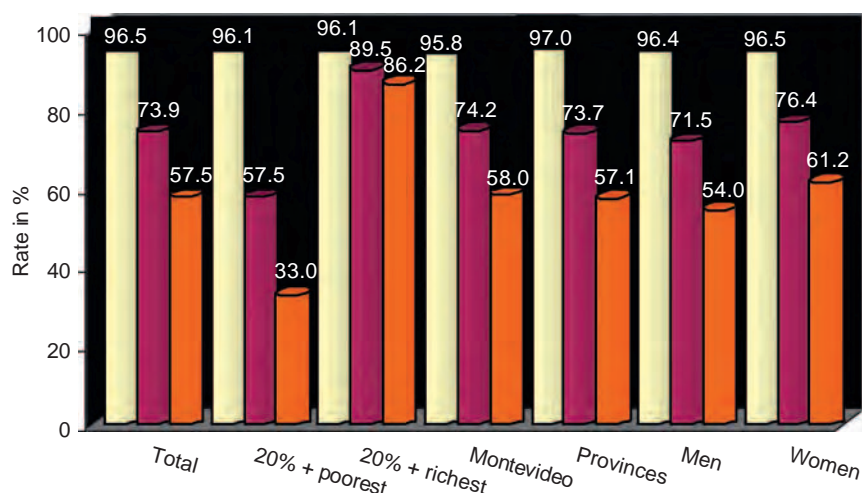


Figure 9 Net schooling rate – year 2005. From own elaboration based on ECH-INE.

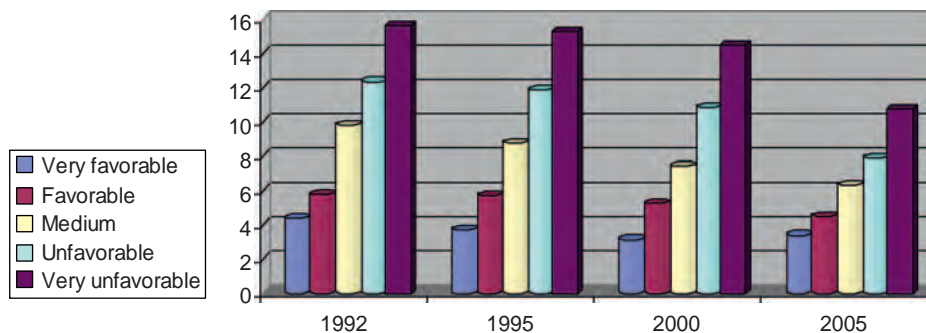


Figure 10 Repeating rates in 1-6 years of primary school, by school's sociocultural context. From Educational Statistic Department of CES.

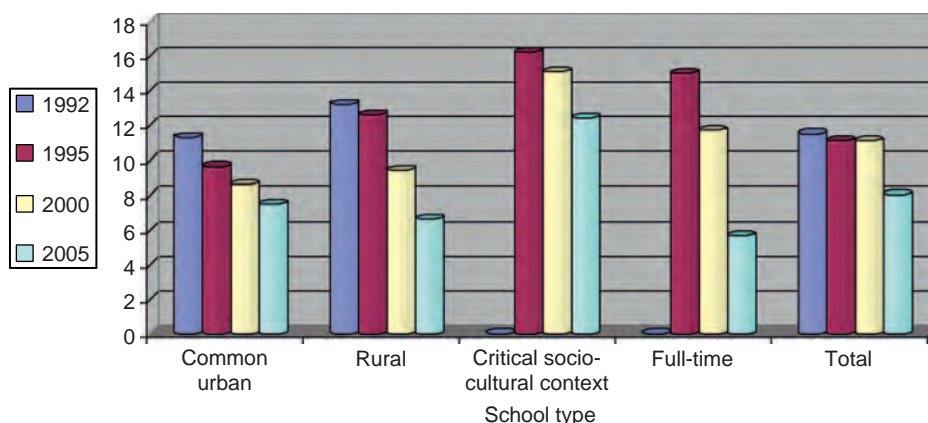


Figure 11 Repeating rates in 1–6 years of primary school, by school type. From Educational Statistic Department of CES.

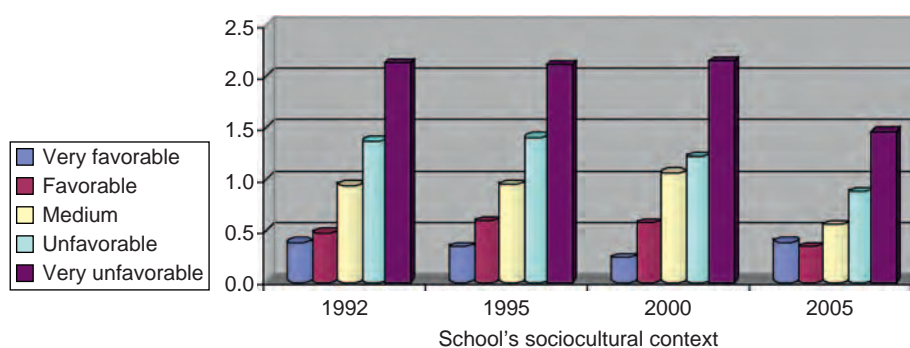


Figure 12 Dropout rates in 1–6 years of primary school, by school's sociocultural context. From Educational Statistic Department of CES.

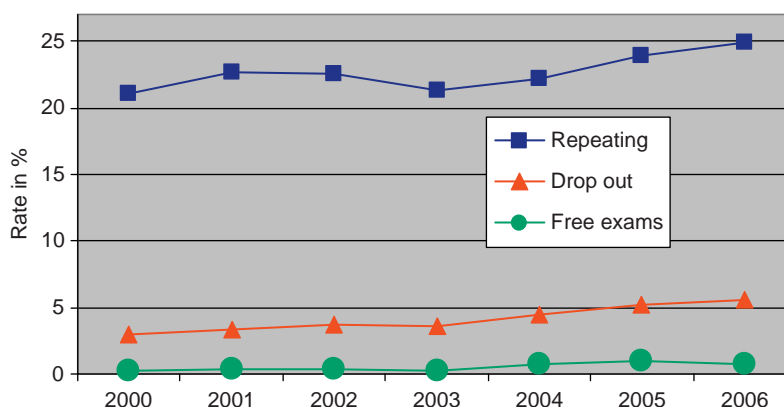


Figure 13 Repeating and dropping out in public secondary education (first to fourth day shift). From Educational Statistic Department of CES.

Uruguay in Programme for International Student Assessment (PISA) 2006

Upon comparison of the two tests (2003–06) it is possible to see that:

- Reading shows the lowest average for all the institutions. In mathematics, scores have not varied significantly.
- There have been decreases in the averages for all the sociocultural contexts, with the exception of the unfavorable one (Table 3).

For 2006:

- Forty-two percent of the Uruguayan students did not reach the competency threshold in science (409 marks),

while 30% did not reach the minimum performance, and 28% passed.

- Forty-six percent of the Uruguayan students did not reach the minimum score (407) in reading, 23% did, and 31% passed.
- Forty-six percent of the Uruguayan students did not reach the minimum score (419) in mathematics, 25% did, and 29% passed.
- Nearly three-quarters of the students from centers located in very unfavorable sociocultural areas scored below the threshold (Table 4).

Notwithstanding the above, Uruguay has had better results than other countries in the region in reading and mathematics (Table 5).

Teaching Profession

The preschool, primary, and middle teacher education courses are offered by the ANEP, through the DFPD established in 1986.

The DFPD has 32 centers which cover 92.1% of the student teachers – 23 296 in 2006 – and one further education and postgraduate center. In 2006, 72% of the students enrolled with DFPD were student teachers for secondary education, 27.7% for primary education and

preschool, and 0.2% for technical education. The first postgraduate course was designed in 2007, to be put into practice together with three others as of 2008.

The oldest centers in the country belong to the primary school teacher education institute (Table 6).

The enrolment at the universities which educate teachers was below 1119 students in 2006, which represented 4.8% of the total enrolment. About 40% of these students entered different postgraduate courses offered by private universities.

In UDELAR, access to a teaching position is gained through a series of competitive examinations in which the candidate's academic background – especially in the corresponding area – teaching experience, work experience, research, and extension are assessed, as along with the candidate's capacity in the educational aspect.

In the year 1968, but gaining momentum only in the 1990s, UDELAR started creating teaching support units at different schools – 16 at present – in charge of training professors in the different didactic and pedagogical aspects (Table 7).

University institutes represent 1.6% of the total enrolment, whereas nonuniversity tertiary institutes gather 1.5% (Tables 8 and 9).

ANEP has 41 139 teachers (2007 census), UDELAR has 7284 professors (2006 estimate), which represent altogether 93.7% of the total academic staff of the different educational institutions, estimated at 51 652.

Table 10 shows some of the characteristics of the ANEP and UDELAR teachers:

The salaries of ANEP teachers and UDELAR professors range from US\$250 to US\$600 for 20 h/week, depending on their seniority and academic degree. In 2002, there was an important loss in the purchasing power, which could be recovered only in 2009. However, the recovery could not fulfill the Union's ambitions, whose goal amounted to 50% of the standard household consumption basket (US\$1390).

Table 3

	<i>Reading</i>		<i>Mathematics</i>	
	2003	2006	2003	2006
Public secondary education	429	410	416	420
Technical-professional education	364	351	370	394
Private middle education	524	495	501	495

From Uruguay in PISA 2006.

Table 4 Averages per area of assessment depending on the sociocultural environment and the Institutional Sector of the Education Center in 2006

<i>Environment</i>	<i>Sector</i>	<i>Science</i>	<i>Mathematics</i>	<i>Reading</i>
Very unfavorable	Public secondary school	371.8	354.6	345.0
	Technical	362.0 (ns)	370.3 (ns)	326.2
Unfavorable	Public secondary school	422.2	415.9	405.9
	Technical	381.0	381.6 (ns)	341.3
Medium	Public secondary school	448.0	452.2	448.3
	Technical	414.5	431.3 (ns)	376.4
Favorable	Private secondary school	441.0 (ns)	435.0 (ns)	456.5 (ns)
	Public secondary school	493.8	502.7	483.6
	Technical	520.9	529.6	452.2
	Private secondary school	473.4 (ns)	474.7 (ns)	478.1 (ns)
Very favorable	Private secondary school	526.3	525.6	517.9

(ns) The difference is not statistically 5% significant regarding the scores obtained by public secondary school students in the sociocultural environment in question.

Public Expenditure in Education

In the year 2006, public expenditure in education represented 3.31% of the GDP, 73% of which was given to ANEP, 19% to UDELAR, and 8% to the MEC. The ratio of public expenditure to GDP increased from 1995 with a decrease in the crisis of 2002–03 (Figures 14 and 15).

The expenditure for education represents between 11% and 12% of the total public expenditure (Figure 16).

In 2006, wages represented 81% of ANEP's expenditure and 84% of UDELAR's. Operational expenditure represented 12% and 13.5%, respectively, whereas investments represented 8% and 2.6%.

Most of ANEP's budget is financed with local resources which come from taxation (88%), 10% through a specific tax for primary education, and 2% through foreign debt. The 52% of the investments are financed through debt (Table 11).

Table 5 Trends in reading and mathematics for the Iberoamerican countries

	Reading			Mathematics		
	2003	2006	Diff. 06–03	2003	2006	Diff. 06–03
Argentina		374		381		
Brazil	403	393	–10	356	370	14
Chile		442			411	
Colombia		385			370	
Spain	481	461	–20	485	480	–5
Mexico	400	410	10	385	406	21
Portugal	478	472	–6	466	466	0
Uruguay	434	413	–21	422	427	5
OECD	498	492	–6	500	498	–2
Average						

In the case of UDELAR, more than 90% of its budget is financed through funds which come from taxation. Finance is also obtained through projects in collaboration with the productive sector.

Two relevant points need to be highlighted:

- The Mutual University Fund was created in 1994, through which UDELAR graduates and tertiary education CETP graduates contribute to financing scholarships for students attending these institutions, as well as MEC's Education and Work Program. The 81% of the scholarships granted in 2005 (3777) were for students from the departments, except Montevideo.
- The legal provision under which, from 2007, ANEP and UDELAR have additional funds to finance educational and investment projects. In the case of investment projects, the sum amounts to 20 million dollars per year.

Private education institutions are financed through the fees paid by the students, as well as by contributions from the private sector.

Research, and Monitoring and Evaluation

Research

In the year 2002, Uruguay had 3120 researchers, 89% of who worked at UDELAR; 42% of these researchers worked in the area of natural and exact sciences, 23% in medical sciences, 12% in engineering and technology, 11% in agricultural sciences, and 12% in humanities and social sciences.

Some support policies in this area are mentioned:

- The creation of the Program for the Development of Basic Sciences (PEDECIBA) in 1986, aimed at training human resources in the basic scientific disciplines and at financing research projects.

Table 6 ANEP – CODICEN teacher education centers–Directorate for Teacher Education and Further Education (DFPD)

Name	Aim	Established	Number and location
Primary School Teacher Education Institute	Education of teachers for primary school and pre-school levels	1882	One in Montevideo
Artigas Teacher Institute	Education of teachers for secondary school	1951	One in Montevideo
Technical Teacher Education Institute	Education of technical teachers for technical education	1964 depending from CETP; 1997 depending from DFPD.	One in Montevideo
Teacher Education Centers	Primary school teacher education and middle school teacher education implying presence or semipresence courses	1977	22 nation wide The semi-presence course center is conducted from Montevideo.
Regional teacher centers	Education of teachers for secondary school	1997	Six nationwide
Further education and higher studies institute	Refresher, specialization and postgraduate courses. The first postgraduate course in education and development started in 2007		One in Montevideo

Table 7 Universities educating teachers

<i>Institution</i>	<i>Course</i>	<i>Aim</i>	<i>Enrolment in 2006</i>
UDELAR – Physical Education Institute	Licentiate in physical education	To educate in the areas of physical education, sports and recreation.	180
UDELAR – School of Humanities and Education Sciences	Licentiate in education sciences, teaching option	To educate teachers for the area of the education sciences.	331 in both options (teaching and research)
UDELAR – Teaching Support Units and Sector Commission for Teaching	Training courses for university professors.	To train professors in the pedagogic and didactic areas.	n/d
Catholic University (UCUDAL)	Licentiate in preschool education	Training for people interested in teaching young learners.	100
	Postgraduate course for university professors	Professional further training for university professors	62
	Masters degree for university professors	To train professors at university level	117 in the different areas
ORT University	Diploma in education	For educators and other professionals interested in educational innovation	65
	Masters in education	To foster the creation of a theoretical–methodological support to allow for the production and the critical analysis of relevant knowledge in the area of education.	79
University of Montevideo	Teacher education courses in: mathematics, literature, history, philosophy, and English	Education of teachers for middle education approved for ANEP.	56
University of the Enterprise	Masters in education	To improve teacher education and to offer areas for analysis and research in education.	83
	PhD in education		46

Table 8 University institutes educating teachers

<i>Institution</i>	<i>Course</i>	<i>Enrolment in 2006</i>
Young Men's Christian Association	Licentiate in physical education, recreation, and Sports	315
Latin-American Human Economy Center	Postgraduate and masters degree in basic education didactics	29
	Postgraduate and masters degree in middle education didactics	34

Table 9 Tertiary education nonuniversity institutes educating teachers

<i>Institution</i>	<i>Course</i>	<i>Enrolment in 2006</i>
National Youth Institute	Social educator	231
Pedagogical Research and Experimentation Center	Preschool educator	114

- The creation of the Sectorial Commission for Scientific Research (CSIC) to foster research within UDELAR.
- The creation of the National Researcher's Fund in 1996, aimed at funding full-time researchers residing in the country, through the MEC.
- The creation of the Technological Development Program in 2001, within the MEC.
- The creation in 2006 of the National Agency for Research and Innovation, to advise the Executive Power in the areas of science, technology, and innovation, to carry out plans, and to coordinate.
- The fund forecasting for research and innovation for the 2007–09 period amounts to 19 million dollars, 30% of which corresponds to foreign debt.

Monitoring and Evaluation

The Constitution of the Republic sets forth the obligation of the public organizations to submit their public 5-year budget plans and annual accountability and execute budgeting reports to the Legislative Power. Annual reports present the amounts spent and the programs and aims associated to the same, as well as their relation to goals established in the strategic 5-year plan. The Legislative

Power may give its opinion on these matters, propose adjustments, and, after the legal terms for analysis and discussion, pass the annual accounts and budget acts.

The country has improved the submission of its annual accountability reports and has made progress in terms of defining strategic guidelines, associating objectives and

goals in executing public programs and policies. ANEP has established its strategic plan for the period 2005–09 and has put the system of indicators for the follow-up of guidelines, objectives, and goals into practice in its programs, based upon which it has been submitting its annual reports since 2006. On the other hand, UDELAR's strategic development plans define different institutional projects which guide the institution's management follow-up.

The participation of ANEP in the PISA program in 2002, the learning assessment census in the last year of primary education, and the educational monitor are crucial follow-up tools.

UDELAR's University Information System, created by the government has, among its aims, the organization of an information system which allows for the planning, follow-up, and evolution of educational management.

In terms of evaluation and accreditation systems, the points on university education are as follows:

- The Central Commission for Institutional Evaluation was created in 1998 in UDELAR, proposing the guidelines for an evaluation program. Based on this, the different member institutions started their own self-evaluation processes.
- In 2001, an external peer committee carried out the assessment of UDELAR's Strategic Plan in the areas of content, adequacy, and degree of fulfillment.
- Participation in the experimental accreditation mechanism for MERCOSUR, Bolivia, and Chile (MEXA). Four universities, nine courses, 11 309 students, and 2305 professors participated in it.
- All the courses presented by Uruguay were credited.
- The ORT University in 2007 was ranked again among the best 500 universities in the world according to the Times Higher Education Supplement of London.

In terms of evaluation and accreditation systems, the points on teacher education are as follows:

- The Coordinating Commission for Self-Evaluation of DFPD was established in 2006. Its aim is to contribute to the development of the academic conditions

Table 10

<i>Characteristics</i>	<i>ANEP (prediction of results for the 2007 census)</i>	<i>UDELAR (2000 census)</i>
Women	80.4%	47%
Men	19.6%	53%
Age 29 and below	20.2%	13.1%
Ages 30–39	28.0%	31.9%
Ages 40–49	27.4%	30.9%
Ages 50–59	20.6%	16.7%
Ages 60–69	3.8%	7.4%
Percentage of professors whose total personal income is their teaching salary	78%	20%
Have a degree	77%	n/d
Specific degree	100% CEP 59% CES 46% CETP 86% DFPD	n/d
Postgraduate degree	n/d	45.2%
Research activities	n/d	46.7% between 1997 and 2000; 60% in the case of those who have postgraduate degrees
Working at only one educational center	58%	n/d
Working at three or more educational centers	42% in DFPD 31% in CETP 28% in CES 1% in CEP	n/d

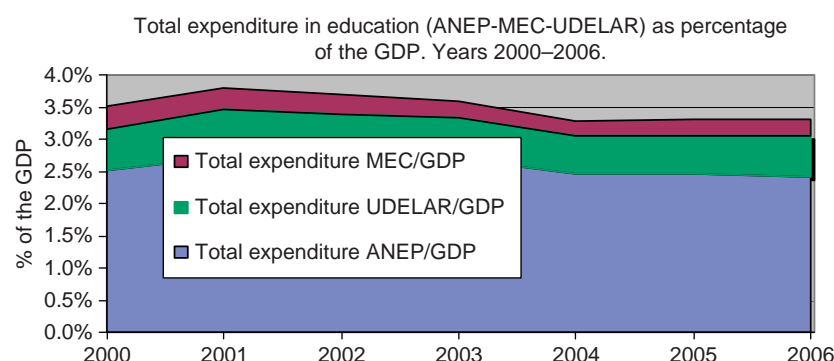


Figure 14 Total expenditure in education (ANEP–MEC–UDELAR) as percentage of the GDP – years 2000–06. Note: MEC includes only educative items. UDELAR includes assistance to the Clinics Hospital. Expenses in education of the Ministry of Defense, Ministry of Interior, Plan CAIF, and INAU are not included. From Executive budgeting balance.

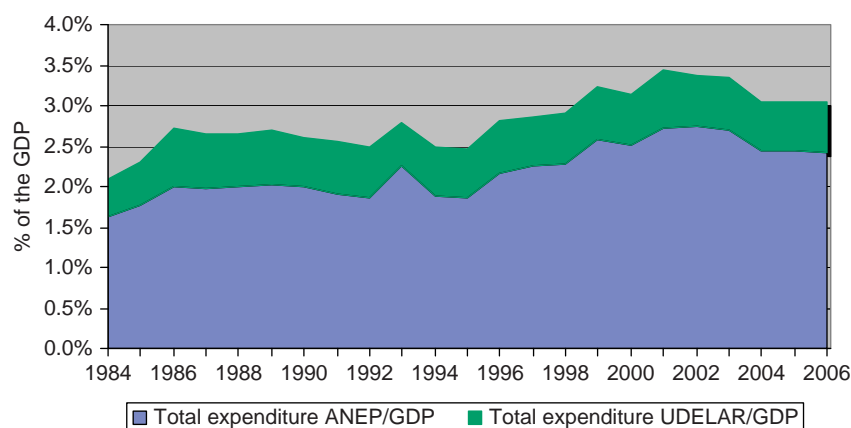


Figure 15 Total expenditure in education (ANEP–UDELAR) as percentage of GDP (years: 1984–2006). From Executive budgeting balance.

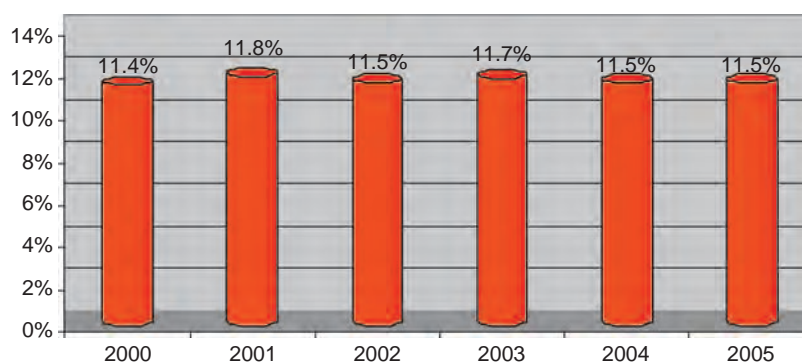


Figure 16 Participation of the public expenditure in education on the total public expenditure. From Executive budgeting balance.

Table 11

Financing sources	Average for the selected years			
	90–94	95–99	00–04	05–06
General taxes	95%	88%	87%	88%
Specific taxes	3%	7%	9%	10%
Foreign debt	2%	5%	4%	2%

From Executive budgeting balance.

necessary to raise the levels of quality and to tend toward the development of an autonomous and self-regulated university teacher education center.

Major Changes and Issues since the 1990s

Certain policies and actions have been adopted by educational institutions for the achievement of the goals and objectives presented above:

- Early education and education for life:
 - There is a project to globalize the enrolment of 4-year-olds and to advance the incorporation of 3-year-olds.

- Ongoing education units have been installed at most of the tertiary-level institutions.
- Attention to students from less-favorable contexts:
 - Full-time schools (CEP since 1992): extension of the pedagogical time, less children per class, socialization workshops, and school meals.
 - Community teachers (CEP since 2005): to favor the educational insertion of children coming from unfavorable contexts through community teachers working with the schoolchildren and their families.
 - Community classrooms (CES since 2005) to achieve the reinserion of 12–15-year-olds who have drifted away from the educational system.
 - Basic professional education for 15-year-olds and above, who have not completed basic secondary education (CEPT 2000).
 - School meals program (CEP from 2001) under which children are provided lunch at public schools.
 - Scholarship plan for students attending teacher education institutes of ANEP.
 - Economic support scholarship and housing program for UDELAR students (since 2002).
 - Scholarship programs at the different private universities under which students are exempted from the payment of the whole or part of the tuition fees.

- Strategy design and development to handle an effective number of students in order to achieve an optimum relationship between the staff and the students in public institutions.
- Incorporation of ICTs as part of the teaching-learning process and for the improvement of the academic and administrative management:
 - Creation of multimedia classrooms in 1995 at different levels to be used as pedagogical spaces.
 - Distance and half-distance educational proposals put into practice at the tertiary-education level.
 - The launch of Plan Ceibal by the present government in one department outside Montevideo to reduce technological illiteracy. By 2009, all children attending public schools shall be integrated into the one laptop per child program not only in its technological dimension but also in the pedagogical one (**Figures 17–19**).
- Territorial and management decentralization and emphasis on the education center:
 - Introduction of regional teacher inspections by ANEP.
 - Strengthening teacher training centers outside Montevideo (started in 1996).
 - Distribution of financial allotments to ANEP's education centers to be managed by them in agreement with the local educational community (started in 2006).
 - UDELAR created the Regional Norte in the department of Salto (1997) in order to strengthen the educational offer in the provinces and broadened its academic offer to 17 graduate courses, making it possible for 1000 students to access university studies.
- Private tertiary institutions have also begun to make educational offers outside Montevideo.
- The proposal of a new Education Act which is under study establishes that the education centers shall be relatively autonomous.
- Creation of coordination and decision-making spaces:
 - Creation of co-government commissions within UDELAR to foster and articulate the different university services and stakeholders involved in research, teaching, and extension.
 - The organization of graduate courses in areas of knowledge within UDELAR since 1999.
 - Joint projects between ANEP and UDELAR.
 - Willingness to organize a national system of public education and a tertiary education network.



Figure 18



Figure 17



Figure 19



Figure 20

- Updating and reformulation of study plans as a strategy to update and reach consensus:
 - In 1996, the Education Reformation was carried out in ANEP. Actions were taken regarding preschool education and full-time schools. Likewise, Plan 1996 was launched for secondary education, which implied a curricular transformation based on teaching in areas of knowledge as well as increasing pedagogical time. The Plan did not have the support of many of the members of the teaching staff. The new administration is designing a new plan referred to as Reformulation 2006 to be implemented progressively in all education centers with active participation of the different groups. Teaching shall be done in subjects again but the strategy of extending the pedagogical time shall be kept.
 - CETP technological baccalaureate: it started in 1997 and it was reformulated in 2007. It provides specific technical training for the work market, and it also enables students to enter university.
 - The graduate course study plans at UDELAR were updated between 1989 and 2002. The largest update since 2002 was the growing offer of postgraduate courses.
- The creation of educational offers that connected students to the world of employment:
 - New technological courses were launched in 2005 by ANEP-UDELAR.
 - Training of technicians for the new production processes at CETP since 2006. Private universities have outlined strategies to provide support for integration into the labor market and to undergo traineeships while studying.
- Progress has been made in terms of conceiving the building infrastructure as an intelligent and flexible



Figure 21

space, which enables pedagogical bonding as well as research. Two examples:

- UDELAR's FARO multifunctional building: three schools located nearby share this building which has large spaces and flexible facilities for different uses and different numbers of students, admitting incremental annexing even when it is in use (2007) (Figure 20).
- ANEP's multimedia learning rooms which combine a library and multimedia resources center within a space designed for teaching, studying, interaction with peers and teachers, and research.
- The creation and search for new institutionalities (Figure 21):
 - Between 1996 and 2004, three of the four existing private universities were created. In the period between 2000 and 2006, these universities had an 81% increase in the enrolment and 86% in the

graduation of students.

- The proposal to turn DFPD into an autonomous pedagogical university;
- The National Debate on Education carried out in 2006–07 in order to generate input for the creation of the new Education Act in an interactive way. The debate ended with the celebration of the National Education Symposium, which saw the participation of 1200 representatives.
- For its achievements, the government has defined a strategic plan and has undertaken to increase the public expenditure in education from 3.3% in 2006 to 4.5% of the GDP by 2009.

Further Reading

- Arocena, R. (2007). Escenarios para el país del futuro. In Arocena, R. and Caetano, G. (eds.) *Uruguay: Agenda 2020*, pp 535–558. Montevideo: Taurus.
- Arocena, R. (2007). Hacia la reforma universitaria. In Arocena, R. (ed.) *Columna periódica del Rector*. University of the Republic. <http://www.udelar.edu.uy> (accessed May, 2009).
- Brovetto, J. (1994). *Formar para lo desconocido: apuntes para la teoría y práctica de un modelo universitario en construcción*. Montevideo: University of the Republic.
- Caraballo, D. (2004). *La formación docente en el Uruguay: evolución, estado de situación y perspectivas*. Venezuela: IESALC/UNESCO.
- Chiancone, A. and Martínez, E. (2007). *La educación superior en Iberoamérica 2006: Informe de Uruguay*. Santiago de Chile: Centro Interuniversitario de Desarrollo.
- De Armas, G. and Filgueira, F. (2007). En busca del tiempo perdido: las tareas del Uruguay social del siglo XXI. In Arocena, R. and Caetano, G. (eds.) *Uruguay: Agenda 2020*, pp 85–132. Montevideo: Taurus.
- Días, M., Lima, L., Ruiz, R., and Trindade, H. (2002). *Evaluación externa del Plan de Desarrollo Estratégico de la Universidad de la República*. Montevideo: University of the Republic.
- Garibaldi, L. (2007). Uruguay 2020: quiénes querrán o podrán aprender qué y para qué. In Arocena, R. and Caetano, G. (eds.) *Uruguay: Agenda 2020*, pp 435–454. Montevideo: Taurus.
- Legislative Power (1985). *Ley de Educación No. 15.739*. Legislative Power. <http://www.parlamento.gub.uy> (accessed May, 2009).
- Mancebo, E. and Torello, M. (2000). *El sistema educativo uruguayo: estudio de diagnóstico y propuesta de políticas públicas para el sector*. Washington, DC: BID.
- Ministry for Education and Culture (2007). *Anteproyecto de Ley General de Educación: Borrador de trabajo*. Montevideo: Ministry for Education and Culture.
- Ministry for Education and Culture (2007). *Anuario Estadístico de Educación*. Montevideo: Ministry for Education and Culture.
- National Administration for Public Education (2000). *Una visión integral del proceso de reforma educativa en Uruguay 1995–1999*. Montevideo: National Administration for Public Education.
- National Administration for Public Education (2005). *Panorama de la educación en el Uruguay. Una década de transformaciones 1992–2004*. Montevideo: National Administration for Public Education.
- National Administration for Public Education (2005). *Proyecto de presupuesto, sueldos, gastos e inversiones 2005–2009*. Montevideo: National Administration for Public Education.
- National Administration for Public Education (2006). *Proyecto de Rendición de Cuentas y Balance de ejecución presupuestal Ejercicio 2005*. Montevideo: National Administration for Public Education.
- National Administration for Public Education (2007). *Balance de dos años de gestión*. Montevideo: National Administration for Public Education.
- National Administration for Public Education (2007). *Gaceta ANEP: Indicadores económicos, sociales y educativos*. Montevideo: National Administration for Public Education.
- National Administration for Public Education (2007). Uruguay en PISA 2006. Montevideo: National Administration for Public Education.
- Pellegrino, A. and Calvo, J. J. (2007). ¿Qué hacer con la población en Uruguay? In Arocena, R. and Caetano, G. (eds.) *Uruguay: Agenda 2020*, pp 61–86. Montevideo: Taurus.
- Rama, G. and Filgueira, C. (1991). *Los Jóvenes uruguayos. Esos desconocidos*. Montevideo: CEPAL.
- University of the Republic (2000). *Plan estratégico de desarrollo de la Universidad de la República*. Montevideo: University of the Republic.
- University of the Republic (2005). *Plan estratégico de desarrollo de la Universidad de la República*. Montevideo: University of the Republic.
- University of the Republic (2005). *Presupuesto Quinquenal 2005–2009*. Montevideo: University of the Republic.
- University of the Republic (2006). *Proyecto de Rendición de Cuentas y Balance de Ejecución presupuestal Ejercicio 2005*. Montevideo: University of the Republic.
- University of the Republic (2007). *Proyecto de Rendición de Cuentas y Balance de ejecución presupuestal*. Montevideo: University of the Republic.
- World Bank (2007). *Uruguay: Equidad y Calidad de la Educación Básica*. Buenos Aires: World Bank.
- Yarzabal, L. (1999). *Consensus para el cambio en la educación superior*. Caracas: IESALC/UNESCO.
- Yarzabal, L. (2005). *Los grandes lineamientos de política educativa para el quinquenio 2005–2009*. Montevideo: National Administration for Public Education.

Relevant Websites

- <http://www.anep.edu.uy> – Administración Nacional de Educación Pública.
- <http://www.bcu.gub.uy> – Banco Central del Uruguay.
- www.fondosolidaridad.org.uy – Fondo de Solidaridad.
- www.latu21.latu.org.uy – Laboratorio Tecnológico del Uruguay.
- www.mides.gub.uy – Ministerio de Desarrollo Social: República Oriental del Uruguay.
- www.mec.gub.uy – Ministerio de Educación y Cultura de Uruguay.
- www.parlamento.gub.uy – Parlamento del Uruguay.
- www.ceibal.edu.uy – Plan Ceibal.
- www.ine.gub.uy – República Oriental del Uruguay Instituto Nacional de Estadística.
- www.ude.edu.uy – Universidad de la Empresa.
- www.udelar.edu.uy – Universidad de la República Uruguay.
- www.um.edu.uy – Universidad de Montevideo.
- www.ort.edu.uy – Universidad ORT Uruguay.
- www.ucudal.edu.uy.

USA

B Pusser, University of Virginia, Charlottesville, VA, USA

M E R Lloyd, College of Charleston, Charleston, SC, USA

© 2010 Elsevier Ltd. All rights reserved.

General Background

The US encompasses approximately 3.54 million square miles of the North American continent, positioned south of Canada and north of Mexico. Its 13 original colonies split from British rule in 1776 and were recognized as an independent, united nation in 1783 under the Treaty of Paris. Today, the US is comprised of 50 states and its capital is the District of Columbia. Its culturally diverse population is estimated to be approximately 303 000 000 people. Currently, the majority of immigrants to the US come from Mexico, the Philippines, Haiti, China, the Dominican Republic, India, Vietnam, Jamaica, Cuba, and South Korea.

The US has a per capita gross domestic product (GDP) of approximately \$46 900, making it one of the world's largest economies. Just over 12% of the population lives below the poverty line, and the country faces a number of challenges that suggest it will be difficult to reduce this percentage. The labor market is becoming more stratified such that those without some college education or professional/technical skills receive lower compensation and employment benefits. The economy also faces significant challenges related to an aging population, financial market crises, and increased medical and retirement expenses.

The first schools in the US varied by geographical region. In the northern colonies, education was first provided in homes for the sake of reading and learning religious principles. By the 1640s, laws were established increasing the number of children attending schools with teachers. In larger northern towns, Latin grammar schools were created for college preparation for boys. Southern colonies relied upon in-home tutoring and private schools for the wealthy.

By 1860, the concept of a tax-supported, free, secular public school system was adopted in the North. Following the Civil War (1865), the South established and accepted this concept for children. Today, a majority of students attend public schools, with private schools continuing to educate a significant portion of the school-aged population.

Major Goals/Historical Shifts in Educational Purposes

Four major educational goals are regarded as paramount in the US today: intellectual development, vocational preparation, training for citizenship, and social mobility. These

purposes, which have often been promoted through a discourse of equality, excellence, and efficiency, have shifted over time in response to social, political, and economic demands. There also have been regular changes in pedagogy and curricula, including shifts in mathematics between the development of basic skills and critical thinking; in literacy education between whole-language and phonics approaches; in early-childhood education between social development and academics; and in higher education between vocational training and liberal arts education.

At the turn of the twentieth century, social and civic goals were regarded as educational responsibilities. Public schools were expected to resolve social conflicts, assimilate immigrants, transition students from school to work, and teach students to be responsible citizens. After the Soviet Union's launching of Sputnik (1957), reforming science and technology education to increase national security became a priority. Subsequent responses, including the National Defense Education Act (1958), continue to shape the evolution of US education.

In the 1960s, legislations such as the Civil Rights Act, the Economic Opportunity Act, the Elementary and Secondary Education Act, and Head Start brought a shift from quality for the talented to equality for all. Throughout the 1970s, efforts were made to enhance curricular relevance, improve teaching, imbue textbooks with more gender-neutral and racially inclusive material, and more fully involve parents and communities in the educational process.

In 1983, following a report by the Commission on Excellence in Education, the public again turned attention to the quality of schooling. Policymakers issued calls for programs to achieve both equality and excellence, with emphasis on preserving federally funded programs directed at improving equality in schools. The next two decades witnessed a political struggle over school choice, organization, and finance. This struggle led to new forms of standards-based assessment and accountability, as well as increased numbers of alternative and charter schools. The political struggle also contributed to the passage in 2002 of the largest federal school reform legislation in the nation's history, the No Child Left Behind Act (NCLB). The goals of the act were to decrease disparities in educational outcomes and increase academic achievement at the elementary-secondary level through stricter school and teacher accountability measures and curricular standards.

Structure and Operation of the Education System

Preschool and Early-Childhood Care

In 1965, the federally funded program Project Head Start was established. The program continues to provide support for preschoolers (typically ages 3–4) characterized as being at risk of social and/or cognitive developmental delays. Despite a lack of funding for all preschool-aged children, since 1970, enrolment rates in center-based early education, including those subsidized by Head Start, have increased from 20% to 56%. In 2005, there were approximately 4 million children enrolled in nursery school, about 49% of eligible children.

Elementary and Secondary (K-12) Education

Attendance laws for elementary and secondary education vary by state such that compulsory schooling begins between ages 5 and 8, and ends between ages 14 and 18. School days average 7.4 h, and the school year lasts 181 days. In 2005, 55.2 million full-time students enrolled in grades PK-12 in a reported 97 382 public and an estimated 34 000 private schools. Of these students, 49.1 million enrolled in public elementary and secondary schools; and 6.1 million enrolled in private elementary and secondary schools. They were taught by some 3.6 million teachers.

Enrolment rates, both in public and private schools, vary by geographic region. In 2005, the South served the most public school students in the US, and the Northeast had the highest regional concentration of private school enrolments. In 2005, minority enrolments constituted 42.4% and 24.6% of public school and private school enrolling populations, respectively.

English-language learners (ELLs) and students with disabilities also contribute to the heterogeneity of school populations. In 1975, The Individuals with Disabilities Education Act (IDEA) was passed and required that public schools serve children aged 3–21 with disabilities. In the fall of 2006, approximately 6.7 million students – 13.5% of enrolled public students – received IDEA services.

In 2006, 10.8 million students spoke a language other than English – most commonly Spanish – in their homes. Of these children, 2.7 million spoke English with difficulty. The number of students speaking another language in the home has nearly tripled since 1979. In order to serve the needs of this population, schools receive federally assisted funding for students with limited English proficiency (LEP).

Elementary education begins in prekindergarten or kindergarten and extends through 8th grade. In 2005, the average student/teacher ratio was 15.8, varying by the size of the elementary school, often with lower ratios in smaller schools. The elementary curriculum varies by

state but generally includes mathematics, science, language arts, social studies, computer/technical skills, and physical education/health.

Secondary education begins in the 9th grade, around age 14, and extends through 12th grade. In 2005, the average student/teacher ratio was 16.6, also varying by school size. The requirements for completing secondary school and earning a standard diploma differ by state and by type of school or program (see **Figure 1**). Graduation rates vary by race, gender, and geographic location. In 2005, some 3.3 million students completed high school, 300 000 of whom were in private schools. The 4-year high school completion rate for all students that year was 74.7%. Another 400 000 individuals completed high school through passing the General Educational Development exam in 2005.

Other Schooling Arrangements

While a majority of school-aged children attend regular public schools, students may also choose a charter, magnet, or private school; be homeschooled; and/or receive supplemental educational services. In 2005, just over 1 million students enrolled in 3780 charter schools. Charter schools are publicly funded but free from many state and local regulations. These schools operate under a charter with the state establishing the curricular and organizational standards for which the school is accountable. In 2005, 2.1 million students enrolled in 2736 publicly funded magnet schools. Funding sources, curricular foci, operational structures, and/or enrolment qualifications vary among magnet schools. Students who spend less than 25 h per week in a private or public school and are educated in part in the home by parents, tutors, or Internet programs are classified as homeschooled. In 2003, 1.1 million students, 2.2% of the school-aged population, were homeschooled.

Teaching and Professional Staff

During 2003–2004, teachers accounted for approximately 57% of the total public school staff and 59% of total private school staff. In public elementary and secondary schools, principals, instructional coordinators and supervisors, librarians and media specialists, as well as counselors, comprised 7% of the total staff. Those same groups comprised 12% of the total staffing of private schools. Nurses, social workers, psychologists, and speech therapists made up approximately 5% of staff in both public and private schools. Instructional and noninstructional aides represented 13% of public and 7% of private school staff. The remaining 18% of both public and private school staffs included clerical support staff and food service, custodial maintenance, and security personnel.

Full-time teachers were predominantly female (74.8%). Among secondary schools, this percentage was lower – 56.5%; while among elementary schools, the percentage was 84.1. Of this, 78.7% were between the ages of 30 and 59. While private schools had a greater percentage of white and Asian teachers, the racial distribution of full-time elementary–secondary school teachers was as follows: 83.3% white, 7.8% black, 6.2% Hispanic, 1.4% Asian, 0.2% Pacific Islander, 0.5% American Indian/Alaska Native, and 0.7% identifying with more than one race.

The US system for preparing teachers for the classroom includes 4-year undergraduate programs in which bachelor's degrees in education are earned; 5-year undergraduate/graduate programs in which bachelor's degrees in a content area and masters of teaching/education are earned; and alternative pathways that may not require traditional

postsecondary coursework, especially in high-demand content areas. During 2003–2004, 1.3% of public school teachers had no degree, 0.4% had an associate's degree, 52.1% had a bachelor's degree, 39.3% had a master's degree, and 6.8% had a specialist or doctoral degree. Within private elementary schools, the percentages of teachers having no degree or a bachelor's degree were higher (5.6% and 65.1%, respectively), while the percentage of those with master's degrees was lower (23.3%).

Teacher salaries generally increase with increases in educational attainment and years of experience. In 2003–2004, public and private school teachers earned on average \$47 700 and \$34 700, respectively.

Teacher attrition is a significant concern for elementary–secondary education. During 2003–2004 and 2004–2005, 8.1% and 5.9% of public and private school teachers,

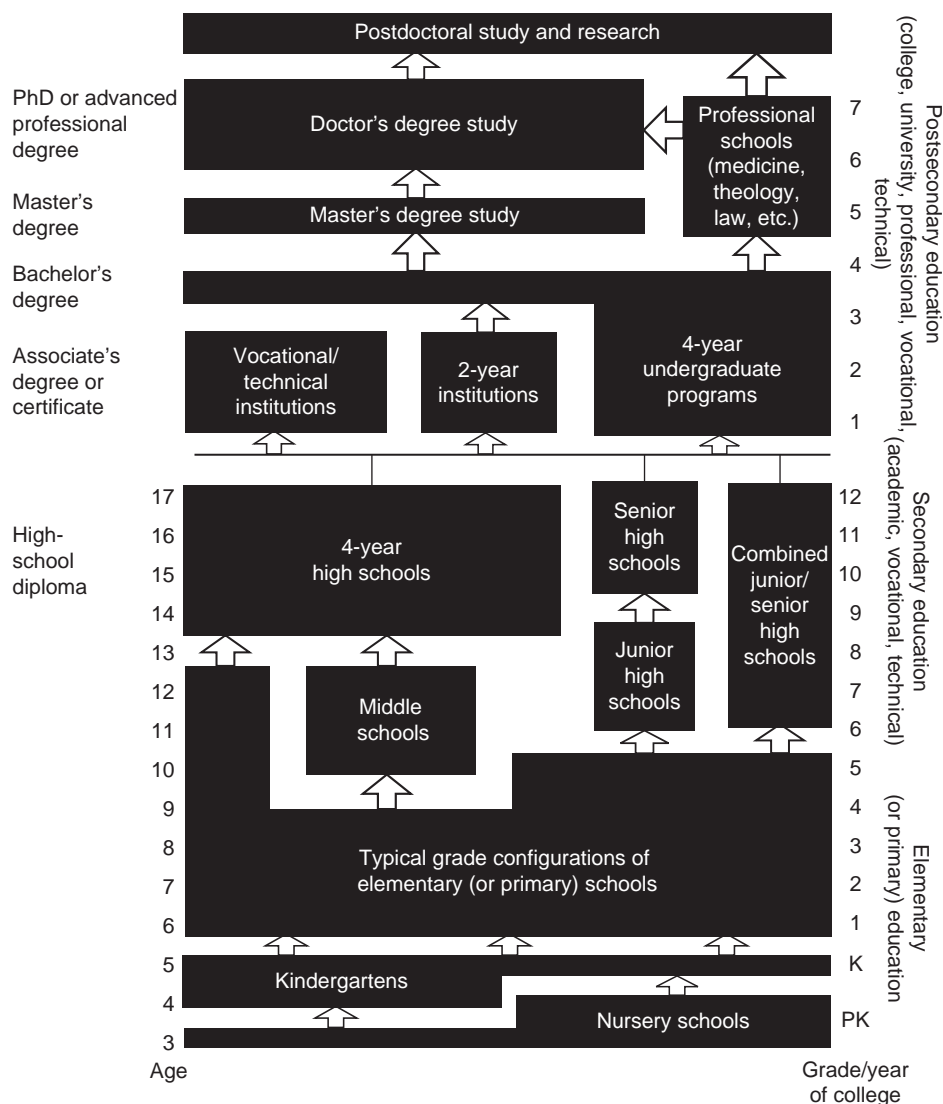


Figure 1 The structure of education in United States. From US Department of Education, National Center for Education Statistics (2007). Digest of Education Statistics: figure 1.

respectively, transferred to different schools and 8.4% of public school teachers and 13.6% of private school teachers left the profession. Schools in urban areas were most affected by attrition, as disproportionately high percentages of teachers transfer or leave urban schools. The same holds for teachers of ELLs, special education, mathematics, and science.

Administrative and Supervisory Structure

Governance at all levels of the educational system is highly decentralized. For public and private elementary education, state and local authorities play a major role in policymaking, regulation, and finance. Through the Department of Education, federal regulations and subsidies also influence school district and school-site behaviors. Public elementary and secondary schools are generally organized into districts governed by boards and administered by superintendents. At school sites, principals are chief administrative officers in public schools, and a principal or headmaster holds a similar position in private elementary and secondary schools.

Public postsecondary institutions are generally directly overseen by a board of trustees or regents, typically made up of appointed members and ex-officio members, such as elected officials and representatives of alumni associations. Boards may oversee single campuses or entire state systems of institutions. Boards have fiduciary responsibility for postsecondary institutions and hire the president or chief executive officer. Except where public institutions have codified autonomy under state charter, the state legislature, governor, and key political leaders play important roles in shaping governance policy, particularly over tuition setting and admissions. Private university boards also have fiduciary responsibility and hire presidents. To a greater degree than for public boards, private trustees are central to university development and fundraising activities.

Educational Finance

Educational expenditures by public and private institutions totaled \$972 billion for school year 2006–2007, about 7.3% of GDP. Of that total, \$373 billion, about 2.8% of GDP, was spent by postsecondary degree-granting institutions. Public elementary–secondary schools spent \$599 billion, while private schools at the elementary–secondary level had expenditures of \$46 billion in 2006–2007. Public postsecondary degree-granting institutions had expenditures of \$239 billion compared with \$134 billion at private postsecondary degree-granting institutions in that same year. The federal government provided approximately \$170 billion for education in 2006. Of that total, \$72.5 billion was spent on elementary–secondary education, \$59 billion was provided to postsecondary education, and \$31.4 billion was directed to postsecondary

research. State and local governments provided \$473.5 billion in current expenditures and capital outlay for elementary–secondary education in 2004–2005. This funding amounted to approximately \$1600 per capita and some 23.6% of total expenditures by state and local governments. These governments also provided \$182.2 billion in current expenditures and capital outlay for colleges and universities in that year, representing approximately \$615 per capita and 9.1% of state and local expenditures on all functions.

A major source of revenue for postsecondary education comes from student tuition and fees. In 2006–2007, the average annual tuition and fees were \$5685 at public 4-year institutions, \$2017 at 2-year nonprofit institutions, and \$20492 at private 4-year colleges and universities. In fiscal year (FY) 2006, tuition revenue totaled \$34.5 billion (17% of total revenue) for public 4-year institutions. Other key sources of institutional revenue for 4-year nonprofit institutions were state appropriations, \$45.5 billion (22.5% of total), and federal grants and contracts of \$25.5 billion (12.6% of total revenue). In that same year, private nonprofit 4-year institutions received \$43.9 billion in tuition and fee revenues (some 28.9% of their total revenue). Revenue for private, nonprofit 4-year institutions in 2006–2007 also came from federal grants and contracts, \$19.6 billion (12.9% of total revenue), private gifts, grants and contracts, \$18.2 billion (12.0% of total revenues) and return on institutional investments, \$35.6 billion (23.4% of total revenues). Two of the largest providers of federal research funding to both public and private research universities were the National Science Foundation and the National Institutes of Health. At 2-year public institutions, income from tuition and fees totaled \$7.2 billion in FY 2006 (some 16.6% of total revenue). State appropriations to that sector totaled \$13.1 billion (30.1% of total revenue), with local appropriations adding \$7.9 billion (18.1% of total revenue). Federal grants and contracts to 2-year public nonprofit institutions, often linked to workforce development and vocational training programs, provided \$4.7 billion, representing 10.9% of total revenue. Private nonprofit 2-year institutions drew \$318.4 million from tuition, 53.6% of total revenue, and \$75.4 million, 12.7% of total revenue from federal appropriations.

The US system of higher education depends to a significant degree on federal, state, and institutional student aid, which totaled \$130.5 billion in 2006–2007. Of that total, federal grants to postsecondary students amounted to more than \$20.7 billion; federal loans comprised \$59.5 billion, institutional grants provided \$26.3 billion, and private loans totaled \$17.1 billion.

Contemporary Assessment

The primary source of national data on elementary–secondary performance for the past 40 years has been the

National Assessment of Educational Progress (NAEP). NAEP tests are administered to public and private school students in the 4th, 8th, and 12th grades in a number of subject areas, including reading, writing, US history, mathematics, and science. Longitudinal federal NAEP data on student performance at ages 9, 13, and 17 in mathematics and science are also collected. Long-term, NAEP results have shown mixed performance with some improvement in both reading and mathematics but considerable variation on the basis of various demographic variables. In the Organization for Economic Cooperation and Development (OECD)'s 2006 Programme for International Student Assessment (PISA) study of the knowledge and skills in science and mathematics of 15-year-olds in 57 countries, the percentage of students in the US at both the highest and lowest levels of proficiency were at or above the average. Overall, the US was ranked statistically significantly below the average on both the science and mathematics scales.

Under NCLB, each school must provide detailed school-level information on the number of students who meet state proficiency standards in areas such as mathematics, language arts, and science. The goal of the Act is that all schools will meet 100% proficiency by the 2013–2014 school year. Those schools that consistently fail to make adequate yearly progress toward meeting state proficiency standards are initially required to develop improvement plans and ultimately subject to restructuring and the imposition of alternative governance plans.

Over the past two decades, assessment has been one of the most politically contested issues at every level of education. Driven by political demands for new forms of access and finance at the elementary–secondary level and by neo-liberal models calling for greater efficiency, effectiveness, and institutional competitiveness in the postsecondary arena, assessment promises to continue to generate conflict in the years to come.

Higher Education

The hallmark of the higher education system has long been the diversity of its institutional types and degree programs. From its origin in the public and private colonial colleges opened in the seventeenth century, the system has evolved to encompass open-access public and private 2-year and 4-year institutions, selective 4-year public and private colleges and universities, and a wide range of proprietary institutions. The private sector is particularly diverse, incorporating small 2-year and 4-year nonprofit institutions, liberal arts colleges, as well as large research universities and for-profit colleges.

In Fall of 2006, there were 17.8 million students enrolled in US degree-granting higher education institutions, 6.9 million in public 4-year institutions, 6.2 million in public 2-year institutions, 4.3 million in private 4-year

institutions, and 293 000 in private 2-year colleges (Tables 1 and 2).

Over the past three decades, the postsecondary system has enrolled an increasingly diverse cohort of students, including a significant proportion of women, adult students, international students, and part-time learners (Table 3). In 2006–2007, minorities comprised 27.9% of enrolments in public 4-year institutions, 37.2% of those in public 2-year institutions, and 27.8% of those enrolled in private 4-year institutions. The percentage of minority students enrolled in higher education increased from 16.1 in 1980 to 30.9 in 2005. In 1980, women comprised 42% of total postsecondary enrolments. By 2006, 58% of those enrolled in higher education were women. In 2006–2007, there were just over 500 000 international students enrolled for the first time in US higher education.

Table 1 Higher education availability

Number of degree-granting institutions in the United States

	Number of institutions		% of Total	
Public	1688		39	
2-year		1045		62
4-year		637		38
Private (not-for-profit)	1615		38	
2-year		107		7
4-year		1508		93
Private (for-profit)	986		23	
2-year		533		54
4-year		450		46
Total	4289		100	

From US Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), Spring 2007, Enrollment component. Table 2.

Table 2 Distribution of students in higher education

Number of students enrolled in higher education

	Number of students		% of Student population	
Undergraduate	15 630 574		85.85	
Full-time		9 929 944		63.50
Part-time		5 700 630		36.50
Graduate	2 231 454		12.26	
Full-time		1 077 397		48.30
Part-time		1 154 057		51.70
Professional	343 446		1.89	
Full-time		309 158		90.00
Part-time		34 288		10.00

From US Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), Fall 2006, Institutional Characteristics component. Table 3.

Table 3 Student demographics in higher education

	<i>Number of students</i>	<i>% of student population</i>
Gender		
Male	7 715 801	42
Female	10 489 673	58
Age		
Under 18	577 163	3.7
18–24	9 830 305	64
older than 25	4 903 941	32
Race/ethnicity		
White (non-hispanic)	10 897 048	59.9
Black (non-hispanic)	2 207 274	12.1
Hispanic	1 897 258	10.4
Asian/Pacific Islander	1 081 628	5.9
American Indian	174 936	1.0
Unknown	1 346 605	7.4
Nationality		
International students		3.3
Internal students (American citizens)		96.7

From US Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), Spring 2006, Enrollment component. Table 37; and US Department of Education, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), Spring 2007, Enrollment component. Table 2.

The US's share of about 25% of all international postsecondary students was the highest in the world. Over 40% of international students studying in the US were from India, China, South Korea, or Japan.

Both public and private postsecondary institutions also offer credentials and degrees in a variety of formats, including Internet-based programs. Degrees include the associate of arts and associate of science, bachelor of arts, bachelor of science, master's degrees, doctoral degrees, and a variety of professional degrees. Students in virtually all types of accredited postsecondary institutions are eligible to seek need-based federal aid and federal student loans. In order to receive various forms of federal financial aid, students must attend colleges and universities that have either regional or national accreditation.

The past three decades have also seen a rapid growth in the endowments of the most prestigious public and private 4-year institutions, along with rapid increases in student tuition. Since World War II, there has also been a massive growth in funded research at the leading research universities. Support for research and development at colleges and universities increased from \$6.9 billion in 1966 to just under \$43 billion in 2004 (all 2004 dollars). The postsecondary system also has the world's most comprehensive system of competitive athletics. In 2006–2007, over 400 000 students participated in competitions sanctioned by the National Collegiate Athletic Association and other governing bodies.

Assessment and accountability have recently become significant issues in the political economy of US higher education. Public institutions, in collaboration with state legislatures, coordinating boards, and regional or national

accreditation agencies, have long held a primary role in determining admissions standards, curricula, standards for credit and transfer of credit, faculty hiring and tenure standards, and student graduation requirements. Private institutions have had similar levels of autonomy with less direct intervention by state or federal elected officials. Despite intensified lobbying efforts and a national review of higher education commissioned by the Secretary of Education, few changes were made to norms of assessment and accountability in the reauthorization of the Higher Education Act signed into law in 2008.

Adult and Nonformal Education

The education system also offers a significant number of access points for adult learners. Adults may complete secondary school through passing a series of tests that lead to a General Educational Development (GED) credential. Many vocational and technical training programs are available for adults in public and private institutions, often supported with state- and federal-funding programs such as Title I of the Workforce Investment Act. Traditionally, the majority of adult learners in postsecondary education has attended community colleges and proprietary schools, or comprehensive 4-year colleges. Until quite recently, few adults have enrolled as undergraduates in residential-life colleges or highly selective 4-year universities. The 1990s witnessed a dramatic increase in adult enrolments in degree-granting proprietary colleges. Significant attention has also been turned to adult education as a vehicle to address a growing skills gap in the

twenty-first-century economy, given predictions that over the next two decades the US will need to substantially increase the number of workers with baccalaureate degrees. Currently, some 4.9 million adult learners are enrolled in degree-granting programs in US higher education.

Emerging Issues

Perhaps no factor has had more influence on education in the US over the past three decades than the rise of neoliberal and market-based approaches to the finance and organization of education. As implemented by conservative administrations at the state and federal levels, the neoliberal project has advocated increased market competition, outsourcing, and consumer choice in key public sector institutions and enterprises. In the elementary–secondary system, this has led to calls for more administrative and management systems modeled after corporate and private sector systems, privatization of school functions, voucher systems for educational finance, and support for charter and other alternative schools.

At the postsecondary level, neoliberal policies have promoted a decrease in state support for higher education, higher tuitions, and market models for the provision of higher education. The increasing pressure to prepare students for global economic competition has led to significant changes in the organization and curriculum of community colleges in particular.

National and international competitiveness remains a key topic in the political economy of education. This has inspired calls for increased education and training for both national economic competitiveness and individual income and social mobility. In 2007, the US ranked 10th among industrialized nations in the percentage of its citizens aged 25–34 with college degrees. In 2004, the median annual income for workers aged 18 and over was \$24 108. For those who had not completed high school, the median income was only \$12 437. For those who had completed high school, earned an associate's degree, or earned a bachelor's degree, median income was \$20 733, \$30 026, and \$38 880, respectively. For those who had earned a master's degree, a professional degree, or a doctoral degree, median income in 2004 was \$50 693, \$74 207, and \$72 073, respectively. As these data suggest, the foremost challenge for the nation remains creating equal educational outcomes for all.

See also: Access and Equity in Higher Education; Adult Basic Education: A Challenge for Vocational Based Learning; Competition and Student Performance; Curriculum and the Education of Cultural and Linguistic Minorities; Data; Economic Approaches to Teacher Recruitment and Retention; Education and Inequality; Educational Privatization; Emergence of For-profit Higher

Education; Evaluation and Accountability; Experienced Teachers' Craft Knowledge; Higher Education: An Overview; Higher Education and the Labor Market; Home-schooling; Impact of Assessment on Classroom Practice; Internationalization of Higher Education; Management of and in Higher Education Institutions; Participation in Adult Learning; Public Policy and Inequality in Postsecondary Opportunity; Educational Statistics and the Failure of Education Reform; School Finance: An Overview; Steering of Higher Education Systems – The Role of the State; Student Test Results in School Accountability; Teacher Education and the Educational Foundations Knowledge Base; Teacher Supply; The Changing Role of the State in Higher Education; The Economics and Finance of Higher Education; The Economics of Charter Schools; The Economics of School Accountability; The Economics of Tuition and Fees in Higher Education; Viewing Private Higher Education: How Much, Where, Why, and What?; Wider Benefits of Adult Education.

Further Reading

- Beattie, I. R. and Arum, R. (2000). *The Structuring of Schooling: Readings in Sociology of Education*. Mountain View, CA: Mayfield Publishing Company.
- Central Intelligence Agency (2008). *The World Factbook: The United States*. <https://www.cia.gov> (accessed July 2009).
- Cubberley, E. P. (1935). Our educational evolution. In Cubberley, E. P. (ed.) *An Introduction to the Study of Education*, pp 1–23. Boston, MA: Houghton Mifflin.
- Gaquin, D. A. and Ryan, M. M. (eds.) (2007). *The Almanac of American Education*. Lanham, MD: Bernan Press.
- Geiger, R. L. (2004). *Knowledge and Money: Research Universities and the Paradox of the Marketplace*. Stanford, CA: Stanford University Press.
- Goodlad, J. I. (1984). *A Place Called School: Prospects for the Future*. New York: McGraw-Hill.
- Healy, R. (2007). Chronicle of Higher Education Almanac Summary (*The Ohio State University, Institutional Research Reports*). http://oaa.osu.edu/irp/Almanac_Summary_08.pdf (accessed July 2009).
- Hoffman, L. (2007). *Numbers and Types of Public Elementary and Secondary Schools from the Common Core of Data: School Year 2005–06* (NCES 2007–354rev). Washington, DC: National Center for Education Statistics, U.S. Department of Education. <http://nces.ed.gov> (accessed July 2009).
- Hoy, W. K. and Miskel, C. G. (eds.) (2008). *Educational Administration: Theory, Research and Practice*, 8th edn. New York: McGraw-Hill.
- Hunt, T. C. (2002). *The Impossible Dream: Education and Search for Panaceas*. New York: Peter Lang.
- Institute of International Education (2008). *Open doors: 2007 Fast Facts*. <http://www.opendoors.iienetwork.org> (accessed July 2009).
- Lazerson, M. (1987). *American Education in the Twentieth Century: A Documentary History*. New York: Teachers College Press.
- Oakes, J. and Lipton, M. (2003). *Teaching to Change the world*, 2nd edn. Boston, MA: McGraw-Hill.
- Organisation for Economic Cooperation and Development (2006). *PISA 2006 science Competencies for Tomorrow's World*. <http://www.oecd.org> (accessed July 2009).
- Planty, M., Hussar, W., Snyder, T., et al. (2008). *The Condition of Education 2008* (NCES 2008–031). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Snyder, T. D., Dillow, S. A., and Hoffman, C. M. (2008). *Digest of Education Statistics 2007* (NCES 2008–022). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Tyack, D. (1974). *The One Best System*. Cambridge, MA: Harvard University Press.

Tyack, D. and Cuban, L. (1995). *Tinkering Toward Utopia*. Cambridge, MA: Harvard University Press.

U.S. Census Bureau, Housing and Household Economic Statistics Division (n.d.). *Poverty: 2007 Highlights*. <http://www.census.gov> (accessed July 2009).

U.S. Department of Education (n.d.). School choices for parents. <http://www.ed.gov> (accessed July 2009).

U.S. Department of Education, National Center for Education Statistics (2007). *The Condition of Education 2007 (NCES 2007-064)*. Washington, DC: U.S. Government Printing Office.

Relevant Website

<http://www.ed.gov> – US Department of Education: NCLB.

Zanzibar

I M Hija, Ministry of Education and Vocational Training, Zanzibar, Tanzania

© 2010 Elsevier Ltd. All rights reserved.

General Background

Zanzibar consists of two main islands, Unguja and Pemba, and several other smaller islands some of which are uninhabited. Zanzibar is located in the Indian Ocean, about 30 km off the East Coast of Africa between latitudes 5° and 7° south of the equator. The island has a total area of 2654 km²; out of this, Unguja, which is the largest, has an area of 1666 km² while Pemba has an area of 988 km².

According to the 2002 population and housing census, Zanzibar had a population of 981 754 with a growth rate of 3.1% and a population density of 370 per square kilometer. Of the total population, 40% lived in urban area and the remaining 60% settled in rural areas. The outburst of the population growth rate was mostly attributed to high fertility rate of 5.3. The projected population in 2005 was 1 072 000.

Zanzibar is part of the United Republic of Tanzania, but is semiautonomous. It has its own government, a legislative assembly known as the House of Representatives, the Executive, headed by the President of Zanzibar and its own judicial system. Zanzibar is divided into five administrative regions (three in Unguja and two in Pemba), ten districts two in each region, and 50 constituencies.

Zanzibar's major economic sectors include agriculture, trade and industries, and tourism. Agriculture is the mainstay of the economy largely due to the clove industry, which is the main foreign currency earner for the Indian Ocean islands of Zanzibar and Pemba. Historically, trade has been second to agriculture but many years of isolation and the socialist policies adopted after the 1964 revolution completely undermined its potential in Zanzibar economy. Recently, tourism has emerged as a possible successor to the ailing clove industry.

Soon after the 1964 Revolution, education was proclaimed free to all Zanzibaris irrespective of color, creed, or gender. Much has changed in the intervening years, a number of policy documents have come up, ostensibly addressing, in the main, the issues of access, equity, and quality. In 1991, the government issued the Zanzibar education policy, which articulated key sectoral objectives and targets. This document was amended in 1995 in order to incorporate a number of goals pronounced in important international conventions and declarations.

In order to implement the amended 1995 policy, in 1996, the Ministry of Education issued the 10-year Zanzibar education master plan (ZEMAP). The key issues

which featured in ZEMAP included those pertaining to access and equity, quality, relevance, promotion of science and technology, and expansion of the provision of early-childhood education.

Vision and Mission of Education

The vision and mission statements of education are:

- Vision: a democratic and peaceful society enjoying a high quality of education and livelihood and committed to lifelong learning to effectively respond to development challenges.
- Mission: to strive for equitable access, quality education for all, and promotion of lifelong learning.

Zanzibar Education System Structure

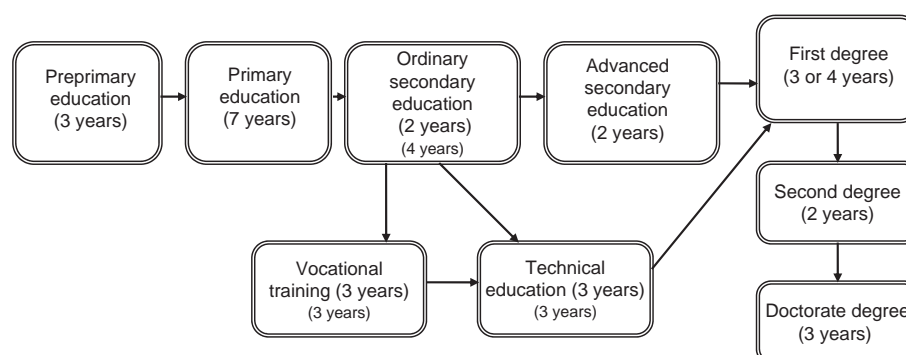
The system is characterized by the existence of several levels (**Case study 1**). These include primary levels which are of 7 years duration, a secondary level which has two channels, one which is subdivided into a first cycle of 3 years and a second cycle of 2 years, and the other a full uninterrupted secondary cycle of 4 years. Both these lead to high school level of 2 years or technical education of 3 years and 3 or more years of higher education.

Early-Childhood Education

Early-childhood education is provided to children of 4 and 6 years by the public preschools, private organizations, associations, and individuals. Currently, there are about 181 preschools providing early-childhood education including 24 public and 154 private. In 2007, 21 538 or 22.2% of the children of this age were enrolled.

Primary Education

Primary education is part of universal free basic education offered to children within the age group of 7–13 years. This level covers the first 7 years of basic education. In 2007, 214 096 pupils, equivalent to 104.6% gross enrolment ratio were enrolled in 224 public and 48 private schools. Rapid increases in enrolment have been possible through community initiatives in the construction of classrooms (**Tables 1 and 2**). Although primary school is free and

Case study 1 Zanzibar formal education system**Table 1** Growth of the number of school-aged children, 2003–2007

Age (years)	Level of education	2003	2004	2005	2006	2007	Growth per year (%)
4–6	Preprimary	91 372	91 254	90 478	88 625	96 925	3.1
7–13	Primary	199 359	201 809	204 138	207 204	204 609	3.1
14–16	Orientation secondary class/ Form 1–Form 2	73 921	77 311	80 092	82 120	83 491	3.1
17–18	Form 3–Form 4	44 177	45 502	46 867	51 058	52 195	3.1

Source: Minister of Education and Vocational Training's Revenue and Expenditure Budget Speech and are open to public after being released to the House of Representatives, 2007/2008.

Table 2 Gross-level enrolment ratio at different levels of education, 2003–2007

Level of education	2003	2004	2005	2006	2007
Preprimary	14.7	13.8	15.9	13.5	22.2
Primary	99.1	100.3	101.3	98.5	104.6
Lower secondary A (OSC–Form 2)	63.5	68.0	69.0	73.9	68.6
Lower secondary B (Form 3–Form 4)	20.6	22.5	27.4	29.9	36.5
Primary and lower secondary A (Primary 1–Form 2)	89.6	91.7	92.6	91.9	94.2
Primary and lower secondary B (Primary 1–Form 4)	79.7	81.8	83.3	83.0	85.4

Source: Minister of Education and Vocational Training's Revenue and Expenditure Budget Speech and are open to public after being released to the House of Representatives, 2007/2008.

compulsory, the net enrolment is relatively low. This implies that there are a number of children of school-age who are either not in school or over-aged.

Secondary Education

Secondary education at the lower level is characterized by a two-pronged system. One prong is of 5 years subdivided

into 3 years of first cycle lower secondary at the end of which students sit for entrance examinations for admission into the secondary cycle of lower secondary lasting for 2 years. At the end of this second cycle, students sit for the certificate of secondary education examination (CSEE). In 2007, 73 417 students were enrolled in 69 public secondary schools, while 3630 were enrolled in private secondary schools (Table 3).

Postbasic Education

The overall aim of the postbasic education is to increase the opportunity for further learning and to prepare students to enter professional careers or acquire marketable skills. It also instills desire for lifelong learning.

Postbasic opportunities include entrance to second phase of secondary education for the successful candidates in the main stream. The majority of students who fail the form two national examinations either join continuing-education classes, vocational training institutions, or join the army of unemployed basic-education graduates.

There are now a number of postbasic education institutions in Zanzibar, both public and private, that offer a range of learning opportunities in both vocational and academic fields.

Table 3 Pupil–teacher ratio at the level of primary and lower secondary education, 2003–2007

Year	<i>Public schools</i>			<i>Private schools</i>			<i>Total</i>		
	<i>No. of students</i>	<i>No. of teachers</i>	<i>Students to teachers</i>	<i>No. of students</i>	<i>No. of teachers</i>	<i>Students to teachers</i>	<i>No. of students</i>	<i>No. of teachers</i>	<i>Students to teachers</i>
2003	230 603	6926	33	6639	467	14	237 242	7393	32
2004	243 092	7633	32	6795	545	12	249 887	8178	31
2005	252 205	8790	29	8410	545	15	260 615	9335	28
2006	258 292	8547	30	8735	508	17	267 027	9027	29
2007	263 222	8448	31	9745	540	18	272 967	8988	30

Source: Minister of Education and Vocational Training's Revenue and Expenditure Budget Speech and are open to public after being released to the House of Representatives, 2007/2008.

Higher Education

The aim of higher education is to equip learners with high level I in intellectual, professional, and managerial capacities necessary for high-level performance. It offers opportunities for decision makers, professional cadres, and for leaders in public and private sectors to acquire new development skills, innovation, and ability to make informed decisions.

Higher-university education is provided in one public university and two private universities. All of these universities are autonomous, self-accrediting institutions responsible for their own governance and coordinated by Tanzania Commission for Universities. Other tertiary institutions in Zanzibar include: Zanzibar Institute of Finance Administration, College of Health Sciences, Kizimbani Institute of Agriculture, and College of Hotel and Tourism. There are also a number of Zanzibari students pursuing courses in Tanzania Mainland higher-learning institutions. In addition to that, there are opportunities for higher education provided through the Open University of Tanzania and outside Tanzania (Table 4).

Education for Learners with Special Needs

Learners with special needs include those with different kinds of disabilities, slow learners, and those who are exceptionally gifted.

The nature and extent of disability among school-going children has not been fully documented. However, impairment of hearing, speech, and sight are the most common-known disabilities. In 2007, there were 46 schools with the classes for children with special needs, with the total enrolment of 1523 pupils. However, this number does not include slow learners and exceptionally gifted ones. Recently, inclusive education has been promoted to ensure that the children with special needs get equal opportunities, barriers to learning are addressed, and the diverse range of learning needs are accommodated. Inclusive-education curriculum in teacher training was introduced and in 2007, a total of 290 teachers were trained on the tactics of teaching in the classes of learners with special needs.

Table 4 Enrolment of Zanzibaris in tertiary and higher education institutions in Tanzania and overseas

<i>Institution</i>	<i>Male</i>	<i>Female</i>	<i>Total</i>
State University of Zanzibar	250	191	441
University College of Education	319	220	539
Zanzibar University	251	90	341
University of Dar es Salaam	6	4	10
Institute of Finance Management	16	10	26
Business College	9	4	13
The Social Work Institute (SWI)	4	2	6
Hubert Kairuki Memorial University	8	3	11
Institute for Rural Development and Planning Dodoma	8	3	11
Dar es Salaam Institute of Technology	7	0	7
University of Mzumbe	1	1	2
Islamic College Morogoro	1	0	1
The Dar es Salaam College of International Relations Diplomacy	0	1	1
Kivukoni College	4	2	6
Bagamoyo Secretarial College	3	2	5
Institute of Judicial Administration (IJA)	3	1	4
Mbeya Institute of Technology	1	0	1
The Dar es Salaam Secretarial College	0	1	1
Tumaini University	0	1	1
Dar es Salaam School of Accountancy	1	0	1
The Primary Health Care Institute (OHCI) Iringa	0	1	1
The Arusha School of Journalism and Mass Communication	0	3	3
Overseas institutions	124	24	148
Total	1016	564	1580

Source: Ministry of Education and Vocational Training's Education Management Information System, June 2007.

Technical and Vocational Education and Training

The need for technical skills in wealth creation is increasing where the application of science and technology is becoming a must. Hands-on learning and knowledge in

mechanization and automation is constantly demanded in the labor market. One way of getting skilled laborers is by providing them with technical and vocational education and training. Technical and vocational education and training is a necessary ingredient toward poverty reduction as it provides opportunity for self-employment.

At present, there is a very limited capacity for technical and vocational education and training in the country. There are few well-equipped institutions which offer this type of education/training in Zanzibar. In 2007, there were four technical secondary schools and one technical college with a total of 644 students, of which 33% of this total enrolment were girls. In addition, there were five formal institutions providing vocational education and training and 68 informal vocational training centers with the total of 1325 learners.

The main fields offered in those institutions are agriculture and fishing industry, construction industry, manufacturing industry, repair and service industry, and trade and service industry.

Nonformal and Adult Education

Nonformal education is designed not only to eradicate illiteracy, but also to increase efficiency and productivity. It plays a major role in poverty reduction. Since nonformal education is a lifelong learning, it needs to be dynamic and flexible to respond to learners' needs.

Adult education is generally provided to adults who because of one reason or another are not reached by the formal system of education and to those who would like to advance their education and skills.

These types of education are provided through a number of programs such as literacy programs, continuing education, vocational training, and alternative education.

Adult education is provided by both public and private institutions. In 2007, there were 61 literacy centers with an enrolment of 5215 learners.

Alternative Education

Basic education is a right which the state is duty bound to make accessible to all children, pupil who for some reasons have missed out on basic education or are unable to complete basic education through normal schooling. They need to be provided with alternative routes to complete their basic education.

The alternative learning program is still at infancy stage and has not spread throughout the country. In 2007, there were 18 centers with an enrolment of 760 learners and 317 learners were mainstreamed into the normal school classes (**Case study 2**).

The Teachers and Teaching Profession

Efficient teacher recruitment, proper deployment, and utilization are necessary to ensure good professionalism. Clearly defined criteria and guidelines to identify, attract, train, and retrain good teachers must be in place so as to facilitate the selection of trainable teachers capable of preparing students to successfully face the emerging local and global challenges.

In 2007, there were 10 520 teaching staff in the public sectors and 540 teachers in private sector of whom 57.7% were females. While the system seems to have enough number of teachers, their distribution or deployment is questionable. Even though the recruitment of teachers for public schools is centralized, the efficiency of the ministry's information management system very often leads to the recruitment and posting of teachers with little regards to demand.

Although the number of teachers has been increasing, the increase has been skewed in favor of some subjects, English, mathematics, science, and technology subjects remain understaffed, resulting in an unfair distribution of teaching load.

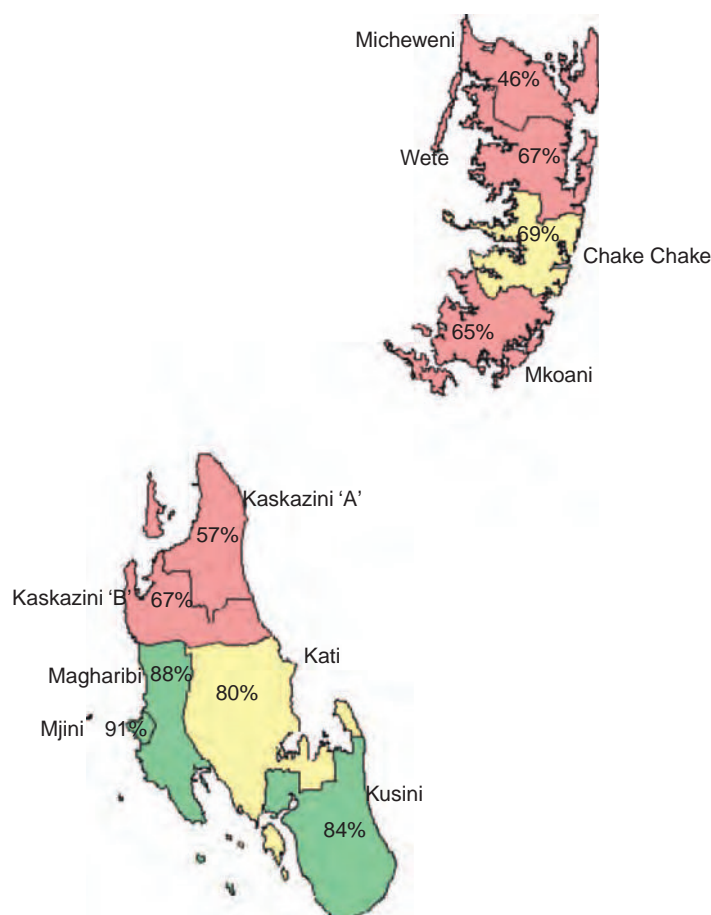
The average pupil teacher ratio in 2007 was 31:1 for public schools while that of private schools was 18:1.

The training of teachers is now undertaken by three universities and three teacher-training colleges and is normally 2 years for primary and 4 years for secondary school teachers. There is also a program for training teachers through distance education.

Ministry of Education and Vocational Training Organizational Structure

Public schools in Zanzibar function under the direct responsibility of the ministry, while the private schools are established and operate under conditions determined by the ministry and the ministry's office of registrar (**Case study 3**). The ministry is currently headed by the minister assisted by the deputy minister. The two have more political rather than professional responsibility for the promotion of education. On matters of policy implementation, the minister is advised by the education council whose members are appointed as per Education Act. Technical and professional tasks are the responsibility of the Principal Secretary who is also the head of civil service as well as chief accountant officer. In the execution of duties, the principal secretary is assisted by deputy principal secretary. In the case of Pemba Island, management responsibilities of the principal secretary are executed by the officer in-charge.

Under the principal secretary, there are two deputy principal secretaries (for academics and for policy, planning, personnel, and administration) and a commissioner supervising 11 directorates.

Case study 2 Percentage of adults literate in any language**Monitoring, Evaluation, and Research**

Education programs have intended outcomes or products that need to be monitored and evaluated. Monitoring and evaluation is central to government's regulatory and quality assurance roles.

The statistics division of the ministry is responsible for monitoring educational programs through data collection, analysis, and dissemination of information. The division administers questionnaires to educational institutions and collects information on student and learner enrolment, attendance, and number of teachers and tutors. Quarterly and annual progressive reports on the status of schools, including physical, material, and academic aspects are also submitted to the division.

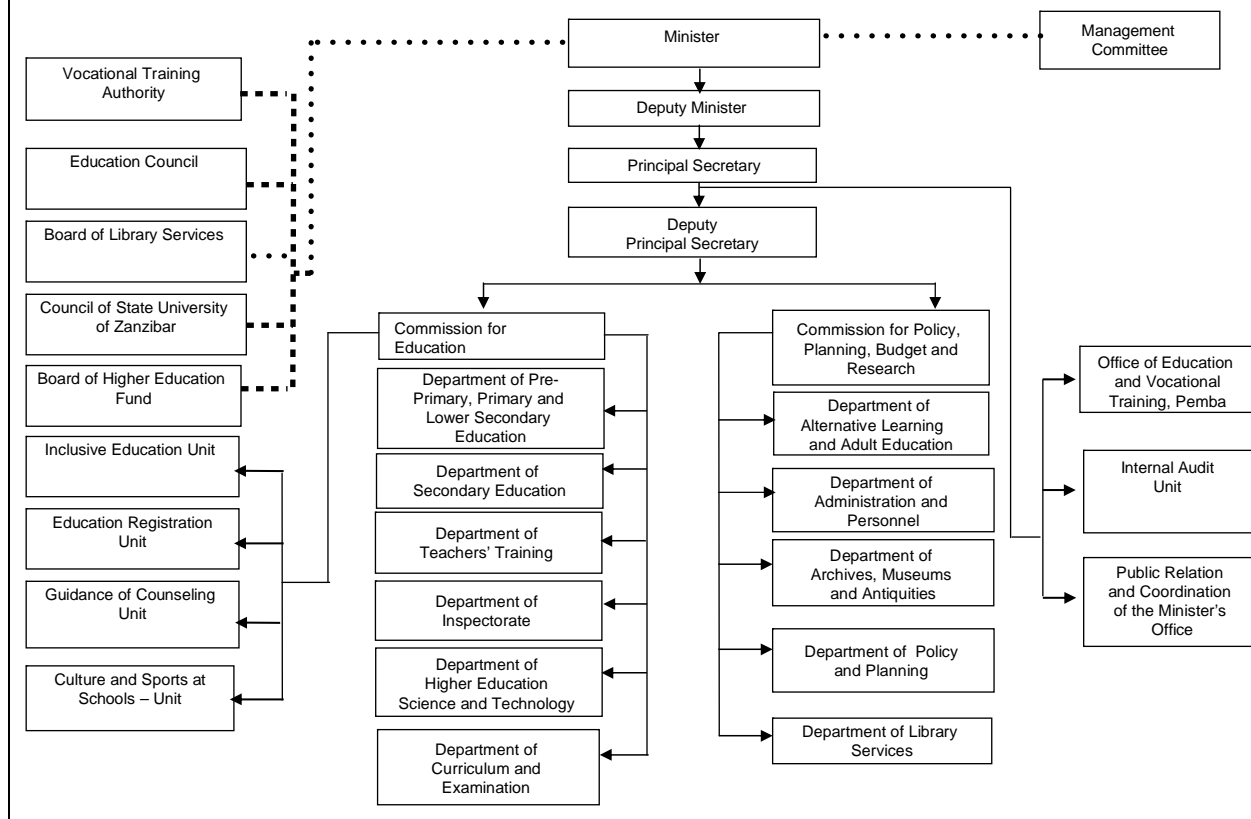
Inspectors, school committees, and school boards also provide data for monitoring purposes. Implementation of development programs is monitored by respective departments in collaboration with the department of planning, policy, and research. Evaluation is very rarely conducted.

Education management information system division also exists in the ministry to develop a sustainable statistical information system for education policymaking. Education management information system division provides information on student enrolment, performance, physical, and human resources. The education management information system division is the main custodian of statistical data and other information systems. Its main activities include development of instruments for collecting, processing, and analysis of data and when required, provision of training to headteachers and school statisticians.

The division, however, has not fully integrated into the planning and management process at the national and local levels due to technical and organizational shortfalls.

Research findings provide information to support the implementation of policy decisions and reforms. Educators and policymakers use researches to improve the education system.

Currently, research activities are being coordinated by the department of planning, policy, and research; however,

Case study 3 Organization Structure of the Ministry of Education and Vocational Training

it has not receive its due recognition, attention, and emphasis in current education efforts. Most of the policy changes in education have been reactive rather than proactive, attempting to solve immediate problems rather than addressing issues based on research findings. Most of the researches are donor funded and focused on curriculum innovation and project implementation.

Education Financing

The Revolutionary Government of Zanzibar is fully committed to the provision and funding of education. Construction of schools, procurement of equipment and educational materials, provision of services, administrative support, and technical backstopping is mainly financed by the government. However, around 90% of the ministry's recurrent budget covers personal emoluments and leave no-salary components, including core items of improving quality of education, grossly underfunded. For a certain time to come, the government will continue to be the major source of financing education.

The government allocated a total of TSh 18.5 billions which is 91.4% of the estimated budget for the financial year 2006/2007 and 20% of the total government budget of the said year up to May 2007.

A higher education fund has also been established to finance higher education, research, and studies and to provide scholarships to the talented needy students.

Community participation has made an important contribution to the provision of essential education materials, building of new classrooms, and assisting in the management of the schools.

Private-sector contribution to education is gradually increasing. Private investors in tourism and hotels as well as local business peoples have contributed significantly to the construction of school buildings and provision of utilities. In community schools, private individuals contribute to the payment of teachers' salaries, tuition, and materials. The private sector also contributes to post-graduate training. In addition, private investors have built and are operating private schools. In general terms, the private sector has acknowledged the importance of education and hence it actively supports and finances its provision.

Table 5 Education recurrent cost projection 2007–16 (billion Tsh)

Level of education	2007/08	2008/09	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16
Preprimary	0.67	0.84	1.11	1.40	1.69	1.47	1.64	1.81	1.95
Primary	12.12	12.21	12.10	12.08	11.83	12.89	12.72	12.53	12.18
Secondary	5.08	5.67	5.58	5.68	5.93	6.50	6.63	6.73	9.67
Tertiary	4.51	4.53	4.53	4.53	4.53	4.53	4.53	4.53	4.53
Vocational and teacher	0.43	0.50	0.51	0.52	0.53	0.54	0.55	0.56	0.58
Nonformal and adult	0.31	0.31	0.32	0.32	0.32	0.32	0.32	0.32	0.33
Administration	4.91	4.96	5.01	5.06	5.11	5.16	5.21	5.26	5.32
Total recurrent costs	28.03	29.03	29.15	29.59	29.95	31.42	31.61	31.76	34.56

Source: Zanzibar Education Development Programme.

Development partners play significant roles in financing development expenditure such as construction and innovation of buildings, production or procurement of school furniture, teaching and learning materials, laboratory equipment, and transport facilities. In the area of technical assistance and capacity building, they have provided scholarships in the field of education to the members of staff. In addition, the development partners finance experts attached to some departments of the ministry, fund education-sector studies, reviews, and innovative projects, and provide links with institutions outside Zanzibar. Development of partners' contributions in education is significant and appears to be increasing, priority being given at the moment to basic and secondary education.

Inherent Challenges

It is evident that the government has strenuously struggled in the face of limited resources to achieve a number of universally cherished goals. It has incorporated in its previous policy documents, key objectives relating to access, equity, and quality. It has expanded provision of basic education to essentially all segments of the society, and can now boast of achieving gender parity in the 10-year compulsory basic education and in the additional 2 years of secondary education.

However, there are a number of goals which remain elusive. For example, only about one in seven children attend some form of preschool before starting primary education, while enrolment in primary schools exposes some worrying weaknesses.

There is also widespread lamentation that the quality of education now offered or attained by young learners has fallen considerably. Graduates of basic education are weak in languages, both the vernacular and English. They lack effective communication skills and their general knowledge in current affairs, science, and arithmetic has serious gaps when compared to children of comparable age in the neighboring countries. Complaints are also heard that

basic-education graduates leave school without usable skills and hence they are unprepared to enter the world of work.

The new education policy (2005) which will be fully implemented by the year 2015, aims at addressing the existing weaknesses of Zanzibar education system by making it more responsive to current needs of the society. It has highlighted the need to set realistic goals and to realign the government's effort with those in the region in accelerating socioeconomic development. Among the essential elements of education reforms are: extending basic education from 10 to 12 years, improving the quality of education, strengthening teachers' training, providing alternative learning for out-of-school youth, and developing general skills.

In order to support the new education policy (Table 5), the Zanzibar education development program (ZEDP) has been developed. ZEDP represents the first comprehensive and sector-wide approach to the educational challenges facing Zanzibar. It identifies, establishes, and plans for substantial targets for the education system, both in terms of equitable access and quality. The ZEDP is strategic in its approach, seeking to develop immediate, medium- and longer-term goals for education which are achievable and sustainable.

The ZEDP is also characterized by a practical focus on where learning takes place and how that environment can be strengthened and enriched. The classroom learning environment, the management structures which support and surround it and the planning and development approaches which guide it remain the critical areas which the ZEDP addresses. By focusing on the fundamentals of learning and teaching, the program draws together all the significant elements necessary for the qualitative changes deemed essential if the learners of Zanzibar are to fulfill their potential over the coming decades.

Further Reading

Ministry of Education and Vocational Training (2006). *Education Policy of 2006*. Zanzibar: Zanzibar Government Press.

Ministry of Education and Vocational Training (2007a). *Education Management Information System Report*. Zanzibar: Zanzibar Government Press.

Ministry of Education and Vocational Training (2007b). *Zanzibar Education Development Program*. Zanzibar: Zanzibar Government Press.

Ministry of Youth, Employment, Women and Children Development (2005). *Vocational Education and Training Policy of 2005*. Zanzibar: Zanzibar Government Press.

Suleiman, H. A. (2007). *Ministry of Education and Vocational Training Revenue and Expenditure Budget Speech for the Year 2007/2008*. Zanzibar: Zanzibar Government Press.

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

This page intentionally left blank

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

EDITORS-IN-CHIEF

PENELOPE PETERSON
School of Education and Social Policy
Northwestern University
Evanston, IL
USA

EVA BAKER
UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

BARRY MCGAW
McGaw Group Pty Ltd
Mont Albert Nth, VIC
Australia



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
The Boulevard Langford Lane, Kidlington, Oxford OX5 1GB, UK

Copyright © 2010 Elsevier Ltd. All rights reserved

LATVIA
Copyright © 2010 A Joma

STUDENT DATA SYSTEMS AND THEIR USE FOR EDUCATIONAL IMPROVEMENT
Copyright © 2010 J C Wayman, V Cho, and M P Richards

QUALIFICATIONS FRAMEWORKS AND THEIR ROLE IN THE REFORM OF EDUCATION AND TRAINING
Copyright © 2010 M Young

LANGUAGE OF VOCATIONAL EDUCATION AND TRAINING
Copyright © 2010 J Misko, J MacKenzie, and R-A Polvere

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44)(0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at (<http://elsevier.com/locate/permissions>), and selecting *Obtaining permissions to use Elsevier material*

Notice

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Catalog Number
A catalogue record for this book is available from the Library of Congress

ISBN: 978-0-08-044893-0

For information on all Elsevier publications
visit our website at books.elsevier.com

PRINTED AND BOUND IN THE UNITED KINGDOM
09 10 11 12 13 10 9 8 7 6 5 4 3 2 1



PREFACE

A preface usually provides a brief introduction to a work, intended to set the stage, provide some background insight, and whet the appetite of the reader. In our case, however, this preface has to address a fundamental question, one that was in our minds at the time we were recruited as Editors-in-Chief for the International Encyclopedia of Education. The question was “Why do we need an Encyclopedia? Its subtext was inspired by the ever-growing, ever-popular Internet. We believe that *this* Encyclopedia is desperately needed and will become a valued resource in education and associated social sciences and arts. The reasons are intellectual and procedural. Anyone with a modicum of knowledge knows that finding and trusting information gleaned from the Internet are two separate actions. The reliance on browsers to help discover references and comments result in resources based on popularity not quality. Pithy titles catch the eye and references rise in the ranks of browser searchers. Related to this is the “editing” in the Internet realm of populist efforts at encyclopedia, references, and other compilations. Once again, after removing offensive material, the accuracy, completeness, lack of bias, and other provenance for entries simply do not exist. Experienced researchers in education can sort through and make intelligent choices. Novices and many journeyman, or practitioners, parents, and policy makers cannot. Contrast how this Encyclopedia was built. Key domains of educational research were identified, and a tentative list of sub-domains or useful applied areas was posited. Then the Editors-in-Chief (apologies for the awkwardness of the term) identified the leading researcher in a particular domain, and with surprisingly little effort, recruited them to participate. They in turn identified the two best researchers in a sub-domain, such as formative assessment or the training of pre-school teachers. The authors of the sections of the Encyclopedia do not represent a collective group of friends and acquaintances, although friendships have been made. Rather they embody a deep and broad scholarly community. The difference from compiled Internet resources is the built-expertise and intellectual engagement of the authors. The summary of the developments and futures in their personal areas of scholarship have been filtered through their years of experience, both as scholars and communicators. Quality, then, is endemic to each piece, developed through this top-down identification of expertise, and made indelible by the bottom-up application of high standards from people leading the sub-domains – the authors, and the domains themselves, the section editors.

On a procedural level, the publishers early committed to the notion that this Encyclopedia would also be an online resource, and access would be available through print, for those with strong bookcases and the persisting love of turning real pages. The Internet version will allow multiple prisms through which the reader may access articles and provide, as it were, an emulation of the Internet in our field, albeit bounded by expertise and high quality.

What must be underscored in the assessment of this effort are the Editors-in-Chief and the publishers’ commitment to find excellence worldwide. We tried very hard to persuade notable scholars from all parts of the world to make contributions. Less than to fulfill the title of “International,” we were on the hunt for perspectives that would enrich the scope and depth of the sections. Our section editors put in enormous time attempting to find the best in the field, wherever they resided. Yet, not everyone is in the volume. Some were overcommitted. Many were not fully confident of their English, and the automated translation software has not yet met standards for technical writing. We believe that such writing and editing tools will make the outreach to an even broader International group of scholars possible in future revisions, or online updates. Furthermore, the birth of the World Educational Research Association (in 2009) will provide a better set of interlocking networks to find and evaluate scholarship from any place on the globe.

Finally, the scope of the effort must be acknowledged: 28 section editors, 926 articles were commissioned, drafted, reviewed, redrafted, edited, and put together in the space of four years. The publishers underwent some internal changes, and alterations in management. We as Editors-in-Chief, changed roles, moved, and also had to keep our own research and development enterprises afloat. Deadlines wobbled; authors dropped from view and had to be replaced.

Yet, at times frustrating as all development is, we find the final product exhilarating. We are enthusiastic not simply because it came into being at all, but because the collective light of the minds that wrote have left a bright resource for the future, one that will impact the way our colleagues understand and experience the educational knowledge, improvement, and impact in the future.

Eva Baker
Penelope Peterson

EDITORS

EDITORS-IN-CHIEF

Penelope Peterson

School of Education and Social Policy
Northwestern University
Evanston, IL
USA

Eva Baker

UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

Barry McGaw

McGaw Group Pty Ltd
Mont Albert Nth
Melbourne, VIC
Australia

SECTION EDITORS

Walter Allen

HEOC Division
University of California, Los Angeles
Los Angeles, CA
USA

Vibeke Grøver Aukrust

Institute for Educational Research
University of Oslo
Oslo
Norway

Dominic J Brewer

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

Kai-Ming Cheng

Faculty of Education
The University of Hong Kong
Hong Kong

Gregory Chung

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Penny Enslin

Department of Educational Studies
University of Glasgow
Glasgow
UK

Noelle Griffin

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Dennis Hocevar

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

David Hopkins

Institute of Education
University of London
London
UK

Mary James

Faculty of Education
University of Cambridge
Cambridge
UK

Sanna Järvelä

Department of Educational Sciences and
Teacher Education
University of Oulu
Oulu
Finland

Thomas Karmel

National Centre for Vocational Education Research
Adelaide, SA
Australia

Harold G Levine

School of Education
University of California, Davis
Davis, CA
USA

Robert L Linn

University of Colorado at Boulder
Ouray, CO
USA

Allan Luke

Faculty of Education
Queensland University of Technology
Brisbane, QLD
Australia

Ann M Mastergeorge

Human Development and Family Studies
University of California, Davis
Davis, CA
USA

Patrick J McEwan

Department of Economics
Wellesley College
Wellesley, MA
USA

Barbara Means

Center for Technology in Learning
SRI International
Menlo Park, CA
USA

Ingrid Moses

University of Canberra
Canberra, ACT
Australia

Bill Mulford

Faculty of Education
University of Tasmania
Hobart, TAS
Australia

Yu Jing Ni

Faculty of Education
The Chinese University of Hong Kong
Hong Kong

Festus Obiakor

Department of Exceptional Education
University of Wisconsin-Milwaukee
Milwaukee, WI
USA

P David Pearson

Graduate School of Education
University of California, Berkeley
Berkeley, CA
USA

Colin Power

Graduate School of Education
University of Queensland
Brisbane, QLD
Australia

Kjell Rubenson

Faculty of Education
The University of British Columbia
Vancouver, BC
Canada

Sandip Sinharay

Education Testing Services
Princeton, NJ
USA

Collette Tayler

Queensland University of Technology
Brisbane, QLD
Australia

Nico Verloop

ICLON, Graduate School of Education
Leiden University
Leiden
Netherlands

HOW TO USE THE ENCYCLOPEDIA

The International Encyclopedia of Education is intended for use by students, research professionals, and interested others. Articles have been chosen to reflect major disciplines in the study of education and common topics of research by academics in this domain. Each article serves as a comprehensive overview of a given area, providing both breadth of coverage for students, and depth of coverage for research professionals. We have designed the encyclopedia with the following features for maximum accessibility for all readers.

The contents of the encyclopedia are arranged alphabetically by section, and within sections, alphabetically by article. The Subject Index is located in Volume 8. Some topics are covered in a multitude of articles from differing perspectives, while other topics may have only one entry. We encourage use of the index for access to a subject area, rather than use of the Contents list alone, so that a reader has a full notion of the coverage of that topic.

The articles include cross-references to other related encyclopedia articles, suggested further readings where applicable, and many contain relevant websites for additional information. We encourage readers to use the cross-references to locate other encyclopedia articles that will provide more detailed information about a subject.

The Further Reading sections include recent secondary sources to aid the reader in locating more detailed or technical information. Review articles and research articles that are considered of primary importance to the understanding of a given subject area are also listed. These suggested further readings are not intended to provide a full reference listing of all material covered in the context of a given article, but are provided as next steps for a reader looking for additional information.

This page intentionally left blank

CONTENTS

Preface	v–vi
Editors	vii–viii
How to use the Encyclopedia	ix

VOLUME 1

A

Adult Education

Adult Education Overview	<i>K Rubenson</i>	1
--------------------------	-------------------	---

Adult Education – Adult Learning, Instruction and Program Planning

Adult Learning	<i>S B Merriam</i>	12
Adult Learning and Instruction: Transformative Learning Perspectives	<i>P Cranton</i>	18
Adult Learning in a Biographic Perspective	<i>L West</i>	25
Adult Learning, Instruction and Programme Planning: Insights from Freire	<i>P Mayo</i>	31
Characteristics of Adult Learning	<i>K Illeris</i>	36
Informal Learning: A Contested Concept	<i>P Hodgkinson</i>	42
Organizational Learning	<i>P-E Ellström</i>	47
Program Planning	<i>A L Wilson and R M Cervero</i>	53
Workplace Learning Frameworks	<i>S Billett</i>	58

Adult Education – as a Field of Study

Class Analysis in Adult Education	<i>T Nesbit</i>	64
Gender Analysis	<i>J Stalker</i>	70
Race and Ethnicity in the Field of Adult Education	<i>J Johnson-Bailey and D Drake-Clark</i>	77
Trends in Workplace Learning Research	<i>T Fenwick</i>	83

Adult Education – Concepts

Lifelong Learning	<i>J Field</i>	89
Rewriting the History of Adult Education: The Search for Narrative Structures	<i>B J Hake</i>	96

Adult Education – Domains and Provision

Adult Basic Education: A Challenge for Vocational Based Learning	<i>R Höghelm</i>	102
Adult Literacy Education	<i>L Tett and R St.Clair</i>	107
Citizenship and Immigrant Education	<i>D Schugurensky</i>	113
Community Based Adult Education	<i>B Connolly</i>	120
Continuing Professional Education: Multiple Stakeholders and Agendas	<i>R M Cervero and B J Daley</i>	127
Financing of Adult and Lifelong Learning	<i>Å Sohlman</i>	133
Labor Education	<i>K Forrester</i>	139
Learning Cities and Regions	<i>C Duke</i>	144
Museums as Sites of Adult Learning	<i>R S Grenier</i>	150
Overview of Lifelong Learning Policies and Systems	<i>D-B Kwon and C H O Daeyeon</i>	156
Popular Adult Education	<i>S Tøsse</i>	162
Provision of Prior Learning Assessment	<i>P Andersson</i>	169
The Age of Learning: Seniors Learning	<i>P Jarvis</i>	175
University Adult Continuing Education: The Extra-Mural Tradition Revisited	<i>M Slowey</i>	183

Adult Education – Economy and Society

Adult Education and Civil Society	<i>J Crowther and I Martin</i>	191
Adult Education and Nation-Building	<i>S Walters</i>	198
Economic Outcomes of Adult Education and Training	<i>A Ferrer and W C Riddell</i>	204
Health and Adult Learning	<i>P Bélanger and M Robitaille</i>	211
Modernization Processes and the Changing Function of Adult Learning	<i>H S Olesen</i>	217
The Political Economy of Adult Education	<i>G Rees</i>	223
Wider Benefits of Adult Education	<i>R Desjardins and T Schuller</i>	229

Adult Education – Participation

Barriers to Participation in Adult Education	<i>K Rubenson</i>	234
Participation in Adult Learning	<i>R Desjardins</i>	240

C**Comparative Education**

Comparative and International Education: English-Language Scholarly Publications	<i>D Post and M Fedri</i>	251
Comparative Education: Philosophical Issues and Concepts	<i>M Mason and C W Evers</i>	257
Comparative Education: Societies and Associations	<i>M Bray</i>	266

Curriculum Development

Curriculum and Religion	<i>C Kapitzke and P Graham</i>	272
Gender and Curriculum	<i>J Marsh</i>	277
Local Teacher-Based Curriculum Development	<i>M D Osborne and H B Barker</i>	283

Curriculum Development – Contexts

Curriculum and Human Rights	<i>C Lohrenscheit</i>	287
Curriculum and the Education of Cultural and Linguistic Minorities	<i>S May</i>	293
Curriculum in Postcolonial Contexts	<i>A Hickling-Hudson</i>	299
Curriculum, Economic and Cultural Development	<i>A Rampal</i>	307
Globalization and Curriculum	<i>R Dale</i>	312
Popular Culture	<i>K Schultz and R Throop</i>	318

Curriculum Development – Evaluation and Research

An Overview of Research in Curriculum Inquiry	<i>F M Connelly and S J Xu</i>	324
Curriculum Evaluation: Approaches and Methodologies	<i>V Klenowski</i>	335
Curriculum Studies, Discourse Analysis, and the Construction of Historical Time	<i>J Laspina</i>	342
Curriculum Use in the Classroom	<i>M Ben-Peretz and B Eilam</i>	348
The Roots and Routes of Teacher-Based Action Research and Curriculum Inquiry: An Historical Perspective	<i>L G Putney and J L Green</i>	355

Curriculum Development – Planning and Development

Curriculum and Syllabus Design	<i>A Woods, A Luke, and K Weir</i>	362
Curriculum and the Publishing Industry	<i>J Larson, A-R Allen, and D Osborn</i>	368
Curriculum and Teacher Change	<i>J G Ladwig</i>	374
Curriculum Governance and Planning	<i>B Levin</i>	379
Curriculum Planning and Systems Change	<i>Z Deng</i>	384
Curriculum, Digital Resources and Delivery	<i>Y Zhao, G Zhang, and C Lai</i>	390
Textbook Development and Selection	<i>P Shannon</i>	397

Curriculum Development – Subject Matter

Curriculum and Globalization: Higher Education	<i>K H Mok</i>	403
Curriculum Development as Subject Matter: Social Studies	<i>J Brophy and J Alleman</i>	411
Curriculum Development in the Area of Reading	<i>C L Madda and T E Raphael</i>	415
Curriculum in the Arts	<i>C M Thompson, L Bresler, and T E Costantino</i>	423
Curriculum: Foreign Language Learning	<i>Angel M Y Lin</i>	428
First Language Learning	<i>T Rogers and L Moffatt</i>	435
Learning a Second Language in First Language Environments	<i>E Bernhardt and D J Tedick</i>	441
Literature	<i>S Hynds</i>	447
Mathematics	<i>B Kaur and C P Vistro-Yu</i>	453
Physical Education and Sports	<i>D Kirk</i>	459
Primary and Elementary/Middle Grades Reading	<i>D E Alvermann and C A Mallozzi</i>	464
Science	<i>Y Kali and M Linn</i>	468

Curriculum Development – Theory

Bilingual Learning (Learning L1 and L2 in an L1 and L2 Environment)	<i>D Schwinge</i>	475
Curriculum and Complex Systems Theory	<i>W-M Roth and J Thom</i>	481

Curriculum and Constructivism	<i>B Davis and D Sumara</i>	488
Curriculum and Critical Theory	<i>L Yates</i>	494
Curriculum and Poststructuralist Theory	<i>J L Miller</i>	499
Curriculum and Structuralist Sociology: The Theory of Codes and Knowledge Structures	<i>J Muller and J Gamble</i>	505
Curriculum Reform	<i>R Gilbert</i>	510
Early Childhood Curriculum and Developmental Theory	<i>S M-Y Lim and C Genishi</i>	514

D

Demography and Social Change

An Evaluation Framework for E-Learning Effectiveness in the Arab World	<i>W AbuSneineh and M Zairi</i>	521
Globalization and Social Justice in Higher Education	<i>E Kimura-Walsh</i>	536
Racial Inequality and the Economic Payoffs of Higher Education	<i>S A Lachtman</i>	542
School Health Education	<i>J S Yang</i>	547

Demography and Social Change – Demography

A Paradigm of Contradictions: Racism and Science Education	<i>B A Brown and J W Mutegi</i>	554
Africana Studies: Past, Present And Future	<i>B R Hare, A Young, J B Stewart, and D P Aldridge</i>	565
Beyond Black Nationalism to Black Internationalism: The Compelling Case for Examining Black Educational Challenges Globally	<i>K Freeman</i>	573
Beyond Racial, Ethnic, and Gender Bias in Education Statistics	<i>J T Lott and C Bennett</i>	581
Caste and Education in India	<i>G G Wankhede</i>	589
Demography, Community, and the Education of Asian American and Pacific Islander Students	<i>R T Teranishi</i>	595
Educational Attainment and Mortality Differentials	<i>H D Horton, L Hixson, and S M Monnat</i>	602
Educational Opportunity and Latino/Chicano College Choice: New Findings, and Theoretical Perspectives	<i>M A Ceja and P A Pérez</i>	610
Ethnic Minority Identity and Educational Outcomes in a Rising China	<i>G A Postiglione</i>	616
Ethnic Studies and Women's Studies: From the Past to the Present, Providing a Space in the Academy for Scholarship Focused on the Histories and Experiences of Marginalized People	<i>C A Brown</i>	623
Gender Equity in Higher Education: Challenges and Celebrations	<i>L Morley</i>	629
Locating Space and Place in the College Access Debate: New Tools for Mapping and Understanding Educational Inequity and Stratification	<i>B Cantwell and J F Milem</i>	636
Public Policy and Inequality in Postsecondary Opportunity: Educational Statistics and the Failure of Education Reform	<i>E P St. John, K Williams, and K Moronski</i>	649
The Education of Indigenous Students	<i>J W Tippeconnic and S C Faircloth</i>	661
The Health Advantages of Educational Attainment	<i>G Y Dinwiddie</i>	667
To Go or Not to Go: Access and Barriers to Ethnic Minority Participation in Higher Education within the United Kingdom	<i>P Brug and T L Oliver</i>	673
Universal Secondary Education and the Two-Tiered School System in the English-Speaking Caribbean	<i>P Anderson and J Devonish</i>	679
Women of Color in Science, Technology, Engineering and Mathematics (STEM): Refining the Concepts, Reframing the Issues	<i>C B Leggon</i>	686

Demography and Social Change – Social Change

Educational Diversity in U.S. Law Schools	<i>C E Daye, A T Panter, W R Allen, M E Deo, and L F Wightman</i>	691
Educational Reform	<i>F Lopez-Segrera</i>	698
Emergence of For-profit Higher Education	<i>X Li</i>	705
Engaged Education: Experiential Learning, Intensive Field Experiences, and Social Change	<i>T R Rone</i>	711
Higher Education and Social Change in Mexico	<i>A Mungaray, G Estrella, P Moctezuma, and M Ramírez-Urquidy</i>	719
The Changing Role of the State in Higher Education	<i>K Gürüz</i>	728
The Role of the State in School Reform: Responding to Increasing Ethnic Diversity in Britain	<i>P Ratcliffe</i>	737
Transforming Mindsets Through Education for Sustainable Development	<i>V M Segovia</i>	746

Demography and Social Change – Sociology

Achievement Motivation in Ethnic Minority Youth	<i>S Graham and E Bryant</i>	753
Affirmative Action and Higher Education in Brazil	<i>J T dos Santos and D M Queiroz</i>	760
Black in White: Black Students at White and Black Colleges	<i>L Evans and J R Feagin</i>	767
Bologna Process: On the way to a Common European Higher Education Area	<i>A V Rezaev</i>	772
Critical Race Theory in Education Research	<i>R Vue and C B Newman</i>	779
Cultural Diversification and Japanese Education: Social Constructions of the New Diversity	<i>R Tsuneyoshi</i>	787
Education in War Countries: A Disempowerment of Society	<i>J L Nkata</i>	793
Educational Attainments of U.S. Black Males and Females: 1971 to 2003	<i>F B Bonner, M-C Jipguep-Akhtar, and R J Harrison</i>	799
Educational Preparation: Fostering the Self-Efficacy and Resilience of Urban Adolescent Youth	<i>M B Spencer and B Tinsley</i>	806
HIV Education for Low-Literate People: Transforming Students and Communities through Paulo Freire's Praxis and the Pedagogy of Action	<i>N Z Haniff</i>	814
How Are Discussions of Interdisciplinary Studies Linked to Diversity Discourse?	<i>C N Spence</i>	823
Identifying Patterns of Doctoral Attrition Across Academic Fields	<i>A James, K J Fosnacht, E B Stolzenberg, and D Thirumalai</i>	827
Male Underachievement in Education Across the Globe: A Shift in Paradigm for Gender Disparities Regarding Academic Achievement	<i>J F L Jackson, J L Moore III, and R A Leon</i>	838
Noncognitive Measures for Higher Education Admissions	<i>W E Sedlacek</i>	845
Peace Education	<i>J Page</i>	850
Remote Control: Africa's Development and Outflow to Developed Nations of Technical Talent	<i>F Haniff</i>	855
Scholar-Baller®: Student Athlete Socialization, Motivation and Academic Performance in American Society	<i>C K Harrison, S Bukstein, J Mottley, E Comeaux, J Boyd, C Parks, and D Heikkinen</i>	860
Service-Learning and Diversity as Mediums to Foster Civic and Community Engagement in Undergraduates	<i>L Oseguera</i>	866
Success and Failure in Education and Criminal Justice: Identifying Common Mechanisms	<i>S Ramirez and A Harris</i>	874
The History of Education: Race and Education	<i>J O Jewell</i>	881

VOLUME 2**E****Early Childhood Care and Education Theory**

- Early Childhood Care and Education: The Family, The Market, and The State *S B Kamerman and M J Neuman* 1
- Gender Issues in Early Childhood Education and Care *B Pocock* 11

Early Childhood Education and Care Governance

- Early Childhood Risk, Protection and Abuse Prevention *E Khoo* 17
- Evaluating Early Childhood Education and Care Programs *M Janus and S Brinkman* 25

Early Childhood Education and Care Investment

- Investing in Early Childhood Education and Care: The Health and Wellbeing Case *T Moore and F Oberklaid* 32
- Cost-effective Early Childhood Programs from Preschool to Third Grade *A J Reynolds, J A Temple, and B White* 38
- Early Childhood in Post-Modern Cultures: Thoughts and Some Concerns *P Gammage* 49
- Investing in Early Childhood Education and Care: Some Policy Implications *J Bennett* 55
- Investing in Early Childhood Education and Care: The Economic Case *G Cleveland and M Krashinsky* 63

Early Childhood Education and Care Programs

- Early Care and Education Programs for Infants and Toddlers *B Nyland* 69
- Literacies in Early Childhood: The Preschool Period *M Fleeer and B Raban* 75
- Participation in Early Childhood Education and Care Programs: Equity, Diversity and Educational Disadvantage *M Vandenbroeck* 81
- Teaching in Early Childhood Centers Instructional Methods and Child Outcomes *I Siraj-Blatchford* 86

Early Childhood Education and Care Provision

- Early Childhood Education and Care in the People's Republic of China *Zeng Xiaodong* 93
- Home-Based and Institutional Early-Childhood Education and Care Services *R Fukkink and H Blok* 98
- Integrated Early Childhood Education and Care Services – Care, Upbringing, Education and Health *S Colley* 104
- Issues of Access and Program Quality *J T Hustedt and W S Barnett* 110
- Kindergarten Transitions and Linkages to Primary School – Readiness Re-conceptualized *A Petriwskyj* 120
- Parent Support in Early Childhood – Approaches and Outcomes *K S Howard, K A Beckmann, and J Brooks-Gunn* 126

Early Childhood Education and Care Research

- Development and Implementation of Early Learning Standards in the United States *C Scott-Little* 132
- Global Status of Early Learning and Development Standards *P R Britto and S L Kagan* 138
- Rethinking Early Childhood Education and Care: Implications for Research and Evaluation *K Thorpe, D Cloney, and C Tayler* 144

Early Childhood Education and Care Theory

- Child Rearing and Early Education: Parents and Professionals: Theoretical and Historical Influences *H Penn* 151

Early Childhood Education and Care Workforce

- Early Childhood Education and Care Professional Profiles in the United Kingdom: Multiagency Teamwork and Transdisciplinary Challenges *A Anning and P Calder* 157
- Early Childhood Education and Care Staff Support: Mentoring and Professional Development in Finland *P Lindberg* 162

Economics of Education

- Education and Health *P Muennig* 169
- Data *M Perez and M Socias* 178
- Empirical Research Methods in the Economics of Education *P J McEwan* 187
- Theoretical Concepts in the Economics of Education *D J Brewer, G C Hentschke, and E R Eide* 193

Economics of Education – Costs and Financing of Schooling

- Cost–Benefit Analysis and Cost–Effectiveness Analysis *C Belfield and H M Levin* 199
- Economic Approaches to Adequacy *J Imazeki* 204
- Economic Approaches to School Efficiency *L L Taylor* 210
- School Finance Reform *T Downes* 216
- School Finance: An Overview *J K Rice, D H Monk, and J Zhang* 222
- The Economics of Tuition and Fees in Higher Education *R G Ehrenberg* 229

Economics of Education – External Benefits of Schooling

- Education and Civic Engagement *T S Dee* 235
- Education and Crime *L Lochner* 239
- Education and Economic Growth *E A Hanushek and L Wößmann* 245
- Parental Socioeconomic Status, Child Health, and Human Capital *J Currie and J Goodman* 253
- The External Benefits of Education *W W McMahon* 260

Economics of Education – Labor Markets, Education and Earnings

- Education and Inequality *J Blanden and S Machin* 272
- Human Capital *E R Eide and M H Showalter* 282
- Race Earnings Differentials *M Carnoy* 288
- Returns to Education in Developed Countries *M Gunderson and P Oreopoulos* 298
- Returns to Education in Developing Countries *H A Patrinos and G Psacharopoulos* 305
- School Quality and Earnings *J R Betts* 313
- Signaling in the Labor Market *M E Page* 321
- The Economics of High School Dropouts *R W Rumberger* 325

Economics of Education – Markets, Incentives and School Choice

- Competition and Student Performance *A A Payne* 331
- Educational Privatization *C Belfield and H M Levin* 337

Student Incentives	<i>E R Eide and M H Showalter</i>	342
The Economic Role of the State in Education	<i>D N Plank and T E Davis</i>	347
The Economics of Catholic Schools	<i>W Sander and D Cohen-Zada</i>	354
The Economics of Charter Schools	<i>R Zimmer and R Buddin</i>	360
The Economics of Parental Choice	<i>R Chakrabarti and J Roy</i>	367
The Economics of School Accountability	<i>D N Figlio and H F Ladd</i>	374
The Efficacy of Educational Vouchers	<i>R Zimmer and E P Bettinger</i>	380
Tiebout Sorting and Competition	<i>T J Nechyba</i>	388

Economics of Education – Production of Schooling

Desegregation, Academic Achievement and Earnings	<i>S Rivkin</i>	394
Education Production Functions: Concepts	<i>D N Harris</i>	402
Education Production Functions: Developed Country Evidence	<i>E A Hanushek</i>	407
Education Production Functions: Evidence from Developing Countries	<i>P Glewwe and S Lambert</i>	412
Family Environment in the Production of Schooling	<i>R Rothstein</i>	423
Neighborhoods and Peers in the Production of Schooling	<i>J Vigdor and J Ludwig</i>	431
Teacher Quality in Education Production	<i>I Pelayo and D J Brewer</i>	438
The Economics of Class Size	<i>D W Schanzenbach</i>	443
The Economics of Early Childhood Interventions	<i>M Nores</i>	450

Economics of Education – Teaching Labor Markets

An Overview of Teacher Labor Markets	<i>H Lankford and J Wyckoff</i>	456
Compensating Differentials in Teacher Labor Markets	<i>J Chambers</i>	465
Economic Approaches to Teacher Recruitment and Retention	<i>S Loeb and J Myung</i>	473
Teacher Incentives	<i>L Santibañez</i>	481
Teacher Supply	<i>P Dolton</i>	489
Teacher Training and Preparation in the United States	<i>M Podgursky, M G Springer, and R Hutton</i>	498
Teachers in Developing Countries	<i>F H Rogers and E Vegas</i>	504
The Economics of Teachers' Unions in the United States	<i>K Strunk</i>	512

Education of Children with Special Needs

Adapted Physical Activity for Students with Special Needs	<i>S R Hodge</i>	519
Assessment Accommodations for Children with Special Needs	<i>R J Kettler and S N Elliott</i>	530
Assistive Technology and Educational Practice	<i>H P Parette and G R Peterson-Karlan</i>	537
Augmentative–Alternative Communication	<i>C Mason and K M Chinn</i>	544
Community Integration and Employment of Youth with Special Needs	<i>B A Ford and A W Daviso III</i>	550
Cooperative Learning for Children with Special Needs	<i>E A Dooley and K R Kossar</i>	555
Co-Teaching	<i>E D Cramer</i>	560
Cultural and Linguistic Diversity and Children with Special Needs	<i>J M Patton, N Day-Vines, and A McSwain</i>	565
Curriculum-Based Assessment and Students with Special Needs	<i>D J O'Shea and J D Katsafanas</i>	571

Diagnosis and Classification	<i>M Sapp</i>	577
Direct Instruction and the Education of Children with Special Needs	<i>T Mehring</i>	584
Disabilities and Gifted Education	<i>V I Daniels and M J McCollin</i>	588
Early Childhood Special Education: Birth to Eight Years Old	<i>T S Lovelace, R Gardner III, J B Axe, and D M Sainato</i>	594
Early Intervention for Students with Special Needs	<i>P A Grant</i>	608
Early Transition of Children with Special Needs	<i>S A Fowler</i>	613
Educating Children Who are Deaf-Blind	<i>J Wood and K M Chinn</i>	621
Educating Children with Deafness and Hearing Impairments	<i>G M Mukuria and C J Eleweke</i>	628
Educating Children with Other Health Impairments (ADHD)	<i>G M Mukuria and F E Obiakor</i>	634
Educating Children with Physical Disabilities	<i>Z P Nkabinde, F E Obiakor, M T Ofor, and D J Smith</i>	640
Educating Children with Specific Learning Disabilities	<i>P A Grant and P B Grant</i>	646
Educating Children with Speech and Language Disorders	<i>S L Green and C D Qualls</i>	654
Educating Students with Autism and Related Disorders	<i>T T Dyches</i>	661
Educating Students with Cognitive Disabilities	<i>J P Bakken</i>	669
Educating Students with Emotional and Behavioral Disorders	<i>S W Smith and G G Taylor</i>	678
Educating Students with Special Needs: An Overview	<i>F E Obiakor, M K Harris, P Harris-Obiakor, and G M Mukuria</i>	688
Educating Students with Traumatic Brain Injury	<i>S Burkhardt and A Rotatori</i>	695
English Language Learners with Special Needs	<i>R Rueda and G Ragusa</i>	701
Functional Behavioral Assessment	<i>J P Bakken</i>	708
History of Special Education	<i>L M Garrick Duhaney and S J Salend</i>	714
Inclusion of Students with Special Needs in General Education Classrooms	<i>M A Prater</i>	721
Instructional Accommodations for Children with Special Needs in Inclusive Settings	<i>A L Yurick, S Musti-Rao, and G Cartledge</i>	727
Interagency Agreements	<i>L A Owens</i>	733
Interdisciplinary Collaboration in Early Childhood	<i>N File</i>	739
Language Development in Children with Special Needs	<i>M S L Stewart</i>	745
Large Scale Assessment and Accountability and Students with Special Needs	<i>M L Thurlow</i>	752
Literacy Instruction for Students with Special Needs	<i>L H Mason</i>	759
Math Instruction for Children with Special Needs	<i>B A Bottge</i>	767
Overrepresentation of Culturally and Linguistically Diverse Students in Special Education	<i>S C Trent</i>	774
Paraprofessionals in Special Education	<i>B Y Ashbaker, S C Dunn, and J Morgan</i>	780
Parent and Family Involvement in the Education of Children with Special Needs	<i>L P Prater</i>	788
Peer Relations and Socialization of Children and Adolescents with Special Needs	<i>L M Bullock, M Wong-Lo, and R A Gable</i>	794
Peer-Tutoring	<i>L J O'Shea and D J O'Shea</i>	802
Postsecondary Participation of Students with Special Needs	<i>R A Stodden and K D Roberts</i>	808
Prereferral Strategies	<i>C G Simpson and J P Bakken</i>	815

Preventing Antisocial Behavior and Delinquency: A Comprehensive School-wide Approach <i>P R Hallam, K R Young, P Caldarella, D G Wall, and L Christensen</i>	820
Programs and Instructional Strategies for Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	829
Related Services for Children with Special Needs <i>S O Obi</i>	834
Risk and Resilience Frameworks in Understanding Special Education <i>W-Y Hsieh and M L Donahue</i>	840
School Dropout Prevention <i>F E Obiakor</i>	845
School-Based Services for Children with Special Needs <i>S O Obi</i>	850
Self-Determination <i>M L Wehmeyer</i>	857
Social Competence <i>L A Owens and S Johnston-Rodriguez</i>	865
The Identification of Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	870
Transition from School to Adult Life <i>M J McCollin and F E Obiakor</i>	876
Visual Impairment <i>W Sapp</i>	880
Workforce Issues in Special Education <i>E Carlson and B Billingsley</i>	886

VOLUME 3

Education of Professionals

An Overview of Accreditation, Certification, and Licensure Processes <i>J J Norcini and M van Zanten</i>	1
Educating America's Military Officers <i>M H Adamshick</i>	7
Education for Architecture in the United States and Canada <i>S Schreiber</i>	13
Education for Medicine <i>M B Anderson</i>	19
Education for the Legal Profession <i>J W Wegner</i>	25
Education of Clergy <i>D O Aleshire</i>	31
Framing Professions Education <i>P S O'Sullivan</i>	37
Issues in Accounting/Business Education <i>W H Gijsselaers and R G Milter</i>	39
Professions – Engineering <i>J Heywood</i>	45
Professions – Health-Related <i>G M Jensen and C B Royeen</i>	51
Professions – Pharmacy <i>D E Beck</i>	58
Teaching for the Complex Practice of Nursing <i>P Benner and M Sutphen</i>	65
Veterinary Education <i>K M Edmondson</i>	71

Education Research Methodology: Quantitative Methods and Research

Abductive Research Methods <i>B D Haig</i>	77
Field Experimentation in Education <i>M A Constat</i>	83
Internal Validity <i>M H Clark and S C Middleton</i>	90
Methods for Approximating Random Assignment <i>B Schneider and S-K McDonald</i>	97
Methods for Approximating Random Assignment: Regression Discontinuity and Propensity Scores <i>L M Scheier</i>	104

Multitrait–Multimethod Designs	<i>M Bong and S Hong</i>	111
Multivariate Longitudinal Data Analysis	<i>H Liu and L Chang</i>	120
Quasi-Experimentation: Two Group Design	<i>E Hong</i>	128
Regression Discontinuity Designs	<i>A Kelava, S Rohrmann, and V Hodapp</i>	134
Sampling	<i>R S Brown</i>	142
Statistical Conclusion Validity	<i>P L Busk</i>	147
Survey Research Methods	<i>A Fink</i>	152

Educational Assessment

An Overview of Educational Assessment	<i>M James</i>	161
Assessment and the Evaluation of Institutional Effectiveness	<i>S M Thomas</i>	172
Assessment in Schools – Dispositions	<i>R D Crick</i>	181
Assessment in Schools Related To Literacy: Reading	<i>K J Topping</i>	189
Impact of Assessment on Students' Learning Strategies and Implications for Judging Assessment Quality	<i>M Segers, F Dochy, and D Gijbels</i>	196

Educational Assessment – Assessment Activities

Alternative Assessment	<i>G Stobart and C Gipps</i>	202
Classroom Assessment Tasks and Tests	<i>J Phelan and J Phelan</i>	209
Dynamic Assessment	<i>J G Elliott, E L Grigorenko, and W C M Resing</i>	220
Instructional System Provided Feedback	<i>H F O'Neil, S Chuang, and T Huang</i>	226
Peer and Self-assessment	<i>C Harrison</i>	231
Portfolio Assessment	<i>V Klenowski</i>	236
Records of Achievement: Beyond Traditional Tests	<i>P Broadfoot</i>	243

Educational Assessment – Assessment in Domains

Assessment in Higher Education	<i>D R Sadler</i>	249
Assessment in Schools – Affective Domain	<i>M Ventura, R D Roberts, and M Zeidner</i>	256
Assessment in Schools – Creative Subjects	<i>T E Costantino and L Bresler</i>	262
Assessment in Schools – Learning to Learn	<i>A Hautamäki, J Hautamäki, and S Kupiainen</i>	268
Assessment in Schools – Literacy Writing (Extended)	<i>J M Parr and L Earl</i>	273
Assessment in Schools – Mathematics	<i>M Brown and J Hodgen</i>	279
Assessment in Schools – Oracy	<i>A L Bailey</i>	285
Assessment in Schools – Primary Science	<i>W Harlen</i>	293
Assessment in Schools – Secondary Science	<i>J Shemwell, A Fu, M Figueroa, R Davis, and R Shavelson</i>	300
Assessment in Schools – Technology Education and ICT	<i>A Jones, B Cowie, and J Moreland</i>	311
Assessment in the Early Years	<i>M J Drummond</i>	316
Assessment in the History, Civics and Social Studies Domains	<i>P G Avery</i>	323
Assessment in the Workplace of Performance, Developing Expertise and Competence	<i>A J Brown</i>	330
Assessment in Vocational Education	<i>K Ecclestone</i>	337

Educational Assessment – Concepts and Issues

Assessing Group Work	<i>M Galton</i>	342
Assessment and the Regulation of Learning	<i>L Allal</i>	348
Challenges of Developing and Implementing Formative Assessment Practices in Schools	<i>R Sutton</i>	353
Formative Assessment	<i>P Black</i>	359
Formative Assessment and Instructional Planning	<i>B Marshall</i>	365
Progression and Assessment: Developmental Assessment	<i>M Forster and G Masters</i>	369
Students' Rights in Assessment and Evaluation	<i>J Gardner and B Holmes</i>	378
Summative Assessment by Teachers	<i>R Daugherty</i>	384
The Multiple Purposes of Assessment	<i>P E Newton</i>	392
The Relationship between Assessment and the Organization and Practice of Teaching	<i>L B Resnick, D William, R Apodaca, and E S Rangel</i>	397

Educational Assessment – Enhancing Quality and Use

Assessment Conversations: Reading and Writing Conferences with Students and with Parents	<i>M Barrs</i>	403
Assessment: Pre-service and In-service Teacher Education	<i>A Davies</i>	410
Classroom Assessment in Policy Context (Australia)	<i>J J Cumming</i>	417
Classroom Assessment in Policy Context (England and France)	<i>M Osborn and M Raveaud</i>	425
Classroom Assessment in Policy Context (French Sources)	<i>D Laveault</i>	432
Classroom Assessment in Policy Context (Hong Kong)	<i>D Carless</i>	438
Classroom Assessment in Policy Context (New Zealand)	<i>T J Crooks</i>	443
Classroom Assessment in Policy Context (Sub-Saharan Africa)	<i>S J Howie, L Zimmerman, and K Draper</i>	449
Moderation of Student Work by Teachers	<i>G S Maxwell</i>	457
School Policies and Practices to Support Effective Classroom Assessment for Learning	<i>D Pedder</i>	464
Student Assessment: Policy and Practice in Eastern Europe	<i>G Bethell</i>	472

Educational Assessment – Key Relationships

Assessment Practice in Policy Context: Latin American Countries	<i>F M Rizo</i>	479
Impact of Assessment on Learner Groups (Boys/Girls)	<i>J Elwood</i>	486
Impact of Assessment on Learner Groups (Disabilities)	<i>M Rouse</i>	492
Impact of Assessment on Students' Test Anxiety	<i>P Buchwald and C Schwarzer</i>	498
Impact of Assessments on Classroom Practice	<i>J L Herman</i>	506
Social Practices in School Assessment and their Impact on Learner Identities	<i>A Pollard and A Filer</i>	512

Educational Evaluation – Concepts, Practice, and Future Directions

Educational Evaluation: Concepts, Practice, and Future Directions	<i>Y J Ni</i>	518
---	---------------	-----

Educational Evaluation – Contemporary Themes in Evaluation

Evaluating E-Learning	<i>D D Williams and C R Graham</i>	530
Evaluating Schools as Learning Communities	<i>S M Thomas</i>	539
Evaluation of Adult Education and Training Programs	<i>G McNamara, P Joyce, and J O'Hara</i>	548
Evaluation of Integrated Health Programs in School	<i>T M Franke</i>	555

Formative Assessment in Teacher Education and Teacher Professional Development	<i>H Tillema</i>	563
Informal Education and Evaluation	<i>T Jeffs</i>	572
Methods to Evaluate Technology	<i>W L Bewley</i>	578

Educational Evaluation – Contexts of Educational Evaluation

Cost Analysis in Evaluation Studies	<i>J D Fletcher</i>	585
Cultural Issues that Can Affect the Validity of Educational Evaluations	<i>R P Durán</i>	592
Evaluation and Accountability	<i>E Sato and S N Rabinowitz</i>	600
Evaluation Governance and Planning	<i>R Picciotto</i>	607
Evaluation Use	<i>V G Thomas</i>	614
Moral and Ethical Issues in Evaluation	<i>M Morris</i>	622
The Role of Stakeholders in Educational Evaluation	<i>S Taut and M C Alkin</i>	629

Educational Evaluation – Evaluation Domains

Curriculum Evaluation	<i>T Levin</i>	636
Evaluating Education in Three Policy Eras	<i>M A Peters, R Britez, and S M Weber</i>	645
Evaluation of Mathematics Education Programs	<i>J Cai</i>	653
Evaluation of Science Education Programs	<i>G Berberoğlu</i>	660
Evaluation of Social Studies and Civic Education Programs	<i>S G Grant</i>	667
Evaluation of Teacher Quality and Practice	<i>W Robinson and J Campbell</i>	674
International Evaluations	<i>A Schleicher</i>	681
National Assessment Programs: The Example of the U.S. National Assessment of Educational Progress	<i>M Schneider</i>	690
Principal Evaluation: Concepts, Needs and Realities	<i>N S Glasman and L D Glasman</i>	698
Program Evaluation	<i>C A Christie and L A Fierro</i>	706
School Based Evaluation: Purposes, Protocols and Processes	<i>J MacBeath</i>	713
School Inspection/External School Evaluation	<i>C Chapman and P Earley</i>	719
The International System for Teacher Observation and Feedback: A Theoretical Framework for Developing International Instruments	<i>L Kyriakides, B P M Creemers, C Teddlie, and D Muijs</i>	726

Educational Evaluation – Evaluation Methods

Evaluation Methodology	<i>M Chatterji</i>	735
Evaluation Reporting and Communicating	<i>R T Torres</i>	746
Longitudinal Evaluation Designs	<i>X Ma</i>	757
Meta-evaluation: Purpose, Prescription, and Practice	<i>L A Wingate</i>	765

Educational Evaluation – Purposes and Models of Educational Evaluation

Defining Quality in Evaluation	<i>P Dahler-Larsen</i>	775
Internal and External Evaluation	<i>D Nevo</i>	781
Needs Assessment in Education	<i>J W Altschuld and T L Lepicki</i>	786
The Purpose of Educational Evaluation	<i>S Mathison</i>	792
Understanding Approaches to Evaluation	<i>A J Love</i>	798

VOLUME 4**Educational Measurement**

Ability Testing	<i>P B Tymms</i>	1
Admissions Testing	<i>R Zwick</i>	7
Appropriateness Measurement: Person Fit	<i>R R Meijer and A M Weekers</i>	15
Automated Essay Scoring: Writing Assessment and Instruction	<i>M D Shermis, J Burstein, D Higgins, and K Zechner</i>	20
Classical Test Theory Reliability	<i>M D Miller</i>	27
Criterion-Referenced Measurement	<i>S A Shrock and W C Coscarelli</i>	31
Differential Item Functioning	<i>M D Hidalgo and J Gómez-Benito</i>	36
Educational Measurement: Overview	<i>R L Linn</i>	45
Equating and Scaling	<i>A A von Davier</i>	50
Fairness	<i>C Gipps and G Stobart</i>	56
Generalizability Theory	<i>R L Brennan</i>	61
Impact of Assessment on Classroom Practice	<i>J L Herman</i>	69
Intelligent Systems	<i>V J Shute and D Zapata-Rivera</i>	75
Item Response Theory	<i>W J van der Linden</i>	81
National Assessments	<i>T J Crooks</i>	89
Norm-Referenced Measurement	<i>H D Hoover</i>	95
Progress Monitoring	<i>L S Fuchs and D Fuchs</i>	102
Rasch Models	<i>D Andrich</i>	111
Scoring Rubrics	<i>J A Arter</i>	123
Second Language Assessment	<i>L F Bachman</i>	140
Self Adaptive Testing	<i>J Revuelta</i>	148
Student Test Results in School Accountability	<i>L S Hamilton</i>	153
Test Development	<i>S M Downing</i>	159
Test Translation and Adaptation	<i>A Allalouf and P Hanani</i>	166
Testing Creativity	<i>M A Runco</i>	170
Testing in History	<i>S Wineburg and B Fogo</i>	175
Validity	<i>R L Linn</i>	181
Validity of Achievement Gains on High-Stakes Tests	<i>D Koretz</i>	186
Validity of Educational Indicators	<i>B P M Creemers and L Kyriakides</i>	193

G**Globalization of Universities and Institutions**

Globalization, Universities and Medium of Instruction	<i>J Lo Bianco</i>	201
The Washington Accord on Engineering Curriculum	<i>E Ko</i>	209

H

Higher Education

Higher Education: An Overview	<i>P Scott</i>	217
-------------------------------	----------------	-----

Higher Education – Management, Leadership and Governance

Academic Managers – Heads of Departments/Schools	<i>S Marshall</i>	229
Leading and Managing the University – Presidents and their Senior Management Team	<i>R Middlehurst, T Kennie, and S Woodfield</i>	238
Management of and in Higher Education Institutions	<i>B Sporn</i>	245
Managerialism and Collegialism in Higher Education Institutions	<i>M Shattock</i>	251
Strategic Planning in Higher Education	<i>L Zechlin</i>	256
The Changing Role of University Presidents, Vice-Chancellors and Rectors	<i>M F Green and P D Eckel</i>	264
The Changing Roles of University Governing Boards and Councils	<i>J Fielden</i>	273

Higher Education – Research and Scholarship

Funding of University Research	<i>G Harman</i>	279
Mobility of PhD Students and Scientists	<i>S Avveduto</i>	286
National Science Policy and Universities	<i>M Henkel and U Kogan</i>	294
Research Commercialization	<i>S Slaughter</i>	300
Research Quality Assessment	<i>I McNay</i>	307
The Management of University Research	<i>J Kirkland</i>	316
The Training of New Scientists	<i>R Neumann</i>	322

Higher Education – Society

Academic Freedom	<i>S F Macintyre</i>	328
Access and Equity in Higher Education	<i>G Baldwin and R James</i>	334
Boundaries of Institutional Autonomy and their Impact on Higher Education	<i>V L Meek</i>	341
Diversity of Higher Education	<i>U Teichler</i>	347
Employability of University Graduates and Graduate Outcomes	<i>J Brennan</i>	354
Higher Education and the Knowledge Society	<i>J Välimaa</i>	360
Higher Education and the Labor Market	<i>J Enders</i>	365
Higher Education and the Transformation of Society	<i>P Scott</i>	370
Quality Assurance in Higher Education – Practices and Issues	<i>D Dill</i>	377
The Role of Accreditation of Higher Education Institutions	<i>J S Eaton</i>	384
Universities and Regional Development	<i>J Goddard and J Puukka</i>	390
Universities' Engagement with Society	<i>D Watson</i>	398

Higher Education – Teaching and Learning in Higher Education

Distance Education and Open Universities	<i>A Kanwar and J Daniel</i>	404
Education Development in Higher Education	<i>Y Ryan and K Fraser</i>	411
Effective Use of Technology in Teaching and Learning in HE	<i>D Laurillard</i>	419

Facilitating Students' Success – Student Support and Learning Resources in Universities	<i>E Woodberry and C Barnes</i>	427
Faculty Values and Expectations and Their Implications for Teaching, Learning and the Organization of Higher Education Institutions	<i>K Harman</i>	433
Flexible Learning in Higher Education	<i>S Alexander</i>	441
Innovation in Teaching and Curriculum Design	<i>J M Dearn</i>	448
Productivity of University Faculty Staff	<i>S Kyvik</i>	455
Promoting Effective Student Learning in Higher Education	<i>K Trigwell</i>	461
Student Attrition from Higher Education Institutions	<i>M Hayden</i>	467
University Students' Experiences of Higher Education	<i>C McInnis</i>	473
What Do University Students Learn?	<i>G D Kuh and B K Bridges</i>	479

Higher Education – Tertiary Education in a Globalized World

Associations of Universities	<i>G F Moodie</i>	485
Collective Bargaining in Higher Education	<i>C Allport, K McAlpine, and S Roberts</i>	491
Financing of Higher Education	<i>B Chapman and P Tulip</i>	499
Higher Education Crossing Borders	<i>J Knight</i>	507
Higher Education in Post-Conflict Societies	<i>M Daxner</i>	514
Higher Education in the New Economies	<i>J Fielden</i>	520
Human Resource Issues in Higher Education	<i>E El-Khawas</i>	527
International Accreditation in Higher Education	<i>A Stella and D Woodhouse</i>	533
Internationalization of Higher Education	<i>M van der Wende</i>	540
National and International Rankings of Higher Education	<i>S Marginson</i>	546
Steering of Higher Education Systems – The Role of the State	<i>K Yokoyama and V L Meek</i>	554
Student and Faculty Transnational Mobility in Higher Education	<i>U Lanzendorf and B M Kehm</i>	559
The Bologna Process in European Higher Education	<i>D Rich</i>	566
The Economics and Finance of Higher Education	<i>A Wolf</i>	573
The Role of the OECD in the Development of Higher Education in a Globalized World	<i>R Yelland</i>	584
Transforming Higher Education in Developing Countries: The Role of the World Bank	<i>J Salmi and R M Bassett</i>	590
UNESCO's Role in the Development of Higher Education in a Globalized World	<i>S Uvalić-Trumbić</i>	597

Higher Education – The Organization and Financing of Higher Education

Higher Education in Federal Systems	<i>M A M Cortés and U Teichler</i>	603
Pathways and Articulation into Higher Education	<i>G F Moodie</i>	609
Unitary, Binary and other National Systems of Higher Education	<i>L Goedegebuure</i>	616
Viewing Private Higher Education: How Much, Where, Why, and What?	<i>D C Levy</i>	622

I

Immigration

Children of Migrant Populations	<i>M M Suárez-Orozco and C Suárez-Orozco</i>	629
Third Culture Kids	<i>R E Van Reken</i>	636

International Comparisons

IEA Studies in ICT	<i>N Law</i>	644
IEA Studies in Mathematics and Science	<i>F K S Leung</i>	650
IEA Study in Civic Education	<i>J Torney-Purta, J-A Amadeo, and J Schwillie</i>	656
IEA: Globalization and Assessment	<i>H Wagemaker</i>	663

International Organizations in Education

Education and the European Union	<i>S Hegarty</i>	669
National University on Educational Planning and Administration, New Delhi	<i>J B G Tilak</i>	676
On School Management – International Council on School Effectiveness and Improvement, Commonwealth Council on Educational Administration and Management	<i>A Walker and T Townsend</i>	681
UNESCO's International Institute for Educational Planning	<i>S J Hite and M Bray</i>	688

L**Leadership and Management**

Leadership and Management Overview	<i>B Mulford</i>	695
------------------------------------	------------------	-----

Leadership and Management – as a Field of Study

Contexts for Leadership at the Beginning of the 21st Century	<i>M Bottery</i>	704
Educational Leadership: Philosophical Issues	<i>C W Evers</i>	710
Ethical/Moral Issues in Educational Leadership	<i>R J Starratt, L Langlois, and P Duignan</i>	717
History of Educational Leadership/Management	<i>R Bates</i>	724
Educational Leadership as a Field of Study: Professional Journals 1994–2006	<i>A R Thomas</i>	731
Research on Educational Leadership – Approaches/Promising Directions	<i>W A Firestone and V M J Robinson</i>	740

Leadership and Management – Leader Preparation

Leaders for Productive Schools	<i>J Murphy, S N Elliott, E B Goldring, and A C Porter</i>	746
New Approaches in Preparing School Leaders	<i>S G Huber</i>	752
Research for Leaders: The National College for School Leadership	<i>G Southworth</i>	762
School Leadership Standards	<i>L Ingvarson</i>	768

Leadership and Management – Leadership Issues

Effective Leadership in Challenging Schools	<i>A Harris</i>	777
Esthetics, Educational Leadership and Management	<i>R Bates</i>	783
Leader Mistreatment of Teachers	<i>J Blase and J Blase</i>	790
Leadership and Gender	<i>J Blackmore</i>	797
Leadership and Student Voice	<i>C Lanskey and J Rudduck</i>	803
Leadership and Urban Education	<i>K A Riley</i>	809
Leadership for Learning	<i>J MacBeath</i>	817
Leadership in Diverse Cultures	<i>A Walker and P K Yu-kwong</i>	824

Leading Diverse Schools	<i>J Ryan</i>	832
Supervision and Personnel Administration	<i>P D Tucker and D G Pounder</i>	839
The Big Challenge: Leadership for Inclusion	<i>M Ainscow and A Sandill</i>	846
The Educational Leader's New Work	<i>P Gronn</i>	852

VOLUME 5

Leadership and Management – Leadership Types

A Distributed Perspective on School Leadership and Management	<i>J P Spillane</i>	1
Leadership: Authentic	<i>P T Begley</i>	7
Leadership: Democratic	<i>J Møller</i>	12
Leadership: Instructional	<i>P Hallinger, K Leithwood, and R H Heck</i>	18
Leadership: Transformative	<i>C M Shields</i>	26
Strategic Leadership	<i>B Davies and B J Davies</i>	34
Teacher Leadership and Organizational Development	<i>A Harris</i>	40
Understanding How Leadership Influences Student Learning	<i>K Leithwood and B Levin</i>	45

Leadership and Management – Politics and Governance

Equity and Educational Effectiveness	<i>P Sammons</i>	51
Leading/Managing Schools in Communities Made Poor	<i>P Thomson</i>	58
Principals Role in Restructuring Schools	<i>G M Crow and K D Peterson</i>	66
School Reform and Restructuring: Self Managing School	<i>B J Caldwell</i>	72
System Leadership	<i>D Hopkins and R Higham</i>	78

Leadership and Management – School and Community

Educational Leadership and Social Capital	<i>J West-Burnham</i>	85
Networked Learning Communities School-to-School Collaboration as an Essential Component of a System Reform Strategy	<i>D Jackson</i>	92
Networks and Communities of Knowledge	<i>T Schuller and H Theisens</i>	100
School Community Relations	<i>R L Crowson and E B Goldring</i>	106
Social Capital, Educational Institutions and Leadership	<i>S Kilpatrick, S Johns, and B Mulford</i>	113

Leadership and Management – School Effectiveness and Improvement

Leadership and Technology	<i>C F Webber</i>	120
Leadership in the Implementation of Innovations	<i>I E Bogotch, T Townsend, and M Acker-Hocevar</i>	128
Leadership: School Improvement	<i>R H Heck and P Hallinger</i>	135
Organizational Learning in Schools	<i>B Mulford and H Silins</i>	143
Professional Learning Community	<i>L Stoll</i>	151
Transformational School Leadership	<i>K Leithwood</i>	158

Learning and Cognition – Issues and Concepts

Attention in Cognition and Early Learning	<i>M Heimann, T Tjus, and K Strid</i>	165
Concept Learning	<i>S J Unsworth and D L Medin</i>	172
Cooperative Learning	<i>R E Slavin</i>	177
Intelligence	<i>R J Sternberg</i>	184
Learning from Multiple Information Sources	<i>H I Strømsø and I Bråten</i>	191
Memory	<i>S Magnussen and T Brennen</i>	197
Metacognition	<i>L Baker</i>	204
Personal Epistemology in Education: Concepts, Issues, and Implications	<i>I Bråten</i>	211
Piaget: Recent Work	<i>L Smith</i>	218
The Adult Development of Cognition and Learning	<i>E A L Stine-Morrow and J M Parisi</i>	225
Vygotsky and Recent Developments	<i>J V Wertsch</i>	231

Learning and Cognition – Issues and Concepts – Focus on Cognition

Cognition and Emotion	<i>F Pons, M de Rosnay, and F Cuisinier</i>	237
Cognition: Overview and Recent Trends	<i>J Wiley and B D Jee</i>	245
Development of Creativity	<i>M A Runco</i>	251
Knowledge Domains and Domain Learning	<i>L Maggioni and P A Alexander</i>	255
Problem Solving and Human Expertise	<i>T J Nokes, C D Schunn, and M T H Chi</i>	265
Problem Solving and Reasoning	<i>R E Mayer</i>	273
Reasoning, Moral and Social	<i>N S Jampol, C B Richardson, and M Killen</i>	279
Remembering as Social Activity	<i>D Middleton</i>	285

Learning and Cognition – Issues, Concepts, Types

Computer-Supported Collaborative Learning: Basic Concepts, Multiple Perspectives, and Emerging Trends	<i>S R Ludvigsen and A I Mørch</i>	290
Learning as Inquiry	<i>S R Goldman, J Radinsky, S Tozer, and D Wink</i>	297
Learning Through Narrative Socialization	<i>A Fasulo</i>	303
Neuroscience Bases of Learning	<i>M H Immordino-Yang and K W Fischer</i>	310
The Neuroscience of Reading	<i>U Goswami</i>	317

Learning and Cognition – Issues, Concepts, Types – Focus on Learning

Learning Strategies	<i>C E Weinstein, J Jung, and T W Acee</i>	323
Learning Through Play	<i>K Aronsson</i>	330
Situated View of Learning	<i>A Collins and J G Greeno</i>	335
The Neuroscience of Aging and Cognition	<i>W J Hoyer and B Bessette-Symons</i>	340

Learning and Cognition – Language, Literacy and Subject-Related

An Overview of Language and Literacy in Educational Settings	<i>V G Aukrust</i>	345
Bilingualism and Learning	<i>L S Verplaetse and E Schmitt</i>	355
Classroom Discourse and Student Learning	<i>L Hemphill</i>	361
Early Writing	<i>B E Hagtvet</i>	367

First Language Acquisition	<i>N B Ratner</i>	375
Foreign Language Learning	<i>C Kinginger</i>	382
Learning Science	<i>J Leach and P Scott</i>	387
Learning to Read	<i>M S Burns and J K Kidd</i>	394
Mathematics Learning	<i>L Verschaffel, B Greer, and E De Corte</i>	401
Music Learning	<i>S Hallam</i>	407
Reading Comprehension: Reading for Learning	<i>C E Snow</i>	413
Second Language Learning	<i>J S Arnfast, J N Jørgensen, and A Holmen</i>	419
Writing: Advanced	<i>R Andrews</i>	426

Learning and Cognition – Other Issues

Gender and Learning	<i>L Yates and J McLeod</i>	432
Parent–Child Relationships in Early Learning	<i>C P Edwards, S M Sheridan, and L L Knoche</i>	438
Peer Interaction and Learning	<i>S Blum-Kulka and S Dvir-Gvirsman</i>	444

Learning and Cognition – Places of Learning

Learning in a Cross-Cultural Perspective	<i>L-F Zhang and R J Sternberg</i>	450
Learning Outside of School	<i>T Nunes</i>	457
Reasoning and Explanatory Talk: Learning in Everyday Settings	<i>D E Beals</i>	464

Learning and Cognition – Theoretical Perspectives – Learning

Apprenticeship Approach to Learning	<i>K Nielsen</i>	469
Connectionism and Learning	<i>T R Shultz</i>	476
Constructivism and Learning	<i>S Sjøberg</i>	485
Cultural–Historical Activity Theory	<i>A Havnes</i>	491
Learning in a Sociocultural Perspective	<i>R Säljö</i>	498
Self-Regulated Learning and Socio-Cognitive Theory	<i>P H Winne and A F Hadwin</i>	503

N

National Systems of Education

Australia	<i>C N Power</i>	510
Austria	<i>M J Gaertner</i>	517
Bangladesh	<i>R Z Dewan</i>	524
Barbados	<i>R Rudder</i>	534
Bulgaria	<i>A Chavdarova and D Vesselinov</i>	540
England	<i>C Whatford</i>	550
Estonia	<i>M Keränen and J Kõrgesaar</i>	555
Fiji	<i>I Senibulu</i>	563
Finland	<i>J E Tuovinen</i>	569
France	<i>M Sellier</i>	576

Germany	<i>A Hüfner and K Hüfner</i>	582
Ghana	<i>R Palmer</i>	590
Greece	<i>K Toura and A Karagiannis</i>	599
Iceland	<i>M Hardardottir and M Gunnlaugsdóttir</i>	605
Indonesia	<i>A C Alwasilah and Furqon</i>	611
Israel	<i>A Volansky</i>	618
Jamaica	<i>E Hannam</i>	627
Latvia	<i>A Joma</i>	630
Lebanon	<i>N A Mneimneh</i>	636
Lithuania	<i>J Navickaitė, L Paurienė, R Zablackė, A Tijūnelytė, and V Kožemiakina</i>	656
Malawi	<i>F R Mkandawire and D Mulera</i>	666
Malta	<i>C Farrugia</i>	673
Mongolia	<i>B Yembuu</i>	681
Namibia	<i>J Ellis</i>	687
New Zealand	<i>R Baker and C Wylie</i>	692
Niger	<i>A A I Daouda</i>	699
Norway	<i>H Farstad</i>	709
Oman	<i>S A Issan</i>	718
Pakistan	<i>N Mahmood and R Malik</i>	727
Poland	<i>A Janowski</i>	737
Qatar	<i>S R Salman</i>	744
Russia	<i>V M Filippov and G F Tkach</i>	750
Rwanda	<i>E Bahizi</i>	758
Scotland	<i>D Raffe</i>	770
Singapore	<i>P T Ng</i>	776
Slovenia	<i>T Plevnik and A B Lakota</i>	782
Solomon Islands	<i>D P Malasa</i>	790
South Africa	<i>P Kgobe and J Pampallis</i>	799
Spain	<i>A Tiana</i>	806
Sri Lanka	<i>B M J Balasooriya, M M Wehella, and S U Wijeratne</i>	813
Suriname	<i>H R Ori</i>	826
Taiwan	<i>C-M Lee</i>	834
Tanzania	<i>R Wedgwood</i>	841
Thailand	<i>C Rukspollmuang</i>	847
The Gambia	<i>Y A-M Jobe</i>	858
The Philippines	<i>E A P Valenzuela</i>	865
Uruguay	<i>B Guinovart and C Troncoso</i>	873
USA	<i>B Pusser and M E R Lloyd</i>	888
Zanzibar	<i>I M Hija</i>	896

VOLUME 6**P****Philosophy of Education**

Philosophy of Education: Overview	<i>P Enslin</i>	1
-----------------------------------	-----------------	---

Philosophy of Education – Contemporary Issues

Neoliberalism, the Market and Performativity	<i>M A Peters</i>	11
Parents and Children's Rights	<i>M Clayton</i>	17
Peace Education	<i>I Gur-Ze'ev</i>	22
Philosophy and Educational Research	<i>D Bridges</i>	29
Race, Critical Race Theory and Whiteness	<i>B Applebaum</i>	36
School Choice and the Common School	<i>G Haydon</i>	44

Philosophy of Education – Philosophical Perspective

African Philosophy of Education	<i>K Horsthemke</i>	50
Feminism	<i>L Stone</i>	56
Hermeneutics	<i>C Bingham</i>	63
Islamic Education	<i>Y Waghid</i>	69
Liberalism and Education	<i>W Feinberg</i>	74
Phenomenology and Existentialism	<i>B Levering</i>	80
Pragmatism	<i>B Thayer-Bacon</i>	86
Psychoanalysis and Education	<i>P Smeyers</i>	92
Social and Cultural Capital in Education	<i>M Print</i>	97
The Aims of Education	<i>R Marples</i>	103
The Analytical Tradition	<i>D Carr</i>	108

Philosophy of Education – Philosophical Themes

Adult Learning: Philosophical Issues	<i>D Beckett</i>	114
Autonomy	<i>C A Wringe</i>	121
Bildung	<i>J Masschelein and N Ricken</i>	127
Critical Theory and Pedagogy	<i>M Mason</i>	133
Critical Thinking	<i>H Siegel</i>	141
Environmental Education	<i>M Bonnett</i>	146
Identity	<i>M S Merry</i>	152
Inclusion	<i>R Cigman</i>	158
Justice and Care	<i>M Tjiattas</i>	164
Lifelong Learning	<i>J Halliday</i>	170
Post-Structuralism and Education	<i>M Mason and M Clarke</i>	175
Reason and Rationality	<i>P Hogan</i>	183

Religious and Spiritual Education	<i>J C Conroy and S J McKinney</i>	189
The Capabilities Approach	<i>L Terzi</i>	194

Primary and Secondary Education – Learning and Teaching in School Age Education

Class Size – Arguments and Evidence	<i>P Blatchford and K C Lai</i>	200
Continuing Professional Development of Teachers	<i>P Earley</i>	207
Homeschooling	<i>P Farenga</i>	214
Innovations and Early Intervention in the Teaching of Literacy	<i>R E Slavin</i>	221
Personalized Learning in School Age Education	<i>D Hopkins</i>	227
Perspectives on Schooling in the Middle Years	<i>V J Russell</i>	233
Reform of the School Workforce	<i>H M Gunter and G Forrester</i>	242
Secondary Education in Developing Countries	<i>P Hughes and A Walker</i>	249

Primary and Secondary Education – Social Justice Issues in School Age Education

Community Focused Schools	<i>A Dyson</i>	255
---------------------------	----------------	-----

Primary and Secondary Education – The Effectiveness of School Systems

Reforming the American School System	<i>M Barber</i>	261
School Effectiveness in Developed Societies	<i>D Reynolds</i>	269
School to School Collaboration: Innovation and Improvement	<i>R Higham</i>	274
Studies of School Improvement in Developing Countries	<i>S E Anderson</i>	281
Sustainable Educational Reform	<i>A Hargreaves</i>	289
The Role of the District in Tri Level Reform	<i>M Fullan</i>	295
Whole School Designs for Enhancing Student Achievement	<i>S Stringfield and J Nunnery</i>	303

Q

Qualitative Research

Action Research in Education	<i>E Stringer</i>	311
Classroom Ethnography	<i>F Erickson</i>	320
Computer Assisted Qualitative Data Analysis	<i>C Silver and A Lewins</i>	326
Conversational Analysis	<i>H G Levine</i>	335
Critical Ethnography	<i>D Foley</i>	337
Critical Race Theory	<i>D Gillborn and G Ladson-Billings</i>	341
Critical Theory	<i>J J Ryoo and P McLaren</i>	348
Cultural Studies	<i>R Gaztambide-Fernández and K Cairns</i>	354
Cultural–Historical Activity Theory	<i>Laboratory for Comparative Human Cognition</i>	360
Design Experiments	<i>A Collins</i>	367
Discourse Analysis	<i>P Nikander and S Vehviläinen</i>	373
Document Analysis	<i>T Rapley and K N Jenkins</i>	380
Ethnography	<i>M Hammersley</i>	386

Ethnomethodology in Education Research	<i>D Macbeth</i>	392
Focus Groups	<i>D L Morgan and K Hoffman</i>	401
Grounded Theory	<i>K Charmaz and A Bryant</i>	406
Hermeneutics	<i>J Dyer</i>	413
Interpretive Research	<i>E Rauscher and M E Graue</i>	419
Interviews and Interviewing	<i>A Marvasti</i>	424
Life History	<i>J A Hatch and S K Newsom</i>	430
Narrative Inquiry	<i>D J Clandinin and J Huber</i>	436
Participant Observation	<i>J Schostak</i>	442
Phenomenology	<i>M van Manen and C A Adams</i>	449
Qualitative Case Studies	<i>S B Merriam</i>	456
Qualitative Data Management	<i>C Grbich</i>	463
Semiotics	<i>M Danesi</i>	475
Symbolic Interactionism, Naturalistic Inquiry, and Education	<i>J A Forte</i>	481
The Quality of Evidence in Qualitative Research	<i>S K Biklen</i>	488
Visual Data in Education Research	<i>J Wagner</i>	498

Qualitative Research – Standards of Proof

Validity: Mapping Diverse Perspectives	<i>G B Rossman, S F Rallis, and A M Kuntz</i>	505
--	---	-----

Quantitative and Qualitative Approaches to Research – Integration

Mixed Methods	<i>A Tashakkori and I Newman</i>	514
---------------	----------------------------------	-----

Quantitative Methods and Research Design

External Validity	<i>G E Matt, A Brewer, and M Sklar</i>	521
-------------------	--	-----

R

Role of NGOs in Globalization

NGOs and Globalization of Education	<i>M Ahmed</i>	529
-------------------------------------	----------------	-----

S

Social and Emotional Aspects of Learning

Achievement Goal Theory: Definitions, Correlates, and Unresolved Questions	<i>T Urdan</i>	537
Affect, Mood and Emotions	<i>S Schnall</i>	544
Anxiety	<i>M Zeidner</i>	549
Attribution Theory	<i>B Weiner</i>	558
Children's Friendship	<i>S R Asher and K L McDonald</i>	564
Coping with Stressful Situations: An Important Aspect of Self-Regulation	<i>M Boekaerts</i>	570
Culture in Motivation Research: A Challenging and Enriching Contribution	<i>K Kimmel and S Volet</i>	576
Early Social Development and Schooling	<i>G M van der Aalsvoort</i>	585

Emotion in Educational Contexts	<i>P A Schutz, P D Quijada, S de Vries, and M Lynde</i>	591
Empathy	<i>R Arnold</i>	597
Flow in Education	<i>J A Schmidt</i>	605
Interest	<i>M Ainley</i>	612
Intrinsic and Extrinsic Motivation	<i>C Levesque, K J Copeland, M D Pattie, and E L Deci</i>	618
Motivating Students in Classrooms	<i>J Brophy</i>	624
Motivation Regulation	<i>C A Wolters and S A Mueller</i>	631
Peer Learning in the Classroom	<i>N M Webb</i>	636
Poverty, Effects of on Social and Emotional Development	<i>N Philipsen Hetzner, A D Johnson, and J Brooks-Gunn</i>	643
Seeking Help as an Adaptive Response to Learning Difficulties: Person, Situation, and Developmental Influences	<i>S A Karabenick and R S Newman</i>	653
Self-Concept in Learning: Reciprocal effects model between academic self-concept and academic achievement	<i>H W Marsh and L F Scalas</i>	660
Self-Efficacy Beliefs	<i>D H Schunk and F Pajares</i>	668
Social and Emotional Outcomes of Learning	<i>J Vorhaus</i>	673
Social Aspects of Collaborative Learning	<i>M Arvaja and P Häkkinen</i>	685
Social Development and Schooling	<i>G W Ladd</i>	691
Social Interaction and Learning	<i>K Littleton</i>	698
Social Networks and the Education of Children and Youth	<i>B Schneider, T G Ford, and L Perez-Felkner</i>	705
Sociocultural Issues in Motivation	<i>R A Walker</i>	712
Socioemotional Aspects of Technology-Supported Learning	<i>S Järvelä</i>	718
Volitional Control of Learning	<i>J Husman and L Corno</i>	724
Wellbeing	<i>T Hascher</i>	732

VOLUME 7

Statistics

An Overview of Statistics in Education	<i>S Sinharay</i>	2
Analysis and Interpretation of Multivariate Data	<i>D J Bartholomew</i>	12
Analysis of Covariance	<i>N T Longford</i>	18
Analysis of Extreme Values in Education	<i>J Beirlant, G Dierckx, K Luwel, and P Onghena</i>	25
Analysis of Variance	<i>B M King</i>	32
Bayesian Statistical Analysis	<i>G Kokolakis</i>	37
Bootstrap Method	<i>K Singh and M Xie</i>	46
Canonical Correlation	<i>A Basu and A Mandal</i>	52
Categorical Data Analysis	<i>S E Fienberg and B W Junker</i>	58
Causal Inference	<i>D B Rubin</i>	66
Cluster Analysis: Overview	<i>S Landau and I Chis Ster</i>	72

Cognitive Psychology and Educational Statistics	<i>H S Kim and S Embretson</i>	84
Computational Statistics	<i>J E Gentle</i>	93
Continuous Probability Distributions	<i>S Sinharay</i>	98
Correspondence Analysis	<i>M Greenacre</i>	103
Data Mining	<i>R S J d Baker</i>	112
Decision Theory	<i>M Mendoza and E Gutiérrez-Peña</i>	119
Design of Experiments	<i>C A Albers and T R Kratochwill</i>	125
Discrete Probability Distributions	<i>S Sinharay</i>	132
Discrimination and Classification	<i>R C Neath and M S Johnson</i>	135
Empirical Bayes Methods	<i>E Ip and G Molenberghs</i>	142
Evaluation Research	<i>T Kellaghan</i>	150
Exploratory Data Analysis	<i>A Unwin</i>	156
Factor Analysis: An Overview and Some Contemporary Advances	<i>F A N Alhija</i>	162
Generalized Linear Mixed Models	<i>S Rabe-Hesketh and A Skrondal</i>	171
Generalized Linear Models	<i>J J Faraway</i>	178
Generating Random Numbers	<i>P J Pashley and A Amodeo</i>	184
Goodness-of-Fit Testing	<i>A Maydeu-Olivares and C García-Forero</i>	190
Graphical Models	<i>R G Almond</i>	197
Growth Modeling	<i>I Plewis</i>	203
Hierarchical Linear Models	<i>D Rindskopf</i>	210
Hypothesis Testing and Confidence Intervals	<i>T R Konold and X Fan</i>	216
Instrumental Variables	<i>E P Bettinger</i>	223
Jackknife Methods	<i>S Sinharay</i>	229
Large-sample Statistical Methods	<i>J M Singer and D F Andrade</i>	232
Latent Class Models	<i>J K Vermunt</i>	238
Markov Chain Monte Carlo	<i>A M Johansen</i>	245
Matrix Algebra	<i>H Yanai and H Ishii</i>	253
Measure of Association	<i>G-H Huang</i>	260
Measures of Central Tendency	<i>D Clark-Carter</i>	264
Measures of Dispersion, Skewness and Kurtosis	<i>D Rindskopf and M Shiyko</i>	267
Meta Analysis	<i>J Sánchez-Meca and F Marín-Martínez</i>	274
Missing Data	<i>C A W Glas</i>	283
Model Selection	<i>S J Haberman</i>	289
Monte Carlo Methods	<i>A M Johansen</i>	296
Multidimensional Scaling	<i>Z Zhang and Y Takane</i>	304
Multiple Comparisons	<i>C Lewis</i>	312
Multivariate Analysis of Variance	<i>N T Longford</i>	319
Multivariate Linear Regression	<i>S Ganesh</i>	324
Multivariate Normal Distribution	<i>J Tacq</i>	332

Nonlinear Regression Analysis	<i>H-H Huang, C K Hsiao, and S-Y Huang</i>	339
Nonparametric Statistical Methods	<i>K Sijtsma and W H M Emons</i>	347
Observational Studies	<i>P R Rosenbaum</i>	354
Order Statistics	<i>I Bairamov</i>	360
Point Estimation Methods with Applications to Item Response Theory Models	<i>F Bartolucci and L Scrucca</i>	366
Principal Components Analysis	<i>D J Bartholomew</i>	374
Probability Theory	<i>T Rudas</i>	378
Recursive Partitioning	<i>S Landau and S Barthel</i>	383
Sampling	<i>J Qian</i>	390
Sequential Probability Ratio Test	<i>T J H M Eggen</i>	396
Sequential Testing	<i>H J Vos</i>	401
Signal Detection Theory	<i>H Abdi</i>	407
Small Area Estimation	<i>A Noble</i>	411
Statistical Analysis of Functional Data	<i>F Ferraty, P Sarda, and P Vieu</i>	417
Statistical Inequalities	<i>H M Mahmoud and J L Gastwirth</i>	426
Statistical Paradoxes	<i>Y-K Tu</i>	429
Statistical Power Analysis	<i>L V Hedges and C Rhoads</i>	436
Statistical Significance Versus Effect Size	<i>X Fan and T R Konold</i>	444
Stochastic Processes	<i>F Baudoin</i>	451
Structural Equation Models	<i>S-Y Lee and X-Y Song</i>	453
Survival Data Analysis	<i>A F De Champlain</i>	459
The Normal Distribution and its Applications	<i>J Tacq</i>	467
Time Series Analysis	<i>P J Brockwell</i>	474
Univariate Linear Regression	<i>M L Hazelton</i>	482
Value-Added Models	<i>D F McCaffrey</i>	489

T

Teacher Education – Evaluation in the Teacher Education Program

Accreditation and Standards in Teacher Education	<i>R J Yinger and K L Daniel</i>	495
Determining Long Term Effects of Teacher Education	<i>C N Brouwer</i>	503

Teacher Education – Evaluation of the Prospective Teacher

The Evaluation of Prospective Teachers in Teacher Education	<i>D E H Tigelaar and J van Tartwijk</i>	511
---	--	-----

Teacher Education – Macro Factors

Diversity in Teacher Education	<i>T Wubbels</i>	518
Social and Cultural Influences on Teacher Education	<i>M Ben-Peretz and R Lotan</i>	525
The Role of Governance in Teacher Education	<i>G E Karlsen</i>	532

Teacher Education – Post-Initial Professional Development

Action Research as a Tool for Teachers' Professional Development	<i>P Ponte</i>	540
Contemporary Approaches to Teacher Professional Development	<i>H Borko, J Jacobs, and K Koellner</i>	548
School Development for Teacher Learning and Change	<i>P Sleegers and K Leithwood</i>	557
Teacher Induction	<i>D Beijaard, J Buitink, and C Kessels</i>	563
Teacher Learning as Workplace Learning	<i>J Imants and K van Veen</i>	569
Teacher Learning with Lesson Study	<i>A Murata</i>	575
Teachers Career Stages and Professional Development	<i>R Fessler and E Rice</i>	582

Teacher Education – Preservice Teacher Education

A Pedagogy of Teacher Education	<i>J Loughran</i>	587
Case Methods in Teacher Education	<i>M Sato and C Rogers</i>	592
Conceptions of Teacher Education	<i>J A Whitcomb</i>	598
Inquiry-Oriented Teacher Education	<i>R Schulz</i>	604
Narratives and Biography in Teacher Education	<i>G Kelchtermans</i>	610
Supervision in Teacher Education	<i>J McCarthy and L F Quinn</i>	615
Taking Prospective Teachers' Beliefs into Account in Teacher Education	<i>M Valcke, G Sang, I Rots, and R Hermans</i>	622
Teacher Education and Models of Teacher Reflection	<i>V K LaBoskey</i>	629
Teacher Education as Teaching for Understanding with New Technologies	<i>M S Wiske and D E Spicer</i>	635

Teacher Education – Preservice: The Knowledge Base

Experienced Teachers' Craft Knowledge	<i>P C Meijer</i>	642
Moral Values in Teacher Education	<i>W Veugelers</i>	650
Pedagogical Content Knowledge	<i>J H Van Driel and A Berry</i>	656
Teacher Education and the Educational Foundations Knowledge Base	<i>F B Murray</i>	662
The Relationship Between Theory and Practice in Teacher Education	<i>F A J Korthagen</i>	669

Teacher Education – Teacher Educators

Characteristics, Scholarship and Research of Teacher Educators	<i>M Lunenberg</i>	676
Professional Development of Teacher Educators	<i>P K Smith</i>	681
Self-Study by Teacher Educators	<i>T Russell</i>	689

Teacher Education – Teaching Specific Domains

Arts Teacher Education	<i>L Beudert</i>	695
Mathematics Teacher Education	<i>K Krainer and S Llinares</i>	702
Music Teacher Education	<i>M W-y Wong</i>	706
Science Teacher Education	<i>J H Van Driel and S K Abell</i>	712
Social Studies Teacher Education	<i>A-L Halvorsen and S M Wilson</i>	719
Teacher Education for Elementary Education	<i>K D Brown and G N Carnes</i>	725

Teacher Education for Teaching the Gifted	<i>A Y Baldwin</i>	732
Teacher Preparation in Mother Tongue Education	<i>M Kooy and H Hulshof</i>	738
Vocational Teacher Education	<i>S Johansson</i>	744

Teacher Education – The Cooperating School

Mentoring in Teacher Education	<i>P D Tomlinson, A J Hobson, and A Malderez</i>	749
Partnerships Between Schools and Higher Education	<i>C Day and L Smethem</i>	757

VOLUME 8

Technology and Learning

An Overview of Technology and Learning	<i>B Means and J Roschelle</i>	1
Educational Data Modeling	<i>S Iksal</i>	11
Student Data Systems and Their Use for Educational Improvement	<i>J C Wayman, V Cho, and M P Richards</i>	14
Technology and Learning: Access in Schools Around the World	<i>R Anderson and J Ainley</i>	21

Technology and Learning – as a Subject of Instruction

Conceptions of Technology Literacy and Fluency	<i>O Erstad</i>	34
--	-----------------	----

Technology and Learning – Assessment

Technology and Formative Assessment	<i>J W Pellegrino</i>	42
Technology for Large-Scale Assessment	<i>R E Bennett</i>	48
Technology Supports for Assessment Design	<i>R J Mislevy, F I Winters, I I Bejar, R E Bennett, and G D Haertel</i>	56

Technology and Learning – Education Reform and Economic Development

National Strategies to Build a Technology Workforce	<i>H-M Cheah</i>	66
Public–Private Partnerships for Educational Reform	<i>V Quah, C P Lim, and C Brook</i>	73
Relating Technology, Education Reform and Economic Development	<i>R B Kozma</i>	81

Technology and Learning – Evaluating Technology

Examining the Effects of Educational Technology Programs: Challenges and Strategies	<i>M Russell</i>	88
---	------------------	----

Technology and Learning – Global Trends

Equity in Technology Access and Opportunities	<i>T Matuchniak and M Warschauer</i>	95
Internet-Based Education	<i>M Bakia</i>	102

Technology and Learning – Issues

Media Use and School Achievement	<i>K Roe</i>	109
Technology and Physical and Social Health	<i>K Subrahmanyam</i>	112

Technology and Learning – Meeting Special Students’ Needs

Universal Design for Learning *D H Rose and J W Gravel* 119

Technology and Learning – School Leadership for Technology Integration

Technology as a Support for School–Community Connections *K M Culp* 125

Technology and Learning – Supports for Managing Instruction

Classroom Uses of Technology to Manage Instruction *W R Penuel* 133

Technology and Learning – Supports for Skill Learning

New Media, Learning from *W Schnotz and H Horz* 140

Reading and Technology *N DeFrance, D Khasnabis, and A S Palincsar* 150

Technological Supports for Acquiring Twenty-First-Century Skills *C Dede* 158

Technology Supports for Acquiring Inquiry Skills *T de Jong* 167

Technology Supports for Acquiring Mathematics *M J Nathan* 172

Technology Supports for Lifelong Learning *A Collins and R Halverson* 184

Technology and Learning – Supports for Student Collaboration

Knowledge Communities in the Classroom *J D Slotta and H Najafi* 189

Technology and Learning – Supports for Subject Matter Learning

Technology Supports for Science Learning *C Delgado and J Krajcik* 197

Technology Supports for Second Language Learning *D Nunan* 204

Technology and Learning – Teacher Learning

Teacher Skills and Knowledge for Technology Integration *N Law* 211

Technology in Preservice Teacher Education *N Davis* 217

Technology Resources for Teacher Learning *J E Hinojosa, C Labbé, and L López* 222

V

Vocational Education and Training

An Overview of Vocational Education and Training *T Karmel* 229

Informal and Non-Formal Learning in Vocational Education and Training *F Beddie and S Halliday-Wynes* 240

School-Based Vocational Education and Training *B Knight and R Sweet* 247

The Economics of Vocational Education and Training *P Carneiro, L Dearden, and A Vignoles* 255

Vocational Education and Training – Individuals and Students

Age Equality in Education and Training *P Taylor* 262

Career Guidance in Vocational Education and Training and in the Workplace *P Plant* 269

Displaced Workers, Unemployed and Vocational Education and Training *B Kriechel* 275

Immigrant Investment in Host Country Vocational Education and Training *I Kogan* 280

Indigenous People and Vocational Education and Training	<i>R G Schwab</i>	289
Persons with a Disability and Vocational Education and Training	<i>F Ferrier and C S Smith</i>	294
Prisoner Education and Training	<i>S Dawe, S H Wynes, and J Misko</i>	300
Vocational Education and Training and the School-to-Work Transition	<i>G Hayward</i>	306

Vocational Education and Training – Industry and Employers

Apprenticeships	<i>E Smith</i>	312
Skills Shortages: Concepts, Measurement and Policy Responses	<i>C Shah and G Burke</i>	320
Training and Learning in the Workplace	<i>R L Jacobs</i>	328
Training in Small Businesses	<i>H Matlay</i>	337

Vocational Education and Training – Research on VET

Language of Vocational Education and Training	<i>J Misko, J MacKenzie, and R-A Polvere</i>	343
Research on Vocational Education and Training	<i>J Robson and B Bailey</i>	352
Vocational Education and Training Information Resources	<i>L-A Harris, R Naidu, and M Saunders</i>	358

Vocational Education and Training – Teaching and Learning

Employability Skills	<i>E Smith</i>	368
Learning Styles in Vocational Education and Training	<i>J Vorhaus</i>	376
Modularization in Vocational Education and Training	<i>H Ertl and G Hayward</i>	383
Personalized Learning and Vocational Education and Training	<i>I Duckett</i>	391
Recognition of Prior Learning and Vocational Education and Training	<i>J Keating</i>	397
The Competency Model	<i>A Gonczi and P Hager</i>	403

Vocational Education and Training – VET in Context

Globalization and Vocational Education and Training	<i>P Loveder</i>	411
Learning to Work for the Future	<i>J Fien, R Maclean, and D N Wilson</i>	416
The Relationship between Technical and Vocational Education and Training and Innovation	<i>K Ruth and L Deitmer</i>	423
The Status of Vocational Education and Training	<i>J Lasonen</i>	429
Vocational Education and Training in Post-Conflict Countries	<i>E Lyby</i>	441

Vocational Education and Training – VET System

Dual System	<i>T Deissinger</i>	448
Industry Involvement in the Vocational Education and Training System	<i>M Hughes and C Hughes</i>	455
Non-Completion in Vocational Education and Training: Issues, Patterns and Possibilities	<i>S Lamb and A Walstab</i>	463
Planning and Policy Development for Technical Vocational Education and Training Systems	<i>R Maclean and M Pavlova</i>	469
Qualifications Frameworks and their Role in the Reform of Education and Training	<i>M Young</i>	476
Quality Assurance in Vocational Education and Training	<i>D Gunning</i>	482

Training Finance	<i>A Ziderman</i>	489
Training Markets	<i>P Ryan</i>	496
Vocational Education and Training Teacher Training	<i>S Nielsen</i>	503
Vocational Education and Training Workforce	<i>D L Wallin</i>	513
Subject Index		519
Contributors		747

P

PHILOSOPHY OF EDUCATION

Philosophy of Education: Overview

P Enslin, University of Glasgow, Glasgow, UK

© 2010 Elsevier Ltd. All rights reserved.

How Do Philosophers of Education Contribute to Educational Enquiry?

What do philosophers of education do and how does their work relate to other ways of studying and reflecting on education? As in all areas of philosophical work, what philosophers of education do, and should do, is a matter of ongoing debate – especially among philosophers themselves. Indeed, philosophy, in general, is characteristically preoccupied with reflection on its own purposes and methods – concerning which there is ongoing dispute. At the same time, philosophy of education is, sometimes, assumed – from outside the discipline – to be a matter of providing definitions of complex concepts, of articulating a personal manifesto or a social group's shared educational ideals, of adopting the educational ideals of one of a number of rival schools of thought or 'isms', such as empiricism and rationalism, or, perhaps, of applying the wise reflections of great thinkers, such as Plato or Rousseau.

Although not all of these possibilities will be thoroughly rejected in the introductory overview that follows, the collection of philosophical articles gathered in this encyclopedia suggests a different and more complex view of what is distinctive about philosophy of education. However, a useful place to start introducing this selection of articles by philosophers of education, the range of approaches they exemplify, and issues they address, is precisely with some attention to what the discipline is not.

Philosophy of education should not be assumed to be a matter of providing definitions, studying the ideas of great thinkers, or of selecting and applying one or another theoretical position to the practice of education. Instead, philosophers of education set out to throw light on the meaning and aims of education, as well as on a

range of problems to which this project gives rise. As a preliminary indication, it can be said that this line of enquiry involves analyzing arguments, identifying and probing assumptions, assessing claims, clarifying concepts, and more – to which we will return in the course of this overview.

The Role of Definitions

The types of problems that the philosophy of education addresses are not amenable to quick solution by offering definitions. A short definition will inevitably be contestable: so complex are many issues in education and other fields that, while a working definition or initial characterization may, sometimes, serve to get a discussion going – probably for later reformulation – a considered account that grapples with conceptual and other complexities is likely to be required (see, e.g., the extended treatment of the concept of experience in 'Adult learning: Philosophical issues,' in which Beckett uses an account of experience to develop a social, intersubjective account of learning).

For example, to define education as being centrally about learning might look succinct and clear. However, what about learning that might be undesirable, such as learning to lie or to be a thief? Also, what of possibly crucial considerations about the nature of education that do not focus just on learning, but also on well-being or the development of capacities such as, say, autonomy? As these examples suggest, a definition may be too broad or too narrow, failing to exclude features necessary to a concept's meaning or failing to include necessary elements. At best, it has a useful role to play in starting exploration of a philosophical issue, rather than settling it (see how Marples' article on 'Aims of education' avoids

defining the aim of education and, instead, explores competing expectations of publicly funded education, i.e., preparation for work, providing knowledge for its own sake, and for well-being).

Philosophy of Education as Expressions of Personal or Group Beliefs

Nor can philosophy of education be equated with a philosophy in the sense of a set of beliefs, whether interpreted as a kind of personal philosophy, as in my personal philosophy of teaching, or a philosophy of life or worldview and its accompanying doctrine of education, shared with a group of others. Such a loose use of the term “philosophy” belies its necessary critical function, its role in probing assumptions and sets of beliefs, however deeply and dearly individual persons might hold them. Yet, having said this, two philosophical articles in this collection do raise current concerns with regard to the significance and status of various groups’ shared views concerning the nature and aims of education (‘African philosophy of education’ and ‘Towards an Islamic philosophy of education’) – an issue to which we will return later.

Schools of Thought

A more viable candidate for the name ‘philosophy’ in educational inquiry than individuals’ or groups’ beliefs could be one of a number of theories that offer more systematic perspectives – often named by a term that includes the noun-forming suffix *ism*, such as ‘rationalism’ or ‘empiricism’ (which differ with regard to the extent to which we depend on experience in acquiring knowledge), or realism and idealism (which differ with regard to whether material things exist independently of minds that perceive them). This common interpretation of philosophy of education was evident in teacher-education curricula across several decades and retains a presence, if less widely so. Accompanying textbooks (e.g., Brubacher, 1962) presented a range of competing schools of thought and readers were commonly invited to consider which best applied to understanding the practice of education, and to extrapolate implications for a range of problems, which could be applied in their work as educators.

One obvious problem with such an approach is that while some such competing theories marked out by the *ism* suffix stand up to critical scrutiny better than others, the range of theoretical positions available for consideration and scrutiny is no longer limited to a set of distinct and clearly opposed positions, as might previously have been the case. Some of the many positions now identified by the *ism* suffix – quite a few of which are relatively new compared with those traditional schools that featured in

textbooks based on a fairly neat typology – are controversial because they are widely agreed among philosophers of education to undermine the educational enterprise. To take a contemporary example – in recent years, philosophers of education have devoted considerable attention to neoliberalism as a contemporary doctrine that has exerted a destructive social influence – not least in education (for an account of neoliberalism and its influence in education, see Peters on ‘Neoliberalism, the market and performativity’ and Halliday’s exploration of the shift in social policy from social justice to economic rationalism in his article on ‘Lifelong learning’). Some such theoretical positions, then, are interesting not because they might offer a possible philosophical stance to adopt as one’s own, but rather because of the importance of subjecting them to ongoing criticism. This is what is wrong about them that often makes them worthy of attention, especially if they are widely held and influential.

Noting the presence of various *isms* in educational discourse prompts us, if considering them philosophically, to raise questions about how we might go about evaluating their competing claims to knowledge concerning the world, including what education is about and how it ought to be pursued. While there is no neutral position from which to evaluate their claims, philosophically regarded, all must be open to criticism, and it is probable that none will successfully claim adherence to the exclusion of all others. To think, furthermore, that one might be able to do so would be to confuse the activity of philosophy with doctrine or dogma.

In addition, there is likely, in any case, to be some diversity and even disagreement among writers who have contributed to the development of any of these theoretical approaches (see Levering’s account of the relationship between Sartre and Camus as existentialists in ‘Phenomenology and existentialism,’ which shows that their views diverge). To assume that a choice could be made among theories as alternative doctrines that might be adopted – or simply learned from – to the possible exclusion of others would be to mistake the nature of the discipline. Feminism is a case in point. As Stone’s contribution (‘Feminism’) illustrates, there is great diversity within the range of theories loosely described as feminist theory, although all share a commitment to ending gender oppression. An informed and sensitive reading of the various approaches to theorizing in education reveals an ebb and flow of, sometimes, overlapping influences (see Mason on ‘Critical theory and critical pedagogy,’ which shows diverse influences on critical theory, including Hegel and Marx as well as elements of Anglo-American thought). The complex interrelationships among the theoretical trends, tendencies, and positions addressed in this collection – which number several dozens – show that developments in social and educational ideas may prompt established theoretical approaches to adjust some aspects

of their position or methodology to accommodate new insights or to meet objections to weaknesses. Some might be obviously opposed but others stand in more ambiguous and arguably more synergistic relationships to one another.

Yet, in spite of these qualifications, it remains the case that a number of theoretical positions and orientations exert considerable influence both on ideas about education and in the discipline of philosophy of education. Of these, pragmatism's continuing influence is a prominent example. Like other major approaches discussed in this collection, pragmatism does not present itself as a single, unified doctrine vying with one or two particular others for our loyalty. In addition, uptake of pragmatism in recent work is not a restatement of earlier ways in which philosophers of education have engaged with this current in philosophy (see Thayer-Bacon on 'Pragmatism').

The usefulness of various theoretical movements shifts over time, both in our interpretation of them and in the multiple ways in which they can illuminate the changing educational scene. Marxism, for example, is less prominently present in educational enquiry than it was in the 1970s and 1980s, arguably as it has been overtaken as an explicit focus of theoretical interest by post-structuralism (see Mason and Clarke on 'Post-structuralism and education') – although its influence as a strand in the history of educational thought continues (see 'Hermeneutics,' 'Critical Theory and critical pedagogy,' 'Peace education') and there has been a recent resurgence of interest in the Marxist theory's relevance to education (Rikowski, 2007).

An approach to understanding philosophy of education, focused on competing schools of thought, shares some difficulties with a related misapprehension, which is that philosophical inquiry about education is a matter of studying the writings of great thinkers – the classic texts in the history of educational thought.

Great Thinkers

Alongside, a tendency to present philosophy of education as a catalog of competing theoretical positions has been a similar approach in which the ideas of great thinkers such as Plato, Kant, or Dewey are presented for respectful consideration and application to contemporary educational issues. Indeed, education has been an issue of interest to many of the major figures in the history of philosophy, broadly viewed across its traditional branches, especially those of metaphysics, epistemology, ethics, and political philosophy.

Although major figures do undoubtedly enjoy a strong presence in educational thought (see, e.g., the frequent references to Wittgenstein, Dewey, and Foucault in this collection), philosophy of education cannot be reduced in this way to an anthology of their ideas – for reasons

already largely indicated above. None has succeeded in articulating a single grand theory of education that successfully addresses all educational issues and controversies, and the very idea that this might be possible has been thoroughly debunked by recent post-structuralist influences (see, e.g., Lyotard, 1984). None has achieved a body of work of such wide reach or consistent success. All have been vulnerable to generative critique from competing theories, or from other philosophers – such as in the cases of Aristotle or Plato or Hegel or Kant – or from others working within a closely related approach who dispute particular issues on which their views and arguments differ. In addition, all four of these philosophers retain an influential presence today – indeed, most of contemporary philosophy of education can be located within either the Kantian or Hegelian traditions. As we see in the current selection of articles on philosophy of education, engagement with the ideas of key figures in the history of educational ideas remains a strong component of what philosophers do.

In philosophy of education, however, such engagement does not take the form of deferential, axiomatic application of the ideas of major thinkers to classroom practice by simple extrapolation from theory to practice. Even the most astute philosophical insights require critical evaluation and interpretation on the way to application, and different issues in policy and practice might benefit best from insights of different writers. Our beliefs, values, knowledge, and educational principles comprise a complex web from which it is not possible to draw conclusions in the form of principles for application in educational situations and practice (Hirst, 1966). Even those who agree on certain beliefs may support different views with regard to education. No straightforward deductive relationship is plausible here. Furthermore, some aspects of the theories of even the most innovative of thinkers are simply more convincing than others and there is much philosophical work to be done, requiring a range of critical tools, in deciding which these are, which may, in turn, depend on the relevant educational and background contexts.

However, to observe that neither major schools of thought nor the work of great thinkers represent what philosophy of education offers to educational enquiry is not to thoroughly dismiss their presence in educational thought. Instead – and while acknowledging the need for a sense of the history of philosophical thought about education – we need to see their role differently, in relation to the use of several critical tools, to which we now turn.

Philosophy of Education as a Distinctive Type of Educational Inquiry

A useful way to distinguish between philosophical and other approaches to educational inquiry is by taking into

account the distinction between first- and second-order questions and discourses. In education, first-order discourse could comprise everyday talk concerning classroom practice, or policy discussion with regard to educational situations and issues. Take, for example, a call for education to be inclusive. Second-order philosophical discourse would focus on the concept of inclusion differently, reflecting on it to develop a critical discourse that tries to clarify the concept of inclusion and similarly complex, related concepts – to scrutinize assumptions and arguments concerning them, and to develop sound arguments about the issues they raise. Thus, Cigman's article ('Inclusion') probes the meaning of inclusion by exploring the conceptual history of inclusive education – from its early formulation in terms of handicap, through attention to the meaning of related concepts such as fairness as equality and fairness as attention to individual or special need, and integration and inclusion.

Philosophers of education work with concepts in several ways. A central feature of analytic philosophy of education (see Carr on 'The analytical tradition') – whose history will be discussed shortly – has been to clarify the meaning of concepts, which can involve several dimensions. The articles in this collection, which exemplify clarification of the meaning of concepts, include Applebaum's analysis of race, racism, and whiteness ('Race, critical race theory, and whiteness'). In order to develop her account of racism, Applebaum emphasizes both the relevance of related concepts such as systematic oppression, and also distinguishes, crucially, between race and ethnicity. A number of articles raise and use well-established conceptual distinctions that continue to illuminate; for example, Terzi reminds us of that between schooling and education ('The capabilities approach'), Hogan of that between instrumental and practical rationality ('Reason and rationality'), and Clayton of the distinction between rights in law and moral rights ('Parents' rights and children's rights'). By contrast, some of the articles question the usefulness of old, accepted distinctions – as in Bridges' observation that the distinction between objectivity and subjectivity is, sometimes, expressed too crudely ('Philosophy and educational research') and in Mason's mention of the Frankfurt School's objection to the separation of questions with regard to values and interests from those concerning the production of what are regarded as objective facts ('Critical theory and critical pedagogy'). Gur-Ze'ev suggests in his proposals for a different focus on peace education that philosophers might begin by analyzing the concepts of peace and education as reciprocal ('Peace education').

In their work on concepts and issues of significance for education, philosophers of education also pay attention to argument – again, in several ways. The most obvious of these is close scrutiny of arguments, their structure, and the assumptions on which they rest, as well as the ways in

which concepts are used. Siegel ('Critical thinking') questions the coherence of reasoned arguments against reason itself in criticisms of critical thinking as an educational ideal. Tjiattas' article constructs an argument against the popular assumption that, in moral education, theories of care and of justice are strongly opposed, rival positions ('Justice and care'). Haydon's contribution ('School choice and the common school') exemplifies both conceptual sensitivity and careful construction of an argument – in his account of the concepts of the common school and school choice – in which he asks whether the topic should be set up as a debate between two sides and continues to examine assumptions with regard to what each involves in order to discern whether the two ideas are mutually exclusive. Bingham's article ('Hermeneutics') shows how the hermeneutic approach challenges the assumption that language is a conduit for transferring thought from one mind to another, while Bonnet ('Environmental education') argues that prevailing background assumptions with regard to nature in educational debate and policy rely too heavily on traditional views of science.

Conceptual exploration and arguments in philosophy and philosophy of education frequently explore examples – often from real or imagined educational issues and contexts. Sometimes, they also draw illuminatingly on what are called thought experiments. In this collection, for example, Hogan ('Reason and rationality') refers to perhaps the most well-known thought experiment in modern political philosophy – Rawls' (1971) thought experiment in which imagined participants derive principles of justice from behind a veil of ignorance, in an original position in which they have no knowledge of their own identities, social position, or commitments. Marples ('The aims of education') refers – in questioning whether happiness could be an aim of education – to Nozick's (1974) experience machine, which could provide us, if plugged in, with pleasurable experiences.

Philosophy of Education Since the Mid-Twentieth Century: Analytic Philosophy and Beyond

Twentieth-Century Trends in Philosophy

While this article opposes the assumption that philosophy of education can be usefully described as predominantly either a study of major theoretical positions or of great thinkers in its history, to understand its present state, one does need to note how the discipline – as it has been described so far – has developed historically, especially since around the mid-twentieth century, when a distinctive approach to philosophy developed and was influential in setting the predominant direction of philosophy of education for decades to come. This marked a shift from the tendency to focus on major theorists and

philosophical doctrines to an analytical approach that reflected major trends in mainstream philosophy and turned, instead, to analyzing key concepts as a central role of the discipline.

The development of a distinct approach to philosophy of education that came to prominence in the 1960s and remains highly influential was fostered by the establishment of linguistic philosophy as the dominant approach in English-speaking countries. With its emphasis on ordinary use, its origins can be traced to Socrates (see 'The analytical tradition'). However, its contemporary development was prompted by the work of figures such as Austin (1962) – through his attention to ordinary language – and, most especially, Wittgenstein (1953) – whose emphasis on language games and forms of life fostered a tradition in philosophy of education that remains vibrant, not just within what came to be called analytical philosophy of education. For his work was later to be productively taken up outside the analytical tradition by some philosophers of education inspired as much by post-structuralist influences as by the analytical approach.

In order to appreciate the philosophical influences on post-war philosophy of education – and for reasons that will become clearer later in this article – the reader need also note the powerful attraction of Logical Positivism – as expressed, for example, in Ayer's *Language, Truth and Logic* (Ayer, 1936). Ayer's suspicion of metaphysical statements – which he described as meaningless because they were unverifiable – was to exert a powerful influence, as was his emotivist theory's dismissal of ethics. Distinguishing between fact and value, there being no moral facts, no cognitive significance could be attached to moral utterances, which Ayer dismissed as mere expressions of attitudes. Although mainstream epistemology was to move away from this brand of empiricism, popular perceptions of metaphysical and ethical statements outside of philosophy and even in the wider community of educational enquiry were to remain wedded to the type of positivism expressed by Ayer, and the persisting influence of positivism has outlived the heyday of logical positivism. Ayer's positivism was echoed in an influential textbook by O'Connor, which set out to apply the revolution that had occurred in philosophy.

In *An Introduction to the Philosophy of Education*, (1957) O'Connor's reflected the climate of the time by likening valuations to superstitions; there are no means available for deciding on ethical issues as, unlike matters of empirical fact, they are not amenable to being settled empirically – whether through observation or experiment. We have been far more successful at addressing problems in mathematics and the natural sciences. Based on this model, philosophy does not offer knowledge beyond the modest role of clarifying and weighing up the logic of evaluative claims. This reductionist dismissal of moral values and metaphysical beliefs was resisted by analytical

philosophers of education, for example, by Hirst – who defended the necessary role of both in educational theory's concern with the rational formulation of principles for practice (Hirst, 1966: 38–42).

O'Connor assigned a major role to psychology in theory of education, given its closer affinity to the natural sciences. In the face of such judgments with regard to the relative respectability of disciplines such as psychology and sociology of education, it is not surprising that philosophers of education would have sought to establish their discipline as having its own distinct methods and criteria of rigor.

R. S. Peters and the London School

The publication of the first of many impressions of O'Connor's narrow construal of the role of philosophy of education was followed by a period in which the discipline flourished, in the conditions of post-war economic prosperity that prevailed in many Western countries. Policy development favored systematic teacher education that prevailed until the increasing influence of neoliberalism, in the 1970s, began to foster a climate that was less encouraging of theoretical reflection on education. In the United States, conditions also favored expansion of schooling and a growth of the discipline – both countries seeing the flourishing of professional associations in philosophy of education as well as the publication of work in the discipline that reflected both an upsurge of philosophical research on education and its growing presence in teacher education.

The leading figure in analytical philosophy of education in Britain during this period was R. S. Peters – whose appointment to a chair in philosophy of education at the University of London, in 1962, opened the way for him and others to foster analytical philosophy of education through the analysis of key concepts such as teaching, reason, indoctrination, and education itself as well as through attention to clear and careful argument as a precondition for the formulation of soundly reasoned principles. Working within the Kantian tradition, Peters' work developed a liberal conception of education – premised on autonomy – most notably in his landmark book *Ethics and Education* (Peters, 1966). Other leading contributions of this era included Hirst's *Knowledge and the Curriculum* (Hirst, 1974) and White's *Towards a Compulsory Curriculum* (White, 1973). Working within the analytical tradition, in the United States of America, Scheffler's publications included *Conditions of Knowledge* (Scheffler, 1965) and *Reason and Teaching* (Scheffler, 1973), although Dewey's influence on philosophy of education in the United States of America marked a difference in approaches taken by philosophers of education in the two countries (see 'Pragmatism').

While Peters and Hirst stressed their focus on the analysis of our use of central concepts in ordinary language, articulating their defining characteristics and the

principles governing their use in order to better understand the logic of education (Hirst and Peters, 1970), Peters himself also addressed ethical questions that fell – in O'Connor's terms – outside a strictly analytical framework. In fact – as a body of work – analytical philosophy of education was to become increasingly interested in ethical and political issues. In doing so, it matched similar developments in the major branches of philosophy – most prominently in Rawls' definitive work in political philosophy (Rawls, 1971, 1993), which has had a huge influence on philosophy of education and is discussed by Tjiattas ('Justice and care') and Feinberg ('Liberalism and education') among other articles in this collection that engage with Rawls' work.

During this period, philosophy of education self-consciously fashioned itself as one of the central disciplines of the systematic study of education, and teacher-education curricula came increasingly to reflect the presence of the disciplinary branches of history and psychology – as well as the more recently developed sociology and philosophy of education – denoted as the disciplinary branches of theory of education (Tibbles, 1966). The defense of philosophy of education as an autonomous discipline remained a commitment of the philosophy of education community, even as a less unified conception of its approach emerged following early criticisms of analytical philosophy of education – from both within and without its membership.

A dissenting voice in the years that ensued following what he called "the heyday of analytic philosophy of education" was that of John Wilson – whose publications, since *Thinking with Concepts* (Wilson, 1963), defended a version of philosophy of education that remained focused on conceptual analysis and the study of education as a logical discipline. Wilson marked out conceptual questions as neither questions of fact nor questions of value (Wilson, 1963: 11) but as concerned with "the uses of words and with the criteria or principles by which those words are determined." While Wilson's call for ongoing attention to concepts continued for decades (e.g., Wilson, 2003), it was this emphasis on use that provoked one of the most powerful reservations of critics of philosophy of education conducted in this style. Although Wilson was to remain an enthusiastic defender of this mode of philosophy of education, his own emphasis on conceptual analysis came to be seen as too limiting of the discipline's focus and content.

Beyond Conceptual Analysis

The analytical movement has had a considerable and enduring influence on philosophy of education. Most of the articles in philosophy of education in this encyclopedia show its influence – if only in their close attention to concepts and careful argument – although a number of them are located within different theoretical traditions.

Yet, its role in the discipline has been far from uncontroversial and key aspects of its approach, and the views of education and its aims that the analytical movement has produced, have been subject to strong criticism – both from within mainstream philosophy of education and without.

From its early stages, philosophers of education in Britain, the United States of America, and elsewhere were skeptical with regard to the methodological directions taken by the analytic movement. For example, Edel (1973) raised doubts, which others shared, concerning attempts to exclude ethical and social dimensions on the assumption that they might be brought into consideration after the process of analysis had been completed. Instead, argued Edel: "they are integral to the analysis at the points of choice throughout, together with the empirical, scientific, and historical considerations" (Edel 1973: 252). Edel (1973: 234) claimed that the "Analytic method was given a certain cast by the dogmas it inherited from logical positivism," particularly the sharp separation between philosophical and empirical enquiry.

Such misgivings with regard to the form that some purist approaches to conceptual analysis took are confirmed by the present-day widespread recognition that ordinary language is highly contextualized, and that to refer to our use of particular words and the way in which we distinguish them from others is to invoke the very specific location in which they have developed and are employed. In retrospect, Dearden's (1984: 31) suggestion that "I do not myself think that philosophy of education stands in any need of a single paradigm," while at the same time describing its analytical role in terms of attention to concepts and arguments, looks mistaken – even if one grants the usefulness and contribution of this approach to philosophy of education. For, in writing this, Dearden was invoking a particular paradigm or approach (Burbules, 2000) – one that was soon to come under fire. For this reason, while this article has argued against viewing philosophy of education – in the first instance – in historical terms centered on a series of theoretical positions or significant philosophers, it is essential to see recently dominant approaches, as well as current alternatives – to which we will turn shortly – as part of an evolving history. No single view of what philosophy of education is about should be assumed to be the final word with regard to what this discipline offers to educational enquiry.

Critical responses to analytic philosophy of education from without the discipline focused not only on its account of its method but also on the substantive positions this approach produced with regard to the meaning of education and associated concepts, as well as its aims. As Wringe shows ('Autonomy'; see also 'The analytical tradition' and 'The aims of education'), philosophers of education – from the 1960s onward (e.g., Peters, 1969; Hirst and Peters, 1970; Dearden, 1972) – developed an account of

education in which the virtues of rationality and autonomy were viewed as achievements developed over generations and whose acquisition requires systematic teaching. Marxist criticisms – in the 1970s and 1980s – tended to view the depictions of education in analytical philosophy of education as highly specific expressions of hegemonic class interests. Their emphasis on the social construction of knowledge, the role of schooling in the reproduction of labor, and the role of ideology in social control (e.g., Bowles and Gintis, 1976; Willis, 1977) reflected the influence of Marxist theory and, particularly, the new sociology of education (Young, 1971).

Criticisms of autonomy as the aim of education – associated with liberalism and liberal education (see Feinberg on ‘Liberalism and education’) – were also inspired by communitarians, who objected to what they saw as the liberal tendency to assume an atomistic, unencumbered subject and to ignore its location within a particular community and its traditions (e.g., Sandel, 1982; MacIntyre, 1981). As philosophers of education in English-speaking countries have come to pay more attention to continental philosophy and its implications for education, consideration has, of late, been given to a wider range of conceptualizations of education. Masschelein and Ricken’s article on ‘*Bildung*’ for example, develops a close account of education as a process in which human beings are incorporated into society through self-formation, by appropriation of the world, although they do note affinities between *Bildung* and notions of liberal autonomy. Bingham’s article (‘Hermeneutics’) offers an alternative to the liberal paradigm’s separation – from a Hermeneutic perspective – of the personal from the cultural, treating the two as inseparable, while also attending to tacit understandings, focusing less on knowledge transmission and more on dialogue and interpretation. In ‘Phenomenology and existentialism,’ Levering echoes the focus of Hermeneutics on the relationship between individuals and the world and explores the role of phenomenology and existentialism in the continental European tradition, noting its attention to upbringing and the family. Drawing on the work of Freud and Lacan, Smeyers’ article on ‘Psychoanalysis and education’ demonstrates the relevance of the unconscious and of the notion of Otherness to our understanding of education. Even within the liberal tradition, there are recent developments that mark out reformulations of the aims and purpose of education. Terzi’s article on ‘The capabilities approach’ explores the implications for education of this normative framework for the evaluation of well-being in which capabilities to function are central to the justice of public institutions, including schools.

Thus, by the early twenty-first century, philosophers of education in the analytic tradition found themselves working in a theoretical climate offering a range of alternative perspectives; however, see Wringer’s article on ‘Autonomy,’ among others, for critical evaluations – both of the veracity

of critiques of analytical philosophy of education’s approach and its view of education as well as an indication of how it has adjusted aspects of its position, including its defense of autonomy, in response to such critiques.

Prospects for Philosophy of Education: The Challenges of Diversity

Philosophers of education are now engaged with a wider array of areas of enquiry relevant to education, such as cultural studies, media studies, literature, and development economics. In drawing on a complex mix of ideas, traditions, and approaches, philosophy of education currently takes on various contentious issues and roles in contemporary education, including a growing role in policy critique, having – by now – moved beyond an erstwhile preoccupation with analysis of concepts – useful though this tool remains in the ongoing task of critical vigilance in the face of education’s vulnerability to fads and fashions. What does the present selection of articles in philosophy of education reveal with regard to the current state of the discipline and likely areas calling for future development? This concluding section takes up these questions in relation to the theme of diversity from several angles.

First, in addition to the variety of theoretical approaches on offer in this collection, the theme of diversity is also evident in many of the topics addressed. ‘Feminism’ is a prominent example; the influence of feminist theory has been significant – both methodologically and for the issues it raises, such as those around distinctions between the public and the private (‘Justice and care’). In this context, the recent development of queer theory and its relevance is to be welcomed for the contribution it makes in bringing to attention issues of gender and sexual orientation in education (see, e.g., Pinar, 1998). Merry’s article explores various theoretical perspectives on ‘Identity,’ including some from outside the discipline of philosophy. Conroy and McKinney (‘Religious and spiritual education’) demonstrate the contested place of faith schools and of religious education in the school curriculum in the context of cultural diversity in postindustrial, Western liberal democracies – drawing on different models of religious education, including a phenomenological one. Taking into account the challenges to social cohesion presented by social and cultural diversity, that article is one of several articles that focus on citizenship. Print’s ‘Social and cultural capital in education’ also raises the problem of declining social capital eroding political engagement, as well as the role of civics and citizenship education in the development of social capital. Siegel’s article on ‘Critical thinking’ emphasizes the role of fostering abilities and dispositions of critical thinking in enabling democracy to flourish (on citizenship education, see also Enslin and White, 2003; McDonough and Feinberg, 2003).

A second and ongoing issue in current philosophy of education concerns the relationship between analytical philosophy of education and what is generally called the continental tradition. Is the diversity marked by these two contrasting approaches to be welcomed, or should disagreements regarding fundamental differences in approach be resolved in favor of one of the two? For a time, philosophers of education in English-speaking countries who took an interest in European philosophy found themselves marginalized by those committed to the analytic approach (Standish, 2007). Hopefully, the selection of articles collected here reflect the current diversity across traditions as a strength, although many resist easy classification and – in spite of this, or perhaps because of it – are able to throw useful light on contemporary issues in education. To some extent, the type of approach best taken to a particular topic in philosophy of education may depend on the questions being asked.

Ironically, those who might take the view that, here, there is a standoff in which one or the other approach to philosophy of education could emerge as the winner, reflect a similar assumption with regard to the nature of philosophy of education to those proposing to frame it by schools of thought or great thinkers. A polarizing tendency to require that, ultimately, reflection on competing orientations should be resolved by deciding to adopt a particular one as one's own, to the exclusion of others, implicitly assumes the ahistorical tendencies of the schools of thought or great thinkers' views of philosophy of education.

Third, what of philosophy of education's presence among other approaches to theory – and especially to research – in education (see 'Philosophy and educational research')? If a strength of contemporary philosophy of education is its diverse openness to other disciplines and to different theoretical orientations within the ranks of philosophers of education, is this view of diversity reciprocated by those working in other domains of educational inquiry? While efforts to establish philosophy of education as a discipline with a systematic and rigorous method may initially have enabled it to compete successfully with other elements of theory of education in teacher-education curricula, it remains vulnerable to the lingering effects of the assumptions that underpinned logical positivism in assessments outside the discipline of what it might offer to educational inquiry as a whole. Bridges' article on 'Philosophy and educational research' considers philosophy as a form of educational research in its own right, proposing how philosophical inquiry has an essential role to play in informing research methodology. This article also draws critical attention to a tendency – in the present-day interest in evidence-based practice – to assume that educational research can best inform policy by providing evidence that excludes philosophical and, especially, normative considerations. In raising these concerns, Bridges reminds

us that the status of philosophy of education is a political question.

Fourth, and finally, we come to the global context in which philosophers of education now work and the challenges of diversity this poses. Growing recent interest in the diversity within Western democracies takes into account the consequences of globalization – as migration has changed relatively homogeneous societies, prompting the attention to diversity and its consequences in philosophy of education, which has already been discussed in this article. Recognition of globalization's consequences also prompts philosophy of education to take postcolonial theories (e.g., Said, 1978; Spivak, 1988) into account – centrally, in their attention to the relationship between Western cultures and economies and peoples in developing countries attempting to deal with the aftermath of colonialism and the hegemony of Western capitalism. While future work in philosophy of education will pay continuing attention to this theme, two articles in the present collection represent a challenge to Western philosophers of education and take us back to the critical position taken toward philosophy of education conceived as world views attached to particular cultural groups. Horsthemke's article on 'African philosophy of education' considers the position that there is a distinctly different African philosophy of education, underpinned by its own concepts and values. Although Horsthemke's response is a skeptical one, an intriguing contrast emerges between this article and the stance taken in Waghid's 'Towards a philosophy of Islamic education' – in which the author's exploration of constitutive features of an Islamic philosophy of education includes drawing some similarities between an Islamic approach to education based on justice and recent theories of deliberative democracy. This implies that there is room for further development and debate in this area, not least with regard to tensions between universality and particularity, and between philosophy of education as worldview and as a critical activity.

In engaging with the many implications of globalization, perhaps the most challenging for Western philosophers of education – if diversity is taken truly seriously – is the role their discipline could play in countering Western cultural hegemony and the injustices of the global economic system. An obvious dimension of the problem of hegemony lies with the dominance of the West in the production of discourses in the philosophy of education – in itself a consequence of the wealth and influence of higher education and associated publishing industries in those countries. The English language dominates international exchanges. Tackling such hegemony suggests the need for philosophers of education to engage further with philosophical issues concerning both cultural recognition and inequitable distribution of resources – in addition to the predominant focus on the largely domestic or Western problems exemplified in the philosophical work published in this collection.

See also: Adult Learning: Philosophical Issues; African Philosophy of Education; Autonomy; Bildung; Critical Theory and Pedagogy; Critical Thinking; Environmental Education; Feminism; Hermeneutics; Identity; Inclusion; Justice and Care; Liberalism and Education; Lifelong Learning; Parents and Children's Rights; Peace Education; Phenomenology and Existentialism; Philosophy and Educational Research; Pragmatism; Psychoanalysis and Education; Race, Critical Race Theory and Whiteness; Reason and Rationality; School Choice and the Common School; Social and Cultural Capital in Education; The Aims of Education; The Analytical Tradition; The Capabilities Approach.

Bibliography

- Austin, J. L. (1962). *How to Do Things with Words: The William James Lectures Delivered at Harvard University in 1955*. Oxford: Clarendon Press.
- Ayer, A. J. (1936). *Language, Truth and Logic*. London: Victor Gollancz.
- Bowles, S. and Gintis, H. (1976). *Schooling in Capitalist America*. London: Routledge and Kegan Paul.
- Brubacher, J. S. (1962). *Modern Philosophies of Education*. New York: McGraw-Hill Book.
- Burbules, N. C. (2000). Philosophy of education. In Moon, B., Ben-Peretz, M., and Brown, S. (eds.) *Routledge International Companion to Education*, pp 3–18. New York: Routledge.
- Dearden, R. F. (1972). Autonomy and education. In Dearden, R. F., Hirst, P. H., and Peters, R. S. (eds.) *Education and the Development of Reason*, pp 448–465. London: Routledge and Kegan Paul.
- Dearden, R. F. (1982). Philosophy of education. *British Journal of Educational Studies* 30(1), 57–71.
- Edel, A. (1973). Analytic philosophy of education at the crossroads. In Doyle, J. F. (ed.) *Educational Judgments: Papers in the Philosophy of Education*, pp 232–257. (1972). London: Routledge and Kegan Paul.
- Enslin, P. and White, P. (2003). Democratic citizenship. In Blake, N., Smeyers, P., Smith, R., and Standish, P. (eds.) *The Blackwell Guide to the Philosophy of Education*, pp 110–125. Oxford: Blackwell.
- Hirst, P. H. (1966). Educational theory. In Tibble, J. W. (ed.) *The Study of Education*, pp 29–58. London: Routledge and Kegan Paul.
- Hirst, P. H. (1974). *Knowledge and the Curriculum: A Collection of Philosophical Papers*. London: Routledge and Kegan Paul.
- Hirst, P. H. and Peters, R. S. (1970). *The Logic of Education*. London: Routledge and Kegan Paul.
- Lyotard, J. -F. (1984). *The Postmodern Condition: A Report on Knowledge*. Minneapolis, MN: University of Minnesota Press.
- MacIntyre, A. (1981). *After Virtue*. Notre Dame: University of Notre Dame Press.
- McDonough, K. and Feinberg, W. (2003). *Citizenship and Education in Liberal-Democratic Societies – Teaching for Cosmopolitan Values and Collective Identities*. Oxford: Oxford University Press.
- Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.
- O'Connor, D. J. (1957). *An Introduction to the Philosophy of Education*. London: Routledge and Kegan Paul.
- Peters, R. S. (1966). *Ethics and Education*. London: Allen and Unwin.
- Peters, R. S. (1969). *Perspectives on Plowden*. London: Routledge and Kegan Paul.
- Pinar, W. F. (ed.) (1998). *Queer Theory in Education*. Mahwah, NJ: Erlbaum.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: The Belknap Press of Harvard University Press.
- Rawls, J. (1993). *Political Liberalism*. New York: Columbia University Press.
- Rikowski, G. (2007). Marxist educational theory unplugged. *Paper for Fourth Historical Materialism Annual Conference*, 9–11 November 2007, School of Oriental and African Studies, University of London.
- Said, E. (1978). *Orientalism*. London: Routledge and Kegan Paul.

- Sandel, M. (1982). *Liberalism and the Limits of Justice*. New York: Cambridge University Press.
- Scheffler, I. (1965). *Conditions of Knowledge: An Introduction to Epistemology and Education*. Chicago, IL: Scott, Foresman.
- Scheffler, I. (1973). *Reason and Teaching*. London: Routledge and Kegan Paul.
- Spivak, G. C. (1988). Can the subaltern speak? In Nelson, C. and Grossberg, L. (eds.) *Marxism and the Interpretation of Culture*, pp 271–313. Urbana, IL: University of Illinois Press.
- Standish, P. (2007). Rival conceptions of the philosophy of education. *Ethics and Education* 2(2), 159–171.
- Tibble, J. W. (ed.) (1966). *The Study of Education*. London: Routledge and Kegan Paul.
- White, J. (1973). *Towards a Compulsory Curriculum*. London: Routledge and Kegan Paul.
- Willis, P. (1977). *Learning to Labour: How Working Class Kids Get Working Class Jobs*. Farnborough: Saxon House.
- Wilson, J. (1963). *Thinking with Concepts*. Cambridge: Cambridge University Press.
- Wilson, J. (2003). Perspectives on the philosophy of education. *Oxford Review of Education* 29(2), 279–293.
- Wittgenstein, L. (1953). *Philosophical Investigations*, Anscombe, G. E. M. (trans.). Oxford: Basil Blackwell.
- Young, M. (1971). *Knowledge and Control*. London: Collier Macmillan.

Further Reading

- Biesta, G. J. J. (2002). *Studies in Philosophy and Education*. <http://www.springer.com/education/educational+philosophy/journal/11217> (accessed September 2009).
- Blake, N., Smeyers, P., Smith, R., and Standish, P. (eds.) (2003). *The Blackwell Guide to the Philosophy of Education*. Oxford: Blackwell.
- Burbules, N. C. *Educational Theory*. published on behalf of the College of Education at the University of Illinois at Urbana-Champaign and co-sponsored by The John Dewey Society, the Philosophy of Education Society, and the College of Education at the University of Illinois at Chicago. <http://www.wiley.com/bw/journal.asp?ref=0013-2004> (accessed September 2009).
- Carr, D. (2003). *Making Sense of Education: An Introduction to the Philosophy and Theory of Education and Teaching*. London: Routledge Falmer.
- Carr, W. (ed.) (2005). *The RoutledgeFalmer Reader in Philosophy of Education*. London: Routledge.
- Curren, R. (2003). *A Companion to the Philosophy of Education*. Oxford: Blackwell.
- Curren, R. (2007). *Philosophy of Education: An Anthology*. Oxford: Blackwell.
- Hirst, P. H. and White, P. (eds.) (1998). *Philosophy of Education: Major Themes in the Analytic Tradition*, 4 vols. London: Routledge.
- Peters, M. A. (2009). *Educational Philosophy and Theory*, published on behalf of the Philosophy of Education Society of Australasia. <http://www.wiley.com/bw/journal.asp?ref=0013-1857> (accessed September 2009).
- Peters, M. A., Ghiraldelli, P. Jr., Standish, P., and Zarnic, B. *Encyclopaedia of Philosophy of Education*. <http://www.vusst.hr/ENCYCLOPAEDIA/main.htm> (accessed September 2009).

Relevant Websites

- <http://www.siuc.edu> – Center for Dewey Studies.
- <http://www.philosophyofeducation.ca> – Canadian Philosophy of Education Society (CPES).
- <http://www.epistemelinks.com> – Episteme Links, Philosophy Resources on the Internet.
- <http://humanum.arts.cuhk.edu.hk> – Humanities Meta-Indices: Philosophy.
- <http://www.ucm.es/info/inpe/> – INPE-International Network of Philosophers of Education, Universidad Complutense de Madrid.

<http://www.philinfo.org> – Philosopher's Information Center, Publisher of The Philosopher's Index.
<http://philosophyofeducation.org> – Philosophy of Education Society.
<http://www.pesa.org.au> – Philosophy of Education Society of Australasia.
<http://www.philosophy-of-education.org> – Philosophy of Education Society of Great Britain (PESGB).

<http://www.iep.utm.edu> – The Internet Encyclopedia of Philosophy.
<http://www.dmoz.org> – The Open Directory Project (dmoz), Philosophy of Education.
<http://tre.sagepub.com> – Theory and Research in Education.
<http://plato.stanford.edu> – The Stanford Encyclopedia of Philosophy.

PHILOSOPHY OF EDUCATION – CONTEMPORARY ISSUES

Contents

Neoliberalism, the Market and Performativity

Parents and Children's Rights

Peace Education

Philosophy and Educational Research

Race, Critical Race Theory and Whiteness

School Choice and the Common School

Neoliberalism, the Market and Performativity

M A Peters, University of Illinois at Urbana–Champaign, Champaign, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Homo economicus (Economic human) –

A concept in neoclassical economics of humans as rational and self-interested individuals who have the ability to make judgments towards their subjectively defined ends.

Methodological individualism – A philosophical method aimed at explaining and understanding broad society-wide developments as the aggregation of decisions by individuals.

Mt Pelerin Society – An influential post-war foundation established by Hayek in 1947 against socialism and the 'big' state and in defence of liberalism as the open society.

Neoliberalism – A doctrine of economic liberalization in global capitalism associated with the rise of the New Right and the so-called Washington Consensus, aimed at structural adjustment, privatization and 'free trade.'

New Right – A right-wing in UK, Australia and NZ who advocated economic liberal and socially conservative policies; strongly associated with Margaret Thatcher.

Performatives – A concept that is related to speech act theory and to the work of J. L. Austin. It accounts for situations where a proposition may constitute or instantiate the object to which it is meant to refer, as in so-called 'performative utterances'.

Performativity – J.-F. Lyotard talks of 'performativity' as that which legitimates knowledge in the

postmodern condition – a logic based on the efficiency of how well something performs, or enables a person to perform. In a cybernetic society, knowledge is legitimated by how performative it is, if it effectively minimizes the various required inputs for the task and maximizes the desired outputs. This conception leaves no room for ethics.

Postmodern condition – Lyotard defines postmodern as incredulity towards metanarratives, a suspicion towards the grand ideological stories of the enlightenment.

Spontaneous order – A doctrine associated with Hayek that describes the spontaneous emergence of order out of seeming chaos and the emergence of various kinds of social order from a combination of self-interested individuals who are not intentionally trying to create order.

Washington Consensus – Refers to a set of policies advocated the privatization of state assets, economic deregulation, and redirection of public expenditure priorities toward fields offering high economic returns; trade liberalization, the liberalization of interest rates and rules for foreign direct investment, and the floating of the exchange rate and tax reform (to lower marginal rates and broaden the tax base) together with enforcement of property rights and the encouragement of greater fiscal discipline.

Performance will be to the twentieth and twenty-first centuries what discipline was to the eighteenth and nineteenth centuries, namely, an onto-historical formation of power and knowledge. (McKenzie, 2001b: 176)

Introduction: The Birth of Neoliberalism

Neoliberalism is a popular label for the doctrine of political and economic liberalism and set of policies, originating in the 1970s, that wielded together classical liberal political theory as exemplified by the Mont Pelerin Society following World War II (WWII) and neoclassical economic theories that became identified with the so-called Chicago school under Milton Friedman, in the 1960s. It is not a unified and coherent doctrine and it has taken on different manifestations at different times and places – sometimes with contradictory results. For an ultimately moral doctrine based on a classical account of political and economic freedom – a marriage of the free market and the open society – paradoxically, beginning with Chile in 1973, administrations and policy regimes based on the minimal state and open global market were brutally established with force and coercion, against the rule of law, and in an antidemocratic way. This imposition became commonplace, during the 1980s, with structural adjustment policies of the International Monetary Fund (IMF) and the World Bank (WB) that forced the transitional economies of Latin America and elsewhere to liberalize trade and monetary systems, to open up their economies, and to privatize state assets and cut-back state welfare.

For analytical purposes, we can postulate several historical stages of neoliberalism:

1. The development of the Austrian, Freiburg, and Chicago schools in neoclassical economics in the first part of the twentieth century.
2. The first globalization of neoliberalism with the establishment of the Mount Pelerin Society, in 1947, through the intellectual conduit of Hayek.
3. The development of the Washington consensus during the 1970s.
4. The Central Intelligence Agency (CIA)-sponsored coup of Salvador Allende in Chile and the imposition of neoliberal market reforms by General Pinochet.
5. The New Right ascendancy of the Thatcher–Reagan years during the 1980s.
6. The emergence of structural adjustment loans and institutionalization of neoliberalism through a series of world policy agencies such as IMF, WB, Organization for Economic Cooperation and Development (OECD), and World Trade Organization (WTO).
7. The transition to knowledge economy and knowledge for development in the 1990s and beyond.

This article briefly considers the role that Friedrich von Hayek played in contemporary neoliberalism prior to detailing the rise of the New Right and the emphasis on accountability and performativity in education.

The Centrality of Friedrich von Hayek

Friedrich von Hayek (1899–1992) was a major circuit for liberal economic and political ideas, as well as an organizer and a social and political thinker following WWII. He began his studies under Eugen von Bohm-Bawerk and von Mises, was invited to the London School of Economics (LSE) by Lionel Robbins, in 1931, and became a British subject in 1938. He took up a professorship at the University of Chicago in 1950, which he held until 1962. He won the Nobel Prize for Economics in 1974 and taught and eventually retired at the University of Freiburg in his later years.

Friedrich Von Hayek is probably the single most influential individual economist or political philosopher to shape what is now understood as neoliberalism, although he is best regarded – and considered himself – as a classical liberal. Hayek's own theoretical direction sprang out of the so-called Austrian School established by Carl Menger, Eugen Boehm-Bawerk, and Ludwig von Mises during first decade of the early twentieth century. What distinguished the Austrian School from the classical school of political economy pioneered by Adam Smith and David Ricardo was their subjective – as opposed to the objective – theory of value. Leon Walras (1834–1910) of the French Lausanne School presented economics as “the calculus of pleasure and pain of the rational individual” and Carl Menger – developing the subjective theory of value – launched what some have called a “neoclassical revolution” in economics.

In 1930, Hayek was invited to the LSE to lecture on trade cycles, where he was, soon after, appointed to a chair in economics and statistics. While at the LSE, Hayek was involved in two famous debates: first, with Keynes over interventionism (and, in particular, Keynes' alleged failure to understand the role that interest rates and capital play in a market economy), and, second – during the early 1920s – with Oskar Lange and others over the nature of socialist planned economy. However, Keynes' star was on the rise during the 1930s and Hayek's criticisms were downplayed by the international economics community. In 1950, Hayek moved to the University of Chicago, where he wrote *The Constitution of Liberty* (Hayek, 1960) – his first systemic treatise on classical liberal political economy. In 1962, Hayek moved to the University of Freiburg where he developed his theory of spontaneous order. The market, he argued, was a spontaneously ordered institution that had culturally evolved in the same way that the institutions of language and morality had evolved. They

were not the product of intelligent design; such social institutions – like their counterparts in the physical world (crystals, snowflakes, and galaxies) – had evolved as spontaneously ordered institutions. The market, then, while the result of many millions of human (trans)actions over many generations – Hayek maintained – is not the result of human design.

Hayek addressed himself, again, to the problems of the nature of the planned socialist economy in one of his most famous and populist works, *The Road to Serfdom* (Hayek, 1944) – a book that suggested that the absence of a pricing system would prevent producers from knowing true production possibilities and costs. It also warned about the political dangers of socialism, in particular, totalitarianism – which he thought came directly from the planned nature of institutions. Following WWII, in the year 1947, Hayek set up the very influential Mont Pelerin Society – an international organization dedicated to restoring classical liberalism and the so-called free society, including its main institution – the free market. Hayek was concerned that even though the Allied powers had defeated the Nazis, liberal government was too welfare-oriented – a situation, he argued, that fettered the free market, consumed wealth, and infringed the rights of individuals. With the Mont Pelerin Society Hayek there gathered around him a number of thinkers committed to the free market, including his old colleague Ludwig von Mises as well as some younger American scholars who were to become prominent economists in their own right. These included Rose and Milton Friedman, James Buchanan, Gordon Tullock, and Gary Becker – economists who went on to establish the main strands of American neoliberalism: the monetarism of the third Chicago School (see e.g., Friedman, 1962), public choice theory (see e.g., Tullock and Buchanan, 1962), and human capital theory (e.g., Becker, 1964). Most significantly, Hayek invited his old friend Karl Popper – who was resident at the University of Canterbury in New Zealand from 1937 to 1945 where he wrote his now classic two-volume work *The Open Society and Its Enemies* (Popper, 1945), which was an attack on historicism and a defense of liberal democracy as the open society. Hayek had arranged for Popper to take up a position in philosophy at the LSE and their association had gone back many years – Hayek refers to Popper's *The Logic of Scientific Discovery* and adopts his falsificationism in his early-1937 *Economica* paper 'Economics and knowledge' (Hayek, 1937). Both were in Vienna at the turn of the century, both were committed to forms of individualism – epistemological and methodological – and both fiercely defended the notion of the free society. Hayek argued that political freedom depended on economic freedom. Together, Hayek and Popper were formidable Cold War warriors who held that Hegel's and Marx's historicism were the basis of twentieth-century totalitarianism. Together, Popper's *The Open Society*, and Hayek's *The Road to Serfdom* (1949) served as polemical tracts that, at one and the same time, warned against going

down the socialist road while extolling the virtues of the open society and its relationship with the open market.

Hayek's liberalism was also very influential in Britain – especially with the Institute of Economic Affairs – and with Margaret Thatcher, who came to power as the leader of the British Conservative Party, in 1979. We might say that neoliberalism, historically, was at its strongest during the era of the trans-Atlantic partnership between Ronald Reagan and Margaret Thatcher – during the decade of the 1980s – and its dominance began to wane in the 1990s.

The Rise of the New Right

Hayek, then, emphasized the limited nature of knowledge: the price mechanism of the 'free' market conveys information about supply and demand that is dispersed among many consumers and producers and cannot be coordinated. In addition, Hayek's liberalism emphasized: methodological individualism; *homo economicus* – based on assumptions of individuality, rationality, self-interest; and the doctrine of spontaneous order. It was during the decade of the 1980s that Hayek's political and economic philosophy was adopted by Thatcher and Reagan to legitimate the New Right attack on big government, State ownership of enterprises, and the bureaucratic welfare state, including the State unions, with a policy mix based on free trade and the establishment of the open economy: economic liberalization or rationalization characterized by the abolition of subsidies and tariffs, floating the exchange rate, the freeing up of controls on foreign investment; the restructuring of the state sector, including corporatization and privatization of state trading departments and other assets, downsizing, contracting out, the attack on unions, and abolition of wage bargaining in favor of employment contracts; and, finally, the dismantling of the welfare state through commercialization, contracting out, targeting of services, and individual responsabilization for health, welfare, and education. On this view, there is nothing distinctive or special about education or health; they are services and products like any other – to be traded in the marketplace.

These policies were designed to restructure or adjust national economies to the dramatic changes to the world economy that have occurred in the last 20 years: the growing competition among nations for world markets; the emergence of world trading blocs and new free trade agreements; an increasing globalization of economic and cultural activities; the decline of the postwar Keynesian welfare-state settlement in Western countries; the collapse of actually existing communism and the 'opening up' of the Eastern bloc; and the accelerated worldwide adoption and development of the new information and communications technologies.

Neoliberalism received its definitive statement by Williamson (2004) who coined the term Washington

Consensus, in 1990, to refer to a set of policies addressed to Latin American countries by Washington-based institutions. These policies essentially involved an attack on big government – a shift in control and resources from the public to the private sector and the substitution of the market or market-like arrangements for the state at every opportunity. The Washington consensus advocated the privatization of state assets, economic deregulation, and redirection of public expenditure priorities toward fields offering high economic returns. It strongly encouraged trade liberalization, the liberalization of interest rates and rules for foreign direct investment, and the floating of the exchange rate. Finally, it proposed tax reform (to lower marginal rates and broaden the tax base) together with enforcement of property rights and the encouragement of greater fiscal discipline. Williamson himself reflecting on these policies a decade later suggests the three big ideas are “macroeconomic discipline, a market economy, and openness to the world (at least in respect of trade and foreign direct investment (FDI)).”

By the 1980s, neoliberalism – both as a political philosophy and policy mix – had taken deep root and was applied both forcefully and manipulatively. For instance in New Zealand, a New Right Labor government instituted sweeping market reforms across the board against the interests of its traditional affiliated trade union party and against sustained democratic protests. During that decade, many governments around the world supported the modernizing reform thrust of neoliberalism – particularly the exposure of the State sector to competition and the opportunity to pay off large and accumulating national debts. By contrast, many developing countries had structural adjustments policies imposed upon them as loans conditions from the IMF and WB. The reforming zeal soon ideologized the public sector *per se* and ended by damaging key national services (including health, education, and telecommunications).

The dogmatism of the neoliberal right had become a serious threat to social justice, to national cohesion, and to democracy itself. Large sections of populations had become structurally disadvantaged, working and living, on the margins of the labor market; rapidly growing social inequalities had become more evident as the rich had become richer and the poor, poorer; companies were failing and underperforming; public services had been stripped down and privatized, some were unable to deliver even the most basic of services; many communities had become split and endangered by the rise of racism, crime, unemployment, and social exclusion.

Neoliberalism, Performativity, and the Culture of Performance

It is one of the uninvestigated central ironies of the new critical science driven by French theory that one of its

major sources of theoretical inspiration is British analytic philosophy and especially the linguistic philosophy of J. L. Austin and Ludwig Wittgenstein. Both Lyotard (1984) and Foucault (1978), for example, use and adapt Wittgenstein’s conception of language games to describe the fragmentary nature of the social bond, the crisis of legitimation of scientific knowledge in the postmodern condition, and the strategic and political nature of games of truth. Derrida and Lyotard – although for different purposes – draw explicitly on Austin’s (1954) notion of the performative first formulated in *How to Do Things with Words*. Derrida (2005) embraces Austin’s category and theory of the performative as the basis for the new humanities and Butler (1990, 1993) extends this use in gender studies to indicate that gender is constituted by performative acts which come to form a coherent gender identity. Lyotard (1984) appropriates Wittgenstein’s philosophy of language games as a basis for his analysis of the problem of the legitimation of knowledge and education in the postmodern condition and draws on Austin to formulate and predict a new culture of performativity for higher education.

Lyotard’s prophetic prognosis on the nature of knowledge and its management in higher education has become a theoretical springboard for educational theorists to analyze performativity as a more general condition affecting not only the whole of the education sector but also, more widely, newly consumer-driven public services across the board. At the same time, management theorists and business schools picking up on similar socioeconomic structural changes in capitalism have developed notions of performance – performance management and performance culture – in a parallel, though generally theoretically unreflexive, discourse.

Jean-François Lyotard is considered by most commentators, justly or not, as the non-Marxist philosopher of the postmodern condition (sometimes referred to as postmodernity). His *The Postmodern Condition: A Report on Knowledge* (1984), originally published in Paris in 1979, became an instant *cause célèbre*. The book crystallized – in an original interpretation – a study of the status and development of knowledge, science, and technology in advanced capitalist societies. *The Postmodern Condition* was important for a number of reasons. It developed a philosophical interpretation of the changing state of knowledge, science, and education in the most highly developed societies – reviewing and synthesizing research on contemporary science within the broader context of the sociology of postindustrial society and studies of postmodern culture. Lyotard brought together, for the first time, diverse threads and previously separate literatures in an analysis which many commentators and critics believed to signal a epoch-making break not only with the so-called modern era but also with various traditionally modern ways of viewing the world.

One way in which his analysis has been prophetic was that the postmodern condition marks the decline of

modernity's grand narratives, and, in particular, the way in which the progress of Western capitalist technoscience has been culturally embedded in deeper metaphysical fables about reason and freedom, rather than globalization, flexibility, social regulation, and performativity. The story of Western capitalist techno-science was supposed to lead to the gradual, but inevitable, unfolding of the progress of knowledge in emancipation. Lyotard, however, hypothesizes that this story became simply too incredible to believe when, in the 1960s, it was clear that the transition to the postindustrial society (or knowledge economy) was no longer a smooth, necessary, inevitable, or desirable movement. He maintained that the postmodern condition was based on the perception that "the existence of a modern era that dates from the time of the Enlightenment and that now has run its course: and this modern era was predicated on the notion of progress in knowledge, in the arts, in technology, and in human freedom as well, all of which were thought of as leading to a truly emancipated society: a society emancipated from poverty, despotism and ignorance. But all of us can see that the development continues to take place without leading to the realization of any of these dreams of emancipation" (Lyotard, 1984). In this transitional period, Lyotard argues, science and technology are falling under the sway of "another language game, in which the goal is no longer truth, but performativity – that is, the best possible input/output equation" (Lyotard, 1984). It is clear that Lyotard is referring to performance and a logic of performativity that is a cornerstone of the new techno-scientific capitalist knowledge economy.

The 'Age of Performativity'

McKenzie (2001a) has offered an explanation of performance considered as a new social attitude in his recently published book *Perform or Else: From Discipline to Performance*. He argues that the concept of performance is now widely recognized in business and the commercial world as a conceptual tool for assessing human and technological standards and that it is fast becoming the dominant social model of evaluation. According to McKenzie (2001a), traditional philosophical distinctions are becoming less influential and performance effectiveness and efficiency are growing in power as the new conventions defining the basis for the measurement of what is right, true, and good.

McKenzie (2001a) theorizes performance as a formation of power and knowledge, and in doing so, warns us of the growing objectification and alienation of human labor. He asks whether we have entered an age of global performance and remarks that performance has emerged as a crucial term in at least three different areas of social life: economics, technology, and art. For McKenzie (2001a), performance management draws upon the paradigm of cultural performance and substitutes for scientific

management or 'Taylorism', which was the dominant organizational paradigm developed in the early part of the last century for a manufacturing-based, nationally oriented, and highly industrialized economy. Performance management, in contrast to scientific management, no longer produces highly centralized bureaucracies, rigid, top-down management styles, or perceived controlling, hierarchical, and conformist organizational cultures.

Performance management does not sell itself as scientific but rather – adopting the paradigm of cultural performance – it re-describes itself as an *ars poetica* of organizational practice, which is evident in texts like *Corporate Renaissance: The Art of Re-engineering* (Cross *et al.*, 1994), *Jamming: The Art and Discipline of Business Creativity* (Kao, 1998), and *Cultural Diversity in Organizations* (Cox, 1993). This new soft power of management theory and practice recognizes performance as having acquired a normative force. McKenzie's (2001a) analysis of it as a formation of power and knowledge enables us to appreciate a theorizing of performative power that extends beyond the realm of cultural production into discourses and practices that have the normative force to structure our organization, and the institutions of work, learning, and leisure. In addition, while there are different semantic ranges involved, together with different sites of pragmatic installation, as McKenzie argues, the soft power of performance also enables us to recognize the integration of cultural, management, and technological systems.

Higher education is a crucial subsector where these types of performative power intersect, especially when framed by the policy template of the knowledge economy – for, in the knowledge economy, the cultural and the symbolic are paramount. This is the very idea behind the so-called sign economy no longer based on raw materials but rather on the transformation of ideas and symbolic resources by means of intellectual, human, and social capital. In this environment, the three spheres of the economic, technical, and cultural are increasingly brought into a close alignment as performative power combines the rational calculation of (high performance) technical systems and databases with the domain of affective management based around personal experience and social interaction. Performance management, in this context, first came to light with the development of performance-measurement systems developed by the performance-indicator movement and later under New Public Management that drew on principal-agency theory and transactional cost analysis. Performance management is an ideal system for knowledge management, especially where one of the main aims for the knowledge manager is to extract knowledge from people's heads (often tacit knowledge that is difficult to codify) and to embed it in intellectual systems or processes as soon as possible – both protecting it as intellectual property under copyright, patent, or international trade law and putting it into commercial service to make a profit. It is classically speaking concerned with the appropriation of the knowledge surplus

and often performance management utilizes the soft psychotherapeutic technologies in the affective domain, alongside traditional peer-review mechanisms and collegial exchange and in combination with simple counts and computer and/or accounting methodologies (including the weighting and the arithmetization of soft variables like ‘reputation’) to produce departmental, faculty, and institutional performance profiles and institutional, national, and international league tables.

See also: Philosophy of Education: Overview.

Bibliography

- Austin, J. L. (1954). *How to Do Things with Words*, 2nd edn. (1975). Cambridge: Harvard University Press.
- Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: National Bureau of Economic Research; distributed by Columbia University Press.
- Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.
- Butler, J. (1993). *Bodies That Matter: On the Discursive Limits of “Sex.”* New York: Routledge.
- Cox, T. (1993). *Cultural Diversity in Organizations*. San Francisco, CA: Berrett-Koehler.
- Cross, K. F., Feather, J. J., and Lynch, R. L. (1994). *Corporate Renaissance: The Art of Reengineering*. Oxford: Blackwells.
- Derrida, J. (2005). The future of the profession or the unconditional university (thanks to the humanities what could take place tomorrow). In Trifonas, P. and Peters, M. A. (eds.) *Deconstructing Derrida: Tasks for the New Humanities*, pp 11–24. New York: Palgrave Macmillan.
- Foucault, M. (1978). La Philosophie analytique de la politique. In Defert, D. and Ewald, F. (eds.) *Dit et écrits 1954–1988*, vol. 3, pp 540–541. Paris: Gallimard.
- Friedman, M. (1962). *Capitalism and Freedom*, with the assistance of Rose D. Friedman. Chicago, IL: University of Chicago Press.
- Hayek, F. (1937). Economics and Knowledge. *Economica* IV. <http://www.hayekcenter.org/friedrichhayek/hayek.html> (accessed September 2009).
- Hayek, F. A. (1944). *The Road to Serfdom*. London: Routledge and Kegan Paul.
- Hayek, F. A. (1960). *The Constitution of Liberty*. London: Routledge and Kegan Paul.
- Kao, J. (1998). *Jamming: The Art and Discipline of Business Creativity*. New York: HarperBusiness.
- Lyotard, J. -F. (1984). *The Postmodern Condition: A Report on Knowledge*. Minneapolis: University of Minnesota Press.
- McKenzie, J. (2001a). *Perform or Else: From Discipline to Performance*. London: Routledge.
- McKenzie, J. (2001b). Performance and Global Transference. *Drama Review* 45(3), 5–7.
- Popper, K. (1945). *The Open Society and Its Enemies*, 2 vols. London: Routledge.
- Tullock, G. and Buchanan, J. M. (1962). *The Calculus of Consent: Logical Foundation of Constitutional Democracy*. Ann Arbor MI: University of Michigan.
- Williamson, J. (2004). *A Short History of the Washington Consensus, Paper Commissioned by the Fundacion CIDOB for the Conference. From the Washington Consensus toward a New Global Governance*. Barcelona, 24–25 September. <http://www.iie.com/publications/papers/williamson0904-2.pdf> (accessed September 2009).

Further Reading

- Hayek, F. (1936). Economics and Knowledge. <http://www.virtualschool.edu/mon/Economics/HayekEconomicsAndKnowledge.html> (accessed September 2009).
- Hayek, F. (1945). The use of knowledge in society. *American Economic Review* XXXV(4), 519–530.
- McKenzie, J. (2004). High performance schooling. *Parallax* 10(2), 50–62. <http://www.english.wisc.edu/mckenzie/docs/highperfschooling.pdf>.

Parents and Children's Rights

M Clayton, University of Warwick, Coventry, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Best-custodian condition – The claim that custodial rights over a child should be allocated to those interested individuals whose parenting can reasonably be expected to be in the child's best interests.

Claim right – A claim right to ϕ is an entitlement to ϕ protected by an injunction that places others under a duty to allow, or enable, one to ϕ .

Distributive justice – The study of the conditions under which everyone receives her due or in which the distribution of opportunities and resources is fair.

Dual-interest view of custody – The claim that custodial rights over a child should be allocated in a way that is sensitive to the interest of adults who hold ambitions to parent a well as the interests of children.

Equality of opportunity – The ideal that everyone enjoys equal life chances.

Liberty right – A liberty right to ϕ is a permissions to ϕ or not to ϕ as one chooses.

Parental neutrality between conceptions of the good – The claim that parental conduct should not be guided by a belief in the intrinsic merits of particular religious or ethical conceptions.

Personal autonomy – The ideal that a person's life should be guided by her own reflective view of what a successful life consists in.

Political legitimacy – The conditions under which citizens have a weighty reason to comply with the law and support their political institutions.

Political morality – The study of how we should live together in society.

Sense of justice – The motivation to comply with a conception of the requirements of justice, or what is due to individuals.

Discussions of the rights of children and parents have centered on whether children and parents have rights, how such rights are justified, the nature and content of these rights, and how to resolve conflicts between the rights of parents and children when they conflict. These discussions proceed on the basis of the Hohfeldian classification of rights into liberties (or privileges), claim-rights, powers, and immunities (Hohfeld, 1919; Sumner, 1987). One has a liberty right to perform an action, to speak for example, if one has the

option of speaking or not as one chooses. Claim-rights are entitlements that generate duties on others: the claim-right to speak involves others being duty-bound not to prevent one from speaking. Powers are permissions to change one's own or, perhaps, someone else's normative status: for example, many claim that parents ought to enjoy the power to accept or refuse a painful operation on behalf of their child. Finally, one enjoys immunity if others are duty-bound not to change one's normative status: for example, children enjoy an immunity against anyone selling them into bonded labor.

One can enjoy legal rights, which are, by definition, enshrined in a legal system or document, such as a legally recognized declaration or convention. Academic discussions of rights, however, typically discuss moral rights. There is no settled understanding of how moral rights should be understood. Some have thought that moral rights are simply the rights that should be recognized by a legal system (Sumner, 1987). However, others insist that it makes sense to assert that someone has a right even when it is impossible or undesirable to enshrine such a right in the law (Feinberg, 1992). If valid, the child's right to be free of religious schooling discussed below may be a right of this kind, a moral right that it is inappropriate to recognize legally.

Children's Rights

Are children, particularly young infants, the sorts of individual to whom moral rights can be ascribed? The answer to this question turns on how rights and, specifically, claim-rights are further specified. So-called choice theorists of rights claim that rights protect a choice on the part of the right-holder, a choice over whether the correlative duty is performed or not. So, for example, if I have the right to \$10 from you I can insist that you honor your duty to give the money to me or, alternatively, I may cancel the duty, as I choose. Yet if rights protect choices, then it is difficult to see how individuals who are incapable of exercising rational choice – neonates and infants – can be right-holders. The choice theory has been challenged by the interest theory of rights, in which rights protect the interests of individuals. Some interest theorists insist that because neonates do indeed possess rights – rights not to be harmed and to be cared for in various ways, for example – the choice theory of right must be mistaken as a general account of rights (MacCormick, 1982). Indeed, the

interest theory seems more plausible because it can accommodate the thought that is central to the choice theory, namely, that many rights protect choices, by citing the important interest we have in leading autonomous lives. However, the interest theory resists the extreme claim that rights protect only choices, for it can also justify rights for individuals who are not capable of choice, such as young children or those who are, or become, incapable of rational judgment. It is open to choice theorists to reply that children can be protected without an appeal to their rights. Indeed, few believe that an account of rights offers a complete conception of how children ought to be viewed and treated. An adequate theorization of the place of children in our moral deliberations must take into account the duty on the part of parents and others to care for children, regardless of whether those duties can be correlated with particular rights (O'Neill, 1988). Nevertheless, it would be unfortunate to eschew the language of rights when addressing issues of the place of children, for if, as many think, the rights of individuals have priority over other kinds of normative claim and should be respected even at considerable social cost, then not conferring rights on infants might place them in a vulnerable position.

If children have rights, how should we describe the rights they enjoy? This question must ultimately be answered by appealing to a conception of morality or political morality. Liberal accounts tend to assert an individual's interests in developing and deploying a sense of justice and the capacity to lead an autonomous life. With respect to the latter some defend what Feinberg (1980) calls "a right to an open future." Children have a right to an education that will equip them with the physical and mental wherewithal to deliberate rationally between different types of lifestyle or conceptions of how to lead good lives. A right to an open future is a claim-right that places others, including parents, under the duty not to school their child in ways that would hinder the child's later autonomous deliberation as an adult. The right also requires others to provide and bear the costs of this education. Furthermore, this right is one that comes without a liberty-right: a child does not enjoy the option of taking up or declining an autonomy-supporting education as she chooses – others may force her to be educated in this way should that prove necessary.

How to elaborate the rights that follow from an interest in autonomy has been the subject of considerable disagreement. The open future elaboration is practically demanding, because it depicts parents who prevent their child from exposure to what they take to be blasphemous or otherwise unworthy lifestyles as wronging their child. Nevertheless, the right to an open future does not rule out schooling that has the aim of imparting controversial religious convictions to a child if this schooling does not inhibit the child's later autonomous reflection in adulthood. A more restrictive right that is defended by some asserts that the interest in autonomy should be elaborated in a way that generates

a duty on parents and other adults not to seek to cultivate controversial religious beliefs in their children. In this alternative elaboration, parents should retain a neutral stance with respect to these matters, leaving them to be addressed by their child when she possesses the capacity for autonomous reflection (Clayton, 2006).

The question of which rights are implied by our interest in developing and deploying a sense of justice has also produced considerable debate among liberal philosophers. Some claim that the interest generates a right to an education that develops the child's sense of justice and certain associated moral and political virtues: the capacity to deliberate rationally on questions of justice, tolerance of religious difference, courage to participate in public affairs and to challenge the political views of others, and a commitment to the ideals of freedom and equality (Callan, 1997). An alternative to this view, defended by Brighouse (1998), proposes a conception of citizenship education that equips children with the knowledge, understanding, and ability to engage in politics without imparting any particular motivation or set of virtues to them. His case rests on the claim that the legitimacy of the state depends upon the free consent of the governed, which would be jeopardized by the cultivation of particular political beliefs.

Nonliberal conceptions of political morality view the rights of children differently. Communitarian or nationalist accounts, for example, view children's interests in terms of an initiation into the particular shared understandings of the community or nation and the educational rights that protect this putative interest will consequently differ from the liberal conception. However, there are certain categories of rights that are common to different conceptions. First, children have developmental rights – rights that generate duties on others to facilitate the growth of the child in terms of health and education. But second, children might have interests that do not refer to their interest in developing into adults. Some of these are rights to be free of certain kinds of treatment, such as being physically harmed, or entitlements to certain kinds of care. Some claim that children enjoy agency rights in virtue of their interest in choosing between different options and pursuing their ambitions, even when their choices lack the level of rationality or deliberation that is typical of adults.

In every society, children lack certain legal rights that are enjoyed by adults: the right to vote, for example. Can such inequalities in the enjoyment of rights be justified? Some have argued that they cannot on the grounds that age is only imperfectly correlated with the ability to vote, and it is objectionably discriminatory to deny the franchise to the competent young while enfranchising older incompetent individuals (Harris, 1982). Others find discrimination on the basis of age to be less objectionable than is discrimination on grounds of sex or race, for example (Daniels, 1988). Whereas sexism and racism produce inequalities that last a lifetime, disenfranchising

the young is a temporary loss of rights for everyone at a certain point in her life.

If age-based discrimination is, in principle, permissible, further questions must be addressed to gauge whether denying children certain rights is justifiable. In the case of the vote, for example, some argue that it is unjust to tax or conscript individuals without extending the franchise to them. Others argue that the entitlement to vote is dependent upon satisfying certain conditions, and there are further debates about the particular personal traits an age group must generally exhibit to be entitled to vote: proposals include the capacity to determine the shape of one's own life or, more demandingly, the ability and willingness to engage competently with political questions (Archard, 2004; Chan and Clayton, 2006). Similar debates attend other rights from which children are standardly excluded: the right to drive vehicles, to marry, and to various freedoms to govern one's own life.

Parents' Rights

Few deny that parents enjoy rights. These rights comprise claim-rights to do various things with, to, or for their children, which generate duties on others not to interfere, or perhaps to facilitate, these activities. Notwithstanding this consensus, the itemization of these rights and their justification have produced considerable controversy. Consider the justification of parental rights first. The central debate here is between those who claim that the rights of parents should be justified on the basis of the interests of the child alone, and those who claim that the interests of parents count independently in determining rights. This debate focuses on the issue of how to allocate custody over children, where custody is understood as constituted by a bundle of rights with respect to a child's life. The natural outcome of purely child-centered justifications of parental rights is what has been called the best custodian condition: childrearing rights with respect to a particular child should be allocated to the interested individuals whose parenting can be expected to be in that child's best interests (Vallentyne, 2003). On this view, the fact that a person begets a child does not imply that she has a right of custody over him. Of course, there might be good reasons to think that having a biological or procreative relationship with a child is generally correlated with being the best custodian, but this need not always be the case (Archard, 1995).

The alternative view, the dual interest view as it has been labeled, asserts that individuals can have a fundamental interest in being a parent, should that be one of her life choices, an interest that should be protected by the enjoyment of custodial rights over the children she begets or adopts. This view might accept that the child's interests carry considerable weight. Consequently, it might concede that parents can forfeit their custodial rights as a

consequence of neglecting or harming their child. Nevertheless, the dual interest view insists that in many cases parents may retain custody even when it is clear that other interested individuals would offer the child in question a better start in life (Brighouse and Swift, 2006; Clayton, 2006).

Both views concerning the justification of parental rights acknowledge that the interests of children and parents can conflict. Child-centered views resolve this conflict in favor of the child, arguing that our interests with respect to parenting are less weighty than our interest in having a better start in life. Yet this position might be too extreme. In other spheres of social policy, the conflict between children and adults does not always favor the former. Medical expenditure, for example, is not overwhelmingly devoted to healthcare, the principal beneficiaries of which are children. Indeed, much of the healthcare budget is devoted to treating diseases and conditions that beset people later in life, even though the health of children might be improved with a reallocation of expenditure. Similarly, the best custodian condition is best for us as children, but the price to be paid for this is that as adults we enjoy a diminished opportunity to be a parent, should that be our ambition.

Parental claim-rights generate duties for others. Most simply, custodial rights place duties on others not to interfere with parental choices with respect to where the child lives, the extent to which he may play with friends, and what kinds of books he reads at home. However, such rights might be more or less limited. For example, custodial rights might be limited by the requirement that the child is educated in an appropriate environment until he reaches a certain age, and the right of parents to choose what the child reads in the home does not necessarily extend to them enjoying a veto with respect to reading material used in school.

Within educational philosophy, disagreements about the extent of parental rights have been played out in two important debates over school choice. The first concerns the extent to which parents have a moral right to choose the content of the curriculum their child receives. Several views have been asserted within this debate. As outlined above, many liberals appeal to the child's right to autonomy. Acknowledging such a right is not sufficient to justify legal constraints on parental choice, because it is compatible with the claim that parents are the best judges of their child's interests. Nevertheless, most autonomy-affirming liberals deny the parents as best judges claim and favor legally constrained parental choice to protect children's interests in becoming autonomous and in acquiring political understanding or a sense of justice (Feinberg, 1992; Brighouse, 2000). However, those who regard initiation into any one of several conceptions of the good life, whether or not that life involves autonomy, tend to favor more extensive parental choice with respect to the school curriculum, though they might remain committed to constrained choice for the sake

of other values, such as the cultivation of various civic virtues (Galston, 2002).

A second important debate over school choice concerns distributive justice. Is it permissible for parents to purchase expensive inequality-generating schooling for their child? In this debate, the worry is not about parents acting in ways that harm their children, but about their school choices setting back the interests of other children whose prospects might be worsened, relatively or absolutely, by these decisions. There are many different considerations that must be addressed to determine whether private schooling should be legally permissible: the importance of equality of opportunity, the effects on others' education or economic prospects of private schooling, and the extent of justifiable parental partiality, for example. For present purposes, it is sufficient to note that granting parents rights of custody over their children does not imply that parents have the right to give their child the best start in life irrespective of the consequences for others of their decisions. More extensive rights of school choice require further justification and have been challenged on grounds of their incompatibility with the ideal of social justice (Swift, 2003).

Whether or not parental claim-rights are limited, they often lack associated liberties. In other words, the fact that parents have rights to make certain choices that affect their child does not imply that they may choose as they like. Claim-rights protect choices against the interference of others, but do not necessarily mean that the protected choice is unconstrained by moral considerations. A liberal view, for example, might protect against interference by others a parent's decision to shield her child from what she takes to be the dangerously anti-religious views expressed in Philip Pullman's *His Dark Materials*. Nevertheless, it would, nonetheless, insist that the parent should not exercise her right in this fashion, because to do so would diminish her child's development or autonomy.

Nevertheless, even without associated options or liberties, parental claim-rights might retain value in protecting the ambition of adults to honor the requirements of parenthood. Many hold the view that even if individuals enjoy very little discretion with respect to how they parent, it remains valuable for them to raise their children well: their ambition is not satisfied just in virtue of their children being raised well by others.

Another aspect of parental rights that has received relatively less attention within educational philosophy concerns who should bear the costs of children. Raising children is costly in terms of time, emotional commitment, and financial resources. Does the right to parent generate a duty on the part of others to bear some of these costs and, if so, should these costs be incurred by other parents or by everyone regardless of whether she parents?

One reason for thinking that the right to parent sometimes includes a right for the costs of parenting to be

shared by others arises from the fact that these costs are unequally distributed due to luck. For example, a child with a congenital disability might be more difficult or costly to bring up. Many argue that it would be unjust for this child or his parents to bear all of the burdens associated with this bad luck. Consider another case. Some individuals might be worse equipped to parent than others are, because they lack a particular set of skills that is useful in parenting or because they enjoy fewer monetary resources through no fault of their own. If they are to have an adequate opportunity to parent, parental rights must include the requirement that resources are transferred from others to enable such individuals to parent effectively. Moreover, if we value children having equal opportunities in life, we might argue for wealth transfers to ensure that differences of family wealth do not translate to inequalities in life chances. For the sake of both parents and children, then, there might be reasons to regard parental rights as justifying duties on others to pay for some of the costs of parenting. However, the extent to which parental rights justify others bearing the costs of parenting is controversial, because certain costs are ones for which individual parents might be responsible and, therefore, do not warrant compensation.

If the costs of raising children should be shared, the question remains between whom they should be shared. Some have argued that because children are a net benefit to others, because they constitute the next cohort that will pay for social services and pensions, everyone, the childless included, are duty bound to contribute to their upbringing. It is consistent with this view to claim that because parents benefit in further ways from their children, from pursuing their ambition to be parents for example, they should bear a greater share of the costs. Nevertheless, on this view parental rights justify duties on everyone to contribute to the raising of their children (George, 1987). A more skeptical view asserts that individuals without children are not required by justice to bear the costs of others' parenting. On this view, becoming a parent is typically a decision taken by individuals motivated by their own ambitions rather than a concern for the long-term welfare of everyone. Since this is the case, it is not obvious why others are duty bound to contribute to the child's upbringing, even if the child's later life has spillover benefits for everyone (Casal and Williams, 2004).

Conclusion

One general observation that might be made about the issue of the rights of parents and children is that their discussion cannot be divorced from wider questions about political morality. Certain rights might be justified by appeal to the fundamental interests we have as children and adults. As children we have fundamental interests in

being protected against various harms, and in being cared for and educated appropriately. As adults, we might have an interest in occupying the role of parent. The rights that are generated by these interests must be accommodated in any plausible theory of politics and education. However, other debates about the rights of parents and children can be resolved only by appeal to a theory of political morality or of what we owe to one another. If we are to determine whether parental rights include an element of child support financed through general taxation, or whether they favor extensive rights of school choice, it seems we have no alternative but to appeal to a theory of political legitimacy or distributive justice. This explains why certain claims concerning the rights of children and parents are so contested. That contestation reflects deeper disagreements about what we owe to one another as human beings.

See also: Autonomy; Justice and Care; Liberalism and Education; Religious and Spiritual Education; School Choice and the Common School; The Aims of Education; The Capabilities Approach.

Bibliography

- Archard, D. (1995). What's blood got to do with it? The significance of natural parenthood. *Res Publica* **1**, 91–106.
- Archard, D. (2004). *Children: Rights and Childhood*, 2nd edn. London: Routledge.
- Brighouse, H. (1998). Civic education and liberal legitimacy. *Ethics* **108**, 719–745.
- Brighouse, H. (2000). *School Choice and Social Justice*. Oxford: Oxford University Press.
- Brighouse, H. and Swift, A. (2006). Parents' rights and the value of the family. *Ethics* **117**, 80–108.
- Callan, E. (1997). *Creating Citizens: Political Education and Liberal Democracy*. Oxford: Clarendon Press.
- Casal, P. and Williams, A. (2004). Equality of resources and procreative justice. In Burley, J. (ed.) *Dworkin and His Critics*, pp 45–69. Oxford: Blackwell.
- Chan, T. W. and Clayton, M. (2006). Should the voting age be lowered to sixteen? Normative and empirical considerations. *Political Studies* **54**, 533–558.
- Clayton, M. (2006). *Justice and Legitimacy in Upbringing*. Oxford: Oxford University Press.
- Daniels, N. (1988). *Am I My Parents' Keeper? An Essay on Justice between the Young and Old*. New York: Oxford University Press.
- Feinberg, J. (1980). The child's right to an open future. In Aiken, W. and LaFollette, H. (eds.) *Whose Child? Children's Rights, Parental Authority, and State Power*, pp 124–153. Totowa, NJ: Rowman and Littlefield.
- Feinberg, J. (1992). In defence of moral rights. *Oxford Journal of Legal Studies* **12**, 149–169.
- Galston, W. A. (2002). *Liberal Pluralism: The Implications of Value Pluralism for Political Theory and Practice*. Cambridge: Cambridge University Press.
- George, R. (1987). Who should bear the cost of children? *Public Affairs Quarterly* **1**, 1–42.
- Harris, J. (1982). The political status of children. In Graham, K. (ed.) *Contemporary Political Philosophy: Radical Studies*, pp 35–58. Cambridge: Cambridge University Press.
- Hohfeld, W. N. (1919). *Fundamental Legal Conceptions as Applied in Judicial Reasoning*, Cook, W. (ed.). New Haven, CT: Yale University Press.
- MacCormick, N. (1982). Children's rights: A test-case for theories of right. In MacCormick, N. (ed.) *Legal Right and Social Democracy*, pp 160. Oxford: Clarendon Press.
- O'Neill, O. (1988). Children's rights and children's lives. *Ethics* **98**, 445–463.
- Sumner, L. W. (1987). *The Moral Foundation of Rights*. Oxford: Clarendon Press.
- Swift, A. (2003). *How Not to Be a Hypocrite: School Choice for the Morally Perplexed Parent*. London: Routledge.
- Vallentyne, P. (2003). The rights and duties of childrearing. *William and Mary Bill of Rights Journal* **11**, 991–1010.

Further Reading

- Archard, D. (2003). *Children, Family and the State*. Aldershot: Ashgate.
- Archard, D. and Macleod, C. (eds.) (2002). *The Moral and Political Status of Children*. Oxford: Oxford University Press.
- Dwyer, J. (1998). *Religious Schools v. Children's Rights*. Ithaca, NY: Cornell University Press.
- Hannan, S. and Vernon, R. (2008). Parental rights: A role-based approach. *Theory and Research in Education* **6**, 173–189.
- Levinson, M. (1999). *The Demands of Liberal Education*. Oxford: Oxford University Press.
- Macleod, C. (1997). Conceptions of parental autonomy. *Politics and Society* **25**, 117–140.

Peace Education

I Gur-Ze'ev, University of Haifa, Haifa, Israel

© 2010 Elsevier Ltd. All rights reserved.

The Uniqueness and Aims of Peace Education

Current peace education does its utmost to be considered a unique educational practice, a field of research, and a discourse that has its own philosophical, social, rhetorical, and political aspects. As an eclectic and ambiguous construction, peace education, at present, enjoys the support and appreciation of most theoretical orientations and international political establishments. Enlightened modernists and soft postmodernists, feminists, critical thinkers, conservatives, and liberals all ardently participate in celebrating peace education. The challenge of legitimizing peace education as an academic discipline or field of research and as a much-needed educational practice is quite demanding. Salomon and Nevo present here a serious question: "What is peace education? What is the core of peace education, its defining attributes? What, if anything, distinguishes its most prototypical instantiations from other, similar fields?" (Salomon and Nevo, 1999: 5–6). According to Ben-Porath "the field entitled 'peace education' is in fact so broad that authors disagree on the description of the problem they wish to address and correspondingly on the proper solution, as well as on the site in which peace education is to take place" (Ben-Porath, 2003: 525–534).

The theoretical stance of the field actually calls for questioning the very justification of peace education as a distinctive, well-defined, self-reflective field of research and as a specific, theoretically justifiable educational practice. It is worth mentioning that, analytically, the reciprocity of its two fundamental concepts – the concept of peace and the concept of education – is hardly ever philosophically analyzed while each is highly problematic, in itself, and the two, together – in their reciprocity (Gur-Ze'ev, 2001). Some attempts to conceptualize peace education have been made in recent years, even if only as very modest preliminary efforts that left the main theoretical work for future scholars to attempt afresh (Salomon, 2004: 257; Cooper, 2005: 2; Page, 2004: 4; Ben-Porath, 2003: 525, 2006). According to Maxine Cooper "there is no one clear agreed upon definition of peace education. Authors like Harris (2004), Vriens (1997), Staub (2003) and Thompson (2003) have defined peace education in a variety of ways. This multiplicity of definitions has led to theorists such as Ben-Porath (2003) and Gur-Ze'ev (2001) stating that current definitions of peace are too wide and nebulous (Cooper, 2005: 2)."

"What is peace education?" Salomon and Nevo (1999) courageously ask the right question. "What is the core of peace education, its defining attributes? What, if anything, distinguishes its most prototypical instantiations from other, similar fields?" They reply: "numerous programs are called 'peace education,' ranging from violence reduction in schools to learning about war and peace, and from democratic education to the cultivation of self-esteem. Subsuming all of these under the category of peace education tends to blur important distinctions, such as between the kind of peace education that is carried out in areas of conflict, such as northern Ireland, and programs designed for more peaceful regions. Similarly, too wide a category tends to lump together programs designed to cultivate a universal peaceful outlook with programs aimed at promoting a peaceful disposition toward a particular group, race or nation to replace collective sentiments of hatred, discrimination, and hostility" (Salomon and Nevo, 1999: 5–6).

In their 2002 paper 'The nature of peace education,' Salomon and Nevo conclude that "neither scholarly nor practical progress can take place in the absence of clear conceptions of what peace education is and what it is to serve" (Salomon and Nevo 2002: 3). It is of vital importance for peace education – as a field of research and as praxis – to enter into this unknown land. A good beginning could have been the elaboration of central categories, such as violence, structural violence, metaphysical violence, counter-violence, power, resistance, revolt, rebellion, and disobedience. Such an elaboration should be realized synchronically and diachronically to be worthy of its name.

Central Trends in Current Peace Education

The central trends in current peace education are: (1) a modern, positivistic trend; (2) a modern critical trend; (3) postmodern, multicultural, and postcolonialist trends; (4) religious (mainly Christian) trends. Some of the modern and some of the postmodern subrends are more socially and critically oriented than others. In current peace education, the social sensibility and the focus of the social context of different types of violence are articulated in different and even within conflicting philosophical, ideological, and educational agendas. While most nongovernmental organizations (NGOs) and the United Nations

Educational, Scientific and Cultural Organization (UNESCO) discussions and agencies are sensitive to the social and conflictual dimensions of the threats to peace – some modern trends that emphasize conflict-resolution skills, and some postmodern and premodern trends, do not share such hopes and sensibilities. The critical sensitivity to the social context of peace education is clearly manifested in a typical UNESCO statement that declares that “there can be no genuine peace when the most elementary human rights are violated, or while situations of injustice continue to exist; conversely, human rights for all cannot take root and achieve full growth while latent or open conflicts are rife . . . Peace is incomplete with malnutrition, extreme poverty and the refusal of the rights of people to self-determination. . . The only lasting peace is a just peace based on respect for human rights. Furthermore, a just peace calls for the establishment of an equitable international order, which will preserve future generations from the scourge of war” (Brock-Utre, 1983: 2).

Some of the modernistic-oriented trends in peace education are more Marxist-oriented than others and some are more positivist-oriented among Marxists and liberals alike. The positivistic-oriented trend manifests a strong belief in teaching conflict resolution strategies. Very little room is reserved here for peace education as an explicit set of ideals and values. Conflict-resolution skills are, in this context, conceived as a matter of objective professional knowledge, good didactics, and fully developed rational participants. In its more mature versions, this kind of peace education enhances sensitivity for the need to cultivate the conditions for the realization of the rational potential for the development and realization of conflict-resolution skills and procedures. Skills are to be taught with the quest for justice, responsibility, common good, or, at least rational utilitarianism. Some, like Chetkow-Yanoov, believe that people might be educated and taught for conflict-resolution skills in the most specific manner (Chetkow-Yanoov, 1996: 12–28). The success of such an education, some claim, is not only possible – it is even measurable in a relatively unproblematic manner (Harris, 1996). This trend is, in the practical sense even if not in principle, in conflict with another non-critical trend – the widely celebrated UNESCO peace-education ideology that emphasizes the need for education for “culture of peace” (UNESCO, 2008).

The critical trend is the one inspired by or realized within the framework of critical pedagogy. It offers peace education a special role in light of its immanent emancipatory commitment and the centrality of the voice and the interests of the victims within this educational agenda. Freire, bell hooks, Rivage-Seul, and many other critical educational theorists like McLaren and Giroux worked hard to give voice and visibility to the perspectives and interests of the marginalized. Rivage-Seul tries to develop a critical peace education based not on Western

universally humanistic-oriented hegemonic perspectives and their interests, but on the contrary – on the locality and self-evidence of the silenced, based on Freire’s concept of moral imagination, which will hopefully transcend “the bounds of technical thought” (Rivage-Seul, 1987: 160). The more modern-humanistic-oriented trends within critical peace education are attacked by postmodern-oriented critical peace educationalists who claim that the traditional (modern-oriented) Freirean framework (Freire, 1972) is universalistic and essentialist, and, ultimately, conceives hierarchical-oppressive relations between teacher and student as a precondition for educational progress (Gur-Ze’ev, 1998a: 463). These critical educational thinkers denote the educational implications of the centrality of anti-universalistic-oriented diversity, epistemology, culture, and politics. Within a deep belief in the centrality of contingency, locality, and nonlinear-never-objective epistemology and philosophy of history, they articulate their alternative peace education.

This is also an opening for a philosophical and political contradiction that challenges postmodern critical peace education. It is a contradiction between (1) the commitment to the agenda of human rights and the defense of democracy as a universally valid set of values and ideals and (2) a strong antiuniversalist and antiessentialist conviction that results in automatic identification with the values, interest and narrative of the victim as against Western/white/American cultural and political violence that allegedly imposes a kind of universalism that destroys the uniqueness and potentials of all other voices and imposes upon them self-misconception, distorted identities, and even resistance solely within the framework of the colonialist-inflicted conceptual apparatus. Philosophically, educationally, and politically, this version of peace education does not end in resistance to dogmatism and violence as such; it concludes in resistance only to one kind of dogmatism while understanding and even supporting other violences – the counter-violence of the oppressed that realizes their self-evidence, closure, and refusal for critical dialog. This commitment of postmodern critical peace educationalists stands in conflict with other trends in current postmodern emancipatory/critical/radical education which emphasize the contingent, temporary, narrativistic, and antiuniversalistic/antiobjectivistic nature of human existence, representation, self-constitution, and communication. This contradiction between modern and postmodern critical peace education and within postmodern critical peace education itself is more dramatic in some of its branches than in others. Strong manifestations of this can be found in current multicultural and post-colonialist peace education.

The multicultural trend in peace education emphasizes diversity as a precondition for peace, in contrast to the liberal agenda of enhancing shared values and a homogeneous type of reflection toward universal solidarity and

responsibility as praised by liberal (Aduan and Bar-On, 2004; Bar-Tal, 2005) and existentialist-oriented peace educators (Gordon, 2005) and by most of the theorists of civil education (Rennebohm-Franz, 1996: 266).

The postcolonialist trend in peace education directly confronts liberal trends that work for the promotion of a humanistic version of pluralism and tolerance that will be nourished by (Western-oriented, universal, liberal) shared values and ideals as a gateway to universal openness, dialogism, pluralism, responsibility, and, eventually, peace. In recent years, there has been increasing sensibility in the liberal agenda to the presence and relevance of cultural diversity and its multicultural educational implications, especially in arenas of dense conflict (Taylor, 1992). Critical education theorists in the postmodernist multicultural and postcolonialist framework – like Peter McLaren, Ward Churchill, and Ilan Pappé – not only understand, but actually support the explicit violence of the oppressed and present it as legitimate counter-violence that echoes only the original colonialist violence and that is free of moral responsibility to its actions that ultimately will build the bridge from colonialism and resistance to postcolonialism and peace. Within this framework effective violence of the oppressed is actually nothing but peace education-in-action.

Multicultural peace education bears a strong family resemblance to postcolonialist and postmodern peace education. The differentiation between these trends is not always unproblematic. Multicultural peace education (as all other members of this family) is explicitly founded on postmodernist theory as against modern national education (Pappé, 1999: 233). It is conceived as a precondition for a worthy peace education (Pappé, 1997: 221). The new, post-modern methodology – in the service of peace education – is founded here on two assumptions: (1) negating the natural sciences as a model for human sciences, and (2) doubting the objectivity of the researcher. For Pappé, “reality is a representative and interpretative issue. The historian, therefore, is allowed to represent it as he or she wishes . . . the historian should give freedom to the sole quality which gives him the power to write about the past – imagination. Not the kind of imagination which supports a reliable representation of the past as modernists understood this conception but rather imagination of the kind that will enable the historian to produce aesthetically the history, according to the known styles in comedy, tragedy and farce” (Pappé, 1999: 235).

This is to occur in the service of the silenced and oppressed voices in a multicultural society (Pappé, 1999: 236). Peace education is made possible here by means of this new methodology for giving voice to silenced ones on the road to peace, which is conditioned by the deconstruction of the hegemony and its structural violence (Pappé, 1997: 236).

The liberal trend in peace education, in its most theoretically developed version, is represented by Johan

Galtung – the prominent founder of peace education as an academic field of research. Actually, Galtung offers the most theoretically developed concept of peace education *per se*. As early as in the 1960s, Galtung challenged the then excessive weight imparted to peace as the presence of the absence of direct violence. In his peace theory, he included the concept of structural violence and offered the conceptions of “positive peace” and “negative peace” – notions that are richer and philosophically more challenging than most conceptualizations in current peace education. According to Galtung “two concepts of peace should be distinguished: negative peace, defined as the absence of organized violence between such major human groups as nations, but also between racial and ethnic groups of the magnitude that can be reached by internal wars; and positive peace, defined as a pattern of cooperation and integration between major human groups” (Galtung, 1975: 29).

In light of his understanding of positive peace, preceding the traditional positive Utopia projects, Galtung understood that he had to develop a vision of this harmonic future of humanity but also a well-defined theory of violence: to address the role of violence in furthering progress in history. An attempt to understand the nature of violence is a precondition to making possible positive peace as a realization of human potentials. Accordingly, Galtung tells us that violence is “present when human beings are being influenced so that their actual somatic and mental realizations are below their potential realization” (Galtung, 1969: 168).

While developing an historical perspective and claiming that “*the needs perspective* has been used as an instrument of the human right tradition” (Galtung, 1994: 107), Galtung includes – in peace education – the commitment to rights such as “the right to sleep,” “the right to co-existence with nature,” and “the right to be free to experiment with alternative ways of life” (Galtung, 1994: 104). He concludes that peace education should also relate to ecological challenges and cultural colonialism while challenging the invisibility of cultural violence and structural violence.

The Concept of Peace

The concept of peace that is so central to each of the various versions of peace education needs to be analyzed diachronically and synchronically, within premodern, modern, and postmodern arenas. Its philosophical, cultural, and political functions may fruitfully be distinctively elaborated in light of (1) a culture that is occupied with the presence of God, (2) in an arena occupied with the awareness of the bursting of the alternative to this rich presence, and (3) in arenas that are free of both God and

the killing of God. Gnosis and postmodernism manifest this trend in distinct ways.

Peace is a theological concept, even in its secular transformations. Within the framework of monotheism, peace obtains its meaning in light of the tension between human exile from the Garden of Eden and the open possibility of redemption within a Godly teleological history. Within the framework of a sacred history, the telos of a monotheistic-oriented progress toward redemption is different from that of (1) secular modern conceptions of progress toward peace and (2) postmodern conceptions of progress toward peace and the relations between conflicting violences and the possibility of harmonious human existence. Simultaneously, we should distinguish between monotheistic, pantheistic, Gnostic, and postmodern concepts of peace. Within the pantheistic tradition of special relevance is the Buddhist conception of peace. Nirvana, one could claim, should not be conceived as the annihilation of life but as a possible realization of peace. Classic Greek conceptions of the relations between violence and peace are also of much value for such an analysis. Within the monotheistic tradition, the Jewish vision of peace has a pivotal role as it is offered by Yeshayahu. Nirvana, one could claim, should not be conceived as annihilation of life but as a possible realization of peace. Isaiah 11:6 offers us a premodern monotheistic utopian vision of peace within which “The wolf also shall dwell with the lamb, and the leopard shall lie with the kid; and the calf and the young lion and the fatling together . . . They shall not hurt nor destroy my holy mountain: for the earth shall be full of the knowledge of the Lord, as the waters covers the sea.”

The monotheistic theological concept of history enables a collective, even cosmic Utopia, which is very different from the Homeric conception of individual *arête* (worthy life). It is important to note that pantheistic-oriented antiquity – in opposition to monotheistic-oriented antiquity – offered many different conceptions of peace, ranging from disregard to heroic refusal of the Jewish conception of peace. The Homeric hero and Greek civilization, in general, “was dominated by a conception of life which was essentially warlike. Whatever the event, whatever the change, conflict in one form or another was always close at hand. Religion often assumed violent overtones” (Zampaglione, 1973: 18). The Gnostic tradition was a significant challenge both to pantheistic and monotheistic premodern traditions and their concept of peace. Accordingly, it offered a very different view of peace education. These traditions transformed themselves in modernity and, even in the postmodern condition, they offer rival conceptions of peace and peace education.

Modern peace education and its concept of peace are settled on the act or on the process of the killing of God. Modern philosophies have secularized theology and have much in common with the premodern concept of God as

it is revealed in central concepts such as progress and peace. The presumption of the human telos as home-returning is replaced by a vision of the killing of God as an act of giving birth – known also as the act/process of critique that is part and parcel of a road leading to positive Utopia. This is a vision of a future harmonious state of affairs, which as in Marxian theory will secularize the eschatological religious vision and will offer a tiger-like leap from “prehistory” into “history” (Marx, 1977: 1) or “peace” as an earthly Garden of Eden. Positive Utopianism (Gur-Ze’ev 1996) made possible by human heroism as a feast of eating the flesh of the Leviathan/father/God and edifying the deification of humanity is a magical spectacle: it transforms natural and historical realities.

The concept of progress as the historical realization of central ideals of humanism, such as freedom, manifests the fundamentality of the home-returning project to the Western tradition – a tradition whose impetus is fear of nature, of bestiality, of thingness, of scarcity, and of meaninglessness. It is dramatically manifested in the philosophies of Condorcet, Hegel, and Marx. According to Condorcet, humanity is under eternal self-rebirth and paves the way for itself to truth and happiness, by constantly relying on human reason and the facts. This Utopia is possible, since nature has placed no limits on human self-edification (de Condorcet, 1955: 4). According to these philosophies – in its total fulfillment, within the framework of the ideal of freedom – humanity realizes peace in the form of effective total violence against external and internal nature and all forms of stagnation, borders, unchallenged power, and the various manifestations of peace/consensus/truth. The idea of freedom as an eternal journey – burning, birth giving, and overcoming – demands total control of human existence and unrestrained effective manipulation of nature. Total rational control and effectively silencing any opposition is taken as peace.

Philosophers and poets – and more so workers, women, colored people, and immigrants – have, during the last three centuries, been brutally faced with different forms and, on different occasions, with the naivety, futility, and hopelessness of the humanistic-oriented home-returning project. They have met the fruits of the various and conflicting modern attempts to re-establishing, on the Earth, the Garden of Eden in the form of replacing total naivety, intimacy, and integration with total control of nature in all its dimensions. In the case of Nietzsche, Spengler, Camus, Adorno, Levinas, and Derrida, this encounter offered an alternative to the quest for peace and to the presumptions, frameworks, and telos of the various home-returning projects in the form of Diasporic philosophy, eternal tragic heroism, and enduring nomadism (Gur-Ze’ev, 2004).

In the post-structuralist tradition, two main manifestations of conceptualizing peace are to be reconstructed. According to one, peace – as a quest, as Utopia, and as

value – is produced, reproduced, represented, distributed, and consumed as a symbol and as a cultural good that is nothing but an echo of the contingent power-relations: a productive illusion governing the stability of a specific context. Peace, in its essence, is nothing but violence that successfully defeated its enemies and was victorious enough to silence even the voices and traces of the oppression and suffering that go along with its triumph (Gur-Ze'ev 2007). According to the other, peace is an ideal worth struggling for. Post-structuralist theory should be a medium and a path for empowering a political, cultural, and educational effort for bettering our world in the face of racism, gender discrimination, colonialism, and ecological threats, as Stuart Hall, Cornel West, Gayatri Chakravorty Spivak, bell hooks, and Peter McLaren tell us. Such efforts, however, should overcome the Eurocentric, phallogocentric, and racist presumptions that made possible Whiteness, patriarchy, and colonialism – which until now have directed traditional (humanist-oriented and Christian-motivated) peace education. Some post-structuralist thinkers, like Rosi Braidotti, Homi Bhabha, and Bracha Ettinger, offer a less political, yet no less practical, practice and vision of peace within an endless heterogeneity and eternal diversity. In this light, Ettinger articulates peace within the framework of co-poiesis and asserts that “the originary event of peace is compassion” (Ettinger, 2008). Braidotti integrates the struggle for sustainability of the Earth, peace education and the possibilities for eternal nomadism in cyberspace in an original manner which is an especially fruitful challenge to current peace education (Braidotti, 2006).

Contradictions Between Explicit and Implicit Aims of Peace Education

Peace education faces a challenging contradiction. While in its various versions it is explicitly pro-life, it represents implicitly a fear of life: an escape attempt into stability and consensus, irrespective of the nature, *telos*, or meaning of the system that it is committed to stabilize. As long as peace education is focused on peace as the elimination or prevention of destabilizing the hegemonic order of things irrespective of the concrete and specific values, ideals, practices, and *telos* of the system – in its essence, it is a conservative, antilife power. Peace education in its central manifestation has neither ambitions nor yardsticks to evaluate peace in essentially different contexts and it fails to judge and resist victories as long as they practice effective violence and have the upper hand, no matter the cost, the fruits, or the *telos* of these victories. As Augustine claimed – earthly peace, actually, represents the “Earthly City” or the more effective secular violence and it is the opposite of genuine “peace” (Augustine, 1957: 415).

Peace education actually challenges the essence of dualism, dialectics, heterogeneity, and life itself (Gur-Ze'ev, 2007: 305–306). It is understood by some scholars as a rescue attempt from the spatial-time factuality of suffering and it manifests a belief in a possible human progress toward Utopia. Montaigne, Haman, Yaacoby, Schelling, Schopenhauer, and Adorno offer a powerful alternative to great modern believers in the possibility of progress toward enlightened education and the realization of peace. The very idea of realizing and sustaining positive universal law, Utopia, or homogeneity was impossible or misleading, according to these critics of universal reason and rational emancipatory powers.

The very notion of peace, one might claim, represents a quest for homogeneity, stability, or victory of Thanatus – a return to nothingness. Ontologically, peace education presupposes the preference and the possibility of the end of dialectics, the end of contradictions, and of difference that makes a difference. This is a legitimate philosophical position and Schopenhauer (1966/1969) and Mainländer (1876) were the most dramatic among the defenders of this philosophical stand.

Counter-Education for Peace?

Eternal Diasporic existence, border-crossing, and improvisation reflect their relation to Eros as an impetus to its enduring responsibility and creativity within co-poiesis that re-affirms life. It is a specific, alternative togetherness with the other, with the self, and with the cosmos. Improvisation, in this context, is a life philosophy in a postmodern era that is teachable to a certain degree and surely a worthy educational endeavor (Peretz, 2003). The realities of present globalizing capitalism and the new technoscientific horizon enable the concrete realization of improvisation in individual and public spheres. It is a way-of-life eternally open, always dangerous, and, for all time, individual. It offers neither a quest for a perfect collective or individual life nor a future of redeemed humanity. It cannot promise us a harmonious future that will realize nirvana, peace, final solutions, ultimate justice, or consensual meaning that appeases everyone and everywhere. It is precisely these philosophical, existential, and historical presumptions and visions that Diasporic counter-education must overcome when relating to peace.

Counter-education is not a mere praxis, neither is it a given theory to be faithfully realized in order to establish the genuine heavenly Jerusalem. Peace according to counter-education is the fruit of normalizing education, as are victories, fulfillment of the nations' historical mission, and other manifestations of self-forgetfulness. At the same time, it does not conceive – with Augustine – peace as an outstanding efficient realization of violence in a predeemed world nor as a future possibility awaiting

the chosen ones when they realize the true faith in its authorized interpretation. Peace as openness that streams out of love and hospitalizes responsibility to ones' self, to the Other, and to the world is of outmost importance for Diasporic counter-education. This is because Diasporic counter-education is not simply and abstractly negative, nor is it an active critique. It cannot be content with critically reconstructing the manners by which the various kinds of normalizing education work: fabricating, reproducing, policing, and edifying the conceptual apparatus, values, and yardsticks that enable and glorify the victory of explicit violence.

Counter-education – in contrast to normalizing education – does not strive for any sentimental peace. Yet, it is at peace with the Diasporic truth, so that in the absence of a nonmanipulative consensus around the final ends, and in light of creative love as the only end/home worthy of nomadism itself, improvisation becomes home. Nomadism in eternal Diaspora religiously improvises itself in love of life. Response-ability and respond-ability toward noncollective, toward presubjective and existential kinds of homelessness, and toward erotic Diasporic existence might offer here-and-now prerational and postdia-logical forms of knowledge and togetherness; it might improvise toward new beginnings and a kind of human dance that we have not yet experienced historically, yet we all share in our blessed moments: a kind of awakening toward becoming-toward-the-world (Gur-Ze'ev, 2005) as against becoming-swallowed-by-the-triumph-of-normalizing-education that is most effective in stable systems and days of tranquility, prosperity, and peace.

Bibliography

- Augustine, St. (1957). *The City of God Against the Pagans*. London: W. Heinemann.
- Aduan, S. and Bar-On, D. (eds.) (2004). *Peace Building under Fire*. Beit Jala: Peace Research Institute in the ME.
- Bar-Tal, D. (2005). *Stereotypes and Prejudice in Conflict – Representations of the Arabs in Israeli Jewish Society*. Cambridge: Cambridge University Press.
- Ben-Porath, S. (2003). War and peace education. *Journal of Philosophy of Education* 37(3), 525–533.
- Ben-Porath, S. (2006). *Citizenship Under Fire – Democratic Education in Times of Conflict*, pp 57–75. Princeton, NJ: Princeton University Press.
- Braidotti, R. (2006). *Transpositions – on Nomadic Ethics*. Cambridge: Polity Press.
- Brock-Utre, B. (1983). *Educating for Peace: A Feminist Perspective*. New York: Pergamon Press.
- Chetkow-Yanoov, B. (1996). Conflict-resolution skills can be thought. *Peabody. Journal of Education* 71, 12–28.
- Cooper, M. (2005). Cultivating wisdom, harvesting peace. *International Symposium on Peace Education*. Griffith University. <http://www.griffith.edu.au/centre/mfc/pdf/symposium20005/maxine%20cooper.pdf>.
- de Condorcet, A. N. (1955). *Sketch for a Historical Picture of the Progress of the Human Mind*. Barraclough, J. (trans.). London: Weidenfeld and Nicolson.
- Ettinger, B. (2008). Compassion and com-passion. <http://underfire.eyebear.org/?q=node/512> (accessed September 2009).
- Freire, P. (1972). *Pedagogy of the Oppressed*, Ramos-Bergman, M. (trans.). Harmondsworth: Penguin Books.
- Galtung, J. (1969). Violence, peace and research. *Journal of Peace Research* 6, 167–184.
- Galtung, J. (1975). *Peace: Research, Education, Action – Essays in Peace Research*, I. Copenhagen: Christian Ejlert.
- Galtung, J. (1994). *Human Rights in another Key*. Cambridge: Polity Press.
- Gordon, H. (2005). *Dance, Dialogue and Despair – Existentialist Philosophy and Education for Peace in Israel*. Ala: University of Alabama Press.
- Gur-Ze'ev, I. (1996). *The Frankfurt School and the History of Pessimism*. Jerusalem: Magnes Press.
- Gur-Ze'ev, I. (1998a). The morality of acknowledgement/not acknowledging the other's Holocaust/genocide. *Journal of Moral Education* 27(2), 161–178.
- Gur-Ze'ev, I. (1998b). Toward a nonrepressive critical pedagogy. *Educational Theory* 48(4), 463–486.
- Gur-Ze'ev, I. (2001). Philosophy of peace education in a postmodern era. *Educational Theory* 51(3), 315–336.
- Gur-Ze'ev, I. (2004). *Toward Diasporic Education – Multiculturalism, Postcolonialism and Counter-Education in a Postmodern Era* (in Hebrew). Tel Aviv: Resling.
- Gur-Ze'ev, I. (2005). Critical theory, critical pedagogy and diaspora today: Toward a new critical language in education (introduction). In Gur-Ze'ev, I. (ed.) *Critical Theory and Critical Pedagogy Today – Toward a New Critical Language in Education*, pp 7–34. Haifa: Faculty of Education the University of Haifa.
- Gur-Ze'ev, I. (2007). *Beyond the Modern-Postmodern Struggle in Education – Toward Counter-Education and Enduring Improvisation*. Rotterdam: Sense Publishers.
- Harris, I. M. (ed.) (1996). *Special Issue: Peace Education in a Postmodern World* *Journal of Education* 71(3), 1–11.
- Harris, M. (2004). Peace education theory. *Journal of Peace Education* 1, 5–20.
- Mainländer, P. (1876). *Philosophie der Erloesung*. Berlin: Theodor Hofmann.
- Marx, K. (1977). *A Contribution to the Critique of Political Economy*. Moscow: Progress Publishers.
- Page, J. S. (2004). Peace education: Exploring some philosophical foundations. *International Review of Education* 50(1), 3–15.
- Pappe, I. (1997). Peace education toward the 2000s – New history, multiculturalism and postmodernism in service of peace education. In Gur-Ze'ev, I. (ed.) *Special Issue: Critical Theory and Education. Studies in Education*, 2(2), 221–248.
- Pappe, I. (1999). Educational implications of Israeli multiculturalism. In Gur-Ze'ev, I. (ed.) *Modernity, Postmodernity and Education* pp 233–251. Tel Aviv: Ramot.
- Peretz, A. (2003). *The Art of Teaching and Improvisation – A Compulsory or Emancipatory Dimension?* (in Hebrew). An MA Thesis. Submitted to the University of Leeds (Israeli branch).
- Rennebohm-Franz, K. (1996). Toward a critical social consciousness in children: Multicultural peace education in a first grade classroom. *Theory into Practice* 35(4), 258–275.
- Rivage-Seul, M. K. (1987). Peace education: Imagination and pedagogy of the oppressed. *Harvard Educational Review* 57(2), 154–181.
- Salomon, G. (2004). Does peace education make a difference in a context of an intractable conflict? *Peace and Conflict* 1(3), 257–274.
- Salomon, G. and Nevo, B. (1999). Peace education: An active field in need for research. *Paper Presented at a Peace Education Conference*, The University of Haifa.
- Salomon, G. and Nevo, B. (eds.) (2002). *Peace Education: The Concept, Principles and Practices Around the World*. Mahwah, NJ: Erlbaum.
- Schopenhauer, A. (1966/1969). *The World as Will and Idea*. New York: Dover Publications.
- Staub, E. (2003). *The Psychology of Good and Evil*. New York: Cambridge University Press.
- Taylor, C. (1992). *Multiculturalism and "the Politics of Recognition"*. Princeton, NJ: Princeton University Press.
- Thompson (2003) (in text).

- UNESCO (2008). Culture of peace. UN resolution A/RES 52/13. http://www3.unesco.org/iycp/uk/uk_sum_cp.htm (accessed September 2009).
- Zampaglione, G. (1973). *The Idea of Peace in Antiquity*. Notre Dame: University of Notre Dame Press.

Further Reading

- Ettinger, L. B. (2005). Coppiesis. *Ephemera* 5(X), 703–713.
- Gur-Ze'ev, I. (ed.) (1999). *Modernity, Post-Modernity and Education* (in Hebrew). Tel-Aviv: Ramot.
- Gur-Ze'ev, I. (1999). *Philosophy, Politics and Education in Israel* (in Hebrew). Haifa: Haifa University Press.
- Gur-Ze'ev, I. (ed.) (1996/1997/1999). *Education in the Era of the Postmodern Discourse* (in Hebrew). Jerusalem: The Magness Press, The Hebrew University.
- Gur-Ze'ev, I. (1996/1997/2000). *The Frankfurt School and the History of Pessimism* (in Hebrew). Jerusalem: The Magness Press, The Hebrew University.
- Gur-Ze'ev, I. (ed.) (2000). *Conflicting Philosophies of Education in Israel/Palestine*. Dordrecht: Kluwer academic Publishers.
- Gur-Ze'ev, I. (2003). *Destroying the Other's Collective Memory*. New York: Peter Lang.
- Gur-Ze'ev, I. (2004). *Toward Diasporic Education*, pp 220. (in Hebrew). Tel Aviv: Resling.
- Gur-Ze'ev, I. (ed.) (2005). *The End of Israeli Academia?* Faculty of Education, University of Haifa, Haifa.
- Gur-Ze'ev, I. (ed.) (2005). *Critical Theory and Critical Pedagogy Today – Toward a New Critical Language in Education*. Faculty of Education, University of Haifa, Haifa.
- Gur-Ze'ev, I. (2007). *Dialectics of Homeland and Diaspora in the Pedagogies of Exile in a Post-modern Era* (in Arabic). Ramalla: Madar.
- Gur-Ze'ev, I. (2007). *Against the Industrialization of Israeli Academia* (in Hebrew). Haifa: University of Haifa.
- Gur-Ze'ev, I. and Lidor, R. (2007). *Sport – Politics and Values* (in Hebrew). Hulon: David Rechgold Publications.
- Marx, K. (1977). *A Contribution to the Critique of Political Economy*. Moscow: Progress Publishers.
- Reardon, B. A. (1988). *Comprehensive Peace Education: Educating for Global Responsibility*. New York: Teachers College Press.
- Roth, K. and Gur-Ze'ev, I. (eds.) (2007). *Education in the Era of Globalizing Capitalism*. Heidelberg: Springer.
- Schmidt, H. J. (1968). Peace research and politics. *Journal of Peace Research* 5(3), 217–232.

Philosophy and Educational Research

D Bridges, University of Cambridge, Cambridge, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Action research – Research carried out by practitioners in the sites of their practice and designed to improve that practice through a repeated cycle of reflection, research and change in the practice.

Empirical – as in ‘empirical enquiry’ – Based on observation and experience (and contrasted with more abstract thought characteristic of philosophy, mathematics, and more speculative aspects of science).

Epistemology – or theory of knowledge –

A branch of philosophy concerned with the nature and justification of belief, truth claims, etc.

Ethics – A branch of philosophy concerned with the nature and justification of beliefs about the right and the good, sometimes referred to as *moral philosophy*. But ethics is also used to refer to codes of (moral) practice governing, for example, professional lives or, as discussed in this article, the conduct of research.

Normative – Related to norms, in this case the values and principles which underpin educational policy and practice.

Phenomenography – Enquiry aimed at eliciting and describing the ways in which people experience the world – an attempt to access, for example, the learner’s or the teacher’s subjective experience of education.

Postmodern – Refers to a body of thought and literature which challenges a whole range of assumptions associated with, in particular, the aspirations of modern science and technology. It is a notion which rather eludes definition (and most of those who espouse its various forms would seek to elude definition), but the paragraph in the text seeks to explain what it means in this context.

Practitioner research – Research (usually small scale) carried out by practitioners, most commonly in this context teachers (see also *action research* above).

The Contributions of Philosophy to Educational Research

This encyclopedia will itself bear witness to the fact that the field of educational research is enormously diverse. Education is not a single form of disciplined enquiry:

rather, it is a field of enquiry which draws on a wide range of the intellectual resources of the academy as well as the accumulation of practical wisdom drawn from the experience of educational practitioners. (There are of course interesting philosophical questions about the relationship between these two sources which are commented on briefly below). Its intellectual resources and its methods of enquiry are drawn from disciplines as diverse as history and neuroscience, literary theory and philosophy, behaviorist psychology and ethnography, law and economics, and museology and the creative arts.

This (in part philosophical) picture of the nature of educational research itself points to three kinds of contribution which philosophy can and does make to this endeavor. First, philosophy is itself among the scholarly, systematic, and sustained disciplines which can – and many would argue must – be drawn upon to contribute to educational enquiry. This is philosophy as educational research. Second, the claims and limitations of all of the other forms of enquiry which offer contributions to educational research need to be examined and understood, and this is philosophical work of an epistemological character, that is, rooted in theory of knowledge. This is referred to as philosophy of educational research. Third, any research which requires engagement with human participants and the redistribution of their knowledge and of knowledge about them raises issues of an ethical and political character which have their roots in a long-standing tradition of philosophical ethics and social and political philosophy. The remainder of this article on the contribution of philosophy to educational research is organized under these three headings.

Philosophy as Educational Research

Philosophers may feel a little uncomfortable about their work being described as research, because this term is more commonly applied in contexts in which people are seeking evidence or data of one kind or another. Philosophers tend to be more comfortable with the language of scholarship or enquiry. This is not to say that philosophers have to stay removed from the grounded world of practice or empirical enquiry, though they are sometimes criticized for failure to engage directly with these worlds (Phillips, 2005). It is rather that what they do with such material – as philosophers – is to reflect upon it, analyze it, and interpret it drawing on the rich tradition of philosophical writing which is at the heart of the subject. However, they have on

the whole accepted their role as educational researchers as the price of a seat at the table of the national and international educational-research community.

Their presence in the educational-research community is an essential one, because philosophical questions are central to the theory and practice of education itself. You cannot go far in the consideration of educational policy and practice without engaging with fundamental questions about the aims of education and about the values and principles which ought to govern this policy and practice. These are themselves rooted in conceptions of human being and flourishing, about the individual and society, and about the lives we want to lead – all of which are at the core of the philosophical tradition. You cannot go far in consideration of the curriculum without engaging in these issues but also with questions about the nature and structure of knowledge itself, about what it is to know or understand something, about the relationship between knowledge and skilled performance, about the authority which can be claimed for different kinds of belief – and about the implications of all of these considerations for what we teach and how we teach it. You cannot go far in pursuit of the social-justice agenda in education without encountering contested interpretations of social justice and conflicts with other social principles and values. How far should the state intervene in the name of justice and at what price to individual freedom and responsibility? How compatible is a pursuit of equality with the pursuit of excellence? How far ought one to respect the rights of minority communities to maintain traditional practices which appear to disadvantage women? (A fuller account of the contribution of philosophy of education to educational enquiry is discussed in the BERA website and elsewhere in the encyclopedia.)

It is not that philosophy will necessarily provide simple answers to such questions: it is perhaps more likely to reveal further layers of complexity. The point is that the questions that are illustrated and many like them are questions which have been and continue to be refined, analyzed, discussed, and substantially investigated in a long tradition of philosophical writing, and if any educational practitioner or enquirer ignores this work, they are destined to rehearse simplistic responses which have long been discredited. A research community which claims to represent the highest standards of intellectual endeavor and whose authority lies precisely in its commitment to critical, systematic, and sustained enquiry cannot restrict these requirements to the empirical aspect of enquiry in a field which, as illustrated above, depends significantly upon philosophical considerations.

Philosophy of Educational Research

Philosophical considerations underpin all or most of the central debates about educational research methods and

methodology, for they are all at base about how we can best know about or understand educational policy and practice in their many shapes and forms. In this section this point is illustrated with reference to four of these debates, though this is by no means a comprehensive treatment of the field.

Research and Educational Practice

A large proportion of academic research in education seeks to inform educational practice – in schools, colleges, and universities in professional training environments like hospitals and also in nonformal settings in the workplace, the home, and the community. However, we have to ask whether all these need research as the academy understands it at all or should we have confidence in the kind of situated practical judgment (a notion derived from Aristotle) and the tacit knowledge (as Polanyi, 1966 might call it) developed by teachers and other practitioners? If some further investigation needs to be done, might this best be in the form of practitioner research or action research developed in the classroom and in intimate relationship with the sites in which change is expected to take place and under the control of those responsible for any such change? Such questions are essentially philosophical ones which require answers rooted in an understanding of the nature of practice and the ways in which it can be informed and transformed; of the extent to which general educational prescriptions can be applied to particular situations; of the relationship between theory and practice. In this last case it is not just a question of whether and how theory can inform practice but also of whether that dichotomy does not already distort a proper understanding of the nature of practical judgment. All of this is well-worn philosophical territory, which nevertheless continues to excite contemporary debate (see, e.g., Carr (1986), Elliott (2001) and the last part of Hammersley (1993)).

Research and Educational Policy

Similarly, we may ask questions about the sort of knowledge which ought to inform wider educational policy (as distinct from questions about the often bizarre considerations which do, as a matter of fact appear to inform the decisions of policymakers). If, as philosophers would certainly argue, policy is always driven by normative considerations, where does this leave us in terms of the contribution which educational research can make? (Just possibly with the conclusion that we had better build in some philosophical consideration of this normative framework.) But, clearly, policy needs to be grounded in evidence of some kind about the actual or likely effects and the consequences of doing this rather than that.

This in turn prompts the question “so to what sort of research should we look to provide such evidence?”

This is a question which has provoked enormous debate in both educational policy and educational-research communities. At one end of the spectrum, it has been answered in very restrictive terms. The What Works Clearing House in the United States has set as the gold standard for what it calls evidenced-based practice research which conforms to the standards of the double-blind controlled experiment which has achieved preeminence in medical research. In the UK, the Evidence for Policy and Practice Information (EPPI) Centre established to conduct systematic reviews of educational research operates with criteria which are not quite so restrictive, but which nevertheless exclude large swathes of the research which is carried on across the educational community. (For a fuller discussion of these issues see Bridges *et al.* (2009) and Hammersley (2007)).

At the same time, as steps are being taken to restrict the range of research which is supposed to inform policy, the wider educational-research community itself has expansively embraced a wider and wider range of intellectual resources drawn from almost every corner of the academy – not only from the social sciences (which have themselves extended and hybridized their repertoire) but from the humanities (biography and autobiography, discourse analysis) and, as Elliott Eisner encouraged in his presidential address to the American Educational Research Association, the creative arts (Eisner, 1993).

It is not proposed to enter these debates in this context. The point to be made is that such debates about what kind of knowledge can and should inform educational policy-making are essentially philosophical ones. The claim that only research which adopts successfully the experimental design of the double-blind controlled experiment and, equally, the claim that educational research can benefit from the contribution of the creative arts are both philosophical claims which require not empirical evidence but careful philosophical argumentation and critique if they are to be sustained. The irony for the What Works movement is that no double-blind experiment will ever demonstrate that the experiment should be the preferred form of educational enquiry.

The Qualitative/Quantitative Debate

A third and related area of essentially philosophical debate about educational research is closely related to this last one insofar as it is another example of argument about what kind of research can best illuminate educational policy, practice, and experience. The debate which is often couched in terms of the opposition between qualitative and quantitative research paradigms is referred to here, though most sophisticated contributors to this debate immediately point out that this is both a crude and misleading dichotomy. It is crude because both terms pick out what are in fact some very diverse forms of enquiry: both large population studies in sociology and small controlled

experiments in classrooms, for example, might come under the quantitative label, just as a long-term ethnographic study of playground rituals, a piece of autobiographical writing by a teacher, and a piece of applied critical theory might all be called qualitative. They are misleading because quantitative data might very well have a place in the sort of thick description of, for example, a school or its neighborhood provided in what are normally thought of as qualitative case studies. Equally, some researchers will subject what might otherwise be regarded as qualitative data, for example a careful transcription of a lesson or an interview, to quantitative analysis while others could apply a different, interpretative analysis to the same data.

Such disassembling of the dichotomy has not, however, ended controversy which is seen by some to have ramifications which extend far beyond the educational-research community and into the very fabric of what is presented by some as a modern(ist) world bent on reducing the social world to a technology whose functioning can be rationally managed, controlled, and (the key to all of this) measured (Smith, 2007) and by others as a postmodern world, fractured, a-rational, defeating (happily perhaps) centralized, technocratic management, a world in which notions of truth and certainty are replaced by a mixture of social relativism, nihilism, fancifulness, irony – or none of these because it refuses to be captured in any of these ways (Stronach and MacLure, 1997; Peters, 2004).

If not all researchers (nor even all philosophers) identify fully with either of these theoretical camps, this wider argument nevertheless adds an intellectual and political frisson to more modest debates – all of which are similarly philosophically rooted. To take what must surely be a central concept in any educational research agenda: “What learning is taking place in this setting?” Drawing on the contemporary range of educational research we might find at least the following illustrative range of responses.

1. An old style behaviorist psychologist interprets learning as a relatively stable change in behavior. (S)he needs to do some systematic observation of the pupils’ behavior and probably conduct a test devised to quantify the learning which has taken place.
2. A neuroscientist (and neuroscience is an increasingly influential source of educational research) will need some more sophisticated observational techniques, involving technology which will enable him or her to observe (and again probably measure) changes in the structure or activity of the brain.
3. A phenomenologist will probably want to conduct an extended interview with a selection of the pupils, exploring their experience of the lessons in questions and their own self-reporting on what they felt they may or may not have been learning, what they benefited from and why.

There may be several different reasons why someone might prefer in general to adopt one of these approaches over another. In part the preference will depend on what one means by learning, though an answer to this question will almost certainly depend on how we think of human beings: Are they reducible to their behavior? Do we want to get rid of what Gilbert Ryle referred to as the ghost in the machine (Ryle, 1973)? What is the relationship between activity of the brain and activity of the mind? How important is it to think of and to relate to human beings as experiencing beings and to understand the world through their eyes? Beyond these fundamental philosophical questions, there are some more practical ones: How useful is it to whom to know about, for example, electrical signals in the brain as compared with a child's own perception of his or her own learning? What can a teacher do with such knowledge?

Objectivity and Subjectivity in Educational Research

The contrasting styles of enquiry illustrated above also raise issues which are sometimes expressed in terms of objectivity and subjectivity (although again this dichotomy is probably too crude – at very least, we need to add the notion of intersubjectivity, of the social character of the construction of knowledge, to the spectrum). This distinction operates at two levels at least. First, we might apply it to the objects of our enquiry, so that the behaviorist and the neuroscientist might claim to be observing the objective world of human behavior and of brain activity, respectively. By contrast, the phenomenologist might be said to be investigating the subjective world of human experience. The distinction gets blurred when, for example, we attempt to describe human behavior, which is almost impossible without reference to the individual and social significance which human beings attach to different forms of behavior (Is the child who holds two fingers up to his teacher indicating a numerical quantity or something rather ruder?) and to human intentionality, both of which are embedded in the kind of perceptions which the phenomenologist seeks to describe and interpret. However, the objectivity/subjectivity distinction continues to frame a lot of educational debate and again we are dealing with matters which are essentially philosophical in character.

The language of objectivity and subjectivity also serves to distinguish the aspirations of different educational researchers and the way in which they deal with the presentation of their research to a readership. At one end of the spectrum, researchers seek to go as far as they can to eliminate through their methodology and their style of presentation the effect, the visibility, and indeed, the very presence of the researcher. This is indeed one of the

aspirations of the double-blind controlled experiment which has been referred to already. Such research goes to great lengths to reduce any impact which the presence of the researcher might have on the research site or participants; it seeks objectivity in the reporting and interpreting of the data; the hand of human authorship is concealed by stylistic devices which replace the first person (I did this and then I observed that . . .) by the third person (The researcher used a standardized test . . .), or employs the passive voice (the students were observed) rather than the active voice (I observed the students . . .).

Elsewhere on the spectrum are advocates of the view that this search after objectivity is ultimately doomed to failure: subjectivity will always enter into educational research, perhaps through the definition of the research agenda or the research questions; through the selection of research methods and methodology; subtly through what is not asked or noted as well as what is, in the collection of data; in the observation; in the analysis; and in the reporting. We deceive ourselves, it is claimed, if we imagine that we can escape such subjectivity. It is “a garment which cannot be removed” (Peshkin, 1988: 7).

This skepticism with respect to the possibility of objectivity is closely linked to a second opinion that, rather than trying to eliminate subjectivity it is better way to enable a reader of research to take account of such subjectivity by providing sufficient information about the researcher – his or her background, interests, ideological attachments, etc. (his or her biographical positioning as it is sometimes called) – and thus enable the reader to take this into account in evaluating the import of the research (see, e.g., Atkinson, 2000). This still, in a sense, represents a search after a kind of objectivity, a version of what things are really like, lying behind what is acknowledged to be one human rendering of this reality.

A third and more radical position suggests that even this kind of access to reality is illusory. All we have are people's subjective experience of the social world, people's perceptions, and people's narratives. These can enlarge our imaginative understanding of possible ways of seeing the world, and some may appeal to us more than others for a variety of reasons, but none can command special authority.

The concern here is not to offer a conclusion to a debate which continues to run through the education research community, but to point out the nature of the argument. These and other related questions about the nature of our knowledge of the social world, about the possibility of and limitations on such knowledge, about the sense in which we can (or cannot) talk about truth and falsity with respect to the claims issuing from educational research – all of these are the very stuff of philosophical writing over two millennia. The debates which rang through Socrates' chambers and the halls of medieval disputations between nominalists and realists are as alive today as they ever were. The passing

centuries may sophisticate the terms in which they are constructed, but they do not seem to resolve them in a way that conclusively sets them aside.

In this context are illustrated only some of the issues in the methodology of educational research (and in this section the focus is on epistemological questions) which invite and have received substantial philosophical attention. There are many more. Indeed lying behind any disciplined form of enquiry whose resources are drawn into educational research, there is a body of literature which relates to the philosophy of this discipline and which addresses issues to do with, for example, the kind of confidence which might be attached to findings generated by that particular form of enquiry, the extent to which one might generalize from such findings, or the extent to which one might confidently apply these findings in a single particular setting. Without making some attempt to engage with these philosophical questions in any research setting, one cannot confidently know how to interpret the research which one reads.

The Ethics and Politics of Educational Research

Educational researchers have become increasingly aware over recent decades of the ethical and, more widely, social and political significance of their interaction with research participants (some of who are of course children), research users, and research sponsors (who include powerful government and corporate organizations). This has been reflected in the establishment within professional research organizations like the American Educational Research Association and the British Educational Research Association of ethical codes to which their membership is expected to subscribe. Individual higher education and research institutions have similarly adopted ethical codes governing the conduct of research with human participants and ethics committees tasked with overseeing conformity with these codes.

Of course, this process has required considerable debate about what are the duties and rights of researchers, of participants in the research, and of sponsors of the research. Among the central issues, perhaps are the following:

- *The issue of informed consent.* To what extent does educational research require the informed consent of those who are going to participate in it? Is covert research excluded? Is the answer the same if one is doing research with vulnerable people or in the setting of powerful government or corporate organizations? Who needs to give consent? If I want to do research in a classroom, is it sufficient to have the agreement of the head teacher or do I also need the agreement of the teacher? The children? The children's parents?

- *The issue of confidentiality.* To what extent can and should the researcher guarantee confidentiality to an institution or individual involved in the research? How does one balance considerations of confidentiality with the public right to know and indeed with participants' rights to acknowledgment of their contribution?
- *The issue of control over data and its interpretation.* To what extent should researchers allow participants to, for example, amend a transcript of a recorded interview or to control the way in which their contribution is interpreted in a research report? Should researchers pay for contributions to their research? Is research a form of theft of other people's knowledge?

The issues indicated thus far generally assume a situation in which participants' rights need to be protected from an unscrupulous researcher. There are other issues framed by situations in which the researcher's own honest enquiry is threatened by the power of controlling organizations – especially, perhaps, where these organizations are paying for the research:

- The issue of the level of control and censorship which those paying for research can exercise over what is researched, how it is researched, and how and where it is reported – and hence ...
- The issue of the responsibility of higher education institutions with respect to the kinds of contract which they enter into and the support they give to their researchers.

There are two things to note, in particular, about these issues.

First, it is easy to see how what at first might be thought of as more narrowly ethical issues carry important social and political significance. Research is about the creation and redistribution of knowledge, which is itself an act with political consequences insofar as it involves a redistribution of power. Such a redistribution has even weightier consequences when, as in the field of education, you are dealing with a social practice which itself carries such a high level of public and political investment and which is itself a matter of intense political controversy. So, beyond the narrower construction of an ethics of educational research, we have to engage with the political principles which should govern such research in a democratic setting. Considerations of social justice and the requirement for informed public opinion jostle with each other and also with considerations about, for example, the right to privacy or confidentiality or at least some kind of credit for the knowledge that participants are sharing (see Griffiths, 1998; Smith, 1999).

Second, although all or most of these issues are addressed in the sort of ethical codes referred to, this does not mean that they are now settled. Many remain controversial. Others, which might appear to be agreed at the level of general principle, turn out to be much more complex when one has to operationalize them in the field – and this

experience has led to a flurry of debate around the situatedness of ethical judgment and the relationship between general principles, embodied perhaps in ethical codes and the resolution of issues in the field.

Both of these sets of considerations drive educational researchers into the philosophical territory, which is always close to the surface in ethical argument. How are we properly to understand these ethical claims? By reference to what principles can we resolve conflicts between these ethical principles? What is the relationship between general principle and situated judgment? All of this is well-worn ground in philosophy and an essential resource for educational researchers seeking to engage seriously and systematically with the sort of issues described.

Conclusion

These last two sections describe some examples of the sort of issues which arise out of the endeavor of educational research. It has been argued that these are inescapably philosophical in character and that they are indeed the sorts of questions which philosophers have engaged with and continue to engage with in their academic work. Educational researchers, like any other citizens can, of course, lead their lives without any attention to the literature which engages in a serious, sustained, and systematic way with these issues, but at enormous cost. After all, are not researchers' particular claims to authority and their claims on public attention based precisely on the expectation that their enquiry is more serious, sustained, and systematic than that which ordinary citizens or even journalists can normally afford? If philosophical questions are, then, central to the research enterprise, then these too must receive the same rigorous attention as is given to the gathering and analysis of empirical data.

However, this argument does not only place an onus of responsibility on educational researchers, but it also points to the importance of philosophers of education engaging in a practical way with the work and life of the wider educational research community – as not only researchers in their own right (as indicated in the first section above) but also as co-workers engaged in a continual conversation about the meaning, justification, and right conduct of the research.

See also: Philosophy of Education: Overview.

Bibliography

- Atkinson, E. (2000). Behind the enquiring mind: Exploring the transition from external to internal inquiry. *Reflective Practice* 1(2), 149–164.
- Bridges, D., Smeyers, P., and Smith, R. D. (2009). *Evidence based Educational Policy. What Evidence? What Basis? Whose Policy?* Oxford: Blackwell.

- Carr, W. (1986). Theories of theory and practice. *Journal of Philosophy of Education* 20(2), 177–186.
- Eisner, E. (1993). Forms of understanding and the future of educational research. *Educational Researcher* 22(7), 5–11.
- Elliott, J. (2001). Doing action research – Doing practical philosophy. *Prospero* 6(3/4), 82–100.
- Griffiths, M. (1998). *Educational Research for Social Justice: Getting off the Fence*. Buckingham: Open University Press.
- Hammersley, M. (1993). *Educational Research (Volume One): Current Issues*. London: Paul Chapman.
- Hammersley, M. (2007). *Education Research and Evidenced Based Practice*. London: Sage.
- Peshkin, A. (1988). In search of subjectivity – One's own. *Educational Researcher* 17(7), 17–21.
- Peters, M. (2004). *Derrida, Deconstruction and Education: Ethics of Pedagogy and Research*. Oxford: Blackwell.
- Phillips, D. C. (2005). The contested nature of empirical educational research (and why philosophy of education offers little help). *Journal of Philosophy of Education* 39(4), 577–597.
- Polanyi, M. (1966). *The Tacit Dimension*. New York: Doubleday.
- Ryle, G. (1973). *The Concept of Mind*. London: Penguin Books.
- Smith, L. T. (1999). *Decolonising Methodologies: Research and Indigenous Peoples*. London: Zed Books.
- Smith, R. D. (2007). As if by machinery: The leveling of educational research. In Bridges, D. and Smith, R. D. (eds.) *Philosophy, Methodology and Educational Research*, pp 31–42. Oxford: Blackwell.
- Stronach, I. and MacLure, M. (1997). *Educational Research Undone: The Postmodern Embrace*. Buckingham: Open University Press.

Further Reading

- Bibby, M. (1997). Using a code of research ethics. *Educational Philosophy and Theory* 15(1), 49–64.
- Bridges, D. (2003). 'Fiction Written under Oath?' *Essays in Philosophy and Educational Research*. Dordrecht: Kluwer.
- Bridges, D. and Smith, R. D. (2007). *Philosophy and Methodology of Educational Research*. Oxford: Blackwell.
- Burbules, N. (1998). Principle and process in the ethics of educational research, reply to Robin Small. *Australian Journal of Education* 42(1), 116–123.
- Hammersley, M. (2002). *Educational Research, Policy Making and Practice*. London: Sage.
- Homan, R. (1991). *The Ethics of Social Research*. Harlow: Longman.
- Howe, K. R. (2003). *Closing Methodological Divides: Towards Democratic Educational Research*. Dordrecht: Kluwer.
- McNamee, M. and Bridges, D. (eds.) (2002). *The Ethics of Educational Research*. London: Blackwell.
- Paul, J. (2004). *Introduction to the Philosophies of Research and Criticism in Education*. New York: Prentice-Hall.
- Pring, R. (2004). *Philosophy of Educational Research*. London: Continuum.
- Sikes, P., Nixon, J., and Carr, W. (eds.) (2003). *The Moral Foundations of Educational Research: Knowledge Enquiry and Values*. Maidenhead/Philadelphia, PA: Open University Press/McGraw-Hill.
- Small, R. (2001). Codes are not enough: What philosophy can contribute to the ethics of educational research. *Journal of Philosophy of Education* 35(3), 345–360.

Relevant Websites

- <http://www.aera.net> – American Educational Research Association, Ethical Standards.
- <http://www.aera.net> – American Educational Research Association, Journal of the American Educational Research Association, Educational Researcher.

<http://www.bera.ac.uk> – British Educational Research Association, Publications and Guidelines; Resources.
<http://eppi.ioe.ac.uk> – EPPI-Centre.
<http://ies.ed.gov> – Institute of Educational Sciences, What Works Clearing House.

<http://www.tlrp.org> – Teaching and Learning Research Programme, Capacity Building: Developing Researcher Expertise.
<http://www.philosophyofeducation.org> and <http://www.philosophy-of-education.org> – Resources and links Google Scholar links to a number of useful sources for this topic.

Race, Critical Race Theory and Whiteness

B Applebaum, Syracuse University, Syracuse, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

While contemporary science confirms that our practices of racial categorization do not correspond to factual biological classifications, racism (and other forms of oppression) continues to play out in schools, educational policy, and educational research in myriad covert (but sometimes still overt) ways. Even after racial segregation has been outlawed (as in the United States' 1954 landmark case *Brown vs. Board of Education of Topeka, Kansas*), savage inequalities (Kozol, 1991) continue to endure in educational systems across the globe. Educational reforms based on ignoring race, on the one hand, and celebrating diversity, on the other, ostensibly implemented to remedy racial inequity in education, are widely criticized not only as ineffective but also as counterproductive in that they perpetuate unjust structures by inoculating the dominant order from any serious challenge. Ladson-Billings and Tate (1995) maintain that educational reforms that attempt to remedy racial inequity get sucked back into the system (p. 62) partly because the conceptual assumptions upon which such reforms are grounded fail to acknowledge that race and racism cannot be adequately assessed without an analysis of power. In order to understand how race functions in the classroom, in schools, and in educational systems and, furthermore, to appreciate what these critics of seemingly well-intended educational initiatives are pointing to, a critical theory lens of race and racism is crucial.

Critical theories of race and racism explore concepts such as race and racism from the framework of systemic oppression and the normalization of whiteness in order to imagine new possibilities in education that challenge rather than reproduce power systems. Critical theories of race and racism are to be distinguished from critical race theory (CRT). The latter is a movement developed by primarily Black, Latino/a, and Asian American legal scholars in the 1990s and advances specific aims. Space limitations preclude an examination of CRT and how it has been applied to education. For some excellent resources, the reader is directed to Ladson-Billings and Tate (1995), Dixon and Rousseau (2006), and Delgado and Stefancic (2001).

The former term, critical theories of race and racism, is more of an umbrella term that includes a broad constellation of scholarship that explores race and racism through the prism of power systems. In what follows, the concepts of race, racism, and whiteness will be addressed from a critical theory of race and racism perspective.

Race as a Social Construction

Although for many years, scientists endeavored to prove that there were innate biological differences related to people of different races, there is much consensus among scholars in such fields as evolutionary biology, anthropology, and other disciplines that biological arguments for racial difference fail to be supported by scientific evidence. Not only is there more genetic variation within than between the so-called races (Lewontin, 1972; also see Appiah (1992) for an excellent review of the scholarship), but also what little differences that have been established between such supposed groups do not reflect the social meanings that are given to racial differences. To be explicit, there is no scientific evidence that one race is superior or inferior to another. The social construction of race also receives support from the fact that the meaning of race is relative to geographical location and historical period. In the United States, the one-drop rule in which having the most minuscule black ancestry designates one as black regardless of physical appearance can result in someone who is defined as black in the United States yet also be regarded as white in other parts of the globe. That race is a fluid and contested notion also gains support with Ignatiev's (1995) historical study that charts the process by which the Irish, raced as nonwhite when they first settled in the United States in the first half of the nineteenth century, became white by embracing racism in the years surrounding the Civil War.

The concept of race is often conflated with the concept of ethnicity. Those who do research from the ethnicity paradigm often understand ethnicity as a purely descriptive concept and ignore how the white ethnic experience often becomes the normative benchmark from which all other ethnic experiences are understood. Charles Gallagher (2003) argues that the ethnicity paradigm functions to support the myth of meritocracy and encourages blaming the victim explanations for social inequality that exonerate whites and allow them to evade any consideration as to how they might be complicit in perpetuating systemic racism. Although race and ethnicity can overlap and while ethnicities often become racialized, race is conceptually a very different type of difference. While ethnicity describes "that aspect of our heritage that provides us with a mother tongue and that shapes our values, our worldview, our family structure, our rituals, the foods we eat" (Dalton, 1995: 107), races are distinctively constructed to exist only in relation to one another. Whiteness is devoid of

meaning without nonwhiteness constituting its boundaries. Blurring the distinctions between the concepts of race and ethnicity obscures how race is a “fault line along which power, prestige, and respect are distributed” (Dalton, 1995: 108). Educators who use race and ethnicity interchangeably, “confuse the positive role of culture . . . with the negative impact of racism” (Derman-Sparks and Phillips, 1997: 13).

The observation that race does not point to a fixed biological or genetic reality but rather is a socially and historically constructed concept whose meaning shifts through time to serve dominant group interests is now widely accepted in most areas of the social sciences (Lopez, 2000; McCarthy *et al.*, 1993). The aforementioned one-drop rule served to ensure that the offspring of a black slave and a white slave owner would be a slave, not free. The rule justified the raping of black slave women by white slave owners who could perceive this act of violence as a profit-making enterprise because it increased the population of black slaves (Payne, 1998). Omi and Winant (1997) explain that groups become racialized when racial meaning is extended to a “previously unclassified relationship, social practice or group” (p. 64). They coined the term racial formation to describe the reciprocal, sociohistorical process “by which social, economic, and political forces determine the content and importance of racial categories, and by which they are in turn shaped by racial meanings” (p. 61). In the United States, racial systems functioned to justify the institution of slavery.

Once these meanings are constructed, social institutions such as law, media, and education continually reify them and mask their origins. In 1977, Susie Guillory Phipps who identified as white but was designated as black on her birth certificate because her great-great-great-grandmother was a black slave in the South (Omi and Winant, 1994) sued to have her ascription as black changed to white. Yet the courts insisted she was black even though a 100% white racial purity has always been a fantasy. The law is one institution that helps to ensure that the racial system is maintained (Lopez, 2000). Education grounded as it is in the myth of meritocracy is another institution that contributes to the reproduction of a racial system in which those who benefit from the racial system as well as those constrained by it come to see it as real. Omi and Winant maintain that there is a racial etiquette in which racial ways of understanding the social world become common sense and “everyone learns some combination, some version, of the rules of racial classification . . . often without obvious teaching or conscious inculcation” (p. 62).

An essential component of understanding race as a social construction, then, is the acknowledgment of the political and ideological institutions, such as education, that maintain the system that makes race matter. Race might not be real, yet a system of privilege and oppression has been built up around the concept of race with real

material, social, economic, and psychological consequences for groups of people. Furthermore, to say that race is socially constructed is to notice that social groups, not individuals, must be the level of analysis through which race and racism are best understood. It is social groups that come into being through racialization. Peter McLaren (2000) articulates this construction thusly,

People do not discriminate against groups because they are different; rather, *the act of discrimination itself* constructs categories of difference that hierarchically locate people as ‘superior’ or ‘inferior’ and then universalizes and naturalizes these differences. (p. 148, italics added)

It is not that groups do not exist prior to language but rather that the meanings we attribute to social groups depend on discourse. Sally Haslanger (1995) refers to the discursive construction in which social groups come to have certain meanings through language. To say that race is socially constructed, however, does not merely refer to the discursive origins that constitute the meanings of social groups but also that race marks social differences that involve the relative location of groups within a system of social relations in which one group is considered the norm and other groups are evaluated by that norm (Minow, 1990).

In her discussion of the dilemma of difference (i.e., “... when does treating people differently emphasize their differences and stigmatize or hinder them on that basis? And when does treating people the same become insensitive to their difference and likely to stigmatize or hinder them on that basis?” (Minow, 1990: 20), Martha Minow analyzes the normative assumptions that constitute difference and draws attention to social structure as the source of the problem rather than the individual. Minow discusses five assumptions that remain unstated but work in powerful ways to define difference. One of these assumptions is that the norm from which difference is demarcated can be taken for granted and does not need to be made explicit. According to Minow, as long as the norm is not made explicit, how and when difference is stigmatized will not be acknowledged.

Individual people are constituted as raced based on their ascribed membership in social groups but also internalize race as identity and their lives are affected accordingly. The white norm that remains unstated in racist systems will be subsequently addressed. If race is a social construction, one must ask, as Omowale Akintunde (1999: 4) urges educational reformers to ask, “why are we still operating in such a system and to whose benefit?”

Definitions of Racism

Race as a social construction presumes that there is a system of privilege and oppression through which racial

social groups are constructed and that individuals experience oppression not as individuals but as members of an oppressed group. Yet, many people view racism strictly as a matter of prejudice and animus. To view racism solely as prejudice and/or discrimination based on race is to ignore the racist system and how white people have systemically benefiting from the system. According to Akintunde (1999), when racism is understood as something that one consciously chooses and enacts then racism can be practiced by anyone. “That is ‘non-Whites’, too, may engage in practicing racism and thus Whites themselves may be victims of racism.” (Akintunde, 1999: 2) Akintunde contends that by narrowly restricting racism to overt, intentional acts or conscious beliefs, those who benefit from the system not only ensure “that the system of White supremacy remains intact but can, in fact, successfully create smoke screens that actually implicate ‘others’ in the maintenance of such a system” (p. 2).

Similarly, in his discussion of antidiscrimination law, Alan David Freeman (1996) explains that racial discrimination can be understood from the victim’s perspective or from the perpetrator’s perspective. While the latter focuses on intent and finding who did what to whom, the former emphasizes the actual consequences including the objective conditions of life of being a member of a marginalized group. Understanding racism from the perpetrator’s perspective alone results in a strong reliance on fault and intention and, thus, serves to draw attention away from the pattern of conditions that a victim perceives to be associated with discrimination. The upshot is that if one is not the direct cause of a racist act, then one is not at fault and bears no personal responsibility for the act.

In her analysis of the concept of oppression, Marilyn Frye (1992) offers a useful metaphor to highlight these patterns of conditions and to distinguish oppression from personal suffering that anyone can experience. Frye suggests that we imagine a birdcage. The experience of oppressed people, Frye maintains, “is that the living of one’s life is confined and shaped by forces and barriers which are not accidental or occasional and hence avoidable, but are systematically related to each other in such a way as to catch one between and among them and restrict or penalize motion in any direction” (p. 4). These objective conditions are like living in a birdcage. If one were to examine the wires of the birdcage one at a time, one might not comprehend how that one wire is restrictive. It would seem that the bird could just fly around the isolated barrier to freedom. Yet, when one steps back and is able to apprehend the entire cage that surrounds the bird, “it is perfectly obvious that the bird is surrounded by a network of systematically related barriers, no one of which would be the least hindrance to its flight, but which, by their relations to each other, are as confining as the solid walls of a dungeon” (p. 5). Oppression, therefore, involves a pattern of barriers that restrict the movement of those

inside the cage. Moreover, individuals are oppressed not because they are individuals but by virtue of their membership in oppressed social groups.

As a system, racism can be understood to work on three mutually supporting levels: individual, cultural, and institutional. The individual level refers to personal attitudes and behaviors that reproduce racial power differentials. Blatant illustrations are racial epithets and jokes directed at people of color and racial discrimination in hiring practices. Personal racism, however, can work in more covert and subtle ways as will be subsequently described in the section on white privilege. Cultural racism refers to the beliefs, symbols, and ideas represented in the media, pop culture, books, standards of beauty, etc. that construct whiteness as normal, neutral, and superior while simultaneously conveying a negative message about all those who are not ascribed whiteness. It is not just a matter of inclusion of negative stereotypes because exclusion stigmatize, as well. Cultural racism and personal racism mutually reinforce each other.

Finally, institutional racism involves the network of structures and policies that sustain these structures in ways that benefit some groups and disadvantage others. The justice system, the education system, employment, housing, banks, government, and healthcare are only some examples of institutions through which racism is maintained and constantly reproduced (Schmidt, 2005). Individuals who participate in these institutions are complicit in the perpetuation of the *status quo* often without their conscious intent or volition.

If racism is a form of oppression that is systematic and involves interlocking barriers that constrain and limit the life outcomes of members of certain social groups, can the victims of racial oppression be racist? Beverly Daniel Tatum (1997) explains that while any person can have racial prejudices and engage in discriminatory actions, she rejects the use of the term racist to describe victims of racism in order to underscore the continuing power disparities between groups and how the system of oppression sustains white supremacy and systemically benefits all white people. Tatum emphasizes, “People of color are not racist because they do not systematically benefit from racism. And equally important there is not systemic, cultural and institutional support or sanction for the racial bigotry of people of color” (p. 10). Individuals, however, belong to many social groups and thus it is possible to be both oppressed and an oppressor. In order to determine when oppression is taking place, Marilyn Frye (1983) suggests looking at the barrier and asking “Who constructs and maintains it? Whose interests are served by its existence? Is it part of a structure which tends to confine, reduce and immobilize some group? Is the individual a member of the confined group?” (p. 14).

Acknowledging racism as a form of systemic oppression that limits individuals qua members of marginalized

groups is to recognize that there is also a group who benefits from the birdcage being in place. Before addressing the issue of whiteness and white privilege, some complex questions about identity must be raised.

Identity Politics, Recognition, and Essentialist Notions of Race

Multicultural education, so prominent in Western educational systems, rests upon the categorization of individuals into groups, specifically focusing on the groups that have been excluded from mainstream society. In this sense, multicultural education and identity politics have many of the same presumptions. In the latter part of the twentieth century in the United States, political movements arose around the acknowledgment that the experience of injustice was systemic and was connected to one's being identified with an oppressed group. Identity became an important tool in political discourse as the phrase identity politics indicates. Common or shared experience was the unifying force that propelled such movements. There is also the assumption that some type of authentic experience connected to group identity has been suppressed but deserves recognition and respect (Taylor, 1994) and must be reclaimed. What made identity politics so significant, Sonia Kruks (2001) explains, is that it involves a demand for recognition "on the basis of the very groups on which recognition has previously been denied: it is *qua* women, *qua* blacks, *qua* lesbians that groups demand recognition" (p. 85).

Although scientific essentialism has been discredited, from the phenomenon of identity politics it is easy to see how essentialism can arise even when race is understood to be a social construction. Essentialism involves the assumption that there is a uniquely distinguishing feature or features that all in the group share. Often, there is an assumption of a commonality of experience that constitutes and unites the group (e.g., the black experience, and the Chicano experience) that can be described independent of aspects of the particular person's life (Harris, 1990; Crenshaw, 1991). While identity politics was useful in empowering marginalized groups, at the same time the approach attracts much criticism.

Identity politics is accused of assuming essentialist notions of identity. The problems of such essentialism involve the normative consequences of commonality (Spelman, 1988). Who gets excluded any time a 'we' is constructed for the purposes of consensus and unity? In the women's movement, what it means to be a woman has been defined on the basis upon the experiences of white, middle-class, able-bodied, heterosexual women resulting in the exclusion and silencing of women of color as well as lesbian and queer women. In his 1994 film,

Black Is . . . Black Ain't, Marlon Riggs powerfully expressed how black women and black gays and lesbians have been excluded from heterosexist and patriarchal conceptions of African-American identity. In his discussion, Chicano nationalist movement, Jose-Antonio Orosco (2002) highlights how Chicano identity has been defined from a male, heterosexist point of view. He demonstrates how a focus on authentic identity does so at the expense of silencing voices that might contest the gendered and heteronormative conception of Chicano identity. In addition, the appeal to commonality of experience can operate as a disciplinary norm within a group that imposes a particular vision of identity onto members of that group. Essentialism assumes that the lives of people who experience multiple forms of oppression can be reduced to an additive model:

straight black women's experience = the experience of racism
+ the experience of sexism;

black lesbian's experience = the experience of racism
+ the experience of sexism + the experience of homophobia

Yet, for instance, the experience of women of color can be erased because they stand at the intersections of different categories. In many cases when black women face discrimination, their experience is not taken seriously because neither white women nor black men experience discrimination in this way.

Perhaps the most radical extension of the critique of identity politics and essentialism has evolved from post-structuralists such as Judith Butler (1990) and Wendy Brown (1995). Both have charged identity politics with assuming a prediscursive self over which forms of socialization impose nonessential characteristics. According to Butler, however, the subject is always already a product of discourse. Any reclamation of one side of binary categories, even a reclaiming of marginal identities, is to reify the framework from within which those categories are constituted. Wendy Brown, for example, contends that identity politics is based on a reclamation of wounded attachments that sustains victimhood continuing to carry marks of subordination instead of challenging the system from which these attachments originate and which they support. Moreover, such politics place subjects in fixed and stable categories that constrain their freedom. Instead of appealing to any type of identity category, both Butler and Brown recommend that all subject positions be decentered and given identity categories be subverted in ways that destabilize the normalizing discursive forces that create those identities.

One of the problems with the post-structuralist position is the question of agency. How is politics possible without identity categories? In addition, reclamations of racial identity can be political and serve as a way to expose

the systematicity of oppression (Taylor, 2000; Alcoff, 2005; West, 1996). According to Paula Moya (2002), some identities have epistemic status and political salience and can be politically progressive “not because they are ‘transgressive’ or ‘indeterminate’ but because they provide us with a critical perspective from which we can disclose the complicated workings of ideology and oppression” (p. 17). What is of particular interest about Moya’s account is that she acknowledges that the simple fact of being born a person of color in the United States does not in itself give one such epistemic privilege. What this claim underscores is that some people have experiences that other people who are not oppressed do not have. In fact, the systemically privileged have other experiences that protect them from even acknowledging the birdcage. When experiences of oppression are politicized, however, they have the possibility to provide information about how power systems operate and are maintained. Political identities, according to Moya, should not be jettisoned but reclaimed.

Whether or not racial identity categories should be reclaimed or eliminated has been an ongoing debate whose complex details are beyond the scope of this overview. It is clear that these debates raise crucial questions about all educational reforms that are based on social group categories. On the one hand, while multicultural education has been vigorously critiqued (Nieto, 1995), the debates around identity categories highlight the need to be vigilant about reifying categories that sustain frameworks of power and privilege rather than challenge them. However, on the other hand, there are dangers in also abandoning racial discourse. To further flesh this out, whiteness and white privilege must be addressed.

Whiteness, White Privilege, and Critical Whiteness Studies

While for generations scholars of color have argued that whiteness lies at the center of the problem of racism and that whiteness is a mutually constitutive aspect of the social construction of race, it is only relatively recent that whiteness within academe has been acknowledged to be a necessary corrective to the study of race that has been exclusively focused on the racialized other. Critical whiteness studies is focused on exposing whiteness as a taken-for-granted norm from which difference is constructed and seeks to analyze whiteness as a determinant of social power. Whiteness is intimately related to the construction of race in that its own meaning is dependent on the process of negation of what is outside its borders. The center and periphery are mutually constituent. As Frankenberg (1996) puts it, “Whiteness comes to self-name . . . simply through a triumphant ‘I am not that’” (p. 7). Quoting from Hazel Carby, Dyer (2000) argues that it is important to study whiteness, to “make visible what is rendered invisible

when viewed as the normative state of existence” (Carby as cited in Dyer, p. 3) in order to dislodge whiteness from its position of dominance. It is impossible then to gain an understanding of systemic racism without understanding how whiteness works and Dyer claims that whiteness, because it is presumed neutral and normal, can only be studied by making it strange (p. 10).

While the definition of whiteness is difficult to pin down, there is widespread agreement that whiteness is a socially constructed category that is normalized within a system of privilege so that it is taken for granted by those who benefit from it (Rodriguez, 1998; Winant, 1997; Roediger, 1991). Cheryl Harris (1993) suggests that whiteness is best understood as a form of property rights that is systemically protected by institutions such as law. Toni Morrison (1993) uses the metaphor of a fishbowl containing fish and water to elucidate the invisibility of whiteness as the condition from which meaning is made. If we focus on the water and the fish, we do not have to see how the fishbowl itself frames where and what happens within it.

... it is as if I had been looking at a fishbowl – the glide and flash of the golden scales . . . and then I saw the bowl, the structure that transparently (and invisibly) permits the ordered life it contains to exist in the larger world. (Morrison, 1993: 17)

White people’s investment in whiteness can obscure how white people even with the best of intentions are complicit in sustaining a racially unjust system.

Much of the study of whiteness focuses on exposing white privilege. White privilege, as McIntosh (1997) writes, is “an invisible package of unearned assets which I can count on cashing in each day, but about which I was ‘meant’ to remain oblivious. White privilege is like an invisible weightless knapsack of special provisions, assurances, tools, maps, guides, codebooks, passports, visas, clothes, . . .” (p. 291). Acknowledging white privilege challenges the ideology of meritocracy grounding many institutions of education in that it is privilege not only merit that results in white people getting ahead in life. The invisibility of white privilege works in conjunction with the persistent belief in meritocracy, that is, anyone can succeed with hard work and adherence to the rules of society, to constitute those who are Othered by ignoring the fact that systems based on white norms privilege white people. Disparities between the experiences of white people and people of color can be perceived as natural and normal, in other words, the fault of the individual and not the result of systematic privileging of the dominant. The marginalization of people of color, therefore, is held in place by white privilege.

White privilege must be understood as relational and also systemic (Leonardo, 2004).

The ideology of colorblindness (that a race-neutral stance is a remedy for eradicating racism) also plays into

the logic of white privilege – if one does not have to consider one's race, one does not have to consider how white norms operate. Policies and practices based on ignoring race that purport to eradicate racism actually sustain it (Lewis, 2001; Schofield, 1997). One of the privileges of whiteness is the ability to be able to ignore in the sense of remaining ignorant, denying, and dismissing the reality experienced by people of color on a daily basis. In schools, the ideology of ignoring color results in teachers and administrators disregarding the relationship between race and tracking, biased testing, discipline etc ... and protects white people from examining how white privilege hides behind normalized whiteness. Ignoring race gives its advocates the sense that they are adhering to the moral demand for egalitarianism yet in actuality this approach leaves white racial privilege and racial disparities intact and secure. White privilege is systemic in that it is connected with a network of privileges both material and psychological. It is also manifest as a way of being in the world. Marilyn Frye (1992) discusses how being whitely exhibits itself in how white people relate to others, their sense of self and their own morality.

White teachers, for example, need to study whiteness and white privilege so that they can understand how to challenge racism as it is evidenced in their classroom. (For an excellent illustration of how white dominance codes nonwhite students as behavior problems see Delpit, 1995.) But the different discursive disguises that white privilege takes on complicate this. Alice McIntyre (1997) describes white talk as “talk that serves to insulate white people from examining their/our individual and collective role(s) in the perpetuation of racism” (p. 45) Examples of white talk that McIntyre describes are “derailing the conversation, evading questions, dismissing counterarguments, withdrawing from the discussion, remaining silent, interrupting speakers and topics, and colluding with each other in creating a ‘culture of niceness’ that made it very difficult to ‘read the white world’” (p. 46). Hytten and Warren (2003) examine how whiteness gets reinscribed in classroom discursive dynamics. In both studies, the white students were trying to be good anti-racist whites yet unintentionally employed discourse that protects whiteness as the privileged center. Whiteness, therefore, can be subtly reinscribed even when it does not appear to be doing so and can be reified even when “we try to disrupt its normative influence” (Hytten and Warren, p. 65). Such white discursive moves complicate calls for dialog and are necessary to explain why separation rather than integration is often preferred by students of color (see Jones, 1999) and expose the need for a politics of listening alongside the study of whiteness and white privilege.

Critical whiteness studies have attracted much criticism. There is a serious concern with recentring whiteness, white people and white feelings and drawing attention away from the needs and interests of people of color.

In his foreward to the book, *White Reign: Deploying Whiteness in America* Michael Apple (1998) cautions,

Having Whites focus on whiteness can have contradictory effects, one of which we need to be well aware. It can enable people to acknowledge differential power and the race nature of everyone. . . . It can just as easily run the risk of lapsing into the possessive individualism that is so powerful in this society. That is, such a process can serve the chilling function of simply saying ‘but enough about you, let me tell you about me’. (p. xi)

If focused on white self-transformation, whiteness studies can result in white people being concerned about salvaging their good moral being. Dyer notes about white people, “the display of our guilt is our cavalry” (p. 11). Whiteness studies can itself become an exercise in white privilege. If critical whiteness studies are focused on white self-transformation then it fails to challenge whiteness as a norm and instead recenters whiteness, white people, and their feelings, and well-being.

An additional concern involves how categories of race are static in much of the whiteness research (Keating, 1995). On the one hand, if interrogations of whiteness require static concepts of identity, the very framework of us/them that from which racial systems are constructed are reified. On the other hand, without marking whiteness how can it be interrogated? Two approaches to the disruption of whiteness have been discussed. One approach advocates that whiteness be rearticulated while the other contends that whiteness be abolished (Giroux, 1997; Ignatiev and Garvey, 1996). A detailed analysis of both these positions is beyond the scope of this overview. The interested reader is directed to Moon and Flores (2000), Flores and Moon (2000), Warren (2001), and MacMullan (2005).

Much of the discussion in this overview has been dominated by a US focus that may or may not be usefully employed in other national contexts. What lessons can be gleaned will depend on how similar and how different race and racism plays out in other contexts. Already there is a growing body of scholarship that is responsive to the racial dynamics in the United States, Canada, United Kingdom, Australia, and New Zealand. In all these contexts, whiteness is always the invisible norm from which the other is constructed. Continued research that interrogates race and racism in education is required. While progressive initiatives in education around racial inequity are important, paying attention to the limitations of such projects is also key in making new and better possibilities imaginable.

See also: *Black in White: Black Students at White and Black Colleges*; *Equity and Educational Effectiveness*; *Race and Ethnicity in the Field of Adult Education*; *The History of Education: Race and Education*.

Bibliography

- Akintunde, O. (1999). White racism, white supremacy, white privilege, and the social construction of race: Moving from modernist to postmodernist multiculturalism. *Multicultural Education* 7(2), 2–8.
- Alcoff, L. M. (2005). *Visible Identities: Race, Gender, and the Self*. Oxford: Oxford University Press.
- Appiah, K. A. (1992). The illusion of race. In Appiah, K. A. (ed.) *In My Father's House: African in the Philosophy of Culture*, pp 28–46. Oxford: Oxford University Press.
- Apple, M. W. (1998). Forward. In Kincheloe, J. L., Steinberg, S. R., Rodriguez, N. M., and Chennault, R. (eds.) *White Reign: Deploying Whiteness in America*, pp ix–xii. New York: St. Martin's Press.
- Brown, W. (1995). *States of Injury: Power and Freedom in Late Modernity*. Princeton, NJ: Princeton University Press.
- Butler, J. (1990). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.
- Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review* 43(6), 1241–1299.
- Dalton, H. (1995). *Racial Healing*. New York: Doubleday.
- Delgado, R. and Stefancic, J. (2001). *Critical Race Theory: An Introduction*. New York: New York University Press.
- Delpit, L. (1995). *Other People's Children: Cultural Conflict in the Classroom*. New York: New York Press.
- Derman-Sparks, L. and Phillips, C. B. (1997). *Teaching/Learning Anti-Racism: A Developmental Approach*. New York: Teachers College Press.
- Dixon, A. D. and Rousseau, C. K. (eds.) (2006). *Critical Race Theory in Education: All God's Children Got a Song*. New York: Routledge.
- Dyer, R. (2000). The matter of whiteness. In Black, L. and Solomos, J. (eds.) *Theories of Race and Racism: A Reader*, pp 539–548. New York: Routledge.
- Flores, L. A. and Moon, D. G. (2000). Rethinking race, revealing dilemmas: Imagining a new racial subject. *Western Journal of Communication* 66(2), 181–207.
- Frankenberg, R. (1996). 'When we are capable of stopping, we begin to see': Being white, seeing whiteness. In Thompson, B. and Tyagi, S. (eds.) *Names We Call Home*, pp 3–17. New York: Routledge.
- Freeman, A. D. (1996). Legitimizing racial discrimination through anti-discrimination law: A critical review of Supreme Court doctrine. In Crenshaw, K., Gotanda, N., Peller, G., and Thomas, K. (eds.) *Critical Race Theory: The Key Writings that Found the Movement*, pp 29–45. New York: New Press.
- Frye, M. (1983). Oppression. In Frye, M. (ed.) *The Politics of Reality: Essays in Feminist Theory*, pp 1–16. Freedom, CA: Crossing Press.
- Frye, M. (1992). White woman feminist. In Frye, M. (ed.) *Willful Virgin: Essays in Feminism*, pp 147–169. Freedom, CA: Crossing Press.
- Gallagher, C. (2003). Playing the ethnic card: Using ethnic identity to negate contemporary racism. In Doane, A. and Bonilla-Silva, E. (eds.) *White Out: The Continuing Significance of Racism*, pp 145–158. New York: Routledge.
- Giroux, H. (1997). Rewriting the discourse of racial identity: Towards a pedagogy and politics of whiteness. *Harvard Educational Review* 67(2), 285–320.
- Harris, A. P. (1990). Race and essentialism in feminist legal theory. *Stanford Law Review* 42(3), 581–616.
- Harris, C. (1993). Whiteness as property. *Harvard Law Review* 106, 1710–1791.
- Haslanger, S. (1995). Ontology and social construction. *Philosophical Topics* 23(2), 95–122.
- Hytten, K. and Warren, J. (2003). Engaging whiteness: How racial power gets reified in education. *Qualitative Studies in Education* 16(1), 65–89.
- Ignatiev, N. (1995). *How the Irish Became White*. New York: Routledge.
- Ignatiev, N. and Garvey, J. (1996). *Race Traitor*. New York: Routledge.
- Jones, A. (1999). The limits of cross-cultural dialogue: Pedagogy, desire, and absolution in the classroom. *Educational Theory* 49(3), 299–316.
- Keating, A. L. (1995). Interrogating 'whiteness,' (de)constructing 'race.'. *College English* 57(8), 901–918.
- Kozol, J. (1991). *Savage Inequalities: Children in America's School*. New York: Harper/Perennial.
- Kruks, S. (2001). *Retrieving Experience: Subjectivity and Recognition in Feminist Politics*. Ithaca, NY: Cornell University Press.
- Ladson Billings, G. and Tate, W. F. (1995). Toward a critical race theory of education. *Teachers College Record* 97(1), 47–67.
- Leonardo, Z. (2004). The color of supremacy: Beyond the discourse of 'white privilege' *Educational Philosophy and Theory* 36(2), 137–152.
- Lewis, A. (2001). There is no race in the school yard: Color-blind ideology in an (almost) all-white school. *American Educational Research Journal* 38(4), 781–811.
- Lewontin, R. (1972). The apportionment of human diversity. *Evolutionary Biology* 6, 381–398.
- Haney Lopez, I. F. (2000). The social construction of race. In Delgado, R. and Stefancic, J. (eds.) *Critical Race Theory: The Cutting Edge*, 2nd edn., pp 163–175. Philadelphia, PA: Temple University Press.
- MacMullan, T. (2005). Beyond the pale: A pragmatist approach to whiteness studies. *Philosophy and Social Criticism* 31(3), 267–292.
- McCarthy, C., Crichlow, W., Dimitriadis, G., and Colby, N. (eds.) (1993). *Race, Identity and Representation in Education*. New York: Routledge.
- McIntosh, P. (1997). White privilege and male privilege: A personal account of coming to see correspondences through work in women's studies. In Delgado, R. and Stefancic, J. (eds.) *Critical White Studies: Looking Behind the Mirror*, pp 291–300. Philadelphia, PA: Temple University Press.
- McIntyre, A. (1997). White talk. In McIntyre, A. (ed.) *Making Meaning of Whiteness*, pp 45–78. Albany, NY: State University of New York Press.
- McLaren, P. (2000). Unthinking whiteness: Rearticulating diasporic practice. In Trifones, P. P. (ed.) *Revolutionary Pedagogies: Cultural Politics, Instituting Education, and the Discourse of Theory*, pp 140–173. New York: Routledge.
- Minow, M. (1990). *Making All the Difference: Inclusion, Exclusion and American Law*. Ithaca, NY: Cornell University Press.
- Moon, D. and Flores, L. A. (2000). Antiracism and the abolition of whiteness: Rhetorical strategies of domination among 'race traitors' *Communication Studies* 5(1), 97–115.
- Morrison, T. (1993). *Playing in the Dark: Whiteness and Literary Imagination*. London: Vintage.
- Moya, P. (2002). *Learning from Experience: Minority Identities, Multicultural Struggles*. Berkeley, CA: University of California Press.
- Nieto, S. (1995). From brown heroes and holidays to assimilationist agendas: Reconsidering the critiques of multicultural education. In Sleeter, C. and McLaren, P. (eds.) *Multicultural Education, Critical Pedagogy and the Politics of Difference*, pp 191–220. Albany, NY: State University of New York Press.
- Omi, M. and Winant, H. (1994). *Racial Formation in the United States: From the 1960's to the 1990's*, 2nd edn. New York: Routledge.
- Orosco, J. -A. (2002). Everybody in the barrio is a nationalist: Chicanismo and the limits of the politics of recognition. In Bailey, A. (ed.) *Community, Diversity, and Difference: Implications for Peace*, pp 65–79. Amsterdam: Rodopi.
- Payne, R. J. (1998). *Getting Beyond Race: The Changing American Culture*. Boulder, CO, Westview.
- Rodriguez, N. M. (1998). Emptying the content of whiteness: Toward an understanding of the relation between whiteness and pedagogy. In Kincheloe, J. L., Steinberg, S. R., Rodriguez, N. M., and Chennault, R. (eds.) *White Reign: Deploying Whiteness in America*, pp 31–62. New York: St. Martin's Press.
- Roediger, D. R. (1991). *The Wages of Whiteness*. New York: Verso.
- Schmidt, S. L. (2005). More than men in white sheets: Seven concepts critical to the teaching of racism as systemic inequality. *Equity and Excellence in Education* 38(2), 110–122.
- Schofield, J. W. (1997). Causes and consequences of the colorblind perspective. In Banks, J. and McGee Banks, C. (eds.) *Multicultural Education: Issues and Perspectives*, 3rd edn., pp 251–271. Needham Heights, MA, Allyn and Bacon.
- Spelman, E. V. (1988). *Inessential Woman: Problems of Exclusion in Feminist Thought*. Boston, MA: Beacon Press.
- Tatum, B. D. (1997). *Why Are All The Black Kids Sitting Together in The Cafeteria?* New York: Basic Books.
- Taylor, C. (1994). *Multiculturalism and the Politics of Recognition. An Essay by Charles Taylor. With Commentary by Amy Gutmann (Editor), Steven C. Rockefeller, Michael Walzer, and Susan Wolf*. Princeton, NJ: Princeton University Press.

- Taylor, P. (2000). Appiah's uncompleted argument: W.E.B. DuBois and the reality of race. *Social Theory and Practice* 26(1), 103–128.
- Warren, J. T. (2001). Performing whiteness differently: Rethinking the abolitionist project. *Educational Theory* 51(4), 451–466.
- West, C. (1996). Our next race question: The uneasiness between Blacks and Latinos. *Harper's Magazine* 292(1751), 55.
- Winant, H. (1997). Behind blue eyes: Whiteness and contemporary US racial politics. In Fine, M., Weis, L., Powell, L., and Wong, L. (eds.) *Off-White: Readings on Race, Power and Society*, pp 40–55. New York: Routledge.
- Kolchin, P. (2002). Whiteness studies: The new history of race in America. *Journal of American History* 89(1), 154–173.
- Sheets, R. H. (2000). Advancing the field or taking center stage: The white movement in multicultural education. *Educational Researcher* 29(2), 15–21.
- Thompson, A. (2003). Listening and its asymmetries. *Curriculum Inquiry* 33(1), 79–100.

Further Reading

- Grillo, T. (2006). Anti-essentialism and intersectionality: Tools to dismantle the master's house. In Hackett, E. and Haslanger, S. (eds.) *Theorizing Feminisms: A Reader*, pp 30–39. Oxford: Oxford University Press.

Relevant Website

<http://cdms.ds.uiuc.edu> – Critical Whiteness Studies.

School Choice and the Common School

G Haydon, Institute of Education, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Common school – A school open to all potential students without discrimination, where the actual student body is representative of diversity in the surrounding community, and the education provided is appropriate for all members of the wider society.

Comprehensive school – A common school within the public education system of the United Kingdom.

School choice – A system or policy that enables parents or students to exercise choice among a number of schools within a publicly funded system.

Separate school – A school that intentionally limits its student body to a certain sector or sectors of the wider community; or a school in which the student body is contingently so limited as a result of demographic factors and parental choice (a *de facto* separate school).

The topic of school choice and the common school has come to seem a natural one for inclusion in a survey of educational issues because of the way the two elements of the topic are linked. Two ideas are sometimes joined into a single topic because they are perceived as fitting together, and sometimes because they are seen as opposed. The latter is the way the pairing of school choice and the common school is often seen: either school choice or the common school. Which should be favored, and why?

Once the topic is set up in this way, it is obvious that a great many considerations, including matters for empirical research and for practical politics, will come into the debate; and many of these considerations will differ from one school system to another.

Since the present article is written within the perspective of philosophy of education, it will not engage with the detail of such considerations. That is not to say that philosophers cannot take account of empirical data and of practical politics; indeed, some of the writers who have entered the debate from a philosophical perspective do pay a good deal of attention to empirical data (e.g., Brighouse, 2000). Still less is it the position of this article that philosophy by itself can settle such a debate. But there is a contribution that philosophy is particularly well placed to make – to clarify the terms of the debate and the nature of the arguments on both sides. Though this may sound like a mere preliminary, it is important because there is a real possibility that different parties in the debate will

have different understandings of what is meant by the common school and school choice.

A logically prior question is whether it is right to set up this topic as a two-sided debate in the first place; but it is hard to answer that question without unpicking the assumptions being made about what is involved in common schools and in school choice. Only after doing that can it be seen whether the two ideas are mutually exclusive.

Even in a philosophical discussion, an abstract comparison of the two ideas, detached from any historical or political context, would barely be possible or illuminating. Where common schools have been advocated, it has been as a practical policy within certain societies; where school choice has been advocated, it has also been as a matter of practical policy, often in reaction against a system in which common schools are predominant. There is a background assumption across most of the debate over common schools and school choice – that the ways in which schooling is provided, funded, and regulated (see Tooley (2000: 9) for this distinction between logically separable forms of state intervention) are matters for political decision. Even when, as now in many societies, there are moves away from common schools and toward greater choice in schooling – and even when such moves are at the same time moves toward lesser governmental control over schooling – it is still within the powers of government to decide how far such moves are to be allowed, encouraged, or actively instituted.

Common Schools and Separate Schools

Within such a context we can give an initial characterization – needing further refinement – of what a common school is. The present discussion draws on Haydon (2007) and on other contributors' chapters in a volume on *The Common School and the Comprehensive Ideal*.

A common school is a school open to and intended to provide education to all potential students in its locality without discrimination on the basis of any differentiating characteristics, such as the ability, gender, or ethnicity of the student or the religion, class, or wealth of the student's parents. Such a statement of intention, however, is not yet sufficient to capture the idea of a common school as it has been understood by its advocates. Two things at least have to be added: (1) that the intention of nondiscrimination should be realized in the actual student body; the actual mix of students within the school should reflect at least as

much diversity as there is in the surrounding population; (2) that the nature of the education provided in such a school should itself be common, in the sense both that it is essentially the same education provided for all within the school, and that it is an education thought to be appropriate for all citizens within the wider society (McLaughlin, 2003). The following paragraphs will focus on point (1) before moving to point (2).

What is opposed to the notion of the common school, understood as one open to all students, is not the idea of school choice as such, but the idea of a school that has a limited mix of students because it does discriminate on the basis of one or more differentiating factors. The term separate schools has been widely used by philosophers of education for such schools (e.g., McLaughlin, 1992). A separate school may, for instance, take students only from a particular faith community; or only girls; or only students with certain kinds of special-educational needs.

Where there are separate schools, this fact does not by itself mean that it is within the scope of choice by students or their parents to attend or not attend a given school. In many of the southern states of the USA, prior to the Supreme Court decision in *Brown versus Board of Education* in 1954, there were separate (and allegedly equal) schools for blacks and whites. In England, for at least two decades after the Education Act of 1944, there were separate schools for students of different measured intellectual ability. In such cases, though there may sometimes have been a degree of parental choice available between different local schools in the same category, it was essentially government (sometimes local government acting within national policy) that decided which kind of school any given student attended.

The English case can serve as an illustration, not only that separate schools are not necessarily open to student or parental choice, but also that what is a common school in intention is not necessarily a common school in its actual mix of students. At a time when many local education authorities in England began to provide schools (known as comprehensive schools) open to any student without selection by ability, it was often the case that other schools which did select from the upper percentiles of the measured ability range were operating in the same geographical area. Since the most able students mostly went to those selective schools, the comprehensive schools did not in fact contain a representative range of students from the surrounding population. Another way in which a notionally common school may be far from comprehensive in the actual composition of the student body is that the demographics of the local area served may be far from representative of the demographics of the wider society. Thus there are notionally common schools in England whose students are almost entirely Asian and Muslim; and notionally common schools in the USA whose students are almost entirely black (Levinson, 2007). These schools are *de facto* separate schools.

These points illustrate that while it is possible to describe a single school as a common school, the debate between common schools and separate schools, as also between common schools and school choice, is a debate over schooling systems and policy. On the one side, the favored system is one in which the norm for the great majority of students is that they attend a common school provided for the whole community; on the other side the favored system is one in which there is a diversity of kinds of school, between which parents have a degree of choice. Exactly what degree of diversity, what level of choice, and through what mechanism choice is exercised, are factors that may depend on local or national conditions. There are also possible variations as to who actually exercises the choice; here, as is usual in this debate, we refer to parental choice (where parents include legal guardians). But it is worth noting that if there is to be choice between schools there is a strong case for students to be involved in the choice, and many parents, especially in choosing a secondary school, will put weight on their offspring's views.

Moving to point (2) above, we can say that a common school is not only one that has a mixed student body, but also one that promotes a common education, that is, an education thought to be appropriate and desirable for all students. (It is worth noting at this point that separate schools may or may not differ in this respect from common schools. A school educating only students from one faith tradition may promote an education that is not common in that it includes initiation into the doctrines of that faith; on the other hand, a school that takes only female students may believe that girls should have the same kind of education as boys.)

What counts as a common education is open to more or less minimal interpretations (Callan, 1997: ch. 7; Haydon, 2007). It might comprise no more than a list of subjects that are more or less universally accepted as essential in anyone's education: numeracy, literacy, and perhaps some basics of scientific and historical understanding. But usually the advocates of common schooling have something fuller than this in mind, for reasons that will be apparent when we consider the arguments offered in favor of common schools.

The majority of such arguments turn on one or both of two considerations: justice to individuals, and the cohesion and flourishing of the community. Separate schools may be unjust to individuals by giving different life opportunities to different individuals for reasons that are in no way dependent on the individuals' own choice or effort. It would be very widely agreed now that the separate-but-equal schools of the southern states were in fact unequal in this way; and the English system of selecting students by ability rapidly came to be seen by many as unjustly prejudicing the life chances of children by assigning them to ability categories at the age of 11 years. In the background to such arguments is a commitment that as far as possible,

all citizens should have an equally good education giving them access to equally good opportunities in life.

More positive arguments spell out more about what such an education would involve and why it is desirable. Often referring to Dewey's (1916) classic *Democracy and Education*, the positive arguments see a common education in common schools as the best if not the only way in which a cohesive democratic community can be forged within large heterogeneous societies. The background assumption here, perhaps even more inescapable a century after Dewey was writing, is that of pluralism in the cultures and traditions of the inhabitants of modern states. In the multicultural society of Britain at the beginning of the twenty-first century, there are arguments for retaining common schools that echo those of Dewey. Alongside arguments about the forging of common values within a plurality of values, there are arguments that only in common schools can the attitudes and practices of a genuinely democratic citizenry be formed (Fielding, 2007). Respect, tolerance, willingness to listen to different points of view, can all best be formed, many would claim, in common schools. Conversely, there have been fears that where parents are able to choose separate schooling for students from different cultural and religious traditions – or where parents of one social class are able to separate their children from those of another class – this will tend to undermine the cohesion of the wider society.

School Choice

Against this background, we can turn to the question of what is to be understood by school choice and what arguments can be made for it. Some simplifying assumptions will be made here. First, although one could consider within the scope of school choice the choice that a child will not attend school at all, the possibility of home schooling will not be considered here. Further, although choosing to pay for one's child to attend a private school could also clearly be counted within the scope of school choice, the debate over the legitimacy and desirability of private schooling will not be considered here (but see Tooley (2003) for an argument that the considerations brought forward in support of school choice within a state system can actually be seen to support full privatization of schooling). The school choice considered here is choice between schools within a publicly funded system of schools.

There are in fact a variety of possible systems that may be said to incorporate school choice, and correspondingly a variety of arguments that may be put forward. Here, a background assumption (not entirely beyond challenge, as we shall see) is that choice presupposes a diversity of schools between which a choice can be made. But how much diversity is one of the variable factors. We can

roughly outline a continuum from quite limited diversity to much greater diversity.

At one end of this continuum is a system in which all schools are common schools, serving a mixed population, and expected to serve certain publicly mandated educational aims, but where there are nevertheless some differences between schools. Within urban areas, it may be practically feasible for parents to choose between a number of accessible common schools.

Why in such circumstances might a policy be favored that gives to parents a choice between several alternatives? We can distinguish here (1) arguments that assume shared educational preferences across all parents and students, and (2) arguments that rest on a diversity of preferences. The two kinds of argument may be used in combination, since parental goals and preferences may coincide in some respects but not in others, but they will be treated separately here.

1. There are arguments from the alleged benefits of competition to the effect that the existence of a plurality of schools, between which a choice can be made, will have instrumental value in raising the level of provision of a good that all parents are assumed to want for their children. For instance, it may be assumed that what all parents want for their children is a good education, where this is assumed to equate to academic achievement and thus to be measurable by test results. Given such a common measure, it is argued that not all schools will be equally effective in achieving good results, while all schools need an incentive to do as well as they can. In a school choice system, where it is assumed that parents will wish their children to go to the schools with the best results, schools have to compete to attract parents and therefore will try to improve their test results. This is a version, then, of the familiar economic argument that competition drives up performance. Some states have also made arrangements by which competition between schools is intended to drive down costs, thus achieving greater efficiency in the use of public funds (see Brighouse (2000: 28–34) for a useful review of this and other arguments for school choice).
2. There are other arguments based on the assumption that preferences, or indeed conceptions of a good education, vary across parents and students. No two schools will be identical in all respects, even if they cannot be distinguished in terms of measurable results, and even indeed if they share an avowed commitment to the promotion of common values and citizenship within a plural society. Some schools are large and vibrant places; others are small and calmer. Some may have an ethos of student deference toward teachers, and others, a culture of friendly two-way relationships between students and teachers. Some may put weight on the delivery of a prearranged curriculum almost to the exclusion of other concerns, while others

allow more flexibility in following up students' interests. Some may aspire to a democratic organization while others will reflect the bureaucratic nature of the wider system, and so on. Though such differences may coexist with similar avowed goals and similar academic results, they can make a large difference to the quality of student experience. If we add the premise that different conditions may be congenial to and appropriate for different individuals, and the further premise that parents will know best what is appropriate for their own children, there are grounds for a presumption that where possible, students and their parents should be able to choose the conditions they find congenial. To this the argument can be added that encouragement to schools to develop their own characteristics gives scope for valuable innovation to emerge.

So far, then, even where a system of common schools is in place, there can be arguments for the desirability of having variation between schools within that system, though by the definitions above, the schools remain common rather than separate. It must be acknowledged that there is also a danger in combining parental choice with common schools, in that parents may take any factors they wish into account in their choice. Where parents choose to send their child to a school where most of the other students are of similar background, this can be the beginning of a slide toward *de facto* separate schools.

The move further along the continuum of diversity involves the formal recognition of separate schools, first with a shared commitment to a common education, and then with no such commitment. At the farthest end of the continuum will be a situation of maximal differences between schools with minimal attempt at government regulation of what happens within the schools. Arguments for moving along this continuum can be made in two stages: arguing for separate schools, and arguing for choice between them. The second stage is the easier, in that if there are separate schools, such as schools promoting one religion, or schools only for one gender, and so on, then in a liberal society, there must be a presumption in favor of parents – and here it is perhaps especially important to add, students too – having either free choice or at least a strong voice in determining the kind of school a child goes to. If certain children were sent to a school promoting, say, a given religion against the parents' wishes, that would be a *prima facie* overriding of liberty and of one element of Clause 26 of the Universal Declaration of Human Rights.

That point does not, however, settle whether it is right that there should be separate schools. The argument that there should be separate schools of whatever kind cannot rest on an abstract appeal to parental rights but has to counter the strong arguments for the desirability of common schools (Brighouse, 2007: 583). The argument also

needs to be made on its merits for any given kind of separation. An argument for separate schools for different ethnic groups, for instance, would be very hard to make within a liberal democratic society where battles against racial discrimination have not yet been finally won. In contrast, the desirability of separate schools for students with certain severe disabilities is a live issue in many countries. An argument for such schools may appeal to the benefits to the individuals concerned that can be provided in specialist schools, while on the other hand it is argued that the deliberate segregation of disabled students can be harming to their self-respect and deleterious to their future integration within the wider community.

An argument for separate schools for different religious communities can appeal to parents' strong interest in bringing up their children within the belief system that they (the parents) believe to be right, and can appeal also to a child's interest in being brought up within a coherent culture (Merry, 2005). On the other side (starting from liberal premises) is the right of every child not to be inducted into a specific worldview, including that of her parents, and the interest of the whole community in having citizens who are open minded and prepared to take seriously the perspectives of others. Many theorists have concluded that the two sides of this argument need not be irreconcilable. It is open to governments, while allowing and perhaps even encouraging certain kinds of separate school, to insist that such schools must subscribe to the common educational goals of the society. Where such a policy is followed, there can be room for parental choice without – in principle – undermining the pursuit of the goals that common schools are meant to promote. Nevertheless, there remains the possibility that the goals of a common education for shared citizenship will be better achieved within those schools that actually are common schools, where students from different backgrounds really do meet and mix with others across differences in beliefs and values, so that this very interaction is part of the educational experience.

Values Underlying the Arguments

If common educational goals are intended as a response to plurality of values, it is also the case that the debate over common schools and school choice reflects a plurality of values. It is not that certain values and certain traditions stand firmly on one side of the debate and others on the other side. Faith communities, for instance, contain some parents who wish their children to be educated only within their faith community, and other parents who acknowledge the desirability of their children being educated alongside those with quite different beliefs. Both sides can, to a considerable degree, recognize the force of the same values while perhaps interpreting differently the

practical consequences of these values. All within this debate can acknowledge the human value of belonging to and being able to identify with a community, though for some it will be the national citizenship community that takes precedence and for others a more specific and local community. All can acknowledge the value of respect for persons: for some, the respect that matters most is that the state respect the rights of parents to live in the ways of their own culture and to bring their children up within it; for others the emphasis is more on mutual respect between cultures that may be best nourished within a common school.

Justice, as we have seen, underpins one argument for common schools, inasmuch as they involve neither discriminating against nor giving preferential treatment to any. But proponents of parental choice can sometimes claim with plausibility that justice is on their side. As regards the schooling of students with severe special needs, as mentioned above, there are strong proponents of their integration into mainstream schooling, but some parents can argue that common schooling is unjust to their children through failing to provide sufficiently for their individual needs. There is also the more general argument, stressed by Brighouse and others that nominal common-school systems unjustly allow choices to parents who can afford to move house to be close to a well-favored school, whereas poorer parents have no such choice.

Liberal philosophical approaches to education commonly give important weight to individual autonomy as a central quality to be developed through education. For many, the common school is the best environment for developing autonomy, because it provides for the child a certain detachment from the home community (Levinson, 1999). But others will claim, putting weight in part on the importance of cultural coherence, that schooling according to the culture of the home community can be a route toward autonomy (McLaughlin, 1992).

Proponents of parental choice can appeal to a principle of liberty, against the coercive intervention of the state into the life of families (Reich, 2007; Brighouse, 2007). But advocates of common schooling may claim to be putting more weight on the potential liberty of the individual student, and also may take a more civic republican view of liberty in which the liberty of the adult within society will not be purely private but will involve engagement within public affairs.

Although what are at least nominally the same values may be appealed to by all participants in these debates, it is possible to discern an underlying difference in what might be called the predominant ethos on either side. The advocacy of school choice is at home within a social ethos in which choice itself, construed on the model of consumer choice in the market, is given high positive value. Many of the arguments that are applied in wider contexts of the marketplace are used here too – that choice

between alternatives empowers the consumer, and gives providers motives to monitor their performance, diminish inefficiency, and be transparent in their operations.

In support of common schooling, it can be said that there are other goods than those that can be chosen on an individual basis, and that the good society cannot be equated simply with a society in which the satisfaction of individual preferences is maximized. Notions of fraternity and solidarity (Fielding, 2007) are at home in the case for the common school as they cannot be in the case for school choice (as usually understood). At the same time, a positive valuation of fraternity and solidarity should not be taken to imply that choice as such is undesirable. Here we need to return to the thought that the common school itself can itself be chosen.

The Common School as an Object of Choice

Common schools can be objects of choice in several ways, differentiated by what it is in preference to which common schools are chosen. First, there is the kind of arrangement already considered, involving a choice between different common schools within a system of common schools. Here, parents choose one common school in preference to another. This has been sufficiently considered above.

Second, a common school can be an object of choice in preference to a separate school. This, of course, is only possible in practice where there is more than one school that a child could attend, including at least one common school and at least one separate school. Such contexts (some examples of which have already been noted) are potentially unstable, since if too many parents choose separate schools, the intake of students to common schools can become distorted to the point where they become only nominally common. To the extent that this happens, whether by deliberate policy or default, parents no longer have the possibility of choosing a common school even though they may have a greater range of separate schools to choose between.

Finally, there is a sense in which common schools can be an object of collective choice. Here, what is chosen is a system of common schools in preference to a system of (diverse kinds of) separate schools. This is not a choice that can be made by individuals acting alone, but is potentially an object of choice for citizens, to be chosen or rejected through democratic political processes. There is perhaps some danger that the penetration of market forces into the political thinking of liberal states will encourage a drift away from common schools that will make the collective choosing of common schooling only an in-principle possibility.

The demise of common schools would itself diminish the range of alternatives open to citizens. This would be a paradoxical outcome in any educational system that

wishes to promote the virtues of shared citizenship within a cohesive society, since it would in effect assume that citizens would only want to make their choices by reference to their private preferences for their children, rather than by any wider public-spirited motivation. The challenge for policymakers and for philosophers of education who seek to influence policy is to devise systems, realizable within the practical politics of given societies, that realize some of the benefits of choice without unduly restricting the very basis on which choices in schooling can be made.

Bibliography

- Brighouse, H. (2000). *School Choice and Social Justice*. Oxford: Oxford University Press.
- Brighouse, H. (2007). Educational justice and socio-economic segregation in schools. *Journal of Philosophy of Education* 41(4), 575–590.
- Callan, E. (1997). *Creating Citizens: Political Education and Liberal Democracy*. Oxford: Oxford University Press.
- Dewey, J. (1916). *Democracy and Education*. New York: Free Press.
- Fielding, M. (2007). On the necessity of radical state education: Democracy and the common school. *Journal of Philosophy of Education* 41(4), 539–557.
- Haydon, G. (2007). In search of the comprehensive ideal. *Journal of Philosophy of Education* 41(4), 523–538.
- Levinson, M. (1999). *The Demands of Liberal Education*. Oxford: Oxford University Press.
- Levinson, M. (2007). Common schools and multicultural education. *Journal of Philosophy of Education* 41(4), 625–642.
- McLaughlin, T. (1992). The ethics of separate schools. In Leicester, M. and Taylor, M. (eds.) *Ethics, Ethnicity and Education*, pp 114–136. London: Kogan Page.
- McLaughlin, T. (2003). The burdens and dilemmas of common schooling. In McDonough, K. and Feinberg, W. (eds.) *Citizenship and Education in Liberal-Democratic Societies*, pp 121–156. Oxford: Oxford University Press.
- Merry, M. (2005). Cultural coherence and schooling for identity maintenance. *Journal of Philosophy of Education* 39(3), 477–497.
- Reich, R. (2007). How and why to support common schooling and educational choice at the same time. *Journal of Philosophy of Education* 41(4), 709–725.
- Tooley, J. (2000). *Reclaiming Education*. London: Cassell.
- Tooley, J. (2003). Why Harry Brighouse is nearly right about the privatisation of education. *Journal of Philosophy of Education* 37(3), 427–447.

Further Reading

- Bridges, D. and Jonathan, R. (2002). Education and the market. In Blake, N., Smeyers, P., Smith, R., and Standish, P. (eds.) *The Blackwell Guide to the Philosophy of Education*, pp 126–145. Oxford: Blackwell.
- Halstead, J. M. and Haydon, G. (2008). *The Common School and the Comprehensive Ideal: A Defence by Richard Pring with Complementary Essays*. Oxford: Blackwell.
- Reich, R. (2003). Common schooling and educational choice. In Curren, R. (ed.) *A Companion to Philosophy of Education*, pp 430–442. Oxford: Blackwell.

PHILOSOPHY OF EDUCATION – PHILOSOPHICAL PERSPECTIVE

Contents

African Philosophy of Education
Feminism
Hermeneutics
Islamic Education
Liberalism and Education
Phenomenology and Existentialism
Pragmatism
Psychoanalysis and Education
Social and Cultural Capital in Education
The Aims of Education
The Analytical Tradition

African Philosophy of Education

K Horsthemke, University of the Witwatersrand, Johannesburg, South Africa

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Amaqihikiza – A type of mentorship program among older and younger girls to ensure sexual abstinence until the latter are ready to take full control of their affairs.

Ubuntu (botho; hunhu) – Humanness.

Ukuhlolwa kwezintombi – Virginity testing in girls.

Umuntu ngumuntu nga banto (motho ke motho ka batho) – A person (a human being) is a person (is human) because of other persons (other human beings).

Education in Africa, like virtually all social interaction, was traditionally characterized by the bounds and bonds of small-scale ethnic communities. Despite their diversity and remoteness, and the sheer geographic distances between these groups, they are seen to share not only the experience of colonization but also (predating this experience) certain common educational concepts, principles, and values.

African philosophy of education is, in some important way, contingent on African philosophy. While existence of the latter is guaranteed by manifestations of the former, the possibility and plausibility of African philosophy does not necessarily imply the same for African philosophy of education. At the heart of these considerations resides the question whether there is a (set of) perspective(s), a body of thought, and/or a particular way of doing philosophy of

education, that can be called African. There are arguably two initial ways of approaching this issue:

- Are there uniquely and distinctly African ways of philosophizing about education?
- Are the component concepts, principles, and values of this philosophy sound?

Whether or not satisfactory answers to these questions are forthcoming, the possibility and plausibility of African philosophy of education might nonetheless be established in terms of its priorities. Given the different historical, geographical, cultural, and social contexts of Africans and education in Africa, it is reasonable to assume that philosophical priorities differ in accordance with these.

Introduction: Modes of African philosophy

Perhaps significantly more than philosophy elsewhere, African philosophy has been marked indelibly by the colonial experience. Historically, and for reasons of graphic illustration, it might be divided into its precolonial and postcolonial manifestations. Precolonial African philosophy had, with very few exceptions (Egypt comes to mind here), an essentially oral tradition. The written tradition came with and succeeded colonialism, exemplified *inter alia* by missionary education. Ethnic philosophy

and sage philosophy characterized the former, while political philosophy and critical (academic or professional) philosophy exemplified the latter.

Ethnic philosophy consisted of folkloric traditions, legends, stories, and myths, and survives in the postcolonial period in both oral and, importantly, written forms. So does sage philosophy, initially the spoken words and teachings of a few wise men or sages, now also documented in writing. Political and critical philosophy, on the other hand, were marked – if not determined – by the colonial experience. The writings and documented speeches of politicians, statesmen, and prominent liberation movement personalities like Kwame Nkrumah, Julius Nyerere, Léopold Senghor, Kenneth Kaunda, and Steve Biko, to name only a few, constituted political philosophy that often also had a nationalist-ideological character. A fourth trend in African philosophizing is the direction pursued by critical or professional philosophy (see Oruka, 1998, 2002). This is a direction associated, for example, with the writings and other contributions of professional philosophers and academics like Kwame Anthony Appiah, Paulin J. Hountondji, Peter Bodunrin, and Kwasi Wiredu.

Although both Henry Odera Oruka and Kwame Gyekye can be credited with recording, concurrently and independently from one another, contemporary nonacademic intellectual traditions (Löfke, 2001: 140), Oruka was the first, in 1978 (Oruka, 2002), to undertake this classification. He has recently (Oruka, 1998: 101, 102) described two additional types of philosophy, the hermeneutic trend and the artistic or literary trend – somewhat unhelpfully, because the former appears to be subsumed by critical or professional philosophy, while the latter contains elements of the other trends identified previously, ethnic philosophy, sage philosophy, and political philosophy. These different kinds of philosophy and philosophizing are illustrated in **Table 1**.

It is important to note that ethnic philosophy and sage philosophy have survived colonialism and that they continue to thrive in postcolonial Africa. Both political and critical philosophy frequently exhibit or seek to validate elements of the former kinds of philosophy. Indeed, the number of academic or professional philosophers

repudiating substantial elements of their African doxastic and conceptual heritage remains fairly small.

A further distinction might be made at this juncture between philosophy as worldview and philosophy as critical activity. Ethnic philosophy and sage philosophy essentially exemplify the former (the worldview in question being either divinely inspired, or by the ancestors, or by the tribal elders). So do – to an extent – the postcolonial visions and ideologies of political leaders and liberation movement personalities (who were/are characteristically not academic or trained philosophers). Only with professional philosophy, at least to a greater extent, has there been a noticeable trend toward critical activity, interrogation not only of the colonial intellectual heritage but also of indigenous worldviews (**Table 2**).

African Traditional Education

African traditional and/or religious worldviews are frequently cited as exemplifying characteristically African philosophy. With regard to education in particular, uniqueness and distinctness are claimed for the component principles of African traditional education: preparationism, functionalism, communalism, perennialism, and wholisticism, principles that, furthermore, yield sound philosophical foundations (Adeyemi and Adeyinka, 2003: 431). Before turning to current trends within African philosophy of education, the author examines briefly each of these traditional principles. Are they uniquely African? Are they defensible?

In customary education, children were equipped with the skills appropriate to their gender, in preparation for their distinctive roles in society. Male education thus produced farmers, warriors, blacksmiths, rulers, and other male-dominated occupations from which women were excluded, while female education was predominantly designed to produce future wives and mothers (Adeyemi and Adeyinka, 2003: 432). Learning through imitation, initiation ceremonies, work, play, oral literature, etc., people were productive as they learned and were smoothly integrated into the community (Adeyemi and Adeyinka, 2003: 432). Third, all members of the society owned things

Table 1 Precolonial and postcolonial philosophies

<i>Precolonial (oral)</i>	<i>Postcolonial (written)</i>
Ethnic philosophy (or ethnophilosophy)	→
Sage philosophy/ philosophical sagacity	→
	Political (nationalist-ideological) philosophy
	Critical (academic or professional) philosophy

Table 2 Philosophy as worldview versus critical activity

<i>Philosophy as worldview</i>	<i>Philosophy as critical activity</i>
Ethnic philosophy (or ethnophilosophy)	
Sage philosophy/ philosophical sagacity	
Political (nationalist-ideological) philosophy	→
	Critical (academic or professional) philosophy

in common and applied the communal spirit to life and work (Adeyemi and Adeyinka, 2003: 432). This was manifest not only in shared manual labor and joint savings but also in the communal upbringing and disciplining of children. The latter were also discouraged from critical interrogation and from experimenting with the unknown, with punitive measures being imposed on those who tried to do so (Adeyemi and Adeyinka, 2003: 433). Education was essentially conservative in nature, perceived as a vehicle for maintaining or preserving the cultural heritage and *status quo* (Adeyemi and Adeyinka, 2003: 433). Finally, the wholistic nature of customary education enabled young people to acquire a variety of skills. Thus, a

child, destined to become a fisherman ... learned not only to catch fish but also to preserve and market it; to make and mend nets; to manufacture canoes and to erect temporary fishing huts. ... An individual [therefore] ... could work as a builder, farmer or fisherman, while a woman was a gardener, housewife and cook, besides being a nurse to her children, etc. (Adeyemi and Adeyinka, 2003: 433)

In a sense, not only African communities but most (if not all) small-scale societies exhibit or are guided by these principles. Moreover, one might think – with all due acknowledgment of their historical and social context – that the androcentrism and gender bias exhibited by some of them, not to mention the more or less tacit endorsement of indoctrination and irrational belief, or superstition, disqualifies a number of these principles from being taken seriously within philosophical discourses in twenty-first-century African education. In fact, as will become clear later, only communalism still enjoys wide currency.

Critical, Academic, or Professional Philosophy: The Content of African Philosophy of Education

If the component principles of African traditional education are either not uniquely African or are indefensible in (post) modern Africa, then what about current philosophical trends in educational discourse on the continent? According to Kwasi Wiredu, who like Adeyemi and Adeyinka is sympathetic to the African traditional conception of education, to be educated (in Akan, as in most other African languages and cultures) means to have knowledge (wisdom and skills), to be tolerant and willing to enter into dialog, and to possess moral maturity (Wiredu, 2004: 17, 18). For Wiredu, this normative conception of education is inspired by the following: indigenous (African) knowledge systems (Wiredu, 2004: 24), traditional African faith in consensus (Wiredu, 2004: 21), and the conceptual and normative priority of community over individuality (Wiredu, 2004: 20). African philosophy of education, thus “must combine all these considerations, which ... reveal the strengths of the traditional African conception of education” (Wiredu, 2004: 24).

The substratum for decolonization of the African mind and for “creating an educational vision capable of serving the legitimate interests of Africa in the contemporary world” (Wiredu, 2004: 24) is for Africans to (learn to) think and/or philosophize in their own language.

The present section takes stock of trends and developments presented both at recent conferences and in recent publications on or within African philosophy of education (see Higgs *et al.*, 2000; Odora Hoppers, 2002; Waghid *et al.*, 2005; in April 2007, the International Society of African Philosophy and Studies (ISAPS) 15th Annual Conference was hosted by the Rhodes University Philosophy Department in Grahamstown, South Africa, and more recently (23–25 October 2007), a conference on African philosophy and the future of Africa took place at St. Augustine College of South Africa). In essence, prevalent themes have been the following:

- the educational significance of indigenous (African) knowledge systems;
- *ubuntu/botbo/hunhu* and communalism;
- the ethical responsiveness of African philosophy (of education); and
- the legacy of colonialism.

The author will discuss each of these in turn, before focusing on an approach that arguably contains the possibility of Africa’s novel contribution to philosophy of education.

Indigenous (African) knowledge systems

The motivation for a focus on indigenous (African) knowledge is fairly easy to explain, especially when one considers the denigration, suppression, and exploitation of traditional knowledge systems during and even after colonialism. The reclamation project that underlies this renewed focus is not only epistemological but also concerned with legislation and social justice. As Mogobe Ramose has put it:

The history of epistemicide in South Africa raises fundamental questions of justice such as the question of epistemological equality of all the existing paradigms of the peoples of South Africa. Epistemological equality is a vital ingredient in the construction of a truly representative South African identity expressed, among others, in the new South African philosophy of education. (Ramose, 2004: 156)

A question that remains largely unaddressed is whether the idea of indigenous knowledge makes any sense. A central problem appears to be the lack of clarity about the meaning or understanding of knowledge. Defenders of this idea distinguish between skills and knowledge – which suggests, in the absence of any definition, that at least part of the understanding concerns propositional knowledge. Insofar as knowledge in this sense includes reference to truth, this invites the perception of the latter also being

indigenous. Bluntly asserting, on more than one occasion during the recent 2007 ISAPS conference, that truth is belief, Wiredu claimed that reference to “infallible” truth is not only a bar to dialogue but that “such a claim to knowledge is also a bar to education” (Wiredu, 2004: 24).

A problem that would need to be addressed is that of relativism (about both knowledge and truth) and of the implications of taking epistemological relativism seriously. A further question concerns the basis, if there is one, for distinguishing between knowledge and superstition within indigenous African belief systems. A common perception, in this regard, is that truth is nothing more than consensus.

African communalism, ubuntu, and consensus

Consensus is seen as desirable, and dissensus as undesirable, both on epistemic and political grounds. In traditional African societies, debate characteristically continues until a compromise is attained and all participants agree with the outcome (Metz, 2007a: 324). A central feature in African philosophy is the consensus-seeking principle provided by *ubuntu*, or humanness, the African principle of mutuality and interdependence. (According to Malegapuru Makgoba, the shared values that are fundamental features of African identity and culture . . . , for example, include hospitality, friendliness, the consensus and common framework seeking principle, *ubuntu*, and the emphasis on community rather than on the individual. These features typically underpin the variations of African culture and identity everywhere (Makgoba, 1997: 198).) According to Ramose, “African philosophy in general, and *ubuntu* philosophy particularly for South Africa, must be inscribed in the research agenda aimed at the construction of a new philosophy of education in South Africa” (Ramose, 2004: 158). *Ubuntu* expresses the view that a person is a person through other persons, or “I am because we are” – *umuntu ngumuntu nga bantu* (or *motbo ke motbo ka batbo*) – and is closely associated but not identical with African communalism.

Wiredu referred to African communalism as the foundation for national reconstruction. Traditional African society, he claimed, was communalist, founded on kinship relations, a system of reciprocity. “Morality is the adjustment of one’s interests to the community under a common guiding principle, like the Golden Rule. On this model, communalism might be characterized as the adjustment of the individual’s interests to those of the community” (Wiredu, 2007). After quoting Nkrumah – “African communalism is a form of socialism” – Wiredu claimed that Western communitarianism took root in individualist systems: “Western communitarianism is compatible with certain forms of cultural individualism, which African communitarianism is not” (Wiredu, 2007). Wiredu suggested that the term communalism therefore be reserved for the latter, but not the former. Communalism, he claimed, is the basis for a good and just society.

Apart from raising the question whether this move does not define communalism into morality, rather than see it as one amongst several orientations in ethics, Wiredu’s account appears to equate individualism with egoism or selfishness. One could ask, furthermore, whether Africa’s misery (see The legacy of colonialism below) might not in part be the result of the preoccupation on the African continent with communalism – which underlies obedience to authority, ancestors, traditional leaders, etc. – and arguably also of the tyranny of consensus. (The idea here is that palaver democracy, which aims at agreement, is considerably less democratic than a system – educational or political – that encourages dissent and critical interrogation.)

Problems that have been identified, with regard to *ubuntu* and communalism, include lack of specificity and action-guiding potential (Metz, 2007b: 342), their seeming incompatibility with deontological considerations, like human rights; as well as their alliance with, and proximity to, questionable (androcentric) practices and traditions. (In the latter regard, see Pitika Ntuli, on the validity and desirability of the *amaqhibikiza* system (a type of mentorship program among older and younger girls to ensure sexual abstinence until the latter are ready to take full control of their affairs) and of *ukublobwa kwezintombi* or virginity testing in girls (that seeks to achieve the goal of purity in the context of the spread of HIV/AIDS; Ntuli, 2002: 61, 62), and Ramose, on the compatibility of *ubuntu* and polygamy: that marriage should not of necessity be monogamous is one of the ancient practices of *ubuntu* philosophy (Ramose, 2002: 329).) Important insights, on the other hand, include the modeling of humanity, and reconciliation/forgiveness, all of which are contained in *ubuntu*. Most strikingly, the following reason was given by one of the victims during the truth and reconciliation hearings as to whether and why she would consider forgiving the perpetrator after his confession: if it means he gets his humanity back (see also Tutu, 1999: 35).

The ethical responsiveness of African philosophy

Wiredu began his plenary address by quoting Nkrumah’s view, on the occasion of Ghana’s independence, that this would be “meaningless unless it involves all of Africa” (Wiredu, 2007). Wiredu took this to encapsulate a preeminent sense of the practical value or worth of philosophy. Commitment to the view that philosophy should be relevant to the transformation of African societies, that is, in the service of ordinary Africans on the continent, in terms of an ethical commitment to their upliftment also characterized the writings and presentations by (Oladipo, 1992: 7, 28), Oladele Abiodun Balogun and Ezekiel Mkhwanazi. According to Oladipo, the task of African philosophers is to be committed to fulfilling their scholarly obligations to their societies. Their primary task should be to begin to create a tradition of thinking and discourse whose main

focus would be on issues affecting the interest and aspirations of the people (Oladipo, 1992: 28; Balogun, 2007: 12).

Such a commitment is surely important, one might respond – but so much for the work of African epistemologists, logicians, philosophers of science, etc. What may be a cause for concern is not only the instrumentalization of philosophy, in other words, the implicit or explicit denial that philosophy has any inherent value, but also the fact that philosophers are presented with a specific task. They are being prescribed with regard to what their intellectual labor should amount to or have as an object. Of course, one might argue that the cognitive, critical, and creative development of ordinary Africans has an important ethical dimension. Clearly it has, but the net effect of such a move would be the dilution of ethics and ethical. Perhaps this is a concern philosophers of education are happier to live with than pure philosophers.

The legacy of colonialism

In the 1990s, with globalization, free enterprise, and unprecedented availability of information, education, knowledge, and learning capacity were identified as key factors of development in Africa. In terms of global educational development, however, Africa remains the taillight, by some distance (Hofmeier and Mehler, 2004: 36–39, *Bildung*).

According to Táíwò (2007), colonialism is to blame for Africa's predicament (and this distinguishes colonialism on the African continent from colonialism elsewhere, that is, in countries that have since become industrial, financial, and even agricultural powerhouses) insofar as it failed to prepare Africa for modernity or modernism.

In this regard, however, it may be worthwhile to consider Axelle Kabou's counter-argument (Kabou, 1991), on the blameworthiness not only of autocratic rulers, traditional leaders, and corrupt elites but also (and especially) of ordinary Africans, that it was ordinary Africans who refused development, who rejected modernity. Moreover, one might ask whether Táíwò is talking about all colonial powers. As divergent as they were, did they all commit the same crime of nonrecognition? According to Táíwò, all had Africans occupying the lowest rung of the human ontological hierarchy. His response to Kabou would presumably be: How could ordinary Africans reject what they did not have access to? Yet, what makes Kabou's account so compelling (her failure to proceed more subtly notwithstanding) is that it takes ordinary African people seriously, as individuals and agents in their own right, and not just as helpless victims, will-less instruments manipulated by the powers that be.

The Promise of African Philosophy of Education

While African philosophy of education possibly errs in postulating distinctly African knowledge (systems), truth,

and values, its uniqueness is arguably constituted by its educational priorities – priorities that exist, at best, to a lesser extent (if at all) elsewhere.

A plausible view appears to be that African philosophy of education shares a range of concerns with philosophy of education elsewhere and that there is a distinctive set of concerns in African philosophy of education, arising from particular historical and sociopolitical circumstances. Thus, it might be claimed that African philosophy of education has different priorities to philosophy of education elsewhere. If it is correct to argue that philosophical and educational priorities will emerge from life experiences and from the ways these are socially articulated, then one might assume that, given that the life experiences of Africans on the African continent are commonly different from those of educators and learners elsewhere, the philosophical and educational priorities will also differ.

Given, for example, the experience of indigenous Africans of physical and mental oppression, it stands to reason that African philosophy of education would have as priorities matters of transformation and redress in education. If philosophical and educational concerns and priorities arise from different forms of social life, then those that have emerged from a social system in which a particular race or group has been subordinate to another must be suspect (see Horsthemke and Enslin, 2009). In addition, given the (especially vicious) history of physical and psychological colonization, it is plausible that one of the educational priorities will be to educate against development of a subordinate or inferior mindset, as well as against a victim and beggar mentality despite the continuing economic crisis and low level of economic growth. Given, too, the ravages of the disease on the continent, HIV/AIDS education has special resonance in (sub-Saharan) Africa: the pandemic does not only snatch away teachers; it also leaves children infected and/or orphaned. An additional priority arises with Africa's low literacy quotients. In many countries, the language of education is the official language of administration: English, French, or Portuguese, in which the majority of children and learners are not competent. Consequently, there exist few successes in learning; quality and efficiency suffer; and high repeat and dropout rates mean a squandering of available resources. While it does not follow that particular historical and socioeconomic circumstances yield or bestow automatic validation or justification of the content and objectives of African philosophy of education, ideas like decolonization of the African mind have a particular resonance here: going back to one's language to think about thinking, to examine one's own ways of conceptualization – in short, to philosophize (Wiredu, 2007).

The promise of African philosophy of education, then, has in part to do with context and locality – but not in terms of any exclusionist, hands-off approach. Rather, it appears to be plausible that the particular historical,

geographic, and sociocultural experiences of Africans give rise to particular priorities that shape African philosophical theory and practice – and also yield conceptual tools that are likely to enrich philosophy of education elsewhere: the idea of mental decolonization, the modeling of humanity, and reconciliation/forgiveness.

See also: Critical Theory and Pedagogy; Identity; Justice and Care; Philosophy and Educational Research; Race, Critical Race Theory and Whiteness; Social and Cultural Capital in Education.

Bibliography

- Adeyemi, M. B. and Adeyinka, A. A. (2003). The principles and content of African traditional education. *Educational Philosophy and Theory* 35(4), 374–382.
- Balogun, O. A. (2007). The task of African philosophers in 21st century Africa. *Paper Presented at the International Society for African Philosophy and Studies 15th Annual Conference*. Rhodes University, Grahamstown, 2–4 April.
- Higgs, P., Vakalisa, N., Mda, T., and Lumumba, N. A. (eds.) (2000). *African Voices in Education*. Lansdowne: Juta.
- Hofmeier, R. and Mehler, A. (eds.) (2004). *Kleines Afrika-Lexikon: Politik – Wirtschaft – Kultur*. Munich: C.H. Beck.
- Horsthemke, K. and Enslin, P. (2009). African philosophy of education: The price of unchallengeability. *Studies in Philosophy and Education* 28(3), 209–222.
- Kabou, A. (1991). *Et si l'Afrique refusait le développement?* Paris: L'Harmattan.
- Lölke, U. (2001). *Kritische Traditionen: Afrika. Philosophie als Ort der Dekolonisation*. Frankfurt: IKO – Verlag für Interkulturelle Kommunikation.
- Makgoba, M. W. (1997). *Mokoko: The Makgoba Affair. A Reflection on Transformation*. Florida Hills, FL: Vivlia.
- Metz, T. (2007a). Toward an African moral theory. *Journal of Political Philosophy* 15(3), 321–341.
- Metz, T. (2007b). The motivation for 'toward an African moral theory.' *South African Journal of Philosophy* 26(4), 342–346.
- Mkhwanazi, E. S. N. (2007). African philosophy: A call to be ethically responsive to the 'other'. *Paper Presented at the International Society for African Philosophy and Studies 15th Annual Conference*. Rhodes University, Grahamstown, 2–4 April.
- Ntuli, P. P. (2002). Indigenous knowledge systems and the African renaissance: Laying a foundation for the creation of counter-hegemonic discourses. In Odora Hoppers, C. (ed.) *Indigenous Knowledge and the Integration of Knowledge Systems: Towards a Philosophy of Articulation*, pp 53–66. Claremont: New Africa Books.
- Odora Hoppers, C. A. (ed.) (2002). *Indigenous Knowledge and the Integration of Knowledge Systems: Towards a Philosophy of Articulation*. Claremont: New Africa Books.
- Oladipo, O. (1992). *The Idea of African Philosophy: A Critical Study of the Major Orientations in Contemporary African Philosophy*. Ibadan: Hope Publications.
- Oruka, H. O. (1998). Sage philosophy. In Coetzee, P. H. and Roux, A. P. J. (eds.) *Philosophy from Africa*, 1st edn., pp 99–108. Cape Town: Oxford University Press Southern Africa.
- Oruka, H. O. (2002). Four trends in current African philosophy. In Coetzee, P. H. and Roux, A. P. J. (eds.) *Philosophy from Africa*, 2nd edn., pp 120–124. Cape Town: Oxford University Press Southern Africa.
- Ramose, M. (2004). In search of an African philosophy of education. *South African Journal of Higher Education* 18(3), 138–160.
- Ramose, M. B. (2002). The ethics of *ubuntu*. In Coetzee, P. H. and Roux, A. P. J. (eds.) *Philosophy from Africa*, 2nd edn., pp 324–330. Cape Town: Oxford University Press Southern Africa.
- Táiwò, O. (2007). Colonialism: A philosophical profile. *Paper Presented at the International Society for African Philosophy and Studies 15th Annual Conference*. Rhodes University, Grahamstown, 2–4 April.
- Tutu, D. M. (1999). *No Future without Forgiveness*. London: Rider.
- Waghid, Y., Van Wyk, B., Adams, F., and November, I. (eds.) (2005). *African(a) Philosophy of Education: Reconstructions and Deconstructions*. Stellenbosch: Department of Education Policy Studies, Stellenbosch University.
- Wiredu, K. (2004). Prolegomena to an African philosophy of education. *South African Journal of Higher Education* 18(3), 17–26.
- Wiredu, K. (2007). Social philosophy in postcolonial Africa: Some preliminaries. *Plenary Lecture Presented at the International Society for African Philosophy and Studies 15th Annual Conference*. Rhodes University, Grahamstown, 3 April.

Further Reading

- Akinpelu, J. A. (1987). *An Introduction to Philosophy of Education*. London: Macmillan.
- Appiah, K. A. (1992). *In my Father's House: Africa in the Philosophy of Culture*. New York/Oxford: Oxford University Press.
- Asante, M. K. (1987). *The Afrocentric Idea*. Philadelphia, PA: Temple University Press.
- Asante, M. K. (1988). *Afrocentricity: The Theory of Social Change*. Trenton, NJ: Africa World Press.
- Biko, S. (1978). *I Write What I Like*. London: Penguin.
- Bodunrin, P. O. (1981). The question of African philosophy. *Philosophy* 56, 161–179.
- Coetzee, P. H. and Roux, A. P. J. (eds.) (1998). *Philosophy from Africa*, 1st edn. Cape Town: Oxford University Press Southern Africa.
- Coetzee, P. H. and Roux, A. P. J. (eds.) (2002). *Philosophy from Africa*, 2nd edn. Cape Town: Oxford University Press Southern Africa.
- Enslin, P. and Horsthemke, K. (2004). Can *ubuntu* provide a model for citizenship education in African democracies? *Comparative Education* 40(4), 545–558.
- Eze, E. C. (ed.) (1997). *Postcolonial African Philosophy*. Cambridge, MA: Blackwell.
- Gyekye, K. (1997). *Tradition and Modernity: Philosophical Reflections on the African Experience*. New York: Oxford University Press.
- Hountondji, P. J. (1996). *African Philosophy: Myth and Reality*. Bloomington, IN: Indiana University Press.
- Waghid, Y. (2005). Revisiting the African-Africana philosophy of education debate: Implications for university teaching. *Journal of Education* 34, 127–142.

Feminism

L Stone, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Education feminism – Bounds and emphasizes the interdisciplinary presence of feminist theory and activism in professional education.

Feminisms – Named in the plural, rather than feminism, this has been used to highlight, theoretical difference, diversity, distinction, and significantly, differentiation.

Feminist movement – Defined across history as entailing both activism and theory.

Introduction

If one were to ask young Western women, particularly those white and middle class, about feminism today, many would respond that they are not feminists – and that there is no longer any need. This is because many women around the world have assumed leadership positions, girls are attending and doing well in school, and both have achieved some degree of security. These marks of equality mask and indeed belie daily struggles still being fought by women and girls, by minorities, the poor, and those residing especially in the third world. Issues, for instance, fear of violence, comparable occupational worth, and conditions and responsibilities of child rearing, are based in discrimination and affect countless females.

The article serves as a reminder that, historically, feminist endeavors were necessary and remain necessary today and feminists (both women and men) know this. Activism and theory exist everywhere. The present focus is on Western, primarily North American theory, rather than activism and theory across the planet (see Goodman, 2004). Acknowledged is the limitation of the author's experience and knowledge, but alongside is the important recognition that each nation, perhaps regions and cultures, have their own feminists whose work is represented herein, albeit very inadequately. Readers are admonished to explore their own lives through those who study and write specifically about them.

The stance of this article is strongly supportive of feminism. While anti-feminist sentiment is mentioned, how could one not support this basic idea: "[Feminism names] a movement for creating a society in which women can live . . . [full self-determined lives]" (Houston, 1996: 215). Feminist theory in what follows is characterized through a

thematic of differentiation, seen as strength. The meaning of differentiation expands across the article, always difference with various formulations of complexity. Today several aspects are central: first, potential of common projects across uncommon histories, experiences, even needs; second, continuity within differences across times and theories; third, commitments not only to individual self-interests but to other interests as well (see Stone, 1995). A central premise is that contemporary politics – the aim of theory – often is engaged across difference.

Overall, this is an overview of feminist theory for philosophy of education, its scholars, teachers, students, and interested public. For this audience, education is constituted of our professional world (institutions of education, undergraduate and graduate programs, teacher training, and the like) even as we draw from sources outside. Again a key premise is that ultimately our work in education must draw sustenance largely from sources within: educators must be agents of change.

Three differentiating frameworks are presented in this article. The first section takes up 'wave feminism,' cross-disciplinary in nature tying feminist theory and philosophy to inquiries that span sciences, social sciences, humanities, and arts. Another section overviews 'feminism in philosophy,' intradisciplinary in character organizing theory relative to modern philosophical traditions. The third section on 'feminism and education' turns specifically to potential frameworks in professional education. A final substantive section is especially important for philosophy of education as it considers the current state of feminist philosophy of education. Throughout what follows examples and exemplars are incorporated; also many references and some additional readings are included to stimulate further inquiry. Finally, the article has been enhanced through contributions from Allison Anders, Ann Diller, Susan Franzoza, Barbara Houston, Nel Noddings, and James Marshall.

To begin, one more framework initiates the discussion: it is the 'history of thought,' serving as a pre or supra frame. Surely it is historically accurate to suggest that across recorded time writings by and about women have been compared to those by and about men. These binary comparisons of difference – but not differentiation – typically established a hierarchy of men over women, of any aspect of life and culture identified with the male over that of the female. Feminist responses to these comparisons have assumed two basic forms, one protesting misogyny and the other promoting equality (rarely superiority). Philosophical treatments often have focused on concepts of human

nature, woman, and rationality. Today millennial writings by women are available but obviously this has not always been so (Mahowald, 1978, 1983; Waithe, 1986–95). In a significant study of patriarchy, Lerner offers the general observation of women's exclusion from "creating symbol systems, philosophies, science, [law, indeed] theory formation" (Lerner, 1986: 5). Referencing philosophy specifically, she points to conditions needed by women scholars: economic independence, educational opportunity, and role models (Lerner 2000; also Witt, 2000/2007).

Framework One: Wave Feminisms

A now-standard cross-disciplinary framework for describing feminism is waves. The general agreement on dates that encapsulate both activism and theory are these: the first wave in the nineteenth and early twentieth centuries, the second wave in the 1960s and 1970s, and the third wave to the present. Across the waves, movement has been toward differentiation, first from sameness to difference and then to dispersion in terms of identities, societal needs, and politics both between women and men and among women.

The first wave feminist theory developed in North America and Europe as part of the movement of liberal enlightenment. Jaggar (1983) pinpoints seventeenth-century England as the birthplace of modern feminism; the term is originally French from sociologist Charles Fourier (see Houston, 1996: 216). The initiating activist emphasis was women's suffrage, at first aligned with movements in abolition, temperance, and workers rights that involved both majority and minority women. Grants of franchise before and by 1920 were recognized as successful culmination of the first wave in various sites – in Britain, the USA, Western Europe, Scandinavia, Australasia especially New Zealand, and Russia. Historic hindsight perhaps indicates a stronger bond between activism and theory in this wave than in the subsequent ones.

Two exemplars illustrate themes of the first wave: Historic bookends, writings from Mary Wollstonecraft, and Simone de Beauvoir are now part of the feminist canon. In the late eighteenth century, in *Vindication of the Rights of Women* (1792, 2006) and *Thoughts on the Education of Daughters* (1787, 1977), Wollstonecraft (1777, 1787, 1792, 2006) builds on rights theory to posit the value of women's place in the private sphere while promoting public co-education (see Richardson, 2002). In mid-twentieth century, Beauvoir locates women's subordination in gendered culture, attends to psychological and sexual normality, and moves women more into the public sphere (see Andrew, 2003). Her classic text is *The Second Sex* (1949/1952). Of note, theoretical discussion continues every today of relationships of these women writers to Locke and Sartre respectively.

The second and third wave feminism/feminisms are developments from and responses to the first wave and to each other; significantly critique has become commonplace.

Once the valuing of diversity was established, several specific trends of differentiation have been apparent: a first is to add to the conventional first wave liberal formulation of women's equality, comparing them to men in terms of deserved resources and positions as well as acknowledged contributions and competencies. A central reconsideration is of nature and capacity (Kroløkke and Sørensen, 2006: 5–6). Part of the first trend also is to radicalize theory through first offshoots. New categorizations then constitute a second trend. By the early twenty-first century, listings account for groupings drawn from activist movements and ideologies, identity and identity politics, theoretical and academic traditions, and cultural and time-bound origins. Memberships and labels, writes Tong about the latter, remain useful as they not only highlight diversity but also point to a significant idea "that, like all other time-honored modes of thinking, feminist thought has a past as well as a present and a future" (Tong, 1998: 1). Examples of labels include socialist, psychoanalytic, eco-feminist, postmodern, and global feminism. A third trend is the emergence of nonmajority feminist theorizing. For example, in the USA groupings include Black, Womanist, and African-American, Latina and Chicana, Asian-American and Native as well as multicultural, and Lesbian feminisms (Tong, 1998, Kravetz and Marecek, 2001).

Differentiation continues into the present day, first, in extensions by Third World, global, and international feminisms based in particular interplays of factor effects (e.g., of colonialism, nationalism, religion, and gender) (Enns and Sinacore, 2001). Second, distinctions are emerging from young feminists in their responses to current societal conditions that they see as different from those past. Third, a feminist backlash and desire to retain more traditional but threatened gender practices is also manifest (Sommers, 1995). Finally, related to these developments, a post-feminist stance whose use is ambiguous is named. This ranges from broad rejection to broad inclusion of feminism, from no longer seeing any need to positing contributions from the lives of women for all (see Coppock, 1995). Still another tradition is post-structuralist through a variety of theoretical particularizations primarily from postmodern, continental writers and fellow travelers (e.g., Oliver and Walsh, 2004).

Across the last several decades, wave feminists have not only written individually and contributed to and been part of commentaries, but also have been collected in general and topical texts that have defined and furthered the broad theoretical field. Two early, classic examples of the latter are *This Bridge Called My Back: Writings by Radical Women of Color* (Moraga and Anzaldúa, 1981; also Anzaldúa and Keating, 2002) and *Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology, and Philosophy of Science* (Harding and Hintikka, 1983, 2003). The first exemplar section below spotlights writers from feminist theory in 'frameworks one and two'; see also the Further Reading section.

Framework Two: Feminism in Philosophy

Locating feminism in the discipline of philosophy, framework two is the basic organizational structure employed by the *Stanford Encyclopedia of Philosophy* (hereafter *Stanford*). As Tuana points out, a contributing influence in feminist philosophy has surely been philosophical training. In this framework, the basic theoretical debate over scholarly relationships of women with men and with other women continues. However, differentiation now becomes distinction as differences in traditions establish bounded approaches to shared concepts and topics and as diversity is most often the purview of individual philosophers (see also Haslanger and Tuana, 2003, 2004). Three approaches are central: analytic, continental, and pragmatist (*Stanford* also includes two intersections, analytic and pragmatist each with continental). Philosophical considerations of method, history, context, politics and social justice, as well as such categories as woman, gender, sex, identity, and subject are distinguished.

Analytic feminism has roots in the tradition of Anglo-American analytic philosophy, in the work of such important philosophers as Frege, Russell, and the early Wittgenstein. Significantly nonpositivist, Quine, Austin, and the later Wittgenstein are important current sources. Most theoretical emphasis has been in epistemology, philosophy of science, and metaphysics and less in political and moral philosophy. Method, subareas of philosophy, and standpoints also are significant. Garry writes that out of the analytic tradition, topics taken up by feminists have been naturalized, socialized, and otherwise have modulated earlier, more abstract treatments (Garry, 2004: 2). Among well-known members are Code, Harding, and Logino.

Continental feminism has roots in the tradition of European continental philosophy developing out of Kant, Hegel, Nietzsche, Freud, and Heidegger. It is itself constituted as a set of evolving subtraditions that include phenomenology, structuralism, and deconstructionism. Like classical pragmatism to follow, each principal theorist can be understood as writing particularist social theory. Of course, epistemology has mattered but perhaps more so have ethics and politics. Cahill (2005) writes that Foucault and Derrida have had the strongest recent influence. Among its members are earlier, Arendt, and more contemporary, Kristeva and Irigaray.

Pragmatist feminism has its roots in the American classical pragmatism of Dewey, James, and Addams as well as progressives such as Gilman and Cooper. Whipps characterizes the approach thus: “For these thinkers, philosophizing was an active process, both as a way to change social realities and to use experience to modify the philosophies themselves” (Whipps, 2004: 1). In this tradition, theory and practice are especially unified; epistemology is dominant, used to further justice and democracy. Its

members include Seigfried, Minnich, and Sklar. Of the three *Stanford* approaches to feminism in philosophy, this is the least developed and known.

Recent Exemplars One

Much contemporary writing by feminist theorists, and especially philosophers, is published worldwide, making it visible and influential. In this section, the writings of three exemplars point to rich theoretical differentiation. Each feminist, in her own way, takes up issues such as race and ethnicity, sex and gender, colonial and postcolonial cultures. Each writes from a feminist orientation and also contributes to theory broadly. All reside in the USA but have strong international appeal. The three are Bell Hooks, Judith Butler, and Gayatri Chakravorty Spivak.

First, American hooks writes on feminism in the context of cultural criticism and activism. In an early work, *Feminist Theory: From Margin to Center* (hooks, 1984), she focuses on relations between black and white women. She asserts, “[privileged] feminists have largely been unable to speak to, with, and for diverse groups of women because . . . [they cannot take the interrelationships of various oppressions] seriously” (Hooks: 14). Later, generally inspired by Freire, she is more inclusive, theorizing across diverse groups for a pedagogy of hope (Hooks, 2000; 2003). Second, in more technical writings from the nineties, American Butler is very instrumental in fueling new movements such as queer and transgender politics (see (Kroløkke and Sørensen, 2006; also Salih, 2004). Writing initially on Hegel, much of her theorizing shows a debt to Foucault; the major text is *Gender Trouble: Feminism and the Subversion of Identity* (see Butler, 1990a). In an essay on pornography and feminist politics, her critical warning is this: “[The real can be understood] as a variable construction which is always and only determined in relation to its constitutive outside: fantasy, the unthinkable, the unreal” (Butler, 1990b, 2004, pp. 185–186). Third, Spivak continues interests in politics as hooks and Butler and the latter’s technical philosophy. Indian born, she was first to translate into English and has been influenced by Derrida. In her postcolonial writings, emblematic in “Can the Subaltern Speak?” she offers, “In seeking to learn to speak to (rather than listen to or speak for) . . . [this] historically muted subject . . . the postcolonial intellectual systematically ‘unlearns’ female privilege” (Spivak, 1988, 1994, p. 91, emphasis in original; also Spivak 1984, 1990; and Landry and Maclean, 1996). For each exemplary feminist, theory facilitates acceptance of provisional certainties in lives both local and global.

Framework Three: Feminist Theory and Education

Feminist theory in education generally has aligned itself with frameworks, trends, and categories in the larger

academy. The principal difference, of course, is a focus on education. Feminists have focused on public and professional activism, others on empirical gender studies, and still others on higher and adult education or pre-collegiate schooling. Writings in feminist theory span the frameworks discussed above, from histories of women's education to specific philosophical connections. Writings on education have been found across time while modern feminist theorizing was initiated by educators beginning in the 1970s. A recent statement in *Stanford*, from above, acknowledges the latter history. By the early 1980s, as Phillips puts it, "feminist philosophers of education were finding their voices" (Phillips, 2008). In this volume on philosophy of education, two potential categorizations for intersections of feminist theory and education are education feminism and feminist philosophy of education.

The category of education feminism was coined in the interdisciplinary text, *The Education Feminism Reader*, to call attention to the place of feminism in professional education (Stone, 1994; Dillabough, 2006; see also Stone, 1997). It remains today the only collection of feminist theory in professional education. To name this theoretical domain, the book's introduction makes several claims: one, a field exists in its own right; two, its scholarship is both derivative and originaive as is all scholarship; and three, professional education incorporates topics related to feminism and also to the discourses, organizations, and practices of education itself. In the text, categories include self and identity, education and schooling, knowledge, curriculum and instructional arrangements, teaching and pedagogy, and diversity and multiculturalism. Although a second volume of readings has not been published, since the mid-1990s the field has continued to develop and evolve. Differentiation not only follows the waves above but also as part of subdisciplines in education such as women's history of education, feminist pedagogy, social sciences theorizing, and research methodology. Majority and minority feminists continue to publish with the latter demonstrating significant development (see, e.g., Bernal *et al.*, 2006).

In a recent essay, Nel Noddings (2009) prefers two centers of concentration in feminist philosophy of education. Cutting through waves theory, these are women's oppression and women's agency. Within each she identifies suborganizations to retain "a kind of philosophy, not a female or feminine activity" (Noddings, citing Walker, 1995, draft p. 2). In oppression, Noddings provides useful discussion of discrimination and potential danger for feminist philosophy and its practitioners, from writings on women's experiences and traditional domains within "the world created by men" (p. 9), and in critiques of the philosophy and of science. Her feminist agency incorporates emphasis on women's cultures, specifically in attending to relation, responding to needs, facilitating familial care, and accepting social responsibility. In her own significant contribution, feminist philosophy and education come together in

connections between care theory and moral education. For this article, feminist philosophy of education might usefully be defined as a subfield of education feminism.

Recent Exemplars Two

Finding voice in feminist writings in education, three American mothers are widely known: Maxine Greene, Jane Roland Martin, and Noddings. They have led now more than one generation of feminists writing in education; indeed, writing in philosophy of education has grown exponentially into the twenty-first century (see Leach, 1991). Each writes critically about education and schooling and each remains scholarly and professionally active today.

Greene is known for her interests in the arts, multiculturalism, pedagogy, and curriculum, and to value the lived lives of all (see Greene, 1978, 1988). Her acknowledged roots are in Dewey, existentialist writings, and often literature. In a 1988 chapter promoting public lives of women, a key theme of freedom links literature, poetry, and specific histories of women leaders. She writes, "[freedom] cannot be conceived apart from a matrix of social, economic, cultural, and psychological conditions. It is within the matrix that selves . . . are created through choice of action in . . . changing situations" (Greene, 1988: 80). Culminating decades of writing, in a 2003 piece co-authored with Morwenna Griffiths, she turns to the personal: "I find myself often perplexed by what feminism signifies for me today – although I am fearfully aware of the old men of the sea, the survival of patriarchal warnings underlying and often crowding out the sound of whatever little story I have to tell" (Greene and Griffiths, 2003: 80).

Distinct from the existential, pragmatist, and phenomenological orientations of Greene and next, Noddings, Martin brings skills from analytic philosophy to women's history within philosophy of education. In addition to the education of women, her interests include broad questions of liberal education, schooling and curriculum. Martin's (1985, 1992) analysis, *Reclaiming a Conversation: The Ideal of the Educated Woman* (1985) remains today a singularly unique contribution. This was followed by an important text, *Schoolhome* (1992); recent writings concern issues of identity and culture (Martin, 2007). In the first text, she establishes a connected conversation through Plato, Rousseau, Wollstonecraft, Beecher, and Gilman. Her position is this: "If conversation about women's education is to be incorporated in the history of educational thought, the definition of that discipline's subject matter must be expanded to include the processes of society with which women's lives have . . . been intertwined" (Martin, 1985: 179).

Noddings's (1984, 1989b) feminism is best known through her career contribution to an ethics of care. Additional interests include curriculum and schooling reform, with recent attention to the potential of women's lives and

cultures. Her classic book, *Caring* (1984) is subtitled 'A feminine approach'. By her next work, *Women and Evil* (1989b), she continues to pay attention to women's experience through working along with and responding to feminist theory. In a 1989 piece republished in 2002, she offers a direct statement of her feminism, naming it social or maternal with forebears that are radical and psychoanalytic. She explains, "This perspective acknowledges that there are still substantial differences between men and women including differences in their views on moral life . . . [Many of them are traced] to centuries of different experience" (Noddings, 1989a, 2002, pp 102–103). As indicated, part of her project is to highlight significant elements of women's lives that "may help all of us lead better lives" (p. 107).

Current Considerations

At the outset of this article a present belief of Western, privileged, and largely majority women that repudiates feminism is reported. This occurs alongside the continued flourishing of feminist theory in the academy broadly and its development in education. Focusing on feminist philosophy of education, a current status check reveals three central aspects: First, feminist theorists in education read their sisters in the academy but the family is dysfunctional and there is little reciprocity. Second, the mothers largely remain emblematic although there is some recognition of progeny and of theoretical dispersion. Third, the state of feminism in philosophy of education is still relatively weak in a field in which shared recognition and governance with male colleagues is still emerging.

A focused comparison between feminist philosophy and feminist philosophy of education assists current assessment. Once again, a USA context is presented, acknowledging its limitations. Feminism in the academy is certainly here to stay even with occasional battles over Women's Studies programs. In philosophy, even minority numbers as faculty cannot obviate the sheer number of potential positions in collegiate departments. Of significance is the Society for Women in Philosophy, with active groups in all divisions of the American Philosophical Association (with a British counterpart). Other philosophical associations have a strong women's presence as well. Moreover, publication by feminist philosophers is renowned everywhere both by individuals and collectivities. Over 500 references in the *Stanford Encyclopedia of Philosophy* referenced above is itself testament. These references, publications, and faculty interests also point to field differentiation found across disciplines.

With growing numbers of publications and presentations since the late 1970s, the state of feminist philosophy of education is nonetheless much less prominent than in the discipline. Finding voice has resulted in some key placements in philosophy of education but most of the positions are filled by men. Feminist philosophers most often teach

in other education domains, particularly teacher or perhaps multicultural education. Another positive sign is leadership presence; women presidents of the Philosophy of Education Society (of North America) have become nearly commonplace. There is also a women's caucus. Many education associations are led by women where feminism appears not to matter. One arena in which feminist presence seems minimal is in specific courses or as notable inclusion. It might even be that there is less reading and studying of feminist theory today than in recent decades. Given girls' success in schooling, gender is passed over in favor of less controversial issues.

Publication presence is mixed. Most of the significant work has come from the mothers although second- and third-generation scholars are producing theoretically important work (e.g., Todd, 1997; Boler, 2003). Recent collections in philosophy of education have included a few women authors but the general pattern is one feminist piece per collection. Kohli's (1995) text, *Critical Conversations in Philosophy of Education*, is an exception (see Blake *et al.*, 2003; Curren, 2007; Siegel, 2009; also Kohli and Burbules, 2009). The most promising development is that today's feminist philosophers of education have diverse theoretical backgrounds and interests that mirror academic counterparts – feminist sisters could well learn from them! (see Li, 1993; Rice, 1999/2008; Titone and Maloney, 1999; Enslin, 2003; Stone, 2004). This seems a mark of field maturity, although one notes that there are few collections that substantiate the larger academic field.

Conclusion

Feminism, as this article attests, needs to be seen as central to intellectual and political life and to philosophy and philosophy of education. As frameworks indicate, it is characterized by various forms of historically evolving theoretical differentiation. What is significant here, of course, is that girls and women must continue to struggle both in the west and elsewhere; a world free of gender bias remains an ideal. Feminist theory in the academy and in education itself has much to contribute to its realization.

Bibliography

- Andrew, B. (2003). Beauvoir's place in philosophical thought. In Card, C. (ed.) *The Cambridge Companion to Simone de Beauvoir*, pp 24–44. Cambridge: Cambridge University Press.
- Anzaldúa, G. and Keating, A. (eds.) (2002). *This Bridge We Call Home: Radical Visions for Transformation*. New York: Routledge.
- Bernal, D., Elenes, C. A., Godinas, F., and Villenas, S. (eds.) (2006). *Chicana/Latina Education in Everyday Life: Feminista Perspectives on Pedagogy and Epistemology*. Albany, NY: State University of New York Press.
- Blake, N., Smeyers, P., Smith, R., and Standish, P. (2003). *The Blackwell Guide to the Philosophy of Education*. Malden, MA: Blackwell.

- Boler, M. (2003). *Pedagogies of Difference: Rethinking Education and Social Change*. New York: RoutledgeFalmer.
- Butler, J. (1990a). *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge.
- Butler, J. (1990b, 2004). The force of fantasy: Feminism, mapplethorpe, and discursive excess. In Salih, S. (ed.) *The Judith Butler Reader*, pp 183–203. Malden MA: Blackwell.
- Cahill, A. (2005). Continental feminism. *Stanford Encyclopedia of Philosophy*. <http://www.platostanfordedu/entries/femapproach-continental/> (accessed September 2008).
- Coppock, V. (1995). *The Illusions of 'Post-feminism': New Women, Old Myths*. London: Taylor and Francis.
- Curren, R. (ed.) (2007). *Philosophy of Education: An Anthology*. Malden, MA: Blackwell.
- Dillabough, J. (2006). Education feminism(s). Gender theory and social thought: Illuminating moments and critical impasses. In Skelton, C., Francis, B., and Smulyan, L. (eds.) *The Sage Handbook of Gender and Education*, pp 47–62. London: Sage Publications.
- Enns, C. and Sinacore, A. (2001). Feminist theories. In Worell, J. (ed.) *Encyclopedia of Women and Gender: Sex Similarities and Differences and the Impact of Society on Gender*, vol. 1, pp 469–480. San Diego, CA: Academic.
- Enslin, P. (2003). Liberal feminism, diversity, and education. *Theory and Research in Education* 1(1), 73–87.
- Garry, A. (2004). Analytic feminism. *Stanford Encyclopedia of Philosophy*. <http://www.platostanfordedu/search/searcherpy?query=analytic+feminism> (accessed September 2008).
- Goodman, R. (2004). *World, Class, Women: Global Literature, Education, and Feminism*. London: RoutledgeFalmer.
- Greene, M. (1978). *Landscapes of Learning*. New York: Teachers College Press.
- Greene, M. (1988). *The Dialectic of Freedom*. New York: Teachers College Press.
- Greene, M. and Griffiths, M. (2003). Feminism, philosophy and education: imagining public spaces. In Blake, N., Smeyers, P., Smith, R., and Standish, P. (eds.) *The Blackwell Guide to Philosophy of Education*, pp 73–92. Malden, MA: Blackwell.
- Harding, S. and Hintikka, M. (1983, 2003). *Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology, and Philosophy of Science*. Dordrecht: Kluwer Academic.
- Haslanger, S. and Tuana, N. (2003, 2004). Topics in feminism. *Stanford Encyclopedia of Philosophy*. <http://www.platostanfordedu/entries/feminism-topics/> (accessed September 2008).
- Hooks, B. (1984). *Feminist Theory: From Margin to Center*. Boston: Southend.
- Hooks, B. (2000). *Feminism for All*. Boston: Southend.
- Hooks, B. (2003). *Teaching Community: A Pedagogy of Hope*. New York: Routledge.
- Houston, B. (1996). Feminism. In Chambliss, J. (ed.) *Philosophy of Education: An Encyclopedia*, pp 215–220. New York: Garland.
- Jaggar, A. (1983). *Feminist Politics and Human Nature*. Totowa, NJ: Rowman & Allanheld.
- Kohli, W. (ed.) (1995). *Critical Conversations in Philosophy of Education*. New York: Routledge.
- Kohli, W. and Burbules, N. (2009). *Feminism and Educational Research*. Lanham, MD: Rowman & Littlefield.
- Kravetz, D. and Marecek, J. (2001). The feminist movement. In Worell, J. (ed.) *Encyclopedia of Women and Gender: Sex Similarities and Differences and the Impact of Society on Gender*, vol. 1, pp 457–468. San Diego, CA: Academic.
- Królókke, C. and Sørensen, A. (2006). Three waves of feminism: From suffragettes to girls. *Gender Communication Theories & Analyses: From Silence to Performance*, pp 1–23. Thousand Oaks, CA: Sage.
- Landry, D. and MacLean, G. (eds.) (1996). Introduction: Reading spivak. In *The Spivak Reader: Selected Works of Gayatri Chakrovorty Spivak*, pp 1–13. New York: Routledge.
- Leach, M. (1991). Mothers of (inter)vention: Women's writing in philosophy of education. *Educational Theory* 41(3), 287–300.
- Lerner, G. (1986). The creation of patriarchy. *Women and History*, vol. 1. New York: Oxford University Press.
- Lerner, G. (2000). Why have there been so few women philosophers? In Tougas, C. and Ebenreck, S. (eds.) *Presenting Women Philosophers*, pp 5–14. Philadelphia: Temple University Press.
- Li, H. (1993). A cross-cultural critique of ecofeminism. In Gaard, G. (ed.) *Ecofeminism: Women, Animals, Nature*, pp 272–294. Philadelphia: Temple University Press.
- Mahowald, M. (1978, 1983). *Philosophy of Woman: An Anthology of Classic and Current Concepts*, 2nd edn. Indianapolis, IN: Hackett.
- Martin, J. (1985). *Reclaiming a Conversation: The Ideal of the Educated Woman*. New Haven: Yale University Press.
- Martin, J. (1992). *The Schoolhome*. Cambridge: Harvard University Press.
- Martin, J. (2007). *Educational Metamorphoses: Philosophical Reflections of Identity and Culture*. Lanham, MD: Rowman & Littlefield.
- Moraga, C. and Anzaldúa, G. (1981). *The Bridge Called My Back: Writings by Radical Women of Color*. Watertown, MA: Persephone.
- Noddings, N. (1984). *Caring: A Feminine Approach to Ethics and Moral Education*. Berkeley: University of California Press.
- Noddings, N. (1989a, 2002). *Educating moral people*. In *Educating Moral People: A Caring Alternative to Character Education*.
- Noddings, N. (1989b). *Women and Evil*. Berkeley: University of California Press.
- Noddings, N. (2009). Feminist philosophy and education. In Siegel, H. (ed.) *Oxford Handbook of Philosophy of Education*. Oxford: Oxford University Press.
- Oliver, K. and Walsh, L. (eds.) (2004). *Contemporary French Feminism*. New York: Oxford University Press.
- Phillips, D. (2008). Philosophy of education. In *Stanford Encyclopedia of Philosophy*. <http://www.platostanfordedu/search/searcherpy?query=philosophy+of+education> (accessed September 2008).
- Rice, S. (1999/2008). Feminism and philosophy of education. In *Encyclopaedia of Philosophy of Education*. http://www.ffst.hr/ENCYCLOPAEDIA/doku.php?id=feminism_and_philosophy_of_education (accessed September 2008).
- Richardson, A. (2002). Mary Wollstonecraft on education. In Johnson, C. (ed.) *The Cambridge Companion to Mary Wollstonecraft*, pp 24–41. Cambridge: Cambridge University Press.
- Salih, S. (ed.) (2004). Introduction. In *The Judith Butler Reader*, pp 1–17. Malden, MA: Blackwell.
- Siegel, H. (ed.) (2009). *Oxford Handbook of Philosophy of Education*. Oxford: Oxford University Press.
- Sommers, C. H. (1995). *Who Stole Feminism? How Women Have Betrayed Women*. New York: Touchstone.
- Spivak, G. (1984, 1990). Criticism, feminism, and the institution. In Grosz, E. (ed.) *The Post-Colonial Critic: Interviews, Strategies, Dialogues*, pp 1–16. New York: Routledge.
- Spivak, G. (1988, 1994). Can the subaltern speak? In Williams, P. and Chrisman, L. (eds.) *Colonial Discourse and Post-Colonial Theory: A Reader*, pp 66–111. New York: Columbia University Press.
- Stone, L. (ed.) (1994). *The Education Feminism Reader*. New York: Routledge.
- Stone, L. (1995). Narrative in philosophy of education: A feminist tale of “uncertain” knowledge. In Kohli, W. (ed.) *Critical Conversations in Philosophy of Education*, pp 173–189. New York: Routledge.
- Stone, L. (1997). Feminism. In Grant, C. and Ladsen-Billings, G. (eds.) *Dictionary of Multicultural Education*, pp 119–122. Phoenix, AZ: Oryx.
- Stone, L. (2004). Julia Kristeva's ‘mystery’ of the subject in process. In Marshall, J. (ed.) *Poststructuralism, Philosophy, Pedagogy* pp 119–139. Dordrecht: Kluwer.
- Titone, C. and Maloney, K. (eds.) (1999). *Women's Philosophy of Education: Thinking Through Our Mothers*. Upper Saddle River, NJ: Merrill.
- Todd, S. (1997). *Learning Desire: Perspectives on Pedagogy, Culture, and the Unsaid*. New York: Routledge.
- Tong, R. (1998). *Feminist Thought: A More Comprehensive Introduction*, 2nd edn. Boulder, CO: Westview.
- Waihe, M. (ed.) (1986–1995). *A History of Women Philosophers* (4 vols). Dordrecht: Kluwer Academic.
- Walker, R. (1995). *To be Real: Telling the Truth and Changing Feminism*. New York: Anchor.
- Whipps, J. (2004). Pragmatist feminism. In *Stanford Encyclopedia of Philosophy*. <http://www.platostanfordedu/search/searcherpy?query=pragmatist+feminism> (accessed September 2008).
- Witt, C. (2000, 2007). Feminist history of philosophy. *Stanford Encyclopedia of Philosophy*.

- Wollstonecraft, M. (1787, 1977). Thoughts on the education of daughters. In Todd, J. (ed.) *A Wollstonecraft Anthology*, pp 27–39. Bloomington, IN: Indiana University Press.
- Wollstonecraft, M. (1792, 2006). *A Vindication of the Rights of Women*. New York: Penguin.

Further Reading

- Alexander, M. J. and Mohanty, C. (eds.) (1997). *Feminist Genealogies, Colonial Legacies, Democratic Futures*. New York: Routledge.
- Annot, M. and Ghaill, M. (eds.) (2006). *The RoutledgeFalmer Reader in Gender and Education*. London: Routledge.
- De Beauvoir, S. (1949/1952, 1953). *The Second Sex*. New York: Knopf.
- Fricke, M. and Hornsby, J. (2000). *The Cambridge Companion to Feminism in Philosophy*. Cambridge: Cambridge University Press.
- Gatens, M. (1991). *Feminism and Philosophy: Perspectives on Difference and Equality*. Bloomington, IN: Indiana University Press.
- James, J. and Sharpley-Whiting, T. D. (eds.) (2000). *The Black Feminist Reader*. Malden, MA: Blackwell.
- Lloyd, G. (1993). *The Man of Reason: 'Male' and 'Female' in Western Philosophy*. Minneapolis, MS: University of Minnesota Press.
- Luke, C. and Gore, J. (eds.) (1992). *Feminisms and Critical Pedagogy*. New York: Routledge.
- McCann, C. and Kim, S. (eds.) (2003). *Feminist Theory Reader: Local and Global Perspectives*. New York: Routledge.
- Mohanty, C., Russo, A., and Torres, L. (eds.) (1991). *Third-World Women and the Politics of Feminism*. Bloomington, IN: Indiana University Press.
- Naples, N. (1998). *Community Activism and Feminist Politics: Organizing Across Race, Class, and Gender*. New York: Routledge.
- Narayan, U. and Harding, S. (eds.) (2000). *Decentering the Center: Philosophy for a Multicultural, Postcolonial, and Feminist World*. Bloomington, IN: Indiana University Press.
- Okin, S. (2002, 2007). Mistresses of their own destiny: Group rights, gender, and realistic rights of exit. In Curren, R. (ed.) *Philosophy of Education: An Anthology*, pp 134–148. Malden, MA: Blackwell.
- Sādāwī, N. (1997). *The Nawal Swaadawi Reader*. New York: St Martin's.
- St Pierre, E. and Pillow, W. (eds.) (2000). *Working the Ruins: Feminist Poststructural Theory and Methods in Education*. New York: Routledge.
- Tuana, N. (2004). Approaches to feminism. *Stanford Encyclopedia of Philosophy*. . <http://www.plato.stanford.edu/search/searcher.py?query=approaches+to+feminism> (accessed September 2008).
- Tuana, N. and Tong, R. (eds.) (1995). *Feminism & Philosophy: Essential Readings in Theory, Reinterpretation, and Application*. Boulder CO: Westview.
- Weiler, K. (ed.) (2001). *Feminist Engagements: Reading, Resisting, and Revisioning Male Theorists in Education and Cultural Studies*. New York: Routledge.

Hermeneutics

C Bingham, Simon Fraser University, Burnaby, BC, Canada

© 2010 Elsevier Ltd. All rights reserved.

Hermeneutics, in its most general meaning, is the art of interpretation. And while philosophy of education has only recently taken up hermeneutics with keen interest, it might be said that it has always been educational, in the sense that one way to become educated about something is to learn how to interpret it. That said, philosophical hermeneutics, especially as has been articulated by Hans-Georg Gadamer, has greatly enriched educational philosophy (Gadamer, 1994, 1976). Gadamer's account of hermeneutics goes far beyond the art of interpretation, offering insight into a central educational concern – human understanding. In what follows, the educational significance of Gadamer's hermeneutics will be detailed. In addition, precedents and related hermeneutic theories will be sketched.

Hermeneutic Precedents

Any educational account of hermeneutics should first take into consideration its rich history. Hermeneutics had its beginnings in, and still owes much of its educational importance to, the interpretation of texts and other symbolic messages. Those beginnings date back to ancient philosophy, with a primary example being Aristotle's work, *On Interpretation*, or *Peri Hermeneias* in Greek. The word itself derives from the Greek *hermeneuein*, meaning to interpret, translate, or express. As an example of its Greek use, one finds Plato calling poets the interpreters (*hermenes*) of the gods. In addition, hermeneutics has long been associated with religious and legal interpretation. Thus, one can place such a figure as Philo of Alexandria (ca. 20 BC–ca. AD 50), who allegorized Old Testament scriptures, in the tradition of hermeneutics. One can also place Christian thinker St. Augustine (354–430), medieval interpreters of the Torah, as well as Protestant reformer Martin Luther (1483–1546) in this tradition.

However, with the advent of modern hermeneutics, especially since the Romantic period, the scope and function of hermeneutics has widened to include elements that go beyond symbolic interpretation. German philosopher Friedrich Schleiermacher (1768–1834) was primary in establishing a system of hermeneutic thought including interpretations of all sorts – religious as well as secular – and widening hermeneutics to include psychological, social, historical as well as linguistic elements (Schleiermacher, 1998). Schleiermacher, who is an influential figure in Gadamer's central text, *Truth and Method*, lectured on

hermeneutics between 1805 and 1833. Among the many hermeneutic themes stressed by him, the following are especially significant: first, the interpretation of an author's words is both a matter of language usage and a matter of individual meaning. That is, one cannot correctly interpret the work of an author by simply knowing what the words he or she used meant at the time they were written. How the author intends words to be is as important as what the words mean in general. Second, misunderstanding a text is likely. It is not likely that the interpreter will understand a text through-and-through given historical, cultural, and psychological differences between author and reader. Third, hermeneutics can be applied to texts of all sorts – secular as well as religious, poetic as well as nonpoetic, foreign-language as well as first-language. Implicit in this third statement is that misunderstanding happens even when communicants share the same language. Fourth, the interpreter may not rely on higher principles such as divine inspiration to guide his or her interpretation, even where religious texts are concerned. These themes, and others, demonstrate how Schleiermacher made hermeneutics into a more general, holistic art of interpretation.

Generally acknowledged as the prime inheritor, and major innovator, of Schleiermacher's hermeneutics, Wilhelm Dilthey (1833–1911) contributed to the philosophical project of interpretation with a general theory of human existence (Dilthey, 2002). According to his understanding of human existence, and especially in his later works, Dilthey introduced a two-pronged description of existence in terms of *erlebnis* and *verstehen*. Regarding *erlebnis* or life, H P Rickman has noted, "The principle that we cannot go behind life is a keystone of Dilthey's whole philosophic position, which is usually described as a philosophy of life" (Rickman, 1963: 15). Whether one is undergoing a humanistic, social scientific, or physical scientific interpretation, one can never escape the horizon of life in order to study that horizon. One's entire experience of life is always implied in every act of interpretation.

A discussion of Schleiermacher and Dilthey brings us to the ontological hermeneutics of Martin Heidegger (1889–1976) (Heidegger, 1962). If Dilthey can be said to have universalized hermeneutics by giving it grounding in the totality of human experience, then Heidegger can best be described as radicalizing this grounding to the point where hermeneutic interpretation is, itself, human experience. By way of describing this radicalization, it is useful to draw on the often-used metaphor of the hermeneutic circle. For Schleiermacher and Dilthey, the hermeneutic circle is

a way of describing the interpretive method of oscillating between the part and the whole. Following this method, the reader (or any interpreter) must interpret each passage (the part) only in relation to the entire work (the whole). He or she must, in turn, interpret a work (the part) only in relation to an entire opus (the whole). And with regard to meaning, the reader (again, any interpreter) must take care to consider the author's intention (the part) as just one element of a larger social/historical context (the whole). Once this circular method has been used, one is finished with the hermeneutic circle until the next exercise in interpretation. In contrast, Heidegger claims that humans are always interpretive beings. Being, itself, gets enacted interpretively. As such, it does not make sense to say that one can step out of the hermeneutic circle since hermeneutic understanding is something human beings do even prior to employing a particular hermeneutic method. This ontological condition of already being inside the hermeneutic circle will be continued in the work of Hans-Georg Gadamer. In fact, Gadamer's entire project of philosophical hermeneutics is most indebted to the work of his teacher, Martin Heidegger.

Heidegger's hermeneutics articulates a difference between the Cartesian objectification-through-reason and a more fundamental, tacit experience of the world that typifies human understanding. Human beings, as part of what it means to be human, find themselves already thrown into a world where things, actions, and intentions are pre-understood. This is why human beings are already interpretive beings. Human beings have pre-conceptions, pre-understandings, and pre-interpretations that are not, under everyday circumstances, subject to questioning. Thus, a hammer, for example, is already pre-understood, pre-interpreted as a hammer, but as such its hammer-ness is not thrown into question. There may, however, be a point when the hammer ceases to function as a hammer. In such a situation, the hammer's hammer-ness may be thrown into question. And so it goes with interpretation of texts and human utterances. Heidegger brought to the hermeneutic project a keen awareness of the pre, of the tacitness of most human experience including the tacitness of textual interpretation. Thus, the question of interpretation is not whether to interpret or how to interpret. It is not a matter of applying some method to the object under interpretation. It is rather a matter of how to foreground, to subject to question, one's tacit interpretations.

Also central to Heidegger's contribution to philosophical hermeneutics is his concern with language itself, with its place in the landscape of human being, rather than with its supposed role as a symbol that conveys human meaning. Heidegger's famous statement that "language is the house of being" intimates this linguistic concern (Heidegger, 1999). At least since Aristotle's *Peri Hermeneias*, language had been seen as a symbol system to be interpreted in order to find an author's, or speaker's,

meaning. Especially in his later work, Heidegger turned this instrumental understanding of language around, asserting that human meaning-making dwells within language. Language is not a tool used by human beings. It is rather a way that human beings turn together to understand something. Language is thus a way of being together while paying attention to how the world presents itself to us. Language is not something to get through in order to interpret the more serious matter of an author's, or speaker's, meaning.

Related Hermeneutic Theories

Given the ancient and modern precedents of hermeneutics, it is possible to note some general hermeneutic themes. These themes will be helpful for understanding various branches of recent hermeneutic thought. First, hermeneutics has long been concerned with how the interpreter might throw a text or utterance into question and thereby problematize the received interpretations of that text or utterance. Second, hermeneutics has been concerned, to varying degrees, with how the world-historical situation of a text or utterance informs one's interpretation of that text or utterance. Third, hermeneutics has paid close attention to the role of language during the interpretive act. Fourth, hermeneutics has long been concerned with the uses and abuses of disciplinary thought. While it began as a practice of textual interpretation focusing mainly on religious, poetic, and legal texts, hermeneutics has been universalized as of late. These themes can be used to distinguish between four major branches of modern/postmodern hermeneutics: conservative, critical, radical, and moderate.

Conservative Hermeneutics

Most commonly identified with the work of E D Hirsch Jr. and Emilio Betti, conservative hermeneutics traces its lineage from the confidence of Romantic hermeneutists Schleiermacher and Dilthey, a confidence that carefully applied hermeneutic method would yield the objective truth of a text (Hirsch, 1967). With regard to hermeneutic questioning, the conservative approach asks not whether a text has a unified meaning but rather what that unified meaning is. About conservative hermeneutics, Shaun Gallagher has observed,

These theorists would maintain that through correct methodology and hard work the interpreter should be able ... to break out of her own historical epoch in order to understand the author as the author intended ... to reach ... objective truth. (Gallagher, 1992: 9)

This is to say, the world-historical situation of the object under study can be grasped given due diligence, and such a grasp will shed helpful light on the true meaning of that object. On this version of hermeneutics, language is

primarily a symbolic representation of an author's meaning. That is to say, language reveals, but does not constitute, the meaning of an author, as well as the author's historical/cultural situation. Reality is one thing; language is another. A methodically rigorous study of language yields objective knowledge of reality. And with regard to disciplinary knowledge, traditional hermeneutics, while it does not restrict the scope of hermeneutics to one or another discipline, uses disciplinary boundaries and specific disciplinary knowledge to enhance the fit between interpretive method and the object under study.

Critical Hermeneutics

Carl-Otto Apel and Jürgen Habermas typify the critical branch of modern hermeneutics. Also described as depth hermeneutics, critical hermeneutics hails from the Freudian/Marxist tradition of the Frankfurt School (Habermas, 1985). Following criticalist principles, there are always two main roadblocks standing in the way of a more accurate interpretation. The first is psychic interference and repression. The second is ideological obfuscation stemming from unjust economic practices. Both of these cloud one's understanding of the object under study to the point where it becomes falsely interpreted. In order to attain an accurate interpretation, psychic interference and ideological obfuscation must be eliminated.

Critical hermeneutics remains fairly sure about which meanings are in need of questioning. The meaning of a text or utterance is questioned, but primarily in regard to psychic and ideological obfuscations. Once these roadblocks are questioned and exposed, the text itself becomes available in all its truth, not unlike the truth revealed by conservative hermeneutics. Regarding world-historical situation, the underlying Marxist orientation of critical hermeneutics once again informs the interpreter's work. While the cultural and historical circumstances of the object under study are significant, more significant are the ways that such circumstances tend to conceal, especially through the workings of capitalist ideology, the ability of the text or utterance to speak its truth. Following the critical hermeneutics of Habermas in particular, language supplies an ideal speech situation only when it has first been purged of ideological and psychic misunderstandings. Thus, Habermas, in his debates with Gadamer, argued that it is naïve to trust language without first purging its ideological underpinnings. A critical understanding of disciplinary knowledge insists, primarily, that no discipline is more objective than another. Even the physical sciences are riddled with ideology and thus require critical, depth hermeneutics.

Radical Hermeneutics

Inspired especially by the genealogical observations of Friedrich Nietzsche, radical hermeneutics is associated

with the work of Michel Foucault and Jacques Derrida, among others. Radical hermeneutics rejects any steadfast grounding for interpretive endeavors (Foucault, 1995; Derrida, 1974). Showing neither confidence that objective truth can be obtained through sound interpretive method nor that truth lies ready to be uncovered once ideology has been unmasked, radical hermeneutics looks to interpret interpretation itself. In so doing, radical hermeneutics rejects the canonizing tendency of most interpretive endeavors, challenging the very borders of the objects it studies. Whereas traditional, moderate, and critical hermeneutics offer various ways of interpreting an object, radical hermeneutics, as it carries out its interpretations, challenges the integrity of the object itself.

This challenge to the object itself can be seen in the ways that a work is put into question. For example, Derrida is famous for stating that "*il n'y a pas d'hors text*" (Derrida, 1988). Roughly translated as "there is nothing outside of the text," this statement typifies the way that Derrida questions his objects of interpretation. Instead of accepting the integrity of an utterance or a text, Derrida often shows that a text almost always unravels precisely what it seems to mean. This is to say, a text cannot keep itself under control. It cannot keep to itself and thus has no true outside or limit. As for the world-historical situation, radical hermeneutics posits that such situations must themselves be subject to interpretation. Foucault, in his early work, posits the existence of epistemic regimes that underlie the ways that it is possible to interpret a world-historical situation. In other words, a historical situation should not be understood at face value because such an understanding does not take into account the sorts of tacit knowledge-claims underlying such a situation in the first place. There are thus clear links between Heidegger's emphasis on tacit-ness and radical hermeneutics. Where language is concerned, radical hermeneutics once again puts interpretation itself under interpretation. Foucault, for example, denounces what he calls the "reign of the signifier," where words are too easily mistaken for the things and ideas they seem to represent. For radical hermeneutics, the very link between words and ideas cannot be taken for granted as power and linguistic slippages undermine such a link at every step. As one might expect, radical hermeneutics is not a friend to disciplinary knowledge. While Foucault posits that such knowledge is no more than a ruse of power, Derrida has almost single-handedly blurred disciplinary distinctions between philosophy, theory, literature, and law, just to name a few.

Moderate Hermeneutics and Educational Thought

Moderate hermeneutics can be understood as a continuation of Heidegger's ontological hermeneutics, but without

the mistrust of the object under study that characterizes radical hermeneutics. Moderate hermeneutics has confidence that texts and utterances have something fairly unified to say to the interpreter. Paul Ricoeur and Hans-Georg Gadamer are primary thinkers along these lines, and Gadamer has been most influential with regard to education (Ricoeur, 1974, 1981; Misgeld and Nicholson, 1992). A detailed, thematic exploration of Gadamer's work indicates its educational usefulness at the same time that it illustrates central tenets of moderate hermeneutics. Though it should be understood that moderate hermeneutics in no way exhausts the scope and function of hermeneutics in its other forms (conservative, critical, and radical), the educational implications of moderate hermeneutics will be used here to illustrate the significance of hermeneutics more generally. The four themes of questioning, world-historical situation, language, and disciplinary knowledge are once again used for illustrative purposes.

Questioning

Questioning is central to any attempt at interpretation; and more specifically, hermeneutics distinguishes between two types of questions, the false and the true. When a person questions with an answer already in mind, such a question is false. Such a question already expects in advance a certain interpretation of a text or an utterance, and thus is not open to the possibility of the unknown. In contrast to the false question, a true question has the profound purpose of finding out something that is not previously known. Hermeneutics reminds us that the true question has serious, even mortal, consequences. The true question reminds one that things might be different than originally thought. That is, things might not be, at least not as they are now. Another aspect of the question is that it can never be wide open. Every question contains conditions that limit the scope of how that question might be answered. For example, when a person asks, "Why do leaves turn red in the fall?" the possible responses to such a question are not endless. The answer to such a question cannot be "There are 60 s in a minute." Every question contains the seeds of its own answer.

The educational implications of hermeneutic questioning are significant. Typically, the questions posed by teachers are not true. Teachers often have answers in mind already, answers that are set by a predetermined curriculum. It is a significant and important educational challenge to try to avoid false questions, in favor of true questions, whenever possible. The self-limiting nature of questioning also poses a significant educational challenge. Heretofore, educational questioning has generally been considered benign rather than authoritative. Hermeneutics reminds educators that questioning enacts its own particular kind of authority. When a teacher questions, he or she is steering the conversation rather than letting

the conversation be steered by the student. Questions are never posed from a blank slate. They entail predilections that are educationally significant.

World-historical situation

As the importance of the cultural backgrounds of students becomes less and less ignored in educational institutions, hermeneutics offers a way to understand the interaction between cultural and personal standpoints. Typically, the educational institution has been construed as a public space wherein students shed their cultural differences in order to engage in the personal goal of becoming educated. This liberal paradigm has the limitation of regarding the personal as an escape from the cultural. Hermeneutics, on the other hand, indicates that the part (the personal) and the whole (the cultural) cannot be separated from each other in such a fashion. When a student learns something new in the classroom, he or she does not shed his or her world-historical horizon. Instead, that larger horizon is the backdrop against which personal learning must make sense. Just as the reader can only interpret one passage of a text in light of the whole text, so too the learner can only interpret school experience in light of a larger cultural horizon.

Hermeneutics also shows that the world-historical situation of the student always orients him or her in tacit ways. These tacit understandings of the world are largely ignored in educational thinking since such thinking often considers the pre-educated mind to be a blank slate or an empty container. From the hermeneutic point of view, understanding is never at a lack, though it is for the most part tacit. Educationally, then, it is important to provoke these tacit understandings. As Gadamer has said, "All understanding is self-understanding." This means that understanding something new always entails provoking something old, something old that one's world-historical situation has tacitly conditioned. In educational terms this means provoking the tacit knowledge of students, knowledge that is informed by students' world-historical situation. Provoking tacit knowledge is a requirement for learning something new.

Language

Perhaps the most important hermeneutic contribution to education is its focus on language. In education, language is most often construed in terms of the conduit metaphor, language being the conduit that transfers thought from one mind to another. Language is taken as a medium by which human beings convey the thoughts that they have in their minds. This understanding of language has influenced teaching methods as well as curriculum construction for centuries. Following this theory of language, teachers are encouraged to make their presentations more clear and interesting because the more effective the language, the more effectively thoughts are conveyed from mind to mind. Also, it is assumed that the organization of

curriculum – through concise logical order and in concise logical language – will ensure the most efficient learning on the part of the student. Both spoken and written languages are taken as conduit for transmitting thoughts of teachers to the minds of students.

The hermeneutic project emphasizes that language cannot be taken in such a way, that educators would do well to rethink what they do along linguistic lines. Gadamer, for example, regards language as essentially conversational. Language, he maintains, is typified by the to-and-fro of partners in a conversation. Words do not start from one person and end up at another. Rather, words are partaken in. They are experienced, rather than used, by interlocutors. Because language is essentially conversational, the abstraction of facts and symbols into logically ordered progressions, indeed, the coveted progression of curricula, is of no value until it is understood and interpreted. Education informed by hermeneutic linguistics thus focuses more on the to-and-fro of dialog and less on knowledge transmission.

The hermeneutic understanding of language challenges one of the central tenets of progressive education, namely, that the authority of the speaking teacher can stultify the freedom of the student. Progressive education, at least since Dewey, has tried to move away from the student-as-listener, encouraging instead the student-as-participant. But Gadamer distinguishes between a person who speaks with authoritarianism and a person who speaks with authoritativeness. The former is harmful, while the latter enacts the laudable authority of wisdom and knowledge. The former uses his or her institutional power to dominate while the other partakes in a learned conversation. Because language is partaken in rather than transmitted, the listener is as much of an active participant as the speaker.

As language is not about understanding another person's meaning, but is about partaking in a conversation, the hermeneutic orientation also offers a very different understanding of second-language education than usually surfaces in educational discussions. Most often, the non-native speaker in a classroom is given special attention because he or she is not likely to understand the meaning of the teacher's words. Such special attention is based on the conduit model of language. If the message is not received correctly, then it is the language that is at fault. In terms of human understanding, hermeneutics reminds us that this sort of remediation puts the cart before the horse. Conversation, even if it is across languages, is always full of meaning and understanding. Education should start with a presumption of such fullness rather than a presumption of lack.

Disciplinary knowledge

Hermeneutics offers insight into disciplinary knowledge that is often overlooked in other approaches to education. Educational practices are guided by disciplinary specialization for two main reasons: because such specialization

provides an efficient way to administer curriculum, and because knowledge has historically been organized along disciplinary lines. However, along with the advantages of disciplinary specialization for curriculum delivery, it has long been acknowledged that learning is facilitated when students make connections across disciplines. Thus, educators perennially negotiate between the efficiencies of disciplinary specialization and the meaningful connections to be made between disparate bodies of thought.

Certainly, this connection between disparate but efficient disciplinary specialization, on the one hand, and making general connections among disciplines, on the other, resonates with the hermeneutic interplay between the part and the whole. But hermeneutics has even more to contribute than the part/whole metaphor. From a hermeneutic perspective, the part and the whole are not simply two entities to be negotiated. Instead, they are elements that are constitutive of all understanding in the first place. So, while educators usually describe this negotiation between the specific and the general in ways that encourage a healthy trade-off between specialization and generalization, hermeneutics reminds educators to take a much harder line. Whatever specialization or generalization characterizes curriculum administration, understanding itself only happens within the to-and-fro movement between the specific and the general.

Summary

Hermeneutics, in general terms, is the art of interpretation. As such, hermeneutics has a rich history and can now be identified with four major strands: conservative, critical, radical, and moderate. Out of these strands, the moderate hermeneutics of Hans-Georg Gadamer has proven to be the most relevant to educational thought. While many hermeneutic themes speak to educational concerns, four themes – questioning, world-historical situation, language, and disciplinary knowledge – are especially relevant. Hermeneutics remains an important, if not yet thoroughly explored, branch of educational philosophy.

Bibliography

- Derrida, J. (1974). *Of Grammatology*, Spivak, G. C. (trans.). Baltimore, MD: John Hopkins University Press.
- Derrida, J. (1988). *Limited Inc.* Mehlman, J. (trans.). Evanston: Northwestern University Press.
- Dilthey, W. (2002). *The Formation of the Historical World in the Human Sciences*. Makkreel, R. A. and Rodi, F. (eds.). Princeton, NJ: Princeton University Press.
- Foucault, M. (1995). *Discipline and Punish: The Birth of the Prison*. New York: Vintage.
- Gadamer, H. -G. (1994). *Truth and Method*. Weinsheimer, J. and Marshall, D. G. (trans.). New York: Continuum.
- Gadamer, H. -G. (1976). *Philosophical Hermeneutics*. Linge, D. (trans.). Berkeley, CA: University of California Press.

- Gallagher, S. (1992). *Hermeneutics and Education*. New York: State University of New York Press.
- Habermas, J. (1985). *The Theory of Communicative Action, Vol. 1: Reason and the Rationalization of Society*. Boston: Beacon Press.
- Heidegger, M. (1962). *Being and Time*. Macquarrie, J. and Robinson, E. (trans.). San Francisco, CA: Harper.
- Heidegger, M. (1999). Letter on humanism. In Krell, D. F. (ed.) *Basic Writings*, 251pp. New York: Routledge.
- Hirsch, E. D. (1967). *Validity in Interpretation*. New Haven: Yale University Press.
- Misgeld, D. and Nicholson, G. (eds.) (1992). *Hans-Georg Gadamer on Education, Poetry, and History: Applied Hermeneutics*. Albany: State University of New York Press.
- Rickman, H. P. (1963). Dilthey and the philosophy of education. *International Review of Education* 8(3/4), 336–343.
- Ricoeur, P. (1974). *The Conflicts of Interpretation: Essays in Hermeneutics*, Domingo, W., et al. (trans.). Evanston: Northwestern University Press.
- Ricoeur, P. (1981). *Hermeneutics and the Human Sciences: Essays on Language, Action and Interpretation*. Thompson, J. B. (trans.). Cambridge: Cambridge University Press.
- Schleiermacher, F. (1998). *Hermeneutics and Criticism*. Bowie, A. (ed. and trans.). Cambridge: Cambridge University Press.

Further Reading

- Bingham, C. (2001). I am the missing pages of the text I teach: Gadamer and Derrida on teacher authority. *Philosophy of Education Yearbook* 2001, 265–272.
- Caputo, J. (1987). *Radical Hermeneutics: Repetition, Deconstruction, and the Hermeneutic Project*. Bloomington, MN: Indiana University Press.
- Grondin, J. (1994). *Introduction to Philosophical Hermeneutic*. New Haven, CT: Yale University Press.
- Grondin, J. (1995). *Sources of Hermeneutics*. New York: State University of New York Press.
- Kerdegan, D. (1998). Hermeneutics and education: understanding, control, and agency. *Educational Theory* 48(2), 241–266.
- O'Neil, L. (2004). Method's web: Gadamer's corrective and educational policy. *Philosophy of Education Yearbook* 2004, 142–149.
- Palmer, R. and Michelfelder, D. (eds.) (1989). *Dialogue and Deconstruction*. New York: State University of New York Press.
- Risser, J. (1997). *Hermeneutics and the Voice of the Other: Re-Reading Gadamer's Philosophical Hermeneutics*. New York: State University of New York Press.
- Wachterhauser, B. R. (ed.) (1986). *Hermeneutics and Modern Philosophy*. New York: State University of New York Press.

Islamic Education

Y Waghid, Stellenbosch University, Matieland, South Africa

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Adl – Universal justice which implies the achievement of justice for every person irrespective of religion, culture, identity and socio-political affiliation.

Ummah – Communal engagement, generally in reference to the unity and connectedness of Muslims all over the world.

shūrā – Public deliberation which creates space for people to engage with each other on the grounds of respect for diversity of opinions.

Jihād – Just striving, including the recognition of the rights of others which is more fundamentally associated with intellectual exertion rather than the exercise of physical and emotional violence.

Ta'līm – Socialization of Muslims learners into an inherited body of Islamic concepts and practices.

Tarbiyyah – Individuation of learners into discourse of reflection and critique.

Ta'līm – Virtuous action mostly related to the achievement of social activism in communities.

Islamic Philosophy of Education and Universal Justice

The Qurān is replete with verses (*āyāt*) which emphasize the significance of achieving justice for all human beings wherever they might be. The most famous of these verses is the one which is recited every Friday when the *imām* (congregational leader) renders the compulsory *khutbah* (sermon): “Surely Allah enjoins the doing of justice and the doing of good (to others) and the giving to the kindred. And He forbids indecency and evil and rebellion; He admonishes you that you may be mindful” (The Bee, 16: 90). Again, in another verse: “O you who believe! Be upright for Allah, bearers of witness with justice, and let not hatred of people incite you to act inequitably; act equitably, that is nearer to piety, and be careful of (your duty to) Allah; surely Allah is Aware of what you do” (The Dinner Table, 5: 8). Likewise, it is stated in the Qurān: “O you who believe! Be maintainers of justice, bearers of witness for Allah’s sake, though it may be against your own selves or (your) parents or near relatives; if he [she] be rich or poor, Allah is nearer to them both in compassion;

therefore do not follow your low desires, lest you deviate; and if you swerve or turn aside, then surely Allah is aware of what you do” (The Women, 4: 135). Now considering that the Qurān is one of the primary sources of a philosophy of Islamic education – another being the Sunnah or life experiences of the Prophet Muhammad (May Allah’s peace and blessings be upon him), and the Qurān clearly emphasizes the importance of achieving justice for all, it would be plausible to claim that the rationale of Islamic education is the achievement of ‘*adl*’ (justice) in relations among people. What the aforementioned verses also foreground is an understanding that justice is not the domain or proprietorship of individuals, but that justice is done to others and in relation with others. These others, of course, might not necessarily be of the same religious, cultural, linguistic, ethnic, political, cultural, or social milieu of the individual or group enacting justice. Simply stated, justice is not reserved for a particular group, but rather for all people wherever they might live. In this sense, justice is universal or global.

The question arises: What does it mean to be treated justly? First, universal justice is conceptually linked to being nonoffensive, nonsubversive, and decent toward others. If so, then one should treat others with dignity and respect without inflicting physical and emotional harm on others. In an educative perspective, this makes sense because learning and teaching cannot take place without people (learners and educators) being made to feel that they deserve one another’s respect as dignified persons. Second, being just requires of one not to act unequally (perhaps through bigotry and resentment). In other words, people should be treated equally and not as some strangers in an unfair manner: specifically, think of the challenges that educators in universities face not to treat students from immigrant communities, for instance, as if they do not deserve our equal and symmetrical attention. Third, justice is linked to recognizing the rights of others (and not just being consumed with one’s own rights) and that their (others’) rights should be assured: think of the rights that students from all communities have to higher education and how university educators should go about ensuring that this right to education is secured. This brings us to a discussion of how treating others with decency (civility), equality, and the recognition and assurance of others’ rights can possibly be achieved through the acts of *ummah* (communal engagement), *shūrā* (public deliberation), and *jihād* (just striving, which includes the recognition of the rights of others) – all considered by the Qurān as acts of justice.

Cultivating Acts of Justice

First, the Qurān (The Chambers, 49: 13) states the following: “O humankind! Surely we have created you of a male and a female and made you tribes and families [communities] that you may know each other; surely the most honorable of you with Allah is the one most careful of his [her] duty; surely Allah is Knowing, Aware.” The point about this verse is that people ought to be recognized as not only having commonalities and differences, but they should also actually share these on the grounds of acting responsibly (with duty, that is to say, with justice). Therefore, the idea of community is aimed at getting people to share with others what they have in common and what they disagree about; that is, it is a matter of getting to know one another. In another Qurānic verse it is stated: “And people are naught but a [global] community, so they disagree; and had not a word already gone forth from your Lord, the matter would have certainly been decided between them in respect of that which they disagree” (Jonah, 10: 19). Therefore, although people are different, the sense of community which ought to exist among them is one whereby they share commonalities and disagreements. However, the fact that humanity is considered as a global community (*ummah*) means that people need to learn to live with the otherness of others, whose ways of being may be deeply threatening to our own (Benhabib, 2002: 130). Moreover, through the idea of *ummah* (community), people endeavor to find a civil space whereby they can enact what they have in common and at the same time make public their competing narratives and significations. In this way, people might develop a real opportunity to coexist. Again, in this way, they would not only establish a community of conversation and interdependence (i.e., they share commonalities), but also one of disagreement (i.e., they do not share commonalities) without holding in disrespect others’ life-worlds (Benhabib, 2002: 35 and 41). In other words, when people are engaged in a community underpinned by interdependence and disagreement, they engage in an educative process with a collective identity – they share commonalities. Besides, educating people to become members of a global community involves creating civil spaces where they can learn to share commonalities and respect the differences of others. Only then, people would be acting justly. In this regard the Qurān states: “You are the best of peoples raised up for the benefit of humanity; you enjoin what is right and forbid the wrong and believe in Allah; and if the followers of the Book had believed, it would have been better for them; of them (some) are believers and most are transgressors” (The Family of Imrān, 3: 110). In essence, when the Qurān advises people to work as a global community (*ummah*), it in fact does so on the grounds that people ought to engage communally with one another’s differences and commonalities. If they do so, their actions would be just,

because they would be for the advancement of humanity. Following such an understanding of *ummah* (communal engagement), one can claim that Islamic education aims to establish opportunities whereby people engage with one another and, in so doing, they share commonalities and differences in the interest of advancing human relations. In this way, they would act justly because participants of a global community remain respectful toward one another despite their differences and disagreements.

In other words, *ummah* (communal engagement) demands that individuals actively engage with the unending struggle and responsibility for the improvement of the economic, social, and political aspects of life (Alibasic, 1999: 234). In this sense, *ummah* is concerned with a long-term and inconclusive commitment to the improvement of human conditions. Such a community is concerned to maintain the freedom and duty of criticism and monitoring of government, to accept criticism in good spirit, to facilitate peaceful change, and to remain united through consensus and disagreement (Alibasic, 1999: 237, 240, 242, and 292) – a clear indication of such a community’s obligation to be critical, to develop self-critical attitudes, and to live peacefully. In addition, communal engagement or *ummah* is also concerned with a plurality of human ideas and not denying the rights of others (Alibasic, 1999: 249 and 271) – thus indicating its recognition of difference.

Second, the Qurān not only encourages people to act justly as a global community, but also suggests a way of *shūrā* (public deliberation) as to how people ought to engage. What does the Qurān say about *shūrā* (public deliberation)? In fact, an entire chapter (*sūrah*) of the Qurān is devoted to a discussion of *shūrā* (public deliberation). The core verse which relates specifically to *shūrā* (public deliberation) is as follows: “And those who respond to their Lord and keep up prayer, and their rule is to take counsel [*shūrā*] among themselves and spend out what we have given them” (Counsel, 42: 38). The value attached to *shūrā* (public deliberation) is so profound that Allah connects the practice to prayer (*salāh*) and almsgiving (*zakāh*). Throughout the chapter, Allah speaks about the importance of engaging others justly (Counsel, 42: 15), that is, with patience, forgiveness, and courage (Counsel, 42: 43). These constitute virtues of public deliberation which would hopefully encourage and persuade people to act justly. In the first instance, public deliberation cannot happen without the patience of listening to the viewpoints of others, albeit that they may be in conflict with one’s own. The point is that public deliberation cannot happen unless we listen attentively to others’ justifications and, in turn, give to others an account of our own justifications. Only then we can justifiably talk about deliberation.

Moreover, in the second instance, the Qurān also states: “Call [engage others] to the way of your Lord with wisdom and goodly exhortation, and have disputations with them in the best manner; surely your Lord best knows those

who go astray from his path, and He knows best those who follow the right way” (The Bee, 16: 125). Whereas public deliberation ought to involve different and contending parties listening to one another’s views, it also needs to invoke disputations. This means that people should also have the courage to take one another’s views into some kind of systematic controversy. In other words, we should not be concerned merely with listening to what others have to say and then agree with them, but also to treat one another’s truth claims critically without, of course, exceeding the limits (The Elevated Places, 7: 55). In this sense, exceeding the limits refers to insulting and demeaning people. However, it does not mean that one cannot fervently disagree with another person’s view. In this sense, disputations do not only have to be feeble; rather, arguments can be articulated ardently without alienating others, more specifically without excluding them from public deliberation. For this reason, the author is somewhat hesitant to equate courage in public deliberation with belligerence, as proposed by Eamon Callan. For Callan, people in deliberation disturb doubts about the correctness of their moral beliefs or about the importance of the differences between what they and others believe (a matter of arousing distress) accompanied by a rough process of struggle and ethical confrontation – that is, belligerence (Callan, 1997: 211). If this happens, belligerence and distress give way eventually to moments of ethical conciliation, when the truth and error in rival positions have been made clear and a fitting synthesis of factional viewpoints is achieved (Callan, 1997: 212). The problem with Callan’s view is that he assumes that all people are necessarily confrontational, which might lead to excluding others from the deliberation who might not be confrontational. One can also arrive at defensible justifications without engaging belligerently with someone else.

However, in the third instance, what is more crucial for the public deliberation to be ongoing is the virtue of forgiveness. When people deliberate, they do not have to argue for a specific point of view after others’ views have proven to be more plausible than theirs. The virtue of forgiveness can mean that a previously held view can be dismissed if implausible, and that the proponent of such a view should not be considered as intellectually slow, but rather be freed from being associated with an indefensible viewpoint. This also implies that people should not ridicule others for a previously held indefensible view. In The Chambers (49: 11) the Qurān states: “O you who believe! let not (one) people laugh at (another) people perchance they may be better than they, nor let women (laugh) at (other) women, perchance they may be better than they; and do not find fault with your own people nor call one another by nicknames; evil is a bad name after faith, and whoever does not turn, these it is that are the unjust.”

Third, this brings us to a discussion of *jibād* (just striving, including the recognition of the rights of the

others) considered as one of the most often misrepresented concepts and which can be considered as a constitutive feature of a philosophy of Islamic education. Why? The Qurān equates *jibād* with seeking closeness to Allah as stated in The Dinner Table (5: 35) as follows: “O you who believe! Be careful of (your duty to) Allah and seek means of nearness to Him and strive hard in His way that you may be successful.” Moreover, the Qurān links *jibād* to working collectively with others in the path of virtuosity: “And We have revealed to you the Book with the truth, verifying what is before it of the Book and a guardian over it, therefore judge between them by what Allah has revealed, and do not follow their low desires (to turn away) from the truth that has come to you; for every one of you did We appoint a law and a way, and if Allah had pleased He would have made you (all) a single people, but that He might try you in what He gave you, therefore strive with one another to hasten to virtuous deeds; to Allah is your return, of all (of you), so He will let you know that in which you differed” (The Dinner Table, 5: 48). If people are encouraged to strive collaboratively to attain justice, then they have to recognize one another’s rights and actually do something about ensuring that their rights are honored. Recognizing and honoring one another’s rights is important in ensuring that people are treated equally, decently, and respectfully. Therefore, the Qurān proclaims that people should be educated about their rights, whether civil, political, or social: “O people! be careful of (your duty to) your Lord, Who created you from a single being and created its mate of the same (kind) and spread from these two, many men and women; and be careful of (your duty to) Allah, by Whom you demand one of another (your rights), and (to) the ties of relationship; surely Allah ever watches over you” (Women, 4: 1). In essence, just striving (*jibād*) is aimed at drawing nearer to a higher good, developing one’s capacities to be morally upright, and respecting the rights of others.

Thus far, we have discussed how the idea of universal justice, the rationale of a philosophy of Islamic education, can be achieved through actions such as *ummah* (communal engagement), *shūrā* (public deliberation), and *jibād* (just striving which includes the recognition of the rights of others). We now move on to how these acts of justice are linked to various conceptions of Islamic education, namely, *ta’līm* (socialization), *tarbiyyah* (individuation), and *ta’dīb* (good action).

Conceptualizations of Islamic Education

Thus far, we have discussed how universal justice constitutes a philosophy of Islamic education. In turn, we have explored three different ways in which universal justice can be achieved, namely, through the acts of *ummah* (communal engagement), *shūrā* (public deliberation), and *jibād* (just striving, including the recognition of the rights of

others). We now examine how these acts of justice guide particular conceptions of Islamic education, namely, *ta'lim* (socialization), *tarbiyyah* (individuation), and *ta'dib* (good action), with reference to their implications for teaching and learning.

First, Islamic education is couched as *ta'lim* (socialization) as illustrated by the following verses in the Qurān: “And He taught (*allama*) Adam all the names, then presented them to the angels; then He said: Tell me the names of those if you are right” (The Cow, 2: 31); “They said: Glory be to Thee! We have no knowledge but that which Thou hast taught us; surely Thou art the Knowing, the Wise” (The Cow, 2: 32); and “Taught man [women] what he [she] knew not” (The Clot, 96: 5). These verses foreground a particular conception of education (*ta'lim*) whereby people learn through being socialized into an inherited body of knowledge. That is, learning takes place when people are taught what they perhaps do not know. Certainly for Muslims, this means being taught how to adhere to their faith, especially those principles associated with being a good person. For instance, being socialized means being taught what it means to believe in Allah, His Angels, His Revealed Books, His Prophets, The Last Day of Judgment, and the separation between good and evil. Likewise, being taught about Islam involves what it means to serve Allah, perform prayer, execute fasting, provide alms to the poor and destitute, and set out on a pilgrimage once in a lifetime, if Muslims have the means to do so. Socialization is also associated with learning the Qurān (including its memorization or passages from it), the Hadīth (sayings related to the life experiences of Prophet Muhammad), the Sīra (the Prophet’s life history and those of his companions), the Islamic sciences such as Shari’ah (law), Fiqh (jurisprudence), and Tawhīd (science of interpretation). Now the problem with *ta'lim* (socialization), at least so it seems, is that this form of education has often been associated with uncritical exegeses of what medieval scholars have said about Islamic knowledge. For instance, some Muslims in the Islamic world often confine their engagement (i.e., their sense of *ummah*) with the primary sources of Islam to the exegeses of past medieval scholars, which in many cases results in a stultification of knowledge and understanding. It is for this reason that claims are often made that Islamic education merely advances doctrinaire learning. This claim is supported by Bagheri and Khosravi (2006: 100), who argue that Islamic education has been used throughout the Muslim world to indoctrinate learners. In the author’s view, this limited view of Islamic education is not commensurate with the notion of *ummah* (communal engagement), whereby Muslims also need to be taught a form of education which invokes criticality. Hence, we now introduce a discussion of *tarbiyyah* (individuation).

Second, whereas *ta'lim* (socialization) aims to introduce people to an inherited body of knowledge (without being

uncritical toward such knowledge), *tarbiyyah* (individuation) specifically invites Muslims to be critical of their learning. In the first instance, the word *rabb* (literally lord, which the author has adapted to educator) occurs approximately more than 1000 times in the Qurān in relation to the provision of mercy, guidance, evidence, and clear proofs. Of concern to the author is the use of *rabb* in relation to proofs (*bayyināt*). In The Cattle (6: 57) it is said: “Say: Surely I have manifest proof from my Lord and you call it a lie; I have not with me that which you would hasten; the judgment is only Allah’s; He relates the truth and He is the best of deciders”; then, again in The Cattle (6: 104): “Indeed there have come to you clear proofs from your Lord; whoever will therefore see, it is for his own soul and whoever will be blind, it shall be against himself and I am not a keeper over you.” Likewise, in The Cattle (6: 157) “Or lest you should say: If the Book had been revealed to us, we would certainly have been better guided than they, so indeed there has come to you clear proof from your Lord, and guidance and mercy. Who then is more unjust than he who rejects Allah’s communications and turns away from them? We will reward those who turn away from Our communications with an evil chastisement because they turned away.” In a specific verse in The Cow (2: 111), Allah (as The Educator) invites people to evaluate His Guidance and that they respond critically with proof of their justifications: “And they say: None shall enter the garden (or paradise) except he who is a Jew or a Christian. These are their vain desires. Say: Bring your proof if you are truthful.” The main point about this verse is that people are invited to come up with their own proofs or justifications, which suggests that a situation or argument can be taken into controversy. Simply stated, the notion of *tarbiyyah* creates scope for critical evaluations and interpretations based on sound reasons for disagreement. The upshot of this view of Islamic education is that people can question and undermine a particular point of view, which suggests that they ought to reflect about the knowledge they receive and construct – a matter of becoming critical. Therefore, *tarbiyyah* (individuation) can be considered as another phase in Islamic learning, that is, once people have acquired knowledge and they are informed they can then begin to challenge and question prevailing understandings. No wonder the Qurān invites people to contemplate and deliberate about educational matters. Somewhere else, the author has specifically discussed how actions such as *tafukkur* (contemplation), *tadabbur* (critical reflection), *fabm* (rational understanding), and *aql* (intellectual inquiry) underscore the practice of deliberation (Waghid 1996b).

Third, with reference to my previous work (Waghid, 1996a) on the matter and in particular the seminal thoughts of Muhammad Naquib al-Attas, another form of Islamic education is under included in the term *ta'dib* (good action) (al-Attas, 1991: 23). For al-Attas, Islamic

education is guided by *adab* or the appropriate use of knowledge (*ilm*), reason (*nutq*), intellect (*aql*), and heart (*qalb*) – more specifically one's physical, intellectual, and spiritual capacities – to perform acts (*amal*) of justice (*adl*). This view of *ta'dīb* (good action) is in line with the rationale of Islamic education discussed earlier, namely that of producing a just person: "The *just* man [women] is he [she] who effects such *adab* unto his [her] self, resulting in his [her] being a *good* man [woman]" (al-Attas, 1991: 24). We concur with this approach to Islamic education on the grounds that one cannot just acquire knowledge of the Islamic sciences and then begin to critically analyze and respond to particular issues on the assumption that one's actions will then result in something worthwhile or appropriate for the global community. The author believes that *ta'dīb* (good action) has in mind actions which can lead to the improvement of the global community's situation. Therefore, one requires *jibād* (just striving which includes the recognition of the rights of others), which can invariably change distorted or improve unsatisfactory situations. For instance, we are specifically thinking of how *ta'dīb* (good action) can contribute potentially toward the eradication of racial bigotry, gender oppression, cultural imperialism, and even terrorism, because good action requires that people pursue actions which can eliminate inhumane acts perpetrated against humanity such as murder, extermination, enslavement, deportation, persecution on political, racial, or religious grounds, war crimes (mistreatment of civilians and noncombatants as well as one's enemy in combat), and genocide (through ethnic cleansing, mass executions, rape, and cruel punishment of the enemy). Simply put, *ta'dīb* (good action) has an emancipatory interest in mind, which can be made possible through a just striving which takes into account the rights of others and that others' rights ought to be assured. Here, we specifically use *jibād* as referring to just actions of the mind rather than the application brute force, which invariably leads to more violence.

In essence, we have shown that the three conceptualizations of Islamic education, namely, *ta'līm* (socialization),

tarbiyyah (individuation), and *ta'dīb* (good action), should be seen as complementary actions of the mind in the pursuit of achieving justice for every person wherever he or she might be. In turn, we have also shown how practices such as *ummah* (communal engagement), *shūrā* (public deliberation), and *jibād* (just striving, including the recognition of the rights of others) can contribute toward achieving universal justice, considered as the rationale of Islamic education.

Bibliography

- Al-Attas, M. N. (1991). *The Concept of Education in Islam*. Kuala Lumpur: International Institute of Islamic Thought and Civilisation.
- Alibasic, A. (1999). The right of political opposition in Islamic history and legal theory: An exploration of an ambivalent heritage. *Al-Shajarah* 4(2), 231–295.
- Bagheri, K. and Khosravi, Z. (2006). The Islamic concept of education reconsidered. *American Journal of Islamic Social Sciences* 23(4), 88–103.
- Benhabib, S. (2002). *The Claims of Culture: Equality and Diversity in the Global Era*. Princeton, NJ: Princeton University Press.
- Callan, E. (1997). *Creating Citizens: Political Education and Liberal Democracy*. Oxford: Oxford University Press.
- Shakir, M. H. (1983). *The Holy Qur'an* ((trans.) Online Book Initiative). Gutenberg: Tahrike Tarsile Qur'an.
- Waghid, Y. (1996a). *Ta'dīb*: Restatement of Islamic education. *Muslim Education Quarterly* 13(4), 32–45.
- Waghid, Y. (1996b). In search of a boundless ocean and new skies: Human creativity is a matter of *a'mal*, *jihad* and *ijtihad*. *American Journal of Islamic Social Sciences* 13(3), 353–365.

Further Reading

- Al-Attas, M. N. (ed.) (1979). *Aims and Objectives of Islamic Education*. Seven Oaks: Hodder and Stoughton.
- Ashraf, A. F. (ed.) (1985). *New Horizons in Muslim Education*. Cambridge: The Islamic Academy.
- Talbani, A. (1996). Pedagogy, power and discourse: Transformation of Islamic education. *Comparative Education Review* 40(1), 66–82.
- Tibawi, A. L. (1972). *Islamic Education: Its Traditions and Modernization into the Arab National Systems*. New York: Crane, Russak.

Liberalism and Education

W Feinberg, University of Illinois at Urbana-Champaign, Champaign, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Prehistory of Liberalism

While the seeds of moral liberalism can be found in both the Hebrew Bible and the New Testament, and especially in the appeal to individual conscience found in the reformation, political liberalism arose in response to a specific historical moment as an instrument to meliorate the religious wars that began in the middle of the sixteenth century. Thomas Hobbes's (1588–1679) attempt to re-legitimate political authority after the doctrine of the Divine Right of Kings was discredited is often seen as the foundational moment of liberalism. However, because he tried to justify the absolute power of the sovereign, Hobbes, himself, is not usually thought of as a liberal.

Hobbes's (1958) ideal was to substitute as the justification of absolute sovereignty of the Crown the rigorous principles of logic, for the will of God. He began by imagining what life was like in a state of nature before governments were formed. He conjectured that such a state would be “a war of all against all” in which life was “nasty, brutish and short” and he speculated that people, recognizing the hopelessness of this situation, eventually came together and gave all of their power over to a single sovereign who then maintained order and govern them all. This sovereign then had absolute power, except in cases where he threatened an individual's life. When this occurred then the relation of the condemned to the sovereign would be as it was in the original state of nature. Hobbes, while not a liberal, made an enormous contribution to the formation of liberalism by affirming government was a human, not a divine, product. It is dependent not on the will of God but, at least in its inception, on the will of the governed. However, for Hobbes the authority of the people, once granted, no longer existed and the authority of the government was virtually absolute. Later liberals would argue that governments continue to be accountable to the will of the people.

Types of Liberalism

The basic tenet of political liberalism is that government resides in and is accountable to the governed. As times have changed, however, this basic idea has taken a number of different and sometimes conflicting forms.

Conventional classification divides liberalism into two forms – classical and contemporary. Classical liberalism emphasizes negative political rights and the need to limit

government authority. It dominated political thought from the days of John Locke (1632–1704) until the Depression of the 1930s when philosophers and economists began to re-think the basic tenets of liberalism. Contemporary liberalism emphasizes positive rights and a more active role for government. The difference between the two can be captured by thinking of classical liberalism as a political doctrine that fears tyranny more than it loves justice, whereas contemporary liberalism may be thought of as a doctrine that loves justice more than it fears tyranny. The first seeks to keep government small; the second to assure that it is helpful. For both, however, the central principles are the same. (1) The individual is real, and society and government is a creation of the governed. (2) Government draws its legitimacy from the governed and is accountable to and can be replaced by it. (3) All individuals have certain rights independent of the government, (4) If the government violates these rights, it loses its legitimacy.

Many liberal theorists also borrow from Hobbes a mode of argumentation that roots the basic principles of liberalism in a contract that arises among people within a state of nature. As we will see, however, the use of the notion of a contract has changed over time from something that was accepted as a literal representation of the history of governments to something that carries but metaphorical and heuristic weight.

In line with political liberalism are certain other ideas; the most important one for education's purposes is that the individual is both the agent and the object of moral decision making. As the agent, the individual must accept the decision as her own if it is to be a moral one. Decisions that are manipulated or forced are not considered moral decisions. This is the root to the liberal commitment to tolerance – beliefs cannot be forced – and to the respect of individuals – persons must be respected and they must be free to choose their own way of life. That is, they must be autonomous. As the object of moral decisions the goodness or badness of any action will be judged not by how well it conforms to some external ideal or to the will of God, but by how well it serves the individual's own idea of the kind of life she should lead. Given these similarities, there is still a great deal of difference between the liberalism that fears tyranny and the liberalism that loves justice.

The real beginning of classical political liberalism is usually associated with the ideas of the British theorist, John Locke. Locke had a considerably more benign view

of the state of nature than did Hobbes. For the most part people lived peacefully with one another, committed to defend their own rights and to respect those of others. Private property arose when men (sic) mixed their labor with nature. To reach up to pick an apple is an example of the mixing of labor with nature and of the beginning of private property. If I plow up unclaimed land in order to plant an orchard, the land and the orchard become mine. Governments arise, not because there is a war of all against all, as Hobbes thought, but as a result of specific conflicts. Say, I did the reaching but you then did the planting. Governments arise when arbiters are selected to resolve the disputes. In this Locke develops the idea that the function of government was limited to protecting individual rights, such as the right to property, and to settling disputes about these rights, such as whose property does it really belong to. In contrast to Hobbes, the role of the government for Locke was not absolute. The governed could change the government, especially if it fails to protect these rights. Nevertheless, as Sabine observed: "The two men fastened on social theory the presumption that individual self-interest is clear and compelling, while a public or a social interest is thin and unsubstantial" (Sabine, 1958: 529).

Given Locke's view that government has but a very limited role – to protect the rights of individuals, rights granted by nature – classical liberals are distrustful of government authority, and often view government itself as a necessary evil. Yet there is also in Locke, as there is in liberalism in general, an implicit appeal to majority rule, and liberal thought will later divide as to whether government must be as vigilant about the possibility of the tyranny of the majority, as it must be over the tyranny of a monarch. As Robert Nozick, in his modern defense of a minimum state, argues, warning against the former, any taxation beyond that required to protect one's natural rights is theft, however good the purpose for which it might be used (Nozick, 1974), or however large the majority favoring it.

Historically, its later association with economic liberalism mitigated the tension within classical political liberalism. Economic liberals such as Adam Smith (1723–90) argued that small government is not only a protection against tyranny, but is also good for everyone. Hence in this view there is no conflict between the state that serves the majority and the state that assures freedom from tyranny. Smith (1904) argued that free commerce would allow producers to compete fairly with one another for the favor of the consumer and out of this competition would come the highest quality at the lowest possible price. The role of government then was to simply assure equal access to the market by guaranteeing that everyone had an opportunity to compete on equal grounds. Any significant government involvement beyond that point would reduce the over all productivity and wealth of the

nation. Hence, a small government was not only a moral imperative as a protection against tyranny, it was also a practical imperative that would increase the wealth and satisfaction of the nation as a whole. Consumers would get the quality they wanted at the price they were willing to pay.

Of course, it is not quite obvious that as the total wealth of a nation increases, the satisfaction of everyone increases along with it. A rising tide may not lift all boats, and the metaphor has its limitations when it comes to economic well-being. A counter example is those instances where as the wealth of a nation grows so too does the uneven concentration of that wealth leads to great inequality and an increase in poverty.

Classical economic liberalism seems unconcerned with responding to the question as to how economic progress is to be judged when there is a discrepancy between the aggregate wealth of a nation and the poverty of the different individuals who compose it. Nevertheless, classical economic liberalism added to classical political liberalism's ideas that a good government must not only protect natural rights, but also, even if only by implication, that a good society must provide its citizens with equal opportunity to compete economically. While these ideas would eventually serve as an important foundation for modern education – the one in terms of a universal right to education and the other as a significant forerunner to the idea of equal educational opportunity – at the time of their formation, politics and economics, not education, were the focus of attention. The failure to understand that an increase in the total wealth of a nation could be attained at the cost of the well-being of many of its citizens served as an important stimulus to the challenges to classical liberalism in both its economic and its political forms and to the development of contemporary liberalism.

The Rise of Contemporary Liberalism

Under classical liberalism, the absence of tyranny was equated with justice, as seen largely in procedural terms. Justice was done when one had a fair hearing or when one was allowed to compete on equal terms with others and without government favoring one producer or consumer over another. All of this would have perhaps been fine if there were not another consideration classical liberalism took seriously – the right that we all had to do what we want with our wealth, including the right to advantage our own children over others – either through inheritance, or through education, or both. This meant that for each new generation the playing field was tilted in favor of the children of the rich and powerful, who were then able to make new laws, which again favored their children. Where these favors resulted in a desperately poor working class, as they did in England and the United States in the

nineteenth and much of the twentieth centuries and where they ignored the disenfranchisement of black people and women, then the fear of destitution and the situation of powerlessness replaced the fear of state tyranny as the primary concern, and real procedural justice, that the classical idea of equal opportunity entails, was seen wanting. The very idea of a minimal government was seen as complicit with the economic tyranny that a largely uncontrolled market was allowed to produce.

Contemporary liberalism arose in response to excesses associated with classical liberalism during the nineteenth century, but it gained currency as the impact of inadequate government intervention became obvious during the Great Depression of the 1930s. While this new liberalism also feared tyranny, it did not believe that tyranny must necessarily be associated with big government. For the average worker, subject to the arbitrary will of an all powerful industrialist, economic tyranny came in the form of the masters of the free market, and the minimal state they seemed to own. In other words, unregulated big business was seen as an even greater danger than big government.

Contemporary Liberalism

Much of the ground work for contemporary liberalism was prepared by Utilitarian thinkers such as John Stuart Mill (1806–1873) who argued that the goodness of society can be measured in terms of the happiness of its members, and who extended the role of government to the performing of those essential tasks that private business would be unable or reluctant to perform. Mill (1962) was generally an advocate of free markets and limited government, as well as of free speech and tolerance. Moreover, like Adam Smith he shifted the intellectual justification for the minimal state away from a deductive argument issuing from certain premises about natural rights to an empirical argument based on the utilitarian principle of maximizing pleasure.

This change in argumentative tactics is important for should this empirical claim prove false, then the foundation of the minimal state would be wanting. As a political reformer arguing for women's rights and against slavery, Mill was aware of the discrepancy between the claims for the minimal state and the actual conditions of working people. While he was reluctant to reject the basic principles of economic liberalism, he did believe that education was the key to realizing the greatest amount of happiness. Having departed from Jeremy Bentham's (1749–1832) view that all pleasures were of equal worth, Mill's view of education would seem to require a large infusion of state involvement to insure that children of workers be exposed to pleasures that their parents did not know or have the opportunity to value. However, it remained for a new breed of liberalism to draw out the implications of this idea.

Contemporary liberalism opens up the possibility that the state and especially state-sponsored education can be instruments for unblocking potential and for advancing human flourishing. Hence, while it builds upon the concern to block tyranny, and thus carries forward a commitment to procedural justice, it also holds that the state has a positive role in enabling human potential as well as protecting rights. It thus promotes the state as both an agent of welfare and an agent of education.

Contemporary liberalism begins with what White (1957) has called the revolt against formalism or what Dewey (1929/1962) has labeled the end to the quest for certainty. It begins with a rejection of the idea that foundational political principles can be deduced by an exercise in formal logic alone, as both Hobbes and Locke seemed to believe, and substitutes probabilistic conclusions about likely consequences for conclusions about logical necessity. This more pragmatic approach provided the grounding for reform movements in many fields. In legal theory, reformers like Holmes (1920), recognizing that changing conditions required new laws, sought to balance the consideration of precedent with a consideration of the consequences of a law. The importance of this shift is difficult to exaggerate since it opened the way for lawyers and judges to seek the opinions of social scientists in a wide range of areas. The shift was also reflected as Keynes (1936/1997) and his arguments for government intervention to stimulate economic growth competed with Adam Smith and his argument for minimal government for favor in university economics departments. The shift also had a profound impact on education.

Education and Contemporary Liberalism

Because of the radical change from a rural society to an urban one as well as the changes in work that the industrial revolution brought about, education became a more significant feature of liberal theory. During the early formulation of liberal ideas when most children learned to work by sharing in the chores on the farm or by apprenticing to a master workman, formal education was but an incidental part of children's lives. For many, but not for all, rudimentary skills were all that was required for the work they were destined to perform, and higher education was important for only a small, largely cultural elite.

While classical economic liberalism advanced the idea that everyone should have an equal opportunity, the ideal was largely restricted to market exchanges. While a few, such as Jefferson, thought the ideal would be a good one to apply to education, in actual fact even high schools and certainly colleges and universities were, with a few exceptions, largely reserved for the children of the elite. The exceptions were more often bright children from marginal groups, who, it was thought, could be assimilated and could then serve as cultural exemplars for their kind.

While this began to change in the United States with the founding of land grant colleges, higher education was not the ground on which originally opportunity was supposed to be apportioned equally. In practice, of course, neither was the economic market, nor many supplementary theories about race, class, and gender, had to be added to make the vast inequalities in economic opportunity fit into the tidy theory that classical economic liberalism provided. Indeed, much of the intellectual work of the first half of the twentieth century, from racial based IQ theory to cultural deficit theory, can be seen as the latter day working through of this supplement to classical economics.

Contemporary liberals recognized that if social institutions and practice were to take advantage of and contribute to the developing technological change, education would need to play a larger role in the equation. As compulsory schooling grew so did the needs of the children in school, and a great many reforms were proposed to address these needs. These ranged from attempts on the part of efficiency experts to match the curriculum to the needs of industry, to the ideas of progressive educators like John Dewey to advance the growth of the whole child, and later to the life adjustment movement which popularized curriculum innovations such as drivers education and sex education.

In the thought of the efficiency expert equality of opportunity meant identifying (largely through IQ type tests) the capacities of individual students to fit into certain slots in the workplace and then to shape a varied curriculum in order to efficiently train students to occupy those slots. For Dewey, it meant providing students with the knowledge and inquiry skills so that they could contribute to their community and decide for themselves how to lead happy, useful, and reflective lives. Progressives like Dewey were also concerned with the psychological qualities that students need to live in liberal, democratic societies, and their pedagogical theory was guided by the implicit idea that the nature of a subject is defined in part by how it is taught, an idea that Dewey had developed more abstractly in his treatment of the concepts of ends and means. To teach a subject as a series of right and wrong answers is to teach that knowledge is fixed. To teach a subject as an area for inquiry is to teach that knowledge is open and subject to revision.

Partly as a result of (1) the war against Hitler's brand of racism, (2) the continuing struggles of women and minorities, and (3) the changing needs of the new economy, racial, gender, and class discrimination in both work and school became more visible and as it did, the principle of equal opportunity migrated from employment and markets to education and schools. In response some elite universities changed their admissions policies favoring private prep school graduates and the sons of the elite, and incorporated tests of academic aptitude, tests which

were billed as objective and impartial. The idea was to cast a wider net for talented students and to provide more opportunities for worthy public school children.

Yet the fact that over time these tests seemed to advance white students with very few spaces for students of color (often women were excluded or relegated to separate schools) led some to re-think both the fairness of the tests and the concept of equal opportunity. Partly in response to this discrepancy, a policy of affirmative action, which began as a policy about hiring, was extended to universities where race and later gender became explicit considerations in selecting faculty and students and in attempting to increase the proportion of students and faculty from different racial and gender background. Affirmative action is still controversial among liberals. Some call it reverse discrimination – implying that “two wrongs don't make a right”. Others express concern that social class is largely left out of the formula and that neither race nor gender is an adequate proxy for social class. Yet the policy also has its defenders. Some believe that diversity itself adds to the quality of education while others view it as either partial payment for a historical debt or an essential tool in breaking down debilitating stereotypes.

Milton Freedman and John Rawls

There is still a robust competition between classical and contemporary liberalism. Classical liberalism has found a persuasive voice in the late Milton Friedman (1912–2006) who argues (2002) that market solutions and small government are the solution to all economic problems. Ultimately, for Friedman all problems are economic problems. This includes education. Here Friedman advanced the idea that government should support parental choice of schools through providing all parents, rich and poor, with vouchers. Friedman seems complacent about any growth in inequality that likely would occur as a result of his policy. As his ideas have spread around the world they have been given the label neoliberalism, a term that involves both a positive and a negative connotation. On the positive side, it suggests economic growth and integration into the world economies for developing countries. On the negative side, it suggests growing inequality and an intensification of the impact of poverty and dislocation.

Contemporary liberalism is represented by the work of the philosopher John Rawls, (1921–2002). Rawls (1971) used a new version of contract theory to refine the idea of liberal justice. He proposed that we imagine that in the state of nature people must choose the principles that will govern their society but that they must do so behind a veil of ignorance where they are deprived of any knowledge of their own individual characteristics and of how they compare to others. He believes that in the absence of knowledge about our own race or religion, or about our physical

strength or intelligence and without knowledge of the wealth or historical epoch of our society, we would all choose a society that will maximize the well-being of the least advantaged person. Hence given a society in which basic individual rights and equal opportunity are secured, societies should strive to maximize the well-being of the least-advantaged person. While Rawls does not provide a lot of detail about how this is to be accomplished, a redistribution of wealth through taxation is certainly not ruled out, and would likely be required to meet the standard of social justice that he advances. While Rawls says little about education, he most likely would support policies that provide the greatest advantage for the least-advantaged child and many contemporary liberals hold that this would support redistributive educational policies.

Critics

While both classical and contemporary liberals are critical of each other, they also are vulnerable to external criticism. Communitarians like MacIntyre (1981) believe that liberalism overemphasizes individual rights and neglects communal connectivity while many Marxist and some feminists critics hold that liberalism fails to address structural inequalities and does not understand the ways in which racist, sexist, and classist institutions both reinforce and are reinforced by racist, sexist, and classist attitudes.

One liberal response to the first criticism is that liberals do not underestimate the significance of communal relations. They are only concerned that adults have the opportunity to develop their own conception of the good. While the cultural group to which the individual belongs may influence this conception, liberals insist that individuals must be free to formulate their own idea of the good life. For educators this means that children must be provided the skills and perspectives needed to lead autonomous lives as adults. Some contemporary liberals argue that liberal states may have an indirect obligation to advance certain cultural formations because of the importance cultures have in the development of individuals. Some, like Kymlicka (1991), believe that this is especially the case with regards to cultural groups that have been involuntarily incorporated into a larger polity with a different cultural orientation. Those educators who believe that it is important for schools to help children maintain their cultural connections through bilingual and bicultural education can find support in this version of liberalism.

Perhaps the most prominent criticism of liberalism today comes from postmodern theory and it comes not in the form of a sustained critique but with an ironic distance that rejects the optimism that it associates with liberalism. What is rejected is not the idea that individuals have certain rights but rather naive optimism sometimes

associated with liberalism that with modernity comes progress. Postmodern theorist remind us of the many voices that have been left out of the political discourse as the norms of liberalism were being shaped. To a large extent this would include women, people of color, as well as gay, lesbian, and bisexual people. This is less a rejection of liberalism as such as it is a reminder of the work that needs to be continued and the people who need to be included in the conversation.

Conclusion

Liberalism has provided a rich and somewhat flexible educational ideal. On the one hand, education is viewed as a means to inform people about and to help secure their basic individual rights while providing them with the skills and attitudes required to earn a living in the modern world. On the other hand, many critics hold that in actual fact schooling has become largely an instrument to reproduce an oppressive unjust social order. The latter view has considerable credibility, especially in countries like the United States and Great Britain where intergenerational mobility, especially among certain minority groups, has been inadequate by most standards of human flourishing. There is also concern that education has lost its public forming mission and has become largely a vocational enterprise. To the extent that liberal theory is used to cover up these concerns, say by insisting that individuals are free to choose their own fate, it is rightly seen as culpable. However, this seems to be a misuse of liberal ideas and certainly there are likely enough intellectual resources within liberalism itself to help us shape a new and more just educational project.

Bibliography

- Dewey, J. (1929). *The Quest for Certainty*. New York: Minton, Balch.
- Friedman, M. (2002). *Capitalism and Freedom*. Chicago, IL: University of Chicago Press.
- Hobbes, T. (1958). *Leviathan*. New York: The Liberal Arts Press.
- Holmes, O. W. (1920). *The Path of the Law, Collected Legal Papers*. New York: Harcourt, Brace.
- Keynes, J. M. (1936/1997). *The General Theory of Employment, Interest and Money*. New York: Prometheus Books.
- Kymlicka, W. (1991). *Liberalism, Community and Culture*. Oxford: Clarendon Press.
- Locke, J. (1690/1956). *The Second Treatise of Government*. New York: The Liberal Arts Press.
- MacIntyre, A. (1981). *After Virtue: A Study in Moral Theory*. Notre Dame, IN: University of Notre Dame Press.
- Mill, J. S. (1962). *On Liberty, in John Stuart Mill, Utilitarianism, On Liberty Essays on Bentham*. New York: New American Library.
- Nozick, R. (1974). *Anarchy, State and Utopia*. New York: Basic Books.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge: Harvard University Press.
- Sabine, G. S. (1958). *A History of Political Theory*. New York: Holt, Rinehart and Winston.

- Smith, A. (1904). *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: Methuen.
- White, M. (1957). *Social Thought in America: The Revolt against Formalism*. Boston, MA: Beacon Press.

Further Reading

- Brighouse, H. (2000). *School Choice and Social Justice*. Oxford: Oxford University Press.
- Bourdieu, P. and Passeron, J. C. (1977). *Reproduction in Education, Society and Culture*. London: Sage.
- Callan, E. (1997). *Creating Citizens*. Oxford: Oxford University Press.
- Feinberg, W. (1998). *Common Schools: Uncommon Identities/ National Unity and Cultural Difference*. New Haven, CT: Yale University Press.
- Gutmann, A. (1987). *Democratic Education*. Princeton, NJ: Princeton University Press.
- Johnston, D. (1994). *The Idea of A Liberal Theory: A Critique and Reconstruction*. Princeton, NJ: Princeton University Press.
- McDonough, K. and Feinberg, W. (2003). *Citizenship and Education in Liberal-Democratic Societies: Teaching for Cosmopolitan Values and Collective Identities*. Oxford: Oxford University Press.
- Strike, K. (1989). *Liberal Justice and the Marxist Critique of Education: A Study of Conflicting Research Programs*. New York: Routledge.
- Taylor, C. (1992). *Multiculturalism and the Politics of Recognition*. Princeton, NJ: Princeton University Press.
- Waldron, J. (1993). *Liberal Rights: Collected Papers, 1981–1991*. Cambridge: Cambridge University Press.

Phenomenology and Existentialism

B Levering, Utrecht University Utrecht, The Netherlands; Fontys Professional University, Tilburg, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Civil disobedience – Revolutionary attitude as opposed to the government to be found in existentialism.

Embodiment – The phenomenological answer to the mind–body problem. The human body is as well object as subject. (Merleau-Ponty).

Essence – The objective of philosophy before phenomenology.

Existence – The human way of being (Heidegger).

Freedom (absolute) – The core of Sartre's early philosophy. Human freedom pushed to the extreme.

Life-world – The human world where all meanings stem from (Husserl).

Natural attitude – Common but wrong conviction in philosophy and science that there is something like a reality behind reality (Husserl).

Pedagogy – Continental view on (philosophy of) education, not restricted to what happens in the school, but focused on upbringing in general.

Responsibility – The important consequence of Sartre's view on human freedom. In the view of Levinas, pedagogical responsibility is not intentional but being held hostage by the helplessness of the child.

Transcendental idealism – The world as – in the view of Husserl – constituted by human consciousness.

Visibility – The property of the world that the invisible can in principle be made visible (Merleau-Ponty).

Visuality – The capacity of human beings to see more of the world than what is caught by the eye (Merleau-Ponty).

“There is no reality behind the reality.” In both philosophy and science, we have to do with reality as we encounter it through our ordinary experience. That is the position of the father of phenomenology, Edmund Husserl (1859–1938). From an epistemological perspective, phenomenology has to be seen as a radical response to the viewpoint of Kant. Precisely what Husserl had in mind is, curiously, not so easy to grasp. The idea that there is a genuine reality concealed behind the reality of phenomena, the so-called natural attitude, is far more familiar to us in one form or another. It is a view that has been

drummed into us since Plato. Husserl's creed *Zu den Sachen Selbst* should not, therefore, be understood as a form of objectivism; more than anything it means “back to the original experience.” However, the structure of this ordinary, everyday experience is anything but easy to reconstruct.

In *Die Krise der Europäischen Wissenschaften und die Transcendentale Phänomenologie*, Husserl (1954) demonstrates very clearly that the idea of a reality behind the reality arises from doubts about the veracity of the reality as we encounter it through our ordinary experience. But on what grounds could we doubt the truth of everyday experience? The idea of the true reality is a fabrication, the result of substruction, as Husserl called it. All meanings originate in the life-world, the concrete human world. With this conclusion and taking this as his starting point, Husserl really set off in a different direction.

At first, Husserl had not completely given up the ambition to get through to objective reality. At that stage, he still regarded phenomenological reduction as putting the question of the relationship between phenomena and genuine reality on hold for the time being. In his later transcendental idealism, Husserl assumed that the world is constituted in the consciousness. A more radical rejection of the natural attitude is not possible. Be that as it may, phenomenological analysis is not concerned with the relationship between phenomena and a true reality behind them, but with the relationship between phenomena themselves. One of the most important lessons that phenomenology can teach us, therefore, is that we have to stop searching everywhere for something behind things. That is a theme that we will encounter again in various different forms.

The triumphal procession of the natural sciences that followed the scientific revolution of the late Middle Ages was based on a concentration on epistemological questions. For the first time in history, epistemological questions were being asked separately from ontological questions. In the classical era, it was customary first to determine the nature of reality in order to then look for adequate ways to acquire knowledge. Now, it became the custom to abandon speculations about the nature of the cosmos and to formulate rules which had to be drawn up based on scientific methods, come what may. Such rules, the primacy of formal logic and the method of doubt, led with René Descartes, for instance, to a strict division between subject and object.

Descartes' method generated interesting problems. The epistemological subject–object division was associated with a strict mind–body division. The epistemological question

"How is knowledge possible?" manifested itself at that level as the question "How can a nonspatial mind have knowledge of a spatial body?" A consequence of Descartes' methodical rules was in fact that mind and body do not resemble each other in any respect. Everything that the mind was, the body was not: immaterial/thinking versus material/nonthinking. How and where would they interact with each other? As we will see later, phenomenology, for the first time in history, came up with a solution for the mind-body problem that had been formulated as long ago as Aristotle. Phenomenologists prefer not to speak in terms of subject and object, but in terms of I and the world. I and the world are not separate but are related to each other, since it is simply impossible to separate I and the world. There is no world without I, because the world we want to talk about exists by the grace of the meaning humans give to it.

The I-world relationship is about unity in diversity. I and the world do not coincide. The world is not in my head. Consciousness is always consciousness of something, but I experience the things that I see or hear, not here but there. Intentionality is called the intelligibility moment of phenomenology. That is to say that the idea of giving meaning lies at the basis of the ability to know the world. This is a two-way matter. Besides giving meaning, there is also borrowing meaning. I give things meaning, but the world also forces its meanings upon me.

It is self-evident that language plays a very important role in the relationship between I and the world. Language allows me to name the things in the world, to indicate what the things in the world mean to me, but the world that I come across does not need me to name it from top to bottom. The things already had names before I was born. Intersubjectivity seems to precede subjectivity, as we will soon see.

All the same, language is the crowning glory as far as the meanings in and of the world are concerned, rather than the most fundamental connection between I and the world. It is human corporeality that creates the original bond between I and the world. Corporeality, particularly as developed by Merleau-Ponty (1945), has little to do with the body as in Descartes, from which the relationship with the mind remained entirely unclear. The body in Merleau-Ponty is not an object, it is object and subject. I can look at my hand and describe it, then it is an object. But I can also grasp objects with my hand, then it is a subject. Because the human body is both subject and object, corporeality solves the epistemological subject-object problem.

The body is indeed, to paraphrase Brand (1971), the pre-eminent concrete *a priori*. Kant was still treating space and time as abstract categories of consciousness. According to phenomenology, people proportion their world in a concrete sense starting from their corporeality. Before the day that a metal bar was laid down in Paris as the metal

standard meter, and much later the meter was defined as the distance traveled by light during an infinitesimal time interval, people had used their thumbs, elbows, and feet to measure things and 5 km was expressed as the walk for 5 h, the time you needed to move your body over that distance. Even when it comes to psychological traits, from time immemorial, these have been expressed in terms which made reference to the body. Before abstract scientific terms like extrovert and authoritarian came into everyday use, we talked, for instance, about people being headstrong and pigheaded when we wanted to describe the psychological traits of others.

In the philosophy of Maurice Merleau-Ponty (1908–61), the relationship between I and the world falls nicely into place by describing the phenomenology of observation in terms of corporeality. We generally only see the things and people around us partially and we never see the front and back sides at the same time. That does not make us nervous; it would be absurd to doubt that things have a reverse side. Our own corporeality completes the things. We really do not need to see it for ourselves, although it is still possible to check whether things have a reverse side. Merleau-Ponty (1964) speaks of *visuality* and *visibility*. We still see the things that are outside our field of view in a certain sense and the things are visible even though there is, in fact, no image on our retina. To put it in everyday terms: our corporeality allows us to look round the corner, as it were.

Visuality and visibility come to play not only in the concrete observation of people and things around us. Our observation of the state of mind of other people is also like being able to look round a corner. After all, we can never directly observe the inner self of another person, we do not know what the inner self is or even whether such a thing as an inner self exists. We observe that a person is not feeling at ease by comparing his or her facial expressions and the circumstances, because of the fact that we know that the person has looked happier on other occasions. Our knowledge of another person's inner self is based on interpretation at the phenomenal level.

Memory and thinking back to earlier times can also be conceived as looking round a corner. When I think about the past, I can still conjure up certain events vividly in my mind, so vividly in fact that they do not seem any less intense that they did at the time they were actually happening. But the past is not really present. Anyone watching me while I am musing on the past will most probably feel that I am absent minded, and, despite the vividness of my recollections, I know that they are memories. Just as I have knowledge of the physical distance to things that I observe from a greater distance, even looking at them can bring me very close to them, as if I were there and not here, so I am aware of the distance in time when I remember, but my thoughts can bring me so close to events in the past that it is as if I have actually spanned the years.

That brings phenomenology into the realm of anthropology. In addressing the epistemological question, we have already said a lot about the specifically human way of knowing, and by doing so about the human condition. The relationship with the world takes us back from an epistemological relationship to a being relationship and that entails a rehabilitation of ontology. If it had been left to Martin Heidegger (1889–1976), we would have started from the other side. Existence, according to Heidegger, precedes essence. We can only fathom the world when we approach it from the human awareness of being there. At the end of his principal work, *Sein und Zeit*, Heidegger connects being with the typically human awareness of time. With the new anthropology comes a new ontology. The new ontology is, however, very different from what the Greeks understood by that term.

Time did not exist for the Greeks, time was a nonbeing. However, human existence is more than a sequence of moments in time that are irrelevant in themselves, human existence is not a continuous present. The human being is a unity of the past, present, and future. That becomes unmistakably clear when we try to understand the meaning of fundamental concrete human experiences like fear, guilt, and regret. To exist is to endure through time. In the case of regret, it is an unpleasant feeling that I have now, that relates to an event in the past, and which forces me to make a resolution for the future. Regret is not a self-chosen recollection of the past or noncommittal contemplation of the future. Regret is a feeling that happens to you at a particular moment, that you cannot prevent, and which connects you unavoidably to your past and your present. Where Husserl was still speaking about consciousness, Heidegger spoke about existence. In Husserl's transcendental idealism, the human being constituted reality; in Heidegger's analysis, the relationship with the world is affective as much as it is cognitive.

From Epistemology to Method and Upbringing

Phenomenology has two general epistemological principles, which not only serve as the starting point for its method, but are also significant for the way the process of education and upbringing takes shape.

The two general epistemological principles of phenomenology are subjectivity and intersubjectivity. There is personal sense-making, which means that every individual human being has his or her own outlook on reality. That personal perspective is bound up with preconceptions and preferences that we have acquired in our own personal past. In addition to this subjectivity, there is intersubjectivity, the body of communal sense-making. These shared meanings are tied up in social rituals and

customs and the communal meanings that are embodied in a language and which are specific to their time and culture.

Subjectivity creates a methodical problem. The phenomenologist bases his analyses on his own experience but the issue is that he is trying to say something about the reality. Phenomenological reduction ought to mean that he will give up his personal prejudices. Those analyzing upbringing do not report on their own personal experiences, even though having been brought up themselves does offer a sure way into the phenomenon. Phenomenological reduction is more of an attitude of wanting to observe things in an unbiased way, rather than a written-up methodical procedure. On close analysis, phenomenological reduction is also actually an impossible operation. After all, if one were to really renounce the meaning one personally gives to reality, that would make reality disappear.

The process known as eidetic reduction dovetails with the epistemological principle of intersubjectivity. It attempts to circumscribe the meaning of the phenomenon to be described through systematic variation. Eidetic reduction can take place by means of concrete comparison or through imaginary manipulation. This allows one to clarify which characteristics must be counted as belonging to a particular phenomenon and which must not. Some say that eidetic reduction cannot actually produce any new knowledge, as it rests on the meanings we already have and is partially embodied in the language that we have acquired in a specific cultural circle. However, if we make the pre-reflexive reflexive, put into words something that has never been put into words before, then we have to ask ourselves whether we can really sustain the view that we knew it all along.

The way the upbringing process takes shape can indeed be described very well in terms of two general epistemological principles of phenomenology. Bringing up children and educating them is always about transferring the culture, introducing them to intersubjective meanings. This means that the child is treated from the outset as a person who is also making sense of the world subjectively, in that process.

Existentialism: How to Live?

Jean Paul Sartre (1905–80), one of the two existentialists whose work and life we want to focus on first here, developed his phenomenology in an intensive exposition of the work of Husserl and Heidegger. In discussing the work of the phenomenologists up to now, their lives have not been addressed. With existentialists such as Sartre and Camus – neither of whom incidentally wanted to be called existentialist – that would be inconceivable. Work and life are so profoundly interwoven that a wrong impression

could be created if one were to pay no attention at all to their lives. Furthermore, in the case of both Sartre and Camus, their literary work formed an excellent introduction to their philosophy. That is why it used to be recommended that people start with the novels to gain a good understanding of these philosophers. For the contemporary reader though, the extreme individualism in the novels has probably become just as abstract as the philosophy, so one may as well start with the epistemology and ontology.

Sartre's philosophical development is marked by his two principal works *L'Être et le Néant* (Sartre, 1943) and the unfinished *Critique de la raison dialectique* (Sartre, 1960). He started the first book that was given the subtitle *Essay d'ontologie phénoménologie* (*Essay on Phenomenological Ontology*) in the early 1930s, before he became acquainted with phenomenology. His second book reveals his turn to Marxism, in particular historical materialism. In *L'Être et le Néant* (*Being and Nothingness*) Sartre substantiates his idea of absolute freedom that is so characteristic of his early philosophy. That absolute freedom is extreme, not only because it does not allow human beings any relativism or any excuse at all for their actions on the grounds of restrictive social conditions. Sartre takes the idea of absolute freedom much further. The man being burned at the stake dies, according to Sartre, not because the fire consumes his body but because he chooses to die. The importance of Sartre's concept of absolute freedom is probably ultimately not the concept itself, but the ethical dimension of human existence that is its consequence: absolute individual responsibility. Sartre aptly expressed the human condition with the words: "man is condemned to freedom." The power of this statement is obvious in that when people are convicted in court, their freedom is generally what they are deprived of, but a human being cannot do anything other than take responsibility.

The title *Being and Nothingness* is readily open to misinterpretation. In that title Sartre is not thinking of being human, his ultimate concern, when he uses the term *Being*; on the contrary, he is thinking of the being of things. The being of things is inherent in itself, it is *en soi*. Nothingness, *le Néant*, is the logical opposite of that being a thing, absolutely nothing. *Nothingness* is consciousness, is *pour soi*: it is geared to being and, for Sartre, that is typically human. This means that human consciousness is purely negative. Sartre coined a new term for this negation: *néantiser*. That is indeed the essence of being human: the freedom, the fundamental indeterminateness. The fact is that people are bound by the consequences of their own decisions: what Sartre called *facticité*. The term existentialism is not found in *Being and Nothingness*. Sartre became especially associated with the term existentialism following his 1946 lecture entitled *L'Existentialisme est un humanisme*, in which he attempted to defend his views against attacks from Catholic and communist quarters. However,

the authenticity that he focused on in that lecture appeared already at the end of his main work.

Sartre does not, however, have his philosophical work to thank for his immense popularity. That was simply too difficult for the ordinary reader and the student. The general public in the 1950s could, however, acquaint themselves with his ideas through his plays and novels, in which the meaning of absolute freedom is insinuated into the consciousness of the reader. In *Huis clos* (*No Exit*) (Sartre, 1938), a play that was later filmed under his direction, three people are condemned to be together for eternity: an army deserter who had cheated on his wife, a woman who had committed infanticide and a lesbian. They conceal their original identities from each other but have all chosen to be what they are. They cannot avoid accepting that what they are is the result of their own choices. In his novel *La nausée* (*Nausea*) (1938) and stories such as *La chambre* (*The Room*) and *Le mur* (*The Wall*), (in *Le mur*, Sartre, 1939), Sartre clarifies philosophical themes such as consciousness, absolute freedom, and responsibility for the ordinary reader.

Sartre is also an outstanding example of the intellectual who for 40 years made no secret of his anti-bourgeois opinions. Throughout all those years he expressed his views on collaboration, the Cold War, colonialism and the exploitation of the Third World, the Vietnam war, Russian prison camps in Siberia, female emancipation, and so on. When he tries to understand people in their social and historical context, the key concepts of his early philosophy, absolute freedom and responsibility, turn out to have shortcomings. He was reluctant to admit to wrong choices that he made and the truth behind political ideologies generally only dawned upon him slowly. When studying phenomenology in Berlin in the early 1930s, he was blind to the emerging fascism. In 1950 he was still reluctant to acknowledge the reality of the Soviet gulags. However, President de Gaulle expressed a judgment of him in no uncertain terms when he said "You don't arrest Voltaire" when Sartre the activist was about to be arrested.

The second existentialist to be discussed here is less a philosopher than the first. Albert Camus (1913–60) only wrote two books that can be broadly termed philosophical books, *Le mythe de Sisyphe*, 1942, and *L'homme révolté* (*The Rebel*), 1951. Both, Camus in 1957 and Sartre in 1964, were awarded the Nobel prize for literature. Sartre turned it down. In the late 1930s, Camus wrote with admiration, but also with detachment, about *La nausée* and *Le mur*. They met in 1944, the beginning of what Sartre called a difficult friendship. Camus never considered Sartre to be his friend. Whatever the relationship was, it ended in 1952 when the periodical *Les Temps modernes*, which Sartre edited, published a scathing criticism of *L'homme révolté* penned by Francis Jeanson.

Camus himself admits that three mythological themes operate in his literary work, those of Sisyphus, Prometheus,

and Nemesis. Sisyphus is for Camus the symbol of the human condition. By the human condition it is not so much about the conditions themselves, but about how people deal with those conditions. In the classical theme of Sisyphean labor – it is the lot of the first king of Ephyra to be punished in the underworld for his wiliness and greed by being made to push a block of marble up a steep hill over and over again for all eternity – the human condition gets its special meaning of absurdity. Absurdity does not exist for the cricket or the ant, only for humanity. However, acknowledging that life is absurd is only the beginning. Camus formulated a number of rules for human behavior. Whatever you do you should not lie and take refuge from the absurdity in religion; that is philosophical suicide. Life is worth the effort because there is a logic of the heart that is stronger than that of the intellect. No lies, no violence, no idealism either, because that costs human lives, but nor can you look on as a disinterested observer. There is a need for *bonnêteté*, which means decency rather than honesty, and tenderness.

It is thanks to French existentialists like Camus and Sartre that civil disobedience came to be part of the thinking about citizenship in the late 1960s. That was something on which they agreed. Where they disagreed was on the limits of protest and especially on the admissibility of violence. Camus set the different positions on violence against each other in his play *Les justes*. Camus was no unconditional pacifist, but he said he would rather be right by not killing anyone than be right standing on the rim of a mass grave. He also said that one can only kill on condition that one is also killed. Because Camus ultimately based resistance to terror on respect for life, he was reproached for having a Red Cross morality. Sartre, who had the last word in the *Les temps modernes* debate, accused Camus of lack of philosophical depth. When it came to the admissibility of violence, Sartre was indeed far more radical as the years passed, but he had his limits. He scathingly described Andreas Baader, a member of the German terrorist Baader-Meinhof group, after visiting him at Baader's request in 1974 in prison, as incredibly stupid.

Freedom and Responsibility in Upbringing and Education

In the continental European pedagogic tradition, in which phenomenology and existentialism play an important role, the term education (German: *Erziehung*, Dutch: *opvoeding*) is emphatically not confined, as it tends to be in the English-speaking world, to what takes place in school. The term pedagogy has been used in Western Europe (and South Africa) from time immemorial to refer to the branch of learning that is concerned with what happens to children on their long journey to adulthood, so this also includes what happens in the family. Not only the relationship

between parents and children, but also the relationship between teacher and pupil are primarily seen as pedagogic relationships. The modern formulation of the purpose of upbringing and education stems from the concept of autonomy, as developed in the work of the philosopher Immanuel Kant. The idea of autonomy, that we encounter again in a radicalised form in the absolute freedom concept of early Sartre, had become untenable by the end of the 1970s, as criticism from postmodernist circles made clear. Absolute freedom is not only restricted by outside forces – social conditions – it is also restricted internally due to internal flaws and inadequacies. For human beings, autonomy is unachievable in the real world. However, autonomy remains essential as a positive fiction for the process of bringing up children. Child-rearing and education continue to aim for autonomy, because human society in general, and especially human society in a democracy, assumes individual autonomy.

On the question of the responsibility that parents and teachers take in the pedagogic relationship, Sartre's interpretation of the concept also falls short. In general one can indeed assume, as Sartre does, that responsibility presumes freedom. In the case of the pedagogic relationship, however, responsibility is not based on freedom but on the lack of freedom, more specifically: the child's lack of freedom. The parent does not have the option whether or not to accept the parenting role. In the terms of the French phenomenologist, Emanuel Levinas (1906–95), the parent is held hostage by the helplessness of the child. Parents cannot do anything other than take this responsibility upon themselves.

Bibliography

- Brand, G. (1971). *Die Lebenswelt. Eine Philosophie des konkreten Apriori*. Berlin: de Gruyter.
- Camus, A. (1942a). *L'Étranger*. (Ward, M. (trans.) (1989). The stranger. New York: Random House.) Paris: Gallimard.
- Camus, A. (1942b). *Le mythe de Sisyphe*. (O'Brien, J. (trans.) (1955). The myth of Sisyphus. London: Penguin Books.) Paris: Gallimard.
- Camus, A. (1951). *L'Homme révolté*. (Bower, A. (trans.) (1991). The rebel: An essay on man on revolt. New York: Vintage.) Paris: Gallimard.
- Husserl, E. (1954). *Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie: Eine Einleitung in die phänomenologische Philosophie: Eine Einleitung in die phänomenologische Philosophie*. (Carr, D. (trans.) (1970). The crisis of European sciences and transcendental phenomenology: An introduction to phenomenological philosophy. Evanston: Northwestern University Press.) (Biemel, W. hrsg) Husserliana VI. Den Haag: Martinus Nijhoff. (urspr. 1936).
- Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. (Smith, C. (trans.) (1962). Phenomenology of perception. New York: Humanities Press.) Paris: Gallimard.
- Merleau-Ponty, M. (1964). *Le visible et l'invisible*. (Lingis, A. (trans.) (1968). The visible and the invisible. Evanston: Northwestern University Press.) Paris: Gallimard.
- Sartre, J.-P. (1943). *L'Être et le Néant. Essai d'ontologie phénoménologique*. (Barnes E. H. (trans.) (1992). Being and nothingness. New York: Washington Square Press.) Paris: Gallimard.

- Sartre, J.-P. (1960). *Critique de la raison dialectique*. Tôme I. Paris: Gallimard (Sheridan-Smith, A. (trans.) (2002). *Critique of dialectical reason*. And in French in text, vol. I. (new edn.). New York: Verso Books.).
- Sartre, J.-P. (1938). *La nausea*. Paris: Gallimard (Alexander, L. (trans.) (1969). *Nausea*. New York: New Directions Publishing Corporation.).
- Sartre, J.-P. (1939). *Le mur*. Paris: Gallimard (Alexander, L. (1991). (trans.) *The wall and other stories*. New York: Directions Publishing Corporation.).

Further Reading

- De Boer, Th. (1966). *De ontwikkelingsgang in het denken van Husserl* ((1978). *The Development of Husserl's Thought*. The Hague: Nijhoff.) Assen: van Gorcum.
- Gorner, P. (2007). *Heidegger's Being and Time: An Introduction*. Cambridge: Cambridge University Press.
- Heidegger, M. (1927). *Sein und Zeit*. (Stambaugh, J. (transl.) (1996). *Being and time*. Albany: State University of New York Press.). Tübingen: Max Niemeyer Verlag.
- Kwant, R. C. (1963). *The Philosophy of Merleau-Ponty*. Pittsburgh, PA: Duquesne University Press.
- Lee, E. N. and Mandelbaum, M. (eds.) (1967). *Phenomenology and Existentialism*. Baltimore, MD: Johns Hopkins Press.
- Levering, B. and Van Manen, M. (2002). Phenomenological anthropology in The Netherlands. In Tymieniecka, A.-T. (ed.) *Phenomenology World-Wide*, pp 274–286. Dordrecht: Kluwer.
- Levinas, E. (1961). *Totalité et Infini. Essay sur l'extériorité*. Den Haag: Martinus Nijhoff (Lingis, A. (trans.) (1969). *Totality and infinity*. Pittsburgh, PA: Duquesne University Press.).

- Levy, B.-H. (2003). *Sartre: The Philosopher of the Twentieth Century*. Cambridge: Polity.
- Lottman, H. R. (1997). *Albert Camus: A Biography*. London: Weidenfeld and Nicolson.
- Peperzak, A. Th. (1997). *Beyond: The Philosophy of Emanuel Levinas*. Evanston: North Western University Press.
- Sartre, J. P. (1945). *Huis clos: suivi les mouches*. Paris: Gallimard (Gilbert, S. (trans.) (1955). *No exit, and three other plays*. New York: Vintage Books.).
- Spiegelberg, H. (1982). *The Phenomenological Movement. A Historical Introduction*, 3rd rev. The Hague: Nijhoff and enlarged edn.
- Zaner, R. M. (1964). *The Problem of Embodiment: Some Contributions to a Phenomenology of the Body*. The Hague: Nijhoff.

Relevant Websites

- <http://www.phenomenologycenter.org> – Center for Advanced Research in Phenomenology.
- http://www.csudh.edu/phenom_studies – Centre for Philosophy and Phenomenological Studies, California State University, Dominguez Hills.
- <http://britishphenomenology.com> – Journal of the British Society for Phenomenology, The British Society for Phenomenology.
- <http://www.spep.org> – Society for Phenomenology and Existential Philosophy.
- <http://britishphenomenology.com> – The British Society for Phenomenology.
- <http://www.ipjp.org> – The Indo-Pacific Journal of Phenomenology.
- <http://www.phenomenology.org> – The World Phenomenology Institute.

Pragmatism

B Thayer-Bacon, University of Tennessee, Knoxville, TN, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Community of inquirers – Peirce's term for those who are most educated in the subject of inquiry, their coming together to share their knowledge with each other in order to advance our understanding to the next generation of scholars.

Critical thinking – Synonymous with reflective thinking for Dewey.

Experience – James defines experience not just in terms of things or events, but also the relationships between things. It involves a doing and an undoing, a trying and undergoing, and recognizes no division between act and thing in its primary integrity, as described by Dewey.

Fallibilism – Peirce's term for the limitations and situatedness of philosophers who cannot make truth claims that are certain.

Habits – James's concept that links the body and mind.

Pluralism – James's term that emphasizes the valuing of diverse cultural perspectives and the possibility that there are multiple ways to describe our world and what we take to be truths.

Radical empiricism – James's term for his theory that the only things debatable among philosophers are those definable in terms of experience. The world we experience is more than we can describe.

Reflective thinking – Dewey's term for the inquiry process that begins with doubt, a felt need, which triggers the thinker to seek a solution by gathering information and making observations and testing out the ideas by application, similar to the scientific method.

Relativism – The view that truth is a function of human practice, and that reality is a function of human belief.

Social behaviorism – Mead's term for his theory, that argues we become individuals out of our social relations with others.

Vicious intellectualism – James's term for philosopher's efforts to describe our experiences in terms of concepts, which are helpful but also limiting. Concepts help us avoid context.

Warranted assertability – Dewey's term for knowledge that focuses on the agreement process that we use to try to establish epistemic claims, emphasizing that the testing of truth claims is a social practice.

Introduction

Euro-Western philosophy has a long history that stretches back to ancient Greece with the contributions of Plato's dialogues and Aristotle's university lectures. American (USA) philosophy owes much to continental philosophy, and its influence is strongly felt in American universities. We cannot lay claim to contributing significantly to philosophical discussion until more recent times, for America is a young country in comparison to other countries, although we do have indigenous views that have developed over a long period of time in the Americas (North, Central, and South), which we are just beginning to appreciate and consider as possible contributors to philosophical thinking. Even though American philosophy is still young, there is a philosophical approach that Americans have developed that continues to contribute to the larger philosophical world, and that is pragmatism. In fact, it can be argued that because of America's youth, pragmatism took root there. However, this youthful suggestion about pragmatism is not meant to suggest that it lacks philosophical depth.

This article describes key ideas that pragmatism contributes to current philosophical discussions, by sketching some of the important contributions made by the founders of pragmatism. Through this discussion it is shown how pragmatism has an influence on philosophers today with neo-pragmatism/postmodernism (Rorty and Foucault), critical race theory (West), and feminism (Seigfried), as well as an influence on philosophy of education (Biesta, Bredo, Garrison, and Thayer-Bacon).

Classic Pragmatism

Charles Sanders Pierce (1839–1914) was credited with coining the term pragmatism by his friend, William James (1842–1910), while they were in graduate school at Harvard University in the 1870s. Peirce later renamed his philosophical views as pragmaticism, in an effort to distinguish himself from James and others. John Dewey (1859–1952) called himself an instrumentalist, or experimentalist, not a pragmatist. Dewey's good friend and colleague, George Mead (1863–1931), called himself a social behaviorist; however, history has brought them together and describes them as the classic pragmatists.

Peirce, a logician, mathematician, and scientist by background, is now known as one of the most original

and versatile American philosophers, the father of not only pragmatism but also of semiotics. Much of his work was unpublished at his death, for he was not able to attain a permanent position at any university and retired while still in his 40s. It was his friend, William James, who not only brought Peirce's work to people's attention, and helped in getting them get published, but also had Peirce's writings housed at Harvard University, where James had a long, successful career. James (1890, 1950) went on to become a well-known professor of philosophy and psychology, and is known as a father of psychology in the United States.

What were Peirce's key ideas that launched pragmatism in the Metaphysics Club that the Harvard graduate students formed? Peirce argued that we cannot separate our ideas from our experiences for there is an inseparable connection between rational cognition and rational purpose, between thought and action, and between thinking and doing. For Peirce: "A belief is that upon which a man [*sic*] is prepared to act" ("The fixation of belief"). Put in the form of his famous maxim, he said, "consider what effects, which might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object" ("How to make our ideas clear"). He contrasted his view of modern philosophy with that of continental philosophy through his devastating criticism of Descartes' doubting method, arguing that we cannot universally doubt, as Descartes attempted to do, for, doubting presupposes something to doubt, therefore we must have prior beliefs. We also cannot find the answer to our doubts in our own individual consciousness, as Descartes attempted to do, for this will lead us to subjectivism, with no way out of our ideas to reality. Peirce showed that Descartes relied on God, in the end, to account for the self and the world. Peirce suggested instead that we determine how clear our concepts are by running them through a functional test, grounding them to experience.

Charles Darwin's *Origin of the Species* had a profound impact on the thinking of the classical pragmatists, and his influence can be found in all of their work. For Peirce, this evolutionary view of science is expressed in terms of wanting philosophy to act like science and be open to critique, not taking anything as given, but instead viewing philosophy as a field of study that is in process, relying on criteria to make judgments that are corrigible. Darwin's evolutionary influence can also be found in Peirce's theory of Truth. Peirce's theory of fallibilism describes philosophers as limited, situated knowers in need of each other to help us find ever-elusive truth(s). Peirce defeated the idea that philosophers could claim to know universal Truth, individually, on their own, with certainty. He made the case for the need of a community of inquirers, those most educated in the subject of inquiry, to work together

and share their knowledge in the hopes of getting clearer in their ideas, and then passing on their knowledge to the next generation of scholars to work on further. For Peirce, Truth is something we are moving closer and closer toward; it is objectively real and in the end of time we will all agree on what it is. However, it is an elusive, regulatory ideal, always in the future, for there is always more we can learn that will help us understand better the world in which we live.

Most philosophers today embrace Peirce's idea of fallibilism, and the need for us to work together to solve our problems, instead of thinking that any of us have a God's eye view of Truth and can discover it on our own. Most philosophers have given up the search for certainty, a clear sign of Peirce's influence. However, concerns have been raised about Peirce's community of inquirers in terms of fears of elitism (who gets to count as a scholar?), his progressivism (that Truth is something we are moving closer to), that at the end of time we will all agree on what Truth is (a consensual, universal view of Truth), and his commitment to realism (that there is a world out there separate from us and what we think about it). Contemporary feminist philosophers (Seigfried, 1996; Thayer-Bacon, 2000, 2003), critical race theorists (West, 1989), and postmodernists (Foucault, 1980; Rorty, 1982, 1989) critique Peirce on these concerns. However, we find disagreements concerning these ideas even with the classic pragmatists.

Peirce's idea of Truth as an emerging absolute, that demands of us endless investigating, changes in the hands of his friend, William James, which is one of the reasons why Peirce tried later to disassociate his pragmatism with James'. James follows Peirce's radical lead of incorporating contingency and revision into a theory of truth, yet he unties his theory of truth from Peirce's assumption that Truths are objectively real. James postulates with Peirce an end of Truth but knows that this is just a belief. "Meanwhile, we have to live to-day by what truth we can get to-day, and be ready tomorrow to call it falsehood" (*Pragmatism*). For James a belief is true if it yields sensibly satisfactory results in experience when acted upon. As James defines truth in *Pragmatism*: "True ideas are those we can assimilate, validate, corroborate, and verify. False ideas are those that we cannot." True ideas are ones we have resolved and no longer doubt. "Truth *happens* to an idea. It *becomes* true, is *made* true by events. Its verity *is* in fact an event, a process; the process namely of its verifying itself, its *veri-fication*." James took the notion of absolute Truth to be dogmatic, a foreshadowing of feminist and postmodern concerns (Foucault, 1980; Rorty, 1982, 1989; Seigfried, 1990, 1996; Thayer-Bacon, 2000, 2003).

The postulation of a mind-independent reality is how Peirce escapes the charge of relativism often leveled against pragmatism, that pragmatism is relativistic due to its relying on a view of reality as a function of human

belief and a view of truth as a function of human practice. One of James's contributions to pragmatism is his willingness to let go of realism. James soundly critiques the epistemological/metaphysical split that worried Peirce and himself as graduate students by critiquing a false metaphysics, that separates experience from existence, and dissolving the absolute/relativist distinction with his theory of radical empiricism and pluralism.

James' (1909, 1975) radical empiricism starts with the postulate that "the only things that shall be debatable among philosophers shall be definable in terms of experience," and goes on to define experience not just in terms of things or events, but "the relations between things, conjunctive as well as disjunctive" (*The Meaning of Truth*). For James (1912, 1976), experience has no inner duplicity; it just is, in its pure thatness. Experience is subjective and objective, it is private and public, it is internal and external, it is thought and thing. What we do with pure experience, when we categorize and separate it and create lines of order for it, is by way of addition, not subtraction to pure experience (*Essays in Radical Empiricism*). Experience can serve different functions and may be different kinds. In different contexts, it plays different parts. With James's radical empiricism, it is not contradictory to say that experience is both absolute and relative; it is both particular and universal.

James (1909, 1977) argues in *A Pluralistic Universe*, that the world we experience is more than we can describe. He describes reality as genuinely continuous and active, it is not a closed system but ontologically open. Our theories are incomplete and imperfect, for our concepts, as valuable as they are, stay on the surface of things. They are not capable of capturing reality's continuously changing nature. James shows over and over again how philosophy is guilty of vicious intellectualism due to its reliance on the use of concepts that are unable to penetrate and capture the flux and depth of concrete reality and experience. It is not that certain concepts are limiting, but that all concepts are. James describes how first we have immediate experience, his radical empiricism, the mere thatness of experiences. Then we frame our experiences, we name them, with concepts. However, our immediate experiences always overflow concepts and logic, conjunctions and disjunctions. Concepts start as a method for us; then they become a habit, and finally a tyranny. Concepts fix things and try to tie them down, and by doing so they help us avoid context. Whenever we try to describe our experiences something always escapes, our logic always omits something.

James represents America's live-and-let-live individualistic approach, in that he argues for the sacredness of individuality. He embraces the value of pluralism in terms of valuing diverse cultural perspectives, thus avoiding the charge of elitism others have made against Peirce. James's individualism caused him to be accused of relativism in his

lifetime; however present-day feminist scholars have demonstrated how he avoids this charge through his dissolution of the absolutism/relativism bifurcation with his theory of radical empiricism and pluralism (Seigfried, 1990). Yet, his individualism leaves him vulnerable to present-day concerns about his blatant sexism and racism that are evident in his writings, both of which are representative of his historical and cultural embedding (Seigfried, 1996). James critiques himself on some of his biases in various essays such as 'On a certain blindness in human beings,' but misses other important limitations, thus confirming Pierce's theory of fallibilism (see Garrison *et al.*, 2002 for present-day essays concerns James).

Along with James' (1899, 1958, 1890, 1950) concept of experience, another key concept of his is habit (See *Talks to Teachers on Psychology*, and *The Principles of Psychology*). James looks at habit in terms of individual habits, how these develop over time and form the basis of learning, how they link the body and the mind, how they help to economize and coordinate our actions, and how an understanding of habit can be used by teachers to not only help manage their classrooms, in terms of students' behaviors, and avoid problems such as students' balky wills, but also help teachers establish helpful habits of inquiry for their students. Children learn to ask questions and to pay attention to their doubts because they are taught to do so and have the chance to practice inquiry skills. Dewey (1910, 1938, 1965) further developed James's habits in terms of habits of thinking (*How We Think*) as well as with his two principles of interaction and continuity (*Experience and Education*). However, before diving into some of John Dewey's key ideas, we turn to his colleague, friend, and neighbor, George Herbert Mead, as Mead's social turn influenced Dewey's thinking as well.

Mead was a philosopher and a scientist, a psychologist and a sociologist during a time when those fields of study had fuzzy boundaries. Mead (1934) describes his theory as social behaviorism in his famous social psychology course (*Mind, Self, and Society*). Mead argues that we become individual selves with minds out of our social relations with others, thus turning upside-down psychology's focus on individuals *qua* individuals as well as classical liberalism's view of the individual as separate and autonomous from society. Mead argues we are first of all social beings who then become individuals. His fundamental approach to understanding human behavior is from the point of view of conduct, particularly conduct as it is observable by others. Again, we find a theory that focuses on action as its fundamental datum.

Mead tries to go back to the early stages of social acts preceding symbols and deliberative communication, apart from some form of consciousness, in order to understand how we develop a mind and self. He looks historically at early human beings' conduct as well as studying infants' experiences, both as they are prior to language. Mead

argues that we must assume that individuals are brought into essential relation with the social process before communication so that contact with the minds of others is possible, in order to avoid solipsism. First, we have experiences within a context – we have the conduct of the individual in the environment. This environment is a social one. It is through communication by a conversation of gestures in a social process that our consciousness develops, our mind arises, and we develop a self. For Mead, gestures are our first means of communicating, the beginnings of social acts. When a gesture means an idea and it arouses that idea in the other, then we have a significant symbol. A significant symbol signifies a certain meaning, and we call this language. When our gestures have shared meaning and become significant symbols, then we are able to think. Mead defines thinking as “simply an internalized or implicit conversation of the individual with himself by means of gestures” (p. 47).

Mead explains that in order for an individual to become conscious of the meaning of her own gestures, she must be able to take the attitude of the second individual toward that gesture, and respond to it in the same way. As we are able to take the attitude of the other toward our own gestures, our consciousness evolves. So that the mind, or intelligence, is defined by Mead as “the taking of the attitude of the other toward one’s self, or towards one’s own behavior” (p. 48). When a gesture indicates to another the same that it indicates to the individual, then it has meaning. This is why Mead says that meaning is a social act. When our gestures call out the same response in the other as ourselves, then we are unconsciously seeing ourselves as others see us. For Mead, the most important gesture we have is the vocal gesture, as vocal gestures allow us reflexivity, the ability to turn back our own experiences upon ourselves. The conversation of gestures is the beginning of communication.

Mead’s theory of self-development is that the self is not there at birth, but develops in the process of social experience and activity by becoming an object to itself. Psychology’s essential problem, according to Mead, is explaining self-consciousness: how can the self get outside of itself to become an object to itself? Mead’s answer to this problem is that the self becomes an object to itself by taking the attitude of others toward her within the social environment. The language process is essential for the development of the self, for communication provides a form of behavior in which the individual may become an object to oneself. As Mead’s theory of self-development is based on social construction, he must address concerns of social determinism. He does this by describing our subjective experiences, our memory, and imagination, as being accessible only to the individual, the I, while the social side of us, the me, stands for the self I am aware of, the conventional self based on the organized attitudes of others. The I is unpredictable and uncertain and exists in

the now; it is unknowable. It is the self’s action over and against the social situation. The me acts like a censor, setting the limits of the self through social control. Together they form a self, which adjusts to others’ attitudes and at the same time changes others’ attitudes.

There are several issues with Mead’s theory with which present-day scholars wrestle with (see Biesta, 1999; Thayer-Bacon, 2003). In the effort to heal splits, Mead continues to assume splits, for example, between the mind/body in his discussion of the mind and self, between reason/emotions in his discussion of thinking and self-reflection, and between the public/private in his discussion of the social me and the qualitatively unique I. However, we find Dewey seeks to heal these splits in his own work as well.

John Dewey and George Mead were good friends and colleagues, and also neighbors living next to each other in Chicago. They greatly influenced each other, as they both were developing their individual theories during the time they worked together. Like Mead, Dewey holds a social behaviorist view of meaning. Dewey, with Mead, recognizes that individuals start out as members of communities. However, Mead’s individual tends to be subsumed by the generalized other, the social me disciplines and censors the I, while Dewey’s individual, even in its state of immaturity as a young child, interacts with and changes general society. Dewey distinguishes his views from Mead’s when he creates a transactional model that describes social groups affecting individuals and individuals affecting social groups (Dewey and Bentley, 1949, 1960). Dewey (1916, 1996) starts his famous *Democracy and Education* with a description of communities and self-development that is very much in agreement with Mead. However, for Dewey, while children are physically helpless and dependent on others, they are socially powerful. They are gifted at getting the cooperative attention of others and of attending with interest to the doings of others. The young, immature, affect the mature and the mature affect the young. Dewey’s significant contribution to democratic theory is based on his recognition of the interactive, interrelational, and interdependent qualities of individuals in relation to others (Thayer-Bacon, 2008). The way we measure the worth of any given mode of social life is by how much the interests of the group are shared by all of its members, and how full and free the group’s interactions are with other groups. For Dewey, democracy is defined as more than a form of government: “it is primarily a mode of associated living, of conjoint, communicated experience.”

In Dewey’s (1922) *Human Nature and Conduct*, he describes his psychology with a social emphasis (habits) based on an evolutionary framework (impulses), where habits constitute the self and mind and serve the regulatory role that the generalized other serves for Mead. Yet, Dewey allows for a greater potential for flux and

adjustability due to impulses. Habits are enforced upon us from the day we are born, by the very fact that we are social beings. Habits are imposed upon us through social custom. Yet, habits are acquired and secondary, not native to us as impulses are. In agreement with James, Dewey argues that impulses are what give us flexibility and diversity and habits are what give our impulses direction and shape.

Not only does James's concept of habit contribute significantly to Dewey's psychological theory, but James's concept of experience is also further developed by Dewey. In *Democracy and Education*, Dewey (1916, 1996) describes experience as having an active and passive element, trying and undergoing. "We do something to the thing and then it does something to us in return; such is the peculiar combination." In *Experience and Nature*, Dewey (1939) describes experience as "a double-barrelled word ... it includes *what* men do and suffer, *what* they strive for, love, believe and endure, and also *how* men act and are acted upon, the ways in which they do and suffer, desire and enjoy, see, believe, imagine ... It is 'double-barrelled' in that it recognizes in its primary integrity no division between act and material, subject and object, but contains both in an unanalyzed totality."

We find this trying and undergoing description of experience in Dewey's educational theory as well. In *Experience and Education* Dewey (1938, 1965) describes two principles, continuity and interaction, that we can use to interpret an experience for its educational significance and value. The principle of continuity emphasizes how every experience "both takes up something from those which have gone before and modifies in some way the quality of those which come after." The principle of interaction "assigns equal rights to both factors in experience – objective and internal conditions. Any normal experience is an interplay of these two sets of conditions." Dewey tells us, "every experience enacted and undergone modifies the one who acts and undergoes, while this modification affects, whether we wish it or not, the quality of subsequent experiences." An experience that has educational value is one that encourages further growth by arousing curiosity, strengthening initiative, and stimulating desires and purposes for effecting change in the environment.

Dewey developed further Peirce's idea of fallibilism and James' radical empiricism with his concept of warranted assertability. Dewey (1938, 1955) tells us in *Logic* that Peirce is the source of his ideas concerning logic. As he takes seriously Peirce's claim that we will never know this Reality, that it is a futuristic Ideal, Dewey distinguishes between ontological Truth (the nature of truth) and epistemic validity (the test of truth). For Dewey, logic is not separate from inquiry; logic is embedded in the contextuality of inquiry. He shows that interpretations of logical forms vary with underlying metaphysical and epistemological assumptions.

As knowledge has come to mean a fixed eternal end, rather than that which satisfactorily terminates inquiry, Dewey adopts the concept warranted assertability to stand for knowledge as a general abstract term related to inquiry in the abstract. Dewey argues that we need to focus on the agreement process that we use to try to establish epistemic claims – inquiry, for this helps us understand that the testing of truth is a social practice. Certainly today's postmodern, feminist, and critical theory work, all support Dewey's argument, and in fact turn to Dewey to help them make the case for the importance of addressing power issues in discussions of truth. Dewey argues that as fallible, contextual human beings, we have access only to truths derived from our own error-prone yet self-correcting procedures. Even our logical forms are developed within our own contextuality. Therefore, our truth claims are forced to be tentative and revisable, and any argument for Truth as corresponding to reality or as coherence must fall back on warranted assertability in practice.

As Dewey's (1938, 1955) central argument is that all inquiry (and logic itself) is affected by philosophical assumptions that are culturally bound, he recommends a view of inquiry that embraces the self-developing and self-correcting nature of scientific inquiry (*Logic*). The scientific model reminds us that individual conclusions are only hypotheses until agreement can be reached. Again we find science's influence on pragmatism. Dewey (1916, 1996) describes his recommended scientific approach to inquiry as reflective thinking in *How We Think* and *Democracy and Education*. As a pragmatist, Dewey shows us that thinking begins with a situation of experience, then a problem develops within this situation (felt need); this triggers the person to seek a solution (hypothesis) by gathering information and making observations (reasoning/solutions), and testing out the ideas by application (testing hypothesis/ideas). His model for reflective thinking is like the scientific method and it has greatly influenced more recent debates concerning critical thinking.

Dewey (1900, 1902) sought to connect his philosophical theory to educational practice, and can be credited as a father to philosophy of education in America. While working at the University of Chicago, he started a lab school that became famous and still exists, a place where educational theories could be tested out in the daily practice of teaching children (see *The School and Society* (1900), and *The Child and the Curriculum* (1902)). His own children attended this school and his wife, Alice, served as principal (Seigfried, 1996). One cannot go to an educational conference in the USA without hearing Dewey's name being referred to, and often his name is in the title of the presenting papers. It is impossible to overstate his influence on American education. However, one will find varying interpretations of Dewey's work, and he is often critiqued for crimes that others argue he never

committed. It is easy to misread Dewey, for he uses common language in unique ways, and one must look carefully for his definitions and examples to make sure the meaning is correctly understood (Biesta and Burbules, 2003; Garrison, 1995, 1997).

Conclusion

It makes sense that pragmatism developed in the USA, for the roots of pragmatism are in the developing culture of the country as well. Pragmatism is about connecting theory to action, showing what the result of that idea is. Pragmatists seek to connect thinking to doing, experience to the outcome of directed action, so that philosophy is not just about abstract ideas, but is also about trying to have an impact, make change, and solve real-life problems that people have. Pragmatists seek to get philosophy out of the ivory tower of the university and out in the community with the philosophers' task viewed as one of social activism. They are dreamers of what should be ideally – poets, prophets, and soothsayers who seek to improve people's everyday lives. Pragmatists take on classical philosophers from the past and offer strong criticisms of their work, rejecting central problems and dichotomies such as mind/body, reason/will, reason/emotion, thought/purpose, facts/values, self/others, appearance/reality, and theory/practice as they seek to heal splits and dualisms that others have created and offer a unifying description of the world. In the healing of dichotomies that have developed over time, pragmatists bring together various fields of study to help them. They turn to history, psychology, sociology, anthropology, for example, to help them in their ameliorative efforts, thus making their work interdisciplinary. They are also pluralists, not only in terms of recognizing the value of diverse perspectives, experiences, values, and meaning, but also in terms of recognizing that there are multiple possible answers or solutions – multiple truths. They seek to work in collaboration with others and create a community of inquirers as they do not think that on their own they are as likely to be able to solve the problems with which they wrestle. Pragmatists want philosophy to act like science in terms of being able to question one's theory at a deep level, all the way down to one's criteria, methodology, and basic assumptions, as well as one's situatedness.

Bibliography

Biesta, G. (1999). Redefining the subject, redefining the social, reconsidering education: George Herbert Mead's course on philosophy of education at the University of Chicago. *Educational Theory* 49(4), 475–492.

- Biesta, G. and Burbules, N. (2003). *Pragmatism and Educational Research*. Oxford: Rowman and Littlefield Publishers.
- Dewey, J. (1900, 1902, 1990). *The School and Society [1900], The Child and the Curriculum [1902]*. Chicago and London: The University of Chicago Press.
- Dewey, J. (1910). *How We Think*. New York: Heath.
- Dewey, J. (1916, 1996). *Democracy and Education*. New York: The Free Press, MacMillan.
- Dewey, J. (1922). *Human Nature and Conduct*. New York: Henry Holt.
- Dewey, J. (1938, 1955). *Logic: The theory of inquiry*. New York: Henry Holt.
- Dewey, J. (1938, 1965). *Experience and Education*. New York: Macmillan.
- Dewey, J. (1939). *Experience and Nature*. New York: G. P. Putnam's Sons.
- Dewey, J. and Bentley, A. F. (1949, 1960). *Knowing and the Known*. Boston, MA: Beacon Press.
- Foucault, M. (1980). *Power/Knowledge*. New York: Pantheon.
- Garrison, J. (ed.) (1995). *New Scholarship on Dewey*. Dordrecht: Kluwer Academic.
- Garrison, J. (1997). *Dewey and Eros: Wisdom and Desire in the Art of Teaching*. New York: Teachers College Press.
- Garrison, J., Podeschi, R., and Bredo, E. (eds.) (2002). *William James and Education*. New York: Teachers College Press.
- James, W. (1890, 1950). *The Principles of Psychology*, vol. 1. New York: Dover.
- James, W. (1899, 1958). *Talks to Teachers on Psychology*. New York: W. W. Norton.
- James, W. (1909, 1975). *The Meaning of Truth*. Cambridge, MA: Harvard University Press.
- James, W. (1909, 1977). *A Pluralistic Universe*. Cambridge, MA: Harvard University Press.
- James, W. (1912, 1976). *Essays in Radical Empiricism*. Cambridge, MA: Harvard University Press.
- Mead, G. H. (1934). In Morris, W. (ed.) *Mind, Self, and Society: The Standpoint of a Social Behaviorist*. Chicago: University of Chicago Press.
- Rorty, R. (1982). *Consequences of Pragmatism*. Minneapolis: University of Minnesota Press.
- Rorty, R. (1989). *Contingency, Irony, and Solidarity*. Cambridge: Cambridge University Press.
- Seigfried, C. H. (1990). *William James's Radical Reconstruction of Philosophy*. Albany, NY: State University of New York Press.
- Seigfried, C. H. (1996). *Pragmatism and Feminism: Reweaving the Social Fabric*. Chicago: University of Chicago Press.
- Thayer-Bacon, B. (2000). *Transforming Critical Thinking: Thinking Constructively*. New York: Teachers College Press.
- Thayer-Bacon, B. (2003). *Relational "(E)pistemologies"*. New York: Peter Lang, Publishers.
- Thayer-Bacon, B. (2008). *Beyond Liberal Democracy in Schools: The Power of Pluralism*. New York: Teachers College Press.
- West, C. (1989). *The American Evasion of Philosophy: A Genealogy of Pragmatism*. Madison, WI: University of Wisconsin Press.

Further Reading

- Peirce, C. S. (1933–58). In Hartshorne, C., Weiss, P., and Burks, A. (eds.) *Collected papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.
- Peirce, C. S. (1958). In Wiener, P. P. (ed.) *Values in a Universe of Chance: Selected Writings of Charles Sanders Peirce (1839–1914)*. Garden City, NJ: Doubleday.

Psychoanalysis and Education

P Smeyers, Ghent University, Gent, Belgium

© 2010 Elsevier Ltd. All rights reserved.

Reconsidering Psychoanalysis

The entire enterprise of education has become colored by means–end rationality. The other, with whom the teacher is confronted, is treated first and foremost as a recipient for knowledge or training whose content has already been set out in advance. Some have, therefore, turned to psychoanalysis as a model for a very different way of relating to the learner. It can be understood as a nontechnical enterprise in which issues of language are in the foreground, where the educational and the ethical no longer go separate ways, and where the encounter with the other precedes understanding. At its core are susceptibility, vulnerability, and openness rather than any kind of knowledge as may be conventionally conceived. The climate of opinion has moved fairly conclusively against psychoanalysis – especially concerning its scientific standing. Yet, Freud often laid claim to being a scientist. It is familiar, for instance, that he is inclined to see the mind in terms of fluid mechanics – as if psychic energy swirls around the system, gaining in pressure the more it is repressed and sublimated, and emerges in dreams, slips of the tongue, and so on like puffs of steam; or, to take another example, that the meaning of dream symbols is fixed as if they were mathematical constants. Freud's followers and translators, generally, perpetuated this picture. Of course – given the prestige of science in Freud's time, as well as in ours – this is unsurprising. Is there a way to understand psychoanalysis, if not as some kind of science?

The views of philosopher Ludwig Wittgenstein are illuminating on this question. He disparages Freud for “constantly claiming to be scientific” (Wittgenstein, 1966: 44) yet is reported (*ibid.*, p. 41) as saying: “I happened to read something by Freud, and I sat up in surprise. Here was someone who had something to say.” The mistake, according to Wittgenstein, is to take physics as our ideal science when studying psychology (Wittgenstein, 1966: 42): “We think of formulating laws as in physics.” We should rather think of Freud as offering us a new mythology to live by:

Freud refers to various ancient myths . . . and claims that his researches have now explained how it came about that anybody should think or propound a myth of that sort . . .

Whereas in fact Freud has done something different. He has not given a scientific explanation of the ancient myth. What he has done is to propound a new myth.

The attractiveness of the suggestion, for instance, that all anxiety is a repetition of the anxiety of the birth trauma, is just the attractiveness of a mythology. (Wittgenstein, 1966: 41)

Freud

In giving us a story about the unconscious, repression, and so on, Freud is doing the same kind of thing as when he reminds us of the story of Oedipus. In both cases, he tells us “life's like this.” Wittgenstein declared approvingly of Freud's work that “It's all excellent similes” (Monk, 1990: 357). On this account, the value of the myth of Oedipus is not that it prefigures or dramatizes the Oedipus complex – the latter being seen as a piece of psychopathology, independently and scientifically knowable, which bears the Greek name for the sake of vividness. To make the acquaintance of the myth is rather, in itself, to learn something helpful – in that the myth depicts a pattern in which aspects of our lives fit and make sense. It is – Wittgenstein says – speculation rather than insight (e.g., into the timeless human condition). “It is something which people are inclined to accept and which makes it easier for them to go certain ways: it makes certain ways of behaving and thinking natural for them. They have given up one way of thinking and adopted another” (1966: 44–45). Mythological explanations have the attraction that they say “this is all a repetition of something that has happened before. And when people do accept or adopt this, then certain things seem much clearer and easier for them. So it is with the notion of the unconscious also” (Wittgenstein, 1966: 43).

Myth supplies a new language, a new framework for looking at the world. It offers both the patient and the analyst an alternative to languages and frameworks which seem necessary, written in the stars, even obsessive; meanings that appear fixed and determinate. An example may help here. A young woman found it difficult to resist her parents' blandishments to return to the family home which she alleged had been a place of abuse during her childhood, and visits back to which correlated accurately with what her psychiatrist diagnosed as subsequent psychotic episodes. She was, on one level, well aware of the damage these visits did her, but repeated that “mum was only trying to be kind,” that it was nice to be in a house that was well decorated (compared to her student accommodation), even if rather fussy, with everything color-coordinated

and ubiquitous potpourri; and that “there was always food available” and that mother constantly joked about “fattening her up.” At one point, the young woman’s counselor commented that home sounded like the gingerbread house (from the fairy tale of Hansel and Gretel: the children are lured there by a witch who only wants to fatten them up to eat them). This apparently gave the young woman a metaphor she found helpful: a way of acknowledging that home was both sweet and threatening, potentially fatal. It appears it released her from the fantasy that home is home after all (this one is, indeed, home, sweet home) and from the difficulty of acknowledging its threats. Wittgenstein says that “it may then be an immense relief if it can be shown that one’s life has the pattern rather of a tragedy—the tragic working out and repetition of a pattern . . .” (Wittgenstein, 1966: 51): or, we might say, the pattern here of a fairy story.

To summarize: in its focus on language, psychoanalysis reminds us that we live in and through language: language is not a neutral instrument. It offers us a new myth in the form of its own characteristic frameworks of understanding (the anal stage, the oral stage, and so on). The technical jargon that it generates is both liberating and dangerous. It is one of the best insights we owe to Freudian psychoanalysis that language is not an indifferent medium through which we conduct our value-neutral scientific interventions. An individual might reflect that the development of a richer and more responsive framework of language (to replace fixed concepts with absolutely determinate meanings) is clearly as much a task for the teacher, at every level, as it is for the psychoanalyst. Moreover – as a major task for the therapist is to see through, and help patients to see through – there is also another interesting parallel. The analyst is not concerned to hand over solutions – readymade or otherwise – but to help the patient learn to find his or her own way to solutions, and this not as a matter of acquiring techniques but of the reorientation of self to the world. What is worth taking seriously in psychoanalysis is, thus, parallel to what is worth taking seriously about some aspects of education.

Lacan

Children’s lives are, to a large extent, led according to rules fixed by adults. Perhaps, this situation is not to be deplored – they are, after all, not yet adults, not yet held responsible in the full-blown sense. Yet, it remains hard to deny that their freedom is limited. People – and children as well – desire particular things, and what comes in the way of fulfilling these is experienced as restraint. In the relationships between teachers and parents – and teachers and children – desire works in different ways. There is, first, what they want for themselves and for this matter from each other, but there is also what they want the other

to desire. In addition, last but not least, there are cases where it is not clear what exactly one wants from the other, where this can only, generally, be indicated as recognition for one’s own desires. The French philosopher and psychoanalyst Jacques Lacan (for a more elaborate general introduction to Lacan, see Benvenuto and Kennedy (1986)) developed a theoretical position in which a number of these matters are dealt with in a highly sophisticated manner. This position sheds light on how others are necessarily always there, and on how their expectations are – for the subject, at the same time – both unavoidable and liberating, that is, they enable her desire to find a pathway for expression. Lacan’s subject is not the subject of the *cogito*, that is, the subject that consciously assures itself of itself in its representations. It is split – separating itself from itself in the very act of self-representation and disappearing into the gap between the *cogito*’s enunciation and its statement: in Lacan’s words, the subject thinks where she is not, therefore, she is where she does not think. Lacan radicalizes Freud’s epistemological and psychoanalytical position – interpreting it not so much as a matter of surging energy or a hydraulic metaphor of the instincts, but rather in terms of particular representations that will recur and recur. Psychoanalysis is itself – in his opinion – above all, based on a fundamental split between the subject and the knowledge she has of herself. Lacan looks at the turning points of the discourse not from a purely linguistic point, but as the effect of a meaning which escapes its own signification. The order of language replaces the subject’s lack of being – from which the desire for the missing object arises. There are always other desires and other objects which present themselves as being able to fill up the lack of being. The formation of the subject (of the one who says ‘I’) is explicated by Lacan through what he calls the mirror phase. The helpless infant – not yet objectively in control of her movements – jubilantly perceives in the mirror the mastery of her bodily unity, which objectively she still lacks. She becomes aware – through seeing her image – of her own body as a totality, as a total form or *Gestalt*. The subject’s identification with her own body as other than herself structures herself as a rival to herself, and the separateness and the rivalry generate aggression. This structure remains present throughout human life and characterizes all human interactions. In dealings with the Oedipus complex, Lacan emphasizes the function of the lack of the object and lack in general. The end of the original unity of mother and child is precisely what leads to the subject’s possibility of becoming a subject of desire. Without the law – which deals with relations with others – the mother can autonomously, in a sense arbitrarily, deal with the child’s wants. In the unmediated realm, the other is not bound by any law and the child is a possible toy of her changing answers. The presence of a third person – one could say (literally or figuratively) the father – breaks this fusion-like unity open. He is also the

one she desires. In this sense, the child cannot be everything for her mother and she is not only there for her. The relationship between mother and child is mediated by a third who represents the symbolic law of the culture.

The separateness already indicated in the mirror stage is further affirmed when the child enters the symbolic – the social order of language and culture. Before the subject can act as a self, it is projected and absorbed in the universal order of language and culture. The child has to settle herself in this order. The symbolic order does not consist only of a system of different meanings (differences), but also of a complex network of discourse, conversations of the Other. The term discourse refers to what is said, to the conversation, to its content, to what is regarded as important. Because of that network, the subject is not only subjected to the formal system of signifiers, but also to the way in which these are organized in the discourses of others, and in other words subjected to a historically determined meaning. The child is absorbed as a part of a network of tales (discourse) told by others about him, some of them even before he was born: “It is a boy. Doesn’t he take after his father?” This network consists of a number of prohibitions and commandments, desires, anticipations, obligations, and value-judgments to which the child has to accommodate herself. These discourses are contradictory and do not give a coherent image of herself. As a consequence, the subject can never satisfy their demands.

The unconscious images – the discourses of the Other that cannot inscribe themselves in the subject and that cannot be accommodated by the subject – stay active in the mind and come to the surface in all kinds of symptoms unavoidable for the subject. They cannot be accommodated – are at war with the already existing images in the subject. The subject, the ‘I’ who speaks of herself, is not the same as the ‘I’ that is spoken of: the meaning surpasses her. The subject (the one who says ‘I’) is not master of herself, but subject of the symbolic order: “*Je est un autre [I is another]*” (Lacan, 1966: 118). The subject arises in relation to desire that is unknown to her. The Other is the real witness and guarantor of the subject’s existence, as it is she who can recognize the subject. In the perception of the subject, the Other is not affected by the same lack, and can be identified with the mother’s original role in relation to the infant. The Other is where the subject is born, not only as a biological entity, but as a subject with a human existence.

From the moment of initiation into the symbolic order, the subject can never reach anymore what she, in the end, desires. That is what keeps the desire going. However, it is also the symbolic law of the language and culture that makes it possible for the child to withdraw herself from the suffocating omnipotence of the Other. There is always the desire of the Other – who recognizes or does not recognize her desire, who approves or rejects, and allows or punishes. To develop in a meaningful way, the desire

needs the recognition of the Other. Thus, the desire of a human being becomes the desire of the Other – “*le désir de l’homme, est le désir de l’autre*” (Lacan, 1966: 693). Unlike Freud’s version in the chain of conscious reasons, the Lacanian unconscious will not provide the answer one cannot find on a conscious level that leaves no gaps. The unconscious reminds the subject that reasons will never really do, it marks the ineliminable impossibility of understanding human desire in a definitive way. The truth of the unconscious – that human desire is not limited to what is good for herself – is something that the subject does not want to know. Because the Other is (necessarily) deceiving her, the psychoanalyst’s function is not to answer the subject’s appeals and demands, but, instead, to act so that the answer comes back to the subject from the analyst as a question: What do you want from me? What the subject lacks or has lost is not present in the Other.

Implications for Education

The message for education may go back to something basic: the importance of being surrounded by people who care and, indeed, demand a lot of those whom they care for. Nothing is, indeed, more violent than being left out, not desired at all by the Other – at the extreme, not being able to enter the symbolic order. Lacan’s position hints at a new ideal of personhood – interpreting what one does and does not do to understand better how and why one desires what one particularly longs for. Maybe all that in order to be able to accept (at last, and each time anew) who one is, to find at least for a moment one’s peace, and one’s truth.

Turning to the context of formal education, one can see that the thoughts and actions of teachers and students are influenced by the imaginary, symbolic world of the school. Even if the student thinks and speaks to herself, the Other exists, choosing the words in language that can best be understood by her. The Other becomes a central concern, ‘it’ controls the infant’s first associations with her parents and forces her to attend to the world, teaching her to see it in terms of logic, rules, and orderliness. Schools make an important contribution to society by reproducing the *status quo* – both in its intellectual and cultural variants. Despite this, they have been able to maintain the fiction of neutrality and autonomy – providing significant ideological services to a capitalist society. In this instance, the culture of the school is conceived as transmitted through pedagogic practices, validated by the political power of the state. These practices, in combination, provide students with a language that defines them and their place in the social system, definitions of what is important for youth growing up in the inner city, for instance. In reality, a litany of traditional values and beliefs are recited to them which seek to define what is good and

bad, who is intelligent and stupid, and what good manners ought to be in the classroom. Good taste and the norms of educational decorum are placed in front of children as a constant reproach: this is how you should act if you wish to be considered an intelligent and worthy person. Thus, the socially correct attitudes and behaviors of classroom life are taught. Children are not to chew, fight, or throw things. They are not to leave their desks or talk without permission. The teacher has the right and duty to control their bodily movements and gestures, can force them to sit still and be silent, and can punish and humiliate them with a word. Rothstein (1993) argues that schools have always been tied to other agencies of capitalism, even as they denied such links. They have arisen from social and political customs and from the functional needs and beliefs of those who funded them. In addition, in pursuing their own purposes, "they [have] developed a special history and practice, borrowing methods from mental hospitals, military and penal institutions, and the charity schools of the early nineteenth century" (Rothstein, 1993: 80). The organization of education systems seems to be opposed to change, insisting on a return to basics and a simpler past.

The insistent demands of Otherness are taught in the classroom, a world of words sustained by teachers where the primary Other is the teacher. It is the teacher who has the power to decide, who is the one pupils must attend to, who speaks where they listen. A collective voice of the Other is constructed – reflecting the background and levels of attainment of the children. Only internalization of the mores, moral understandings, and language of the classroom permits students and teachers to speak and to be understood. It is, however, not something they consciously plan or execute. Early in their schooling, children become aware of the Otherness that teachers and other students represent. The Other, in these instances, is never an actual person, but always an unconscious construct in which the word is located and used by children and teachers alike. With their servile silence and submissive gestures, students seek to soften the coerciveness and authority of the all-powerful adult, summoning up and transferring behaviors and identifications from previous experiences with all-powerful parents. Mostly students seek recognition and acceptance as worthwhile and competent persons, and some measure of assurance and predictability in their relationship with the demanding teacher. 'A very good answer' is not only an encouragement and praise, but also creates the dangers of dependency and submissiveness. As long as the teacher ignores her role as a state-employed worker – as long as she is ignorant of the social functions of schooling in modern life, she cannot understand her true relationship to her students, Rothstein (1993) argues.

In this case, the unconscious is that part of an educational system's history that is marked by mythology, folklore,

and amnesia. It is the proscribed past that has forgotten its own roots and social functions. It can be uncovered in the words and phrases that teachers and students use when they speak to one another. It can be discerned in the traditions and folklore of schooling, and in the normative and ethical structures used to validate pedagogic practices. Every experience of failure and rejection in classrooms leaves personal scars. The humiliations of student experiences are, often, repressed and unavailable for conscious retrieval. But what has been forgotten still exists and affects the way students and teachers act. As the context of education reveals itself as serious and important, the child represses her social needs and desires. She accepts her subordinacy and regimentation because all of her own kith and kin have advised her to do so, and because of her own dependent status in her family and communal relations.

When small children are doing well in mathematics, they might, sometimes, become fearful if confronted with more complicated problems. The child needs to be disturbed out of this kind of complacency and needs to be made to confront the mathematics problem. Instead of giving the correct answer, the teacher might try to help them to feel more secure, that they are up to this – thus not taking this task away from them. At the other extreme, in the writing of a PhD thesis, the adult student – while also working at a full-time job – frequently struggles with the question of where it all leads to. Such a student will go through periods where the whole thing seems impossible – she thinks she is not able, she does not have the time to do the topic justice. However, perhaps, the struggle is internally related to the quality of the learning and to what is produced; a student's PhD work can come too easily. This may also more broadly be true of education. It points to the inappropriateness of separating what is learned and the learning involved from the person of the learner in a broad sense, although not of all learning, of course – for example, learning to type. Here, the supervisor should not try to answer but return the question and its difficulties to the student. Although she might try to re-channel the candidate's energy when she goes through a difficult time, it would be wrong for her to offer answers. Constantly available and maintaining that there are ways forward, she should refrain from giving one – instead, creating the opportunities for the student to find one for herself. It would be a betrayal to her real interests if the student were not frustrated, comforted but not confronted with real learning itself. Another case is the youth who fantasizes about a future career. There is an incredible range of choice in higher education and there is a tendency with the rapidly improving information and guidance on these matters to treat the decision in an excessively rational way. Again, the student ought to become frustrated. The careers counselor should acknowledge her desire and indicate alternatives, but she should stay away from indicating the

answer. Her project ought to be a negative one, as, again, it is the student who has to find for herself what she wants to do. As for the pilgrim of earlier days, it is most important to travel. Here, as elsewhere, there is the message that you cannot know what you really want unless you are confronted with something. In addition – as in sport – frustration and failure are, sometimes, desirable in teaching. This is at odds with finding the correct answer as suggested by the technological spirit in education; rather, it points to the idea of self-realization in progressive education.

The Lacanian position and psychoanalysis, more generally, thus challenge several pictures that held us captive – not the least of which is that we know what we really want. In stressing the third-person perspective and the way the infant is fixed into this order, Lacan finds a subtle balance. He reminds us over and over again that there are many rooms in the house of being and that transgressing this border by trying to delimit them is dangerous. An estrangement from herself will destroy the subject and put humanity itself at risk. (For an elaboration see Blake *et al.*, (1998, 2000) and Smeyers *et al.* (2007) on which this article is based.)

Bibliography

- Benvenuto, B. and Kennedy, R. (1986). *The Works of Jacques Lacan. An Introduction*. London: Free Association Books.
- Blake, N., Smeyers, P., Smith, R., and Standish, P. (1998). *Thinking Again: Education after Postmodernism*. New York: Bergin and Garvey.
- Blake, N., Smeyers, P., Smith, R., and Standish, P. (2000). *Education in an Age of Nihilism*. London: Falmer Press.
- Lacan, J. (1966). *Écrits*. Paris: Editions du Seuil.
- Lacan, J. (1973). Seminar on "the purloined letter" Mehlman, J. (trans.) *Yale French Studies*, **48**, pp 38–72.
- Monk, R. (1990). *Ludwig Wittgenstein. The Duty of Genius*. London: Jonathan Cape.
- Rothstein, S. W. (1993). *The Voice of the Other. Language as Illusion in the Formation of the Self*. Westport: CT Praeger.
- Smeyers, P., Smith, R., and Standish, P. (2007). *The Therapy of Education*. Houndsmills: Palgrave Macmillan.
- Wittgenstein, L. (1966). *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*. (Compiled from notes taken by Smythies, Y., Rhess, R., Taylor, J., and Barrett, C. (ed.)). Oxford: Basil Blackwell.

Social and Cultural Capital in Education

M Print, University of Sydney, Sydney, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cultural capital – A broad theory to explain differences in societies such as class differences.

Formal curriculum – The formally planned and organized set of learnings for students which usually takes the form of the subjects that compose the school curriculum.

Informal curriculum – The nonschool subject learning in schools that includes activities such as participation in meetings, clubs, student governance, and fundraising which are largely unplanned by the school compared with the formal curriculum.

Social capital – This refers to connections among individuals that facilitate collective action and includes trust, norms, and networks of association representing any group which gathers consistently for a common purpose.

Many forms of capital – economic, human, cultural, and social – have been identified by social scientists. These are mostly seen as input measures to explain variations in phenomena in society. For example, differential amounts of capital may explain differences between classes, or why some students perform better at school than others and so forth. This article examines the relationships that social and cultural capital have with education in general and schools in particular. In these contexts education is usually considered as an input variable that affects the level of social and cultural capital. Consequently, increasing levels of education, it could be predicted, should produce higher levels of social and cultural capital. While this is an intriguing possibility this article will concentrate upon the nature of the relationship between social capital and education and particularly on the potential of the school in facilitating the accumulation of social capital by individuals and society.

Social Capital

Social capital is a seductive concept. It offers much, tantalizing researchers and policy makers with a procedural solution to unsolvable problems. As Portes commented "... social capital has evolved into something of a cure-all for the maladies affecting society at home and abroad." (1998, p. 2) Its ability to deliver solutions, however, has

been seriously questioned (Portes, 1998; Baron *et al.*, 2002; Dika and Singh, 2002). In the 1990s social capital burst upon the academic scene driven by the popular work of Robert Putnam and its appeal to major intergovernmental agencies such as the World Bank and the Organization for Economic Cooperation and Development (OECD) as it addressed major issues of civil society, economic development, and poverty (World Bank, 1999; Woolcock, 2001). For the World Bank social capital is a useful organizing concept to explain levels of social cohesion which, in turn, is critical for economic development and sustainability (World Bank, 1999).

It was soon realized, however, that the construct was neither new nor without controversy. Some have traced its roots to even the late nineteenth century, though most researchers attribute the work of Bourdieu (1986) and Coleman (1988) as the meaningful beginnings. While most writers treat social capital as highly positive, it can also have a dark side (Bourdieu, 1986; Print and Coleman, 2003). Interestingly, in recent years social capital has, to a limited extent, consolidated its presence in a diversity of applied areas (such as welfare, urban planning, engineering, health, economics, organizational behavior, sociology, natural resources management) and yet waned academically as it appears to have failed to live up to its high expectations as a research panacea for explaining social phenomena. So what is this construct and why is it seductive?

Defining Social Capital

How one defines social capital depends in large measure which position one takes relative to the three key theorists – Bourdieu, Coleman, and Putnam. From a socialist perspective Bourdieu saw social capital as elitist, helping to explain the process of social reproduction. He envisioned social capital as a critical tool, with an analytical function that would contribute to a deeper understanding of the way different forms of capital interacted. Coleman argued that nonelites could also benefit from acquiring social capital, taking a more nuanced, sociological perspective to the construct. Putnam, however, took a political science perspective and used social capital to help explain political systems, especially democracy and how people might gain from increasing their stock of social capital.

Putnam's view that social capital refers to connections among individuals that facilitate collective action has received wide, though equivocal, support. These resources include trust, norms, and networks of association

representing any group which gathers consistently for a common purpose. In a political sense social capital is often seen as the glue that keeps democracy together. It is essentially about trust, cooperation, and networking. As Putnam argues, social capital enables people to resolve collective problems more easily, facilitates trust and interactions in communities, enhances interconnectedness amongst people, and facilitates the flow of information.

Many researchers have argued that healthy democracies are characterized by high levels of social capital. For example a norm of a society high in social capital is reciprocity, which encourages bargaining, compromise, and pluralistic political outcomes. Another norm is belief in the equality of citizens, which encourages the formation of groups that transcend traditional societal groupings such as class.

In a more pragmatic sense the World Bank (1999) refers to the norms and networks that enable collective action. It encompasses institutions, relationships, and customs that shape the quality and quantity of a society's social interactions. For the World Bank social capital consists of five dimensions – groups and networks; trust and solidarity; collective action and cooperation; social cohesion and inclusion; and information and communication.

A common theme amongst these definitions is that social capital facilitates social functioning. As such it consists of four main components – it resides within groups; it involves civic virtues such as trust, cooperation, and civility; it requires of participants that they be actively engaged; and it is mutually beneficial for members and the group (Print and Coleman, 2003).

However, to suggest that there is unanimity in definition or even conceptualization, let alone whether the construct is meaningful, is far from reality. Contested definitions abound (Portes, 1998; Newton, 1999; Lin, 2001) usually arising from a different perspective or the argument that earlier definitions are incomplete or inadequate.

Types of Social Capital

The literature, and particularly the extensive work of Putnam (1993, 2000), Woolcock (2001), and Field (2003), identifies at least three types of social capital and one could argue for a fourth:

1. Bonding social capital is more exclusive. This includes social networks amongst homogeneous groups of people such as family, friends, and neighbors. Bonding is referred to as dense, tight-knit, homogeneous social networks. For Putnam (2000) the best metaphor was sociological superglue. While envisaged as a positive form of social capital, bonding can also include negative social capital as in the case of criminal gangs.
2. Bridging social capital is more inclusive and includes social networks between socially heterogeneous groups

such as close friendships, workmates, and formal groups. These are more access oriented to enable access to valuable resources and information outside immediate networks. Putnam (2000) saw this as a form of sociological lubricant. This form of social capital is more positive as it builds bridges between different groups and so contributes towards more effective functioning of society, governments, and communities.

3. Linking social capital is where unlike people, in dissimilar situations, join together for mutual advantage (Woolcock, 2001). This refers to vertical associations that provide upward links to more powerful groups, agencies, and people. For example, a community might link with valuable outside members to build a wider pool of resources.
4. These are not discrete forms of social capital and more commonly societies would experience various combinations of the above.

Appeal of Social Capital

Why is social capital so alluring? Why is it sufficiently important to be adopted by intergovernmental agencies such as the World Bank? Social capital has wide appeal in the social sciences, and other disciplines, as a means of explaining differences between individuals and groups. In particular social scientists seek to understand and predict relationships, hopefully causal relationships, between social capital and other aspects of society. For example the following hypotheses suggest the appeal of social capital, especially to policy makers, although many social scientists would be tentative in applying them:

- The greater the social capital, the greater the confidence in government (and other institutions) which enable societies to function more effectively and prosper.
- The greater the social capital, the easier to organize support to resolve collective problems. People will be better able to solve collective problems if they cooperate, trust each other, and build networks together. They may perceive advantages in not cooperating but this is short-sighted. Addressing global warming is an example.
- The greater the social capital, the higher the percentage of problem-solving outside the governmental sector, while less social capital requires greater reliance upon authoritative controls. When people join together they tend to be more tolerant, more empathic to others' problems, and less cynical, and consequently, require less government persuasion to live harmoniously together.
- The greater the social capital, the more prevalent the norm of reciprocity (bargaining, compromise, pluralism) amongst people. This enables them to resolve problems more effectively.
- Similarly, the more social capital is present in society the higher the priority of the norm of equality. In high

social capital areas public space is friendlier, safer, cleaner, and has less crime.

- The greater the social capital the more it will assist those in disadvantaged areas and will lead to an improvement in the quality of their life and mitigate the insidious effects of socioeconomic disadvantage.
- In general, being involved in groups is beneficial to people and associational life (and its quality) builds social networks, enhances society, strengthens democracy, and can improve health and happiness.

The application of social capital has particular interests in providing explanation for a current phenomenon plaguing western democracies – declining political support and increasing political disengagement. Many social scientists have explored this approach in some depth (Putnam, 1993, 2000; Woolcock, 2001; Print and Coleman, 2003; Jennings and Stoker, 2004), as they search to explain why, particularly in western societies over the past few decades, social ties and political engagement have diminished. As Putnam argues, decline in social capital erodes political participation, interpersonal trust, as well as political trust, all of which have become abundantly evident more so in the United States, Canada, and Britain but also in varying degrees in Western Europe as well.

Education and Social Capital

Putnam (2000) and more recently his associates (Helliwell and Putnam, 2007) contend that education and social capital are closely linked. Education they believe is the best predictor of many forms of social and political engagement. They also note that while education levels have risen dramatically within the United States over the past half century, levels of social and political participation have not risen accordingly and on many measures have declined.

This paradoxical puzzle has concerned political scientists for some time though recent research reinforces the argument that as education levels rise so do key measures of social capital. The conundrum remains, partly because Putnam and others use gross indicators of education such as years of completed schooling and partly because there is little evidence of understanding what happens within schools (Print and Coleman, 2003). Greater insight into the contribution and impact of civics and citizenship education subjects and noncurricula aspects of schools would enable a deeper understanding of how schools may affect social capital development (Print and Coleman, 2003; Print, 2007).

Schools and Social Capital

As social capital can be acquired and expanded, schools could, and arguably should, play an important role in

building social capital especially in terms of affecting citizenship and democracy. For example, Print and Coleman contend that “Schools offer a potentially powerful opportunity for both learning about the elements of social capital and enacting components of social capital through related activities. [including] some form of civics and citizenship education” (2005, p. 129). The contribution of schools can occur through three sources – the formal curriculum, the informal curriculum, and the extra curriculum. The hidden curriculum may also contribute but given its nature it is not considered here.

Formal Curriculum

When Putnam suggests that schools may contribute substantially to building social capital through enhancing trust, networking, and participation, he, and others, are thinking about social capital in relation to the formal curriculum, that is, the subjects that compose the school curriculum (Croniger and Lee, 1991; Dika and Singh, 2002; Print and Coleman, 2003). That is, schools can contribute to social capital through teaching certain subjects (such as civics, citizenship, service learning) which builds knowledge, skills, and values within students. Citizenship education, for example, contributes to knowledge of society and government, building skills of cooperation and networking as well as facilitating values such as trust and civic virtues (Print and Coleman, 2003). In this way teachers facilitate the growth of bridging, bonding, and linking forms of social capital. This of course, assumes that schools and teachers understand social capital, how it works and why they should purposively facilitate it.

Putnam (2000) also addresses the relationship between social capital and schools from the other direction. He contends that children will have more positive academic and nonacademic outcomes (lower drop-out, higher academic performance, less truancy) if social capital is strong at home and parents engage with schools. Similarly, Croniger and Lee (2001) found that teachers are an important source of social capital for students. Where these teachers help students personally, influencing the quality of the student’s social networking, students are significantly less likely to drop out of school.

A group of Scottish researchers in the Applied Educational Research Scheme are investigating social capital in the context of school outcomes. They are exploring how teachers and schools may utilize social capital to enhance student school outcomes. A current study is investigating how students may build their personal stock of social capital, particularly to use in overcoming backgrounds of disadvantage.

Informal Curriculum

Within the school, social capital may be generated substantially in ways other than through formal curriculum.

The informal curriculum refers to the non-school subject learning in schools that includes activities such as participation in meetings, clubs, student governance, and raising funds for charities (Print and Coleman, 2003). These activities, largely unplanned compared with the formal curriculum, are ways of building social capital for, as these authors argue “It is specifically through the informal curriculum that students may acquire participatory skills and values, as well as knowledge, from engaging in activities such as conducting student councils, running school parliaments, raising funds on special days, interacting with the community through service learning and the like” (2003, p. 134).

Extra curricular activities may also contribute to building social capital through individual and group activities which are beyond the confines of the school curriculum. These activities are conducted outside of school hours and retain only minimal contact with the school, such as debating, charity fundraising and, in many countries, many forms of community service.

Issues in Social Capital

Social capital is far from unproblematic, both as a social science tool and as an explanation of social phenomena. Further theoretical work and empirical research is required to enhance the value of this construct. Meanwhile several issues can be raised about its effectiveness.

In terms of schools, building social capital should be seen as an important, integral task of education. This occurs, though not as much as expected and certainly less than many may like. Many parts of the formal school curriculum are undervalued in this role including school subjects such as civics and citizenship education.

Similarly, the informal school curriculum could play an even more important role in generating social capital within schools. Yet, as Print and Coleman note, this is problematic given the low status of the informal curriculum and “Significant further research is required in this area to gauge the impact of the informal curriculum on student potential to build social capital” (2003, p. 134).

There are also, more generally, potential problems with social capital should the capital component be excessively emphasized. Within social phenomena there are clearly many factors that cannot be economically isolated or expressed in terms of capital. Human behavior does not necessarily follow a logical path, such as the maximization of capital, so that accumulating social capital may not occur even though it might be logical to do so. For example, many people do or will not join associations or social networks, for a variety of reasons, and thereby restrict their bonding and bridging social capital.

Finally, social capital is inevitably treated as a positive asset and is portrayed in a positive light. Yet the acquisition of social capital may also occur amongst gangs and criminal groups or in societies which are fundamentally

divided such as Northern Ireland or the Balkans or in countries dominated by malevolent leadership such as Nazi Germany or the Khmer Rouge in Cambodia. Thus the potential for a dark side of social capital needs further consideration.

Cultural Capital

Cultural capital is a very general theory, somewhat elusive, and employed largely by sociologists to explain culturally generated differences in societies. As a theory, cultural capital recognizes that a multicausal approach is required to understand the complexity of a social phenomenon, thereby making it more valuable as a realistic explanation of society. The concept of cultural capital is perceived as a useful means for researchers to explore phenomena such as how social inequalities are organized in culture-drenched societies. It has been particularly influential in sociology to explain the ways the middle classes distinguish themselves from the working classes through their distinctive cultural tastes, knowledge, and competencies.

For example, cultural capital may be anything in one's personal/social background that contributes to or diminishes one during one's life. It could be something like family background or educational qualifications or wealth/income. These attributes help explain differences between people in a multicausal way. There is, in effect, no limit to what can or cannot be considered cultural capital as it really depends on the context being studied. However, as a theory it has a weakness in that it is frequently difficult to identify the relative influence of particular cultural factors and so explain social phenomena effectively.

Bourdieu (1986) referred to three types of capital in his seminal work – economic, social, and cultural capital. The last he saw as forms of knowledge, education, skills, or similar advantages that persons might have which gives them a higher status within society. As such, cultural capital helps explain why classes exist within societies and why they are essentially sustained. The capital element of the construct is applied because, as with other forms of capital, cultural capital may be accumulated, built, or diminished.

Types of Cultural Capital

Bourdieu (1986) identified three subtypes of cultural capital: embodied, objectified, and institutionalized:

Embodied. Cultural capital may be embodied within the individual, such as the inherited and acquired properties of one's self. Inherited cultural capital would be the cultural traditions from one's family and social class. This is acquired over time and is strongly linked to one's personal character and way of thinking.

Objectified. Cultural objects which are owned, such as artworks, may be transmitted symbolically as cultural capital as well as the more common transmission physically or by sale. Bourdieu claimed that while one can own objectified cultural capital through possession, those objects can only be utilized if people have the correct type of embodied cultural capital.

Institutionalized. Institutional recognition of cultural capital is mainly understood in relation to the labor market in the form of academic credentials or qualifications. Institutionalized cultural capital allows easier conversion to economic capital by assigning a monetary value for a certain institutional level of achievement.

Cultural Capital and Education

From an ideological perspective Bourdieu's thesis is that the role of education is mainly one of social reproduction that enables the dominant social class to reproduce its power, wealth, and privilege in a legitimate manner. Consequently, schools, as part of the political/ideological superstructure in capitalist society, help to perpetuate social and economic inequalities across the generations.

Using cultural capital Bourdieu demonstrated how the working classes were systematically blamed for their relative failure within the education system. He argued that access to resources and information as found in schools is a form of cultural capital restricting the working classes. If an equal opportunity to schooling was available to all, then success or failure must be a consequence of the individual rather than the way in which the system is structured to favor one class over another.

Further, Bourdieu argued that children are not simply socialized into societal values as a whole, but rather, they are socialized into the culture that corresponds to their class. This becomes their set of values, beliefs, norms, attitudes, experiences, and so forth that equip people for their life in society and represents their cultural capital.

More recently in terms of education, cultural capital may theoretically explain differential student achievement utilizing a wide range of influences. The past two decades have witnessed an increasing emphasis upon formal, systemic assessment of student achievement in many countries. The outcomes of these assessments reveal, unsurprisingly, that students from disadvantaged backgrounds perform at lower levels than other students. Cultural capital can provide a theoretical framework for explaining this phenomenon by addressing questions of power and ideology which authors contend are central to the differential achievement debate. For example, in the context of differential student achievement the explanation is traditionally linked to individual ability and some school effects rather than by cultural resources reflecting family influences and in the process the transmission of privilege is legitimized and not perceived as handicapping the disadvantaged child. Social capital

refers to connections among individuals that facilitate collective action and has received wide, though equivocal, support. These resources include trust, norms, and networks of association representing any group which gathers consistently for a common purpose. In her study of cultural capital on English school students Sullivan (2001) found strong evidence that cultural capital was transmitted from parents to their children which helps explain differential performance.

Issues in Cultural Capital

Multiple problems have been identified in both the theoretical understanding and practical application of cultural capital to explain social phenomena. Foremost there are conceptual difficulties concerning how cultural capital is to be identified and measured given that most measures are proxies. For example, which cultural activities constitute cultural capital? If reading books and newspapers as well as television viewing habits are included as Sullivan (2001) suggests, what of music habits, both listening and playing or participation in sport?

Similarly, technical difficulties abound concerning how cultural capital can be applied practically to address sociological problems. Will reading more books, for example, enhance cultural capital? Or is some measure of quality in books important in contributing to an individual's stock of cultural capital?

The impact of cultural capital is valuable as an explanatory concept but is more problematic when considered as a major theory of cultural reproduction. As Sullivan found, when disaggregating the effects of cultural capital and social class on student attainment, cultural capital provided only a partial explanation of social class differences in the educational achievement of students. As with social capital, cultural capital appears more useful at first blush and both were initially welcomed by social scientists with acclaim. But subsequent exploration, both theoretically and empirically, has found both concepts wanting as broad based explanation of social phenomena.

Bibliography

- Baron, S., Field, J., and Schuller, T. (2002). *Social Capital: Critical Perspectives*. Oxford: Oxford University Press.
- Bourdieu, P. (1986). The forms of capital. In Richardson, J. (ed.) *Handbook of Theory and Research for the Sociology of Education*, pp 241–258. New York: Greenwood Press.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology* **94**, 95–120.
- Croniger, R. and Lee, V. (2001). Social capital and dropping out of school. *Teachers College Record* **103**(4), 548–581.
- Dika, S. and Singh, K. (2002). Applications of social capital in educational literature: A critical synthesis. *Review of Educational Research* **72**, 31–60.
- Field, J. (2003). *Social Capital*. London: Routledge.

- Helliwell, J. and Putnam, R. (2007). Education and social capital. *Eastern Economic Journal* **33**(1), 1–19.
- Jennings, K. and Stoker, L. (2004). Social trust and civic engagement across time and generations. *Acta Politica* **39**, 342–379.
- Lin, N. (2001). Building a network theory of social capital. In Lin, N., Cook, K., and Burt, R. (eds.) *Social Capital Theory and Research*, pp 3–31. New York: Aldine De Gruyter.
- Newton, K. (1999). Social capital and democracy. *American Behavioral Scientist* **40**(5), 575–586.
- Portes, A. (1998). Social capital: Its origins and application to modern sociology. *Annual Review of Sociology* **24**, 1–24.
- Print, M. (2007). Citizenship education and youth participation in democracy. *British Journal of Educational Studies* **55**(3), 325–345.
- Print, M. and Coleman, D. (2003). Towards understanding social capital and citizenship education in divided societies. *Cambridge Journal of Education* **33**(1), 123–149.
- Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.
- Putnam, R. D., Leonardi, R., and Nanetti, R. Y. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.
- Sullivan, A. (2001). Cultural capital and educational attainment. *Sociology* **35**(4), 893–912.
- World Bank (1999). What is social capital? *Poverty Net*. <http://www.worldbank.org/poverty/scapital> (accessed August 2009).
- Woolcock, M. (2001). The place of social capital in understanding social and economic outcomes. *Isuma: Canadian Journal of Policy Research* **2**(1), 1–17.
- Halpern, D. (2005). *Social Capital*. Malden, MA: Polity Press.
- Hooghe, D. and Stolle, M. (eds.) (2003). *Generating Social Capital: Civil Society and Institutions in Comparative Perspective*. New York: Palgrave Macmillan.
- Jackman, R. W. and Miller, R. A. (1998). Social capital and politics. *Annual Review of Political Science* **1**, 47–73.
- Putnam, R. D. (1995). Tuning in, tuning out: The strange disappearance of social capital in America. *Political Science and Politics* **28**, 664–683.
- Pye, L. (2001). Civility, social capital and civil society. In Rotberg, R. (ed.) *Patterns of Social Capital*, pp 337–374. Cambridge: Cambridge University Press.
- Stolle, D. and Hooghe, M. (2004). The roots of social capital: Attitudinal and network mechanisms in the relation between youth and adult indicators of social capital. *Acta Politica* **39**, 422–441.
- Van Deth, J., Maraffi, M., Newton, K., and Whitely, P. (eds.) (1999). *Social Capital and European Democracy*. London: Routledge.
- Woolcock, M. (2000). *Using Social Capital: Getting the Social Relations Right in the Theory and Practice of Economic Development*. Princeton, NJ: Princeton University Press.

Further Reading

- Bourdieu, P. and Passeron, J. -C. (1990/1977). *Reproduction in Education, Society and Culture*. London: Sage.
- Dalton, R. (2004). *Democratic Challenges Democratic Choices*. New York: Oxford University Press.
- Fukuyama, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity*. New York: Free Press.

Relevant Websites

- <http://www.aers.ac.uk> – Applied Educational Research Scheme (AERS).
- <http://www.abs.gov.au> – Australian Bureau of Statistics.
- <http://www.cpn.org> – Civic Practices Network.
- <http://www.oecd.org> – Organization for Economic Co-operation and Development.
- <http://www.worldbank.org> – Poverty, The World Bank.
- <http://www.hks.harvard.edu> – Saguaro Seminar – Civic Engagement and Social Capital in America.
- <http://www.socialcapitalgateway.org> – Social Capital Gateway.
- <http://www.statistics.gov.uk> – Social Capital guide: Office for National Statistics.
- <http://www.bowlingalone.com> – The Saguaro Seminar: Civic Engagement in America.

The Aims of Education

R Marples, Roehampton University, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Anyone with any serious concern about the aims of education during the past 50 years will have been influenced by the work of Peters (1973), who, during his tenure as professor of philosophy of education at the University of London, had a profound influence in devising courses on which teachers were forced to come to grips with the question of whether or not education, or even schooling, had any overall aim or purpose. In an article concerned with the question “what is the aim of education?”, he suggested that an appropriate answer would be the same as the answer to the question “what is the aim of life?” namely, that it has not got one, while at the same time acknowledging that in the case of education, a longer story needs to be told if only to satisfy the taxpayer. Both here, and in his hugely influential *Ethics and Education*, Peters (1966: 28) argues that questions about the aims of education are “... a way of getting people to get clear about and focus attention on what is worth while achieving. It is not to ask for the production of ends extrinsic to education which might explain their activities as educators.” Education, he insists, is logically connected to what is thought to be worthwhile; its value is intrinsic, and not contingent upon something else to which education is a means.

Such a claim certainly has a *prima facie* plausibility in virtue of the fact that instrumental reasoning is forced to end somewhere; some things just are, as it were, of ultimate significance. Whatever it is to be educated, it is not necessarily a state which derives its value from any instrumental purposes it might serve. Nevertheless, it would be invalid to conclude from this warrantable premise that education serves no instrumental purposes whatsoever. There are countless reasons why society might want to invest in education. Probably more than any other institutions, schools exert a profound influence in shaping people's belief systems and characters, which in turn have an incalculable impact on the kind of society in which we conduct our lives. It is therefore incumbent upon those with responsibility for curriculum construction and delivery to try and ensure that pupils are equipped with the wherewithal to make the sort of contribution to society that a liberal democracy has a right to expect.

If asked to specify the aims and purposes of schooling, many a hard-nosed Victorian would have had few qualms about the desirability of two different kinds of schooling, with radically different aims. Working-class children should learn to read, write, and count, with the patronizing

aim of teaching them to appreciate their position, while their middle-class counterparts required little more than an unrelieved diet of Latin and Greek, with the very different aim of preparing them for positions of leadership throughout the Empire. It is now generally accepted that something more is required from a system of publicly funded education, and it is the concern of what follows to spell out what this might entail. We shall, of necessity, be forced to restrict the discussion to three possible candidates. These are: knowledge for knowledge's sake, preparation for work, and education for well-being.

Knowledge for Knowledge's Sake

According to Peters, education is the process of initiation into intrinsically worthwhile activities, or those activities which one might engage in for their own sake. These include things such as history, science, mathematics, religion, and esthetic awareness. Those who are successfully initiated into such forms of thought, and in some sense make it their own, find, according to Peters, “the contours of the public world... transformed. The process of initiation into such modes of thought and awareness is” he says, “the process of education” (Peters, 1966: 51). He went on to develop this idea in collaboration with Hirst (1965, 1973), who had earlier argued that the principal aim of education should be the development of mind for its own sake, through a liberal education conceived as an initiation into such forms of knowledge and understanding.

As fine-sounding as this is, there are enormous problems associated with determining what is to count as intrinsically worthwhile. What some might take to be a valuable activity appears boring or pointless to others. In *Ethics and Education*, Peters goes to considerable lengths to articulate the notion of worthwhile activity, relying on a transcendental argument reminiscent of Hirst's attempt to justify the importance of knowledge and understanding as the bedrock of a liberal education. According to Hirst, “to ask for the justification of any form of activity is significant only if one is committed already to seeking rational knowledge. To ask for a justification of the pursuit of knowledge therefore presupposes some commitment to what one is seeking to justify” (Hirst, 1965: 126). In an educational context, transcendental arguments are notoriously problematic. But were we able to demonstrate the worth of particular activities or forms of understanding by this or any other means, it remains to be seen whether or

not rational curriculum planning should include such activities at all, or whether there is more to the notion of a liberal education than the pursuit of knowledge. Propositional knowledge (knowing that) is, after all, only one way of knowing. It is also possible to know how to do things – communicate in Mandarin, drive a car, use PowerPoint, re-wire a house, and countless other things. It might well be argued that schools have for too long concentrated on knowledge for knowledge's sake at the expense of equipping children with the skills they are likely to require in adult life. In reality, however, the problem is one of emphasis only, because most skills are impossible to master without knowing that certain things are true or false, and even if the teaching of skills of various kinds could be justified as part of a liberal education, it is inconceivable that one could live a fulfilling life in a society such as ours, without a great deal of factual knowledge and understanding. According to Bailey (1984: 29), a liberal education is characterized by its capacity to liberate a person from the here and now, by involving pupils in what is fundamental and general, and this, he argues, requires an education firmly within the framework as articulated by Peters and Hirst.

As indisputable as the value of knowledge undoubtedly is, the fact remains that until we are clear as to what there is about such knowledge that makes it indispensable to a person's long-term interests, we should not just assume that it is the only thing of any value in a child's education, or indeed that it should be regarded as having supreme importance. Again, if we are to take seriously the aims underpinning the current English National Curriculum, which include the transmission of enduring values, with the intention of promoting pupils' integrity and autonomy, helping them become responsible and caring citizens committed to sustainable development, promoting their self-esteem, emotional well-being, their ability to form satisfying relationships, to relate to others for the common good, and to be able to make informed choices throughout their lives (DfEE/QCA, 1999), it is unclear how a knowledge or subject-based curriculum is expected to realize such aims.

A number of philosophers, in recent years, have not only expressed reservations about Hirst's analysis of knowledge (subsuming it as he does, under a number of different forms such as mathematics, physical sciences, knowledge of persons, literature and the fine arts, morals, religion and philosophy, with each form distinguishable from the other by reference to its own unique set of concepts, logical structure, and tests for truth), but have questioned the desirability of restricting the idea of a liberal education to the pursuit of truth, on the grounds that such an obsession with the intellectual ignores other equally important aspects of personhood, such as the emotional and the practical, with which education must have an equally legitimate concern (see, e.g., Martin, 1981).

People do not exist in ivory towers, pursuing arguments wherever they may lead in the manner of the proverbial professor. Ignoring such a truism would have dire consequences, especially for those students who fail the initiation test and are left floundering at the gate with other barbarians. There may well be a place for the pursuit of knowledge for its own sake within a school setting, but unless those who advocate its having priority over other possible aims are to avoid the charge of elitism, they will need to demonstrate to both individual learners and the society of which they are members, the benefits that knowledge is likely to bestow.

We must now address other claims as to what the aims of education might be. After addressing the competing merits of vocational versus liberal education, we shall return to the question of what constitutes an appropriate education for a whole person.

Education and Preparation for Work

Discussions relating to the aims of education are all too frequently plagued by the assumption that criteria of relevance should be determined by reference to the extent to which school leavers are prepared for the world of work, or the real world. The whole point of schooling, and the investment in education in general, is often seen as a means to the obtaining of qualifications leading to paid employment. Knowledge, (or book-learning, as it is often disparagingly called) is all very well, but can the billions we spend on education be restricted to its pursuit?

Some of the advocates of what is generally known as vocational education see it as having relevance for all pupils, especially those in the latter stages of secondary education, while others would like to see some kind of educational divide at around the age of 14, whereby the academic sheep are separated from the vocational goats, on the grounds that some children are unmotivated in the quest for prestigious academic qualifications and need something else to show for all their years in compulsory schooling. While it is no doubt true that many pupils would relish the idea of opting out of academic courses, and would no doubt perceive the vocational alternative to have greater relevance, such a solution to the problem of the disenchanting underachiever or disgruntled student is both facile and fraught with dangers.

First, there is the problem of identifying the vocational goats. The fact that a pupil hates her history or science lessons does nothing to undermine the claim that she has needs and interests in common with the rest of us to which, as a person, she has a right, and for which those charged with her education have an obligation to provide. Second, there are complex issues relating to the kind of work for which she might be expected to be trained. Many jobs are not only boring, they are useless in the sense that

they require people to produce things for which nobody has any real need, as well as have adverse consequences for the environment or for opportunities to exercise individual judgment or imagination or the possibility of enjoying a fulfilling family life and a host of other activities for which someone may have talents or enthusiasms. A great many people have very little choice about working for others who are privileged and wealthy, and for very low wages. The injustices associated with writing off people at an absurdly young age, before they are fully aware of the range of significant alternatives before them, and while their critical powers of evaluation are insufficiently developed, are obvious. Unless we conceive of vocational education in more liberal terms than that associated with vocational training, teachers should refuse to conspire in the process of selling children short by refusing to alert them to the nonegalitarian, oppressive, and all too frequently unrewarding features associated with so many jobs in a modern industrial society.

The problems associated with how to identify those children who might be thought most suitable for vocational training notwithstanding, it has long been thought to be the case that meaningful work is indispensable to a flourishing life and that schools are under a moral obligation to prepare people for it in some, as yet, unspecified way. However, as White (1997) has persuasively argued, there are reasons for doubting this. In a wide-ranging discussion on the role of education in relation to work, he makes a number of important distinctions, including that between: (1) autonomous work (where it is one of one's major goals to produce something) and heteronomous work (where it is not), (2) work in which one willingly engages, from that done unwillingly, and (3) paid work and voluntary work. He goes on to question the assumption that work (of any kind – meaningful, autonomous, or whatever) must, of necessity, find a place in a fulfilling life, given the plethora of alternative possibilities in which one might enthusiastically engage, such as gardening, spending time with friends, cooking, playing the piano, or just thinking. Given the all-absorbing nature of such pursuits, it is unclear why one needs to find time for work. Persuasive or not, the fact remains that an unacceptably large number of young people end up working in tediously repetitive jobs from which they would almost certainly quit upon winning the lottery.

Teachers might well respond in a number of ways. They could ignore the facts, ostrich-like, and concentrate on trying to initiate pupils into the forms of knowledge, or they could face up to the fact that work is so much a part of most people's lives after leaving school that they had better give serious thought to what counts as preparation for it, even if they find themselves in the uncomfortable position of having to subscribe to the values to which so many employers are attached, such as deference and obedience; values which they themselves may abhor.

Alternatively, they could acknowledge the truth in something to which Dewey (1916) drew attention almost a century ago, and that is the absurdity of the hard-line distinction between the theoretical and the practical.

Pring (1995) is one philosopher among many who is highly critical of vocational training, on the grounds that its prime concern is with mere competence, as opposed to the liberal ideal of the educated person having something to do with what it is to be fully human. While there is nothing wrong in preparing people for the world of work, such preparation may be conceived as genuinely educational only if it helps people to make sense of the social and economic context in which they find themselves. Pring is at pains to demonstrate that vocational education can not only be liberating, but can also provide opportunities for people to gain access to all kinds of knowledge and ways of seeing the world. Through work, such as building or motor mechanics, one may come to appreciate something from an esthetic, scientific, or even moral point of view, which, if true, serves to dispel the false dichotomy between liberal and vocational education.

In trying to specify the aims of education, we have had reason to be skeptical about two widely held views; that education, or schooling, should aim at nothing more than producing knowledgeable people, and that criteria of relevance can be seen in purely vocational terms. As we have seen, the authors of the English National Curriculum believed that schools had other functions than that of merely producing knowledgeable or employable people. They were only too well aware that there was more to human well-being than this, and that education should make significant inroads in trying to help young people flourish as people. In what follows, we shall explore what this might entail, by reference to which notions such as self-esteem, the ability to make rational choices between competing alternatives, and concern for others, gain application.

Education for Well-Being

The idea of what it is to flourish as a person has occupied the attention of philosophers since Plato and, as might be expected, there is considerable disagreement as to what it amounts to. There are at least three aspects under which it might be conceived; as happiness, desire-satisfaction, and having the capacity and opportunity to live autonomously as a moral being. This latter notion is both complex and troublesome, but it is hoped that a proper understanding of what it involves will provide a morally acceptable basis for those concerned with preparing young people for life in the twenty-first century.

Utilitarianism is the theory according to which well-being is seen in terms of happiness. Unfortunately, happiness is a notoriously subjective phenomenon. Nothing

better illustrates this than a thought experiment postulated by Nozick (1974), who asks us to imagine what it would be like to be connected to a machine capable of providing us with any experience whatever. While plugged in, one could believe that one was loved or famous or popular; one could satisfy more or less any fantasy that one had and remain blissfully happy throughout. But how can it possibly be the case, Nozick asks, that one can flourish if, in actual fact, one is friendless, ignorant, despised, exploited or in any other state manifestly incompatible with personal well-being. The principal objection to any such mental-state account of well-being is that it fails to distinguish between believing that one is flourishing and actually flourishing; in other words, there is more to flourishing than simply feeling happy. Moreover, as an aim of education, the pursuit of happiness is a nonstarter. It is precisely because of its subjective status that it is impossible for teachers to gear what they do toward its promotion. In so far as it is attainable, it is the result of a subjective judgment that one's life is, for the most part, going well.

A variety of desire-satisfaction accounts of well-being have been constructed in order to meet the difficulties associated with the simplistic reductionism associated with mental-state accounts. One such theory construes well-being in terms of getting whatever it is that one happens to want, the initial plausibility of which rests firmly within liberal democratic theory whereby we should remain unconstrained in the pursuit of our own version of the good life, however idiosyncratic that might be, with the proviso that in so doing, we avoid preventing others from pursuing theirs. Such a view is problematic in several respects. Even if it were possible to structure one's numerous and all too frequently conflicting desires into some kind of order of priority, the theory has counterintuitive implications. It would be compatible with such a view that my well-being could be promoted without my knowing it if, say, as a result of my charitable work other people, the fate of whom was unknown to me, thrived in ways not possible without my contributions. Having satisfied my desire to improve the quality of other people's lives albeit unknowingly, it is questionable whether such beneficence on my part could be said to have any positive impact whatsoever on my own well-being. Again, having successfully obtained some of the things I happened to want, I might well end up feeling indifferent or ashamed, while other (satisfied) desires may be based on such false beliefs that my actions have no rational foundation, providing a sufficient reason for doubting that my welfare might possibly be enhanced in trying to satisfy it in the first place.

In response to such objections, a number of philosophers have attempted to construe well-being in terms of the satisfaction of those desires about which one could be said to be fully informed, in the sense that one is cognizant of both her origins as well as the consequences

of satisfying them. Unfortunately, there is no guarantee that such a ploy will work. However informed one might be about a particular desire, satisfying it might well be incompatible with one's flourishing. Knowing all there is to know about the lethal consequences of smoking, one may well continue to satisfy the desire to smoke. Those wishing to defend the view that well-being can be cashed out in terms of the satisfaction of informed desires are wont to restrict the notion to only those desires a person would have, were she to fully appreciate what is at stake, while all too frequently failing to provide an account of what it is to appreciate such information that is not circular (see Thomson, 1987).

If this were not enough to undermine desire-satisfaction accounts of well-being, we have to acknowledge that desires do not appear from nowhere; for the most part, they arise as a result of the aspect under which their objects are seen or evaluated. Our preference for *x* as opposed to *y* is, as often as not, a direct result of acknowledging the desirability of *y*, which itself depends on something other than merely wanting it. On the assumption that not just anything merits the epithet desirable, teachers have a major role in helping young people appreciate the significance of what Taylor (1985) calls the language of contrastive evaluation, in accordance with which values, in areas of life as diverse as the political, the esthetic, and the moral, acquire their import. We attribute qualities of character to someone by reference to moral idioms such as courage, integrity, loyalty, and kindness on the one hand, and malicious, depraved, pusillanimous, and cruel on the other, and it is foolish to pretend that we could employ them in any way we choose. Rational choice and decision making, as well as the evaluations upon which they are based, must rely on more objective foundations than the extent to which we happen to want something. There is a fundamental disagreement between philosophers on this matter however. Those who are inclined to see values as essentially subjective, tend to argue that the order of explanation is from desire to value, while those who believe that values are, in some sense of that term, objective, are inclined to believe that desire, as an indispensable element in attributing value to something, is unnecessary. The issues are complex, however, and cannot be pursued here. (For a full discussion of the issues involved, see Griffin, 1986.)

The history of philosophy is replete with accounts of human flourishing in terms of a single *summum bonum*, and it is all too easy to assume that the only alternative to desire-satisfaction accounts of well-being is some form of perfectionism and the associated specter of authoritarianism. Nevertheless, there are limits to what can possibly amount to a flourishing life; limits, for the most part, determined by a life that is recognizably human, with its legitimate concern for harm-avoidance. A person may be said to have been harmed if her needs remain unmet, but not necessarily when her desires are merely frustrated.

What a person needs in order to flourish will undoubtedly vary from one social context to another, but as Raz (1986) has argued, in a society such as our own, whereby enormous value is attached to being able to make choices in significant areas of one's life such as with whom one might live, whether or not to subscribe to a religious faith, choice of career and such like, it is inconceivable that one could do so without a measure of autonomy in such matters. Teachers thus have a particularly important role to play in promoting the autonomy, both intellectual and emotional, of their pupils.

A person may be said to be autonomous to the extent that she is in a position to appreciate the degree to which her beliefs and actions have been shaped by the forces of socialization, as well as having the courage and strength of will to reject them if necessary. She must also be able to provide some coherence to her choices over time, in accordance with which her life may be distinguished from the episodic and infantile. An autonomous person must also have the ability and opportunity to make choices from a significant array of options. Anyone who is deficient in these respects is more likely to fall victim to the will of others and her life will be largely inauthentic.

Important as personal autonomy undoubtedly is within contemporary accounts of well-being, it would be wrong to assume that people could flourish as autonomous egoists. As Benn (1988: 259) says: "We are species-beings in the sense that our enterprises are necessarily embedded in some continuing tradition, in terms of which we discover what is worth doing. . .". Our interests are rarely entirely self-interested; there are countless respects in which they might be said to be bound up with those of others, including those to whom we are intimately related as well as to things as distant as the ecological balance of the rain forests. As Lear (1984) has demonstrated, a significant form of human flourishing partly consists in promoting flourishing generally, whereby morality provides its own motivation and the distinction between altruism and egoism is less-sharply defined than might otherwise appear.

The implications for education are legion, with teachers confronted with the daunting task of trying to promote sensitive perception within children as well as getting them to care about the needs of others as well as themselves. In this respect, moral education is more complex than that proposed/promoted by those wedded to the teaching of knowledge and understanding for its own sake, whereby a morally educated person is little more than someone able to understand the meaning of moral concepts or, at best, as someone trying to regulate her actions by reference to a set of principles. As Anthony O'Hear (1981: 129–130) says: "Moral behaviour is not so much a matter of an isolated rational agent acting on freely chosen principles, as a development of one's sense of a shared humanity." While the promotion of autonomous

agency is undoubtedly one of the principal aims of education, it cannot suffice in the absence of a moral and social education appropriate to the fostering of the skills and dispositions associated with participatory democracy (see, e.g., Gutmann, 1987). All this, of course, presupposes a considerable amount of knowledge and understanding, but the purpose of instilling this into children derives its legitimacy from the perspective of the altogether deeper underlying values associated with human flourishing, with the implications for rational curriculum planning being recognized accordingly.

Bibliography

- Bailey, C. (1984). *Beyond the Present and Particular: A Theory of Liberal Education*. London: Routledge and Kegan Paul.
- Benn, S. I. (1988). *A Theory of Freedom*. Cambridge: Cambridge University Press.
- Dewey, J. (1916). *Democracy and Education*. New York: Free Press.
- DfEE/QCA (1999). *The National Curriculum Handbook for Primary/Secondary Teachers in England*. London: DfEE.
- Griffin, J. (1986). *Well-Being: Its Meaning, Measurement and Moral Importance*. Oxford: Clarendon Press.
- Gutmann, A. (1987). *Democratic Education*. Princeton, NJ: Princeton University Press.
- Hirst, P. H. (1965). Liberal education and the nature of knowledge. In Archambault, R. D. (ed.) *Philosophical Analysis and Education*, pp 113–138. London: Routledge and Kegan Paul.
- Lear, J. (1984). Moral objectivity. In Brown, S. C. (ed.) *Objectivity and Cultural Divergence*, pp 135–170. Cambridge: Cambridge University Press.
- Martin, J. R. (1981). The ideal of the educated person. *Educational Theory* 31(2), 97–109.
- Nozick, R. (1974). *Anarchy, State and Utopia*. Oxford: Basil Blackwell.
- O'Hear, A. (1981). *Education, Society and Human Nature*. London: Routledge and Kegan Paul.
- Peters, R. S. (1966). *Ethics and Education*. London: Allen and Unwin.
- Peters, R. S. (1973). The aims of education. *London Educational Review*.
- Pring, R. (1995). *Closing the Gap: Liberal Education and Vocational Preparation*. London: Harper and Row.
- Raz, J. (1986). *The Morality of Freedom*. Oxford: Oxford University Press.
- Taylor, C. (1985). What is human agency? In *Human Agency and Language*, pp 15–44. Cambridge: Cambridge University Press.
- Thomson, G. (1987). *Needs*. London: Routledge and Kegan Paul.
- White, J. (1997). *Education and the End of Work: A New Philosophy of Work and Learning*. London: Cassell.

Further Reading

- Brighouse, H. (2006). *On Education*. London: Routledge.
- Hirst, P. H. and Peters, R. S. (1973). *The Logic of Education*. London: Routledge and Kegan Paul.
- Marples, R. (ed.) (1999). *The Aims of Education*. London: Routledge.
- Pring, R. (2004). *Philosophy of Education: Aims, Theory, Commonsense and Research*. London: Continuum.
- Standish, P. (2006). The nature and purpose of education. In Curren, R. (ed.) *A Companion to the Philosophy of Education*, pp 221–245. Oxford: Blackwell.
- White, J. (2004). *Rethinking the School Curriculum*. London: RoutledgeFarmer.
- White, J. (2006). *What Schools Are for and Why*. London: Philosophy of Education Society of Great Britain.
- Wringe, C. (1988). *Understanding Educational Aims*. London: Allan and Unwin.

The Analytical Tradition

D Carr, University of Edinburgh, Edinburgh, UK

© 2010 Elsevier Ltd. All rights reserved.

The Analytical Tradition in Philosophy

From an analytical perspective, although we may attach some sense to the idea that there is a separate and distinct analytical approach to philosophy, there is equally – as we shall shortly indicate – much confusion in this notion. To be sure, going back to the person who might well be considered the founding father of analytical philosophy – the Athenian philosopher Socrates – it would appear that Socratic philosophizing was of a rather different order from that which characterized what has come to be called (significantly enough) pre-Socratic philosophy. By and large, the various pre-Socratic philosophers – Anaxagoras, Anaximander, Thales, Heraclitus, Parmenides, Zeno (as well as such sophists as Gorgias and Protagoras with whom Socrates so often took issue) – were concerned with the construction of grandiose speculations on the nature of the cosmos or human society that might nowadays appear rather closer to the theories of modern natural or social scientists. Of course, Socrates' own great pupil, Plato, was not himself averse to ambitious cosmological speculations and political theorizing; but even Plato generally follows the rather different view of his mentor, Socrates, of the fundamental role of philosophical reflection in the construction of such theories. For while Socrates agrees with previous thinkers that it is a key aim of any human enquiry to get to the truth of things, he also thinks that it is crucial to the achievement of such truth that the terms in which such enquiry is conducted are rationally clear and coherent. For Socrates, the chief impediments to truth are conceptual confusion and invalid argument and for him it is the major task of the philosopher to clear such impediments from the path of knowledge.

The key philosophical task, as Socrates sees it (and as most subsequent analytical philosophers have come to view it), is the clarification of concepts – or, as it might also be nowadays said – the precise definition of terms. Before we can begin to talk or speculate clearly about anything, we must make sure that we are using the basic terms of debate or discussion in a clear and unambiguous way. Indeed, Socrates is perhaps the first great philosopher to realize (though he clearly learned much in this respect from the Eleatics) the intimate connection between such conceptual clarity and valid inference. Socratic elenchus, for example, seems to be largely directed to exposure of the general fallacy that later philosophical logicians have come to call *reductio ad absurdum*. In Plato's *Republic*

(Hamilton and Cairns, 1961), for example, Socrates adopts his usual strategy of asking interlocutors to define a given philosophically problematic term – in this case, justice – and a respondent offers a particular definition, *p*, of justice (in this case, telling the truth and paying one's debts). Briefly, Socrates then proceeds to describe a circumstance, *q*, in which doing what is indicated in *p* would not (by common agreement) be just. Since *q* therefore entails not *p* and not *p* is inconsistent with *p*, *p* cannot be a conceptually coherent concept of justice, and in so far as arguments or inferences based upon such a concept of justice are liable to inconsistency, they must therefore be invalid.

At all events, such close Socratic attention to the logical form of argument and to clarification of the terms in which such argument is conducted might well be considered the backbone of the analytical approach. From this viewpoint (and for present purposes), one might identify three key moments in the subsequent post-Socratic development of this approach. The first of these would seem to be the earliest systematic attempt by Plato's pupil, Aristotle, to develop a formal theory of inference, based on certain fundamental laws of thought (of identity, non-contradiction, and excluded middle), according to which any and all valid reasoning would need to proceed. In the works collectively known as the *Organon* (see McKeon, 1941), containing his theory of the syllogism (the basic form of deduction), Aristotle set out to identify the rules governing valid inference from separate premises *A* and *B* to some further distinct conclusion *C*, such as would thereby block inference of true conclusions from false premises. Aristotle's theory of syllogism (despite some ancient rivals and/or variants) was widely regarded as the last word on logical theory until virtually the late nineteenth century. However, in what might for present purposes be considered as the second key moment of analytical method, the great German logician and mathematician, Gottlob Frege, devised revolutionary techniques for the logical analysis of complex statements of mathematics, science, and natural language – the logic of quantification of so-called predicate calculus – that virtually replaced Aristotelian syllogistic (see Frege, 1978; Geach and Black, 1960). From this viewpoint, Frege – particularly along with the British philosopher and logician, Bertrand Russell (Russell, 1956), who showed how Fregean quantification theory could be used to illuminate hitherto intractable metaphysical and ontological problems – may be considered the main founder of modern philosophical analysis.

However, among the important twentieth-century philosophers influenced by Frege – including the Anglo-American and Germanic logical positivists – was the Austrian philosopher, Ludwig Wittgenstein. Wittgenstein's early *Tractatus Logico-Philosophicus* (Wittgenstein, 1961) was deeply influenced by both Frege and logical positivism, but his later *Philosophical Investigations* (Wittgenstein, 1953), seeks to develop a key idea of Frege in a way that effectively demolishes positivism. For generally, the logical positivists inherited from such earlier British empiricists as David Hume, an idea – also, as we have seen, to be found in Socrates and Plato – that the philosophical search for meaning is basically a search for definitions: in short, that one needed to fix the unique and invariant sense of a given term, before one could meaningfully use it. In the introduction to his important work, *The Foundations of Arithmetic*, however, Frege had said that the meaning of a term could only be grasped “in the context of a proposition” (Frege 1978, Introduction p. xe). In a nutshell, Wittgenstein's revolutionary application of this point in his later philosophy consists in showing that one cannot generally understand a (philosophically problematic or other) term by deciding its meaning in advance of usage – precisely because one needs to know how a term is used in order to understand its meaning: hence, the slogan, usually associated with Wittgenstein's philosophy – “Do not look for the meaning, look for the use.” Wittgenstein perceived, as previous ancient and modern philosophers had not (though this insight is arguably foreshadowed in Aristotle), that philosophically intractable terms seldom have a single univocal use. From this viewpoint, it seems philosophically mistaken to look (as previous philosophers had) for a single unitary definition of good given the rather different uses of the term in the utterances, “That's a good thermometer” and “Good morning!”

During his long association with Cambridge, first as a student and later as professor of philosophy, Wittgenstein had a profound influence on the development of Anglophone philosophy, although the extent to which he directly influenced such leading Oxford philosophers of the time as Gilbert Ryle and J. L. Austin (or whether these other philosophers were independently moving in the same broad direction as Wittgenstein) is a moot point. But generally (and despite differences), all three of these philosophers came to be regarded as leading lights of the particular philosophical movement known (by way of some contrast with the logical analysis of Russell, logical positivists, and many American pragmatists) as ordinary-language philosophy. Broadly, ordinary-language philosophers shared with more formal approaches the basic analytical view that the key philosophical task of conceptual analysis is primarily that of understanding the logical workings of the forms of discourse through which human beings have tried to understand their world. However, whereas the latter were inclined to regard forms of human discourse that could

not be precisely regimented (broadly in terms of Frege's predicate calculus) for scientific purposes as unworthy of serious philosophical attention, ordinary-language philosophers were wont to consider such looser forms of discourse (e.g., the languages of ethics, esthetics, and religion) of considerable philosophical interest. Indeed, it is noteworthy that the later Wittgenstein tended to dispense with the term logic (with all its connotations of formal precision) in his characterization of philosophical work, in favor of the term grammar. The job of the philosopher is precisely to investigate the particular grammatical complexities of large areas of usage that – despite being resistant to precise scientific regimentation – may, for all that, express matters of the very highest human value and concern.

Analytical Philosophy of Education

There can also be little doubt that by the middle of the twentieth century, a broadly ordinary-language approach to conceptual analysis had yielded impressive philosophical results in such traditionally problematic (and often less systematically codifiable) fields of philosophy as ethics, esthetics, and philosophical theology, and – in the wake of widespread postwar political, social, and educational trends and reforms in Britain, America, and elsewhere – it was only a matter of time before the techniques of philosophical analysis came to be applied to the discourse of education. The philosopher most readily credited with developing analytical philosophy in Britain was Richard Stanley Peters who already had an established reputation in mainstream philosophy (for extensive work in philosophical psychology, ethics, and political philosophy) prior to his appointment as Professor of Philosophy of Education at the London Institute of Education in 1962. At the Institute, however, Peters quickly gathered around him a group of younger philosophically minded educational theorists – including, notably, Paul Hirst, Robert Dearden, and John White – with a view to the fairly radical overhaul of educational philosophy in particular, and of educational theorizing in general, in Britain and elsewhere. Indeed, Peters' pioneering analytical approach to educational reflection – strongly influenced by the largely ordinary-language drift of Oxbridge philosophy in which he had himself been schooled – had a much wider influence in the English speaking world, not least in such British Commonwealth countries as Canada, Australia, and South Africa. However, this Oxbridge approach was less favored in USA, where Peters' American counterpart, Israel Scheffler, drew more on a pragmatist tradition inspired by the late-nineteenth- and early-twentieth-century work of such home-grown philosophers as Charles Sanders Peirce, William James, and John Dewey. Still, although the influence of Wittgenstein (despite some important American disciples) and other Oxbridge philosophers was not so extensive in

USA, and much mainstream pragmatism inclined to the formal extremes of analytical philosophy, there was considerable agreement between Peters and Scheffler and their respective followers on what was broadly appropriate for postwar educational philosophy in terms of analytical aims and methods.

In this regard, one of the past academic trends that the analytical approach of Peters in Britain sought to overturn was the doctrines of the great educators approach (as Peters called it) to professional educational theorizing in general, and philosophy of education in particular, characteristic of professional teacher education in the British colleges of education. This approach largely consisted of a mixed and uncritical diet (taught mainly by ex-practitioners with occasional professional masters' degrees) of educationally relevant history, philosophy, and social science, in which Plato, Rousseau, and Dewey rubbed rather promiscuous theoretical shoulders with Pavlov, Piaget, and Basil Bernstein. To whatever extent such perspectives were taken seriously by professional teacher trainers as theoretically relevant to past or present educational policy, they were almost certainly regarded as quite irrelevant to actual classroom practice by the vast majority of teacher trainees. However, although Peters certainly regarded Plato, Rousseau, Dewey, and other such past great educators as educationally relevant – writing insightfully on many of them (Peters, 1981) – he envisaged a quite different analytical role for educational philosophy in the professional armory of teachers. In short, for the new analytical philosophers, the key role of philosophy was not to learn the views of past philosophers, but to acquire the skills of critical analysis of past or present theory, policy, and practice in the interests of greater conceptual clarification of the professional task. The early leaders of this analytic movement also set the highest possible standards of such work in a wide range of pioneering studies. In Britain, in addition to Peters' (1973) own pioneering conceptual and normative work on the aims of education – among many other topics – his colleagues Paul Hirst, R. F. Dearden, and John White produced seminal work on (more or less, respectively) knowledge and the curriculum, primary education, and on issues regarding the psychology of learning (e.g., intelligence and indoctrination) and, in USA and elsewhere, there was equally important work by Scheffler and others on reason, teaching, and a host of other educationally central issues and topics.

Once again, the key point in all of these was not to learn about philosophies of education but to develop a set or repertoire of professionally relevant skills that would enable reflective teachers and policymakers to come to a clearer and more coherent view of the point and purpose of their practice. Indeed, one important brainchild of the London Institute revolution in philosophy of education was the development of a whole new discipline-based

conception of professional teacher education for the new post-Robbins Bachelor of Education degrees in Britain. On this view, first sketched by Peters (1966) in *Ethics and Education* and more fully developed by Paul Hirst in various papers (e.g., Hirst, 1983), the new professional degree ought to be grounded in the systematic specialist-taught study of such key educationally relevant academic disciplines of history, psychology, sociology, and (of course) philosophy of education. However, the point above all was that philosophy of education should be practically useful to educational practitioners in clarifying both their rational understanding and intelligent conduct of education and teaching. Moreover, the reason why the new analytical philosophers considered some such facility for philosophizing to be indispensable to teachers was precisely because education and teaching would seem, on the face of it, to be problematically unclear – or at least conceptually and normatively complex – enterprises. Thus, while some of the expertise needed by teachers (such as the craft skills of classroom practice) might be of a more straightforward technical kind, the kinds of professional decisions in which they were likely to be daily implicated would also be such as to require quite ethically complex reflection and judgment on a variety of normative and evaluative issues. Although philosophical analysis could not, of course, relieve practitioners of the burdens of professional judgment, it was held that such burdens might be considerably eased by the clarity of thought and argument that analytical techniques of philosophical analysis could afford.

The three key tasks of analytical philosophy of education were also arguably well exemplified by works published by well-known mainstream philosophers within a few years of each other. The first and most fundamental of these tasks – that of straight conceptual analysis of basic educational concepts, of distinguishing between different senses of terms, and teasing out the logical or conceptual implications of a given expression – was well represented by a work of the distinguished Australian philosopher, John Passmore (1980), published under the title *The Philosophy of Teaching*. Indeed, this work might still be considered an evergreen philosophical treatment of this important educational topic. However, the second time-honored task of analytical philosophy – on which latter-day educational philosophers have long been (perhaps from Socrates on) profitably engaged – is that of the critique of received theories of education and learning. This task had been an early concern of Peters (1958) as a pure philosopher in his trenchant demolition of psychological theory in an influential work entitled *The Concept of Motivation*, but related points against psychological theory were given explicit educational application by Peters' erstwhile Birkbeck colleague, David Hamlyn (1978), in his penetrating critique of Piaget and others in *Experience and the Growth of Understanding*. However, the third central task of analytical philosophers in general, and of educational philosophers in particular, might be characterized as the

clarification of normative perspectives on education and the exploration and/or critique of the (practical) reasons for holding this rather than that position. Just as it has always been considered part of the mainstream moral philosopher's job to evaluate arguments for and against capital punishment, euthanasia, or abortion, so one might expect educational philosophers to be interested in the reasons offered for educational selection, separate schooling, or the use of corporal or other punishment in schools. From this viewpoint, the British philosopher, David Cooper's (1980) defense of selective education in his publication, *Illusions of Equality*, could be considered (whether or not one disagreed with it) a philosophically clear example of this sort of work.

Criticisms of Conceptual Analysis and of Analytical Philosophy of Education

Still, there can be little doubt that the early 1980s saw the beginnings of a shift away from analytical philosophy of education in the style of Peters, Passmore, Hamlyn, and Cooper. While it is not easy to identify and/or account for all sources of this shift, it may suffice for the moment to highlight two key dimensions of it. The first involved widespread growing interest in philosophers and theorists who had not been traditionally associated with mainstream analytical traditions – and who, indeed, it had been common for early- and mid-twentieth-century Oxbridge analytical philosophers to discount as any sort of philosophers at all. To be sure, such (diverse) thinkers as Hegel, Kierkegaard, Marx, Nietzsche, Heidegger, Sartre, Levi-Strauss, Lacan (and later), Derrida, Foucault, Lyotard, and Levinas were linked with such largely continental European currents of thought as Marxism, phenomenology, existentialism, structuralism, post-structuralism, and postmodernism that had ever been prone to unfavorable contrast with Anglo-American conceptual analysis. Indeed, it was partly such contrasts that were to lend weight to the widespread notion that there are different traditions of, or approaches to, philosophy. At the same time, however, it was becoming fashionable in some educational quarters to hold that there was something profoundly mistaken about conceptual analysis as a philosophical method an idea that gained ground not only under the influence of post-structuralism and postmodernism but also with increased educational attention to latter-day historicist trends in the philosophy of science and (from the early 1980s onward) to more neo-idealist or post-analytical social ethics.

Briefly, the dominant idea in this connection was the view – owing most perhaps to nineteenth-century German idealism and its Marxist and pragmatist progeny – that since human knowledge and truth are socially constructed in the light of rather diverse local circumstances, the absolute and/or universal meaning and truth which bygone philosophers had sought could be only illusion. There can also be no

doubt that this idea has in various recent guises had a fairly intoxicating effect on contemporary philosophers. It is no less clear that, new garments notwithstanding, it is almost as old as the hills. Indeed, a highly subjective form of this essentially relativist view – according to which all truth is perspectival – was both defended by the ancient Greek sophist Protagoras and (according to Plato in his *Theaetetus*) opposed by Socrates. Moreover, of the numerous modern educationally influential forms and versions of this idea, one might mention three. The first is that of the French post-structuralist philosopher, Michel Foucault, who characterizes the task of philosophy in terms of what he calls (borrowing from Nietzsche) genealogy (Foucault, 1980, 2004). This is essentially a form of historical sociology that seeks to recognize – in, so far as one can tell, a fairly neutral and nonjudgmental way – the diverse ways in which key social concepts of madness, punishment, and sexuality have been conceptualized from past to the present. A second is associated with the postmodern anti-theorist, J-F Lyotard, who – on much the same sociological grounds – urges us to accept that there are no grand recits or overarching meta-narratives (Lyotard, 1984). The third (albeit a good deal nearer to analytical philosophy than these other two) is the moral and social theory of the British philosopher, Alasdair MacIntyre, who has lately defended a unique brand of neo-idealist Aristotelianism (or perhaps, more accurately, Thomism), that denies the possibility of any external moral adjudication between diverse rival social traditions of moral enquiry (MacIntyre, 1981, 1987, 1992).

In the interests of avoiding some confusion, however, there are pressing questions to be asked about the extent to which such views are inconsistent with either an analytical approach to philosophy or (for that matter) the idea of absolute or universal truth. Thus, to begin with, is analytical philosophy at all at odds with the view of Foucault and others, that key social and other concepts are sociohistorically constructed? The answer is, of course, that it is surely not. In fact, it is probably safe to say that all serious modern (if not also many past) philosophers subscribe to what the British analytical philosopher, Michael Dummett (1978), has called the thesis of “the social character of meaning” – precisely the idea that human ideas have complex and philosophically interesting social histories. Indeed, after Wittgenstein, one might well expect those modern analytical philosophers of usage who influenced the postwar revolution in philosophy of education to be sensitive to such linguistic developments – which, indeed, this author thinks they have generally been. However, does this mean that modern analytical philosophers would have to deny any and all possibilities of universal, analytical, or necessary truth? Again it clearly does not, and any such denial would in any case be spectacularly foolish. Indeed, the first effective attempt to refute subjectivist or social constructivist denials of absolute or universal truth was made (in Plato's *Theaetetus*) by Socrates who pointed

out that such denials are themselves absolute claims (to the effect that there is no absolute truth). Likewise, a philosopher such as MacIntyre who claims that there can be no single cross-cultural conception of morality or moral education, presumably holds that this universally – rather than just occasionally – holds.

It is also worth asking how Socrates might have responded to the postmodern claim of Lyotard – often paraded by educational theorists as a self-evident truth – that there are no overarching meta-narratives. It is likely that Socrates might first have asked Lyotard to clarify the precise sense in which he intended this claim to be taken. Could the denial of meta-narrative mean that no one today can possibly believe in any grand theory? (which is, more than likely, a false empirical claim); could it mean that no single theory can be taken to explain everything? (a claim that, while true, would seem virtually trivial – for who would ever have supposed this to be so?); could it mean that no theory could be regarded as successfully or effectively explaining what it sets out to explain? (which would also seem implausible – since many theories appear to explain things rather well); could it mean that any and all explanations are final, so that there is nothing left to say on this or that matter? (which may well be a false claim – but also one that is unlikely to have been widely endorsed); and so on for other possible interpretations. However, the key point for a Socrates would be that such general skeptical sloganizing about the nature and status of theories or explanations can be taken at its face value only at the price of much muddle and confusion of different issues requiring careful articulation and distinction (and that once these have been so distinguished such generalities may cease to be at all coherent or plausible). Indeed, an analytical philosopher's trouble with the sweeping claims of a Lyotard (apart from the fact that they look suspiciously like meta-narratives themselves) is much the same trouble that Socrates discovered in the no-less-sweeping skeptical claims of various sophists and pre-Socratic philosophers: that, once carefully examined, they lack any clear or coherent sense.

However, in this light, the widely held claim that there are traditions of or approaches to philosophy or philosophy of education – such as Marxism, phenomenology, existentialism, critical theory, post-structuralism, postmodernism – that offer viable alternatives to conceptual analysis begins to look less plausible. Indeed, this view is prey to the dilemma that either Marxists, critical theorists, or post-structuralists are engaging in the critical analysis of concepts and arguments or they are not; if they are, then it is difficult to see how they might constitute different forms of or approaches to philosophy; if they are not, then it is difficult to see why one should want to describe what they are doing as philosophy. From this perspective, moreover, the short answer may well be that sometimes they are and sometimes they are not. In so far as Marxists,

phenomenologists, and post-structuralists do sometimes seem to have been in the business of making sweeping quasi-empirical claims about the social provenance of concepts and knowledge, there would seem to be something to the received analytical suspicion that all of this is not so much philosophy as bad speculative sociology. On the other hand, there are evidently moments at which such non-analytical philosophers seem to be quite clearly involved in the clarification of concepts and arguments – which rather undermines any claim that such work rests on an alternative approach to that found in the analytical tradition. On this view, indeed, while it seems clear enough that much nineteenth-, twentieth-, and twenty-first-century pragmatism is about as analytical as it gets – and is therefore of considerable philosophical interest – it may also be that some Marxism, existentialism, and post-structuralism is not (whatever its sociological or other interest) really philosophy as such at all.

Recent Philosophy of Education

In so far as all this is so, it would seem that there is at least some confusion in any alleged repudiation of the analytical tradition: for, if this means a rejection of the hard-won analytical techniques bequeathed to us by the likes of Plato, Aristotle, Aquinas, Descartes, Kant, Russell, Quine, Wittgenstein, Ryle, Austin, and others – or, yet worse, some kind of denial of the laws of thought and logic themselves – then it is difficult to see how anyone might subscribe to this and still claim that what they are doing is philosophy. To be sure, some feminists have come close to claiming that the laws of logic are to be regarded as mere expressions of the power and control of males by males – but how might we either agree or disagree with those who refuse to recognize the basic rules of argument? What is true, of course, is that some of the above philosophers who have evidently contributed to the development of analytical philosophical techniques are also philosophers who philosophized in the light of particular metaphysical, epistemic, and moral perspectives or commitments. Thus, for example, Thomists are Roman Catholics who are committed to a certain perspective on Catholic theology and moral teaching deriving from Aquinas, and Quineans are pragmatists committed to no less distinctive perspectives on knowledge and truth. That said, many modern Thomists have been avowedly analytical in their philosophical approach and few pragmatists would regard themselves as departing from the overall analytical drift of philosophical method. It is also not hard to see how there might well be analytical Marxists or existentialists – and some phenomenologists and critical theorists seem to have been analytical enough.

However, it is also possible that what those philosophers who would want to regard themselves as opposed to

analytic method or the analytic tradition have wanted to deny is just the idea that there can be a rationally neutral or value-free conception of, or approach to, philosophy that does not presuppose metaphysical, epistemic, or moral commitments. This is certainly a hard claim to evaluate – not least because, from an analytical viewpoint, it is again not altogether clear what it means. On the one hand, one wants to ask who exactly is supposed to have denied this – since it is generally as clear to analytical philosophers as anyone else that although arguments may be more or less valid, the basic premises on which arguments are based may not in turn have been arrived at by any further argument (although they may be supported by better or worse intuition or evidence). On the other hand, it seems equally clear that philosophers (and other people) have been driven to change their minds or commitments precisely on the grounds that they no longer found the arguments for such commitments entirely compelling: thus, for example, there are well-documented instances of philosophers rejecting their Catholic or Christian faith because they no longer regard the arguments commonly advanced for such faith to be compelling – as well as of atheists becoming theists under the same compulsion. In short, a (relatively) detached analytical view of a matter and an appreciation that our arguments may not be as strong as we formerly took them to be can help us to see more clearly and to correct our views – which would be precisely a major analytical claim.

It is in precisely this spirit that many contemporary philosophers of education have persisted in deploying analytical methods in the interests of further clarifying a range of key educational concepts, questions, or arguments that are often no doubt prone to ambiguity, conflation, and other sorts of unclarity. Some of these philosophers are Christians, others are Marxists or socialists, others are secular liberals (or even neoliberals) and few of these would see these different commitments as impediments to serious intellectual engagement with those of other commitments. What is of concern to such educational philosophers is the more dogmatic adherence to Marxism, feminism, or post-structuralism which might claim that no argument can be anything other than bourgeois or patriarchal hegemony, or more radical interpretations of Nietzsche or existentialism that deny that there is anything to human value beyond arbitrary personal

choice. For, if we are forced to conclude that any and all aspiration to unbiased and/or objective argument and truth is in vain, this surely spells not only the end of philosophy but of anything much worth calling education.

See also: Formative Assessment in Teacher Education and Teacher Professional Development; Liberalism and Education; Philosophy of Education: Overview; Pragmatism; Reason and Rationality; The Aims of Education.

Bibliography

- Cooper, D. E. (1980). *Illusions of Equality*. London: Routledge and Kegan Paul.
- Dummett, M. (1978). *Truth and Other Enigmas*. London: Duckworth.
- Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings 1972/1977*, Gordan, C., Marshall, L., Mepham, J., and Soper, K. (trans.). New York: Pantheon Books.
- Foucault, M. (2004). *The Archaeology of Knowledge*, Sheridan Smith, A. M. (trans.). London: Routledge.
- Frege, G. (1978). *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*. Oxford: Blackwell.
- Geach, P. T. and Black, M. (eds.) (1960). *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell.
- Hamilton, E. and Cairns, H. (eds.) (1961). *Plato: The Collected Dialogues*. Princeton, NJ: Princeton University Press.
- Hamlyn, D. W. (1978). *Experience and the Growth of Understanding*. London: Routledge and Kegan Paul.
- Hirst, P. H. (ed.) (1983). Educational theory. *Educational Theory and its Foundation Disciplines*, pp 3–29. London: Routledge and Kegan Paul.
- Lyotard, J. -F. (1984). *The Post-Modern Condition: A Report on Knowledge*, Bennington, G. and Massumi, B. (trans.). Manchester: Manchester University Press.
- MacIntyre, A. C. (1981). *After Virtue*. Notre Dame: Notre Dame Press.
- MacIntyre, A. C. (1987). *Whose Justice, Which Rationality ?* Notre Dame: Notre Dame Press.
- MacIntyre, A. C. (1992). *Three Rival Versions of Moral Enquiry*. Notre Dame: Notre Dame Press.
- McKeon, R. (ed.) (1941). *The Basic Works of Aristotle*. New York: Random House.
- Passmore, J. (1980). *The Philosophy of Teaching*. London: Duckworth.
- Peters, R. S. (1958). *The Concept of Motivation*. London: Routledge and Kegan Paul.
- Peters, R. S. (1966). *Ethics and Education*. London: George Allen and Unwin.
- Peters, R. S. (ed.) (1973). Aims of education. *The Philosophy of Education*, pp 11–29. Oxford: Oxford University Press.
- Peters, R. S. (1981). *Essays on Educators*. London: George Allen and Unwin.
- Russell, B. (1956). *Logic and Knowledge: Essays 1901 – 1950*. London: George Allen and Unwin.
- Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.
- Wittgenstein, L. (1961). *Tractatus-Logico-Philosophicus*. London: Routledge and Kegan Paul.

PHILOSOPHY OF EDUCATION – PHILOSOPHICAL THEMES

Contents

Adult Learning: Philosophical Issues

Autonomy

Bildung

Critical Theory and Pedagogy

Critical Thinking

Environmental Education

Identity

Inclusion

Justice and Care

Lifelong Learning

Post-Structuralism and Education

Reason and Rationality

Religious and Spiritual Education

The Capabilities Approach

Adult Learning: Philosophical Issues

D Beckett, The University of Melbourne, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Experience and Andragogy

If a single concept marks out the learning of adults from the learning of children, it is experience. The pursuit and accumulation of learning, and its refinement, by adults, is usually underpinned by some assumptions about the integrity and persistence of experience. If an individual has lived longer (i.e., has passed through childhood), they are regarded, initially, as more adept at identifying what they have learned, in the past, and what they want to learn in the future. Moreover, adult individuals are also assumed to have insights into how they learn best now, and in the past, and expect to learn, in the future.

Experience, closely connected to the life span of individuals, is therefore a fruitful way into a review of the main philosophical issues involved in adults' learning. There are clearly, also, intimate conceptual links from this singular life span across into the parallel area of lifelong learning, which will emerge below. The immediate task is to set out how the fruitfulness of a single concept (experience) should be distinguished from the singularity of its manifestation, that is, in individuals, all of whom experience experience differently. As Candy puts it, the assumption that experiences are only located in individuals is "so deeply entrenched in the ethos of adult

education as to be thought 'obvious' or 'self-evident' and to thus be beyond question" (1987: cited in Usher *et al.*, 1997: 93).

What is so special about adults' experiences, as adults? Distinguishing between children's and adults' learning was crisply and famously done by Knowles (1970), who was not a philosopher but gave significant shape to the field of practice called adult education, at least in North America. His andragogy (not androgogy) theorized adults' learning through the explicit utilization of experience. Adults learn best, according to Knowles, when they

learn how to take responsibility for their own learning through self-directed inquiry, how to learn collaboratively with the help of colleagues rather than compete with them, and, especially, how to learn by analyzing one's own experience. ...[This] is the essence of the human relations laboratory. (Knowles, 1970: 45)

By contrast, Knowles' pedagogy marked out children's learning, in which instruction continued to be central, due to the limited and naive nature of youthful experience. Later, Knowles modified this crisp distinction in favor of more experientially inclusive learning for individuals of any age.

So if we take experience as the point of entry to theorizations of adult learning, we can readily locate it

within humanist scholarship, more generally. Knowles quite explicitly drew upon the psychologists, Carl Rogers and Abraham Maslow, but philosophers can agree that advocating experience as the bedrock concept for adult learning immediately looks Deweyan, and educators in general have been drawing on his vast publications well before andragogy surfaced in the 1970s and 1980s (Dewey, 1938). Of course, Dewey mainly had school-based education in his sights, but his advocacy of organic, holistic learning (and teaching) has attracted many adult educators.

However, before we go on, getting some idea of where nonschool education occurs is important in understanding the impact of Deweyan (and Knowlesian) experience. There is no single sector of the education industry in which one could find adults' learning (unlike schooling, where children mainly and compulsorily learn). We can immediately identify formal organizations and institutions, such as vocational colleges and universities, but adult learning occurs informally, in daily life: at paid work (in employment), unpaid work (in the home and in community settings), public discourse (mass media including infotainment), and political advocacy (e.g., the green agenda). There is, then, an amorphous and distributional quality about adult learning, which a simple reliance on experience under-recognizes. We will return to that inherent contextual messiness shortly.

Shaping Foundations

Giving shape to the messiness of theorizations of adult learning has been splendidly done by Elias and Merriam (2005) over the past three decades. They list and describe various main philosophical foundations of adult education: liberal, progressive, behaviorist, humanistic, radical and critical, analytic, and postmodern. It is worth noting that Knowles provided the foreword to the first edition in 1980, beginning with the claim that "[d]uring the entire time I have been in the field of adult education, which is almost half a century, I have heard it being criticized for not having a philosophical foundation (xi)." Elias and Merriam do not attempt to provide such a foundation, but instead provide an accessible overview of several main writers in the various approaches which are frequently drawn upon to underpin policies and practices in adult education. As expected, these various approaches are generically educational, so giving shape to the messiness of adults' education is assumed, rightly, to be a subset of these.

Experience – that is to say, how adult learning is best conceptualized – fares differently within each of these foundations or approaches, as Elias and Merriam (2005) show. Liberal writers draw upon the classical Greeks, the values of medieval schooling, and the character formation

thought to evolve from the contemplative study of an intellectual canon: literally the classics. By contrast, progressivists, such as William James and John Dewey, engage the relationship of subjective experience to the worlds of work, democracy, and of sociality, generally, with change as the driving force.

Behaviorism is well known through the work of Pavlov, Watson, and Skinner, where experiences are manifest as modified behavior, achievements of outcomes, and reinforcement (or not) of these. Humanism has been already introduced, above, through the Knowlesian contribution to adult education, but its heritage is much broader and deeper: Heidegger, Sartre, Camus, Marcel, and Buber, in European scholarship, and Dewey, in North America, for example. In adult education specifically, Brookfield, Mezirow, and Jarvis have each shaped richer contributions to theory, policy, and practices from within this tradition (yet none would claim to be philosophers).

Radical or critical adult education takes sociopolitical stances on learning, and advances it in the spirit of ideologues such as Marxism, some feminisms, anarchism, critical theory, environmentalism, and so on. Holt, Goodman, Illich, and mainly Freire are significant contributors here.

Analytic philosophy is represented by Lawson and Paterson, who have made influential early contributions in the United Kingdom (about the same time as Knowles, in the USA) drawing upon the London School of philosophy of education, well known through the work of R.S. Peters and Paul Hirst, for example (Curran *et al.*, 2003).

Elias and Merriam (2005), draw attention, finally, to postmodern approaches, which "offer... a trenchant critique of the entire enterprise of adult education," (p. 14) in that a host of scholars in a vast range of fields, across the arts and social sciences, have interrogated the modernist assumptions of unilateral and monocultural progress, in human affairs in general. Postmodern perspectives in adult education were evidenced in the second paragraph of this article you are now reading: Usher *et al.* (1997: ch 5 *passim*) refuse to take experience as a given (bedrock) phenomenon. We may ask, in the spirit of postmodernity: Whose experience? In what contexts? By whom is it perceived? These are three interrogations of the modernist experiential narrative which was our point of entry, and in asking these sorts of questions, postmodernists seek to expose the various constructions of learning, learners, and of teachers, from which adult education can, itself, be interrogated.

Experience is, then, manifest diversely and intersubjectively (among, not merely between, each other). It follows that all constructions are relational: they build upon, and between, varieties of subjectivities. Postmodernity works the relationality of experiences even harder than that. Identities and practices are themselves

constituted in and by the networks and nodes by which humans associate. We live, move, and have our being through such associations; and postmodern perspectives in philosophy of education, and in adult learning, in particular, are serious about diverse and locally devised interrogations of problems, issues, and challenges.

Embracing Diversity

So Elias and Merriam do us a fine service in mapping the main foundations of adult education, and introducing persisting and influential bodies of scholarship to practitioners and other theoreticians. Overall, we can agree that there is no one-size-fits-all philosophy of adult education, nor is there one best approach to adult learning. In their chapter in a well-received book covering all of adult education and training, Fenwick and Tennant (2004: ch 5), set out to understand adult learning, by making three assumptions:

First, no one theory of learning or of facilitating learning trumps the others. There is no generic essentialised “adult learner”. . . . Second, learning is not a mental process occurring in a vacuum. The context of a person’s life – with its unique cultural, political, physical, and social dynamics – influences what learning experiences are encountered and how they are engaged. Furthermore, “context”. . . is active and dynamic. Third, the “learner” is not an object separable from the “educator” in teaching-learning situations. The positionality of the educator (whether the expert, coach, liberator, observer, arbiter, commentator, guide, decoder) affects how learners perceive, feel, behave, and remember. (Fenwick and Tennant, 2004: 55.)

Drilling down even more sensitively into adults’ experiences often requires a phenomenological approach. Here, bringing one’s awareness of one’s consciousness to the surface, and articulating it, is essential to the realization of a sense of personhood (Stanage, 1987). This Husserlian approach holds much potential for adult education in, for example, nurses’ leadership of antenatal classes, where utterly personal expectations, and real experiences, of pain, can be trawled.

When faced with the messiness of theories, we can conclude so far that a postmodern spirit, rather than embrace a singular postmodernism, enables an eclectic and particularistic approach to central philosophical issues (subjectivities and their construction in and around sociopolitical associations), that themselves revolve around adult learning’s traditional focus on the integrity and persistence of experience.

Now we return to the messiness of contexts. Sensitivity to diversity of contexts of adult learning is essential, but is difficult when formal and informal adult learning

may well work against each other, and frequently do. An apprenticeship requires on-the-job assimilation of masterful skill (the informal) and typically also 1 or 2 days a week in a vocational college (the formal), over several years. Yet these experiences may not dovetail. Similarly, one of the main criticisms of traditional university teaching (formalized learning) at least in the West has been its disdain for those experiences brought to the lecture hall and tutorials, not just by school leavers, but also by mature age or further education students (informal learning).

Know-How at Work

The formality of what philosophers call propositional knowledge – necessary though it is for a worthwhile education – is now regarded as insufficient for genuinely lifelong learning, in particular, for enhancing a graduate’s employability across the life span beyond the compulsory years of schooling. How, for example, a university or a vocational college, engages students’ informal learning, which is inevitably and desirably experiential, raises fundamental issues of teaching, program design, and assessment for that institution. Ryle’s (1949) important distinction between knowing that X (where X is a proposition) and knowing how X (where X is an experience) has become problematic for many educators as lifelong-learning policy discourses traverse experience as a human phenomenon, trawling it for meaning.

Know-how is emerging as a fruitful site of analysis in its own right. For example, Polanyi (1966) picked up the tacit (inarticulable) significance of knowing how: we know more than we can say. Decades on, we are now more willing to acknowledge that powerful adult learning resides in what have hitherto been low-status experiences, often called intuition, common sense, women’s ways of knowing, and nous. The epistemological significance of these is now under evaluation, especially in the light of Deweyan accounts of how learning processes, such as reflection, both enhance, and are constitutive, of action. This is in the global context of adult learning at and for work, which has emerged as a major new component in the new world of free trade, rampant consumerism, and market-driven organizational changes, with the collapse of Communism in the 1980s. As Boud and Garrick (1999) put it:

Learning at work has become one of the most exciting areas of development in the dual field of management and education. . . . Modern [sic] organisations ignore learning at the cost of their present and future success. . . . learning has moved from the periphery. . . to the lifeblood that sustains them. (Boud and Garrick, 1999: 1)

Significant philosophical issues are raised by the vocationalization of adult learning, to which we now turn.

Livelihoods, Competencies, and the New Model Worker

Adults learning at and from work undergo experiences only some of which are educative, and only some of which are of enterprise, or strategic, significance. So what, then, is worthwhile workplace learning? And how can it be recognized? A good deal of attention has been given to the calibration of such learning, in various competence structures in most developed economies, and in many developing economies. The extent to which any vocational or workplace competencies can capture the really worthwhile is highly debatable, but so also is the debate – what is really worthwhile? Making a livelihood is essential: how does learning contribute to that? Only in the past couple of decades has this question achieved such prominence. After all, any educational significance that learning at and for work may have has often not been the main motive governments and organizations have in embracing competencies. Learning is not the main activity of most organizations and enterprises. Industrial, labor market, and productivity considerations are important motivators to be weighed among the educational and philosophical, here.

What helps to shape such debates is the acknowledgment of an epistemological and ethical tension. At one polarity, is a reliance on behaviorism issues in low-level checklisting of skills performed, without regard for the rich practical judgments often implied by such performances, and invites reductive, outcomes-driven accounts of human learning. Teaching and training, on this behavioral approach, consist of modifying behavior by repetition, motivators, and incentives which may have little to do with intrinsic satisfactions. An organization, or national government, can run entire learning programs on this approach.

The other polarity is humanist, which assumes the whole person is the worker, not merely a set of hands, but ascribes to the workplace the expectations of fulfillment attributable to human life in general, accompanied by the vocational self-direction to achieve these fulfillments. Again, an entire organization or national government can run entire learning programs on this approach. National generic or key competencies can be identified, often without regard for contextual sensitivities, against which varieties of more or less reductive performance evidence can be assembled, and, in some more enlightened case, some Aristotelian-influenced practical judgments can be inferred.

More commonly, controversies accrue around the fluidity and admixture of behavioral and humanist assumptions in particular contexts. Does the so-called new model worker (ideally self-directedly flexible, well-rounded, creative) get caught up in the ideology of individual success to the detriment of her or his integrity as a person? Usher and Edwards (1994) argue, following Foucault, that “the exercise of modern disciplinary power is exemplified

in the panopticon” (p. 101), which was the nineteenth-century circular prison designed as cells radiating spoke-like, from a hub, with the guards in the hub, such that surveillance was always assumed, by the prisoners, even if the hub was unstaffed.

Usher and Edwards critique the new model workplace, and the competencies that calibrate the work, if workers are learning to be complicit in their own organizational or professional surveillance. This raises substantial ethical and epistemological issues for adult learning, provoked by the very shape and structure of accountabilities for worthwhile vocational learning. Although these issues arise at national, organizational, and enterprise level, they are best resolved locally and in particular cases, with due recognition of contextual sensitivities from those with the experiences that matter most. In this way, a postmodern spirit is played out in judgments of worthwhile adult learning.

Knowing That, Knowing How . . . , and Knowing Why

Epistemology and ethics are only part of the adult learning story. The already amorphous and distributional quality of adult learning is exacerbated by the immediacy of the self-direction which Knowles’ traditional andragogy highlights. Adults learn best when they see the point of it.

Vocationalization within adult learning (as part of a globalizing lifelong-learning policy scenario), such as has just been discussed, impels us to consider the extent to which adults learn best when they can see its direction: this is to know why X. . . (where X is a reason, purpose, or goal).

As we have seen above, translating this across formal and informal contexts is complex: individuals will have a variety of motivations, often in any one day, and often derived from sociocultural allegiances. Through work, home, community, and national institutions, individual adults locate their sense of what is worth achieving – not only for themselves as individuals, but also for their sense of their shared self-hood. What kind of worker, partner, parent, or citizen do I, or rather, we, want to be? Once the arena for learning is broadened out, we can locate ourselves as distributed among and between diverse allegiances. This lateral dimension complements the linear dimension – the journey of the individual across the life span.

In this way, we can add to the traditional Rylean analysis, a teleological perspective: we learn, from each other, what it is to know why we have particular goals and motives. These might come from our employment (such as a corporate vision or strategy), or from our home or community allegiances (such as spiritual or activist values), or from national imperatives (policy discourses which attract or repel).

On this tripartite analysis, worthwhile adult learning is constituted, coextensively, by know-what, know-how, and

know-why. This adds to propositional knowledge the significance of sensitivity to our own learning experiences (knowing how X...), and also adds the directionality of this learning, or the teleological dimension (knowing why X...). This analysis locates the individual as an intersubjectivity among not only other individuals, but also as relationships of associations and allegiances, in groups of diverse sociocultural significance.

Adult learning, analyzed this way, is a wide and deep phenomenon, with astonishingly diverse manifestations, only some of which fit with traditional assumptions of the learner as an individual, or subjective self, enrolled in formal studies in an educational institution. On the contrary, this analysis requires philosophical enquiry into diverse contexts where, almost case by case, particularistic approaches can be tailored to prospects of successful learning. So we need to look closely at educative practices, from which adults can learn.

Practices and Identities

The practice turn in education has taken much attention in recent years (Hogan, 2008), and not just with adults. However, here we are advancing an experiential analysis of adults' learning in particular contexts of practice, so it is important to bear in mind some actual learning activities which could be and often are assembled into programs and practices for groups of adults, both formally and informally:

- problem solving/workshopping; learning-to-learn/double-loop questioning; critical thinking/evaluation exercises;
- negotiation/collaboration/interactivity/interpersonal skill formation via groupwork;
- literacy and numeracy through real-life excursions/reporting;
- case studies/informal presentations;
- simulations/role playing;
- journaling;
- work placements – real on-the-job learning/training with episodes of expert instruction/guidance (coaching);
- peer-group instruction/guidance (debriefing/consulting/mentoring);
- audio-visual presentations and resultant groupwork;
- project-based team activities.

Diversity is not self-justifying. Here, the purposes in assembling diverse activities within particular contexts expose adults to diverse ways of learning – experiences – from which they build their own capacity for self-direction.

Implicit in this very Deweyan concept of organic growth is both self-direction and self-identity. Identities, or subjectivities, are constructed and reconstructed, in

practices – such as in classrooms, anywhere in the world. The way, then, to clarify important aspects of experience for adult learning, given the amorphous and distributed nature of those experiences, is to pay close attention to the practical – even as a philosophy. In the policy context of lifelong learning, Leicester *et al.* (2007) argues for practical diversity in this way:

Our notion of “practical philosophy,” in encouraging attention to the diversity of voices and experiences in the real world, also has this political tendency [that is, to be more inclusive of the interests of all people]. Postmodernist epistemology recognises that knowledge is validly constructed from the intersubjective agreement in the experiences of oppressed groups and not just from that of the educated group (the group which writes papers) which has tended to exclude these voices hitherto. However...we are also suggesting that practical philosophy, recognising that meaning is rooted in the (complex and context-dependent) uses of a word, has implications for conceptual analysis regardless of any political commitments or implications. (Leicester, 2007: 263)

Both Leicester's constructive practical philosophy and the author's listing of a range of learning activities (listed at the beginning of this section) have in common an experientially diverse approach to the practice turn. Most important in this is that both imply a richer and hitherto model of agency. Adult learning is underpinned by assumptions of the integrity and persistence of human experience, as we noted at the outset, but in this era of lifelong learning, individuals are assumed, and indeed required to exercise, in their very self-direction, and to advance their self-identity, sophisticated autonomy, which assumes agency. It is in this sense that Schon's (1987) reflective practitioner gets a new lease of life: socially located, amidst the intersubjective messiness of daily work life, and fully agentive. The revitalization of Aristotle's *phronesis* in clinical medical practice is a case in point (Montgomery, 2006).

The Project of the Self: Always Work In Progress

If, as has been argued above, and building upon a Rylean analysis, a tripartite, more intersubjective, more social notion of adult experience is required, then agency itself needs to be rethought. If adults – and, indeed, all humans – learn not just individually, but powerfully from each other, then agency as expressed in what can be called the project of the self, needs recasting. Self-direction becomes self-management, but less of me, and more of us.

The vocationalization of higher education has been rightly criticized if it merely reduces adults' learning to employability. Such reduction of the head to the hands

preserves a Cartesian dualism at the very time adult and lifelong learning is taking the organic and holistic approach that was argued by Dewey and many others in adult education across the twentieth century. Reductive attempts to produce the new model worker – utterly flexible, compliant, and creative – deserve criticism when, as often occurs, competencies and shallow skill-talk drive the educational and policy discourse. It may be that a character-driven capabilities approach (following Nussbaum) is a more defensible formulation of the project of the self.

However, what is interesting beneath all this is the need to revitalize agentive practices: these are opportunities for adult learners to make decisions and judgments in particular contexts which meet the particular purposes of those contexts. Practice-based learning activities immerse participants in intersubjective relationships with various forms of association (e.g., a team, mentorship). Moreover, they can take skill acquisition seriously – beyond behavioristic training, and often meld with broad organizational development initiatives, and can do this inclusively – with the respect for constructively diverse experiences, which Leicester claims for them, above.

In addition, the embodied nature of adult learning is crucial. Rather than leave materiality of the learning to the doing of the hands (which preserves a Cartesian dualism), some current research is bringing to prominence the whole person with a due regard for how what we find ourselves doing expresses not merely our learning, but our identities as well (O'Loughlin, 2006).

Thus, agency and identity are intertwined in deliberately providing a (strategic, purposeful) diversity of learning activities for particular formal and informal contexts. Intersubjectivity requires the selection of social, reflective (both public and private), and personal experiences of the most profound kind. In this way, the practice turn demonstrates what a new approach to agency might look like when adults gather to learn. Over time, and explicitly heading toward assessment tasks, or some other agreed outcomes, the intention is to grow self-direction and self-identity. Choices – decisions – are negotiated, and in these ways, human decision-making capacity is shown to be valued.

Beckett and Hager (2002), argued for the centrality of practical judgments, as a relational way of advancing a new epistemology of practice, one which decenters the traditional Cartesian and even Platonic epistemology. Educators have traditionally assumed that coming to know something is to arrive at a state of the mind as evidenced in accounts of what is cognitively the case – this is about whether the propositions are in place there. Yet adults learn best when their experiences are taken seriously, so a tripartite Rylean epistemology should supplement the traditional account. The low status of the tacit, intuitive, reflective, phenomenological, embodied, and the socially efficacious leaves much human

experience out of the educational vision, wherever it is, but especially for adults.

By taking seriously the holistic nature of particular everyday experiences, such as those of the workplace and the pedagogically diverse classroom, adult educators, whether they are practitioners (who have real expertise in inclusive learning strategies) or researchers (who have interests in relational practices), can find philosophically rich ways through the messiness of adult learning.

See also: Autonomy; Critical Theory; Critical Theory and Pedagogy; Feminism; Identity; Liberalism and Education; Lifelong Learning; Phenomenology and Existentialism; The Analytical Tradition; The Capabilities Approach.

Bibliography

- Beckett, D. and Hager, P. (2002). *Routledge International Studies in the Philosophy of Education 14: Life, Work and Learning: Practice in Postmodernity*. London: Routledge.
- Boud, D. and Garrick, J. (eds.) (1999). *Understanding Learning at Work*. London: Routledge.
- Curran, R., Robertson, S., and Hager, P. (2003). The analytic movement. In Curran, R. (ed.) *A Companion to the Philosophy of Education*, pp 176–191. Oxford: Blackwell Publishing.
- Dewey, J. (1938). *Experience and Education*. New York: Macmillan.
- Elias, J. L. and Merriam, S. B. (2005). *Philosophical Foundations of Adult Education*, 3rd edn. (orig. pub., 1980). Malabar, FL: Krieger Publishing.
- Fenwick, T. and Tennant, M. (2004). Understanding adult learning. In Foley, G. (ed.) *Dimensions of Adult Learning: Adult Education and Training in a Global Era*, pp 55–73. Sydney: Allen and Unwin.
- Hogan, P. (ed.) (2008). Philosophical perspectives on educational practice in the 21st century. *Studies in Philosophy and Education* 27 2–3.
- Knowles, M. (1970). *The Modern Practice of Adult Education: Andragogy versus Pedagogy*. New York: Association Press.
- Leicester, M., Twelvetrees, R., and Bowbrick, P. (2007). Philosophical perspectives on lifelong learning: Insights from education, engineering and economics. In Aspin, D. (ed.) *Lifelong Learning Book Series 11: Philosophical Perspectives on Lifelong Learning*, pp 258–274. Dordrecht: Springer.
- Montgomery, K. (2006). *How Doctors Think: Clinical Judgment and the Practice of Medicine*. New York: Oxford University Press.
- O'Loughlin, M. (2006). *Embodiment and Education: Exploring Creatural Existence*. Dordrecht: Springer.
- Polanyi, M. (1966). *The Tacit Dimension*. New York: Doubleday.
- Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.
- Schon, D. (1987). *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*. San Francisco, CA: Jossey-Bass Publishers.
- Stanage, S. (1987). *Adult Education and Phenomenological Research: New Directions for Theory, Practice, and Research*. Malabar, FL: Krieger Publishing.
- Usher, R., Bryant, I., and Johnston, R. (1997). *Adult Education and the Postmodern Challenge: Learning Beyond the Limits*. London: Routledge.
- Usher, R. and Edwards, R. (1994). *Postmodernism and Education*. London: Routledge.

Further Reading

- Archer, M. (2000). *Being Human: The Problem of Agency*. Cambridge, UK: Cambridge University Press.
- Bagnall, R. (1999). *Discovering Radical Contingency: Building a Postmodern Agenda in Adult Education*. New York: Peter Lang.

- Beckett, D. (2004). Embodied competence and generic skill: The emergence of inferential understanding. *Educational Philosophy and Theory* **36**(5), 497–509.
- Belenky, M., Clinchy, B., Goldberger, N., and Tarule, J. (1986). *Women's Ways of Knowing: The Development of the Self, Voice and Mind*. New York: Basic Books.
- Bright, B. (ed.) (1989). *Theory and Practice in the Study of Adult Education: The Epistemological Debate*. London: Routledge.
- Brockett, R. (ed.) (1988). *Ethical Issues in Adult Education*. New York: Teachers' College Press.
- Brookfield, S. (1986). *Understanding and Facilitating Adult Learning*. San Francisco, CA: Jossey-Bass.
- Hager, P. and Halliday, J. (2006). *Lifelong Learning Book Series 7: Recovering Informal Learning: Wisdom, Judgement and Community*. Dordrecht: Springer.
- Jarvis, P. (1997). *Ethics and Education for Adults in a Late Modern Society*. Leicester: National Institute of Adult Continuing Education (England and Wales).
- Morris, G. and Beckett, D. (2004). Performing identities: A new focus on embodied adult learning. In Kell, P., Singh, M., and Shore, S. (eds.) *Adult Education @21st Century: Global Futures in Practice and Theory*, pp 121–136. New York: Peter Lang Publishing.
- Mulcahy, D. (2000). Body matters in vocational education: The case of the competently-trained. *International Journal of Lifelong Education* **19**(6), 506–524.
- Van Manen, M. (1990). *Researching Lived Experience: Human Science for an Action Sensitive Pedagogy*. Albany, NY: State University of New York Press.
- Welton, M. (ed.) (1995). *In Defense of the Lifeworld: Critical Perspectives on Adult Learning*. Albany, NY: State University of New York Press.

Autonomy

C A Wringe, Keele University, Keele, UK

© 2010 Elsevier Ltd. All rights reserved.

The Concept of Autonomy

Autonomy was originally a political term meaning sovereign or self-governing, as applied by the Greeks to independent city-states, and it is still used today when we speak of autonomous regions in such countries as China or Spain. In the context of present-day moral and educational philosophy, however, it is usually ascribed to individuals whose conduct, way of life, and belief systems are determined by their own choices and in the light of their own judgment.

Although Kant (1785) is normally credited with popularizing our modern usage, the notion of rational autonomy and its justification, though not the term itself, was already an important part of Enlightenment thinking in political and other writing that sought liberation from the authority of the ancients and tradition in matters of belief and conduct. The faculty of rationality or reason was widely held to be equally available to all in matters of factual or religious belief (Hobbes, 1642) and if political obligation was to be considered valid, it had to take a form under which individuals in a state of nature, or guided by reason alone, would voluntarily agree to put themselves (Hobbes, 1642; Locke, 1689). Human beings were held to enter the world with no prior indication of their authority or subordination (Locke, 1689) and all were said to be the best judges of their own best interests. Such thinking remains central to our understanding, not only of liberal democracy but also of our social norms, social policies; health education and health promotion (Heathcote, 2000); medical ethics; as well as our freedoms of religion, speech, the press, enquiry, research, teaching, and learning.

The state of not being autonomous is termed heteronomy, controlled by others or, importantly in the moral field, by one's whims, inclinations, and instincts. These, Kant held, represented the part of ourselves that belongs to our material being and is subject to the material laws of nature rather than our rational humanity which is capable of exercising moral freedom by prescribing laws for itself. According to Kant, one's autonomy is one's power to follow a self-chosen moral law arrived at in the light of reason alone.

Autonomy may be distinguished from the wider concept of freedom, which in its so-called negative form usually refers to the mere absence of external constraint or coercion. The term positive freedom, may be used with unfavorable connotations by liberal writers (Berlin, 1969) to imply that the person concerned has been liberated

from illusion, error, or false consciousness in order to pursue or adopt a particular cause or belief system, but with the help of some fairly robust guidance or re-education, not to say brainwashing or coercion. The dependence of autonomy on rationality distinguishes it from authenticity, which may likewise be cited as an educational aim (Bonnett and Cuypers, 2003). Proponents of authenticity may, indeed, point to the arguments of rationality as barriers to the genuinely individual will. Truly authentic choices, it may be held, may be essentially criterionless and attempts to make them according to canons of reason or even consistency may be seen as a form of bad faith.

Moral autonomy is an essentially mental quality enabling the rational agent to overcome such nonrational influences as emotion, obsession, addiction, and weakness of the will, as well as error and falsehood or indoctrination, both in making individual judgments and in choosing, pursuing, and amending his or her conception of the good life.

Autonomy as an Educational Aim

In certain respects, the concern of post-1960s philosophers of education with autonomy goes some way to provide a systematic rationale for the earlier writings of progressive educators who attacked not only the coerciveness and rigid assumptions of transmission-based traditional education but also aspired to enable their pupils to think and act independently for themselves (Dewey, 1916; Neill, 1962). Crucially, however, the child-centered assumptions of many progressive educators that the unconstrained and virtually uninstructed child would naturally develop both moral and intellectual qualities, is fundamentally rejected by philosophers of education from the 1960s onward (Peters, 1969; Hirst and Peters, 1970; Dearden, 1972). Both the virtue of autonomy and the rationality that underpins it were seen by them as sophisticated social achievements, which have taken many generations to develop and must be systematically taught rather than simply allowed to unfold.

Promoting the learner's autonomy and by implication his or her powers of reasoning came to be seen as one of the central aims of education during the largely Kantian era of philosophy of education in the 1960s and 1970s (Aviram, 1986). For many it was, and for some indeed it may remain, part of the very definition of education itself. Learning processes, which promote belief and compliance

without ensuring the learner's autonomy in respect of belief and action tend to be termed, at best training or socialization, at worst indoctrination (White, 1970). The tight connection between the concepts of education and autonomy has led many writers to adopt the supposedly neutral term schooling in relation to a process that may be seen as enslaving rather than empowering and leaves unanswered the question of whether critical scrutiny is encouraged, or even possible (Illich, 1971; Barrow, 1981). The disparaging use of the term socialization, when contrasted with education, goes even further with its Marxian implications of engendering false consciousness and subordination of the true interest which the autonomous learner would rationally choose to follow, to that of others.

The concept of autonomy has been fundamental to much thinking about both the justification and the curriculum of education (White, 1982, 1990, 2007). In one key text it is proposed that, given the nature of the human predicament, we unlike other creatures, are constantly obliged to make choices for ourselves (Peters, 1973). Often these will be choices with moral implications and, given the view of moral judgment then widely prevailing, good moral choices were seen as a matter of applying higher-order moral principles to the dilemma with which one was confronted (Hirst, 1972). Various writings on moral education and moral development take it as axiomatic that maturity of moral development and success in moral education are achieved when young people are able to do this without hindrance from the various forms of heteronomy by which human beings are beset (Kohlberg, 1968).

The necessity upon human beings of making choices and the related goal of promoting autonomy of the individuals who are to make and be affected by them, is also taken as a justification for the knowledge curriculum of education. If we are to make valid choices about our lives we need to know both about the range of options available to us and the implications of our choices in the real world: what satisfactions, valuable pursuits, and worthwhile ways of life are open to us; what will be the effects of our choices and what significance may be attributed to them. In the modern world, it has been argued that this need is particularly strong in view of the greatly increased scope for personal and vocational freedom available to individuals, as well as the range of leisure pursuits and other satisfactions which the modern world presents (White, 2007).

We cannot here discuss in detail how the various so-called forms of knowledge, or forms of human sensibility and understanding, contribute in their various ways to the meeting of this need. Of significance, however, is the degree of importance attached, in the educational discussion of the forms of knowledge or disciplines, to the supposed tests for truth or validity central to each.

Autonomous choosers not only need a great deal of content knowledge about the world, but they also need to be in a position to subject what they are told to rational criticism and, if necessary, generate new knowledge and understandings for themselves (Hirst, 1965; Dearden, 1975). This has obvious implications, not only for the curriculum content to be selected but also for the manner in which it is to be communicated, for relations between teachers and learners, and for the organization and management of educational institutions (Hirst and Peters, 1970).

Although the development of moral autonomy and intellectual autonomy in the pursuit of individual disciplines may be important in all areas of study, both have a particular place in the field of citizenship education in a liberal democratic society. Central to the conception of such a polity is the right of individual citizens to choose, pursue, and amend their own ways of life. Though there is no absolute obligation to participate actively in the political life of one's community, all have the right to do so and those who do participate actively may be regarded as responsible citizens making a positive contribution to their community's life. All may contribute to the direction of society's affairs by the formal act of voting and through their expressed attitudes regarding their national and local governments' policies. Such contributions can only be made by morally and intellectually autonomous citizens capable of judging the moral validity of outcomes that are proposed and criticizing the factual and ideological premises by which they are supported (Bridges, 1997; Smith, 1997; Wringe, 1992, 2006).

A brief mention must finally be made of the way in which, in the context of higher and further education, autonomous learning may refer to the manner and organization of learning rather than its larger aims. By this stage it may, rightly or wrongly, be supposed that the essential core elements of a liberal education are in place and that what is studied will relate more or less closely to the diverse life choices of particular students. These may reasonably expect to be free to prepare themselves to follow this or that vocation or way of life and, in view of their various aspirations, resources, and commitments, may need to arrange their studies with a degree of flexibility. At this level too, it may often be an efficient use of time, expertise, and other resources for teachers to guide their students in the pursuit of their own learning rather than act as direct transmitters of educational content. In a world in which knowledge needs may be more diverse than can be served by the resources of individual institutions and the availability of new knowledge may advance rapidly, the ability to continue to learn outside periods and places of formal study is important. In this situation, however, the ability to continue to apply autonomous intellectual and, in some cases, moral scrutiny remains crucial.

Some Reservations Regarding Rational Autonomy and its Place in Education

Neither the view that reason alone can or should provide the motivating grounds for the actions of individuals, nor the central place ascribed to it in education have escaped criticism. Attention may be drawn to the view of Hume (1739) regarding the relationship of reason, emotion, and action. Far from being an essentially heteronomous element which should play no part in our practical reflections, Hume argues that it is the passions which determine our goals while the function of reason is merely to choose the most efficient means of achieving them. More recent writers have also drawn attention to the difficulty of bridging the gap between so-called grounding reasons which may indicate in rational terms what we ought to do and motivating reasons which ultimately lead us to act (Bond, 1983). One attempt to overcome this problem has been to posit a hierarchical model of autonomy according to which our first-order or everyday desires may be restrained by so-called second-order desires when these conflict with some other, to us, more important if less immediate desire (Frankfurt, 1971, 1999). Typically we come to recognize our second-order desires as a result of rational reflection. Critics of this view have, however, asked where the authority of second-order desires comes from and pointed out that recourse to a level of third-order desires would, of course, prefigure an infinite regress. It has also been suggested that so-called second-order desires, insofar as they are desires at all, are logically indistinguishable from those of the first order (Cuypers, 1992; Bonnett and Cuypers, 2003).

One of the problems in defending rational autonomy is that, at least in its pure Kantian form, it may seem to suggest that one course of action or way of life is indisputably rational and that others must therefore be less so, if not actually irrational. This, far from promoting individual freedom, would seem to leave the agent with very little scope for choice of action and suggest that an agent taking an alternative view was in need of restraint. It may even seem to open the way to a measure of coercion or moral bullying, if not to the excesses of so-called positive freedom referred to above. Critics of this interpretation of rational autonomy may point to the observation of Mill (1859) that as no one size or style of footwear fits all, it is even less likely that all will be suited by a single way of life.

In recent years postmodernists, communitarians, feminists, and virtue ethicists have all expressed reservations about the central role attributed to rational autonomy or its equivalent in political and moral as well as educational philosophy. The grand narratives of progress, justice and democracy, truth, and even reason itself no longer command the unquestioned acceptance they did from the philosophers of the Enlightenment. We no longer think that the problems of moral and political philosophy can be

solved by a single overriding theory analogous to the newtonian law of gravity, as Kant (1785), Hobbes (1642), and no doubt many other Enlightenment philosophers supposed. Access to the facts behind any general theory about the social or political world is difficult to obtain (Baudrillard, 1989; Fish, 1989) and these are, in any case, subject to interpretations, which may owe more to contingent relations of power than to reality if, indeed, it is meaningful to speak of reality at all. Bodies of argument or justification relying on such Enlightenment foundations as reason, truth, justice, and individual identity may be dismissed with such deconstructive terms as rationales, rhetorics, discourses, or phrase regimes (Foucault, 1978; Lyotard, 1979; Rorty, 1989).

Communitarians have also denied the very coherence of an intellectually and socially unencumbered individual autonomously developing, choosing, and evaluating a personal way of life, or even making important moral choices independently of the established values, practices, and traditions of his or her community (MacIntyre, 1981; Sandel, 1982). With this in mind one writer (Allen, 1982) goes so far as to claim that, far from securing a person's freedom, attempts to instill the practice of rational autonomy along the lines envisaged by some philosophers of education would risk entirely destroying it.

Many philosophers, including philosophers of education, no longer think of moral conduct as being arrived at by the application of moral rules deduced from certain higher-order principles which are themselves subject to rational scrutiny and validation. Good conduct may now often be seen as that which embodies certain virtues whose appropriateness or the appropriate manner of whose exercise will vary according to contexts of time, place, and tradition, the role of the agent in relation to particular others or more generally within his or her society (Slote, 1992; Carr, 1991). The view that inclination should have no place in our moral reflections may be unfavorably contrasted with the Aristotelian view (Aristotle *Nicomachean Ethics*) that the virtuous person is one who loves virtue and pursues it precisely because he or she is aware of its attractions.

The virtues of independence and rational decision making, especially where this is conceived of as the application of impersonal rules and principles, have also been identified as predominantly masculine traits (Griffiths and Smith, 1989). The Piagetian/Kohlbergian (and essentially Kantian) account of moral development according to which all individuals potentially develop through successively rational, or supposedly rational stages, culminating in the ability to make moral judgments in the light of principles which, at the highest level, are themselves subject to rational scrutiny, has been challenged by Gilligan (1982) among others. These writers argue that the mature stages of moral development in many women may take an alternative course leading to the different and equally

estimable, if not superior virtues of flexibility, cooperativeness, and the preservation of caring interpersonal relations. On this view, an education which accords supreme importance to predominantly masculine forms of development shows undue gender bias and undervalues women and their contribution to social, professional, and everyday life.

In Defense of Rational Autonomy as an Educational Aim

The hesitations considered in the previous section are valuable in problematizing the Enlightenment assumptions that reason alone can find unitary solutions to moral as well as scientific questions. It should be said, however, that they tend to focus on an extreme version of the notion of rational autonomy and may not be universally accepted as conclusive. Nor should it be thought that they provide convincing reasons to abandon the aim of encouraging and, as far as possible enabling, young people to become morally and intellectually independent. It may be felt that the theoretical assumptions which these criticisms presuppose are themselves vulnerable when taken to their logical conclusions. The following remarks, however, can do no more than indicate some of the more obvious counterarguments that may be advanced.

A life totally without inclinations, desires, and even passions, in which everyone's actions and words expressed nothing but the purest rationality would certainly be impoverished. As Kant recognized, we are in part emotional and physical beings. Unlike Kant, however, we recognize that this part of ourselves may enrich rather than damage the flourishing life. We also have more reason than Kant to be aware of the harm that may be wrought by the injudicious repression of this part of ourselves. Equally, however, we can scarcely deny the baleful effects both socially and upon individual lives if all yielded to their immediate desires and passions or fecklessly followed their passing inclinations. Little needs be said of the ills of over-ready acquiescence in the promptings of others or the uncritical acceptance of the beliefs it may suit others to foist upon us. It is not obvious that the disjunction between reason and desire is quite as clear as Hume and others have suggested or that the gap between grounding and motivating reasons is quite as unbridgeable as more recent writers have supposed unless, that is, the two terms are defined so tightly that reason cannot, in the nature of the case, give rise to action. Certainly, reasoned judgment can enable us to choose between conflicting desires and may sometimes equip us to see that a particular goal is more or less desirable than we previously supposed. It is also possible to speak of rational passion, as when we are strongly motivated to do something precisely because it seems

the right, proper, or most worthwhile thing to do. Of course, if it is then said that once we want to do something which has these properties, this then automatically becomes exclusively definable as a desire and therefore no longer a reason, then the conclusion that reason and desire are irremediably distinct is demonstrated, but vacuous.

The criticism that higher-order desires are logically indistinguishable from first-order ones appears to be based on a misunderstanding. Second-order desires do not provide the grounds for first-order ones in the way that higher-order principles are sometimes said to justify individual moral injunctions. Their function is, rather, to coordinate our various desires and enable us to choose between them, giving reasons to prefer some to others in particular contexts of time and place. Once again, of course, it is always possible, but vacuous, to claim that since so-called higher-order desires determine action they must necessarily be of a kind with other desires of the more immediate kind.

Hesitations arising from Mill's claim to see freedom as the ability to follow the promptings of our own nature, and to some extent also those suggested by the communitarians, are less troublesome to counter. It must be acknowledged that even if our essentially human nature resides in our rationality, it is nevertheless irrefutably the case that we inhabit different bodies, are possessed of different capacities, have different past histories, are in different situations, and have differing material resources. These are the givens of individual human lives which we may to some extent challenge as we challenge the uncritical assumption that they debar us from doing what we think is right or desirable to do, but in the end they must be taken into account when individuals decide how to conduct their lives. To fail to do so would not indicate autonomy but heteronymous perversity. Autonomy does not demand that we perform the impossible and what is possible will depend on circumstances.

We may readily accept the communitarian view that society, rather than pure reason, is the source of values in the light of which our moral decisions and lifestyle choices are made. The pluralistic modern world, however, offers us a vast range of values and possibilities of fulfillment. There is no predetermined fixity in the matter of who we choose to be or, to adopt MacIntyre's (1981) terminology, of what narrative our life forms a part, or to which tradition the life we live will eventually be seen to belong. Autonomy is the ability to critically assess the possibilities our society offers and to choose judiciously between them.

Postmodernist criticism of the disreputable origins of some of our values and assumptions and the problematizing of Enlightenment narratives are helpful in alerting us to the dangers of moral and political complacency. Some courses of action may nevertheless appear more

satisfactory than others, even if it is always possible that our major choices will later come to be seen as having been based on illusion. Insofar as this is the inevitable lot of humankind (if that is indeed the case) this would seem to be all the more reason for helping individuals to create or choose their own destinies and, as far as possible, to do so in the light of plausible, if provisional, justification. Though virtues are perhaps not to be identified by the application of universal rules and must owe much to one's community and local traditions and practices, they are, as Aristotle recognized, far from making the use of reason and rational choice entirely redundant. This is particularly evident when undoubted virtues appear to clash as when, for example, the virtue of personal loyalty appears to clash with those of truthfulness or public spirit. The autonomous individual may also have recourse to reason in order to distinguish the virtuous mean from the vices of deficiency or excess as his or her community defines them.

It may also be the case that in some societies including our own, the making of decisions and the application of rules and principles happen to be associated with individuals of one gender while, for whatever reason, flexibility, cooperativeness, and the preservation of caring interpersonal relations are more associated with those of the other. These capacities, however, are desirable in all human beings, irrespective of gender and courses of action involving flexibility, cooperativeness, and the preservation of good interpersonal relations may be as freely chosen in the light of reasoned reflection as decisiveness or the rigorous following of rules. It is at least questionable whether these two sets of qualities are to be regarded as innately different in persons of different genders rather than the result of differential socialization which as a society we may quite rationally determine to abandon.

The modern world is one which offers many lifestyles and is subject to rapid change; individuals may receive little guidance from the practices and traditions of earlier generations and are bombarded with information and persuasive claims, which they must critically assess in their own interests and as responsible citizens. Despite the various reservations mentioned above and others which may be advanced in the future, it is difficult, in such a world, to envisage reputable programs of upbringing and education in which developing a measure of moral and intellectual autonomy does not play a central role.

See also: Academic Freedom; Citizenship and Immigrant Education; Religious and Spiritual Education; The Aims of Education.

Bibliography

- Allen, R. T. (1982). Rational autonomy: The destruction of freedom. *Journal of Philosophy of Education* 16(2), 199–208.
- Aviram, A. (1986). The paradoxes of education for democracy or the tragic dilemmas of the modern liberal educator. *Journal of Philosophy of Education* 20(2), 187–200.
- Barrow, R. (1981). *The Philosophy of Schooling*. Brighton: Wheatsheaf.
- Baudrillard, J. (1989). *Selected Writings*, Poster, M. (ed.). Cambridge: Cambridge University Press.
- Berlin, I. (1969). *Four Essays on Liberty*. Oxford: Oxford University Press.
- Bond, E. J. (1983). *Reason and Value*. Cambridge: Cambridge University Press.
- Bonnett, M. and Cuypers, S. (2003). Autonomy and authenticity in education. In Blake, N., Smeyers, P., Smith, R., and Standish, P. (eds.) *The Blackwell Guide to the Philosophy of Education*, pp 326–340. Blackwell: Oxford.
- Bridges, D. (1997). Personal autonomy and practical competence: Developing politically effective citizens. In Bridges, D. (ed.) *Education, Autonomy and Democratic Citizenship*, pp 127–137. London: Routledge.
- Carr, D. (1991). *Educating the Virtues*. London: Routledge.
- Cuypers, S. (1992). Is personal autonomy the first principle of education? *Journal of Philosophy of Education* 26(1), 5–18.
- Dearden, R. F. (1972a). Autonomy and education. In Dearden, R. F., Hirst, P. H., and Peters, R. S. (eds.) *Education and the Development of Reason*, pp 448–465. London: Routledge and Kegan Paul.
- Dearden, R. F. (1972b). Education as a process of growth. In Dearden, R. F., Hirst, P. H., and Peters, R. S. (eds.) *Education and the Development of Reason*, pp 65–84. London: Routledge and Kegan Paul.
- Dearden, R. F. (1975). Autonomy as an educational ideal. In Brown, S. C. (ed.) *Philosophers Discuss Education*, pp 3–18. London: Macmillan.
- Dewey, J. (1916). *Democracy and Education*. New York: Macmillan.
- Fish, S. (1989). *Doing What Comes Naturally: Change, Rhetoric and the Practice of Theory in Literary and Legal Studies*. Durham, NC: Duke University Press.
- Foucault, M. (1978). *Discipline and Punish*, Sheridan, A. (trans.). New York: Pantheon.
- Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. In Frankfurt, H. G. (ed.) *The Importance of What We Care About*, pp 23–45. Cambridge: Cambridge University Press.
- Frankfurt, H. G. (1999). *Necessity, Volition and Love*. Cambridge: Cambridge University Press.
- Gilligan, C. (1982). *In a Different Voice*. Cambridge, MA: Harvard University Press.
- Griffiths, M. and Smith, R. (1989). Standing alone: dependence, independence and interdependence in the practice of education. *Journal of Philosophy of Education* 23(2), 283–294.
- Heathcote, G. M. (2000). Autonomy, health and aging: Transnational perspectives. *Health Education Research* 15(1), 13–24.
- Hirst, P. H. (1965). Liberal education and the nature of knowledge. In Archambault, R. D. (ed.) *Philosophical Analysis and Education*, pp 113–140. London: Routledge and Kegan Paul.
- Hirst, P. H. (1972). *Moral Education in a Secular Society*. London: University of London Press.
- Hirst, P. H. and Peters, R. S. (1970). *The Logic of Education*. London: Routledge and Kegan Paul.
- Hobbes, T. (1642). *De Cive*, translated as *The Citizen*. In Gert, B. (ed.) *Thomas Hobbes, Man and Citizen*. (1972). London: Harvester.
- Hume, D. (1739). *A Treatise of Human Nature*, Selby-Bigge, L. A. (ed.). (1968). Oxford: Oxford University Press.
- Illich, I. (1971). *Deschooling Society*. London: Calder Boyars.
- Kant, I. (1785). *Fundamental Principles of the Metaphysic of Ethics*, Abbott, T. K. (trans., 1962). London: Longmans.
- Kohlberg, L. (1968). Stage and sequence: The cognitive-developmental approach to socialisation. In Goslin, D. (ed.) *Handbook of Socialisation*, pp 94–113. New York: Rand McNally.
- Locke, J. (1689). Second treatise of government. In Laslett, P. (ed.) *Locke's Two Treatises of Government*, pp 299–478. (1963). Cambridge: Cambridge University Press.
- Lyotard, J. -F. (1979). *La Condition Postmoderne: Rapport sur le Savoir*. Paris: Éditions de Minuit.
- MacIntyre, A. (1981). *After Virtue*. London: Duckworth.

- Mill, J. S. (1859). *On Liberty*. (1913). London: Longmans Green.
- Neill, A. S. (1962). *Summerhill: A Radical Approach to Education*. London: Gollancz.
- Peters, R. S. (ed.) (1969). *Perspectives on Plowden*. London: Routledge and Kegan Paul.
- Peters, R. S. (1973). The justification of education. In Peters, R. S. (ed.) *The Philosophy of Education*, pp 239–268. London: Oxford University Press.
- Rorty, R. (1989). *Contingency, Irony and Solidarity*. Cambridge: Cambridge University Press.
- Sandel, M. (1982). *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press.
- Slote, M. (1992). *From Morality to Virtue*. New York: Oxford University Press.
- Smith, R. D. (1997). The education of autonomous citizens. In Bridges, D. (ed.) *Education, Autonomy and Democratic Citizenship*, pp 127–137. London: Routledge.
- White, J. P. (1970). Indoctrination: A reply to I.M.M. Gregory and R. G. Woods. *Proceedings of the Philosophy of Education Society of Great Britain* 4, 107–120.
- White, J. P. (1982). *The Aims of Education Restated*. London: Routledge and Kegan Paul.
- White, J. P. (1990). *Education and the Good Life*. London: Kogan Page.
- White, J. P. (2007). *What are Schools for and Why?* London: Philosophy of Education Society of Great Britain.
- Wringe, C. A. (1992). The ambiguities of education for active citizenship. *Journal of Philosophy of Education* 26(1), 29–38.
- Wringe, C. A. (2006). *Moral Education: Beyond the Teaching of Right and Wrong*. Dordrecht: Springer.

Further Reading

- Bridges, D. (ed.) (1997). *Education, Autonomy and Democratic Citizenship*. London: Routledge.
- Callan, E. (1994). Autonomy and alienation. *Journal of Philosophy of Education* 28(1), 35–54.
- Christman, J. (ed.) (1989). *The Inner Citadel: Essays on Individual Autonomy*. Oxford: Oxford University Press.
- Dworkin, G. (1988). *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.
- Frankfurt, H. G. (1999). *Necessity, Volition and Love*. Cambridge: Cambridge University Press.
- Hobbes, T. (1651). *Leviathan*. In Macpherson, C. B. (ed.). Harmondsworth: Penguin.
- Lindley, R. (1986). *Autonomy*. London: Macmillan.
- Winch, C. (2004). *Education, Autonomy and Critical Thinking*. London: Routledge.

Relevant Websites

- <http://www.plato.stanford.edu> – Autonomy in Moral and Political Philosophy, Stanford Encyclopedia of Philosophy.
- <http://www.plato.Stanford.edu> – Personal Autonomy, Stanford Encyclopedia of Philosophy.
- <http://www.warwick.ac.uk> – Teacher education for teacher-learner autonomy, The University of Warwick.

Bildung

J Masschelein and N Ricken, Katholieke Universiteit Leuven, Leuven, Belgium; Universität Bremen, Bremen, Germany

© 2010 Elsevier Ltd. All rights reserved.

Bildung could be regarded as one of the most central as well as controversial concepts in (German) educational and cultural discourse. Although there is no single notion of Bildung, it functions as a particularly modern paradigm that conditions specific theoretical perspectives as well as practical strategies, which differ strongly from Greek-antique *paideia* and Christian-medieval *imitatio* (but are not without parallels to liberal education). Considering this paradigm, we first indicate some general features of the concept of Bildung and its complicated history. However, we want to concentrate on the way in which the concept was part of various discourses. Starting from the question of which space of experience and thought is opened up by the notion of Bildung, we want to investigate the way in which the notion allows for educational phenomena to be shaped and to appear. Precisely because Bildung implies a certain relationship between the individual and itself, others and the world, and a certain idea about the dynamics of this relationship, we intend to ask how and in what way Bildung asks us to look at ourselves, to look at ourselves in relation with others and with the world, that is, how we have to think about ourselves. Concretely, we will elaborate the question about the strategic function of the notion of Bildung in five steps. We will first (1) give a historical and (2) semantic clarification, before we (3) sketch the space of experience and thought that is opened up by Bildung and indicate (4) briefly the use of the concept in present discourse. We conclude (5) by asking whether we should look at Bildung itself as being part of a particular power formation, whether the anthropological matrix offered by Bildung is obfuscating our radical vulnerability, and, finally, whether the space of experience it opened up is still ours.

The History of Bildung

Seen from a historical perspective, the concept of Bildung has had a remarkable and unique career. The concept was still very rarely used in the first third of the eighteenth century and when used, it was done so exclusively in relation to questions regarding origin and generation in the context of discussions on the issue of creation. However – albeit within a very different meaning and relatively late compared to the cultural and esthetic discourse (Horlacher, 2004) – by the middle of the eighteenth century, it was already very evident in educational discourse. By the end

of eighteenth century, it had turned into one of the key concepts of (educational) Enlightenment (Assmann, 1993). Originally coming out of a theological-mythical context and strongly connected to the notions of Bild (image), Gebilde (building, structure), Imago, and Forma (see Meyer-Drawe, 1999), the early notion of Bildung referred to the process of a becoming-in-time, of growing, and decay of creatures in the context of basic assumptions concerning creation. In the middle of the eighteenth century, the concept becomes more and more secularized and becomes related to the “allgemeine Emporbildung der inneren Kräfte der Menschennatur zu reiner Menschenweisheit” (“the general raising of the inner forces of human nature towards pure human wisdom”) as a still normative concept of development (Pestalozzi, 1780/2001: 32).

At the end of the eighteenth century, the concept of Bildung was developed as a central category for thinking about human beings (about what was called then: the destination of mankind – *die Bestimmung des Menschen*) which was, in particular, introduced through the work of Johann Gottfried Herder and Wilhelm von Humboldt. Based on the theoretical conception of *perfectibilité* as it was developed by Rousseau (1762/1969) in contrast to and as a transformation of the idea of perfection, Bildung was formulated as a critical principle of education to refuse specific and determining expectations of the old and new civil society. It indicated an epigenetic (cultural) structure, in contrast to a genetic conception in which the origin is linked with man's development: human beings are not determined by their own nature or their metaphysical and religious origin as creatures, but by their own practices. Consequently, it is not possible to specify any definite aim of human development. In fact, Bildung was given the endless task of developing, unfolding, and enlightening the human mind and making real the independence of human will and action from natural and social determinations, coercion, and constraints.

Within Wilhelm von Humboldt's conception of the “true meaning of men” as the “highest and most proportional edification of his (cf. human) forces into a whole” (“höchste und proportionierlichste Bildung seiner Kräfte zu einem Ganzen”) (Humboldt, 1792: 131), Bildung becomes a process of self-production through self-activated dealing with the world, implying that the process in which the individual gets his or her form can be understood above all as a work of myself (*Werk meiner selbst*) (Fichte) and is no longer to be seen as the realization of a pre-given

(innate) structure. From this perspective, *Bildung* is an intransitive term and has to be spelled as *Sich-Bilden* (self-construction and self-education) – with the consequence that each world appropriation (*Weltaneignung*) has to be reconsidered as a specific way of producing oneself (*Selbstervorbringung*) and vice versa. This double relation of the self and the world (*Wechselwirkung*) could be described as the task of involving oneself in the world (*Einarbeitung in Welt*) as a production of world (*Ausarbeitung von Welt*) by self-production (*Selbstervorbringung in Welt*), which means both to encompass the world in oneself (*die Welt in sich einzubilden*) and to establish and express oneself in the world (*sich in die Welt auszubilden*) (Humboldt, 1793/1960). In this way of conceiving *Bildung* as *Selbstbildung*, which, however, still needs the outer action through education that is itself bound to public state institutions in the concept of *Bildsamkeit* (Herbart, 1841/1997: 186), we get a clear rupture in the representations of development, in the sense that the human being is not any more just becoming what s/he was (as was still the idea with Goethe), but now became, through *Bildung*, what s/he made of her/himself. However, the horizon on which this rupture came about was, of course, still one in which the issue of who s/he should (ought to) be remained evident (cf. Hegel, 1807/1970).

It was precisely the enormous integrative and synthesizing force of the concept of *Bildung* which expressed itself in the reconciliations between opposites, which it allowed for and which had been inconceivable before – such as the opposition between *Eigentümlichkeit und Brauchbarkeit* (“the individual distinctiveness or identity and the social or economical usefulness”) (Niethammer, 1808/1968: 118), *Entwicklung und Bestimmung* (“development and destination”) (Pestalozzi, 1780/2001: 32), *Vernunft und Natur* (“reason and nature”) (Kant, 1803/1964: 697), *Wissen und Glauben* (“knowledge and belief”) (Schleiermacher, 1822/1988: 324ff.), *Individuellem und Allgemeinen* (“the individual and the general”) (Humboldt, 1960/1793: 235), and, finally, *Freiheit und Ordnung* (“freedom and order”) (Hegel, 1807/1970: 359ff.) – that makes clear why this concept could become one of the central concepts of Enlightenment and Modernity. In the Brothers Grimm’s classic *Deutsches Wörterbuch* (1854–1960), the word *Bildung* takes four meanings indicated by the Latin expressions: (1) *imago* (image), (2) *forma* (form), (3) *cultus anima, humanitas* (cultivation of the soul), and (4) *formatio, institutio* (edification or formation).

To understand the really spectacular process through which the *Bildung* discourse managed to establish itself as the main intellectual discourse in the late eighteenth century, we have to refer to three different contexts in which it played the role of guiding concept and of program draft (see Koselleck, 1990). In all three contexts, we can discern a process of destruction of the traditional order and – hand in hand – an increase and generalization of an experience of uncertainty and insecurity so that we

could say that increasing awareness of (radical) contingency as well as its possible compensation in the idea of progress constitutes the central characteristic of upcoming modernity.

First, in the context of a radical change in the generational constellation and relation between generations, and in contrast to the traditional (Christian) *imitatio* and the (Greek antique) *paideia*, *Bildung* replaces the traditional circle of life (in which the younger follow the elder) and establishes a new order of life (Rousseau) in which the elder generation is superseded in an endless process of higher development (*Höherentwicklung*, Humboldt) of the younger generation. According to this new dynamism and the general idea of progress, *Bildung* became the exclusive educational goal, which was articulated in the European wide establishment of the school system and marked a more or less paradoxical cultural and pedagogical orientation to facilitate self-determination by a disciplinary logic.

Second, in the context of a radical change in the political constellation, initiated by the French revolution and leading toward a new order in Europe based on nation states, *Bildung* was considered to be the royal road to the construction and reconstruction of the nation and of national culture. This was a complicated matter since *Bildung* was originally conceived as a critical and emancipatory enterprise, that is, as a process in which human beings became truly free and emancipated themselves from all kinds of power including the power of the actual given State. This is also true for Humboldt even though he believed for a time that this process could be organized by the State if it held back from intervening in the actual content of the process and even when, in fact, the process of *Bildung* was leading toward the establishment of a new (national-German) culture and a new (nation) State (Bollenbeck, 1994).

Third, in the context of a radical change in the cultural and religious constellation, in which not only traditional Christian representations of order lost their overall meaning but also (scientific) knowledge about the world gained increasing importance, *Bildung* became the notion under which the appropriation of knowledge was promoted and legitimized by the same token.

Although, as we have seen, the concept was introduced at the end of the eighteenth century, it was only toward the end of the nineteenth century that it became articulated as the ‘specific structure of education.’ Since Dilthey (cf. Dilthey, 1884/1986), it has occupied a central place in educational practice and thought, referring to the incorporation of human beings into society through their individual self-production and spiritual self-formation.

Semantic Field

The enormous range of meaning covered by the concept of *Bildung* has meanwhile given way to its tremendous

broadening (some say dissolving), although this did not lead to a totally different use or to its actual disappearance. On the contrary, the concept of Bildung seems to be omnipresent as much in daily language as in professional educational discourse, certainly in Germany, and now also increasingly in other countries and regions. Indeed, notwithstanding the fact that one of the decisive roots of Bildung has to be traced back to the English moral philosopher Shaftesbury (1671–1713) and his concept of ‘inner’ or ‘inward form’ which was translated by Herder into German as Bildung (Horlacher, 2004), it seemed that the notion was tied very strongly to the particular history of Germany, both as the history of the formation and development of the nation and as the history of the German thought and language (including the broader sphere of its influence in Western and Eastern Europe), and that there was no point to it outside Germany. However, in the last decade – and although there had been sporadic interest in the notion (and history) of Bildung – we witness increasing attention to Bildung in the Anglo-Saxon philosophy of education (e.g., special issues on Bildung in the *Journal of Philosophy of Education and in Educational Philosophy and Theory*). Strikingly in this context, Bildung, which some propose to translate as edification, as cultivation or ‘formation,’ has been related to various conceptions of liberal education (Løvlie and Standish, 2002) in an attempt to recover, strengthen, or renew its critical force with regard to actual developments in educational policy at all levels (as also discussion below). However, before going into this critical ambition, which has been associated with Bildung for a long time, let us return to its field of meaning.

Roughly stated, the notion of Bildung refers as much to the process in which the individual acquires its form as to the product of this process of formation of soul and spirit. If we have a closer look at both process and product, we can distinguish between six typological areas of meaning. (1) Bildung can refer to the body of knowledge which is generally mandatory (binding) in a certain time (the canon), including reflective knowledge. (2) The notion is also used to refer to qualifications and certificates obtained through a process of knowledge appropriation and acquisition. These are commonly seen as measures of Bildung and regulate very concrete opportunities of access to and participation in social, economic, and cultural life (educational privileges, so to say). (3) In addition to these two notions and as their necessary transformation, Bildung can indicate the disposition or attitude of the Gebildete (the educated), which is the product of this process and which has to be clearly distinguished from simply being an expert or somebody who disposes of a lot of knowledge. In this sense, Bildung could be spelled out as a habitus or an inner horizon of references, which remains when all objects of knowledge are forgotten. In the same direction, Adorno explains Bildung as the subjective inside

of the objective outside of culture (Adorno, 1959/1972). (4) The process character of Bildung itself, which has no end and is a process of continuous higher development (without a final destination) and permanent self-transformation. (5) The self-referentiality of this process, so that knowledge is not (and has not to be) appropriated for its own sake but for the sake of self-improvement and self-perfection – a process which can be measured by the increase of the individual degree of freedom. As a result, Bildung is sometimes reduced to a plea for a humanistic individualism and used as critically opposed to social and especially economic needs. (6) Finally, the hope that this appropriation of knowledge with regard to producing an increase of freedom also implies a raising of morality and ethics, so that Bildung is often related also to respect, prudence, and caution. How this hope, at least regarding its assumption of automaticity, has proved to be an illusion is shown by the developments of German fascism and its experiences of human destruction. Indeed, since World War II and taking into account that an important part of the leadership of Nazi-Germany could be called ‘Hoch-Gebildete’ (highly educated) in all of the meanings just mentioned, it seems that Bildung should be approached with suspicion (see Adorno, 1959/1972).

The Anthropological Matrix

Taking up the issue of the strategic function of the discourse of Bildung and asking how the figure of Bildung asks us to think about ourselves, our relations to others and the world, we can see how it offers a kind of matrix in which these relations are configured and prefigured. Two things have to be remembered here. (1) Bildung refers neither to a process of unfolding out of a prefigured inner self nor to a development process which would be directed by actions from outside. The term Bildung is used only when the intended process can be understood as a relational process – a process of self-production through appropriation of the world. This fits the meaning of Bildung as the inner part of culture, that is, the part of appropriation and acquisition of culture (Adorno, 1959/1972) and it echoes Herder’s translation of Shaftesbury’s concept of inward form. (2) Moreover, and very importantly, Bildung opens up a different experience of time in as much as it has largely contributed, in the historical context of the dissolving of foundational, timeless substances and essences, to a particular consciousness of our temporary existence. Indeed, on the one hand it increasingly opened up the forward side of the time axis – time became, seen and experienced in terms of progress without, however, a clear identifiable end of history. On the other hand, it carried, on the reverse side of the time axis, so to say, a historical consciousness which sorted out traditions and structured time experiences. In this it

bounded the openness of the future (as progress) back to the handed down history (as tradition). Indeed, this historical consciousness or self-understanding, implying an experience of being part of a historical process, with its promise of emancipation and progress (and of reconciliation of mankind) – notwithstanding that this process has its own obstacles and possible relapses, its own contingency, is one of the main features of the space of experience and thought opened up by the notion of *Bildung*.

Returning to the way in which *Bildung* makes up human beings or creates a way for people to be, we can understand *Bildung* as an anthropological matrix in a threefold way (Ricken, 2006): (1) As the formation of the self in relation to itself, in as far as the self sees itself to be set in a relation which it has to think and construct as to be constantly overcome or transcended. (2) As the formation of the social, that is, as a particular formation of the self in relation to others, in as far as the relationship to others is to be thought of in terms of neither commonality nor species, but on the horizon of the General (*Das Allgemeine*). (3) As being in relation to the world, which is not so much to be understood as the sustaining ground and the encompassing and endless transcending world, but as something which offers itself as an object, something which is before us, to be appropriated in view of our own ends – and which for educational reasons can be partitioned in disciplines and (teaching) subjects.

Therefore, the particular space of experience and thought opened up by the concept of *Bildung* can be sketched as a particular configuration of a power–knowledge complex (Foucault, 1975), which at once defines the self and produces it.

The Present Use of the Concept

Against this background, the contemporary use of the concept of *Bildung* is quite ambivalent. On the one hand, the notion of *Bildung* is still in wide use, especially in the public discourse on the meaning and function of education and the education system (including the educational sciences) within the currently emerging large-scale transformations of societies that are often described under the key term ‘globalization process’ and increasingly associated with the establishment of ‘learning societies.’ In this perspective, *Bildung* has become a key term in politics representing the different competences which seem to be required for survival in such a learning society. It is used and invoked by administrators and politicians to accompany and legitimate the deep transformation processes of (national) educational systems. On the other hand, the concept of *Bildung* is often neglected and refused as an antique and no longer regarded as a fitting idea because we seem to be moving in different spaces of experiences and thought. In particular, its supposed

normativity provokes distance and permanently leads to different concepts such as majority (*Mündigkeit*), identity, competence, or capability.

Additionally, in educational discourse, the meaning of the concept of *Bildung*, its utility, and its adequacy is increasingly contested (see Thompson, 2008). On the one hand, there are still traditional interpretations of *Bildung*, for example, firstly indicating the cultural content of education formulated as a specific canon or secondly sketching *Bildung* as a specific sort of self-experience and self-development summarized in the phrasing of *Sich-Bilden* or self-education. Although *Bildung* has become in some sense the key term of the educational sciences, there are many theorists in educational sciences who question its semantic and theoretical consistency and its practical as well as its political function.

According to Lenzen, for example, the concept has become superfluous and useless because, due to its dependency on the practical educational context in which it appears, it no longer allows us to create clarity or meaningful differentiation within the discipline and its theoretical discourse. He therefore argues that we should abandon the concept of *Bildung* and introduce the concept of autopoiesis or self-organization (Lenzen, 2000). In contrast to Lenzen, Tenorth pleads for maintaining the concept, since it would be the only concept which is sufficiently poly-referential and multivalent to locate the different disciplinary and nondisciplinary, and practical and professional pedagogical domains in relation to each other and since it could function as a leading concept for the sciences of culture in a broad sense (Tenorth, 2000). However, both Lenzen and Tenorth seem to forget that academic disciplines are themselves very specific projects whose interpretations penetrate deeply into our lives, that is, they seem to neglect the role of disciplines and their concepts in the development of power.

In fact, the concept of *Bildung* now occupies a lasting and prominent but ambivalent critical role. Gruschka’s article in one of the leading journals of the German educational scene, the *Zeitschrift für Pädagogik*, is an indication of this ambivalent role which is also present in the title: ‘*Bildung*: Unvermeidbar und überholt, ohnmächtig und rettend’ (‘*Bildung*: unavoidable and outdated, powerless and redeeming,’ Gruschka, 2001). Although Gruschka admits that the old idea of *Bildung* and of the *Gebildete* (the educated woman or man) has become actually outdated and dysfunctional, the same idea remains a central means for analyzing and criticizing what he diagnoses, following Adorno (cf. Adorno, 1972) and Heydorn (cf. Heydorn, 1979), as *Halbbildung* (half-edification) or *Unbildung* (un-edification). In this context, the concept of *Bildung* is used, for example, to criticize the concepts of learning present in the discourse on the learning society and also in criticism of many didactic and constructivist theories, accusing them of forgetting or misunderstanding the relationship

of knowledge to the world and the own rights and exigencies of that world, which are recognized in Bildung (this critical use parallels the way in which the notion 'liberal education' is often used for similar aims; see Løvlie and Standish (2002)). In fact, Gruschka's article is illustrative of a very common argument. The idea of Bildung here remains related to a kind of forgotten, negated, neglected, reduced, distorted, unfulfilled, or even not yet articulated promise which is, however, misused by some writers as a license for introducing differentiations of rank and distinction into education. Even though as Ruhloff (cf. Ruhloff, 1996) states, Bildung can be retained only as a nonnormative concept since absolute foundations and teleological ambitions have to be rejected, this does not show that we can conceive of Bildung without any normative connotation. The concept of Bildung always contains an understanding of what is human (an understanding of humanity) and how we can attain this. Bildung always implies an idea of humanity and a knowledge or representation of that which is unsatisfactory or insufficient. This idea of humanity which is contained in Bildung remains fundamentally oriented toward the choice of one's own form of life as an independent, emancipated, and self-directed life. It remains related to the Enlightenment ideas of autonomy and consensus (as the agreement between autonomous persons) since educated men, the *Gebildete*, are essentially autonomous persons. The ideal of Bildung can serve then as a point of resistance or a critical orientation which can be invoked against all kinds of distortion and negation of this independence and autonomy (in thinking and acting).

Against this background, and taking into account that Bildung meanwhile constitutes an omnipresent and culturally rooted interpretation scheme, it is not so surprising that the more recent reflections on Bildung neither want to simply revive traditional approaches nor to replace or abandon the concept. Rather they try to take up the tradition using its contradictions and ambivalences as an opportunity to problematize the paradoxical figure of learning and development processes (Schäfer, 1996; Meyer-Drawe, 1999; Thompson, 2008).

Some Questions

Even if it is not convincing to question the usefulness of the concept of Bildung in general, it could be productive to reflect on three questions. One could ask (1) whether we should look at Bildung itself as being part of particular power formation (rather than being the opposite of power), (2) whether the anthropological matrix offered by Bildung is obfuscating our vulnerability, dispossession, and deliveredness, and finally, (3) whether the space of experience it opened up is still ours or whether we are meanwhile part of a totally different space in which the

notion of Bildung is still present, but no longer effective with regard to our experience and thought.

1. Because the educational sciences in the broadest sense can be analyzed as truth games involved in the formation of actual power relations, we have suggested elsewhere that, starting from the analytical framework offered by Foucault, it is possible to analyze Bildung as a specific matrix of power and to reveal and unravel a complicity between Bildung (in practice and theory) and the birth of the modern subject (Masschelein and Ricken, 2003; Ricken, 2006). We could look indeed at both the history of Bildung and the history of educational discourse in which it is as central as being part of the history of the ways in which human beings conduct and govern themselves and others in the light of specific truth games. Here the notion of Bildung is regarded as connected to a certain kind of knowledge about and relation toward ourselves and others. The idea of Bildung then implies a very specific kind of interpellation, that is, a very specific way in which we are addressed as social beings and in which we are supposed to address ourselves and others. Bildung, with its epigenetical (cultural) and social connotation, appears as a social program formulated in a specific historical and social context in which it becomes the key term of bourgeois society. This program articulates itself as a transformation from the anthropological perspective (what kind of beings are we), from the social perspective (how, through which techniques and strategies are we related to the others), and both linked together in a transformation of the epistemological perspective (what kind of knowledge is possible and expected).
2. In line with this first analysis of Bildung as a specific modern matrix of subjectivation, it could be interesting to reconstruct the character of experience in the concept of Bildung. In contrast to the widespread conviction that Bildung is strongly linked with the idea of self-determination and self-assertion, there are some elements allowing a description of Bildung as a paradoxical process of individuation as self-consciousness on the one hand and self-alienation as a specific dispossession or exposition on the other hand. Following Foucault's explanation of experience as a paradoxical process of subjectivation and desubjectivation, Bildung could be reconstructed as a performative concept of transformation and transgression – including the existential problem of human vulnerability and deliveredness to others. In this perspective, Bildung means the fragile situation of being exposed to others, of experiencing one's own dispossession and of becoming different to oneself (see Thompson, 2008).
3. Finally, a totally different question is whether today the notion of Bildung is still able to give way to what we called the experience of a historical consciousness

(Simons and Masschelein, 2008). One might wonder whether today our space of experience and thought is first of all characterized by an environmental, that is, a spatial, consciousness (moving in networks) which means that Bildung now appears in terms of a resource or a (always temporary) competitive advantage in dealing with the demands of ever changing environments, rather than as an operation that situates us in history, that is, in relation to a tradition and in view of a future emancipation.

Bibliography

- Adorno, T. W. (1959/1972). Theorie der Halbbildung. In Adorno, T. W. (ed.) *Gesammelte Schriften: Soziologische Schriften I*, pp 93–121. Frankfurt am Main: Suhrkamp.
- Assmann, A. (1993). *Arbeit am nationalen Gedächtnis. Eine kurze Geschichte der deutschen Bildungsidee*. Frankfurt am Main: Campus.
- Bollenbeck, G. (1994). *Bildung und Kultur. Glanz und Elend eines deutschen Deutungsmusters*. Frankfurt am Main: Suhrkamp.
- Dilthey, W. (1884/1986). Grundlinien eines Systems der Pädagogik. In Misch, G. and Groethuysen, B. (eds.) *Gesammelte Schriften. Band 9: Pädagogik – Geschichte und Grundlinien des Systems*, pp 167–238. Göttingen: Vandenhoeck & Ruprecht.
- Foucault, M. (1975). *Surveiller et punir. Naissance de la prison*. Paris: Gallimard.
- Gruschka, A. (2001). Bildung: Unvermeidbar und überhohlt, ohnmächtig und rettend. *Zeitschrift für Pädagogik* **47**(5), 621–639.
- Hegel, G. W. F. (1807/1970). Phänomenologie des Geistes. In von Moldenhauer, E. and Michel, K. M. (eds.) *Werke*. Frankfurt am Main: Suhrkamp.
- Herbart, J. F. (1841/1997). Umriss pädagogischer Vorlesungen (2. Auflage). In Benner, D. (ed.) *Johann Friedrich Herbart – Systematische Pädagogik. Band 1: Ausgewählte Texte*, pp 185–196. Weinheim: Beltz/Deutscher Studien Verlag.
- Heydorn, H. -J. (1979). *über den Widerspruch von Bildung und Herrschaft. Bildungstheoretische Schriften (Band 2)*. Frankfurt am Main: Syndikat.
- Horlacher, R. (2004). Bildung – a construction of a history of philosophy of education. *Studies in Philosophy and Education* **23**, 409–426.
- Humboldt, W. v. (1960/1793). Theorie der Bildung des Menschen. Bruchstück. In Humboldt, W. v. (ed.) *Werke Bd. 1: Schriften zur Anthropologie und Geschichte*, pp 234–240. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Humboldt, W. v. (1792). Wie weit darf sich die Sorgfalt des Staates um das Wohl seiner Bürger erstrecken? *Neue Thalia* **1**(2), 131–169.
- Kant, I. (1803/1964). Über Pädagogik. In von Weischedel, W. (ed.) *Werke in sechs Bänden. Band 6: Schriften zur Anthropologie, Geschichtsphilosophie, Politik und Pädagogik*, pp 693–761. Frankfurt am Main: Insel.
- Koselleck, R. (1990). Einleitung – Zur anthropologischen und semantischen Struktur der Bildung. In Koselleck, R. (ed.) *Bildungsbürgertum im 19. Jahrhundert. Teil 2: Bildungsgüter und Bildungswissen*, pp 11–46. Stuttgart: Klett-Cotta.
- Lenzen, D. (2000). Bildung im Kontext. Eine nachgetragene Beobachtung. In Dietrich, C. and Müller, H. R. (eds.) *Bildung und Emanzipation. Klaus Mollenhauer weiterdenken*, pp 73–86. Weinheim: Juventa.
- Løvlie, L. and Standish, P. (2002). Introduction: Bildung and the idea of liberal education (*Special Issue: Educating Humanity: Bildung in Postmodernity*). *Journal of Philosophy of Education* **36**(3), 391–408.
- Masschelein, J. and Ricken, N. (2003). Do we (still) need the concept of Bildung? *Educational Philosophy and Theory* **35**(2), 139–155.
- Meyer-Drawe, K. (1999). Zum metaphorischen Gehalt von 'Bildung' und 'Erziehung'. *Zeitschrift für Pädagogik* **45**, 161–175.
- Niethammer, F. I. (1808/1968). Streit des Philanthropinismus und Humanismus in der Theorie des Erziehungsunterrichts unserer Zeit. In Hillebrecht, W. (ed.) *Niethammer: Philanthropismus – Humanismus. Texte zur Schulreform*, pp 79–445. Weinheim: Beltz.
- Pestalozzi, J. H. (1780/2001). *Ausgewählte Schriften. Studienausgabe mit einer Lebensdarstellung, Anmerkungen und Register*. In von Wilhelm, F. (ed.). Weinheim: Beltz.
- Ricken, N. (2006). *Die Ordnung der Bildung. Beiträge zu einer Genealogie der Bildung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Rousseau, J. -J. (1762/1969). *Émile ou De l'éducation*. Paris: Éditions Gallimard.
- Ruhloff, J. (1996). Bildung und problematisierender Vernunftgebrauch. In Borelli, M. and Ruhloff, J. (eds.) *Deutsche Gegenwartspädagogik, Band 2*, pp 148–157. Baltmannsweiler: Schneider Hohengehren.
- Schäfer, A. (1996). *Das Bildungsproblem nach der humanistischen Illusion*. Weinheim.
- Schleiermacher, F. D. E. (1822/1988). *Dialektik*. In von Odebrecht, R. (ed.). Darmstadt: Wissenschaftliche Buchgesellschaft.
- Simons, M. and Masschelein, J. (2008). From schools to learning environments: The dark side of being exceptional. *Journal of Philosophy of Education* **42**(3–4), 687–704.
- Tenorth, H. E. (2000). Bildung – was denn sonst? In Dietrich, C. and Müller, H. R. (eds.) *Bildung und Emanzipation. Klaus Mollenhauer weiterdenken*, pp 87–102. Weinheim: Juventa.
- Thompson, C. (2008). *Bildung und die Grenzen der Erfahrung. Randgänge der Bildungsphilosophie*. Habilitationsschrift der Martin-Luther-Universität Halle-Wittenberg.

Further Reading

- Bauer, W. (ed.) (2003). *Special Issue: Bildung. Educational Philosophy and Theory* **35**(2), 133–137.
- Biesta, G. (2002). Bildung and modernity: The future of Bildung in a world of difference. *Studies in Philosophy and Education* **21**, 343–351.
- Bruford, W. (1975). *The German Tradition of Self-Cultivation: Bildung from Humboldt to Thomas Mann*. Cambridge: Cambridge University Press.
- Cleary, J. and Hogan, P. (2001). The reciprocal character of self-education: Introductory comments on Hans-Georg Gadamer's address 'education is self-education'. *Journal of Philosophy of Education* **35**(4), 519–527.
- Kettler, D. and Lauer, G. (eds.) (2005). *Exile, Science, and Bildung: The Contested Legacies of German Emigre Intellectuals (Studies in European Culture and History)*. London: Palgrave MacMillan.
- Lichtenstein, E. (1971). Art. *Bildung. Historisches Wörterbuch der Philosophie*, pp 921–937. Basel/Darmstadt: Schwabe.
- Løvlie, L., Mortensen, K. P., and Nordenbo, S. E. (eds.) (2002). *Special Issue: Educating Humanity: Bildung in Postmodernity*. *Journal of Philosophy of Education* **36**(3).
- Masschelein, J. (2004). How to conceive of critical educational theory today? *Journal of Philosophy of Education* **38**(3), 351–367.
- Meyer-Drawe, K. (2007). 'Du sollst dir kein Bildnis noch Gleichnis machen ...'. Bildung und Versagung. In Koller, H. -C., Marotzki, W., and Sanders, O. (eds.) *Bildungsprozesse und Fremdheitserfahrung. Beiträge zu einer Theorie transformatorischer Bildungsprozesse*, pp 83–94. Bielefeld: Transcript.
- Ricken, N. (2000). In den Kulissen der Macht. Anthropologien als figurierende Kontexte pädagogischer Praktiken. *Vierteljahrsschrift für wissenschaftliche Pädagogik* **76**(4), 425–454.
- Schäfer, A. (2007). Bildungsprozesse. Zwischen erfahrener Dezentrierung und objektivierender Analyse. In Koller, H. -C., Marotzki, W., and Sanders, O. (eds.) *Bildungsprozesse und Fremdheitserfahrung. Beiträge zu einer Theorie transformatorischer Bildungsprozesse*, pp 95–108. 95–108.
- Thompson, C. (2005). The non-transparency of the self and the ethical value of Bildung. *Journal of Philosophy of Education* **39**(3), 519–533.

Critical Theory and Pedagogy

M Mason, Hong Kong Institute of Education, Hong Kong, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

In at least one sense it might be argued that any pedagogy worth its name is necessarily critical, for if it were not, it would be especially vulnerable to accusations of indoctrination, or it would, at best, be a form of pedagogy that intended pupils merely to become well versed or well drilled in the propositional claims or procedural skills identified with the subject being taught. However, the association of critical pedagogy with critical theory implies a particular understanding of the term critical, one associated most strongly with the Frankfurt School.

At the same time, it should be noted that not all critical pedagogy, critical practice, critical research, or critical discourse in education is associated with the critical theory of the Frankfurt School. The term critical is widely used (and abused) in the discourse of education, not least in examining why and how education fails, how education reproduces unjust economic relationships, how education reproduces state power and other unequal power relationships in society, and the role education plays in social and cultural reproduction. Comparatively little of this discourse, however, has much to do with the Frankfurt School, other than in the loosest sense or at a rather general level. Kincheloe (2008: 5 ff.), for example, includes the following as among the central characteristics of critical pedagogy, none of which could be claimed as necessarily derived from the Frankfurt School: it is “grounded on a social and educational vision of justice and equality, ... constructed on the belief that education is inherently political, ... dedicated to the alleviation of human suffering, ... [and is] a pedagogy that prevents students from being hurt.”

Some strands of this critical discourse may share Marxian roots with the Frankfurt School; others may claim a shared paradigm of conflict theory. Those that draw explicitly on the critical theory of the Frankfurt School constitute the major focus of this article.

Critical Theory and the Frankfurt School

The Institute of Social Research, from which the Frankfurt School developed, was founded at the University of Frankfurt in 1923. Its most active years, from 1930 (when Max Horkheimer assumed the directorship) to 1944, coincided with the rise to power and prominence of Nazism and Fascism: the Institute was accordingly moved to Geneva in 1933, and to Columbia University in New York in 1935. It was not until 1953 that it was properly

reestablished in Frankfurt. Horkheimer, Theodor Adorno, Herbert Marcuse, and Jürgen Habermas are perhaps its most notable members, but the importance of the contributions of the likes of Walter Benjamin, Erich Fromm, Leo Löwenthal, Franz Neumann, Friedrich Pollock, and Albrecht Wellmer should not be missed. The most influential early members of the Frankfurt School, Horkheimer, Adorno, and Marcuse, developed a form of Marxist theory known as critical theory, which, following Bottomore (1984), exerted a substantial influence in social theory on the study of class, politics, culture, and ideology. Critical theory also, following Held (1980), engaged with Marx's critique of political economy, Freudian psychoanalysis, esthetics, and the philosophy of history; contributed to the analysis of capitalism, the family, and the individual; and pursued new directions in epistemology and research methodology – their aim being “to lay the foundation for an exploration, in an interdisciplinary research context, of questions concerning the conditions which make possible the reproduction and transformation of society, the meaning of culture, and the relation between the individual, society and nature” (Held, 1980: 16). For Geuss (1981), the Frankfurt School's critique of ideology was central to these concerns, a question which will be considered shortly.

The Frankfurt School's primary intellectual sources lay in Kant, Hegel, Marx, Weber, Lukács, and Freud, although Habermas looked also at particular strands of Anglo-American thought, especially linguistic philosophy and the more recent philosophy of science. Kant and Hegel, and German idealism (and its critics) more generally, provided them with a better understanding of the philosophical foundations of Marx's work. Weber's work on the processes associated with rationalization and bureaucratization was important for their theorization of contemporary society. Freud's writings informed their investigations of human subjectivity. Importantly, it was Marx's early writing, especially the 1844 *Manuscripts*, to which they turned for an understanding of the critical grounds and orientation of his ideas. In Lukács and Korsch they found critiques of strongly determinist and positivist interpretations of historical materialism and its emphasis on the inevitability of historical development as the inexorable and logical expression of an apparently autonomous economic base of society. As Held (1980: 21), in this regard, puts it, the Frankfurt School maintained that “the traditional standpoint of orthodox Marxism ... fails to grasp the significance of examining both the objective conditions of action and the ways in which

these conditions are understood and interpreted. By underplaying human subjectivity and consciousness Marxists missed the very factors which were so central in preventing the emergence of a revolutionary agent.”

The point should not be missed, however, that the Frankfurt School were, while rejecting dialectical materialism, also critical of an overly humanist interpretation of Marx. As Horkheimer (1968: 202) wrote, “There can be no formula which lays down once and for all the relationship between the individual, society and nature. Though history cannot be seen as a uniform unfolding of human nature, the opposite fatalistic formula that the course of events is dominated by necessity independent of Man is equally naive.”

The Frankfurt School drew extensively on Lukács’s argument that the process of reification posed a substantial impediment to the development of revolutionary consciousness. Lukács had in turn drawn on Marx’s account of the fetishism of commodities in *Capital*, on Simmel’s analysis of the commodification of culture, and on Weber’s description of the process of rationalization to define a process of reification, in terms of which work becomes meaningless: the process and products of people’s labor or productive activity are construed as alien to them – as, in other words, existentially worthless. This definition of reification is closely related to the more common understanding of the term – the process whereby that which is socially constructed is perceived as natural, and therefore as objective, inevitable, permanent, and immutable – which would have been equally important to the Frankfurt School in understanding the ossifying processes contributing to the reproduction of the *status quo*.

The members of the Frankfurt School anticipated that the outcomes of their research would contribute “to the making of history with will and consciousness,” intending their findings “to become a material force in the struggle against domination in all its forms” (Held, 1980: 35). It is worth noting in this regard that Horkheimer, Adorno, and Marcuse viewed the exploitation typical of capitalist relations of production as only one, specific form of domination. The history of civilization reveals a history of domination in all its forms: “First, domination over one’s self, over one’s own nature. . . ; second, domination of the labour achieved by such disciplined and controlled individuals; and third, domination of outward nature, science and technology” (Marcuse, 1970: 12). This theme underlies and permeates many of the Frankfurt School’s main research projects, which, following Held (1980: 35), included questions such as: why European labor movements had not produced a united workers’ struggle; how the political and the economic were related in capitalist societies; why authoritarianism and bureaucracy were increasingly prevalent in society; how Nazism and Fascism had come to dominate Central and Southern

Europe with such apparently large-scale support; how social relationships and individual development appeared to be changing, possibly along with the development of a new type of ideology in the cultural sphere, which appeared increasingly open to direct manipulation (in this, a foreshadowing of a society dominated by consumerism); and whether, given the fate of orthodox Marxism, there existed a social agent capable of initiating progressive change toward an equitable socialist society.

These and other key research foci of the Frankfurt School were generated by their concern with instrumental rationality, a form of rationality focused more on technical than on normative questions, on the limitation of practical problems to technical questions. In its analysis they drew substantially, as indicated above, on Weber’s concept of rationalization. Enlightenment notions of reason had in their view become distorted in a modernity steeped in rationalization and bureaucratization into a form of rationality obsessed with means, with method and efficiency, rather than with ends and purposes. Important normative questions, such as “Why should?,” with their implicit humanist orientation, are in this process displaced by questions of procedure and technique such as “How to?”. Questions about values and interests are thus unacceptably, in the view of the Frankfurt School, separated from the production of the so-called objective fact. It is this view that underlies the Frankfurt School’s rejection of correspondence theories of truth and their critique of positivism and its associated belief that true knowledge is scientific and consists only of those statements which are objectively and empirically verifiable. Their concerns about the shaping of research practice in the human and social sciences according to a positivist model of the natural sciences widened to a critique of the ways in which scientific rationality had been extended to the conduct of life itself (Weber, 1972: 293), where a means–end rationality becomes a form of domination in itself, with means becoming ends in themselves, and with the consequent predominance of a rule-oriented disposition vitiating the possibility of critique. As Geuss puts it, a key aim of the Frankfurt School involved “the criticism of positivism and the rehabilitation of ‘reflection’ as a category of valid knowledge” (Geuss, 1981: 2).

Challenges to the critical theory of the Frankfurt School have been numerous, but frequently start from, or are related to, the criticism that critical theory remained too steeped in German idealism and engaged insufficiently with Marx’s materialist critique of political economy. Related to this is the charge that members of the Frankfurt School were more concerned with philosophical and theoretical problems and questions than they were with practical and political issues, and that they failed to connect in any integrated way their studies of individual and social psychology with material political economy and

societal critique. Closely connected is the criticism, from a harder Marxist perspective, that the Frankfurt School was too attentive to phenomena in the superstructure of society – culture and esthetics, for example – and insufficiently focused on factors associated with the economic base. What was almost worse is that Horkheimer and his colleagues were seen to be remote and far-removed from working-class politics, from the daily life experiences of ordinary people. These perceived shortcomings probably led to the visceral challenge Adorno suffered at the hands of his students at the University of Frankfurt in April 1969. Shortly after starting his lecture before almost 1000 students, he was challenged to engage in self-criticism, accused of defending capitalism, and confronted by three female students who bared their breasts, showered him with roses and tulips, and tried to embrace him. He left the theater in humiliation; within 4 months he had died of a heart attack. The event came to be known as the Busenaktion, or breast action, and typified the challenges offered by the increasingly praxis-oriented German New Left to the intellectuals whose theories lay behind their movement, but who were now perceived as too theoretically removed. It should be noted, however, that Marcuse, who did not return to Frankfurt when the Institute was reestablished there in the early 1950s but remained in the USA, continued to enjoy the confidence of the New Left in the 1960s and into the early 1970s. His sustained commitment to social and political struggle contributed to the esteem in which his *One-Dimensional Man* was held by the left.

Once a research assistant of Adorno's, Habermas has, since he first began extending and further developing the concepts of critical theory in the late 1950s, become the Frankfurt School's leading exponent. He himself understands his project to include: the development of a theory of society motivated by the normative purpose of the self-emancipation of its members from domination; a reinvigoration of the possibility of making history, in the words of one of his collaborators, Wellmer, "with will and consciousness" (Wellmer, 1974: 53); and the development of a critical theory that all but collapses the epistemological/axiological divide in its recognition of the inseparability of truth and virtue, of facts and values, and of theory and practice. In Wellmer's words, it is a project defined in terms of a "struggle for the critical soul of science" and "the scientific soul of criticism" (Wellmer, 1974: 53).

As intimated earlier by reference to Geuss (1981), the Frankfurt School's critique of ideology is central to these concerns, and it is at this point that consideration of their notion of *Ideologiekritik* most appropriately provides a context in which to understand a central strand of Habermas's thought. Perhaps the Frankfurt School's greatest interest in Marx's theory of society lay in its epistemological

implications, which underlie their critical theory, the essential distinguishing features of which are summarized by Geuss as follows:

1. Critical theories have special standing as guides for human action in that:
 - a. they are aimed at producing enlightenment in the agents who hold them, i.e. at enabling those agents to determine what their true interests are;
 - b. they are inherently emancipatory, i.e. they free agents from a kind of coercion which is at least partly self-imposed, from self-frustration of conscious human action.
2. Critical theories have cognitive content, i.e. they are forms of knowledge.
3. Critical theories differ epistemologically in essential ways from theories in the natural sciences. Theories in natural science are 'objectifying'; critical theories are 'reflective'.

A critical theory, then, is a reflective theory which gives agents a kind of knowledge inherently productive of enlightenment and emancipation. . . . The very heart of the critical theory of society is its criticism of ideology. Their ideology is what prevents agents in the society from correctly perceiving their true situation and real interests; if they are to free themselves from social repression, the agents must rid themselves of ideological delusion. (Geuss, 1981: 1–3)

The implications of this epistemology become clearer when contrasted with positivism's (a major target of critical theory, as indicated earlier) denial that theories could be both reflective and cognitive. While positivism may be useful in the natural sciences, its expansion into the human and social sciences could only, from the Frankfurt School's perspective, have a deleterious effect on that major vehicle of human emancipation, a critical theory of society and its potential to enable *Ideologiekritik* to that end. With this in mind, we can consider a central aspect of Habermas's thought – the relationship between knowledge and human interests.

In his major exposition of this theme in *Knowledge and Human Interests* (1971), Habermas suggests that "knowledge is neither a mere instrument of an organism's adaptation to a changing environment nor the act of a pure rational being removed from the context of life in contemplation" (Habermas, 1971: 197). Knowledge is constituted by our cognitive interests, because these orientations "shape and determine what counts as the objects and types of knowledge: they determine the categories relevant to what we take to be knowledge, as well as the procedures for discovering and warranting knowledge claims" (Bernstein, 1976: 192). Habermas's concern, in other words, is not merely epistemological: it is with the cognitive interests,

more broadly conceived than as in the interests of private individuals or those of politically motivated groups, which ultimately influence the constitution of knowledge. He identifies three primary cognitive interests: the technical, the practical, and the emancipatory, to which correspond three types of disciplinary fields: “The approach of the empirical-analytic sciences incorporates a *technical* cognitive interest; that of the historical-hermeneutic sciences incorporates a *practical* one; and the approach of the critically oriented sciences incorporates the *emancipatory* cognitive interest” (Habermas, 1971: 308). As Bernstein clarifies, “each of these cognitive interests is grounded in one dimension of human social existence: work, interaction, or power. Work corresponds to the technical interest which guides the empirical-analytic sciences; interaction, to the practical interest which guides the historical-hermeneutic disciplines; [and] power, to the emancipatory interest which guides the critical disciplines – the critical social sciences” (Bernstein, 1976: 193).

The empirical-analytic sciences, and the historical-hermeneutic sciences (which Habermas also describes as the “systematic *sciences of social action*, that is economics, sociology and political science” (1971: 310)) have, in Habermas’ view, the goal of producing nomological knowledge, the laws of nature. However, he asserts, “a critical social science . . . is concerned with going beyond this goal to determine (not only) when theoretical statements grasp invariant regularities of social action, . . . (but also, more importantly) when they express ideologically frozen relations of dependence *that can in principle be transformed* (emphasis added)” (1971: 310). A critical theory of society is then, according to Habermas, able to criticize a set of beliefs or worldview as ideological by showing, as summarized by Geuss:

- a. that the agents in the society have a set of epistemic principles which contains a provision to the effect that beliefs which are to be sources of legitimation in the society are acceptable *only* if they could have been acquired by the agents under conditions of free and uncoerced discussion;
- b. that the *only* reason the agents accept a particular repressive social institution is that they think this institution is legitimized by a set of beliefs embedded in their world-picture;
- c. that those beliefs could have been acquired by these agents *only* under conditions of coercion. (Geuss, 1981: 68)

Of key importance here are these conditions of free and uncoerced discussion. Habermas describes these conditions as the ideal speech situation, a transcendental position, based in his views about the preconditions for language use, that refers to a situation of “absolutely uncoerced and unlimited discussion between completely free and equal human agents” (translated by Geuss (1981: 65)

from Habermas and Luhmann (1971: 101ff., 113ff., 135ff.) and Habermas (1973: 252ff.)):

To be a human agent is to participate at least potentially in a speech community. . . . But no agent can be even potentially a member of a speech community who cannot recognize the difference between true and false statements. . . . But what it means for a statement to be true is that it would be the one on which all agents would agree if they were to discuss all of human experience in absolutely free and uncoerced circumstances for an indefinite period of time.

This ideal speech situation serves Habermas (translated by Geuss (1981: 66) from Habermas and Luhmann (1971: 139, 224))

as a transcendental criterion of truth, freedom and rationality. Beliefs agents would agree on in the ideal speech situation are *ipso facto* true beliefs, preferences they would agree on are rational preferences, interests they would agree on are real interests. The agents are free if their real situation is one which satisfies the conditions of the ideal speech situation.

The ideal speech situation is, according to Habermas, “neither an empirical phenomenon nor simply a construct, but a reciprocal supposition unavoidable in discourse, . . . a critical standard against which every actually realized consensus can be called into question and tested” (translated by Held (1980: 344) from Habermas (1973: 258)). It is the conditions of the ideal speech situation that critical pedagogs are frequently concerned to replicate as far as possible in their classrooms.

Critical Pedagogy

In Habermasian terms, the study of education is probably quite commonly understood as a historical-hermeneutic science incorporating a practical interest corresponding to the field of human interaction. However, from the perspective of critical theory, the study of education is conceptualized as a critical social science, incorporating an emancipatory interest focused on the distribution of power and its associated attributes: economic wealth, political influence, cultural capital, social prestige and privilege, and the like. Critical theory and pedagogy, in other words, are premised on the normative commitments associated with the exposure and transformation of, and the emancipation of humans from, oppressive economic, political, social and cultural formations, and the associated inequalities and injustices that are ubiquitous and pervasive in all domains of human interaction. Critical pedagogy is a term commonly used to describe the theorization and operationalization of the pedagogical challenges grounded in opposition to these formations.

Following Ellsworth (1989: 298), other terms used in the field include pedagogy of critique and possibility, pedagogy of student voice, pedagogy of empowerment, radical pedagogy, pedagogy for radical democracy, and pedagogy of possibility. Ellsworth's (1989: 300, 307) review of the literature reveals that critical pedagogy supports classroom analysis and rejection of oppression, injustice, inequality, the silencing of marginalized voices, and authoritarian social structures (see Fine, 1987; Giroux, 1986; Simon, 1987). Radical educators committed to critical pedagogy recognize and, in turn, help students to recognize and name injustice; they empower students to act against their own and others' oppressions (including oppressive school structures); and they criticize and transform their own conceptualizations in response to those of their students (see Giroux and McLaren, 1986; Shor and Freire, 1987). The goals of critical pedagogy include a critical democracy, individual freedom, social justice, and social change, realized not least through a revitalized public sphere which is characterized by citizens capable of confronting public issues critically through ongoing forms of public debate and social action (see Shor and Freire, 1987; Giroux, 1988). Students would thus be empowered by social identities that affirmed their race, class, and gender positions, and that provided the basis for moral deliberation and social action (see Liston and Zeichner, 1987). Such empowerment would be for human betterment (Parker, 1986: 227), for expanding "the range of possible social identities people may become" (Simon, 1987: 372), and for "making one's self present as part of a moral and political project that links production of meaning to the possibility for human agency, democratic community, and transformative social action" (Giroux, 1988: 68–69).

This review of the literature reveals the key role that Henry Giroux has played in the development of critical pedagogy. His *Theory and Resistance in Education: A Pedagogy for the Opposition* (Giroux (1983) is a key text in the field, not least in its solidification of the concept of a radical (theory of) pedagogy in the discourse in North America and more widely afield.

A key figure prior to Giroux in the emergence of critical pedagogy in the 1960s and 1970s is Paulo Freire (Freire 1970, 1973). His *Pedagogy of the Oppressed*, originally published in Portuguese in 1968 and in English in 1970, and his *Education for Critical Consciousness*, published in English in 1973, parts of which were originally published in Portuguese in 1969 as *Education as the Practice of Freedom*, are landmarks in the field of critical pedagogy. Freire shares the Frankfurt School's focus on the early Marx and the role of human consciousness and agency in contributing to social change; in addition, his work is influenced by existential and phenomenological themes, and by Catholic liberation theology. Freire was a Brazilian literacy educator who worked with rural peasants until the military dictatorship that took power of that country

in 1964, seeing him as a threat to wealthy landowners and to themselves, imprisoned and subsequently deported him. Well known for its critique of conventional forms of education as akin to banking, where teachers deposit knowledge into the passive minds of students only to withdraw it subsequently in examinations, Freirean pedagogy – problem posing and dialogical in nature – is probably synonymous with education as a process of conscientization, whereby unjust economic, political, social, and cultural relations, reified into a state of perceived permanence, as natural and inevitable, are raised to a level of conscious critical questioning in a process in which students learn to name their world and the factors contributing to their oppression. For Freire, literacy means both reading the word and reading the world. As Kincheloe (2008: 74) has drawn on Freire, through such a process of learning, students "read their reality and write their [own] lives." Freirean pedagogy is thus thoroughly political and inextricably engaged with the students' world, with their being in it, with the injustices, both explicit and implicit, which they suffer, and against which they learn to act. Freire's influence on the field has been such that McLaren (2000: 1) has referred to him as the inaugural philosopher of critical pedagogy.

Critical pedagogs, as indicated above, have frequently been concerned to replicate as far as possible in their classrooms the conditions of the ideal speech situation, where "all members have equal opportunity to speak, all members respect other members' rights to speak and feel safe to speak, and all ideas are tolerated and subjected to rational critical assessment against fundamental judgments and moral principles" (Ellsworth, 1989: 314). This commitment to dialog is a basic assumption of critical pedagogy, and has been posited as a necessary condition of democratic education and, in turn, a democratic society. In Giroux's words, in order for genuine dialog to be possible, students and teachers must show "trust, sharing, and commitment to improving the quality of human life" (Giroux, 1988: 72). Giroux and McLaren describe the critical educator's classroom as a form of public sphere in which "students and teachers can engage in a process of deliberation and discussion aimed at advancing the public welfare in accordance with fundamental moral judgments and principles School and classroom practices should, in some manner, be organized around forms of learning which serve to prepare students for responsible roles as transformative intellectuals, as community members, and as critically active citizens outside of schools" (Giroux and McLaren, 1986: 237).

Recent developments in the theory of critical pedagogy offer a description, following Kincheloe (2007: 21–24, and 2008: 50–59), of some of its current concerns and indicate subtle refinements in some of its key concepts. The notion of critical emancipation, for example, tends latterly to be used rather cautiously, following recent

acknowledgments that no one is ever completely emancipated from his or her context; the term is also frequently used with a skeptical orientation to the claim that this involves access to what some theorists refer to as Western reason. There have, in particular, been refinements in the conceptualization of power in the theory of critical pedagogy. Given that in contemporary society institutions such as the media and the architects of political ‘spin’ play an increasingly important role in widening acquiescence to inequitable distributions of power, wealth, and opportunity through the legitimation of socially constructed economic, political, social, and cultural relations by their depiction as natural and inevitable, critical pedagogy has become more cognizant of the importance of Gramsci’s (1971) notion of hegemonic power, associated with rule by consent of the ruled rather than by physical force. Contemporary theory of critical pedagogy has also become increasingly convinced of the accuracy of Foucault’s conceptualization of power: that, in its capillary-like nature, it is a fundamental and ubiquitous constituent of social relations, with both oppressive and productive capacity and consequences. Critical pedagogs are, further, all the more aware of the nuanced role that ideology plays in the establishment of hegemonic power relations. Far from understanding ideology as blunt, monolithic, and unidirectional propaganda, they realize that “ideological hegemony involves [at least] the cultural forms, the meanings, the rituals, and the representations that produce consent to the *status quo*” (Kincheloe, 2007: 23, 2008: 55). Also in a related vein, linguistic and discursive theories of power have become more prominent following the influence of post-structuralism: critical theory and pedagogy are now less convinced of the Habermasian position that language is, at least potentially, a vehicle for transcending ideologies and engaging in a free and equal exchange of communicative reason in an ideal speech situation, and more convinced by the positions of discourse theory that language is a socially shaped resource, steeped in culturally and historically sedimented values and assumptions, whose distillations of social meaning comprise multiple discourses, which in turn, produce, or become constitutive of, social reality, and legitimate particular discourses of power while excluding others.

As critical theory and pedagogy have become increasingly concerned with discourse and the analysis of discursive practices, so has their concern with critical hermeneutical analysis and interpretation, which are focused on revealing the dynamics of power and the normative hermeneutics – prescriptive norms of meaning and their legitimation – that constitute social and cultural texts. Critical hermeneutics might, for example, be concerned to understand why the attributes associated with a particular form of cultural capital, or with a particular form of symbolic violence (see Bourdieu, 1977a, 1977b; Bourdieu and Passeron, 1990), would carry the

meanings that they do. Related to Bourdieu’s concept of cultural capital, and following and further developing the Frankfurt School’s exploration of high culture, popular culture, youth culture, and the relationship of human culture to the natural environment (a theme of increasing importance in critical pedagogy, albeit very recently, thanks in most part to the work of Bowers, 1987; Bowers and Apffel-Marglin, 2004), the domain of culture is perceived in contemporary critical theory and pedagogy to be particularly important in understanding relationships of power and domination. The field of cultural studies has played an important role in this domain, which has taken on new levels of importance in media-saturated societies dominated by consumerism, the latter not least because of the contemporary ubiquity of consumer goods made possible over the last two decades in particular by the globalization of production (in turn made possible by huge differentials in costs across continents, by cheap containerization and transportation, and by the proliferation of information and communications technology). Culture is hence viewed as a domain of struggle over the production and dissemination of knowledge, since “dominant and subordinate cultures deploy differing systems of meaning based on the forms of knowledge produced in their cultural domain” (Kincheloe, 2007: 23, 2008: 56). Cultural pedagogy has accordingly come to play an increasingly important role in critical pedagogy. As cultural production disseminates particular forms of knowledge, socializes, and constructs identities (and contributes to both domestication – through the legitimation of particular hegemonic ideologies – and the potential for emancipation), so it is understood as a form of pedagogy. Cultural pedagogy refers here to the ways in which “dominant cultural agents produce particular hegemonic ways of seeing [in a] . . . corporate-dominated, media-based pedagogical process” (Kincheloe, 2007:24, 2008: 59). Hence, and again, in media-saturated societies dominated by consumerism, such cultural pedagogies have, of late, received increasing attention from critical educators.

This ongoing and lively debate in the field notwithstanding, the term critical remains, as indicated in the opening paragraphs of this article, widely used and abused in the discourse of education. Blake and Masschelein (2003) warn that the way in which the somewhat vague aim of critical competence, for example, has become a standard educational end and even a “standard prerequisite of operating and functioning within the given order” risks robbing critique of its “element of transcendence of the given order.” It has substituted a “positivist ameliorism for a critical awareness of the ‘intolerable’ experience of injustice” (Blake and Masschelein, 2003: 55). What is clear in the rather widespread trivialization and domestication of the terms critical and critique is that as long as what Blake and Masschelein refer to as “society’s anaesthetization to injustice” (2003: 54) remains an unfortunate

hallmark of contemporary social interaction, critical theory and critical pedagogy will remain important in education.

Bibliography

- Bernstein, R. J. (1976). *The Restructuring of Social and Political Theory*. Philadelphia, PA: University of Pennsylvania Press.
- Blake, N. and Masschelein, J. (2003). Critical theory and critical pedagogy. In Blake, N., Smeyers, P., Smith, R., and Standish, P. (eds.) *The Blackwell Guide to the Philosophy of Education*, pp 21–37. Oxford: Blackwell Publishing.
- Bottomore, T. (1984). *The Frankfurt School*. London: Tavistock Publications.
- Bourdieu, P. (1977a). *Outline of a Theory of Practice*. Cambridge: Cambridge University Press.
- Bourdieu, P. (1977b). Cultural reproduction and social reproduction. In Karabel, J. and Halsey, A. H. (eds.) *Power and Ideology in Education*, pp 487–510. Oxford: Oxford University Press.
- Bourdieu, P. and Passeron, J. (1990). *Reproduction in Education, Society and Culture*, 2nd edn. London: Sage.
- Bowers, C. A. (1987). *Elements of a Post-Liberal Theory of Education*. New York: Teachers College Press.
- Bowers, C. A. and Apfel-Marglin, F. (2004). *Re-Thinking Freire: Globalization and the Environmental Crisis*. New York: Erlbaum.
- Ellsworth, E. (1989). Why doesn't this feel empowering? Working through the repressive myths of critical pedagogy. *Harvard Educational Review* 59(3), 297–324.
- Fine, M. (1987). Silencing in the public schools. *Language Arts* 64, 157–174.
- Freire, P. (1970). *Pedagogy of the Oppressed*. New York: Continuum.
- Freire, P. (1973). *Education for Critical Consciousness*. New York: Continuum.
- Geuss, R. (1981). *The Idea of a Critical Theory: Habermas and the Frankfurt School*. Cambridge: Cambridge University Press.
- Giroux, H. A. (1983). *Theory and Resistance in Education: A Pedagogy for the Opposition*. South Hadley, MA: Bergin and Garvey.
- Giroux, H. A. (1986). Radical pedagogy and the politics of student voice. *Interchange* 17, 48–69.
- Giroux, H. A. (1988). Literacy and the pedagogy of voice and political empowerment. *Educational Theory* 38, 61–75.
- Giroux, H. A. and McLaren, P. (1986). Teacher education and the politics of engagement: The case for democratic schooling. *Harvard Educational Review* 56, 213–238.
- Gramsci, A. (1971). *Selections from the Prison Notebooks*. New York: International Publishers.
- Habermas, J. (1971). *Knowledge and Human Interests*. Boston, MA: Beacon Press.
- Habermas, J. (1973). Wahrheitstheorien. In Fahrenbach, H. (ed.) *Wirklichkeit und Reflexion: Festschrift für Walter Schulz*, pp 211–265. Pfullingen: Neske.
- Habermas, J. and Luhmann, N. (1971). *Theorie der Gesellschaft oder Sozialtechnologie – Was leistet die Systemforschung?* Frankfurt: Suhrkamp.
- Held, D. (1980). *Introduction to Critical Theory: Horkheimer to Habermas*. Berkeley, CA: University of California Press.
- Horkheimer, M. (1968). *Bemerkungen zur philosophischen Anthropologie* (Remarks on philosophical anthropology). *Kritische Theorie*, vol. 1. Frankfurt: S. Fischer Verlag.
- Kincheloe, J. L. (2007). Critical pedagogy in the twenty-first century. In McLaren, P. and Kincheloe, J. L. (eds.) *Critical Pedagogy: Where Are We Now?*, pp 9–42. New York: Peter Lang.
- Kincheloe, J. L. (2008). *Critical Pedagogy*, 2nd edn. New York: Peter Lang.
- Liston, D. P. and Zeichner, K. M. (1987). Critical pedagogy and teacher education. *Journal of Education* 169, 117–137.
- Marcuse, H. (1970). *Five Lectures*. Boston, MA: Beacon Press.
- McLaren, P. (2000). *Che Guevara, Paulo Freire, and the Pedagogy of Revolution*. Lanham, MD: Rowman and Littlefield.

- Parker, W. C. (1986). Justice, social studies and the subjectivity/structure problem. *Theory and Research in Social Education* 14, (Fall), 277–293.
- Shor, I. and Freire, P. (1987). What is the 'dialogical method' of teaching? *Journal of Education* 169, 11–31.
- Simon, R. (1987). Empowerment as a pedagogy of possibility. *Language Arts* 64, 370–382.
- Weber, M. (1972). *From Max Weber: Essays in Sociology*, Gerth, H. and Mills, C. W. (ed. and trans.). New York: Oxford University Press.
- Wellmer, A. (1974). *Critical Theory of Society*. New York: Seabury Press.

Further Reading

- Adorno, T. W. (1973). *Negative Dialectics*. New York: Seabury Press.
- Adorno, T. W., Frenkel-Brunswik, E., Levinson, D., and Sanford, N. (1950). *The Authoritarian Personality*. New York: Harper and Row.
- Althusser, L. (1969). *For Marx*. New York: Vintage Books.
- Apple, M. (1999). *Power, Meaning and Identity: Essays in Critical Educational Studies*. New York: Peter Lang.
- Apple, M., Au, W., and Gandin, L. A. (eds.) (2009). *The Routledge International Handbook of Critical Education*. New York: Routledge.
- Arato, A. and Gebhardt, E. (1978). *The Essential Frankfurt School Reader*. New York: Urizen Books.
- Benhabib, S. (1986). *Critique, Norm and Utopia: A Study of the Foundations of Critical Theory*. New York: Columbia University Press.
- Bourdieu, P. and Passeron, J. (1990). *Reproduction in Education, Society and Culture*, 2nd edn. London: Sage.
- Cole, M. (2007). *Marxism and Educational Theory: Origins and Issues*. London: Routledge.
- Darder, A., Baltodano, M., and Torres, R. (eds.) (2009). *The Critical Pedagogy Reader*, 2nd edn. New York: Routledge.
- Freire, P. (1970). *Pedagogy of the Oppressed*. New York: Continuum.
- Gibson, R. (1986). *Critical Theory and Education*. London: Hodder and Stoughton.
- Giroux, H. A. (1983). *Theory and Resistance in Education: A Pedagogy for the Opposition*. South Hadley, MA: Bergin and Garvey.
- Gramsci, A. (1971). *Selections from the Prison Notebooks*. New York: International Publishers.
- Gur-Ze'ev, I. (ed.) (2005). *Critical Theory and Critical Pedagogy Today: Toward a New Critical Language in Education*. Haifa: University of Haifa: Studies in Education Series.
- Habermas, J. (1970). *Toward a Rational Society: Student Protest, Science, and Politics*. Boston, MA: Beacon Press.
- Habermas, J. (1971). *Knowledge and Human Interests*. Boston, MA: Beacon Press.
- Habermas, J. (1976). *Legitimation Crisis*. London: Heinemann.
- Habermas, J. (1979). *Communication and the Evolution of Society*. London: Heinemann.
- Habermas, J. (1981). *The Theory of Communicative Action*. London: Beacon Press.
- Habermas, J. (1985). *The Philosophical Discourse of Modernity*. Cambridge, MA: MIT Press.
- Held, D. (1980). *Introduction to Critical Theory: Horkheimer to Habermas*. Berkeley, CA: University of California Press.
- Honneth, A. (2008). *Reification: A New Look at an Old Idea*. Oxford: Oxford University Press.
- Horkheimer, M. (1972). *Critical Theory: Selected Essays*. New York: Herder and Herder.
- Horkheimer, M. (1974). *Eclipse of Reason*. New York: Seabury Press.
- Horkheimer, M. and Adorno, T. W. (1972). *Dialectic of Enlightenment*. New York: Herder and Herder.
- Jay, M. (1973). *The Dialectical Imagination: A History of the Frankfurt School and the Institute of Social Research, 1923–1950*. Boston: Little, Brown.
- Kanpol, B. (1997). *Issues and Trends in Critical Pedagogy*. Cresskill, NJ: Hampton Press.
- Karabel, J. and Halsey, A. H. (eds.) (1977). *Power and Ideology in Education*. Oxford: Oxford University Press.
- Kellner, D. (ed.) (1989). *Critical Theory and Society: A Reader*. New York: Routledge.

- Kincheloe, J. L. (2008). *Critical Pedagogy*, 2nd edn. New York: Peter Lang.
- Marcuse, H. (1951). *Eros and Civilization: A Philosophical Inquiry into Freud*. Boston, MA: Beacon Press.
- Marcuse, H. (1964). *One-Dimensional Man: Studies in the Ideology of Advanced Industrial Society*. Boston, MA: Beacon Press.
- Marcuse, H. (1968). *Negations: Essays in Critical Theory*. Boston, MA: Beacon Press.
- Marx, K. (1969). *Early Writings*. New York: McGraw-Hill.
- McCarthy, T. (1978). *The Critical Theory of Jürgen Habermas*. Cambridge, MA: MIT Press.
- McLaren, P. and Kincheloe, J. L. (2007). *Critical Pedagogy: Where Are We Now?* New York: Peter Lang.
- Morrow, R. and Torres, C. A. (2002). *Reading Freire and Habermas: Critical Pedagogy and Transformative Social Change*. New York: Teachers College Press.
- Popkewitz, T. and Fendler, L. (eds.) (1999). *Critical Theories in Education: Changing Terrains of Knowledge And Politics*. New York: Routledge.
- Rasmussen, D. (ed.) (1996). *The Handbook of Critical Theory*. Oxford: Blackwell.
- Shor, I. (1992). *Empowering Education: Critical Teaching for Social Change*. Chicago, IL: University of Chicago Press.
- Wexler, P. (ed.) (1991). *Critical Theory Now*. New York: Falmer Press.

Relevant Websites

<http://www.tonywardedu.com>– Tony Ward's Resources for Critical Theory and Pedagogy.

Critical Thinking

H Siegel, University of Miami, Coral Gables, FL, USA

© 2010 Elsevier Ltd. All rights reserved.

This article draws extensively upon Siegel 2003 and Bailin and Siegel 2003.

Critical thinking enjoys a preeminent position among the many educational aims and ideals advocated by educators and educational theorists today. It has enjoyed this status since the earliest days of philosophical thinking about education, at least in the Western tradition originating in Greece nearly 2500 years ago. Although the expression 'critical thinking' is relatively new, the underlying ideal it names – often expressed in terms of the cultivation of reason or the fostering of rationality – has been regarded as a fundamental educational ideal by most of the historically significant philosophers of that tradition who turned their attention to education. No other proposed aim of education – caring, civic-mindedness, community, creativity, happiness, knowledge, obedience to authority, social solidarity, spiritual fulfillment, the fulfillment of potential, etc. – has enjoyed the virtually unanimous endorsement of historically important philosophers of education that critical thinking, reason, and rationality have (Siegel, 2003).

In contemporary discussions (e.g., Nussbaum, 1997; Siegel, 1988, 1997), critical thinking continues to be defended by many as an important educational aim or ideal. Unlike some historical predecessors, contemporary advocates of the ideal do not understand reason as a special psychological faculty; in defending rationality, they do not align themselves with the historical movement known as continental rationalism, according to which knowledge is based on the perception or intuition afforded by such a faculty. What then, exactly, is that underlying ideal? How is the ideal of critical thinking to be understood?

Critical Thinking: What Is It?

Key aspects of critical thinking, as currently advocated by contemporary theorists, include (1) the claim that the notion is essentially normative in character and (2) the claim that critical thinking involves two distinct components: both (a) skills or abilities of reason assessment and (b) the dispositions to engage in and be guided by such assessments. These are discussed in turn next.

Normativity

Advocates of efforts to foster critical thinking in schools sometimes conceive it narrowly, in terms of imparting

skills which will enable students to function adequately in their jobs, and in so doing to be economically productive. More often, however, proponents of the educational aim of critical thinking have in mind a broader view of critical thinking as more or less equivalent to the ideal of rationality.

So understood, critical thinking is a sort of good thinking. Therefore, the notion of critical thinking is fundamentally a normative one, thus distinguishing this understanding of critical thinking from those, common in psychology, which treat the notion as descriptive, identifying particular psychological processes (Bailin *et al.*, 1999). To characterize a given episode of thinking as critical is to judge that it meets relevant standards or criteria of acceptability, and is thus appropriately thought of as good. Most extant philosophical accounts of critical thinking, in addition to the account by Bailin *et al.*, emphasize such criteria. Robert H. Ennis, for example, defines critical thinking as "reasonable reflective thinking that is focused on deciding what to believe and do" (Ennis, 1987: 10), and offers a detailed list of abilities, skills, and dispositions which thinking (and thinkers) must manifest if it is (they are) to qualify as critical. Matthew Lipman defines critical thinking as thinking that facilitates judgment because it relies on criteria, is self-correcting and is sensitive to context (Lipman, 1991). Richard Paul analyses critical thinking in terms of the ability and disposition to critically evaluate beliefs, their underlying assumptions, and the worldviews in which they are embedded (Paul, 1990). Harvey Siegel characterizes the critical thinker as one who is "appropriately moved by reasons" (Siegel, 1988: 23), and emphasizes the critical thinker's mastery of epistemic criteria which reasons must meet in order to be rightly judged to be good reasons, that is, reasons that justify beliefs, claims, judgments, and actions. Other authors, including John McPeck (1981, 1990), similarly emphasize this normative dimension of the concept. While these authors' accounts of critical thinking differ in many respects, and have their own emphases, they are nevertheless agreed on its essentially normative character (Bailin and Siegel, 2003).

Skills/Abilities and Dispositions

While some early treatments of critical thinking defined it only in terms of particular skills – for example, Ennis' early

definition of it as “the correct assessing of statements” (Ennis, 1962: 83) – almost all more recent philosophical discussion of it (including Ennis’ more recent discussions) regards critical thinking as involving both (1) skills or abilities of reason assessment and (2) a cluster of dispositions, habits of mind, and character traits, sometimes referred to collectively as the critical spirit (Siegel, 1988). According to the advocates of this broader conception of critical thinking, education should have as a fundamental aim the fostering in students of (1) the ability to reason well, that is, to construct and evaluate the various arguments, and the reasons/premises and inferences of which they are composed, which have been or can be offered in support or criticism of candidate beliefs, judgments, and actions; and (2) the disposition or inclination to be guided by reasons so evaluated, that is, actually to believe, judge, and act in accordance with the results of such reasoned evaluations. Students (and people generally) are rational, or reasonable, or critical thinkers, to the extent that they believe, judge, and act on the basis of (competently evaluated) reasons. Thus, being a critical thinker is a matter of degree. To regard critical thinking as a fundamental educational aim or ideal is to hold that the fostering in students of the ability to reason well and the disposition to be guided by reasons is of central educational importance. These two aspects of the ideal deserve further comment.

The reason assessment component

Thinking is critical just to the extent that it manifests and reflects due attention to, concern for, and competence in assessing the probative strength of relevant reasons. In this respect, critical thinking can be understood as the educational cognate (Siegel, 1988: 32) of rationality, since both rational thinking and critical thinking are coextensive with the relevance of reasons (Scheffler, 1965: 107). Beliefs, judgments, and actions are rational just to the extent that the believer/actor has good reasons for so believing, judging, or acting; consequently, being able to think critically involves the ability to ascertain the epistemic or evidential goodness of candidate reasons. Consequently, a central task involved in educating for critical thinking is that of fostering in students the ability to assess the probative strength of reasons.

Any such account of critical thinking needs to be supplemented by an account of the constitution of good reasons which the proponent of the ideal is obliged to provide. How do we determine the degree to which a proposed reason for some belief, judgment, or action is a good or forceful one? What are the guidelines, or principles, in accordance with which the goodness of candidate reasons are to be ascertained? What is the nature of such principles? How are they themselves justified? Related questions arise concerning the criteria by which the goodness of candidate reasons is determined. How are these criteria chosen, and

who chooses them? How are they themselves justified – and indeed, can they be justified, even in principle, in a non-circular or question-begging way? What is the source of their epistemic authority? Are they absolute or relative? Are they really epistemic or rather political, constituting tools of power and oppression? These questions and others like them are epistemological in nature; they call for a general account of the relationship between a putative reason and the belief, judgment, or action for which it is a reason. Such an epistemological account will have to grapple with deep questions concerning the nature of epistemic justification, the relationship between justification and truth (and so the nature of truth), the relativity (or absoluteness) of principles of reason evaluation, and so forth. In this sense, the educational ideals of reason and rationality depend, for their own articulation and justification, on an adequately articulated and defended underlying epistemology. (For further discussion see Bailin, 1992, 1995, 1998; Siegel, 1988, 1989, 1997, 1998.) This also supplies a reason for thinking that epistemology should itself be taught in schools (Siegel, 2008).

The critical spirit

Having the ability to determine the goodness, or probative force, of candidate reasons for belief, judgment, or action may be necessary, but cannot be sufficient, for critical thinking, since a given thinker may have the ability but not (or not systematically or routinely) use it. Accordingly, most theorists of critical thinking argue that, along with the skill or ability to assess the probative force of reasons, critical thinkers must also have relevant dispositions. The primary dispositions are those of valuing good reasoning and its fruits, and of seeking reasons, assessing them, and governing beliefs and actions in accordance with the results of such assessment. In addition, most theorists outline a subset of dispositions or traits which are also necessary for critical thinking, including open-mindedness, fair-mindedness, independent-mindedness, intellectual modesty and humility, an inquiring attitude, and respect for others in group inquiry and deliberation (Bailin *et al.*, 1999; Hare, 1979, 1985). This two-component conception of critical thinking – according to which critical thinking encompasses both a reason-assessment component and a dispositional, critical-spirit component – is endorsed by most theorists.

The second aspect of the ideal – the disposition or inclination actually to be guided by the results of the reasoned evaluation of reasons – has broader philosophical implications. Here, the ideal recommends not simply the fostering of skills or abilities of reason assessment, but also the fostering of a wide range of attitudes, habits of mind, and character traits, thought to be characteristic of the rational or reasonable person (Scheffler, 1989; Siegel, 1988). This extends the ideal beyond the bounds of the

cognitive, for, so understood, the ideal is one of a certain sort of person. In advocating the fostering of particular dispositions, attitudes, and character traits, as well as of particular skills and abilities, the proponent of this educational aim denies the legitimacy, or at least the educational relevance, of any sharp distinction between the cognitive and the affective, or the rational and the emotional. The ideal calls for the fostering of certain skills and abilities, and for the fostering of a certain sort of character. It is thus a general ideal of a certain sort of person, which sort of person it is the task of education to help to create. This aspect of the educational ideal of rationality aligns it with the complementary ideal of autonomy, since a rational person will – at least ideally – also be an autonomous one, capable of judging for himself/herself the justifiedness of candidate beliefs and the legitimacy of candidate values.

Critical Thinking as a Fundamental Educational Ideal

As noted above, the cultivation of reason has been regarded by many philosophers of education in the Western tradition as a fundamental aim, and overriding ideal, of education. Today, the fostering of critical thinking (and so rationality) is often regarded in the same way. To so regard it is to hold that educational activities ought to be designed and conducted in such a way that the construction and evaluation of reasons (in accordance with relevant criteria) are paramount, throughout the curriculum. As Israel Scheffler puts the point:

Critical thought is of the first importance in the conception and organization of educational activities (Scheffler, 1989: 1).

Rationality . . . is a matter of *reasons*, and to take it as a fundamental educational ideal is to make as pervasive as possible the free and critical quest for reasons, in all realms of study (Scheffler, 1989: 62, emphasis in original).

The fundamental trait to be encouraged is that of reasonableness. . . . In training our students to reason we train them to be critical (1989: 142, 143).

To so take it is to regard the fostering of the abilities and dispositions of critical thinking in students as the prime educational directive, of central importance to the design and implementation of curriculum and educational policy. It is to hold that educational activities should be designed and conducted in such a way that the construction and evaluation of reasons (in accordance with relevant criteria) are paramount, throughout the curriculum. This is not to say that other aims and ideals might not also be of serious importance, but that none outranks the

primary obligation of educational efforts and institutions to foster critical thinking (Siegel, 1988: 136–137).

Justification of the Ideal

Why should the fostering of critical thinking be thought to be so important? There are at least four reasons for thinking so. First, and most importantly, striving to foster critical thinking in students is the only way in which students are treated with respect as persons. The moral requirement to treat students with respect as persons requires that we strive to enable them to think for themselves, competently and well, rather than to deny them the fundamental ability to determine for themselves, to the greatest extent possible, the contours of their own minds and lives. Acknowledging them as persons of equal moral worth requires that we treat students as independent centers of consciousness, with needs and interests not less important than our own, who are at least in principle capable of determining for themselves how best to live and who to be. As educators, treating them with respect involves striving to enable them to judge such matters for themselves. Doing so competently requires judging in accordance with the criteria governing critical thinking. Consequently, treating students with respect requires fostering in them the abilities and dispositions of critical thinking.

A second reason for regarding critical thinking as a fundamental educational ideal involves education's generally recognized task of preparing students for adulthood. Such preparation cannot properly be conceived in terms of preparing students for preconceived roles; rather, it must be understood to involve student self-sufficiency and self-direction. In this, the place of critical thinking is manifest. A third reason for regarding the fostering of critical thinking as a central aim of education is the role it plays in the rational traditions, which have always been at the center of educational activities and efforts – mathematics, science, literature, art, history, etc. All these traditions incorporate and rely upon critical thinking; mastering or becoming initiated into the former both requires, and is basic to the fostering and enhancement of, the latter. A fourth reason involves the place of careful analysis, good thinking, and reasoned deliberation in democratic life. To the extent that we value democracy, we must be committed to the fostering of the abilities and dispositions of critical thinking, for democracy can flourish just to the extent that its citizenry is sufficiently critical (Siegel, 1988, ch. 3).

These four reasons have been spelled out at greater length by several authors (e.g., Bailin, 1998; Robertson, 1995, 1999). They are sufficiently powerful to justify regarding critical thinking as a fundamental educational

ideal. Efforts to foster critical thinking aim at the promotion of independent thinking, personal autonomy, and reasoned judgment in thought and action; these particular aims are themselves in keeping with broader conceptions of knowledge, reasons, and persons: for example, that all knowledge is fallible, that it is possible to objectively evaluate the goodness of reasons, and that personal autonomy is an important value (Bailin, 1998: 204). These aims, and the broader conceptions in terms of which they are grounded, are philosophically contentious; it is no surprise, then, that they – and the educational ideal of critical thinking itself – have been challenged.

Criticisms of the Ideal

There are many extant criticisms and thoughtful critics of the ideal of critical thinking. Critics charge that critical thinking:

1. privileges the values and practices of dominant groups in society and devalues those of groups traditionally lacking in power;
2. privileges rational, linear thought over intuition;
3. is aggressive and confrontational rather than collegial and collaborative;
4. neglects or downplays emotions;
5. deals in abstraction and devalues lived experience and concrete particularity;
6. is individualistic and privileges personal autonomy over community and relationship; and
7. presupposes the possibility of objectivity and thus does not recognize an individual's situatedness (for details, discussion and references, see Bailin, 1995; Bailin and Siegel, 2003).

These criticisms, often made from feminist and/or post-modernist perspectives, must in the end be considered on their own merits. However, it is of considerable comfort to friends of the ideal that there is a general reply available to all attempts to reject the ideal, one that appears to be effective against them all, and which is manifested in the discussions of each of the specific critiques considered above. This reply, if successful, establishes the impossibility of rationally rejecting reason – and so preserves the legitimacy of regarding its cultivation as an educational ideal. This discussion concludes by rehearsing the reply and assessing its effectiveness.

Suppose that one wishes to reject the ideal of reason. One can reject it without thought or argument – indeed, one can reject it without ever recognizing or addressing the question of whether it should be rejected – or one can reject it on the basis of some reasoned challenge to it (e.g., that it fosters patriarchy, aids and abets oppression, depends upon a problematic individualism, rests on an inadequate conception of objectivity, or whatever). In the

former case, one's rejection does not threaten the legitimacy of the ideal, since no challenge is made. It is the latter, philosophical sort of rejection that genuinely challenges the ideal's cogency.

However, if such a challenge is made, it will be forceful, and successful, just to the extent that it is based upon good reasons for rejecting the ideal. The challenger is arguing, in effect, that there is good reason to reject the ideal of reason. Any such argument against reason, if successful, will itself be an instance of its successful application or execution. That is, the reasoned rejection of the ideal is itself an instance of being guided by it. In this sense, the ideal appears to be safe from successful challenge: any successful challenge will have to rely upon it; any challenge which does not cannot succeed. While challenges to the ideal might succeed in refining our understanding of it, none can succeed in overthrowing it. Thus, the ideal cannot be successfully challenged.

Transcendental arguments like this one are notoriously controversial philosophically; I cannot provide a general defense of them here. However, I should note that the argument does not prove too much. It does not suggest that other ideals are not important. Nor does it suggest that people cannot live contrary to it – although that they can (and do) does nothing to challenge the legitimacy of the ideal, or the claim that they ought to be guided by it. The argument obviously will not be persuasive to one who rejects reason, but offers no argument against it. But such an argument-less rejection fails as a critique, since it offers no criticism of the ideal or argument in favor of its rejection.

The proponent of rationality and its cultivation must, to be consistent, regard challenges to it as centrally important, and must regard the obligation to take such challenges seriously as integral to rationality itself. Insofar, deep criticisms of the ideal, and reasoned consideration of both its praiseworthy characteristics and its indefensible ones, are exactly what the ideal itself recommends. Whether the ideal survives extant criticisms will always be, in some sense, an open question; such criticisms may well succeed in altering our understanding of it. Nevertheless, there is a limit beyond which any proposed criticism of rationality cannot go without undermining itself. In so far, the ideal of rationality (at least in some formulation of it) cannot be coherently rejected (Siegel, 1996, 1997, ch. 5, Epilogue; 2003).

This self-justifying feature of rationality might itself be thought to provide some reason for regarding its cultivation as desirable. However, the main purpose served by the fundamental reply just rehearsed is not that of justifying the ideal directly – that task is more directly performed by the four reasons (respect, preparation for adulthood, initiation into the rational traditions, and democratic life) offered in favor of the ideal above. Rather, the main purpose is to make plain just how difficult it is to challenge the

ideal. Once the unchallengability of reason is clear, the desirability of its cultivation – on the basis of those four reasons (and quite possibly others as well) – is manifest.

Bibliography

- Bailin, S. (1992). Culture, democracy and the university. *Interchange* 23(1), 63–69.
- Bailin, S. (1995). Is critical thinking biased? Clarifications and implications. *Educational Theory* 45(2), 191–197.
- Bailin, S. (1998). Education, knowledge and critical thinking. In Carr, D. (ed.) *Education, Knowledge and Truth: Beyond the Postmodern Impasse*, pp 204–220. London: Routledge.
- Bailin, S., Case, R., Coombs, J. R., and Daniels, L. B. (1999). Conceptualizing critical thinking. *Journal of Curriculum Studies* 31, 285–302.
- Bailin, S. and Siegel, H. (2003). Critical thinking. In Blake, N., Smeyers, P., Smith, R., and Standish, P. (eds.) *The Blackwell Guide to the Philosophy of Education*, pp 181–193. Oxford: Blackwell.
- Ennis, R. H. (1962). A concept of critical thinking. *Harvard Educational Review* 32(1), 81–111.
- Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In Boykoff-Baron, J. and Sternberg, R. J. (eds.) *Teaching Thinking Skills: Theory and Practice*, pp 9–26. New York: W. H. Freeman.
- Hare, W. (1979). *Open-Mindedness and Education*. Kingston: McGill-Queen's University Press.
- Hare, W. (1985). *In Defense of Open-Mindedness*. Kingston: McGill-Queen's University Press.
- Lipman, M. (1991). *Thinking in Education*. Cambridge: Cambridge University Press.
- McPeck, J. E. (1981). *Critical Thinking and Education*. New York: St. Martin's Press.
- McPeck, J. E. (1990). *Teaching Critical Thinking*. New York: Routledge.
- Nussbaum, M. (1997). *Cultivating Humanity: A Classical Defense of Reform in Liberal Education*. Cambridge, MA: Harvard University Press.
- Paul, R. (1990). *Critical Thinking: What Every Person Needs to Survive in a Rapidly Changing World*, Binker, A. J. A. (ed.). Rohnert Park, CA: Center for Critical Thinking and Moral Critique.
- Robertson, E. (1995). Reconceiving reason. In Kohli, W. (ed.) *Critical Conversations in Philosophy of Education*, pp 116–126. New York: Routledge.
- Robertson, E. (1999). The value of reason: Why not a sardine can opener? In Curren, R. (ed.) *Philosophy of Education 1999*, pp 1–14. Urbana, IL: Philosophy of Education Society.
- Scheffler, I. (1965). *Conditions of Knowledge*. Chicago, IL: Scott Foresman.
- Scheffler, I. (1989). *Reason and Teaching* (first published by Routledge and Kegan Paul, 1973). Indianapolis, IN: Hackett.
- Siegel, H. (1988). *Educating Reason: Rationality, Critical Thinking, and Education*. London: Routledge.
- Siegel, H. (1989). Epistemology, critical thinking, and critical thinking pedagogy. *Argumentation* 3, 127–140.
- Siegel, H. (1996). Reason and rationality. In Chambliss, J. J. (ed.) *Philosophy of Education: An Encyclopedia*, pp 536–539. New York: Garland Publishing.
- Siegel, H. (1997). *Rationality Redeemed?: Further Dialogues on an Educational Ideal*. New York: Routledge.

- Siegel, H. (1998). Knowledge, truth and education. In Carr, D. (ed.) *Education, Knowledge and Truth: Beyond the Postmodern Impasse*, pp 19–36. London: Routledge.
- Siegel, H. (2003). Cultivating reason. In Curren, R. (ed.) *A Companion to the Philosophy of Education*, pp 305–317. Oxford: Blackwell.
- Siegel, H. (2008). Why teach epistemology in schools? In Hand, M. and Winstanley, C. (eds.) *Philosophy in Schools*, pp 78–84. London: Continuum.

Further Reading

- Bailin, S. (1987). Critical and creative thinking. *Informal Logic* 9(1), 23–30.
- Bailin, S. (1990). Argument criticism as creative. In Trapp, R. and Schuetz, J. (eds.) *Perspectives on Argumentation: Essays in Honor of Wayne Brockriede*, pp 232–240. Prospect Heights, IL: Waveland Press.
- Burbules, N. C. (1991). The virtues of reasonableness. In Buchmann, M. and Floden, R. E. (eds.) *Philosophy of Education 1991: Proceedings of the Forty-Seventh Annual Meeting of the Philosophy of Education Society*, pp 215–224. Normal, IL: Philosophy of Education Society.
- Burbules, N. C. (1995). Reasonable doubt: Toward a postmodern defense of reason as an educational aim. In Kohli, W. (ed.) *Critical Conversations in Philosophy of Education*, pp 82–102. New York: Routledge.
- Ennis, R. H. (1996). *Critical Thinking*. Upper Saddle River, NJ: Prentice-Hall.
- Ennis, R. H. (1998). Is critical thinking culturally biased? *Teaching Philosophy* 21, 15–33.
- Hare, W. (1987). Russell's contribution to philosophy of education. *Russell* 7, 25–41.
- Hare, W. (1995). Content and criticism: The aims of schooling. *Journal of Philosophy of Education* 29(1), 47–60.
- Kohli, W. (1995). Educating for emancipatory rationality. In Kohli, W. (ed.) *Critical Conversations in Philosophy of Education*, pp 103–115. New York: Routledge.
- Martin, J. R. (1992). Critical thinking for a humane world. In Norris, S. P. (ed.) *The Generalizability of Critical Thinking*, pp 163–180. New York: Teachers College Press.
- McClennan, J. E. Jr (1999). Keep hope alive. In Curren, R. (ed.) *Philosophy of Education 1999*, pp 19–21. Urbana, IL: Philosophy of Education Society.
- Norris, S. P. (ed.) (1992). *The Generalizability of Critical Thinking*. New York: Teachers College Press.
- Scheffler, I. (1960). *The Language of Education*. Springfield, IL: Charles C. Thomas.
- Scheffler, I. (1991). *In Praise of the Cognitive Emotions*. New York: Routledge.
- Siegel, H. (1992). Education and the fostering of rationality. In Talaska, R. A. (ed.) *Critical Reasoning in Contemporary Culture*, pp 89–112. Albany, NY: State University of New York Press.
- Siegel, H. (1999). What (good) are thinking dispositions? *Educational Theory* 49(2), 207–221.
- Talaska, R. A. (ed.) (1992). *Critical Reasoning in Contemporary Culture*. Albany, NY: State University of New York Press.
- Thompson, A. (1999). Who thinks like this? In Curren, R. (ed.) *Philosophy of Education 1999*, pp 15–18. Urbana, IL: Philosophy of Education Society.

Environmental Education

M Bonnett, University of Bath, Bath, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

While environmental education has been recognized as an important aspect of education for some time (the term itself perhaps dating back to a meeting of the International Union for Conservation of Nature and Natural Resources in Paris 1948), it is only comparatively recently that it has received sustained philosophical scrutiny. This is now establishing itself as an area of growing interest.

There are a number of reasons why the topic is attractive to philosophers of education. The first, shared by philosopher and non-philosopher alike, is the growing extent and urgency of what are perceived to be serious environmental problems affecting humanity both locally and globally. It is clear that these are becoming acute for citizens of the early twenty-first century and that education has a responsibility to address them. The goals, assumptions, and general approach of educational programs designed to do this therefore are of more than purely academic interest to philosophers. Second, some of the central ideas and assumptions that inform and orientate much current educational policy and debate regarding the environment are rapidly rendered problematic by philosophical examination. This is illustrated by reference to what is currently one of the most powerful orientating ideas in the arena of environmental education: sustainable development. Third, there is a very extensive and rich literature (including philosophical) on the environment to be examined and sifted for its educational significance. This is a major task. Fourth, the issues that are raised by environmental concern and environmental education are profound; they invite us to review a range of ideas and assumptions about the nature of knowledge, learning, and human well-being that have implications not simply for some discrete curriculum area or theme, but for the content and ethos of education as a whole. Indeed, many who have taken a strong interest in environmental education have wished to explore its potential as a vehicle for innovation in education, sometimes linking it with other reform movements such as school democratization and action research. Hence the clarification of key concepts and frameworks used in underpinning research, their relationship with learning theory more generally, and the adequacy of the measures to be employed in assessing the success of environmental education is also an important task (see Meyer, 2006).

The following sections identify some of the main ideas, strands of argument, and – most important – questions that characterize a growing field of educational debate.

General Issues Concerning the Aims of Environmental Education

Orr (1994) was perhaps one of the most influential in making the point that environmental concern raises not simply problems in education, but the problem of education – that is, conventional Western education that increasingly focuses on producing students who are effective operators in a global market economy premised on perpetual growth. For Orr, students so equipped, lacking ecological literacy, would perpetuate – indeed, render ever more rapacious – an economic system that is a heavy contributor to our current environmental predicament in terms of the strain being placed on our planetary ecosystem. This then raises a general point concerning the character of the aims of environmental education: they are not necessarily such as to sit comfortably within or alongside existing taken-for-granted practice.

A further issue arises from a practical impulse underlying environmental education: a tension between a desire on the part of many educational policymakers to achieve certain tangible behavioral outcomes that are taken to be environmentally friendly and a commitment to educate in the sense of developing a critical understanding of issues and possibilities. It is not merely that these demands will lead to different emphases in what is taught and expected, but that the independence of mind developed by the latter approach might lead to a questioning – and possible rejection – of the behaviors specified as desirable by the former. In environmental education, the issues are highly value-laden and not infrequently highly controversial. For example, not only are there serious tensions between environmental concern and capitalism as indicated above, but also what have been offered as environmental imperatives can conflict with key principles of liberal democracy, such as freedom and non-specification of the good life. (See Bell (2004) for some discussion of this.) Again, it becomes important for the philosophy of environmental education to address the issue of the general character of its aspirations. For example, should its focus be to convey information, to clarify competing ideas, to modify attitudes, or perhaps to inculcate some environmental ethic that parallels a widely accepted basic human ethic? In turn, such alternatives make wider assumptions about whether the solutions to environmental problems are to be achieved through technocratic authority or democratic participation.

They also raise questions about intergenerational learning and the role that children should play in addressing the

environmental situation. For example, should pupils be prepared to act as initiators of environmental education at home and in the community? This is an issue that environmental education raises in a particularly cogent form: social action is desperately needed, but should education be regarded as a vehicle for producing this? – as, indeed, arguably it has been on, say, the problems of racism, inequality, and economic competitiveness. By way of counter, it has been argued that not only is it unjust to transpose the responsibility for solving environmental problems from the generations that produced them to the generations that inherit them, but also that school activities must be evaluated on the basis of their educational value and thus according to educational criteria.

A school does not become 'green' by conserving energy, collecting batteries or sorting waste. The crucial factor must be what the students learn from participating in such activities, or from deciding something else

(Jensen and Schnack, 1997).

Finally, while education unavoidably needs to address the question of what our response should be to perceived environmental problems, it has been argued that at the most fundamental (and educative) level this means not merely addressing how we deal with discrete problems, but considering the underlying issue of what should be our relationship with nature. For this, it is held, pervades and conditions what we perceive as problems and sets the parameters for what can count as satisfactory solutions. Such views arise in the context of analyses that problematize ideas of nature and of the environment.

The Idea of the Environment

The term environment signifies a relational context: that in which something is located or embedded, the surrounding conditions in which it operates or has its being. Hence, for humans, it embodies the issue of our place in the world and the quality of our relationship with it. On such an account, our engagement with the environment necessarily will be manifold: sensuous as well as intellectual, esthetic as well as moral, involving affect and conation as well as pure cognition. Thus, environmental education would seem to refer to gaining some embodied understanding of ourselves in our surroundings, our world. Cooper (1992) has suggested that in many ways the idea of the environment is best understood as one's intentional field of significance: that web of entities and phenomena with which one has an intimate familiarity – the world in which one lives. This may be contrasted with the global environment and even one's geographically local environment – with which one might have scant familiarity. It might be argued that as necessarily the context in which personal understanding and action are

grounded, this lifeworld is where education has to start – and in a certain sense end – education must seek to enlarge and refine it if rather myopic and too partial environmental responses are to be avoided.

Some have seen this to raise afresh and in a forceful way the issue of place: the importance of locale to the significance of action, our sense of identity, and our sense of responsibility toward the environment. It is noted by environmental psychology that nothing we do is unplaced; hence, humans are always and already geographical beings. If place is an interactive environment that influences and responds to whatever is within it, places themselves can be understood as nodes in a nested holarchy. Our interactions with nature and place (our emplaced contingent existence) condition our worldview. Here again the importance of the lifeworld of the individual as the space where environmental phenomena are encountered and interpreted is emphasized.

In parallel with this, often a distinction is drawn between the built and the natural environment, and while the former may be an important educational topic – and itself inter-relates with the latter – ultimately, it is the issue of our relationship with the natural environment that is foregrounded by aspects of environmental concern that are currently experienced as most pressing, for it is the impact of human behavior on natural systems as evidenced by our concerns over things such as resource depletion, pollution, climate change, and wilderness and species extinction that motivates our present sense of a looming environmental crisis. Of course, addressing such issues has extensive implications for the built environment – especially if we were to include under this our social/political/economic environment – but the fundamental locus lies in our attitudes toward, and the impact of our behavior upon, the ecological systems that constitute our natural environment. Thus when we encounter educational talk of learning through, for, and in the environment, usually the implicit reference is to the natural environment.

But even if this delimitation is accepted we must be aware of the plurality of the natural world. Clearly, there are significant differences in the entities that compose it, both in terms of level (is the concern with the biosphere, an ecosystem, a species, a population, an individual, or a particular habitat or landscape?) and in terms of quality (such as human, conative, sentient, biotic, material), such that there would be corresponding differences in the way that each category is to be valued and treated.

Some of these issues have been taken up through the widely encountered idea of environmental or ecological literacy. Drawing on the broader literacy debate, Andrew Stables (1998) has identified three senses of environmental literacy: (1) a functional awareness of the scientific facts of nature (equivalent to decoding–encoding of text, understanding of literal meaning); (2) a received awareness of cultural significances and inputs regarding ideas and aspects

of nature and environmental issues; and (3) a capacity for critically evaluating personal meaning and underlying social/ideological determinants of issues in ways that can be translated into action.

As with a number of other views, this approach emphasizes the fact that environmental issues are human issues (or issues viewed from a human perspective) and therefore not only matters for natural science but for the arts and humanities. Its implicit suggestion that the environment can be regarded as a text that can be read has also been seen as opening the way to a full-blooded semiotic approach that problematizes the human–nonhuman boundary and leads to a questioning of aspects of the special position that humans frequently attribute to themselves.

To what extent is it appropriate to regard nature as a text? The question of the reality of nature goes to the heart of environmental education and is clearly highly philosophical in character. Is nature (merely) a social construct, redolent with ideological bias and therefore having no foundational reality? In an attempt to address this question, the point has been made that while of course, our concept(s) of nature are socially produced, we experience nature as precisely not socially produced. Its essential independence of human intention is central to its meaning, such that indeed its fundamental experiential quality is that of being self-arising (Bonnett, 2004: ch 5). It has been claimed that this experience of the autonomy – and hence the mystery – of nature is so deeply embedded in, and orientating of, our engagement with the world at large that it can in no way be regarded as optional or disposable as some postmodern views have suggested (Rorty, 1994: ch 8). Nature therefore can serve in a foundational role, though of course there is a constant need to be alert for the way in which naturalizing certain categories (such as man, woman, child) has the potential to authenticate, for example, questionable power relationships (Haraway, 1991, ch 8). Debate on such issues continues and is again illustrative of the very wide significance of the questions that are invited by a philosophical examination of the grounds of environmental education.

Furthermore, in emphasizing the central importance of our relationship with nature, the view both argues for seeking a systemic wisdom that transcends systematic scientific knowledge and alerts against naturalising environmental problems that lie not in nature (which has no problems) but in the character of our relationship with it. It therefore seeks to reveal the deep motives (such as mastery) that inform human practices in the modern and late modern era.

Education for Sustainable Development

In addition to frequently being taken to encapsulate what our response to the environment should be, the idea of sustainable development provides rich illustration of the

philosophical issues that are opened up for education. The notion first came to prominence with the publication in 1987 of the report of the World Commission on Environment and Development, *Our Common Future*. Here sustainable development was defined as a development that meets the needs of the present generation without jeopardizing the ability of future generations to meet their needs. This definition was widely taken up and was consolidated as an educational matter at the Earth Summit Conference held in Rio de Janeiro in 1992 attended by delegates from over 170 countries and whose centerpiece agreement was Agenda 21 which included the proposal to introduce sustainable development into the educational programs of signatory nations. Thus, it has found its way into the core curriculum of many nations.

Now the question is: how adequate a notion is sustainable development, so defined, as a basis for environmental education? Advocates make clear that it has the benefit both of a positive focus – on what we can do, rather than overweening gloom of ever-burgeoning problems – and in its recognition that human economic and material aspirations must be included in any proposed solutions. From the perspective of philosophy of education the question is in part modulated through three further questions: first, what does sustainable development mean?; second, how defensible are the values and assumptions it expresses?; and third, what are its educational implications? For some, pursuit of these questions has proved to reveal an approach to environmental education that is itself highly problematic and perhaps open to the objection that it risks undermining what should be at the very heart of any education that takes environmental concern seriously. In addition to problems concerning identifying non-paternalistic criteria for judging human needs both present and future in the context of a world population that is highly diverse geographically, culturally, and economically, there is the issue of an overweening anthropocentric stance. While some argue that allying concern for the environment with human self-interest is its best protection, others, fearing that the idea of development in sustainable development will inevitably be interpreted in accordance with Western market models (Shiva, 1992), point out its compatibility with economic and epistemological postures that may be at the very heart of current environmental problems – making sustainable development, so interpreted, a net contributor to them. Again, such arguments have led attention to the importance of our underlying attitude toward nature as an issue for environmental education, the educational focus moving from sustainable development as a policy to sustainability as a frame of mind (Bonnett, 2004: ch 9). Here, in somewhat Heideggerian vein, arguments have been mounted to suggest that understood as the place where things show up – are let be – authentic human consciousness is intrinsically involved in sustainability and that education

should be concerned to promote and nurture this by disrupting the metaphysics of mastery that currently holds sway in Western-style culture.

Knowledge, Behavior, and Action

One of the central questions facing environmental education is the kind of knowledge that it requires. Often, the philosophies underlying different approaches have emphasized qualities of knowledge that challenge much conventional educational practice.

The long-running OECD Environment in Schools Initiative (ENSI) is a good example of this. Its movement toward an ecologization of schools by placing environmental issues at the heart of the curriculum and its development of the notion of action competence, which involves students identifying and addressing environmental problems as a member of the school and wider community, requires that pupils not only acquire relevant factual knowledge, but that they personally evaluate issues and the relevance of knowledge, and learn the practical ability to work effectively with others in a way that respects the views of all concerned. Here, the aim is not the acquisition of pre-specified subject organized knowledge, but critically reflective environmental action framed in the context of students' own lifeworlds and understandings. In such a context, they are as likely to be the generators of knowledge as the recipients. This approach has been viewed by those involved as radically transgressive in the way it disrupts many boundaries that structure conventional education such as those between childhood dependency and adult responsibility, knowledge users and knowledge producers, knowing and acting, and facts and values. It is argued that the inherently complex, contextualized, frequently controversial, and often piecemeal occurrence of environmental issues in real-life situations precludes a traditional school curriculum and requires students to participate in shaping the social and economic conditions of their existence in society (Elliot, 1999).

This highly dynamic model of environmental education is clearly removed from the traditional model that seeks to transmit from above pre-decided environmentally good attitudes and behavior. It provides a powerful example of the extent to which environmental education might impact upon conventional views of pupils, teachers, and educational institutions. There is clearly much philosophical work to be undertaken in reviewing such possibilities, including the adequacy of a student rationality tacitly conditioned by the current motivational climate to identify key issues. This leads to two further points concerning the kind of knowledge to be valorized by environmental education.

The first of these is the importance and interpretation of systemic wisdom as an educational concern. A strong theme underlying much environmental debate

is that we must regard ourselves as thoroughly (though perhaps not exclusively) embedded in nature conceived as an ecological whole and hence that our actions reverberate throughout this whole having unlimited and often unintended consequences. Not infrequently, the whole visits us with unanticipated repercussions. On such an account, there is a strong need to think in a way that embraces this greater whole – and is open to it in a way that transcends our self-orientated purposiveness. But human consciousness is always selective. How is the greater whole to be understood? What are the appropriate metaphors? A variety of candidates are current: a causal system, an information flow and feedback (i.e., cybernetic) system, a realm governed by abstract laws, a domain of dialogical encounters attended by mystery. The helpfulness of any or all of these relates to the second point.

There has been an extensive debate in the literature concerning the traditional assumption that science offers the most authentic insight into the reality of nature. The central elements of this debate concern the aggressive and manipulative motives taken by some to be inherent in modern experimental science and the possibility that these are antipathetic to revealing the manifoldness of nature properly understood, such that other portals must be sought – a further argument for broad curricular involvement. Recognition of the need for more receptive-responsive, dialogical, forms of knowing reveals wide-ranging implications for knowledge transformation, curriculum, pedagogy, and the culture of the school. This has been interpreted as providing an entrée into educational debate for indigenous perspectives, romantic philosophy, as well as the more measured views of early environmentalists such as Thoreau, Emerson, Leopold, Muir, and ecofeminist points concerning *inter alia* knowing dominated by a masculine rationality (Plumwood, 1995).

In a different but not unrelated vein, the so-called co-evolutionary approach sees the relationship between society and nonhuman nature as one of ongoing reciprocal change, each impacting on the other in ways that can never be fully anticipated. Drawing on economic theory, Gough (2002) suggests that this perspective in which uncertainty is systemic invites a real-options approach to environmental planning. Here one attempts to calculate the option value of some aspect of the environment at different decision points over a given time period – the impetus being to think in a way that tends toward increasing the future option value of the environment and that is permanently prepared to be surprised by the way that things turn out. Education would need to develop the ability to conceive, recognize, and manage an ever-increasing range of possibilities and change perceptions so that future options that are systematically discounted in a political context that characteristically focuses on the short term can become more highly valued. It would need

to prepare all to contribute to this in their decision-making in the different fields and institutions in which they work.

Interestingly, such an approach invites a heuristic rather than economist interpretation of the much-vaunted learning society in which we allow the sense-giving frameworks that we employ in understanding the world to be open to change, “surrendering ourselves to the inescapable creativity and open-endedness of our grapple with the emergent” (Foster, 2002). In thus being radically open to life it will be radically precautionary in order to leave room for our fallibility and accommodation to the unexpected. These are some of the deep structural conditions for sustainability conceived as living with the grain of nature, recognized as a well-spring of surprises as much as an object of prediction.

The Social/Political Dimension

In addition to bringing pupils to an understanding of the sociopolitical values and practices that have been heavy contributors to the current environmental situation, it has been argued that there is an equally pressing need to address the sociopolitical consequences of these values and practices. An important cluster of these and their educational implications has been foregrounded through the notion of an eco-justice pedagogy. On the account developed by Bowers (2002), this has three main foci: (1) to develop an awareness of the environmental racism and class discrimination involved in the way that the deleterious environmental impacts fall disproportionately on ethnically and economically marginalized groups; (2) the recovery of non-commodified aspects of community through a reversal of an ever-increasing dependency on meeting life's daily needs through consumerism rather than through self-reliance within the family and within networks of mutual support within communities; and (3) to develop a sense of responsibility toward future generations and a corresponding self-limitation by an expansion of nonconsumptive relationships and opportunities to develop personal talents that enrich the community. Key to achieving these aims, for Bowers, is a recognition of the root metaphors embedded in language that shape the way that we understand the world. He argues that currently, the metaphors that hold sway are those that underlay the industrial revolution such as individual and linear progress and that these systematically undermine the value of tradition and therefore of intergenerational knowledge and continuity; they are in conflict with the root metaphor underlying an eco-justice pedagogy which is ecology and which foregrounds the relational and independent nature of our existence as cultural and biological beings.

On such a view, the educational aim that would have greatest impact would be to get students to see how language carries forward culturally specific ways of thinking

(orchestrated by root metaphors) and to reflect upon how the cultural and environmental patterns connect. In this way, a cultural space could be created for criticizing the language patterns of different Western cultures that create the individual psychology that accepts consumer dependency and environmental degradation as a necessary trade-off for achieving personal conveniences and material success.

Finally, there is a further important social cultural element to be addressed. Plausibly, it has been argued that politically sustainability has become exhausted. In the late-modern context – and two decades after *Our Common Future* – the pervasive dominance of liberal democratic consumer capitalism has left it as a piece of hollow rhetoric, and that in fact we now live in a post-ecologist paradigm of the politics of unsustainability (Bluhdorn, 2007). A hyper-environmentalism conceals the fact that the emphasis has shifted from life reform aiming to bring social practices into line with categorical, that is, non-negotiable, eco-imperatives to adapting these imperatives to systemic needs and lifestyle preferences which have themselves acquired the status of non-negotiability. Insofar as today, environmental education has been shaped largely as a response to concerns over environmental degradation and seeks an understanding that focuses on the sociopolitical sources of this situation, such an analysis of sociopolitical reality is of considerable relevance to its enterprise and the framing of its agenda.

Summary

In all there are many ways in which environmental concern leads to a questioning of the general anthropocentrism and Enlightenment humanism that dominates education, suggesting a need to consider the possibilities of post-humanist approaches. In this and in cognate ways considered above, a philosophical examination of environmental education prompts us to address the metaphysics of education as a whole.

Bibliography

- Bell, D. (2004). Creating green citizens? Political liberalism and environmental education. *Journal of Philosophy of Education* 38(1), 37–53.
- Bluhdorn, I. (2007). Hyper-environmentalism and post-ecologism: Eco-politics after the consensus on climate change. *Presented at Workshop Die Zukunft der Ökologiepolitik*. Institute for Human Sciences, Vienna, 29–30 November.
- Bonnett, M. (2004). *Retrieving Nature. Education for a Post-Humanist Age*. Oxford: Blackwell.
- Bowers, C. A. (2002). Toward an eco-justice pedagogy. *Environmental Education Research* 8(1), 21–34.
- Cooper, D. (1992). The idea of environment. In Cooper, D. and Palmer, J. (eds.) *The Environment in Question*, pp 165–180. London: Routledge.

- Elliot, J. (1999). Sustainable society and environmental education: Future perspectives and demands for the educational system. *Cambridge Journal of Education* **29**(3), 325–340.
- Foster, J. (2002). Sustainability, higher education and the learning society. *Environmental Education Research* **8**(1), 35–41.
- Gough, S. (2002). Increasing the value of the environment: A 'real options' metaphor for learning. *Environmental Education Research* **8**(1), 61–72.
- Haraway, D. (1991). *Simians, Cyborgs, and Women. The Reinvention of Nature*. London: Free Association Books.
- Jensen, B. B. and Schnack, K. (1997). The action competence approach in environmental education. *Environmental Education Research* **3**(2), 163–178.
- Meyer, R. (2006). Environmental learning: Reflections on practice, research and theory. *Environmental Education Research* **12**(3–4), 459–470.
- Orr, D. (1994). *Earth in Mind: On Education, Environment, and the Human Prospect*. Washington, DC: Island Press.
- Plumwood, V. (1995). Nature, self, and gender: Feminism, environmental philosophy, and the critique of rationalism. In Elliot, R. (ed.) *Environmental Ethics*. Oxford: Oxford University Press.
- Rorty, R. (1994). *Philosophy and the Mirror of Nature*. Oxford: Blackwell.
- Shiva, V. (1992). Recovering the real meaning of sustainability. In Cooper, D. and Palmer, J. (eds.) *The Environment in Question*, pp 177–193. London: Routledge.
- Stables, A. (1998). Environmental Literacy: Functional, cultural, critical. The case of the SCAA guidelines. *Environmental Education Research* **4**(2), 155–164.

Further Reading

- Bateson, G. (2000). *Steps to an Ecology of Mind*. Chicago, IL: University of Chicago Press.
- Bonnett, M. (2004). *Retrieving Nature. Education for a Post-Humanist Age*. Oxford: Blackwell.
- Bonnett, M. (2008). Systemic wisdom, the 'selving' of nature, and knowledge transformation: Education for the 'greater whole'. *Studies in Philosophy and Education* **28**(4), 39–49.
- Carr, D. (2004). Moral values and the arts in environmental education: Towards an ethics of aesthetic appreciation. *Journal of Philosophy of Education* **38**(2), 221–239.
- Chapman, R. (2007). How to think about environmental studies. *Journal of Philosophy of Education* **41**(1), 59–74.
- Cooper, D. and Palmer, J. (1998). *The Spirit of the Environment*. London: Routledge.
- Giddens, A. (1994). *Beyond Left and Right*, ch 8. Cambridge: Polity Press.
- Huckle, J. and Sterling, S. (eds.) (1996). *Education for Sustainability*. London: Earthscan.
- Luke, T. (2001). Education, environment and sustainability: What are the issues, where to intervene, what must be done? *Educational Philosophy and Theory* **33**(2), 187–202.
- McKenzie, M., Bai, H., Hart, P., and Jickling, B. (eds.) (2009). *Fields of Green: Re-Storying Education*. Cresskill, NJ: Hampton Press.
- Postma, D. W. (2006). *Why Care for Nature? In Search of an Ethical Framework for Environmental Responsibility and Education*. Dordrecht: Springer.
- Reid, A. and Scott, W. (eds.) (2008). *Researching Education and the Environment: Retrospect and Prospect*. London: Routledge.

Identity

M S Merry, University of Amsterdam, Amsterdam, The Netherlands; Nieuwe Prinsengracht 130, 1018 VZ, Amsterdam, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Published research and theory on identity far exceeds reasonable expectation. Thousands of articles and books have been written on the subject, yet navigating this topic continues to be something of a conceptual minefield. Like diversity, the term identity has been used in so many contexts, many of them polemical and political, that it is difficult to find any common understanding on which scholars agree. Culture typically has been the way that anthropologists have talked about identity, yet critics charge that anthropological models of identity too often proffer essentialist interpretations of their object, masquerading as objective description. Sociological models of identity, notwithstanding their bold structural claims and attention to environmental and institutional forces, are thought by some to have a thin empirical basis. Doubtless psychology has dominated the field of identity (and self) studies, yet psychological models, for all of their personal and interpersonal promise, are believed by many to have a weak theoretical foundation.

Meanwhile, identity in the philosophical literature has until recently been confined mostly to ruminations on the self: as soul in Plato; rationality connected to, but not coterminous with, the body in Descartes; uninterrupted consciousness in Locke; a stream of experiences which a thing has in relation to itself in Hume; an emotive life in Rousseau; as noumenal self about which we can know little in Kant; as will in Schopenhauer; as an elusive but nonspecific something in Wittgenstein; and as the continuous, purposive struggle of creative power against an historically contingent concatenation of social, political, and cultural forces in Nietzsche and Foucault (Glover, 1988; Perry, 1975; Marshall, 2001).

Yet what follows is not a summary of metaphysics or philosophy of mind. There also is not sufficient space here to examine identity in each of the separate disciplines. Instead, what follows are some prefatory discursive remarks about identity; then, owing to the dominance of psychological theories on identity, a brief overview of some of its salient themes will be adumbrated; this will be followed by a discussion on identity formation and maintenance both in political philosophy and philosophy of education.

Contemporary interpretations of identity run the gamut, ranging from essentialist readings found in the many varieties of culturalist discourse (e.g., aboriginal and indigenous studies), to postmodern dismissals of the term itself. Thus from the one end, essentialists ascribe inherent, lucid meanings to identity which lay beneath

social or political particularities. While essentialists appreciate that circumstances and historical contexts vary, they nevertheless insist that persons know, experience, and relate to something constant about themselves beyond or beneath the linguistic constructions or cultural narratives to which these merely give expression. Essentialist readings of identity, then, are reductionist inasmuch as some particular aspect of identity is believed to be irrevocably and timelessly true. This interpretation of identity assumes that there is something inalterable and determined about who one is on the basis of genealogy, culture, race, gender, or sexual orientation; as such, this interpretation risks erring on the side of identity politics.

Identity politics describes the fact that persons of various minority groups may feel trapped within an identity constructed by others, leading to the expectation that one must represent a particular set of ideals, interests, and behaviors in order to be authentic (see Merry and New, 2008). Essentialism is amply on display in the fields of psychology, sociology, and, perhaps most especially, in anthropology, which gave birth to the notion of cultural identity and difference. Yet the notion that one must identify exclusively with a specific identity marker (e.g., gender) – with all sorts of concomitant purposes and meanings in tow – usually comes at the expense of creating or exploring alternatives, or of critically examining how one's identity is both historically and continuously constructed. Here, Appiah (2005: 113) astutely observes that while the “contours of identity are profoundly real, [they are] no more imperishable, unchanging, or transcendent than other things that men and women make.”

Conversely, postmodernists argue that no such thing as a self exists; identity only describes indeterminate, mutable narratives of persons located within the social or cultural roles they are called upon to play. Subjectivities (in the preferred parlance) are contingent, malleable, hybrid assemblages that never cease evolving and adapting to any number of contexts. As Hall and du Gay (1997: 6) write:

Identities are, as it were, the positions which the subject is obliged to take up while always knowing (the language of consciousness here betrays us) that they are representations, that representation is always constructed across a lack, across a division, from the place of the Other, and thus can never be adequate – identical – to the subject processes which are invested in them.

For postmoderns, then, identity names little more than ephemeral positions and attachments we assume within

the various discourses that circumscribe who we are at any given moment. On this view, all identities – whether historical or cultural – are manufactured, disseminated, and politically deployed (Anderson, 2006; Hollinger, 1994). However, this approach renders insignificant the ways in which persons actually experience their identities, underneath it all, as substantive and real. That is, persons can sense that something about them exists beneath or outside of their cultural or social narratives; that it is they who are the authors of such stories; and that they can sense themselves as being in time and social space.

Indeed, however socially constructed identities are, one nevertheless feels herself to be Jewish, transsexual, black, or whatever. Of course there are myriad ways in which one is Jewish, transsexual, or black, and these markers of identity do not describe all that we are (the blunder of identity politics), but this point is frequently missed by critics of identity: how we see ourselves, and how we are seen by others, matters, and perhaps especially to those whose (constructed) identities enjoy little if any public recognition. The politics of recognition makes this very point: everyone should have the right to be acknowledged publicly for the identities that matter to them. Recognition is central to social anthropology in the context of ascribed and assumed identity. One is what one is seen to be, and one is as one believes oneself to be.

Both personal and social aspects shape identity but most theorists on the subject generally agree that identities do not evolve from pristine cultures; rather, identities are products of multiple, oftentimes conflicting, attachments. Indeed, the stable and fixed identity so favorable to earlier psychologists is increasingly being challenged by others who argue that our identities, far from being fixed and secure, are constantly evolving, adopting new meanings, and appropriating habits, customs, and beliefs according to contextual circumstance and need. Indeed, all persons possess hybrid identities that combine, mix, and separate identity components to adapt to different environments. Hybrid identities – what Salman Rushdie called the mongrel self – reflect the plural cultures and societies we live in, and we do not so much discard one identity for another so much as we interchange multiple, not entirely consonant, identities (Benhabib, 2002; Holland *et al.*, 1998).

In industrialized, Western societies, identities are understood as intensely personal and individual, stressing uniqueness and the independent self. Accordingly, the aim of exploring and questioning inherited values, opinions, and habits generally is encouraged. Conversely, identity in more traditional cultures – including, in East Asia, those which are highly industrialized – is typically circumscribed by one's age, caste, gender, or status. Accordingly, what some have called the interdependent self inclines toward similarity, conformity, and group harmony (Gross and Gore, 2003; Markus and Kitayama, 1991).

Of course, even describing these differences is a symptom of essentialism, and increasingly these cultural dichotomies are being challenged. It is likely that various levels of individual, relational, and collective identity are to be found in most persons irrespective of where they were born and how they were raised. Yet whether one is an independent or an interdependent self, Hogg (2003: 462) observes that group membership plays a crucial role in identity development:

Groups [influence] the type of people we are, the things we do, the attitudes and values we hold, and the way we perceive and react to people around us. Groups furnish us with an identity, a way of locating ourselves in relation to other people. Indeed, our sense of self derives from the groups and categories we belong to, and in many ways individuality may merely be the unique combination of distinct groups and categories that define who we are.

Identity Defined

Granting its manifold nuances, meanings, and applications, identity normally refers to the complex and ever-evolving expressions of self-understanding that describe how persons relate, and form attachments, to their historical–social–cultural environment over a lifetime and consciously or unconsciously arrange their priorities and commitments to reflect those, sometimes conflicting, attachments. Attachments are formed with other persons as well as one's environment (hence the attachment to one's homeland, for instance).

Identity denotes an understanding of who we are, as individuals, and not merely as products of history, culture, or caste. Josselson (1987: 12–13) describes identity as a “dynamic fitting together of parts of the personality with the realities of the social world so that a person has a sense both of internal coherence and meaningful relatedness to the real world.” One's identity is invariably multifaceted and may involve tension; identities, too, may be ranked (and reordered) in importance, some aspects being discarded or constructed in light of new and changing circumstances and experience. Sen (2006: 19) concurs:

Identities are robustly plural, and [one] identity need not obliterate the importance of others [...] a person [also] has to make choices – explicitly and by implication – about what relative importance to attach, in a particular context, to the divergent loyalties and priorities that may compete for precedence.

Some identities may be temporary, while others endure. Any number of different selves may be salient depending on the persons or environment with whom/which a person may be relating. Which aspects of identity one consciously or unconsciously summons will largely be determined by her context and circumstances, but also her temperament and emotional needs.

Yet, while identity is rarely settled or fixed, it vaguely describes how persons see themselves and how they express what is important to them. However variable one's self-concept is there remains a very particular sense in which persons understand themselves, and others, as specific, historical persons in space and time. Who one thinks she is, is unavoidably connected to the self that she was, is becoming, and will be, however little she remembers, understands, or is able to predict about those distinct phases of growth and development. This is consistent with what some psychoanalysts believe to be true of a healthy ego identity, namely, there must be a "subjective sense of continuity of being the same person over time and in different situations" (Côté and Levine, 2002: 121). This means that personal identity will manifest characteristics of having a past, a present, and a future, even if the present or future self no longer identifies as the same person she once was. This, then, is one of the great paradoxes of identity: experience changes who we are, and thus how we see the world as well as how we are seen by others.

Identity Formation

Identity formation has been studied from a number of perspectives. For example, there are models of racial identity formation (Cross, 1971), sexual identity formation (Cass, 1979), and minority identity formation (Atkinson *et al.*, 1983). Whatever the case, identity formation describes the conscious process of (re)examining one's feelings, thoughts, behaviors, and ways of relating to others who may or may not share similar commitments and habits. It is to reflect upon "our place in the universe, the meaning of life and death, and our purpose for being here" (Chickering and Reisser, 1993: 207). Identity formation also describes the way in which individuals deal with uncertainty and ambiguity.

Notwithstanding a variety of disciplinary approaches, contemporary studies on identity formation continue to assume a psychological cast, and none has towered over the field like Erik Erikson. Identity formation, for Erikson, names a process at the core of the individual and at the core of her communal culture, a process which establishes a single identity that links them together. "Identity grows and is nurtured or frustrated in a complex bonding of self and society" (Hoover *et al.*, 1997: 21). The formation of identity involves both competence and integrity. Competence is achieved by one's efforts and is validated by the recognition of others, while integrity is a state of mind about who one is in relation to oneself and others.

Identity for Erikson is both a personal and social construction, for there is a strong interplay between the psychic self and the social self. It is personal inasmuch as identity is

developed through the integration of various identifications with significant others and reference groups, and it is social inasmuch as it is developed through the internalized roles and appraisals of others. Central components of identity include: (1) a sense of personal continuity over time and across situations; (2) a sense of inner agency; (3) a commitment to certain self-representations as self-defining; (4) a commitment to certain roles as self-defining; (5) an acknowledgment of one's role commitments and views of self by significant others; (6) a commitment to a set of core values and ideal self-standards, and (7) a commitment to a worldview that gives life meaning (Erikson, 1959, 1968).

For Erikson, identity begins long before there is self-awareness, for it is in the nascent bonds of intimate relations (primarily with one's parents) that identity assumes its earliest expression. Intimate relationships, particularly the maternal bond with children, supply the mutual trust and recognition necessary for security and trust. Over time, and with consistent care and attention, it is within these trusting bonds that one comes to identify in a particular way with a set of attachments, habits, and thought patterns. This also describes the manner in which persons are enculturated, which is to say that most persons gradually come to identify with a way-of-being as natural, self-evident, and correct. The foundations of trust give rise to greater possibilities for personal well-being; well-being in this sense describes the experience of being accepted by others as well as a sense of security, satisfaction, and confidence about one's being-in-the-world. A healthy self-concept describes those who are comfortable with their self-image, with how others see them, with the roles they have chosen for themselves, or even which others have chosen for them. (Erikson's notions also supply the foundations for many sociological and anthropological understandings of what makes groups of individuals cohere.)

Yet an absence, or shattering, of foundational trust, for Erikson, portends a looming identity crisis. This crisis (which in late adolescence he describes as identity diffusion) frequently results from persistent doubts about one's ethnic, gender, religious, sexual, or racial identity, and describes the inability to resolve a profound personal challenge when faced with it. On the other hand, another type of crisis known as moratorium, is one that persons must navigate in order to achieve identity at each stage of development. Either way, identity crises arise when the acceptance of one's identity is questioned or rejected by oneself or others. Young persons are particularly susceptible to peer pressure; anxious to be accepted as a group member, intolerance may be expressed towards non-members, that is, outsiders. In Erikson's words:

It is difficult to be tolerant if deep down you are not quite sure that you are a man (or a woman), that you will ever [be] attractive, that you will be able to master your drives,

that you really know who you are, that you know what you want to be, that you know what you look like to others, and that you will know how to make the right decisions without, once for all, committing yourself to the wrong friend, sexual partner, leader, or career (Erikson, 1959: 93).

If the crisis proves too difficult to overcome, some kind of identity pathology may set in (e.g., narcissistic personality disorder), these normally being the “result of impaired ego functioning caused by a functional or organic disorder” (Côté and Levine, 2002: 154). The result, for psychoanalysts, is that a failure to thrive at any level represents a failure to successfully navigate a much earlier conflict (e.g., trust vs. mistrust, autonomy vs. doubt) so that a personality disorder is the result of a more primary problem with attachment that is manifested in that period when intimacy is the primary demand for growth, that is, adolescence and early adulthood.

Marcia (1966) extended the work of Erikson by examining the various ways in which identity formation occurred in adolescence. Marcia, however, was less certain that identities were either resolved or confused. The crisis that Erikson described was not, for Marcia, an emergency but rather a stimulus for overcoming challenges and therefore leading to growth. The result would be greater individualization and differentiation. Marcia also concurred with most psychologists on this point: identity arises from any number of competing – and not necessarily conflicting – influences. Though not an exhaustive list of identity-forming effects, birth order, peer group, gender assignment, sexual orientation, religion, occupation, and culture all influence, to one degree or another, how one comes to understand her relationship to others, including the way in which one shares a set of communally based commitments or practices.

Needless to say, identities will not arise within, or map onto, neatly prescribed categories. Côté and Levine (2002: 46) observe:

[In] order to understand an individual's personal identity, one needs to know more about a person than his or her ego constructions and sociohistoric location and opportunities – one needs to know about the emergent interpersonal circumstances affecting his or her behavior, including others' perceptions of past personal-identity displays, labels that might have been affixed to him or her by others, prejudices faced, gossip relevant to the person, multiple and contradictory pressures to conform, and so forth.

Perhaps one truism seldom noted by psychological identity theorists is that many identities also are formed in resistance to peer or parental expectations, as well as to one's inherited group identity or membership. Further, preferring an interdependent and relational model of

identity formation, some feminist critics (Gilligan, 1982) of Erikson – and moral development theorist Lawrence Kohlberg – have objected to their purported androcentric preference for autonomy and independence as signs of identity achievement.

Philosophical Trends

In political philosophy, communitarians have long argued that personal identity is entirely dependent on one's inherited cultural context. Parekh (2000: 159) expresses this well:

[C]ulture gives coherence to our lives, gives us the resources to make sense of the world, stabilizes our personality, and so on. Its values and ideals inspire us, act as our moral compass, and guide us through life; its arts, rituals, songs, stories and literature fill us with joy and add colour and beauty to our lives; and its moral and spiritual wisdom comforts and helps us cope with the inevitable tragedies of life.

Some cultural identities are recognized and protected by dominant beliefs and practices and enjoy a level of identity capital that those in minority groups do not. Identity capital describes the various investments that persons make about who they are, but the term also expresses a series of transactions and exchanges with others that validate or invalidate one's personal or social identity. Yet where identities appear to be under threat owing to a weakened cultural base, well-being itself appears threatened to some, thus leading to efforts to protect and maintain (fixed) cultural identities. Some have even argued that the state ought to underwrite the protection of certain imperiled cultures in order that the identities of their members may remain intact (Margalit and Halbertal, 1994). Meanwhile, postmodern critics, such as Appadurai (1996), argue that minorities in many parts of the world are at least as artificial as the majorities they seem to threaten, while liberal critics have noted that the problem with cultural rights is that it seeks to uncritically protect groups and not the individuals within them. More will be said about this in the following text.

Kymlicka's (1989) *Liberalism, Community and Culture* marked a watershed. By far his most interesting and controversial claim has been that culture – and by extension, cultural identity – is, in Rawlsian terms, a primary good. That is, like income, basic liberties, opportunities, and a social basis for self-respect, culture is necessary in order to flourish as a human being. This is because cultures supply us with our most basic identities, facilitate trust between ourselves and others, and promote general well-being among others who share our way of life. Cultures also supply persons with the context for choice, and thus enable freedom and autonomy.

The implications for identity are clear: identities arise from within a given cultural context. To be sure, many choices are already made for us, inasmuch as our cultures provide a limited range of options. Importantly, however, while our cultural identities may be given, for Kymlicka it remains for persons to form, pursue, and revise their conceptions of the good life, and this includes how we see ourselves as individuals and in relation to others. Interestingly, Kymlicka argues that some groups, owing to past and present discrimination, are not effectively able to choose or to act unless corrective measures are taken by the state (e.g., affirmative action) to level the playing field. These corrective measures also may require considerable cultural protection of indigenous (though not immigrant) groups. Yet, as a philosophical liberal, Kymlicka is ultimately concerned with individual freedom and autonomy, and therefore argues that persons need to be free to revise or even cast off their cultural identity if they are unable to live it well from the inside. As such, one's identity is not fixed or determined by cultural membership.

Kymlicka's views have been strongly criticized from several directions. First, communitarians argue that culture provides such depth of meaning and direction that one's identity is both circumscribed and bound by it. In Sandel's (1982) terms, the self is irrevocably encumbered. Additionally, there is resistance to what some interpret as a liberalizing agenda in Kymlicka's insistence that persons ought to autonomously subscribe to their cultural identities, that is, rationally and critically reflect upon who they are. Most cultural identities, they argue, are not open to revision, nor should they be.

Liberal critics of Kymlicka mostly target two areas. Some (e.g., Barry, 2001) argue that Kymlicka's cultural defense assumes too unspoiled a notion of how cultures actually are formed and operate, and further, ignores the fact that identities are unavoidably hybrid and complex. Arguing against the one person-one culture model, Waldron (1996) insists that while a person needs cultural attachments for a meaningful identity to develop, a homogeneous, stable, or unaltered culture is not required, nor is one likely to find it. Further, cultural characteristics do not tell us all we need to know about individuals. Rather, positing the one person-many fragments substitute, Waldron argues that in this postmodern, post-Fordist, globalized world the idea of absorbing only one culture – let alone the idea that any univocal culture exists – is implausible.

Concerning Kymlicka's claim that some vulnerable cultures may need state protection in order to help preserve threatened identities, critics note that cultures are not static entities that can be preserved by state intervention except in the most artificial way. Indeed, state protection will only sanction contestable expressions favored by a minority group's leadership. Further, state intervention in order to protect the identity of a minority population

suggests that persons are unable (or unlikely) to adapt, and thus need assistance in order to weather the storm of cultural change. Indeed, Waldron suggests that there is something very paternalistic about the assumption that some are better able than others to acclimate to change.

Still others, notably Okin (1998), point to the problems with defending minority group rights, however mistreated these national or indigenous minorities may have been in the past, or continue to be. Her concern is with a cultural group's most vulnerable members, women and children. She points out that sexual discrimination is rarely overt and states are rarely able or willing to patrol the private sphere. Finally, all liberal critics argue that cultural protection conflates the interests of the adult members with those of the children, and this portends difficulties with the younger generation being able to decide for themselves whether they wish to identify with the culture of their parents.

In philosophy of education, identity formation and maintenance has been a theme most frequently taken up in debates over religious (and recently, home-) schooling. An education that aims to strengthen identity and the commitments that derive from it can be described as a culturally coherent education (Merry, 2005). Cultural coherence describes a continuity of commitments, values and beliefs between the home and school which reinforces a child's cultural or religious identity and builds self-confidence and trust with others who share these commitments. While not without its critics, arguments for cultural coherence operate on the assumption that persons experience emotional stability by identifying primarily (though perhaps not exclusively) with a particular notion of the good life. Further, advocates of culturally coherent education recognize that limiting the range of identity choices/ options, particularly for young children, helps to solidify identity and enable agency to act upon the meanings one attaches to those identities, including the capacity to pursue and revise their self-understandings and attachments.

One advocate of this approach, Burt (1994), argues that excepting for cases involving abuse or neglect (a contested terrain if there ever was one), parents ought to have unchallenged prerogatives to raise children in ways specific to cultural or religious attachments. This is necessary, she argues, for equipping children with the necessary psychological resources to make sense of their inherited world and to live well. As a liberal, Burt stresses the importance of cultivating autonomy but argues that this can be done in a manner consonant with the parent's beliefs or cultural and religious practices. She also acknowledges that freedom to exit one's community is important, but believes that the resources to do this will come about either through exposure to outside influences or, more likely, by engaging with counter narratives within one's own tradition.

With culturally coherent school models, there are worries that parents and teachers may not only discourage critical thinking and dissent, but also use coercion and fear to achieve conformity. The obligatory authority of certain religious texts, or the charismatic authority of community leaders often distinguishes acceptable (e.g., heterosexual) identities from unacceptable ones. Further, critics of this approach to parenting and education insist that too little attention is given to the cultural construction and historical development of any assumed identity. Thus, children must not only be exposed to views outside of one's family and immediate community, but they must meaningfully engage with others whose views are decidedly different before it can be said that they have an informed understanding of what it means to live an identity from the inside. Finally, critics of culturally coherent parenting and education worry over the conflation of the interests, attachments and identities of children with those of their parents. Debates among philosophers of education continue over whether an upbringing to anchor identity during the primary grades is desirable, even necessary, in order to provide children with a foundation upon which to make comparative judgments about alternative identities, or whether children ought to be instructed and acculturated in ways that depart more manifestly from the opinions and habits of the parents (Feinberg, 2006; MacMullen, 2007; Merry, 2007).

See also: Liberalism and Education.

Bibliography

- Anderson, B. (2006). *Imagined Identities*. London: Verso Press.
- Appadurai, A. (1996). *Modernity at Large: Cultural Dimensions of Globalization*. Minneapolis, MN: University of Minnesota Press.
- Appiah, K. A. (2005). *The Ethics of Identity*. Princeton, NJ: Princeton University Press.
- Atkinson, D. R., Morten, G., and Wing Sue, D. (1983). *Counseling American Minorities*. Dubuque, IA: Wim C. Brown.
- Barry, B. (2001). Culture and Equality. Cambridge: Cambridge University Press.
- Benhabib, S. (2002). *The Claims of Culture: Equality and Diversity in the Global Era*. Princeton, NJ: Princeton University Press.
- Burt, S. (1994). Religious parents, secular schools: A liberal defense of an illiberal education. *Review of Politics* 56, 51–70.
- Cass, M. (1979). Homosexual identity formation: A theoretical model. *Journal of Homosexuality* 4, 219–235.
- Chickering, A. W. and Reisser, L. (1993). *Education and Identity*, 2nd edn. San Francisco, CA: Jossey-Bass.
- Côté, J. and Levine, C. (2002). *Identity Formation, Agency, and Culture*. Mahwah, NJ: Erlbaum.
- Cross, W. E. (1971). The Negro-to-Black conversion experience. *Black World* 20, 13–27.
- Erikson, E. (1959). *Identity and the Life Cycle*. New York: International Universities Press.
- Erikson, E. (1968). *Youth, Identity and Crisis*. New York: Basic Books.
- Feinberg, W. (2006). *For Goodness Sake*. New York: Routledge.
- Gilligan, C. (1982). *In a Different Voice*. Cambridge: Harvard University Press.
- Glover, J. (1988). *The Philosophy and Psychology of Personal Identity*. London: Penguin Press.
- Gross, S. E. and Gore, J. S. (2003). Cultural models of the self. In Leary, M. R. and Tangney, J. P. (eds.) *Handbook of Self and Identity*, pp 536–564. New York: Guilford Press.
- Hall, S. and du Gay, P. (eds.) (1997). *Questions of Cultural Identity*. London: Sage.
- Hogg, M. A. (2003). Social identity. In Leary, M. R. and Tangney, J. P. (eds.) *Handbook of Self and Identity*, pp 462–479. New York: Guilford Press.
- Holland, D., Lachicotte, W. J. R., Skinner, D., and Cain, C. (1998). *Identity and Agency in Cultural Worlds*. Cambridge: Harvard University Press.
- Hollinger, R. (1994). *Postmodernism and the Social Sciences: A Thematic Approach*. Thousand Oaks, CA: Sage Publications.
- Hoover, K., Marcia, J., and Parris, K. (1997). *The Power of Identity: Politics in a New Key*. Chatham, NJ: Chatham House.
- Josselson, R. (1987). *Finding Herself: Pathways to Identity Development in Women*. San Francisco, CA: Jossey-Bass.
- Kymlicka, W. (1989). *Liberalism, Community and Culture*. Oxford: Oxford University Press.
- MacMullen, I. (2007). *Faith in Schools? Autonomy, Citizenship and Religious Education in the Liberal State*. Princeton, NJ: Princeton University Press.
- Marcia, J. (1966). Development and validation of ego identity status. *Journal of Personality and Social Psychology* 3, 551–558.
- Margalit, A. and Halbertal, M. (1994). Liberalism and the right to culture. *Social Research* 61, 491–510.
- Markus, H. R. and Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review* 98, 224–253.
- Marshall, J. (2001). A critical theory of the self: Wittgenstein, Nietzsche, Foucault. *Studies in Philosophy and Education* 20, 75–91.
- Merry, M. S. (2005). Cultural coherence and the schooling for identity maintenance. *Journal of Philosophy of Education* 39, 377–497.
- Merry, M. S. (2007). *Culture, Identity and Islamic Schooling: A Philosophical Approach*. New York: Palgrave.
- Merry, M. S. and New, W. (2008). Constructing an authentic self: The challenges and promise of African-centered pedagogy. *American Journal of Education* 115, 35–64.
- Okin, S. (1998). Feminism and Multiculturalism: Some tensions. *Ethics* 108, 661–684.
- Parekh, B. (2000). *Rethinking Multiculturalism: Cultural Diversity and Political Theory*. Cambridge: Harvard University Press.
- Perry, J. (ed.) (1975). *Personal Identity*. Berkeley: University of California Press.
- Sandel, M. (1982). *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press.
- Sen, A. (2006). *Identity and Violence: The Illusion of Destiny*. New York: W.W. Norton.
- Waldron, J. (1996). Multiculturalism and mélange. In Fullinwider, R. (ed.) *Public Education in a Multicultural Society: Policy, Theory, Critique*, pp 90–118. Cambridge: Cambridge University Press.

Further Reading

- Ambrosio, J. (2008). Writing the self: Ethical self-formation and the undefined work of freedom. *Educational Theory* 58, 251–267.
- Berzonsky, M. D. (1990). Self-construction over the life span: A process perspective on identity formation. *Advances in Personal Construct Psychology* 1, 155–186.
- Kenny, M. (2004). *The Politics of Identity*. Cambridge: Polity Press.
- Parekh, B. (2008). *A New Politics of Identity: Political Principles for an Interdependent World*. New York: Palgrave.
- Taylor, C. (1989). *Sources of the Self: The Making of the Modern Identity*. Cambridge: Harvard University Press.

Inclusion

R Cigman

© 2010 Elsevier Ltd. All rights reserved.

A Brief Conceptual History of Inclusive Education

Segregation Versus Integration

The first polarity in the series is between segregated and integrated education, embodied in the UK Education Acts of 1944 and 1981 respectively. The former distinguished between normal and handicapped children, categorizing the latter according to their medical and psychological defects, and providing for them (insofar as it was considered possible to do so) in separate institutions. Many children, considered ineducable, received no provision at all. Categories of handicap included maladjusted and educationally subnormal, which many understandably found offensive.

The 1981 Education Act followed the Report of the Committee of Enquiry into the Education of Handicapped Children and Young People (DES, 1978), under the chairmanship of Mary (later Baroness) Warnock. Subsequently known simply as the Warnock Report, this was commissioned by the UK government to review an area of education that was increasingly seen as unjust. Inspired by a wish to normalize children who had formerly been marginalized educationally, the Warnock committee argued that no child is ineducable, and characterized education in terms of a set of goals toward which every child without exception may aspire. These goals were independence, enjoyment, and understanding, and the idea was that every child without exception must travel down a path toward these common goals. The difference between handicapped and normal children was that the former encountered all sorts of obstacles on the path, whereas the latter traveled with relative ease.

This led to the idea that, instead of classifying some children in terms of their medical or psychological handicaps, we should think of them in terms of their need for help in reaching the common goals. The 1978 Warnock Report, and the Education Act that followed, recognized a universal right to such help, and the concept of a special educational need replaced the concept of a handicap. The right to help imposed on schools and local authorities a duty to provide whatever assistance is required.

The conceptual shift from handicap to special educational need was a defining moment in our ethical and political history. Instead of seeing some children as defective and others as normal, the idea was that every human being has needs, but some are more pressing than others. The aim was to emphasize our common humanity, in

terms of universal educational goals as well as our shared neediness and vulnerability.

This shift was formulated in terms of a series of linguistic proposals. Not only was handicap off the agenda; no longer were we to categorize children according to their deficiencies, for example, their inability to see, hear, or learn quickly. Instead, we were to talk about the difficulties (visual, auditory, learning, etc.) they experienced, and the Warnock committee's concern to avoid negative labeling is the background to a complex debate which continues to this day.

Inclusion Versus Integration

Despite the desire to normalize children hitherto seen as inescapably different, Warnock's conceptualization gradually attracted criticism. The labels might have changed, but labeling continued, and was the basis for special school provision. Warnock had been optimistic that the change in nomenclature would reduce stigma. Doubts were raised about this, and searching questions started being raised about our tendency to differentiate between children by labeling them (in whatever form), and providing for them in special schools.

An effect of the 1981 Education Act was a gradual decline in the number of children in special schools. Many were integrated in the manner recommended by Warnock, with statements of special educational need in many cases, bringing an entitlement to additional help. However, all was not well. Though need and difficulty are arguably less stigmatizing than handicap, a fatal flaw in these concepts – a flaw to which handicap is less susceptible – is their fluidity. There exists a broad spectrum of needs, from those which are impossible to ignore to those which are debatable or even indulgent. Indeed the advantage of the concept of need, from Warnock's point of view, had been precisely this breadth, which spans the normal/abnormal distinction. However, it is impossible to meet all needs, and questions arose about which needs impose duties and which do not. Children hitherto described as blind, now described as having visual difficulties, have obvious needs. Children who struggle to read, on the other hand, have needs that are far less clear: do they, in every case, need additional help, extra time in exams, and so on? Answers to such questions are not always forthcoming; not, at least, in a way that commands general agreement. This was a weakness in the conceptualization of special education, and the practical effect was an explosion in the numbers of children identified as having

special educational needs. Parents fought for this label, and the statements that often accompany it, in order to receive extra support. Of course, the more successful parents tended to be those with social and economic advantages, and special education gradually became a social justice issue.

In this context, special educational needs (SEN) was seen by many as a label with the power to secure disproportionate amounts of resources. When this label secured admission to a special school – far more expensive than additional help in the mainstream – the resource injustice was seen as even greater. As the policy of integration reduced numbers in special schools, questions started to be raised about how the existence of such schools was justified. Why not (people asked) use this money to enhance the capacity of mainstream schools to respond to diversity? Questions were also raised about why some children are seen as having special difficulties, and the answer that presented itself to many was that schools create and sustain these by designing their educational delivery around so-called normal children. An alternative approach would be to re-conceptualize schools around the concept of diversity, in which difference is seen as universal – every individual is different from every other – and no one is seen as normal.

This was the philosophy of inclusion, as distinct from the philosophy of integration, and it was first articulated by UNESCO in a document that became known as the Salamanca Statement (UNESCO, 1994). This was a proclamation by delegates from 92 governments and 25 international organizations for the World Conference on Special Needs Education at Salamanca in Spain. It says:

Regular schools with this inclusive orientation are the most effective measures of combating discriminatory attitudes, creating welcoming communities, building an inclusive society and achieving education for all; moreover, they provide an effective education to the majority of children and improve the efficiency and ultimately the cost-effectiveness of the entire education system. (Section 2, p. ix)

This was a new conceptualization of schools, and it marked a significant shift from the principles of the 1978 Warnock Report. It was reinforced by a conceptual shift that was being promoted by disabled people, from a medical to a social model of disability. In 1975 the Union of the Physically Impaired against Segregation (UPIAS) had written “It is society that disables people,” and this gradually led to the imperative to change society by changing schools. The medical and social models of disability underpin the inclusion debate as we know it today, and these will be discussed now.

Medical Versus Social Model of Disability

The distinction between the medical and social models of disability emerged as a question: where is disability

located? Is it located within individuals or within society? Disability Equality in Education (2005) says:

We believe that the problem is not in the child and their impairment, but in the social and attitudinal barriers in the education system. This ‘social model’ draws on the thinking of disabled people and underpins all inclusive education (T. E. S. 08.07.05)

Inclusive education is a social response to the medicalization of disability. Instead of seeing some but not others as deficient in various ways, the idea is that society makes some people unable to function as they are capable of functioning. Install ramps and lifts in buildings, and people in wheelchairs will no longer be functionally disabled. Change attitudes toward wheelchair users, so that they are perceived as having different rather than impaired mobility and are welcome in society, and we are talking about inclusion.

The shift from a medical to a social model is synonymous, for many, with justice and equality for disabled people. Yet it has not been immune to criticism, even from people with disabilities. Lord Low (2007), who is visually impaired, says:

If education is about anything, it is about influencing and indeed changing the individual child. One may do this by modifying the social environments in which the child is placed, but one cannot eliminate the individual dimension altogether. (p. 9)

Norwich (1996) makes a similar point:

...it is one thing to attribute the disadvantage of a disabled person only to individual characteristics, it is quite another to exclude the role of individual characteristics. (p. 27)

Many sociologists and educationists see the medical and social models as mutually exclusive, and the former as responsible for discrimination against disabled people. Low and Norwich suggest, on the contrary, that these models need to be combined in some way if we are to do justice to all children. However, the idea persists that the medical and social models are exclusive, and this polarization now extends to the question of educational placement.

It makes more sense to see placement as a matter of either–or than it does to see models this way. After all, children attend either a mainstream or a special school (although special units in mainstream schools break down this polarity to an extent). What makes less sense, perhaps, is the idea that schools either exclude or include children, rather as disability models are either medical or social. This idea implies that special schools are necessarily exclusive, in the sense of being discriminatory, divisive, and unjust. This suggests that inclusion/exclusion should be seen as a polarity between types of institution.

Inclusion (Mainstream Provision) Versus Exclusion (Special Provision)

The Salamanca Statement introduced an argument about rights, saying that every child has a right to an education, which will be most effectively provided for the majority in mainstream schools which adapt to children's diverse characteristics and needs. This proclamation was seized upon by defenders of the social model, who interpreted the right to an education as a right to an inclusive education, which is only possible in a mainstream school. This is not what the Statement says; indeed, it clearly states that in exceptional cases, children should be educated in "special schools or special classes within ordinary schools."

The right to an inclusive mainstream education is not universally accepted, and it is legitimate under such circumstances to ask: where does this right come from? By whom or what is it authorized? It is not a legal right, so if anything it must be a moral right. But it is one thing to claim such a right, another to establish it as trumping all other rights, and imposing duties on others. If it exists, the right to an inclusive mainstream education (understood as a trumping right) imposes a duty on all parents to seek this kind of education on behalf of their children. Those who fail to do this will then be seen as violating their children's rights. This idea is naturally found provocative by parents who feel that their children need a nonmainstream environment, and that they are doing the best for their children by sending them to special schools.

The right to an inclusive mainstream education imposes a duty on mainstream schools to provide for all pupils without exception, that is, to welcome them and adapt to their diverse needs. However, this is problematic, for it is far from universally agreed (indeed many seriously doubt) that it is possible for all schools to do this (see, for example, Pirrie and Head, 2007; Rogers, 2007a, b; Moore, 2007). Some children (e.g., those with social, emotional, and behavioral difficulties) are disruptive to the extent that they compromise other children's rights to an education. Others (e.g., those on the autistic spectrum) find mainstream schools a social nightmare; their difficulties are social and communicative, and they seem to require a quiet, ordered environment very different from that of a mainstream school.

The implied duty of mainstream schools to adapt to all children is not (logically could not be) a duty to do this successfully. At most, it is a duty to attempt to do this. Doubts about the possibility of achieving this are often met by the idea that inclusion is a process of learning to adapt to diversity which will not succeed so long as parents withdraw their children from the mainstream when difficulties are experienced (see Ainscow, 1993). This is, in effect, a loophole for failure. Failures include children who, because of their particular difficulties, feel more excluded than included in inclusive mainstream

schools, and are consequently miserable and unable to learn. Concerns about exclusion within inclusion were recently articulated by Baroness Warnock, and we turn to these now.

Inclusion (Feeling Included) Versus Exclusion (Feeling Excluded)

Concerns about children who feel excluded within inclusively orientated schools gave rise to what could be termed a counter-movement that defended special schools for some children and offered a definition of inclusion that was independent of mainstream provision. It was argued that the range of abilities and needs is so great that some children will inevitably lose out on an education in mainstream schools, however inclusively oriented. This does not mean that inclusion is impossible. Rather inclusion should be taken to mean, as the UK National Association of Head Teachers (NAHT, 2003) puts it:

a process that maximises the entitlement of all pupils to a broad, relevant and stimulating curriculum, which is delivered in the environment that will have the greatest impact on their learning. All schools, whether special or mainstream, should reflect a culture in which the institution adapts to meet the needs of its pupils and is provided with the resources for this to happen. (Quoted in Warnock, 2005: section 3, p 41)

Inclusive education means being educated, the NAHT paper goes on:

... in the most appropriate setting. This will be the one in which [children] can be most fully included in the life of their school community and which gives them a sense both of belonging and achieving. (Warnock, 2005: section 3, p 41)

In 2005, Mary Warnock published a booklet called *Special Educational Needs: A New Look*, in which she quoted the above passages from the NAHT paper, and argued that the concept of inclusion is a disastrous legacy of the 1978 Report. There was an outcry in the media and the special educational world. Educationists and disability activists accused Warnock of trying to segregate society, and she was attacked with particular vehemence by organizations like Parents for Inclusion who wanted to bring about the closure of all special schools by 2020. This organization exemplifies the idea that special is synonymous with exclusive in passages like this:

The 2020 Campaign is based on the experience of disabled adults who went to special education schools and colleges. They experienced abuse, isolation and failure that emotionally scarred thousands for life. End this shameful exclusion that ruins lives. (Parents for Inclusion, 2004)

In effect, Warnock was leading the way for a moderate philosophy of inclusion, as distinct from the universalist philosophy which seeks to place all children in the same schools. In her 2005 booklet, Warnock drew attention to certain groups, such as autistic children, many of whom, she said, are bullied relentlessly and suffer all the pains of the permanent outsider when placed in mainstream schools. She wrote:

The reality seems to be, in many cases, that [inclusion] is experienced as a painful kind of exclusion. (p. 43)

This is a new concept of exclusion – exclusion as a painful experience of nonbelonging – and it is one that is likely to be of particular concern to parents. Baroness Warnock was arguing for a new polarity, between feeling included and feeling excluded. She was saying that the former, rather than mainstream attendance, is an essential component of a good education.

This sensitivity to the feelings of children, and tendency to cast some as particularly vulnerable, was seen by universalists as playing into the hands of indulgent middle-class parents who favor selective education and expect more than their fair share of resources. Warnock spearheaded a deep distress among many parents (and, it must be said, many teachers too) about the neglect of some children, albeit a small minority, in the promotion of inclusive education. The distress had been mounting during the 1990s and early 2000s, and it came to a head with Warnock's 2005 publication. At the launch of this publication, disability activists tried to bar the entrance, and a parent who had been struggling against the closure of special schools openly broke down in tears. At the heart of the debate was a violent clash of views about the meaning of special schools. Are they lifelines for some, which any decent society must provide? Or do they represent shameful exclusion, as Parents for Inclusion were arguing?

It is against the background of such questions that we finally turn to the philosophy of inclusion. According to one argument, there is no inclusion in special schools. According to another, inclusion means being educated in the most appropriate setting, which has the greatest impact on learning, and this may be special or mainstream. Which is more defensible? In order to answer this question, we need to examine the values and assumptions at the heart of these positions, and ask what they imply about education.

The Inclusion Debate: A Philosophical Perspective

The debate about inclusion is a fierce debate about values. What values are in dispute? Are there any values about

which universalists and relativists/particularists? or thin and thick, weak and strong universalists? moderates agree? When they call themselves inclusionists, do they use that word with the same or different meanings?

In a recent publication (Cigman, 2007a) about the post-2005 inclusion debate, Brahm Norwich wrote:

It is rare to find arguments against inclusion, as it is rare to find arguments against democracy. Where disagreement lies is in the nature and extent of inclusion. (p. 71)

It is true that it is rare to find arguments against inclusion, but unless we understand the meaning of this term, it is possible that disagreements are misleadingly couched in terms of a single concept. We want to suggest that there is a common concept, a basic value shared by all those who support educational inclusion in some sense. Unfortunately, this common meaning is often buried in the heat and rhetoric of the debate.

Recalling segregation in the post-war period, it seems clear that one reason why there has been no call for a return to this system is that it is seen as demeaning to people who are cast as handicapped. Like sending children to secondary modern schools because they are deemed unintelligent, or to racially segregated schools because they are not white, the mandatory placement of a child at an institution that caters for abnormality is indeed education by deficit. Our objection to this is that it is deeply disrespectful, because it disregards the interest we all have in being accepted and welcomed in society. The shift to integration wherever possible acknowledged this disrespect, and tried to de-stigmatize groups that had suffered under such a system.

It gradually became obvious, first, that segregated education was often a matter of convenience rather than necessity, and second, that many people suffered terribly from the feeling of being excluded. Indeed, they were in a real sense excluded, particularly if they were unable to benefit from the kind of education that was offered to the majority of children. The barriers to inclusion (an expression now in frequent use by inclusionists) were an unwillingness to make appropriate physical or technological adjustments, and more fundamentally, negative or unwelcoming attitudes.

Philosopher Margalit (1996) has argued that a decent society is one "whose institutions do not humiliate people" (p. 1). Humiliation, he says, is defined as "any behaviour or condition that constitutes a sound reason for a person to consider his or her self-respect injured (p. 2)." We have argued (Cigman, 2007a) that our objection to segregated education – that is, the objection of most members of our society – is rooted in the sense that we do not want to belong to a society that institutionally humiliates its members. Post-war segregated schools did this, and their continued presence would have prolonged an indecency that we want no part of. What it means to

say that such schools institutionally humiliate their members is that children who attended such schools had a sound reason to consider their self-respect injured. This is not the same as simply feeling that one's respect is injured, which can be based on an exaggerated or misplaced sense of one's own dignity.

The shift from segregated to integrated education was widely welcomed because it accorded with the fundamental moral principle that every human being should be treated with unconditional respect. To send a child to a segregated school because her needs are difficult to deal with is to disregard her basic interest in being accepted as a member of society. This, many would argue, gives a person a sound reason to consider her self-respect injured, and the imperative not to do this lies, we submit, at the heart of the philosophy of inclusion. The recent history of inclusion is a history of divergent interpretations of this basic principle.

Interpretations are divergent because people disagree about what constitutes a sound reason to consider one's self-respect injured. Specifically, they disagree about whether special school placement constitutes such a reason. Universalists justify their argument that special school placement is inherently humiliating by saying that it is never really voluntary. It is and can only be chosen as the lesser evil in a society that fails to provide a properly inclusive education for every child.

However, there is a weakness in this argument. If someone chooses a special school, believing they do so freely, the argument that this was not really a free choice depends on the view that such a choice is in every case contrary to human interests. It can only be chosen, according to this view, insofar as the universally desired mainstream option is either unavailable or not inclusive enough. However, the mainstream option is not universally desired. This could be because a parent has selective impulses – she wants a better school for her own child than she is prepared to accept for other children – or it could be because she sees mainstream education as contrary to the basic interests of her child. Some parents argue that mainstream education is by its nature inadequate for some children, and good intentions will never overcome this. Furthermore, they experience the denial that parents can and should make judgments about the basic interests of their children as disrespectful.

We have written that the universalist position:

... is universalist in aspiration but not in meaning. It aspires towards a society in which *all* children attend mainstream schools, but it is based on an understanding of what special schools *mean* which is conspicuously unshared. (Cigman, 2007b: xviii)

The discrepancy between aspiration and meaning is a serious problem for the universalist position. To a universalist, special means exclusion, segregation, and humiliation.

To many others it means the only setting in which a child with particular difficulties may feel included and learn effectively. This difference of perspective should prompt a meta-question: how is the difference to be explored or negotiated? Instead of addressing this question, the universalist reiterates undefended claims, like the claim that every child has a right to an inclusive mainstream education.

This disagreement about the meaning of special is underpinned by a disagreement about the meaning of disability. The social model which underpins the universalist concept of inclusive education says that all disability is located in society, rather than in individuals. Philosophically, this disjunction is dubious. Wittgenstein's later philosophy (see, for example, Wittgenstein, 1953: *passim*) should make us wary of the idea that a concept like meaning is intelligibly discussed in terms of location. Practically, too, the social model is problematic, because it wrongly assumes that everyone is at ease with this interpretation of disability. In fact, the interpretation is disputed, not simply by professionals who seek gains from the medicalisation of disability, but also by disabled people themselves. The following is by a disabled woman:

... there is a tendency within the social model of disability to deny the experience of our own bodies, insisting that physical differences and restrictions are entirely socially created. While environmental barriers and social attitudes are a crucial part of our experience to disability—and do indeed disable us—to suggest that this is all there is to it is to deny the personal experience of physical or intellectual restrictions ... (quoted Oliver, 1998: 38)

Interestingly, this passage is quoted by Mike Oliver, a disabled activist, one of the first and most vociferous promoters of the social model. By quoting it, he effectively acknowledges the limitations of the social model and the possibility, as Norwich (1996) puts it (quoted above), that:

... it is one thing to attribute the disadvantage of a disabled person only to individual characteristics, it is quite another to exclude the role of individual characteristics. (p. 27)

To exclude the role of individual characteristics is to perpetuate the offence that the inclusion debate seeks to address: treating people with disrespect. It is disrespectful to deny people's personal experiences of bodily and mental restrictions. The anguish of children forced into mainstream schools is ignored in the service of a theory which says that all differences are to be subsumed under the heading of diversity. This exclusion of individual experiences has elicited outraged responses precisely because (according to my argument here) it is experienced as deeply disrespectful.

At the heart of the universalist position is denial of the dilemma of difference, according to which:

[w]e either treat all children as essentially the same, which means treating them as fairly as possible but with the risk of neglecting individual differences. Or we treat them differently, with the consequence that some are better off than they would otherwise have been, but there is a risk of being unfair by devoting more resources or expertise to some than others. (Cigman, 2007b: xxii)

Inclusive education is complex, conceptually, politically, and ethically, because it presents us with the conflicting values of fairness as equality and fairness as attention to individual or special need. This dilemma does not go away when a decision is made to emphasize the first at the expense of the second. Both values need due attention, and the way forward is to engage in good, respectful dialog between all interested parties, and pay proper attention to empirical evidence.

Bibliography

- Ainscow, M. (1993). Beyond special education: Some ways forward. In Visser, J. and Upton, G. (eds.) *Special Education in Britain after Warnock*, pp 166–182. London: David Fulton.
- Cigman, R. (2007a). A question of universality: Inclusive education and the principle of respect. *Journal of Philosophy of Education* **41**(4), 775–794.
- Cigman, R. (2007b). Editorial introduction. In Cigman, R. (ed.) *Included or Excluded? The Challenge of the Mainstream for Some SEN Children*, p xv. Oxford: Routledge.
- DES (1978). *Report of the Committee of Inquiry into the Education of Handicapped Children and Young People: The Warnock Report*. London: Her Majesty's Stationery Office.
- Disability Equality in Education (2005). In defence of inclusion: Warnock challenges the rights of disabled children to inclusion. *Times Educational Supplement* July 8th.
- Low, C. (2007). A defence of moderate inclusion and the end of ideology. In Cigman, R. (ed.) *Included or Excluded? The Challenge of the Mainstream for Some SEN Children*, pp 3–14. Oxford: Routledge.
- Margalit, A. (1996). *The Decent Society*. Cambridge, MA: Harvard University Press.
- Moore, C. (2007). Speaking as a parent: Thoughts about educational inclusion for autistic children. In Cigman, R. (ed.) *Included or Excluded? The Challenge of the Mainstream for Some SEN Children*, pp 34–41. Oxford: Routledge.
- NAHT (National Association of Head Teachers) (2003). *Policy Paper on Special Schools*. <http://www.naht.org.uk/welcome/resources/>

- key-topics/special-education-needs/policy-paper-on-special-schools (accessed August 2009).
- Norwich, B. (1996). *Special Needs Education, Inclusive Education or Just Education for All?* Inaugural Lecture. London: Institute of Education University of London.
- Parents for Inclusion (2004). Parents for Inclusion: Parents helping parents so their disabled children can learn, make friends and have a voice in ordinary schools and throughout life. <http://inclusion.uwe.ac.uk/csie> (accessed August 2009).
- Pirrie, A. and Head, G. (2007). Martians in the Playground: researching special educational needs. *Oxford Review of Education* **33**(1), 19–31.
- Rogers, C. (2007a). *Parenting and Inclusive Education: Discovering Difference, Experiencing Difficulty*. Basingstoke: Palgrave Macmillan.
- Rogers, C. (2007b). Experiencing an “inclusive” education: Parents and their children with “special educational needs”. *British Journal of Sociology of Education* **28**(1), 55–68.
- UNESCO (1994). *The UNESCO Salamanca Statement and Framework for Action on Special Educational Needs Education*. Paris: UNESCO.
- Warnock, M. (2005). *Special Educational Needs: A New Look*, 1st edn. London: Philosophy of Education Society of Great Britain.
- Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.

Further Reading

- Clark, C., Dyson, A., and Millward, A. (eds.) (1998). *Theorising Special Education*. London: RoutledgeFalmer.
- Farrell, M. (2004). *Inclusion at the Crossroads Special Education – Concepts and Values*. London: David Fulton.
- Jaeger, P. L. (2008). *Understanding Disability: Inclusion, Access, Diversity, and Civil Rights*. Westport, CT: Praeger.
- Lewis, A. and Norwich, B. (2005). *Special Teaching for Special Children? Pedagogies for Inclusion*. Oxford: Oxford University Press.
- Mitchell, D. (2007). *What Really Works in Special and Inclusive Education: Using Evidence-Based Teaching Strategies*. London: Routledge.
- Norwich, B. (2007). Dilemmas of inclusion and the future of education. In Cigman, R. (ed.) *Included or Excluded? The Challenge of the Mainstream for Some SEN Children*, pp 69–84. Oxon: Routledge.
- Norwich, B. and Kelly, N. (2005). *Moderate Learning Difficulties and the Future of Inclusion*. London: RoutledgeFalmer.
- Oliver, M. (1996). *Understanding Disability: From Theory to Practice*. Basingstoke: Palgrave Macmillan.
- Sainsbury, C. (2000). *Martian in the Playground: Understanding the Schoolchild with Asperger's Syndrome*. London: Sage.
- Shapiro, A. (2000). *Everybody Belongs: Changing Negative Attitudes Toward Classmates with Disabilities*. London: RoutledgeFalmer.
- Sinclair, J. (1998). Why I dislike person first language. http://web.syr.edu/~jsincla/person_first.htm (accessed August 2009).
- Terzi, L. (2008). *Justice and Equality in Education: a Capability Perspective on Disability and Special Educational Needs*. London: Continuum.

Justice and Care

M Tjiattas, North Carolina State University, Raleigh, NC, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Autonomy – Capacity for self-determination; ability to guide one's conduct on the basis of one's own judgment and reasons.

Care ethics – Moral theory for which caring relations between particular individuals are central to all morality.

Contractualism – The view that morality is best understood as based on a contract or agreement (actual or hypothetical).

Dignity – Moral status appropriate to autonomous beings.

Distributive justice – Fair allocation of social and economic benefits and burdens.

Empathy – Capacity to adopt the point of view of another, "put oneself in another's shoes."

Impartial – Give equal consideration to all without favoring some over others.

Original position – John Rawls's hypothetical choice situation in which parties, from an impartial point of view, choose principles of justice that will govern the fundamental institutions of society.

Particularism – View that denies a central role to moral principles and general rules in morality; moral attitudes display unmediated responses to particular circumstances, including unmediated concern for particular individuals.

Reasonability – Capacity and willingness to consider reasons others can share when evaluating and justifying actions, positions etc.

Respect – The attitude of recognizing the rights and dignity of autonomous agents; for many moral theories, the fundamental attitude toward persons.

Rights – Morally justified claim or entitlement that protects liberty or autonomy against interference from others, or provides justified expectation to assistance from others.

Sentimentalism – The theory that moral attitudes are grounded in emotions or feelings rather than in reason.

Universalizability – The principle that moral claims applying to any individual must apply to all whose circumstances are relevantly similar to those of the individual concerned.

Vulnerability – Personal features or circumstances making a person susceptible to unfair treatment or exploitation.

Introduction

It is not self-evident that justice and care are mutually exclusive. The contrary view is due to early proponents of the ethics of care (Gilligan, 1982; Noddings, 1984), who attempted to provide a radical alternative to what they took to be the traditional, male-dominated moral philosophy which they referred to as the ethics of justice. This is characterized by a commitment to universal rights, impartiality, disinterested evaluation, rationality, and to universal principles of conduct. The ethics of care in direct contrast embraces particularism, takes moral judgment to be motivated and informed by caring attitudes or sentiments, and is concerned about protecting and sustaining attachments and relations with particular individuals.

The justice–care debate emanates from this assumed bifurcation of moral perspectives. Engagement in the debate has been uneven. From the beginning, it has lacked a well-defined set of alternatives on which disputants could join issue, and many developments within the ethics of care have blurred what were initially presented as stark differences. Although debate talk cannot be entirely avoided in discussing the relations between justice and care given its centrality to the original formulations of care theory, it is important to keep in mind that questions of their conceptual relations are far from settled.

Justice has not lacked able defenders. It is widely recognized that John Rawls's *A Theory of Justice* is the defining work of contemporary political philosophy, setting its basic parameters and guiding its articulation and development. Social justice theory, its centerpiece, focuses on the most basic and formative institutions of society (social, political, and economic), which exert pervasive influence on the choices, life prospects and opportunities of individuals. Its central aim is to specify principles concerning the conditions under which the basic structure of society is just, ensuring that activities that they regulate result in just outcomes, or, as Rawls puts it, that the burdens and benefits of social cooperation are fairly distributed. Principles of justice define the fundamental terms of association, agreement, and forms of social cooperation which regulate and determine the distribution of wealth, income, goods, and opportunities. In addition, they mediate both intimate and nonintimate relations between people, providing the means by which they discharge at least some of their moral duties to one another. For Rawls, this requires that the principles be endorsable by free, equal, reasonable, and rational people. For our purposes, the

salient features of Rawls's theory of justice are its commitment to moral equality, respect for the dignity of persons through recognition of rights, the moral constraints (within which individuals may pursue their interests and principles) that determine the fair distribution of benefits and burdens of social and economic activity, and reasonable agreement as the central method of justification.

Against this background, the ethics of care stands out as a radical approach to morality that challenges traditional positions not only on substantive matters but also on what is involved in moral thinking and experience. It rejects the priority of justice and favors particularist values on the ground that personal relationships and attachment are much more significant than relations with strangers, and accordingly that establishing and sustaining such relationships should be taken as far more important, morally speaking, than co-coordinating interests or resolving conflicts impartially on the basis of principles. For the ethics of care, morality is based, not on principles, but on unmediated relations with others, and its main concern is the welfare, rather than the moral powers, or dignity, of others. Moreover, some care ethicists suggest that this contrast marks a gender divide: men and women tend to conceive morality in different ways. The former are drawn to autonomy, rights, justice and the latter to connection and caring.

These seemingly abstract concerns raise many questions central to moral education. Given its traditional emphasis on autonomy and critical thinking, moral education derives much of its impetus from the ethics of justice. For example, civic education emphasizes the inculcating of capacities required in order for individuals to fulfill responsibilities and enjoy the rights of democratic citizenship. Autonomy, civic toleration, and a critical attitude toward contending political claims are deemed essential for participation in liberal society (Gutmann, 1995). These competences and virtues are also seen as inherently valuable or at least instrumental to the well-being of individuals (Nussbaum, 2000) and therefore central to moral education. Epistemic capacities like knowing how to think and reflect critically and being aware of nonrational influences on processes of belief formation have the dual effect of enhancing both personal well-being and the performance of civic responsibilities. Requirements of justice (arising especially from equality of opportunity) have also structured deliberations about educational resources and their distribution (Brighouse, 1998). The outcome of the debate, or at least developments in response to challenges to traditional moral philosophy from the ethics of care, thus has repercussions for moral and political education. If, on the one hand, the ethics of care replaces traditional concerns of justice, then corresponding upheavals in philosophy of education can be expected to follow. If, on the other hand, an integrated theory of justice and care ethics is attained, this would encourage a richer account of such entrenched concerns in philosophy of education as

the aims of moral education, the grounds of authority or legitimacy in education, and the distribution of educational resources.

The Ethics of Justice

Conceptions of Justice

The formal notion of justice (people should be treated equally unless there are relevant differences between them) and its correlates, namely impartiality (like cases should be treated alike) and reciprocity (others should be treated as one would expect to be treated by them), have a long pedigree, going back at least to Aristotle. Rawls's work preserves the core of this conception, while specifying it in terms of principles of fair distribution of the benefits and burdens of social life.

For Rawls, justice, the first virtue of institutions, is about securing fair terms of participation in a society made up of inescapably interdependent members. His conception of justice, formulated as two ordered principles, is tailored to (1) guarantee fundamental individual liberties, assigning them inalienability and absolute priority, and (2) embody an egalitarian ideal of social cooperation based on reciprocity, ensuring fair equality of opportunity and demanding that any inequalities benefit the worst off.

Thus construed, justice provides each person with the conditions and resources required to meet one's fundamental interests, to develop and exercise one's fundamental capacities (moral powers of free, responsible, and rational agency), and to realize one's conception of the good. The first principle secures that status of equal citizenship, and the second, fair opportunity and adequate resources for all.

The Argument for Justice as Fairness

In order to establish the legitimacy of these principles, Rawls employs a contractualist argument to show that they would be agreed to by reasonable people under fair conditions of choice. He describes a hypothetical original position in which parties to the agreement (representatives of actual individuals) operate under constraints on choice (corresponding to exclusion of information about morally irrelevant individual interests) with the result that they have no option but to consider matters from the point of view of all, thereby making decisions and choices that are fair.

Since the original position is merely hypothetical, Rawls addresses separately the question of whether the principles adopted in the original position are realizable. To this end, he provides an account of how citizens in a well-ordered society come to acquire a sense of justice. Contrary to what many of his critics allege, the account is not solely or even largely rationalistic. Central to learning to appreciate the point of view of others and developing

virtues of impartiality and considerateness, and ultimately a sense of justice are the love, trust, affection, and guidance experienced in the family. Reciprocity, according to Rawls's account of moral psychology, is contingent on experiencing and recognizing that others intend our good. On such experiences the desire to reciprocate, the desire to form attachments to persons and institutions, and social cooperation are based. For the purposes of this article, it is important to note the emphasis Rawls places on forming of attachments for the development of commitment to the moral point of view and the availability of a common vantage point from which fundamental political issues can be addressed by citizens in a reasoned and principled way.

The Ethics of Care

A Different Voice

The ethics of care at its inception was a predominantly feminine or feminist approach aspiring to foreground women's moral experience in developing a distinctive moral philosophy. Gilligan's founding work did not target Rawls or other theorists of justice directly but was a response to Kohlberg's canonical six-stage theory of moral development (which presented the morality of justice as the highest stage). Gilligan questioned Kohlberg's assumptions and conclusions, in particular, his finding that the ultimate stage of moral maturity involves a contractarian respect for persons, embracing and applying general principles, and taking fairness and rights to be fundamental values, and his relegating moral evaluations based on caring and relationships to a lower stage of development. She took these judgments to reflect a bias in favor of male values, interests, and psychological proclivities. Kohlberg's procedures and findings, Gilligan argued, captured the male approach to moral judgment but not the female voice of care, responsibility, and connection with others.

For the ethics of care, responsibility and relationships rather than rights and rules are central and these are inextricably bound to concrete circumstances and contexts. The ethics of care thus focuses on caring relations and on responsiveness and attentiveness to need, while the ethics of justice, concerned as it is with mediating and resolving conflicts, takes as central fairness, equality, individual rights, and abstract principles. Although Gilligan holds both to be normatively adequate, they are not compatible because moving from one to the other is like effecting a gestalt switch. A problem is construable either as one of justice or one of care, but not both.

The Particularity of Caring

Noddings' starting point is that traditional moral philosophy neglects people who, due to personal or social

circumstances, have not been able to attain the status of autonomous agents and so offers at best a partial treatment of central issues of the moral life. She seeks to remedy this by providing a permanent place within moral theory and practice for the cared for.

This leads her to deny that principles and rules are central to morality. Traditional ethics, centered on moral reasoning and the establishment of principles, employs the language of the father evidenced by the centrality of justice and fairness. It fails to recognize that sympathy or empathy feeling with the other, is the wellspring of moral conduct. Sentiments of natural sympathy are the fundamental source of moral conduct and the caring relation is both morally and ontologically basic (the desire to maintain and recapture the most caring moments undergirding the ethical ideal). Caring is rooted in receptivity, relatedness, and responsiveness, the feminine alternative to Kohlberg's last stage.

The radical nature of Noddings' position stems from her rejection of all general elements and procedures, both rationalist and sentimentalist. The ethical relationship is not an expression of universalizability or general benevolence, but is a relationship between two parties and two parties only: the one caring and the one cared for, a concrete process of caring for distinguishable from generally caring about. It involves actual people, who, in order to achieve caring for, need to fulfill specific requirements: on the part of the one caring, of engrossment in the other (e.g., empathizing with her, stepping into her frame of reference, and being concerned for her welfare) and on the part of the cared for, requirements of recognizing and responding to caring. In this scenario, principles are not only redundant, but also obstructive, since the essential elements of caring are located in the relation between the one caring and the one cared for. The desired moral response should be immediately elicited in attending to the cared for. Any moral decision making requires concretization rather than abstraction.

Noddings shuns talk of moral justification. She says that caring itself and the ethical ideal that strives to maintain and enhance it guide us in moral decisions and conduct. The marks of ethical behavior are feeling, thinking, and conducting oneself as the one caring. In part, ethical conduct is determined by one's actual receptivity. One is irredeemably evil when one intentionally rejects the impulse to care. Justice plays an ancillary, second-best role: it is needed only when it is impossible to exercise caring for. Noddings furthermore insists that the ethics of care can provide alternative notions of reciprocity and obligation based on the relatedness of caring, and so does not have to make concessions to the ethics of justice in order to account for these admittedly important elements of morality.

A Conflict?

Proponents of justice are inclined to assume or assert that care and justice are compatible, while adherents of the ethics of care tend to deny it, presenting care as a comprehensive ideal that provides answers to all moral and political questions. This section examines the issue by considering four points of conflict alleged by care theorists:

1. Care ethicists point to the different moral psychologies adopted by the two positions. The ethics of justice foregrounds rationally grounded, universal moral principles applied impartially across cases, while the ethics of care centers on empathically informed responsiveness to particular situations. Reservations about abstract reasoning in the moral domain range from meta-ethical concerns about adopting universalizability (understood as mandating uniformity and so excluding partial judgments based on empathy, caring, and connection) as a fundamental principle in moral reasoning, to worries that abstraction, while posing as a universal value, promotes the interests of dominant, male social agents. These suspicions are extended to justice and rights.

In response to this alleged conflict, Okin (1989, 2004), Herman (1993) and others have argued that the contrast between the reasoning employed in the two approaches is overdrawn, and is often based on a misconstrual of theories of practical reasoning. Proponents of universalizability do not deny the role of particular judgments in moral reasoning, recognizing that it involves identifying more restricted principles for application of abstract ones, and requires particular, contextual details. There is ample room for contextually sensitive elements to play a determining role in moral judgment. On the other hand, principles are indispensable to isolating morally salient features of a situation. For example, many instances of caring in the descriptive sense are clearly intrusive, abusive, or exploitative (Friedman, 1987). To determine whether an activity constitutes caring in the moral sense, it is not sufficient to rely only on the intentions or avowals of the caregiver. Engrossment is not self-authenticating; it cannot identify the normative features of a situation.

2. Impartiality, widely taken to define the moral point of view, is believed by care theorists to endanger attachment, and so requiring it is not only psychologically burdensome but also morally deleterious (Slote, 2007). Read as proscribing special obligations to intimates and associates, it leads to a morality that ignores the central importance of interpersonal relations in our lives, undervaluing love and other intimate feelings. The biases inextricable from empathy, for example, in favor of dealing with those in spatial and temporal proximity, favoring one's own children, are precisely

what make considerations morally salient. On this matter, it might appear that there is a real dispute between justice and care ethicists. However, on closer consideration, evidence of a deep divide evaporates. Moral philosophers who support impartiality are concerned to defend it as a second-order criterion, that is, one to be applied to the basic rules and principles of morality and justice, while care theorists attack it as a first-order criterion which most justice theorists would not want to defend (Barry, 1995).

3. The justice perspective conceives of individuals as autonomous, that is, responsible for and in control of their own lives, and as equal members in a moral community. This, some care theorists allege, reflects an erroneous view of human nature and circumstances based on male stereotypes of self-reliance. It reflects a failure to recognize human dependency, the prevalence of involuntary relations between people and, indeed, of the fact that individuals do not create themselves. In devaluing connectedness in favor of separateness, it fails to recognize that ethical activity is unmediated by general principles and inextricably bound to particular relationships. However, objections along these lines are at least partly due to misunderstanding of the thesis of separateness of persons, which is advanced by theorists of justice against the utilitarian move of aggregating individual interests, allowing individual benefits to be sacrificed for collective ones often without restriction. Its advocates (see especially Rawls) make it clear that separateness is meant neither as a metaphysical thesis nor as a rallying cry for self-reliance, and hence that autonomy does not necessarily imply sociological atomism (Okin, 2004).

It is difficult to see how the claim that autonomy is detrimental to caring can be upheld or how an adequate conception of care could jettison autonomy. Effective caretaking requires both that caretakers be autonomous and that the autonomy of the ones cared for be recognized and respected as well (Grimshaw, 1986). In short, respect for others is a non-negotiable requirement of adequate care (Friedman, 1987).

4. A central claim of early care theorists was that the ethics of justice ignores or does not take up the perspective on morality peculiar to or favored by women, resulting in their marginalization and exclusion from decision making in the public domain. But the ethics of care has itself occasioned feminist concerns. Its uncritical acceptance of all self-described relations of caring as ethically valuable puts it in danger of perpetuating existing gender inequalities, making a virtue of the contingent fact that women have borne the burden of caring in society. Moreover, the commitment to particularism has meant that even if care theorists were concerned about structural inequalities involved in discriminatory systems, they would not have the

resources to address them. Sensitivity and responsiveness to need alone seem too slight to handle social problems such as domestic violence and coercion or broader issues of structural inequality. The restriction to relations between pairs of interacting people, and the insistence on contextuality seem to limit care to a personal ethic concerned with meeting individual needs and maintaining particular relations in concrete situations.

A Possible Synthesis: Extending the Boundaries and Justifications of Care

The most recent general discussions of the ethics of care, presaged by important work by Tronto (1993), reflect a willingness to explore common ground with more traditional approaches.

Recognizing that care requires resources that are available only within the framework of appropriate institutional structures, Tronto argues that it cannot be confined to relations between pairs of individuals but must be located in a wider social and political context. This also counters the perception of the work of care as exclusively, except in the private sphere, concerned with the marginalized and needy, which has had the effect of foreclosing questions about the just distribution of care work.

Tronto responds also to what she deems the normative deficit of early-care theory by focusing, in analyzing care, on questions of justification. Disavowing one of the characteristic claims of the first wave of care theory, namely that the ethics of care is incompatible with universalistic moral reasoning or considerations of justice, she asserts that caring well involves making complex moral judgments for which wide-ranging considerations are relevant. Care and justice are not conceptually opposed or incompatible, nor do they hold sway in exclusive domains (private and public, respectively). Rather, they are complementary, neither being complete without the other. The role of care theory is to frame issues that justice has identified more concretely and immediately by drawing attention to dependency as a universal feature of human existence. It should inform practices of democratic citizenship through attentiveness to needs and augment the set of liberal values (rights and due process) making citizens more thoughtful, aware, and in short, better democratic citizens.

The trend evident in Tronto's work to bring the ethics of care into closer interaction with standard moral theories is advanced in the most recent literature in care ethics (e.g., Held, 2006; Slote, 2007; Engster, 2007) all of which develop aspects of her programmatic reconstrual. The notable shift in recent discussions of the ethics of care allows universality to be recognized as central to moral justification, endorses the augmentation of its scope to the public domain, is skeptical of the dichotomy of reason

and emotion, and treats justice, rights, and autonomy as indispensable moral concepts, without relegating care to a subordinate role. Several theorists (Held, 2006; Slote, 2007) argue that the capacity to care and be cared for provides a preferable criterion of moral status to autonomy, the candidate of choice of traditional universalist theories. Some also hold that care is the more inclusive, or causally fundamental notion (Held, 2006; Engster, 2007), presupposed by justice, rights, and autonomy (but not vice versa). Respectful relationships, required by justice, depend on a social fabric of trust and connectedness; families and intimate relations more generally are bases of wider social cohesiveness.

Reclaiming these foundational liberal concepts also involves reconceptualization in terms of care theory. This may involve a curtailing of their form or scope: for example, Slote (2007) argues that the liberal account of autonomy should be significantly weakened to exclude the requirements of critical vigilance of one's ends, because this puts at risk relationships and feelings. More positively, a distinctive care theory of obligation that focused on vulnerabilities stemming from dependency rather than promoting freedoms could lead to an augmented theory of obligation better able to protect directly activities aimed at meeting basic needs and functioning such as education and health-care (Engster, 2007; Held, 2006).

Educational Applications

Contemporary moral education has been strongly influenced by justice-oriented positions. Hence, the focus has been on the twin concerns of fair distribution of educational resources guided by the goal of ensuring equal opportunity, and on what justice requires of educational practice, in order to support and sustain a culture of justice, namely the developing of autonomous and well-informed citizens of a constitutional democracy. Care theorists maintain that the latter emphasis especially has been detrimental to the cultivation of responsiveness and receptivity. Noddings' proposals for pedagogical reforms of the educational system aim to rectify this. The goal of education is to produce caring people. Teachers should model caring behavior by interpreting and reflecting students' conduct in the best possible light. In addition, they should ensure that assignments cater to their interests so that they are willingly received and executed. At the policy level, curriculum requirements, especially in middle and high school, should not force students to take classes they do not find interesting, for example, algebra and geometry, but should reflect the dominance of care, both with respect to content and organization.

These reflections usefully emphasize a (recently) neglected but important pedagogical concern, namely the manner of education. Attentiveness and sensitivity no doubt enhance the experience of learning. But the suggestion that students'

responses be the sole criterion of adequate teaching, and that their current interests be allowed to determine the content of the curriculum raises legitimate concerns, not the least of which is that it is unlikely to result either in the development of crucial capabilities or of caring tendencies.

Later-care theory, in extending its scope, and accommodating autonomy, allows for more promising applications. Its recognition of dependency and vulnerability highlights the development of basic capabilities and functions as crucial preconditions both of well-being and the development of autonomy. Its insistence that those in authority meet standards of empathic care should have a salutary impact both on pedagogical practices and on educational results. Empirical research (Eisenberg, 1992) suggests that children who are treated respectfully display both responsibility for their behaviors and caring tendencies as opposed to those treated in an authoritarian manner who tend to behave heteronomously and have little empathy with others.

A synthesis may be expected to yield theoretical results as well. Improved accounts of both autonomy and dependency in moral development, avoiding, on the one hand, excessively strong construals of autonomy, and on the other, counterintuitive claims that assumptions of inevitable dependency should drive educational thinking, would focus on the common ground of capability development as a central educational aim. One way of putting the point is that autonomy-facilitating education would be a (circumscribed) part of capability-facilitating education (Engster, 2007; Nussbaum, 2000; Brighouse, 1998).

With respect to broader issues like the distribution of resources, the case for synthesis is even more obvious. From the perspective of theories of justice, education is privileged because it is crucial for equality of opportunity. This is attainable only if children are provided with access to institutions that foster the development of the capacities that enable them to live rich and rewarding lives, including autonomy-related ones. A broader-based ethics of care, concerned with providing an education that develops central capacities with the aim of ensuring that people do not suffer avoidably from their vulnerabilities, would, in practice, support justice-based endeavors.

Bibliography

- Barry, B. (1995). *Justice as Impartiality*. Oxford: Oxford University Press.
 Brighouse, H. (1998). Civic education and liberal legitimacy. *Ethics* 108(4), 719–745.

- Eisenberg, N. (1992). *The Caring Child*. Cambridge, MA: Harvard University Press.
 Engster, D. (2007). *The Heart of Justice. Care Ethics and Political Theory*. Oxford: Oxford University Press.
 Friedman, M. (1987). Beyond caring: The de-moralization of gender. *Canadian Journal of Philosophy* 13, 87–110.
 Gilligan, C. (1982). *In a Different Voice*. Cambridge, MA: Harvard University Press.
 Grimshaw, J. (1986). *Philosophy and Feminist Thinking*. Minneapolis, MN: University of Minnesota Press.
 Gutmann, A. (1995). Civic education and social diversity. *Ethics* 105(3), 557–579.
 Held, V. (2006). *The Ethics of Care. Personal, Political, and Global*. Oxford: Oxford University Press.
 Herman, B. (1993). *The Practice of Moral Judgment*. Cambridge, MA: Harvard University Press.
 Noddings, N. (1984). *Caring: A Feminine Approach to Ethics and Moral Education*. Berkeley, CA: University of California Press.
 Nussbaum, M. (2000). *Women and Human Development: The Capabilities Approach*. Cambridge: Cambridge University Press.
 Okin, S. M. (1989). Reason and feeling in thinking about justice. *Ethics* 99(2), 229–249.
 Okin, S. M. (2004). Justice and gender: An unfinished debate. *Fordham Law Review* 72(5), 1537–1567.
 Slote, M. (2007). *The Ethics of Care and Empathy*. London: Routledge.
 Tronto, J. C. (1993). *Moral Boundaries. A Political Argument for an Ethic of Care*. New York: Routledge.

Further Reading

- Clement, G. (1996). *Care, Autonomy, and Justice*. Boulder, CO: Westview Press.
 Darwall, S. (2002). *Welfare and Rational Care*. Princeton, NJ: Princeton University Press.
 Goodin, R. E. (1985). *Protecting the Vulnerable: A Reanalysis of Our Social Responsibilities*. Chicago, IL: University of Chicago Press.
 Held, V. (ed.) (1995). *Justice and Care*. Boulder, CO: Westview Press.
 Jaggar, A. M. (1995). Caring as a feminist practice of moral reason. In Held, V. (ed.) *Justice and Care*, pp 179–202. Boulder, CO: Westview Press.
 Kittay, E. F. and Meyers, D. T. (eds.) (1987). *Women and Moral Theory*. Lanham, MD: Rowman and Littlefield.
 Kohlberg, L. (1973). The claim to moral adequacy of a highest stage of moral judgment. *Journal of Philosophy* 70, 630–646.
 Larrabee, M. J. (ed.) (1993). *An Ethic of Care: Feminist and Interdisciplinary Perspectives*. New York: Routledge.
 Macedo, S. (1995). Liberal civic education and religious fundamentalism: The case of God v. John Rawls? *Ethics* 105(3), 468–496.
 Noddings, N. (1992). *The Challenge to Care in Schools: An Alternative Approach to Education*. New York: Teachers College Press.
 O'Neill, O. (1996). *Towards Justice and Virtue*. Cambridge: Cambridge University Press.
 Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
 Sevenhuijsen, S. (1998). *Citizenship and the Ethics of Care*, Savage, L. (trans.). New York: Routledge.

Lifelong Learning

J Halliday, University of Strathclyde, Glasgow, UK

© 2010 Elsevier Ltd. All rights reserved.

Policies for Lifelong Learning

The concept of lifelong learning seems to be unproblematic. It has been theorized extensively by Dewey (1966) and others concerned that too much emphasis is placed on learning at school and too little on learning in other places and at other times. The concept became problematic when it formed part of the focus for educational and social policy in the 1960s. It became unclear whether lifelong learning was the same as lifelong education or whether a new term was required which referred to a desirable type of society which came to be known as a learning society. It also became unclear whether the concept referred to learning beyond school or as a new type of master concept which transcended the preschool, school, postschool sectors of education altogether. Wain (2004), who is an important current theorist of lifelong learning, terms these references the minimalist and maximalist conceptions, respectively. As contributors to Aspin *et al.* (2001) note, inevitable conceptual conflict was present from the introduction of the concept of lifelong learning into social and educational policy worldwide.

In the first section of the article, some origins of policies for lifelong learning are explored. It is argued that policies with a humanistic concern for social justice shifted toward economic rationalism and the idea developed that lifelong learning primarily equips people with the skills needed to compete in a globalized and supposedly ever-changing series of workplaces. The section explores some of the issues arising from this shift, including the issue of whether lifelong learning should be more than an enhanced form of adult education (the minimalist conception). The section also explores whether such learning might be seen as the basis for a different kind of society within which a range of established dichotomies are no longer appropriate. For example, a humanistic concern with social justice need not necessarily be opposed to a concern with economic efficiency. Formal and informal learning, school and postschool institutional structures may be reformed (the maximalist conception).

The concept of lifelong education was proposed by both the United Nations Educational, Scientific and Cultural Organisation (UNESCO) and the Council of Europe “as involving a fundamental transformation of society so that the whole of society becomes a learning resource for each individual” (Cropley, 1979: 105). The vision outlined in the Faure Report *Learning to Be* (UNESCO, 1972) was that well-organized lifelong education would enable

all citizens to participate fully in a more just and egalitarian society. The Council of Europe saw lifelong education as a means of promoting European integration through preservation and renewal of the European cultural heritage. Both preservation and renewal were seen as public and private goods realized through new forms of educational provision. For UNESCO and the Council of Europe during the 1970s and early 1980s, the main focus for policies concerned with lifelong education was greater social justice.

Through such policies the concepts of lifelong learning, lifelong education, and a socially just learning society became conflated. The operational details of policy which depended on these concepts were not clear and a minimalist conception of lifelong learning prevailed. The term lifelong learning came to be used as an umbrella term for all those sectors of education not primarily concerned with compulsory schooling – further, higher, recurrent, adult, continuing education. Primary and secondary schooling was not generally conceptualized under the lifelong umbrella.

Education and learning are not the same however and a socially just society is not necessarily the product only of education or learning policy. Forms of learning are not equally worthwhile and as a minimum educational requirement it is important to have some idea of the kind of things that are worth learning. Politically too it is important to have some idea of the kind of things that are worth supporting with public money and other things that are best left to private preference. Moreover while lifelong learning might be seen as a means of countering the exclusion suffered by those who do not succeed at the school curriculum, such a vision depends upon there being a system of qualifications that values learning beyond school as much as learning in school. The so-called academic vocational divide may be seen to be based on the difference in value placed on what are often supposed to be two types of learning. This supposition persists, further complicating the conceptual complexities that surround a humanistic conception of lifelong learning. For many people success in an academic curriculum at school forms the gateway to successful and rewarding work whereas for others success in vocational subjects after school does not.

At the same time as these humanistic conceptions of lifelong learning were becoming embedded within the policy discourse at an international level, neoliberalism was gaining increasing political support in Britain

and elsewhere as a guide to education policy in general (Halliday, 1990). The idea that globalization produces such rapid changes in the world of work that learning must be ongoing to cope with it offers an attractive alternative and superficially coherent conception of lifelong learning which appears to bypass the distinction between education and learning. From around the early 1990s, this alternative conception has come to the fore.

According to this conception, it is the individual's responsibility to learn throughout life so that they and the State may prosper economically. Yet the learning opportunities on offer are largely controlled for them within a minimalist conception of lifelong learning. The market in which they are supposedly competing is also beyond their control according to the logic of globalization. The question of what to learn seems to be answered in terms of what the market requires. However of course, the requirements of the market must be unknown according to this conception which depends upon belief in an uncertain globalized future. Hence, this conception has been subjected to some serious sustained criticism.

Conceptions of lifelong learning are highly contested and normative. They necessarily involve ideas not only of what should be learned but also the type of society learning is supposed to encourage. The replacement by economic rationalism of the normative force of the earlier humanistic conception of learning for social justice is seen by many (Aspin *et al.*, 2007) to be in need of urgent reversal. A strong theoretical attempt has been made to reclaim the concept of lifelong learning from its neoliberal dominance within current policy in some parts of the world. There is by no means one common voice however. For example, British neoliberalism may be contrasted with French humanistic approaches. Social justice and economic efficiency are not necessarily incompatible goals for policies concerned with lifelong learning and in the following section a summary of the work of some leading social theorists is given in an attempt to show how they might be reconciled.

While none of the major theorists whose work is discussed in the following section is primarily concerned with the concept of lifelong learning or a learning society, they do shed light on how these concepts could become embedded within the kind of society each theorist thinks is desirable or possible. Wain (2004) is one among many (Usher and Edwards, 1994) who argue that today's world is in a postmodern condition and that in order to make sense of the concept of lifelong learning, it is necessary to see how the modernist assumptions of many current education systems are irrelevant to postmodernism. Hence for Wain (2004: x), it is necessary to move beyond modernist assumptions in which social justice and economic efficiency are seen as competing goals and to look instead at "the postmodern turn towards a non ideological world (with the collapse of any viable alternative to capitalism

and liberal democracy)." For this task Wain examines the work of some leading social theorists and a summary of part of his examination is given below.

MacIntyre's Educated Public

MacIntyre diagnoses a crisis in modernity as an inability to reconcile two tasks set for modern education systems: that of socializing the young and that of rendering them autonomous individuals able to go beyond and contribute to the reordering of that society, independent of it. He contends that the two purposes are only possible where there already exists what he calls an educated public or what others would call a learning society. For MacIntyre (1987: 18) such a society requires the existence of "a tolerably large body of individuals educated both into the habit and the opportunity of active rational debate, to whose verdict appeal is being made by the intellectual protagonists" and who "in their communication with one another . . . must recognise themselves as constituting a public". It also requires "shared assent, both to the standards by appeal to which the success or failure of any particular thesis or argument is to be judged, and to the form of rational justification from which those standards derive their authority" (MacIntyre, 1987: 19).

His notion of a practice is central. This notion provides an answer to the question of what ought to be learned. Learning a practice is similar to an apprenticeship and MacIntyre (1981) recognizes that there are many examples where people learn to do things under decreasing supervision of a master who provides the standards and whose authority must be recognized for learning to take place. The aim of such apprenticeship is to enable learners to come to love the practice so that they engage in it for the sake only of achieving the excellences internal to it and not primarily for some external reward. Practices do however enable economic productivity as many practices form the basis of paid work. Rather than completing a job in the minimum time compatible with being paid, a good practitioner completes a job with due regards to the standards of excellence internal to that practice. In many cases, it would mean that in the long term remuneration from the work is increased.

The rules and standards of the practice are not fixed however. Practices have traditions. Moreover, no one is just one type of practitioner. Rather at any time people should seek to raise the standards of excellence within practices in which they are involved still higher. A sense of purpose provides the learner with a life that has the unity of an unfolding narrative for MacIntyre. On MacIntyre's view, there cannot be one set of practices into which all learners can be initiated for practices themselves are varied and localized. For him a learning society could consist of a number of small-scale learning societies each with its own

forms of practical reasoning about the ends to which it should serve.

In this way, MacIntyre wants to preserve contesting voices in a pluralistic world yet enable them to communicate with each other and theorize each others purposes and methods. The spirit of such communication is conflict and combat, and the framework of communication to be sought is one of constraint and agreement to disagree on fundamentals. MacIntyre's work is highly relevant to theorists of lifelong learning because only a small number of practices can be learned at school. Communities themselves including working communities provide opportunities for learning far more. It is through practices that the dual aims of enculturation and independent practical reasoning can be achieved and a conception of social justice realized that depends upon the recognition that others are bound to differ in their view of the human good.

Habermas

Unlike MacIntyre, Habermas is sympathetic to both social democratic liberalism and modernism through the recovery of a critical public sphere. For Habermas (1988) late modernity is characterized by an increasing colonization of what he calls the life world of ordinary people by the system world of economic and political institutions. Like Dewey, he thinks that the dominance of learning at school is unhelpful and that such learning is part of the system world. He would like to see a reversal of such colonization. He argues for what he calls communicative action guided by a hypothetical notion termed an ideal speech situation. An ideal speech situation is one in which all participants have equal access to the decision-making arena and truth is determined through rational argumentation alone. Even though such a situation is hypothetical, the fact that it can be envisaged is sufficient to provide normative criteria for what a desirable society would be like for Habermas.

Habermas sees the increasing systematization of the life world as being in the interests of economic efficiency but going against a view of a socially just democratic society. For him, theory and the culture industry have come to act together to obliterate the need for a critical public sphere. Habermas (1989: 47) sees the media not so much or at all concerned with informing a critical public sphere as catering for increased sales driven by the need for relaxation and entertainment. The culture industry "produces suitable material for every taste and ability rather than upgrading the abilities of those previously excluded from it" (Habermas, 1989: 166). In this way, publicity loses its critical function and the possibility even for schooling to combat the pernicious effects of the media is unlikely. What is needed is some means of establishing or re-establishing a

critical public sphere within which rational debate and argumentation can take place.

Habermas has come to accept the idea of multiple publics and that the system and life worlds are dynamically interdependent. He holds on however to the necessity of a deliberating critical society that

no doubt abounds just as much in conflicts and meaning generated forms of life. . . . a secularised society that has learned to deal with its complexity consciously and deliberately. . . . the communicative mastery of these conflicts constitutes the sole source of solidarity among strangers – strangers who renounce violence and, in the cooperative regulation of their common life, also concede one another the right to remain strangers. (Habermas, 1998: 308)

For this to happen, Habermas relies upon an undisturbed life world capable of wild initiatives which disturb the system world. For Habermas and others (Field, 2006), it is through unpredictable informal initiatives that a learning society in the form of a critical public is enabled. While it might seem fanciful to expect that governments could actually promote the taking of wild initiatives out with the sphere of formal educational institutions, it could be hoped that governments might do nothing to discourage their formation and development, that is, what Habermas means by undisturbed subjects. Wain (2004) points to Japan as an example of a society with a government that supports forms of informal learning through exemptions, financial assistance, tax breaks, and other policies.

In Britain and elsewhere, the Japanese example has not been followed. Rather the possibility for informal learning depends crucially on how much money individuals have to support their learning. Those who already have acquired sufficient money through the primary opportunity to secure positional advantage through school-based examinations are in an excellent position to take advantage of opportunities for learning beyond school. Others are not in a financial position to do this. What has been left out is the Deweyan idea that learning together educates and what emerges according to Field (2006) are successive attempts to govern lifelong learning rather than letting it grow. A concern with social justice only features in policies for lifelong learning to the extent that an economically prosperous society is more likely to be a socially just one through government intervention. Such intervention is not likely to encourage the disruption of the system world by the life world however. Rather it may lead to what Wain calls a panopticon society.

A Panopticon Society

A panopticon is a structure which enables people to be incarcerated without knowing when they are being watched. In effect they may not be being watched at all but the

possibility that they might be is sufficient to deter them from doing something that is not considered normal. The possibility that the absolute power of government can be released upon individuals at any time is sufficient to convince most of them to accept the normalization required of them. If they should do the unexpected, then an educational institution serves to correct the deviation from the norm. Foucault's (1977) description of the emergence of modern institutions such as educational institutions may be seen precisely as a way governments seek to prevent the disruption of the system world. Whereas informal learning always risks such disruption, mechanisms of curriculum development and management within institutions control formal learning.

There are three primary techniques of control: hierarchical observation, normalizing judgment, and the examination. Within educational institutions, control can be maintained through forms of panoptic incarceration which disables people from knowing when they are under observation. Thus, they must act as if they were being observed. Moreover, they can be judged against what is regarded as normal through a series of techniques of which the examination or formal assessment is most powerful. For those who fail at such examinations, records of achievement can be used as means of getting them to confess what they thought they should be knowing and measuring themselves up against the criteria that they have set for themselves (Usher and Edwards, 1994). The record of achievement may be seen as a powerful means of surveillance of formal learning. It formalizes the informal through detailed records of the extent to which people are normalized into what is expected of them by the system world, that is, in contrast to encouraging the disruption of the life world by resisting the tendencies to formalization. While theoretically both maximalist and minimalist conceptions of lifelong learning can provide specific challenges to the normalizing agenda began at school, a maximalist conception is more likely to do so.

Theories of Learning

Sfard (1998) argues that two basic ideas have underpinned many theories of learning. The first is the idea of learning as acquisition and the second is learning as participation. Learning as acquisition dominates schooling. The idea that children should go to school to acquire knowledge, skills, and attitudes that will help them cope with later life and earn a living in a productive way is deeply held. A maximalist conception of lifelong learning challenges this idea however. The idea that people learn through participating at work, at home, in pastimes of various kinds, through relationships throughout life seems much more appropriate. The former idea is based on an even more fundamental idea that it is first necessary to acquire something in order to live. The latter is based on the idea

that living itself under the right conditions is educative (Dewey, 1966). On the latter view education is not so much a matter of a state of socially approved knowledge, skills, and attitudes. Rather it is a characteristic of a type of journey.

On the former view products are acquired in relatively standard contexts of school or other types of educational institution. These contexts follow fairly standard patterns of tables, chairs, paper, computers, pens, books, and so on. On the latter view learning takes place in a myriad of settings involving a multiplicity of tools artifacts and special relationships. Learning as participation emphasizes contextuality. It also emphasizes the important role of judgments in effecting successful action. Very often, learning is not written down or expressed verbally. Rarely does it make sense to say that something has been acquired.

For Hager and Halliday (2006) a recovery of informal learning for everyone is essential both for economic efficiency and social justice. These researchers develop a theory of informal learning as "a developing capacity to make sound contextually sensitive judgements." For them learning is best enabled when the learner has opportunity to participate in a wide range of practices which involve the multiplicity of artifacts and tools previously mentioned. If formal learning is defined as that which takes place intentionally within educational institutions of one sort or another, then creating opportunities for informal learning is central to any robust account of lifelong learning in the maximalist sense. Arguments in favor of such recovery including the ones advanced in this article are widely supported (Coffield, 1999; Field, 2006). Yet such arguments are often resisted at the policy level. It is worth speculating why this might be the case.

First, the idea that things can be learned once and for all in a relatively stable context for future application in a multiplicity of contexts appears economically efficient, enables control of what is learned, is amenable to standardized forms of assessment, and enables a dominant economic rationalist view of learning. Second, it is much easier to provide mechanisms for accounting for public money when a supposed product of public investment can be measured and tested. Third, the metaphor of learning as acquisition is firmly embedded within ordinary discourse and there is some agreement on the kind of things that ought to be learned in formal educational institutions. Even if resources were redirected away from formal learning toward providing opportunities for informal learning of a variety of sorts, the question would arise as to what kind of opportunities should be provided. It is possible to learn to be a bricklayer or a thief – which is most worthwhile? To put the question less provocatively, should the state provide public money to provide opportunities for a few people to learn to be bricklayers or more people to learn to play tennis or some people to learn philosophy or everyone to participate in the political process or what?

MacIntyre provides some such guides to answers to such questions through his distinction between practices and institutions. Practices are concerned with internal goods as explained earlier, whereas institutions are characteristically and necessarily concerned with external goods. Even though practices and institutions are intimately related (MacIntyre, 1981: 194), it is clear that practices are to be preferred. That is because doing something well for the sake of that action and the benefits it brings to everyone is preferred to doing something only for the sake of something else and the benefit that brings individually. So for MacIntyre, learning as acquisition should not be encouraged.

On this view governments might support those forms of voluntary associations open to everyone through which authority of experts is recognized without recourse to standards and goods external to those associations. Hobbies, clubs, societies, crafts, sports, informal political associations, and games would be examples of arrangements that enable practices to flourish. Moreover while work practices are always tied up with institutions and external goods to some extent, it is possible to recognize and encourage the practical aspects of organizations in the interests of learners. An increasing privatization of what could be communal spaces also works against MacIntyre's account of a desirable or learning society that promotes lifelong learning.

The outcomes of lifelong learning remain indeterminate and individualized. In many instances, it must remain unclear why some people participate in some ways and others in other ways. Means ends rationality is not a central guide to the learning that might be encouraged. Learning is opportunistic; throughout life opportunities arise to participate in different ways without there being any obvious guides other than the ones provided by MacIntyre as to which opportunities should be grasped. Hager and Halliday (2006) use the notion of wisdom as an indicator of what it might mean to be able to go on learning in a measured way. For MacIntyre the notion of the narrative unity of a life performs a similar function. The view that learning is acquisition through a series of forced encounters in formal educational institutions leads toward a panopticon society according to Hager and Halliday (2006). Such a bleak prospect might be avoided by recovering informal learning and making such recovery a central part of policies for lifelong learning.

Bibliography

Aspin, D. N., Chapman, J., Hatton, M., and Sawano, Y. (eds.) (2001). *International Handbook of Lifelong Learning*. Dordrecht: Kluwer.

- Aspin, D. N. (ed.) (2007). *Philosophical Perspectives on Lifelong Learning*. Dordrecht: Springer.
- Coffield, F. (1999). *Speaking Truth to Power*. Bristol: Policy Press.
- Cropley, A. J. (ed.) (1979). *Lifelong Education: A Stocktaking*. Hamburg: UNESCO Institute for Education.
- Dewey, J. (1966). *Democracy and Education* (originally published, 1916. New York: Macmillan). New York: Free Press.
- Field, J. (2006). *Lifelong Learning and the New Educational Order*. Stoke on Trent: Trentham.
- Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison*. London: Penguin.
- Habermas, J. (1988). *The Theory of Communicative Action, Vol 2. Lifeworld and System: A Critique of Functionalist Reason*. (McCarthy, T., trans.). Cambridge: Polity Press.
- Habermas, J. (1989). *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society* (Burger, T. with the assistance of Frederick Lawrence, trans.). Cambridge: Polity Press.
- Habermas, J. (1998). *Between Facts and Norms*. Cambridge, MA: MIT Press.
- Hager, P. and Halliday, J. S. (2006). *Recovering Informal Learning: Wisdom, Judgement and Community*. Dordrecht: Springer.
- Halliday, J. S. (1990). *Markets, Managers and Theory in Education*. London: Falmer.
- MacIntyre, A. (1981). *After Virtue*. London: Duckworth.
- MacIntyre, A. (1987). The idea of an educated public. In Haydon, G. (ed.) *Education and Values*, pp 15–36. London: Institute of Education.
- Stard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher* 27(2), 4–13.
- UNESCO (1972). *Learning to Be (The Faure Report)*. Paris: UNESCO.
- Usher, R. and Edwards, R. (1994). *Postmodernism and Education*. London: Routledge.
- Wain, K. (2004). *The Learning Society in a Postmodern World: The Education Crisis*. New York: Peter Lang.

Further Reading

- Coffield, F. (ed.) (2000). *The Necessity of Informal Learning*. Bristol: Policy Press.
- Colley, H., Hodgkinson, P., and Malcolm, J. (2003). *Informality and Formality in Learning*. London: Learning and Skills Research Centre.
- Collins, R. (1979). *The Credential Society: An Historical Sociology of Education and Stratification*. Orlando, FL: Academic Press.
- Edwards, R. (1997). *Changing Places? Flexibility, Lifelong Learning and a Learning Society*. London: Routledge.
- Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education* 26(2), 247–276.
- Fejes, A. and Nicoll, K. (eds.) (2008). *Foucault and Lifelong Learning: Governing the Subject*. London: Routledge.
- Field, J. (2005). *Social Capital and Lifelong Learning*. Bristol: Polity Press.
- Halliday, J. S. and Hager, P. (2002). Context, judgement and learning. *Educational Theory* 52(4), 429–445.
- Holford, J., Jarvis, P., and Griffin, C. (eds.) (1998). *International Perspectives on Lifelong Learning*. London: Kogan Page.
- Jarvis, P. (2007). *Globalisation, Lifelong Learning and the Learning Society: Sociological Perspectives*. London: Routledge.
- Sutherland, P. and Crowther, J. (2006). *Lifelong Learning: Concepts and Contexts*. London: Routledge.

Post-Structuralism and Education

M Mason, Hong Kong Institute of Education, Hong Kong, People's Republic of China

M Clarke, University of New South Wales, Sydney, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

Structuralism, Post-Structuralism, and Postmodernism

Structuralist analyses of education and other social phenomena have typically been characterized by: a concern with wholes, understood as complex structures; the notion of embeddedness, by which is implied the nesting of particular phenomena within broader networks or structures; ideas of presence and absence, which stress that what may be present or visible is usually dependent on deeper structures or on a host of complexly networked factors (as in the view that Charlene's failure in math is because, as a black girl from a low socioeconomic status background, she is a member of the underclass in a society whose education system reflects the interests of the elite); and a focus on relationships, among phenomena and the structures which underlie them. Typical of this theoretical approach in sociological analysis is, for example, the situation, conceptualization, understanding, and explanation of schooling within its wider societal context. Criticisms of structuralism, for its sometimes rigid, static, ahistorical, and reductionist perspectives on social phenomena and, in some cases, for its assumptions of the objectivity of social phenomena, its positing of abstract macro-level generalizations about human behavior, and its employment of the empirical methods of the natural sciences to generate these laws, have contributed to the development of post-structuralist analyses of education and other social institutions. Such post-structuralist analyses nevertheless display significant levels of continuity with structuralism, not least in their shared critique of humanism's rational and autonomous subject and their concern to de-center that subject. Both analyses would, for example, pay less attention to individual factors that might contribute to Charlene's failure in math – that she might be unmotivated, lazy, or not particularly capable intellectually – and more to underlying structural causes such as those to do with her socioeconomic background, ethnicity, and gender, and how these structures are related to broader structures associated with the distribution of opportunity, wealth, and privilege in her society. Post-structuralism would, however, treat these structural factors in a less systemic and a less deterministic manner than would structuralism, and would be more focused on the complex and probably recursive interplay among the networks of economic, political, social, and cultural factors involved.

Our concern here is with post-structuralism, a term frequently conflated with postmodernism, not least because of the theoretical perspectives they share. Post-structuralism shares postmodernism's skepticism of meta-narratives (Lyotard, 1984), of transcendental or totalizing theories of truth in the social sciences, principally of the paradigm of enlightenment rationality and its faith in scientific reason as the bearer of the promise of emancipation through knowledge. The linguistic turn associated with post-structuralism and postmodernism does not, as Peters and Wain (2003: 65) put it, "warrant the assumption of a metalinguistic neutrality or foundational epistemological privilege." Post-structuralism shares the postmodern perspective that is, in Bauman's conceptualization, concerned with the unmasking of the illusions of modernity: the essence of the postmodern approach to epistemology and ethics lies in "the rejection of ... the philosophical search for absolutes, universals and foundations in theory" (Bauman, 1993: 4). Our search for these absolutes, universals, and foundations has probably been tempered by our realization, as a consequence of the multicultural spaces we now inhabit in an increasingly globalized world, that ours is a plural world, with a diversity of perspectives and claims to truth, beauty, and goodness. Post-structuralism shares with postmodernism the latter's challenges to the claims associated with the enlightenment that our knowledge of society is holistic and cumulative; that we can attain rational knowledge of society; that such knowledge is universal and objective; and that acting upon social scientific knowledge can lead to emancipation and social upliftment (McLennan, 1994). Post-structuralism shares with postmodernism a genealogical ancestor in Nietzsche's (1886/1990) view that truth is little more than the solidification of old metaphors, that it is neither foundational nor universal, but merely the sedimented beliefs of a particular cultural worldview. Nietzsche's (1886/1990) claim of the death of God was, after all, a claim about the absence of any universal, foundational, or transcendental source of existential meaning. Derrida's deconstruction, a major philosophical thrust of post-structuralism, is primarily concerned with the unpacking of these old metaphors which have been solidified into truth, deconstructing them into multiple realities articulated as the discourses of particular knowledge communities, each with its own truth claims that are likely to be incommensurable with the discourse of any other community – hence the linguistic turn referred to above. Foucault's post-structuralist

approach to historical analysis, his genealogical or archaeological approach to the social sciences, accordingly dispenses with notions of linearity and causality and focuses instead on knowledges that is local, discontinuous, disqualified, and illegitimate (Foucault, 1984). Rorty's pragmatism suggests, therefore, that rather than seeking ahistorical and transcontextual rules, philosophy should search for new forms of description, rather than prescription, in a world marked by such radical contingency.

At the most general level, post-structuralism differs from postmodernism in the different theoretical object of the study of each. As Peters and Wain (2003) suggest, the object of study of the former is structuralism; while that of the latter is modernism, associated with which are postmodernism's critiques of modernity.

Post-Structuralism, Critical Theory, and Reason

In approximately the last quarter of the twentieth century, the so-called post discourses – including those of post-structuralism, postmodernism, postcolonialism, and critical feminist discourses – have offered various critiques of the approaches to critical theory typical of the Frankfurt School, which have extended these into new territory. Scholars associated with the Frankfurt School had, at the risk of oversimplification, been committed to reclaiming a notion of reason more consistent with Enlightenment notions of universal reason from the typically instrumentalist forms of reason into which the formations and practices of modernity had in their view degenerated. The post discourses have challenged this continuing commitment to reason and share a skepticism of the attempts to reclaim any notion of universal reason. In considering these more recent approaches to the critical theorization of education, we focus more substantially on post-structuralism and on the work of Michel Foucault, who, although he never referred to himself as a post-structuralist, has, through his work on discourse, power, and disciplinary bodily training, probably had more impact on the critical theorization of social institutions – not least education – than any theorist since Marx and the leading members of the Frankfurt School.

In his *What is Enlightenment?* (1984/1997), Foucault reflected on the history and ethos of the present through a consideration of questions such as the meaning of critique and of enlightenment. He framed his reflections in the form of a dialog with Kant, who had 200 years earlier written a text with the same title, in which he too engaged with notions of enlightenment, critique, and the meaning of the present moment in history. Foucault characterized the difference between his own and Kant's responses to these questions in terms of a contrast between a focus on critique that circumscribed the activities of reason and his

own view of critique as a practice of exposing the limits of what Judith Butler (2002: 217) refers to as the epistemological horizon, or, critique as exposing the limits of reason. Perhaps as importantly, he modeled a way of thinking as an ongoing dialog, involving responses to the challenges and questions pondered by Kant and his successors – a dialogical approach that is central to the concerns and practices of post-structuralist critical educators.

Post-Structuralism and Language, Knowledge, and Truth

Post-structuralism and many other post discourses share a substantial research focus with Habermas and the Frankfurt School on language. But, whereas for Habermas, language is, at least potentially, a vehicle for transcending ideologies and engaging in a free and equal exchange of communicative reason in an ideal speech situation, for post-structuralist-influenced critical pedagogy, language is a socially shaped resource, steeped in culturally and historically sedimented attitudes, values, and assumptions, which precedes and exceeds any single individual. It is these coagulations of social meaning that comprise discourse, in the sense of "ways of being in the world, or forms of life which integrate words, acts, values, beliefs, attitudes and social identities" (Gee, 1996); and it is these discourses that, in post-structuralist theory, produce, or become constitutive of, social reality. Moreover, rather than regarding language as providing a transparent mirror of nature (Rorty, 1979), where words simply and clearly represent objects or concepts, post-structuralism, in common with its forebear, structuralism, sees meaning as deriving from differences within the linguistic system, so that dog means not cat or not horse, sacred means not profane, with the consequence that words or concepts are always defined in relation to other words, *ad infinitum*. Following Saussure (1959), linguistic signs act relationally rather than referentially. Hence, for post-structuralism, meaning is always partial and provisional, emerging as it does from an endless process of difference and deferral within the linguistic system, rather than from direct reference to the real world. This linguistic turn has significant implications for key educational notions of truth, progress, and emancipation: post-structuralism's claim "that all truths are textual, that the way we see the world is 'always already' infected by language" (MacLure, 2003: 4), means that language becomes both constitutive and – critically for education – a site of contestation, since there is no unproblematic one-to-one match between language and reality. Hence, the meaning and implications of terms like democracy and freedom invite and merit ongoing debate and negotiation, rather than being settled and simply requiring transmission to future generations.

Far from being committed to reason, then, post-structuralist thought is committed “to discourse, literally narratives about the world that are admittedly *partial*” (Aronowitz, 1987: 103).

Post-Structuralism and Power

The constitutive role of language-as-discourse in relation to knowledge raises issues of power and social (in)justice, insofar as education involves ultimately contestable knowledge, as well as asymmetric relations between holders and seekers of the symbolic capital provided by education (Bourdieu and Passeron, 1990). Questions arise as to who has the right to determine the meanings of key concepts and terms and whether the voices of educators should carry greater weight than the voices of those being educated. Post-structuralist pedagogy thus embraces an ongoing interrogation of the relations between knowledge and power. However, post-structuralism conceives power rather differently to more conventional formulations, seeing it as something that circulates in all human relations; as a potentially productive, rather than just a repressive, force – and as something that can be used “as a point of resistance and a starting point for an opposing strategy” (Foucault, 1978: 101). Foucault mentions the relationship between a teacher and student as an example of such potentially productive power relations (1988). This suggests that if power is an inevitable and potentially productive feature of social relations, then the challenge for educational practice is more one of strategically managing, rather than overcoming or simply ignoring, such power relations. In Foucault’s words,

The thought that there could be a state of communication which would be such that games of truth could circulate freely, without obstacles, without constraint and without coercive effects, seems to me to be Utopia. It is being blind to the fact that relations of power are not something bad in themselves, from which one must free one’s self. I don’t believe there can be a society without relations of power, if you understand them as means by which individuals try to conduct, to determine the behaviour of others. The problem is not trying to dissolve them in the utopia of a perfectly transparent communication, but to give one’s self the rules of law, the techniques of management, and also the ethics, the ethos, the practice of self, which would allow these games to be played with a minimum of domination. (Foucault, 1988: 18)

Post-Structuralism and Subjectivity

Foucault proceeds to draw connections between these games of truth – involving knowledge and power relations – and ethics. Ethics, or, as he constructs it, practices of the

self, draws attention to the issue of subjectivity, or identity, the third component of his three-dimensional grid of intelligibility (Dean, 1994: 194), which post-structuralist theory sees as co-constituted along with knowledge and power relations in social and educational practice. Accordingly, “language is the place where actual and possible forms of social organization and their likely social consequences are defined and contested. Yet it is also the place where our sense of selves, our subjectivity, is constructed” (Weedon, 1997: 21). In this reading, post-structuralist critical education is not just about questioning particular constructions of knowledge, or the social power relations operating in educational or pedagogical practice, but also about questioning or problematizing the creation of particular educational subjects or individuals. For example, albeit at the risk of oversimplification, didactic and inquiry-based pedagogical methods are each likely to produce different types of learners. It is important to recognize that this co-construction of knowledge, power relations, and subjectivity does not occur through the operation of disembodied ideas alone. In particular, post-structuralist theories consider the ways in which emotion and desire play a critical role in educational practice (McWilliam, 1999; Zembylas, 2007a, b), as well as the ways in which discourse as power/knowledge operates on and through the body via disciplinary practices and means of training, as ideas operate and circulate in conjunction with physical, spatial and temporal arrangements, techniques, and practices. We can see this, for example, if we consider how early literacy education operates through bodily, as much as cognitive, training and produces particular disciplined individual subjects (e.g., in the form of pupils who know how to comport themselves in educational settings, particularly in relation to books and other texts) and particular social power relations (e.g., in the form of the teacher’s right to be heard and students’ obligations to listen), as well as constructing and conveying particular bodies of knowledge (e.g., the ability to manage and extract meaning from written texts) (A. Luke, 1992).

Post-Structuralism and Critique

In critical post-structuralist educational research, the focus is generally on a historical, or genealogical, tracing or uncovering of the ways in which knowledge, social power relations, and subjectivity are constructed and inscribed within institutional practices – not so much to determine whether these ways are right or wrong, good or evil, but rather in the spirit of a “recurrent disturbance and negotiation of the numerous paradoxes of [political] life over attempts to conceal, resolve or suppress them” (Connolly, 1998: 124). From this perspective, reason and the practices of rationality, often seen as the essence of criticality, are themselves open to historical, genealogical critique,

rather than being understood as transcendent, universal, and timeless forces for good (communicative reason) or bad (instrumental reason). For some authors, however, the possibility of a critique of reason is unhelpful: Best and Kellner, for example, are skeptical about “the rational critique of rationality, totalizing rejections of totality, and the subjective hubris of dismissing the category of subjectivity” (Best and Kellner, 1991: 281). It might be argued in response to this that Foucault does not call for a rejection of reason but for an acknowledgment that, like many phenomena that humans tend to naturalize, it has taken different forms in the past, and could consequently operate differently in the future. Indeed, one way of characterizing the differences between Habermas and Foucault is that Habermas is interested in the theory of reason, whereas Foucault is interested in the historicity of reason (Hoy, 1998). In this sense, Foucault’s post-structuralist critical theory can be read as complementing that of Habermas.

[Post-structuralist] critical theory conceived genealogically may unmask substantive injustices, but it need not justify this unmasking through the methodological picture of inquiry presented by traditional theory. It need not construe itself only as seeing through illusions and constructing a timeless, context-transcendent theory of rationality from which we can then measure present society. Instead, when practiced genealogically, critical theory can present itself as offering new interpretations. Along the way it may be unmasking previous interpretations. Since what is unmasked is self-interpretation, this unmasking through genealogical critical history can now be seen not simply in traditional epistemological terms as ‘revealing reality’, but also modally as ‘deconstructing necessity’. That is, genealogical research will show that self-understandings that are taken as universal, eternal, and necessary have a history, with a beginning, and therefore, possibly, an end, (Hoy, 1998: 31)

The consequences of this emphasis on the contingent and constructed nature of some cherished ideas, values, and structures of social organization have implications for critique, which in this view should assume an attitude of humility coupled with a ceaseless interrogation, while resisting the temptation to be judgmental: for “judgements operate ... as ways to subsume a particular under an already constituted category, whereas critique asks after the occlusive constitution of the field of categories themselves” (Butler, 2002: 213).

Critical Challenges

Genealogical critique – what Hoy (2004) refers to as postcritique or genealogical deconstruction – in education entails in this view broader, deeper, and more

sustained forms of vigilance than those directed by earlier forms of critical education toward falsehoods or ideologies; for if education as a social practice entails “the triple fabrication of the subject ... [as] not simply an effect of domination but a complex result of practices and techniques of power, knowledge, and ethics” (Dean, 1994: 113), then pedagogical practice accordingly needs to problematize issues and probe horizons in multiple directions. Among the complex critical pedagogical issues facing educators, the following can be considered a small sample:

1. How to acknowledge “minority discourses, diasporic voices, texts and statements that are ‘written out’ and over by dominant institutions” (Luke, 2002).
2. How to address the complex, subtle, ways that power operates through/on bodies and space beyond indoctrination into belief systems (Olssen, 2006).
3. How to acknowledge, challenge, and extend students’ investments in particular discourses and positions (Misson, 1996).
4. How to harness the creative energy of desire and resistance without domestication (Lankshear and Knobel 2006; Misson and Morgan, 2006).
5. How to expand our sense of possibilities in the present rather than just imagining alternative social orders (Clarke, 2008; Connolly, 2002; Sawicki, 1998).
6. How post-structuralist educators might ensure that critical pedagogy does not become just another regime of truth (Ellsworth, 1989; Gore, 1993), as oppressive in its own way as the pedagogies it critiques.

Each of these challenges reflects an overarching problem of how post-structuralist educators are to negotiate the complex intertwining of knowledge, power, and subjectivity at individual, institutional, and societal levels. Each can be seen by these educators as encapsulating Foucault’s challenge: “the endeavour to know how and to what extent it might be possible to think differently, instead of legitimating what is already known ... , to learn to what extent the effort to think one’s own history can free thought from what it silently thinks, and so enable it to think differently” (Foucault, 1985: 9).

Post-Structuralism and Critical Pedagogy

A distinction should be drawn at this stage between critical thinking and critical pedagogy (Burbules and Berk 1999). They of course share an aim to arm people against delusion, but while critical thinking understands such delusions in terms of false or unreasonable assumptions, invalid reasoning and irrational thought processes, post-structuralist and other forms of critical pedagogy understand such delusions in terms of a metaphorical colonization by entrenched and unjust social assumptions, discourses, and practices. Critical thinking, frequently grounded in an epistemological

distinction between facts and values and an ontological focus on the individual, aims to enable individuals to develop as autonomous rational agents; post-structuralist and other forms of critical pedagogy, on the other hand, are skeptical about the possibility of distinguishing facts from values and, because they construe the social world as prior to, and the origin of, the individual, their target is taken-for-granted assumptions and socially received ideas, practices, and arrangements (Popkewitz, 1999). More particularly, critical pedagogy frequently draws on post-structuralist views of language in relation to knowledge, power, and subjectivity to challenge prevailing assumptions and practices in relation to, for example, class, gender, ethnicity, and sexual orientation.

In her important article, "Why doesn't this feel empowering? Working through the repressive myths of critical pedagogy", Ellsworth (1989) makes the telling claim that "while the literature states implicitly or explicitly that critical pedagogy is political, there have been no sustained research attempts to explore whether or how the practices it prescribes actually alter specific power relations outside or inside schools" (1989: 301). In response to this critique, she premises the shift to a post-structuralist critical pedagogy on "[the] recognition . . . that all knowings are partial, that there are fundamental things each of us cannot know – a situation alleviated only in part by the pooling of partial, socially constructed knowledges in classrooms – [which] demands a fundamental retheorizing of 'education' and 'pedagogy'" (1989: 310). This post-structuralist critical pedagogy has been substantially influenced by antiracist and feminist theories, each of which has emphasized that any individual's voice, any individual's knowing, will always be, in the words of Alcoff (1988), partial, multiple, and contradictory. These influences, including, not least, feminism's "reject[ion of] the 'universal subject'" (Luke and Gore, 1992: 5), have rendered problematic earlier forms of critical pedagogy that focused on helping each student to identify his or her authentic voice, even multiple authentic voices, inside him- or herself, as a means to empowerment. Ellsworth has remarked that such a notion of an inner voice ignores the

often contradictory intersection of voices constituted by gender, race, class, ability, ethnicity, sexual orientation, or ideology, . . . [as well as] the fact that the particularities of historical context, personal biography, and subjectivities split between the conscious and the unconscious will necessarily render each expression of student voice partial and predicated on the absence and marginalization of alternative voices (1989: 312).

Worse than this, in Ellsworth's view, conventional notions of dialog assume what is impossible: "rationalized, individualized subjects capable of agreeing on universalizable 'fundamental moral principles' and 'quality of life' that [apparently] become self-evident when subjects cease

to be self-interested and particularistic about group rights" (1989: 316). In a feminist post-structuralist perspective on critical pedagogy, "social agents are not capable of being fully rational and disinterested" (Ellsworth, 1989: 316), constituted as they are by the conscious and the unconscious, by multiple and frequently conflicting social positioning, by their own particularities. Post-structuralist perspectives deny, moreover, the possibility of the existence of any absolute or universal fundamental moral principles. Spivak (1988) refers to the search for such a coherent narrative as counterproductive, asserting instead the importance of persistent critique (1988: 272), as echoed by Butler, above. These perspectives need to be understood, in the words of Luke and Gore, in terms of

post-structuralism and postmodernism, [both of which] take issue with the centuries-long rule of Enlightenment epistemology and the fictions of the individual that it spawned. Both reject the self-certain subject, . . . the fixity of language, and the functionalist order imputed to the social and to theories of the social (1992: 5).

While post-structuralist feminist theories of critical pedagogy share many epistemological and axiological commitments with post-structuralist and postmodern theories of critical pedagogy, there are distinct differences, not least in post-structuralist feminism's assertion of distinct theoretical foundations, such as their grounding of their epistemology "on the foundation of difference, . . . on a politics of location and identity [crucial in] . . . the feminist project of standpoint" (Luke and Gore, 1992: 7), a commitment that in some respects places this discourse at odds with postmodern and post-structuralist rejections of such foundations.

Similarly, while antiracist theories of critical pedagogy share many epistemological and axiological commitments with (post-structuralist and postmodern) theories of critical pedagogy, there are distinct differences, not least in emphasis. Why, asks Allen (2005: 54), has critical pedagogy been based for so long upon class rather than race? "For critical pedagogy to become anti-racist", he asserts, "it will need to be much more serious about the race-radical philosophies of people of colour around the world and move away from the comforts and constrictions of a Marxist Eurocentricity" (Allen, 2005: 54). The arguments of W. E. B. DuBois (1902) in *The Souls of Black Folk* and of Carter G. Woodson (1933) in *The Mis-Education of the Negro* are, after all, frequently cited for their early contributions to the field. Frantz Fanon's (1963/1968) *The Wretched of the Earth* has not been without influence either. Contemporarily, writers such as Gloria Ladson-Billings, bell hooks and Zeus Leonardo have had a substantial influence on the field.

Feminist post-structuralist critical pedagogy is thus "profoundly contextual (historical) and interdependent (social)" (Ellsworth, 1989: 323), as it tries to find "a commonality in

the experience of difference without compromising [the] distinctive realities and effects [of difference]" (Gentile, 1985: 7). Greene (1992) describes it as "relational, practice-centred, contextualized, open-ended, ... demand [ing] confrontations with discontinuities, particularities, and the narratives that embody actual life stories, ... [and] renewed attentiveness to the construction of knowledge and the life of meaning" (1992: x). Luke and Gore premise it in an epistemology that views knowledge as "always provisional, open-ended and relational" (1992: 7). Ellsworth thus suggests a more humble starting point for feminist post-structuralist critical educators in their classroom practice aimed at a kind of communication across differences:

If you can talk to me in ways that show you understand that your knowledge of me, the world, and the 'Right thing to do' will always be partial, interested, and potentially oppressive to others, and if I can do the same, then we can work together on shaping and reshaping alliances for constructing circumstances in which students of difference can thrive (1989: 324).

Conclusion

Post-structuralist critical pedagogy is not without its own internal inconsistencies. (This fact might not, of course, be construed by post-structuralist critical pedagogs as especially problematic.) Ellsworth, for example, claims that "the antagonist [is] power itself, as it is, [for example], deployed within classrooms" (1989: 322). This is inconsistent with Foucault's conceptualization of power, as we considered earlier. He did not believe there could be a society without relations of power, understanding them not as something bad in themselves, from which one must free one's self (Foucault, 1988: 18). As indicated above, he saw power as something that circulates in all human relations, as a potentially productive, rather than just a repressive or antagonistic, force, as a point of resistance and a starting point for an opposing strategy (Foucault, 1978: 101).

Against the post-structuralist/postmodern skepticism of enlightenment notions of truth and rationality in contemporary theories of critical pedagogy, Siegel (1997) has argued that if we were to

give up our 'conservative' understandings of truth, warrant, rationality and rational justification, then we [would be] unable to establish, for example, that apparent victims of marginalization or oppression at the hands of a hegemonic dominant culture are actual victims. Similarly, we [would be] unable to establish that such victimization, even if actual, is a bad thing. In order to establish these judgements as true and/or justified, we must have recourse to conceptions of truth, justification, etc., which provide us with the conceptual resources to establish these claims. Without those resources, there is no possibility of acting, in a

morally motivated and justified way, so as to alleviate the suffering wrought by injustice (1997: 152).

We can, according to Siegel, reject these epistemological presuppositions of evidence, reason, rational justification, and truth "only at the price either of incoherence or arbitrariness. For to reject them *justifiably* is to reject them for reasons which satisfy the very conditions concerning rational justification which one wants to reject" (1997: 147).

Mason (2005) has argued that

we need to talk about the limits of reason, rather than its limitations. *Sophie's Choice*, in the novel by that name, about which of her children to give up to death at the hands of the Nazis is not susceptible to a rational process of decision-making. But that says less about the limitations of reason and more about its limits, the extent of the domain in which it is applicable. ... [It says more] about domains beyond the scope of reason where problems are tractable only and ultimately by moral choices (2005: 313).

Although educational critique has recently sought to incorporate, as we have described, post-structuralist, post-modern, antiracist, and feminist insights, it probably still remains vulnerable to criticism for what C. Luke (1992) has referred to as its attempts to subsume these different perspectives within just such a master narrative of critique, and for what Buckingham (1998) has claimed as an overly strong emphasis on rationalistic analysis and a sometimes simplified worldview of oppressors and dupes. Buckingham notes that the research shows that "it is more productive to relate critical analysis to students' own concerns, tastes and identities rather than engaging in the more abstract analyses of ideology that have traditionally been prevalent" (Buckingham, 2007: 163). He suggests that in an age when digital media texts are a central part of students' experiences beyond the school, the critical understanding sought by teachers "can [also] be developed – in some instances, more effectively and enjoyably – through the experience of creative production" (Buckingham, 2007: 163). This entails an emphasis on constructing, as well as deconstructing, texts in various media so as to develop critical and creative literacy in how textual elements work singly and in concert to produce particular social realities, relations, and identities.

This assertion of the importance of cultural production by students that might describe an alternative to the consumerism-dominated and consumer-oriented cultural production of late-modern societies reflects Maxine Greene's concern with the role of esthetic literacy, of creativity and the imagination, in a post-structuralist feminist pedagogy, and it is with her words that we conclude:

We cannot negate the fact of power. But we can undertake a resistance, a reaching out towards becoming *persons* among other persons ... To engage with our students

as persons is to affirm our own incompleteness, our consciousness of spaces still to be explored, desires still to be tapped, possibilities still to be opened and pursued. At once, it is to rediscover the value of care, to reach back to experiences of caring and being cared for (as Nel Noddings writes) as sources of an ethical ideal. It is, Noddings (1984) says, an ideal to be nurtured through “dialogue, practice and confirmation”, processes much akin to those involved in opening a public sphere. We have to find out how to open such spheres, such spaces, where a better state of things can be imagined; because it is only through the projection of a better social order that we can perceive the gaps in what exists to try to transform and repair It demands the exercise of imagination, enlivened by works of art, by situations of speaking and making I would like to think that this can happen in classrooms, in corridors, in schoolyards, in the streets around (2009: 95).

See also: Critical Theory and Pedagogy.

Bibliography

- Alcoff, L. (1988). Cultural feminism versus post-structuralism: The identity crisis in feminist theory. *Signs* 13(Spring), 405–437.
- Allen, R. L. (2005). Whiteness and critical pedagogy. In Leonardo, Z. (ed.) *Critical Pedagogy and Race*, pp 53–68. Oxford: Blackwell.
- Aronowitz, S. (1987). Postmodernism and politics. *Social Text* 18 (Winter), 99–115.
- Bauman, Z. (1993). *Postmodern Ethics*. Oxford: Blackwell.
- Best, S. and Kellner, D. (1991). *Postmodern Theory: Critical Interrogations*. New York: Guilford Press.
- Bourdieu, P. and Passeron, J. (1990). *Reproduction in Education, Society and Culture*, 2nd edn. London: Sage.
- Buckingham, D. (1998). Introduction: Fantasies of empowerment? Radical pedagogy and popular culture. In Buckingham, D. (ed.) *Teaching Popular Culture: Beyond Radical Pedagogy*, pp 1–17. London: University College London Press.
- Buckingham, D. (2007). *Beyond Technology: Children's Learning in the Age of Digital Culture*. Cambridge: Polity Press.
- Burbules, N. and Berk, R. (1999). Critical thinking and critical pedagogy: Relations, differences, and limits. In Popkewitz, T. and Fendler, L. (eds.) *Critical Theories in Education: Changing Terrains of Knowledge and Politics*, pp 45–65. New York: Routledge.
- Butler, J. (2002). What is critique? An essay on Foucault's virtue. In Ingram, D. (ed.) *The Political: Readings in Continental Philosophy*, pp 212–228. London: Blackwell.
- Clarke, M. (2008). The ethico-politics of teacher identity. *Educational Philosophy and Theory* 41, 185–200.
- Connolly, W. (1998). Beyond good and evil: The ethical sensibility of Michel Foucault. In Moss, J. (ed.) *The Later Foucault*, pp 108–128. London: Sage.
- Connolly, W. (2002). *Identity/Difference: Democratic Negotiations of Political Paradox*, 2nd edn. Minneapolis, MN: University of Minnesota Press.
- Dean, M. (1994). *Critical and Effective Histories: Foucault's Methods and Historical Sociology*. London: Routledge.
- de Saussure, F. (1959). *Course in General Linguistics*. Baskin, W. (trans.). Bally, C., and Sechehay, A. (eds.). New York: Philosophical Library.
- DuBois, W. E. B. (1902). *The Souls of Black Folk*. Chicago, IL: A.C. McLurg.
- Ellsworth, E. (1989). Why doesn't this feel empowering? Working through the repressive myths of critical pedagogy. *Harvard Educational Review* 59(3), 297–324.
- Fanon, F. (1963/1968). *The Wretched of the Earth*. New York: Grove Press.
- Foucault, M. (1978). *The History of Sexuality: Volume 1. An introduction*. Hurley, R. (trans.). New York: Pantheon Books.
- Foucault, M. (1984). *The Foucault Reader*. Rabinow, P. (ed.). New York: Pantheon Books.
- Foucault, M. (1985). *The Use of Pleasure: The History of Sexuality: Volume 2*. Hurley, R. (trans.). New York: Pantheon Books.
- Foucault, M. (1988). The ethic of care for the self as a practice of freedom. In Bernauer, J. and Rasmussen, D. (eds.) *The Final Foucault*, pp 1–20. Cambridge, MA: MIT Press.
- Foucault, M. (1997). *What is Enlightenment? Ethics, Subjectivity and Truth: Essential Works of Foucault 1954–1984* vol. 1. New York: New Press.
- Gee, J. P. (1996). *Social Linguistics and Literacies: Ideology in Discourses*, 2nd edn. London: Taylor and Francis.
- Gentile, M. (1985). *Film Feminisms: Theory and Practice*. Westport, CT: Greenwood Press.
- Gore, J. (1993). *The Struggle for Pedagogies: Critical and Feminist Discourses as Regimes of Truth*. New York: Routledge.
- Greene, M. (1992). Foreword. In Luke, C. and Gore, J. (eds.) *Feminisms and Critical Pedagogy*, pp ix. New York: Routledge.
- Hoy, D. (1998). Foucault and critical theory. In Moss, J. (ed.) *The Later Foucault*, pp 18–32. London: Sage.
- Hoy, D. (2004). *Critical Resistance: From Poststructuralism to Post-Critique*. Cambridge, MA: MIT Press.
- Lankshear, C. and Knobel, M. (2006). *New Literacies: Everyday Practices and Classroom Learning*, 2nd edn. Buckingham: Open University Press.
- Luke, A. (1992). The body literate: Discourse and inscription in early literacy training. *Linguistics and Education* 4(1), 107–129.
- Luke, A. (2002). Beyond science and ideology critique: Developments in critical discourse analysis. *Annual Review of Applied Linguistics* 22, 96–110.
- Luke, C. (1992). Feminist politics in radical pedagogy. In Luke, C. and Gore, J. (eds.) *Feminisms and Critical Pedagogy*, pp 25–53. New York: Routledge.
- Luke, C. and Gore, J. (eds.) (1992). *Feminisms and Critical Pedagogy*. New York: Routledge.
- Lyotard, J.-F. (1984). *The Postmodern Condition: A Report on Knowledge*. Bennington, G. and Massumi, B. (trans.). Minneapolis, MN: University of Minnesota Press.
- MacLure, M. (2003). *Discourse in Educational and Social Research*. Buckingham: Open University.
- Mason, M. (2005). Critical theory and critical pedagogy after postmodernism. In Gur-Ze'ev, I. (ed.) *Critical Theory and Critical Pedagogy Today: Toward a New Critical Language in Education*, pp 300–316. Haifa: University of Haifa, Studies in Education Series.
- McLennan, G. (1994). The enlightenment project revisited. In Hall, S., Held, D., and McGrew, T. (eds.) *Modernity and Its Futures*, pp 327–377. Cambridge: Polity Press/Blackwell/The Open University.
- McWilliam, E. (1999). *Pedagogical Pleasures*. New York: Peter Lang.
- Misson, R. (1996). “Despite all my rage I'm still just a rat in a cage”: The limits of critical literacy. *Fine Print* 18(3), 3–9.
- Misson, R. and Morgan, W. (2006). *Critical Literacy and the Aesthetic: Transforming the English Classroom*. Urbana, IL: National Council of Teachers of English.
- Nietzsche, F. (1886/1990). *Beyond Good and Evil: Prelude to a Philosophy of the*, Hollingdale, R. J. (trans.). London: Penguin.
- Noddings, N. (1984). *Caring: A Feminine Approach to Ethics and Moral Education*. Berkeley, CA: University of California Press.
- Olsen, M. (2006). *Michel Foucault: Materialism and Education*. updated edn. Boulder, CO: Paradigm Publishers.
- Peters, M. and Wain, K. (2003). Postmodernism/post-structuralism. In Blake, N., Smeyers, P., Smith, R., and Standish, P. (eds.) *The Blackwell Guide to the Philosophy of Education*, pp 57–72. Oxford: Blackwell.
- Popkewitz, T. (1999). Critical traditions, modernisms, and the “posts” In Popkewitz, T. and Fendler, L. (eds.) *Critical Theories in Education: Changing Terrains of Knowledge and Politics*, pp 1–13. New York: Routledge.
- Rorty, R. (1979). *Philosophy as the Mirror of Nature*. Princeton, NJ: Princeton University Press.

- Sawicki, J. (1998). Feminism, Foucault and 'subjects' of power and freedom. In Moss, J. (ed.) *The Later Foucault*, pp 93–107. London: Sage.
- Siegel, H. (1997). 'Radical' pedagogy requires 'conservative' epistemology. *Rationality Redeemed? Further Dialogues on and Educational Ideal*, pp 141–154. New York: Routledge.
- Spivak, G. C. (1988). Can the subaltern speak? In Nelson, C. and Grossberg, L. (eds.) *Marxism and the Interpretation of Culture*, pp 271–313. Urbana, IL: University of Illinois Press.
- Styron, W. (1979). *Sophie's Choice*. London: Jonathan Cape.
- Weedon, C. (1997). *Feminist Practice and Poststructuralist Theory*, 2nd edn. London: Blackwell.
- Woodson, C. G. (1933). *The Mis-Education of the Negro*. Washington, DC: Associated Publishers.
- Zembylas, M. (2007a). A politics of passion in education: The Foucauldian legacy. *Educational Philosophy and Theory* **39**(2), 135–149.
- Zembylas, M. (2007b). Risks and pleasures: A Deleuzo–Guattarian pedagogy of desire in education. *British Educational Research Journal* **33**(3), 331–347.
- Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings*. New York: Pantheon Books.
- Greene, M. (2009). *In Search of a Critical Pedagogy*. In Darder, A., Baltodano, M., and Torres, R. (eds.) *The Critical Pedagogy Reader*, 2nd edn, pp 97–112. New York: Routledge.
- Hooks, B. (1994). *Teaching to Transgress: Education as the Practice of Freedom*. New York: Routledge.
- Ladson-Billings, G. (1995). Toward a critical race theory of education. *Teachers College Record* **97**, 47–68.
- Lather, P. (1991). *Getting Smart: Feminist Research and Pedagogy with/in the Postmodern*. New York: Routledge.
- Leonardo, Z. (ed.) (2005). *Critical Pedagogy and Race*. Oxford: Blackwell.
- Marshall, J. (ed.) (2004). *Poststructuralism, Philosophy, Pedagogy*. Dordrecht: Kluwer Academic Publishers.
- Peters, M. (ed.) (1998). *Naming the Multiple: Poststructuralism and Education*. Westport, CT: Bergin and Garvey.
- Peters, M. and Besley, T. (2006). *Why Foucault? New Directions in Educational Research*. New York: Peter Lang.
- Peters, M., Lankshear, C., and Olssen, M. (eds.) (2003). *Critical Theory and the Human Condition: Founders and Praxis*. New York: Peter Lang.
- Popkewitz, T. and Fendler, L. (eds.) (1999). *Critical Theories in Education: Changing Terrains of Knowledge and Politics*. New York: Routledge.
- Schrift, A. (1995). *Nietzsche's French Legacy: A Genealogy of Poststructuralism*. New York: Routledge.

Further Reading

- Ball, S. (ed.) (1990). *Foucault and Education: Disciplines and Knowledge*. New York: Routledge.
- Cherryholmes, C. (1988). *Power and Criticism: Poststructuralist Investigations in Education*. New York: Teachers College Press.

Reason and Rationality

P Hogan, National University of Ireland, Maynooth, Ireland

© 2010 Elsevier Ltd. All rights reserved.

Introduction

To deal with this extensive theme in an encyclopedia article calls for a degree of conciseness that requires important points to be treated in a summarized way. In the first part of the article, central philosophical issues are identified and briefly reviewed. The second part then sketches some educational implications and consequences of these earlier reviews.

Some Central Issues

In everyday conversation, reason and rationality are often used interchangeably. For instance, people frequently say: “where’s the rationality in that?” for “where’s the reason in that?”, or “his outlook lacks rationality” for “his outlook lacks reason.” Staying more strictly with dictionary definitions however, an important distinction can be drawn between reason and rationality. The Oxford Dictionary states that the noun reason refers to an intellectual faculty, or capacity, characteristic of human beings, that enables them to draw logical conclusions from premises. As for the noun rationality, this is not listed as a separate entry in the dictionary, but appears as a cognate noun of the adjective rational, where rational itself has a range of meanings, including, endowed with reason, reasoning, sane, not foolish or absurd, or rejecting what is unreasonable or cannot be tested by reason in religion or custom. On this stricter account, reason refers to the human capacity and rationality to its use; namely to good or less-than-good embodiments of that capacity in human thought and action. Not surprisingly then, while reason remains for the most part a term of approval, rationality attracts more critique and criticism, chiefly because it reveals itself in contrasting kinds of embodiments. These embodiments take place not just in the thoughts and actions of individuals but also in cultural traditions, in social and economic arrangements, in work practices, in systems of justice, and so on. This point is brought home to us by the tensions among the unprecedented diversity of rationalities of which we are conscious today, not only internationally but also within particular countries or societies.

Many find the plural form of the word (rationalities rather than rationality) a bit disconcerting, holding that rationality is simply that which is in accordance with reason and irrationality that which is not. To speak of

rationalities however, as for instance Alasdair MacIntyre does in detail in his book *Whose Justice? Which Rationality?* (MacIntyre, 1988) is not only to claim that the use of reason can produce contrasting or conflicting stances, but to suggest that it regularly does so, even necessarily does so. Such claims would be rejected by rationalists, who are invariably keen to claim that reason is singular, not plural, in nature, and that conflicts of rationality can be traced either to faulty premises or to faulty logic in arguing from premises, or to a combination of both faults. For rationalists, a proper understanding of reason would acknowledge that it is an *a priori* capacity, uncolored and unclouded by bias, and accordingly that anything worthy of the term rationality would be a model of consistency and even-handedness.

Debates about the merits of one form of rationality over another normally come down to two key issues, to the justifiability of the premises and to the adequacy of the associated reasoning – in other words, how rational (as distinct from biased or prejudiced) one’s starting premises are, and how consistent and thorough is the body of arguments issuing from these premises. The European Enlightenment of the seventeenth and eighteenth centuries is a good focus for attention here as it marks a high point for critiques of a range of historically inherited rationalities, or if you prefer, of contrasting kinds of rationality. Most notably these were rationalities that granted privileged standing to monarchy, to ecclesiastical authority, to various kinds of nobility, and more generally to a cosmic order, a great chain of being, where people and things had their allocated places, with accompanying expectations and practices.

Not only the rationalities inherit from the past, but also reason itself, became the subject of critique for the Enlightenment, particularly for Kant (1724–1804). His three major critiques (*Critique of Pure Reason* 1781, *Critique of Practical Reason* 1788, and *Critique of Judgement* 1790) sought to provide thorough investigations of reason itself and comprehensive foundations for its valid exercise in a range of domains. Arising from these investigations, Kant announced the Enlightenment’s major educational theme with the declaration “Enlightenment is man’s emergence from his self-incurred immaturity.” Such immaturity he described as “the inability to use one’s own understanding without the guidance of another” (Kant, 1784). Among the new directions for a rationality appropriate to education proposed by Enlightenment thinkers were ones like the following sample:

1. In education, as in other forms of action, the claims of long-established authority and tradition can no longer remain unchallenged, as many of these claims have stronger roots in prejudice than in rationally justifiable premises. All such claims must therefore be submitted to the judgment seat of autonomous human reason.
2. Education must enable learners to make full use of their own understanding. Learning must therefore be a form of emancipation. It must cultivate the rational autonomy of the individual rather than be an instrument for securing conformity in thought and belief.
3. As humans have a capacity for reason, they have a unique standing among other beings. While everything in nature conforms blindly to laws, rational beings can formulate laws, to which they themselves can then conform or not conform. Humans therefore have a status which is incomparably higher than anything in nature, and this higher status commands moral respect.

Challenges such as these, Enlightenment thinkers maintained, could be successfully tackled only to the degree that unwarranted assumptions or overlooked preconceptions could be brought to light and capably dealt with by pure reason. Deficient forms of reason, it was claimed, could be identified in much of previous philosophy, including the major metaphysical systems of Plato, Aristotle, and Aquinas. In particular, certain articles of faith, whether springing from pagan religion, from Greek beliefs about politics, or from Christian doctrines, were seen to be active in an influential way in metaphysical philosophy, but remained outside of the critical focus of metaphysical enquiry itself.

How far the Enlightenment and its intellectual legacies have succeeded in establishing a conception of reason that is freed from all assumption and presupposition however, remains controversial. Indications of this can be seen not just in contemporary discourse on ethics, politics, and religion, but in rationalities embodied in different domains of human work; for instance, in the rationalization of industry (e.g., from human assembly lines to robotic manufacture), of agriculture (from the land enclosures of history to today's technological agribusiness), and of education (from the monitorial system of Bell and Lancaster to current forms of e-learning). All of these have involved forms of rationality that draw sharp distinctions between means and ends – in effect between prized ends on the one hand and the use of humans (or substitutes for humans) as instruments on the other. Max Weber's famous distinction between instrumental rationality (*Zweckrationalität*) and practical rationality (*Wertrationalität*) comes to mind here. It should be stressed that the term practical rationality, as used by Weber and by philosophy more generally since Aristotle, refers to moral-practical reasoning, not to a kind of practical mentality that sidelines moral issues in its keenness to achieve certain practical goals. This latter would be an example of

instrumental rationality. The premises informing instrumental rationality (e.g., the maximizing of profit, the defeating of enemies, or the securing of power for one's party) can quickly be seen as laden with different kinds of taken-for-granted assumptions. Although issues of justification are not always irrelevant here, they are removed to the background. Reason is already spoken for, as it were, as it is already embodied in procedures and practices designed to accomplish, without questioning or interruption, certain previously decided purposes or goals. Practical rationality, by contrast, is concerned with questions of how one should act, and of what kind of reasonable belief could inform such action. In this case, the issue of presuppositions and their justification remains critically at the center.

Apart from illicit activities which do not have even a hint of honor among thieves, virtually all human occupations involve some combination of instrumental and practical rationality. In occupations like medicine, nursing, social work, and not least education, because of their nature as caring professions, demands for explicit justification are likely to arise more frequently than they would, for instance, in the exercise of management functions in public administration or in industry and commerce. Even where procedures are already decided on, and may thus look like instrumental rationality, these procedures regularly come up for review, or fresh ones have to be considered for new situations and circumstances. Where practitioners in such fields lean too heavily toward instrumental forms of rationality (e.g., where a profession has become heavily bureaucratized) it is likely that the values underlying practice itself will become semiautomatic in their operation. Assumptions which at one point had to be debated, justified, and decided upon by practitioners, can become ingrained in practice over the course of time. To the extent that they do, an instrumental rationality prevails in a taken-for-granted way and practitioners themselves can all too often become creatures of habit.

Resistance to the kinds of debates that involve practical reasoning can be quite common among members of different professions, sometimes because practitioners see them as endless, or believe some colleagues talk too much at meetings, or want to go into each issue twice over, and so on. Much of the apparent interminability of practical reasoning however arises from the failure – so far at least – to deliver on the Enlightenment's expectations of pure, or unbiased, reason. That is to say, humankind still seems to lack an effectual conception of reason from which hindrances like preconceptions, prior assumptions, not to say prejudices, have been removed. The most that seems to have been achieved along these lines are temporary or stipulative conceptions, where different parties agree for present purposes to certain definitions, or meanings.

In different ways, the attempts of major philosophers such as Descartes, Kant, and Husserl sought to provide

a methodical procedure whereby rational thinking could become free of deceptive predisposing influences. Success in such attempts would secure a foundational or irrefutable conception of reason, and furnish universally valid thought procedures for the most important of human enquiries. In the case of each of the three thinkers, however, the goals being pursued by their attempts became so intractable that reason itself had to be disconnected in one or more ways from human experience, or more precisely, from the fullness or integral-ness of human experience. Descartes had to create a methodically doubting human subject with a contrived style of thinking that radically disengaged the thinker from others. Kant created an even more elaborate system that included categories of understanding that were distinct from categories of reason, analytic *a priori* judgments that were distinct from synthetic *a priori* judgments, as well as a range of other highly abstract distinctions. Despite Kant's best intentions, these distinctions led inquiry elsewhere than toward an illumination of what befalls human experience itself, and what befalls reason itself, as reasoning is attempted. Husserl, keen to retain sight of this fuller experiential context, developed a discipline of phenomenological reduction that sought to bracket out the distorting influences of uncritical, or pre-reflective consciousness. Again, however, Husserl's efforts became involved in intractable difficulties that led to successive new starts on his own part, until thinkers such as Heidegger, Wittgenstein, and Dewey brought inquiries into reason and thinking on quite different paths.

Before reviewing the fruits of these latter enquiries, it is important to point out that efforts to secure a unitary conception of rationality that would keep preconceptions at bay, and that would have universal application, still continue, not least where practical rationality is concerned. Among the most influential of such efforts in recent times has been that of John Rawls in his book *A Theory of Justice*. This book is not an investigation of reason as such but an attempt to establish a universal rationality for justice. Rawls seeks to provide a standpoint that can overcome all partisan positions and furnish neutral or impartial grounds that manifest such a universal rationality. He presents a thought experiment involving the adoption of an original position in matters of justice. The thought experiment seeks to establish what principles of justice might come to the fore if one placed oneself behind a veil of ignorance, thus depriving oneself of any knowledge of the beliefs, gender, social standing, ethnic background, religion, or abilities of others. Adopting the veil of ignorance thus means ridding oneself of all assumptions and preconceived ideas that might otherwise cloud one's thinking and prevent one from taking an original position. The original position would therefore be a purely disinterested stance. Critics of Rawls point out that, whatever its merits as a thought experiment, his argument fails in its claim to universality. The charge is that Rawls's theory implicitly privileges

a stance of individual liberalism over others and that it ignores the social and historical factors that give substance to principles of justice – factors that remain inescapable features of human rationality itself.

But what if an original position, or disinterested stance, is actually unattainable by the rational efforts of humans? What if the attempts to seek such a position are mistaken? Such were the thoughts that began to exercise the young Heidegger in the 1920s as he struggled with the difficulties bequeathed by efforts such as those of Descartes, Kant, and his own teacher Husserl. Heidegger concluded that human understanding is not something separate from human reason, human judgment, or human interpretation, but includes all three, albeit deficiently, from the start. More startlingly, he argued that interpretation and prior judgments remain inescapably active in all understanding. He concluded, uncontroversially, that critical reflection can helpfully identify and discipline many of the predisposing influences that orient human understanding. More challengingly however, and against all rationalist epistemologies, he argued that such influences cannot be finally overcome and laid aside through the conquest of some supposed neutral starting point. On this account, human rationality must accept a more modest view of its own accomplishments and of the possibilities of human reason. What the pursuit of these possibilities achieves now becomes acknowledged as something that is at best partial, and in both senses of that word: incomplete and burdened by bias.

Arguments like these advanced by Heidegger have subsequently been developed into major themes in the philosophical hermeneutics of Gadamer and Ricoeur. They also find parallels and resonances in the works of philosophers as different as Dewey, Wittgenstein, and Popper, and more recently in writings such as those of Charles Taylor, Seyla Benhabib, and Richard Bernstein. These philosophers argue that the search for secure and certain foundations for reason and its exercise must be given up and replaced by something more self-critical – something with a better understanding of the implicated-ness of history in human reason and a more discerning assessment of human possibility and limitation. The accounts produced by thinkers like those mentioned in this paragraph still make universal claims however, albeit in a provisional, or fallibilist sense, as distinct from any *a priori* sense.

Such thinkers are broadly called nonfoundational philosophers. But there are also more radical forms of nonfoundational philosophy, associated chiefly with postmodern currents of thinking that arose in the last quarter of the twentieth century. These are associated particularly with the writings of Jean François Lyotard and Michel Foucault, and also in some respects with those of Jacques Derrida and Richard Rorty. The characteristic tendency of postmodern thinking is not only to dismiss metaphysical and epistemological accounts of reason and rationality, but also

to reject any philosophical claims to universality. With this goes an insistence that all reasoning must be merely local, as must all accounts of reason itself (small narratives rather than grand narratives). Such a stance is recommended by postmodernists in order to: (1) expose the often oppressive features in hegemonic forms of rationality that claim universality; and (2) destabilize any emergent successors to forms of such rationality that have historically enjoyed a place in the sun.

Critics of both kinds (hermeneutic and postmodernist) of nonfoundational thinking often argue that allowing interpretation such a central role in the exercise of human reason does away with the notion of objective truth and invites the charge of relativism. A related version of this criticism is that independent norms of practical reasoning (i.e., about how to act and what to believe) are now abandoned and that reason itself falls prisoner to the conditioning influences of historical and social circumstances. Responses to this criticism would differ from both forms of nonfoundational philosophy that we have just considered. Postmodernist responses would likely say good riddance, both to objective truth and to independent norms of practical reasoning. Both would be seen as involving pretensions that are no longer credible and that have in the past all too often led to invidious social and political consequences. In the case of the other forms of nonfoundational philosophy, the broadly hermeneutic ones, there would likely be a variety of responses, but with a concentration on arguments like the following: objective truth, rather than being dismissed, now features as something that can be suggestively imagined, but hardly attained, by human efforts – much as omniscience can be humanly imagined, but not humanly attained. Truth is seen on this account as something to which reasoning remains on the way, rather than something possessed with certainty, and for good. For the purposes of practical reasoning, truth is thus associated with what Dewey called warranted assertibility (Dewey, 1938/1986: 15), or with what Popper described as theories that have so far withstood refutation, but remain open to it. This is not to devalue the importance of truth however. If anything, it is to do the opposite, and in a threefold way: first, to suggest that pure reason is something more godly than human, but that still beckons the best efforts of human enquiry; second, to disclose the predisposed nature, including the possibilities and the limitations, of the human capacity for reason; and third, to remind us of the provisional character of the forms of rationality accomplished by the exercise of that capacity.

Educational Consequences and Effects

We have noted the radical challenge of the Enlightenment for education in Kant's declaration about humankind's emergence from its self-incurred immaturity, and in his aspiration that understanding might be used without the

guidance of another. In this, Kant was influenced by Rousseau's memorable and forceful arguments in his *Émile* (Rousseau, 1762/1993) for promoting independence of mind among learners (albeit male learners). Kant announced the motto of Enlightenment as "*sapere aude!*" (dare to know). Have courage to use your own understanding!" (Kant, 1784). Cultivating such understanding and the courage to use it thus become predominant educational aims in those forms of rationality championed by the Enlightenment. Enlightenment thinkers were keenly aware that this marked a sharp break with paternalistic and custodial forms of rationality that were hitherto influential in the history of Western education. These earlier forms would include not only those that had origins of one kind or another in Plato's metaphysics. Crucially, they would also include the long traditions of teaching in cathedral and monastic schools prior to the Reformation. These were based on a theological–philosophical rationality that owed its effective allegiance to the papacy. During the Reformation and its aftermath, this kind of rationality endured much turbulence, before becoming reconstituted along mutually opposed denominational–cum–political lines: Catholic–political, most notably in Spain; Lutheran–political, as in northern Europe; Calvinist–political in Geneva, Scotland, and elsewhere. From a critical Enlightenment perspective however, such breakup and reconstitution meant that the traditional paternalistic rationality that had long underlain European education became strengthened rather than weakened in becoming plural. Each denominational tradition developed a form of rationality that became doctrinally more exact, and more exacting, with new sectarian dimensions reinforcing the already authoritarian constraints on teaching and learning.

In the wake of the Enlightenment itself, however, the rational autonomy championed by Enlightenment thinkers as a goal of learning fared ill, at least for a few generations, in European schools and colleges. Efforts by Pestalozzi in Switzerland to encourage learners to be more active agents in their own learning met with some moderate success, as did those by Froebel in Germany to transform pedagogical thinking from a paternalistic rationality to a more pedagogically perceptive one. Neither Pestalozzi nor Froebel were advocates however of the Enlightenment's bolder kind of rational autonomy. Wilhelm von Humboldt's measures in the early nineteenth century to reorganize university education in Berlin and higher-secondary education in Prussia along Enlightenment lines successfully broke new ground, but these measures were exceptions in a Europe that grew increasingly reactionary after the fall of Napoleon in 1815. In fact, it was not until the twentieth century that notions like encouraging students to dare to know made substantial progress in Western countries. The first wave of this progress was associated with the work of John Dewey in the USA, though in an urbane rather than a revolutionary way. Dewey argued that educational practice should be

constituted as far as possible on democratic lines – that practice itself should exemplify a democratic rationality. He tried to embody such a rationality in his Laboratory School, attached to the University of Chicago. In the context of deeply conservative attitudes in educational policy circles in the United States however, much of Dewey's work was misunderstood or misconstrued. The educational philosophy he developed over his long lifetime found an influential following, including some uncritical devotees; but it was also frequently caricatured or dismissed, including by President Eisenhower in 1959.

Elsewhere, rational autonomy began to make some strides as an explicit research theme in education. For instance, in the 1970s in the United Kingdom, much research in the philosophy of education was concerned with exploring rational autonomy as an educational aim. Associated in particular with the writings of analytic philosophers Richard Peters, Paul Hirst, and Robert Dearden, this research focused characteristically on an analysis of the logical properties of the concept of rational autonomy, and on the conditions that needed to be satisfied for a person to be called rationally autonomous. It also gave much attention to elucidating forms of rational thinking into which learners were to be initiated through education. A key mark of an educated person would thus be his or her ability to make rational choices based on public criteria of judgment. Some critics of this analytic approach, which featured strongly in many teacher-education courses during the 1970s and early 1980s, argued that it paid insufficient attention to the importance of personal authenticity (in addition to public criteria) in any account of rational autonomy (Bonnett and Cuypers, 2003). Others charged that this approach, while retaining a declared allegiance to the disinterested stance and second-order procedures of analytic philosophy, often ignored its own rules: that its philosophical accounts – of curriculum, of development, of teaching, of rationality itself – were frequently value laden, and were more of a first-order than a second-order character (O'Connor, 1973; Carr, 2004).

Toward the end of the twentieth century, a new interest in Dewey's work became evident, in the United States and internationally. This was helped by the emergence of a more receptive climate for many of his ideas. This was not because there was now a widespread appreciation among policymakers of the kind of rationality that Dewey envisaged for healthy environments of learning. Rather, it was for more pragmatic reasons, such as the importance given to learner-centered education as a desirable strategy to promote newly ascendant international goals like the knowledge society and lifelong learning. It is difficult to say how this revival of interest in Dewey's educational thought will fare in the future, or in what directions it will lead. It seems clear however from the many recent reports on education by bodies like the

Organization for Economic Cooperation and Development (OECD), the United Nations Educational, Scientific and Cultural Organization (UNESCO), and the European Commission, that themes like learner-centered education, lifelong learning, and the knowledge society will remain central in the context of intensifying policy debates on globalization.

Developments like these provide timely opportunities for exploring the theme of educational rationality as a distinctive type of rationality. MacIntyre's book mentioned earlier examines at length four contrasting kinds of rationality – Aristotelian, Augustinian, Scottish Enlightenment, European Enlightenment – and reviews some others in less detail. Among the hallmarks of MacIntyre's approach are his illustrations that particular rationalities are historical human accomplishments; accomplishments that are embodied in, and developed through traditions of enquiry. Rationalities that are thus embodied can be strengthened, or alternatively have their weaknesses exposed, MacIntyre claims, by engaging in active debate with rival traditions. He gives educational institutions, particularly universities, a special importance here, but only if they speak as protagonists of one contending party (MacIntyre, 1988: 401). Liberal universities cannot do this successfully, MacIntyre claims, because they conduct their research and teaching "as if there were indeed shared standards of rationality, accepted by all teachers and accessible to all students" (p. 399). MacIntyre insists, in common with both forms of nonfoundational philosophy mentioned earlier, that it is an illusion to hold that there are such universally shared standards, but he adds that this is an illusion on which the liberal university itself is based. In denominational universities by contrast, MacIntyre continues, traditions of research and teaching are associated with the shared standards of a particular rationality – Catholic, Protestant, or other as the case may be. Their members are thus better placed than their counterparts in liberal universities to engage in the combative dialogue through which particular rationalities are further articulated and sustained against the claims of their rivals. Here, MacIntyre locates himself, and his philosophical efforts of recent decades, within an Aristotelian-Thomistic, or Catholic tradition.

MacIntyre holds that subscribing to a form of rationality embodied in a particular religious tradition is what makes it possible for teachers in a denominational college or university to share the kinds of standards he talks about. On this kind of account however, it would be possible for people who understand themselves first and foremost as teachers to do something that MacIntyre does not envisage – to share, for educational purposes, certain provisionally acknowledged standards for the conduct of teaching and learning. These would be standards that do not claim to be universal in any *a priori* sense, but that claim to be worthy of assent, to the best of our fallible knowledge to date. In fact, this is the more convincing

rationale, not only for the liberal university, but for any educational institution that seeks to withstand charges of indoctrination. Where educational practice, in liberal universities or elsewhere, fails to meet the standards of such a rationale, it is often because there is insufficient attention paid by teachers and students to their own preconceptions; preconceptions that are often all the more powerful in their influence because they are undeclared or even undetected by those who hold them. Far from faulting the universalistic pretensions of liberal universities, a more promising way to remedy this is to try to work out more fully what an educational rationality, properly so called, would look like. This is a large undertaking, the tenor of which has been initially touched on in some earlier remarks, but a few key features can be indicated here by way of conclusion.

An educational rationality that hopes to be adequate to its task seeks, in the first place, to be alert to its own predisposing influences, especially those preconceptions that habit and custom in institutions of learning might occasion it to overlook. Such a bringing to light and disciplining of presuppositions cultivates a self-critical disposition in would-be teachers and learners. This needs to be accompanied however by positive action to promote distinctive environments of teaching and learning – environments that give priority to the exercise of constructive criticisms and to enquiry itself as a community endeavor by learners. Such environments would also be marked, in a special way, by a hospitality to imaginative venturing.

See also: Autonomy; Hermeneutics; Philosophy of Education: Overview; The Analytical Tradition.

Bibliography

- Bonnett, M. and Cuypers, S. (2003). Autonomy and authenticity in education. In Blake, N., Smeyers, P., Smith, R., and Standish, R. (eds.) *The Blackwell Guide to the Philosophy of Education*, pp 326–340. Malden: Blackwell.
- Carr, W. (2004). Philosophy of education. In Dunne, J. and Hogan, P. (eds.) *Education and Practice: Upholding the Integrity of Teaching and Learning*, pp 127–143. Malden: Blackwell.
- Dewey, J. (1938/1986). *The Later Works of John Dewey*, pp 1925–1953. 1925–1953: 1938, *Logic: The Theory of Inquiry*, vol. 12 (Collected Works of Dewey, J.), Boydston, J. A. (ed.), with Introduction by Nagel, E. Carbondale, IL: Southern Illinois University Press.
- Kant, I. (1784). An answer to the question: What is enlightenment? <http://theliterarylink.com/kant.html> (accessed September 2009).

- MacIntyre, A. (1988). *Whose Justice? Which Rationality?* London: Duckworth.
- O'Connor, D. J. (1973). The nature and scope of educational theory. In Langford, G. and O'Connor, D. J. (eds.) *New Essays in the Philosophy of Education*, pp 47–65. London: Routledge and Kegan Paul.
- Rousseau, J. -J. (1762/1993). *Émile*, B. Foxley (trans.). London: Dent.

Further Reading

- Bernstein, R. (1991). *The New Constellation: The Ethical–Political Horizons of Modernity/Postmodernity*. Cambridge: Polity Press, in association with Oxford: Basil Blackwell.
- Benhabib, S. (1992). *Situating the Self: Gender, Community, and Postmodernism in Contemporary Ethics*. New York: Routledge.
- Dearden, R. F., Hirst, P. H., and Peters, R. S. (1972). *Education and the Development of Reason*. London: Routledge and Kegan Paul.
- Descartes, R. (1637/1999). *Discourse on Method and Related Writings*, Clarke, D. M. (trans.). London: Penguin.
- Gadamer, H. -G. (1960/1975). *Truth and Method*, Barden, G. and Cumming, J. (trans.). London: Sheed and Ward.
- Heidegger, M. (1926/1973). *Being and Time*, Maxquarrie, J. and Robinson, E. (trans.). Oxford: Basil Blackwell.
- Husserl, E. (1936/1970) *The Crisis of European Sciences and Transcendental Phenomenology*, Carr, D. (trans.). Evanston: Northwestern University Press.
- Kant, I. (1781/1963). *Critique of Pure Reason*, Kemp-Smith, N. (trans.). London: Macmillan.
- Kant, I. (1788/1997). *Critique of Practical Reason*, Gregor, M. (trans.). Cambridge: Cambridge University Press.
- Kant, K. (1790/1978). *Critique of Judgement*, Meredith, J.C (trans.). Oxford: Oxford University Press.
- Popper, K. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge and Kegan Paul.
- Rawls, J. (1971). *A Theory of Justice*. Oxford: Oxford University Press.
- Ricoeur, P. (1981). *Hermeneutics and the Human Sciences*, Thompson, J. B. (trans., ed., and introduced). Cambridge: Cambridge University Press.
- Taylor, C. (1995). *Philosophical Arguments*. Cambridge, MA: Harvard University Press.
- Weber, M. (1947). *The Theory of Social and Economic Organization*, Parsons, T. (ed., Introduction). New York: Free Press.
- Wittgenstein, L. (1953). *Philosophical Investigations*, Anscombe, G. E. M. (trans.). Oxford: Basil Blackwell.

Relevant Websites

- <http://www.ffst.hr> – Encyclopaedia of Philosophy of Education, *Filozofski fakultet u Splitu*.
- <http://www.iep.utm.edu> – Internet Encyclopedia of Philosophy.
- <http://plato.stanford.edu> – Stanford Encyclopedia of Philosophy, Table of Contents.
- <http://www.nuim.ie> – Teaching and Learning for the 21st Century (TL21), National University of Ireland, Maynooth.

Religious and Spiritual Education

J C Conroy and S J McKinney, University of Glasgow, Glasgow, UK

© 2010 Elsevier Ltd. All rights reserved.

Religious Education – Contested Concept and Competing Expectations

It is arguable that religious education is both more contested and more regularly denigrated than any other area of the school curriculum across a range of polities. This contestation arises as a consequence of much diversity in the conceptualization and the pedagogical realization of religion, and because the relationship between religious communities and public policy in late industrial political communities has become fraught as the church and state have become increasingly separated.

There are those who, supporting the existence of religious education as a necessary feature for a rounded education in the public school, argue that, religion is an important dimension of human life in both historical and contemporary contexts; religious ideas have been hugely influential in the creation of cultures and civilizations, in literatures, and in modes of thought; religion is an important component of personal and collective identity for many people; religion provides exemplars of ethical codes that enable a virtuous life. Moreover, a number of scholars (Conroy and Davis, 2008; Wright, 2007) have argued that the growth of religious illiteracy is a central educational problem for late industrial liberal democratic polities. On the other hand, those opposed to any concord between religious communities and the state believe that religious education merely extends and enfranchises epistemological and moral infantilism whereby individual moral responsibility is usurped by authoritarian regimes. Another related refraction of such arguments is rooted in the post-Enlightenment rejection of transcendentalism.

Nevertheless, it is hard to escape the claim that cultures, including secular cultures, are themselves bred out of a sense of the religious (Scuton, 2007). Hence, the influence of Christianity in the development of Western civilization is particularly marked in historical structures of governance and codes of law and ethics. It has supported technologies, such as mass print (to disseminate the word of God), which themselves produce certain kinds of intellectual shifts; the Renaissance gives rise to the Reformation which in turn is midwife to the Enlightenment and eventually produces late-nineteenth-century atheisms. Religion, then, even in its negative manifestation, is an inescapable and important dimension of human life, worthy of study at various levels of education, including school.

There are equally robust national and international philosophical debates about the aims, purpose, and scope

of religious education in schools located within denominational, nondenominational, and multid denominational school systems informed by historical religious traditions, modernist and postmodernist theoretical frameworks, and multireligious perspectives. The diversity of conception and practical realization of religious education is marked not only in debates among academics or teaching practitioners but, perhaps more significantly, also between legislators. What constitutes religious education in different polities differs quite markedly. In the United States, for example, there being little or no religious education in public schools, it is normally seen as a synonym for religious nurture or catechesis in religiously denominational schools. In other Anglophone cultures, such as Britain and Ireland, South Africa, and Australia, it may take quite different shapes and have quite different purposes. Across Britain, despite national differences, there has been some broad general agreement about the nature of religious education as primarily a place to encourage the study of a range of world faiths. In Australia, there are certificated programs in religious education, although these have not been widely supported in state schools; some would argue that this is because there exists continued right of access by representatives of religious communities who can teach, on a voluntary basis, those whose parents actively choose to have their children introduced to their own faith. Yet again, across Europe, there are a range of provisions which elide any clear distinction between the public and the private, between the educational and the evangelical or catechetical (e.g., UK). The complexity of the debate is reflected in the nomenclature within different school systems which often provide a key to understanding these localized conceptions of the subject and the compromises reached. While the term religious education is popular throughout the Western world, there are many variations (e.g., religious instruction in Estonia; knowledge about Christianity in Denmark; and religion in Romania) and, in some cases, a broadening out of the intended scope (religions and ethics in Slovenia; Christian knowledge and religion and ethical education in Norway; and religious moral and philosophical studies in Scotland). In some countries there are options that provide a nonreligious alternative (confessional religious education or civic education in Serbia and Montenegro; and religious education or ethics in Slovakia). Often religious education, in whatever form, is compulsory for school children up to a certain stage, although there may be a conscience clause and the possibility to opt out (Kuyk *et al.*, 2007).

As the subject is often compulsory, contemporary debates, especially within the nondenominational and multi-denominational school systems, can be the sites of inter- and intra-religious and cultural maneuvering as different religious groups or denominations position themselves to claim a stake for their interest group in the construction of rationales and syllabi for national and local religious education in the classroom (e.g., Germany and Romania). This claim can also be for a stake in state education, and hence some form of state recognition and public legitimization of the religious group (McKinney, 2006). Indeed, it is difficult to see how any other subject on the curriculum would be susceptible to such pressures.

Religious education is often perceived as a subject that can make a contribution to whole-school initiatives, although Broadbent and Brown (2002) warns that, as a consequence, religious education could lose its focus as a discrete subject. In recent years, strong links have been established between religious education and forms of personal, social, and health development, moral and values education, citizenship education, and spiritual education. In such instances, religious education functions as a surrogate for other forms of education. This is of course a move that parallels similar philo-theological shifts in approaches to religious language, evident from the middle of the nineteenth century on and finding a strong voice in R. B. Braithwaite's celebrated essay (Braithwaite, 1971), which suggested that religious language serves as a form of reified moral language. Arguably, such links help to explain competing expectations of religious education from different stakeholders in society and in school education – expectations that are not simply related to examination results and uptake of the subject by school pupils but related more to a variety of constructions of the value of religion, religiously influenced attitudes, values, and lifestyles (Schreiner in Kuyk *et al.*, 2007). Civic society can harbor expectations that the subject should engender mutual tolerance and respect between different religious and cultural groups. Parents can expect the subject to provide a background and a literacy in religion that many of them are no longer able or willing to provide. Religious communities can often expect religious education in schools to aid them in the formation and development of religious identity and can attempt to impose exclusively orthodox representations of their religion to ensure adherence and continuity in their tradition (McKinney, 2006). Politicians often expect religious education to promote ethical conduct and support the creation of good citizens. This range of expectations can place enormous pressure on religious education as the different stakeholders judge the merits of the subject and its position in the school curriculum according to the results that are linked to their expectations. The unreasonably diverse and unattainable nature of these expectations can lead to a discourse, or discourses, of failure that can destabilize public confidence in the validity and efficacy of

religious education in schools. Perhaps, more importantly, so busy are those who try to co-opt religious education to their particular mission that they ignore what it is that constitutes the religious.

Models of Religious Education

A useful starting point for the mapping of religious education is Hull's proposal that religious education can promote education into religion, about religion, or education from religion, while acknowledging the simplistic nature of these distinctions and the creation of artificial boundaries between potentially complementary and overlapping activities (Schreiner in Kuyk *et al.*, 2007). Religious education into religion is into one specific religion (often a Christian denomination). Religious education about religion involves learning about beliefs, values, and practices of a religion. Religious education from religion helps children develop their own views by considering different answers to major religious and moral issues. The next sections explore three models of religious education that exemplify religious education into, about, and from religion: the confessional approach; the phenomenological approach; and the interpretive approach. These sections should be read with the caveat that these three models are not necessarily mutually exclusive nor do they exhaust the range and complexity of the models in religious education.

Confessional Approach

Reinforcing and simultaneously emanating from the controversies alluded to at the beginning of this article are the models of religious education currently adopted in faith schools (private or state funded), often referred to as catechetical, confessional, or faith formational models of religious education (e.g., Belgium and Scotland). Of course, it may be true that such approaches are generally limited to faith-based institutions in the twenty-first century, but this has not always been the case. Until the 1970s, in Britain, for example, most religious education was rooted in bible knowledge with the express aim of nurturing children in the Christian tradition – a formal position that has its roots in the 1870 Education Act. While it is the case that faith schools have traditionally belonged to a Christian denomination (a classic example of the catechetical or faith formational approach would be the worldwide network of Roman Catholic schools), Western liberal democracies have witnessed a growing diversity in religious schools, including the emergence of an increasing number of Islamic schools (e.g., Netherlands, Denmark, and England) (Kuyk *et al.*, 2007). Religious education in faith schools, often perceived within faith schools as a vehicle for religious nurture, is frequently suspected by outsiders of being overly transmissive and coercive, despite much diversity in the models of

faith schooling (ranging from liberal to conservative). This discussion of religious education in faith schools is further complicated by the difficulty of extricating religious education within the faith school from the related issues of collective or corporate acts of worship and the religious character of the school (Parker-Jenkins *et al.*, 2005).

A number of key questions emerge in the academic debate concerning the position and role of religious education in faith schools where religious education slips into religious practice and nurture. Some academics, such as Feinberg (2006), a proponent of liberal education, have challenged the models of religious education within faith schools. Feinberg questions how some of the aims of liberal education can be achieved by religious education in a faith school. While he acknowledges the diversity of religious education within, for example, American Roman Catholic schools, he wonders how religious education in a faith school can address moral novelty (new problems raised in science and technology) and allow children to be authentic and make authentic choices for their own good? He doubts if there is sufficient scope for critical reflection and is concerned that faith schools are not open enough to other religious perspectives.

A number of points can be raised here. First, this kind of position is located within the chain of argument that has moved from the claims for rational autonomy of the child being brought up in a religious family to be protected to the more recent claims that the rights of the child are compromised in faith schools (Manson in Gardner *et al.*, 2005). Second, such a line of argument often fails to fully recognize the wide variety of faith schools and the variegation within specific faiths that range from theologically conservative to the liberal (McKinney, 2006). The model of religious education as conceived and implemented within some faith schools may readily be perceived as compatible with the aims of liberal education. In other cases, a more restrictive, theologically dependent approach is adopted. Yet, even here, there is some evidence that some of these more conservative institutions have been influenced by religious, cultural, and ethnic plurality and their approach to religious education is becoming ever more open to other faiths and worldviews, embracing some degree of critical reflection (Schreiner, 2007).

The Phenomenological Approach to Religious Education

The move in public schools from an implicitly Christian-based, biblically rooted religious education to a more descriptive reified pedagogy has its roots in philosophical theology, which, in its own turn, goes back to nineteenth-century hermeneutics. The central questions of the hermeneutical endeavor: How am I to stand in relation to this text? How does this text appear to me? gave rise to the phenomenological excursions of such early-twentieth-century

thinkers such as Edmund Husserl. Such phenomenology represents one attempt in a line going back to Kant, to disclose *das ding an sich* (the thing-in-itself); in other words, to get to the essence of things. Accompanying this journey to unearth the inscape of things, we have seen the emergence of late-nineteenth-century anthropology with its drive to map the topography of others beliefs and practices. Against the backdrop of these developments, the study of religion or religious studies emerges as a discipline. *The Religious Experience of Mankind* (Smart, 1969), with its focus on a morphological approach (mapping perceived common features of religion), becomes a standard text as departments of theology or divinity (traditionally Christian) begin an evolution into departments of religious studies. At the same time, in the postwar period in Britain, immigration came to play an increasingly important role in shaping the curriculum. How were we to offer religious education to children whose own religious beliefs and practices are quite different from those of the majority indigenous population? The shift from theology to the study of religion(s) was pioneered in Britain by Ninian Smart, who also chaired an influential Schools Council Working Party, which, in its turn, recommended the adoption of a phenomenological approach to religious education. However, the practical application of phenomenology as a ground for a radically different justification and pedagogy for the conduct of religious education was rooted in morphological description rather than the search for essences. The practical consequence of this shift was that religious education came to be seen as a means for comparing and contrasting a variety of religious belief systems. Frequently, in the classroom, this meant a comparison between ritual practices with little-enough attention paid to the more-difficult epistemological and ontological questions which suffuse doctrine and belief.

The Interpretive Approach to Religious Education

Over the subsequent three decades, increasing disquiet emerged about the efficacy of a strictly phenomenological approach (although it is not entirely clear that the diluted practices in schools count as phenomenology). This gave rise to a number of modifications or alternatives. One important shift has seen the emergence of the interpretive approach, propounded by Robert Jackson and his colleagues at the University of Warwick, which is grounded in the conceptualizations and methodology of ethnographic research (Jackson, 1997, 2004). Rooted in, but distinct from, the phenomenological tradition Jackson attempts to chart a *via media* between a pedagogy hypothecated on the presentation of prescribed and normative beliefs and practices on the one hand, and the deconstruction of religion on the other, arguing for the preeminence of the personal narrative. Jackson argues that the interpretive

approach enables a more nuanced understanding of religion that recognizes a variety of, sometimes contested, expressions of religious belief and practice that are often interrelated or inextricably bound with culture and identity. An important aspect of the approach is the explicit recognition of the religious and ideological backgrounds of the pupils, and the children are encouraged to draw on their own experiences and those of their peers. This should enable them to derive their own personal positions within key debates and questions concerning religion and religious plurality. There are clear advantages to this approach that are focused on a model of religious education that draws from the complexity of religious situatedness. This, he claims, enables religious education to break out of the stranglehold of stereotypical orthodox discourses and the dominance of the Christian discourse. It allows the children some insight into the diffuse and variegated nature of religious expression within particular religions (especially the Eastern religions, such as Hinduism). It also promotes a greater understanding of religion as evolving and engaging with new sociocultural contexts. There may be a perception that this approach relies too heavily on experience, and the balance of the *via media* would have to be carefully monitored as this approach could easily become restricted to personal narrative. Some religious groups may be uncomfortable with this approach as it places less value on orthodoxy and orthodox representation but is more open to a variety of expression and representation. Some critics (Wright, 2007) argue that this approach fails to sufficiently recognize the importance of religious literacy, although Jackson would argue that his approach complements others.

Key Issues for Religious Education in the Twenty-First Century

Perhaps a key issue for the delivery of religious education is the manner in which professionals have attempted to construct conceptual maps of religion for children in school curricula. Whether these are grounded in organizing principles, emanating from pedagogy or religion, they tend to be rooted in a Christian worldview. In this, history is hard to escape – many religious education programs in Western Europe have their roots in a Christian imperative as a consequence of the historic strength of denominational schooling. Despite conscientious and well-intentioned attempts to incorporate a world-religions perspective in curricular developments, these maps can often be refracted, to varying degrees, through an implicit or explicit Christian lens. Sometimes this is demonstrated in the paradoxical, and ultimately meaningless, use of language and concepts borrowed from Christianity to construct frameworks that can be applied to all mainstream religions. As knowledge and understanding of Eastern religion in Western society

has deepened, and the influx of immigrants from diverse religious traditions (and denominations within those traditions) has increased, the use of these frameworks has become increasingly problematic and contested. Perhaps a series of more complex or sophisticated heuristic frameworks is required? Does this pose an unwelcome dilemma for those who wish to retain the dominance of Christianity in religious education?

This problem of perspective has been exacerbated by a recent challenge yet to be properly recognized by many religious educators, which has emerged out of the growth of a range of groups and spiritual movements that can lay claim to the description religious. The paradigm of six mainstream world religions, identified by academics such as Cole (1998), has been modified by the inclusion of other faiths in interfaith dialogs and networks (e.g., Baha'i; and Zoroastrianism in the Inter Faith Network for the UK); however, even this wider frame is facing serious challenge from the emergence of new religious movements and new-age spiritualities (Cowan and Bromley, 2008). Such terms are crude generalizations but nevertheless point to a wide variety of expressions of religion from liberal or Western developments of mainstream faiths (e.g., engaged Buddhism and neo-Hinduism) to the groups previously (and pejoratively) identified as cults (e.g., Church of Scientology and Unification Church), and those viewed with historical suspicion (e.g., Wicca and witchcraft) (Clarke, 2006). The understanding of what constitutes a religion is being challenged as some of these groups and movements lobby for charitable status and official participation in interfaith networks and dialogs (Clarke, 2006). A key question for the twenty-first century will be how models of Western religious education, primarily focused on the six mainstream religions (often through a Christian lens), can negotiate this shift in the religious landscape and recognize or even accommodate these increasingly popular groups and movements without some form of possible re-entrenchment or recalibration of the nature of religious education.

Another major issue for the continuation of religious education in state school education is not the much-publicized and premature prediction of the terminal decline of religion in Western society, but the tension caused by the political and ideological pressure to create a permanent disjuncture between the state and organized religion at all levels of society. While liberal democracies defend the rights of individuals to hold and practice religious beliefs, these beliefs and practices should not impinge on or influence the social polity. Religious education in state schools can become a visible focus of the invidious intrusion of religion into public life and the unwelcome challenges of religious worldviews for the principles of liberalism. This is compounded by the emergence and growth of religiously motivated terrorism in the early twenty-first century, which has created an unfounded and unjust perception in some quarters, that religious

adherence is somehow synonymous with, or will inevitably lead to, fundamentalism. This, in turn, can lead to claims for a greater prominence for intercultural education, citizenship education, and values education in school education – forms of education that are explicitly grounded in educational ideologies that promote social tolerance and political participation but can be perceived to compete with, rather than complement, religious education (Jackson, 2004; Keast in Gardner *et al.*, 2005).

Conversely, some academics in the liberal tradition, such as Feinberg (2006), argue that it may be advantageous to a secular liberal democracy to include religious education in the curriculum in state schools. These academics suggest that a state influenced religious education that is balanced and adopts a critical stance toward religion may be able to counter possible excesses and fundamentalist tendencies that can emerge in informal/formal faith formation in religious communities. Similarly, Brighouse (in Gardner *et al.*, 2005) warns of the dangers of unregulated religious education in some of the private faith schools in America that can be closed to plurality in religion and critical engagement with both politics and religion.

Concluding Comments

Self-evidently, religion has had a formative influence on Western liberal democratic thought. Moreover, it remains equally true that, despite being considered misguided by numbers of their fellow citizens, many continue to be attached to religion as not just a life enhancement but an epistemological and ontological reality. It is for them a fundamental option that directs and influences their life options. As such, religion remains an important aspect of human life and, from one perspective, if children are to learn what it is to be human, it remains worthy of study. From another perspective, religion is an irrational endeavor that enslaves people with false soteriological and eschatological hopes that inhibit authentic rational autonomy and engagement in the furtherance of a human rather than theistic utopia. Religious education simply serves to authenticate and perpetuate this irrationality and enslavement. It is within these kinds of contesting discourses, located at the center of the search for meaning in human existence, that

religious educators continue to argue for the importance of their subject and its place in the school curriculum.

Bibliography

- Braithwaite, R. B. (1971). An empiricist's view of the nature of religious belief. In Mitchell, B. (ed.) *The Philosophy of Religion (Oxford Readings in Philosophy)*, pp 72–91. Oxford: Oxford University Press.
- Broadbent, L. and Brown, A. (eds.) (2002). *Issues in Religious Education*. London: Routledge.
- Clarke, P. (2006). *New Religions in Global Perspective*. London: Routledge.
- Cole, W. O. (ed.) (1998). *Six World Faiths*. London: Cassell.
- Conroy, J. and Davis, R. (2008). Citizenship, education and the claims of religious literacy. In Peters, M., Blee, H., and Britten, A. (eds.) *Global Citizenship Education*, pp 187–200. Rotterdam: Sense.
- Cowan, D. E. and Bromley, D. G. (2008). *Cults and New Religions*. Malden: Blackwell.
- Feinberg, W. (2006). *For Goodness Sake*. New York: Routledge.
- Gardner, R., Cairns, J., and Lawton, D. (eds.) (2005). *Faith Schools, Consensus or Conflict?* London: Routledge Falmer.
- Jackson, R. (1997). *Religious Education: An Interpretive Approach*. London: Hodder and Stoughton.
- Jackson, R. (2004). *Rethinking Religious Education and Plurality*. London: Routledge Falmer.
- Kuyk, E., Jensen, R., Lankshear, D., Loh Manna, E., and Schreiner, P. (eds.) (2007). *Religious Education in Europe*. Oslo: Iko Publishing House.
- McKinney, S. J. (2006). Faith based schools: Challenging assumptions and stereotypes. *Journal of Moral Education* 35(1), 105–115.
- Parker-Jenkins, M., Hartas, D., and Irving, B. A. (2005). *In Good Faith, Schools, Religion and Public Funding*. Aldershot: Ashgate Publishing.
- Scuton, R. (2007). *Culture Counts: Faith and Feeling in a World Besieged*. New York: Encounter Books.
- Smart, N. (1969). *The Religious Experience of Mankind*. New York: Scribner.
- Wright, A. (2007). *Critical Religious Education, Multiculturalism and the Pursuit of Truth*. Cardiff: University of Wales Press.

Further Reading

- Gates, B. E. (2007). *Transforming Religious Education: Beliefs and Values Under Scrutiny*. London: Continuum.
- Grace, G. and O'Keefe, J. (eds.) (2007). *International Handbook of Catholic Education*. Dordrecht: Springer.
- Husserl, E. (1964). *The Paris Lectures*. The Hague: Martinus Nijhoff.
- Jackson, R. (2003). *International Perspectives – Citizenship, Education and Religious Diversity*. London: Routledge.
- Schreiner, P., Kraft, F., and Wright, A. (eds.) (2007). *Good Practice in Religious Education in Europe*. Munster: LIT Verlag.
- Sandmark, S. (2000). *Is World View Neutral Education Possible and Desirable?* Carlisle: Paternoster Press.
- Watson, B. and Thompson, P. (2006). *The Effective Teaching of Religious Education*. Harlow: Longman.
- Ziebertz, H. G. (2003). *Religious Education in Plural Western Society: Problems and Challenges*. Munster: LIT Verlag.

The Capabilities Approach

L Terzi, Roehampton University, London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Agency – One's reflections and actions in shaping one's own life goals.

Capability/capabilities – The real, effective freedoms to pursue one's goal; genuine opportunities for functionings.

Development – The expansion of freedoms.

Equality – The fundamental principle of social justice; equal distribution of whatever goods are deemed valuable.

Functionings – The valued states and actions that are meaningful to people.

Normative framework – A framework specifying ethical principles, that is, how things ought to be.

Practical reason – The capacity to form and revise one's conception of the good.

Originally formulated by philosopher and welfare economist Amartya Sen, and further developed by philosopher Martha Nussbaum, the capabilities approach has increasingly influenced academic research as well as policymaking in the last two decades, (Robeyns, 2006a: 351). The capabilities approach is a normative framework for the evaluation of human well-being, the justice of institutional arrangements, and the level of individual advantage or disadvantage in contemporary societies. The fundamental innovative insight of the approach is that the evaluation of people's quality of life should be based on their capabilities to function, that is, on their real freedoms (their capabilities) to be and to do what they value being and doing (their functionings). According to Sen and Nussbaum, what a person is able to be (e.g., being well or poorly educated) and to do (for instance, doing a rewarding job), determines a person's quality of life and his/her well-being. In their view, the freedom to choose among possible functionings is fundamental to the possibility of leading good lives, and this insight should inform processes of social evaluation, as well as the analysis of equality. In focusing attention on capabilities and functionings, the approach provides a compelling alternative to both welfare economic models based on utilitarian considerations and philosophical approaches, to equality, which concentrate primarily on an individual's preference, satisfaction or happiness, or on income and wealth (Sen, 1982, 1992; Nussbaum, 2000, 2006b).

The philosophical roots of the approach can be traced back to Aristotelian and Marxist insights, as well as, to the works of Adam Smith and John Stuart Mill among other (Sen, 1980, 1992, 1999; Nussbaum, 2000, 2006a; Robeyns, 2005). Certainly, the theoretical reach of the capabilities approach and its inherent interdisciplinary nature (Robeyns, 2005, 2006a) are attested by the considerable interest it has attracted from different academic disciplines and from policymaking. Political philosophers, development economists, scholars in gender studies, as well as in sociomedicine and disability studies, have all engaged the ideas of the approach. Furthermore, the approach has been operationalized in the United Nations' Human Development Reports (1990–2004) and has more recently informed the final report of The Equalities Review (2007), commissioned by the UK Prime Minister to assess the state of equality in the country.

It is only relatively recently that educationalists have started to analyze the potentially fruitful insights of the capabilities approach for educational theory, policy, and practice. Perspectives applying the capabilities approach to education emphasize primarily its contribution to the conceptualization of social justice and educational equality, and to the theoretical and empirical analysis of educational provision, both in developed and developing countries. Attention is also devoted to pedagogical and curricular elements, and to their importance for the expansion of people's capabilities and their overall freedom to lead flourishing lives.

The article offers an overview of the capabilities approach and its relevance for education. A critical discussion of the main normative insights of the approach, as differently developed by Sen and Nussbaum, is followed by an analysis of the role of education and of the conceptions of educational justice and equality informed by these insights. The analysis is based on the main theoretical and empirical research developed in education, and outlines some of the themes and the future possibilities of the capabilities approach in education.

Normative Insights

Sen's Capability Approach

In critically engaging perspectives in welfare economics, mainly based on commodities and the philosophical debate on what kind of equality institutions should seek to achieve, Sen's work aims to provide a more adequate framework for conceptualizing human development,

as well as a compelling form of egalitarianism (Sen, 1992, 1999). Key concepts in Sen's approach are the space of capability, the specific metric informing interpersonal comparisons of advantage, the concept of human diversity, and finally, the democratic decision process entailed by the approach. Each of these concepts requires some elaboration, and we begin with the space of capability.

According to Sen, evaluations of poverty, inequality, and the just design of social institutions should take place in the space of capability. This is the space of the effective freedoms that people have in which to achieve objectives that are meaningful to their lives. Within this space, capabilities are defined as the real and genuine opportunities available to individuals to pursue their valued ends, and thus choose one type of life or another (Sen, 1992: 40). Functionings are valued states and actions (Sen, 1992: 39); being healthy and being educated, reading, going to work, and participating in the life of the community, are all examples of functionings. Capabilities and functionings are essential to well-being, Sen maintains, and as such, any evaluation of individual well-being should take the form of an assessment of these elements (Sen, 1992: 39).

Sen endorses equality of capabilities – or, conversely, the elimination of unambiguous inequalities in capabilities (Sen, 1992: 7) – as the appropriate objective of social policy, and asserts the importance of addressing interpersonal comparisons in the space of capability. The emphasis is therefore placed on the equalization of people's genuine opportunities, rather than on their achieved functionings. In line with broad liberal views, Sen supports a perspective that equalizes the opportunities available to people, thus allowing for the pursuit of their objectives in accordance with their different conceptions of the good life. This is further confirmed by the importance of the role of reasoned agency in forming valued life plans, since agency relates to choosing one's own broad life goals and commitments, and to actively engaging in their achievement.

The assessment of poverty and disadvantage, according to Sen, should be based on a specific metric for interpersonal comparisons, which sees the concept of human diversity as central, thus reinstating considerations of individual differences as fundamental to this process. Human diversity is conceptualized in relation to internal features, such as gender, age, and physical and mental abilities, as well as external factors, including the design of the physical, social and cultural environment that people inhabit. Furthermore, and importantly, Sen highlights people's differences in converting resources and commodities into valued objectives (Sen, 1992: 85). He illustrates this element of diversity with the well-known example of children and lactating women who, for their functionings in their daily lives, require a higher intake of proteins than other people. Hence, in his view, this different conversion of resources into functionings should inform any comparison of individual advantage.

On the basis of these insights, Sen identifies the capability approach as a framework of thought and not as a theory of justice, while declining, in light of the variability of human ends, to specify a definitive list of capabilities or functionings. Sen leaves these processes to public choice, reasoning, and democratic procedures, that are themselves the most freedom-preserving means by which social policy can be determined (Terzi, 2005). Hence, his approach has a deliberately underspecified character.

This underspecified nature of the approach, however, has been the subject of critiques. These primarily highlight the difficulties in choosing what capabilities are deemed important, and how they can be evaluated. Consider, for instance, the capabilities of a visually impaired person and those of a person suffering from asthma: How can these two sets of capabilities be compared in order to determine disadvantage? Furthermore, difficulties arise in evaluating these sets of capabilities in relation to those of many other individuals, who may have, and value, certain capabilities and not others. The unspecified character of the approach, some critics maintain, leads to the difficulty of assessing and indexing the value of capabilities. Sen replies to such critiques by highlighting, on the one hand, his reference to some basic, fundamental capabilities – such as being well nourished, sheltered, healthy, and educated – which are truly essential to human well-being, and which he repeatedly identifies as such in his work (Sen, 1992, 1999). At the same time, he emphasizes the role of democratic processes in selecting and listing relevant capabilities in relation to specific and determined contexts. This latter defense, however, leaves open the problem of securing prior democratic principles and constitutional procedures that would allow the advocated process of participatory selection to take place.

Martha Nussbaum's version of the capabilities approach, as discussed in the next section, addresses some of these concerns.

Martha Nussbaum's Capabilities Approach

Nussbaum endorses Sen's idea of capabilities as the correct variable for comparisons of quality of life, but articulates and specifies the philosophical dimension of the approach in some important ways. In particular, she identifies a list of central human capabilities that have a universal dimension – in that they can be shared by people holding otherwise very different conceptions of the good – and sets a threshold of adequacy in the possession of these capabilities, which should be guaranteed to all. These elements form the basis for constitutional principles to be adopted by all governments (Nussbaum, 2000: 12).

The central human capabilities proposed by Nussbaum include life, bodily health, bodily integrity, senses, imagination, and thought, emotions, play, control over the environment and other species, and practical reason and

affiliation (Nussbaum, 2000: 78–80). While emphasizing that her list is open to changes, Nussbaum considers these capabilities as fundamental to human well-being. Furthermore, she assigns a central role to practical reason, in its Aristotelian meaning of being able to form one's conception of the good, and to affiliation, or the ability to establish meaningful relationships, with each seen as both foundational to other capabilities and essential for leading lives that are truly human (Nussbaum, 2000: 82).

A further, important specification in Nussbaum's version of the approach relates to her articulation of the concept of capabilities in basic, internal, and combined capabilities. This addresses the critique of underspecification of the approach. Thus, basic capabilities, in her account, relate to basic innate capacities of individuals, whereas internal capabilities are developed states sufficient for the exercise of functionings. These internal states of the person, unlike innate capabilities, are developed in relation to external conditions and circumstances. For instance, according to Nussbaum, most people have the internal capability for freedom of speech, but whether they can exercise it and achieve the related functionings depends on external conditions (Nussbaum, 2000: 84). Combined capabilities amount to the internal states and the external conditions for the exercise of the relevant freedoms.

Nussbaum's focus on central human capabilities is grounded on the intuitive ideal of the equal moral worth and dignity of each and every individual, seen as an end in themselves (Nussbaum, 2000: 5), and on the possibility of a political overlapping consensus on the importance of each capability on the list. She maintains that when we ask the question central to the capabilities approach, "What is this person able to be and to do?", we are evaluating what this person's liberties and opportunities are, and how the resources at her disposal allow her to function in a truly human way (Nussbaum, 2000: 71, 74). Thus, the central capabilities act as guidance for political action as well as for policy design, since governments should provide both the conditions for the development of internal capabilities and the external conditions for their exercise, in ways and to levels that reflect the dignity of the human life.

In specifying a list of universal capabilities and in setting a threshold level that should be guaranteed to all, Nussbaum provides an answer to the charge of vagueness and underspecification leveled at the approach. However, her version of the capabilities approach is open to other critiques. In particular, these have been expressed in relation to the universality of her list of capabilities, and to the method used in formulating the list, which is seen as departing from the democratic consultation processes that could give it a more general endorsement. In reply, Nussbaum maintains that the list is indeed open to adaptations, and that people who otherwise belong to very different cultural and social environments share the capabilities on the list, thus reinstating the validity of her perspective (Nussbaum, 2000: 115).

While these and related matters remain far from settled, Nussbaum's approach contributes fundamentally to the theory of justice, in particular with her formulation of precise understandings of the constitutive elements of well-being and the good life. Further, in her more recent work, she extends these considerations to previously unexplored issues of justice, such as justice beyond national borders and justice for cognitively disabled citizens (Nussbaum, 2006a).

Both Sen's and Nussbaum's versions of the capabilities approach have influenced educationalists who have recently started to apply these ideas to the theory, policy, and practice of education. The next sections present the main insights which the capabilities approach offers to the field of education.

The Capabilities Approach and Education

Both Sen (1999, 2002) and Nussbaum (2000, 2004, 2006a, b) highlight the importance of education in the capabilities approach. They emphasize the contribution that being educated makes to the formation and expansion of capabilities and hence the contribution it makes to people's lives. Sen includes the capability to be educated among the small group of basic capabilities that are fundamental to well-being and to development and points out, for instance, the broadening of women's freedoms resulting from education and its related benefits, such as a reduction in infant mortality (Sen, 1999: 198). In her work, Nussbaum reflects extensively on the value of education, which, she maintains, should be seen "not merely as a provider of useful technical skills, but also, and more centrally, as a general empowerment of the person through information, critical thinking, imagination" (Nussbaum, 2006a: 323). She stresses the importance of education for democracy, for equality, and political action, as well as for the flourishing of human life (Nussbaum, 2006a: 322). This view of education as fundamental to the development of capabilities is further explored in her defense of the cultivation of humanity, regarded as central to a project of liberal higher education. The promotion of forms of critical self-examination, the idea of world citizenship, and the fostering of processes of narrative imagination, all contribute to individual development as well as to a political aim of global citizenship (Nussbaum, 1997).

Inscribed in these positions are some of the fundamental insights of the capabilities approach to education, which, for the purpose of this analysis, can be subsumed in three main areas. First, the approach emphasizes, together with its instrumentality, the intrinsic value of education and its functions, both for the individual and for society. Second, it contributes a valuable framework for the assessment of inequalities in education, while suggesting a conception of equality in terms of equal, effective, opportunities for the achievement of important functionings which are promoted

by education. Finally, it informs pedagogy and models of curriculum that enhance freedom and agency. Each of these insights has been recently explored in the growing literature on the capabilities approach in education, and each is analyzed in more detail in what follows.

The Value and Functions of Education

In contrast to current approaches mainly concerned with the instrumental role of education, such as human capital theory – with its accent on the acquisition of skills and competencies for the promotion of economic goals – the capabilities approach provides a broader view of education as having both an important intrinsic value and a range of different instrumental values for the individual, and for society more generally (Dréze and Sen, 2002; Robeyns, 2006b; Saito, 2003). The emphasis of the approach on well-being and the expansion of individual freedoms relates to an understanding of education as intrinsically valuable. Being educated provides and enhances the possibility of engaging in activities that contribute to one's fulfillment in life and are not simply instrumental in securing better jobs or positions in society. The capabilities approach highlights how, for instance, being introduced through education to the appreciation of literature, or science, or different kinds of music, leads to forms of personal fulfillment and personal reflection that have their own inherent value. On the other hand, the approach importantly points to the several instrumental aims served by education, which are fundamental in attaining better life prospects and career opportunities, as well as in promoting forms of social and economic development. Education, and specifically formal schooling (however imprecise and questionable this distinction may be), is crucial in securing standards of living, since it allows the acquisition of knowledge and skills that play an important role both for well-being and for the exercise of agency. In addition to this, as Robeyns notes (Robeyns, 2006b: 73), the roles of education are valuable also at the collective level. Economic growth and higher standards of living are promoted by educated people, as are more tolerant and diversified views of what constitutes a good life, thus indirectly enhancing the well-being and freedom of all.

The fundamental functions of education in the capabilities approach lead to a further set of considerations which see the approach as an innovative framework for the assessment of inequalities and for conceptualizing educational equality.

The Capabilities Approach and Justice in Education

One of the most distinctive insights of the capabilities approach to education concerns both its normative contribution to the theorization and analysis of inequalities in education, and its possible operationalization in evaluating

the justice of current educational systems. As we have seen, the approach provides a normative framework for the assessment of the just design of social and institutional arrangements, and suggests an idea of capabilities equality as equal opportunities for valued functionings. Applied to education, this framework shifts attention from the amount of resources spent on education, or from a consideration of education as a resource in itself, aimed at unilaterally promoting economic goals. It also provides a richer, and more articulated framework than those based on assessing educational outcomes, in terms of exam results and levels of qualifications. Instead, the approach suggests a comprehensive framework that has at its center the equalization of the actual effective opportunities for the achievement of important functionings in education, such as being able to read and to write, or to concentrate and accomplish tasks, or to reflect critically on one's own actions. While the exact specification of these fundamental functionings may, to some extent, be considered as the result of democratic processes similar, for instance, to those exercised in devising national curricula, the actual educational provision in terms of equal opportunities for functionings has a distinctively normative force. Here, effective and real opportunities are to be understood as including educational resources, settings such as school buildings and facilities, and external conditions such as policies and legislation that are necessary to promote educational achievement (Terzi, 2008). An example may illustrate this point. Consider, for instance, Emily and Emma, who have achieved different outcomes in mathematics. Suppose they are otherwise similar, and have both attended an excellent school. However, Emily has spent most of her time in leisure activities, whereas Emma has devoted more attention to her studies. The different achieved functionings in this case are not a matter of justice for the capabilities approach, since a degree of personal choice relates to the different outcomes. Suppose instead that these different achieved functionings are the result of Emma attending a much better equipped school than Emily's: in this case, the difference in functionings is a matter of justice, since it relates to a clear inequality of capabilities. The attention to capabilities, and their relation to achieved functionings, constitutes an innovative contribution to educational thinking about issues of justice. However, the actual operationalization of this framework to more practically oriented aims is still rather unexplored.

Furthermore, it must be noted that while referring to capabilities is paramount in relation to what governments can provide through public policy, the question becomes more complex with regard to the education of children, given their restricted agency, freedom, and choice. On this issue, Sen (Saito, 2003) highlights the future dimension entailed by education, while Nussbaum (2006b) specifies the importance of considering functionings, rather than capabilities, when enacting forms of compulsory education for children.

Pedagogy and Curriculum

Rather less explored than the previous two dimensions, the potential contribution of the capabilities approach to forms of pedagogy and the design of curricula relates to the expansion of individuals' freedoms and their reasoned agency. A focus on critical understanding and reflection on valuable knowledge, as well as on the capabilities for the formation of valued aims and life plans, is central to both capabilities-based pedagogy and curriculum. Questions arise in relation to the distinctive institutional designs and the forms of teaching and learning that would promote valued capabilities, while taking into due account the different sociocultural elements that might influence learning.

Nussbaum (1997, 2006b) explores the pedagogical dimension of an education aimed at freedom, in relation to her defense of a liberal project (as mentioned above), specifically, but not limited to, higher education. The three elements that are foundational to that project also provide valuable suggestions for pedagogical and curricular design. The critical examination of oneself, and of one's own and other traditions, informs not only pedagogy that privileges dialog and critical dispositions, but also logical reasoning and rigorous judgment. The understanding of oneself as a citizen of the world indicates the importance of forms of multicultural education, of historical and political knowledge, and of other languages. Finally, Nussbaum highlights ways in which the ability of narrative imagination is specifically cultivated by the arts and literature (Nussbaum, 2006b: 391). The importance of creativity relates also to the human-central capabilities of senses, imagination, and, finally, play. Nussbaum reinforces the idea that the exercise of these functionings is crucial for the formation of the adult's future capabilities, thus further guiding the design of curricula and of forms of pedagogy that enhance, rather than hinder, their development (Nussbaum, 2000).

While these are only some indications of ways in which pedagogy and curricula may be informed by a capabilities perspective, they nevertheless strongly suggest that forms of teaching and contents that undermine, exclude, devalue, or hinder human freedom are incompatible with the insights of the approach. However, more research is necessary for an application of these ideas to the policy and practice of education.

Summary

The centrality of well-being and the expansion of freedom advocated by the capabilities approach, its engagement with issues of justice and questions of equality, as well as with the value of democratic processes are all fundamental insights for educational theory, policy, and practice. While it is important to bear in mind that the capabilities approach is not an educational framework, and has not emerged in relation to educational concerns,

it is nevertheless equally important to highlight ways in which its theoretical and normative insights provide extremely important understandings of the value of education, of issues of educational justice, and of forms of pedagogy and the design of curricula that place individuals' well-being and their flourishing as central.

The insights suggested by the approach have the potential to contrast and provide valid counterarguments to current approaches to education, which narrowly focus on the acquisition of skills and competences unilaterally aimed at the furthering of economic goals, and to forms of pedagogy and curriculum that center on performativity rather than on the expansion of human freedom.

However, further theoretical and empirical research is needed to advance thinking along these lines.

See also: Globalization and Social Justice in Higher Education; Liberalism and Education; Wellbeing.

Bibliography

- Drèze, J. and Sen, A. (2002). *India: Development and Participation*. New Delhi: Oxford University Press.
- Nussbaum, M. C. (1997). *Cultivating Humanity: A Classic Defense of Reform in Liberal Education*. Cambridge, MA: Harvard University Press.
- Nussbaum, M. C. (2000). *Women and Human Development: The Capabilities Approach*. Cambridge: Cambridge University Press.
- Nussbaum, M. C. (2004). Women's education: A global challenge. *Sign: Journal of Women and Culture in Society* 29(2), 325–255.
- Nussbaum, M. C. (2006a). *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge: Cambridge University Press (The Tanner Lectures on Human Values).
- Nussbaum, M. C. (2006b). Education and democratic citizenship: Capabilities and quality education. *Journal of Human Development* 7(3), 385–396.
- Robeyns, I. (2005). The capability approach – a theoretical survey. *Journal of Human Development* 6(1), 93–114.
- Robeyns, I. (2006a). The capability approach in practice. *Journal of Political Philosophy* 14(3), 351–376.
- Robeyns, I. (2006b). Three models of education: Rights, capabilities and human capital. *Theory and Research in Education* 4(1), 69–84.
- Saito, M. (2003). Amartya Sen's capability approach to education: A critical exploration. *Journal of Philosophy of Education* 37(1), 17–33.
- Sen, A. (1980). *Equality of What? The Tanner Lectures on Human Values*, McMurrin, S. (ed.). Salt Lake City, UT: University of Utah Press.
- Sen, A. (1982). *Choice, Welfare and Measurement*. Oxford: Blackwell.
- Sen, A. (1992). *Inequality Reexamined*. Oxford: Clarendon Press.
- Sen, A. (1999). *Development as Freedom*. Oxford: Oxford University Press.
- Sen, A. (2002). Response to commentaries. *Studies in International Comparative Development* 27(2), 78–86.
- Terzi, L. (2005). A capability perspective on impairment, disability and special needs: Towards social justice in education. *Theory and Research in Education* 3(2), 197–223.
- Terzi, L. (2008). *Justice and Equality in Education: A Capability Perspective on Disability and Special Educational Needs*. London: Continuum.
- The Equalities Review (2007). *Fairness and Freedom: The Final Report of the Equalities Review*. London: Crown Publications.

Further Reading

- Alkire, S. (2002). *Valuing Freedoms: Sen's Capability Approach and Poverty Reduction*. Oxford: Oxford University Press.

- Berges, S. (2007). Why the capability approach is justified. *Journal of Applied Philosophy* 24(1), 16–25.
- Crocker, D. (1995). Functionings and capabilities: The foundations of Sen's and Nussbaum's development ethics. In Nussbaum, M. C. and Glover, J. (eds.) *Women, Culture and Development: A Study of Human Capabilities*, pp 153–199. Oxford: Clarendon Press.
- Deneulin, S., Nebel, M., and Sagovsky, N. (2006). *Transforming Unjust Structures: The Capability Approach*. Dordrecht: Springer.
- Kaufman, A. (ed.) (2006). *Capabilities Equality: Basic Issues and Problems*. New York: Routledge.
- Nussbaum, M. C. and Glover, J. (eds.) (1995). *Women, Culture and Development: A Study of Human Capabilities*. Oxford: Clarendon.
- Nussbaum, M. C. and Sen, A. (eds.) (1993). *The Quality of Life: A Study Prepared for the World Institute for Development Economics Research (WIDER) of the United Nations University*. Oxford: Clarendon Press.
- Sen, A. (2004). Capabilities, lists and public reason: Continuing the conversation. *Feminist Economics* 10(3), 77–80.
- United Nations Development Program (2000–2004). *The Human Development Report*. New York: United Nations.
- Unterhalter, E. (2007). *Gender, Schooling and Global Social Justice*. London: Routledge.
- Walker, M. (2006). *Higher Education Pedagogies: A Capabilities Approach*. Maidenhead: Open University Press.
- Walker, M. and Unterhalter, E. (eds.) (2007). *Amartya Sen's Capability Approach and Social Justice in Education*. New York: Palgrave Macmillan.
- Watts, M. (2006). What is wrong with widening participation in higher education. In Martin, W. and Satterthwaite, J. (eds.) *Discourse, Resistance and Identity Formation*. Stoke on Trent: Trentham Books.
- Wolff, J. and De-Shalit, A. (2007). *Disadvantage*. Oxford: Oxford University Press.

Relevant Websites

- <http://www.capabilityapproach.com> – The official website of the Human Development and Capability Association.
- <http://www.ophi.org.uk> – The official website of the Poverty and Human Development Initiative at the University of Oxford.

PRIMARY AND SECONDARY EDUCATION – LEARNING AND TEACHING IN SCHOOL AGE EDUCATION

Contents

Class Size – Arguments and Evidence

Continuing Professional Development of Teachers

Homeschooling

Innovations and Early Intervention in the Teaching of Literacy

Personalized Learning in School Age Education

Perspectives on Schooling in the Middle Years

Reform of the School Workforce

Secondary Education in Developing Countries

Class Size – Arguments and Evidence

P Blatchford, Institute of Education, University of London, London, UK

K C Lai, The Hong Kong Institute of Education, Hong Kong, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In many countries over the world, there has been a hotly contested and widely reported debate over the educational consequences of class-size differences. Opinions vary from those academics and policymakers who argue that class-size reduction is not cost effective to those who argue that it should be a cornerstone of educational policy. In some countries, policy has changed in favor of small classes. In the USA alone, 33 states have enacted legislation for class-size reduction (CSR) programs. Current government policy in England and Wales is for a maximum class size of 30 at reception and KS1 (4–7 years), and larger cuts are planned in Scotland. There have been initiatives involving class size or pupil to adult ratio reductions in the Netherlands and New Zealand. In East Asia, starting from the mid-1990s, many countries and cities (including Mainland China, Hong Kong, Korea, and Japan) have implemented policies to reduce class sizes. The controversy about class size has been very noticeable in Hong Kong where, in the context of a sharp decline in birth rates and school closure, mounting political pressure on the government to reduce class sizes led to the commissioning of a 3-year research project, only for political developments to lead to implementation of CSR in primary schools starting from 2009/10.

Despite the important policy and practice implications of the topic, there has been a wide gap between professional belief – that small classes are likely to lead to better teaching and learning – and research findings which have

not always been clear. There have been a number of comprehensive reviews of the class-size literature: Anderson (2000); Biddle and Berliner (2002); Blatchford and Mortimore (1994); Blatchford *et al.* (1998); Blatchford *et al.* (2009); Ehrenberg *et al.* (2001); Finn *et al.* (2003); Galton (1998); Grissmer (1999); Hattie (2005). Given a limited word length, in this article, we provide a selective review, describing important studies and identifying issues that seem to be important to us at this time. We address the issues of: (1) whether class-size differences affect children's educational attainment and learning, (2) whether class size affects classroom processes like teaching and pupil behavior, and (3) educational implications, in particular ways in which teachers can make the most of small (and large) classes. We provide a more international perspective than many reviews which have tended to be dominated by research in the USA. We cover recent developments in East Asia, where class-size effects have figured prominently in research and development. We also provide suggestions for future research.

Class Size and Educational Performance

Overall, it has often been pointed out that much previous research has not had designs strong enough to draw reliable conclusions about the effects of class-size differences (Blatchford *et al.*, 1998; Ehrenberg *et al.*, 2001; Finn *et al.*, 2003).

Correlational Designs

There has been considerable interest in the consistently superior performance of East Asian students in large-scale international studies, even though they have larger class sizes than most Western students (e.g., Pong and Pallas, 2001; Woessmann, 2006). This has led many to query the benefits of small classes. We take up below the issue of cultural factors in interpreting research findings, but here we make the general point that simple correlational designs, which examine associations between a measure of class size or pupil–teacher ratios (PTRs) on the one hand, and measures of pupil attainment on the other, are misleading. These studies – which often find that pupils in larger classes do better than pupils in smaller classes – are difficult to interpret, because we often do not know whether the results can be explained by another factor, for example, that poorer-performing pupils are placed in smaller classes.

Many econometric studies tend to use a measure of PTR rather than class size. It is not always realized that these two measures are not the same thing. PTRs are usually calculated by dividing the number of pupils on a school's roll by the full-time equivalent number of qualified teachers, and do not take into account noncontact time. It should not be assumed, therefore, that teachers who entered into the calculation are teaching for the whole time. PTRs are important from an administrative or economic point of view because they are closely related to the amount of money spent per child. However, in terms of pupil learning, the number of students physically present is more important. Researchers need to derive measures of class size as close as possible to that experienced by teachers and pupils. Here, we review several main studies of class-size effects on academic outcomes.

Student Achievement Guarantee in Education

The Student Achievement Guarantee in Education (SAGE) program began in Wisconsin in 1996 for K–3 pupils and involved several parallel interventions that targeted more disadvantaged schools and included reducing PTRs to 15 students. Results suggest that SAGE students significantly improved as a result of being in smaller classes, the gains for African-American students being greater than for white students (Ehrenberg *et al.*, 2001). One of the limitations of this study was that it involved changes to PTRs not class size as such and also that the effect of PTR reductions was not separable from other interventions in the program.

California

This began in 1996 and involved a massive state-wide implementation of CSR in grades K–3 at a cost of well

over 1.5 billion dollars and involving 1.8 million students. The overall aim was to reduce class sizes from 30 to 20 or less. Results showed a disappointingly small advantage in achievement for students in smaller classes but it is commonly recognized that it is very difficult to interpret results because the research design was limited. The intervention depended on the hiring of many unaccredited and inexperienced teachers and the use of portable classrooms and nonclassroom space (Bohrnstedt and Stecher, 1999).

To arrive at sounder evidence, two kinds of research design are needed. We concentrate on two studies which represent each approach.

Student/Teacher Achievement Ratio

It is often assumed that the problems of correlational research are best overcome by the use of experimental research or randomized controlled trials. Such studies are unique in potentially being able to provide clear evidence of causal direction. This is one reason for the great attention paid to the Tennessee Student/Teacher Achievement Ratio (STAR) project. A cohort of pupils and teachers at kindergarten through to grade 3 were assigned at random to three types of classes within the same school: a small class (around 17 pupils), a regular class (around 23 students), and a regular class with a teacher-aide. In brief, the researchers found that in both reading and mathematics, pupils in small classes performed significantly better than pupils in regular classes, and children from minority ethnic group backgrounds benefited most from small classes (Finn and Achilles, 1999). In fourth grade, the pupils returned to regular classes and the experiment ended, but gains were still evident after a further 3 years, that is, grades 4–6. Although some aspects of the project are still contentious, reanalysis of the data, using more sophisticated techniques, supported the central finding of a difference between small and regular classes (Goldstein and Blatchford, 1998).

Class Size and Pupil–Adult Ratio

Although experimental studies can provide sound evidence of causality, it should be recognized that they are extremely difficult and expensive to set up, and have validity problems of their own, not always appreciated (Goldstein and Blatchford, 1998). An alternative approach was adopted in a large-scale UK study the Class Size and Pupil–Adult Ratio (CSPAR) project (see e.g., Blatchford (2003); Blatchford *et al.* (2005); Blatchford *et al.* (2003); Blatchford *et al.* (2002). This studied the effect of class size on pupils' academic attainment and on classroom processes such as teaching, pupil attention, and pupil relations. It tracked over 10 000 pupils in over 300 schools from school entry (at 4/5 years)

to the end of the primary school stage (11 years). It employed a nonexperimental multimethod longitudinal design, measuring the effects of natural variations in class size with multilevel regression statistical analyses in order to determine effects of class size controlling for other factors, such as pupil prior attainment.

There was a clear effect of class size on children's academic attainment over the reception year (4/5 years), in both literacy and mathematics, even after adjusting for other possible confounding factors. The effect is comparable to that reported by the STAR project. Small classes (below 25) worked best in literacy for children who were most in need academically. Effects were still evident on literacy progress at the end of the second year of school (Year 1), though by the end of the third year, the effects were not clear. There were no clear longer-term effects of class-size differences on mathematics achievement. Although this indicates that the early benefits wash out after 2 years in school, there were no restrictions in terms of which size of class they moved to from year to year. Moving to a class of a different size, especially a larger class, had a negative disruption effect on progress.

Some issues connected to class-size effects on attainment

Results from both experimental and longitudinal observational studies are therefore consistent in showing positive effects of smaller classes on academic outcomes, in the first years of schooling. But there are several complicating factors when considering effects of class size on pupil attainment.

Overall class sizes

Research evidence is usually based on class sizes normally experienced in countries within North America, Europe, and Australasia. However, average class sizes can vary greatly between countries and in some countries can be very much larger. Hattie (2005) has pointed out that different class sizes will affect what is considered effective in teaching. Very different styles of teaching will be necessary, and different class size effects can be expected, when faced with class-size bands as different as 80+, 30–80, and 15–30 students in a class.

Age of student

The STAR and CSPAR projects are consistent in showing that effects on academic outcomes are clearest with the youngest students and there would seem to be clear support for policies involving CSR in the first years of school. There is no evidence that smaller classes benefit students later in their school careers. There is, though, still debate about whether CSR effects are best seen as age dependent or a start-up effect. If the latter, then it may be that CSR is advantageous soon after strategic points

of transition in student's school lives, for example, primary to secondary education. Research evidence to settle this debate is needed.

Other characteristics of students

Studies such as STAR, SAGE, and CSPAR suggest that CSR tends to benefit some students more than others, particularly lower attaining and disadvantaged pupils. This is probably because extra individual attention is especially beneficial.

Threshold effects?

It is often assumed that class sizes need to fall below a certain number (the figure of 20 is often mentioned) before they can have an impact on educational outcomes. However, this is likely to have much to do with the class sizes chosen in research. The STAR project, for example, compared classes of about 17 with class sizes of about 23 – and this is a main reason why the mid-point between the two is seen as important. However, this range of class sizes is not common in many countries, even in the USA, and an alternative approach is to examine effects of class size across the full range of class sizes, rather than presuppose class sizes likely to be important. In the CSPAR, this approach suggested that 25 or less was important for lower-attaining pupils in literacy.

In general, it is probably oversimplistic to talk about optimal class size in an exact way. Teachers' views about preferred class sizes are likely to be affected by what they have experienced and what they perceive as realistically achievable. Judgments are also likely to be affected by culturally bound views about teaching and about learning, and for these reasons it would not be surprising if views and effects differed between countries.

Alternatives to CSR

Conclusions about the efficacy of CSR are often made on the basis of comparisons with other interventions. Hattie (2005) argues that we should consider effects of class size not in relation to zero – that is, having no effect – but in comparison with other interventions, for example, one-to-one tutoring, phonics training, and Success for All. In general, CSR does not do well in these comparisons. But this is not a fair test in the sense that CSR is not an intervention like phonics training but simply involves changing the number of people in a room, with no control over what happens in the room. A fairer test is to compare it with effects of other, alternative classroom contextual changes (see below and Blatchford *et al.*, 2009).

School and cultural effects

For the most part, studies have examined effects of class size on processes and performance isolated from school and other local contexts within which they take place. But

teachers and pupils in any given class will operate in the context of the school ethos, social background, and local policies. Lai (2007) has shown the important role of school principals in providing a supportive environment for teachers to prepare for small-class teaching and engage in professional development.

One needs to be cautious about generalizing class-size effects in isolation from other factors in education systems (Biggs, 1998). In East Asia, research findings from the US and UK have been extensively cited by both proponents and critics of CSR, despite the fact that the class sizes studied were much smaller. Biggs (1998) has argued that Westerners often wrongly perceive large classes in Confucian heritage cultures (CHCs), such as China, Japan, Korea, and Singapore, as conducive to teacher dominance and passive and rote learning. Meaningful learning can take place in whole-class teaching as children in CHCs are socialized in ways that make them amenable to work in large classes.

Effects of Class Size in Relation to Classroom Processes

It is now widely appreciated that attention needs to move from studies of the effects on academic outcomes to better understanding of the classroom processes that might be involved (Anderson, 2000; Finn *et al.*, 2003; Grissmer, 1999). Unfortunately there are methodological weaknesses in much research in this area. Studies have been relatively anecdotal with a reliance on teacher opinion and comment (Blatchford *et al.*, 1998). Finn *et al.* (2003) have pointed out that there is relatively little strong systematic observational research which would permit objective study of relationships between class size, teaching practices, and student behavior. Much of the evidence from China and Hong Kong is anecdotal in nature (e.g., Lai and Ip, 2007).

Teacher's Individual Attention to Pupils

Perhaps the most consistent finding is that the most important classroom process, affected by reduced class size, is individualization of teaching (Ehrenberg *et al.*, 2001). Results from the CSPAR systematic observation studies (e.g., Blatchford *et al.*, 2005) showed that though there was a heavy reliance on whole-class teaching and individual work in primary schools; pupils in small classes were more likely to experience one-to-one teaching and were more often the focus of a teacher's attention. Separate research has shown this trend continues into secondary schools (Blatchford *et al.*, 2008). In this vein, Anderson (2000) sees small classes encouraging a more personalized and appropriate curriculum for individual pupils.

Easier Classroom Control and Management

A number of studies have reported that pupil discipline tends to be more difficult in large classes and more of an intrusion into the teaching and learning process. In contrast, smaller classes tend to be quieter and more easily managed. Bourke (1986) found that more nonacademic procedural arrangements were necessary in large classes. Once again, cultural factors may be important: Lai (2007) found some teachers in Hong Kong and Shanghai felt overwhelmed by the increased level of student activity and movement in small classes, compared with traditionally large classes under tight teacher control.

Less Teacher Stress and Better Morale

Many studies report that teachers are put under more strain when faced with large classes. Teachers' morale is better in smaller classes (Biddle and Berliner, 2002). This probably comes about because of the increased demands on them and also because they are faced with compromises to their preferred pedagogy (Blatchford *et al.*, 2007). The underlying point, which informs much research stemming from teacher report, is that teachers feel better about teaching in small classes.

Better Relationships with and Knowledge of Pupils

Teachers in small classes tend to experience better relationships with, and have more knowledge of, individual pupils (Finn *et al.*, 2003). In smaller classes, it can be easier for teachers to spot problems and give feedback, identify specific needs, and gear teaching to meet them, and set individual targets for pupils.

Active Involvement with Teacher

Larger classes can lead to pupils having a passive role in class. Research in the UK found that children in large classes were more likely to simply listen to the teacher but in smaller classes, pupils were more likely to interact in an active, sustained way with teachers (Blatchford *et al.*, 2005). Bourke (1986) found more student questions to teachers in large classes but these were mostly requests for clarification or for other help from the teacher.

Pupil Attentiveness/On-Task Behavior

Finn *et al.* (2003) developed a theoretical and empirical case for why student classroom engagement is the key process that explains why smaller classes lead to better attainment. They conclude that class size seems to affect student engagement more than teaching. Research in England has found student engagement increases with smaller classes, but this particularly applied to low-attaining

students (Blatchford *et al.*, 2008). However, the CSPAR study found no effects on pupil attentiveness in 10/11-year-old pupils, and Bourke (1986) found no class-size effect on primary school student engagement.

Peer Relations

It might be expected that in larger classes there would be more negative and aggressive behaviors between children. This is supported by some reviews (e.g., Finn *et al.*, 2003) but other research with older pupils seems less clear. In the CSPAR, it was not found that pupils in smaller classes had better peer relations; indeed, peer relations were if anything worse (Blatchford, 2003).

Size and Number of within-Class Groups

Class size and grouping of pupils in the classroom are closely linked. As the size of the class increases, the size and/or number of groups necessarily increases. Group size can have effects on teaching, for example, larger groups can result in more off-task behavior, and miss out on a teacher's attention (Blatchford, 2003). Bourke (1986) found that teachers in larger classes tended to form more groups during mathematics lessons and that this led to fragmentation of the lesson and inefficient use of the teacher's time. Worryingly, research finds a tendency for less groupwork to take place in smaller classes, no doubt because teachers are availing themselves of the opportunities for more contact with individual pupils (Blatchford *et al.*, 2009).

Curriculum Effects

Class-size effects on classroom processes can vary by school subject. Rice (1999) found that in mathematics, but not science, as class size increased, less time was spent on small groups and individuals, innovative instructional practices, and whole-group discussions. In the CSPAR study, the overall effects of class size on individualized attention were found in all subjects but English, probably because English is the most discussion-based subject and relies less on questioning of individual pupils. (Blatchford *et al.*, 2005). Interim findings from classroom observations in the Hong Kong government's longitudinal study (Hong Kong Education and Manpower Bureau, 2007) revealed signs, particularly among teachers of Chinese and mathematics, that participant teachers were changing their practice in ways that resulted in higher levels of problem-solving questioning and a greater range of feedback responses, although, overall, there was little evidence of a dramatic change in teaching modes.

One direction for future research would be to identify more precisely ways in which class-size effects vary in relation to particular school subjects and student age, and to explore factors that explain any differences found.

Research has tended to identify associations between class size and individual teacher and pupil variables, rather than position these in an overall conceptual framework. With some exceptions, for example, Finn *et al.* (2003), research has tended to be atheoretical and there is a lack of an overriding conceptual understanding of class-size effects. Following the lead of Anderson (2000), it is important for future advances in this field that possible frameworks are advanced and tested. Further, much more effort needs to be put into testing causal connections linking class size, type of pupil, teaching practices, pupil behavior, and academic outcomes.

Implications for Practice

Research therefore suggests that small classes can benefit pupils academically. There can also be benefits in terms of classroom teaching but it has often been pointed out that teachers do not necessarily change the way they teach when faced with smaller classes. This might well account for the relatively modest effects of class size on achievement. One reason for this lack of change is that existing pedagogical and curricular approaches put constraints on teachers. There is a need for more research to understand how teachers can take advantage of reduced class sizes. Lai (2007) investigated teachers' experiences in changing from teaching in large to small classes, and found that teachers had different approaches to change. In Chinese cities implementing small-class teaching, there are many action-research reports and reflective accounts published by principals and teachers in books and journals and presented in seminars (e.g., Nanjing Education Bureau, 2007; Zhao, 2002). Though valuable, it is difficult to determine the scale of impact in the absence of large-scale empirical studies. Future work could do well to coordinate larger-scale and/or linked studies.

How should teachers change the way they teach? As Galton (1998) has said, one problem is that we do not have a lot of knowledge about effects of class size on teaching on which to base practical advice. In addition, and more fundamentally, we do not have a well-worked-through theory underpinning teaching and pedagogy, and teachers do not have a theory to draw on. It may be helpful to distinguish those processes most likely to follow from differences in class size and those that are potentially likely to follow but which will depend on what a teacher makes of a smaller class. Drawing from the review above, it seems likely that the following things are likely to follow from smaller classes:

1. more individualization of teaching;
2. easier classroom control;
3. more time for marking, assessments, and planning; and
4. less teacher stress.

One needs to be careful, however, about the tendency toward more individualization in smaller classes because classroom management can become very teacher centered and there can be little development of student independence and control of learning. This is why Blatchford *et al.* (2009) have suggested several ways in which CSR can be accompanied by pedagogical changes to enhance beneficial effects for students. These include taking advantage of the possibilities of more-effective collaborative learning between pupils, and more adventurous and flexible teaching. The experiences in Chinese cities implementing CSR show that small-class teaching has played a key role in teachers' professional development and set the conditions for student-centered learning, creativity, inquiry learning, assessment change, and delivery of the new curriculum (Zhang, 2007). This can be further strengthened by professional-development networks (Lai and Ip, 2007).

Overall, it may be too simplistic, especially in an international context, with its huge variety of resourcing, structures, cultures, and pedagogies, to seek to identify particular educational practices that will be affected by class size and which should be adopted in small classes. Rather, we may do better to be clear as educators about the educational aims that we consider important and then think through carefully where CSRs can help. Teachers will vary in their educational goals for very good reasons – because of the age of pupils, the subject area, the emphasis at a given point in the year, in response to a particular cohort of pupils, and because of fundamental differences of opinion about appropriate pedagogical approaches. Teachers may be better equipped, when given the opportunities afforded by small classes, if they consider educational principles, for example, increasing mastery orientation and teaching for understanding, rather than specific practices, and then to consider how small classes can help. This is line with Achilles (1999) who argues that small classes can allow teachers to do what they feel is the right thing to do.

Some have argued that teacher professional development is a better investment than CSR, but as illustrated by the examples in East Asia it is better not to see them in opposition. Rather, professional development should be used to help teachers harness the opportunities of small classes, and help teachers develop strategies for realizing educational objectives in large classes. We feel that there could be a much bigger role in professional development for a close consideration of classroom-contextual features, of which the number of children in the class is one. CSR should also be addressed at the whole-school level, to ensure that structural and instructional approaches make the most of CSR, and that planning with classroom teachers ensures that their efforts complement each other. Smaller classes can facilitate opportunities for teachers to teach and students to learn.

We therefore suggest that CSR should be accompanied by pedagogical changes to enhance beneficial effects for students. We also suggest that CSR should be evaluated in conjunction with these particular interventions, that is, it is the combined effect of CSR along with appropriate pedagogical and curricular changes that is of most relevance, and which needs to be evaluated.

Bibliography

- Achilles, C. A. (1999). *Let's Put Kids First, Finally: Getting Class Size Right*. Thousand Oaks, CA: Corwin Press.
- Anderson, L. (2000). Why should reduced class size lead to increased student achievement? In Wang, M. C. and Finn, J. D. (eds.) *How Small Classes Help Teachers Do Their Best*, pp 3–24. Philadelphia, PA: Temple University Center for Research in Human Development and Education.
- Biddle, B. J. and Berliner, D. C. (2002). What research says about small classes and their effects. Part of series. In *Pursuit of Better Schools: What Research Says*. <http://www.WestEd.org/policy/perspectives> (accessed September 2009).
- Biggs, J. (1998). Learning from the Confucian heritage: So size doesn't matter? *International Journal of Education Research* **29**, 723–738.
- Blatchford, P. (2003). *The Class Size Debate: Is Small Better?* Maidenhead: Open University Press.
- Blatchford, P., Bassett, P., and Brown, P. (2005). Teachers' and pupils' behaviour in large and small classes: A systematic observation study of pupils aged 10/11 years. *Journal of Educational Psychology* **97**(3), 454–467.
- Blatchford, P., Bassett, P., and Brown, P. (2008). Do low attaining and younger students benefit most from small classes? Results from a systematic observation study of class size effects on pupil classroom engagement and teacher pupil interaction. *Paper to American Educational Research Association Annual Meeting*, New York.
- Blatchford, P., Bassett, P., Goldstein, H., and Martin, C. (2003). Are class size differences related to pupils' educational progress and classroom processes? Findings from the Institute of Education class size study of children aged 5–7 Years. *British Educational Research Journal* **29**(5), 709–730. Special Issue: *In Praise of Educational Research*. Gorrard, S., Taylor, C. and Roberts, K. (eds.).
- Blatchford, P., Goldstein, H., and Mortimore, P. (1998). Research on class size effects: A critique of methods and a way forward. *International Journal of Educational Research* **29**, 691–710.
- Blatchford, P., Moriarty, V., Edmonds, S., and Martin, C. (2002). Relationships between class size and teaching: A multi-method analysis of English infant schools. *American Educational Research Journal* **39**(1), 101–132.
- Blatchford, P. and Mortimore, P. (1994). The issue of class size in schools: What can we learn from research? *Oxford Review of Education* **20**(4), 411–428.
- Blatchford, P., Russell, A., Bassett, P., Brown, P., and Martin, C. (2007). The effect of class size on the teaching of pupils aged 7–11 years. *School Effectiveness and Improvement* **18**(2), 147–172.
- Blatchford, P., Russell, A., and Brown, P. (2009). Teaching in large and small classes. In Saha, L. J. and Dworkin, A. G. (eds.) *International Handbook of Teachers and Teaching*, pp 779–790. New York: Springer.
- Bohrnstedt, G. W. and Stecher, B. M. (eds.) (1999). *Class Size Reduction in California: Early Evaluation Findings, 1996–98*. Palo Alto, CA: American Institutes for Research.
- Bourke, S. (1986). How smaller is better: Some relationships between class size, teaching practices, and student achievement. *American Educational Research Journal* **23**(4), 558–571.
- Ehrenberg, R. G., Brewer, D. J., Gamoran, A., and Willms, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest* **2**(1), May.

- Finn, J. D. and Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis* **21**(2), 97–109.
- Finn, J. D., Pannozzo, G. M., and Achilles, C. M. (2003). The 'why's' of class size: Student behaviour in small classes. *Review of Educational Research* **73**(3), 321–368.
- Galton, M. (1998). Class size: A critical comment on the research. *International Journal of Educational Research* **29**, 809–818.
- Goldstein, H. and Blatchford, P. (1998). Class size and educational achievement: A review of methodology with particular reference to study design. *British Educational Research Journal* **24**(3), 255–268.
- Grissmer, D. (1999). Class size effects: Assessing the evidence, its policy implications, and future research agenda. *Educational Evaluation and Policy Analysis* **21**(2), 231–248.
- Hattie, J. (2005). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research* **43**, 387–425.
- Hong Kong Education and Manpower Bureau (2007). *Updated Background Brief Prepared by the Legislative Council Secretariat for the Meeting on 16 April 2007: Small Class Teaching*. Paper Submitted to the Legislative Council Panel on Education. No. CB(2) 1545/06-07(02).
- Lai, K. C. (2007). Understanding teacher change in small class teaching. *Presentation at the International Symposium on Class Size: Research, Policy and Practice*, The Hong Kong Institute of Education (February 10).
- Lai, K. C. and Ip, K. Y. (2007). *Seeing Big from Small: Portfolios of Small Class Teaching* (in Chinese). Hong Kong: Step Forward Press.
- Nanjing Education Bureau (2007). *Towards Quality Education: An Exploration of Teachers' Professional Development in Small Class Teaching* (in Chinese). Nanjing: Nanjing Education Bureau.
- Pong, S. L. and Pallas, A. (2001). Class Size and eighth-grade math achievement in the United States and abroad. *Educational Evaluation and Policy Analysis* **23**(3), 251–273.
- Rice, J. K. (1999). The impact of class size on instructional strategies and the use of time in high school mathematics and science courses. *Educational Evaluation and Policy Analysis* **21**(2), 215–229.
- Woessmann, L. (2006). Where to look for student sorting and class-size effects: Identification and quest for causes based on international evidence. In Howie, S. J. and Plomp, T. (eds.) *Contexts of Learning Mathematics and Science: Lesson Learned from TIMSS*, pp 353–368. New York: Routledge.
- Zhang, M. X. (2007). Teachers' professional development in the context of small class teaching. *Presentation at the Second Changjiang Delta Symposium on Small Class Teaching*, (in Chinese), Shanghai, 30 November.
- Zhao, C. Y. (ed.) (2002). *"Small" Class Size, "Large" Education* (in Chinese). Shanghai: Shanghai Yangpu District Education Bureau.
- International Handbook of Educational Research in the Asia-Pacific Region*, pp 741–754. Dordrecht: Kluwer.
- Cooper, H. M. (1989). Does reducing student-to-teacher ratios affect achievement? *Educational Psychologist* **24**(1), 79–98.
- Day, C., Toley, H., Hadfield, M., Parkin, E., and Watling, G. R. (1996). *Class Size Research and the Quality of Education: A Critical Survey of the Literature Related to Class Size and the Quality of Teaching and Learning*. Haywards Heath: National Association of Head Teachers.
- Evertson, C. M. and Randolph, C. H. (1989). Teaching practices and class size: A new look at an old issue. *Peabody Journal of Education* **67**(1), 85–105.
- Finn, J. D. and Wang, M. C. (2002). *Taking Small Classes One Step Further*. Greenwich, CT: Information Age Publishing/Laboratory for Student Success.
- Glass, G., Cahen, L., Smith, M. L., and Filby, N. (1982). *School Class Size*. Beverley Hills, CA: Sage.
- Robinson, G. E. and Wittebols, J. H. (1986). *Class Size Research: A Related Cluster Analysis for Decision Making*. Arlington, VA: Educational Research Service.
- Shapson, S. M., Wright, E. N., Eason, G., and Fitzgerald, J. (1980). An experimental study of the effects of class size. *American Educational Research Journal* **17**, 144–152.
- Slavin, R. E. (1989). Class size and student achievement: Small effects of small classes. *Educational Psychologist* **24**(1), 99–110.
- Wang, M. C. and Finn, J. D. (eds.) (2000). *How Small Classes Help Teachers Do Their Best*. Philadelphia, PA: Temple University Center for Research in Human Development and Education.
- Wilson, V. (2006). Does small really make a difference? An update. *A Review of the Literature on the Effects of Class Size on Teaching Practice and Pupils' Behavior and Attainment*. Scottish Council for Research in Education (SCRE) Centre: University of Glasgow, Glasgow.

Relevant Websites

- <http://www.breaktime.org.uk> – Breaktime Project.
- <http://www.ied.edu.hk> – Centre for Development and Research in Small Class Teaching, The Hong Kong Institute of Education.
- <http://www.classsizematters.org> – Class Size Matters.
- <http://www.classsize.org> – CSR Research Consortium.
- <http://www.heros-inc.org> – Health and Education Research Operative Services Inc.: Class Size Research.
- <http://www.classsizeresearch.org.uk> – Institute of Education: Class Size Research project.
- <http://www.ioe.ac.uk> – Institute of Education: Psychology and Human Development.
- <http://www.njxbh.cn> – Nanjing Small Class Education Website (Primary).
- <http://www.reduceclasssizenow.org> – Reduce Class Size Now.
- <http://www.SPRinG-Project.org.uk> – SPRinG Project.
- <http://www.supportstaffresearch.org.uk> – Support Staff Research Project, Institute of Education.
- <http://www.weac.org> – Wisconsin Education Association Council, SAGE Resource Page.

Further Reading

- Betts, J. R. and Shkolnik, J. L. (1999). *Special Issue: The Behavioral Effects of Variations in Class Size: The Case of Math Teachers*. *Educational Evaluation and Policy Analysis* **21**(2), 193–213.
- Blatchford, P. and Catchpole, G. (2003). Class size and classroom processes. In Keeves, J. P. and Watanabe, R. (eds.)

Continuing Professional Development of Teachers

P Earley, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

The ongoing or continuing professional development (CPD) of teachers is crucial for organizational growth and school improvement. One of the hallmarks of being identified as a professional is to continue to learn throughout a career and teachers' CPD is a key component for developing children's learning and also for the implementation of policy, be it government or school policy. As Day and Sachs (2004) state in the introduction to their *International Handbook of the Continuing Professional Development of Teachers*, it is often used to describe all the activities in which teachers engage during the course of a career which are designed to enhance their work; yet, they state "this is a deceptively simple description of a hugely complex intellectual and emotional endeavour which is at the heart of raising and maintaining standards of teaching, learning and achievement in a range of schools, each of which poses its own set of special challenges" (2004: 3). Bolam and McMahon (2004) go further and argue for a conceptual map which they say despite a huge CPD literature has not been forthcoming. In part, they suggest this is because CPD studies:

are written from diverse perspectives, for instance focusing on individual teacher self-development, classroom or subject teaching, school organisation and leadership, national policy-making or theoretical development. It is also because they draw upon a range of theoretical sub-fields, some of which have a strong discipline base (for example, sociology, cognitive psychology, social psychology and occupational psychology) and some of which are more eclectic (such as teacher education and training, andragogy, evaluative research, change theory, management studies and policy studies). (Bolam and McMahon, 2004: 52)

One consequence of this is that there are too many rather than too few approaches and theories, a situation of conceptual pluralism. Bolam and McMahon distinguish between four categories of CPD literature:

- 'knowledge for understanding,' which includes theoretical and critical policy analysis
- 'knowledge for action,' which includes evaluations
- 'policy makers,' which includes national and state level policy statements, and
- 'practitioner' literature, which includes two subsets – instrumentalist and reflexive – and consists mainly of accounts of practice and methods (Bolam and McMahon, 2004: 53).

Given this conceptual pluralism and diversity of perspectives it is necessary to state clearly that in considering CPD this article draws from all categories of literature and is written predominantly from a management and policy studies perspective. More specifically, this article begins by asking what is CPD, before considering entitlements and responsibilities, what we know about what works, and how we evaluate the impact of teachers' CPD.

What is CPD?

CPD has become the phrase or term widely used for the ongoing education and training of teachers, while in England the term workforce development is becoming more common in recognition of the fact that schools increasingly employ as many, if not more, support staff than teachers. If teaching is seen as a profession – and a case for this has long been argued – an important characteristic or hallmark of a member of a profession is the commitment its members show toward self-improvement or development. This is not, however, for its own sake but to ensure that the beneficiaries or clients – pupils and parents – are provided with the best possible service. The responsibility for securing individual professional development of teachers is not, however, the exclusive concern of the employer – teachers themselves must expect to play a key role – and professional development opportunities must be available to help them become better practitioners.

However, what do we mean by the term CPD and is it different from personal development or staff development or in-service education and training (inset)? Broadly speaking, CPD encompasses all formal and informal learning that enables individuals to improve their own practice. It involves "reflective activity designed to improve an individual's attributes, knowledge, understanding and skills" (TDA, 2007) and it means maintaining, improving, and broadening relevant knowledge and skills in subject specialisms and teaching so that it has a positive impact on practice and learner experience.

Professional development is an aspect of personal development and, wherever possible, the two should interact and complement each other. The former is mainly about occupational role development, whereas personal development is about the development of the person, often the whole person, and it almost always involves changes in self-awareness. As Waters explains: "It is the development

that can occur when teachers are construed first and foremost as people, and is predicated on the premise that people are always much more than the roles they play” (Waters, 1998: 30).

A definition of CPD might refer to: “any professional development activities engaged in by teachers which enhance their knowledge and skills and enable them to consider their attitudes and approaches to the education of children, with a view to improve the quality of the teaching and learning process” (Bolam, 1993). In this sense, it is perhaps a little different from how some have defined in-service training or staff development. The seminal James Report (DES, 1972) defined INSET as: “the whole range of activities by which teachers can extend their personal education, develop their professional competence and improve their understanding of education principles and techniques.”

An analysis of the literature does, however, reveal a number of nuances and slight differences between the different concepts used. A simple but most useful conceptual breakdown is offered by Bolam (1993) where he makes use of a threefold distinction among:

- professional training, for example, short courses, workshops and conferences emphasizing practical information and skills;
- professional education, for example, long courses and secondments emphasizing theory and research-based knowledge;
- professional support, for example, activities that aim to develop on the job experience and performance.

CPD embraces those education, training, and support activities that teachers engage in, following their initial certification, which aim to add to their professional knowledge; improve their professional skills; help clarify their professional values; and enable pupils to be educated more effectively (Bolam, 1993).

CPD is an ongoing process, building upon initial teacher training (ITT) and induction, including development and training opportunities throughout the career and concluding with preparation for retirement. At different times and at different stages, one or other may be given priority; however, the totality can be referred to as CPD. Development is about improvement, both individual and school improvement.

In a survey of continuing education for the professions, Madden and Mitchell (1993) state that CPD can fulfill three functions:

- updating and extending the professional’s knowledge and skills on new developments and new areas of practice – to ensure continuing competence in the current job;
- training for new responsibilities and for a changing role (for example, management, budgeting, teaching) – developing new areas of competence in preparation for a more senior post;

- developing personal and professional effectiveness and increasing job satisfaction – increasing competence in a wider context with benefits to both professional and personal roles.

Other definitions (e.g., Day, 1999: 4) are more wide ranging and go well beyond the mere acquisition of knowledge or skills. Bubb and Earley (2007) see staff development as:

an on-going process encompassing all formal and informal learning experiences that enable all staff in schools, individually and with others, to think about what they are doing, enhance their knowledge and skills and improve ways of working so that pupil learning and wellbeing is enhanced as a result. It should achieve a balance between individual, group, school and national needs; encourage a commitment to professional and personal growth; and increase resilience, self-confidence, job satisfaction and enthusiasm for working with children and colleagues. (Bubb and Earley, 2007: 4)

Alternatively and put more simply, it is about creating opportunities for adult learning, ultimately for the purpose of enhancing the quality of education in the classroom. To summarize, CPD is an ongoing process of education, training, learning, and support activities which is:

- taking place in either external or work-based settings;
- engaged in by qualified, educational professionals;
- aimed mainly at promoting learning and development of teachers’ professional knowledge, skills, and values; and
- helps decide and implement valued changes in their teaching and learning behavior so that they can educate their students more effectively, thus achieving an agreed balance between individual, school, and national needs.

It is clear that long gone are the days when initial training and induction were seen as a total or final preparation for a career in teaching; nowadays they have to be seen as merely providing a platform on which further professional development will be built. Nevertheless, the initial period in teaching is crucial as the experience of the first year is most formative. There is therefore a need to set high expectations and standards when there is the greatest receptiveness and willingness to learn and develop.

Perhaps the single most important feature of CPD is to encourage and promote a commitment on the part of the individual to professional growth. Leading and managing human resource development – making CPD work – therefore means providing structures and procedures to coordinate developmental opportunities so as to promote such growth and to help staff develop and improve their workplace performance.

The CPD of teachers has to be seen as a collective responsibility – the responsibility of both individual staff and the schools and colleges in which they work. Individuals and their places of employment should take joint responsibility for professional development and training, which should be for the benefit of both. Growing attention has been given to organizational cultures and the emphasis that may or may not be given to the training and development or the learning of its members. The experience and expertise of staff – both teaching and support – are generally recognized to be the school's most important and most expensive resource. The term learning organization or professional learning community is becoming more commonly known and attempts are being made in many national education systems to ensure that more schools are aware of what this means for themselves. Leading and managing people and their development have to be seen as a central part of the responsibility of managing the school's total resources. Governments throughout the world are now turning their attention to this particularly as schools become more autonomous and have a major responsibility in managing their people.

An Entitlement to CPD

The notion of an entitlement to CPD is something that has been promoted in many countries in the developed world. The General Teaching Council in England and Wales sees an entitlement to professional development as “career long and sustained so that on entry to the profession a teacher has a clear expectation of continuing, relevant and planned professional development.” England's General Teaching Council (GTC, 2003) has drawn up an entitlement to professional learning and states that teachers need the opportunity to:

- Have structured time to engage in sustained reflection and structured learning;
- Create learning opportunities from everyday practice such as planning and assessing for learning;
- Develop their ability to identify their own learning and development needs and those of others;
- Develop an individual learning plan;
- Have school-based learning as well as course participation, recognized for accreditation;
- Develop self-evaluation, observation and peer review skills;
- Develop mentoring and coaching skills and their ability to offer professional dialogue and feedback;
- Plan their longer-term career aspirations (GTC, 2003: 6).

In some education systems, teachers and employers are seen as having different but complementary responsibilities in relation to CPD (Jones, 2003):

- *Employers.* To provide professional development opportunities for teachers to support a broad range of priorities which occur during the normal work cycle and an entitlement to professional development which focuses on the individual professional and personal needs and objectives of the teacher is emphasized.
- *Teachers.* To develop themselves as reflective professionals by reflecting on their work and by identifying new ways of working. These activities should be undertaken as part of a teachers' work.

In others, the emphasis continues to be on seeing CPD as a way of meeting national policies or organizational objectives. In more mature systems, there is a focus on the individual, the organization and national/local priorities (see **Figure 1**; Jones, 2003: 37).

Meeting teachers' needs while also addressing whole-school and collective needs is a challenge. A more personalized approach is crucial if limited CPD funds are deployed most effectively. Teachers' previous knowledge and experience needs to be taken into account and the school will need to respond, if possible, to individuals' preferences and learning styles. Bolam and Weindling (2006) in a detailed review of 20 recent UK research studies into CPD provide strong evidence “that the more influence teachers have over their own CPD the more likely they are to consider it effective. More generally, the (findings) lend support to the importance of teachers' professionalism and agency as key components of effective CPD” (Bolam and Weindling, 2006: 3).

With entitlements come responsibilities and professional development is seen as a responsibility throughout a teacher's career. However, in many education systems with the change of emphasis on the importance of training and development and greater control from the center, there has been a corresponding move toward a higher degree of prescription and statutory requirements regarding CPD. In some other professions there is specified mandatory CPD for all its members. Since 2007, for example, teachers and lecturers in the English further education (post 16) sector were expected to clock up 30 h of CPD, something which teachers in Scotland have been required to do for some time. What is key however is not the specified amount of CPD undertaken but rather its effectiveness.

What is Effective CPD?

Any discussion of what constitutes effective professional development needs to begin by asking effective for whom – the individual, the school and its pupils, or the education system as a whole? Also, for what? The beneficiaries of CPD may not always overlap, although in many cases they will; what is helpful for an individual is also likely to be of benefit both for pupils and for the school. On the other

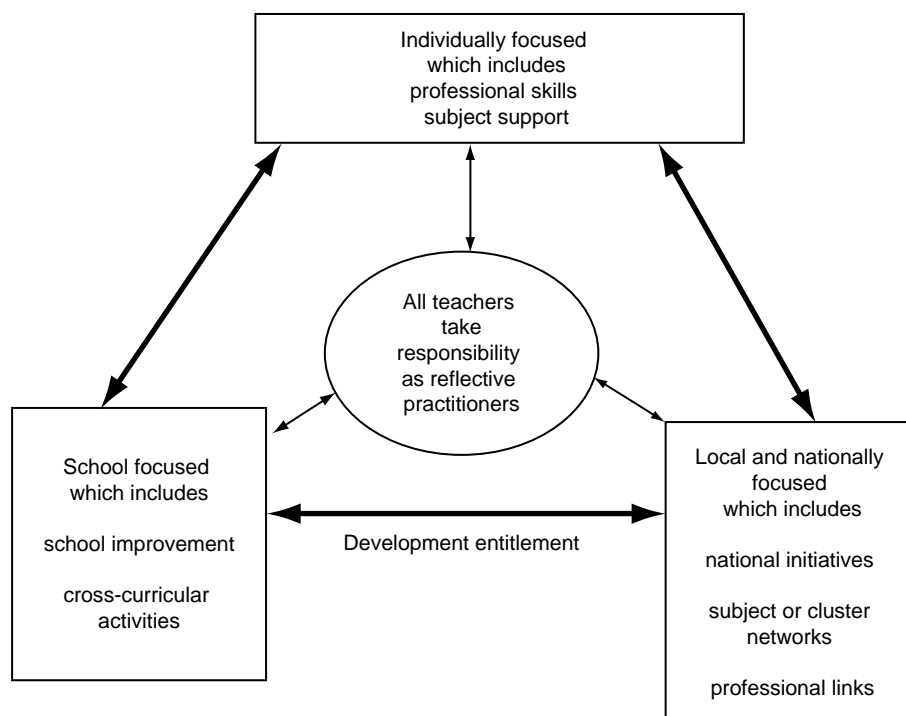


Figure 1 Continuing professional development framework. From Jones, T. (2003). Continuing professional development in Wales: An entitlement for all. *Professional Development Today* 6(1), 37.

hand, training to meet the latest government initiative may be effective for implementing a particular strategy but not necessarily for the overall long-term development of staff or the school. Effective CPD is likely to consist of that which first and foremost enhances pupil outcomes, but which also helps to bring about changes in practice and improves teaching, as well as management and leadership skills and qualities. It should be fit for purpose and add something to the school's overall capacity to develop; the school should be able to build upon the collective learning of its people. Exposure to and participation in a wide range of professional development opportunities are likely to bring about change to individuals' beliefs, values, attitudes, and behaviors, and these may well lead to changes in classroom and school practices.

Much has been written about what constitutes effective CPD but it should be remembered that the characteristics influencing effectiveness are multiple and highly complex. It is worth noting an important caveat: the technical and methodological problems of measuring the direct effects or impact of CPD, especially on classroom performance – of both staff and pupils – are considerable. This is, among other reasons, because the term embraces a wide variety of modes and methods, and includes professional training, education, and support. The number of variables to be taken into account is considerable.

Teachers' CPD can only have an indirect impact on pupil learning and outcomes. Very few research studies have considered the relationship between the

characteristics of professional development and any change in classroom teaching practice or the gains in pupil achievement or learning outcomes more broadly. Despite this, there is a growing body of research evidence and informed professional judgment that gives a broad indication of the essential characteristics of effective CPD. These include the clear identification of aims and objectives, and an analysis of training needs to ensure training and development activities matched existing levels of expertise. Opportunities for reflection are also important, as are action research, ongoing evaluation, and follow-up work. Effective training was seen to form part of a coherent program and was not a one-off activity. The optimum use of existing resources and facilities was most likely when CPD activities were well planned, helping to promote targeted and tailor-made training, avoiding overload and minimizing disruption (Bubb and Earley, 2007). The practice in schools however is often quite different.

The range of professional development activities which teachers can access is huge and can be categorized as on-the-job, off-the-job, and close-to-the-job opportunities. A recent report from inspectors in England has described the management of professional development as a logical chain of procedures, which entails "identifying school and staff needs, planning to meet those needs, providing varied and relevant activities, involving support staff alongside teachers, monitoring progress and evaluating the impact of the professional development." They note that schools which had designed their CPD effectively and integrated it

with their improvement plans found that teaching and learning improved and standards rose (Ofsted, 2006: 2).

Fragmented one-shot workshops and conferences at which staff listen passively to experts rarely make a difference. These traditional approaches to professional development appear insufficient to foster learning which fundamentally alters what teachers teach or how they teach. The following types of professional development activities are more likely to offer sustained learning opportunities:

- study groups in which staff are engaged on regular, structured and collaborative interactions around topics identified by the group;
- observation – observing others and being observed;
- coaching or mentoring arrangements, where staff work one-on-one with an equally or more experienced teacher;
- networks, which link people either in person or electronically, to explore and discuss topics of interest, pursue common goals, share information and address common concerns;
- immersion in inquiry, in which people engage in the kinds of learning that they are expected to practice with their students (Boyle *et al.*, 2003: 3).

Involvement in development activities by staff is, to a large extent, voluntary but there are a number of factors that will encourage participation. Boring, repetitive, and dependent work discourages professional development and growth, whereas challenging, variable, and independent work encourages it – certainly there are parallels with the classroom here. Equally, personal factors and life changes can cause individuals to reconsider career priorities and goals. Professional development can help enhance performance through improved self-esteem – this, along with personal well-being and a sense of professional control, are all essential components of job satisfaction.

The kinds of teacher professional development which make the most difference to practice are based on professional dialog about teaching and learning and the improvement of practice through a variety of activities, including coaching, mentoring, shadowing, and peer support. Awareness-raising events are useful for absorbing information and updating knowledge but are not likely to lead to skills development. Joyce and Showers (2002) concluded that, for training to be truly effective, it needs to include the following five components or stages:

- theory – where the new approach is explained and justified;
- demonstration – to give a model of how this can be put into practice;
- practice – so that the teacher can try out the new approach;
- feedback on how well the new approach is working; and

- coaching – to help the teacher discuss the teaching in a supportive environment and consider how it might be improved.

Their research showed that, without the opportunity to receive feedback and coaching, there is no measurable impact on classroom practice. However, once these two components are added, particularly the final coaching stage, there is a large and measurable impact on practice. Attention to evaluating impact is given in the next section.

The Impact of Teachers' CPD

Evaluating the impact of teachers' professional development and training is a neglected area partly because it is perceived to be difficult. Few evaluation studies concerning CPD make any reference to its impact on teacher behavior or pupil learning outcomes. However, evaluation is necessary to provide a sound basis for improving and upgrading teacher development programs but it needs to be relatively easy and inexpensive, otherwise it may be seen as diverting scarce resources away from other more important activities. Evaluation was found to be effective when the professional development activity had a clear, predefined outcome and a suitable method for collecting evidence of its impact.

Over the years, a number of models to evaluate the impact of training and development have been devised. One of the first and perhaps the best known is the framework for evaluating training developed by Kirkpatrick (1959). This has been adapted and developed over the years but Kirkpatrick's four-step framework remains the model or framework for evaluating most training programs in business and commerce. The model explores the relationship between training and the workplace at four levels: reactions, learning, behavior, and results. These are measured at suitable points during the training process. The first three are essentially trainee based, while the fourth changes emphasis and centers on the effectiveness of the training for the organization.

Another American, Thomas Guskey, has developed and refined Kirkpatrick's model more specifically for education. He is clear that: "We need to make evaluation an integral part of the professional development process ... Systematically gathering and analysing evidence to inform our actions must become a central component in professional development technology" (Guskey, 2000: 92).

Guskey talks in terms of five levels of evaluation of CPD with improved pupil outcomes being the desired result. These five levels are:

1. participants' reactions;
2. participants' learning;
3. organization support and change;

4. participants' use of new knowledge and skills; and
5. pupil learning outcomes.

Level 3 – organization support and change – is not found in the Kirkpatrick model and refers to the key role that the school can play in supporting or sabotaging any CPD efforts. The focus needs to be on the attributes and organizational features of the school that are necessary for success. Were changes at the individual level encouraged and supported at all levels? Were sufficient resources made available, including time for sharing and reflection? Were successes recognized and shared? (Guskey, 2002: 47).

Each level builds on those that come before but reversing these five levels can be useful in planning for teachers' professional development. Therefore, planning might begin by asking, "What improvements in pupils do we want and how will we know when they're achieved?" Then ask, "If that's the impact we want, what needs to change?" Next, consider what types of organizational support or change are needed, and so forth (Guskey, 2005).

Guskey notes that, in some cases, information on pupil learning outcomes is used to estimate the cost effectiveness of CPD, sometimes known as return on investment (Guskey, 2002: 49).

Conclusion

If the ultimate purpose of CPD for teachers is to bring about improvement in the quality of teaching and learning in schools and classrooms – and thus impact positively on pupil outcomes – then it is important to have evaluation processes in place which give an idea whether this is happening or not. Evaluation should be considered at the outset and not at the end of a professional development activity.

Bolam and McMahon (2004) discuss the developments in CPD over the last 20 years and remark how it is striking just how much has changed and yet other things remain the same. For example, technological developments such as e-learning, Teachers TV, networks (both virtual and real), and the Internet have made possible new CPD methods and approaches. They comment:

The increased emphasis on school-focused CPD which was evident in the late 1980s has developed into a much more sophisticated approach which takes account of school culture, work-based learning and professional learning communities (p. 51).

Yet, other things remain largely unaltered, such as the potential tension between the needs of teachers and those of the organization or national policy, and "the evidence from research and experience in several countries demonstrates that national needs have been dominant since the mid 1980s" (Bolam and McMahon, 2004: 51). It is clear

that "governments and systems are using CPD as a means of achieving broader education reform agendas" (Bolam and McMahon, 2004: 27).

CPD is about ongoing or lifelong learning which will help teachers and other staff respond to ever-changing situations and exercise judgment in informed and creative ways but, as Pachler and Field state, it should also be seen as a means "to rejuvenate practice, to expand our professional repertoire, increase our self-esteem, self-confidence and enthusiasm for teaching or, for example, our level of criticality and, thereby, achieve enhanced job satisfaction" (Pachler and Field, 2004: 2). As Bolam and McMahon note, "the image of the teacher as an autonomous professional is being replaced by the concept of the school as a learning community in which the teacher works collectively as a member of a wider group of staff in a joint effort to improve the quality of learning" (Bolam and McMahon 2004: 55). Finally, Bolam and McMahon remark that the longer-term health of schools and students depends on well-educated teachers, not just well-trained and well-informed ones, asking crucially: What part should democratic values play in the CPD of teachers? and what kind of teachers are needed for the twenty-first century?

Bibliography

- Bolam, R. (1993). Recent development and emerging issues. In Williams, M. and Bolam, R. (eds.) *The Continuing Professional Development of Teachers*, pp 33–63. London: GTC.
- Bolam, R. and McMahon, A. (2004). Recent developments in CPD. In Day, C. and Sachs, J. (eds.) *International Handbook of the Continuing Professional Development of Teachers*, pp 23–46. Maidenhead: Open University Press.
- Bolam, R. and Weindling, D. (2006). *Synthesis of Research and Evaluation Projects Concerned with Capacity-Building through Teachers' Professional Development*. London: GTC.
- Boyle, B., White, D., and Boyle, T. (2003). A longitudinal study of teacher change: What makes professional development effective? *Paper Given to American Education Research Association*, Chicago, April.
- Bubb, S. and Earley, P. (2007). *Leading and Managing Continuing Professional Development*, 2nd edn. London: Paul Chapman/Sage.
- Day, C. (1999). *Developing Teachers: The Challenge of Lifelong Learning*. London: Falmer.
- Day, C. and Sachs, J. (eds.) (2004). Professionalism, performativity and empowerment: Discourses in the politics, policies and purposes of continuing professional development. In *International Handbook of the Continuing Professional Development of Teachers*, pp 3–32. Maidenhead: Open University Press.
- Department of Education and Science (DES) (1972). *Teacher Education and Training (James Report)*. London: HMSO.
- General Teaching Council (GTC) (2003). *Teachers' Professional Development Learning Framework*. Birmingham: GTC.
- Guskey, T. (2000). *Evaluating Professional Development*. New York: Corwin.
- Guskey, T. (2002). Does it make a difference? Evaluating professional development. *Educational Leadership* 59(6), 45–51.
- Guskey, T. (2005). A Conversation with Thomas R. Guskey. *Evaluation Exchange* XI(4), 12–15.
- Jones, T. (2003). Continuing professional development in Wales: An entitlement for all. *Professional Development Today* 6(1), 35–42.

- Joyce, B. R. and Showers, B. (2002). *Student Achievement through Staff Development*, 3rd edn. Alexandria, VA: Association for Supervision and Curriculum Development.
- Kirkpatrick, D. (1959). The Kirkpatrick model for training evaluation. *Journal of the American Society of Training Directors* **13**, 11–12.
- Madden, C. and Mitchell, V. (1993). *Professions, Standards and Competence: A Survey of Continuing Education for the Professions*. Bristol: Department for Continuing Education, University of Bristol.
- Ofsted (Office for Standards in Education) (2006). *The Logical Chain: Continuing Professional Development in Effective Schools*. London: Ofsted.
- Pachler, N. and Field, K. (2004). Continuing professional development. In Capel, S., Leask, M., and Turner, T. (eds.) *Starting to Teach in the Secondary School*, pp 233–246. London: Routledge/Falmer.
- TDA (Training and Development Agency for Schools) (2007). *What is Effective CPD?* London: TDA.
- Waters, M. (1998). Personal development for teachers. *Professional Development Today* **1**(2), 29–38.

Further Reading

- Bolam, R. and Weindling, D. (2006). *Synthesis of Research and Evaluation Projects Concerned with Capacity-Building through Teachers' Professional Development*. London: GTC.
- Bubb, S. and Earley, P. (2007). *Leading and Managing Continuing Professional Development*, 2nd edn. London: Paul Chapman/Sage.
- Day, C. (1999). *Developing Teachers: The Challenge of Lifelong Learning*. London: Falmer.
- Day, C. and Sachs, J. (eds.) (2004). *International Handbook of the Continuing Professional Development of Teachers*. Maidenhead: Open University Press.
- Guskey, T. (2000). *Evaluating Professional Development*. New York: Corwin.
- Joyce, B. R. and Showers, B. (2002). *Student Achievement through Staff Development*, 3rd edn. Alexandria, VA: Association for Supervision and Curriculum Development.

Homeschooling

P Farenga, Medford, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

An Introduction to Homeschooling

Books about learning at home were categorized under domestic education by the United States Library of Congress until recently. This category holds books and resources that colonists, pioneers, and city dwellers used to help their children learn to read, write, and calculate at home, on the farm, and in their local communities, as well as books about private tutors, governesses, and related topics. In the late 1980s, due to the rapid growth of books and materials that help parents teach their own children at home, a new category was created in the Library of Congress, Home Schooling.

Why Do People Choose to Homeschool?

Though homeschooling represents a small percentage of the total school-age population, it serves as a growing counter-trend to educators and others who insist that children need to spend more time in school, more time studying for tests, and less time pursuing subjects such as music, art, religion, philosophy, foreign languages, and literature, let alone extracurricular activities. However, it is not correct to conclude that the growth of homeschooling indicates a rejection of the public school system because private schools are losing enrollments to homeschooling as well. The growth of homeschooling suggests that more factors are at play. Personalized learning, closer family ties, inviting children to learn among adults doing real work, and an emphasis on the delight of learning are common factors among homeschoolers regardless of the educational philosophies they hold, and these are factors homeschoolers do not find in their local schools.

In 1981, the author and teacher John Holt cited these reasons as to why parents choose homeschooling: they feel that raising their children is their business and not the government's; they do not want their children to be hurt morally, spiritually, or physically; and they enjoy being with their children and do not want to give that pleasure up to others. While these continue to hold true today, reasons cited more recently by homeschoolers include a desire for alternatives to conventional learning for their children (see **Table 1** for more information).

Policy Context and Issues

Many philosophical and social themes are entwined whenever educational matters are discussed and homeschooling

provides a unique lens for exploring institutional versus family concerns about school. Indeed, these issues have long been embedded in discussions about compulsory school laws, but they were largely not of much public concern until the resurgence of homeschooling in the late twentieth century. Since homeschooling itself was never illegal in any state in America, if schools took homeschoolers to court it was typically on educational neglect or truancy charges, not because they were homeschooling. Children are compelled to attend school by law, but school can take different forms, though few families opted to call their home a school until recently. The right of parents to select a child's education is cited in both British and American law, and US Constitutional Amendments one, nine, and fourteen support individual rights and the right to privacy for a family to choose alternative forms of education (Arons, 1983; Whitehead and Bird, 1986).

Homeschooling was the primary way people were educated until the mid-nineteenth century when public schools were established and universal compulsory attendance laws were passed. The tensions caused by claims from professional educators and the state for dominion over students' lives and learning against similar claims by parents and child advocates have increased substantially with the emergence of homeschooling as a way to satisfy compulsory attendance laws (Arons, 1983). Homeschooling can be seen not only as an educational movement toward individualized learning, but also as a political statement against the delivery of mandatory public services into one's private family life.

The growth of homeschooling was slow but steady throughout the late 1970s and 1980s, but in the 1990s it grew very quickly. By 1999, the US Department of Education claimed that 850 000 children were being taught at home. In 2007, the number of homeschooled children was 1.5 million, representing 2.9% of the total United States school-age population (National Center for Education Statistics, 2008) (**Table 2**).

Approaches to Homeschooling

Homeschooling can be broadly divided into two streams of educational philosophy, the nontraditional and the traditional. The nontraditional approach emphasizes a learner's autonomy and encourages questioning authority. For instance, a mother writing to New York City school officials in 1923 claims that she will teach her son at home

Table 1 Reasons for homeschooling^a

<i>Reasons for homeschooling</i>	<i>Percentage of homeschooled students whose parents reported particular reasons for homeschooling: 2003 and 2007</i>	
	<i>2003 Percent</i>	<i>2007 Percent</i>
Concern about the school environment ^b	85	88
Dissatisfaction with academic instruction at other schools	68	73
A desire to provide religious or moral instruction	72	83
Nontraditional approach to child's education	n/a	65
Child has a physical or mental health problem	16	11
Child has other special needs	29	21
Other reasons ^c	20	32

^aExcludes students who were enrolled in school for more than 25 h a week and students who were homeschooled only because of a temporary illness.

^bThese include safety, drugs, or negative peer pressure.

^cParents homeschool their children for many reasons that are often unique to their family situation. Other reasons parents gave for home schooling include: it was the child's choice; to allow parents more control over what child was learning; and flexibility.

Percentages do not sum to 100 percent because respondents could choose more than one reason.

Adapted from US Department of Education, National Center for Education Statistics, Parent and Family Involvement in Education Survey of the 2003 and 2007 National Household Education Surveys Program (NHES).

so he can truly learn how democracy and self-determination work. Some of the chapter titles in the book she wrote about her homeschooling experiences summarize the non-traditional approach well; for instance: "In which it is plainly to be seen that democracy cannot be learned in a place where it is not lived" and "Wherein is shown how by guiding a child's natural interests, he enters into that heritage whereby he 'begins to learn everything the moment he begins to learn anything'" (Anonymous, 1923). These sentences illustrate an educational approach that eschews standardized curricula and seeks to fit schooling to the child rather than the other way around.

In contrast, the traditional approach emphasizes structured curricula and obedience to authority. For instance, the British educator Charlotte Mason advocated teaching one's own children in the late nineteenth century. Though her approach is rooted in the medieval university and its Trivium curriculum she is no taskmaster; her attitude toward children emphasizes understanding and gentleness. However, both approaches to homeschooling declined by the mid-twentieth century as the vast majority of children were taught in traditional schools while their parents worked or ran the household. In the late 1970s, this situation began to reverse itself as families sought other ways of helping their children learn besides sending them to school all day.

John Holt, whose books were influential in the school-reform movement of the 1960s, became dissatisfied with the progressive school-reform movement's lack of success and, throughout the 1970s, he sought new ways to help children learn and grow. Already influenced by the ideas of social and education critic Paul Goodman, who advocated student-initiated learning, Holt became particularly

Table 2 Number of homeschoolers

<i>Total number of homeschooled students, ages 5 through 17 with a grade equivalent of kindergarten through 12th grades: 1999, 2003, and 2007</i>			
<i>Homeschooled students</i>			
Year	1999	2003	2007
Total	850 000	1 096 000	1 500 000

Adapted from US Department of Education, National Center for Education Statistics, Parent and Family Involvement in Education Survey of the 2003 and 2007 National Household Education Surveys Program (NHES).

impressed with the ideas of the philosopher and activist Ivan Illich, who wrote that we did not necessarily need schools to educate people, and we should instead invest in other social institutions and projects that improve people's lives and help them learn to do things better (Illich, 1971). Holt's support for children's innate abilities to learn and grow through informal education meshed well with Illich's ideas about our need to reform social institutions into smaller, local, more responsive entities. When Holt learned that some parents were simply refusing to send their children to school and were teaching them at home, he founded the first journal about homeschooling, *Growing without Schooling*, in 1977 as a way to connect and support people who were already homeschooling or who wanted to do so. Holt identified homeschooling with the burgeoning ecology movement, back to the land communarians, do-it-yourself enthusiasts, advocates for religious and individual rights, and those who held a strong skepticism

of institutional authority. Holt did not like the terms deschooling or homeschooling to describe the learning children did without going to school, so he invented the word unschooling to better describe learning that does not have to occur at home and that does not resemble school. However, unschooling, as a word, did not catch on with the public and Holt himself used the word homeschooling often to describe learning at home. In 1980, homeschoolers who were uncomfortable with Holt's secular outlook and advocacy for children formed their own publication, *The Teaching Home*, and organized tight-knit groups to foster homeschooling that identified itself closely with religious instruction.

Dr. Raymond Moore, a professional educator and researcher, became a major figure in Christian homeschooling when he created his own program for homeschooling based on delaying academics until children were 8–10 years old. Moore and Holt corresponded and worked together on homeschooling issues, exemplifying how liberals and conservatives can join to fight attacks on homeschooling. Holt died in 1985 and Moore gradually became less influential as other Christian homeschooling advocates became popular and the homeschooling movement became more polarized along secular and religious lines. By the early 1990s, many Christian homeschooling groups required signed statements of faith for membership and these groups became an overtly conservative political force that promoted only certain methods of homeschooling and supported political issues that went well beyond homeschooling. Influential conservative Christian evangelists, such as James Dobson and Pat Robertson, became vocal supporters of homeschooling to their audiences. Michael Farris, a founder of The Home School Legal Defense Association, engaged the political situation head on by creating Patrick Henry College in 2000. Patrick Henry College is both a college specifically for homeschoolers and in general for any Christian who believes in the inerrant Bible and who wants to transform America through political service.

Secular and religious homeschoolers who were uncomfortable with the hard right political positions put forth by major Christian homeschooling groups tried to organize counterbalancing groups to promote a more inclusive vision of homeschooling to the public, but they never achieved the political momentum that right-wing homeschoolers did (Stevens, 2001).

During the 1990s, the number of homeschooling families increased to the point that many schools, institutions, and businesses recognized it as a niche market and they began to market a multitude of learning opportunities for home learners. Homeschoolers who desire to follow a specific learning philosophy for their children, such as Montessori or Waldorf/Steiner, find they can adapt these concepts or purchase materials to make them work in homeschooling settings. Homeschooling

families can join groups or schools that provide support for many different approaches to learning at home, including support for homeschooling gifted or special-needs children.

Homeschooling refers to all forms of teaching and learning that are used to help children learn, but researchers and homeschoolers themselves often create categories to identify different models of learning. So, in between the approaches noted earlier, there are eclectic homeschoolers who use anything they think will work; classical homeschoolers who incorporate the study of Latin and emphasize the use of classic books; pagan, Catholic, Mormon, and many other groups who see homeschooling as a way to not just teach children reading, writing, and arithmetic, but also as a way to pass on values that they feel are, at best, not addressed or, at worst, belittled in conventional schooling.

Nontraditional approaches to homeschooling, such as unschooling, emphasize the child as one who learns all the time and these families work with each child's interests and abilities on a much longer and broader schedule than conventional schooling permits. Unschoolers do not use a predetermined curriculum; instead they use learning opportunities that are embedded in the activities the children engage in. Unschoolers often use real-life experiences, such as traveling and volunteering, acting in shows or building projects, and unrestricted reading and questioning as the primary means of learning in lieu of seat-time in class. If school transcripts are required of them, unschoolers can supply portfolio evaluations, personal references, work samples, and dynamic assessments to establish completion of work and competency, or they can engage outside evaluators to assess their children's work.

Traditional approaches emphasize completing the same work as their age mates in school do, often using the same curricula, textbooks, and grading systems as schools. Many of the major Christian Independent School curriculum suppliers began selling their conventional school materials to homeschoolers in the early 1980s and eventually created products and curricula just for home use, well before the larger, more mainstream curriculum providers did. The lines between these two approaches, and the many positions in between, are blurred though. For instance, it is not unusual for an unschooler to decide that he/she wants to take a class in order to learn something in more depth or for a traditional homeschooler to decide to take a break from a subject or drop a course to take advantage of a new opportunity. The most common element among all homeschooling approaches to teaching and learning is flexibility of scope and sequence. This shows that the use of packaged all-in-one curricula among homeschooling families is declining and a strong majority of parents are choosing to tailor curricula for their children by handpicking major components (Ray, 1997).

Government schools are gradually working with homeschoolers, but private enterprises and new forms of private schools, such as online distance-learning schools, are actively creating alternatives that are attractive to families. For example, families who want their children to be fluent with current-day media and technology are finding many opportunities to do so at home, while many schools are not providing their students with such access.

Today, the Internet contains numerous cyber schools and other distance-learning programs that are marketed to homeschoolers; some public schools allow homeschoolers to participate in their classes and activities; numerous independent and alternative schools have programs designed to award homeschoolers their diplomas for work done at home; and a few public schools have created cooperative education programs that provide school textbooks and resources to homeschoolers who agree to follow the school's curriculum at home (Lines, 2003). Learning centers that cater to all sorts of niches, from how to write science fiction to learning world history using film, have found many members among homeschoolers, as have libraries and museums.

These options are not necessarily an abundance of riches for homeschoolers since they blur the distinction between one's private home and public school when companies seek public-education funds for enrolling homeschooled students in their programs. For instance, in exchange for tax-funded school programs at home, families typically agree to adhere to the school's curricula and grading system. This is a trade-off some families are willing to make, but most homeschooling families seek something different than school at home and wish to avoid the strings that come attached to government funding. These new programs are creating another type of schooling experience, one that is not totally independent from public schooling as homeschooling can be.

Teaching and Learning in Homeschooling

Most parents who homeschool do not possess teaching credentials and the vast majority of them do not have college degrees, but that does not prevent them from helping their children learn. Rather than view themselves as teachers of all subjects, homeschool parents often cast themselves in the role of facilitator, connecting their children to classes and teachers as needed. These parents learn to be better teachers of their children by doing it every day; they are free to discard texts and classes that are not working and to try something different quickly. One study was able to compare the test scores between an extreme of teacher contact – certified teachers who homeschooled their children – to nonteacher parents and found “an insignificant relationship between the test scores of

homeschooled students and whether the parent-teachers were state-certified teachers” (Ray and Wartes, 1991).

Many homeschoolers begin by purchasing a curriculum, following its instructions and, if it works for them, they continue using it. Most homeschooling parents are confident teaching the basic school subjects to young children, especially for the early elementary years, but if they get in over their heads, they are able to find help. There are curriculum fairs, private consultants, support networks and local groups, private tutors, and other resources for homeschooling parents to fall back on.

Teaching Children of Different Ages at the Same Time

Most homeschooling families are larger than the average family, so very often they have one or more school-age children being taught at home and a preschooler to care for at the same time. How does a homeschooling parent handle such different needs at once? By maximizing his or her use of available resources and differentiating the instruction for each learner.

Addressing Each Child's Unique Way of Learning

Without reference to education theory, many homeschooling parents nonetheless find ways to help their children learn due to their intimate knowledge of their children's personalities, whereas educators must use a more formal approach to discover how a particular child learns best. For instance, the theory of multiple intelligences identifies eight intelligences that we use in order to learn:

- linguistic intelligence;
- logical–mathematical intelligence;
- spatial intelligence;
- bodily–kinesthetic intelligence;
- musical intelligence;
- interpersonal intelligence;
- intrapersonal intelligence; and
- naturalist intelligence.

Here is how a homeschooling parent might put these ideas into practice whether or not she is aware of multiple-intelligence theory. Mom knows that her 9-year-old son loves to touch and build things, so she uses building blocks to explore and explain arithmetic with her son. Math manipulatives can be useful tools for bodily–kinesthetic learners like her son. After she explains the concept of fractions using the blocks, she asks her son to make a building then break the building into halves, fourths, and so on. While he builds and divides, she can work with her other children, responding to her son as needed. Establishing different centers of activity and then moving among them was a common technique used in one-room

schoolhouses that is also used today by homeschoolers with large families.

Since her teenager likes to learn by reading, analyzing, and discussing what she read, her mom emphasizes linguistic and logical–mathematical learning for her. She will read the same materials as her daughter and discuss them with her. But what if her teenager wants to learn chemistry and mom knows very little about it? Here is where mom shifts roles from being the sage on the stage to the guide on the side. Mom can help her daughter find chemistry courses online or at the local community college. She can post requests for a chemistry tutor online, at local stores, and at local schools. She can also evaluate various chemistry books and kits with her daughter, choose one and learn chemistry together, or have her daughter go through the text or kit on her own, asking for help as necessary.

Encourage Independent Learning

Making children figure things out for themselves does not have to be treated as the last resort of an exasperated parent. Indeed, parents usually take great pride in watching infants and toddlers independently learn to speak, walk, think, and socialize with little formal instruction from adults and homeschoolers can continue to nurture that ability. Homeschoolers can allow plenty of quiet time for their children to reflect or daydream since they are not changing classes every hour. Reading a book or a website, watching a video, listening to books on tape, drawing, writing in journals or creating a story, making a video or play, practicing an instrument or sport, working on a model or other project, are common ways homeschoolers encourage independent learning throughout their day. By talking with their children about what they are doing these parents help their children make connections to other topics or people that can expand their learning.

Engage the Children to Help Each Other

The homeschooling mom mentioned above might try to get her 4-year-old to play with the building blocks alongside her son, or perhaps have her son read to the young child while mom works with the teen. She can also ask the teen to work on arithmetic with her son while mom plays with the 4-year-old. Inviting other kids to their house to play or study is another way homeschoolers foster independent learning and make time for parents to have time with their other children.

Research about Homeschooling

There is not enough unbiased quantitative data about homeschooling to justify making direct comparisons about the academic performance of homeschooled

children to school children, but what research does exist so far indicates there are no significant differences in achievement among them (Thomas and Pattison, 2007; Ray and Wartes, 1997). However, the lack of research or concern among parents for seat-time in school for their children has not stunted the growth of homeschooling. Further, data consistently show that homeschooling is not a harmful or irrevocable decision. Homeschoolers matriculate back to school with no significant difficulties when they wish to attend. Indeed, homeschooling can be a phase of schooling in most children's lives rather than a constant: just 15% of secular homeschoolers and 48% of religious homeschoolers continue to homeschool after 6 years (Isenberg, 2007).

Many parents considering homeschooling find the conviction to do it not from academic research but from their own experiences and those of others they know or read about. History shows that children are able to learn social skills from their parents and culture without being forced by law to attend school; alternative schools successfully help children learn in different ways than conventional schools allow, and the wealthy have always used private tutors, special classes, and travel to provide elite educations for their children; so homeschooling has significant antecedents.

However, since school has become the primary place for children to spend their time in recent decades, many people think that socialization skills are best developed in schools, so a primary objection to modern homeschooling has been that homeschooled children will lack socialization skills. Many homeschoolers note the antisocial structure of school and the bad behavior of many schoolchildren as a primary reason for removing their children from school or for not sending them. Studies in the United States consistently show the major reason for homeschooling is a concern about the school environment (Table 1).

Homeschool advocates note how homeschooled children are socialized not in classes where everyone is the same age but by being around adults and children of mixed ages within a variety of groups and activities. There are studies that demonstrate homeschooled children are socially well adjusted, as well as studies conducted on adults who were homeschooled. The latter indicate homeschooled adults are engaged in their communities, enjoy life, and are gainfully employed. However, all such studies are to date limited by the self-selection and self-reporting of homeschoolers who participate and the difficulty of comparing standardized school system results to individualized family results.

Colleges and universities now encourage homeschoolers to apply, with some even sending recruiters to homeschooling conferences. A majority of American colleges and universities have admission policies for homeschooled applicants (Jones and Gloeckner, 2004).

Summary

We have seen how homeschooling has been successfully used in lieu of conventional schooling well before compulsory laws were enacted in the mid-nineteenth century. Toward the end of the twentieth century, homeschooling had a resurgence from the efforts of liberal and conservative parents and educators that continues to gain adherents in the twenty-first century. Businesses and private schools have supported the growth of homeschooling by identifying and selling to this market, while families who are less focused on consuming education products support the growth of homeschooling by encouraging self-reliance and creating opportunities within one's local community. Though there is considerable research about homeschooling, there is no true random sample of homeschoolers that one can definitively compare to a random sample of public schoolers; so claims about the academic achievement of homeschoolers are hard to generalize. However, there is no evidence to support claims that homeschoolers fare worse than those who go to school, particularly with regard to employment and college admissions. The myriad methods and philosophies that support homeschooling also drive considerable debate about the role of compulsory education in our lives, sometimes resulting in political action about the rights of families, parents, and children.

As more children move in and out of school and homeschool, new conceptions of where, how, and from whom learning takes places can be studied and explored. In fact, many homeschooling families also have children who attend public or private school full or part-time, so it would be wise for schools to not alienate homeschoolers and instead form fruitful partnerships with them. New methods of evaluation, new scopes and sequences for learning, and new ideas about standard school assumptions and practices can emerge from such partnerships (Thomas and Pattison, 2007). The growth of homeschooling proves that parents, regardless of their educational background, can be trusted to help children learn and that whether or not schools decide to support their endeavors, homeschooling will continue. Homeschooling parents are deeply involved in their children's educations and though schools openly claim they want parental involvement, they have yet to figure out how they can work positively with this vivid example.

Homeschooling is not a solution to school problems since it is so different from school: the skills one needs to manage a class of children from diverse backgrounds who all need to be taught the same thing at the same time are totally different from the skills a parent needs to help their own children learn in the comfort of their home. Nonetheless, homeschooling can inform the future direction of schooling not because it is better than school, but, as John Holt wrote,

because it is not a school at all. It is not an artificial place, set up to make "learning" happen and in which nothing except "learning" ever happens. It is a natural, organic, central, fundamental institution; one might easily and rightly say the foundation of all other institutions. . . What I am trying to say, in short, is that our chief educational problem is not to find a way to make homes more like schools. If anything, it is to make schools less like schools. (Holt and Farenga, 2003)

See also: Informal Education and Evaluation; Learning Styles in Vocational Education and Training; Personalized Learning and Vocational Education and Training.

Bibliography

- Anonymous (1923). *A Mother's Letters to a Schoolmaster*. New York: Alfred Knopf.
- Arons, S. (1983). *Compelling Belief: The Culture of American Schooling*. Amherst, MA: University of Massachusetts Press.
- Holt, J. and Farenga, P. (2003). *Teach Your Own: The John Holt Book of Homeschooling*. Cambridge, MA: Perseus Books.
- Illich, I. (1971). *Deschooling Society*. New York: Harper and Row.
- Isenberg, E. J. (2007). What have we learned about homeschooling? *Peabody Journal of Education* **82**, 387–409.
- Jones, P. and Gloeckner, G. (2004). *Special Homeschool Issue: A Study of Admission Officers' Perceptions of and Attitudes Towards Homeschool Students*. *Journal of College Admission* **185**, 12–21.
- Lines, P. (2003). *Support for Home-Based Education: Pioneering Partnerships between Public Schools and Families Who Instruct Their Children at Home*. Eugene, OR: Educational Resources Information Center, University of Oregon.
- Ray, B. (1997). *Strengths of Their Own: Homeschoolers across America*. Salem, OR: National Home Education Research Institute.
- Ray, B. and Wartes, J. (1991). The academic achievement and affective development of home-schooled children. In Van Galen, J. and Pitman, M. (eds.) *Home Schooling: Political, Historical, and Pedagogical Perspectives*, p 59. Norwood, NJ: Ablex.
- Stevens, M. (2001). *Kingdom of Children: Culture and Controversy in the Homeschooling Movement*. Princeton, NJ: Princeton University Press.
- Thomas, A. and Pattison, H. (2007). *How Children Learn at Home*. London: UK: Continuum International Publishing Group.
- Whitehead, J. W. and Bird, W. R. (1986). *Home Education and Constitutional Liberties*. Westchester, IL: Crossways.

Further Reading

- Armstrong, T. (2000). *In Their Own Way: Discovering and Encouraging Your Child's Multiple Intelligences*. New York: Tarcher.
- Bielick, S. (2008). 1.5 million homeschooled students in the United States in 2007. *US Dept. of Education, National Center for Education Statistics, Issue Brief NCES 2009030*.
- Gathercole, R. (2007). *The Well-Adjusted Child: The Social Benefits of Homeschooling*. Denver, CO: Mapletree.
- Gaither, M. (2008). Homeschooling goes mainstream. *Education Next* Winter 2009, 11–18.
- Herr, M. (ed.) (2008). *Everywhere All the Time: A New Deschooling Reader*. Oakland, CA: AK Press.
- Knowles, J. G. and Muchmore, J. A. (1995). Yep! We're grown-up home-school kids – and we're doing just fine, thank you. *Journal of Research on Christian Education* **4**(1), 35–56.
- Llewellyn, G. (2005). *Real Lives: Eleven Teenagers Who Don't Go to School Tell Their Stories*. Eugene, OR: Lowry House.
- Macaulay, S. S. (1984). *For the Children's Sake: Foundations of Education for Home and School*. Wheaton, IL: Crossway Books.

- Mayberry, M., Knowles, J. G., Ray, B., and Marlow, S. (1995). *Home Schooling: Parents as Educators*. Thousand Oaks, CA: Corwin Press.
- Meighan, R. (2007). *John Holt*. London, UK: Continuum.
- Miller, R. (2002). *Free Schools, Free People: Education and Democracy after the 1960s*. Albany, NY: State University of New York Press.
- Moore, R. S. and Moore, D. N. (1994). *The Successful Homeschool Family Handbook*. Nashville, TN: Thomas Nelson.
- Ray, B. (2003). *Home Educated and Now Adults: Their Community and Civic Involvement, Views about Homeschooling, and Other Traits*. Salem, OR: National Home Education Research Institute.
- Shyers, L. E. (1992). A comparison of social adjustment between home and traditionally schooled students. *Home School Researcher* 8(3), 1–8.
- Webb, J. (1990). *Children Learning at Home*. London, UK: Falmer Press.

Relevant Websites

- <http://homeschooling.gomilpitas.com> – A to Z Home's Cool Homeschool.
- <http://www.holtgws.com> – HoltGWS.com.
- <http://www.home-school.com> – Homeschool World.
- <http://www.homeedmag.com> – Homeschooling.
- <http://gaither.wordpress.com> – Homeschooling Research Notes.
- <http://www.indiana.edu> – Homeschooling Research & Scholarship.
- <http://www.nheri.org> – National Home Education Research Institute.
- <http://www.patfarenga.com> – Pat Farenga.com.
- <http://www.learninginourownway.com> – Pat Farenga Presents The Learning In Our Own Way Conference.
- <http://www.infed.org> – The Encyclopedia of Informal Education.

Innovations and Early Intervention in the Teaching of Literacy

R E Slavin, Johns Hopkins University, Baltimore, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

From the time they enter kindergarten through their first years of elementary school, children substantially define themselves as learners. Those who end third grade reading well are not guaranteed success in school and in life, but they have cleared a major hurdle. Those who do not succeed during this critical period, however, are likely to have serious problems throughout their subsequent school careers. For example, Juel (1988) found that almost all 7-year-olds who had reading difficulties also had reading difficulties as 10-year-olds. Lloyd (1978) reported that high school drop-out could be predicted to a substantial degree based on the learning levels of 9-year-olds, but could not be accurately predicted based on characteristics of 6-year-olds, supporting the idea that early school learning success (or failure) is a key factor in long-term outcomes of schooling.

In recent years, research has found that the outcomes of early schooling can be substantially affected by the programs and practices adopted by elementary schools. Recognizing the importance of personal and home factors and of preschool experiences, there is no doubt that elementary schools can make a profound difference in children's learning trajectories during the crucial early years.

Beginning-Reading Programs

The most important criterion for success in the early elementary years is success in reading. Reading is the basis for success in all other skills. Poor progress in reading is the main indicator of need for special education, for example, and for retention in grade.

In recent years, a consensus has emerged among most researchers about the importance of systematic, synthetic phonics in the teaching of beginning reading. That is, children have been found to learn to read best if they are taught to apply consistent strategies for blending letter sounds into words, rather than trying to learn whole words. Definitive reviews in the US by Snow *et al.* (1998) and the National Reading Panel (2000) emphasized the strong evidence base for synthetic phonics, and the Rose Report in the UK (Rose, 2006) came to the same conclusion. However, it is not enough for teachers to be given textbooks that emphasize phonics, or to have policies promoting phonics. Research finds that particular approaches to synthetic phonics and particular professional-development approaches are more effective than others.

Slavin *et al.* (in press) recently carried out a review of research on the learning outcomes of core

beginning-reading programs. In order to be included in the review, studies had to meet the following standards:

1. They evaluated core reading programs that began in kindergarten or first grade (remedial programs are discussed in a separate section).
2. They compared children who used the program to those in matched or randomly assigned control groups.
3. The study took place over at least 12 weeks, but usually a year or more.
4. On pretests, the experimental and control groups were no more than a half standard deviation apart, and were well matched on demographic variables.

Studies of programs that posttested at the end of kindergarten were reviewed separately, as such studies often find positive effects just because programs teach phonics or reading skills that the control group has not yet been taught. However, by first grade, all children are being taught to read; so experimental-control comparisons are meaningful. Programs that began in kindergarten and reported end of first grade (or later) outcomes were included in the main review.

An exhaustive search of published and unpublished articles written since 1970 produced a total of 63 studies that met the inclusion standards and posttested in first grade or later. The qualifying studies were done in the US, Canada, Norway, Denmark, and Germany. They involved more than 22 000 children. The reading programs were divided into four categories: reading curricula, technology, instructional-process approaches, and combined curricula and instructional-process approaches.

Beginning Reading: Curricula

Beginning-reading curricula have been evaluated in seven studies, five of which used randomized quasi-experiments. These studies evaluated four core basal-reading programs, Open Court Reading, Reading Street, and Scholastic Phonics Readers with Literacy Place, plus three supplemental programs, the Open Court Phonics Kit, Phonics in Context, and Elements of Reading: Phonics and Phonemic Awareness.

With the exception of a small study of the Open Court Phonics Kit, none of the programs had effect sizes in excess of +0.20. The sample size-weighted mean effect size across all seven was +0.12, with the four studies of core basal programs reporting a weighted mean effect size of +0.11 and the three studies of supplementary programs

with a weighted mean of +0.12. Effect sizes averaged +0.23 for decoding measures, but only +0.09 for comprehension/total reading measures.

Beginning Reading: Technology

Thirteen studies of instructional technology for beginning reading met the inclusion standards. These were divided into three categories: (1) supplemental technology programs, such as Waterford, Wireless Internet Center for Advanced Technology (WICAT), and Phonics-Based Reading, which are programs that provide additional instruction at students' assessed levels of need to supplement traditional classroom instruction; (2) mixed-method models, represented by Writing to Read, are methods that use computer-assisted instruction (CAI) along with non-computer activities as students' core reading approach; and (3) embedded multimedia, represented by Reading Reels, provides video content embedded in teachers' whole-class lessons.

The weighted mean effect size for all technology approaches in beginning reading was only +0.09. A large, randomized study by Dynarski *et al.* (2007) and Campuzano *et al.* (2009) found no impact of five current supplemental CAI models. This study's findings greatly affected the weighted mean of nine studies of supplementary CAI, estimated at +0.08. The weighted mean effect size for decoding measures, also substantially affected by the Dynarski/Campuzano findings, was only +0.05, although comprehension/total reading effects (not measured in the Dynarski/Campuzano study) averaged +0.20. Large effect sizes were reported in small, matched studies of Waterford and WICAT. Reading Reels, which uses multimedia embedded in teachers' class lessons, had modest positive effects in two large randomized experiments (weighted mean ES = +0.20). With these potentially promising exceptions, research on the use of technology in beginning reading instruction does not show positive achievement effects of the types of software that have been most commonly used.

Beginning Reading: Instructional-Process Approaches

Instructional-process programs are methods that focus on providing teachers with extensive professional development to implement specific instructional methods. These fall into three categories. Cooperative learning programs (Slavin, 1995, 2009) use methods in which students work in small groups to help one another master academic content. Phonological-awareness training is an approach that gives teachers specific classroom strategies for building phonics and phonemic-awareness skills. Phonics-focused professional development models, including Reading and Integrated Literacy Strategies (RAILS),

Sing, Spell, Read, and Write, Ladders to Literacy, Early Reading Research, and Orton Gillingham, provide training to teachers to help them effectively incorporate phonics, phonemic awareness, and other elements in beginning-reading lessons.

Effects for instructional process programs were very positive. Across 17 studies, five of which were randomized quasi-experiments, the weighted mean effect size for instructional process approaches in beginning reading was +0.37. The mean was +0.47 for decoding measures and +0.30 for comprehension/total reading measures. In particular, positive effects were seen on cooperative learning programs such as Peer-Assisted Learning Strategies (PALS) and Classwide Peer Tutoring (mean ES = +0.46), phonics-focused professional development programs such as Sing, Spell, Read, and Write, Early Reading Research, and RAILS (mean ES = +0.43), and teaching of phonological awareness to kindergartners (mean ES = +0.22 on tests at the end of first or second grade).

Beginning Reading: Combined Curriculum and Instructional Process Approaches

Two programs, Success for All and Direct Instruction (DI), both provide teachers with structured, phonetic reading materials and extensive training in the teaching of systematic phonics and other reading skills. These are by far the most extensively evaluated of all beginning-reading approaches, with 23 mostly large qualifying studies of Success for All and three large studies of DI.

Success for All is a whole-school reform model mainly used with high-poverty elementary schools. It provides schools with a K–5 reading curriculum that focuses on phonemic awareness, phonics, comprehension, and vocabulary, beginning with phonetically regular minibooks in kindergarten and first grade. Cooperative learning is extensively used at all grade levels. Children are frequently assessed on curriculum-based measures, and struggling first graders may receive one-to-one tutoring. Teachers receive extensive professional development and follow up, and a full-time facilitator in each school helps all teachers use the model effectively. A solutions team works with parents to help them support their children's achievement and to deal with issues such as attendance and behavior problems. Evaluations of Success for All, including a 3-year national randomized experiment, have found substantial positive effects of the approach, averaging an effect size of +0.29.

DI is a structured approach to beginning reading that emphasizes a step-by-step introduction of synthetic phonics, use of decodable text, and extensive professional development. A large, national evaluation of DI found only small positive effects on reading outcomes. A large Houston study found substantial positive effects, but a Baltimore study found few differences.

Kindergarten-Only Studies

As noted earlier, studies that began and ended in kindergarten were treated separately because of the problem of determining whether any program effects are due to early teaching of skills ordinarily taught later. However, several kindergarten-only programs are at least promising in their outcomes. The kindergarten-only program with the strongest positive effects is an instructional process approach called Ladders to Literacy, which provides extensive professional development on the teaching of phonics. Two large studies of this model found positive effects on reading outcomes, especially when Ladders to Literacy was combined with PALS. Two studies of a phonics program called Voyager Universal Learning System found conflicting results, with one showing positive effects and one no differences. Kindergarten studies of the phonetic Waterford CAI model also had conflicting outcomes, while another CAI model called Destination Reading had significant negative effects.

Conclusions: Beginning-Reading Studies

The most important overall conclusion of the Slavin *et al.* (in press) review is that systematic, synthetic phonics are necessary but not sufficient to build strong reading skills in the early elementary years. All of the programs that were successfully evaluated had a strong emphasis on systematic phonics, but so did many of the programs that were not found to be effective. Beyond phonics, what made the greatest difference is professional development. The most successful models provided extensive training and in-classroom follow-up in cooperative learning, phonetic teaching strategies, or (usually) both. The most extensively evaluated and successful model, Success for All, provides about 26 person-days of in service and follow-up in the first year, as well as an on-site facilitator. In contrast, none of the programs that provide only phonetic textbooks without extensive training or follow-up were found to have educationally significant positive effects. The same pattern was found in reviews of research on upper-elementary reading approaches (Slavin *et al.*, in press) and secondary-reading models (Slavin *et al.*, 2008a): instructional-process approaches with extensive professional development generally produced positive outcomes in rigorous evaluations, while textbook approaches made little difference. Instructional technology approaches showed more promise in beginning reading than in upper-elementary or secondary reading, but the amount of research on technology approaches was limited. Particularly promising were embedded multimedia approaches used with the whole class, in contrast to the individualized approaches that have been more typical applications of CAI.

Cost Effectiveness

The only cost-effectiveness analysis of any of the beginning-reading programs was done in a longitudinal study of Success for All by Borman and Hewes (2003) that followed Baltimore children who had been in Success for All or control schools in grades K–5. At eighth grade, the Success for All students still scored significantly better than controls, and they had been retained or assigned to special education substantially less than control students. The authors found the cost-effectiveness of the program to be greater than that of other interventions that had been studied longitudinally, the Perry Preschool, the Abecedarian Program, and reductions in class size to 15.

Beyond the Basics: Programs for the Upper-Elementary Grades

From second to fifth grade, children go through a critical transformation as readers. Most beginning second graders are able to decode, recognize key sight words, comprehend simple texts, and read with some degree of fluency. The tasks that lay ahead of them, however, are qualitatively different from those they have navigated so far. They must consolidate and extend their basic skills, to be sure, and they must become fluent, confident readers. But most importantly, children in the upper-elementary grades must become strategic comprehenders of increasingly sophisticated text. They must build a vocabulary of words and concepts as well as a vocabulary of cognitive and metacognitive approaches to texts. While decoding skills may develop in a fairly step-by-step progression, the skills mastered in the upper-elementary grades emerge as children read in many genres and learn how to make sense of what they read in a less-straightforward process. Stage theorists (e.g., Chall, 1983) point out that the upper-elementary years are when children transition from learning to read to reading to learn.

Due to the different objectives and requirements of the upper-elementary grades, programs that are effective in building beginning-reading skills are not necessarily optimal in the upper-elementary grades, and vice versa. For this reason, in reviewing research on effective reading programs, it is important to review programs at each of these levels separately. Slavin *et al.* (2008a) reviewed research on upper-elementary reading programs. The following sections summarize the findings of this review.

Upper-Elementary Reading: Curricula

The reading curricula category includes seven qualifying studies of core basal textbooks and nine studies of supplementary texts used as initial instruction with all students. Professional development is typically provided

with these curricula, but there is far less of it than would be typical of the programs categorized in this review as instructional-process programs or combined curriculum and instructional-process programs. In the reading-curriculum programs reviewed in this section, the theory of action is that improved content and curriculum-embedded assessments aligned with national or state standards will improve students' reading achievement. Reading outcomes of core and supplementary textbooks for the upper-elementary grades have not been previously reviewed. The Slavin *et al.* (2008b) review of middle and high school reading programs did not find any qualifying studies of reading curricula.

Both core and supplemental reading curricula for the upper-elementary grades have been studied in high-quality evaluations. Among 16 studies, there were six randomized experiments as well as three randomized quasi-experiments. These studies found few effects on student reading achievement. The weighted mean effect size for core reading curricula was only +0.06, and for supplementary curricula it was +0.07, with an overall weighted mean of +0.06. The only curriculum with promising effects was Open Court, but in both of the studies of this program, teachers received far more professional development than that usually provided, and in both studies Open Court was used for 2½ h per day while control students had 90 min of reading.

Upper-Elementary Reading: CAI

The effectiveness of CAI has been extensively debated over the past 20 years, and there is a great deal of research on the topic. Kulik (2003) concluded that research did not support use of CAI in elementary or secondary reading, although Chambers (2003) came to a more positive conclusion, giving a mean effect size of +0.25. A large randomized evaluation of various computer software programs by Dynarski *et al.* (2007) found no significant effects on the reading achievement of first and fourth graders.

Thirty-one studies of CAI met the standards for the review. These were divided into three categories. Supplemental CAI programs, such as Jostens/Compass Learning, Academy of Reading, LeapTrack, My Reading Coach, and CCC/Successmaker provided additional instruction at students' assessed levels of need to supplement traditional classroom instruction. Computer-Managed Learning Systems included only Accelerated Reader. This program uses computers to assess students' reading levels, assign reading materials at students' levels, score tests on those readings, and chart students' progress, but students do not work directly on the computer. Innovative Technology Applications included Fast ForWord and Lightspan.

A total of 31 qualifying studies evaluated various forms of CAI, eight of which used random assignment to treatments. The studies involved a total of more

than 10 000 students. Overall, the sample size-weighted mean effect size was very small ($ES = +0.06$). The randomized evaluations ($n = 8$) had a weighted mean effect size of +0.05. These findings support Kulik's (2003) conclusion that effects of CAI in reading are minimal.

Outcomes were similar across the three categories of CAI programs. Across 25 studies of supplemental programs (such as Jostens and CCC), the weighted mean effect size was +0.05. Two studies of Accelerated Schools had a mean effect size of +0.06, two studies of Fast ForWord had a mean effect size of +0.21, and a study of Lightspan had an effect size of +0.42.

Upper-Elementary Reading: Instructional-Process Programs

Instructional process programs are methods that focus on providing teachers with extensive professional development to implement specific instructional methods. In upper-elementary reading, instructional process programs are quite diverse. Thirty-two studies, six of which used random assignment, evaluated a broad range of approaches. Cooperative learning programs (Slavin, 1995, 2009) use methods in which students work in small groups to help one another master academic content.

Strategy-instruction programs teach students cognitive and metacognitive skills such as summarization, graphic organization, and prediction to help them comprehend text. Strategy instruction is often combined with other methods, especially cooperative learning and peer tutoring. Structured phonetic-intervention programs are approaches emphasizing phonics, systematic instruction, and frequent assessment of student progress. Phonics-focused professional-development programs are ones that teach teachers the National Reading Panel (NRP) elements, especially phonics and phonemic awareness, mostly in workshops. Integrated language arts programs are less structured and less phonetic, and focus on integrating reading and writing, literature study, and pleasure in reading. Cross-age tutoring programs involve older children working with younger ones, and same-age tutoring involves having children take turns tutoring one another. Classroom management and motivation programs focus on building a positive learning environment.

Both the methods and the findings of instructional process programs for upper-elementary reading were quite diverse. Across 33 experimental-control comparisons, involving more than 17 000 students, the weighted mean effect size was +0.21. These include four randomized and two randomized quasi-experimental studies.

Ten of the studies evaluated two forms of cooperative learning. These had a weighted mean effect size of +0.21. All but one of the cooperative learning studies evaluated Cooperative Integrated Reading and Composition (CIRC), which involves students in well-structured cooperative

groups within which they help each other master and apply metacognitive learning strategies. CIRC was the basis for middle school reading programs called Student Team Reading and The Reading Edge, which had a weighted mean effect size of +0.29 in four secondary studies. The consistent positive effects of this family of cooperative learning approaches support the idea that programs focusing on professional development in structured activities that engage children in discussions about reading, giving them opportunities to help each other learn and use metacognitive skills, may have particular promise for enhancing reading achievement from the second grade onward. Positive effects were also found for cross-age tutoring programs ($ES = +0.26$ in 4 studies) and for same-age tutoring ($ES = +0.26$ in 2 studies), reinforcing the conclusion that interaction among students on reading strategies is an effective approach. Another promising category was programs emphasizing metacognitive strategy instruction, such as Reciprocal Teaching and Thinking Maps, which had a weighted mean effect size of +0.32 in five studies. In these programs, students were taught skills such as prediction, summarization, and self-evaluation.

Upper-Elementary Reading: Combined Curriculum and Instructional-Process Programs

Five reading programs, Direct Instruction, Corrective Reading, Spell Read, Wilson Reading, and Project Read, were categorized as combining curriculum and instructional process approaches. In each case, the programs provide complete curriculum materials, replacing (rather than supplementing) basal texts. Yet, all five also provide extensive professional development, far more than that provided with basal texts, and their theories of action depend on creating significant changes in teachers' daily instructional behaviors. All five programs strongly emphasize phonics, and all five provide step-by-step teacher's manuals, training, follow-up, coaching, and frequent assessment of educational progress.

Three of the programs (Corrective Reading, Spell Read, and Wilson Reading), as well as Failure-Free Reading, were evaluated in the same randomized evaluation by Torgerson *et al.* (2006). Each program was compared to its own control group, so these are treated as four separate studies.

The six qualifying evaluations of combined curricula and instructional process programs involved a total of only 867 students. However, the evidence from these studies suggests that combining phonetic curricula and extensive professional development in instructional strategies is effective. The sample-size weighted mean effect size was +0.29. There were particularly positive effects for Direct Instruction/Corrective Reading in three studies ($ES = +0.34$).

Programs for Struggling Readers

Beyond the studies of core reading programs, there is extensive research on programs targeted to individual pupils who are struggling in reading. This research was reviewed by Slavin *et al.* (2009). The review concludes that one-to-one tutoring is very effective in improving reading performance. Tutoring models that focus on phonics obtain much better outcomes than others. Teachers are more effective than paraprofessionals and volunteers as tutors. Small-group phonetic tutorials can be effective, but are not as effective as one-to-one phonetically focused tutoring. Classroom instructional-process programs, especially cooperative learning, can have very positive effects for struggling readers. CAI generally had few effects on reading. Taken together, the findings support a strong focus on improving classroom instruction and then providing one-to-one phonetic tutoring to students who continue to experience difficulties.

Conclusions

Research on reading approaches for the elementary grades finds several programs with significant promise for increasing children's learning and school success. The evidence particularly supports the use of approaches such as cooperative learning, which provide extensive professional development to teachers to help them make their teaching more active, motivating, and interactive for all pupils. Particular uses of technology may also contribute to children's success, and tutoring approaches for struggling pupils can help them get off to a good start in school.

What is most important about the evidence cited here is that it clearly indicates that schools need not accept the current reading levels of young children. Proven, replicable approaches are available and others along similar lines could be developed. A judicious policy of using what works, while expanding the number and variety of available programs, could make a broad and meaningful difference in students' success during the crucial early elementary years.

See also: Cost-Benefit Analysis and Cost-Effectiveness Analysis; Literacy Instruction for Students with Special Needs; Reading and Technology; The Economics of Early Childhood Interventions.

Bibliography

- Borman, G. and Hewes, G. (2003). Long-term effects and cost effectiveness of success for all. *Educational Evaluation and Policy Analysis* 24(2), 243–266.

- Campuzano, L., Dynarski, M., Agodini, R., and Rall, K. (2009). *Effectiveness of Reading and Mathematics Software Products: Findings from Two Student Cohorts*. Washington, DC: US Department of Education.
- Chall, J. S. (1983). *Stages of Reading Development*. New York: McGraw-Hill.
- Chambers, E. A. (2003). *Efficacy of Educational Technology in Elementary and Secondary Classrooms: A Meta-Analysis of the Research Literature from 1992–2002*. Unpublished doctoral dissertation, Southern Illinois University at Carbondale.
- Dynarski, M., Agodini, R., Heaviside, S., et al. (2007). *Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort*. Washington, DC: Institute of Education Sciences.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology* **80**(4), 437–447.
- Kulik, J. A. (2003). *Effects of Using Instructional Technology in Elementary and Secondary Schools: What Controlled Evaluation Studies Say. SRI Project Number P10446.001*. Arlington, VA: SRI International.
- Lloyd, D. N. (1978). Prediction of school failure from third-grade data. *Educational and Psychological Measurement* **38**, 1193–1200.
- National Reading Panel (2000). *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. Rockville, MD: National Institute of Child Health and Human Development.
- Rose, J. (2006). *Independent Review of the Teaching of Early Reading*. London: Department for Education and Skills.
- Slavin, R. E. (1995). *Cooperative Learning: Theory, Research, and Practice*, 2nd edn. Boston: Allyn and Bacon.
- Slavin, R. E. (2009). Cooperative learning. In McCulloch, G. and Crook, D. (eds.) *International Encyclopedia of Education*. Abington, UK: Routledge.
- Slavin, R. E., Cheung, A., Groff, C., and Lake, C. (2008a). Effective programs in middle and high school reading: A best-evidence synthesis. *Reading Research Quarterly* **43**(3), 290–322.
- Slavin, R. E., Lake, C., Chambers, B., Cheung, A., and Davis, S. (in press). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research*.
- Slavin, R. E., Lake, C., Cheung, A., and Davis, S. (2008b). *Beyond the Basics: Effective Reading Programs for the Upper Elementary Grades*. Baltimore, MD: Center for Research and Reform in Education, Johns Hopkins University.
- Slavin, R. E., Lake, C., Madden, N. A., and Davis, S. (2009). Effective Programs for Struggling Readers. Manuscript submitted for publication. <http://www.bestevidence.org>.
- Snow, C. E., Burns, S. M., and Griffin, P. (eds.) (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.
- Torgerson, C. (2006). The quality of systematic reviews of effectiveness in literacy learning in English: A tertiary review. *Journal of Research in Reading* **29**(2), 1–29.

Personalized Learning in School Age Education

D Hopkins, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Personalized learning is an idea that is capturing the imagination of teachers, parents, and young people around the world. It is an idea that has its roots in the best practices of the teaching profession and has the potential to make every young person's learning experience mind-enlarging, creative, fun, and successful.

Personalization is the guiding motif that allows a system to evolve from one based on delivery of services to one that emphasizes mass customization and co-production. It is about putting citizens at the heart of public services and enabling them to have a say in the design and improvement of the organizations that serve them. In education this can be understood as personalized learning – the drive to tailor schooling to individual need, interest, and aptitude. This emphasis provides a bridge from prescribed forms of teaching, learning skills, curriculum, and assessment to an approach to classroom practice that is predicated on enabling every student to fulfill their potential. In his pamphlet *Learning about Personalisation: How Can We Put the Learner at the Heart of the Education System?* Leadbeater (2004: 16) clearly and sensitively links the concept of personalization with personalized learning as the key driver for the transformation of schooling:

The script of a system characterised by personalised learning . . . would start from the premise that the learner should be actively engaged in setting their own targets, devising their own learning plans and goals, choosing from a range of different ways to learn. . . . By making learning the guiding principle of the system, personalisation challenges some of the current divide and boundaries that exist – for example between formal and informal learning; between academic and vocational learning and between different ages and types of learners. (Leadbeater, 2004:16)

In one sense, personalization represents a logical progression from the standards and accountability reform strategies of the 1990s. These strategies marked an important first phase in a long-term large-scale reform effort, but in order to sustain systemwide improvement, societies are increasingly demanding strategies characterized by diversity, flexibility, and choice.

It is moral purpose that drives personalization. We see it most vividly in the concern of the committed, conscientious teacher to match what is taught, and how it is taught to the individual learner as a person. That is not just a question of sufficient challenge of aligning pedagogy to the point of progression that each learner has

reached, even though that is vitally important. It is also part of the teacher's concern to touch hearts as well as minds, to nourish a hunger for learning, and help equip the learners with a proficiency and confidence to pursue understanding for themselves.

In exploring the concept of personalized learning in this article, we shall:

- define the concept a little further and review its main components;
- explore the curricula implications of taking personalized learning seriously;
- emphasize the meta-cognitive aspects of personalized learning; and
- in the coda briefly discuss how to move personalized learning to scale.

What Is Personalized Learning?

Personalized learning is not a new idea. Many schools and teachers have tailored curriculum and teaching methods to meet the needs of children and young people with great success for many years. What is new is the drive to make the best practices universal. It is reimagining the education system around the learning needs and talents of young people that is the basis for every school becoming great.

To build a successful system of personalized learning, we must begin by acknowledging that we should be giving every single child the chance to be the best they can be, whatever their talent or background. Personalized learning means high-quality teaching that is responsive to the different ways students achieve their best. There is a clear moral and educational case for pursuing this approach. A system that responds to individual pupils by creating an education path that takes account of their needs, interests, and aspirations will not only generate excellence but will also make a strong contribution to equity and social justice.

One can summarize this approach to personalized learning as follows (Hopkins 2007):

- As an educational aspiration, personalized learning reflects a systemwide commitment to moral purpose, high excellence, high equity, and to every school being or becoming great.
- As an educational strategy, personalized learning relates to and builds on the learner's experience, knowledge, and cognitive development; develops their confidence

and competence; and leads toward autonomy, emancipation, and self-actualization.

- As an approach to teaching and learning, personalized learning focuses on individual potential, develops the individual's learning skills (particularly ICT), and enhances creativity and social skills.
- As a curriculum orientation personalized learning offers an approach to subject teaching that balances societal aspirations and personal relevance and unifies the curriculum offer across sectors and age groupings.

This leads directly to the following implications that can help guide day-to-day practices:

- for children and young people, it means clear learning pathways through the education system and the motivation to become independent, e-literate, fulfilled, life-long learners;
- for schools, it means a professional ethos that accepts and assumes every child comes to the classroom with a different knowledge base and skill set, as well as varying aptitudes and aspirations; and because of that, there is a determination for every young person's needs to be assessed and their talents developed through diverse teaching strategies;
- for school governors, it means promoting high standards of educational achievement and well-being for every pupil, ensuring that all aspects of organizing and running the school work together to get the best for all pupils;
- for the national and local authorities, it means a responsibility to create the conditions in which teachers and schools have the flexibility and capability to personalize the learning experience of all their pupils, combined with a system of intelligent accountability so that central intervention is in inverse proportion to success; and
- for the system as a whole, it means the shared goals of high quality and high equity.

The rationale of these principles is clear: to raise standards by focusing teaching and learning on the aptitudes and interests of pupils and removing any barriers to learning. The key question is how to collectively build this offer for every pupil and every parent.

There are six key components that can help to deepen and extend this personalization of education. **Figure 1** illustrates the central components of personalized learning within and beyond the school. They are:

1. Assessment for learning implies:

- that the school, its teachers, and the system develop a high-level capacity for using data to promote student learning;
- the process of seeking and interpreting evidence for learners and teachers to decide where the learners are in their learning, where they need to go, and how best to get there; and

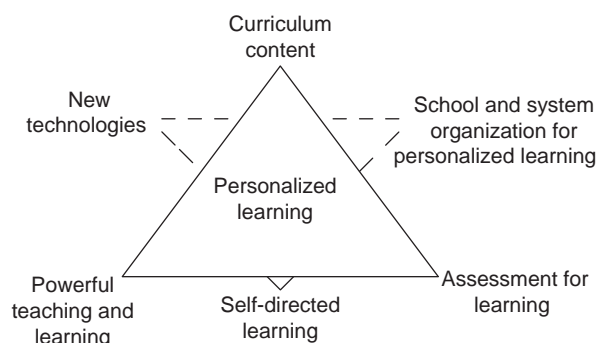


Figure 1 Central components of personalized learning within and beyond the school.

- shared objectives, feedback that identifies targets for improvement, higher order questioning, and self and peer assessment.
2. Powerful learning and teaching implies:
- the curriculum should focus on understandings and competencies that have enduring and intrinsic value;
 - high expectations and challenging targets should be set for all, but while standards should remain constant, time and support should be varied according to individual student need; and
 - teachers should show students how to incorporate new information into their existing knowledge through activities that induce critical thinking with conceptual problems.
3. Self-directed learning implies:
- that such learning contracts provide the basis for project work that is an essential and ongoing feature of the curriculum offer;
 - individual learning profiles; and
 - counterintuitively, strong emphasis laid on cooperative group learning and social interaction.
4. Customizing the curriculum offer implies:
- modifying the framework of the national curriculum to ensure a continuum for personalized learning across the three phases of education—foundation, middle, and 14–19;
 - using the enquiry into subjects within the context of standards as the building block of curriculum provision; and
 - involving students in the formulation of their own educational goals. This is the key to them establishing engagement in learning as well as a long-term commitment to schooling.
5. The contribution of the new technologies implies:
- opportunity for personal creativity, the ability to match curriculum to individual learning styles, and putting the pace of learning under the individual's control;
 - concurrent and extended learning opportunities outside of normal school day; and

- building diagnostic assessment for learning with different pathways to follow.
6. Organizing schools for personalized learning implies:
- differentiating the workforce for student learning, enhancing the role of the learning mentor, and providing each learner with the link to an adult;
 - block timetabling and grouping students on basis of learning need, within and between schools to ensure network and community learning; and
 - establishing a system of transferable learning profiles and credits to underpin assessment and ensure flexibility.

Obviously, all of these six components are important as they intend to shape a modern pedagogy centered on learning, but they are not necessarily exclusive. Here, however, two aspects of personalization that are crucial to ensuring that all students reach their potential are highlighted. The first is customizing the curriculum offer in order that breadth of study, personal relevance and flexible learning pathways can be delivered throughout the education system. The second is meta-cognition or learning how to learn.

Toward the Personalized Curriculum

It is the curriculum that provides the means for every school becoming a great school and all students realizing their potential. Yet in many educational systems the curriculum is often a barrier to achieving the forms of personalized learning so necessary for such a transformation. One way of both pinpointing the curriculum issue and drawing the argument together is to pose the question “What does it mean to be educated” at any particular phase of education.

Being educated at any particular age has four central elements:

- a breadth of knowledge gained from a curricula entitlement;
- a range of skills on a developmental continuum that reflects increasing depth at ages 7, 11, 14, 16, and in many cases, 18;
- a range of learning experiences; and
- a set of key products, projects, or artifacts.

It also means that students are sufficiently articulate to:

- sustain employability through basic skills;
- apply their knowledge and skills in different contexts;
- choose from and learn in a range of post-14 study (assuming an entitlement curriculum up until then); and
- draw on wider experiences to inform further learning and choice.

We suggest that most national curricula do not meet these desiderata. In general, central authorities set out

mandatory programs of study in all subjects, with schools generally free to determine their own timetables and design their own programs of teaching. In particular, the following issues seem to be shared across many jurisdictions and militate against the realization of personalized learning:

- curriculum congestion – the curriculum suffers from content and process overload;
- expected levels – do clear levels of attainment exist in the mother tongue, maths, and science?
- poor catch-up provision or progression for low-attaining pupils;
- inadequate embedding of core skills;
- lack of clarity for students on the common learning skills;
- insufficient stretch for the most able; and
- staffing problems.

Although the following proposals are based on the author’s understanding of the Key Stage 3 (11–14) curriculum in England, they represent a broader attempt to imagine a structure that enables schools and teachers to personalize the curriculum across all stages of education.

The first is to focus on core study. Functional literacy, numeracy, and communication could be clarified as expected attainment at the end of the KS3 curriculum. ICT would also need to be explicitly added to a suggested core of mother tongue, math, and science. Functional skills would similarly need to be embedded across the curriculum.

The second would be a condensed statutory curriculum in noncore subjects combined with an optional entitlement. In many countries, this is referred to as the essential curriculum. This means that the statutory curriculum content and processes in noncore subjects would be reduced. As a rule of thumb, in most national systems this would mean that the content removed would be approximately 20–25% of current specifications. The reduction could be redesignated as an optional entitlement.

Third, the flexibility of an optional entitlement would allow schools to guarantee time to:

- secure essential knowledge and teach common learning skills through the curriculum; and
- organize the curriculum to meet the needs of a range of abilities, tailoring support for underachieving and underperforming students, and to stretch gifted and talented students.

Fourth, there needs to be clarity on common learning skills. This requires that a common framework of skills is identified across the whole curriculum, as is seen in the following section.

Finally, there is the need to champion effective pedagogy. There needs to be external support to help schools organize the curriculum to meet the needs of a range of

abilities. It must also help teachers bring curriculum knowledge and common learning skills together in the classroom.

The clear prize from pursuing these actions would be a curriculum tailored to the needs, talents, and aptitudes of all students. This would ensure that every student had the core and common skills required to learn at each stage of education, and that the best students were properly stretched.

In this section, WE have taken a view on the necessary reform of any national or local curriculum to accommodate personalized learning. But underpinning the importance of structural curriculum change is the necessity that it accommodates not just curricular knowledge, but also learning how to learn. This essential aspect of personalized learning is dealt with ahead.

Meta-Cognition and Learning How to Learn

Meta-cognitive skills enable students to develop the capacity to monitor, evaluate, control, and change the way they think and learn. There is clear evidence that the acquisition of these skills can significantly increase achievement. To ensure more students gain these skills we need:

- first, teaching strategies that consistently and strategically develop students' learning skills. For example, instead of simply presenting information for knowledge acquisition, teachers can ensure that in tandem with learning new knowledge, students also extract ideas, memorize information, build hypotheses and theories, use metaphors to think creatively, and work effectively with others.
- second, a framework of common learning skills, as there is currently a lack of clarity for students on the skills they should acquire and how they can develop these as they progress. These skills would need to be identified and taught coherently across the curriculum.

We explore the framework for learning skills in this article. Inevitably, there is overlap between these two vital aspects of personalizing learning as they are crucially the opposite sides of the same coin. This point was made by Joyce *et al.* (2002: 7) by arguing that it is the teacher's task not simply to teach but to create powerful contexts for learning. It is a truism that no one can teach anyone anything: the best that can be done is to help another to learn. That idea and the essence of personalized learning was expressed thus:

Learning experiences are composed of content, process and social climate. As teachers we create for and with our children opportunities to explore and build important areas of knowledge, develop powerful tools for learning, and live in humanising social conditions.

It is the integration of “content, process and social climate” that puts the power into the personalized learning experience. But at the heart of personalized learning is its impact, not just on test scores and examination results, but also on the students' learning capability. If the teacher can teach the students how to learn at the same time as assisting them to acquire curriculum content, then the twin goals of learning and achievement can be met simultaneously.

A fundamental aspect of personalized learning is therefore the ability of learners to respond successfully to the tasks that they are set as well as the tasks they set themselves – in particular to:

- integrate prior and new knowledge,
- acquire and use a range of learning skills,
- solve problems individually and in groups,
- think carefully about their successes and failures,
- evaluate conflicting evidence and to think critically, and
- accept that learning involves uncertainty and difficulty.

The deployment of such a range of learning strategies is commonly termed meta-cognition, which can be regarded as the learners' ability to take control over their own learning processes. The key point is that within whatever context learning takes place, it involves an active construction of meaning. This carries implications for the management of learning opportunities, in particular that an active construction of meaning requires practical, cognitive, and other learning strategies. As learning is interactional it can occur only as the learner makes sense of particular experiences in particular contexts. This making sense involves connecting with an individual's prior knowledge and experience. Thus, new learning has to relate to, and ultimately fit with what individuals already understand (Vygotsky, 1962).

If we are serious about personalized learning, then we need to be clear as to the typology of the skills students should gain in order to develop their personal effectiveness and employability. These skills fall into three categories:

- functional skills,
- thinking and learning skills, and
- personal skills.

Functional skills are literacy, numeracy, and ICT. In most countries, these are regarded as key priorities and therefore tend to be taught and assessed in the core subjects of mother tongue, math, and ICT. We have already discussed how they can be integrated into a revised curriculum structure.

Thinking and learning skills are the skills young people need to acquire in order to become effective learners. Gaining mastery of these skills equips students to raise their achievement by developing their ability to:

- improve their achievement by applying a wide range of learning approaches in different subjects;
- learn how to learn, with the capability to monitor, evaluate, and change the ways in which they think and learn; and
- become independent learners, knowing how to generate their own ideas, acquire knowledge, and transfer their learning to different contexts.

Personal skills are the skills young people need to acquire in order to develop their personal effectiveness. Gaining mastery of these skills equips students to manage themselves and develop effective social and working relations, initiative, and self-motivation.

This framework provides clarity on the skills students should gain. But to ensure students gain mastery of these skills there needs to be agreement on how the skills should be:

- embedded in teaching and learning, especially since in most systems these skills are not well-taught in all schools and consequently students' skill development is patchy and
- developed coherently across the curriculum especially as again in most systems, these skills are not specified or developed in a systematic way and progression is assumed and not explicit.

This skills framework is consistent with trends in other countries. For example, the OECD's (2005) Definition and Selection of Competencies (DeSeCo) Project, which classified individuals' key competencies for a successful life into three broad categories:

1. Use tools interactively (both physical and sociocultural ones):
 - a. use language, symbols, and texts interactively,
 - b. use knowledge and information interactively, and
 - c. use technology interactively.
2. Interacting in heterogeneous groups and specifically to:
 - a. relate well to others,
 - b. cooperate and work in teams, and
 - c. manage and resolve conflicts.
3. Acting autonomously:
 - a. act within the big picture,
 - b. form and conduct life plans and personal projects, and
 - c. defend and assert rights, interests, limits, and needs.

So in summary, the clarity provided by a single skills framework allied with better guidance and training on pedagogy will itself create greater coherence across any national curriculum. This is the necessary foundation for ensuring that the essence of personalization is available for every student. This in turn is a critical building block in ensuring that every school is a great school. In the final section we examine further how these initiatives can be taken to scale.

Coda – Moving Personalized Learning to Scale

In this final section of this article, we discuss briefly how to move personalized learning to scale. At a minimum three complementary sets of activities are required to successfully move personalized learning to scale:

- First, the personalized learning offer needs to be developed and made the centrepiece of the range of national policy options. In parallel, curriculum content and subject specialism need to be reviewed in light of the key components of personalized learning and relentlessly connected to the standards agenda. For example, **Table 1** illustrates how this work was carried forward in England as part of the Key Stage 3 (lower secondary, 11–14 years old) national support program.
- Second, the personalized learning offer needs to be the central agenda of all national and local partnership arrangements, for example, national/local negotiations, the developing of a consensus amongst national support organizations, and mobilizing nongovernmental organizations.
- Third, personalized learning and moral purpose must be made synonymous, that is, personalized learning is important because it will enable all young people to reach their potential. It becomes the key message in national and local government's communication policy and conferences as well as the recipient of a series of symbolic (and substantive) policy initiatives, for instance, learning vouchers for every 14-year-old.

Table 1 Key Stage 3 and personalized learning

Assessment for learning	<ul style="list-style-type: none"> • Setting personal pupil targets • Introducing self and peer assessment • Developing more effective feedback to pupils
Powerful learning and teaching	<ul style="list-style-type: none"> • Developing thinking and learning skills across the curriculum • The learning challenge program • Mentoring skills
Self-directed learning	<ul style="list-style-type: none"> • Teaching study and research skills • Introducing learning contracts
Curriculum	<ul style="list-style-type: none"> • Guidance on different curriculum models for KS3 • The 2-year KS3 pilot – creating time for the tailored curriculum
New technologies	<ul style="list-style-type: none"> • Implementing ICT across the curriculum • Teaching the use of the Internet • Interactive teaching programs
Organizing schools for personalized learning	<ul style="list-style-type: none"> • Models and materials for catch-up provision • Involving parents project • Teaching strategies that include and challenge everyone

To meet the full range of individual needs and aspirations inherent in the goal of excellence for all requires extensive, but disciplined, innovation of many different kinds and at different levels in educational provision and professional practices. This entails changes (1) at the level of the classroom, for example, through ICT training for classroom teachers and leaders to build on the ICT infrastructure; (2) at the level of the school, for example, through workforce reform to ensure that teachers spend more time actually teaching; as well as at (3) system level such as the creation of federations between schools and other collaborative arrangements. It should by now be clear that the dissemination of the outcomes of innovation and new methods of achieving excellence cannot be achieved by a center–periphery model alone, but requires new mechanisms of lateral transfer through networks. The challenge for personalized learning is to create a

learning system that is capable of adapting to deep changes in our economy and society by pursuing universal participation and achievement.

Bibliography

- Hopkins, D. (2007). *Every School a Great School*. Buckingham: McGraw Hill/Open University Press.
- Joyce, B. R., Calhoun, E., and Hopkins, D. (2002). *Models of Learning – Tools for Teaching*, 2nd edn. Buckingham: Open University Press.
- Leadbeater, C. (2004). *Learning about Personalisation: How Can We Put the Learner at the Heart of the Education System?* Nottingham: Department for Education and Skills.
- OECD (2005). *The Definition and Selection of Key Competencies*. <http://www.oecd.org/dataoecd/47/61/35070367.pdf> (accessed September 2009).
- Vygotsky, L. S. (1962). *Thought and Language*. Cambridge, MA: MIT Press.

Perspectives on Schooling in the Middle Years

V J Russell, The University of Melbourne, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction: Betwixt and Between

The beginning and final years of schooling, with their focus on foundational skills or school completion achievements and opportunities, commonly act as magnets drawing strong public, parental, political, media, and educational attention. Less public interest has been generated by the years of schooling in between. These in-between years have been termed the middle years (or middle grades/phase) of schooling; they, generally encompass students of ages 10–15 or, in a K–12 system, grades 5–9, with some flexibility in the upper and lower limits of these groupings. The term middle schooling implies a characteristic orientation to the education in the middle years, centered on early adolescents' needs, interests, and capacities and linked to certain psychological, sociological, philosophical, and educational concepts. Another relevant concept is the middle school. This refers not to an orientation to schooling but to an organizational unit—a separate school, or a separate campus or section of a larger school. The middle school is a structural concept. Middle schools do not necessarily implement a middle schooling approach.

In this article, consideration is given first to recent interest in the middle years of schooling, amid changing global contexts, increasing complexity of stated goals of learning, and expanding research-based knowledge of young adolescent students and their learning. From the diversity of individual school and system approaches to schooling in the middle years, two different characteristic types of approach are identified, the first focusing on the middle school as structure and the second focusing on middle schooling as practice. A description is given of what is thought to constitute high-quality middle schooling. The question of an evidence base for middle schooling approaches is an important one. Two aspects are discussed: the extent to which middle schooling approaches have actually been implemented and the effectiveness of such approaches in achieving stated goals of learning. Evidence is far from abundant, although what there is suggests that better designed macro-level research would support existing, sound, micro-level research to provide the evidence base that is sought.

Increasing Interest in the Middle Years

The last two decades have witnessed a groundswell of interest in the middle years of schooling and a deepening

of understanding of what constitutes good middle years schooling. From the late 1980s, there has been a spate of publications in this area, including journal articles with regard to fundamental psychological processes relevant to early adolescence; special journal issues regarding middle schooling; establishment of targeted journals; investigations and reports linking adolescence, engagement, and approaches to schooling; and advice concerning appropriate forms of middle schooling. Several factors have contributed to this renewed interest.

Reduced Progress During the Middle Years

Evidence has been accumulating that the growth trajectory of students' learning flattens during the middle years; a decline in student engagement underpins the reduced growth in achievement. A large, longitudinal Australian study based on repeated measures of 5 year-level cohorts revealed a lack of growth (even some decline) in literacy and reducing enjoyment of school in grades 5–8 (Hill *et al.*, 1996, cited in Hill and Russell, 1999). Similar dips in achievement and engagement of students have been reported in a range of countries, particularly in the lower secondary years, with socioeconomically and ethnically disadvantaged students being most affected (Whitby *et al.*, 2006). The role of transition from primary to secondary school in such dips has been highlighted (Galton *et al.*, 1999), although the decline begins prior to transfer. Improvement resumes at the end of the compulsory years, perhaps reflecting the fact that disengaged students have left school.

Declining Engagement During the Middle Years

Engagement, or directed energy in action, is evidenced in students' actual learning behavior. It is student engagement rather than underlying motivation to achieve that reduces markedly during the middle years (Russell *et al.*, 2003). Student engagement is responsive to the nature of the learning environment; maintenance of a stimulating, positive environment is a factor under school control.

Indicators of declining engagement in school vary from the less to the more serious, for example, from boredom or passivity to delinquency or self-harm. All can act as indicators of potential early school leaving. An English study of Key Stage 3 students suggested to Barber (1996) that nearly 20–30% of students were bored, 10–15% were behaving badly and truanting regularly,

while 2–5% had given up school altogether. The consequences for individual students of lost opportunities for realizing their potential are clear; so, too, are the economic consequences for the country. The disadvantaged are affected disproportionately.

Adolescence

An earlier theoretical and descriptive focus on the negative in adolescence, the storm and stress of troubled youth and their risky behavior, led many educators to the view that declining adolescent engagement in learning was inevitable and therefore immutable. In recent decades, a shift to interest in the capacities of adolescents, particularly their expanding cognitive potential, coping strategies, individual agency, and potential for positive development, has been accompanied by well-honed, methodologically sophisticated, often longitudinal research on specific aspects of adolescent behavior.

The needs and challenges of adolescence continue to be acknowledged: adjustment to physical, social, emotional, and intellectual changes; development of a positive self-concept, identity, and values; growth of independence, agency, and responsibility; establishment of positive relationships with peers and adults; and growth in the capacity to respond constructively to the physical, social, and political environment. However, the present research emphasis on the capacities of adolescents points to educational approaches that improve the course of adolescent experience and development (Steinberg and Lerner, 2004).

Deeper Understanding of Learning Processes

The strongest school contribution to student learning is commonly acknowledged to be the classroom teaching and learning process. Understanding of the learning process has deepened, moving beyond dependence on simple concepts of drives and rewards or discovery and activity, to focus on the perceptions, thoughts, and beliefs of the learner, the way the learner takes in information, interprets and processes it, draws meaning from it, and stores it. Learning is seen as an active process of building meaning; it takes place in all situations, even the least promising, but the best-quality learning takes place when students have the opportunity to think, use and manipulate ideas, create new connections, develop deeper understandings of existing knowledge, construct more complex mental models of the world, and gain greater confidence and capacity to regulate their own learning. As the basic understanding of constructivism has increased, so too has the quality of its translations into classroom practice. Contributions also came from the pursuit of thinking-skill development, careful brain-based research, deeper understandings of student motivation, engagement and self-regulation, and recognition of the value of the student voice, giving rise

to more productive pedagogical approaches within a context of positive classroom climate and student–teacher relationships.

Changing Contexts and the Goals of Learning

The developing globalized knowledge society necessitates rethinking of the goals of schooling and the means of meeting them. Change is rapid, continuing, and pervasive: change in knowledge, information, work, communication, mobility, diversity, and relationships. In such a context, educational goals that focus on literacy, numeracy, and current factual knowledge are inadequate. One of the main goals in this context of change should be the development of people with the capacity to adapt – lifelong learners who want to continue to learn and who know how to go about doing so. Many related skills, characteristics, and qualities are also required (Russell *et al.*, 2003). To achieve these goals, young people need to enjoy their learning and complete their education. This means taking reform in middle years' schooling seriously. Such reform would affect the whole of schooling, transforming the "whole ecology of formal education" (Hill and Russell, 1999: 180).

Approaches to Schooling in the Middle Years

Although national education systems, based on different values and structures, vary in their approaches to middle schooling, common issues do exist. Schooling in the middle years is commonly seen as the weak link in a review of eight European systems (Hirsch, 1994), showing a tendency to emphasize examinations assessing core knowledge transmission via didactic teaching to the detriment of attention to the developmental needs of early adolescents. Another multinational study of 16 systems (Greenaway, 1999) highlighted the unclear role of lower secondary education and negative effects of transfer between phases of schooling.

The English system illustrates many of the common challenges. The lower secondary years (grades 7–9), are a cause of considerable concern (Barber, 1999). English middle schools (encompassing both upper primary and lower secondary grades) have been accused of insufficient academic rigor and thus of being part of the problem, yet it is the transition from primary to secondary schools that is commonly viewed as producing a decline in student progress (Galton *et al.*, 1999). Although English middle schools are being closed, improvement is unlikely via a structural solution. In the lower secondary Key Stage 3 Strategy, emphasis has been placed on potentially more productive approaches aimed at increasing student engagement and positive transition experiences through improving the quality of pedagogy and teacher learning.

Among the school systems that have sought change from the traditional primary student-centered or secondary subject-centered approaches in the middle years, two characteristic orientations are evident. One has focused historically on school structure, the other on school practice. These two orientations are not mutually exclusive.

Middle School Structures

The USA has long been an active participant in the provision of middle schooling. Its dominant orientation for many years was a structural one. Several waves of middle years' reform are discernible. Serious national economic issues stimulated the creation of junior high schools (grades 7–9) at the beginning of the twentieth century, in order to reduce early leaving and improve prevocational preparation and employment. Progressive reforms in curriculum were heralded but did not emerge, as junior highs aped their more powerful senior colleagues and became imitation high schools (Cuban, 1992).

Educators' dissatisfaction with junior highs in the 1960s and 1970s led to rapid development of middle schools (grades 5/6 to 8) aimed at meeting the needs of early adolescents. Organizations such as the Association for Supervision and Curriculum Development, the National Association of Secondary School Principals, and the National Middle School Association (NMSA) provided vision, publications, and other support (e.g., *This We Believe*, first published by NMSA in 1982, with the latest revision in 2003) (NMSA, 2003). However, by the late 1980s, little seemed to have changed in terms of school curriculum, organization, and pedagogy (Cuban, 1992). Criticism was also leveled at middle schools by those who perceived them to lack academic rigor and discipline.

In a climate of economic uncertainty and concern about equity in a changing population, the 1989 publication of the Carnegie Council report, *Turning Points*, was influential (Carnegie Council on Adolescent Development, 1989). Its assertion of a mismatch between middle school organization and curriculum and the needs of young adolescents, both intellectual and emotional, provoked serious consideration of their recommendations for transforming middle schools: creation of small learning communities, teaching a core academic program, ensuring success for all, empowering school-based decision-making, employing specialist middle years teachers, fostering health and fitness, and involving families and community in schools.

In the following decade, some changes were detected, although these tended to concern within-school structural changes to enhance relationships and well-being rather than improvements of core student experience in curriculum, pedagogy, and assessment (Jackson and Davis, 2000). Concern was expressed about reform being confined to

pockets of excellence, lack of growth in student learning (reinforced by USA performance on international studies of achievement), and on-going lack of equity in achievement (Haycock and Ames, 2000). *Turning Points 2000* (Jackson and Davis, 2000) aimed to rectify the situation by providing a coherent framework for middle schooling that balanced caring relationships with more powerful and appropriate approaches to curriculum, instruction, and assessment. At present, the balance seems to be altering, with language use and goals highlighting rigorous public academic standards and high achievement. What happens will depend on how these concepts are spelled out in practice. A narrow interpretation could see the return of traditional forms of curriculum and classrooms. This was not the intention of Jackson and Davis (2000).

Middle Schooling as Practice

In Australia, discussion of early-adolescent education has been more recent and has focused on the orientation of practices used in the middle years rather than on school structure. From the early 1990s, a body of mainly national but also state reports stimulated debate, research, and adoption of changed practice in middle schooling. Motivating this activity were concerns with regard to schooling for early adolescents in grades 5–9 (grade levels that crossed the primary–secondary school divide) that did not match their needs or capacities, with consequent negative effects on engagement, retention, employment, and skill development for the globalized knowledge society. In contrast to the USA situation, there was little focus on minority urban group disadvantage or extreme, risky behaviors.

Victoria and Queensland are two Australian states that have developed coherent and comprehensive approaches to the middle years. Local research findings in the early 1990s stimulated Victorian concern regarding the middle years: the grade 5–8 plateau in student progress, decline in engagement, and greater between-class than between-school differences, pointing to the vital role of classroom pedagogy (Hill and Russell, 1999). The government established the *Middle Years Research and Development (MYRAD) Project* (1998–2001), aimed at evaluating a whole-school approach to improving student outcomes in grades 5–9. Evidence from the project shaped the subsequent middle schooling approach in Victoria.

Emphasis was placed on classroom climate and practices that established a balance between developmentally appropriate supportive relationships and challenging, authentic pedagogy, curriculum, and assessment. Other processes were seen to be necessary if changes inside the classroom were to be developed and sustained, for example, cooperation, planning, and consistency between clustered primary and secondary schools; use of a whole-school approach to vision and planning; adoption of an evidence-based approach; and system–policy coherence and alignment in

curriculum, assessment, pedagogical, and accountability dimensions of schooling (Russell *et al.*, 2003). An outline of the key strategic intentions of the approach can be found in Hill and Russell (1999). Subsequent research and development projects fleshed out the MYRAD recommendations, while funded programs, such as *Schools for Innovation and Excellence*, scaled up government–school adoption of middle schooling and a new government P–10 curriculum framework was developed, incorporating these principles. The coherence of the Victorian middle years reform agenda could be seen in the policy documents, reports, resource materials, and system level support that were developed.

In 2003, Queensland published its *Middle Phase of Learning State School Action Plan*. This policy document encapsulates the middle schooling approach, adopted in that state for grades 4–9. It is predicated on the concepts of the needs, challenges, and expanding horizons of early adolescents; on the lack of learning progress, decline in engagement, and increased gap between high and low achievers during that period; and on the importance of developing more sophisticated thinking and related skills that are needed in the knowledge economy. Central to the middle schooling outlined in the *Action Plan* is alignment of authentic curriculum, pedagogy and relationships, and assessment. It commits the Queensland system to ensure that every student in grades 4–9 engages in purposeful, intellectually challenging, and successful learning, with particular support in transition periods, and interaction with developmentally appropriate and individualized teaching.

The Queensland government's policy was able to draw on extensive and intensive Queensland research, including research undertaken by Luke and colleagues that targeted middle years literacy and numeracy, *Beyond the Middle* (2002), and the *Queensland School Reform Longitudinal Study* conducted by Lingard *et al.* (2001). The latter gave rise to the Productive Pedagogies framework, which together with the real-world,

Rich Tasks, forms the conceptual pivots of Queensland's New Basics curriculum framework. The consistency and comprehensiveness of Queensland's approach to the middle years can be seen in its provision of systemic support and range of helpful publications, reports, and resource materials, in addition to the policies it has established.

High-Quality Middle Schooling

Separate middle school institutions are neither necessary nor sufficient for high-quality middle schooling. In fact, such restructuring can divert attention from and exhaust energy for the task of creating and embedding a successful middle schooling approach. There is no neat or even agreed definition of what constitutes high-quality middle schooling. The middle schooling approach is often described in piecemeal fashion – a series of signifying practices, such as teacher teaming, integrative curriculum, cooperative learning, and flexible timetabling, that lack a coherent framework to bind them together conceptually.

A whole-of-school conceptualization of middle schooling provides such a framework. It rests on the assumption that middle schooling involves not just separate add-on programs or strategies, but a transformation of all of the critical, interactive elements of school experience, centered on critical beliefs and understandings and creating a change in the whole ecology of schooling (Hill and Russell, 1999). The whole-school design approach used in the MYRAD project illustrates this (Figure 1). A description of each design element is set out in Table 1.

Other well-known examples are the *Turning Points* 2000 design (Jackson and Davis, 2000) and the 14 middle school practices and cultural characteristics delineated by the NMSA. The elements in the former design are eight principles or recommendations, the central one being to ensure the success of every student. Others concern responsive

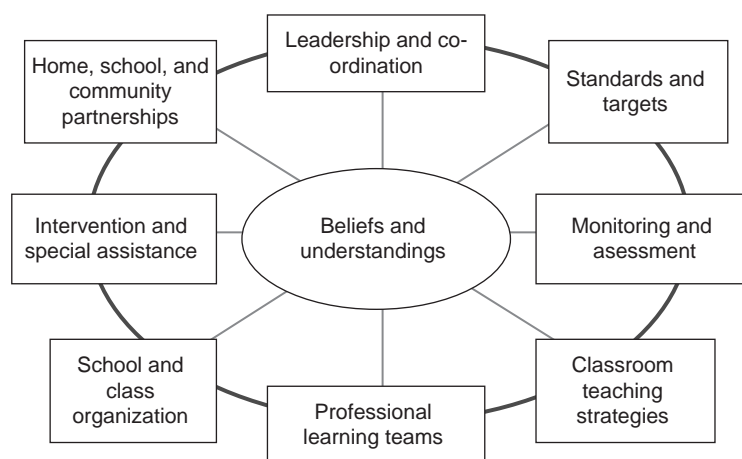


Figure 1 Whole-school design for improving learning outcomes. From Hill, P. W. and Crevola, C. A. M. (1997). *The Literacy Challenge in Australian Primary Schools*. IARTV Seminar Series No. 69. Jolimont: IARTV.

Table 1 Whole-school design elements and their description

<i>Beliefs and understandings</i>	<i>Classroom teaching strategies</i>
<p>Whole-school agreement regarding teacher belief in</p> <ul style="list-style-type: none"> professional efficacy – capacity to make a difference to student learning capacity of all students to learn, given right support and instruction role of effort (<i>cf</i> ability and demographic background) in student learning importance of not giving up on students teacher responsibility (individual and collective) for student learning importance of understanding and being responsive to the developmental needs of early adolescents <p><i>Leadership and coordination</i></p> <p>Leadership team members play a critical role in</p> <ul style="list-style-type: none"> development of a strong, positive school culture through ensuring <ul style="list-style-type: none"> a clear, agreed vision and goals for student learning high expectations of staff and students supportive, fair, open relationships collaborative, team-based professional learning for staff high quality of teaching a safe, orderly school environment constructive school policies involvement of school community members in decision-making development of a strong, student-centered instructional focus in the school by ensuring <ul style="list-style-type: none"> implementation, embedding, and monitoring of a whole-school design approach staff to coordinate design approach on a day-to-day basis establishment of clear instructional goals coordination of the curriculum across school grade levels targeting resources to meet school learning goals protection of class time from interruption; arrangement of appropriate timetabling arrangement of appropriate class size and composition encouragement of staff collaboration improvement of teacher effectiveness establishment of standards and targets <p><i>Standards and targets</i></p> <ul style="list-style-type: none"> Specific targets set against system-wide standards for the whole school, each grade level, and each class group Individual learning targets or levels of achievement set for each student <p><i>Monitoring and assessment</i></p> <ul style="list-style-type: none"> Regular, frequent, systematic, and detailed monitoring and recording of student learning progress Use of assessment to inform teaching: understanding of the individual's learning styles, difficulties, interests, and zone of proximal development Monitoring of student progress in all areas, to ensure students are seen as whole persons 	<p>Use of approaches attuned to adolescent needs:</p> <ul style="list-style-type: none"> a repertoire of strategies and use of evidence to meet individual needs authentic curriculum and work: important, meaningful, interesting, connected to students' lives and concerns inclusion of problem-based, trans-disciplinary, real-world Queensland rich tasks, fertile questions, integrative curriculum curriculum depth rather than breadth student agency and involvement in learning, curriculum and assessment decision-making stimulating, active, varied learning experiences challenging, higher-order, and critical thinking activity cooperative as well as independent learning alignment between pedagogy, curriculum, and assessment <p>Creation of a classroom climate and culture characterized by</p> <ul style="list-style-type: none"> respectful, supportive, fair, and friendly relationships; teacher knowledge of individuals clear, high expectations about work, social behavior safe, orderly environment high level of engaged learning time deep learning and mastery achievement goals <p><i>Professional learning teams</i></p> <p>Based on commitment to change and improvement of teaching practice:</p> <ul style="list-style-type: none"> professional learning in teams to increase professional collaboration focus of professional learning on deep understanding of pedagogy and practices sustained, ongoing, theoretically based but practically situated learning in-school, shared observation, reflection, and feedback with team members use of individual teacher-development plans <p><i>School and class organization</i></p> <ul style="list-style-type: none"> Small interdisciplinary teacher team/small group structure to enhance teacher–student relationship Co-location of team to enhance collaboration Team planning time Block timetabling for sustained, deep learning Stable class composition across curriculum Dedicated home room Mixed-ability classes; flexible groupings in classes <p><i>Intervention and special assistance</i></p> <ul style="list-style-type: none"> Close, evidence-based monitoring of individuals Early intervention on basis of diagnostic evidence Team monitoring and action planning for individual students <p><i>Home/school/community links</i></p> <ul style="list-style-type: none"> Close professional links between secondary and feeder primary schools to ensure shared knowledge of students, curriculum continuity, successful transition, and shared teaching approaches Interaction of home and school to ensure regular, informative school reporting, shared knowledge of students, consistent approaches to learning and behavior Cooperative school–community relationships to increase students' real-life experiences; holistic provision of community services through schools

curriculum with rigorous standards; authentic and personalized classroom pedagogy and assessment; specialist middle grade teachers who continue to learn professionally; caring relationships with students in an intellectually challenging climate; democratic school governance and evidence-based decision-making about students; a safe and healthy school environment; and a collaborative relationship among the school, home, and community.

There is much in common among whole-school designs, particularly the balancing of the two broad interactive dimensions of schooling Newman and colleagues (Newmann *et al.*, 1992) define as essential to adolescent engagement and achievement: first, school and classroom culture and relationships that enable students to feel connected to school; second, student learning experiences that are challenging, interesting, meaningful, appropriate to their needs, and nontrivial. However, the issues highlighted reflect the particular middle years context of the systems concerned. For example, there is more concern to establish mixed ability grouping and cooperative learning in the US context, where tracking is standard practice and often considered publicly to be evidence of effort for high achievement, than in Australia where tracking is less common. Specialist middle grades teacher preservice qualifications are also a greater issue in the US context, where middle schools provide targeted employment.

Reflection on the description of high-quality middle schooling leads one to question whether this approach is unique to students in the middle years. In fact, the approach appears to be characteristic of good learning, teaching, and schooling for students of any age. The emphasis on tailoring practices to meet the developmental characteristics and needs of young adolescents is unique, but then good learning and teaching are responsive to the particular developmental needs of students, irrespective of their age group. Middle schooling could well stimulate improvement of teaching and learning at other levels.

The Evidence Base

Research evidence is required with regard to two aspects of middle schooling: first, the extent to which middle schooling has been implemented and second, the effectiveness of middle schooling in improving learning outcomes. There are complexities in attempting to investigate each of these. Writers commonly remark on the paucity of middle schooling research. There is much that is simply not known.

Extent of Implementation

It would be unreasonable to expect that a whole-school approach to middle schooling could spring to life in a school fully realized, given that design elements encompass

all aspects of school life. Implementation occurs developmentally (Russell *et al.*, 2003). Research evidence is required with regard to the following: variation in the development of middle schooling achieved within schools; aspects of middle schooling that are most and least likely to develop; the best starting point for successful development; the time scale required for full implementation; and variation in quality of development.

Some evidence comes from a study in which Felner and Jackson (1997) developed an index of middle school transformation to differentiate schools with high, medium, and low implementation levels of the *Turning Points* (1989) eight design recommendations. Scoring was based on indicators of implementation quality in each design area. Variation in implementation was evident. Of the 31 schools, nine were classed as high implementers, 12 as partial implementers, and ten as having achieved little if any implementation, each group being a representative demographic mix of schools. The authors commented on the value of having a whole-school design both for comprehensive school planning and for research purposes.

The most successful developmental sequencing began with key leadership practices and staff commitment to common attitudes, beliefs, and understandings regarding middle schooling. Other elements could then be established effectively on this basis. Much is yet to be discovered about how elements interact, for example, how the presence, absence, or intensity of element implementation affects development of other elements. It was noted that implementation takes substantial time – nearly 3 years for processes to be embedded. Thus, for a middle school program encompassing grades 6–8, it would take nearly 5 years before a cohort of students would have access to a full middle schooling experience.

An Australian study of primary, secondary, or P–12 schools engaged in middle schooling by Pendergast *et al.* (2005) concluded that there are three phases in the implementation of middle schooling reform: an initiation phase (1–2 years), a developmental phase (2–5 years), and a consolidation phase (5–10 years). The report emphasized that school reform takes much longer than usually allowed for. Pendergast and colleagues endorsed the view that leadership innovation and an agreed school vision for learning are critical variables. Enhanced pedagogies, especially the provision of greater intellectual challenge in classroom work, were deemed vital for progression from the developmental to the consolidation phase, thus reinforcing the present understanding that improving actual teaching and learning approaches is the most important strategy to implement, although the one most difficult to achieve.

What is the norm in the middle years of schooling these days? While case studies and extended experience with reform programs (Haycock and Ames, 2000), as well as systematic research, point to high-quality middle schooling being achievable, there is very little evidence

to indicate whether the fully realized middle schooling approach is widespread within interested education systems, even in systems that provide policy, targeted funding, and other systemic support.

Effectiveness

Evidence of the effectiveness of middle schooling is essential. Does it enable the range of current goals of schooling to be met better than a traditional approach? There is research evidence of the effectiveness of component processes of middle schooling. For example, leadership teams' school improvement practices, especially their vision for improved student learning in the school, are linked positively to teachers' perception of the quality of their own professional learning culture (Russell *et al.*, 2003). Evidence-based professional learning challenges and changes teachers' attributional beliefs about learning, increasing teachers' acceptance of responsibility for student learning and their preparedness to intervene (Timperley and Robinson, 2001). Students are more engaged and use better, in-depth learning strategies within classroom cultures characterized by task-focused mastery goals rather than ability-focused performance achievement goals (Meece *et al.*, 2006). Multiple affective and cognitive aspects of classroom learning environments are significantly predictive of student behavior, level of student task participation, and quality of thinking (Matsumura *et al.*, 2008).

Research evidence is also available with regard to individual signifying practices of middle schooling, such as teacher teaming, block or flexible timetabling, problem-based and transdisciplinary curriculum, and authentic pedagogy and assessment. Summaries of research on such middle schooling practices have been published (e.g., Anfara, 2001), while evaluations of large-scale middle schooling initiatives provide useful evidence (Mertens *et al.*, 1998), and design-based approaches to middle schooling such as *Turning Points 2000* (Jackson and Davis, 2000) use research evidence to underpin the practices they advocate.

There is less evidence, however, with regard to the effects of an integrated, whole-school design approach to middle schooling. More is needed. Felner and Jackson (1997), in the study already outlined, examined results cross-sectionally and longitudinally. Differences based on schools' level of implementation of middle schooling were found in sixth- and eighth-grade student achievement levels in reading, mathematics and language arts, as well as in teacher-rated student behavior, and student self-reported measures of self-esteem and affect. Longitudinal analysis (although over a 2-year period only) showed that gains increased as schools moved to higher levels of implementation and students had longer exposure to middle schooling approaches.

There is much to be done to improve research into middle schooling. The strongest evidence achieved with

methodological rigor is at the micro level. Research at the greater macro level is mainly qualitative, individualistic, nonreplicative in nature, and inconclusive in its results. Sound research designs, valid operational definitions of middle years practices encompassing the complexities involved (e.g., team teaching), joint use of quantitative and qualitative approaches together with more sophisticated forms of data analysis, 5–10-year longitudinal studies, and use of valid measures of complex student outcomes would enhance the evidence base considerably. Too often effectiveness is still evaluated in terms of standardized literacy and numeracy achievement test results, disregarding other skills, knowledge, and dispositions necessary for life in the twenty-first-century knowledge society. Use of inappropriate outcome measures can make nonsense of research in this area.

Concluding Comments

The perspectives on schooling in the middle years discussed in this article point to the complexity of the issues involved in designing, implementing, and evaluating purposeful, systematic, and appropriate education for young adolescents. Consideration has been given to the following issues: factors that have stimulated interest in the middle years in recent decades; the two major orientations to the middle years of schooling, one focusing on structure and the other on practice; the characteristics of high-quality middle schooling; and the quality of evidence concerning the extent to which middle schooling is being implemented and its effectiveness assessed.

Middle schooling implementation constitutes a subset of the more general areas of school effectiveness and school improvement research; much that is instructive for middle schooling can be gained from the study of this research. Characteristic approaches to middle schooling are, in turn, highly relevant to the achievement of effective school improvement. Luke *et al.* (2003: 139) advised that we should be "mainstreaming the middle." Three interpretations of this maxim are proposed. First, implementation of middle schooling, when undertaken by a school, should be fully realized, systematically transforming the whole school over time. Second, such a middle schooling approach should be implemented in all schools educating young adolescents. Third, the essential concepts and design of middle schooling should be implemented in age-appropriate ways at all levels of schooling, not just in the middle years.

If there is to be success in achieving the complex goals of learning demanded by a globalized knowledge society, if young people are to be enabled to develop the capacity to adapt effectively to rapid, continuing, and pervasive change, if students are to be helped to become motivated and able lifelong learners, then it is essential that adolescents'

experience of schooling in the middle years be a vital one, continuing and expanding their learning capacities, motivations, and satisfactions seamlessly from their early years through to their final years of schooling. The concepts underpinning high-quality middle schooling, in essence, concern excellent teaching and learning. Because of this, the implementation of such an approach to middle schooling has the capacity, in turn, to impact on, improve, and transform schooling at all age levels.

See also: Achievement Goal Theory: Definitions, Correlates, and Unresolved Questions; Globalization and Curriculum; Interest; Leadership for Learning; Professional Learning Community; School Development for Teacher Learning and Change; Sustainable Educational Reform.

Bibliography

- Anfara, V. A. (ed.) (2001). *The Handbook of Research in Middle Level Education*. Greenwich, CT: Information Age.
- Barber, M. (1996). *The Learning Game*. London: Victor Gollancz.
- Barber, M. (1999). Taking the tide at the flood: Transforming education in the middle years. *Paper Presented at the Middle Years of Schooling Conference*, Melbourne.
- Carnegie Council on Adolescent Development (1989). Turning points: Preparing American youth for the 21st century. *The Report of the Task force on Education of Young Adolescents*. New York: Carnegie Corporation of New York.
- Cuban, L. (1992). What happens to reforms that last? The case of the junior high school. *American Educational Research Journal* **29**(2), 227–251.
- Felner, R. D. and Jackson, A. W. (1997). The impact of school reform for the middle years: Longitudinal study of a network engaged in *Turning Points*-based comprehensive school transformation. *Phi Delta Kappan* **78**(7), 528–542.
- Galton, M., Gray, J., and Ruddock, J. (1999). *The Impact of School Transitions and Transfers on Pupil Progress and Attainment*. London: DfEE.
- Greenaway, E. (1999). *Lower Secondary Education: An International Comparison*. London: International Review of Curriculum and Assessment Frameworks.
- Haycock, K. and Ames, N. (2000). Where are we now? Taking stock of middle grades education. *Paper Presented at the National Educational Research Policy and Priorities Board's Conference on Curriculum, Instruction and Assessment in the Middle Grades: Linking Research and Practice*. Washington, DC.
- Hill, P. W. and Crevola, C. A. M. (1997). *The Literacy Challenge in Australian Primary Schools*. IARTV Seminar Series No. 69. Jolimon: IARTV.
- Hill, P. W., Rowe, K. J., and Holmes-Smith, P. (1996). *The Victorian Quality Schools Project: A study of school and teacher effectiveness*. Report to the Australian Research Council, Vol. 1. Centre for Applied Educational Research, Faculty of Education, The University of Melbourne, Melbourne.
- Hill, P. W. and Russell, V. J. (1999). Systemic, whole-school reform of the middle years of schooling. In Bosker, R. J., Creemers, B. P. M., and Stringfield, S. (eds.) *Enhancing Educational Excellence, Equity and Efficiency*, pp 167–196. Dordrecht: Kluwer.
- Hirsch, D. (1994). *Schooling for the Middle Years: Developments in Eight European Countries*. (No. ED383471). Washington, DC: Carnegie Council on Adolescent Development.
- Jackson, A. W. and Davis, G. A. (2000). *Turning Points 2000: Educating Adolescents in the 21st Century*. New York: Teachers College Press.
- Lingard, B., Ladwig, J., Luke, A., et al. (2001). *Queensland School Reform Longitudinal Study: Final Report*. Brisbane: Education Queensland.
- Luke, A., Elkins, J., Weir, K., et al. (2003). *Beyond the Middle: A Report about Literacy and Numeracy Development of Target Group Students in the Middle Years of Schooling*. Canberra: Commonwealth Department of Education, Science and Training.
- Matsumura, L. C., Slater, S. C., and Crosson, A. (2008). Classroom climate, rigorous instruction and curriculum, and students' interactions in urban middle schools. *Elementary School Journal* **108**(4), 293–312.
- Meece, J. L., Anderman, E. M., and Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology* **57**, 487–503.
- Mertens, S. B., Flowers, N., and Mulhall, P. F. (1998). The Middle start initiative phase 1: A longitudinal analysis of Michigan middle-level schools. *The Center for Prevention Research and Development*. University of Illinois.
- Newmann, F. M., Wehlage, G. G., and Lamborn, S. D. (1992). The significance and sources of student engagement. In Newmann, F. M. (ed.) *Student Engagement and Achievement in American Schools*, pp 11–39. New York: Teachers College Press.
- NMSA (2003). This we believe: Successful schools for young adolescents: A summary of a position paper. <http://www.nmsa.org/> (accessed August 2009).
- Pendergast, D., Flanagan, R., Land, R., et al. (2005). *Developing Lifelong Learners in the Middle Years of Schooling*. Canberra: Ministerial Council on Education, Employment, Training and Youth Affairs.
- Russell, J., Mackay, T., and Jane, G. (2003). *Messages from MYRAD: Improving the Middle Years of Schooling*. Jolimon: IARTV.
- Steinberg, L. and Lerner, R. M. (2004). The scientific study of adolescence: A brief history. *Journal of Early Adolescence* **24**(1), 45–54.
- Timperley, H. S. and Robinson, V. M. J. (2001). Achieving school improvement through challenging and changing teachers' schema. *Journal of Educational Change* **2**(4), 281–300.
- Whitby, K., Lord, P., O'Donnell, S., and Grayson, H. (2006). *Dips in Performance and Motivation: A Purely English Perspective?* London: National Foundation for Educational Research.

Further Reading

- Australian Curriculum Studies Association (1996). *From Alienation to Engagement: Opportunities for Reform in the Middle Years of Schooling*. Canberra: National Advisory Committee for the Student Alienation During the Middle Years of Schooling Project.
- Barratt, R. (1998). Shaping middle schooling in Australia: *A Report of the National Middle Schooling Project*. Canberra: Australian Curriculum Studies Association.
- Chadbourne, R. (2001). *Middle Schooling for the Middle Years: What Might the Jury be Considering?* Melbourne: Australian Education Union. <http://www.aeufederal.org.au/Publications/Middleschooling.pdf> (accessed August 2009).
- Davis, G. A. (2001). Point to point: Turning points to turning points 2000. In Anfara, V. A. (ed.) *The Handbook of Research in Middle Level Education*. Greenwich, CT: IAP.
- Department of Education and Training, Australian Capital Territory (2005). *Teaching and Learning in the Middle Years in the ACT: A Study to Support Schools meet the Learning Needs of Adolescent Students*. Canberra. http://www.det.act.gov.au/__data/assets/pdf_file/0010/17992/middle_years_pts1_2_3_4.pdf (accessed August 2009).
- Eyers, V., Cormack, P., and Barratt, R. (1992). The education of young adolescents in South Australian government schools: *Report of the Junior Secondary Review*. Adelaide: Education Department of South Australia.
- International Baccalaureate Organization (2002). *Middle Years Programme. A Basis for Practice*. Geneva. <http://www.ibo.org> (accessed August 2009).
- lisahunter (ed.) (2007). *Machinations in the middle. Special Issue on the Middle Years of Schooling Australian Educational Researcher* **34**(2), 1–6.

- Petzko, V. N. (2004). Findings and implications of the NASSP National Study of Leadership in middle level schools, volumes 1 and 11: Teachers in middle level schools. *NASSP Bulletin* 88(638), 69–88.
- Seashore Louis, K. (2000). Teachers' professional development for vital middle schools: What do we know and where should we go? *Plenary Address to the National Educational Research Policy and Priorities Board's Conference on Curriculum, Instruction, and Assessment in the Middle Grades: Linking Research and Practice*. Washington, DC.
- Takanishi, R. and Hamburg, D. A. (eds.) (1997). *Preparing Adolescents for the Twenty-First Century: Challenges Facing Europe and the United States*. New York: Cambridge University Press.
- Thompson, S. C. (ed.) (2004). *Reforming Middle Level Education: Considerations for policymakers*. Greenwich, CT: Information Age.

Relevant Websites

- <http://www.acsa.edu.au> – Australian Curriculum Studies Association.
- <http://www.cprd.uiuc.edu> – Center for Prevention Research and Development.
- <http://www.inca.org.uk/> – International Review of Curriculum and Assessment Frameworks.
- <http://www.dcsf.gov.uk> – Key Stage 3, Department for Education and Skills, UK.
- <http://www.middleweb.com> – MiddleWeb.
- <http://education.qld.gov.au> – Middle Years of Schooling, Education Queensland.
- <http://www.nmsa.org> – National Middle Schools Association.
- <http://www.turningpts.org> – Turning Points.

Reform of the School Workforce

H M Gunter, University of Manchester, Manchester, UK

G Forrester, Liverpool John Moores University, Liverpool, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The status, workload, and expertise of teachers in publicly funded education within Western-style democracies remain a salient political issue. Reports identify common issues regarding heavy workload (e.g., Dibbon, 2004; Gardner and Williamson, 2005) affecting recruitment and retention (e.g., Smithers and Robinson, 2003). It seems that the advent of site-based decision making has led to increased demands for performance combined with public audits. The Organization for Economic Cooperation and Development (OECD) recognizes the importance of confronting supply and demand issues, not least because “teacher quality is a critical factor in determining student learning” (Santiago, 2002: 8).

The emerging policy strategy to secure and sustain quality, that is, in Ozga’s (2005) terms, traveling around the world, is the simultaneous incorporation of teachers as members of a diverse workforce combined with regulation through centralized performance leadership regimes. In schools in England those with qualified teacher status (QTS) are becoming the smallest group in many schools (Butt and Gunter, 2007) and are facing controls through inspections and the delivery of national reforms through local leadership roles. How such reforms are engaged with, locally, does depend on what Ozga calls embedded cultures and practices. For example, in Scotland Ozga (2005) emphasizes the professional role of teachers in comparison with the use of teacher assistants who teach in England; in South Africa managerialist strategies are in tension with the aim to build democracy (Berkhout, 2007); whereas in Australia (Vidovich, 2007) and New Zealand (Fitzgerald, 2007) policy borrowing from England, while highly problematic, seems to be normalized. While the demands of globalization seem to suggest that workforce reform is a traveling and totalizing discourse about teachers and their work, with demands for a flexible, accountable workforce which follows good practice, in reality there are local histories, cultures, and value systems that shape how the debate is being constructed and played out. Our intention is to focus on England in order to show the complexities of the origins and development of a particular approach to workforce reform. In this way we intend to make a contribution to problematizing the promotion of what Barber (2007), as the architect of New Labour education reform, called a world-class education system.

England: A Study in Rapid Reform

Over the two decades since the 1988 Education Reform Act a form of site-based performance leadership emerged, which combines the centralized direction of the purposes of schools and the curriculum with the local responsibility and accountability for delivery (see Gunter, 2008). Policy texts show that a central concern for successive governments has been to improve national standards by: (1) sufficiency of workers through recruitment and retention (Morris, 2001); (2) the credentials and the requirement to have qualified teacher status (QTS) (DfES/PricewaterhouseCoopers, 2007); (3) a workforce trained to implement national priorities and reforms (DfES, 2004a); and (4) a workforce which can be performance managed through local leadership of national reforms (DfEE, 1998).

Educational standards dominated government strategies from the 1970s onward, with Sir Keith Joseph in 1984 telling the North of England Conference that “this Government’s aim is to raise standards” (Joseph, 1984: 138). Whereas Joseph talked about the need for “open discussion” (p. 140) New Labour in 1997 presented a non-negotiable standards agenda where “each school will have its own challenging targets to raise standards, and will be held responsible for achieving them” (DfEE, 1997: 6). The approach was based on what Barber (2007) describes as deliverology where “... in one way or another, the minister can influence what happens inside the head of an eleven-year-old in, for example, Widnes” (p. 85). This regulated and audited standards regime needed a workforce that would operate as one link within a delivery chain (p. 85), and so New Labour reworked the concept of teaching professionalism in two main ways: (1) through remodeling the school workforce; and (2) through remodeling head-teacher work, and subsequently all teachers’ work, as a form of performance leadership.

Remodelling the School Workforce

Remodelling the School Workforce was launched from January 2003 with the National Agreement (DfES, 2003) between the government, employers, and unions (except the National Union of Teachers). A social partnership was formed so that workforce planning in relation to government policy priorities could take place. What had to be resolved was that site-based performance leadership

had made teaching an unattractive job, with reports of a working week over 50 h (PwC, 2001; DfES, 2000; Thomas *et al.*, 2004). The General Teaching Council (GTC) reported that “one in three teachers expect to leave the profession within five years, protesting about workload, government interference and poor pupil behaviour” (Woodward, 2003: 1). Remodeling was meant to handle work overload and the poor image of the job in two main ways: (1) by defining 25 tasks that teachers should not normally do such as routine clerical work, cover, and examination invigilation; and (2) increasing the status of teachers by constructing their role as leading non-QTS members of the workforce.

Remodeling was phased in from September 2003 (ahead of the evidence from the government’s own pilot project evaluation, see Gunter and Butt, 2007a, b), with requirements, for example, to limit teachers covering for absent teachers, to use the wider workforce to invigilate examinations, and to guarantee time for planning, preparation, and assessment. There is some evidence that it has worked well in places with gains made in school by moving certain types of work from teachers to the bursar, site manager, teaching and administrative assistants (see Gunter *et al.*, 2005; Butt and Gunter, 2007; Higham *et al.*, 2007). Indeed many schools were ahead of the process because site-based management from 1988 was premised on the school being able to identify curriculum need and plan the type and deployment of the workforce. The growth of information and communications technology (ICT) has created opportunities for efficiency, and the workforce has been encouraged to access teaching resources and lesson plans from the Internet in order to cut workload. However, there is also evidence that the claims for increasing the status of teachers as managers of learning through the use of ICT, personalized learning, and teaching assistants has not been based on understandings about professional knowledge, identity, and the reality of the job (Gunter, 2007). In particular, there are issues about the growing role of non-QTS members of the workforce taking over professional work such as teaching, assessment, and pastoral care, with challenges to the assumption that the design of teaching (by teachers and/or through resources obtained via the Internet or government curriculum strategy packages) can be separated from its delivery (Edwards, 2007; Gunter and Butt, 2007a, 2007b; Gunter, 2007; Yarker, 2005).

Since September 2003 there has been no requirement for a headteacher to have QTS. Indeed, the DfES/Price-waterhouseCoopers (2007) report regards the credentializing of leadership in this way as a form of monopoly, and so the top job (e.g., chief executive) should be open to anyone with proven leadership credentials. Plans have been announced to cut the time needed to train as a teacher to 6 months in order to facilitate mature entrants

(BBC, 2009). The role of the school as an isolated provider of education is undergoing change through the Every Child Matters (DfES, 2004b) agenda where the emphasis is on integrated services including education, health, welfare, police, and community services. The label of workforce to encompass teachers together with clerical, teaching assistant, and bursar roles can be easily extended to embrace a wider group of public workers (e.g., police, librarians, welfare officers, doctors, and nurses) who provide local services (see Gunter and Rayner, 2007).

Performance Leadership

Another connected intervention into workforce practice has been through the promotion and delivery of performance leadership. The comprehensivization of schools from the 1960s created an imperative for school leadership, where larger schools (over 1000 students) and a mixed ability student intake, meant that more attention had to be given to organizational arrangements. The 1988 Education Reform Act intensified the need for organizational leadership in order to deal with the budget and staffing due to competition based on open enrolment within a quasi-market. From the early 1990s onward the emphasis on performance leadership through the publication of output data, inspections, and performance-related pay meant that systems for the gathering, analysis, and use of data to examine individual student, teacher, and school compliance with national standards had to be developed. New Labour from 1997 tightened this performance regime through a focus on the headteacher as leader who must deliver national reforms (Gunter and Forrester, 2008). In *Excellence in Schools* (DfEE, 1997) policy is explicit:

The vision for learning set out in this white paper will demand the highest qualities of leadership and management from headteachers. The quality of the head often makes the difference between the success or failure of a school. Good heads can transform a school; poor heads can block progress and achievement. It is essential that we have measures in place to strengthen the skills of all new and serving heads. (p. 46)

New Labour drew on school improvement and effectiveness research to increase the status of the head as effective leader, connected pay to whole school performance, and provided a model of leadership that emphasized the transformational agency of the single charismatic and inspirational leader who could build commitment to the localized implementation of national curriculum strategies for literacy and numeracy. Such investment was symbolic in the form of attention given to headteachers through speeches and conferences, and financial in the form of pay and investment in training.

The work of headteachers was redesigned around national standards based on the appropriate attributes, skills, and behaviors of effective leaders, and largely drawn from private-sector models and produced by private-sector consultants. Interconnected with this is the training agenda, were the programs set up by the previous Conservative Governments (e.g., National Professional Qualification for Headship) were incorporated into a national training framework housed within a National College for School Leadership (NCSL) from 2000. The NCSL is a nondepartmental public body and as such it enables centralized control to operate through a remit to deliver on national policy priorities, together with clear accountability mechanisms for supplying the next generation of school leaders (Gunter and Forrester, 2009).

The NCSL scoped the field of knowledge and research, and presented a model of effective school leadership (Hopkins, 2001), and through training, research, and conferences, has sought to structure the practice and performance of school leaders. There have been successes in regard to access to training and the production of resources to support the NCSL role in reform implementation, but there have also been concerns about its role in colonizing the field (see Gunter and Fitzgerald, 2008; Weindling, 2004). Furthermore, there has been increased recognition that the transformational leader model was causing additional workload, and was unattractive to many headteachers who did not see themselves as charismatic visionaries. Notions of distributed and shared leadership have emerged, in order to ensure that middle leaders, teacher leaders, and the wider workforce (e.g., bursar/financial manager) are directly involved, and consistent with remodeling, in the delivery of outputs. The lack of research evidence and conceptual clarity of distributed leadership has not prevented it from being advocated, and what is emerging is a form of distribution that is more about job redesign, or even a form of localized steering where the headteacher is required to manage the risk of implementation (Gunter, 2005). Indeed, as Hartley (2007) has shown, distribution to the school is in the form of tactics or how to implement within context, and is not about distributing decisions about strategy because policy control remains centrally regulated.

Challenges also remain in regard to headteacher recruitment, and currently the NCSL is working on the delivery of the DfES/PricewaterhouseCoopers (2007) shift from effective school and headteacher leadership toward effective executive leadership for the local delivery of children's services. What remains a constant in New Labour strategy has been the single person who delivers locally, but what is changing is the type of person who might inhabit that role. Remodeling leadership means that an expert in teaching and learning may not be the chief executive, and so the language and approach is about leadership as a generic task and set of behaviors which

others from the public, voluntary, and private sectors can do. The endurance of the single leader who is responsible and accountable for delivery is a feature across the public and private sectors, and so the culture and practice in education and children's services is congenial with such interagency and intersector movement. As Ball (2007) has shown, the delivery of public services through and with private-sector roles, practices, and cultures is now endemic within education and so non-QTS appointments making educational decisions has been normalized. A final point on this is that as Coffield *et al.* (2007) have shown the recent reform strategy (Prime Minister's Strategy Unit, 2006) is about the need to encourage bottom-up market responses and provision through parents and the community. Within the advocated approach the term used is leadership, and public-sector experts such as teachers are absent (Fitzgerald and Gunter, 2009).

Explanations for Workforce Reform

Following Raffo and Gunter (2008), we examine these rapid and far-reaching changes through a particular approach to understanding modernization. What is modern, new, innovative, twenty-first century, and pathfinding is empty until filled by those who have a commitment to do this (Gunter and Butt, 2007a). New Labour's approach to modernizing has shifted from job redesign (i.e., changing work from teaching and learning to organizational effectiveness) to a form of incremental upgrading (i.e., training headteachers to be effective school leaders), toward a reconceptualization of effective leadership (i.e., as generic transferable skills and behaviors). There are two ways of explaining this: (1) functional and (2) socially critical. Both recognize education, change, and improvement as essential, but both differ in terms of how the workforce is understood and engaged with.

Functional Position

The functioning of schools as effective and efficient organizations requires a workforce that is sufficient in terms of numbers and training to deliver on national and organizational goals in order to produce a world-class system (DFES, 2002). Functionalism in its current format has been captured by neoliberals who want to challenge the state and neoconservatives who want to attack socially just ideas about access to education as a public good. This alliance means the market and economic productivity is paramount in the discourse, and it is seen as means by which particular interest groups (class based, religious groups, etc.) can gain an advantage for a preferred form of schooling.

The literatures tend to challenge the state, bureaucracy, and welfare protection (e.g., Bobbitt, 2002; Chubb

and Moe, 1990; Osborne and Gaebler, 1992), where the argument is made that the workforce must be fit for purpose in a globalizing economy, is endlessly trainable, and deployable. There are no demarcation lines or claims to professional practices and knowledge, and instead there is a can-do culture where everyone is ready and capable to take on what is necessary to be successful. Barber and Mourshed (2007) argue that exclusivity and making teaching harder to get into, rather than higher pay, is what is needed to make it a high-status job. Indeed, in an education system where particular interest groups can control schools (e.g., philanthropists, religious groups) then commitment to values (e.g., sectarian beliefs, entrepreneurship) is what matters, and so teachers can be attracted to an exclusive brand, and so be trusted to implement as engaged collaborators.

Hence explanations regarding workforce reform tend to focus on teachers as the problem and the dysfunctions that are caused for children, parents, communities, and the economy, as a result. The narrative tends to go like this:

Teachers have used their professional knowledge and historical control of their work to protect themselves and to promote ideas that parents do not want. So they can make claims about teaching and learning that are not true, and they create work for themselves and each other by planning lessons, taking the register and covering lessons for absent teachers. Teachers have sought to prevent reforms, and have largely been unaccountable for what they do and the quality of the outcomes from their teaching. As a result parents and employers are concerned about levels of literacy and numeracy, the promotion of radical ideas which challenge parental control, and children are absenting themselves in large numbers. Teachers need to operate in the real world, and give better value for money, not least because they get so many holidays.

Consequently, interventions tend to be about tasks, structures, and cultures where strategic control over the work of teachers has been largely taken from them, and they now have to operate according to prescribed curriculum ring binders, assessments, and performance systems. The emphasis is on contractual compliance with targets, job descriptions, and reform priorities. Training is about enabling this delivery to be efficient and effective, and enables teachers to move from one reform initiative to another. Teachers are also expected to talk and engage in ways that show commitment and excitement about their work. In this way, teachers can also be employed by a range of educational providers, not least the new academies and trust schools where private-sector investors as sponsors can control the curriculum and the conditions of employment. Increasingly, the workforce may be recruited, trained, and developed within a particular school or academy brand based on the vision of the sponsors.

Socially Critical

This position also sees workforce reform as potentially beneficial to the workforce and to education, but claims are made that the drive for a world-class system masks inequity. Concerns are raised about the ability of functional interventions to work in the ways expected, but more importantly, dysfunctional outcomes are identified. In addition, the socially critical work position is based on values related to social justice and that schools are a public good. What is regarded as critical is twofold: (1) through confronting how existing power structures operate in ways that secure advantages for private interests (wealth and religion) (see Woods *et al.*, 2007); and, (2) through creating socially just ways of providing education and learning opportunities (Lingard *et al.*, 2003; Thomson, 2003).

Hence, explanations tend to focus on how teachers have been unfairly defined as the main problem. The consequences are that functionalism cannot operate in the ways intended because performance leadership creates the low morale and failure to recruit and retain that it claims to solve. Socially critical work therefore focuses on how teachers can be enabled to take more control of the purposes of education, to recognize how the alliance of neoliberal and neoconservative functional interventions and narratives position them and hence enable them to work for a more socially just approach to professionalism (Smyth, 2003). The narrative tends to go like this:

Teachers are unable to develop their professional contribution to teaching and learning, and to lead on school improvement because they are excluded from decisions about the purposes and practices of education. Consequently teachers are positioned in contradictory settings where they have to implement functional tasks such as targets and measure their valued added but the socio-economic context in which the school is located impacts on their ability to do this. Children and parents do not enter willingly into such processes, and so the relevance of such structures to their lives remains problematic. Teachers should be in a position to develop their professionalism through working with students and parents to develop the curriculum through community based action, resources, and seeking to improve the learning environment. Teacher development isn't separate from student involvement and the starting point for learning is to begin with students' interests so that capacity to learn is through building a project. This means that teacher professionalism needs to embrace student participation in the design, experience and assessment of learning.

Consequently, interventions in practice is through the professional development of teachers as researchers of their own and other's practice, with postgraduate study and reading as central to the ownership of the evidence

base. The generation of innovative pedagogic practices is therefore based on the interplay between agreed educational priorities, research, and local innovation (see Lingard *et al.*, 2003) and raising the status of children as active participants in the design of learning (see Smyth, 2006). This is located in both re-imagining schools and schooling with a commitment to education as a public good and the nation's children, where disadvantage is confronted and more socially just practices worked for.

Summary

The English experience of reform shows the reworking of the public provision of education through constructing small businesses where functional interventions are made in the composition, deployment, and accountability of workers. Tactical gains have been secured at the local level through more efficient and effective deployment of the workforce, with new ways of working being adopted. Leadership has been conceptualized, trained, and promoted as the means to secure change. The single leader as chief executive together with a remodeled workforce in a redesigned structure can deliver national reforms locally. While claims can be made for the outputs of delivery, and the fear of not delivering can be all consuming (see Barber, 2007), there remain concerns. First, there is a lack of reliable and accurate data regarding teacher supply and retention, which means that interventions may not be evidence based (Santiago, 2002); second, reforms are not evidence based (e.g., remodeling went ahead before the government's own pilot was completed) and so there is a combination of conviction politics with reforms based on functional beliefs combined with selective and *post hoc* evidence construction taking place; third, there are problems in the system that are not being directly engaged with, not least the issue of social justice representations in the workforce such as gender and ethnicity, which means that while the system functions it may be replicating disadvantage (e.g., Gunter *et al.*, 2005); fourth, reform is a product of elite adults outside of the school, making interventions into the practice of adults in school, and it does not begin with learners and learning, which means that those who are meant to be the beneficiaries of reform are excluded from the process in ways that are affecting the well-being and achievements of students (e.g., Smyth, 2006); fifth, the cultures of reform are based on caring about functional interventions in task completion and contractual compliance, and this is at odds with teacher-workforce identities where caring for children is more important than caring about paperwork (see Forrester, 2005); and, finally, functionalism is based on what tasks teachers should no longer do (e.g., photocopying) and it is not based on a defensible model of

practice that is rooted in pedagogic practices and expertise about learning (see Wilkinson, 2005).

Socially critical approaches can reveal the embedded nature of how localities respond to globalized good practice, and as Stevenson (2007) has shown there are big issues such as the graduate status of teaching that need to be confronted by teachers, otherwise they will disappear, like QTS for headteachers, without comment. Such issues can be confronted not only by defending graduate status but by providing an alternative approach to how workforce reform might be engaged with. Research has already produced interesting studies into rethinking school improvement (Hollins *et al.*, 2006; Thomson and Gunter, 2006) and the role of teachers in the development of their work and their identity (Gunter, 2005; Sachs, 2003). When combined with the growing international evidence about how children can participate in school reform (Gunter and Thomson, 2007; Smyth, 2006), meaningful opportunities for change can be opened up. If teachers have been an absent presence (Ball, 2003) in functional workforce reform then they are in danger of becoming just absent through the remodeling process in England. What is happening is what Thomson and Blackmore (2006) identify as work redesign “without a coherent and meaningful ethical and political (spiritual and moral perhaps) purpose” (p. 175). The irony is that the success of functionalism has been to create the conditions in which this has happened and through which teachers have complied. Evidence shows that school leaders can read policies differently, challenge and create site-based policies (Forrester and Gunter, 2009), and that there are local models in play where leadership is not assumed to be a function of a role but is communal (Grubb and Flessa, 2006). Hence socially critical work is taking place, and so the potential exists for such activism to not only challenge functionalism but also to recreate different narratives and practices.

See also: An Overview of Teacher Labor Markets; Evaluating Education in Three Policy Eras; History of Educational Leadership/Management; Teacher Supply.

Bibliography

- Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of Education Policy* 18(2), 215–228.
- Ball, S. J. (2007). *Education PLC*. London: Routledge.
- Barber, M. (2007). *Instruction to Deliver. Tony Blair, Public Services and the Challenge of Achieving Targets*. London: Politico's.
- Barber, M. and Mourshed, M. (2007). *How the World's Best-Performing School Systems Come Out On Top*. London: McKinsey and Company.
- BBC (2009). *Six Months Training Plan*. www.news.bbc.co.uk/1/hi/education/7933690/stm.
- Berkhout, S. (2007). Democratization and the remaking of teachers in South African schools. In Butt, G. and Gunter, H. M. (eds.) *Modernizing Schools: People, Learning and Organizations*, pp 149–162. London: Continuum.
- Bobbitt, P. (2002). *The Shield of Achilles*. London: Penguin.

- Butt, G. and Gunter, H. M. (eds.) (2007). *Modernizing Schools: People, Learning and Organizations*. London: Continuum.
- Chubb, J. E. and Moe, T. M. (1990). *Politics, Markets and America's Schools*. Washington, DC: Brookings Institute.
- Coffield, F., Steer, R., Allen, R., et al. (2007). *Public Sector Reform, Principals for Improving the Education System*. London: Bedford Way Papers, Institute of Education, University of London.
- DfEE (1997). *Excellence in Schools*. (Cm 2681) London: The Stationary Office.
- DfEE (1998). *Teachers: Meeting the Challenge of Change*. (Cm 4164) London: DfEE.
- DfES (2000). *The Ninth Report of the School Teachers' Review Body*. London: STRB.
- DfES (2002). *Time for Standards: Reforming the School Workforce*. London: DfES.
- DfES (2003). *Raising Standards and Tackling Workload: A National Agreement*. London: DfES.
- DfES (2004a). *National Standards for Headteachers*. London: DfES.
- DfES (2004b). *Every Child Matters: Change for Children*. London: HMSO.
- DfES/PriceWaterhouseCoopers (2007). *Independent Study into School Leadership*. London: DfES.
- Dibbon, D. (2004). It's About Time!! *A Report on the Impact of Workload on Teachers and Students*. Newfoundland: Memorial University of Newfoundland.
- Edwards, C. (2007). "'I can't do maths.' 'You'll be fine' said the head" *The Guardian*, 4th September. The author's wrote this account using a pseudonym.
- Fitzgerald, T. (2007). Remodelling schools and schooling, teachers and teaching: A New Zealand perspective. In Butt, G. and Gunter, H. M. (eds.) *Modernizing Schools: People, Learning and Organizations*, pp 163–175. London: Continuum.
- Fitzgerald, T. and Gunter, H. M. (2009). Contesting the orthodoxy of teacher leadership. *International Journal of Leadership in Education* 11(4), 331–340.
- Forrester, G. (2005). All in a day's work: Primary teachers 'performing' and 'caring'. *Gender and Education* 17(3), 271–287.
- Forrester, G. and Gunter, H. M. (2009). School Leaders: Meeting the challenge of change. In Chapman, C. and Gunter, H. M. (eds.) *Radical Reforms: Public Policy and a Decade of Educational Reform*, pp 67–79. London: Routledge.
- Gardner, C. and Williamson, J. (2005). Teaching is the not the job it used to be. *Paper Presented to the American Educational Research Association Conference*. Montreal.
- Grubb, W. N. and Flessa, J. J. (2006). "A job too big for one" multiple principals and other non-traditional approaches to school leadership. *Educational Administration Quarterly* 42(4), 518–550.
- Gunter, H. M. (2005). *Leading Teachers*. London: Continuum.
- Gunter, H. M. (2007). Remodelling the school workforce in England: A study in tyranny? *Journal of Critical Education Policy Studies* 5(1). www.jceps.com (accessed August 2009).
- Gunter, H. M. (2008). Policy and workforce reform in England. Special Edition of *Educational Management Administration and Leadership*. ERA 1988: from local management to privatisation. *A Critical Review* 36(2), 253–270.
- Gunter, H. M. and Butt, G. (2007a). The challenges of modernization. In Butt, G. and Gunter, H. M. (eds.) *Modernizing Schools: People, Learning and Organizations*, pp 1–6. London: Continuum.
- Gunter, H. M. and Butt, G. (2007b). Whither modernization? In Butt, G. and Gunter, H. M. (eds.) *Modernizing Schools: People, Learning and Organizations*. Continuum: London.
- Gunter, H. M. and Fitzgerald, T. (2008). The future of leadership research. *School Leadership and Management* 28(3), 263–280.
- Gunter, H. M. and Forrester, G. (2008). New labour and school leadership 1997–2007. *British Journal of Educational Studies* 55(2), 144–162.
- Gunter, H. M. and Forrester, G. (2009). Institutionalised governance: The case of the National College for School Leadership. *International Journal of Public Administration* 32(5), 349–369.
- Gunter, H. M. and Rayner, S. (2007). Remodelling headteachers in England: Is it the end of educational leadership? *International Electronic Journal for Leadership in Learning* 11. www.ucalgary.ca/iejll (accessed August 2009).
- Gunter, H. M., Rayner, S., Thomas, H., et al. (2005). Teachers, time and work: Findings from the evaluation of the transforming the school workforce pathfinder project. *School Leadership and Management* 25(5), 441–454.
- Gunter, H. M. and Thomson, P. (2007). But, where are the children? *Management in Education* 21(2), 23–28.
- Hartley, D. (2007). The emergence of distributed leadership in education: Why now? *British Journal of Educational Studies* 55(2), 202–214.
- Higham, R., Hopkins, D., and Ahtaridou, E. (2007). *Improving School Leadership: Country Background Effect*. London: Leadership Centre for Leadership in Learning, Institute of Education, University of London.
- Hollins, K., Gunter, H., and Thomson, P. (2006). Living improvement: A case study of a secondary school in England. *Improving Schools* 9(2), 141–152.
- Hopkins, D. (2001). Think tank. *Report to Governing Council*. NCSL: Nottingham.
- Joseph, K. (1984). Speech by the Rt Hon Sir Keith Joseph, Secretary of State for Education and Science, at the North of England Education Conference", Sheffield, on Friday 6th January 1984. *Oxford Review of Education* 10(2), 137–146.
- Lingard, B., Hayes, D., Mills, M., and Christie, P. (2003). *Leading Learning*. Maidenhead: Open University Press.
- Morris, E. (2001). *Professionalism and Trust, Social Market Foundation Speech*. London: DfES.
- Osborne, D. and Gaebler, T. (1992). *Reinventing Government*. Reading, MA: Addison-Wesley.
- Ozga, J. (2005). Modernizing the education workforce: A perspective from Scotland. *Educational Review* 57(2), 207–219.
- Prime Minister's Strategy Unit (2006). *The UK Government's Approach to Public Service Reform: A Discussion Paper*. London: Cabinet Office.
- PwC (2001). *Teacher Workload Study*. London: DfES.
- Raffo, C. and Gunter, H. M. (2008). Leading schools to promote social inclusion: Developing a conceptual framework for analysing research, policy and practice. *Journal of Education Policy* 23(4), 363–380.
- Sachs, J. (2003). *The Activist Teaching Profession*. Buckingham: Open University Press.
- Santiago, P. (2002). Teacher demand and supply: Improving teaching quality and addressing teacher shortages. *OECD Education Working Paper No. 1*. Paris: OECD.
- Smithers, A. and Robinson, P. (2003). Factors affecting teachers' decisions to leave the profession. *Research Report RR 430*. London: DfES.
- Smyth, J. (2003). Undamaging 'damaged' teachers: An antidote to the 'self managing school'. *Delta* 55(1 and 2), 15–42.
- Smyth, J. (2006). Educational leadership that fosters 'student voice' *International Journal of Leadership in Education* 9(4), 279–284.
- Stevenson, H. (2007). Restructuring teachers' work and trade union responses in England: Bargaining for change? *American Educational Research Journal* 44(2), 224–251.
- Thomas, H., Butt, G., Fielding, A., et al. (2004). The evaluation of the transforming the school workforce pathfinder project. *Research Report 541*. London: DfES.
- Thomson, P. (2003). *Schooling the Rustbelt Kids: Making the Difference in Changing Times*. Sydney: Allen and Unwin.
- Thomson, P. and Blackmore, J. (2005). Beyond the power of one: Redesigning the work of school principals. *Journal of Educational Change* 7, 161–177.
- Thomson, P. and Gunter, H. M. (2006). From 'consulting pupils' to 'pupils as researchers': A situated case narrative. *British Educational Research Journal* 32(6), 839–856.
- Vidovich, L. (2007). Navigating 'global' modernization policies in education: Responses from Australia. In Butt, G. and Gunter, H. M. (eds.) *Modernizing Schools: People, Learning and Organizations*, pp 189–202. London: Continuum.
- Weindling, D. (2004). *Funding for Research on School Leadership*. Nottingham: NCSL. www.ncsl.org.uk (accessed August 2009).
- Wilkinson, G. (2005). Workforce remodelling and formal knowledge: The erosion of teachers' professional jurisdiction in English schools. *School Leadership and Management* 25(5), 421–439.

- Woods, P., Woods, G., and Gunter, H. M. (2007). Academy schools and entrepreneurialism in education. *Journal of Education Policy* **22**(2), 263–285.
- Woodward, W. (2003). A third of teachers plan to quit. *Guardian*. 07/01/03.
- Yarker, P. (2005). On not being a teacher: The professional and personal costs of workforce remodelling. *Forum* **47**(2 and 3), 169–174.

Further Reading

- Chapman, C. and Gunter, H. M. (2009). *Radical Reforms: Perspectives on an Era of Educational Change*. London: Routledge.
- Gunter, H. M. and Forrester, G. (2010). School leadership and policymaking in England. *Policy Studies* **31**(1), 495–511.

Secondary Education in Developing Countries

P Hughes, Australian National University, Canberra, ACT, Australia

A Walker, Hong Kong Institute of Education, Hong Kong SAR, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

Glossary

BEFA – Basic education for all.

Declaration of Human Rights – The UN declaration in 1945.

The Delors Report – The International Commission on Education for the Twenty-First Century Report.

Education for All (EFA) – The UNESCO program scheduled for achievement by 2015.

MGD, The Millennium Development Goals – The eight goals accepted by the UN for achievement by 2015.

TVET – Technical and Vocational Education and Training.

UN – The United Nations Organization.

UNESCO – The United Nations Educational, Scientific and Cultural Organization.

Education in the Developing Countries – A Changing Scene

Education in developing countries is varied and rapidly changing. Some countries have made massive advances, while the systems in others slide further into crisis. A common theme is the high priority given to education and particularly the commitment to achieve an effective primary education for all (EFA). Patterns in secondary education are not only affected by the major aim for primary education but also play a major role in that quest.

Most developing countries emerged into independence after World War II. At the end of the war, the allied nations acknowledged the importance of joint international efforts to avoid the threat of future wars, establishing in 1945 the United Nations Organization (UN). The UN as one of its early emphases sponsored the Declaration of Human Rights placing education as one of the essential rights of all people:

Everyone has the right to education. Education shall be free, at least in the elementary and fundamental stages. Elementary education shall be compulsory. Technical and professional education shall be made generally available and higher education shall be equally accessible to all on the basis of merit. (UN, 1945)

The commitment to achieve EFA has been difficult to deliver. One of the early actions of the UN was to establish the United Nations Educational, Scientific, and Cultural Organization (UNESCO) to assist countries to fulfill relevant aspects of the Human Rights Declaration. Since these early days, UNESCO has played a unique and important role, one which has grown exponentially with the emergence of the many new countries after 1946.

UNESCO was originally an organization of some 51 nations with a strong European–North American focus. It now has almost 200 member nations embracing all areas of the world with many more developing than developed countries. The UNESCO member nations originally had a total population of 2.1 billion: currently that population approaches 7 billion most of who are in developing countries. In 1950, there were 300 million students in schools, now there are more than 1.4 billion. However, almost 100 million children are still denied access to formal schooling, and many of those who are involved are taught in very large classes with few facilities (Dakar 1, 2000).

Given the central role played by UNESCO in education in developing countries, an understanding of its role is crucial. The founding meeting for UNESCO in London late in 1945 declared the central purpose in words that are still relevant:

The purpose of the Organization is to contribute to peace and security by promoting collaboration among nations through education, science and culture in order to further universal respect for justice, for the rule of law and for the human rights and fundamental freedoms which are affirmed for the peoples of the world, without distinction of race, sex, language and religion, by the Charter of the United Nations. (UNESCO, 1945)

As part of that thrust toward peace and security the countries saw education as an essential instrument in their economic and social development. Their activities are supported through the introduction to the world scene of major development agencies such as the World Bank, the International Monetary Fund (IMF), and the Asian Development Bank.

It is in the context of international efforts that the full story of education for developing countries can be told. The aim of EFA has been accepted by all countries, developed and developing, but remains unfinished in spite of major efforts over 60 years. It has become enshrined in the Millennium Development Goals (MDGs) of the UN,

signed by 147 nations, with a target date for achievement of 2015. The eight goals are:

- Eradicate poverty and extreme hunger; (reduce by half those living on a dollar per day and by half those living in hunger).
- Achieve universal primary education.
- Promote gender equality and empower women.
- Reduce child mortality; (by two-thirds of those under five).
- Improve maternal health.
- Combat malaria, HIV/Aids and other diseases.
- Ensure environmental sustainability.
- Develop a global partnership for development. (UN, 2008).

The goal of universal primary education (UPE) formed the central strategy for education development and EFA. The achievement of this goal remains incomplete after more than 60 years, and is increasingly discussed in tandem with sectors such as technical and vocational education and training (TVET), higher education, and secondary education. Over the past two decades, secondary education has come more and more into focus. The Delors Report, *Education for the Twenty-First Century* (Delors, 1996), described this sector as “the crossroads of education,” seeing it as a vital yet undervalued factor in the success of the EFA.

A Broader Concept of EFA

Three major international initiatives have emphasized the need for extensions of the concept of EFA: *The World Conference on Education for All*, (Jomtien, 1990); *The International Conference on Education for the 21st Century*, (Delors, 1996); and, the Dakar meeting in 2000, the *World Education Forum*. Initiatives emerging from these forums express a much more crucial role for secondary education.

At Jomtien in 1990, the commitment to EFA was reinforced by the agreement of the 200 nations attending. The report from Jomtien was bleak. In spite of major efforts, nationally and internationally, the situation was, in many ways, worse. The 1970 figure for illiterates in the age group 15+ was 760 million; by 1980, it was 824 million; and by 1990, there were 882 million, with an estimate of 912 million by 2000 (Jomtien, 1990). Access to primary education was progressing far too unevenly across developing countries.

The apparent failure of the world community’s extensive efforts toward EFA is partly attributable to massive population increases. The poor literacy and school attendance figures were paralleled by other social indicators: birth-rates, infant mortality, agricultural productivity, life expectancy, housing quality, availability of clean water, employment opportunities, and political participation. Where there was deterioration in one indicator there

tended to be deterioration in all others. The Conference called this grim conjunction “the convergence of disadvantage.” Education was the essential tool for breaking this cycle of misfortune, with a high priority being set for action. The concept of UPE was expanded to take a wider focus, basic education for all (BEFA). While not shifting the center of activity from primary education, this was clear recognition of the need to view educational and social development, holistically, thereby elevating interest in secondary as well as primary education. The report from Jomtien states:

These needs comprise both essential learning tools (such as literacy, oral expression, numeracy and problem-solving) and the basic learning content (such as knowledge, skills, values, and attitudes) required by human beings to be able to survive, to develop their full capacities, to live and work in dignity, to participate fully in development, to improve the quality of their lives, to make informed decisions, and to continue learning. (Jomtien, 1990)

In 1993, UNESCO set out to develop further its policy directions for the future of education. This was done through the International Commission on Education for the Twenty-First Century, *The Delors Report*. The report recognized a continuing commitment to good-quality basic education for everyone but stressed that this should include extension beyond primary to secondary education. It is important to note that this expanded commitment targeted not only the expansion of secondary education, but also the need to reform it. The Delors Report states:

Many of the hopes and criticisms aroused by the formal systems seem to focus on secondary education. On the one hand, it is often regarded as the gateway to social and economic advancement. It is accused on the other hand, of being inequalitarian and not sufficiently open to the outside world and generally failing to prepare adolescents not only for higher education but also for the world of work. In addition, it is also argued that the subjects taught are irrelevant and that not enough attention is paid to the acquisition of attitudes and values. It is now generally recognised that, for economic growth to take place, a high proportion of the population has to have received secondary education. It would thus be useful to clarify what secondary education needs to do to prepare young people for adulthood. (Delors, 1996)

At Dakar in 2000, mechanisms were established to oversee EFA until 2015. The nations reinforced the aim and set a target date of 2015 for the goal to be achieved; that is:

The commitment of EFA is to basic education. But this is not a fixed or clear-cut concept . . . Most but by no means all have chosen to restrict ‘basic’ to primary schooling. . . . which so many have still to attain. ‘Basic’ in an increasing

number of countries, however, now connotes not only early childhood care and education and primary schooling, in some it now encompasses junior secondary schooling and in others it extends to a full secondary education. (EFA Forum, 2000)

The term basic education was further redefined by the director-general of UNESCO:

Basic education denotes the minimum skills and knowledge needed in order to be able to make a full contribution to one's environment and to be in control of one's life. In an increasingly interdependent world, the contents, and therefore the very notion of the 'quality' of basic education are evolving. It can no longer be reduced to learning reading, writing and arithmetic. It must also teach individuals to be, to do, to learn and to live together. (Maatsuura, 2000)

Delors also stressed that basic education should also be made available to the 900 million illiterate adults, 100 million children not in school, and the more than 100 million children who drop out of school. The comments of Delors remain relevant for all countries. While the shortcomings of secondary education are obviously more apparent in developing countries, developed countries also fall short in equity terms as a substantial percentage of students leave school with no formal qualification or with poor standards of literacy and numeracy. Delors defined basic education essentially as a "passport to life"; or an enabling foundation on which diversity can be built; diversity in course and teaching-learning processes; and, diversity in context, including the alternating of study and work.

The existing secondary system was perceived as far too inflexible for current needs and a variety of formats was suggested, including alternation between school and work. The pressing need for societies is to make this stage of learning work for all, rather than using it as a filtering system, leaving a large proportion of students with a feeling of failure and frustration. A variety of pathways is necessary between these different educational opportunities – pathways which allow students to move in and out at different times depending on circumstances and needs.

In June 2007, the UN provided a progress report on the achievement of the MDGs. The assessment showed mixed results. None of the eight goals had made sufficient progress to ensure their achievement by 2015. The proportion of people living on less than a dollar per day had fallen from 23.4% in 1999 to 19.8%, broadly on target but with massive variations. The eradication of hunger was still remote with 46% of the under-fives in Africa registered as underweight: infant mortality is down by one-sixth, as against the target of two-thirds reduction. Whereas China and India and other parts of Asia had increased school attendance, 12% of children globally and 30% in sub-Saharan Africa were still out of school. Other targets,

particularly in Africa, displayed similar patterns. According to UN (2007):

The world is failing in the battle to combat hunger, reduce infant mortality and to put every child in school. The results show that there have been some gains and that success is still possible in most parts of the world but they also point to how much remains to be done.

The Dakar World Education Forum was particularly important in terms of shifting emphasis beyond primary education and redefining basic education to include secondary education. A focused session during the conference asked directly "After Primary Education: What?" (MacLean, 2003). The question captured the essence of the way ahead and acknowledged the fact that secondary schooling correlates positively with lower fertility and infant and maternal mortality rates. It also provides adolescents more powerful means to participate actively in social, cultural, and political life. When pulled together these, as governments around the world have openly acknowledged, have important implications for economic development.

The World Bank has also highlighted the importance of secondary education to economic and social development. Experience in newly industrialized countries has shown that growth in secondary education allows the formation of a pool of a better-educated workforce which can play an important role in attracting foreign investment. According to World Bank:

Secondary education... has been shown to contribute to individual earnings and economic growth. It is associated with improved health, equity, and social conditions. It buttresses democratic institutions and civic engagement. And the quality of secondary education affects the levels above and below it – primary and tertiary education. (World Bank, 2006: 17)

UNESCO has initiated a broad range of activities in an attempt to address the complexities of secondary education in developing countries. Maclean (2003) lists the major priorities as expanding access, assuring quality, reforming curriculum, and formulating alternate delivery systems. Curriculum reform emphasizes human immunodeficiency virus (HIV)/acquired immune deficiency syndrome (AIDS) education, linkages between school and work, values education, and science and technology. Other priority concerns include equity issues, especially those related to gender, effective pedagogies, student dropouts, learning and teaching resources, recruitment, and professional development.

In short, building secondary education helps adolescents develop abstract reasoning, observation, analysis and synthesis competencies, and other higher-order thinking skills. The development of a broader and deeper base in the sciences, technology, languages, and other

subjects can not only provide the immediate skills needed by developing economies but also lay a foundation for future lifelong training.

Issues in Secondary Education

Whereas UNESCO and the World Bank develop priorities and delineate program areas, in line with a more holistic approach, they see them more and more as closely related – inevitably interacting with one another in an ongoing dynamic fusion. Here, we address three of these interactive areas – access and expansion, content and context, as well as delivery and relevance.

Access and Expansion

Perhaps the key drivers of change and development are those related to the expansion of secondary education and subsequent opportunities, or a lack thereof, of access to it to significant proportions of young people. The unprecedented expansion of secondary education has been driven by the achievement of UPE and forceful government agendas to provide open access. For example, India, Pakistan, and China have moved aggressively toward a secondary system providing for all students. The Indian Prime Minister Manmohan Singh clearly outlined this agenda:

The government will make secondary education in India universal, with at least one school for every cluster of villages over the next two–three years. The program will build on the success of Sarva Shiksha Abhiyan education for all programme) and will cover the entire country in two–three years. These publicly funded 6000 schools will establish benchmarks for excellence in public schooling, which would then be role models for the rest of the public educational system. (Gulf Times, 2007)

Policy implementation still has a long way to go. For example, in 2006, while China had achieved a participation rate in secondary education of just over 70%, India had just breeched the 50% mark and Pakistan approximately 27% (Kennedy and Lee, 2008).

A number of Arab countries have also instituted ambitious reforms to expand access to secondary education. In Qatar, the Rand Corporation has undertaken a long-term development program for K-12 education in the country. Currently in Qatar, 25% of primary students do not complete their schooling so that a major program at this level is planned as a basis for the reform at the secondary level (UNESCO, 2005). In Bahrain, a National Education Reform Project has been introduced.

We need to work closely with the private sector to ensure that students have access to courses that are required by the economy and that graduates are competitive once they

enter the job market because of the skills and experience provided by the education system. In order to achieve this certain steps need to be taken such as improving the quality of education institutions – focusing on teachers' training and education; providing alternative vocational education opportunities for students; and establishing an independent Quality Assurance Authority that will regularly review the performance of all education institutions. (Forrest, 2007)

In Oman, in 1970, there were only three schools, with enrolments of 909 pupils, all male. By 1990, there were 838 schools, with 365 000 students, both male and female. By 1997, the number of schools had grown to 1062 and the number of students to 529 000. Building on this base, there were major technical and higher education developments. Similarly, there have been striking advances in adult literacy rates. The country, which has a high proportion of migrant workers, aims to supply its own needs for skilled people and also contribute to other countries.

The dramatic expansion of secondary education requires substantial financial input. Given the relative current wealth differentials between societies, this has been a problem for many African countries where the advances in primary education have been disappointing and secondary development is still largely a matter of plans rather than results.

In Kenya, the debate on provision of free education to secondary schools has become a point of focus with several presidential aspirants using it as a key rallying point to woo voters. In what some critics label as a campaign tool, the government has provided a Sh4.3 billion subsidy for public secondary schools to cater to tuition fees from January 2008. Presidential candidate Raila Odinga has promised to go beyond subsidized education to a free and compulsory primary and secondary education.

A study by the state-owned Kenya Institute for Public Policy Research and Analysis (Kippra) on the feasibility of free secondary education (FSE), says such a bold undertaking is a good public initiative that, if implemented, would increase access to secondary education while relieving the cost burden on parents. This will allow households to increase spending on other needs, such as health and post-secondary education and training, hence impact positively on poverty reduction. The findings reveal that the number of teachers required will double in the first 4 years of implementation of FSE. If it is effected next year, the number will shoot up from the current 48 425 to 53 828 and peak at 99 115 in 2011. The country will also need 8600 additional classrooms as well as new day schools and an expansion of the existing schools. The Nation, 2007 states:

While this sounds encouraging, a comparison between estimated increased total Government revenue and the financing gap reveals that implementing a full FSE is quite a sacrifice with the FSE financing gap likely to consume 28 per cent of the increases in total revenue in

the initial year. This will rise to 50 per cent of any revenue increase by the fourth year of implementation.

The report warns that allocation of any revenue increase to FSE will reduce investment in other competing sectors of the economy. In Nigeria, the curriculum of the nation's senior secondary schools (SSSs) is to undergo a complete overhauling with students learning a trade as well as civic education. Professor Godswill Obioma, executive secretary, Nigerian Educational Research and Development Council (NERDC), noted that the overhaul of the curriculum of the SSSs was part of the reforms in the education sector and a follow-up on the restructuring of the 9-year basic education curriculum. The review of the curriculum is to ensure that children are properly educated and have entrepreneurial skills to be able to live a productive life (Nigerian Tribune, 2007).

Content and Context

A further challenge to developing secondary education is finding ways to shape, negotiate, deliver, and embed effective learning for all. This is difficult given that as much diversity exists within as between developing societies. Such diversity may relate to culture, politics, gender, race, or religion – all of which can influence the type of learning opportunities available and useful to students.

The shape of secondary education relates increasingly to globalization. Although there is little doubt that increased global interdependence means that schools can remain isolated within a single society, this does not mean that secondary education will or should adopt a one-size-fits-all model. Although key trends in secondary education, such as decentralization, quality education, outcomes-based education, and increased marketization, mirror trends common in developed countries, they can look very different at both policy and school levels.

At a policy level, marketization underpins broad moves to upgrade secondary education in many countries, but the meaning of market varies. The market in China is directly and explicitly regulated by the state. Educational policy continues to stress dedication to communist/socialist ideology across separate initiatives. At the school level, culture may determine to a greater or lesser extent the way secondary schools are governed and what happens in classrooms. Recognition of difference means that there is no global recipe for shaping secondary education at the operational level. Without such recognition, global initiatives by bodies such as UNESCO or the World Bank, regardless of their intentions, may struggle to take root in schools.

The growth in primary education has put extra pressures on secondary education. The changes will be fundamental if secondary education is to become an effective part of the whole. However, as developed countries continue to discover, the nature of a secondary education which covers

the entire age group is very different from that for a selective system, as is predominantly the norm in developing societies, essentially geared to prepare students for further education.

Delivery and Relevance

Spreading secondary education in terms of both quality and quantity is closely linked to the ways such education is delivered. The delivery of secondary education is an increasing issue as countries seek to expand their enrolments. Countries such as Brazil, Argentina, Mexico, India, Zambia, and Indonesia have significant portions of their student population receiving secondary education through distance-learning programs. Distance education is seen as an alternative way of organizing secondary education in order to reach a greater number of pupils than permitted by traditional means.

While recognizing the importance of differing values and other contexts, alternative models are needed to deliver secondary education. This may entail moving beyond understandings and structures currently engrained in more developed societies. Ironically, developing countries may benefit from more technologically savvy approaches being touted in the more developed world. Alternate delivery mechanisms may include nonformal means, the use of modern communications technologies and the vernacular, as well as the use of English as the medium of instruction. They may also integrate with more traditional community-based modes which were misplaced during periods of colonization. For example, programs in Argentina and Mexico provide lower secondary education in remote rural areas in a nonconventional way. Relying on tutors, teaching materials, and itinerant teachers, such programs allow disadvantaged children to attend school at relatively low cost.

When looking more to secondary education, we need also to look specifically at the following questions: Do we have the division between primary and secondary right? Should traditional divisions between primary, secondary, and further education be retained? How can secondary education be made more relevant to the labor market? What links between schools and employers should be established to facilitate the transition from school to work, and what role can TVET play in this respect? What strategies can be developed for young people at risk?

Where to Go?

An interesting byproduct of the knowledge age is that all countries, regardless of their stage of development, are reexamining their approach to and understanding of secondary schooling. Common drivers for this are broadening conceptions of EFA and recognition of the interrelated

and interactive nature of learning, education levels, and age. Reworked ways of thinking require that secondary schools take on multiple functions: preparation for higher education, preparation for work, preparation for citizenship, together with the two additional emphases identified so strongly in the Delors Report – learning to live together and learning to be. Being the endpoint of basic education, the last stage of EFA students presents problems few countries have solved.

This shift in direction has exacerbated the problems surrounding secondary education in developing countries. Perhaps the sharpest of the many dilemmas faced by developing countries and international support agencies is that while economic performance is necessary for social improvement and major systems upgrading, educational improvement is impossible without a major increase in expenditure. Where bodies, such as the World Bank, once followed a strictly economic agenda they now see the necessity to provide a substantial boost to education as an enabling measure for economic and social advance. The Delors Report stresses the urgency of a debate “at a time when educational policies are being sharply criticised or pushed-for economic and financial reasons-down to the bottom of the agenda.” This may well drive the continued search for better, more equitable, and meaningful secondary education in developing, and in developed, countries over the next 20 years.

Bibliography

Dakar 1 (2000). *World Education Forum, Final Report*. Paris: UNESCO.
 Delors, J. (1996). Learning: The treasure within. *Report of the International Commission on Education for the 21st Century*. UNESCO: Paris.

EFA Forum (2000). *The Dakar Framework*. Paris: UNESCO.
 Forrest, M. (2007). Spotlight on education reform. *Bahrain Daily News*. 29/07/07.
 Gulf Times (2007). *Secondary Education for All in Three Years*. June 30, 2007.
 Jomtien (1990). *World Conference on Education for All*. Paris: UNESCO.
 Kennedy, K. and Lee, C. K. (2008). *The Changing Role of Schools in Asian Societies*. London: Routledge.
 Maatsuura, K. (2000). Is the globalization of the economy creating values for a new civilization? *Prospects* 30(4), 399–404.
 Maclean, R. (2003). Secondary education reform in the Asia-Pacific Region. In Keeves, J. and Watanabe, R. (eds.) *The Handbook of Educational Research in the Asia-Pacific Region*, pp 73–92. Dordrecht: Kluwer Press.
 Nigerian Tribune (2007). *Restructuring the Curriculum*. 16 July, 2007.
 The Nation (2007). *Oura Future in Education*. The Nation, Nairobi, 11 July, 2007.
 UN (1945). *Article 1 of the UNESCO Charter*. New York: UN.
 UN (2007). *The Millennium Goals: Mid-Term Assessment*. New York: UN.
 UN (2008). *Millennium Development Goals*. New York: UN. <http://www.un.org/millenniumgoals>.
 UNESCO (1945). *Article 1 of UNESCO Charter*. Paris: UNESCO.
 UNESCO (2005). *Global Education Digest*, UNESCO Institute of Statistics, Montreal.
 World Bank (2006). *Global Monitoring Report, 2006*. New York: World Bank.

Further Reading

Dakar 2 (2000). *The Dakar Framework for Action*. Paris: UNESCO.
 Dakar 3 (2000). *Statistical Document*. Paris: UNESCO.

PRIMARY AND SECONDARY EDUCATION – SOCIAL JUSTICE ISSUES IN SCHOOL AGE EDUCATION

Community Focused Schools

A Dyson, University of Manchester, Manchester, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Despite the transformations in education that have taken place in most countries over the past 100 years, the characteristics and functions of schools have remained remarkably unchanged. Schools are institutions whose function is to educate children and young people. They are staffed by teachers and support staff necessary to assist the teachers, carry out administrative tasks, or maintain the building. Typically, students attend for a limited number of hours each day, a limited number of days each week, and a limited number of weeks each year. When students are present, other family and community members have little access to the school, and when students are not present, the building may well be closed.

This pattern has been so common as to attract little comment. However, it overlooks two major considerations. First, the resource locked up in schools – the buildings and facilities, the intellectual resource of the teaching staff, the cultural resource of values and expectations, and the employment opportunities on offer – constitute a major investment in the areas where they are located. Second, those resources, powerful as they are, remain relatively weak in relation to other factors in the lives of children that impact on their well-being, their educational achievements, and their life chances. Family dynamics, local cultures, and material conditions remain beyond the reach of the school, while the health, social care, welfare, and other services that might impact on these factors operate, at best, in a way that is only loosely coordinated with the work of the school.

These considerations are important even where resources are abundant and students in general do well. However, they become crucial in contexts where social resources are scarce, or where there are deeply entrenched disadvantages and students who face multiple personal, family, and social difficulties. In these situations, it is arguable that the existing investment in schools is underutilized

in relation to its potential benefits for local communities, yet underpowered in relation to the needs of children and young people.

Not surprisingly, therefore, there have been attempts in many places to think about school resourcing differently. These attempts have been enormously diverse, but basically involve some mixture of enhancing the resources available in the school for working with students while making the school's resources available to families and local people. Often, these attempts result from the initiative of particular school principals attempting to meet what they see as the needs of their students and of local communities. Sometimes, however, they take the form of more or less well-organized programs sponsored by national or local governments, or by nongovernmental organizations (NGOs). In these cases, the preferred model of provision may be dignified with a title – full-service schools, community schools, extended schools, or schools plus, for instance. However, there is no internationally agreed term for schools that work in this way, and any label inevitably disguises the real diversity of approach.

With this caveat in mind, we shall refer to these schools here as community focused. This term is borrowed from the education system in Wales, where a community-focused school is defined as:

... one that provides a range of services and activities, often beyond the school day, to help meet the needs of its pupils, their families and the wider community. (National Assembly for Wales, 2003: 1.2)

Whatever the shortcomings of this label and definition, it at least draws attention to the recognition by these schools that there are communities beyond the school gates, that the school may have an obligation to those communities, and that the effective education of students may require the school to engage with communities and the resources available therein.

What Do Community-Focused Schools Do?

Rather than a single blueprint for community-focused schools, it is probably closer to the truth to say that there is a menu of activities and forms of provision. In broad terms, the menu has the following components:

- activities outside the standard school day and year to extend the curriculum offer to students;
- additional forms of support for students' learning, social, and health needs;
- opportunities for students to develop leadership skills and community engagement;
- support for families on personal, social, health, and welfare issues;
- opportunities for families to become involved with the school and, particularly, to support their children's learning; and
- opportunities for community members to use school facilities, and engage in arts, leisure, learning, and vocational development activities.

Delivering this menu in turn implies that schools have to make a series of changes to their practices. For instance, they may have to:

- change the use of the school site so that it is open beyond the school day and year and is accessible to parents and community members;
- change their staffing structures to employ more non-teachers able to work with students, families, and community members;
- develop new relationships with other education providers, child and family services, community organizations, and business in order to deliver a wider range of services on site or in collaboration; and
- develop new styles of leadership and governance, perhaps involving community members.

In practice, schools select from these menus and undertake changes selectively in line with how they understand their particular contexts, the requirements of any programs in which they are participating, their access to funding streams, and a whole range of other local factors. Two examples illustrate this point. Both are of community-focused schools, yet what this means in practice is very different in each case:

1. *The Arturo Toscanini Complex (ATC)*. It is a campus in New York City hosting three middle (grades 6–8) schools. It follows a model of community schooling promoted by the Children's Aid Society – a model which is as well developed and well evaluated as any in this field. ATC's provision recognizably derives from the menu outlined above, but does so in ways that meet the disadvantaged and multi-ethnic characteristics of its inner-city context. It includes:

- extended day provision, including homework support, literacy tuition, creative writing, and tutoring;
- programs focusing on personal and social development;
- fitness and health programs;
- performing-arts activities;
- access for students and families to health, dental health, and mental health services;
- social work support for students and families;
- opportunities for student involvement in leadership activities and community issues; and
- English as a second language, welfare assistance, family support, health-insurance advice, and cultural and leisure activities for parents and community members.

2. *The Evergreen School*. It is a small school in England catering exclusively for students with high levels of special educational need. Its concerns are not with the socio-economic disadvantages experienced by students, families, and local communities, but with the opportunities for its disabled students to integrate with their nondisabled peers, and the support available to students' families. Hence, its program includes:

- weekend clubs for students;
- arts projects in which the schools' students work alongside children from regular schools;
- access to the school's facilities for children from other schools and to adults from local communities;
- summer-school provision for the school's students alongside other members of their families; and
- support groups for parents of students and of younger children with special educational needs (see 4Children, 2007: 10).

Contrasting examples such as these could be multiplied many times. What differentiates these schools is the response to local contexts and, in particular, to the perceived needs of their students and their students' families and communities. What links them is the intention to bring additional resources to bear on the school's work with students, and to bring the school's resources to bear on family and community issues.

Rationales

If we are to understand the diversity of community-focused approaches, it is necessary to unpack a little further the idea that those approaches arise from schools' responses to their local contexts. As in the examples above, schools do indeed work in contexts where there are marked material differences and, particularly, where students, families, and communities have quite different characteristics. However, it is clear that the leaders of community-focused initiatives also interpret those differences in relation to a series of assumptions – among other things, about what counts as a

need, which needs are greatest, and what schools can and should do in response to those needs. The interaction between the material conditions of local contexts and the interpretations of those contexts give rise to different rationales on which community-focused approaches are based.

Again, two examples will serve to illustrate the point. The first is a rationale articulated by Joy Dryfoos, one of the foremost advocates of full-service schools in the United States:

... schools are failing because they cannot meet the complex needs of today's students. Teachers cannot teach hungry children or cope with young people who are too distraught to learn. Anyone working in an inner-city school, in a marginal rural area, or even on the fringes of suburbia will tell you how impossible her or his job has become. The cumulative effects of poverty have created social environments that challenge educators, community leaders, and practitioners of health, mental health, and social services to invent new kinds of institutional responses. (Dryfoos, 1994: xvii)

For Dryfoos, the problem is essentially one of the enormous social stresses produced by poverty. These stresses are reflected in the turbulent lives and unmet needs of children, who, as a consequence, come to school unable to learn. Traditional forms of service delivery cannot, she argue, meet these overwhelming needs. Instead, what are needed are full-service schools, where a range of child and family services will be co-located on the school site so that they can respond quickly and effectively to the difficulties that children face.

The second is taken from a memorandum of 1925 by Henry Morris, then chief education officer of the predominantly rural county of Cambridgeshire in England, proposing the establishment of village colleges. Morris' vision has been every bit as formative in the English context as Dryfoos' has in the United States, and village colleges and their like continue to flourish to this day. Morris' concerns, however, were very different from those of Dryfoos. He believed that the countryside was losing out, educationally and economically, to the towns, not least because its education and leisure provision was scattered, uncoordinated, and difficult for many people to access. His proposed village colleges would, therefore, house a range of community education and leisure facilities under one roof. In this way:

The isolated and insulated school, which has now no organic connection with higher education, would form part of an institution in which the ultimate goal of education would be realized. As the community centre of the neighbourhood the village college would provide for the whole man, and abolish the duality of education and ordinary life. It would not only be the training ground for the art of living, but the place in which life is lived, the

environment of a genuine corporate life. The dismal dispute of vocational and nonvocational education would not arise in it, because education and living would be equated. It would be a visible demonstration in stone of the continuity and never ceasingness of education. (Morris, 1925: XV)

These statements of purpose share the view that what goes on inside the four walls of schools cannot be separated from what goes on in the wider communities that schools serve and where their students live. However, each takes a different view of the nature of those communities, of their needs, and of the needs of learners within them. For Dryfoos, communities are ultimately aggregations of poverty, needs are to do with health and hunger, and the aim of full-service provision is to make students teach-able. For Morris, on the other hand, communities are the location for a viable social and cultural life, needs are for quality of life, and the purpose of village colleges is to make communities in this broad sense sustainable.

Again, these examples could be multiplied many times over. For instance, in Saskatchewan, Canada, community schools have the same focus as Dryfoos' full-service schools on children and communities at risk. Here, however, the target groups also include First Peoples, and the approach is based not simply on providing services, but on cultural affirmation and community empowerment. In South Africa, full-service schools are seen as a means of promoting inclusive education by locating services in the regular school that might otherwise be available only in special schools. In some of the new democracies of Central and Eastern Europe, community schools are seen as agents for fostering democratic engagement, involving students and adults in solving social problems rather than remaining dependent on state intervention. Even within the same initiative, there may be quite different rationales. For instance, a full-service schools initiative in Australia was accompanied by proposals that such schools might alternatively recreate a sense of community in the face of global change, break the cycle of disadvantage, or empower young people and families to become participants in the creation of dynamic communities (Kirner *et al.*, 1998: iii).

It is possible to see these different rationales as lying along two dimensions: a student-community dimension and a deficit-empowerment dimension. The first is to do with how the community-focused approach is targeted and who are seen as its intended beneficiaries. Some rationales – Dryfoos' being an obvious example – focus on the potential benefits for students. By concentrating resources on the school site, students can be offered higher levels of support or a richer array of opportunities. The community-focused approach is thus, in effect, a means of enhancing the school's capacity to play its traditional role in the education of the young. Other approaches – such as Morris' – focus by contrast

on benefits for families and the wider community. By opening up the school's resources to wider use, conditions beyond the school gates can be changed and local people can benefit by having their needs met, or by accessing enhanced opportunities.

The second, deficit–empowerment dimension is to do with whether the intended beneficiaries of community-focused approaches are seen, to borrow a phrase from Semmens and Stokes (cited in Kirner *et al.*, 1998: 26), as patients or as citizens. For some – Dryfoos again being a prime example – the students and families served by community-focused schools are, like patients, in medicine, defined by their needs. If they were not in such dire circumstances, if they did not present with such severe and multiple problems, there would be no need for schools to reorient their work. For others, the beneficiaries of community-focused approaches are not targeted because of their evident needs but because, as actual or potential citizens, they have rights and are entitled to a rich array of possibilities. For Morris, for instance, village colleges are not about addressing the problems and disadvantages of needy people, but about making available to rural citizens the facilities that their urban counterparts take for granted, and about ensuring the sustainability of the rural way of life.

The location of rationales along these dimensions, of course, itself implies yet more fundamental assumptions. One review of developments in France, Sweden, and the United States, for instance, found that the diversity of approach between these countries was grounded in the way they:

... **choose** to understand childhood and the relationship between children, family, local community and wider society. (Moss *et al.*, 1999: 39, emphases in original)

In particular, the two dimensions outlined above can be seen to emerge out of fundamentally different ways of understanding the role of schools in building the good society. Across the world and, particularly, across the countries of the affluent West, governments are concerned about using education to enhance countries' competitiveness in a globalized economy, and, particularly, about narrowing the gaps in achievements and capacity to compete between the most disadvantaged members of society and the majority. In this context, schools that combine powerful strategies for raising achievement with strategies for addressing students' personal difficulties, tackling family problems, and ameliorating some of the dysfunctional aspects of the communities where they live seem very attractive to policymakers. It is no coincidence that in England, for instance, a recent wave of interest in extended schools originated as part of an attempt to address the social exclusion of disadvantaged groups and, particularly, to promote the renewal of urban areas where those groups were concentrated.

This contrasts markedly with what some might call a communitarian view. Here, the focus is on building communities based on shared values and democratic participation, with social breakdown understood not in terms of socioeconomic disadvantage alone, but of alienation and disengagement. The role of education in general and schools in particular, is to promote social participation. As two American writers on school-community relations put it:

Participatory democracy provides an institutional mechanism in public schools whereby the condition of postmodern chaos and disorientation in the local community and indeed the wider society can be dealt with more effectively. What results, one hopes, are "communities of hope". (Mirón and St. John, 2003: 306, in-text citations omitted)

These fundamental differences are of much more than theoretical interest. For practitioners, policymakers and, indeed, researchers, the purposes of developing community-focused schools can often seem to be self-evident. As a result, underlying assumptions tend not to be made explicit, and activities tend to be embarked on without any systematic and rigorous analysis of the purposes they are expected to fulfill. At the very least, this may mean that actions are poorly focused and less effective than they might otherwise be. At worst, it may mean that well-intentioned professionals impose on local people their implicit and unproblematic views of where their difficulties originate and what it is they need.

Evidence of Impact

Impressive claims are made about the impact of community-focused schools. A review of evaluations from 20 US initiatives (Blank *et al.*, 2003) is typical in concluding that benefits can be anticipated for children, their families, local communities, and schools themselves. In brief, children learn more effectively, parents offer better support to their children's learning and engage more fully with schools, communities become more sustainable, and schools function more effectively. However, such claims need handling with care. Other reviews are much more cautious about what can be claimed. Keyes and Gregg (2001), for instance, are skeptical about whether there are any impacts on children's learning, while others point to the paucity and poor quality of evaluation studies. As Wilkin *et al.* conclude, there has been "little systematic, rigorous evaluation of the concept and its implementation" (Wilkin *et al.*, 2003: v). Instead, the literature tends to focus on advocacy and guidance – from an international perspective, it is not helpful that it is dominated by US publications.

It seems, however, that poor quality is not the sole problem. Given the differing rationales outlined above, and the differences in interpretation from school to school, it is far from clear what community-focused schools are

supposed to achieve and how success should be judged. For many potential outcomes there are no obvious indicators, and evaluations tend therefore to focus on more readily measurable outcomes, such as children's attainments. Even where outcomes measures are available, schools' community-focused approaches tend to be complex and to be implemented in turbulent environments. Attributing changes in outcomes to those approaches can, therefore, be highly problematic.

In this situation, some critics conclude that the claimed impacts of community focused approaches are, in fact, illusory (Rees *et al.*, 2007). However, this seems harsh. The weight of the evidence – if not always its quality – suggests that positive impacts are likely, particularly since community-focused schools frequently adopt programs and interventions that have been shown to be effective in other contexts. Moreover, where evaluations take proper account of the complexity of community-focused initiatives and the nature of intended outcomes, findings become more convincingly positive. For instance, the national evaluation of full-service extended schools in England (Cummings *et al.*, 2007) was atypical both in using a multistrand approach to evaluation and in incorporating a theory-of-change methodology that could be customized to the intended outcomes of different schools. As a result, it was able to demonstrate the significant positive effects these schools were having on highly disadvantaged students and families, and the (marginally) superior achievement outcomes compared to other, similar schools. It was also able to show the benefits of these schools in relation to their costs, and to indicate how they might come to have widespread effects on engagement with learning and other indicators of well-being in communities.

In this situation, it is possible not only to be optimistic about the likely impacts of community-focused schools, but also to be realistic about the outcomes that can be claimed. Such schools are likely to bring appreciable benefits to their students, to families, and to local communities. Those benefits may be significant for some groups and individuals, but are unlikely to bring about large-scale transformations. Moreover, the kinds of outcomes that can be anticipated depend very much on the focus of each school's work. Above all, further robust evaluations are urgently needed and, in their absence, the leaders of initiatives need to build their own theories of change and ensure they are in a position to evaluate them.

Beyond the Community-Focused School

Community-focused approaches are essentially about strengthening schools' access to community resources and communities' access to school resources. It is remarkable, then, how few of them form part of wider strategies

for community development, area regeneration or, indeed, social justice. Despite the emphasis on community at the level of rhetoric, in practice most remain as school-centered initiatives, driven by school leaders, or leaders in local education departments or the education arms of NGOs. This may be because of the dominance of US initiatives and thinking in this field. As Moss, Petrie, and Poland point out in their review of international developments, American initiatives tend to be seen as ways of reversing the economic inequalities and social dislocations that in some other countries (they refer specifically to Sweden) are addressed through other public policies (Moss *et al.*, 1999: 37).

One problem with this reliance on school-level action is that, however successful they are in reconfiguring community resources, schools remain underpowered in the face of structural inequalities and disadvantages. The implication is that any move toward community-focused approaches in schools has to be part of, rather than a substitute for, much more wide-ranging policies for social justice. Ultimately, these are matters for national governments. However, intermediate steps are possible. For instance, schools can adopt the more positive view of disadvantaged communities that we outlined earlier and become centers not merely of service provision for local people, but of community activism. There are also examples of schools working within the framework of local strategies for community development and area regeneration. Such strategies have the potential for taking individual schools' community-focused approaches, supporting them, and multiplying their effects by aligning them with a wide range of other interventions.

As the proposal for one such strategic development puts it:

A wider view has to be taken beyond that which applies in any one school. The impact of difficulties faced by an individual school and the outcomes for individual children are felt by other higher level schools, by alternative education & training providers and by the wider community at an economic, social and behavioural level. The stakeholders in education have to be prepared to come together to pro-actively plan for better outcomes taking the view that children and young people need to be helped throughout all their educational phases. Their success at the end is not 'owned' by any one stakeholder but is a shared achievement and one in which everyone involved can take pride in. (Ryan, 2004: 57)

It may well be, therefore, that rethinking the role of schools in relation to the communities they serve should not stop at the proliferation of community-focused schools in their current form. Instead of simply broadening the work of the school, we may now need to integrate that work within coherent cross-sectoral strategies for addressing disadvantage and promoting well-being at the local level.

Bibliography

- 4Children (2007). *Special Schools in the East Midlands Delivering the Full Core Offer of Extended Services: An Examination of Best Practice*. London: 4Children for the East Midlands Regional Partnership.
- Blank, M., Melaville, A., and Shah, B. (2003). *Making the Difference: Research and Practice in Community Schools*. Washington, DC: Coalition for Community Schools, Institute for Educational Leadership.
- Cummings, C., Dyson, A., Muijs, D., et al. (2007). *Evaluation of the Full Service Extended Schools Initiative: Final Report. Research Report RR852*. London: DfES.
- Dryfoos, J. (1994). *Full-Service Schools*. San Francisco, CA: Jossey-Bass.
- Keyes, M. C. and Gregg, S. (2001). *School-Community Connections: A Literature Review*. Charleston, WV: AEL.
- Kirner, J., Mukherjee, D., Semmens, R., and Stokes, H. (1998). *Three Perspectives on Schools, Communities and Citizenship. Thinking About Full Service Schools No. 1*. Erskineville, NSW: Australian Centre for Equity Through Education.
- Mirón, L. F. and St. John, E. P. (2003). Implications of the new global context for urban reform. In Mirón, L. F. and St. John, E. P. (eds.) *Reinterpreting Urban School Reform: Have Urban Schools Failed, or Has the Reform Movement Failed Urban Schools?* pp 299–312. Albany, NY: State University of New York Press.
- Morris, H. (1925). *The Village College. Being a Memorandum on the Provision of Educational and Social Facilities for the Countryside, with Special Reference to Cambridgeshire*. <http://www.infed.org.uk/archives/e-texts/morris-m.htm>.
- Moss, P., Petrie, P., and Poland, G. (1999). *Rethinking School: Some International Perspectives*. Leicester: Youth Work Press for the Joseph Rowntree Foundation.
- National Assembly for Wales (2003). *Community Focused Schools. Guidance Circular 34/2003*.
- Rees, G., Power, S., and Taylor, C. (2007). The governance of educational inequalities: The limits of area-based initiatives. *Journal of Comparative Policy Analysis* 9, 261–274.
- Ryan, G. (2004). *Get It Right First Time: An Education Strategy for Ballymun*. Dublin: Ballymun Partnership.
- Wilkin, A., White, R., and Kinder, K. (2003). *Towards Extended Schools: A Literature Review*. London: DfES.
- of Children and Young People from Disadvantaged Areas: A Report from the Schools Plus Policy Action Team 11. London: DfEE.
- Dryfoos, J. G. and Maguire, S. (2002). *Inside Full-Service Community Schools*. Thousand Oaks, CA, Corwin Press.
- Dryfoos, J. G., Quinn, J., and Barkin, C. (eds.) (2005). *Community Schools in Action: Lessons from a Decade of Practice*. Oxford: Oxford University Press.
- Dyson, A. and Raffo, C. (2007). Education and disadvantage: The role of community-oriented schools. *Oxford Review of Education* 33, 297–314.
- Dyson, A. and Robson, E. (1999). *School, Family, Community: Mapping School Inclusion in the UK*. Leicester: Youth Work Press for the Joseph Rowntree Foundation.
- Forbes, J. and Watson, C. (eds.) (2009). *Service Integration in Schools: Research and Policy Discourses, Practices, and Future Prospects*. Rotterdam: Sense Publishers.
- Henderson, A. T. and Mapp, K. L. (2002). *A New Wave of Evidence: The Impact of School, Family, and Community Connections on Student Achievement: Annual Synthesis 2002*. Austin, TX: Southwest Educational Development Laboratory.
- Piper, J. (ed.) (2006). *Schools Plus to Extended Schools: Lessons from the Last Five Years*. Coventry: ContinYou.
- The Dutch Community School Steering Group (2004). *The Community School 0–12 Years*. <http://www.youthpolicy.nl/youthpolicy/docs/pdf/The%20Community%20School.pdf> (accessed September 2009).
- van Veen, D. (2001). *Leading Cities in Educational Renewal: From School Improvement & Designing Multi-Service Schools Towards Improving the Social Infrastructure*. Leuven: Garant.

Relevant Websites

- <http://www.communityschools.org> – Coalition for Community Schools (US).
- <http://www.cisnet.org> – Communities in Schools (US).
- <http://www.cs-network.ru> – Community School Movement (Russia).
- <http://www.continyou.org.uk> – ContinYou (UK).
- <http://www.dest.gov.au> – Department of Education, Employment and Workplace Relations: Schools Education Publication Indexes (Australia).
- <http://www.education.gov.sk.ca> – Government of Saskatchewan Ministry of Education: SchoolPLUS Initiative (Canada).
- <http://www.sedl.org> – National Center for Family and Community Connections with Schools (US), SEDL.
- <http://www.ncea.com> – National Community Education Association (US).
- <http://www.education.manchester.ac.uk> – School of Education, The University of Manchester: The Centre for Equity in Education.
- <http://www.teachernet.gov.uk> – Teachernet Extended Schools Site (England), Extended Services.
- <http://www.aecf.org> – The Annie E. Casey Foundation (US).
- <http://www.childrensaidsociety.org> – The Children's Aid Society (US).

Further Reading

- Ball, M. (1998). *School Inclusion: The School, the Family and the Community*. York: Joseph Rowntree Foundation.
- Crowson, R. L. (ed.) (2001). *Community Development and School Reform*. London: JAI.
- Department for Education and Employment (DfEE) (1999). *Schools Plus: Building Learning Communities. Improving the Educational Chances*

PRIMARY AND SECONDARY EDUCATION – THE EFFECTIVENESS OF SCHOOL SYSTEMS

Contents

Reforming the American School System
School Effectiveness in Developed Societies
School to School Collaboration: Innovation and Improvement
Studies of School Improvement in Developing Countries
Sustainable Educational Reform
The Role of the District in Tri Level Reform
Whole School Designs for Enhancing Student Achievement

Reforming the American School System

M Barber

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In considering the remarks the author makes, three distinct perspectives have been brought to bear. First, as an outsider with some knowledge of education systems around the world, the author comments on how the school system in the United States looks against international benchmarks. (Think of this as speaking from the head.)

Second, as a battle-hardened veteran of the Blair government's controversial school reforms, the author draws some lessons from that broadly successful but also messy, error-strewn experience, which took us from below average to above average but not yet world class. These are thoughts, as the great Theodore Roosevelt put it, "from the man in the arena whose face is marred by dust and sweat and blood" rather than from one of those critics who stand on the sidelines, "cold timid souls who know neither victory nor defeat." (Think of this as speaking from the gut.)

Third, as a committed friend of this country who majored in American History at Oxford and who has been fascinated by trying to unravel its mysteries ever since, the author comments on the struggle – across over two centuries – to close the gap between the American Dream and the American reality. One of the virtues of a good friend should surely be that he or she can say what, otherwise, might go unsaid. (Think of this as speaking from the heart.)

This author believes that school reform in this country is at a critical juncture. In the next year or so, it will be necessary to choose between two broad options: on the one hand, a retreat to the comfortable, introverted, input-focused, evidence-light approach that characterized education reform in the last three decades of the twentieth

century, during which time Americans tried and failed to live up to the towering ambitions of the civil rights movement; on the other, an advance to the demanding, outward-looking, results-focused, evidence-informed approach toward which some promising progress has recently been made.

That choice will have to be made in governors' mansions, state capitols, city halls, and school boards across the country but, symbolically and substantively, the reauthorization of the No Child Left Behind (NCLB) Act – the most important piece of education legislation for many years and the most equitable legislation of the new century so far – will be the moment of truth.

The decisive factors in the making of this choice will be the accumulating evidence of what works (and what does not) and the courage of those who lead public education and who shape opinion within the system and among businesses and communities. This room is brimful of those courageous leaders and this author salutes their achievements so far – but they are only a beginning. In the words of the great poet, two roads are diverging in a wood and the choice this country makes – you make – will make all the difference.

As one prepares to face that fateful choice, three things need to be done – a glance back at the past, an assessment of the present, and a sketch of the future.

Section 1: The Past

In 1955, the year General Motors achieved a US market share of 50% and 2 years before the launch of Sputnik undermined America's post-war confidence, the American

high school reached its zenith – at least for white kids. A year earlier, the Supreme Court had momentarily decided that the education those white kids received should ultimately be available to all – setting the terms of debate for the ensuing decades.

Up to that time – and indeed beyond – the US had a huge comparative advantage over all other countries in the provision of universal, general education, as Claudia Goldin and Larry Katz have recently demonstrated. “During the first three quarters of the [20th] century educational attainment rose rapidly”, they argue.

This was, largely, due to the existence of universally available high school education but also to the growing availability of college. Because good schooling brings very long-term benefits, America’s educational leadership over the rest of the world brought substantial relative gains in economic growth right through to the end of the twentieth century. Even at present, the United States leads the world in the college-graduate share of those in the age range 55–64 years.

However, in the last quarter of the twentieth century, educational attainment stagnated in this country. Countries, which – educationally speaking – had been trailing in America’s wake for most of the twentieth century began to catch up. As Goldin and Katz explain:

The slowdown in the educational attainment of young Americans at the end of the twentieth century is especially striking when compared with the acceleration of schooling among many nations in Europe and parts of Asia, where educational change has been exceedingly rapid.

This relative slide in the educational performance of the United States has had – and will continue to have – economic consequences. Since “A greater level of education results in higher labour productivity [and] . . . tends to foster a higher rate of aggregate growth” relative weakness in education puts at risk long-term growth rates. The recent work of Eric Hanushek and others reinforces the case by demonstrating the strong, positive correlation between the performance of countries in Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS) and their rates of economic growth. Having looked at the international comparisons in science and math over the past 40 years, they conclude: “higher levels of cognitive skill appear to play a major role in explaining differences in economic growth.” They show that the United States has fallen, relatively, in these international comparisons and is now, at best, average.

Some suggest – given strong economic growth in the United States over the last two decades – that this does not matter or, as Gerald Bracey has argued, “Our schools are better than the critics claim.” This is a dangerously complacent line to take – the time lag in the relationship between

schooling and economic growth is long. Look at the data of Andreas Schleicher of the Organization of Economic Cooperation and Development (OECD)-PISA. In the 1960s, the United States led the world in high school qualifications and Korea was 27th. At present, Korea leads the world and the United States is 13th and falling. As recently as 1995, the United States was second in the world on college-level graduation rates; just a decade later, it has slipped to 14th – slightly below the OECD average.

This slippage is not the result of a lack of investment, which remains relatively high in the United States. Rather, it reflects – to use hard economic terms – a lack of productivity. The point is reinforced by the fact that, in international comparisons of younger children, the United States does relatively well which – given the country’s wealth – is what one would expect. The problem is that, as they get older, children make less progress each year than children in the best-performing countries. In this instance, the discussion is not just about poor kids in poor neighborhoods; it is about most kids in most neighborhoods.

Moreover, there is no comfort in the belief that future economic success depends not so much on the overall levels of cognitive skill in the population but rather on ensuring that at least a few brilliant rocket scientists come through. Hanushek *et al.* show convincingly that, in the twenty-first century, having “a substantial cadre of high performers” and “near universal basic skills” are both essential. In short, the choice is a false one and the debate a distraction.

Summarizing then, long term, the future success of the American economy will depend on significantly improving the US school (and college) system with all the urgency that can be mustered. Indeed, because of the inevitable time lag, even with the most rapid imaginable education reform, it will be some years before the impact is felt on economic growth. No wonder business leaders often play a leading role in driving school reform in this country; they see the hard edge of these issues. It is one thing in the global economy to offshore unskilled jobs because labor is cheaper elsewhere; quite another to offshore highly skilled jobs simply because the qualified workforce cannot be found; but, too often, this has become the reality.

Equally importantly, however, school reform in this country has never been just a question of economics – important though that is. From the beginning of the Republic, education was seen as fundamental to building democracy and extending freedom. Jefferson was giving expression to a universal belief when he said, “If a nation expects to be ignorant and free in a state of civilisation, it expects what never was and never will be.” This idealism was one of the reasons why the United States made much

earlier and more rapid progress toward universal public education than many countries in Europe. The State of New York, for example, established universal public education in 1812 – not matched even remotely in England until 1870. In fact, in 1812, the British were more preoccupied with shelling the White House! The roots of the comparative advantage Goldin and Katz identified in the twentieth century lie deep in the previous century.

From the 1950s onward, the realization of this ideal – an ideal which underpinned the American Dream – became central to the burgeoning civil rights movement, at first in the Deep South where grotesque educational inequality was placed firmly on the agenda by the Brown versus the Board of Education decision and later in the northern cities as the hopes of the diaspora, all too often, turned to despair.

There were, of course, many strands to the civil rights movement – all wonderfully woven together in Taylor Branch's monumental trilogy *America in the King Years*. Two things in his account are both striking and relevant to our agenda, at present. One is that the movement in the 1960s, while symbolized by great set-piece speeches or new legislation, was – in practice – the accumulation of numberless acts of often unrewarded, unnoticed heroism, which, like ripples on an ocean, combined to become an irresistible tide. One can see equivalent inspiring acts of educational heroism in schools and communities across this country now; they too need to become an ocean tide.

The second is that the leaders of the civil rights movement believed firmly that once equal access to school and college had been achieved, equal success would follow. This is evident when one listens to King speaking to African American school students in Cleveland in 1964; “doors of opportunity are opening now that were not opened to your mothers and fathers. The great challenge facing you is to be ready to enter those doors.” If they are ready, he implies, access will be enough.

Similarly, one can listen to President Johnson promising “every child a place to sit and a teacher to learn from” – he makes the same assumption that access would be enough. While some civil rights leaders expected it to take decades to make up the ground – Bob Moses estimated it would take “fifty years to work this through” – all expected that the ground would be made up in time. In relation to education, those expectations have not been realized.

Dig deeper than the headlines and the evidence is compelling. Not only does the United States perform somewhat below OECD averages in the recent PISA, it also suffers from a very high socioeconomic impact on student performance. In other words, rather than overcoming the social differences children bring with them when they start school, the US system – similar to the one in the United Kingdom – tends to reinforce

them. As Goldin and Katz have argued, “The slowdown in the growth of educational attainment . . . is the single most important factor increasing wage differentials since 1980 and is a major contributor to increased family inequality.”

Those who led the civil rights movement, whether in Congress, churches, or communities, must surely be devastated by the actual outcomes 40 or 50 years later. One knows, at present, that access to school is not enough. It is success in school that matters. It is also understood that, at the heart of education reform, at the heart of success, it is not just access, it is the quality of what happens in classrooms – the skills and knowledge, the expectations and ambitions, and the consistency and dedication which teachers bring to the task of enabling students, whatever their background, to achieve the standards necessary for life, work, and citizenship in the twenty-first century. Though no one believes it would be enough on its own, achieving the ambitions for 2014 set by NCLB would be a great start. This far-from-easy task has become, one would venture to suggest, the emerging frontier in the drive to realize civil rights in this country.

History suggests, therefore, that – at present – there is a great challenge facing America. Both future economic success and the wider aspirations at the heart of the very idea of America depend on vastly improving the outcomes of public education. The great threat to the country's future is that, for a range of reasons, it might fail to rise to this challenge. Then, for many, the American dream will never be more than a dream. The great opportunity is that a combination of business and civil rights leaders – along with the cross-party consensus that passed NCLB – could become unstoppable. At present, the issue hangs in the balance.

Section 2: The Present

This section commences by assessing the negative side of the ledger – the author's fears. One worry the author has is the sheer difficulty of getting things done in this country compared to many other countries, including his own. Of course, successful large-scale change is never easy but, in the Blair era, when the government had a large majority in Parliament and significant popular support, rapid progress was possible – as, of course, were rapid blunders! Partly in conscious reaction to that very British constitution, your founding fathers separated powers between three branches and two levels; a similar dispersion of power is usually found at state level. Add to that a culture which is historically suspicious of the very concept of government – as if George III is lurking behind every filing cabinet – and the challenge becomes much greater than in most European or Asian countries.

Then, recognize the power and organization of those who defend the *status quo*, face up to the legacy of failed attempts to bring about bold reform and the result is a widespread sense of defeat in people's heads before they even begin. Across the country, people sigh, along with the Russian Prime Minister – in the 1990s – who said, on leaving office, “We tried to do better but everything turned out as usual.”

Moreover, these organizational and cultural barriers within the system are compounded by the worrying lack of anxiety among the American people with regard to public education. It appears that the public is resigned to the state of their public schools rather than satisfied or delighted with them. *Education Next's* Fall issue (2008) finds that if parents could issue letter grades to the system – as schools do to students – just 20% would give an A or B. People are significantly more satisfied with their police forces and post offices. Even so, there is little recognition that, unless public education significantly improves in the near future, there is a disaster in the making. Education systems do not fail with the suddenness of a natural disaster but the consequences can be just as devastating. In a moment of despair, James Baldwin once observed that civilizations are destroyed not by wickedness but by spinelessness.

How many Americans see education as the top priority in facing up to the country's economic challenges? If public education really is the frontier of the civil rights movement, where are the modern equivalents of the Freedom Rides and the Freedom Summer? Where is the clamor? How is it that, recently, a leading elected official could say that if one youth was beaten by the police there were mass protests but if thousands of youths were failed by the schools no one lifted a finger?

John F Kennedy first made his name in 1940, by publishing an essay with regard to the appeasement of Hitler in the 1930s and called it *Why England Slept*. This author hopes, in 10 year's time, no one looking back on this country's attempt to grapple with education reform will feel the need to write *Why America Slept*.

Fortunately, these anxieties are balanced by very real grounds for hope. To start with, all over in America, one can sense a growing recognition among the country's leaders at local, state, and national level that public education needs fixing. Furthermore, many of these leaders are ready to look abroad as well as at home for solutions. The growing interest from states in the American Diploma Project is a case in point.

Moreover, not the least consequence of the NCLB Act is much greater clarity in the data with regard to the extent of the problem. The diagnosis is becoming clearer and, while this does not automatically lead to the cure, it is a major step forward.

In addition – again assisted by the data – we are increasingly well-informed about, to use Tony Blair's

favorite phrase, what works. We have chains of schools, such as KIPP, Aspire, and Green Dot, that demonstrably succeed where many – in the past – have failed. The United States has whole systems such as Boston, Chicago, and New York City which are driving bold reform and delivering results. There are organizations, such as Teach for America (TFA), the New Teacher Project, and New Leaders for New Schools, showing how apparently insurmountable human capital challenges can, in fact, be surmounted. There are also not-for-profit organizations, such as Education Trust and Achieve with deep expertise in crucial areas. There are also foundations – Gates, Broad, and Dell, for example – willing to take risks and invest substantially in bold alternatives to the inadequacies of the present. At the state and local level, there are community organizations and foundations acting with a similar sense of purpose. Never before has there been so much insight into how to bring about successful change, nor such substantial capacity to deliver it. The question is whether political and educational leaders can seize that insight and capacity and bring irreversible progress.

There is a further point; the very international comparisons that make such depressing reading for the United States (and challenge the United Kingdom too), also indicate a way forward. They demonstrate what can be done – with evident progress in less than a decade – with a combination of the right strategy and courageous, sustained leadership. Singapore's story over 40 years is truly inspirational. Thus, in an entirely different culture, is Finland's over 30 years. Poland made remarkable progress in the last decade. The reforms in Alberta and Ontario, just across the US northern border, are working too. It can be done.

Which leads to the author's final ground for hope – the NCLB Act itself. This was legislation which reached across political divides and set ambitious goals. It put no ceiling on educational performance but, for the first time, it fixed in legislation a high floor. It has put the achievement gap on the agenda from sea to shining sea. To set a date for delivery as early as 2014 was aspirational, certainly; some critics say it is unrealistic and the due date should be postponed, perhaps indefinitely.

Others say, “provide the preschool and the capacity first and we'll come to the accountability later.” Even if the American people would accept a pay-first-and-ask-questions-later approach, this argument fails to recognize the degree of urgency. Moreover, the fact that there are schools – at present – achieving those goals surely suggests significant delay would be a mistake. In any case, from the perspective of the Freedom Summer of 1964, far from looking too soon, 2014 appears a half-century too late. As the then vice-president said rejecting pleas for patience, on the 100th anniversary of the Gettysburg Address, “It is empty to plead that the solution to the

dilemmas of the present rests on the hands of a clock. The solution is in our hands.”

Across the country there are promising, albeit early, signs that – prompted by NCLB – progress is being made. Performance is improving – not enough, but at least heading in the right direction. Achievement gaps are narrowing – not yet transformatively, but increments are better than nothing.

Therefore, NCLB is a source of hope. The question it raises to my mind is not “Should it be reversed or abandoned?” but “How can it be refined and followed through?” In the last section of this article, the discussion turns to this.

Section 3: The Future

It is clear that, for all its qualities, NCLB does need refinement. It is rare for any law to pass and for legislature in the world to be without blemish and rarer still not to find ways of improving it once implementation begins. There are plenty of people – many of them in this room – who are much better informed than the present author with regard to the finer points of this particular piece of legislation and the Commission on NCLB (which the Aspen Institute supported) has published a distinguished and instructive report on the subject. A number of refinements, in particular, stand out from an international perspective.

First, the power of the Act depends crucially on the quality of the assessments being used. Where poor tests are used, the information they provide will be misleading with potentially dire consequences for the students themselves when they leave school and meet the real world coming the other way. One option for solving this problem was advocated recently by the *New York Times*, and would involve asking National Assessment of Educational Progress (NAEP) to create a rigorous test to be given free to states – with those that choose not to use it being publicly identified and, therefore, asked to explain themselves. Whether this is the best solution is not sure, but the direction is certainly right.

Second, the introduction of growth models has brought welcome refinement to the accountability requirements and it makes sense to build on this development. Growth models are helpful as a measure of progress and explanation of the scale of the challenge; they are dangerous when they become a justification for poor performance or lower expectations. Our experience in England suggests growth or value-added models should be combined with a continuing focus on the absolute outcomes which are all that matter to students when they leave school. They need to be part of a refined, modern, student-based data system which puts the evidence at the fingertips of every professional at every level.

Third, districts and states need to develop the capacity to act decisively in response to the data. They need to be able to recognize and reward those who succeed, especially those who succeed in challenging circumstances; they also need to intervene effectively where progress is not what it needs to be. In England, not without difficulty and controversy, this latter capacity was developed. Under pressure from us in central government, our equivalent of school districts did learn how to intervene effectively in failing schools. They did not apply pre-packaged interventions in mechanical sequence; instead, they diagnosed the problem in each school and tailored the solution accordingly, answering the only question that matters in these circumstances – “what do we need to do to get these kids a good education as fast as possible?”

Sometimes, it meant closure of the school and dispersal of the students to other schools; sometimes, it required the introduction of new providers and, sometimes, the replacement of a principal. The districts varied in their technical capacity to do this; they also varied in the degree of political will they brought to it. It is always tempting in these difficult circumstances to give the benefit of the doubt – but doing so, in this author’s experience, is almost always a mistake. After all, it raises the question, “How come you were so doubtful?” Meanwhile, at the central government level, we intervened in districts which lacked the capacity to drive school improvement. This was done successfully in almost 10% of the total, which, among other things, incentivized the rest.

This raises the fourth point of this discussion: successful education reform is as much about means as it is about ends. Getting the policy right is difficult, to be sure; but it is relatively easy compared to making it happen, consistently and effectively, so that the benefits are felt in every classroom. Policy failure is as often a failure of implementation as it is of concept. Systems need to develop both the technical capacity and the necessary mindset to deliver results. This is what, in the Blair administration we self-mockingly called “deliverology,” but, when we applied it systematically, it worked. An American equivalent is needed across the country to ensure the ambitions of NCLB are realized rather than eroded. As one of our more hapless kings – Charles I – once observed (shortly before they cut off his head), “There is more to the doing than bidding it be done.”

Fifth, there is much more to do to ensure there is a highly effective (more important than highly qualified) teacher in every classroom and a highly effective principal in every school. It is especially important to ensure that the schools facing the toughest challenges have access as soon as possible to the most talented teachers and leaders. Doing so requires root-and-branch reform of inherited, traditional, bureaucratic systems of recruiting and training teachers and leaders, of paying and rewarding them and of shaping their incentives, both short term and long

term, including pension arrangements. There needs to be a constant focus on developing talent and building capacity. At the moment, all around America one can see fine examples of what is required – TFA, the National Institute for School Leadership, KIPP's leadership development programs and New York City's Leadership Academy, for example – but, at the moment, these remain exceptions to the rule, not the rule itself.

In relation to human capital, there is a more profound, underlying question. As McKinsey's Report *How the World's Best Performing School Systems Come Out on Top* makes clear, the world's best systems are recruiting teachers who have both the right personal qualities and come from the top-third of the graduate distribution; in the United States, most teachers are from the bottom-third. Without improving their underlying capacity to attract talented people, education systems will struggle to compete in the future.

In England, to address the massive teacher shortage we faced a decade or more ago, we completely overhauled our teacher training – making it more classroom based, more accessible to people in mid-career who wanted to switch into teaching, better aligned with the school reform agenda and much better quality assured. We offered students incentives to go into teacher-education programs and varied those incentives according to the degree of shortage in particular subjects. Inspired by David Puttnam, the film producer, we introduced the National Teacher Awards – covered annually on prime-time television. Once the product had improved, it was promoted vigorously through advertising with the slogan, *Those Who Can, Teach*. The result has been several years of sustained improvement in the numbers and quality of recruits into teaching, assisted by our own version of TFA called Teach First. Teachers' pay, too, was increased but evidence from around the world shows that increasing pay on its own brings poor returns.

For the United States, the question of where, in the long run, it will find sufficient teachers of real quality – especially in science and math – remains unanswered. There is a medium-sized state with great universities that produced just one new physics teacher last year – just one. Increasing the supply of talent into teaching will require in addition to major changes in policy, a change in the way teachers are perceived. TFA and similar programs are beginning to bring that about – but only beginning to do so.

Changing a culture also requires leadership from those in government, business, and the not-for-profit sectors. The chances of success would be greatly enhanced if teacher leaders too became advocates of reform. In the early 1990s, when this author worked for a teacher union in England, he suggested turning the traditional case for investment in teachers on its head. Instead of continuing to argue that if the government increased our pay we

might consider improving the system, this author suggested that accountability should be embraced, the system improved, and then the public told, "Look, now you can see a system worth investing in."

Accepting this case, at the time, required too great a leap of faith. However, history is instructive. In spite of teacher opposition, accountability was imposed, schools did improve, and, as a result, the biggest ever increases in investment in education – including in teachers' pay and professional development – did follow. Blair called it "Investment for Reform." While teachers did not always enjoy the journey, they arrived at a much better, if more challenging, destination. Given the wide range of opportunities available, talented people will not flock into a profession with lockstep conditions and a beleaguered image. Nor will skeptical citizens continue to invest precious tax dollars in a system that does not seem to be working. Citizens the world over, similar to good businesses, prefer to invest in success.

This raises the sixth point for discussion; the extent and distribution of funding for public schools. International benchmarks suggest that America's overall expenditure on schools is above the average but, compared to other countries, two questions of distribution stand out. The first is that, even after funding equity suits, often more money per student is spent in wealthier areas than poorer ones. It is easy to see how this came about; but to an outside observer in the early twenty-first century, this disparity makes no sense at all. If all young people are to reach high standards, as NCLB envisages, then the system has to provide greater support to those with the furthest to go. A child of age 4 of professional parents will have heard 45 million words; a child aged 4 years of welfare parents just 13 million words. Overcoming such a massive language deficit is possible but not if the school system merely reinforces the advantages of the already advantaged. The inescapable practical implication is that, over time, states will have to play a larger role in raising and distributing funds for public education.

US funding distribution is also skewed in another way which receives less comment. A much lower proportion of it actually reaches the classroom than in the best-performing systems; much more of it is tied up in administration. Of course, good administration matters but every dollar spent on unnecessary administration is a dollar that could have assisted that welfare child reach for the stars.

In England, the districts were required, by law, to devolve the vast bulk of the funding for schools to the schools themselves. Each school now has a 3-year delegated budget, based on a published formula. This direction was set by the Thatcher government and continued by all its successors. In effect, the burden of proof was reversed – a pound (or dollar) should be spent at the level of the school unless there is a convincing case for spending it elsewhere.

Any government in England that sought to return to the old opaque system would have a riot on its hands.

Along with the money, schools need to be given responsibility for how it is spent. The PISA evidence shows that increased management autonomy at school level is associated with better results but this lesson remains to be learnt in many parts of the country. Recently, in a northern US city, this author came across a good school principal setting out to turn around a failing school. She had all the right ideas but had no control over which teachers were employed at her school. She even needed permission from a teacher to visit a classroom. What chance did she have? The contrast with the nearby charter school, in similar social circumstances, was dramatic. Accountability and autonomy need to go together. The question for the United States is not just how many charter schools it wants but how soon all schools can have charter-like autonomy.

This leads to the seventh and final point of this article. The PISA evidence indicates strongly that there are benefits in holding schools to account through standards-based external assessments. In the global economy, the question of national standards inescapably arises. The emerging bipartisan alliance in favor of common or national – as distinct from federal – standards suggests growing recognition that the United States needs them to be fewer, clearer, and higher. Moreover, *Education Next's* recent poll showed ~70% support in the population for national standards – and over half of all public school teachers agreed.

The truth is that, across the world, standards in math, science, and English will inevitably be set by global benchmarks in a globalized economy. Quite simply, to succeed, countries will need world-class standards; algebra and geometry do not change at the Rio Grande or the 49th Parallel.

From this perspective, the question of national standards is straightforward; they will arrive anyway eventually. The only questions are whether they do so by accident or design; haphazardly or systematically; sooner or later. As they used to say in the civil rights decade, “If not now, when?”

Conclusion

The choice facing the United States is, as this article began by saying, a stark one. It will shape America's capacity to succeed over the decades ahead in the profoundly challenging global economy; it will shape too whether the American Dream in a stunningly diverse America, is genuinely open to every citizen. This is why progressive business leaders and civil rights leaders stand together. They know that the aspirations unlocked in the dawn of the civil rights movement have only partially been fulfilled. The school system has improved in many

ways since that time but surely no American can be satisfied by the outcomes. There is no ceiling on what individual Americans can achieve but, though enshrined in legislation, the high floor – on which economic success and social justice depend – has yet to be built. For those committed to a vibrant, successful America where the American Dream and the American reality more closely coincide, there can surely be, in Martin Luther King's phrase of old, “neither rest, nor tranquility.” It will be a long, hard road. Many of us – in the arena now – know this only too well. More and more people are watching us, and hoping.

Counting on your success in this endeavor are not just children and families across this great country, not just the future of the American economy, not just the idea of the American Dream, but all of us around the world. In the twentieth century, a strong, generous, outward-looking America was a decisive factor in enabling humanity to rise and meet the challenges it faced. As Goldin and Katz have demonstrated, public education was central to making that possible. How much more important for everyone then is US education in the twenty-first century when the world is so much more complicated and the clock is ticking?

At the bottom of the staircase in No.10, just outside the Prime Minister's office, there is an exceptional photograph of Winston Churchill. Facing the camera, he glowers with such defiance that, even at that uncertain hour, war-time defeat must have seemed inconceivable to the onlooker. In fact, one is told the real cause of his mighty frown is that the photographer had forced him to put down his cigar! Be that as it may, the time has surely come to heed his famous words, “America always does the right thing but only after it has exhausted all the alternatives.” In education reform, those alternatives have, indeed, been exhausted. It is time for America to do the right thing.

Acknowledgment

This article was prepared as a speech given by the author at Aspen Institute's Education Summit in Washington DC, 15 September 2008.

Further Reading

- Baldwin, J. (1998). *Collected Essays*. New York: Library of America.
- Barber, M. (2008). *Instruction to Deliver: Fighting to Transform Britain's Public Services*. London: Methuen.
- Barber, M. and Mourshed, M. (2007). *How the World's Best Performing School Systems Come Out on Top*. New York: McKinsey.
- Branch, T. (1988). *Parting the Waters: America in the King Years, 1954–63*. New York: Simon and Schuster.
- Branch, T. (1998). *Pillar of Fire: America in the King Years, 1963–65*. New York: Simon and Schuster.

- Branch, T. (2007). *At Canaan's Edge: America in the King Years, 1965–68*. New York: Simon and Schuster.
- Dallek, R. (2004). *John F. Kennedy: An Unfinished Life*. London: Penguin.
- Goldin, C. and Katz, L. (2008). *The Race between Education and Technology*. Cambridge, MA: Harvard University Press.
- Hanushek, E., Jamison, D., Jamison, E., and Woessman, L. (2008). Education and economic growth. *Education Next* 8(2), 62–70.
- Howell, W., et al. (2008). Americans speak out. *Education Next*.
- Loomis, C. (2006). The tragedy of general motors. *Fortune*. February 6.
- New York Times Editorial* (12 August 2008).
- OECD (2007). *Pisa Report*.
- Roosevelt, T. (2004). *Theodore Roosevelt: Letters and Speeches*. In Auchincloss, L. (ed.). New York: Library of America.
- Schleicher, A. (2008). *Seeing the US School Systems through the Prism of International Benchmarks, Presentation to the Hunt Institute*.
- Thompson, T. and Barnes, R. (2007). *Beyond NCLB: Fulfilling the Promise to Our Nation's Children*. Washington, DC: The Aspen Institute.

School Effectiveness in Developed Societies

D Reynolds, University of Plymouth, Plymouth, UK

© 2010 Elsevier Ltd. All rights reserved.

The History of School Effectiveness

School effectiveness (SE) is a relatively recent educational discipline that has grown in many developed societies at a very rapid pace. It wishes to understand the functioning of schools, and their effects on their students. Some members of the discipline wish to go further than this and directly try to improve school functioning. The discipline set up its own organization to promote research – The International Congress of School Effectiveness and Improvement (ICSEI) only in 1988 – and its own journal, *School Effectiveness and School Improvement*, only in 1990. Reviews of the field are available in Teddlie and Reynolds (2000), and Sammons (1999).

The general intellectual, and particular policy, setting that saw the rise of SE was not a particularly hospitable one. In the 1960s, there had been a widespread belief that education could not compensate for society, and that the failure of the liberal education reforms of enhanced expenditure and school organizational change to achieve the hopes of their proponents showed that schools made no difference.

However, beginning in the United States, where it was initially a practitioner movement but became more research orientated, and in the United Kingdom, where it came initially from the research community, the field is now represented in probably 40 or 50 developed societies across the world, and in some developing ones too. In the United States, the School Board Superintendent, Edmonds (1979) made the best-known initial contributions, and in the United Kingdom, the researchers Michael Rutter (Rutter *et al.*, 1979) and David Reynolds (1976) were responsible for its initial offerings.

The range in the particular contributions to SE knowledge and practice from different societies is marked. The United Kingdom brought a well-developed focus, from its well-established sociology of the school within its sociology of education, upon school-level organizational characteristics and also was responsible for the generation of much of the methodological revolution in statistical methods that occurred in the mid-1980s. Its school improvement (SI) perceptions and links were also well developed, and virtually all SE work has been notable for the very wide range of student outcomes that have been measured in addition to the core one used internationally for academic achievement.

The United States has been notable for the high quantity and quality of its work done on effective teaching, and

for the associated successful teacher improvement programs of the 1970s and 1980s. It has also pioneered the emphasis upon any contextual specificity of what may be effective practices, with its sampling strategies that attempted to look at schools from diverse socioeconomic environments (Teddlie and Stringfield, 1997), unlike the United Kingdom which has sampled almost exclusively from disadvantaged communities because of its predominant equity concerns. The United States has also been notable for the development of special improvement programs to implement SE and other bodies of knowledge, as with the initial Effective Schools Programs of the 1980s and the federally funded Special Strategies Programs of the 1990s.

Elsewhere in North America, Canada (Fullan, 1991) has been notable for the development of improvement practice more than for SE research, although it has evidenced numerous useful studies on SE situated within the quantitative, organizational paradigms. The Scandinavian societies too have been mostly improvement orientated with particularly the generation of culturally sophisticated programs from Norway.

In Europe, the biggest contribution other than from the United Kingdom has come from the Netherlands, where interestingly a large volume of quantitatively orientated research studies have been paralleled with attempts to generate middle-range theories of school functioning. Dutch SE material has also been quite sophisticated at the classroom level, reflecting probably the historical popularity of learning and instruction perspectives from European social psychology within that country.

Elsewhere in the world, Australia has been notable for its rare attempt in the early 1990s to survey what key stakeholders, including pupils and parents, thought was effective, and has evidenced useful teaching effectiveness material also.

While in some parts of the developed world SE has yet to make its mark in any major fashion (Italy, Germany, and France evidence little penetration of the SE paradigm), virtually all the developed societies of the world participate in the studies that emanate from the IEA, which involve the ranking of societies on their educational effectiveness through use of their students' test scores and related attempts to explain the resulting international differences. Studies such as the Third International Mathematics and Science Study (TIMSS) from the mid-1990s and the Programme for the Evaluation of Student Achievement (PESA) from the mid-2000s have taken the

SI paradigm to many parts of the world, further encouraging its development.

It has perhaps been the enhanced educational focus upon the need for educational quality, necessary for any developed society to compete in a world where productive forces are subject to the pressures of globalization, which has spurred increased interest in SE from researchers, practitioners, and, especially, governments. Related concerns about educational efficiency, value for money, accountability, choice, and educational evaluation have also created a fertile ground for the discipline to grow. The associated enhanced governmental interest across the world in better management of schools, the professional development of teachers, and educational innovation/change more generally have also been potent for the development of knowledge upon good schools and how to make all schools good that is the intellectual and practical core of SE, although there have been critics in many developed societies who have criticized SE's contribution to what is seen as a narrow managerialism, a discourse of performativity, and a climate of blaming of the educational system.

The Methodology of SE

SE took a number of its core methodological characteristics from the American educational finance, educational production function, and educational administration specialisms that were already in existence in the early 1970s, particularly its core focus upon student inputs and how these were acted upon by school organizational processes to become educational outputs. In this framework, inputs are seen as the students and their characteristics in terms of their levels of prior achievement, their family background, and the nature of their communities, as they enter different schools.

Outputs are seen as their characteristics and attributes as they leave, in terms of their academic achievement, their social/affective characteristics, and, in some formulations, their psychological attributes. Therefore, public examination results, achievement test scores, levels of self-esteem, propensity for criminal behavior, school attendance rates, levels of problematic behavior in schools, and even levels of post-school employment have all been used as output measures. The concern in many societies to widen the mission of the schools into the tackling of economic and social disadvantages directly is stretching the array of those measures used in SE studies even more.

The third area where SE concentrates upon is that of school processes. Most SE research has looked at the formal organizational characteristics of schools in the search for factors to explain their variation in impacts, rather than at the cultural or informal characteristics. The leadership strategy and methods of the headteacher/principal, the

academic organization of the school, its pastoral/welfare organization, its facilities/resources, its involvement with parents and the other outside school stakeholders, its strategies for motivating children, and the methods, efficiency, and effectiveness of its classroom teaching/instructional level have all been major foci of interest.

Early methods of analysis in the SE of the 1970s and early 1980s involved simple multiple regression analysis in which outputs were regressed on inputs, with the resulting residuals (or differences in how schools actually performed compared to their inputs predictions of how they should do) being related to their organizational processes noted above. However, from the mid-1980s the use of multilevel modeling (MLM) became widespread, and is now the axiomatic method, used internationally. These methods permit the analysis of multiple levels of the education system with specific estimates of the effects of variance in, for example, the school level, the classroom level, and the district/local authority levels, permitting the disaggregating of the educational system into its component parts that simple regression analyses could not achieve.

However, it has been pointed out that the estimates attained using multilevel analyses are not dissimilar from those obtained using conventional regression analyses, and that multilevel methods also do not permit the study of interactions between levels that may be important in certain areas (e.g., school/classroom interactions). In addition, multilevel methods presuppose direct paths of educational influences upon outcomes, while multiple, interactive, and indirect relationships also seem to be likely. For these reasons, enthusiasm is now turning toward structural equation modeling (SEM) as a tool of analysis that gives advantages in the areas dealt with problematically by the multilevel methods mentioned above.

Whatever their precise nature, all these statistical methods adopt a value-added approach, whereby schools' outputs are not used raw but in which schools are assessed by the value they have added to their inputs.

Key Findings of SE

From the large number of SE studies conducted – 2000 reviewed in Teddlie and Reynolds (2000) and probably 3000 in existence now – a number of clear trends can be seen.

First, the influence of schools upon their student outcomes is always considerably less than that of their students' prior achievement and/or their family/community backgrounds, with perhaps 10–15% of variation in student achievement explained by the school and classroom levels together. For certain subjects, where exposure is usually exclusively from within the educational system, such as mathematics, educational effects are greater. Educational

effects are also greater for younger-age children and for those from socially disadvantaged areas. While schools do make a difference, in SE analyses, they do not make the major difference.

Second, school performance can be varied over time. Schools vary in their effectiveness level year on year, and some schools improve rapidly consequent on their organizational changes, particularly when generated by a new principal/headteacher. Nevertheless, probably two-thirds of any countries' schools are average over time, defined in terms of performing where they should be, with only a limited proportion adding considerably more or less value than expected. However, in many developed societies, there is a substantial size group of negative value-added, low-performing schools that have perplexed policy-makers and practitioners by their consistent poor performance and by their inability to respond to conventional school-improvement attempts. These schools in challenging circumstances form a contemporary focus for SE in many countries.

Third, there is a degree of variation as to how different kinds of students perform even within the same school. This will apply to boys/girls, ethnic groups, children of different achievement levels, and children of different social backgrounds: schools will be differentially effective for their subgroups of students. More recently, particularly in the United Kingdom, there is increasing discussion of within-school-variation (WSV) in the effectiveness of the different subject areas that children study, with some analysis suggesting that WSV dwarfs school-against-school variation by 4 or 5 times.

Lastly, there is also a degree of variation in the effectiveness of different schools on their different kinds of outcomes. Social or affective outcomes seem to be partially independent of academic/cognitive outcomes, and to be associated with different educational processes within the school/classroom.

The Process of Effective Schooling

In spite of the clear difficulties about making easy, global judgments as to what works in schools that are consequent on the substantive findings above, this has not stopped the generation of multiple descriptions of effective schools and of multiple visions about how to make all schools effective schools (see reviews in Reynolds *et al.*, 1996; Reynolds *et al.*, 1994).

Factors such as the varying financial resource levels between, and within, societies do not seem to determine whether a school is effective or not. Rather, it seems to be the ways in which schools actually use their resourcing that matters, although there are some hints that resource effects may be greater for schools with more challenging students. Building age, school-site quality, and the

quantity of educational technologies likewise do not seem important – it is site upkeep and the patterns of use of their plant and educational equipment that seem to be the keys to effectiveness.

The crucial factor in determining effectiveness is the quality of the leadership provided by the principal/headteacher, with effective schools possessing leaders who can set institutional directions, who have a clarity in their mission, who motivate and involve both senior management and their line teachers, who acquire cultural and financial support from external stakeholders, and who use educational change from outside the school to broker changes in the instructional attitudes and behaviors of the teaching force within it. Such effective leadership crucially appears to be focused on the classroom – effective leaders manage by walking about (MBWA) so they are aware of their teachers' strengths and any needs for support. They communicate high expectations about the importance of teaching, use programs of continuing professional development to potentiate it, and are particularly concerned with maximizing the quality of their teacher supply.

It may seem a truism, but effective schools also possess effective teaching. Class time is well managed, with no loss of instructional time due to lessons running over their allotted times or to time-wasting disciplinary interventions. There is a climate of high expectations about what students can achieve, and an academic press that generates high achievement. Lessons have clear purpose, are related to prior learning, and have clear, well-managed structuring. Classroom discipline is both firm and fair, with rules that are understood and clear to all. The students are involved in the learning process, both through generating and discussing curriculum content and through developing as reflexive learners who help determine their teaching and learning methods. Assessment data are used both summatively – to promote feedback – and formatively, to improve learning.

The students are involved in the school, as well as in the classrooms of effective schools. The school will possess multiple clubs, societies, sports teams, and associations to multiply the chances of student involvement. Prefect/monitorial systems and/or consultation systems and committees will be used to generate further involvement. The effective school will be incorporative of its student clients.

Parents will also be involved in the effective school both formally in systems of parent/teacher associations and more informally through the possession of a school ethos that encourages parents to become involved in lessons as helpers, in fundraising and in generally visiting the school. Information will be fully provided to parents.

The effective school itself will be intelligent in its use of data and in its management information system (MIS). A wide range of data will be used to track students, with data systems that are sensitive enough to quickly pick up any variation in how students are doing over time and in

different subjects. Such systems also usually involve the prediction of performance and the calculation of actual performance, with appropriate interventions to optimize outcomes. They may also involve the collation – at individual, class/teacher, and year level – of additional student data or student voice data that both motivates individual students through encouraging reflexivity and also provides the authorities of the school with valuable consumer-response material that can be used to check on the quality and reliability of their systems as seen from the learning level.

From SE to SI

Although there now clearly exists considerable knowledge about the characteristics of effective schools, the application of that knowledge to improve schools has been somewhat patchy across the developed countries of the world, with little evidence of any consistent improvement through utilizing it. In the United States, SE boomed and busted in the 1980s, when the early five-factor theories were tried in many schools and districts, but the knowledge base was unsophisticated and attention moved on to more transformational approaches. In the United Kingdom, SE informed the policies of the Labour Government from 1997, being reflected in the national strategies that ran from 1999 and in the number of SE researchers that advised government. However, this phase of prescription of what works was not matched by any subsequent improvement additional to the initial surge in UK national educational test scores that occurred from 1999 to 2002. In most other societies, SE has had a less-pervasive impact within their educational systems, although in perhaps 40 or 50 developed countries across the world there has been some take-up of SE practices at society, district and/or individual school level, often associated with the influence of SE researchers in their higher education establishments.

Part of the reason for limited take-up has been the historic separation between the activities of SE researchers and SI practitioners/researchers, which particularly meant that SE lacked some of the SI understandings that might have helped it to achieve greater reach. SI had been concerned more with school culture than school organization, was particularly keen to involve teachers in determining what should happen in any school being improved, and, in its most advanced formulations, was full of insight about the complex socio-psychological currency within schools that needed to be addressed to actually institutionalize improvement.

SE's more managerialist and organizational paradigm made it, perhaps, the paradigm of choice for the principal and the headteacher more than of the classroom teacher, and perhaps made it the policymakers' friend rather than that of any broader constituency of educationalists.

Contemporary Developments in SE

Perhaps the most likely reason for the somewhat limited take-up of SE, and its correspondingly limited effects, was that, as a knowledge base it was not particularly sophisticated in its first flowerings in the early to mid-1990s. Interestingly, since then a great deal of progress – intellectual and practical – has been made in a number of areas, virtually all of which have direct implications for education policies and for the SI attempts of practitioners.

First, it is now clear from recent SE research that the range of organizational variation within schools is considerably greater than that between schools when value-added assessments are made, by a scale of 3–4 times. In secondary schools, this can be seen in the variation in the academic outcomes of subject departments – in primary schools this can be seen in variation between the outcomes of individual teachers. While the reasons for the existence of these variations may be unclear – the effect of initial training or later professional development opportunities have been suggested as possibilities – the use of this variation to improve schools promises a stronger educational lever than whole-school reform.

These within-school differences are within the same organization, so should be amenable to easy diagnosis and remediation. Students in the different subject areas, or different classes in primary/elementary schools, will be relatively the same in their attributes, making alibis or excuses from teachers for any failure less likely. Schools may also be more able to develop a positive tone about improvement based upon studying their own best practice, rather than relying on outside school sources of improvement.

Second, attention in SE research has shifted away from the school organizational level to that of the classroom, with most analyses showing that the classroom, instructional, or learning level explains much more variation in student outcomes than the school level. The teaching behaviors and characteristics that potentiate classroom learning seem to be quite large in number – 50 or 60 in most formulations – and consequently effective teaching appears to be the result of doing a large number of things well, rather than doing correctly a smaller number of major things. Effective teachers seem to have both the appropriate behaviors and related associated beliefs, which appear to have additive effects, as in the case of the importance of factors such as a positive locus of control, positive attitudes to students, and strong belief in the power of education.

Third, SE has begun to explore the variation in what works by context that might affect commitments to what are effective educational practices. Building on the American work noted earlier, variations in the socioeconomic status of schools' catchment areas, their rural/urban status, effectiveness level, improvement trajectory,

governance structure, and, indeed, their country have all been shown to have implications for what is needed to generate effectiveness.

Conclusions and Future Research/ Practice Needs

Even though there has been rapid progress in some areas of SE, there are numerous other disciplinary areas that await attention. There are very few attempts to create theories about why certain educational factors, or groups of factors, have effects. Many of the new school outcomes that are regarded as essential in a world where simple knowledge is rapidly redundant, such as learning-to-learn skills, and some of the psychosocial attributes necessary in a more-stressful world for young people, such as personal resilience, for example, have not received the place in the SE knowledge base that they need. We still know very little about which of the SE correlates travel across cultural contexts, and why.

However, the increased pressure on the educational systems of developed societies internationally, consequent upon pressure on public expenditure across virtually all of them, will ensure that issues concerning educational efficiency and effectiveness will continue to be high profile, creating further opportunities for SE research. The rise of the new cognitive neuroscience paradigm is likely to create further knowledge about how student learning can be potentiated. The increased number of experiments of nature in which national governments try to change the governance, processes, and outcomes of their educational systems is creating an international laboratory of education. The prospects for SE could not be brighter.

Bibliography

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership* 37, 5–11.

- Fullan, M. G. (1991). *The New Meaning of Educational Change*. London: Cassell.
- Reynolds, D. (1976). The delinquent school. In Hammersley, M. and Woods, P. (eds.) *The Process of Schooling*, pp 217–229. London: Routledge and Kegan Paul.
- Reynolds, D., Creemers, B. P. M., Stringfield, S., et al. (1994). *Advances in School Effectiveness Research and Practice*. Oxford: Pergamon Press.
- Reynolds, D., Sammons, P., Stoll, L., Barber, B., and Hillman, J. (1996). School effectiveness and school improvement in the United Kingdom. *School Effectiveness and School Improvement* 7(2), 133–158.
- Rutter, M., Maughan, B., Mortimore, P., Ouston, J., and Smith, A. (1979). *Fifteen Thousand Hours: Secondary Schools and Their Effects on Children*. London/Boston, MA: Open Books/Harvard University Press.
- Sammons, P. (1999). *School Effectiveness Coming of Age in the 21st Century*. Lisse: Swets and Zeitlinger.
- Teddlie, C. and Reynolds, D. (2000). *The International Handbook of School Effectiveness Research*. London: Falmer Press.
- Teddlie, C. and Stringfield, S. (1993). *Schools Make a Difference: Lessons Learned from A Ten Year Study of School Effects*. New York: Teachers College Press.

Further Reading

- Campbell, R. J., Kyriakides, L., Muijs, R. D., and Robinson, W. (2004). *Assessing Teacher Effectiveness: A Differential Model*. London: Routledge Falmer.
- Creemers, B. P. M. (1994). *The Effective Classroom*. London: Cassell.
- Creemers, B. P. M. and Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness*. London: Routledge.
- Fitz-Gibbon, C. T. (1996). *Monitoring Education: Indicators, Quality and Effectiveness*. London: Cassell.
- Goldstein, H. (1980). *Multilevel Statistical Models*, 3rd edn. London: Arnold.
- Harris, A. (2008). *Distributed School Leadership*. London: Routledge.
- Hopkins, D. (2001). *School Improvement for Real*. London: Routledge Falmer.
- Muijs, D. and Reynolds, D. (2005). *Effective Teaching*, 2nd edn. London: Sage.
- Reynolds, D., Creemers, B. P. M., Stringfield, S., Teddlie, C., and Schaffer, G. (eds.) (2002). *World Class Schools: International Perspectives in School Effectiveness*. London: Routledge Falmer.
- Scheerens, J. and Bosker, R. (1997). *The Foundations of School Effectiveness*. Oxford: Pergamon Press.
- Teddlie, C. and Stringfield, S. (1993). *Schools Make a Difference: Lessons Learned from A Ten Year Study of School Effects*. New York: Teachers College Press.

School to School Collaboration: Innovation and Improvement

R Higham, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

A hypothesis that has long held appeal is that there are solutions elsewhere in the education system for challenges faced in individual schools. We have not always known, however, where these solutions are or how best to apply them into other contexts. Recent advances may have made such aims a little more realistic. Over the last decade or so, teachers have become increasingly expected and/or willing to be observed, to observe others, and to discuss their pedagogic practice. Schools commonly use comparative data in their own self-evaluation. A wide range of local partnerships and networks bring teachers and leaders together to innovate and share ideas. We may also be finding out more about how schools actually work collaboratively. Approaches under the banner of knowledge transfer, for instance, have not always been found appropriate in collaborative settings. Schools working together collegially may also be co-constructing new practices to common challenges (Fielding *et al.*, 2005).

Within this emerging context, this article explores the work of school-to-school collaboration as a lever for innovation and mutual improvement. To do so, it focuses on the Leading Edge Partnership Program (LEPP) as a case study for analysis. The article progresses in several main ways. First, it considers the emergence of school-to-school collaboration in policy and practice in England. Second, it describes the nature and work of Leading Edge partnerships. Third, it outlines the main collaborative advantages achieved by a sample of Leading Edge partnerships. Fourth, it explores the collaborative methodologies these partnerships have employed to do so. Fifth, it analyses the additional factors that appear important in developing a collaborative impact on student outcomes. Finally, the article concludes by arguing that where an ethos for mutual improvement remains undeveloped, collaborative advantage will be difficult to achieve. In partnerships where progress is made, it appears to be correlated with strong leadership and organization, the full involvement of partner schools and increasingly rigorous methodologies for collaborative work.

A Recent History of School Collaboration in England

It is estimated that nearly all schools in England are now involved in some form of networking (Hill, 2006). This would have been hard to anticipate during the late 1990s. A New Labour government came to power in England in

1997 having coined the now famous slogan 'Education, Education, Education'. The promise was for educational renewal aimed at improving school standards and social equity simultaneously. In reality, at the systemic level, New Labour's main approach was to evolve the more radical reforms introduced in the 1980s under Margaret Thatcher, albeit in a context of significantly higher public spending. Thatcherism had introduced an educational market to unleash competition-driven improvements. New Labour sought to advance market effectiveness through greater parental choice and institutional differentiation by curriculum specialisms. Similarly, where the Conservatives had introduced the National Curriculum and national tests, New Labour developed accompanying National Strategies that summarized a range of pedagogic approaches and promoted a minimum set that schools were strongly encouraged to implement.

It was at the more specific level of schools facing challenging circumstances that New Labour initially developed a range of funded initiatives aimed at school collaboration. The Excellence in Cities program, for instance, sought to share capacity for teaching, learning, and community engagement across urban schools. The Leadership Incentive Grant aimed to strengthen leadership in schools through collaborative professional development and mentoring. A growing number of specialist schools were expected to work with other schools to spread good practice and raise standards. This might have appeared contradictory to the broader approach of market-led competition. But this was New Labour and its third-way philosophy that prioritized eclectic pragmatism over ideological chastity. The forces of market competition were to be combined with collaboration and the sharing of best practice.

By 2003, the government felt sufficiently confident to argue that there were now system-wide benefits to schools working in partnership (DfES, 2003: 12). A range of reasons exist for this. First, faced with the limits of command-and-control policies, as well as the system inequalities that can result from competition, networks held the appeal of greater professional engagement, lateral working, and system coherence (Glatter, 2003). Second, there was an existing need for joint working in practice to resolve interconnected challenges in, for example, student welfare (Connolly and James, 2006). Third, this was to some extent a rediscovery of the professional collaborations that had been facilitated by local education authorities in the 1970s and 1980s (Glatter, 1995; Stevenson, 2007).

A decade or so on from 1997, there is a growing sense of the potential benefits of collaboration between schools as well as with other agencies. Different outcomes relate to local aims, activities, and ways of working. However, a range of studies and evaluations have found a relatively high degree of similarity in the benefits that can be achieved where partnerships are effective. These are summarized in **Table 1**.

Another common message is that achieving these benefits is far from easy. As Vangen and Huxham (2003: 62) found, reports of unmitigated success are not common. Instead, partnerships can develop collaborative inertia in which only hard-fought or negligible progress is made. In part, this results from the continuing coexistence of collaboration and competition. At the extreme, for instance, there remains “a spirit of intense competition and mutual suspicion” (Arnold, 2006: 6). This inevitably has an impact on the ability and desire of schools to overcome other common obstacles to collaboration. These include, as set out in **Table 2**, time, resource constraints, and distance. It also becomes more difficult to develop practices commonly associated with effective partnerships, including building trust, a willingness to compromise, and bearing the costs of partnership work usually before benefits occur.

What is significant is that, in many places in England, obstacles to collaboration appear to be being surmounted and a new interconnectedness developed. In research with urban schools, Ainscow and West (2006: 137) found head-teachers were beginning to identify shared principles around which their staff could be drawn together. This was generating a new impetus for change across schools, wider ownership of the improvement agenda, and reduced

polarization. They concluded that, despite the longer-term deepening of socioeconomic inequalities by market reforms, there are reasons for optimism as “the system has considerable untapped potential to improve itself” (Ainscow and West, 2006: 131). This exists in the accumulated skills, knowledge, and creativity within and between schools and their local communities.

How these skills and resources might be mobilized to support innovation and improvement beyond the individual school remains, however, a challenge. Indeed it is unclear what ways of working and collaborative leadership might best enable such work (Glatter and Harvey, 2006). This article seeks to consider these issues with reference to the LEPP. After introducing LEPP, it goes on to consider what:

- collaborative advantage is actually achieved in a sample of partnerships;
- methodologies these schools developed and deploy to work together; and
- other key factors related to a partnership’s effectiveness.

Leading Edge

The LEPP was launched by the government in 2003. It was a successor to the Beacon Schools Programme that had existed since 1998. The initial program had funded high-achieving schools to disseminate aspects of their practice to other schools, but usually without any sustained support. LEPP sought to learn from the limitations of this approach by demanding a clearly defined

Table 1 Commonly quoted benefits of collaboration

Good practice:	share effective practice or expertise identify shared problems and work collaboratively on solutions
Professional development:	provide mutual and informed support enhance quality of staff development and critical reflection joint staffing, wider career structures, solving staffing shortages improve leadership quality and support and/or whole school systems
Direct student benefits:	wider curriculum choice and learning pathways improved transition of pupils into secondary school raised student expectations and (in some cases) attainment
Local strategic leadership:	increase equity and reduce polarisation of schools promote coherent provision for local communities increased community involvement ensure the survival of rural schools
Resources:	drawing in additional funding and resources developing efficiency and economies of scale reduce risk and uncertainties of innovation and new projects

Table 2 Commonly quoted factors supporting effective collaboration and obstacles to it.

Factors supporting effective collaboration

Ethos:	Trust, honesty, respect, openness; a sense of joint ownership, with different views taken into account; staff values supporting co-operation; ability to compromise while seeing collaboration in one’s interests
Leadership:	Senior leadership commitment; a clearly identified and realist focus that is predominantly shared; monitoring and evaluation of progress
Activities:	A degree of consensus on the methods that will lead to success; purpose directly connected with needs of specific learners; a focus on goals that individual partners could not achieve alone
<i>Obstacles to effective collaboration</i>	
Resources:	time / distance; lack of funding; costs often occurring before benefits
Leadership:	apprehension of staff not acknowledged; poor communication; silo mentality; unwillingness to negotiate sovereignty
History:	a culture of competition; difficulty of working across old structures

local partnership in which a lead school and several partner schools could work together.

LEPP now connects about 200 lead secondary schools with about 800 partners. Lead schools receive around £60 000 per annum from the government on behalf of their partnership. It is intended that these funds are used to support both innovation and the raising of standards in partner schools. When spread across an average partnership size of four or five schools, this does not constitute a large contribution to school budgets. However, it does enable schools to seed new projects, backfill staff working out of school, and/or to take on coordination costs. With over 200 partnerships, covering about a third of secondary schools in England, LEPP represents a relatively significant systemic commitment.

In 2006, the central leadership of the program was migrated out of the government (and specifically the Innovation Unit) to the Specialist Schools and Academies Trust (SSAT). The Innovation Unit had, perhaps unusually for the government, allowed schools to develop projects organically without much guidance. The SSAT sought to combine this initial mandate with slightly more accountability for partnership progress and a clearer central offer, including national conferences, workshops, online tools, and visits by consultant heads.

This article draws primarily on research with five LEPP partnerships. The research included 1 day at the lead school with semistructured interviews with the headteacher and at least five members of staff and half a day each at two partner schools with interviews with the headteacher, a head of department, and a teacher. Where possible, a collaborative activity was also observed. To test the wider generalizability of the findings, a survey of the 32 consultant heads who make annual visits to all 200 partnerships was conducted. An earlier draft of this article was also presented to a workshop of about 40 school leaders at the 2007 LEPP annual conference in both York and London.

Collaborative Advantage

The partnerships were engaged in a wide range of work, but there were also several important similarities. These included:

- a collaborative ethos, with the majority of lead and partner schools reporting positively on each other's commitment to partnership working and their own willingness to contribute to the partnership's development;
- evidence of trust, with many of the partnerships demonstrating among a majority of schools a relatively high level of trust in one another to deliver their strategic and/or operational responsibilities;
- a history to partnership membership, with previous collaborative work often having taken place between a

number of the schools and with the explicit inclusion of one or more historically lower-achieving schools; and

- a coexistence of the partnership, with a range of other local collaborations that were emerging or ongoing, but that did not necessarily include all of the partnership's members.

There were also a range of positive impacts for professionals from collaborative work in the majority of partnerships. In increasing degrees of collaborative depth, these included:

- improved professional support, with the majority of teachers reporting the benefits of an extended network of curricular-relevant assistance, and heads of department finding a means to overcome isolation;
- increased opportunities to gain inspiration and pool ideas for practice with colleagues in an informal setting and based on professional dialog and judgment;
- professional development in the workplace with the opportunity to generate a wider learning community among staff, and support and coaching for leaders at a variety of levels; and
- the collaborative development of new practice, with not only the transfer or refinement of what was already perceived to be good practice but also the joint identification of challenges and research and innovation to develop responsive practices.

These latter innovative practices coalesced around five main categories:

- The development of new teaching and learning strategies and the use of information and communication technology (ICT) as a support for learning with, for example, podcasts posted on virtual learning environments for gifted and talented students, catch-up, and/or revision.
- Assessment for learning and the use of student data, with work to enable students to access and review their own progress and teachers to analyze attainment data to identify learning needs and review teaching strategies.
- Wider learning pathways for greater student enjoyment and interest. For instance, the collaborative provision of a substantial 14–16-year-old work-related learning offer and vocational qualifications.
- Support, guidance, and re-engagement for students with, for instance, students undertaking learning walks to develop a document for teachers on what they perceived to be effective teaching and learning.
- School organization becoming more responsive to student needs with, for instance, new transition arrangements and a primary link of teacher accompanying student to their secondary school to provide support and parent liaison.

There was some evidence of an impact of these activities on student outcomes. Two forms were most common. First,

a contribution to whole school improvement in a partner school that had previously been underachieving. Where this had occurred, the key strategies stemming from collaborative work that appeared to be most significant were: student tracking and assessment for learning; wider curriculum pathways; leadership support and mentoring; and approaches to challenging low expectations. Second, improved attainment, behavior, attendance, and/or engagement by students in a target group. In these instances, the strategies that had been most effective included: targeting students due to under-performance; supported and mentored small groups; and re-engaged student in new curriculum pathways.

There were some important caveats, however, to these perceived impacts. Identifiable progress was rarely partnership-wide, with usually only some schools moving forward. Where progress had occurred, Leading Edge work was often a contributing factor but it would have been difficult to demonstrate its proportional impact. Where a specific group of students had been targeted, it was rare for all to have moved forward, with (more usually) progress for a majority but less impact for a minority. There was also a range of work for which no improvement in student attainment could be demonstrated. Perhaps the most important reason for this was that the partnerships had initially viewed LEPP as not being bound to the strictures of short-term increases in standards. They had included longer-term challenges or objectives more amenable to professional judgment rather than those judged solely by increases in student attainment (at least in the short term).

Transferring Knowledge and Innovating Collaboratively

Across this range of collaborative work, two main methodological approaches were apparent. The first concerned the direct transfer of practice from a lead to a partner school. This was usually as a direct lever for school improvement, especially in partnerships with lower-achieving schools. This had strong parallels to what Fielding *et al.* (2005) termed the “delivery of ‘validated’ packages of pre-formed practice seen by others to be good for the recipient.” Fielding and co-workers critique such transfer as having “little validity amongst teachers.” However, in the majority of cases where this was occurring, the partner school’s leadership team had actively sought out such support from their lead school. The latter was often trusted and seen as more capable of providing effective solutions than other external advisers. Moreover, rather than simply being a one-off transfer of knowledge, the structure provided by the partnership often meant that this was an ongoing process that included a period of refinement in which transferred practice was developed into the new context. Lead schools also reported benefiting

from such work by, in several cases, reincorporating new ideas back into their own practice.

These were examples of success. There were cases where knowledge transfer had not been effective or sustained in practice. There were also obstacles to transfer. These included:

- a partner school pulling away from collaboration when it experienced difficulties;
- a partner school’s staff not wanting to receive transferred practice and/or considering (rightly or wrongly) the lead school’s practice to (only) be effective when serving students from different (usually more advantaged) backgrounds; and
- the lead school not actively seeking to engage in knowledge transfer and/or the partnership’s culture being at odds with such work.

The second and more typical approach to partnership working was collaborative innovation. This fits well with Fielding *et al.*’s (2005) alternative concept of co-construction and the useful notion of a mutuality of the process of innovation. Here, a range of methods was being employed, including:

- A series of informal planning meetings, with staff from each school researching and discussing new ideas before returning to their classrooms to try new approaches, with feedback given in subsequent meetings.
- Learning walks, as inspired by the Pittsburgh Institute, which included the processes of identifying a focus, undertaking a walk with observation and dialog, giving feedback, and undertaking a project informed by the insights gained. These were used both to look inwards at departmental practices and to system gaze for new ideas and inspiration in other schools.
- Formal research, with teachers encouraged to undertake in-school research to diploma and master’s levels under supervision from university colleagues. Staff often chose subjects of personal interest with a link to the school development plan.
- Innovation trios, where three teachers involved with the same teaching group, but in different subjects, worked to build more effective teaching strategies for the group and support each other in doing so.
- Conferences for staff to conclude and disseminate findings to other staff at the end of a term or project strand.

Across these methods, evaluation of the effectiveness of new practices was predominantly based on professional judgment. This was variously described as: a gut feeling during teaching; the reaction of students as to whether they were engaged; and professional dialog at review meetings. Some went further to structure the use of student tracking into projects as well as more formal reporting. However, a common question many leaders asked was how to maintain authentic professional engagement, and

informal networking, while bringing greater rigor into their approach. In particular, their concern was for evaluating the impact on student outcomes and documenting innovation so that it could be shared at a school level. With some degree of inductive development, **Figure 1** summarizes an emerging consensus across several partnerships about what greater rigor might look like.

The process of collaborative innovation is likely to be far messier in practice. It is also governed by professional judgment and thus the proposed cycle is not conceived as a strict flow but, rather, as a guide with openness to steps being rearranged or leading backwards. Each step is seen, however, to have its own value, with:

- analysis of information – bringing together professional judgment, contextual data, student voice, and/or research evidence to jointly interrogate current practice across the partnership;
- identification of the focus or target group – with collaborative development of aims and methods, be that at a whole-school, departmental, or classroom level, but with a common theme across the partners;
- joint planning of strategies – with time for collaborating teachers and school leaders to come together to plan strategies or interventions;
- experimental implementation in one's own classroom/school – with project members trialing and adapting the innovations in their own contexts;
- peer observation and feedback – with opportunities for individuals to observe each other and discuss their experimental implementation;
- feedback from students – with students engaged either as respondents (through surveys or questioning) or as researchers themselves;
- (back to) analysis of information – with this completion of the cycle focusing on reflection and evaluation of implementation; and

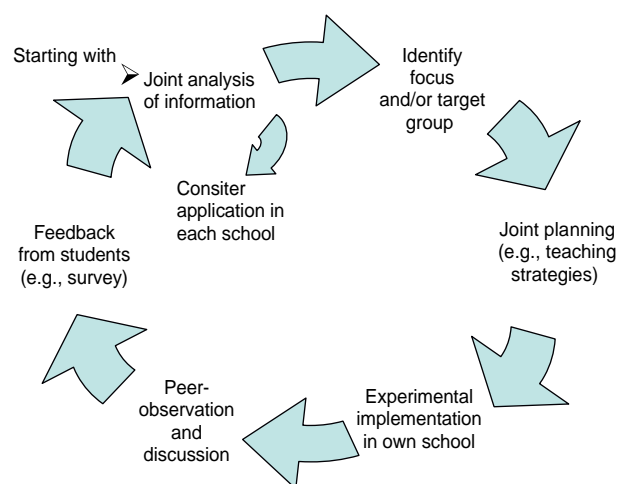


Figure 1 Innovating collaboratively.

- (and then, potentially) a wider applicability within each school – with the sharing of innovation with colleagues and/or the development of new whole-school practices.

This approach, or similar approaches in practice, corresponds well with the research evidence on action research. Kemmis and McTaggart (1988), for instance, proposed a circular planner for action research which comprises the steps: plan, act, observe, and reflect (and back to plan again). (McKernan (1996) proposes a similar but more detailed cycle of: define problem, needs assessment, hypothesize ideas, develop action plan, implement plan, evaluate action, and make decisions (reflect, explain, and understand action).)

Hopkins (2008) proposes a set of principles to guide classroom research by teachers that include:

- that research should not disrupt a commitment to teaching as a primary focus;
- data collection must not be too demanding on teachers' time;
- the methodology employed must be reliable enough to enable the formation of hypotheses and to develop strategies for classroom practice; and
- the research focus should be one to which the teacher is committed.

Leading Innovation Partnerships

These guidelines would have found acceptance among a majority of Leading Edge partnerships. However, there were at least four further factors that appeared important in developing a collaborative impact on student outcomes.

Building Capacity

The partnerships identified a number of collaborative capacity-building components that underpinned their joint work. These included:

- purposefully deploying staff motivated to lead the innovation process;
- supportive relationships, both informal and formal, nurtured by partners;
- well-established systems and procedures for collaboration; and
- the alignment of internal professional development to support collaborative innovation.

Collective practice also required senior leaders to commit existing resources, particularly in the initial years. Delegated funding from government covered some of this, but in the majority of cases, schools also needed to deploy their own resources. This took the form of staff time and goodwill, the use of whole-school in-service

training days, the provision of specific training courses, and, inevitably, staff knowledge, skills, and leadership resources. These investments came with a level of risk given the reliance on other partners. Indeed, each partnership had experienced events that depleted the emerging capacity among schools. These included:

- the unexpected shocks of key staff lost to long-term illness or to career progression elsewhere;
- the unraveling of projects following the arrival of a new headteacher with different priorities;
- the energy drained by a lack of full engagement by one or more partners; and
- the diminishing of trust from the souring of relationships.

New Models of Leadership

In response, the challenge a number of leaders faced, particularly in lead schools, was to become more flexible, more able to engage externally, and less vulnerable to staff turnover or changing priorities. In several more-advanced partnerships, the response could be described as bringing together three relatively common concepts in educational leadership:

- system leadership that, in these contexts, concerned leading beyond an individual school to contribute student outcomes at the wider local level;
- distributed leadership to effectively “orchestrate the skills of colleagues” (Arnold, 2006: 7), draw them into the decision-making process, and build their ability to take on wider leadership roles at the whole-school level;
- leadership of learning with the focus this implies on the classroom and “what constitutes effective learning . . . and how teaching and new technologies are best harnessed to support learning” (Stoll, 2001: 6).

The way these concepts were commonly being brought together is summarized in **Figure 2**.

Lateral Accountability

While leadership development was important in sustaining internal school capacity and partnership working, collaboration had also presented the challenge of holding partners to account. A common concern was that doing so could damage longer-term relationships. These issues were often complex and had emerged over time, including:

- the apathy of one or more partner schools, through absenteeism at meetings and events, or a gradual unwillingness to fulfill agreed actions; and
- to a lesser extent, concern about the ethos of the partnership, including the approach and focus of work, a perceived lack of support from the lead school, or

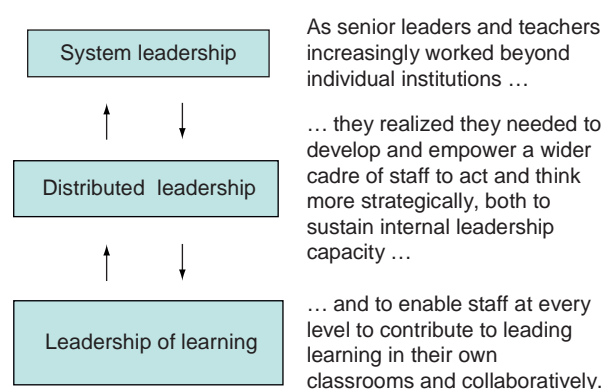


Figure 2 Leadership for collaborative capacity building.

uncertainty about how LEPP funding was being deployed by the lead.

For several partnerships, the solution lay in clearer lateral accountability. Invariably, this meant clearer expectations on issues of engagement and on developing benchmarks for progress. The strategies being developed by partnership leaders, either individually in lead schools or collectively with partner schools, included:

- more flexible involvement in specific activities, with not all schools needing to engage in every project;
- clarification of expectations with, at the most unambiguous, a minimum number of agreed yearly activities to be undertaken by each school;
- the holding back of LEPP funding by the lead school where specific actions had not been completed by individual partner schools; and
- inactive partners being left out when the partnership changed focus or moved to a new phase (e.g., following redesignation of the lead school).

Partnership Ethos

Perhaps the most powerful form of lateral accountability, however, was a widely shared collaborative ethos for mutual improvement. This was the conviction that partnerships should be collegial and have an impact on the identified needs of all partners. This did not imply an abdication of leadership by the lead schools. In most cases, they set the agenda for the partnership, usually collaboratively. It did concern moral purpose and a willingness to share expertise and resources, particularly with partners in need. This is not solely about the distribution of delegated funds. Rather it equates more directly to whether a partnership enables the full involvement and ownership by partner schools, how capacity for progress is built in practice, and, ultimately, whether student outcomes are improved.

Conclusion: Collaboration and Improvement

Enhancing learning and teaching remains a key priority for schools. Trends toward personalizing learning to individual student needs and interests, coupled with a greater responsibility for student well-being, represent real challenges for schools and their leaders as they attempt to continue to raise school standards and offer a broad, balanced, and engaging education. This article has explored how school-to-school collaboration might offer part of the solution. In particular, it appears collaboration can help to:

- build professional learning communities, within and across schools, that develop and widen learning and teaching strategies to respond to a range of student learning needs;
- use the full potential of school workforces to share and innovate practice, create new curricular and learning pathways, and extend services;
- develop new models of leadership to appropriately distribute an increasing range of responsibilities to a wider and differentiated pool of expertise; and
- impact on student outcomes, either for a targeted group of students or at a whole-school level, particularly in lower-achieving schools.

In partnerships where there is less willingness to work for the success of other partners, such collaborative advantages remain difficult to achieve. In cases where there is a lack of leadership capability or whole-school capacity, collaborative progress can be made, but it may often be a harder task and/or less widespread across the partnership. Where progress is made, it appears to be correlated with more advanced partnership organization, the fuller involvement of partner schools, and increasingly rigorous methodologies for collaboration. In the case of Leading Edge, the thrust for such effective collaboration is usually provided by the lead school's leadership team. However, partner schools are also involved, either in what one might call system working, as a spur to improve their own provision, or more directly in collaborative leadership.

See also: Action Research as a Tool for Teachers' Professional Development; Networked Learning Communities School-to-School Collaboration as an Essential Component of a System Reform Strategy; Partnerships Between Schools and Higher Education; System Leadership; The Educational Leader's New Work.

Bibliography

- Ainscow, M. and West, M. (2006). Drawing out the lessons: Leadership and collaboration. In Ainscow, M. and West, M. (eds.) *Improving Urban Schools*, pp 130–144. Berkshire: Oxford University Press.
- Arnold, R. (2006). Schools in collaboration: Federations, collegiates and partnerships. *EMIE, report 86*. Slough: National Foundation for Educational.

- Connolly, M. and James, C. (2006). Collaboration for school improvement: A resources dependency and institutional framework of analysis. *Educational Management, Administration and Leadership* 34(1), 69–87.
- Department for Education and Skills (DfES) (2003). *A New Specialist System: Transforming Secondary Education*. Nottingham: DfES Publications.
- Fielding, M., Bragg, S., Craig, J., et al. (2005). *Factors Influencing the Transfer of Good Practice*. Nottinghamshire: DfES Publications.
- Glatter, R. (1995). Partnership in the market model: Is it dying? In Macbeth, A., McCreath, D., and Aitchison, J. (eds.) *Collaborate or Compete? Educational Partnerships in a Market Economy*, pp 25–35. London: Falmer Press.
- Glatter, R. (2003). Collaboration, collaboration, collaboration: The origins and implications of a policy. *Management in Education* 17, 10–16.
- Glatter, R. and Harvey, J. (2006). *New Models of Headship Varieties of Shared Headship: A Preliminary Exploration*. Nottingham: NCSL.
- Hill, R. (2006). *Leadership that Lasts: Sustainable School Leadership in the 21st Century*. London: ASCL.
- Hopkins, D. (2008). *A Teacher's Guide to Classroom Research*, 4th edn. Maidenhead: Open University Press.
- Kemmis, S. and McTaggart, R. (1988). *The Action Research Planner*, 3rd edn. Geelong: Deakin University.
- Stevenson, H. (2007). Improvement through collaboration and competition – can the government have it both ways? *Management in Education* 21(1), 29–33.
- Stoll, L. (2001). *Enhancing Internal Capacity: Leadership for Learning*. <https://forms.ncsl.org.uk/media/604/DF/enhancing-internal-capacity.pdf> (accessed September 2009).
- Vangen, S. and Huxham, C. (2003). Enacting leadership for collaborative advantage: Dilemmas of ideology and pragmatism in the activities of partnership managers. *British Journal of Management* 14, 61–76.

Further Reading

- Atkinson, M., Springate, I., Johnson, F., and Halsey, K. (2007). *Inter-School Collaboration: A Literature Review*. Slough: NFER. <http://www.nfer.ac.uk/research-areas/pims-data/summaries/review-of-collaboration-between-schools.cfm> (accessed September 2009).
- Cordingley, P. and Bell, M. (2007). *Transferring Learning and Taking Innovation to Scale*. Nottingham: Innovation Unit.
- Evans, E., Castle, F., Cooper, D., Glatter, R., and Woods, P. (2005). Collaboration: The big new idea for school improvement? *Journal of Education Policy* 20(2), 223–235.
- Fullan, M. (2004). *Leadership & Sustainability: System Thinkers in Action*. London: Sage.
- Hargreaves, D. (2003). *Education Epidemic: Transforming Secondary Schools through Innovation Networks*. London: DEMOS.
- Hill, R. and Matthews, P. (2008). *Schools Leading Schools: The Power and Potential of National Leaders of Education*. Nottingham: NCSL.
- Huxham, C. and Vangen, S. (2005). *Managing to Collaborate: The Theory and Practice of Collaborative Advantage*. London: Routledge.
- O'Leary, D. and Craig, J. (2007). *System Leadership: Lessons from the Literature*. Nottingham: NCSL.
- Rudd, P., Lines, A., Schagen, S., Smith, R., and Reakes, A. (2004). *Partnership Approaches to Sharing Best Practice*. Slough: NFER.
- Spender, B. (2006). Linking up for success. *Curriculum Briefing* 4(2), 28–31.

Relevant websites

- <http://www.teachernet.gov.uk/management/tsp/primary/toolkit/develop/viability/fedcollab/collab/>
- <http://www.ledge.org.uk> – Leading Edge Partnership Programme.
- <http://www.ncsl.org.uk> – National College for School Leadership (New Models of Leadership Project).

Studies of School Improvement in Developing Countries

S E Anderson, University of Toronto, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

School effectiveness and improvement is an international preoccupation. Parents everywhere pin their hopes for the future economic well-being of their children on their participation in formal schooling. National governments see participation in primary and secondary education as vital to economic development, social welfare, and civic engagement. International agencies, such as the United Nations Educational Scientific and Cultural Organization (UNESCO), the World Bank, and many government sponsored and nongovernmental international development-aid organizations direct attention to the role of schools not only as key contributors to regional economic and social development, but also as critical players in the resolution of critical cross-national issues of poverty reduction, public health, and human rights. Concerns about school improvement arise not merely from common beliefs about the role and importance of schools in ensuring the quality of life, but also from the recognition of wide variation in the degree to which all schools and school systems are fulfilling these public expectations.

This article provides an overview of some key historical trends in approaches to and in the study of school improvement in developing countries over the past half century, including: basic resource-capacity building; the adoption and implementation of changes in curriculum content, materials and teaching strategies; changing the support system for improvement; decentralization and school-based management (SBM); school choice; and comprehensive school reforms. The article focuses on the knowledge base for improvement at the school level. It does not address country-level indicators of education quality and improvement, such as those generated by aggregate scores of student academic performance on international tests of literacy and numeracy, or through national statistics on access, attendance, retention, and completion rates. Nor does it take up initiatives focused primarily on improvement in initial-teacher-education programs.

Basic Resource-Capacity Building

Although there are disparities in material and human-resource inputs among schools in the developed world, basic inputs are essentially guaranteed in both public- and private-sector schools. Students are not sitting on the ground in the open air or forced to share desks with 50–100 peers in poorly ventilated classrooms originally built for 30 students. The students have access to textbooks and a supply of supplementary learning materials (e.g., school

libraries). They are not learning simply by copying and reciting what the teacher says or writes on a chalkboard that serves as the primary material instrument of teaching and learning. Sanitary toilets and running water are available for the children and teachers. The teachers have more than a primary or secondary school education, and the vast majority are professionally trained and certified to teach. Teachers are regularly paid. While conditions have improved since the middle of the twentieth century, in many developing-world contexts, school-improvement efforts are still heavily focused on building adequate facilities, getting appropriate and sufficient teaching and learning materials into the hands of teachers and students, and making sure that teachers have sufficient subject-matter content knowledge and teaching skills to guide the students' learning. Unfortunately, this situation has been exacerbated by international efforts to achieve universal access to free quality primary education by the year 2015. In Uganda, for example, primary school enrolment more than doubled from 3 million to 7.3 million after the government adopted its Universal Primary Education policy, while in Malawi, enrolment rose from 1.9 million to 2.9 million in response to the government's free primary education policy in the 1990s.

Comparative research in developing countries on student participation and achievement in schools that vary in the availability and quality of these kinds of inputs show, indeed, that the provision of basic material and human-resource inputs does make a difference to school effectiveness (Fuller, 1987; Lockheed, 1993; Reynolds, 2000). While satisfaction of basic resource needs is a key factor in the overall level of student learning in schools, it does not guarantee high-quality student performance, or equity in the quality of performance across schools within a school system. The problem and challenge then becomes one of finding ways to continue to elevate the standard and levels of student learning for greater numbers of students and more and more schools. The remaining approaches to school improvement reviewed here highlight strategies intended to enhance the quality of teaching and learning beyond that attained by ensuring basic resources.

Adoption and Implementation of Changes in Curriculum and Instruction

This approach to school improvement is premised on the idea that the quality of learning outcomes is a function of

the quality of the curriculum content, teaching and learning materials, and instructional practices that teachers put into practice. Thus, the way to make schools more effective is to replace or supplement existing curriculum content, materials, or teaching methods with others thought to be of higher quality, or perhaps a better fit to local needs and contexts. The new instructional programs, materials, and methods are typically developed and disseminated by external experts (government agencies, textbook publishers, university professors, aid agencies, etc.). In theory, the quality of education in schools will incrementally and cumulatively improve as a result of implementing new programs and practices often limited to particular subjects or grades at a given time. Scholars of educational change characterize this as the innovation adoption and implementation approach to change (e.g., Fullan, 2001).

Research on the process and outcomes of innovations in curriculum and teaching in schools demonstrates that neither the form an innovation takes in practice nor its effects on student learning can be predicted from the characteristics of the product or practice as originally adopted. Some innovations never actually get implemented, especially when support for their use in the form of teacher training and assistance, material resources, funding, time, and supervision is lacking or poorly planned and executed. Typically, when new programs are put into practice, the implementing educators modify the innovations in response to unique needs and conditions of their local context or to more closely resemble what they were doing before. Education change theorists refer to this as mutual adaptation. Researchers have also found that when innovations are implemented, there is no guarantee that their use will be sustained over time. The lessons of research on both positive and negative instances of innovation implementation in schools are well reviewed and reported (e.g., Fullan, 2001; Hall and Hord, 2006). The key to successful implementation of changes in curriculum content, materials, and teaching methods lies partly in the quality and practicality of the innovations and their fit with local needs and context, but most importantly, in the effectiveness of the support provided by school-system authorities and other external change agents to those expected to put the innovations into practice (e.g., resources, initial and continuing training, supervision, and incentives). In studies of innovation adoption and implementation, improvement is often equated with evidence that the changes are being used as per the developer or policymaker's intent, rather than with reliable measures of impact on student learning at either the individual student or school levels. That said, the adoption and implementation of discrete new programs and practices are fixtures in the educational change and school improvement landscape, although often embedded within other conceptions and approaches to school improvement as described in subsequent sections of this article.

A distinct history of educational innovation and research in the developing world began in the 1970s (Havelock and Huberman, 1978). Early educational improvement work in the developing world often exemplified the innovation adoption and implementation approach to change (e.g., Verspoor, 1989; Lewin and Stuart, 1991). Verspoor (1989), for example, analyzed 21 high-outcome education improvement projects selected from a sample of 282 World-Bank-funded projects between 1963 and 1984 in Africa, Latin America, South Asia, and Asia. His findings echoed findings on innovation implementation elsewhere. In-service teacher training, for example, was critical to implementation success, and effective teacher-training designs provided local training (at the school or school-cluster level) to support implementation of change programs, with regular classroom visits, support, and assistance from supervisors and program staff. Some of Verspoor's findings foreshadowed later trends in approaches to school improvement. More successful projects combined project assistance with interventions to develop the capacity of education systems to plan, coordinate, monitor, and support innovation on a continuous basis. He also found that change projects that focused on a single component of education provision (e.g., curriculum, teacher training, school management, and material resources) had lower outcomes than projects that addressed multiple components in a coordinated manner. Finally, Verspoor described four pathways to change: progressive innovation (multiple linked innovations introduced over a period of time); incremental expansion (innovation introduced in a small number of settings with gradual increase in the scope of use); discrete change (isolated projects in limited settings and uncoordinated with other projects); and permanent pilot (an experimental program that never gains political support and resources for broader use). While not discounting the benefits for the few schools involved, discrete change and pilot projects remain a common phenomenon in the developing world, often linked to efforts by nongovernmental organizations (NGOs) to stimulate and model innovations in curriculum, teaching, and learning (e.g., Anderson, 2002). This has contributed to a major preoccupation by governments and aid agencies with how to scale up and sustain successful small-scale changes across the larger system.

Changing the Support System for Innovation

The realization that change and improvement do not happen simply by presenting school personnel with new programs, materials, and introductory training workshops, has led government and other education-change agents in developing countries to experiment with alternative organizational systems for supporting the implementation of change in curriculum and instruction. These efforts to

change the ways innovations in teaching and learning are professionally supported represent another approach to school improvement, though they are often instituted in conjunction with particular pedagogical change initiatives. Many of these changes focus on alternative in-service teacher-development processes, such as school-based action research, peer coaching, mentoring of new teachers, or structured opportunities for joint work among teachers focused on sustaining and improving the quality of teaching and learning (see Shaeffer, 1990, for overview of participatory approaches to teacher development and the section titled 'Further reading' for examples of research on specific strategies in developing-country contexts). These kinds of changes often do not require change in organizational policies and structures so much as a shift from reliance on periodic external workshops and courses to more school-based and collegial means and norms of ongoing professional learning. In-service teacher-development projects involving nontraditional teacher-learning strategies are often reported as discrete initiatives affecting a few schools, rather than as school district or broader-system initiatives with reliable indicators of their impact system wide on school improvement.

Other support system changes involve the creation and implementation of new organizational structures designed to enable more sustained introductory and follow-up support to school personnel involved in change. In developing countries, intermediate levels of educational organization similar to school districts in the United States and Canada, or local education authorities in England, often operate as a bureau of the municipal government whose primary purpose is to manage the flow of funds, resources, and information from central government agencies to schools, to hire and appoint school personnel, to carry out formal school-supervision procedures on behalf of the government, and to oversee public-school examination processes. They often do not have the professional capacity or explicit function of support for school improvement. In rural areas, the working relationships and communication between district and school educators may be further complicated by poorly kept roads and lack of transportation.

Two organizational support system innovations that have been experimented with intermittently in developing countries over the past quarter of a century are teacher centers and school clusters. Teacher centers are distinct facilities typically staffed by one or more government-appointed master teachers. They are often set up in unused portions of a school. The centers serve as sites for workshops and other professional-development events run or organized by the center staff for schools within a designated catchment area accessible to neighboring schools by foot or public transportation. They are typically expected to maintain professional libraries for teachers, and may provide materials and facilities for teachers to

come individually or in groups to create learning materials. Teacher centers are administered by local management committees that represent the schools and any partner organizations (e.g., local teacher's college). Teacher centers were popularized in East Africa with support from the British Council in the 1980s and 1990s. In Kenya, for example, several hundred primary school teacher-advisory centers were created, though the central government has had persistent difficulty developing and sustaining the operational capacity of the centers to support schools (for an overview of the East African experience with teacher centers see Knamiller *et al.*, 2000). The school-cluster approach is similar. It involves a local network of neighboring schools which are expected to work together on an ongoing basis on curriculum and teacher-development initiatives. One of the schools, however, is typically designated as a lead or coordinating school for organizing and hosting cluster activities, and for facilitating contacts with external administrative and support services from government agencies, teacher-training institutions, and education-focused nongovernmental aid organizations (for an example of a school cluster-based support system see Tsang and Wheeler, 1993).

The cascade model for introducing large-scale changes in curriculum and teaching at the national level has been commonly practiced and reported in developing countries. In this approach, governments set up and implement temporary systems through which a core group of experts involved in the development of the programmatic change at the national level introduce and train regional teams in the content of the change and expectations for practice. The regional teams are responsible for replicating the introductory training processes with selected trainers at a more local level, who then deliver the introductory training to key school personnel, which finally cascades down to program delivery by these key teachers to the targeted implementers in the school (e.g., for an application of this model see Ibrahim, 1991). While the cascade approach works in a way as a mobilization strategy for delivering and creating some basic awareness and knowledge of new programs and practices at a large scale, the emphasis is on getting an innovation into the hands of those expected to put it into practice, not on providing ongoing support and expertise during implementation. The model needs to be combined with local organizational designs for continuing support and expert assistance at the school level.

In sum, as noted by Verspoor early in the history of school improvement efforts in developing-world countries, one of the major obstacles to improvement has not simply been the provision of basic resources, and high-quality programs and teaching methods, but also the creation of sustainable support systems capable of providing effective professional assistance to schools. Some of the difficulties of integrating new support system

positions and organizational structures into existing government-regulated administrative structures and norms at the local level are illustrated in a book of case studies of school district-level school-improvement initiatives sponsored by an international NGO in East Africa (Anderson, 2002).

Decentralization and SBM

Decentralization trends are premised on the belief that centralized control and regulation of economic activity and social services by government authorities inhibits organizational effectiveness. Decentralization policies have been strongly advocated by some international donor agencies that are key actors in providing aid to developing countries, such as the World Bank. Decentralization trends extend beyond education to all sectors of government-regulated economic and social services activity. In education, the push toward decentralization has manifested itself in two key ways. One is to adopt policies and organizational arrangements that devolve more decision-making authority for school management, improvement, and in some cases, funding, to the local level. The second is to create a more competitive marketplace of schools characterized by greater school choice and privatization (see the next section of this article).

The SBM approach to school improvement, often referred to as a form of restructuring, became popular in the 1980s and 1990s. Proponents of this approach argued that the needs and circumstances of schools were too varied to mandate a single solution and plan for improvement, and that schools were constrained from innovation by education bureaucracies and policies that required standardization in structure and practice (for overviews of the restructuring movement see Elmore, 1990). They further held that the solution was to allow school-level stakeholders (e.g., educators, parents, and taxpayers) greater freedom (and responsibility) to determine their needs for improvement, as well as greater control over the use of resources (e.g., money, staffing, and time) and organizational arrangements (e.g., professional roles and duties, decision processes, and grouping of students and teachers).

In North America and other developed countries, advocates of SBM have argued that this approach honors the professional expertise of those working closest to the problems of teaching and learning to determine what and how to improve. SBM policies and initiatives typically consist in requiring schools (often on a voluntary basis) to establish school teams of administrators, teachers, and parents to produce school-improvement plans. Sometimes, this is accompanied by waivers that permit the schools to opt out of certain school district and government-policy regulations. Government policies mandating the establishment of school councils dominated by parent and

community representatives, not by teachers and principals, are another common approach to SBM. The logic of improvement differs from the professionalization arguments, because this strategy shifts greater power and authority over school decision making to school clients not to education professionals. In practice, many of these policies, at least in North American schools, only assign an advisory role to school councils, reserving management decisions (e.g., hiring principals and teachers or allocating school budget) and educational decisions (e.g., school goals or program change) to professional educators. A decade of experimentation and research on SBM in North American schools yielded little evidence that SBM strategies in and of themselves made any significant difference in school effectiveness in the short or long term on a recurrent basis (Leithwood and Menzies, 1998).

SBM initiatives in the developing world have placed more emphasis on strengthening parent and community influence on school decision making than on expanding the scope of authority of school personnel (for syntheses of research see Task Force on Education and Gender Equity, 2006, United Nations Millennium Project). The arguments for increased local control center on increasing parent and community pressure on school personnel to carry out their duties in a professional manner, enhancing parental commitment to sending the children to school, and strengthening parental involvement in managing and perhaps supplementing school funding. This typically takes the form of school-management committees that are formally empowered with decision-making authority over staffing (hiring and firing of principals and teachers), teacher supervision, school finances, school maintenance, and to a lesser degree, curriculum and instruction. When the right to hire and fire teachers (and principals) shifts from education authorities to school parents, research on decentralization effects shows improvements in teacher attendance. When parents take responsibility for the management and upkeep of their children's school, they are more likely to ensure regular attendance. Use of funds for the purchase of school resources is more likely to happen when parents have ways to ensure that principals are accountable for funding. It stands to reason that student achievement will improve if teachers are consistently on the job teaching, if students regularly attend classes, and if school funds are transparently spent on essential teaching and learning resources. It is not the act of decentralization, however, that improves schools, but the quality of implementation at the local level, including supports such as management training for school committees. Schools operating under SBM systems can be run just as ineffectively as schools operating under more centralized policies and systems. Finally, there is little research on the impact that decentralization through SBM has on school quality and improvement over time. Local parent/community management may help raise the baseline level

of school performance, but does not guarantee ongoing improvement unless combined with measures that more directly affect the quality of teaching and learning.

School Choice and Privatization

School choice is a cover term that encompasses an array of strategies and policies based on the premise that competition for clients (students and families) among schools would compel school personnel to find ways to get better at what they do. The literature on school choice identifies two key policy strategies:

1. Public choice refers to the provision of greater choice among schools in the public sector through such mechanisms as open enrolment among schools within or between school districts; access to alternative schools or programs serving students with special needs or interests (e.g., arts or technology); and parallel publicly funded school systems based on religious preferences or language of instruction.
2. Market choice refers to the use of funding mechanisms such as educational vouchers and tax credits to create competition between public and private school providers for students and government funding. Of these, voucher programs are the most widely known, implemented, and debated. The idea is that a government allocates a voucher worth a set amount of dollars per child, that parents choose where to send their children from an open market of public and privately managed schools, and that government funding follows the child to the chosen school. In theory, a voucher system would lead schools to compete for students and dollars in order to survive and thrive in the education marketplace. School-choice policies are silent, however, on what mechanism and knowledge schools might use to respond to the pressures to improve.

School-choice programs have been widely debated, but only implemented on a large scale in a relatively small number of developing and developed countries (see Plank and Sykes, 2003, for overview of choice policies and case studies from Chile, South Africa, China, New Zealand, Australia, England, Sweden, the Czech Republic, and Hungary). In Chile, for example, since the 1980s, the government has funded a national voucher program whereby a set amount of money is allocated per pupil, and students can elect to attend either public schools or private schools (except those that charge tuition). Funds are allocated based on enrolments and actual attendance. Findings from research on the implementation and outcomes of strategies for school choice (e.g., vouchers or open boundaries) internationally are ambiguous and do not generally support the claim that increased market-oriented choice strategies lead to more effective schools

overall. It is not uncommon to find that schools of choice modestly outperform and cost less than mainstream public schools (the former may not offer same pay scale and benefits to teachers as public schools); however, the effects of greater choice are found to exacerbate the problems of low-performing schools often serving the most socioeconomically disadvantaged communities. More academically motivated students (whose parents tend to be more educated and to have higher aspirations for their children) and wealthier families are most likely to exercise choice options, leaving the source schools with an even greater proportion of hard-to-serve students, with fewer funds to do it, and with reputations guaranteed to deter recruitment of quality teachers. Competition for admissions may also lead higher-performing schools to restrict entry to the academically strongest applicants, creating an illusion of school improvement that is really more a function of student selection than of organizational strategies to improve existing practices.

The issues of choice are more salient and complicated in developing countries where extremely large numbers of private schools may account for a substantial proportion of available primary and secondary schools (e.g., see Srivastava and Walford (2007) for an overview of issues and studies of private schooling in developing countries). Tooley and Dixon (2007), for example, share findings from a comparative study of private-education provisions in low-income communities in Nigeria, Ghana, India, and Kenya. They report startling statistics on the number of children enrolled in private schools, including surprisingly large numbers of nonsubsidized and nonregistered schools. In the Lagos State, Nigeria, for example, 34.3% (185) of schools in low-income areas were publicly funded government schools, while 65.75% (355) were private unaided schools. In Hyderabad, India, 65% of children in low-income areas were reportedly attending private unaided schools. These numbers pale in comparison to the situation in Bangladesh where a parallel system of private schools in low-income communities run by a large NGO (Bangladesh Rural Advancement Committee (BRAC)) operates over 30 000 community schools without government aid. While some developing countries have created policies to encourage and even subsidize private-school alternatives to government-run public schools, this is happening more because the public demand for universal primary and secondary education outstrips the government financial and resource capacity to provide, not as a school-improvement strategy. Comparative studies of outcomes and costs of government and private schools tend to support the case for higher-quality outcomes in the private-sector schools. Some would argue that the main role of government should be limited to establishing common expectations for student learning outcomes, and to creating school accreditation and accountability systems, leaving responsibility for the actual delivery (and perhaps even

funding) to private providers and their clients. The moral issues and controversy surrounding the growth of private education providers (regulated and unregulated, for profit and philanthropic) are too complicated to attempt to cover here. It will suffice to note that some observers maintain that government-run school systems in many developing countries have proven themselves persistently ineffective, inefficient, and incapable of achieving quality education on a large scale, and that the only viable alternative to improving the quality of education is to support the increasing privatization of primary and secondary education. Complete privatization, thus, represents a more drastic form of decentralization than SBM of public schools, and a more drastic version of choice than that envisioned and embodied in voucher and government-subsidized private-school policies and options.

Comprehensive School Reform

The comprehensive school-reform (CSR) movement emerged in the United States in the 1990s (Murphy and Datnow, 2003). The basic idea is that change has to be whole school, involving all participants, and the improvement process needs to encompass a comprehensive array of interrelated factors – for example, program (curriculum, materials, and program structure), teaching methods, school management and leadership, teacher development, funding, parent involvement, monitoring, and evaluation (echoing Verspoor's arguments from back in the 1980s). A CSR model provides a blueprint, process, and support system for school development across multiple components of school governance, teaching, and learning. The models identify themselves with a brand name, such as Success for All, Accelerated Schools, the Coalition of Essential Schools (Murphy and Datnow, 2003). In the CSR approach to school improvement, individual schools voluntarily choose to adopt and implement a CSR model (which may come with requirements for additional funding). School personnel receive implementation training, on-site assistance, and program materials from the model developers, which may be NGOs or based in universities. CSR schools become members of networks of schools using the same model, and the model developers facilitate information sharing and assistance across the network. The CSR developers and networks function like a kind of alternative but dispersed school system. The technical capacity and expertise of formal school district or government education agencies to programmatically support these schools, however, is unclear. While the evaluation research evidence shows variable impact of different CSR models, the CSR approach to school improvement fares better than the other approaches to school improvement reviewed here when outcomes are considered on a school-by-school basis. As

an approach to improvement at the school-system level, however, it is problematic due to its reliance on voluntary participation by individual schools, and due to its uncertain fit with government-mandated program expectations and government-managed support and accountability systems for improvement.

Independent of these developments in the US, similar programs of CSR with similar tales of success have evolved in various regions of the developing world, though descriptive and evaluative research on these models of schooling and school improvement is not readily accessible (see Further Reading for specific reports on the examples listed below). The two best known are *Escuela Nueva*, which began in Colombia in 1979 and has since been replicated in Central America, and the BRAC network of community-based schools in Bangladesh. Other CSR models have emerged elsewhere in the developing world. The Aga Khan Foundation, for example, sponsored the evolution of a whole school-improvement program (SIP) through a series of school-improvement initiatives in East Africa beginning in the mid-1980s and extending into the current century (Anderson, 2002). In Northern Pakistan, faculty from the Aga Khan University Institute for Educational Development created a Whole School Improvement Program (WSIP) that operates in a network of government and NGO-sponsored schools in predominantly rural and remote mountain communities. An international NGO, Save the Children, has sponsored community schools' initiatives in Mali, Malawi, Ethiopia, and Uganda since 1988. In Chile, *La Fundación Chile* has sponsored the implementation of a CSR model called *Mejor Escuela* in about 50 high-poverty low-performing schools with clear gains in student outcomes in participating schools. Farrell (2002) provides a comparative perspective on the emergence of alternative models of schooling that have yielded positive student-learning results in these and other regions of the developing world. In sum, the CSR approach to school improvement represents another active and promising approach to school improvement in the developing world. CSR programs and networks, however, typically depend on external nongovernmental funding sources and technical assistance (even when implemented in publicly funded schools, not just when set up as independent schools). Cooperation and coordination between public school systems and the CSR networks can be politically, financially, and logistically problematic.

Conclusion

During the latter half of the twentieth century, many developing-world countries were just beginning to create mass public-education systems. The predominant initial concern in these contexts has been to develop the material (e.g., school buildings, curriculum, and instructional

materials) and human-resource capacity (e.g., trained teachers, educational managers, and supervisors) to deliver public primary and secondary education on a large scale, and to extend access to that education to all children (e.g., urban and rural, boys and girls, rich and poor, and majority and minority cultural/linguistic groups). Resolution of the fundamental material and human-resource-capacity issues remains an enduring challenge in many areas of the world in the context of ongoing national efforts to extend universal access to primary and secondary education, and in the face of regional conflicts and other humanitarian crises (e.g., the human immunodeficiency virus (HIV) AIDS epidemic in Africa or natural disasters) that disrupt the continuity and provision of basic education to children on a large scale.

In addition to investments in the basic infrastructure for and access to primary and secondary school education, national governments with assistance from international and local stakeholders and donor agencies have undertaken various policies and programmatic approaches to improve the quality of primary and secondary education. This article provides an overview of six common approaches to school improvement in developing countries: basic resource-capacity building, adoption and implementation of changes in curriculum and instruction, changing the support system for innovation, decentralization and SBM, school choice and privatization, and the creation and implementation of CSR models. While these are conceptually distinct approaches to improvement, in practice they frequently overlap. CSR models, for example, require the adoption and new programs and practices (the innovation adoption and implementation approach). School choice and privatization overlap with policy trends toward decentralization. Changes in organizational support systems for change are necessary for effective implementation of programmatic changes in curriculum and teaching, and different strategies may be required to provide effective assistance for continuous improvement in centralized and decentralized systems. While research on school improvement has yielded a substantial knowledge base for how to improve teaching and learning at a classroom and on a school-by-school basis, application of this knowledge at a large scale remains a daunting and ongoing political and practical challenge.

Bibliography

- Anderson, S. E. (ed.) (2002). *Improving Schools through Teacher Development: Case Studies of the Aga Khan Foundation Projects in East Africa*. Lisse, NL: Swets and Zeitlinger.
- Elmore, R. (1990). *Restructuring Schools: The Next Generation of Educational Reform*. San Francisco, CA: Jossey-Bass.
- Farrell, J. (2002). The Aga Khan Foundation experience compared with emerging alternatives to formal schooling. In Anderson, S. E. (ed.) *Improving Schools through Teacher Development: Case studies of the Aga Khan Foundation Projects in East Africa*, pp 247–270. Lisse, NL: Swets and Zeitlinger.
- Fullan, M. (2001). *The New Meaning of Educational Change*, 3rd edn. New York: Teachers College Press.
- Fuller, B. (1987). School effects in the Third World. *Review of Educational Research* 57(3), 255–292.
- Hall, G. and Hord, S. (2006). *Implementing Change: Patterns, Principles, and Pitfalls*, 2nd edn. Boston, MA: Pearson Education.
- Havelock, R. G. and Huberman, M. (1978). *Solving Educational Problems: The Theory and Reality of Innovation in Developing Countries*. Paris: UNESCO.
- Ibrahim, N. A. (1991). Inservice training in Malaysia for the new primary curriculum. In Lewin, K. and Stuart, J. (eds.) *Educational Innovation in Developing Countries: Case Studies of Changemakers*, pp 95–126. London, UK: MacMillan.
- Knapp, G. W., Fairhurst, M. G., Gibbs, W., et al. (2000). *The Effectiveness of the Teacher Resource Centre Strategy (Education Research Series Serial No. 34)*. London, UK: Department for International Development.
- Leithwood, K. and Menzies, T. (1998). A review of research concerning the implementation of school-based management. *School Effectiveness and School Improvement* 9(3), 233–285.
- Lewin, K. and Stuart, J. (1991). *Educational Innovation in Developing Countries: Case Studies of Changemakers*. London, UK: MacMillan.
- Lockheed, M. (1993). The condition of primary education in developing countries. In Levin, H. and Lockheed, M. (eds.) *Effective Schools in Developing Countries*, pp 20–40. London, UK: Falmer.
- Murphy, J. and Datnow, A. (2003). *Leadership Lessons for Comprehensive School Reform*. Thousand Oaks, CA: Corwin.
- Plank, D. and Sykes, G. (2003). *Choosing Choice: School Choice in International Perspective*. New York: Teachers College Press.
- Reynolds, D. (2000). School effectiveness: The international dimension. In Teddlie, C. and Reynolds, D. (eds.) *The International Handbook of School Effectiveness Research*, pp 232–256. New York: Falmer.
- Shaeffer, S. (1990). Participatory approaches to in-service teacher training. In Rust, V. and Dalin, P. (eds.) *Teachers and Teaching in the Developing World*, pp 95–114. New York: Garland.
- Srivastava, P. and Walford, G. (2007). *Private Schooling in Less Economically Developed Countries: Asian and African Perspectives*. Oxford, UK: Symposium Books.
- Task Force on Education and Gender Equity (2006). Towards universal primary education: Investments, incentives and institutions. A *United Nations Millennium Project Report*. New York: United Nations Millennium Project.
- Tooley, J. and Dixon, P. (2007). Private education for low-income families: Results from a global research project. In Srivastava, P. and Walford, G. (eds.) *Private Schooling in Less Economically Developed Countries: Asian and African Perspectives*, pp 15–40. Oxford, UK: Symposium Books.
- Tsang, M. C. and Wheeler, C. (1993). Local initiatives and their implications for a multi-level approach to school improvement in Thailand. In Levin, H. and Lockheed, L. (eds.) *Effective Schools in Developing Countries*, pp 108–129. London, UK: Falmer.
- Verspoor, A. (1989). Pathways to change: Improving the quality of education in developing countries. *World Bank Discussion Papers* #53. Washington, DC: The World Bank.
- Glassman, D., Naidoo, J., and Wood, F. (eds.) (2007). *Community Schools in Africa: Reaching the Unreached*. New York: Springer.
- Greenland, J. (2002). Evolution in school improvement with the Aga Khan Foundation from 1984 to 2000. In Anderson, S. E. (ed.) *Improving Schools through Teacher Development: Case Studies of the Aga Khan Foundation Projects in East Africa*, pp 21–46. Lisse, NL: Swets and Zeitlinger.
- Haiplik, B. (2004). *BRAC's Non-Formal Education Program (NFEP)*. Washington, DC: Academy for International Development.
- Halai, A. (2006). Mentoring in-service teachers: Issues of role diversity. *Teaching and Teacher Education* 22(6), 700–710.

Further Reading

Harvey, S. (1999). The impact of coaching in South African primary science InSET. *International Journal of Educational Development* **19**, 191–205.

Levin, H. and Lockheed, M. (eds.) (1993). *Effective schools in developing countries*. London, UK: Falmer.

Rust, V. and Dalin, P. (1990). *Teachers and Teaching in the Developing World*. New York: Garland.

Shamim, F. (2009). Impact and sustainability of the Whole School Improvement Program. In Querishi, R. and Shamim, F. (eds.) *Schools and Schooling Practices in Pakistan: Lessons for Policy and Practice*, pp 211–234. Oxford, UK: Oxford University Press.

Stuart, J. (1991). Classroom action research in Africa: A Lesotho case study of curriculum and professional development. In Lewin, K. and Stuart, J. (eds.) *Educational Innovations in Development Countries*, pp 127–152. London, UK: MacMillan.

Relevant Website

<http://www.mejorescuela.cl> – Segundo seminario regional sobre educación y calidad.

Sustainable Educational Reform

A Hargreaves, Boston College, Chestnut Hill, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

This article discusses the nature and importance of sustainability and unsustainability in primary and secondary education and in educational reform. It provides a definition of sustainability and contrasts it with synonyms and analogs in the field. Three historical and sequential ways of addressing the challenge of sustainability over the past half-century are reviewed – along with their strengths and limitations – and then principles of sustainability that define a fourth age of educational change are defined, described, and exemplified.

Education for Sustainable Development

Sustainability and sustainable development have two prime uses in educational change. The first use – and the chief focus of this article – concerns the sustainability of particular changes and also of broader change processes, especially in terms of whether these changes last or spread. The second use is connected more closely and obviously to the original meanings of sustainability and sustainable development. This section deals with the more obvious aspects of educational and environmental sustainability prior to linking them to the ensuing issues of human and organizational sustainability.

The global crisis of climate change has brought the ecological aspects of sustainability to the forefront of almost everyone's attention (Giddens, 2009). One educational response to these issues has been the emergence of Education for Sustainable Development (ESD). In some respects, ESD is a re-working of environmental or global education – teaching young people about the environment's interconnectedness and how to care for it. It has spawned initiatives in ESD curriculum, and in movements to build sustainable or green schools that promote innovative architecture, prudent and responsible use of energy, and environmental practices such as recycling, throughout the school's community. Influential advocates and intellectuals, such as Senge (forthcoming), point to the necessity of schools and young people engaging in real-world environmental activism to develop the systems-thinking that enables and empowers them to act as informed, concerned, and committed citizens in an imperiled world.

International organizations, such as United Nations Educational, Scientific and Cultural Organization (UNESCO), advocate for an even more inclusive approach to ESD that encompasses the needs of sustainable communities as well as natural environments – communities characterized by

gender equity, multicultural inclusiveness, the rights of basic education for all, elimination of poverty, and so on.

This essential understanding of ESD makes clear that sustainability of the natural environment cannot be disconnected from social justice in the organizational and human environment. ESD is not just a prelude or sidebar to the discussions of organizational sustainability that concern the endurance and spread of reforms. It strikes at the nature and effects of those reforms themselves. Thus, reforms that restrict the curriculum or standardize teaching to such an extent that the enriching, engaging, and empowering elements of ESD are sacrificed to raising test scores in the basics of literacy and numeracy at any cost are inherently unsustainable – whether or not they endure or last – for they transgress the fundamental purposes and principles of sustainable development in learning and life.

In this respect, it is important to recall and renew the ethical and environmental origins of sustainability and sustainable development. In the early 1990s, Lester Brown founded the World Watch Institute and defined a sustainable society as one that is able to satisfy its needs without diminishing the opportunities of future generations to meet theirs. The later idea of sustainable development was historically defined by the Brundtland Report of the World Commission on the Environment and Development as the ability to “meet the needs of the present without compromising the ability of future generations to meet their own needs” (World Commission on Environment and Development, 1987). Adapting this original definition, Dean Fink and this author developed the first body of work that theoretically elaborated and empirically supported a definition and philosophy of sustainability in education and educational change:

Sustainable educational leadership and improvement preserves and develops deep learning for all that spreads and lasts, in ways that do no harm to – and, indeed, create positive benefit for others around us – now and in the future (Hargreaves and Fink, 2003). Michael Fullan's *Leadership and Sustainability* (2004) sees sustainability as the link between systems thinking (grasping the interconnectedness of things) and moral purpose (taking responsibility for the connections) (Fullan, 2004).

The remainder of this article uses these definitions and the associated ideas of ESD as touchstones to review historic changes in educational reform strategy and its varied approaches – and failure – to achieve sustainability. It then identifies and exemplifies seven key principles of

educational sustainability in the emerging reform era and considers their implications for the future of educational change.

Three Ages of Unsustainability

In the 1960s and early 1970s, an age of innovations in curriculum programs, innovative schools, and open classrooms was stimulated by the US response to Sputnik and the need for more advanced teaching and learning in mathematics and science. Research on educational change, during this period, primarily examined what happened to innovations as things, and how they could be spread or diffused from central scholarly experts to educational practitioners. However, it was discovered, knowledge was rarely utilized in a rational way. People assigned their own meanings to innovations depending on the parts of it they experienced and the sense they made of the innovations in relation to their own roles and background experiences. The same innovations were assigned very different meanings according to the characteristics of users (Louis, 1998) and of users' social networks. Innovative schools are subject to the same fate. Typically, innovations fade, lighthouse schools lose their luster, and attempts to spread initiatives across a wider, more skeptical system – scaling them up – meet with little success.

As a result, the focus of educational change shifted from orchestrating the rational mechanics of dissemination and diffusion, to understanding the meaning of change for different participants and finding ways to manage that meaning (Fullan, 1982).

The inconsistencies of the *Age of Innovation* were, therefore, followed by a quest for purpose and coherence, and for managing the meaning of educational change in order to increase commitment to it. Sustainability – as we might now see it – was approached by the attempt to develop common meaning. First, there was an attempt to set out some central principles, parameters, and even broad standards of educational purpose that provided stronger direction while still allowing professionals to exercise discretionary judgment in how these purposes were interpreted.

Second, there were efforts to restructure the roles and responsibilities of schooling to reduce the restrictive effects of bureaucracy, to create more incentives for school-based initiative, and to reconfigure teacher interactions so they crossed traditional boundaries. Decentralized decision making, magnet schools, and charter schools were devised as ways to raise achievement for the children of aspirational families and to release schools from the one-size-fits-all controls of administrative bureaucracies and teacher unions in local school districts.

However – for the classroom teacher – this combination of centralized frameworks and initiatives with

decentralized roles and responsibilities often seemed bewildering. Portfolio assessments were paralleled by standardized tests. Interdisciplinary initiatives ran alongside subject-based report cards. Magnet schools targeted to particular populations also had to include others with special educational needs to meet federal civil rights guidelines. Teachers, in some schools, succeeded in maintaining their missions while still addressing the standards and accommodating to the structures. Many others, however, were fragmented and frustrated. The problem was not the structure, but the culture of teaching.

The third element of educational change in this period, therefore, was one of reculturing the working relationships among teachers, and managing the meaning they attached to teaching, learning, and school reform. In first devising and defining the concept of reculturing, this author contrasted it with restructuring in terms of its focus on rules and relationships rather than roles and responsibilities (Hargreaves, 1994). Reculturing is more about the meaning of change than its technical mechanisms. It is a response to long-standing concerns with the traditional culture of teaching characterized by noninnovative cultures of conservatism, individualism, and presentism – a fixation on the short term (Lortie, 1975).

Reculturing has promoted more teacher collaboration and inquiry. Collaborative cultures are generally associated with increased student success and improved retention among new teachers. They provide the mutual learning and moral support that stimulate teachers and sustain them through the difficulties of change.

In the end, however, reculturing failed to achieve consistent success or sustainability because of inconsistent leadership and insufficient emphasis on results. Under strong and effective leadership, collaborative cultures could make sense of complex purposes together and make sure that joint professional efforts led to improved outcomes in learning and results. Under weak or ineffective leadership, however, collaborative cultures were not always connected to learning and achievement and, under autocratic leadership, mandated coaching and collaboration often turned genuine teacher inquiry into rituals of contrived or enforced collegiality that actually made teachers inclined to collaborate less (Hargreaves, 1994).

Frustrations with the technical strategy of planned diffusion, and the cultural strategy of developing shared meaning, led to a more politically driven *Age of Large-Scale Systemic Reform* characterized by centrally coordinated prescription, control, intervention, and accountability.

Michael Fullan heralded the return of large-scale reform (LSR) at the turn of the century (Fullan, 2000), but it had already had more than a decade's track record in many Anglo-American nations where schools were, increasingly, placed in a competitive system of market choice for students and their parents, in relation to more-and-more detailed standards linked to high stakes

tests that were widely publicized in league tables of performance and, often, combined with weakened levels of resourcing, as well as impossibly accelerated timelines for implementation. This political strategy of educational change was, subsequently, also adopted within the bipartisan, federal No Child Left Behind (NCLB) legislation in the US.

In the Age of LSR, the benefits of sustainability-as-spread were evident in the emergence of clearer focus, greater consistency, and attention to all students with a stronger sense of urgency. However, many negative effects of unsustainability were evident beyond the criterion of spread. Achievement gains often occurred for a year or two in most cases, but then soon reached a plateau. Parents had more choice, but it was the affluent ones who knew how to work the system to advance their interests and protect their privileges. Standards raised the bar but shortfalls of professional support did not help children reach it. As measures of performance rose in tested literacy, rates of reading for pleasure actually fell. The costs to the quality, depth, and breadth of children's learning to increased dropout and reduced innovation, as well as to the caliber of teachers and leaders the profession could recruit and retain, were considerable (Hargreaves, 2003).

In response to these limitations of unsustainability, a revamped version of LSR – or LSR 2.0 – was launched instead (Barber, 2007; Hopkins, 2007). LSR 2.0 is even tighter in its imposition of ends, yet considerably more flexible in the orchestration of means.

In LSR 2.0, government establishes a small number of specific goals – such as the closely specified “Adequate Yearly Progress” indicators in the US, or system-wide literacy and numeracy targets in Ontario and England – and provides greater oversight at every level in their prescription and pacing. Test-score data are the focus of professional learning communities of inquiry in schools and districts that identify gaps and inconsistencies and design interventions accordingly. Collegial coaching and leadership supervision through walk-throughs and other methods provide technical support to teachers while also ensuring they comply with or demonstrate fidelity to the reforms. League tables printed in newspapers and digital media inform the public about student achievement results, and parents in underachieving schools are given opportunities to transfer their children to others with better results. Educators are encouraged to build lateral learning networks to drive change and the public has access to information with regard to teacher quality and student achievement levels. The government sponsors outsourcing of some educational provisions, such as the supplementary educational services provided by NCLB to students in struggling schools. Diverse provider models enable parents and students to shop among a variety of school options. Politically imposed timelines for improvement are linked to short-term election cycles, and the

failure of schools to meet these leads to increasing amounts of intervention, so that, in general, intervention is inversely related to success.

Advocates of LSR 2.0 claim increased standards in measurable improvement, narrowed achievement gaps, enhanced professional quality and motivation, and system-wide impact. In terms of sustainability, they point to depth or penetration of impact on achievement and spread of impact on a large scale.

Critics acknowledge these gains but also highlight significant limitations that create unsustainability such as an overemphasis on tested and targeted basics that marginalizes attention to arts, social studies, innovation, and creativity that are essential for competitive success in twenty-first-century knowledge; an inverse relationship between narrowly tested achievement and the development of the whole child and its overall well-being; a preoccupation with data-driven improvement that distracts teachers from deeper engagements with teaching and learning; a tendency for large-scale changes to lose their sophisticated and deeper qualities when they are hurriedly transplanted from one city or country to another; an overwhelmingly short-term orientation that leads to cynical strategies to improve results that secure only temporary success; and disincentives for schools to assist their weaker neighbors when school-by-school competitiveness persists (summarized in Hargreaves and Shirley, 2009).

Given these weaknesses, LSR 2.0 has difficulties meeting the sustainability criteria of depth, endurance, and benign impact on others in the surrounding environment, even though it might achieve some success in terms of spread.

A Fourth Age of Sustainability

Dean Fink and this author have been the first to develop an integrated theory and evidence-based elaboration of educational sustainability (Hargreaves and Fink, 2003). The theory of sustainability consists of seven interrelated principles that emerged from evidence collected in a study of educational *Change Over Time* funded by the Spencer Foundation (Hargreaves and Goodson, 2006). The study examined the experiences of educational change over more than 30 years in eight US and Canadian secondary schools – four innovative, and four more traditional – through the eyes of teachers and leaders who had taught there since the 1960s and 1970s. It encompassed more than 250 interviews with three cohorts of teachers in these schools who had entered the profession and also these schools at different points in time. It also included ethnographic observations as well as interpretation of detailed archival and demographic data from the schools and their surrounding districts and communities.

Through grounded theorizing, the project identified themes and periods in relation to changes that did and did not persist and why. The existing literature of educational change usually defined these questions of persistence in terms of sustainability, but, in turning to the wider literature of environmental and organizational sustainability, the discovery was that, in education, sustainability was being used as a synonym for maintainability. More detailed and interactive engagement of this wider literature with the evidence of our study then gave rise to our theory of educational and organizational sustainability and the seven essential and interrelated principles that comprise it.

Sustainable improvement, it was found, is characterized by depth of purpose in learning, caring, and real achievement rather than superficially tested performance; length of impact over the long haul – beyond individual leaders and short-term results – through longer-term planning and effectively managed leadership succession; breadth of influence – where leadership and improvement become a distributed responsibility; justice in ensuring that improvement efforts do no harm to, and actively benefit, students in other classes and schools; diversity that replaces monocultural standardization and alignment with networks and cohesion; resourcefulness that conserves and renews teachers' and leaders' energy and does not burn them out; and conservation that builds on the best of the past to create an even better future.

Sustainable Reform in Practice

In recent years, this author's research program has been concentrating on investigating and articulating the basic principles of demonstrably successful and sustainable change and reform in schools, districts, networks, provinces, and countries, as well as other sectors outside education. Two examples of these follow: a report for the Organization for Economic Cooperation and Development (OECD) on the relationship between leadership and school improvement in the world's highest performing country on the Program for International Student Assessment (PISA) tests of educational performance – Finland (Hargreaves *et al.*, 2008); and an ongoing cross-sector project co-directed with Alma Harris on *Performing Beyond Expectations*, in education, health, business, and sport, that includes the London Borough of Tower Hamlets (Hargreaves and Shirley, 2009).

Finland is the world leader on results in PISA tests of sophisticated, applied knowledge in mathematics, science, and literacy, as well as on international ratings of economic competitiveness according to the International Monetary Fund (IMF). This is remarkable given Finland's

economic collapse, in 1992, when unemployment levels reached ~20%.

Finland avoids short-sighted and unsustainable standardization along with national standardized tests altogether and reaches high levels of educational achievement and economic success by attracting highly qualified teachers with supportive working conditions, strong degrees of professional trust, and an inspiring mission of inclusion and creativity. This mission connects back to Finns' understandings of themselves as a creative and craft-like people and is embedded in a curriculum that not only values science and technology but also promotes creative and performing arts for all students until the end of secondary school with musical instruments paid for by the state.

Our OECD research team found that school principals in Finland have been able to lead communities of highly qualified teachers who develop their school curriculum together in each municipality within broad national guidelines. They and their teachers work in cultures of trust, cooperation, and responsibility, seeing themselves as one of a society of experts – who work with fellow professionals and neighboring schools – to achieve compelling and inspiring purposes, together, that rebuild their communities and their nation around knowledge society principles of inclusiveness and creativity. Indeed, in the city of Tampere, high schools principals told us that when they jointly agree on an important initiative for their community, if one school is short of resources, the principal can call the others and one of them will say “We have a little bit extra, would you like some of ours?” Finnish teachers think about more than ‘me and my class.’ In addition, Finnish principals think and act beyond ‘me and my school.’

Finns' inspiring purpose enables them to connect their hi-tech future to their creative past (conservation) and to concentrate on learning and caring for every student rather than boosting scores on testing (depth). Finns are not distracted by the necessity for short-term lifts in test scores to satisfy the temporary demands and transient initiatives of passing governments (endurance). Distributed responsibility among teachers and with principals ensures breadth of impact within and between schools. Schools help their neighbors because of a compelling and transcending commitment to a common community (justice) and they work together in networks of curriculum development and support (diversity). Free school meals and instruments for all children along with sufficient salaries and adequate working conditions give teachers the energy for change that is not sapped by unwanted external interventions (resourcefulness). For 15 years, the Finnish system of improvement and change in education and in the society has not only been successful but also sustainable.

The London Borough of Tower Hamlets with its large Bangladeshi community is an equally compelling example of system-wide sustainable improvement. Tower Hamlets' Bengali community suffered from high unemployment rates and some of the greatest incidences of poverty in the country with more children on free school meals than almost anywhere else. Educators' aspirations for student achievement were startlingly low and, in 1997, Tower Hamlets was proclaimed the country's worst-performing Local Education Authority (LEA), with the lowest performing primary school in the nation.

Ten years later, the transformation of the schools in Tower Hamlets is dramatic. The schools perform around and above the national average. On standardized achievement tests, high school examination results, and rates of students going on to university, the borough ranks as the most improved local authority in England. It has significantly reduced achievement gaps in relation to children with special educational needs, those from cultural minorities, and those on free school meals. These gains have been achieved with, largely, the same population and in comparison with the more modest national gains posted in the same time period.

Refusing to apply punitive interventions or sponsor market competition, Tower Hamlets has had an orderly succession of inspiring district leaders who have communicated high expectations that poverty is not an excuse for poor outcomes, that aspirations should be extremely high, that efforts to meet these aspirations should be relentless, and that everyone should work on this together. These system leaders have developed collaborative and trusting relationships with leaders in their schools. Testing is always on teachers' and leaders' radar but engaging students in their learning definitely comes first. Schools set ambitious performance targets together in a culture of target setting so that everybody owns them and these targets exceed those that are handed down to them by government.

District administrators are very present in the schools. They report that "lots of our schools work very closely together and with the local authority" and inspectors' reports refer to the "enthusiasm and high level of morale among the workforce". This grounds intervention in consistent and direct personal knowledge and communication more than in the numerical data that eventually appear on spreadsheets.

When schools fall behind, others rally round to help. Better teachers are cultivated and kept as a result of positive partnerships with local teacher-training providers. Positive and responsible partnerships have been developed with the corporations in the Docklands development of Canary Wharf, and droves of paid teaching assistants – drawn from the local community – work alongside classroom teachers, easing the workload, and

developing active trust and committed engagement with parents and the community.

In Tower Hamlets, as in Finland, a carefully managed succession of leaders (endurance) with a robust and resilient sense of purpose (depth) work closely with their schools and develop shared targets for improvement (breadth). The local authority has developed a collaborative ethic of schools helping schools and the strong supporting the weak (justice and diversity) even within a national political system that is designed to be highly competitive. Two proud pasts – of Tower Hamlets and its Bengali community – are connected to the borough's inspiring transformation for the future (conservation). These leaders think and work sustainably beyond their isolated schools – with other schools, teacher training-providers, and also the surrounding communities. This not only develops distributed responsibility beyond individual schools and beyond the schools in general, but it also increases the energy, support, and resourcefulness for change. One of Tower Hamlet's leaders sums it up well. It is "not just about the data. It's actually knowing the school, knowing the community, knowing about history, knowing about the staff – all of that."

Conclusion

Sustainability means more than maintainability. It means developing changes that matter, spread, and last and do no harm to others. Three ages of educational reform have failed to achieve sustainability. Innovations as things did not spread or endure. Efforts to reculture the teaching profession were undone by insufficient investment in leadership and insubstantial attention to results. More recent movements toward LSR have achieved greater spread and affected the details of practice, but, oddly, the practice that has been affected and enforced has reinforced traditional practices rather than disrupting them. More than this, LSR 2.0 has typically enjoyed only transient success and, often, at the expense of others.

A deeper philosophy and practice of sustainability returns to and embraces the environmental roots of the concept. It emphasizes, celebrates, and advocates for learning in depth about one's place in and contribution to the world; for developing individual interdependence on each other in cultures of trust and responsibility; for building a lasting legacy beyond short-term cycles of political popularity; for ensuring that competition does not destroy or diminish one's necessary capacity for helping one's neighbors; and for elevating change by professionalism and democracy over reform through market competition and bureaucracy. Our future has to be sustainable. So must our schools be (Havelock, 1971).

See also: Finland; Reforming the American School System; School Reform and Restructuring: Self Managing School.

Bibliography

- Barber, M. (2007). *Instruction to Deliver: Fighting to Transform Britain's Public Services*. London: Methuen.
- Fullan, M. (1982). *The Meaning of Educational Change*. New York: Teachers College Press.
- Fullan, M. (2000). The return of large-scale educational reform. *Journal of Educational Change* 1(1), 5–27.
- Fullan, M. (2004). *Leadership and Sustainability*. Thousand Oaks, CA: Corwin Press.
- Giddens, A. (2009). *Politics of Climate Change*. Cambridge, UK: Polity Publishers.
- Hargreaves, A. (1994). *Changing Teachers, Changing Times: Teachers' Work and Culture in the Postmodern Age*. London/New York / Toronto: Cassells/Teachers' College Press/University of Toronto Press.
- Hargreaves, A. (2003). *Teaching in the Knowledge Society: Education in the Age of Insecurity*. New York: Teachers College Press.
- Hargreaves, A. and Fink, D. (2003). Sustainable leadership. *Phi Delta Kappan* 84(9), 693–700.
- Hargreaves, A. and Goodson, I. (eds.) (2006). Educational change over time? The sustainability and non-sustainability of three decades of secondary school change and continuity. *Educational Administration Quarterly* 42(1), 3–41.
- Hargreaves, A., Halász, G., and Pont, B. (2008). The Finnish approach to system leadership. In Pont, B., Nusche, D., and Hopkins, D. (eds.) *Improving School Leadership, Vol. 2: Case Studies on System Leadership*, pp 69–109. Paris: OECD.
- Hargreaves, A. and Shirley, D. (2009). *The Fourth Way*. Thousand Oaks, CA: Corwin Press.
- Havelock, R. G. (1971). The utilization of educational research and development. *British Journal of Educational Technology* 2, 84–97.
- Hopkins, D. (2007). *Every School a Great School: Realising the Potential of System Leadership*. Maidenhead: Open University Press.
- Lortie, D. C. (1975). *Schoolteacher: A Sociological Study*. Chicago, IL: University of Chicago Press.
- Louis, K. S. (1998). Reconnecting knowledge utilization and school improvement: Two steps forward, one step back. In Hargreaves, A., Lieberman, A., Fullan, M., and Hopkinset, D. (eds.) *International Handbook of Educational Change*, pp 1074–1095. Dordrecht: Kluwer.
- Senge, P. M. (forthcoming). Education for an interdependent world: Developing systems citizens. In Hargreaves, A., Fullan, M., Hopkins, D., and Lieberman, A. (eds.) *The Second International Handbook of Educational Change*. Dordrecht: Springer.
- World Commission on Environment and Development (1987). *Our Common Future*. New York: United Nations General Assembly.

Further Reading

- New Commission on the Skills of the American Workforce (2007). *Tough Choices or Tough Times: The Report of the New Commission on the Skills of the American Workforce*. Washington, DC: National Center on Education and the Economy.

The Role of the District in Tri Level Reform

M Fullan, University of Toronto, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

The most distinctive feature of educational reform, over the past decade, has been the increasing focus on system-wide effectiveness. This has been called the tri-level reform: what has to happen at the school and community level, the mid- or district level, and at the state. This new work is not so much about the intralevel developments, but rather about strong multi-way interrelationship across the levels – in other words, systemic reform.

This article focuses particularly on the role of the district – linking both downward (to its schools) and upward to the state. In North America, districts play a more direct line authority role than they do, for example, in England.

The role of local authority in England is more complex with the evolution of the cross-cutting forces of the local autonomy, and more central accountability. The increase in the scope of responsibilities to include all of children's services has further complicated the scene. It is, in this author's view, very difficult at this time to sort out the causal role of local authorities in England. It is too much of a moving and complex target. One empirical study of the role of local authorities in England (the data on which it is based is now 3-years old, which itself is a problem) concluded "the Education Authority attended by pupils has almost no relevance to their progress" (Tymms *et al.*, 208: 261).

The role of the district in North America is more influential. It is true that the impact of the classroom is most important, the school second most, and the district third most, and so on. Nonetheless, this article claims that the infrastructure does matter a great deal.

The main two sections of this article present the case and the data, drawn from this author's own research and the research in the field as a whole. These issues are pursued, first, by considering the role of the district (and its schools), and, second, by placing districts in the context of the state or system as a whole.

The Role of the District

If one takes a quantitative approach, the majority of districts are not effective. To be fair, stimulating, coordinating, and sustaining coherent development across many schools is exceedingly difficult because it requires balancing top-down and bottom-up forces. This article traces the evolution of the role of districts in school reform using informal language to capture the three themes: getting somewhere; not so fast; and what's next?

Getting Somewhere

Since nearly 1990, there has been a growing body of work that points to common characteristics and strategies that successful districts use to raise student achievement. Rosenholtz's (1989) study of 78 elementary schools classified schools as "stuck," "moving," and "in-between." Rosenholtz also found that a disproportionate number of stuck schools came from certain districts; likewise, moving schools were clustered in certain other districts. This prompted her to write a chapter on stuck and moving districts. Rosenholtz comments:

The contrast between stuck and moving districts, nowhere more apparent than here, underscores how principals become helpful instructional advisors or maladroit managers of their schools. It is also clear that stuck superintendents attribute poor performance to principals themselves, rather than accepting any responsibility to help them learn and improve. This again may indicate their lack of technical knowledge and subsequent threats to their self-esteem. If districts take no responsibility for the in service needs of principals, of course, principals become less able colleagues, less effective problem-solvers, more reluctant to refer school problems to the central office for outside assistance, more threatened by their lack of technical knowledge, and, most essential, of substantially less help to teachers. (p. 189)

Anderson (2006), a colleague of mine, reviewed the research on district effectiveness and named 12 key strategic components.

1. District-wide sense of efficacy.
2. District-wide focus on student achievement and the quality of instruction.
3. Adoption and commitment to district-wide performance standards.
4. Development and adoption of district-wide curricula and approaches to instruction.
5. Alignment of curriculum, teaching, and learning materials, and assessment to relevant standards.
6. Multi-measure accountability systems and system-wide use of data to inform practice, hold school and district leaders accountable for results, and monitor progress.
7. Targets and phased focuses of improvement.
8. Investment in instructional leadership development at the school and district levels.

9. District-wide, job-embedded professional development focuses and supports for teachers.
10. District-wide and school-level emphasis on teamwork and professional community (including in several cases positive partnerships with unions).
11. New approaches to board–district relations and in-district relations.
12. Strategic relations with state reform policies and resources.

One would think, then, that there is a growing consensus and that it is just a matter of going to town on what we know. One would be wrong.

Not So Fast

So a district should get the standards right, align curriculum to them, conduct assessments on the new alignment, provide solid and continuous professional development on curriculum and instruction, set up a data system that can be used for both assessment for and assessment of learning, and engage with the local community and state reform policies. It may surprise many readers that these steps, by themselves, are not sufficient and, at best, may represent a waste of resources, and, at worst, do more harm than good.

The experience of the San Diego City Schools District is a good place to start with respect to the not-so-fast theme. Coming off a highly successful experience in District 2 in New York City from 1988 to 1996, Tony Alvarado was hired as Chancellor of Instruction in 1997 to join a new high-profile superintendent – Alan Bersin – in San Diego. In a sense, the question was – if you could take the best knowledge, and add resources and political clout, could you get results in a large urban district within a 4-year period, and then keep going, in this case, moving from success in 45 schools (District 2) to 175 schools (San Diego)? The answer, incidentally, is yes, but it requires good strategies and a good deal of finesse.

The San Diego reform story is probably the most closely watched reform initiative in the history of urban school improvement. Here, this article draws on the excellent account by Hubbard *et al.* (2006). The San Diego strategy was well detailed and explicit from day 1. It consisted of the following three main components:

- Improved student learning: closing the achievement gap.
- Improved instruction: teacher learning through professional development.
- Restructuring the organization to support student learning and instruction.

The focus was on literacy, and the strategies were highly specific. Teachers received support from literacy coaches and principals who were positioned to be leaders of

instruction, with day-to-day support, and monthly full-day in-service sessions by area superintendents whose new roles (and new people) were re-created as instructional leaders.

This article does not provide the space to enter a detailed account of the San Diego experience, but the main outcomes and reasons can be identified (for a full account, see Hubbard *et al.*, 2006). To cut to the chase, literacy achievement increased, somewhat, at the elementary level in the 1997–2001 period, had a limited impact in middle schools, and was a dismal failure in high schools. Momentum was lost by 2001, Alvarado was asked to leave in 2002, and Bersin – after slowing down the nature and pace of reform in 2003–04 – was replaced by the school board when his term expired in June 2005. What happened?

One could say that it was a political problem – the board was divided from the beginning (3:2 in favor of the reform initiative), and the teacher union that opposed the reform from the beginning eventually carried the day. There is some truth to this, but the deeper explanation is closer to the theme of our interest in meaning and motivation relative to pace, the too-tight/too-loose problem, and the depth of instructional change and thinking required to make a difference. Hubbard *et al.* (2006) expressed the basic problem in terms of three challenges that the strategy failed to address: “The need to accomplish deep learning within the constraints of a limited time frame; principals’ and coaches’ limited understanding of the concepts they were trying to teach; and the difficulty of reaching common ground between school leaders and teachers” (p. 128).

All this, despite plenty of classroom visits, walk-throughs involving all schools, frequent problem-solving sessions, and an emphasis on job-embedded professional learning. The San Diego case is an exercise in the dilemmas faced by leaders with an urgent sense of moral purpose and considerable knowledge of what should happen in classroom instruction. However, it also points to how the strategies employed must be much more respectful of how deep change happens. Much good was done in improving literacy achievement in elementary schools, but it was not deep enough or owned enough to go further. The San Diego strategy failed because the pace of change was too fast, the strategy was too unidirectional from the top, relationships were not built with teachers and principals, and, above all, the strategies did not really build capacity – which is the development of the collective knowledge and understandings required for ongoing instructional improvement that meets the needs of each child.

San Diego is also one of the better examples of attempted reform. Most districts do not focus their efforts on district-wide reform. In addition, when they do, they encounter limits to what can be accomplished despite considerable effort and resources.

Another confirmation of our not-so-fast worry comes from the Cross City Campaign for Urban School

Reform (2005), which has been cross-referenced above and examines major reform initiatives in Chicago, Milwaukee, and Seattle. All three school systems had the attention of political leaders at all levels of the system and focused on many of the right things, such as literacy and math; all of the systems used current choice strategies such as concentration on assessment-for-learning data, invested heavily in professional development, developed new leadership, and focused on system-wide change.

In addition, they had money – Seattle had \$35 million in external funds, Milwaukee had extra resources and flexibility, and Chicago had multimillions. There was huge pressure, but success was not expected overnight. Decision-makers and the public would have been content to see growing success over a 5- or even 10-year period. The upfront conclusion of the case-study evaluation was that for many of the principals and teachers interviewed, “the districts were unable to change and improve practice on a large scale” (Cross City Campaign for Urban School Reform, 2005: 4).

The issues in the Chicago, Milwaukee, and Seattle reforms help to identify the missing ingredient, even though those districts appear to have gotten most components right. Chicago, for example, appeared to have an impressive strategy: Academic standards and instructional frameworks, assessment and accountability systems, and professional development for standards-based instruction are among the tools of systemic reform that are used to change classroom instruction (Cross City Campaign for Urban School Reform, 2005: 23).

This appears to be a standards-based, system-wide reform that sounds like it should work. The failure, in this author’s opinion, is that the strategy lacks a focus on what needs to change in instructional practice. In Chicago, teachers did focus on standards, but in interviews, they “did not articulate any deep changes in teaching practice that may have been under way” (p. 23). Furthermore, “instructional goals were articulated more often in terms of student outcomes or achievement levels than in terms of instructional quality, that is, *what the schools do* to help students achieve” (p. 29, emphasis in original). Milwaukee reveals similar problems in achieving instructional improvements while using greater decentralization in the context of system support and competitive choice. The focus was on literacy; a literacy coach was housed in every school in the district and considerable professional development and technical support services were available. Education plans for each school were to focus on literacy standards through (1) data analysis and assessment and (2) subject-area achievement targets – including literacy across the curriculum. Sounds like a convincing strategy. However, what is missing, again, is the black box of instructional practice in the classroom. The case writers observe: “We placed the Education Plan in the indirect category due to its non-specificity regarding regular or

desired instructional content and practices” (Cross City Campaign for Urban School Reform, 2005: 49).

More generally, the report concludes that while these serious district-wide reform initiatives appeared to prioritize instruction, they did so indirectly (through standards, assessment, and leadership responsibilities). However, in the experience of principals and teachers, the net effect was that “policies and signals were non-specific regarding intended effects on classroom teaching and learning” (p. 65).

The third case, Seattle, is a variation on the same theme. The game plan looks good. Standards defined the direction, while the district’s Transformational Academic Achievement Planning Process “was designed as a vehicle for helping schools develop their own strategy for (1) helping all students meet standards, and (2) eliminating the achievement gap between white students and students of color” (p. 66). Similar to Milwaukee, the district reorganized to support site-based management, including the allocation of considerable resources to schools. The case writers observe:

The recent effort to become a standards-based district was one of the first sustained instructional efforts with direct attention to teaching and learning. However, the conversations district leaders had about standards *were rarely connected to changes in instruction.*

(Cross City Campaign for Urban School Reform, 2005: 69; emphasis added)

The report continues: “At the school level, finding teachers who understood the implications of standards for their teaching was difficult” (p. 72).

This article cites one more case, which – in some ways – is more encouraging but still proves this author’s main conclusion that instructional change is going to require different strategies that help develop and shape collective capacity and shared commitment to engage in continuous improvement. Supovitz (2006) conducted an excellent case study of the reform effort in Duval County, Florida. The title of his book captures the emphasis of his analysis – *The Case for District-Based Reform*. Supovitz chronicled the district-wide reform effort from 1999 through 2005. Duval County has 142 schools. The reform strategy is now familiar to us.

1. Develop a specific vision of what high-quality instruction should look like.
2. Build both the commitment and capacity of employees across the system to enact and support the instructional vision.
3. Construct mechanisms to provide data at all levels of the system that will be used both to provide people with information that informs their practices and to monitor the implementation of the instructional vision.
4. Develop the means to help people continually deepen their implementation and to help the district continually refine this vision and understand its implications.

With a sustained 5-year focus on the four strategic components, the district made significant gains in student achievement. For example, the number of schools receiving C or better on the state assessment system went from 87 (of 142) in 1999 to 121 by 2003. In addition – for the first time in a 7-year period, in 2005 – no school in the district received an F on the state accountability system.

The strategy was driven by a strong superintendent who helped to orchestrate the development of district-wide capacity according to the four core components described above. The strategy was enacted with considerable action and focus. As Supovitz reports, “Duval County leaders repeatedly stated their vision and the strategies for achieving it in public venues” (p. 43). Supovitz argues that the spread and deepening of district-wide success is as much “gardening” as it is “engineering” (p. 63). And that the balance requires “advocacy without mandate” (p. 66), “fostering urgency” (p. 68), and “building existing proof” of success (p. 69). One sees a similar array of strategies as in San Diego, but with less heavy-handedness: direct training of teachers, school standards coaches, district standards coaches, principal leadership development, and district leadership development.

With 6 years of consistent effort and with an explicit emphasis on professional learning communities as a strategy, Supovitz comments: “The possibilities of professional learning communities—rigorous inquiry into the problems and challenges of instructional practice and the support of that practice—seemed only to be occurring in pockets of the district” (p. 174). Much was accomplished in Duval County, but it was, by no means, deep or durable after 6 years. So the not-so-fast observation presented above is an apt worry. Even with comprehensive strategies and relentless focus over a 5- or 6-year period, one is still not getting it right.

What’s Next?

It has been seen that even the most ambitious efforts fall short, and these initiatives involve only a small minority of districts. Most are not nearly so active. This author believes that these efforts are on the right track, but the approach needs considerable refinement. To state what is needed upfront, there is need for a focus on instruction, standards, assessment, continuous feedback and use of data, and instructional leadership at the district and school levels. However, also needed is a process of interactive capacity building and commitment building within and among schools, and between schools and the district. Above all, this, increasingly, must de-privatize teaching so that learning in context can occur, and the district must stay the course over a period of 10 or more years. This work does not necessarily require the same superintendent over two or more terms, but does require continuity of good direction over two or three superintendencies. This article

cites three examples, from three different countries, of what this means in practice.

York Region District School Board just outside Toronto, Ontario, is a multicultural district with a growing and diverse population, and over 100 different languages spoken in the schools. There are 145 elementary schools and 30 secondary schools. We have been working in partnership with York for the past 5 years, including monitoring the processes and results as we go (see, e.g., Sharratt and Fullan, 2006). The focus is on literacy in an initiative called the Literacy Collaborative (LC). The basic approach is designed to shape and reshape district-wide continuous improvement – what this author calls “capacity building with a focus on results.” Key features of the approach include:

- A clearly articulated vision and commitment to literacy for all students, which is continually the subject of communication in the district.
- A system-wide comprehensive plan and framework for continuous improvement.
- Using data to inform instruction and determine resources.
- Building administrator and teacher capacity to teach literacy for all students.
- Establishing professional learning communities at all levels of the system and beyond the district.

All schools – including all secondary schools – joined the LC in a phased-in fashion, with school-based teams being the focal point for capacity building. At the elementary level, teams consisted of the principal (always the principal), the lead literacy teacher (a leadership role within the school, with a teacher released for 0.5 to 10 time to work with principals and teachers), and the special education-resource teacher. High school teams were slightly larger and focused on literacy – especially in the ninth and tenth grades. The LC model has evolved to contain 13 parameters, which are not listed here but include embedded literacy teachers, timetabled literacy blocks, a case-management approach focusing on each student, cross-curricular literacy connections, and so on (see Sharratt and Fullan, 2006). There is constant interaction, action research, and capacity building through formal monthly sessions, and many learning-in-context interactions carried out daily by school and district leaders within and across schools.

The results – as measured by province-wide assessments – were significant after a 3-year period (2001–04), but not as substantial as district leaders had hoped. On a closer examination of the initial cohort of 17 schools, it was found that nine of the schools had implemented the 13 parameters more deeply compared with the other eight. When the latter schools were separated, the results showed that the nine schools – despite starting below the York Region and Ontario provincial average in 2001 – had risen above both averages by 2004. In the meantime, the district

was working with all 167 schools. Province-wide results, in 2005, showed that York Region increased by a full 5%, on the average, in literacy across its 140 elementary schools. High schools also did well for the first time on the tenth-grade literacy test.

In terms of what is new, the theory of action reflected in the approach in York Region can be considered. First, we have many of the elements we have seen previously – standards, assessment of and for learning, instructional leadership, and so on – but one also sees two new significant emphases. One is that the leaders have taken a long-term perspective; they realize that it takes a while for change to kick in; they frequently speak of staying the course, and persistence but flexibility; the pace is steady, even pushy, but not overwhelming; they expect results, not overnight, but also not open ended. The other new aspect is that leaders are careful not to judge slow or limited progress in given schools. They take what this author calls a “capacity building first, judgment second” stance. Large-scale change is all about moving the whole system so that more and more leaders permeate the system and take daily actions that build capacity and ownership.

This is an entire district that is on the move. There has been one director (superintendent) – Bill Hogarth – throughout the 8-year process, and a strong rapport between the board and the district leadership. As a strong collaborative culture has been built, the chances of continuing this direction when he leaves are greatly increased. As this author said, one does not need the same superintendent over 8–12 years, but one does need continuity and deepening of good direction.

In further work in the York Region, the Superintendent of Curriculum – Lyn Sharratt – and this author examined the prospects for “sustaining leadership in York Region” (Fullan and Sharratt, 2007). When school principals were asked how they sustain their focus on continuous improvement, five major themes were identified. For effective building for the future, school principals said that they simultaneously focus on five interrelated components:

1. Shared beliefs, goals and vision
2. Distributed leadership and professional learning cultures
3. Data-based decisions/impact measures/celebrating success
4. Mobilization of resources (time, ideas, expertise, money)
5. School/community/home relations. (Fullan and Sharratt, 2007: 126)

While it is observed that sustainability will always be problematic, it can be concluded that York Region has done better than most in establishing the conditions that make sustainability more likely. In particular, four propositions have been formulated:

- Proposition One: Sustainability is not about prolonging specific innovations, but rather it concerns establishing the conditions for continuous improvement.

- Proposition Two: Sustainability is not possible unless school leaders and district leaders are working together on the same agenda.
- Proposition Three: Proposition Two notwithstanding, sustainability is not furthered by school and district leaders simply agreeing on the direction of the reform. Any temporary agreements must be continually tested and extended in the crucible of implementation with school and district leaders being equally influential.
- Proposition Four: Despite the clear signs in York Region and despite being able to identify favorable conditions, we still do not know what will happen when district leadership changes. (Fullan and Sharratt, 2007: 134–135).

A second good example is the decade-long reform initiative of the 58 000 student Boston Public Schools (BPS) under the leadership of Superintendent Tom Payzant. McLaughlin and Talbert (2006) describe the basic plan as based on six essentials: effective instruction as the core essential, student work and data, professional development, shared leadership, resources, and families and community. Again the words are familiar, but it is the theory of action – and careful cultivation over a decade – that is the real story. Professional development, for example, takes place at the school level and features a coaching strategy involving collaborative teaching groups. In collaborative coaching, teachers learn by analyzing one another's work under the guidance of skilled coaches. The idea is not just to observe one another's teaching and share ideas, but to critique lessons in a way that links to improving student learning. As with the York Region, a well supported and easy-to-access database on student learning is used to help teachers examine their teaching in light of ongoing results, integrating data into professional learning. Substantial outreach to parents and community is a major component of teachers' and schools' work.

Education Week published a profile on BPS and Payzant on the occasion of his announcing his retirement, in 2007, following 11 years as superintendent (“Time on his side,” Allen, 2006). In addition to reporting on the activities and approach just described, *Education Week* gets beneath the strategy.

Mr. Payzant did not bring the Boston schools to this point overnight. He rolled out initiatives not all at once, but only when they made sense. The idea was to start small, test things out, and retool them. In addition, he focused on building consensus. All were radical notions in an era of hard-charging, quick turnaround leaders (p. 31).

The impact of the Boston strategy brought significant results in student achievement. In the tenth-grade English language arts and in mathematics, scores have increased steadily since 1999 for all four race and ethnic groups (black, white, Asian, and Hispanic), with some leveling off in the 2004 and 2005 years. McLaughlin and Talbert (2006) summarize the positive impact in these words.

Multiple evaluations show that Boston's approach to instruction [and] to collaborative coaching and learning are benefiting students and teachers. Student outcomes have improved, as have relationships between teachers and students and among teachers. Boston sees other positive system-level consequences of their strategy—enhanced coherence, increased accountability at all levels, and increased buy in from district educators. (pp. 126–127)

Again, one sees a more sophisticated theory of action carrying the day. It is not that one is seeing flawless strategies. The pace of change was, likely, not fast enough in Boston. Put another way, few superintendents would be allowed to take this amount of time in 2009. However, the point remains. Too fast is a more likely negative scenario. Balancing pace – press for improvement with corresponding capacity building – carefully assessed as you go, is required. Payzant's own reflective lessons are revealing. He says he left some areas of work "too much to chance." He said he should have allowed fewer programs for teaching literacy. Likewise, he said it was a mistake to let high schools come up with their own plans for creating more personalized learning environments for students ("Time on his side," Allen, 2006).

The performance in Boston, along with substantial improvements, also reveals a plateauing effect in the last 2 years – a phenomenon that is normal, not to be lamented, but requiring new, deeper strategies. Elmore (2004) and this author (Fullan, 2006) have both commented on the plateau effect as a natural and (depending on what one does next) valuable opportunity to consolidate and then go deeper. In addition, while all four race and ethnic groups have gained, the gap has not closed and, in some cases, has increased. The next critical question for Boston is: who will be Payzant's successor? This author has said that, in these cases of being on the right track, it is crucial for districts to hire for continuity and deepening of good direction.

We move to England for our third example – Knowsley Local Education Authority. Note here, the contrast to the negative findings from Tymms *et al.* (2008), which found that the local authority made little difference in the education lives of students; qualitative case studies are likely to find examples of success (and failure) as opposed to large quantitative studies.

In any case, Knowsley Local Education Authority (called Local Authority, since 2005) is a metropolitan district just east of Liverpool. It is defined as the sixth most-deprived authority in the country. In 1999, Knowsley consisted of 59 primary schools, 11 secondary schools, and seven special schools. The district was audited in that year as part of the national inspection scheme conducted by the Office for Standards in Education (OFSTED). The assessment found serious weaknesses on most basic dimensions of performance: student achievement, capacity to

improve, relationships between the district and the schools, and linkage to the community. A new Director of Education, Steve Munby, was appointed in 1999. A second inspection was conducted in 2003, which found major improvements. What happened in 4 years to transform a very low-performing, disadvantaged, discouraged system into one vastly improved and ready to do more?

One can begin with OFSTED's 2003 findings (OFSTED, 2003).

Recent developments and the implementation of well-thought-through initiatives have resulted in Knowsley establishing itself as a local education authority (LEA) of some significance. It has improved over the past 3 years and has shown how vision and leadership, together with excellent relationships with schools, can revive an education service (p. 2).

As OFSTED further noted, both literacy and numeracy scores increased at a time when national averages were flatlined. OFSTED also observes that "the new administration has taken partnerships and collaborative working to an unusually high level." In addition, "head-teachers of individual schools see themselves as part of a wider team with responsibility for the education service throughout the borough" (p. 2).

Munby (2003) states that the drivers for change are low student performance, new leadership, external funding, and a moral commitment to narrow the gap between the highest and lowest performing schools. Munby then lists what he calls his "priorities for sustainability":

- Establishing an innovative, coherent, and comprehensive policy framework that provides direction for instruction and professional learning
- The training of "lead learners" to support school learning
- Deploying lead learners to work with clusters of schools to embed new practice
- Cluster-based work—action learning, observing and sharing learning, supporting small-scale action research to provide evidence of impact on pupil motivation, and engagement with the learning process
- Encouragement and support of the further development and embedding of a culture of co-planning, co-teaching, co-review, and co-coaching in schools, everyone a leader of learning. (p. 2)

In January 2005, Munby was appointed the chief executive officer (CEO) of the National College of School Leadership. A time for the continuity of good direction. The new director, Damian Allen, was appointed from within Knowsley, having been deputy director. Because Munby had employed a strategy of co-development of leadership, Allen was already immersed in the strategy, and, in fact, had helped shape it. By the time he was appointed in January 2005, the new Every Child Matters agenda had become a reality, with all children's services – including schools – coming

Table 1 A comparison of Knowsley's strategy with the government's white paper

Co-leadership between LA and schools	Vs.	Individual independent specialisms
Collaboration and federation as standard	Vs.	Collaboration and federation as a response to weakness
System performance	Vs.	Individual school performance
Failure driven out by challenge and support	Vs.	Failure driven out by early intervention and closure
Development of co-leadership	Vs.	Schools need autonomy
High support and engagement with schools	Vs.	Light-touch monitoring

under the LA. Allen became the first executive director, Children's Services. Knowsley has continued with the directional strategy of having an ambitious agenda for children, but forging ahead with co-leadership and capacity building. The district proceeded to introduce a remarkable secondary school reform that involved closing all 11 high schools, and reopening them with eight brand-new schools – complete with new state of the art buildings called learning centers. Knowsley did this without any rancor and, indeed, considerable enthusiasm – partly because of the co-leadership strategy, partly because new national money was available for new buildings, and partly because it was already experiencing success (e.g., the percentage of 15-year-olds passing five or more GCSEs – a mark of advanced placement courses for further education – had doubled from 22% in 1998 to 45% in 2005, while the national average moved from 47% to 57%).

Ever conscious of the theory of action that had gotten them there, Allen (2006) made a presentation at a national meeting in which he compared the Knowsley strategy with the strategy embedded in the recent white paper from the government (Department for Education and Skills, 2005). After pointing out the inconsistencies in the white paper, he noted the following comparisons (Knowsley on left; white paper on right, **Table 1**).

What is new, in this author's view, is the creating of partnerships of engagement that mobilize the entire district. It is still early in Knowsley's journey, but one can see a consistent, adaptable strategy in which successive leaders build on good direction, interacting with internal and external contexts.

The District and the State

It is beyond the terms of reference in this article to take up the complex and new emergence in the role of the state/central governments whether they be federal and state/provincial, or single national entities. There is, certainly, a growing focus on system strategies such as Michael Barber's *Instruction to Deliver* in England, and Hargreaves and Shirley's (forthcoming) *The Fourth Way*. Issues of market model, education and citizen empowerment, accountability, the new teacher professionalism are all intermingled in this debate.

For our purposes in this particular context, one can take the narrower question of the implications of the

previous section on the role of the state. The first thing to say is that the district should be focused and proactive in its own right, irrespective of the state's role.

Second, and more directly, this author's conclusion is that there is a direct analogy between the relationship of the schools and their districts, and the relationship between districts and state. In other words, in the same way that schools and the district must work as a single system, so too must the state and its schools/districts.

We have had a direct experience in shaping such a system in Ontario, Canada, since 2003. Ontario has two million students, 4900 schools, and 72 districts. We have focused, with considerable success, on improving literacy, numeracy, and high school graduation. The principles are similar to those that were reviewed in the last section (but obviously more complex, given the increase in size and component parts).

The relevant policies, strategies, and results, so far, are well documented in Levin *et al.* (2008) and Fullan (2008). Thus, Levin *et al.* describe the main sustaining elements as:

- Respect for staff and professional knowledge
- Comprehensiveness (whole system)
- Coherence and alignment through partnership with district and schools.

This author has described the actual action strategy as based on six elements:

1. Direction and sector engagement
2. Capacity-building with a focus on results
3. Supportive infrastructure and leadership
4. Managing the distractors
5. Continuous evaluation and inquiry
6. Two-way communication. (Fullan, 2008: 278)

It is encouraging to find that the same lessons about smaller system effectiveness (districts) apply with respect to whole-system reform. It can safely be predicted that the next phase of reform – assume 2009–14 – will focus deeply on system-wide reform on what this author has called tri-level reform. It is not just that the goal is to achieve new effectiveness across the whole larger system, but rather that the strategies involve system components at all three levels working in two-way and multi-way partnerships. One is, indeed, heading toward greater comprehensiveness, greater focus, more capacity building, and greater precision in zeroing in on core goals of literacy, numeracy, and high

school graduation. The next phase should accelerate our learning and knowledge with regard to school and system effectiveness.

Indeed, we and others are experiencing considerable success in increasing student achievement in literacy and numeracy and high school graduation rates across whole state systems (Fullan, 2010a, b).

Bibliography

- Allen, D. (2006). A new role for local authorities – from direct provide to strategic commissioner. *Paper Presented to Capita White Paper Conference*, London.
- Anderson, S. (2006). The school district's role in educational change. *International Journal of Educational Reform* **15**(1), 13–37.
- Cross City Campaign for Urban School Reform (2005). *A Delicate Balance: District Policies and Classroom Practice*. Chicago, IL: Author.
- Elmore, R. (2004). *School Reform from the Inside Out*. Cambridge, MA: Harvard University Press.
- Fullan, M. (2006). *Turnaround Leadership*. San Francisco, CA: Jossey-Bass.
- Fullan, M. (2008). Have theory will travel: A theory of action for system change. In Hargreaves, A. and Fullan, M. (eds.) *Change Wars* pp 275–293. Bloomington, IN: Solution Tree.
- Fullan, M. (2010a). *Motion Leadership*. Thousand Oaks, CA: Corwin Press.
- Fullan, M. (2010b). *All Systems Go*. Thousand Oaks, CA: Corwin Press.

- Fullan, M. and Starratt, L. (2007). Sustaining leadership in complex times. In Davies, B. (ed.) *Developing Sustainable Leadership* pp 116–136. London: Paul Chapman.
- McLaughlin, M. and Talbert, J. (2006). *Building School-Based Teacher Learning Communities*. New York: Teachers College Press.
- Office for Standards in Education (OFSTED) (2003). *Inspector Report: Knowsley LEA*. London: Author.
- Rosenholtz, S. (1989). *Teachers Workplace*. New York: Longman.
- Starratt, L. and Fullan, M. (2006). Accomplishing districtwide reform. *Journal of School Leadership* **16**(5), 583–595.
- Supovitz, J. (2006). *The Case for District-Based Reform*. Cambridge, MA: Harvard Education Press.
- Tymms, P., Merrell, C., Heron, T., et al. (2008). The importance of districts. *School Effectiveness and School Improvement* **19**(3), 261–274.

Further Reading

- Archer, J. (2006). Time on his side. *Education Week* **25**(3), 31–32.
- Barber, M. (2008). *Instruction to Deliver* (revised paperback). London: Methuen.
- Fullan, M. (2007). *The New Meaning of Educational Change*, 4th edn. New York: Teachers College Press.
- Munby, S. (2008). *Broad and Deep: A Whole Authority Approach to Motivating and Learning*. Mersey, UK: Knowsley Local Education Authority.

Relevant Website

<http://www.michaelfullan.ca> – Michael Fullan.

Whole School Designs for Enhancing Student Achievement

S Stringfield, University of Louisville, Louisville, KY, USA

J Nunnery, Old Dominion University, Norfolk, Virginia, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Change is the normal condition in both the physical and the social world. A generation ago, Toffler (1970) noted that the rate of change in our social world was increasing geometrically, and Freidman (2005, 2008) regularly reminds us of the international implications of these rapidly increasing changes. Increasing changes in the social world are causing a great need for dramatic educational improvements across nations.

Increased education is associated with a wide variety of desirable outcomes, ranging from increased voting rates to the acceptance of leadership in social and religious organizations to income. As an example of the later trend, in the United States the difference in incomes between 25–34-year-old male high school dropouts and 25–35-year-old male college graduates stood at 30% in 1949. In the twenty-first century, it stands at over 150%. Today's young college graduate typically makes over \$2.50 for every dollar earned by a high school dropout, and the gap continues rising. Such changes over time in the impact of education have profound implications for both the economy and civil society. In this rapidly changing environment, with the consequences of educational failure at their highest levels in history, it is hardly surprising that nations are increasingly focusing on improving educational attainments.

The focus of this article is on results from relatively large-scale, multiyear studies of diverse school-reform efforts that include measures of academic outcomes. While individual case studies are interesting and often enlightening, they cannot provide information as to generalizability of findings.

Further, we focus largely on studies conducted in the United States. There are a range of interesting projects in other countries, such as the Improving the Quality of Education for All efforts in diverse countries. However, many such studies have not focused clearly on measures of student achievement. By contrast, the United States offers a deep history and a substantial number of mixed-method, outcome-focused studies. We invite readers to consider the findings and to independently estimate their relevance to their countries' contexts.

The United States as Context

There are several characteristics of US education that make it relevant for study. With over 300 000 000 citizens

from around the world, the United States is both large and diverse. The US education system is unusually decentra-
lized, with 50 state governments and over 15 000 local education authorities (LEAs) governing schools and attempting to hold schools accountable for a wide array of outcomes and measures. On achievement measures ranging from those used for the National Assessment of Educational Progress (NAEP) to the tests of the Third International Math and Science Studies (TIMSS), states and LEAs range from not statistically different from the highest-achieving countries worldwide to achieving at levels comparable to Third World nations.

Throughout America's history it has also been the case that Americans have been perpetually unsatisfied with the *status quo* and hence highly likely to change a broad range of governmental and other social conditions.

More specifically, Americans are perpetually unsatisfied with the quality of their schooling. At the end of the nineteenth century, Harvard's much-admired president C.W. Eliot complained that the main characteristic of instruction in American schools was a lack of stimulating, interesting tasks. Seventy-five years later John Goodlad summarized data from his large-scale studies and concluded that student boredom was the norm of American schooling.

At a policy level, the 30 years following Goodlad's finding have seen repeated large-scale efforts to change core schooling practices. Striking examples include the New American Schools initiative begun by the first President Bush in the early 1990s and the federally funded Comprehensive School Reform program. Given the aforementioned national penchant for change, perhaps it is not surprising that both of those efforts subsequently have been abandoned by the federal government – Americans perpetually change their changes.

The question remains, however, about how to best approach change to produce consistent improvement in adults' educational practices and student outcomes in schools.

Large-Scale School Improvement Studies in the United States

Fortunately, an extensive body of research on large-scale efforts to reform American schools, spanning most of the twentieth century and continuing, has yielded consistent findings regarding key issues in whole-school change.

Several of these findings are robust despite greatly differing historical conditions, reform goals, change strategies, and cultural norms within which the studies were conducted. Several of the most significant of these US studies are briefly described below. These are followed by cross-study findings and implications for future research and achievement-enhancing reform.

The Eight-Year Study of the 1930s employed a quasi-experimental design to examine the impact that substantial high school reforms would have on students' achievement in college. Thirty experimental schools were matched to 30 similar comparison schools, and arrangements were made with 300 colleges and universities to alter their admission standards to allow freedom for the experimental schools to develop local, innovative curricula. External change agents, known as the Curriculum Associates, used the reform goals espoused by the Progressive Education Association to work with the experimental schools to facilitate whole-school reform. In addition to collecting qualitative data regarding the type of reforms attempted and the change process, experimental and comparison student performance in college were examined.

The RAND Change Agent study of the mid-1970s examined implementation of school programs and curricular innovations that received federal support for a trial period of implementation. The first phase of the study examined how the characteristics of the reform interacted with local conditions to affect project initiation and implementation in 293 sites, with a more intensive focus on 29 sites to permit comparisons of similar projects implemented in diverse settings. The second phase investigated project continuation in 100 sites after federal support ended.

The Follow Through Classroom Observation Evaluation (FTCOE) of the 1970s evaluated the implementation and outcomes of seven school restructuring models that had been developed by design teams receiving federal support. The goal of each model was to ameliorate the well-documented fade out of positive effects from federally funded preschool programs. This stoppage of fade was to be achieved by providing continuing intervention for disadvantaged children in the early elementary grades. FTCOE employed a quasi-experimental design, using low-inference classroom observations, cognitive and affective student outcome measures, and teacher questionnaires to determine whether (1) each model was implemented as intended, (2) variability in classroom instruction was a function of which model was implemented, and (3) the models had an impact on student outcomes.

One of the most comprehensive studies of federally supported educational innovation that focused specifically on the relative efficacy of various change

strategies was the Dissemination Efforts Supporting School Improvement (DESSI) from the late 1970s and early 1980s. DESSI was intentionally designed to provide a basis for comparing four fundamental strategies for encouraging innovation:

1. interpersonal linkage of validated practice,
2. commercial distribution of innovative curricula,
3. state-administered programs, and
4. local invention and development.

The DESSI findings were based on observations, interviews, questionnaires, site visits, and analysis of documents gleaned from nearly 5000 participants in 146 project sites.

Like DESSI, the Urban and Suburban/Rural Special Strategies for Educating Disadvantaged Students of the mid-1990s examined implementation of both locally developed and externally developed reforms, but, like FTCOE, also looked at the effectiveness of the strategies in terms of impact on student achievement. The externally developed reforms examined in the Special Strategies study included nationally disseminated whole-school programs that employed very different nationally developed approaches to change ranging from highly specified programs, like Success for All, to such loosely structured reform principals as the Coalition of Essential Schools. The locally developed reforms included school-wide projects (locally developed whole-school change efforts), extended-day and extended-year programs, and peer tutoring.

Special Strategies examined 25 sites that were nominated as exemplary on the basis of their implementation of the various strategies in terms of impact on student achievement, and the perceptions of students, teachers, administrators, and parents. Observations continued for 3 years.

In the last decade, large-scale studies of implementation and effects of the New American Schools designs and reviews of the Comprehensive School Reform literature have provided additional insight into externally facilitated whole-school reform efforts in the United States.

Most recently, efforts have been made to conduct meta-analyses of the effects of diverse reforms on the academic achievements of students. The largest and most thorough of those meta-analyses concluded that there were only three externally developed reform designs which had strong, consistent research support: Success for All, Direct Instruction, and the Comer School Development Program. At least ten other reform designs were producing net positive effects, but that the research bases in support of the designs were not yet strong. The second striking finding from that work was that reforms seemed to get their greatest effects if they persevered in schools for at least 5 years.

Findings from Large-Scale School Improvement Efforts

The previous section has illustrated the historic and intellectual breadth of studies of whole-school reform in the United States. The next section summarizes conclusions that can be drawn from those large-scale and other studies of whole-school reform efforts:

1. Virtually all educational improvement efforts include academic gain as one of their stated goals, but most studies of change efforts fail to find substantial evidence of academic gain resulting from most reform types. This frequent failure to achieve desired outcomes is significant and should receive more discussion. Meaningful improvements evaded the medical field for over 100 years until the majority of researchers concluded that most medicines had yielded no positive benefits. This realization led to a century of fundamental reexamination of what was known, hard research, and our current, often breathtaking, level of research-based medical advances. From the Eight-Year Study through studies of the New American Schools, most studies of most reforms have not yielded intended results. Acknowledging this fact proved greatly energizing in medicine, and we should do the same in education.
2. Externally developed reforms have proven more likely to produce desired outcomes than locally developed reforms. There have been instances in which locally developed innovations have been successfully implemented to good effect, but, at least in the published research, these instances are rare. The US Department of Education's What Works Clearinghouse recently released a report on 'Turning around chronically low-performing schools' in which they sought out strong evidence for turn-around school characteristics. The authors' conclusion was that the evidence for generalizability of any common characteristic of locally developed schools' academic improvement effects was low. This was in part because the What Works Clearinghouse reviewers excluded all externally developed designs from their analyses. By so doing, they left themselves with a least-likely-to-succeed data set. Across over a half-century's research, the best available evidence is that outcomes from reforms conceived and implemented purely locally are more likely to result in changes that are trivial, and often are more costly, than those that are the result of collaboration with external facilitators. Several of the previously noted large-scale studies found that schools were more successful at changing reading practices if they adopted and then adapted external programs. Locally developed programs attempted less-ambitious change, and staff

members became overwhelmed when confronted with the daunting task of development.

There are several possible explanations for this, and they are not mutually exclusive. Locally developed reform efforts tend to spend a year developing the reform. By the time they get to implementation, they may have exhausted their enthusiasm or their leadership may have moved away. Alternately, the group of local educators may choose a set of reform elements that, however attractive in the abstract, do not have track records of producing positive effects. Hence, even when well implemented, the reforms may not produce desired effects. Third, meaningful change is hard, and local educators may not be willing to push one another with sufficient force to cause necessary changes.

A fourth explanation is that the external group has gone through the change process many times in many locations. The processes and stresses of change are familiar to them, and they may have developed time-proven strategies for dealing with large-group change as a part of their job. By contrast, most local educators go through whole-school change efforts once or twice in a career. Regardless of the reasons, in studies over several decades, externally developed designs have proven more likely to produce specific, desired outcomes.

Across studies that included both whole-school and targeted-to-specific-subgroup reforms, whole-school reforms have tended to produce more effects. Pull-out programs, however attractive in the abstract, do not tend to have good track records. Among other plausible explanations for this finding, there is the issue of time for coordination between the regular classroom teacher and the pull-out specialist. US schools allocate precious little teacher time for lesson preparation and student materials development and scoring. They are left with virtually no time for coordination of lessons with pull-out teachers. A student receiving two sets of uncoordinated lessons in one subject area could hardly be blamed for not accelerating his learning.

3. A few specific whole-school reform designs – not all – have produced reasonably consistent evidence of positive effects. In the United States, Success for All, Direct Instruction, and the Comer School Development program have produced the most nearly stable evidence of positive impact on student achievement. Perhaps a dozen other reform designs have produced positive, although less-compelling evidence.
4. To be effective, external reforms and related professional development must have specificity. Multiple studies of the New American Schools redesign models concluded that designs requiring extensive local development were less likely to experience implementation success – typical schools simply did not have the

capacity to do their own development on the scale required for substantive change. Reform approaches that leave the lion's share of development to the school tend to result in increased variability in instructional practices and often result in staff anxiety and burnout.

External assistance in the form of change facilitation, wherein schools still bear the onus of development of specific program components, has not proved sufficient to produce change in most schools. External assistance that focuses on issues such as enhanced collaboration, redistribution of power, and facilitative leadership in the absence of specific guidance relative to curriculum, instruction, and assessment has repeatedly proven unlikely to produce substantive change. Reforms that are highly complex and ill-defined, yet require substantial change, typically result in reduced expectations in terms of goals, behavioral change, and reform outcomes.

Reform strategies that lack specificity and direction regarding instructional practices – such as peer coaching, site-based management, teaching teams, and organizational development models – generally fail to deliver changes in teaching or student outcomes because they are not linked to specific practices and do not link organizational changes to changes in individual teacher behavior.

Follow Through researchers found that the programs that were most likely to be successfully implemented and to have positive effects on student achievement were highly structured or even prescriptive. Likewise, the three corporate social responsibility (CSR) models that a recent meta-analysis indicated most likely to have strong positive effects on student achievement are characterized by high levels of specificity regarding core schooling functions. Both DESSI and Special Strategies researchers found that, although there was variation in implementation both between and within schools, externally developed models could be implemented with fidelity to their key characteristics. Indeed, the authors of DESSI felt that the most robust finding of the entire study was that teaching practices changed substantially in at least 50% of the schools that adopted externally developed programs.

5. In addition to specificity, whole-school change efforts must be accompanied by user-oriented assistance that provides support and guidance to school staffs as they implement the reform. The Rand Change Agent study found a strong relationship between extensive pre-implementation training, ongoing workshops, and locally available technical assistance. Support and assistance that address teachers' specific concerns related to implementing new strategies, delivered by a credible person who has classroom experience with the strategies, provide a powerful inducement to change in teaching behavior.

6. There are many more examples of successful change in classroom processes and student outcomes in the primary grades than in secondary schools. There are several plausible explanations for this finding as well. Secondary schools tend to be substantially larger than primary schools, and it is harder to change large organizations. Secondary students are older and perhaps less open to change. The content of high school instruction is inherently more advanced and more specialized; perhaps this requires more direct instruction. Within the folk wisdom of educators, high school teachers are regarded as having seen a greater range of student behaviors, and as having become more cynical about change in general. For whatever reason, high schools have proven to be particularly challenging environments in which to achieve lasting changes.
7. While a reform design team or other provider of external assistance presumably holds expert knowledge about both the intended changes and the processes of change, the experts on the realities of any given school are the professionals working within that school. Successful improvement efforts both recognize and build from both pools of expertise. The authors of the Rand Change Agent study noted that in schools there was either mutual adaptation between the design team and the in-school professionals or no meaningful change. Reflecting on recent multiyear studies of diverse reform efforts in large urban districts, scholars have referred to the same phenomenon in a more active voice as co-construction of reform. We are aware of no examples of sites in which a local school literally and uncritically adopted a reform *in toto*. Rather, in study after study, teachers and administrators obtained the greatest multiyear effects for their students by becoming active, engaged partners in the definition and conduct of improvement efforts in their schools.
8. LEAs have not received as much attention in research on educational change as have teachers or schools. However, emerging bodies of research suggest that LEAs can have substantial effects on schools' ability to implement any given reform. At the extreme, LEAs can mandate specific reforms, or the discontinuation of any given reform effort. More commonly, LEAs indirectly but clearly signal presence or absence of support for any given school-level reform efforts. Any of the above can be effective at enhancing or discontinuing a change effort.

Sustainability and Fidelity/Reliability

Two additional topics have received much less attention, and hence require a bit more speculation. The first relates to the issue of implementation fidelity/reliability. Since the 'eight-year study' of the 1930s, it has been clear that

those schools that engage in any given reform most fully and uniformly are the most likely to obtain desired changes in student outcomes. This has led to increasing calls from researchers and funders that local educators implement external reforms with high levels of fidelity. Several of the findings from large-scale research noted above probably are associated with success specifically because they improve the fidelity of an innovation. Whole-school efforts, specificity of desired changes, and ongoing professional development are examples of reform components likely to enhance implementation levels and, hence, fidelity.

Clearly without some sort of faculty fidelity to something beyond the *status quo*, school reform cannot happen. However, it is possible that calls for greater emphasis on fidelity to a specific, often abstract reform design are somewhat misdirected. Teachers and principals are not chemical reagents in a beaker. They are adults who have autonomy and agency. They tend to guard those characteristics as part of their perception of being members of a profession. Reforms are, in part, abstractions. By contrast, teachers deal with students in the concrete. No reform provides guidance for every situation students present and teachers face. Hence, teachers and administrators necessarily adapt the reforms they may be attempting to implement. Local professional educators co-construct the daily realities of reforms. These co-constructions will take somewhat different forms in classrooms beside one another and in schools thousands of miles apart.

It is possible that what is needed to achieve strong reform effects is to assist local educators in working with whole-school reform groups in reliably co-constructing reforms. Whereas high fidelity to specific components of a reform may not always be desirable, high reliability in implementing general principles may be.

There is 20 years of research on organizations that must work extremely reliably. Regardless of their place in the larger society, these very different organizations share characteristics that would be very useful in school-reform efforts. Highly reliable organizations (HROs) have a sharp focus on a small number of clear goals, have standard operating procedures, and, at the same time, they honor persons who find flaws in those procedures. They are vigilant to lapses of all sorts. HROs recruit and (re)train constantly, and they insist that efficiency drives do not reduce the reliability with which core tasks are performed. Interestingly, even within narrowly bounded types of organizations, such as air traffic control towers, no two HROs look exactly alike.

Although limited in quantity, extant research on efforts to help schools and districts become more like HROs has yielded promising, sometimes dramatic results. The implication is that school reformers may be well advised to acknowledge and honor local educators in their efforts to implement school-restructuring designs.

Finally, there is limited research on the sustainability of reforms and reform outcomes. If a reform is not going to produce lasting effects, it is probably not worth implementing. Studies of specific reforms tend to document a reform dropout rate, exceeding 50% over any given 5-year period.

Reforms go away for a variety of reasons. They may be viewed as expensive, and district administrators may conclude that they cannot afford them post-start-up funding. Alternately, a core group of teachers may retire or move to positions outside the school. A supportive principal or superintendent may receive a promotion. Often, the reform is replaced by the next great thing. Given such a history, teachers and administrators can and should be excused if they decline to tie their professional reputations and positions to any given reform effort.

On the more optimistic side of the sustainability issue is evidence that when reforms do last, their effects generally increase with time, particularly after the first 4 years. There are two plausible explanations for this. One is that the school or school system has finally, fully learned and integrated the reform, and its effects are most fully evident over the longer span of time. This is clearly true in some cases.

Equally true, the schools and districts most likely to drop a reform, or most likely to drop it first, are the schools and districts that either were unable to mount a credible change effort or, having mounted the effort, did not realize the hoped-for changes in student outcomes. If only the successful continue, then all who continue will appear successful, and the mean effect among the remaining participants will rise.

This later point should not be interpreted as taking anything away from the fact that some schools constitute existence proofs that positive, long-term change is possible.

Implications for Practice

Research on school reform has produced a range of implications for educational practitioners interested in achieving specific goals. The first several might be thought of as a gut check. Before beginning any reform effort, both the local authority and the school should ask themselves several questions: Can a consensus be achieved on a finite number of common goals? Do all concerned understand that the work of change will be much harder than they initially will hope? Do teachers, administrators, and funders possess the will and the resources to support change over several years? Are all willing to go at it hard? Are all willing to examine data on an ongoing basis to determine what fine-tuning is needed? Do they understand that change is resource intensive over time? Over the long haul, are all willing to stick to the change and resist the next new, shiny thing until they have mastered and grown

with the current reform? If the answer to any of the above is 'no' or 'maybe,' then the group should meet and discuss until all understand and agree to what change will require of all participants.

Assuming that the school and district are in agreement on goals and are prepared for the hard work of change, there are several additional implications of research on school change. A school and school district are more likely to succeed in producing overall student gains if they do the following:

1. use a research-proven whole-school reform;
2. work with experienced external-change facilitators;
3. support teachers and administrators in actively co-constructing the reform in their local context – passivity will not produce change;
4. monitor progress constantly – create data/information-rich environments, and use the data at multiple levels to help guide reform efforts; and
5. stick with the reform, and keep refining it over several years.

Implications for Research

Three of the most pressing implications of the above for future research are as follows:

1. *Build on what has come before.* One of the most striking and frustrating phenomena one encounters when reviewing school reform research is the regularity with which new research teams reinvent the wheel. There is a great deal of knowledge and wisdom that has come before, and new teams would be well advised to study it before gathering new data.
2. *Schools and districts are large, complex, multilayered organizations.* School reform researchers should choose methods that match the phenomena under investigation. This almost necessarily calls for mixed methods, multilevel research designs.
3. *In the area of educational reform, studies that only last 1 or two years are not worth doing, and may be misleading.* Given that previous research suggests that implementation takes years – and shifts over years, and given that reform effects seem to grow over time, studies of less than 3 years are likely to yield misleading conclusions, either positive or negative. Scholars would be well advised to plan 5-year (or longer) studies.

Summary

While progress in creating more effective educational reforms and progress in creating quality evaluations of school reform efforts have been equally slow, we know much more today about both the specific requirements

and the processes needed to create more effective-bearing reforms. As a profession, we educators can and should make much better use of the knowledge gained through decades of hard challenging research and practice. It is equally important that we do a better job of rigorously gathering data so that our future efforts can become yet more building blocks for the next generation of teachers, leaders, and educational reformers.

Bibliography

- Friedman, T. (2005). *The World is Flat*. New York: Farrar, Straus, and Giroux.
- Friedman, T. (2008). *Hot, flat, & crowded*. New York: Farrar, Straus, and Giroux.
- Toffler, A. (1970). *Future Shock*. New York: Random House.

Further Reading

- Aikin, W. M. (1942). *The Story of the Eight Year Study*. New York: Harper.
- Berman, P. and McLaughlin, M. W. (1977). *Federal Programs Supporting Educational Change: Vol. VI, Implementing and Sustaining Innovations*. Santa Monica, CA: RAND.
- Borman, G. D., Hewes, G. M., Overman, L. T., and Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research* **73**, 125–230.
- Comer, J. P. (1988). Educating poor minority children. *Scientific American* **259**, 42–48.
- Crandall, D. P. and Loucks, S. F. (1983). *A Roadmap for School Improvement: Executive Summary of the Study of Dissemination Efforts Supporting School Improvement*. Andover, MA: NETWORK.
- Datnow, A. (2005). The sustainability of comprehensive school reform models in changing district and state contexts. *Educational Administration Quarterly* **41**(1), 121–153.
- Datnow, A., Lasky, S., Stringfield, S., and Teddlie, C. (2006). *Systemic Integration for Effective Reform in Racially and Linguistically Diverse Contexts*. New York: Cambridge University Press.
- Datnow, A. and Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed At Risk* **5**(1–2), 183–204.
- Fullan, M., Hill, P., and Crevola, C. (2006). *Breakthrough*. Thousand Oaks, CA: Corwin.
- Goodlad, J. (1984). *A Place Called School*. New York: McGraw-Hill.
- Harris, A. and Chrispeels, J. (eds.) (2006). *Improving Schools and Educational Systems*. Oxon, UK: Routledge.
- Herman, R., Dawson, P., Dee, T., et al. (2008). *Turning Around Chronically Low-Performing Schools: A Practice Guide* (NCEE # 2008-4020). Washington, DC: National Center for Education Evaluation and Regional Assistance. Institute of Education Sciences, US Department of Education. <http://ies.ed.gov/ncee/wwc/practiceguides> (accessed September 2009).
- Hopkins, D. (ed.) (2005). *The Practice and Theory of School Improvement: International Handbook of Educational Change*. Dordrecht, NL: Springer.
- National Assessment of Educational Progress (NAEP) (2008). *The Nation's Report Card*. <http://nces.ed.gov/NATIONREPORTCARD> (accessed September 2009).
- Nunnery, J. A. (1998). Reform ideology and the locus of development problem in educational restructuring. *Education and Urban Society* **30**(3), 277–295.
- Reynolds, D., Hopkins, D., Potter, D., and Chapman, C. (2001). *School Improvement for Schools Facing Challenging Circumstances: A Review of Research and Practice*. Exeter, UK: University of Exeter.
- Slavin, R. and Madden, N. (2001). *One Million Children: Success for All*. Thousand Oaks, CA: Corwin.

- Stallings, J. and Kaskowitz, D. (1974). *Follow Through Classroom Observation Evaluation (1972–1973)*. Menlo Park, CA: Stanford Research Institute.
- Stringfield, S., Millsap, M., and Herman, R. (1998). Using “promising programs” to improve educational processes and student outcomes. In Hargreaves, A., Lieberman, A., Fullan, M., and Hopkins, D. (eds.) *International Handbook of Educational Change* pp 1314–1338. Dordrecht: Kluwer.
- Stringfield, S., Reynolds, D., and Schaffer, E. (2008). Improving secondary students’ academic achievement through a focus on reform reliability: 4- and 9-year findings from the high reliability. *Schools Project School Effectiveness and School Improvement* **19**(4), 409–428.
- Teddlie, C. and Reynolds, D. (2001). Countering the critics: Responses to criticisms of school effectiveness research. *School Effectiveness and School Improvement* **12**(1), 41–82.
- Townsend, T. (ed.) (2007). *International Handbook of Research on School Effectiveness and Improvement*, pp 131–166. Dordrecht, NL: Springer.
- US Department of Education (2008). What Works Clearinghouse. What Works Clearinghouse. <http://www.ed.gov/about/offices/list/ies/ncee/wwc.html> (accessed September 2009).

Relevant Website

<http://timss.bc.edu> – Trends in International Mathematics and Science Study (TIMSS).

This page intentionally left blank



QUALITATIVE RESEARCH

Contents

Action Research in Education
Classroom Ethnography
Computer Assisted Qualitative Data Analysis
Conversational Analysis
Critical Ethnography
Critical Race Theory
Critical Theory
Cultural Studies
Cultural–Historical Activity Theory
Design Experiments
Discourse Analysis
Document Analysis
Ethnography
Ethnomethodology in Education Research
Focus Groups
Grounded Theory
Hermeneutics
Interpretive Research
Interviews and Interviewing
Life History
Narrative Inquiry
Participant Observation
Phenomenology
Qualitative Case Studies
Qualitative Data Management
Semiotics
Symbolic Interactionism, Naturalistic Inquiry, and Education
The Quality of Evidence in Qualitative Research
Visual Data in Education Research

Action Research in Education

E Stringer, Curtin University of Technology, Perth, WA, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Critical theory – A theoretical perspective that provides the descriptive and normative basis for approaches to social inquiry aimed at

decreasing domination, and increasing freedom in all their forms.

Dialogical – Dialogs in which two parties exchange arguments over a central claim.

Ideology – A system of political ideas.

Living theory – Explanations derived from an individual's lived experienced and embodied and expressed in practice and behavior.

Paradigm – An interpretive framework guided by a set of beliefs and feelings about the world, and about how it should be understood, studied and presented.

Participatory research – The perspective that suggests that those formally known as subjects of study should participate in every part of the research process – planning, data collection, analysis, and reporting.

Phenomenology – The study of structures of consciousness as experienced from the first person point of view.

Postmetaphysical – Often associated with Habermas, this philosophical perspective suggests theories of society and communication grounded in the inherent complexities of social interreaction – a plurality of language games, differing modes of power, and the structure of politics at play at any one time.

Pragmatism – A philosophical tradition focusing on the need to trace the practical consequences of a concept or hypothesis.

Primary stakeholders – Those most immediately affected by the issue investigated.

Stakeholder – A person or group having an interest in, or being affected by an issue.

Introduction: Situating Action Research

Action research has recently gained increasing credibility as an approach to research that challenges the prevailing orthodoxies of qualitative and quantitative research. Despite considerable resistance from an academic establishment that has built complex systems of administration and professional reward around the basic assumptions of the traditional paradigms, action research has rapidly created new spaces within the research community. It has moved the boundaries and goalposts of investigation, and provided windows into forms of inquiry and new ways of conceiving what counts as legitimate academic knowledge. As Guba (2007) noted "... there is indeed more in heaven and earth than has been dreamed of in the received philosophy."

Like qualitative research that emerged from relative obscurity in the period from the late 1960s onward to challenge experimentalism as the only acceptable form of research, action research is now increasingly acknowledged as a legitimate approach to investigation. Widely accepted in academic journals, and the subject of many university texts, increasing sales clearly indicate its

growing use in academic courses. Action research has achieved particular acceptance in education, where a wide variety of resources speak of its expanding popularity in teacher education and professional-development programs, and presentations and papers related to action research continue to proliferate in academic journals and conferences.

Action research is particularly attractive to educational practitioners. It enables them to break free from the complex systems and patterns of thought associated with more traditional approaches to inquiry, providing simple and coherent formats that are easily applied to problems and issues within classroom, school and other educational contexts. Whether these processes are sufficiently stable to have generalized use over a broad range of contexts, or whether it is an inherently localized production suited only to a limited context is still subject to debate. What is clear, however, is that the move to action research provides a clear indication of the need for processes of investigation that deliver effective solutions to many of the enduring problems confronting practitioners in educational contexts.

The History of Action Research in Education

The genesis of action research is often attributed to Kurt Lewin, a social scientist who sought constructive solutions to social conflict, particularly in situations when poverty and minority exploitation were predominant issues. His approach to action research (Lewin, 1946, 1948, 1952; Lewin and Lewin, 1942) drew on theories of progressive education from philosophers like Dewey (1938) who saw learning as emerging from processes of inquiry, questioning, and dialog. Lewin was interested in assisting people to learn how to solve their own problems through self-education that would enable them to improve their situation. Lewin presented action research as a cyclical or reiterative process that included fact finding, planning, action, reflection, evaluation, and refinement of the problem.

These types of investigation emerged in a variety of forms in a number of locations and were later applied to education by Stenhouse (1975, 1983) and Elliott (1987, 1990, 1991) in Britain, and Kemmis and McTaggart (1988) in Australia. The general thrust of their activity supported the notion of teacher as researcher, investigation being oriented to issues relevant to everyday classroom practices and problems. These types of development emerged more broadly from the profession's frustrations with research that failed to provide effective solutions to problems of professional practice and more obdurate problems within education systems – equity, class, race, gender, and so on.

Broad dissatisfaction with the inefficacy of the abstract and generalized experimental and quasi-experimental research that originally dominated university investigations of the period following World War II led to the emergence of qualitative research. This signaled a movement toward more localized research efforts typified by ongoing developments within qualitative research/naturalistic inquiry. The reason, suggested Guba (2007), lay in “the gap between general laws and specific applications; such laws can have, at best, only probabilistic implications for specific cases. The fact, for example, that 80% of patients presenting a given set of symptoms are likely to have lung cancer does not imply that a particular patient with those symptoms ought to immediately be rushed to surgery.” “It ought to be apparent by now,” he went on to say, “that generalized one-size-fits-all solutions do not work. . . . Without intimate knowledge of local context, one cannot hope to devise solutions to local problems. *All* problems are de facto local; inquiry must be decentralized to the local context” (p. xi).

While qualitative research initially showed promise of illuminating critical features and dynamics of the educational world, it ultimately failed to provide the means to develop effective solutions to educational issues that had dogged the educational establishment for many decades. The movement from qualitative to action research was at first slow, but is well chronicled in a number of texts and papers (e.g., Adelman, 1993; Foshay, 1994; King and Lonquist, 1992; McTaggart, 1991). Noffke’s (1997) review of this literature revealed a burgeoning field, and since that time a veritable explosion of activity has seen action research become an increasingly visible entity within educational research. Whether this movement signals the means to make headway against those issues and problems is yet to be realized, although there are signs that in localized contexts, rigorously applied action-research processes often produce effective outcomes that resolve the issues or problems on which investigation has focused.

Defining Action Research

Definitions of action research differ considerably, with writers responding variously to the diverse contexts, histories, and issues that provide the boundaries of their experiences and perspectives. Reason and Bradbury (2008) in the *Handbook of Action Research* define action research as:

... a participatory process concerned with developing practical knowing in the pursuit of worthwhile human purposes. It seeks to bring together action and reflection, theory and practice, in participation with others, in the pursuit of practical solutions to issues of pressing concern to people, and more generally the flourishing of

individual persons and their communities” (Reason and Bradbury, 2008: 4).

In education, action research has been defined as:

- ... a systematic approach to investigation that enables people to find effective solutions to problems they confront in their everyday lives (Stringer, 2007: 1).
- ... any systematic inquiry conducted by teacher researchers, principals, school counsellors, or other stakeholders in the teaching/learning environments to gather information about how their particular schools operate, how they teach, and how well their students learn (Mills, 2007: 5).
- ... a form of collective, self-reflective enquiry undertaken by participants in social situations in order to improve the rationality and justice of their own social or educational practices, as well as their understanding of these practices and the situations in which these practices are carried out (Kemmis and McTaggart, 1988: 6).

The diversity of definitions can clearly be attributed to the contexts in which its various practitioners worked, the diverse processes used, and the different types of outcomes they sought. As is evident below, differing theoretical and methodological orientations are often associated with the different distinctive purposes to be achieved. Although common threads permeate these definitions, the most evident being the pragmatic purposes to be attained through systematic investigation, there is even more variation in the enactment of these definitions. As becomes evident in the following section, philosophical and political orientations create wide disparities in the way action research is both conceptualized and practiced.

Orientations to Action Research

For some, action research is a purely pragmatic activity that focuses primarily on teacher practices, with classroom instruction and student performance outcomes being the major purpose of engaging action research (e.g., Mills, 2007; Koshy, 2005; Holly *et al.*, 2008; Baumfield *et al.*, 2008). The methods reflect an extension of commonly accepted classroom teaching practices, wherein teachers choose a promising practice that might enhance student learning. These are then implemented to see if they achieve the desired outcomes – often improved performance scores. This approach is particularly popular in preservice teacher education, since it requires little theoretical or methodological sophistication and links easily to readymade lessons for novice teachers.

This family of approaches tends to be modeled on first-person investigations (Reason and Bradbury, 2007), practitioners contending with no perspectives other than there

own, except for the voices inherent in the associated academic or professional literature. Clearly, this mode of operation reflects directive approaches to classroom instruction, or orientations to research that valorize the literature review. Models associated with this orientation range from the relatively simple plan–act–reflect–evaluate (an often simplistic application of a model suggested by Carr and Kemmis, 1986), to more sophisticated versions that include feedback loops (Mills, 2007). The bipolarity of this orientation to action research is evidenced in the vexed question of many researchers whether to attend the ‘Action Research’ special interest group or the ‘Teacher Researcher’ special interest group, which are usually scheduled at the same time at annual meetings of the American Educational Research Association.

These more pragmatic forms of action research differ considerably from the more ideologically driven approaches developed by Kemmis and his associates at Deakin University (Carr and Kemmis, 1986; Kemmis and McTaggart, 1988, 2005). Grounded in the critical theory of Jurgen Habermas, their work evolved to present a participatory, critical approach to action research that seeks to take into account many of the agendas emerging from recent scholarship in qualitative research. It incorporates elements of the work of Lewin (1952) that involves participants collectively in the process of investigation, Stenhouse’s (1975) notion of teacher as researcher, and Schon’s (1983, 1987, 1991) reflective practitioner (Kemmis, 2008).

Kemmis (2008) critiques much current educational action research as containing naive views that truth can be apprehended by knowing subjects, alluding to the common view that expertise derived from professional training can provide sufficient understanding to solve workplace problems. He applies a Habermasian post-metaphysical philosophy in which truth becomes manifest only through exploration in communicative action, where participants arrive at intersubjective agreement, mutual understanding, and consensus as the basis for practical action. This clearly postmodern view casts doubt on the ability of the master texts of government and other centralizing agencies to organize decisive unitary action because of the contested interactions that operate at all levels of these systems.

Kemmis therefore suggests the need for action research to work not just on self-reflections and realizations of individuals, or on the operation of organizations and institutions, “but in the interstices between people and organizations, and across boundaries between life-worlds and systems” (p. 123). His definition of action research includes “research undertaken collectively by participants in social practice to achieve historical self-consciousness . . .” (p. 135), a praxis that is, as morally informed, committed, action-oriented, and informed by tradition. The focus is therefore on collective deliberation aimed at shared self-understanding about particulars of a practical situation.

This approach to investigation research clearly has much in common with that presented by Stringer (2007, 2008) whose suggests a phenomenologically oriented participatory process grounded in a methodology that takes into account the sensitivities emerging from postmodern, critical, and feminist scholarship. Taking a more pragmatic stance, Stringer suggests that effective and sustainable action can only emerge from processes that incorporate the voices of all stakeholders, focusing particularly on primary stakeholders as the foundational text on which ongoing participatory processes of investigation are grounded. This orientation, suggests Stringer, provides the context by which democratic, empowering practices emerge as a natural consequence of the application of action research to the knowledge-production process, thus breaking down the ability of power elites to impose their definitions of the situation (Foucault, 1972). Like Kemmis, however, Stringer grounds his research in processes by which interacting individuals are able to construct mutual understanding and consensus as the basis for practical action, the pragmatism of Rorty (1981, 1989) therefore lying clearly in the background. Stringer suggests that social change emerges not from application of predetermined political agendas, but by the democratic operation of participatory processes that enable the voices of the less powerful to change the actions, procedures, and operational structures of organizations and institutions that act as stakeholding entities within the research process. Stringer is affected by his long experience working alongside Australia’s marginalized Aboriginal peoples in community and organizational contexts, a position from which he constantly challenges the master texts of the academic, political, and organizational world.

Although both Stringer and Kemmis relate their work to critical theory, they differ considerably in orientation from the social activist approaches to action research presented by adherents to participatory action research (PAR) who take a more conflictual, confrontative approach to action research. Based on a Marxian or neo-Marxian perspective, PAR advocates emerge from traditions of social activism grounded in historical political struggles common to the contexts class and union struggles or in the exploitation and disempowerment experienced by many people in developing nations. The work of Friere (1970, 1998a, 1998b) often provides an orientation to this perspective, but activist scholars (Fals Borda, 1979; Rahman, 2008; Fals Borda and Rahman, 1991) directly engage an approach to action research that consciously and openly challenges the structures and social processes sustaining the interests of power elites.

Within education, Kincheloe and McLaren (2005) provide examples of this theoretical orientation, although its application to action research also is exemplified in the work of Brydon-Miller *et al.* (2004) whose primary focus is on interrogation of systems of power and privilege within

social, organizational, and institutional environments. In a similar voice, Fine and Torre (2008) present an activist approach to action research that focuses on questions of audience, product, and provocation. Working with and from the perspectives of those with least power, they seek products that will most effectively provoke action leading to social change, or changes within a system. They seek to redress injustices through interrogating the operation of those systems, reflecting a confrontational and overtly political orientation that reflects much political life in the Americas. Typical of those who work within this genre, Fine and Torre therefore struggle for social justice in institutions with deeply entrenched operational systems that do nothing to change the social structures from which troubles and crises emanate. They quote Bordieu (1998) who speaks of "... a crisis of politics ... [in which we encounter] ... despair at the failure of the state as guardian of the public interest." The focus of PAR-oriented action research is therefore on an activist rather than dialogical critical theory, oriented to social organizing within unions and collectives, with the intent of interrogating, challenging, and disrupting the operation of institutional systems.

A more reflective approach to investigation is presented by Whitehead and McNiff (2006) who base action research on 'living theory' derived from people's lived experience. They distinguish the explanations of action researchers from the general explanations of propositional theory associated with the social and behavioral sciences. Living theory is that which individuals generate to explain educational influences in their own experience, and is grounded in a flow of life-affirming energy and relationally dynamic awareness that give meaning and purpose to the lives of individuals. The primary purpose of exploring living theory is to develop forms of understanding that illuminate features of educational life or generate ideas to improve teaching practices.

These differing approaches are encapsulated in the recent release of the *Handbook of Educational Action Research* (Noffke and Somekh, 2009). This presents a rich overview of the different traditions of action research, describing the diverse rationales and practices applied in educational settings, and broadly differentiating them according to professional, personal and political orientations inherent in the educational action research literature.

Applications of Action Research in Educational Settings

The dramatic increase in publications related to educational action research is evidenced by the wide range of texts now available. An incomplete list, covering a variety of approaches to action research includes authors from Britain, the United States, and Australia. Texts that focus

on the use of action research to improve teacher practices and/or student outcomes include those by Johnson (2002, 2007), Koshy (2005), Mertler (2005), Calhoun (1994), Brown and Dowling (1998), and Burnaford *et al.* (2001). In the same vein, Meyers and Rust (2003) and McNiff and Whitehead (2006) describe how teachers can use action research to engage classroom issues, formulate professional-development programs, and develop effective instructional practices.

Texts that focus on broader curriculum, ethical and legal issues in school settings include those by Holly *et al.* (2008), Armstrong (2004), Bray *et al.* (2000), Tomai (2003), Anderson, *et al.* (2007), and Somekh (2005). Phillips and Carr (2006) further extend the array of applications by describing how action research can increase the formation of professional identity for preservice teacher-education students, and Mclean *et al.* (2005) reveal the way teachers and administrators can employ it to find solutions to school problems.

Sagor (2004) uses action research for classroom improvement while Glanz (2003) and Hendricks (2008) suggest how it might be applied to school improvement. Some of this type of literature is clearly directed to educational and social change (O'Hanlon, 2003; Brown and Jones, 2002; Berge and Ve, 2000; Atweh *et al.*, 2005; and Christiansen *et al.*, 1997). Pedraza and Rivera (2005) likewise focus on school reform through educational research within Latino communities. A number of texts focus on specific areas of educational life – English-language teaching (Burns, 1999; Wallace, 1998), Teachers of English to Speakers of Other Languages (TESOL) (Edge, 2000) media within school libraries Sykes (2002), educational policy (Hollingsworth, 1997; McTaggart, 1997; Caro-Bruce *et al.*, 2007; Pedraza and Rivera, 2005; Ladson-Billings and Tate, 2006; Cammarota and Fine, 2008), Black Education (King, 2005), teacher education (Kitchen and Stevens, 2008), and educational leadership (James *et al.*, 2007).

These citations represent but a small proportion of the now voluminous literature featuring applications of action research to a whole range of educational contexts, endeavors, populations, and professional purposes. Most journals that focus on qualitative research in education incorporate action-research studies and, as becomes evident below, the dramatic increase in relevant websites further extends the range of this increasingly popular methodology.

Representation in Educational Action Research: Theater, the Arts, and Media

This diversity of literature reflects changes in all facets of educational research, requiring research facilitators and participants to take seriously the movements in qualitative research that evolve from questions of representation (Denzin, 1997, 2005). The participatory impetus common

to many action researchers has impelled them to find appropriate and effective ways of communicating both the processes and outcomes of action research projects to their diverse stakeholders and audiences. Academic journals and texts that have been the major vehicle for communicating research outcomes are now complemented by diverse formats that act for this purpose. Thus, we find theater (Guhathakurta, 2008), music, art, poetry, role play, video, and multimedia presentations providing powerful and effective means for transmitting the major messages emerging from action-research processes in educational settings (Stringer, 2008, Jack Whitehead's Web Page).

Whitehead (2009) focuses particularly on multimedia presentations and representations incorporating visual narratives that provide the means to communicate meanings not possible through printed texts. This approach to research is particularly significant as it opens up possibilities to incorporate and communicate voices that have traditionally been silenced by the text-based reports. The possibilities of including voices of parents, students, marginalized groups, and others are considerably enhanced by the ability to provide visual and aural narratives in a variety of creative forms. These movements away from purely textual modes of representation reflect an ideological position that challenges the right on an academic elite to control the processes of knowledge production. They also signal an ideological position that acknowledges the diverse ways of knowing required to give voice to, or represent the perspective of the disparate groups participating in an action research project (Heron and Reason, 2008; Kowlaski, *et al.*, 2008).

Website Resources

The desire to represent and communicate knowledge in diverse forms is facilitated by the burgeoning impact of electronic media, a continually expanding number of sites and resources providing access to information related to educational action research. A small sample of sites that have general relevance to action research in education includes sites for the Centre for Action Research in Professional Practice at the University of Bath, AR Expeditions, and the *Action Research Journal*. These types of sites provide a broad range of resources, comprising articles describing action-research projects and strategies for conducting action research. Other general sites include Action Research Resources that provides an online journal, access to an online course, useful papers, and other general resources, and Action Research at Queen's University that supplies links to programs, conferences, other action-research sites and resources, publications, and reports. Action Research on the Web Help (AROWHELP) provides a useful set of resources, including journals and links for teacher researchers. Similar resources can be accessed at the Action Research Page

from the University of Colorado at Denver School of Education, Jack Whitehead's Home Page that offers a wide range of offerings especially pertinent to schools and education, including examples of action-research theses, and Research for Action (RFA) the site of a non-profit organization engaged in educational research and reform to improve educational opportunities and outcomes for all students. These general resources are complemented by many others, including AROW (Action Research Open Web), and Links2Go: Action Research.

Sites relevant to teacher action research have recently proliferated and the following providing but a small sample of the type of resources available:

1. What is Teacher Action Research? Presents guidelines for teacher researchers.
2. Action Research Network provides resources and links for teacher researchers.
3. Action Research Rochester Teacher Centre Teacher Research Program provides a variety of useful resources for teacher researchers
4. Action Research Overview presents resources for teacher researchers that are oriented toward the development of a democratic, student-oriented learning community for improving teaching and learning for all.
5. Teacher Research (Action Research Resources) lists magazine and journal articles about action research, some of which are presented in full text form.

More specialized sites are also available:

1. Science Education Action Research provides information about instructional processes and teaching/learning materials for science education.
2. Early Childhood Education presents information about the use of action research in early-childhood education.
3. School Improvement describes how a district instructional leadership team used action research to conduct a 2-year study of three aspects of the district strategic plan.
4. School Renewal describes how a consortium of schools used action research processes as a framework for school renewal.
5. Action Research Listserv (AELACTION): Action Research in Schools and Classrooms describes five action research projects that show how participants select a focus, choose data collection methods, and reflect on findings.
6. An Action Research Module describes action research in primary and secondary schools.

While many of these sites are relatively stable, the general picture is one of a quantum foam of sites, some appearing and disappearing randomly, according to fluctuations in personnel, funding, and priorities at the institutions and organizations responsible for their maintenance.

Nevertheless, websites now provide what is rapidly becoming a major source of information about the trajectory and operation of action research in education.

Action Research in Higher Education

With the spectacular growth of action research as an approach to investigation that is superficially appealing to many researchers and professional practitioners, it is clear that it would suffer the pains of unsustainable growth. Many of the texts and projects that now emerge in the literature are not well grounded in any sense, but have been enacted within the common-sense worldviews of naive practitioners, or traditional researchers who merely impose their own frameworks of inquiry onto what is essentially quite a different approach to investigation. These become embedded in texts and related university programs and take on a life of their own.

Levin (2008) takes task with the superficiality of much of this research, and outlines the skills and knowledge needed by action researchers in order to ensure rigorous and effective implementation of this approach to inquiry. He suggests that practitioners need to develop the ability to support involvement in actions as well as the capability to critically reflect on process and outcomes of action research. He identifies the proficiency to work with social change to solve participants' problems, the skill to sustain collaborative learning processes, and the competencies required to effectively communicate the outcomes of a research process. These, he suggests, demand development of a broader spectrum of capacities than those required of more traditional un-involved approaches to research, with concurrent need to incorporate them into professional-development processes that are an integral part of higher education programs.

This theme is part of a wider call for reform of universities to provide them with the capacity to take advantage of developments in research methodologies and develop practical and enduring solutions to seemingly intractable social problems. (Levin and Greenwood, 2008; Sarason, 1990). Levin (2008: 211) suggest action research is one of the means to overcome the disconnection between the university and the public good, providing a "knowledge generation that is intrinsically capable of producing public goods through concrete and practical problem-solving..." (2008, p. 211) and that this may be achieved by using action research processes to restructure the teaching and research role of universities.

International Educational Action Research

Although this review speaks mainly to English-speaking countries, educational action research has spread widely

across the world and is practiced in countries as diverse as Scandinavia, Greece, Mexico, China, Korea, Japan, Thailand, Singapore, India, and others in Latin America. It is likely that activity in these countries will, because of social and cultural differences, challenge many of the orthodoxies of action research, and bring to question some of the underlying issues that sit comfortably beneath the surface of English-speaking countries. This has already been signaled in previous sectors of this article, where the revolutionary impulse embedded in the history of some South American countries has spawned a strongly critical approach to action research. It is likely that the Confucian traditions embedded in Chinese contexts (Sung-Chan *et al.*, 2008) and traditional structures of authority in tribal settings (Kirk, 2008) will likewise bring to question the egalitarian and collaborative perspectives common to many adherents to action research. The continued spread of action research across the globe is sure to both enrich and enhance the development of the genre.

Conclusion

Although there is dispute within the field about the nature of action research – whether it is essentially qualitative, can be quantitative, or both – its emergence has served to spawn rich and enlightening debate that pushes the academic world to clarify the nature, purposes, methods, and outcomes of forms of investigation that wish to be recognized as legitimate research. Action research enables academic and professional researchers to enhance their practice by engaging research processes that are constrained by the strictures and rigidities of experimental method, or perplexed by the complexity of multi-voiced narratives emanating from the postmodern turn in qualitative research. To use a cosmic metaphor, is it possible that the inflating and undifferentiated universe of perspectives of early big-bang postmodernism might be brought by action research to a more ordered accumulation of matter that forms more stable systems of thought? What is clear, however, is that action research continues to emerge as a significant research entity in education, providing the means to engage in rigorous and systematic inquiry that provides practical solutions to many educational problems.

See also: Critical Theory; Ethnography; Interpretive Research; Visual Data in Education Research.

Bibliography

- Adelman, C. (1993). Kurt Lewin and the origins of action research. *Educational Action Researcher* 1(1), 7–25.
- Anderson, G., Herr, K., and Nihlen, A. (2007). *Studying Your Own School: An Educator's Guide to Qualitative Research*, 2nd edn. Thousand Oaks, CA: Corwin.

- Armstrong, F. (2004). *Action Research for Inclusive Education*. New York: Routledge Falmer.
- Atweh, B., Weeks, P., and Kemmis, S. (2005). *Action Research in Practice: Partnerships for Social Justice in Education*, 2nd edn. New York: Routledge.
- Baumfield, V., Hall, E., and Wall, K. (2008). *Action Research in the Classroom*. Thousand Oaks, CA: Sage.
- Berge, B.-M. and Ve, H. (2000). *Action Research for Gender Equity*. Buckingham: Open University Press.
- Bordieu, P. (1998). *Acts of Resistance: Against the Tyranny of the Market*. New York: New Press.
- Bray, J., Lee, J. A., Smith, L. L., and Yorks, L. (eds.) (2000). *Collaborative Inquiry in Practice: Action, Reflection, and Making Meaning*. Thousand Oaks, CA: Sage.
- Brown, A. and Dowling, P. (1998). *Doing Research/Reading Research. A Mode of Interrogation for Education*. London: Falmer.
- Brown, T. and Jones, L. (2002). *Action Research and Postmodernism: Congruence and Critique*. Buckingham: Open University Press.
- Brydon-Miller, M., Maguire, P., and MacIntyre, A. (2004). *Travelling Companions: Feminism, Teaching, and Action Research*. Westport, CN: Praeger.
- Burnaford, G. E., Fischer, J., and Hobson, D. (eds.) (2001). *Teachers Doing Research: The Power of Action through Inquiry*, 2nd edn. Mahwah, NJ: Erlbaum.
- Burns, A. (1999). *Collaborative Action Research for English Language Teachers*. Cambridge, UK: Cambridge University Press.
- Calhoun, E. (1994). *How to Use Action Research in the Self-Renewing School*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Cammarota, J. and Fine, M. (2008). *Revolutionizing Education: Youth Participatory Action Research*. New York: Routledge.
- Caro-Bruce, C., Flessner, R., Klehr, M., and Zeichner, K. (2007). *Creating Equitable Classrooms through Action Research*. Thousand Oaks, CA: Corwin.
- Carr, W. and Kemmis, S. (1986). *Becoming Critical: Education, Knowledge, and Action Research*. Philadelphia, PA: Falmer.
- Christiansen, H., Goulet, L., Krentz, C., and Maeers, M. (eds.) (1997). *Recreating Relationships: Collaboration and Educational Reform*. Albany: SUNY Press.
- Denzin, N. K. (1997). *Interpretive Ethnography*. Thousand Oaks, CA: Sage.
- Denzin, N. (2005). Emancipatory discourses and the ethics and politics of interpretation. In Denzin, N. and Lincoln, Y. (eds.) *Handbook of Qualitative Research*, 3rd edn., pp 933–958. Thousand Oaks, CA: Sage.
- Dewey, J. (1938). *Experience and Education*. New York: Macmillan.
- Edge, J. (2000). Action research (Case studies in TESOL practice). Teachers of English to Speakers of Other Languages (Amazon.com).
- Elliott, J. (1987). Educational theory, practical philosophy, and action research. *British Journal of Education Studies* 35, 149–169.
- Elliott, J. (1990). Teachers as researchers: Implications for supervision and for teacher education. *Teaching and Teacher Education* 6(1), 1–26.
- Elliott, J. (1991). *Action Research for Educational Change*. Philadelphia, PA: Open University Press.
- Fals Borda, O. (1979). Investigating reality in order to transform it: The Columbian experience. *Dialectical Anthropology* 4, 33–55.
- Fals-Borda, O. and Rahman, M. (1991). *Action and Knowledge: Breaking the Monopoly with Participatory Action Research*. New York: Apex.
- Fine, M. and Torre, M. E. (2008). Theorizing audience, products and provocation. In Reason, P. and Bradbury, H. (eds.) *Handbook of Action Research*, 3rd edn., pp 407–719. Thousand Oaks, CA: Sage.
- Foshay, A. (1994). Action research: An early history in the United States. *Journal of Curriculum and Supervision* 9, 317–325.
- Foucault, M. (1972). *The Archaeology of Knowledge*. New York: Random House.
- Friere, P. (1970). *A Pedagogy of the Oppressed*. New York: Seabury Press.
- Friere, P. (1998a). *Pedagogy of the Heart*. New York: Continuum International.
- Friere, P. (1998b). *Teachers as Cultural Workers: Letters to Those Who Dare*. Boulder, CO: Westview.
- Glanz, J. (2003). *Action Research: An Educational Leader's Guide to School Improvement*, 2nd edn. Norwood, MA: Christopher Gordon.
- Grossman (2008).
- Guba, E. (2007). Foreword. In Stringer, E. (ed.) *Action Research*, 3rd edn, pp ix–xiv. Thousand Oaks, CA: Sage.
- Guhathakurta, M. (2008). Theatre in participatory action research: Experiences from Bangladesh. In Reason, P. and Bradbury, H. (eds.) *Handbook of Action Research*, 3rd edn., pp 510–521. Thousand Oaks, CA: Sage.
- Hendricks, C. (2008). *Improving Schools through Action Research: A Comprehensive Guide for Educators*, 2nd edn. Columbus, OH: Allyn and Bacon.
- Heron, J. and Reason, P. (2008). Extending epistemology within co-operative inquiry. In Reason, P. and Bradbury, H. (eds.) *Handbook of Action Research*, 3rd edn., pp 366–380. Thousand Oaks, CA: Sage.
- Hollingsworth, S. (ed.) (1997). *International Action Research: A Casebook for Educational Reform*. London: Falmer.
- Holly, M., Arhar, J., and Kasten, W. (2008). *Action Research for Teachers: Travelling the Yellow Brick Road*, 3rd edn. Upper Saddle River, NJ: Prentice Hall.
- James, E., Milenkiewicz, M., and Bucknam, A. (2007). *Participatory Action Research for Educational Leadership: Using Data-Driven Decision Making to Improve Schools*. Thousand Oaks, CA: Sage.
- Johnson, A. (2002). What Every Teacher Should Know about Action Research. Boston, MA: Allyn and Bacon.
- Johnson, A. (2007). *A Short Guide to Action Research*, 3rd edn. Boston, MA: Allyn and Bacon.
- Kemmis, S. (2008). Critical theory and participatory action research. In Reason, P. and Bradbury, H. (eds.) *Handbook of Action Research*, 3rd edn., pp 121–138. Thousand Oaks, CA: Sage.
- Kemmis, S. and McTaggart, R. (1988). *The Action Research Planner*. Geelong, VIC: Deakin University Press.
- Kemmis, S. and McTaggart, R. (2005). Participatory action research: Communicative action and the public sphere. In Denzin, N. and Lincoln, Y. (eds.) *Handbook of Qualitative Research*, 3rd edn., pp 559–604. Thousand Oaks, CA: Sage.
- Kincheloe, J. and McLaren, P. (2005). Rethinking critical theory and qualitative research. In Denzin, N. and Lincoln, Y. (eds.) *Handbook of Qualitative Research*, 3rd edn., pp 303–342. Thousand Oaks, CA: Sage.
- King, J. (2005). *Black Education: A Transformative Research and Action Agenda for the New Century*. Mahwah, NJ: Lawrence Erlbaum.
- King, J. and Lonquist, M. (1992). *A Review of Writing on Action Research*. Minneapolis, MN: Centre for Applied Research and Educational Improvement, University of Minnesota.
- Kirk, P. (2008). *African Leadership: Meanings, Connotations and Application*. PhD Thesis, University of West England.
- Kitchen, J. and Stevens, D. (2008). Two teacher-educators practice action research as they introduce action research to preservice teachers. *Action Research* 6(1), 7–28.
- Koshy, V. (2005). *Action Research for Improving Practice: A Practical Guide*. London: Paul Chapman.
- Kowalski, R., Yorks, L., and Jelinek, M. (2008). The workplace stress and aggression project: Ways of knowing – Our Rosetta Stone of practice. In Reason, P. and Bradbury, H. (eds.) *The Handbook of Action Research*, pp 497–509. Thousand Oaks, CA: Sage.
- Ladson-Billings, G. and Tate, W. (2006). *Education Research in the Public Interest: Social Justice, Action, and Policy*. New York: Teachers College Press.
- Levin, M. (2008). The praxis of education action researchers. In Reason, P. and Bradbury, H. (eds.) *Handbook of Action Research*, 3rd edn., pp 669–681. Thousand Oaks, CA: Sage.
- Levin, M. and Greenwood, D. (2008). The future of universities: Action research and the transformation of higher education. In Reason, P. and Bradbury, H. (eds.) *Handbook of Action Research*, 3rd edn., pp 211–251. Thousand Oaks, CA: Sage.
- Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues* 2(4), 34–46.
- Lewin, K. (1948). *Resolving Social Conflicts*. New York: Harper.
- Lewin, K. (1952). Group decision and social change. In Swanson, G., Newcomb, T., and Hartley, E. (eds.) *Readings in Social Psychology*. New York: Holt.
- Lewin, G. and Lewin, K. (1942). Democracy and the school. *Understanding the Child* 10, 7–11.

- McLean, J., Herman, J. L., and Herman, J. J. (2005). *Improving Education through Action Research: A Guide for Administrators and Teachers (Roadmaps to Success)*. Thousand Oaks, CA: Corwin Press.
- McNiff, J. and Whitehead, J. (2006). *Action Research for Teachers – A Practical Guide*. Abingdon, UK: David Fulton.
- McTaggart, R. (1991). *Action Research: A Short Modern History*. Geelong, VIC: Deakin University Press.
- McTaggart, R. (ed.) (1997). *Participatory Action Research: International Contexts and Consequences*. Albany, NY: SUNY Press.
- Mertler, C. (2005). *Action Research: Teachers as Researchers in the Classroom*. Thousand Oaks, CA: Sage.
- Meyers, E. and Rust, F. (2003). *Taking Action with Teacher Research*. Portsmouth, NH: Heinemann.
- Mills, G. (2007). *Action Research: A guide for the Teacher Researcher*, 3rd edn. Upper Saddle River, NJ: Pearson.
- Noffke, S. (1997). *Professional, Personal and Political Dimensions of Action Research*. *Review of Educational Research*, 22. Washington, DC: AERA.
- Noffke, S. and Somekh, B. (2009). *Handbook of Educational Action Research*. Thousand Oaks, CA: Sage.
- O'Hanlon, C. (2003). *Educational Inclusion as Action Research*. Maidenhead: Open University Press.
- Pedraza, P. and Rivera, M. (2005). *Latino Education: An Agenda for Community Action Research*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Phillips, D. K. and Carr, K. (2006). *Becoming a Teacher through Action Research: Process, Context, and Self-Study*. New York: Routledge.
- Rahman, M. A. (2008). Some trends in the praxis of participatory action research. In Reason, P. and Bradbury, H. (eds.) *Handbook of Action Research*, 3rd edn, pp 49–62. Thousand Oaks, CA: Sage.
- Reason, P. and Bradbury, H. (2008). *Handbook of Action Research*, 3rd edn. Thousand Oaks, CA: Sage.
- Rorty, R. (1981). *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press.
- Rorty, R. (1989). *Contingency, Irony and Solidarity*. Cambridge, UK: Cambridge University Press.
- Sagor, R. (2004). *The Action Research Guidebook: A Four-Step Process for Educators and School Teams*. Thousand Oaks, CA: Corwin.
- Sarason, S. (1990). *The Predictable Failure of Educational Reform: Can We Change Course before it's too Late?* San Francisco, CA: Jossey-Bass.
- Schon, D. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.
- Schon, D. (1987). *Educating the Reflective Practitioner*. San Francisco, CA: Jossey-Bass.
- Schon, D. (1991). *The Reflective Turn: Case Studies in and on Educational Practice*. New York: Teachers College Press.
- Somekh, B. (2005). *Action Research (Doing Qualitative Research in Educational Settings)*. Buckingham: Open University Press.
- Stenhouse, L. (1975). *An Introduction to Curriculum Research and Development*. London: Heinemann Educational.
- Stenhouse, L. (1983). *Authority, Education and Emancipation*. London: Heinemann Educational.
- Stringer, E. (2007). *Action Research*, 3rd edn. Thousand Oaks, CA: Sage.
- Stringer, E. (2008). *Action Research in Education*. Upper Saddle River, NJ: Pearson.
- Sung-Chan, P., Yuen-Tsang, A., Yadama, G., and Sze, A. (2008). Action research and social development in China. *Action Research* 6(2), 193–212.
- Sykes, J. (2002). *Action Research: A Practical Guide for Transforming Your School Library*. Portsmouth NH: Libraries Unlimited.
- Tomai, D. (2003). *Action Research for Educators*. Lanham, MD: Scarecrow Education.
- Wallace, M. (1998). *Action Research for Language Teachers*. Cambridge, UK: Cambridge University Press.
- Whitehead, J. (2009). Generating living theory and understanding in action research studies. *Action Research* 7(1), 85–99.
- Whitehead, J. and McNiff, J. (2006). *Action Research: Living Theory*. Thousand Oaks, CA: Sage.

Further Reading

- Christiansen, H., Goulet, L., Krentz, C., and Maeers, M. (1997). *Recreating Relationships: Collaboration and Educational Reform*. Albany: SUNY Press.

Relevant Websites

- <http://www.ael.org> – AELACTION: Action Research in Schools and Classrooms.
- <http://www.bath.ac.uk/carpp> – Centre for Action Research in Professional Practice (University of Bath).
- <http://www.coe.uga.edu> – College of Education – UGA: Action Research Overview.
- <http://www.enc.org> – Eisenhower National Clearinghouse: Science Education Action Research.
- <http://educ.queensu.ca> – Faculty of Education, Queen's University: Action Research at Queen's University.
- <http://ericps.ed.uiuc.edu> – Helping Early Childhood Teacher Education Students Learn about the Internet.
- <http://www.bath.ac.uk> – Jack Whitehead's Home Page (University of Bath).
- <http://www.links2go.com> – Links2Go: Action Research.
- <http://www.open.k12.or.us/arowhelp> – Oregon Public Education Network: Action Research Open Web Help (AROWHELP).
- <http://www.pekin.net> – Pekin Area Chamber of Commerce: School Improvement.
- <http://www.researchforaction.org> – Research for Action (RFA).
- <http://www.rochesterteachercenter.com> – Rochester Teacher Center: Action Research. Rochester Teacher Centre Teacher Research Program.
- <http://www.acsa.edu.au> – The Australian Curriculum Studies Association Inc: Action Research Network.
- <http://www.bath.ac.uk> – University of Bath: Action Research Module.
- <http://ucerc.edu> – University of California Educational Research Center: Teacher Research (Action Research Resources).
- <http://www.cudenver.edu> – University of Colorado Denver: Action Research Page.
- <http://www.extension.umn.edu> – University of Minnesota Extension: Action Research Resources.
- <http://www.fau.edu> – Welcome to Florida Atlantic University: School Renewal.
- <http://badger.state.wi.us> – Wisconsin.gov: What is Teacher Action Research?

Classroom Ethnography

F Erickson, University of California, Los Angeles, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

It is necessary to put some boundaries around the notion of classroom ethnography. As qualitative, narratively descriptive studies of teaching and of everyday life in classrooms developed in the late 1960s and early 1970s, some researchers referred to any such studies as ethnography. That is too broad a characterization – it blurs some important distinctions between ethnography and other forms of qualitative inquiry. On the other hand, a rigidly canonical definition of classroom ethnography could be too narrow – it might rule as out-of-bounds studies that were ethnographic in spirit but not in form. That would be easy to do, since the term ethnography, as applied to classroom research, is a metaphoric characterization rather than a literal one. Classroom ethnography does not correspond exactly to the classical methods and content of general ethnography. Yet classroom research that is ethnographic in intent bears certain family resemblances to general ethnography. We hope to make clear here what makes classroom ethnography ethnographic, and to present and discuss some key examples of classroom studies that have been undertaken along ethnographic lines. Another way to say this is that ethnography, as employed in classroom research, is not so much a set of techniques or methods as it is a perspective, a particular intellectual stance (see the discussion of Wolcott (2008: 67–89) on ethnography as a way of seeing). Accordingly, we need to consider the intellectual history out of which the perspective and stance of ethnography has developed.

Origins of General Ethnography

The word ethnography was invented in the late nineteenth century as a new term based on the combination of two Greek words: *graphein*, the verb for to write, and *ethnoi*, a plural noun for the nations – the others. Liddell and Scott's *Greek Lexicon* defines the singular noun *ethnos* as "A number of people accustomed to live together, a company, a body of men." But this definition slides by a distinction that makes a difference. For the ancient Greeks, *ethnoi* were not just groupings of people – in whatever scale of grouping one might want to consider – they were the groupings of people who were not Greek. *Ethnoi* was the contrast term for *'ellenoi*. Hellenes were referred to as we and *ethnoi* were referred to as they – Thracians, Scythians, Phrygians, Persians, Etruscans, Egyptians, and Mesopotamians. The Greeks

were more than a little xenophobic, so that *ethnoi* carries pejorative implications. To see the force of this, we can consider that in the Greek translation of the Hebrew Scriptures the Hebrew term for them – *goyim* – was translated as *ethnoi*, and in modern as in ancient Hebrew the term *goy* is not a compliment. All this is to say that the most accurate definition of the term ethnography, given its etymology and its initial use in the nineteenth century to refer to descriptive accounts of the lifeways of non-Western people, is "writing about other people."

Another term that became current in the late nineteenth century was ethnology. This means the comparative study of the meanings of differing patterns of organization and custom across differing human groups. Ethnology looks across separate ethnographic case studies of particular human groups for similarities and differences across the groups. Descriptive data from ethnography thus becomes the grist for the analytic mills of ethnology. Ethnography is always conducted with a comparative frame around it – the assumption is that what is being seen is not simply natural but arbitrary – that local lifeways are constructed. Each separate society and its lifeways are considered against the backdrop of comparison with all other human societies, known from contemporary research and from archeology, history, and prehistory.

In the ancient world Herodotus, the Greek scholar writing in the fifth century BCE, had interests that were ethnographic and ethnological as well as historical. Writing in the second century CE, the Greek skeptical philosopher Sextus Empiricus conducted a cross-cultural survey of morality, showing that what was considered right in one society was considered wrong in others. He worked from the accounts of travelers, and these continued to provide the primary basis for comparative knowledge about human lifeways until the late nineteenth century.

Ethnography claimed to be more thorough and comprehensive in its description than had been the reports previously written by travelers, soldiers, and colonial officers. Perhaps the first monograph of the kind that would become modern ethnography comes from urban sociology. DuBois (1899) conducted a study of a particular census tract in Philadelphia which was then the primary neighborhood of residence for African-Americans in the city. His report, titled *The Negro in Philadelphia*, combined demographic data, area maps, recent community history, and surveys of local institutions and community groups, with some descriptive accounts of the conduct of daily life in the neighborhood. His purpose in writing what can be

considered the first ethnography was to make visible the lives – and the orderliness in those lives – of people who had been heretofore invisible. A similar purpose and descriptive approach, combining demography and health statistics with narrative accounts, was taken in the reports of working-class life in East London that were prepared in the 1890s by Charles Booth, together with Beatrice and Sidney Webb (see Booth, 1891). Even more emphasis on narrative description was found in *How the Other Half Lives*, an account of the everyday life of immigrants on the lower East Side of New York City, written by the journalist Jacob Riis and illustrated with photographs (Riis, 1890). All these authors were social reformers. They were not simply producing description for its own sake – they were describing in order to advocate and inform social change.

Although their descriptions were not value neutral, these early practitioners of what can be called ethnography did not claim to be describing everyday life from the points of view of those who lived it. Their descriptions, in other words, were conducted from an etic point of view – based on a descriptive language of facts that were presented as self-evidently accurate and objective, behaviorally. They did not claim to be identifying behavioral differences that made a difference for subjective meaning among the people whose lives they were describing, that is, they did not claim that their descriptions had emic epistemological status. The interpretive significance of certain behaviors for everyday meaning, as Geertz (1973: 6) says what distinguishes an eye blink from a wink, was not what they were aiming at.

To portray social action (as wink) rather than behavior (as eye blink) – that is, to describe the conduct of everyday life in ways that make contact with the subjective orientations and meaning perspectives of those whose conduct is being reported – is the fundamental shift in interpretive (hermeneutical) stance within ethnography that was claimed to have been accomplished by Malinowski (1922) in his monograph *Argonauts of the Western Pacific*. He said that ethnographic description should not only be factually accurate, but that it should represent “the native’s point of view, his relation to life, to realize his vision of his world” (Malinowski 1922: 25). Malinowski, who was Polish, began fieldwork in the Trobriand Islands in 1914 and was then interned by the British colonial authorities because they suspected he might be a spy. Forced to remain in the Trobriands for the next 4 years, Malinowski later made a virtue of necessity and claimed that his long-term fieldwork and knowledge of the local language enabled him to write a report that not only encompassed the system of everyday life in its entirety but which accurately represented nuances of local meaning in its daily conduct. As people would say later, his descriptive reporting had emic validity, or interpretive validity. After Malinowski, this became a hallmark of ethnography in anthropology – reporting whose descriptions made contact with the meaning perspectives of those whose daily actions were being described. (This aim is never

fully realizable. The ethnographer can never completely discover or communicate the vision of the world as it is held by the people he or she studies – nor is there necessarily a single vision shared identically among those that are studied. However, approximating the meaning perspectives of those studied is a defining intention in ethnographic work.)

To conclude this brief overview of general ethnography, here is a description of ethnography by the contemporary anthropologist Conklin (1968: 172): “[ethnographic data] derive ultimately from the direct observation of customary behavior in particular societies. Making, reporting, and evaluating such observations are the tasks of ethnography . . . [An ethnography] requires a long period of intimate study and residence in a small, well-defined community, knowledge of the spoken language, and the employment of a wide range of observational techniques including prolonged face-to-face contacts with members of the local group, direct participation in some of that group’s activities, and a greater emphasis on intensive work with informants than on the use of documentary or survey data.” More abstractly Conklin said in another place, “an adequate ethnography is here considered to include the culturally significant arrangement of productive statements about the relevant relationships obtaining among locally defined categories and contexts (of objects and events) within a given social matrix. These . . . should comprise, essentially, a cultural grammar.” (Conklin, 1964: 25). An even more modern definition of ethnography would be that it considers the setting that is studied as a local community of practice and identifies the full range of variation in practices that occur there, within the full range of culturally significant situations or activity types, considering the meaning perspectives that underlie the conduct of the cultural practices that take place (see also the discussion in Anderson-Levitt, 2006).

Classroom Ethnography

One of the chief differences between classroom ethnography and classic general ethnography is that classroom ethnography is usually done by researchers who have themselves spent time there as children. This is what can be called domestic ethnography in contrast to exotic ethnography. Classic general ethnography was done by researchers who went to places with which they were unfamiliar – either to village groupings or tribal groupings of so-called primitive people in the colonies ruled by Western empires or to slums inhabited by poor people in large European or American cities. The aim was to make visible the lifeways of people who were literally off the map and to communicate this portrayal to audiences in polite Western society generally and in academia more specifically.

The core task of exotic ethnography, then, is to make the strange and unknown familiar and intelligible. This is

very different from a study of a school classroom that is similar to those previously attended by the ethnographic researcher and also by the researcher's readers. In such domestic ethnography, the core task for the researcher as a fieldworker – and ultimately for the researcher's audience – is to make the familiar strange, and in so doing to make habitual actions visible. The reason to do ethnographic case studies of familiar places like classrooms is that they are too familiar for us to understand at first glance – habit makes daily practice partially and sometimes wholly invisible to its practitioners – thus we need to strangify the familiar in order to see it. Marx said that for scholarly inquiry “the problem is not to understand the world but to change it.” Yet we must be able to see daily practices before we can decide whether or not to change them. The critical inquiry that is inherent in domestic ethnography means that we study everyday practices with possible change in mind, yet we withhold judgment on changing things until we have made a detailed and comprehensive description of the everyday practices themselves. (this point is discussed in the Conclusion section).

It was said at the outset that the connections were not direct between classic general ethnography – portraying the whole way of life of a naturally occurring local grouping of people, with research typically being done by an outsider. Classroom ethnography portrays a topic-focused account of some aspects of a particular institutional setting in which none of the participants live their entire lives, with research typically being done by someone who, if not exactly an insider, still knows much about the setting, including knowledge that comes from prior life experience in similar settings. Having reviewed principal characteristics of general ethnography it is now appropriate to summarize key aspects of classroom research that might be ethnographic, bearing family resemblance to general ethnography if not one-to-one correspondence with it.

We want to present a deliberately generous definition – one that would include many studies although it would exclude some. A classroom study that was ethnographic in approach would have the following characteristics:

1. Long-term, close observation of and participation by the researcher in routinely recurring daily activities.
2. Consideration of the setting, both during fieldwork and in reporting, as a local ecosystem of relations of simultaneous and mutual influence among differing aspects or components.
3. Identification of the total cast of characters in the setting, the variation that obtains locally in ways of enacting their roles, and the structured relationships of power and authority in relation to one another (a portrayal of the overall social organization, formal and informal).
4. Identification of the full range of activities and social situations that take place in the setting (with special emphasis on the spatial and temporal location and

organization of each activity as well as on the culturally significant practices, verbal and nonverbal, of the various participants in those activities).

5. Identification of the meaning perspectives that are entailed in the conduct of the everyday practices that occur.

(Implicit and explicit beliefs, values, and identities, including local ontologies and epistemologies. These are systems of belief about what is real in the world that, in a classroom, take the place of a folk religious world-view – what truly exists, what is right and wrong, and how we can know that, what is knowledge, what are learners, what is learning, what is evidence of learning, and basic postulates such as “order must be established before learning can take place.”)

Unlike a general ethnography, a classroom ethnography would not include description of the practices by which subsistence is maintained and economic relations are conducted, just as there would be no literal description of a legal system or of folk medicine. Yet an ethnographic classroom study might document the exchange of goods in a symbolic economy (student effort as exchanged for classroom rewards such as grades and teacher regard), local notions of fairness and due process, and local notions of and standards for physical, emotional, and intellectual health and growth. (For an earlier discussion along these lines, see Erickson, 1973/1984.)

Given the emphasis on holism in general ethnography, an ethnographic classroom study might most appropriately be done in early grades classrooms, where teacher and students participate together across an entire day, in a variety of activity settings, confronting various subject matters – conditions that are more analogous to the intimacy of acquaintance and commonality of horizon in daily life that are found in a naturally occurring small community. In middle school, high school, and college classrooms, where only one subject is taught and the students meet as a class for a single instructional period and then move on to other classrooms, the analogy with daily life in a small community such as a village is more loose, and it may be less appropriate to consider such classrooms as settings for study that is ethnographic in spirit. Still there could be a family resemblance with ethnography in such settings.

In our judgment, approaches to classroom study that would not be ethnographic would include the following:

1. Studies of the formal subject matter being taught that do not include the hidden curriculum that accompanies subject matter instruction.

(When math or reading or any other particular subject matter is being taught, so are assumptions about the nature of knowledge and of knowers, gender, racial, ethnic, and language identity in relation to learning subject matter, power relationships with the teacher, relative privilege differences among students.)

2. Studies of relations between the teacher(s) and students that consider those social identity categories as unitary and fixed, rather than as multidimensional and in dynamic ecological relationship.

(Usually there is more than one kind of student in a class – various student characters – and more than one kind of way of enacting being the teacher. Moreover, power relations are reflexively related – teachers do not simply control students through classroom management – students push back, and upon occasion it is students who manage the teacher rather than vice versa.)

3. Studies conducted short-term – a 2-week – observational study, let alone a single class period's or a single school day's observation, would not be ethnographic.

(One might prepare a narrative descriptive account from such drive-by ethnography but genuinely ethnographic work presumes that local custom – local cultural practices – are nuanced and locally distinctive and that such subtleties in the organization of the conduct of everyday life in a local setting require repeated learning attempts by the fieldworker – repeated visits across substantial strips of time – in order to answer such questions as “What is the actual full range of different ways of being a student in this particular classroom?”

This is all the more necessary when the observer has surface familiarity with the classroom and thus needs time to make the familiar strange and visible.)

Let us consider a few illustrative examples of classroom ethnography, early and current. First, two case studies conducted in the early 1950s by George Spindler and published later. The first case study is of a fifth-grade classroom whose teacher was given the pseudonym Roger Harker (Spindler and Spindler, 1982). Spindler's description highlighted differences between formal and informal social organization in the classroom. He found that students who were doing well academically were all seated on one side of the room, while students who were not doing so well were all seated on the other side – and that was the side where the students of minority and/or of low family income sat. The second case study is of a girl from that classroom, called Beth Anne, who was chosen because of her reputation in the school as a well-adjusted child (see Spindler and Spindler, 1990). As Spindler observed Beth Anne across a wide range of activity settings in the classroom and on the playground, it became apparent that she was subtly anxious. She was a more complex, multiple-faceted person than she had been typecast as, on the basis of surface appearances and on the basis of seeing her in only a limited range of activity settings.

In a volume by Jules Henry titled *Culture against Man* containing chapters on elementary classrooms, the theme of student anxiety over achievement and over relationships among peers was continued and extended (Henry, 1963). Henry's portrayal of the classroom shows how competition among students and carping criticism of one

another's' performances were ubiquitous, as was the expression of affection by the teacher as a means of intensifying the students' anxiety over achievement. As in the Spindler's portrayals, the emphasis is on contrast between the official, formal social organization of the classroom and its unofficial informal organization, and the hidden curriculum of social relations and emotionality around learning is foregrounded, as well as consideration of the manifest curriculum of subject matter instruction. Another pioneering work was Philip Jackson's book length report *Life in Classrooms* (1968). In it he considered classroom practices and social organization from the perspective of ethology, the observational study of naturally occurring behavior among animals. Among the points Jackson emphasized was that as a child enters a school classroom, one of the things the child has to do (in contrast from family life at home) is learn how to be a member of a crowd. Also the child has to learn how to wait in the crowd. The development of boredom management techniques follows from this. Once again, informal social organization and hidden curriculum were being emphasized.

Perhaps the most comprehensive of the early attempts at ethnographic study of classrooms was the book length account of an entire year in an inner city seventh-grade self-contained classroom, co-authored by Louis M. Smith, the ethnographer, and William Geoffrey (pseudonym), the classroom teacher. Published in 1968 and titled *The Complexities of an Urban Classroom* (Smith and Geoffrey, 1968), this account emphasized the development of special roles by various children, in a distinctly local classroom social order that evolved over the first days and weeks of school – thus an emphasis on informal social organization. The book also described the teacher's approaches to personalizing instruction – relating differently to different students depending upon their distinct strengths and needs – while still working within the framework of traditional curriculum and use of standard textbooks.

Teacher-authored accounts of classroom life have often been done from a holistic ethnographic perspective. An early example of this is Kohl's (1968) *Thirty-Six Children*, and a more recent example is Cynthia Ballenger's account of literacy instruction in a bilingual classroom (Ballenger, 1999).

Frank (1999) has published a guide to teachers' research in their own classrooms titled *Ethnographic Eyes*. Hammersley (1990) is a good general resource on classroom ethnography, illustrating ethnographic studies of classrooms and discussing methods for accomplishing them. The discussion in Castanheira *et al.* (2001) is also useful.

Conclusion

In recent years, there has been a burgeoning of classroom studies, especially concerning instruction in literacy,

science, and mathematics, that describe themselves as ethnographic in orientation – far too many to review here. Some of these studies are ethnographic only in the sense that they are narratively descriptive, or in that they present a transcript of classroom discourse. Some of these incompletely ethnographic studies focus entirely on subject matter instruction – the manifest curriculum – and ignore the hidden curriculum. Many take a broader view that is indeed ethnographic; a holistic view of everyday scenes of classroom life as learning environments, socioculturally organized, with both official and unofficial aspects, with both manifest and hidden curriculum. These are studies that view the classroom as at once locally constructed and as connected with and influenced by wider spheres of social control and its contestation in society more generally. Routine classroom practices are portrayed as socially and culturally constructed, considering the particular local construction at hand as a specific variation within a world-wide range of possibilities. That ethnological backdrop for ethnographic case studies of classroom life continually suggests possibilities for change – it implies that what has been constructed locally in a particular way holds continuing possibility for reconstruction.

Yet the ethnographic emphasis on the specifics of local social and cultural ecology suggests that single-factor approaches to change – quick fixes – will not be successful. Rather, multiple components of the classroom ecosystem must be changed, in concert with one another. For example, let us consider a teacher who wanted to change mathematics instruction from an emphasis on procedures for finding the right answer to an emphasis on deep conceptual understanding of basic mathematical ideas. If students were to be able to discuss alternative ideas about mathematics, practices in the conduct of discussion would need to change, in order to alter the system of carping criticism, mockery, and overall competition among students as described by Jules Henry. As a student it is more face threatening to have to say what you think and defend it than it is to call out a right answer. The teacher, as manager of social relations in the classroom, would need somehow to make it safe for students to disagree with one another.

Teacher beliefs about learners would also need to change – students would need to be seen by the teacher as capable of understanding key concepts thoroughly and of articulating those understandings in talking and writing. Changes in other subject matter instructional practices might also follow. If right answers are no longer the primary focus in mathematics instruction, what about allowing invented spelling and punctuation in writing – emphasizing the primacy of the sense of what is being written over the canonical form of the writing? What about basic postulates? The teacher's assumption that order must be established

before learning can take place, as mentioned earlier, might well be replaced by an assumption that order usually follows student interest and understanding. Finally, what about assessment? If high stakes testing in an accountability system external to the classroom emphasizes student knowledge of facts and simple skills as evidence of learning, the teacher's attempt to teach mathematics for understanding might become difficult to implement – not impossible, but it would require the teacher to swim upstream against strong cultural currents.

Changes in the social organization of a classroom learning environment, the ethnographer would assume, would necessarily accompany changes in the subject matter content of instruction. In other words, the social-ecological perspective of classroom ethnography suggests that when change happens in one aspect of classroom practices, changes in a whole system of classroom practices would likely follow.

In sum, the ethnographic perspective on classrooms as learning environments assumes that what humans have made they can change, even though that may well involve changes in multiple aspects of practices. It further assumes that if enduring change is to happen in classrooms, local autonomy is necessary to be exercised on a daily basis by the teachers and students who live there. This is an inherently critical stance for classroom research, yet it is criticism with empathy for and understanding of members' points of view, and with respect for their capacities for sense-making even when the researcher may not fully agree with the sense that is being made. It is criticism presented by a researcher who was both stranger and friend to those among whom he or she became closely acquainted with their locally distinctive way of life.

See also: Critical Ethnography; Ethnography; Hermeneutics.

Bibliography

- Anderson-Levitt, K. (2006). Ethnography. In Green, J., Camilli, G., and Elmore, P. (eds.) *Handbook of Complementary Methods in Education Research*, pp 279–295. Washington, DC/Mahwah, NJ: American Educational Research Association/Erlbaum.
- Ballenger, C. (1999). *Teaching Other People's Children: Literacy and Learning in a Bilingual Classroom*. New York: Teachers College Press.
- Booth, C. (1891). *Labour and Life of the People of London*. London: Williams and Norgate.
- Castanheira, M. L., Crawford, T., Dixon, C., and Green, J. (2001). Interactional ethnography: An approach to studying the social construction of literacy practices. *Linguistics and Education* 11(4), 353–400.
- Conklin, H. (1964). Ethnogeneological method. In Goodenough, W. (ed.) *Explorations in Cultural Anthropology: Essays in Honor of George Peter Murdock*, pp 25–55. New York: McGraw-Hill.

- Conklin, H. (1968). Ethnography. In Sills, D. (ed.) *International Encyclopedia of the Social Sciences* 5, pp 172–178. New York: MacMillan.
- DuBois, W. (1899). *The Philadelphia Negro: A Social Study*. New York: Schocken.
- Erickson, F. (1973/1984). What makes school ethnography 'ethnographic'? *Anthropology and Education Newsletter* 4(2), 110–114. [Reprinted in 1984 *Anthropology and Education Quarterly* 15(1), 51–66].
- Frank, C. (1999). *Ethnographic Eyes: A Teacher's Guide to Classroom Observation*. Portsmouth, NH: Heineman.
- Geertz, C. (1973). *The Interpretation of Cultures*. New York: Basic Books.
- Hammersley, M. (1990). *Classroom Ethnography: Empirical and Methodological Essays*. Toronto: OISE Press.
- Kohl, H. (1968). *36 Children*. New York: New American Library.
- Malinowski, B. (1922). *Argonauts of the Western Pacific*. New York: E.P. Dutton.
- Riis, J. (1890). *How the Other Half Lives: Studies among the Tenements of New York*. New York: Charles Scribner's Sons.
- Smith, L. and Geoffrey, W. (1968). *The Complexities of an Urban Classroom: An Analysis toward a General Theory of Teaching*. New York: Holt, Rinehart, and Winston.
- Spindler, G. and Spindler, L. (1982). From familiar to strange and back again: Roger Harker and Schoenhausen. In Spindler, G. (ed.) *Doing the Ethnography of Schooling*, pp 20–46. New York: Holt, Rinehart, and Winston.
- Spindler, G. and Spindler, L. (eds.) (1990). Schooling in the American cultural dialogue. In *Schooling and American Culture*, pp 56–69. London: Falmer.
- Wolcott, H. (2008). *Ethnography: A Way of Seeing*. Lanham, MD: Altamira Press.

Further Reading

- Girod, M. (2008). Deepening understanding of the teaching and learning context through ethnographic analysis. *Teacher Educator* 43(3), 216–237.
- Henning, E. (1993). Ethnography as classroom research methodology. *South African Journal of Education* 13(3), 112–118.
- Trueba, H., Guthrie, G., and Au, K. (1981). *Culture and the Bilingual Classroom: Studies in Classroom Ethnography*. Rowley, MA: Newbury House.
- Van Lier, L. (1988). *The Classroom and the Language Learner: Ethnography and Second-Language Classroom Research*. London: Longman.
- Watson-Gegeo, K. (1997). Classroom ethnography. In Hornberger, N. and Corson, D. (eds.) *Encyclopedia of Language and Education, Vol. 8, Research Methods in Language and Education*, pp 135–144. Dordrecht: Kluwer.

Computer Assisted Qualitative Data Analysis

C Silver and A Lewins, University of Surrey, Guildford, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Autocode – To use software to search for words, strings, phrases, and code the results, possibly together with surrounding context.

CAQDAS – This stands for Computer-Assisted Qualitative Data Analysis Software.

Hyperlink – A direct link between two pieces of text, or out of the software to an external file or website.

Introduction

In recent years, the process of conducting qualitative research has become increasingly technologically driven. From the digitized collection and storage of data, through their transcription, analysis, and the presentation of results, technology is never far away. Computer-Assisted Qualitative Data Analysis Software (CAQDAS) packages are used across a wide range of academic and applied settings, in disciplines including education, psychology and sociology, business and management, government research, health studies, geography, and art history. Some packages and aspects of their functionality are referred to below, but only to illustrate potential differences or the individuality of some important tools within some packages. Details of functionality are transient because of the speed of technological advances, so the article deals with broad issues and does not attempt to be a comprehensive review of packages mentioned.

Technological developments generally occur more quickly than methodological ones, and in the field of CAQDAS, it is pertinent to ask to what extent technology is driving methodology. The computer does not analyze data for the researcher, and therefore the interpretation of results rests with the researcher. There are however certain tasks enabled by CAQDAS packages that would be difficult to replicate by more traditional craft methods and there are new types and combinations of data that can be incorporated because of technology.

The term CAQDAS was coined by Ray Lee and Nigel Fielding at the time of the first Surrey Research Methods conference in 1989, which brought together pioneers in the field. Their edited volume of papers by methodologists and software developers provides an historical context and is therefore a seminal work in the development of this technology and its adoption by researchers

(Fielding and Lee 1991/1993). An important typology was offered by Weitzman and Miles (1995), which categorized the then available software as text retrievers, code-and-retrieve packages, and theory-building software, but these categories lost relevance as each software offers new functions cutting across the typologies. Technological advances have also led to methodological innovations, many of which have been applied beyond traditional disciplinary boundaries. The applications of software tools have broadened. There are now many packages available which support various aspects of, and approaches to qualitative analysis. Alongside the market leaders of commercial software are a range of low-cost or freely available alternatives, and some bespoke tools providing a smaller range of tools for specific analytic tasks or data types. In particular, the analysis of visual data is becoming better supported, with some tools focusing specifically on multimedia analysis. Others which hitherto focused on textual analysis now incorporate more flexible ways of handling visual data.

It is increasingly difficult to apply a clear typology or to define what we mean by the CAQDAS label as some software that offer such qualitative functions also encompass tools that support a more quantitative approach to qualitative data, including basic content analysis methodology in which the statistics of word frequencies are given. Some CAQDAS packages, such as QDA Miner offer two quite different philosophical approaches to qualitative data with more sophisticated content (text-mining) analysis tools alongside good qualitative tools. We might then at least encourage a reconceptualization of the term CAQDAS to refer to any software tool, which includes specific support for the qualitative analysis of qualitative data. Such a tool would include first, thematic coding tools to organize and interpret concepts, processes, interactions, etc., in order to interrogate and identify meaningful patterns and relationships. Other necessary tools would be text-level tools with rich annotation, memo, and text-linking devices.

The aim here is to provide an up-to-date summary of the broad task-based areas of functionality in CAQDAS packages reflecting recent technological and methodological trends.

Qualitative Tasks Supported by CAQDAS Packages

In contextualizing software tools we refer to **Figure 1**. This points to the structure of the article and graphically

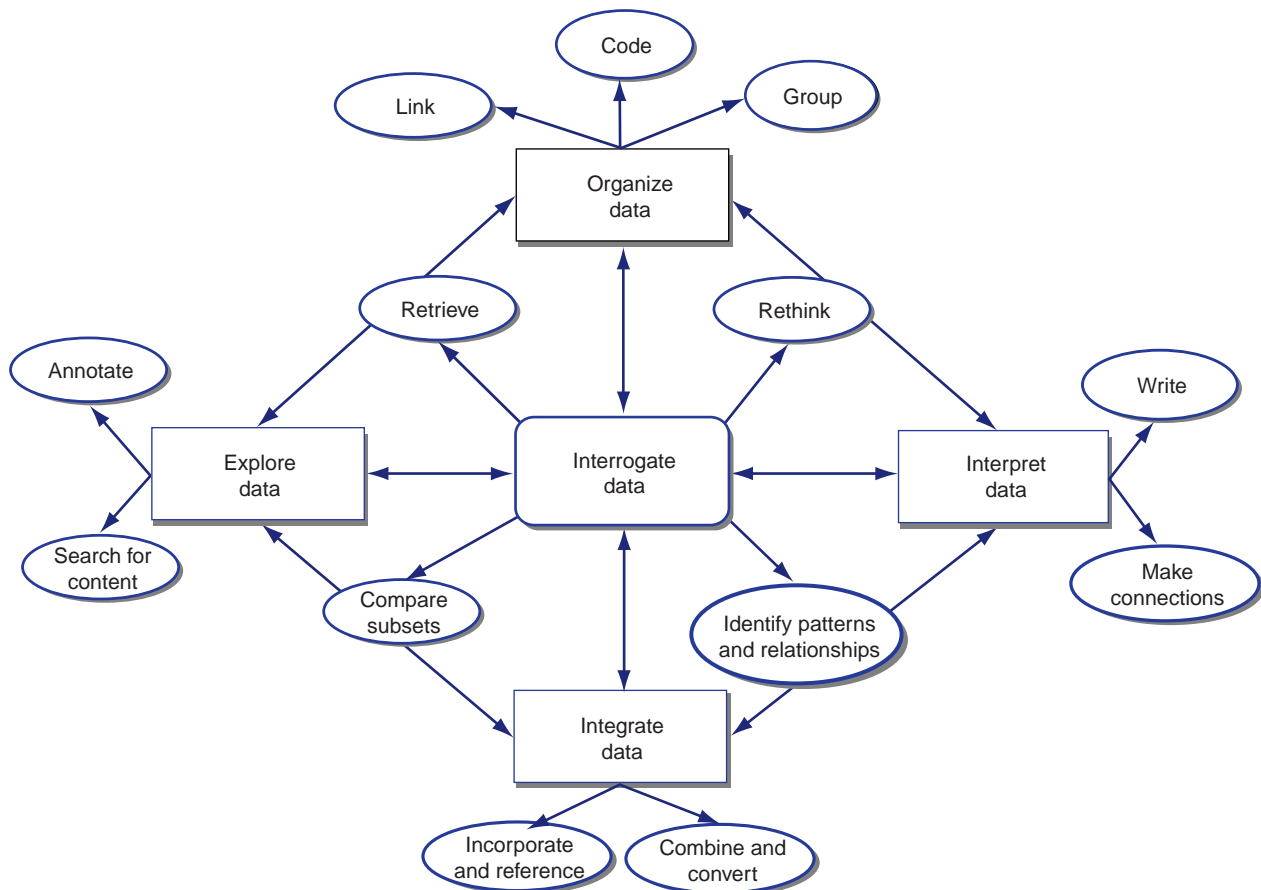


Figure 1 Qualitative tasks supported by CAQDAS packages. Adapted from Lewins, A. and Silver, C. (2007). *Using Software in Qualitative Research: A Step-by-Step Guide*. London: Sage.

illustrates linkages between the main qualitative tasks of integrating, organizing, exploring, interpreting, and interrogating data. Linked to the main tasks are indications of the generic tools in CAQDAS packages that support those main areas of work. **Table 1** summarizes some of these tools in more detail in the context of the practical aspects of project work as enabled by the software. The researcher can pick and choose which tools support a chosen analytic path in order to be creatively and efficiently in control of the software.

While these packages are generically referred to as qualitative data analysis tools, they are perhaps more accurately described as qualitative data management tools. Like most computer programs, the user remains in control of the executed tasks and it is important to stress that the tool is not undertaking the analysis, but simply storing the data and enabling the researcher to view, interrogate, and retrieve them in different ways.

Project Planning and Working in a Team

Planning which software to use and how and when to familiarize with it should be part of the wider research

project planning. This is especially true when a team of researchers are working together on a project.

Projects in software cannot usually be worked on concurrently by different users, and therefore, a strategy for dividing work needs to be developed (concurrent use in one software project, especially packages designed to handle textual data is not generally possible at the time of publication). Separate software projects can often be merged at various stages but the procedures for doing so can be quite involved. Only one or two software (and those only for multimedia such as video) now provide multiuser versions that do allow several researchers to work simultaneously on the same project. In these cases, users can see what each other is doing in real time and can communicate with one another using basic text messaging tools.

Another aspect of team working may be the negotiation and development of a shared coding schema, which can be a considerable task. Several CAQDAS packages now include integrated coder reliability tools that facilitate this process, providing a statistical measure of agreement between codings.

Table 1 Practical aspects of project work enabled by CAQDAS packages

<i>Practical aspects</i>	<i>Analytic rationale</i>
Planning and managing your project	Store and keep together the different aspects of your work in one place, that is, individual data files and your thinking about them. Aid continuity. Construct an audit trail as a natural part of the process of coding, and particularly by using memos.
Writing analytic memos	Manage your developing interpretations by keeping track of ideas as they occur, and building on them as you progress . . .
Writing memos about process	Reinforce record keeping about your thinking and improve continuity by logging tasks undertaken in a disciplined way and putting reminders about follow-up work for next work session. Log into this memo at the beginning and end of each session.
Read, mark, annotate, or comment on data	Discover and mark interesting aspects in the data as you see them. Tools variably enabled for different types of data and in different software.
Searching (for strings, words, phrases, etc.)	Explore data according to its content, observing how context may differ where words or phrases are found. Exploring prevalence and usage of jargon or respondents' particular ways of describing things.
Developing a coding scheme	Manage your ideas about your data, in themes, concepts, etc. Structure and function may depend on methodology and style of working. Create codes up front or as you see them emerge from what seems significant in the data, or in both ways.
Coding	Capture what is going on in your data in a consistent way. Bring together similar data according to themes, concepts, etc. Assign codes to chunks of data as required, include a little surrounding context in the data to enable better understanding when in retrieval mode.
Retrieval of coded segments	Revisit coded data chunks or segments to assess similarity and difference, influences, undercurrents, and context. Read and reread to consider how coding is helping your understanding and analysis, and whether it is helping you in where to go next in developing analysis.
Recoding	Recode into broader or narrower themes or categories if appropriate and necessary. Perhaps bring coded data back together in new groupings and think about them differently.
Organization of data	Organize data according to known facts and descriptive features to enquire into particular subsets of respondents, or compare across them. These aspects may be a central part of your sample and will contribute to your analysis.
Hyperlinking	Link data to other text segments and/or to other files to track process, contradiction, etc. Variable tools in this respect provided in software packages.
Searching/querying the database and the coding schema	Test ideas, proximity/co-occurrence of codes, interrogate subsets for similarity and difference, or generate another level of coding.
Mapping	Manage analytic processes by visualizing connections in a graphic way to display or remind of relationships, patterns, processes, ideas observed in the data. A way to step back from the data.
Generating output	Report on different aspects of your progress and the project at any stage. Save as files to capture status at an analytic stage or to work in other applications. Print off to get away from the computer and think and work in more traditional ways. Types of output reduce with multimedia.

Adapted from Lewins, A. and Silver, C. (2007). *Using Software in Qualitative Research: A Step-by-Step Guide*. London: Sage.

Integrating and Incorporating Data

A key benefit of using a customized software to manage qualitative research projects is the potential it provides to integrate data. Methodologically one can identify two ways in which data integration may take place when using qualitative software:

- (1) simple incorporation of diverse data types alongside one another within a software project to inform different aspects of the project, likely analyzed separately or sequentially; and
- (2) integrating the analysis of data types, either by combining qualitative and quantitative data within an

analysis or by converting qualitative data to quantitative variables for use in a statistical analysis.

Incorporating and Referencing Data

Many qualitative projects use several different forms of data which can be incorporated within or referenced from the chosen package in order to be handled together or separately as required. The software project either contains all the data you wish to work with, or links to the files which remain outside the software can be created (see Lewins and Silver (2007) for a discussion of the implications of each system).

Packages often allow a range of data formats to be directly handled often including a variety of multimedia, for example, graphic, audio and video formats, as well as textual formats, for example, .rtf, .doc (Check software developer websites for up-to-date lists of supported data formats.) Some packages also handle .pdf files, others now provide the ability to connect geographical references from Google Earth with (parts of) texts or codes.

Material relevant to the research project which either cannot be directly incorporated into the CAQDAS package because it is in the wrong format or which the researcher wishes to keep elsewhere, can be referenced from within the software project. These may include hard copy books or articles, handwritten diaries, television programs, films, etc. Material which remains external to the software project may be summarized in note form internally or incorporated in lists, and subsequently this information treated like any other data record. It can be annotated or cross-referenced by coding. Keeping records of routine reading of literature or writing a literature review can be a simple application of some of the tools in most CAQDAS packages.

Whatever data are incorporated or referenced in the qualitative software project, once work begins, access to them or the notes you have made about them become almost instantaneous, thereby enabling seamless flicking between relevant materials which may otherwise only be accessed through different programs. Such basic organizational tools facilitate continuity and keep the researcher in reliable and continuous contact with all manner of information.

Multimedia Data

Enhanced recording and storage facilities enable researchers to generate volumes of digitized data in audio, video, and graphic formats, allowing the capture of large quantities of high-quality images. The social world itself is both more technologically and visually driven, with multimedia data being an increasingly normal part of everyday life.

In response, software developers focus on effective and customized means of handling multimedia data for the purposes of qualitative analysis. Some of these are very low cost or even free. Transana (open source, low cost) was one of the first bespoke packages to provide specific tools for the analysis of audio-visual data, initially providing a sophisticated transcription tool, subsequently developing flexible annotation, coding, and play back tools. The Digital Replay System (DRS – open source and free) similarly provides transcription and coding support for multimedia, while also enabling other related or synchronous forms of digital data to be viewed, coded, and interacted with alongside. Teamwork is well supported by some of the low-cost multimedia packages Transana and MiMeg (open source and free) provide, a means by which teams can tap into a data set from different places at the same time.

Increasingly commercial packages such as ATLAS.ti, HyperRESEARCH, and now NVivo have sought to support work with multimedia alongside the textual data they were initially created for.

The utility of given software packages in multimedia analysis depends largely on two closely related factors; the status of that data within the research project and the methodological approach to its analysis. Audio visual data provides access to very different accounts than textual data, offering the researcher multifaceted perspectives of the realities of social phenomena. Projects in which such data is supplementary or provides background context to textual data have very different requirements than those in which the audio visual data is the primary source. For example, a key issue in analyzing audio visual data relates to the need for development and use of a written transcript or log. Most software now allow the audio or video file to be marked, annotated, and coded directly, without the need for a written log of any sort. This can be particularly useful when analyzing nonverbal interactions and body language, for example, between teachers and students, or doctors and patients. However, the development of a written log (which may vary from a fully verbatim transcription, to a summary overview, or an analytic memo), may be an important analytic task in its own right. It will certainly provide an additional dimension to the data. While audio visual data is incredibly rich, it can only be viewed and heard in real-time, and it is therefore difficult to consider as a whole without some written summary. Abstracting from the data by working with a written log as well as directly with the multimedia file can provide the researcher with an additional perspective and means of analytic consideration. Playback options and tools that allow shuffling through help to deal with this, but nevertheless, a written transcription of some sort usually acts as a valuable data source. Many software that handle multimedia data now enable full synchronization between a written transcript and the corresponding audio visual file.

Another recent development is the ability to integrate geo-referencing tools with CAQDAS software. The MAXqda and ATLAS.ti have been early pioneers in this integration. These packages allow geo-reference points accessed through Google Earth to be viewed and saved as data within the software. It is also possible to create hyperlinks from points in text files within the software, to geo-reference points in Google Earth for quick movement between applications. These developments provide new ways to handle the spatial elements of social research.

Integration of Qualitative and Quantitative Analyses

Mixed methods projects are commonly undertaken and can be increasingly facilitated by the use of CAQDAS packages. More sophisticated methodological integration

requires the transformation of qualitative data into variables. Where codes derived from qualitative data are recorded as the presence/absence of the code in each case or as a frequency of the code's occurrence, a case-by-variable matrix can be generated. While still within the CAQDAS software there might be interactivity between each cell in the matrix to the underlying qualitative data. On export of the quantified data to the quantitative package, contact with the qualitative data is lost but statistical techniques such as cluster analysis, correspondence analysis, and multidimensional scaling can be applied. A number of hybrid techniques to interrelate quantitative and qualitative methods have been demonstrated in single studies but are not widely used (Bazeley, 2006).

Many CAQDAS packages have long provided the facility to import quantitative information in the form of spreadsheets to link to qualitative records. An example would be to conduct a qualitative analysis of open-ended questions to survey data. The formatting of both the spreadsheets and the qualitative records needs to be considered for these processes to occur seamlessly and protocols vary according to the CAQDAS package being used. Depending on the nature of the qualitative records, it may be necessary to take the first step of autocoding repeated structures within or across data files in order to link up the quantitative information correctly. These repeated structures may represent speakers in a focus group or responses in open-ended survey data. Such facilities are available for instance in NVivo and MAXqda. But several leading CAQDAS packages allow the import of quantitative data to categorize whole files such as interviews.

Exploring Data

Searching the content of textual files for keywords, phrases, or character strings may be a simple explorative task, form part of a closer analytical reading, or be an essential component of a quantitative content analysis. Becoming aware of what is interesting and significant in the data is a process which begins during the first moments of data collection and while transcribing, as well as during the more formally analytic phases of annotating, coding, retrieving, etc. Exploring the content of data files and one's own notes in the form of memos or annotations can be iterative and consideration of the content as well as the meanings of texts can help reflect on them from different perspectives (see Lewins and Silver, 2007).

Word Frequency Tools

Many packages now include basic data mining tools such as word frequency counts which index textual strings in tabular format, providing counts by file and across (parts of) the data set. With roots in the quantitative

content analysis tradition, such tools may be criticized by qualitative researchers who view them as crude and irrelevant in uncovering meaning. Quantitative summaries of qualitative data should rightly be treated with caution, so as not to extrapolate beyond their relevance. For example, one limitation of these tools is their failure, in the tabular view, to discriminate between the origin of words (e.g., who used certain words in an interview file). Nevertheless, they offer alternative ways of viewing and accessing textual data which are extremely useful in many projects, in particular when analyzing documentary evidence or using software to facilitate a literature review.

Word frequency tools differ between packages. There is a difference between packages that provide word frequency tools which give interactive movement between the words and the text they are contained in (or Key Word in Context/KWIC) and those that do not. Again, the individual user should be aware of the varying levels of statistical information and how they are presented, in case it matters; see earlier section on Integration of Qualitative and Quantitative Data.

Text Search Tools

Building on word frequency tools, text search tools allow the reader to specify which strings, words, or phrases to search for, and be taken directly to those positions within the relevant textual files. Such tools not only provide access to keywords within the source file but usually allow for a user-specified context around each hit to be automatically coded (e.g., the sentence, paragraph, or other unit). Most CAQDAS packages enable the use of wildcards such as the asterisk character to find alternative endings or prefixes. Boolean operators and other functions refine instances for retrieval. Caution again needs to be applied as these tools will only locate when the searched-for words appear, and not where the text refers to the topic without using that word or phrase. Gaps in the data are often as analytically meaningful as those which contain certain words, yet projects with an extensive data corpus will find the ability to quickly and automatically code data according to the presence of key words a useful way of initial exploration.

Annotation Tools

Alternatively, close analytic reading of written texts may be the main focus of the analysis; as in, for example, forms of discourse, narrative, and conversational analysis phenomenology and the related interpretive phenomenological analysis (IPA). Historically, such analysts have been somewhat reluctant to use CAQDAS packages, often due to a perception of their reliance on coding tasks (see the following discussion in relation to linking tools). Recent technological developments however, mean that such methodological approaches are increasingly well-served

by CAQDAS packages (see Silver and Fielding (2008) for more discussion and illustration of this). Additionally, where audio visual data is directly analyzed, annotation tools can provide a key way of marking data and unpacking the meanings of identified processes or interactions.

Annotations are directly linked to the data segments which they are about. In textual documents, they act similarly to footnote functions in word processors and can usually be outputted in this way at the level of the whole data source, or when outputting data coded in certain respects. Depending on methodology, researchers may rely more or less heavily on these tools, some wishing to use them to the exclusion of coding tools. They add to the ways data can be effectively managed in general – ways implementing stable connections between data and the researcher's thoughts about them.

Organizing Data

Data organization is a key aspect of handling and analyzing qualitative data. The CAQDAS packages provide three main ways to organize data. Grouping data according to their known characteristics or structurally coding parts of files according to who is speaking are closely connected. More conceptually, coding is also key to organizing (and interpreting) data based on pertinent themes, etc. Related to conceptual work is the linking of data segments to track processes or interactions. The manner and extent of organizing data will largely depend on the methodological approach and forms of data involved.

Grouping and Filtering Data

Organizing data into groups enables the isolation of parts of the data corpus in order to focus on (combinations of) subsets of records or respondents. It is usually possible to group either whole files which represent a case or respondent such as interview files, published reports, or video files etc., or parts of files where, for example, several cases or respondents occur within a file, such as in focus group, survey, or field data. Whatever the situation, known characteristics, often conceptualized as sociodemographic attributes or variables, can be used to group whole data files or relevant parts of them. This enables the subsequent retrieval and interrogation of data based on an individual grouping, or combinations of them.

Characteristics which are known about the data at the outset, and perhaps sampled for, such as the sociodemographic attributes of respondents, for example, age, gender, and marital status, are common and powerful bases for grouping data. However, any analytically meaningful feature which is identified through the analysis can be handled at this level within CAQDAS packages, where it applies to the chosen organizational unit of analysis.

For example, in a project exploring the nonverbal interactions of university lecturers and their students in video-recorded tutorial sessions, analysis may reveal that some lecturers use certain hand gestures more often than others to reinforce particularly salient points. Where the individual is the organizational unit of analysis, an attribute which differentiates between lecturers according to their identified style of hand gestures, and another to reflect the frequency of their use, could be applied to all instances of individual lecturers' data. This would enable the subsequent grouping of lecturers who displayed similar hand gestures, in order to consider whether they also had other communicative techniques in common, and so on. In this way analysis can build incrementally as the results of a particular observation or interpretation are incorporated into the software and used to inform further enquiry.

Interpretive Coding

Many qualitative analyses involve organizing data by way of conceptually or thematically coding them. Coding is a main aspect of managing interpretation. There are many different approaches to qualitative coding which are not possible to discuss in detail here, but in general terms coding is the process by which segments of data are identified as relating to, or being an example of, a more general idea, instance, theme, or category.

Qualitative researchers may take an inductive, deductive, or combined approach to generating and applying code labels, each of which are usually well supported by software. From a technological point of view, all CAQDAS packages provide flexible means by which to code data. For example, codes can be generated from the data level, or independently from it. The same code can be applied to any number of data segments derived from any data source. In most of the packages designed originally for textual analysis, the same or overlapping or embedded data segment can be coded by any number of relevant codes. The amount of data coded can be increased or decreased. Codes can be merged or grouped; previously coded data can be recoded or uncoded. Codes can be defined, renamed, and lists printed or exported. Coded data can be easily retrieved and outputted (see the following text).

The coding functionality of most CAQDAS packages is similar, although there are some subtle variations which may be important to researchers from certain disciplinary backgrounds. One example is the way in which coding schema are visually presented within the chosen software. Researchers vary in their expectations of coding schema structures, some preferring visually hierarchical systems, others feeling constrained by perceived notions of hierarchy. What is more useful than any set structure is that you can often regroup, reorder, and visualize codes in alternative ways. This reflects the way that researchers, having broken data down by initial coding processes, often need

to combine codes and their data differently simply in order to output these new permutations of data for reexamination. Sometimes this can be part of a necessary stepping back from data having been immersed closely in them for too long.

Linking Data

Some approaches to qualitative data analysis view coding as limiting because it fragments and reduces the data to an extent which obscures the dialectic relationship between reading text (or viewing multimedia data) and writing. Associative trails through data are created rather than collecting similar segments at thematic or conceptual codes which abstracts away from the data. In these approaches, the analytic focus is the relationships between topics or objects rather than codes or concepts. Such work is best supported by hyperlinking and annotation tools.

Hyperlinking is a free association approach to building up an analytic chain of reasoning embedded in the data. It facilitates an unstructured, nonlinear approach to identifying, writing about, and linking ideas, whereby multidirectional associative trails can be created. Hyperlinking tools may be used to track processes as diverse as narrative, sequence, time, and interaction. To some extent the presence of these and other annotating tools in one or two CAQDAS packages counters arguments that the technology has cemented the dominance of code and retrieve as an overly dominant qualitative approach. A key debate concerning ethnographic approaches and the development of methodology alongside technology can be seen in *Sociology Online* (Coffey *et al.*, 1996; Lee and Fielding, 1996).

There are some customized software applications for handling data in this way and these developments have occurred separately from CAQDAS. These applications enable the creation of hypertexts; electronic multidimensional documents in which several writing spaces are viewable simultaneously, and multiple paths through the data are offered – see Silver and Fielding (2008) for more discussion on hyperlinking tools in CAQDAS packages.

Managing Interpretive Work

Interpretation is the core business of qualitative analysis. It takes different forms according to methodological approach, requirements of the research project, and the researcher's personal style of working. While it is important to reiterate that the software does not analyze or interpret data in itself, it does provide a range of tools which can facilitate these processes. As well as various ways of interrogating the data set discussed in the following interrogating data section, writing tools and visual ways to make connections offer powerful means of facilitating the interpretive process. Thematic coding and linking

between points in the data may already have contributed to the management of analysis itself (see the section on interpretive coding).

Writing

Taking notes, making annotations, and writing longer explanations about what is seen in the data are analytic processes which are undertaken throughout the process of qualitative research.

In addition to annotation tools already discussed, memo systems provide larger and more central writing spaces which can be used for a range of purposes. Recording the research progress, writing an analytic description, fleshing out a theoretical idea, summarizing what seems important about an interview, and referring back to research questions might all be managed in a well-designed memo system. The final written report can begin as separate memos within the software. The benefit of using integrated memo systems rather than keeping memos outside of the CAQDAS package relates to the systematic management and retrieval of writing, alongside the integration with data. This will be particularly so with methodologies relying on thick description. Familiarity with any software package which has a memo system enhances continuity, recall, and reflection.

Making Connections

Many CAQDAS packages include integrated mapping or networking tools enabling visual representations of aspects of work. The way these tools work vary considerably between packages and therefore offer quite different ways of working. Nevertheless, there are three key ways in which mapping tools can be particularly useful when undertaking qualitative analysis using CAQDAS packages. First, the user can graphically represent hypotheses, hunches, or theoretical models at the outset of a more deductive analytic project. Second, the user can represent the linkages which exist in the project, based on the tasks already undertaken. One may, for instance, use mapping tools to visually represent a developing theory or explanation having taken an inductive, bottom-up analytic direction. A less analytic use of mapping might be to illustrate the coding scheme in a graphic way to see at a glance the way it is structured. Where objects are illustrated, such as codes, documents, or memos, there is usually live interactivity back to the coded text, the document text, or the memo represented.

Interrogating Data

The CAQDAS packages offer various means to interrogate the data set based on earlier work, and these tools are the power house of moving beyond the basic tasks of code

and retrieve which characterize much qualitative work. Functionality ranges from simple to complex, not all of which can be included here.

Retrieving Data

Basic retrieval of coded or linked data provide means of reviewing progress across the whole database in order to reconsider a body of data, and perhaps to work on a more conceptual and abstract level. It is usually possible to retrieve data for these purposes without recourse to creating complex queries. Coded data can be viewed in context or lifted out of context to view in isolation, either within the software, or as output, for example, in a word processing application. This facilitates general consideration of all instances of a theme or concept, horizontally across all data files.

It is often also analytically useful to consider one file at a time, perhaps where a file represents an individual interview transcript or the field notes about a particular research setting. Viewing data vertically by all or selected codes provides a way of considering a case or situation as an independent entity. This can be achieved visually within the software, using a margin view illustrating where particular codes occur, or by generating an output report which lifts out all the segments within that file coded by each code.

Whatever the approach to coding qualitative data it will often be necessary to recode coded segments. If coding in a broad-brush or deductive way then this can be facilitated by most CAQDAS packages by first retrieving data coded at a general theme, and then recoding into more detailed codes. This process adds additional codes to the original data, thereby allowing the researcher to analyze the theme in more depth.

Most packages also provide a more quantitative overview of how data have been coded in various respects. Tables showing for example, frequency of code application by file, or number of words (or duration of an audio-video file) coded across the data set are usually either provided automatically, or can be generated very easily. Such tables are often fully interactive within the software; that is, each cell in the table provides direct link to the data regardless of the way it is counted in the table. Additionally, these tables can be exported to spreadsheet applications for more complex statistical analysis.

Identifying Patterns and Relationships

As discussed earlier, mapping tools can provide visual ways of making connections or viewing patterns and relationships in the data at an abstract level. Search or Query tools provide more sophisticated ways to interrogate the data set according to the presence and absence of codes as they have been applied to the data. The traditional CAQDAS packages initially developed for the analysis of textual data often provide the most sophisticated options.

Queries may be run at early stages of coding in order to test a hunch, or get a general overview of the current situation. Queries can usually be saved to be rerun at a later stage and doing so is a way of creating reminders of interesting aspects which deserve revisiting. They can contribute to the iterative and incremental coding process. It may be at the latter stages, however, when the researcher is confident the main coding process has been completed consistently, that the results of queries are used as the basis of analysis.

Boolean and proximity query operators are common to many CAQDAS packages, allowing the researcher to find where two or more codes occur together or separately in the data.

Comparing Subsets

Many qualitative projects include a comparative element, for example in cross-case analysis where it is necessary to compare how different cases or types of respondent discuss certain aspects, see section on Organizing data as discussed earlier.

The ease with which results, sometimes viewed as matrices, can be generated varies but many CAQDAS packages provide the facility to compare groups of codes by socio-demographic characteristics, such as gender, age, teacher, pupil, parent, etc. Alternatively, it may be useful to create a qualitative cross tabulation of codes by codes in order to see where particular codes occur together in the data.

Conclusion

This article has provided an overview of how CAQDAS packages can support the common tasks of qualitative data analysis. This is a rapidly changing area with new software developments occurring frequently. As basic CAQDAS tools have become well embedded, developers are able to stretch methodological and disciplinary boundaries by offering tools which provide means of handling and analyzing data in ways hitherto difficult or impossible. These developments provide the researcher with a range of options, some of which may be welcomed, others which may not be required by individual projects. The researcher determines the way the software is used, the sequencing of tasks, and the interpretation of the data. While software can certainly enable more transparency in process and offers means by which analytic rigor and the robustness of conclusions can be tracked, the quality of the output is not guaranteed or enhanced merely by use of the software. There is however great potential for parts of the analytic process to be more searching and ambitious and for the processes of analysis to be more flexible when using a CAQDAS package.

See *also*: Action Research in Education; Curriculum and Syllabus Design; Discourse Analysis; Participant Observation.

Bibliography

- Bazeley, P. (2006). The contribution of computer software to integrating qualitative and quantitative data and analyses. *Research in the Schools* **13**(1), 64–74. http://www.msera.org/rits_131.htm (accessed August 2009).
- Coffey, A., Holbrook, B., and Atkinson, P. (1996). Qualitative data analysis: Technologies and representations. *Sociological Research Online* **1**(1). <http://www.socresonline.org.uk/socresonline/1/1/lf.html> (accessed August 2009).
- Fielding, N. and Lee, R. (eds.) (1991/1993). *Using Computers in Qualitative Research*. London: Sage.
- Lee, R. M. and Fielding, N. F. (1996). Qualitative data analysis: Representations of a technology: A comment on Coffey, Holbrook and Atkinson. *Sociological Research Online* **1**(4). <http://www.socresonline.org.uk/socresonline/1/4/lf.html> (accessed August 2009).
- Lewins, A. and Silver, C. (2007). *Using Software in Qualitative Research: A Step-by-Step Guide*. London: Sage.
- Silver, C. and Fielding, N. (2008). Using computer packages in qualitative research. In Willig, C. and Stainton-Rogers, W. (eds.)

The Sage Handbook of Qualitative Research in Psychology, pp 334–351. London: Sage.

Weitzman, E. and Miles, M. (1995). *A Software Source Book: Computer Programs for Qualitative Data Analysis*. Thousand Oaks, CA: Sage.

Further Reading

- Bazeley, P. (2002). *Issues in Mixing Qualitative and Quantitative Approaches to Research*. 1st International Conference – Qualitative Research in Marketing and Management, University of Economics and Business Administration, Vienna. <http://www.researchsupport.com.au/MMIssues.pdf> (accessed August 2009).
- Friese, S., Gibbs, G., and Mangabeira, W. C. (eds.) (2002). Using Technology in the Qualitative Research Process. *Forum Qualitative Sozialforschung / Forum (FQS)*, **3** (2) May. <http://www.qualitative-research.net/index.php/fqs/issue/view/22> (accessed August 2009).
- Gilbert, N. (ed.) (2008). *Researching Social Life*, 3rd edn. London: Sage.
- Lewins, A. (2008). Computer assisted qualitative data analysis (CAQDAS). In Gilbert, N. (ed.) *Researching Social Life*, 3rd edn., pp 302–323. London: Sage.
- Lewins, A. and Silver, C. (2006). Choosing a CAQDAS package, *CAQDAS Networking Project Working Paper*. <http://caqdas.soc.surrey.ac.uk/ChoosingLewins&SilverV5July06.pdf>.

Conversational Analysis

H G Levine, University of California, Davis, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Conversational analysis (CA) is the study of interaction, both everyday, ordinary conversation, as well as the more formalized talk that occurs in institutional settings such as classrooms, courtrooms, and medical clinics. CA uncovers the underlying interactional rules and principles that guide the participants in a conversation to make intersubjective meaning possible. CA, as either a method of empirical inquiry or as theory of enacted social organization, has had an impact in a wide variety of social science disciplines and research arenas.

CA has its genesis in the late 1960s and early 1970s based, in large part, upon the earlier pioneering work of sociologists Harold Garfinkel and Erving Goffman. Their conceptual writings provided the intellectual justification for including everyday, commonplace activities carried out through mundane – but purposeful – interaction as core subjects for sociological inquiry. Garfinkel's further contribution was to argue that conversation represents an “institutional order *suigeneris*” (Heritage, 2005: 103) with a knowable set of interactional rules and practices of the participants as they construct meaning for each other in particular real-world contexts. For most writers in CA, conversations may reference or index sociological constructs/variables, such as social class, race, and gender; but do not depend upon, or necessarily derive their ordering from, them.

Harvey Sacks, Emanuel Schegloff, and Gail Jefferson are credited with founding CA as an approach to the study of talk-in-interaction as an inherently human practice that is constitutive of the lived social order. Sacks argued that the study of interaction was preferred over the then-standard preoccupation among academic sociologists with abstract conceptual structures that, by definition, were at a distance from – and, Sacks argued, could only be vaguely tied to – the specific events and actions in a group's life that the field of sociology was designed to understand.

For Sacks and others, there are three important assumptions that underlie CA (Heritage, 1984):

1. Interaction is structurally organized and this structure can be discerned through careful analysis – primarily with the aid of audiotaped recordings of natural interaction and a robust transcription system (which was developed to meet this need) that would allow the researcher to pinpoint not only what was said, but

how (e.g., with simultaneous or overlapping utterances, pauses, and intonation particularities).

2. Conversation is always contextual. Individuals bring context with them; that is, their utterances can only be understood (by them, as well as by researchers) in terms of what came before. An individual in an interaction then reaffirms this context for what comes next, thereby building a “turn-by-turn sequential organization of interaction,” one of the early discoveries of conversational analysts (Psathas, 1995: 13). These paired orderings are known in the field as adjacency pairs, and they have been identified and studied in a large variety of settings, including cross-culturally (e.g., Moerman, 1988).
3. Because of (1) and (2) above, no detail of interaction can be dismissed, on *a priori* grounds, as irrelevant. A further contribution of CA researchers has been to demonstrate how seemingly deviant or irregular instances of adjacency pairs recorded in real conversations can further demonstrate the underlying normative structure by which they operate.

One of the earliest, and still frequently cited, applications of CA to education is the work of Mehan (1974). Mehan studied classroom discourse and was able to demonstrate the ordering principles of student–teacher interaction. Mehan argued that, to understand how schooling impacts students and their learning, it is essential to start with “descriptions of the actual processes of education” (p. 5). Both the sociological facts of race, status, gender, and so on and so forth, as well as educational constructs, such as teacher quality or opportunity-to-learn, must be instantiated in the actualities of interaction.

See also: Classroom Ethnography; Discourse Analysis.

Bibliography

- Heritage, J. (1984). *Garfinkel and Ethnomethodology*. Cambridge: Polity Press.
- Heritage, J. (2005). Conversation analysis and institutional talk. In Fitch, K. L. and Sanders, R. E. (eds.) *Handbook of Language and Social Interaction*, pp 103–148. Mahwah, NJ: Erlbaum.
- Mehan, H. (1979). *Learning Lessons: Social Organization in the Classroom*. Cambridge, MA: Harvard University Press.
- Moerman, M. (1988). *Talking Culture: Ethnography and Conversation Analysis*. Philadelphia, PA: University of Pennsylvania Press.

Psathas, G. (1995). *Conversation Analysis: The Study of Talk-in-Interaction*. Thousand Oaks, CA: Sage.

Further Reading

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Hammersley, M. (2003). Conversation analysis and discourse analysis: Methods or paradigms? *Discourse and Society* **14**(6), 751–781.

Sacks, H., Schegloff, E., and Jefferson, G. (1974). A simplistic systematics for the organization of turn-taking for conversation. *Language* **50**, 696–735.

Critical Ethnography

D Foley, The University of Texas at Austin, Austin, TX, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Action anthropology – Sol Tax's 1950s style of engaged, politically active ethnography.

Collaborative ethnography – The twenty-first-century experimental ethnographic practices that empower the research subjects to produce accessible texts.

Cultural broker – An ethnographer who seeks to mediate between powerful institutions and powerless subjects.

Cultural critiques – A late twentieth century style of policy-relevant ethnography that critiques public institutions.

Decolonizing ethnography – Australian Aborigine model of ethnography that seeks to protect tribal interests.

Feminist perspective of science – A philosophy of science that deemphasizes a grandiose notion of absolute truth and scientific method.

Investigative ethnography – A sociological style of ethnography that uses covert field methods to critique societal institutions.

Participatory action research – An activist type of educational ethnography based on the philosophy of Paulo Friere and Fals Borda.

Positivism – A philosophical doctrine of social science that tries to emulate physical science methodologies.

Self-critical reflexivity – An alternative approach to positivist methods that emphasizes the interrogation of the ethnographer's subjectivity and field relationships.

Introduction

The purpose of this article is to highlight the post-1960s shift in educational ethnography toward more critical, policy-oriented studies of American society and its public schools. Initially, critical ethnography was often based on classic Marxism or neo-Marxist critical theory (Carspecken, 1996). Several recent literature reviews of educational ethnography (Levinson and Holland, 1996; Foley *et al.*, 2001; Villenas and Foley, 2002) note that the philosophical base of critical ethnography has expanded to include race,

gender, sexual identity, and postcolonial perspectives. These developments underscore the widespread disenchantment with the positivist notion of the social science as objective and value-free.

This academic and methodological revolt against positivism is political in two senses: first, it generally seeks to transform the knowledge production of the academy, and second, it seeks to use the knowledge produced to make society more egalitarian. This transformation of knowledge production is rooted in new ways of understanding social scientific and educational research. Educational philosopher Schwantz (2000) argues that the post-1960s methodological revolt has interpretive, hermeneutic, and constructivist alternatives. Elsewhere, we have chronicled the following self-critical reflexive practices in post-positivist critical ethnography: confessional, theoretical, intertextual, and deconstructive (Foley, 2002).

These new reflexive ethnographic practices are often grounded in a feminist perspective of science. Donna Haraway (1988) and Sandra Harding's (1998) concept of science allows politically progressive critical ethnographers to make strong, yet more focused knowledge claims. Harding's discussion of standpoint theory and Haraway's notion of situated knowledge are so well known that there is little need to elaborate here. Suffice to say that many critical ethnographers have replaced the grand positivist vision of speaking from a universalistic, objective standpoint with a more modest notion of speaking from a historically and culturally situated standpoint. Interpreting from a historically specific standpoint acknowledges the impossibility of what Haraway aptly calls the god trick of an omnipotent standpoint. If an investigator subscribes to this philosophy of science perspective, they see themselves as mere culture-bound mortals speaking from very particular race, class, gender, and sexual identity locations. Since all standpoints represent particular interests and positions in a hierarchical society, they are partial, thus ideological in the sense of transparent positionality.

Once an ethnographer abandons the positivist fallacy that specific research techniques and methodologies can produce a detached, objective, or omnipotent standpoint, it also opens the door to use more intuitive or subjective ways of knowing. Hence, contemporary critical ethnographers of various philosophical stripes are beginning to practice introspection, memory work, autobiography, and even dreams as important ways of knowing. In the current experimental moment, the road to more defensible accounts goes through the ethnographer's reflections on

her subjectivity and intersubjective relationships (Denzin, 1997). The new, more reflexive critical ethnographer explores the intense self–other interaction of fieldwork and the constructed nature of ethnographic narratives.

Consequently, many twenty-first-century critical ethnographers have lost faith in the belief that knowledge production is innocent or neutral in a society marked by class, racial, and sexual conflict. Anthropologists in post-1960s era began calling for reinventing the field (Hymes, 1999), studying up (Nader, 1996), and studying people without history (Wolfe, 1982). They began advocating cultural critiques of modern society and its institutions (Marcus and Fischer, 1986). Such cultural critiques often studied ruling groups and ruling ideologies and/or the struggles of various oppressed groups (Foley and Moss, 2001). Investigators subscribing to this perspective are generally committed to value-oriented research that promotes an egalitarian society.

One pre-1960s engaged anthropologist, who helped initiate anthropology's historical transformation, was Sol Tax (Foley, 1999). His action anthropologists operated without the sponsorship of government bureaucracies or private nongovernment organizations (NGOs). They found independent funding and worked more directly with and for the people they were studying. Tax argued that because action anthropologists became accepted insiders, they were positioned to collect better data on social change and acculturation than detached scientific ethnographers. Consequently, action anthropologists would help the community while they wrote trustworthy ethnographies. In effect, Tax envisioned a social science that created knowledge that was as practical and useful as it was theoretical and universal. For him, academic social scientists had produced a false notion of science and knowledge that privileged the theoretical over applied, practical knowledge.

In effect, anthropologists subscribing to action anthropology and the subsequent anthropology-of-cultural-critique approach began to occupy the methodological and ideological terrain that the earlier sociologists occupied. Vidich and Lyman (2000) note that in the 1930s and 1940s community sociologists such as Lynd (1956) and the Chicago School of Sociology were writing critiques of social class inequality and positive portraits of marginalized, stigmatized groups. One of the most powerful critical sociologists who emerged from this era was C.W. Mills (1959). He urged positivistic sociologists to abandon arid grand theories and to historicize and situate their studies in the everyday realities of power structures and social movements. Mills wanted to produce useful knowledge that addressed societal problems. An excellent contemporary illustration of this approach in education is Hugh Mehan's (2008) career as a design sociologist. After writing numerous critiques of American public education, Mehan helped create a California public school that detracks and

promotes the academic achievement of racially and culturally different populations.

In a complimentary fashion, existentialist sociologist Jack Douglas (1976) preferred an investigative posture to study societal problems. Tapping into muckraking new journalism (Wolfe, 1974), Douglas advocated operating covertly to expose corrupt bureaucrats or hate group leaders. He argued that in a politically corrupt, conflict-filled society, any means used to get the story was justifiable, if it exposed harmful public practices. Marxist educator Michael Apple's (2006) recent philosophical and empirical reflection on the religious right's attempt to privatize American public education illustrates nicely the investigative impulse in much critical educational ethnography.

After reviewing many contemporary studies of American culture, Foley and Moss (2001) concluded that late twentieth-century critical ethnography had developed a complex philosophical basis in the continental philosophies of post-Marxism, postmodernism, and feminism. The current crop of critical ethnographers seems to be focusing more on dramatic public issues, and they are finding ways to reach wider audiences. The work of Peggy Sanday (1990, 1996) on campus date rape and her involvement and coverage of rape trials is a case in point. Nancy Scheper-Hughes (1992) and Scheper-Hughes and Sargent (1998) study of child welfare issues and Third-World organ harvesting is also exemplary.

In the field of educational anthropology, we tried to document the gradual shift toward more critical qualitative and ethnographic approaches from the mid-1980s on (Foley *et al.*, 2001). Kemmis and McTaggart (2005) provide an excellent introduction to activist educational research using more collaborative methodologies called participatory action research (PAR). PAR researchers often base their approach on the philosophy of Latin American social activists Paulo Friere and Fals Borda. These researchers have affinities with the more activist-oriented applied anthropologists (Eddy and Partridge, 1987). They often play the role of democratic facilitator and consciousness-raiser, or cultural broker between powerful institutions and the disenfranchised citizens. The exemplary educational studies of Linda McNeil (2000) and Angela Valenzuela (2005) have forcefully critiqued the present federal policy called No Child Left Behind. These scholars have worked tirelessly with local teachers, community educational leaders, and state legislators for Texas school reform, and they have publicized their critique on national TV shows such as *Sixty Minutes*.

Increasingly, anthropologists who seek to serve the public interest are advocating a more politicized type of policy studies (Levinson and Sutton, 2001; Okongwu and Mencher, 2000; Kane and Mason, 2001). These surveys of the field, and a recent school of American research conference on critical ethnography (Marcus, 1999), describe a

host of new politically relevant sites of cultural critiques: corporate agriculture, environmental pollution, pharmaceutical dumping, transnational labor migration, the publishing industry, cyberspace hackers, the AIDS crises, and media and legal system demonization and criminalization of urban street life and informal economies based on drugs, sex, and cultural rebellion. These new critical ethnographers are beginning to write more accessible, less jargon-filled accounts. They are also finding additional ways to step outside the narrow publication circles of academia to influence public opinion. This may mean appearing as experts on talk and news shows. It may also mean providing reporters with expert testimony for their journalistic exposes, and working as policy advisors with state legislators, thus influencing educational legislation (Foley and Valenzuela, 2004).

A recent Wenner-Gren conference reaffirmed this explosion of politically and policy relevant critical ethnographies (Fox and Fields, 2008). The Wenner-Gren Foundation is the anthropological equivalent of the Spencer Foundation. It is the most prestigious foundation that sponsors research and conferences. During a recent theme conference on where anthropology as a field is headed, participants Charles Hale and Joan Rappaport advocated a much more politically engaged type of ethnographic practice. In their version of critical ethnography, activist ethnographers become deeply involved in progressive social movements and community-based reforms. These ethnographers reported using much more collaborative field methods, which invite the research subjects to co-theorize and co-author the ethnographies. Luke Lassiter (2005) has written an excellent historical review of how collaborative ethnographic methodologies developed in anthropology. Although still not the dominant mode of ethnographic research, collaborative and activist approaches are definitely gaining ground in the field. A recent special issue of the *Anthropology and Education Quarterly* (2008) on activist educational ethnography reports on a variety of projects in which ethnographers empowered youth and immigrant populations through ethnographic research, oral history, film production, drama, and writer's workshops. Many of these ethnographic projects used highly collaborative methodologies to produce written and multimedia accounts of their communities.

Conclusion

Overall, it would seem that progressive social scientists and educational researchers are creating a more publicly activist role for themselves. The cultural diversification of the academy and the social sciences is part of this change process. People of color, women, gays, and working-class academics are slowly replacing the upper-middle-class

white male gentleman scholars of early twentieth-century anthropology. Further, the emergence of the interdisciplinary field of critical cultural studies has created many new journals and special series in university presses. A market for more critical, investigative ethnographies that expose relations of power and exploitation has clearly evolved. However, these developments have their limits.

All young scholars are aware that Division I research institutions operate under a publish-or-perish logic. One either publishes articles in the refereed journals of his/her field – and books if his/her department is a book department – or he/she is not awarded tenure. The rub for many critical ethnographers is that their scholarship must be political in an academically acceptable manner. Consequently, many progressive academics tend to produce knowledge that other academics recognize as having universal or theoretical virtue. This may lead them away from producing local knowledge for activist groups that helps win lawsuits or rent strikes or that decolonizes the power of outside researchers that Maori scholar Linda Tuhiwai Smith (1999) advocates. Somehow, activist critical ethnographers have to find a way to satisfy the demands of the academy and local groups.

As many participants in the aforementioned Wenner-Gren conference noted, this has proven to be no simple task (Fox and Fields, 2008). A number of the conference participants pointed out interesting career shifts. These conversations left the participants wondering if critical ethnographers, who become deeply involved in local political struggles, have stopped writing academic books and articles. Contrary to concerns that radicals are taking over academia, the conference participants felt that activist colleagues often failed to get tenure, or they left the academy before such decisions were made. Unfortunately, we know precious little about where pushed-out activists go. Are they teaching in community colleges? Are they writing articles for local newspapers? Or have they succumbed to political disillusionment? Most critical ethnographers rarely chronicle the psychological and monetary price they pay for their political activism. There are few good intellectual histories of how the twenty-first-century academy is changing.

Overall, we have tried to raise some issues and make some distinctions that will move self-proclaimed critical ethnographers to interrogate their current ethnographic practices. Moreover, we have suggested many ways to question our notions of purpose, positionality, collaboration, and writing styles. We see many promising new varieties of critical ethnography emerging from the post-1960s revolt against positivism and a social scientific/educational establishment that was content with producing arid, a-historical studies that serve academic interests rather than the public interest. Transforming the academic knowledge production industry obviously requires much more than challenging the ideology of positivism

and scientism. We also need to change the way academic publishing is organized and controlled, and the way the academy awards promotion and tenure for publication and public service. In addition, we must continue to open up the academy to underrepresented groups.

See also: Action Research in Education; Critical Theory; Ethnography; Interpretive Research.

Bibliography

- Apple, M. (2006). *Educating the 'Right' Way: Markets, Standards, God and Inequality*, 2nd edn. London: Routledge.
- Carspecken, P. (1996). *Critical Ethnography in Educational Research: A Theoretical and Practical Guide*. London: Routledge.
- Denzin, N. (1997). *Interpretive Ethnography: Ethnographic Practices for the 21st Century*. Newbury Park, CA: Sage.
- Douglas, J. (1976). *Investigative Social Research*. Beverly Hills, CA: Sage.
- Eddy, E. and Partridge, W. (eds.) (1987). *Applied Anthropology in America*, 2nd edn. New York: Columbia University Press.
- Foley, D. (1999). The Fox project: A reappraisal. *Current Anthropology* 40(2), 171–191.
- Foley, D. (2002). Critical ethnography: The reflexive turn. *International Journal of Qualitative Studies in Education* 15(4), 469–491.
- Foley, D., Levinson, B., and Hurtig, J. (2001). Anthropology goes inside: The new educational ethnography of ethnicity and gender. In Secada, W. (ed.) *Review of Research in Education*, pp 37–87. Washington, DC: American Educational Research Association Publications.
- Foley, D. and Moss, K. (2001). Studying American cultural diversity: Some non-essentializing perspectives. In Susser, I. and Patterson, T. (eds.) *Cultural Diversity in the United States*, pp 130–152. Oxford: Blackwell.
- Foley, D. and Valenzuela, A. (2004). Critical ethnography: The politics of collaboration. In Denzin, N. and Lincoln, Y. (eds.) *The Handbook for Ethnographic Research*, 3rd edn., pp 217–234. Beverly Hills, CA: Sage.
- Fox, R. and Fields, L. (eds.) (2008). *Anthropology Put to Work*. Oxford: Berg.
- Haraway, D. (1988). Situated knowledges: The science question in feminism as a site of discourse on the privilege of partial perspective. *Feminist Studies* 14, 575–599.
- Harding, S. (1998). *Is Science Multicultural? Postcolonialisms, Feminisms, and Epistemology*. Bloomington, IN: Indiana University Press.
- Hymes, D. (ed.) (1999). *Reinventing Anthropology*, 2nd edn. Ann Arbor, MI: University of Michigan Press.
- Kane, S. and Mason, T. (2001). AIDS and criminal justice. *Annual Review of Anthropology* 30, 457–479.
- Kemmis, S. and McTaggart, R. (2005). Participatory action research: Communicative action and the public sphere. In Denzin, N. K. and Lincoln, Y. S. (eds.) *The Sage Handbook of Qualitative Research*, 3rd edn., pp 553–603. Thousand Oaks, CA: Sage.
- Lassiter, L. (2005). *The Chicago Guide to Collaborative Ethnography*. Chicago, IL: University of Chicago Press.
- Levinson, B. and Holland, D. (1996). The cultural production of the educated person an introduction. In Levinson, B., Foley, D., and Holland, D. (eds.) *The Cultural Production of the Educated Person: Critical Ethnographies of Schooling and Local Practice*, pp 1–54. Albany, NY: State University of New York Press.
- Levinson, B. and Sutton, M. (eds.) (2001). *Policy as Practice: Towards a Comparative Sociological Analysis of Educational Policy*. Westport, CT: Ablex.
- Lynd, R. (1956). *Middletown: A Study of American Culture*, 2nd edn. New York: Harcourt Brace.
- Marcus, G. (ed.) (1999). *Critical Anthropology Now: Unexpected Contexts, Shifting Constituencies, Changing Agendas*. Santa Fe: School of American Research Press.
- Marcus, G. and Fischer, M. (1986). *Anthropology as Cultural Critique: An Experimental Moment in the Human Sciences*. Chicago, IL: University of Chicago Press.
- McNeil, L. (2000). *Contradictions of School Reform: Educational Costs of Standardized Testing*. New York: Routledge.
- Mehan, H. (2008). Engaging the sociological imagination: My journey into design research and public sociology. *Anthropology and Education Quarterly* 39(1), 77–91.
- Mills, C. W. (1959). *The Sociological Imagination*. New York: Oxford.
- Nader, L. (1996). *Naked Science: Anthropological Inquiry into Boundaries, Power, and Knowledge*. New York: Routledge.
- Okongwu, F. and Mencher, J. P. (2000). The anthropology of public policy: Shifting terrains. *Annual Review of Anthropology* 29, 107–124.
- Sanday, P. R. (1990). *Fraternity Gang Rape: Sex, Brotherhood and Privilege on Campus*. New York: New York University Press.
- Sanday, P. R. (1996). *A Woman Scorned: Acquaintance Rape on Trial*. New York: Doubleday.
- Scheper-Hughes, N. (1992). *Death without Weeping: The Violence of Everyday Life in Brazil*. Berkeley, CA: University of California Press.
- Scheper-Hughes, N. and Sargent, C. (eds.) (1998). *Small Wars: The Cultural Politics of Childhood*. Berkeley, CA: University of California Press.
- Schwandt, T. A. (2000). Three epistemological stances for qualitative inquiry. In Denzin, N. K. and Lincoln, Y. S. (eds.) *Handbook of Qualitative Research*, 2nd edn., pp 37–84. Thousand Oaks, CA: Sage Publications.
- Special Issue (2008). *Activist Educational Ethnography*. *Anthropology and Education Quarterly* 39(1).
- Tuhiwai Smith, L. (1999). *Decolonizing Methodologies: Research and Indigenous Peoples*. London: Zed Books.
- Valenzuela, A. (ed.) (2005). *Leaving Children behind: Why Texas-Style Accountability Fails Latino Youth*. Albany, NY: State University of New York Press.
- Vidich, A. J. and Lyman, S. (2000). Qualitative methods: Their history in sociology and anthropology. In Denzin, N. and Lincoln, Y. (eds.) *The Handbook of Qualitative Research*, 2nd edn., pp 37–84. Thousand Oaks, CA: Sage.
- Villenas, S. and Foley, D. (2002). Chicano/Latino critical ethnography of education: Borderlands cultural productions from La Frontera. In Valencia, R. R. (ed.) *Chicano School Failure and Success: Past, Present, and Future*, 2nd edn., pp 195–226. London: RoutledgeFalmer.
- Wolfe, E. (1982). *Europe and People without History*. Berkeley, CA: University of California Press.
- Wolfe, T. (1974). *The New Journalism*. New York: Harper and Row.

Critical Race Theory

D Gillborn, Institute of Education, London, UK

G Ladson-Billings, University of Wisconsin–Madison, Madison, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Color-blindness – An approach that claims to treat everyone the same regardless of their race. Although this view asserts a liberal and equitable intent, by refusing to acknowledge the significance of historic and contemporary race inequity, color-blind policies act to protect the status quo.

Contradiction-closing case – A perspective, developed by Derrick Bell and Richard Delgado, highlighting the limited and uncertain long term effects of landmark civil rights breakthroughs. Although such cases appear to close the contradiction between the rhetoric of equal opportunities and the reality of racial injustice, radical changes are typically thwarted by obstruction and delay.

Critical race theory – A multi-disciplinary approach that combines social activism with a critique of the fundamental role played by White racism in shaping contemporary societies.

Differential racialization – Refers to the constantly changing and malleable nature of racist stereotypes. Mainstream society constructs different stereotypes of different ethnic groups, but also, those stereotypes can change over time.

Interest convergence principle – An analysis, first proposed by Derrick Bell, noting that advances in race equality have historically come about only when White elites see the changes as in *their* own interests; especially as a means of limiting civil unrest and protecting against fundamental change.

Intersectionality – An approach that is especially sensitive to the ways in which multiple axes of oppression interact. For example, analyzing how racism operates within and across factors such as social class and gender.

Praxis – A combination of critical theoretical analysis and applied practical strategies of resistance.

Racial realism – A perspective that insists on the need to interrogate the workings of race and racism in their historical context in the real world rather than as hypotheticals in an abstract analytic form.

Social constructivism – A perspective that views phenomena such as race, class, and gender as socially constructed categories. Although ‘race’

masquerades as fixed and natural, it is a constantly changing, malleable and contested category.

White supremacy – Commonly refers to individuals and groups who engage in the crudest, most obvious acts of race hatred (such as extreme nationalists and neo-Nazis). In critical race theory (CRT) the term refers to a regime of assumptions and practices that constantly privilege the interests of White people but are so deeply rooted that they appear normal to most persons in the culture.

Critical race theory (CRT) is the name taken by a radical perspective on the nature and significance of race inequality in society. The approach assumes a social constructivist view of race and races as products of social thought and relations. Not objective, inherent, or fixed, they correspond to no biological or genetic reality; rather, races are categories that society invents, manipulates, or retires when convenient (Delgado and Stefancic, 2001: 7). The name critical race theory was coined in the 1980s to describe a movement of legal scholars, mostly people of color, who were dissatisfied with existing approaches to the field. Subsequently, the approach grew in popularity and transcended its disciplinary boundaries, moving into educational studies in the mid-1990s. The approach is now established as an important part of educational studies and, although it was originally focused entirely on the US, CRT is being taken up by critical scholars internationally.

The article examines the origins of CRT and identifies the key defining elements that characterize the approach. Some common myths and misunderstandings are also addressed.

The Origins and Development of Critical Race Theory

CRT has its roots in the centuries-old diasporic experiences and struggles of people of color, especially enslaved Africans and their descendants in the US. The perspective builds on this tradition in numerous ways, including the central role it devotes to political struggle, its concern for storytelling (and other narrative approaches), and the significant position accorded to key Black intellectual figures

of the nineteenth and twentieth centuries such as Frederick Douglass and W. E. B. Du Bois (Baszile, 2008; Mills, 2003).

As a self-consciously new and oppositional form of anti-oppressive theory, CRT began in the 1970s and 1980s in the work of legal scholars. It began as an alternative to dominant perspectives, not only the conservative mainstream but also the ostensibly radical tradition of critical legal studies (CLS) which, in the words of West (1995: xi) deconstructed liberalism, yet seldom addressed the role of deep-seated racism in American life. Frustration with the silence on racism prompted CRT scholars to foreground race and to challenge not only the foci of existing analyses, but also the methods and forms of argumentation that were considered legitimate. The foundational critical race theorists in law include Derrick Bell, Kimberlé Crenshaw, Richard Delgado, Alan Freeman, Angela Harris, Charles Lawrence, Mari Matsuda, and Patricia Williams. These writers represent a range of racial/ethnic heritages (including people of African American, Latina/o, Asian American, and White backgrounds) but share a commitment to analyze and oppose the workings of race inequality in legal culture and, more generally, in US society as a whole.

There is no single canonical statement of CRT; the perspective is built upon a series of key insights which are constantly refined through their application analytically and practically. In this sense, critical race theorists view social theory as a work in progress. But this does not mean that CRT is any less serious about the importance of theory – quite the contrary. From its first iteration as a new approach, critical race scholars have staked a claim to the conceptual importance of their work. Kimberlé Crenshaw, for example, describes how she and colleagues sought to find a form of words that could be used to describe (and provide a rallying point for) the new ideas they were developing as they began to organize what was to become the first ever CRT workshop (held at the University of Wisconsin, Madison, in July 1989):

Turning this question over, I began to scribble down words associated with our objectives, identities, and perspectives, drawing arrows and boxes around them to capture various aspects of who ‘we’ were and what we were doing (...) we settled on what seemed to be the most telling marker for this peculiar subject. We would signify the specific political and intellectual location of the project through ‘critical’, the substantive focus through ‘race’, and the desire to develop a coherent account of race and law through the term ‘theory’ (Crenshaw, 2002: 1360–1361).

This practical and strategic orientation reflects a perspective that Bell (1992) terms racial realism, that is, a determination to continually interrogate the workings of race and racism in the real world rather than as hypotheticals in an abstract analytic context. The real-world focus of CRT

should not be seen as in any way lessening its claim to be taken seriously as a major innovation in social theory. From its inception, CRT has encountered patronizing and dismissive responses from academics who found its focus on racism distasteful and/or threatening (Crenshaw, 2002: 1357).

CRT quickly began to move beyond law schools and was introduced into educational studies in the mid-1990s by Ladson-Billings and Tate (1995). Subsequently, the approach has been adopted by numerous scholars, especially people of color working with qualitative methods, most notably Thandeka Chapman, Adrienne Dixon, Marvin Lynn, Laurence Parker, Celia Rousseau, Daniel Solórzano, David Stovall, Edward Taylor, and Tara Yosso. CRT is also building an international presence, including work in the UK (Gillborn, 2005, 2008; Hylton, 2008) and Australia (McDonald, 2003; Moreton-Robinson, 2004).

Tenets of CRT

From its earliest formulations, CRT has generally been united by a dual concern to understand and oppose race inequality. In an influential statement of the approach, Kimberlé Crenshaw and colleagues state:

Although Critical Race scholarship differs in object, argument, accent, and emphasis, it is nevertheless unified by two common interests. The first is to understand how a regime of white supremacy and its subordination of people of color have been created and maintained ... The second is a desire not merely to understand the vexed bond between law and racial power but to change it (Crenshaw *et al.*, 1995: xiii).

Within CRT, the term white supremacy is used in a particular way that differs from its usual understanding in mainstream writing: whereas the term commonly refers to individuals and groups who engage in the crudest, most obvious acts of race hatred (such as extreme nationalists and Neo-Nazis), in CRT the more important, hidden, and pervasive form of White supremacy lies in the operation of forces that saturate the everyday mundane actions and policies which shape the world in the interests of White people:

[By] ‘white supremacy’ I do not mean to allude only to the self-conscious racism of white supremacist hate groups. I refer instead to a political, economic, and cultural system in which whites overwhelmingly control power and material resources, conscious and unconscious ideas of white superiority and entitlement are widespread, and relations of white dominance and non-white subordination are daily reenacted across a broad array of institutions and social settings (Ansley, 1997: 592).

Many critical race scholars view White supremacy, understood in this way, as central to CRT in the same way that the notion of capitalism is to Marxist theory and

patriarchy to feminism (Stovall, 2006). This perspective on the nature and extent of contemporary racism is one of the key defining elements of CRT.

The Centrality of Racism

CRT begins with a number of basic insights. One is that racism is normal, not aberrant, in American society. As racism is an ingrained feature of our landscape, it looks ordinary and natural to persons in the culture (Delgado and Stefancic, 2000: xvi).

CRT views racism as more than just the most obvious and crude acts of race hatred; it focuses on the subtle and hidden processes that have the effect of discriminating regardless of their stated intent. In the political mainstream racism tends to be associated with acts of conscious and deliberate race hatred; discrimination is assumed to be an abnormal and relatively unusual facet of the education system. In contrast, CRT suggests that racism operates much more widely; often through the routine, mundane activities and assumptions that are unquestioned by most practitioners and policymakers, for example, through the design of the curriculum, the operation of certain forms of assessment, and the selection and training of teachers who overwhelmingly replicate dominant cultural norms and assumptions about race and racial inequality (Ladson-Billings, 2004).

Critical race theorists do not view racism as a simple or unchanging aspect of society. CRT challenges ahistoricism by stressing the need to understand racism within its social, economic, and historical context (Matsuda *et al.*, 1993: 6). The notion of differential racialization refers to the constantly changing and malleable nature of racist stereotypes. For example, a group once seen as conservative and conformist might be redefined as competitive and threatening at another time, for example, Japanese workers in the US and Asian groups in the UK during the twentieth century.

The focus on racism in CRT does not operate to the exclusion of other forms of social inequality. Indeed, a key aspect of CRT is a concern with intersectionality, that is, an attempt to analyze how racism operates within and across other axes of differentiation such as social class and gender (Crenshaw, 1995; Tate, 1997).

A Critique of Liberalism

Another distinctive theme is CRT's critique of liberalism. CRT portrays dominant legal claims of neutrality, objectivity, color-blindness, and meritocracy as camouflages for the self-interest of powerful entities of society (Tate, 1997: 235). In the education system, for example, racism is figured in the distribution of material and educational resources and even in teachers' notions of ability and motivation (Gillborn, 2008). In this situation, the adoption

of color-blind approaches (which refuse to acknowledge racial reality) and an emphasis on supposed merit (as measured by dominant assessments) may appear open and equitable, but the playing field is not level. Minoritized students are more likely to attend poorly funded schools, with less-highly qualified teachers and, because of socio-economic inequalities, they are less likely to enjoy additional educational resources at home, such as Internet access (Ladson-Billings, 2006). Under such unequal conditions, a color-blind insistence on a simple merit standard will not only ensure that race inequalities continue but also present them as fair and just.

The Call to Context (Experiential Knowledge and Storytelling)

CRT places a special importance on the experiential knowledge of people of color. There is no assumption that minoritized groups have a singular or true reading of reality; rather, there is recognition that by experiencing racial domination, such groups perceive the system differently and are often uniquely placed to understand its workings. Delgado (1989) is one of the leading advocates of the need to name one's own reality. Inspired by the scholarship of Derrick Bell, and the centuries-old traditions of storytelling in minoritized communities, Delgado argues forcefully for the use of narrative and counter-storytelling as a means of presenting a different reading of the world, one that questions taken-for-granted assumptions and destabilizes the framework that currently sustains, and masks, racial injustice.

A Revisionist Critique of Civil Rights Progress (the Interest Convergence Principle)

Detractors have sought to present CRT as disrespectful of civil-rights campaigns and their victories but this misreads the approach. CRT is not critical of the campaigns or the people who sacrificed so much to advance race equality (Crenshaw *et al.*, 1995). Rather, CRT examines the limits to reform via law and policymaking, and shows how even apparently radical changes are reclaimed and often turned back over time. A key element here is the concept of interest convergence. Put simply, this view argues that advances in race equality come about only when White elites see the changes as in their own interests. Bell (2004: 59), who coined the interest-convergence principle, summarizes the idea like this:

Justice for blacks vs. racism = racism.

Racism vs. obvious perceptions of white self-interest = justice for blacks.

It is important to note that interest convergence does not envisage a rational negotiation between minoritized

groups and White power holders, where change is achieved through the mere force of reason and logic. Rather, history suggests that advances in racial justice must be won through protest and mobilization, so that taking action against racism becomes the lesser of two evils for White interests. For example, the moves to outlaw segregation in the 1960s are usually thought of as signs of enlightenment and a landmark civil-rights victory. But they must be understood within the context of the Cold War and the need for the US to recruit friendly African states (Dudziak, 1988):

No such decision would have been possible without the world pressure of communism which made it 'simply impossible for the United States to continue to lead a "Free World" with race segregation kept legal over a third of its territory W. E. B. Du Bois (1968 quoted in Bell, 2004: 67).

The obvious signs of segregation – such as separate toilets and lunch counters – may have gone but the reality of ingrained racism continues in economic, residential, and educational terms. It has been argued that more African Americans now attend segregated schools than they did in 1954 at the time of the Supreme Court decision in the *Brown versus Board of Education* case (Delgado and Stefancic, 2001: 33). Delgado and Stefancic (2001: 24) describe the process like this:

... after the celebration dies down, the great victory is quietly cut back by narrow interpretation, administrative obstruction, or delay. In the end, the minority group is left little better than it was before, if not worse. Its friends, the liberals, believing the problem has been solved, go on to something else ... while its adversaries, the conservatives, furious that the Supreme Court has given way once again to undeserving minorities, step up their resistance.

Landmark victories may actually come to operate in ways that protect the racist *status quo*: these are sometimes known as contradiction-closing cases, which operate like a safety valve to provide a solution when the gap grows too large between, on one hand, the liberal rhetoric of equal opportunities and, on the other hand, the reality of racism.

[contradiction-closing cases] are a little like the thermostat in your home or office. They assure that there is just the right amount of racism. Too much would be destabilizing – the victims would rebel. Too little would forfeit important pecuniary and psychic advantages for those in power (Delgado, 1995: 80).

Landmark cases – such as the *Brown* desegregation case in the US, and the Stephen Lawrence Inquiry in the UK (Macpherson, 1999) – appear to have addressed blatant race inequalities but, in reality, little or nothing changes. Indeed, such cases are sometimes used as yet another weapon against further reform because they:

... allow business as usual to go on even more smoothly than before, because now we can point to the exceptional case and say, 'See, our system is really fair and just. See what we just did for minorities or the poor' (Delgado, 1999: 445).

CRT is a wide-ranging perspective that can be applied to numerous different issues within the field of education. For example, Ladson-Billings (1998) has noted CRT's relevance to the curriculum (where the selection of material typically promotes a sanitized version of historical and contemporary race relations), instruction, and pedagogy (where teachers' assumptions often disadvantage particular minoritized groups who are seen as a likely source of conflict rather than academic achievement), assessment (where the instruments used to grade and select students frequently encode unrecognized race bias), and in school funding (where, for example, African American students are more likely to attend schools with the lowest per-pupil funding and where teachers have less secure prospects).

Myths and Misunderstandings

Like any new perspective, CRT has been subject to a range of responses and critiques. Some of the engagement has been positive and constructive, pushing critical race scholars to clarify their arguments and develop further analyses. Other responses, however, have sought to reassert traditional assumptions (in the guise of scientific rigor) or dismiss CRT as misguided or simplistic (see Stovall (2006) for a review and response). As we have noted above, CRT offers a view of the world that is fundamentally at odds with mainstream assumptions and so it is no surprise that the approach is sometimes misunderstood. Three myths seem particularly common: first, that CRT views race as the source of all major inequalities; second, that CRT views white people as uniformly racist and privileged; and finally, that CRT promotes a sense of hopelessness and pessimism.

Despite its central focus on racism, CRT does not insist that race is always the single most important factor in every situation. CRT argues that race/racism is always relevant to an understanding of wider social inequalities but it is not the only element. Indeed, race inequity often cannot be fully understood in isolation from other axes of differentiation such as class and gender. As Stovall (2006: 252) notes:

Vital to this misinterpretation is the semantics of referencing CRT as a critique solely of 'race'. In no CRT literature is there a claim to the unanimity of race. The critique has and continues to be one of the functions of White supremacy and the complexities of race.

Similarly, it is sometimes argued that by identifying the underlying forces that legitimate and support white supremacy CRT imagines all white people to be the same: in fact the criticism betrays a one-dimensional reading of CRT. Critical race theorists do not imagine that White people are uniformly privileged and racist, nor do they feel that all Whites benefit equally from White supremacy. Such a position is patently ludicrous, especially in view of the fact that foundational CRT writers have repeatedly noted how interest convergence usually operates to defend White elites at the expense of lower-class Whites (Bell, 2004). However, CRT does show how even working-class Whites draw advantage from their Whiteness (Harris, 1993). Whites do not benefit equally, but they do all benefit from Whiteness to some degree (McIntosh, 1992). For example, when the attainment of the most economically disadvantaged White students in the UK dipped marginally below that of their Black peers, the media responded with stories blaming “the race-relations industry” and claiming that neo-Nazi groups would gain advantage. The stories failed to mention that White students outperformed virtually every minoritized group among the 86% of the school population not counted as living in poverty (Gillborn, 2008). Hence, even for the White students living in greatest poverty, their race meant that the media perceived a national scandal if their achievement was not greater than similarly disadvantaged minoritized peers.

A third myth about CRT is that, by viewing racism as such a deep-rooted aspect of contemporary society, the perspective promotes hopelessness. Bell (1992: ix) recalls an incident when he was challenged at a public reading of his work:

Professor Bell, you have achieved much despite racial discrimination. How dare you now deny our children the hope that they may enjoy a success like yours?

The author responded that “it was the society and not me” that closes down opportunities for African Americans; he did not create the situation, he simply chronicled what society had done and was likely to do. In fact, far from promoting a sense of hopelessness, CRT insists on the vital importance of active resistance against racism.

Bell argues that a total victory over racism may prove elusive but sees a duty to combat injustice (against all oppressed groups) as a central component of what he calls a life fulfilled (Bell, 1992: xi). The history of racism and education in the UK, for example, clearly demonstrates that all meaningful advances in race equality have come about as a result of community action (Tomlinson, 2008). Antiracist action may never entirely remove racism but in the absence of resistance, it is certain that racist inequity would worsen. As Frederick Douglass observed more than 150 years ago: “If there is no struggle, there is no progress ... Power

concedes nothing without a demand. It never did, and it never will” (quoted in Crenshaw, 2002: 1372).

Delgado and Stefancic respond to the accusation that CRT is a theory of despair by asking “Is medicine pessimistic because it focuses on diseases and traumas?” (2001: 13). Indeed, Delgado turns the accusation on its head and identifies the lie at the heart of liberal perspectives which appear optimistic but disguise the true scale and nature of contemporary racism:

Suppose I am sent to an inner city school to talk to the kids and serve as role model of the month. I am *expected* to tell the kids that if they study hard and stay out of trouble, they can become a law professor like me. That, however, is a very big lie: a whopper. When I started teaching law sixteen years ago, there were about thirty-five Hispanic law professors, approximately twenty-five of which were Chicano. Today, the numbers are only slightly improved (...). Despite this, I am expected to tell forty kids in a crowded, inner city classroom that if they work hard, they can each be among the chosen twenty-five (Delgado, 1991: 1228 original emphasis).

Continuing Debates and Unresolved Issues

CRT is gaining increasing attention but it is by no means a finished and settled set of approaches. CRT is a living and changing perspective, not a monolithic structure. There are, for example, many spin-off movements from traditional CRT, including critical race feminism and LatCrit – a version of CRT that focuses on the particular experiences and struggles of Latina/o communities (see Delgado and Stefancic, 1998; Dixson and Rousseau, 2006; Solórzano and Yosso, 2001; Wing, 1997). Although CRT in the US began with work that often focused on the position of African American communities, it is not the case that CRT adopts (or has ever supported) a simple racial binary perspective that views the world as divided between Whites and a unitary racial Other.

There are many important debates within CRT about the best way of conceiving its work and, in particular, the most effective means of moving things forward through a critical praxis, that is, a combination of theoretical analysis and applied practical strategies of resistance (Lynn and Parker, 2006). Many of these debates raise issues that are relevant to a number of different perspectives and are by no means unique to CRT. For example, there is discussion about the level of group-identification/abstraction that is appropriate for different analytic and political purposes: sometimes it may be best to organize around a collective signifier that includes numerous minoritized groups, while at other times a more specific identity may be preferred (national, linguistic, or religious).

There is a continuing concern within CRT to understand the numerous, complex, and changing ways in which race/racism intersects with other axes of oppression, such as class, gender, disability, and sexuality. This concern with intersectionality is especially strong in critical race feminism (Wing, 1997; Youdell, 2006). Indeed, building on Crenshaw's work, UK scholars Brah and Phoenix (2004) argue that intersectionality itself can provide a useful focus that offers numerous advances on current single-issue thinking. As Crenshaw (1995) argues, rather than viewing intersectionality as a kind of problem to be solved, the best way ahead may be to use intersectionality as a key means of understanding how White supremacy operates and how to mount effective resistance.

Summary

CRT began as an oppositional movement, mostly involving scholars of color, in US law schools in the 1970s and 1980s. The approach has developed rapidly and is now an interdisciplinary perspective with a growing presence in educational studies internationally.

There is no single canonical statement of CRT but certain elements have emerged as central themes that characterize the movement. The first of these is the central role accorded to racism, which is seen as a subtle and pervasive force in society that is so deep rooted as to appear normal to the majority. CRT is also characterized by a critique of liberalism, which points to the failure of notions such as merit, neutrality, and color-blindness which masquerade as fair and just but, because of the uneven playing field of contemporary racist society, they actually function to ensure the continuation of race inequality.

Building on a long tradition of oral histories and subversive storytelling, CRT writers sometimes adopt a narrative approach and, distinctively, they give particular prominence to the experiential knowledge of people of color. This has been a point of controversy with the academic mainstream but reflects CRT's constructivist view of knowledge and its determination to challenge the common-sense assumptions that often encode majoritarian interests.

This deep commitment to promoting real change in the position of minoritized groups is a central tenet of CRT and fuels its disenchantment with traditional notions of civil-rights progress. Critical race scholars draw inspiration from their activist predecessors but they are far from content with the scale of changes that have been won to date. Indeed, a central concept is the interest-convergence principle which notes the benefits to White people at the heart of even the most celebrated civil-rights cases.

This perspective on the enduring nature of racism has led to CRT being condemned by some detractors as a council of despair. Such a reading, however, oversimplifies the approach and fails to appreciate the activist theme that lies at the core of critical race analyses. Similarly, CRT is sometimes caricatured as being concerned with race inequity alone and viewing all white people as uniformly racist and privileged: both these myths reflect a failure to grasp the complexities of CRT which continues to grow and develop in new ways. In particular, the notion of intersectionality is increasingly important, providing a basis for critical race scholars to examine the changing and complex intersections between racism and other axes of oppression, such as class, gender, and disability. As CRT spreads internationally, and generates off-shoot movements such as critical race feminism, it is likely that the approach will generate further interest and controversy.

See also: Critical Theory; Narrative Inquiry.

Bibliography

- Ansley, F. L. (1997). White supremacy (and what we should do about it). In Delgado, R. and Stefancic, J. (eds.) *Critical White Studies: Looking behind the Mirror*, pp 592–595. Philadelphia, PA: Temple University Press.
- Baszile, D. T. (2008). Beyond all reason indeed: The pedagogical promise of critical race testimony. *Race Ethnicity and Education* **11**, 251–265.
- Bell, D. (1992). *Faces at the bottom of the Well: The Permanence of Racism*. New York: Basic Books.
- Bell, D. (2004). *Silent Covenants*. Oxford: Oxford University Press.
- Brah, A. and Phoenix, A. (2004). Ain't I a woman? Revisiting intersectionality. *Journal of International Women's Studies* **5**, 75–86.
- Crenshaw, K. W. (1995). Mapping the margins: Intersectionality, identity politics, and violence against women of color. In Crenshaw, K., Gotanda, N., Peller, G., and Thomas, K. (eds.) *Critical Race Theory: The Key Writings that Formed the Movement*, pp 357–383. New York: New Press.
- Crenshaw, K. W. (2002). The first decade: Critical reflections, or 'a foot in the closing door' *UCLA Law Review* **49**, 1343–1372.
- Crenshaw, K., Gotanda, N., Peller, G., and Thomas, K. (eds.) (1995). *Critical Race Theory: The Key Writings that Formed the Movement*. New York: New Press.
- Delgado, R. (1989). Storytelling for oppositionists and others: A plea for narrative. *Michigan Law Review* **87**, 2411–2441.
- Delgado, R. (1991). Affirmative Action as a majoritarian device: Or, do you really want to be a role model? *Michigan Law Review* **89**, 1222–1231.
- Delgado, R. (1995). *The Rodrigo Chronicles: Conversations about America and Race*. New York: New York University Press.
- Delgado, R. (1999). Rodrigo's committee assignment: A skeptical look at judicial independence. *Southern California Law Review* **72**, 425–454.
- Delgado, R. and Stefancic, J. (eds.) (1998). *The Latino/a Condition: A Critical Reader*. New York: New York University Press.
- Delgado, R. and Stefancic, J. (2000). Introduction. In Delgado, R. and Stefancic, J. (eds.) *Critical Race Theory: The Cutting Edge*, 2nd edn., pp xv–xix. Philadelphia, PA: Temple University Press.
- Delgado, R. and Stefancic, J. (2001). *Critical Race Theory: An Introduction*. New York: New York University Press.
- Dixon, A. D. and Rousseau, C. K. (eds.) (2006). *Critical Race Theory in Education: All God's Children Got a Song*. New York: Routledge.

- Du Bois, W. E. B. (1968). *The Autobiography of W.E.B. Du Bois: A Soliloquy on Viewing My Life from the Last Decade of Its First Century*. New York: International Publishers.
- Dudziak, M. L. (1988). Desegregation as a cold war imperative. *Stanford Law Review* **41**, 61–120.
- Gillborn, D. (2005). Education policy as an act of white supremacy: Whiteness, critical race theory and education reform. *Journal of Education Policy* **20**, 485–505.
- Gillborn, D. (2008). *Racism and Education: Coincidence or Conspiracy?* London: Routledge.
- Harris, C. I. (1993). Whiteness as property. *Harvard Law Review* **106**, 1707–1791.
- Hylton, K. (2008). *'Race' and Sport: Critical Race Theory*. London: Routledge.
- Ladson-Billings, G. (1998). Just what is critical race theory and what's it doing in a nice field like education? *International Journal of Qualitative Studies in Education* **11**, 7–24.
- Ladson-Billings, G. (2004). New directions in multicultural education: Complexities, boundaries, and critical race theory. In Banks, J. A. and Banks, C. A. M. (eds.) *Handbook of Research on Multicultural Education*, 2nd edn., pp 50–65. San Francisco, CA: Jossey-Bass.
- Ladson-Billings, G. (2006). From the achievement gap to the educational debt: Understanding achievement in US schools. *Educational Researcher* **35**, 3–12.
- Ladson-Billings, G. and Tate, W. F. (1995). Toward a critical race theory of education. *Teachers College Record* **97**, 47–68.
- Lynn, M. and Parker, L. (2006). Critical race studies in education: Examining a decade of research on US schools. *Urban Review* **38**, 257–290.
- Macpherson, W. (1999). *The Stephen Lawrence Inquiry*. CM 4262-I. London: The Stationery Office. <http://www.archive.official-documents.co.uk/document/cm42/4262/4262.htm> (accessed August 2009).
- Matsuda, M., Lawrence, C., Delgado, R., and Grenshaw, K. (1993). *Words that Wound: Critical Race Theory, Assaultive Speech, and the First Amendment*. Boulder, CO: Westview.
- McDonald, H. (2003). Exploring possibilities through critical race theory: Exemplary pedagogical practices for Indigenous students. *Paper Prepared for NZARE/ AARE Joint conference*. http://eprints.jcu.edu.au/951/01/mcdonald_1.pdf (accessed August 2009).
- McIntosh, P. (1992). White privilege and male privilege: A personal account of coming to see correspondences through work in women's studies. In Delgado, R. and Stefancic, J. (eds.) *Critical White Studies: Looking behind the Mirror*, pp 291–299. Philadelphia, PA: Temple University Press.
- Mills, C. W. (2003). *From Class to Race: Essays in White Marxism and Black Radicalism*. New York: Rowman and Littlefield.
- Moreton-Robinson, A. (2004). The possessive logic of patriarchal white sovereignty: The High Court and the Yorta Yorta decision. *Borderlands e-journal* **3**. <http://eprints.qut.edu.au> (accessed August 2009).
- Solórzano, D. and Yosso, T. (2001). Critical race and LatCrit theory and method. *International Journal of Qualitative Studies in Education* **14**, 471–495.
- Stovall, D. (2006). Forging community in race and class: Critical race theory and the quest for social justice in education. *Race Ethnicity and Education* **9**, 243–259.
- Tate, W. F. (1997). Critical race theory and education: History, theory, and implications. In Apple, M. W. (ed.) *Review of Research in Education*, pp 195–247. Washington, DC: American Educational Research Association.
- Tomlinson, S. (2008). *Race and Education: Policy and Politics in Britain*. Maidenhead: Open University Press.
- West, C. (1995). Foreword. In Crenshaw, K., Gotanda, N., Peller, G., and Thomas, K. (eds.) *Critical Race Theory: The Key Writings that Formed the Movement*, pp xi–xii. New York: New Press.
- Wing, A. K. (ed.) (1997). *Critical Race Feminism: A Reader*. New York: New York University Press.
- Youdell, D. (2006). *Impossible Bodies, Impossible Selves: Exclusions and Student Subjectivities*. Dordrecht: Springer.

Further Reading

- Ladson-Billings, G. (ed.) (2003). *Critical Race Theory Perspectives on Social Studies: The Profession, Policies, and Curriculum*. Charlotte, NC: Information Age Publishing.
- Parker, L., Deyhle, D., and Villenas, S. (eds.) (1999). *Race Is ... Race Isn't: Critical Race Theory and Qualitative Studies in Education*. Oxford: Westview.
- Taylor, E., Gillborn, D., and Ladson-Billings, G. (eds.) (2009). *Foundations of Critical Race Theory in Education*. New York: Routledge.

Critical Theory

J J Ryoo and P McLaren, University of California, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Critical pedagogy – A teaching approach that brings together critical theory and educational practice in order to help students question and challenge dominant beliefs and practices for higher social, critical consciousness.

Dialectic – Initially conceptualized as a Socratic technique of employing discussion, reasoning, and, therefore, dialog to expose false beliefs, Hegel applied the dialectic to mean a critical investigation of the process of change resulting in the theoretical application of such a process to the social sciences.

Domination – The exercise of ruling power with supremacy or preeminence over another.

Exploitation – The act of making use of another in an unfair or mean manner for one's own advantage.

Inquiry – A systematic investigation and examination of facts or principles.

Praxis – A union of action and reflection and of theory and practice; critical praxis in education refers to a threefold process of self-reflection, reflective action, and collective reflective action for social change.

Critical Theory: An Overview

Critical theory has a long history of informing and shaping diverse qualitative research practices. Such a theoretical lens – both for educational researchers and otherwise – has proven particularly valuable in – first, challenging qualitative researchers to question more positivist approaches that can inaccurately assume scientific neutrality among objective observers while, second, engaging researchers in more reflexive research practices that question the biases from which they view the world. Through using critical theory, qualitative researchers can push educational reform for the better of all humanity.

Although it is difficult to point to one, universally accepted critical theory that has been unproblematically embraced by all critical philosophers, one could still recognize a shared notion of critical theory as referring to both the legacy of Western theoretical traditions following the Frankfurt School, as well as an evolving school of thought and process of critique.

As a philosophy within the social sciences, critical theory refers to the prolific work of several generations of German thinkers belonging to Western European and Marxist traditions who can be broadly categorized as members of the Frankfurt School. Philosophers of the Frankfurt School moved away from traditional theory by seeking to radically reconstruct the meaning of human emancipation through what Max Horkheimer has described as “liberate[ing] human beings from the circumstances that enslave them” (Horkheimer, 1972: 244). Critical theory is rooted in historicizing, critiquing, and exposing the relationships of domination and subordination as well as the contradictions in which humankind is entrenched and thus, in essence, is a liberatory philosophy. Such critique focuses on deconstructing how knowledge is produced, whose knowledge is valued, and how control of such knowledge equates to power in society, with the purpose of preventing people from becoming slaves to the ideas and activities that societal institutions push upon them. As a result of this search for a theory that – through critique of social constructs – could articulate freedom for all, while arguing against the suppression of human consciousness, multiple and diverse critical theories have developed since the birth of the Frankfurt School, including (but not limited to): critical race theory, critical pedagogy, critical literacy, critical feminist theory, etc. (see the section titled Global Critical Philosophies).

Viewed as a process of critique or methodology, critical theory may also be defined as a type of self-conscious means of addressing the need for dialog and a dialectic approach around transformative, emancipatory practice. Such a process of critique pushes theorists to continuously question how they look at – and depict – the world in comparison to how the world actually exists, and, as a result, challenges theorists to question the very theories they form through such a process. According to critical theory, when a philosophy is embraced without question and fails to problematize wider society, it has become an ideology in need of critique. The Frankfurt School valued the use of such inquiry or critique as a tool articulating praxis – therefore, engaging a relationship between theory and society.

A Brief History of Critical Theory and the Frankfurt School

The best way to understand critical theory is to first locate it within Western philosophy's historical trajectory.

In doing so, one will find that critical explorations of the nature of knowledge production that form the framework of critical theory today can be traced back as far as 2500 years before the first critical scholars.

Classical Philosophers

A critical approach to understanding the relationship between knowledge and power in society can be found in the work of the Greek philosophers Socrates (469–399 BCE), Plato (428/427–348/347 BCE), and Aristotle (384–322 BCE). Through Plato's dialogues – most notably *Gorgias* and *The Apology* – one can see how Socrates (Plato's mentor) questioned the way the Sophists (classical Greek philosophers who exercised great power in Athens) used rhetoric to persuade people to manufacture and believe in false truths, rather than use rhetoric to search for ultimate truth for the better of humankind. Similarly, Plato's determination to record Socrates' stories through the written word – despite Socrates' skepticism about the technology of writing – reflects the ways in which Plato recognized the importance of language as a meaning-making tool for the development of critical thought. Following Plato, Aristotle theorized and deconstructed the power of knowledge creation in language manipulation through *The Art of Rhetoric* and *Poetics*. Aristotle outlines the process of writing persuasive speeches that appeal to *logos* (reason), *pathos* (emotion), and *ethos* (confidence in the orator's personal character). Although thousands of years would pass between the lives of these Greek philosophers and the foundations of critical theory which grew out of the Enlightenment, the examination of linguistic tools in knowledge production that explored in Socrates, Plato, and Aristotle's work proved invaluable to Western philosophy traditions.

Kant, Hegel, Marx, and Modern Foundations of Critical Philosophy

Following the Middle Ages – that marked the hegemony of the Catholic Church – and the Renaissance, that marked the break from this Church's control, came the birth of the Enlightenment in Europe and America. During this era, Immanuel Kant (1721–1804) highlighted the value of critique as the only way to find true knowledge and combat illusory ideologies by separating the self from the dogmas surrounding us and by questioning those concepts that are presented to us as facts. Georg Wilhelm Friedrich Hegel (1770–1831) also theorized about knowledge production but wished to push beyond Kant's separation of mind and self when deconstructing ideologies. Hegel, instead, sought to historicize the process of arriving at truth. Hegel's approach suggested that history could be viewed as progress toward a greater humanity – and Utopia – through a higher consciousness of self in and

with the world. Hegel's great contribution to the foundations of critical theory exists in his use of dialectic reasoning to reach perfection, which involves engaging in dialog to overcome the contradictions between thesis and antithesis toward a synthesis that would, in turn, be challenged repeatedly until perfect synthesis (Utopia) could be reached.

Following Hegel, Karl Marx's (1818–83) work with Friedrich Engels with regard to the historical and material conditions of capitalism, is recognized as the most influential to shaping critical theory. Although seen, primarily, as an economic philosopher, Marx's elucidation of how the proletariat (working class) has been alienated and exploited by the bourgeoisie (ruling class) – who benefit through the accumulation of wealth – helped develop an understanding of how ideologies reinforce false consciousness that can be used to maintain power in the hands of those in control. Thus, Marx provided the basic idea on which critical theory is constructed: those who have economic power in society also control the cultural superstructure of that society. Marx recognized how production and consumption of media (through language) controlled class consciousness and how, simultaneously, such a tool could also be used against oppressive capitalist ideologies for a better humanity.

The Frankfurt School

The original home of the Frankfurt School was in the Institute for Social Research (Das Institut für Sozialforschung) that was originally founded in Frankfurt, Germany, in February 1923, but was forced to move under Nazi threat, in 1933. In 1930, the Institute came under the directorship of Max Horkheimer and grew to include critical theorists such as Erich Fromm, Herbert Marcuse, Theodor Adorno, and Walter Benjamin who shifted its theoretical focus from an analysis of bourgeois society's socioeconomic substructures to bourgeois cultural superstructures. Although primarily influenced by Marxism, the Frankfurt School rejected an orthodox reading of Marx and represented a different form of Marxism than their Bolshevik contemporaries, as they did not envision an inevitable proletariat revolution nor focus solely on economic production. Instead, the Frankfurt School sought to understand the relationship between theory and society through a dialectical framework that also examined the realm of culture “including not only the so-called spiritual contents of science, art, and religion, but also law, ethics, fashion, public opinion, sport, amusement, life style, etc.” (Horkheimer, 1972). Thus, the Frankfurt School became both a representation of Western Marxism and a critique of it.

Central to the Frankfurt School's articulation of critical theory was its critique of the Enlightenment's conceptualization of instrumental reason, rationality, and positivism (as explored by diverse thinkers including

Saint-Simon, Comte, the Vienna Circle, and others). The Frankfurt School believed there was a breakdown in reason due to a “crisis of science” that reduced the concept of knowledge to a limited scientific methodology of “description, classification, and generalization of phenomena, with no care to distinguish the unimportant from the essential” that could only be resolved through the creation of “a correct theory of the present social situation” (Horkheimer, 1972). Echoing Nietzsche’s belief that it was not science but the scientific method that proved triumphant during the nineteenth century, the Frankfurt School sought to rescue the world from the basic assumptions of logical positivism – that valued objective observation as the only way to produce truth – by envisioning critical theory as an improved approach to human reasoning, supporting a more self-conscious critique of human action for transformation. The Frankfurt School believed that positivist rationality and its narrow view of science endangered human ability to think critically and subjectively with regard to the world – freezing human history behind selected, organized, and fetishized facts in false-value neutrality, rather than recognizing how natural reality and science relate both to human consciousness and lack of consciousness as well as subjectivity and objectivity positions. Furthermore, the Frankfurt School articulated how those engaged in using positivist rationality were not necessarily conscious of the ways in which such an approach could maintain an oppressive power imbalance or the *status quo* in work labeled as scientific, and thus urged philosophers to consider their positionality as producers of knowledge.

In order to address these problems with logical positivity and Enlightenment rationality, the Frankfurt School formulated a critical theory that first recognized the relationship between individuals and society as a whole. In this way, critical theory develops a capacity for meta-theory by acknowledging the cultural–historical positionality of interested, non-neutral parties whose political or social interests affect how they navigate and present knowledge or information. With such a self-critical and self-conscious lens, critical theory engages in dialectical thought that lives within the tension between critique and theoretical reconstruction. As noted by Marcuse (1960), “Dialectical thought starts with the experience that the world is unfree; that is to say, man and nature exist in conditions of alienation, exist as ‘other than they are.’ . . . Thought ‘corresponds’ to reality only as it transforms reality by comprehending its contradictory structure.” Dialectical methodologies synthesize philosophy, the social sciences, and radical politics to help illustrate how society, state, economy, culture, and the individual are all related to one another. Thus, critical theory should work towards an improved society devoid of injustice and oppression and “emancipation from slavery” (Horkheimer, 1972) through equal engagement of theory and practice in critical praxis.

The Frankfurt School also used critical theory to reject essentializing and hierarchical perspectives of culture that were supported by sociologists and orthodox Marxists of the time. The Frankfurt School rejected definitions of culture as a stagnant, objectified ideology, and, instead, embraced culture’s processual and politically and economically situated role in the development of human relationships over time. Such was visible in the work of Horkheimer and Adorno.

More specifically, Theodor Adorno (1903–69) powerfully shaped the future of cultural theory through his work in the Frankfurt School by recognizing how major culture industries used popular culture – such as film and radio – to advance the oppressive ideologies of the dominant, capitalist society. Through Adorno’s work, critical theory grew to question the dehumanizing relationship between media culture and its consumers, while simultaneously recognizing the power of using new media literacies against such control. In *One Dimensional Man* (Marcuse, 1964), Herbert Marcuse (1898–1979) critiques industrialism’s oppression of individuals for the maintenance of systems of production. Through this work, Marcuse pushed critical theory to demonstrate how individuals are completely controlled by a system of production and consumption. In other work, Marcuse also helps to elucidate the purpose of dialectical thought as a connection between thought and action in *Reason and Revolution* (Marcuse, 1960), where he writes how “Thought ‘corresponds’ to reality only as it transforms reality by comprehending its contradictory structure . . . to break down the self-assurance and self-contentment of common sense, to undermine the sinister confidence in the power and language of facts.” Max Horkheimer’s (1895–1973) work focused primarily on the function of reason within philosophy and the creation of oppressive ideologies in larger society which proved particularly significant during a period in which the Nazis were using evil and dehumanizing methods to manipulate society to engage in the mass murder and torture of Jewish people. Horkheimer used critical theory as a tool to fight such destructive mass ideologies in a rapidly changing world. Habermas’ contributed to critical theory in his development of the theory of communicative reason or communicative rationality couched in discourse ethics and idea exchange. Walter Benjamin (1892–1940) engaged in critiques of literary and artistic work – building on aesthetic theory while pushing critical theory to discuss the realm of culture and its historical context. Thus, critical theory was created to give people the agency to challenge taken-for-granted assumptions through tools such as dialectical reasoning and praxis that promote free-thinking and self-determination for all individuals in society. As elucidated by Horkheimer, a truly critical theory “has as its object human beings as producers of their own historical form of life” (Horkheimer, 1993: 21).

Global Critical Philosophies

Although critical theory itself was born of a Western and European philosophical history, other great philosophies have also developed worldwide that should be recognized as part of the critical theory canon in its current, living form. Some such traditions have borrowed from Western philosophies while others have developed more independently and, while not all such work has been unearthed due to language barriers and imbalanced power relationships between Western imperialists and scholars of other nations, an attempt will be made here to bring some of these global critical philosophies to the mainstream of Eurocentric critical thought.

Within the African American tradition, one can see important developments in critical theory through the work of W.E.B. DuBois (1868–1963) in his famous work – *The Souls of Black Folk* (originally published in 1903, DuBois, 1903). Through his deep exploration of race in America, DuBois theorizes about “double consciousness” experienced in the African American who sees herself/himself through the gaze of those who hate people of color, and describes the torment this places on the spirits of all people in a racist society. DuBois recognizes how ideologies of race are formed through dominant languages, but how a critical approach to education could help transcend such oppressive ideologies for social change. Following the Harlem Renaissance and the growth of amazing art and literary works that helped engage humanity in a critique of America’s social structures and institutions, Carter G. Woodson (1875–1950) published *The Miseducation of the Negro* (Woodson, 1933) which challenged the ways classical education incorporated forms of scientific racism by revising world histories to enforce ideas about the racial inferiority of people of color. By removing African Americans and African peoples from school knowledge with regard to history and science, classical education dehumanized African Americans and continued to oppress and mold them to the racist, hierarchical American fabric. Woodson calls for a critical education in which people of color could question the schooling they receive for individual and community liberation. Such works not only influenced innumerable educators, scholars, artists, and activists toward a critique of society for the purpose of social change as is visible in the work of Frederick Douglass, Langston Hughes, Marcus Garvey, Malcolm X, and the Black Panther Movement – all of who engaged in challenging how knowledge is produced and used to manipulate truth in a racially hierarchical society.

During the 1980s, other great scholars – such as Edward Said and Homi Bhabha – pushed critical theorists to think in new ways by developing studies in colonialism and postcolonialism. Recognizing how Western colonialism maintained its hegemony through the control of knowledge and knowledge production, postcolonial

theorists engage critical theory in ways most applicable to the current reality of globalization. Although often associated with postmodern theory in its methodology under the influence of Jacques Derrida’s concepts of deconstruction, Jacques Lacan’s psychoanalysis, and Michel Foucault’s discourse analysis, postcolonial theory still relates to critical theory in the ways that colonized, indigenous peoples have found agency in critiquing the oppressive societies in which they live. In *Orientalism* (Said, 1978), Said critiques the ways in which the West continues to romanticize Asian and Middle Eastern nations as the archaic and unchanging ‘other’, while using such false ideologies to justify continued imperialism throughout the Orient. Similarly, in *Nation and Narration* (Bhabha, 1990), Bhabha problematizes how postcolonial countries continue to be treated as one, homogeneous group by their ex-colonizers, with concepts of nationhood narrativized through power imbalances between ex-colonizing nations and their ex-colonial states.

Frantz Fanon (1925–61) made immeasurable contribution to the furthering of critical theories through his own work in Africa focusing on Algerian resistance to French rule. Believed to be influenced by the 1930s Négritude movement – that rejected French colonial racism and was led by Martinican poet Aimé Césaire, Guianan Léon Damas, and future Senegalese President Léopold Sédar Senghor – Fanon explored how the psychology of oppression and the impact of colonialism affected the mental health of colonized people in works like *A Dying Colonialism* (Fanon, 1959) or *Black Skin, White Masks* (Fanon, 1967). By critiquing the way colonial powers control colonial ideology production and dissemination and how media is used to manipulate colonized people into believing in Western racial hierarchies of language, custom, or skin color, Fanon illustrates how indigenous peoples can deconstruct dominant language and texts toward ideology change for liberation. Fanon’s work highly influenced the revolutionaries of other nations – most notably Fidel Castro in Cuba, Paulo Freire in Brazil, and the Black Panthers in the United States.

Of course, Latin American and Central American development of new critical theories can be traced through the work of Liberation Theology that united indigenous people with Catholic clergy against the oppressive, racist structures of the European Catholic church, or in the work of Ernesto Ché Guevara – Minister of Education in post-revolution Cuba and organizer of the guerrilla mountain schools – but the most notable development of critical theory can be found in the work of Paulo Freire (1921–97), in Brazil. Freire’s most famous work – *The Pedagogy of the Oppressed* – explores the ways the educational system proves central to the dehumanizing processes of capitalist and neocolonial societies through banking education that treats students as empty vessels that passively collect the oppressive norms and knowledge of teachers. Freire focused on literacy and literacy education while teaching illiterate adults how to

interrogate and deconstruct dominant texts within a racist, classist Brazil. Freire believed in the power of praxis and dialog to push critical thought against oppressive forces that control schooling structures.

While critical theory has surely influenced revolutionaries, scholars, and activists throughout East Asia, the continued imbalance of power relationships that maintain hierarchies with regard to race, class, color, and language in the Western world continues to negate and oppress East Asian critical scholarship. Little of East Asian critical theory has been translated and shared with Western academia and this issue should be immediately addressed.

Critical Theory in Education

The critical theory of the Frankfurt School and other global critical theories prove crucial to educational reform and can be used to revolutionize today's schools toward a more just, humanizing, and democratic system. First of all, the ways in which critical theory challenges positivist rationality that assumes political neutrality in educational spaces is invaluable to opening a space for educators to recognize the hidden curricula of current standardized schooling. Second, critical theory values historical consciousness as fundamental to engaging in critical thought and critical literacy. Recognizing the sociohistorical framework in which knowledge is produced and the pedagogical tools used by teachers in today's schools proves essential not only to deconstructing what students are learning, but also how students are learning. Third, engaging in historical consciousness through critical theory in education simultaneously forces us to re-mediate the ways in which educators and administrators of schools view race, class, gender, language, sexuality, and culture. Critical theory perspectives allow for people to engage in an informed critique toward reforming education so that it treats each student with respect toward the funds of knowledge and perspectives from which she/he comes without essentializing or reifying her/his experiences. Fourth, critical theory can be usefully applied to education through the use of dialectical thought and praxis which pushes teachers and students alike to actively use social inquiry to question oppressive, dominant structures in society while moving toward a more liberatory practice. Fifth, critical theory in education believes in the power of schooling as a tool for social justice by engaging in the needs of students' home communities. Thus, using critical theory not only as a means of meaning-making, but also as a means of connecting diverse students and teachers to each other in shared practices through education can build a new community of learners who can change the world. As recognized by Freire and Macedo (1987), critical theory can be engaged in education through critical praxis that involves "reading

the world . . . through reading the word" (p. 35) as students gain new critical literacies by being taught to engage in questioning social relations existent in society.

Following this critical theory tradition, developers of critical pedagogy – including, but not limited to, Paulo Freire, Peter McLaren, Ernest Morrell, Henry Giroux, Ira Shor, Antonia Darder, bell hooks, Joe Kincheloe, Paula Allman, Donald Macedo, and Maxine Greene – encourage educators to apply the ideas of critical theory to schooling structures in ways that revolutionize the ways we think about both the purpose and practices of education. Although diverse in its applications to schooling, critical pedagogy challenges the conservative policies and practices of banking education that – in the current era – are tied to standardized testing and curricula in a classist, racist, sexist, and heterosexist world. By embracing the power of student, educator, and researcher agency within an oppressive American school system, critical pedagogy grounds social justice education using tools of critique that question how knowledge is produced, controlled, and disseminated. Critical pedagogy proves a valuable example of how critical theory can be applied to education. Critical revolutionary pedagogy (see the work of Peter McLaren, Paula Allman, Glenn Rikowski, Mike Cole, and Dave Hill) has also been developed in opposition to more domesticated and sanitized versions of critical pedagogy and critical theory found in educational reform efforts that do not contest neoliberal capitalism, and attempts to unite critical pedagogy and classical Marxist theory.

Critical theory pushes educators and educational researchers to rethink the power dynamics that both influence and work within schooling systems and, as explored by Antonio Gramsci (1891–1937) – who believed that all humans were capable of intellectual and rational faculties – critical theory can help us recognize the ways in which education, similar to media, is an ideological apparatus of the state that reproduces the hegemony of those in control. Thus, educational researchers can find agency through critical theory to engage in social inquiry that transforms the space of schools, the practice of teachers, the treatment of students, the sharing of diverse knowledge, and the improvement of society at large.

See also: Critical Ethnography; Critical Race Theory; Discourse Analysis; Hermeneutics.

Bibliography

- Bhabha, H. (ed.) (1990). *Nation and Narration*. New York: Routledge.
 DuBois, W. E. B. (1903/1982). *The Souls of Black Folk*. New York: Signet.
 Fanon, F. (1959). *A Dying Colonialism*. New York: Grove Press.
 Fanon, F. (1967). *Black Skin, White Masks*. New York: Grove Press.

- Freire, P. and Macedo, D. (1987). *Literacy: Reading the Word and the World*. New York: Continuum.
- Horkheimer, M. (1972). *Critical Theory*. New York: Seabury Press.
- Horkheimer, M. (1993). *Between Philosophy and Social Science*. Cambridge: MIT Press.
- Marcuse, H. (1960). *Reason and Revolution*. Boston, MA: Beacon Press.
- Marcuse, H. (1964). *One Dimensional Man*. Boston, MA: Beacon Press.
- Said, E. (1978). *Orientalism*. New York: Vintage.
- Woodson, C. G. (1933/1990). *The Miseducation of the Negro*. Washington, DC: Africa World Press.

Further Reading

- Adorno, T. (1987). *Negative Dialectic*. New York: Continuum.
- Bourdieu, P. and Passeron, J. (1992). *Reproduction in Education, Society and Culture*. London: Sage.
- Darder, A., Baltodano, M., and Torres, R. D. (eds.) (2003). *The Critical Pedagogy Reader*. New York: RoutledgeFalmer.
- Dubiel, H. (1985). *Theory and Politics: Studies in the Development of Critical Theory*. Cambridge, MA: MIT.
- Freire, P. (1970). *Pedagogy of the Oppressed*. New York: Continuum.
- Freire, P. (1994). *Pedagogy of the Oppressed*, anniversary edn. New York: Continuum.
- Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings*. New York: Pantheon.
- Giroux, H. (1983). *Theory and Resistance in Education*. South Hadley, MA: Bergin and Garvey.
- Green, M. (1988). *The Dialectics of Freedom*. New York: Teachers College.

- Held, D. (1980). *Introduction to Critical Theory: Horkheimer to Habermas*. Berkeley, CA: University of California Press.
- Horkheimer, M. (1972). *Critical Theory: Selected Essays*. New York: Herder.
- Kellner, D. (1989). *Critical Theory, Marxism, and Modernity*. Baltimore, MD: Polity and John Hopkins University Press.
- Marcuse, H. (1966). *One Dimensional Man*. Boston: Beacon.
- Marx, K. (1969). *Early Writings*. New York: McGraw-Hill.
- McLaren, P. and Giarelli, J. (1995). *Critical Theory and Educational Research*. New York: SUNY Press.
- Morrell, E. (2008). *Critical Literacy and Urban Youth: Pedagogies of access, dissent, and liberation*. New York: Routledge.
- Rasmussen, D. (ed.) (1996). *The Handbook of Critical Theory*. Oxford: Blackwell.
- Ryder, M. (2009). Contemporary Philosophy, Critical Theory and Postmodern Thought. http://carbon.cudenver.edu/~mryder/itc_data/postmodern.html#resources (accessed September 2009).
- Sim, S. and Van Loon, B. (2001). *Introducing Critical Theory*. Royston, UK: Icon Books.
- Wiggershaus, R. (1994). *The Frankfurt School*. Cambridge, UK: Polity Press.

Relevant Websites

- <http://www.humanities.uci.edu> – Critical Theory Institute, University of California, Irvine.
- <http://www.uta.edu> – Illuminations, The University of Texas at Arlington.
- <http://www.marxists.org> – Marxists Internet Archive.

Cultural Studies

R Gaztambide-Fernández and K Cairns, Ontario Institute for Studies in Education, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Critical media literacy – A branch of media education designed to foster a critical approach to popular texts, with an emphasis on how media images connect to broader cultural understandings. In addition to developing techniques of interpretation, critical media literacy approaches cultural production as a valuable site of learning.

Ethnography – Methodology through which anthropologists and some sociologists study cultural phenomena. It involves prolonged periods of observation and participation within a particular cultural context.

Frankfurt School – Originally an institute of social research founded in 1923, led by prominent Marxist scholars Theodor Adorno, Walter Benjamin, Max Horkheimer, and Herbert Marcuse. Their work inspired the evolution of ‘critical theory,’ and is best known for critiques of ideology and industries of mass culture.

Hegemony – Concept developed by Italian scholar Antonio Gramsci to explain how unequal relations of power are maintained through popular consent and taken-for-granted assumptions. Hegemony refers to the process whereby members of a society participate in reproduction of dominations through the assumptions of ‘common sense’ values that benefit those who have power.

Hybridity – In cultural theory, hybridity refers to the process through which two or more cultural forms come together to create a new form of cultural practice. It is often associated with cultural practices that emerge from the encounter of multiple groups through processes of migration, globalization, and dissemination of cultural forms through media.

Postcolonial theory – A theoretical approach that seeks to counter the racism that underlies the systems of knowledge and cultural production that support colonial or imperial conquest. It calls into attention the ways in which both scientific knowledge as well as cultural production are employed as tools of conquest and oppression, as well as the ways in which colonized communities respond to conquest through complicated cultural practices.

Queer theory – Building on the notion that both gender and sexuality are social constructs that require the daily practices – or ‘performances’ – of

individuals to come into being, it draws on feminist and psychoanalytic theories to explore questions of embodiment, desire, and the ways in which sexual subjectivities are shaped by, while also resisting, power dynamics.

Tyler rationale – First articulated by Ralph Tyler in 1949, the Tyler rationale refers to the dominant approach of curriculum development based on a technical view of learning and teaching. It is based on four key steps: defining learning goals, identifying experiences that will yield those goals, organizing those experiences, and assessing how well the goals were met.

Cultural studies developed as a scholarly field over the second-half of the twentieth century. There are multiple ways of defining cultural studies, in part because it has developed into different streams of scholarship influenced by an array of academic disciplines and theoretical frameworks. Cultural studies scholars engage a varied range of cultural phenomena within many fields of practice – including education, mass media, science, literature, film, consumer culture, architecture, leisure, corporate culture, and the academy, to name a few. These fields are characterized by a range of cultural practices involving different institutional arrangements, discourses, and material as well as symbolic resources that to some extent or another shape the experiences of those who participate in these fields. In turn, individuals and groups engage these cultural practices in ways that often support, sometimes resist, and others redefine these arrangements. In general terms, cultural studies scholars are concerned with understanding these cultural practices in contemporary societies and the ways individuals engage and make meaning of their relationships with others and with the institutions that mediate those relationships.

Cultural studies is interdisciplinary in scope, bridging sociology, anthropology, literature, philosophy, linguistics, and political science. Cultural studies scholars draw on theoretical frameworks from critical theory, postmodernism, feminism, psychoanalysis, and, more recently, from queer and postcolonial theory. Each of these frameworks provides a different lens through which to understand the relationship between individual experience and the various structures and institutions around which cultural activity is organized.

During the early part of the twentieth century, academics as well as community-based and political organizers around the world turned their attention to culture as the locus of social and political change (Wright, 2004). For community and political organizations, this meant engaging their communities through cultural practices in an attempt to challenge oppressive social and political institutions by revealing their otherwise hidden cultural logics. For academics, this meant turning to the analysis of the cultural lives of various groups in cotemporary society. As Wright (2004) has argued, cultural studies as a field emerged out of the confluence of these social and academic movements in their efforts to recenter cultural politics.

Despite these worldwide movements, most accounts of the emergence of cultural studies credit the work of British scholars, such as Richard Hoggart, Raymond Williams, and E.P. Thompson, who were primarily interested in the cultural lives of the British working class. Their work led to the establishment of the Birmingham Centre for Cultural Studies (BCCS) in 1964, thus naming a field that had been evolving for over 50 years in other parts of the world. BCCS scholars rejected the two opposing views of culture that predominated academic work; the first drawn from anthropology, which looked for culture in the ways of living of non-European societies, and the second drawn from fine arts and literary criticism, which looked for culture in what were deemed the great accomplishments of European civilization. Instead, these scholars turned to the Marxist insight that there is a relationship between social structures and cultural practice, while rejecting the Marxist tenet that social structures overdetermine or predefine cultural experience.

Drawing on the intellectual legacy of Italian Marxist Antonio Gramsci, these early cultural studies scholars viewed culture as the dynamic interplay between social structures and the everyday practices of individuals. For Thompson (1963), for example, the working class was not an ahistorical structural formation, but rather, it came alive through the cultural practices of individuals. It is in and through these cultural practices that the possibility for social transformation resides. For cultural studies scholars, then, cultural analysis is not simply about understanding culture, but about fomenting social change through the transformation of cultural practice.

Against the essentialist view of culture as an inherent characteristic of different groups, cultural studies views culture as a practice that is always contextual – always negotiated within specific relationships and under specific circumstances – and not always predictable. Schools are one of those institutional contexts in which cultural experience evolves and where various actors engage each other through cultural negotiation. Indeed, cultural studies raises crucial questions about the relationship between schooling and society. Cultural studies scholars have made important contributions to our understandings of cultural processes in

school settings as well as in nonformal educational contexts. In particular, youth and youth cultural practice has been one of the central foci of cultural studies. Cultural studies has held an important place in certain areas of educational scholarship, particularly within curriculum studies, critical pedagogy, critical multicultural education, language and culture, and media education.

However, the fields of cultural and educational studies remain in tension with each other (Gaztambide-Fernández *et al.*, 2004). While a few education scholars have embraced cultural studies as a foundation of their work, cultural studies remains largely marginal to contemporary thinking about education in general and schools in particular. One reason for this marginality is that insights from cultural studies challenge some of the fundamental assumptions made about education. Here, we introduce the field of cultural studies through an exploration of three ways in which it has challenged dominant views of education. These challenges revolve around different (1) approaches to institutions, (2) conceptions of youth, and (3) relationships to social and political change. For each of these, we first discuss what cultural studies has contributed to this particular dimension. Second, we discuss how these contributions exist in tension with how education has traditionally framed the issue. Third, we argue that rather than obstacles, these tensions present opportunities for deepening our understanding of the relationship between culture and education. Our discussion of the opportunities that a cultural studies lens opens for educational thinking are premised on the marginality of cultural studies within educational discourse while also drawing on the work of education scholars that explore these opportunities.

Framing Schools

Until the 1960s, most sociologists framed schools as black boxes and studied them only on the basis of the relationship between inputs, such as school resources and family background, and outputs, mainly academic achievement. Challenging this model, scholars, informed by Marxist cultural theory, argued that structures and practices within schools played a crucial role in the reproduction of economic and social inequality. As they explored these internal mechanisms, they paid little attention to how students themselves participated and made sense of their experience within schools.

In 1977, building on the work of the BCCS on youth culture, Paul Willis turned his attention to the lives of youth at a British working-class school (Willis, 1977). Drawing on a long tradition of ethnographic fieldwork, Willis was interested in how young people's daily cultural practices were related to social class reproduction. In his book *Learning to Labour*, Willis reveals that it is not only (or mainly) school structures and practices that lead working-class kids to

working-class jobs. Rather, he describes the complicated ways in which working-class youth resist schooling and how this resistance is implicated in the apparently inevitable outcome of class reproduction. It is not through compliance, but through resistance that working-class youth develop and cultivate the cultural dispositions necessary for factory work.

The scholarship inspired by Willis' work demonstrates that schools are not simply sites of social reproduction; rather, cultural processes yield certain unpredictable outcomes that challenge the black-box model of schooling. This framing of schools, as the settings in which cultural production yields specific economic arrangements, has predominated within cultural studies scholarship. At the same time, Willis' work has been criticized for the way it positioned young women and its lack of attention to sexism in the lives of working-class youth. McRobbie (1991), whose early work focused on how young women engaged with magazines, brought a feminist lens to the analysis of youth culture. Other scholars, such as Kobena Mercer and Paul Gilroy, brought a focus on racism and racial politics to cultural analysis. It is in this context of analytic expansion into matters of race and gender that cultural studies traveled to the United States, where scholars such as Wendy Luttrell, Cameron McCarthy, and Kevin Kumashiro, have built on these initial insights to explore how dynamics of race, gender, and sexuality also play out and intersect in the context of schools.

The focus on cultural practice is in tension with the notion that schools are the great equalizer, a place where the playing field is leveled and where only hard work and aptitude matter in determining outcomes and achievement. Education scholars have largely framed schools as, at least, benign institutions that socialize students into their proper place and, at most, sites for the potential transformation of society and social change. Cultural studies scholars, on the other hand, have explored schools as sites where power dynamics operate to produce and sustain particular inequalities. Underlying these different approaches to schooling are distinct conceptions of the notion of culture. Educational studies tend to presume that culture preexists schools and that through curriculum and pedagogy established social patterns can be transformed into more equitable ones. In contrast, cultural studies suggests that culture happens in schools and that the institutional arrangements of schools are implicated in the reproduction of inequality.

One way in which this tension plays out is around the question of multicultural education. Conventional educational approaches to multiculturalism are often framed simplistically as a celebration of diversity. From this perspective, multicultural education involves learning about the food, dress, and customs of various ethnic groups, with the goal of fostering openness to differences among students. A cultural studies approach shares this interest in

difference, but rather than taking cultural identifications for granted, it asks how understandings of difference are produced. This approach engages culture not as a predetermined attribute, but rather as a set of practices through which individuals make sense of themselves in relation to those around them. Critical of educational approaches that view culture as a discrete identity, cultural studies scholars have drawn on postcolonial theory and the notion of hybridity in order to explore the multiple and complex identifications that students perform in schools. Even as these scholars embrace possibilities for fluidity and agency in students' identity constructions, they continue to foreground relations of power. As a result, a cultural studies approach to multicultural education explores how racism, sexism, and other forms of oppression shape students' schooling experiences (Carlson and Dimitriadis, 2003).

Cultural studies does not deny the possibility that schools can serve as a site of social transformation, but it does challenge scholars of education to examine the power relations that structure this practice and to examine the complex relations of power and resistance. Thus, the tension that emerges from these two conceptions of schooling provides an opportunity for exploring schools as complex spaces, where inequalities are produced alongside possibilities for social justice. Cultural studies foregrounds the work of cultural processes, as scholars attend to the ways in which notions of identity and difference are relationally constructed, rather than the product of some predefined attribute. Cultural studies brings to studies of education a more fluid notion of what culture means, and views schools as places where boundaries are negotiated and redefined, and where the possibility of social transformation is opened, but not guaranteed. Schools have the potential of being the site of cultural transformation, even as they are also primarily the site of social reproduction. Reflecting the generative nature of this tension, this complex conception of schooling has prompted scholars to reconsider conventional approaches to youth, shifting from the notion of behavior to that of practice in order to capture the fluidity of cultural processes.

Framing Youth

Educational scholars have historically framed youth as a developmental stage. Early educational psychologists examined adolescence as a period of personal growth on the path to adulthood. This developmental conception of youth has led to students being framed as the objects of pedagogy. By contrast, cultural studies attends to youth not as a period of psychological development, but rather as active producers of meaning. Scholars in this field have foregrounded the question of agency within youth studies and have sought to explore the extent to which young

people's actions can be understood as responses to certain social circumstances rather than a manifestation of psychological development.

Early cultural studies research at the BCCS focused on the notion of resistance, by examining how groups of young people engage with and contest relations of cultural hegemony and the demands of economic capital (Hall and Jefferson, 1976). Situating youth's struggles within historical and structural relations of power, cultural studies scholars demonstrate how young people forge personal identifications through collective experiences of resistance. In 1979, Dick Hebdige brought the lens of cultural resistance to the study of self-presentation. His book *Subcultures* explores how young people use style as a way to symbolically subvert dominant systems of meaning. Similar to Willis and Hebdige, early work on youth subcultures tended to focus narrowly on the experiences of heterosexual white men living in urban centers. More recent work on youth cultural practice has turned its attention to notions of desire, and has examined the relationship between sexuality and consumption, as well as racial discourses and identity politics.

While education studies tends to privilege the concerns of formal curriculum and pedagogy, cultural studies suggests that in the context of schools, what teachers teach and how they teach is just a fraction of what youth actually learn. In fact, youth cultural practices are a substantial part of the curriculum and of students' school experiences. This view of learning has prompted cultural studies scholars to critique conventional models of curriculum and pedagogy whereby, as Freire (1970) famously observed, students are viewed as empty vessels into which information is deposited. In contrast, cultural studies provides a lens for understanding how youth make contributions to their own learning.

One tension produced by these two contrasting conceptions of youth is in the different ways that education and cultural studies view the pedagogical possibilities of popular culture. Recently, education scholars have acknowledged popular texts as a source for engaging students by connecting the formal curriculum to young people's interests and experiences. Yet, for the most part, these scholars continue to conceive of popular culture as a lesser form of cultural expression, viewing the role of schooling as one of inoculating students against potentially corrupting influences. In contrast, cultural studies takes seriously young people's investments in popular culture. As a pivotal field in which young people interpret, contest, and rearticulate meanings, popular culture can be a generative site, in which youth are not simply consumers, but are also active cultural producers of their own representations.

Recent work has explored transnational processes through a cultural studies lens, and emphasized the impact of global media flows on youth's understandings

of themselves and of their own social and geographical locations. At the heart of these processes is the issue of representation, a key concept in cultural studies. The work of Hall (1997) has been central to how dynamics of representation are understood within cultural studies and education. For Hall, representations are never simply free-floating signs that have inherent or predetermined meaning. Rather, how cultural products come to have meaning and to stand for ideas and people is a constantly contested process through which signifying practices evolve and are negotiated. Representations are never settled and their meanings are always in flux. As different people engage images and ideas in different contexts and under different circumstances, they reconfigure the meanings attached. This is most evident in the constant flux of images associated with popular media.

Critical media literacy provides an example of how these competing understandings of youth and popular culture create a tension that is productive for educational research. Buckingham (2003) argues for engaging popular culture as a pedagogical site where educators can move beyond the conventional objective of cultivating savvy consumers and toward a more dynamic practice of cultural production. Based on empirical research in media education classrooms in the United Kingdom, he demonstrates the pedagogical possibilities that arise from young people's playful approach to media production, as students rework popular texts in order to produce new forms of meaning. In doing so, Buckingham highlights the importance of popular culture as a resource that youth draw upon to construct and perform shifting identities. Even as he promotes youth media production as a promising site of learning, Buckingham dispels the myth that such projects are inherently transformative. His research reveals that students produce cultural texts that both challenge and uphold stereotypical representations of gender and race. Thus, critical media education must be grounded in an analysis of power that attends to the ways in which individuals reproduce as well as challenge dominant structures through their own meaning-making practices. This work also has implications for how youth become active participants in a democratic civic sphere, as they define new and alternative modes of engagement in public discourse.

Cultural studies emerged as an attempt to challenge both the notion that youth behavior stemmed from psychological processes of development and also the perception that young people were unaware actors in the social world. An approach to education that views youth practice as a form of active meaning-making suggests a rethinking of curriculum and pedagogy that parallels long-standing debates in curriculum studies, particularly in relation to the critique of the Tyler rational and traditional forms of curriculum design. Highlighting the value of popular culture as a site of learning, cultural studies

scholars demonstrate that engaging with students in their preferred forms of cultural expression can cultivate opportunities for youth to contribute to the public sphere. This expanded conception of education views cultural production as a form of democratic participation, and reflects cultural studies' enduring insistence that teaching and learning are intimately connected to political practice.

Relationship to Social and Political Change

From its inception, cultural studies sought to problematize the notion of the political and the relationship between culture and social change in at least two ways. First, as articulated above, the work of community-based organizations around the world in the early part of the twentieth century was grounded on the idea that cultural practice could undermine political and economic domination. Likewise, cultural studies scholars drew insight from the work of the Frankfurt School of critical theory and the ways in which Gramsci turned the Marxist frame on its side to suggest that cultural practice is not just the outcome of economic arrangements, but rather plays a role in those arrangements. These scholars were also convinced that academic work also had a fundamental role to play in social transformation. This belief is at the heart of the notion of public intellectuals that cultural studies scholars, such as Said (1994), have vehemently promoted, and continues to be central to the discourse of cultural studies. To the extent that academic work is itself a form of cultural practice, scholars play a role in reproducing, redefining, and/or resisting the social order and are therefore directly or indirectly implied in the oppression of others.

Dominant educational discourse generally fails to recognize teaching and learning as political. This is in part because of a narrow definition of politics that is limited to political parties, elections, and government policies. More fundamentally, however, this failure to understand how teaching and learning are fundamentally political has to do with a view of teaching as a neutral task that is purely technical and prescribed. By contrast, education scholars who draw on cultural studies have persisted in their commitment to thinking about teaching as a political act. As a form of change-directed engagement with individuals and groups, teaching is fundamentally political, whether it is as a form of reproduction or as a form of transformation. In her 1994 book, *Teaching to Transgress*, Bell Hooks elaborates the role that teachers can play in the transformation of consciousness and in the process of engaging social change toward justice and equity (Hooks, 1994). Drawing on Freire's (1970) understanding of teachers as cultural workers and teaching as a political project,

Hooks invites educators to realize that their daily engagement with students is a constant opportunity to imagine things otherwise.

The idea that teaching is political is not new, and education scholars have argued as much throughout the twentieth century, albeit in different ways. Yet, dominant education discourse has remained impervious to this insight and has insisted in framing teaching in terms of accountability, scientifically based evidence, and curriculum mandates. The tension between viewing teaching as a purely technical task, which still predominates in schools of education, and the view of teachers as intellectuals and political agents persists. The traditional educational approach suggests that teachers implement programs that are assumed to be a redress for social inequality, yet teachers are simply technicians. By contrast, cultural studies recognizes that teaching and schooling are inherently implied in the idea of inequality, while empowering teachers to engage in political practices in which teachers are not just implementers of programs but rather are directly engaged in political struggle.

Over the last 20 years, as cultural studies has become institutionalized within the academy, critics have suggested that it has lost its commitment to social justice. This presents an interesting irony – on the one hand, cultural studies provides theoretical lenses through which to understand teaching and learning as cultural practice and illuminate the ways in which they are inherently political; on the other hand, cultural studies scholars, in having to deal with the institutional structures that shape their work, have tended to lose sight of their political implications as they engage in these processes. Critical education scholars have sought to revitalize the political project of cultural studies (e.g., Giroux, 2000; Dimitriadis and Carlson, 2004; McCarthy, 1998), even as others have challenged some of the dynamics underlying how cultural studies has evolved (e.g., Weiler, 2001; Wright, 2004). In this sense, educators have a lot to offer cultural studies because their direct engagement and experience in education contexts provides ample evidence of the challenge of maintaining commitments to social justice and anti-oppressive education while working within institutional constraints. This tension is productive because it demands that cultural studies recognize the realities and challenges of institutional contexts and the immediate issues facing students and teachers, rather than escape to the intellectual realm of academia. In addition, it yields a view of teaching and educational practice that is collaborative, generative, and recursive, and that takes culture seriously as a site of social transformation.

Conclusion

The most important contribution that cultural studies makes to education is a view of culture as always in the

making – as a practice. This opens the door for seeing teachers as participants in the production of culture, and for an understanding of that process as essential for social transformation. Cultural studies provides a set of lenses through which to make sense of the experiences of students and teachers in school contexts as well as their educational experiences outside of school walls. It allows us to understand what students and teachers do as cultural practice and to consider how these practices have the potential to reproduce, resist, and redefine – sometimes at the same time – the structures that organize educational experience. In this way, cultural studies has been intimately related to recent developments in curriculum studies. Indeed, curriculum itself is now understood as a practice. Curriculum is an experience that emerges through encounters between learners (and, sometimes, teachers), in various educative contexts, whether formal or informal (even, sometimes, in schools).

Many education scholars have not missed the key insights of cultural studies, and many already engage in educational practices that view schools, youth, and teaching and learning in complex ways. Dominant educational discourse, in its attempt to universalize, essentialize, standardize, and measure all educational practice, has successfully ignored the challenges of cultural studies. Yet, many educators, in their search to humanize, contextualize, politicize, and bring more complexity to their teaching, have come to understand that schools are not neutral institutions, that youth are active participants in the world that surrounds them, and that teaching is a powerful form of political activism. The success of their work is evidence that cultural studies has much to offer to education.

See also: Critical Theory; Critical Theory and Pedagogy; Gender and Learning.

Bibliography

- Buckingham, D. (2003). *Media Education: Literacy, Learning, and Contemporary Culture*. Medford, MA: Polity.
- Carlson, D. and Dimitriadis, G. (2003). *Promises to Keep: Cultural Studies, Democratic Education and Public Life*. New York: RoutledgeFalmer.
- Freire, P. (1970). *Pedagogy of the Oppressed*. New York: Continuum.
- Gaztambide-Fernández, R. A., Harding, H. A., and Sorde-Martí, T. (eds.) (2004). *Cultural Studies and Education: Perspectives on Theory, Methodology and Practice*. Cambridge: Harvard Educational Review.
- Giroux, H. (2000). *Impure Acts: The Practical Politics of Cultural Studies*. New York: Routledge.
- Hall, S. (ed.) (1997). *Representation: Cultural Representation and Signifying Practice*. London: Sage.
- Hall, S. and Jefferson, T. (eds.) (1976). *Resistance through Ritual: Youth Subcultures in Post-War Britain*. London: Hutchinson University Press.
- Hooks, B. (1994). *Teaching to Transgress. Education as the Practice of Freedom*. London: Routledge.
- McCarthy, C. (1998). *The Uses of Culture: Education and the Limits of Ethnic Affiliation*. New York: Routledge.
- McRobbie, A. (1991). *Feminism and Youth Culture*. London: Macmillan.
- Said, E. (1994). *Representations of the Intellectual: The 1993 Reith Lectures*. New York: Pantheon Books.
- Thompson, E. P. (1963). *The Making of the English Working Class*. London: Victor Gollancz.
- Weiler, K. (ed.) (2001). *Feminist Engagements: Reading, Resisting, and Revisioning Male Theorists in education and Cultural Studies*. New York: Routledge.
- Willis, P. (1977). *Learning to Labour: How Working Class Kids Get Working Class Jobs*. Saxon House: Farnborough.
- Wright, H. (2004). *A prescience of African Cultural Studies: The Future of Literature in Africa is Not What it Was*. New York: Peter Lang.

Further Reading

- During, S. (ed.) (1999). *The Cultural Studies Reader*. London: Routledge.
- Edgerton, S. (1996). *Translating the Curriculum: Multiculturalism into Cultural Studies*. New York: Routledge.
- Gaztambide-Fernández, R. A. and Gruner, A. (eds.) (2003). Popular culture and education: [Special issue]. *Harvard Educational Review* 73(3), 253–465.
- Giroux, H. A. and Shannon, P. (eds.) (1997). *Education and Cultural Studies: Toward a Performative Practice*. New York: Routledge.
- Grossberg, L., Nelson, C., and Treichler, P. (eds.) (1992). *Cultural Studies*. New York: Routledge.
- Hoggart, R. (1957). *The Uses of Literacy: Aspects of Working-Class Life, with Special References to Publications and Entertainments*. London: Chatto and Windus.
- Kellner, D. (1995). *Media Cultures: Cultural Studies, Identity and Politics between the Modern and the Postmodern*. New York: Routledge.
- Maira, S. and Soep, E. (eds.) (2005). *Youthscapes: The Popular, the National, the Global*. Philadelphia, PA: University of Pennsylvania Press.
- Nayak, A. and Kehily, M. (2007). *Gender, Youth and Culture: Young Masculinities and Femininities*. New York: Palgrave MacMillan.
- Surber, J. P. (1998). *Culture and Critique: An Introduction to the Critical Discourses of Cultural Studies*. Boulder, CO: Westview Press.

Relevant Websites

- <http://www.ctheory.net> – CTheory.net.
- <http://theory.eserver.org> – EServer: Cultural Studies and Critical Theory.
- <http://www.popcultures.com>
- <http://www.media-awareness.ca> – Media Awareness Network.
- <http://www.theory.org.uk> – Media/Identity/Resources and Projects.

Cultural–Historical Activity Theory

Laboratory for Comparative Human Cognition, University of California, San Diego, San Diego, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

A long-standing challenge in educational research is to describe and explain the complex dynamics of learning and development that occur in educational settings. This article summarizes ways in which qualitative methods are essential to this enterprise from the perspective of scholars who approach the issues using the theoretical lens of cultural–historical activity theory (CHAT). After summarizing basic principles of this theoretical approach we provide four examples involving different levels of analysis and methodologies. (Methodology is used to refer to the ensemble of methods that mediate between theoretical statements and data used to evaluate them.)

To some researchers who employ qualitative methods, the very fact that we enter into this topic guided by a theoretical framework disqualifies our claim to be qualitative researchers. Smith argues that “qualitative approaches in psychology are generally engaged with exploring, describing and interpreting the personal and social experiences of participants. An attempt is usually made to understand a small number of participants’ own frames of reference or view of the world rather than trying to test a preconceived hypothesis on a large sample” (Smith, 2003: 2).

Our approach involves small samples, and we are interested in participants’ own understandings; however, we do operate from a preconceived theoretical base and in that sense we have preconceived hypotheses. Moreover, the approach we espouse does not preclude quantification. However, such quantification is more likely to be used for purposes of comparative analysis of qualitatively different activities (Cole *et al.*, 1978) or summary evaluations of products than for a deep analysis of the process of change (cf., Hayes, 1997).

The Theoretical Framework

CHAT refers to an interdisciplinary approach to studying human learning and development associated with the names of the Soviet Russian psychologists, L. S. Vygotsky, A. R. Luria, and A. N. Leontiev. (There has been a lively debate in recent years about the extent to which these three thinkers represent a single theoretical perspective. According to one line of interpretation, those who follow Vygotsky have focused attention on processes of mediation, adopting mediated action in context as a basic unit of analysis (Wertsch *et al.*, 1995). This line of work is often referred to as sociocultural research. By contrast, followers

of Leontiev are said to choose activity as a basic unit of analysis (Kaptelinin, 1996). For our present purposes, these distinctions are not central and we will treat the differing formulations as expressions of a single family of theoretical commitments.) The following are some theoretical principles of this approach:

1. *Mediation of experience through artifacts.* The initial premise of the cultural–historical school was that human psychological processes are bound up with a form of behavior in which material objects (e.g., hammers, pictures, gestures, and vocal sounds) and corresponding ideal objects (e.g., meanings and values) are incorporated into human actions and modified over generations as a means of regulating humans’ interactions with the world and each other. Artifacts incorporated into human action do not both change a person’s conditions of existence, and react back to that person’s psychological processes. Consequently, such artifacts are both symbolic and material mediators. Vygotsky referred to this kind of mediated action as the cultural habit of behavior which enables human beings to begin to regulate themselves from the outside.
2. *Activity as the essential unit of analysis.* The analysis of human psychological functions must be situated in relation to historically accumulated forms of human activity which are the proximal loci of human experience. The early Russian CHAT theorists demonstrated that at least in some institutional settings (among which schools and classroom activities were a major focus of attention) it is possible to make mediated-actions-in-activity/context a genuine object of study. Contemporary research has enormously broadened the range of activities and institutions to which scholars have been able to turn their attention (Engeström *et al.*, 1999).
3. *The cultural organization of human life.* Implied by the dual emphasis on mediation and activity is the centrality of culture in human life. Culture is present in the form of the tools, signs, cultural practices, architectural arrangements, social institutions, etc., that mediate human activity. It consists of all the material/ideal artifacts accumulated over the social group’s history, whether that history is of long or short duration.
4. *The primacy of the social.* From a CHAT perspective, children’s relationship to the world is, from the beginning, a social relation. However, especially in the beginning, children are maximally dependent upon adults not only

because of their physical immaturity but because they have no knowledge whatsoever of the cultural tool kit of the social group into which they are born. Thus, social others have a primary role in the development of psychological processes because it is only by already-enculturated adults arranging for the child to appropriate the cultural heritage of the social group that specifically human, culturally mediated, forms of psychological life become possible.

5. *Genetic analysis.* Vygotsky (1978) used the notion of genetic in the sense of seeking the origins of current phenomena by studying their history. Individual human development (ontogeny), he held, is the emergent outcome of processes of phylogenetic, cultural–historical, and microgenetic history.

CHAT Approaches to Learning and Development

A critical issue in CHAT approaches to education is the relationship between learning and development.

Learning

Vygotsky (1978) discusses learning in terms of the ideas of E. L. Thorndike, whose theory represented the original framework of behaviorism. According to Thorndike, learning is the result of trial and error during which associations form between stimuli and responses. Such associations or habits become relatively stronger or weaker when certain responses come to dominate others because they lead to rewards. According to Vygotsky, Thorndike proposed that:

[...] learning is the acquisition of many specialized abilities for thinking about a variety of things ... which does not alter our overall ability to focus attention but rather develops various abilities to focus attention on a variety of things. According to this view, special training affects overall development only when its elements, material, and processes are similar across specific domains. (p. 83)

Development

Vygotsky (1978) strongly differentiated learning from development. In his words:

Our concept of development implies a rejection of the frequently held view that cognitive development results from the gradual accumulation of separate changes. We believe that child development is a complex, dialectical process characterized by periodicity, unevenness of the development of different functions, metamorphosis or qualitative transformations of one form into another,

intertwining of external and internal factors, and adaptive processes which overcome impediments that the child encounters. (p. 73)

Vygotsky and his colleagues emphasized that not only are there changes in specific psychological functions (e.g., word meaning or memory) but that “the entire history of the development of higher psychological functions is nothing else than the alteration of primary interfunctional relations and ties, and the appearance and development of new psychological functional systems” (Vygotsky and Luria, 1930/1994: 167).

The Relationship of Learning to Development

Vygotsky (1934/1987) famously argued that school instruction should be organized in such a manner that “instruction moves ahead of development. When it does, it awakens a whole series of functions that are in a maturation” (p. 212). The kinds of functions Vygotsky had in mind included logical memory, abstraction, comparison, and differentiation. “These complex mental process,” he argued, cannot simply be learned in completed form, “like a mental habit” (p. 170).

But how is such arranging to be done? The problem does not arise for those who believe that learning equals development. For them the answer is clear: by direct instruction in which teachers provide appropriate stimuli and students are rewarded when they provide appropriate responses, that is, drill and practice with clear-cut consequences. However, as many critics of this form of instruction have noted, this method is inadequate:

The teacher who attempts to use this approach achieves nothing but a mindless learning of words, an empty verbalism that simulates or imitates the presence of the concepts in the child. Under these conditions, the child learns not the concept but the word, and this word is taken over by the child through memory rather than thought. Such knowledge turns out to be inadequate in any meaningful application. (Vygotsky, 1934/1987: 170)

It is to solve this problem that Vygotsky introduces the idea of a zone of proximal development (ZPD). The task confronting the teacher is to create a system of culturally mediated social interactions, organized in such a way that the student can, drawing upon prior and ongoing learning, achieve a qualitatively distinct new understanding, a conceptual re-organization that enables conscious control over the newly acquired knowledge. Creating such instructional circumstances is easier said than done. Although Vygotsky offered various suggestions about how to implement a ZPD, he never offered a systematic methodological system. (As Kozulin and Gindis (2007) comment, “Consequently, One can ... be true to the word and meaning of Vygotsky’s theory of the ZPD, but

one cannot follow or deviate from Vygotsky's ZPD assessment methodology for the simple reason that he never spelled it out" (p. 353).

Examples of CHAT Research

Recognition that the development side of the learning/development dance that constitutes successful academic education is a complex, dialectical process orients us to study the qualitative changes over time (longitudinally), at an appropriate genetic time scale. These dynamics involve the kinds of complex interactions between individuals in groups that routinely occur in instructional settings. CHAT-inspired research captures this complexity by focusing simultaneously on the activities or practices that contextualize these interactions. It also focuses on micro-genetic changes, ontogenetic changes, and cultural historical changes in relation to one another. Appropriate methods of data collection include digital voice and video recording, interviewing, structured and informal observation, participant observation, and quasi-experiments. Each example below represents the study of qualitative changes that occur in interaction under different levels of constraint in different social settings (dyadic teaching, a classroom lesson, weaving instruction in a Mexican village, and a small-group reading intervention).

Example 1: Puzzles

An early American study illustrating the centrality of structured observational method to CHAT research was carried out by James Wertsch and his colleagues (Wertsch *et al.*, 1980). In this work with middle-class families whose children attended preschool, mothers were asked to assist their infants (1–4 years of age) to copy a jigsaw puzzle depicting a cargo truck created from different regular shapes and colors. The researchers' particular interest was in the dynamics of adult and child discourse in relation to the child's puzzle-constructing actions. They hypothesized that these dynamics would change in conjunction with various adult communicative behaviors aimed at helping the child copy the puzzle.

Numeric registers alone, such as the number of times that a child glanced at the model when seeking to put a piece into the copy would yield a puzzling finding: The children all gazed at the model for roughly an equal amount of time. Careful analysis of the pattern of interactions indicated that the children had all learned names for the colors and shapes of the pieces of the puzzle, and the mother's often used the terms shape and color when seeking to assist them. However, it turned out that there was a regular change with age in who regulated the gaze at the model – sometimes the child gazed spontaneously, and sometimes the gaze seemed to require some sort of communicative gesture from the mother. In this respect, the children differed greatly: the youngest children were

least likely to attend spontaneously to the model. In Vygotskian terms, these data reveal a change in the interfunctional relations between attention and problem solving that had to be understood in terms of the developing child in relation to the enculturating mother. The mother's prior cultural knowledge, embodied in child-directed speech in conjunction with the children's prior learning (which in turn was mobilized by the mothers' when the children experienced difficulties), served to constitute the ZPD.

Example 2: Gazzinta and the process of long division

Petitto (1986) observed the process of teaching long division to fourth graders who were encountering it for the first time. Classroom interactions were videotaped. The data permitted the researchers to explore students' ZPDs as well as qualitative transformations in the interactions.

The children were expected to have learned their multiplication facts ($6 \times 7 = 42$) and simple division as reverse multiplication ($? \times 7 = 42$), as a preamble to the development of a qualitatively new form of arithmetic understanding – division that leaves a remainder. The teacher told the children that the seven *gazzinta* (goes into) 46 problem is solved in the same way as simple division, by finding a number which, when multiplied by seven produces a number close to 46 except it does not go over and then subtracting to find the remainder. The fourth graders seemed to have a difficult time grasping the *gazzinta* relation; however, with repeated examples using multiplication tables, over time the children began to catch on to what was being requested of them.

During a conference attended by several researchers from Japan, Petitto described the children's difficulty understanding the goes-into operation. Her description made perfect sense to members of the research team who had been discussing the work for some time, but the Japanese researchers were baffled. Eventually, one of them raised his hand to ask for an explanation of this new concept, *gazzinta*. Andrea was nonplussed. "Goes into" she said, slowly and with exaggerated intonation, and then demonstrated the entire procedure for determining how often seven *gazzinta* 46.

A fourth grader encountering long division for the first time faces a far more difficult task than the Japanese colleagues who knew the concepts involved in division. The child hears the word, *gazzinta*, and seeks, like the Japanese visitors, to figure out its meaning. But teachers do not say what *gazzinta* means. In fact, it probably is not possible to give an unambiguous explanation of *gazzinta*. *Gazzinta*, for long division with remainders, involves an iterative estimation procedure that is a combination of multiplication and subtraction carried out in no specifiable sequence on the number line. Faced with the difficulty of explaining the concept of *gazzinta*, teachers create a procedure to assist

the child, often putting the requisite times tables nearby. This offloads the work of retrieving the fact that $6 \times 7 = 42$, which is too little, and $7 \times 7 = 49$, which is too much, allowing the child to realize that seven *gazzinta* 46 six times, and that after subtraction, four will be remaining. The child must form the concept of *gazzinta* under the constraints provided by the adult and the cultural conventions for representing long division.

Researchers also noticed a strange phenomenon that further demonstrated the problematic relation between direct instruction and conceptual development. There were times when the teacher and a child were unknowingly talking about different parts of the same problem (e.g., the teacher said something about the quotient, the child understood it to be about the dividend) but there was no noticeable disruption in the conversation and the child arrived at a correct answer. It was as if the teacher and child were in close enough coordination, despite local discrepancies in the precise part of the problem they were referring to, to permit the action to unfold.

Petitto's work shows how instruction operates within a medium created by graphic symbols and classroom routines that provide constraints for, but cannot provide direct instruction for, mastery of a new concept.

Example 3: Remediating reading difficulties

Despite significant differences among them, modern approaches to reading have distinguished two, presumably distinct, major components of the reading process: decoding (the process by which letters of the alphabet are associated with corresponding acoustic patterns) and comprehension (the process by which meaning is assigned to particular visual/acoustic representations). Within this seemingly obvious dichotomy, theorists differ on the question of how to sequence instruction (code emphasis first vs. meaning emphasis first), how best to help children break the code (by teaching phonetic analysis or by teaching whole words), and how to motivate children to engage in reading (Burns *et al.*, 1999).

The following example seeks to solve these problems by creating a scripted, small group activity in which all of the theoretically important aspects of reading are brought together in a routine, coordinated, manner. The core of the procedure (described in more detail in Cole, 1996) was to create a set of roles or division of labor, each of which specified a different role in the overall process of reading. The roles were printed on index cards and every participant was responsible for fulfilling at least one role. The roles were:

- the person who asks about words that are hard to say;
- the person who asks about words that are hard to say what they mean;
- the person who asks a question about the main idea of the passage being read;

- the person who picks the person who has to answer the question at hand; and
- the person who asks about what is going to happen next in the text.

All the participants were given pencil and paper to jot down words, phrases, or notes (in order to answer questions about the text) and their card to remind them about their role. There was also a kitchen alarm to signal the start of reading time, and when it was over, the scripted activity of going through the roles in order to come up with a question about the main idea of the passage began. After a few sessions, the children and adults were able to engage in question–asking–reading, including children who still were unable to read for comprehension. The whole act of reading was not the responsibility of any one participant, but rather, was constructed for all in the act of working through the roles collectively and coming up with a question about the main idea.

Evidence for the way in which this procedure worked is derived from several sources to reveal qualitative differences in the children's ability to read and changes in those abilities over time: videotaped recordings of the instructional sessions, children's written work on the quizzes that completed each session, and various test results. Although data were gathered from the beginning of the first session, the crucial data for analysis of the process of reading development came after several sessions when the children had learned the overall script so that the group was working as a coordinated structure of interaction. Under such circumstances, different children became disoriented in different ways. Some children had difficulty saying how a written word sounded; others had difficulty understanding the meaning of the word; all had difficulty, at first, figuring out what the main idea was or using the information in the text to anticipate what would follow.

Data were also analyzed for instances in which misunderstandings moved up from decoding the comprehension of individual words to emerging comprehension of the text as a whole. Both forms of data helped the researchers to identify children's ZPDs and how the persons involved, the scripted procedures, and the cultural tools used contributed to change over time in both the participants and the activity.

Example 4: Developmental changes in learning to weave

In the 1960s, Patricia Greenfield and colleagues began to study the social organization and cognitive consequences of learning to weave among the Zinacantan Maya of Chiapas, Mexico (Greenfield and Childs, 1977). The researchers carefully described the way that mothers introduced young girls into weaving, analyzed mother–child interactions, and of the kinds of woven products produced. In the 1990s, they returned to the same village and conducted

parallel observations of parents (former child subjects) inducing their children into weaving. Greenfield emphasized the interconnectedness of historical change in economic activity, including exposure to new products and practices from contact with the modern sector of Mexican society, socialization practices (in particular, modes of socializing girls into weaving), and cognitive processes involving the mental representation of the patterns in woven cloth (Greenfield, 2004).

The instructional mode characterizing the mother-child weaving sessions in 1970 emphasized a long process of gradual apprenticeship involving many roles preparatory to weaving itself. When children first began to weave, mothers hovered close by and guided children with their own hands and bodies, using little verbal instruction. The entire system appeared to focus on maintenance of tradition and was characterized as interdependent cultural learning. In the 1990s, mothers who were more involved in the modern economy (e.g., weaving products for sale) instructed their children verbally from across the room. The mothers sometimes asked older children to take over instruction and only kept an eye on the processes, which were characterized by a good deal of explicit verbal instruction. Over time, the types of interactions changed as well as the number of artifacts used as a part of instruction. The gradually transformed participation in production of the 1970s gave way to increased verbal instruction and work with simplified weaving tools in the late 1990s.

In conjunction with changing verbal and nonverbal instructional practices, young girls were provided with simplified weaving tools of two levels of complexity. The simpler of the two was a tool for winding thread that maintains the orientation of the threads that would later be used in weaving the cloth; the more complex tool involved doubling the long (warp) threads around a dowel. This more complex approach required the weaver to visualize the extended warp (undoubled) rather than simply see it. Greenfield and her colleagues argue that the complex warping frame required the ability to engage in mental transformations while the simplified winding frame did not.

The changes in instructional practices came with an increased respect paid to individual innovation, seen in changes in the variety of products. In the late 1960s, the variety of products was limited, reflecting a very small set of right ways to weave. By the 1990s, there was no longer a small set of simple, correct patterns, but an efflorescence of patterns of great variety and complexity. This change was also shown in the way children represented weaving patterns in an experimental task, with children who attended school being more likely to create novel patterns. Overall, these results nicely illustrate several general theoretical claims derived from a CHAT perspective: the interconnections between cultural-historical change and ontogenetic experience, the ways in which microgenetic, minute-by-minute forms of interaction provide the

proximal locus for ontogenetic change, the primacy of the social in organizing specific cultural practices, and the changing nature of the artifacts that are employed to mediate these developmental processes at different scales of time.

Discussion

We have now presented four examples of research in the CHAT framework. They are characterized by two common elements. First, there is a shared focus on change over time in individuals, groups, and the activities in which they interact, even as the temporal and spatial dimensions of the activities under study vary greatly. Second, all four studies are explicitly theoretically informed. In **Table 1**, we summarize how each of the studies exemplifies different levels of analysis, different time and spatial scales, and different methods appropriate to each study.

These research examples highlight how qualitative methods are used to test the theoretical constructs we described as central to a CHAT approach. In our examples, the unit of analysis is not the individual learner, the teacher, or the learning environment. In all four examples, the unit of analysis is joint-mediated activity, which includes two or more individuals in interactions mediated by cultural artifacts (e.g., puzzles and terms for shapes and colors, a blackboard, math notation, times tables, texts, and a scripted process for reading and asking questions, and weaving tools and procedures).

All four examples, involved data collection that permitted analysis on at least two levels. In example 1, the role of social interaction in the development of problem solving involved analysis of primary social relations which are the source of the cultural resources the child will be expected to obtain and deploy as an adult. This example also illustrates the need to take into account, modes of discourse as they relate to qualitative changes in individual ontogenetic development associated with the acquisition of higher (e.g., culturally mediated) psychological functions. Methodologically, the brief temporal extent of example 1 indicates that it is focused on relatively short-term (microgenetic) interactions provoked by a puzzle task in a laboratory setting. It might also be considered a quasi-experiment.

In example 2, learning/instruction in a classroom, in particular, long division and the ineffable process labeled *gazzinta*, is the complex phenomenon being studied. The methods used included a series of observations and videotapes of math lessons on long division and subsequent analysis of discourse. In addition to the location of the study in a math classroom, the interaction data collected focused on microgenetic processes embedded within routine lessons.

Originally, analysis focused on the process of instruction, changes in the children's understanding, and the group processes and tool use that mediated those changes.

Table 1

<i>Example</i>	<i>Level 1: Sociocultural phenomenon</i>	<i>Level 2: Areas of specific focus</i>	<i>Temporal/spatial scale</i>	<i>Methods used</i>
1. Puzzles	Development of primary social relations	Mother–child interactions, discourse	Microanalysis: Short-term interactions in lab	Structured observation
2. Long division (<i>gazzinta</i>)	School-based math instruction	Classroom instruction, teacher–students	Microanalysis: Sequence of class periods, <i>in situ</i> in math classes	Natural observation
3. Question – asking–reading	Practice of reading school-based reading	Scripted activity as medium for experience of mature reading	Microanalysis of artifact-mediated social interactions over time, mesoanalysis of the scripted learning activity, ontogenetic changes in reading ability.	Design intervention
4. Weaving	Practice of weaving	Mother–child interaction, practice of instruction, practice, and artifacts of weaving	Microanalysis: Sequence of instructional interactions over time, <i>in situ</i> in homes Cultural–historical: Changes to the practices of weaving and instruction of weaving	Ethnography, and structured observation

Questions from Japanese visitors prompted analysis of the assumptions underlying the assumption that *gazzinta* was transparent as a term and a concept. A focused analysis of the complex cognitive routines involved in *gazzinta* – the meaning of *gazzinta* – remains to be conducted.

In example 3, rather than a quasi-experimental study in a laboratory or an observational study *in situ*, the research involved design of a learning environment in which a set of scripted procedures was used by adults and children to mediate the development of reading comprehension. The learning environment and the scripted procedures comprised an intervention, which the researchers hypothesized would permit participating children to experience qualitative change in their reading comprehension. In this study, group interactions were tracked over time in order to gain understanding of the microgenetic social and mediational processes related to individuals' changes in reading comprehension. At the same time, the development of the learning environment itself over time, a mesogenetic process that is longer than microgenetic interactions, but of a much shorter time scale than the cultural history of the practice of school-based reading, was tracked and analyzed. This allowed for a process of iterative design and ongoing refinement of the learning environment itself, even as the social interactions and individual changes in reading behaviors and comprehension provided particularly acute insight into the distinct strengths and weaknesses of each participating child, providing an ideal medium for remediation of their difficulties.

The timescales in example 4 are both microgenetic (instructional interactions) and cultural historical (change in the practices of weaving and weaving instruction over 30 years). The methods used included observation and

video recording *in situ* as well as ethnography that included both households and the political economy of the national context in which the activities of those households took place. This example illustrates the value of a historical lens in making sense of human interactions and practices such as instruction, which change in relation to societal changes, that is often not considered directly in educational research.

In all four example studies, the processes by which artifacts and social interaction with adults or more experienced peers helped to regulate children's behavior was of interest. In each case, the studies focused on the ways in which actions and the arrangements of artifacts were organized in order to promote or restrain learning.

In each case, there is a dual-mediational process at play. On the one hand, learning was mediated by more capable partners – mothers' communicative gestures, a teacher building on prior knowledge of multiplication tables and simple division, adult guides using a game-like script for reading instruction, and mothers and siblings using the traditional practice, verbal instruction, and modified tools for weaving. In all four examples, the children internalized or appropriated the use of symbolic tools (e.g., terms for color and shape, math language, the procedural rules in a script, written text, and visual and verbal instructions) in order to coordinate their roles in the different activities with the adults and other children. In all four examples, there were qualitative changes in observed behaviors, that is, the children's participation in the activities that indicated development. Observation of these changes and the social and mediational processes related to them required systematic collection of multiple forms of data over time and rigorous analysis of those data.

Conclusion

CHAT is a genetic and generative approach to understanding the living processes of learning and development. Research in the CHAT tradition is complex, but the rewards are many. As CHAT methods are deliberate models of cultural practices, there is an immediate reciprocal relationship between theory and practice. The ultimate criteria for success are not the discoveries and explanations that the research generates, but the uptake and the deployment of this knowledge by the educational communities crowding our country where education is woefully inadequate to the challenges of modern economic life.

Bibliography

- Burns, M. S., Griffin, P., and Snow, C. E. (eds.) (1999). *Starting out Right: A Guide to Promoting Children's Reading Success*. Washington, DC: National Academy Press.
- Cole, M. (1996). *Cultural Psychology: A Once and Future Discipline*. Cambridge, MA: Harvard University Press.
- Cole, M., Dore, J., Hall, W., and Dowley, G. (1978). Situational variability in the speech of preschool children. *Annals of the New York Academy of Science* **318**, 65–105.
- Engeström, Y., Lompscher, J., and Ru'ckreim, G. (eds.) (1999). *Putting Activity Theory to Work: Contributions from Developmental Work Research*. Berlin: Lehmanns Media.
- Greenfield, P. M. (2004). *Weaving Generations Together: Evolving Creativity in the Maya of Chiapas*. Santa Fe, NM: School of American Research.
- Greenfield, P. M. and Childs, C. P. (1977). Weaving, color terms and patter representation: Cultural influences and cognitive development among the Zinacantecos of Southern Mexico. *Inter-American Journal of Psychology* **11**, 23–28.
- Hayes, N. (1997). *Doing Qualitative Analysis in Psychology*. Hove: Psychology Press.
- Kaptelinin, V. (1996). Computer-mediated activity: Functional organs in social and developmental contexts. In Nardi, B. A. (ed.) *Context and*

- Consciousness: Activity Theory and Human Computer Interaction*, pp 45–68. Cambridge, MA: MIT Press.
- Kozulin, A. and Gindis, B. (2007). Sociocultural theory and education of children with special needs: From defectology to remedial pedagogy. In Daniels, H., Cole, M., and Wertsch, J. V. (eds.) *The Cambridge Companion to Vygotsky*, pp 332–362. New York: Cambridge University Press.
- Petitto, A. (1986). Division of labor: Procedural learning in teacher-led small groups. *Cognition and Instruction* **2**, 233–270.
- Smith, J. A. (ed.) (2003). *Qualitative Psychology: A Practical Guide to Research Methods*. Thousand Oaks, CA: Sage.
- Vygotsky, L. S. (1978). In Cole, M., John-Steiner, V., Scribner, S., and Souberman, E. (eds.) *Mind in Society: The Development of Higher Mental Processes*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. S. (1934/1987). *Thinking and Speech*. New York: Plenum.
- Vygotsky, L. S. and Luria, A. R. (1930/1994). Tool and sign in the development of the child. In van der Veer, R. and Valsiner, J. (eds.) *The Vygotsky Reader*, pp 99–174. Cambridge, MA: Basil Blackwell.
- Wertsch, J. V., del Rio, P., and Alvarez, A. (1995). Sociocultural studies: History, action, and mediation. In Wertsch, P. del Rio and Alvarez, A. (eds.) *Sociocultural Studies of Mind*, pp 1–34. New York: Cambridge University Press.
- Wertsch, J. V., McNamee, G. D., McLane, J. G., and Budwig, N. A. (1980). The adult-child dyad as a problem-solving system. *Child Development* **51**, 1215–1221.

Further Reading

- Cole, M. and The Distributed Literacy Consortium (2006). *The Fifth Dimension: An After-School Program Built on Diversity*. New York: Russell Sage.
- Engeström, Y. (1987). *Learning by Expanding: An Activity – Theoretical Approach to Developmental Research*. Orienta-Konsultit: Helsinki.
- Greenfield, P. M. (2002). The mutual definition of culture and biology in development. In Keller, H., Poortinga, Y. H., and Schölerick, A. (eds.) *Between Culture and Biology: Perspectives on Ontogenetic Development*, pp 57–76. New York: Cambridge University Press.
- Hedegaard, M., Chaiklin, S., and Jensen, U. J. (1999). Activity theory and social practice: An introduction. In Chaiklin, S., Hedegaard, M., and Jensen, U. J. (eds.) *Activity Theory and Social Practice: Cultural-Historical Approaches*, pp 12–30. Aarhus: Aarhus University Press.

Design Experiments

A Collins, Northwestern University, Evanston, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Design experiment – A formative experiment carried out to evaluate a particular design for a learning environment. It usually involves multiple iterations in order to identify problems with the design and to make refinements to improve the design.

Randomized-control design – The design of a comparative intervention experiment, where different groups of learners are assigned randomly to different conditions.

The evolving methodology of design experiments began as a reaction to traditional psychological experimentation (Brown, 1992; Collins, 1992) which had dominated educational research in teaching and learning. The methodology of psychological experimentation was based upon notions of controlling variables in order to be able to precisely determine what causes different effects. This led to conducting experiments under laboratory conditions according to carefully defined procedures. Design experiments, in contrast, attempt to carry experimentation into real-life settings, in order to refine the design so that it works in practice. This means giving up the notion of controlling variables, and therefore necessitates developing a new methodology for carrying out research.

While the initial work on this topic was carried out under the rubric of design experiments or design studies, more recently the terms design research and design-based research have come into common usage. Most of the research carried out in this vein concerns the design of computer-based learning environments. Designers of computer-based systems often go through multiple versions of a system until all the bugs are worked out, and in applying this approach to the design of learning environments, they bring a product-design mindset to the enterprise.

The novelty of the design-experiment methodology is seen most strikingly by comparing it to the experimental methodology used to study human learning in psychological literature. Learning research started before the turn of the century with the German psychologist Hermann Ebbinghaus, who invented the nonsense syllable in order to be able to study learning in its purest form. He identified many of the most important variables that affect learning, such as the similarity of stimuli to each other and the nature of the activity between learning and recall.

This tradition of research on learning continues to this day and has evolved to address questions about how humans learn to solve problems and carry out complex tasks. It has produced many important findings about the conditions that affect both learning and transfer.

Seven major differences can characterize how design-experiment methodology that is currently evolving differs from the kind of psychological methodology that has dominated education research heretofore:

1. *Laboratory settings versus messy situations.* Experiments conducted in laboratories avoid contaminating effects. Learners concentrate on the task without any distractions or interruptions. The materials to be learned are well defined and are presented in a standardized manner, rather than the manner a particular teacher may choose at any given moment. In fact, the presentation is usually one directional, rather than relying on interactions between teachers and learners. In short, learning in a laboratory does not look anything like what goes on in a typical classroom, workplace, or home, where most learning occurs in life. Design experiments are set in the messy situations that characterize real-life learning, in order to avoid the distortions of laboratory experiments.
2. *A single dependent variable versus multiple dependent variables.* In most psychological experiments, there is one dependent variable, such as the number of items recalled or the percent correct on a test of some kind. In design experiments, there are many dependent variables that matter, though the experimenter may not pay attention to them all. They fall into three types of variables: (a) climate variables, such as engagement of the learners, cooperation among learners, and risk taking by learners, (b) outcome variables, including the learning of content, skills, strategies, and dispositions, and (c) system variables, such as spread of use, sustainability, and ease of adoption.
3. *Controlling variables versus characterizing the situation.* Psychological experiments use a methodology of controlling variables borrowed from early physics. The goal is to identify a few independent and dependent variables, and hold all the other variables in the situation constant. Therefore, for example, if the experimenter regards amount of learning as the dependent variable, the goal will be to hold motivation constant. But the goal of teachers in classrooms is to find ways to motivate students, so that they learn something. Thus, holding

motivation constant fundamentally undermines the usefulness of the results. In design experiments, there is no attempt to hold variables constant, but instead the goal is to identify all the variables, or characteristics of the situation, that affect any dependent variables of interest. The goal is not only to characterize what affects any dependent variable, but also to identify the nature and extent of the effect.

4. *Fixed procedures versus flexible design revision.* Psychological experiments follow a fixed procedure that is carefully documented, so that it can be replicated by other experimenters. Design experiments, in contrast, start with planned procedures and materials, which are not completely defined, and which are revised depending on their success in practice. For example, Brown (1992) developed a design called Fostering a Community of Learners, where elementary school children worked in groups to learn about ecology. In design experiments, the experimenter should characterize what happens as completely as possible, and document any changes made in the plans, together with the reasons for the changes. The goal is to start with teaching methods that are most likely to succeed, but to monitor how they are working and to modify them when appropriate. This progressive refinement is standard practice in the product-design community, as can be seen in the many refinements that are made in products over time. However, until recently, progressive refinement was not the approach taken with education innovations, because of the strictures for replicability on the experimental methods, inherited from psychology.
5. *Social isolation versus social interaction.* In most psychological experiments, the subjects are learning in isolation. There is no interaction with other learners and usually no interaction with a teacher or expert; the material to be learned is simply presented by text or video. By contrast, design experiments are set in complex social situations, such as a classroom where students may be working in groups (Brown, 1992). In consequence, students are sharing ideas, distracting and making fun of each other, being interrupted in their work, trying to make life difficult for the teacher, etc. Design experiments have to cope with all the noisy data that arise from such situations.
6. *Testing hypotheses versus developing a profile.* In psychological experiments, the experimenter has one or more hypotheses, which are being tested by systematically varying the conditions of learning. In design experiments, the goal is to see what conditions lead to different effects. Design experiments ideally are much more like what consumer reports do when they evaluate the quality of different automobiles. The goal is to look at many different aspects of the design and develop a qualitative and quantitative profile that characterizes the design in practice. There are a large number of

contextual variables that determine the success of an innovation, such as the setting and professional development needed. It is best if evaluation is done with respect to a number of dimensions in a comparative fashion, as when consumer reports evaluate different products.

7. *Experimenter versus co-participant design and analysis.* In psychological experiments, the experimenter makes all decisions about the design and analysis of the data in order to maintain control of what happens and how it is analyzed. In design experiments, there is an effort to involve the different participants in the design, in order to bring their different expertise into producing and analyzing the design. Thus, teachers, curriculum designers, technology experts, cognitive psychologists, and anthropologists may all be involved in developing the design and evaluating its effects. Design experiments require many resources to be staged and hence it makes sense to bring to bear wide expertise in their design and evaluation.

Methodology of Design Research

The design-research community is developing criteria for carrying out and reporting on design experiments. Not every design experiment embodies these criteria, but they characterize the elements the design-research community is responsible for. In an ideal world, design research will move in the direction of embodying many of the practices we outline here. But it will take teams of researchers and accessible archives documenting design experiments to make these practices possible.

Implementing a Design

Each implementation of an education design is different. Therefore it is important to identify the theory behind the design, the critical elements of the theory for the design, and how the elements fit together. In order to evaluate any implementation, one needs to analyze each particular case in terms of these key elements and their interactions. Some elements will be implemented more or less as the designers intended, some will be changed to fit the circumstances, and some will not be implemented at all. What is needed is a profile for each implementation as to how each of the critical elements were implemented and how well the elements in the implementation worked together toward the designers' goals.

Modifying Designs

A goal of design research is to improve the way a design operates in practice. The teacher or researchers may see that an element of the design is not working in the course of the experiment. It is important to analyze why it is not

working, and take steps to fix whatever problems appear to be the reasons for failure. In this way we collect information about failures, plus information gathered from the attempted repairs to the design, and whether they succeed or fail. It is critical to document the failures and revisions, as well as the overall results of the experiment.

The experimental methods inherited from psychology that assume a fixed procedure are used throughout the experiment. Design research assumes continuous refinement. This difference has deep ramifications and requires changes in the way researchers analyze and report what is done. They should document their designs in detail, recording all major changes in design. These design changes mark the borders between phases. The goal then is to characterize the design elements that are in place in each phase and the reasons for the transitions from each phase to the next. Data relevant to research questions should be collected in each phase. For example, if there were four phases in a particular implementation, then it would be good if there were an intermediate assessment of learning outcomes between phases 2 and 3, as well as pretests and posttests. A detailed design history of this kind allows research audiences to evaluate the credibility of design decisions and the quality of lessons learned from the research.

Multiple Ways of Looking

Rogoff (1995) calls for analysts of learning environments to attend to three critical aspects: the personal layer (the experience of the individual), the interpersonal layer (one-on-one interactions), and the community layer. In the context of design experiments, researchers must additionally attend to interactions of learners with elements of the environment. There are many different aspects of what makes for an effective design, and so both designers and evaluators need to wear many hats in order to design and assess educational interventions. Consider some of the different aspects that are relevant to educational designs:

- *Cognitive level.* What do learners understand before they enter a particular learning environment, and how does that understanding change over time? Some of the tools for analysis at this level include observations of thinking through learners' representations and explanations. Through visual and verbal descriptions of ideas, researchers ask learners to expose their thinking. Are the explanations clear? Do representations capture important relationships?
- *Interpersonal level.* This viewpoint addresses how well teachers and students interact personally. Is there sharing of knowledge? Have the students bonded with each other so that they respect and help each other?

Researchers use ethnographic techniques to observe these kinds of interactions.

- *Group or classroom level.* This viewpoint addresses issues of participant structure, group identity, and authority relationships. Is everyone participating? Is there a sense of the goals and identity of the group? Again, ethnography is an effective approach to analysis.
- *Resource level.* This level deals with what resources are available to learners and if they are easy to understand and use. How accessible are the resources? How well are they integrated into the activities?
- *Institutional or school level.* At this level, issues arise as to communication with outside parties and support from the entire institution. Are parents happy with the design? Do administrators support it strongly? What are the micro-political issues that impact the design?

These levels are very much intertwined. To design and assess these different issues require many different kinds of expertise: teachers, administrators, psychologists, anthropologists, media designers, etc. Conceivably one person can address all these different perspectives, but it helps to have them all represented explicitly.

Characterizing Dependent Variables

Success or failure of an innovation cannot simply be evaluated in terms of how much students learn on some criterion measure. Different kinds of evaluation are necessary for addressing questions such as how sustainable the design is after the researchers leave, how much the design emphasizes reasoning as opposed to rote learning, how the design affects the attitudes of students, etc. To evaluate different variables, it is necessary to use a variety of evaluation techniques, including standardized pretests and posttests, survey and interview techniques, and systematic scoring of observations of the classrooms. Both qualitative and quantitative evaluations are essential parts of design-research methodology.

At least three types of dependent variables are important to assess: (1) climate variables, such as engagement, cooperation, risk taking, and student control; (2) outcome variables, such as content knowledge, skills, dispositions, metacognitive strategies, and learning strategies; and (3) systemic variables, such as sustainability, spread, scalability, ease of adoption, and costs.

Evaluating climate variables requires observational techniques, either by producing field notes while observing the intervention in practice, or collecting video records of the intervention and scoring those records subsequently. For example, these techniques might be used to evaluate three kinds of climate variables: the degree of engagement of students in learning in the classroom, the degree of cooperation among students in the classroom, and the degree of effort students are making to

understand the curriculum topic. To evaluate these variables, one might collect videos of different classes spread out over the time the teacher is carrying out the designed intervention. These videos can be scored systematically by multiple raters using a five-point scale for each specified interval in the lesson. Raters would be trained using benchmark lessons for which scores have been calibrated with experts.

Outcome variables are best assessed by collecting pretest and posttest measures. For example, pretests and posttests can be used to evaluate three kinds of learning variables: content, reasoning, and dispositions. To evaluate learning of content and reasoning, it is possible to use short-answer or essay questions, oral interviews, or multiple-choice items. By using items from standardized tests, it is possible to compare performance to national norms for the items. To evaluate learning of dispositions, one might apply instruments developed by Dweck (1986) to assess whether there are changes in students' beliefs reflecting a move from performance goals to learning goals. There have been such changes reported in a design experiment carried out by Scardamalia *et al.* (1994).

Systemic variables are best evaluated by interviews and surveys. For example, one might evaluate systemic variables, such as the ease of adoption of a design into the curriculum, the degree to which it is sustained in subsequent years, and the spread of use to other teachers and students. These can be measured by surveys and structured interviews with teachers and students. It is possible to develop a questionnaire that addresses the advantages and difficulties teachers encountered in adopting a design in their classroom. The other variables can be evaluated by surveys administered to both teachers and students at regular intervals. The surveys will ask about what aspects of the design are being sustained and are spreading, and which aspects are not.

Characterizing Independent Variables

In evaluating any design, there are a large number of independent variables that may affect the success of the design in practice. It is important to determine what general aspects of the situation researchers need to consider in order to decide what is affecting the success of the design. The contextual variables that can determine the success of an innovation include the following:

- *Setting.* The setting of the learning environment is a critical variable in how any design fares. The setting might vary over homes, workplaces, museums, schools, or colleges; elementary, middle or high schools; public or private schools; urban, suburban, or rural schools; elite or community colleges; etc. How broadly applicable an innovation is can be determined only by trying it out in many different settings.

- *Nature of the learners.* Critical variables about the learners include things such as their age, socioeconomic status, turnover rate, attendance rate, etc. For example, some innovations may work with weaker students and some with gifted students. So it is important to determine for which type of learners the design is effective, and in what ways.
- *Required resources and support for implementation.* In order to carry out any design, there will be a need for resources and supports of various kinds, including materials, technical support, administrative support, and parent support. If a design requires teachers to gather materials, spend time in preparation or other activities, enlist administrators or parents to make the design succeed, then these requirements need to be identified.
- *Professional development.* Often in order for a design to be successful, teachers (and perhaps others) need to be provided with professional development of various kinds. These can encompass workshops, design meetings, courses, videos of exemplary practice of the design, guided practice with expert practitioners, reflective meetings with colleagues, etc. Identifying what teachers need to implement the design successfully is an important aspect of designing an innovation.
- *Financial requirements.* Any intervention adds costs that need to be tracked, including equipment costs, service costs, professional support and development costs, replacement costs, etc. Very often substantial costs, such as technical support and replacement costs, are ignored when calculating the cost of a technological innovation.
- *Implementation path.* This term covers the variables involved in implementing a design, such as how the innovation is introduced, the time devoted to it, the difficulties teachers may face in introducing the design, etc. There is a structure to the introduction and evolution of a design that needs to be characterized in analyzing any implementation.

Reporting on Design Research

The experimental literature developed a conventional structure for reporting on experiments that evolved over time. The structure consists of four parts: background to the problem, experimental method, results, and discussion. As design research reconceptualizes the experimental process, there needs to evolve a different structure for reporting, perhaps including five sections in reporting on design experiments:

- *Goals and elements of the design.* An important aspect of reporting on design experiments is to identify the theory behind the design, the critical elements of the theory, and how they fit together to accomplish the goals of the design. The critical elements of a design may be materials, activities, a set of principles, or some

combination of all these. It is equally important to describe the goals of the design and how all the elements are meant to work together to attain those goals. Goals, critical elements, and their interactions need to be described in enough detail, so that it is possible to evaluate how well the design was implemented in different settings.

- *Settings where implemented.* The description of the settings needs to include all the information relevant to the success of the design outlined in the section titled 'Characterizing independent variables' above. Differences between how the design was implemented in each setting should be detailed, so that readers can evaluate how faithfully the design was carried out in each setting.
- *Description of each phase.* The design is likely to go through a different evolution in each setting, so it is necessary to describe each phase in each setting. When changes are made in a setting, the reasons for the changes should be specified along with the effects of making the changes. It also makes sense to describe how the critical elements of the redesign accomplish the goals of the original design or how the goals have changed.
- *Outcomes found.* The outcomes should be reported in terms of a profile of values on the dependent variables in the different settings, just as qualitative and quantitative data are reported about different products in consumer reports. These should be included to the extent that intermediate data describing the different phases were collected.
- *Lessons learned.* Considering what happened in the different implementations, the report should attempt to pull together all the findings into a coherent picture of how the design evolved in the different settings. It is important to describe the limitations and failings of the design, as well as the successes, both in implementation and outcomes.

Implications for Summative Evaluation

While design experiments were conceived as a formative evaluation strategy, the principles involved do have implications for summative evaluation. Any such assessment of educational innovations must carry out both quantitative and qualitative assessments, using a randomized-control design as Cook (2002) advocates, and comparative analysis, which consumer reports use. For example, to compare how effective two reading programs are, one would need to carry out comparative analyses in a variety of different settings, such as urban, suburban, and rural schools, and perhaps even homes, workplaces, and military settings. In such studies there must be a fixed experimental procedure, unlike the flexible design revision necessary for formative evaluation. The assessment should produce a

profile that shows the strengths and weaknesses of the designs being compared. Hence, different designs might be found to be more effective in some settings or with regard to some outcomes.

In order to have a sound assessment process, educational researchers as a community should develop a consensus process to determine what variables to look at and how to assess them. The assessment should address the multiple concerns of different stakeholders, including developers, and so they should be included in the consensus process. The design-research methodology argues that we need to look at multiple contextual and dependent variables, as described earlier.

To carry out such evaluations effectively, the country would need to invest in an independent agency, in the style of Consumer's Union, with the expertise to carry out comparative evaluation. Such an agency could develop the expertise and methods for looking in a cost-effective manner at innovations in use, in a way that best informs the many different stakeholders.

Conclusion

Brown (1992) felt that laboratory experiments, ethnographies, and large-scale studies are all valuable methodologies to study learning, but that design experiments fill a niche these methodologies do not address. Specifically, they allow researchers to evaluate and refine learning environments that are designed on particular principles and then revise the environment and the principles. Tharp and Gallimore (1982, 1988) have elegantly described how different methodologies can most effectively work together.

It is clear from the spread of these kinds of research methods (Barab and Kirshner, 2001; Barab, 2004, 2006; Edelson, 2001; Design-based Research Collective, 2003; Kelly, 2003; Sandoval and Bell, 2004) that design research is becoming an established practice. But design experiments often lead to the collection of large amounts of data that go unanalyzed. Hence, it makes sense for the design-research community to establish an infrastructure that would allow researchers at other institutions to analyze the data collected in design studies, in order to address their own questions about learning and teaching. This would require the community to honor such reanalysis of data with the same status as original research and it would require research journals and tenure committees to take such work seriously. Other fields, such as child language, have developed widely available archives of data, enabling researchers to discuss and analyze the same data from many different perspectives. The design-research community should strive to set up an infrastructure that can support researchers at different sites in analyzing the large data sets that design experiments are now producing.

Bibliography

- Barab, S. A. (ed.) (2004). Special issue: Design-based research: clarifying the terms. *Journal of the Learning Sciences* **13**(1), 1–128.
- Barab, S. (2006). Design-based research: A methodological toolkit for the learning scientist. In Sawyer, R. K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 153–170. New York: Cambridge University Press.
- Barab, S. A. and Kirshner, D. (eds.) (2001). Special issue: Rethinking methodology in the learning sciences. *Journal of the Learning Sciences* **10**(1 and 2), 1–222.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions. *Journal of the Learning Sciences* **2**(2), 141–178.
- Collins, A. (1992). Toward a design science of education. In Scanlon, E. and O'Shea, T. (eds.) *New Directions in Educational Technology*, pp 123–130. Berlin: Springer.
- Cook, T. D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Educational Evaluation and Policy Analysis* **24**(3), 175–199.
- Design-Based Research Collective (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher* **32**(1), 5–8.
- Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist* **41**(10), 1040–1048.
- Edelson, D. C. (2001). Design research: What we learn when we engage in design. *Journal of the Learning Sciences* **11**(1), 105–121.
- Kelly, A. E. (ed.) (2003). Theme issue: The role of design in educational research. *Educational Researcher* **32**(1), 3–34.
- Rogoff, B. (1995). Observing sociocultural activity on three planes: Participatory appropriation, guided participation, and apprenticeship. In Wertsch, J. V., Rio, P. D., and Alvarez, A. (eds.) *Sociocultural Studies of Mind*, p 252. Cambridge: Cambridge University Press.
- Sandoval, W. A. and Bell, P. (eds.) (2004). Special issue: Design-based research methods for studying learning in context. *Educational Psychologist* **39**(4), 199–260.
- Scardamalia, M., Bereiter, C., and Lamon, M. (1994). CSILE: Trying to bring students in world 3. In McGilley, K. (ed.) *Classroom Lessons: Integrating Cognitive Theory and Classroom Practice*, pp 201–228. Cambridge, MA: MIT Press.
- Tharp, R. G. and Gallimore, R. (1982). Inquiry processes in program development. *Journal of Community Psychology* **10**, 103–118.
- Tharp, R. and Gallimore, R. (1988). *Rousing Minds to Life*. New York: Cambridge University Press.

Further Reading

- Bereiter, C. (2002). Design research for sustained innovation. *Cognitive Studies, Bulletin of the Japanese Cognitive Science Society* **9**(3), 321–327. <http://ikit.org/people/~bereiter.html> (accessed September 2009).
- Collins, A., Joseph, D., and Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences* **13**(1), 15–42.
- Confrey, J. (2006). The evolution of design studies as methodology. In Sawyer, R. K. (ed.) *Handbook of the Learning Sciences*, pp 135–152. New York: Cambridge University Press.
- diSessa, A. A. and Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *Journal of the Learning Sciences* **13**(1), 77–103.
- Tharp, R. and Gallimore, R. (1982). Inquiry processes in program development. *Journal of Community Psychology* **10**, 103–118.

Discourse Analysis

P Nikander, University of Jyväskylä, Jyväskylä, Finland

S Vehviläinen, University of Helsinki, Helsinki, Finland

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Conversation analysis – The systematic study of talk in interaction in everyday and institutional settings. Teases out the moment-to-moment, turn-by-turn sequential unfolding of discourse as a generic order in interaction.

Critical-discourse analysis – A specific tradition within discourse studies that (often following the Foucauldian notion) approaches language and discourse practices not as an innocent, transparent, or neutral means of description, but rather, as a means of regulating, controlling, and governing discourse subjects.

Discourse analysis – A broad umbrella term for a variety of research traditions across disciplines that take talk, texts, and interactions as the analytic foci.

Discursive pattern – A recurrent feature or sequence in text or talk that structures certain discourse environments. For example, the initiation, response, and evaluation sequence in the classroom that furbishes speakers with certain positions.

Discourse and discourse analysis have become central catchwords in social and education sciences in recent decades. Closely linked to the constructionist or linguistic turn across disciplines, research in the field of education increasingly resorts to a range of discursive analyses that take language in context – texts, talk, and social interaction – as the starting point for their qualitative inquiry. In this article, we lay out for view the range of analytic foci under the broad umbrella term discourse analysis. Center stage is given to the variety of theoretical approaches and analytic emphases, and to the differences, dialog, and points in common between different traditions and schools of thought.

First, discourse analysis is best understood as an interdisciplinary collection of methodologies with varying theoretical standpoints within and across disciplines. Educational researchers have for decades drawn on a mix of theoretical and methodological developments, concepts, and analytic styles developed in linguistics, linguistic anthropology, post-structuralism, sociology, discursive psychology, and sociolinguistics when turning toward the study of discourse. Following a discursive epistemology (Nikander, 2008), researchers in education have covered an impressive

range of research topics, ventured outside the confines of teacher–student classroom interaction, branched out to questions of nonverbal, multimodal interaction, and tackled crucial and critical questions of how identities, knowledge, power, and ideology are constructed in face-to-face encounters and in-and-through historically unfolding educational and institutional texts and practices.

One decisive difference between the theoretical starting points and methodological brands and flavors of discourse analysis is based on whether the researcher foregrounds the microsituation in which an interaction takes place, or the wider political underpinnings of educational encounters, texts, and practices, and the more wide-reaching ideological effects of discourse. Based on these criteria, **Table 1** presents one way of breaking down the general rubric of discourse analysis into three traditions, and aims to pinpoint some key differences between them. **Table 1** should be taken as a broad-brush depiction of major developments and analytic foci. Out of necessity, some nuances within this area of research are silenced and some of the complexities lost.

Combining an interest both in the close analysis of language and communication acts with culture and context that surround them, ethnography of education has produced early insight into how race, class, and other cultural orders in communities outside the classroom have consequences on the dynamics and interactional processes in and across educational contexts. Large ethnographic, anthropological research projects based on close analysis of discourse, participant observation, and other means of data collection, have provided a nuanced picture of how notions of academic success, situated identity performance, and competence among others, are related to race, class, and cultural backgrounds.

Drawing on critical linguistics and post-structuralist discourse theory, critical-discourse analysis (CDA), seeks to identify the constitutive and disciplining force of educational discourses. Following the Foucauldian notion, language and discourse practices, in this tradition, are seen not as an innocent, transparent, or neutral means of description, but rather, a means of regulating, controlling, and governing discourse subjects. According to Luke (1996), the emergence of multicultural, multilingual, and global Western communities, and their rapidly changing demographic, ethnic, gender, and class profiles call for critical analysis of power in pedagogic surroundings, and genealogical studies of the ideological functions of curriculum and educational policies.

Table 1 Discourse analytic traditions, analytic foci, and typical data sets

<i>Tradition</i>	<i>Analytic focus</i>	<i>Approach and data sets</i>
Ethnography of communication	Speech events, practices, and repertoires	Analysis of linguistic data and participant observation in and outside the classroom
Critical discourse analysis (CDA)	Analysis of how discourses of/ within education are infused with ideology and power	Analysis of official texts and documents Deconstruction of power relations and the constitutive forces of discourse
Micro-level analyses, for example, conversation analysis, linguistic analyses of discourse	The turn-by-turn unfolding of interaction and situated action	Videotapes of institutional interaction within and outside the classroom and in other education-related settings

The third approach in **Table 1**, micro-level analysis of interaction *in situ*, is by far the most detailed approach to discourse. Research within this tradition focuses on mapping the architecture and the multimodal nature of our everyday and institutional interaction.

Based on videotaped materials from microsettings, and the detailed transcriptions produced thereafter, analyses seek to tease out the moment-to-moment sequential unfolding of discourse as a generic order in interaction.

Discourse Analysis and Education

From the above, one can deduce that the broad field of discourse analysis in education can be delineated by the division into (1) micro-level analyses of social interaction and verbal communication in various social settings and the social construction of activities, knowledge, competence, and identities thereof; and (2) macro-level genealogical analyses of institutional regimes and knowledge structures, the analysis of the production of realities and facticities. Third and more recently, feminist and postcolonial analyses have focused on situated accounts on individual learning experiences and the development of identities and positions in the society (e.g., Luke, 1996). In what follows, we further flesh out the picture of discursive research in education and discuss some empirical examples and future challenges. More specifically, we take a closer look at some of the topics and themes relevant in one of these basic fields, namely the micro-level interactional analyses. We start from the most descriptive, basic linguistic analyses and move toward more applied studies that aim to improve educational practices.

Micro-level analysis of interaction and speech stand at the intersection of several disciplines (anthropology, linguistics, sociology, social psychology, communication studies, and gender studies). Many scholars of classroom discourse have chosen to focus on this setting for reasons other than the interest of improving educational practice or even learning as such. For instance, for linguists Sinclair and Coulthard (1975), the classroom provided a setting for constructing an analytic framework for the systematic linguistic analysis of verbal discourse in general. For some

researchers, on the other hand, discourse analysis is a means of examining learners' interaction so as to identify the most successful practices for enhancing collaborative learning at school and to challenge the participatory patterns and epistemic division of labor in the traditional classroom (Mercer, 2004).

The Linguistic and Ethnographic Analyses of Verbal Communication

The roots of the study of classroom interaction and discourse reach to the 1970s and the multidisciplinary linguistic turn. The main historical influences have been the interactional sociology of Erving Goffman, ethnomethodological sociology as the analysis of mundane meaning making and of social activity as participants' practical accomplishment (Mehan, 1979), and conversation analysis as the systematic study of social interaction as a turn-by-turn, sequential accomplishment. This approach, sharing many core assumptions with discourse analysis, has a tradition for analyzing institutional discourse (for conversation analyses on classroom settings, see McHoul, 1978; Seedhouse, 2005). A prominent influence has been provided by the ethnography of communication, with the notion of speech communities and related communicative competence, as well as the development linguistics and the analysis of discourse as a linguistic phenomenon (Sinclair and Coulthard, 1975). In the 1970s, the so-called new sociology of education in Britain had also begun to produce micro-level analyses of the reproduction of social inequalities through language use in the classroom. (For overviews of these historical developments see Cazden, 1986; Luke, 1996; Warriner, 2008).

This branch of study has brought about a line of inquiry, combining linguistic and ethnographic orientations and methods to what is mostly a qualitative, data-driven analysis of naturally occurring social interaction, verbal communication, and activities in the classroom as well as in other pedagogical or education-relevant settings. In these studies, audio- or video recordings of naturally occurring social situations are generally used, as well as ethnographic data like observations, interviews, and various

textual materials consisting of assignments, tests, and student essays. The discursive practices are treated as contextual situational accomplishments, while the context itself – the shared frame for interpreting a given utterance or action – is viewed as a dynamic accomplishment of the participants. Context is invoked in various ways by participants and therefore should not be treated as a prefigured frame for analysis. Furthermore, it is assumed that classroom discourse is organized in terms of particular participation structures, with specific rights and obligations for the participants' communication and contributions, as well as particular access to knowledge. Participation in classroom activities – and thereby demonstrations of learning – take place through this, often implicitly acquired, set of assumptions. These studies, thereby, demonstrate and examine the socially achieved and dynamic nature of the classroom (or other educational activities and settings) as well as the identities, competencies, and subjectivities that are invoked and performed through these constellations of participation (Cazden, 1986).

Linguistically oriented microethnographic analyses of classroom interaction have evolved into a vast field of literature on various aspects of talk and interaction. Most prominently, this field has produced analyses of various events, actions, or patterns in interaction, of how these events act as constituents of the classroom activity, shape the participation and contributions of the participants, create possibilities or obstacles for communication and participation, and shape identities, memberships, and competencies of participants (Baker, 1997; Mehan, 1979; McHoul, 1978; Cazden, 1986; Stubbs, 1983; Lemke, 1990; Mercer, 2004). Some studies place more emphasis on detailed qualitative analyses of particular actions, discursive techniques, or activity types, while others include quantitative evidence of their distribution, thereby presenting evidence of general tendencies of their use during lessons and the entire school day.

Perhaps the most studied discursive phenomenon in the classroom is the three-partite sequence that typically organizes most of the dialog between the teacher and the class, namely the teaching cycle. This cycle, or the initiation, response, and feedback (IRF) or initiation, response, and evaluation (IRE) sequence represents the typical recurring unit, organizing the formal classroom interaction (Sinclair and Coulthard, 1975; Mehan, 1979; McHoul, 1978). It consists of the teacher's question, student's response, and the teacher's reaction (evaluation or feedback) to that response. Through this pattern, the teacher is able to control the topic of the talk and thereby the joint focus of the attention of pupils. This pattern is also a vehicle for controlling and allocating the pupils' participation, for monitoring their competence and providing feedback to individual pupils as well as to the entire class.

The dynamics of the teaching cycle sequence provide the teacher with the position of questioner: the controller

of the topic, as well as the evaluator of the content of the response and the competence of the responding speaker. The teacher's position as the epistemic authority is incarnate in this pattern, while the pupil is positioned as the speaker who demonstrates knowledge and competence, and offers it for the teacher's checking, control, and authorization. This pattern has been viewed as the basic element and constituent of the lesson and its components. The participation structure it provides for the counterparts has also been heavily criticized. Responding to questions can be seen as a fairly limited way of participating in learning activities. It can be stated, for instance, that instead of demonstrating their competence or understanding of the content or issue at hand, students end up demonstrating their ability to recognize the teacher's cues for the correct answer or an acceptable performance.

All three elements or turn types of this sequence have merited numerous studies of their own. For instance, in the studies on the questioning activities of the teacher, various ways of modifying the question have been identified. When pupils have problems in producing an answer, teachers are found to modify and reformulate their questions, using various ways of eliciting the appropriate answer thereby discursively scaffolding the pupil's performance. Teachers are shown to orient to the importance of having pupils themselves make some attempt at the question. Thus, this three-partite sequence is also an important way of involving the students and eliciting participation from them.

The IRE sequence is, of course, but one of the discursive patterns used in the classroom (cf. Lemke, 1990), and not all situations are organized as exchanges between the teacher and the class as a collective. A formal situation where a group is meant to learn by means of listening to an individual's monolog and reacting to it or responding to questions about it is nevertheless a common one, not only in schools but also in various adult settings. For instance, Tracy (1997) analyzed the academic departmental colloquium as a communicative situation, focusing on types of discursive actions in relation to the participants' discourse positions. Discourse positions refer to the various interests, entitlements, and constraints related to participants' institutional (academic) identities and their task of appearing both knowledgeable in academic-subject learning from each other and gaining new knowledge. When a novice (doctoral student) has given a departmental presentation, a senior staff member (such as a professor) may present a question signaling a problem of understanding without risking her or his epistemic authority or competence: the problem is likely to be treated as the presenter's problem. The situation is quite different with the roles turned around: if a student indicates problems of understanding, it is likely to be treated as her or his own. Thus, competence and expertise, as identities, are social accomplishments managed and negotiated in the discourse of institutional encounters.

An early focus was to analyze the interactional patterns in the classroom and the formal discourse between the teacher and the class. In recent years, the scope of the analysis has evolved from the teacher–pupil/teacher–class formal interaction toward various types of peer interaction and collaborative work in groups of pupils, as well as in their informal moments during breaks or meals (Goodwin, 2006). In addition, analytic dimensions have evolved: not only verbal and nonverbal communication is focused on, but the analysis extends to gestural and postural action, movement in space, the use of visual resources and material artifacts. Furthermore, the focus has moved from the examination of solely cooperative and coherently organized formal situations toward the analysis of conflict, negotiation, and improvisation – in other words, the interplay between organization and disorganization in the life of a classroom.

The Discursive Production of Memberships, Identities, and Differences

A key focus in the linguistically oriented discourse analysis has been the discursive production of the subjects of the pedagogical relationship: the pupil, the teacher, and the parent. These analyses focus on the categorizations and assigned obligations, responsibilities, moral expectations, and category-bound attributes invoked in face-to-face situations, in talk about absent parties, or in texts (e.g., Baker, 1997).

Variation of practices in discourse has been central within linguistically oriented ethnography of discourse. Speaking connects the speaker to a community, but may also disconnect her or him from others. In this vein, a key focus within micro-level analyses of classroom discourse has been to identify the construction and management of differences in speech and communication. At issue is how differences are recognized and interpreted in classrooms and other educational settings, and how they can be used to support the learning of different students. Similarly important is the question of how these interpretations work to maintain unequal treatment of pupils or students from different backgrounds (Cazden, 1986).

A seminal example of a discourse analysis of both construction of identities and linguistic variation in an educational setting is Erickson's and Schultz's (1982) microethnography of American college counseling. Erickson and Schultz point out that the counselor's work is essentially about defining, through the counseling interview, who the student is, in other words, categorizing the student for institutional purposes. Using videotapes of encounters and participants' commentaries on the tapes, the authors studied various phenomena such as verbal and nonverbal contextualizing cues or the rhythm of speech, identifying several interactional features that are likely to affect the way individuals experience each other and their encounter,

and the ways in which these experiences are related to the social (especially ethnic or speech community) identities of the participants. One of their main arguments was that communication styles, and situationally defined co-membership in a social category (speech community or other) have consequences on the friendliness of the encounter and the amount of personal attention and help the student receives. These studies provided evidence that college counseling, in fact, functions as a gate-keeping activity, and that the local management of social identities – ethnicity and various locally invoked co-memberships – influence the counselors' gate-keeping, and thereby also the potential outcome of counseling.

Discourse Analysis and Pedagogical Applications

Some approaches to discourse analysis not only describe the construction of meaning, participation, and knowledge in interaction, but also connect discourse analytic understanding of the interaction to the learning outcome. Studies within these approaches are usually informed by activity theory, Vygotskian psychology, and sociocultural or situational theories of learning. Social interaction is viewed as a constituent of learning itself. Many researchers have come to agree that learning should be viewed as an intrinsic dimension of participation in social activities, practices, and communities (Lave and Wenger, 1991). The sociocultural approach views learning, not as a function of individual cognition, but as a function of constant change in the constellation of participations in communities of practice. Along with the focus on the social nature of learning activity, there is a demand for a firmer empirical grasp on social interaction as the constituent of learning activity (Mondada and Pekarek Doehler, 2004; Martin, 2004).

One application of empirical work in this area is found in Mercer (2004) and his colleagues' work. The usefulness of discourse analysis is both in the possibilities for describing learning-relevant discourse as it unfolds, by the use of data-driven categories, and, based on these analyses, in creating typologies of discourse that enable teachers to improve their practice. Within this approach, language is viewed as a cultural and psychological tool for joint activity – a means and media for thinking and solving problems together. The overall interest in Mercer's and his colleagues' studies (for overview, see Mercer, 2004) has been the way in which different, relatively distinguishable discursive practices provide for learning, participation, and epistemic agency for children in classrooms. The focus is both on the teacher–student interaction and on the peer-group collaboration, and both qualitative and quantitative approaches are utilized. At issue is how participants of a learning situation use language to create meaning together.

Mercer's analyses distinguish discursive techniques used by teachers – types of elicitations (such as questions),

responses (such as evaluations or confirmations following responses to questions), and various kinds of recaps (statements describing aspects of shared experience). Importantly, Mercer and colleagues have also differentiated between three kinds of discourse in children's group situations and task-oriented work: disputational, cumulative, and exploratory discourse. In disputational talk, speakers mainly focus on maintaining their own position and resisting that of others. Speakers tend to disagree, using a lot of assertions and challenges. In Mercer's example below, pupils try to decide how to get about their group task and who should do what:

Example (from Mercer, 2004: 148, Transcript 1):

Carol: Just write in the next letter. "Did you have a nice English lesson".

Jo: You've got to get it on there. Yes that's you. Let's just have a look at that. "Hi Alan did you have a nice English lesson. Yes thank you, Yeah, yes thank you it was fine".

Carol: You've got to let me get some in sometimes.

Jo: You're typing.

Carol: Well you can do some, go on.

Jo: "yes thank you"

Carol: [unintelligible]

Jo: You're typing. "yes thank you" "I did, yeah, yes thank you I did"

Carol: You can spell that

Jo: Why don't *you* do it?

Carol: No, because *you* should.

In cumulative talk, on the other hand, participants build on each other's contributions without criticism or argument. Common knowledge is created by accumulation, using agreement, repetition, confirmation, and elaboration. The following excerpt is found from a Finnish classroom group work session (eighth to ninth graders), where pupils plan and coordinate their work.

Example (Turkia, 2007, 218 Example 2, translated from Finnish by SV; conversation analytic transcript simplified by SV, () = a pause, [] = overlap):

Toni: which part shall we present then () on the presentation day

Tuomas: I don't know (looks at a book)

Toni. well we should () have [something]

Tuomas: [lets see () a movie]

(Tuomas looks at Toni)

Tuomas: because those Niskavuori's have been done seven times or

Toni: let's take a piece of that [movie and show it]

Tuomas: [five]

()

Tuomas: yeah,

()

Toni: [that's a good idea]

Tuomas: [(yes) and besides we won't need to read anything.]

The third type can be both accumulative and confrontational; ideas are argumentatively examined and knowledge is made publicly accountable. This type of talk makes reasoning available to the participants. Many researchers have been interested in how to enhance such explanation, argumentation, and joint testing of ideas among groups. The following example comes from a Finnish teacher education setting, where second-year teacher students are learning about the concept of solubility through an experiment carried out in groups. The authors analyze the following episode as a case of joint construction of meaning of the concept of dissolving. The episode unfolds after a series of attempts to understand the experiment as well as what solubility means.

Example (from Kaartinen and Kumpulainen (2002), 201, Table 4, Episode 6, data translated from Finnish by the authors):

Juho: I do not know what dissolving is... whether one substance gives something to another

Paula: You can never read about it

Jarmo: Many times in high school, but I have forgotten

Elina: On the other hand, in dissolving it seems to me that the crystal structure gets smaller in salt so that the sodium

Juho: Sodium chloride breaks down

Elina: Sodium and chloride those ions go there

Jarmo: Yes the idea is like that those crystals in a way break down

Elina: Yes

Jarmo: And when water is being evaporated those ions join together again

Elina: that is how it seems to be

Jarmo: I think it is also a chemical explanation on the basis of formulae.

Sociocultural discourse analysis identifies these types of discourse, but also facilitates analyses on related applied issues. It is possible to examine, for instance, to what extent children of different ages engage in explorative discourse, and how their discourse helps them to accomplish their task. Another focus, directly relevant to teaching, is how children's discourse changes over time and whether teacher's facilitation enables such change.

Future Challenges of Discursive Research in Education

The diffuse nature of discursive analysis as a set of varied methodological and theoretical approaches to the systematic study of language in interaction presents a challenge to anyone new to the field. Scholars interested in utilizing discourse analysis need to distinguish between the multiple intellectual and theoretical histories, conceptual demarcations and emphases associated with discourse as a term, and to take care when situating their own (under)standing

and approach in this field of traditions. Discourse analysis as an approach may bring to the foreground the minute-to-minute construction of talk, identity, and meaning in the flux of microsituations, as well as historical, structural, and ideological power policies and their links to the construction of texts and interactions in and out of school settings. The researcher's gaze focusing on both the foreground and background – the zooming between the micro and macrostructures of educational encounters, discourses, and texts – thus needs to be based on an educated judgment of the field at large, the research question, notions of context, and the relevant unit of analysis (Warriner, 2008).

Numerous methodological and topical challenges to the field arise from shifting population demographics, multiculturalism, and new information technologies. Global social relations, new sources of information and learning expand the confines of the curriculum and provide new alternatives for the content, direction, and control of learning and teaching. As new technologies push the boundaries of teaching and learning beyond the classroom walls, discursive analysis also needs to continue moving beyond this comfort zone toward the analysis of multimodal aspects of learning, toward human-machine-based learning and the adoption, use, and interactional dynamics of new technologies in education. Simultaneously, qualitative and discursive analyses move ahead from the already-known basic dynamics of teaching and learning in interaction and increasingly focus on some of the more nuanced aspects of topics like resistance, embodiment, and emotions across educational situations. These developments mean that previous understandings of what discourse data consist of also continuously expand. In addition, our understanding of student positioning in and through discourse needs to be enriched by analyses of how class, race, and gender are re-negotiated and mixed into new intersectional youth identities and what the potential consequences of this are for educational institutions and for teacher identities and qualifications.

Basic discursive research on the phenomenon of teaching and learning are also essential for the future. The cultural and physical surroundings and modes through which these take place, however, are expanding in ways that need constant tackling and research effort. Simultaneously, practice-oriented discursive research, that not only seeks to understand and depict educational discourse processes, but also aims at direct practical impact or intervention, is imperative. Ongoing and future developments, evolving new structures, interaction processes, and forms of textual existence of educational institutions call for equally alert and informed development of discourse analytic tools. The future development of discursive research in the field of education thus depends on how scholars new to the field adopt and understand existing methods and traditions, how they further add to, refine, and extend the existing empirical and theoretical knowledge pool, and how new insightful

means of analyzing discourse in educational contexts are introduced and implemented. Theoretical and methodological advances in the field of discursive research in education will, also in the future, result from the continuous, unprejudiced, and well-informed cross-fertilization and dialog between analytic traditions.

See also: Classroom Discourse and Student Learning; Classroom Ethnography; Conversational Analysis; Critical Ethnography; Emotion in Educational Contexts; Ethnomethodology in Education Research; Identity; Learning Through Play; Multiple Comparisons; Peer Interaction and Learning; Reasoning and Explanatory Talk: Learning in Everyday Settings; Remembering as Social Activity; Social Interaction and Learning; Visual Data in Education Research.

Bibliography

- Baker, C. D. (1997). Ticketing rules: Categorization and moral ordering in a school staff meeting. In Hester, S. and Eglin, P. (eds.) *Culture in Action: Studies in Membership Categorisation Analysis*, pp 77–98. Lanham, MD: International Institute for Ethnomethodology and Conversation Analysis and University Press of America.
- Cazden, C. B. (1986). Classroom discourse. In Wittrock, M. C. (ed.) *Handbook of Research on Teaching*, pp 432–463. New York: Macmillan.
- Erickson, F. and Schultz, J. (1982). *Counselor as Gatekeeper. Social Interaction in Interviews*. New York: Academic Press.
- Goodwin, M. H. (2006). *The Hidden Life of Girls. Games of Stance, Status, and Exclusion*. Oxford: Blackwell.
- Kaartinen, S. and Kumpulainen, K. (2002). Collaborative inquiry and the construction of explanations in the learning of science. *Learning and Instruction* 12, 189–212.
- Lave, J. and Wenger, E. (1991). *Situated Learning. The Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Lemke, J. L. (1990). *Talking Science. Language, Learning and Values*. Norwood, NJ: Ablex.
- Luke, A. (1996). Text and discourse in education: An introduction to critical discourse analysis. *Review of Research in Education* 21, 3–48.
- Martin, C. (2004). *From Other to Self. Learning as Interactional Change*. Uppsala: Acta Universitatis Upsaliensis.
- McHoul, A. (1978). The organization of turns at formal talk in the classroom. *Language in Society* 7, 183–213.
- Mehan, H. (1979). *Learning Lessons: Social Organization in the Classroom*. Cambridge, MA: Harvard University Press.
- Mercer, N. (2004). Sociocultural discourse analysis: Analysing classroom talk as a social mode of thinking. *Journal of Applied Linguistics* 1, 2.
- Mondada, L. and Pekarek Doehler, S. (2004). Second language acquisition as situated practice: Task accomplishment in the French second language classroom. *Modern Language Journal* 88(4), 501–518.
- Nikander, P. (2008). Constructionism and discourse analysis. In Holstein, J. A. and Gubrium, J. F. (eds.) *Handbook of Constructionist Research*, pp 413–428. New York: Guilford Press.
- Seedhouse, P. (2005). *Interactional Architecture of the Language Classroom. A Conversation Analysis Perspective*. London: Blackwell.
- Sinclair, J. and Coulthard, M. (1975). *Towards an Analysis of Discourse*. London: Oxford University Press.
- Stubbs, M. (1983). *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*. Oxford: Blackwell.
- Tracy, K. (1997). *Colloquium: Dilemmas of Academic Discourse*. Norwood, NJ: Ablex.
- Turkia, N. (2007). Pienryhmäkeskustelun toimintajaksot. (Sequences of action in group work interaction). In Tainio, L. (ed.) *Vuorovaikutusta luokkahuoneessa. Näkökulmana keskusteluanalyysi*. Helsinki:

- Gaudeamus (Interaction in the classroom: conversation analytic perspective).
- Warriner, D. S. (2008). Discourse analysis in educational research. In King, K. A. and Hornberger, N. H. (eds.) *Encyclopedia of Language and Education Vol. 10: Research Methods in Language and Education*, 2nd edn., pp 203–215. New York: Springer.

Further Reading

- Burman, E. (1994). *Deconstructing Developmental Psychology*. London: Routledge.
- Butler, C. W. (2008). *Talk and Social Interaction in the Playground*. London: Ashgate.
- Cazden, C. (1988). *Classroom Discourse: The language of Teaching and Learning*, 2nd edn.
- Cole, K. and Zuengler, J. (eds.) (2007). *Research Processes in Classroom Discourse Analysis: Current Perspectives*. Mahwah, NJ: Erlbaum.
- Edwards, D. and Mercer, N. (1987). *Common Knowledge: The Development of Joint Understanding in the Classroom*. London: Methuen.
- Hester, S. (1998). Describing 'deviance' in school: Recognizably educational psychological problems. In Antaki, C. and Widdicombe, S. (eds.) *Identities in Talk*, pp 133–150. London: Sage.
- Lemke, J. L. (1998). Analysing verbal data: Principles, methods, and problems. In Tobin, K. and Fraser, B. (eds.) *International Handbook of Science Education*, pp 1175–1189. London: Kluwer.
- Luke, C. and Gore, J. (eds.) (1993). *Feminism and Critical Pedagogy*. New York: Routledge.
- Mercer, N. (2000). *Words and Minds: How We Use Language to Think Together*. London: Routledge.
- Peräkylä, A. and Vehviläinen, S. (2003). Conversation analysis and the professional stocks of interactional knowledge. *Discourse and Society* 14(6), 727–750.
- Rogers, R. (2005). Critical discourse analysis in education: A review of the literature. *Review of Educational Research* 75, 365–416.
- Schegloff, E. A. (2007). *Sequence Organization in Interaction*. Cambridge: Cambridge University Press.
- Tholander, M. and Aronsson, K. (2003). Doing subteaching in school group work. Positionings, resistance, and participation frameworks. *Language and Education* 17, 208–234.
- Vehviläinen, S. (2003). Avoiding providing solutions: Orienting to the ideal of student's self-directedness in counselling interaction. *Discourse Studies* 5(3), 389–414.
- Wortham, S. and Rymes, B. (eds.) (2003). *Linguistic Anthropology in Education*. Westport, CT: Praeger.

Document Analysis

T Rapley and K N Jenkins, Newcastle University, Newcastle upon Tyne, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Documents are a seen-but-unnoticed part of our everyday lives. Whether held in one's hand or viewed on a screen, they are staple elements of all areas of our lives. They shape our thoughts, actions, talk, and writing. We not only consume them in vast quantities, routinely refer to them in conversations, but also produce them. However one conceives of them, they are essential to the formation of our contemporary societies and institutions.

Document-based research in qualitative educational research, as in all areas of the social sciences, is a relatively small-scale enterprise. The notable exception is historically orientated research – where given the lack of living sources of evidence – the document is the only potential source of evidence. In addition, the practice of reviewing the literature relies on collecting, analyzing, and speaking about a range of documents. Commonly, this is often relegated to the front sections of articles and reports, prior to a primary focus on interview-based talk or observations of others' interactions.

In this article, we begin to explore some of the variety of ways that qualitative researchers have actively sought to engage with documents as a central part of their research practice. We explore some more general issues about conducting research on and around documents. We then focus on two main areas. First, we look at research that focuses on the substantive content of documents – the relatively new area of systematically reviewing qualitative research articles – and the more discursive work that looks at how documents, be they newspaper articles or policy documents, create specific versions of the world. Second, we look at research that focuses on the situated use and creation of documents – how interview-based and ethnographic studies have also looked at the place of documents in the lives and routines of those they study and on the more ethnomethodological work, that seeks to describe in detail the practical ways that documents are used and created in specific instances of interaction.

Approaches to Analyzing Documents

In this article, we are using the term documents to refer to both paper-based and computer-mediated texts. This includes both the written elements of texts alongside the extra-textual elements – images, photos, graphs, and diagrams – that are routinely embedded in documents.

The post-structural and postmodern turn in academia has led to an expansion of the meaning of the term text, to include buildings, bodies, clothing, alongside artifacts, devices, and other aspects of material and technical culture. Simultaneously, interview transcripts, field notes, and video recordings are referred to and analyzed as texts, but we will not focus on either of these genres of work here. Instead, we are concentrating on four relatively distinct areas of work around documents (see **Table 1**). The table proves an at-a-glance comparison of the four analytic methods in terms of their general focus; a more descriptive account of their analytic focus; examples of their typical research questions; the data collected to answer these questions; the key methods involved in data collection; the methodologies often related to these approaches; and the type of results each produces. These four analytic approaches are described individually in greater detail below.

Analytic work on documents can be loosely divided into two areas (see **Table 1**, focus):

1. work that focuses on the actual textual and extra-textual content of documents (meta-synthesis and discourse analysis) and
2. work that focuses on the use, role, and function of documents in interactional and organizational settings (ethnography and ethnomethodological ethnography).

The first focuses on the document as an object in its own right, as an orphaned or docile text, as a container of knowledge. Such work forgoes empirical observation of how people actually read, refer to, or use the documents in question. Whereas the second area is primarily observational, seeking to understand how documents are active agents in organizational and/or interactional life. Obviously, the same topic can be approached in multiple ways and within the same project. Therefore, with a research project on, say, student evaluations of lecturers, one might focus on: the written comments students give (via discourse analysis); how results of these evaluations are fed back on websites, internal memoranda or official reports (via discourse analysis); how lecturers introduce, distribute, collect, and read such forms (via observation); what students, lecturers, and managerial staff think is the value of such documents (via observation and interview); the role of such evaluations in creating institutional change (via observation and interview); and so on.

Whatever approach people take, and irrespective of the methodological traditions they follow, the practical work of analyzing data related to documents closely echoes that

Table 1 Approaches to analyzing documents

	<i>Meta-synthesis</i>	<i>Discourse analysis</i>	<i>Ethnography</i>	<i>Ethnomethodological ethnography</i>
Focus	Content of documents		Documents-in-action	
Analytic focus	Intertextuality	Language and intertextuality	Social interaction and organizations	Language and social interaction
Typical research questions	What is the evidence for a specific intervention?	What versions of the world do documents create?	What part do documents play in the organization of a social institution?	How do documents shape (and are reflexively shaped by) ongoing social interaction?
Data	Academic outputs	Any	Any	Any
Method	Collect documents	Collect documents	Collect documents, field notes, and interviews	Collect documents, field notes, interviews, and audio-visual data
Methodologies	Meta-ethnography, meta-synthesis, and thematic synthesis	Discourse analysis, critical discourse analysis, Foucauldian discourse analysis, and semiotics	Interactional ethnography, grounded theory, phenomenology, and symbolic interactionism	Conversation analysis and ethnomethodology
Results	Synthesis of findings of a collection of documents	How documents construct and reflect a specific version of an argument, truth, social identities, or institutions	Use, role, and function of documents in specific organizational contexts	Use, role, and function of documents in specific interactional contexts

of all forms of qualitative research. Therefore, the data – whether they are documents, interview, field notes, or audio–video recordings (see **Table 1**, method) – will be subject to some form of sampling, as it is collected it will be coded, and some version of the constant comparison method and deviant case analysis undertaken, as well as some level of reflection on the role of the analysts in the process.

The analysis of documents is also related to the research tradition of content analysis (Weber, 1990). In this, researchers gather specific documents, then establish a coding frame and apply that coding frame to the documents to count the number of times particular words, phrases, or themes are used. This enables descriptive and statistical findings to be established. However, this is often seen by qualitative researchers as a relatively limited, albeit systematic approach, in that it predefines the key aspects of analysis alongside removing words and phrases from their context. Some researchers use content analysis alongside other more discursive methodologies, in part to gain a general overview of their data, although this is not discussed here.

Meta-Synthesis: Documents and Creating Evidence

The social, medical, and political sciences have seen the rise of evidence-based policy and practice. Education research has been central in this trajectory, and this shift

has not been without its critics. The growth in systematic reviewing and meta-analysis has also been significant in prescribing a new format for the textual analysis of research (see **Table 1**, first column). In relation to reviews of nonexperimental research, Noblit and Hare's (1988) work on meta-ethnography has been influential. They sought to combine the findings of six ethnographies on the impact of desegregation on urban schools in the US in the late 1970s. They translated the relatively diverse findings of the case studies into a general theory about the impact of desegregation. Following this intertextual work, a range of approaches to synthesizing the findings of qualitative research have emerged. They have been referred to under various terms including: meta-synthesis (Sandelowski and Barroso, 2007), critical interpretative synthesis (Dixon-Woods *et al.*, 2006), and thematic synthesis (Thomas and Harden, 2007). They all rely on insights from Noblit and Hare's (1988) original work, and all seek to move beyond simply aggregating and summarizing findings from qualitative research papers to actively create new conceptual models or theories and to provide evidence for future interventions.

As with quantitative systematic reviewing, this qualitative synthesis work involves searching, evaluation, data extraction, and presentation. However, given the non-numerical nature of the findings, they have adapted each stage in the process. Importantly, the process echoes and draws on key ideas from how one conducts empirical, primary, qualitative research. For example, when searching the literature one seeks a broad range of heterogeneous

studies, purposively sampling studies that cover a range of approaches and findings to the substantive topic of the review. One may engage in further rounds of sampling to find further documents that challenge or extend one's initial findings. The evaluation of the quality of studies is also complex, in that there is no hierarchy of evidence available for qualitative research. In this way, inclusion and exclusion criteria and judgments about quality are more flexible with an initial focus on how relevant the article is to the topic and to the development of the synthesis, and only papers deemed to have major flaws are rejected at the outset. Currently, we have a range of approaches for judging quality that draw on and combine the key aspects of various quality guidelines that exist.

With data extraction, the focus is on collecting data that enables the researcher to outline the key ideas, concepts, and themes within the article. This also includes extracting verbatim quotes from interviews, field notes, and other sources alongside summarizing key aspects. The final stage of synthesizing the data shifts the focus beyond just summarizing the extracted data, to producing explicit interpretations. Again, this process is similar to that undertaken in analysis of primary data in qualitative research. The extracted data is coded in some detail, coded sections are compared and contrasted, and one generates a coding frame which outlines the main themes and issues and relationship between them. These recurring lines of argument (Noblit and Hare, 1988) are drawn from both the new conceptualization of the original data alongside some of original concepts used in the articles.

This style of working with documents is still very much in its infancy, much of the work has been undertaken in relation to health-related topics, and currently we have a range of complimentary methods and methodologies. What they do share is a concern to compare, combine, and critically interpret the findings of different qualitative articles. In this way, they attempt to generate new micro- and meso-level conceptual models of the topic of the review.

Discourse Analysis: Documents Create Realities

Rather than see the documents as neutral sources of information to be searched, evaluated against some external quality criteria, and then the data abstracted from the context, another way that researchers analyze documents is to undertake a relatively close and detailed analysis of the language and meaning within the document (see **Table 1**, second column). This style of research, known as discourse analysis, focuses on how documents produce specific versions of reality. Within this approach, language (written or spoken) is never treated as a neutral, transparent, means of communication. Instead, language is understood as

performative and functional. So, for example, describing something as systematic review instead of as, say, a review of the literature or a child as gifted and talented over, say, bright for their age does specific work, and directs thoughts and actions. Researchers are interested in the specific versions of the world that documents produce, the effects they have both on the potential reader and on the wider culture, society, and institutions.

Within this style of research, the focus is on how words are used. Methodologically, there are two main research trajectories – discourse analysis (Potter and Wetherell, 1987) and critical discourse analysis (Rogers, 2004) – although in practice they often somewhat overlap. Whatever approach is taken, when studying documents in this way one is interested in the rhetorical work of the text, how the specific issues it raises are structured and organized, and how it constructs a specific version of the world.

One might try to understand how the document seeks to persuade someone about the authority of its understanding on an issue. So, for example, one might focus on the range of sources of knowledge and evidence a document draws on – so within a newspaper article one may have verbatim quotes from leading practitioners or academics, references to policy documents or academic articles, and report on conflicting sides of the argument. Such intertextual work helps build the factual status of the account. Moreover, one might focus on the forms and modes of language, knowledge and evidence that are employed. So, for example, systematic reviews and meta-analysis rely on statistics, graphs, and tables – especially forest plots – alongside a detailed account of the methods employed. We have no way of knowing whether in practice any of these methods was actually followed in just this way, but trust and authority are built in and through presenting the case in just this way. Centrally, one is trying to make sense of how the document produces a specific argument, the positions it takes in relation to a topic alongside the other alternate or contradictory positions that are excluded or silenced.

Another central question asked of documents from work in this tradition is how are specific ideas, practices, or identities produced, sustained or negotiated within texts? Some researchers seek to understand and describe the historical trajectory of the contemporary ideas we all currently take for granted. People often work quite closely with some of the writings of the French philosopher Foucault (1977). One of Foucault's interests was in how specific discourses, specific formations of what he called power/knowledge (e.g., medical, psychological, or pedagogical discourses) produce, shape, and enable specific subjects (e.g., attention-deficit/hyperactivity disorder (ADHD)-diagnosed children) and specific actions (e.g., prescribing medications like Ritalin). For example, Rafalovich (2004) has focused on the the conceptual antecedents of the contemporary diagnosis of ADHD. Drawing on articles and books from 1877 to 1929, Rafalovich looked at the

trajectory of three concepts – idiocy; imbecility; and encephalitis lethargica – and shows the initial attempts by medicine to establish an organic basis for children's immoral, or unconventional behavior.

Such histories of the present remind us that what we take for granted often has complicated and esoteric beginnings. This style of work may lead us to the question why exactly we act in just this way or why certain groups of people have the knowledge and power to understand and act on others. Importantly, it seeks to map the trajectories of these discourses, to outline the specific strategies that led to our present as well as those that were rejected or seen to fail. Obviously, given its historical nature, this style of work often uses documents to explore specific topics or questions, although saying that such histories may focus on something that has emerged in the recent past and may also rely on other forms of data – a multi-method approach. Therefore, in Rafalovich's (2004) study of ADHD, he also undertook interviews with parents, children, teachers, and clinicians. In addition, rather than focus in great detail on a small sample of texts these studies can focus on a broad range of texts as they try to show the styles of thought as they emerge, consolidate, and compete across and between texts often over a large timescale. Importantly, such work asks "What are the assumptions in specific documents?"

Ethnography: Documents, Interactions, and Organizations

The ethnographic approach to documents (see **Table 1**, third column) can cover a broad spectrum of activities and sites, although it routinely focuses on the use, role, and function of document in the context of schools. In general terms, such work seeks to combine a more micro-focus on moments of social interaction with and around documents in a range of classroom settings, a meso-focus on the organizational ecology of schools and a macro-focus on broader community and political factors. One of the most obvious roles documents play in education is reading, especially reading in class. Ethnographers are not really concerned with reading in terms of the cognitive processes involved and do not understand it as a monadic activity – that is, that reading can only be understood as something people undertake in isolation. Rather, they focus on reading as a practice situated within and related to the broader contexts of classroom activities and organizational cultures.

For example, the video-based ethnography of literacy by Castanheira *et al.* (2001) demonstrates the situated roles and relationships that people can establish around documents. They show how one teacher positions a specific workbook as the authoritative source of knowledge and how the student orientated to this text and their partner as

potential resources for defining and completing the task. Whereas in a different lesson, a teacher directly engages with the student, asking questions, encouraging them, draws a graph on the student's paper, etc. In this context, the student and teacher collaborate on the problem at hand, thus the student is positioned as central in producing knowledge with the workbook supporting him. Such ethnographic work also seeks to make sense of such situated reading work within the organizational ecology of education. Therefore, one may compare the use, role, and function of workbooks over time, over single classes, over different students, or over multiple subject areas; one may research the background of the teachers or how the workbooks get selected as well as the schools or regional policies around such workbooks. In this way, such research focuses on the embeddedness of reading documents in relation to both student's classroom interactions and their broader educational and social worlds (Green and Meyer, 1991).

As education is bureaucratically organized – and routinely involves coordinating a diverse array of professionals, administrators, and technicians alongside learners over a range of activities, times, and spaces – documents become central, a glue that binds these people and their work together. Mehan's (1993) ethnographic study of the identification, classification, and labeling of educationally disabled children shows us the central role that documents can play in the coordination of educational interactions, decision making, and organizational work. His study draws on a range of methods; these include observing, interviewing, video-recording meetings, and lessons alongside asking teachers to reflect on the videos of these lessons. He uses the sequential paper trail that is created by a student (in class work, examinations, and educational testing sessions with the school psychologist) and created about a student (in their personal school records, school appraisal team committee, and eligibility and placement committee meetings) to focus the fieldwork on the range of sites where bureaucratic assessment and classification work occurs in the special-education referral process.

Documents, in the form of student records, are continually created and discussed by a range of professionals in multiple sites and spaces. Centrally, these records come to represent and stand-on-behalf of the student. In meetings about students, discussions routinely draw on and highlight aspects of the paper trail created by and about the student. The decisions that emerge from these meetings are central in classifying the child and, therefore, are central in creating their specific educational trajectory. In his study, Mehan (1993) does not present us with the documents themselves, rather he focuses on the distributed talk and actions of those people who create and use them. He shows us how the documents created by and about students become objectified and how the technical talk and decisions they facilitate are given much higher status than other forms of evidence such as a parental voice.

Ethnomethodological Ethnography: Documents-in-Interactions

Research on documents within the tradition of ethnomethodological ethnography – and we include conversation analysis within this area – focuses on describing in great detail the moment-by-moment social organization of interactions with, around, and about documents (see **Table 1**, fourth column). Like ethnographic work outlined in the last section, it is not so much interested in the content of documents, but in the organization of interactions where documents take some role. However, in general terms, there are two central differences. First, for ethnomethodological work, the immediate, here-and-now, to-and-fro of the participants interactions is the central object of interest, description, and analysis. Broader contextual factors, like participants' backgrounds or organizational policies, are generally not drawn on as a resource to explain why an interaction happens in a specific way. Instead, the focus is on what the participants in the interaction show themselves to actually do, the resources they draw on in just that moment of interaction. Second, ethnomethodological work increasingly relies on audio or video recordings of naturally occurring interactions. These form the centerpiece of analysis, are subjected to repeated viewings, and sequences of interaction – including verbal and nonverbal features – are transcribed in very fine detail, analyzed, and revised in conjunction with the original recordings.

An example of this method in practice is Heap's various analyses of audiotapes of young children reading in classrooms (e.g., Heap, 1985, 1991). Rather than assuming we already know what reading looks like, or when it occurs in classrooms – through some *a priori* theory or criteria – he focuses on the broad range of interactional activities that produce the ongoing activity as reading. For him, children learn what counts as reading criterially, by learning what counts as reading procedurally, in and through taking part in interactions where good reading is shown to be taking place. For example, Freebody and Freiberg (2001) show how in a moment of a parent-child interaction, where a child is reading a book aloud, the parent works to praise, instruct the child to sound the letters, correct a sounding, etc. In this moment of reading a document, the central interactional task is reading out loud a written text, and good reading is orientated to and produced as correct word-saying. In other contexts, say a university seminar on a specific article, in and through debating and discussing, students learn what good academic reading involves – finding fault, contrasting with other documents, quoting and referencing from this and other documents, reading footnotes, following up references, etc.

Such work focuses researchers' attention on unpacking, in detail, how specific tasks are undertaken.

Documents are not only paper-based, single-user, objects and ethnomethodological ethnographic research

has focused on how mundane and digital technologies shape the world of educational activities. Such research has focused on how people interact, work with, and draw on such document-related technologies as white- and blackboards, overheads, computer-based presentations, and videos. Rendle-Short's (2006) study of computer-science seminar presentations focuses on the academic monolog. She shows us how presenters work to interact with the audience not only through their talk, but also through their gestures, gaze, and bodily movements. Using a collection of videotapes of seminars and transcribing in great detail the moment-by-moment verbal and nonverbal behavior, she focuses on how presenters coordinate a range of documents and technologies as they seek to engage with the audience. As presenters talk, they time their slides to focus, support, and supplement the issues they are raising. The text-based and visual images on the slides work to illustrate some aspects of their ongoing talk, and presenters through their hand gestures work to direct the audiences' attention to a specific aspect or issue. In this way, we begin to see how, what appears as a relatively simple task – giving a seminar – is saturated with a complex range of interactional work. The documents the presenters work with, be it notes available to them, or slides available to all, are brought to life and shaped through the concerted, moment-by-moment, organization of talk, gesture, and technology.

Such ethnomethodological ethnographic research findings can seem focused on the quite obvious, in that it can show us what we take for granted. However, researchers in this tradition argue that it is only through an understanding the details of what we actually do (the details of which can largely escape our notice), over what people tells us they do, or what we think they do, that we can then develop theories or interventions that are actually relevant to current practice.

Conclusion

Given the ubiquity of paper and computer-based documents in contemporary social and organizational life, the range of potential types of documents open to qualitative analysis is immense. However, the primary sources for educational research are policy documents (be they national, regional, or local), textbooks, and internal organizational documents; the secondary sources are from the academic canon. Documents in any broader sense – be they blogs, diaries, magazines, newspapers, or websites – are underused and underanalyzed. As noted above, the lack of focus on documents is not restricted to qualitative educational research, but is endemic to the social sciences. On the one hand, this is quite surprising, given the centrality of documents to everyday experiences, the readiness-to-hand of document-based raw data, and that it does not have to

require large amounts of time in the field collecting data. However, ideas around hearing peoples' voices, witnessing things with your own eyes, have been increasingly positioned as a central requisite of authentic and high-quality qualitative research and increasingly involve documents.

Nevertheless, in relation to critical discourse analysis style studies of education, Rogers *et al.* (2005) see the turn away from a focus solely on written documents toward a focus on talk as a positive direction, in which new theoretical, methodological, and analytic possibilities emerge. Notwithstanding these potentials for development alongside the existing tradition of ethnographic work focusing on the use, role, and function of documents-in-action, we are still left with a relatively limited corpus of work that focuses on the social and educational lives of documents. Documents shape, and are reflexively shaped by, our perceptions, interactions, institutions, policies, and society. They are central in the production, reproduction, and transformation of our educational landscapes. As such, they deserve a more sustained and systematic analytic focus.

See also: Classroom Ethnography; Conversational Analysis; Discourse Analysis; Ethnography; Ethnomethodology in Education Research.

Bibliography

- Castanheira, M. L., Crawford, T., Dixon, C., and Green, J. (2001). Interactional ethnography: An approach to studying the social construction of literate practices. *Linguistics and Education* **11**, 353–400.
- Dixon-Woods, M., Bonas, S., Booth, A., *et al.* (2006). How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research* **6**, 27–44.
- Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison*. New York: Pantheon.
- Freebody, P. and Freiberg, J. (2001). Re-discovering practical reading activities in schools and homes. *Journal of Research in Reading* **24**, 222–234.
- Green, J. L. and Meyer, L. A. (1991). The embeddedness of reading in classroom life: Reading as a situated process. In Baker, C. D. and Luke, A. (eds.) *Towards a Critical Sociology of Reading Pedagogy: Papers on the XII World Congress on Reading*, pp 141–160. Amsterdam: John Benjamins.
- Heap, J. L. (1985). Discourse in the production of classroom knowledge: Reading lessons. *Curriculum Inquiry* **15**, 245–259.
- Heap, J. L. (1991). A situated perspective of what counts as reading. In Baker, C. D. and Luke, A. (eds.) *Towards a Critical Sociology of Reading Pedagogy: Papers on the XII World Congress on Reading*, pp 103–139. Amsterdam: John Benjamins.
- Mehan, H. (1993). Beneath the skin and between the ears: A case study in the politics of representation. In Chaiklin, S. and Lave, J. (eds.) *Understanding Practice: Perspectives on Activity and Context* pp 241–268. Cambridge: Cambridge University Press.
- Noblit, G. W. and Hare, R. D. (1988). *Meta-Ethnography: Synthesizing Qualitative Studies*. Newbury Park, CA: Sage.
- Potter, J. and Wetherell, M. (1987). *Discourse and Social Psychology*. London: Sage.
- Rafalovich, A. (2004). *Framing ADHD Children: A Critical Examination of the History, Discourse, and Everyday Experience of Attention Deficit/Hyperactivity Disorder*. Latham, MD: Rowman and Littlefield-Lexington Books.
- Rendle-Short, J. (2006). *The Academic Presentation*. Aldershot: Ashgate.
- Rogers, R. (ed.) (2004). *An Introduction to Critical Discourse Analysis in Education*. Erlbaum: Mahwah, NJ.
- Rogers, R., Malancharuvil-Berkes, E., Mosley, M., Hui, D., and O'Garro Joseph, G. (2005). Critical discourse analysis in education: A review of the literature. *Review of Educational Research* **75**, 365–416.
- Sandelowski, M. and Barroso, J. (2007). *Handbook for Synthesizing Qualitative Research*. New York: Springer.
- Thomas, J. and Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology* **8**, 45.
- Weber, R. P. (1990). *Basic Content Analysis*, 2nd edn. Newbury Park, CA: Sage.

Further Reading

- Livingston, E. (1995). *An Anthropology of Reading*. Bloomington, IN: Indiana University Press.
- Prior, L. (2003). *Using Documents in Social Research*. London: Sage.
- Scott, J. (1990). *A Matter of Record: Documentary Sources in Social Research*. Cambridge: Polity Press.
- Smith, D. E. (1990). *Texts, Facts and Femininity: Exploring the Relations of Ruling*. London: Routledge.

Ethnography

M Hammersley, The Open University, Milton Keynes, UK

© 2010 Elsevier Ltd. All rights reserved.

The terms 'ethnography' and 'ethnographic' have been in common use within research on education since at least the 1950s (Spindler, 2000). Initially, they had a relatively specific meaning, referring to anthropological research that focused on the process of cultural transmission within schools, studied against the background of local communities, and usually with a particular interest in the experience of minority or subordinated groups. However, from the 1960s onward, within education and beyond, the word 'ethnography' began to be used in other disciplines and in a wider range of ways, with some other methodological labels treated as near-synonyms – including 'case study', 'field research', 'interpretive inquiry', and 'qualitative method'. Moreover, where previously the label had referred to studies employing participant observation over relatively long time periods, now the data collection tended to be of shorter duration and some studies relied primarily if not exclusively upon in-depth interviews. Diversity in orientation also increased: some of this more recent ethnographic work formed part of evaluation projects, some drew on sociological theory and on sociolinguistics, and some of it was influenced by Marxism, feminism, and other critical and activist approaches – though critical and activist strands can be traced within US anthropology of education from quite early in its development (Yon, 2003). There was also dispersion in geographical terms: prior to 1960 almost all ethnographic work in education was carried out by US scholars, but from that decade onward there was considerable growth in this kind of research elsewhere, most notably in the UK, but also in Australia and on the European continent (for the case of Sweden, see Larsson, 2006). The result is that there is now a huge amount of work that labels itself as ethnography.

Over the second half of the twentieth century, there was also diversification within anthropological research in education, for example with some of it developing an international comparative dimension, while another flourishing genre focused on language-use, stimulated by the ethnography of communication and other work in sociolinguistics. A further influential strand developed within the Chicago tradition of US sociology, with studies of schoolteachers (for instance, Lortie, 1975; and the work of Becker, see Burgess, 1995), and of students (e.g., Geer *et al.*, 1961). From the 1960s onwards, outside of anthropology and sociology, there were many investigations of life in classrooms (Jackson, 1968), some of which could be described as ethnographic. In the UK, some early ethnographic work focused on the effects of streaming or tracking in secondary

schools (e.g., Lacey, 1970). Subsequently, in the 1970s and 1980s, developments around the new sociology of education led to a considerable range of studies concerned with both teacher and student perspectives in the context of classroom interaction (see Hammersley, 1999). Under the influence of feminism and anti-racism, this work spread out into a range of investigations concerned with inequitable treatment within the education system. More recently, in the UK and elsewhere, much ethnographic work has been concerned with investigating the character and consequences of managerialist forms of educational policymaking (for instance, Woods *et al.*, 1997; Troman *et al.*, 2006). There has also been a small amount of work that has extended the focus beyond teachers and students in state-funded mainstream schools, for example, looking at religious schools (Peshkin, 1988), at vocational training (Atkinson *et al.*, 1981), and at educational processes in non-institutional settings (Delamont, 2006).

The Meaning of Ethnography

In etymological terms, 'ethnography' means writing about a people, and came to refer to producing an account of the way of life of a particular community or society. In early twentieth-century anthropology, what was aimed at was a descriptive account that captured a distinctive form of social organization or culture. Initially, ethnography was contrasted with ethnology, an influential form of nineteenth-century anthropological work which focused on the historical and comparative analysis of societies, and was usually based on accounts produced by travelers and missionaries. Over time, the term 'ethnology' fell out of favor, and 'ethnography' came to refer to theoretical interpretation of cultures on the basis of firsthand investigation carried out by anthropologists and other social scientists themselves. Moreover, the word has a double meaning, referring both to a form of research and to the product of that research: in other words, ethnography as a practice produces ethnographies as published accounts. A contrast has also sometimes been drawn between doing ethnography and using ethnographic methods. This distinction has been employed by some anthropologists in an attempt to mark off their own practice from what passes for ethnographic work within sociology and evaluation studies (Wolcott, 1999). However, the distinction raises a more general issue: that the various methodological ideas and strategies associated at any particular time with

ethnography have been by no means universal or fixed. So, if we look back from the twenty-first century a hundred years or more, across different disciplines and different countries, we find considerable variation in the nature of ethnographic principle, practice, and products. Even anthropological research has changed considerably over this period, and become more diverse in methodological terms. For many anthropologists in the past, ethnography required living with a group of people for an extended period – for a year or even several years – in order to document and explain their distinctive way of life, the beliefs and values integral to it, the social institutions (including those relating to education) characteristic of it, and so on. However, today, ethnography has come to be defined in more specific methodological terms as the use of participant observation and/or in-depth interviewing; though even this meaning has been breached by the notion of virtual or internet ethnography, where the primary form of data is online rather than face-to-face (Silva, 2002).

Despite variation in meanings given to the word, in practical terms of method ethnography usually involves many of the following features:

1. People's actions and accounts are studied primarily in everyday contexts, rather than under conditions created by the researcher – such as in experiments or in highly structured interview situations. In other words, research takes place in the field or is naturalistic in character.
2. Data are gathered from a range of sources. While participant observation and/or relatively informal conversations are usually the main ones, documents, artifacts, and even statistical data may also be employed.
3. Data collection is usually relatively unstructured, in the sense that it does not involve following through a fixed and detailed research design setup at the beginning. Nor are the categories that will be used for interpreting what people say or do generally built into the data collection process itself – via prestructuring of observation, interviews, or documentary analysis. Instead, they are to be discovered or constructed during the course of inquiry.
4. The focus is usually on a small number of cases, perhaps a single setting or group of people, occasionally just one person, as with some forms of life history and auto-ethnography (Reed-Danahay, 2001). Sometimes a larger number of cases are studied, notably where these are small-scale, for example, school lessons.
5. The analysis of the data involves interpretation of the meanings and functions of human actions, and usually also how these are implicated in local, and wider, contexts. What are produced by ethnographic analyses, for the most part, are verbal descriptions, explanations, and theories; quantification and statistical analysis play a subordinate role at most.

There have been continuing disputes about whether any of these features is essential, and about the relationship with other work of various kinds that is placed today under the broader heading of 'Qualitative method'. There is no definitive answer to these questions, they are a matter of dispute.

Ethnography as a Distinctive Methodological Orientation

As a practical approach, ethnography is not far removed from the means that we all use in everyday life to make sense of our surroundings. However, it involves a more deliberate and systematic attitude, as with any form of research, and also a distinctive mentality. This can perhaps best be summarized as starting from an attempt to make the strange familiar – in the sense of finding intelligibility and rationality in what is initially inexplicable – and, at the same time, making the familiar strange – by suspending those background assumptions that immediately give apparent sense to much of what we experience, at least in contexts with which we are well-acquainted (Hammersley and Atkinson, 2007).

Over the course of its development, ethnography has been influenced by a range of methodological and theoretical movements. Early on, within anthropology, it was shaped by German historicist ideas about the difference between the human and the natural sciences, by folk psychology, but also by nineteenth-century positivism. Subsequently, in the form of the case study approach of the Chicago School, it was informed by philosophical pragmatism, while in more recent times phenomenology, hermeneutics, structuralism, and post-structuralism have all played an important role. As noted earlier, it has also been influenced by various critical orientations: Marxism, feminism, anti-racism, disability activism, and queer theory. The forms that ethnographic work has taken, and the particular influences operating on it, have varied across different fields and different countries, as well as over time.

Despite these diverse influences, at an abstract level ethnography tends to be characterized by a few distinctive methodological ideas about the nature of the social world and how it can be understood. These have influenced how ethnographers have studied educational structures and processes. They can be summarized as follows:

1. Human behavior is not an automatic product of either internal or external forces or stimuli. People's responses to the world are constructed and reconstructed over time and, across spaces, in ways that reflect the biographies and socio-cultural locations of the actors, how they interpret the situations they face, and how these situations develop over time.

2. There are diverse cultures that can inform human behavior, and these vary not just between societies or local communities but also within them.
3. Human social life is not structured by fixed, law-like patterns, but displays emergent processes of various kinds that involve a high degree of contingency.

While these generic ideas have informed much ethnographic work, they have been interpreted in a variety of ways, and have generated some tensions.

Tensions Within Ethnography

In recent times, there has been some dispute over the character of the phenomena that ethnographers study and how they should study them. There are a number of dimensions to this.

One is a tension between naturalism and constructionism. For the first, the task of ethnography is to document stable cultures, patterns of social interaction, institutions, and so on, as they exist in the world independently of the researcher. By contrast, constructionism is concerned with the interactional or discursive processes whereby cultures and institutions are ongoingly, and contingently, produced and sustained. Indeed, in its more radical forms, constructionism treats the phenomena studied by ethnographers as effectively constituted in and through the research process itself, and especially through the process of writing (see Clifford and Marcus, 1986).

These disputed assumptions about the nature of the social world are closely linked with ideas about how we can understand it. And here too significant differences in approach come to the surface. One of these concerns the nature of context. On the one hand, some ethnographers focus on the details of what happens in specific, small-scale contexts on particular occasions, and perhaps on how participants themselves define these contexts. On the other hand, there are ethnographers who insist on the need to locate what has been studied within a theoretical understanding of some larger social whole. Over the past few decades there has been a trend towards more micro-focused ethnographies (Erickson, 1992), but there has long been, and remains, a counter-tradition which stresses the need to locate what is studied in a wider context, whether a national society or the global pattern of social relations (Burawoy *et al.*, 2000).

Parallel to this have been criticisms of much ethnography for being ahistorical. It is sometimes portrayed as preoccupied with describing and explaining what happens at some particular place and time, thereby neglecting longer-term trends. One response to this has been to advocate longitudinal ethnographies, for example following the development of students' lives over several years as they traverse the education system, investigating the

changes experienced, the adaptations made, and the outcomes (see Pollard, 2007). Also relevant here are restudies (e.g., see Smith, 1983; Burgess, 1987), and life history investigations that trace, for instance, the patterns of teachers' careers and the factors shaping these (Goodson and Sikes, 2001).

A third tension within ethnographic thinking is between a focus on the unique and the use of comparative analysis: between seeking to study the distinctive aspects of particular cases and being concerned with producing generalisations or building theories. Ethnographers vary considerably in their position on this dimension, but most seek to satisfy both demands simultaneously in one way or another, to at least some degree. The concept of thick description (Geertz, 1973) represents one sort of trade-off, where theories are primarily means for understanding what is going on in particular cases; but are developed in the course of this. Toward the other end of the spectrum are grounded theorizing and analytic induction, where the intended product of ethnographic work is some kind of general theory, albeit evidenced through data from particular cases (see Hammersley, 1989, 2008).

Another issue concerns whether the emphasis is on description or explanation. For some, the primary ethnographic task is explicating the perspectives, or cultural orientations, of the people being studied in their own terms. For others, the goal is to explain why people see the world and act in the ways that they do, and perhaps also to account for the consequences of this. The first approach emphasizes the role of careful description, of understanding the meanings people give to the situations they face and to their own identities, perhaps even seeking to amplify their voices. By contrast, the second often produces accounts that raise questions about the validity of people's beliefs about themselves and their world. This may involve explaining their attitudes and actions in terms of causal factors whose existence or significance they do not acknowledge, or even explicitly deny. Sometimes these two approaches have been applied selectively within the same study, with the perspectives of some actors being presented as at least partly representing genuine understanding of the world while other aspects of their perspectives, or the perspectives of other people, are treated as ideological (see Hammersley, 1998).

A related variation concerns attitudes toward the distinction between appearance and reality. Some ethnographers see their work as challenging official or public appearances, the fronts people display, in order to find out what people really believe or what is really going on. A somewhat different orientation involves viewing social life as a matter of socio-cultural performance, with the task being to study the processes or strategies by which people bring off particular performances on particular occasions (Bloome *et al.*, 1989; Atkinson and Coffey, 2002). From this second point of view, there is no true

or fundamental reality behind appearances, only constitutive, interactional processes that generate one set of phenomena rather than another.

Even for those ethnographers who place emphasis on documenting people's perspectives, there are questions about the nature of understanding. How far it is ever possible or necessary for ethnographers to understand participants' perspectives 'from the inside?' It has been suggested that this involves reducing the Other to the Same, forcing what is different into terms that are familiar. At the same time, ethnography has also sometimes been accused of Othering, of rendering non-Western societies or marginalized groups within Western societies exotic and alien, a criticism that parallels Said's discussion of orientalism (Said, 1978). Closely related are criticisms of the totalizing orientation of much older ethnography, where cultures are described as if they were objects in the world that are internally homogeneous, and as if membership of a culture determined everything of importance about any individual person.

A further dimension of difference in orientation concerns whether an appreciative or a critical stance is judged to be most appropriate. In some influential forms, ethnography has involved a concern to capture the beliefs and actions of the people being studied in such a way as to minimize the effects of the research process and of the attitudes of the researcher. Here, ethnography was usually distanced from any concern with practical improvement or social intervention, and therefore adopted a nonjudgmental or appreciative orientation (Matza, 1969). However, in the mid-twentieth century there developed forms of applied anthropology that treated ethnography as a basis for interventions designed to improve the lives of the people being studied. Later, some ethnographers adopted Marxist or other critical perspectives in which the phenomena studied were to be located within a political perspective that generated evaluations and recommendations for social change. At the same time, the impact of post-structuralism and postmodernism has challenged reliance upon political positions involving meta-narratives, notably but not exclusively Marxism, in favor of subordinating ethnographic work to local struggles, with one of the tasks being to liberate those repressed forms of knowledge that have been banished to the margins of conventional society. These developments have also raised doubts about, but in practice also sometimes reinforced, the idea that at least part of the ethnographic task is to give voice to those treated as low status or marginalized within particular societies and communities: for example, students or parents from minority ethnic groups.

Closely associated with some of these developments have been pressures to do ethnographic work *with* people rather than *on* them, along the lines of various participatory forms of inquiry or action research (Reason and Bradbury, 2006). In some cases this has built on a commitment to

advocacy by anthropologists, and on the notion of indigenous ethnography; while, elsewhere, it also derived from feminist and other approaches to research ethics which have challenged what is seen as the hierarchical relationship between researcher and researched in conventional forms of ethnography and other kinds of social research. However, there is a tension here not only with the commitment of older forms of ethnography to appreciation and understanding but also between subordinating research to participants' orientations and using it as a means of raising their consciousness in order to generate desirable social change. Within the field of education, there has been considerable work drawing on ethnographic methods that has been aimed at working with practitioners, or enabling the latter to do research themselves, notably under the banner of educational action research, but some of it also involving a critical orientation (e.g., see Gitlin *et al.*, 1989).

A final, related, trend worth mentioning is increasing pressure to recognize the extent to which, and ways in which, all research, including ethnography, plays a political role in the world. To some degree this began long ago with criticism of how anthropology was implicated in Western imperialism. In more recent times, the concern with the politics of ethnography has become much broader, reflecting the influence of new social movements of various kinds, and wider socio-political circumstances. For some commentators, the whole enterprise of research is political through and through, in the sense that it cannot but involve reliance on value assumptions, and these cannot but reflect the identity, commitments, and social location of the researcher as a person. This runs against earlier forms of ethnography where research was treated as concerned simply with producing objective scientific knowledge about diverse communities and cultures, an orientation that is now regarded by many, though not all, ethnographers as simply an ideological disguise for political interests that serve the status quo.

Some Further Developments

As noted earlier, ethnography refers not just to a process of inquiry but also to a particular type of product: to the sort of account generated by ethnographic research. Prior to the early 1980s, the task of writing up ethnographies was given relatively little attention in the methodological literature. Most of the focus was on problems surrounding data collection and analysis. However, in the past three decades there has been considerable interest in this topic, not just from a practical point of view but also in terms of analyzing how ethnographic accounts represent or effectively constitute the social contexts and people investigated. Epistemological, political, and ethical concerns are intermingled in what has come to be seen as a crisis of representation (see Hammersley, 2008).

Developments in technology have also had an important impact on ethnographic work. In the second half of the twentieth century, the availability of increasingly portable audio- and video-recorders meant that fieldnotes came to play a subordinate role in much ethnography. Furthermore, the use of video-recording has built on earlier developments in visual ethnography that employed photographs and film. These technologies may have encouraged the spread of an increasingly micro-focused concern with the details of what is said and done on particular occasions.

Advances in computer technology, and in software for processing qualitative data, are another important area of development, one where there is disagreement about whether the technology serves or distorts ethnographic practice. What seems clear, though, is that digitization of data, and the increased capacity of computers to handle multimedia material, will open up considerable opportunities for ethnographers, as well as no doubt also raising new problems, or old problems in new forms. Closely related here is the development of the internet, and the opportunities that this provides not just as a source of information but also for the investigation of virtual communities.

Finally, it is worth mentioning a significant feature of the changing environments in which ethnographers seek to carry out their work. Both anthropologists and sociologists have encountered increasing barriers in gaining access to settings in many societies, and this includes those relevant to education. These stem from a variety of factors, among which are managerialist forms of regulation within both privately owned and publicly funded organizations, and increasing commercialization. A related external factor is increasing ethical regulation, the ethical codes on which this is based often assuming a model of research that is at odds with both the theory and the practice of ethnography (Lincoln and Tierney, 2004).

The rise of the notion of evidence-based policymaking and practice, and associated demands that educational research be reformed to provide more effective evidence about what works, has become a significant feature of the environment in which ethnographic inquiry is now carried out, in many countries. What is involved here is not just a push toward more applied kinds of work but also the imposition of distinctive research criteria. Even where randomized controlled trials are not treated as the gold standard, the methodological orientation associated with the evidence-based practice movement in education is at odds with ethnography in some key respects, for example, in demanding closely specified initial research designs. Of related significance here is increasing strategic management of research and modes of research training within universities, which are also sometimes informed by a similar methodological model.

In conclusion, then, term 'ethnography' now refers to a range of flourishing approaches within educational

research and social science. Much is shared among these approaches, but there are also some significant differences and tensions, as well as important challenges.

See also: Action Research in Education; Classroom Ethnography; Computer Assisted Qualitative Data Analysis; Participant Observation.

Bibliography

- Atkinson, P. and Coffey, A. (2002). Revisiting the relationship between participant observation and interviewing. In Gubrium, J. F. and Holstein, J. A. (eds.) *Handbook of Interview Research*, pp 801–814. Thousand Oaks CA: Sage.
- Atkinson, P., Shone, D., and Rees, T. (1981). Labouring to learn: Industrial training for slow learners. In Barton, L. and Tomlinson, S. (eds.) *Special Education: Policy, Practices and Social Issues*, pp 235–257. London: Paul Chapman.
- Bloome, D., Puro, P., and Theodorou, E. (1989). Procedural display and classroom lessons. *Curriculum Inquiry* 19(3), 265–291.
- Burawoy, M., Blum, J., George, S., et al. (2000). *Global Ethnography*. Berkeley, CA: University of California Press.
- Burgess, R. G. (1987). Studying and restudying Bishop McGregor School. In Walford, G. (ed.) *Doing Sociology of Education*, pp 51–79. Lewes: Falmer.
- Burgess, R. G. (ed.) (1995). *Howard Becker on Education*. Milton Keynes: Open University Press.
- Carr, W. and Kemmis, S. (1986). *Becoming Critical*. Lewes: Falmer.
- Clifford, J. and Marcus, G. (eds.) (1986). *Writing Culture: The Poetics and Politics of Ethnography*. Berkeley, CA: University of California Press.
- Creese, A. (2008). Linguistic ethnography. In King, K. A. and Hornberger, N. H. (eds.) *Encyclopedia of Language and Education*, 2nd edn, vol. 10, pp 229–241. New York: Springer.
- DeCastell, S. and Walker, T. (1991). Identity, metamorphosis, and ethnographic research: What kind of story is *Way with Words*? *Anthropology and Education Quarterly* 22, 3–20.
- Delamont, S. (2006). The smell of sweat and rum: Teacher authority in capoeira classes. *Ethnography and Education* 1(2), 161–175.
- Dicks, B., Mason, B., Coffey, A., and Atkinson, P. (2005). *Qualitative Research and Hypermedia: Ethnography for the Digital Age*. London: Sage.
- Erickson, F. (1992). Ethnographic microanalysis of interaction. In LeCompte, M., Millroy, W., and Preissle, J. (eds.) *Handbook of Qualitative Research in Education*, pp 201–225. San Diego, CA: Academic Press.
- Geer, B., Hughes, E. C., Strauss, A., and Becker, H. S. (1961). *Boys in White*. Chicago, IL: University of Chicago Press.
- Geertz, C. (1973). On thick description. In Geertz, C. (ed.) *The Interpretation of Culture*, pp 3–32. New York: Basic Books.
- Gitlin, A. D., Siegel, M., and Boru, K. (1989). The politics of method: From leftist ethnography to educative research. *Qualitative Studies in Education* 2(3), 237–253.
- Goodson, I. and Sikes, P. (2001). *Life History in Educational Settings*. Buckingham: Open University Press.
- Hammersley, M. (1989). *The Dilemma of Qualitative Method*. London: Routledge.
- Hammersley, M. (1994). Ethnographic writing. *Social Research Update* 5. Guildford: University of Surrey.
- Hammersley, M. (1995). *The Politics of Social Research*. London: Sage.
- Hammersley, M. (1998). Partisanship and credibility: The case of antiracist educational research. In Connolly, P. and Troyna, B. (eds.) *Researching Racism in Education: Politics, Theory and Practice*, pp 21–33. Buckingham: Open University Press.
- Hammersley, M. (ed.) (1999). *Researching School Experience: Ethnographic Studies of Teaching and Learning*. London: Falmer.
- Hammersley, M. (2000). *Taking Sides in Social Research: Essays on Partisanship and Bias*. London: Routledge.

- Hammersley, M. (2008). *Questioning Qualitative Inquiry*. London: Sage.
- Hammersley, M. and Atkinson, P. (2007). *Ethnography: Principles in Practice*, 3rd edn. London: Routledge.
- Hine, C. (2000). *Virtual Ethnography*. London: Sage.
- Jackson, P. (1968). *Life in Classrooms*. New York: Holt, Rinehart and Winston.
- Jeffrey, B. and Walford, G. (2004). *Ethnographies of Educational and Cultural Conflicts: Strategies and Resolutions*. Oxford: Elsevier.
- Kelle, U. (2004). Computer-assisted qualitative data analysis. In Seale, C., Gobo, G., Gubrium, J. F., and Silverman, D. (eds.) *Qualitative Research Practice*, pp 473–490. London: Sage.
- Lacey, C. (1970). *Hightown Grammar*. Manchester: Manchester University Press.
- Larsson, S. (2006). Ethnography in action: How ethnography was established in Swedish educational research. *Ethnography and Education* 1(2), 177–195.
- Lincoln, Y. S. and Tierney, W. G. (2004). Qualitative research and institutional review boards. *Qualitative Inquiry* 10(2), 219–234.
- Lortie, D. C. (1975). *School Teacher*. Chicago, IL: University of Chicago Press.
- Matza, D. (1969). *Becoming Deviant*. Englewood Cliffs, NJ: Prentice-Hall.
- Peshkin, A. (1988). *God's Choice: Total World of a Fundamentalist Christian School*, 2nd edn. Chicago, IL: University of Chicago Press.
- Pollard, A. (2007). The identity and learning programme: "Principled pragmatism" in a 12-year-longitudinal ethnography. *Ethnography and Education* 2(1), 1–19.
- Reason, P. and Bradbury, H. (eds.) (2006). *Handbook of Action Research*, Concise edn. London: Sage.
- Reed-Danahay, D. (2001). Autobiography, intimacy and ethnography. In Atkinson, P., Delamont, S., Coffey, A., Lofland, J., and Lofland, L. H. (eds.) *Handbook of Ethnography*, pp 407–425. London: Sage.
- Said, E. (1978). *Orientalism*. New York: Pantheon.
- Silva, R. (2002). Doing electronic educational ethnography: Issues of interpretive quality and legitimacy in virtual reality. In Walford, G. (ed.) *Debates and Developments in Ethnographic Methodology*, pp 123–150. Amsterdam: JAI Press.
- Smith, L. M. (1983). *Innovation and Change in American Education: Kensington Revisited*. Institute of Education, Washington University.
- Spindler, G. (ed.) (1982). *Doing the Ethnography of Schooling*. New York: Holt, Rinehart and Winston.
- Spindler, G. (ed.) (2000). *Fifty Years of Anthropology and Education 1950–2000: A Spindler Anthology*. Mahwah NJ: Erlbaum.
- Troman, G., Jeffrey, B., and Beach, D. (2006). *Researching Educational Policy: Ethnographic Experiences*. London: Tufnell Press.
- Walford, G. (ed.) (2008). *How to Do Educational Ethnography*. London: Tufnell Press.
- Wolcott, H. F. (1999). *Ethnography: A Way of Seeing*. Walnut Creek, CA: Altamira.
- Woods, P., Jeffrey, B., Troman, G., and Boyle, M. (1997). *Restructuring Schools, Reconstructing Teachers*. Buckingham: Open University Press.
- Yon, D. A. (2003). Highlights and overview of the history of educational ethnography. *Annual Review of Ethnography* 32, 411–429.

Further Reading

- Hammersley, M. and Woods, P. (eds.) (1984). *Life in School: The Sociology of Pupil Culture*. Milton Keynes: Open University Press.
- Hargreaves, A. and Woods, P. (eds.) (1984). *Classrooms and Staffrooms: The Sociology of Teachers and Teaching*. Milton Keynes: Open University Press.

Ethnomethodology in Education Research

D Macbeth, The Ohio State University, Columbus, OH, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Formal analysis – A gloss for the expert practices of social science analysis that aim to reproduce either or both the measurement program of the natural sciences, or its discoveries of formal causative structures. But unlike those achievements in the natural sciences, formal analysis in social science – whether statistical or theoretic – must specifically ignore the endogenous organizations of the affairs it aims to formally render. The hearable (indexical) contingency of an interview, for example, is lost to every scheme of codes.

Indexical expressions – Expressions whose definite sense and meaning are tied to the occasions of their use. Natural language is shot through with indexical expressions. This means that common understanding is achieved, and has an analytic foundation. Speaking and listening are analytic exercises. In our unrelieved use of these expressions, we achieve the “rational properties of indexical expressions,” and worlds in common.

Reflexivity – Ethnomethodological reflexivity refers to relations of co-constitution. Questions, for example, are reflexive to answers: We know what was said was a question because we hear an answer next. Its canonical formulation is the “essential reflexivity of accounts,” or how our accounts of the world are constitutive of the affairs they render accountable. Among other interests, a constitutive reflexivity spells the end of representational models of language. Talk does not represent; it acts.

Situated action – Action that is not situated, that is, it is not formulated in local settings that give it cause and sense, is difficult to imagine. For EM, situated action is a unifying formulation. It is part of an argument for the local study of cases-in-context.

Introduction

This article takes up a certain intellectual history in matters of educational study. It is certainly not the only one, or even, perhaps, a central one, for understanding the central tendencies of the research literature. However, it is

an important one for making sense of where we have come in understanding the relationship of social science to professional education, and what claims, promises, and offers of assistance can usefully, sensibly be made from the one to the other. There are two tasks it hopes to accomplish: To bring into view a period of great conceptual innovation in our thinking about the ordered character of social worlds for a current generation of young scholars who may know it less well. Those innovations struck to the core of the promise of a social science, and became the curriculum for the generation of education scholars who wrote the qualitative turn in education in the 1970s. The second is to clarify what is persistently relevant about those conceptual innovations in the particulars of ethnomethodology (hereafter, EM). EM is not well understood in the textbook literature, although it has been a central participant to an invisible college of commentaries, criticisms, and alternative proposals that have rewritten social science and educational study over the last 40 years.

There are several instructive accounts of EM in the sociological literature (see especially Heritage, 1984; Lynch, 1993) and in education (see Baker, 1997; Watson, 1992). But the motivated reader should also turn to its founder, Garfinkel (1967, *passim*) and his 50-year corpus of work. Garfinkel has been articulate and detailed about the program he intends, although by many accounts his articulations do not make for easy reading. Included is the story of his coinage of the term ethnomethodology, having to do with the intellectual currents of the early 1950s and the developing fields of the ethnosciences, as in ethnobotany, ethno-astronomy, and ethno-pharmacology. Garfinkel found their demonstrations kindred to his interests in how ordinary worlds owned practices of disciplined inquiry long before the first graduate student was ever minted.

The ethnosciences were organized by the rubrics of modern science (astronomy, and the like). But Garfinkel was taking interest in methodic practices far more quotidian, and foundational. Ethnomethodology is a covering statement for the endlessly diverse member-methods that give form and recurrence to the most ordinary social-organizational things in the world (e.g., the order of a queue at the supermarket, or a classroom lesson), and how those things are jointly produced, and thus contingent, and therefore achieved. The members here are cultural members, and at every turn, social life and order show our methodic actions in concert. In the measure, EM articulates a

program of studies that is quite different from the culture and conceptualizations of social science and educational research as we know them.

By many accounts, EM is a puzzle, a source of confusion for some, even consternation for others. A reputational penumbra remains from its early reviews having to do with vexing texts, odd experiments, and relentless though perhaps penetrating discontents with normative social science. All could be true. However, in addition, the most troublesome readings seem to follow from the very term ethnomethodology itself.

EM tends to be read as yet another professional methodology, and thus another proposal for technical methods that might bring distinction to the field. Such a reading takes EM as something to apply, on the promise that this application will serve our purposes better than another. The reading thus relies on our familiar expectations for scientific methodology, and the place of science in society as a site of high technical expertise. It is a reading that preserves the promise of science and its appointments, and thereby misses entirely EM's commentary on those self-same ambitions.

EM refers not to scientific or social scientific methods. It proposes a very different understanding of a world that already owns methods of inquiry and analysis. In full, EM is the topic for EM studies. EM is EM's object, interest, and task to describe. EM sights a world of diverse methods of inquiry, action, and instruction, from laboratory science to bird watching to learning a second language. These are the first methods on the scene, and thus it is their practitioners – cultural members – who are our first methodologists. Yet, to say that EM is the topic of EM inquiries begs the question of how those inquiries might be organized. The question returns us to the presumptive authority of formal methodology, as though an inquiry could not be credible unless its methods were regular, repeatable, and vetted in advance. It is this very culture of methodism that EM sets aside. One might agree or not, but if we understand how EM is talking about methods, we understand how it is saying that we proceed with the life of the quarry in mind. Our professional-study methods, whether interviews, audio tapes, or document collections, are measured to our occasions of inquiry.

In this light, the lunchroom at noon takes on a methodic orderliness. So also, for teaching quadratic equations and poetic forms. EM writes a comprehensive view of social order as a methodic, praxeological assemblage, and a proposal to study its diverse tasks and occasions for what is methodic and disciplined about them (Garfinkel, 1967). On the face of it, its relevance for educational studies is quite substantial. For a world of disciplined practices, pedagogy is a central exercise. Competent worlds – worlds competent to storytelling and computer games – must already own methods for teaching

and learning the orderly affairs they display. How they do is perhaps the great ignored topic of the educational research literature.

The Mid-Century

EM was a development of mid-century sociology. Harold Garfinkel was party to an extraordinary generation of conceptual innovation in matters of social science, including critical commentaries on its very possibility. The registers ranged from sociology to anthropology to philosophy, and how they were all unavoidably joined at the hip of natural language study. Lists will not do those developments justice, and space will not permit a fair accounting, but between the mid-1950s and early 1970s, Anglo-American social science experienced a profusion of excavations, proposals, and demonstrations from Herbert Blumer to Erving Goffman to Clifford Geertz, and from Thomas Kuhn to Charles Taylor to Peter Winch and Wittgenstein.

From these various directions, the formal structures and distinctions that underwrite the natural-science model for social science became the objects of critical inquiry that yielded sociology's qualitative turn of the 1950s and 1960s. Central to the rethinking was the question of what purchase the model could have for understanding meaningful social worlds (or science, for that matter) in any actual case. One need not insist on understanding actual cases to see how the familiar empiricism of social science and educational research can say little, if anything, about them. As a familiar example, for all kinds of practical administrative purposes, it can be very useful to know that the average American family produces 2.2 children. For program-planning purposes, for the professional administrative offices of modern life, this can be very useful stuff to know; and it is entirely descriptive, although in a peculiar way: It tells us – and can tell us – nothing of any actual family. It tells us nothing of the formative histories that yield its aggregate finding. It can be of no use for understanding the very affairs it reports (though the report can be of interest to other affairs). For these reasons, should we take interest in the affairs that produce such data in the aggregate, we will have to look to them, and do our looking in a very different fashion.

Within social science, Garfinkel's EM was perhaps the most analytically radical proposal among the commentaries, radical in that it recast the exercise of sociological analysis and its relationship to the worlds we study. The formulation of member methods broke the conceptual monopoly on where and how we might find methods of inquiry, and who a methodologist or analyst could be. His work drew on an eclectic yet thematic collection of prior work, in Husserlian and Heideggerian phenomenology, conversations and readings with Alfred Schutz and Aaron Gurwitsch,

readings in philosophical pragmatism, the liminal works of Merleau Ponty, the formal architectures of Talcott Parsons, his advisor at Harvard, alongside the sociological canon. He set out to topicalize, and then write an alternate to sociology's consensus project of formal, synthesizing analyses whose aim and achievement would be expressed in synoptic theory. Parsons' theory of social action became an exemplar for the explication of what Garfinkel refers to as "formal analysis" and its identifying exercise of "generic representational theorizing" (Garfinkel, 1996).

As for the promise of formal analysis and its synoptic field of view – the macro-end of the familiar macro–micro divide, and the confident panoptics it delivers – it was observed in return that every observation of a distant landscape is a local reckoning. To see a learning disability, best practice, or hegemonic complicity is to collect local reckonings by the basketful, to rely on them, arrange them, submit them to certification procedures, and render them as practical evidences of a structure that was there all along. When we see this work, the comforts of distinctions such as form and content, structure and function, and macro and micro, collapse into fields of reflexive relations of co-constitution. Content becomes constitutive of form. Structure lives in function's local occasions. Rather than genealogies, we find constitutive relations. Rather than causal chains, we find demonstrable sense. (Notions of reflexivity are now familiar in the literature. We have the reflexive practitioner and the reflexive ethnography. These formulations trade on the wisdoms of the early-moderns, who counseled know thyself. The counsel is beyond reproach. We should. But the reflexivity of constitutive relations is quite different, as a topic and task of analysis (Garfinkel, 1967).)

This is analytically radical fare, in that to take it seriously is to re-think our received analytic appointments and ambitions. Central to this account is the place of constitutive detail for the analysis and understanding of social action, or what Garfinkel refers to as the concreteness of the plenum (Garfinkel, 1996). Husserl and Schutz spoke of the *lebenswelt*. Each refers to the life world, the profusion of sense, meaning, action, and others that fills our lives in the natural attitude of everyday life. Neither idiosyncratic nor subjective, the plenum refers to the meaningful character of evident – and thus objective – worlds. EM proposes to describe and understand the production of this public, witnessable order. Yet it is a central premise of our familiar research methods that the social world as we find it naturalistically, in its mundane presenting forms as, for example, the chatter of a classroom during third period, is without useful analytic possibilities. If we are to make good on the promise of a social science, we cannot work with it that way, in its actual, temporal durations and material productions. The contextual coherence of students engaged in chit chat defeats the promise of science.

For that reason, we rely on formal methods to craft stable representations of those ordinary worlds, by the familiar devices of code and variable constructions, set in play within fields of mathematical – or theoretical – relations, that might then yield models or logics that can, in turn, reveal unseen order or structure. This is the family of practices Garfinkel refers to as formal analysis, and by the expression he means to topicalize the unspoken premise that there is no order in the concreteness of actual occasions. Order, on the formal account, lies elsewhere. But one need not doubt that there are indeed structures unnoticed or out of view – as in the vector of a flu transmission or the collapse of a housing market – in order to take interest in the local enactments that yield such things as vectors or collapses, or the work of finding them.

Conversation Analysis

The subsequent and closely aligned development of conversation analysis (CA) (Sacks, 1992; Sacks *et al.*, 1974) gave these arguments a vivid field of demonstrations, vivid in the profusion of circumstantial detail recorded in tapes and on transcript. Conversation shows itself as a primordial members' method. In the organization of natural conversation, turn by turn, we can see what such vernacular analyses could be. (The formative history and conceptual relations between EM and CA is a regular topic in the EM/CA literature, for example, Heritage, 1984.)

Sequential analysis gave detail to the proposition that meaning owns a practical–analytic fabric, as in how every next turn to a conversation – without exception – displays an understanding of the turn that has gone before, and shapes the sequential horizon, or context, of next turns (Sacks *et al.*, 1974). Indeed, to speak next, whether at the dinner table or in the classroom, is to evidence an understanding of the speaking just done, what action it produces, what horizon of relevant next actions it projects, and, of all things, where the ongoing turn might end, so that we might begin our own. To speak next is to analyze all these things, whose analyses are revealed in the production of an apt next turn, on time (Moerman and Sacks, 1971/1988). The very formulation of turn taking in classroom discourse studies owes a very large debt to these first studies.

EM studies thus aim to show that there is indeed order in the plenum, an ignored, constitutive orderliness. To understand and describe it is to have use for the circumstantiality of meaning's productions. The point is not that somehow one should have use for this worldliness, because it has been ignored. It is rather and only that the order and structure of ordinary worlds turns upon it. In our everyday lives, we do not remotely ignore this circumstantiality. We live by it.

Invisible Colleges and the College of Education

There developed around these initiatives, a cohort of students and colleagues taking interest in this very different vision of what social-science inquiry might be. EM argued the primacy of meaning for the production of order, structure, and recurrence. Besides, if meaning were the solar furnace of orderly worlds – if order follows meaning – it was not a great leap to see that order and its complements took root not in distal formal structures, but in local, situated ones, as meaning does. Such a reading is neither a critique nor a dismissal of history, as though the history of childhood, for example, does not show us great innovations in matters of what the cultural category of child means. It is rather that history writing is itself an exercise in Garfinkel's treatment of Mannheim's documentary method of interpretation, wherein fragments and shards are seen for larger forms, and larger forms organize our seeing of fragments and shards. There is no relief from this constitutive reflexivity. We take interest not to be skeptical or to pose as arbiters, but to see how it is done. (Historians take the measure of their own work.)

For the college of education, these developments began at a great distance, and their relevance was a generation away. The tracks lead in several directions, and the notion of invisible colleges is a useful one for framing them. Invisible colleges are those identifiable, yet indefinite, communities of thought and interest that underwrite disciplinary history. The phrase traces to the early-modern science scholars of the seventeenth century who produced, among other things, the British Royal Society. Invisible colleges are the social histories of disciplinary formations. The field of sociolinguistics was once one. So too were child language-acquisition studies and post-Kuhnian studies of science. The formative period of the qualitative turn in educational studies was also one. In 1986, Fred Erickson wrote the first chapter on qualitative research to appear in the *Handbook of Research on Teaching*. The (invisible) conversation was of course much older (as seen in an early EM study of education by Cicourel and Kituse, 1963).

For its first 50 years, education's conceptual-analytic orbit had everything to do with psychology. Its first borrowings were there, mediated by the aims and promises of Thorndike and his students. (See McDermott and Hood (1982), on the analytic dispositions that followed.) However, in the 1970s, the disciplines beyond psychology began to matter a great deal for a cohort of young scholars who would write the qualitative turn for educational studies. As it happened, their degrees were, in the main, elsewhere: in anthropology, linguistics, developmental psychology, and sociology. But their interests were intently educational. They were bringing their readings of what were then exotic literatures for the college of education, and insisting that there was

news to be found that had direct bearing on matters of classroom life and instruction. Their studies were empirical, rather than empiricist, and had everything to do with the study of cases. Not all were reading EM or conversation analytic studies, but many were. Some were reading them very closely while others were reading them into the mix of their graduate and postdoctoral curricula.

The list of this transformational cohort includes Courtney Cazden, Michael Cole, Fred Erickson, James Heap, Ray McDermott, Hugh Mehan, Marilyn Merritt, Elliot Mishler, and many of their students and colleagues. The list is hardly exhaustive, and there were contemporary developments in the British educational literatures, for example, the works of Wes Sharrock, Rod Watson, George Payne, David Hustler, David Francis, Stephen Hester, and others, all of whom were studying EM/CA with great energy, and bringing its program to classroom studies. (See Hester and Francis (2000) for a diverse collection of EM educational studies.) These lists only suggest the serious trading that was going on between the disciplines and the college of education. Intellectual history is made of such tradings, and EM was party to the mix (see Erickson, 1992).

Its influence in educational studies began to show itself in formulations of situated action (Cicourel *et al.*, 1974), sense making and ethnographic adequacy (McDermott, 1976), constitutive ethnography (Mehan, 1979), situated instruction (Cook-Gumperz, 1977), in a growing interest in the sequential analysis of classroom discourse (McHoul, 1978) and, more generally, in the most unremarkable and unremarked upon order of the room. The everyday life of classroom education came into view as an un-examined plenum. Studies of the early grades flourished, with close attention to the jointly constructed order of, for example, circle time and kindred organizational curricula. Journals took interest and dissemination venues developed, for example, the series *Working Papers in Sociolinguistics* published by the Southwest Educational Development Laboratory in Austin, Texas, and the journals *Anthropology and Education Quarterly*, the *British Journal for the Sociology of Education*, and *Discourse Processes*, among others.

It is not being suggested that EM solely wrote these developments. It did not. There were kindred developments in sociolinguistics, the ethnography of communication, and child language and developmental studies. But EM was party to a turn in the analytic culture of educational study that now had use for the most ordinary things in the world of classroom life and instruction. It was a major resource for rewriting the kinds of questions we might ask, where we might pursue them, and how we might take interest in what we find there.

Central to EM's contribution was the understanding that the hinges of classroom order, and therefore instruction, were to be found in filaments of meaning, and that meaning owned a local, temporal, and public organization.

The insight was not simply a celebration of context. Context turned out to be more than an oblique invocation of what everyone knows. It was more than a normative repository, or the frame to a picture. The insight was effectively captured in the title of Mishler's benchmark publication, "Meaning in context: Is there any other kind?" (Mishler, 1979), and if the answer is no, then we are condemned both to the study of meaning and its contextually reflexive relations. Context becomes a generative field, prospectively and retrospectively. Each next turn of talk trades upon and (re)writes it. In the bargain, our maps of the life of the room – and of interaction – are rewritten. The study of cases in their constitutive detail displaces the promise of formal, causative analysis.

It is difficult to overstate the implications of the difference, both conceptually, programmatically, and for the very terms of relationship between the professional and research communities. In such fashion, a 90-year program for developing a generalizing theory of learning or instruction or curriculum discovered a pedagogical landscape of diverse locales. The promise of a unified instructional method that would be normative and prescriptive (Bruner, 1966), would then go the way of a unified scientific one. EM was a formative influence in shaping these currents; it wrote an early and articulate critique of a world ordered for the convenience of academic genealogies, and sighted an alternative.

The Local Study of Situated Action

We sometimes cannot tell the history of our ways of speaking, although they surely have one, and seldom only one. The programmatic initiatives collected in this section are those of local study, on the one hand, and EM's interest in the coherence of situated action, on the other.

Local study was of course the mark of the qualitative turn. It was the ethnographic impulse, the turn to cases in context, to Malinowski's "imponderabilia of actual life". In its more familiar expressions, local study yields stories we did not know existed. At its most penetrating, it rewrites a very large conceptual map: The consensual macro-micro divide is inverted, and then dissolved. The local is not derivative of the distal; it is constitutive. Formal structure owes to ordinary, practical enactments and ordinary organizational things (Garfinkel and Sacks, 1970), and in studies of situated action, we can see something of the dissolution, or how reflexive relations do not own firsts. Every meaningful expression already has a world attached, and re-writes it.

Garfinkel often uses the example of freeway traffic, for how the order of the flow is reflexive to the competence of the drivers who produce it. Their competence is visible in the orderly traffic they assemble, and formal accounts of traffic flow are a gloss for the situated analyses of competent drivers engaged in driving. Drivers thus produce the

order that others then search for evidence of autonomous rules and structure. EM sights instead an order of practice.

Questions of order, as in classroom order, quickly yield to questions of meaning. For EM, the organizations of meaning *in situ* – the achievements of common understanding via grammars of natural language – are the central problematic. These are not a grammarian's grammar. They are rather a grammars of social action, played out on fields of relentlessly indexical expressions, expressions whose definite sense and meaning are tied to the occasions of their production and use. As it turns out, there is very little that we say – or can say – whose sense is not occasioned. Every expression is indexical in that way, even $2 + 2$, for whether it is a problem to be solved, a joke, or remonstrance, or a phrase in the argot of design. Classroom instruction is unimaginable absent the exercise of disciplining indefinite expressions, as when a teacher asks, "Do you remember what we were doing last week?" and the students reply in unison, "Yes." Propositionally, of course, the question is simply impossible; too many things were done last week, and none of them definitely so. But the question is understood for what it is doing, namely, introducing what we are about to do now. The production and recognition of the question's sense are contextually embedded products of situated action and analysis. Thus, Garfinkel proposes to analyze the "demonstrably rational properties of indexical expressions," or how indefinite talk yields definite sense and meaning, and thus worlds in common. "The appropriate image of a common understanding is therefore an operation rather than a common intersection of overlapping sets" (Garfinkel, 1967: 30).

If so, meaning has an actionable basis. It is therefore produced and achieved, and action's temporal parameters are included in the production. The meaning of a pause following a teacher's question to a student is a useful example. It begins as an unremarkable duration, a time to formulate a considered reply. But as it continues, it becomes a puzzle of interpretation – did the student hear the question or not? And should it continue further, the duration gives palpable evidence of trouble, or a failure to acquit the task, and thus becomes a remark on the student's capacities, identity, or standing. As a student in my corpus excitedly observed of another's ongoing pause in response to a question about a homework problem, "He can't do it." The conclusion is an analytic one, a remark on the meaningful shape of a developing temporal field.

By these understandings, meaning is neither private nor subjective. It is rather public, methodic, and communitarian. Meaning underwrites the separation of action from behavior, the difference between Gilbert Ryle's winks and blinks. In this light, we can see how a praxeology of social action underwrites every normative glossary. This is not an account of contingency run wild. It is rather an account of the production of normative order as the work of situated action.

Situatedness is now a familiar formulation in educational studies. We find it in the literatures of situated cognition and learning, and activity theory, as arguments on behalf of social action as the foundational field for our inquiries. There is very good sense to them, and taken together they constitute a profound challenge to the psychological metaphors that have anchored and organized educational research. The braiding of these several intellectual histories, however, is less clear.

Among its most familiar discussions in the educational literature is the publication *Situated Cognition* by Brown et al. (1989). It is widely cited and has been enormously influential. Yet, there is some irony in the observation that situatedness was in its earlier formulations, part of a critique of the cognitive metaphor. It was a sociological formulation, and the conceptual arguments found in EM and CA point to public spaces, rather than cognitive ones.

We also tend to find readings of situated action wherein some actions or occasions are situated, and some not. In short order, we can be led to understand that occasions can be *authentically* situated, or not. We have now fashioned a kind of moral metric from a concept that was, at the outset, analytic rather than invidious. A formulation that comes to us as a unifying concept, wherein drill and practice is understood as no less a situated production than whole-language instruction, is rendered a device for parsing and evaluation. This is understandable, given education's professional charge to continually assess, improve, and decide what's best. Every new conceptual resource is examined for what it may give to that effort, and overwhelmingly, interest in situated action has been taken up for how it might leverage change in the efficiencies of the institution and/or the performances of the students it serves. It may be an instructive case in how a conceptual innovation can be hitched to a deeply familiar task, and then, perhaps, lost.

Classroom Studies

EM classroom studies tend to proceed with a different purpose. James Heap, for example, has written a penetrating alternative to the analytic habits of research on reading instruction (Heap, 1982, *passim*). He takes up how classroom reading instruction is organized, *in situ*, and points out the profound consequences of misdescription in classroom research. When we fail to faithfully describe what the parties are doing – the order and grammar of their affairs for them – the consequences for our analyses and understandings of those affairs are devastating. Description is a conceptual matter, and his texts are not simply critiques, but conceptual reformulations of the tasks of reading research, and instruction. Rather than leveraging efficiencies, they attempt to uncover the conceptual knots that yield failed descriptions, and write new ones to replace them. As a way of casting

the difference, EM classroom studies are written for the instruction of the adults in the room.

Similarly, EM studies of science education take interest in the practical classroom enactments that carry it off. Here, especially, we see what may be misplaced about the measure of authenticity. Calls for real science in the classroom date to the nineteenth century. Yet, authentic science is what scientists do (and postdoctoral students struggle to do). The tasks and judgments of science students are not identical with those of a discovering science (although the relationship is more interesting still: there is ample measuring, mixing, watching, and giving of accounts going on in each). To insist on or promise authentic science in the classroom authorizes a daunting gap-closing exercise. It begs the question of what these sciences might be, and elides the work of instruction we actually find there.

Instead, by EM's reading, a curriculum for novices routinely shows itself as a kind of mock-up or version of the affairs it speaks of (as in the difference between driving and driver's education). In a startlingly brief and clarifying discussion, Garfinkel and Sacks (1970) develop the pedagogy of the mock up, and how it makes "deliberately false provision" for certain features of the affairs it represents. But these false provisions are not, then, candidates for correctives. On the contrary, the instructional effectiveness of the mock-up turns on them, as in their example of a plastic engine showing pistons, crankshafts, and connecting rods. It teaches us the internal movements of an engine by not being one. There will be ample time for real engines later. So also, it is for textbook formulations of the causes of the American Civil War, or the well-formed paragraph. Each is a mock-up of topics and competencies that own a lifetime of professional complexity, but must be rendered as a curriculum that novices can do.

Note especially that the difficulty for the authenticity program is not that fifth graders, or undergraduates, are not engaged in inquiry; surely they are, whether in math instruction, language arts, or geography. Instruction is a practical analytic task for students and teachers alike. However, it is precisely these practical forms of inquiry that are missed when we are in hot pursuit of professionally authorized ones. EM classroom studies are bringing these endogenous forms of inquiry into view, and among the forms missing from the literature are the disciplined inquiries that turn our instructions into courses of action we can do. This is the work of following instructions (Amerine and Bilmes, 1988), and when we cannot go on with them, whether in our homework or in assembling a piece of IKEA furniture, instruction founders. Instruction turns on diverse forms of inquiry, if only because the achievements of common understanding that underwrite it are themselves practical tasks of inquiry.

EM has a strong presence in the general mix of classroom discourse analysis through the 30-year corpus of

studies of natural conversation developed by Sacks *et al.* (1974). The linguistic turn announced for philosophy in the 1950s seems to have arrived in education, though surely not the same one. Various iterations of discourse analysis, some affiliated to Gumperz or Hymes, some to Halliday or Sinclair and Coulthard, have taken root in education. But perhaps the most familiar forms derive not from natural language study at all, but from theoretical and continental treatments of Discourse. The most familiar register of classroom discourse studies is found in various expressions of “critical discourse analysis” (and Bakhtin’s theory of dialog). Yet, notwithstanding the attractions of theory-driven analyses, there persists an interest in the sequential analyses of students and teachers engaged in talking their way through their lessons, as we see in a robust literature in foreign and second-language studies. (Firth and Wagner, 1997).

Lee (2006), for example, reexamines the notion of communicative competence. Competence becomes not only the outcome of instruction – the thing we are supposed to learn – but instruction’s very grounds. Students may not know their curriculum, but they must be competent to the interactional co-production of their lessons, and the fixing of misunderstanding, for their learning to proceed. Their competence to those tasks reminds us of an ignored orderliness in instruction’s every enactment, and EM studies of classroom lessons tend to look there, to what is taken for granted about instruction’s work.

The corpus is not limited to primary and secondary classrooms. It includes studies of medical school instruction, law-school pedagogy, and MBA classrooms, alongside studies of undergraduate science and second-language education. In every case, one will find that instruction premises common understanding, and it may be that instruction’s tasks and problematics never stray too far from the achievements of common understanding. In the measure, the order of instruction never strays far from the order of natural language.

Conclusion

This article sketches an account of how a very different vision of disciplined inquiry found its way into a college that has pursued an applied science with great energy for a very long time. EM offers a comprehensive view of the order of everyday worlds. It treats phrases such as situated action or the social construction as real work in the world, and sets out to describe them in the material detail of actual occasions. It has produced a corpus of studies that shows us what such things look like, and has written descriptions that revise many taken-for-granted confidences about things like science, technical reason, and the formal administration of order. These last phrases are among the desiderata that have organized the history

of educational study, and EM writes a very different account of them. At the same time, however, and no matter how keen the critique, the conceptual commitments it criticizes seem no less firmly held.

The attachment to science in matters of education is clearly still with us. Although it has continually escaped our grasp, it apparently feels not quite so distant as the last time we reached for it. Thus, the moral re-authorizations of testing, measurement, and clinical trial models, the dusting off of familiar phrases about what works, and perhaps less noticed – because it is now most pervasive within the qualitative research literature – the return to grand theory, the muse of formal analysis and autonomous structures. In these ways, the strength of our cultural commitments to the technical administration of education seems undiminished.

In sociology, EM has produced departmental communities of scholarship over the last 40 years at University of California, Los Angeles (UCLA), the University of California at Santa Barbara, Boston University and elsewhere in the US, and at Manchester University, Lancaster University, and elsewhere in the UK and Commonwealth countries. Contemporary scholarship has developed in France, Scandinavia, and throughout Europe, and in Japan and Korea, and includes the latest mark of invisible colleges, the LISTSERV. EM/CA is a section of the American Sociological Association, and has a strong presence in science, technology and work place studies. (See, e.g., Button, 1993; Livingston, 2008; and Lynch, 1993.) Yet, EM has never, as a practical matter, challenged the prevailing analytic forms of sociological or educational study.

This is not surprising. EM offers no vision of corporate solutions, no promises of what the research has shown, or complaints about how schools have failed to make good on those promises. Instead, it recommends studies of how students and teachers encounter and acquit their daily tasks of producing the order and instruction of their affairs, and how they show us what is irreducibly social and local about teaching and learning. The question, of course, persists of what on earth would we ever do with such accounts, in the presence of our seemingly perpetual climate of crisis and blame. It should be no surprise that Dewey took interest in the same question. Rather than an administered order grounded in technical expertise, he had in mind a professional cohort of classroom teachers in full possession of “independent judgment” (Dewey, 1965/1904). But if that were so, the familiar pyramid of technical expertise would yield to something quite different; pyramids would become continua. To the researcher, theorist, or consultant, the practical analyst would be no less analytic. There would be a leveling of the epistemic field.

For such a world, EM descriptions and analyses of educational settings and occasions might be a very fine curriculum. For EM, social science is not an arbiter of

things like best practices, only a keenly interested party. Its task is not to trump the professional or moral narratives that are already in place, but to understand them, on the hunch that such understandings might be instructive. But to be clear, the instruction it offers is not for the children in the room. In matters of educational study, it is our instruction that may be the most neglected.

Others hold confidence that this is not so, that there are best practices, and none too many of them. If we could find, reproduce, and install them, we might finally reap the efficiencies promised by the other wing of the progressive movement. This has been the modernist vision – and conundrum – for educational research. Although his topics were different, thematically they were kindred when Wittgenstein summed his discussion of forms of expression and how we use them in the following image:

In the actual use of expressions we make detours, we go by side-roads. We see the straight highway before us, but of course we cannot use it, because it is permanently closed (Wittgenstein, 1958: 426).

EM has been writing this road sign for 50 years, as signage for a world of local meaning, action, order, and competence. EM sights a world we might study to see how it works, and thus find instruction in how indeed it does. Garfinkel speaks of EM studies as tutorials. As with all instruction, however, there are no warranties attached. What we take from them is always a pending discovery. This, of course, is widely known by professional teachers, even university faculty. Yet, it is these very terms of instruction – of every kind and venue – that education's program of formal analysis cannot quite allow. It aims at reproducible certainty, and in these several ways, we are likely to continue to find what is both penetratingly relevant, but improbable, about EM's program for educational study.

See also: Classroom Ethnography; Conversational Analysis.

Bibliography

- Amerine, R. and Bilmes, J. (1988). Following instructions. *Human Studies* 11, 327–339.
- Baker, C. (1997). Ethnomethodological studies of talk in educational settings. In Davies, B. and Corson, D. (eds.) *Encyclopedia of Language and Education*, vol. 3. Dordrecht: Kluwer.
- Brown, J. S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* 18(1), 32–42.
- Bruner, J. (ed.) (1966). Theorems for a theory of instruction. In *Learning about Learning: A Conference Report*. Washington, DC: US Dept. of HEW.
- Button, G. (1993). *Technology in working order: Studies of work, interaction and technology*. New York: Routledge.
- Cicourel, A. and Kituse, J. (1963). *The Education Decision Makers*. Indianapolis, IN: Bobbs-Merrill.
- Cicourel, A. V., Jennings, K. H., Jennings, S. H. M., et al. (1974). *Language Use and School Performance*. New York: Academic Press.

- Cook-Gumperz, J. (1977). Situated instruction: Language socialization of school age children. In Mitchell-Kernan, C. and Ervin-Tripp, S. (eds.) *Child Discourse*, pp 103–121. New York: Academic Press.
- Dewey, J. (1965/1904). The relation of theory to practice in education. In Borrowman, M. (ed.) *Teacher Education in America*, pp 140–171. New York: Teachers College Press.
- Erickson, F. (1992). Post-everything: The world of the moment and how we got here. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Francisco, CA.
- Firth, A. and Wagner, J. (1997). On discourse, communication and [some] fundamental concepts in second language acquisition research. *Modern Language Journal* 81(3), 285–300.
- Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice Hall.
- Garfinkel, H. (1996). Ethnomethodology's program. *Social Psychology Quarterly* 59(1), 5–21.
- Garfinkel, H. and Sacks, H. (1970). On formal structures of practical actions. In McKinney, J. C. and Tiryakian, E. A. (eds.) *Theoretical Sociology*, pp 337–365. New York: Appleton-Crofts.
- Heap, J. (1982). Understanding classroom events: A critique of Durkin, with an alternative. *Journal of Reading Behavior* 14(4), 391–411.
- Heritage, J. (1984). *Garfinkel and Ethnomethodology*. Cambridge, UK: Polity Press.
- Hester, S. and Francis, D. (eds.) (2000). *The Local Education Order*. Amsterdam: John Benjamins.
- Lee, Y. A. (2006). Towards respecification of communicative competence: Condition of L2 instruction or its objective? *Applied Linguistics* 27(3), 349–376.
- Livingston, E. (2008). *Ethnographies of Reason*. Ashgate Publishing. Hampshire, UK.
- Lynch, M. (1993). *Scientific practice and ordinary action*. Cambridge, UK: Cambridge University Press.
- McDermott, R. (1976). *Kids Make Sense*. Stanford University, Palo Alto, CA.
- McDermott, R. and Hood, L. (1982). Institutionalized psychology and the ethnography of schooling. In Gilmore, P. and Glatthorn, A. (eds.) *Children in and out of School*, pp 232–249. Washington, DC: Center for Applied Linguistics.
- McHoul, A. (1978). The organization of turns at formal talk in the classroom. *Language in Society* 7, 183–213.
- Mehan, H. (1979). *Learning Lessons*. Cambridge, MA: Harvard University Press.
- Mishler, E. (1979). Meaning in context: Is there any other kind? *Harvard Educational Review* 49(1), 1–19.
- Moerman, M. and Sacks, H. (1971/1988). On 'understanding' in the analysis of natural conversation. In Moerman, M. (ed.) *Talking Culture*, pp 180–186. Philadelphia, PA: University of Pennsylvania Press.
- Sacks, H. (1992). *Lectures on Conversation*. In Jefferson, G. (ed.) Vol. I and II. Oxford: Blackwell.
- Sacks, H., Schegloff, E., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language* 50, 696–735.
- Watson, R. (1992). Ethnomethodology, conversation analysis and education: An overview. *International Review of Education* 38(3), 257–274.
- Wittgenstein, L. (1958). *Philosophical Investigations*. In Anscombe, G. E. M. (trans.), 3rd edn. New York: Macmillan.

Further Reading

- Atkinson, P. and Delamont, S. (1977). Mock-ups and cock-ups. In Woods, P. and Hammersley, M. (eds.) *School Experience: Explorations in the Sociology of Education*, pp 87–108. London: Croom Helm.
- Burns, S. (1997). Practicing law: A study of pedagogic interchange in a law school classroom. In Travers, M. and Manzo, J. (eds.) *Law in action: Ethnomethodological and conversation analytic approaches to law*, pp 265–287. Dartmouth: Ashgate Press.
- Davies, B. (1983). The role pupils play in the social construction of classroom order. *British Journal of Sociology of Education* 4(1), 55–69.

- Fox, S. (2008). That miracle of familiar organizational things: Social and moral order in the MBA classroom. *Organizational Studies* **29**, 733–761.
- Griffenhagen, C. and Sharrock, W. (2006). Mathematical relativism: Logic, grammar, and arithmetic in cultural comparison. *Journal for the Theory of Social Behavior* **36**, 97–117.
- He, A. (2004). CA for SLA: Arguments from the Chinese language classroom. *Modern Language Journal* **88**(4), 568–582.
- Heap, J. (1992). Ethnomethodology and the possibility of a metaperspective on literacy research. In Beach, R., Green, J., Kamil, M., and Shanahan, T. (eds.) *Multidisciplinary Perspectives on Literacy Research*, pp 35–56. Urbana, IL: NCTE.
- Hestor, S. and Francis, D. (2004). *An Invitation to Ethnomethodology: Language, Society and Social Interaction*. London: Sage.
- Koschmann, T. and LeBaron, C. (2002). Learner articulation as interactional achievement: Studying the conversation of gesture. *Cognition and Instruction* **20**(2), 249–282.
- Lynch, M. and Macbeth, D. (1998). Demonstrating physics lessons. In Greeno, J. (ed.) *Thinking Practices: A Symposium on Mathematics and Science Learning*, pp 269–298. Palo Alto, CA: Erlbaum.
- Macbeth, D. (1994). Classroom encounters with the unspeakable: 'Do you see Danelle?' *Discourse Processes* **17**, 311–335.
- Macbeth, D. (2003). Learning lessons reconsidered: On the differences between the naturalistic and critical analysis of classroom discourse. *American Educational Research Journal* **40**(1), 239–280.
- McHoul, A. and Watson, R. (1984). Two axes for the analysis of 'commonsense' and 'formal' geographical knowledge in classroom talk. *British Journal of Sociology of Education* **5**(3), 281–302.
- Mondada, L. and Pekaret-Doehler, S. (2004). Second language acquisition as situated practice: Task accomplishment in the French second language classroom. *Modern Language Journal* **88**(4), 501–518.
- Payne, G. and Hustler, D. (1980). Teaching the class: The practical management of the cohort. *British Journal of Sociology of Education* **1**(1), 49–66.
- Seedhouse, P. (2004). *The Interactional Architecture of the Language Classroom: A Conversation Analytic Perspective*. Malden, MA: Blackwell.
- Sherman, W. (2004). Science studies, situatedness, and instructional design in science education: A summary and critique of the promise. *Canadian Journal of Science, Mathematics, and Technology Education* **4**(4), 443–465.

Focus Groups

D L Morgan and K Hoffman

© 2010 Elsevier Ltd. All rights reserved.

Researchers use focus groups to generate qualitative data on their topic of interest by creating a conversation about that topic among a set of participants. While the interaction between participants provides focus groups with their unique identity as a method, it is important to recognize that these are not naturally occurring conversations. Instead, the researcher explicitly creates these groups as a means for collecting research data. A related issue is whether there is a difference between group interviews and focus groups. Although there are various examples of group interviewing throughout the history of the social sciences, the most common practice over the past two decades has been to refer to all forms of group interviewing as focus groups.

Although the current popularity of focus groups in the social sciences is a relatively recent phenomenon, they were originally developed in the 1940s by two sociologists, Robert Merton and Paul Lazarsfeld (Merton *et al.*, 1990). After World War II, while focus groups almost disappeared from the social sciences, they also became the most common qualitative method in marketing research. Starting in the 1980s, contacts with marketing researchers led to the reintroduction of focus groups in the social sciences. This renewed interest in focus groups corresponded to the growing importance of qualitative research in education and many other fields. As a result, focus groups are now used so widely that they appear in over 1000 articles per year (according to searches that combine databases such as ERIC, PsychInfo, PubMed, etc.).

One of the main reasons for the increased popularity of focus groups is the variety of not only different forms they can take but also different purposes they can serve (Barbour and Kitzinger, 1999; Krueger and Casey, 2009; Morgan, 1997; Morgan and Krueger, 1998). In a typical focus group, six to eight participants are led by an interviewer – known as the moderator – in a discussion that lasts between 1 and 2 h. The researcher chooses the topics for the conversation, and the participants typically share a similar background with regard to those topics. In addition, focus groups can be conducted in either a relatively unstructured fashion that is oriented toward exploration and discovery, or in a more structured style that is built around the research team's predetermined agenda. This flexibility means that there is no one right way to do focus groups. Instead, there are a number of different factors influencing the design of any research project using focus groups, depending on the goals that the focus groups must meet within that project.

Within qualitative research as a whole, participant observation and individual interviewing are the two most likely alternatives to focus groups as a method for collecting data. For participant observation, the fundamental strength of this method is the ability to provide data about behavior that occurs in a more natural context. Hence, whenever researchers need to know how people actually behave in specific situations, there is an obvious advantage to observing such behavior, rather than listening to discussions about it. In particular, although focus groups rely on the observation of interaction among the participants, it is important to recognize that they fundamentally are interviews. As such, they produce talk data rather than indicators of actual behavior in real contexts.

Shifting to the comparison between individual interviews and focus groups, the first one to be considered is the logistics of bringing together an appropriate group of people. In some situations, it is difficult to locate participants who share an interest in the research topic. In other situations, it may be possible to locate potential participants but difficult to bring them together in the same time at the same place. Whenever those problems are severe, individual interviews have an obvious advantage over focus groups.

At a more substantive level, focus groups and individual interviews differ in the amount of information that they provide about each participant. As focus groups consist of several participants and typically last between 1 and 2 h, a corresponding set of individual interviews would provide a much deeper and more detailed understanding about each participant. There is thus a trade-off between the strengths of focus groups and individual interviews. On the one hand, focus groups offer the opportunity to hear several participants discuss their similarities and differences, but they yield correspondingly less information about each participant. On the other hand, individual interviews offer richer data about the lives of specific participants, but it is up to the researcher – rather than the participants themselves – to make comparisons across participants' responses to the research topic.

Even though focus groups and individual interviews each have their own distinctive set of strengths, it is important not to overstate the differences between the two. In particular, there are many situations where either method could pursue the same goals, so the choice between them often depends on practical considerations. In addition, researchers can often combine focus groups and individual interviews in the same project, instead of facing a forced

choice to use only one method or the other (Morgan, 1997). For example, individual interviews with key informants can provide vital information for designing a focus group research project. Alternatively, focus groups that follow-up on the analysis of individual interviews can serve as a useful form of member checking. This ability to combine focus groups and individual interviews is yet another illustration of the flexible range of uses that have led to the rapid increase in the popularity of focus groups.

Uses for Focus Groups

There are two main reasons why focus groups have found such a wide range of uses. First, focus groups are especially useful for applied research, although these practically oriented uses have been less common in education than in other fields, such as healthcare and social services. Hence, this section includes a number of hypothetical examples of potential uses for applied focus groups within education. The second reason for the popularity of focus groups is their value across a wide range of research goals and research designs. In particular, focus groups work well as either a self-contained method or as part of a project that combines both qualitative and quantitative methods.

In combinations with quantitative methods, focus groups can serve as either inputs or follow-ups to the quantitative portion of the project. In fact, contributing inputs for the content of survey questionnaires was one of the earliest and still best-known uses for focus groups. In this role, focus groups allow the survey researchers to hear the perspective and language of their respondents, prior to developing the actual survey instrument. Focus groups can also serve a similar function as input to experimental program interventions. In that case, the qualitative inputs help the research team learn more about their clients' perspectives prior to designing a program for those clients. In both cases, this input is especially useful for projects working with unfamiliar research populations or new research areas. For example, a preliminary set of focus groups with students could shed light on the relatively new phenomenon of online bullying. Both the issues and the language captured in the focus groups could be especially useful in writing a survey for the larger student body.

Using focus groups as a follow-up to surveys or program interventions typically emphasizes learning more about the results from the quantitative portion of the overall project. When the quantitative results match the researchers' expectations, the follow-up focus groups can put a human face on the data by illustrating the actual experiences associated with the quantitative results. Alternatively, when the goal is to learn more about the quantitative results, exploratory focus groups can provide more depth and detail about the actual processes involved. In both cases, the broad goal of the follow-up focus groups

is to say more about the quantitative results, including how and why those results occur.

For example, consider a school district that wants to pilot test a free breakfast program, based on teachers' reports that students who regularly eat breakfast are more alert and have longer attention spans, leading to better test scores and attendance rates than children who often skip breakfast. One way to enhance the effectiveness of the pilot program would be to conduct focus groups with teachers and parents after the first month, and then use this early feedback to improve portions of project. Later, in the final month of the project, a further set of interviews with both teachers and parents could investigate their responses to the program. In these later groups, the primary topic for both sets of participants would be their feelings about the program. For the teachers, these later groups could also ask about their perceptions of the program's effects on the students' learning and behavior, while additional questions for the parents could ask about any effects on their children's health or behavior. Overall, these two sets of follow-up focus groups could both improve the program's chance of success as well as provide insights into the sources of the program's ultimate outcomes.

As the examples in this section demonstrate, focus groups can be used on their own or in conjunction with other research methods, as well as before or after a larger data-collection effort. Of course, successful focus group projects require careful decisions about research design, regardless of the specific uses for the data from those groups.

Research Design for Focus Groups

One fundamental principle that applies to all aspects of research designs for focus groups is the importance of balancing both the researchers' and the participants' needs. At the most fundamental level, the researchers' needs do come first because they determine the topic for the group discussion. Even so, this starting point is most likely to produce useful data when researchers explicitly consider the needs of the participants. This balancing process occurs in each of the basic aspects of focus group research design:

- selecting the research participants;
- writing the interview questions; and
- moderating the focus groups.

The following sections pay equal attention to each of these three aspects of design. This approach differs from many discussions of focus groups, which devote most of the coverage to moderating as the key to success. Although the role of the moderator is a distinctive aspect of focus groups, this does not mean that skillful moderating is

any more important than selecting appropriate sets of participants and choosing effective questions to guide the discussion. Thus, even the best moderators will struggle if they have to ask poorly written questions to inappropriate participants.

Selecting Research Participants

From a research design point of view, decisions about who to choose as participants is a matter of determining the group composition. The principle of balancing both researchers' and participants' needs is central in making choices about group composition, because of its impact on the interaction that is so essential to focus groups. At a minimum, the participants must feel comfortable discussing the research topic with each other. This is, however, only a minimal requirement; high-quality discussions are most likely to emerge from a group composition that brings together people who are truly interested in the research topic. One common standard for group composition is to find the kind of participants who are just as interested in the topic as the researchers are. Even more effective discussions happen when the participants are not just individually interested in the topic, but also interested in what the others have to say about that topic. Although it is obvious that group composition which meets this standard will increase the quality of the data, another point that is often overlooked is how much easier it is to moderate groups where the participants are highly motivated to share and compare their points of view. Thus, one of the first steps to successful moderating is careful attention to selecting the group composition.

The most common way of facilitating the kind of conversation described above is through recruiting participants to emphasize homogeneity. The value of homogeneity in the group composition comes from the ease of interacting when participants share relatively similar perspectives on the research topic. Thus, the kind of homogeneity emphasized here involves things that the participants have in common with regard to the topic, rather than matching demographic characteristics (such as age or gender). Of course, background characteristics can also be quite relevant to the topic in question. For example, issues related to racial advantage and disadvantage would be a good candidate for using racially homogeneous groups. Focus groups do not, however, have to be limited to homogeneous groups. In particular, groups composed of participants who represent different stances on an issue can provide information about both the sources and implications of those differences – as long as the participants meet the fundamental requirement of feeling comfortable about discussing this topic within the group.

When the design for the research project as a whole is explicitly built around comparing predetermined differences in group composition, this is known as segmentation.

In a project that relies on segmentation, the research team will carefully select the composition of each group in ways that create useful comparisons across the full set of groups. Once again, there is an element of balancing needs by separating the groups in sets that are both homogeneous. From the participants' point of view, the similarity in group composition facilitates their interactions; from the researcher's perspective, the differences between groups create analytic comparisons.

For example, consider a situation where administrators in a school district were interested in implementing Spanish/English dual-language immersion (sometimes known as two-way immersion) in kindergarten classrooms. In particular, if another district in the state had recently proposed this model but met with opposition from stakeholders such as school principals, teachers, and parents, then the administrators in the current district could run into opposition. First, they would want to better understand the dynamics of the previous attempt in the district where the dual-immersion proposition had failed. A method of generating this information would be to do a pair of focus groups with key informants who were either proponents or opponents of that earlier attempt. Within the proponents group, questions could focus on: Why participants felt inclined to support dual-immersion? Where and how they encountered opposition? Also, what they might have done differently to gain the acceptance of stakeholders? Within the opponents group, questions would focus on: Why they did not support the proposal? What they objected to? Also, were there different approaches that might have increased their buy-in? The researchers could use this information to create a revised plan for introducing this language-immersion program, and then conduct a set of focus groups in their own district to hear whether their own stakeholder groups would have a more favorable response.

As a final topic on selecting participants for focus groups, no presentation on this issue would be complete without considering recruitment. Although recruiting may seem like a relatively routine task, it is far and away the most common cause of problems in focus group research. These problems arise when recruitment is treated as merely clerical work, in comparison to the supposedly more important work of designing the project and moderating the groups. Yet, even the best research design and most accomplished moderators are of no use if not enough participants show up or if they arrive with misguided expectations. Fortunately, recognizing the potential magnitude of this problem is the most important step in solving it, and there are several useful descriptions of how to do effective recruitment (e.g., Krueger and Casey, 2009; Morgan, 1998).

Writing Interview Questions

The full set of questions for a focus group is referred to as the interview guide. These guides typically vary along

one basic dimension: from less structured to more structured. Less-structured interview guides are especially well suited to projects where the goals are broad and exploratory, because the participants are encouraged to discuss a relatively small number of questions from their own point of view. In contrast, more-structured guides are most appropriate when the goal is to generate depth and detail on a predetermined agenda, because the discussion consists of a larger number of well-defined questions that match the researcher's own agenda. Focus groups work equally well with either less-structured, exploratory questions or more-structured, agenda-driven questions; the decision between these two approaches is a function of the goals for the research project, rather than the strengths and weaknesses of focus groups as a method.

Many research projects call for both more exploratory and more agenda-driven questions, and the most common approach for this combination is a funnel format. The funnel metaphor represents the idea that the interview begins with a set of broad, exploratory questions before shifting to a more tightly focused set of questions. The less-structured questions at the top of the funnel are thus devoted to hearing the participants' perspectives in their own terms, while the more-structured questions in the later section are devoted to learning about the researcher's pre-selected topics. One of the principal advantages of the funnel format for interview guides is that the early section shows the extent to which the participants are interested in the same issues as the researchers, while the later sections ensure that the research team hears about the topics that are most central to their own interests.

Regardless of the degree of structure in the guide, it is once again important to consider both the participants' and the researcher's needs. The core issue is how appropriate the questions are for these participants as a basis for discussing this topic. Thus, the kind of conversation that will be most useful for the researcher's interests also has to consider what will interest the participants. Finding an approach that will make it easier for the participants to talk about the research topic is especially important for the first question, that is, the one that starts the actual group discussion.

As an example of questioning strategies, consider a case where the administrators at a state college are interested in female students' low levels of enrolment and high rates of attrition in science and math and engineering (SME) classes. The goal for these focus groups would be to learn more about the barriers and strategies for engaging and retaining women in SME programs. A research design for this project could compare newly admitted female students with equivalent Scholastic Aptitude Test (SAT) scores and high school grade point average (GPA), where one segment had enrolled in SME classes and the other had

not. In this case, it would be appropriate to use a funnel approach, beginning with general curriculum interest questions, such as "What were some things you considered when you chose which courses to take?" From the participants' point of view, this would be a relatively easy and interesting question to talk about, thus starting a free-flowing discussion. For the researchers, this would allow them to hear what students had to say about SME courses, before any of the later questions in the funnel explicitly introduced this topic. Finally, narrower, follow-up questions could ask about more specific issues related to both SME classes and the appeal of an SME-career-associated lifestyle.

Moderating Focus Groups

Like interview guides, approaches to moderating typically vary from less structured to more structured. With a less-structured moderating style, the moderator provides rather little active direction during the discussion, to create a conversation that concentrates on the participants' thoughts and feelings about the topic. To accomplish this less-structured approach, moderators will typically use the introduction at the start of the session to explain that their primary role will be to listen to and learn from the participants, so the participants themselves will be generating and carrying on the conversation. In contrast, a more-structured moderating style requires an active role in directing the discussion, so that participants concentrate on topics that match predetermined research goals. For this approach, moderators would typically introduce their role as ensuring that each question gets covered as thoroughly as possible, even if that means the moderator must occasionally either remind the participants to stay on topic or move them on to the next question. Once again, focus groups work equally well with both less-structured and more-structured approaches to moderating, so the choice depends on the purposes for the research. For example, a less-structured moderating style would be more appropriate for an exploratory study aimed at hearing how parents of special needs children want classrooms to be improved to meet their children's challenges, while a more-structured style would be more appropriate for investigating how those same parents evaluated a program that had already redesigned classroom participation for their special needs children.

As a final comment on moderating, it is useful to return to the point that was made at the beginning of this section on research design: moderating is only one component of successful focus groups; the selection of participants and the writing of interview questions are just as important. In particular, as any experienced moderator will tell you, it is much easier to conduct a focus group that consists of participants who are genuinely interested in the questions being asked.

Analysis

In addition to data collection, another important issue is how the data will be captured during the sessions. Audio recording is by far the most common technique for generating the actual data from focus groups. In contrast, video recording is relatively rare due to the difficulty of creating camera setups that provide sufficient detail on all the participants. Relying strictly on written notes is another option, especially when ethical issues make audio recording inappropriate. Written notes are also the preferred method for relatively small preliminary projects, such as using focus groups to help design specific questions for a survey questionnaire. In that case, the note taker is usually an assistant who sits in and then debriefs with the moderator after the group. This debriefing process usually consists of writing up a summary of the discussion that can be distributed to the rest of the team the next day. In addition, another valuable role for note-taking assistants can be to keep a log of who said what during the focus group. This kind of log is crucial for analyses that examine the contributions from individual participants, because that information requires a knowledge of who said what.

The actual methods for analyzing focus group data require relatively little discussion, because they are quite similar to individual qualitative interviews in this regard. Specifically, the most common approach is to transcribe the audio recordings and then apply coding and other interpretative processes to the text data. For example, one common approach is to use content analysis as a basis for attaching codes to sections of the conversation that meet prespecified criteria. Alternatively, there are more inductive approaches, such as grounded theory, where the analyst begins with open coding to create initial codes which are then grouped into interpretative categories, and so on.

The main element that requires unique attention in analyzing focus groups is the obvious fact that the data do come from groups of individuals. In particular, the data from the various participants are not independent due to the influence that participants have on each other's contributions to the discussion. The key issue in considering data from a set of focus groups is determining the topics that mattered most to the full set of participants across all the groups. The first tactic for addressing this situation is what is known as group-to-group consistency, where the topics that are important in one group are also important in other groups. Beyond that, the importance of a topic is reinforced by the number of participants who join in the relevant discussion, where widespread interest clearly counts for more. The third criterion is the level of energy that participants devote to the discussion of that topic,

because it is entirely possible for most of the participants in most of the groups to acknowledge a topic without becoming engaged in the discussion of that topic. Thus, the easiest way to make the case that a particular topic was important to the participants consists of three elements: the topic should be discussed with enthusiasm by nearly all the participants in nearly all of the groups. Note, however, that this process is designed to locate topics that were important to the participants, which is not the same as saying that they will be central to the researcher's analysis. In particular, if you are conducting an inductive analysis that is oriented toward locating themes and building theory, then information about what the participants considered important may be only one aspect of your broader interpretive goals.

Conclusions

The final point that we would like to emphasize about focus groups is the widespread acceptance they have received as a general-purpose method throughout the social sciences. They are thus useful for both classic qualitative goals such as generating theory and for more applied purposes such as program design and evaluation. In addition, they are equally appropriate for less-structured, exploratory research and more-structured, in-depth examinations of researcher-determined topics. Finally, they work equally well as a stand-alone method for generating purely qualitative data and as a contribution to projects that combine qualitative and quantitative methods.

One of our goals throughout this article has been to encourage the increased use of focus groups within the field of education. In particular, if we have successfully demonstrated the value of focus groups by combining a design-oriented presentation with a series of practical examples, then that will help accomplish this goal.

Bibliography

- Barbour, R. and Kitzinger, J. (1999). *Developing Focus Group Research*. Thousand Oaks, CA: Sage.
- Krueger, R. A. and Casey, M. A. (2009). *Focus Groups: A Practical Guide for Applied Research*, 4th edn. Thousand Oaks, CA: Sage.
- Merton, R. K., Fiske, M., and Kendall, P. L. (1990). *The Focused Interview*, 2nd edn. New York: Free Press.
- Morgan, D. L. (1997). *Focus Groups as Qualitative Research*, Sage Qualitative Research Methods Series, 2nd edn., vol. 16. Thousand Oaks, CA: Sage.
- Morgan, D. L. (1998). Planning focus groups. In Morgan, D. L. and Krueger, R. A. (eds.) *The Focus Group Kit*, vol. 2. Thousand Oaks, CA: Sage.
- Morgan, D. L. and Krueger, R. A. (1998). *The Focus Group Kit*. Thousand Oaks, CA: Sage.

Grounded Theory

K Charmaz, Sonoma State University, Rohnert Park, CA, USA
A Bryant, Leeds Metropolitan University, Leeds, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Constructivist grounded theory – A twenty-first-century revision of Barney G. Glaser and Anselm L. Strauss's classic statement of grounded theory. This theory adopts grounded theory strategies without the positivist epistemological underpinnings of the original statement. Instead, constructivist grounded theory (1) is founded on a relativist epistemology, (2) views the method as interactive as well as comparative and inductive, (3) takes into account the multiple and shifting standpoints of the researcher and the researched, (4) builds on the method's pragmatist heritage, and (5) emphasizes reflexivity.

Theoretical sampling – A grounded theory strategy that involves sampling for the development of the researcher's emergent theory. Grounded theorists engage in theoretical sampling to fill out the properties of their theoretical categories, to refine relationships between categories, and to identify variation and difference in their emerging theories. Theoretical sampling is only used to develop or test categories late in the analytic process; it is not an initial sampling procedure for beginning data collection.

Grounded theory has two related meanings: (1) a set of systematic methodological strategies that constitute a distinct method for conducting research and analyzing inductive data and (2) the product of this process, the completed theoretical analysis of these data. Here, we focus on the method and the construction of the product, typically a grounded theory that illuminates a substantive area of research interest. Grounded theory strategies consist of systematic, but flexible, guidelines for data gathering, coding, synthesizing, categorizing, and integrating concepts for the explicit purpose of generating middle-range theory. Most fundamentally, grounded theory strategies invoke comparative methods for analyzing data and entail an iterative process of simultaneous data collection and analysis. Grounded theorists engage in emergent analysis to direct their subsequent data collection, which, in turn, they use to develop and check their emerging ideas.

This paper discusses the history and development of grounded theory; compares the contemporary version of

the method, constructivist grounded theory, with positivist versions; outlines its main methodological strategies; and suggests its future directions. The originators of grounded theory, sociologists Glaser and Strauss (1967), explicitly developed analytic strategies for constructing theory from qualitative data; hence, they proposed grounding emergent theories in the data. Thus, the resulting grounded theories explain the data that give rise to them, rather than being separate from the collected data or imposed on them.

History and Development of Grounded Theory

The development of grounded theory reflects its historical location in the 1960s and the intellectual standpoints of its founders. Glaser and Strauss (1967) defended qualitative research at a time when quantitative methods had gained dominance in sociology and throughout most of the academy. Many of Glaser and Strauss's quantitative contemporaries saw qualitative research as useful only for exploratory research or for constructing more precise quantitative instruments, if they saw any value in it at all. Vibrant University of Chicago qualitative traditions of life-history research and field studies had significantly informed the discipline throughout earlier decades in the twentieth century but had become imperiled by the 1960s.

At that time, qualitative researchers attempted to defend their practice through the framework of quantitative inquiry with its emphases on reliability and validity. Qualitative sociologists focused on making their studies objective through the accuracy and thoroughness of their data collection. Novices had few methodological texts available to guide their work. Instead, they learned about conducting qualitative inquiry primarily through mentoring and immersion in fieldwork.

Glaser and Strauss (1967) took up the methodological debates of their day and answered criticisms of qualitative inquiry in their initial statement of grounded theory, *The Discovery of Grounded Theory: Strategies for Qualitative Research*. They realized that other researchers could use the methods they had developed in their studies of the social organization of dying. In this book, Glaser and Strauss explicitly confronted their discipline and also implicitly revealed the traditions that each brought to the research process. Strauss

brought his University of Chicago background to grounded theory with its Chicago School traditions of naturalistic inquiry, pragmatist assumptions, and symbolic interactionist perspective. Thus, Strauss emphasized field research, viewed individuals as active agents, saw interaction as open ended, took into account language and meaning, and focused on action, all of which informed grounded theory.

Simultaneously, Glaser brought systematic methods, an explicit logic, and a concept-indicator model to grounded theory. Glaser's training at Columbia University in rigorous quantitative methods influenced his goal to create an analogous codification of qualitative methods like his mentor, Paul Lazarsfeld, had established in quantitative inquiry. Glaser supplied much of the logic and language of grounded theory and built the method on assumptions of externality, neutrality, and parsimony.

The two originators converged on five main points. Both of them:

1. emphasized constructing emergent theories with new ideas;
2. contended that qualitative research could generate theory;
3. viewed grounded theory as a method for conducting rigorous, processual analyses;
4. advocated using comparative methods throughout the analytic process; and
5. intended to provide specific tools for theory construction.

Glaser and Strauss (1967) challenged the nature and purpose of sociological theory and attempted to demystify what theorizing entailed. By the 1960s, theorizing in sociology had become the purview of a few elite scholars who seldom based their theoretical notions on empirical research. The division of labor separating empiricists and theorists had widened into a chasm. Most theorists aimed to create grand theories that would explain social structure and change at societal levels. Those quantitative researchers who attended to theory invoked a logico-deductive model to derive hypotheses from the narrow range of concepts that these grand macro-sociological theories offered. Many empirical researchers, whether quantitative or qualitative, worked on concrete research problems such as consumer surveys or ethnographic description. Glaser and Strauss called for qualitative researchers to move their work from description to theoretical analyses, and in so doing they sought to extend the reach and significance of qualitative inquiry.

Grounded theory played a significant role in creating and advancing the qualitative revolution that began in the late 1960s, grew in the decades thereafter, and broadened research in education. The arguments in the *Discovery* book repositioned and revitalized qualitative research. Glaser and Strauss attacked the dominant methodological views that relegated qualitative research to a marginalized position in sociology. They contended that

1. qualitative analysis could be rigorous and systematic, rather than impressionistic and idiosyncratic;
2. theory and method could be joined;
3. the sharp division of labor between theorists and methodologists was arbitrary;
4. systematic qualitative analysis of data could produce fresh, useful theories;
5. sociologists should aim for empirically grounded substantive and formal theories of the middle range rather than airy macro-level theories unrelated to social life; and
6. theorizing would be democratized because grounded theory methods gave ordinary researchers tools of theory construction.

From the beginning, Glaser and Strauss intended to create systematic theories that fit the studied realities were modifiable as further data became manifest and were useful to policymakers, practitioners, and research participants as well as to disciplinary colleagues. They argued for constructing analyses that would give research participants a theoretical explication of a fundamental social process that affected them. Glaser and Strauss were convinced that researchers could make their most significant contributions by offering abstract ideas about data that were thoroughly grounded in these data.

Glaser and Strauss made a powerful case for legitimizing qualitative research in sociology that appealed to researchers in sociology and other disciplines alike. Glaser and Strauss's work inspired generations of graduate students in the social sciences and professions to pursue qualitative research and to claim grounded theory as their method of choice. Most of these students treated grounded theory as a justification for conducting inductive qualitative research. Even if they adopted several grounded theory strategies, they seldom used them as Strauss and, particularly, Glaser had envisioned. Glaser and Strauss's (1967) book made rallying claims and proposed explicit methodological strategies, although the latter remained opaque to most readers. Glaser's (1978) book, *Theoretical Sensitivity*, made the methodological strategies more explicit but his dense writing, dearth of clear definitions, and lack of an index reduced the book's effectiveness as a research resource for most readers who had not studied with him.

Strauss's (1987) publication of *Qualitative Analysis for Social Scientists* and the 1990 book, *Basics of Qualitative Research*, co-authored with Juliet Corbin, provided grounded theory instructions for thousands of teachers and graduate students. Strauss saw grounded theory as a method of verification, and he and Corbin introduced several techniques to apply to the data analysis. Glaser's (1992) acrimonious rebuttal, *Basics of Grounded Theory Analysis: Emergence vs. Forcing*, countered Strauss and Corbin's claims and objected to their new techniques. Glaser argued that Strauss and Corbin had abandoned the emergent analytic

process of theory construction inherent in grounded theory in favor of conceptual description and preconceived techniques. When stripped of their acrimony, several of Glaser's main charges were correct. The first two editions of *Basics of Qualitative Research* seemed procedural and prescriptive and lacked the emphases on flexible guidelines and emergent inquiry in the earlier works. Most researchers, however, remained untroubled by the apparent discontinuities between the *Discovery* book and Strauss and Corbin's texts (Bryant and Charmaz, 2007).

Charmaz's (2000) and Bryant's (2002) critiques took grounded theory in new directions. They each viewed both Glaser's (1978, 1992) and Strauss and Corbin's (1990, 1998) versions of grounded theory as tied to positivist epistemologies. Unlike earlier critics of grounded theory, Charmaz and Bryant each argued that researchers could adopt the methodological guidelines of grounded theory without importing positivist assumptions with these guidelines. The guidelines could be used from a variety of epistemological starting points and initial theoretical standpoints, as is evident in educational studies. Thus, Charmaz and Bryant aimed to retain key grounded theory strategies but to reposition them in twenty-first-century epistemologies and ontologies. Publication of Charmaz's (2006) book, *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*, offered an accessible explication of how to conduct a grounded theory study founded on a different set of epistemological principles. The subsequent *Handbook of Grounded Theory* (Bryant and Charmaz, 2007) covered the range of versions of grounded theory from positivist to constructivist and showed the varied ways in which grounded theory informs the contributors' thinking and research practice.

Comparing Constructivist and Positivist Grounded Theory

Constructivist grounded theory builds on the original statements and strategies that Glaser and Strauss (1967) first articulated. The contrasts in epistemology between constructivist and positivist versions of grounded theory, which we outline here, are greater than those between constructivist grounded theory and Strauss and Corbin's (1990, 1998) postpositivist versions (Corbin and Strauss, 2007). Grounded theory is an evolving method, not simply because of different methodologists' visions but because Glaser (e.g., 1998, 2003) and Corbin (Corbin and Strauss, 2007) have revised their earlier versions of the method. Nonetheless, Glaser has maintained a highly consistent logic and epistemology over four decades.

In brief, constructivist grounded theory arises from a relativist epistemology, challenges positivist assumptions in earlier versions of grounded theory, and aligns the method with interpretive inquiry. It treats grounded theory

strategies as flexible guidelines that serve as heuristic devices. Constructivist grounded theory has roots in social constructionism, takes action as a central concern, builds on the pragmatist legacy of Anselm Strauss, and assumes that the researcher is part of the research process. Subsequently, constructivist grounded theorists engage in reflexivity throughout the research process, which fosters attending to the varied standpoints of both the researcher and the researched. Constructivists argue that all knowledge flows from standpoints, both those preceding inquiry and occurring within it.

The relativism inherent in constructivist grounded theory assumes multiple, layered realities that shift and change under different conditions. Realities may then be elusive and the researcher's actions become part of them. Fact and value merge. Constructivist grounded theorists define facts through the values that allow them to see. For example, a nuanced understanding of poverty may alert an educational researcher as to how its invisible consequences affect the relative effectiveness of program initiatives in the schools. A middle-class researcher without this sensitivity may not see the same facts as the first researcher.

The relativity of constructivism contrasts with the objectivism inherent in positivism. Grounded theorists, who adhere to positivism, treat reality as unitary and, for the most part, self-evident. Fact and value are separate in this view. Positivist grounded theorists attempt to find objective facts in an external reality and question the subjectivism that may enter constructivist grounded theory through the researcher's choice of topics and methods, such as intensive interviews. Constructivists attempt to recognize the subjective origins of method – and knowledge – rather than to erase their existence. Thus, constructivists try to place subjectivity in its social locations and examine it reflexively.

Constructivist and positivist grounded theorists take divergent positions on data collection. Constructivist grounded theorists advocate gaining an insider's view of the research problem, setting, and participants, which means gathering extensive rich data about research participants' lives and worlds through sustained interaction rather than limited interviews or isolated visits. The depth and extent of data facilitate the researcher's efforts to go beneath the surface and enter the liminal world of research participants' implicit actions and meanings. Constructivists follow the pragmatist prescription of paying attention to language as a way to learn participants' implicit meanings and to understand their actions. For constructivists, the detailed work involved in gaining an insider's view allows discovering variation in the studied process and identifying difference in the analyses. Gathering extensive rich data is also the first step toward making the subsequent research product credible.

In contrast with constructivists, positivist grounded theorists do not attend closely to data collection. They

believe that detailed data distract researchers from discovering and pursuing the analytic focus. Hence, several proponents argue against transcribing interviews, a step most qualitative researchers see as a given. Positivist grounded theorists emphasize gathering only enough data to explicate their emergent category or categories. These researchers tend to take language and meanings for granted without examining their implicit meanings and relationship to action.

The objectives of constructivist and positivist grounded theory differ, particularly those in Glaser's (1978, 1998; Glaser and Strauss, 1967) version of grounded theory. He aims to identify variables, make parsimonious generalizations, and offer useful explanations, all of which transcend the particularities of historical and social locations. Constructivist grounded theory, in contrast, aims for interpretive understanding and situated knowledge. This approach locates the research process and product in its historical, social, and situational contexts. Hence, situated knowledge takes into account positionality and particularities (Clarke, 2005) and informs the researcher's interpretations of the data.

Given the relativity and reflexivity that constitute constructivist grounded theory, it follows that constructivists view data as co-constructed with research participants, whether these data consist of interviews or documents or anything else. The co-construction of an interview may be readily discernible as interviewer–respondent interactions proceed. Nevertheless, co-construction occurs with other forms of data as well. We may not construct documents, but how we read them depends on our view of their purposes and, moreover, the questions and frames we bring to them.

As Bryant and Charmaz (2007) point out, grounded theory has become a general method albeit with several versions. It reaches multiple disciplines and fields and stretches across the globe. Several grounded theory strategies, particularly qualitative coding and simultaneous involvement in data collection and analysis, have become generalized as routine practices in qualitative inquiry. Other qualitative researchers may not use specific grounded theory methodological strategies as grounded theorists do, but nonetheless adopt aspects of them. Many educational researchers adopt grounded theory coding strategies for synthesizing data and defining themes but not for theoretical analyses of processes.

Despite its differences from other versions of grounded theory, constructivist grounded theory practice shares much in common with them. All versions of grounded theory:

1. begin with an inductive logic;
2. emphasize the analytic process;
3. endorse explicit analytic guidelines, although we differ on which guidelines we adopt;
4. aim for abstract conceptualization to advance theory construction;

5. engage in an iterative process to advance the analysis; and
6. intend to encourage innovation.

Perhaps ironically, the recent explication of constructivist grounded theory has advanced all variants of the method because it concretely showed how to employ grounded theory logic and strategies.

Grounded Theory Strategies and Educational Research

The grounded theory method consists of flexible, yet distinct, strategies that distinguish it from other qualitative approaches. Educational researchers have adopted and adapted certain strategies, particularly grounded theory coding, to fit their research problems and practice. A major contribution of educational researchers is to integrate grounded theory strategies with ethnography, case studies, and often multisite case studies that use several types of data collection.

The relationship between extant theory and grounded theory construction and the place of the literature review remain contested by both grounded theory critics and its proponents. Classic grounded theorists eschew relying on extant theory and enjoin researchers to delay the literature review until they develop their own analyses. Both strategies are intended to help researchers avoid preconceiving their ideas by forcing data into extant concepts.

The constructivist response to these strategies is consistent with grounded theory practices in educational research that draw upon earlier ideas but subject them to empirical inquiry, such as Coburn and Talbert's (2006) use of sense-making and institutional theory while starting with descriptive codes with little interpretation. Constructivist grounded theorists argue that researchers:

1. already possess a fund of knowledge and experience before they begin;
2. may draw on broad ideas from their experience or discipline as starting points for data collection but not as ending points analysis;
3. should remain open to the empirical world; and
4. must subject all ideas about it to rigorous scrutiny, including their own emerging theoretical notions.

Constructivists position their research in relevant literatures and explain how it advances knowledge. This approach fits theoretically driven studies in education, which begin inquiry from theoretical positions, such as Thornberg's (2008) inclusion of domain theory in moral education, Qin and Lykes (2006) starting with feminist theory, and Valadez's (2008) use of structuration.

The specific methodological strategies of grounded theory support analyzing processes. Three main strategies

comprise the grounded theory method: coding, memo-writing, and theoretical sampling. Coding means applying a short-hand label to sort, synthesize, and conceptualize data. It involves two stages: initial coding, which emphasizes gaining an analytic handle on the data by defining them, and focused or selective coding, which entails using the most frequent and/or significant initial codes to sort and synthesize data. Memo-writing consists of writing about tentative ideas and emergent categories and includes the crucial intermediate stage of writing between coding data and writing the first draft of a paper. Theoretical sampling is the strategy that grounded theorists use after developing some tentative conceptual categories to fill out these categories, although most do not take their studies into explicit theory construction.

During initial coding, grounded theorists look for processes in the data and attempt to show processes through choosing words as codes that describe what people are doing and what is happening in the settings. Educational researchers often pursue practical problems, which their codes reflect. Eich (2008) aimed to discover the elements of effective student leadership development programs and began by coding for program attributes, actions, and outcomes.

Glaser (1978) described grounded theory as sociology of gerunds because these words depict actions. Using gerunds helps researchers to see what people are doing and what is occurring. Furthermore, coding with gerunds helps researchers to identify a significant process that conceptualizes and integrates the studied phenomena.

Qin and Lykes (2006) not only capture the grounded theory emphasis on a significant process in their study, 'Reweaving a fragmented self', but they also show how specific subprocesses, such as "integrating traditional values" (p. 185) and "breaking the web" (p. 187), and "rethinking 'the homeland'" (p. 191) become analytic categories that contribute to this overall process. Qin and Lykes delineate the conditions under which Chinese women students' understanding of themselves (1) had become fragmented before leaving China, (2) became fragmented after coming to the United States, and (3) became woven into a transformed self. Qin and Lykes's use of grounded theory reflects constructivist principles of using a theoretical perspective as an initial lens to look at data, acknowledging standpoints and starting points, maintaining reflexivity throughout the research process, and subjecting their theoretical notions to rigorous scrutiny. They end their analysis by challenging grand meta-narratives of feminist theory.

Initial line-by-line coding is a heuristic device that leads the researcher to study each line of data to discern the action it indicates. When researchers code in gerunds and keep their codes short, precise, and analytic, as Qin and Lykes's categories indicate, they can discern relationships between codes and begin to see larger processes unfold.

This type of coding differs from most qualitative coding because the latter typically defines general topics and themes, rather than actions and processes. Furthermore, grounded theorists compare data with data, as Eich compared the attributes and actions of programs with each other. The next stage of coding, focused coding, takes the comparative process a step further. Here, researchers use the most frequent and/or significant codes to sort and synthesize data. Hence, focused coding prompts researchers to scrutinize data through the lens of selected initial codes and thus puts these codes to test.

Memo-writing proceeds from the start and becomes more analytic and directed as inquiry ensues. Initially, a researcher's memos might consist of hunches, questions, and areas to explore during subsequent data collection. Gradually, researchers write more analytic, abstract memos on their emerging theoretical categories. A major purpose of memo-writing is to explore the potential of selected focused codes as tentative categories. Subsequently, researchers define these codes by the properties they see in the data. These definitional aspects are crucial. The defined properties do not inhere in the data. Rather, researchers define them through interacting with the data during the analytic process as well as during actual data collection. Through memo-writing, a researcher can break a tentative category open as well as articulate the conditions under which it develops or changes, its relationship to other categories, and its consequences. Memo-writing allows constructivist grounded theorists to explore research participants' tacit as well as overt meanings and actions.

Writing memos keeps the researcher active in the analytic process as well as alert to unanswered questions about the data. These memos demonstrate the links between the researcher's emerging theoretical analysis and the data. Not only does a memo capture the researcher's ideas about a tentative category, but it also contains and weighs evidence for this category. Memo-writing encourages the researcher to make systematic comparisons between data as Thornberg (2008) compared children's responses to moral questions for their similarities and differences.

As an intermediate but imminently correctable stage of writing, memo-writing shortens the distance between coding and writing first drafts. Advanced memos provide the substance for framing papers and chapters and, when taken collectively, may indicate a logical ordering. If grounded theorists define a major process in the data, their explication of the process can order the memos.

By engaging in memo-writing, researchers often find that they need more data to illuminate their categories or to find variation in the studied process. Hence, they return to their research participants and gather more data or build further questions into subsequent data collection. In short, they engage in theoretical sampling to fill out the properties of their categories. Perhaps because the term

theoretical sampling borrows language from quantitative research, it is the most misunderstood of the grounded theory strategies. Theoretical sampling is a sampling that develops the researcher's emergent theoretical analysis; it is not representational or initial sampling. Having developed a set of categories is prerequisite for conducting theoretical sampling. In short, theoretical sampling provides a means for researchers to check, elaborate, and assess their emerging categories and to obtain the data to help them demonstrate how their analytic categories fit together. Thus, conducting theoretical sampling buttresses researchers' claims to making their studies grounded.

Future Directions

Grounded theory offers educational researchers a method that complements varied forms of qualitative data collection and that will expedite their work. Adopting more grounded theory strategies will enable educational researchers to further the theoretical reach of their studies and to make tacit meanings and processes explicit. Constructivists have not only re-envisioned grounded theory, but also revised it in ways that make the method more flexible and widely adoptable than its earlier versions.

In the past, grounded theory has often been viewed as separate from other methods. Now, the constructivist version makes the usefulness of combining grounded theory with other approaches more apparent, as is evident in grounded theory studies in education. Grounded theory can make ethnography more analytic, interview research more in-depth, and content analysis more focused. Several computer-assisted qualitative data analysis programs are built on grounded theory, and this method can add innovation to mixed methods research. Grounded theory emphasizes focusing data collection and checking and developing analytic ideas. Hence, grounded theory offers the tools for building strong evidence within the analysis and for explicating processes. Consequently, grounded theorists in education have a bright future for making powerful arguments in areas such as curricular studies, educational leadership, and educational policy.

See also: Computer Assisted Qualitative Data Analysis; Ethnography; Interpretive Research; Interviews and Interviewing; Life History; Mixed Methods; Participant Observation; Pragmatism; Qualitative Data Management; Validity; Mapping Diverse Perspectives.

Bibliography

- Bryant, A. (2002). Re-grounding grounded theory. *Journal of Information Technology Theory and Application* 4, 25–42.
- Bryant, A. and Charmaz, K. (2007). Grounded theory in historical perspective: An epistemological account. In Bryant, A. and Charmaz, K. (eds.) *Handbook of Grounded Theory*, pp 31–57. London: Sage.
- Charmaz, K. (2000). Constructivist and objectivist grounded theory. In Denzin, N. K and Lincoln, Y. (eds.) *Handbook of Qualitative Research*, 2nd edn, pp 509–535. Thousand Oaks, CA: Sage.
- Charmaz, K. (2006). *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. London: Sage.
- Charmaz, K. (2007). Constructionism and grounded theory. In Holstein, J. A and Gubrium, J. F. (eds.) *Handbook of Constructionist Research*, pp 319–412. New York: Guilford.
- Clarke, A. E (2005). *Situational Analysis: Grounded Theory after the Postmodern Turn*. Thousand Oaks, CA: Sage.
- Coburn, C. E and Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education* 112, 469–495.
- Corbin, J. and Strauss, A. (2007). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, 3rd edn. Los Angeles: CA: Sage.
- Eich, D. (2008). A grounded theory of high-quality leadership programs: Perspectives from student leadership development programs in higher education. *Journal of Leadership and Organizational Studies* 15, 176–187.
- Glaser, B. G. (1978). *Theoretical Sensitivity*. Mill Valley, CA: The Sociology Press.
- Glaser, B. G. (1992). *Basics of Grounded Theory Analysis*. Mill Valley, CA: The Sociology Press.
- Glaser, B. G. (1998). *Doing Grounded Theory: Issues and Discussions*. Mill Valley, CA: Sociology Press.
- Glaser, B. G. (2003). *The Grounded Theory Perspective II: Description's Remodeling of Grounded Theory*. Mill Valley, CA: The Sociology Press.
- Glaser, B. G. and Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL: Aldine.
- Qin, D. and Lykes, M. B. (2006). Reweaving a fragmented self: A grounded theory of self-understanding among Chinese women students in the United States of America. *International Journal of Qualitative Studies in Education* 19, 177–200.
- Strauss, A. L. (1987). *Qualitative Analysis for Social Scientists*. Cambridge: Cambridge University Press.
- Strauss, A. and Corbin, J. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, CA: Sage.
- Strauss, A. and Corbin, J. (1998). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*, 2nd edn. Thousand Oaks, CA: Sage.
- Thornberg, R. (2008). School children's reasoning about school rules. *Research Papers in Education* 23, 37–52.
- Valadez, J. R. (2008). Shaping the educational decisions of Mexican immigrant high school students. *American Educational Research Journal* 45, 834–860.

Further Reading

- Carreón, G. P., Drake, C., and Barton, A. B. (2005). The importance of presence: Immigrant parents' school engagement experiences. *American Education Research Journal* 42, 465–498.
- Charmaz, K. (2008a). The legacy of Anselm Strauss for constructivist grounded theory. In Denzin, N. K. (ed.) *Studies in Symbolic Interaction* 32, pp 127–141. Bingley, UK: Emerald.
- Charmaz, K. (2008b). Reconstructing grounded theory. In Bickman, L., Alasuutari, P., and Brannen, J. (eds.) *Handbook of Social Research Methods*, pp 461–478. London: Sage.
- Charmaz, K. (2008c). Grounded theory as an emergent method. In Hesse-Biber, S. N. and Leavy, P. (eds.) *The Handbook of Emergent Methods*, pp 155–170. New York: Guilford.
- Creswell, J. W. (2005). *Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research*, 2nd edn. Upper Saddle River, NJ: Pearson Merrill Prentice-Hall.
- Glaser, B. G. and Strauss, A. L. (1965). *Awareness of Dying*. Chicago, IL: Aldine.

- Green, D. O., Creswell, J. W., Shope, R. J., and Clark, V. L. P. (2007). Grounded theory and racial/ethnic diversity. In Bryant, A. and Charmaz, K. (eds.) *Handbook of Grounded Theory*, pp 472–492. London: Sage.
- Morse, J., Stern, P., Corbin, J., et al. (2008). *Developing Grounded Theory: The Second Generation*. Walnut Creek, CA: Left Coast Press.

Relevant Websites

<http://sbs.ucsf.edu> – Anselm Strauss website.
<http://www.groundedtheory.org> – The Grounded Theory Institute.
www.dahsm.medschool.ucsf.edu – University of California, San Francisco: School of Medicine.

Hermeneutics

J Dyer, Memorial University of Newfoundland, St. John's, NL, Canada

© 2010 Elsevier Ltd. All rights reserved.

The ancient messenger god Hermes transforms that which is beyond human understanding into intelligible form and invents language and writing, the tools of intelligibility.

The field of hermeneutics refers to a manifold of methodologies of interpretation, understanding, and the related issues of translation, communication, and information exchange. It includes practices of interpretation and theories of understanding, where both involve intellectually comprehending the meaning of things and justifying that meaning to others. For this reason, hermeneutics is fundamental to contemporary theories of education because they are guided by the same concerns.

Ancient and Medieval Background to Hermeneutical Inquiry

The ancient Greek terms *hermeneuein* and *hermeneia* mean to assert, to express, and to translate and they are a topic of discussion in texts by Euripides, Xenophon, Plutarch, Epicurus, Longinus, and Plato. Plato (428–328 BC) uses *hermeneuein* to refer to the stated meaning of poetry or rhetoric, distinguishing it from the conscious reflection of *sophia*. Aristotle pushes the relation of hermeneutic enquiry toward truth in his text on language and logic, the *Peri hermeneias*, by placing *hermeneia* with the intellect, where it refers to the mind's operation of perceiving truth value in statements. Philo of Alexandria (20 BC–50 AD) takes up *hermeneia* as a practice of systematic interpretation that can reveal allegorical meanings. In Augustine (AD 354–430), hermeneutics is a practice of systematically revealing layers of meaning in the Bible. The practice of understanding and communicating divine truth involves a basic threefold of will or love, mind, and knowledge. Understanding begins with an intuition of divine truth that imprints on but is hidden in the mind. In order to rediscover this truth, the self requires a way to identify it. For Augustine, this way is love. Love brings together self and divine other and only when love binds the two together can mind realize knowledge of divine truth in an ongoing process of continual self-realization with the other. Augustine's scriptural hermeneutic of understanding brings together language, interpretation, and self-understanding in a triadic model that foreshadows the later theories of Friedrich Schleiermacher and of Hans-Georg Gadamer.

Medieval interpretations of Scripture develop Philo and Augustine by calling attention to the allegorical levels

of sacred texts. For the Church Fathers, understanding as interpretation is both an individual and an inter-subjective experience guided by tradition. Tradition operates as a set of accepted rules used to clarify scriptural interpretation. While tradition is open to misunderstandings, it reduces misunderstandings because it accounts for the possibility of historical change that can break with accepted doctrine. The problem of history gives rise to the hermeneutics of medieval theology and law, where a normative, authoritative tradition of interpretive practices establishes a canon of rules or methods of right practice rather than a theory of understanding.

There are obvious connections between the ancient and medieval practices of understanding and the modern theories of understanding. The former establishes as a part of hermeneutics the practice of consciously reflecting on meaning, the notion of layers of meanings, the problems of authorial intention, historical transmission and the role of tradition, and the concept of a mediating principle which relates self and other, or interpreter and text. Before the modern move into a theory of understanding, however, Reformation and Humanist thinkers problematized the roles of tradition and history, a development fundamental to the transformation of hermeneutics from a practice into a theory.

Humanism and The Introduction of History into Hermeneutics

Hermeneutics as a theory of understanding takes on its modern meaning after the Reformation. Martin Luther stressed that faith and the Bible are the only sources of religious authority and cannot be mediated by tradition. This led to the Reformation insistence on literal exegesis of what the text directly states and on the study of Greek and Hebrew, bringing a historical dimension into scriptural study. The emphasis on literal interpretation opens the possibility of individual interpretation of scripture, where meaning is produced by and for the individual reader of the Bible. The Protestant principle of Scripture explaining itself replaces tradition with an historical understanding of scripture buttressed by faith. Eighteenth-century theologians extend this rejection of tradition and develop the historical dimension of texts, considering scripture a historical document to be understood in its historical context, thereby making interpretation increasingly critical and historically defined. The upshot of this account is that the object of

hermeneutics, rather than the practice, becomes problematized and history becomes central.

The centralization of history is re-enforced by Humanist thinkers, such as Giambattista Vico (1668–1744) and Benedict de Spinoza (1632–77). Neither outlines a system of interpretation, but both emphasize the importance of history for all understanding, thereby laying the ground for the Romantic hermeneuticians. Vico argues that the true is the constructed, such that we can only have certain knowledge about things we create, such as cultural, political, linguistic, economic, and intellectual structures. The natural sciences offer approximate truths because they imitate natural processes, but the social sciences provide exact knowledge because their objects are created. Vico makes history primary and, in doing so, he strengthens the fundamental position of the interpretive subject, for if we are historical beings, then to understand is to understand oneself. The world is not mind independent but identified with the historical subject. Historical, interpretive analysis becomes the apex of scientific knowledge.

Similarly, Spinoza locates the structure of understanding in a historical context when he claims that understanding must place its objects in their historical framework. His *Tractatus Theologico-Politicus* treats the Bible as a historical document that reflects the intellectual practices of its culture; further, God and nature are held to be one substance. Spinoza thereby proposes an analogical relation between the natural and human sciences in which the method of scientific inquiry is made into a historical, proto-dialectical process of relating parts to their contextual wholes, such as Scripture to its context, or natural processes to God, where each can only be understood in relation to the other.

The Romantic Tradition: The Precursor to Modern Hermeneutics

A number of influential figures develop theories of interpretation out of humanism. For example, the eighteenth-century theologian J.S. Semler (1725–91) enforces the distinction in interpretative practice between personal contingent interests and universal permanent meanings while the philologist J.A. Ernesti (1707–81) brings the grammatical to the historical and logical readings of texts. The crux of their analyses is that the linguistic is as explanatory as the historical in objectively interpreting historical documents. It is Johan Martin Chladenius (1710–59), however, who provides a truly systematic treatment of interpretation theory. His hermeneutics is a theory of understanding that offers an account of interpretive procedure. First, texts address readers and are intended to induce complete understanding. Second, understanding requires the use of reason, located resolutely in the human mind, in conjunction with common sense. Third, the reader must be suspicious of

everything that is not self-evidently true. Chladenius takes up the notion of perspectivism (articulated in G.H. Leibniz' monadism) and emphasizes the hermeneutic significance of point of view. Misunderstanding is due to historical, subjective changes in perspective, so Chladenius outlines a typology of points of view intended to clarify opaque meanings. Reading, he argues, requires skeptical interpretation, achieved through trial and error and the continual negotiation of points of view to determine the true meaning of a text. Reading is a fundamental science that results in objectively correct interpretations which can be justified. For Chladenius, hermeneutics is necessary because texts are opaque; his theory predicts the twentieth-century hermeneutics of suspicion.

G.F. Meier's (1718–77) theory of signs holds that signs are meaningful only in relation to other signs. The meaning of a word is not dependent on the object it represents, but on the entire differential structure of words. Hermeneutic theory takes up Meier's focus on interpretation in terms of the differential linguistic structure of texts without appeal to extra-linguistic aspects such as authorial intention, Divine mind, or the object-referent of the sign. Meier's account initiates the practice of relating sign and structure, or parts and wholes in ongoing interpretative activity.

Still more influential are the pre-Romantic classical philologists Friedrich Ast (1778–1841) and Friedrich August Wolf (1759–1824), both of whom extend the domain of understanding from the interpretation of historical texts (Chladenius) and the relational interplay of signs in a system (Meier) to knowledge and understanding as a whole. Ast argues that the history of understanding must be put into the context of the entire history of reason and that the meaning of individual texts can be understood only when interpreted in the context of world history. Wolf's classical philology is developed as a science of interpretation in which historical, cultural, and linguistic elements are critically understood only in the context of the whole of world history. Influenced by Idealism and the emerging Romanticism, both thinkers push the bounds of the hermeneutic framework to include not just the author's point of view and historical situation but the entire historical horizon of human reason.

The Romantic Emergence of Modern Hermeneutics

The theologian G.W.F. Schleiermacher (1768–1834) is a key figure in the development of hermeneutics. In *Hermeneutics and Criticism*, he argues that hermeneutics is the universal condition of all understanding and his general hermeneutic theory is meant to overcome misunderstanding by combining a scientific, methodical practice of interpretation with empathy. Hermeneutic practice

unites the objective, linguistic-historical, and the subjective theological dimensions of interpretation because, he claims, in order to understand a text we must understand it even better than its author did. Schleiermacher thus advocates re-experiencing the text from the perspective of the author, but with the benefit of knowing the author's historical context. Interpretation interrelates the text's grammatical sense with the author's intentional or technical meanings. Grammatically, language is the shared medium of understanding, so to grasp the text's meaning, the interpreter must understand its syntax and grammar. In his technical re-experiencing of a text, interpretation requires an intuitive leap or act of divination to relive the consciousness of the author and grasp her style or treatment of language. There is no directing tradition for this divination; Schleiermacher's subjective idealism grants the interpreting subject the intuitive capacity to move from universal to particular. In this way, Schleiermacher's contribution is in defining interpretation as an interrelational process of continually moving from the universal to the individual, grammatical to technical in a circular process of development. The goal of correct interpretation may not be fully realizable, but it nevertheless requires a sustained and focused comprehension of the author's historical context and unique subjectivity. Only when interpreters can put aside the obstacle of their own presuppositions and relive the author's historical experiences can the meaning fully be comprehended. Schleiermacher's process of interrelating comparative and intuitional, syntactical, and intentional aspects of interpretation develops what will become the hermeneutic circle of understanding and interpretation, or speculation and reflection, developed later by Martin Heidegger, Gadamer, and Paul Ricoeur.

Schleiermacher's systematic general hermeneutics was taken up by thinkers such as Alexander von Humboldt, Friedrich Carl von Savigny, Leopold von Ranke, and Johann Gustav Droysen and applied to their various disciplines. Ranke's (1795–1886) historiographic appropriation of hermeneutics explains historical understanding through the continual interrelation of particulars and universals with the aim of achieving consciousness of historical life. Droysen (1808–84) contends hermeneutics allows for objective understanding of the telos of history by manifesting the mediation of historical being through traditions of understanding. However, it is Wilhelm Dilthey (1833–1911), Schleiermacher's biographer, who clarifies his general hermeneutics by continuing Schleiermacher's emphasis on revealing the conscious knowledge of the subject within its surrounding living totality. Dilthey's *Formation of the Historical World in the Human Sciences* broadens the scope of these insights, reasoning that history and the conscious, intentional subject share a reliance on symbolic expression or mediation. Meaningful expression is a form of symbolic mediation; therefore, understanding the world requires understanding its

mediation through symbolic expressions, a point which is developed with reference to two key concepts: *Erlebnis* (self-understanding) and *Verstehen* (understanding others). There is no understanding of self without an understanding of the self as mediated through the world shared by others. *Erlebnis* and *Verstehen* mediate each other, indicating a universal human nature. Moreover, Dilthey rejects Schleiermacher's psychological intentionalism by claiming that all experience is mediated by pre-reflective symbolic expressions. This means symbolic mediation reveals universal history. Dilthey maintains an inextricable, ongoing relationship of self and other in his theory of understanding, opening up the human sciences to a study of the objective conditions of life experience, pointing toward the twentieth-century ontological turn in hermeneutic theory.

The Twentieth-Century Ontological Turn in Hermeneutics

The philosopher Martin Heidegger (1889–1976) is influenced by Dilthey's attempt to re-live the experiences of past human expressions, but he rejects Dilthey's subjectivism, claiming that the entirety of historical experience cannot be represented in historical self-understanding. Heidegger's hermeneutics of facticity, presented in his work *Being and Time*, recognizes that understanding is primarily about practical experience, not conceptual meaning. Heidegger argues that there is no getting outside historical situatedness. The world is made intelligible through our everyday, practical and pre-reflective projections onto the world of our presumptions, expectations, and categories, the horizon of experience that informs understanding. Understanding is neither a methodology nor a realm of experience but the human mode of being, or *Dasein*, a means of orienting ourselves in the world that is limited by the facticity of historical existence. Reflective interpretation is therefore always already shaped by this horizon of understanding. While understanding discloses the world as a totality in which being is at home, we cannot access anything beyond. Interpretation brings understanding to reflective consciousness by revealing both the facticity of the world and its limits, for there is no understanding of otherness without interpretively appropriating it into our horizon. This is the crux of Heidegger's hermeneutical situation of existence: understanding and interpretation mediate each other and there is no escaping that mediation. The understanding of our mode of being is mediated by the world, and understanding the world is mediated by our mode of being, how the world is disclosed to us. *Dasein* has a circular hermeneutical structure whereby interpretations alter our understanding, which is modified by continually re-interpreting the totality of world. Heidegger's hermeneutics is ontological because it is structured by *Dasein*.

The meaning of things is tied neither to linguistic, propositional truth values nor to the alien historical horizon of past experiences but to the ongoing relational process of the world's disclosure to us. His hermeneutic circle of understanding and interpretation is existential and productive. However, it fails to address the problem of the unappropriable, unintelligible other.

Heidegger's difficulty is addressed by his student Gadamer (1900–2002), who asks in his fundamental work *Truth and Method* what we can do with things that cannot be intelligibly appropriated into our world. Gadamer holds that our understanding is formed by the prejudices of our horizon of understanding, but that these prejudices are the basis of intelligibility, or understanding and misunderstanding. Prior to any reflective identification of meaning there is the world-disclosing interaction of understanding and interpretation revealed through language. That is, we are linguistic beings who recognize, interpret, and change in relation to the world's self-disclosure to us through language, a process of understanding with four components.

First, Gadamer includes others in his hermeneutic circle of understanding and interpretation. Understanding a thing is a matter of projecting it onto our horizon of experience; prejudice or, better, tradition mediates rather than obscures understanding. To be is to be practically immersed in a tradition of meanings, values, and expectations that shape understanding. Inevitably, however, tradition cannot fully interpret the thing, throwing understanding back to reflectively modify its grounds of intelligibility. Hitting an interpretive impasse reveals that tradition is limited, but flexible. The hermeneutic circle of understanding and interpretation is a process that constantly discloses and changes the limits of intelligibility. That is, the traditions historically handed down as our horizon of understanding bear the potential to be modified, crucial to the possibility of further understanding, and in modifying our traditions we open their potential toward understanding the alien object.

Second, this does not imply Heidegger's appropriation of the object into our world because interpretation expands tradition by critically reconsidering its limits. Engagement with retractable objects begins a dialog with the other that requires our openness to the unrealized potential effects of history. History has an authority as both part of and more than, preceding and informing tradition such that tradition is a never completely realized basis of intelligibility that is open to change. Gadamer calls it effective history, the continual expansion of tradition that enriches self-understanding; in understanding the other, we change.

Thus, thirdly, Gadamer describes the mode by which we engage with the other as play. Without appropriating the unintelligible, understanding plays with it by submitting to or fully immersing itself in its structure as we would in, say, a game or conversation. Play is an open relation to the other which begins interpretation, suggesting that the

process of interpretation and understanding is dialogic and open-ended.

Fourth, this means that playing with the unintelligible productively opens our horizon of understanding or prejudices to another's. It enables a fusion of horizons. The hermeneutic circle is now an interplay of effective history and responsible interpretation that neither absorbs the other into one tradition nor denies the possibility of new understanding. In this respect, Gadamer's hermeneutics is an indefinitely ongoing process of dialogical transformation of both object and subject, or text and reader. Like Heidegger, Gadamer's account can be considered idealist because it deals only with the human world. But Gadamer's is also a practical philosophy of enriching our world.

Gadamer's hermeneutical ontology was rejected entirely by Emilio Betti (1890–1968) who charges Gadamer with destroying the classical hermeneutical ideal of objective understanding by denying the independent objectivity of the object. Texts, for instance, are objective representations of an author's intentions with only one true meaning that is not connected to the interpreter's subjective horizon. Returning to Schleiermacher, Betti insists that interpretation must recreate the author's intended meaning. Similarly, E.D. Hirsch (b.1928) argues that Gadamer ignores the possibility of right and wrong interpretations. Gadamer, however, is not engaged in distinguishing interpretation rules but in analyzing the structure of understanding. His hermeneutics is ontological, outlining the structure of being itself.

Jurgen Habermas (b.1929) critically considers Gadamer's hermeneutics with reference to the problem of validity. Habermas' famous debate with Gadamer concerns Gadamer's emphasis on the authority of tradition at the expense of critical judgment. Habermas holds that hermeneutics requires a distinction between distorted and true communication found only in the standards of validity of communicative reason. Gadamer responds that hermeneutics is universal because situatedness in a tradition is the possibility condition of any intellectual inquiry and the revision of tradition is thereby critical. Still, Habermas insists that because inquiry is also always directed by sets of interests (practical, technical, or emancipatory), no theoretical inquiry is universal. Intellectual inquiry which aims for self-knowledge must engage in a critical hermeneutics to account for ideological distortions of tradition. This is the critique of ideology found only in the critical social sciences because they have the goal of emancipation. The critique of ideology questions the traditions it takes up, exposing and amending distortions. Karl-Otto Apel (b.1922) shares Habermas' concern with the emancipatory character of the critical social sciences; Apel calls for an ideal communicative community similar to Habermas' complete communicative competence, namely a regulative ideal of freedom underlying the critique of ideology that questions history's

authority. Without a critical hermeneutics, Gadamer's fusions of horizon can be an unrecognized fusion of misunderstandings.

Phenomenological Hermeneutics and the Semiotic Turn

Paul Ricoeur's (1913–2005) phenomenological hermeneutics mediates between Gadamer's hermeneutics of tradition and Habermas' critique of ideology by installing critique within the hermeneutic process. When the unintelligibly other admits to no fusion of horizons, neither thinker explains how we can responsibly come to understand the other. Ricoeur answers by expanding the hermeneutic circle to include explanation, understanding, and appropriation. Explanation determines the object's structure, since the formal structure of, say, a text is mind independent, open to certain interpretations and impervious to others. Understanding discerns the meaning of the subject matter, such as the author's chosen words. Interpretation compares our understanding of the text's meaning with the explanation of its form to approach an increasingly determinate interpretation out of its possible meanings. Interpretation is open and never fully determinate, for interpreters engage different traditions and subjective associations with the text. Ricoeur's hermeneutic circle sustains the objective independence of the other in relation to the subject, while maintaining their dialogical openness to change.

In *Time and Narrative*, Ricoeur develops the hermeneutical role of tradition, admitting that there is no stepping outside of tradition and that even Habermasian critique is an authoritative form of tradition. However, Ricoeur calls for a hermeneutics of historical consciousness which reveals the implicit distinctions within tradition by proposing another threefold of traditionality, traditions, and tradition. First, all thinking is situated in traditionality because it is embedded in a communicated history of prejudices. Traditionality is similar to explanation, for both concern formal structure. It is transcendental, for it is the purely formal, universal structure of situatedness in some tradition. Second, traditions are the particular symbolic heritages of meaning that cultures inherit differently from history. Third, tradition is specifically one's own heritage that is accepted as the legitimate tradition above all others. Whatever tradition we find ourselves in is *de facto* the normative tradition or Heideggerian world with which our interpretation engages, even when dialogical engagement changes it. Alternately, tradition is what Ricoeur calls narrative, that which engages with the symbolic capacity of others.

The relation of narrative and symbol is crucial to Ricoeur's hermeneutic theory and the basis of his hermeneutics of suspicion, the discernment of what a text hides.

On the one hand, symbols are objects of interpretation, inciting interpretation because they suggest more than their self-evident facticity and they can only be understood when dialogically engaged. Initially, speculation interprets the symbol pre-reflectively in terms of our own tradition, but upon reflection we realize the potential for further meanings not encompassed by tradition. Like explanation, speculation is directed by the symbol's materiality while, like interpretation, reflection opens those traditional meanings and associations to further possible understandings. Symbols show there is more to know than tradition suggests. On the other hand, narrative concerns the possibilities of one's own being. In a return to hermeneutic's ontological thrust, Ricoeur argues that narrative is the specificity of the interpreter's tradition. Our lives are narratives which evince unified structures not entirely explicable by our subjective intentions because they are structurally open to affect and be affected by others. For Ricoeur, hermeneutics concerns continually realizing new ways to exist in the world.

Despite Ricoeur's attempt to incorporate objective determination into the hermeneutic circle, his approach remains firmly with the idealism of his predecessors, for Ricoeur's determination of meaning is still grounded in the interpreting subject's willing, dialogical engagement with others. In his *Of Grammatology*, Jacques Derrida (1930–2004) takes issue with the dialogic approach, arguing that the appropriation of Heidegger's ontology misunderstands its power. With echoes of Meier, Derrida contends that meaning is not dependent on the fusions made in open-ended dialog but on their lack. What drives interpretation is precisely the misunderstandings upon which Gadamer bases his hermeneutics, the absence of meaning that impels us to make connections within language's interminably differential system of signs. Derrida questions the possibility of responsibly understanding the other because interpretation never allows full understanding.

The enduring influence of hermeneutics can be found in a variety of thinkers in the latter half of the twentieth century, such as Richard Rorty, Richard Shusterman, John McDowell, Donald Davidson, and Mary Hesse. Recently, hermeneutics has been challenged as both an interpretive practice and a theory of understanding by the semiotics of C.S. Peirce (1839–1914) and followers like Umberto Eco (b.1932). They understand interpretation as having a threefold structure of sign, object, and interpretant. Against the idealist orientation of hermeneutics, this structure is held to constitute the real nature of all things: the universe is a triadic system of information exchange or communication, of which human activities are but an instance. Ultimately, however, from Plato through to Ricoeur, hermeneutics offers a deeper understanding of the educational process, as Shaun Gallagher argues in his *Hermeneutics and Education*. In relation to educational

theory, hermeneutics is employed as a dialogical mode of learning, for it shares with education theory the basic tenets of analyzing and developing notions of understanding and interpretation, authority and tradition, communication and distortion, and self-realization and objectivity.

See also: Critical Theory; Discourse Analysis; Interpretive Research; Narrative Inquiry; Phenomenology; Semiotics.

Further Reading

- Bleicher, J. (ed.) (1980). *Contemporary Hermeneutics*. London: Routledge and Kegan Paul.
- Bradley, J. (2009). Beyond hermeneutics: Peirce's semiology as a trinitarian metaphysic of communication. *Analecta Hermeneutica* 1(1), 45–57.
- Bruns, G. L. (1993). *Hermeneutics Ancient and Modern*. New Haven, CT: Yale University Press.
- Derrida, J. (1976). *Of Grammatology*, Spivak, G. C. (trans.). Baltimore, MD: The Johns Hopkins University Press.
- Dilthey, W. (2002). *The Formation of the Historical World in the Human Sciences*, Makreel, R. A. and Rodi, F. (eds.). Princeton, NJ: Princeton University Press.
- Gadamer, H. -G. (1994). *Truth and Method*, Weinsheimer, J. and Marshall, D. G. (trans.). New York: Continuum.
- Gallagher, S. (1992). *Hermeneutics and Education*. New York: State University of New York Press.
- Grondin, J. (1994). *Introduction to Philosophical Hermeneutics*. New Haven, CT: Yale University Press.
- Heidegger, M. (1962). *Being and Time*, Macquarrie, J. and Robinson, E. (trans.). San Francisco, CA: Harper.
- Hirsch, E. D., Jr. (1967). *Validity in Interpretation*. New Haven, CT: Yale University Press.
- Ormiston, G. L. and Schrifft, A. D. (eds.) (1990). *The Hermeneutic Tradition from Ast to Ricoeur*. New York: State University of New York Press.
- Palmer, R. E. (1969). *Hermeneutics: Interpretation Theory in Schleiermacher, Dilthey, Heidegger and Gadamer*. Evanston, IL: Northwestern University Press.
- Palmer, R. E. and Michelfinder, D. (eds.) (1989). *Dialogue and Deconstruction*. New York: State University of New York Press.
- Piercey, R. (2004). Ricoeur's account of tradition and the Gadamer-Habermas debate. *Human Studies* 27(3), 259–280.
- Ricoeur, P. (1984). *Time and Narrative* (3 vols.), McLaughlin, K. and Pellauer, D. (trans.). Chicago, IL: University of Chicago Press.
- Schleiermacher, F. (1998). *Hermeneutics and Criticism*, Bowie, A. (ed. and trans.). Cambridge: Cambridge University Press.
- Shalin, D. (2007). Signing in the flesh: Notes on pragmatist hermeneutics. *Sociological Theory* 25, 193–224.

Interpretive Research

E Rauscher and M E Graue, University of Wisconsin–Madison, Madison, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Chronotope – The manner in which literature represents time and space.

Constructivism – An approach to inquiry that positions knowledge as constructed rather than received. Knowledge is culturally and historically situated and replete with values.

Critical theory – A theory that emphasizes the importance of integrating knowledge across disciplines in order to illuminate the many ways in which power functions in societies.

Empiricism – A view that gives primacy to sensory experience as a source of knowledge.

Epistemology – A theory of knowledge. Positivism and constructivism are examples.

Modernist – A point of view that embraces Enlightenment ideals such as empiricism and progress through science.

Positivism – A philosophy that posits that objective truth is the essence of knowledge.

Postmodernist – Related to the theory that involves a radical reappraisal of modern assumptions about classification, culture, identity, history, and language.

Post-positivism – An approach to inquiry that positions human knowledge as a series of conjectural theories, which may be recursively revised with the discovery of new information.

Transformationist – Related to the theory that ideas, beings, identities, histories, languages, and cultures have originated by the modification of some other previously existing forms.

Introduction

Over the past few decades, scholars throughout the field of education have increasingly engaged in what is generically termed interpretive research. While interpretive research has amassed considerable currency in social research writ large, divergent perspectives have emerged as researchers have variously parsed what it means to conduct interpretive research and how such research ought to be distinguished from other forms of inquiry. This article provides an introduction to central concepts and tensions that characterize the genre of interpretive research in contemporary

education inquiry. Specifically, drawing upon the work of key theorists from the social and human sciences within and beyond the field of education, this article identifies three archetypal approaches to interpretive research as represented by three groups of thinkers and doers who share ideas while staking distinct turf in the research landscape: the signpost group, the big tent group, and the hybrid identities group. Before addressing each group in turn and concluding with a discussion of the viability of the interpretive research genre, this article begins with an overview of the subject of interpretation.

The Matter of Interpretation

The term interpretive research invites a conspicuous question: What do interpretive researchers set out to interpret? Although answers to this question abound throughout disciplines concerned with social and cultural theory, most education research textbooks find firm footing in anthropology when tracing the antecedents of contemporary interpretive inquiry. Specifically, the words of Clifford Geertz, an American cultural anthropologist, are commonly cited:

Interpretive explanation ... trains its attention on what institutions, actions, images, utterances, events, customs, all the usual objects of social science interest, mean to those whose institutions, actions, customs, and so on they are. ... Inquiry is directed toward cases or sets of cases, and toward the particular features that mark them off; but its aims are as far-reaching as those of mechanics or physiology: to distinguish the materials of human experience. (Geertz, 1983: 22)

The overarching intent of Geertz' project was to describe (and thereby interpret) the constitutive components of human experience in such a way that aspects of a group's culture could be understood by people outside that culture. One classic example of his work consisted of describing a Balinese cockfight and arguing how it can be used as a window for understanding gender roles in Balinese culture (1973). The significance of this particular work was two-fold. First, by describing in evocative detail the practice of cockfighting engaged by the Balinese people, Geertz invited readers to experience it vicariously. Second and, moreover, through this experience, readers are compelled to understand how the cultural text of the cockfight is not merely a game but a rulebook for masculinity and power.

In education, many researchers take a similar approach. They describe human experience as it relates to their fields and subfields. They immerse themselves in the culture of a classroom, school, or other relevant organization and provide their audiences with a rich, multilayered account of what occurs in that particular context. The model for this descriptive work is derived from the British philosopher Gilbert Ryle's notion of thick description (introduced to many US audiences by Geertz in 1973). A thick description is a detailed narrative account produced through prolonged fieldwork engaging multiple sources of data and multiple participants. However, like Geertz, interpretive researchers in education do not solely provide highly descriptive accounts of lived experience. They also attempt to make sense of how norms, actions, customs, gestures, dispositions, etc., work together in that (school, classroom, or institutional) culture, thereby generating meaning through their interpretations.

The connection between thick description and meaning generation can be seen by considering the following classic example. Two boys are rapidly contracting their right eyelids. What does this rapid eyelid contraction mean? Through focused investigation the researcher learns that one of the boys has a twitchy eye (i.e., his movements are involuntary) and the other boy is winking at a friend. Although their actions are physically similar, it is only because the researcher is fully embedded in his/her culture that he/she can distinguish between the two actions. A third boy enters the scene and is also contracting his right eyelid. This boy is parodying a wink for comedic purposes. Thin or simple description would not and could not distinguish among these three actions, while the goal of thick description is to do just that. This is because what distinguishes interpretive research from other forms of description is its interpretive intent, namely, its commitment to understanding that focuses on the meanings of actors in local settings.

In summary, interpretive researchers set out to interpret the meaning inherent in concrete, observable phenomena – a position that is exemplified and continuously advanced by the signpost group and divergently extended by the big tent and hybrid identities groups.

The Signpost Group

The signpost group of interpretive researchers in the field of education is made up of individuals who link their identities to what many call the interpretive turn in social research – a concept about which Rabinow and Sullivan (1987) offer several key points:

The *interpretive turn* refocuses attention on the concrete varieties of *cultural meaning*, in their particularity and complex texture, but without falling into the traps of historicism or cultural relativism in their classic forms.

For in the human sciences both the *object of investigation* – the web of language, symbol, and institutions that constitutes signification – and *the tools* by which investigation is carried out share inescapably the same pervasive context that is the human world. Interpretation begins from the postulate that the web of meaning constitutes human existence to such an extent that it cannot ever be meaningfully reduced to constitutively prior speech acts, dyadic relations, or any predefined elements. Intentionality and empathy are rather seen as dependent on the prior existence of the shared world of meaning within which the subjects of human discourse constitute themselves. (Rabinow and Sullivan, 1987: 6, emphasis added)

First, they note that interpretive research focuses on local lived experience and that its description is produced through a multilayered interpretive process – participants and researchers are mutually interpreting their interactions, in dialog with their histories and social networks. This means that all interpretive research is located in specific times and specific places. As a result, interpretation at every level is referenced – meanings are in relation to particular things. Charles Taylor specified three elements of meaning:

- (a) Meaning is for a subject: it is not the meaning of a situation *in vacuo*, but its meaning for a subject, a specific subject, a group of subjects, or perhaps what its meaning is for the human subject as such . . . (b) Meaning is of something; that is, we can distinguish between a given element – situation, action, or whatever – and its meaning. . . (c) Things only have meaning in a field, that is, in relation to the meanings of other things. (Taylor, 1979: 41)

This focus on situated interpretation shifts the researcher from a passive role describing a separate, given reality, to an engaged actor who uses personal and cultural understandings in interpreting experience. This is the essence of the interpretive turn.

Within the interpretive turn is the introduction of a particular mode of authority for the interpretive researcher. Enacting Taylor's elements of meaning, the value of an interpretive account comes out of a sense of being there, an ability to convince the reader that the analysis comes out of the researcher's lived experience in the research context:

In fact, what gives the ethnographer authority and a text a pervasive sense of concrete reality is the writer's claim to represent the world as only one who has known it firsthand can, which thus forges and intimate link between ethnographic writing and fieldwork. (Rabinow, 1986: 23)

By using the rhetorical power of a text depicting field experience, interpretive research locates meaning in a context and persuades through descriptive portrayal. However, this approach has a notable drawback:

Which brings us, finally, to theory. The besetting sin of interpretive approaches to anything – literature, symptoms, culture – is that they tend to resist, or to be permitted to resist, conceptual articulation and thus to escape systematic modes of assessment. You either grasp an interpretation or you do not, see the point of it or you do not, accept it or you do not. Imprisoned in the immediacy of its own detail, it is presented as self-validating or, worse, as validated by the supposedly developed sensitivities of the person who presents it; any attempt to cast what it says in terms other than its own is regarded as a travesty – as, the anthropologists severest term of moral abuse, ethnocentrism. (Geertz, 1973: 24)

The Big Tent Group

Written in the late 1980s as qualitative approaches to educational issues were finding their way into the mainstream, Frederick Erickson's chapter on qualitative methods provides an historical bridge between the psychologically oriented perspectives on teaching to the interpretively oriented viewpoints coming out of anthropology, sociology, and other disciplines using qualitative methods:

I will use the term *interpretive* to refer to the whole family of approaches to participant observational research. I adopt this term for three reasons: (a) It is more inclusive than many of the others. . . . (b) it avoids the connotation of defining these approaches as essentially nonquantitative; . . . and (c) it points to the key feature of the family resemblance among the various approaches – central research interest in human meaning in social life and in its elucidation and exposition by the researcher. (Erickson, 1986/1990: 77–78)

At the time, interpretive research was an up-and-coming alternative to the statistical inference and causally motivated approaches that predominated research on teaching. Erickson provided an articulate case for the contributions of interpretive research, positioning his argument in relation to quantitatively oriented approaches and from within an interpretive framework.

For Erickson, interpretive research is differentiated from other forms of inquiry in its intentions and substance. Method does not serve to distinguish interpretive research; instead, its critical attribute is that its purpose is to examine meaning. Erickson's discussion of interpreting the patterned nature of lived experience highlights the social theorization so important in interpretive research:

Humans, the interpretive perspective asserts, create meaningful interpretations of the physical and behavioral objects that surround them in the environment. We take action towards the objects that surround us in the light of our interpretations of meaningfulness. Those interpretations, once made, we take as real – actual qualities of the objects

we observe. . . . We see the ordinary world as if it were real, according to the meanings we impute to it. (Erickson, 1986: 97)

This approach operates such that the focus is on concrete particulars rather than on abstract universals. Returning to the example of the boys who were rapidly contracting their right eyelids, the particular details of each boy's situation must be learned in order to fully understand why he is performing the action. If the second boy is unable to speak with his friend due to the situation they are in (e.g., school during an examination), the boy could be communicating a number of things by winking. Perhaps in the school, it is a custom for groups of students to produce practice tests in preparation for an exam. If they attempt to accurately guess the questions on the examination, the wink could indicate one of them guessed correctly and they will plan to work together again. To understand the situation, both the concrete particulars (the boy winked, to a friend, during an examination) and the abstract universals (people wink to get each other's attention in a friendly way) would have to be known. Importantly, the researcher could not claim this to be the case in all schools. His/her findings are not generalizable. Interpretive researchers work a tension among universals, limited generalizations, and local uniqueness, noting, "The paradox is that to achieve valid discovery of universals one must stay very close to concrete cases" (p. 109). Building from the interpretive turn, Erickson introduces interpretive research to a cultural group socialized to value a post-positivist approach to research, linking key concepts to research across paradigms.

Another example of a broadly inclusive framing of interpretive research is Howe's (1998) philosophical exploration of education's engagement in the interpretive turn (and its aftermath). Under an epistemological umbrella of constructivism, Howe places interpretivism in relation to empiricism (which he links to positivism). His definition of this approach to inquiry is derived from the interpretive turn and represented by Taylor's rejection of value-free science communicated with impersonal language: "We have to think of man as a *self-interpreting* animal. . . . [T]here is no such thing as the structure of meanings for him independently of his interpretation of them" (Howe, 1998: 13). This provides a context for his view of interpretivism as a "broad epistemological turn" focused on the culturally embedded identities and interests (p. 13). While this perspective might seem quite similar to those suggested by those in the interpretive signpost group, what distinguishes it is the way it frames the rest of the world.

Placing interpretivist research in opposition to positivist research, Howe's definition sets multiple posts for a big tent encompassing all research endeavors. This is clear in his inclusion of two distinct approaches to interpretivism, what he calls postmodernist and transformationist. The former he links to a rejection of the modernist core of

Enlightenment philosophies and the latter he sees as holding true to the commitment to modernist notions of progress. This big tent philosophy toward interpretive research takes the notion of interpretive practice so far that it makes ascertaining what is not interpretive rather difficult. Arguably, the recognition that researchers do not have direct access to lived experience and that they interpret it as participants and investigators stands to benefit more from fine-tuning than from judgment regarding its placement inside or outside that tent. For this reason, the next group – the hybrid identities group – can be understood as being characterized by more permeability.

The Hybrid Identities Group

Critiques of interpretive research have highlighted the conceptual differences in assumptions among theoretical groups. Garrick (1999) examined the foundations of interpretive research, identifying three sets of theoretically situated concerns. From the positivist perspective, interpretive research is limited because interpretation implies subjectivity, which short-circuits the mechanisms of generalization. For critical theorists, interpretive research does not have the tools to interrogate historical, social, economic, structural, and environmental forces on the individual that can be used to investigate issues of power. Finally, postmodernists find interpretivism's focus on individual experience problematic because of its contention that individuals are largely autonomous and agentive subject: "Human beings do not live in worlds entirely of their own devising" (p. 152).

Together, these critiques illustrate the limited contribution of framing interpretive research as all things that interpret. Researchers whose affinity lies with the hybrid identities group assert that the interpretive landscape is much more nuanced than the big tent metaphor would imply and that researchers lose the generative power of the contributions that each theoretical perspective provides.

A variety of philosophers and qualitative researchers have parsed the terrain of inquiry, with attention to the ways that epistemology shapes the practice and criteria for judging research. Bredo and Feinberg (1982) contrasted positivist, interpretivist, and critical theory in terms of the relation between the researcher and the researched, which range from absolute separation to mutually constituting each other. Their portrayal of interpretivism as a theoretical perspective that creates meaning through interaction makes it subject to the same critiques as those suggested by Garrick.

Crotty (1998) framed social inquiry in terms of methods, methodology, theoretical perspective, and epistemology and argued that good scholarship has these elements coherently related to one another. He depicted the world epistemologically in terms of objectivism, constructionism, and subjectivism, and then described an array of theoretical perspectives. Interpretivism is one of those theoretical

perspectives, "conceived in reaction to the effort to develop a natural science of the social. Its foil was largely logical empiricist methodology and the bid to apply that framework to human inquiry" (Schwandt, 1994: 125). For Crotty, interpretivist research "looks for culturally derived and historically situated interpretations of the social life-world" (p. 67). While this definition may sound very similar to earlier views, what distinguishes his framework is its approach to mapping that loosely links elements into groups. He reminds researchers to maintain awareness of how tools (like methods) have deeply embedded meanings related to ways of seeing the world. By extending this mapping beyond the quantitative-or-not distinction, Crotty not only complicates but also enriches researchers' understandings of interpretive research. Interpretivism is more fully linked with specific theoretical frameworks – in Crotty's typology it is paired with hermeneutics or symbolic interactionism in relation to post-positivism, critical theory, feminism, and postmodernism.

Kamberelis and Dimitriadis (2005) developed a map through an historical analysis of theories foundational to qualitative inquiry that recognize relations in explicit and contradictory ways. Specifically, they posit four overlapping categories or chronotopes – snapshots of time with permeable boundaries and related histories – that "index historical realities that constitute what is common, natural, and expected by collectives of social scientists who conduct particular kinds of qualitative research" (p. 25):

1. objectivism and representation;
2. reading and interpretation;
3. skepticism, conscientization, and praxis; and
4. power/knowledge and defamiliarization.

In essence, these chronotopes are eras or stages of qualitative inquiry. A reading of Kamberelis and Dimitriadis' distinctions between the second chronotope (reading and interpretation) and the other three recognizes a reality that is socially constructed through experience and language while maintaining a core that can be identified through consensus. In particular, in contrast to the two subsequent chronotopes (skepticism, conscientization, and praxis and power/knowledge and defamiliarization), the reading and interpretation chronotope poses language as a natural mediator, one that constitutes our worlds but that is not particularly interested in power and its play in the world. Interpretivists are committed to describing lived experience in ways that enable the illumination of what is uniquely human but that have limited attention to forces that privilege some and marginalize others. Reflecting its roots in an empiricist past (the prior chronotope, objectivism and representation) with colonialist practices, interpretive research can be seen as relativist and realist, socially oriented yet individually focused in many cases.

Adherents of the hybrid identities group would argue that this kind of portrayal of the research landscape allows

researchers to appreciate the complexity at play in qualitative research today. By distinguishing inquiry in this historically reflective way, what could be seen as promoting Balkanization is instead suggesting hybrid descriptions that are more nuanced and interested in diverse dimensions of inquiry reflective of the practice of research. Their permeable boundaries are illustrative of Geertz' notions of blurred and blurry genres, originally written to describe the border crossing between anthropology and literary theory but now layered to include multiple ways to theorize and practice research. Interpretive research, rather than being all things that are qualitative, represents a particular set of theoretical and therefore methodological tools interested in examining the social construction of meaning in particular settings and with specific participants. Its focus on concrete particulars and a consensus model of truth make it a transitional model of inquiry, a theoretical bridge between the modernism of post-positivist approaches and postmodern multiplicity. Its unique blend of realism and relativism gives the researcher a particular power as an author who recounts his/her field experience to create a sense of meaning. Jennifer Greene's description is particularly apt:

Interpretivist social inquiry is an opportunity to give voice to one's self, to offer a view of human experience that promotes one's own values and ideals. Interpretivism retains its essential relativism on this issue, such that it can be used as a platform for many different, even competing, advocacy stances and beliefs. (Greene, 1992: 39)

Conclusion

With the myriad affinity groups circulating various perspectives on interpretive research, is interpretive research a viable term today? In other words, does the existence of multiple meanings ascribed to the term interpretive research mute or amplify its importance to the field of education inquiry? An answer to this question lies in the reader's interpretation of this article, for it is very much an object lesson in interpretive theory and practice. Rather than generating a single definition of interpretive research, this article provides a set of interpretations made through its authors' experiences with the literary discourse espousing a multiplicity of perspectives concerning interpretive research. To be sure, the ideas shared in these pages may do more to

frustrate than to clarify, an experience analogous to an allegory famously presented by Geertz:

There is an Indian story – at least I heard it as an Indian story – about an Englishman who, having been told that the world rested on a platform which rested on the back of an elephant which rested in turn, on the back of a turtle, asked (perhaps he was an ethnographer; it is the way they behave), what did the turtle rest on? Another turtle. And that turtle? “Ah, Sahib, after that it is turtles all the way down.” (Geertz, 1973: 28–29)

The adventure through the interpretive literature would say that the interpretive research term is indeed viable because it is, like turtles, all the way down. However, as this article makes clear, its viability rests on the capacity of the inquirer to identify the perspective used to make the interpretation.

Bibliography

- Bredo, E. and Feinberg, W. (eds.) (1982). *Knowledge and Values in Social and Educational Research*. Philadelphia, PA: Temple University Press.
- Crotty, M. (1998). *The Foundations of Social Research. Meaning and Perspective in the Research Process*. Thousand Oaks, CA: Sage.
- Erickson, F. (1986/1990). Qualitative methods in research on teaching. In Linn, R. L. and Erickson, F. (eds.) *Quantitative Methods. Qualitative Methods*, pp 77–194. New York: Macmillan.
- Garrick, J. (1999). Doubting the philosophical assumptions of interpretive research. *Qualitative Studies in Education* 12(2), 147–156.
- Geertz, C. (1973). *The Interpretation of Cultures*. New York: Basic Books.
- Geertz, C. (1983). *Local Knowledge: Further Essays on Interpretive Anthropology*. New York: Basic Books.
- Greene, J. C. (1992). The practitioner's perspective. *Curriculum Inquiry* 22(1), 39–45.
- Howe, K. R. (1998). The interpretive turn and the new debate in education. *Educational Researcher* 27(8), 13–20.
- Kamberelis, G. and Dimitriadis, G. (2005). *On Qualitative Inquiry: Approaches to Language and Literacy Research*. New York: Teachers College Press.
- Rabinow, P. (1986). Representations are social facts: Modernity and post-modernity in anthropology. In Clifford, J. and Marcus, G. E. (eds.) *Writing Culture: The Poetics and Politics of Ethnography* pp 234–261. Berkeley, CA: University of California Press.
- Rabinow, P. and Sullivan, W. M. (eds.) (1987). *Interpretive Social Science: A Second Look*. Berkeley, CA: University of California Press.
- Schwandt, T. A. (1994). Constructivist, interpretivist approaches to human inquiry. In Denzin, N. K. and Lincoln, Y. S. (eds.) *Handbook of Qualitative Research*, pp 118–137. Thousand Oaks, CA: Sage.
- Taylor, C. (1979). Interpretation and the sciences of man. In Rabinow, P. and Sullivan, W. M. (eds.) *Interpretive Social Science*, pp 25–72. Berkeley, CA: University of California Press.

Interviews and Interviewing*

A Marvasti, Penn State Altoona, Altoona, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Active interview – A type of unstructured approach to data collection and analysis that treats the interview as a social occasion for reality production.

Constructionism – A philosophical orientation that focuses on how human practices create social reality.

Ethnographic interview – A variant of unstructured interviews that is particularly sensitive to the social context of the respondents' lives.

Focus group – Interviewing a group of respondents at the same time with the goal of stimulating discussion.

In-depth interview – A variant of unstructured interviews with special emphasis on revealing the respondents' inner feelings and attitudes.

Interview – A technique of collecting information by asking questions and receiving answers through face-to-face, phone, or Internet interaction.

Positivism – A theoretical model for conducting social research modeled after the natural sciences and based on a search for universal laws and strictly empirical evidence.

Postmodernism – An interdisciplinary movement that is based on the relentless critique of established social conventions.

Structured interview – An interviewing technique that relies on closed-ended questions and strict procedures for collecting data.

Unstructured interview – An interviewing technique based on open-ended questions that encourage respondents to elaborate on their statements.

Interviewing is a basic form of data collection based on a question-answer format. However, one does not need to be a social scientist to have participated in this mode of communication. Teachers, for example, conduct informal interviews with students who have missed assignments, asking them why they missed the assignment before deciding whether to allow the student to complete the work. Similarly, in a physician's office, patients participate in intake interviews where they answer specific questions about their medical history. Job applicants typically undergo a face-to-face interview before they are selected

for a position. Television and popular magazines entertain their audiences with celebrity interviews.

The setting and the purpose of the interview aside, the format is fairly consistent: questions are posed to individuals and they are expected to provide meaningful responses. Arguably, information gathering through interviewing is so ubiquitous that it typically goes unrecognized. Given the prevalence of interviews as a mode of communication in all facets of social life, we may refer to our social world as an interview society (Atkinson and Silverman, 1997). Yet, from a historical standpoint, this seemingly natural mode of inquiry is a relative newcomer in the world of collecting data about human behavior (Gubrium and Holstein, 2002). That is to say, the interview is based on certain assumptions and role expectations, which in turn inform how the interview is conducted and for what purpose. Let us briefly consider these taken-for-granted assumptions and how they relate to methodologies of qualitative and quantitative research.

Conceptualizing the Interview

Let us begin by noting that the so-called quantitative and qualitative research methodologies are two extremes on a continuum of research with much variation in between. To the extent that there are real differences between these two research methods, the distinctions have more to do with theories of the social world and reality than research techniques *per se*. Specifically, qualitative researchers tend to be interested in: (1) the fluidity of social life, (2) individual and situational variations of human experience, and (3) the deliberate or subjective work of conveying human experience. These points of emphasis collectively may be referred to as constructionism (an orientation aimed at understanding variable constructions of knowledge). By comparison, quantitative researchers emphasize a static cause-effect model of social reality and accordingly seek: (1) explanations of human behavior that are universally constant and (2) representations of human experience that are objective, true, or free from subjective distortions or biases. The philosophical paradigm informing quantitative research may be termed positivism (an orientation inspired by the natural sciences and committed to discovering positive, conclusive knowledge about the social world). Each standpoint then requires certain techniques of discovery that complement its theoretical view. For example, the twin sciences of survey

* Adapted from Marvasti, A. (2003). *Qualitative Research in Sociology*. Thousand Oaks, CA: Sage.

research and statistics allow quantitative investigators to summarize opinions and social trends for large aggregates. Such findings are then shared with readers in the form of numerical tables, or static representation of social reality. By comparison, qualitative researchers use techniques that show the depth and fluidity of experience as their variations across social settings. Qualitative data often are represented in the form of narratives, interviews, or field notes.

The interview method is used by both qualitative and quantitative researchers. While it is possible to distinguish the usage in purely technical terms (structured interviews for quantitative and unstructured for qualitative), the differences imply deeper understandings of research participants and their relationship to the investigators. Before delving into a discussion of research techniques, in the next section, we consider how the two camps might conceptualize the interview as a process of communication between the researcher and research participants.

Structured Interview: Extraction

From a quantitative standpoint, the research interview is primarily about extracting information from a subject. What matters first and foremost is that the information remains pure, or free from potential distortions or diversions. The data-collection process and its techniques are therefore designed to reduce biases that could originate from the data collection instrument, the investigator, or the participants. This type of objectivity is achieved through a structured research design (e.g., random samples and closed-ended questions) that provides consistency, uniformity, and structure to researcher–subject interactions while the information is being extracted.

Gubrium and Holstein (2002) argue that this particular view of the research interview is founded on three premises. First, this interview format assumes that human beings share a common experience, which can be articulated by any random member of society. The authors refer to this assumption as “the democratization of opinions” (p. 4), which implies every person’s opinion of the world is valid and that the statistical summary of these views paints a reasonably complete picture of social reality. According to Gubrium and Holstein, the second assumption of conventional interviews is the division between the formalized roles of the researcher and the respondent. In a positivistically-oriented interview, these roles and expectations resemble a leader–follower relationship. The interviewer is the leader; he or she asks questions and in doing so decides the topic, the pace, and the relevance of what will be discussed. The respondent’s primary responsibility in turn is to provide coherent, presumably truthful, answers when prompted to do so. The third premise of a positivistic interview format is that respondents are essentially vessel of answers (Gubrium and Holstein, 2002), or

proverbial fountains of information that could be turned on by the right questions. In this positivistic framework, the subjects’ involvement in the interview is limited to only answering questions and that researcher’s primary job is to extract answers. Positivistic researchers treat the context and the nature of the researcher–respondent interaction as a sort of background noise that interferes with true findings. Such interferences are controlled or manipulated to improve the objectivity of the data. As David Silverman points out,

For positivists, an observation that interview responses might be an outcome of the interview setting would be heard as a charge against the reliability of the technique. To the extent that this possibility arises, checks and remedies are built into the research design. Similarly, for positivists, the language of the interviewee serves primarily as an instrument for the communication of social or psychological facts. (Silverman, 2001: 88)

As a whole, these assumptions regarding democratization of opinions, formalized roles, and respondents as vessels of answers help shape the structured interview and its related data-collection techniques. Let us now consider a qualitative orientation, and theory informs the interviews and interviewing.

Unstructured Interview: Interaction

For most qualitative researchers, the traditional model of structured interviewing lacks the analytical tools for understanding how the social context and the interactional dynamics of the interview shape the research findings. From a qualitative standpoint (particularly derived from a constructionist view), research participants cannot and should not be treated as mere conduits or vessels of information to be used and disposed of at the interviewers’ whim. This emerging sensitivity to social context and the interactional nature of interviewing is best captured by the following edict: “Do not assume that the subject behind the respondent is merely there for the asking” (Dunbar *et al.*, 2002: 295). Thus, for a growing number of sociologists and other social scientists, the interview process is no longer limited to the simple give and take of asking and answering questions. While gathering information about people remains a central purpose of interviewing, exceedingly, qualitative researchers try to couple technical and procedural matters with matters of meaning making, social interaction, and social context.

The most notable attempt to incorporate these concerns into a unified approach can be found in Holstein and Gubrium’s (1995, 1997) active interviewing. According to this perspective, the interview is a social occasion, or an event, in its own right whereby researchers and respondents jointly create social reality through interaction. As Holstein and Gubrium (1997) put it, “From this

perspective, interview participants are practitioners of everyday life, constantly working to discern and communicate the recognizable and the orderly features of experience" (p. 121).

Viewed from an active-interview orientation, the elaborate procedural guidelines of survey interviews take on a different meaning. Instead of producing less-biased, more-objective data, the structured interview and its rigid rules can be seen as creating simply another version of truth, one that reflects the interviewer's assumptions (e.g., what topics are relevant, what is bias, and what is a proper answer), as much as it tells us about real experiences or attitudes. A positivistic perspective on the interview suggests that the subjects' knowledge is a thing that can be extracted from them, but from an interactional or constructionist view, interviews and interviewing are relative to time and place. Instead of seeing the interview as a social vacuum where the general principles of human interaction are suspended in favor of truth finding, qualitative researchers approach the interview as a social encounter with its own set of constructive practices and interactional dynamics.

Having said that, it should be noted that the open-ended, unstructured format of qualitative interviewing does not mean anything goes. On the contrary, this alternative format requires greater attention and analysis of a wide range of issues that tend to be ignored by the more standardized interview format. Among other things, unstructured interviewing demands that researchers be more reflective and deliberate about the reality they are jointly constructing with their participants. As Gubrium and Holstein (1997) put it, seeing the interview as an active interaction means: "All participants in an interview are inevitably implicated in making meaning" (p. 126).

Given these broad theoretical distinctions, let us now consider some interviewing models often associated with qualitative research.

Range of Qualitative Interviews

As the name implies, unstructured interviews are less stringent about the assumptions of interviewing. Also referred to as open-ended interviews, they allow more fluid interaction between the researcher and the respondent. In most qualitative interviews, respondents are not required to choose from a predesigned range of answers; instead, they can elaborate on their statements and connect them with other matters of relevance. In fact, in some published manuscripts, this data-collection procedure is simply referred to as talking, signifying its informal and conversational style. The following is an example of an open-ended interview with a nursing home resident.

Jay: Everybody has a life story. Why don't you tell me a little about your life?

Rita: Well there's not much. I worked in a telephone

company as a telephone operator before I was married. After I got married I moved to New Jersey and had two boys . . . (Gubrium, 1993: 20)

As seen in this example, unstructured interviewers are procedural minimalists: they simply provide a general sense of direction and allow respondents to tell their stories.

Two variants of the unstructured format are in-depth and ethnographic interviews. It is important to note, however, that the terms unstructured, in-depth, and ethnographic interviewing are sometimes used interchangeably in social science literature. Both in theory and in practice these orientations overlap. Furthermore, every study could and often does either create its own version of these techniques or uses them in combination. With these qualifications in mind, the following section offers an overview of these two interviewing techniques.

In-Depth Interviews: Manifesting the Inner Self

In-depth interviewing, as the name suggests, is founded on the notion that delving into the subject's deeper self produces more authentic data. John M. Johnson (2002) suggests that in-depth interviews are based on a number of assumptions. First, understanding the deeper self in this context means seeing the world from the respondent's point of view, or gaining an empathic appreciation of his or her world. In-depth interviewers aim to gain access to the hidden perceptions of their subjects, or as Johnson puts it,

[In-depth interviewing] begins with commonsense perceptions, explanations, and understandings of some lived cultural experience . . . and aims to explore the contextual boundaries of that experience or perception, to uncover what is usually hidden from ordinary view or reflection or to penetrate to more reflective understandings about the nature of that experience. (p. 106)

Another common assumption of in-depth interviewing is that it can and should be mutually beneficial to the subject and the researcher. That is, in addition to helping the subject uncover suppressed feelings through the interview process, the researcher also gains knowledge of his or her own hidden or conflicting emotions (p. 106). Lastly, according to Johnson (2002), in-depth interviewing provides a multi-perspective understanding of the topic. In other words, by not limiting respondents to a fixed set of answers, in-depth interviewing has the potential to reveal multiple, and sometimes conflicting, attitudes about a given topic.

In sum, the procedural guidelines of in-depth interviewing encourage mutual self-disclosure in the context of an emotionally charged atmosphere where the interviewer and interviewee freely express their views about an issue (Douglas, 1985). The questions are designed to go beyond the presumed surface level of respondents' feelings and into the deeper levels of their consciousness. That is to say, the

inquiries are directed at the unseen or the hidden dimensions of the self. Not surprisingly, all this gives this particular brand of in-depth interviewing the quality of talk therapy. Its procedures are reminiscent of Freudian psychoanalytic techniques aimed at uncovering the subconscious through free association, or random expressions of thoughts.

Ethnographic Interviews: Recording Life in Context

Ethnographic research has its roots in the discipline of anthropology where interviews tend to be conducted in everyday settings and greater attention is given to the context of social life. Indeed, early ethnographies were essentially anthropological travelogs that provided accounts of exotic people and their cultures (Tedlock, 2000). Its ethnocentric flaws notwithstanding, anthropology's emphasis on culture has greatly contributed to the development of ethnographic research, which is regarded as uniquely suited for connecting empirical observations with the contingencies of the setting.

In particular, ethnographic interviewing differs from other approaches in that it takes place in, or is related to a particular physical setting, sometimes referred to as the field. The ethnographic field is the social context that guides the interview in terms of what questions are asked, which people are interviewed, and how their answers are interpreted. Ethnographic researchers typically rely on informants for assistance in navigating the field. In addition, ethnographic interviewers use observations from the field to assess the meaning and relevance of their interview data.

In the course of conducting ethnographic interviews, greater attention is given to where, when, and with whom the interviews are conducted. The ethnographic research enterprise, which begins with selecting a research site and gaining access and rapport, uses interviewing to supplement field observations and other sources of data (e.g., official documents). Therefore, the interview questions, its timing and place, and the choice of respondents depend on what the ethnographer needs to know at a given stage of the research. For example, an ethnography of illicit drug culture in an urban high school might begin by asking students about illicit drug use but could gradually shift to their parents' and teachers' attitudes about the topic. As the field work progresses, the researcher may have to develop new questions and recruit a wider range of participants than originally planned.

In essence, the ethnographic interview is typically unstructured like most other qualitative interviews. The main difference is that the researcher, or ethnographer, typically has greater rapport and understanding of the social context of the respondents' lives. Furthermore, the interview itself is more likely to be conducted in the field, or places where the respondents actually live or work.

Focus Group Interviews: Encouraging Group Interaction

In focus groups, the researcher asks questions from a number of respondents at the same time to "stimulate discussion and thereby understand (through further analysis) the meanings and norms which underlie those group answers" (Bloor, *et al.*, 2001: 43). According to Fontana and Frey (2002: 651), historically, focus groups owe much of their popularity to marketing researchers and political candidates who wanted to gauge the opinions of their consumers or constituents about a particular product or issue (see also Bloor *et al.*, 2001: 1–3). From there, focus groups gradually made headway into the social sciences and now occupy a well-respected position among the various data-collection methods.

The format ranges from very structured with respondents taking turns answering each and every question to a more flexible brainstorming session where participants voice their opinions at will. The interactional nature of focus groups can stimulate respondents' memory of specific events and facts (Fontana and Frey, 2002: 651). As a whole, as Rubin and Rubin state,

In focus groups, the goal is to let people spark off one another, suggesting dimensions and nuances of the original problem that any one individual might not have thought of. Sometimes a totally different understanding of a problem emerges from the group discussion. (Rubin and Rubin (1995: 140), quoted in Berg (2001: 115))

Other advantages are that focus groups can be very stimulating to the respondents (they will not become bored), and provide participants with the opportunity to elaborate on each other's answers (Fontana and Frey, 2002: 252). However, the approach is not without its problems. Some of the challenges of focus group interviewing include:

1. one person could dominate the group;
2. respondents may be reluctant to discuss sensitive topics in the presence of others, or they could distort their answers in an effort to appear socially desirable;
3. some individuals may be shy and thus require more encouragement to participate; and
4. the interviewer has to be skilled at managing the group dynamics and asking questions simultaneously (Merton *et al.* (1956), cited in Fontana and Frey (2002: 652)).

Respondent Characteristics

Interviewers are becoming more sensitive to how the respondents' characteristics influence their data-collection procedures and outcomes. An important consideration for researchers in the area of education is how to deal with

obvious inequality of power and status between researchers and their underage respondents. As Eder and Fingerson (2002) note, interviewers who work with children and adolescents have to take special care to minimize the risk of exploitation. The recommended solution is to make certain that the young respondents' participation in the research is of some value to them and their communities. As Eder and Fingerson state,

The researcher's desire to gain information from child participants without giving something in return reflects an underlying sense of the adult researcher's privilege. However, by giving something in return for receiving information, researchers can reduce the potential for power inequality. (p. 185)

According to the authors, this can be done in two ways. First, the interview questions and procedures can be designed to help children and adolescents learn something about themselves or become empowered. For example, in a study of young adolescent girls, respondents reported higher levels of self-awareness and enhanced communication skills as a direct result of participating in interviews. As one subject reported,

But since the question came up, it let me know how I felt. I think that's good. I can do this forever you know . . . keep on going . I'll bring a lot up with just easy questions that you would ask anybody, you know. It lets you know about yourself. (Taylor *et al.* (1995: 129), quoted in Eder and Fingerson (2002: 186))

Another way in which social scientists can give something back, or reciprocate, is through action-oriented research, or research that produces concrete positive changes in the respondents' lives and their community. For example, Eder and Fingerson note how Valenzuela's (1999) work with Mexican-American adolescents in their classrooms helped reduce tension and misunderstanding between them and their teacher.

Other Considerations

This article began with a comparison of the traditional understanding of an interview as a research tool based on questions and answers. We gradually moved to the more analytically complex idea that an interview is a social occasion that creates a particular version of social reality. Interviewing, a once simple and seemingly innocuous data-gathering method, is being critiqued and expanded into new territories. According to some commentators, this is due to the influence of postmodernism (Fontana, 2002), which for our purposes can be defined as an interdisciplinary movement that is relentlessly critical of established social conventions. Among the various questions raised by postmodernism are: How could the traditional

interview model be transformed into something more liberating and empowering for the respondents? Who owns the text and the stories that emerge from an interview? Is it the researcher's story to write as he or she wishes or is it the respondents' version? (Gubrium and Holstein, 2002: 16–17). These are just a few of the many concerns that postmodernists have brought to the realm of social science research.

As is the case with many interests in sociology, these debates will likely continue well into the next century. On a more practical level, it is important to keep in mind that one's choice of interview technique should be in sync with the topic of one's interest and the questions one wishes to answer. If one is interested in the relationship between income and grade point average among college students, a survey interview will do the job efficiently and inexpensively. If, on the other hand, one wants to study how the privileges of social class shape a person's identity and worldview, he or she might proceed differently. For example, one might ask low- and high-income people an open-ended question such as, "What is the most important thing you want people to know about you?" and compare their answers; or one could supplement structured interviews from a large sample with unstructured in-depth interviews from a smaller group of respondents. The important point is not to impose arbitrary limitations on one's curiosity about the social world because of methodological orthodoxies.

Bibliography

- Atkinson, P. and Silverman, D. (1997). Kundera's immorality: The interview society and the invention of self. *Qualitative Inquiry* 3(3), 324–345.
- Berg, B. L. (2001). *Qualitative Research Methods for the Social Sciences*, 4th edn. Boston, MA: Allyn and Bacon.
- Bloor, M. J., Frankland, M. T., and Robson, K. (2001). *Focus Groups in Social Research*. London: Sage.
- Douglas, J. D. (1985). *Creative Interviewing*. Beverly Hills, CA: Sage.
- Dunbar, C. Jr., Rodriguez, D., and Parker, L. (2002). Race, subjectivity, and the interview process. In Gubrium, J. and Holstein, J. (eds.) *Handbook of Interview Research: Context and Method*, pp 279–298. Thousand Oaks, CA: Sage.
- Eder, D. and Laura, F. (2002). Interviewing children and adolescents. In Gubrium, J. and Holstein, J. (eds.) *Handbook of Interview Research: Context and Method*, pp 181–201. Thousand Oaks, CA: Sage.
- Fontana, A. (2002). Postmodern trends in interviewing. In Gubrium, J. and Holstein, J. (eds.) *Handbook of Interview Research: Context and Method*, pp 161–180. Thousand Oaks, CA: Sage.
- Gubrium, J. and Holstein, J. (2002). From the individual interview to interview society. In Gubrium, J. and Holstein, J. (eds.) *Handbook of Interview Research: Context and Method*, pp 3–32. Thousand Oaks, CA: Sage.
- Holstein, J. and Gubrium, J. (1995). *The Active Interview*. Thousand Oaks, CA: Sage.
- Holstein, J. and Gubrium, J. (1997). Active interviewing. In Silverman, D. (ed.) *Qualitative Research: Theory, Method, and Practice*, pp 113–129. London: Sage.
- Johnson, J. M. (2002). In-depth interviewing. In Gubrium, J. and Holstein, J. (eds.) *Handbook of Interview Research: Context and Method*, pp 103–119. Thousand Oaks, CA: Sage.

- Marvasti, A. (2003). *Qualitative Research in Sociology*. Thousand Oaks, CA: Sage.
- Merton, R. K., Fiske, M., and Kendall, P. L. (1956). *The Focused Interview*. Glencoe, IL: Free Press.
- Rubin, H. J. and Rubin, I. S. (1995). *Qualitative Interviewing: The Art of Hearing Data*. Thousand Oaks, CA: Sage.
- Silverman, D. (2001). *Interpreting Qualitative Data: Methods for Analysing Talk, Text and Interaction*, 2nd edn. London: Sage.
- Taylor, J. M., Gilligan, C., and Sullivan, A. M. (1995). *Between Voice and Silence: Women and Girls, Race and Relationship*. Cambridge, MA: Harvard University Press.
- Tedlock, B. (2000). Ethnography and ethnographic representation. In Denzin, N. and Lincoln, Y. (eds.) *Handbook of Qualitative Research*, 2nd edn., pp 455–486. Thousand Oaks, CA: Sage.
- Valenzuela, A. (1999). *Subtractive Schooling: U.S.-Mexican Youth and the Politics of Caring*. Albany, NY: State University of New York Press.
- Handbook of Qualitative Research*, 2nd edn., pp 645–672. Thousand Oaks, CA: Sage.
- Hammersley, M. and Atkinson, P. (1990). *Ethnography: Principles in Practice*. London: Routledge.
- Litchman, M. (2006). *Qualitative Research in Education: A User's Guide*. Thousand Oaks, CA: Sage.
- Mann, C. and Stewart, F. (2002). Internet interviewing. In Gubrium, J. and Holstein, J. (eds.) *Handbook of Interview Research: Context and Method*, pp 603–628. Thousand Oaks, CA: Sage.
- Morgan, D. L. (1988). *Focus Groups as Qualitative Research*. Newbury Park, CA: Sage.
- Silverman, D. (1997). *Qualitative Research: Theory, Method and Practice*. London: Sage.
- Tierney, W. and Dilley, P. (2002). Interviewing in education. In Gubrium, J. and Holstein, J. (eds.) *Handbook of Interview Research: Context and Method*, pp 453–471. Thousand Oaks, CA: Sage.

Further Reading

- Fontana, A. and Frey, J. (2000). The interview: From structured questions to negotiated text. In Denzin, N. and Lincoln, Y. (eds.)

Life History

J A Hatch and S K Newsom, University of Tennessee, Knoxville, TN, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Life history – A mode of qualitative research that seeks to apprehend, understand, and render individual life stories within their personal, social, economic, political, and historical contexts.

Life history is a mode of qualitative research that seeks to apprehend, understand, and render individual life stories within their personal, social, economic, political, and historical contexts. Its emphasis on theorizing individual biographies in relation to the pervasive influences of the contexts that surround them is the hallmark of this kind of social research. Life historians use interviews, observations, artifacts, and other qualitative data to generate life stories from individual participants; but they go beyond narrative accounts of individual lives and undertake in-depth analyses of the collective contextual influences in which those lives are situated. Life history has been called “sociologically read biography” (Bertaux, 1981), “life story located within its historical context” (Goodson, 1992), “genealogies of context,” and “stories of action within theories of context” (Goodson, 1995). Each of these expressions emphasizes the place of both story and context. If both are not theorized in relation to the other, the product is something other than a life history.

This article describes the unique contributions of life-history research; provides a brief history of life history, tracing its roots to a variety of disciplines; compares life history to other qualitative approaches, narrative research in particular; describes data collection and analysis possibilities for life-history work; and discusses issues related to doing life-history research.

Unique Contributions

Life-history research is positioned to make unique contributions to the knowledge base in education and the social sciences because of its focus on relationships between individual subjective experience and the complex contexts in which that experience is embedded. Life-history research illuminates both the logic of individual decision making and the impact of societal-level conditions within which decisions are made. It provides ways to

study connections among individual lives, culture, and social structure that are not possible using other research methodologies – quantitative or qualitative. Life-history researchers are committed to examining perspectives on culture and social structure that are not available to those applying a wider methodological lens. By making individual lives the centerpiece of the research enterprise, life-history researchers generate findings that reveal the lived experiences of social actors. These findings make possible a dialog between abstract theoretical and empirical descriptions of cultural and social phenomena and the subjective encounters of real people with the real-world expressions of those phenomena.

By taking the perspectives of individual participants as a point of departure, life-history studies disrupt the assumptions of traditional social science research. Life-history researchers’ concern with the perceptions of individual social actors elevates subjectivity to the status of a methodological necessity, rather than a flaw to be controlled at all costs (Cole and Knowles, 2001). Further, the emphasis on collecting rich data with a few informants runs counter to newly revived positivist contentions that such strategies are not scientifically based. Life-history projects are valuable because they emphasize detailed contextualized analyses of the subjective experiences of selected individuals. It is assumed that the perspectives of these individuals can supply different, perhaps better, knowledge of prevailing social conditions than the detached observations of traditional social scientists (Casey, 1992).

Life-history research is unique because it allows for systematic inquiry into the human condition through a close examination of living in the world as experienced by individuals. The logic of this kind of research is based on the maxim that the general can best be understood through analyses of the particular. For life-history researchers, to understand the intricacies of one life is to gain insights into the collective. That is not to say that to understand one is to understand all; but understanding one life helps to shed light on the complexities of lives in communities (Cole and Knowles, 2001). Life-history researchers capture stories of lives set within social, political, socioeconomic, and historical contexts. These stories contribute special kinds of knowledge that enriches understandings of individual and collective experience.

Life-history research provides a forum that allows authentic voices to speak for themselves. In this kind of research, the voices of participants are of central importance to the stories being told – the more authentic the

voices, the more powerful the stories. Life-history research typically asks questions not from the point of view of powerful actors, but from the perspective of those being acted upon (Becker, 1970). The goal is to elicit expressions of what people living in particular social worlds already know about those worlds. They are the experts at understanding the meaning structures and discourses that define their experiences, and life-history research seeks to utilize their expert voices to help educate the rest of us. Individuals from all groups have the right to speak for themselves and have their voices accepted as legitimate (Gilligan, 1982/1993). Life-history work depends on the authentic voices of individuals to make research findings meaningful.

History of Life-History Research

The foundations of life-history research can be traced within the histories of a number of scholarly disciplines, including anthropology, sociology, psychology, and education. The popularity and acceptance of life-history approaches in these and other social science fields of inquiry have waxed and waned since the early twentieth century. The period between 1920 and World War II is recognized as a period of early substantial interest in the study of life histories (Runyan, 1982). Anthropologists who collected the life stories of Native Americans were among the first to produce life-history research. Sociologists from the University of Chicago, who viewed individuals and their stories as living illustrations of the social conditions of the times, powerfully influenced the academic landscape of the 1920s and 1930s. Thomas and Znaniecki's (1920) *The Polish Peasant in Europe and America* is widely acknowledged by anthropologists and sociologists as the most influential early publication of life-history work. Psychologists of the same period used idiographic methods to study individuals and their stories in order to understand personality development and psychological maladjustments.

From World War II through the mid-1960s, a significant decline in life-history research was seen (Runyan, 1982). This decline was evident across the social sciences, and scholars of the day attributed it to the ascendancy of abstract theory, experimental research designs, and quantitative data analyses across the disciplines (Bertaux, 1981). While the study of life histories did not completely stop during this period, such work was less highly valued, less widely recognized, and less well funded than research based on testing hypothetical relationships among measurable variables.

A resurgence in interest in studying life histories occurred in the late 1960s, leading social scientists to explore areas such as psychopathology, personality development, life-span development, aging, social mobility, and socialization (Runyan, 1982). Life-history work of this

period emphasized exploring how the life course is influenced by social, demographic, and historical conditions. During the 1980s and 1990s, many feminist researchers and those who research aspects of sexuality began to use life-history methods in their studies, in large part because of the ways life histories are able to give expression to hidden or silenced lives. In the 1980s, educational researchers, many of them from the United Kingdom and Canada, began to apply life-history methods to the study of teachers' lives. Throughout the 1980s and 1990s, a small but influential cohort of scholars in the field of education continued to produce life-history accounts of administrators, teachers, and pupils, along with generating important theoretical and methodological texts (Goodson, 2001).

From the late 1980s onward, the influence of postmodern thought had an impact on life-history research across the academic disciplines. On the one hand, the postmodern challenge to the master narrative of objective science made the inherently subjective nature of life history an attractive alternative. On the other, the postmodern notion of plural, fragmented, nonlinear selves created for life historians a crisis of representation – how could researchers ever capture the complexities of the lives they were studying? (Hatch and Wisniewski, 1995). Early in the twenty-first century, a return to modernist definitions of what constitutes science has occurred in the United States and other nations. Based on positivist models used in medical studies, the federal government in the United States has stipulated what is considered to be scientifically based research. As life-history research is not based on experimental methods with large numbers of subjects, double-blind treatment designs, and control groups, its future is under challenge.

Relationships with Other Qualitative Approaches

Life-history research is one of several qualitative research approaches that fit under the larger category of narrative research. Along with life history, narrative research includes such approaches as life-story research, oral history, biography, personal experience methods, and narrative inquiry. As a narrative approach, life history shares the following characteristics that distinguish narrative from other forms of qualitative inquiry:

- Focus on individual stories – understanding individual lives through individual stories is central to the processes and products of life history and narrative research.
- Personal nature of research processes – researchers and participants must work closely together to come to a shared understanding of the participants' stories.

- Practical orientation – because of their goal of capturing real lives as lived, research outcomes make it possible to connect understandings to the everyday world.
- Emphasis on subjectivity – more than other qualitative methodologies, life history and narrative go beyond scientific and empiricist standards, relying on the authentic voices of participants to generate confidence in research findings (Hatch and Wisniewski, 1995).

These characteristics mean that life history and narrative approaches fit within some qualitative research paradigms and not others. Because of their emphasis on subjectivity and co-constructed understandings, life history and narrative do not align with characteristics of the postpositivist research paradigm, which assumes that approximations of reality can be discovered through rigorous data collection and analysis procedures. There is a better fit for life history and narrative within constructivist, critical/feminist, or post-structuralist paradigms, which assume different relationships between researchers and participants, different connections between sociopolitical positionings and research, and different possibilities for representing the complexities of postmodern life. Excellent life-history and narrative research has been done within the assumptions of each of the constructivist, critical/feminist, and post-structuralist paradigms (Hatch, 2002).

Life history is a kind of narrative research, but it is distinct from other narrative approaches. All life histories are narratives; but not all narratives are life histories. All narrative research is focused on collecting individual stories, but what distinguishes life-history work is its broad purpose. Life-history research goes beyond personal accounts and individual interpretations and examines those accounts and interpretations within an array of social, historical, and cultural contexts. While other narrative approaches seek to make meaning of individual experiences, life-history work draws on individual experiences to make sense of broader social phenomena (Cole and Knowles, 2001).

Data Collection and Analysis

Data collection in life-history studies utilizes many of the techniques of anthropological research, including informant interviewing, participant observation, and document and artifact collection. As data need to be collected that allow for individual life stories to be told in social–historical context, data gathering methods that get at both personal biographies and contextual influences need to be employed. Overlap exists between the data-collection techniques that are used to capture the personal and those used to describe the contextual. In the following sections, data for life-history projects are organized into three general types: researcher-generated materials, participant-generated

materials, and other-generated materials (Cole and Knowles, 2001). Data of each type can be used to generate life stories and understand contextual influences.

Researcher-generated materials include data such as interview transcripts, observation field notes, and videotapes or photographs taken by the researcher. Life-history interviews are characterized by reciprocal relations between interviewers and interviewees. Interviewers ask their respondents to tell the stories of their lives and interpret the historical, social, and political contexts in which the stories are embedded. Often, data are collected through a series of open-ended, informal interviews that resemble conversations between friends more than structured interviews between researcher and subject. Later interviews build on previous conversations and other data collection to build contextualized life-history accounts. Interviews are almost always audiotaped and transcribed, and life historians sometimes interview people who are acquainted with or know the contexts surrounding the individuals they are studying as a way to enrich the data of their studies.

Life-history researchers often study lives in context by acting as participant observers in the settings experienced by their informants. By directly experiencing the life circumstances of those being studied, life-history researchers are better able to understand and relate the influence of setting on individual experience. Some researchers shadow their informants in order to make a record of events that make up their daily experience. Participant observers typically take copious notes during the time they are in the settings being observed, then they fill in their notes with details soon after leaving the research scene. Researchers sometimes make video or photographic records as part of their data collection.

Participant-generated materials are a second general type of life-history data. These data include materials such as reflective journals and diaries; time lines; photographs and family albums; and letters, e-mails, and other work-related or personal documents. Journals and diaries can be those that study participants have been keeping on their own, or it might be part of the research design to ask informants to keep reflective records of their experiences. Some researchers (e.g., Goodson, 2001) have participants construct time lines of key events in their lives and use the time lines as a device for facilitating life-history interviews. Photographs that participants have collected or which they have taken as part of the research process can be important data in life-history studies. In addition, participant-generated letters, memos, e-mails, and other documents can be valuable sources of information when trying to construct a contextualized story of a life.

Other-generated materials include documents and artifacts that have a bearing on the individuals or contexts of the study, but are not produced by the researchers or their participants. These data include institutional

records (e.g., memos, reports, agendas, minutes, and press releases), artifacts (e.g., physical objects such as tools or toys that represent something important in the life or context under examination), and historical records (e.g., archives or newspaper records that shed light on individual experience or social contexts). Collecting such data can help fill in spaces in time, trigger memories in participants, and bring coherence to the creation of a life-history report.

Data analysis in life-history research parallels data analysis done in other qualitative research approaches. While it is understood that there is an artistic, intuitive dimension to making sense of the large amounts of data that are typically collected, life historians are careful to insure that their interpretations are grounded in empirical evidence. Like many other qualitative researchers, life-history workers are reluctant to assign formulaic methods to their data analyses. They are more likely to note that each project and data set will lead to different data analysis processes, applying different analytic frameworks. However, no matter what the framework, life-history analysis is focused on making sense of the information the researcher has decided to count as data, including information about the life stories of individuals and the contexts in which the life stories are contained. While many life historians are against rigid analytic schemes, most would agree that data analysis must be based on “systematic and disciplined attention” (Cole and Knowles, 2001: 101).

Wolcott’s (1994) well-known classification of data transformation into description, analysis, and interpretation has been used to distinguish different dimensions in life-history data analysis (Cole and Knowles, 2001). For example, at the descriptive level, researchers may search for data excerpts that allow participants’ direct accounts to tell their own stories. At the analysis level, researchers might look for potential patterns, categories, themes, models, or typologies, then systematically return to the data to confirm or disconfirm their existence. At the interpretive level, researchers construct artfully written life histories, using their data to bring together the stories of individual lives in relation to relevant socio-historical contexts and discipline-based theories. As Wolcott explained, all qualitative research has some elements of each dimension – it’s the proportion of each that changes within individual studies. So it is for life-history data analyses: some emphasize description, some analysis, and some interpretation.

Life-history scholars agree that researchers must be reflective and reflexive throughout the data collection and analysis process, self-consciously keeping track of their own place in generating and making sense of life-history data. Since a life history is always contextual and researchers are part of the context, it is the researchers’ obligation to keep track of their own place in what

happened and to share that information with their readers. Data analysis in life-history research requires researchers to monitor their place in the study’s descriptive, analytic, and interpretive processes. Life-history reports are enriched when researchers are explicit about their data collection and analysis practices and their reasons for using them, including a reflexive exposition of their positioning throughout the research.

Issues in Life-History Research

As with any methodology, scholars from inside and outside the community of scholars doing and writing about it have identified several issues related to life-history research. Some of the most prominent of these issues are discussed next. The nature of the issues chosen for discussion range from modernist concerns for the reliability, validity, and generalizability of life-history reports to postmodernist critiques of life historians’ epistemological and ontological positionings. Here, the effort is not to resolve the issues raised but to identify salient tensions in the field.

As life history is a qualitative research approach that relies on subjective interpretations of individual life stories, critics who operate within the assumptions of the positivist research paradigm argue that life-history reports lack sufficient rigor to be counted as scientifically based research. Such reports are thought to be based on less-than-valid data. These data take form in retrospective and introspective accounts that are perceived to be seriously limited by participants’ abilities to remember and their willingness to provide objective descriptions. The usefulness of life histories is automatically suspect if one assumes that it is unwise to generalize from single subjects to larger populations. Further, as in other qualitative approaches, life-history researchers are the principal data-collection instruments in the research process, and their perceived lack of objectivity is judged by some to be a significant limitation.

Life historians assume the primacy of subjectivity in their studies, but they also have concerns about the place of the author in the production of life-history texts. It is the researcher/author who finally puts the life history into text, and that person’s power in shaping how data are collected and interpreted and how stories are reported is an issue in life-history work. Some argue that the feigned invisibility of authors in telling life stories is a thin facade that distorts the reality of the research process (Fine, 1994). Others call for an open exposition of authorial positionings so that readers can decide for themselves the impact researchers may be having on the life stories being told (Tierney, 1998). Still others have challenged life historians’ ability to translate lives as lived into lives as text. They see the complexities involved in attempting to

represent others as coherent unitary selves as very difficult to overcome (Munro, 1998). Based on the thinking of Derrida (1972), Denzin (1989) explains, "there is no clear window into the inner life of a person, for any window is always filtered through the glaze of language, signs, and the processes of signification" (p. 29). The attending crisis of representation continues to be an issue in life-history research.

Another issue is tied to the risks involved for participants because of close relationships developed between themselves and researchers as life-history research is undertaken. Because of the necessity of close proximity and the advantages of sharing sensitive information, life-history work has been called "licensed voyeurism" (Mearor and Sykes, 1992: 209). Researchers ask their participants to share the intimate details of their lives, and the more intimate the information, the more risk is assumed by the sharer of that information. The obvious concern is that life-history takers will exploit their participants to serve the researchers' own ends.

Another layer of exploitation is identified by postmodern theorists as an additional issue in life-history research. Some scholars have argued that life-history work has the potential to disempower those who have been silenced historically and who appear to be given voice through participation in life-history projects. Tierney (1998) has characterized life-history work that objectifies participants as examples of particular groups as neocolonialist, essentially serving to reinscribe traditional notions of power. Denzin (1992) goes even farther, arguing that some researchers are agents of the modernist surveillance state, collecting life histories as information that keeps the myth of autonomous individualism alive.

Finally, some postmodern thinkers have challenged the metaphysical assumptions on which life-history work has been grounded. Cary (1999) points out that life history's reliance on contextualized and collaboratively constructed stories represents a turn toward a realist ontology that seeks truth through reliance on method. She argues that the need for rigor through contextualization and collaboration reaffirms realist assumptions and thereby contradicts postmodern understandings about what is knowable and ways to do research. Along with Lather (1996), Cary chides life historians for their failure to problematize the desire for totality that continues to frame their research.

Summary

Life history has a long history as an important research method in the social sciences. Its focus on contextualized life stories places life-history research in a unique position to provide important insights into the experience of

social phenomena through the eyes of individual social actors. Utilizing data collection and analysis methods adapted from other qualitative approaches, high-quality life-history work generates artistic portrayals of real-life experience that are grounded in carefully collected empirical evidence. Like other qualitative research methodologies early in the twenty-first century, life history faces challenges from both sides of the modernist/postmodernist divide, and its future is difficult to predict.

See also: Interpretive Research; Interviews and Interviewing; Narrative Inquiry; The Quality of Evidence in Qualitative Research.

Bibliography

- Becker, H. (1970). *Sociological Work: Method and Substance*. Chicago, IL: Aldine.
- Bertaux, D. (ed.) (1981). *Biography and Society: The Life History Approach in the Social Sciences*. Beverly Hills, CA: Sage.
- Cary, L. J. (1999). Unexpected stories: Life history and the limits of representation. *Qualitative Inquiry* 5, 411–427.
- Casey, K. (1992). Why do progressive women activists leave teaching? Theory, methodology and politics in life-history research. In Goodson, I. F. (ed.) *Studying Teachers' Lives*, pp 187–209. London: Routledge.
- Cole, A. L. and Knowles, J. G. (eds.) (2001). *Lives in Context: The Art of Life History Research*. Walnut Creek, CA: AltaMira Press.
- Denzin, N. K. (1992). Deconstructing the biographical method. *Paper Presented at the American Educational Research Annual Meeting*, San Francisco, CA, USA.
- Derrida, J. (1972). Structure, sign and play in the discourse of the human sciences. In Macksey, R. and Donato, E. (eds.) *The Structuralist Controversy: The Language of Criticism and the Stories of Man*, pp 247–265. Baltimore, MD: Johns Hopkins University Press.
- Fine, M. (1994). In working the hyphens: Reinventing self and other qualitative research. In Denzin, N. and Lincoln, Y. (eds.) *Handbook of Qualitative Research*, pp 70–82. London: Sage.
- Gilligan, C. (1982/1993). *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University Press.
- Goodson, I. F. (1992). Studying teachers' lives: An emergent field of inquiry. In Goodson, I. F. (ed.) *Studying Teachers' Lives*, pp 1–17. London: Routledge.
- Goodson, I. F. (1995). The story so far: Personal knowledge and the political. In Hatch, J. A. and Wisniewski, R. (eds.) *Life History and Narrative*, pp 89–98. London: RoutledgeFalmer.
- Goodson, I. F. (2001). *Life History Research in Educational Settings: Learning from Lives*. Buckingham, UK: Open University Press.
- Hatch, J. A. (2002). *Doing Qualitative Research in Education Settings*. Albany, NY: State University of New York Press.
- Hatch, J. A. and Wisniewski, R. (1995). *Life History and Narrative: Questions, Issues, and Exemplary Works*. In Hatch, J. A. and Wisniewski, R. (eds.) *Life History and Narrative*, pp 113–135. London: RoutledgeFalmer.
- Lather, P. (1996). Troubling clarity: The politics of accessible language. *Harvard Educational Review* 66, 525–545.
- Mearor, L. and Sykes, P. (1992). Visiting lives: Ethics and methodology in life history. In Goodson, I. F. (ed.) *Studying Teachers' Lives*, pp 209–233. London: Routledge.
- Munro, P. (1998). *Subject to Fiction: Women Teachers' Life History Narratives and the Cultural Politics of Resistance*. Buckingham: Open University Press.
- Runyan, W. M. (1982). *Life Histories and Psychobiography: Explorations in Theory and Method*. New York: Oxford University Press.

- Thomas, W. and Znaniecki, F. (1920). *The Polish Peasant in Europe and America*, 2nd edn. Chicago, IL: University of Chicago Press.
- Tierney, W. G. (1998). Life history's history: Subjects foretold. *Qualitative Inquiry* 4, 49–70.
- Wolcott, H. F. (1994). *Transforming Qualitative Data: Description, Analysis, and Interpretation*. Thousand Oaks, CA: Sage.

Further Reading

- Ball, S. J. and Goodson, I. F. (eds.) (1985). *Teachers' Lives and Careers*. London: Falmer.
- Casey, K. (1993). *I Answer with My Life: Life Histories of Women Teachers Working for Social Change*. New York: Routledge.
- Cohen, R. M. (1991). *A Lifetime of Teaching: Portraits of Five Veteran High School Teachers*. New York: Teachers College Press.
- Denzin, N. K. (1989). *Interpretive biography*. Newbury Park, CA: Sage.
- Dollard, J. (1935). *Criteria for the Life History, with Analysis of Six Notable Documents*. New Haven, CT: Yale University Press.
- Etter-Lewis, G. (1993). *My Soul is My Own: Oral Narratives of African American Women in the Professions*. New York: Routledge.
- Knowles, J. G. (1993). Life history accounts as mirrors: A practical avenue for the conceptualization of reflection in teacher education. In Calderhead, J. and Gates, P. (eds.) *Conceptualizing Reflection in Teacher Education*, pp 70–92. London: Falmer.
- Lawrence-Lightfoot, S. (1994). *I've Known Rivers: Lives of Loss and Liberation*. New York: Addison-Wesley.
- McLaughlin, D. and Tierney, W. G. (eds.) (1993). *Naming Silenced Lives: Personal Narratives and the Process of Educational Change*. New York: Routledge.
- Radin, P. (ed.) (1926). *Crashing Thunder: The Autobiography of an American Indian*. New York: D. Appleton.
- Schmidt, M. and Knowles, J. G. (1995). Four women's stories of failure as beginning teachers. *Teaching and Teacher Education* 11, 429–444.

Relevant Website

<http://www.sussex.ac.uk> – Centre for Life History and Life Writing Research.

Narrative Inquiry

D J Clandinin, University of Alberta, Edmonton, AB, Canada

J Huber, University of Regina, Regina, SK, Canada

© 2010 Elsevier Ltd. All rights reserved.

Emergence of Narrative Inquiry on the Research Landscape

Narrative inquiry is a ubiquitous practice in that,

Human beings have lived out and told stories about that living for as long as we could talk. And then we have talked about the stories we tell for almost as long. These lived and told stories and the talk about the stories are one of the ways that we fill our world with meaning and enlist one another's assistance in building lives and communities. What feels new is the emergence of narrative methodologies in the field of social science research

(Clandinin and Rosiek, 2007: 35).

Given this newness, it is important to define terms and how they are lived out in data collection and analysis within the emerging field of narrative inquiry. This is the focus of this article.

Terms and Definitions

Even though Reissman and Speedy (2007) point out that "narrative inquiry in the human sciences is a 20th century development; the field has 'realist', 'postmodern', and constructionist strands, and scholars and practitioners disagree on origin and precise definition" (p. 429), there is some agreement on the following definition:

People shape their daily lives by stories of who they and others are and as they interpret their past in terms of these stories. Story, in the current idiom, is a portal through which a person enters the world and by which their experience of the world is interpreted and made personally meaningful. Narrative inquiry, the study of experience as story, then, is first and foremost a way of thinking about experience. Narrative inquiry as a methodology entails a view of the phenomenon. To use narrative inquiry methodology is to adopt a particular view of experience as phenomenon under study

(Connelly and Clandinin, 2006: 375).

Commonplaces of Narrative Inquiry

Narrative inquiry is a way of understanding and inquiring into experience through "collaboration between researcher and participants, over time, in a place or series of places,

and in social interaction with milieus" (Clandinin and Connelly, 2000: 20). Three commonplaces of narrative inquiry – temporality, sociality, and place – specify dimensions of an inquiry and serve as a conceptual framework. Commonplaces are dimensions which need to be simultaneously explored in undertaking a narrative inquiry. Attending to experience through inquiry into all three commonplaces is, in part, what distinguishes narrative inquiry from other methodologies. Through attending to the commonplaces, narrative inquirers are able to study the complexity of the relational composition of people's lived experiences both inside and outside of an inquiry and, as well, imagine the future possibilities of these lives.

Temporality

"Events under study are in temporal transition" (Connelly and Clandinin, 2006: 479). Directing attention temporally points inquirers toward the past, present, and future of people, places, things, and events under study. The importance of temporality in narrative inquiry comes from philosophical views of experience where the "formal quality of experience through time is [seen as] inherently narrative" (Crites, 1971: 291). Drawing on philosophers such as Carr (1986) who shows that "we are composing and constantly revising our autobiographies as we go along" (p. 76), narrative inquirers need to attend to the temporality of their own and participants' lives, as well as to the temporality of places, things, and events.

Sociality

Narrative inquirers attend to both personal conditions and, simultaneously, to social conditions. By personal conditions, "we mean the feelings, hopes, desires, aesthetic reactions and moral dispositions" (Connelly and Clandinin, 2006: 480) of the inquirer and participants. Social conditions refer to the milieu, the conditions under which people's experiences and events are unfolding. These social conditions are understood, in part, in terms of cultural, social, institutional, and linguistic narratives. A second dimension of the sociality commonplace directs attention to the inquiry relationship between researchers' and participants' lives. Narrative inquirers cannot subtract themselves from the inquiry relationship.

Place

Connelly and Clandinin (2006) define place as "the specific concrete, physical and topological boundaries of

place or sequences of places where the inquiry and events take place” (p. 480). The key to this commonplace is recognizing that “all events take place some place” (p. 481). Indeed, for narrative inquirers such as Marmon Silko (1996), our identities are inextricably linked with our experiences in a particular place or in places and with the stories we tell of these experiences.

Possible Starting Places for Narrative Inquiries

While most narrative inquiries begin with telling stories, that is, with a researcher interviewing or having conversations with participants who tell stories of their experiences, “a more difficult, time-consuming, intensive, and yet, more profound method is to begin with participants’ living because in the end, narrative inquiry is about life and living” (Connelly and Clandinin, 2006: 478). Furthermore, from either starting point, narrative inquirers situate themselves in more or less relational ways with their participants. Some narrative inquirers see themselves and their participants as co-composing each aspect of the inquiry as well as their lives as they live out the inquiry. Other narrative inquirers see themselves and their participants at more of a distance, and acknowledge the relational aspects as less important. We discuss a special form of narrative inquiry, autobiographical narrative inquiry, in a separate section.

Within each section, we outline ways of analyzing field texts. These ways of analyzing are framed by thinking narratively, that is, by inquiring within the three commonplaces: temporality, sociality, and place.

Beginning with telling stories

Most narrative inquiries begin with asking participants to tell their stories, either in one-to-one situations or in groups. In one-to-one situations, participants are asked to tell their stories in a variety of ways: by responding to more or less structured interview questions; by engaging in conversation or dialog; by telling stories triggered by various artifacts such as photographs or memory-box items. In group situations, two or more participants meet together with the inquirer to tell stories of their experience when they have lived through similar situations. Texts are created from the told stories and these texts are analyzed using different analytic frames. Chase (2005) identified five diverse approaches for analyzing told stories: a psychosocial developmental approach; an identity approach with a focus on how people construct themselves within institutional, cultural, and discursive contexts; a sociological approach with a focus on specific aspects of people’s lives; a narrative ethnographic approach; and an autoethnographic approach. While Chase’s approaches do not have clear borders distinguishing one approach from

another, they give a sense of the diversity of approaches used in analyzing texts when the starting point is telling stories.

Beginning with living stories

As noted above, some narrative inquiries also begin with participants’ living stories although telling or told stories also take their place within such studies. Craig and Huber (2007) summarize the tensions within narrative inquiries undertaken from this starting point. Others such as Bach (2007) using participants’ photographs of their unfolding lives and Nelson (2008) highlighting change in participants’ lives through engaging in narrative inquiry also begin with living stories. Analysis and interpretation of living stories use some of the same approaches as narrative inquiry beginning with telling stories although tensions, bumping places, and temporal threads are more commonly used as analytic tools.

Autobiographical Narrative Inquiry

Autobiographical narrative inquiry is a special form of narrative inquiry and is closely linked to autoethnography. Understanding life as narrative led Bruner (2004) to posit that “the stories we tell about our lives ... [are] our ‘autobiographies’” (p. 691). Yet, narrative inquirers understand that telling stories is not an untethered process. How people tell their stories and what their stories tell is shaped by “cultural conventions and language usage ... [and] reflect the prevailing theories about ‘possible lives’ that are part of one’s culture” (p. 694). Audience also shapes autobiographical narrative inquiry. Who the characters are in people’s stories, the plotlines people choose to tell, and the audiences to whom they tell, all influence autobiographical narrative inquiry. As Freeman (2007) writes about autobiographical narrative inquiry, “the interpretation and writing of the personal past ... is ... a product of the present and the interests, needs, and wishes that attend it. This present, however – along with the self whose present it is – is itself transformed in and through the process” (pp. 137 and 138). These ideas, highlighted in autobiographical narrative inquiry, are also present in narrative inquiries undertaken with others but are often less visible.

Research Design Considerations

Whether a narrative inquiry begins with telling or living stories and is more or less relational, there are generally agreed-upon considerations in designing narrative inquiries. A more detailed description of these considerations is found in Connelly and Clandinin (2006) and Clandinin *et al.* (2007). We outline the most salient design considerations in what follows.

Justification

In narrative inquiry it is important to think about justifying the research in three different ways.

Personal justification

Narrative inquirers begin with personal justification, that is, by justifying the inquiry in the context of their own life experiences, tensions, and personal-inquiry puzzles. Personal justification is commonly only thinly described in published narrative inquiries. Narrative theses and dissertations include more detailed personal justification for the inquiry.

Practical justification

In order to justify narrative inquiry practically, researchers attend to the importance of considering the possibility of shifting or changing practice. For example, practical justifications are sometimes made in narrative inquiries around teacher-education puzzles concerning the kinds of situations in which preservice students might undertake practicum, deepening their understandings of who they are in relation with children and families or in medical education around puzzles concerning the conditions under which medical residents engage in reflecting on their clinical practice.

Social justification

Narrative inquiries are socially justified in terms of addressing the so-what and who-cares questions important in all research undertakings. We can think of social justification in two ways: theoretical justification as well as social action and policy justifications. Theoretical justification comes from justifying the work in terms of new methodological and disciplinary knowledge. Social action or policy justification comes in terms of social action such as making visible the intergenerational impact of residential schools on Aboriginal youth.

Naming the Phenomenon

Thinking narratively about the phenomenon throughout the inquiry

Thinking narratively about a phenomenon, key to undertaking narrative inquiries, entails thinking within the three commonplaces of narrative inquiry – temporality, sociality, and place. Thinking narratively about the phenomenon is necessary throughout the inquiry from framing the research puzzle, to being in the field, to composing field texts, and finally, to composing research texts. For example, as Paley (1997) thinks narratively about children's experiences in preschool settings she attends to social, cultural, and institutional narratives in which particular children's lives unfold. Thinking in this way, Paley

attends to the particularities of the places in which each child lives and goes to school, to each child's particular interactions and relationships and how each child responds in particular esthetic, emotional, and moral ways. Thinking in this way highlights the shifting, changing, personal, and social nature of the phenomenon under study. Thinking narratively about a phenomenon challenges the dominant story of phenomenon as fixed and unchanging throughout an inquiry. Thinking narratively also influences the living of a narrative inquiry. Many narrative inquirers draw on ideas such as self-facing, liminality, relational knowing, world-traveling, truth as communal, and unknown and not-knowing to describe their own and their participants' living throughout an inquiry.

Framing a research puzzle

Framing a research puzzle is part of the process of thinking narratively. Each narrative inquiry is composed around a particular wonder and, rather than thinking about framing a research question with a precise definition or expectation of an answer, narrative inquirers frame a research puzzle that carries with it "a sense of a search, a 're-search', a searching again", "a sense of continual reformulation" (Clandinin and Connelly, 2000: 124).

Living the Narrative Inquiry

Narrative inquiry is a process of entering into lives in the midst of each participant's and each inquirer's life. What this draws attention to is the importance of acknowledging the ongoing temporality of experience when it is understood narratively. Narrative inquiry always begins in the midst of ongoing experiences. In this process, inquirers continue to live their stories, even as they tell stories of their experiences over time. Inquiries conclude still in the midst of living and telling, reliving and retelling, the stories of the experiences that make up narrative inquirers' and participants' lives, both individual and social. The process of narrative inquiry is described as a recursive process of being in the field, composing field texts, drafting and sharing interim research texts, and composing research texts.

From field to field texts

Living in the midst of stories in the field is not an easy undertaking. The field can be the ongoing conversations with participants where they tell their stories or the living alongside participants in a particular place or places. Being in the field, then, involves settling into the temporal unfolding of lives. Sarris (1993) notes that stories are often not shared in "chronological sequence" (p. 1) and hooks (1997) explains that people's lived and told stories are not linear – they do not necessarily "move from point A to

point B". These narrative qualities of lived and told stories arise from the temporal nature of experience in which people are simultaneously participants in and tellers of their life stories (Carr, 1986).

There are multiple ways to gather, compose, and create field texts (data) from studying the experiences of participants and inquirers in a narrative inquiry. Field texts can include transcripts of conversations, field notes, family stories, memory-box artifacts, photographs, and other texts that are composed by narrative inquirers and participants to represent aspects of lived experience. Whether narrative inquirers are listening to participants' told stories or living alongside participants as their lives unfold in particular contexts, interpretation of the stories lived and told is an essential, ongoing aspect. Being attentive to the relational aspects of working with participants within the conceptual frame of the commonplaces requires that narrative inquirers and participants acknowledge that they are always interpreting their pasts from their present vantage points. In this way, narrative inquirers actively attend to and listen to participants' stories knowing that they "give shape to what . . . [they] hear, mak[ing] over . . . [participants'] stories into something of . . . [their] own" (Coles, 1989: 19).

From field texts to interim research texts

"Dissection is an essential part of scientific method, and it is particularly tempting to disassemble" (Bateson, 1989: 10) people's experience when narrative inquirers leave the field and begin analysis and interpretation at a distance from participants. Narrative inquirers work to resist this temptation. The movement from composing field texts to composing interim research texts is a time marked by tension and uncertainty for narrative inquirers. While some interpretations are always underway as the inquiry continues to be lived out with participants in the field, at some point there is a move away from the close intensive contact with participants. Given the quantity of field texts from interview and conversation transcripts and possibly artifacts, documents, photographs, and field notes, all composed with attention to temporality, sociality, and place, beginning the analysis and interpretation by drafting interim research texts allows narrative inquirers to continue to engage in relational ways with participants. In composing interim research texts, narrative inquirers continue to think narratively, that is, positioning field texts within the commonplaces. Interim research texts are often partial texts that are not closed to allowing participants and researchers to further co-compose storied interpretations open to negotiation of a multiplicity of possible meanings. Bringing back interim research texts to further engage in negotiation with participants around unfolding threads of experience is central in composing research texts. The dialog with participants around interim research texts can lead the inquirer back for more intensive work with the participant if further field texts are

needed in order to compose a more complex account of the participants' experiences.

From interim research texts to research texts

Mishler, in conversation with Clandinin and Murphy (2007) and reflecting on the enormous quantity of field texts that narrative inquirers compose with participants, notes that what becomes shared in research texts is usually only a small portion of the overall data. Mishler encourages narrative inquirers to make visible in their research texts the process by which they chose to foreground particular stories. As earlier described, there are multiple approaches to analyzing field texts. However, as Gergen (2003) cautions, an "analytical method of deconstructing stories into coded piles" could undermine "the aims of the research" (p. 272) by directing attention away from thinking narratively about experience.

Working with metaphors, creating visual and textual collages, found poetry, word images, and photographs, narrative inquirers create research texts that show the complex and multilayered storied nature of experience. In this way, they create research texts that represent the complexity of people's lives and experiences.

Ongoing negotiation with participants allows narrative inquirers to create research texts that both critically and deeply represent narrative inquirers' and participants' experiences while also maintaining each person's integrity and their relationship into the future. When, for example, a narrative inquiry shows the bumping up of participants' lives (and narrative inquirers' lives) with dominant cultural and institutional narratives, various ways of working with fictionalization in research texts can enable these stories to be told without harming participants' lives or the relationships composed by narrative inquirers and participants. Signature and voice both shape research texts. It is important that the voice of the inquirer does not write over the voices of participants in the final research texts by using an overly dominant researcher signature.

Positioning

In relation to other research

Some forms of qualitative research focus on a search for common themes across participants' stories or use participants' stories to develop or confirm existing taxonomies or conceptual systems. As narrative inquirers attend to individual's lives as they are composed over time in relation with people and situations in a particular place or places, the focus remains on lives as lived and told throughout the inquiry. The knowledge developed from narrative inquiries is textured by particularity and incompleteness – knowledge that leads less to generalizations and certainties

(Clandinin and Murphy, 2007) and more toward wondering about and imagining alternative possibilities.

In relation to research undertaken from differing epistemological and ontological assumptions

Working from a metaphor of borderlands between narrative inquiry and research undertaken from other epistemological and ontological assumptions, such as those underlying postpositivism, poststructuralism, and Marxism, Clandinin and Rosiek (2007) delineated ways in which narrative inquirers work from different assumptions. Beginning with ways experience is often viewed differently by narrative inquirers and researchers of other methodologies, Clandinin and Rosiek trace how a Deweyian view, in which experience “is understood as the continuous interaction of human thought with our personal, social, and material environment” (p. 39), shapes “the kinds of questions asked and methods employed” (p. 43) across methodologies. This understanding of experience also shapes ways in which the inquiry is both lived through and subsequently shared with a broader audience. Differences in views of reality, knowledge developed from an inquiry, the relationship between experience and context, and the relationship between researchers and participants all shape borders. Clandinin and Rosiek encourage narrative inquirers to understand and to learn from differing epistemological and ontological assumptions so as to strengthen the future of narrative research.

Ethical Considerations

Narrative inquirers comply with the legal and procedural aspects of ethics held by institutional research boards. However, because of the relational aspects of narrative inquiries, ethical considerations are of prime importance throughout the inquiry. Lieblich urges narrative inquirers to move beyond the institutional narrative of do no harm by learning an attitude of empathic listening, by not being judgmental and by suspending their disbelief (Clandinin and Murphy, 2007: 647) as they attend to participants’ stories.

Often woven deeply by “fidelity to relationships” (Noddings, 1986), the ethical considerations in narrative inquiries are commonly thought about as responsibilities negotiated by participants and narrative inquirers at all phases of the inquiry (Clandinin and Connelly, 2000). These relational responsibilities are increasingly understood as long term, that is, as attentive to participants’ and narrative inquirers’ lives both as the inquiry is undertaken and research texts are written and, also as their lives continue to unfold into the future (Huber *et al.*, 2006).

The relational aspects of narrative inquiries compel narrative inquirers to pay attention to particular ethical matters as research texts are written. Narrative inquirers understand that a person’s lived and told stories are who

they are and who they are becoming and that these stories sustain them. This understanding shapes the necessity of negotiating research texts that respectfully represent participants’ lived and told stories.

Negotiating research texts creates a space where participants’ narrative authority is honored. Issues of anonymity and confidentiality take on added importance as the complexity of lives are made visible in research texts. Strategies such as fictionalizing and blurring identities and places are often used. Narrative-inquiry research texts often call forward increased attentiveness to ethical matters.

Issues in Representation

As noted earlier, voice and signature are key considerations in composing research texts. There are a range of possible narrative forms. However, it is important to attend to forms that fit the lives of the participants and narrative inquirers being represented. Sometimes, particular metaphors or genres become apparent in the field texts and are used in representational forms in final research texts. However, these cannot be imposed on the field texts *a priori*. It is often helpful for narrative inquirers to participate in response communities where, alongside narrative inquirers as well as researchers from differing methodological backgrounds, they share and respond to one another’s thinking or writing in progress in ways that are attentive to the lives being represented.

Issues of audience are also important. Narrative inquirers need to be attentive to the features of the discourse communities where research texts are shared so that the lives represented are respected. Given that narrative inquiry is a new methodology, some audiences are unfamiliar with criteria to judge and respond to narrative-inquiry research texts. Criteria for judging narrative inquiries follow from the definition of narrative inquiry and the conceptual frame for thinking narratively.

Change Dimension in Narrative Research Living, Telling, Retelling, and Reliving Stories

Narrative inquirers see change as part of the process of narrative inquiry. Linking back to personal, practical, and social justification, change is seen as possibly occurring in multiple dimensions. Through engaging with participants, narrative inquirers see themselves and participants as each retelling their own stories, and as coming to changed identities and practices through this inquiry process. Change also occurs as phenomena under study are understood in new ways and, in this way, new theoretical understandings emerge. In this midst, much possibility exists for social change, that is, for the creation of shifted social, cultural, institutional, and linguistic narratives.

See also: Interviews and Interviewing; Participant Observation.

Bibliography

- Bach, H. (2007). Composing a visual narrative inquiry. In Clandinin, D. J. (ed.) *Handbook of Narrative Inquiry: Mapping a Methodology*, pp 280–307. Thousand Oaks, CA: Sage.
- Bateson, M. C. (1989). *Composing a Life*. New York: Atlantic Monthly Press.
- Bruner, J. (2004). Life as narrative. *Social Research* 71(3), 691–710.
- Carr, D. (1986). *Time, Narrative and History*. Bloomington, IN: Indiana University Press.
- Chase, S. E. (2005). Narrative inquiry: Multiple lenses, approaches, voices. In Denzin, N. K. and Lincoln, Y. (eds.) *The Sage Handbook of Qualitative Research*, 3rd edn, pp 651–679. London: Sage.
- Clandinin, D. J. and Connelly, F. M. (2000). *Narrative Inquiry: Experience and Story in Qualitative Research*. San Francisco, CA: Jossey-Bass.
- Clandinin, D. J. and Murphy, S. (2007). Looking ahead: Conversations with Elliot Mishler, Don Polkinghorne, and Amia Lieblich. In Clandinin, D. J. (ed.) *Handbook of Narrative Inquiry: Mapping a Methodology*, pp 632–650. Thousand Oaks, CA: Sage.
- Clandinin, D. J., Pushor, D., and Orr, M. A. (2007). Navigating sites for narrative inquiry. *Journal of Teacher Education* 58(1), 21–35.
- Clandinin, D. J. and Rosiek, J. (2007). Mapping a landscape of narrative inquiry: Borderland spaces and tensions. In Clandinin, D. J. (ed.) *Handbook of Narrative Inquiry: Mapping a Methodology*, pp 35–75. Thousand Oaks, CA: Sage.
- Coles, R. (1989). *The Call of Stories: Teaching and the Moral Imagination*. Boston, MA: Houghton Mifflin.
- Connelly, F. M. and Clandinin, D. J. (2006). Narrative inquiry. In Green, J., Camilli, G., and Elmore, P. (eds.) *Handbook of Complementary Methods in Education Research*, pp 375–385. Mahwah, NJ: Erlbaum.
- Craig, C. and Huber, J. (2007). Relational reverberations: Shaping and reshaping narrative inquiries in the midst of storied lives and contexts. In Clandinin, D. J. (ed.) *Handbook of Narrative Inquiry: Mapping a Methodology*, pp 251–279. Thousand Oaks, CA: Sage.
- Crites, S. (1971). The narrative quality of experience. *Journal of the American Academy of Religion* 39(3), 291–311.
- Freeman, M. (2007). Autobiographical understanding and narrative inquiry. In Clandinin, D. J. (ed.) *Handbook of Narrative Inquiry: Mapping a Methodology*, pp 120–145. Thousand Oaks, CA: Sage.
- Gergen, M. (2003). Once upon a time. A narratologist's tale. In Daiute, C. and Lightfoot, A. (eds.) *Narrative Analysis: Studying the Development of Individuals in Society*, pp 267–285. Thousand Oaks, CA: Sage.
- hooks, b. (1997). *Wounds of Passion: A Writing Life*. New York: Henry Holt.
- Huber, M., Clandinin, D. J., and Huber, J. (2006). Relational responsibilities as narrative inquirers. *Curriculum and Teaching Dialogue* 8(1–2), 209–223.
- Marmon Silko, L. (1996). *Yellow Woman and a Beauty of the Spirit: Essays on Native American Life Today*. New York: Touchstone.
- Nelson, C. (2008). We are all affected: considering the recovery of HIV/AIDS infected and affected children. *International Journal of Children's Spirituality* 13(3), 203–211.
- Noddings, N. (1986). Fidelity in teaching, teacher education, and research for teaching. *Harvard Educational Review* 56(4), 496–510.
- Paley, V. G. (1997). *The Girl with the Brown Crayon. How Children Use Stories to Shape Their Lives*. Cambridge, MA: Harvard University Press.
- Reissman, C. K. and Speedy, J. (2007). Narrative inquiry in the psychotherapy professions: A critical review. In Clandinin, D. J. (ed.) *Handbook of Narrative Inquiry: Mapping a Methodology*, pp 426–456. Thousand Oaks, CA: Sage.
- Sarris, G. (1993). *Keeping Slug Woman Alive: A Holistic Approach to American Indian Texts*. Los Angeles, CA: University of California Press.

Further Reading

- Anzaldúa, G. (1987). *Borderlands/La Frontera: The New Mestiza*. San Francisco, CA: Aunt Lute Books.
- Atkinson, R. (2007). The life story interview as a bridge in narrative inquiry. In Clandinin, D. J. (ed.) *Handbook of Narrative Inquiry: Mapping a Methodology*, pp 224–245. Thousand Oaks, CA: Sage.
- Basso, K. (1996). *Wisdom Sits in Places: Landscape and Language among the Western Apache*. Albuquerque, NM: University of New Mexico Press.
- Chan, E. (2006). Teacher experience of culture in curriculum. *Journal of Curriculum Studies* 38(2), 161–176.
- Charon, R. and Montello, M. (eds.) (2002). *Stories Matter: The Role of Narrative in Medical Ethics*. New York: Routledge.
- Clandinin, D. J., Huber, J., Huber, M., et al. (2006). *Composing Diverse Identities: Narrative Inquiries into the Interwoven Lives of Children and Teachers*. New York: Routledge.
- Connelly, F. M. and Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational Researcher* 19(5), 2–14.
- Heilbrun, C. G. (1999). *Women's Lives: The View from the Threshold*. Toronto, ON: University of Toronto Press.
- Hollingsworth, S. and Dybdahl, M. (2007). Talking to learn: The critical role of conversation in narrative inquiry. In Clandinin, D. J. (ed.) *Handbook of Narrative Inquiry: Mapping a Methodology*, pp 146–176. Thousand Oaks, CA: Sage.
- Hollingsworth, S., Dybdahl, M., and Minarik, L. (1993). By chart and chance and passion: The importance of relational knowing in learning to teach. *Curriculum Inquiry* 3(1), 5–35.
- Kerby, A. P. (1991). *Narrative and the Self: Studies in Continental Thought*. Bloomington, IN: Indiana University Press.
- Li, Y. Mitton-Kükner, J., and Yeom, J. S. (2008). Keeping hope alive: Reflecting upon learning to teach in cross-cultural contexts. *Reflective Practice* 9(3), 245–256.
- Lieblich, A. (1996). Some unforeseen outcomes of conducting narrative research with people of one's own culture. In Josselson, R. (ed.) *Ethics and Process in the Narrative Study of Lives (The Narrative Study of Lives, Vol. 4)*, pp 172–184. Thousand Oaks, CA: Sage.
- Lugones, M. (1987). Playfulness, 'world-traveling' and loving perception. *Hypatia* 2(2), 3–19.
- Pinnegar, S. and Daynes, G. (2007). Locating narrative inquiry historically: Thematics in the turn to narrative. In Clandinin, D. J. (ed.) *Handbook of Narrative Inquiry: Mapping a Methodology*, pp 3–34. Thousand Oaks, CA: Sage.
- Rogers, A. (2007). The unsayable, Lacanian psychoanalysis, and the art of narrative interviewing. In Clandinin, D. J. (ed.) *Handbook of Narrative Inquiry: Mapping a Methodology*, pp 99–119. Thousand Oaks, CA: Sage.
- Steeves, P. (2006). Sliding doors – opening our world. *Equity and Excellence in Education* 39(2), 105–114.
- Steeves, P., Yeom, J., Pushor, D., et al. (2009). The research issues table: A place of possibility for the education of teacher educators. In Craig, C. and Deretchin, L. F. (eds.) *Teacher Education Yearbook XVII: Teacher Learning in Small Group Settings*. Toronto, Canada: Scarecrow Education/Rowman and Littlefield.
- Torgovnick, M. (1994). *Crossing Ocean Parkway*. Chicago, IL: The University of Chicago Press.
- Vinz, R. (1997). Capturing a moving form: 'Becoming' as teachers. *English Education* 29(2), 137–146.

Relevant Website

<http://www.narrativesig.cahs.colostate.edu> – Narrative and Research Special Interest Group of the American Educational Research Association.

Participant Observation

J Schostak, Manchester Metropolitan University, Manchester, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Action research – A research methodology increasingly popular with professionals, in all walks of life, who engage in researching their own practice in the workplace in order to improve the quality of professional understanding, judgment, decision making, and action. More broadly, it can be employed by anyone in any context as a strategy to stimulate change and a sense of empowerment and emancipation from the limitations imposed by current conditions on action.

Intentionality – In phenomenology, this term refers to the directedness of acts of consciousness. Whether thinking, dreaming, desiring, touching, or, otherwise, sensing something, there are states of mind that seem to point toward – or focus on – an object, qualities of things, states of affairs, or events. This feature of intentionality is useful in developing analyses of people's talk concerning their life-worlds and how they relate to people, objects, states of affairs, locations, and the production of events. It is particularly useful in analyzing the typical relationships that obtain between actors in terms of intentional networks.

Intertextuality – Texts do not exist in isolation from other texts. The meaning of a given text takes on further meanings as it is placed into relation with other texts. In this way, meanings are loosened from the author's or speaker's intentions as a particular text takes on meanings derived from other texts. For example, a teacher may talk about the need to discipline a child for his/her own good. In the context of other texts that discourse on the essential sinfulness, or dangerous instincts of human beings, or indeed the sexual connotations of discipline, the original context of talking about disciplining a child in school may take on darker meanings in relation to the broad range of texts in which the term discipline is used. Every text exists in relation to every other text and, in that sense, is intertextual. Text – in its broadest sense – does not just refer to the written word but to any sign, symbol, gesture, or mark that, in some way, can be made to mean something to people. In this way, everyday life is fundamentally intertextual where each meaningful sign relates to and takes on different meanings and a range of possible

interpretable across a range of different contexts. Kristeva coined this term.

Phenomenology – Broadly, in philosophy, this refers to the appearance of phenomena to consciousness. The phenomenological philosophy that influenced Schutz's sociological phenomenology derived from Husserl. Husserl made a distinction between the "natural attitude" of everyday life and the philosophical attitude that is employed to suspend any position for or against the truth or falsity of what is said to be known and real. It is this distinction that Schutz employed in developing his view of the sociological implications of Husserl's phenomenology. Where Husserl developed the philosophical methodology focusing upon the domain that arose following the suspension of belief – or epoché, as Husserl called it – Schutz developed the sociological domain where belief is not systematically suspended but rather is taken for granted as being true for all practical purposes. This focus on the practicality or pragmatism of peoples' typical attitude toward states of affairs in everyday life also drew upon the pragmatism, in particular, of William James and George Herbert Mead and, thus, has much in common with the later developments of symbolic interactionism. In the study of psychological, social, cultural, and political developments in phenomenology has drawn from a multitude of influences – Marxism, psychoanalysis, existentialism, and Hegelian phenomenology among many others.

Triangulation – On a map, a position can be defined by correlating – or triangulating – its latitude and longitude or its position in a grid. Metaphorically, the validity, objectivity, and, indeed, reliability of an observation, and object, a category, a meaning, a state of affairs, or an event is increasingly established through correlating – or cross-checking – a multiplicity of viewpoints directed toward it over time and across contexts. What looks like a street of houses at first sight, might – on further inspection – turn out to be simply a facade constructed for a film set. By walking up to the facade, trying the doors, walking around the back of it, talking with people working at the site, and so on, its nature can be assessed. To come to an appropriate judgment with regard to the validity of the street, more than one

method and more than one perspective had to be employed. In social settings, participant observers can talk with – or, more formally, interview – a range of people, in order to see whether they share understandings about given phenomena. These understandings can then be compared with what is seen during observations where the phenomena described in the interviews are supposed to occur. The more agreements there are between different methods and viewpoints on social phenomena, the stronger the triangulation.

Participant observation claims a special access to insights and understandings concerning the lives of individuals, groups, communities, and peoples. Evolving from the writings of adventurers and missionaries and, later, the early anthropologists, it was adopted by sociologists who wanted to explore the meanings and experiences of people rather than measuring and counting (deWalt and deWalt, 2002). At present, participant observation is employed in any research, methodological perspective, or discipline in order to gain insight into the lives of others. In many ways, it is more than a method for the production of knowledge, understanding, and theory. It is a way of life which – for some – may last a lifetime.

According to Spradley (1980: 58–62), there are different degrees of participation – complete, active, moderate, passive, and nonparticipation. As a passive participant, the observer is present but not engaged – more a bystander looking at a scene of action rather than being part of it. Rather than merely hanging around, loitering, or being a bystander, one can then try to act like the members by moving on from the stage of recording what went on to actually joining in – but not enough to be acceptable as a member. Going further than this involves entering the world of the other as an accepted member – like Whyte (1943) when he joined Doc's gang in Cornerville. However, to be a complete participant observer, Doc would already have had to be a gang member who decided to undertake research while participating. The parallel would be, say, a teacher who decides to research the classroom – perhaps, focusing on their own practice in order to improve it as in Action Research.

In each case, participant observation involves being with others – whether living with people far from home whose language, customs, or behaviors are not known, as did Malinowski in his study of the Trobriand Islanders, or, perhaps, like Whyte (1943), where the participant observer goes to live a slum area, or join a criminal gang or study power elites. In these cases, there is a clear demarcation line between my home and their home, my way of life and their way of life, and where I go to visit strangers with the object of becoming familiar with their way of life, their

view of the world. The demarcation becomes less obvious when what is observed is already a part of the researcher's everyday life as, for example, in observing what happens in one's life, one's own community – say in a school or in one's own place of work (as in action research) and the virtual worlds of globalized communications technologies. Whenever, wherever, and however people interact, share experiences, and, in some way, organize their lives in relation to each other, it becomes a possible focus for the participant observer. Understanding the nature and process of being a participant observer begins only at the moment of joining.

Joining

To know the world of others – to understand a particular group – involves joining them in some way, being with people in their everyday locations and activities, and learning how to do what they do in the ways that they do it in the locations where they do it. Joining with, witnessing what happens, and acting alongside others makes participant observation more than just a method or technique for data collection. The participant observer is part of the data itself – that is, responsible for how this data is to be represented, analyzed, interpreted, and used, as in this example from a narrative account of fieldwork by Schostak (1980):

Entering the tutor room, a chemistry laboratory, under the gaze of Jacko's mates, I feel inverted – watched, examined. Where shall I sit?

Carol, a girl I know, smiles, invites me to Jacko's group.

The tutor, a teacher I know well enters. He ignores me.

"Jacko. All of you, come and sit at the front."

No one moves. I feel uneasy.

"Come on now!"

"But there's gas taps at that table."

"I don't care, I've told you to sit down here. Now do it."

Carol moves, the rest follow and I am last.

After the formalities of the tutor period – registration, reading a pupil news letter – the teacher approaches smiling. "So you're following Crawley around today?"

"Yes," I turn away as best I can.

"Good luck," the teacher retreats to the doorway where for the remaining time he blocks the escape of three girls who push against him.

Being there and living it, the researcher is a direct participant. It is an approach owing much to the sociologists who, during the first-half of the twentieth century, became known as the Chicago School. Their attention was on the city as a natural laboratory of human behavior (Park and Burgess, 1925). For example, Thomas studied the communities of Polish immigrants (1918) – and, Whyte, the Italian community (1943) – as the natural

areas produced by human habitation and use. In any city, there are many such areas, such as the business district, the Red Light area, or the territories of particular gangs, as well as the secluded areas of the wealthy. In each case, the initial concern is to gain entrance. This, typically, involves finding a gatekeeper— that is, someone whose knowledge and authority can ease the passage into the world of others by hanging around sufficient times in a place where the target group tends to meet until someone begins a conversation. Such gatekeepers are easier to identify and contact in formal organizations – for example, the headteacher or principal of a school, the leader of a youth group or sports team, the head of a committee or director of a business. In other circumstances, greater secrecy may be vital as in the study of a criminal gang. To openly identify oneself might cost one's life, or, at least, physical health. Whether open or covert, participant observation requires developing strategies for meeting, joining, and engaging with others in their everyday scenes of action.

Ethics, *Dramatis Personae*, Locations, and Scenes of Action

Joining with others means getting to know the members of scenes of action – the *dramatis personae* – getting to know their values, their reasons for action, and their ways of acting. For example, being with Jacko meant going where he went, meeting other people along the way – pupils and teachers – and, when engaging with these others, reflecting ethically on the decisions concerning what to say, how to act, and what and how to record. Observing Jacko's (A) actions toward his mates (C, D, and E) and say a teacher (B) as they approach or keep their distance, patterns can be observed and noted (Figure 1).

Each individual acts in some way toward another – whether in friendship, hostility, or authority. In phenomenology, this directedness toward another is called an

intentional relation. The diagram, thus, describes an intentional network (Schostak, 2002) of people as they repeatedly engage with each other. In Figure 1, Jacko's (A) attention is directed, in various ways, toward B, C, D, and E. B also has a particular relationship of mediation with G (say, a senior teacher, a parent, or a police officer, or a rival gang leader). This mediation may play out in events such as B telling G about A or misleading G about A's real intentions. Or, instead of people, the focus could be upon valued object B that A and G both desire hence leading to a conflict of interest and thus contributing to a study of the ways in which such conflicts are handled. The participant observer becomes a part of this structure, say G, as participants recognize and trust the researcher and as a valid part of the scene able to apply the ethics of the participants – that is, what is right, good, fair in their eyes.

Following Jacko throughout a particular day was part of a larger study of a school and its community (Schostak, 1983). The ethical principles for the conduct of the project had to be agreed. To do this, the participants needed to have an idea of the purposes of the project, what it involved and how data was going to be used. They needed to be assured that they could not be identified and that what they said and did would be treated confidentially so that others could not use the information to harm them. Such principles can be seen more formally in the guidelines provided by discipline-based organizations such as the American Educational Research Association or the British Sociological Association. Indeed, particular organizations – such as hospitals and universities – typically, have their own ethics committees that oversee research involving their institutions. In practice – for each organization – ethical protocols need to be tailored for the specific circumstances of projects. With a formal organization such as a school, general principles may be agreed in writing with its governing committee. Letters describing the research and assuring confidentiality and the anonymity of names in the use of data will be sent to parents as a basis for agreement to be involved in the research. Participants may be asked to sign letters that confirm their understanding of and agreement to the conditions of participation. However, ethical practice is not simply reducible to – nor does it stop with – satisfying the formal ethical procedures of committees.

The different circumstances of each meeting continually pose ethical issues. Not all of the people – the *dramatis personae* – involved in a given scene of action in the various locations of the project may be aware of the status of the participant observer as researcher rather than as an ordinary member of the scene. Indeed, Hilbert (1980) argues that there is always a degree of covertness in research, as aspects of the researcher's biography and later uses of the data remain obscure in the eyes of participants – whether through lack of understanding, familiarity with academic discourse, or disinterest. In the extreme circumstance of

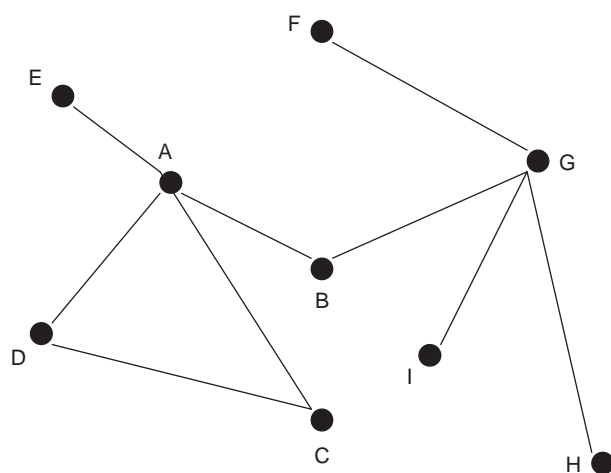


Figure 1 Intentional networks.

joining a gang, or being a football hooligan, it may be argued that not only may acts be seen that are criminal, but also, to gain trust as a *bona fide* member, illegal acts may have to be committed requiring a rethinking of what it means to act ethically (Ferdinand *et al.*, 2007). Without trust, no access to the world of the members would be granted, thus no research into their life-world is possible.

The researcher needs to demonstrate a commitment to the reality of those being observed, however they define this reality as being true, natural, or right – it is what Quine (1948/1949) would call an ontological commitment. Commitment to what there is in the world is revealed through people's vocabulary, ways of talking, ways of justifying beliefs and actions, as well as in carrying out the kinds of acts expected of members. In this way, a shared sense of reality is constructed that acts as the natural, taken-for-granted, familiar basis for undertaking social action. Typically, when this familiar world of shared ontological commitments is disturbed, its constructed features are revealed.

The Familiar and the Strange

Schutz (1964) described how being a stranger sensitizes the researcher to practices, knowledge, beliefs, and values that participants take for granted as being true, real, or right. In learning how to belong – to be a part of a scene of action – a distinction between two worlds is posed: that of the stranger and that of the members of a given group or society that the stranger attempts to enter. Participant observers, similar to strangers, can take nothing for granted – a standpoint paralleling the phenomenological attitude of suspending any position for or against the truth and reality of the world. Presupposed by this phenomenological attitude is a domain of taken-for-grantedness – or paramount reality that is placed in suspension and becomes the focus for the sociological phenomenology of Schutz. As with the symbolic interactionism (Mead, 1934; Blumer, 1969) of the Chicago School, to understand the everyday worlds of people and learn the meanings that actors gave to actions, objects, and the scenes of action in their everyday lives, participant observation was a natural strategy. For symbolic interactionists, it is not enough simply to observe behavior. Behavior has a symbolic dimension. Its meaning to an actor and its relevance to a social situation cannot be known without asking the actor(s) involved. The focus of the symbolic interactionist, then, is upon the symbolic nature of the actions and interactions of people who engage with each other in social settings. Mead (1934) is considered the father of symbolic interactionism, although the name was coined by Blumer (1969).

Being a participant observer, however covertly, heightens awareness of the artificiality or socially constructed nature of the ways in which people behave in everyday life.

Whether standing in bus queues, waiting at hotel reception desks, or relaxing at home – what are the regular behaviors that occur? Why do they occur in that way? Could they be changed as in changing the rules of chess? Garfinkel (1967) demonstrated the fragile nature of social rules in his experiments where, for example, a student might go home and act like a stranger, thus transgressing the tacit rules of normal behavior – often, leading to angry reactions by family members. Less dramatically, when participating in Jacko's activities during the school day, the researcher – as a 32-year old – could not pass as a 14-year-old pupil. What becomes noticeable then, are the moves of inclusion as well as those of exclusion made by different members of the *dramatis personae* in the classroom, in the corridors, and out in the playground. Although Jacko had given permission to be followed, he always placed a distance between himself, his friends, and the researcher, although others invited the researcher to join them. Through such a process of seeking permissions, invitations to join, or distancing and excluding, social space becomes dynamically organized as boundaries between who is in, who is out, who is close, and who is distant. This negotiation of boundaries also took place between the researcher and the teachers and the placing of boundaries varied according to circumstances. Thus, the physical reality of streets, houses, and parks are continually mapped and remapped by the ways in which individuals project and negotiate boundaries around them over time. To see it requires distinguishing between “the scene of action, the human picture of that scene, and the human response to that picture working itself out upon the scene of action. It is like a play suggested to the actors by their own experience, in which the plot is transacted in the real lives of the actors, and not merely in their stage parts” (Lippmann, 1922). In this early statement can be seen the seeds of studying social action like a performance on a stage where the analysis of performance, talk and narrative structure, and a focus on biography becomes important rather than static descriptions of states of affairs (Schostak, 2006).

The projection of a world and its boundaries is most obvious when there is a sense of disjunction or clash between alternative perceptions. This can be seen in the following transcript extract from an action research project involving creative practitioners (artists of various kinds), teachers, and pupils (funded by CapeUK Creative Partnerships for Education (CAPE) 2004–05; see also Schostak and Schostak, (2008: 48–50)). The focus was on making representations of local communities – both real and imagined. The pupils – aged about 10 years – had just watched a video that had been made of their area:

Boy: . . . if you looked at the surroundings they were like, you know ‘oh I wouldn’t like to live there’. Cos if you looked at all the houses they were all boarded up and, all

the graffiti and smashed bottles (. . .)

JFS: So would you have done a different story, you know for (place name)?

Boy: yeah.

JFS: What would you have shown?

Boy: Well I'd have shown the good bits, like the park and . . . (laugh) hmmm, there's a point.

Girl: The good parks, the local parks are all trashed up so we don't get anywhere to play.

Immediately there is a clash between the picture in the head and the picture on the video. There are also the explicit and implicit viewpoints of the *dramatis personae*: the boy, the girl, the creative practitioners, the teachers, and the unnamed people who trashed the parks. The participant observer can elaborate the multiple pictures in the head by meeting, observing, and talking to representatives of the different viewpoints. How these viewpoints relate to each other can be represented as intentional networks in **Figure 2** where the children are represented by group C who are able – through their reflection on their community – to imagine alternative visions of community that are, perhaps, seen to be enjoyed by group B. The depressing reality experienced by group B is represented by the trashed parks. However, group B never enter the children's location. For them, it is – for all practical purposes – invisible.

The central bar symbolizes a boundary dividing the two groups – B and C – and thus two ways of projecting a picture of the world. By participating in each group, the

researcher A is able to build up pictures (a_1 , a_2) of how each constructs their ways of seeing the world with their different ontological commitments: whether the hard-and-fast realities of the social organization of gang territories or the views and fears of locals residents who want to freely walk the streets, or the police and other figures of social and political authority who have quite different conception of how the world should be organized. If group B represents the dominant voice in the social, political, and economic organization of society, it will see itself as exemplars of good order. The boundary that delimits their picture of the world acts like a mirror reflecting back their own values (represented by the double arrow in the diagram). For B, C is rendered marginal at best – perhaps an underclass – invisible at worst.

Rendering the invisible visible means transgressing the bar between the dominant and the powerless. The role of the participant observer is already transgressive. Between these multiple camps, the participant observers are always, to a degree, out of place because they are always between worlds. The danger is in slipping into a position of authority granted by some superior insight into what are the right ways of seeing.

Fieldwork as a Basis for Validity, Generalization, Reliability, and Objectivity

Fieldwork involves making records of what is seen, heard, and done in particular locations under particular conditions: making extensive notes, maps, and drawings; committing to memory what happened; recording sounds and images; collecting artifacts and documents, etc. It involves decisions regarding how obvious the process of recording will be. Note taking when being a member of a gang is unlikely to be perceived as natural but, as a student in a classroom (Hilbert, 1980), it is perfectly natural. Thus, depending upon circumstances, making notes may take place either during or following the scenes of action.

Even as a bystander, or recipient of stories, the participant observer becomes a member of the *dramatis personae* of a scene of action and thus as a member of the different *dramatis personae* that engage with each other in scenes of action across a range of locations and becomes a carrier of the local stories similar to the other members – each with their views. The multiplicity of views constructing stories of what happened, with whom, how, where, and why are retold for particular purposes and become vital elements in the sense of reality and drama of everyday life.

Each story retold connects to other stories, weaving each actor, the scenes of action, and the objects involved in the action into a reality, a sense of my place, my people, and my world where what happens becomes familiar, expected, routine, and ritualized and where the unusual

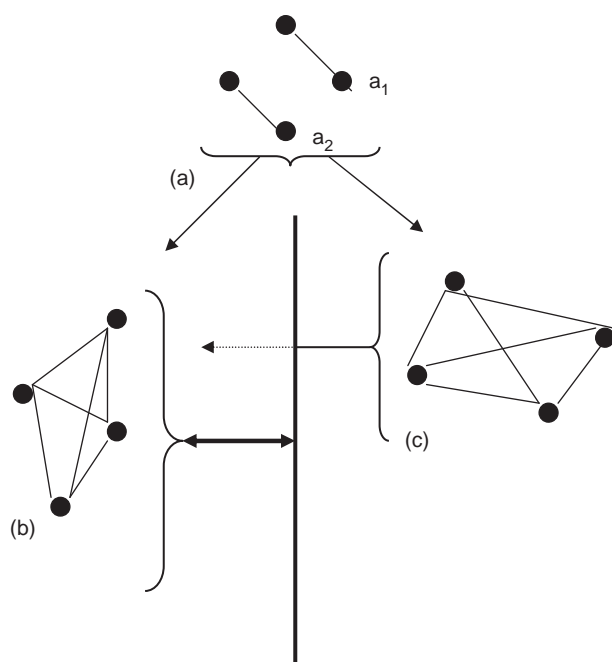


Figure 2 Mirror groups. From Schostak, J. F. and Schostak, J. R. (2008). *Radical Research. Designing, Developing and Writing Research to Make a Difference*, p 172. London: Routledge.

stands out. Aspects of one story may recall other stories – stories that elaborate on the character of one of the actors, or add to a sense of pride, frustration, or tragedy whether it is “teachers always pick on you like that” or “that’s what always happens to people like us.” This connectedness – or intertextuality – composed of shared and sharable anecdotes, local histories, and biographies becomes the vital data of the participant observer (Schostak, 2006) who also contributes to the stories through various degrees of participation. This intertextuality, with its multiplicity of viewpoints scanning shared or disputed locations, territories, scenes of action, and events, contributes an intersubjective sense of generalizability, validity, reliability, and objectivity of observations and accounts described during fieldwork. More specifically – in phenomenological terms – across all the variations of viewpoint some relatively invariant or typical relations, actions, objects, and meanings attributed to them can be identified, thus leading to judgments as to the generalization of the observations and analyses made. The process is, typically, referred to as triangulation. Features seen with one group and in one scene of action can be compared and contrasted with other groups, scenes of action, and locations. With the aim of identifying the range and limits of variation across individuals, forms of social organization, and locations, the participant observer can seek out comparative and contrastive situations, people, and locations in order to test out emergent understandings, analyses, and theories in a process called theoretical sampling.

The participant observer, then, is continually learning, making analyses, and searching for opportunities to test out emergent understandings and theories concerning the social world being studied. As well as making field-notes to describe observations of people, events, and places, analytic memos can be written in order to formulate potential categories by which insiders describe the circumstances of a situation and provide motives for their action (cf. Mills, 1940) and theoretical memos on emergent theory that can be tested out through further observations (Schostak, 2002). The participant observer is interested not just in a single subject’s point of view but in how the sense of a common world, shared reality, and orchestrated behavior is brought about as an intersubjective reality in the lives of people.

When the researcher is able to represent the intersubjectively constructed world(s) of actors in ways that they recognize as their world, their definitions, and their way of doing things – whether through performance as a member or in spoken and written accounts – this achieves a degree of validation and reliability. More critically, participant observation by seeking out the multiple voices and perspectives and learning from their viewpoints provides a critical resource for researchers in their debates with regard to what counts as knowledge. Whether it is the view that there is the inevitability of a distortion – a

violence even – in the interpretations of the researcher as shown by the critiques of feminists or by critical race theory or poststructuralism, the debate is continually renewed, shifted, and challenged by a return to the field.

The Significance of Participant Observation

In summary, as Thomas (in Thomas and Thomas, 1929: 571–572) put it: “If men define situations as real they are real in their consequences.” The task is to learn how situations are defined as a member, how reality is constructed as a member, what it means to be a member, or even to pass as a member of a particular group. This is not to say that the social realities of members are homogeneous. Indeed, where there are disputes or disagreements, these are managed or resolved according to the rules, rituals, habits, beliefs, values, and social positions of the members. For that reason – from an ethical and democratic standpoint – the participant observer records the range of disagreements as well as agreements in order to produce a record that the French political philosopher Jacques Rancière calls being faithful to the disagreements (cf., Schostak and Schostak, 2008). This includes not only a representation of the world of those being studied but also extends to the public debates concerning what counts as knowledge and whose knowledge and perspectives are to be included in public debates that affect people’s lives. In this way, participant observation contributes to the creation of public debate about what counts as knowledge, justice, and the good society.

The only way of learning participant observation is to do it, be there, and live it.

Bibliography

- Blumer, H. (1969). *Symbolic Interactionism: Perspective and Method*. Englewood Cliffs, NJ: Prentice-Hall.
- deWalt, K. M. and deWalt, B. R. (2002). *Participant Observation. A Guide for Fieldworkers*. Walnut Creek: AltaMira Press.
- Ferdinand, J. M., Pearson, G., Rowe, M., and Worthington, F. (2007). A different kind of ethics. *Ethnography* 8(4), 521–544.
- Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Hilbert, R. A. (1980). Covert participant observation: On its nature and practice. *Journal of Contemporary Ethnography* 9, 51.
- Lippmann, W. (1922). *Public Opinion*. New York: Harcourt Brace.
- Mead, G. H. (1934). *Mind, Self and Society*. Chicago, IL: University of Chicago Press.
- Mills, C. W. (1940). Situated actions and vocabularies of motive. *American Sociological Review* 5, 904–913.
- Park, R. E. and Burgess, E. W. (1925). *The City*. Chicago, IL: University of Chicago Press.
- Quine, W. V. O. (1948/1949). On what there is. *Review of Metaphysics* 2, 21–38.
- Schostak, J. F. (1980). A day with Jacko: An experience of fieldwork in school. <http://www.enquirylearning.net/ELU/Issues/Research/jacko.html> (accessed September 2009).

- Schostak, J. F. (1983). *Maladjusted Schooling: Deviance, Social Control and Individuality in Secondary Schooling*. London, Philadelphia: Falmer.
- Schostak, J. F. (2002). *Understanding, Designing and Conducting Qualitative Research in Education. Framing the Project*. Buckingham: Open University.
- Schostak, J. F. (2006). *Interviewing and Representation in Qualitative Research Projects*. Buckingham: Open University.
- Schostak, J. F. and Schostak, J. R. (2008). *Radical Research. Designing, Developing and Writing Research to Make a Difference*. London: Routledge.
- Schutz, A. (1964). The stranger: An essay in social psychology. In Brodersen, A. (ed.) *Collected Papers. Vol. II. Studies in Social Theory*, pp 91–105. The Hague: Martinus Nijhoff.
- Spradley, J. P. (1980). *Participant Observation*. New York: Holt, Reinhart and Winston.
- Thomas, W. I. and Thomas, D. (1928). *The Child in America: Behavior Problems and Programs*. New York: Alfred A. Knopf.
- Whyte, W. F. (1943). *Street Corner Society. The Social Structure of an Italian Slum*. Chicago, IL: University of Chicago Press.
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Garden City, NY: Doubleday.
- Goodson, I. (2001). The story of life history: Origins of the life history method in sociology. *Identity* **1**(2), 129–142.
- Maines, D. R. (1993). Narrative's moment and sociology's phenomena – toward a narrative sociology. *Sociological Quarterly* **34**(1), 17–37.
- McRobbie, A. (1982). *The Politics of Feminist Research: Between Talk, Text and Action Feminist Review* **12**, 46–57.
- Patrick, J. (1973). *A Glasgow Gang Observed*. London: Methuen.
- Schostak, J. F. (1985). Creating the narrative case record. *Curriculum Perspectives* **5**(1), 7–13.
- Somekh, B. (2006). *Action Research: A Methodology for Change and Development*. Buckingham: Open University.
- Suler, J. (1999). The psychology of cyberspace. <http://www-usr.rider.edu/~suler/psycyber/partobs.html> (accessed September 2009).
- Tope, D., Chamberlain, L. J., Crowley, M., and Hodson, R. (2005). The benefits of being there. Evidence from the literature on work. *Journal of Contemporary Ethnography* **34**(4), 470–493.
- Villenas, S. and Deyhle, D. (1999). Critical race theory and ethnographies challenging the stereotypes: Latino families, schooling, resilience and resistance. *Curriculum Inquiry* **29**(4), 413–445.
- Willis, P. (1977). *Learning to Labour*. Farnborough: Saxon House.

Further Reading

- Brewer, R. M. (1989). Black women and feminist sociology: The emerging perspective. *American Sociologist* **20**(1), 57–70.
- Glaser, B. G. and Strauss, A. L. (1967). *The Discovery of Grounded Theory. Strategies for Qualitative Research*. Atherton, CA: Aldine.

Relevant Website

<http://www.enquirylearning.net> – The enquiry Learning Unit website.

Phenomenology

M van Manen and C A Adams, University of Alberta, Edmonton, AB, Canada

© 2010 Elsevier Ltd. All rights reserved.

Phenomenological research in education is a project of careful and systematic reflection on the lived experience of educational (pedagogical, psychological, teaching, learning, and parenting) phenomena. Careful, in the sense that reflecting on experience must be as much as possible free from theoretical, prejudicial, and suppositional intoxications. Systematic, in the sense that phenomenology in education is an approach to inquiry that is guided by philosophical phenomenological methodology and assisted by human science procedures and techniques. However, phenomenology is not just a method that one can employ; it is also an attitude that relies on the perceptiveness, creative insight, interpretive sensitivity, scholarship, and writing competence of the researcher.

The phenomenological attitude comprises a fascination with the uniqueness, the particularity of an experience or phenomenon. When I am in love and I reflect on the meaning and significance of this love, then I am compelled not by abstractions but by the concreteness of my experiences: the sweet taste of that last kiss this morning, the tenderness I feel when I look in my love's face, the longing I experience when reading the love letter, the desire I feel to be the object of my lover's desire, the arousal of voluptuousness. Thus, a phenomenology of love is not primarily pursued through a theoretical discourse or a conceptual analysis of the notion of love. It is pursued through attempts to awaken the experience as we live it, and make contact – through concrete examples and reflection – with the living sensibility of its uniqueness.

Phenomenological research is oriented to the lifeworld as we immediately experience it – prereflectively, rather than as we conceptualize, theorize, categorize, or reflect on it. It is the study of lived or experiential meaning and attempts to describe and interpret these meanings in the ways that they emerge and are shaped by consciousness, language, our cognitive and noncognitive sensibilities, the ontics of meaning, and our personal, social, and cultural preunderstandings. Phenomenology can be adopted to explore the unique meanings of any educational experience or phenomenon. For example, it may study what it is like to have a classroom conversation, how students experience difficulty in learning certain concepts, what it is like to read and write poetry, how tests and examinations are experienced, how digital media technologies such as PowerPoint shape the teaching-learning relation and the knowledge that students learn, how young people encounter success and failure, how students experience

recognition and respect (or the lack of it) in schools, and so forth. Phenomenological inquiry also addresses larger questions of educational significance such as the technological presumptions of curriculum, the increasing instrumentalities of teaching and learning, the scientization of everyday life in schools and classrooms, or the rationalistic–technocratic ontotheology that progressively dominates our thinking. Besides, phenomenology may concern itself with shifts in pedagogy due to new technologies, such as reading and writing online, and the changing quality of privacy, intimacy, and contact in virtual learning environments. In a broad sense, any human experience may become the focus of phenomenological research.

Maurice Merleau-Ponty explained that a proper understanding of phenomenology is only accomplished through doing it. Phenomenological understanding needs to be practiced as method, and identified as a style of thinking – a manner of orienting to experience as we live through it. Within the domain of phenomenological philosophy there exist a variety of traditions and orientations. These traditions have spawned different expressions in the manner that phenomenology is pursued in education and other professional disciplines. In addition, in education certain liberties are taken with the nomenclature of phenomenology. Some research texts classify ethnography, biography, narrative inquiry, and arts-based and action research approaches as falling under the purview of phenomenology, in part because all these social and human sciences have an interest in studying experience and have adopted some of the vocabulary of phenomenology. However, from a stricter phenomenological perspective as reflected in the thoughts and writings of scholars such as Edmund Husserl, Martin Heidegger, and Merleau-Ponty, phenomenology differs from other qualitative approaches in that it remains primarily concerned with the prereflective, preconceptual, prepredicative, and pretheoretic dimensions of existence.

Phenomenological research aims for a certain effect, one that can lead us to suddenly see or grasp a human phenomenon in a way that enriches our understanding of everyday life experience. Such seeing may transform our being and thus our practices. The production of insight is not merely a cognitive affair but must proceed through the creation of a research text that appeals to our cognitive and noncognitive sensibilities. Thus, phenomenological understanding is distinctly existential, emotive, enactive, embodied, situational, and nontheoretic. A powerful

phenomenological text thrives on an irrevocable tension between what is unique and what is shared, between particular and transcendent spheres of the lifeworld. Without this tension, the qualitative research text tends to turn flat, shallow, boring, because it loses the power to break through the taken-for-granted dimensions of everyday life.

Phenomenology may be described as the study of lived experience. The term lived experience is presently rather widely used among the disciplines, but it possesses special methodological significance for phenomenology. The term lived experience derives from the German *Erlebnis* – experience as we live through it and recognize it as a particular type of experience. It could be argued that human experience is the main epistemological basis for many other qualitative research traditions, but the notion of lived experience, as used in the works of Husserl, Merleau-Ponty, and like-minded phenomenologists, announces the intent to explore directly the originary or prereflective dimensions of human existence. So, phenomenology prioritizes and investigates how the human being experiences the world; for example, how the teacher experiences the pedagogical encounter, how the student experiences a moment of failure, how a child experiences distress, and so forth.

From a phenomenological perspective, we always exist (act, think, feel, etc.) in the present of the now. Even when we dwell in memories or in anticipations, or when we dream or hallucinate, we do so in the experiential immediacy of the present moment or the living now. Yet when we try to capture and reflect on the now of our experience, we are always too late. The strange fact is that we always live in the present of the temporal now and yet we can never capture this present. Or perhaps it is better to say that the present as we constantly recall and reconstruct it has never really existed. Phenomenology does not just research human experiences and phenomena, it also continually questions the assumptions and presuppositions that prevent us from adequately understanding and expressing in words the living moments of experience – no matter how conceptually powerful or poetically evocative our words may be.

So, phenomenological research or inquiry is interested in recovering somehow the living moment of the now – even before we put language to it or describe it in words. At the same time, phenomenology tries to show how our words, concepts, and theories always shape (distort) and give structure to our experiences as we live them. For example, it is one thing to get lost in a novel but it is another to retrospectively capture what happened to us, just now, as we slipped into this textual space and began to dwell in the story. Similarly, we may identify and rate with empirical descriptors the nature and intensity of various forms of pain, but the actual experience of pain somehow seems to be beyond words.

Phenomenological Traditions

From a historical and philosophical perspective, phenomenology consists of a complex and diversified web of traditions starting well before Hegel in Germany, spreading all over Europe and eventually branching out to North America, Asia, and other continents and countries. Often these traditions are strongly associated with renowned phenomenological scholars. Here we mention a few of these traditions and briefly focus on selected terminology and the implications for research methods.

Transcendental phenomenology is the name of the tradition that begins with Husserl. Transcendence refers to the realization that we can never see a thing from all sides or perspectives at once, so the full essence of a thing can only be appropriated in transcendental or pure consciousness – in some sense abstracted from the perception of the experiential world. According to Husserl, consciousness has a transcendental structure. So phenomenology orients to the way that consciousness structures or constitutes the phenomena of the world. Early Husserlian phenomenology is understood as eidetic description: determining the essential nature and acts of consciousness. Phenomenological research proceeds through transcendental reflection as practiced through the eidetic reduction (bracketing) or epoché. In the transcendental reduction the researcher withdraws from the natural attitude of the taken-for-grantedness of the everyday world and of objective science and turns toward the intersubjective level of the transcendental ego. Husserl stresses that the phenomena (persons, things, objects, events, ideas, etc.) of which we are conscious are not simply in consciousness (as in a box), rather they are constituted as being what they are for us and as what they mean for us. Therefore, transcendental phenomenology could also be called constitutive phenomenology. It should be added that for Husserl, phenomenological intuition (the imaginative variation of the eidetic reduction) grasps the essence of things with all the vagueness that belongs (essentially) to them, and these are necessarily inexact essences. Only mathematical essences are exact.

Ontological phenomenology inquires into the nature of human existence or modes-of-being in the world. Heidegger distanced himself from the Husserlian preoccupation with *eidetics*, consciousness and intentionality, in favor of an ontological and hermeneutic perspective. In *Being and Time*, Heidegger argues that human existence (*Dasein*) is always already embedded in a world of meanings. Things are not first of all phenomena that are constituted in consciousness; rather we encounter them immediately in the world where we use them. In his early work Heidegger shows through a variety of concrete topics how phenomenology captures in language how things such as a jug, a bridge, a tool, show themselves. In his later writing Heidegger uses an increasingly evocative and poetic

discourse. Phenomenology is the study of what shows itself in the unique manner that it shows itself to us. Every mode of being (such as the mode of being a student, a teacher, a reader, a scientist, a parent) is always simultaneously a way of understanding the world. These modes of being in the world need to be interpreted. So, ontological phenomenology becomes hermeneutical when its method is taken to be interpretive, rather than purely descriptive as in transcendental phenomenology. But the contrast between descriptive and interpretive phenomenology is sometimes over-simplified by researchers in the professional disciplines. Heidegger says that all description is always already interpretation. Every form of human awareness is interpretive. Hans-Georg Gadamer, Heidegger's student, continues the development of a hermeneutic phenomenology, especially in his famous text, *Truth and Method*. Although Heidegger and Gadamer do not offer a method for conducting phenomenological inquiry, their works are examples both in their form and content. In *Truth and Method*, Gadamer (1975) carefully explores the role of language, the nature of questioning, the phenomenology of human conversation, and the significance of prejudice, historicity, and tradition in the project of human understanding. All these topics have relevance for educational researchers.

The works of Jean-Paul Sartre, Maurice Blanchot, Simone de Beauvoir, and Merleau-Ponty are known also as different forms of existential phenomenology. The relevance of existential phenomenology for education lies in its focus on the embodied, linguistic, gendered, and intersubjective dimensions of human existence. In his preface to the *Phenomenology of Perception*, Merleau-Ponty (1962) asks, "What is phenomenology?" and suggests that phenomenology begins in awakening and describing the basic experience of the world. Merleau-Ponty's clarification of the notions of intentionality, the reduction, wonder, and phenomenological reflection should be mandatory reading for those interested in developing more pragmatic approaches to phenomenological research.

In the post-structuralist writings of Jacques Derrida and his French colleagues such as Julia Kristeva and Hélène Cixous, we can address a radical linguistic phenomenology. Derrida aims to show that meaning is always primarily linguistic. Meaning resides in language and the text rather than in the subject. His famous claim that there is "nothing outside of the text" illustrates this well. For Derrida intersubjectivity is intertextuality. In contrast to Husserl's search for an indubitable ground of human understanding in the cogito, Derrida points out the essentially unstable and undecidable character of the nature of signs and meaning. Through the method of deconstruction Derrida aims to demonstrate, not the invariance (essence) of human phenomena but the essential variance, the *différance* that destabilizes all meaningful distinctions and discernable identities.

Ethical phenomenology originates with Max Scheler, a contemporary of Husserl. It also finds its origin in Jean-Paul Sartre's concern with ethical themes of freedom, responsibility, and choice. However, ethical phenomenology is especially associated with the original and influential work of Emmanuel Levinas intending to radicalize the thinking of Husserl and Heidegger into a phenomenology of otherness. For Levinas, the Husserlian focus on the essence of things and Heidegger's preoccupation with the modalities of being in the world all are manifestations of the primacy of the self or mine-ness in traditional philosophical phenomenology. For a truly profound understanding of the human reality, one must not ask for the meaning of being, self, or presence but for the meaning of what is otherwise than being, alterity, or difference. Levinas finds the phenomenological power of this question in the encounter with the face of the other who makes an appeal on us. In the vulnerability of the face of the other we experience an appeal: we are being called, addressed. Our response to the vulnerability of the other is experienced as a responsibility. This is an ethical experience, an ethical phenomenology.

Phenomenology as a Human Science

Next to these philosophic traditions, phenomenology became popular as a qualitative research approach in the social sciences. Although ethnography and ethnomethodology are characterized by their own distinct methodologies and epistemological assumptions, these disciplines prepared the way for a reception of phenomenological inquiry in North America. Ethnography offered ways of examining how subjects construct their own meanings and cultural reality; ethnomethodology enabled the social sciences to study the practices of everyday life and the meanings associated with those practices. At first, the rise of these qualitative approaches encountered considerable opposition. The root of the commotion was the challenge to traditional social science regarding their taken-for-granted assumptions about everyday life.

It is especially through disciplines such as education, pedagogy, and psychology that phenomenology was introduced in the practical, applied, or professional fields. For example, in psychology there was the influence of the Duquesne University scholars such as psychologist Adrian van Kaam and social psychologist Rolf von Eckartsberg. The Duquesne School became especially known for publishing qualitative methodological explications of phenomenological research that lent itself for application to the more practical fields of counseling and clinical psychology.

Before there was any significant interest in phenomenology in North America, a unique experiment had taken place in the Netherlands, Belgium, and Germany. For

example, the University of Utrecht School – a loosely associated group of phenomenologically oriented psychologists, educators, pedagogs, pediatricians, sociologists, criminologists, jurists, psychiatrists, and other medical doctors – can be considered a genuinely original contribution to the international discussion about phenomenology as a research perspective in the professions. Among these, scholars such as J. H. van den Berg wrote about the changing nature of childhood; the pedagogue-philosopher O. F. F. Bollnow wrote on the pedagogical atmosphere; the educator M. J. Langeveld established the field of phenomenological pedagogy.

In recent years, further developments in phenomenological methodology, inspired by continental scholars, are found in all the major professional disciplines. Practically oriented explications of phenomenological research methods are found in psychology through the efforts of Amadeo Georgi and Clark E. Moustakas, and in education through the writings of Max van Manen. Early phenomenological authors in education carved their own unique paths through the complexity of strands and traditions. Maxine Greene focuses especially on the writing of Hannah Arendt. Donald Vandenberg orients to the Husserlian philosopher Stephen Strasser and the educational thoughts of O.F. Bollnow. Madeleine Grumet and William Pinar experiment with gender based and post-structuralist phenomenologies. More recently, there is a growing interest in the pedagogical import of the phenomenology of technology and media, such as in Catherine Adams' study of PowerPoint and the pedagogy of digital media technology, and Iain Thomson's critique of the culture of technology and modern university education in light of Heidegger.

The Method of the Reduction

The idea of the reduction occupies a special place in phenomenological method. Reduction is the technical term that describes the phenomenological device which permits us to discover the experiential surge of the lifeworld. The aim of the reduction is to reach a direct and primal contact with the world as we experience it rather than as we conceptualize it. But the discovery of the prereflective lifeworld through the technique of the reduction always transcends the lifeworld: when we bracket lived experience we experience meaning. The method of reduction is meant to bring the aspects of meaning that belong to the phenomena of our lifeworld into nearness. In particular, it aims to bring into focus the uniqueness of the particular phenomenon to which we are oriented.

There exist many philosophical investigations and explications of the reduction that can make this topic complex and confusing. This is not surprising in view of

the fact that the project of phenomenology can be understood in a variety of ways. Below several levels or dimensions of the reduction are distinguished for their eclectic value and methodological usefulness: wonder or heuristic reduction, openness or hermeneutic reduction, concreteness or experiential reduction, and universality in contingency or eidetic reduction. Each of these dimensions of the reduction needs to be practiced as if in concert. Yet it may be helpful to deal with them separately while keeping the integrity of the larger phenomenological project in view.

At the most basic level, the heuristic reduction consists of the attitude or mood of wonder in the face of the world. What does this mean? It implies an approach that can shatter the taken-for-grantedness of our everyday reality. Wonder is the unwilling willingness to meet what is utterly strange in what is most familiar. To wonder is to step back and let things speak to us, a radical passive receptivity to let the things of the world present themselves in their own terms. When we are struck with wonder we seem to have evaporated momentarily our present preoccupations. We are suddenly struck by the strangeness of this thing, this phenomenon. Perhaps it is strange to speak of wonder as a method. But if we understand method as *methodos*, as path or way, then we may indeed consider wonder an important motive in human science inquiry. The way to knowledge and understanding begins in wonder. So methodologically the heuristic reduction requires discovering the miraculous moment of wonder; and in this moment a question may emerge that addresses us. The heuristic reduction involves the awakening of a profound sense of amazement at the mysteriousness of the belief in the world. This fundamental amazement may animate one's questioning of the meaning of the lived experience of the world. In terms of the particular research project in which one is engaged the heuristic reduction challenges the researcher to be receptive and awakened to a profound sense of wonder. But it also challenges the researcher to write in such a way that the reader of the phenomenological text is similarly stirred to the same sense of wondering attentiveness to the topic under investigation. Phenomenological inquiry continually edifies a wondering attitude of attentiveness.

At the level of the hermeneutic reduction, the phenomenologist needs to reflect on his or her own preunderstandings, frameworks, and biases regarding the psychological, political, and ideological motivation and the nature of the question. This is a search for genuine openness to engage in a conversational relation with the phenomenon. On the one hand, this means that one needs to practice a critical self-awareness with respect to the assumptions that prevent one from being as open as possible to the sense and significance of the phenomenon. The researcher needs to forget as it were vested interests and preunderstandings. On the other hand, it means that one needs to realize that forgetting

one's preunderstandings is not really possible and, therefore, these various assumptions and interests may need to be explicated so as to exorcise them in an attempt to let speak that what wishes to speak. Practically, the hermeneutic reduction consists of reflectively examining and turning over in one's textual labor the various preunderstandings that seem to impinge on the reflective gaze. This does not mean that one must hope to arrive at some pure vantage point, as if such a pure gaze were possible. But it requires that the various dimensions of lived meaning of the selected human experience are investigated for their different sources and layers of meaning, rather than being overlaid with a particular frame of meaning. Phenomenological inquiry is continually open to questioning assumptions and preunderstanding.

The experiential reduction requires that one avoids abstraction, theorizing, and generalization. Indeed, for any research project one must examine the available theories and discuss the body of knowledge about the topic. Theories need to be reviewed for how they inform (but fail) concreteness. Many theories contain some phenomenological material, or they are built on certain intuitions that presume phenomenological understandings. In the phenomenological reduction one needs to strip away the theoretical or scientific conceptions and thematizations that overlay the phenomenon one wishes to study, and that prevent one from seeing the phenomenon in a non-abstracting manner. The way in which to bracket theoretical meaning is not to ignore it but to examine it for possibilities of extracting phenomenological sensibilities. It is helpful to examine how the theories of conceptualizations gloss or hide the experiential reality upon which they ultimately must be based. Theories tend to explain phenomena that are not really understood in a lived or concrete sense. So one must ask: how is this topic actually experienced? What are examples of possible incidents or events? Phenomenological inquiry is continually oriented to the beginning, to the concrete, to experience as lived.

The eidetic reduction is the most central to the phenomenological method, especially for those traditions that borrow from Husserl. The researcher asks: What makes this experience uniquely different from other related experiences? In the eidetic reduction one needs to see past or through the particularity of lived experience toward the essence or *eidōs* that lies on the other side of the concreteness of lived meaning. The idea of phenomenological essence or *eidōs* does not refer to some immutable universal or generalization about human nature of human life. This would be committing the fallacy of essentialism. Phenomenological inquiry is only concerned with possible human experiences – not with experiences that are presumed to be universal or shared by all humans irrespective of time, culture, gender, or other circumstance. In addition, phenomenological determination of meaning is allusively and ultimately always indeterminate, always tentative, always

incomplete, always inclined to question assumptions by returning again and again to lived experience itself, the beginnings of phenomenological inquiry. The eidetic reduction is partially accomplished by comparing the phenomenon with other related but different phenomena. For example, in exploring the phenomenology of secrecy one would practice the eidetic technique of variation in imagination. How is the experience of secrecy different from the experience of privacy or the experience of reserve? What makes keeping a secret different from lying? Are there different kinds of secrecy? How is keeping a secret different from lying? What are concrete examples of this experience? and so forth.

In the eidetic reduction, patterns of meaning or themes seem to emerge. These are not themes in the sense of theoretical or conceptual abstractions; they do not belong to existing theories, taxonomies, genres, paradigms, philosophies, or conceptual frameworks. Phenomenological themes are the working material for phenomenological writing. The eidetic reduction differs from concept analysis in that the reduction does not claim to clarify linguistically the boundaries of a phenomenon or how a concept is being used in different contexts. Rather, the reduction attempts to offer intimations of meaningfulness. The eidetic reduction asks: Does this piece of text bring the experience into view? Does this phrase resonate with our prereflective sensibilities? Are these portrayals of lived meaning recognizable? Do they evoke something unique about this human experience? The eidetic reduction is not a simplification, fixation, or contraction of the world into a system of eidetic concepts – rather it is the exact opposite: the eidetic reduction makes the world appear as it precedes every cognitive construction: in its full ambiguity, irreducibility, contingency, mystery, and ultimate indeterminacy.

Human Science: Empirical and Reflective Methods and Procedures

The reduction is the method that is central to the phenomenological study of the lifeworld. But as phenomenology is adopted by other disciplines, empirical and reflective methods are imported that are derived from the humanities and the social sciences. Empirical methods such as interviewing, observation, eliciting written descriptions, and borrowing from literary and artistic sources are now used to gather experiential material. Reflective methods of thematization, etymological analysis, and meaning analysis are used to facilitate the challenges involved in the various forms of the reduction.

Our personal life experiences are immediately accessible to us in a way that no one else's are. However, the phenomenologist does not want to trouble the reader with purely private, autobiographical facticities of one's life.

In drawing on personal descriptions of lived experiences, the phenomenologist knows that the patterns of meaning of one's own experiences are also the possible experiences of others, and therefore may be recognizable by others. To conduct a personal description of a lived experience, the researcher aims to describe a phenomenon (lived experience) as much as possible in experiential terms. The focus is on direct description of a particular situation or event as it is lived through, without offering causal explanations or interpretive generalizations. It is to the extent that a personal experience can be recognized by others that the phenomenologist wants to be reflectively aware of certain experiential meanings.

To gain access to the experience of young children, it may be important to play with them, talk with them, puppeteer, paint, draw, follow them into their play spaces and into the things they do while the researcher remains attentively aware of the way it is for children. Participatory activities and close observation may generate different forms of experiential material than is obtained through the written or the interview approach. Observational method may require that one be a participant and an observer at the same time, that one maintains an orientation of reflectivity while guarding against the more manipulative and artificial attitude that a reflective attitude tends to insert in a social situation and relation. Literature, such as novels and short stories, are sometimes excellent sources for experiential material. The phenomenological value of a novel, for example, is determined by what may be called the perceptiveness and the intuitive sensitivity of the author. Through a powerful novel one is given the chance of living through an experience that provides the opportunity of gaining insight into certain aspects of the human condition.

Whereas empirical methods aim to explore the range and varieties of prereflective experiential material that is appropriate for the phenomenon under study, reflective methods aim to interpret the aspects of meaning or meaningfulness that are associated with the phenomenon and that assist with the reduction. Researchers commonly use the device of thematization to explore the qualitative dimensions of a phenomenon. For example, when we are interested in the phenomenology of reading a novel, we may soon notice some possible themes: (1) When we open a book we experience this wondrous sensation that this thing-like object, the book, can draw us into the other-worldly space of the text. (2) When we begin to read a book, we enter it, as it were. (3) Reading a novel means that we begin to care for the people who make up the novel. (4) While we read a story we experience action without having to act ourselves. (5) When we interrupt a book, we exit the world created by the word, etc. These kinds of themes are only fasteners or foci around which a web of phenomenological descriptions of the experience of reading a novel can be constructed.

Ultimately the concept of theme is itself primarily of heuristic importance. It may be considered simply as a means to get at the phenomenon that the researcher is addressing. Thematic reflection can provide a measure of control and a sense of order in our research and writing. Our lived experiences and the structures of meanings (themes) in terms of which these lived experiences can be described and interpreted constitute the immense complexity of the lifeworld. We can even speak of the multiple and different lifeworlds that belong to different human existences and realities. For example, the lifeworld of the child has different experiential qualities from the lifeworld of the adult. There are the lifeworlds of the elderly, the sick, the man, the woman, the researcher, and so forth. Each of us may be seen to inhabit different lifeworlds at different times of the day, such as the lived world of school and the lived world of the home.

One can also speak of existential themes that pervade the lifeworlds of all human beings, regardless of their historical, cultural, or social situatedness. Existential themes that may prove especially helpful as guides for reflection in the research process are lived space (spatiality), lived body (corporeality), lived time (temporality), and lived human relation (relationality). They are productive categories for the process of phenomenological questioning, reflecting, and writing.

See also: Action Research in Education; Ethnography; Ethnomethodology in Education Research; Hermeneutics; Interpretive Research; Interviews and Interviewing; Life History; Narrative Inquiry.

Bibliography

- Gadamer, H. -G. (1975). *Truth and Method*. New York: Seabury.
Merleau-Ponty, M. (1962). *Phenomenology of Perception*. London: Routledge & Kegan Paul.

Further Reading

- Arendt, H. (1958). *The Human Condition*. Chicago, IL: The University of Chicago Press.
Benner, P. (ed.) (1994). *Interpretive Phenomenology*. Thousand Oaks, CA: Sage.
Derrida, J. (1976). *Of Grammatology*. Baltimore, MD: Johns Hopkins University Press.
Giorgi, A. (ed.) (1985). Sketch of a psychological phenomenological method. In *Phenomenology and Psychological Research*, pp 8–22. Pittsburgh, PA: Duquesne University Press.
Greene, M. (1973). *Teacher as Stranger*. New York: Wadsworth.
Greene, M. (2000). *Releasing the Imagination: Essays on Education, the Arts, and Social Change*. Hoboken, NJ: Wiley.
Heidegger, M. (1962). *Being and Time*. New York: Harper & Row.
Heidegger, M. (1977). *Basic Writings*. New York: Harper & Row.
Husserl, E. (1970). *The Crisis of European Sciences and Transcendental Phenomenology*. Evanston: Northwestern University Press.

- Levinas, E. (1969). *Totality and Infinity: An Essay on Exteriority*. Pittsburgh, PA: Duquesne University Press.
- Lingis, A. (1998). *The Imperative*. Bloomington, IN: Indiana University Press.
- Løgstrup, K. E. (1997). *The Ethical Demand*. Notre Dame: University of Notre Dame Press.
- Moustakas, C. (1994). *Phenomenological Research Methods*. Thousand Oaks, CA: Sage.
- Sartre, J. P. (1956). *Being and Nothingness*. New York: Philosophical Library.
- Scheler, M. (1973). *Formalism in Ethics and Non-Formal Ethics of Values*. Evanston, IL: Northwestern University Press.
- Sokolowski, R. (2000). *Introduction to Phenomenology*. Cambridge: Cambridge University Press.
- Strasser, S. (1963). *Phenomenology and the Human Sciences*. Pittsburgh, PA: Duquesne University Press.
- Thomson, I. (2005). *Heidegger on Ontotheology: Technology and the Politics of Education*. Cambridge: Cambridge University Press.
- Toombs, S. K. (ed.) (2001). *Handbook of Phenomenology and Medicine. Philosophy and Medicine Series*. Dordrecht: Kluwer.
- Tymieniecka, T. (ed.) (2002). *Phenomenology World-Wide*. Dordrecht: Kluwer.
- Vandenberg, D. (1971). *Being and Education*. Englewood Cliffs, NJ: Prentice-Hall.
- van Manen, M. (1991). *The Tact of Teaching: The Meaning of Pedagogical Thoughtfulness*. London, ON: Althouse Press.
- van Manen, M. (1997). *Researching Lived Experience: Human Science for an Action Sensitive Pedagogy*. London, ON/Albany, NY: Althouse Press/SUNY Press.

Qualitative Case Studies

S B Merriam, University of Georgia, Athens, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

Case studies can be found in all subfields of education and most of us have encountered them in our professional training and in our practice. However, despite the prevalence of case studies in our literature, there is still some confusion as to exactly what a case study is. The fact that a lawyer, a social worker, an economist, a medical doctor, and even a detective can be involved in research on a case further confounds the issue as to what constitutes case study research. Until the late 1970s, the only time textbooks on research design included a chapter on case study was as a catchall for any study not statistical or experimental (such as historical, ethnographic, and biographic). Even today, as Stake (2006: 8) observes, “here and there, researchers will call anything they please a case study.” This article is intended to clearly define what a case study is and, in particular, what a qualitative case is. To that end, this article is organized into the following sections: definition and characteristics of a qualitative case study, design and implementation of a qualitative case study, and writing up a qualitative case study.

Definition and Characteristics

What makes a case study a case study is the unit of analysis; that is, a case study is an in-depth description and analysis of a bounded system. While some authors define a case study as the process of conducting a case study (Yin, 2003) and/or the end product of an investigation (Wolcott, 1992), we have concluded that it is the unit of analysis that makes something a case study. By unit of analysis we mean a bounded system (Smith, 1978), a single entity, a unit that is selected for study around which there are boundaries. We can fence in what we are going to study. The case then, could be a single person, a program, a group, an institution, a community, or a specific policy (see, e.g., Bennett’s (2002) case study of Indiana University’s project to increase ethnic diversity in their teacher education program or Koosimile’s (2002) study of a new science curriculum in Botswana). Miles and Huberman (1994) think of the case as “a phenomenon of some sort occurring in a bounded context” (p. 25). They graphically present it as a circle with a heart in the center. The heart is the focus of the study, while the circle “defines the edge of the case: what will not be studied” (p. 25).

The unit of analysis, not the topic of investigation, defines a case study. For example, a study of how teens negotiate online social relationships would likely be a qualitative study but not a case study; the unit of analysis would

be the teens’ experiences, and there could be an indefinite number of teenagers and their experiences negotiating online relationships selected for the study. In other words, the sample is without boundaries. If there is no end, actually or theoretically, to the number of people who could be interviewed or to observations that could be conducted, then the phenomenon is not bounded enough to qualify as a case. For it to be a case study, one particular Internet social site or one particular teenager selected on the basis of typicality, uniqueness, success, etc., would be the unit of analysis. Stake (2006: 1) explains:

A case is a noun, a thing, an entity; it is seldom a verb, a participle, a functioning. Schools may be our cases – real things that are easy to visualize. . . . Training modules may be our cases – amorphous and abstract, but still things, whereas “training” is not. Nurses may be our cases; we usually do not define “nursing activity” as the case. “Managing,” “becoming effective,” “giving birth,” and “voting” are examples of functioning, not entities we are likely to identify as cases. For our cases, we may select “mangers,” “production sites,” “labor and delivery rooms,” or “training sessions for voters.” With these cases we find opportunities to examine functioning, but the functioning is not the case.

The bounded system is of course embedded in the larger sociohistorical context in which it exists. As Geertz (1996: 262) points out, “No one lives in the world in general. Everybody, even the exiled, the drifting, the diasporic, or the perpetually moving, lives in some confined and limited stretch of it – ‘the world around here.’” Accounting for this context is one of the strengths of case study research.

The Qualitative Case Study

Certainly an in-depth study of a bounded system can employ a variety of quantitative and/or qualitative data. For example, a case study of a community could be based solely on demographic and economic data and thus be considered quantitative. Case studies might also include a mix of quantitative and qualitative data as is often the situation in evaluation case studies. An evaluation of a hospital’s disaster readiness and response, for example, might include timed responses, mapping of personnel deployment, interviews with supervisors, and on-the-ground observations. In keeping with the focus of this section on qualitative research methods, this article is devoted to case studies that are qualitative.

The question then becomes what makes a case study qualitative? Qualitative case studies share several characteristics with other forms of qualitative research. Qualitative researchers are interested in understanding how people interpret their experiences, how they construct their worlds, and what meaning they attribute to their experiences. Drawing from interpretive and constructivist notions of reality, it is assumed that reality is socially constructed, that is, there is no single, observable reality. Rather, there are multiple realities, or interpretations, of a single event. Researchers do not find knowledge; rather, they construct it. Constructivism is a term often used interchangeably with interpretivism. Cresswell (2007) explains:

In this worldview, individuals seek understanding of the world in which they live and work. They develop subjective meanings of their experiences. . . . These meanings are varied and multiple, leading the researcher to look for the complexity of views. . . . Often these subjective meanings. . . are formed through interaction with others (hence social constructivism) and through historical and cultural norms that operate in individuals' lives. (pp. 20, 21)

For example, rather than finding out what types of jobs college graduates get and if there is any correlation with their major, one could investigate the transition from full-time study to full-time work – what was the process, how were obstacles negotiated, what are the changes in daily activities, how have the historical and economic contexts shaped the process, and so on. These questions are about understanding their experiences and would call for a qualitative design. The overall purposes of qualitative research are to achieve an understanding of how people make sense out of their lives, to delineate the process (rather than the outcome or product) of meaning-making, and to describe how people interpret what they experience. These constructions or interpretations are context dependent with context broadly defined to include social, historical, economic, and cultural dimensions, all of which intersect at any particular point in time.

A second characteristic of all forms of qualitative research is that the researcher is the primary instrument for data collection and analysis. Since understanding is the goal of this research, the human instrument, which is able to be immediately responsive and adaptive, would seem to be the ideal means of collecting and analyzing data. Other advantages are that the researcher can expand his or her understanding through nonverbal as well as verbal communication, process information (data) immediately, clarify and summarize material, check with respondents for accuracy of interpretation, and explore unusual or unanticipated responses.

Often qualitative researchers undertake a qualitative study because there is a lack of theory or an existing theory fails to adequately explain a phenomenon. Therefore,

another important characteristic of qualitative research is that the process is inductive; that is, researchers gather data to build concepts, hypotheses, or theories rather than deductively testing hypotheses as in positivist research. Qualitative researchers build toward theory from observations and intuitive understandings gleaned from being in the field. Bits and pieces of information from interviews, observations, and/or documents are combined and ordered into larger themes as the researcher works from the particular to the general. Typically, findings inductively derived from the data in a qualitative study are in the form of themes, categories, typologies, concepts, tentative hypotheses, and even theory about a particular aspect of practice.

Finally, the product of a qualitative inquiry is richly descriptive. Words and pictures, rather than numbers, are used to convey what the researcher has learned about a phenomenon. There are likely to be descriptions of the context, the participants involved, and the activities of interest. In addition, data in the form of quotes from documents, field notes, and participant interviews, excerpts from videotapes, electronic communication, photos, or a combination of these are always included in support of the findings of the study. These quotes and excerpts contribute to the descriptive nature of qualitative research.

A qualitative case study then, is an in-depth study of a bounded system in which meaning and understanding of the phenomenon of interest are sought. The case study researcher is the primary instrument of data collection and analysis, findings are inductively derived from the data, and the end product is richly descriptive.

Types of Case Studies

Bogdan and Biklen (2007) differentiate among historical organizational case studies, observational case studies, and life histories. The first type, the historical organizational case study, is exactly what the name implies – it is a study of the development of a particular organization over time. The key to historical case studies, organizational or otherwise, is the notion of investigating the phenomenon over a period of time. The researcher still presents a holistic description and analysis of a specific phenomenon (the case) but presents it from a historical perspective.

Historical research is essentially descriptive, and elements of historical research and case study often merge. Yin (2003) discusses the two approaches:

The case study is preferred in examining contemporary events, but when the relevant behaviors cannot be manipulated. The case study relies on many of the same techniques as a history, but it adds two sources of evidence not usually included in the historian's repertoire: direct observation of the events being studied and interviews of the persons

involved in the events. Again, although case studies and histories can overlap, the case study's unique strength is its ability to deal with a full variety of evidence – documents, artifacts, interviews, and observations – beyond what might be available in the conventional historical study. (pp. 7, 8)

An observational case study is one in which “the major data-gathering technique is participant observation (supplemented with formal and informal interviews and review of documents) and the focus of the study is on a particular organization (school, rehabilitation center) or some aspect of the organization” (Bogdan and Biklen, 2007: 60). Using the workplace as an example, an observational case study could focus on a particular place in an organization (e.g., the director's office, or the staff break room), a specific group of people (e.g., the evening shift, or mid-level managers), or a particular activity (e.g., implementing a new computer system, or team meetings). An observational case study can be exploratory, descriptive or explanatory (Yin, 2003). The third type of case study described by Bogdan and Biklen is the life history. Here “the researcher conducts extensive interviews with one person for the purpose of collecting a first-person narrative” (p. 63). This type of case study goes by several names, such as life story, biographical case study, and portraiture (Patton, 2002).

Yet another typology of types of case studies is the one by Stake (2005). He identifies three types of case studies, differentiated by the researcher's interest – intrinsic, instrumental, and collective. The intrinsic case study is undertaken when the researcher is interested in the particular case itself – it is intrinsically interesting. “The purpose is not to come to understand some abstract construct or generic phenomenon;” nor is the purpose “theory building.” Rather, the “study is undertaken because of an intrinsic interest in, for example, this particular child, clinic, conference, or curriculum” (p. 445). An instrumental case study, on the other hand, “is examined mainly to provide insight into an issue or to redraw a generalization. The case is of secondary interest, it plays a supportive role, and it facilitates our understanding of something else” (p. 437). Finally, in a collective or multiple case study, a number of cases are studied “to investigate a phenomenon, population, or general condition” (p. 445).

Finally, several writers point to case study's usefulness in evaluation research (LeCompte *et al.*, 1993; Patton, 1996, 2002). Evaluative case studies involve description, explanation, and judgment. Yin (2003) has identified several different applications of case studies in evaluation research with the most important being “to *explain* the presumed causal links in real-life interventions that are too complex for the survey or experimental strategies” (p. 15, emphasis in original).

Design and Implementation of a Qualitative Case Study

As stated above, a case study is an in-depth analysis of a bounded system. For a case study to be qualitative, the focus is on understanding a phenomenon from the perspectives of those in the case. How do they make meaning of what is going on? How do they interpret their experiences? How is the context an important component in this understanding? In designing a qualitative case study, one must first identify a problem. A research problem is actually a gap in our knowledge and understanding of some phenomenon. The research problem is always lodged in the literature on the topic and forms the theoretical framework for one's study (Merriam, 2006). From previous scholarship in an area, we can identify models, theories, frameworks, and we can find out what we do know. When we identify something we do not know, there's a gap in our knowledge. It is this gap that one's study is designed to address. For example, there is a voluminous literature on math anxiety, including hundreds of studies. Parker (1997) however, could find no research on how adults who were math anxious were able to overcome their handicap and become comfortable with math tasks in their daily lives. Parker's qualitative study was on the transition these adults went through from being math anxious to comfortable with math. An example of a qualitative case study is Corvey's (2003) study based on Wenger's (1998) theory of communities of practice. His theory posits that learning is a social activity in which we collectively make meaning as we mutually engage in some activity. Further, that learning not only impacts our practice, but also changes who we are, our identity. To see if this theory held up in practice, Corvey (2003) studied an online community of practice of clinical nurses. The overall purpose of her study was to determine how Listserv membership in a community of practice of advanced practice nurses fostered both their practice and their professional identity.

Case Selection

Once the problem to be investigated is identified, the next step is to select the case. This selection is purposive or purposeful, not random. The logic of purposeful sampling is that one has to select the case that will yield information that can best address the study's purpose. Two levels of sampling are usually necessary in qualitative case studies. First, one must select the case to be studied. Then, unless he/she plans to interview, observe, or analyze all the people, activities, or documents within the case, there will be a need to do some sampling within the case.

A case is a single unit, a bounded system. As Stake (1995) points out, sometimes selecting a case turns out “to be no ‘choice’ at all . . . It happens when a teacher decides

to study a student having difficulty, when we get curious about a particular agency, or when we take the responsibility of evaluating a program. The case is given” (p. 3). Other times, we have a general question, an issue, or a problem that we are interested in, and we feel that an in-depth study of a particular instance or case will illuminate that interest.

To find the best case to study, one would first establish the criteria that will guide case selection and then select a case that meets those criteria. For example, if one’s interest is in schools that appear to be successful in integrating a multilingual, multicultural student body, he/she would establish criteria for what constitutes a successful program; then one would select a school that meets those criteria. This particular school would be the case. For multicase or comparative case studies one would select several cases based on relevant criteria. One of the criteria might be that one would want as much variation as possible; hence, one would be employing a maximum-variation sampling strategy in the selection of cases. Using the above-mentioned example, one might seek out successful schools representing a range of socioeconomic neighborhoods or that address a wide range of linguistic or cultural diversity.

Thus, the researcher first identifies the case – the bounded system, the unit of analysis – to be investigated. The case can be as varied as a second-grade classroom, the training department of a company, a system-wide model science program, or a patient education clinic at a local hospital. Within every case, there exist numerous sites that could be visited (as in the model science program), events or activities that could be observed, people who could be interviewed, and documents that could be read. A sample within the case needs to be selected either before the data collection begins or while the data are being gathered (ongoing or theoretical sampling). Random sampling of a particular unit, for example, teachers in a school can be used within the case. More commonly, however, purposeful sampling as outlined earlier is used to select the sample within the case, just as it is used to select the case itself. However, a second set of criteria is usually needed to purposefully select whom to interview, what to observe, and which documents to analyze.

Data Collection and Analysis

Since a case study is an in-depth description and analysis of a bounded system, rich and detailed data are crucial to being able to describe the case and answer the research questions. For most qualitative case studies, data are collected through interviewing, observations, and document analysis. Case-study researchers are encouraged to use all three methods of data collection where possible, as multiple sources of data enhance the validity of the findings. That is, if we as researchers hear people tell us about the

phenomenon, if we also observe it, and if we read about it in supporting documents, we can feel some confidence that we are capturing their understandings/interpretations of the phenomenon as closely as possible.

Interviews range from highly structured, where specific questions and the order in which they are asked are determined ahead of time, to unstructured, where one has topic areas to explore but neither the questions nor the order are predetermined. Most interviews fall somewhere in between. The semi-structured interview contains a mix of more and less structured questions. Usually, specific information is desired from all the participants; this forms the highly structured section of the interview. The largest part of the interview is guided by a list of questions or issues to be explored, and neither the exact wording nor the order of the questions is determined ahead of time. The quality of the data obtained through interviews is dependent upon the questions asked. Fewer, more open-ended questions, such as “Tell me about a time when. . .” or “How would you describe that. . .” or “How did you feel about that. . .” will generate the best data.

Data collected through observations represent a first-hand encounter with the case rather than a secondhand account obtained in an interview. Observation is the best technique when an activity, event, or situation can be observed firsthand, when a fresh perspective is desired, or when participants are not able or willing to discuss the phenomenon under study. Like interviewing, there is a range here also from being a complete observer to being an active participant. A complete observer is unknown to those being observed and a very active participant observer might be someone who is a member of the group or organization and is thus participating while observing. In most case studies, the observer is known to the participants and is minimally, if at all, involved in the activity being observed. In some case studies, somewhat informal interviews are conducted in conjunction with observations. This joint data collection strategy is labeled fieldwork or field study.

Since case studies are of a single bounded unit, such as an individual, an event, or an organization, what is loosely called documents are invaluable sources of information and insight about the case. These documents can be written, oral, visual (such as photographs), or cultural artifacts. Public records, personal documents, and physical material are other types of documents. The strength of documents as a data source lies with the fact that they already exist in the situation; they do not intrude upon or alter the setting in ways that the presence of the investigator might. Nor are they dependent upon whims of human beings whose cooperation is essential for collecting data through interviews and observations.

Interviews, observations, and documents are standard sources of data in qualitative case study research. Data can

also be collected online. Web pages, papers available online, etc. can be considered documents; artifacts in the form of illustrations and games can be downloaded; interviews can be conducted by e-mail; and researchers can observe online chat rooms. Correll's (1995) ethnographic case study of an online lesbian community is a good example of data collection through this medium. She observed the online postings and discussion of the community and conducted interviews via e-mail.

In qualitative case studies, as with all qualitative research, data analysis is best done simultaneously with data collection. Analysis should begin with the first interview, the first observation, or the first document accessed for the study. Simultaneous data collection and analysis allow the researcher to make adjustments to collect the best data. For example, once an interview transcript or field notes from an observation have been at least cursorily analyzed, it might be determined that people representing other interests need to be included or different activities should be observed. At the very least, an analysis of a transcript or set of field notes is likely to suggest other, better questions to ask, or adjustments in what is being observed. Most importantly, tentative findings or answers to one's research questions can be weighed against subsequent data. This iterative process of qualitative data analysis leads to solid and well-supported findings.

Data analysis is not only iterative but inductive. One begins with a unit of data (any word, phrase, or narrative that possibly seems to address/answer the research questions guiding the study). These data are compared to another unit of data, and so on, all the while looking for common patterns across the data. These patterns are given labels and are refined and adjusted as the analysis proceeds. In multiple case studies, one should first analyze the data within each case, then conduct an analysis across the cases. Bogdan and Biklen (2007) recommend doing fieldwork one site at a time, rather than simultaneously collecting data from several sites. "The reason for this is mainly that doing more than one site at a time can get confusing. There are too many names to remember, too much diverse data to manage. After you finish your first case, you will find that in multicase studies subsequent cases are easier. . . . the first case study will have provided a focus to define the parameters of the others" (p. 70).

Writing Up the Case

Since a case study is an in-depth description and analysis of a bounded system, the write up of a case study includes a detailed description of the case as well as the analysis of the data collected in the case. Many case-study write-ups have two parts; first, a description and, second, the findings in some form of an organizing scheme of categories, themes, concepts, or typology. Some writers present

a descriptive narrative first, followed by analysis and interpretation. This is what Lightfoot (1983) did in her well-known study of what makes a good high school. She studied six different high schools and her findings are presented first as six individual case studies (or portraits as she calls them); she then offers a cross-case analysis leading to generalizations about what constitutes a good high school. Another common organization is to integrate descriptions and vignettes with commentary. This is how Abramson (1992) wrote the case study of his Russian immigrant grandfather.

The descriptive component affords the reader the vicarious experience of having been there. "The case study," Patton (2002) writes, "should take the reader into the case situation and experience – a person's life, a group's life, or a program's life" (p. 450). Of course, in order for a reader to vicariously experience a phenomenon, the writer must transport the reader to the setting. This is done through writing a vividly descriptive narrative of the setting and the situation. These descriptions can be of any length, depending on whether they are describing the context generally, or illustrating a specific point. Detailed description of particulars is needed so that the reader can vicariously experience the setting of the study; detailed description is also necessary for the reader to assess the evidence upon which the researcher's analysis is based.

Donmoyer (1990) offers three compelling rationales for conveying the vicarious experience of a case study to the reader. First is the advantage of accessibility. "Case studies can take us to places where most of us would not have an opportunity to go" (p. 193). This does not have to refer to exotic places. Case studies allow us to experience situations and individuals in our own settings that we would not normally have access to. A second advantage to case studies is seeing through the researcher's eyes. By this, Donmoyer means that case studies may allow us to see something familiar but in new and interesting ways. The third advantage he identifies is decreased defensiveness. "Vicarious experience is less likely to produce defensiveness and resistance to learning" (p. 196). People can learn from a case study, perhaps more willingly than from actual experience. "Resistance to accommodating novelty" in a case study, for example, "will not be as great as when a threat is experienced in real life" (p. 197).

Stake (1995) points out that the case report usually falls somewhere between storytelling and the traditional research report. According to Stake, the development of most case-study reports follows one of several paths: "a chronological or biographical development of the case; a researcher's view of coming to know the case"; or "description one by one of several major components of the case" (p. 127). He also offers an outline tailored to case studies, including an entry vignette, the purpose and method of the study, narrative description of the context,

key issues for understanding the case, data in support of the issues and assertions about the case, and a closing vignette. In summary, writing up a qualitative case-study report is not a lot different from writing up any report of qualitative research. In both situations, it is crucial to consider the audience, which in turn, helps determine the style and voice of the report, as well as how much attention should be devoted to each of the necessary component parts. Case studies, which have as their goal to convey understanding, must contain enough description to provide a vicarious experience for the reader.

Generalizing from Qualitative Case Studies

The worth of qualitative case studies for contributing to the knowledge base of a field and to improving practice has been challenged over the years. Most recently, the federal government's preference for evidence-based research, that is, research that is experimental or quasi-experimental, has led to some questioning of all forms of qualitative research. However, those who conduct research in education for the most part fully appreciate what can be learned from qualitative research and qualitative case studies. In a recent presentation critiquing the new gold standard of randomized controlled trials in educational research, Shields (2007) argues for qualitative case studies: "The strength of qualitative approaches is that they account for and include difference – ideologically, epistemologically, methodologically – and most importantly, humanly. They do not attempt to eliminate what cannot be discounted. They do not attempt to simplify what cannot be simplified. Thus, it is precisely because case study includes paradoxes and acknowledges that there are no simple answers, that it can and should qualify as the gold standard" (p. 13).

Education is a social science that involves real people in real time and an in-depth snapshot of some aspect of educational practice can be enormously instructive to others in the field. Such studies can also lead to models or theories that inspire even more research. For example, Piaget's theory of cognitive development, which has inspired literally thousands of studies over the years, was derived from case studies of his two children. Moreover, in citing single cases, experiments, and the experiences of Galileo, Newton, Einstein, Bohr, Darwin, Marx, and Freud, Flyvbjerg (2004) makes the point that both human and natural sciences can be advanced by a single case. He also argues that formal generalizations based on large samples are overrated in their contribution to scientific progress.

The problem of generalizing from a qualitative case study is a problem only if one thinks from a positivist understanding of knowledge construction and relies on

the statistical notion of generalizing from a random sample to a population. There are other ways to think about generalizability more congruent with the constructivist philosophy underlying qualitative research. Perhaps because a case study focuses on a single unit, a single instance, the issue of generalizability looms larger here than with other types of qualitative research. However, much can be learned from a particular case. Readers can learn vicariously from an encounter with the case through the researcher's narrative description (Stake, 2005). The colorful description in a case study can create an image – "a vivid portrait of excellent teaching, for example – can become a prototype that can be used in the education of teachers or for the appraisal of teaching" (Eisner, 1991: 199). Further, Erickson (1986) argues that since the general lies in the particular, what we learn in a particular case can be transferred to similar situations. It is the reader, not the researcher, who determines what can apply to his or her context. Stake (2005: 455) explains how this knowledge transfer works: case researchers "will, like others, pass along to readers some of their personal meanings of events and relationships – and fail to pass along others. They know that the reader, too, will add and subtract, invent and shape – reconstructing the knowledge in ways that leave it...more likely to be personally useful."

Summary

The term case study is not used precisely in the literature and is still used by some as a catchall category for studies that are clearly not experimental, survey, or historical. Further, case study is sometimes used interchangeably to refer to qualitative research. In this article, we have delineated the nature of qualitative case studies. We begin with the definition of a case study as an in-depth description and analysis of a bounded system. Case studies are case studies because the unit of analysis is a single bounded system. When the question asked of this single bounded system or the case, is about meaning and understanding of a phenomenon from the perspectives of those within the case, we have a qualitative case study. All of the assumptions underlying qualitative research apply then to this case study. Sample selection, data collection, and data analysis are all congruent with qualitative research. Also discussed is the write up of a case-study report. The article concludes with a discussion of the issue of generalizability of qualitative case studies.

See also: Ethnography; Grounded Theory; Interpretive Research; Interviews and Interviewing; Participant Observation; Qualitative Data Management; Validity; Mapping Diverse Perspectives.

Bibliography

- Abramson, P. R. (1992). *A Case for Case Studies*. Thousand Oaks, CA: Sage.
- Bennett, C. I. (2002). Enhancing ethnic diversity at a big ten university through project TEAM: A case study in teacher education. *Educational Researcher* 31(2), 21–29.
- Bogdan, R. C. and Biklen, S. K. (2007). *Qualitative Research for Education: An Introduction to Theories and Methods*. Boston, MA: Pearson.
- Correll, S. J. (1995). The ethnography of an electronic bar: The lesbian café. *Journal of Contemporary Ethnography* 24(3), 270–298.
- Convey, R. J. (2003). *The Role of Listserv Participation in the Professional Development of a Nursing Community of Practice*. Unpublished doctoral dissertation, University of Georgia, Athens.
- Cresswell, J. W. (2007). *Qualitative Inquiry and Research Design*, 2nd edn. Thousand Oaks, CA: Sage.
- Donmoyer, R. (1990). Generalizability and the single-case study. In Eisner, E. W. and Peshkin, A. (eds.) *Qualitative Inquiry in Education: The Continuing Debate*, pp 175–200. New York: Teachers College Press.
- Eisner, E. W. (1991). *The Enlightened Eye: Qualitative Inquiry and the Enhancement of Educational Practice*. Old Tappan, NJ: Macmillan.
- Erickson, F. (1986). Qualitative methods in research on teaching. In Whittrock, M. C. (ed.) *Handbook of Research on Teaching*, 3rd edn., pp 119–161. Old Tappan, NJ: Macmillan.
- Flyvbjerg, B. (2004). Five misunderstandings about case-study research. In Seale, C., Gobo, G., Gubrium, J. F., and Silverman, D. (eds.) *Qualitative Research Practice*, pp 420–434. London: Sage.
- Geertz, C. (1996). Afterword. In Field, S. and Basso, K. H. (eds.) *Sense of Place*, pp 259–262. Santa Fe, NM: School of American Research Press.
- Koosimile, A. T. (2002). Access negotiation and curriculum change: Lessons from Botswana. *Qualitative Studies in Education* 15(2), 205–223.
- LeCompte, M. D., Preissle, J., and Tesch, R. (1993). *Ethnography and Qualitative Design in Educational Research*, 2nd edn. Orlando, FL: Academic Press.
- Lightfoot, S. L. (1983). *The Good High School*. New York: Basic Books.
- Merriam, S. B. (2006). Transformational learning and HIV-positive young adults. In Anfara, V. A. Jr. and Mertz, N. T. (eds.) *Theoretical Frameworks in Qualitative Research*, pp 23–38. Thousand Oaks, CA: Sage.
- Miles, M. B. and Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*, 2nd edn. Thousand Oaks, CA: Sage.
- Parker, S. L. B. (1997). *Overcoming Math Anxiety: Formerly Math-Anxious Adults Share Their Solutions*. Unpublished doctoral dissertation. University of Georgia, Athens, GA.
- Patton, M. Q. (1996). *Utilization-Focused Evaluation*, 3rd edn. Thousand Oaks, CA: Sage.
- Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods*, 3rd edn. Thousand Oaks, CA: Sage.
- Shields, C. M. (2007). Can case studies achieve the “Gold Standard”? Or when methodology meets politics. *Paper Presentation at the Annual Meeting of the American Educational Research Association*. Chicago, IL.
- Smith, L. M. (1978). An evolving logic of participant observation, educational ethnography and other case studies. In Shulman, L. (ed.) *Review of Research in Education*, pp 316–377. Itasca, IL: Peacock.
- Stake, R. E. (1995). *The Art of Case Study Research*. Thousand Oaks, CA: Sage.
- Stake, R. E. (2005). Qualitative case studies. In Denzin, N. K. and Lincoln, Y. S. (eds.) *The Sage Handbook of Qualitative Research*, 3rd edn, pp 443–466. Thousand Oaks, CA: Sage.
- Stake, R. E. (2006). *Multiple Case Study Analysis*. New York: Guilford Press.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. NY: Cambridge University Press.
- Wolcott, H. F. (1992). Posturing in qualitative inquiry. In LeCompte, M. D., Millroy, W. L., and Preissle, J. (eds.) *The Handbook of Qualitative Research in Education*, pp 3–52. Orlando, FL: Academic Press.
- Yin, R. K. (2003). *Case Study Research: Design and Methods*, 3rd edn. Thousand Oaks, CA: Sage.

Further Reading

- Hancock, D. R. and Algozzine, B. (2006). *Doing Case Study Research*. New York: Teachers College Press.
- Merriam, S. B. (2002). *Qualitative Research in Practice: Examples for Discussion and Analysis*. San Francisco, CA: Jossey-Bass.
- Merriam, S. B. (2009). *Qualitative Research: A Guide to Design and Implementation*. San Francisco, CA: Jossey-Bass.
- Yin, R. K. (2008). *Case Study Research: Design and Methods*, 4th edn. Thousand Oaks, CA: Sage.

Qualitative Data Management

C Grbich, Flinders University, Adelaide, SA, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Monoepestemological – A single aspect utilized to generate knowledge.

Polyepistemological – Many aspects utilized to create knowledge.

Philosophical Issues

Good organization of data in qualitative research is essential if researchers are not to end up swamped by a room full of files about which they may well have forgotten important details. Management needs to start early when the first interview/observation/document is collected. Aspects which will impact on how best to manage qualitative data during data collection lie in the following:

- differences among research designs relating to the traditions from which particular approaches emerge;
- the impact of frames which the researcher brings to data collection and analysis;
- the positions the researcher takes in the process of data collection along the subjective–objective dimension; and
- where the form of data analysis and the style of presentation places the audience – the future readers of the documents that will be produced from this research.

All these will influence decisions as to how data will be managed.

Research Styles

Taking each of these aspects in turn, researcher approaches to qualitative research slide along a continuum from subjective to iterative to investigative to enumerative and each has a different focus. Subjective approaches would favor more intact data in the autoethnographic, post-modern, phenomenological, oral, historical, and narrative traditions where intact stories, with a beginning, a middle, and an end, are of importance; iterative will overlap with subjective in terms of the process of constant feedback enriching the holistic view sought, but will often fragment data in order to reassemble them in another form, for example, as in grounded theory, action and evaluative approaches; investigative is even more selective in terms of aspects to be analyzed and has a focus on language,

particularly that of discourse, deconstruction, content, conversation, and sociolinguistic narratives; while enumerative utilizes quasi-statistical and matrix-type approaches, counting incidences and applying predesigned coding formula.

Researcher Characteristic and Beliefs

Frames are the principles of organization which govern social events and the actors' subjective involvement in them (Goffman, 1974: 10). Researchers' views of the world come from their socialization in a particular culture and these experiences influence their interpretation of data collected. Life experiences which may impact on interpretation include class location, age, gender, ethnicity, personality, and religious or political affiliations. Other influences lie in the disciplines researchers have been trained within, be it philosophy or psychology or sociology together with the literature relevant to their research question which researchers have examined prior to data collection. Understanding which particular aspects of life and educational experiences influence one's perceptions is important and exposing these in a diary in the process of self-reflexivity, is necessary when data collection and analysis are undertaken.

Researcher Position

Where the researcher stands is also important. This may well develop as a result of the research question, the approach decided upon, frames accumulated, and anticipation of the potential audience. For example, a phenomenological approach requires the researcher to achieve an in-depth understanding of the experiences of interviewees of the phenomenon under investigation – here a passive empathetic position is essential and the researcher simply collects and displays data with minimal analysis. An autoethnography would require one to go deep into one's own experiences and emotions in order to achieve the complete insider view and then have the capacity to move to a more distant position in order to interrogate oneself about these experiences. At the other extreme, analysis of existing documentation such as television programs or policy documents would require a distant pseudo-objective stance where researcher feelings and emotions would be of lesser importance and data fragmentation and relinking of greater importance.

The Reader

The audience to whom the final outcome of data analysis will be displayed also influences decisions about the form analysis may take and the degree to which data segmentation/intactness occurs. The role the reader can assume is created both by the tone of the report or article that is produced from the research and also by the form of presentation of data. If the researcher takes a strong authoritative voice along the lines of “I, as an expert in this field have undertaken this piece of research and here are my findings for you to digest,” then the reader is placed in a passive take it or leave it position. In this situation, support data for researcher findings will be displayed but this display may be fairly minimal and the reader would have to trust that the researcher has displayed the most relevant data rather than that which best supports her/his particular view. On the other hand, the reader may be encouraged to come close to the raw data in order to see more of what the author has experienced and to judge the interpretation of this through hyperlinks or the display of extensive quotes, case studies, vignettes, and the exposure of many participant voices.

The issues of validity (demonstrating that the author has gained truths regarding the research question) and reliability (demonstrating sound design to gain these truths) are also of importance and are clarified via the notion of transparency. Detailed explanations need to be provided so that the reader can judge what has been done, and how particular conclusions have been arrived at.

Managing Data during Collection

There are four different types of data that can be collected: individual interviews, focus group, observation data, and the collation of existing documentation, be it written, visual, or material. Although the actual organization of transcription approaches for individual pieces of data is idiosyncratic to suit each researcher, there are various mechanical processes which tend to be used widely.

Managing Interview Data

For interviews, it is desirable that transcription occurs as soon as possible after the interview has been undertaken so that the nonverbals can be recalled and inserted appropriately. Transcription is usually undertaken double-spaced for ease of reading and with a column of about one third the width of the page being left blank on one side for preliminary data analysis to be written in. Within the text of the transcript various indicators are used that is, . . . indicates silences or pauses for thinking where each dot represents a period of time defined by the researcher – usually one second per dot. Nonverbals are usually indicated in brackets and with italics (e.g., *(smiles)*

or *(coughs)*). Emphasis is added in bold (i.e., . . . he said that but **it is not a view I share**). Clean copies (without any form of analysis can be sent back to the interviewee for verification of the transcript).

Managing Focus Group Data

Focus group data are not usually transcribed – partly not only because it is not necessary to send transcriptions back to participants but also because the immediate cross-referencing that occurs in the group makes further verification unnecessary. Another factor is that transcription is difficult to do and many transcribers refuse to attempt this because of the difficulty of unscrambling individual voices. With the introduction of videoed focus groups, the need for written transcription has disappeared.

Managing Observational Data

Observations are usually either taped or written up without discussion and preferably within 24 h after the observation has been undertaken. Discussion is thought to start the process of interpretation often bringing in the perceptions of others. Each observation may be accompanied by a brief overview map which indicates movement within the setting of both the researcher and those under observation. This provides a useful aid for prompting memory and a visual diagram for the reader.

Managing Textual Data

Textual/verbal documentation once collected will either be left as it is, apart from chronological compilation or sorting into broad groupings, or it may be scanned or transferred onto a disk for greater ease of management and analysis.

Preliminary Data Analysis

This process is undertaken in some form or another regardless of the design approach which has been chosen. The purpose of preliminary data analysis is twofold; first, to track data as it is being collected so that any gaps that may be emerging can be dealt with early and second, to start the process of identifying major trends in the data as they emerge. It is an ongoing process of engagement with the text in order to gain a greater understanding of the meanings, values, and actions which are being portrayed.

There are various idiosyncratic ways of undertaking preliminary data analysis but most approaches include some aspects of the following six points as the researcher reads through each interview, observation, and document as they are collected:

- Questioning one's own assumptions and those which appear in the data – this can be done by asking why? or

what about...? or what if... or well if this, then how does that fit...? From these critiques, a list of further questions to pursue can be developed.

- Comparing the data with one's own experiences, with findings which appeared in the literature review and with any theoretical/conceptual explanations which look as though they may shed light on what is emerging to date. Brainstorming the ideas these comparisons bring up is helpful here.
- Reading the data in a different order or even haphazardly in order to gain different perspectives.
- Chunking data – starting to group together segments of data, for example, experiences in the classroom, and creating broad labels and sub-labels for these groupings, that is, as negative experiences or positive experiences.
- Writing analytic memos about each segment incorporating ideas, questions, and additional information recalled from the actual collection of these data (setting, context, other events, processes, interactions, etc.).
- Taking some of the groupings as they become more established and turning them into forms of presentation, for example, poems, vignettes, or narratives in order to identify orientations and information missing, and to trial approaches of data presentation.

Examples of preliminary data analysis

Figures 1–4 provide examples of preliminary data analysis.

Face Sheets

Following completion of the above processes of preliminary data analysis for each piece of data collected, it is

useful to collate the major points in some easily accessible form such as a face sheet which can be attached to the transcript or document. This face sheet should identify who/what the data comprise, for example,

Interview No 1 with 'John' School Principal. Date Location. and Time in the setting. Or, if it is an observation; Observation number ... of ... undertaken (date), (time in setting). Any special circumstances which may have an impact on the data collection should also be noted: the day was very hot/cold/I was late/X was late. Interruptions occurred here... and here.... If the data is an existing document; Document name... Source location ... Date accessed ... Verification of the authenticity of this document made by...

Then a summary usually in dot-point format should clarify what the researcher has found of interest in the data with regard to:

- the major issues emerging; and
- aspects which need to be followed up.

Assuming the above example from an interview (Figure 1) was a complete interview, the face sheet could look like this (see Figure 5):

Focus group data tend to be analyzed soon after the exercise using a replay of the video/audio recording so that developing themes can be visually/manually identified. But aspects of preliminary data analysis will also be built into this process and questions developed for exploration in the next focus group. These will be pursued in depth in face-to-face interviews with individuals identified from the group who have indicated they would be happy to provide further information.

Interview transcript segment from an interview with the director of an adult training center for young people with intellectual disability	Preliminary data analysis
<p>Question: What better ways might there be of integrating people with intellectual disabilities into community activities?</p> <p>Answer: Well regarding physical activity, everyone functions better and are more mentally alert if they are physically fit. I think for the severely disabled it is essential they get more from this part of the program than any other. For years, most of them have just had a ball thrown at them. we've got a basketball team that used to stand on the court and wait for a ball to be thrown - its all a matter of education, with community access, you just add a grade to the local district association so they play at their skill level and the system is open to them to move on up if they are capable.</p>	<p>How do you get these people to shift from passive to active activity? What forms of education/processes were used here?</p> <p>Re-grading, is this really integration? Aren't we just creating another segregated layer for these people who are already marginalized to the fringes of society and who will continue to be separated on the basis of their limited capacity?</p>

Figure 1 Interview data. From Grbich, C. and Sykes, S. (1989). *What about Us! A Study of Access to School and Work of Young Women with Severe Intellectual Disabilities*. Melbourne: Monash University, Krongold Centre for Exceptional Children.

Observation transcript segment from a case study of drug education in colleges.	Preliminary data analysis
Set on the windswept plains of one of the outer limits of X suburban sprawl, this new school is being developed to meet the needs of a housing estate. The estate is a fenceless collection of architect-designed display houses and standard homes with a mixture of bald turf and newly turned clay gardens. It is an isolated development with a long walk to the shops and a bus to connect with the major public transport systems. Most of the houses seemed empty in the mid-morning with closed curtains and no one in a yard or on the streets. The school is still being built. The buildings are spaced apart with brick-paved courtyards and the occasional well-protected shrub. It is lunchtime, students huddle together in small groups in a yard in the lee of buildings and movement between groups appears to be hindered by a fierce wind. There are no games being played. Not even the usual handball game is evident. The unpaved areas of the school are deep in wet clay.	<p>What are the demographics of this estate?</p> <p>How do students entertain themselves during school breaks and after school in this community?</p>

Figure 2 Observational data. Adapted from Garrard, J. and Northfield, J. (1987). *Drug Education in Victorian Post-Primary Schools*. Melbourne, VIC: Monash University, Victoria.

Textual document segment from policy documents identifying aims, values, and beliefs in junior school education	Preliminary data analysis
<p>Values are internalized sets of beliefs or principles of behavior held by individuals or groups. They are expressed in the ways in which people think and act. No schooling is value free. Values are mostly learned through students' experience of the total environment, rather than through direct instruction. The content of a school's curriculum reflects what is valued by a society and a school community.</p> <p>A belief in education, at home and at school, is a route to the spiritual, moral, social, cultural, physical, and mental development, and thus the well-being, of the individual. Education is also a route to equality of opportunity for all, a healthy and just democracy, a productive economy, and sustainable development. Education should reflect the enduring values that contribute to these ends. These include valuing ourselves, our families, and other relationships, the wider groups to which we belong, the diversity in our society, and the environment in which we live. Education should also reaffirm our commitment to the virtues of truth, justice, honesty, trust, and a sense of duty.</p>	<p>The purpose of education in New Zealand appears to consolidate the student into the values of her/his community. Does this mean that achievement beyond this group or criticism of these values is not encouraged?</p> <p>In the United Kingdom, education is seen as a way of providing equality of opportunity as individual well-being and to foster the greater good of the wider community. How many of those in the lowest socioeconomic class would move up through education? Or does education fail to equalize opportunity?</p>

Figure 3 Existing documentation. From New Zealand Ministry of Education (1993). *The New Zealand Curriculum Framework. New Zealand of Education*. Wellington: Learning Media; and Qualifications and Curriculum Authority (QCA) (1999). *The National Curriculum: Handbook for Primary Teachers in England Key Stages 1 and 2*. London: DTEE and QCA.

Data Summaries and Early Theory Generation

These are seen as useful ways to start consolidating data as they are collected. Every three to six data sets, a summary is produced which starts teasing out related issues which

are starting to appear as recurring themes. The twin advantages of undertaking this process are to further identify any holes in individual themes which may require further data collection and to start the process of writing


Visual document segment	Preliminary data analysis
 <p>Nursery children from a crèche at a cylinder factory dancing in a forest with their teachers in Russia 1935.</p>	<p>Young children dancing in the forest probably during summer in the early depression era. Why are they wearing knitted caps but no shirts? They look well fed despite the depression era. What was happening in Russia at this time?</p> <p>Do the knee-length white clothes and bobbed hair of the teachers reflect women's fashion in 1935? Or are they unique to this country?</p>

Figure 4 Visual documentation. Free photo from <http://www.clipartguide.com/>

Face sheet
<p>Identifiers: Interview number 1 with director of an adult training center for young people with intellectual disability</p> <p>Location of interview: the director's office</p> <p>Date: 02/03/89. Time: 3.00 – 4pm. Initial coffee and unrecorded introduction. Recorded interview lasted 45 min. Debrief and chat.</p> <p>Special circumstances: None</p> <p>Major issues emerging:</p> <ul style="list-style-type: none"> ▪ People with intellectual disability may benefit from specific educational programs to change attitudes and maximize their potential ▪ Integration may simply mean isolation and segregation under a different label. <p>Aspects to be followed up</p> <ul style="list-style-type: none"> ▪ Information required regarding the content of education programs to convert passive games response of people with intellectual disability to an active response. Success rate? ▪ Interview needed with state/national basketball organizers to explore the feasibility of integration through an additional grade of players ▪ Has genuine integration occurred in any sport?

Figure 5 Face-sheet example.

up and interpreting results. Early re-presentation of data may also be trialed here (poems, etc.).

The following example (**Figure 6**) consolidates six sets of interview and observation data in a study which sought to explore the integration of young people with severe intellectual disabilities into school and community settings. Early themes from parents' and teacher's views are interpreted through various theoretical concepts such as normalization and integration.

Following the processes of both preliminary data analysis and data summaries, the researcher should be able to provide some answers to their research question and should be on top of her/his database, avoiding ending up in the unenviable situation of being surrounded by a pile of data with no clear idea of how to make sense of it.

Some researchers, particularly those working within the subjective tradition, may well decide not to pursue processes of data analysis beyond this point as their

Perceptions of integration in schools: Researcher's observations, parents' and teachers' views

One of the central issues emerging so far is how different people view the integration of special school students into regular schools (maximum of one half to 2 days per week). Parents (of X, Y, and Z) all favor this, seeing it as a glimmer of hope, an indication of improvement and a move across the great divide from special to regular schools with accompanying connotations of normality. All are very anxious their son/daughter should remain in the integration program despite fitting (epileptic fits) experienced on regular school days by X (none on special school days), isolation and aloneness of Y when his mate was withdrawn from the program (teacher's perception). X seems to be settling down over time. He likes the pop music in practical classes and although he can't speak, he can sing along to the radio/cassette/CD player and the girls think he is cute and encourage him to sing. Y is ignored (nonverbal and almost blind) and Z has full aide support and a home room to retreat to (which he shares with all the special students being integrated in this setting). Regular students are encouraged to come in and play cards with Z and other students with disabilities in their home room. This is very popular on wet days as it beats standing around in the windswept covered ways which is the only other option. Z seems calm and settled but the presence of an aide may be preventing interaction with regular students in the classroom setting. The school environment and forms of/lack of support do appear to impact on the integrated students' experiences.

Teachers in special settings fear integration, seeing it as leading to a loss of students with the potential of job loss for them – they feel that if the students are starting to move to regular settings they should go too, as they could provide expertise to support regular teachers. The latter view integration as ideologically sound but pointless in practice unless there is full aide support so they don't have to worry about the presence of X, Y, and Z or adjust their curricula (and none have so far). X and Y's teachers in the regular classrooms express concerns about what X and Y are gaining from the integrated experience, viewing it as negative and stressful for both them and the student. It seems clear that schools that are more academically oriented (like the ones X and Y attend) are more resistant to any disturbance caused by integrating low-performing students whereas less academically oriented schools, such as the one where Z is located, express less hostility and are more oriented to giving it a go.

Figure 6 Example of a data summary. From Grbich, C. and Sykes, S. (1989). *What about Us! A Study of Access to School and Work of Young Women with Severe Intellectual Disabilities*. Melbourne: Monash University, Krongold Centre for Exceptional Children; and Grbich, C. (2007). *Qualitative Data Analysis: An Introduction*. London: Sage.

preference will be to keep the data as intact as possible. Careful preliminary data analysis together with data summaries which consolidate information emerging will have provided all the analytical processes necessary to manage their data.

Other researchers, particularly those with large databases or data that they feel require a more rigorous process of analysis will move to thematic analysis. Rather than a vertical building up of the database as in preliminary data analysis, thematic analysis is more a process of horizontal segmentation and regrouping.

Managing Data after Collection

Thematic Analysis

Once all the data have been collected and the processes of preliminary data analysis have ensured that any major gaps have been plugged, the researcher/s should take a break in order to achieve a more distant perspective.

It is also useful at this point to pass a clean copy (without preliminary data analysis) of the transcribed or collated data to a friend, colleague, partner, or another member of the research team and encourage them to go through and identify the issues which they see as being present. This will help locate aspects which the researcher may have overlooked or deliberately discarded as not fitting into their view of what is important.

The researcher then starts the active process of moving across the data and selectively grouping particular aspects. These groups may comprise a set of answers to a particular question? For example, "How do you try to meet the different intellectual levels of students in your class?" Or grouped stories associated with "When did you first understand that your child had a disability?" Or, the researcher may consolidate emerging themes from the database. These themes and responses may be collated either using the block-and-file approach in which the researcher can choose to either keep data intact or to fragment them, or, they may prefer to use a conceptual mapping approach.

The Block-and-File Approach (Intact Data)

The first stage involves identifying the relevant responses to a question such as, “What was your experience of gaining a diagnosis and support for your child?” and placing them in a file.

The second stage involves underlining or shading the different themes within these responses and placing them in separate columns and attaching general headings.

The responses have been kept intact and overlapping is not an issue (Figure 7). The headings given may stay in this form but may well change as new data are added. Consolidation and the development of new columns are desirable. For example, the column containing responses of health professionals to parents may need to be broken

down into responses of doctors and responses of other health professionals’ or even further into male and female health professionals’ responses as more data are added.

Block- and File-Approach (Data Fragmentation)

A slightly different approach to thematic analysis involving further fragmentation and some de-contextualization might focus in on a more refined aspect of the data such as doctor’s interactions with parents. The process involves identifying relevant key words and listing these, regrouping and labeling them, and developing a working proposition about what has been found. See Figure 8.

Length of time to gain diagnosis	Responses of health professionals to parents	Responses of parents to the situation
1. For the first 10 months no one helped. I went to several doctors and a pediatrician.	1. He (pediatrician) told me “He’s retarded. I haven’t time to answer your questions; I’m on the phone to the Physiotherapist. I hope I haven’t ruined your weekend. Try an intervention program. Some kids it helps, some it doesn’t. The second pediatrician we saw at X hospital was nice but hasn’t given us advice on where to go. I have to go and find everything myself. He didn’t even tell me there was a developmental disability nurse at the hospital.	1. Its not a very caring hospital. Information is not given freely. They don’t bother to remember a child’s name – don’t give any extra attention. I wouldn’t leave my child overnight there. I did once, he was in a very bad way when I came back in the morning. They told me he was very spoilt.
2. We finally were referred to a pediatrician at 6 months.		
3. I knew there was something wrong at 6 weeks but they didn’t test her until she was two and a half.	2. He was absolutely atrocious – he treated the situation as though it was a damaged brain independent of the child. He didn’t refer to the services in the hospital, let alone outside. There was a total lack of information as to what cerebral palsy was.	2. There is a lack of co ordination of services when disability is diagnosed. It should be function of the hospital to provide a support base from which to move out into the community to whichever services are most appropriate.
4. Two days after her birth the doctor came and told me my child had Downs Syndrome	3. I was told she had curvature of the spine and that’s all I was told. When she was 4 years old I became dissatisfied and went to a private physiotherapist. She got my daughter’s medical history and on my next visit she told me my daughter had cerebral palsy.	3. When she was 4 years old I became dissatisfied and went to a private physiotherapist.
	4. Two days after her birth, the doctor came and told me my child had Down’s Syndrome and we were strongly advised to put her in a home and forget her because she wouldn’t live past 15 and would just be a vegetable	4. I was so upset – she was my own flesh and blood - so I took her and left the hospital and vowed to ignore this doctor. Now look at her – she can walk, talk, and see with glasses; she is now 16 and in good health.

Figure 7 Block- and file-approach (four intact interview responses).

List relevant data:

Rude and abrupt, I haven't time to answer your questions, 'He's retarded'...try an intervention program some it helps others it doesn't, the pediatrician was nice but gave no advice on where to go, he didn't even tell me there was a disability nurse at the hospital, information isn't given freely, absolutely atrocious, referred to damaged brain separate to the child, he didn't refer to the services at the hospital let alone outside, total lack of information, there is a lack of coordination of services when a disability is diagnosed, it should be a function of the hospital to provide a support base and link to community services, I was told she had curvature of the spine and that is all I was told, we were strongly advised to put her in a home as she would only ever be a vegetable.

Group under particular labels:

Uncaring attitude: rude and abrupt, I haven't time to answer your questions, 'He's retarded'...try an intervention program some it helps others it doesn't; absolutely atrocious - referred to damaged brain separate to the child, we were strongly advised to put her in a home as she would only ever be a vegetable

Unhelpful attitude: the pediatrician was nice but gave no advice on where to go, he didn't even tell me there was a disability nurse at the hospital, I was told she had curvature of the spine and that is all I was told

Connection to services: information isn't given freely, he didn't refer to the services at the hospital let alone outside, total lack of information, he didn't even tell me here was a disability nurse at the hospital, there is a lack of coordination of services when a disability is diagnosed, it should be a function of the hospital to provide a support base and link to community services.

Develop a summary proposition to which case examples can be attached in the writing-up process:

Proposition: Doctors' attitudes to this group of parents of a young child with a disability have been uncaring and unhelpful providing poor information and failing to connect parents either with existing hospital or with community services.

Figure 8 Block-and-file approach (focused and fragmented data).

Conceptual Mapping

This is another way of tracking the outcomes of thematic analysis. This approach is useful when a broad overview or a specific aspect is required. More than one map will need to be developed in order to represent different parts of the database.

From the above examples, parents' experiences with doctors could be displayed in a conceptual map (**Figure 9**):

Coding

If the researcher has a large database or plans to utilize computer management of data he/she will need to undertake a process of coding involving the attachment of numbers to the groupings identified in the database. This process may be undertaken prior to thematic analysis especially when preliminary data analysis has clearly identified the major groupings. However, it is suggested that it is better undertaken following thematic analysis when the researcher has finalized all the groupings in

the data. The application of codes prior to thematic analysis can lead to information being randomly assigned to particular coding groups and the researcher then has to spend considerable time sorting through these groups to weed out irrelevant data and develop new codes and subcodes. In addition, the development of a coding framework too early in the piece may lead to the researcher becoming so attached to it that all subsequent data are forced into this framework regardless of fit. In this manner, new directions in the database are often overlooked completely.

Codes should include all the themes identified as well as the concepts linking to theoretical perspectives which may be useful in the interpretative process. These may have been generated by the researcher or are grounded in the data and used by participants. Other aspects which may be useful to code include: methods used, important events, aspects of the setting, for example, school organization or structures which may impact on the data collected, the political situation which may influence funding or values, relationships developed between researcher and researched, etc. The number of codes will

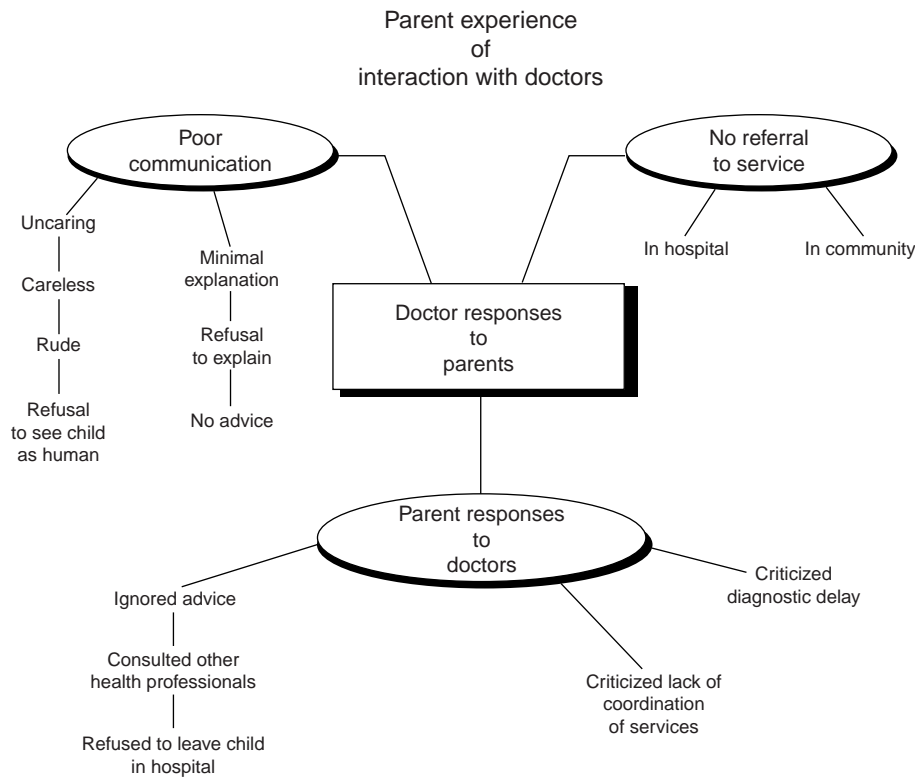


Figure 9 Parents' experiences with doctors, in a conceptual mapping.

depend on the database but more than 50 codes and subcodes in a moderate-sized database would be regarded as unwieldy.

Example

In the study examining integration of young people with intellectual disability which involved 80 randomly chosen multiple-member case studies (the individual, his/her parents, regular school teachers and special school teachers, and principals/sheltered workplace organizers, and halfway house carers) the following broad codes were developed: demographic details (age, sex, living situation, and attendance at school/work), diagnosis, communication regarding diagnosis, developmental issues, level of independent living skills, social networks, social activity, disability funding, family income, services accessed, family composition, effect on family, schooling/work history and capacities, family expectations, sexual activity (actual and expected), home activities, school/workplace tasks, transport, and community integration. These formed the broad frameworks the data became grouped into but within each were a series of subgroups. For example, independent living skills included the subcodes, cooking skills, telephone use, toileting, self-washing, feeding, washing clothes, and communication level: verbal/nonverbal/communication/board, as well as dressing, showering,

coping with menstruation, sexuality issues, budgeting and other, in order to identify the range of living skills and the issues therein evident in this group.

Data Management Using Computer Programs

Brief History

Although computer packages cannot analyze qualitative data, only manage it, there has been a proliferation of these programs in recent years. The main clearing house for these programs is Computer Assisted Qualitative Data Analysis (CAQDAS) (see the section titled 'Relevant web-sites' section for the link to this where a number of programs are listed and links and demo programs provided).

The development of these packages has undergone three distinct developmental eras since their emergence 30 years ago.

- Systematization: code and retrieve,
- Theory generation, and
- Content analysis.

Systematization

The notion of systematic data emerged in the 1970s following an earlier push to make qualitative research

look more like quantitative and to be more transparent regarding methodological approaches. This resulted in pressure for more rigorous exposure by researchers in order to justify their findings. The programs developed in this tradition focused on the storage and retrieval of data and the provision of subviews of the data through the assigning of codes and categories. These early programs were termed code and retrieve.

Code-and-retrieve programs are generally single-file systems and require the segmenting of the text and the attaching of codes to these segments. These coded segments are then filed with some form of identification, for example: "Interview No 1 with xy, Date..." and some programs also have the capacity to attach memos to individual segments and to undertake Boolean searches of and, or, or not. The capacity to define variables and links to statistical software such as Statistical Product and Service Solutions (SPSS) is often provided to allow frequency counts. Files can then be printed out for further examination and consolidation. Coding can occur either on or off screen. On-screen coding has always attracted debate because of the limited screen views and also because of the error rate of reading on screen.

Ethnograph 5 (PC) is an example of this type of program being the most well developed of the cheaper code-and-retrieve programs. It can create and import data files, format the file, and automatically save it as an ETH file and import an unlimited number of data files. It can also rename, move, merge, duplicate, delete, or back up projects as desired, apply several codes at once, create a master code list, organize words into code families and code family trees, attach memos of up to 32 pages long to specific lines in the data file, search by Boolean codes, search by author-created code, and cross reference segments. Code frequencies can also be identified via export to SPSS.

Theory Generation

During the mid-1980s, researchers at qualitative computing conferences recognized the benefits of the first generation of code-and-retrieve programs but pointed out their limitations, particularly their inability to facilitate theoretical interpretations. Another perceived limitation was the incapacity to support the combining of qualitative and quantitative data. These debates resulted in the development of a second generation of computer-management programs. These theory-generation programs comprise a two-file system (data and literature) and are often underpinned by the structural framework of grounded theory, in particular the constant comparative process which the researcher can operate between the two files using Boolean logic. This process facilitates the pitting of propositions against previous literature and existing theoretical interpretations in the generation of new explanations.

Atlas/ti v 5.2.15 (PC) has been a popular program within this tradition. It has an object-oriented graphical user interface for processing of textual, graphical, audio, and video data, on-screen coding (drag and drop), no fixed definition of data segments, simultaneous display of data segments in context, codes, and memos, virtually unlimited number of documents and segments, integration of all relevant material – primary texts, annotations, theories – into separate files, mind mapping and graphical network editing, semi-automatic coding with multi-string text search and pattern matching, theory building, hypertext links, Boolean logic, hypotheses, end-on SPSS, presentation templates, and networking.

The development of these more sophisticated theory-generation programs prompted further discussion regarding the methodological and theoretical implications of the use of computers to manage qualitative data. Initial concerns included the possible impact on the craft of qualitative research and the moves toward control rather than diagnosis and toward explanation and method direction rather than method evolution and interpretation.

Content Analysis

These packages are useful tools for breaking into documents. Most have the capacity to undertake the following: word frequencies to display how often each word occurs in a document; category frequencies which group synonyms into categories and indicate how many times each category occurs in the document; key word in context (KWIC) which displays, in alphabetical order, each word together with a number of words on either side to provide information on its context in the document; cluster analysis where groups of words can be identified as being utilized in similar contexts; and co-occurrence of pairs of words. Dictionaries and cultural grammars are often included and the more sophisticated of these programs are developing the capacity for more complex semiotic analyses.

Textpack 7 (PC) is a basic content-analysis package originally developed to cope with open-ended questions in surveys. It undertakes word frequencies, KWIC and keyword-out-of-context, cross-references, and concordances, and can undertake word comparison of two texts.

Management by Computing Packages: Some Concerns

Despite their usefulness in the management of large databases, the concerns which emerged in the 1980s regarding these packages have persisted. The following is a summary of the major issues which center round the construction of knowledge.

Processes of interpretation

The interpretive processes of data management and analysis which occur without computer management are fluid, involving an *ad hoc* process of building up and fitting together groups of information while regularly standing aside to allow the light of previous literature and the lens of various theoretical perspectives to provide insight. The structure of computer-management programs limits the chaotic nature of this process. In order to facilitate computer management, the database often has to be carved up into lines or paragraphs and this segmentation process completely changes the way the researcher interacts with data. The more sophisticated programs allow for researcher-defined text segments but even with this facility, the framing devices of structured coding, categorizing, relinking, and reinterpreting required by most computer programs cannot fail to have an effect.

Theoretical underpinnings

In addition, most programs are underpinned by their own particular theoretical structures. The impact of the inverted tree-framing structure of NUD* 1ST parallels the development of cognitive frames for knowledge where networks of nodes and relations hang off a defined aspect or category. Adaptation of nodes and replacement frames through matching and accommodation are part of the continuing development of new knowledge. The consolidation of this knowledge is drawn from branches to the singular trunk, leading to a definitive and fairly linear conclusion. In contrast, ATLAS/ti uses a nonhierarchical horizontal rhizomatic frame which stretches to infinity and is constantly evolving. This latter approach fits more closely with the expectations of qualitative research that transformation and change are intrinsic. Whatever form they take, these program underpinnings must inevitably shape the outcomes and perceptions of data collected.

Program structures

Tools constructed for a particular program must also structure thought processes and texture the data in particular ways. Each tool creates artifacts and metaphors which are not neutral in effect and which change our ways of thinking and seeing. The procedural aspects of coding, categorizing, and relinking are fundamentally reductionist and affect our views of the data as they move from a complex, multifaceted reality of intersecting aspects embedded in rich contexts, to a simplified version viewed as discrete groups of representation and subrepresentation.

Once the researcher starts to think in a reductionist manner and to use high-frequency logic, the speed of data atomization will far exceed the speed at which data would be contemplated in the reflexive process off screen. Although speed and a capacity for logic may be advantageous in coping with computer technology, these may well foster superficial and decontextualized interpretations of qualitative data.

De-contextualization

The de-contextualization of data is clearly an issue. The production of narrowly coded segments confines the data separating them from both the researcher and the context. Discovery then tends to be linear rather than chaotic and complex. It seems inevitable that these procedures must simplify what we see, limiting the potential for change and transformation of both researcher and data in the research process.

The introduction of the third-to-fifth-generation languages incorporating Boolean logic and true-false dichotomies foster an emphasis on objectivity, minimization, variable analysis, and enumeration. These approaches have links with positivism. Views of reality tend to become algorithmic and sequential favoring deterministic logic. The representations of reality that are produced are based primarily on discrete objects, although fuzzy logic (where variables are less precise and can overlap) has been used in a couple of programs. The current emphasis in qualitative research has been away from stage-based coding programs such as traditional Straussian grounded theory and toward postmodern versions of all methods particularly ethnography and feminist approaches where the emphasis is on data which are intact and contextualized rather than atomized.

Apart from texturing, fragmenting, and de-contextualizing reality, the emphasis in computer programs on mathematical logic tends to distance the multi-perspective, creative approaches which form the backbone of qualitative research. The thin segmented data which eventuate from computer programs are more likely to be forced into a particular outcome to fit a particular purpose rather than allowing them to challenge current knowledge.

Database size

The increasing capacities of many programs encourage the collection of larger and larger databases and there is a tendency to emphasize the common and shared properties of a large number and to make the framing device more real than its content. In this manner, meaning is lost, people are de-personalized, events neutralized, and the variations in the minutiae of detail of small numbers of participants lost. Another result may be the application of the principle of garbage in and gospel out (GIGO). The reification of computer technology often incorrectly convinces researchers that qualitative research cannot be undertaken without a program and that the codes they develop within these programs can rise to the state of variables capable of quantitative analysis.

Data quantification

Quantitative analysis through the introduction of end-on SPSS and other statistical packages fails to take into account the differing paradigms between quantitative and qualitative data. Where the same data set is involved, the different sampling techniques between the two styles

and lack of variable definition and control in qualitative approaches are relevant. The qualitative emphasis on diversity, nonrepresentativeness, small numbers, self-selection, and thick ethnographic description, suggests that drawing quantitative outcomes from qualitative data is contraindicated and that such a transformation produces unrealistic versions of the original. Where the quantitative data come from a separate data set with the designs kept separate, this is not an issue.

Theory Generation Using Qualitative Data

Regardless of whether computer management is, or not, used, the process of theory generation involves taking the consolidated data groupings derived from preliminary data analysis, thematic analysis, and coding, and subjecting them to two processes. The first involves contrasting the findings with existing literature in order to question, challenge, and add knowledge to the field. The second is to interpret the results found from a more abstract position utilizing whatever theories, models, or concepts are relevant to the discipline where the author is situated, or to the anticipated readership. For example, the study on intellectual disability cited earlier in this article was initially published as a government-funded report; so interpretation was placed within practice-based models of integration within the concepts of integration and normalization. When parts of the study were published in academic journals, both these frameworks and Max Weber's ideal democracy and his concept of economic rationalism were also utilized.

Conclusion

Managing data adequately is crucial not only for preventing the researcher from drowning in his/her data but also for demonstrating researcher trustworthiness and dependability. The subjectivity of the researcher and the complexity of the analytical and interpretive processes require considerable transparency if the reader is to be convinced that the project has been managed properly. The construction of knowledge and how the information researchers collect within a particular research design transfers into results and then is interpreted into conclusions and recommendations is a process we need to be fully aware of. Data-management benefits from starting at the beginning of data collection through the vertical processes of preliminary data analysis, face-sheet development, and regular data summaries. Post data collection, further data segmentation of a horizontal nature can be

undertaken through thematic analysis. In large databases or those to be managed through computer programs, coding should occur but in order to prevent oversimplification and de-contextualization, this should preferably follow thematic analysis rather than precede it.

Computer programs provide a useful management tool for larger databases, but the use of these continues to be contentious. When data become segmented and ordered in monoepistemological computer programs rather than in polyepistemological mind processes away from computers, then we need to consider very carefully the frames, processes, and structures we are imposing. We must always be wary of systems which have the capacity to direct and simplify the views of researchers and ultimately those of readers.

Bibliography

Goffman, E. (1974). *Frame Analysis: An Essay on the Organization of Experience*. New York: Harper and Row.

Further Reading

- Ezzy, D. (2002). *Qualitative Analysis: Practice and Innovation*. Thousand Oaks, CA: Sage.
- Garrard, J. and Northfield, J. (1987). *Drug Education in Victorian Post-Primary Schools*. Melbourne, VIC: Monash University, Victoria.
- Grbich, C. (2007). *Qualitative Data Analysis: An Introduction*. London: Sage.
- Grbich, C. and Sykes, S. (1989). *What about Us! A Study of Access to School and Work of Young Women with Severe Intellectual Disabilities*. Melbourne: Monash University, Krongold Centre for Exceptional Children.
- Kelle, U. (ed.) (1995). *Computer Aided Qualitative Data Analysis: Theory, Methods and Practice*. London: Sage.
- New Zealand Ministry of Education (1993). *The New Zealand Curriculum Framework. New Zealand of Education*. Wellington: Learning Media.
- Qualifications and Curriculum Authority (QCA) (1999). *The National Curriculum: Handbook for Primary Teachers in England Key Stages 1 and 2*. London: DTEE and QCA.
- Woolcott, H. (1994). *Transforming Qualitative Data: Description, Analysis and Interpretation*. Thousand Oaks, CA: Sage.

Relevant Websites

- <http://www.caqdas.soc.surrey.ac.uk> – CAQDAS Networking Project: Links.
- <http://bama.ua.edu> – Content Analysis Resources: Software, The University of Alabama.
- <http://www.qualisresearch.com> – The Ethnograph.
- <http://www.atlasti.com> – ATLAS.ti.
- <http://www.social-science-geis.de/en/software/textpack/index.htm>.
- <http://www.textanalysis.info> – Text Analysis Info - QDA software.
- <http://homepages.vub.ac.be> – VUB/ULB Computing Centre.

Semiotics

M Danesi, University of Toronto, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Connotation – The extended or acquired meanings of a sign.

Denotation – The referential or intentional meaning of a sign.

Markedness theory – The theory which posits that one of the two elements in an opposition is the default (unmarked) form and the other the marked form (the one that stands out).

Opposition – A binary relation that keeps signs distinct from each other through opposition (night versus day).

Paradigmatic structure – Any aspect of signs that keeps them distinct.

Semiosis – The process of producing and understanding signs.

Signification – The process of using signs for various psychological and social-cultural reasons.

Signified – The conceptual part of the sign; what the sign encodes and refers to.

Signifier – The physical part of a sign; also known as form.

Syntagmatic structure – The combinatory structure of signs (how they have been constructed).

and the like; the latter include not only words, but also gestures and the various symbols that humans invent to serve their psychological, social, and communicative needs (Deely, 2001: 24–56). It was the British philosopher John Locke (1632–1704) who first put forward the proposal of incorporating the study of conventional signs into philosophy in his *Essay Concerning Human Understanding* (Locke, 1690). However, the idea of fashioning an autonomous discipline of sign study did not emerge until the late nineteenth century, when the Swiss philologist Ferdinand de Saussure proposed such an idea in his *Cours de linguistique générale* (de Saussure, 1916). Saussure suggested that the main goal of a science of signs was to understand how signs stand conventionally for things in specific social environments and how they permit people to interact meaningfully.

The first sophisticated classification of signs is due to the American pragmatist philosopher Charles S. Peirce (1839–1914), who identified 66 species of signs in total. Some of these are simply terminological strategies that provide detail as to how signs function cognitively. For example, as its name implies, a qualisign is a form that draws attention to some quality of its referent (the object it represents). In language, an adjective is a qualisign since it draws attention to the qualities (color, shape, size, etc.) of nouns. In other sign systems, qualisigns include colors (painting), harmonies and tones (music), etc. A sinsign is a form used to single out a particular object – a pointing finger and the words here and there are examples of sinsigns. A legisign is a form that designates something by convention (literally by law). Legisigns include symbols, emblems, such as those used on flags, and various other forms that are used according to specific social rules or conventions. However, there are three signs identified by Pierce that are now part and parcel of every semiotician's methodological toolkit – icons, indexes, and symbols. Signs originating in the perception of some property in a referent are icons. Iconic signs stand for their referents by resemblance, imitation, simulation, or emulation. Portraits are icons of human faces (and what they mean); onomatopoeic words are icons of sounds made by certain objects or actions (drip, bang, etc.); and so on. The meanings of icons can generally be figured out even by those who are not a part of the culture that use them. Signs that relate referents, sign users, and even other sign forms in some way are known as indexes. The pointing finger is a perfect example of an index. When we point to something, we are in fact relating it to our location as pointers. If it is close by, we refer to it as

Classifying Signs

The first fundamental task of any science is the classification of its objects of investigation. This is also the case with semiotics. The first classification system of signs goes right back to the ancient world, starting with the definition of sign by Hippocrates (c. 460–370 BCE), the founder of Western medicine, as a symptom. Hippocrates argued that the particular physical form that a symptom takes – a *semeion* (mark) – constitutes a vital clue for finding its etiological source. Shortly thereafter, philosophers started referring to signs as being either natural (produced by the body or nature) or conventional. Among the first to tackle this basic typology was St. Augustine (354–430 CE) in his *De Doctrina Christiana*, in which he describes natural signs (*signa naturalia*) as forms lacking intentionality and conventional ones (*signa data*) as forms produced by human intentions. The former include not only symptoms, but also such phenomena as plant coloration, animal signals,

near or here. If not, we refer to it as far or there. Finally, signs that stand for referents in historically based or conventional ways are symbols. The V-figure standing for peace has symbolic valence, that is, it is interpreted conventionally for peace as a result of some historically or culturally based event.

Signs and Meaning

Semiotic inquiry is guided by two fundamental models of the sign – Saussurean and Peircean. Saussure saw the sign as a binary form with two components – a physical part, such as the sounds that make up the word cat, which he called the signifier, and the mental concept that the sign elicits, which he called the signified (de Saussure, 1916: 101). He claimed, moreover, that a sign is created purely by convention and its connection to a referent is arbitrary. As discussed above, Peirce saw some signs as originating instead in perceptual modeling tendencies (icons). In contrast to Saussure, he claimed that iconicity is a basic tendency in sign-creation (known as semiosis) and that conventional and arbitrary semiosis are a later development, phylogenetically and ontogenetically. Peirce also put forward a triadic model of the sign – a structure consisting of the actual physical sign form, called the representamen, the thing to which it refers, termed the object, and the meanings that it elicits in real-world situations, called the interpretant (Peirce, 1931–1958: vol. 2, 247–248, 307). The interpretant is itself a derived mental sign, because it entails the further production of meanings (senses) arising from the context in which the sign is used. In our culture, a cat is considered to be a domestic companion, among other things; in others, it is viewed primarily as a sacred animal; and in some others, it is considered to be a source of food. Thus, while the sign refers to virtually the same mammal in different cultural contexts (no matter what name is used), its interpretant varies considerably.

A major problem for semiotics is to determine what meaning is and how it manifests itself through signs. To avoid ambiguity, the terms semiosis and signification are used by semioticians. The former refers to the psychological process itself of producing and recognizing specific forms as signs (Fisch, 1978: 32, 41); the latter refers to the mental images that crop up in the brain when a sign is used. Signification involves two main meaning-making processes – reference and sense. Reference is the process itself of identifying what the sign form captures; sense is what that form elicits psychologically, historically, and socially. Signs (such as words) may refer to the same (or similar) things, known as referents, but they have different senses. For example, the ‘long-eared, short-tailed, burrowing mammal of the family Leporidae’ can be called rabbit or hare in English. Both words refer essentially to the same kind of mammal. However, there is a difference

of sense between the two words – hare is the more appropriate term for describing the mammal if it is larger, has longer ears and legs, and does not burrow. Another difference is that a rabbit is often perceived to be a pet, while a hare is unlikely to be perceived as such (Danesi, 2007: 34).

The terms denotation and connotation are preferred to reference and sense. Consider, again, the word cat. The word elicits an image of a ‘creature with four legs, whiskers, retractile claws’, etc. This is its denotative meaning. As this shows, the function of denotation is referential – that is, it allows users of the sign to determine if something real or imaginary under consideration is a cat or not. Denotative meaning thus divides the world of reference into yes–no domains – something is either a cat or it is not. All other meanings of the word cat are connotative – “He’s a real cool cat,” “She let the cat out of the bag,” and so on. These meanings are products of the historical associations forged between cats and socially significant concepts or processes. Connotation thus connects the world of reference to larger historical processes. Connotation is not an option, as some traditional philosophical and linguistic theories of meaning continue to sustain to this day; it is something we are inclined to extract from a sign. The numbers 7 and 13 have specific denotative meanings – each digit stands for a specific quantity of things. However, in our culture, both invariably reverberate at the same time with connotative meanings such as fortune, destiny, bad luck, and so on. In 1957, the psychologists Osgood, Suci, and Tannenbaum showed how connotative meaning is built into everyday concepts by using a technique that they called the semantic differential. This consists in asking a series of evaluative questions to subjects about a particular concept – Is X good or bad? Should Y be weak or strong? etc. – which they are then told to rate on seven-point scales. The ratings are collected and analyzed statistically in order to sift out any general pattern that they might bear. Suppose that subjects are asked to rate the concept “ideal American president” in terms of the following scales: Should the president be: (1) Young or old? (2) Practical or idealistic? (3) Modern or traditional? (4) Male or female? and so on. A subject who feels that the president should be more youngish than oldish would place a mark toward the young end of the top scale; one who feels that a president should be bland, would place a mark toward the bland end of the attractive–bland scale; and so on. Research using the semantic differential has shown that connotation is invariably culture specific.

Structure

In order to extract meaning from a sign, sign system, or sign assemblage (text), one must be able to recognize it as such in the first place. This means that it must have

structure. Specifically, a sign (or sign assemblage such as a text) is recognizable as such if: (1) it is physically distinctive and (2) it is organized in a patterned way. The former implies paradigmatic and the latter syntagmatic structure. The phonic difference between cat and rat, for example, is a paradigmatic feature, allowing the native speaker to perceive the difference as a cue to meaning differentiation. Paradigmatic structure, as this simple example shows, involves some form of meaning-bearing differentiation in the physical makeup of a sign. In music, major and minor chords of the same key are perceivable as significantly different on account of a half-tone difference in the middle note of the chord; raising the index and middle fingers in a vertical orientation can mean victory, or peace (among other meanings), but aiming the same two fingers in a horizontal way at someone would be interpreted instead as a threat; and so on. Now, the way in which the sounds of the word cat are put together is a further cue as to its signifying value. Simply stated, the sounds of the word have been put together in line with English syllable structure. This is an example of syntagmatic structure. In music, for instance, a melody is recognizable as such only if the notes follow each other in a certain way (e.g., according to the rules of harmony); writing words in English involves putting the letters in a left-to-right form; and so on. In essence, a sign is perceivable as a meaning-bearing form if it has paradigmatic (differential) and syntagmatic (combinatorial) structure.

Another main tenet of semiotics is signs can vary in size, that is, in actual physical form. A sign can thus be something small, such as two fingers raised in a V-form; or it can be something much larger, such as a mathematical equation or a narrative. The interpretation of the sign is in all cases holistic, not discrete (decoded in terms of its constituent parts). If asked what $c^2 = a^2 + b^2$ means, a mathematician would say that it stands for the Pythagorean theorem, not for specific digits captured by the letters used (even if this is also true). If we ask someone who has just read a novel what it was all about, we would receive an answer that reveals a perception of the novel as a form containing a singular (holistic) message or purpose. The larger signs constructed by combining smaller ones (such as words in the novel) are called texts and the meanings we extract from them are called messages, rather than just signifieds. The term text embraces such things as conversations, letters, speeches, poems, myths, novels, television programs, paintings, fashion styles, scientific theories, mathematical equations, musical compositions, and so on. Texts are composite structures (signs made up of smaller signs) that are not interpreted in terms of their constituent parts, but holistically as single meaning-bearing structures.

Texts function primarily as representation devices, intended to relate, depict, portray, or reproduce something perceived, sensed, imagined, or felt as requiring signification by means of a complex (combinatorial) form.

This implies knowledge of how signs cohere into systems or codes. To enter into a conversation, for example, one would need to know the language code involved – the system of sounds, words, syntactic rules, etc. that allows for conversation to occur. Language, dress, music, and gesture, among many other things, are examples of codes. These can be defined, more formally, as systems of signs that can be used over and over to encode and decode texts. There are many kinds of codes. For example, intellectual codes contain signs in them (numbers, words, symbols, etc.) that allow for representational activities of a logical, mathematical, scientific, or philosophical nature. Social codes (dress, gender, food, space, etc.) contain sign structures for making messages about oneself in socially appropriate ways and for regulating interpersonal activities. Food codes, for example, underlie how people interpret certain foods as signifiers of various rituals, meanings, etc. It is important to note that once a text is constructed, it takes on its own paradigmatic and syntagmatic properties – that is, it is differentiable from other texts in terms of the kinds of signs that constitute it (verbal, pictorial, etc.), and it is the result of specific syntagmatic properties associated with the code or codes utilized to construct it (language, narrative, and so on).

To identify forms as meaning-bearing structures, Saussure introduced the notion of *différence*. This was expanded by the so-called Prague Circle of linguists in the 1920s and 1930s into the technique known as opposition, which has been used across disciplines, from psychology to linguistics. For the sake of historical accuracy, it should be mentioned that this notion is, implicitly, an ancient one, being characteristic of myths (good vs. evil, youth vs. old age, etc.). The Prague Circle linguist Trubetzkoy (1936, 1968), however, was the first to apply it formally to the study of the structure of language. According to psychologist Ogden (1932: 18), an early promoter of opposition theory, opposition had general applicability because it offered “a new method of approach not only in the case of all those words which can best be defined in terms of their opposites, or of the oppositional scale on which they appear, but also to *any* word.” In the 1930s and 1940s, semioticians started noticing that oppositional structure was not confined to language. It cropped up in the analysis of nonverbal systems and codes as well – in mathematics, for example, fundamental oppositions include positive versus negative, odd versus even, and prime versus composite; in music, they include major versus minor and consonant versus dissonant; and so on.

The Prague Circle linguists also argued that there were orders of oppositions that went beyond the purely binary. In mathematics, for example, the addition-versus-subtraction opposition is a basic or first-order one, while the multiplication-versus-division opposition is a derived or second-order one – since multiplication is repeated addition and division repeated subtraction. French semiotician

Algirdas J. Greimas later introduced the notion of the semiotic square to connect sets of oppositions (Greimas, 1987). Given a word such as *rich*, Greimas suggested that we determine its overall meaning by opposing it to its contradictory, not *rich*, its contrary, *poor*, and its contradictory, not *poor*, in tandem. Moreover, as work with the semantic differential showed in the 1950s (mentioned above), there seem to be gradations within the binary oppositions themselves, which are due to culture-specific connotative processes. Anthropologist Lévi-Strauss (1958) further expanded opposition theory by showing that pairs of oppositions often cohere into sets forming recognizable units. In analyzing kinship systems, he found that the elementary unit of kinship was made up of a set of four oppositions: brother versus sister, husband versus wife, father versus son, and mother's brother versus sister's son. Lévi-Strauss suspected that similar sets, or orders, characterized oppositions in other cultural systems.

Although the idea goes right back to Aristotle, the cognitive importance of opposition was noticed by the psychologists Wilhelm Wundt (1832–1920) and Edward B. Titchener (1867–1927), who were the ones to term it opposition. Its use in semiotics and linguistics is due, as mentioned, to the Prague Circle linguists (Hjelmslev, 1939; Benveniste, 1946; Jakobson, 1939). The theory was criticized from the start as being useful but highly artificial and not really consistent with human psychology. However, already in the 1940s, Jakobson (1942) empirically showed that the theory of opposition predicts the sequence of phone acquisition in children, since he found that the sound oppositions that occur frequently are among the first ones learned by children, while those that are relatively rare are among the last ones to be acquired by children.

Another early criticism of opposition theory was the fact that it seemed to neglect the importance of another structural relation that is vital to understanding semiosis – associative structure. The study of such structure came, actually, to the forefront in the latter part of the 1970s, becoming a major trend within linguistics itself (Pollio, *et al.*, 1977; Lakoff and Johnson, 1980, 1999; Fauconnier and Turner, 2002). American linguist George Lakoff and philosopher Mark Johnson set the ball rolling by claiming that a simple linguistic metaphor such as “My friend is a gorilla” is really not an idiomatic exception to basic literal meaning, but a systematically formed utterance; it is a token of a cognitive associative structure that they called a conceptual metaphor (Lakoff and Johnson, 1980). This is why we can also say that John or Mary or whoever we want is a gorilla, snake, pig, puppy, and so on. Each specific linguistic metaphor (“John is a gorilla,” “Mary is a snake,” etc.) is an instantiation of a more general associative cognitive structure – people are animals. The question becomes: How are such concepts formed? They are formed through

image schemata, as the two scholars argued (Lakoff, 1987; Johnson, 1987). The source for the people are animals conceptual metaphor seems to be an unconscious perception that human personalities and animal behaviors are linked in some way. In other words, it is the consequence of an ontological opposition: humans as animals. This is a perfect example of how opposition manifests itself as an associative mechanism, not just a binary or multi-order one (as discussed above). In this case, the two poles in the opposition are not contrasted (as in night versus day), but equated: humans as animals. This suggests that oppositional structure operates in a noncontrastive way at the level of figurative meaning.

Post-Structuralism

The most severe critiques of opposition theory have revolved around the relative notion of markedness (Tiersma, 1982; Eckman *et al.*, 1983; Andrews, 1990; Battistella, 1990). In oppositions such as night versus day, the question becomes: Which of the two poles is the more basic one (culturally or psychologically)? This is called the unmarked pole (the default one) and the other pole, the marked one (the one that stands out). It is easy to identify day as the unmarked form and night as its marked counterpart, since generally speaking we sleep at night and carry out conscious activities in the day. For example, we ask, “How many days are left before your birthday?” not “How many nights are left before your birthday?” Now, the problem is deciding which pole is marked and unmarked in a problematic opposition such as the male-versus-female one. The answer seems to vary according to the social context to which the opposition is applied. In patrilineal societies the unmarked form is male; but in matrilineal ones, such as the Iroquois one (Alpher, 1987), it appears to be female. Markedness, thus, seems to have social consequences. The late French semiotician-philosophers Michel Foucault (1926–84) and Jacques Derrida (1930–2004) claimed that this whole approach to sign study is flawed, which in itself is the unwitting source of social inequalities (Foucault, 1972; Derrida, 1976). Their refutations led to the movement known as post-structuralism, starting in the late 1950s and gaining prominence in the 1970s. In this movement, the oppositions identified by linguists and semioticians are seen as the result of an endemic logocentrism on the part of the analyst, not the result of some tendency present in the human brain. Saussure claimed that every sign was understandable in terms of its difference from other signs. In contrast to Saussure's idea of *différence*, Derrida coined the word *différance* (spelled with an ‘a,’ but pronounced in the same way), to intentionally satirize Saussurean theory. With this term, Derrida aimed to show that Saussure's

so-called discoveries were unwitting, biased ones, because a science of language cannot succeed since it must unfold through language and thus partake of the slippage (as he called it) it discovers.

Derrida argued further that all sign systems are self-referential – signs refer to other signs, which refer to still other signs, and so on *ad infinitum*. Thus, what appears to be descriptive or objective to the structuralist turns out to be illogical and paradoxical. Post-structuralism has had a profound impact on many fields of knowledge, not just semiotics and linguistics. As written language is the fundamental condition of knowledge-producing enterprises, such as science and philosophy, post-structuralists claim that these end up reflecting nothing more than the writing practices used to articulate them. However, in hindsight, there was (and is) nothing particularly radical in post-structuralism. Already in the 1920s, Jakobson and Trubetzkoy started probing the relativity of language oppositions in the light of their social and psychological functions. Basing their ideas in part on the work of German psychologist Karl Bühler (1879–1963), the Prague Circle linguists posited that language categories mirrored social ones. The goal of a true semiotic science, they claimed, was to investigate the isomorphism that manifested itself between sign and social systems. In other words, opposition theory was the very technique that identified social inequalities, not masked them.

Concluding Remarks

Semiotics is ultimately a form of inquiry into how humans shape raw sensory information into knowledge-based categories through sign creation, that is, by using forms that stand for the categories. Signs are selections from the flux of information that is taken in by our senses allowing us to encode what we perceive as meaningful in it and, thus, to remember it. We are born into a system of sign use, called the semiosphere by the late Estonian semiotician Juri Lotman (1922–1993) – that will largely determine how we come to view the world around us (Lotman, 1991). The semiosphere, like the biosphere, regulates human behavior and shapes social evolution. However, although they can do little about the biosphere, humans have the ability to reshape the semiosphere any time they want. This ability to create new signs and sign systems is what distinguishes human semiosis from all other kinds of sign and signal systems in other species. Our textual resources, for instance, stimulate us to seek new meanings and new ways of seeing the world. These open up the mind, encourage creativity, and engender freedom of thought. As Charles Peirce often wrote in his correspondence, it would seem that as a species we are inclined to “think only in signs.”

See also: Critical Ethnography; Ethnography.

Bibliography

- Alpher, B. (1987). Feminine as the unmarked grammatical gender: Buffalo girls are no fools. *Australian Journal of Linguistics* 7, 169–187.
- Andrews, E. (1990). *Markedness Theory*. Durham: Duke University Press.
- Battistella, E. L. (1990). *Markedness: The Evaluative Superstructure of Language*. Albany, NY: State University of New York Press.
- Benveniste, E. (1946). Structure des relations de personne dans le verbe. *Bulletin de la Société de Linguistique de Paris* 43, 225–236.
- Danesi, M. (2007). *The Quest for Meaning: A Guide to Semiotic Theory and Practice*. Toronto: University of Toronto Press.
- Deely, J. (2001). *Four Ages of Understanding: The First Postmodern Survey of Philosophy from Ancient Times to the Turn of the Twentieth Century*. Toronto: University of Toronto Press.
- Derrida, J. (1976). *Of Grammatology*, Spivak, G. C. (trans.). Baltimore, MD: Johns Hopkins Press.
- de Saussure, F. (1916). *Cours de linguistique générale*, Bally, C. and Sechehaye, A. (eds.), Baskin, W. (trans.) (*Course in General Linguistics*) New York: McGraw-Hill, 1958.
- Eckman, F. R., Moravcsik, E. A., and Wirth, J. R. (eds.) (1983). *Markedness*. New York: Plenum.
- Fauconnier, G. and Turner, M. (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.
- Fisch, M. H. (1978). Peirce's general theory of signs. In Sebeok, T. A. (ed.) *Sight, Sound, and Sense*, pp 31–70. Bloomington, IN: Indiana University Press.
- Foucault, M. (1972). *The Archeology of Knowledge*, Sheridan Smith, A. M. (trans.). New York: Pantheon.
- Greimas, A. J. (1987). *On Meaning: Selected Essays in Semiotic Theory*, Perron, P. and Collins, F. (trans.). Minneapolis, MN: University of Minnesota Press.
- Hjelmslev, L. (1939). Note sur les oppositions supprimables. *Travaux de Cercle Linguistique de Prague* 8, 51–57.
- Jakobson, R. (1939). Observations sur le classement phonologique des consonnes. *Proceedings of the Fourth International Congress of Phonetic Sciences*, Ghent 31–41.
- Jakobson, R. (1942). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist and Wiksell.
- Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination and Reason*. Chicago, IL: University of Chicago Press.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago, IL: University of Chicago Press.
- Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: Chicago University Press.
- Lakoff, G. and Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.
- Lévi-Strauss, C. (1958). *Structural Anthropology*. New York: Basic Books.
- Locke, J. (1690). *An Essay Concerning Human Understanding*. London: Collins.
- Lotman, Y. (1991). *Universe of the Mind: A Semiotic Theory of Culture*. Bloomington, IN: Indiana University Press.
- Ogden, C. K. (1932). *Opposition: A Linguistic and Psychological Analysis*. London: Paul, Trench, and Trubner.
- Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The Measurement of Meaning*. Urbana, IL: University of Illinois Press.
- Peirce, C. S. (1931–1958). *Collected Papers of Charles Sanders Peirce*, In Hartshorne, C. and Weiss, P. (eds.), Vols. 1–8. Cambridge, MA: Harvard University Press.

- Pollio, H., Barlow, J., Fine, H., and Pollio, M. (1977). *The Poetics of Growth: Figurative Language in Psychology, Psychotherapy, and Education*. Hillsdale, NJ: Erlbaum.
- Tiersma, P. M. (1982). Local and general markedness. *Language* **58**, 832–849.
- Trubetzkoy, N. S. (1936). *Essaie d'une théorie des oppositions phonologiques*. *Journal de Psychologie* **33**, 5–18.
- Trubetzkoy, N. S. (1968). *Introduction to the Principles of Phonological Description*. The Hague: Martinus Nijhoff.

Further Reading

- Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fischer.

Symbolic Interactionism, Naturalistic Inquiry, and Education

J A Forte, Salisbury University, Salisbury, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Contextualism – Theoretical assumption that each moment is a unique and complex event but one situated in a historical timeline and a distinctive spatial context.

Culture – The interrelated set of perspectives that persons who interact together intensively and extensively develop to meet role and institutional challenges.

Life history – The documentation of the interplay of identity, commitments, situational demands, and strategies across a career by the use of various qualitative research tools especially interviewing to make sense of critical events and the life course.

Naturalistic inquiry – An approach to the study of self, relationship, and community processes as they occur requiring minimal interference but maximal immersion by the investigator.

Perspective – The coordinated set of images, ideas, and actions a student uses in dealing with some problematic situation.

Teaching strategy – A creative package of intentions and actions used to achieve goals and respond to the constraints experienced while enacting the teacher role.

Verstehen – Interpretive understanding.

Symbolic Interactionism

Symbolic interactionism (SI) is a theoretical perspective and school of thought created between 1890 and 1910 by philosophers, sociologists, and social workers aligned with the University of Chicago in the United States. The naming of the tradition is credited to Herbert Blumer (Forte, 2006). The perspective focuses attention on the symbolic and the interactive together as they are organized and experienced in everyday life. It looks at how the meanings attached to symbols emerge during social interaction, how agreements about meanings are negotiated, how people use these shared and conventional meanings to do things together, and how meanings are transformed over time. Interactionist theorists and researchers give attention to symbol creation, use, and change as conditioned by interaction, cultural memberships, and social order. Applied

interactionists are interested in how members of groups and organizations redefine objects and events, create new meanings, alter their plans of action, and change the direction and quality of their lives (Forte, 2003).

History

As a school, SI is a network of scholars appreciative of the legacy of the founders (George Herbert Mead, John Dewey, and Jane Addams), committed to a particular paradigm of empirical reality and a set of assumptions about scientific inquiry, and dedicated to transmitting the interactionist vision, theories, and methods across generations (Forte, 2001). Notable members of the school include W. I. Thomas, Herbert Blumer, Ernest Burgess, Robert Park, Everett Hughes, Manfred Kuhn, Carl Couch, Anselm Strauss, Norman Denzin, Helene Lopata, and Erving Goffman.

Several sub-schools have been identified (Forte, 2001). A major dividing line is the conception of the social self and science. The Chicago School affirms the theoretical dispositions of Blumer, emphasizes the human capacity for creative and novel action, and uses a set of core empirical studies, ethnographies and life histories, especially those conducted or sponsored by Burgess and Park between 1917 and 1942, as models (Deegan, 2001). The Iowa School developed under the tutelage of Kuhn theorizes about the solid, persistent aspects of self and the shaping influence of social roles and social organizations, and prefers survey and experimental methods. In the last decade, the identification of interactionists with particular sub-schools has become rare. This article characterizes the Chicago School and its contribution to qualitative research.

Assumptions

Interactionists assume that human beings have species-specific capacities, distinctive strengths (Forte, 2003). Human learners are different from and more complex than members of other species in their capacities for self-awareness and reflection, emotionality, intelligent problem-solving behavior, symbolic interaction, and tool use. Humans are capable, unlike animals, of reflective participation in membership processes and free, responsive choices. The interactionist embrace of bio-evolutionary tenets regarding human differences demands suitable methods for inquiry. Consequently, interactionists developed approaches, such

as personal document analysis, life history, and participation observation sensitive to human characteristics (Denzin, 1992). Interactionist educators also endorse this humanistic assumption. For example, Quicke (2003) asserted: "We should study the learning process from the perspective of the persons at the center of the web of interaction (pupils and teachers) and appreciate that these persons are 'reflective agents'" (p. 52).

Interactionists assume that humans live in a symbolic environment (Forte, 2003). Effective action requires the interpretation of the particular objects (persons, things, events, and other social constructions) constituting one's field of operations. Effective interaction emerges when participants in an encounter interpret each other's conduct (appearance, gestures, expressive and discursive symbols, accounts, and narratives) in a roughly consensual way allowing for the coordination of lines of action. Scientific inquiry, consequently, is a process of *verstehen* or interpretive understanding. Prus (1996), for example, advises ethnographers to use Mead's method of "taking the role of the other" to gain the intimate familiarity with subjects necessary to interpret their situated meanings. Becker summarized educational field work as interpretation: "we try to establish the meanings of various activities, objects, or words by looking at them as they occurred, as they were used, or as they were engaged in a great variety of settings" (Debro, 1986: 43). Woods (1983) research agenda includes topics such as "how teachers and pupils interpret school processes, personnel, and organization such as lessons, the curriculum, their peers, and each other?" and "What factors bear on these interpretations?" (p. 17).

Interactionists take a pragmatic position regarding knowledge production (Forte, 2003). Theory and practice cannot be separated. Knowledge is valuable to the extent that, for example, it can be used practically in social reform and advocacy campaigns. Science is conceptualized in accord with the vision of Charles Sanders Peirce, a founder of American pragmatism, as free and unrestricted communication among a community of inquirers sharing standards for investigation, committed to the revision of understandings following real-world tests, and determined to solve problems of collective concern. Consequently, scientific inquiry is judged by the pragmatic triad of critical scrutiny, experimental consequences, and moral valuation. Does a pedagogical theory satisfy a critical teacher's concerns about its reasonableness and usefulness? Will a pedagogical theory contribute to the achievement of preferred and predicted learning consequences in field experiments? Can the pedagogical theory be used in ways concordant with ethical ideals of the school and community?

Interactionists assume that theorists and researchers should be engaged and critical contributors to bettering personal, social, political, and economic processes

(Forte, 2003). Many interactionists, for example, affirm the progressivism of Dewey, Mead, and Addams. Many have indicated an affinity with the critical theory of Jurgen Habermas (Forte, 2001). Consequently, scientific inquiry takes sides. Becker (Debro, 1986), for example, rejected a definition of the educational researcher role as impartial agents of their employers indifferent to power inequities. Ethnographers should study, he suggested, "everyone connected with the school" (Becker, 1995: 118), even if this is disturbing to some teachers and administrators, and ethnographers should investigate the discrepancies between an educational establishment's goals and actual performances. Good ethnographers report the facts about the causes of school system troubles, "the existence of which no one there would admit" (Becker, 1995: 113), in ways not restricted to students' shortcomings and even when the causes incriminate conditions or processes that administrators do not want to do something about. Science should also be a process generating ideas and images that reconstruct problematic situations and make possible new solutions. Becker characterized scientific inquiry as a search for radical results, "a radical result is one that rises above current orthodoxies, whether they are political, moral, institutional, scientific" (p. 38). Such inquiry is an important prelude to creative policy advocacy.

Finally, interactionism has been built on the root metaphor or assumption of contextualism (Forte, 2006). Each moment is a unique and complex event but one situated in a historical timeline and a distinctive spatial context. Personal processes (human development, learning, and individual action) can best be understood when connected to social processes, practices, and forces as these are related to relevant cultural, ecological, and historical constraints and opportunities. Learning, for example, is considered an event that emerges out of dynamic social interaction between person, setting, and temporal contingencies. Consequently, scientific inquiry must maintain sight of foreground, the acting person, and background, the ecology. The interactionist researcher immerses himself/herself in the teachers' or students' context; appraises the interconnectedness of subjectivity, intersubjectivity, temporal factors, and situational factors; and translates the context-specific meanings for users of the research. Applied interactionists have a preference for case studies: individualized, detailed, and problem-driven accounts of what worked in particular educational contexts.

Concepts

The interactionist perspective includes a set of concepts that can sensitize theorists and researchers to the tradition's central concerns, and also guide observation and data collection in specific educational settings (Forte, 2001).

Identity-related concepts include self, I, me, meanings, identity, identity salience, identity commitments, self-esteem, significant other, reference group, and generalized other. Role-related concepts include roles, counter-roles, formal and informal roles, careers, role-taking, role-making, and socialization. Interaction-related concepts include self-presentation, impression management, significant others, definition of the situation, negotiation, working consensus, aligning actions, strategies, and vocabulary of motives. Context-related concepts include negotiated order, meso-structure, interaction order, symbolic order, and social worlds.

Scholarly Association

The traditions of SI are now celebrated and refined by members of the Society for the Study of Symbolic Interaction (SSSI). This association is an international organization of scholars and practitioners interested in the study of a wide range of issues from the interactionist perspective. The society holds an annual meeting that includes paper presentations, distinguished lectures, business gatherings, and an awards ceremony. The journal, *Symbolic Interaction*, and the research annual, *Studies in Symbolic Interaction*, are the major vehicles for the dissemination of reports on research, theory, and practice.

Exemplary Models

Three interactionists have special value as exemplars to educators and educational researchers. These exemplars represent three different generations.

George Herbert Mead

Recent historical scholarship has documented that Mead, the originator of the tradition, contributed significantly to educational practice, theory, and research (Biesta and Trohler, 2008; Deegan, 1999). Mead played a leadership role in many civic organizations and committees devoted to educational reform. He was a member of City Club of Chicago for almost 20 years and chairman of the Club's Committee for Public Education. He served as president of the Parent's Association for the University of Chicago's School of Education (1902–03), and worked as editor and editorial note writer for the *School Review* and *The Elementary School Teacher*. Mead taught a philosophy of education course regularly between 1905 and 1911, translating his social psychological ideas about self, meaning, and reflective consciousness into language useful to educators. George Herbert Mead was a colleague of John Dewey; and Dewey asserted that he derived many of his seminal ideas about education from exchanges with Mead.

Mead's scholarship addressed topics, such as science education, vocational training, and the role of play. For

example, Mead theorized that play was an essential ingredient in the education of children. Effective education, he argued, blends work, activity to achieve a predesignated end, and play, free and spontaneous activity leading to exploration. Mead recommended objective setting and kindergarten design organized around the use of children's imagination and capacities for spontaneous activity. Teachers should relate learning tasks to children's interests and motivations in ways that channel students' physical energies while supporting experimentation. Such an approach fosters the natural unfolding of students' potentials.

Mead developed an innovative and substantial theoretical framework. He suggested that education is a process of symbolic communication between occupants of teacher and student roles, a process dedicated to the creative formation and transformation of meaning in a way linking learner and society. Teachers should promote a method of thought, that is, scientific reflection, which prepares novice members for effective, competent, responsible, satisfying participation in groups, organizations, and institutions. The medium for teaching is communication by symbols (language, images, gestures, and objects), and this is characterized as a social, intersubjective, and cooperative partnership between teacher and student. The materials of education are more than the lesson plans and learning activities brought to the classroom by the teacher; the meanings jointly constructed during the symbolic interaction in relation to the plans and activities are the critical materials. Mead related teaching to phases of human action. The affective phase requires the activation of the learner's emotion of interest. During the esthetic phase, the teacher helps the learner to attend to and grapple with the presented image or idea and its esthetic or sensuous qualities. In the intellectual phase, the teacher promotes the analyzing and appraising of the image or idea with special reference to its specific and planned uses. Mead's pragmatic theory of consciousness informs the teaching process also. Consciousness is intermittent according to pragmatists, and consciousness is activated when inner or environmental obstacles impede forward action. A curriculum that presents intellectual problems that confound students' action inclinations will compel reflection. As a correlate, Mead adds that learning problems identified and accepted as a problem by the learner rather than imposed by an outside authority increase the reflective qualities of concentration and focus essential to the problem-solving effort.

Mead and his associates also investigated and reported on the public library system, vocational education in public schools, and on the funding and administration of local education. His sponsorship and guidance of two Chicago stockyard district studies, for example, resulted in findings about the employment opportunities of teenage dropouts, the reasons that children and parents gave

for early withdrawal from school, and parental attitudes toward school. These studies, like later interactionist research, used interview methods to solicit the perspective of their participants. Also like later pragmatic and interactionist studies, the research of Mead and his associates included suggestions for improved policies and programs such as increasing the age of compulsory education and increasing opportunities for industrial education.

Howard Becker

Becker, a second-generation interactionist, began his study of education in the doctoral program at University of Chicago and became a diligent interactionist. In his lifetime, he produced “a series of articles on schools and schooling that influenced generations of researchers in Britain and the USA” (Burgess, 1995: 1). The pioneering *Boys in White* (Becker *et al.*, 1961) study investigated the student culture of medical school and the transition of interns from novice to competent status. *Making the Grade* (Becker *et al.*, 1968) examined and detailed the perspectives used by college students to deal with academic work in the context of college.

Becker’s conceptualization of perspectives, culture, and situational adjustment are significant contributions to educational theory (Burgess, 1995). Becker characterized a perspective as the coordinated set of images, ideas, and actions a student uses in dealing with some problematic situation (Becker *et al.*, 1961). For example, medical students have more schoolwork than available time. Many develop a clinical experience perspective. The medical student perceives that clinical experience is the essential prerequisite for a career as a doctor. School activities that provide clinical experience are judged desirable and prioritized. Progress depends on attention to and the accumulation of clinical lessons; and other educational obligations merit minimal effort. Culture, Becker theorized, is one way to understand the relationship of person and organization. Culture is defined as the interrelated set of perspectives that persons who interact together intensively and extensively develop to meet role and institutional challenges (Becker *et al.*, 1968). It helps students coordinate their actions with administrators, teachers, and staff. Once developed, the culture provides situational definitions, understandings of problems and possible solutions, and guidelines regarding preferred ways of prevailing against obstacles. Those who are socialized into a culture relate to the larger organization in a somewhat predictable way. Becker discovered, for example, that culture directs and regulates the degree of student effort. Culture provides a framework for action, permitting creative adjustment to organizational constraints but allowing deviance from rules that does not result in academic failure. Becker’s labeling approach to deviance also has utility for educators and education research (Burgess, 1995). Becker explained that no particular act (e.g., classroom conduct or performance) is inherently deviant. An act becomes deviant

when persons with socially powerful statuses or positions in a hierarchy of credibility label it as deviant. If the label sticks, the actor’s self-concept, interaction, and career are changed as others adjust their behavior to the new and damaging definitions.

Asked about his career, Becker (Ben-Yehuda *et al.*, 1989) asserted “most of my work has been done in education, studying educational institutions, and also in . . . qualitative methodology” (Ben-Yehuda *et al.*, 1989: 485). About design, Becker suggested that theory and method should inform but not dictate a researcher’s approach. His team (Becker *et al.*, 1961), for example, started the medical school study embracing SI theoretical assumptions. Rather than developing detailed research agendas or hypotheses, the team agreed to examine the matters of importance to the people studied. Becker contrasted the fixed design of quantitative methodologies to the flexibility of field research, and argued that the “main advantage of so-called qualitative methods is that you get to change your mind . . . if I learn something useful from my field research, my observation, I can go out tomorrow and use it. You lose the advantages of standardization and you gain the advantage of flexibility” (Mullan, 1996: 133). Becker also valued the hard-earned knowledge obtained through participation observation. He said that “the hardest, most tedious work” often results in the amalgamation of all the materials gathered into compelling empirical arguments and valuable insights into group cultures (Mullan, 1996: 142). Becker is credited with teaching researchers how to make the familiar strange when studying schools. He also inspired extensive inquiry into how teachers observe, classify, react to, and label pupil behavior (Burgess, 1995).

Peter Woods

Woods is a contemporary interactionist with a long record of scholarship relevant to education. From a base of interactionist theory and ethnography, he has written research reports, treatises on the methodological dilemmas faced by qualitative researchers, practical guides for educators, policy analyses, and comprehensive overviews of the SI approach to education and educational research (Woods, 1983, 1996).

Woods’ (1980) major theoretical work develops the interactionist approach to strategies. Teaching strategy, he suggests, is a creative package of intentions and actions used to achieve goals and respond to the constraints experienced while enacting the teacher role. The effective use of strategies enhances a teacher’s adjustment to and survival in the organization. The construction of a teaching strategy is individually motivated and influenced by teacher self-awareness, identity, and commitment to the profession. Each teacher selects and uses teaching strategies, however, in ways influenced by the pool of available cultural symbols, the stability of daily educational situations, support from others, and characteristics of the school and larger social

structure. Woods has conducted much research and supported the research of others on teaching strategies. For example, Woods summarized his findings as a typology of teaching strategies, including domination, negotiation, fraternization, absence, removal, ritual, routine, and moral boosting. Careful observation also revealed the patterned, coherent, and relatively permanent nature of many teachers' choice of preferred strategies.

Woods has also been an active editor and interactionist researcher. With Hammersley, he provided a collection of ethnographic studies of students' interpretations of teachers, pedagogy, and school organization (Hammersley and Woods, 1984). He has pioneered the use of interactionist life histories for purposes of teacher development (Woods, 1993). A life history uses various research tools, especially interviewing to make sense of critical events and the life course. For teachers, a life history can document the interplay of identity, commitments, situational demands, and strategies across a career. Life histories have restorative functions. A life history can help a teacher better understand the distinctive and common aspects of his or her disciplinary knowledge, the transformative processes leading to the contemporary self, the meaning of critical career events and turning points, and the influence of organizational conditions (e.g., a school's gender- and race-based structure) on personal and professional development.

Naturalistic Inquiry and Educational Research

The Blumer (1969) and Denzin (1970, 1971) methodological stance and set of methodological directives undergird the preferred interactionist approach to science, naturalistic inquiry. Qualitative research, ethnography, and interpretative research are terms associated with this approach. Blumer (1969) conceived of social science as a process yielding "verifiable knowledge of human group life and human conduct" (p. 21). Due to the distinctive symbol-creating and-using capacities of humans, Blumer argued, the scientific study of human life requires a methodology radically different from that used by positivist social science. To "respect the nature of the empirical world" (Blumer, 1969: 60), naturalistic inquiry calls for the collection of naturally occurring data in naturally occurring situations for natural time spans by a naturalistic researcher.

Naturalistic inquiry studies self, relationship, and community processes as they occur when there is minimal interference but maximal immersion by the investigator (Denzin, 1970). Mechanical and intrusive investigative methods used by an authoritative social scientist will not capture the emergent and continually constructed nature of social reality. Instead, the researcher attempts to become an insider or group member. George Herbert Mead, John Dewey, W. I. Thomas, and other educators at the University

of Chicago Department of Sociology worked closely and frequently with Jane Addams and Hull House social workers. The settlement house worker, therefore, is a role familiar to the first symbolic interactionists. The role provided a template for their conception of the research process. By living among community members, learning their customs and language, sharing meals with them, participating in cultural traditions, and honoring their significant symbols and practices, these social workers became privy to and participants in diverse social worlds. They shared their knowledge with the larger public to gain sympathy and support. Immersion of this kind increases the likelihood that the naturalistic researcher will witness and understand central interaction processes, such as stage setting, self-presentation, perspective taking, meaning negotiation, novice socialization, culture creation, policymaking, and problem solving. Role boundaries and distinctions dissolve, and the researcher intentionally becomes a native.

Naturalistic inquiry prefers research sites in the community, places where group members act, interact, and develop cultures rather than in scientific laboratories. Blumer (1969) contended that actors could not be understood independently of their social context, "the world of objects, the sets of meanings, and the schemes of interpretation that they already possess" (p. 20). For interactionists, actors act according to the meanings assigned to objects composing their particular fields of interaction. Group patterns are made intelligible and the veil covering the subjects' interpretations can best be lifted *in situ* (Blumer, 1969; Denzin, 1970). Naturalistic inquiry can be compared to the acquisition of a new language and the translation of this language to nonspeakers. Language mastery occurs fastest when the researcher obtains an invitation to the language user's home, workplace, neighborhood, and other locales. Through careful and tactful observation of language use, flexible and conversational interviewing, and participation in daily routines, the researcher learns the language: the system of significant symbols, for example, related to self-conceptions, grades, teachers, school lunch, classrooms, or educational policies. The research report is a translation from the natural language of the members studied to the language of the science-speaking audience.

Naturalistic inquiry requires a time commitment because insight into meaningful symbols and interaction patterns develop slowly. The point-in-time approach of surveys or short-term engagement in most experimental studies is rejected. As each educational setting is a somewhat unique configuration of selves, temporality, and place, the researcher can only make minimal use of prior knowledge and needs time to grasp the distinctive features of the setting. Taking the other's role and then, viewing the world from the other's point of view enable the educational researcher to achieve a depth of understanding. However, achieving the degree of closeness and trust necessary for perceptive and accurate role does not occur

quickly. Symbolic interactionists often produce natural histories. Ethnographies that extend months or years make possible discoveries about meaning-making across time as related to biological aging, developmental progression, learning achievements, student and teacher careers, educational sequences, problem formulation and resolution steps, and mental health trajectories (Denzin, 1971).

Naturalistic inquiry requires a particular stance toward the research role and approach to data collection and analysis, the grounded theory method (Glaser and Strauss, 1967). The naturalistic researcher is the research instrument rather than a survey or standardized measure, and the researcher reflects continuously on the fit of his or her lines of action with the study subjects, and adjusts his or her self, self-presentation, and action accordingly. The researcher maintains a state of responsiveness to the values and influences emerging naturally in symbolic interaction at the research site. Theorizing does not precede the data gathering and possibly render artificial findings. Data collection and data analysis are not separated as in positivist methods. The naturalistic researcher generates theory from the qualitative data with the theorizing and analysis grounded at every stage of the research process in real-world observations (Lincoln and Guba, 1985).

SI, Naturalistic Inquiry, and Qualitative Methods: Conclusion

The charge of this article has been to demonstrate the relevance of SI as a paradigm for educators and educational researchers. Space limits preclude discussion of theoretical and research advances into the micro-aspects of conduct (cognition and emotion) or the macro-processes and structures conditioning meaning-making (power, class, and culture). Interactionist's alignments with other traditions, including semiotics, critical theory, Vygotsky's sociocultural approach, and feminism as well as the recent interactionist responses to criticisms of the naturalist approach (e.g., issues of generalizability, utility for policy advocacy and practice, and the rhetoric of science writing), cannot be reviewed.

SI represents a distinctive tradition of scientific inquiry and applied science offering educators an assumptive base supportive of interpretive, engaged, and practical theorizing about human subjectivity and situated conduct; an opus of work rich with grounded theories, methodological innovations, and naturalistic studies; a set of exemplary theorists and researchers; a framework for the amelioration of educational problems and betterment of educational institutions; and an approach to inquiry that can make sense of the qualities (immediacy, complexity, uniqueness, and symbolic mediation) related to the perspectives, actions, and interaction of members of diverse learning communities.

See also: Grounded Theory; Interpretive Research; Pragmatism.

Bibliography

- Becker, H. S. (1995). Studying urban schools. In Burgess, R. G. (ed.) *Howard Becker on Education*, pp 113–122. Buckingham: Open University Press.
- Becker, H. S., Geer, B., and Hughes, E. C. (1968). *Making the Grade: The Academic Side of College Life*. New York: Wiley.
- Becker, H. S., Geer, B., Hughes, E. C., and Strauss, A. (1961). *Boys in White: Student Culture in Medical School*. Chicago, IL: University of Chicago Press.
- Ben-Yehuda, N., Brymer, R. A., Dubin, S. C., et al. (1989). Howard S. Becker: A portrait of an intellectual's sociological imagination. *Sociological Inquiry* 59(4), 467–489.
- Biesta, G. and Trohler, D. (2008). George Herbert Mead and the development of a social conception of education. In Biesta, G. and Trohler, D. (eds.) *The Philosophy of Education: George Herbert Mead*, pp 1–16. Boulder, CO: Paradigm.
- Blumer, H. (1969). *Symbolic Interactionism*. Englewood Cliffs, NJ: Prentice-Hall.
- Burgess, R. G. (1995). Critical introduction: Problems, theories and methods in the work of Howard Becker. In Burgess, R. G. (ed.) *Howard Becker on Education*, pp 1–27. Buckingham: Open University Press.
- Debro, J. (1986). Dialogue with Howard S. Becker (1970). In Becker, H. S. (ed.) *Doing Things Together: Selected Papers*, pp 25–46. Evanston, IL: Northwestern University Press.
- Deegan, M. J. (ed.) (1999). *George Herbert Mead: Play, School, and Society*. New York: Peter Lang.
- Deegan, M. J. (2001). The Chicago school of ethnography. In Atkinson, P., Coffy, A., Delamont, S., Lofland, J., and Lofland, J. (eds.) *Handbook of Ethnography*, pp 11–25. London: Sage.
- Denzin, N. K. (1970). *The Research Act: A Theoretical Introduction to Sociological Methods*. Chicago, IL: Aldine.
- Denzin, N. K. (1971). The logic of naturalistic inquiry. *Social Forces* 50, 166–182.
- Denzin, N. K. (1992). *Symbolic Interactionism and Cultural Studies: The Politics of Interpretation*. Cambridge, MA: Blackwell.
- Forte, J. A. (2001). *Theories for Practice: Symbolic Interactionist Translations*. Lanham, MD: University Press of America.
- Forte, J. A. (2003). Applied symbolic interactionism: Meanings, memberships, and social work. In Reynolds, L. T. and Herman, N. J. (eds.) *Handbook of Symbolic Interactionism*, pp 915–936. Walnut Creek, CA: Altamira Press.
- Forte, J. A. (2006). *Human Behavior and the Social Environment: Models, Metaphors, and Maps for Applying Theoretical Perspectives to Practice*. Belmont, CA: Thomson Brooks/Cole.
- Hammersley, M. and Woods, P. (1984). *Life in Schools: The Sociology of Pupil Culture*. Stony Stratford: Open University Press.
- Lincoln, Y. S. and Guba, E. G. (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage.
- Mullan, B. (1996). *Sociologists on Sociology*, 2nd edn. Aldershot: Avebury.
- Prus, R. (1996). *Symbolic Interaction and Ethnographic Research: Intersubjectivity and the Study of Human Lived Experience*. Albany, NY: State University of New York Press.
- Quicke, J. (2003). Educating the pupil voice. *Support for Learning* 18(2), 51–57.
- Woods, P. (1980). Strategies in teaching and learning. In Woods, P. (ed.) *Teacher Strategies: Explorations in the Sociology of the School*, pp 18–33. London: Croom Helm.
- Woods, P. (1983). *Sociology and the School: An Interactionist Viewpoint*. London: Routledge and Kegan Paul.
- Woods, P. (1993). Managing marginality: Teacher development through grounded life history. *British Educational Research Journal* 19(5), 447–465.
- Woods, P. (1996). *Researching the Art of Teaching: Ethnography for Educational Use*. London: Routledge.

Further Reading

- Baugh, K., Jr. (1990). *The Methodology of Herbert Blumer: Critical Interpretation and Repair*. New York: Cambridge University Press.
- Charmaz, K. (2005). *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Newbury Park, CA: Sage.
- Denzin, N. (1989). *Interpretive Interactionism*. Newbury Park, CA: Sage.
- Lofland, J., Snow, D., and Anderson, L. H. (2005). *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*, 4th edn. Belmont, CA: Wadsworth.
- O'Donoghue, T. O. and Punch, K. (2003). *Qualitative Educational Research in Action: Doing and Reflecting*. London: RoutledgeFalmer.
- Plummer, K. (2001). *Documents of Life 2: An Invitation to Critical Humanism*. London: Sage.

Relevant Websites

- <http://www.unm.edu> – Qualitative Research in Education, The University of New Mexico.
- <http://www.espach.salford.ac.uk> – Society for the Study of Symbolic Interaction, ESPaCH – School of English, Sociology, Politics and Contemporary History.
- <http://www.und.nodak.edu> – Teaching & Learning: *The Journal of Natural Inquiry and Reflective Practice*. Grand Forks, ND: College of Education and Human Development, University of North Dakota.
- <http://www.brocku.ca> – The Mead Project 2.0, Brock University.

The Quality of Evidence in Qualitative Research

S K Biklen, Syracuse University, Syracuse, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Assessing the quality of evidence in qualitative research is not straightforward. While scholars do say, “This is what good evidence looks like” and “This is what it is not like” – suggesting that data that are descriptive, detailed, and complex are preferable to field notes that are thin, summarizing, and superficial – these statements are limited. There is a complicated and tension-filled relationship between data and evidence, the context for claiming data as evidence of a particular finding or analytic is always significant, and evidence looks different and may take different forms in the many kinds of scholarship that are included under the umbrella term qualitative research including ethnography, life history, narrative inquiry, and visual methods.

Since qualitative approaches study well the complexity of everyday life, any discussion of evidence must address the complications involved in representing complexity. Funders of major qualitative-research projects may typically want more definitive findings from qualitative researchers than they are sometimes prepared to give, and they want it in the form of useful nuggets that definitively point to specific truths. While published qualitative studies often do make straightforward claims about their focus, equal numbers of studies articulate processes that occurred in complicated ways over time in a setting, capturing the complications rather than highlighting specific claims. Both sorts of projects involve evidence, but the evidence is represented in contrasting ways. In addition, qualitative researchers often must account for conditions where many more people in a setting will assert a particular truth while fewer people who articulate views quite contrary to the first position are closer to what is going on and know more about it. In other words, sometimes the most insightful informants are not the most numerous. Researchers must continually make judgments about how to understand and evaluate what people say.

Discussions of the quality of evidence in qualitative studies, then, in relation to fieldwork, involve wide-ranging concerns such as whether or not the fieldworker trusts the research participants, how to account for what participants assume and know but do not talk about, how what informants do not know about their subjectivities may be a dominant issue, what the particular investments are that research participants have in their perspectives, and what discourses are prominent and available for them as resources – and these are only a sample of concerns. These caveats do not suggest that researchers should retreat from discussions of what constitutes evidence, because assuming the transparency of research participants’

meanings and leaping too quickly to assume understanding of what informants say is dangerous. But we must find ways of grappling with these concerns without developing too narrowly conceived criteria that mimic a positivist view of the world. Evidence is political as well as epistemological and politics must always be taken into account.

Evidence is political on several different levels. First, evidence is political in the sense that it is related to the state. There are particular kinds of evidence that the state supports when it promotes certain forms of research. The state’s interest over the first 10 years of the twenty-first century in randomized clinical trials, for example, that has been so central to defining the evidence-based research movement, has meant that it has little interest in funding qualitative projects, seeing them as too messy for its interests. In this sense, politics references the power of the government to determine what counts as evidence because it limits what methods can be used to collect it. Politics shapes what vested groups want studied, and hence what evidence will even be considered.

Second, evidence is also political because it relates to power relations. From this perspective, the politics of evidence refers to concerns such as whose perspectives qualitative researchers study, who is considered an expert, how language is understood, and the investments people have in particular readings of evidence. Evidence, in other words, is not value neutral. As Denzin (2009) has written, “standards for assessing quality are forms of interpretive practice that enact a politics of evidence and truth” (p. 139).

The question of evidence in qualitative research, in other words, is complex. Its consideration in the postpositivist moment, where criteria for standards for proof in qualitative research are questioned, demands the negotiation of many tensions: what it means to be in the field when the field might refer to someone’s office where an interview takes place; what different groups want qualitative methods to do for them; how audiences understand the question of voice in relation to representation; the dangers and attractions of particular forms in relation to evidence; and the question of theory. This article examines the question of evidence in qualitative methods from an historical perspective, illustrating shifting ideas of evidence and the vocabularies in use; discusses the investments that different groups have in how the social reality of qualitative research in education is constructed (these are not separate tasks, even if they appear so here); and engages some criteria for evaluating evidence as part of the empirical research process.

What Is Evidence?

Evidence for a claim, an argument, or even a portrait involves and relies on data, but data and evidence are not the same. Data are necessary for evidence, but they only become evidence when they relate to a specific question or concern: “Evidence... is data brought to bear on specific questions, theories or experiences. Evidence is data with a purpose. The purpose may be historical, theoretical, evaluative or descriptive, but it is arrayed to some larger end, and it is arrayed in a specific kind of order. Evidence represents data to which have been added a layer – or multiple layers – of interpretation...” (Lincoln, 2002: 5). Data become evidence not only in relation to the question investigated but also in relation to the paradigm or theoretical orientation (articulated or not) that animates the research project. Evidence is structured data. Evidence is always related to interpretation but evidence is of varying quality, and the variation and the interpretive aspects must always be addressed.

To articulate that evidence is data interpreted – or even that observations or interview transcripts must be interpreted to be named as data – does not indicate clearly how the process of interpretation occurs. It might sound as if the frameworks that qualitative researchers use for interpretation determine what the researcher will notice and hence doing that fieldwork comes to equate filling in the empty spaces of a view of the world that is already known to the researcher. Interpretation covers basic processes such as recognizing and considering, where recognizing refers to the researcher’s decision that something seen was worth recording as data, and considering includes the work of figuring out what something recorded as data might mean. At the same time, researchers who spend significant time in the field frequently come to see what they did not expect, or learn a way of making sense of the world that was not previously on their radar. In fieldwork, researchers may hear what is new to them. But they are positioned for this to occur. Finding evidence involves a process of negotiation, a kind of encounter between researchers with specific social locations that orient them to the world in certain and specific ways and those whose ways of seeing the world they try to account for and reflect on. This process of reflection includes imagining alternative ways that readers will interpret an analysis and working to account for them or respond to them.

Contexts

What is it that we seek evidence about? The topics that qualitative researchers study in relation to education travel an extremely broad range, from educational reforms, to student or teacher identity, literacy curricula, punishment, to classroom innovation, to white privilege and color-blind

ideology, and to questions of identity, discomfort, and voice. But whatever is studied, as Mischler (1979) so cogently put it, for the qualitative researcher, meaning must always be made, ascertained, analyzed, and viewed in context. People do not make meaning outside of particular contexts, but contexts are not always local. Social discourses about what is right, how the world is organized, what matters, and who is an authority, for example, are broader than particular locations even as they interact with them. So the researcher articulates and frames the context(s), within which sense making is brought to the data. Contextualizing the data helps transform it into evidence.

When Willard Waller published *The Sociology of Teaching* in 1932, he relied on qualitative data from in-depth interviews, life histories, participant observation, diaries, letters, personal documents, and case records to capture the social worlds of schools. Waller believed that “children and teachers are not disembodied intelligences, not instructing machines and learning machines, but whole human beings tied together in a complex maze of social interconnections. The school is a social world because human beings live in it” (Waller, 1932: 1). His influential work was undertaken at a time when the field of educational sociology, like the current moment in government-sponsored research, was highly quantified. Researchers tended to dismiss the use of approaches such as the life history, describing it as more suitable for social-service work in fields such as social work, rather than for sociology of education. But his approach to focus on the complicated contextualized person and situation, certainly what Mills (1959) defined as the intersection of biography and social structure, demanded an approach where evidence could build a picture.

Characteristics of Qualitative Research

Qualitative research is empirical, a method for studying social life; and what is central to it, shifts over time. When students and faculty did qualitative research during the time of the Chicago School approach, they customarily stayed in the field studying communities for 2 or 3 years (see Becker, 1970a: 53, note 20). Today, graduate students and scholars conducting funded research might interview 10 or 20 people on several occasions for their data collection and consider that a qualitative study. Some students and scholars still undertake the kind of longitudinal work that qualitative research was originally known for, but there is a much wider variation in what counts as adequate time in the field than before. What counts as adequate evidence is one criterion that is not self-evident and that shifts over time. While the life history, the study of a single person that almost takes the form of biography but is a kind of qualitative research, has been around for over 50 years, the standard for numbers of informants that qualify as adequate for studies other than the life history has dropped. When Howard Becker

did his doctoral research, using qualitative interviews with urban schoolteachers, his sample was 50. Many scholars in education used that sample size 20 or 30 years ago as a marker for what was possible. In many circles today, that number would be seen as high. As the approaches for doing qualitative research have expanded in interpretive methods, agreement about what a sample size should be depends on what the local context accepts as adequate. At the same time, ideas about how to evaluate adequacy are also changing. The power of the conservative federal government's politically conservative emphasis on scientific evidence that has marked research in education also affects how qualitative researchers design their studies, what kind of work gets funded, and what research is seen as significant. That is, while funding agencies are willing to fund more traditional qualitative studies, scholars on the margins try to push the limits of what counts as research so that research becomes more inclusive of varying cultural modes.

A range of methods is encompassed under the qualitative label. Traditionally, qualitative researchers in the field gather where research participants normally spend their time; they collect data that are descriptive in narrative form; their focus is on process rather than outcomes; their interest centers on the meaning-making process, that is, on how things, actions, interactions, and situations come to have the meanings that research participants assign to them; and the analysis of the data tends toward the inductive (Bogdan and Biklen, 2007). Newer forms of work emphasize or de-emphasize particular aspects of these characteristics. Institutional ethnography, for example, explores the social relations that structure people's everyday lives, emphasizing how the relations of ruling, that is, hierarchical power relations, can be traced through official texts and discourses into the lives of individuals (Smith, 2005). Auto-ethnography centers the narrator's story and may dispense with the idea of the field. Qualitative methods based on critical race theory assume a world where racism is understood as a central rather than peripheral feature of American life. Qualitative methods using critical race theory start from the assumption that racism exists. Research may look at different forms it takes, the varied vocabularies that represent it, or the kinds of communities that support it, but the researcher does not explore whether racism exists in American culture. Silverman (2004) argues that qualitative researchers need to "broaden our conception of qualitative research beyond issues of subjective 'meaning' and towards issues of language, representation and social organization" (p. 1). This view insists that it is not adequate to study the perspectives of research participants as the participant understands them, but to examine larger social discourses that shape how informants speak. Participatory action research dispenses to some degree with research subjects, informants, or participants (at least as they are traditionally understood), and works toward collaborative interventionist approaches for social change.

Qualitative researchers gather evidence in multiple ways. Strategies of interviewing and participant observation have predominated over the years, although how people have listened and to whom they have listened have changed. These traditional methods of gathering evidence, that is, have not remained static. Harding (1987), for example, has suggested that in interviews, feminist qualitative researchers have listened "critically to how traditional social scientists conceptualize women's and men's lives" (p. 2). Additionally, they have paid attention to behaviors that might earlier have been thought unworthy of observing. Even though qualitative methods have always been oriented to everyday life, what is included within that frame should not be taken for granted.

One sort of qualitative study that has witnessed growth is problematic in several ways. The sort of qualitative study that has expanded is the attraction to work that employs qualitative technique rather than the interpretive approach. This expansion may have occurred because of the desire to combine qualitative and quantitative approaches in the same study or because the federal emphasis on a more positivist approach has had these effects. The use of qualitative technique occurs when researchers collect narrative data to questions they have already predetermined. Their questions are already set before they begin their project. In this form, qualitative is simply a synonym for research where the evidence takes narrative form. The data which researchers using qualitative techniques amass are in response to their questions. The participants in this research approach serve as a kind of answer-providers, giving responses and information – even interpretations – but are shut out from helping to make certain that the focus and emphasis of the study connect to their daily lives. Here, the narrative form of the data is really the reason that it is called qualitative (i.e., not statistical or experimental).

That is, it does not take advantage of the strengths of interpretive approaches. Qualitative research is capable of representing social processes. One of its most significant strengths is that it can elicit what people take for granted and frame perspectives within this context. It studies well the logic of people's views. It beautifully addresses the complexity of lives, situations, and institutions, and therefore, unlike approaches that work to simplify environments, it is useful for actual policy development.

Interpretive qualitative evidence emerges from work that has postponed defining the questions' parameters. In this mode, researchers have a general question when they enter the field, but refine it through the process of being in the field or by talking with participants. This is why a broad question is the strongest one with which to enter the field. Evidence based on qualitative techniques results in more conventional and superficial findings because the researchers sidestep the process of negotiating with informants about what the study is. In the interpretive approach, researchers must learn from the people they

encounter in the field about a range of issues – what the organization, community, or setting is like; what is on people's minds; what kinds of vocabularies people employ to talk about concerns; what seems common sense to them; and what kinds of knowledge are valued. This sort of qualitative research does not privilege professional knowledge over the knowledge of ordinary people because it rejects the hierarchy of credibility. The researcher must take questions of power into account: "In any system of ranked groups, participants take it as a given that members of the highest group have the right to define the way things really are" (Becker, 1967). This means that qualitative researchers always study something from a particular vantage point and must communicate what that vantage point is. It is important for qualitative researchers not to reproduce what is taken for granted in these hierarchies, that those who are more educated or more trained are more insightful. At the same time, researchers should not romanticize those without power, assuming that their standpoint is both innocent and a window to truth. In disability-studies research using a qualitative approach, therefore, the perspectives of people with disabilities cannot be missing from the research. At the same time, their perspectives should not be conceptualized as either transparent or uncomplicated.

Previously, when the collection of evidence was accomplished primarily through participant observation, researchers negotiated with informants over the direction of the research because the length of time researchers spent in the field and the scope of the projects meant that researchers could not know the story to tell in advance. Today, with the press away from longitudinal studies on the one hand, and the insistence in many schools of education on the other that the literature has to drive the shape of questions that doctoral students pose for their dissertation research, students end up doing more limited, narrow, and in some ways weaker, less-robust studies. The increased conventionality means a weaker study because the opportunities are diminished for collecting evidence on the assumptions that informants have about their lives.

Evidence is gathered, mulled over, sorted, analyzed, and used to illustrate arguments. In these different aspects of the research process, thinking about evidence as strong or weak, thick, or thin, involves particular concerns. They involve different standards of judgment. When gathering evidence, for example, questions about the relationship of researchers to their informants must be addressed. Traditionally, researchers used the category of rapport to describe the formation and nurturing of this relationship. Did the researcher create a trusting relationship with the informant? When conducting interviews, we worry and write about whether the interviewer asked leading questions. Evidence is less trustworthy, that is, if interviewers use versions of "Don't you think that. . ." to start a question. It is also less trustworthy if they ask questions that

force interviewees to choose between two competing ways of approaching an issue rather than allowing the interviewee to decide how a response will be framed. Attention to the process of fieldwork is still significant even if evidence is understood in the interpretive frame.

Another question about evidence during fieldwork concerns how the researcher understands the norm, or the social construction of the norm. When, for example, fieldworkers talk about research participants on the one hand and black interviewees on the other, this vocabulary illustrates that the researchers assume that whiteness is a norm for a research participant and black people are constructed as other, or a deviation from this norm. This point is significant as a contradiction for the old argument that qualitative researchers once made, that just being in the field for a long period of time was a good criterion for establishing the worth of the data. Researchers' class, race, gender, sexual orientation, and age, to name a few significant identity markers follow them into the field and contribute to the construction of how they see and how they label what they see.

In the analysis stage, questions about the trustworthiness of evidence are visible in considerations of the context. Did researchers investigate how the particular context in which the researcher observed or interviewed and the specific time period in which the research was conducted come into play? If the data were analyzed as if they were context free, the study becomes less believable because the evidence is not grounded in time or place. Contexts can be defined geographically, socially, and linguistically. Any qualitative study that examines what participants take for granted, and particularly research that emphasizes discourse analysis negotiates multiple contexts, the local and the nationally regulating.

From the Traditional to the Postpositivist

Historically, qualitative researchers were not themselves satisfied with the idea that just being in the field was enough. Scholars discussed how it was that people came to say what they said. In their study of medical students, for example, Becker and his colleagues found that the students talked differently depending on the context and to whom they were speaking. In public groups, for example, the students tended to express their more cynical side, speaking little about the idealistic motivations they may have had or still continued to have about being doctors. In interviews or conversations with the researchers, though, they spoke more about it. Becker (1958) discussed how in fact people express different parts of themselves in different settings and that researchers can account for these contradictions rather than work to figure out which mode of speaking is true. This kind of discussion about the multiple allegiances of people being studied presages

more poststructural ideas about identity and subjectivity as both unstable and multiplicative.

Traditionally, qualitative researchers, to use the anthropologist James Spradley's metaphor, saw ethnographers as attempting to hold a mirror to the world. The idea was that ethnography should reflect the world that the ethnographer studied. This work was not simple as one captured informants' views through the lens of culture, and much cultural knowledge was either hidden from view or not even recognized as knowledge. Yet, tensions existed in this mirror view as qualitative anthropologists asserted that we represent one culture in terms of another, referring to how central the anthropologist's own culture was in the analysis. However, another anthropologist, James Clifford, firmly planted in the postpositivist world, argued that the metaphor of the mirror was false because one could only learn partial truths (Clifford, 1986). By partial truth, Clifford meant that there is no location that the researcher can occupy that affords a vantage point from which 'the truth,' and by this he meant a single truth, could be determined. Additionally, because the product of research is a written text, results become representations. Research efforts, then, lead only to partial truths.

In addition to the mirror metaphor, researchers relied on several ways of framing what was studied that emphasized the social construction of reality. The social construction of reality refers to how knowledge comes to be socially established as reality (Berger and Luckmann, 1967, 3). What people assume and take for granted is always central to the qualitative project. In articulating this view, Berger and Luckmann argued that ordinary people live in worlds that they understand as coherent. Further, the reality of everyday life appears already objectified (p. 21), that is, sense-making schemes seem embedded or integral. Engaging with the world includes figuring out what the coherence depends on, who benefits from the arrangements, and whose views are silenced. Of course, since humans participate in multiple worlds, this is not a single place. The definition of the situation is one aspect of the social construction of reality that is central to the ideas of symbolic interactionism, one of the threads of the history of qualitative methods. The definition of the situation refers to the idea that people act in situations they find themselves in on the basis of how they interpret the situation because there are no inherent meanings in the situations themselves. These interpretations come from people's interactions with others and with institutions and written texts, and they represent how they have learned what the situation means in the particular setting.

In the postpositivist era, the researchers' views and values, and their social location must also be examined. Of particular concern is the issue of power relations. Harding (1987) argued that the "inquirer . . . be placed in the same critical plane as the overt subject matter, thereby recovering the entire research process for scrutiny in the

results of research" (p. 9). This means that the researcher's identity markers of race, class, gender, disability, and age need to be considered as partly defining the researcher's investment and influence on the research. Evidence, in other words, is political, epistemological, socially constructed, and invested in certain frames.

When Donna Haraway addressed the idea of the inevitability of situated knowledge in 1991, she wrote that engaging the term objectivity was inescapable. While she was speaking specifically about tasks for feminists engaged in research, the term is something that qualitative researchers must address when they explain qualitative methods to quantitative-research audiences because of the different assumptions the groups make about research. Discussions of objectivity are also inescapable at this current moment because of the government's insistence on funding only research that employs randomized experimental designs (see Cho and Trent, 2006). These designs rest on a particular view of objectivity, one that emphasizes the positivist scientific tradition. The traditional definition, Haraway explained, seems to suggest that those who conduct objective research are not embodied, while those whose research is often accused of bias are not allowed not to have a body, and so are inevitably disqualified because they are seen to have a finite point of view (183). Objectivity, she argued, "turns out to be about particular and specific embodiment" (190). This alternative view of objectivity, she wrote, is the only kind of responsible knowledge claim because views that assert that they see the world from a neutral position are in fact irresponsible. They pretend that an unmediated account of the world is possible.

The significance of the question that Haraway proposed has been addressed by others. Harding (1987) argued that researchers need to represent themselves in their texts as actual "historical individuals[s] with concrete, specific desires and interests" (p. 9). These actions – engaging in greater reflexivity and representing these reflections in the text – open up the text for greater scrutiny by the audience and hence increase the narrator's trustworthiness and lessen possible distortion. Harding, who distinguishes between objectivity and objectivism, a "stance that attempts to make the researcher's cultural beliefs and practices invisible while simultaneously skewering the research objects' beliefs and practices to the display board" (p. 9), insists that the researcher's views and beliefs become part of the evidence that readers can evaluate. Since the judgments of different readers are central to evaluations of the quality of qualitative research, this movement has significant weight. Introducing subjectivity into the analysis increases the objectivity of the research (Harding, 1987). Therefore, evaluating the quality of evidence is partially related to the visibility of the social location of the researcher, where the social location refers to the researcher's race, class, gender, and other identity markers and how these markers are

invested in the design, understanding, and production of a qualitative project. These are by no means the only significant aspects of a project, but they count. Historically, qualitative researchers in the Chicago tradition suggested in their methodological discussions that simply having respect for the world and studying it closely to try and figure out how people in their particular contexts see it and act can enable any researcher to understand any research participant. Researchers never discussed in their methodological commentaries the possibility that their access to understanding could be limited by their social locations.

For qualitative researchers who are somewhat alienated from the idea of objectivity, framing this problem that Haraway and Harding discuss relates to a researcher's reflexivity and its representation in published accounts. Posed this way, we could say that reflexivity increases not the objectivity of the research but its trustworthiness. Since the worth of a qualitative study depends so heavily on the audience's response to it, trustworthiness takes on heightened importance. If we pay attention to the subject position of the researcher through reflexivity, this attention works to make it visible and addresses the subjectivity question rather than leave it unaddressed. In *Trouble on Memory Lane*, for example, this researcher addressed what happens when narrators of qualitative studies are not reflexive about their age in relation to their research participants when they study adolescents (Biklen, 2004). Many researchers in studies of adolescents draw unproblematically upon their own experiences as youth to illustrate to readers that they understand youth because they were once youth themselves, ignoring the different time periods where they lived through adolescence, and the social changes that have created a world that contrasts significantly from their own time as teenagers. They also describe their own experiences of youth, often in an unreflective way, to bolster their narrative authority.

Readers seeking to evaluate the quality of evidence in a study must be able to assess how authors construct their narrative authority in written texts. Narrative authority means that narrators of written texts have authority with readers because of how they represent themselves in the text. There are a variety of strategies that narrators use to gain this authority, including sharing vulnerability, articulating theoretical sophistication, using various rhetorical strategies, sharing details about significant time spent in the field; so having narrative authority does not rely on being forceful or authoritarian. One can gain narrative authority as Lareau (1989) does by sharing all of the mistakes she made in fieldwork. As readers we trust her because we see her as willing to lay her cards on the table. She is not protecting an image of herself as the sophisticated and naturally capable researcher. Another important strategy for establishing narrative authority with some audiences is to discuss how one's identities (gendered, raced, and classed) shaped the interactions with informants, the problems that

arose in the field, and the shape of the project. This kind of discussion builds on the notion of the embodied researcher that Harding portrays and emphasizes the notion of partial truths (see, e.g., De Andrade, 2000; Pascoe, 2007).

Before the postpositivist moment, qualitative methodologists assumed to some extent that doing empirical work meant that they could, to use James Spradley's metaphor, portray the world they studied as if it were almost a mirror reflection. The social location, the investments – the positionality of the researcher – did not engender much interest partly because of the points that Haraway (1991) so forcefully made, that researchers did not particularly think of themselves as embodied. Becker (1970c), for example, described participant observation in the following way:

The participant observer gathers data by participating in the daily life of the group or organization he studies. He watches the people he is studying to see what situations they ordinarily meet and how they behave in them. He enters into conversation with some or all of the participants in those situations and discovers their interpretations of the events he has observed. (p. 25)

Becker suggests here that doing participant observation is a fairly straightforward task.

Several aspects of these sentences stand out. First, they seem to assume that the researcher is male. Becker had long experience working with Blanche Geer, his co-author on studies (e.g., Becker *et al.*, 1961; Becker *et al.*, 1968), and respect for Helen Hughes and other academic women who were sociologists; so speaking in this way did not come from his research experience. But it may have come from the standards of writing at the time which subsumed women under the category of men. The assumptions Becker made to produce this sentence contradicted some of his experiences and reflected others.

Second, Becker describes the nature of this work in extremely simple language, seeming to challenge the idea that doing research requires jargon, specialized ways of knowing, and a professional demeanor. This style of presentation attracted many students to this method because it seemed to emphasize the significance of the world rather than a science to study it. The method relies on ordinary skills of looking, listening, and talking, as this way of writing suggests. A significant part of doing research, it suggests, involves hanging out.

Third, the observer conducts research, and in the process 'discovers' the perspectives of informants, perspectives that are already there. The research effort and experience are not seen as a process of production – the researcher and the informant do not together produce accounts – but, rather the researcher's time in the field and the careful recording of what is said and done enable the researcher to learn what informants think and why they develop the accounts of their world that they do.

Fourth, the researcher is involved in an enterprise that is inductive. The researcher enters the field with some idea of what will be studied, but must spend time in the setting in order to understand enough to understand what must be understood. Becker's view downplays the ways of making sense of the world, the ideologies and values that make particular aspects of a setting visible and other aspects invisible. Elsewhere, he and many other researchers have argued that the amount of data that rigorous qualitative researchers collect means that the researcher's values and perspectives will not dominate the research because of the weight of qualitative evidence. These researchers spend significant time in the field and so collect enormous amounts of data.

Finally, the researcher who has observed what really happened in the field works to understand what the interpretations of the people who acted in their worlds are on the events that have been observed. Becker does not argue that the researcher stops with the perspectives of informants, but gaining these perspectives is an important task for the qualitative researcher. In a later publication, Becker did say that he almost always studied issues that were immediately relevant to his life (Becker, 1996) suggesting that he did not only report the research participants' perspectives because he already had an investment in the topic.

Becker's comments on participant observation came near the beginning of an essay on 'Problems of inference and proof in participant observation,' first published in 1958 (Becker, 1958). His comments reflect what ways of thinking about qualitative methods have changed significantly since he wrote those words on method so many years ago and what assumptions no longer carry credibility. Just as qualitative methodologists argue that the good researcher studies meaning making in context, qualitative researchers themselves write in context. At the time he wrote, Becker argued that one of the problems in presenting evidence in qualitative studies is that there are no agreed-upon guidelines for presenting evidence: how much, what makes it convincing, and what the rules are for its display. In the end, one of the most important questions to ask in relation to evidence is whether or not the interpretation is credible. The next section of this article examines some of the foundational issues about evidence in empirical qualitative studies. It addresses questions about the amount of data considered necessary, the importance of the relationship between researchers and their research participants, and the discourses that shape the research effort.

Evidence and the Field

Qualitative methods are, taken together, a form of empirical investigation. This means that they are related to data. They are different from other interpretive works, say in

the humanities, where the rules for inquiry have different foundations. Data and interpretation are interrelated and connected. What kinds of empirical issues are important in evaluating evidence?

How much data? As discussed above, how much is a time-sensitive question that has changed over the years. One of the significant characteristics of this approach that does affect evidence is that qualitative approaches get at complications rather than work to simplify them. Therefore, researchers have to have a way to spend enough time or talk to enough people in a setting not only to get a feel for the place and situation, but also to be able to say in some way, "this is how things work around here for the people I am studying." One can use member checks to make sure that the researcher understands what people think and feel, but it always takes time. So-called *blitzkrieg* ethnography (Rist, 1980) produces only superficial readings of situations, people's perspectives, and settings. Imagine moving to a country where the temperature is expressed in terms of Celsius rather than Fahrenheit. Figuring out the temperature when it is expressed in terms that the traveler does not have an intrinsic feel for demands some conscious work that will make the translation known. Brought up in a Fahrenheit society, the citizen knows what 72° feels like, but has to work to make the connections, in a Celsius society, to what 22° feels like. Qualitative researchers work to find out what it feels like to know what 72° or 22° feels like.

Whatever changes have shaped how we consider data and their analysis, spending time in a setting or with participants in interview research is a minimum requirement for strong evidence. It does not guarantee a thoughtful, elegant piece of research, but without time in the field and a significant quantity of field notes, the evidence cannot be strong. Detail does not in itself yield thick description (Geertz, 1973). Field notes are not rich just because their data are detailed. Thick description refers to a particular understanding of social life that the ethnographer works to capture: "social life is densely coded and performed through multiple frames of reference. There are multiple forms through which the social is enacted and accounted for. It is rendered through actions, narratives, texts, visual representations, and material artifacts. These do not cumulate to convey one simple set of cultural 'messages,' but they constitute multiple semantic domains, multiple arenas of action, and multiple sources of significance" (Atkinson *et al.*, 2003: 114). Thick description is the result of engaging people's complicated and layered lives in order to communicate something significant about them without losing the complexity.

Evidence and Social Location

How does the relationship between the researcher and the informant matter? The kind of evidence that results from

qualitative in-depth interviews depends on the sorts of questions that get asked. Asking leading questions pushes informants to respond to the interviewer's ideas and assumptions about what is significant and how the world is ordered. Asking questions in a way that presses the informant to choose the frame from which to respond offers greater access to the informant's sense-making frames. Just asking open-ended questions, in itself, does not secure this access as sometimes a reliance on particular vocabularies and assumptions within, say, a 'how' question can lead informants out of a frame that might have been attractive to them. There is also the issue that many informants want to please the researcher and so they work to tell a story that the researcher wants to hear. That is why the one-time or one-shot interview is a weaker form of data collection that results in evidence that is less trustworthy.

Interviewers not only ask questions, but they also listen. How researchers hear or understand what is being talked about, even what is enacted in the interview process, is shaped by the frameworks within which they listen. One way that scholars who write about methodology refer to these frameworks is with the vocabulary of theory. When interviewers listen, that is, they listen to what informants say through a lens or filter that orders the interviewee's words; for example, positioning some words as important and some words as unimportant, according to ideas about how the world works and what matters. Theory – that is theories about human life – is this lens. How does the researcher understand power relations? How does the researcher understand the relationship between an individual's biography and the larger events of history? How does a researcher see or take into account the identity of the person being interviewed. This is a question not only of how the researcher listens to the informant but of how the researcher understands the identity of the research participant being interviewed. Theory refers to the most basic aspects of identity. Does the researcher see the different markers of an informant's identity as separate or as inter-connecting? How researchers theorize this question shapes what they ask, what they hear, what they even consider to be data. In a study of gender and elementary school teachers, for example, I understood the racial and gendered identities of the research participants to be separate rather than intersecting. I assumed that I could study women teachers, whether black, white, or Latina teachers as women. This affected how I listened to the informants.

One example about questioning and listening is instructive. An instructional specialist, one of very few Black teachers I interviewed multiple times, had children who were about my children's age. Like my children, they were in day care. Whenever I entered the building to do participant observation and encountered her, or had an interview arranged with her, she always spent significant time asking about my children, how they were doing, and talking about

her children and how they were doing. I read this conversation as chatting, and categorized this kind of talk as rapport building. From my perspective, I wanted to get down to work. I experienced talking about our children as an effort to place me in a domestic realm. I felt resistant to this pressure because I saw myself as working on a funded research project where I needed to collect meaningful data. I already had a good relationship with her, I thought, so why did we need to continually develop rapport?

What I did know about my informant was that she was one of the few teachers who spoke clearly about her ambitions to be a principal. She assumed that adult women had continuous employment, and she connected her ambitions with strategies about how to advance in the workforce. These were the issues that I saw as central to my study, and that I wanted to explore. At the same time, she was one of the only teachers who had children in day care. Had I interpreted her talk about children and childcare as connected to her work, as in fact it was, I would have taken these interchanges as serious engagement about the topic of my research rather than as rapport building or chatting. I could also have asked her in greater detail about her mother traveling from the south to help with childcare so that she would not have to miss too much work when her children were born. How much time was enough? Did her mother help her with both children? How did she understand her mother's views about her daughter's ambitions for the principalship? All of these significant aspects of her perspective were missing because I listened to her talk through theory that separated work and family. I might have been one of the only people she could have spoken with about her children as she and I both held shared beliefs in the value of public child care (see Biklen (1995); and Biklen *et al.* (2008) for a longer discussion of these concerns).

The Question of Distance

In the modernist frame, representing reality truthfully meant that one had to have distance from the object of study (Potter, 1996). Closeness to the informants could mean that the researcher was too invested to get it right. In qualitative methods, however, closeness is a significant requirement for knowing. Lofland (1971) evaluated this by four criteria: physical proximity to the people studied, staying connected to informants or to the community studied over time, closeness in terms of 'social intimacy' and a sense of 'confidentiality' in the relationship, and finally, closeness in terms of recording the 'minutiae' and details of everyday life (p. 3) so that they appear in field notes. This last point about recording the details in field notes suggests that the articles written from the study directly quote from what people in the field say. If the texts resulting from qualitative fieldwork do not

contain these direct quotations, their believability is lessened. This perspective emphasizes both the significance of the reader's or audience's role in relation to deciding the credibility of an interpretation and the strategies that researchers use to gain the reader's trust in their insights. Qualitative studies depend on field notes the researcher writes up from fieldwork and from samples of interview transcripts to illustrate and reconfirm analysis. This evidence can be woven throughout the text as in Diamond's (1992) study of nursing homes, or separated out in individual chunks to show how the author's view rests on the data as Lacy (2007) does in her study of Black middle-class communities.

Conclusion

The vocabulary frequently used to refer to the credibility of interpretations is validity, which researchers understand as the trustworthiness of inferences drawn from data (Eisenhart and Howe (1992), 644 cited in Freeman *et al.* (2007)). Validity as most scholars use this shorthand refers to the part of the equation that insists that there is a 'there there,' and frames how the research represents it. It is not as useful in adequately addressing assertions about theoretically framing a project or in addressing what counts as evidence and why. But it lurks adjacent to all published accounts, particularly during times when the federal government has moved discussions about research value toward a particularly constructed scientific model, or model of scientism. Readers bring questions about the trustworthiness of an account to each text, so researchers need to communicate not only what counts as evidence in the project and why, but how they represent themselves and the evidence in the text as they do. Evidence is not only related to research design, fieldwork, and data analysis but to writing, rhetorical strategies, and textual representation as well.

The quality of evidence in a published text reporting qualitative research is important for multiple reasons. One reason is that in order for an account to be understood as the author of the account intends, the writer must be able to convince the reader that the account is believable and that the narrator's authority is trustworthy. The quality of the evidence is constructed by the author and the reader together as the reader makes judgments about its quality. When readers consider the quality of the evidence, they continually consider its believability. If audience members do not share understandings with the research participants, they need to have enough examples of what informants say to believe the narrator. They also have to believe the narrator's interpretations of what those who were studied say.

This article opened with the complications of evaluating evidence after the interpretive turn. This does not

suggest that because qualitative researchers tend to operate on the basis of the social construction of reality, the implications of social constructions are not real. As Potter (1996) suggests, there is a "there there." In addition to basic demands that researchers spend time in the field, know their informants well, understand that people's lives are complicated and layered, and that the goal of these methods is to capture those complications, there are additional concerns about how the worlds that researchers and informants live in are structured. Questions about evidence are embedded in issues of power relations, identity markers, investments, and enormous inequalities. In order to make judgments about whether or not a qualitative study is believable, readers must take all of these factors into account.

See also: Critical Race Theory; Ethnography; Life History; Narrative Inquiry; Symbolic Interactionism, Naturalistic Inquiry, and Education.

Bibliography

- Atkinson, P., Coffey, A., and Delamont, S. (2003). *Key Themes in Qualitative Research: Continuity and Change*. Walnut Creek, CA: Altamira Press.
- Becker, H. (1996). The epistemology of qualitative research. In Jessor, R., Colby, A., and Schweder, R. (eds.) *Ethnography and Human Development: Context and Meaning in Social Inquiry*, pp 53–72. Chicago, IL: University of Chicago Press.
- Becker, H. (1970a). Field work evidence. In *Sociological Work*, pp 39–62. New Brunswick, NJ: Transaction Books.
- Becker, H. (1958). Problems of inference and proof in participant observation. *American Sociological Review* 23 652–660. (Reprinted in Becker, H. (1970c). *Sociological Work*, pp 25–38. New Brunswick, NJ: Transaction Books.
- Becker, H. (1970b). Life history and the scientific mosaic. *Sociological Work*, pp 63–74. New Brunswick, NJ: Transaction Books.
- Becker, H. (1970c). *Sociological Work*. New Brunswick, NJ: Transaction Books.
- Becker, H. (1967). Whose side are we on? *Social Problems* 14, 239–247.
- Becker, H., Geer, B., and Hughes, E. C. (1961). *Boys in White*. Chicago: University of Chicago Press.
- Becker, H., Geer, B., and Hughes, E. C. (1968). *Making the Grade: The Academic Side of College Life*. New York: John Wiley.
- Berger, P. and Luckmann, T. (1967). *The Social Construction of Reality*. Garden City, NY: Anchor Books.
- Biklen, S. (1995). *Schoolwork: Gender and the Cultural Construction of Teaching*. NY: Teachers College Press.
- Biklen, S. (2004). Trouble on memory lane: Adults and self-retrospection in researching youth. *Qualitative Inquiry* 10(5), 715–730.
- Bogdan, R. and Biklen, S. (2007). *Qualitative Research for Education*, 5th edn. Boston, MA: Allyn & Bacon.
- Cho, J. and Trent, A. (2006). Validity in qualitative research revisited. *Qualitative Research* 6(3), 319–340.
- Clifford, J. (1986). Introduction: Partial truths. In Clifford, J. and Marcus, G. (eds.) *Writing Culture: The Poetics and Politics of Ethnography*, pp 1–26. Berkeley, CA: University of California Press.
- De Andrade, L. (2000). Negotiating from the inside: Constructing racial and ethnic identity in qualitative research. *Journal of Contemporary Ethnography* 29(3), 268–290.
- Dezin, N. (2009). The elephant in the livingroom: Or extending the conversation about the politics of evidence. *Qualitative Research* 9, 139–160.

- Diamond, T. (1992). *Making Gray Gold: Narratives of Nursing Home Care*. Chicago, IL: University of Chicago Press.
- Eisenhart, M. A. and Howe, K. R. (1992). Validity in educational research. In LeCompte, M. D., Millroy, W. L., and Preissle, J. (eds.) *The Handbook of Qualitative Research in Education*, pp 644–680. San Diego, CA: Academic Press.
- Freeman, M., deMarrais, K., Preissle, J., Roulston, K., and St. Pierre, E. (2007). Standards of evidence in qualitative research: An incitement to discourse. *Educational Researcher* 36(1), 25–32.
- Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In *The Interpretation of Cultures*, pp 3–30. NY: Basic Books.
- Harding, S. (1987). Is there a feminist method? In Harding, S. (ed.) *Feminism and Methodology*, pp 1–14. Bloomington: Indiana University Press.
- Haraway, D. (1991). Situated knowledges: The science question in feminism and the privilege of partial perspective. In *Simians, Cyborgs and Nature*, pp 183–201. NY: Routledge.
- Lacy, K. (2007). *Blue-Chip Black: Race, Class, and Status in the New Black Middle Class*. Berkeley, CA: University of California Press.
- Lareau, A. (1989). *Home Advantage*. London: Falmer Press.
- Lincoln, Y. (2002). On the nature of qualitative evidence. *Paper Presented at the Annual Meeting of the Association for Study of Higher Education*, Sacramento, California, November 2002.
- Lofland, J. (1971). *Analyzing Social Settings*. Belmont, CA: Wadsworth.
- Mills, C. W. (1959). *The Sociological Imagination*. London: Oxford University Press.
- Mischler, E. G. (1979). Meaning in context: Is there any other kind? *Harvard Educational Review* 49, 1–19.
- Pascoe, C. J. (2007). *Dude, You're a Fag: Masculinity and Sexuality in High School*. Berkeley, CA: University of California Press.
- Potter, J. (1996). *Representing Reality: Discourse, Rhetoric and Social Construction*. London: Sage.
- Rist, R. (1980). Blitzkrieg ethnography. *Educational Researcher* 9(2), 8–10.
- Silverman, D. (2004). Introducing qualitative research. In Silverman, D. (ed.) *Qualitative Research: Theory, Method and Practice*, 2nd edn., pp 1–8. London: Sage.
- Smith, D. (2005). *Institutional Ethnography*. Walnut Creek, CA: Alta Mira Press.
- Waller, W. (1932). *Sociology of Teaching*. New York: Wiley.

Further Reading

- Bourdieu, P. (1990). *In Other Words: Essays Towards a Reflexive Sociology* (Matthew Adamson, Trans.). Stanford, CA: Stanford University Press.
- DeVault, M. (1990). Talking and listening from women's standpoint: Feminist strategies for interviewing and analysis. *Social Problems* 37(1), 96–116.
- DeVault, M. (1991). *Feeding the Family: The Social Organization of Caring as Gendered Work*. Chicago, IL: University of Chicago Press.
- Miller, J. and Glassner, B. (2004). The inside and the outside: Finding realities in interviews. In Silverman, D. (ed.) *Qualitative Research: Theory, Method and Practice*, 2nd edn., pp 125–139. London: Sage.

Visual Data in Education Research

J Wagner, University of California, Davis, Davis, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Brown versus Board of Education, or Brown decision

– The decision by the U S Supreme Court in Brown versus Board of Education to overturn the legal precedent of separate but equal in favor of abolishing racially segregated schooling.

Conscientization – A term proposed by Paolo Freire to describe the empowerment of indigent populations through political education and consciousness development.

Emic (and etic) perspective – Accounts of natural phenomena in terms that are meaningful to insiders, participants, or subjects of a study – contrasts with etic.

Etic (and emic) perspective – Accounts of natural phenomena in terms that are meaningful to outsiders, observers, or researchers – contrasts with emic.

Image-based research – Research that considers visual images and imagery as a key dimension of culture and social life.

Microethnography – The application of ethnographic perspectives and methods to investigate social interaction in relatively small-scale settings and exchanges – that is, conversation, performance, teacher–student questioning, game playing, etc.

Photo elicitation – Using photographs as interview prompts.

Projective interview – An interview designed to reveal something about the person being interviewed (in contrast to interviews that focus on the interviewee as a source of information about external matters).

Visual anthropology – A perspective within anthropology that emphasizes the analysis of visual dimensions of culture and social life and/or uses visual imagery to develop or explicate anthropological knowledge.

Visual data – Data that take the form of visual images, including visible artifacts, photographs, and videotape recordings, maps, drawings, etc.

Visual sociology – A perspective within sociology that emphasizes the analysis of visual dimensions of culture and social life and/or uses visual imagery to develop or explicate sociological knowledge.

Education researchers have used visual data to investigate diverse phenomena and to communicate about their research to colleagues, policymakers, practitioners, and the public. The forms of visual data most common to these purposes include videotape recordings, photographs, drawings, material artifact collections, and graphic figures, logs, and tracings.

In application to education research, visual data of this sort support two different modes of research. In the first mode, researchers examine phenomena independent of the human agency and experience that might bring them into being. In the second, visual data are used to explore the knowledge, values, and perspectives of individuals and groups.

Research cast in the first mode involves investigators as external observers who record and analyze what they see in terms that make sense to other researchers. Photographs, videotapes, drawings, and visual artifacts can provide useful data for studies of this sort and contribute to what anthropologists refer to as the etic perspective, or outsider's point of view. In the second mode, research questions focus on how phenomena are experienced, shaped, and perceived by individual and groups. Visual materials can also provide valuable data for studies aimed in this direction and that reflect what anthropologists call the emic perspective, or insider's point of view.

Of special interest to educators, visual data can play a role in research-based interventions – either locally or farther afield. A videotape recording of classroom interaction, for example, could be analyzed solely for its methodological significance to other researchers who make videotape recordings, but it could also be used to stimulate reflection among the students and teachers recognizable in the recording – or among others who identify with them – or to document the work of an individual teacher for a certification review. Visual studies of textbooks or classrooms could be framed by historical questions that have few contemporary implications, but they could also play a role in multi-million dollar law suits and challenges to state or district school policies.

These general observations about visual data apply in part to numerical and text data as well. All three forms of evidence can support research directed toward etic or emic accounts, and all three can inform interventions and social action. For many social scientists, however, visual imagery is linked more closely with an emic or semiotic perspective, numbers with a natural science or etic approach, and texts

more or less with both. In anthropology and sociology, these associations have led researchers working with visual data to establish separate professional organizations, journals and conferences. Visual studies in education have been shaped somewhat by these developments in allied fields, but they also reflect the distinctive ethos of a profession that values both interdisciplinary and applied inquiries.

The Visual Heritage of Education Research

The work of Lewis Hine during the early part of the twentieth century provides a useful starting point for considering these matters further. Hine is best known today as a documentary photographer of child laborers, immigrants, and the Empire State Building, but his camera work began as an assistant teacher of nature study and geography at the Ethical-Cultural School in New York City. Hine came to this position in 1904 after completing undergraduate work in sociology and anthropology at the University of Chicago and while working toward a master's degree in education at New York University. Shortly after he arrived at the school, he was asked by Principal Frank Manny to serve as the official school photographer.

The Ethical-Cultural School supported, and was supported by, a remarkable group of educators, social researchers, and social reformers, including John Dewey (two of his children attended the school) and the anthropologist Ruth Benedict. Consistent with the progressive ethos of the school, Hine expanded his school-photographer role to make images of culture and social life outside the school that he hoped could enrich instruction within it. This ideal led him to begin photographing – and encouraging his students to photograph with him – immigrant families as they arrived at Ellis Island. The visual data he used to teach students about immigration and cultural diversity stimulated Hine's interest in using photography to teach the public about similar issues. With that in mind, in 1909 he left teaching for photography and generated an oeuvre of visual data that served both researchers and reformers.

Somewhat similar connections between fieldwork, visual documentation, pedagogy, and social reform characterized Charles Hamilton Houston's work for the *National Association for the Advancement of Colored People* (NAACP). The Association was founded in 1909 with strong support from the Ethical-Cultural Society, the parent organization of the school where Hine had been teaching. Some 25 years later, after a term as dean of the Howard University Law School, Houston was appointed special counsel to the association. His signal contribution was to craft a legal strategy for undermining the separate but equal principles used to legalize racial segregation since the US Supreme Court's landmark 1896 decision in *Plessy versus Ferguson*. Houston's proposal to the NAACP

membership was not to challenge this principle directly, but to illustrate the social and economic costs of holding institutions to the standard of truly equal facilities for both black and white citizens. This strategy required legal challenges to be pursued in the courts, but it also required good evidence about what was equal and what was not.

Houston's first step toward mobilizing NAACP support for this strategy was to lead a film crew in documenting separate and unequal schooling in South Carolina. Houston screened the film as part of his formal address to the association. When a majority of the members voted in favor, the NAACP embarked on a decade-long series of legal challenges to racial segregation. These challenges culminated (shortly after Houston's death in 1950) in the Supreme Court's decision, in *Brown versus Board of Education*, to overturn *Plessy* and call for the abolition of racially segregated public schools.

Field Studies, Experiments and Interviews

With varying emphases on documentation, inquiry, instruction, and advocacy, visual field studies evolved during the mid-century decades as a valued approach in research on education and schooling. Some studies of this sort were closely allied with related work in anthropology, sociology, and psychology. In the 1930s and 1940s, for example, the anthropologists Gregory Bateson and Margaret Mead reported on their extensive photographic studies of Balinese culture, a substantial strand of which focused on adult-child interaction in support of learning and socialization. As a sequel to her work with Bateson, Mead collaborated with Frances MacGregor in analyzing the same photo archives to explore cross-cultural dimensions of child-development theories articulated by psychologist Arnold Gesell.

Image-based education research also drew inspiration from psychological investigations that used visual materials – photographs, drawings, abstract shapes, or artifacts – as prompts for clinical and research interviews. The Rorschach and the Thematic Apperception Test (TAT) were well-known variants of this projective interview genre. In application to education and schooling, however, a more notable example was a series of studies during the late 1930s and early 1940s by Kenneth and Mamie Phipps Clark. The Clarks used a black doll and white doll as visual prompts for interviewing black and white children. With the dolls displayed before them, children were asked which of the two they preferred playing with, which was the good doll, and which doll was most like them. The Clarks interpreted the preference for white dolls among both black and white children, and the visible distress exhibited by the black children in response to the forced-choice questions, as projective evidence of the psychological costs to children of racial segregation and inequality. They testified to this effect in several school hearings – including the Briggs

versus Elliott case that was incorporated into Brown versus Board of Education – and the US Supreme Court made explicit note of their research in its rationale for the Brown decision.

The Clarks' doll studies shifted the focus of projective interviewing from individual personality toward culture and social relations. John Collier, Jr. pushed it even farther in that direction through what he called the photo-elicitation interview, an approach in which photographs of local environs, activities, and materials were used as prompts for discussion with research subjects. Collier noted, in particular, the potential of this approach to elicit accounts from research subjects that reflected their own language, concepts, and priorities rather than mirroring those of the researcher.

Collier had worked as a documentary photographer for the Farm Security Administration in the 1940s and extended the lessons of that experience into his anthropological fieldwork. Through a series of studies in the 1950s and 1960s, he used photo-elicitation interviews to explore perceptions of those he studied and systematic photographic inventories to document their material possessions, homes, and communities. He was keenly interested in studies of both school organization and student experience and affirmed the etic and emic as complementary perspectives. His own interest in education research was apparent in his seminal work, *Visual Anthropology: Photography as a Research Method*, and from his film recordings of native Alaskan classrooms.

Another notable application of image-elicitation techniques to education appeared in the literacy and political education projects Paulo Freire and his associates pursued with indigent populations in Brazil and other Latin American countries. As the first step of a widely imitated intervention protocol, they used traditional fieldwork techniques to identify key circumstances of local culture and social life. As a second step, they shared photographs or drawings of these circumstances with local residents to stimulate community discussions as a foundation for politicization and literacy development. Within the process that Freire called conscientization, social-documentary images functioned as instructional materials. As an intimation of future application of visual data to both research and practice, the subjects of Freire's interventions were also collaborators in analyzing the images themselves.

Microethnographies of Classroom and Language Use

By the mid- to late 1970s, the heritage of these varied approaches had demonstrated the potential value of visual data to studies of the behavior, psychology, social life, culture, and politics of subject populations and to both educational and social reform. Interest in this dual potential

continued as portable videotape recorders and digital media ushered in new forms of visual data and new kinds of visual studies.

One of the most vital applications of portable videorecording emerged among educational anthropologists and sociologists such as Frederick Erickson, Jeffrey Shultz, and Hugh Mehan who were interested in fine-grained investigations of schooling processes and functions. Much of this work focused on different forms of school-related communication – presenting class lessons, teacher–student consultations, classroom interaction, and so on. These phenomena had received prior research attention, but videotape technologies allowed researchers to examine them more systematically and in greater detail. They also made it possible to review the same interaction sequences again and again, code for more subtle cues, observe the interplay of verbal and nonverbal messages, and extend data analysis protocols to colleagues who had no direct involvement in a given field setting.

Many of these studies (but by no means all) were attuned to the ideal of improving experiences and outcomes for students who were disadvantaged in school by their culture, language, race, or social class. Indeed, an enduring contribution of visual studies of this sort was to document how interaction patterns in and around classrooms both reflected and supported structural school inequalities.

Videotape equipment not only made recording easier and more portable, it also made possible new forms of editing, playback, and collaboration. In the mid- to late 1980s, for example, George and Louise Spindler presented videorecordings of classroom interaction as prompts for group discussions in their cross-cultural investigations of schooling in Wisconsin and Germany. As an extension of shared video viewing between researchers and subjects, other researchers engaged research subjects as image-makers in their own right. The work of John Adair, Sol Worth, and Richard Chalfen with native Navaho filmmakers provided a template for subsequent studies in other venues and with other populations, including youth media productions directed toward both education research and youth-development outcomes.

Digital Media and Visual Studies

In the decades just before and after the millennium, the growth of digital media has stimulated additional attention to visual data in education research and in the classroom. For more than a century – at least since Hine's work at the Ethical–Cultural School–teachers and students could pursue image-based field studies as part school curricula, but this practice increased dramatically as digital still and video cameras became increasingly affordable and available. Teachers can now assign students visual-documentation projects and review the results immediately in class. Students can also combine digital image making at

school with related projects in their homes and communities. This expanding genre of student work (including video-based oral history and digital storytelling) includes strands of both documentary and projective imaging and has attracted the attention of researchers interested in literacy, classroom instruction, and youth development.

The capacity of computers to generate visual logs of how individuals use them has also created new possibilities for examining how students learn, what they know, and how they think. In developing the Green Globbs mathematics education software program, for example, Sharon Dugdale designed separate interfaces for students and for teachers and researchers. Within the former, students create algebraic equations to trace a line on their computer screen through predetermined points, or globbs. By presenting different sets of points, the software challenges students to experiment with different coefficients and different kinds of equations. Through the instructor's interface, however, a teacher or researcher can also review hits and misses as diagnostic information about students' mathematical understanding.

Computer-generated visual representations of how students process information have played an increasingly important role in studies of cognition and language development. Neurophysiologists, for example, have used magnetic resonance imaging (MRI) scans to reveal patterns of brain activity associated with learning, memory, and problem solving. As a microbehavioral complement, researchers have used computerized sensing instruments to investigate how students decode and process visual information. In a relatively recent strand of reading research, for example, an eye-tracking device and computer screen operate in sync to generate a visual map of how students scan words on a computer screen.

The instrumentation associated with MRI's and eye tracking seems more closely aligned with experimental research than with fieldwork. However, instrumented interactions with computers are now a common feature of computer-assisted tutoring, social networking, and recreational gaming. In a format that combines elements of both experimental and field studies, some researchers have captured changes in how information is displayed on a computer screen (including changes made by students themselves) in connection with audio-video recordings of students interacting with the computer, either alone or with tutors or peers.

In addition to naturalistic and experimental research studies, digital media and the Internet have precipitated broad changes in how visual materials are distributed and accessed. Online archives from museums, newspapers, historical societies, and libraries have created opportunities for research on school populations, facilities, and rituals as they were photographed and otherwise documented at the time. Websites and networking technologies have also stimulated inquiries into visual dimensions of

adolescent social life, distance and hybrid learning, and student-identity formation.

The Internet has also made it possible to incorporate richer visual data in communication among education researchers and between researchers and practitioners. In form, if not in substance, Collier's photo-elicitation interviews and Freire's efforts to enrich subjects' understanding of key social and political processes can now be pursued online. An early step in this direction was taken when the Child Language Data Exchange System (CHILDES) project created an online archive of audio and videorecordings of children's language use that is now part of a larger TalkBank system. Another step occurred when analysis tools from the Third International Mathematics and Science Study (TIMSS) were reconfigured to support websites where videotape recordings of teaching and learning could be uploaded for review and comment by others engaged in similar pursuits. In a less structured and rapidly evolving way, YouTube has been used by an increasing number of educators and scholars to share visual documents of teaching and learning. As a related, but more structured and officially sanctioned enterprise, annotated video documents of classroom instruction have emerged as an integral feature of teacher credential and licensing programs, including certification by the National Board for Professional Teaching Standards (NBPTS).

Research Contributions and Study Formats

Broadly defined and illustrated by studies such as those noted above, visual data can contribute to education research in four key dimensions:

- extending and enriching observations of natural phenomena that are difficult to document, analyze, and represent with textual and numerical data alone;
- instantiating forms of thinking, feeling, belief, and understanding that are difficult to examine empirically in other ways;
- helping to challenge and refine theoretical constructs for which significant visual dimensions have been overlooked; and
- stimulating and enriching communication among researchers and between researchers, practitioners, and policymakers.

Anne Dyson's study of how children developed literacy skills necessary for successful school performance illustrates several such contributions. Prior studies emphasizing individual linguistic competence represented students as young cogitators trying to apprehend grammar rules and vocabulary presented by teachers. Consistent with this view, policy attention focused on building teacher skills in diagnosing reading and writing difficulties,

designing appropriate lessons, and helping students stay focused on what they could learn from their teachers.

Dyson's direct observations revealed that students' literacy skills might also develop through communication with their peers, that is, within the social worlds they created and found in school classrooms. To explore the depth of these apparent effects, Dyson had to examine not only what children read and wrote, but also what they said to each other and the visual materials they created and responded to in and out of school, including their own drawings a wide range of mass-media imagery. By directly observing how children used these materials and images, she noticed a neglected relationship between mass culture, children's drawings, classroom social life and literacy development – a relationship that the children did not always notice themselves.

The data Dyson collected about the visual materials children attended to in their literacy development made possible a kind of natural scientific inquiry that would be difficult to pursue in other ways, but visual data can also be applied to semiotic analyses and the construction of emic accounts. For a study of cross-cultural perspectives on preschooling, for example, Joseph Tobin, David Wu, and Dana Davidson needed vivid portraits of what some preschools actually looked like. Toward that end, they made videotape recordings at several preschools where they had been conducting fieldwork in Japan, China, and the US. They focused these recordings on activities that were routine for each school, but they also made a special effort to include interactions that reflected conflict, ambiguity, and emotion. They then edited the raw tapes into three 20-min videos portraits, each of which revealed some dimensions of the culture and social organization of one school in each country.

For the second phase of their study, the researchers scheduled public viewings in each country of its corresponding videotape preschool portrait and recorded audience responses, questions, and discussion. These quasi-focus group events helped the researchers to assess the typicality of their video portraits. More importantly, they revealed ideals that adults from each country had about their own preschools. These ideals were refined further through a second round of inquiry in which audiences viewed and commented on preschool video portraits of the other two countries. Taken together, within-country and cross-country video-elicitation activities helped researchers understand how people in each of the three countries regarded the lives of young children and revealed some of the contrasting ideals that supported different preschool configurations among them.

Visual evidence can also provide grounds for challenging the taken-for-granted correspondence between some kinds of data and the phenomena to which they refer, at times in ways that redefine phenomena themselves. Contributions of this sort are illustrated by the three-decade, history of the

International Mathematics and Science Studies. In the first of these studies, national schooling phenomena were defined by standardized test scores, according to which students from some countries did much better than students from others. Critics noted quite rightly, however, that without knowing more about what concepts had been taught to students, and when, comparative assessments of knowledge acquisition and teaching effectiveness were indeterminate. To correct for this shortcoming, a second study in this series combined test-score data with information about the scope and sequence of mathematics curricula in each participating country. The official documents collected as evidence of these variations broadened the conception of schooling being compared but were also somewhat problematic in focusing on espoused curricula that might or might not correspond to curricula in use.

In hopes that it could deepen these international comparisons, a classroom videotape component was included as an adjunct to the Third International Mathematics and Science Study or TIMSS (since extended under the same acronym as Trends in International Mathematics and Science Study). Designed by James Stigler and his associates to sample some of the same classrooms participating in the larger study, the videotape component revealed cross-national differences in how mathematics and science lessons were presented to students, the allocation of class time among different tasks and activities, and patterns of student and teacher interaction. These contrasting features were significant in comparing schooling across different countries, but they were difficult to represent through numerical and textual data alone. The TIMSS videotape study stimulated a cascade of related research and policy investigations focusing on classroom instruction, mathematics and science education, the preparation of teachers, and cultural foundations of schooling. It also challenged the presumption of using either test scores or official curricula as proxies for effective teaching and learning. This outcome prompted the National Research Council's Board on International Comparative Studies to recommend the expanded use of videotape in cross-national studies of schooling.

For each of these three studies, visual images were not only examined as empirical data about the particulars of education and schooling. They also were used to report research findings to external audiences (and in the preschool study, as prompts for eliciting audience commentary). Dyson's writing and presentations typically included photocopies of student work as both evidence and illustrations of her research findings. Tobin, Wu, and Davidson made videotapes not only for interviewing adult subjects but also for stimulating public discussion and policymaking. Visual data generated through the TIMSS research were also integral to both research and public reports about the project. Sample video clips of classroom instruction from countries participating in the initial video study,

for example, were distributed as supplements to written reports. Technologies for analyzing these data were also incorporated into what has become a proprietary lesson lab project for supporting professional development among teachers and other educators.

Symmetry between the analysis of visual data for research purposes and inclusion of visual data in research reports does not necessarily apply to education research in general. Many researchers have analyzed visual data without including them in published reports. In conducting his elementary school ethnography, for example, Jan Nesper reviewed photographs of his own and those made by students and also student-drawn community maps, but only a few of the maps are reproduced in the book based on this work, *Tangled up in School*. The inverse also applies, as many researchers who do not analyze visual data use visual imagery in presenting their work to others.

Diverse Phenomena and Varied Purposes

Education studies are noteworthy for using visual data to investigate a wide range of phenomena and to serve an equally wide range of research and development purposes. Videorecordings of classroom interaction, for example, have been used to study student–teacher interaction, the presentation of course content, and – as interview prompts – to investigate adult views of appropriate school behavior. The same kinds of data have also been used to support professional teaching communities or conduct-certification reviews. Visual artifacts of student schoolwork have been examined to answer questions about student abilities and effort, the cultural knowledge of different school populations, teaching effectiveness, and school–home–community interactions, but they are also used routinely to support parent–teacher conferencing about individual students.

These diverse applications suggest that visual data, analysis tools, and evidence are relatively normalized within the conduct of education research. Visual data in education, for example, have not been subsumed within a separate domain of inquiry that corresponds to visual sociology or visual anthropology. They appear instead as pragmatic, visual dimensions of empirical inquiry in a broad range of studies. The evolution of visual recording technologies, media, and distribution systems could challenge this normalcy. However, for over a century now, education researchers have found ways of using new imaging tools to stimulate research and theorizing about education and schooling and to support fieldwork, pedagogy, and social advocacy. There is every indication they will be adding to this heritage for some time to come.

See also: Critical Ethnography; Ethnography; Semiotics.

Further Reading

- Banks, M. (2001). *Visual Methods in Social Research*. Thousand Oaks, CA: Sage.
- Bateson, G. and Mead, M. (1942). *Balinese Character: A Photographic Analysis*, vol. 2. New York: New York Academy of Sciences.
- Becker, H. (1986). Photography and sociology. *Doing Things Together: Selected Papers*, pp 271–272. Evanston, IL: Northwestern University Press.
- Board on International Comparative Studies in Education (2001). In Ulewicz, M. and Beatty, A. (eds.) *The Power of Video Technology in International Comparative Research in Education*. Washington, DC: National Research Council: National Academy Press.
- Coles, R. (1992). *Their Eyes Meeting the World: The Drawings and Paintings of Children*. New York: Houghton Mifflin.
- Coles, R. (1997). *Doing Documentary Work*. New York: Oxford University.
- Collier, J. J. (1967). *Visual Anthropology: Photography as a Research Method*. New York: Holt, Rinehart & Winston.
- Collier, J. (1973). *Alaskan Eskimo Education: A Film Analysis of Cultural Confrontation in the Schools*. New York: Holt, Rinehart and Winston.
- Dugdale, S. and Kibbey, D. (1990). Beyond the evident content goals. Part 1: Tapping the depth and flow of the educational undercurrent. *Journal of Mathematical Behavior* 9(3), 201–228.
- Dyson, A. H. (1989). *Multiple Worlds of Child Writers: Friends Learning to Write*. New York: Teachers College Press.
- Dyson, A. H. (1993). *Social Worlds of Children Learning to Write in an Urban Primary School*. New York: Teachers College Press.
- English, F. W. (1988). The utility of the camera in qualitative inquiry. *Educational Researcher* 17(4), 8–15.
- Erickson, F. (1975). Gatekeeping and the melting pot. *Harvard Educational Review* 45(1), 44–70.
- Erickson, F. and Schultz, J. (1981). When is a context? Some issues and methods in the analysis of social competence. In Green, J. and Wallat, C. (eds.) *Ethnography and Language in Educational Settings*, pp 147–160. Norwood, NJ: Ablex.
- Freedman, R. (1994). *Kids at Work: Lewis Hine and the Crusade against Child Labor*. New York: Clarion.
- Freire, P. (1981[1973]). *Education for Critical Consciousness*. New York: Continuum.
- Grady, J. (1996). The scope of visual sociology. *Visual Sociology* 11(2), 10–24.
- Harper, D. (2000). Reimagining visual methods: Galileo to Neuromancer. In Denzin, N. and Lincoln, Y. (eds.) *Handbook of Qualitative Research*, pp 717–732. Thousand Oaks, CA: Sage.
- Kulich, M. (1990). In Elwood, W. (ed.) *The Road to Brown*. US: California Newsreel.
- Margolis, E. (1999). Class pictures: Representations of race, gender, and ability in a century of school photography. *Visual Sociology* 14(1), 7–38.
- Mead, M. (1995). Visual anthropology in a discipline of words. In Hockings, P. (ed.) *Principles of Visual Anthropology*, pp 3–10. New York: Mouton de Gruyter.
- Mead, M. and Macgregor, F. C. (1951). *Growth and Culture: A Photographic Study of Balinese Childhood*. New York: Putnam.
- Mehan, H. (1979). *Learning Lessons: The Social Organization of Classroom Instruction*. Cambridge, MA: Harvard University Press.
- Nesper, J. (1997). *Tangled up in School: Politics, Space, Bodies and Signs in the Educational Process*. Mahwah, NJ: Erlbaum.
- Norris, S. (2004). *Analyzing Multimodal Interaction: A Methodological Framework*. New York: Routledge.
- Pink, S. (2001). *Doing Visual Ethnography*. London: Sage.
- Prosser, J. (ed.) (1998). *Image-Based Research: A Sourcebook for Qualitative Researchers*. Bristol, PA: Falmer.
- Rayner, K. (1998). Eye movements in reading and information processing: Twenty years of research. *Psychological Bulletin* 124, 372–422.
- Ruby, J. (2000). *Picturing Culture: Explorations of Film and Anthropology*. Chicago, IL: University of Chicago.
- Schwartz, D. (1989). Visual ethnography: Using photography in qualitative research. *Qualitative Sociology* 12(2), 119–154.

- Smith-Shank, D. L. (2003). Lewis Hine and his photo stories: Visual culture and social reform. *Art Education* 56(2), 33–37.
- Spindler, G. (1987). Cultural dialogue and schooling in Schoenhausen and Roseville: A comparative analysis. *Anthropology and Education Quarterly* 18(1), 3–16.
- Spindler, G. and Spindler, L. (1987). Schonhausen revisited and the rediscovery of culture. In Spindler, G. and Spindler, L. (eds.) *Interpretive Ethnography of Education: At Home and Abroad*, pp 143–170. Hillsdale, NJ: Erlbaum.
- Stanczak, G. (ed.) (2007). *Visual Methods and Social Research*. Thousand Oaks, CA: Sage.
- Stigler, J. W., Gonzales, P. A., Kawanka, T., Knoll, S., and Serrano, A. (1999). *The TIMSS Videotape Classroom Study: Methods and Findings from an Exploratory Research Project on Eighth-Grade Mathematics Instruction in Germany, Japan, and the United States (Research and Development Report: Third International Mathematics and Science Study (TIMSS) No. NCES 1999074)*. Washington, DC: National Center for Education Statistics.
- Thomas, P. (ed.) (2008). *Doing Visual Research with Children and Young People*. New York: Routledge.
- Tobin, J. J., Wu, D. Y. H., and Davidson, D. H. (1989). *Preschool in Three Cultures*. New Haven, CT: Yale University Press.
- Wagner, J. (ed.) (1979). *Images of Information: Still Photography in the Social Sciences*. Beverly Hills, CA: Sage.
- Wagner, J. (ed.) (1999). *Visual Sociology* 14(1&2): *Seeing Kids' Worlds* (vol. 14): IVSA.
- Worth, S., Adair, J., and Chalfen, R. (1997). *Through Navajo Eyes: An Exploration in Film Communication and Anthropology*. Albuquerque, NM: University of New Mexico Press.

Relevant Websites

- www.childes.psy.cmu.edu – CHILDES: Child Language Data Exchange System.
- www.coe.uh.edu – Educational Uses of Digital Storytelling.
- www.greenglobs.net – Green Globes and Graphing Equations.
- www.nces.ed.gov – Highlights from the TIMSS 1999 Video Study of Eighth-Grade Mathematics Teaching.
- www.iea.nl/timss-r_video.html – International Association for the Evaluation of Educational Achievement, TIMSS-R Video.
- www.visualsociology.org – International Visual Sociology Association.
- www.lessonlab.com – Lesson Lab: A Pearson Education Company.
- www.nbpts.org – National Board for Professional Teaching Standards.
- www.paulofreireinstitute.org – Paulo Freire Institute.
- www.pacttpa.org – Performance Assessment for California Teachers.
- www.societyforvisualanthropology.org – Society for Visual Anthropology.
- www.talkbank.org – TalkBank.
- www.memory.loc.gov – The Library of Congress: American Memory.
- www.loc.gov – The Library of Congress: The American Folklife Center.
- www.nces.ed.gov – Trends in International Mathematics and Science Study (TIMSS).
- www.tandf.co.uk – Visual Anthropology.
- www.visualanthropology.net – Visual Anthropology News and Resources.
- www.people.virginia.edu – Visual Anthropology Review.
- www.tandf.co.uk – Visual Studies.

QUALITATIVE RESEARCH – STANDARDS OF PROOF

Validity: Mapping Diverse Perspectives

G B Rossman and S F Rallis, University of Massachusetts Amherst, Amherst, MA, USA

A M Kuntz, University of Alabama, Tuscaloosa, AL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Connoisseurship (researcher as) – Researcher as expert; asserts evidence as support for expert judgment or interpretation.

Consequential ethics – Actions are judged good or bad because of their results within particular contexts; implies the ends justify the means.

Credible evidence – That which provides justification for researched assertions; reveals researcher's biases concerning what can be known and how one may come to such knowledge.

Nonconsequentialist ethics – Universal standards exist to guide all behavior regardless of consequences related to specific contexts.

Postmodernism – Anti-foundational perspective that explicates historically privileged research practices in order to disrupt commonsensical and unexamined ways of practicing research.

Postpositivism – Argues for truth claims that are plausible and based on testable causal explanations.

Poststructuralism – Questions all authoritative truths and the ability of language to fix meaning. Asserts power and knowledge as intimately linked, knowledges as contingent and historically produced.

Probity – Recognizes holistic truth in need of multiple perspectives for definition; asserts the integrity of the research project above all else.

Realism – Postpositivist perspective that asserts the existence of a real world yet inability for humans to fully observe causal mechanisms and underlying structures of that world. Aims to systematically exclude rival hypotheses as implausible.

Relational ethics – Foregrounds ethical implications of research; questions the meaning of actions and

evidence has for all individuals involved in research projects.

Social constructivism – Meaning is socially constructed, locally situated, and developed through individual experiences.

Trustworthiness – Assumptions about research design, process, and effects that generate assertions of merit, worth, and rigor.

Judging the soundness and credibility of research is central to scholarship, policy development, and program design. What to believe – what to put one's trust in – is foundational to the development of further research, sound policies, and programs that are designed on credible knowledge. The trustworthiness of qualitative inquiry, as with all research, depends on the appropriateness of the study's design for the research questions posed, the depth and rigor displayed in data collection, the richness and integrity of the analyses, and finally on the truth claims put forward, based on the evidence offered. These concerns are subsumed under the criteria or standards for judging the rigor and validity of a study, as well as the reliability with which the data were gathered.

Historically, reliability, validity, objectivity, and generalizability were viewed as the standards against which to judge all research, including qualitative inquiry. However, as the postmodern turn/s and their compelling critiques of normal social science have taken up residence in most social science discourses, these four canonical criteria have been challenged, as has the very notion of criteria or standards itself. Multiple and sometimes conflicting assumptions about the nature of truth (what we claim we know), reality (what we claim we know about), knowledge production (how we produce those claims), and evidence (what we argue supports those claims) now drive theorizing about inquiry as well as on-the-ground research practice. Standards are emergent and often contested, as advocates for a particular genre of qualitative inquiry argue that their are the proper or correct ones. The four canonical standards no longer obtain, at least in their original form. Yeats's (1920) comment is quite appropriate here: "Things fall apart; the

center cannot hold;/ Mere anarchy is loosed upon the world". Thus, judging the quality of qualitative inquiry is no longer a simple process or a formulaic one.

A range of approaches to validity, or trustworthiness, has evolved, reflecting differing assumptions about knowledge, truth, and strategies for observing the agreed-upon canons. We have clustered these into positivism, postpositivism, and postmodernity and poststructuralism. Within the variegated discourse, however, is little explicit discussion of the role of language or narrative form, specifically in terms of argument and evidence, which we discuss next.

The Validity Question: Argument, Evidence, and Authorship

Standards of proof within qualitative inquiry gain meaning and legitimacy from the larger critical contexts from which they emerge. Consequently, the very meaning of validity or truth, for example, is shaped by the assumptions inherent in critical social movements associated with positivism, postpositivism, and postmodernity. The positivist and, to some extent, the postpositivist perspectives focus on procedural rules for establishing the truth claims of research. Phillips (2006) invites us to move beyond methodological procedural rules to a consideration of how the truth-claim argument is constructed from the evidence presented and the warrants that connect the evidence to the assertions. The postmodern critiques offer a challenge to the author-free assumption of traditional social science writing. Central to this critique is the assertion that all texts are authored by a historically situated, political, gendered individual. Thus the notion (which seems obvious today) that the author is present in the text, that, in fact, the author has constructed the text, invites a critical read of the ways in which evidence is brought forward, assertions are supported, and how the evidence warrants those assertions. The focus shifts to the author as well as the content. All this comes under the purview of argumentation which fosters a close reading of the text *qua* text to discern how and in what ways the author has developed a convincing case for the tentative conclusions put forward. Thus, the validity question moves to a more complex place where the soundness/strength/validity of the argument, as well as the soundness/strength/validity of the research itself, becomes paramount. Much of this discussion is without the methodological discourse but, as Phillips (2006) notes, the main question to be asked is:

has the overall case made by the investigator been established to a degree that warrants the tentative acceptance of the theoretical or empirical claims that are being put forward? (p. 24, italics in original).

He notes that building an evidentiary case entails "scientists as making convincing cases, cases that appeal to a wide variety of evidence" (p. 1). The making of a case depends on the argumentation put forward: its logic, clarity, sources of evidence, warrants connecting the evidence to the assertions, and relevance. However, what is convincing varies across time, space, and sociopolitical context.

This article maps this contested terrain by first describing the historical period when the four canonical criteria held sway – the positivist persuasion. Based on the assumptions that one true reality exists and that social scientists can have direct, unmediated knowledge of that social world, this period witnessed the direct application of the canons developed in the natural sciences to qualitative inquiry. Resting on the philosophy of logical positivism, adherents argued for standards for inquiry that provided evidence of cause and effect to foster predictive statements about the social world, as natural scientists could (it was then presumed) do so about the forces of nature. As discussed below, some of these assumptions and normative demands for a positivist kind of validity remain in today's methodological discourse.

We next discuss the reaction – postpositivism – which softened the positivist's belief in a fully apprehensible world and introduced notions of a contingent, tentative knowing of the social world. Those of the postpositivist persuasion argue for inquiry that produces interpretations or truth claims that are plausible, based on the evidence put forward – thus the shift from proof to plausibility. Many nuances can be found within this large category; we discuss realism, social constructivism, connoisseurship, and relational ethics as key perspectives that adopt a less-absolutist ontology and focus more on individual sense making in the research process.

We then shift to the postmodern turn/s more explicitly and the development of poststructural critiques. Within this large, sometimes cacophonous group, are both gentle as well as radical arguments about validity. While some methodological scholars sought to displace the centrality of the term validity within discussion of qualitative inquiry, others utilize poststructural theory to interrogate validity as a contestatory site of multiple and competing discourses. The many contexts inherent in the inquiry process are now investigated with increased urgency as the lines which demarcate the boundaries of the research study blurred to become a site of inquiry themselves.

Historical Perspectives

The positivist period in qualitative inquiry suffered from what could be called physics envy, directing researchers' gaze to the processes and standards of rigor developed in the natural sciences which derive from a correspondence theory of truth. Reliability, validity, and generalizability

became the “holy trinity . . . , worshipped with respect by all true believers in science” (Kvale, 1996: 229). Early texts seamlessly applied these canons from the natural sciences to social science inquiry with little or no critique of the epistemological or ontological assumptions undergirding these standards. For example, Taylor and Bogdan’s 1984 text, *Introduction to Qualitative Research Methods: A Search for Meaning* (2nd edn.), notes that “qualitative researchers emphasize validity” (p. 7), providing procedural rules to ensure validity. These include describing the methods used in some detail (the research design, duration and intensity of fieldwork, number, and type of settings and informants, and analysis procedures including checks of informants’ statements) (pp. 151, 152). Precisely how a researcher is to check on informants’ statements as well as what to check for remain opaque.

A critical moment in the history of qualitative inquiry occurred when Lincoln and Guba published *Naturalistic Inquiry* in 1985. This text engaged with central notions that frame social interactions as well as all research: Do we believe in the claims that a text, a person, or the media put forward? On what grounds do we find these assertions credible? What evidence is put forward? How was that evidence gathered; by what standards do we evaluate it? Are these claims potentially useful for the issues at hand? These questions capture natural-science concerns with validity, reliability, objectivity, and generalizability, but broaden and deepen them. Lincoln and Guba reclaimed ordinary language terms (Kvale, 1996: 231), putting forward the constructs of credibility, dependability, confirmability, and transferability as alternatives to the canon. They argued that standards for trustworthiness should address the credibility of a study, whether it was systematically and rigorously conducted, and whether the study has potential usefulness and significance (see also Rossman and Rallis, 2003). They also offered procedural rules to help ensure that the standards would be met. To help meet the validity/credibility criterion, they urged the qualitative researcher to stay in the setting or with the participants for a long time (prolonged engagement); to conduct member checks where participants validate (or not) the emerging analyses and findings; and to engage in peer debriefing with trusted and yet critical colleagues to ensure that analyses are grounded in the evidence. This latter, along with creating detailed notes of the process for conducting an audit, helps to meet the dependability/reliability and objectivity/confirmability criteria foregrounded by earlier scholars.

Lincoln and Guba’s work was generative. Their arguments directly challenged the positivist cannon, noting specifically that objectivity is a chimera, as all social interactions are mediated by a perceiving individual. Thus, sanitized, person-free inquiry, objective research is not possible. They also introduced a challenge to the traditional canon’s emphasis on generalizability which

was possible only if the probabilistic conditions of randomly selected sampling occurred. Their argument that alternative logics – notably those of comparison and contrast as seen in legal reasoning and medical reasoning on a case-by-case basis – better suited the small sample size of much qualitative research (see also Kennedy, 1979).

Much of the subsequent discourse addressing rigor, validity, and generalizability in qualitative inquiry invokes their work, whether as a starting point for critique or to deepen the ideas that they proposed. In fact, there are few major texts about qualitative inquiry that do not cite this generative text. (See, e.g., Bodgan and Biklen, 2007; Creswell, 2002, 1998; Kvale, 1996; Marshall and Rossman, 2006; Patton, 2001; Wolcott, 2001; Yin, 2003.) Today, calls for meeting the positivist understandings, especially the realist ontology, of these canons continue, but with some softening, as we discuss below. For example, Morse and Richards (2002) assert that “determining reliability and validity remains the qualitative researcher’s goal” (p. 168) and that “claim[ing] that reliability and validity have no place in qualitative inquiry is to place the entire paradigm under suspicion” (p. 168). What is contested is how these crucial constructs are to be defined, by whom, for what research project, and for which audience/s. Addressing these questions, not too long after the publication of *Naturalistic Inquiry* (1985), Guba and Lincoln (1989) critiqued their earlier work as being “parallel, quasi-foundational, and clearly intended to be analogs to conventional criteria” (Schwandt, 1997: 165). They called for an emphasis on the outcomes of research, as well as the politics of both conducting and sharing results.

Postpositivist Perspectives: Keeping the Faith

As introduced above, with its confidence in apprehendible reality and testable causal explanations, positivism promised discovery of truth and a predictable world. The ruling standard of proof was validity; if the study had validity, the results could be trusted (Campbell and Stanley, 1966). The alternative epistemologies of postpositivist perspectives suggest different considerations of what counts as proof and evidence. To a postpositivist, reality is “only imperfectly and probabilistically apprehendible” (Guba and Lincoln, 2005: 195), and any claim to knowledge is both fallible and contingent. Yet, at times the language that some postpositivists use for determining what is legitimate and acceptable research does little beyond borrowing and building on the established positivist criteria. This section begins by discussing what might be called the realist postpositivist perspective, drawing specifically from the work of Maxwell (1996). It then moves to social constructivist perspectives which demonstrate

the shift to the community of discourse or practice as validators of research, then to connoisseurship and its emphasis on transparency of method, then the ethical imperative of probity. All, one way or another, articulate either procedural rules or guiding criteria to help ensure the trustworthiness of the research.

Realist Postpositivism

While the term postpositivism can include various philosophies, each recognizes the role of theory and interpretation in how we define reality and know the world. Realism, for example, asserts that while a real world exists, causal mechanisms and underlying structures are not directly observable; thus, humans use a variety of sense-making techniques to mediate between what is observed and what actually exists. The realist approach, however, still aims to produce an accurate representation of reality; so validity is an exercise to explore “how might you be wrong?” (Maxwell, 1996: 86). From this perspective, the qualitative researcher’s task is to rule out validity threats “using evidence, collected during the research itself, to make rival hypotheses implausible” (p. 88). This language is remarkably similar to that used in the classical positivist text on social science experiments by Campbell and Stanley (1966).

To elaborate, Maxwell (1996) describes distinct threats to validity in qualitative understanding: inaccurate or incomplete data threaten the validity of description; imposing the researcher’s own framework or meaning threatens valid interpretation; and theoretical validity is threatened by establishing vague and abstract propositions without consideration of discrepant data or alternative explanations. Drawing on his realist frame, Maxwell proposes a checklist of validity tests:

- The *modus operandi* method searches for alternative explanations or causes for each event.
- Searching for discrepant evidence and negative cases seeks to falsify any proposed conclusion; and unexplained instances can suggest modifications to a finding.
- Triangulation reduces risk of chance association and systematic biases.
- Soliciting feedback both from those who are familiar with the setting and from strangers may surface biases as well as flaws in logic or methods.
- Member checks offer specific feedback to rule out misinterpretation and to uncover alternative perspectives.
- Rich data (i.e., detailed and complete) provide full and revealing pictures on which to test one’s interpretations.
- Quasi-statistics allow assessment of the amount of evidence that bears on any given conclusion.
- Comparison provides a background against which to test a conclusion (pp. 92–98).

Finally, Maxwell declares that internal generalizability is critical for qualitative case studies. He asks whether descriptive, interpretive, and theoretical conclusions within the setting or group be generalized to the case as a whole. “If you are studying the patterns of interaction between the teacher and students in a single classroom, your account of that classroom as a whole is seriously jeopardized if you have selectively focused on particular students or kinds of interactions and ignored others” (p. 97). In summary, Maxwell, the realist methodologist, seeks to uncover the real arguments or evidence behind an event or interaction.

Social Constructivism

A broad, loosely connected set of perspectives, social constructivists, move away from a strict realist ontology and epistemology: knowledge is local; reality consists of individual experiences, the meaning of which is co-constructed within some community. This perspective stands in contrast with the realist approach above where demonstrating validity lies within the researcher’s and his/her text’s purview. Social constructivists explicitly argue that the strength of a study is determined socially – within a group. Still, many researchers of these persuasions rely on procedural rules for trustworthiness that are reminiscent of approaches to the positivist notion of validity: rigor, systematicity, and competent practice are words used as standards for credible evidence or proof. These approaches still focus on method to ensure interpretive rigor: Are data sources, perspectives, methods, and investigators triangulated? Was there prolonged engagement? Were member checks and peer-debriefing employed?

Some social constructivists, however, offer a somewhat different view of determining trustworthiness. Those of this set of persuasions view knowledge as a set of social artifacts, the reality of which can be observed in social interaction. Trustworthiness, or validation, lies within a “community of scientists” (Mischler, 1990: 422) who identify shared exemplars to guide further research. The essential criterion for judging a study to be trustworthy is “the degree to which we can rely on the concepts, methods, and inferences of a study as the basis for our own theorizing and empirical research” (p. 419). This approach implies an elite discursive validity, socially constructed among and applied by researchers to research.

Another group of constructivists reject any permanent, unvarying, or foundational standards on which knowledge is grounded or by which truth can be universally known; they propose a quite different conceptualization of research as social inquiry. Truth (with a capital T rather than a lower-case one) and any agreement regarding what is valid knowledge is seen as rising from the relationship between members of some stakeholding community and is thus subject to negotiation (Lincoln, 1995). Schwandt

(1996) suggests that instead of criteria to judge social inquiry – which he calls “regulative norms for removing doubts and settling disputes about what is correct or incorrect, true or false” (p. 59) – we practice inquiry as practical philosophy, that is, deep questioning or moral critique. This perspective evaluates the researcher and her research by how well her reports of the inquiry support the “training or calibration of human judgment” (p. 69) and “the capacity for practical wisdom” (p. 70). If this approach uses any criterion, it is the utility of the findings to produce social change (Schwandt, 2008).

Schwandt is not alone in linking credible evidence to use. Spillane and Miele point out that evidence use is an issue of evidence construction: “research does not construct evidence – people do” (Spillane and Miele, 2007: 47). A set of facts becomes evidence only when people interpret the facts as confirming, contradicting or complicating a proposed definition or solution to a problem. The meaning and value of the evidence lies in its use in what they call work practice (p. 58). People construct evidence together within the context in which it is observed or collected, not alone. Credible evidence, then, is a product of collective action while examining practice settings where “new information is encountered (or overlooked), interpreted as evidence of one thing or another (or dismissed as irrelevant), and eventually put to use (or ignored)” (p. 58). This symbolic interaction perspective (Blumer, 1969) assumes that meaning is constructed and reconstructed during human interaction.

Connoisseurship and Critique

The researcher as connoisseur (one who appreciates) or critic (one who discloses) stands as an expert who has had a great deal of experience with the persons, objects, situations, or activities being considered. This researcher questions the need and usefulness, indeed the appropriateness, of the literal and restricted definition of knowledge that the so-called scientific researchers follow: “scientific knowledge is seldom true in the literal sense. Especially in the social sciences where metaphor, analogical reasoning, and hypothetical constructs abound, literal truths are scarce” (Eisner, 1991: 108). As the connoisseur or critic deals with “matters as complex and subtle as the description, interpretation, and evaluation” of human interaction (p. 109), validity is seen as probable or judged by what is reasonable. Eisner (1991), the spokesperson for this perspective, posits that “we can secure no unmediated grasp of things as they ‘really are’”, so “we are ‘stuck’ with judgments and interpretations” (p. 109). The appreciative or critique perspective sees evidence, then, not as providing proof for a given truth but as support for a particular interpretation or judgment. The connoisseur or critic clearly does not believe that anything goes in qualitative inquiry. In fact, Eisner (1991) offers three evidentiary bases

for judging credibility: structural corroboration, consensual validation, and referential adequacy.

The evidentiary bases serve to persuade the reader in favor of the critic’s interpretation. Structural corroboration is, like triangulation, a process “through which multiple types of data are related to each other to support or contradict the interpretation and evaluation of a state of affairs” (Eisner, 1991: 110). The weight and coherence of evidence allows the researcher to build a credible case that is tightly argued and persuasive. Consensual validation implies that competent others agree with the description, interpretation, and evaluation put forth by the researcher or critic. Evidence of the plausibility of the argument, description, and judgment lies in the overlap among those who have described and evaluated the event, object, or phenomenon. The consensus is won among readers who are persuaded by what the critic or researcher has to say. Thus, transparency of method and criteria used is essential to persuade the reader to accept any particular interpretation. A referentially adequate criticism enables the reader to “locate in its subject matter the qualities and the meanings [the critic] ascribes to them” (Eisner, 1991: 114). Put simply, the reader can see what the critic sees and why. Again, transparency of method and criteria is essential. All three of these evidentiary bases work together; a good critic uses multiple sources and types of data, puts forward her methods and criteria so readers can judge their agreement (or not), and offers a picture sufficiently detailed to reveal the grounds for her interpretation.

Relational and Ethics Probity

The previous discussions have considered validity *qua* validity and have addressed ethics as a separate, sometimes equally important consideration for judging the quality of qualitative inquiry. In a departure from this norm, but still within the postpositivist perspective, two of the authors (Rossman and Rallis, 2003; Rallis and Rossman, 2004) argue that the trustworthiness of qualitative research is judged by two sets of standards: Does the study meet agreed-upon canons for acceptable and competent practice (including validity), and does it demonstrate sensitivity to ethical issues? We argue that both considerations are intimately linked because a study might meet accepted standards for practice but, if not ethically conducted, the work falls short. From this perspective, an unethical study is not a trustworthy study. “Bad science makes for bad ethics” (Rosenthal, 1994: 128), and bad ethics makes for bad science. This ethical perspective claims that trustworthiness should depend not only on whether the researcher got the technical matters right, but it must also be judged by how well the researcher got the relational matters right. Relational matters – the long-term, thoughtful, and sensitive interactions, built through probity – are seen as central to ethical considerations.

Probity implies wholeness and integrity; it acknowledges that truth is a holistic concept, is more than meets the eye, and demands multiple avenues toward full understanding. Moreover, probity requires rigorous reasoning through moral principles to determine moral soundness. “The point of moral principles is to regulate interactions among human beings” (Strike *et al.*, 1988: 41) – and what is social science research but an interaction among human beings? Thus, this perspective asks: How are ethical considerations embedded in the demonstration of evidence to support the claims made? What moral principles guide choices for collecting, analyzing, and interpreting the data to inform the findings? What reasoning lies behind the choice of principles? How does the researcher demonstrate and enact the principles in both planning and implementing the study?

This perspective argues that ethical theories, grounded in moral principles, provide direction for researchers to help ensure the overall trustworthiness of a study. Philosophers and ethical theorists have put forward various theories to analyze or direct ethical behavior, and these theories can be grouped into two broad categories – consequentialist and nonconsequentialist – according to the criteria each uses to determine the moral correctness of a particular action, in this case, in a research context.

Consequentialist ethics use the results of actions to determine their rightness or wrongness. Such reasoning implies that ends justify means. Any particular action is either good or bad because of its results in a particular context – its consequences. The best-known example of this category is utilitarianism, “the doctrine that the greatest good of the greatest number should be the guiding principle of conduct” (Brown, 1993: 3534). The moral principle of utilitarianism declares “that actions are right if they are useful or for the benefit of a majority” (Brown, 1993: 3534). Among themselves, utilitarian philosophers debate about what is good and for whose good. Consequentialists may rely on usefulness – both degree and type – as an evidentiary base. They ask: What happens when the research findings are used? Whom do they affect? In what ways? In short, what are their consequences?

In contrast, nonconsequentialist ethical theories derive from the moral principle that universal standards exist to guide all behavior, regardless of the consequences in a specific situation. Three nonconsequentialist moral principles are individual rights and responsibilities, social justice, and care. The ethic of rights and responsibilities upholds the unconditional worth of all human beings and the respect to which they are entitled. This ethic also mandates the corresponding obligations to protect those rights. Reasoning from this moral principle, the trustworthiness of a study is judged by the degree to which and in what ways the conduct of the research and its written product respect the participants’ rights, not by its outcomes or consequences.

The ethic of justice argues for the redistribution of resources and opportunities to promote equity above equality. It relies on principles of fairness and equity to judge the trustworthiness of research decisions and actions. Justice aims to ensure that everyone is better off, even though the allocation of a benefit may differ across groups. Such apparently unequal treatment is justified because not attending to the least is to hurt the whole. Specifically from this perspective, then, the purpose of research is to benefit those who have been excluded or deprived (less to the privileged; more to the needy). This perspective demands that qualitative researchers be diligent in collecting evidence that represents participants who have been marginalized or previously silenced. As noted above, Schwandt (2008) espouses an ethic of justice when he argues that trustworthy research leads to social change whether in the conduct of a study or in its application (resonating with a consequentialist perspective, also).

Finally, this relational perspective draws on another nonconsequentialist moral principle that emphasizes the interdependence of people. Articulated as an ethic of care, this principle addresses the effect any action in a research context is likely to have on human relationships. This ethic honors and respects the intimate connections inherent in human interaction. Gilligan (1982) and Noddings (1984) are two caring theorists who first defined the relational-self in contrast to the individual-self. As they see individuals through their relationships with others, actions have multiple meanings depending upon the particular attachments formed. Rather than relying on general laws or principles, the caring researcher focuses on the particular setting and relationships formed in the research context. Trustworthy research considers what meaning the action and resulting evidence would have for the individuals involved.

These nonconsequential ethical theories are concerned with ongoing actions and interactions. From this perspective, therefore, trustworthy evidence is found not in presumed or hoped-for use but in daily practice from a moral stance. All the postpositivist perspectives discussed above, while there is variation among them, tend to hold to the belief that guiding principles or procedural rules can assure the validity and trustworthiness of qualitative inquiry. We turn now to the postmodern critique and, specifically, poststructural challenges to these more traditional notions of validity and its related questions of truth, reliability, and evidence.

Postmodern Turns and Poststructural Interrogations: Deconstructing, Disentangling, and Reconsidering Validity

The postmodern turn in the humanities and social sciences has spawned an interrogation of traditional forms of all social science inquiry, particularly qualitative inquiry.

Issues of authorship (who is speaking/writing? by what authority?), representation (who is speaking for whom about what?), and truth claims (on what grounds is truth presented?) are entangled in a complex web. Central to these interrogations are issues of evidence and legitimacy, frequently applied to the traditional, highly formalized genre of academic writing. Thus, academic writing itself – the production of text that represents the processes and results of an inquiry project – has come under increasing scrutiny. Much of this critique comes from social theory and literary criticism and is seen in a generative experimentation with alternative forms of representation and evidence. The justification for this experimentation – other than purely for purposes of critique – comes from the insight that, if postmodern and critical perspectives take seriously one of their fundamental assumptions – that the social world is multifaceted, contradictory, confusing, and political in nature – then representations of this multiplicity should be witnessed in multiple, equally legitimate forms.

Specifically, the postmodern turn within qualitative inquiry brings an antifoundational stance and a suspicion of the regulatory norms that govern the true, real, or valid. Consequently, within the postmodern frame, standards of evidence are evaluated based on their reproduction of, or deviance from, historically privileged discourses of truth. In short, postmodernity calls into question the assumptions which make validity possible. As Corbin and Strauss (2007) note:

The notion of judging the quality of research seemed so clear before postmodernist and constructionist thinking pointed out the fallacies of some of our ways. Now I wonder, if findings are constructions and truth a ‘mirage,’ aren’t evaluative criteria also constructions and therefore subject to debate? (p. 297).

Postmodernity has invoked many qualitative researchers, including some of those discussed above, to eschew the terms validity and reliability, due to their historical alliance with quantitative research, and favor credibility and trustworthiness (e.g., Agar, 1986; Corbin and Strauss, 2007). However, some poststructural qualitative methodologists seek to reorient the notion of validity as it pertains to the process of doing research rather than the findings and assertions of the research product (e.g., Beach, 2003; Koro-Ljungberg, 2008; Kvale, 1995; Lenzo, 1995). With roots in postpositivist qualitative methodological writings (see the discussion above on social constructivism, connoisseurship, and relational ethics), this more critical poststructural group situates validity as an historical, power-laden discourse, often used to demarcate a line between science and not-science, and is perhaps best exemplified by the work of Patti Lather.

In her analysis of validity within contemporary times, Lather (1993) asks the simple question, “what do you do with validity once you’ve met poststructuralism?” (p. 676).

Lather (2001) recognizes the question of validity as a continually emergent discourse that can never be avoided or resolved fully. Rather than simply abandoning the notion of validity, Lather (1993) asks qualitative scholars to interrogate the term, questioning what a fluid, context-sensitive understanding of validity might look like. She proposes situated methodologies that convey the complexity of individuals interacting with sociohistorical discourses. Ultimately, Lather and others, who take on a poststructural perspective in qualitative research, focus a critical lens on research practices as a means for evaluating validity within qualitative work.

Poststructuralism carries with it several critical assumptions:

- There is no authoritative truth or reality.
- Binaries, such as true/false, self/other, and researcher/participant are artificial, socially constructed, and usefully collapsed.
- Knowledge is contingent, constructed within sociohistorical contexts.
- Truth claims are limited and partial, never fixed.
- Language insufficiently articulates meaning.
- Power and knowledge are intimately linked, both historically produced and enacted/maintained by discursive practices.
- Prescriptive assertions evoke modernist perspectives of the world as knowable.

One example of how such assumptions alter one’s practices within qualitative inquiry is a shift away from triangulation toward the metaphor of crystallization. The triangle is seen as a rigid, two-dimensional and fixed shape that assumes a static object to be triangulated. As Richardson (1997) writes, the crystal

combines symmetry and substance with an infinite variety of shapes, substances, transmutations, multidimensionalities, and angles of approach . . . Crystals are prisms that reflect externalities *and* refract within themselves, creating different colors, patterns, arrays, casting off in different directions. What we see depends upon our angle of repose (p. 92; original emphasis).

This critical perspective on validity as contingent, a crystal that gains meaning through its variety, calls forth important distinctions between positivist or postpositivist prescriptions for validity and poststructural inquiry. Lather (2001) juxtaposes regulatory validity (that which demarcates the borders of science and not-science) with constitutive practices of validity (those relational and social contexts in which knowledge is claimed) and correspondence validity (emphasizing essential beliefs connected to a broader reality) with transgressive practices of validity (attempts to deconstruct inquiry as a cultural practice) (p. 246). Validity is thus a discursive site, the intersection of diverse knowledge, which never

produces certainty, only the complexity of contextualized truth claims.

Validity in poststructural qualitative research also requires that a methodology incorporate its own mechanisms for self-critique (Lather, 1986, 1993; Lenzo, 1995). Such reflexivity must situate a study within local contexts. Consequently, research practices “cannot be prescribed ahead of time, but can be negotiated in ways that address poststructural problematics involving power, language, representation, politics, and ethics” (Lenzo, 1995: 17). Context, in this sense, builds on Foucault’s (1984) assertion that sociohistorical discourses shape our understanding of language, as well as produce normative interpretations of reality; knowledge is made possible through discursive exercises of power.

Perhaps the poststructural construct of validity that is best known is Lather’s (1986) conception of catalytic validity: the manner in which the process of research re-orientes participants to their reality to stimulate transformative possibilities. Scholars such as Beach (2003) assert catalytic validity as a productive counter to traditional notions of validity, which resists the capitalistic tendencies of current methodological practices. Participants actively collaborate on research projects and co-create the knowledge such projects present. Research projects that take seriously catalytic validity involve the active documentation of researcher and participant interaction with the project context, the types of discourses in which such interactions occur, and the events that give shape to particular constructions of knowledge (Koro-Ljungberg, 2008). Catalytic validity requires that researchers self-reflectively recognize the ontological and epistemological assumptions which govern their interactions with their projects, the very methodological practices they employ.

Final Words

As this article has demonstrated, the markers that determine the worth of qualitative work are contested and set against a larger backdrop of critical social movements. How one determines the validity, trustworthiness, reliability, or even worth of a qualitative study depends on the foundational assumptions that govern the critical tradition to which one subscribes. We have sought to map out how key social movements and scholars have addressed the contested terrain of validity, demonstrating how such perspectives draw their definition in relation to one another. This is not simply an evolving debate – the definitions of such key terms do not systematically become more precise and succinct with the progression of time. Qualitative inquiry takes seriously the complexities of social interaction, the dynamic relations of human meaning making. Consequently, those inquirers who engage in qualitative study necessarily take part in a messy endeavor – a social process heavy with meaning. Whatever theoretical

orientation informs the modes of inquiry, the assumptions guiding the research are diligently articulated through an ongoing set of daily practices. To close, we return to the generative work of Lincoln and Guba (1985) who remind qualitative researchers that a trustworthy study is one whose findings are “worth paying attention to, worth taking account of” (p. 290). As this article has demonstrated, who determines the worth is contested.

See also: Grounded Theory; Hermeneutics; Narrative Inquiry; Phenomenology.

Bibliography

- Agar, M. (1986). *Speaking of Ethnography*. Beverly Hills, CA: Sage.
- Beach, D. (2003). A problem of validity in education research. *Qualitative Inquiry* 9(6), 859–873.
- Blumer, H. (1969). *Symbolic Interactionism*. Englewood Cliffs, NJ: Prentice Hall.
- Bodgan, R. C. and Biklen, S. K. (2007). *Qualitative Research for Education: An Introduction to Theory and Methods*, 5th edn. Boston, MA: Pearson.
- Brown, L. (ed.) (1993). *Oxford English Dictionary*, vol. 1. Oxford, UK: Clarendon.
- Campbell, D. T. and Stanley, J. C. (1966). *Experimental and Quasi-Experimental Designs for Research*. Skokie, IL: Rand McNally.
- Corbin, J. and Strauss, A. (2007). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 3rd edn. Thousand Oaks, CA: Sage.
- Creswell, J. W. (1998). *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*. Thousand Oaks, CA: Sage.
- Creswell, J. W. (2002). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Upper Saddle River, NJ: Merrill-Prentice Hall.
- Eisner, E. W. (1991). *The Enlightened Eye: Qualitative Inquiry and the Enhancement of Educational Practice*. New York: Macmillan.
- Foucault, M. (1984). Polemics, politics, and problematisation. In Rainbow, P. (ed.) *The Foucault Reader*, pp 381–390. New York: Southern.
- Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women’s Development*. Cambridge, MA: Harvard University Press.
- Guba, E. G. and Lincoln, Y. S. (1989). *Fourth-Generation Evaluation*. Newbury Park, CA: Sage.
- Guba, E. G. and Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions, and emerging confluences. In Denzin, N. K. and Lincoln, Y. S. (eds.) *Sage Handbook of Qualitative Research*, 3rd edn., pp 191–215. Thousand Oaks, CA: Sage.
- Kennedy, M. M. (1979). Generalizing from single case studies. *Evaluation Quarterly* 12, 661–678.
- Koro-Ljungberg, M. (2008). Validity and validation in the making in the context of qualitative research. *Qualitative Health Research* 18(7), 983–989.
- Kvale, S. (1995). The social construction of validity. *Qualitative Inquiry* 1(1), 19–40.
- Kvale, S. (1996). *Interviews: An Introduction to Qualitative Research Interviewing*. Thousand Oaks, CA: Sage.
- Lather, P. (1986). Issues of validity in openly ideological research: Between a rock and a soft place. *Interchange* 17(4), 63–84.
- Lather, P. (1993). Fertile obsession: Validity after poststructuralism. *Sociological Quarterly* 34(4), 673–693.
- Lather, P. (2001). Validity as an incitement to discourse: Qualitative research and the crisis of legitimation. In Richardson, V. (ed.) *Handbook of Research on Teaching*, 4th edn., pp 241–250. Washington, DC: American Education Research Association.
- Lenzo, K. (1995). Validity and self-reflexivity meet poststructuralism: Scientific ethos and the transgressive self. *Educational Researcher* 24(4), 17–45.

- Lincoln, Y. S. (1995). Emerging criteria for quality in qualitative and interpretive research. *Qualitative Inquiry* 1, 275–289.
- Lincoln, Y. S. and Guba, E. G. (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage.
- Marshall, C. and Rossman, G. B. (2006). *Designing Qualitative Research*, 4th edn. Thousand Oaks, CA: Sage.
- Maxwell, J. A. (1996). *Qualitative Research Design: An Interactive Approach*. Thousand Oaks, CA: Sage.
- Mischler, E. G. (1990). Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard Education Review* 60, 415–441.
- Morse, J. M. and Richards, L. (2002). *Read Me First for a User's Guide to Qualitative Methods*. Thousand Oaks, CA: Sage.
- Noddings, N. (1984). *Caring, a Feminine Approach to Ethics and Moral Education*. Berkeley, CA: University of California Press.
- Patton, M. Q. (2001). *Qualitative Research and Evaluation Methods*, 3rd edn. Thousand Oaks, CA: Sage.
- Pelto, P. J. and Pelto, G. H. (1978). *Anthropological Research: The Structure of Inquiry*, 2nd edn. Cambridge: Cambridge University Press.
- Phillips, D. C. (2006). A guide for the perplexed: Scientific educational research, methodolatry, and the gold versus platinum standards. *Educational Research Review* 1, 15–26.
- Rallis, S. F. and Rossman, G. B. (2004). Trustworthiness: Beyond procedures to probity. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Diego, CA.
- Richardson, L. (1997). *Fields of Play: Constructing an Academic Life*. New Brunswick, NJ: Rutgers University Press.
- Rosenthal, R. (1994). Science and ethics in conducting, analyzing, and reporting psychological research. *Psychological Research* 5(3), 127–134.
- Rossman, G. B. and Rallis, S. F. (2003). *Learning in the Field: An Introduction to Qualitative Research*. Thousand Oaks, CA: Sage.
- Schwandt, T. A. (1996). Farewell to criteriology. *Qualitative Inquiry* 2, 58–72.
- Schwandt, T. A. (1997). *Qualitative Inquiry: A Dictionary of Terms*. Thousand Oaks, CA: Sage.
- Schwandt, T. A. (2008). *Reflections on the Civic Obligations of Qualitative Inquirers*. Invited address, Division D, American Educational Research Association, March 27, Chicago.
- Spillane, J. P. and Miele, D. B. (2007). Evidence in practice: A framing of the terrain. In Moss, P. A. (ed.) *Evidence and Decision Making, the 106th Yearbook of the National Society for the Study of Education, Part I*, pp 46–73. Malden, MA: Blackwell.
- Strike, K., Haller, E., and Soltis, J. (1988). *The Ethics of School Administration*. New York: Teachers College Press.
- Taylor, S. J. and Bogdan, R. (1984). *Introduction to Qualitative Research Methods: A Search for Meaning*, 2nd edn. New York: Wiley.
- Wolcott, H. F. (2001). *Writing Up Qualitative Research*, 2nd edn. Thousand Oaks, CA: Sage.
- Yeats, W. B. (1956). The Second Coming. In *The Collected Poems of W. B. Yeats* (Definitive edition with the author's final revisions). New York: Macmillan.
- Yin, R. K. (2003). *Case Study Research: Design and Methods*, 3rd edn. Thousand Oaks, CA: Sage.

Further Reading

- Denzin, N. K. and Lincoln, Y. S. (eds). *Sage Handbook of Qualitative Research*, 3rd edn. Thousand Oaks, CA: Sage.
- Eisner, E. W. (1991). *The Enlightened Eye: Qualitative Inquiry and the Enhancement of Educational Practice*. New York: Macmillan.
- Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher* 33(2), 3–11.

QUANTITATIVE AND QUALITATIVE APPROACHES TO RESEARCH – INTEGRATION

Mixed Methods

A Tashakkori and I Newman, Florida International University, Miami, FL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Inference – A conclusion, understanding, or explanation derived from the results of data analysis in a research study.

Inference quality – The degree to which the interpretations and conclusions of a study are seen as valid, credible, acceptable, and more plausible than other possible explanations of the findings.

Inference transferability – The generalizability or applicability of conclusions made on the basis of the results of a study to other individuals, entities, settings, situations, time periods, or methods of observation/measurement.

Mixed method – A type of research design in which the findings of qualitative and quantitative methods/approaches are integrated in order to gain a fuller understanding of the phenomenon under investigation.

Mixed-methods research systematically integrates quantitative and qualitative approaches to answer complex research questions that might benefit from such a combination. One of its major strengths is that it empowers the researcher to use the most appropriate techniques for answering the research questions. Using an integrated approach has gained increasing recognition in the past two decades. Such recognition is evident in a relatively large number of books (e.g., Bergman, 2008b; Creswell and Plano Clark, 2007; Greene, 2007; Ridenour and Newman, 2008; Tashakkori and Teddlie, 2003; Teddlie and Tashakkori, 2009), numerous articles in professional and academic journals, dissertations, conference presentations, and other outlets (also see the section titled 'Further Reading'). Establishment of the *Journal of Mixed Methods Research*, and the successful continuation of the International Mixed Methods Conference for the 5th year in The School of Healthcare, University of

Leeds, UK (July, 2009), are other examples of the widespread popularity and acceptance of mixed-methods research in the social, behavioral, educational, and health sciences.

Traditionally, quantitatively oriented researchers have worked within the postpositivist worldview in the past few decades, and are predominantly interested in collecting numerical data with structured data-collection methods. They analyze their data using statistical techniques, and maintain a degree of objectivity in their interpretations. On the other hand, qualitatively oriented researchers emerged as a distinct reaction to what was seen, in the 1980s and 1990s, as the dominant and excessive practices of quantitative research. Qualitative scholars follow a constructivist worldview and are predominantly interested in collective narrative data using open-ended (holistic) procedures. They typically analyze their data via thematic analysis, and make holistic, value-based, and subjective inferences on the basis of the findings. Although qualitative research is often assumed to be exploratory, both qualitative and quantitative research may be exploratory/descriptive or explanatory/confirmatory. The dichotomizing based on this aspect of research is too simplistic to be valid (see Bergman, 2008a; Teddlie and Tashakkori, 2009).

Negating the qualitative–quantitative dichotomy, mixed methods has emerged as a third alternative – allowing researchers to take relative stances on a continuum between the two. Mixed-methods researchers identify themselves as integrative and eclectic, work primarily within a pragmatic worldview, collect both narrative and numerical data, employ both structured and emergent designs, analyze their data both via statistical and content analysis, and make meta-inferences as answers to their research questions by integrating the inferences gleaned from their qualitative and quantitative findings.

According to Greene *et al.* (1989) (see the section titled 'Further Reading'), mixed-methods researchers combine the characteristics of quantitative and qualitative traditions that yield answers to questions that neither method alone can answer. Mixed methods potentially enable the

researchers to examine research problems from multiple perspectives and types of (qualitative and quantitative) evidence from multiple sources, leading to conclusions that may be complementary and/or confirmatory. Using an integrated approach allows one to generate new theoretical explanations or expand the current ones, as well as test these understandings in a single study, if needed (Ridenour and Newman, 2008).

Despite the general assumption that integrated methodology has emerged as an alternative to the strict dichotomy of the qualitative and quantitative approaches of the past century, in truth, it is well rooted in the eclectic approaches of social/behavioral science researchers in the first half of the twentieth century (see Brewer and Hunter, 2006; Teddlie and Tashakkori, 2009; Waszack and Sines, 2003 for more details). Consistent with this inherent return-to-roots characteristic, the dominant tenet of mixed-methods research is that the research purpose and question (not the paradigm) dictates the method of inquiry.

Mixed-methods scholars, however, do not ignore the possible impact of investigators' culture, worldview, and sociopolitical context on choosing and formulating research questions. These influences are often inherent in the purposes of the study, which, in turn, shape the research questions, and eventually lead the researcher's methods of study (see Newman *et al.*, 2003). Understandably, only some research purposes and questions are amenable to mixed methods. Many others might be answered more parsimoniously by utilizing either qualitative or quantitative approaches, as appropriate.

Purposes and Utility of Mixed-Methods Research

Admittedly, one of the main reasons for the popularity of mixed methods is that it is more effective for answering some research questions. These are often the type of questions that are not readily answerable by either qualitative or quantitative methods, or the purposes underlying them are not reachable via qualitative or quantitative approaches, "mental models (Greene, 2007)," and techniques. At least seven reasons are often given for using such methods:

- **Complementarity:** Using both mixed methods for the purpose of integrating two expertly different but related answers to the research question, one gleaned from using a qualitative and the other from a quantitative approach.
- **Completeness:** Using mixed methods in order to obtain a fuller understanding of the phenomenon under investigation. The expectation is that such a more

complete understanding will emerge if the inferences from qualitative and quantitative strands of a study are merged effectively (such an effective integration has been called "integrative efficacy" of mixed methods inferences).

- **Development:** Mixed methods is conducted with the explicit (preplanned or emergent) purpose of obtaining research questions, sampling framework, or data sources of a second (e.g., a qualitative) strand of the study from the first strand (e.g., quantitative).
- **Expansion:** Same as 'developmental' discussed above, but with the purpose of expanding the answers already obtained in a previous strand of a study.
- **Corroboration/Confirmation:** Utilizing integrated methods in order to evaluate the credibility of inferences obtained from a (qualitative or quantitative) strand of a study. The research questions of the first strand are often exploratory, while the research questions of the second strand are often explanatory/confirmatory.
- **Compensation:** Utilizing mixed methods with the express purpose of compensating the weaknesses of one approach (e.g., data correction errors/biases) with the strengths of the other.
- **Diversity:** Conducting mixed methods with the hope/purpose of comparing and contrasting divergent pictures of the same phenomenon. (Dialectic approach in which the answers are pitted against each other; Greene, 2007).

Research Questions in Mixed-Methods Research

As we discussed earlier, the main purpose for utilizing a mixed-methods design is to answer research questions that are not effectively answerable by either a qualitative or quantitative approach alone. Such questions often have interrelated components – each following qualitative or quantitative orientations, such as questions including 'what and how', or 'what and why'.

Mixed-methods questions are often seen in two non-mutually exclusive ways: (1) an over-arching question that includes both qualitative and quantitative type of sub-questions, or (2) a collection of separate qualitative and quantitative type of questions, followed by a question with regard to the nature of integration (i.e., how do the answers to the two types of questions relate to each other; see Creswell and Plano Clark, 2007). The first variation is more frequently seen in the literature. For example, one might ask: "What are the effects of teaching method X on achievement and self-perceptions of groups A and B?" Such a question might be followed by three subquestions asking: (1) "Are groups A and B different in achievement and self-perceptions?" (2) "What are the perceptions and constructions of participants in

groups A and B regarding teaching method X and its impact?” and (3) How does teaching method X work differently in the two groups (if it indeed did)?” The subquestions are answered in separate strands (phases) of the study (Teddle and Tashakkori, 2009: 133).

The second approach mentioned above includes explicitly separate qualitative or quantitative type of questions that are clearly interrelated. They are followed by an explicit question about the nature of integration (for more information, also see Tashakkori and Creswell, 2007). Such a question might be stated upfront, or might emerge during the course of a study, such as in emergent sequential designs discussed above. Examples of such preplanned or emergent questions include: “do the findings of the two strand of the study agree?” or “how do the two sets of answers contribute to a comprehensive understanding of the phenomenon under study?”

Design of Mixed-Methods Studies

A grossly simplified framework for classifying mixed-methods designs is to divide them into four families that simultaneously include purposes and expected outcomes, and the sequence of procedures (parallel, sequential, and conversion). Within each family of designs, there is variation on the basis of units of study (e.g., including individuals, social units, or both). Three basic families of design have been described as parallel, sequential, and conversion. A fourth family (fully integrated) includes combinations of the other three basic (simpler) families of designs (see Tashakkori *et al.*, 2009).

Decision to adopt any of these families emerges from an investigator’s answers to two broad questions – one addressing the sequence of strands/phases, and the other exploring the possibility of converting one type of data to another:

1. Which sequence of steps will provide a better opportunity to answer my research questions: Collecting

and analyzing data sequentially (sequential mixed-methods study) or in a parallel manner (parallel mixed-methods design)? In parallel designs, two sets of data (one qualitative and one quantitative) are collected and analyzed independently at the same time or with a time lag (see **Figure 1**), in order to answer a mixed research question. In sequential designs, the second round of data collection and analysis is rooted in (initiated from, made possible by, or modified according to) the results of the first set (see **Figure 2**).

In parallel mixed studies, units of data collection and analysis might be selected in three forms: (1) the sample might consist of two related groups of individuals from the same population (parallel mixed-methods design, multisample); (2) all or a subset of individuals in the first sample are selected for the second strand (parallel mixed-methods design, same/subsample); and (3) the two samples might be at two different levels of organization (parallel mixed-methods design, multilevel). For example, an investigator might decide to select a stratified sample of 30 classroom teachers for in-depth ethnographic interviews, while also administering a structured questionnaire to all students in their classes.

In sequential mixed designs, the purpose, questions, sample, data, or other components of the second strand are rooted in the results of the first strand (see **Figure 2**). For example, based on the results of the first strand of a study that tries to identify the predictors of success in certain groups of economically disadvantaged students, you might also decide to select few highly successful and highly unsuccessful students on the basis of the results, and interview them in depth. The second strand of the study (qualitative in-depth interviews) is conducted in order to explain/expand/confirm the results of the first strand. The second strand of a sequential study might be preplanned, or the need for it might emerge because the results of the first strand are inconsistent, unexpected, or need expansion (see Tashakkori *et al.*, in press for further details).

Similar to parallel mixed studies, in sequential mixed-methods designs the samples might be selected

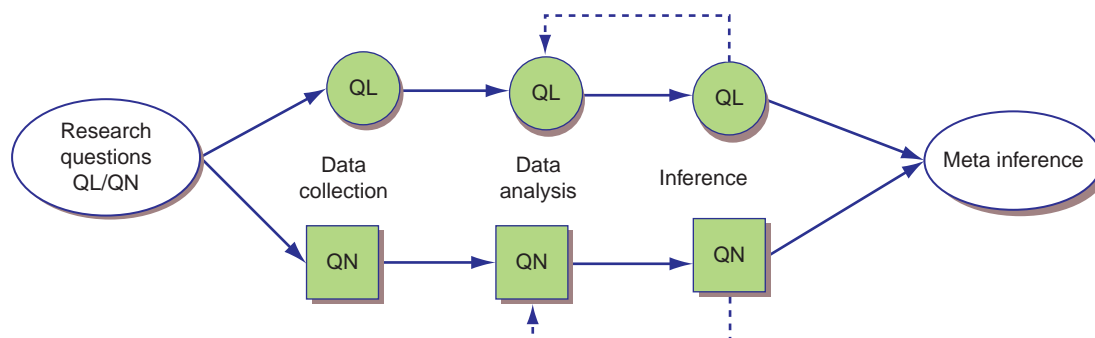


Figure 1 Graphic presentation of parallel (QL + QN) mixed-methods designs with multiple questions, inferences gleaned from the findings of two strands, and a final set of integrated meta-inferences.

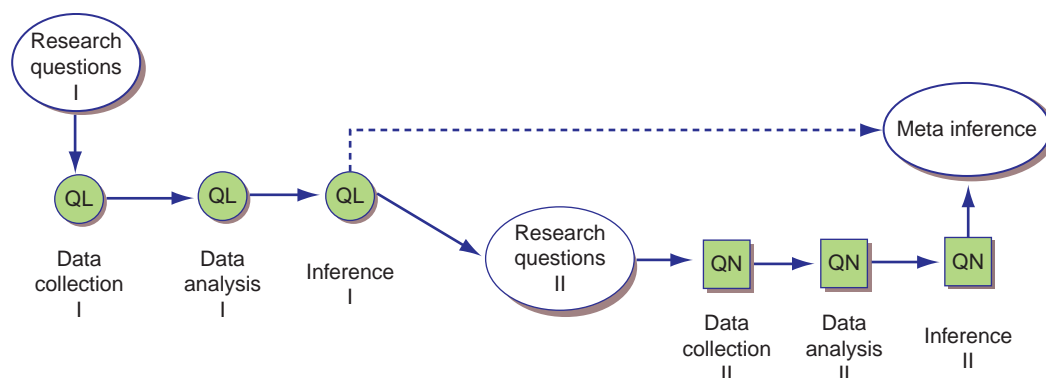


Figure 2 Graphic presentation of a sequential (QL → QN) mixed-methods design in which research questions of the second strand emerge from the findings of the first. A final set of integrated meta-inferences are made on the basis of both sets of findings. (Please note that this family of design also includes a QN → QL sequence.)

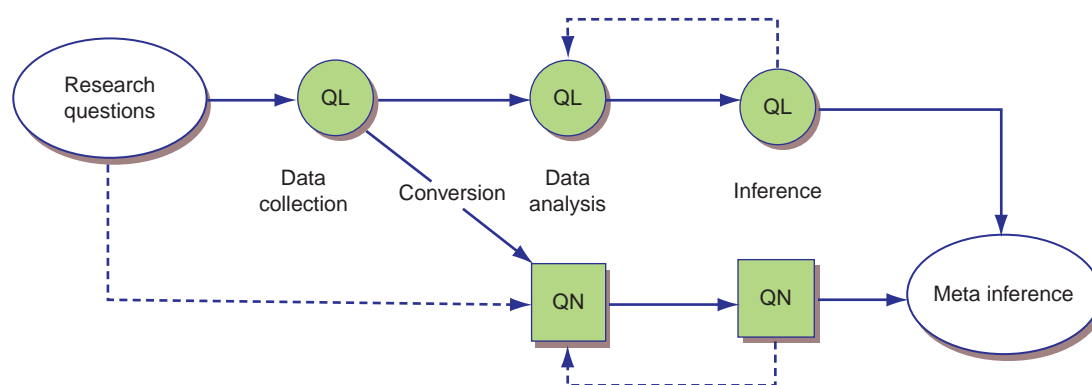


Figure 3 Graphic presentation of conversion (QL ↔ QN) mixed-methods designs with multiple questions – one type of data that are also transformed and reanalyzed. Final integrated meta-inferences are made on the basis of both sets of findings. Please note that this family of design also includes conversion of quantitative data to qualitative (QN ↔ QL).

independently (sequential mixed-methods designs, multisample), one might be the same or a subset of another (sequential mixed-methods designs, same/subsample), or one sample might be at a different level of the same organization (sequential mixed-methods design, multi-level). Unlike parallel studies, in sequential designs the selection of the second sample is dependent on the results obtained from the first. In other words, the second sample cannot be selected until the results of the first are analyzed and evaluated.

2. Will I have an opportunity to convert one type of my data to another, and reanalyze it in order to get more credible answers? If the answer is affirmative, a conversion mixed-methods design is employed (see **Figure 3**). For example, you may collect qualitative data (open-ended interviews, classroom observations, etc.), do a thematic analysis of the narrative, and make conclusions. After this, you may convert the obtained themes to numerical indicators that may be analyzed statistically. In conversion mixed designs, the qualitative data are transformed (quantitized) into numerical indicators that may be analyzed separately using statistical techniques.

Alternatively, in these studies, quantitative indicators (e.g., test results) might be transformed (qualitized) into qualitative data (e.g., profiles) that are analyzed separately as they originated from a separate sample (see Teddlie and Tashakkori, 2009 for more details). Obviously, the findings gleaned from both sets of data analysis are compared, contrasted, and integrated for better understanding of the phenomenon under study.

Depending on the answers to these questions, the investigator might choose one of the three basic (simple) families of designs described above: parallel, sequential, or conversion. In practice, many mixed-methods studies use a combination of these three families – in an iterative and dynamic manner – as needed. These studies have been labeled iterative mixed studies, using a fully integrated mixed-methods design (Teddlie and Tashakkori, 2009), depicted in **Figure 4**. This is the ideal design in mixed-methods studies because it allows the researcher to dynamically utilize both qualitative and quantitative approaches/methods, and use the insights from one type of data/process to modify/enhance the other within and across stages of the study.

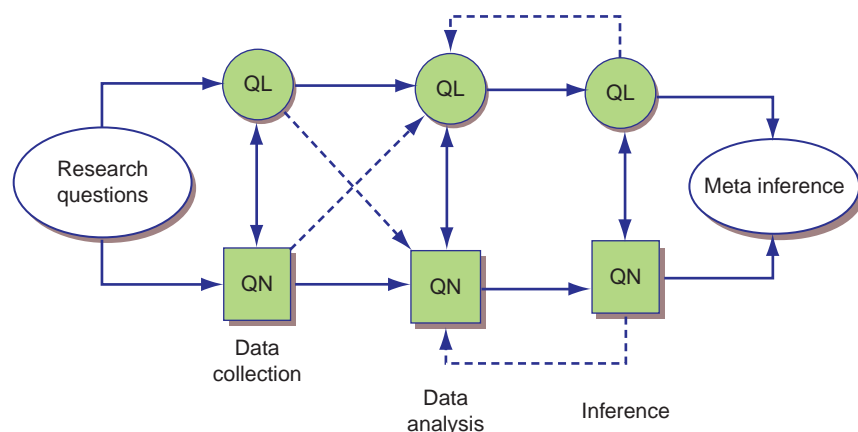


Figure 4 Graphic presentation of a fully integrated mixed-methods design with multiple questions, data, data analysis (including data conversion), and inferences within each strand. Elements/stages of the design are iterative both within and across the two strands (i.e., one is modified on the basis of the other). Final integrated meta-inferences are made on the basis of all the findings.

Data-Collection Strategies in Mixed-Methods Research

During the course of a mixed-methods study, an investigator collects a variety of narrative and numerical data, using a combination of qualitative (open-ended, naturalistic, and holistic) and quantitative (close-ended, structured, and variable-specific) data. It is possible to collect both qualitative and quantitative data in a single data-collection strategy. For example, in a mixed-methods study, observations may be recorded by taking extensive notes to record as many of the interactions as possible, while also using highly structured protocols yielding numerical data. Most observations include a combination of both unstructured (open-ended) recording and structured (closed-ended) items to record the behaviors/events. Similarly, interviews and questionnaires may be used to collect both types of data in a mixed-methods study. Open-ended (qualitative) interviews or questionnaires provide qualitative data. Structured (closed-ended) interviews and questionnaires provide quantitative data. Once again, most questionnaires and interviews in research do some of each.

In mixed-methods studies, interviewers conduct funnel-sequenced interviews (Teddlie and Tashakkori, 2009) that start from general questions/topics (“tell me about your school”) and are gradually focused on more specific emerging or preplanned issues. These interviews are conducted individually or in focus groups. Focus groups combine interviews, discussions, and observations of group interactions by the researcher(s), and, often, provide mixed data. Mixed questionnaires provide two types of data as well, although they are less dynamic (interactive) than interviews. Computerized questionnaires provide a greater potential for adapting each item/prompt on the basis of previous responses to close-ended questions.

In addition to using one method of data collection for obtaining both qualitative and quantitative data, mixed data might be obtained through multiple modes of data collection that each provide either qualitative or quantitative data (e.g., close-ended questionnaires with unstructured observations).

Data-Analysis Strategies in Mixed-Methods Research

Highly consistent with the type of design, in mixed methods the qualitative and quantitative data may be analyzed independently, or in an interdependent sequential manner. A detailed discussion of these analyses is beyond the scope of this article. Hence, we gloss over some of the main issues, and refer you to other sources (e.g., Creswell and Plano Clark, 2007; Teddlie and Tashakkori, 2009).

An example of data analysis in parallel mixed designs is when the investigator statistically compares an experimental and control group, while also conducting a thematic analysis of postexperimental interviews (or open observation notes, if the individuals/groups were observed). An example for data analysis in sequential designs might be the formation of groups of people/settings on the basis of qualitative data-analysis results, and then comparing these groups statistically (analysis of variance, discriminant analysis, etc.), using the available (or newly collected) quantitative data.

Integrated Inferences in Mixed-Methods Studies

As mentioned above, integration of the findings and conclusions are the primary objective of most methods studies. Although a degree of consistency between the findings

of qualitative and quantitative strands of a mixed-methods study is desirable, such consistency is not required. Integration in mixed methods denotes elaboration, complementarity, completeness, confirm/disconfirmation, and so forth. Inconsistency is examined carefully by mixed-methods researchers, to see if it might be a result of possible problems in data collection and analysis, or the conclusions derived from the results. If such an examination does not reveal problems in the design or the findings, lack of agreement might indicate that the two sets of findings/inferences are revealing two different aspects of the same phenomenon, one set might provide the conditions for the applicability of the other, or there is a possibility that there are two plausible, but different, answers to the question.

Mixed methodologists have discussed the importance of developing validity criteria and audits that incorporate and expand the qualitative and quantitative concepts. For example, Tashakkori and Teddlie (1998) (see the section titled 'Further Reading') have proposed using the term inference quality as a term to incorporate the term internal validity used by quantitative researchers and credibility used by qualitative researchers. Employing such an umbrella term gives mixed-methods researchers the opportunity to incorporate elements of quality assessment and strategies to deal with possible problems in making inferences (e.g., threats to quality) into a single integrative framework (Tashakkori and Teddlie, 2008). In such a framework, qualitative concepts – such as prolonged engagement, peer debriefing, triangulation, audit trail, negative case analysis, and others – can be used to enhance the quality of inferences gleaned from quantitative findings.

Similarly, concepts such as consistency between elements of the design, design quality, construct validity, investigator expectancy effects, and participant reactivity may be used as safeguards when making inferences on the basis of qualitative findings. In other words, using such an umbrella framework sensitizes the researcher to multiple perspectives when considering the quality/validity of their interpretations. A similar logic is applicable to evaluating and enhancing the quality of data that the findings and inferences are based on (i.e., data quality).

Integrating alternative (but not mutually exclusive) concepts from both qualitative and quantitative traditions, Tashakkori and Teddlie (2008) have offered a broad and twofold definition of inference quality: Design quality, and interpretive rigor. Design quality is the degree to which the interpretations and conclusions made on the basis of the results meet the professional standards of rigor, trustworthiness, and acceptability. Interpretive rigor is the degree to which alternative plausible explanations for the obtained results can be discounted or ruled out (please see Tashakkori and Teddlie (2008) and Teddlie and Tashakkori (2009) for details with regard to the components of these two quality criteria).

Transferability of Inferences in Mixed-Methods Studies

Inference transferability is a mixed-methods umbrella incorporating the qualitative concept of transferability and quantitative concept of external validity. It is broadly defined as the degree of applicability or generalizability of conclusions of a mixed-methods study to other individuals or entities (population transferability), other settings or situations (ecological transferability), other time periods (temporal transferability), or other methods of observation/measurement (operational transferability) (see Teddlie and Tashakkori, 2009 for details).

Conclusions

Mixed methodology puts each component of a research project (questions, data, data analysis, design, inferences, etc.) on a continuum between alternative qualitative and quantitative approaches/conceptualizations. This is consistent with the position of Popper (1959) (also see the section titled 'Further reading' below) – the eminent philosopher of science – who described how information grows scientifically through inductive and deductive approaches. Mixed-methods research is often based on the premise that the research purpose/question – not the paradigm – dictates the method of inquiry.

An advantage of mixed-methods is that it encourages researchers to think in terms of multiple continua rather than a single dichotomy of qualitative and quantitative approaches, employ multiple perspectives as needed, use multiple methods that are rooted in qualitative and quantitative traditions, and interpret the findings freely without the imposed constraints of either of the two broad approaches to research.

A challenge for the mixed-method researcher is that they must learn specific techniques that have been traditionally identified as being qualitative or quantitative. They must also learn how to integrate understandings gleaned from these complementary or competing methods in order to reach meta-inferences that better inform other professionals and stakeholders. There is an underlying assumption that the more perspectives one can get of a particular situation, the more complete the understanding inferences are likely to be. The value of mixed methods to researchers may not solely be in empowering them to use all possible tools for answering their research question, but also in becoming better informed as a result of integrating the findings of these methods.

Bibliography

- Bergman, M. M. (2008a). The straw man of qualitative–quantitative divide and their influence on mixed methods research. In Bergman, M. M. (ed.) *Advances in Mixed Methods Research: Theories and Applications*, pp 11–21. London: Sage.

- Bergman, M. M. (ed.) (2008b). *Advances in Mixed Methods Research: Theories and Applications*. London: Sage.
- Brewer, J. and Hunter, A. (2006). *Foundations of Multimethod Research: Synthesizing Styles*, 2nd edn. Thousand Oaks, CA: Sage.
- Creswell, J. W. and Plano Clark, V. L. (2007). *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage.
- Greene, J. C. (2007). *Mixing Methods in Social Inquiry*. San Francisco, CA: Jossey-Bass.
- Greene, J. C., Caracelli, V. J., and Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis* **11**, 255–274.
- Newman, I., Ridenour, C., Newman, C., and DeMarco, G. M. P., Jr. (2003). A typology of research purposes and its relationship to mixed methods research. In Tashakkori, A. and Teddlie, C. (eds.) *Handbook of Mixed Methods in Social and Behavioral Research*, pp 167–188. Thousand Oaks, CA: Sage.
- Popper, K. R. (1959). *The Logic of Scientific Discovery*. New York: Basic.
- Ridenour, C. S. and Newman, I. (2008). *Mixed Methods Research: Exploring the Interactive Continuum*. Carbondale, IL: Southern Illinois University Press.
- Tashakkori, A., Brown, L. M., and Borghasie, P. (2009). Integrated methods for studying a systemic conceptualization of stress and coping. In Collins, K. M. T., Onwuegbuzie, A. J., and Jiao, Q. G. (eds.) *Toward a Broader Understanding of Stress and Coping: Mixed Methods Approaches*, pp. 31–57. Greenway, CT: Information Age.
- Tashakkori, A. and Creswell, J. (2007). Exploring the nature of research questions in mixed methods research. *Journal of Mixed Methods Research* **1**(3), 207–211.
- Tashakkori, A. and Teddlie, C. (1998). *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. Thousand, CA: Sage.
- Tashakkori, A. and Teddlie, C. (2003). *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks, CA: Sage.
- Tashakkori, A. and Teddlie, Ch. (2008). Quality of inferences in mixed methods research. In Bergman, M. (ed.) *Advances in Mixed Methods Research: Theories and Applications*. London: Sage.
- Teddlie, C. and Tashakkori, A. (2009). *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Thousand Oaks, CA: Sage.
- Waszak, C. and Sines, M. (2003). Mixed methods in psychological research. In Tashakkori, A. and Teddlie, C. (eds.) *Handbook of Mixed Methods in Social and Behavioral Research*, pp 557–576. Thousand Oaks, CA: Sage.

Further Reading

- Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research* **6**(1), 97–113.
- Caracelli, V. J. and Greene, J. C. (1993). Data analysis strategies for mixed methods evaluation designs. *Educational Evaluation and Policy Analysis* **15**(2), 195–207.
- Creswell, J., Shope, R., Plano-Clark, V., and Green, D. (2006). How interpretive qualitative research extends mixed methods research. *Research in the Schools* **13**(1), 1–11.
- Erzberger, C. and Kelle, U. (2003). Making inferences in mixed methods: The rules of integration. In Tashakkori, A. and Teddlie, C. (eds.) *Handbook of Mixed Methods in Social and Behavioral Research*, pp 457–490. Thousand Oaks, CA: Sage.
- Howe, K. R. (1988). Against the quantitative–qualitative incompatibility thesis or dogmas die hard. *Educational Researcher* **17**, 10–16.
- Johnson, B. and Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher* **33**(7), 14–26.
- Maxwell, J. and Loomis, D. (2003). Mixed methods design: An alternative approach. In Tashakkori, A. and Teddlie, C. (eds.) *Handbook of Mixed Methods in Social and Behavioral Research*, pp 241–272. Thousand Oaks, CA: Sage.
- Morgan, D. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research* **1**(1), 48–76.
- Newman, I. and Benz, C. R. (1998). *Qualitative–Quantitative Research Methodology: Exploring the Interactive Continuum*. Carbondale, IL: University of Illinois Press.
- Niglas, K. (2004). *The Combined Use of Qualitative and Quantitative Methods in Educational Research*. Tallinn, Estonia: Tallinn Pedagogical University Dissertation on Social Sciences.
- Popper, K. R. (1962). *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Basic.
- Teddlie, C., Tashakkori, A., and Johnson, B. (2008). Emergent techniques in the gathering and analysis of mixed methods data. In Leavy and Hesse-Biber (eds.) *Handbook of Emergent Methods*, pp 389–414. Guilford Press.

QUANTITATIVE METHODS AND RESEARCH DESIGN

External Validity

G E Matt, A Brewer and M Sklar, San Diego State University, San Diego, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Analog models – Studying a phenomenon in a controlled setting (e.g., scale model of a bridge, animal model of Alzheimer's disease) that mirrors what one would observe in the field (e.g., actual 200 m suspension bridge) or target system (e.g., elderly humans).

Domains of generalizability – Universes of units (*U*), treatments (*T*), outcomes (*O*), or settings (*S*) with regard to which inferences are sought.

Group theory – A broadly applicable and highly abstract mathematical theory designed to describe generalizations of all kinds.

Meta-analysis – Combining findings from many different, yet related, studies to foster empirical knowledge with regard to causal associations that are more trustworthy than those possible from any single study.

Principle of causal explanation – Seeks to strengthen the validity of generalizable knowledge claims by investigating the causal mediating mechanisms that underlie a relationship of interest. Causal explanation is valuable because it explains how and why an effect occurs, and, consequently, provides information about when and where the relationship can be replicated.

Principle of empirical interpolation and extrapolation – Seeks to strengthen the validity of generalizable knowledge claims concerning the unsampled range of values on a particular variable.

Principle of heterogeneous irrelevancies – Seeks to strengthen the validity of generalizable knowledge claims by demonstrating the robustness – or invariance – of a causal association across substantively irrelevant conditions.

Principle of proximal similarity – Seeks to strengthen the validity of generalizable knowledge claims by demonstrating the correspondence between samples and the target universe to foster

generalizability of inferences from sample to universe.

Purposive sampling – Selecting instances from a target universe to make the correspondence between sample and universe an explicit and rational exercise rather than the result of a random process.

Principle of discriminant validity – Seeks to strengthen the validity of generalizable knowledge claims by demonstrating that a finding holds for a target construct and not for a cognate. This involves examining the conditions under which an association changes in magnitude and direction.

Technology-transfer models – Relying on a progression of studies to bridge the gap between the controlled, artificial laboratory conditions and field applications.

Introduction

The term external validity was first introduced more than 50 years ago in a seminal paper by Campbell (1957) titled 'Factors relevant to the validity of experiments in social settings.' For Campbell, internal validity and external validity were the two major criteria for evaluating the validity of research designs examining causal propositions. In his definition, internal validity asked whether an experimental stimulus (e.g., treatment) made some significant difference in a specific instance. In contrast, external validity asked questions about representativeness or generalizability: To what populations, settings, and variables can an effect be generalized?

The distinction between internal and external validity evolved in subsequent writings by Campbell and Stanley (1963), Cook and Campbell (1979), Campbell (1986), and Shadish *et al.* (2002) into a four-partite model of statistical conclusion, internal, construct (treatment and effect), and external validity. The latter two are closely related, in that both involve inferences with regard to more abstract

constructs or universes based on the manifest instances present in a particular study. The following sections briefly define key concepts, review traditional approaches for justifying generalized inferences, introduce Cook's pragmatic principles of generalized causal inferences, and highlight the importance of meta-analysis in justifying the external validity of causal claims. The article concludes by outlining future directions for theory and practice of causal generalization.

What We Generalize About

In Campbell's original formulation, external validity concerned the generalizability of the causal association between a treatment (i.e., experimental stimulus) and an outcome (i.e., response). Following Cook and Campbell (1979) and Cronbach (1982), the external validity of this causal association can be examined with respect to the domains of experimental units (U), treatments (T), outcomes (O), and settings (S).

Carrying forward the notation introduced by Cronbach (1982), U refers to the universe of units with regard to which inferences are sought. Units are the students, classrooms, communities, districts, etc., on which observations are made, that are subjected to different treatments, and for which causal mechanisms are specified to explain the effect of an intervention. In contrast, u represents the specific sample of units in a study.

T represents the universe of treatments about which inferences are sought. The treatments may be interventions, programs, experimentally manipulated conditions, or naturally occurring events. In contrast, t refers to a specific treatment or group of treatments, as realized in a study.

O refers to the universe of outcomes with regard to which inferences are sought. This universe consists of all admissible measures used for collecting data on a particular effect construct. The specific measure or group of measures used in a study is denoted with o .

S refers to the universe of settings in which a study is implemented. Settings include the social, cultural, organizational, and historical context in which a study is conducted. While S refers to the entire universe of admissible settings, s identifies the specific setting of a study.

The Three Main Questions of External Validity

Three major types of external validity questions have been distinguished in the literature. The first – and perhaps most common – is Campbell's original question: To what populations, settings, and variables can an effect be generalized? This question involves inductive inferences from the particular sample of persons, treatment

implementation, outcome measures, and settings (i.e., $utos$) to the universes of interest (i.e., $UTOS$), of which the particular instances included in a study are a part.

The second external validity question asks: Across which universes (or subuniverses) of units, treatments, outcomes, and settings can an effect be generalized? This question concerns inferences about the robustness of an effect and the conditions that moderate its strength or direction. Potential moderators may be found in the characteristics of persons, treatments, outcomes, and settings.

The third question concerns inferences about novel universes of units, treatments, outcomes, and settings that have not been studied yet: To what yet unstudied universes of units, treatments, outcomes, and settings can an effect be generalized? This question asks to make extrapolations or projections from domains in which an effect has been studied to novel domains.

Traditional Approaches to Justifying Generalized Inferences and Their Challenges

Causal Explanation

In the tradition of experimental science, complete knowledge of the mediating and moderating principles provides the strongest justification for generalizing causal relationships to target universes, across subuniverses, and to novel universes (external validity questions 1, 2, and 3). For instance, a complete understanding of the mediating and moderating principles underlying the effects of homework assignments on learning outcomes among seventh graders would allow teachers to identify the optimal homework for a particular group of students to achieve a particular learning objective – given the students' personal characteristics, and family and school support networks. Complete understanding requires substantive theories detailing the explanatory principles, well-specified constructs, valid measurements of these constructs, and the absence of an equal or better-fitting model (Cook, 1993). Unfortunately, complete understanding of a causal association and its mediating and moderating principles is hardly attainable in the social and behavioral sciences, and therefore, cannot provide a solid foundation to justify the generalization of causal inferences.

Statistical Sampling Theory

In the tradition of observational research, generalizations to target universes (external validity question 1) are best justified through the correspondence between samples and the universes they represent. When universes are clearly designated and their elements can be sampled

with known probability, statistical sampling theory provides the strongest justification for this correspondence. Although designating human populations (e.g., students, teachers, classrooms, and schools) to sample from is relatively straightforward (e.g., rosters of students, directory of schools), it is rarely practiced in experimental and quasi-experimental research, in which the primary emphasis is on internal rather than external validity. Further, sampling theory may be outright impossible to apply when selecting instances from the universes of treatment (e.g., homework assignments), outcomes (e.g., learning outcomes), and settings (e.g., sociocultural context). These universes are, often, defined as theoretical constructs with fuzzy – rather than crisp – boundaries (Matt, 2005), generally have multiple and partially conflicting definitions, and lack commonly accepted sampling frames. As valuable as statistical sampling theory can be when it is applied to designated universes and valid sampling frames, it does not provide a panacea for all generalizability domains. It also provides no support for generalizations across subuniverses and extrapolations (external validity questions 2 and 3).

Purposive Sampling

Purposive sampling is a common approach in efforts to validate treatment and outcome constructs. The goal of purposive sampling is to select instances from a target universe that make the correspondence between sample and universe an explicit and rational exercise rather than the result of a random process. The exercise begins with a description of a target construct – including its prototypical and more marginal components, features shared with cognate constructs, and boundaries shared with related constructs. Then, specific instances – which belong to the target construct and demonstrably share most of its prototypical components – are selected. Variants are then purposively selected to rule out the confounding of substantive interpretations and irrelevancies and to better understand potential features that may moderate the association between treatment and effect constructs. Purposive – as opposed to random – sampling ensures construct coverage (Cronbach *et al.*, 1972), and allows researchers to specify and test hypotheses about the nature of treatment and effect constructs. Thus, purposive sampling can be used to justify generalizations from specific instances of a sample to a target universe (external validity question 1) and is widely practiced in construct validation studies of treatments (*T*) and outcome measures (*O*).

Analog Models

Analog models – common in education (e.g., mock classrooms in preservice teacher education programs), engineering (e.g., scale models), biotechnology (e.g., genetic

knock-out mice), medicine (e.g., first-aid practice mannequin, animal studies), and some areas of behavioral science (e.g., computer simulations) – use purposive sampling to justify inferences about a target model drawn from data collected on an analog model. Similar to purposive sampling methods described above, analog systems do not need to provide a probabilistic representation of the target; they must only share relevant properties of the target system. That is, a successful analog system may look very different, but must behave just like a target system.

Analog models make it possible to study a phenomenon in a controlled laboratory setting (e.g., scale model of a bridge, animal model of Alzheimer's disease) that mirrors what one would observe in the field (e.g., actual 600-ft suspension bridge) or target system (e.g., elderly humans). The critical task is to create a valid analog model that corresponds to the actual generalizability target in all relevant features such that findings based on the analog model by extension also apply to the generalizability target. Analog models provide a promising direction to justify inferences about novel universes (external validity question 3).

Technology Transfer Models

Technology transfer models rely on a progression of studies to bridge the gap between the controlled, artificial laboratory conditions and field applications. They are closely related to the idea of analog models and purposive sampling in that earlier studies serve as analogs to justify inferences with regard to yet-unstudied organisms or domains.

A good example of the implementation of the technology transfer model is the approval process that guides the development of new pharmaceutical products in the USA (US FDA (US Food and Drug Administration (FDA)), 1998). Experimental new drugs are first tested in preclinical studies for safety and efficacy in controlled laboratory conditions. These studies often involve cell cultures, computer models, and animal analogs. This is followed by Phase I clinical studies in humans, short-term studies on small samples, focusing on safety, and, often, involve healthy subjects. If these initial studies demonstrate a drug's safety, Phase II studies follow – in which short-term efficacy and drug safety are investigated in larger samples. If Phase II studies continue to demonstrate a drug's safety and efficacy, large-scale Phase III clinical trials are conducted with a focus on drug dosage, long-term effectiveness, and drug safety. Data collected in the preclinical and clinical trials are again reviewed by FDA expert panels to approve, request additional studies, or deny approval of a new drug. Following FDA approval, monitoring systems are put in place to detect adverse reactions.

The technology transfer model is also evinced in the field of education. University-based educational research

centers bring together researchers interested in studying issues surrounding the teaching and learning of mathematics, science, languages, and other subjects. Studies are conducted in controlled laboratory conditions to investigate basic cognitive processes involved in developing new mathematical concepts (e.g., rate of change) or in acquiring a second language through dual immersion. Research findings are then used to inform the development of materials for the professional development of educators, as well as educational curriculum that is ultimately implemented in actual classrooms.

Through a programmatic series of research studies, the technology-transfer model can support inferences with regard to target universes (e.g., target populations, target effect, side effects; i.e., external validity question 1), robustness and moderating conditions (e.g., drug indications and contraindications; external validity question 2), and novel universes (e.g., off-label use; i.e., external validity question 3).

Meta-Analysis

Meta-analysis combines findings from many different – yet related – studies to foster empirical knowledge about causal associations that are more trustworthy than those possible from any single study. This benefit arises for two main reasons. First, combining findings from parallel studies promises to increase statistical power and precision for estimating the magnitude of a causal association. More importantly, however, is the potential of meta-analysis to strengthen external validity by identifying the realm of application of a causal association – that is, meta-analyses are most useful when they allow us to examine whether a causal association (1) holds with specific populations of persons, settings, times, and ways of varying the cause or measuring the effect; (2) holds across different populations of people, settings, times, and ways of operationalizing a cause and effect; and (3) can even be extrapolated to other populations of people, settings, times, causes, and effects than those that have been studied to date – that is, meta-analyses offer opportunities to probe external validity questions 1, 2, and 3.

In 1978, Glass and Smith published a noteworthy meta-analysis examining the relationship between class-size and academic achievement – an issue that was previously surrounded by inconclusive research findings. This meta-analysis synthesized the results of 77 separate studies that included 725 comparisons of academic achievement in smaller versus larger classes. It received considerable attention because it was the first meta-analysis to find clear evidence that reduced student-to-instructor ratios significantly improved academic achievement (Cooper, 1989). Examples from this meta-analysis will be presented to illustrate how Cook's five principles justify generalization in the absence of formal sampling.

Five Pragmatic Principles for Justifying Generalized Causal Inferences

Cook (1991) put forth a pragmatic alternative approach to statistical sampling theory for justifying the validity of generalized causal inferences in the absence of formal probability sampling. The proposed approach relies on five principles, adopted from research on construct validation (e.g., purposive sampling), and existing analog and transfer models of generalization. It can be used to justify generalization to target constructs (e.g., treatment, outcomes), as well as generalization to target populations of units and setting. The approach emphasizes purposive sampling for theoretical ends rather than random sampling to represent a population.

The Principle of Proximal Similarity

The principle of proximal similarity describes the correspondence between samples and the target universe. Proximal similarity requires an explication of which components are – and are not – central to the targets of the generalization (*UTOS*). Because matching cannot be feasibly achieved on all features, it is most critical to demonstrate similarity on prototypical components (Cook, 1993). In research syntheses, proximal similarity is achieved by ensuring that the relevant properties of the target universe are adequately represented by the samples included in the meta-analysis (Matt, 2003). Because Glass and Smith (1978) were interested in generalizing inferences about the effect of reduced student-to-instructor ratios (*T*) on achievement (*O*) among school-aged children (*U*) in elementary and secondary-education classrooms (*S*), they ensured that the prototypical components inherent to each universe were present in the sampled studies.

The Principle of Heterogeneous Irrelevancies

The principle of heterogeneous irrelevancies concerns the robustness, or invariance, of a causal association across substantively irrelevant conditions. Robust empirical findings suggest broad main effects of interventions – the type of effects that are particularly useful for policy decisions affecting large and diverse constituencies (Abelson, 1995). The greater the range of substantive irrelevancies across which a causal association has been found to be robust, the more confident one can be that the causal association will hold under yet-unstudied conditions.

The principle of heterogeneous irrelevancies highlights the importance of probing causal associations under many different conditions, especially those that are deemed irrelevant to the phenomena under investigation. Doing so protects against the possibility a causal relationship is confounded by a factor common to all the studies in which it was examined (Matt, 2003). For example, if all studies

examined by Glass and Smith (1978) administered the same standardized achievement test, then the relationship between class-size and achievement could be confounded by outcome-method invariance. Consistent with the principle of heterogeneous irrelevancies, Glass and Smith (1978) ensured variance in assessment method and subject area, pupil demographics and intelligence quotient (IQ), geographic location, and other irrelevancies across studies. Further, the studies date back to 1900 – indicating that the effect of class-size on achievement was neither confounded by time, nor associated with an event that occurred at a particular point in time. The range of irrelevancies encompassed by the studies in this meta-analysis parallels the heterogeneity of the *UTOS* to which authors intended to generalize the causal inference.

The Principle of Discriminant Validity

In addition to demonstrating robustness of a causal association across substantively irrelevant conditions, external validity is strengthened when we can show that a finding holds for a target construct and not for a cognate. Demonstrating discriminant validity requires exploring the boundaries of target constructs and understanding the conditions moderating an association. The principle of discriminant validity requires identifying the substantively relevant conditions under which an association holds, fails to hold, and changes in magnitude or direction. To illustrate, the relationship between class-size and achievement held across the range of irrelevancies listed above; however, it was slightly weaker in elementary, as compared with secondary classrooms (Glass and Smith, 1978). By examining the relationship across a greater range of *utos* than any single study is capable of, Glass and Smith (1978) pinpointed a moderating factor that might have otherwise gone unrecognized.

The Principle of Empirical Interpolation and Extrapolation

The principle of empirical interpolation and extrapolation addresses the issue of generalizing to an unsampled range of values on a particular variable (Cook, 1993). For example, if the effect was observed in second and fifth graders, empirically bound generalization based on interpolation would infer that the effect would also be observed at unstudied levels in-between (i.e., third and fourth graders). Interpolation relies on the assumption that the relationship between cause and effect is known (e.g., linear) between the studied levels of a variable. In contrast, extrapolation involves generalizing beyond the range of sampled values in either direction (e.g., generalizing findings from second and third graders to first and/or fourth graders). According to Cook (1993), there is greater confidence in generalization based on interpolation than on extrapolation, and there is

stronger justification for either when (1) a greater range of values is studied and (2) the effect is consistent across this range. Further, shorter extrapolations are less problematic than larger inferential leaps. An advantage of meta-analysis is that it encompasses a wider range of person and/or treatment variable values than is typically observed in any single research study (Cook, 1993). Therefore, interpolation is much more common in meta-analysis than is the riskier practice of extrapolation.

The Principle of Causal Explanation

The principle of causal explanation requires investigation into the causal mediating mechanisms that underlie a relationship of interest. Causal explanation is valuable because it explains how and why an effect occurs, and consequently, provides information regarding when and where the relationship can be replicated. Complete understanding of an entire causal mediating process is the holy grail of experimental scientific research because it provides the knowledge necessary to identify the conditions under which a causal association will and will not occur. A complete blueprint of a causal association may be an unrealistic goal; however, even partial knowledge of the causal process can strengthen generalized causal inference. For example, if the causally efficacious component of reduced class-size is one-on-one time with the teacher, then a similar increase in achievement might be attained by scheduling more individualized instruction time in classrooms with fixed student-to-instructor ratios. Similarly, if one-on-one time improves achievement by enhancing self-esteem (i.e., a mediator), then other interventions targeting self-esteem may indirectly affect achievement. Unfortunately, the major limitations of using meta-analysis to identify causally relevant treatment components and mediating factors are that (1) primary studies, often, do not provide sufficient details about the treatment and (2) studies rarely investigate the same mediating mechanism (Cook, 1993). For example, in regard to the latter, in order for Glass and Smith (1978) to infer that the effect of class-size on achievement was mediated by self-esteem, a number of primary studies would have had to similarly investigate this connection.

Outlook

Statistical sampling theory provides a powerful theoretical framework for generalizing from samples to corresponding populations and is most relevant when generalizing to populations of units and settings (external validity question 1) that can be enumerated and are under the control of the researchers. In research practice, however, random sampling from designated universes of students, classrooms, schools, etc. is uncommon in most studies that involve

causal associations and random assignment of experimental units to different treatment conditions. The relevance of sampling theory to generalizations about the treatments and outcomes and to other questions of external validity is less clear. This situation leaves researchers in a difficult position: If sampling theory is not often used or is not able to answer important generalizability questions, how should we justify the external validity of causal relationships? What is needed is a viable new theory of generalization that can guide the design of primary studies and the design of research syntheses.

The principles introduced by Cook (1991, 1993) and further elaborated by Shadish *et al.* (2002) can serve as a starting point. They borrow strategies and techniques commonly used to justify construct-validity claims and have been applied in programmatic approaches to support the transfer of research and technology from labs to applications (US FDA, 1998). They provide practical guidelines for exploring and building arguments around generalizability claims and can be applied very well to research syntheses (Matt and Cook, 2009).

Meta-analysis provides the most promising research design to explore generalizability claims using Cook's five principles as no single study provides as many opportunities as research synthesis to examine the robustness and moderating conditions of causal associations. Matt and Cook (2009) examine, in detail, the validity of generalized causal inferences based on research syntheses.

Another promising direction for a comprehensive theory of generalization comes from a branch of mathematics called group theory. In fact, group theory is a broadly applicable and highly abstract mathematical theory designed to describe generalizations of all kinds (Livio, 2006). A mathematical group is defined as any set whose members obey certain rules with respect to particular operations, and therefore, act similarly in response to such operations. The members of a group may include numbers, words in the English language, a pair of jeans, or a causal association. Operations can be as diverse as 'rotate 120°,' 'do nothing,' 'turn inside out,' and 'followed by,' 'add,' or 'multiply.' A group exists if the operations are applied to the members and the following properties hold: (1) closure, (2) associativity, (3) identity element, and (4) inverse.

In contrast to sampling theory, group theory does not characterize a generalization based on the objects it involves (e.g., a set of exemplars defining a universe settings), but by whether an operation changes it or leaves it unchanged. Similar to Cook's pragmatic principles, group theory relies on studying the invariance to transformation (i.e., robustness) and the identification of moderators (i.e., discriminant validity). Translating the tenets of group theory to the generalizability of causal association could provide important insights into the design of primary studies and the analysis of meta-analyses that facilitate generalizable proposition. Based on Cook's five principles and the

cursorry introduction of group theory, drawing generalizable inferences will benefit from efforts to examine causal associations under many diverse conditions, both substantively irrelevant as well as substantive relevant characteristics of units, treatments, outcomes, and settings. Research synthesis and programs of research bear most promise for providing the empirical warrants for such generalized inferences.

See also: Causal Inference; Class Size – Arguments and Evidence; Classical Test Theory Reliability; Design of Experiments; Educational Measurement: Overview; Exploratory Data Analysis; Field Experimentation in Education; Generalizability Theory; Internal Validity; Meta Analysis; Multitrait–Multimethod Designs; Sampling; Sampling; Statistical Conclusion Validity; Statistical Power Analysis; The Economics of Class Size; Validity.

Bibliography

- Abelson, R. P. (1995). *Statistics as Principled Arguments*. Hillsdale, NJ: Erlbaum.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin* **54**, 297–312.
- Campbell, D. T. (1986). Relabeling internal and external validity for applied social scientist. In Trochim, W. M. K. (ed.) *Advances in Quasi-Experimental Design and Analysis*, pp 67–77. San Francisco, CA: Jossey-Bass.
- Campbell, D. T. and Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- Cook, T. D. (1991). Meta-analysis: Its potential for causal description and causal explanation with program evaluation. In Albrecht, G., Otto, H. U., Karstedt-Henke, S., and Bollert, K. (eds.) *Social Prevention and the Social Sciences: Theoretical Controversies, Research Problems, and Evaluation Strategies*, pp 245–285. New York: Walter de Gruyter.
- Cook, T. D. (1993). A quasi-sampling theory of the generalization of causal relationships. *New Directions for Program Evaluation* **57**, 39–82.
- Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago, IL: Rand McNally.
- Cooper, H. M. (1989). Does reducing student-to-instructor ratios affect achievement? *Educational Psychologist* **24**(1), 79–98.
- Cronbach, L. J. (1982). *Designing Evaluations of Educational and Social Programs*. San Francisco, CA: Jossey-Bass.
- Cronbach, L. J., Gleser, G. C., Nanda, H., and Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: Wiley.
- Glass, G. V. and Smith, M. L. (1978). *Meta-Analysis of the Relationship of Class Size and Student Achievement*. San Francisco, CA: Far West Laboratory for Educational Research.
- Livio, M. (2006). *The Equation that Couldn't Be Solved*. New York: Simon and Schuster.
- Matt, G. E. (2003). Will it work in Münster? Meta-analysis and the empirical generalization of causal relationships. In Holling, H., Böhning, V., and Schulze, R. (eds.) *Meta-Analysis*, pp 113–139. Berlin: Springer.
- Matt, G. E. (2005). Uncertainty in the multivariate world: A fuzzy look through Brunswik's lens. In Beauducel, A., Biehl, B., Bosniak, M., *et al.* (eds.) *Multivariate Research Strategies. A Festschrift in Honor of Werner W. Wittmann*, pp 123–152. Maastricht, NL: Shaker Publishing.
- Matt, G. E. and Cook, T. D. (2009). Threats to the validity of generalized inferences from research syntheses. In Harris Cooper, L. V. H. and Valentine, J. C. (eds.) *The Handbook of Research Synthesis and Meta-Analysis*, 2nd edn., pp 503–520. New York: Russell Sage Foundation.

- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Design for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.
- US FDA (US Food and Drug Administration) (1998). *The CDER Handbook*. Washington, DC: U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research.

Further Reading

- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher* 5(10), 3–8.

This page intentionally left blank

R

ROLE OF NGOs IN GLOBALIZATION

NGOs and Globalization of Education

M Ahmed, Brac University, Dhaka, Bangladesh

© 2010 Elsevier Ltd. All rights reserved.

Globalization and Education

Globalization in its generic and literal sense is the process of transformation of local or regional phenomena into global ones. It refers to the idea of the people of the world being connected and unified in multiple ways. It is not a new phenomenon, but the issue has come to the fore in recent decades with economic aspects of globalization receiving the most attention. In the economic context, it refers to the reduction and removal of barriers between national borders in order to facilitate the flow of goods, capital, services, and labor. Another dimension of globalization is cultural integration, again an ancient phenomenon, intensified by recent revolutionary changes in communication and transportation technologies.

Cultural globalization, driven in its recent manifestations by digital communication technology, has led to both a process of homogenization and an enhanced potential for promoting and appreciating cultural diversity. Effective and aggressive marketing of American and Western cultural industries has raised the specter of the global domination of Western culture at the expense of national and indigenous cultural diversity. However, the spread of communication technologies has also supported the emergence of movements in defense of local uniqueness, individuality, and identity (Croucher, 2004).

A critical aspect of cultural globalization, or the influence of the globalizing forces beyond the economic sphere, is the development of a global perspective and global interaction on education. This includes influence of globalization on educational content, objectives, and delivery of services.

A Global Perspective on Education

In the post-World War II era, under the auspices of the United Nations, a different form of globalization, in contrast to economic globalization driven by the free market principles, has developed. This other form of globalization is based on the recognition of the need for an international order to guide action on common interests of humanity. Principal tenets of this order include peaceful resolution of conflicts; common responsibility to promote and protect human rights and human dignity, justice for all, and development of the human potential everywhere; protecting and nurturing the planetary resources for the benefit of the present and future generations; and inculcating a spirit of human solidarity and responsibility.

The establishment of the United Nations itself is a manifestation of the phenomenon of globalization arising from interdependence and connectedness of peoples and nations. Globalization guided by the principle of human solidarity, since the creation of UN, has found a concrete expression in the regime of international treaties and conventions which have emerged to establish and protect human rights and promote human development. International and national NGOs and civil society organizations have played a prominent role in shaping the content of these international agreements and lobbying for the adoption and ratification of these internationally and in countries.

The Universal Declaration of Human Rights adopted in 1948 was the first global statement about the inherent dignity and equality of human beings. In adopting the declaration, the governments representing the international community accepted a set of principles and obligations

covering a wide range of human endeavor, which applied to all nations universally.

The principles directly related to education are set out in the declaration's Article 26, which affirms that: (1) everyone has the right to education, (2) education shall be directed toward the full development of the human personality and to the strengthening of respect for human rights and fundamental freedoms, and (3) parents have a prior right to choose the kind of education that shall be given to their children.

The universal declaration served as the touchstone for subsequent international agreements in the form of declaration, treaties, and conventions addressing common concerns and interests of humanity, which called for collective and global action. For example, the universal declaration directly led to the negotiation and adoption of the two binding treaties in 1966, recognized as the twin pillars of global human rights standards and provisions – the International Covenant of Civil and Political Rights and the International Covenant of Economic, Social and Cultural Rights. These agreements impose a moral obligation, and many place a legal obligation, on the signatory countries to abide by these agreements. They constitute a body of legal and institutional structure of common responsibility for common good of humanity (United Nations, 2002).

Major international treaties include a provision for states to report to a treaty body on progress made and steps taken in implementing the agreements by national governments who have become signatories to these treaties. There are, for example, the Human Rights Committee and the Committee on Economic, Social and Cultural Rights, which serve as treaty bodies for the relevant agreements. Similar bodies exist for reviewing progress in countries on implementation of the Convention to End all forms of Discrimination against Women (CEDAW) and the Convention on the Rights of the Child (CRC). NGOs have often served as an effective watchdog in respect of efforts of countries to implement the agreements and adequate and objective reporting by countries to the treaty bodies. Often, NGOs and other civil society bodies have prepared alternative reports to draw public attention, from their perspective, to lack of good-faith efforts to live by the treaty obligations.

The creation of the United Nations Educational, Scientific and Cultural Organization (UNESCO) itself represents a commitment of the international community “to contribute to peace and security by promoting collaboration among the nations through education, science and culture in order to further universal respect for justice, for the rule of law and for the human rights and fundamental freedoms which are affirmed for the peoples of the world, without distinction of race, sex, language or religion, by the Charter of the United Nations” (Article 1, Constitution of UNESCO).

In the case of education, several international treaties and agreements have been adopted since 1945 by international conferences convened either by the United Nations or by UNESCO. Probably the best-known examples in the field of education are the World Declaration on Education for All, adopted by the World Conference on Education for All (EFA), held in Jomtien, Thailand, in 1990 and, as its follow-up, the World Education Forum in Dakar in 2000. A framework for EFA with specific global goals for 2015 was adopted in Dakar. The framework has served as both the guide for formulation of national plans of action and the focus of advocacy for national and international NGOs. (See below the description of advocacy activities of the Global Campaign for EFA.)

The United Nations General Assembly has from time to time adopted resolutions proclaiming an international day, year, or decade relating to education, the purpose being to focus world opinion on a particular aspect of education considered to be deserving of attention and support. Examples include International Literacy Day (September 8), World Teachers Day (October 5), International Literacy Year (1990), and United Nations Decade for Human Rights Education (1995–2004). These proclamations and events and activities related to these, on the one hand, reflect a broad international consensus on educational issues; on the other hand, these serve as the basis for advocacy and mobilization activities on educational issues by national and international NGOs and other civil society bodies.

Development in education over the years since the beginning of the process of establishing global norms and goals, starting with the proclamation of the Universal Declaration of Human Rights, has been dramatic. In the late 1940s, a little more than 50% of the world's adults were literate at a rudimentary level and less than 50% of the children had access to formal education. Today, most children go to school and the majority has access to beyond primary education. Four out of five of the adults have acquired literacy skills at least at the basic level, although women still lag significantly behind (UNESCO, 2000).

As noted, important contribution of the globalization forces in the progress achieved in education has been in articulating the norms and goals that all shared and in assessing and comparing the country situations in global league tables. NGOs and the civil society have played an important role in advocacy, awareness raising, and capacity building, and, in many instances, extending educational opportunities in creative ways to those left out from the mainstream educational services. What do all these add up to? It is reasonable to argue that the international consensus and expression of commitments, cooperation and sharing of lessons and experience, the global recognition of the right to education and the obligation to fulfill this right, and the work of NGOs in driving the messages home, have considerably strengthened the momentum for the secular trends of educational expansion that existed.

Changing Educational Paradigm in the Context of Globalization

The demands of globalization in terms of competitive market, skills and capabilities that young people must acquire, dramatic changes in knowledge dissemination possibilities, and interdependence of people and nations have altered the conventional centuries-old thinking and models about learning content, objectives, and delivery of educational services. At the same time, it has become more critical than ever that learners are aware of and that they interact effectively with the world in which they live and appreciate and claim the common heritage of human thought, action, and, creativity.

The need for change in the educational paradigm has been argued cogently in the report of the UNESCO-appointed International Commission on Education for the Twenty-first century led by Jacques Delors. The Commission identified four key purposes as the pillars of learning on which the educational systems must be built. These are:

1. learning to know,
2. learning to do,
3. learning to live together, and
4. learning to be.

These pillars refer to the imperatives of acquiring the skills of learning (to learn rather than to accumulate facts), being purposefully productive, being responsible and effective members of the human community locally and in the world, and, at the same time, being confident of one's identity (UNESCO, 1998).

The point is often made that advances in information technology and telecommunications have created unprecedented opportunities for learning for every one; that the barriers of time, distance, language, pedagogic skills, and resources can be less critical. However, this potential can be realized and transformation of learning can happen, only if the new paradigm of knowledge, education, and learning can incorporate components that are neglected in the current models. The Global Information Infrastructure Commission suggests the following key elements as essential components of the new learning paradigm (Cogburn, 1998):

- focusing on concepts and understanding, rather than accumulation of factual information;
- using a holistic and iterative approach in pedagogy, rather than a linear development of knowledge and skills;
- emphasizing on enhancing the learners' ability to manipulate symbols;
- enhancing the student's skills to acquire and utilize knowledge;
- producing an increased quantity of scientifically and technically trained persons;

- blurring the distinction between mental and physical labor;
- encouraging learners to work in teams;
- overcoming the barriers of time and space in learning; and
- developing an agile and flexible system of education.

Recognizing the global implications of local action and vice versa, educational content has been developed, frequently in higher education, and at secondary and primary levels as well, under the label of global education or international education. Global education places particular emphasis on the changes in communication and relationships among people across nations and regions. It sheds light on how major societal phenomena, such as human conflict, economic systems, human rights and social justice, human commonality and diversity, the common legacy of literature and culture, and technological revolution affect people and communities. Without denying the value of traditional branches of specialist knowledge, global education seeks to blur the boundaries between disciplines. It encourages interdisciplinary and multidisciplinary studies, which can help offer solutions of complex human problems.

The term international education, sometimes used as the substitute for global education, may mean different things, but two meanings are generally attached to the concept. The first refers to education that transcends national borders through the exchange of people, as in study abroad. The other meaning reflects a comprehensive approach to education that attempts to prepare students to be effective participants in an interconnected world.

The premise underlying the approach of global education and international education is that the understanding of society, nations, and the human condition is deepened through examination of cultures, languages, environmental situations, governments, political relations, religions, geography, and history of the world. Professionals and students particularly wishing to be engaged in international educational development are mostly the participants in international education academic programs (Cummings and McGinn, 1997).

Commodification of Education

One effect of economic globalization is that education may be treated as a marketable commodity. The traditional view of education as a public good financed largely by public funding is being challenged by the criterion of efficiency of the market. The neoliberal view – the belief that society's needs can be best met and problems can be best solved by the market with minimal government regulation and public sector involvement – has spread from

the economic sphere to what have been generally regarded as public services and the domain of the public sector.

At least three elements of the neoliberal market approach may be found in the design for educational provisions:

1. attempting to make the provision of education cost efficient by designing them as marketable products;
2. standardizing the educational services and experiences as well as assessing their performance by objective and standardized testing of results; and
3. focusing on marketability of outputs.

The three elements are reflected in policies, such as cut-backs in the public sector, closing inefficient programs that do not directly meet market needs, and the use of technology, including distance learning to bring about greater efficiency, comprising courses and degrees that are packaged for delivery over the Internet by for-profit corporations (Tabb, 2007).

The intergovernmental body World Trade Organization (WTO), established in 1994, is dedicated to the goal of removing trade barriers among nations. The General Agreement on Trade in Services (GATS), a product of the WTO, is aimed at deregulating international markets in services, including education. Two key operating principles of GATS are: the most favored nation and national treatment. The former requires that all countries be treated identically with regard to import or export, while the latter enjoins countries to treat foreign companies at least as favorably as their equivalent national competitors. The purpose of these principles is the creation of an open, global market place where education and other services can be sold to the highest bidder.

GATS covers educational services which are not exclusively provided by the public sector, or those which have commercial purposes. Since education systems are rarely exclusively public and many programs are partially commercial, large parts of the education systems of countries fall under the purview of GATS.

Internationally, education is estimated to be a trillion-dollar industry. Corporations have seen the prospects of a deregulated education sector and of having a market share in the industry, to the extent education can be turned into a tradable commodity (Frase and O'sullivan, 2007).

This aspect of globalization is a direct challenge to the public good character of education. It is a crucial political and educational policy issue in many countries to decide how the balance should be struck between the public good character of education and education as a market commodity. Civil society bodies and NGOs have a role in advocacy and awareness raising in respect of molding public opinion about the political choice that must be made (Tabb, 2007).

NGOs and Globalization

There is no one widely accepted definition of the term nongovernmental organization.

The diversity of NGOs, with hugely varied goals, structures, and the driving force behind them, makes it difficult to agree on such a definition.

While NGOs vary widely, their activities often fall into two broad categories: they engage in operational activities related to delivery of services and/or they undertake advocacy activities designed to influence policies and programs. Their size and scope of activities can be limited to a local community or extend across nations. Many NGOs in developing countries receive international funding for their work in areas such as research, advocacy, relief, environment, human rights, and health and education services.

Education is an important arena of public good in which NGOs have played a prominent role both as provider of services and in undertaking advocacy. The scope and nature of their involvement in education and the impact and significance in respect of globalization of education in a particular national, subnational, and international context vary in line with the diversity of NGOs themselves. However, the overall significance of NGOs in shaping globalization of education is impressive, as will be seen below (The Commission and Non-Governmental Organisations, 2000).

Economic globalization and advances in communication and transportation technology have influenced the growth of international NGOs and global civil society institutions. Although international religious and academic networks can be traced back to the Middle Ages, international NGO activity began to spread during the middle of the nineteenth century. They multiplied since the last quarter of that century (Seary, 1996; Charnovitz, 1997). It has been argued that governments followed the lead of international NGOs in addressing international issues (Charnovitz, 1997).

In the 1980s, funding for NGOs from bilateral aid providers, such as the US Agency for International Development (USAID) and the official European donors as well as private Northern donors, grew significantly. The prevailing neoliberal and pro-market policies of US and other Northern governments promoted the reduction of the role of the state and expansion of the role of the private sector, including NGOs.

While economically the role of NGOs was to complement the provisions for public goods, politically NGOs served as building blocks of global civil society (Spiro, 1995; Mathews, 1997). Participation of NGOs in the UN system deliberations on global issues has helped the process of the formation of global civil society. It has been observed that there are often strong differences of viewpoint between Northern and Southern NGOs in these

deliberations. For example, at the UN Conference on Environment and Development (UNCED) in Rio in 1994, Southern NGOs raised concerns about the debt crisis and multinational corporations contributing to environmental destruction, while Northern NGOs were more interested in commitments to preserve certain resources (Clark *et al.*, 1998). This diversity of voices representing genuine stakeholder interests added to the legitimacy and credibility of NGOs.

NGO Response to Globalized Education – Two Examples

Building Resources Across Countries

BRAC (formerly known as the Bangladesh Rural Advancement Committee, renamed in 2008 as Building Resources across Countries, but well known just by the acronym BRAC) illustrates the significant role of NGOs in national development, including the development of education. BRAC, exceptionally among Southern NGOs, has also become an international organization working in several Asian and African countries.

BRAC was established as a relief organization to assist war-displaced people in the aftermath of the Bangladesh Liberation War. It now operates in all 64 districts of Bangladesh as well as in Afghanistan, Sri Lanka, Pakistan, Tanzania, Uganda, Southern Sudan, Sierra Leone, and Liberia. In Bangladesh, BRAC has over 7 million micro-finance group members, 37 500 nonformal primary schools, and more than 70 000 health volunteers. It employs over 120 000 full-time and part-time staff, the majority of whom are women. Among others, BRAC operates its own bank and runs a well-known university. BRAC USA and BRAC UK have been established as affiliates both to generate funding support, especially for the international programs, as well as to operate programs aiming at the disadvantaged immigrant populations in the two countries (BRAC, 2008).

BRAC's Non-Formal Primary Education (NFPE) program provides the 5-year primary education course in 4 years to poor, rural, disadvantaged children and drop-outs, who cannot access formal schooling. These one-room-one-teacher schools are for children between 8 and 14 years of age. Each school typically consists of 33 students who as a cohort are taken through the 4-year program, before another cohort is recruited. Core subjects include mathematics, social studies, and English in line with the national curriculum objectives. The schools also offer extracurricular activities. As of June 2008, 37 500 primary schools and 24 750 pre-primary schools were in operation enrolling nearly 3 million children, 65% of whom are girls. The schools have a dropout rate of less than 5%.

BRAC has set up centers for adolescent girls called Kishori Kendra that provide reading material and serve as a gathering place for adolescents where they are educated about issues sensitive to the Bangladeshi society such as reproductive health, early marriage, and women's legal rights. BRAC has also set up about a 1000 community libraries, 20% of which are equipped with computers (Smilie, 2009).

The Global Campaign for Education

The Global Campaign for Education (GCE) is a civil society movement, established originally with the initiative of Education International, the worldwide federation of teachers' organizations. Its aim is to end the global education crisis. Its mission is to hold governments to account for their promises made to provide EFA and to ensure that they act to deliver the right of children and adults to quality public education.

Since its formation in 1999, organizations and people from over 100 countries – civil society organizations, trade unions, child rights campaigners, teachers, parents, and students – have united under its auspices to demand universal education.

GCE was a strong civil society voice in the World Education Forum in Dakar 2000 in setting the EFA agenda for 2015. The GCE's membership is open to national coalitions and international and regional organizations that advocate for EFA and subscribe to GCE's principles and goals. The main advocacy approaches and activities of GCE are summarized below.

Global action week

GCE coordinates the Global Action Week on education with all its national and international partners in April every year. Activities included the World's Biggest Lesson and Send My Friend to School. In 2008, close to 9 million people took part in the World's Biggest Lesson, including national leaders and celebrities aimed at drawing attention to the state of education and actions that must be taken.

Year-round campaign

GCE represents civil society on education at key education decision-making bodies, including the Education Fast Track Initiative Steering Committee, the EFA UNESCO-led Working Group and High Level Group. It also guides national campaigns to lobby the annual G8 and World Bank and International Monetary Fund (IMF) meetings to demand adequate long-term financing for education.

EFA class of 2015

GCE helped launch the EFA Class of 2015 at the UN high-level meeting on the Millennium Development Goals on 25 September 2008. This class brings together governments, faith leaders, corporations, civil society, and

education advocates who are committed to making sure that the EFA goals become a reality by 2015.

Global advocacy

GCE encourages policy research on different aspects of the EFA agenda, what is needed to achieve EFA and rate the performance of national governments. It has produced a school report that ranks governments' efforts toward achieving EFA.

Recent Developments

There have been always winners and losers from economic globalization. Who the winners and who the losers are from the inexorable forces of globalization are determined by whether people live in the global North or in urban and developed parts of a country, whether they are skilled and educated, whether people have acquired the capacity to adjust to change and new situations easily, whether they possess assets, their gender and ethnic status, and such factors over which people have little control (Streeten, 2001).

Globalization-induced improvements in a nation's standard of living also can promote women's status and raise household income, education, nutrition, and life expectancy, while lowering birthrates. The key issues clearly are: discerning the factors which create the winners and losers; identifying the vulnerable groups affected by these factors; determining the conditions and actions that would reduce the vulnerability of people; and promoting and supporting the creation of these conditions and enabling people to engage in these actions. Enhancement of people's capabilities to discern and assess the options and make the choices can happen through effective access to education, knowledge, and information.

The quality of human resources determines success for a nation in the era of intensified globalization. The intellectual capital of a nation has become more important than natural resources and financial capital to a country's ability to compete in the global economy.

The economic crisis of 2008–9, manifested in the meltdown of the financial market, causing huge contraction of capital, decline in consumption, and increase in unemployment has posed new challenges to the legitimacy of the premises underlying economic globalization. This situation creates new opportunities for the civil society and NGOs to reinforce the international coalitions and compromises that are necessary to mitigate the negative aspects of globalization and enhance the positive effects – putting a higher premium on global human solidarity than on unimpeded global competition. Education and learning, in multiple ways and in the broadest sense, have to be regarded as a critical element in this effort.

Bibliography

- BRAC (2008). *Annual Report*. Dhaka: BRAC.
- Charnovitz, S. (1997). Two centuries of participation: NGOs and international governance. *Michigan Journal of International Law* **18**(2), 183–286.
- Clark, A. M. (1995). Non-governmental organizations and the influence on international society. *Journal of International Affairs* **48**, 507–525.
- Clark, A. M., Friedman, E. J., and Hochstetler, K. (1998). Global civil society: A comparison of NGO participation in United Nations world conferences on *World Politics*, vol. 52, No. 1, p 136.
- Cogburn, D. L. (1998). Globalisation, knowledge, education, and training in the information age. *International Forum on Information and Documentation* **23**(4), 23–29.
- Croucher, S. L. (1994). *Globalization and Belonging: The Politics of Identity in a Changing World*. London: Rowman & Littlefield.
- Cummings, W. K. and McGinn, N. F. (eds.) (1997). *International Handbook of Education and Development: Preparing Schools, Students, and Nations for the Twenty-First Century*. Kidlington: Elsevier Science.
- European Commission (2000). The commission and non-governmental organisations: Building a stronger partnership. http://ec.europa.eu/civil_society/ngo/en/communication.pdf (accessed October 2009).
- Frase, T. and O'Sullivan, B. (2000). *The Future of Education under the WTO*. Madison, WI: 180/Movement for Democracy and Education Clearing House.
- Potier, B. (2004). Globalization and education explored at GSE. *Harvard University Gazette*, 4 March, 2004.
- Seary, B. (1996). The early history: From the Congress of Vienna to the San Francisco Conference. In Willetts, P. (ed.) *The Conscience of the World: The Influence of Non-Governmental Organizations in the UN System*, pp 15–30. London: Hurst.
- Smilie, I. (2009). *Freedom from Want: The Remarkable Success Story of BRAC, the Global Grassroots Organization That's Winning the Fight against Poverty*. W. Hartford, CT: Kumarian Press.
- Spiro, P. J. (1995). New global communities: Nongovernmental organizations in international decision-making institutions. *The Washington Quarterly* **18**(1), 45–56.
- Streeten, P. (2001). *Globalization: Threat or Opportunity?* Copenhagen: Copenhagen Business School Press.
- Tabb, W. (2007). Globalization and education as a commodity. *Clarion. Public Service Congress*, City University of New York, Summer (2007).
- UNESCO (1998). Learning: The treasures within. *Report of the International Commission on Education for the 21st Century*. Paris: United Nations Educational, Scientific and Cultural Organization.
- UNESCO (2000). *World Education Report 2000. The Right to Education: Towards Education for All Throughout Life*. Paris: United Nations Educational, Scientific and Cultural Organization.
- United Nations (nd) United Nations Documents: Research Guide. <http://www.un.org/Depts/dhl/resguide/resins.htm> (accessed October 2009).
- Wolf, M. (2005). *Why Globalization Works*. New Haven, CT: Yale University Press.

Further Reading

- Chabbott, C. (2003). *Constructing Education for Development: International Organizations and Education for All*. New York: Routledge Falmer.
- Drucker, P. F. (1997). The Global Economy and the Nation-State. *Foreign Affairs* **76**(5), 159–171.
- Gordenker, L. and Weiss, T. G. (eds.) (1996). *NGOs, the UN and Global Governance*. Boulder, CO: Lynne Reinner.
- Nadjam, A. (1998). Searching for NGO Effectiveness. *Development Policy Review* **16**, 305–310.

- Rossello, P. (1970). Historical Note. In UNESCO/IBE *International Conferences on Public Education: Recommendations*. Paris: United Nations Educational, Scientific and Cultural Organization.
- Spring, J. (2008). *Globalization of Education: An Introduction*. Oxford, UK: Taylor and Francis.
- Suarez-Orozco, M. (ed.) (2007). *Learning in the Global Era: International Perspectives on Globalization of Education*. Berkeley, CA: University of California Press.
- Valderama, F. (1995). *A History of UNESCO*. Paris: United Nations Educational, Scientific and Cultural Organization.

Relevant Websites

- <http://www.un.org> – GA Resolutions, International Instruments (1946 onwards), UN Documentation: Research Guide, United Nations (UN).
- <http://www.campaignforeducation.org> – Global Campaign for Education.
- <http://ec.europa.eu> – The commission and non-governmental organizations: Building a stronger partnership. *Commission Discussion Paper* by Prodi and Kinnock, European Commission.

This page intentionally left blank

S

SOCIAL AND EMOTIONAL ASPECTS OF LEARNING

Contents

Achievement Goal Theory: Definitions, Correlates, and Unresolved Questions

Affect, Mood and Emotions

Anxiety

Attribution Theory

Children's Friendship

Coping with Stressful Situations: An Important Aspect of Self-Regulation

Culture in Motivation Research: A Challenging and Enriching Contribution

Early Social Development and Schooling

Emotion in Educational Contexts

Empathy

Flow in Education

Interest

Intrinsic and Extrinsic Motivation

Motivating Students in Classrooms

Motivation Regulation

Peer Learning in the Classroom

Poverty, Effects of on Social and Emotional Development

Seeking Help as an Adaptive Response to Learning Difficulties: Person, Situation, and Developmental Influences

Self-Concept in Learning: Reciprocal effects model between academic self-concept and academic achievement

Self-Efficacy Beliefs

Social and Emotional Outcomes of Learning

Social Aspects of Collaborative Learning

Social Development and Schooling

Social Interaction and Learning

Social Networks and the Education of Children and Youth

Sociocultural Issues in Motivation

Socioemotional Aspects of Technology-Supported Learning

Volitional Control of Learning

Wellbeing

Achievement Goal Theory: Definitions, Correlates, and Unresolved Questions

T Urdan, Santa Clara University, Santa Clara, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Research on achievement goals has emerged over the last three decades as one of the most prominent fields of motivational research (Elliot, 2005; Weiner, 1990). Achievement goals are defined somewhat differently by

various researchers, but there is general agreement that goals refer to the perceived purposes of achievement. Two broad categories of purposes have been most commonly examined: the goal of developing competence through

learning, understanding, and building new skills and the goal of demonstrating competence, either to oneself or others, according to social-comparative norms. The first class of goals has been most commonly called mastery goals (or task goals), whereas the second category of goals is currently most often referred to as performance goals (it has also been called ability, relative ability, and ego goals). Achievement goals have also been distinguished according to their valence properties. In the mid-1990s, a number of researchers demonstrated that achievement goals, particularly performance goals, differed according to whether they represented an appetitive desire to approach success (performance-approach goals) or a fear-based desire to avoid appearing incompetent (performance-avoidance goals; Elliot and Harackiewicz, 1996; Middleton and Midgley, 1997; Skaalvik, 1997). A performance-approach goal represents a desire to outperform others, whereas a performance-avoidance goal reflects a goal of avoiding being outperformed by others. More recently, mastery goals have also been divided into approach and avoidance dimensions where a mastery-approach goal represents a goal of gaining competence and skills and a mastery-avoidance goal represents a concern with not doing as well as one could or with deteriorating skills (Pintrich, 2000).

Roots of Achievement Goal Theory

The achievement goal framework grew out of, and in response to, a number of motivational theories prominent in the 1970s. The need-based theories of Atkinson, McClelland, and their colleagues that held sway from the 1950s to the 1970s raised important questions about the stability of motivational tendencies. Need for achievement (nAch) and fear of failure (FOF) were believed to represent stable, personality-like needs of individuals. As social-cognitive theories such as attribution theory and self-efficacy research began to gain footholds, a group of researchers at the University of Illinois (Carole Ames, Carol Dweck, Martin Maehr, and John Nicholls) developed achievement goal theory. They conceptualized achievement goals as motivational dispositions that were not only partly stable orientations but also somewhat malleable and subject to the influence of messages in the achievement context regarding the definition and purposes of achievement in those contexts. Whereas achievement motivation researchers described competitive, avoidance, and achievement orientations as stable aspects of personality, goal theorists generally ascribed greater responsibility to situational factors (Elliot, 2005; Urdan, 1997).

Achievement goal theory also has foundational roots in attribution theory. Carol Dweck and her colleagues examined why elementary school children often differed in their responses to failure. They argued that some children

explained their failure to causes they perceived as stable (i.e., a lack of intelligence or ability), leading to a helpless response to failure. Others attributed their failure to unstable, controllable causes (e.g., a lack of effort), leading to effortful and optimistic responses to failure (Diener and Dweck 1978; Dweck and Leggett, 1988). These different explanations for failure were associated with different achievement goals. When one believes that ability can be improved through effort, the goal of developing ability (i.e., a mastery goal) will be adopted. In contrast, when one believes that ability is trait like, with little hope of altering it through effort, one becomes concerned with demonstrating superior or ability or hiding a relative lack of ability (i.e., performance goals).

The development of achievement goal theory also has strong ties to progressive ideals. Nicholls (1989) was unabashed in his belief that the democratic ideal of American education could best be realized through the pursuit, and promotion, of mastery goals. As mastery goals include a definition of achievement that is internal rather than social-comparative, all children can achieve mastery goals, regardless of their ability. After all, everyone can improve and learn, even if these occur at different rates for different students. Since performance goals define achievement comparatively, any system that promotes performance goals creates a situation where some will lose and others will win. Nicholls argued that although the pursuit of performance goals may lead to achievement in school, such an association is indicative of school systems promoting the wrong (i.e., competitive) values rather than the inherent benefits of performance goals.

The last decade or so has seen a return to the achievement motivation roots of achievement goal research. Elliot and his colleagues have argued that the nAch and the FOF are the causal antecedents of mastery and performance goal adoption (Elliot, 1997; Elliot and Harackiewicz, 1996). According to this model, mastery goals are driven primarily by an nAch, performance-avoidance goals by an FOF, and performance-approach goals by a combination of nAch and FOF. They argue that achievement goals represent the proximal, cognitive influences of achievement-related behavior that channel the more general, unconscious motive dispositions of nAch and FOF.

Correlates of Achievement Goals

A wealth of research, mostly using survey and experimental methods, has revealed a pattern of associations among achievement goals, self-regulatory strategy use, affective and motivational variables, and achievement (for reviews of this research, see Elliot, 2005; Harackiewicz *et al.*, 2000; Midgley *et al.*, 2001; Urdan, 1997). Most of this research has found a generally positive constellation of correlates with mastery goals. For example, when pursuing mastery

goals, students generally persist longer when faced with difficulty, are more willing to attempt difficult or challenging tasks, use more deep-level cognitive processing strategies, are more intrinsically motivated, and feel better about school and school work. Despite these positive motivational, affective, and cognitive correlates of mastery goals, researchers have not consistently found positive associations between mastery goals and achievement, both in school and on experimental tasks (Elliot, 2005). Recent research examining the association between mastery goals and student performance on assessments designed to match instruction in actual classrooms has revealed modest positive correlations between achievement and mastery goals (Linnenbrink, 2005; Nolen, 2003).

In contrast to the correlates of mastery goals, pursuing performance-avoidance goals is usually associated with a negative pattern of motivational beliefs and behaviors. When students are performance-avoidance goal oriented, they are more likely to give up when faced with difficult work or confronted with failure, use more shallow-level cognitive strategies (such as rote memorization), are less likely to seek help when they need it, and are more likely to engage in self-defeating practices like self-handicapping. As performance-avoidance goals represent a concern with being outperformed by others and with appearing incompetent, the primary emotions associated with these goals are shame and fear. Individuals who are able to achieve well despite their performance-avoidance goal orientation are most likely to feel relief rather than pride.

Whereas the maladaptive motivational and achievement consequences of pursuing performance-avoidance goals have been found consistently, the effects of performance-approach goals on motivation and achievement have been ambiguous. Some research has found that performance-approach goals are associated with lower persistence after failure (Elliott and Dweck, 1988), greater use of surface learning strategies and less use of deep cognitive processing strategies (Nolen, 1988), greater feelings of self-consciousness (Roeser *et al.*, 1996), negative affect in school (Urdu and Midgley, 2003), and the tendency to attribute failure to lack of ability (Dweck and Leggett, 1988). However, there is also a substantial body of research that has found significant positive associations between performance-approach goals and adaptive or beneficial motivational and achievement variables, including intrinsic motivation (Elliot and Harackiewicz, 1996), teacher-assigned grades (Harackiewicz *et al.*, 2002; Roeser *et al.*, 1996), achievement motivation (Elliot, 1997; Nicholls *et al.*, 1985), engagement in science class (Meece *et al.*, 1988), valuing of school (Midgley *et al.*, 1996; Wolters *et al.*, 1996), self-efficacy (Roeser *et al.*, 1996; Wolters *et al.*, 1996), and self-regulation (Wolters *et al.*, 1996). Indeed, a number of studies have found that performance-approach goals are positively correlated with mastery goals

(Nicholls *et al.*, 1985; Roeser *et al.*, 1996; Nolen and Haladyna, 1990).

It appears that the effects of a performance-approach goal orientation on motivation, affect, and achievement depend on a number of factors, including how the goals are measured, how they combine with the simultaneous pursuit of mastery goals (Barron and Harackiewicz, 2001; Meece and Holt, 1993; Pintrich, 2000; Wolters *et al.*, 1996), and personal characteristics of students including their gender (Urdu, 1997b), ethnicity (Midgley *et al.*, 1996), culture (Urdu, 2004), and grade level (Midgley *et al.*, 2001). The mixed pattern of associations with performance-approach goals makes global statements about the positive or negative effects of a performance-approach goal orientation difficult to support. Indeed, the ambiguous results have spawned a debate among achievement goal researchers regarding the potential benefits and harms of performance-approach goals (see Midgley *et al.*, 2001; Harackiewicz *et al.*, 2002). Midgley and her colleagues argued that the beneficial effects of performance-approach goals may be limited primarily to older students (i.e., college) with a history of high academic achievement. Molden and Dweck (2000) argued that performance-approach goals are beneficial as long as students are succeeding, but an orientation toward demonstrating superior ability can quickly change to a fear of demonstrating relative incompetence when students encounter failure or difficulty. Recently, Brophy (2005) argued that researchers' focus on the relative merits and drawbacks of performance goals may be misguided as observational and interview research suggests that students rarely mention social-comparative concerns in school settings unless directly prompted by researchers, as through the use of surveys. At present, the most definitive statement that can be made regarding performance goals is that much remains unresolved regarding their definition, operation, and influence on motivation and achievement.

Classroom Goal Structures

The achievement goals that have been discussed so far refer specifically to the personal goals that individuals pursue. In addition to personal goals, there are also goal structures. Goal structures refer to messages in the environment (e.g., experimental situation, classroom, and school) that make certain goals salient. Such goal-related messages can come from a variety of sources in the achievement context, including how the purposes or definitions of success are presented, the types of tasks that are assigned, how much time students are given to complete the task, and how individuals are recognized and rewarded in the classroom, school, work, or athletic context. For example, it is common in experimental manipulations of goals to provide research participants with

instructions that make one type of goal salient, and these instructions constitute the goal structure for that situation. Similarly, in a classroom, teachers can emphasize the ability differences between students (performance goal structure) by grading on a curve or emphasize the development of competence by using portfolio assessment to chart the progress of each student over time (mastery goal structure). Most researchers who have discussed goal structures have referred specifically to the classroom- or school-level goal structures (e.g., Ames, 1992; Anderman and Midgley, 1997).

Researchers examining goal structures have experimentally manipulated the goal messages in the achievement context (most commonly a laboratory setting), assessed students' and/or teachers' perceptions of the classroom goal structure in the classroom or school using surveys, conducted observations in classrooms to see which goals were emphasized, or some combination of survey and observational methods. Experimental manipulations of goals have yielded the most consistent results, generally producing an adaptive motivational pattern when mastery goals were emphasized, lower motivation and achievement when performance-avoidance goals were emphasized, and a mixed pattern of motivation and achievement when performance-approach goals were emphasized (Elliott and Dweck, 1988; Elliot and Harackiewicz, 1996). Some researchers have questioned whether the results of research using experimental manipulations of goals in laboratory settings will generalize to authentic achievement contexts such as classrooms. In the lab, it is easy to emphasize that a single achievement goal and performance on the assigned tasks generally have few, if any, long-term consequences for participants. In the classroom, students receive multiple goal messages and are usually engaged in activities that influence their grades, an outcome of significance for most students.

Research on the influence of classroom goal structures on motivation, affect, and achievement that has been conducted in classrooms and schools has produced much more ambiguous results than has laboratory-based experimental research (see Urdan (2004b) for a review). For example, survey research has revealed that students' perceptions of a mastery goal structure in the classroom are positively correlated with their own personal mastery goals, positive affect in school, self-regulatory strategy use, and sometimes course grades (Ames and Archer, 1988; Urdan and Midgley, 2003). However, other researchers have reported a lack of association between perceived classroom mastery goal structures and a variety of outcomes, including persistence, choice of challenging work, or course grades (Wolters, 2004), between changes in the perceived mastery goal structure and changes in students' valuing of mathematics and English over time (Anderman *et al.*, 2001), or expected negative associations between classroom mastery goal structures and self-handicapping (Urden *et al.*, 1998). Similarly, some research has found a

pattern of associations between perceived classroom performance goal structures and a variety of maladaptive motivational outcomes. For example, classroom performance goal structures have been found to correlate with procrastination (Wolters, 2004), self-handicapping (Urden *et al.*, 1998; 2004a), and attributing failure to lack of ability (Ames and Archer, 1988). However, perceived classroom performance goal structures have also been found to correlate with positive cognitive outcomes, such as attributing success to effort and failure to inappropriate strategy use (Ames and Archer, 1988) and have been found to be unrelated to other variables, such as cognitive and meta-cognitive strategy use (Wolters, 2004), and self-handicapping (Turner *et al.*, 2002).

There are several possible explanations for the inconsistent pattern of results found when examining classroom goal structures. One is that classroom goal structures may simply not be very strong, or clear, in most classrooms. As mentioned previously, genuine achievement contexts such as classrooms usually contain multiple achievement goal messages. Research suggests that students' perceptions of classroom goal structures are largely subjective creations rather than objective realities; therefore, students' existing personal goal orientations may shape their perceptions of the classroom goal structures to a significant degree (Brophy, 2005; Urden, 2001). Observational research and survey research using hierarchical linear modeling reveal that students in the same classroom vary widely in their perceptions of the classroom goal structures. A second plausible explanation for the mixed pattern of results is that researchers vary widely in their operational definition of classroom goal structures. The survey measures employed by Church *et al.* (2001), Ames and Archer (1988), Urden and Midgley (2003), Urden (2004a), Nolen (2003), and Greene *et al.* (2004) are so different from each other that comparisons across these studies are not likely to yield interpretable conclusions. It is not only clear that goal-related messages in achievement contexts can influence the personal goals students adopt, but also clear that research examining classroom goal structures is in its early stages and still needs significant advancement in conceptualization and measurement.

Remaining Questions about Achievement Goals

Although research on achievement goals has grown rapidly to become perhaps the dominant paradigm in the field, a number of important questions remain. Among these, perhaps none is more important than the question of what, exactly, an achievement goal is. A precise definition of achievement goals is critical for understanding the results of past research and to help the field advance. For

example, much of the debate discussed previously regarding the benefits and costs of performance goals may hinge on how performance goals are defined, both conceptually and operationally. For achievement goal research to be useful to educators, coaches, and parents, the definition of achievement goals must be clear.

Elliot (2005) argued that achievement goals have been defined as specific, desired end states (i.e., to master a concept or skill, to outperform other students in a class, to win a competition) and as broader, overarching orientations that include cognitive and affective components (see Ames, 1992; Dweck, 1986). In a recent qualitative study, Urdan and Mestas (2006) found that high school students often reported quite different reasons for pursuing the same achievement goal. For example, some students said they wanted to outperform others (i.e., a performance-approach goal) because they wanted to silence other students who had questioned their intelligence, whereas others said they wanted to outperform others to make their parents proud. Elliot (2005) argued that achievement goal researchers should separate achievement goals from the reasons behind their pursuit in order to define goals as clearly and simply as possible. However, even when goals are separated from their reasons, there is no consensus yet about the defining features of achievement goals. Grant and Dweck (2003) found that performance goals have been operationally defined as a desire to validate one's ability, to outperform others, or as a desire to simply do well on a task, such as getting a good grade. Using factor analysis on a survey measure they designed, Grant and Dweck found that performance goals divided into outcome goals (i.e., doing well), ability goals (i.e., validating ability), and normative goals (i.e., doing better than others and validating superior ability). Goal researchers have not yet agreed about the defining feature, or features, of performance goals. Differences in operational definitions used by goal researchers may explain the different patterns of results researchers have reported when examining the consequences and correlates of personal performance goals. Additional questions about the definition of achievement goals, including whether they are stable orientations or more situation specific, also need resolution.

In addition to the defining features of mastery and performance goals, questions remain regarding how many and which goals count as achievement goals. Several researchers have argued that social goals, including goals such as making or maintaining friendships, being a good citizen in the classroom, pleasing the teacher, and gaining social status among peers, are important goals that students pursue in the classroom (Ryan, 2001; Urdan and Maehr, 1995; Wentzel, 1991). Social goals were included in early goal measures, such as the ego/social goals construct of Nicholls *et al.* (1985). However, others have argued that social goals are fundamentally different from

mastery and performance goals because the latter goals are focused specifically on competence. Although there are dozens of goals that individuals might pursue (to be a good parent, to own a nice home, to lose weight, etc.), only those goals that focus on competence in achievement-related activities – such as school, sport, and work – should be included as achievement goals (Elliot, 2005). Other goals that have sometimes been included in research on achievement goals, including social goals and work-avoidance goals (i.e., the desire to get by with as little effort as possible), may be different in kind than are mastery and achievement goals.

In addition to questions about the precise definition of achievement goals, there remain a number of questions regarding how they operate. Most research has found that mastery and performance goals are either moderately correlated or orthogonal, meaning that individuals can, and do, pursue multiple goals simultaneously. Barron and Harackiewicz (2001) argued that mastery and performance may combine in three different ways: independently, additively, or interactively. In the independent model, mastery and performance goals have unique associations with different outcomes, but these associations are not affected by the presence or absence of the other goals. For example, students with a strong mastery goal may value mathematics more strongly than do students with a low mastery goal, regardless of whether they have a strong or weak performance-approach goal. In the additive model, mastery and performance goals combine to create stronger effects than do either goal alone. For example, students with strong mastery and performance-approach goals may have higher achievement than do students with either strong mastery or performance goals, but not both. Finally, the interactive model suggests that mastery and performance goals interact to influence outcomes. For example, perhaps performance-approach goals are positively related to persistence after failure when mastery goals are high but negatively related to persistence when mastery goals are low.

Research examining the associations between multiple goal configurations and outcomes has not yet yielded consistent results. Some studies have found that the optimal pattern of motivation and achievement is produced when students are high in their pursuit of mastery goals and low in their performance goal pursuit (Meece and Holt, 1993; Pintrich and Garcia, 1991). Others have found that the most adaptive pattern of outcomes is associated with the simultaneous endorsement of both mastery and performance goals (Elliot and Church, 1997; Pintrich, 2000). Barron and Harackiewicz (2001) examined multiple goals among college students in two ways: (1) using an experimental manipulation of goals and (2) simply asking participants to self-report their goals. They found that for self-reported goals, mastery predicted intrinsic motivation variables but not achievement on a math task, whereas

performance goals predicted achievement in success, but not difficult, conditions. They did not find any unique effects of multiple goal pursuit. These results suggest that mastery goals are predictive of some outcomes and performance goals are predictive of other outcomes; therefore, the authors concluded that pursuing both goals is adaptive because it allows students the best opportunity to succeed on multiple outcomes. In the manipulated goals experiment, Barron and Harackiewicz found that holding multiple goals may benefit both high- and low-achievement motivation students when they are in situations that do not match their achievement motive well.

In an ambitious quasi-experimental study, Linnenbrink (2005) conducted an experimental manipulation of the classroom goal structures in ten upper-elementary school classrooms during a 5-week mathematics unit. She discussed achievement goal theory with the five participating teachers and then assigned teachers to one of three conditions: mastery, performance, or combined mastery and performance. Linnenbrink gave teachers a set of strategies they could use to promote the appropriate goal or combination of goals and then observed them to see if they faithfully implemented strategies to promote the appropriate goal emphasis. She also used surveys to assess students' motivation (self-efficacy, interest value, and utility value), emotional well-being (positive affect, negative affect, and test anxiety), cognitive engagement (the quantity and quality of self-regulatory strategy use), and help-seeking behavior before the unit began and again immediately upon its conclusion. An end-of-unit achievement test was also administered pre- and postintervention. Linnenbrink found that scores on the unit test improved the most (from pretest to posttest) among students in the performance and combined performance-mastery conditions relative to the mastery goal condition. This difference between the groups decreased by the time of the follow-up test administered 5 weeks after the conclusion of the math unit. She also found that students in the performance-approach condition increased their use of expedient help seeking (i.e., just asking for someone to give them the correct answer), whereas students in the other two groups decreased slightly in their expedient help-seeking behavior from pre- to posttest. There were no goal-structure-group differences, either in mean level or change over time, for motivation, emotional well-being, cognitive engagement, or adaptive or avoidance help seeking. From the research examining multiple goals, it appears that mastery goals are predictive of some outcomes and performance goals are predictive of others. When combined, their additive and interactive power does not appear to be substantial. More research on the combined effects of mastery and performance goals, over time, under success and failure conditions, and with populations of different ages is needed before firm conclusions can be drawn.

Finally, more research is needed that examines possible differences by age, culture, and achievement level in the definition and function of achievement goals. A limited body of research indicates that cultural factors such as race (Midgley *et al.*, 1996) and generational status (Urdan, 2004a) may influence the way students interpret goal-related messages in the classroom and the association between achievement goals and various outcomes. As noted earlier, some have argued that achievement goals, particularly performance goals, have different consequences for young (i.e., elementary) and older (i.e., college) students (Midgley *et al.*, 2001) and for high achievers compared to low achievers (Dweck and Leggett, 1988). Once clear conceptual and operational definitions of achievement goals have been established, research examining these cultural, age, and achievement level differences should prove most interesting.

Conclusion

The achievement goal framework has produced an incredibly productive and varied body of research. As the field reaches middle age, a number of important conceptual and empirical questions remain. The continuing maturation of research in this area will undoubtedly produce findings of importance to motivation researchers and educators alike.

See also: Culture in Motivation Research: A Challenging and Enriching Contribution; Emotion in Educational Contexts; Interest; Intrinsic and Extrinsic Motivation; Learning Strategies; Motivating Students in Classrooms; Motivation Regulation; Self-Concept in Learning: Reciprocal effects model between academic self-concept and academic achievement; Self-Efficacy Beliefs; Self-Regulated Learning and Socio-Cognitive Theory; Volitional Control of Learning.

Bibliography

- Ames, C. A. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology* **84**, 261–271.
- Ames, C. and Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology* **80**, 260–267.
- Anderman, E. A., Eccles, J. S., Yoon, K. S., *et al.* (2001). Learning to value mathematics and reading: Relations to mastery and performance-oriented instructional practices. *Contemporary Educational Psychology* **26**, 76–95.
- Anderman, E. M. and Midgley, C. (1997). Changes in personal achievement goals and the perceived classroom goal structures across the transition to middle level schools. *Contemporary Educational Psychology* **22**, 269–298.
- Barron, K. E. and Harackiewicz, J. M. (2001). Achievement goals and optimal motivation: Testing multiple goal models. *Journal of Personality and Social Psychology* **80**, 706–722.
- Brophy, J. (2005). Goal theorists should move on from performance goals. *Educational Psychologist* **40**, 167–176.

- Church, M. A., Elliot, A. J., and Gable, S. L. (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology* **93**, 43–54.
- Diener, C. I. and Dweck, C. S. (1978). An analysis of learned helplessness: Continuous changes in performance, strategy and achievement cognitions following failure. *Journal of Personality and Social Psychology* **36**, 451–462.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist* **41**, 1040–1048.
- Dweck, C. S. and Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review* **95**, 256–273.
- Elliot, A. J. (2005). A conceptual history of the achievement goal construct. In Elliot, A. J. and Dweck, C. S. (eds.) *Handbook of Competence and Motivation*, pp 52–72. New York: Guilford.
- Elliot, A. J. and Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology* **70**, 461–475.
- Elliot, E. S. and Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology* **54**, 5–12.
- Grant, H. and Dweck, C. S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology* **85**, 541–553.
- Greene, B. A., Miller, R. B., Crowson, H. M., Duke, B. L., and Akey, K. L. (2004). Predicting high school students' cognitive engagement and achievement: Contributions of classroom perceptions and motivation. *Contemporary Educational Psychology* **29**, 462–482.
- Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, J. A., and Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology* **94**, 638–645.
- Linnenbrink, E. A. (2005). The dilemma of performance-approach goals: The use of multiple goal contexts to promote students' motivation and learning. *Journal of Educational Psychology* **97**, 197–213.
- Meece, J. L., Blumenfeld, P. C., and Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology* **80**, 514–523.
- Meece, J. L. and Holt, K. (1993). A pattern analysis of students' achievement goals. *Journal of Educational Psychology* **85**, 582–590.
- Middleton, M. and Midgley, C. (1997). Avoiding the demonstration of lack of ability: An under-explored aspect of goal theory. *Journal of Educational Psychology*.
- Midgley, C., Arunkumar, R., and Urdan, T. (1996). "If I don't do well tomorrow, there's a reason:" Predictors of adolescents' use of academic self-handicapping strategies. *Journal of Educational Psychology* **88**, 423–434.
- Midgley, C., Kaplan, A., and Middleton, M. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, and at what costs? *Journal of Educational Psychology* **93**, 77–86.
- Molden, D. C. and Dweck, C. S. (2000). Meaning and motivation. In Sansone, C. and Harackiewicz, J. M. (eds.) *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*, pp 131–159. New York: Academic Press.
- Nicholls, J. G. (1989). *The Competitive Ethos and Democratic Education*. Cambridge: Harvard University Press.
- Nicholls, J. G., Patashnick, M., and Nolen, S. B. (1985). Adolescents' theories of education. *Journal of Educational Psychology* **77**, 683–692.
- Nolen, S. B. (1988). Reasons for studying: Motivational orientations and study strategies. *Cognition and Instruction* **5**, 269–287.
- Nolen, S. B. (2003). Learning environment, motivation, and achievement in high school science. *Journal of Research in Science Teaching* **40**, 347–368.
- Nolen, S. B. and Haladyna, T. M. (1990). Personal and environmental influences on students' beliefs about effective study strategies. *Contemporary Educational Psychology* **15**, 116–130.
- Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology* **92**, 544–555.
- Pintrich, P. R. and Garcia, T. (1991). Student goal orientation and self-regulation in the college classroom. In Maer, M. L. and Pintrich, P. R. (eds.) *Advances in Motivation and Achievement, Goals and Self-Regulatory Processes*, Greenwich, CT: JAI Press.
- Roeser, R. W., Midgley, C., and Urdan, T. (1996). Perceptions of the school psychological environment and early adolescents' self appraisals and academic engagement: The mediating role of goals and belonging. *Journal of Educational Psychology* **88**, 408–422.
- Ryan, A. (2001). The peer group as a context for the development of young adolescent motivation and achievement. *Child Development* **72**, 1135–1150.
- Skaalvik, E. M. (1997). Self-enhancing and self-defeating ego orientation: Relations with task and avoidance orientation, achievement, self-perceptions, and anxiety. *Journal of Educational Psychology* **89**, 71–81.
- Turner, J. C., Midgley, C., Meyer, D. K., et al. (2002). The classroom environment and students' reports of avoidance strategies in mathematics: A multimethod study. *Journal of Educational Psychology* **94**, 88–106.
- Urdan, T. (1997a). Achievement goal theory: Past results, future directions. In Maehr, M. L. and Pintrich, P. R. (eds.) *Advances in Motivation and Achievement*, vol. 10, pp 99–141. Greenwich, CT: JAI Press.
- Urdan, T. (1997b). Achievement goals and the orientation of friends toward school in early adolescence. *Journal of Contemporary Educational Psychology* **22**, 165–191.
- Urdan, T. (2004a). Predictors of academic self-handicapping and achievement: Examining achievement goals, classroom goal structures, and culture. *Journal of Educational Psychology* **96**, 251–264.
- Urdan, T. (2004b). Can achievement goal theory guide school reform? In Pintrich, P. R. and Maehr, M. L. (eds.) *Advances in Motivation and Achievement*, vol. 13.
- Urdan, T. and Maehr, M. L. (1995). Beyond a two-goal theory of motivation: A case for social goals. *Review of Educational Research* **65**, 213–244.
- Urdan, T. and Mestas, M. (2006). The goals behind performance goals. *Journal of Educational Psychology* **98**, 354–365.
- Urdan, T. and Midgley, C. (2003). Changes in the perceived classroom goal structure and pattern of adaptive learning during early adolescence. *Contemporary Educational Psychology* **28**, 524–551.
- Urdan, T., Midgley, C., and Anderman, E. (1998). The role of classroom goal structure in students' use of self-handicapping strategies. *American Educational Research Journal* **35**, 101–122.
- Weiner, B. (1990). History of motivational research in education. *Journal of Educational Psychology* **82**, 616–622.
- Wentzel, K. R. (1991). Social competence at school: Relation between social responsibility and academic achievement. *Review of Educational Research* **61**, 1–24.
- Wolters, C. A., Yu, S. L., and Pintrich, P. R. (1996). The relation between goal orientation and students' motivational beliefs and self-regulated learning. *Learning and Individual Differences* **8**, 211–238.
- Wolters, C. A. (2004). Advancing achievement goal theory: Using goal structures, and goal orientations, to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology* **96**, 216–235.

Further Reading

- Harackiewicz, J. M., Barron, K. E., Tauer, J. M., Carter, S. M., and Elliot, A. J. (2000). Short-term and long-term consequences of achievement goals in college: Predicting continued interest and performance over time. *Journal of Educational Psychology* **92**, 316–330.

Affect, Mood and Emotions

S Schnall, University of Plymouth, Plymouth, UK

© 2010 Elsevier Ltd. All rights reserved.

Definitional Issues

Despite a considerable body of literature dedicated to the scientific study of affect, there is surprisingly little consensus on some of the most basic concepts related to affect. To some extent, definitions relating to affective processes are specific to certain research traditions. For example, whereas some theorists consider the conscious experience of emotion one of its central features, others propose that emotional processes can also take place entirely outside of conscious awareness. There is broad agreement, however, that affect comprises both emotion and mood, which differ regarding their object and temporal constraints. An emotion typically follows a specific eliciting stimulus or event, and is intense but limited in duration. Certain action tendencies, such as whether one is more likely to approach or avoid the target of one's current emotional feeling, are associated with specific emotions. These action tendencies do not imply that the person will necessarily engage in that action; however, in principle, a readiness to engage in the action is present. A mood, on the other hand, is usually not attributable to a specific stimulus, is of low intensity, and of longer duration. Since the source of a mood is often ambiguous, the feeling can be incorrectly attributed to a stimulus under consideration, thus leading to various types of affective influences on evaluation and cognitive processing, as are discussed in more detail further below.

Emotions can be classified along two orthogonal dimensions, namely valence and arousal. Valence refers to whether the experience is positive or negative, and thus reflects the value of a given object or situation. Arousal refers to whether the person experiences a state of activation or deactivation, and signifies whether the event is important or trivial. Other classifications of emotions have involved categorizing them with regard to specific action tendencies, such as approach and avoidance. For example, although anger and sadness are both subjectively negative experiences, anger is considered an approach emotion because an angry person is inclined to confront the source of their anger, whereas sadness is considered an avoidance emotion because a sad person is likely to withdraw from other people.

Since affect can either pertain to emotion or mood, affective feelings refer to the subjective experience of emotions or moods. Such affective feelings can be distinguished from those that are nonaffective in nature, such as hunger, thirst, or boredom. Usually, convergent affective cues are

provided by multiple response systems, including subjective experience, cognitive processes, hormonal changes, physiology, motor behavior, and expressive behavior such as facial expressions, postures, and vocal signs. Both cortical and subcortical neural subsystems are thought to coordinate these multiple response systems. Affective processes are scientifically measured by tools closely relating to the different response systems. Such measures include self-reports (e.g., emotion rating scales), observational data (e.g., of facial expressions and instrumental actions), psychophysiological measurements (e.g., heart rate, skin conductance, and electromyography), and brain imaging techniques (e.g., functional magnetic resonance imaging (MRI) and electroencephalogram (EEG)).

Many issues in the study of affect have created great controversy and are still being debated. For example, whether there is a fixed number of basic emotions continues to be a matter of dispute. Some researchers have argued that evidence for culturally universal facial expressions of happiness, sadness, anger, fear, disgust, and surprise suggests a special status for these emotions, relative to others. Other researchers maintain that facial expressions are, however, one of many facets of the emotional experience, and do not provide enough evidence to suggest a privileged status of those relative to other emotions. Indeed, definitional issues of emotion are intricately linked to theoretical positions, with varying emphasis placed on the necessary and sufficient components of an affective experience. As will be reviewed below, different research traditions have emphasized functional, embodied, or cognitive components, and although they differ in the extent to which they consider one or the other primary, they have contributed to the general understanding of affective phenomena.

Components of Affect

Functional Aspects of Affect

Affective processes have historically often been contrasted with rational processes; however, it has become increasingly apparent that this is not a useful distinction because many affective phenomena are, at least in principle, functional in nature. Charles Darwin was among the first to scientifically study emotion from a functional perspective by exploring its evolutionary underpinnings.

There is widespread agreement that the function of emotion is to facilitate social interactions with conspecifics,

and to facilitate rapid responses toward potentially harmful aspects of the environment. As far as social relationships are concerned, comparative research has established that in animals and humans alike, early experiences with the primary caregiver are critical for affective functioning in later life. The attachment system ensures a warm and caring relationship between the child and caregiver, and this primary relationship results with the child developing an internal model of social relationships. Deficits in this early relationship can have detrimental effects on social and romantic relationships later in life. Although the attachment style is relatively stable and can have profound consequences on one's emotional responses within social relationships, it can be altered through specific efforts, for example, in the context of psychotherapy.

Presumably because of their evolutionary origins, affective responses toward certain stimuli are stronger than toward others. For example, human and nonhuman primates are especially likely to develop negative attitudes to snakes and spiders, whereas this is not the case for neutral stimuli such as flowers or mushrooms. The evolutionary argument for this preparedness in learning is that primates who differentially learned to avoid snakes, spiders, and other potentially harmful stimuli had a greater chance of survival than those who did not.

In addition to these basic ways in which affective processes can be considered evolutionarily adaptive, affective influences on cognitive processing have also been discussed as providing useful informational feedback about the state of the environment. On a general level, affective feelings provide feedback about whether something is good or bad. More precisely, they provide one with information about the momentary value of objects and situations. When making evaluative judgments and deciding whether something is desirable or not, people attend to their own feelings, as if asking themselves: How do I feel about it? The experience of these felt evaluations serves as information. Thus, people generally like what they feel good about and dislike what they feel badly about. As a consequence, affective feelings have been shown to influence ratings of life satisfaction estimates of risk, and many other judgments. In addition, many attitudes are shaped by affective processes.

Affective cues, however, are not only informative when interpreted as evaluations of objects and situations, but can also be interpreted as performance feedback while working on a task: positive feelings serve as success feedback, indicating that the current cognitive strategy is adequate, whereas negative feelings serve as failure feedback, indicating that a different cognitive strategy should be pursued. As a result, being in a good mood makes people more likely to process incoming information in relation to easily accessible knowledge, expectations, and heuristics; that is, a good mood facilitates top-down processing. In contrast, being in a bad mood makes people more likely to rely on

uninterpreted perceptions and analytical processing; that is, a bad mood facilitates bottom-up processing. Affective feelings, therefore, directly concern adaptive actions as well as which actions work well in pursuit of a goal or which ones might better be abandoned.

Contextual constraints can qualify emotional experiences on several levels. First, affect is embedded in a social, cultural, and historical context, and the meaning of affective practices is constructed in relation to this context. Modern-day research has confirmed that some aspects of emotion, such as the facial expressions of a number of emotions, are universal. Cultural variability exists, however, with different culture display rules regulating the contexts in which it is acceptable to publicly express an emotion. For example, in Western societies where a considerable emphasis is placed on feeling unique and independent, emotions like pride are more socially acceptable than in some Eastern cultures (e.g., Japan) where fitting into an interdependent community is more important. In these interdependent cultures, emotions that are defined by interpersonal demands, such as shame or guilt, serve to maintain social relationships.

Second, once emotions are experienced within this contextual background, they provide important information to the individual concerning what is good or bad. However, exactly how this information serves as evaluative feedback depends on the situation in which the emotional feelings are interpreted, or more precisely, toward what they are experienced. For specific emotions, such as feelings of anger, fear, happiness, etc., what object they are felt toward is usually clear. In contrast, more subtle moods, such as feeling mildly good or mildly bad, usually do not have a clear object attached to them and, thus, the source of these moods remains ambiguous. Since moods lack specificity and are therefore unconstrained regarding their informational value, they can be misinterpreted, or misattributed in various ways: people tend to report feeling more satisfied with their lives on a sunny day than on a rainy day; that is, people's judgments can be influenced by irrelevant affect from an external source such as the weather, provided they are not focally aware of this irrelevant source as the cause of their feelings. However, simply asking about the weather before obtaining judgments of life satisfaction is sufficient to eliminate the effects of weather on these judgments. Thus, a central finding in the affect literature is that irrelevant affective states influence judgments and cognitive processing only as long as their source is kept ambiguous. When this is the case, feelings are experienced as relevant to whatever is currently on one's mind, and the meaning of affective feelings depends on the current agenda of a person.

Overall, the functionality of affect is evident directly in the way people deal with certain stimuli in the environment, and indirectly in the way people cognitively process information about the environment.

Embodied Aspects of Affect

Emotion research has a long history of studying embodied aspects of experience. William James argued that the experience of emotion reflects the experience of bodily changes. According to this view, bodily sensations determine the subjective experience of emotion. Although this view has been criticized at times, it currently enjoys renewed support as a consequence of recent developments in the research area of embodied cognition. A direct consequence of the view that embodied cues constrain subjective experience is that emotional feelings can be regulated, and even initiated, by intentionally manipulating a person's physical state. For example, when a person is induced to put on a smile, the person is likely to report feeling happy. Similar evidence also has been obtained when research participants are induced to adopt postures characteristic of certain emotional states, such as fear, anger, or sadness. This research suggests that people read their emotional bodily behavior, and may experience their emotional physical states as emotional feelings. In other words, rather than being epiphenomenal to the experienced feeling of emotion, some theorists have argued that these bodily changes actually constitute the emotion and causally influence the feeling.

In addition to obvious emotional bodily cues such as facial expressions and emotional postures, other bodily cues can also be interpreted as providing affective information in appropriate contexts. For example, motor behaviors that are associated with agreement or disagreement have shown to influence attitudes. When research participants were induced to produce a vertical head movement, resembling head nodding, while listening to a communication about increasing tuition at their university, they were later more likely to agree with the tuition increase than participants who had been induced to produce a horizontal head movement resembling head shaking. Similarly, performing simple motor behaviors that are typically associated with approach behavior results in more liking of neutral stimuli than engaging in avoidance behavior. Presumably, people typically approach positive things and avoid negative ones, and finding themselves engaging in these behaviors produces the inference that an attitude object is more or less positive.

Considerations of the physicality of emotional experiences have increasingly involved the study of the brain. In particular, subcortical structures in humans and other mammals are considered fundamental in the subjective experience of emotion because of their direct connections with the hypothalamus, which in turn controls functions of the autonomic nervous system (heart rate, arousal, etc.) and the hormonal system. Subcortical structures are thought to be involved in the primary evaluation of the affective stimulus's valence as either positive or negative, and its importance of being personally relevant or

irrelevant. Although other subcortical structures are also critically involved, the amygdala appears especially important in this respect. The amygdala, which has been studied extensively in the processing of fear-eliciting stimuli, influences the cortex directly in order to provide an initial evaluation of a stimulus situation outside of conscious awareness. Evidence for this notion comes from the finding that whereas fear can be experimentally conditioned in the absence of cortical involvement, a conditioned fear response is absent after damage or removal of the amygdala. In addition, the amygdala activates cortical arousal systems of the brain which alert the organism to prepare itself for appropriate action. In other words, the amygdala serves two purposes at the very onset of an emotional experience: Direct projections to the cortex provide specific information about the stimulus situation, and nonspecific cortical arousal makes the organism ready for potential action.

The notion of cortical involvement in emotional processing is widely accepted, presumably because it has become clear that cognitive appraisals (see below) can modify a person's physical and subjective emotional experience. An early demonstration of the importance of the cortex in emotional experience was the famous case study of Phineas Gage, who, as a consequence of an accident, lost large sections of his prefrontal cortex, leading to profound deficits in emotional functioning for the rest of his life. Other individuals suffering from prefrontal brain damage have been studied using modern brain imaging techniques to localize specific brain lesions, and how these relate to abnormal affective experience. These patients tend to exhibit flat affect, that is, they do not appear to be experiencing any emotions at all, based on what can be inferred from their openly expressed emotions. Although it was initially suspected that perhaps only the expressive, rather than the experiential, component of emotion might be impaired, data regarding physiological reactivity suggested otherwise. Whereas healthy individuals show characteristic physiological reactions to an emotional stimulus, patients with specific damages confined to the prefrontal cortex do not show, for example, skin conductance responses to emotionally disturbing stimuli. However, these patients still indicate that they cognitively know that the stimuli are disturbing and ought to elicit negative emotions. Thus, although the knowledge about appropriate emotional reactions is still intact and accessible, conscious emotional feelings cease to exist in patients with certain prefrontal cortical impairments. Interestingly, in addition to their impaired affective functioning, these patients also show deficits in decision making because they appear unable to generate hypothetical mental scenarios that would help them to selectively choose among several options. Thus, contrary to the assumption that

emotional feelings need to be controlled to allow for rational thinking, the absence of emotional feelings can be even more debilitating as far as efficient, adaptive decisions are concerned.

Cognitive Aspects of Affect

In the 1960s, researchers advanced the notion that emotional experiences involve an evaluation to determine their relevance for the individual. This conception was later expanded to include a primary appraisal of determining personal relevance, and a secondary appraisal that involves assessing one's ability and resources to cope with the emotional situation. Further elaborations of appraisal theories have explicated how specific appraisals lead to the experience of distinct emotional feelings. Cognitive approaches of affect, including appraisal theories, suggest that affective states can be modified by changing the cognitive component of the emotion. Many emotion-eliciting events can be cognitively reappraised or given a different meaning; as a consequence, the quality of the experience changes. For example, the feeling of anger after waiting in vain for a friend can give rise to the feeling of worry after learning that the friend was in a car accident. Thus, the same phenomenological experience, the bodily cues associated with an emotion, can be interpreted differently depending on the situational context and one's cognitive appraisal. Research on the self-regulation of emotion indicates that it is more beneficial to reappraise emotion-eliciting stimuli rather than trying to suppress an emotional reaction once this response has already occurred.

Appraisal processes have been implied as one component of emotional intelligence, which is considered a person's ability to appropriately identify and utilize emotional responses. Emotional intelligence consists of four aspects: (1) perceiving the emotion, and accurately identifying it, both in oneself and in others; (2) using appropriate emotions to solve problems and enhance cognitive processes; (3) understanding the elaborate interconnection of various emotions, and their potential consequences for oneself and others, for example, in the context of flexible planning; and (4) regulating or managing one's emotions effectively, which not only aims at not getting overwhelmed by emotions, but also not suppressing or denying them when they occur. Indeed, suppressing the outward expression of emotion can be associated with impoverished cognitive functioning (e.g., decreased memory performance), presumably because controlling one's emotional responses is effortful and takes up resources that could otherwise be dedicated to cognitive processing.

Modeled along widely used intelligence tests, assessments of emotional intelligence have been developed to test individual differences in the extent to which people understand and utilize their emotions. Such test scores

have been successfully used to predict the quality of social relationships, as assessed by individuals themselves and their family members and peers. In addition, there appears to be a negative relationship between emotional intelligence and various deviant and antisocial behaviors. For example, people scoring high in emotional intelligence are less likely to engage in bullying behavior or abuse drugs than those who score low. Additional research efforts will need to clarify the developmental trajectories of emotional intelligence, and the extent to which it can be taught in schools and other educational settings.

In addition to differences in emotional intelligence, variability in affective styles has been observed, with some individuals apparently being more prone to experience positive emotions than others who are more prone to experience negative ones. Further findings have established that these differences are relatively stable, with some people reporting consistently higher levels of happiness than others throughout the course of their lives. Surprisingly, once the basic human needs such as adequate shelter, food, and social ties are taken care of, wealth, material possessions, and other situational factors appear to only be weakly associated with increased happiness, suggesting that a person's level of happiness has a strong genetic component and is not influenced heavily by contextual factors. These differences in affective style have been documented as being associated with specific types of brain activity. Both in infants and adults, hemispheric dominance has been invoked in the experience of positive and negative emotions, with left-sided activation being associated with higher levels of self-reported positive affect. In general, the left hemisphere of the brain shows relatively greater activation with approach-related, positive emotions, whereas withdrawal-related, negative emotions show greater right-sided activation.

Summary

Although some central definitional issues in the field of affective studies remain to be controversial, the field has been accumulating considerable empirical evidence for certain core findings. This includes emphasizing relative contributions of bodily and cognitive influences and how they might have evolved in the course of evolution. In addition, findings through research with subcortical structures and the influence of bodily cues highlight the embodied aspects of affect. Similarly, approaching affect from a cognitive perspective illustrates how it can be altered through cognitive processes and appraisals. These varying theoretical perspectives of affect and their corresponding findings illuminate its complex nature, and it is only with more research that affect can be fully understood and these findings may be integrated.

Bibliography

- Bowlby, J. (1969). *Attachment and Loss: Attachment* vol. 1. New York: Basic.
- Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam.
- Davidson, R. J. (2000). Affective style, psychopathology, and resilience: Brain mechanisms and plasticity. *American Psychologist* **55**, 1196–1214.
- Davis, M. and Whalen, P. J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry* **6**, 13–34.
- Diener, E. (2000). Subjective well-being: The science of happiness, and a proposal for a national index. *American Psychologist* **55**, 34–43.
- Ekman, P. and Davidson, R. J. (eds.) (1994). *The Nature of Emotion: Fundamental Questions*. New York: Oxford University Press.
- Feldman Barrett, L. and Russell, J. A. (1999). The structure of current affect: Controversies and emergent consensus. *Current Directions in Psychological Science* **8**, 10–14.
- Lazarus, R. S. (1991). *Emotion and Adaptation*. New York: Oxford University Press.
- Lewis, M. and Haviland-Jones, J. M. (eds.) (2000). *Handbook of Emotions*, 2nd edn. New York: Guilford.
- Martin, L. L. and Clore, G. L. (eds.) (2001). *Theories of Mood and Cognition: A User's Handbook*. Mahwah, NJ: Erlbaum.
- Mayer, J. D., Salovey, P., and Caruso, D. R. (2004). Emotional intelligence: Theory, findings, and implications. *Psychological Inquiry* **15**, 197–215.
- Ortony, A., Clore, G. L., and Collins, A. (1988). *The Conceptual Structure of Emotions*. New York: Oxford University Press.
- Öhman, A. and Soares, J. J. F. (1998). Emotional conditioning to masked stimuli: Expectancies for aversive outcomes following nonrecognized fear-relevant stimuli. *Journal of Experimental Psychology: General* **127**, 68–82.
- Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotion*. New York: Oxford University Press.
- Schwarz, N. and Clore, G. L. (2007). Feelings and phenomenal experience. In Kruglanski, A. and Higgins, E. T. (eds.) *Social Psychology: Handbook of Basic Principles*, 2nd edn., pp 385–407. New York: Guilford.

Further Reading

- Clore, G. L. and Schnall, S. (2005). The influence of affect on attitude. In Albarracín, D. and Johnson, B. (eds.) *Handbook of Attitudes and Attitude Change: Basic Principles*, pp 437–490. Mahwah, NJ: Erlbaum.
- Feldman-Barrett, L., Niedenthal, P., and Winkelman, P. (eds.) (2005). *Emotion and Consciousness*. New York: Guilford.
- Frank, R. H. (1988). *Passions within Reason: The Strategic Role of the Emotions*. New York: Norton.
- Frijda, N. H. (1992). The laws of emotion. *American Psychologist* **43**, 349–358.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* **108**, 814–834.
- Laird, J. D. (2006). *Feelings: The Perception of Self*. New York: Oxford University Press.
- Lane, R. D. and Nadel, L. (eds.) (2000). *Cognitive Neuroscience of Emotion*. New York: Oxford University Press.
- LeDoux, J. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon and Schuster.
- Markus, H. and Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review* **98**, 224–253.
- Neumann, R., Förster, J., and Strack, F. (2003). Motor compatibility: The bidirectional link between behavior and evaluation. In Musch, J. and Klauer, K. C. (eds.) *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*, pp 7–49. Mahwah, NJ: Erlbaum.
- Prinz, J. J. (2006). *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford University Press.
- Richards, J. M. (2004). The cognitive consequences of concealing feelings. *Current Directions in Psychological Science* **13**, 131–134.
- Scherer, K. R., Schorr, A., and Johnstone, T. (eds.) (2001). *Appraisal Processes in Emotion*. New York: Oxford University Press.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist* **35**, 151–175.

Anxiety

M Zeidner, University of Haifa, Haifa, Israel

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cognitive-focused interventions – Emphasize the mediating role of cognitive processes in sustaining or eliminating test anxiety and refer to a wide array of therapeutic approaches directed toward modifying the worry and irrational thought patterns of test-anxious clients.

Emotionality – Consists of perceptions of autonomic reactions evoked by evaluative stress.

Emotion-focused interventions – Primarily aim at reducing the arousal and heightened emotional reactions of test-anxious persons when faced with stressful evaluative situations.

Situation-specific personality trait – In the context of test anxiety research, this trait refers to the individual's disposition to react with extensive worry, intrusive thoughts, mental disorganization, tension, and physiological arousal when exposed to evaluative situations.

Test anxiety – Refers to the set of phenomenological, physiological, and behavioral responses that accompany concern about possible negative consequences or loss of competence on an examination or similar evaluative situation. Test-anxious behavior is typically evoked when a person believes that his/her intellectual, motivational, and social capabilities are taxed or exceeded by demands stemming from the test situation.

Worry – Primarily refers to cognitive concern about the consequences of failure.

typically required to adjust to a variety of novel and challenging academic and social stressors and demands that may evoke anxiety. These include demanding coursework assignments; heavy workload; time pressures in meeting deadlines for submission of papers; insufficient finances and work opportunities; demanding examinations administered under stringent time pressures; poor student–faculty rapport; inadequate and crowded study conditions; and conflict between meeting academic demands, leisure, and extracurricular pursuits, and family responsibilities. The anxiety experienced by many students as a result of school or academic stress is such an unpleasant and painful experience that it is not surprising that anxiety often interferes with student learning, well-being, and health.

Anxiety is a universal human experience, intrinsic to the human condition. The construct is defined by a loosely coupled ensemble of cognitive, affective, somatic arousal, and behavioral tendency components, evoked in response to mental representations of future threat or danger in the environment (Zeidner, 2007). However, the nature of the specific environmental stimuli evoking anxiety appears to have changed considerably over the years (Zeidner, 1998). Whereas in ancient times, wild beasts, natural catastrophes, and the like plausibly served as major sources of environmental stress, in the modern achievement-oriented society, stress and anxiety are evoked largely by a wide array of social-evaluative encounters (e.g., appearing for a college aptitude test, defending a dissertation, and taking a vocational aptitude test battery).

Modern conceptualizations view anxiety as an important adaptation that signals to us what is potentially harmful, dangerous, or threatening in a stressful encounter, with critical value for human survival (Rachman, 2004). Thus, anxiety in educational contexts serves as a call for action, taking precedence before all other activities and shifting attention to the sensed threats and potential losses and dangers at hand. When an individual experiences anxiety this is important information suggesting that something in the environment has been appraised as threatening or harmful to one's well-being and some action needs to be taken to ward off the threat (Lazarus, 1999).

Anxiety is frequently cited as a key villain in the ongoing drama surrounding educational testing and evaluation, and is claimed to be among the factors at play in determining a wide array of unfavorable outcomes and contingencies (e.g., poor cognitive performance, scholastic underachievement, low satisfaction with school or college, and psychological distress and ill health; Zeidner, 1998).

Overview

Anxiety is incontestably the most researched of all emotional states in education (Schutz and Pekrun, 2007). Educational settings are contexts in which students at all levels are ubiquitously exposed to a wide array of potentially stressful and anxiety-evoking experiences. Thus, school-aged students are frequently exposed to stressors such as parental pressures for high achievement, fierce classroom competition for high grades, experiences of frustration and failure, teacher disapproval, peer conflict, social isolation and rejection, and physical and verbal aggression and abuse. Comparably, college students are

Clearly, many students have the potential to do well but perform poorly because of their debilitating levels of anxiety, thus limiting educational or vocational development. The loss to society of the full contribution of potentially capable students through anxiety-related distress and somatic ailments, underachievement and failure at school, or performance decrements constitutes an important problem for educational practitioners.

Conceptualizations

In contrast to early mechanistic views of anxiety as a unified construct, it is currently construed as a complex multidimensional construct embodying a series of interrelated cognitive, affective, and behavioral tendency components and reactions. Its complex nature, coupled with the fact that anxiety encompasses worry and self-preoccupation, physical upset, disruptive feelings, and maladaptive behaviors, makes it particularly difficult for educational researchers to sort out all these components. Anxiety has been variously conceptualized as an antecedent stimulus condition, a latent mediating process (e.g., as a probability of a harmful future outcome), and a response (physiological, affective, behavioral, etc.) to a stressful condition.

Since the early 1950s, the anxiety construct was dramatically advanced by a number of important conceptual distinctions, which helped refine thinking and research in the area. One useful distinction differentiates between anxiety as a relatively stable personality trait and as a more transitory state reaction to specific ego-threatening situations (Spielberger, 1972). Thus, trait anxiety refers to relatively stable individual differences in anxiety proneness, whereas state anxiety is a palpable, temporary reaction to a stressful event (e.g., final examinations) characterized by subjective feelings of tension, apprehension, nervousness, and worry, as well as by the activation or arousal of the nervous system. Whether or not students who differ in trait anxiety will show corresponding differences in state anxiety in the school or college setting depends on the extent to which each of them perceives a specific situation (e.g., college algebra examination) as psychologically threatening, and this is influenced, in turn, by each individual's constitution (e.g., numerical ability) and past experiences (e.g., number of mathematics courses taken).

Another important conceptual and methodological contribution to the evaluative anxiety literature is the distinction between facilitating and debilitating anxiety (Alpert and Haber, 1960). Accordingly, facilitating and debilitating anxiety, respectively, are claimed to lead to task-related and task-irrelevant behaviors during evaluative ego-threatening situations. A particularly useful conceptual distinction differentiates between worry and emotionality components of anxiety (Liebert and Morris,

1967). Worry, the cognitive component of anxiety, was viewed primarily as a cognitive concern about the consequences of failure on evaluative tasks (e.g., college aptitude examinations). By contrast, emotionality, the affective component of anxiety, was construed as perceptions of autonomic reactions evoked by stress. These two components are empirically distinct, though correlated, and worry relates more strongly to performance decrements than does emotionality.

Lazarus's transactional theory of stress and coping (Lazarus and Folkman, 1984; Lazarus, 1991) provides a contemporary and fundamental conceptual framework for the analysis of stress and anxiety in educational settings. According to this perspective, emotions, such as anxiety, reveals something of a person's goal hierarchy and belief system and how events in the immediate environment are appraised by the person. Thus, any evoked emotion reflects a high-level synthesis of several appraisals relating to the individual's adaptational status in the current environment. The core theme in anxiety is a danger or threat to ego or self-esteem, especially when a person is facing an uncertain, existential threat. Thus, the very presence of anxiety in an evaluative encounter is informative because it communicates that an existential threat has not been controlled very well, thus providing the researcher and educational specialists (counselors, school psychologists, etc.) with critical diagnostic information.

Evaluative Anxiety

A host of different types of anxiety may be relevant to specific educational settings (test anxiety, math anxiety, computer anxiety, social anxiety, etc.). These forms of anxiety are frequently encountered in education and share the prospect of personal evaluation in real or imagined social situations, particularly when a person perceives a low likelihood of obtaining satisfactory evaluations from others (Leitenberg, 1990). Next, we discuss two prevalent forms of evaluative anxiety in education – test and math/computer anxiety.

Test Anxiety

Test anxiety refers to the set of phenomenological, physiological, and behavioral responses that accompany concern about possible negative consequences or poor performance on an examination or a similar evaluative situation (Zeidner, 1998). Test-anxious behavior is typically evoked when a student believes that his/her intellectual, motivational, and social capabilities are taxed or exceeded by demands stemming from the test situation.

Test anxiety has taken on a variety of different meanings throughout its relatively brief history as a scientific construct. In the early days of research, the construct was

defined in motivational terms, either as drive level, goal interruption, or a need to avoid failure. Subsequently, it was conceptualized as a relatively stable personality disposition linked to cognitive–attentional phenomena. Accordingly, the highly anxious person is one who attends excessively to evaluative cues concerning personal competence, and to feelings of physiological arousal. Test anxiety may also be a concomitant of self-handicapping employed to preserve one's self-merit in the face of potential failure (Zeidner and Matthews, 2005). Cybernetic self-regulative models have seen test anxiety as resulting from a conflict between competing reference values (Zeidner, 2007).

Recent theorizing (Zeidner, 1998) emphasizes the distinction between test anxiety as an attribute of the person and as a dynamic process. From the first perspective, dispositional test anxiety may be construed as a contextualized personality trait. Accordingly, test anxiety refers to the individual's disposition to react with extensive worry, intrusive thoughts, mental disorganization, tension, and physiological arousal when exposed to evaluative contexts or situations. The more transient-state expressions of anxiety may be assessed separately from the more stable trait. From the second, process-oriented perspective, test anxiety depends on the reciprocal interaction of a number of distinct elements at play in the ongoing stressful encounter between a person and certain parameters of an evaluative situation. These elements include the specific educational context, individual differences in vulnerability (trait anxiety), threat perceptions, appraisals and reappraisals, state anxiety, coping patterns, and adaptive outcomes.

Math and Computer Anxieties

Both math and computer anxieties, respectively, are conceptually related to test anxiety through a common theme of concerns about evaluation (e.g., Rosen and Maguire, 1990). Math anxiety is defined by feelings of tension, helplessness, mental disorganization, and associated bodily symptoms that are evoked in mathematical problem-solving situations (Ashcraft, 2002). Math anxiety is claimed to interfere with the manipulation of numbers and the solving of complex mathematical problems in a wide variety of ordinary life and academic situations. Statistics anxiety, referring to the feeling of anxiety encountered when taking a statistics course or working on statistical analysis, has frequently been construed as a subset of math anxiety (Zeidner, 1991). Math anxiety, coupled with objective cognitive difficulties experienced in learning mathematics, may lead people to reject goals, such as scientific career choices, for which studying mathematics is instrumental.

Computer anxiety (sometimes termed computer phobia, technophobia, or cyberphobia) may be decomposed into anxiety about present or future interactions with computers or computer-related technologies; specific

negative cognitions or self-critical internal dialogs when interacting with the computer or when contemplating future computer interaction; and negative global attitudes about computers, their operation, or their societal impact (Weil *et al.*, 1990). The effects of computer anxiety on the utilization of computer-based technology may incur serious economic costs estimated at the level of billions of dollars per year (Bozionelos, 2001).

Math and computer anxieties may relate not just to the obvious stimulus attributes of mathematics/numbers and computers, but also to deeper personal concerns. Thus, math anxiety focuses not only on the evaluative nature of mathematics tests, but also concerns mathematical content (symbols, operators, etc.), its distinctive features as an intellectual activity (inductive and deductive reasoning, problem solving, etc.), and its meanings for many persons in our society (Richardson and Woolfolk, 1980). Similarly, computer anxiety is evoked by the consideration of the broader implications of computer use for perception of the self, society, and culture. Computer-anxious persons may also suffer from a more generalized technophobia, which itself is evident before adulthood (Weil *et al.*, 1990).

Similar to the state-trait distinction for test anxiety, trait math anxiety reflects relatively stable individual differences in the tendency to perceive situations involving the manipulation of numbers and the use of mathematical concepts and data as threatening or harmful. Persons high in trait math anxiety respond to these situations with elevations in state anxiety, involving both heightened emotion and interfering worry responses (Anton and Klisch, 1995). State math anxiety refers to elevations in worry, apprehension, and arousal in a situation involving mathematical content or reasoning. Likewise, in contrast to the dispositional nature of trait computer anxiety, state computer anxiety is aroused by specific objects (personal computer, scanner, printer, etc.) or situations (computer error). Individuals high in trait computer anxiety are especially vulnerable to state anxiety responses (Gaudron and Vignoli, 2002).

All forms of evaluative anxiety are quite common, with prevalence estimates in adults ranging from 20 to 50% for math and computer anxieties (e.g., Bozionelos, 2001). Experiencing various forms of evaluative anxiety in educational settings is a near-universal phenomenon across people differing in age, gender, and culture. Thus, meta-analyses of test anxiety data from various national sites show that although mean test anxiety levels vary to a certain extent across cultures, test anxiety is a prevalent and relatively homogenous cross-cultural phenomena. Furthermore, women tend to report higher levels of evaluative anxiety (test, math, and social) than men; however, the gender difference often does not translate into objective performance differences. In addition, as discussed below, evaluation anxiety has frequently been linked to performance decrements in educational settings.

Table 1 Tentative typology of test-anxious students

Type	Brief description
I. Students deficient in study and test-taking skills	Characterized by a major deficiency in study and test-taking skills. Their poor examination performance results from deficits that include problems in acquisition (encoding), organization/rehearsal (study skills), and retrieval/application during a test.
II. Students experiencing anxiety blockage and retrieval problems.	These students have efficient study skills but suffer from anxiety blockage, consequently encountering problems in retrieving information during the examination. These anxious students study effectively, but cannot handle the stresses and pressures of evaluative situations.
III. Failure-accepting students	Failure-accepting students are characterized by a personal history of repeated test failures. They come to accept low ability as the primary explanation of their failures. As a consequence, they become accepting of failure, exhibiting apathy, resignation, and a sense of defeat, not unlike reactions traditionally associated with learned helplessness.
IV. Failure-avoiding students	Failure-avoiding students are driven to achieve primarily as a means of protecting themselves against beliefs that they lack ability. For these students, effort is truly a double-edged sword. They may strive for success through meticulous preparation; yet, failure despite high efforts increases the probability that one's ability will be considered low, thus inducing anxiety reactions.
V. Self-handicappers	These students avoid diagnostic information about intellectual tasks by reducing effort or avoiding the test situation. Accordingly, if a low score is obtained, the self-handicapping student can rely on the debilitating effects of anxiety as an excuse to escape responsibility for actions, thus reducing otherwise burdensome expectations others hold for that person.
VI. Perfectionistic overstrivers	These overstriving perfectionists are characterized by high personal standards of academic success, perception of high or even exaggerated expectations, perceived doubt regarding the quality of academic performance, and a need for order and organization in their academic work. No effort is ever sufficient as the perfectionistic examinee seeks approval and acceptance and tries to avoid errors and failure through an endless cycle of self-defeating overstriving.

Discussions of evaluative anxiety in the literature are commonly guilty of a uniformity myth, conveying the impression that evaluative anxiety is a rather homogeneous category. In the domain of test anxiety research, Zeidner (1998) has sketched some distinct, yet potentially overlapping categories, of subjects with test anxiety (see **Table 1**). In fact, as this tentative typology of test-anxious students demonstrates, test anxiety has a variety of sources and, similarly, its behavioral consequences vary with contextual and personal factors.

Measurement and Assessment

We now briefly discuss a number of issues in anxiety assessment, focusing on subjective self-reports, which are by far the most popular observational procedure for mapping out the phenomenology of anxiety in educational settings.

Subjective Self-Report Measures

Subjective reports include any direct report by the person regarding his/her own anxiety experience and responses in a particular setting (learning mathematical operations, using new computer programs, taking examinations, engaging in social interactions, etc.). These assessments typically employ single-item rating scales (e.g., "Please indicate how anxious you were speaking before the entire

class, employing the following 7 point rating scale: 1 = not at all anxious, 7 = extremely anxious"); multi-item questionnaires (e.g., Spielberger's 20-item Test Anxiety Inventory, 1980); or oral interviews before, during, or after an important stressful event in the educational context.

Self-report instruments are now popular because they are considered to provide the most direct access to a person's subjective experiences in ego-threatening situations, possess good psychometric properties, are relatively inexpensive to produce, and are simple to administer and score (Zeidner, 1998). Self-report paper-and-pencil questionnaire measures of state anxiety ask individuals to report which of the relevant symptoms of anxiety they are currently experiencing in a particular situation, whereas trait measures ask subjects to report symptoms they typically or generally experience in a particular class of situations (e.g., public speaking, classroom examination, social interaction, and sports competition). Unfortunately, many studies use self-report data exclusively, without any attempt to measure salient behavior (e.g., through observational procedures), thus either under- or overestimating anxiety levels.

Rather fortunately, most popular anxiety inventories have satisfactory reliability coefficients, typically in the high 0.80s to low 0.90s. Among the factors influencing reliability are test length, test-retest interval, variability of scores, and variation within the test situation. However, at present, we have no infallible or perfectly objective criterion against which to validate anxiety scores. Scores

Table 2 Some alternative measures for assessing anxiety

<i>Type of assessment</i>	<i>Examples</i>
Physiological measures	Accretion levels of corticosteroids, adrenaline products, sugar, cholesterol, and free fatty acids.
Performance measures	Examination scores, semester grade point averages, and latency and errors in recall of stress-relevant stimulus materials.
Systematic observations of specific behaviors	Perspiration, excessive body movement, hand wringing, fidgety trunk movements, and inappropriate laughter when subjects were engaged in examination situations.
Trace measures	Amount of chewed traces on the pencil or ruler, sweat smudges on examination papers, and personal diaries.
Think-aloud procedures	Relating thoughts and emotions following or during stressful experience (e.g., "Please list as many thoughts and feelings as you can recall having during this algebra examination").

on ability tests, grade point average, observer ratings, behavior in structured evaluative situations, and the like have been employed as measures of criterion behaviors. A number of alternative measures of anxiety appear in Table 2.

Anxiety and Cognitive Performance

Scores of studies have investigated the complex pattern of the relations between anxiety and different kinds of performance. Various forms of evaluative anxiety (test, mathematics, computer, statistics, etc.) have been found to interfere with competence both in laboratory settings as well as in true-to-life test situations in school or collegiate settings. Processing deficits that relate to test anxiety, including general impairments of attention and working memory, together with more subtle performance changes, such as failure to organize semantic information effectively.

Hembree's (1998) meta-analytic study, based on 562 North American studies, demonstrated that test anxiety correlated negatively, though modestly, with a wide array of conventional measures of school achievement and ability at both high school and college levels. Data collected on students from upper elementary school level through high school show that test anxiety scores were significantly related to grades in various subjects, although the correlation was typically about -0.2 . Cognitive measures (i.e., aptitude and achievement measures combined) correlated more strongly with the worry than the emotionality component of test anxiety. Higher effect sizes were reported for low-rather than high-ability students and for tasks perceived as difficult rather than those perceived as being easy. Another meta-analysis reported by Ackerman and Heggestad (1997) showed a mean correlation of -0.33 between test anxiety and general intelligence test performance. Test anxiety was also correlated in the -0.20 to -0.30 range with other broad intellectual abilities including fluid and crystallized intelligence, learning and memory, visual perception, and mathematics ability.

There is a large literature on anxiety as a predictor of information processing in laboratory studies. The information-processing components sensitive to anxiety relate to input (encoding and acquisition of information), central processing (e.g., memory, language processing, conceptual organization, judgment, and decision-making), and output (e.g., information retrieval, response selection, and execution). These anxiety-related deficits, at various stages of processing, suggest some general impairment in attention and/or working memory. These various performance deficits are often attributed to high levels of worry and cognitive interference.

Both cognitive interference and cognitive bias appear to be pervasive in evaluative anxiety, influencing various stages of information processing (Eysenck, 1992). Anxiety often leads to scanning of the environment for threat (generating distractibility and attentional impairment), followed by focusing of attention on sources of threat (generating attentional bias). In addition, competence deficits may also be a consequence of poor skill acquisition. For example, deleterious effects of test anxiety may reflect not just cognitive interference, but also deficits in study habits and test-taking skills.

Behavioral avoidance generated in part by performance-avoidance goals plays a key role in the maintenance of evaluative anxiety and concomitant skill degradation. Evaluative anxiety leads to procrastination, motivated by fear of failure in learning specific subject matter or the aversiveness of the test situation or material. Procrastination, such as failure to complete homework assignments or study for the test, leads to failure to acquire the knowledge required. In turn, this lack of preparation leads to poor performance and anxiety in the test situation (Naveh-Benjamin, 1991), increasing subsequent test anxiety and avoidance of study.

Studies also identify moderator variables that accentuate or reduce deficits in performance. For example, negative feedback appears to be especially detrimental to anxious students, whereas providing reassurance and social support may eliminate the deficit. However, there have been sufficient instances of nonconfirmation of predicted deficits to suggest that high anxiety does not

automatically generate lower achievement outcomes. Generally, anxiety is more detrimental to attentionally demanding tasks, and may even facilitate performance on easy tasks. There may also be more subtle effects related to the qualitative nature of the task.

Interventions

A bewildering array of anxiety-treatment programs has been developed and evaluated over the past three decades. Current attempts to reduce debilitating levels of anxiety and enhance scholastic performance have typically focused either on treatments directed toward the emotional (affective) or cognitive (worry) facets of evaluative anxiety.

The emotionally oriented therapies primarily aim at reducing the arousal and heightened emotional reactions of anxious persons when faced with stressful evaluative situations. Based on the assumption that anxiety comprises a physiological component, attempts to alleviate anxiety symptoms should prove successful, in part, if they focus on reducing levels of arousal or on altering ways in which people appraise their arousal in evaluative situations.

In general, these emotion-focused treatments rely on key behavioral learning principles (counterconditioning, reciprocal inhibition, extinction, observational and coping skill learning, etc.) They also draw from an arsenal of behavioral techniques, such as deep muscle relaxation, guided imagery, and graduated hierarchies. For example, relaxation and guided imagery are not unique to a particular behavioral intervention method, but are employed in several methods, including relaxation as self-control, systematic desensitization, and anxiety management training. Procedures designed to reduce emotionality, while clearly useful in modifying subjectively experienced anxiety, by these methods, appear to have little effect on cognitive performance. Overall, emotion-focused treatments appear to be relatively ineffective in reducing evaluative anxiety unless these treatments contain cognitive elements.

Recent years have witnessed a proliferation of cognitively oriented intervention programs that emphasize the mediating role of cognitive processes in sustaining or eliminating anxiety. Cognitive therapy is a generic term that refers to a wide array of therapeutic approaches directed toward modifying the worry and irrational thought patterns of anxious clients. Broadly speaking, cognitively oriented approaches to anxiety intervention are quite similar in assuming that cognitive processes are determining factors in anxiety, although they differ in terms of actual intervention procedures. A fundamental assumption shared by contemporary cognitive models of test anxiety is that cognitive processes mediate the person's emotional and behavioral responses to stressful evaluative situations. It follows that to modify the negative emotional reactions of anxious clients to evaluative

situations, therapy needs to be directed at reshaping the faulty premises, assumptions, and negative attitudes underlying maladaptive cognitions of anxious subjects. A brief summary of key emotion-focused, cognitive-focused, and skill-focused treatment techniques and methods, and their reported effectiveness, is presented in **Table 3**.

The choice of which therapy to use will be influenced not only by the diagnosis of the specific nature of the client's problem and type of test anxiety, but also by the broader diagnostic picture, the immediate and long-term goals of treatment, and the therapeutic orientation adopted. For example: although relaxation may not increase the performance of test-anxious students with study-skill deficits, it may be prescribed by the school psychologist in order to help the student achieve the immediate goal of achieving control over test anxiety – as a first step toward academic problem solving. Thus, once the anxiety that interferes with learning new study skills is removed, the following step would be training the student in efficient study skills. Furthermore, there are different ways that a therapist may view his/her students' problem (distorted thinking styles, poor problem-solving skills, etc.) In addition, each of these views may give rise to different treatment procedures.

Summary and Conclusions

Anxiety is one of the most ubiquitous and researched emotions in education. Anxiety is a multifaceted construct, involving cognitive, affective, and behavioral components. Although different forms of anxiety discussed above are distinguished by the antecedent conditions and contexts evoking the anxiety (e.g., tests, and mathematics/computers), they have important structural similarities (worry and arousal) and are governed by similar cognitive and motivational processes (apprehension of being evaluated and fear of not meeting standards).

The nature of the anxiety–performance relationship is best viewed as reciprocal in nature. Thus, high levels of anxiety, accompanied by elevated levels of worry and cognitive interference, absorb part of the capacity needed for attention, working memory, problem solving, or other cognitive processes required for successful completion of a task. Evaluative anxiety also produces certain aversive patterns of motivation, coping, and task strategies that interfere with learning and performance. The result is that competence and self-efficacy suffers, thus leading to further anxiety over time and generating a vicious circle of increasing anxiety and degrading competence.

Overall, the assessment of anxiety in educational settings has not kept pace with the theoretical advances in conceptualizing the construct. Thus, much of the construct domain (e.g., task-irrelevant thinking, off-task thoughts, and poor academic self-concept) is underrepresented in current measures of anxiety. Stressful situations

Table 3 Some focal emotion-focused and cognitive-focused anxiety intervention techniques

<i>Treatment</i>	<i>Description</i>	<i>Effectiveness</i>
<i>I. Emotion-focused interventions</i>		
Biofeedback	Use of instrumentation (e.g., a physiograph) to provide a person with immediate and continuous information about one or more physiological processes (e.g., skin conductance, temperature, heart rate, blood volume pulse, respiration, and electromyograph). Biofeedback teaches highly test-anxious persons to monitor and modify the physiological processes associated with their emotional reactions.	A large body of literature supports the notion of increased physiological control when using physiological feedback and self-regulation. However, biofeedback alone is not effective in reducing anxiety (nor does the addition of biofeedback training improve the efficacy of other forms of treatment). Given the potential cost and inconvenience of using biofeedback training, it may not be the treatment of choice for anxiety intervention.
Relaxation training	Recommended on the premise that maintaining a relaxed state, via deep breathing and muscle relaxation exercises, would counteract a person's aroused state. Presumably, if a person knows when and how to apply relaxation, it will be applied directly as a counterresponse to anxiety.	Meta-analytic research tends to support the effectiveness of relaxation therapy. However, the effects on performance tend to be negligible.
Systematic desensitization	Situation-specific anxiety is viewed as a classically conditioned emotional reaction resulting from a person's aversive experiences in aversive situations. Systematic desensitization proposes that anxiety reactions to threatening situations may also be unlearned through specific counter-conditioning procedures. The anxious client is typically trained in a deep muscle relaxation procedure and, while relaxed, instructed to visualize an ordered series of increasingly stressful scenes (an anxiety hierarchy). The client imaginably proceeds up the hierarchy until he/she is able to visualize the most stressful scenes on the list without experiencing anxiety. Through repeated pairings of imaginal representations of threatening evaluative situations with deep relaxation, the bond between the threatening evaluative scenes and anxiety is expected to be weakened.	Meta-analytic data lend support to the effectiveness of systematic desensitization in reducing anxiety, particularly test anxiety, in school children and college students. It is shown to be as, if not more, effective in reducing test anxiety than a variety of other treatments, including relaxation training, hypnosis, and skills training. However, systematic desensitization fares less well when cognitive performance (e.g., academic achievement) is the criterion or when outcome is being assessed.
Anxiety management	Teaches highly anxious subjects to recognize their situation-specific related arousal responses as they are building, and then to use them as cues for initiating the coping response of relaxation in threatening situations.	A body of research supports the effectiveness of this technique in reducing anxiety. Thus, anxiety management training appears to be as, if not more, robust and effective than related interventions. Reductions in debilitating anxiety were maintained for follow-up periods ranging several weeks to months.
Modeling	Involves the live or symbolic (e.g., through videotape) demonstration of desired coping behaviors in a stressful situation such that they can be subsequently imitated by the anxious person. It is assumed that exposure to models displaying adaptive behavior may play a positive role in facilitating performance. Clients are instructed to vividly imagine the stressful evaluative scene and focus on the anxiety and associated response-produced cues (e.g., racing heart, neck and shoulder tensing, dryness of the mouth, and catastrophic thoughts). Clients are then trained to use these cues to prompt adaptive coping skills to actively relax away tension, and reduce anxiety before it mounts too severely.	A body of research lends support to the effectiveness of modeling in treating anxiety. In particular, exposure to models who are task oriented and provide attention-directing cognitive structuring clues is beneficial to the performance of anxious persons. Of additional benefit is evidence in the behavior of the model that he/she is successfully coping with the worry and tension associated with anxiety.

Continued

Table 3 Continued

<i>Treatment</i>	<i>Description</i>	<i>Effectiveness</i>
<i>II. Cognitive-focused interventions</i>		
Cognitive-attentional training	Cognitive attentional training provides specific training in the redirection of attention to task-focused thinking and emphasizes the inhibition of task-irrelevant thinking and nonproductive worry. The cognitive attentional approach relates performance decrements to the diversion of attention to self-focused thinking, coupled with the cognitive overload caused by the worry component of anxiety. By redirecting attention to the task and reducing worry and task-irrelevant thinking, cognitive resources are freed and, when redirected to the task, performance is improved. Attentional training programs traditionally provide clients with instructions to attend fully to the task and to inhibit self-relevant thinking while working on a variety of academic tasks.	The beneficial effects of attentional instruction on the anxiety and cognitive performance of highly anxious students is supported by some empirical research. Task instructions that provide examinees with information about appropriate problem-solving strategies, and away from self-preoccupied worry, may be particularly helpful to the anxious individual's cognitive functioning.
Cognitive restructuring	The rationale is that anxious persons will be able to master their anxiety by learning to control task-irrelevant cognitions that generate their anxiety and direct attention from their task-directed performance. The two most prominent cognitive therapeutic methods in test anxiety intervention are Rational Emotive Therapy and Systematic Rational Restructuring. Both forms of treatments are based on the premise that anxiety or emotional disturbance is a result of illogical or irrational thinking. Two key irrational beliefs that maintain anxiety are that one must succeed at all costs, and that success is equivalent to self-worth. Anxious individuals are taught how to recognize, vigorously challenge, question, and dispute their irrational beliefs, and replace their maladaptive internal dialog with more rational structures and beliefs. Presumably, by modifying irrational beliefs and schemas, negative emotional reactions will be reduced, and performance improved. Systematic Rational Restructuring aims at helping test-anxious clients to discover the worrisome task-irrelevant thoughts they entertain, to eclipse such thoughts, and to substitute positive self-statements that redirect their attention to the task at hand.	Research indicates that whereas cognitive restructuring reduces anxiety, there is no concomitant improvement in performance. A number of studies provide evidence showing that these techniques may be effective in reducing anxiety. However, concomitant improvements in cognitive performance are observed with far less consistency.
Cognitive Behavioral Modification	A multifaceted program merging both cognitively-focused and emotionally focused techniques (as well as skill training in many cases), thus offering the test-anxious client the best of many worlds, so to speak. This multimodal treatment attempts to deal with the multiple manifestations of anxiety, including negative motivational or affective tendencies, irrational thought patterns, and skills deficits, and emphasizes the application, and transferring of acquired coping skills to <i>in vivo</i> test situations. Given its dual emphasis on modifying both emotional processes and irrational thoughts and cognitions, this results in a powerful approach that merges emotionally oriented and cognitively oriented techniques to alleviate clients' anxiety and enhance their performance. This procedure is based on the premise that reducing a person's level of anxiety involves both anxiety-reduction training as well as detailed cognitive restructuring of certain faulty beliefs or misconceptions.	Multimodal treatment packages, such as cognitive-behavior modification, are most likely to be effective by their support for the inclusion of multiple domains related to anxiety. These procedures are relatively effective in reducing self-reported levels of debilitating anxiety, and are equally effective, more or less, in reducing both cognitive and affective components of anxiety. These procedures increase test performance, on average, by about half a standard deviation in school-aged samples, and elevate grade point average by close to three-quarters of a standard deviation.

would typically have effects on various response systems (i.e., verbal, physiological, and cognitive performance), and each measurement method possesses unique functions in anxiety assessment and is characterized by specific and unique limitations. It is desirable to obtain measures from a number of systems and triangulate any observed effects by means of converging operations.

Evaluative anxiety is more than a combination of physiological arousal, negative self-preoccupation, deficit in stress-related coping skills, and poor study habits. It is the complex interaction among these diverse components that seems to define anxiety. As the cognitive, affective, and behavioral components of anxiety interweave in contributing to the problem of evaluative anxiety and its treatment, it is predicted that an induced change in one system would generally be followed by a change in the other. Therapeutic approaches, which emphasize cognition, often extend to the emotional life too, and vice versa. For example, it is likely that emotion-focused training (e.g., progressive relaxation) may make the client less anxious and result in a decrease in anxiety-focused, task-irrelevant ideation. Similarly, some forms of cognitive therapy may provide anxious subjects with an increased sense of perceived control, which might spill over into the emotional domain and result in lower emotional arousal in an evaluative situation.

Anxiety assessments need to be understood within the context of a student's life and social milieu. Thus, understanding the results of a score on an anxiety measure requires an appreciation of the possible multiple and interactional influences on anxiety. These include the subject's past affective and academic history, and current social, emotional, vocational, and economic adjustments, as well as behavior during the examination. When a life history (no reported math anxiety in the past) is in disagreement with the results of a math anxiety scale, it is best to pause before making a diagnosis or decision on the basis of the anxiety scale alone, as the former is generally a more reliable criterion.

It is now readily apparent that interventions should be based on a careful theoretical analysis of the nature of evaluative anxiety and its key components and manifestations. Traditionally; however, interventions have mainly evolved from interest in specific behavioral treatment techniques rather than from an analysis of the nature and effects of anxiety. Indeed, most investigators who have applied behavioral methodology to the reduction of evaluative anxiety have generally paid little attention to relating the treatment process to important theoretical conceptions. The current diversity of treatments, while supplying the clinician with a rich variety of treatment

options to choose from in rendering services, also reflects a state of uncertainty marked by the lack of consensus regarding the most effective method for treating anxiety in educational settings.

Bibliography

- Alpert, R. and Haber, R. N. (1960). Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology* **61**, 207–215.
- Anton, W. D. and Klisch, M. C. (1995). Perspectives on mathematics anxiety and test anxiety. In Spielberger, C. D. and Vagg, P. R. (eds.) *Test Anxiety: Theory, Assessment, and Treatment. Series in Clinical and Community Psychology*, pp 93–106. Philadelphia, PA: Taylor and Francis.
- Eysenck, M. W. (1992). *Anxiety: The Cognitive Perspective*. Hove, UK: Erlbaum.
- Hembree, R. (1988). Correlates, causes and treatment of test anxiety. *Review of Educational Research* **58**, 47–77.
- Lazarus, R. S. (1999). *Stress and Emotion: A New Synthesis*. New York: Springer.
- Lazarus, R. S. and Folkman, S. (1984). *Stress, Appraisal, and Coping*. New York: Springer.
- Liebert, R. M. and Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports* **20**, 975–978.
- Naveh-Benjamin, M. (1991). A comparison of training programs intended for different types of test-anxious students: Further support for an information-processing model. *Journal of Educational Psychology* **83**, 134–139.
- Rachman, S. (2004). *Anxiety*, 2nd edn. Oxford: Blackwell.
- Schutz, P. and Pekrun, R. (eds.) (2007). *Emotions in Education*. Beverly Hills, CA: Sage.
- Spielberger, C. D. (1972). *Anxiety: Current Trends in Theory and Research: I*. Oxford, UK: Academic Press.
- Spielberger, C. D. (1980). *Test Anxiety Inventory: Preliminary Professional Manual*. Palo Alto, CA: Consulting Psychologists Press.
- Zeidner, M. (1998). *Test Anxiety: The State of the Art*. New York: Plenum.
- Zeidner, M. and Matthews, G. (2005). Evaluation anxiety: Current theory and research. In Elliot, A. J. and Dweck, C. S. (eds.) *Handbook of Competence and Motivation*, pp 141–163. New York: Guilford.

Further Reading

- Covington, M. V. (1992). *Making the Grade*. New York: Cambridge University Press.
- Phillips, B. N. (1978). *School Stress and Anxiety: Theory, Research and Intervention*. New York: Human Sciences Press.
- Sarason, I. G. (ed.) (1980). *Test Anxiety: Theory, Research and Applications*. Hillsdale, NJ: Erlbaum.
- Sieber, J. E., O'Neil, Jr. H. F., and Tobias, S. (eds.) (1977). *Anxiety, Learning, and Instruction*. Hillsdale, NJ: Erlbaum.
- Spielberger, C. D. and Vagg, P. R. (eds.) (1995). *Test Anxiety: Theory, Assessment, and Treatment*. Washington, DC: Taylor and Francis.
- Zeidner, M. (1995). Adaptive coping with test situations: A review of the literature. *Educational Psychologist* **30**, 123–133.
- Zeidner, M. (2007). Test anxiety: Conceptions, findings, conclusions. In Schutz, P. and Pekrun, R. (eds.) *Emotions in Education*, pp 165–184. Beverly Hills, CA: Sage.

Attribution Theory

B Weiner, University of California, Los Angeles, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

An amalgam of assumptions, beliefs, data, hypotheses, and theories related to phenomenal causality fall under the rubric of attribution theory. Hence, attribution theory refers to a field of inquiry rather than to a specific scientific conception. Inasmuch as the focus of study concerns inferences about the causes of events and outcomes, including achievement-related success and failure, and the consequences of these beliefs, the field of attribution is of central concern to educators. Indeed, a great deal of life in the classroom can be examined from an attribution perspective.

The origin of attribution thinking for psychology was Fritz Heider and his seminal book: *The Psychology of Interpersonal Relationships* (Heider, 1958). However, there had been a traditional philosophic interest in the study of causal reasoning and even prior to Heider a number of empirical studies were concerned with perceived causality (e.g., Michotte, 1946). Heider's contribution was relatively neglected prior to the influential attribution analyses offered by Jones and Davis (1965) and Kelley (1967). In addition to these contributions, the book *Attribution: Perceiving the Causes of Behavior* (Jones *et al.*, 1972) was also responsible for ushering in nearly two decades (1970–90) of attribution dominance in the field of social psychology. Attribution theory, often associated with naive psychology, replaced in impact the noncommon sense theory of cognitive dissonance, and subsequently was supplanted within social psychology by a focus on unconscious processes and priming. During and following its ascendance, attributional thinking penetrated other fields of study as well, including educational psychology.

This article presents aspects of attribution thinking that are of relevance to education, confining this discussion to the perceived causes of success and failure in achievement-related contexts. In so doing, a great deal of attribution-related literature is omitted, particularly from cognitive psychology pertaining to causal reasoning, from clinical psychology related to coping and mental health, and from social psychology concerning interpersonal attitudes and behavior.

The article proceeds as follows. Initially, the perceived causes of success and failure and their underlying properties or characteristics are considered. This is followed by a presentation of some antecedents of causal beliefs, and then a more extensive examination of their cognitive, emotional, and behavioral consequences.

The Perceived Causes of Success and Failure

Explanations may be based on reasons or causes, which need to be distinguished, although this differentiation is at times murky and fraught with philosophical intricacies (see Buss, 1978). If a student is asked why he or she enrolled in a particular course, he or she might call forth reasons such as: "I need it for my major" "I heard the teacher is great" "It meets at the perfect time;" and so forth. These explanations or justifications make the choice understandable and intelligible. In explaining everyday action, people typically call forth current reasons, which are associated with incentives (costs and benefits) and volitional choice (see Malle, 2004).

Attribution theory, on the other hand, is centered on causes. Causes are invoked to explain outcomes or end results, such as success and failure, rather than actions; they are antecedents instead of (in addition to) justifications and can apply to intended or unintended outcomes and to factors that may or may not be controllable. Failing in mathematics because of perceived poor aptitude is considered here to be an attribution, or a causal antecedent, rather than a reason, while enrolling in a class to be with a friend is a reason (incentive) and is not regarded a cause.

It must also be noted that not all outcomes elicit a search for understanding and causality, for this requires cognitive work. For example, if one succeeds at a simple puzzle in the newspaper, it is unlikely that time is taken to answer the question: "Why did I succeed?" On the other hand, unexpected failure at an important task is particularly likely to give rise to a search to identify the cause (see Gendolla and Kolle, 2001). Attribution analyses thus are subsumed within cognitive functionalism in that future successful actions often depend on perceiving the causes of past failures.

What, then, are the causal ascriptions for success and failure? First, it must be remembered that attribution theory is concerned with phenomenal causality, rather than seeking the true causes. Perceived causes vary as a function of situational context. For example, the perceived causes of success and failure at sports (e.g., strength, a windy day) differ from the causes of success and failure at math (e.g., math aptitude). Math outcomes may even elicit different dominant causal beliefs (e.g., aptitude) than do performance results at a history quiz (memorization effort). Causal beliefs also vary between age groups,

cultures, and depend on whether the causal target is the self or someone else.

The potential causes of achievement-related outcomes therefore are numerous, diverse, and often idiosyncratic – one must be wary of generalizations. Nonetheless, there is a set of causal beliefs that appears in many contexts. The most common causes of success and failure are aptitude, ability (or a learned skill), immediate and long-term effort, task characteristics (such as ease or difficulty), intrinsic motivation, teacher characteristics (such as competence), mood, and luck (see Weiner, 1985, 1986). Furthermore, within this delimited list, aptitude (which is presumed here to include ability) and effort dominate causal beliefs. Hence, one succeeded because one is smart and/or tried hard, and failed due to being incompetent and/or not exerting effort. At the very heart of the contribution of attribution theory to education is a conceptual analysis of the distinction between ability and effort and the contrasting linkages of these two causal beliefs to other cognitions, emotions, and actions.

Causal Properties

Inasmuch as causes are diverse and numerous, there have been attempts to identify their underlying properties or characteristics. For example, although it is evident that ability and effort are not the same, in what ways do they differ? Progress from description to classification allows causes to be compared and contrasted quantitatively, rather than merely indicating that they differ qualitatively.

Certainly three, perhaps four, causal dimensions have been isolated (see Weiner, 1985, 1986). One characteristic of causes is their locus, or location, within or outside of the actor. This causal property is most associated with Julian Rotter (1966), the originator of the concept referred to as locus of control. The locus label grew from research on skill-versus chance-determined tasks and is most associated with a scale used to classify individuals according to their beliefs regarding personal causality. Aptitude and effort are similar in locus, both being internal to the actor (although aptitude is likely to be regarded as more internal than effort), and differ from causes such as chance or task ease, which are regarded as located in the environment, or external to the actor.

A second property on which causes can be compared and contrasted is controllability. Aptitude, which is internal to the person, is not subject to volitional control and change, whereas effort expenditure is considered subject to personal control – it could be otherwise.

Endurance or stability is the third known property of causes. Some causes, such as aptitude, are viewed as stable over time, whereas others, such as chance, are not

enduring. Effort also is typically considered unstable, although recall attribution theory deals with phenomenal causality so that labeling another lazy, or industrious, implies stability in effort expenditure.

Yet, a fourth causal characteristic relates to the globality of the cause, or the extent to which the cause generalizes across situations. For example, it could be contended that perceived general intelligence exerts an influence over a wide array of academic outcomes, whereas perceptual reasoning ability has more limited application.

All causes, then, have multiple properties and can be classified within a taxonomic system. Aptitude, for example, is internal to the actor, stable, uncontrollable, and often considered global, whereas effort also is internal to the actor, but likely believed to be unstable, controllable, and specific. In a similar manner, chance tends to be perceived as external to the actor, unstable, uncontrollable, and specific. These groupings are important because causes that differ qualitatively (e.g., aptitude and chance) may share some consequences but differ on others. For example, failing in math because of perceived lack of aptitude or due to bad luck both capture the belief that “I could not do anything about it” and give rise to the results of uncontrollable thinking. On the other hand, these causes are linked to different expectations about the likelihood of future success because ability is stable whereas luck is unstable.

Causal Antecedents

Attributional accounts of the determinants of causal beliefs have been derived from an analysis of reasoning processes. The concepts most invoked are covariation, conditional probabilities, causal rules such as necessary and sufficient causality, and the like (see Kelley, 1967). Discussions also often consider prescriptive or correct reasoning, which is contrasted to the flawed conclusions reached by the person on the street. This body of knowledge, related to epistemology, has applicability in educational settings but is not discussed in this context.

More pertinent here is a consideration of some psychologically based antecedents that give rise to causal beliefs, particularly hedonic concerns. In addition, causal determinants that involve communications from teachers to their pupils also are of central importance in educational contexts. Finally, in the following section I consider impression management techniques that attempt to alter the causal beliefs of others.

Biases in Causal Reasoning

It has been repeatedly documented that individuals overestimate themselves on virtually all positive characteristics.

Just as 75% of the mothers believe that their children are in the upper 25% of the class, individuals regard themselves as smarter, more moral, better drivers, having more common sense, and so on, than others and compared to what they really are (see Dunning *et al.*, 2004). A similar conclusion can be reached regarding causal reasoning: individuals relatively believe that they caused their own success, whereas failure was due to outside forces. This phenomenon has been given various labels, including the self-serving attribution bias, ego-enhancement, and most prominently, the hedonic bias. It is to be expected, then, that children and their parents blame the teacher or the school for failure, while the teacher holds the child and/or his or her parents responsible. Disagreement and conflict regarding the causes of success and failure are normative.

Other biases have been hypothesized and have been very popular topics for attribution research. One considered determinant of causal beliefs is labeled the actor–observer perspective. It has been suggested that actors attribute their behavior to the situation, whereas observers ascribe an action to a characteristic of the actor. For example, if I hit someone it is because that other person provoked me (situational causality), whereas when someone else acts with aggression it is because that person is hostile. Associated with this discrepancy is a dispositional bias, or the fundamental attribution error, which is the tendency to underestimate the situational influence on the behavior of others and overestimate the importance of traits or perceived characteristics of the person.

These presumed tendencies, however, are now being called into question (see Malle, 2006). One factor undermining these hypotheses is the established hedonic bias. If individuals tend to take personal credit for success, then positive behaviors will not be ascribed by the actor to the situation. On the other hand, in situations of failure, both the hedonic bias and an actor–observer discrepancy go hand in hand, with over-attribution by the actor to the situation, relative to an observer.

Teacher Communications

Teachers may inadvertently communicate causal information to their pupils. This information can be conveyed through expressed emotions, praise and blame, and help giving.

As will be elaborated later, emotions are linked with causal beliefs. Anger following the failure of another tends to be elicited by controllable causes, such as lack of effort, whereas uncontrollable causality, such as lack of aptitude, elicits sympathy. Compare, for example, parental reactions to the failure of their child due to lack of studying as opposed to being caused by a mental handicap. Teachers who express anger following failure convey to the pupils that the failure was their fault and is changeable, whereas heightened sympathy, and particularly pity, conveys “it is

not your fault and nothing could have been done.” Thus, a positive or pro-social emotional expression could have negative behavioral consequences (see Graham, 1990).

In a similar manner, praise for success at an easy task, and the absence of reprimand for failure at such tasks, are cues that one has low ability. Furthermore, uncalled-for help may be a low-ability signal from a teacher. In sum, a variety of causal information regarding ability and effort has the teacher as its source.

Impression Management Techniques

It is intuitively evident that individuals do not want others to blame them for failure or untoward actions. To manipulate the causal beliefs of others, disparate impression management techniques are used. Four impression management strategies have been identified: denial, excuse, justification, and confession. In denial, the actor does not acknowledge a particular outcome (e.g., “I did not fail the test”; “I was not smoking”). This is a rather primitive technique, often used by younger children, for it can be readily disconfirmed if untrue.

A more common strategy is to provide an excuse to change causal beliefs (ex = from; cuse = cause). The actor attempts to shift the perceived controllable cause (e.g., lack of effort) to an uncontrollable one (e.g., the bus was late). In so doing, it is anticipated that blame and personal responsibility will be averted. Justification, a third identified impression management strategy, has a similar goal, although in this case allegiance to a higher moral goal is invoked (“I did not study because I had to take my father to the hospital”). Finally, confession of a misdeed is possible. In so doing, the wrongdoer acknowledges a bad behavior; however, confessing puts forth the impression that he or she is a good person who will not engage in the act in the future (“I did not study for the test; I am sorry and it will not happen again”).

Teachers and their pupils at times disagree on the acceptableness of various excuses and justifications. For example, pupils believe that their alarm failing to ring is a reasonable cause for missing an appointment, whereas teachers do not regard this as acceptable to promote forgiveness and lack of anger (see review in Weiner, 1995). It also is the case that some apparent excuses are truthful. However, an observer cannot conclude with any accuracy whether a student in fact missed the test because the bus was late, or had the flu, as opposed to being negligent.

Causal Consequences

Of particular relevance to educators are the consequences of causal thinking on the expectations, emotions, and behavior of pupils. These are addressed in turn.

Expectancy of Success

Motivation theorists have established that behavior is in part guided by future expectancy of success. Regardless of the intensity of a need, if the expectancy of goal attainment is low, then instrumental action directed toward that goal will not be undertaken. For example, regardless of how hungry one is, a restaurant known to be closed will not be a target of choice.

To a great extent, expectancy of future success in achievement contexts is determined by causal beliefs about the cause of prior failure. Stable or enduring causes give rise to the conviction that the future will be no different than the past. Hence, attributions of failure to lack of aptitude, an uncaring school environment, an absence of support, and so on promote expectancy of future failure. This lowers motivation and results in a tendency to leave that setting. On the other hand, ascription to changing factors provides hope that the future may differ from the past. Attribution to lack of effort, or to a teacher who is leaving, promotes positive expectations of the future inasmuch as the perceived causes of the prior failures are unstable or changing.

In an attribution-based change program designed to keep first-year college students from dropping out, it was merely communicated that grades increase the longer the student remains in school (Wilson and Linville, 1982). That is, over time, there is grade inflation. There was no attempt to change the study habits, attendance, note taking, or any other instrumental actions of the pupils. The data suggest the effectiveness of this program in improving school performance and decreasing dropout rates, presumably because expectation of success had increased.

Emotions

The appraisal approach to the understanding of emotion is based on the simple assumption that feelings are determined by thoughts, that is, thoughts are necessary and sufficient determinants of affective states. For example, negative self-related beliefs such as “I am a bad person” or “I am incompetent” are sufficient to promote negative feeling states and decrease personal esteem.

Causal beliefs have far-reaching emotional consequences, both in regard to the self and feelings about others. To understand the emotional lives of school children, it is essential to invoke causal ascriptions. Here, I present an overview of the implications of causal thinking for some feeling states (see Hareli and Weiner, 2002).

It has already been indicated that hope for the future and, similarly, the opposing feeling of hopelessness are guided by the perceived stability of a prior cause of failure. The other causal properties also have affective implications. Attribution of success to the self gives rise to pride in accomplishment and increments in self-esteem, whereas this is not the case for external causal

beliefs. Hence, ascriptions of success to ability or effort both promote pride and positive views of the self, whereas attribution to luck, or to help from others, does not contribute to these self-directed affects. Indeed, attribution of success to others fosters gratitude (if the help provided by the other was volitional). One consequence of the simultaneous influences of the hedonic bias as well as the desire to manage the thoughts of others is that public expressions of gratitude exceed private beliefs about the deservedness of this expression.

Perceived causal controllability, the third property of causes, has a complex relation to feeling states, with the linked affects dependent on the target of the emotion. Ascription of failure to a controllable cause such as lack of effort, given a desire to reach the goal, elicits guilt and regret. On the other hand, lack of aptitude as the cause of failure promotes humiliation and shame, which also involve social comparison with competent others.

These same attributions for the behavior of others, however, elicit a different set of emotions. If another failed due to controllable causes, then anger tends to be elicited. On the other hand, nonattainment of a goal due to uncontrollable causes tends to promote sympathy. Self- and other-related emotions are entangled within a complex interconnected system – communicated anger, if accepted, implies a controllable cause, which should promote guilt, whereas communicated pity implicates uncontrollable causality and promotes feelings of humiliation.

A number of other achievement-related affects also depend on causal thinking. Envy, for example, tends to be elicited by internal and uncontrollable causality for the success of others. One is envious of another’s intellectual capabilities, or beauty, but not of their hard work, since one can also work hard. Admiration, on the other hand, can be evoked by the effort as well as the ability of successful others.

Affect-related personality inferences also are linked to perceptions of the causal role of effort and ability. For example, if the success of another is perceived as due to ability, while the other communicates effort or external factors as the cause, then modesty is inferred. Modesty is a desirable inference associated with liking. On the other hand, if ability is conveyed as the cause of success, then arrogance is inferred, even if it is the true cause. If Einstein says “I am an Einstein,” he nonetheless will be regarded as arrogant, which is an unfavorable trait. Thus, school-based competence has some negative consequences in that others may be envious and perceive the capable person as arrogant. This is perhaps responsible for pejorative labels, such as nerds and geeks, which convey negative attitudes toward competence.

Behavior

Expectancy of future success, along with affective states, guides future action. That is, goal anticipations and

feelings bridge the gap between causal thinking and behavior. For example, assume that a student failed in an exam and ascribes that failure to lack of ability. Inasmuch as ability is a stable cause, expectancy of future success is low; and being an internal and uncontrollable cause, self-esteem decreases while shame and humiliation are experienced. Low expectancy of success, low self-esteem, and humiliation are motivational inhibitors that contribute to a decision to leave the setting.

On the other hand, lack of effort is an unstable, controllable cause, so that attribution to this factor maintains hope and positive anticipations, while arousing guilt, which is a motivator to make amends for the past. Thus, an attribution of failure to lack of effort tends to be adaptive and functional.

Along with the attributions, emotions, and behaviors of actors, this same sequence also describes observers of the actor. An observer making attributions of failure of another to lack of ability tends to experience sympathy, which promotes help-giving. On the other hand, ascriptions to lack of effort evoke anger and an antisocial reaction (see Weiner, 1995, 2006).

As the attributions of ability versus effort are linked with disparate reactions, attribution-change programs have focused on having students make insufficient effort rather than low-ability ascriptions in situations of failure (see Perry *et al.*, 1993). These programs vary but often provide models in film clips portraying adaptive attribution patterns. The attribution interventions have proved successful and provide promising tools to improve school-related performance.

Some Concluding Comments

Attribution theory was developed and nurtured by social psychologists. However, the centrality of thoughts about causality in everyday life transported this body of knowledge to other fields of study. For education, attributional analyses provided a ready fit inasmuch as parents, pupils, and teachers focus so much attention on evaluation and the causes of good and poor performance.

Two broad questions emerge from this approach: How does one know or determine one's level of ability, the adequacy of effort expenditure, and the like? Thus, existential questions, such as "Who am I?" (as well as, "Who are you?"), are addressed. In this article, three inferential sources were pointed out: biases that enhance the self, communications from teachers, and impression management techniques used by others.

The second extended question asked is: "So what? What difference does it make if success or failure was caused by low ability, as opposed to lack of effort?" These questions point out the functional and motivational significance of causal thinking. Here, the classification of

causes into fundamental properties was introduced. These underlying characteristics were shown to influence expectancy of success, emotional reactions, and personality inferences. Of special note is the array of affects linked to causal thinking: admiration, anger, envy, guilt, gratitude, hope, hopelessness, pity, pride, regret, self-esteem, shame, and sympathy were mentioned, and this is not the full list of causally linked feelings.

Expectancy and affect, in turn, are two key mediators of achievement-related behavior. Thus, increased school motivation and decreased school dropouts can be affected by alteration of perceived causality. In sum, the study of education processes is intimately linked with attributional analyses.

See also: Cognition and Emotion; Mathematics Learning; Motivating Students in Classrooms; Volitional Control of Learning.

Bibliography

- Buss, A. R. (1978). Causes and reasons in attribution theory: A conceptual critique. *Journal of Personality and Social Psychology* **36**, 1311–1321.
- Dunning, D., Heath, C., and Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest* **5**, 69–106.
- Gendolla, G. H. E. and Koller, M. (2001). Surprise and causal search: How are they affected by outcome valence and importance? *Motivation and Emotion* **25**, 327–349.
- Graham, S. (1990). Communicating low ability in the classroom: Bad things good teachers sometimes do. In Graham, S. and Folkes, V. S. (eds.) *Attribution Theory: Applications to Achievement, Mental Health, and Interpersonal Conflict*, pp 17–36. Hillsdale, NJ: Erlbaum.
- Hareli, S. and Weiner, B. (2002). Social emotions and personality inferences: A scaffold for a new direction in the study of achievement motivation. *Educational Psychologist* **37**, 183–193.
- Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.
- Jones, E. E. and Davis, K. E. (1965). From acts to dispositions: The attribution process in person perception. In Berkowitz, L. (ed.) *Advances in Experimental Social Psychology*, vol. 2, pp 219–266. New York: Academic Press.
- Jones, E. E., Kanouse, D. E., Kelley, H. H., *et al.* (eds.) (1972). *Attribution: Perceiving the Causes of Behavior*. Morristown, NJ: General Learning Press.
- Kelley, H. H. (1967). Attribution theory in social psychology. In Levine, D. (ed.) *Nebraska Symposium on Motivation*, vol. 15, pp 192–240. Lincoln, NE: University of Nebraska Press.
- Malle, B. F. (2004). *How the Mind Explains Behavior: Folk Explanations, Meaning, and Social Interaction*. Cambridge, MA: MIT Press.
- Malle, B. F. (2006). The actor–observer asymmetry in attribution: A (surprising) meta analysis. *Psychological Bulletin* **132**, 895–919.
- Michotte, A. (1946). *The Perception of Causality*. New York: Basic Books.
- Perry, R. P., Hechter, F. J., Menec, V. H., and Weinberg, L. E. (1993). Enhancing achievement motivation and performance in college students: An attributional retraining perspective. *Research in Higher Education* **34**, 687–723.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs* **80**(1), 1–28.

- Weiner, B. (1985). An attributional theory of achievement-related emotion and motivation. *Psychological Review* **29**, 548–573.
- Weiner, B. (1986). *An Attributional Theory of Motivation and Emotion*. New York: Springer.
- Weiner, B. (1995). *Judgments of Responsibility: A Foundation for a Theory of Social Conduct*. New York: Guilford.
- Weiner, B. (2006). *Social Motivation, Justice, and the Moral Emotions*. Mahwah, NJ: Erlbaum.
- Wilson, T. D. and Linville, P. W. (1982). Improving the academic performance of college freshmen: Attribution theory revisited. *Journal of Personality and Social Psychology* **42**, 367–376.

Children's Friendship

S R Asher, Duke University, Durham, NC, USA

K L McDonald, University of Maryland, College Park, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

Friendship is a close dyadic relationship characterized by a shared history, mutual affection, and a recognition by both participants that the relationship has a special status, which sets it apart from mere acquaintanceship or colleagueship. Friendship carries with it certain expectations, such as companionship, cooperation, instrumental help, loyalty (reliable alliance), confiding, and emotional support. Furthermore, friendships tend to be characterized by relatively equal power between relationship partners, especially compared to the parent-child relationship. For most people, a close friendship is their first nonfamilial intimate relationship that is freely chosen. A comparison between friendships and sibling relationships is also instructive. When siblings argue or have a falling out, they continue to be siblings even if they do not engage in relationship repair. On the other hand, as friendships are typically voluntary, friends need to work at maintaining the relationship or they risk the friend withdrawing from the relationship.

Although being a friend inevitably carries expectations, these expectations differ somewhat as a function of age. When interviewed or surveyed, young children report friendship expectations that center on play and companionship. In middle childhood, there is an increased emphasis on sharing, cooperation, and helpfulness, and in adolescence young people report greater expectations for intimacy, shared confidences, and emotional support. However, interviews and surveys may underestimate the richness of even young children's friendships. When children as young as 3 years of age are observed playing with their best friend, many examples arise of children cooperating with one another, providing emotional support, and sharing confidences. The fact that some of these processes occur within the context of fantasy play adds to their charm but does not detract from the evidence that even young children can display more developmentally advanced friendship processes.

Characteristics of friendships may differ to some degree as a function of gender. Girls and boys do not differ in their reports of how much companionship or conflict they have in their friendships; however, girls do report more help and guidance, emotional support, intimacy, and ease of conflict resolution. Yet boys and girls do not differ in the number of close friends they have in their class, their satisfaction with their friendships, or their loneliness in school, suggesting that we need to more extensively observe the friendships of boys, in order to better understand interaction processes in their friendships.

Characteristics of friendships also vary across cultures. In many cases, friendships are voluntary relationships that children choose to form and can terminate if they wish. However, in some cultures, friendships may become formalized relationships that are unbreakable, such as ceremonial friendships or 'blood brotherhoods.' In addition, the expectations held of friends and the importance placed on certain features of friendships can vary widely from culture to culture. For example, Indonesian children describe instrumental aid as a more central feature of their friendships than do American children, whereas American children describe enhancement of worth, companionship, and reliable alliance as being more important in their friendships than do Indonesian children.

Measuring Friendship in School and Distinguishing Friendship from Peer Acceptance

There is a long tradition extending back to the late 1920s of using sociometric measures, or measures of interpersonal attraction, to identify classroom friendships. Typically children are asked to name their best friends in their class or grade. Researchers can ask children for a limited number of nominations (e.g., their best three friends) or can use an unlimited nomination measure in which children can name all their friends in their class or in their grade. Several improvements have taken place in sociometric friendship assessment over the years. First, children are typically given a roster with names of schoolmates to circle rather than relying on children's ability to recall and write down their classmates' first and last names. Second, for young children, photographs of schoolmates are used instead of a class roster, again to overcome literacy or memory constraints. Third, researchers have distinguished between unilateral and reciprocal nominations, and most typically define a friendship as existing when two children mutually nominate one another as opposed to one child nominating the other but the other failing to reciprocate the nomination. With this methodology, children can be assigned a score based on the number of reciprocated friendship nominations they have.

Note that the number of friends that children have does not speak to the stability or the more qualitative features of their friendships. To measure the stability of a friendship, repeated sociometric surveys are needed to learn whether friendships are maintained. Furthermore, to learn about

the qualitative features of friendships (e.g., levels of companionship, emotional support, or conflict), children can be observed or interviewed about their friendships. Research on stability and qualitative features of friendship indicates that these are extremely important dimensions. For example, children who have more stable friendships and higher-quality friendships are less lonely at school.

Furthermore, it is also important to distinguish children's participation in friendships from their overall acceptance by the peer group. Peer acceptance refers to how well a child is liked by the group as a whole. Peer acceptance is typically measured in one of two ways. One approach is to ask children to nominate three children in their class or grade they like the most and three children they like the least. Researchers then typically calculate a social preference score, which is the number of liked least nominations subtracted from the number of liked most nominations a child receives. This score can be standardized by grade or class and can also be standardized within gender. Alternatively, children can be asked to complete a sociometric rating-scale measure of acceptance in which they rate each of their classmates or grademates on a Likert scale in terms of how much they like each person, or how much they like to play or work with each person. With this measure, acceptance scores are the average ratings received by peers, typically standardized within classroom or grade and gender.

Research using a rating-scale measure of acceptance and mutual friendship nominations to index friendship indicates that children can be well-liked by their peers but not have close friendships at school. Likewise, children can be generally disliked by classmates but still have one or more friends (subgroups of highly aggressive children are likely to be in this situation). Overall, the correlation between the number of mutual friends children have and their acceptance by peers is approximately 0.50, suggesting that there is unique variance associated with each index of peer adjustment. Evidence in support of the friendship-acceptance distinction comes from several lines of inquiry. First, having friends and being liked by peers make distinct contributions to children's early school adjustment. Second, having friends and being accepted by peers independently predict to feelings of loneliness at school and to whether or not children are victimized by peers. Third, children are less likely to form friendships with peers of a different race than they are to like or accept them. Finally, social skills intervention studies have repeatedly found that children can make gains in peer acceptance yet not make gains in their number of good friends.

In the section that follows, the focus is on the effects of friendship on school adjustment; however, several of the supplementary readings may be useful for readers interested in learning more about acceptance and school adjustment.

The Effects of Friendship on School Adjustment

Children's friendships provide validation, emotional support, a context for self-disclosure, companionship, help and guidance, and reliable alliance. These provisions in turn serve protective functions in children's lives at school. Indeed, children with friendship difficulties are at risk for a variety of social-emotional adjustment problems in school, including increased social anxiety, lower self-esteem, and increased likelihood of loneliness and depression. One reason that children lacking friends may feel worse than children with friends is that having friends can protect children from being victimized by peers. Children with friends at school are less likely to be victimized, especially when their friends are strong and protective.

It should be noted, however, that the benefits of friendship may be moderated by friendship quality and stability. Having friendships that are high in positive qualities and that are relatively stable is associated with increases in self-esteem over time for children; however, if friendships are high in negative qualities, this is less likely to be the case. Further, if children are in friendships in which their friend is treating them badly, it can be particularly distressing. When friends interact in negative or hostile ways with one another, children show increases in internalizing problems such as anxiety, social avoidance, and loneliness.

One of the compelling findings about friendship and school adjustment is that children who are not nominated by any of their peers as being a friend are much more likely to drop out of school years later. Undoubtedly, having friends is part of the glue that keeps children connected to school. The academic benefits of friendship are evident as early as in the first years of school. Kindergarten children who enter school with friends or who make friends during the school year are better adjusted to school (on indexes such as attitudes toward school, school anxiety, and absenteeism) at the end of the year than children who do not.

Here, too, the association of friendship with school-adjustment may be moderated by the quality of a child's friendship. For example, kindergartners who report receiving help from a friend increase in their liking for school more than those children who did not receive help from a friend. Further, children who report more positive qualities in their friendships also report being more involved in class and feeling more accepted by peers. Conversely, having friendships that are high in unresolved conflict is predictive of being more disruptive in class and being less involved in school.

The characteristics of a child's friends may also affect the child's academic success and achievement motivation. The academic performance of a child's friends is associated with the later academic performance of the child. For instance, if a child's close friends get higher

grades then the child is more likely to improve, but if a child's close friends perform poorly, then his or her performance is also likely to suffer. This influence may occur through a variety of processes, such as tangible help (or lack thereof) that friends provide each other, discussions about schoolwork, or through increasingly similar beliefs and attributions about academic achievement and the causes of academic success and failure. It is also important to note that even though being friends with children who are high achievers may increase a child's academic performance, this friendship may also increase the child's negative self-evaluations due to social comparison between the self and the friend.

Just as friends influence the achievement orientation of one another, they also influence children's tendencies to be prosocial or antisocial in their behavioral orientations in school. Children who affiliate with antisocial peers are more likely to engage in antisocial behavior themselves. Antisocial children positively reinforce each other for antisocial behavior even in their conversations about previous delinquent behavior. Since high levels of antisocial behavior are predictive of school dropout, friends who engage in 'deviancy training' with one another are, in effect, contributing to one another's eventual withdrawal or suspension from school.

As children move into adolescence, their friendships are embedded in cliques and larger groups called crowds (e.g., athletes, academic nerds, populars, and druggies). Children within cliques and crowds have shared interests in particular activities and also frequently share beliefs about the value of school achievements. Schools, too, can differ markedly in the extent to which the peer group values achievement. Accordingly, cliques, crowds, and overall peer culture at schools make their own independent contribution to school adjustment.

The Social Tasks of Friendship

Since acceptance by peers and participation in friendship are positively correlated, many of the interpersonal skills related to getting along with peers are also related to being able to make and maintain friendships. In general, it benefits children to be prosocial (i.e., friendly, cooperative, kind, and helpful), to be competent academically and athletically, and to have a good sense of humor. Nonetheless, friendship is a rich and complex context that poses specific kinds of challenges for children that go beyond the social tasks involved in maintaining harmonious relationships with the group as a whole. The friendship tasks that children face include forming a friendship (which means moving a relationship beyond an acquaintanceship or colleagueship), managing conflict and responding to friendship transgressions, providing help and reliable alliance, and self-disclosure.

Forming Friendships

At the core of friendship is companionship, recreation, shared activities, and having fun together. Children who are good at generating enjoyable activities, are good at games and sports, knowledgeable about peer culture, have a good sense of humor, and interact with a spirit of fairness are more likely to be enjoyable play partners.

Forming friendships often calls upon children to initiate interactions and a relationship with another child. This task involves being able to approach another child in a positive manner, to be able to initiate conversation, and/or suggest a joint activity. Part of the social task of initiating interaction is coping with the task of group entry. Children in school contexts are faced with the continuing problem of how to get access to a group of children who are playing together. Children who are skillful at taking the perspective of other group members and thereby gently easing their way into the ongoing interaction are generally more successful than children who more egocentrically attempt to plow ahead and disrupt the ongoing activity. Successful initiations of interactions also involve being responsive to others' initiations. Children who respond warmly and positively to others' initiation attempts are more likely to be sought out to play by peers. Even in infancy and toddlerhood there are large individual differences in how children respond to the initiations of others.

In the early childhood years, parents are likely to play a more active role in initiating contact with other children outside of the school context. For example, they may arrange a play date at home or at a local park. Indeed, parents who are skillful planners of their children's social world and engage in indirect, nonintrusive but competent monitoring of children's play, have children who are more successful at making friends. As children grow older, however, initiating play dates becomes more of their responsibility, although parents maintain an important role by providing transportation or enrolling children in programs such as youth sports, church groups, or dance classes. Neighborhood characteristics also play important roles. For example, where heavy concentrations of children live in close proximity, less scaffolding by parents is required. Furthermore, environmental features such as access to neighborhood parks, public transportation, and safe bike routes affect the extent to which children can initiate their own friendship contacts.

It should also be noted that friendship formation between children is facilitated by physical propinquity. When children are asked to name their 'best friends in the world,' a high proportion of the children named are actually in their classroom or neighborhood. Physical propinquity is a powerful force in children's friendships, much the same as in adults' friendships. With regard to similarity, the role of demographic characteristics is not to be underestimated. For example, the great majority of

elementary-school-age children's friendships is with children of the same gender. Likewise, children tend to form friendships with children of their own race and social class. Similar interests and activities (a shared interest in sports, video games, music, etc.) are also bases for friendship formation. Indeed, children whose interests are highly discrepant from the interests of the mainstream peer group may have difficulty forming friendships within that group. In relation to this, one important role that parents play in their children's social lives is fostering interests in their children that help them connect with other children and thereby form friendships.

Beyond initiation, children need to be able to sustain interaction with peers to form friendships. Even in the preschool years, children can be observed engaging in extended play bouts, often in the context of shared fantasy (e.g., playing firefighters or playing house). For play to continue harmoniously as children grow, they need to develop increased communication skills, such as listening well to their partner, communicating in a contingent way (such that their utterances relate in a meaningful and relevant way to those of their partner), and amicably resolving disagreements.

Managing Conflict and Responding to Friendship Transgressions

Children can be successful at initiating friendships but are still not successful at maintaining friendships. Conflicts are virtually inevitable in friendships and friendships often rise or fall based on how well the parties resolve disputes when they arise. Responding to conflict is particularly challenging for some children and there are large individual differences in the goals and the behavioral strategies that children enact in conflict situations. Children who focus on maintaining a good relationship with their partner and who engage in cooperative and compromising strategies are more successful in their friendship than those who focus exclusively on getting their way. Further, children who are vengeful in conflict-of-interest situations with friends have fewer friendships and, if they do have a friend, their friendships are of lower quality. Indeed, children's goals and strategies in conflict situations predict success at friendship even when statistically controlling for children's levels of peer acceptance. Here, again, adults (both as parents and teachers) can play a valuable role by helping to mediate disputes or, better still, by teaching children conflict-resolution strategies.

Just as conflicts are virtually inevitable in friendship, it is not possible for friendships to extend over time without incidents of minor, or even major, friendship transgressions. As discussed above, friendship carries expectations for companionship, cooperation, instrumental help, loyalty, confiding, and emotional support. These expectations, usually strongly held, mean that children are vulnerable to feeling

sad, hurt, or angry when a friend violates a core expectation of friendship. Examples of transgressions include friends canceling play dates, preferring to be with another friend on a particular day, refusing to provide help when help is needed, or sharing a secret with others. When transgressions occur, children make inferences about the meaning behind the action of the friend. A fundamental question is whether the friend meant to cause harm or simply had other needs, goals, or external pressures that were temporarily interfering with the ability to be a good friend. Children who attribute hostile intent or infer that the other person's actions are a sign of lack of caring, rejection, or disregard are far more likely to want to get even or terminate the relationship. As part of this reaction they are also more likely to engage in negative behavior, including physical, verbal, and relational aggression (intended harm through the damage of another person's relationships or social standing).

Friendships are embedded in the larger peer network of cliques and crowds. Knowing how to effectively manage dyadic relationships in this larger peer context may be particularly challenging for some children. For instance, friends have other friends and if children have high expectations of exclusivity in their friendships they may become jealous. Jealousy is a cognitive, emotional, and behavioral reaction set off by the perception that one's partner prefers a relationship with someone else. Young adolescents who are more jealous in their friendships report lower self-worth and more loneliness. The jealous child conceives of friends having other friends as a friendship transgression rather than an expected part of social life.

Providing Help and Reliable Alliance

As children move into the primary school years, there is an increasing expectation that friends will provide help, guidance, and tangible forms of assistance. This can range from small things such as loaning a friend a pencil in class to larger tasks such as helping a friend with homework or a chore. It also involves providing advice and/or emotional support when a friend is coping with familial, school, or peer relationship difficulties.

Help-giving requires special cognitive and social skills. As in the task of group entry, perspective taking plays a key role. In addition, children need effective communication skills to elicit important information before offering advice. They also need a broad knowledge base of effective solutions to problems that can arise in everyday life. Further, when no easy solution is in sight, friends can help by distracting a child by providing reliable and enjoyable companionship.

Friends also help each other by serving as reliable allies. Friends count on each other, stand up for each other in threatening (e.g., bullying) situations, and have each other's best interests at heart. Having both emotional

and instrumental support from friends can increase children's sense of belonging in school, protect children from stress, and foster a sense of security which in turn facilitates the self-disclosure that occurs when friends confide in one another.

Self-Disclosure

Self-disclosure, or the sharing of personal information with another person, is a central feature of many friendships, particularly in adolescence. This is consistent with a major concern that arises at this stage of development, namely self-exploration and the development of identity. Sharing high levels of intimate information can be risky unless the friend can be relied upon to treat information as confidential and not share it with others. Through reciprocated self-disclosure, children form a sense of emotional closeness and acquire extensive knowledge about one another. Part of a shared history of a friendship is not just the activities or events that friends share together but the detailed and intimate knowledge they have of the other person. As a result, friends know far more about one another than do nonfriends.

One interesting complexity about self-disclosure deserves comment. Although friendships that are high in self-disclosure are more likely to be close friendships, co-ruminating with friends, or excessively discussing problems and focusing on the negative feelings associated with those problems, may have negative emotional consequences for children and adolescents, especially girls. To illustrate, although co-rumination is associated with higher-quality friendships for boys and girls, it is also associated with higher levels of anxiety and depression for girls.

Interventions to Promote Friendships

Given the powerful role of friendship in children's social-emotional and academic adjustment, effective intervention methods are needed for helping children who lack friends and for children whose friendships are of lower quality. Interventions to promote friendship have taken various forms although, on the whole, far more research is needed. Evidence exists that placing children in cooperative work groups can lead children to form friendships across racial lines, although the long-term maintenance of newly formed friendships has not been investigated. There is also evidence that children who lack friends can make a friend when put in a school context where they get to interact with classmates during a pleasurable activity (e.g., planning for and putting on a skit for the class). However, in this case, follow-up assessments suggest that the positive friendship outcomes of this type of intervention do not endure once the activity has ended. This evidence, along with other research indicating that children who lack friends are less socially skilled, suggests

that providing opportunities for enjoyable interaction may not be sufficient and that instead some form of social relationship skill learning opportunities is needed.

Interestingly, several social skills training studies carried out with 8- to 10-year-old children in elementary schools have found that teaching certain social relationship concepts leads to significant gains in overall acceptance by peers, but not necessarily gains in best friendship. In these studies, children were taught core concepts of social interaction (e.g., staying involved in the game or activity, cooperating, and being friendly and supportive) and given opportunities to try out the ideas by playing a game or doing an activity with a peer. Following these interactions, the child was given the opportunity to reflect further on the concepts in light of his or her immediately preceding peer interaction experiences. A potential next step for educational researchers could be to learn whether this kind of direct instruction approach for promoting social competence could be successfully adapted to help children develop friendships and to achieve friendships that have positive qualities. Such research is certainly needed given the important role that friendship plays in children's academic success and emotional well-being.

See also: Neighborhoods and Peers in the Production of Schooling; Parent and Family Involvement in the Education of Children with Special Needs; Peer Interaction and Learning; Social and Emotional Outcomes of Learning; Social Competence; Social Development and Schooling; Social Interaction and Learning; Social Networks and the Education of Children and Youth; Wellbeing.

Bibliography

- Altermatt, E. R. and Pomerantz, E. M. (2003). The development of competence-related and motivational beliefs: An investigation of similarity and influence among friends. *Journal of Educational Psychology* **95**, 111–123.
- Altermatt, E. R. and Pomerantz, E. M. (2005). The implications of having high-achieving versus low-achieving friends: A longitudinal analysis. *Social Development* **14**, 61–81.
- Asher, S. R., Parker, J. G., and Walker, D. L. (1996). Distinguishing friendship from acceptance: Implications for intervention and assessment. In Bukowski, W. M., Newcomb, A. F., and Hartup, W. W. (eds.) *The Company They Keep: Friendship in Childhood and Adolescence*, pp 366–405. New York: Cambridge University Press.
- Berndt, T. J. and Keefe, K. (1995). Friends' influence on adolescents' adjustment to school. *Child Development* **66**, 1312–1329.
- Bigelow, B. J. (1977). Children's friendship expectations: A cognitive-developmental study. *Child Development* **48**, 246–253.
- Crick, N. R. and Nelson, D. A. (2002). Relational and physical victimization in friendships: Nobody ever told me there'd be friends like these. *Journal of Abnormal Child Psychology* **30**, 599–607.
- Dishion, T. J., Andrews, D. W., and Crosby, L. (1995). Antisocial boys and their friends in early adolescence: Relationship characteristics, quality, and interactional process. *Child Development* **66**, 139–151.
- French, D. C., Pidada, S., and Victor, A. (2005). Friendships of Indonesian and United States youth. *International Journal of Behavioral Development* **29**, 304–319.

- Hodges, E. V. E., Malone, M. J., and Perry, D. G. (1997). Individual and social risk as interacting determinants of victimization in the peer group. *Developmental Psychology* **33**, 1032–1039.
- Keefe, K. and Berndt, T. J. (1996). Relations of friendship quality to self-esteem in early adolescence. *Journal of Early Adolescence* **16**, 110–129.
- Ladd, G. W. (1990). Having friends, keeping friends, making friends, and being liked by peers in the classroom: Predictors of children's early school adjustment? *Child Development* **61**, 1081–1100.
- Ladd, G. W., Kochenderfer, B. J., and Coleman, C. C. (1996). Friendship quality as a predictor of young children's early school adjustment. *Child Development* **67**, 1103–1118.
- Parker, J. G. and Asher, S. R. (1987). Peer relations and later personal adjustment: Are low-accepted children at risk? *Psychological Bulletin* **102**, 357–389.
- Parker, J. G. and Asher, S. R. (1993). Friendship and friendship quality in middle childhood: Links with peer group acceptance and feelings of loneliness and social dissatisfaction. *Developmental Psychology* **29**, 611–621.
- Parker, J. G., Low, C. M., Walker, A. R., and Gamm, B. K. (2005). Friendship jealousy in young adolescents: Individual differences and links to sex, self-esteem, aggression, and social adjustment. *Developmental Psychology* **41**, 235–250.
- Rose, A. J. and Asher, S. R. (2009). *Boys' friendships: What's good about being male?* Paper submitted for publication.
- Rose, A. J., Carlson, W., and Waller, E. M. (2007). Prospective associations of co-rumination with friendship and emotional adjustment: Considering the socioemotional trade-offs of co-rumination. *Developmental Psychology* **43**, 1019–1031.
- Wentzel, K. R. and Caldwell, K. (1997). Friendships, peer acceptance, and group membership: Relations to academic achievement in middle school. *Child Development* **68**, 1198–1209.
- tasks and social goals perspective. In Maehr, M. L., Karabenick, S. A., and Urdan, T. C. (eds.) *Advances in Motivation and Achievement* (vol. 15: *Social Psychological Perspectives*), pp 357–390. Bingley, UK: Emerald Press.
- Asher, S. R. and Parker, J. G. (1989). Significance of peer relationship problems in childhood. In Schneider, B. H., Attili, G., Nadel, J., and Weissberg, R. P. (eds.) *Social Competence in Developmental Perspective*, pp 5–23. Norwell, MA: Kluwer Academic Publishers.
- Bukowski, W. M., Newcomb, A. F., and Hartup, W. W. (eds.) (1996). *The Company They Keep: Friendship in Childhood and Adolescence*. New York: Cambridge University Press.
- Dunn, J. (2004). *Children's Friendships: The Beginnings of Intimacy*. Malden, MA: Blackwell.
- Gottman, J. M. and Parkhurst, J. T. (1980). A developmental theory of friendship and acquaintanceship processes. In Collins, W. A. (ed.) *Minnesota Symposium on Child Psychology*, vol. 13, pp 197–253. Hillsdale, NJ: Erlbaum.
- Hartup, W. W. and Stevens, N. (1997). Friendship and adaptation in the life course. *Psychological Bulletin* **121**, 355–370.
- Juvonen, J. (2007). Reforming middle schools: Focus on continuity, social connectedness, and engagement. *Educational Psychologist* **42**, 197–208.
- Krappmann, L. (1996). Amicitia, drujba, shin-yu, philia, Freundschaft, friendship: On the cultural diversity of human relationships. In Bukowski, W. W., Newcomb, A. F., and Hartup, W. W. (eds.) *The Company They Keep: Friendship in Childhood and Adolescence*, pp 19–40. New York: Cambridge University Press.
- Ladd, G. W. (2005). *Children's Peer Relationships and Social Competence: A Century of Progress*. New Haven, CT: Yale University Press.
- Newcomb, A. F. and Bagwell, C. L. (1995). Children's friendship relations: A meta-analytic review. *Psychological Bulletin* **117**, 306–347.
- Parker, J. G., Rubin, K. H., Price, J. M., and DeRosier, M. E. (1995). Peer relationships, child development and adjustment: A developmental psychopathology perspective. In Cicchetti, D. and Cohen, D. J. (eds.) *Developmental Psychopathology*, pp 96–161. New York: Wiley.
- Rubin, K. H., Bukowski, W., and Parker, J. (2006). Peer interactions, relationships, and groups. In Eisenberg, N. (ed.) *Handbook of Child Psychology: Social, Emotional, and Personality Development*, 6th edn., pp 571–645. New York: Wiley.

Further Reading

- Asher, S. R. and Coie, J. D. (eds.) (1990). *Peer Rejection in Childhood*. New York: Cambridge University Press.
- Asher, S. R., MacEvoy, J. P., and McDonald, K. L. (2008). Children's peer relations, social competence, and school adjustment: A social

Coping with Stressful Situations: An Important Aspect of Self-Regulation

M Boekaerts, Leiden University, Leiden, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Coping with stressful situations includes coping with difficulties, obstacles, impediments, failure, shortcomings, impairments, handicaps, and liabilities. Several attempts have been made to classify the stressors that children and adolescents encounter in daily life. A distinction was made between acute and chronic stimuli that exert a taxing demand on the individual and require an adaptive response (Compas, 1987). Acute demands are caused by cumulative life events (e.g., parental divorce) and specific events, such as irritating, frustrating, and anxiety-provoking situations that occur in the classroom, at home, and on the way to school. These new demands may disrupt ongoing and upcoming activities, because the youngster must meet a sudden stream of threats and challenges. A successful attempt to cope with the acute source of stress signifies that a bridge has been built between the stressor and the existing coping repertoire. By contrast, several unsuccessful coping attempts may turn a situation into a chronic stressor.

Chronic stressors refer to recurring stressful life events that cause aversive conditions in the home (such as poverty, maternal depression, sibling rivalry) or the school environment (suboptimal learning conditions, teacher harassment, peer rejection, and feelings of academic inadequacy). Specific personal conditions may also create an enduring handicap or liability for the student, such as suffering from a chronic disease, physical handicaps, or learning problems.

Several questionnaires have been constructed to explore the major life events and daily hassles that students encounter in a school context. The data collected revealed that school stressors can be divided into social conflicts (e.g., unpleasant social exposure: rejection, bullying, harassment, and jealousy) and academic stressors (e.g., failures, overloads, exam stress). Both acute and chronic stressors impact on students' functioning, but the degree of the impact depends on the students' interpretation of the stressor and the coping strategies that they can bring to bear on the stressor. Research has described various adverse outcomes that are associated with lack of effective coping strategies, including several types of psychosomatic complaints (e.g., insomnia, headache, tummy ache, nervousness, feeling dizzy, and low back pain), absenteeism, fatigue, drop-out, and maladaptive social behavior. Interventions have been designed to help teachers and educators coach students' coping attempts. Before discussing some of these interventions, some the components of the coping process are discussed.

What Does Coping with Problematic Situations Entail?

When under stress, students report that their ongoing behavior is interrupted suddenly. They experience increased levels of arousal. Often they report negative emotions, such as anxiety, irritation, disappointment, shame, guilt, and sadness. They also report intrusive, repetitive thoughts about the event that caused the stress, and about the possible consequences of not being able to reduce the threat. They might inform about bodily changes as well, such as increased heart rate and perspiration, and about behavioral changes (e.g., aggressive or passive behavior).

It is important to realize that stress does not lie in the problematic situation itself. Rather, it concerns the way a potential stressor is interpreted. Each person comes to a stressful situation with either direct or vicarious experiences with the stressor. These prior experiences are activated at the time the stressor occurs and form a sort of internal context to interpret the problematic situation. At the same time, the person perceives and interprets the situation in which the stressor is currently embedded (external context). The significance that youngsters attach to a stressor and the way they interpret it depends on both the internal and external context.

An interesting question that was raised in this respect is: what is effective coping and which external and internal conditions promote (in)adequate coping? Questions such as these will never find a definite answer, because no solution is valid for all individuals at all times. An answer that sounds plausible is that personal functioning, development, and well-being are rooted in personal resources (e.g., self-efficacy, competence, resilience) and contextual affordances that facilitate the pursuit of one's salient goals and helps one to be resistant to life stressors and daily impediment.

Later in this article, some interesting issues about the effect of socialization and context on the development of (mal)adaptive forms of coping are discussed. For now, it is sufficient to understand that there are several conditions that should be met for a youngster to cope adaptively with classroom stress. In the first place, students should be able to appraise a problematic situation and determine whether that situation is changeable and controllable. Two other important conditions that should be met have to do with the selection of adequate coping strategies.

On the one hand, students need to regulate their emotional arousal. On the other hand, they should come up with a solution that produces a sense of mastery (Compas *et al.*, 1999). These important conditions are addressed next.

Appraising a Problematic Situation

Lazarus and Folkman (1984), the founders of transactional stress theory, view stress as the transaction between a person and the environment. They defined stress in terms of the individual's appraisal of the situation. More specifically, Lazarus and Folkman argued that any new, unexpected situation triggers a primary and a secondary appraisal. The primary appraisal determines how significant the situation is for well-being. Is it benign, threatening, or harmful? Does it harbor a loss of some kind? Situations are appraised as problematic when individuals detect a discrepancy between the perceived task demands and their own resources to meet these demands. Secondary appraisal determines the degree of the threat, challenge, harm, or loss. More specifically, the person assesses whether his or her coping resources are sufficient to deal with the situation. The degree of stress that is experienced depends on the balance achieved between primary and secondary appraisal processes.

Perrez and Reicherts (1992) proposed a way to explore the mental representation that a person makes of the stressor. They described five objective characteristics that individuals use to describe and categorize a stressful situation. These are: valence (how stressful or intense is the situation?), re-occurrence (will the situation re-occur in the future?), changeability (will the stressor go away by itself, without any actions being undertaken?), controllability (do I have control over the situation), and ambiguity (is there sufficient information to interpret the situation?).

Selecting a Coping Strategy

Based on their mental representation of the stressor, students select one or more coping strategies from their repertoire. What types of coping strategies can they select from? Lazarus and Folkman (1984) proposed a basic dichotomy between problem-focused and emotion-focused coping strategies. The former refers to efforts to change the situation by acting on the source of stress (e.g., work harder, problem solving). Emotion-focused coping refers to efforts to reduce the emotional distress by moving away from the source of stress either mentally or actually (e.g., denial, distraction). Other researchers described two slightly different universal ways of dealing with stressors, such as approach versus avoidance; active versus passive forms of coping; and primary versus secondary control coping strategies. They all argued that individuals tend to use habitual and preferential ways of dealing with specific types of

stressors, but that there is no way of telling which coping mode is most effective for a specific stressor. It is almost impossible to make any comments on the effectiveness of a person's coping strategies without knowledge of his or her coping goal. This issue is discussed further, later in the article.

Stress researchers working in educational psychology have used an empirical approach; they factor-analyzed the different coping strategies that students report using. The literature can roughly be divided into how students cope with unspecified stressors and with specific daily hassles. One of the most valid coping questionnaires is that of Frydenberg (2004). Frydenberg and her colleagues reported that 18 coping strategies are commonly used by adolescents, namely seek social support, social action (join people with the same concerns), seek to belong (improve relationships), invest in close friends, wishful thinking, worry, work hard, keep to oneself, seek relaxing diversions, physical recreation, focus on the positive, ignore the problem, tension reduction, seek professional help, spiritual support, self-blame, and not coping. Summarizing the results of several studies, Frydenberg reported age and gender differences. The younger students in her samples (11–13 years of age) reported more hard-work strategies and less self-blame and tension-reduction strategies than did the older students (15–16 years). Girls tended to address problems immediately and used more social support seeking, wishful thinking, and tension releasing than did boys. Boys adopted a wait and see strategy, preferring to manage by themselves rather than seek social support.

Bridging the Gap between New Stressors and One's Coping Repertoire

Children and adolescents are in a period of cognitive, emotional, physical, and social development. It is clear that situations that they consider problematic at one point in time may not be problematic a few weeks or months later. For example, at major transition periods (e.g., changing body image, changing schools) most youngsters appraise their daily life as chaotic and problematic. They cannot form an adequate mental representation of the new situation and find it difficult to deal adequately with the stressor. Often, they characterize the stressor as highly significant, re-occurring, unchangeable, uncontrollable, and ambiguous (e.g., I cannot find my way in the new school; I do not know the rules). Although such transitions are part of normal development, prior experiences of uncertainty and insecurity may be activated at the time the stressor occurs, thus forming an unfavorable internal context to interpret the new situation. At the same time, the student may perceive and interpret the actual situation in which the stressor is currently embedded as suboptimal (e.g., I do not have any friends here).

Fortunately, most students will pass these transitions smoothly, mainly because their habitual coping strategies prove to be adequate in the new context.

It is important to realize, however, that selecting a coping strategy to deal with a new stressor is not a one-shot selection process. It is a dynamic process that involves a series of transactions. For example, it is plausible that a student, who is confronted with a social conflict in class, uses emotion-focused coping before switching to problem-focused coping, or vice versa. In stressful situations, students have – very often – a direct, urgent need to restore their well-being, as well as a long-term intention to resolve the problem or conflict. During the coping episode, these short-term and long-term goals compete for dominance in the goal system as the following example illustrates.

Emily's teacher has just given class a challenging group assignment. Emily wants to start right away, but Maggie pressurizes group members to work in a manner they do not like. Emily protests and asks Maggie to explain why her method is better. Maggie ignores her and starts to give directions. Emily protests again, raising her voice. Other group members avert her gaze. Emily starts chatting with Lucy about her new skirt. They seemed to enjoy the interaction for a while and then joined the group in the problem solving process.

In order to interpret Emily's successive ways of dealing with the social stressor, we need to know the functions that different coping strategies serve during the coping episode. At first, Emily's focus was on the learning process, but Maggie's social pressure triggered negative emotions. Emily wanted to stand her ground and argued against the proposed working method, because she felt that one of her long-term goals was frustrated (not to let other students make decisions for her). At the same time, she felt a strong urge to avoid conflict and tension in her group, particularly when she noticed that the rest of her group did not openly support her protest. Hence, Emily opted for problem-solving coping initially but switched to emotion-focused coping later on. Research has shown that youngsters who rely on both types of strategies exhibit better overall adjustment than those who are limited to a single type of coping.

Coping and Volition: Two Essential Aspects of Self-Regulation

The notion of the purpose and function of coping processes is fundamental to the distinction that Lazarus and Folkman made between emotion-focused and problem-focused coping strategies. However, we need to know more than the students' appraisal of the stressor and their choice of coping strategies. We also need to consider

their intentions and goals in relation to the stressor, as well as their intentions and goals concerning the activity that was interrupted by the stressor.

At this point in the discussion, with a little digression, a point to note is that there are moments in the history of a discipline when concepts and models, stemming from a neighboring discipline, start penetrating into one's own theories and models. At this junction in time, a great deal of confusion, due to the simultaneous use of old and new labels, might be prevented if somebody compares and contrasts these different models and brings order into the multitude of constructs.

Elsewhere (e.g., Boekaerts, 2006), it is proposed to blend constructs and research findings from the coping and volition literature and integrate them into the self-regulated learning model. There is a distinction made between three types of self-regulation, namely top-down, bottom-up, and volition-driven self-regulation. Top-down self-regulation refers to goal pursuit that is driven by the students' own values, needs, interests, and goals. Bottom-up self-regulation occurs when specific cues in the learning environment create a mismatch with the students' current learning intention. Perception of a mismatch triggers cue-driven self-regulation with the purpose of engaging in activities to explore the nature of the mismatch. Discovery of obstacles and difficulties that are a threat to well-being elicit negative emotions. At that point, students' attention is drawn away from the activity they were previously engaged in. They either get involved in activities, which serve the purpose of regulating their emotional arousal, or engage in efforts to continue goal pursuit. The former strategies are called emotion-regulation strategies. They are akin to the emotion-focused coping strategies that are described in the coping with stress literature. The latter strategies, which are called volition-driven strategies, are akin to the problem-focused strategies that were described previously.

The three self-regulation strategies are visualized in Boekaerts' (2006) dual processing self-regulation model. The top part of **Figure 1** represents the formation of a mental representation, or appraisal, of a learning situation. The bottom part depicts the three self-regulation pathways. The broken line visualizes top-down self-regulation. Students start activity on the mastery pathway, when their appraisal of the learning situation is such that the learning goal is congruent with their personal goals. This match elicits dominantly positive emotions (e.g., Emily is interested in the task. She wants to increase her competence).

Activity on the well-being pathway (dotted line) begins when students initially appraise the learning situation as problematic or when they detect cues that signal a threat to well-being during the learning process. In our example, the pressure that Maggie put on the group created a mismatch with Emily's current learning intention and with her need for autonomy. This triggered negative

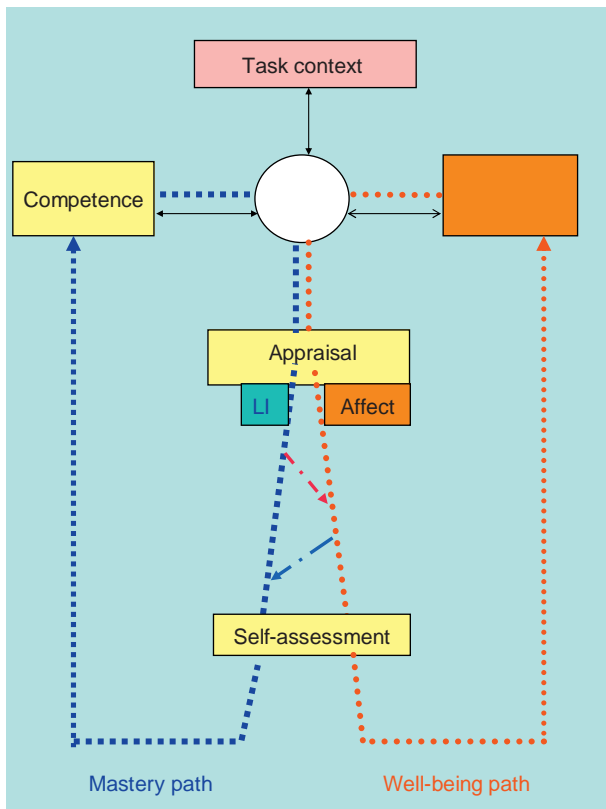


Figure 1 The dual processing self-regulation model. From Boekaerts, M. (2006). Self-regulation and effort investment. In Sigel, E. and Renninger, K. A. (eds.) *Handbook of Child Psychology, Vol. 4, Child Psychology in Practice*, pp 345–377. Hoboken, NJ: Wiley.

emotions and initiated a coping episode. At first, Emily stayed task-focused and tried to re-negotiate the working method (i.e., she used volitional strategies to stay on the mastery pathway). Her emotional arousal increased when she observed that Maggie ignored her and that the group did not support her either. Emily raised her voice, but when that did not help she opted for a different coping strategy, namely looking for distraction. Emily experienced positive emotions while chatting with Lucy and this gave her enough energy to redirect her attention to the learning process. This re-routing is visualized in **Figure 1** by a path that runs from the well-being pathway to the mastery pathway.

Emily's example illustrated that she made use of different strategies to regulate her emotions and to re-fuel the learning process. The next two sections review briefly what is known about these two types of bottom-up strategies.

Regulating Emotional Arousal

Emotions are an inherent part of human functioning. Emotions are tightly linked to the context and what the student is trying to do. Experiencing negative emotions signals to the

individual that something is wrong and that it is important to monitor the context carefully. By contrast, positive emotions signal that everything is fine and that there is room for exploration and play. Why can some students regulate their emotions more effectively than others? The ability to regulate one's emotions refers to the capacity to understand one's own emotions and their expression. Students may or may not understand that emotions are sources of energy and they may or may not be able to override them when salient goals should be pursued.

Inability to temper the intensity and duration of one's emotional arousal (e.g., worry, self-blame) hinders functioning in a social context, whereas the capacity to temper one's emotions (e.g., tension reduction, relaxation, and distraction) facilitates functioning in a social context. It helps students to feel self-efficacious, appraise the classroom environment in a positive way, and the learning and problem-solving process as constructive. It is important to realize that modulating one's emotions does not necessarily imply that one has to play them down. At times, it may be beneficial to express one's emotions, amplifying them so that people present in the situation can take account of one's feelings. Children realize that the way they communicate their emotions affects their peers' subsequent reactions to their behavior. For example, Emily raised her voice, hoping that Maggie would listen to her protest. Maggie's reaction informed her that this was not a good strategy to use and she switched to distraction.

Eisenberg and Spinrad (2004: 4) showed that not all students are able to regulate their emotions strategically. They cannot temper their emotions in such a way that they can continue the (social) activities that they were engaged in at the moment the emotion was triggered. Research shows that parental modeling and coaching plays an important role in the way children regulate their emotions. Children who grow up in families with intense marital conflict develop either aggressive coping strategies and externalizing behavior problems (particularly boys) or acute distress reactions (particularly girls). Ramsden and Hubbard (2002) showed that the way children rated their own emotion regulation predicted aggressive behavior best. Interestingly, the overall degree of negative expressivity in the family coincided with a low rating of emotion regulation, whereas family acceptance of the child's emotions was associated with higher emotion regulation. These researchers pointed out that children, who grow up in families with moderate structure and availability of social support, have a larger repertoire of coping strategies to choose from.

Volitional Strategies: Finding a Solution That Produces a Sense of Mastery

Many common educational practices reduce students' intrinsic interest in academic learning, such as a focus on evaluations and grading, and presenting tedious, ambiguous,

and uncontrollable tasks. Why can some students remove obstacles to intrinsic sense making and persist in the face of difficulties whereas others cannot? Volition-drive self-regulation refers to the ability to maintain focus and effort toward learning goals despite obstacles and distractions. The capacity to get started on learning tasks or continue working on them, when one gets sidetracked by competing goals, facilitates functioning in a classroom and homework context. For example, it helps students to protect their intentions, specifically when difficult work must be completed. It is important to realize that having access to volitional strategies, which Randi and Corno (2000: 5) call good work habits, is essential for re-routing activities from the well-being to the mastery pathway. These include, how to organize one's work, keep distractions at bay, make a time schedule and stick to it, how to resolve conflicts, how to set goals and sub-goals and prioritize them, and how to monitor progress. Socialization practices impact on the development of volitional strategies and it is important that teachers and parents enhance will-power by modeling it and discussing its effectiveness with students. An interesting study by Hill and Craft (2003: 5) reported that parental involvement in school tasks also affects performance. They found that the effect of parental involvement on math performance was mediated by emotional regulation, the children's tendency to accept authority, and their ability to initiate activities and stay on track.

How Can We Help Students Cope with Stressful Situations in the Classroom?

In the previous sections, it has been argued that students need to know which conditions trigger stress and need to practice how they can deal strategically with increased levels of arousal. Often, initial attempts to deal with a stressor fail because students do not know which strategies are most effective in a given context. Alternatively, they might have this knowledge but fail to generate productive strategies at the point of use. In the remainder of this article, two intervention studies that provided insight into how teachers and educators might coach students' coping attempts (for a more extensive review, see Pincus and Friedman, 2004) are described.

Stress Interventions Coach Students' Coping Attempts

Cunningham *et al.* (2002) argued that many adolescents manage their coping resources ineffectively. They hypothesized that low coping efficacy is associated with a pessimistic attributional style (i.e., positive outcomes are viewed as specific, temporary, and caused by luck) and nonproductive coping strategies. These researchers set up an intervention to decrease nonproductive, emotion-focused coping

strategies (worry, wishful thinking, tension reduction, ignoring the problem, self-blame, and keeping to oneself). After an 8-week training period, the students had a greater sense of control over their internal states as compared to controls, showed significantly less depressive attributions, and a decrease in nonproductive coping strategies. Contrary to expectation, no change was found in the use of productive strategies. The reason could be that the researchers' focus was on what the students should not do, rather than on the use of productive coping strategies.

An interesting study by Zimmerman and Kitsantas (2002) showed that students need to have the chance to build up conditional knowledge about the types of obstacles they may encounter during skill development and about the coping strategies that are effective to deal with these obstacles. They found that students, who had observed how a model flawlessly performed a new skill (here: dart throwing and writing-revision strategies) surpassed students who did not have the chance to observe a model perform the task. However, students, who had been exposed to a coping model before they had to practice the skill, outperformed the former group. Why? Watching a coping model struggle to implement emotion regulation and volitional strategies informed the students that the new skill is complex and difficult to implement and that stressors may come in one's way, yet realize that the stressor is changeable and controllable. This information is stored as conditional knowledge and allows students to bootstrap their strategy use and efficacy beliefs. Zimmerman and Kitsantas reported that the coping model condition produced not only better observational learning but also more effective practice experiences and superior acquisition during post-testing than the mastery model or no model condition.

It is easy to implement the coping model condition in the classroom, allowing students to build up information about the personal resources that are necessary and sufficient to deal with the obstacles that may come on their way and about the contextual affordances that facilitate the pursuit of the learning goals. The final point is that there is an urgent need to train teachers to coach volitional and coping strategies.

Bibliography

- Boekaerts, M. (2006). Self-regulation and effort investment. In Sigel, E. and Renninger, K. A. (eds.) *Handbook of Child Psychology, Vol. 4, Child Psychology in Practice*, pp 345–377. Hoboken, NJ: Wiley.
- Compas, B. E. (1987). Coping with stress during childhood and adolescence. *Psychological Bulletin* **101**(3), 393–403.
- Compas, B. E., Connor, J., Saltzman, H., Thomsen, A. H., and Wadsworth, M. (1999). Getting specific about coping: Effortful and involuntary responses to stress in development. In Lewis, M. and Ramsey, D. (eds.) *Soothing and Stress*, pp 229–256. Mahwah, NJ: Erlbaum.

- Cunningham, E. G., Brandon, C. M., and Frydenberg, E. (2002). Enhancing coping resources in early adolescence through a school-based program teaching optimistic thinking skills. *Anxiety, Stress and Coping* **15**(4), 369–381.
- Eisenberg, N. and Spinrad, T. L. (2004). Emotion-related regulation: Sharpening the definition. *Child Development* **75**(2), 334–339.
- Frydenberg, E. (2004). Coping competencies: What to teach and when. *Theory into Practice* **43**(1), 14–22.
- Hill, N. E. and Craft, S. A. (2003). Parent–school involvement and school performance: Mediated pathways among socio-economically comparable African American and Euro-American families. *Journal of Educational Psychology* **95**(1), 74–83.
- Lazarus, R. S. and Folkman, S. (1984). *Stress, Appraisal and Coping*. New York: Springer.
- Perrez, M. and Reicherts, M. (1992). *Stress, Coping, and Health*. Seattle, WA: Hogrefe and Huber.
- Pincus, D. B. and Friedman, A. G. (2004). Improving children's coping with everyday stress: Transporting treatment intervention to the school setting. *Clinical Child and Family Psychology Review* **7**(4), 223–240.
- Ramsden, S. and Hubbard, J. (2002). Family expressiveness and parental emotion coaching: Their role in children's emotion regulation and aggression. *Journal of Abnormal Child Psychology* **30**(6), 657–667.
- Randi, J. and Corno, L. (2000). Teacher innovations in self-regulated learning. In Boekaerts, M., Pintrich, P. R., and Zeidner, M. (eds.) *Handbook of Self-Regulation*, pp 651–685. San Diego, CA: Academic Press.
- Zimmerman, B. and Kitsantas, A. (2002). Acquiring writing revision and self regulatory skill through observation and emulation. *Journal of Educational Psychology* **94**(4), 660–668.

Culture in Motivation Research: A Challenging and Enriching Contribution

K Kimmel and S Volet, Murdoch University, Murdoch, WA, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Culture – In its broad, ethnographic sense, culture, or civilization, is that complex whole which includes knowledge, belief, art, morals, law, custom, and any other capabilities and habits acquired by people as a members of society.

Emic/etic – An emic account of behavior is a description of behavior in terms that are meaningful (consciously or unconsciously) to the actor, whereas an etic account is a description of a behavior in terms that are familiar to the observer. Scientists interested in the local construction of meaning and local rules for behavior rely on emic accounts, while those interested in facilitating comparative research and making universal claims rely on etic accounts.

Motivation – Commonly defined as an internal state or condition and sometimes described as a desire or want that drives people's behavior and gives it direction. Based on the expectancy-value theory, what motivates behavior is a function of the expectancies one has and the value of the goal toward which one is working.

Aim, Focus, and Structure

This article aims to capture the current zeitgeist of research on culture and motivation, embedded in its historical development. Key motivational constructs have been selected and reviewed to illustrate the diversity and richness of culture-based theorizing as well as the range of empirical studies that have examined motivation from a cultural perspective. The article also highlights the salient research trends that have emerged in the last decade and the significant contribution that culture has made to motivation research.

As a background to understanding the development of recent research on culture and motivation, the first section provides a brief overview of critical milestones in the development of culture research in the broader field of psychology. The following section examines five key motivational constructs that have attracted a significant amount of research from a cultural perspective. The choice of constructs and the grouping of studies are, to a

large extent, arbitrary, the aim being to illustrate a range of unique theoretical and empirical contributions that a culture-based perspective has made to motivation research. Two examples of cultural psychology research have been added to show that emic research from non-Western settings can make a unique contribution by unveiling new dimensions of learning and motivation. The article concludes with a brief discussion on the shortcomings of research and future directions.

Background and Historical Development of Culture-Based Research

The modern epoch of cross-cultural psychology with a coherent research agenda began only in the mid- to late 1960s. Earlier, according to Adamopoulos and Lonner (2001), cultural studies were largely the domain of anthropologists. The main purpose of early cross-cultural psychology research was to test theories, initially developed and validated in Euro-American contexts, in a range of other cultural contexts so that they could claim universality. This culture-label research (typically using a country or an ethnic group as the independent variable) has revealed numerous cross-cultural differences in target variables. However, inherent to their research design, studies representing an essentialist conceptualization of culture could not explain the variations that were observed.

To address this issue, many cross-cultural psychologists have shown interest in identifying and measuring cultural variables that may account for cross-cultural differences. The most well-known cultural variables involve individualism/collectivism (Hofstede, 1980; Triandis, 1994), basic human values (Schwartz, 1994), and independent/interdependent constructions of the self (Markus and Kitayama, 1991). Methodological problems have also been addressed, for example, the psychometric equivalence of data from cross-cultural settings (Van de Vijver and Poortinga, 1982). Overall, however, and regardless of whether the research involves a culture label or adopts a culture-measured approach, this research is problematic according to cultural psychologists because there is an underlying assumption that culture is static and homogenous, and that it can be treated as an antecedent of psychological phenomena.

For cultural psychologists, differences across cultures are the product of unique cultural contexts, where culture

and mind can only be conceived as inextricably connected since they mutually constitute each other (Adamopoulos and Lonner, 2001). It is argued that cultures need to be treated as dynamic processes (Greenfield, 1997) or open systems (Kitayama, 2002) that spread across geographical borders, evolve over time, and are constantly in flux due to changing contextual characteristics (Hong and Chiu, 2001; Kagitcibasi, 1996; Kashima, 2001; Zusho and Pintrich, 2003). The dynamic, complex interface of culture, context, and social cognition is a relatively new trend in cultural and educational psychology research. Accordingly, mainstream motivation research has just begun to explore how the interplay of the cultural aspects of contexts and personal dimensions co-contribute to producing different motivational patterns.

Additionally, similar views to the position of cultural psychologists have been expressed from an indigenous psychology perspective. Although the main motive of the indigenous psychologists is to address the prevailing dominance of the Euro-American mainstream cross-cultural psychology research, their position is consistent with that of cultural psychologists. Their call for the development of theories from within, that is, those developed within non-Western contexts that are therefore more appropriate for such milieus (Sinha, 1998; Yang, 2000), is consistent with the view that cultural systems should be the unit of analysis (Kim and Berry, 1993) in psychological research.

Overall, the major shift in culture-based research from a static, homogenous, decontextualized conceptualization to a dynamic, complex, and contextualized perspective has emerged almost in parallel to recent developments within mainstream research on motivation and learning (Volet, 2004). In the latter, the shift is from studying psychological phenomena with an exclusive focus on the individual (cognitive) toward a person-in-context perspective (sociocognitive), taking account of the location of mental processes in social activities that are embedded in broader social, cultural, and historical contexts (situative, sociocultural perspective) (Pintrich, 2000; Turner and Meyer, 2000; Volet, 2001). In addition, this perspective (situative, sociocultural) stresses the significance of mutual, dynamic interactions between individuals and culturally constituted contexts for the emergence of cognitive, motivational, and learning orientations. Thus, both cultural psychology and the situative perspective highlight the emerging, complex, dynamic, and contextualized nature of culture and motivation.

Motivational Constructs Investigated from a Cultural Perspective

The search through conceptual and empirical material for this article revealed that motivation research from a

cultural perspective is very diverse and spans across a range of theoretical perspectives and constructs. Here, we provide illustrations of the type of theorizing and empirical work that has emerged in recent years with the view to highlighting its richness and significance. As mentioned in the introduction, the selection of motivational constructs and the grouping of studies are, to some extent, arbitrary; therefore, this article does not represent any systematic mapping of the field.

Achievement Motivation, Its Relationship to Effort and Ability

In the 1960s and 1970s, extensive cross-cultural research related to achievement motivation (McClelland, 1961) was conducted in Asian contexts. The purpose was to investigate the roots of achievement motives to better understand the essential underlying psychological mechanisms of achieving societies. A number of empirical studies (e.g., McClelland, 1961, 1965) revealed that Asian samples scored lower on achievement motivation compared to American samples, which was conceptualized at the time as a relatively stable personality disposition learned through independence and mastery training. This type of research was highly criticized in the 1970s for being unable to provide insight into the specific cultural context in which achievement motivation was generated. Maehr (1974), for instance, proposed “a framework . . . that stresses the importance of contextual conditions in eliciting achievement motivation” (p. 887). Similarly, Salili *et al.* (1976) pointed out that motivational patterns could manifest differently across diverse cultural contexts due to varying sociocultural influences. In the 1990s, Yu and Yang (1994) also criticized McClelland’s work on the ground that this achievement motivation theory was based on Western middle-class values, which are not generalizable to an Asian setting. Similarly, Salili’s (1994) research using a repertory grid technique (based on Kelly’s 1955 personal construct theory) suggested that different cultures could share the same dimensions of achievement, but their conceptions of achievement, for example, the meaning attached to success, could vary.

This illustration shows how, in early research on achievement motivation, culture was viewed as the unique, contextual frame for the development of motivational patterns. These patterns, therefore, were assumed to reflect the values and beliefs of the specific sociocultural setting. Simply stated, the view that culture is context was used to explain why motivational orientations and processes have different manifestations across distinct cultures.

Recent work related to achievement motivation (Hufton *et al.*, 2002a) has also endeavored to reveal causal relationships between culture and motivation; however, culture and context have been treated as distinct, and deliberate efforts have been made to interpret the

meaning of the findings in relation to the characteristics of the cultural context. The empirical work of Hufton *et al.*, conducted with American, English, and Russian students, revealed ambiguous relations between the self-perception of academic competence and attribution of achievement to effort and ability. The former two groups appeared more likely to view effort as the cause of high achievement, although they significantly displayed lesser endeavor than their Russian counterparts, who, in contrast and despite working significantly harder, were more likely to ascribe high achievement to ability.

It is noteworthy that Hufton *et al.* (2002b) did not limit the interpretation of their findings in light of the prevailing culture alone. Context was treated separately from culture in terms of schooling and classroom practices, on the grounds that these are expected to shape and build the basis for the emergence of distinct motivation and learning patterns. The authors, moreover, highlighted the significance of not only understanding the meanings attached to the notions of effort, ability, and achievement, but, more importantly, also acknowledged that these may vary within and between cultures. To this effect, they combined their survey with an in-depth qualitative component. Their exploration of the meanings attached, by Russian students, to effort and ability revealed that although a strong emphasis was laid on effort, working hard or effort was considered to be the norm, leading the authors to conclude that "individual differences in ability may be a more salient and discriminating factor" (Hufton *et al.*, 2002b, p. 282). However, based on their additional findings of the conceptions of ability as the result of effort in Russian students, the authors criticized any simplistic and dichotomous conceptualizations of effort/ability.

The value of combining cross-cultural surveys with qualitative studies has also been advocated by Bempechat and Drago-Severson (1999), who called for a qualitative shift in cross-cultural research on achievement motivation. They stressed the significance of exploring the context and culture-specific beliefs about learning, achievement, and motivation to elicit the underlying meanings that individuals attach to these constructs. In their view, subjective perspectives are critical to gaining a deeper and fuller understanding of why learning and motivation patterns are consistent or vary inter- and intraculturally.

The following section pursues the examination of the dichotomous conceptualizations of motivational constructs in a cultural perspective, this time in regard to the self-determination theory, and the constructs of extrinsic and intrinsic motivation.

Self-Determination Theory, and the Constructs of Extrinsic and Intrinsic Motivation

From a Western perspective, the bipolar construct of extrinsic and intrinsic motivation is traditionally conceived

in terms of intrinsic motivation being more beneficial and efficient for learning as it serves as an antecedent for deep learning strategies (Watkins, 2000). This widely shared belief was recently contradicted by Ramburuth and McCormick's (2001) findings, which revealed evidence that the surface strategies were connected with intrinsic motivation for Asian students, whereas extrinsic motivation was linked to deep strategies for Australian students.

Other studies, such as Iyengar and Lepper's (1999), contributed to raising doubts about the Western-based conceptualization of intrinsic motivation rooted in the self-determination theory (Deci and Ryan, 1985, 2000). In short, the self-determination theory postulates that an individual's intrinsic motivation will be higher in situations where options of personal choice are given. Iyengar and Lepper's study found that while this assumption did hold true for their American participants, it failed to predict the motivational tendencies of their Asian sample. The latter group displayed higher intrinsic motivation when task choice was made by significant others, such as their mothers or a valued in-group member, than when they were given the opportunity to choose a task themselves. The authors interpreted their findings in light of Markus and Kitayama's (1991) theory of independent and interdependent self-construal. They argued that while free choice for people from independent societies corresponds to their need for autonomy and personal control, it might harm the need for relatedness of people coming from interdependent societies. In other words, it entails the risk that their personal choice may not be in line with the beliefs and values of important others, and therefore may cause conflict or jeopardize group belongingness.

In this regard, Katz and Assor (2007) argued that having a choice could be motivating when the options meet the choosers' needs for autonomy, competence, and relatedness. Of these, relatedness refers to congruence with the values of the cultures of the choosers, which, as discussed above, is of special importance for members of interdependent/collectivist societies. This assumption is supported by the research of Roth *et al.* (2006), which revealed that on the relative autonomy continuum theorized by the self-determination theory, conformity exists as an intermediate level between external regulation and introjection. The authors concluded that the need for autonomy might be less compatible with Eastern cultures that embrace collectivist values. They suggested future research to examine the relations between conformity, well-being, and performance in collectivist societies.

In light of these findings, which also explored the significance of culture as context in terms of pan-cultural dimensions and self-systems (independence/interdependence), it may be timely to reconsider the bipolar construct of extrinsic-intrinsic motivation. If choice does not play a critical role for people coming from more socially interdependent societies, the fundamental assumptions underlying

this concept can no longer be viewed as appropriate to explain human motivation across cultural contexts. It seems rather, that while the pursuit of self-determination may enhance intrinsic motivation in more independent societies, pursuit of social conformity fulfils that same function in more interdependent societies. However, and as alluded to in the introduction, caution needs to be exercised when explaining differences in motivational tendencies exclusively in light of a global cultural dimension, as it has only limited potential to acknowledge the dynamic nature of motivation across contexts and situations within cultures.

This issue is addressed in the following section when discussing the impact of culturally specific self-beliefs on achievement and motivational orientation.

Attribution Theory, Self-Beliefs about Achievement and Motivation Orientation

In the Euro-American literature, there is a widely shared belief that having confidence in oneself and thinking positively essentially helps and enables people to be successful and to perform at their best. The validity of these beliefs was supported, for example, by Bandura's (1982) research, which revealed that a positive sense of self-efficacy often results in enhanced achievement.

In recent years, however, a number of researchers active in cultural research have cast doubt on this perspective. They suggested that the positive impact of self-confidence might not sustain in cultural contexts other than the Euro-American one (e.g., Fiske *et al.*, 1997; Heine and Lehmann, 1997). Their empirical work in East Asian settings, such as Japan, revealed that, in contrast to the aforementioned shared assumption, it was a person's self-critical view that was positively associated with achievement and motivation.

These findings suggest that motivational beliefs may play out differently across contexts depending on culturally specific self-views. Again, this research treated culture (national culture) as one form of context and assumed it would provide specific opportunities, constraints, and affordances for the development of motivational orientations. For instance, research by Kitayama *et al.* (1997) and Heine *et al.* (2001) provided strong evidence that while the independent self-view (e.g., predominant in the United States) is positively related to self-enhancing motivational strategies, the interdependent self view (e.g., prevalent in Japan) is more likely to be associated with self-improving motivational approaches. Heine *et al.* found that their Japanese participants worked harder on a second task if they had failed to complete the first one successfully, which implies that they were strongly focusing on their weaknesses (self-criticism). In contrast, their American participants worked harder the second time if they previously succeeded in task completion, which suggests that they were motivated by their strengths (self-enhancement).

Notably, the authors' interpretation of these findings was not exclusively based on cultural dimensions of self-construal. Some attention was also given to each group's respective cultural-educational environment, in terms of school culture and university entrance qualification systems, on the grounds that these would have shaped people's lay theories of the self. Dweck and colleagues (e.g., Chiu *et al.*, 1997; Hong *et al.*, 1999) have coined the terms entity theory and incremental self-theories. They argued that while the former refers to the cultural belief that the self is relatively fixed and stable, the latter views the self as adjustable, fluid, and improvable. These culturally based self-beliefs may help explain the different motivational strategies displayed by American and Japanese participants following failure. While the American participants attributed their failure to lack of ability, a fixed, inherited characteristic, the Japanese participants thought that task completion was just a matter of more effort.

Despite these striking differences, it is important to highlight that although diverging motivational strategies were adopted, both groups were striving toward the same goal, namely to do their best. To sum up, it suggests that humans may share similar goals, needs, and desires. However, while the strategies that people adopt in pursuit of these, and the ways in which these are constructed, may depend on cultural elements to some extent, the multiple contextual factors prevalent in specific contexts are also significant. These were only minimally taken into consideration in the research examined so far. The importance of culture and context as distinct constructs is discussed in the following sections on goal orientation as well as agency and self-efficacy.

Goal Orientation, Social Diversity, and Educational Practices

Culture-based research on goal orientation has revealed that the basic constructs of the goal theory show remarkably similar structures across cultural groups (e.g., McInerney, 2003; McInerney *et al.*, 1997; Nelson *et al.*, 2006). Based on that research, it would appear that goal orientation and achievement motivation may not be that dissimilar around the globe.

However, research conducted in the context of multicultural classrooms, for example, by Kaplan and Maehr (1999), showed strong evidence that while task goals were positively related to the sense of school belonging, perceived competence, and self-esteem in minority students, by contrast, ego goals correlated negatively. In this regard, Maehr and Yamaguchi (2001) argued that school cultures that stress and encourage task goals may play an important role in reducing negative and inhibiting aspects associated with social diversity. Most importantly, this strengthens the view that motivational orientations are

malleable, and thus may change in response to specific educational practices.

The respective significance of culture and multiple contextual variables on students' learning and motivational orientations has been investigated by, for example, Salili *et al.* (2001). The authors conducted research with three groups of students, namely Chinese in Hong Kong, European-Canadian, and Chinese-Canadian. The findings were interpreted in light of the sociocultural setting as well as the context of schooling practices (e.g., grading and assessment systems). The study revealed that Chinese-Canadian and European-Canadian students who participated in the same learning environment displayed different learning attitudes and goals. For example, Chinese-Canadian students spent more time studying, received higher marks, and indicated family-oriented goals more often than European-Canadian students. Importantly, observed differences in self-efficacy scores between Chinese-Canadian and Hong Kong Chinese students could not simply be explained by culture alone. Different schooling practices, such as assessment processes and success criteria, were found to play a vital role, with a significant impact on students' self-efficacy beliefs and motivational patterns. Hong Kong students doubted their self-efficacy because despite working very hard, they only received low marks. In contrast, the efforts made by Canadian students were reflected in their examination results.

The findings of this study by Salili *et al.* (2001) illustrate the importance of interpreting motivational orientations in light of the specific cultural milieu and the multiple contextual characteristics that afford and constrain particular learning and behavior patterns. Moreover, the study highlights that a contextualization of research has the potential to reveal stability and change in individual motivation patterns across situations and over time.

The final section addresses the situated nature of individuals' agency and self-efficacy beliefs, and discusses how these evolve differently within and across cultures and contexts.

Agency and Self-Efficacy, Separating Culture and Context

In recent years, a number of researchers, for example, Hernandez and Iyengar (2001) and Kitayama and Uchido (2005), have suggested that cultural differences in motivation may be best explained in terms of distinct agentic modes. The idea is that people coming from cultures that stress independence and autonomy are more personally agentic and their behaviors attributed to dispositional characteristics, whereas people from cultures that emphasize interdependence are more collectively agentic and their behaviors attributed to situations or are even viewed as directed by groups. More specifically, it is argued that

personal agents view the self as the source of agency and essentially display higher intrinsic motivation in situations that involve self-initiated and self-directed actions. In contrast, collective agents perceive agency as emerging from the collective and, in turn, exhibit higher intrinsic motivation for behaviors that are rooted in and directed toward the collective (Kitayama and Uchido, 2005; Markus and Kitayama, 2004).

In the same line of thought, three modes of agency are distinguished in the sociocultural theory: personal, proxy, and collective agency. While personal agency is exercised individually, proxy agency is in operation when individuals influence others to take actions for them. In collective agency, individuals act in accordance with each other to produce collectively desired outcomes (Bandura, 2002). However, Bandura made it explicit that it is of utmost importance to realize that although "the determinants and agentic blends of individual, proxy and collective agency vary cross-culturally . . . all these agentic modes need to be enlisted to make it through the day, regardless of the culture in which one happens to reside" (Bandura, 2002, pp. 269–270). Consequently, he also argued that cultural variations in the behaviors of individuals may be best explained in terms of the relative importance attributed to each type of agency in a particular cultural context, rather than the result of entirely bipolar individualist or collectivist modes of agency. Moreover, Bandura stated that regardless of which mode of agency is exhibited, one underlying mechanism is omnipresent, namely efficacy beliefs. The core belief that one has the power and the ability to achieve desired outcomes serves as a baseline for a range of factors that may guide the behaviors of individuals. In other words, cognitive, motivational, affective, and decisional processes are viewed as essentially shaped by an individual's self-efficacy beliefs. However, again according to Bandura, although efficacy beliefs have generalized functional meanings, their emergence, structure, exhibition, and purpose vary across cultural contexts. Additionally, as mentioned above, not all efficacy beliefs are limited to an individual perspective. Collective efficacy beliefs refer to situations where group members act in accordance to a shared belief that a desired outcome can be achieved by joint actions. Again, however, although the focus shifts from the individual to the collective, the basic underlying functions and processes of efficacy beliefs are the same. Consequently, based on Bandura's position, it can be argued that regardless of the cultural context, there is universal commonality in human agency and mechanisms of operations. It is the ways these mechanisms are put in practice and the shapes they adopt that can vary cross-culturally.

Furthermore, researchers such as Bandura (2002), Hernandez and Iyengar (2001), Hong and Chiu (2001), and Kashima (2001) have stressed the contextualized nature of human behavior. They view context as a combination of

a variety of contextual attributes and situational factors that can engender or inhibit human motivation, regardless of the prevalent cultural milieu. For instance, depending on whom a person is interacting with and what the interaction is about, intra-individual differences in behavioral and motivational patterns can emerge. This idea is supported by Freeman and Bordia's (2001) findings that depending on the reference group (e.g., peer, family, academic, and national), participants exhibited different levels of individualistic and collective orientations.

These findings therefore support the significance of situational and contextual factors on the development of behavioral and motivational orientation. Furthermore, they highlight the importance of taking intra-individual as well as within-culture variations into account. Thus, focusing on context and situation-specific characteristics as fundamental for the thoughts and actions of the individuals, rather than exclusively relying on the global construct of culture, essentially enhances the understanding of the dynamic nature of social behavior. In this regard, Volet's (1999) research provided strong support for the benefits of systematically separating culture and context, allowing for the examination of stability and change in the motivational patterns of Singaporean and Australian students in the same educational context and over time. The characteristics of the specific learning settings as well as students' subjective perceptions of these settings were assumed to afford and constrain particular learning and motivational patterns, regardless of the students' cultural backgrounds. Volet found that while the overall achievement motivation remained relatively stable in the group of Singaporean students, more contextualized aspects of their motivation, such as self-efficacy and goal orientation for particular learning activities, changed over time in the new academic setting with its unique configurations of contextual features. Thus, macro-(sociocultural background) and micro-(classroom and instructional practices) level contextual influences, as well as students' subjective perceptions of these, crucially shape the development of motivational orientations and processes that consequently become congruent with the particular cultural-educational context.

Kitayama and Uchido (2005) stated that motivation is universal in all cultures and the ways in which it is constructed depends on cultural values and characteristics. However, while traditional theories of human motivation have postulated personal agency and self-determination as the central drive of all human actions, a growing body of culture-based research suggests that these fundamental assumptions may not be as relevant among members of more interdependent cultures. As elaborated above, it seems that motivational variations may be best explained in light of prevailing sociocultural factors and the interplay of multiple contextual variables, rather than in terms of static, bipolar cultural dimensions that cannot account

for the dynamic nature of culture and motivation, which constantly change over time and across situations.

In the following section, we briefly illustrate how indigenous psychology research can make a unique contribution to expanding and enriching mainstream theories of motivation through unveiling culturally specific conceptions of learning, motivation, and achievement.

Cultural Research from Within

Two Asian scholars, Li (2002) and Ho (1998), provide examples of cultural research from an indigenous perspective. Li's (2002) research, situated in China and focused on indigenous, traditional conceptions of learning and achievement, proposed a new dimension of motivated learning that cross-cultural research had not captured. Her model of "heart and mind for wanting to learn" stresses that knowledge seeking and the cultivation of lifelong learning is of greater value for Chinese learners than achievement itself. The overall aim of Chinese learners, according to her, is to achieve breadth as well as depth of knowledge and, simultaneously, to personally and morally grow through the learning process. Furthermore, the meanings attributed to success/failure and effort/ability in her model remarkably differ from the commonly shared understanding of these notions in Western models. The difference, according to her, emerges as a result of the dialectical reasoning style characteristic of Chinese culture; thus, failure and success are both perceived as essential components of the learning experience. From this perspective, failure is not perceived negatively, but as a sign to work on and improve particular skills, which in turn will help to achieve success in the future. Consequently, Li's Chinese model of learning stresses the impact of effort to achieve learning outcomes. Innateness of ability is not neglected; however, it is also not viewed as a determinant factor because lack of ability can be compensated by substantial effort (Li, 2002). Interestingly, Li's conception of effort/ability appears consistent with the findings by Hufton *et al.* (2002b) in a Russian educational context.

While Li recommends the accommodation of the Chinese conceptions of learning and achievement in existing frameworks for theory enhancement, she does not claim the universality of these dimensions. In contrast, Ho (1998) proposed methodological relationalism as a universally generalizable approach to consider human relationships as culturally defined and, thus, to interpret individual behavior more accurately. Ho grounds his idea of methodological relationalism and the significance of relational dimensions in Asian social psychology. According to Ho, the methodological relationalism approach is critical to capture the inherently social nature of human actions that are invariably embedded in relational contexts. Ho (1998) recommends person-in-relations and persons-in-relation as

useful, universal units of analysis to acknowledge the impact of relational contexts on the thinking and actions of individuals.

These two constructs, postulated by Ho, explain how cultural research from within can provide a unique insight into indigenous perspectives on learning and motivation. This insight is viewed as a crucial prerequisite for understanding and explaining cross-cultural variations because it unveils how participants themselves construct and value the phenomena under investigation.

Research Shortcomings and Future Directions

Focusing on the five key motivational constructs was useful for providing an illustrative picture of the challenging and enriching contribution of culture in motivation research. It highlighted how current theories of motivation, mainly developed and validated through Western lenses, are not always useful for explaining and predicting motivational orientations in cultural milieus different from those in which they originated. For example, the relatively simplistic, dichotomous conceptualizations of ability/effort, success/failure, and extrinsic/intrinsic motivation appear unable to adequately explain cultural variations across contexts, and therefore need to be reexamined and redefined. The moderating influence of culture on relationships between variables was revealed, and emic positions were found useful to highlight how variables can differ across cultures.

The research on culture and motivation is characterized by a number of methodological shortcomings. These involve a dominance of single-context studies that use cross-sectional designs and rely on questionnaire data. Such studies have limited potential to capture the complexity and interacting nature of personal, cultural, and situational influences on emerging motivational patterns. In contrast, studies that involve longitudinal designs and that investigate motivation in multiple contexts have greater potential to reveal stability and change in motivational orientations over time, within and across contexts (e.g., Volet, 1999). This is important because it is only if cultural variations are found across diverse contexts that we can be confident that these represent a real main effect. Acknowledging the moderating influences of culture on the relationships between variables as well as the dynamic, situated nature of both culture and motivation require research designs and methodologies that reflect these more complex conceptualizations. Such approaches also have the potential to unveil intra-individual as well as within-culture differences (e.g., Bandura, 2002).

The current undue reliance on questionnaire data will also need to be reexamined and complemented by other approaches. Van de Vijver and Poortinga (2002) have

argued that multiple approaches are needed to progress with the daunting task of studying the relationship of development and cultural context. They recommend combining qualitative and quantitative research methodologies as well as the use of a variety of models, ranging from simple main effects to dynamic interaction models. A similar approach would benefit research on culture and motivation. There is little doubt that questionnaire data are not well suited to the task of capturing the significance of culturally constituted contexts for individuals' processes of understanding and meaning making of the phenomena under investigation (e.g., Elliott and Bempechat, 2002). Two possible ways of empirically examining and improving the validity of survey methodology in cross-cultural contexts are Rasch measurement (e.g., Andrich, 1978) and cognitive pretesting (Karabenick *et al.*, 2007). While the former allows for the measurement of qualitative differences in the responses of individuals to psychometric items, the latter empirically examines the extent to which respondent's interpretations of items are consistent with the meaning that a particular construct is intended to capture.

To conclude, both sociocognitive and situated theoretical perspectives on motivation have stressed that achievement, self-efficacy, and agency beliefs (to name just a few) are socially constructed. Social constructions include cultural constructions, and both call for qualitative, ethnographic, and emic approaches to complement traditional survey methodologies. As discussed in this article, culture has already made a challenging and enriching contribution to motivation research. Further enhancement in our understanding of the respective contribution of culture and contextual dimensions on motivation requires continuous reexamination of theoretical assumptions within and across multiple cultural contexts, and a variety of research approaches that reflect the complex and dynamic nature of culture and motivation.

See also: Achievement Goal Theory: Definitions, Correlates, and Unresolved Questions; Intrinsic and Extrinsic Motivation; Sociocultural Issues in Motivation.

Bibliography

- Adamopoulos, J. and Lonner, W. J. (2001). Culture and psychology at a crossroad: Historical perspective and theoretical analysis. In Matsumoto, D. (ed.) *Handbook of Culture and Psychology*, pp 11–34. New York: Oxford University Press.
- Andrich, D. (1978). Application of a psychometric rating model to ordered categories which are scored with successful integers. *Applied Psychological Measurement* 2(4), 581–594.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist* 37(2), 122–147.
- Bandura, A. (2002). Social cognitive theory in cultural context. *Applied Psychology: An International Review* 51(2), 269–290.
- Bempechat, J. and Drago-Severson, E. (1999). Cross-national differences in academic achievement: Beyond etic conceptions of

- children's understandings. *Review of Educational Research* **69**(3), 287–314.
- Chiu, C., Hong, Y., and Dweck, C. S. (1997). Lay disposition and implicit theories of personality. *Journal of Personality and Social Psychology* **73**(1), 19–30.
- Deci, E. L. and Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.
- Deci, E. L. and Ryan, R. M. (2000). The “what” and “why” of goal pursuit: Human needs and the self determination theory of behavior. *Psychology Inquiry* **11**(4), 227–268.
- Elliott, J. G. and Bempechat, J. (2002). The culture and contexts of achievement motivation. *New Directions in Child and Adolescent Development* **96**, 7–26.
- Fiske, A. P., Kitayama, S., Markus, H. R., and Nisbett, R. E. (1997). The cultural matrix of social psychology. In Gilbert, D. T., Fiske, S., and Lindzey, G. (eds.) *Handbook of Social Psychology*, 4th edn., pp 915–981. New York: McGraw-Hill.
- Freeman, M. A. and Bordia, P. (2001). Assessing alternative models of individualism and collectivism: A confirmatory factor analysis. *European Journal of Personality* **15**, 105–121.
- Greenfield, P. M. (1997). Culture as a process: Empirical methods for cultural psychology. In Berry, J. W., Poortinga, Y. H., and Pandey, J. (eds.) *Handbook of Cross-Cultural Psychology: Theory and Methods*, vol. 1, pp 301–346. Needham Heights, MA: Allyn and Bacon.
- Heine, S. J., Kitayama, S., Lehman, D. R., et al. (2001). Divergent consequences of success and failure in Japan and North America: An investigation of self-improving motivations and malleable selves. *Journal of Personality and Social Psychology Monograph Supplement* **81**(4), 599–615.
- Heine, S. J. and Lehman, D. R. (1997). The cultural construction of self-enhancement: An examination of group-serving bias. *Journal of Personality and Social Psychology* **72**(6), 1268–1283.
- Hernandez, M. and Iyengar, S. S. (2001). What drives whom? A cultural perspective on human agency. *Social Cognition* **19**(3), 269–294.
- Ho, D. Y. F. (1998). Interpersonal relationships and relationship dominance: An analysis based on methodological relationalism. *Asian Journal of Social Psychology* **1**(1), 1–16.
- Hofstede, G. (1980). *Culture's Consequence: International Differences in Work-Related Values*. Beverly Hills, CA: Sage.
- Hong, Y. Y. and Chiu, C. Y. (2001). Toward a paradigm shift: From cross-cultural differences in social cognition to social-cognitive mediation of cultural differences. *Social Cognition* **19**(3), 181–196.
- Hong, Y., Chiu, C., Dweck, C. S., Lind, D. M., and Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology* **77**(3), 588–599.
- Huften, N., Elliott, J. G., and Illushin, L. (2002a). Achievement motivation across cultures: Some puzzles and their implications for future research. *New Directions for Child and Adolescent Development* **2002**(96), 65–86.
- Huften, N., Elliott, J., and Illushin, L. (2002b). Educational motivation and engagement: Qualitative accounts from three countries. *British Educational Research Journal* **28**(2), 265–289.
- Iyengar, S. S. and Lepper, M. R. (1999). Rethinking the value of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology* **76**(3), 349–366.
- Kagitcibasi, C. (1996). *Family and Human Development Across Cultures: A View from the Other Side*. Mahwah: Lawrence Erlbaum.
- Kaplan, A. and Maehr, M. L. (1999). Enhancing the motivation of African American students: An achievement goal theory perspective. *Journal of Negro Education* **68**(1), 23–41.
- Karabenick, S. A., Wolley, M. E., Friedel, J. M., et al. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist* **42**(3), 1–17.
- Kashima, Y. (2001). Culture and social cognition: Toward a social psychology of cultural dynamics. In Matsumoto, D. (ed.) *Handbook of Culture and Psychology*, pp 325–360. New York: Oxford University Press.
- Katz, I. and Assor, A. (2007). When choice motivates and when it does not. *Educational Psychology Review* **19**(4), 429–442.
- Kelly, G. A. (1955). *The Psychology of Personal Constructs*. New York: Norton.
- Kim, U. and Berry, J. W. (1993). *Indigenous Psychologies: Experience and Research in Cultural Context*. Newbury Park, CA: Sage.
- Kitayama, S. (2002). Culture and basic psychological processes – toward a system view of culture: Comment on Oyserman et al. (2002). *Psychological Bulletin* **128**(1), 89–96.
- Kitayama, S., Matsumoto, H., Markus, H. R., and Norasakkunkit, V. (1997). Individual and collective processes in the construction of the self: Self-enhancement in the United States and self-criticism in Japan. *Journal of Personality and Social Psychology* **72**(6), 1245–1267.
- Kitayama, S. and Uchido, Y. (2005). Interdependent agency: An alternative system for action. In Sorrentino, R., Cohen, D., Olsen, J. M., and Zanna, M. P. (eds.) *Culture and Social Behavior: The Ontario Symposium*, vol. 10, pp 137–164. Mahwah, NJ: Erlbaum.
- Li, J. (2002). A cultural model of learning: Chinese “Heart and mind for wanting to learn” *Journal of Cross-Cultural Psychology* **33**(3), 248–269.
- Maehr, M. L. (1974). Culture and achievement motivation. *American Psychologist* **29**(12), 887–896.
- Maehr, M. L. and Yamaguchi, R. (2001). Cultural diversity, student motivation and achievement. In Salili, F., Chiu, C., and Hong, Y. (eds.) *Student Motivation: The Culture and Context of Learning*, pp 123–148. New York: Plenum.
- Markus, H. R. and Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review* **98**(2), 224–253.
- Markus, H. R. and Kitayama, S. (2004). Models of agency: Sociocultural diversity in the construction of action. In Murphy-Berman, V. and Berman, J. J. (eds.) *Cross-Cultural Differences in Perspectives on the Self: Nebraska Symposium on Motivation*, vol. 49, pp 1–57. Lincoln, NE: Nebraska University Press.
- McClelland, D. C. (1961). *The Achieving Society*. Princeton, NJ: Van Nostrand.
- McClelland, D. C. (1965). Toward a theory of motive acquisition. *American Psychologist* **20**(5), 321–333.
- McInerney, D. M. (2003). Motivational goals, self-concept, and sense of self – what predicts academic achievement? Similarities and differences between Aboriginal and Anglo Australians in high school settings. In Marsh, H. W., Craven, R., and McInerney, D. M. (eds.) *International Advances in Self Research*, pp 315–346. Greenwich, CT: Information Age.
- McInerney, D. M., Roche, L. A., McInerney, V., and Marsh, H. W. (1997). Cultural perspective on school motivation: The relevance and application of goal theory. *American Educational Research Journal* **34**(1), 207–236.
- Nelson, G. F., O'Mara, A. J., McInerney, D. M., and Dowson, M. (2006). Motivation in cross-cultural settings: A Papua New Guinea psychometric study. *International Education Journal* **7**(4), 400–409.
- Pintrich, P. R. (2000). Educational psychology at the millennium: A look back and a look forward. *Educational Psychologist* **35**(4), 221–226.
- Ramburuth, P. and McCormick, J. (2001). Learning diversity in higher education: A comparative study of Asian international and Australian students. *Higher Education* **42**, 333–350.
- Roth, G., Assor, A., Kanat-Maymon, Y., and Kaplan, H. (2006). Assessing the experience of autonomy in new cultures and contexts. *Motivation and Emotion* **30**(4), 365–376.
- Salili, F. (1994). Age, sex and cultural differences in the meaning and dimensions of achievement. *Personality and Social Psychology Bulletin* **20**(6), 635–648.
- Salili, F., Chiu, C., and Lai, S. (2001). The influence of culture and context on students' motivational orientation and performance. In Salili, F., Chiu, C., and Hong, Y. (eds.) *Student Motivation: The Culture and Context of Learning*, pp 221–247. New York: Plenum.
- Salili, F., Maehr, M. L., and Gillmore, G. (1976). Achievement and morality: A cross-cultural analysis of causal attribution and evaluation. *Journal of Personality and Social Psychology* **33**(3), 327–337.
- Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues* **50**(4), 19–45.

- Sinha, D. (1998). Changing perspectives in social psychology in India: A journey toward indigenization. *Asian Journal of Social Psychology* **1**, 17–31.
- Triandis, H. (1994). Theoretical and methodological approaches to the study of collectivism and individualism. In Kim, U., Triandis, H., Kagitcibasi, C., Choi, S.-C., and Yoon, G. (eds.) *Individualism and Collectivism: Theory, Method and Applications: Cross-Cultural Research and Methodology Series*, pp 41–51. Thousand Oaks, CA: Sage.
- Turner, J. C. and Meyer, D. K. (2000). Studying and understanding the instructional contexts of classrooms: Using our past to forge our future. *Educational Psychologist* **35**, 69–85.
- Van de Vijver, F. J. R. and Poortinga, Y. H. (1982). Cross-cultural generalization and universality. *Journal of Cross-Cultural Psychology* **13**(4), 387–408.
- Van de Vijver, F. J. R. and Poortinga, Y. H. (2002). On the study of culture in developmental science. *Human Development* **45**, 246–256.
- Volet, S. (1999). Motivation within and across cultural–educational contexts: A multi-dimensional perspectives. *Advances in Motivation and Achievement* **11**, 185–231.
- Volet, S. (2001). Understanding learning and motivation in context: A multidimensional and multi-level cognitive–situative perspective. In Jaervelae, S. and Volet, S. (eds.) *Motivation in Learning Contexts: Theoretical Advances and Methodological Implications*, pp 57–82. Oxford: Pergamon and EARLI.
- Volet, S. E. (2004). Understanding learning and motivation in context: What do alternative research traditions have to offer? In Wosnitza, M., Frey, A., and Jaeger, R. S. (eds.) *Lernprozess, Lernumgebung und Lerndiagnostik. Wissenschaftliche Beiträe zum Lernen im 21. Jahrhundert. (Learning Process, Learning Environment and Learning Diagnostics. Scientific Contributions to Learning in the 21st Century)*, pp 276–293. Landau, Germany: VEP.
- Watkins, D. (2000). Learning and teaching: A cross-cultural perspective. *School Leadership and Management* **20**(2), 161–173.
- Yang, K. S. (2000). Monocultural and cross-cultural indigenous approaches: The royal road to the development of a balanced global psychology. *Asian Journal of Social Psychology* **3**, 241–263.
- Yu, A. B. and Yang, K. S. (1994). The nature of achievement motivation in collectivist societies. In Kim, U., Triandis, H. C., Kagitcibasi, C., Choi, S. C., and Yoon, Y. (eds.) *Individualism and Collectivism: Theory, Method, and Applications*, vol. 18, pp 239–250. Newbury Park, CA: Sage.
- Zusho, A. and Pintrich, P. R. (2003). A process-oriented approach to culture: Theoretical and methodological issues in the study of culture and motivation. In Salili, F. and Hoosain, R. (eds.) *Teaching, Learning and Motivation in a Multicultural Context*, pp 33–65. Greenwich, CT: IAP.

Further Reading

- Kagitcibasi, C. (1996). *Family and Human Development Across Cultures: A View from the Other Side*. Mahwah, NJ: Erlbaum.
- Matsumoto, D. (ed.) (2001). *Handbook of Culture and Psychology*. New York: Oxford University Press.
- Miller, J. G. (2002). Bringing culture to basic psychological theory – beyond individualism and collectivism: Comment on Oyserman *et al.* (2002). *Psychological Bulletin* **128**(1), 97–109.
- Salili, F., Chiu, C., and Hong, Y. (eds.) (2001). *Student Motivation: The Culture and Context of Learning*. New York: Plenum.

Early Social Development and Schooling

G M van der Aalsvoort, Utrecht University, Utrecht, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Conflict management skills – Conflicts are part of friendships between peers. Conflict management skills are associated with forming and maintaining friendships, and therefore with increased social understanding.

Context – The immediate setting, such as neighbourhood, school, and community, and different types of interpersonal relationships, that is, peers, family, and school-based.

Friendships – A special form of peer relationship when children engage in many more positive interactions with each other that are characterized by talking, cooperation, and positive affect, than they would with other peers.

Joint Attention – The ability to share attention with others.

Peer relationships – The degree that children are liked by children in the larger peer group, and to peer friendships, that apply to dyadic relations between two children.

Social Competence – A combination of effective use of social skills, absence of maladaptive behaviour, positive relations with others, and accurate, age-appropriate social cognition, or social understanding.

Social Development – A set of behavioural patterns, feelings, attitude, and understanding of the relationship to others that becomes apparent over time. It is the result of both intrinsic child characteristics and socialization.

Socialization – Guiding the child in such a way that reciprocal compliance, a child that is willing to be guided by parent directives and a parent that aims at fostering socialization, results in a social competent child.

Social play – Collaborative interactions of peers during play.

Social understanding – Accurate, age-appropriate social cognition.

Temperament – The whole of an endogenously organized personality trait that has a persuasive effect on the quality and frequency of social exchanges and interpersonal relations.

Transactional model – A continuous attunement between intrinsic child characteristics and elements

in the child's environment, such as circumstances in the home, the school, the neighbourhood, and former life experiences.

Introduction

This article deals with social development, which refers to a set of behavioral patterns, feelings, attitude, and understanding of the relationship to others that becomes apparent over time. We consider social development as a result of transactions between child characteristics and environment. This transactional view on social development means that we look at social skills as abilities that develop during socialization. Processes of socialization are related to the child's parents and the cultural habits of the parents, and contextual influences, for example, family relations, and neighborhood. Social development starts at birth when the child's skills to share attention with others increases. The results of joint attention and increasing skills to use language allow the child to share actions and emotions with his caregivers. The child's interactions with caregivers gradually develop to more sophisticated verbal and nonverbal interactions. These allow the child to turn his attention to other adults and peers gradually. Both the results of socialization and the child's characteristics, especially temperament, add to increased social development. When the child goes to school, the results of socialization up to that point in time, and the child's achieved social skills further influence his social understanding to become a socially competent member of his culture.

The very first sign of social awareness is sharing attention. This capability is typically human as opposed to chimpanzees, and it refers to the first social behaviors that children show shortly after birth. The meaning of this skill is therefore the first topic that we present as one of the main elements of social development with regard to children aged 0–6 years. The other elements are development of peer interactions, social play, and social competence with regard to school. We based the choice for these essential factors upon extensive literature searches related to issues on social development. First, we define each of the items, and we describe recent theoretical insights, which are based upon recent findings of studies related to the theme. We consider temperament as a factor, which

relates to all of the items, and we therefore present it in a separate paragraph. We will start with a historical overview of interest in social development to offer a frame for the article content, thereafter we present the elements of social development one after the other. In the conclusive remarks, we draw attention once again to early social development as a concept that is related to effective schooling.

Historical Overview of Interest in Social Development

Social development is the result of both intrinsic child characteristics and socialization. Both elements have received considerable attention over the years. The theoretical considerations and studies to test hypotheses derived from these theories reflect a growing insight in the way children become socialized given the genetic traits, and the environment in which they grow up. A first concept of socialization was that children are reinforced to form habits by their parents: As the child observes his parents, and as his parents deliver continuous feedback on intentional and nonintentional behavior while being together, the child develops increasing insight in how to behave. Later on, psychodynamic views came into focus. Children are active learners as they initially experience control by their parents but they gradually develop internal social control. This growing ability to understand social behavior, in turn, affects how they perceive social events. Studies on visualizing social behaviors showed that both parenting style and attachment affect socialization, and that the reciprocal nature of these factors leads to ongoing changes over time. Next, a new element of socialization came into view. This was the capacity to self-regulate, and it added a new series of studies to understand processes of socialization. It became clear that both child characteristics and environmental influences decide the outcome of socialization. Parents need to guide their child in such a way that reciprocal compliance, a child that is willing to be guided by parent directives, and a parent that aims at fostering socialization, results in a socially competent child. Consequently, research methodologies became more sophisticated to address bidirectionality in the study of social behaviors, especially with regard to the sequential properties of social interactions between child and parent. This sense of bidirectionality guides the description of the elements considered as main factors in early social development and schooling.

Main Factors in Early Social Development and Schooling

Joint Attention

The ability to share attention with others is a developmental task that precedes social competence. Behaviors

that are indicative of the capacity for joint attention begin to emerge about 3 months after birth. Infants at that age already show visual preference for members of their own race compared to members of a different race. They also favor the sound of the native language when spoken to, since they tend to look longer at the person who speaks their language, and they are attracted to movements of the human face. More elaborated and differentiated forms of attention develop in the next 15 months. The tendency to fixate social stimuli and to engage in social interaction is socially motivated: The child seeks a broader context that provides comfort, security, and satisfaction. Perceiving what another person is looking at facilitates referential communication, which includes three distinct skills of joint attention: responding to joint attention, or proto-imperative behavior; initiating joint attention, or proto-declarative behavior; and initiating behavior regulations. Successful joint attention is socially rewarding for the child, although individual children may differ in tendency to share positive affect with others. Moreover, eye contact marks awareness that others have powers of perception as well.

Joint attention is a skill that develops gradually based upon child and caregiver characteristics. Being motivated for interpersonal and emotional dyadic sharing is reciprocally rewarding, and the relationship between the child's daily social interactions and later social understanding is strong. Studies about joint attention by 12-month-old infants showed that the more frequent these children responded to, and initiated joint attention with an unfamiliar tester, the more often parents reported that their child showed optimal social competence at 30 months of age. As these results came forward even after excluding measures of cognition and language, it indicates that joint attention as a skill is an important predictor of later social competent behavior.

Studying joint attention is often included when investigating how children develop shared intentionality as part of cognitive development. In several studies, young children and chimpanzees have been compared with regard to the way they develop skills to share intentions. This information seems important to understand the meaning of developing joint attention for the human species. Four skills were compared between young children and young chimpanzees: The first one is gaze following and joint attention. Young children and chimpanzees know what others see, but only children attempt to share attention with others, and they know that they are doing this whereas chimpanzees do not. A second skill is that both young children and chimpanzees gesture in order to communicate with others, but only children understand its underlying cooperative motive. Group activity and collaboration can be observed with young children and chimpanzees, but only children actively encourage the adult to assist them to take part in the activity. Finally, social learning occurs only with young children and an adult as he or she demonstrates or teaches them. However, adult

chimpanzees do not demonstrate things to their youngsters. Instead, the young animals observe and copy the adult's behavior. Joint attention as an expression of shared intentionality therefore, is "a direct expression of the biological adaptation that enables children to participate in the cultural practices around them" (Tomasello and Carpenter, 2007). At the same time, these cultural practices alert to child to "specific conditions under which the built-in default assumption of universal knowledge and omniscient other minds must be suspended or inhibited" (Gergely *et al.*, 2007). Joint attention, therefore, is a first step in social referencing which develops gradually and is experience driven. Improved social referencing allows children to build peer relationships.

Peer Relationships

From a relational perspective, the child's early relationships with his parents initiate the child into a system of reciprocal and mutual interactions. These affect the child's connection to his parents and later on to his brothers and sisters, and to his peers. Three features of relational quality seem to be meaningful to successful socialization: warmth, security, and mutual reciprocity (Laible and Thompson, 2007). Secure attachment, or a positive relational quality between parent and child, predicts the child's competence in peer relations. As secure attachment is a main result of early parenting, adults play a significant role in the socialization of their child. Children become interested in peers throughout the second year. At first, these interactions mainly include looking, offering, and taking toys. By the 14th month, toddlers start using objects to lengthen the time of interacting, and to increase opportunities to reciprocal communication. During the preschool years, children's improved opportunities to use language for communication with peers allow longer and more intensive interactions, which enhance the development of interpersonal understanding. This understanding characterizes peer relationships and it proceeds along four levels that can be observed when pairs of preschool children are invited to play together. The first level is that immediate will is pursued with little consideration for any other factors; unilaterally expressing one's own needs or wishes is the second level; then reciprocal strategies which includes coordinating perspectives through persuasion occur, and finally collaboration by developing mutual understanding of both partners' perspective shows up. Social sensitivity of preschool peers is not the same between pairs of children. A range of developmental levels occurs while watching preschoolers playing together. Moreover, children differ remarkably in the way they profit from each other's social developmental level.

Peer relations refer to peer acceptance, which is the degree that children are liked by children in the larger peer group, and to peer friendships, that apply to dyadic relations between two children. Core aspects of the

dynamics of interaction processes in peer interaction are the child's effectiveness in social interaction, and his social power, or status of popularity. Both elements of peer interaction are reflected by sociometric status. Measuring a child's status allows expressions of the child's social competence: The children are nominated by their peers to derive scores for social preference and social impact. Next, sociometric categories are computed to decide the category that a child belongs to. These are popular, rejected, neglected, controversial, or average. The findings of recent studies question whether sociometric status is an attribute of individual children. Studies on revealing how social development of rejected children unfolds over time show that rejected children who are not aggressive are less at risk for poor social development outcomes than their peers who are both rejected and aggressive. The same accounts for the other categories of sociometric status. Therefore, an increased emphasis on the fit of the child's characteristics in relation to the expectations of the larger group is taken into account in studies on identifying processes of peer functioning.

A special form of peer relationship is friendship. Children can have dyadic relationships with peers, but it is called friendship only when they engage in many more positive interactions with each other, and these are characterized by more talking, cooperation, and positive affect, than those of other peers. Friendships start early in childhood. Toddlers can be observed in showing specific interest in one child, and spending time with that child by frequent initiatives to share playthings, being close to one another, and share laughs and actions. During preschool, friendships show up in sustained coordinated play and role play. In kindergarten, playing together is still an important part of friendship but there are also more often verbal interactions. These are characterized by many positive and animated exchanges including sharing of preferred activities. Making and keeping friends requires several social skills, such as perspective taking, self-regulation, and understanding desires and beliefs of the other. Having friends is not a critical determinant of social competence, however, as it also depends upon the child's other close relationships in the family, or the level of stressfulness in the child's life. For children who grow up in unstable families, friendships often serve as socialization agents that are indispensable for healthy social development. School is an important place to develop friendships. Studies show that adjustment to school requirements is most important in the early school years "when classrooms are relatively self-contained and being able to interact effectively with a variety of peers is significant" (Gifford-Smith and Brownell, 2003). The role of context, therefore, is more important than earlier presumed.

Conflicts are part of friendships between peers. Conflict management skills are associated with forming and maintaining friendships, and therefore with increased social understanding. The skills develop from simple

skills, which is insisting to one's own strategy or using a strategy that does not allow compromise and conciliation, to elaborate skills. In that case, a child reasons with his peer and attempts to compromise by giving the peer information about his perspective, and offering resolutions that the peer may find acceptable. A study with 5-year children from preschool showed that they use simple assertion and compliance as main strategies in a conflict situation. However, these strategies are situated: simple assertion is mainly used when the opponent insists, and compliance is used more often when the opponent uses a justification, or gives an alternative proposal. Children apply aggressive assertion when their opponent uses physical aggression. Conflict strategies are, however, part of the social skills that children develop while interacting with peers. Episodes of physical aggression decline over the period from 2 to 4 years of age. Use of aggressive behavior is, therefore, an expression of age-related behaviors that are part of becoming more socially competent. These skills are used in relationship to the peer they are interacting with. It follows that being able to develop relationships with peers allows children to enjoy playing together with peers, and to develop further understanding of one's own mind and that of others during social play.

Social Play

Social play refers to collaborative interactions of peers during play. The interactions are the result of productive engagement through the sharing of knowledge, and they provide opportunities to engage in role taking and social perspective taking. These activities lead to adjustments in inter-individual cooperation. As a result, children jointly develop rules that guide their activity, as peers are often approximately equal in status and competence during social play. Acquiring social skills that result in successful interactions during social play is primarily a developmental task. Social play develops from including social bids, such as smiling to the partner and smiling back, to role reversals in which children take over the role of their play partner, to pretend-play when children act as if, and finally, to meta-play when the children discuss the next step of a story line, and discuss changes of roles. The lowest level of social play occurs by 13 months, and the highest level can be observed from 30 months and onward. The more children are experienced to play with peers, the higher is their level of engagement in social play. There is also a significant effect of age with respect to length of sequences of social play. A longer history of shared social play leads to more successful role play. Moreover, child-directed play or free play in groups without external interventions leads to significant longer social play.

There are, however, individual differences in the degree to which children initiate social play, and are willing to take part in play. Child characteristics, such as

temperament, and parental influences, for example, parenting style, and cultural background, influence whether and how children play together. Experience in social play also increases a child's ability to interact with peers while playing together. The more partners are involved in social play, and treat it as a joint act and understanding, the more likely they will learn from it, resulting in positive adjustment, and consequently, improved social competence of each of them. By the time a child visits school 5 days a week, skills that allow him to understand and cooperate with peers and his teacher in the classroom offer opportunities to respond to social demands successfully.

Temperament

Temperament is visible from the child's earliest years. It is defined as the whole of an endogenously organized personality trait that has a persuasive effect on the quality and frequency of social exchanges and interpersonal relations. There are clear associations between temperament and immediate and later adjustment. Three dimensions of temperament are widely accepted: the first dimension is reactivity, which refers to irritability or distress to limitations, and distress to novelty; the second one is self-regulation consisting of effortful control of attention and of emotions; and the final one is sociability, or the tendency to approach novel situations. Temperament determines a child's adjustment as it is moderately stable over time. It contributes strongly to the development of abilities to cope with emotions; it directly and indirectly affects social behavior; it is an antecedent of collaboration quality, and is gender specific. Four models explain the developmental processes through which temperament affects social development. The first one is that temperament has direct linear effects on social development (e.g., since the child is upset easily by novel situations, he tends to withdraw from them; therefore, the child misses out opportunities to practice social skills that foster peer acceptance and play with peers). In the second model, temperament has an indirect effect on social development (e.g., the child's mother is distressed easily by novel situations, and consequently tends to keep her child away from these situations, which in turns affects the child's behavior toward novel situations). The third model refers to an interaction between temperament and social development, and includes a goodness-of-fit hypothesis (e.g., the child's mother is highly sociable while the child is easily distressed by novel situations. The possibility that the child is overexposed to situations that unnerve him is therefore big and affects the child's emotional stability in a negative way). A transactional model of the relationship between temperament and social development refers to a continuous attunement between intrinsic child characteristics and elements in the child's environment, such as

circumstances in the home, the school, the neighborhood, and former life experiences. In these continuous and long-lasting transactions, temperament is either a risk or a protective factor for adequate social development.

Social Competence in School

The results of social development by the time that children enter preschool and kindergarten are reflected in skills that children apply in social interactions with peers and adults surrounding them. We refer to these skills as a part of social competence. Adequate social competence is defined as a combination of effective use of social skills, absence of maladaptive behavior, positive relations with others, and accurate, age-appropriate social understanding. Social competence refers to “a long term characteristic as opposed to social effectiveness which accounts for short term interactions mainly” (Steenbeek and Van Geert, 2007). Socially competent children succeed in integrating the behavior of the self with others during interactions, and they are able to self-monitor and correct errors during interactions. The higher a child’s social competence is, the more positive his developmental outcomes are, since successful interpretation and response to other’s affective communications, elicits positive interactions with peers and others.

School settings require specific skills related to social competence. These are engaging in prosocial interactions, regulating behavior to complement that of others, and delaying personal gratification. Moreover, children’s relationships with teachers are less personal and intimate than those of their parents, and finally, evaluations of children’s academic and social skills are ongoing (Wentzel and Looney, 2007). Social competence in school is related to the school as a context that affords skills to be and to become skilled as a social individual. An example of the relevance of context is that, especially for younger children, mixed age groups in preschool are beneficial in affecting their abilities to communicate, and to socialize with classmates. This advantage disappears by the age of 4–5 years. This suggests that older children profit more from same-age experiences than younger children, as these allow sharing of thoughts and activities that stabilize ongoing relationships with peers. Social competence, therefore, is an ability that is contingent upon opportunities and affordances that allow a child to become and stay socially integrated as well as self-assertive.

Task behavior is a specific form of social competence as it refers to being able to listen and follow instructions, show adequate behavior in the classroom, and staying task oriented. It is a skill that improves by practicing. It follows that children who have attended childcare are more experienced to interact with adults who are not their caregivers, tend to be more skilled to communicate with peers, and more easily show task behavior. Moreover,

children who like school at the start of entering school are more receptive to the role of being a student, and they more often express behaviors related to being a good student, such as complying with classroom rules, and responsibilities as a student. Classmates are important sources of information about ways to be socially effective.

Teachers play a central role in structuring learning environments that afford social goals for the students in their class. Preschool programs that offer combinations of child-initiated and teacher-initiated programs are more successful in fostering social competence. Teachers that apply a combined approach in their classroom are more sensitive to children who are not yet able to self-regulate activities. These teachers compensate for differences between children, and promote social skills by teaching them explicitly to children who tend to avoid situations in which they are to use social skills. Teachers can promote social competence by educating children about emotions. They can tell stories in which the characters experience conflicts, invite them to use feeling words that express the character’s feelings, and discuss strategies that regulate these feelings in appropriate ways to deal with the conflict.

Conclusive Remarks

This article dealt with social development as a set of behavioral patterns, feelings, attitude, and understanding of the relationship to others that become apparent over time. We described social development from a transactional developmental perspective to allow inclusion of both processes of socialization and of development of social skills. We identified four main elements in early social development starting at birth when the child’s skills to share attention with others increases as a very first sign of social awareness. The other elements we presented were development of peer interactions, social play, and social competence related to school. We based the choice for these essential factors upon extensive literature searches related to issues on social development. As we considered temperament as a factor, which relates to all elements of social development, we described this child characteristic in a separate paragraph. In the final paragraph on social competence, we again emphasized the important role of context for social skills developing in the school.

See also: Children’s Friendship; Early Childhood Curriculum and Developmental Theory; Early Childhood in Post-Modern Cultures: Thoughts and Some Concerns; Learning Through Play; Social Aspects of Collaborative Learning.

Bibliography

- Gergely, G., Egyed, K., and Király, I. (2007). On pedagogy. *Developmental Science* **10**, 139–146.
- Gifford-Smith, M. E. and Brownell, C. A. (2003). Childhood peer relationships: Social acceptance, friendships, and peer networks. *Journal of School Psychology* **41**, 235–284.
- Laible, D. and Thompson, R. A. (2007). Early socialization. A relationship perspective. In Grusec, J. E. and Hastings, P. D. (eds.) *Handbook of Socialization*, pp 181–207. New York: Guilford.
- Steenbeek, H. W. and Van Geert, P. (2007). An empirical validation of a dynamic systems model of interaction: Do children of different sociometric statuses differ in their dyadic play? *Developmental Science* **11**, 253–258.
- Tomasello, M. and Carpenter, M. (2007). Shared intentionality. *Developmental Science* **10**, 121–125.
- Wentzel, K. R. and Looney, L. (2007). Socialization in school settings. In Grusec, J. E. and Hastings, P. D. (eds.) *Handbook of Socialization*, pp 382–403. New York: Guilford.

Further Reading

- Blair, K. A., Denham, S. A., Kochanoff, A., and Whipple, B. (2004). Playing it cool: Temperament, emotion regulation, and social behavior in preschools. *Journal of School Psychology* **42**, 419–443.
- Broadhead, P. (2006). Developing an understanding of young children's learning through play: The place of observation, interaction and reflection. *British Educational Research Journal* **32**, 191–207.
- Carpendale, J. and Lewis, C. (2006). *How Children Develop Social Understanding*. Malden, MA: Blackwell.
- DeLisi, R. and Golbeck, S. L. (1999). Implications of Piagetian theory for peer learning. In O'Donnell, S. M. (ed.) *Cognitive Perspectives on Peer Learning*, pp 3–37. Mahwah, NJ: Erlbaum.
- Göncü, A. (1999). *Children's Engagement in the World*. New York: Cambridge Press.
- Grusec, J. E. and Hastings, P. D. (eds.) (2007). *Handbook of Socialization*. New York: Guilford.

- Lloyd, B. and Howe, N. (2003). Solitary play and convergent and divergent thinking skills in preschool children. *Early Childhood Research Quarterly* **18**, 338–355.
- Lutz, M. N., Fantuzzo, J., and McDermott, P. (2002). Multidimensional assessment of emotional and behavioral adjustment problems of low-income preschool children: Development and initial validation. *Early Childhood Research Quarterly* **17**, 22–41.
- Maccoby, E. E. (2007). Historical overview of socialization. In Grusec, J. E. and Hastings, P. D. (eds.) *Handbook of Socialization*, pp 13–41. New York: Guilford.
- McClelland, M. M. and Morrison, F. J. (2003). The emergence of learning related social skills in preschool children. *Early Childhood Research Quarterly* **18**, 206–224.
- Pianta, R. C. and Stuhlman, M. W. (2004). Conceptualizing risk in relational terms: Associations among the quality of child–adult relationships prior to school entry and children's developmental outcomes in first grade. *Educational and Child Psychology* **21**, 32–46.
- Sanson, A. and Hemphill, S. A. (2004). Connections between temperament and social development. *Social Development* **13**, 142–170.
- Sheridan, S. M., Buhs, E. S., and Warnes, E. D. (2003). Childhood peer relationships in context. *Journal of School Psychology* **41**, 285–292.
- Spinrad, T. L., Eisenberg, N., Harris, E., et al. (2004). The relation of children's everyday nonsocial peer play behavior to their emotionality, regulation, and social functioning. *Developmental Psychology* **40**, 67–80.
- Szewszyk-Sokolowsky, B. and Bost, K. K. (2005). Attachment, temperament, and preschool children's peer acceptance. *Social Development* **14**, 379–397.
- Van der Aalsvoort, G. M. and Van der Leeden, R. (2008). Offering opportunity to play by students developing at-risk: Better quality of play and school performance? *Tijdschrift voor orthopedagogiek, kinderpsychiatrie en klinische kinderpsychologie* **33**(1), 19–33.
- Vaughan Van Hecke, A., Mundy, P. C., Acra, C. F., et al. (2007). Infant joint attention, temperament, and social competence in preschool children. *Child Development* **78**, 53–69.

Emotion in Educational Contexts

P A Schutz, P D Quijada, S de Vries and M Lynde, University of Texas at San Antonio, San Antonio, TX, USA

© 2010 Elsevier Ltd. All rights reserved.

Emotions in Educational Contexts

Educational activities tend to generate affective experiences for those involved in such activities. As such, emotions permeate educational contexts and affect everyone in the schooling process. A recent example is the accountability movement in the United States that has brought with it an increase in the use of high-stakes testing and associated emotional experiences (pleasant and or unpleasant) for students, teachers, and parents (Nichols and Berliner, 2007). Specifically relating to teachers, the current high attrition and early retirement rates of teachers in a number of countries around the world have been linked to unpleasant affective states such as anger, stress, anxiety, and burnout (Wilhelm *et al.*, 2000; Wisniewski and Gargiulo, 1997). Thus, understanding the nature of emotions in educational settings may be a key to successful educational experiences for students, teachers, and parents.

The increased interest by scholars in issues related to emotions in education has resulted in a growing number of educational journals devoting special issues to this topic (e.g., *Educational Psychologist*, 2002; *Learning and Instruction*, 2005; *Teaching and Teacher Education*, 2006; *Educational Psychology Review*, 2007) and several recently published books regarding emotions in education (Boler, 1999; Schutz and Pekrun, 2007; Zembylas, 2005). This suggests a perceived need to better understand the nature of emotions in education.

In this article, we describe some current inquiries related to emotions in education. We begin by discussing the concepts we used to describe emotional experiences and explicate some of the current research on student emotions. This is followed by a discussion of the current work concerning teachers' emotional experiences. We conclude with suggestions for potential future directions for research on emotions in education.

What Are Emotions?

Due to the variety of terms used in the academic literature and everyday speech, we must first clarify our approach on emotional experiences. We currently conceptualize affective experiences as being organized into three interrelated constructs – affective tendencies, core affects, and emotional experiences.

Affective Tendency

Affective tendency is a predisposition toward certain ways of being in the world (Lazarus, 1991; Rosenberg, 1998). This propensity for particular ways of being tends to develop through transactions among a variety of sources such as socialization, individual beliefs, temperament, approach/avoidance motives, and personality. These affective tendencies provide a lens through which individuals view their transactions in the world. Thus, if one is predisposed to view the world as a scary place, one is more likely to interpret various situations as being potentially frightening and may have more frequent fear-type emotional experiences.

Core Affect

Core affect is essentially how we feel at any particular point in time. Researchers in this area suggest that core affect is fundamentally a combination of two types of feeling continuums – valiance (pleasant to unpleasant) and arousal (low activation to high activation) (Linnenbrink, 2007; Russell, 2003; Russell and Barrett, 1999). Thus, at any particular point in time we may feel calm (low activation and pleasant), tense (high activation and unpleasant), or happy (pleasant, high activation). Mood, which is one example of core affect, can be viewed as a prolonged core affect without an object. Russell (2003) suggests that core affect is object free but can become directed at an object through attributions or appraisals. In these instances, the affect experience is usually described as an emotional experience.

Emotional Experience

Schutz *et al.* (2006) described emotional experiences as “socially constructed, personally enacted ways of being that emerge from conscious and/or unconscious judgments regarding perceived successes at attaining goals or maintaining standards or beliefs during transactions as part of social-historical contexts” (p. 344). There are two key aspects of this definition that are relevant to our discussion here.

Appraisals and emotional experiences

First, emotions involve judgments or appraisals (Boekaerts, 2007; Pekrun *et al.*, 2007; Schutz *et al.*, 2007). As suggested above, core affect tends to become an emotional experience

via appraisal or attribution regarding what is happening during a particular episode. Individuals' goals, values, and beliefs, as well as their social network are the referent points used to judge where they are in relation to where they want to be (Carver and Scheier, 2000; Schutz and DeCuir, 2002; Schutz and Davis, 2000; Powers, 1971). These goals, values, and beliefs represent ways individuals, as members of social groups, position themselves during a particular life event (Boekaerts, 2007; Ford, 1992; Markus and Nurius, 1986; Schutz *et al.*, 2001).

Thus, appraisals involve students' or teachers' perceptions of how the pursuit of a goal progresses during an academic transaction. In most cases, these judgments tend to occur outside of the students' or teachers' awareness, yet these judgments are seen as being key to the emotional experience (Frijda, 1993; Lazarus, 1991; Pekrun *et al.*, 2007; Schutz and Davis, 2000; Smith, 1991). Lazarus (1991, 1999) made a distinction between primary and secondary appraisals. He indicated that primary appraisals deal with how important the outcome is perceived to be by the teacher or student. For example, if a student does not see homework as important, there is probably little potential for emotions related to homework. For primary appraisal, Lazarus (1991, 1999) talks about judgments related to goal relevance (i.e., is it important to the student's goals?), goal congruence (i.e., is it going how the student hoped it would?), and the type of ego involvement (i.e., how much of the student's self or one's identity is involved?).

Secondary appraisals are the judgments a student or teacher would make about their potential to handle the particular situation. Key secondary appraisals are related to agency or control (Lazarus, 1999; Pekrun *et al.*, 2007; Schutz and Davis, 2000) and problem efficacy (Bandura, 1997; Schutz and Davis, 2000), or what Lazarus (1991) referred to as coping potential (i.e., the person's confidence about their ability to handle the situation). These secondary appraisals help differentiate among potential emotions. Consequently, a situation that has been appraised as goal important (i.e., "It's vital for my students to do well on this standardized test") and goal incongruent (i.e., "My students did really badly on that test"), with secondary appraisals of self-blame (i.e., "I knew I should have done a better job of teaching") may result in shame (Turner and Waugh, 2007). Alternatively, the same situation with secondary appraisals of other blame (i.e., "These standardized tests don't measure what my students know") may result in anger.

Social construction of emotional experiences

Schutz *et al.* (2006) also suggest that emotions are socially constructed and emerge from particular social-historical contexts. In other words, emotions are relational, such that emotional experiences do not exist as exclusive features of a person or of an environment (Denzin, 1984; Lazarus, 1991; Lazarus and Folkman, 1984; Meyer and Turner,

2007; Op 'T Eynde *et al.*, 2007). Particular emotional experiences involve person-environment transactions. Consequently, in most cases there are both a social dimension of an emotional experience, and the person's enactment of the particular emotional way of being.

Emotional experiences are also influenced by the particular social-historical context in which the transaction occurred (DeCuir-Gunby and Williams, 2007; Markus and Kitayama, 1994; Ratner, 2007; Schutz *et al.*, 2007; Stearns and Stearns, 1985). For example, the emotional experience we label guilt is based upon developing knowledge of the ethical and legal values of a culture (Ratner, 2007; Weiner, 2007). As such, the student or teacher must adopt the cultural value or belief of personal responsibility in order for a guilt emotional experience. This suggests that the appraisals and attributions that students and teachers make are reflective of the social-historical context in which they and their social groups are embedded.

As such, when there are emotional experiences in the classroom, they reflect person-environment transactions as well as the social-historical contexts in which those transactions occur. Illustrative of this concept is an emotional school situation described in a study by Decuir-Gunby and Williams (2007). After attending an assembly featuring a local civil rights leader, some students had intensive emotional experiences, including guilt, anger, and sadness. One student, in particular, described her classmate as being "upset, and crying. She was upset that he was saying she was a bad person simply because she was white". This student was experiencing an emotion (sadness) as a result of both an environmental transaction (listening to the civil rights leader's speech) and her own social-historical context (the perceived meaning of being white).

Educational Research on Students' Emotional Experiences

Researchers' efforts to study students' emotional experiences can be broadly organized into two general areas – students' self-regulation and motivation and the related area of students' learning and achievement.

Students' Self-Regulation, Motivation, and Emotional Experiences

In the motivation and self-regulation literature, students' emotional experiences emerge during activities that are embedded within personal as well as social-historical contexts. This personal history includes students' goals, values, and beliefs regarding academics in general as well as their beliefs about the subject and the particular activity (Op 'T Eynde *et al.*, 2007; Schutz *et al.*, 2006).

For example, a student may possess a general affective tendency to avoid a challenging task for fear that it may

show ignorance relative to his/her peers (i.e., performance avoidance goal orientation) (Elliot and Pekrun, 2007; Linnenbrink, 2007). In this situation, the student may begin an activity fearful and anxious and, as a result, may attend more closely to aspects of the activity that may be interpreted as a lack of understanding. Thus, through fear and anxiety the student may look to the teacher for help (i.e., external regulation) (Linnenbrink, 2007; Pekrun *et al.*, 2004).

Researchers have also found that pleasant emotional experiences tend to be associated with perceived self-regulation, intrinsic motivation, and the tendency to engage in meta-cognitive or self-regulatory strategies (Linnenbrink, 2007; Meyer and Turner, 2007; Pekrun *et al.*, 2002). This suggests that emotional experiences (pleasant or unpleasant) tend to provide multidirectional feedback for other processes involved in self-regulation (Turner and Waugh, 2007). For example, having successful experiences that students attribute to themselves may result in pleasant emotions, which can increase interest and focus on the activity (Ainley, 2007; Efklides and Petkaki, 2005). This link between emotional experience and self-regulatory processes was demonstrated by Boekaerts (2007), who found that pleasant affect was associated with increases in competence and value judgments as well as effort, whereas unpleasant affect lowered competence and value judgments and was associated with less effort.

The influence of emotional experiences on self-regulation and other motivational and learning processes has resulted in researchers investigating the process involved in emotional regulation. Emotional regulation involves various processes that are directed at monitoring, evaluating, and modifying emotional experiences (Schutz and Davis, 2000; Schutz and DeCuir, 2002; Thompson, 1994). In the area of testing, Schutz *et al.* (2004) were able to explain 55% of the variance in the Worry subscale of the Revised Test Anxiety scale (Benson, 1998) with the emotional regulation during test-taking scale (ERT). In a follow-up study Schutz *et al.* (2007) accounted for 56% of the variance in pleasant and 87% of unpleasant test emotions. These results suggest what students do during tests or other classroom activities may influence their emotional experience.

Students' Learning, Achievement, and Emotional Experiences

Interest in relationships among learning, achievement, and emotional experiences has a long history. The focus of much of this research has been on the test anxiety construct. Test anxiety has been associated with a number of adverse effects on students: poor cognitive performance, scholastic underachievement, psychological distress, and ill health (Zeidner, 2007). Zeidner (2007) goes as far as to suggest that:

Indeed, a student's performance on a classroom exam may be as much an indicator of the students' ability to cope with high levels of evaluative stress and anxiety in the classroom as a reflection of the ability or achievement the exam aims at measuring. Thus, the measurement of any particular ability or proficiency will be confounded with anxiety. (Zeidner 2007: 160)

Recently, researchers interested in emotions in education have also begun to examine other emotional experiences. They paint an interesting, yet complex, picture of the role of emotional experiences in the classroom. As expected, in general, pleasant emotional experiences tend to be associated with the use of more effective learning strategies, task-focused attention, and academic performance (Pekrun *et al.*, 2002, 2004), whereas unpleasant emotional experiences, such as anxiety, boredom, and hopelessness tend to reduce academic performance (Pekrun *et al.*, 2002; Zeidner, 2007). However, other research suggests that unpleasant emotional experiences can also improve achievement. In one such study, Pekrun and Hofmann (1996) found that for some students anxiety was not related to improvement in achievement-related agency over time but for others, there was a positive relationship.

Others have obtained similarly inconsistent findings. Linnenbrink (2007) indicated that general measures of pleasant affect tended to be consistently unrelated to students' learning. This suggests that emotional experiences do not occur in a personal vacuum. As indicated, emotional experiences emerge out of the appraisals students make regarding where they perceive themselves to be in relationship to their goals, values, and beliefs. As a result, what is key to the nature of a particular emotional experience, as well as the potential learning and self-regulation during and after an activity, are the students' perceptions of control and efficacy as well as the attributions they make during and after the activity (Pekrun *et al.*, 2007; Weiner, 2007).

Thus, emotional experiences occur in real time and therefore, are susceptible to changes that occur during an event (Turner and Waugh, 2007). As such, an emotional experience can change very quickly (e.g., joy can rapidly become frustration or anger), which may explain some of the inconsistent findings reported. This also reinforces the need for researchers to continue looking at emotional experiences as they are occurring (Ainley, 2007; Meyer and Turner, 2007; Op 't Eynde *et al.*, 2007; Turner and Waugh, 2007).

Educational Research on Teachers' Emotional Experiences

Researchers who foreground teacher emotions in educational contexts have tended to focus on issues related to teachers' efforts to cope with unpleasant, difficult

emotional experiences, and the related topic of emotional rules and labor.

Unpleasant Emotional Experiences

Researchers suggest that emotions tend to be pervasive within the service of teaching (Meyer and Turner, 2007; Williams *et al.*, in press). Meyer and Turner (2007), demonstrated that teachers who care tend to be sensitive and use humor to create classrooms where students are less likely to avoid tasks and increase performance. However, similar to most professions that require a positive work relationship with the public, teachers must juggle their own emotions, while trying to contend with their students' needs and attend to other administrative duties. As such, teaching is an emotionally charged situation and, if not regulated appropriately, may lead to anxiety, depression, anger, or simply becoming burned out on teaching.

Liljestrom *et al.* (2007) reported that some teachers become angry when required to carry out additional administrative duties that interfere with their curriculum and teaching, although the subsequent behavioral reactions varied across individuals. For instance, some teachers will rebel by becoming less cooperative with administrators, while some advocated for policy changes to higher-level administrators. Still others choose to leave the profession entirely.

The prevalence of unpleasant emotions was demonstrated by Sutton (2007) who, in a diary of teachers' anger and frustration, found that each week teachers reported a median of two anger experiences and seven emotional experiences labeled as frustration. Even more illustrative of the nature of teaching was that many of the teachers reported that these were experiences that sometimes lasted for more than an hour (Sutton, 2007). Reports such as these have resulted in researchers focusing on emotional labor and the emotional rules associated with teaching.

Emotional Rules and Labor

While researchers agree that some emotions are universal, it is important to keep in mind that emotions are socially constructed and therefore behavioral, emotional, and physiological reactions tend to fluctuate across cultures. To understand these social-historical influences, researchers have suggested that there are normative emotional rules that indicate what emotional displays are or are not appropriate (Hargreaves, 2001; Liljestrom *et al.*, 2007; Morris and Feldman, 1996; Zembylas, 2007). Thus, when examining normative rules for teachers, it is important to keep in mind that institutions such as schools embody norms that have been established historically and culturally. Because of the normative nature of these rules, teachers often feel they must conform to socially acceptable standards of displaying emotions or be perceived as deviant. This can be illustrated

by comparing a teacher who teaches to the test and shows neutrality of emotions to a teacher that embodies pedagogical practices that engage students in authentic learning. Depending on the particular school context, each of these teachers could be considered deviant.

Normative rules and consequences related to emotions are not only affected by environmental factors, but also reinforce a reciprocal relationship. Continuing with the above example, a teacher who deviates from the institutionally constructed norm of emotional neutrality and teaching to a standardized test, may be emotionally affected by feeling shame and doubt about how they are teaching. Williams *et al.* (2008) found evidence of this when they found that many teachers deem it necessary to suppress or avoid the display of emotions in the classroom. As such, teachers' emotional experiences are tied to construction of their own teachers' identities (Schutz *et al.*, 2007). The dynamic and transactional nature of emotional experience is influenced by current teachers' identities; this may restrict their affective displays based on how they want to be seen, they will tend not to exhibit emotions that contradict this image.

The idea that teachers are expected to display different emotions in particular ways has been associated with the idea of emotional labor (Hargreaves, 2001; Hochschild, 1990; Morris and Feldman, 1996; Zembylas, 2007). Morris and Feldman (1996) define emotional labor "as the effort, planning, and control needed to express organizationally desired emotions during interpersonal transactions" (p. 987). Emotional labor, under some circumstances, has been associated with emotional exhaustion (a key component of burnout), job satisfaction, and adverse health symptoms (Hochschild, 1990; Morris and Feldman, 1996) thereby demonstrating again the importance of understanding the nature of emotional experiences in the classroom.

Where Do We Go from Here?

As indicated, we currently view emotional experiences as transactional processes that occur in real time. In addition, these transactional processes are embedded within particular, ever-changing social-historical contexts. Given this perspective, it would seem important to approach the study of emotions in education using multiple methods and methodologies (Pekrun and Schutz, 2007) at the various levels of research (Schutz *et al.*, 2004).

One crucial problem area in need of exploration is how issues of class, race, and ethnicity, and gender transact with emotional experiences within the educational setting. In many societies, such as the United States, issues such as race and gender often impact the learning and development that occurs within the school context (DeCuir-Gunby and Williams, 2007; Liljestrom *et al.*, 2007; Steele, 1997). Further exploration of these issues

may help demonstrate the impact class, race, and ethnicity, and gender prejudices have in the educational setting, as well as provide an avenue for researchers and educators to help address these issues.

A second area in need of attention is the area of emotional development. Psychologists have researched the nature of emotional development (Aviles *et al.*, 2006); however, developing an understanding of how unpleasant and complex emotional experiences develop within the school context will require more contextual and domain specific approaches.

A related topic is the development of ways for teachers to scaffold emotions in the classrooms (Meyer and Turner, 2007). It seems clear that by scaffolding and encouraging students' emotions in class, a teacher can help to forge a trusting bond with the student, demonstrate and model respect for the student, and facilitate better engagement with learning tasks. Therefore, more research and eventually implementation research is needed in this area.

Another area involves the emotional labor associated with teaching. Emotional labor has been associated with job satisfaction, adverse health symptoms, and emotional exhaustion – key components of burnout and teacher attrition (Schutz *et al.*, 2007; Zembylas, 2007). A better understanding of the nature of emotional labor and the emotional display rules for teachers will be important to facilitate teachers' emotional well being.

The final area related to teaching is the area of skills at reading emotions of others or other directedness (Nowicki and Carton, 1993; Snow *et al.*, 1996). As mentioned earlier, for the most part, goals and our self-directed attempts at accomplishing them are predominantly social endeavors. They emerge from transactions in a social environment and we pursue them within our social-historical context. The expression of emotions and our ability to read others' emotions are key components of the self-regulation process. For example, the ability of a teacher to tell when students are anxious or frustrated about a task or not understanding the questions, may be the first step in helping students to learn how to regulate those emotions. This is an area where there has been very little inquiry but which has tremendous potential to help the educational process.

Conclusions

Emotions are an integral part of the educational activity setting and, therefore, an understanding of the nature of emotions in the school context is an important goal. In this article, we have attempted to examine some current research on emotional experiences in education. As this area matures, we see the continual examination of the theoretical foundations of emotions in education as fundamental to healthy development of our program of inquiry as well as for that of the larger community.

Bibliography

- Ainley, M. (2007). Being and feeling interested: Transient state, mood, and disposition. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion in Education*, pp 141–157. New York: Elsevier.
- Aviles, A. M., Anderson, T. R., and Davila, E. R. (2006). Child and adolescent social-emotional development within the context of school. *Child and Adolescent Mental Health* **11**(1), 32–39.
- Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York: Freeman.
- Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practices* **17**, 10–17.
- Boekaerts, M. (2007). Understanding students' affective processes in the classroom. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion in Education*, pp 33–52. San Diego, CA: Elsevier.
- Boler, M. (1999). *Feeling Power: Emotions and Education*. New York: Routledge.
- Carver, S. C. and Scheier, M. F. (2000). On the structure of behavioral self-regulation. In Boekaerts, M., Pintrich, P. R., and Zeidner, M. (eds.) *Handbook of Self-Regulation*, pp 41–84. San Diego, CA: Academic Press.
- DeCuir-Gunby, J. T. and Williams, M. R. (2007). The impact of race and racism on students' emotions: A critical race analysis. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion in Education*, pp 199–213. San Diego, CA: Elsevier.
- Denzin, N. K. (1984). *On Understanding Emotion*. San Francisco, CA: Jossey-Bass.
- Efklides, A. and Petkaki, C. (2005). Effects of mood on students' metacognitive experiences. *Learning and Instruction* **15**, 415–431.
- Elliot, A. and Pekrun, R. (2007). The hierarchical model of approach–avoidance achievement motivation and emotion. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion in Education*, pp 53–69. San Diego, CA: Elsevier.
- Ford, M. E. (1992). *Motivating Humans: Goals, Emotions and Personal Agency Beliefs*. Newbury Park, CA: Sage.
- Frijda, N. H. (1993). The place of appraisal in emotion. *Cognition and Emotion* **7**, 357–387.
- Hargreaves, A. (2001). Emotional geographies of teaching. *Teachers College Record* **103**, 1056–1080.
- Hochschild, A. R. (1990). Ideology and emotion management: A perspective and path for future research. In Kemper, T. D. (ed.) *Research Agendas in the Sociology of Emotions*, pp 234–266. Oxford, UK: Basil Blackwell.
- Lazarus, R. S. (1991). *Emotion and Adaptation*. New York: Oxford University Press.
- Lazarus, R. S. (1999). *Stress and Emotions: A New Synthesis*. New York: Springer.
- Lazarus, R. S. and Folkman, S. (1984). *Stress, Appraisal, and Coping*. New York: Springer.
- Liljestrom, A., Roulston, K., and deMarrais, K. (2007). 'There's no place for feeling like this in the workplace': Women teachers' anger in school settings. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion in Education*, pp 267–283. San Diego, CA: Elsevier.
- Linnenbrink, E. A. (2007). The role of affect in student learning: A multi-dimensional approach to considering the interaction of affect, motivation, and engagement. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion in Education*, pp 101–118. San Diego, CA: Elsevier.
- Markus, H. R. and Kitayama, S. (1994). The cultural construction of self and emotion: Implications for social behavior. In Kitayama, S. and Markus, H. R. (eds.) *Emotion and Culture: Empirical Studies of Mutual Influence*, pp 89–130. Washington, DC: APA.
- Markus, H. and Nurius, P. (1986). Possible selves. *American Psychologist* **41**, 954–969.
- Meyer, D. K. and Turner, J. C. (2007). Scaffolding emotions in classrooms. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion in Education*, pp 235–249. San Diego, CA: Elsevier.
- Morris, J. A. and Feldman, D. C. (1996). The dimensions, antecedents, and consequences of emotional labor. *Academy of Management Review* **21**(4), 986–1010.

- Nichols, S. L. and Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.
- Nowicki, Jr. S. and Carton, J. (1993). The measurement of emotional intensity from facial expressions. *Journal of Social Psychology* **133**, 749–750.
- Op 'T Eynde, P., De Corte, E., and Verschaffel, L. (2007). Students' emotions: A key-component of self-regulated learning? In Schutz, P. A. and Pekrun, R. (eds.) *Emotion in Education*, pp 179–198. San Diego, CA: Elsevier.
- Pekrun, R., Frenzel, A. C., Goetz, T., and Perry, R. P. (2007). The control value theory of achievement emotions: An integrative approach to emotion in education. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion in Education*, pp 9–32. San Diego, CA: Elsevier.
- Pekrun, R., Goetz, T., Perry, R. P., Kramer, K., and Hochstadt, M. (2004). Beyond test anxiety: Development and validation of the test emotions questionnaire (TEQ). *Anxiety, Stress and Coping* **17**, 287–316.
- Pekrun, R., Goetz, T., Titz, W., and Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of quantitative and qualitative research. *Educational Psychologist* **37**, 91–106.
- Pekrun, R. and Hofmann, H. (1996). Affective and motivational processes: Contrasting interindividual and intraindividual perspectives. *Paper presented at the Annual Meeting of the American Educational Research Association*. New York.
- Pekrun, R. and Schutz, P. A. (2007). Where do we go from here? Implications and future directions for inquiry on emotions in education. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion in Education*, pp 303–321. San Diego, CA: Elsevier.
- Powers, W. T. (1971). *Behavior: The Control of Perception*. Chicago, IL: Aldine.
- Ratner, C. (2007). A macro cultural-psychological theory of emotions. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion in Education*, pp 85–100. San Diego, CA: Elsevier.
- Rosenberg, E. L. (1998). Levels of analysis and the organization of affect. *Review of General Psychology* **2**(3), 247–270.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review* **110**(1), 145–172.
- Russell, J. A. and Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology* **76**, 805–819.
- Schutz, P. A., Benson, J., and DeCuir, J. T. (2008). Approach/avoidance motives, test emotions, and emotional regulation during testing. *Anxiety, Stress and Coping: An International Journal* **21**(3), 263–281.
- Schutz, P. A., Cross, D. I., Hong, J. Y., and Osbon, J. N. (2007). Teacher identities, beliefs and goals related to emotions in the classroom. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion in Education*, pp 215–233. San Diego, CA: Elsevier.
- Schutz, P. A., Crowder, K. C., and White, V. E. (2001). The development of a goal to become a teacher. *Journal of Educational Psychology* **93**, 299–308.
- Schutz, P. A. and Davis, H. A. (2000). Emotions and self-regulation during test taking. *Educational Psychologist* **35**, 243–255.
- Schutz, P. A. and DeCuir, J. T. (2002). Inquiry on emotions in education. *Educational Psychologist* **37**, 125–134.
- Schutz, P. A., DiStefano, C., Benson, J., and Davis, H. A. (2004). The development of a scale for emotional regulation during test taking. *Anxiety, Stress and Coping: An International Journal* **17**, 253–269.
- Schutz, P. A., Hong, J. Y., Cross, D. I., and Osbon, J. N. (2006). Reflections on investigating emotions among social-historical contexts. *Educational Psychology Review* **18**, 343–360.
- Schutz, P. A. and Pekrun, R. (2007). *Emotion in Education*. New York: Elsevier.
- Smith, C. A. (1991). The self, appraisal and coping. In Snyder, C. R. and Forsyth, D. R. (eds.) *Handbook of Social and Clinical Psychology: The Health Perspective*, pp 116–137. Elmsford, NY: Pergamon.
- Snow, R. E., Corno, L., and Jackson, D. (1996). Individual differences in affective and conative functions. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 243–310. New York: MacMillan.
- Stearns, P. N. and Stearns, C. Z. (1985). Emotionology: Clarifying the history of emotions and emotional standards. *American Historical Review* **90**, 813–836.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist* **52**(6), 613–629.
- Sutton, R. E. (2007). Teachers' anger, frustration, and self-regulation. In Schutz, P. A. and Pekrun, R. (eds.) *Emotions in Education*, pp 259–274. San Diego, CA: Elsevier.
- Thompson, R. A. (1994). Emotional regulation: A theme in search of definition. In Fox A. (ed.) *The Development of Emotional Regulation: Biological and Behavioral Considerations*, pp 25–52. Chicago: The University of Chicago Press.
- Turner, J. E. and Waugh, R. M. (2007). A dynamical systems perspective regarding students' learning processes: Shame reactions and emergent self-organizations. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion in Education*, pp 119–139. San Diego, CA: Elsevier.
- Weiner, B. (2007). Examining emotional diversity in the classroom: An attribution theorist considers the moral emotions. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion in Education*, pp 70–84. San Diego, CA: Elsevier.
- Wilhelm, K., Dewhurst-Savellis, J., and Parker, G. (2000). Teacher stress? An analysis of why teachers leave and why they stay. *Teachers and Teaching* **6**(3), 291–304.
- Williams-Johnson, M. W., Cross, D. I., Hong, J. Y., Aultman, L. P., Osbon, J. N., and Schutz, P. A. (2008). "There is no emotion in math": How teachers approach emotions in the classroom. *Teacher College Record* **110**(8), 1574–1612.
- Wisniewski, L. and Gargiulo, R. (1997). Occupational stress and burnout among special educators: A review of the literature. *Journal of Special Education* **31**(3), 325–346.
- Zeidner, M. (2007). Test anxiety in educational contexts: Concepts, findings, and future directions. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion and Education*, pp 159–177. San Diego, CA: Elsevier.
- Zembylas, M. (2005). *Teaching with Emotion: A Postmodern Enactment*. Greenwich, CT: Information Age.
- Zembylas, M. (2007). The power and politics of emotions in teaching. In Schutz, P. A. and Pekrun, R. (eds.) *Emotion and Education*, pp 285–301. San Diego, CA: Elsevier.

Further Reading

- Efkides, A. and Volet, S. (2005). Emotional experiences during learning: Multiple, situated and dynamic. *Learning and Instruction* **15**, 377–380.
- Linnenbrink, E. L. (2006). Emotion research in education: Theoretical and methodological perspectives on the integration of affect, motivation, and cognition. *Educational Psychology Review* **18**, 307–314.
- Schutz, P. A. and Lanehart, S. L. (2002). Introduction: Emotions in education. *Educational Psychologist* **37**, 67–68.
- van Veen, K. and Lasky, S. (2005). Emotions as a lens to explore teacher identity and change: Different theoretical approaches. *Teaching and Teacher Education* **21**, 895–898.
- Zembylas, M. (2005). *Teaching with Emotion: A Postmodern Enactment*. Greenwich, CT: Information Age.

Empathy

R Arnold, University of Sydney, Sydney, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Attunement – Actively listening to, observing, and focusing on another to resonate with their thoughts and feelings.

Empathy – An act of heartfelt, thoughtful imagination to help determine another's cognitive and affective experience.

Intersubjective experience – Engaging with the internal and expressed thoughts and feelings of others.

Intrasubjective experience – Engaging with one's own internal thoughts and feelings through reflection.

Mirroring – The affirmation of one's feelings and thoughts by another through their words, facial gestures, tone of voice, or other expressive features.

Introduction

For a long time, effective educators and others have known tacitly that the relationships and contexts within which humans learn are significant to the quality of that learning. It is timely to elaborate the nature of such aspects, particularly, but not exclusively, as they relate to school learning. Mindful that much significant learning of language and socialization occurs before children enter formal education, the nature of infant learning is relevant to understand early influences on attitudes to learning and self as a learner. In early childhood, affective and inter-/intrasubjective templates are formed which influence learning readiness and interpersonal responsiveness (O'Connor and McCartney, 2007; Repacholi and Meltzoff, 2007; Stern, 1985; Winnicott, 1965).

Empathy through inter-/intrarelatedness, developed as empathic intelligence (Arnold, 2005), is a useful frame for reflecting how social and emotional experiences can affect pedagogy. In the pedagogical framework of this article, empathy is theorized as a function of mind, brain, and feeling within contexts of care and relatedness. By creating a dynamic between thinking and feeling in a climate perceived as caring by students, teachers can mobilize learners' intrasubjective experiences in the service of learning. Similarly, understanding the nature of empathy and how quality relationships can influence learning intensifies educators' attunement to their own

thoughts and feelings and those of others to whom they relate and influence.

Intrasubjective engagements are the thoughts and feelings which come into play as we try to make sense of experience. For example, understanding ambivalent reactions to a situation, or discerning in order to make an important decision are processes which touch on internalized prior experiences. Since we are all influenced in our psychic development by significant others in life, such as parents, siblings, teachers, and friends, our internalized thoughts and their emotional aura continue to influence present thoughts, feelings, and responses.

Empathic intelligence posits that in educators and those involved in interpersonal and intrapersonal work, particular desirable behaviors, such as attunement, warmth, sensitivity, and capacity for mirroring, can have positive effects on those with whom they engage. Importantly, educators' empathic disposition needs to be matched with other abilities, such as a capacity to engage, an ability to create a dynamic between thinking and feeling, coupled with professional expertise, enthusiasm, and intelligent caring. While empathic intelligence is particularly apt for those involved in teaching, training, and educating, it is equally applicable for those involved in a range of professions engaged in relationships and learning. Education is conceptualized here to include formal and informal teaching and learning in public and private settings across all age ranges. In learning contexts and organizations, which function within a climate of empathic intelligence, participants' tacit abilities to learn and relate effectively can be mobilized.

Some background to the development of empathy is outlined here, but the focus of this article is on empathy and empathic intelligence in teaching and learning. Theorization of the concept of empathy has been influenced by work in disciplines such as philosophy (Buber, 1965; Greene, 1995; Kohlberg, 1978; Noddings, 1984, 1988, 1992; Nussbaum, 1995, 1997; Polanyi, 1959, 1969, 1974, 1983; Verducci, 2000), biology (Darwin, 1965), psychology (Barnes and Thagard, 1997; Barnett, 1987; Bruner, 1972, 1986, 1990; Csikszentmihalyi, 1990; Eisenberg and Strayer, 1996; Feshbach and Feshbach, 1987; Freud, 1922; Gardner, 1983, 1993, 1997; Goleman, 1995; Isen, 1984; Kincheloe *et al.*, 1999; Mayer and Salovey, 1997; Piaget, 1926, 1981; Wadsworth, 1989; Vygotsky, 1978), psychoanalysis (Kohut, 1959, 1971, 1979, 1982, 1985; Winnicott, 1965), and infant studies (Lichtenberg, 1983, 1984; O'Connor and McCartney, 2007; Repacholi and Meltzoff, 2007; Stern, 1985; Trevarthen, 1977, 1979). The relevance of empathy to learning has only recently surfaced in

education though it has been foreshadowed in liberal, democratic, student-centered educational practice, influenced by Dewey (1916, 1963, 1971). The brain-based research of Damasio (1994, 2000, 2003), LeDoux (1992), and Williams (2001) offers important insights into the complexity of cognitive/affective interplays in the mind, signaling that learning is not primarily a cognitive task but very much influenced by social and emotional aspects.

Historical Antecedents

Research on the nature of empathy (Barnett, 1987; Barnes and Thagard, 1997; Davis, 1994; Duan and Hill, 1996; Ekman, 2003; Eisenberg, 1989; Eisenberg and Strayer, 1996; Kohut, 1971, 1982, 1985; Verducci, 2000) serves to highlight some of its conceptual complexity. Some salient features of the concept will be highlighted, particularly as they relate to education as an outcome of interpersonal relatedness.

Apart from a reference in Aristotle's *Rhetoric* to *empathiea*, as Verducci (2000a) points out, empathy first appears in nineteenth-century German esthetics in the work of Friedrich Vischer (1847–1933), and following them, Martin Buber and Theodore Lipps. Friedrich Vischer called empathy as a process of 'symbolic interjection of emotions into objective forms', such as engaged in by viewers of works of art. Verducci (2000: 67) notes that Vischer's son Robert named the process *Einfubling*, which was later translated as empathy. Robert Vischer described the process as projecting one's own life into the lifeless form, "only ostensibly do I keep my own identity... I am mysteriously transformed into this Other" (Friedrich Vischer 1994: 19–20). The philosopher Martin Buber (1965) describes the experience of feeling strong connectedness with subjects or objects,

To glide with one's own feeling into the dynamic structure of an object... to 'transpose' oneself over there and in there. Thus it means the exclusion of one's own concreteness, the extinguishing of the actual situation of life, the absorption in pure aestheticism of the reality in which one participates (Buber, 1965: 97).

Although Buber's description of empathy here is far removed from the concept of empathy in human relatedness, it does describe a capacity for imagination for decentering and opening up to and trusting experience which is part of empathic responsiveness.

Lipps (1851–1914) went further than the Vischers and Buber, saying "empathy is the fact here established, that the object is myself and by the very same token this self of mine is the object... the antithesis between myself and the object disappears, or rather does not yet exist" (Verducci, 2000: 68). He saw the phenomenon of empathy as projective, imaginative, and primarily affective. Verducci (2000: 67) notes that in these early ideas of empathy,

"The seed of this process lies in the imagination, the flower in the viewer's affective life. The imagined mental representation of an object and the viewer's feelings become inseparable." In these rudimentary views of empathy, concentration and the capacity for absorption are also functioning. What is not identified is a capacity for cognitive distancing or awareness of a self engaging in the experience, arguably, necessary predispositions for those engaged in educating others. The experience as described by these philosophers is esthetically pleasing though no judgment or cognition is identified in the process. Nor is empathy associated with unpleasant feeling such as disgust or rage which can be mobilized, along with other unpleasant feelings such as shame and distress, in an empathic response to the events witnessed. As Verducci (2000) notes "empathy as sole grounding for the aesthetic experience cannot account for the phenomenon of aesthetic judgment" (Verducci, 2000: 69), signaling the cognitive aspect present in some empathic experiences.

Empathy in Psychoanalytic Literature

It is surprising that Freud had little explicitly to say about empathy, apart from acknowledging that it "plays the largest part in our understanding of what is inherently foreign to our ego in other people" (Freud, 1922: 69–144). Here Freud seems to be suggesting empathy as an ability to recognize what and which we perceive in others to be different from ourselves. Earlier followers of Freud such as Theodor Reik and Robert Fliess regarded empathy as essentially a cognitive process of identifying the feelings of their therapy patients, while attempting to maintain an emotionally distanced therapeutic relationship with them. Little is reported about patients' responses to what might well seem like dissonant engagements where therapists encourage patients to express or disclose their emotions, while rigorously suppress their own.

In common parlance, empathy means being able to imagine, often intuitively, how the other thinks and feels. Heinz Kohut (1959) described empathy as "vicarious introspection". As a more complex process than identification, vicarious introspection includes both affective attunement and the cognitive capacity to judge best how to respond insightfully and feelingly to the other's psychic state. Kohut argued that empathy should be examined and evaluated 'in an empirical context as a mental activity' (Kohut, 1982: 397). He differentiated two levels of empathy, one an "information-gathering activity" and the other "a powerful bond between people" (Kohut, 1982: 397). This differentiation is a useful precursor to the development of empathy as both a cognitive and an affective process, capable of both developing and reflecting relatedness and understanding. Kohut argues that the presence of empathy is beneficial both in a clinical setting and in human life,

as it “suggests an explanation for certain observable contents and/or sequences of events in man’s (*sic*) psychic life” (Kohut, 1982: 397). Most importantly, the self-soothing and empathic responses of parents and others, promotes the development of a nuclear self into a mature, cohesive self (Kohut, 1971). In psychotherapeutic terms, the task of being empathically responsive is complex. Guided by empathic attunement, the therapist’s capacity for introspection and sensitivity to the state of the therapeutic alliance functions to soothe the client’s feeling state (if he/she is responsive to the therapist’s mood), and to stimulate awareness of masked feelings, in the expectation that understanding and even insight can be developed. Kohut warned, “empathy is used non-intuitively, ploddingly, if you wish, by trial and error. I did not write about empathy as being always correct and accurate” (Kohut, 1982: 396). Arguably, it is not the accuracy, or otherwise, of the empathic response which is necessarily therapeutic as much as the therapist’s perceptible wish to be attuned and responsive which cues patients’ feelings of positive self-regard. Similarly, children with positive attachment experiences with caretakers, particularly mothers, are likely to enter preschool and formal schooling, and are able to develop positive relationships with responsive teachers and peers.

Empathy and Infancy

A number of infant studies (Lichtenberg, 1981, 1983; Stern, 1977; Trevarthan, 1979, 1980) support the view that infants are programmed to seek engagement with others and with their environment. These studies focus attention on the infant’s development of a subjective self. The work of Daniel Stern (1985) on the role of empathy in infancy illuminates the importance of empathic attunement in early learning, emotional development, and socialization. Stern writes about the development of “the domain of inter-subjective relatedness” (Stern, 1985: 27), that ability to experience one’s self as a separate being from others, but as a dependent being too whose sense of self can shift according to others’ responsiveness or otherwise.

Stern theorizes that in dyadic engagements between infant and mother in which the mother’s empathic responsiveness (or otherwise) amplifies, soothes, validates (or deadens) the infant’s affect states, there develop what Stern theorizes as representations of interactions that have been generalized (RIGs), (Stern, 1985: 99). These might be thought of as psychically coded emotional experiences which provide the infant with a unified sense of a core self and awareness of a core other: “The existential bedrock of interpersonal relations” (Stern, 1985: 125). The degree to which the major affect states (interest, joy, surprise, anger, distress, fear, contempt, disgust, and shame) are mirrored back to the child by the mother and other caretakers influences how they are encoded psychically.

Thus codified, they become available as emotional templates influencing responsiveness to events experienced as similar by the child. These social and emotional templates can be reinforced, modulated, and nuanced throughout life, depending upon one’s ability to be self aware and deeply reflective. Emotional response to a stimulus can range from the mild pleasure of cursory attention to the intense pleasure of deep absorption in it, or flow (Csikszentmihalyi, 1990). Angry feelings can range from mild irritation to intense rage or fury, shame can be experienced as mildly embarrassing or searingly humiliating, distress can be momentarily discomforting or completely terrifying.

Stern’s point is that in early childhood and beyond, the way others reflect back or mirror the child’s emotional states, affects the internalization and elaboration of those states. The infant’s initial pleasure in throwing food on the floor is modulated by the caretaker’s response to that behavior, hence relatedness and emotional responsiveness influence learning. Stern’s concept of RIGs is helpful in explaining how early experiences influence children’s attitudes and predispositions to learn. The child for whom learning language, listening to stories, and exploring through play is associated with pleasurable relationships and emotions is likely to be predisposed to learning. Long before children enter school, they have developed feelings about learning and themselves as learners, modulated by their sense of relatedness with those modeling behaviors and attitudes. This is evidenced in the experimental work of Repacholi and Meltzoff (2007) who demonstrated that infants’ actions were influenced by their memory of the affect of adults’ behavior. Furthermore, “the infants’ actions varied as a function of whether they were currently in the Emoter’s visual field” (Repacholi and Meltzoff, 2007: 503). The researchers concluded that infants “learn from emotional eavesdropping and their subsequent behaviour depends on the Emoter’s orientations toward them” (Repacholi and Meltzoff, 2007: 503). By 12 months of age, infants can use emotional cues such as tone of voice and facial expression to regulate their own actions (see Feinman *et al.*, 1992, cited in Repacholi and Meltzoff, 2007: 50 for a review of the literature on infant social referencing).

Teachers and schooling can amplify or dampen feelings of expectation, joy, anxiety, or dread around learning experiences. Empathic teachers can attune to these feelings and relate with their students in constructive and emotionally engaging ways. While students may not necessarily articulate the qualities of effective teachers in complex ways, they will experience them in felt ways depending upon the teachers’ emotional sophistication and attunement to their needs. How do we know this? As research into infant social referencing indicates that infants in their early stage know how to read emotional cues and can use such cues to modify their behavior. Repacholi and Meltzoff’s (2007) study provides important insights into what might be called as infants’ capacities to learn in social contexts through

affective experiences. O'Connor and McCartney's (2007) study of 1364 children from birth to sixth grade found positive associations between quality of teacher-child relationships and achievements. In addition, they found that high-quality teacher-child relationships buffered children from the negative effects on achievement of insecure or other maternal attachment issues. Secure attachments "are marked by high levels of maternal sensitivity and child trust in maternal availability and support" (O'Connor and McCartney, 2007: 343). The effect of quality teacher-child relationships on achievement was mediated through child and teacher behaviors in the classroom.

Several studies show that children develop attachment relationships with teachers which are associated with their long-term achievement (Birch and Ladd, 1997; Pianta, 1994; Pianta and Nimetz, 1991). Peer relationships also influence achievement (Azmitia and Montgomery, 1993; Birch and Ladd, 1996; Ladd *et al.*, 1999). As O'Connor and McCartney (2007) concluded from their study:

Educating teachers as to how to develop high-quality relationships with children may provide strategies for teachers working with children who are at risk of lower levels of achievement. . . Informing teachers as to the influence of relationship quality on classroom behavior may increase teacher awareness and in turn prevent children with low quality relationships from engaging in harmful behaviors in the classroom. (O'Connor and McCartney, 2007: 364)

Narratives and Empathy

Maxine Greene argues passionately and persuasively for the role of imagination in education. She says,

One of the reasons I have come to concentrate on imagination as a means through which we can assemble a coherent world is that imagination is what, above all, makes empathy possible. It is what enables us to cross the empty spaces between ourselves and those we teachers have called "other" over the years. . . of all cognitive capacities, imagination is the one that permits us to give credence to alternative realities. (Greene, 1995: 3)

Narratives offer particular social and emotional experiences which can promote empathic development. Martha Nussbaum (1997) argues that three capacities, above all, are essential for the cultivation of humanity in today's world:

First is the capacity for critical examination of oneself and one's traditions-for living what, following Socrates, we may call "the examined life". . . (Second) Citizens who cultivate their humanity need. . . an ability to see themselves not simply as citizens of some local region or group but also, and above all, as human beings bound to all other

human beings by ties of recognition and concern. . . The third ability of the citizen, closely related to the other two, can be called the narrative imagination. This means the ability to think what it might be like to be in the shoes of a person different from oneself, to be an intelligent reader of that person's story, and to understand the emotions and wishes and desires that someone so placed might have. (Nussbaum, 1997: 9-11)

Nussbaum (1997: p. 14) says empathy means "learning how to be a human being capable of love and imagination." Further, she argues that when a child and a parent learn to tell stories together, sharing a sense of wonder, the child is acquiring essential moral capacities "...stories interact with (children's) own attempts to explain the world and their own actions in it. A child deprived of stories is deprived, as well, of certain ways of viewing other people. For the insides of people, like the insides of stars, are not open to view" (Nussbaum, 1997: p. 89). "The habits of wonder promoted by storytelling thus define the other person as spacious and deep, with qualitative differences from oneself and hidden places worthy of respect" (Nussbaum, 1997: 90).

Narratives (and metaphors/symbolic experiences) can function to develop empathic attunement and perspective taking. Some stories draw us back into the past; others project us into the future. Some illuminate the day; others take us into the darkness of human behavior. The best storytellers lighten and enlighten the paradoxes and ambiguities of life. In narratives, the past and present exist in a seemingly timeless realm in which human characters confront physical challenges, emotional and moral dilemmas in some form, and tacitly suggest to us, as readers, how we might best choose to conduct our lives.

Empathic Intelligence

Empathy is defined here as an ability to understand the thoughts and feelings of self and others. It is a sophisticated ability involving attunement to one's own thoughts and feelings, a capacity to decenter and distinguish between projection and introjection and a capacity for deep introspection. It is, fundamentally, an act of thoughtful, heartfelt imagination whereby understanding of the other is provisional and nonintrusive. Empathy seeks to understand rather than to judge.

Empathic intelligence is more complex than empathy itself but necessarily includes it. Empathic intelligence theorizes how an empathic disposition and associated qualities of enthusiasm, expertise, and capacity for engagement can be enacted in the effective practice of pedagogy. Empathic intelligence articulates aspects of intersubjective and intrasubjective phenomena of pedagogy.

Empathic intelligence is a sustained system of psychic, cognitive, affective, social, and ethical functioning derived from:

- an ability to differentiate self-states from others' states ("who owns what in the interaction");
- an ability to engage in reflective and analogic processing to understand and mobilize a dynamic between thinking and feeling in self and others;
- attunement to the potential integration (or disintegration) of experiences created through that dynamic (recognizing whether learning occurs or not);
- an ability to use mirroring and modeling to good effect; and
- a commitment to the care, well-being, and development of self and others.

In various combinations and with different strengths according to context, these qualities inform the enactment of empathic intelligence. Since these social, emotional, and personal qualities are complex and often subtle in their manifestations, they can be evidenced often in the enthusiasm, capacity to engage, expertise, and empathy of teachers. In a sense, these four qualities provide a shorthand way of thinking about empathic intelligence and recognizing it in professional practice.

It is helpful to conceptualize empathic intelligence as manifest in five behaviors: enthusiasm, expertise, capacity to engage, intelligent caring, and empathy itself. While all these behaviors, and the values and attitudes consistent with them, can help to identify an empathically intelligent educator, it is important to consider them as having some distinctive qualities but sharing in common an informing belief that they function in the service of the other, albeit they can also provide important substance to the professional well-being of their adherents. This is because they are inherently intra/intersubjective behaviors which function in the interpersonal domain of experience and thrive upon mirroring, rigorous self reflection and sensitivity to dynamics.

Enthusiasm

Enthusiasm reflects a sense of inner spiritedness. In a pedagogical context, it can convey pleasure in the act of teaching and influencing, along with pleasure in engaging with students. Such spiritedness can have a positive effect on students, provided it is modeled sensitively and not overwhelmingly. It signals commitment, a sense of purpose and belief in the worth of the persons and enterprise of teaching and learning, which can be emotionally affirming for students and help to engage their attention.

Enthusiasm is easy to recognize in this highly expressive form where the enthusiast is demonstrative physically, and perhaps vocally. In a more subtle form, it can manifest itself

also in quiet concentration upon a task or purpose, or even in a centeredness which can be as captivating to observe as its more demonstrative forms.

Engagement

The skill of engaging others can depend on all kinds of personality factors and even the use or abuse of power and status factors. In this model of empathic intelligence, engagement depends on the ability to create within the other, a justifiable belief in the worth of paying attention to the one seeking the engagement. Experienced writers know the importance of engaging readers in the first couple of paragraphs of a book, mindful that readers are not particularly forgiving or enthralled with writers who fail to draw them into the web of language early on. Surprise, a tone of authority, a description of an interesting time, place, or character, or similar strategies, can work well to keep the reader believing there is something for them in the pages which follow. Similarly, the engaging educator needs to be sophisticated in choosing from a repertoire of strategies, those likely to keep the audience believing in the worth of the engagement. It certainly helps if that expectation of worth in the engagement is honored and fulfilled. Not only does it satisfy the audience, it also provides the engager with positive feedback about the strategies selected, and indeed, about the worth of the enterprise.

Expertise

In teaching and learning, expertise requires a range of educative abilities and qualities (Berliner, 1994, 2002; Bucci, 2003; Hattie, 2003) but particular ones are highlighted here as they relate to empathic intelligence. In the model of empathic intelligence, the expert can mobilize students' imagination and encourage them to look at experience from a number of perspectives in order to hypothesize about cause and effect.

The expert educator is theoretically informed about professional practices and discipline content. This includes understanding child development and the nature of learning. Such educators understand that learning occurs in complex ways and progress is not always straightforward or easily discernible. An expert educator has an expansive repertoire of professional practices, tested and modified through reflection on practice. The expert educator can model best practices and can tolerate own and others' mistakes.

Intelligent Caring

Intelligent caring can identify the right balance between dependency and independence. It offers the other close

attention and sustained engagement, providing an awareness of possible outcomes. It positions the welfare of the other at the heart of the engagement but seeks to mobilize the other's psychic energy, thought, and feeling in constructive ways (Arnold, 2005: 57). It establishes reasonable boundaries to promote autonomy. The function of care in moral development has been extensively theorized (Hoffman, 2000; Noddings, 1984, 1988, 1992).

Empathy and Brain-Based Research

The work of Antonio Damasio (1994, 2000, 2003) provides strong support for the argument here that empathic intelligent teaching mobilizes a dynamic between thought and feeling, recognizing the complementarity of thought and feeling in intellectual and personal development. He argues that human consciousness is actually consciousness of the feeling and experiencing of self.

"Contrary to traditional scientific opinion, feelings are just as cognitive as other percepts" (Damasio, 1994: p. xvii). "(Emotion and feeling) provide the bridge between rational and nonrational processes, between cortical and subcortical structures" (Damasio, 1994: p. 128).

Laboratory work on brain-based research shows that emotion is integral to the processes of reasoning and decision making. As Damasio says,

The neurological evidence simply suggests that selective absence of emotion is a problem. Well-targeted and well-deployed emotion seems to be a support system without which the edifice of reason cannot operate properly. . . . These (results) also made it possible to view emotion as an embodiment of the logic of survival. (Damasio, 2000: 42)

Damasio (1994) illustrates the story of one of his patients, Elliot, whose brain damage affected his normal ability to generate response options to social situation and to consider spontaneously the consequences of particular options. Damasio remarks that when treating Elliott,

I began to think that the cold-bloodedness of Elliot's reasoning prevented him from assigning different values to different options, and made his decision-making landscape hopelessly flat. (Damasio, 1994: 51)

I found myself suffering more when listening to Elliot's stories than Elliot himself seemed to be suffering. In fact, I felt I suffered more than he did just by thinking of those stories. (Damasio, 1994: 44)

Damasio's own philosophical and ethical commitment to both understanding and enhancing the lives of his patients, even those whose mental functioning is very minimal, reflects a professional functioning applicable widely in professional practice. His willingness to engage in a sensitive and feeling way to his patients provided a source of significant scientific insights.

Furthermore, he embodies the capacity to create and to understand a dynamic between thinking and feeling, in a climate of care. The care is characterized as intelligent caring for the well-being of the patient and for the realization of the truth of the situation. It suggests that the phenomena of each individual experience are created uniquely in each particular context. As he remarks,

Were it not for the possibility of sensing body states that are inherently ordained to be painful or pleasurable, there would be no suffering or bliss, no longing or mercy, no tragedy or glory in the human condition. . . . Feelings form the base for what humans have described for millennia as the human soul or spirit. (Damasio, 1994: xvii-xviii)

Steven Rose, a practicing scientist working at the forefront of medical research, provides a similar perspective. He describes his experience of discovering that his feelings interfered with his game of chess, a game which he previously believed involved purely cognitive and logical skills (Rose, 1993: 36). From his scientific work and his reflections upon his own learning experiences, he argues that cognition cannot be divorced from affect, try as one might (Rose, 1993: 36). He continues,

The problems that it (the link between affect and cognition) illuminates are fundamental to my research strategy, just as much as their resolution. However, even today I find myself frequently in danger of forgetting that lesson, though it ought to be fundamental to a strategy for living. (Rose, 1993: 36)

Howard Gardner, arguing for a more complex view of intelligence than commonly prevailed, remarked:

The roots of a sense of self lie in the individual's exploration of his (*sic*) own feelings and in his emerging ability to view his (*sic*) own feelings and experiences in terms of the interpretative schemes and symbol systems provided by the culture. (Gardner, 1983: 294)

Many educators work tacitly in sympathy with these neurological insights but it is affirming and encouraging to professional practice to understand why certain processes are effective. Learning is equally an affective, cognitive, imaginative, and empathic enterprise, and teaching practices need to reflect that. It is important to recognize that it is within the intersubjective and intrasubjective experience of imagined worlds that feeling and reason can psychically engage, deepening one's sense of self and consciousness of others. The empathically intelligent teacher harnesses social and emotional experiences in the service of students' learning.

See also: Affect, Mood and Emotions; Emotion in Educational Contexts; Flow in Education; Motivating Students in Classrooms; Social and Emotional Outcomes of Learning.

Bibliography

- Arnold, R. (2005). *Empathic Intelligence: Teaching, Learning, Relating*. Sydney, NSW: University of New South Wales Press.
- Azmitia, M. and Montgomery, R. (1992). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development* 2, 202–221.
- Azmitia, M. and Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development* 2(3), 202–221.
- Barnes, A. and Thagard, P. (1997). Empathy and analogy. *Dialogue: Canadian Philosophical Review* XXXVI(4), 705–720.
- Barnett, M. A. (1987). Empathy and related responses in children. In Eisenberg, N. and Strayer, N. (eds.) *Empathy and Its Development*, pp 146–162. Cambridge: Cambridge University Press.
- Berliner, D. (2002). Learning about and learning from expert teachers. *International Journal of Educational Research* 35, 463–482.
- Berliner, D. C. (1994). Expertise: The wonder of exemplary performances. In Mangieri, J. N. and Block, C. (eds.) *Creating Powerful Thinking in Teachers and Students: Diverse Perspectives*, pp 161–186. Fort Worth, TX: Holt, Rinehart and Winston.
- Birch, S. H. and Ladd, G. W. (1996). Interpersonal relationships in the school environment and children's early school adjustment: The role of teachers and peers. In Jaana, J. and Wentzel, K. R. (eds.) *Social Motivation: Understanding Children's School Adjustment*, pp 199–225. New York: Cambridge University Press.
- Birch, S. and Ladd, G. (1997). The teacher–child relationship and children's early school adjustment. *Journal of School Psychology* 35(1), 61–79.
- Bruner, J. (1972). *The Relevance of Education*. Harmondsworth: Penguin.
- Bruner, J. (1986). *Actual Minds, Possible Worlds*. Cambridge, MA: Harvard University Press.
- Bruner, J. (1990). *Acts of Meaning*. Cambridge, MA: Harvard University Press.
- Buber, M. (1965). *Between Man and Man*. Smith, R. G. (trans.) New York: Macmillan.
- Bucci, T. T. (2003). Researching expert teachers: Who should we study? *Educational Forum* 68, 82.
- Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper and Row.
- Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Harper Collins.
- Damasio, A. R. (2000). *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. London: Vintage.
- Damasio, A. R. (2003). *Looking for Spinoza: Joy, Sorrow and the Feeling Brain*. Orlando, FL: Harcourt Books.
- Darwin, C. (1965). *The Expression of Emotions in Man and Animal*. Chicago, IL: University of Chicago Press.
- Davis, M. H. (1994). *Empathy: A Social Psychological Approach*. Madison, WI: W.I. Brown and Benchmark.
- Dewey, J. (1916). *Democracy and Education: An Introduction to the Philosophy of Education*. New York: Macmillan.
- Dewey, J. (1963). *Freedom and Culture*. New York: Capricorn Books.
- Dewey, J. (1971). *The Child and the Curriculum and the School and Society*. Chicago, IL: University of Chicago Press.
- Duan, C. and Hill, C. (1996). The current state of empathy research. *Journal of Counselling Psychology* 43(3), 261–274.
- Ekman, P. (2003). *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. New York: Henry Holt.
- Eisenberg, N. (ed.) (1989). *Empathy and Related Emotional Responses*. San Francisco, CA: Jossey-Bass.
- Eisenberg, N. and Strayer, J. (eds.) (1996). *Empathy and Its Development*. Cambridge: Cambridge University Press.
- Feinman, S., Roberts, D., Hsieh, K., Sawyer, D., and Swanson, D. (1992). A critical review of social referencing in infancy. In Feinman, S. (ed.) *Social Referencing and the Social Construction of Reality in Infancy*, pp 3–13. New York: Plenum.
- Feshbach, N. D. and Feshbach, S. (1987). Affective processes and academic achievement. *Child Development* 58, 1335–1347.
- Freud, S. (1922). Group psychology and the analysis of the ego. In Strachey, J. (ed.) *Complete Works of Sigmund Freud*, vol. 18, pp 69–144. London: Hogarth Press.
- Gardner, H. (1983). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.
- Gardner, H. (1993). *Creating Minds*. New York: Basic Books.
- Gardner, H. (1997). *Extraordinary Minds*. London: HarperCollins.
- Goleman, D. (1995). *Emotional Intelligence: It Can Matter More than IQ*. London: Bloomsbury.
- Greene, M. (1995). *Releasing the Imagination: Essays on Education, the Arts and Social Change*. San Francisco, CA: Jossey-Bass.
- Hattie, J. (2003). Teachers make a difference. What is the research evidence? Australian Council for Educational Research, October. http://www.det.nsw.edu.au/proflearning/docs/pdf/qt_hattie.pdf Retrieved from Department of Education and Training NSW website, 11 May 2009.
- Hoffman, M. L. (2000). *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge: Cambridge University Press.
- Isen, A. (1984). Towards understanding the role of affect in cognition. In Wyer, R. and Srull, T. (eds.) *Handbook of Social Cognition*, pp 179–237. Hillsdale, NJ: Erlbaum.
- Kincheloe, J., Steinberg, S., and Villavende, L. (eds.) (1999). *Rethinking Intelligence: Confronting Psychological Assumptions about Teaching and Learning*. London: Routledge.
- Kohlberg, L. (1978). The cognitive-developmental approach to moral education. In Sharf, P. (ed.) *Readings in Moral Education*, pp 36–51. Minneapolis, MN: Winston Press.
- Kohut, H. (1959). Introspection, empathy and psychoanalysis: An examination of the relationship between mode of observation and theory. *Journal of the American Psychoanalytic Association* 7, 459–483.
- Kohut, H. (1971). *The Analysis of the Self*. Madison, WI: International University Press.
- Kohut, H. (1979). The two analyses of Mr Z. *International Journal of Psycho-Analysis* 60(1), 3–27.
- Kohut, H. (1982). Introspection, empathy, and the semi-circle of mental health. *International Journal of Psycho-Analysis* 63, 395–407.
- Kohut, H. (1985). *Self Psychology and the Humanities*. New York: W.W. Norton.
- Ladd, G. W., Birch, S. H., and Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development* 70(6), 1373–1400.
- LeDoux, J. (1992). Emotion and the amygdala. In Aggleton, A. P. (ed.) *The Amygdala: Neurobiological Aspects of Emotion, Memory and Emotional Dysfunction*, pp 339–351. New York: Wiley-Liss.
- Lichtenberg, J. (1983). *Psychoanalysis and Infant Research*. Hillsdale, NJ: Analytic Press.
- Lichtenberg, J., Bornstein, M., and Silver, D. (eds.) (1984). *Empathy 1 and Empathy 11*. Hillsdale, NJ: The Analytic Press.
- Mayer, J. and Salovey, P. (1997). What is emotional intelligence? In Salovey, P. and Sluyter, D. (eds.) *Emotional Development and Emotional Intelligence: Implications for Educators*, pp 3–31. New York: Basic Books.
- Noddings, N. (1984). *Caring: A Feminine Approach to Ethics and Moral Education*. Berkeley, CA: University of California Press.
- Noddings, N. (1988). An ethic of caring and its implications for instructional arrangements. *American Journal of Education* 96(2), 215–230.
- Noddings, N. (1992). *The Challenge to Care in Schools: An Alternative Approach to Education*. New York: Teachers College Press.
- Nussbaum, M. (1995). *Poetic Justice: The Literary Imagination and Public Life*. Boston, CA: Beacon Press.
- Nussbaum, M. (1997). *Cultivating Humanity: A Classical Defence of Reform in Liberal Education*. Cambridge: Harvard University Press.
- O'Connor, E. and McCartney, K. (2007). Examining teacher–child relationships and achievements as part of an ecological model of development. *American Educational Research Journal* 44(2), 340–369.
- Piaget, J. (1926). *The Language and Thought of the Child*. New York: Harcourt Brace.

- Piaget, J. (1981). *Intelligence and Affectivity: Their Relationship during Child Development*. Palo, CA: Annual Reviews.
- Pianta, R. C. (1994). Patterns of relationships between children and kindergarten teachers. *Journal of School Psychology* **32**, 15–32.
- Pianta, R. C. and Nimetz, S. (1991). Relationships between children and teachers: Associations with classroom and home behaviour. *Journal of Applied Developmental Psychology* **12**, 379–393.
- Polanyi, M. (1959). *The Study of Man*. Chicago, IL: University of Chicago Press.
- Polanyi, M. (1969). *Knowing and Being – Essays by Michael Polanyi*. In Greene, M. (ed.). Chicago, IL: University of Chicago Press.
- Polanyi, M. (1974). *Personal Knowledge: Towards a Post-Critical Philosophy*. Chicago, IL: University of Chicago Press.
- Polanyi, M. (1983). *The Tacit Dimension*. Gloucester, MA: Peter Smith.
- Repacholi, B. and Meltzoff, A. (2007). Emotional eavesdropping: Infants selectively respond to indirect emotional signals. *Child Development* **78**(2), 503–521.
- Rose, S. (1993). *The Making of Memory*. Toronto, ON: Bantam Books.
- Stern, D. (1985). *The Interpersonal World of the Infant*. New York: Basic Books.
- Stern, D. (1997). *The First Relationship: Mother and Infant*. Cambridge: Harvard University Press.
- Trevarthen, C. (1977). Descriptive analyses of infant communicative behaviour. In Schaffer, H. R. (ed.) *Studies in Mother–Infant Interaction*, pp 227–270. London: Academic Press.
- Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In Bullowa, M. (ed.) *Before Speech*, pp 321–347. Cambridge: Cambridge University Press.
- Verducci, S. (2000). A conceptual history of empathy and a question it raises for moral education. *Educational Theory* **50**(1), 87–99.
- Vischer, F. (1994). *Empathy, Form and Space: Problems in German Aesthetics, 1873–1893*. In Mallgrave, H. F. and Ikonomou, E. (eds.). Los Angeles, CA: The Getty Center for the History of Art and the Humanities.
- Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Wadsworth, B. J. (1989). *Piaget's Theory of Cognitive and Affective Development*. New York: Longman.
- Williams, L. (2001). *The emotional brain*. Address to Science Forum, University of Sydney, 4th April.
- Winnicott, D. (1965). *The Maturation Processes and the Facilitating Environment: Studies in the Theory of Emotional Development*. London: Hogarth Press/The Institute of Psychoanalysis.

Further Reading

- Feshbach, N. D. (1984). Empathy, empathy training and the regulation of aggression in elementary school children. In Kaplan, R. M., Konecni, V. J., and Novaco, R. (eds.) *Aggression in Children and Youth*, pp 192–208. The Hague: Martinus Nijhoff Publishers.
- Ickes, W. (1997). *Empathic Accuracy*. New York: Guilford.
- LeDoux, J. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Touchstone.
- Olney-Friedrich, M. (2001). Empathy: Recognizing a teachable moment. *Teaching Pre K-8* **32**(3), 60–61.
- Rose, S. (ed.) (1998). *From Brains to Consciousness: Essays on the New Science of the Mind*. Princeton, NJ: Princeton University Press.
- Salovey, P. and Schlesinger, H. J. (1984). The process of empathic response. In Lichtenberg, J., Bornstein, M., and Silver, D. (eds.) *Empathy II*, pp 187–216. Mahwah, NJ: Erlbaum.
- Salovey, P. and Sluyter, D. (eds.) (1997). *Emotional Development and Emotional Intelligence: Educational Implications*. New York: Basic Books.
- Siegel, D. J. (1999). *The Developing Mind: How Relationships and the Brain Interact to Shape Who We Are*. New York: Guilford.
- Thompson, E. (2001). Empathy and consciousness. *Journal of Consciousness Studies* **8**(5–7), 1–32.
- Wertsch, J. (1985). *Vygotsky and the Social Formation of the Mind*. Cambridge, MA: Harvard University Press.

Flow in Education

J A Schmidt, Northern Illinois University, DeKalb, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Flow – A state of optimal experience characterized by total absorption in the task at hand; a merging of action and awareness in which the individual loses track of both time and self.

Brief Definition and History

Nearly every educator holds the primary, though often elusive, goal of facilitating students' deep engagement in learning activities. In recent decades, Mihaly Csikszentmihalyi's flow model has advanced our understanding of the experience of deep engagement, as well as the individual and contextual factors that may promote it. The model has implications for both research and practice, and has seen application in fields, including education, psychology, psychiatry, anthropology, and business. This article provides a description and analysis of the flow model, and summarizes the potential and actual applications of flow in education.

Born out of a desire to understand intrinsically motivated activity, flow refers to a state of optimal experience characterized by total absorption in the task at hand; a merging of action and awareness in which the individual loses track of both time and self. The flow state is experientially positive, and out of the flow experience emerges a desire to replicate the experience. The concept of flow was developed nearly four decades ago by psychologist Mihaly Csikszentmihalyi while he was observing students at an elite art school in the United States. In an effort to understand what led individuals to be passionate about their creative pursuits, Csikszentmihalyi observed the students as they worked on their artistic creations. He quickly noted that students often became so engrossed in their work that they would effectively tune out any outside distractions or obligations. Students would lose track of time, missing meetings or mealtimes, working well into the night, apparently sustaining deep levels of both concentration and enjoyment for extended periods. Almost paradoxically, once the painting or sculpture was completed it seemed to lose its value to the student: The creation was often hauled back to a dark corner of the studio where it would sit indefinitely. It became clear that for these art students, the value in art was not to be found in the final product so

much as it was in the process of creating it: once the final brush stroke was applied students rarely took the time to admire their own creation, but instead were eager to get involved in a new project. Csikszentmihalyi came to characterize these experiences as autotelic, where the goal (telos) lies in participation itself (auto). In the nearly four decades that have elapsed since this initial discovery, Csikszentmihalyi and colleagues have studied the flow experience in multiple contexts, and have identified a shared phenomenology in that participants consistently describe optimal states of complete absorption, focus, and enjoyment (for reviews see Csikszentmihalyi, 1990; Nakamura and Csikszentmihalyi, 2002).

The Flow Model and Its Use in the Study of Education

Studies of the flow experience have revealed consistency in the conditions under which these optimal states most often occur. Numerous researchers have confirmed commonalities both in optimal experiences and in their underlying conditions (Csikszentmihalyi, 1975/2000, 1990, 1996; Jackson, 1995, 1996; Massimini and Carli, 1988; Perry, 1999).

The Flow Experience

The optimal state described by individuals is most commonly characterized by: (1) intense concentration on the task at hand; (2) a deep sense of involvement and merging of action and awareness; (3) a sense of control over one's actions in dealing with the task at hand; (4) enjoyment or interest in the activity; and (5) a distorted sense of time (usually that time has passed very quickly). During the process of gathering these descriptions, several interviewees described themselves as being in flow or flowing. Thus, experiences characterized by such descriptions have become known as flow experiences or flow states.

The Flow Conditions

The specific activities from which individuals derive flow experiences vary widely. Interviews with males and females of different ages, classes, and cultural backgrounds have revealed that the flow state can emerge from involvement in a variety of activities, including athletics, performing surgery, tending cattle, haggling in the marketplace, working

on a factory line, reading, and writing. While there is considerable variation in the particular activities that lead people to experience flow, there are a number of phenomenological conditions that are typically present when flow does occur, regardless of the specific activity in which one is engaged. These conditions include: (1) engagement in activity chosen for its own sake – not a necessary, but a facilitative condition; (2) perceived challenges of the task at hand that are relatively high and in balance with one's perceived skills; (3) clear proximal goals that are regarded as important; (4) immediate feedback indicating one's success at meeting these goals; and (5) highly focused, rather than divided or scattered, attention.

Over the years, challenge and skill have emerged as two primary conditions for the flow experience: research has consistently shown that when challenges and skills are relatively high and in balance, the experience of flow likely ensues (Csikszentmihalyi and Csikszentmihalyi, 1988; Csikszentmihalyi, 1990, 1997). The flow model is often depicted by a chart similar to that shown in **Figure 1**. The figure describes four different channels of experience, each defined by the relative relationship between challenge and skill. When challenges and skills are both high, individuals tend to experience flow. Instances marked by high challenges but low skills tend to produce anxiety, while instances marked by low challenges but high skills produce relaxation (or at the extreme, boredom). Finally, instances of low challenges and low skills tend to produce feelings of apathy. Some researchers have further parsed these basic four channels of experience into 8 or even 16 channels (see Massimini and Carli, 1988), but in the interest of simplicity, and to preserve the focus on the flow state, only the more basic model is presented here.

The flow model itself is dynamic and is designed to account for changes in ability and circumstance: as an individual becomes more skilled, a given activity becomes decreasingly challenging, with the result that a person may cease to experience flow. In order to maintain one's

state of flow, the challenges of one's activities must be increased by choosing a more difficult task, setting a higher goal, or otherwise manipulating one's circumstances to bring challenges and skills back into a state of balance. Likewise, an individual may take on a task that is more challenging than one's skills are able to address at the moment, producing a state of anxiety. This anxiety can be relieved either by taking action to rapidly improve one's skills and bring them into alignment with the current challenges, or by adjusting one's activities and/or goals downward, so that the challenges of the activity are more in line with one's current skills. When viewed in an educational context, this dynamic nature of the flow model is similar in many ways to the zone of proximal development (Vygotsky, 1962; Gray and Feldman, 2004).

Some activities by their very nature are structured in such a way that proximal goals and feedback are more salient, challenges can be manipulated to best match one's skills, and distractions are minimized to focus attention. Indeed, there is some evidence that certain activities (e.g., making music and competitive athletics) are more likely than others to produce flow (Csikszentmihalyi, 1990). Nevertheless, flow refers only to a subjective phenomenology, suggesting that what matters most is that these conditions are salient to the individual, not necessarily inherent to the activity itself. Individuals have the capacity to identify challenges in seemingly unchallenging situations, define proximal goals and rules for engagement, and focus attention in such a way as to create the conditions for flow even when such conditions are absent from the task at hand. What is most essential for the experience of flow appears to be one's subjective perception of challenge, skills, goals, feedback, autonomy, and focused attention. Accordingly, numerous investigations have documented the flow state among individuals while doing daily household chores (Csikszentmihalyi, 1990), working in factory jobs (LeFevre, 1988), living in concentration camps (Logan, 1985), and in other situations that might appear on the surface to be counterproductive to the experience of flow (for a review, see Csikszentmihalyi, 1990).

The experience of flow may be related to personal characteristics. Adolescents who are highly optimistic and those with higher self-esteem generally report higher levels of flow than their peers (Schmidt *et al.*, 2007). Adolescent females tend to report more flow than males (Shernoff *et al.*, 2000). Turning specifically to individual characteristics related to flow in classrooms, American 12th graders report more flow in school than 10th graders, ethnic minority students report more flow than white students, and students from lower socioeconomic backgrounds appear to report more flow in school than their more advantaged peers (Shernoff and Schmidt, under review).

The momentary experience of flow is also linked to sustained engagement in tasks. Individuals who experience flow are likely to seek out these optimal experiences again,

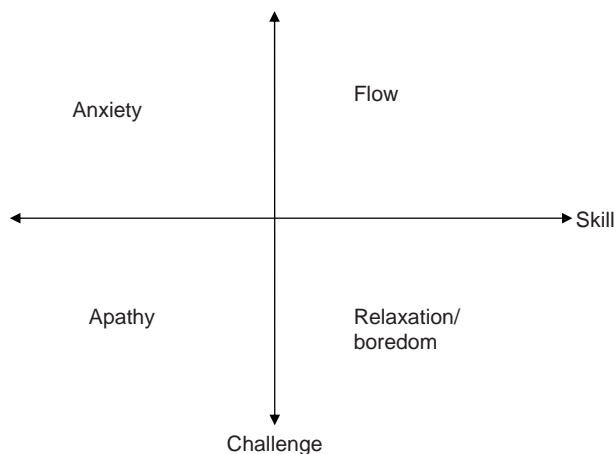


Figure 1 Channels of subjective experience.

with the result that they develop commitments to certain tasks or fields of engagement. This phenomenon has been referred to as emergent motivation, because one's desire to continue to pursue an activity emerges from the very experience of engaging in the activity. This perspective has been applied to the study of long-term academic pursuits. For example, Csikszentmihalyi, *et al.* (1993) demonstrated that students who have positive subjective experiences in a particular academic domain demonstrate greater commitment to that domain several years later. Csikszentmihalyi and Schneider (2000) demonstrated similar links between classroom experiences and students' future aspirations.

The Paradox of Flow in Educational Contexts

While the flow model was developed as a model of engagement more generally, it is not difficult to see the numerous applications to academic engagement. Many of the conditions that facilitate flow are present and manipulable in most classroom environments. While most academic activities are not freely chosen, there is some room to offer choice in learning activities. The balance between challenges and skills can certainly be adjusted to promote a state of optimal balance for students. Goal setting and feedback are both regular parts of most classroom settings. Finally, teachers can minimize distractions to promote attention to the task at hand. According to the model, if some or all of these conditions could be facilitated in classrooms, several things would be more likely to occur; students would likely experience flow when engaged in academic tasks, having positive subjective learning experiences. These positive experiences would motivate students to continue their deep engagement in tasks, as they would desire to maintain their positive flow state. As it is a dynamic process that keeps one in a flow state, learning would occur in order to keep one's skills in balance with the ever-increasing challenges presented by the classroom environment. Over time, such engagement would develop into long-term commitments to particular academic endeavors, and ideally, to increased success with related academic tasks.

Research has confirmed that some of the conditions that tend to promote flow are often present in educational settings. Specifically, adolescents experience high levels of challenge and skill more often in school than in any other context of their lives (Csikszentmihalyi, 1990). Recent research in American schools has suggested that while the challenge and skill conditions are often present in schools, students generally do not feel the deep concentration, involvement, and enjoyment that typically follow from these conditions. (Schmidt, *et al.*, 2007) Similar findings were obtained in a study of Italian students (Delle Fave *et al.*, 2002). The context of school is a unique setting in that the flow experience does not seem to follow from these two basic conditions as is the case in other contexts.

It appears to be academic classes, in particular, where the link between the flow conditions and the flow experience fails to exist. Research examining traditional middle and high school classrooms from the perspective of flow theory suggests that the instructional practices that are typically used in classrooms may lack certain other conditions for flow. Most high school students still spend a majority of their classroom time doing individual seatwork or listening to lecture, and while these activities may present significant challenges in which children can apply their skills, they tend not to be very involving for students (Shernoff, *et al.*, 2003).

Turner and colleagues (1998), similarly, have argued that specific instructional practices may promote flow experiences more often than others. They found that middle school students experienced the most flow when their teachers provided more scaffolded instruction (thus manipulating the balance of challenge and skill) and when they used instructional practices that fostered intrinsic motivation, such as providing choices and taking interests of students into account. The problem is that at the middle and high school level, the practices that produce the flow experience tend to be few and far between.

Schweinle, *et al.* (2006) provide further explanation for this disconnect between the flow conditions and the experience of flow in classrooms. In a study of elementary school mathematics classes, the researchers found that activities generally perceived by students to be challenging, were largely viewed as a threat to students' self-efficacy and were not seen as an opportunity to develop new skills. The authors speculate that children in elementary grades may have a different and more negative definition of challenge than older individuals. It may not be until individuals are older that they can view challenges more positively, after they have experienced the opportunities that optimal challenges can provide.

These studies suggest that while challenges and skills are often present, some of the other conditions for flow tend to be lacking in classrooms. Conditions for the flow experience also include: recall that focused attention, clear short-term goals, and the feeling that one is engaged in an activity that is freely chosen. Arguably, many traditional classroom environments might be characterized as places where students have little choice about their learning activities, and where the focus is on longer-term goals (e.g., completing the assignment, getting a good grade) rather than short-term goals (e.g., completing one small step at a time). In addition, a bustling classroom where students shift their attention from subject to subject after only a short period of engagement may not be the ideal place to promote focused attention. Indeed, there is some empirical evidence that in the rare cases where these additional conditions are present in classrooms, the flow experience is more likely to occur (Shernoff *et al.*, 2003, Schmidt *et al.*, 2007).

There are certain settings that do appear to consistently promote flow. When in nonacademic classes like fine arts or music, students experience greater levels of flow than in almost any other context, inside or outside of school (Schmidt, *et al.*, 2007). Similarly, in those relatively rare instances where students are involved in more hands-on, collaborative work in their academic classrooms, the experience of flow is more frequent (Shernoff *et al.*, 2003). Given that the flow state is more likely when students are engaged in these active learning environments, it is not surprising that students in nontraditional learning environments that emphasize active learning tend to experience more flow. Rathunde and Csikszentmihalyi (2005a, 2005b) compared the experience of students in traditional public middle schools to students in a Montessori school. While students in the two school types did not differ from one another in their subjective experience of nonschool activities, their experience of school was very different in that Montessori students experienced flow much more often than students in a traditional school. In 2004, Johnson obtained similar results in a study that compared the experience of students in a traditional American high school with that of students in an alternative high school with greater emphasis on student autonomy.

Andersen (2004, 2005a, 2005b, 2007) has conducted comparative studies in several Scandinavian countries and in Japan in which he observed classrooms and interviewed students about their flow experiences. Primary-grade students in Denmark experience particularly high levels of flow in school compared to students in other countries. These findings may be due to an emphasis on student autonomy, interest, and an appropriate balance between teacher-led and student-led learning activities (Andersen, 2004). These observations are consistent with the findings discussed earlier in which challenge and skill alone appear to be insufficient to produce flow in school contexts. Autonomy and control appear to be particularly critical conditions for fostering flow in school. It is important to note, however, that although Danish students were by far the most engaged in school, they lagged behind students from other countries (specifically Finland) in their performance of basic academic skills (Andersen, 2005a). These results suggest that schools should be warned against promoting engagement in school at the expense of demanding competence when it comes to basic skills.

Anderson's (2005b) work in Japan has identified several elementary schools and afternoon programs that are specifically designed to promote optimal learning among students. Central to these programs is a collaborative learning environment in which both students and teachers are expected to learn and develop through the use of computer-based activities and other advanced problem-solving activities. Classroom activities require frequent changes not only in activity, but also in students' physical

placement in the room, moving from group discussions, to hands-on projects, to teacher lecture-discussions, to computer sessions, all focused on the same subject matter. Andersen reports that students who participate in these programs not only experience a high degree of flow, but also develop deep understanding of the academic material. Unlike the Scandinavian programs studied, these Japanese programs have achieved a balance between the promotion of engagement and focus on learning outcomes.

Putting Flow theory into Practice: The Key School

While researchers have been able to identify those moments when students experience flow in school and link them to specific personal characteristics, instructional practices, or academic programs, most of the work presented thus far does not represent intentional application of flow theory *per se*. However, there are individual teachers, and indeed entire schools or school programs that have done substantial work to incorporate the conditions for flow into their learning environments. One of the longest-standing and most successful models of this is the Key Learning Community in Indianapolis, Indiana in the United States. The Key School is a magnet school that is part of the Indianapolis Public School System. Begun in 1987 as an elementary school, Key now serves elementary, middle, and high school students on two separate campuses. From its inception, the school has been structured to promote flow and intrinsically motivated behavior among its students. This mission has been carried out in several ways, including the development of courses and curricula that place equal emphasis on Gardner's (1993) multiple intelligences.

Perhaps the most intriguing application of flow theory has been the school's Flow Center. Students are scheduled to spend time in the Flow Center each week, just as they would spend time in any other class. The purpose of the center is to provide students with the opportunity to experience flow, so that they may recognize the positive nature of this state and be motivated to seek it out in their other scholastic activities. Knowing that many of the conditions for flow are themselves highly individual and subjective, the designers of the Flow Center recognize that the flow experience simply cannot be created for all students in a classroom context by the manipulation of activities or expectations by some external person (usually the teacher). What is more valuable then is to provide students with the tools to create flow for themselves, so that they can facilitate their own flow in a variety of environments. In order to help students experience flow, the Flow Center is equipped with a variety of games, puzzles and challenging activities for students to engage in. Students are free to choose their own activities in the Flow Center on the one condition that

the time spent in there is not treated as recess or downtime but is intended for problem-solving activities. The mission of the flow room is to help students realize that they can become deeply engaged in activities that are educational, and that the same processes they used in the flow room (whatever they are) can be applied in other areas of their studies (see Whalen and Csikszentmihalyi, 1991 for further description).

Evidence suggests that these modifications in the school environment translate into academic success. In addition to an uncommon sense of excitement about learning observed by visitors to the school, more standard measures point to success as well. Standardized test scores for students at the elementary and high school level are consistently higher than those of students at other schools in the district, though it is important to note that the high school program has existed only a short time, with the first students graduating in 2003. While only time will reveal the long-term success of these students, all initial indications are positive. In the years for which data are available, the school posts graduation rates at or near 100%, with the vast majority of graduates pursuing postsecondary education.

Looking Back, Looking Ahead

An examination of students' educational experience from the perspective of flow theory is informative to both researchers and educators in that it helps us understand the personal and situational factors that promote students' deep engagement in learning. There is enormous potential for students to experience flow in schools, but this potential is often thwarted because some key conditions for flow are largely missing from many of today's classrooms. While students generally experience two of the main conditions in that they are using their skills to address increased levels of challenges in the classroom, they often fail to experience flow. This may be because school environments generally offer few opportunities for choice. Moreover, the goals and feedback in classrooms may be focused on a target that is less immediate than the goals that are usually salient in flow experiences. Finally, classrooms may need to be structured in a way that helps students focus their attention on the tasks at hand.

Most encouraging is the fact that flow is consistently observed in a number of unique school contexts. Students often feel flow in their nonacademic classes, most of which provide a number of flow conditions like choice, autonomy, and focus that are typically absent in academic subjects. Likewise, students tend to experience flow when engaged in hands-on learning tasks. Additionally, a number of non-traditional schools and school programs are very successful in facilitating flow among their students – this appears to

be attributable to greater focus on many of the conditions that produce flow. Teachers interested in increasing their students' engagement in classrooms may learn a lot by examining the structure of nonacademic classes, as well as the nontraditional programs mentioned here, to take some cues about how to create more of the conditions for flow in students' everyday experience.

Looking to the future of research on flow in education, there is still much work to be done. First, more research needs to be done linking the experience of flow in academic pursuits with a variety of learning outcomes, as the few studies that have addressed this issue have produced mixed results. Future research must further examine the nature of links between the experience of flow and specific learning outcomes. Additionally, more research is needed to understand the link between flow in a given subject area, and long-term commitment to that field.

Second, relatively little work has been done to examine the role of flow in learning environments beyond the classroom. One emerging application of flow theory in education concerns the use of computers and video games. Children and adolescents frequently experience flow when engaged in video and computer games (Bassi and Delle Fave, 2004). In recent years, researchers and educators alike have attempted to use the appeal of videogames to construct interactive computer technology for learning. A growing body of evidence suggests not only that these e-learning environments can be intensely engaging, but also that such engagement is linked to a variety of positive learning outcomes (Coller and Shernoff, 2006; Pearce, 2005; see Scoresby and Shelton, 2007 for a review). An examination of the e-learning experience from the perspective of flow theory will assist in the understanding and design of these increasingly used educational tools. Likewise, researchers should continue to examine the role of flow as it relates to learning in other extracurricular environments.

Third, the vast majority of research on flow in education has involved children and adolescents. While the experience of flow appears to be consistent across age groups, it is possible that certain conditions for flow, similar to challenge or autonomy, might be differentially salient to learners of different ages. Thus, examining a broader age range of learners, including adults, would be informative. In general, research on flow in schools should take into consideration the developmental stage of the students under investigation.

Finally, we must not forget that teachers play a key role in educational processes – an examination of teachers from the perspective of flow theory would be informative as well. Dissertation work by Di Bianca, in 2000, suggests that when teachers report the most flow, students generally report the least. In other words, those moments that are most engaging to teachers are least engaging to students. Further research is needed to corroborate and explain these findings.

The flow model and related research provides a solid base of knowledge regarding how students might become more engaged in their learning and how they feel when they are so engaged. This knowledge base is of practical use to educators interested in increasing student engagement. There is, however, much work remaining to be done in this field in order to more fully understand the complexities of flow's role in educational processes.

See also: Constructivism and Learning; Development of Creativity; Emotion in Educational Contexts; Interest; Intrinsic and Extrinsic Motivation; Motivating Students in Classrooms; Motivation Regulation; Volitional Control of Learning.

Bibliography

- Andersen, F. O. (2005a). *Exploring the Roots of Optimal Learning: A Story of Successful Primary and Special Needs Education in Finland*. Bilund: Lego Learning Institute.
- Andersen, F. O. (2005b). "Kids on Campus" – an Optimal Japanese Concept for Learning. Copenhagen: The Danish University of Education.
- Andersen, F. O. (2007). *Creativity – and Creative Thinking – as an Integrated Part of Optimal Learning Environments*. Copenhagen: The Danish University of Education.
- Bassi, M. and Delle Fave, A. (2004). Adolescence and the changing context of optimal experience in time: Italy 1986–2000. *Journal of Happiness Studies* **5**, 155–179.
- Coller, B. D. and Shernoff, D. J. (2006). *Teaching Dynamic Systems and Control with a Video Game to Mechanical Engineering Undergraduates: A Proposal to the Course, Curriculum and Laboratory Improvement Program of the National Science Foundation*. DeKalb, IL: Northern Illinois University.
- Csikszentmihalyi, M. (1975/2000). *Beyond Boredom and Anxiety*. San Francisco, CA: Jossey-Bass.
- Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York: HarperPerennial.
- Csikszentmihalyi, M. (1996). *Creativity: Flow and the Psychology of Discovery and Invention*. New York: HarperCollins.
- Csikszentmihalyi, M. (1997). *Finding Flow: The Psychology of Engagement with Everyday Life*. New York: Basic Books.
- Csikszentmihalyi, M. and Csikszentmihalyi, I. S. (eds.) (1988). *Optimal Experience: Psychological Studies of Flow in Consciousness*. New York: Cambridge University Press.
- Csikszentmihalyi, M. and Schneider, B. (2000). *Becoming Adult: How Teenagers Prepare for the World of Work*. New York: Basic Books.
- Delle Fave, A., Bassi, M., and Massimini, F. (2002). Quality of experience and daily social context of Italian adolescents. In Comunian, A. L. and Gielen, U. P. (eds.) *It's All About Relationships*, pp 159–172. Lengerich: Pabst Science Publishers.
- Gardner, H. (1993). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.
- Gray, P. and Feldman, J. (2004). Playing in the zone of proximal development: Qualities of self-directed age-mixing between adolescents and young children at a democratic school. *American Journal of Education* **110**, 108–143.
- Jackson, S. (1995). Factors influencing the occurrence of flow state in elite athletes. *Journal of Applied Sport Psychology* **7**, 138–166.
- Jackson, S. (1996). Toward a conceptual understanding of the flow experience in elite athletes. *Research Quarterly for Exercise and Sport* **67**, 76–90.
- LeFevre, J. (1988). Flow and the quality of experience during work and leisure. In Csikszentmihalyi, M. and Csikszentmihalyi, I. S. (eds.) *Optimal Experience*, pp 307–326. New York: Cambridge University Press.
- Logan, R. (1985). The "flow experience" in solitary ordeals. *Journal of Humanistic Psychology* **25**, 79–89.
- Massimini, F. and Carli, M. (1988). The systematic assessment of flow in daily experience. In Csikszentmihalyi, M. and Csikszentmihalyi, I. S. (eds.) *Optimal Experience: Psychological Studies of Flow in Consciousness*, pp 266–287. New York: Cambridge University Press.
- Nakamura, J. and Csikszentmihalyi, M. (2002). The concept of flow. In Snyder, C. R. and Lopez, S. J. (eds.) *Handbook of Positive Psychology*, pp 89–105. Oxford: Oxford University Press.
- Pearce, J. M. (2005). Engaging the learner: How can the flow experience support e-learning? *Paper Presented at the E-Learn 2005 Conference*. Vancouver, BC, Canada.
- Perry, S. K. (1999). *Writing in Flow*. Cincinnati, OH: Writer's Digest Books.
- Rathunde, K. and Csikszentmihalyi, M. (2005a). Middle school students' motivation and quality of experience: A comparison of Montessori and traditional school environments. *American Journal of Education* **111**, 341.
- Rathunde, K. and Csikszentmihalyi, M. (2005b). The social context of middle school: Teachers, friends, and activities in Montessori and traditional school environments. *Elementary School Journal* **106**, 59.
- Shernoff, D. J., Csikszentmihalyi, M., Schneider, B., and Steele-Shernoff, E. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly* **18**, 158–176.
- Schweinle, A., Meyer, D. K., and Turner, J. C. (2006). Striking the right balance: Students' motivation and affect in elementary mathematics. *Journal of Educational Research* **99**, 271–293.
- Shernoff, D. J., Knauth, S., and Makris, E. (2000). The quality of classroom experiences. In Csikszentmihalyi, M. and Schneider, B. (eds.) *Becoming Adult: How Teenagers Prepare for the World of Work*, pp 141–164. New York: Basic Books.
- Scoresby, J. and Shelton, B. E. (2007). Visual perspectives within educational computer games: Effects on presence and flow within virtual learning environments. *Paper Presented at the Annual Meeting of the American Educational Research Association*. Chicago, IL, USA.
- Schmidt, J. A., Shernoff, D. J., and Csikszentmihalyi, M. (2007). Individual and situational factors related to the experience of flow in adolescence: A multilevel approach. In Ong, A. D. and van Dulmen, M. (eds.) *The Handbook of Methods in Positive Psychology*, pp 542–558. Oxford: Oxford University Press.
- Turner, J. C., Meyer, D. K., Cox, K. E., et al. (1998). Creating contexts for involvement in mathematics. *Journal of Educational Psychology* **90**, 730–745.
- Vygotsky, L. S. (1962). *Thought and Language*. Cambridge, MA: MIT Press.
- Whalen, S. P. and Csikszentmihalyi, M. (1991). *Putting Flow Theory into Educational Practice: The Key School's Flow Activities Room. Report to the Benton Center for Curriculum and Instruction*. Chicago, IL: University of Chicago.

Further Reading

- Hektner, J. M., Schmidt, J. A., and Csikszentmihalyi, M. (2007). *Experience Sampling Method: Measuring the Quality of Everyday Life*. Thousand Oaks, CA: Sage.
- Nakamura, J. and Csikszentmihalyi, M. (2002). The concept of flow. In Snyder, C. R. and Lopez, S. J. (eds.) *Handbook of Positive Psychology*, pp 89–105. Oxford: Oxford University Press.
- Shernoff, D. J. and Csikszentmihalyi, M. (in press) Flow in schools: Cultivating engaged learners and optimal learning environments. In Gilman, R., Huebner S., and Furlong, M. J. (eds.) *Promoting Wellness in Children and Youth: A Handbook of Positive Psychology in the Schools*. Mahwah, NJ: Erlbaum.

- Schmidt, J. A., Shernoff, D. J., and Csikszentmihalyi, M. (2007). Individual and situational factors related to the experience of flow in adolescence: A multilevel approach. In Ong, A. D. and van Dulmen, M. (eds.) *The Handbook of Methods in Positive Psychology*, pp 542–558. Oxford: Oxford University Press.
- Shernoff, D.J. and Schmidt, J. A. (in press). Further evidence of an engagement-achievement paradox among US high school students.

Relevant Websites

- <http://www.616.ips.k12.in.us> – Key Learning Community in Indianapolis, Indiana, US.
- <http://www.legolearning.net> – Lego Learning Institute (contains work of F.O Andersen).
- <http://www.ppc.sas.upenn.edu> – Positive Psychology Center at University of Pennsylvania.

Interest

M Ainley, University of Melbourne, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Avatar/animated pedagogical agent – An agent within a computer game or simulation; in game environments avatars are usually the players game character; in learning software animated pedagogical agents are humanoid assistants or animated characters who facilitate learning.

Deep processing strategies – Processing strategies that require transformation and elaboration of learning material; require more processing effort than strategies relying solely on rote memory or surface processing.

Engagement – Connection and involvement; used to denote active, positive connection with schooling; task engagement; school engagement; distinctions made between behavioural, cognitive and affective engagement.

Mastery achievement goals – Learning goals where the purpose is to improve skills, achieve understanding, or expand competencies; cf, performance achievement goals.

Seductive details – Task information or details that students find highly interesting but which is not part of the important information content of a task.

Seeking system – An evolutionary emotional brain system that gears the animal for goal seeking action including exploration and information seeking; interest is an expression of this system (see Panksepp, 2000).

Self-regulated learning – Learning that is controlled and regulated by the learner; a self-regulated learner has clear learning goals, monitors progress toward achievement of their goals, and reflects on the quality of their own learning achievements.

Trait curiosity – Curiosity from the perspective of an individual disposition or personality trait; distinguishing between individuals on the basis of their typical style of reaction to novel and puzzling phenomena; cf, state curiosity.

A couple of years prior to the publication of the last volume of the encyclopedia, Suzanne Hidi published a seminal paper describing interest as a mental resource for learning (Hidi, 1990). Although Hidi was not the only person writing about interest at that time, her paper struck a chord in the educational community and is widely cited.

Two complementary perspectives on interest were identified and this distinction has formed the basis for most of the research and thinking on the nature of interest and the role it plays in learning and development. Hidi's review article was followed up in 1992 with a volume that presented a collection of the major writers and researchers who were addressing the issue of interest and its contribution to learning (Renninger *et al.*, 1992). Another significant feature of this volume was the bringing together of researchers from around the world. The work emanating from the school of German researchers (Hans Schiefelé and Erhardt Todt) with their person-object (POI) model of interest was presented alongside work that had been growing in the US and Canada. What followed has been an intense research focus on the characteristics of interest and its role in learning.

Interestingly, the resurgence in attention to the role of interest in learning and development claimed heritage from Dewey's writings in the early twentieth century. For example, in *Interest and Effort in Education* Dewey wrote: "the root idea of the term seems to be that of being engaged, engrossed, or entirely taken up with some activity because of its recognized worth" (1913 cited Boydston, 1979, p. 160); another work referred to interest as the "vital union between the student and his study" (p. 253). Berlyne's work in the 1940s and 1950s was another important influence on the development of ways of thinking about interest. Berlyne defined three ways in which the term interest was used in psychology: a state of the organism, a quality of the stimulus, and a relation between organism and stimulus (Berlyne, 1949). As will become clear, each of these has in some way influenced current thinking. We will consider current perspectives on interest by focusing on the three most widely used perspectives; interest as psychological state, situational interest, and individual interest.

The Character of Interest

What exactly does interest mean? One of the simplest ways to represent the range of usage of interest terminology is to look at the level of generality in the behavior mentioned. At the most specific level, interest refers to a psychological state in which attention is focused on a particular object or event. For example, a student has her gaze fixed in concentrated attention on a computer screen where she has just discovered a website showing the first images of a newly discovered planet. Close observation reveals that

she is absorbed by what is in front of her and as we watch she continues to explore the website and other images of the planet, ignoring the conversation of her friends at the next table. In this example, we are witnessing interest as a state and this phenomenon raises a number of questions that contemporary research programs are addressing. What specific processes are involved? What function does it serve in learning and achievement?

Interest as State

The state of interest has been shown to involve positive affect. There is still some debate over whether this state of positive affect is an emotion. Some current neuropsychological theories of emotion suggest that interest is the information-seeking strand of a basic seeking system (Panksepp, 2000). We know that positive affect is an important element in the state of interest and that it may also be accompanied by other feelings which are not necessarily positive. For instance, most people can appreciate the mix of positive and negative affect associated with students watching their first animal dissection in biology classes. The state of interest is predictive of persistence, exploratory, or information-seeking activity, and has been shown to be associated with knowledge acquisition and performance. Interest is a dynamic state and so can be expected to change over relatively short periods of time. Students' persistence with curriculum tasks often depends on whether the task sustains interest. When interest in a task is sustained or even increased, students are more likely to persist with the task and show higher levels of learning and performance.

Situational Interest

Shifting the lens slightly directs the focus to the character of interest as determined by sources or factors that trigger interest. Specific kinds of events or contextual factors can trigger a state of interest and, when triggered in this way, interest has been referred to as situational interest. The young female student described above happened upon the planet images by chance. She was searching for a site on the Internet featuring a popular music group and happened to click on the planet link by chance. Immediately, her attention was caught by the colorful swirls of the orbiting planet and she investigated it further. Her interest in the planet is appropriately described as situational interest.

Situational interest can be triggered by a wide range of environmental events. Physical properties of an event, such as change, novelty, ambiguity, and uncertainty, have all been linked with the triggering of situational interest. This fact has often been used in educational settings to engage students with curriculum tasks. It is clear from a large number of research findings that interactive

computers and digital information can be used to trigger situational interest in a variety of tasks for a wide range of students. In relation to learning and development, the problem then becomes one of holding or sustaining students' interest so that they engage with the learning content. There is no doubt that the new medium captures interest of the students. However, very often that interest rapidly wanes, and often disappears when what follows does not match the enticing first impressions. The interest triggered by the swirling color and movement of the planet in orbit is likely to dissipate if our student is reminded of her intention to find the music group's website.

A substantial proportion of the research on situational interest has investigated the influence of interest on reading text. General findings provide clear evidence of a significant effect of interest on performance indicators, for example, comprehension and recall of text information. It is also well documented that interest affects deeper processing as shown by scores on learning measures that require processes such as elaboration, transfer, or application of information to a new setting. These effects imply that techniques for increasing situational interest might be employed to improve learning outcomes. There are a number of text features that can be used to influence situational interest. Structural features of the text, such as novelty, coherence, ease of comprehension, and vividness, can be modified to increase interest. At an individual level, personally meaningful or valued content can be used to increase interest, while at the broader group level, themes of universal personal significance (e.g., life, death, or sex), sometimes referred to as absolute interests, have predictable effects on interest (Schraw and Lehman, 2001; Wade, 2001).

While providing some guidance for teachers, there are also some cautions to be heeded. When not connected to the important informational content of the text these themes may take on the character of seductive details. Rather than improving performance they may interfere with it by diverting attention away from significant text themes and meaning. It is when interest and important text elements coincide that interest has a positive influence on learning. Similar findings have been reported in other learning domains, for example, learning from science texts and learning with multimedia resources.

Individual Interest

At a more general behavioral level, personal or individual interests have been identified as an important type of interest that influences students' learning. Individual interests are predispositions or trait-like personal organizations that have been developed over time. A topic, a domain, a school subject, or a type of activity, may each become the object or content of an individual interest. For example, a second young student has also landed on the website displaying the first images of the new planet. He also

shows the unmistakable signs of experiencing interest; his demeanor leaves no doubt that he is enjoying what he is seeing. His attention is concentrated on the screen and he looks as if he will be glued to that website for some time. However, the interest being observed in this student is different from that of the young girl who discovered the planet images by accident. This student came to the site with a well-developed interest in astronomy. He had heard that the images of the new planet had just been released and was eager to check them out. From as early as 6 years of age, he was attracted to pictures of planets, has learnt their names, and details of their composition and orbits. Hence, although displaying a similar state in response to the images of the new planet, his interest is different. He is seeking different kinds of information about the new planet, information that will expand his already well-developed knowledge of the universe. Hence, the characteristics of individual interest, sometimes referred to as personal interest, are that there is a well-developed personal organization of knowledge, value, and affect. Our astronomy enthusiast has a well-structured core of knowledge about the domain. The subject is important to him and his feelings, when pursuing his interest, are likely to include some or all of fascination, excitement, enjoyment, surprise, wonder, and delight.

Clearly, when the object of interest is a school domain or has close links with a school domain, problems of motivation and engagement or problems with achievement outcomes are likely to be minimal. But by the same token, when individual interest in school domains is lacking or is lower than required for effective learning, this is soon reflected in low achievement levels and lack of motivation. It is well documented that in many developed countries significant numbers of young adolescent students are showing low interest in schooling domains. Awareness of this problem has prompted questions concerning the development of individual interest which will be taken up in a later section.

Vocational Interests

One form of interest, not yet considered in this article, concerns interest at a level more general than individual interest – generic interests. These interests are the broad personal organizations of preferences for types of activities and experiences that constitute vocational interests. Vocational interests are generally treated as personality traits. They can be observed in children's preferences early in their schooling and function as organizing themes in the way children and adolescents react to new situations, to people and to tasks (Betz and Borgen, 2000). Holland's classification of vocational interests into realistic, investigative, artistic, social, enterprising, and conventional (RIASEC) themes is used in a wide range of research and in practical settings as a model of vocational

interests. In the same way that individual interests develop through experience over time, it has been proposed that vocational interests are based on two main forms of self-perceptions that accrue over time. These are individuals' self-perceptions of efficacy or confidence in relation to pursuing tasks and activities, and their perceptions of the likely outcomes of those tasks and activities. Schooling experiences at all levels contribute to and consolidate students' vocational interests.

Development of Interest

How interest in schooling domains develops is central to educational philosophy and practice. If interest, in any or all of its forms, is a key component of learning as our review suggests, then understanding the factors that initiate and support the development of interest are essential for successful educational outcomes. This question has come to prominence recently and a number of models of interest development have been proposed. All emphasize an important role of experience in supporting the development of interest. Hidi and Renninger (2006) have proposed a four-phase model of interest development that identifies a sequence which begins with the triggering of situational interest by a specific feature of the environment. When this or similar experiences occur a number of times, the situational interest is maintained or strengthened. External support, for example in the form of teacher encouragement, is important at this stage of interest development. Further experiences over time strengthen and deepen the intrapersonal organization of thinking and feelings associated with the activity. By now, a wider range of similar activities or even a broader domain has become the content of the interest. It has progressed through phases of being an emergent individual interest to being a well-developed individual interest with stronger and deeper knowledge, value, and feeling components. Although still responsive to external support, it is not essential at this stage. If we go back to the two students who were observing planet images on the computer screen, we can locate them at different positions in their development of an interest in the solar system. The young girl who happened on the planet images by chance would be identified as being at the earliest phase, a triggered situational interest. This may be the start of a new interest or it may be an event that is never repeated and her interest in the solar system may go no further. On the other hand, the young boy already has a well-developed interest in the solar system and his experience is about consolidating and extending his knowledge within that domain.

A three-stage model of development from situational to individual interest (situational interest, a stabilized situational interest, and an individual interest) based on

the POI theory of interest (Krapp, 2003) has also been advanced. This theory is used to map developmental trajectories of interest, especially in relation to schooling domains and to explore the basis for the typically lower levels of interest in school that occur as students move from childhood to early adolescence.

Gender patterns of interest development have also been investigated. Besides showing specific relationships with school subjects domains, gender patterns of interest have also been shown to influence the way students respond to specific topics within school subjects and to ways that topics are presented. These patterns have important implications for teachers. For example, in physics classes it has been shown that the interest levels of boys and girls are influenced by the way that physics topics are presented. When learning the principles of a pump, girls showed higher levels of interest when the topic was pumping the blood around the human body than when the topic was pumping gas out of the ground (Hoffmann and Haussler, 1998). These findings raise a number of important questions for educators and researchers. What are the limits to this approach? Can all of the physics content that is important in a school curriculum be presented in ways that trigger interest for all students? Can interest triggered in this way provide the basis for the development of individual interest in physics, sufficient to sustain further exploration and involvement with a wide range of physics concepts and topics?

Yet, another perspective on interest development focuses on its contribution to knowledge acquisition and the development of expertise in specific domains. Alexander (2004) links situational and individual interest to different stages in a model of domain learning (MDL) that describes the trajectory from novice to expert. At the novice stage, situational interest supports knowledge and skill acquisition. However, later in the knowledge-development sequence individual interest in conjunction with deep-processing strategies is more likely to support development of skill and expertise.

These models signal the potential of interest theory and research to contribute to understanding how students learn and to identify how teachers might use their understanding of interest processes to support and deepen students' learning.

Interest and Related Motivational Concepts

Interest is only one of a number of motivational concepts currently being researched and some comment is needed on the place of interest among motivation concepts. Most commonly, interest is grouped with concepts such as engagement, curiosity, flow, and intrinsic motivation. All refer in some way to students becoming involved

with an activity such that they are focused on the task or activity for its own sake rather than simply as a means to some other end outside of the task. In education, the contrast is drawn between interest in the activity (intrinsic motivation) and working on the activity to achieve some external end, such as a high grade, or to receive some form of praise (extrinsic motivation). One significant direction in contemporary thinking on motivation is the articulation of how these concepts contribute to self-regulated learning. Hence, an important question in relation to interest is an understanding of how it is related to the many processes that contribute to adaptive, self-regulated learning.

Self-regulated learning has been defined in terms of the processes whereby personal learning goals are translated into action. While there are a number of specific models of self-regulated learning, they all imply a sequence from activation and task planning, to monitoring and control of on-task behavior, and finally reflective evaluation of task outcomes. Interest, whether considered from the perspective of on-task state, situational interest or individual interest, plays an important role in the sequence of processing, that is, self-regulated learning.

At the task level, the general pattern emerging is that positive orientations to learning (e.g., trait curiosity, mastery achievement goals, or individual interest in a cognate domain) make it more likely that the new task will trigger student's interest. Once interest has been triggered it is then more likely that students' will experience positive affect, will choose to persist with the task, and are more likely to use deep-processing strategies. Another self-regulatory variable that has been shown to be closely linked with the triggering of interest is self-efficacy. When students feel confident about their ability to perform the task, they are more likely to have their interest triggered.

Similar patterns of association have been shown when the research is focusing on individual interest. For example, in college students mastery achievement goals have been shown to be associated with higher levels of individual interest in the subject and this in turn is associated with further participation by choosing courses within that domain. Hence, individual interest is an important component in student engagement over extended time intervals or students' continued participation with the domain content.

Yet, another aspect of the self-regulatory function of interest is operating when students actively make a boring task interesting. It has been shown that when students are committed to continuing with a boring task, they will find ways of making the tasks interesting and in this way regulate their task motivation (Sansone and Thoman, 2005). In the next few years, investigations of exactly how these variables combine to produce effective self-regulated learning will expand our current understanding of ways to support and enhance student learning.

Measurement of Interest

In interest research, as with many areas of educational processes, there is a great reliance on questionnaire measures and this is likely to continue to be an appropriate way to measure individual interest and vocational interests. However, recent developments especially in the measurement of the state of interest are capitalizing on the potential of interactive computer technologies in innovative ways. For example, measures inserted into a task requiring a quick response and minimal interruption to the flow of the task are being used to monitor how the state of interest changes across a task. Using these techniques with reading tasks, it has been demonstrated that gender interacts with text topic to influence whether interest is maintained or extinguished across the course of reading passages of narrative and expository text. Using interactive technologies, complex problem scenarios from domains such as mathematics, social issues, and biology, have been used to identify complex relationships between sets of motivational characteristics (including interest) and to determine how they function in sequences of task behavior connecting persistence, processing strategies, and performance. Simultaneously, these new technologies are also being investigated as a medium for learning and their impact on students' interest in their learning is one of the important factors being addressed.

Educational Technologies and Interest

Part of the appeal of using modern informational technologies as an instructional medium is the often reported observation that students find this medium, more interesting. The assumption is then that students will learn more effectively using modern information technology than when using more traditional media. This assumption has been tested in research comparing information technology and more traditional instructional media (e.g., textbooks) and a number of important findings have been reported. There is no doubt that many of the features of multimedia presentations that are possible by using the latest information technology trigger student interest. Complex combinations of color, movement, and sound have a powerful effect on students. Interest is triggered and in the same way as our earlier student had her interest triggered by the color and movement of the planet images, instructional materials that make use of the potential of new information technologies have a ready audience. At one level, the state of interest is a response to novelty. Interest is triggered and the students' attention is focused on the novel event. However, this effect is short lived and the student is then prey to the next novel event that triggers their interest.

It has been widely shown that students will engage further with the instructional content, that is, their interest will be sustained if what follows the initial novel experience is personally meaningful. However, what is personally meaningful will vary according to the age and experience of students. Instructional content that builds on students' personal or individual interests is likely to sustain the interest initially triggered by the novel medium. As with the research on student interest and text learning, the introduction of seductive details into the multimedia instructional content to hold interest has a harmful rather than beneficial effect on learning. Using scientific text material supported with graphics, it has been demonstrated that seductive details function by diverting the reader. Seductive details prime the student by bringing to mind prior knowledge that is not the main content of the task which then guides selection and organization of new content. In this way, important content is likely to be overlooked (Harp and Mayer, 1998).

On the other hand, some features of new technologies have the potential to improve learning through their effects on student interest. For example, instructional systems have incorporated social agents (referred to as avatars or animated pedagogical agents) who support students' learning. When students reach a plateau in their problem solving, the social agent becomes active and through utterances and/or actions they provide scaffolds for the student's learning. One of the ways this facilitation effect operates is through an increase in students' interest in the task (Moreno *et al.*, 2000). With the rapid expansion in use of information technology, the role of interest as one of the important processes underlying the effectiveness of these technologies in student learning will be a priority question for further investigation.

The Next Decades of Interest Research

The relationship between interest and academic achievement has been referred to a number of times in this article. Although there is variability across studies using different learning domains, different students, and different schooling levels, interest typically accounts for approximately 10–15% of the variation in achievement. It is also clear that interest is associated with deeper processing. As has already been suggested, the next decades are likely to see increased attention on the ways that interest processes combine with other motivational variables to contribute to effective self-regulated learning. In addition, modern information technologies will continue to expand our knowledge of how interest processes can support students' learning as well as being an important tool for learning.

See also: Affect, Mood and Emotions; Emotion in Educational Contexts; Flow in Education; Intrinsic and

Extrinsic Motivation; Motivating Students in Classrooms; Motivation Regulation; Self-Efficacy Beliefs; Volitional Control of Learning.

Bibliography

- Alexander, P. (2004). A model of domain learning: Reinterpreting expertise as a multidimensional, multistage process. In Dai, D. Y. and Sternberg, R. J. (eds.) *Motivation, Emotion and Cognition: Integrative Perspectives on Intellectual Functioning and Development*, pp 273–298. Mahwah, NJ: Erlbaum.
- Berlyne, D. E. (1949). 'Interest' as a psychological concept. *British Journal of Psychology* **39**, 184–195.
- Betz, N. E. and Borgen, F. H. (2000). The future of career assessment: Integrating vocational interests with self-efficacy and personal styles. *Journal of Career Assessment* **8**(4), 329–338.
- Boydston, J. A. (ed.) (1979). *John Dewey: The Middle Works, 1899–1924. Vol. 7: 1912–1914*. London: Southern Illinois University Press.
- Harp, S. F. and Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology* **90**(3), 414–434.
- Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research* **60**, 549–571.
- Hidi, S. and Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist* **41**(2), 111–127.
- Hoffmann, L. and Haussler, P. (1998). An intervention project promoting girls' and boys' interest in physics. In Hoffmann, L., Krapp, A., Renninger, K. A., and Baumert, J. (eds.) *Interest and Learning: Proceedings of the Seeon Conference on Interest and Gender*, pp 301–316. Kiel: IPN.
- Krapp, A. (2003). Interest and human development: An educational-psychological perspective. In Smith, L., Rogers, C., and Tomlinson, T. (eds.) *BJEP Monograph Series, Series II(2): Development and Motivation*, pp 57–84. Leicester: The British Psychological Society.
- Moreno, R., Mayer, R. E., and Lester, J. C. (2000). Life-like pedagogical agents in constructivist multimedia environments: Cognitive consequences of their interaction. *Paper Presented at the World Conference on Multimedia, Hypermedia, and Telecommunications, (ED-MEDIA)*. Montreal.
- Panksepp, J. (2000). Emotions as natural kinds within the mammalian brain. In Lewis, M. and Haviland-Jones, J. M. (eds.) *Handbook of Emotions*, 2nd edn., pp 137–154. New York: Guilford.
- Renninger, K. A., Hidi, S., and Krapp, A. (eds.) (1992). *The Role of Interest in Learning and Development*. Hillsdale, NJ: Erlbaum.
- Sansone, C. and Thoman, D. B. (2005). Interest as the missing motivator in self-regulation. *European Psychologist* **10**, 175–186.
- Schraw, G. and Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review* **13**, 23–52.
- Wade, S. E. (2001). Research on importance and interest: Implications for curriculum development and future research. *Educational Psychology Review* **13**, 243–261.

Further Reading

- Ainley, M., Hidi, S., and Berndorff, D. (2002). Interest, learning and the psychological processes that mediate their relationship. *Journal of Educational Psychology* **94**(3), 545–561.
- Bergin, D. A. (1999). Influences on classroom interest. *Educational Psychologist* **34**, 87–98.
- Boekaerts, M. (2002). Bringing about change in the classroom: Strengths and weaknesses of the self-regulated learning approach – EARLI Presidential address, 2001. *Learning and Instruction* **12**, 589–604.
- Hidi, S. (2001). Interest, reading and learning: Theoretical and practical considerations. *Educational Psychology Review* **13**, 191–209.
- Hidi, S. and Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research* **70**, 151–179.
- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review* **16**(4), 385–407.
- Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In Sansone, C. and Harackiewicz, J. M. (eds.) *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*, pp 375–407. New York: Academic Press.
- Sansone, C., Weir, C., Harpster, L., and Morgan, C. (1992). Once a boring task always a boring task? Interest as a self-regulatory mechanism. *Journal of Personality and Social Psychology* **63**(3), 379–390.
- Schiefele, U. (1999). Interest and learning from text. *Scientific Studies of Reading* **3**, 257–280.
- Silvia, P. (2006). *Exploring the Psychology of Interest*. New York: Oxford University Press.

Intrinsic and Extrinsic Motivation

C Levesque, K J Copeland, and M D Pattie, Missouri State University, Springfield, MO, USA
E L Deci, University of Rochester, Rochester, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Individuals are motivated for different reasons. Throughout the day, and over a lifetime, individuals act and behave differently depending on the circumstances surrounding their decisions. One who feels supported in an educational environment might feel satisfied and challenged with a difficult class project. On the other hand, one who feels controlled in an educational setting might consequently miss a lot of classes and appear disengaged while in class. When one feels supported and engaged, positive forms of motivation such as intrinsic motivation are likely to be sustained. In a supportive context, individuals are more likely to value what they do and feel satisfied while doing the activity. In contrast, when one feels controlled and pressured, negative forms of motivation such as external regulation are likely to be triggered. In a controlling context, individuals are more likely to do only what they are told and feel dissatisfied and disengaged while doing so.

The difference between intrinsic and extrinsic motivation is a well-known dichotomy between motivational types that has been discussed in the psychology literature for almost four decades. The following examples illustrate this dichotomy. Tommy likes reading, and he reads every chance he gets. His room is filled with books! Tommy's parents encourage his reading and get involved whenever they can. On the other hand, Melanie does not like to read, and she only does so when she is forced by her teacher and parents to read for an upcoming book report. The difference in their environments fosters different types of motivation which supports different learning outcomes.

Intrinsic Motivation

Intrinsic motivation is the most self-determined, or autonomous, form of motivation. It underlies activities that are performed purely for the joy gained from the activities themselves. In other words, individuals who are intrinsically motivated choose to engage in activities solely for the pleasure, interest, enjoyment, and satisfaction derived from performing those very activities. A child who plays with construction blocks for hours at a time, fully absorbed in the process of building towers and houses, is intrinsically motivated toward that activity (Deci, 1971, 1975). He does not do it in order to gain something from it, but only for the simple, pure pleasure of the activity itself.

deCharms's (1968) notion of perceived locus of causality speaks directly to the autonomous nature of intrinsic

motivation as it describes the tendency of individuals to be motivated to feel like they are at the origin of their behaviors. That is, individuals want to perceive themselves as choosing to perform behaviors out of their own volition. For example, students are more likely to feel like the behavior originated from within if they were encouraged to develop their projects and worked on them at their own pace. In contrast, students are more likely to feel their behavior is controlled if a certain project topic were imposed upon them with a very strict deadline. deCharms' concept of perceived locus of causality is, furthermore, linked to the basic psychological need for autonomy discussed in self-determination theory (SDT) (Deci and Ryan, 1985, 2000). The need for autonomy does not signify independence. Autonomy signifies choice and volition. The need for autonomy is satisfied when individuals feel they have choice and volition while engaging in a particular behavior. The need for autonomy underlies the development of intrinsic motivation.

Another basic need that supports intrinsic motivation is the need for competence. The concept of competence derives from White's (1959) assertion that people often engage in activities simply because they want to experience themselves as competent or efficacious. This is linked to the basic psychological need for competence as described in the SDT (Deci and Ryan, 1985, 2000). The need for competence is satisfied when individuals feel like they are developing skills and mastering the activities they are performing.

The satisfaction of the basic psychological needs for autonomy and competence are instrumental for the maintenance and enhancement of intrinsic motivation. Environments that provide support for competence and autonomy will foster intrinsic motivation. For example, a classroom environment that encourages students to become involved in how they will learn new materials and that provides them with the opportunity to frequently experience small successes will foster intrinsic motivation. In contrast, a classroom environment that is dictated, and, therefore, does not provide support for autonomy and competence will diminish intrinsic motivation and not encourage student involvement. The provision of rewards has been found to decrease an individual's level of intrinsic motivation when the rewards are salient and given for performing an activity that the individual finds interesting. This is the case because rewards are usually seen as an attempt to control one's behavior, thus jeopardizing one's sense of autonomy. Take

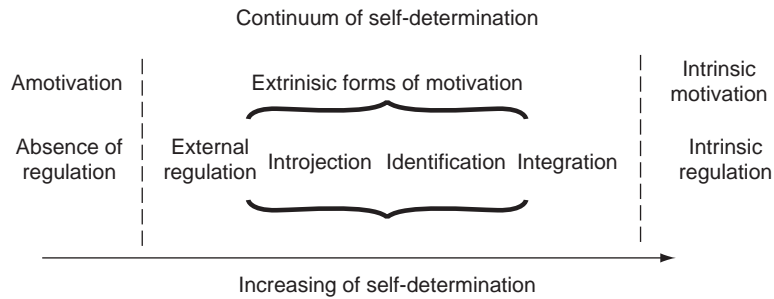


Figure 1 Visual representation of the self-determination continuum.

for example 6-year-old Lisa who really likes to read *The Cat in the Hat* book. Just to showcase her reading skills at the next family reunion, Aunt Sally begins to give money to Lisa for reading the book to family and friends. After a few family reunions and a few dollars spent, Lisa does not want to read *The Cat in the Hat* on her own anymore. She will only read it if she is given money for doing so. This is a book that Lisa used to enjoy reading, on her own, without any external incentives. In other words, she was intrinsically motivated toward reading the book. From a motivational perspective, the provision of the reward for doing an enjoyable activity shifted the perceived locus of causality from internal to external, making Lisa feel less autonomous about reading the book – feeling less like an originator of her behavior. Lisa is no longer doing the activity because she enjoys it but because she wants to obtain the reward. Reading this book is now more extrinsically motivated than intrinsically motivated, and it was her environment that triggered this change.

Extrinsic Motivation

Extrinsic motivation refers to the forms of regulation that underlie activities that are performed as a means to an end. Little Lisa who is now reading *The Cat in the Hat* because she wants the money is extrinsically motivated toward reading that book. A fifth grader who does his homework in order to gain the privilege to go see a movie at his friend's house afterward is extrinsically motivated toward doing his homework. This child is doing homework in order to obtain a reward. The same would be true of a teenage girl who cleans her room in order to avoid losing car privileges over the weekend. She is cleaning her room because she does not want to lose the car for the weekend, not because she just wants to clean and because she enjoys a clean room. Taken as a whole, extrinsically motivated activities are performed to attain a goal, to obtain a reward, or to avoid a penalty or a negative consequence. When extrinsically motivated, individuals perform the activities not because they simply derive enjoyment from them, but because the activities are instrumental in reaching a goal, or avoiding an undesirable outcome or

consequence. The activities may still be valuable to the person engaging in them, but the motivator is not the enjoyment but rather what can be obtained by engaging in the activity.

Although motivation can be dichotomized as intrinsic and extrinsic, SDT (Deci and Ryan, 1985, 2000) further specifies four types of extrinsic motivation. The different types of extrinsic motivation vary in terms of the degree to which they have been internalized and thus are self-determined. **Figure 1** depicts the placement of the four forms of extrinsic motivation and the one form of intrinsic motivation along a continuum of self-determination.

External Regulation

This is the form of motivation commonly referred to as extrinsic motivation, for it involves behaving to obtain a reward or avoid a punishment. This form of extrinsic motivation is the least self-determined and is thus located on the far left end of the continuum. Behaviors that are externally regulated are performed because of pressures, obligations, and constraints coming from external sources. When externally regulated, individuals try to obtain a reward or positive consequence, or they try to avoid some undesirable consequence. Externally regulated behaviors are only performed in the presence of specific external contingencies, such as controlling teachers, strict school requirements, and pressures from peers that are maintaining the behaviors themselves. In the absence of these contingencies, individuals will not remain engaged in the activity, will eventually quit or abandon the activity, or will not be able to persist when the task becomes more difficult. For example, students who show up for an eight o'clock physics test may do so because they do not want to fail the class and are trying to get a good grade on the test. If external regulation is their main motivation for being in class at eight in the morning, many students will not show up for class unless attendance is required or there is a test on a certain day. As constraints, pressures, and rewards shift the perceived locus of causality outside of the individual, external regulation is opposite to intrinsic motivation and consequently not self-determined (Deci *et al.*, 1999).

Introjected Regulation

This underlies behaviors that are performed out of guilt, ego involvement, or other kinds of internal pressures. Similar to external regulation, powerful pressures maintain these behaviors; however, unlike external regulation, the pressures associated with introjection are internal as opposed to external. Introjected regulation is not self-determined because the behaviors are pressured and controlled even though that pressure and control come from an internal source. Individuals who behave mainly out of introjection engage in behaviors because they want to avoid feeling guilty or bad about themselves or because performing the behavior will allow them to aggrandize themselves. For example, a young man is asked to perform in a spelling bee because of how well he would represent the school. Although he does not feel very comfortable with the idea of competing in front of a lot of people, he accepts because he would feel guilty for not participating. He does not want to disappoint his teacher. With introjection, the contingencies maintaining the behaviors have been partially internalized. That is, they are within the person but not fully endorsed by the self. The regulation of the behavior is not yet fully integrated with the individual's motivations, cognitions, and affects into a coherent whole, reflected by the fact that the individual does not really want to perform the activity and does not choose to do it.

Identified Regulation

This exists when the contingencies maintaining a certain behavior become endorsed by the self and the behavior itself becomes valued, thus no longer requiring the specific contingencies. At this point, the behavior is said to be regulated through identification. When individuals are motivated mostly out of identification, they identify with their behaviors, value them, and find them important. The identified regulations are endorsed by the self, which means that although the behaviors are still instrumental in attaining goals (thus still extrinsic), the goals are valued by the individual who identifies with the importance of the activities performed. At this point, the behaviors are considered to be self-determined or autonomous though they are not considered intrinsically motivated. For example, a student might be working on an extracurricular project requiring him to stay after school 3 nights a week. Even though the activity is demanding and might interfere with other things going on at that time, he gladly participates because this extracurricular project is important to him. He does not necessarily find it interesting (and thus not intrinsically motivated) but it is valued.

Integrated Regulation

This is the form of motivation that represents the most self-determined form of extrinsic motivation. When a behavior

is regulated through integration, it is not only valued but also integrated within the self and is in harmony and coherence with other aspects of the self. A graduating senior might talk about her academic experience by discussing the growth she has experienced over the years, the friends she has made, the connections she now has with faculty members, and how all these experiences will benefit her in her chosen career and in becoming the person she aspires to be. The motivation that underlies such statements is a fully self-determined extrinsic motivation. The motivation is still extrinsic because the experiences discussed are in the service of the goal of her future life and career, but these experiences are fully integrated within the self.

By further examining the continuum of self-determination, one can see that intrinsic motivation is located at the far right of the continuum and is also separated from the other types of motivation on the continuum. This is so because intrinsic motivation is inherent to the person rather than resulting from internalization and is based on interest in the activity itself rather than on the importance of the activity for self-selected values and goals. Intrinsic motivation is the prototype of self-determination signifying engagement in activities because the activities themselves are rewarding.

Process of Internalization

Considering the various forms of motivation, including intrinsic and extrinsic motivation, is very useful when trying to explain why someone would behave or act a certain way. Distinctions between the different types of motivation allow researchers to explain a considerable range of human behaviors and experiences. Understanding these various forms of motivation also enables researchers and practitioners to identify antecedents that will foster these different types of motivations and the consequences that may follow these motivated behaviors.

According to SDT, individuals have a natural inclination toward growth and toward integrating the different aspects of the self into a single unified sense of self. The process of internalization is an example of the functioning of this natural movement toward a greater integration of the self and toward more self-determined forms of extrinsic motivation to accompany people's intrinsic motivation. Integration refers to the natural propensity of individuals to take in or internalize the contingencies underlying their behaviors and then to integrate them into their core self.

Antecedents of the Development of Intrinsic and Extrinsic Motivation

Certain social and environmental contexts provide what is essential for the development of self-determination and

intrinsic motivation. In addition to individuals' natural propensity toward internalization, autonomy-supportive settings foster the development of intrinsic motivation. Social and environmental contexts that are autonomy supportive acknowledge people's perspective and provide them with choices and encouragement. Such settings allow self-initiative and provide a rationale for performing requested behaviors. In addition, and very importantly, settings that are autonomy supportive rarely use pressures, controlling strategies, and external contingencies to motivate behaviors; rather, in these settings, empathy for others is frequently expressed and appropriate and timely feedback is offered in order to confirm competence within behaviors or tasks.

Controlling contexts make use of explicitly pressuring and coercive strategies to motivate behavior. For example, these strategies may include salient rewards, competition, deadlines, the threat of punishment, imposed goals, surveillance, and controlling language such as words and phrases like *should*, *have to*, and *ought to* (see Deci *et al.* (1999) for a review). These strategies are most often perceived as controlling because they force individuals to engage in behaviors rather than inviting their engagement. They shift the locus of causality from being perceived as internal to external leading the individual to no longer perceive himself or herself as the causal agent of the behavior.

Environmental factors and social contexts that foster the development of intrinsic motivation and self-determination are also the ones that satisfy the basic psychological needs for autonomy, competence, and relatedness. To be psychologically healthy, individuals must be engaged on a daily basis in activities that contribute to the satisfaction of these three needs.

As mentioned earlier, autonomy signifies choice and volition rather than independence. Autonomy means the endorsement of one's behaviors even though these behaviors might not reflect independence. A girl could decide to get involved in a group project where she will have to share the decision-making power. She will be working with others, not independently, and yet because it is by her own choice, she may feel completely volitional about her decision. In contrast, a boy could feel pressured to study by himself in the library for an important test. Though this behavior would be done independently, he would not want to perform it and would, therefore, not feel a sense of autonomy while doing so.

Competence refers to a sense of mastery. Students need to feel competent while engaged in various academic and nonacademic activities. A sense of competence increases the level of interest and involvement in the activities performed. The need for competence is most likely to be satisfied when students work on developing and mastering skills. An adequate level of challenge, often referred to as an optimal level of challenge, is also necessary in order to

satisfy the need for competence. Again, when this need is satisfied, as when any of the three needs are satisfied, the development of intrinsic motivation and self-determination is furthered.

Relatedness refers to the need to connect with and relate to other people. Students derive a lot of satisfaction from seeing their friends at school everyday, meeting new people, and getting to know their professors. Being with friends in school becomes an important part of going to school and contributes to the extent to which students find school satisfying and enjoyable (Ryan *et al.*, 1994). Furthermore, when students develop healthy relationships with their teachers and other staff members, their productivity and interest level in the classroom improves. This also leads to greater satisfaction and actual enjoyment within a student's academic experience.

Consequences of Intrinsic and Extrinsic Motivation

In general, individuals who behave out of intrinsic motivation or self-determined extrinsic motivation (identification and integration) have been found to experience a greater proportion of positive outcomes with their behaviors. In contrast, individuals who mostly behave out of non-self-determined extrinsic motivation (introjection and external regulation) have been found to experience a much greater proportion of negative outcomes. For example, intrinsic motivation, integration, and identification have been associated with a variety of benefits including academic achievement, school competence, and in general, higher levels of well-being (Boggiano *et al.*, 1993; Levesque *et al.*, 2004; Soenens and Vansteenkiste, 2005). In contrast, introjected regulation and external regulation have been found to lead to a host of negative outcomes including anxiety, burnout, school dropout, and in general, lower levels of well-being and vitality. Paradoxically, the harder the teachers and parents attempt to push students to do their homework, perform well in school, and achieve, the more negative outcomes are experienced without improvements in academic achievement (Deci and Ryan, 2002). By trying to force and push students to perform in school, the opposite results are often obtained. When parents and teachers are autonomy supportive, however, and foster choice and self-determination regarding school-work, students then tend to be more creative, perform better in school, and achieve more academically.

Interplay Between Intrinsic and Extrinsic Motivation

If the movement toward greater internalization and more self-determination is innate, why is it that we do not see

self-determined behaviors to a greater extent in school? A recent meta-analysis summarizing the literature on the effects of rewards on intrinsic motivation highlighted the powerful effects of using rewards to motivate behavior (Deci *et al.*, 1999). Providing rewards for behaviors that are intrinsically motivated or self-determined has been found to undermine the motivation for the activity by shifting the locus of causality outside the self (Deci *et al.*, 1999). To illustrate this process, take the example of Mary, a 5-year-old girl about to begin kindergarten. She has been talking about going to school for the past year, and she is very excited about finally being able to go to school everyday. After her first day at school, she can not stop talking about all the fun games, the school, the teachers, and all that she has done that day. She can't wait to go back! Mary shows intrinsic interest in school for the entire year and is very excited to go back to school and start first grade the following year. However, during that year, something changes. Her teacher begins talking about doing well more often. She makes use of conditional regard in the classroom and has a tendency to compare the students with one another. The teacher offers plenty of rewards for good behaviors and at the beginning of the year instituted a strict behavior-management plan in order to prevent bad behavior in the classroom. Good behavior is rewarded and bad behavior punished. In third grade, Mary is introduced to standardized tests and is constantly reminded by her teacher how important it is to do well on these tests. If asked, Mary would now say that she does not like school and does not want to go anymore. She is afraid and worried on test dates and becomes physically ill sometimes. More generally, the emphasis on standardized tests is perceived as controlling and fosters an external locus of causality, thus the development of non-self-determined extrinsic motivation. The students in Mary's class are criticized when they attempt to find the solution to a problem in a way that differs from the teacher's expectations or the predetermined answer in the textbook. Students in this class feel forced to perform well on assignments and tests, but they do not yet fully understand the concepts they need to apply. As a result, they do not feel comfortable asking questions. They refrain from asking for help because they are afraid they will be perceived negatively. This dynamics will obviously lead to poor performance on assignments and/or tests.

The emphasis on rewards in the traditional school system often works to undermine students' level of intrinsic motivation and self-determination toward school. There is less room for creative activities and exploration because the emphasis is on teaching the standardized tests to the students. The creation of autonomy-supportive learning environments, where choices and options are offered and where creativity is encouraged, would foster the development and maintenance of intrinsic motivation

and self-determination (Levesque *et al.*, 2006). Autonomy supportive environments would enable satisfaction of the needs for autonomy, competence, and relatedness which would then help students to better engage in materials, feel better about themselves, and perform better in the long run (Boggiano *et al.*, 1993; Grolnick *et al.*, 1991). The connection with teachers that is fostered in an autonomy-supportive environment allows students to feel comfortable asking more questions which then leads to overall better performance.

Looking Ahead

There is no doubt that within our education system there is an environment that is inherently restricted by rules and regulations. Grades need to be assigned, skills need to be mastered, and performance needs to meet certain standards of proficiency. However, if the environment in which these constraints are present is consciously crafted to be more autonomy supportive, then intrinsic motivation and self-determination will be more likely to flourish. This in turn will allow students to grow and develop, to reach their potential, and to experience greater levels of well-being (see Vansteenkiste *et al.*, 2006, for a review).

Bibliography

- Boggiano, A. K., Flink, C., Shields, A., Seelback, A., and Barrett, M. (1993). Use of techniques promoting students' self-determination: Effects on students' analytic problem-solving skills. *Motivation and Emotion* **17**, 319–336.
- deCharms, R. (1968). *Personal Causation*. New York: Academic Press.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology* **18**, 105–115.
- Deci, E. L. (1975). *Intrinsic Motivation*. New York: Plenum.
- Deci, E. L., Koestner, R., and Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* **125**, 627–668.
- Deci, E. L. and Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.
- Deci, E. L. and Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry* **11**, 227–268.
- Deci, E. L. and Ryan, R. M. (2002). The paradox of achievement: The harder you push, the worse it gets. In Aronson, J. (ed.) *Improving Academic Achievement: Contributions of Social Psychology*, pp 59–85. New York: Academic Press.
- Grolnick, W. S., Ryan, R. M., and Deci, E. L. (1991). The inner resources for school performance: Motivational mediators of children's perceptions of their parents. *Journal of Educational Psychology* **83**, 508–517.
- Levesque, C. S., Sell, G. R., and Zimmerman, J. A. (2006). A theory-based integrative model for learning and motivation in higher education. *To Improve the Academy* **24**, 85–103.
- Levesque, C., Zuehlke, A. N., Stanek, L. R., and Ryan, R. M. (2004). Autonomy and competence in German and American university students: A comparative study based on self-determination theory. *Journal of Educational Psychology* **96**, 68–84.

- Ryan, R. M., Stiller, J., and Lynch, J. H. (1994). Representations of relationships to teachers, parents, and friends as predictors of academic motivation and self-esteem. *Journal of Early Adolescence* **14**, 226–249.
- Soenens, B. and Vansteenkiste, M. (2005). Antecedents and outcomes of self-determination in three life domains: The role of parents' and teachers' autonomy support. *Journal of Youth and Adolescence* **34**, 589–604.
- Vansteenkiste, M., Lens, W., and Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational Psychologist* **4**, 19–31.
- White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review* **66**, 297–333.

Further Reading

- Ryan, R. M. and Deci, E. L. (2006). Self-regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? *Journal of Personality* **74**, 1557–1585.

Relevant Website

- <http://www.psych.rochester.edu> – Self-Determination Theory, University of Rochester.

Motivating Students in Classrooms

J Brophy[†]

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This article focuses on the challenges facing teachers seeking to motivate their students to learn, along with the strategies they might use for doing so. Research on motivation in education has mushroomed, yielding many principles that are both well grounded in good theory and well supported in empirical research. Unfortunately, most of these are difficult to implement consistently in classrooms, so guidelines for teachers emphasize managing conflicting agendas and achieving the best possible compromise, rather than creating and sustaining the ideal.

Motivation in Classrooms as Expectancy \times Value Reasoning within the Social Context of a Learning Community

Much of what has been learned about motivation can be organized within an expectancy \times value model (Wigfield and Eccles, 2000), which holds that the effort that people are willing to expend on a task is a product of the degree to which they (1) expect to be able to perform the task successfully if they apply themselves and (2) value the task itself or whatever rewards successful task performance will bring. Effort investment reflects the product rather than the sum of the expectancy and value factors because both factors must be present to at least a threshold level. People do not invest in tasks that do not offer enjoyment or valued outcomes, even if they know they can perform the tasks successfully. Nor do they invest in even highly valued tasks if they believe that they cannot succeed, no matter how hard they try.

In addition to their subjective expectancy and value reasoning, students' motivation is affected by their interactions with teachers and classmates. Some classroom climates are supportive of motivation to learn, but others interfere with it. Several recently developed lines of theory and research converge on the conclusion that it is important for teachers to establish their classrooms as collegial learning communities in which students focus on collaborating and supporting one another's learning rather than on self-aggrandizement and competition.

Sociocultural theorists speak of learners as novices undergoing cognitive apprenticeship, under the supervision of mentors, within communities of practice. Everyone

in the community both mentors and learns from others as they collaborate in carrying out the community's activities (Rogoff *et al.*, 2001; Tharp *et al.*, 2000; Wells, 1999). Social constructivist models of learning also depict it as a socially collaborative activity, in which students strive to make sense of new input by relating it to their prior knowledge and collaborating in dialog with their teacher and peers to construct shared understandings (Brophy, 2002; Moll, 1990; Newman *et al.*, 1989). Achievement goal theorists talk about establishing mastery goal structures in classrooms (which focus students on developing knowledge and skills) but avoiding performance goal structures (which encourage students to become more concerned with preserving their self-worth and competing with peers than with mastering the curriculum) (Ames, 1992; Covington, 1992; Ryan and Patrick, 2001; Turner *et al.*, 2002).

These and many other emerging ideas about optimal social contexts in classrooms center around the concept of learning community. Common definitions point to the two key ideas emphasized in the term itself. First, learning implies something more than merely completing academic tasks or even passing tests. School learning is supposed to be enriching and empowering for students, equipping them with important knowledge, skills, values, and dispositions. Second, this learning occurs within a community in which people have social connections and responsibilities toward one another and the group as whole. This implies that the learning will be collaborative as community members encourage and support one another's efforts. Within such a social context, students can feel comfortable asking questions, seeking help, and responding to questions when unsure of the answer because confusion and mistakes are understood as natural parts of the learning process (Baker *et al.*, 1997).

Challenges Facing Teachers in Motivating Students to Learn

Students are motivated to learn when they find classroom lessons and learning activities worthwhile and try to gain the intended learning benefits from them. They may not find these activities pleasurable, exciting, or even interesting, but they engage in them seriously with the intention of pursuing learning goals because they perceive it as worth the effort to do so.

Learning can be enjoyable, but is not fun in the sense that recreational activities are fun. It requires sustained

[†] Deceased

concentration and goal-oriented mastery efforts. In classroom settings, learning situations are complicated in ways that create special motivational challenges for teachers.

First, school attendance is compulsory and the curriculum reflects what society wants students to learn, not what students would choose for themselves. Consequently, teachers often are faced with trying to motivate students to engage in activities they do not value. Second, teachers working with 20 or more students cannot always meet individuals' needs. Some students are often bored while others are often frustrated. Third, classrooms are social settings, so failures often produce not only personal disappointment but also public embarrassment. Fourth, assignments and tests are graded, and reports are sent to parents, which magnifies the consequences of poor performance for struggling students. Finally, teachers and students often settle into familiar routines that become the daily grind. Attention is focused on what must be done to complete activities rather than the knowledge or skills that the activities were designed to develop. Task completion and performance replace motivation and learning as focal concerns (Brophy, 2004).

Expectancy Issues: Supporting Students' Confidence as Learners

The expectancy aspects of motivation involve students' beliefs about their own capabilities. Dweck (1991, 1999) showed that students' engagement in learning activities is affected by their implicit theories about learning ability. Those who subscribe to the entity theory think of ability as a fixed entity over which they have no control, so they are prone to give up and develop helplessness perceptions when they do not succeed easily. In contrast, those who subscribe to the incremental theory believe that ability can be increased incrementally through effort, so they are more likely to persist in their efforts to attain learning goals. Dweck's work implies that teachers should portray learning activities as opportunities to acquire (not just display) knowledge or skill, remind students that mistakes are part of learning, and talk about achieving mastery through successive approximations rather than attaining it quickly and easily.

Weiner (1992, 2001) focused on students' causal attributions – the explanations they generate to explain their successes or failures. Effort and persistence are greater when students attribute their performance to internal and controllable causes. Concerning successful performance, optimal patterns of motivation are associated with attributing success to the combination of sufficient ability and reasonable effort. When students believe that they possessed whatever abilities the task required and were able to meet its demands by applying reasonable effort, they can be confident that they will be

successful on similar tasks in the future. They would have less reason for confidence if they attributed their success to more external, less controllable causes (e.g., the task was easy, I lucked into the solution, I got unexpected help that I probably will not get again).

Concerning unsuccessful performance, effort and persistence are greater when students attribute failures to internal but controllable causes such as insufficient knowledge (of task-relevant information or response strategies) or insufficient effort (they did not prepare or concentrate as much as they should have). These failure attributions enable students to believe that they can improve and ultimately achieve success (by acquiring the needed knowledge or increasing their level of effort). They would have less basis for confidence if they attributed their failures to external causes (e.g., poor textbook or teacher), or worse, to the internal cause of low ability (especially if they had an entity theory of ability). Teachers can encourage productive attributions by portraying learning as acquired gradually through persistent efforts and avoiding any suggestion that failures are due to fixed ability limitations or other factors beyond their students' control.

Self-efficacy perceptions are beliefs in one's capabilities to meet the demands of achievement situations. Students who possess self-efficacy perceptions believe that they can accomplish what the situation calls for, whereas those who lack these perceptions are unsure whether they can succeed or even convinced that they cannot (Bandura, 1997; Bong and Skaalvik, 2003; Pajares, 1996). Teachers can support self-efficacy perceptions by encouraging their students to set specific and challenging but attainable goals, modeling and cueing effective response strategies, providing informative feedback, and helping them to appreciate that they are developing their abilities by accepting challenges and applying consistent effort (Schunk and Ertmer, 2000).

Teachers can support their students' confidence as learners and dispositions to approach learning activities with productive goals and strategies through their approaches to curriculum, instruction, and evaluation.

Curriculum: Program for Success

The simplest way to ensure that students expect success is to make sure that they achieve it consistently. This requires setting tasks at the appropriate level of difficulty (challenging but doable) and then scaffolding their learning efforts through guidance and feedback. Comments on students' progress should stimulate appreciation for their accomplishments to date and imply confidence that they will attain ultimate goals. Both goal setting and feedback are more effective if they focus on the processes or strategies needed to accomplish the task, especially during earlier stages of learning (McNeil and Alibali, 2000; Schunk and Ertmer, 2000; Zimmerman and Kitsantis, 2002).

Instruction: Teach Goal Setting, Performance Appraisal, and Self-Reinforcement

Students may not fully appreciate their accomplishments unless teachers help them to identify and use appropriate evaluation standards. The process begins with setting goals that are proximal rather than distal, specific rather than global, and challenging rather than either too easy or too hard (Locke and Latham, 2002). Students may need help in formulating challenging but reachable goals, especially when perfect performance is unlikely. Richly specified goals help learners to focus on important aspects of the task, motivate their mastery efforts, and empower them with criteria they can use to assess and if necessary adjust their strategies as they work (Page-Voth and Graham, 1999).

In giving feedback, teachers can help students evaluate their performance appropriately, comparing it with absolute standards or their own previous performance rather than with the performance of classmates (Shih and Alexander, 2000). Students who have been empowered with concepts and language needed to evaluate accurately are in a position to reinforce themselves for their successes (that is, attribute them to their willingness to make the needed effort). Students should be taught to think of effort as investment rather than risk: learning may take time and involve confusion or mistakes, but persistence and careful work build knowledge or skills and prepare them to handle more challenging tasks in the future.

Assessment: Emphasize Informative Feedback

Getting and following up productively on informative feedback are integral parts of the learning process, but classroom evaluation and grading systems often undermine students' motivation and learning strategies. If testing and grading are not well matched to everyone's knowledge and skill levels, some students can obtain high grades without exerting much effort and others cannot no matter how hard they try. Assessment information should measure progress toward major instructional goals. Ordinarily, this requires authentic tasks that call for students to integrate and apply what they are learning, not just respond to matching or fill-in-the-blank items (Ames, 1992; Butler, 1987; Crooks, 1988; Wiggins, 1993).

Most motivational researchers advise teachers to include safety nets by allowing failing students to take an alternative test (following a period of review and relearning) or to earn extra credit by producing some product to indicate that they have overcome the deficiencies identified in their test performance. This encourages struggling students to make the extra efforts needed to accomplish the instructional goals (Brophy, 2004).

Certain students become highly anxious and perform considerably below their potential in tests or during any test-like situation in which they are monitored and

evaluated. Several strategies have been developed for minimizing these test anxiety problems: Explain each test's general scope and nature, and how students can best prepare for it; be friendly and encouraging when administering the test; avoid time pressures unless they are truly central to the skill being taught; portray tests as opportunities to assess progress rather than as measures of ability; and teach students effective test-taking skills and attitudes (Hembree, 1988; Hill and Wigfield, 1984; Neveh-Benjamin, 1991; Wigfield and Eccles, 1989; Zeidner, 1998).

Value Issues: Helping Students See Learning Activities as Meaningful and Worthwhile

Whereas the expectancy aspects of motivation surround the question, "Can I do this?", the value aspects surround the question, "Why should I?" Eccles and Wigfield (1985) suggested that subjective task value has four major components: (1) attainment value: the importance of attaining success on the task in order to affirm our self-concept or fulfill our needs for achievement, power, or prestige; (2) intrinsic or interest value: the enjoyment we obtain from engaging in the task; (3) utility value: the role that engaging in the task may play in advancing our career or helping us to reach other larger goals; and (4) cost: the time, effort, and other resources that must be committed to the task, as well as the lost opportunities to devote these resources to other agendas.

Addressing the value aspects of motivation involves inducing students to view their engagement in lessons and learning activities as worthwhile (benefits exceed costs). Traditionally, teachers have been advised to accomplish this either by offering incentives for good performance (extrinsic motivation approach) or by emphasizing content and activities that students find enjoyable (intrinsic motivation approach).

Extrinsic Motivation Approaches

Extrinsic rewards are popular because teachers enjoy giving them and students enjoy receiving them. However, they support motivation to learn only under certain circumstances. First, it is important to deliver rewards in ways that provide students with informative feedback and call attention to significant achievements. This encourages students to apply themselves to their studies because doing so empowers them with knowledge and skills, not just because it can lead to extrinsic rewards (Sansone and Harackiewicz, 2000).

In addition, rewards can act as motivators only for those students who believe that they have a chance to gain them. Opportunity to earn a reward by obtaining a high grade will be motivating to high achievers, but demotivating to

students who have little chance to earn such a grade. Thus, using rewards effectively requires individualized success criteria that allow all students to have comparable access to the rewards. An alternative that avoids this problem is to give rewards only to the class as a whole (“I know that you all put a lot of work into your projects. As a token of my appreciation for your efforts. . .”). Celebrations of everyone’s efforts and progress also are more in keeping with the spirit of a learning community.

Teachers’ praise is another potential source of extrinsic motivation, but again, it is important to deliver it effectively. Students are likely to be motivated by sincere praise delivered privately or through notes on returned assignments, but not by being singled out publicly, especially for things that are not significant achievements (such as sitting up straight and paying attention). Effective praise and encouragement communicate appreciation and informative feedback (not evaluative judgments). They focus on the effort and care that students put into their work, their gains in knowledge or skill, or their achievement’s noteworthy features (meanwhile avoiding attributions of success to high aptitude) (Brophy, 1981; Caffyn, 1989; Henderlong and Lepper, 2002).

Intrinsic Motivation Approaches

The intrinsic motivation approach involves emphasizing content that students are interested in and activities that they enjoy. Interesting activities provide learners with forms of input or opportunities for response that they find rewarding and want to pursue (Renninger and Hidi, 2002; Schraw and Lehman, 2001). This may be because the activities are perceived as relevant or useful; are simply enjoyable or sources of fun; provide self-actualization potential by allowing students to feel empowered or creative; are meaningful or satisfying because they allow students to experience new understandings or gain new skills that they value; or provide opportunities for identification or self-projection (as when they identify with the hero of a story). For example, most students enjoy collaborating in pairs or small groups, activities that provide opportunities to use a wide variety of skills (e.g., conducting and reporting research) rather than limiting them to boring repetition (e.g., filling in blanks on a worksheet), and activities that allow them to create a product that they can point to and identify with (e.g., a display or report) (Alleman and Brophy, 1993–94; Askill-Williams and Lawson, 2001).

Other intrinsic motivation approaches involve adapting school content or activities to students’ interests (Renninger, 2000; Schraw *et al.*, 2001). For example, Hidi and Baird (1988) found that interest in texts was enhanced when their main ideas were elaborated through insertions that featured: (1) character identification (information about people with whom the students could identify, such as the inventors whose discoveries led to the knowledge under

study); (2) novelty (content that was new or unusual); (3) life theme (connections to students’ lives outside of school); and (4) activity level (intense activities or strong emotions). Teachers also can stimulate students’ curiosity or whet their anticipation by introducing content in ways that create a need to resolve some ambiguity or obtain more information about the topic, or by asking questions to put students into an information-processing or decision-making mode (Blank and White, 1999; Reeve, 1996).

Teachers’ beliefs about effective motivation strategies emphasize intrinsic approaches: cooperative learning, stimulating tasks, opportunity to make choices, simulations, projects, learning games, relating content to current events, hands-on activities, and personalized content (Hootstein, 1995; Nolen and Nicholls, 1994). However, teachers typically focus on finding enjoyable activities or adding interesting elements to content rather than on helping students to develop appreciation for the content itself (Zahorik, 1996). Hands-on activities will not produce important learning unless they include minds-on features that engage students in thinking about big ideas.

The self-determination theory of Deci and Ryan (1985, 2000) has been a major source of ideas about supporting intrinsic motivation in classrooms. As part of developing their core thesis that in order to experience a sense of well-being, people require social contexts that meet their three basic needs for autonomy, competence, and relatedness, self-determination theorists have been exploring what is involved in creating autonomy-supportive contexts in classrooms. Operating from intrinsic motivation typically involves exercising autonomy and making free choices. In contrast with controlling contexts, which pressure students to think, act, or feel in particular ways, autonomy-supportive contexts minimize extrinsic performance pressures, provide students with choices, encourage them to solve problems in their own ways rather than insisting on a single method, and invite them to ask questions and suggest ideas for learning activities. Autonomy-supportive teachers empathize with the student’s perspective, seek to facilitate independent thought and decision making, and provide meaningful rationales when choice is constrained. In contrast, control-oriented teachers overmanage their students using detailed instructions backed by rewards, grades, and threats (Reeve *et al.*, 1999; Valas and Sovik, 1994; Vansteenkiste *et al.*, 2006).

Teachers often can allow their students to choose among task alternatives or to exercise autonomy in pursuing alternative ways to meet curricular requirements. For example, they might allow students to select topics for book reports, composition assignments, or research projects, and perhaps also to select from alternative ways of representing their work. When students are likely to make undesirable choices if left completely on their own, teachers can provide a menu of choices to select from or help them make choices that are well suited to their interests and reading levels (Starnes and Paris, 2000; Worthy *et al.*, 2002).

Motivating Students to Learn

Mitchell (1993) distinguished between catching students' interest and holding it. He found that presenting students with brainteasers or puzzles or allowing them to work on computers or in groups were effective for catching initial interest, but not for holding that interest in ways that led to significant learning. The best outcomes were associated with meaningful content (students could appreciate its applications to life outside of school) and instructional methods that fostered involvement (active learning and application activities, not just watching and listening). Other research similarly concluded that the key to motivating students to learn is to structure the curriculum around big ideas and develop them with emphasis on their connections and applications to life outside of school (Hickey *et al.*, 2001; Newton, 2000; Pugh, 2002; Reeve *et al.*, 2002).

Students do not need to enjoy school activities in order to be motivated to learn from them, but they do need to perceive these activities as meaningful and worthwhile. Therefore, teachers need to make sure that their curriculum content and learning activities are in fact meaningful and worthwhile, and then develop the content and scaffold students' engagement in the activities in ways that enable students to appreciate their value.

This begins with content selection and representation. If planning is guided by purposes and goals phrased in terms of desired student outcomes, the curriculum should feature content that students can appreciate as worthwhile and activities that they can appreciate as authentic. The bulk of such a curriculum will consist of coherent networks of connected content structured around powerful ideas. These ideas will be developed in sufficient depth to promote deep understanding of their meanings and connections, appreciation of their significance, and exploration of their applications to life outside of school. As much as possible, this learning will occur through engagement in authentic activities that require using what is being learned for accomplishing the very sorts of life applications that justify inclusion of the content in the curriculum in the first place.

Good content development includes conveying reasons why the content's big ideas are worth learning, explaining when and why they might be used, and modeling how it looks and feels when we use them. Coaching should include goal reminders and encouragement of students' appreciation for the learning domain, and feedback should call attention to developments in students' knowledge or skills, to signs of artistry or craftsmanship in their work, or to unique signature elements that reflect their personal style of operating in the domain (Brophy, 2004).

Optimally mediated learning experiences raise students' consciousness of the purposes and goals of each activity and help them learn not only with understanding but also with

appreciation and life application. Appreciation connotes that students not only understand what they are learning but also value it because they realize that there are good reasons for learning it. These reasons may include not only practical applications but also ways that the learning might enrich the students' repertoires of insights and recognitions or otherwise enhance the quality of their inner lives. Life application implies that the students experience authentic activities that will enable them to apply what they are learning to their lives outside of school.

Conclusion

In conclusion, most discussions of motivation in classrooms emphasize extrinsic praise and rewards or intrinsic motivation strategies for making learning fun, but research findings point to structuring the content around big ideas developed with emphasis on their connections and applications as the key to motivating students to learn. The best results will occur when teachers combine this approach to the value aspects of motivation with an approach to the expectancy aspects that supports all students' confidence as learners and an approach to classroom management that emphasizes learning community principles.

See also: Self-Efficacy Beliefs; Sociocultural Issues in Motivation.

Bibliography

- Alleman, J. and Brophy, J. (1993–94). Teaching that lasts: College students' reports of learning activities experienced in elementary school social studies. *Social Science Record* **30**(2), 36–48; **31**(1), 42–46.
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology* **84**, 261–271.
- Askell-Williams, H. and Lawson, M. (2001). Mapping students' perceptions of interesting class lessons. *Social Psychology of Education* **5**, 127–147.
- Baker, J., Terry, T., Bridges, R., and Winsor, A. (1997). Schools as caring communities: A relational approach to school reform. *School Psychology Review* **26**, 586–602.
- Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York: Freeman.
- Blank, M. and White, S. (1999). Activating the zone of proximal development in the school: Obstacles and solutions. In Lloyd, P. and Fernyhough, C. (eds.) *Lev Vygotsky: Critical Assessments*, pp 331–350. London: Routledge.
- Bong, M. and Skaalvik, E. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review* **15**, 1–40.
- Brophy, J. (1981). Teacher praise: A functional analysis. *Review of Educational Research* **51**, 5–32.
- Brophy, J. (ed.) (2002). *Social Constructivist Teaching: Affordances and Constraints*. New York: JAI Elsevier Science.
- Brophy, J. (2004). *Motivating Students to Learn*, 2nd edn. Mahwah, NJ: Erlbaum.

- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology* **79**, 474–482.
- Caffyn, R. (1989). Attitudes of British secondary school teachers and pupils to rewards and punishments. *Educational Research* **31**, 210–220.
- Covington, M. (1992). *Making the Grade: A Self-Worth Perspective on Motivation and School Reform*. New York: Cambridge University Press.
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research* **58**, 438–481.
- Deci, E. and Ryan, R. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.
- Deci, E. and Ryan, R. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry* **11**, 227–268.
- Dweck, C. (1991). Self-theories and goals: Their role in motivation, personality, and development. In Dienstbier, R. (ed.) *Perspectives on Motivation: Nebraska Symposium on Motivation 1990*, vol. 38, pp 199–235. Lincoln, NE: University of Nebraska Press.
- Dweck, C. (1999). *Self-Theories: Their Role in Motivation, Personality, and Development*. Philadelphia, PA: Taylor and Francis.
- Eccles, J. and Wigfield, A. (1985). Teacher expectations and student motivation. In Dusek, J. (ed.) *Teacher Expectancies*, pp 185–226. Hillsdale, NJ: Erlbaum.
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research* **58**, 47–77.
- Henderlong, J. and Lepper, M. (2002). The effects of praise on children’s intrinsic motivation: A review and synthesis. *Psychological Bulletin* **128**, 774–795.
- Hickey, D., Moore, A., and Pellegrino, J. (2001). The motivational and academic consequences of two innovative mathematics environments: Do curricular innovations and reforms make a difference? *American Educational Research Journal* **38**, 611–652.
- Hidi, S. and Baird, W. (1988). Strategies for increasing text-based interest and students’ recall of expository texts. *Reading Research Quarterly* **23**, 465–483.
- Hill, K. and Wigfield, A. (1984). Test anxiety: A major educational problem and what can be done about it. *Elementary School Journal* **85**, 105–126.
- Hootstein, H. (1995). Motivational strategies of middle school social studies teachers. *Social Education* **59**, 23–26.
- Locke, E. and Latham, G. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist* **57**, 705–717.
- McNeil, N. and Alibali, M. (2000). Learning mathematics from procedural instruction: Externally imposed goals influence what is learned. *Journal of Educational Psychology* **92**, 734–744.
- Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology* **85**, 424–436.
- Moll, L. (ed.) (1990). *Vygotsky and Education: Instructional Implications and Applications of Socio-Historical Psychology*. Cambridge, UK: Cambridge University Press.
- Neveh-Benjamin, M. (1991). A comparison of training programs intended for different types of test-anxious students: Further support for an information-processing model. *Journal of Educational Psychology* **83**, 134–139.
- Newman, D., Griffin, P., and Cole, M. (1989). *The Construction Zone: Working for Cognitive Change in School*. Cambridge, UK: Cambridge University Press.
- Newton, D. (2000). *Teaching for Understanding: What It is and How to Do It*. London: Routledge/Falmer.
- Nolen, S. and Nicholls, J. (1994). A place to begin (again) in research on student motivation: Teachers’ beliefs. *Teaching and Teacher Education* **10**, 57–69.
- Page-Voth, V. and Graham, S. (1999). Effects of goal setting and strategy use on the writing performance and self-efficacy of students with writing and learning problems. *Journal of Educational Psychology* **91**, 230–240.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research* **66**, 543–578.
- Pugh, K. (2002). Teaching for transformative experiences in science: An investigation of the effectiveness of two instructional elements. *Teachers College Record* **104**, 1101–1137.
- Reeve, J. (1996). *Motivating Others: Nurturing Inner Motivational Resources*. Boston, MA: Allyn and Bacon.
- Reeve, J., Bolt, E., and Cai, Y. (1999). Autonomy-supportive teachers: How they teach and motivate students. *Journal of Educational Psychology* **91**, 537–548.
- Reeve, J., Jang, H., Hardré, P., and Omura, M. (2002). Providing a rationale in an autonomy-supportive way as a strategy to motivate others during an uninteresting activity. *Motivation and Emotion* **26**, 183–207.
- Renninger, K. A. (2000). Individual interest and its implications for understanding intrinsic motivation. In Sansone, C. and Harackiewicz, J. (eds.) *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*, pp 373–404. San Diego, CA: Academic Press.
- Renninger, K. A. and Hidi, S. (2002). Student interest and achievement: Developmental issues raised by a case study. In Wigfield, A. and Eccles, J. (eds.) *Development of Achievement Motivation*, pp 173–195. San Diego, CA: Academic Press.
- Rogoff, B., Turkkanis, C., and Bartlett, L. (2001). *Learning Together: Children and Adults in a School Community*. New York: Oxford University Press.
- Ryan, A. and Patrick, H. (2001). The classroom social environment and changes in adolescents’ motivation and engagement during middle school. *American Educational Research Journal* **38**, 437–460.
- Sansone, C. and Harackiewicz, J. (eds.) (2000). *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*. San Diego, CA: Academic Press.
- Schraw, G. and Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review* **13**, 23–52.
- Schraw, G., Flowerday, T., and Lehman, S. (2001). Increasing situational interest in the classroom. *Educational Psychology Review* **13**, 211–224.
- Schunk, D. and Ertmer, P. (2000). Self-regulation and academic learning: Self-efficacy enhancing interventions. In Boekaerts, M., Pintrich, P., and Zeidner, M. (eds.) *Handbook of Self-Regulation*, pp 631–649. San Diego, CA: Academic Press.
- Shih, S. and Alexander, J. (2000). Interacting effects of goal setting and self- or other-referenced feedback on children’s development of self-efficacy and cognitive skill within the Taiwanese classroom. *Journal of Educational Psychology* **92**, 536–543.
- Starnes, B. and Paris, C. (2000). Choosing to learn. *Phi Delta Kappan* **81**, 392–397.
- Tharp, R., Estrada, P., Dalton, S., and Yamauchi, L. (2000). *Teaching Transformed: Achieving Excellence, Fairness, Inclusion, and Harmony*. Boulder, CO: Westview.
- Turner, J., Midgley, C., Meyer, D., et al. (2002). The classroom environment and students’ reports of avoidance strategies in mathematics: A multimethod study. *Journal of Educational Psychology* **94**, 88–106.
- Valas, H. and Sovik, N. (1994). Variables affecting students’ intrinsic motivation for school mathematics: Two empirical studies based on Deci and Ryan’s theory on motivation. *Learning and Instruction* **3**, 281–298.
- Vansteenkiste, M., Lens, W., and Deci, E. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational Psychologist* **41**, 19–31.
- Weiner, B. (1992). *Human Motivation: Metaphors, Theories and Research*. Newbury Park, CA: Sage.
- Weiner, B. (2001). Intrapersonal and interpersonal theories of motivation from an attribution perspective. In Salili, F., Chiu, C., and Hong, Y. (eds.) *Student Motivation: The Culture and Context of Learning*, pp 17–30. New York: Kluwer Academic/Plenum.
- Wells, G. (1999). *Dialogic Inquiry: Towards a Sociocultural Practice and Theory of Education*. New York: Cambridge University Press.
- Wigfield, A. and Eccles, J. (1989). Test anxiety in elementary and secondary school students. *Educational Psychologist* **24**, 159–183.
- Wigfield, A. and Eccles, J. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology* **25**, 68–81.

- Wiggins, G. (1993). Assessment: Authenticity, context, and validity. *Phi Delta Kappan* **75**, 200–214.
- Worthy, J., Patterson, E., Salas, R., Prater, S., and Turner, M. (2002). 'More than just reading': The human factor in reaching resistant readers. *Reading Research and Instruction* **41**, 177–202.
- Zahorik, J. (1996). Elementary and secondary teachers' reports of how they make learning interesting. *Elementary School Journal* **96**, 551–564.
- Zeidner, M. (1998). *Test Anxiety: The State of the Art*. New York: Plenum.
- Zimmerman, B. and Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology* **94**, 660–668.

Further Reading

- Alderman, M. K. (2004). *Motivation for Achievement: Possibilities for Teaching and Learning*, 2nd edn. Mahwah, NJ: Erlbaum.
- Schunk, D., Pintrich, P., and Meese, J. (2007). *Motivation in Education: Theory Research, and Applications*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.
- Stipek, D. (2002). *Motivation to Learn: From Theory to Practice*, 4th edn. Boston, MA: Allyn and Bacon.

Motivation Regulation

C A Wolters, University of Houston, Houston, TX, USA
S A Mueller, Pflugerville ISD, Austin, TX, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Students' motivation within academic tasks is consistently viewed as a critical determinant of their learning and achievement. Students who report increased levels or more adaptive forms of motivation tend to exhibit greater learning and higher levels of achievement. Unfortunately, academic tasks are replete with obstacles that make it difficult for students to be motivated. Students complete cognitively challenging academic tasks, learn material that has little personal relevance, and repeatedly practice basic decontextualized skills. Further, the classroom environment within which these activities are situated is often characterized by multiple tasks occurring at one time, a high level of noise and distractions, and many opportunities for off-task behavior. Hence, even students who begin academic tasks eager to work and be successful may suffer declines in motivation. Motivational demands for completing academic work outside the classroom can be even more burdensome. Students may have to complete homework perceived as difficult, unimportant, or boring, and do so without the guidance or social supports available in the classroom. Despite these problems, many students do exhibit high levels of motivation for academic tasks both in school and outside the classroom context.

Motivational regulation is one explanation for how students' are able to rise above these obstacles and sustain or increase their own motivation for learning. The purpose of this article is to review the conceptual understanding of motivational regulation and to explain its importance to students' academic functioning. To achieve this goal, the article is divided into five sections. In the opening section, we provide a basic conceptual definition of motivational regulation and distinguish it from the process of motivation. In the following section, we identify the theoretical roots of motivational regulation within models of volition and self-regulated learning. Next, we discuss three major dimensions of motivational regulation. Finally, we touch on research examining the antecedents and the outcomes associated with students' motivational regulation.

General Definition and Distinction from Motivation

At the broadest level, the term motivational regulation could be used to describe any efforts to manage students'

achievement motivation. From this perspective, teachers and parents engage in motivational regulation on a regular, if not daily, basis when they purposefully design interesting tasks, promise rewards for hard work, or make statements to bolster students' self-confidence. This more expansive viewpoint, however, is typically supplanted by one that focuses on regulatory efforts that are initiated and directed by students themselves. In line with this focus, the emphasis here is on motivational regulation that is purposeful and is initiated and directed by the student. We define motivational regulation or the regulation of motivation generally as the process through which individuals purposefully manage either their level of motivation or the underlying processes through which motivation is determined. Motivational regulation, therefore, includes thoughts and behaviors through which students act to initiate, maintain, or supplement their willingness to start or to provide effort toward completing academic activities (Wolters, 2003).

Motivational regulation is closely related, but conceptually distinct from motivation itself. Achievement motivation most often refers to students' willingness or desire to engage in, persist, or work hard at academic tasks or to account for goal-directed behavior. Motivation also encompasses the cognitive or affective processes through which this willingness is determined (Anderman and Wolters, 2006). Behaviorally, theories of achievement motivation seek to explain students' choice of activities, the intensity or quality of their effort, and their persistence at academic activities. Most contemporary models explain these outcomes using cognitive constructs such as students' causal attributions, perceptions of self-competence, goals, values, interests, feelings of self-determination, or reasons for engaging in an activity.

Motivation and the regulation of motivation differ most acutely with regard to the awareness and purposefulness of students' thoughts and actions. Regulation of motivation describes the thoughts and actions through which students consciously and intentionally manage their motivation regarding a particular activity (Boekaerts, 1996; Wolters, 2003). This process has a deliberateness or plan-fullness that is unnecessary and absent in most explanations of motivation. Cognitive theories of motivation do not typically propose that students must ever be aware of the underlying processes that determine their motivation. In addition, models of motivation do not advocate that students ever purposefully intervene in these processes.

As an example, students in a class may experience a heightened level of personal interest based on the topic of a particular lesson. This increased interest may, in turn, foster greater attention and persistence for tasks associated with that lesson. Students may not and need not be cognizant of their increased interest for it to affect their behavior. In contrast, motivational regulation would postulate that students understand the positive effects of increased interest and intentionally take steps that will facilitate the interestingness of the lesson. Boekaerts (1992) described this distinction as subjective versus active control. The former refers to actions influenced through motivational beliefs and perceptions, and latter refers to actions determined more by a conscious intent to manipulate.

Theoretical Roots of Motivational Regulation

Research on motivational regulation has emerged from a number of different theoretical traditions. Parts of this process have been investigated by researchers interested in volition, personality, language development, delay of gratification, cognitive and emotional development, behavioral conditioning, and self-regulated learning. These diverse and extensive theoretical roots speak of the importance of motivational regulation as a fundamental aspect of effective functioning in both academic and nonacademic settings. Space limitations prohibit a full and comprehensive consideration of the contributions from each of these perspectives to the overall understanding of motivational regulation. We focus our discussion on volition and self-regulated learning, the two theoretical frameworks most frequently used to understand this process within academic settings.

Motivational regulation is a prominent feature within the research and theory of examining volition. Historically, theories of volition focus on understanding how people are able to maintain the pursuit of goals they have selected in light of competing opportunities and behavioral obstacles (e.g., Corno, 2001; Kuhl, 1985). Corno (2004; Corno and Kanfer, 1993) has served as a primary ambassador for using a volitional framework to understand students' learning and achievement behaviors in academic settings. As a whole, this work emphasizes a distinction between motivational and volitional processes. Motivational processing is viewed as responsible for identifying which goals or behavioral outcomes students decide to pursue. Volition, in contrast, is viewed as critical for protecting those intentions once they have been formed. For example, motivation would account for a student's decision to sit and begin a homework problem set. Volition would account for the student's efforts to remain engaged in this activity until it was finished despite distractions in the environment and/or difficulty with the task. Within this perspective, the switch from

forming intentions to protecting those intentions is considered an important qualitative change that is difficult to undo.

According to volitional theories, students maintain their intention to pursue goals they have selected through various regulatory or control strategies. Based on which aspect of the learning process is being controlled, these strategies have been divided into strategies associated with motivation, as well as attention, encoding, cognition, emotion, and environmental distractions (Corno, 2001). A student who has adopted a goal of preparing for an upcoming exam might go to a quiet place to study (environmental control), use flashcards to memorize course material (encoding control), and reduce anxiety through self-talk (emotion control). Most relevant to the current discussion, students' control of motivation has been described as one type of volitional strategy central to students' efforts to increase their persistence or time on task (Corno, 2004; Kuhl, 1985).

Models of self-regulation emphasize individuals' active, purposeful, and reflective role in their own functioning and behavior within domains that include employment, clinical well-being, chronic illness, and general physical health. The term self-regulated learning is most often applied within social cognitive models that articulate how this process operates in academic contexts. In these contexts, self-regulated learners are viewed as autonomous and efficient learners who have the cognitive abilities as well as the motivational beliefs and attitudes needed to understand and direct their own achievement-related behaviors (Boekaerts, 1996; Pintrich, 2004; Zimmerman, 2000). In particular, these students are knowledgeable about many different cognitive learning strategies and have the metacognitive skills needed to select, monitor, and control their use of these strategies effectively. They also have motivational beliefs and attitudes that drive their engagement and persistence in academic tasks.

Regulation of motivation has become a key feature within several social cognitive models of self-regulated learning (Boekaerts and Cascallar, 2006; Pintrich, 2004; Wolters, 2003). For instance, Pintrich (2004) argued that self-regulated learning entails students' ability to have forethought, monitor, control, and reflect on four areas of academic functioning one of which was motivation. Models advanced by Zimmerman (2000) and Boekaerts (1996) also depict students' understanding and active management of their own motivational processing as an important part of self-regulated learning. Within each of these major models of self-regulated learning, motivation or motivational processing can be the target of students' efforts to plan, monitor, control, and reflect. These efforts, furthermore, are thought to pay dividends to students in terms of increased attention, effort, and persistence that ultimately lead to improved learning and higher achievement.

Dimensions of Motivational Regulation

Drawing from across theoretical perspectives, there appears to be at least three distinguishable dimensions to motivational regulation. One facet reflects the meta-level knowledge or understanding needed to regulate motivation (Boekaerts, 1996; Wolters, 2003). In the research on metacognition, this type of knowledge has been differentiated based on whether it relates to the person, the task, or strategies (Pintrich *et al.*, 2000). Students' knowledge of what topics, domains, or activities they personally find interesting, fun, or motivating in general would reflect person-related meta-motivational knowledge. Insight into the types of tasks that are more or less motivating in general or that may make it difficult to be motivated because they are boring, frustrating, or lack usefulness would be task-related meta-motivational knowledge. Consistent with views of metacognition, meta-level knowledge related to motivational regulation strategies is likely to include declarative, procedural, and conditional forms of knowledge. Some students, for example, have been found to use self-talk to highlight features of completing the task that they find motivating (Wolters, 1998). Before engaging in this sort of strategy, however, students must know about it as a potential strategy, must know how to enact the steps needed to accomplish it properly, and must believe that using the strategy will lead to some desired effect on motivation within a particular situation. Students lacking any of this knowledge are unlikely to use this particular motivational regulation strategy.

A second dimension of effective motivational regulation is the monitoring of one's level or state of motivation. Students can become aware or self-assess their motivation before a task begins (prediction of motivation), during a task (experience of motivation), and after a task has been completed (reflection on motivation). This consideration of one's level of motivation is necessary to identify when motivation is waning, and is thus a prerequisite to any active intervention designed to bolster one's motivation. As an example, a person is unlikely to take steps that will make a task more enjoyable if he/she does not first become aware that he/she is bored and losing interest in finishing the task. This type of self-awareness may include monitoring both the level of motivation as well as the nature of one's motivation. The former ensures that the willingness or quantity of effort needed to complete a task is available whereas the latter may be necessary to ensure that the type of motivation driving one's engagement will lead to the desired outcomes. Much like the monitoring of one's comprehension or understanding outlined as part of metacognition, this process may continue without much conscious effort until a problem is encountered. Nonetheless, without effective monitoring, students may not successfully regulate their motivation for academic tasks.

A third dimension of effective motivational regulation is the purposeful or active efforts to intervene and control

one's own motivation for a task. This process encompasses the actual strategies one engages in, to control the level or nature of motivation. For instance, once a student recognizes that she has little interest in tackling a required task she must do something to increase her interest, or to foster some other form of motivation. Given the diverse nature of motivation, the particular methods that might be used to affect motivation for a task are likely to be quite varied. Prior work examining motivational regulation, in fact, has identified a number of strategies that students might use to control their motivation (Wolters, 2003). These strategies include attempts to regulate various motivational beliefs that have been discussed in the achievement motivation literature such as goal orientation, self-efficacy, task value, and interest in the task. When initiated in order to control outcomes such as effort and persistence, students' management of their affect, environment, and behavior might also be considered forms of motivational regulation (Boekaerts, 1996; Wolters, 2003).

Antecedents of Motivational Regulation

The consideration of factors that determine or influence whether students engage in motivational regulation is important for both its theoretical and practical implications. Consistent with views about self-regulation and volition more generally (Boekaerts and Corno, 2005; Pintrich, 2004), motivational regulation is likely a function of both stable individual differences among students, as well as more situational influences within the instructional context and the broader environmental context. Recent work in this area also argues that all aspects of self-regulation, including motivational regulation, are developed through social and cultural interactions/influences. In this section, we briefly identify factors from within each of these areas that have been discussed as potential influences on students' motivational regulation.

Individual differences among students, including aspects of their cognitive development, personality, gender, and ethnic identity may affect the extent to which they regulate their motivation. Developmentally, research suggests that forms of motivational regulation may emerge at a very young age. Children as young as 3 years old use strategies to creatively modify academic tasks to regulate their motivation, and even before reaching school age there is evidence that children use strategies to block distractions in order to complete tasks (Corno, 1994; Metcalfe and Mischel, 1999). Still, as suggested by the extant research on metacognition and self-regulated learning, the most sophisticated forms of motivational regulation may not develop until adolescence. Moreover, students who are more conscientious or who identify more closely with doing well in school may be more apt to regulate their motivation for academic tasks.

Another determinant of students' efforts to regulate their motivation is likely the motivational beliefs and attitudes they bring to a particular situation or task. Past work on the more cognitive aspects of self-regulated learning has firmly established that motivation impacts students' use of meta-cognitive regulatory strategies. As motivational regulation represents a similar expression of self-management, it follows that students who hold more adaptive motivational beliefs would be more apt to utilize motivational-regulation strategies than students who hold less-adaptive motivational beliefs (Boekaerts, 1996; Corno, 1994). The view that motivation can be both a cause and an outcome of students' regulatory processing is consistent with the view of self-regulated learning as a complex, iterative process. Further, prior research provides some evidence supporting this additional link between motivation and motivational regulation. Wolters and Rosenthal (2000), for instance, found that students who viewed the material they were learning as more important or useful, or who focused on mastery goals tended to report greater use of motivational regulation strategies. In this same study, however, self-efficacy was not an important predictor of students' motivational regulation. Hence, a more complete understanding of these relations remains an important goal for future research.

Situational factors, including the instructional environment specifically, represent another possibly potent influence on students' motivational regulation. For instance, teachers can structure their classrooms to encourage motivational-regulation strategies by allowing students time to reflect, the autonomy needed to control aspects of the task, and by providing opportunities for students to observe models of this process. In addition, parents can foster the different facets of motivational regulation through modeling and by providing opportunities for self-management, and through other parenting practices. Students can be trained to use strategies for motivational regulation, like other forms of volition and self-regulation, through more direct interventions.

Broader sociocultural processes can also be viewed as an important influence on the development of motivational regulation. For instance, Jarvela and Volet (2004) describe motivation as a dual psychological-social phenomenon because it involves the development of individual abilities through shared, social, or interactive processes. Others have argued that motivational regulation and other similar processes are best conceptualized using the term co-regulation because they grow from the scaffolding, intersubjectivity, and cultural supports provided through social interactions (McCaslin *et al.*, 2006). Overall, these views emphasize that the knowledge, beliefs, attitudes, and skills individuals need to regulate their motivation emerge from shared, dynamic, and socially supported experiences within culturally meaningful contexts.

Outcomes Associated with Motivational Regulation

Successful motivational regulation should, of course, have a positive influence on students' engagement, effort, or persistence for academic tasks. Ultimately, students who regulate their motivation most effectively should also exhibit better academic outcomes than students who fail to self-regulate their motivation. These benefits may be even more acute when a situation is particularly detrimental to ongoing motivation. Empirical evidence that expressly documents these anticipated relations, however, continues to be in short supply.

Some evidence for these links comes from research investigating more global indicators of students' regulatory activities. For instance, prior research has shown a link between overall measures of volitional control and positive educational outcomes (Corno and Kanfer, 1993; Kuhl, 1992). These studies have not, however, tended to assess and test motivational control or the regulation of motivation separately from other forms of volition. Moreover, these models have assumed a fairly broad view of motivational regulation. They have not accounted for potential differences in the types of motivational-regulation strategies that students use.

Two studies by Wolters (1998, 1999) contribute more direct evidence for the positive effects of motivational regulation. The college students studied by Wolters (1998) reported using many different types of motivational-regulation strategies. Findings from this study indicated that students who reportedly exerted some purposeful control over their level of motivation for academic tasks tended to have more positive academic outcomes than other students. Similarly, Wolters (1999) found that adolescent students' motivational regulation was related positively to their use of learning strategies, effort, and classroom performance. Not all studies examining these relations, however, have found a positive relation between indications of students' motivational regulation and achievement (Xu and Corno, 2003). An increased understanding of the specific motivational-regulation strategies students use at different ages, and how these strategies are related to other self-regulatory processes and to academic performance is an important goal for future research.

Conclusions

Motivation is an important process that undergirds students' engagement and achievement in academic contexts. Students' ability to understand and purposefully influence their motivation, therefore, is a potentially critical influence on their learning and achievement. Motivational regulation

has emerged and proven useful as a model for conceptualizing and understanding this process. Further, evidence specifically documenting the nature and importance of this process has begun to emerge. Students' ability to regulate their motivation has not, however, received the same level of attention as their ability to manage their cognitive processing and there is much left to investigate. More attention is needed for a better understanding of the different components of motivational regulation, as well as the most critical antecedents and substantial outcomes of this process both in the classroom and in academic tasks completed in other contexts.

See also: Achievement Goal Theory: Definitions, Correlates, and Unresolved Questions; Attribution Theory; Coping with Stressful Situations: An Important Aspect of Self-Regulation; Interest; Intrinsic and Extrinsic Motivation; Metacognition; Motivating Students in Classrooms; Self-Efficacy Beliefs; Self-Regulated Learning and Socio-Cognitive Theory; Volitional Control of Learning.

Bibliography

- Anderman, E. and Wolters, C. (2006). Goal, values, and affect. In Alexander, P. and Winne, P. (eds.) *Handbook of Educational Psychology*, 2nd edn, pp 369–389. Mahwah, NJ: Erlbaum.
- Boekaerts, M. (1992). The adaptable learning process: Initiating and maintaining behavioral change. *Applied Psychology: An International Review* **41**, 377–397.
- Boekaerts, M. (1996). Self-regulated learning at the junction of cognition and motivation. *European Psychologist* **1**(2), 100–112.
- Boekaerts, M. and Cascallar, E. (2006). How far have we moved toward the integration of theory and practice in self-regulation? *Educational Psychology Review* **19**, 199–210.
- Boekaerts, M. and Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology: An International Review* **54**, 199–231.
- Corno, L. (1994). Student volition and education: Outcomes, influences and practices. In Schunk, D. and Zimmerman, B. (eds.) *Self-Regulation of Learning and Performance: Issues and Educational Applications*, pp 229–251. Hillsdale, NJ: Erlbaum.
- Corno, L. (2001). Volitional aspects of self-regulated learning. In Zimmerman, B. and Schunk, D. (eds.) *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*, 2nd edn, pp 191–225. Mahwah, NJ: Erlbaum.
- Corno, L. (2004). Introduction to the special issue work habits and work styles: Volition in education. *Teachers College Record* **106**, 1669–1694.
- Corno, L. and Kanfer, R. (1993). The role of volition in learning and performance. In Darling-Hammond, L. (ed.) *Review of Research in Education*, vol. 19, pp 301–341. Washington, DC: AERA.
- Jarvela, S. and Volet, S. (2004). Motivation in real-life, dynamic, and interactive learning environments: Stretching constructs and methodologies. *European Psychologist* **9**, 193–197.
- Kuhl, J. (1985). Volitional mediators of cognition-behavior consistency: Self-regulatory processes and action versus state orientation. In Kuhl, J. and Beckman, J. (eds.) *Action Control: From Cognition to Behavior*, pp 101–128. New York: Springer.
- Kuhl, J. (1992). Recurrent issues in self-regulation research: A rejoinder. *Applied Psychology: An International Review* **41**(2), 160–173.
- McCaslin, M., Bozack, A., Napoleon, L., et al. (2006). Self-regulated learning and classroom management: Theory, research, and considerations for classroom practice. In Evertson, C. and Weinstein, C. (eds.) *Handbook of Classroom Management: Research, Practice, and Contemporary Issues*, pp 223–252. Mahwah, NJ: Erlbaum.
- Metcalfe, J. and Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review* **106**, 3–19.
- Pintrich, P. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review* **16**, 385–407.
- Pintrich, P., Wolters, C., and Baxter, G. (2000). Assessing metacognition and self-regulated learning. In Schraw, G. (ed.) *Metacognitive Assessment*. Lincoln, NE: University of Nebraska Press.
- Wolters, C. (1998). Self-regulated learning and college students' regulation of motivation. *Journal of Educational Psychology* **90**, 224–235.
- Wolters, C. (1999). The relation between high school students' motivational regulation and their use of learning strategies, effort, and classroom performance. *Learning and Individual Differences* **11**, 281–299.
- Wolters, C. (2003). Regulation of motivation: Evaluating an underemphasized aspect of self-regulated learning. *Educational Psychologist* **38**, 189–205.
- Wolters, C. and Rosenthal, H. (2000). The relation between students' motivational beliefs and attitudes and their use of motivational regulation strategies. *International Journal of Educational Research* **33**, 801–820.
- Wolters, C., Pintrich, P., and Karabenick, S. (2005). Assessing academic self-regulated learning. In Moore, K. and Lippman, L. (eds.) *What Do Children Need to Flourish?: Conceptualizing and Measuring Indicators of Positive Development*, pp 251–270. New York: Springer.
- Xu, J. and Corno, L. (2003). Family help and homework management reported by middle school students. *Elementary School Journal* **103**, 503–517.
- Zimmerman, B. (2000). Attaining self-regulation: A social cognitive perspective. In Boekaerts, M., Pintrich, P., and Zeidner, M. (eds.) *Handbook of Self-Regulation*, pp 13–39. San Diego, CA: Academic Press.

Further Reading

- Boekaerts, M., Pintrich, P., and Zeidner, M. (eds.) (2000). *Handbook of Self-Regulation*. San Diego, CA: Academic Press.
- Pintrich, P. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research* **31**, 459–470.
- Schunk, D. and Zimmerman, B. (1994). *Self-Regulation of Learning and Performance: Issues and Educational Applications*. Hillsdale, NJ: Erlbaum.
- Zimmerman, B. (1994). Dimensions of academic self-regulation, A conceptual framework for education. In Schunk, D. and Zimmerman, B. (eds.) *Self-Regulation of Learning and Performance: Issues and Educational Applications*, pp 2–21. Hillsdale, NJ: Erlbaum.
- Zimmerman, B. and Schunk, D. (2001). *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*, 2nd edn. Mahwah, NJ: Erlbaum.

Peer Learning in the Classroom

N M Webb, University of California, Los Angeles, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Peer Learning in the Classroom

Recognizing that students can learn by working with and helping each other, school districts, state departments of education, national research organizations, and curriculum specialists recommend, and sometimes mandate, the use of peer-based learning. Undergirding these recommendations is a large research literature showing positive effects of peer-based instructional methods on student achievement, compared with other forms of instruction that involve little interaction between students such as teacher-led whole-class instruction or individual work.

However, simply placing students in small groups does not guarantee that learning will take place. Rather, the extent to which students benefit from working with other students (variously referred to as cooperative learning, collaborative learning, peer tutoring, and peer-based or peer-directed learning) depends on their interaction. This article explores the mechanisms by which working with other students is thought to benefit student learning, and the many ways in which small-group work might be orchestrated to bring about those benefits.

Social-Behavioral Perspectives on the Benefits of Peer Interaction

Social-behavioral perspectives hypothesize that working together, helping each other, and supporting each other's contributions will lead to increased effort, greater learning, and more liking of the task and other students than instructional settings without such opportunities for student interaction.

Motivational Perspectives

One way to motivate students to collaborate is to have groups work toward a common goal (Deutsch, 1949). In a group or cooperative goal structure, group members can attain their own personal goals only if the group is successful. A group goal may be based on rewards or incentives. In Slavin's (1992) Student Teams-Achievement Divisions (STAD), for example, teams are rewarded based on the improvement of group members' achievement. Since the team's success depends on the progress of all group members, these group rewards ensure individual accountability, a feeling of personal responsibility for what happens in the

group. This motivates students to work hard toward the group goal and encourage and help others do the same.

To limit undue focus on the reward itself and maintain focus on learning and task performance, the group incentive (e.g., a team certificate or a mention in a classroom newsletter) is not designed to be highly desirable or have high stakes. To ensure student effort and reduce pressure on less-capable group members, task difficulty is tailored to the capabilities of individual team members and no one is assigned overly difficult tasks. Cooperative learning methods that use both group rewards and individual accountability show positive effects on achievement compared to cooperative learning methods that include only one or neither of these components.

Social Cohesion Perspectives

Creating socially cohesive groups is also thought to motivate students help each other because they care about the group and its members. To promote a sense of group identification and concern for others, some cooperative learning methods, for example, Johnson and Johnson's (1994) *Learning Together and Alone: Cooperative, Competitive, and Individualistic Learning* uses teambuilding and development of social skills. Teambuilding activities help students get to know each other and experience success as a team. Training in social skills such as active listening, stating ideas freely, accepting responsibility for one's behaviors, taking turns, sharing tasks, making decisions democratically, and trying to understand others' perspectives helps group members trust, accept, and support each other, communicate accurately and effectively, and resolve conflicts constructively.

Another way to build group cohesion is to provide feedback on group functioning. By discussing their respective group's interaction and how they might improve it (group processing), groups can identify and solve general communication problems (e.g., disruptive or bullying behavior) and reinforce each other's collaborative efforts.

Cognitive/Developmental Perspectives on Learning from Peers

Piagetian Perspective

According to Piaget (1932) and his followers, cognitive conflict leads to higher levels of reasoning and learning. The conflict arises when learners perceive a contradiction between their existing understanding and what they hear

or see in the course of interacting with others. To resolve the conflict, learners reexamine and question their own ideas and beliefs, seek additional information, and try out new ideas. Exchanges with peers are more likely to promote learning than exchanges with adults because students are more likely to cooperate as equals and challenge each other, exercise mutual control over the interaction, and share each other's language and point of view.

Supporting evidence of the benefits of cognitive conflict comes from research on the cognitive development of pairs of children working on conservation tasks to discuss whether some characteristics of objects (e.g., the volume of liquid) remain the same when others (e.g., the width or height of the container) change. When children who have not yet learned the principle of conservation are paired with children who have mastered it, the former often gain whereas the latter rarely regress to giving nonconserving responses. Moreover, pairs of nonconservers progress when they hold conflicting, but still incorrect, ideas, but not when they hold the same incorrect conception. Benefits of resolving conflicting perspectives also appear in studies of larger groups in which students initially hold different conceptions.

Vygotskian perspective

According to Vygotsky (1978), learning can occur when a more-expert person helps a less-expert person. With the help of a more-skilled person, a process of negotiation and transformation enables the less-competent person to carry out a task or solve a problem that the latter student could not perform without assistance, a process sometimes referred to as scaffolding or guided participation. Through this process, the less-proficient student can internalize skills and knowledge that he/she has practiced and developed such that they become part of his/her individual repertoire.

Students can often provide effective scaffolding and support due to their understanding of, and familiarity with, each other's misunderstandings and their ability to explain concepts in familiar terms. Students being helped will benefit most when they have, and use, opportunities to apply explanations received to solve the problem or carry out the task for themselves.

Co-Construction of Knowledge

Students can also learn by co-constructing knowledge with their peers, for example, when students contribute different pieces of information or build upon others' explanations to jointly create a complete idea or solution. Students can collaboratively build knowledge and problem-solving strategies that no group member has at the start by acknowledging, clarifying, correcting, adding to, building upon, and connecting each others' ideas and suggestions.

Such interactions have been shown to help children co-construct and internalize strategies and concepts in a wide variety of areas including, for example, identifying chemicals in chemical reactions, understanding of place value, learning how to multiply numbers, and constructing mental models of the nature of matter.

Co-construction may require a high degree of coordination among group members. In highly coordinated interaction, students acknowledge each other's ideas, repeat others' suggestions, and elaborate on others' proposals (Barron, 2000). Speakers' turns are tightly connected, and group members pay close attention, and respond, to what other members do and say, give space for others' contributions, and monitor how the unfolding contributions relate to the problem-solving goal.

Cognitive Elaboration Perspective

From a cognitive elaboration perspective, interacting with others may encourage students to restructure their own knowledge and understanding. Specifically, explaining the material to others may promote learning by encouraging explainers to rehearse information, reorganize and clarify material, recognize misconceptions, fill in gaps in their understanding, strengthen connections between new information and previously learned information, internalize and acquire new strategies and knowledge, and develop new perspectives and understanding. When formulating an explanation, students may think about the salient features of problems and generate self-explanations that help them internalize principles, construct specific inference rules for solving problems, and repair imperfect mental models (Chi, 2000). This process may help students monitor their own understanding and develop a metacognitive awareness of what they do and do not understand.

The social component of explaining to others is important. Consistent with the idea that the process of explaining to someone else leads to more differentiated, complex, unified, and organized cognitive structures than does merely learning the material for oneself are findings that vocalizing to a peer (presumably to teach that person) produces greater concept attainment than vocalizing to an experimenter (presumably only to demonstrate mastery of the material). By accommodating explanations to the difficulties of other students, helpers may construct more elaborate conceptualizations than they would when solving the problems for themselves.

The strong relationship between giving explanations and achievement in small groups has been well documented (Webb and Palincsar, 1996). Moreover, giving more-elaborated explanations may be more effective for learning than giving less-elaborated explanations, for example, providing multiple reasons versus a single reason about the role of resistors in electric circuits.

Debilitating Interpersonal Processes

Groups may not function in ways that are optimal for learning. A number of detrimental processes have been documented.

Unequal Participation

Some students participate more than others due to personality characteristics (extroverted and energetic members may do most of the talking) or status characteristics. High-status students, especially on academic standing or popularity, tend to be more active and influential than low-status individuals, while low-status individuals tend to be less assertive, talk less, and give fewer suggestions and less information than high-status individuals. Both real and artificial differences in task-related competence (e.g., classifying students' competence on the basis of fictitious test scores) can create imbalances in activity and influence. Status may also be linked to social characteristics, such as gender or race, with boys and white students being more active than girls and colored students.

Sometimes, students choose not to participate. One or more group members may sit back and let others do the work (referred to as social loafing or diffusion of responsibility). Individuals may go along for a free ride if they believe that their efforts cannot or will not be identified or are dispensable. The free-rider effect can turn into the sucker effect when the group members who were doing all of the work discover that they have been taken for a free ride and stop working to avoid being suckers.

Students who are not involved in group interaction will not experience the benefits of active participation described earlier. Students who do participate will not benefit from the knowledge and perspectives of the passive students.

Many cooperative learning methods encourage participation of all group members. In *The Jigsaw Classroom* (Aronson *et al.*, 1978), students are assigned responsibility for mastering a portion of the material, discussing that material with other students assigned the same topic, and teaching their topic to the other members of their groups. The group-investigation method by Sharan and Hertz-Lazarowitz (1980) requires students to carry out research on a piece of a group project and then work together as a team to integrate their findings and plan their class presentations.

Failure to Seek and Obtain Effective Help

Another set of debilitating processes concerns the failure of students to seek help when they need it or to obtain effective help when they seek it. Students may fail to seek help because they lack the metacognitive skills necessary to monitor their own comprehension, or may watch their

teammates solve a problem or accomplish a task and assume that they can do it too (Nelson-Le Gall, 1992).

Students may decide not to seek help for fear of being judged academically or socially incompetent, or they may not want to feel indebted to those giving the help (Newman, 1998). Alternatively, students may believe that help seeking is undesirable (as a result of classroom norms to be quiet and work alone without disturbing others) or may have received antagonistic or unsatisfactory responses to previous help-seeking attempts. Students may believe that no one in the group has the competence or resources to help, or they may lack a sense of responsibility or motivation to do the work. Finally, they may believe themselves to lack the competence to benefit from help that others may provide.

Even if students do seek help, their help-seeking strategies may be ineffective. Students may select helpers who are nice or kind, or who have high status, rather than those who have task-relevant skills. Conversely, they may ask vague, indirect, or unfocused questions, which are not likely to elicit explanations, rather than questions that are explicit, precise, direct, and targeted to a specific aspect of the problem or task. Asking precise questions makes it easier for other group members to identify the student's misconceptions or areas of confusion and formulate effective help accordingly. Precise questions may also signal to the group that the help seeker wants to learn, and has sufficient understanding to be able to profit from explanations provided, and thus motivate the group to help. General questions, in contrast, may signal a lack of ability or of effort (it takes less effort to declare general confusion than to formulate precise questions), or both. Students who appear to be loafing or who give up easily may be less likely to receive help than those who appear to be working hard.

Even if students are willing to help their teammates, they may not have the skills to provide effective explanations. Help givers may have misconceptions themselves, not be able to translate their thinking into appropriate language, use confusing language, not provide enough detail, dictate how to solve a problem or complete a task without referencing the needs of the help seeker, or they may not monitor other students' comprehension and thus be unaware of specific misconceptions that need to be addressed.

Whether students obtain help may depend on the group's composition and a student's relative position in the group. For example, in groups that are academically heterogeneous, middle-ability students may be left out of teacher-learner relationships that emerge between high-ability and low-ability group members.

Too Little or Too Much Cognitive Conflict

Although students can learn by resolving discrepancies in ideas, too little or too much conflict may be detrimental. Infrequent conflict may reflect suppression of

disagreements, either from the domination of one group member over the others or from social pressures not to challenge others. Too much conflict may prevent group members from moving forward, especially if group members engage in an adversarial or conflictual style of argumentation instead of a co-constructive style in which group members work together to critique suggestions and create new solutions.

Lack of Coordination

Group functioning may also suffer from uncoordinated communication (Barron, 2000), marked by low levels of attention to, and uptake of, members' suggestions (even correct ones), and by students advocating and repeating their own positions and ideas, ignoring others' suggestions, rejecting others' proposals without elaboration or justification, interrupting others, or talking over them simultaneously. Lack of coordination and joint attention may undermine many of the processes by which individuals can gain by collaborating with others, such as resolving conflicts and co-constructing knowledge, as well as reduce group cohesion and students' motivation to work together.

Other Negative Socioemotional Processes

Other negative socioemotional processes, such as rudeness, hostility, and unresponsiveness, may also impede the participation and learning of group members. Rudely disagreeing with others and ignoring their suggestions may prevent groups from solving problems correctly. Aggressiveness, hostility, and insulting behavior may lead to unconstructive and bitter arguments and may cause students to withhold knowledge and ideas from the group or decide not to seek help.

Approaches to Promoting Beneficial Peer Interaction

Researchers have designed a variety of collaborative approaches to promote beneficial peer interaction and inhibit detrimental group dynamics. Many, if not most, peer-based methods incorporate one or more of the strategies described here.

Preparing Students to Work Collaboratively

Altering expectations and status relationships

To promote equal participation among students in heterogeneous (especially multiracial) groups, Cohen and Lotan (1995) developed methods of minimizing status effects by altering high-status students' expectations about low-status students' competence. By training low-status students on academic and nonacademic tasks which they then teach to

high-status students, high-status students change their perceptions about low-status students' competence. A related approach is the multiability intervention, which raises students' awareness of the multiple skills necessary to perform the task. The teacher discusses with students the multiple abilities needed to solve complex problems (e.g., visual thinking, intuitive thinking, and reasoning) and, when groups work on these tasks, points out the particular contributions of specific students (particularly of low status) and how these are important and valuable.

These approaches have shown success in reducing the relationship between status (based on language background, race, socioeconomic status, and academic ability) and behavior in small groups. For example, the more frequently teachers talk about the multiple abilities needed for a task (and the fact that no one has all of the abilities) and assign competence to low-status students (e.g., observing and commenting on a non-English speaker's ability to build structures based on drawn diagrams), the greater is the participation rate of low-status students, and the smaller is the gap between the participation rates of high-status and low-status students.

Instruction in explaining and group reasoning skills

As an adjunct to providing social skills training designed to improve communication in groups, a number of studies have incorporated instruction in academic helping, explaining, and help-seeking behaviors, such as asking clear and precise questions, giving explanations instead of answers, monitoring understanding of teammates, checking others' answers, and giving specific feedback on the problem-solving strategies of their teammates. Some training approaches focus on helping students make their reasoning explicit and engage constructively with others' ideas (e.g., sharing relevant information, reaching agreement, taking responsibility for decisions, providing reasons, challenging others, and discussing alternatives before making decisions). Compared to untrained groups, this preparation for working in groups has been shown to produce more explaining and often higher achievement (Gillies and Ashman, 1996).

Structuring Peer Interaction

To supplement training in communication and explaining skills, some peer-based methods structure group interaction in specific ways or implement activities to guide groups' collaboration. These methods require groups to carry out certain strategies or activities, or assign students certain roles to play, or both.

Reciprocal teaching

Palincsar and Brown (1984) developed teacher scaffolded instruction, referred to as reciprocal teaching, to help students carry out certain strategies designed to improve comprehension of text: generating questions about the

text they have read, clarifying what they do not understand, summarizing the text, and generating predictions. Teachers initially take the leadership in explaining the strategies and modeling their use, and gradually help students become proficient (by asking students to demonstrate the strategies and giving them feedback) so that they can carry out the strategies in their groups. Students using reciprocal teaching methods have shown improvement in reading comprehension and better performance than students receiving other kinds of instruction with and without peer interaction.

Explanation prompts

Some peer-learning approaches give students specific prompts to encourage them to exchange elaborated explanations. Students may be prompted to describe what happened in their experiments, find patterns in their results, and explain why their results occurred. Alternatively, they may be prompted to construct explanations, justify answers and beliefs, relate what they learned in class to the task at hand, use distinguish between scientific and everyday definitions and explanations, and to compare real-world experiences to class learning. The use of these explanation prompts produces conceptually advanced explaining and more accurate and complete understanding of the material.

Guided reciprocal questioning

In guided reciprocal questioning, students ask each other high-level questions about the material to help them monitor their own and each others' comprehension and encourage students to describe and elaborate their thinking. King (1992), for example, gave students how and why question stems to guide their discussions of text (e.g., "Why is . . . important? How are . . . and . . . similar?"; p. 113). In problem-solving contexts, students may ask each other questions to help them reflect on problems before solving them (e.g., identifying known and unknown information), make connections between current problems and previous ones, and generate and defend their choice of problem-solving strategies. The use of high-level questions by the groups increases the frequency of elaborated explaining and student achievement.

Structured controversy

To promote the benefits that arise from resolving conflicts, Johnson and Johnson's (1995) structured controversy approach subdivides groups into teams, requires them to master material on different sides of an issue and debate the issue with the other team, and then to work as a group to synthesize the two positions. Groups required to debate the issues often carry out more high-level discussion of material and show higher achievement than groups required to seek concurrence by working cooperatively and compromising.

Cognitive role specialization

Some approaches assign students roles to play (usually alternated or rotated) based on specific cognitive activities. Students may play the role of recaller (also referred to as learning leader or summarizer) or listener (also referred to as active listener, learning listener, or listener/facilitator), which are sometimes incorporated into scripts for groups to follow (O'Donnell, 1999). The recaller summarizes the material and the listener is responsible for detecting errors, identifying omissions, and seeking clarification. Students work together to elaborate on the material, and then change roles for the next part of the task. A large body of research shows that such scripted cooperation usually produces greater elaboration of ideas and higher student achievement than does unstructured cooperation.

Tutor and tutee roles are widely used. In reciprocal peer tutoring, students receive training in how to model strategies such as summarizing text, and how to give explanations, corrections, and feedback about other students' work, and then alternate tutor and tutee roles during pair work. Some approaches pair more-skilled and less-skilled learners, whereas others pair students randomly or with similar proficiency. Some approaches specifically focus on promoting a high-level of discourse during paired discussions, such as training tutors to give highly elaborated conceptual rather than algorithmic explanations to their partners, or training tutors to push tutees to give high-level explanations that make connections among ideas.

Manipulating the group-work task

Equal student participation can also be encouraged through the use of complex tasks or open-ended problems without clear-cut answers or procedures that require the combined expertise of everyone in the group. Such tasks encourage groups to value the different contributions that students can make, whereas narrowly defined tasks or problems, especially those that can be completed by one student with the requisite skills, may limit the participation of some students.

The Teacher's Role

Developing Classroom Goal Structures and Norms

The teacher can influence peer interaction by working with students to mutually construct norms for student engagement. He/she can raise the level of discussion by monitoring and intervening in small-group dialogs to remind students about their obligations (e.g., to share their thinking and solution methods with others and to challenge each other's solutions) and make specific suggestions for (e.g., stop another student and ask for help), and by using actual and hypothetical situations to initiate

discussions with the class about students' responsibilities in collaborative work, and to show examples of genuine dialog between students (Yackel *et al.*, 1991).

Modeling Desired Discourse

Teachers also communicate their expectations for students' behavior through their own discourse. When teachers use a recitation approach in which they assume primary responsibility for solving the problem, ask students to provide only answers to discrete steps, and rarely encourage students to verbalize their thinking, groups may adopt an interactional style in which help-givers do most of the work and infrequently monitor other students' level of understanding. In contrast, training teachers to use specific skills to challenge students' thinking encourages students to probe each other's opinions, acknowledge each other's ideas, and attempt to relate new information to ideas previously discussed.

Further Research

While much is known about how and when students learn from their peers, as well as how small-group work might be orchestrated to produce desired learning outcomes, a number of issues have received less attention or are only beginning to be studied systematically. Some of these topics include how to use student characteristics (e.g., ability, gender, cultural or racial background, and personality) when creating group assignments to optimize group functioning for all students; the correspondence between mechanisms that promote academic and nonacademic outcomes (e.g., interpersonal attitudes, liking of the task, and values); the impact of different task structures (e.g., division of labor) on learning outcomes; peer interaction and learning in computer-mediated environments; and how to help teachers modify their teaching practices and classroom environments to facilitate peer learning.

See also: Social Aspects of Collaborative Learning; Social Interaction and Learning.

Bibliography

- Aronson, E., Blaney, N., Stephan, C., Sikes, J., and Snapp, M. (1978). *The Jigsaw Classroom*. Beverly Hills, CA: Sage.
- Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences* 9, 403–436.
- Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In Glaser, R. (ed.) *Advances in Instructional Psychology: Educational Design and Cognitive Science*, pp 161–238. Hillsdale, NJ: Erlbaum.

- Cohen, E. G. and Lotan, R. A. (1995). Producing equal-status interaction in the heterogeneous classroom. *American Educational Research Journal* 32, 99–120.
- Deutsch, M. (1949). An experimental study of the effects of cooperation and competition upon group process. *Human Relations* 2, 199–231.
- Gillies, R. M. and Ashman, A. F. (1996). Teaching collaborative skills in primary school children in classroom-based work groups. *Learning and Instruction* 6, 187–200.
- Johnson, D. W. and Johnson, R. T. (1994). *Learning Together and Alone: Cooperative, Competitive, and Individualistic Learning*, 4th edn. Boston, MA: Allyn and Bacon.
- Johnson, D. W. and Johnson, R. T. (1995). *Creative Controversy: Intellectual Challenge in the Classroom*. Edina, MN: Interaction Book Company.
- King, A. (1992). Facilitating elaborative learning through guided student-generated questioning. *Educational Psychologist* 27, 111–126.
- Nelson-Le Gall, S. (1992). Children's instrumental help-seeking: Its role in the social acquisition and construction of knowledge. In Hertz-Lazarowitz, R. and Miller, N. (eds.) *Interaction in Cooperative Groups: The Theoretical Anatomy of Group Learning*, pp 49–68. New York: Cambridge University Press.
- Newman, R. S. (1998). Students' help seeking during problem solving: Influences of personal and contextual achievement goals. *Journal of Educational Psychology* 90, 644–658.
- O'Donnell, A. M. (1999). Structuring dyadic interaction through scripted cooperation. In O'Donnell, A. M. and King, A. (eds.) *Cognitive Perspectives on Peer Learning*, pp 179–196. Hillsdale, NJ: Erlbaum.
- Palincsar, A. S. and Brown, A. L. (1984). Reciprocal teaching of comprehension fostering and monitoring activities. *Cognition and Instruction* 1, 117–175.
- Piaget, J. (1932). *The Language and Thought of the Child*, 2nd edn. London: Routledge and Kegan Paul.
- Sharan, S. and Hertz-Lazarowitz, R. (1980). A group-investigation method of cooperative learning in the classroom. In Sharan, S., Hare, P., Webb, C. D., and Hertz-Lazarowitz, R. (eds.) *Cooperation in Education*, pp 14–46. Provo, UT: Brigham Young University Press.
- Slavin, R. E. (1992). When and why does cooperative learning increase achievement: Theoretical and empirical perspectives. In Hertz-Lazarowitz, R. and Miller, N. (eds.) *Interaction in Cooperative Groups: The Theoretical Anatomy of Group Learning*, pp 145–173. New York: Cambridge University Press.
- Vygotsky, L. S. (1978). In Cole, M., John-Steiner, V., Scribner, S., and Souberman, E. (eds.) *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Webb, N. M. and Palincsar, A. S. (1996). Group processes in the classroom. In Berliner, D. and Calfee, R. (eds.) *Handbook of Educational Psychology*, pp 841–873. New York: Macmillan.
- Yackel, E., Cobb, P., and Wood, T. (1991). Small-group interactions as a source of learning opportunities in second-grade mathematics. *Journal for Research in Mathematics Education* 22, 390–408.

Further Reading

- Hatano, G. (1993). Time to merge Vygotskian and constructivist conceptions of knowledge acquisition. In Forman, E. A., Minick, N., and Stone, C. A. (eds.) *Contexts for Learning: Sociocultural Dynamics in Children's Development*, pp 153–166. New York: Oxford University Press.
- Hogan, K., Nastasi, B. K., and Pressley, M. (2000). Discourse patterns and collaborative reasoning in peer and teacher-guided discussions. *Cognition and Instruction* 17, 379–432.
- Lou, Y., Abrami, P., and d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research* 71, 449–521.
- Mercer, N. (1996). The quality of talk in children's collaborative activity in the classroom. *Learning and Instruction* 6, 359–377.
- O'Donnell, A. M. and O'Kelly, J. (1994). Learning from peers: Beyond the rhetoric of positive results. *Educational Psychology Review* 6, 321–349.

- Rohrbeck, C., Ginsburg-Block, M., Fantuzzo, J., and Miller, T. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology* **95**, 240–257.
- Salomon, G. and Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research* **13**, 89–99.
- Slavin, R. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology* **21**, 43–69.
- Topping, K. (2005). Trends in peer learning. *Educational Psychology* **25**, 631–645.

Poverty, Effects of on Social and Emotional Development

N Philipsen Hetzner, A D Johnson and J Brooks-Gunn, Columbia University, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Absolute poverty measure – In the United States, it is a threshold determined by a number generated in 1960 that has been carried forward to the present while accounting for increases in cost of living.

Attention – The ability to maintain arousal, disengage and refocus attention, and to resist competing demands for attention.

Cross-sectional study – An observational study that examines a subset of the population at a single point in time.

Delay of gratification – The ability to inhibit impulse behavior and shift attention away from tempting objects to forgo an immediate reward in order to gain a more substantial one later.

Empathy – The ability to experience thoughts and feelings from the perspective of another person.

Environmental chaos – Generally refers to the level of noise, degree of crowding and foot traffic, the routine, predictability, and organization of the home environment.

Exogenous shock or Natural experiments – An observable phenomenon that happens naturally and approximates the properties of a controlled experiment.

Experiments – A method of investigating causal relationships in a controlled environment. Participants are randomly assigned to treatment and control groups that differ only on the variable of interest.

Fixed effects study – Fixed effects methods attempt to control for all stable characteristics of an individual in order to minimize or eliminate selection bias.

Inhibitory control – The ability to control one's behavior by inhibiting responses to irrelevant stimuli while pursuing a goal.

Longitudinal study – An observational study that examines a subset of the population over more than one point in time. If a variable of interest is measured at many points in time, it can be controlled for at an earlier point in order to determine change in the outcome over time.

Relative poverty measure – The extent to which a household's financial resources fall below the

median income of the all the other households in the measurement area.

Selection bias – Spurious associations between variables of interest in a research study often due to unmeasured characteristics.

Introduction: Poverty and Child Development

In 2005, nearly one-fifth of America's children were living in poverty. American individuals and families are considered to be living in poverty if their total household pre-tax income is less than the federal poverty threshold for a family of the same size and composition.

In the United States, the federal poverty threshold is defined in absolute terms based on a number generated in 1960 that has been carried forward accounting for increases in the cost of living. However, in Europe, a measure of relative poverty is used. Relative poverty is the extent to which a household's financial resources fall below the median income of all other households in that measurement area. Regardless of what standard of poverty is used, it is clear that child poverty in the US is alarmingly high. On a measure of relative poverty, the US ranks 24th out of the 25 industrialized countries with respect to child poverty; using an absolute poverty measure, the US occupies the 12th place (see **Figure 1**).

Within the US, a family of three living at 100% of the federal poverty level in 2005 had an annual pre-tax income of \$15 720 (US Census Bureau, 2006a). Those living with incomes below 50% of the federal poverty level experience extreme poverty; in 2005, 7.2% of families with children were living in extreme poverty. Although the number of minority children in the general population is smaller than the number of white children, black and Hispanic children are overrepresented in the poverty population; approximately 25% of black children and nearly 22% of Hispanic children were poor in 2005, compared to 8.3% of white children (US Census Bureau, 2006b).

The prevalence of child and family poverty in the United States has contributed to a flurry of research in recent years around the impact of economic deprivation on the developmental trajectories of children. Not surprisingly, studies have found negative associations between poverty and child development in families who cannot

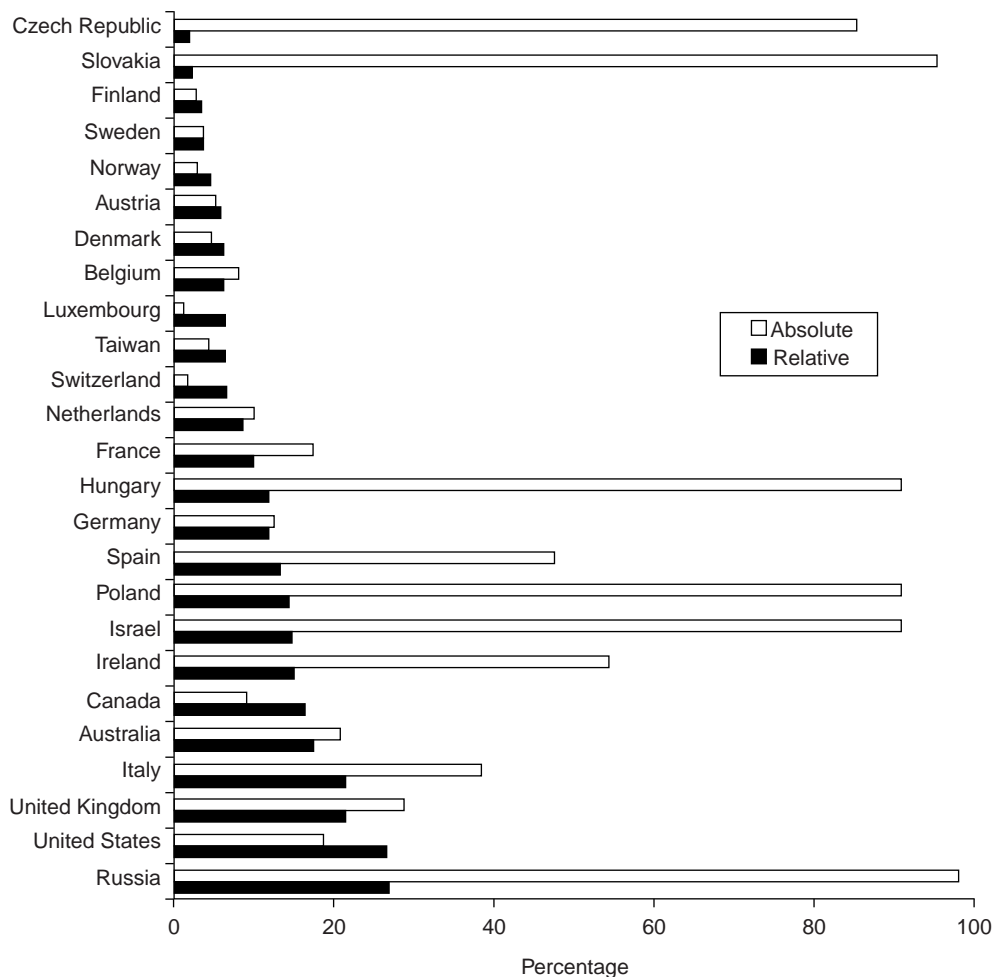


Figure 1 Percentage of children living in poverty in industrialized nations using relative and absolute measures. Relative poverty rate is 50% of overall national median income for survey year. Absolute rate is \$15 299, based on 1995 US official poverty line for a two-parent family with two children. Adapted from Bradbury, B. and Jäntti, M. (1999). Child poverty across industrialised nations. *Innocenti Occasional Papers, Economic and Social Policy Studies*, no. 71. UNICEF International Child Development Centre, Florence, Italy.

afford basic goods and services such as food, clothing, and healthcare (e.g., Aber *et al.*, 1997; Brooks-Gunn and Duncan, 1997; Betson and Michael, 1997; Havemen and Wolfe, 1994). (The majority of poverty research is based on correlational work which largely limits the ability to draw causal conclusions. Therefore, debates occur as to whether income matters at all, or whether income effects can be attributed to selection bias – spurious associations due to some unmeasured characteristics. For this reason, researchers place more credence on some research designs than others; in order from least to most: cross-sectional studies, longitudinal studies where initial outcomes are controlled at an earlier time point so a change in outcomes is analyzed (Duncan *et al.*, 1994), studies looking at change in income and in outcomes (Chase-Lansdale *et al.*, 2003), fixed effects or sibling effects, exogenous shock or natural experiments, and experiments (Fauth and Brooks-Gunn, in press). A number of studies have examined the negative role that poverty plays in shaping children's cognitive skills and school achievement outcomes. In particular, such research

suggests that poor children are more likely to score lower than their middle-class peers on standardized achievement tests (Smith *et al.*, 1997), and more likely to experience developmental delays and disabilities than their nonpoor classmates (Brooks-Gunn and Duncan, 1997). In fact, the achievement gap between low-income children (as young as 2 years) and their middle- and upper-class peers persists and even widens as children grow older (Baydar *et al.*, 1993; Phillips *et al.*, 1998; Rouse *et al.*, 2005).

As links between income and children's cognitive outcomes have generally been more consistent than links between income and children's social and emotional outcomes (Bradley and Corwyn, 2002), and because socio-emotional outcomes tend to be more difficult to measure than cognitive skills, much of the research on poverty and development has focused on school readiness and achievement skills. However, a growing body of literature suggests that poverty also hinders the healthy development of social and emotional skills in children of all ages (e.g., Bradley and Corwyn, 2002; Brooks-Gunn and Duncan, 1997; Evans,

2004; McLoyd, 1998; Duncan *et al.*, 1994; Korenman *et al.*, 1995; McCabe *et al.*, 2004; McLeod and Shanahan, 1993; Sameroff *et al.*, 1987).

This article explores links between poverty and children's social and emotional development. We begin with a brief definition of social and emotional development as it is currently understood in the research literature along with its association with poverty. Next, we examine the moderating role that depth, persistence, and timing of poverty have on the association between income and socio-emotional development and the mechanisms through which poverty and its correlates impinge on early development: parents, the home environment, early education and care settings, and neighborhoods (Romano *et al.*, 2005). Finally, we conclude with a brief discussion of the policy implications that extend from research findings on the impact of poverty on social and emotional development, as well as suggestions for future reading.

Poverty and Child Social and Emotional Development

Developmental psychologists and researchers generally agree on a broad definition of social and emotional development that includes a host of behaviors, typically grouped along two dimensions: internalizing behaviors such as anxiety, withdrawal, and depression (three distinctive behaviors that are often combined when studying children) and externalizing behaviors such as aggression, fighting, and acting out (Brooks-Gunn and Duncan, 1997). In addition, the definition of social and emotional development also includes emotion regulation (including inhibitory control, a child's ability to control one's own behavior, and delay of gratification, a child's ability to inhibit impulsive behavior and shift attention away from tempting objects (Blair, 2002; Cole *et al.*, 2004; Eisenberg, 2001; Eisenberg and Fabes, 1998; Fantuzzo *et al.*, 2005; Li-Grining, 2007; McCabe *et al.*, 2004; Raver, 2004) and attention. Attention refers to a child's ability to maintain arousal, ability to disengage and refocus attention, and the ability to resist competing demands for attention (Posner and Rothbart, 2000).

Data collected on children's social and emotional development are often based on observations of children in structured situations in order to elicit inhibitory control, delay of gratification, emotional knowledge, attention, and empathy. Data may also be collected through parent and teacher reports of children's emotional and behavioral symptoms. Symptoms generally fall on a continuum, allowing researchers to study children who may at some point have mild or moderate levels of psychopathological symptoms in order to determine how they influence particular outcomes. Although less common in large-scale, representative research, classifications of social and emotional diagnosable disorders allow

researchers to study people in terms of groups based on a set of criteria. Some studies have found that the effect of symptoms is as meaningful as psychiatric diagnosis for later consequences of childhood disorders (Jensen *et al.*, 1999).

As the associations between poverty and child outcomes can emerge as early as pregnancy (Wood, 2003), longitudinal studies help us understand links between early environmental conditions and subsequent developmental consequences that emerge over time. On average, children in poverty display more internalizing and externalizing social and emotional developmental problems (Brooks-Gunn *et al.*, 1999; Duncan *et al.*, 1994; McLeod and Shanahan, 1993; McLeod and Shanahan, 1996), as well as decreased emotional regulation and attention (Izard *et al.*, 2008).

Although income is positively correlated with child outcomes, the trend is not linear. Deep poverty is indeed associated with the most negative child outcomes; however, the slope of the line is steeper at the bottom of the income distribution, meaning that one dollar matters more for those children than it does for those higher on the income distribution (Duncan *et al.*, 1998). Exploring how poverty matters differently for social and emotional outcomes based on the depth (or severity) of poverty, the timing of poverty during childhood, and the persistence or length of time poverty is experienced, leads into a discussion of the moderators between poverty and child social and emotional outcomes (Brooks-Gunn and Duncan, 1997).

Depth of Poverty

The finding that deep poverty is especially deleterious for early development is bolstered by evidence from different studies demonstrating that children at the lowest end of the income distribution are most dramatically affected by limited resources (Dearing *et al.*, 2001; Duncan *et al.*, 1998). As mentioned earlier, those in deep poverty are less likely to have access to basic resources and services such as healthy food, safe shelter, medical care, transportation, and stimulating home environments. Children living in deep poverty are likely to encounter a greater number of risk factors than are their less- or nonpoor peers, therefore enduring accumulating risks on development. Research on the National Longitudinal Survey of Youth (NLSY), a national survey spanning from 1986 to 1994 with an overrepresentation of economically disadvantaged respondents, suggests that young children in deep poverty display increased anxiety, withdrawal, and depression compared to less- or nonpoor children, and that the gap between children of different socioeconomic levels widens as children age. For slightly older children (5–8-year-olds), deep poverty was associated with increased aggression, fighting, and acting out (Brooks-Gunn *et al.*, 1999). However, increases in income may also have the greatest positive impact for children living in the deepest poverty. Data from the National

Institute of Child Health and Human Development Study of Early Child Care (NICHD SECC), a longitudinal study designed to follow the childcare experiences of more than 1000 families and their children from birth, reflect that children in poverty who experience a rise in income during the first 3 years of life scored similarly to nonpoor children on behavioral outcome measures. While increases in income seem to have little association with the outcomes of children who were not living in poverty, a small (10%) increase in family income was associated with a decline in negative behavior of children who were poor (Dearing *et al.*, 2001; Taylor *et al.*, 2004), also suggesting that monetary increases are most important for families with the lowest incomes.

Persistence of Poverty

Similarly, the associations of poverty and development vary with the length of time poverty is experienced, such that families in chronic poverty face more prolonged and extensive negative life events and stressors than do families living in transitory poverty (Linver *et al.*, 2002). Children in the NLSY who experienced long-term poverty demonstrated increased internalizing problems (dependence, anxiety, and unhappiness), while those who experienced transient poverty displayed hyperactivity and headstrong behavior (McLeod and Shanahan, 1993). Additionally, the internalizing problems were observed in a follow-up study by the same authors several years later, regardless of further changes in income, suggesting that the negative impact of poverty on social and emotional development may endure over time (McLeod and Shanahan, 1996). The finding that long-term poverty was linked to internalizing problems in children of the NLSY is supported by evidence from another longitudinal study, the Infant Health and Development Program (IHDP; a randomized, multisite early intervention program for premature, low birth-weight children from families of varying socioeconomic levels). IHDP children who experienced enduring poverty demonstrated increased externalizing and internalizing behavior problems (Duncan *et al.*, 1994). Conversely, when families are lifted out of poverty, children may experience improvements in social and emotional outcomes. Income over time was studied using a natural experiment that evolved during a longitudinal study of American Indian and predominately white children living on a reservation in a rural community. Unlike the longitudinal, nonexperimental studies described thus far, this study was able to examine the links between poverty and child outcomes under experimental conditions, therefore ruling out many alternative explanations and spurious associations often characteristic of longitudinal data. After a casino opened on the Indian reservation, the incomes of the American Indian families increased while those of the white families stayed the same. Increased family income was associated with a significant decrease in observed psychiatric symptoms for those children whose

families moved out of poverty as a result of the introduction of the casino. Moreover, children whose families were lifted out of poverty with the additional income demonstrated reduced behavioral problems such that they matched children whose families had never been poor. Children's emotional problems, however, were unaffected by the change in income (Costello *et al.*, 2003).

Timing of Poverty

Research findings on the impact of the timing of poverty on development have been mixed. On the one hand, there is evidence that poverty in early childhood is more harmful to long-term behavioral outcomes than is poverty in later childhood or adolescence (Duncan and Brooks-Gunn, 2000). Some have hypothesized that, because very young children are biologically vulnerable, early deprivation has particularly negative consequences. Evidence suggests that, as a period of rapid and neurological development occurs between the ages of 0 and 3, children who do not have access to rich, stimulating environments during this crucial period are more likely to demonstrate deficits in healthy development (Shonkoff and Phillips, 2000).

On the other hand, there is evidence that implies children who experience poverty later in life have less favorable developmental outcomes than do their younger counterparts. For instance, research suggests that children who experienced poverty in middle childhood (between the ages of 4 and 9) demonstrated increased social and behavioral problems (Hofferth, *et al.*, 2000; NICHD Network, 2005), while a study on the well-being of children whose mothers returned to work after the 1996 welfare reforms found heightened social and emotional problems in the adolescent children of welfare recipients (Gennetian *et al.*, 2004). Therefore, it remains unclear what the impact of timing of poverty may be.

Mechanisms/Pathways

The pathways (or processes) of poverty refer to the mechanisms that explain how poverty works to produce negative child outcomes, and offer points of intervention for policies and programs. Research on these pathways can be organized into four categories: parenting, the home environment, education and care settings, and neighborhoods.

Parenting: The family stress model

Positive child outcomes typically flow from warm parent-child interactions, cognitive stimulation, clear limit-setting, and adequate monitoring (Bornstein, 2002). However, the psychological stress that accompanies poverty, resulting from the disruption of family relationships, unpredictable work schedules, and unstable income, can negatively affect parenting behavior which may subsequently contribute to poor child outcomes. Families that experience either

short-term financial loss or persistent poverty struggle to supply food, shelter, safety, and clothing; these struggles lead to increased levels of depression and anxiety for parents (McLoyd, 1990); notably, linkages between parenting stress and poor parenting practices are more pronounced for families in deep poverty (Petterson and Albers, 2001). In general, poverty has been connected to harsh parental behaviors associated with risk such as physical disciplining practices (Dodge *et al.*, 1994; Linver *et al.*, 2002), decreased attentive and responsive parenting (Dodge *et al.*, 1994; Jackson *et al.*, 2000; Smith *et al.*, 2001), and decreased levels of parental support (Jackson *et al.*, 2000). These parenting behaviors can lead to the deterioration of the parent-child relationship, especially in disadvantaged families (Petterson and Albers, 2001). In sum, the links between financial strain, parental stress, the disruption of family relationships, and ultimately negative child outcomes represent the crux of the family stress model.

Researchers have applied the family stress model to low-income adolescents, school-aged children (McLeod and Nonnemaker, 2000; Sampson and Laub, 1994), and young children (Dodge *et al.*, 1994; Duncan *et al.*, 1994; Jackson *et al.*, 2000; Linver *et al.*, 2002; McLeod and Shanahan, 1993; Yeung *et al.*, 2002), and have found evidence that families in stress can either exacerbate or serve as buffers against the negative child outcomes associated with poverty. For instance, a study by Linver *et al.* (2002) documented that income was associated with behavior problems and that authoritarian and harsh parenting, resulting from maternal emotional distress, accounted for the link between income and behavior problems for 3- and 5-year-olds in the IHDP sample. Similarly, in a study on 93 working mothers (who were poor) and their children, financial strain was found to be associated with increased depressive symptoms which were negatively linked with parenting quality and in turn led to negative child behavioral outcomes (Jackson *et al.*, 2000).

On the other hand, parents who engage in positive parenting despite stress factors due to poverty can buffer the negative association between poverty and child outcomes. In a classic study from the Great Depression, Elder and Caspi (1988) report that emotionally stable fathers were less likely to resort to negative parenting despite dramatic income loss, in turn positively influencing their children's self-esteem. In a second study, findings from parents in a poor rural community who were classified as nurturing and involved suggest that such positive parenting buffered the association between income stress and their children's development (Conger and Conger, 2000). Fortunately, social services, community-based programs, and early intervention programs, including home visiting components, are all potential sources of support, assistance, and information for parents that can improve parenting and, as a result, the child socio-emotional well-being (although results differ as a function of the intervention, with some

programs not finding effects; Barnes *et al.*, 1995; Benasich *et al.*, 1992; Brooks-Gunn *et al.*, 2000; Brooks-Gunn and Markman, 2005). Thus, the family stress model addresses an important pathway through which poverty operates on child social and emotional outcomes, and offers an opportunity for intervention and enhancement through social programming.

The home environment

While the family stress model emphasizes the association between income, parenting, and family relationships, the home environment itself is another pathway through which poverty may impact child development. As mentioned earlier, poverty almost certainly limits a family's ability to purchase material goods and experiences, to spend quality time with children, and to provide a safe and calm living environment that is conducive to healthy social and emotional development.

Research suggests that children from economically impoverished families have limited access to a variety of learning materials, such as books and educational toys (Bradley *et al.*, 2001), but that exposure to print materials and stimulating games may mitigate some of the negative links between poverty and child outcomes (Yeung *et al.*, 2002), as well as provide an opportunity for social exchanges between children and adults (Bradley and Corwyn, 2002).

In addition to having limited access to resources, a study using data from the NLSY found that the physical environments of families in poverty are generally less safe, less clean, darker, and more cluttered than those of nonpoor families (Bradley *et al.*, 2001). Furthermore, the lack of structure and routine, combined with increased environmental chaos that is more prevalent in low-income households, has been shown to have negative associations with child social and emotional development (Evans *et al.*, 2005).

Unfortunately, low-income households often lack important learning materials and resources in the home environment associated with positive child social and emotional outcomes. However, balancing monetary and time-related needs is especially challenging for low-income families because poor parents who work sacrifice time with their children without gaining much buying power from income in exchange (Ryan *et al.*, 2006).

Early childhood education and care

Yet a third pathway through which poverty may impact child development is by way of children's early childhood education and care (ECEC) environments. Specifically, there is promising evidence that quality ECEC settings are linked to improved child outcomes (e.g. Schweinhart *et al.*, 2005). Quality in ECEC refers to aspects of developmental environments recognized by researchers to promote and enhance early learning, including, but not limited to, low child-to-adult ratios, high levels of caregiver

education and qualification, stimulating and responsive interactions between the child and caregiver, and the availability of appropriate toys and learning materials; such aspects of quality in ECEC settings have been linked to improved cognitive and social outcomes for young children (Blau, 2001; Burchinal and Cryer, 2003; Lamb and Ahnert, 2006; NICHD Network, 2000; Shonkoff and Phillips, 2000). For children in poverty, high-quality ECEC may improve school readiness and subsequent chances for school success, financial independence, and social stability (Heckman and Lochner, 2001), and studies have found that the poorest children may benefit the most from high-quality ECEC programs (Brooks-Gunn, 2003; Brooks-Gunn *et al.*, 1993). Unfortunately, not all ECEC environments are of high quality. Recent findings from the Early Childhood Longitudinal Study Birth cohort (ECLS-B), report that 75% of infants and toddlers in center-based care were in care that was of low-to-medium quality, while only 24% were in high-quality centers (Mulligan and Flanagan, 2006). For poor children, low-quality care may add to the multitude of risk factors already present in a child's life.

Despite discouraging evidence regarding the quality of available care, there is reason to believe that when care is of high quality, low-income children experience long-term benefits. As space does not allow for a comprehensive review of the ECEC literature, we highlight findings from random-assignment experimental evaluations of high-quality model ECEC programs. When studies involve random assignment of subjects to treatment conditions (in this case, to receive or not to receive quality ECEC services), the possibility that noted differences after the conclusion of the study are due to the program being tested is higher than it would be if subjects chose, or selected themselves into, the different conditions. Thus, observed differences in performance or outcomes between subjects who were randomly assigned to a given condition or to the control group (children who did not participate in the high-quality ECEC program) are less likely to be due to unobserved, preexisting differences between groups and more likely to be attributed to effects of the program itself.

Beginning in the 1960s, researchers have conducted several experimental evaluations of the so-called model ECEC programs. These programs, which enroll low-income youngsters, typically offer children cognitively stimulating educational curricula directed by qualified staff in classrooms with low child-to-staff ratios, and often include parenting classes, health and nutrition components, and a home-visiting dimension. Low-income children who participated in the Infant Health and Development Program (IHDP) exhibited short-term positive changes in behavior; however, there is evidence that participation in high-quality ECEC programming can contribute to sustained socio-emotional benefits as well. For instance, children who attended Perry Preschool demonstrated fewer conduct and behavioral problems than did control-group children

up to 15 years after the conclusion of the intervention (Schweinhart *et al.*, 1993). Program children continued to experience lower rates of criminal arrest and decreased delinquent behavior through age 40 (Schweinhart *et al.*, 2005). Children who received ECEC services from the Carolina Abecedarian Project were less likely than those who did not to demonstrate the negative social and behavioral outcomes typically associated with childhood poverty, such as teen pregnancy and delinquent behavior, at ages 15 and 21 (Campbell and Ramey, 1994).

The neighborhood

Families, homes, and ECEC settings all exist within communities or neighborhoods and are inextricably linked to the characteristics of that area. However, aside from these elements, other neighborhood characteristics are also linked with child development. Therefore, the neighborhoods in which families reside are another possible pathway of influence through which income operates on child social and emotional outcomes.

The structural and demographic features of neighborhoods and communities are likely to affect child and adolescent outcomes, through community-level social and cultural processes such as community monitoring, the number and quality of social ties, organizational participation, and value consensus. For example, neighbors may serve as role models and exercise social control, helping young people to internalize social norms and learn the boundaries of acceptable behavior (Gephart, 1997; Jencks and Mayer, 1990; Xue *et al.*, 2005). Alternatively, low-income families are often limited in their choice of neighborhoods and find themselves in areas with social disorganization such as crime, unemployed adults, and few resources such as playgrounds and parks (Brooks-Gunn and Duncan, 1997). Overall, research has found that living in areas of concentrated poverty is linked with a negative impact on child and adolescent development (Brooks-Gunn *et al.*, 1997; Jencks and Mayer, 1990; Leventhal and Brooks-Gunn, 2000), as residence in low-income neighborhoods has been associated with increased child behavior problems, anxiety, and depression over and above family poverty (Chase-Lansdale *et al.*, 1997).

The ways that neighborhoods function to impact children in poverty have been studied using two large, longitudinal data sets. The Project on Human Development in Chicago Neighborhoods (PHDCN) is a multi-level, longitudinal study of a representative sample of children aged 5–11 years recruited from 80 neighborhoods and conducted in the late 1990s. In the PHDCN sample, concentrated neighborhood disadvantage was associated with more frequent and more severe mental health problems than other neighborhood socio-demographic characteristics after controlling for family demographic characteristics, maternal depression, and earlier child mental health scores (Xue *et al.*, 2005). This finding is

consistent with prior research on the same sample, linking concentrated disadvantage to neighborhood violence (Sampson, *et al.*, 1997), and suggesting that the social emotional health of children in poverty who reside in neighborhoods of concentrated disadvantage may suffer. However, data from the Moving to Opportunity Program, another randomized study of low-income families living in poor neighborhoods, found mixed results in terms of neighborhood wealth or poverty and social and emotional development (Leventhal and Brooks-Gunn, 2005).

Policy Implications

Childhood is a time of great vulnerability, in part because the developing social and emotional competencies of poor children are especially susceptible to the obstacles that growing up in poverty presents. However, childhood is also a time of great opportunity; the developmental trajectories of children and young adolescents are malleable, and thus these sensitive periods offer windows for intervention. Given what we know about the depth, persistence, and timing of poverty, programs that increase household income even modestly, that intervene to break up what would otherwise be extended cycles of poverty, and that offer social safety nets for children at all stages of development would be beneficial investments for society.

Moreover, research that has illuminated the range of pathways through which poverty presents obstacles to the healthy development of poor children has also highlighted a variety of options for enhancing the lives of poor children. With respect to parenting, the federally funded Early Head Start (EHS) program, which includes parenting classes, parental support groups, and home-visiting components targeted at improving parenting practices, shows some promise regarding more supportive parent-child interaction and increased episodes of parent-child reading compared to families who did not receive EHS services (Love *et al.*, 2005). Likewise, programs that incorporate home-visiting elements may also enhance the home environment of children in poverty.

With respect to ECEC, it is clear that high-quality ECEC programs that begin early, offer a comprehensive set of services, and that adhere to recognized standards of quality have a great deal of potential to boost developmental outcomes of children in poverty. Increased access to subsidies for childcare would increase the purchasing power of low-income families so that they might be able to buy into higher-quality care. In the meantime, continued support for the federal Head Start Program is crucial; begun in the 1960s, the comprehensive, wrap-around, center-based services offered by Head Start include ECEC that is, on average, of higher quality than a majority of other preschool programs (Currie, 2001). Additionally, recent research suggests that Head Start does have the

potential to substantially decrease the school readiness gap that exists between minority children (who tend to be low income), and their white peers (Magnuson and Waldfogel, 2005). However, a concern for those interested in intervening early in the lives of poor children is that only families below the federal poverty line are eligible for childcare subsidies and Head Start enrollment; families who are near-poor and endure many of the same stressors as poor and deep-poor families are nonetheless ineligible for many of the social services that would enhance their children's development. Thus, making ECEC programming universally available is an approach to intervention that can benefit extremely poor as well as near-poor children. Indeed, state-level initiatives, such as those in Oklahoma and Georgia to implement universal prekindergarten to all 4-year-old children, have demonstrated positive impacts on child development (Gormley and Phillips, 2005; Henry *et al.*, 2001). More effort is needed on behalf of policymakers to increase accessibility and affordability of quality care for children too young for programs like Head Start and universal prekindergarten.

Regarding the impact of neighborhoods as a conduit between poverty and child socio-emotional development, community development initiatives and programs that seek to build social capital in communities have great potential for alleviating many of the negative dimensions of poverty prevalent in low-income neighborhoods. For instance, high degrees of collective efficacy and community cohesion are associated with reduced violent crime and disorder; research findings reflect that collective efficacy within a community mediates the association between concentrated disadvantage and violence (Sampson *et al.*, 1997). As the association between both depression and anxiety in children, and exposure to violence, is well established and collective efficacy may protect children from these neighborhood threats (Xue *et al.*, 2005), a viable goal for community interventions and initiatives may be to increase efficacy.

In sum, it is imperative that policymakers attend to the extensive body of research literature emphasizing linkages between poverty and social and emotional development, as well as the range of options for intervening in and enhancing the lives of poor children. One possibility suggested by recent research is to apply an investment model to early childhood, where dollar-for-dollar increases in societal investments in children are linked to equal or greater pay-offs in terms of societal benefits. Indeed, some have argued that viewing social programs that support early childhood development as economic development initiatives may be the most effective way to marshal resources for interventions critical in setting healthy developmental trajectories (Rolnick and Grunewald, 2003; Schweinhart *et al.*, 2005). Regardless of the approach to intervention, policymakers must implement programs that help, rather than hinder, children's development by supporting families and

children. Without appropriate support mechanisms and intervention strategies, it is likely that poor children will continue to lag behind their nonpoor peers.

Bibliography

- Aber, J. L., Bennett, N. G., Conley, D. C., and Li, J. (1997). The effects of poverty on child health and development. *Annual Review of Public Health* **18**, 463–483.
- Barnes, H. V., Goodson, B. D., and Layzer, J. I. (1995). *Review of Research on Supportive Interventions for Children and Families*. Cambridge, MA: Abt Associates.
- Baydar, N., Brooks-Gunn, J., and Furstenberg, F. F. (1993). Early warning signs of functional illiteracy: Predictors in childhood and adolescence. *Child Development* **64**(3), 815–829.
- Benasich, J. S., Brooks-Gunn, J., and Clewell, B. C. (1992). How do mothers benefit from early intervention programs? *Journal of Applied Developmental Psychology* **13**, 311–362.
- Betson, D. M. and Michael, R. T. (1997). Why are so many children poor? *Future of Children* **7**, 25–39.
- Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist* **57**, 111–127.
- Blau, D. M. (ed.) (2001). *The Child Care Problem: An Economic Analysis*. New York: The Russell Sage Foundation Press.
- Bornstien, M. H. (2002). *Handbook of Parenting*. Mahwah, NJ: Erlbaum.
- Bradbury, B. and Järntti, M. (1999). Child poverty across industrialised nations. *Innocenti Occasional Papers, Economic and Social Policy Studies*, no. 71. UNICEF International Child Development Centre, Florence, Italy.
- Bradley, R. H. and Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology* **53**, 371–399.
- Bradley, R. H., Corwyn, R. F., Burchinal, M., Pipes-McAdoo, H., and Coll, C. G. (2001). The home environments of children in the United States Part II: Relations with behavioral development through age thirteen. *Child Development* **72**, 1868–1886.
- Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report* **17**(1), 1–14.
- Brooks-Gunn, J. and Duncan, G. J. (1997). The effects of poverty on children. *Future of Children* **7**, 55–71.
- Brooks-Gunn, J. and Markman, L. B. (2005). The contribution of parenting to ethnic and racial gaps in school readiness. *Future of Children* **15**, 139–168.
- Brooks-Gunn, J., Klebanov, P., Liaw, F., and Spiker, D. (1993). Enhancing the development of low-birthweight, premature infants: Changes in cognition and behavior over the first three years. *Child Development* **64**, 736–753.
- Brooks-Gunn, J., Greg, J. D., and Aber, J. L. (eds.) (1997). *Neighborhood Poverty: Contexts and Consequences for Children* vol. 1. New York: Russell Sage Foundation.
- Brooks-Gunn, J., Leventhal, T., and Duncan, G. J. (1999). Why poverty matters for young children: Implications for policy. In Osofsky, J. D. and Fitzgerald, H. E. (eds.) *WAIMH Handbook of Infant Mental Health*. Vol. 3: *Parenting and Child Care*, pp 92–131. New York: Wiley.
- Brooks-Gunn, J., Berlin, L. J., and Fuligni, A. S. (2000). Early childhood intervention programs: What about the family? In Shonkoff, J. P. and Meisels, S. J. (eds.) *Handbook of Early Childhood Intervention*, 2nd edn., pp 549–587. New York: Cambridge University Press.
- Burchinal, M. R. and Cryer, D. (2003). Diversity, child care quality, and developmental outcomes. *Early Childhood Research Quarterly* **18**, 401–426.
- Campbell, F. A. and Ramey, C. T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development* **65**, 684–698.
- Chase-Lansdale, P. L., Gordon, R. A., Brooks-Gunn, J., and Klebanov, P. K. (1997). Neighborhood and family influences on the intellectual and behavioral competence of preschool and early school-age children. In Brooks-Gunn, J., Duncan, G. J., and Aber, J. L. (eds.) *Neighborhood Poverty: Context and Consequences for Children*, vol. 1, pp 79–118. New York, NY: Russell Sage.
- Chase-Lansdale, P. L., Moffitt, R. A., Lohman, B. J., et al. (2003). Mothers' transitions from welfare to work and the well-being of preschoolers and adolescents. *Science* **299**, 1548–1552.
- Cole, P. M., Martin, S. E., and Dennis, T. A. (2004). Emotion regulation as a scientific construct: Methodological challenges and directions for child development research. *Child Development* **75**, 317–333.
- Conger, R. and Conger, K. J. (2000). Resilience in midwestern families: Selected findings from the first decade of prospective longitudinal study. *Journal of Marriage and the Family* **64**, 361–373.
- Costello, E. J., Compton, S. N., Keeler, G., and Angold, A. (2003). Relationships between poverty and psychopathology: A natural experiment. *Journal of the American Medical Association* **290**, 2023–2029.
- Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives* **15**, 213–238.
- Dearing, E., McCartney, K., and Taylor, B. A. (2001). Change in family income-to-needs matters more for children with less. *Child Development* **72**, 1779–1793.
- Dodge, K. A., Pettit, G. S., and Bates, J. E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development* **65**, 649–665.
- Duncan, G. J. and Brooks-Gunn, J. (2000). Family poverty, welfare reform, and child development. *Child Development* **71**, 188–196.
- Duncan, G. J., Brooks-Gunn, J., and Klebanov, P. K. (1994). Economic deprivation and early childhood development. *Child Development* **65**, 296–318.
- Duncan, G. J., Yeung, W., Brooks-Gunn, J., and Smith, J. R. (1998). How much does childhood poverty affect the life chances of children? *American Sociological Review* **63**, 406–423.
- Eisenberg, N. (2001). The core and correlates of affective social competence. *Social Development* **10**, 120–124.
- Eisenberg, N. and Fabes, R. A. (1998). Prosocial development. In Damon, W. and Eisenberg, N. (eds.) *Handbook of Child Psychology: Social, Emotional, and Personality Development*, vol. 3, pp 701–767. NY: Wiley.
- Elder, G. H. (1999). *Children of the great depression: Social change in life experience*. Boulder, CO: Westview Press.
- Evans, G. W. (2004). The environment of childhood poverty. *American Psychologist* **59**, 77–92.
- Evans, G. W., Gonnella, C., Marcynyszyn, L. A., Gentile, L., and Salepkar, N. (2005). The role of chaos in poverty and children's socioemotional adjustment. *Psychological Science* **16**(7), 560–565.
- Fantuzzo, J. W., Bulotsky-Shearer, R., Fusco, R. A., and McWayne, C. (2005). An investigation of preschool classroom behavioral adjustment problems and social-emotional school readiness competencies. *Early Childhood Research Quarterly* **20**, 259–275.
- Fauth, R. C. and Brooks-Gunn, J. (2008). Are some neighborhoods better for child health than others? In Schoeni, R. F., House, J. S., Kaplan, G. A., and Pollack, H. (eds.) *Making Americans healthier: Social and economic policy as health policy*, pp. 334–376. New York: Russell Sage.
- Gennetian, L., Duncan, G., Knox, V., et al. (2004). How welfare policies affect adolescents' school outcomes: A synthesis of evidence from experimental studies. *Journal of Research on Adolescence* **14**, 399–423.
- Gephart, M. A. (1997). Neighborhoods and Communities as Contexts for Development. In Brooks-Gunn, J., Duncan, G. J. M., and Aber, J. L. (eds.) *Neighborhood Poverty: Context and Consequences for Children*, pp 1–43. New York: Russell Sage Foundation.
- Gormley, W. T. and Phillips, D. (2005). The effects of universal pre-k in Oklahoma: Research highlights and policy implications. *Policy Studies Journal* **33**(1), 65–82.
- Haveman, R. and Wolfe, B. (1994) *Succeeding Generations: On the Effects of Investment in Children*. New York: Russell Sage Foundation.
- Heckman, J. J. and Lochner, L. (2001). Rethinking education and training policy: Understanding the sources of skill formation in a modern economy. In Danziger, S. and Waldfogel, J. (eds.) *Securing*

- the Future: Investing in Children from Birth to College*, pp 47–83. New York: Russell Sage Foundation.
- Henry, G. T., Gordon, C. S., Mashburn, A., and Ponder, B. D. (2001). *Pre-K Longitudinal Study: Findings from the 1999–2000 School Year*. Applied Research Center, Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA, USA.
- Hofferth, S. L., Smith, J., McLoyd, V. C., and Finkelstein, J. (2000). *Achievement and Behavior among Children of Welfare Recipients, Welfare Leavers, and Low Income Single Mothers*. Institute for Social Research, University of Michigan, Ann Arbor, MI, USA.
- Izard, C. E., King, K. A., Trentacosta, C. J., and Morgan, J. K. (2008). Accelerating the development of emotion competence in Head Start children: Effects on adaptive and maladaptive behavior. *Development and Psychopathology* **20**(1), 369–397.
- Korenman, S., Miller, J. E., and Sjaastad, J. E. (1995). Long-term poverty and child development in the United States: Results from the NLSY. *Children and Youth Services Review* **17**, 127–155.
- Jackson, A. P., Brooks-Gunn, J., Huang, C.-C., and Glassman, M. (2000). Single mothers in low-wage jobs: Financial strain, parenting, and preschoolers' outcomes. *Child Development* **71**, 1409–1423.
- Jencks, C. and Mayer, S. (1990). The social consequences of growing up in a poor neighborhood. In Lynn, L. and McGreary, M. (eds.) *Inner-City Poverty in the United States*, pp 111–186. Washington, DC: National Academy Press.
- Jensen, P. S., Brooks-Gunn, J., and Graber, J. A. (1999). Dimensional scales and diagnostic categories: Constructing crosswalks for child psychopathology assessments. *Journal of the American Academy of Adolescent and Child Psychiatry* **38**, 118–120.
- Lamb, M. E. and Ahnert, L. (2006). Nonparental child care: Context, concepts, correlates, and consequences. In Damon, W., Lerner, R. M., Renninger, K. A., and Sigel, I. E. (eds.) *Handbook of Child Psychology, Vol. 4. Child psychology in practice*, 6th edn., pp 950–1016. New York: Wiley.
- Leventhal, T. and Brooks-Gunn, J. (2000). The neighborhoods they live in: The effects of neighborhood residence upon child and adolescent outcomes. *Psychological Bulletin* **126**, 309–337.
- Leventhal, T. and Brooks-Gunn, J. (2005). Neighborhood and gender effects on family processes: Results from the moving to opportunity program. *Family Relations* **54**, 633–644.
- Li-Grining, C. P. (2007). Effortful control among low-income preschoolers in three cities: Stability change and individual differences. *Developmental Psychology* **43**, 208–221.
- Linver, M. R., Brooks-Gunn, J., and Kohen, D. E. (2002). Family processes as pathways from income to young children's development. *Developmental Psychology* **38**, 719–734.
- Love, J., Kisker, E. E., Ross, C., et al. (2005). The effectiveness of early head start for 3-year-old children and their parents. *Developmental Psychology* **41**(6), 885–901.
- Magnuson, K. A. and Waldfogel, J. (2005). Early childhood care and education: Effects on racial and ethnic gaps in school readiness. *Future of Children* **15**(1), 169–196.
- McCabe, L. A., Cunningham, M., and Brooks-Gunn, J. (2004). The development of self-regulation in young children: Individual characteristics and environmental contexts. In Baumeister, R. and Vohs, K. (eds.) *Handbook of Self-Regulation: Research, Theory, and Applications*, pp 340–356. New York, NY: Guilford Press.
- McLeod, J. D. and Nonnemaker, J. M. (2000). Poverty and child emotional and behavioral problems: Racial/ethnic differences in processes and effects. *Journal of Health and Social Behavior* **41**, 137–161.
- McLeod, J. D. and Shanahan, M. J. (1993). Poverty, parenting, and children's mental health. *American Sociological Review* **58**, 351–366.
- McLeod, J. D. and Shanahan, M. J. (1996). Trajectories of poverty and children's mental health. *Journal of Health and Social Behavior* **37**, 207–220.
- McLoyd, V. C. (1990). The impact of economic hardship on black families and children: Psychological distress, parenting, and socioemotional development. *Child Development* **61**, 311–346.
- McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist* **53**, 185–204.
- Mulligan, G. M. and Flanagan, K. D. (2006). *Age 2: Findings from the 2-Year-Old Follow-Up of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)*. Washington, DC: National Center for Education Statistics.
- NICHD Early Child Care Research Network. (2000). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Science* **4**(3), 116–135.
- NICHD Early Child Care Research Network. (2005). Duration and developmental timing of poverty and children's cognitive and social development from birth through third grade. *Child Development* **16**, 795–810.
- Petterson, S. M. and Albers, A. B. (2001). Effects of poverty and maternal depression on early child development. *Child Development* **72**, 1794–1813.
- Phillips, M., Crouse, J., and Ralph, J. (1998). Does the Black-White test score gap widen after children enter school? In Jencks, C. and Phillips, M. (eds.) *The Black-White test Score Gap*, pp 229–272. Washington, DC: The Brookings Institution.
- Posner, M. and Rothbart, M. (2000). Developing mechanisms of self-regulation. *Development and Psychopathology* **12**, 427–441.
- Raver, C. C. (2004). Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Development* **75**, 346–353.
- Rolnick, A. and Grunewald, R. (2003). Early childhood development: Economic development with a high public return. *Region* **17**, 6–12.
- Romano, E., Tremblay, R., Boulerice, B., and Swisher, R. (2005). Multi-level correlates of childhood physical aggression and prosocial behavior. *Journal of Abnormal Child Psychology* **33**, 565–578.
- Rouse, C., Brooks-Gunn, J., and McLanahan, S. (2005). Introducing the issue. *Future of Children* **15**, 5–13.
- Ryan, R. M., Fauth, R. C., and Brooks-Gunn, J. (2006). Childhood poverty: Implications for school readiness and early childhood education. In Spodek, B. and Saracho, O. N. (eds.) *Handbook of Research on the Education of Young Children*, pp 3–10. Mahwah, NJ: Erlbaum.
- Sameroff, A. J., Seifer, R., Barocas, B., Zax, M., and Greenspan, S. (1987). IQ scores of 4-year old children: Social environmental risk factors. *Pediatrics* **79**, 343–350.
- Sampson, R. J. and Laub, J. H. (1994). Urban poverty and the family context of delinquency: A new look at structure and process in a classic study. *Child Development* **65**, 523–540.
- Sampson, R. J., Raudenbush, S. W., and Earls, F. (1997). Neighborhoods and violent crime: A multi-level study of collective efficacy. *Science* **277**, 918–924.
- Schweinhart, L. J., Barnes, H. V., and Weikert, D. P. (1993). *Significant Benefits: The High/Scope Perry Preschool Study through Age 27*. Ypsilanti, MI: High/Scope Press.
- Schweinhart, L. J., Montie, J., Xiang, Z., et al. (2005). *Lifetime Effects: The High/Scope Perry Preschool Study through Age 40*. Ypsilanti, MI: High/Scope Press.
- Shonkoff, J. and Phillips, D. A. (eds.) (2000). *From Neurons to Neighborhoods*. Washington, DC: National Academy of Sciences.
- Smith, J. R., Brooks-Gunn, J., and Klebanov, P. K. (1997). The consequences of living in poverty for young children's cognitive and verbal ability and early school achievement. In Duncan, G. J. and Brooks-Gunn, J. (eds.) *Consequences of Growing up Poor*, pp 132–189. New York: Russell Sage Foundation Press.
- Smith, J. R., Brooks-Gunn, J., Kohen, D., and McCarton, C. (2001). Transitions on and off AFDC: Implications for parenting and children's cognitive development. *Child Development* **72**, 1512–1533.
- Taylor, B. A., Dearing, E., and McCartney, K. (2004). Incomes and outcomes in early childhood. *Journal of Human Resources* **39**, 980–1007.
- US Census Bureau (2006a). Families by age of householder, number of children, and family structure: 2005. http://pubdb3.census.gov/macro/032006/pov/new04_100_01.htm (accessed April 2009).

- US Census Bureau (2006b). Income, poverty, and health insurance coverage in the United States: 2005. <http://www.census.gov/prod/2006pubs/p60-231.pdf> (accessed April 2009).
- Wood, D. (2003). Effect of child and family poverty on health in the United States. *Pediatrics* **112**, 707–711.
- Yeung, W., Linver, M. R., and Brooks-Gunn, J. (2002). How money matters for young children's development: Parental investment and family processes. *Child Development* **73**, 1861–1879.
- Xue, Y., Leventhal, T., Brooks-Gunn, J., and Earls, F. J. (2005). Neighborhood residence and mental health problems of 5- to 11-year-olds. *Archives of General Psychiatry* **62**, 554–563.

Further Reading

- Case, A., Lubotsky, D., and Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *American Economic Review* **92**, 1308–1334.
- Leventhal, T., Fauth, R., and Brooks-Gunn, J. (2005). Neighborhood poverty and public policy: A five year follow up of children's educational outcomes in the New York city moving to opportunity demonstration. *Developmental Psychology* **41**, 933–952.

Seeking Help as an Adaptive Response to Learning Difficulties: Person, Situation, and Developmental Influences

S A Karabenick, University of Michigan, Ann Arbor, MI, USA

R S Newman, University of California – Riverside, Riverside, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Adaptive help seeking – Help seeking which is self-regulative, strategic, and ‘smart’.

Autonomous help seeking – See also adaptive help seeking.

Excessive help seeking – See also expedient help seeking.

Executive help seeking – See also expedient help seeking.

Expedient help seeking – Help seeking which is effort-avoidant and unnecessary.

Instrumental help seeking – See also adaptive help seeking.

Mastery-oriented students – Students motivated by a desire to learn, understand, and master difficult tasks.

Performance-avoid goals – Desire to avoid an appearance of incompetencies.

Social affiliation goals – Desire for friendship and intimacy.

Social status goals – Desire for peer approval and popularity.

Help Seeking and Learning: An Evolving Perspective

Students experience and deal with difficult academic tasks in diverse ways. Some exert little effort, sit passively, and give up prematurely, whereas others respond to failure with extra resolve and resourcefulness. One manifestation of such resolve involves seeking help. Indeed, recent research and theoretical advances indicate that help seeking can be an important form of behavioral self-regulation (Pintrich and Zusho, 2002), part of the tool kit of cognitively, behaviorally, and emotionally engaged learners (e.g., Butler, 1998; Karabenick, 2003, 2004; Karabenick and Knapp, 1991; Karabenick and Newman, 2006; Nelson-Le Gall and Resnick, 1998; Newman, 2000; Skinner and Zimmer-Gembeck, 2007; Zimmerman and Martinez-Pons, 1990; Zusho *et al.*, 2007). Seeking help may contribute to a general pattern of student resilience in overcoming obstacles to learning and achievement (see Newman, 2000, 2002).

This characterization of help seeking challenges the belief that it should be discouraged out of concern that it perpetuates dependency. Continued dependence on others could be a consequence of less adaptive help seeking (also referred to as expedient, excessive, or executive), which is effort-avoidant and unnecessary (Nelson-Le Gall, 1981, 1987). More adaptive help seeking (also referred to as instrumental, strategic, or autonomous; Butler, 1998; Karabenick, 1998; Nelson-Le Gall, 1981; Newman, 2000), however, can have just the opposite effect by increasing students’ capacity to surmount learning difficulties when encountered subsequently. Just as seeking help that is work avoidant or excessive may not be in students’ long-term best interests, avoiding seeking help when that help is truly necessary also can be nonadaptive (see Marchand and Skinner, 2007; Newman, 2007). The distinction between more and less adaptive forms (or goals) of help seeking must be considered as we examine the person characteristics and features of the learning context that influence its effective use.

Toward that end, we begin with a normative, or ideal, description of learners who employ adaptive help seeking. Such learners carefully monitor how they are doing academically, are attuned to difficulties they cannot overcome independently, and have the skills, motivation, and emotions to address those difficulties by approaching a potential helper and requesting what is needed. Resources required for adaptive help seeking include: (1) cognitive competencies (e.g., knowing when help is necessary and knowing how to formulate linguistically a specific question that yields exactly what is needed); (2) social competencies (e.g., knowing which instructors and classmates are more knowledgeable and can potentially help them and communication skills, such as how and when to approach helpers and how and when to thank them); (3) affective–motivational resources (e.g., academic and social goals, self-beliefs, and emotions that allow the student to tolerate difficulty and uncertainty and the ego strength required to withstand possibly negative perceptions in the eyes of classmates); and (4) contextual and interpersonal resources (i.e., classroom and home affordances such as teachers’ goals, grading system, collaborative activities, rules of student–teacher engagement, and teachers’ and parents’ expectations for the student) that support students’ cognitive and social competencies and affective–motivational resources.

Person and Contextual Influences

As in other areas of self-regulation, help seeking has been extensively studied within the framework of achievement goal theory, which differentiates between mastery-focused and performance-focused approaches to learning (Arbreton, 1998; Butler, 1998; Butler and Neuman, 1995; Karabenick, 2003, 2004; Newman, 2002; Pintrich, 2000; Ryan *et al.*, 1997; Ryan and Pintrich, 1997, 1998). At the individual level, studies have consistently shown that instrumental/adaptive help seeking is more likely among mastery-oriented students, who tend to construe seeking help within the broader goal of understanding and self-improvement (i.e., developing competence). Theirs is a learning orientation, which lends itself either to seeking help adaptively, or to working independently when that would be more effective (Butler, 1998). Adaptive help seeking is less likely, however, for students who are concerned about appearing incompetent (i.e., a performance-avoid orientation). If such students do seek help, it is often in the form of expedient or executive help seeking (e.g., Karabenick, 2004; Karabenick *et al.*, 2005).

Students' achievement goals at any point in time are a function both of past experiences and features of the contemporaneous learning context. Achievement goal structure refers to how students construe their classrooms and courses of study in terms of the contextual emphasis on mastery and/or performance goals (Ames and Archer, 1988; Midgley, 2002). Studies using hierarchical modeling consistently have found that students' perceptions of their classes' achievement goal structure influence their tendencies to seek or to avoid seeking help when needed (Church *et al.*, 2001; Midgley, 2002; Urdan *et al.*, 2002). Elementary school classes that students collectively judge as more focused on mastery are less likely to avoid seeking needed help (Turner *et al.*, 2002). Although young children are concerned about not appearing incompetent by asking for help, not until middle school do such concerns influence whether they will ask (Newman, 2000). Presumably a consequence of increased evaluation pressures that begin with the transition to middle school (Eccles and Midgley, 1989; Ryan and Pintrich, 1997, 1998), performance goal-related classroom characteristics, in addition to perceived classroom mastery goals, affect middle school students' tendencies to seek or to avoid seeking help (Karabenick *et al.*, 2005; Newman, 2002; Ryan *et al.*, 1998). By the time students are in college, available evidence suggests that classroom mastery goals are not as relevant to help seeking; rather, students in classes they perceive to be focused on avoiding demonstrations of incompetence (performance-avoid goals) are less likely to seek needed help, or they seek expedient help (Karabenick, 2004).

Cultural Influences

Culture can influence whether and in what situations students seek help, in particular, the degree of stress on individualism versus collectivism (Triandis, 1994). Learners in the US especially are socialized to idealize individualism and deplore dependency (Fischer and Torney, 1976; Sears *et al.*, 1957), which adds to the threat posed by help seeking. Individualistic values were codified in early theories of achievement motivation in which seeking help was considered incompatible with an achievement motive (Beller, 1957; Winterbottom, 1958). Learners in collectivist societies presumably are not as subject to the same prohibitions and should accordingly be less reluctant to seek help. This prediction was verified in that Israeli students raised collectively on kibbutzim were more likely to seek help than those socialized in individualistic-oriented cities (Nadler, 1998). Such cultural influences extend to learning and performance in the workplace as well, as evidence indicates that collectivistic (vs. individualistic) norms facilitate help seeking due to the perceived safety that results from collectivist organizational norms (Sandoval and Lee, 2006).

When examining the effects of culture on help seeking, however, it is important to avoid essentialist generalizations. This means taking into consideration characteristics of tasks, specific learning contexts, and whether the help is sought in public (as in classrooms) or privately (after class or in faculty offices) (Karabenick and Knapp, 1988a). For example, the cultural differences found by Nadler depended on whether students worked on tasks individually or in groups. Japanese college students' collectivistic acculturation, which stresses cooperation, dependency, and empathy, does facilitate seeking assistance from peers outside of the classroom. However, due to culturally induced deference to authority in the form of relationships with instructors, students are hesitant to ask their instructors questions in class (Shwalb and Sukemune, 1998). The same deference to authority occurs in other collectivist societies such as Malaysia (Hashim *et al.*, 2003). Situation influences have also been demonstrated among US and Australian college students whose intentions to seek help from peers decreased as the cultural difference between them and their peers increased (Volet and Karabenick, 2006). The more students were culturally unlike other students, the less likely they would approach them for needed assistance with their studies. The effect of cultural difference between those in need of assistance and help providers was moderated, however, by students' perceived support from their teachers for intergroup acceptance: cultural differences were minimized to the extent their teachers supported intercultural interactions among students in their classes.

Fostering Adaptive Help Seeking in the Classroom

Teacher Influences

Help seeking in the classroom is a social transaction (Newman, 1998a). Teachers establish – and students internalize – patterns of discourse in the classroom. Teachers who respond to requests for help with hints and contingent instruction (vs. direct and controlling answers) are likely to have students who not only accomplish difficult tasks but, in addition, learn that questioning is an invaluable academic strategy. In contrast, teachers who take on the role of expert (e.g., who present to the class an explanation without discussion and then expect students simply to practice) are likely to support overly dependent executive/expedient help seeking. When teachers personally demonstrate that uncertainty can be tolerated – and perhaps even transformed into intellectual challenge – students are likely to realize it is normal not to be able to solve all problems independently (McCaslin and Good, 1996). It is expected that, when teachers scaffold learning experiences and socialize the normalcy of academic difficulty, need for collaboration, and expectation of answers to their questions, students internalize a personal sense of empowerment and voice (Nelson-Le Gall and Resnick, 1998).

Ideally, students learn the value, usefulness, and skills of questioning that are important for monitoring, diagnosing, and fixing misconceptions. The frequency with which teachers call on students, the amount of time they wait for a response, and the amount and type of praise they give vary from student to student (Eccles and Wigfield, 1985). Teacher feedback helps students know when they need help. Giving no more assistance than is necessary may help students learn the difference between adaptive and nonadaptive (i.e., expedient) help seeking. Encouraging students to go back to an incorrect problem and try to re-solve it may convince them of the importance in determining if they need further assistance. Further, it may be instrumental in students' coming to appreciate the function of questioning and help seeking in the ongoing process of self-monitoring and learning.

As noted above, when both classroom and personal goals emphasize learning and developing competence, students are especially likely to seek help adaptively, whereas when both types of goals emphasize performance, students are reluctant to do so. When students who are concerned about grades and looking smart are placed in a learning-goal classroom, they may tend to overcome – and compensate for – their personal tendencies to avoid help. Thus, by being attuned to individual students' personal goals, teachers can assist those who otherwise might give up in the face of adversity (Newman, 1998b). Teachers can also try to accommodate students' social goals

(e.g., social affiliation, social status) that influence help seeking (Ryan *et al.*, 1997). The task of goal-coordinator is not easy, as multiple personal (i.e., both achievement-related and social) goals and multiple classroom goals can complement or conflict with one another. Things become even more complex when one considers that responsive teachers try to support student autonomy while at the same time satisfying their own personal (i.e., both achievement-related and social) goals and need for autonomy within the constraints of public school settings (Butler, 2006). In classrooms in which teachers share with children their time, energy, and nurturance, students tend to be attentive, effortful, self-expressive, and interested in learning.

Teachers who are interpersonally involved with students and attuned to the student's purpose typically establish classrooms that facilitate adaptive help seeking. When teachers and students are on the same page, teachers are especially able to take the student's perspective and understand his or her thinking (e.g., regarding a particular academic task) and, based on this understanding, appropriately guide the student's learning. Teachers who are perceived as friendly and caring tend to demonstrate democratic interaction styles, with lines of communication open to students; they listen, ask questions, inquire if students need help, make sure students understand difficult material, and provide help in a nonthreatening way (Wentzel, 1997). When they experience this type of communication, students learn that teachers are trustworthy helpers. Low achievers, who often have poor self-perceptions of ability and low self-esteem, typically are reluctant to seek academic help in class (Karabenick and Knapp, 1988b, 1991). For these students, especially, teachers who believe their responsibility is to attend to students' academic as well as social and emotional needs can counter student disengagement (Ryan *et al.*, 1998).

Teacher involvement forms the basis of students' beliefs and feelings about the benefits – and costs – of help seeking. Early-elementary-aged students generally feel comfortable approaching their teacher for assistance because of global, affective traits of the teacher (e.g., niceness and kindness). By the middle of elementary school, students tend to view teachers as helpful when they show an awareness of their problems and give them advice, time, energy, and encouragement to ask questions in class (Newman and Schwager, 1993). In classes where teachers are perceived as supporting collaboration, student questioning, teacher fairness, respect, and caring, middle and high school students are especially likely to seek adaptive help and not avoid seeking help (Karabenick *et al.*, 2005). Their approaches to teaching, in terms of their own achievement goals, can also influence how supportive they are perceived by students (Butler, 2007). Perceived support for student questioning facilitates help seeking, in particular, at the

college level (Karabenick and Sharma, 1994). However, as early as grade 2, students often are fearful of teachers' negative reactions (e.g., I think she might think I'm dumb if I ask for help) if they ask for help (Newman and Goldin, 1990). Perceived costs are heightened when teachers are unwilling to help (e.g., if you had paid attention, you wouldn't need to ask that question). Children weight the relative benefits and costs of help seeking, with the integration process becoming increasingly complex over the school years, whereas older students increasingly struggle in deciding what to do when they need academic assistance (Newman, 1990).

The Influence of Peers: Social Relations and Comparison Processes

Generally, classroom goals and task structure determine how students influence one another's competence and self-perceptions of competence required for adaptive help seeking. In some learning contexts, students are allowed to collaborate – asking and answering one another's questions – but in other contexts, they are not. In contrast to individual classroom activity (where teachers usually expect students not to need assistance) and whole-class activity (where questions generally flow in the direction of teacher-to-student rather than student-to-teacher), small-group activity generally reduces social comparison and promotes peer collaboration. During small-group collaboration, children can turn to one another when they need assistance.

With experience working together, children gradually become better at asking each other good questions in good ways. Adaptive help seeking requires that students skillfully request help from one another, for example, checking to see if the potential helper is paying attention to them before actually making a request. Ideally, students make requests that are direct, sincere, polite, and clear about what exactly is being requested (Cooper *et al.*, 1982). Requests are often revised and clarified if they are initially unsuccessful in obtaining a response. When students make vague requests but persist by reformulating and clarifying the requests, academic performance tends to improve. Also, when they ask for – and receive – elaborated help (e.g., explanations rather than direct answers) and when they then use the help in a constructive way, children are most likely to learn (Webb and Palincsar, 1996).

Especially at upper-elementary and middle school, collaborative activity provides students a chance to think in public and exchange with one another their thoughts. Built into many collaborative activities are opportunities for students to ask – and be asked – questions for purposes of monitoring their own and others' understanding and for requesting clarification, justification, and elaboration of other students' ideas. Questions potentially allow an exchange of perspectives among individuals who are

working on relatively equal footing. As students observe the effectiveness of peers' questions in resolving difficulties, they are likely to learn that different individuals contribute unique skills and knowledge. They may learn how, in the future, to choose helpers according to both their own needs and others' competencies (Webb *et al.*, 2006).

Although social comparison can be detrimental if that process facilitates performance-avoid goals, comparing one's performance with that of others can have a positive influence on help seeking. Social comparison offers information about others' strengths and weaknesses and thus helps children evaluate peers' capacity to be effective helpers (Ruble and Frey, 1991). It helps children make realistic judgments about whether they personally have tried enough before turning to others. With development, children are increasingly able to judge when assistance is truly necessary so they can request the right amount of help – not too little and not too much.

Typically there are concomitant changes in how children conceptualize ability; students come to believe that smart children who do not have to try very hard and dumb children who have to work extra hard can get a similar grade on an assignment. Thus, students come to perceive children who need help as not very smart (Nicholls and Miller, 1984). By publicly seeking help, adolescents put their self-worth at risk (Covington, 1992). Social comparison strongly influences one's achievement motivation, for it affects why one works hard, or cheats, or gives up: Is it to satisfy one's personal desire to learn, or to prevent others from seeing a weakness, or to impress classmates and parents (see self-determination theory; Ryan and Deci, 2000). By managing positive and negative aspects of social comparison (e.g., according to grading practices, types of classroom activity, goal structures), classroom teachers can potentially affect the extent to which students influence one another's sense of autonomy needed for adaptive help seeking (Newman, 2000).

Friends, by definition, assist and support one another (Berndt and Keefe, 1996). Quality friendships are characterized by mutual support as well as certain features (e.g., reliability, affection, intimacy, and lack of conflict and rivalry) that tend to mediate efforts to seek help. In close relationships, children are relatively unconcerned about self-disclosure, threat to self-esteem, and indebtedness to those who help them. In a friendly context, children find it easier to manage and negotiate social demands of interactions and focus their shared efforts on learning and problem solving. In contrast, children in conflictual relationships typically are reluctant to disclose difficulties to one another and probably would not expect help to be forthcoming even if they requested it.

Social goals are related to academic help seeking. The more strongly they strive for goals of social affiliation (i.e., desire for friendship and intimacy), the more students value and use help seeking as a strategy for dealing with academic difficulties (Ryan *et al.*, 1997). These goals

however do not guarantee that help seeking is adaptive. Children may appear to work together and request help from one another but really just goof off. Indeed, requests among friends are sometimes socially inappropriate (e.g., shouting questions across the room) and cognitively inappropriate (e.g., requesting unnecessary help). The more strongly students strive for goals of social status (i.e., desire for peer approval and popularity), the more they are embarrassed to ask for help in the classroom (Ryan *et al.*, 1997). At the transition to middle-school, when students typically are concerned about their self-image in addition to the increase in evaluation pressures, social status goals are likely to inhibit help seeking, especially if an individual's self-esteem is easily threatened and his or her peer group does not value academic success. Importantly, however, inhibition is likely to be minimized if the student has a strong sense of self and a peer group that does value learning. Such influences may also play a role when students transition to college. If nothing else, their peers during that time are a convenient and even preferred source of assistance (Knapp and Karabenick, 1988).

The Role of Technology

Help seeking increasingly takes place by other than face-to-face contact. More so than voice, computer-mediated communication (CMC) potentially affords students an unlimited landscape of individuals, virtual communities, information, and intelligent systems as resources (Keefer and Karabenick, 1998). Such availability can dramatically reduce the cost of seeking help both in terms of sheer time and effort but also in creating safe contexts by virtue of increased (or total) anonymity, which reduces the threat to self-worth and increases the likelihood that students will seek help (Karabenick and Knapp, 1988a). The wide availability of resources, however, also increases the potential for seeking help that is expedient and work avoidant and raises significant concerns that technology-assisted help seeking may tend to be less adaptive. Recent work is exploring this issue in students' use of online tutoring systems, which employs comprehensive models of the help-seeking process to provide context sensitive help (Aleven *et al.*, 2003, 2006). One emerging concern is that students tend to game the tutoring system in order to quickly complete online tasks rather than gain greater understanding of the material. It appears that the technology-mediated help-seeking process is subject to the same motivation-related characteristics of contexts in which it is situated (Schofield, 1995). Accordingly, whether learning occurs in a mastery-versus performance-focused instructional setting may be just as important when seeking help using advanced technologies as it is from teachers in the classroom.

Summary

Help seeking is a unique social and behavioral strategy of self-regulated learning that involves a social transaction with teachers, classmates, parents, and resources that may be technologically mediated. Its social dimension renders seeking help more complex than other regulation strategies such as rehearsing, organizing, self-checking, and self-testing, strategies that students can carry out independently. The promotion of strategic help seeking necessitates that teachers and other help providers such as counselors, tutors, or parents understand the distinction between help that is instrumental or adaptive from that undertaken to avoid work (expedient), such as seeking help excessively (Alexitch, 2006; Collins and Sims, 2006).

Teachers and peers play important roles in children becoming self-regulated learners. Adaptive help seeking is facilitated by fostering mastery goals. When teachers stress the intrinsic value of learning in their classrooms rather than just getting good grades (or avoiding bad grades), students tend to ask task-related questions in order to truly understand their work. Knowing how one is doing in relation to classmates can have both positive (e.g., it's normal to need help) and negative (e.g., concern about social status). Teachers can potentially affect processes of social comparison. Sensitive and responsive teachers buffer students from factors (e.g., potential embarrassment) that typically inhibit help seeking. They play an instrumental role in the development of children's personal beliefs about the costs and benefits of help seeking. They can help establish collectivistic (vs. individualistic) classroom norms and facilitate the degree to which peer involvement provides children opportunities to experience social aspects of learning. Teachers need to be aware of how technology, in the form of communications or online tutoring, can facilitate help seeking by reducing costs (time, effort, and embarrassment), but they also need to be aware of the potential for its excessive misuse. Increasingly, learners as well need to develop better adaptive help-seeking strategies, which would be important for teachers and other providers of help to foster as they would other cognitive and metacognitive learning strategies.

Bibliography

- Aleven, V., McLaren, B., and Koedinger, K. (2006). Toward computer-based tutoring of help-seeking skills. In Karabenick, S. A. and Newman, R. S. (eds.) *Help Seeking in Academic Settings: Goals, Groups and Contexts*, pp 259–296. Mahwah, NJ: Erlbaum.
- Aleven, V., Stahl, E., Schworm, S., Fischer, F., and Wallace, R. M. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research* 73(2), 277–320.
- Alexitch, L. R. (2006). Help seeking and the role of academic advising in higher education. In Karabenick, S. A. and Newman, R. S. (eds.) *Help Seeking in Academic Settings: Groups, Goals, and Contexts*, pp 175–202. Mahwah, NJ: Erlbaum.

- Ames, C. and Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology* **80**, 260–267.
- Arbreton, A. (1998). Student goal orientation and help-seeking strategy use. In Karabenick, S. A. (ed.) *Strategic Help Seeking: Implications for Learning and Teaching*, pp 95–116. Mahwah, NJ: Erlbaum.
- Beller, E. (1957). Dependency and autonomous achievement striving related to orality and anality in early childhood. *Child Development* **28**, 287–315.
- Berndt, T. J. and Keefe, K. (1996). Friends' influence on school adjustment: A motivational analysis. In Juvonen, J. and Wentzel, K. R. (eds.) *Social Motivation: Understanding Children's School Adjustment*, pp 248–278. Cambridge: Cambridge University Press.
- Butler, R. (1998). Determinants of help seeking: Relations between perceived reasons for classroom help-avoidance and help-seeking behaviors in an experimental context. *Journal of Educational Psychology* **90**, 630–644.
- Butler, R. (2006). An achievement goal perspective on student help seeking and teacher help giving in the classroom: Theory, research, and educational implications. In Karabenick, S. A. and Newman, R. S. (eds.) *Help Seeking in Academic Settings: Goals, Groups, and Contexts*, pp 15–44. Mahwah, NJ: Erlbaum.
- Butler, R. (2007). Teachers' achievement goal orientations and associations with teachers' help seeking: Examination of a novel approach to teacher motivation. *Journal of Educational Psychology* **99**(2), 241–252.
- Butler, R. and Neuman, O. (1995). Effects of task and ego achievement goals on help-seeking behaviors and attitudes. *Journal of Educational Psychology* **87**, 261–271.
- Church, M. A., Elliot, A. J., and Gable, S. L. (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology* **93**, 43–54.
- Collins, W. and Sims, B. C. (2006). Help seeking and the role of academic advising in higher education. In Karabenick, S. A. and Newman, R. S. (eds.) *Help Seeking in Academic Settings: Groups, Goals, and Contexts*, pp 203–223. Mahwah, NJ: Erlbaum.
- Cooper, C. R., Marquis, A., and Ayers-Lopez, S. (1982). Peer learning in the classroom: Tracing developmental patterns and consequences of children's spontaneous interactions. In Wilkinson, L. C. (ed.) *Communicating in the Classroom*, pp 68–84. New York: Academic Press.
- Covington, M. V. (1992). *Making the Grade: A Self-Worth Perspective on Motivation and School Reform*. Cambridge, MA: Cambridge University Press.
- Eccles, J. S. and Midgley, C. (1989). Stage-environment fit: Developmentally appropriate classrooms for young adolescents. In Ames, C. and Ames, R. (eds.) *Research on Motivation in Education*, vol. 3, pp 139–186. New York: Academic Press.
- Eccles, J. S. and Wigfield, A. (1985). Teacher expectations and student motivation. In Dusek, J. B. (ed.) *Teacher Expectations*, pp 185–226. Hillsdale, NJ: Erlbaum.
- Fischer, P. L. and Torney, J. V. (1976). Influence of children's stories on dependency: A sex-typed behavior. *Developmental Psychology* **12**(5), 489–490.
- Hashim, R. A., Yaakub, N. F., Hashim, H. J., Othman, A. H., and Ali, R. M. (2003). Correlates of academic help-seeking behaviours among adolescents: Insights from a cognitive-motivational perspective. Presented at the Biennial Meeting of the European Association for Research on Learning and Instruction. Padova, Italy.
- Karabenick, S. A. (ed.) (1998). *Strategic Help Seeking: Implications for Learning and Teaching*. Mahwah, NJ: Erlbaum.
- Karabenick, S. A. (2003). Help seeking in large college classes: A person-centered approach. *Contemporary Educational Psychology* **28**, 37–58.
- Karabenick, S. A. (2004). Perceived achievement goal structure and college student help seeking. *Journal of Educational Psychology* **96**, 569–581.
- Karabenick, S. A. and Knapp, J. R. (1988a). Effects of computer privacy on help-seeking. *Journal of Applied Social Psychology* **18**(6), 461–472.
- Karabenick, S. A. and Knapp, J. R. (1988b). Help-seeking and the need for academic assistance. *Journal of Educational Psychology* **80**, 406–408.
- Karabenick, S. A. and Knapp, J. R. (1991). Relationship of academic help seeking to the use of learning strategies and other instrumental achievement behavior in college students. *Journal of Educational Psychology* **83**(2), 221–230.
- Karabenick, S. A. and Newman, R. S. (eds.) (2006). *Help Seeking in Academic Settings: Goals, Groups, and Contexts*. Mahwah, NJ: Erlbaum.
- Karabenick, S. A. and Sharma, R. (1994). Perceived teacher support of student questioning in the college classroom: Its relation to student characteristics and role in the classroom questioning process. *Journal of Educational Psychology* **86**, 90–103.
- Karabenick, S. A., Zusho, A., and Kempler, T. M. (2005). Help seeking and perceived classroom context. Paper Presented at the Biennial Meeting of the European Association for Research on Learning and Instruction. Nicosia, Cyprus.
- Keefer, J. A. and Karabenick, S. A. (1998). Help seeking in the information age. In Karabenick, S. A. (ed.) *Strategic Help Seeking: Implications for Learning and Teaching*, pp 219–250. Mahwah, NJ: Erlbaum.
- Knapp, J. R. and Karabenick, S. A. (1988). Incidence of formal and informal academic help-seeking in higher education. *Journal of College Student Development* **29**(3), 223–227.
- Marchand, G. and Skinner, E. (2007). Motivational dynamics of children's academic help seeking and concealment. *Journal of Educational Psychology* **99**, 65–82.
- McCaslin, M. and Good, T. L. (1996). The informal curriculum. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 622–670. New York: Simon and Schuster Macmillan.
- Midgley, C. (ed.) (2002). *Goals, Goal Structures, and Patterns of Adaptive Learning*. Mahwah, NJ: Erlbaum.
- Nadler, A. (1998). Relationship, esteem, and achievement perspectives on autonomous and dependent help seeking. In Karabenick, S. A. (ed.) *Strategic Help Seeking: Implications for Learning and Teaching*, pp 61–93. Mahwah, NJ: Erlbaum.
- Nelson-Le Gall, S. (1981). Help-seeking: An understudied problem-solving skill in children. *Developmental Review* **1**, 224–246.
- Nelson-Le Gall, S. A. (1987). Necessary and unnecessary help-seeking in children. *Journal of Genetic Psychology* **148**(1), 53–62.
- Nelson-Le Gall, S. and Resnick, L. (1998). Help seeking, achievement motivation, and the social practice of intelligence in school. In Karabenick, S. A. (ed.) *Strategic Help Seeking: Implications for Learning and Teaching*, pp 39–60. Hillsdale, NJ: Erlbaum.
- Newman, R. S. (1990). Children's help-seeking in the classroom: The role of motivational factors and attitudes. *Journal of Educational Psychology* **82**, 71–80.
- Newman, R. S. (1998a). Adaptive help seeking: A role of social interaction in self-regulated learning. In Karabenick, S. A. (ed.) *Strategic Help Seeking: Implications for Learning and Teaching*, pp 13–37. Mahwah, NJ: Erlbaum.
- Newman, R. S. (1998b). Students' help seeking during problem solving: Influences of personal and contextual achievement goals. *Journal of Educational Psychology* **90**, 644–658.
- Newman, R. S. (2000). Social influences on the development of children's adaptive help seeking: The role of parents, teachers, and peers. *Developmental Review* **20**, 350–404.
- Newman, R. S. (2002). What do I need to do to succeed...when I don't understand what I'm doing? Developmental influences on students' adaptive help seeking. In Wigfield, A. and Eccles, J. (eds.) *Development of Achievement Motivation*, pp 285–306. San Diego, CA: Academic Press.
- Newman, R. S. (2007). The motivational role of adaptive help seeking in self-regulated learning. In Schunk, D. and Zimmerman, B. (eds.) *Motivation and Self-Regulated Learning: Theory, Research, and Application*, pp 315–337. Mahwah, NJ: Erlbaum.
- Newman, R. S. and Goldin, L. (1990). Children's reluctance to seek help with schoolwork. *Journal of Educational Psychology* **82**, 92–100.
- Newman, R. S. and Schwager, M. T. (1993). Student perceptions of the teacher and classmates in relation to reported help seeking in math class. *Elementary School Journal* **94**, 3–17.
- Nicholls, J. G. and Miller, A. (1984). Reasoning about the ability of self and others: A developmental study. *Child Development* **55**, 1990–1999.

- Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology* **25**, 92–104.
- Pintrich, P. R. and Zusho, A. (2002). Student motivation and self-regulated learning in the college classroom. In Smart, J. C. and Tierney, W. G. (eds.) *Higher Education: Handbook of Theory and Research*, vol. XVII, pp 55–128. New York: Agathon Press.
- Ruble, D. N. and Frey, K. S. (1991). Changing patterns of comparative behavior as skills are acquired: A functional model of self-evaluation. In Suls, J. and Wills, T. A. (eds.) *Social Comparison: Contemporary Theory and Research*, pp 79–113. Hillsdale, NJ: Erlbaum.
- Ryan, A., Gheen, M., and Midgley, C. (1998). Why do some students avoid asking for help? An examination of the interplay among students' academic efficacy, teachers' social-emotional role, and classroom goal structure. *Journal of Educational Psychology* **90**, 528–535.
- Ryan, A. M., Hicks, L., and Midgley, C. (1997). Social goals, academic goals, and avoiding help in the classroom. *Journal of Early Adolescence* **17**, 152–171.
- Ryan, A. M. and Pintrich, P. R. (1997). "Should I ask for help?" The role of motivation and attitudes in adolescents' help seeking in math class. *Journal of Educational Psychology* **89**, 329–341.
- Ryan, A. M. and Pintrich, P. R. (1998). Achievement and social motivational influences on help seeking in the classroom. In Karabenick, S. A. (ed.) *Strategic Help Seeking: Implications for Learning and Teaching*, pp 117–139. Mahwah, NJ: Erlbaum.
- Ryan, R. M. and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* **55**, 68–78.
- Sandoval, B. A. and Lee, F. (2006). When is seeking help appropriate? Now norms affect help seeking in organizations. In Karabenick, S. A. and Newman, R. S. (eds.) *Help Seeking in Academic Settings: Groups, Goals, and Contexts*, pp 151–173. Mahwah, NJ: Erlbaum.
- Schofield, J. W. (1995). *Computers and Classroom Culture*. Cambridge, MA: Cambridge University Press.
- Sears, R. R., Maccoby, E. E., and Levin, H. (1957). *Patterns of Child Rearing*. Oxford: Row, Peterson.
- Shwalb, D. W. and Sukemune, S. (1998). Help seeking in the Japanese college classroom: Cultural, developmental, and social-psychological influences. In Karabenick, S. A. (ed.) *Strategic Help Seeking: Implications for Learning and Teaching*, pp 141–170. Mahwah, NJ: Erlbaum.
- Skinner, E. and Zimmer-Gembeck, M. (2007). The development of coping. *Annual Review of Psychology* **58**, 119–144.
- Triandis, H. (1994). Theoretical and methodological approaches to the study of collectivism and individualism. In Kim, U., Triandis, H., Kagitcibasi, C., Choi, S. -C., and Yoon, G. (eds.) *Individualism and Collectivism: Theory, Method and Applications: Cross-Cultural Research and Methodology Series*, pp 41–51. Thousand Oaks, CA: Sage.
- Turner, J. C., Midgley, C., Meyer, D. K., et al. (2002). The classroom environment and students' reports of avoidance strategies in mathematics: A multimethod study. *Journal of Educational Psychology* **94**, 88–106.
- Urdu, T., Ryan, A. M., Anderman, E. M., and Gheen, M. H. (2002). Goals, goal structures, and avoidance behaviors. In Midgley, C. (ed.) *Goals, Goal Structures, and Patterns of Adaptive Learning*, pp 55–84. Mahwah, NJ: Erlbaum.
- Violet, S. and Karabenick, S. A. (2006). Help seeking in cultural context. In Karabenick, S. A. and Newman, R. S. (eds.) *Help Seeking in Academic Settings: Groups, Goals, and Contexts*, pp 117–150. Mahwah, NJ: Erlbaum.
- Webb, N. M., Ing, M., Kersting, N., and Nemer, K. M. (2006). Help seeking in cooperative learning groups. In Karabenick, S. A. and Newman, R. S. (eds.) *Help Seeking in Academic Settings: Goals, Groups, and Contexts*, pp 65–121. Mahwah, NJ: Erlbaum.
- Webb, N. M. and Palincsar, A. S. (1996). Group processes in the classroom. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 841–873. New York: Simon and Schuster Macmillan.
- Wentzel, K. R. (1997). Student motivation in middle school: The role of perceived pedagogical caring. *Journal of Educational Psychology* **89**, 411–419.
- Winterbottom, M. R. (1958). The relation of need for achievement to learning experiences in independence and mastery. In Atkinson, J. W. (ed.) *Motives in Fantasy, Action, and Society*, pp 453–478. New York: Van Nostrand.
- Zimmerman, B. J. and Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology* **82**, 51–59.
- Zusho, A., Karabenick, S. A., Bonney, C. R., and Sims, B. C. (2007). Contextual determinants of motivation and help seeking in the college classroom. In Perry, R. P. and Smart, J. C. (eds.) *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*, pp 611–659. Dordrecht: Springer.

Further Reading

- Butler, R. (2007). Teachers' achievement goal orientations and associations with teachers' help seeking: Examination of a novel approach to teacher motivation. *Journal of Educational Psychology* **99**(2), 241–252.
- Karabenick, S. A. and Moosa, S. (2005). Culture and personal epistemology: Middle Eastern students' beliefs about scientific knowledge and knowing. *Social Psychology of Education* **8**, 375–393.
- Nelson-Le Gall, S. and Glor-Scheib, S. (1986). Academic help-seeking and peer relations in school. *Contemporary Educational Psychology* **11**(2), 187–193.
- Newman, R. S. (2007). The motivational role of adaptive help seeking in self-regulated learning. In Schunk, D. and Zimmerman, B. (eds.) *Motivation and Self-Regulated Learning: Theory, Research, and Application*, pp 315–337. Mahwah, NJ: Erlbaum.
- Ryan, A., Gheen, M., and Midgley, C. (1998). Why do some students avoid asking for help? An examination of the interplay among students' academic efficacy, teachers' social-emotional role, and classroom goal structure. *Journal of Educational Psychology* **90**, 528–535.

Self-Concept in Learning: Reciprocal effects model between academic self-concept and academic achievement

H W Marsh and L F Scalas, Oxford University, Oxford, UK

© 2010 Elsevier Ltd. All rights reserved.

Self-concept is one of the oldest and most important constructs in the social sciences, dating back to at least the seminal work by William James (1890/1963). Today, self-concept enhancement is a major goal in many fields, including education, child development, health, sport/exercise sciences, social services, and management. This concept is a multidimensional hierarchical construct with highly differentiated components such as academic, social, physical, and emotional self-concepts in addition to a global self-concept component. It is also an important mediating factor that facilitates the attainment of other desirable outcomes. Particularly in education settings, a positive academic self-concept (ASC) is both a highly desirable goal and a means of facilitating subsequent learning and other academic accomplishments.

Multidimensionality of Self-Concept

Definition of Self-Concept

Historically, self-concept measurement, theory, research, and application have been plagued by the poor quality of both theoretical models and self-concept measurement instruments (e.g., Shavelson *et al.*, 1976; Wells and Marwell, 1976; Wylie, 1974, 1979). In an attempt to remedy this situation, Shavelson *et al.* (1976) reviewed existing self-concept research and instruments, proposed a new theoretical model of self-concept, and provided a blueprint for the development of multidimensional self-concept instruments (see review by Marsh and Hattie, 1996).

Self-concept, broadly defined by Shavelson *et al.* (1976), is a person's self-perceptions formed through experience and interpretations of one's environment. It includes feelings of self-confidence, self-worth, self-acceptance, competence, and ability. It is influenced especially by evaluations by significant others, reinforcements, and attributions for one's own behavior. Self-concept is not an entity within the person, but a hypothetical construct that is potentially useful in understanding individuals and how they behave. Shavelson *et al.* noted that self-concept is important both as an outcome and as a mediating variable that helps to explain other outcomes. Self-perceptions influence the way important outcomes and these outcomes in turn influence one's self-perceptions.

Multidimensional, Hierarchical Model of Self-Concept

Shavelson *et al.* (1976) identified seven features that were critical to their definition of the self-concept construct. The most important for present purposes are that self-concept is multifaceted and hierarchically organized with perceptions of personal behavior in specific situations at the base of the hierarchy, inferences about self in broader domains (e.g., social, physical, and academic) at the middle of the hierarchy, and global self-concept (also known as self-esteem) at the apex (see **Figure 1**).

Remarkably, at the time Shavelson *et al.* proposed their model, there was little support for the multidimensionality of self-concept. The resistance to the multidimensional aspect of the model in particular was so strong that leading researchers of that period (e.g. Coopersmith, 1967; Marx and Winne, 1978) argued that self-concept was either a unidimensional construct or that the facets of self-concept were dominated so heavily by a general factor that they could not be differentiated adequately. Recently, Byrne (1984: 449–450) noted that: “Many consider this inability to attain discriminant validity among the dimensions of SC to be one of the major complexities facing SC researchers today.” In retrospect – as clearly articulated by Shavelson *et al.* (1976) – the renaissance of self-concept research was floundering due to the lack of a solid basis of measurement to support it.

However, the Shavelson *et al.* model provided a blueprint for the development of a whole new generation of multidimensional self-concept instruments that have provided overwhelming support for the multidimensionality of self-concept. There are several widely used inventories for measuring multiple dimensions of self-concept that, to some extent, differ in the self-concept dimensions included (e.g., Bracken, 1996; Harter, 1998; Marsh, 1990b; see review by Byrne, 1996b). Typically, however, these instruments include at least one or more factors representing the specific academic (e.g., mathematical and verbal self-concept), social (e.g., relations with friends and relations with parents), physical (e.g., physical competence and attractiveness), and emotional domains of self-concept, and a global self-esteem scale as posited in the Shavelson *et al.* (1976) model. Among the various instruments, external reviews (see Boyle, 1994; Byrne, 1996b;

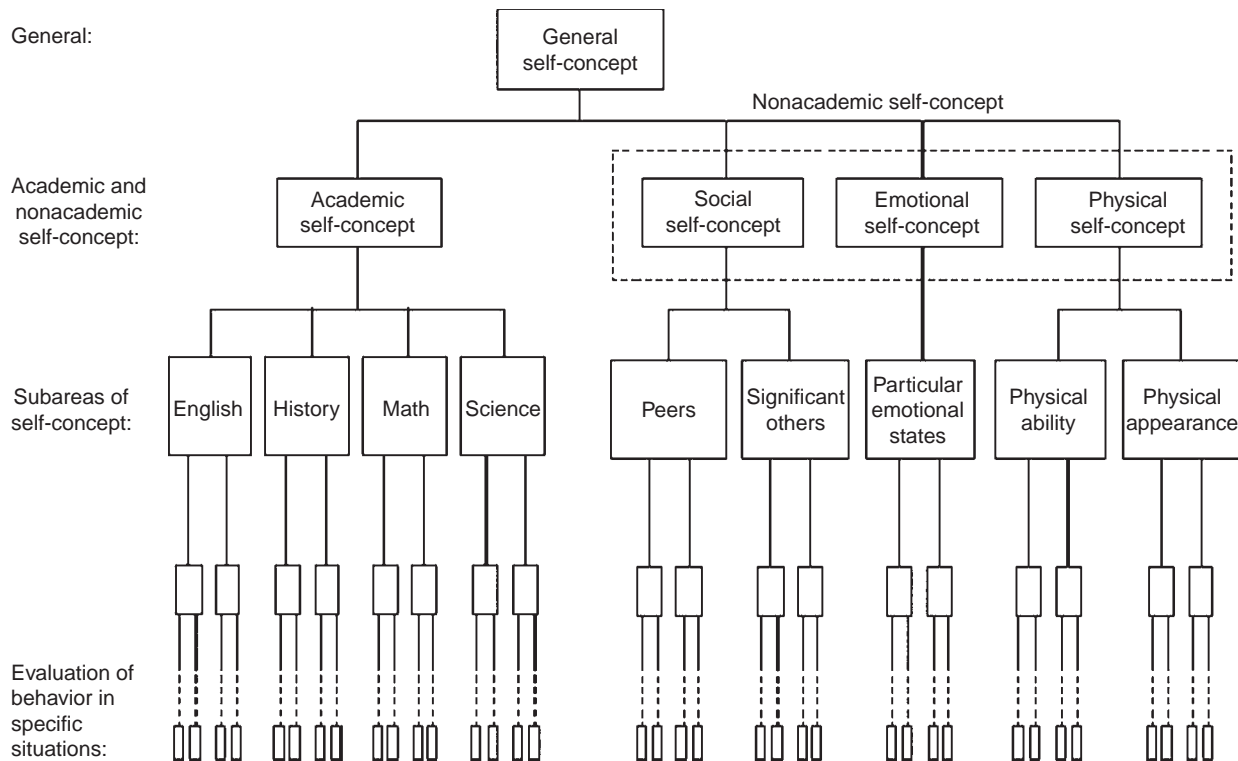


Figure 1 Pictorial representation of the multidimensional, hierarchical model of self-concept posited by Shavelson *et al.* (1976) and Marsh and Shavelson (1985). The box consisting of dashed lines around the nonacademic self-concept factors is used to distinguish these from the academic self-concept factors, but does not imply that there is a single higher-order nonacademic factor, as is hypothesized for the academic factors. The unlabeled boxes in the bottom of the hierarchy are used to show that the model posits additional levels in the hierarchy and even more domain-specific components of self-concept than those that are explicitly presented (e.g., mathematical self-concept might be broken into different mathematical topics such as algebra, trigonometry, or calculus, and each of these could be further subdivided into specific components relevant to each of the mathematical subjects). Reprinted from Shavelson, R. J., Hubner, J. J., and Stanton, G. C. (1976). Validation of construct interpretations. *Review of Educational Research* 46, 413.

Hattie, 1992; Wylie, 1989) suggest that the Self Description Questionnaire (SDQ) instruments are the strongest multidimensional self-concept instruments for children (SDQI), adolescents (SDQII), and young adults (SDQIII), respectively. In less than a decade, the accepted wisdom in self-concept research moved from a unidimensional perspective in which the multidimensionality of self-concept was all but dismissed to a multidimensional perspective that was particularly embraced in educational research (see reviews by Marsh and Hattie, 1996; also see Byrne, 1996a, 1996b; 2002). The hierarchical aspect of the Shavelson model linked these two approaches, including both a global component of self-concept at the apex of the model and increasingly differentiated multiple dimensions of self-concept at lower levels of the model.

Marsh and Craven (2006) argued that the acceptance of a multidimensional rather than a unidimensional perspective of self-concept varies substantially across various social science disciplines and within subdisciplines in psychology. However, its broadest acceptance and strongest

support comes from educational psychology with its focus on ASC (Marsh, 1993) and its relation to academic achievement, school grades, student learning, and other academic outcomes.

Support for Multidimensionality: Correlations between ASC and Achievement

Various measures of academic achievement are substantially correlated to corresponding measures of ASC, but are nearly uncorrelated (or even negatively correlated) with nonacademic components of self-concept and self-esteem (Marsh, 1993; Marsh and Craven, 2006). Thus, for example, Marsh and O'Neill (1984) related mathematics and English achievement to responses by high school students to the SDQIII instrument. Mathematics achievement correlated 0.58, 0.27, and 0.11 with math, general academic, and verbal self-concepts, respectively, whereas English achievement correlated 0.42, 0.24, and 0.19 with verbal, general academic, and math self-concepts,

respectively. Remarkably, none of the nine nonacademic scales – including global self-esteem – was significantly related to either of the achievement scores. Demonstrating the generalizability of these effects, Marsh *et al.*, (1988) found that correlations between math and English self-concepts based on each of three different instruments were close to zero, that math achievement was substantially correlated with math self-concept but not English self-concept, and that English achievement is substantially correlated with English self-concept but not math self-concept. Similarly, Marsh (1992) established that relations between ASCs in eight specific school subjects were substantially related to school grades in the matching school subjects ($r_s = 0.45\text{--}0.70$), offering support for the external validity of specific facets of ASC. In contrast, self-esteem was nearly uncorrelated with school grades in all the school subjects, indicating that it had no validity in relation to this criterion. More recently, Marsh *et al.* (2006) demonstrated a predictable pattern of substantial relations between eight academic criterion variables (grades, test scores, and coursework selection in different school subjects) and corresponding ASCs, whereas self-esteem was nearly uncorrelated with all these criteria ($r_s = -0.03$ to $0.05 >$).

Although self-concept and school grades are substantially correlated, there are also important differences between the two constructs. Thus, for example, academic achievements in different school subjects are substantially correlated, indicating a strong hierarchical ordering in which much of the variance in specific subject areas can be explained by a global achievement factor (or intelligence quotient (IQ)). In contrast, self-concepts in different school subjects are highly differentiated. In fact, even though math and verbal achievement are highly correlated (r_s of $0.5\text{--}0.8$), math and verbal self-concepts are nearly uncorrelated. Marsh and Hau (2004) demonstrated support for a theoretical model designed to explain this juxtaposition between ASC and achievement domains generalized across nationally representative samples of students from 26 different countries. This extreme differentiation among math and verbal self-concepts also led to the Marsh/Shavelson revision of the original Shavelson model in which self-concepts in core academic subjects are represented by two higher-order ASC factors (math/academic and verbal/academic) rather than one (Marsh, 1992; Marsh *et al.*, 1988).

Furthermore, consistent with this distinction between self-concept and achievements in different school subjects are relations between ASC, achievement, and other academic criteria. For example, Marsh and Yeung (1997) demonstrated that whereas self-concepts in different school subjects and matching school grades were substantially correlated, the specific components of ASC predicted subsequent coursework selection much better than did school grades or more general components of self-concept. These results provide empirical evidence calling into

question the usefulness of a general self-esteem construct in educational psychology research, and offer strong support for the multidimensional perspective. In summary, ASC and achievement are not only systematically related, but also very distinct constructs.

ASC/Learning Relationship: Reciprocal Effects Model

Causal-Ordering ASC and Achievement: Reciprocal Effects Model

Do changes in ASC lead to changes in subsequent academic performance? Correlational studies provide convincing evidence of the strong relations between ASC and various measures of academic accomplishment, achievement, and learning. However, a much more demanding question is whether ASC causes achievement or achievement causes ASC. Historically, the causal ordering of self-concept and performance has been, perhaps, the most vexing question in self-concept research (Byrne, 1984; Marsh, 1993; Marsh and Craven, 2006). This critical question has important theoretical and practical implications, and has been the focus of considerable research – particularly in educational settings.

Byrne (1984) emphasized that much of the interest in the ASC/achievement relation stems from the belief that ASC has motivational properties such that changes in ASC will lead to changes in subsequent academic achievement. As self-concept and academic achievement are not readily amenable to experimental manipulations, most research relies on longitudinal panel data in which both self-concept and achievement are measured on at least two occasions (i.e., a two-wave, two-variable design) and preferably three or more (see **Figure 2**). Calsyn and Kenny (1977) contrasted self-enhancement and skill development models of this relation. The self-enhancement model posits self-concept as a primary determinant of academic achievement (i.e., self-concept \Rightarrow achievement) and would support self-concept enhancement interventions. In contrast, the skill development model implies that ASC emerges principally as a consequence of academic achievement (i.e., achievement \Rightarrow self-concept) so that the best way to enhance ASC would be to develop stronger academic skills.

In a review and critique of this research, Marsh (1990a, 1990b, 1993) argued that much of this research was methodologically unsound and inconsistent with self-concept theory (Marsh *et al.*, 1999). He emphasized that according to both common sense and theory, prior academic achievement was one determinant of ASC. Hence, the critical question was whether there also existed a significant path from prior ASC to subsequent achievement – regardless of whether or not this path was larger than the path from prior academic achievement to subsequent ASC. Marsh (1990a)

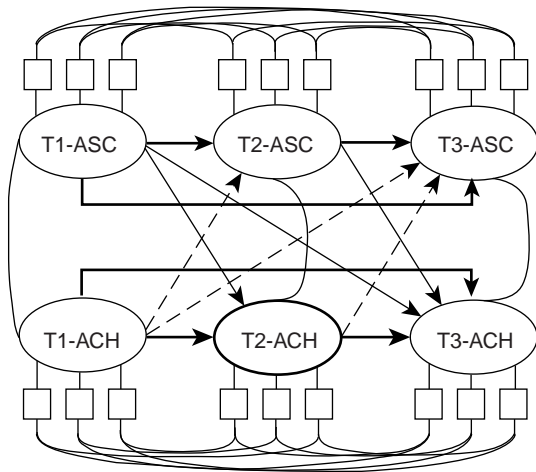


Figure 2 Prototype causal-ordering model for testing self-enhancement, skill development, and reciprocal effects models. In this full-forward, multiwave, multivariable model, multiple indicators of academic self-concept (ASC) and achievement (ACH) are collected in three successive waves (T1, T2, and T3). Each latent construct (represented by ovals) has paths leading to all latent constructs in subsequent waves. Within each wave, academic self-concept and achievement are assumed to be correlated; in the first wave, this correlation is a covariance between two latent constructs, and in subsequent waves, it is a covariance between residual factors. Curved lines at the top and bottom of the figure reflect correlated uniqueness between responses to the same measured variable (represented by boxes) collected on different occasions. Paths connecting the same variable on multiple occasions reflect stability (the solid gray paths), but these coefficients typically differ from the corresponding test-retest correlations (which do not include the effects of other variables). Dashed lines reflect effects of prior achievement on subsequent self-concept, whereas solid black lines reflect the effects of prior self-concept on subsequent achievement.

further argued that a more realistic compromise between the self-enhancement and skill development models was a reciprocal effects model (REM) in which prior self-concept affects subsequent achievement and prior achievement affects subsequent self-concept (Marsh and Craven, 2006; Marsh *et al.*, 1999). He further argued that this early emphasis on an either/or model of causal ordering was due largely to limitations in statistical techniques for testing these models in the 1980s that are no longer relevant.

Bringing together stronger theoretical and statistical bases for addressing these issues, Marsh (1990a) proposed the REM of ASC. **Figure 2** presents a prototypical REM designed to test the causal ordering of ASC and achievement. Self-concept and academic achievement are each measured at least 3 times in this longitudinal panel design. The critical issue is whether there are statistically significant paths leading from prior self-concept to subsequent achievement (in support of self-enhancement predictions) and from prior achievement to subsequent self-concept (in support of skill development predictions). Support for the

REM requires that both sets of paths are statistically significant; however, from the perspective of self-concept theory and practice, the linkages from prior self-concept to subsequent achievement are particularly important.

A growing body of research, reviewed by Marsh and Craven (2006), has established support for the REM of relations between ASC and academic achievement (Marsh *et al.*, 1999). In a recent meta-analysis of relevant research, Valentine *et al.* (2004) also concluded that there was clear support for predictions based on the REM over those derived from self-enhancement and skill development models.

This reciprocal pattern of relations between self-concept and performance posited in the REM is also represented in many other theoretical accounts of self-concept and related self-belief constructs (e.g., Bandura, 1997; Byrne, 2002; Eccles, 1983; Eccles and Wigfield, 2002; Harter, 1998, 1999; Hattie, 1992; Skaalvik, 1997; Valentine and DuBois, 2005; Valentine *et al.*, 2004; Wigfield and Eccles, 2002) as well as in the broader themes of reciprocal patterns of relation in developmental psychology (e.g., Lerner, 1982, 1996). Thus, for example, expectancy-value theory (Eccles, 1983; Eccles and Wigfield, 2002) hypothesized academic self-beliefs to be a function of prior academic successes and to affect subsequent academic success directly or indirectly through their influence on other mediating constructs. More generally, in their review of theoretical research and meta-analysis of empirical research, Valentine and DuBois (2005) concluded that the posited reciprocal effects relating academic self-beliefs and achievement are consistent with theories of learning and human development that view the self as a causal agent (e.g., Bandura, 1997; Carver and Scheier, 1981; Deci and Ryan, 1985). Indeed, Valentine and DuBois concluded that support for the REM was equally strong for domain-specific ratings of ASC and self-efficacy.

Extensions of the REM

In their review of theoretical and empirical support for the REM, Marsh *et al.* (1999; also see Marsh and Craven (2006)) argued for the need for further research to test the generalizability of the REM over nationality and culture (support was based largely on studies done in Western, English-speaking countries), age (support was based primarily on research with adolescents), and content area (research was based largely on studies in a limited number of academic domains – particularly mathematics).

Cross-cultural generalizability

Partly in response to Marsh *et al.* (1999), recent research demonstrated that this support for the REM of ASC and achievement generalized to different cultural/national settings in a large nationally representative sample of

Hong Kong students (Marsh *et al.*, 2002) and large samples of East and West German students at the time of the fall of the Berlin Wall (Marsh and Köller, 2003; Marsh *et al.*, 2001). Support for the generalizability also comes from research based on French-speaking Canadian primary students (Guay *et al.*, 2003) and the German high school students (Marsh *et al.*, 2005). Hence, there is strong cross-national and cross-cultural support for the REM.

Generalizability over age

Based on developmental theory, some researchers have suggested that the reciprocal pattern of relations in support of the REM found with adolescents is unlikely to generalize to preadolescents (see Wigfield and Karpachian, 1991). However, two reviews of this literature (Marsh *et al.*, 1999; Valentine *et al.*, 2004) concluded that there was not sufficient good-quality research with young children to support this conclusion. Guay *et al.* (2003) addressed this issue about developmental trends in REM research. They used a multicohort–multioccasion design for responses by students in grades 2, 3, and 4 (i.e., three age cohorts aged 8–10 years of age, each with three measurement occasions). The structural equation model (SEM) for the total sample supported an REM for the first two waves of data (paths leading from prior self-concept to subsequent achievement, and from prior achievement to subsequent self-concept) and a self-enhancement effect (paths leading prior self-concept to subsequent achievement) between the second and the third waves. This pattern was replicated in tests of the invariance of the SEM across the three age cohorts, demonstrating support for the generalizability of the REM across these preadolescent ages.

Generalizability to the physical domain

Although there is a growing body of research based on ASC and academic achievement, Marsh *et al.* (1999) noted that there were few tests in nonacademic domains. More recently, Marsh and colleagues have undertaken a number of studies evaluating the generalizability of the REM to the physical domain for both general populations and elite athletes.

Marsh *et al.* (2005) evaluated the generalizability of the REM for gymnastics self-concept and gymnastics performance (independently evaluated by expert judges viewing videotapes of a standardized performance test). Consistent with *a priori* predictions in support of the REM, the effect of T1 gymnastics self-concept on T2 gymnastics performance (0.20) and the effect of T1 gymnastics performance on T2 gymnastics self-concept (0.14) were both highly significant. Consistent with the REM, gymnastics self-concept and gymnastics performance were both determinants and consequences of each other.

Recognizing the critical importance of health-related physical self-concept in children and adolescents,

Marsh *et al.* (2006) adapted the REM in a study of the causal ordering of physical self-concept and exercise behavior. The study was based on a large sample of primary and secondary Greek physical education students (2786 students, 200 classes, 67 teachers) and data collected early (T1) and late (T2) in the school year. There was clear support for the REM as there were significant effects of T1 physical self-concept on T2 exercise behavior and T1 exercise behavior on T1 physical self-concept. Physical self-concept is both an effect and a cause of exercise behavior.

Does physical self-concept influence subsequent physical performance during preadolescence? Marsh *et al.* (in press) combined the need to test the generalizability of the REM with children and in the physical domain. They used longitudinal data for young boys and girls ($N = 1135$; M age = 9.67 years) to show that physical self-concept is both a cause and a consequence of physical accomplishments. After controlling for prior physical performance (physical-performance-based test and teacher assessments in grade 3, primary school), physical self-concept had a positive effect on subsequent physical performance in both grade 4 and, following the transition to secondary school, grade 6. Despite the fact that physical self-concept is a gender-stereotyped construct (with boys having systematically higher physical self-concepts; Crain, 1996; Marsh, 1989), support for the REM generalized over gender. In addition, gender differences favoring boys in grades 4 and 6 were largely mediated by prior differences in grade 3. Coupled with previous REM research based largely on studies of adolescents in the academic domain, this study supports the REM's generalizability over gender, self-concept domain, preadolescent ages, and the transition from primary to secondary school.

How well does support for the generalizability of the REM generalize to elite athletes? Marsh and Perry (2005) tested the effects of sport self-concept on subsequent performance in 270 elite swimmers from 30 countries participating in the Pan Pacific Swimming Championships and the World Short Course Championships. Whereas subsequent championship performance was highly related to prior personal best performances ($r = 0.90$), SEMs demonstrated that elite athlete self-concept contributed significantly to the prediction of subsequent championship performance, explaining approximately 10% of the residual variance after controlling for personal best performances. As each swimmer typically competed in at least two different events, we were able to show that support for the REM was nearly identical for both events.

Marsh and Craven versus Baumeister *et al.* Debate

Support for the importance of self-concept and the REM is part of a revolution sweeping psychology, one that

emphasizes a positive psychology focusing on how healthy, normal, and exceptional individuals can get the most from life (e.g., Seligman and Csikszentmihalyi, 2000). Positive self-beliefs are at the heart of this revolution (Hunter and Csikszentmihalyi, 2003; Marsh and Craven, 2006).

In a potentially serious threat both to this positive psychology movement and to the REM, Baumeister *et al.* (2003, 2005) challenged the prevailing optimistic perspective of the value of positive self-beliefs in a highly influential review commissioned for *Psychological Science in the Public Interest*. In their reviews, they posed the question: “Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles?” Drawing a negative conclusion to their question, Baumeister *et al.* (2003) concluded that “self-esteem per se is not the social panacea that many people hoped it was” (p. 38), a point reiterated by Baumeister *et al.* (2005) in their *Scientific American* article in which they concluded “that efforts to boost people’s self-esteem are of little value in fostering academic achievement or preventing undesirable behavior” (p. 84).

Particularly in relation to models of causal relations between self-concept and academic achievement – a major focus of the Baumeister *et al.* (2003, 2005) reviews – Marsh and Craven (2006) provided convincing evidence for the consistent positive effects of ASC on subsequent achievement after controlling the effects of prior achievement. They argued that conclusions drawn by Baumeister and colleagues were based largely on research studies, statistical methodology, and theoretical conceptualizations of self-concept that are no longer current. Although there were many points of agreement between the reviews by Marsh and Craven (2006) and those by Baumeister and colleagues (e.g., the need for longitudinal studies, cross-lagged designs, and appropriate statistical analyses), they differed particularly in relation to Baumeister *et al.*’s sole reliance on self-esteem and an implicit unidimensional perspective of self-concept compared to the explicitly multidimensional perspective taken by Marsh and Craven. Integrating the two apparently contradictory perspectives, Marsh and Craven argued that it was quite reasonable and consistent with their multidimensional perspective that there are almost no reciprocal relations between academic achievement and self-esteem (as reported by Baumeister *et al.*), whereas there are consistent reciprocal relations between ASC and achievement (as reported by Marsh and Craven). Indeed, there was almost no overlap between the older, multiple regression studies of self-esteem considered by Baumeister *et al.* and the more recent SEM studies of ASC emphasized by Marsh and Craven. Consistent with this rapprochement, Valentine and Dubois’ (2005) meta-analysis showed that the effect on subsequent school performance was stronger for academic self-beliefs, such as ASC, than for more global self-belief constructs, such as global self-esteem,

and even stronger when the self-belief construct logically matched the achievement construct in terms of domain specificity (i.e., mathematics achievement and math self-concept).

Summary and Implications

The results of the causal modeling studies in particular provide a clear affirmative answer to the question: Do changes in ASC lead to changes in subsequent academic achievement? This research is critically important in that it has established that increases in ASC lead to increases in subsequent academic achievement and other desirable educational outcomes. Hence, not only is self-concept an important outcome variable in itself, but it also plays a central role in mediating the effects of other desirable educational outcomes. These findings have significant implications for international educational policy and practice. It is important to emphasize that the direction of causality between ASC and achievement also has very important practical implications for educators. If the direction of causality were from ASC to achievement (the self-enhancement model), then teachers might be justified in placing more effort into enhancing students’ self-concepts rather than fostering achievement. On the other hand, if the direction of causality were from achievement to self-concept (the skill development model), then teachers should focus primarily on improving academic skills as the best way to improve self-concept. In contrast to both these apparently overly simplistic (either-or) models, the REM implies that ASC and academic achievement are reciprocally related and mutually reinforcing. Improved ASCs will lead to better achievement and improved achievement will lead to better ASCs. For example, if teachers enhance students’ ASCs without improving achievement, then the gains in self-concept are likely to be short lived. However, if teachers improve students’ academic achievement without also fostering students’ self-beliefs in their academic capabilities, then the achievement gains are also unlikely to be long lasting. If teachers focus on either one of these constructs to the exclusion of the other, then both are likely to suffer. Hence, according to the REM, teachers should strive to simultaneously improve both ASC and achievement.

Bibliography

- Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York: Freeman.
- Baumeister, R. F., Campbell, J. D., Krueger, J. I., and Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest* 4, 1–44.

- Baumeister, R. F., Campbell, J. D., Krueger, J. I., and Vohs, K. D. (2005). Exploding the self-esteem myth. *Scientific American* **292**, 84–92.
- Boyle, G. J. (1994). Self-description questionnaire II: A review. *Test Critiques* **10**, 632–643.
- Bracken, B. A. (ed.) (1996). *Handbook of Self-Concept: Developmental, Social, and Clinical Considerations*. New York: Wiley.
- Byrne, B. M. (1984). The general/academic self-concept nomological network: A review of construct validation research. *Review of Educational Research* **54**, 427–456.
- Byrne, B. M. (1996a). Academic self-concept: Its structure, measurement, and relation to academic achievement. In Bracken, B. A. (ed.) *Handbook of Self-Concept: Developmental, Social, and Clinical Considerations*, pp 287–316. New York: Wiley.
- Byrne, B. M. (1996b). *Measuring Self-Concept across the Life Span: Issues and Instrumentation*. Washington, DC: American Psychological Association.
- Byrne, B. M. (2002). Validating the measurement and structure of self-concept: Snapshots of past, present, and future research. *American Psychologist* **57**, 897–909.
- Calsyn, R. and Kenny, D. (1977). Self-concept of ability and perceived evaluations by others: Cause or effect of academic achievement? *Journal of Educational Psychology* **69**, 136–145.
- Carver, C. S. and Scheier, M. F. (1981). *Attention and Self-Regulation: A Control-Theory Approach to Human Behavior*. New York: Springer.
- Coopersmith, S. A. (1967). *The Antecedents of Self-Esteem*. San Francisco, CA: Freeman.
- Crain, R. M. (1996). The influence of age, race, and gender on child and adolescent multidimensional self-concept. In Bracken, B. A. (ed.) *Handbook of Self-Concept: Developmental, Social, and Clinical Considerations*, pp 395–420. New York: Wiley.
- Deci, E. L. and Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.
- Eccles, J. S. (1983). Expectations, values and academic behaviors. In Spence, J. T. (ed.) *Achievement and Achievement Motivations*, pp 75–146. San Francisco, CA: Freeman.
- Eccles, J. S. and Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology* **53**, 109–132.
- Guay, F., Marsh, H. W., and Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology* **95**, 124–136.
- Harter, S. (1998). Developmental perspectives on the self-system. In Eisenberg, N. (ed.) *Handbook of Child Psychology*, 5th edn., vol. 3, pp 553–618. New York: Wiley.
- Harter, S. (1999). *The Construction of the Self*. New York: Guilford Press.
- Hattie, J. (1992). *Self-Concept*. Hillsdale, NJ: Erlbaum.
- Hunter, J. and Csikszentmihalyi, M. (2003). The positive psychology of interested adolescents. *Journal of Youth and Adolescents* **22**(1), 27–35.
- James, W. (1963). *The Principles of Psychology*. (orig. publ. 1890) New York: Holt, Rinehart and Winston.
- Lerner, R. M. (1982). Children and adolescents as producers of their own development. *Developmental Review* **2**, 342–370.
- Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early-adulthood. *Journal of Educational Psychology* **81**, 417–430.
- Marsh, H. W. (1990a). The causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology* **82**, 646–656.
- Marsh, H. W. (1990b). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology* **82**, 623–636.
- Marsh, H. W. (1992). The content specificity of relations between academic achievement and academic self-concept. *Journal of Educational Psychology* **84**, 35–42.
- Marsh, H. W. (1993). Academic self-concept: Theory measurement and research. In Suls, J. (ed.) *Psychological Perspectives on the self*, vol. 4, pp 59–98. Hillsdale, NJ: Erlbaum.
- Marsh, H. W., Byrne, B. M., and Shavelson, R. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology* **80**, 366–380.
- Marsh, H. W., Byrne, B. M., and Yeung, A. S. (1999). Causal ordering of academic self-concept and achievement: Reanalysis of a pioneering study and revised recommendations. *Educational Psychologist* **34**, 154–157.
- Marsh, H. W., Chanal, J. P., Sarrazin, P. G., and Bois, J. E. (2005). Self-belief does make a difference: A reciprocal effects model of the causal ordering of physical self-concept and gymnastics performance. *Journal of Sport Sciences* **27**, 53–70.
- Marsh, H. W. and Craven, R. G. (2006). What comes first? A reciprocal effects model of the mutually reinforcing effects of academic self-concept and achievement. In Marsh, H. W., Craven, R. G., and McInerney, D. M. (eds.) *International advances in self research*, vol. 2, pp 15–52. Greenwich, CT: Information Age.
- Marsh, H. W., Gerlach, E., Trautwein, U., Lüdtke, U. and Brettschneider, W.-D. (in press). Physical self-concept and performance: Tests of reciprocal effects model of causal ordering in preadolescence. *Child Development*.
- Marsh, H. W. and Hattie, J. (1996). Theoretical perspectives on the structure of self-concept. In Bracken, B. A. (ed.) *Handbook of Self-Concept*, pp 38–90. New York: Wiley.
- Marsh, H. W. and Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalisability of the internal-external frame of reference predictions across 26 countries. *Journal of Educational Psychology* **96**, 56–67.
- Marsh, H. W., Hau, K. T., and Kong, C. K. (2002). Multilevel causal ordering of academic self-concept and achievement: Influence of language of instruction (English compared with Chinese) for Hong Kong students. *American Educational Research Journal* **39**, 727–763.
- Marsh, H. W. and Köller, O. (2003). Bringing together two theoretical models of relations between academic self-concept and achievement. In Marsh, H. W., Craven, R. G., and McInerney, D. M. (eds.) *International Advances in Self Research*, vol. 1, pp 17–48. Greenwich, CT: Information Age.
- Marsh, H. W., Köller, O., and Baumert, J. (2001). Reunification of East and West German school systems: Longitudinal multilevel modeling study of the big-fish-little-pond effect on academic self-concept. *American Educational Research Journal* **38**(2), 321–350.
- Marsh, H. W. and O'Neill, R. (1984). Self Description Questionnaire III (SDQ III): The construct validity of multidimensional self-concept ratings by late-adolescents. *Journal of Educational Measurement* **21**, 153–174.
- Marsh, H. W., Papaioannou, A., and Theodorakis, Y. (2006). Causal ordering of physical self-concept and exercise behavior: Reciprocal effects model and the influence of physical education teachers. *Health Psychology* **25**(3), 316–328.
- Marsh, H. W. and Perry, C. (2005). Does a positive self-concept contribute to winning gold medals in elite swimming? The causal ordering of elite athlete self-concept and championship performances. *Journal of Sport and Exercise Psychology* **27**, 71–91.
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., and Baumert, J. (2005). Academic self-concept, interest, grades and standardized test scores: Reciprocal effects models of causal ordering. *Child Development* **76**, 297–416.
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., and Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of Personality* **74**, 403–456.
- Marx, R. W. and Winne, P. H. (1978). Construct interpretations of three self-concept inventories. *American Educational Research Journal* **15**, 99–108.
- Seligman, M. and Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist* **55**, 5–14.
- Shavelson, R. J., Hubner, J. J., and Stanton, G. C. (1976). Validation of construct interpretations. *Review of Educational Research* **46**, 407–441.
- Skaalvik, E. M. (1997). Issues in research on self-concept. In Maehr, M. L. and Pintrich, P. R. (eds.) *Advances in Motivation and Achievement*, vol. 10, pp 51–98. Greenwich, CN: JAI Press.
- Valentine, J. C. and DuBois, D. L. (2005). Effects of self-beliefs on academic achievement and vice-versa: Separating the chicken from the egg. In Marsh, H. W., Craven, R. G., and McInerney, D. M. (eds.)

- International Advances in Self Research*, vol. 2, pp 53–78. Greenwich, CT: Information Age.
- Valentine, J. C., DuBois, D. L., and Cooper, H. (2004). The relations between self-beliefs and academic achievement: A systematic review. *Educational Psychologist* **39**, 111–133.
- Wells, L. E. and Marwell, G. (1976). *Self-Esteem: Its Conceptualization and Measurement*. Beverly Hills, CA: Sage.
- Wigfield, A. and Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In Wigfield, A. and Eccles, J. S. (eds.) *Development of Achievement Motivation*, pp 173–195. San Diego, CA: Academic Press.
- Wigfield, A. and Karpachian, M. (1991). Who am I what can I do: Children's self-concepts and motivation in achievement situations. *Educational Psychologist* **25**, 233–261.
- Wylie, R. C. (1974). *The Self-Concept*, rev. edn., vol. 1. Lincoln: University of Nebraska Press.
- Wylie, R. C. (1979). *The Self-Concept* vol. 2. Lincoln: University of Nebraska Press.
- Wylie, R. C. (1989). *Measures of Self-Concept*. Lincoln: University of Nebraska Press.
- Psychology: The Journal of the Hellenic Psychological Society* **9**, 459–493.
- Marsh, H. W. and Craven, R. (1997). Academic self-concept: Beyond the dustbowl. *Handbook of Classroom Assessment: Learning, Achievement, and Adjustment*, pp 131–198. Orlando, FL: Academic Press.
- Marsh, H. W. and Hau, K. T. (2003). Big-fish-little-pond effect on academic self-concept: A cross-cultural (26 country) test of the negative effects of academically selective schools. *American Psychologist* **58**, 364–376.
- Robbins, S. B., Lauver, K., Le, H., et al. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin* **130**(2), 261–288.
- Skaalvik, E. M. and Valas, H. (1999). Relations among achievement, self-concept, and motivation in mathematics and language arts: A longitudinal study. *Journal of Experimental Education* **67**, 135–149.
- Swann, W. B., Chang-Schneider, C., and McClarty, K. L. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. *American Psychologist* **62**, 84–94.
- Trautwein, U., Lüdtke, O., Köller, O., and Baumert, J. (2006). Self-esteem, academic self-concept, and achievement: How the learning environment moderates the dynamics of self-concept. *Journal of Personality and Social Psychology* **90**, 334–349.
- Yeung, A. S., Chui, H. S., Lau, I. C. Y., et al. (2000). Where is the hierarchy of academic self-concept? *Journal of Educational Psychology* **92**, 556–567.

Further Reading

- Aronson, J. (2002). *Improving Academic Achievement: Impact of Psychological Factors on Education*. San Francisco, CA: Elsevier.
- Lawrence, D. (2006). *Enhancing Self-Esteem in the Classroom*. London: Paul Chapman.
- Marsh, H. W. (1986). Global self esteem: Its relation to specific facets of self-concept and their importance. *Journal of Personality and Social Psychology* **51**, 1224–1236.
- Marsh, H. W. (1995). A Jamesian model of self-investment and self-esteem – comment. *Journal of Personality and Social Psychology* **69**, 1151–1160.
- Marsh, H. W. (2002). A multidimensional physical self-concept: A construct validity approach to theory, measurement, and research.

Relevant Websites

- <http://www.apa.org> – American Psychological Association.
- <http://www.eric.ed.gov> – Education Resources Information Center.
- <http://www.ericdigests.org> – ERICDigests.org.
- <http://self.uws.edu.au> – University of Western Sydney.

Self-Efficacy Beliefs

D H Schunk, The University of North Carolina at Greensboro, Greensboro, NC, USA
F Pajares[†]

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Outcome expectations – Beliefs about the expected outcomes of actions.

Reciprocal determinism – Interacting influences of cognitions, behaviors, and environmental variables.

Self-efficacy – Perceived capabilities to learn or perform behaviors at designated levels.

Social cognitive theory – The theory of behavior emphasizing cognitive, vicarious, self-regulatory, and self-reflective processes in human adaptation and change.

Overview of Social Cognitive Theory

With the publication of *Social Foundations of Thought and Action: A Social Cognitive Theory* in 1986, Albert Bandura advanced a view of human functioning that accords a central role to cognitive, vicarious, self-regulatory, and self-reflective processes in human adaptation and change (Bandura, 1986). From this social cognitive perspective, human thought and action are viewed as products of a dynamic interplay among personal, behavioral, and environmental influences. How people interpret the results of their own behaviors informs and alters their environments and the personal factors they possess that, in turn, inform and alter subsequent behaviors. The view that (1) personal factors in the form of cognition, affect, and biological events; (2) behaviors; and (3) environmental influences create interactions that result in a triadic reciprocity is the foundation of Bandura's conception of reciprocal determinism.

The reciprocal nature of the determinants of human functioning in social cognitive theory makes it possible for therapeutic and counseling efforts to be directed at personal, environmental, or behavioral factors. Strategies for increasing well-being can be aimed at improving emotional, cognitive, or motivational processes; increasing behavioral competencies; or altering the social conditions under which people live and work. In school, for example, teachers face the challenge of improving the academic learning and confidence of their students. Using social cognitive theory as a framework, teachers can work to

improve their students' emotional states, correct their faulty self-beliefs and habits of thinking (personal factors), improve their academic skills and self-regulatory practices (behaviors), and alter the school and classroom structures that may work to undermine student success (environmental factors).

Social cognitive theory is rooted in a view of human agency in which individuals are proactively engaged in their own development and can make things happen by their actions. Individuals are imbued with certain capabilities that define what it is to be human. Primary among these are the capabilities to symbolize, plan alternative strategies (forethought), learn through vicarious experience, self-regulate, and self-reflect. For Bandura, however, the capability that is most distinctly human is that of self-reflection; hence, it is a prominent feature of social cognitive theory. Through self-reflection, people make sense of their experiences, explore their own cognitions and self-beliefs, engage in self-evaluation, and alter their thinking and behavior accordingly.

Self-Efficacy Beliefs

Of all the thoughts that affect human functioning, and standing at the very core of social cognitive theory, are self-efficacy beliefs, or individuals' judgments of their capabilities to learn or perform courses of action at designated levels. In essence, self-efficacy beliefs are the self-perceptions that individuals hold about their capabilities. These beliefs of personal competence provide the foundation for human motivation, well-being, and personal accomplishment. This is because unless people believe that their actions can produce the outcomes they desire, they have little incentive to act or to persevere in the face of difficulties.

Self-efficacy should not be confused with constructs such as self-concept or self-esteem, which are broad evaluations of one's self, complete with judgments of self-worth that accompany such evaluations. Self-efficacy beliefs revolve around questions of can (Can I write this essay? Can I solve this problem?), whereas self-concept/self-esteem beliefs reflect questions of feel (Do I like myself? How do I feel about myself as a writer?). Moreover, one's beliefs about what one can do may bear little relation to how one feels about oneself. Many bright students are able to engage their academic tasks with

[†]Deceased.

strong self-efficacy even while their academic skills are a source of low self-esteem, having been labeled by their classmates as nerds or geeks.

Research examining the empirical properties of these two constructs has shown that they differ in important ways. Self-efficacy beliefs are cognitive, goal-referenced, relatively context-specific, and future-oriented judgments of competence that are relatively malleable due to their task dependence. Self-concept beliefs, on the other hand, are primarily affective, heavily normative, typically aggregated, hierarchically structured, and past-oriented self-perceptions that are relatively stable due to their sense of generality. Self-efficacy acts as an active precursor of self-concept development.

People's self-efficacy beliefs should not be confused with their judgments of the consequences that their behavior will produce, or outcome expectations. Typically, self-efficacy beliefs help determine the outcomes one expects. Confident individuals anticipate successful outcomes. Students confident in their academic skills expect high marks on examinations and expect the quality of their work to reap personal and professional benefits. Conversely, students who lack confidence in their academic skills envision a low grade before they begin an examination or enroll in a course. The expected results of these imagined performances will be differently envisioned: greater academic success and subsequent career options for the former and curtailed academic possibilities for the latter. However, self-efficacy beliefs also can be inconsistent with the outcomes one expects. A high sense of efficacy may not result in behavior consistent with that belief, for example, if the individual also believes that the outcome of engaging in that behavior will have undesired effects. A student highly self-efficacious in his/her academic capabilities may elect not to apply to a particular university whose entrance requirements are such as to discourage all but the hardest souls.

The notion of perceived control also differs from self-efficacy. People who believe they can control what they learn and perform are more apt to initiate and sustain behaviors directed toward those ends than are individuals who hold a low sense of control over their capabilities. Perceived control is generic; thus, it is meaningful to speak of this control over learning or performing and over outcomes. Further, it is only one aspect of self-efficacy. Other factors that influence self-efficacy include perceptions of ability, social comparisons, attributions, time available, and perceived importance. People may believe that they can control their use of learning strategies, effort, and persistence, yet hold a low sense of self-efficacy for learning because they feel that the learning is unimportant and not worth the investment of time.

Since individuals operate collectively as well as individually, self-efficacy is both a personal and a social construct. Collective systems develop a sense of collective

efficacy – a group's shared belief in its capability to attain goals and accomplish desired tasks. For example, schools develop collective beliefs about the capability of their students to learn, of their teachers to teach and otherwise enhance the lives of their students, and of their administrators and policymakers to create environments conducive to these tasks. Organizations with a strong sense of collective efficacy exercise empowering and vitalizing influences on their constituents.

Sources of Self-Efficacy Beliefs

Individuals form their self-efficacy beliefs by interpreting information primarily from four sources. The most influential source is the interpreted result of one's previous performances, or mastery experiences. The process of forming self-efficacy beliefs from mastery experiences is simple and intuitive: individuals engage in tasks and activities, interpret the results of their actions, use the interpretations to develop beliefs about their capability to engage in subsequent tasks or activities, and act in concert with the beliefs created. Outcomes interpreted as successful raise self-efficacy; those interpreted as failures lower it.

In addition to interpreting the results of their actions, people form their self-efficacy beliefs through the vicarious experiences of observing others perform tasks. This form of efficacy information is particularly powerful when people observe models who they believe possess similar capability as themselves. Observing the successes of such models contributes to the observers' beliefs about their own capabilities (If they can do it, so can I!). Conversely, watching models with perceived similar capabilities fail can undermine the observers' beliefs about their own capabilities to succeed. When people perceive the model's capabilities as highly divergent from their own, the influence of vicarious experiences is greatly minimized. It is noteworthy that people seek out models with qualities they admire and capabilities to which they aspire. Significant models in one's life can help instill self-beliefs that will influence the course and direction that life will take.

Individuals also create and develop self-efficacy beliefs as a result of the social persuasions they receive from others. These persuasions can involve exposure to the verbal judgments that others provide. Persuaders play an important part in the development of an individual's self-beliefs. However, social persuasions should not be confused with knee-jerk praise or empty inspirational homilies. Effective persuaders must cultivate people's beliefs in their capabilities while simultaneously ensuring that the envisioned success is attainable. Just as positive persuasions may work to encourage and empower self-efficacy beliefs, negative persuasions can work to defeat and weaken the same. It is usually easier to weaken self-efficacy beliefs

through negative appraisals than to strengthen such beliefs through positive encouragement.

Physiological and emotional states such as anxiety, stress, arousal, and mood states also provide information about efficacy beliefs. People can gauge their degree of confidence by the emotional state they experience as they contemplate an action. Strong emotional reactions to a task provide cues about the anticipated success or failure of the outcome. When they experience negative thoughts and fears about their capabilities, these affective reactions can lower self-efficacy perceptions and trigger additional stress and agitation that help ensure the inadequate performance they fear. One way to raise self-efficacy beliefs is to improve physical and emotional well-being and reduce negative emotional states. As individuals have the capability to alter their own thinking and feeling, enhanced self-efficacy beliefs can, in turn, powerfully influence the physiological states themselves.

The sources of self-efficacy information are not directly translated into judgments of competence. Individuals interpret the results of events, and these interpretations provide the information on which judgments are based. The types of information people attend to and use to make efficacy judgments, and the rules they employ for weighting and integrating them, form the basis for such interpretations. The selection, integration, interpretation, and recollection of information influence judgments of self-efficacy.

Effects of Self-Efficacy Beliefs

Self-efficacy beliefs enhance human accomplishment and well-being in countless ways. They influence the choices people make and the courses of action they pursue. Individuals tend to select tasks and activities in which they feel competent and confident and avoid those in which they do not. Unless people believe that their actions will have the desired consequences, they have little incentive to engage in them. How far will an interest in architecture take a student who feels hopeless in geometry? Irrespective of the factors that operate to influence behavior, they are rooted in the core belief that one has the capability to accomplish that behavior.

Self-efficacy beliefs also help determine how much effort people will expend on an activity, how long they will persevere when confronting obstacles, and how resilient they will be in the face of adverse situations. People with a strong sense of personal competence approach difficult tasks as challenges to be mastered rather than as threats to be avoided. They set challenging goals and maintain strong commitment to them, heighten and sustain their efforts in the face of failure, and recover their sense of efficacy more quickly after setbacks. High self-efficacy helps create feelings of serenity in approaching difficult tasks and activities. Conversely, people with low

self-efficacy may believe that things are tougher than they really are, a belief that fosters anxiety, stress, depression, and a narrow vision of how best to solve a problem. As a consequence, self-efficacy beliefs can powerfully influence the level of accomplishment that one ultimately achieves. This function of self-beliefs also can create the type of self-fulfilling prophecy in which one accomplishes what one believes one can accomplish; that is, the perseverance associated with high self-efficacy is likely to lead to increased performance, which, in turn, raises one's sense of efficacy and spirit, whereas the giving in associated with low self-efficacy helps ensure the very failure that further lowers confidence and morale.

Of course, human functioning is influenced by many factors. The successes and failures that people experience as they engage in the myriad tasks that comprise their life naturally influence the many decisions they must make. In addition, the knowledge and skills they possess play an important role in what they choose to do. However, because past attainments, knowledge, and skills are always interpreted by the individual, it is the interpretations that form the foundation for the beliefs that are developed about subsequent capabilities. As a consequence, people's accomplishments are generally better predicted by their self-efficacy beliefs than by their previous attainments, knowledge, or skills.

Development of Self-Efficacy Beliefs

The first influences on an individual's self-efficacy take place within the family. Parents and caregivers provide their children with the first experiences that differentially influence self-efficacy beliefs. These home influences that help children interact effectively with the environment positively affect self-efficacy. When the home environment is rich in activities and materials that arouse the children's curiosity and offer challenges that can be met, children are motivated to work on the activities and learn new information and skills.

Parents who provide a warm, responsive, and supportive home environment, encourage exploration and stimulate curiosity, and provide play and learning materials, accelerate their children's intellectual development. Parents also are the key providers of self-efficacy information. Since mastery experience is the most powerful source of self-efficacy information, the parents who arrange for their children to experience varied mastery experiences develop more efficacious youngsters than do parents who arrange fewer opportunities. Such experiences occur in homes enriched with activities and in which children have the freedom to explore. With respect to vicarious experiences, parents who teach children diverse ways to cope with difficulties and model persistence and effort strengthen their children's self-efficacy. Family members are also

prime sources of persuasive information. Parents who encourage their children to try different activities and support and encourage their efforts help to develop children who feel more capable of meeting challenges.

As children grow, peers become increasingly important. Parents who steer their children toward efficacious peers provide further vicarious boosts in self-efficacy. Peers themselves influence children's self-efficacy in various ways. One means is through model similarity. Observing others, who are similar to the observers, succeed can raise the self-efficacy of the observers and motivate them to perform the task if they believe that they, too, will be successful. Conversely, observing others fail can lead students to believe that they lack the competence to succeed, and dissuade them from attempting the task.

Self-efficacy beliefs tend to decline as students advance through school. There are several reasons for this, including greater emphasis on competition, more norm-referenced grading, less teacher attention to individual student progress, and the stresses associated with school transitions. These and other school practices can weaken academic self-efficacy, especially among students who are less academically prepared to cope with increasingly challenging academic tasks. Lock-step sequences of instruction frustrate some students who fail to grasp skills and increasingly fall behind their peers. Ability groupings can lower self-efficacy among those relegated to lower groups. Classrooms that allow for much social comparison tend to lower the self-efficacy of students who find their performances inferior to those of their peers.

Periods of transition in schooling also bring other factors that affect self-efficacy into play. As elementary students remain with the same teacher and peers for most of the school day, children receive much attention, and individual progress is stressed. Typically, however, several elementary schools feed into the same middle school, and children begin to move from class to class for specific subjects. Thus, middle school students become exposed to peers whom they do not know. Evaluation becomes normative, and there is less teacher attention to individual progress. The widely expanded social reference group, coupled with the shift in evaluation standards, requires that students reassess their academic abilities. As a consequence, perceptions of academic competence typically begin to decline during middle school.

In self-efficacy research, it is not uncommon for children to report overconfidence about accomplishing difficult tasks. Even being provided with feedback indicating low performance may not decrease self-efficacy. The incongruence between children's self-efficacy and their actual performance may be due to various causes. Children often lack task familiarity and do not fully understand what is required to execute a task successfully. As they gain experience, their accuracy improves. Children may also be unduly swayed by certain task features and decide based on these that they can

or cannot perform the task while ignoring other features. In subtraction, for example, children may focus on how many numbers the problems contain and judge longer problems to be more difficult than those with fewer numbers, even when the longer ones are conceptually simpler. As their capability to focus on multiple features improves, so does their accuracy.

Children sometimes do not know what they are capable of accomplishing. In writing, for example, it is difficult for them to know how clearly they can express themselves or whether their writing skills are improving. A teacher's feedback – especially at the elementary level – is intended to encourage and stress what children do well. Children may believe they can write well when, in fact, their writing is below normal for their grade level. As they develop, children gain task experience and engage more often in peer social comparisons, which improve the accuracy of their self-assessments. The correspondence between self-efficacy and performance also can be increased by providing children with instruction and opportunities to practice self-evaluation and with instructional interventions that convey clear information about children's skills or progress.

Self-Efficacy, Motivation, and Academic Achievement

Bandura situated self-efficacy within a social cognitive theory of personal and collective agency that operates in concert with other sociocognitive factors in regulating human well-being and attainment. He also addressed the major facets of agency – the nature and structure of self-efficacy beliefs, their origins and effects, the processes through which such self-beliefs operate, and the modes by which they can be created and strengthened. Bandura has reviewed a vast body of research on each of these aspects of agency in diverse applications of the theory. Researchers have demonstrated that the self-efficacy beliefs of individuals powerfully influence their attainments in diverse fields.

A search for the term self-efficacy in most academic databases reveals that, at the start of 2007, nearly 4 000 articles have been written on this important psychological construct. In a Google Internet search, the term returned over 1 million webpages. Self-efficacy has generated research in areas as diverse as life-course development, education, business, athletics, medicine and health, media studies, social and political change, moral development, psychology, psychiatry, psychopathology, and international affairs. Self-efficacy has been especially prominent in studies of educational outcomes such as academic achievement, attributions of success and failure, goal setting, social comparisons, memory, problem solving, career development, and teaching and teacher education. Researchers have established that self-efficacy beliefs and behavior outcomes

are highly correlated and that self-efficacy is an excellent predictor of academic motivation and performance.

Researchers also have demonstrated that self-efficacy beliefs mediate the effect of skills, previous experience, mental ability, or other motivation constructs on subsequent achievement, which implies that they act as a filter between prior determinants and subsequent accomplishments. Meta-analyses have found that the average weighted correlation between self-efficacy and work-related performance was $(G)r = 0.38$, which transforms to an impressive 28% gain in task performance. In education, a meta-analysis of studies published between 1977 and 1988 revealed that efficacy beliefs were positively related to academic achievement. Self-efficacy beliefs were related to academic outcomes ($r_u = 0.38$) and accounted for approximately 14% of the variance. Correlations between self-efficacy and academic performances in investigations, in which self-efficacy is analyzed at the item- or task-specific level and corresponds to the criterial task, have ranged from 0.49 to 0.70; direct effects in path analytic studies have ranged from $\beta = 0.349$ to 0.545. Self-efficacy beliefs have also been found highly predictive of college students' selection of majors and career choices. Variables such as perceived control, outcome expectations, value, attributions, and self-concept are types of cues used by individuals to assess their efficacy beliefs.

In general, empirical evidence amply supports Bandura's contention that self-efficacy beliefs touch virtually every aspect of people's lives – whether they think productively, self-debilitatingly, pessimistically, or optimistically; how well they motivate themselves in the face of adversities; their vulnerability to stress and depression; and the life choices they make. Self-efficacy is also a critical determinant of how individuals regulate their own thinking and behavior. Particularly in psychology and education, self-efficacy has proven to be a more consistent predictor of behavioral outcomes than have other motivational variables.

Summary

Bandura's social cognitive theory of human functioning emphasizes the critical role of self-beliefs in human cognition, motivation, and behavior. Social cognitive theory gives prominence to a self-system that enables individuals to exercise a measure of control over their thoughts, feelings, and actions. In putting forth this view, Bandura reinvigorated the nearly abandoned focus on the self in the study of human processes that William James initiated nearly a century earlier. Social cognitive theory is an agentic and empowering psychological perspective in which individuals are proactive and self-regulating rather than reactive and controlled either by environmental or biological forces. Instead, the beliefs that people have about themselves are key elements in their

exercise of control and of personal, cultural, and social achievement. It is because of their beliefs about their own capabilities – their self-efficacy beliefs – that people are able to exercise the self-influence required to contribute to the types of persons they become and their achievements.

See also: Anxiety; Attribution Theory; Motivating Students in Classrooms; Peer Learning in the Classroom; Self-Regulated Learning and Socio-Cognitive Theory.

Bibliography

- Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York: Freeman.
- Bandura, A., Barbaranelli, C., Caprara, G. V., and Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. *Child Development* **67**, 1206–1222.
- Bong, M. and Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review* **15**, 1–40.
- Hackett, G. (1995). Self-efficacy in career choice and development. In Bandura, A. (ed.) *Self-Efficacy in Changing Societies*, pp 232–258. New York: Cambridge University Press.
- Meece, J. L. (1997). *Child and Adolescent Development for Educators*. New York: McGraw-Hill.
- Multon, K. D., Brown, S. D., and Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology* **38**, 30–38.
- Pajares, F. (1996). Self-efficacy beliefs in achievement settings. *Review of Educational Research* **66**, 543–578.
- Pajares, F. (1997). Current directions in self-efficacy research. In Maehr, M. and Pintrich, P. R. (eds.) *Advances in Motivation and Achievement*, vol. 10, pp 1–49. Greenwich, CT: JAI Press.
- Pajares, F. and Schunk, D. H. (2002). Self and self-belief in psychology and education: A historical perspective. In Aronson, J. (ed.) *Improving Academic Achievement: Impact of Psychological Factors on Education*, pp 3–21. San Diego, CA: Academic Press.
- Pajares, F. and Urdan, T. (eds.) (2006). *Adolescence and Education, Vol. 5: Self-Efficacy Beliefs of Adolescents*. Greenwich, CT: Information Age.
- Pintrich, P. R. and Schunk, D. H. (2002). *Motivation in Education: Theory, Research, and Applications*, 2nd edn. Upper Saddle River, NJ: Merrill/Prentice-Hall.
- Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research* **57**, 149–174.
- Schunk, D. H. (1995). Self-efficacy and education and instruction. In Maddux, J. E. (ed.) *Self-Efficacy, Adaptation, and Adjustment: Theory, Research, and Application*, pp 281–303. New York: Plenum.
- Schunk, D. H. and Pajares, F. (2002). The development of academic self-efficacy. In Wigfield, A. and Eccles, J. (eds.) *Development of Achievement Motivation*, pp 15–31. San Diego, CA: Academic Press.
- Stajkovic, A. D. and Luthans, F. (1998). Self-efficacy and work-related performances: A meta-analysis. *Psychological Bulletin* **124**, 240–261.
- Wigfield, A., Eccles, J. S., and Pintrich, P. R. (1996). Development between the ages of 11 and 25. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 148–185. New York: Simon and Schuster Macmillan.
- Zimmerman, B. J. (1995). Self-efficacy and educational development. In Bandura, A. (ed.) *Self-Efficacy in Changing Societies*, pp 202–231. New York: Cambridge University Press.

Relevant Website

<http://des.emory.edu> – Self-Efficacy: A Community of Scholars.

Social and Emotional Outcomes of Learning

J Vorhaus, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In this article, evidence on the social and emotional outcomes of learning, causal and other explanations for their development, and the frameworks in which explanations and outcomes are located, are reviewed.

The relationships between learning and outcomes in health, well-being, citizenship, parenting, and criminality are explored. Attention is also paid to social relations, and these relationships over the life-course and across generations are explored.

This study is not a systematic review. Three sources of evidence are treated with particular interest: (1) Heckman and colleagues on cognitive and non-cognitive development over the life-course; (2) Eccles and Roeser on schools and adolescent development; and (3) Wilczenski and colleagues on social and emotional learning outcomes. We also draw heavily on the evidence the author and his colleagues at the Centre for Research on the Wider Benefits of Learning (WBL) collected and reviewed over the last 7 years. The first three sources are given priority: first, because the authors have themselves conducted substantial reviews of the literature; second, because each has made a significant contribution to the field; and third, because together they represent a variety of disciplines and methods – including interdisciplinary methods – that are necessary to call upon to produce a sound theoretical framework for understanding the relationship between learning and its social and emotional effects.

This article is organized as follows: we first deal with explanatory frameworks. These provide the context for the presentation of evidence in sections that follow. Priority is given to the drivers of the outcomes for individuals from and within the experience of learning, categorized according to whether their primary source lies in compulsory or post-compulsory schooling.

A review of this kind is bound to be strictly circumscribed: we say nothing about the closely related subject of economic returns; in the main we consider social and emotional variables as outcomes and not determinants of learning, and the range of acknowledged learning activities is largely confined to formal environments.

Several omissions deserve notice. Description of method is kept to a minimum; there is no discussion of pre-school or out-of-school programs; and several prominent lines of enquiry are barely touched upon. For example, Heckman *et al.* (2004) present evidence that both

cognitive and non-cognitive skills affect the returns to schooling: increasing non-cognitive ability over the same decile range as cognitive ability has a greater effect on many outcomes than increasing cognitive ability over the same range. This applies, for example, to a reduced likelihood of smoking, spending time in jail by the age of 30, and nonmarital pregnancy outcomes (Cunha: 19). But the literature on the returns to enhancing non-cognitive skills cannot be surveyed.

Many other issues are also omitted, including race, learning difficulties, and bilingualism. In these cases also, there is a large literature attached to each, and that is more than this article can begin to do justice to. That should not be taken to imply that outcomes of these kinds are any fewer or matter any less than those that are explored.

A Framework for Analysis

Complementarity, Self-Productivity and the Skill Multiplier

Cunha *et al.* (2005) consider the life cycle as the proper context in which to interpret evidence on the development of outcomes. As against the influential work of Becker and Tomes (1979), childhood is conceived as a multi-stage process during which early investments feed into later investments.

Over the life-course, genetic, environmental, and personal inputs are variously productive of outputs, in the form of achieved levels of skill. Points of greatest productivity are sensitive periods, and these become critical if one stage alone is productive of a given skill or ability. Skills produced at one point augment skills achieved later on (self-productivity), and skills produced at one stage raise the productivity of investment at subsequent stages (complementarity). Levels of skill at different ages bolster one other, producing a skill multiplier.

This framework explains significant findings from the literature: that skill begets skill and abilities beget abilities (*ibid.*: 5), and that the formation of skills is a dynamic process, whereby those formed at one stage in the life cycle affect the productivity of learning in the next stage (*ibid.*: 12).

Multi-Level Contexts

The work of Heckman and colleagues illuminates research by Eccles and the WBL, whose two models for

conceiving and analyzing learning outcomes are mutually complementary.

Eccles and Roeser examine how schools promote academic, social, and psychological competence. Following Bronfenbrenner, schools are conceived as comprising multi-level contexts; these influence adolescent social-emotional and behavioral development, through organizational, social, and instructional processes, each of which operate at different levels. (Eccles and Roeser, 2003)

The school context is conceived as a “series of hierarchically-ordered, interdependent levels of organization,” beginning at the level of the classroom and moving up in complexity to the school as “an organization system embedded in a larger cultural system” (*ibid.*: 3). There are multiple levels of regulatory processes, usually dynamic, interrelated across levels of analysis, and which develop or change as children move through different school levels. These multi-level processes are interwoven into one another, and relations between different levels of organization in the school can be complementary or contradictory (*ibid.*: 4).

A central point is that the influence of schools on adolescent development is by means of multiple dynamic impacts of different levels of school interacting with each other, rather than the static resources or characteristics of the curriculum, teachers, or school *per se* (*ibid.*). These are not specified in this article (see Eccles and Roeser for a detailed conceptual framework).

The specifics of school-related processes change across the course of children’s and adolescents’ development as they progress through primary, secondary, and tertiary education, and beyond into learning in adulthood. However, throughout development and maturation, interactions between the learner and the learning environment are essential to learning (Eccles and Midgley, 1989; Eccles: 5).

In parallel with the work of Eccles, the WBL has developed a broadly similar multi-level framework for analyzing the wider benefits of learning, recognizing social and emotional outcomes at the level of the individual, family, community, and nation. The model highlights the range of outcomes influenced by education, the multiple, dynamic relationships operating at and between the different levels, and the formative significance of the multiple contexts in which they operate (Preston *et al.*, 2005 and WBL Research Report, No. 11).

A distinguishing feature of the WBL analysis is the concept of social productivity. This concept denotes the capacity of education to support outcomes of social value at each of the levels mentioned above, and to prevent outcomes such as crime, intolerance, mental health problems, disengagement, and social immobility.

Socially valuable outcomes developed in the educational process include skills, capabilities, and social networks. These emerge from interactions in numerous

contexts; they are not simply the product of attendance at educational institutions. Hence, the effects of education will depend on interactions with peers, teachers, and others; on the impact of identity, beliefs, and values; and on the ethos, pedagogy, and curricula comprising the learning environments. The socially productive capacity of learning is evidenced by its impact on variables such as self-concept, self-efficacy, and social capital.

So, for example, Putnam (1993) views education as a valuable source of social capital formation. At the level of primary education, learning can promote societal cohesion and strengthen citizenship when individuals of all socioeconomic backgrounds are enrolled in the public education system. Learning experience can provide opportunities to gain and practice social capital skills, such as participation and reciprocity; to extend and deepen social networks; and to support the development of shared norms and the values of tolerance, understanding, and respect. Learning experience can also provide a forum for community activity and for students to learn how to participate responsibly in their society (Heyneman, 1998; Schuller *et al.*, 2004 and WBL Research Report No. 3).

Organizing the Evidence

In keeping with the framework set out in this section, the evidence in sections 3–7 are considered in relation to the primary levels of analysis – the individual, family, and community – and the primary stages of lifelong learning: childhood, adolescence, and adulthood. No level or stage is intrinsically more important than any other but for reasons of space priority is given to individual-level outcomes for children and adolescents. And while some themes are treated in the context of one level or stage rather than another, this is largely because the evidence reviewed tends to adopt that temporal focus; it is not meant to imply that the themes themselves do not prominently feature at other points in the life-course.

Individual-Level Effects: Childhood and Adolescence

Introduction

The effects treated in this section have been hypothesized as sources of causal impact on social and emotional outcomes. They are individual level effects in the sense that they are construed as the effects of individual agents, operating, for example, as managers, teachers, or deliverers of a curriculum. At the same time it is also recognized that there are alternative legitimate construals, according to which these effects operate at other levels, such as the level of the school or at a level involving other social and organizational agents.

Teacher Expectations

When students are aware of teachers' high expectations, students achieve more, experience higher esteem, and are less involved in problem behavior during childhood and adolescence (Eccles *et al.*, 1993; Rutter, 1983; Roeser *et al.*, 1998; Weinstein, 1989). Teachers who believe they can, reach out to the most difficult students, communicate positive expectations to those students, thereby enhancing the students' confidence in their ability to master academic material (Ashton, 1985; Midgley *et al.*, 1989). When teachers have a low sense of their effectiveness, this can lead to students exhibiting helpless responses to failure in the classroom, and to the development of depressive symptoms (Cole, 1991; Roeser *et al.*, 1999).

Teachers often have varying expectations from students in any one class. Rosenthal (1969) argues that teacher-expectancy effects depend on whether teachers structure activities differently, and interact differently, with high- and low-expectancy students, and on whether these differences are perceived by the students (Brophy, 1985; Parsons *et al.*, 1982; Weinstein, 1989). One focus of research is differential treatment in respect of gender, ethnicity, and social class. Studies report small but consistent undermining effects of low-teacher expectations from girls (mathematics and science), minority groups (all subject areas), and children from low social-class backgrounds (all subject areas). (Eccles and Roeser, 2003; Brophy and Good, 1974; Ferguson, 1998; Jussim *et al.*, 1996; Valencia, 1991).

Classroom Management

Research on classroom management gives evidence of the effects of orderliness and predictability. Student achievement and conduct improves when teachers provide feedback, and establish smoothly run and efficient procedures for monitoring student progress and work completion (Blumenfeld *et al.*, 1983; Eccles *et al.*, 1998; Pintrich and Schunk, 1996).

Boggiano, Deci, and Ryan suggest that overly controlling, autonomy-inhibiting environments undermine intrinsic motivation, ability self-concepts, and self-direction, and, instead, encourage a learned helpless response to difficult tasks. Laboratory and field-based studies provide support for this view (Boggiano *et al.*, 1992; Deci *et al.*, 1981; Grolnick and Ryan, 1987).

While direct instruction has a positive impact on student achievement, research also suggest that students exhibit more stress in didactic as compared with student-centered contexts (Dunn and Kontos, 1997). More open classroom contexts encourage students to be more creative, hold better self-concepts, and more positive attitudes toward teachers and school, and to demonstrate greater independence and curiosity (Dunn and Kontos, 1997; Peterson, 1979; Reynolds, 1975).

The Curriculum

The significance of providing learning materials that are meaningful to students is borne out by the association between low interest in and perceived irrelevance of the school curriculum on the one hand, and poor attention, disengagement, and alienation from school on the other (Jackson and Davis, 2000; Finn, 1989; Larson and Richards, 1989). Curricula that fail to represent the voices and experience of traditionally underrepresented groups can also explain the alienation of some group members from the educational process (Fine, 1991; Sheets and Hollins, 1999; Barton).

Many school programs aim to improve outcomes by supporting students' self-image and self-esteem. But Wilczenski argues that enhancing social and emotional outcomes requires a focus on mastery in the classroom rather than student self-perceptions (Wilczenski *et al.*, 2001: 4). Socioemotional learning cannot be separated from the instructional mission of schools, because social and emotional adjustment mediates academic competence and resilience (Shapiro, 2000).

Peer Collaboration, Tutoring and Mentoring

Collaboration among peers fosters socioemotional development (Slavin, 1980; Sharan, 1980), including better attitudes toward learning, better self-concepts, better attitudes toward others, and better racial relationships (Bossert, 1988/1989; Slavin, 1990). Peer-tutoring programs frequently show socioemotional gains: tutees are described as more cooperative and respectful toward peers and teachers, while also exhibiting higher self-esteem (Gensemer, 2000; Phillips *et al.*, 1994; Roswal *et al.*, 1995).

There is evidence of the effectiveness of mentoring and motivational programs for disadvantaged teenagers, largely stemming from the benefits of fostering motivation and improving non-cognitive skills (Karoly *et al.*, 1998; Currie and Blau, 2005; Heckman, 2000). Studies of Big Brothers/Big Sisters (BB/BS) and Philadelphia Futures sponsor-a-scholar (SAS) programs show positive social effects on school-aged children and adolescents. BB/BS paired unrelated adult volunteers with young people from single-parent households. Tierney and Grossman (1995) found that 18 months after being matched with a mentor, Little Brothers and Sisters (aged 10–16 at the time of the match) were less likely to have used drugs or alcohol, to skip class or a day of school, or to lie to their parents; they were more likely to feel competent in their school work and to report a better relationship with parents (Cunha, 2005).

Ability Grouping

Individuals may be grouped by ability either within a class or between classes. The evidence on socioemotional outcomes appears to favor ability grouping for high ability

but not low-ability students (Fuligni *et al.*, 1995; Gamoran and Mare, 1989; Kulik and Kulik, 1987).

The argument in favor of ability grouping includes the point that students are more likely to learn when material is well matched to their competence. There is evidence consistent with this for children in high-ability classrooms and high within-class ability groups (Dreeben and Barr, 1988; Fuligni *et al.*, 1995; Pallas *et al.*, 1994). Shields (1996) found that, compared to heterogeneously grouped gifted students, homogeneously grouped gifted students showed higher self-concepts, self-acceptance, and independence. Feldhusen and Moon (1992) argue that heterogeneous grouping for gifted students leads to lower motivation and poorer attitudes toward school.

Eder and Felmlee (1983, 1984) explored attention patterns in low- and high-ability first-grade reading groups. They controlled for individual characteristics that included personal attention patterns, and found that students in low-ability groups became inattentive at four times the rate of students in high-ability groups. Eder and Felmlee observed that teachers of high-ability groups quickly managed inattentiveness or disruption whereas teachers of low-ability groups tended to ignore this. Peer effects were observed: high-ability groups applied pressure to maintain attention during interruptions, while low-ability groups tended to use interruptions as an opportunity to distract attention from the current task. Schwartz (1981) observed the behavior in low-ability classes among elementary and junior high school students, and found it to be characterized by challenges to teacher authority, obstruction of academic activities, and misuse of educational resources. These findings indicate that social influences in low-ability groups may overwhelm any instructional advantages to grouping (Wilczenski, 1981: 10).

The Lou *et al.* (1996) meta-analysis of studies of within-class grouping found small but positive effects of placing students in groups; they were more positively disposed toward schools and had higher self-concepts than students in ungrouped classes. However, the same studies show negative socioemotional effects of grouping low-ability students together, in part owing to the low-expectations of teachers and the absence of positive role models. Being placed in between-class low academic groups is also associated with poor attitudes toward school, feelings of incompetence, and problem behaviors (Oakes *et al.*, 1992), effects often accounted for by inferior educational experience (Dreeben and Barr, 1988; Pallas *et al.*, 1994; Vanfossen *et al.*, 1987).

School-Level Effects: Childhood and Adolescence

Social Climate

The academic focus of a school can affect mental health. A series of studies found that a belief that school is

ability-focused can lead to a decline in students' self-esteem, and an increase in anger, depressive symptoms, and school truancy when moving from seventh to eighth grade (Roeser and Eccles, 1998; Roeser *et al.*, 1998). An emphasis on ability can alienate many of those students unable to perform at the highest levels, leading to anxiety, anger, and disenchantment (Eccles and Midgley, 1989; Finn, 1989). Schools emphasizing effort, improvement, and the expectation that all students can learn, appear to include more adolescents in the learning process, while also reducing depression and the anxiety that achievement settings can give rise to (Eccles and Roeser, 2003: 18).

Figueira-McDonough (1986) compared two high schools that were similar with respect to intake and achievement rates but different with respect to academic orientation and rates of delinquent behavior. The school putting more emphasis on competition and high grades had higher delinquency rates and delinquent behavior was predicted by low grades. For the school with more diverse goals, taking a greater interest in students' non-academic needs, school attachment was greater on average, with higher delinquency rates associated with low-levels of attachment. The emphasis on motivation and diversity may have promoted an attachment to school, thereby discouraging delinquency.

Transitions

There is evidence that academic motivation and achievement decline during early adolescence (Eccles, 1994; Eccles and Midgley, 1989; Eccles *et al.*, 1993; Maehr and Midgley, 1996), often coinciding with the transition into middle school or junior high school. There is a decline in some early adolescents' school grades as they move into junior high school (Simmons and Blyth, 1987), and similar declines occur for interest in school (Epstein and McPartland, 1976), intrinsic motivation (Harter, 1981), and self-concepts (Eccles *et al.*, 1989; Wigfield *et al.*, 1991). There are also increases in test anxiety (Wigfield and Eccles, 1989) and both truancy and school dropout (Rosenbaum, 1976). Eccles and Midgley (1989) suggest that junior high schools typically lack educational environments that are developmentally appropriate for early adolescents.

Teaching practices

Researchers observe the combination of increasing maturity of high school students and greater teacher emphasis on control and discipline, providing fewer opportunities for student decision making as compared with elementary school teachers (e.g., Brophy and Evertson, 1978; Midgley and Feldlaufer, 1987; Moos, 1979). The evidence confirms what stage-environment fit theory suggests: the discrepancy between the wish for autonomy and control, and the actual opportunities for students in the classroom, leads to

a decline in motivation and interest in school (Mac Iver and Reuman, 1988).

Students entering junior high school face more in the way of whole-class task organization and between-classroom ability grouping (Eccles and Midgley, 1989; Oakes *et al.*, 1992). This is likely to increase social comparison and competitiveness (see Eccles *et al.*, 1984; Rosenholtz and Simpson, 1984), while the use of grading and public assessment has a negative affect on early adolescents' self-perceptions and motivation.

Classroom practices can be distinguished by the priority given to either task mastery or to performance goals. Teachers and learners testify that, as adolescents move from elementary to middle school, the school environment is increasingly focused on performance – competition, relative ability, and social comparison (Midgley *et al.*, 1995; Roeser *et al.*, 1994). How far teachers were task-focused at both levels predicted students' and teachers' sense of personal efficacy, this being lower among middle school participants than among elementary school participants.

Anderman *et al.* (1999) compared two groups of young adolescents: the first moved into a middle school emphasizing task-focused instructional practices, the second into a middle school emphasizing performance-focused instructional practices. The two groups of students were found to differ in their motivational goals following the transition, although these had not differed beforehand; the adolescents moving into the first school were less likely to show an increase in extrinsic motivational and performance-oriented motivational goals.

Eccles and Midgley argue that changes in school environment are often harmful for young adolescents: "the nature of these environmental changes, coupled with the normal course of individual development, is likely to result in a developmental mismatch so that the 'fit' between the early adolescent and the classroom environment is particularly poor, increasing the risk of negative motivational outcomes, especially for adolescents who are having difficulty succeeding in school academically" (Eccles and Midgley, 1989).

Individual-Level Effects: Adult Learning

Introduction

The term adult learning refers to the learning that follows the years of compulsory schooling. In the United Kingdom this is referred to as post compulsory or post-16 learning, and includes, among other contexts, higher education, further education, the workplace, and adult and community learning.

Health, Well-Being, and Social Inclusion

Analysis by the WBL (WBL Research Report No. 6) of the 1958 National Child Development Study (NCDS)

and the 1970 British Cohort Study (BCS70) suggests considerable health advantages associated with having higher-level qualifications. For example:

- Graduates are less likely to smoke: those educated to level 2 or below are 75% more likely to be smokers at age 30, compared to similar individuals educated to degree level or higher.
- Those educated to degree level or higher are between 70% and 80% more likely to report excellent health, compared to a similar individual educated to level 2 or below.
- Graduates are between 35% (women) and 55% (men) less likely to suffer from depression, compared to a similar individual educated to level 2 or below.

The WBL also found in the British Household Panel Study (BHPS), a strong relationship between educational success and progress at 16 and uptake by women of recommended cervical smear tests (WBL Research Report No. 12).

Civic engagement

Feinstein and Hammond (2004) and WBL Research Report No. 8) used the 1958 cohort to examine the contribution of adult learning to a wide range of health and social-capital outcomes. Analysis focused on changes between the ages of 33 and 42 in life outcomes for adults, controlling for their development and context up to age 33.

It was found that participation in adult learning has positive effects on a wide spectrum of health and social-capital outcomes with statistically significant effects of participation on nine of the 12 outcomes looked at. These outcomes are changes in smoking, exercise taken, life satisfaction, race tolerance, authoritarian attitudes, political cynicism, political interest, number of memberships, and voting behavior.

Research on race tolerance combined with authoritarianism (Preston *et al.*, 2005 and WBL Research Report No. 11) found that adult education may be important in sustaining non-extremist views, but it does not appear to be associated with a transformation away from extremist positions.

School-age signals

The WBL has explored how far-emerging aspects of child development have implications for their long-term health and well-being as adults. Using the 1958 cohort, eight proxy measures were used for failure to flourish at secondary school. These encompass attainment, attendance, social adjustment, and attitudes. The adult variables covered aspects of well-being and mental health and some physical conditions and health behaviors. The WBL estimated the associations between failure to flourish at secondary school and health and well-being at age 33,

controlling for social, psychological, and economic factors up to the age of 7.

Those who flourished at school had better outcomes in all main aspects of well-being than those who did not flourish. The magnitudes of adjusted associations were substantial and apply to well-being, mental and physical health, and health behaviors.

Both attainment and engagement at secondary school were markers of adult health and well-being. Those with poor attainment but who were quite engaged at school had relatively poor levels of health and well-being in adulthood compared to cohort members who had attained well at school, but those with poor attainment and poor engagement had worse adult health and well-being than either group.

Literacy and numeracy

The National Research and Development Centre for Adult Literacy and Numeracy (NRDC) analyzed literacy and numeracy data from the 2004 survey of 9664 members of BCS70 (NRDC reports: *New light on literacy and numeracy*; *Illuminating Disadvantage* 2007).

Substantial differences in life chances, quality of life, and social inclusion were evident between individuals at or below entry level 2 in the English National Qualification Framework compared with others at higher levels of literacy and numeracy. Entry level 2 skills were associated with poor health prospects and lack of social and political participation. Gender differences were marked in some of these relationships including the tendency for men with poor skills to lead a solitary life without children in their mid-30s. In contrast, women with the same levels of skills were also more likely to be without a partner, but more typically were parents and often with large families.

At age 16, men and women with entry level skills were the most likely to be alienated from school and to have parents with low aspirations for them. 20% of women with entry level 2 literacy and 10% with entry level 2 numeracy had experienced a spell of homelessness compared with 6% women at level 1 or higher. Women with entry level 2 skills were more than twice as likely as women with level 1 or higher skills to have been a teenage mother (18% to 8%) and three times more likely to have four or more children by age 34 (11%–3%).

Self-efficacy and confidence

Hammond and Feinstein (LRE) investigated links between participation in adult learning and self-efficacy, particularly for adults with low levels of achievement at school. Analyses of data from NCDS, controlling for a range of features of the individual and their context up to the age of 33, found an association between taking courses and transformations from low to good levels of self-efficacy between ages 33 and 42 for all cohort

members. The association is greatest for those who had low achievement levels at school.

This correlational evidence does not prove that adult learning causes these positive changes; there remains the need to identify how far adults who are motivated to participate in adult education are already more likely to have higher levels of confidence and self-efficacy.

Family-Level Outcomes

Intergenerational Effects

The WBL reviewed the evidence on the role of parental education in child development (WBL Research Report, No. 10) and hypothesized that parental education is transmitted inter-generationally through six pathways:

1. by impacting on key distal factors such as income and poverty;
2. by moderating the effect of each distal factor, acting protectively and providing resilience in the family so that they can better deal with difficulties such as poverty;
3. by impacting on the characteristics of contexts, such as parents' cognitions or their mental health and well-being;
4. by supporting individuals and families in managing a set of characteristics and hence moderating their effects;
5. by impacting on proximal processes such as learning behaviors in the home; and
6. by moderating the effects of proximal processes.

The WBL found strong support for the view that education influences most of the factors that have been found to affect children's attainments. Besides having a direct influence on most of the characteristics and parent-child interactions, parental education can also moderate the effects of risk factors and ease the effect of them on interactions between parents and children.

Post-16 Participation and Benefits for the Next Generation

Although, there is considerable evidence to suggest that parenting is an important mediator, WBL research found that the association of educational participation and parenting skills and educational behaviors does not appear to be causal.

Using data from over 11 000 members of the 1958 British Cohort Study (WBL Research Report) the WBL estimated the causal effects of education on parenting skills. In the analysis that did not take account of the possible endogeneity of education, there was evidence of a clear gradient between educational background and elements of parenting. However, this relation between

education and parenting disappeared under a more robust causal instrumental variable specification. The WBL concluded that ability and positional ambition in society could be strong confounders of the relation between staying on in post-compulsory education and subsequent parenting skills.

WBL (Research Report) investigated the relationship between a mother's education and her parenting using data from the child supplement of the 1958 NCDS. By considering data across generations, the dataset allows for an estimation of the size of the bias in the relationship between education and parenting from failing to account for background characteristics, early cognitive development, and mother's own parenting experiences. Results indicate a confounding bias of 73% for cognitive stimulation and 89% for emotional support. This confounding bias is larger for females than for males.

The evidence on the effects of education on parenting skills and educational behaviors is insufficiently robust to problems of selection bias. It cannot be assumed that years of schooling is an appropriate measure for the formation of the identity capital skills that underlie parenting skills.

Community-Level Outcomes

At the community level, outcomes that indicate good functioning and well-being include community cohesion, low levels of crime and anti-social behavior, and trust and other aspects of social capital. A significant issue at the community level, not always apparent at lower levels of aggregation, is that social networks start to matter more explicitly in the determination of outcomes.

Individuals live and learn in multi-faceted and related contexts and these contexts interact. For example, family background interacts with schooling, and these interactions are important for the attainment and engagement of young people and the wider benefits that follow. Other contexts are also important, such as peer groups, child-care, and neighborhoods.

Crime

A US study (Lochner and Moretti) showed that a ten percentage point rise in the rate of high school graduation would cut the murder (arrest) rate by between 14% and 27%. A one percentage point increase in the graduation rate would lead to a reduction in crime of between 34 000 and 68 000 offences.

The WBL (Research Report No. 14) evaluated the impact on male juvenile burglary conviction rates of two UK government interventions, the reducing burglary initiative (RBI) and educational maintenance allowances (EMAs). The EMA program was piloted in some local education

authorities of England, creating a quasi-experimental setting. The RBI was introduced by the Home Office between 1999 and 2002. The aim of this initiative was to reduce burglary nationally by targeting areas with the worst domestic burglary problems. The WBL found that in areas where both initiatives were introduced convictions for 16- to -18-year-olds for burglary fell between 1.1 and 1.5 per 1000 relative to areas where neither program was introduced. This was also a much greater crime reduction than for areas that introduced the EMA or the RBI singly.

National-Level Benefits: Educational Inequality

At the national level, the benefits of learning are manifest as impacts on a wide range of outcomes, such as crime rates, health mortality and morbidity rates, levels of social cohesion, and engagement with public life.

Using international datasets the WBL (Research Report No. 7) explored the relationship between education and social cohesion. Education was found to be a powerful generator of social capital in some contexts, and the most significant predictor of individual propensity to trust, join, and engage in politics, even when controlling for other factors such as age, gender, and income. The report found that there is no significant relation between mean levels of education and societal cohesion. However, excluding the outliers, Norway and Germany, there is a negative and significant correlation of -0.765 between societal cohesion and education inequality. Creating a more cohesive society is likely to require policies which are designed to increase educational equality.

Final Thoughts

There is evidence of significant learning outcomes along all the socioemotional dimensions explored, including health, well-being, social cohesion, social capital, and citizenship. The best explanations situate outcomes in the context of the life-course, emphasizing the dynamic, interrelated and multi-level nature of context and process in relation to which outcomes materialize. The longitudinal aspect permits analysis of how learning outcomes are later built upon, neglected, or undermined. The dynamic aspect draws attention to the interactivity between contexts, whose causes and effects are shaped by factors variously proximate or distal, and which operate at many more levels than one.

We can only illustrate the many questions for further research: social and other types of capital theory promise further insight; there is scope for exploring further how far a model emphasizing self-productivity and complementarity (Heckman) supports or contradicts a model emphasizing context, process, and interactivity (Eccles

and the WBL). And the identification of learning outcomes often runs ahead of a secure explanation of their genesis, including the profile and direction of causality.

Bibliography

- Anderman, E. M., Maehr, M. L., and Midgley, C. (1999). Declining motivation after the transition to middle school: Schools can make a difference. *Journal of Research and Development in Education* **32**, 131–147.
- Ashton, P. (1985). Motivation and the teacher's sense of efficacy. In Ames, C. and Ames, R. (eds.) *Research on Motivation in Education, Vol. 2. The Classroom Milieu*, pp 141–171. Orlando, FL: Academic Press.
- Becker, G. and Tomes, N. (1979). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy* **87**(6), 1153–1189.
- Blumenfeld, P., Hamilton, V. L., Bossert, S., Wessels, K., and Meece, C. (1983). Teacher talk and student thought: Socialization into the student role. In Levine, J. and Wang, U. (eds.) *Teacher and Student Perceptions: Implications for Learning*. Hillsdale, NJ: Erlbaum.
- Boggiano, A. K., Shields, A., Barrett, M., et al. (1992). Helplessness deficits in students: The role of motivational orientation. *Motivation and Emotion* **16**, 271–296.
- Bossert, S. T. (1988/1989). Cooperative activities in the classroom. *Review of Research in Education* **15**, 225–252.
- Brophy, J. (1985). Teachers' expectations, motives, and goals for working with problem students. In Ames, C. and Ames, R. (eds.) *Research on Motivation in Education, Vol. 2: The Classroom Milieu*, pp 175–213. New York: Academic Press.
- Brophy, J. E. and Evertson, C. M. (1978). Context variables in teaching. *Educational Psychologist* **12**, 310–316.
- Brophy, J. E. and Good, J. L. (1974). *Teacher-Student Relationships*. New York: Holt, Rinehart and Winston.
- Cunha, F., Heckman, J., Lochner, L., and Masterov, D. V. (2005). Interpreting the evidence on life cycle skill formation. In Hanushek, E. and Welch, F. (eds.) *Handbook of the Economics of Education*, pp 697–805. Amsterdam: North-Holland.
- Deci, E. L., Schwartz, A. J., Sheinman, L., and Ryan, R. M. (1981). An instrument to assess adults' orientations toward control versus autonomy with children: Reflections on intrinsic motivation and perceived competence. *Journal of Educational Psychology* **73**, 642–650.
- Dreeben, R. and Barr, R. (1988). Classroom composition and the design of instruction. *Sociology of Education* **61**, 129–142.
- Dunn, L. and Kontos, S. (1997). Developmentally appropriate practice: What does research tell us? (ERIC Document Reproduction Service No. ED413106).
- Eccles, J. S., Lord, S. E., Roeser, R. W., Barber, B. L., and Jozefowicz, D. M. H. (1997). The association of school transitions in early adolescence with developmental trajectories through high school. In Schulenberg, J., Maggs, J., and Hurrelmann, K. (eds.) *Health Risks and Developmental Transitions during Adolescence*, pp 283–320. New York: Cambridge University Press.
- Eccles, J. S. and Midgley, C. (1989). Stage/environment fit: Developmentally appropriate classrooms for early adolescents. In Ames, C. and Ames, C. (eds.) *Research on Motivation in Education*, vol. 3, pp 13–44; 139–181. New York: Academic Press.
- Eccles, J., Midgley, C., and Adler, T. (1984). Grade-related changes in the school environment: Effects on achievement motivation. In Nicholls, J. G. (ed.) *The Development of Achievement Motivation*, pp 283–331. Greenwich, CT: JAI Press.
- Eccles, J. and Roeser, R. (2003). Schools and Adolescent Development.
- Epstein, J. L. and McPartland, J. M. (1976). The concept and measurement of the quality of school life. *American Educational Research Journal* **13**, 15–30.
- Feldhusen, J. F. and Moon, S. M. (1992). Grouping gifted students: Issues and concerns. *Gifted Child Quarterly* **36**, 63–67.
- Fine, M. (1991). *Framing Dropouts: Notes on the Politics of an Urban Public High School*. Albany, NY: State University of New York Press.
- Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research* **59**, 117–142.
- Figueira-McDonough, J. (1986). School context, gender, and delinquency. *Journal of Youth and Adolescence* **15**, 79–98.
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin* **101**, 171–191.
- Ford, M. E. (1982). Social cognition and social competence in adolescence. *Developmental Psychology* **18**, 323–340.
- Gamoran, A. and Mare, R. D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? *American Journal of Sociology* **94**, 1146–1183.
- Gensemer, P. (2000). Effectiveness of cross-age and peer mentoring programs. (ERIC Document Reproduction Service No. ED438267).
- Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental Psychology* **17**, 300–312.
- Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics* **54**(1), 3–56.
- Jackson, A. W. and Davis, G. A. (2000). *Turning Points 2000: Educating Adolescents in the 21st Century*. New York: Teachers College Press.
- Jussim, L., Eccles, J. S., and Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. In Berkowitz, L. (ed.) *Advances in Experimental Social Psychology*, pp 281–388. New York: Academic Press.
- Karoly, L. A., Greenwood, P. W., Everingham, S. S., et al. (1998). *Investing in Our Children: What We Know and Don't Know About the Costs and Benefits of Early Childhood Interventions*. Santa Monica, CA: RAND.
- Kulik, J. A. and Kulik, C. L. (1987). Effects of ability grouping on student achievement. *Equity and Excellence* **23**, 22–30.
- Larson, R. and Richards, M. (eds.) (1989). *Special Issue: The Changing Life Space of Early Adolescence*. *Journal of Youth and Adolescence* **18**, 501–626.
- Maehr, M. L. and Midgley, C. (1996). *Transforming School Cultures to Enhance Student Motivation and Learning*. Boulder, CO: Westview.
- Midgley, C., Anderman, E., and Hicks, L. (1995). Differences between elementary and middle school teachers and students: A goal theory approach. *Journal of Early Adolescence* **15**, 90–113.
- Midgley, C. and Feldlaufer, H. (1987). Students' and teachers' decision-making fit before and after the transition to junior high school. *Journal of Early Adolescence* **7**, 225–241.
- Midgley, C. M., Feldlaufer, H., and Eccles, J. S. (1989). Changes in teacher efficacy and student self- and task-related beliefs during the transition to junior high school. *Journal of Educational Psychology* **81**, 247–258.
- Moos, R. H. (1979). *Evaluating Educational Environments*. San Francisco, CA: Jossey-Bass.
- Mortimer, J. T., Shanahan, M., and Ryu, S. (1994). The effects of adolescent employment on school-related orientation and behavior. In Silbereisen, R. K. and Todt, E. (eds.) *Adolescence in Context: The Interplay of Family, School, Peers, and Work in Adjustment*, pp 304–326. New York: Springer.
- Oakes, J., Gamoran, A., and Page, R. N. (1992). Curriculum differentiation: Opportunities, outcomes, and meanings. In Jackson, P. (ed.) *Handbook of Research on Curriculum*, pp 570–608. New York: MacMillan.
- Pallas, A. M., Entwisle, D. R., Alexander, K. L., and Stulka, M. F. (1994). Ability-group effects: Instructional, social, or institutional? *Sociology of Education* **67**, 27–46.
- Parsons, J. S., Kaczala, C. M., and Meece, J. L. (1982). Socialization of achievement attitudes and beliefs: Classroom Influences. *Child Development* **53**, 322–339.
- Peterson, P. L. (1979). Direct instruction reconsidered. In Peterson, P. L. and Walberg, H. J. (eds.) *Research on Teaching: Concepts, Findings, and Implications*. National Society for the Study of Education Series on Contemporary and Educational Issues. Berkeley, CA: McCutchan.
- Pintrich, P. R. and Schunk, D. H. (1996). *Motivation in Education: Theory, Research, and Application*. Englewood Cliffs, NJ: Prentice-Hall.

- Roeser, R. W. and Eccles, J. S. (1998). Adolescents' perceptions of middle school: Relation to longitudinal changes in academic and psychological adjustment. *Journal of Research on Adolescence* **88**, 123–158.
- Roeser, R. W., Eccles, J. S., and Freedman-Doan, C. (1999). Academic functioning and mental health in adolescence: Patterns, progressions, and routes from childhood. *Journal of Adolescent Research* **14**, 135–174.
- Roeser, R. W., Eccles, J. S., and Sameroff, J. (1998). Academic and emotional functioning in early adolescence. Longitudinal relations, patterns, and prediction by experience in middle school. *Development and Psychopathology* **10**, 321–352.
- Rosenbaum, J. E. (1976). *Making Inequality: The Hidden Curriculum of High School Tracking*. New York: Wiley.
- Rosenholtz, S. J. and Simpson, C. (1984). The formation of ability conceptions: Developmental trend or social construction? *Review of Educational Research* **54**, 301–325.
- Roswal, G. M., Mims, A. A., Evans, M. D., et al. (1995). Effects of collaborative peer tutoring on urban seventh graders. *Journal of Educational Research* **88**, 275–279.
- Rutter, M. (1983). School effects on pupil progress: Research findings and policy implications. *Child Development* **54**, 1–29.
- Schuller, T., Preston, J., Hammond, C., Brassett-Grundy, A., and Bynner, J. (2004). *The Benefits of Learning: The Impact of Education on Health, Family Life and Social Capital*. New York: RoutledgeFalmer.
- Schwartz, F. (1981). Supporting or subverting learning: Peer group patterns in four tracked schools. *Anthropology and Education Quarterly* **12**, 99–121.
- Sharan, S. (1980). Cooperative learning in small groups: Recent methods and effects on achievement, attitudes, and ethnic relations. *Review of Educational Research* **50**, 241–271.
- Shields, C. M. (1996). To group or not to group academically talented or gifted students? *Educational Administration Quarterly* **32**, 295–323.
- Slavin, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research* **60**, 471–499.
- Tierney, J. and Grossman, J. (1995). *Making a Difference: An Impact Study of Big Brothers/ Big Sisters*. Philadelphia, PA: Public/Private Ventures.
- Weinstein, R. (1989). Perceptions of classroom processes and student motivation: Children's views of self-fulfilling prophecies. In Ames, C. and Ames, R. (eds.) *Research on Motivation in Education: Vol. 3. Goals and Cognitions*, pp 13–44. New York: Academic Press.
- Wigfield, A., Eccles, J., Mac Iver, D., Reuman, D., and Midgley, C. (1991). Transitions at early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology* **27**, 552–565.
- Wilczenski, F. L., Bontrager, T., Ventrone, P., and Correia, M. (2001). Observing collaborative problem-solving processes and outcomes. *Psychology in the Schools* **38**, 269–281.
- Wilczenski, F. L., Konstam, V., Ferraro, B., Kaplan, L., and Bontrager, School and Classroom Organisations, ERIC ED 462 659.
- Anderman, E. M. and Maehr, M. L. (1994). Motivation and schooling in the middle grades. *Review of Educational Research* **64**, 287–309.
- Arunkumar, R., Midgley, C., and Urdan, T. (1999). Perceiving high or low home-school dissonance: Longitudinal effects on adolescent emotional and academic well-being. *Journal of Research on Adolescence* **9**, 441–466.
- Bandura, A. (1994). *Self-Efficacy: The Exercise of Control*. New York: Freeman.
- Barber, B. L., Eccles, J. S., and Stone, M. R. (2001). Whatever happened to the jock, the brain, and the princess? Young adult pathways linked to adolescent activity involvement and social identity. *Journal of Adolescent Research* **16**, 429–455.
- Barker, R. and Gump, P. (1964). *Big School, Small School: High School Size and Student Behavior*. Stanford, CA: Stanford University Press.
- Baron, S., Schuller, T., and Field, J. (2000). *Social Capital: Critical Perspectives*. Oxford: Oxford University Press.
- Becker, G. and Tomes, N. (1986). Human capital and the rise and fall of families. *Journal of Labor Economics* **4**(3), S1–S39.
- Blossfeld, H. -P. and Shavit, Y. (1993). *Persistent Inequality: Changing Educational Attainment in Thirteen Countries*. Boulder, CO: Westview.
- Ben-Porath, Y. (1967). The production of human capital and the life cycle earnings. *Journal of Political Economy* **75**(4), 352–365.
- Blackwell, L. and Bynner, J. (2002). Learning, family formation and dissolution. Wider Benefits of Learning Research Report No.4. *Centre for Research on the Wider Benefits of Learning*.
- Blumenfeld, P. C. (1992). Classroom learning and motivation: Clarifying and expanding goal theory. *Journal of Educational Psychology* **84**, 272–281.
- Bowers, C. A. and Flinders, D. (1990). *Responsive Teaching: An Ecological Approach to Classroom Patterns of Language, Culture, and Thought*. New York: Teachers College Press.
- Bransford, J. D., Brown, A. L., and Cocking, R. R. (eds.) (1999). *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academy Press.
- Brofenbrenner, U. (1979). *The Ecology of Human Development*. Cambridge, MA: Harvard University Press.
- Brookover, W., Beady, C., Flood, P., Schweitzer, J., and Wisenbaker, J. (1979). *School Social Systems and Student Achievement: Schools Can Make a Difference*. New York: Praeger.
- Brooks-Gunn, J., Berlin, L. J., and Fuligni, A. S. (2000). Early childhood intervention programs: What about the family. In Meisels, S. and Shonkoff, J. (eds.) *The Handbook of Early Childhood Interventions*, pp 549–588. New York: Cambridge University Press.
- Brophy, J. (1988). Research linking teacher behavior to student achievement: Potential implications for instruction of chapter 1 students. *Educational Psychologist* **23**, 235–286.
- Brophy, J. E. (1979). Teacher behavior and its effects. *Journal of Educational Psychology* **71**(6), 733–760.
- Brown, B. B. (1990). Peer groups and peer culture. In Feldman, S. S. and Elliott, G. R. (eds.) *At the Threshold: The Developing Adolescent*, pp 171–196. Cambridge, MA: Harvard University Press.
- Bryk, A. S., Lee, V. E., and Holland, P. B. (1993). *Catholic Schools and the Common Good*. Cambridge, MA: Harvard University Press.
- Bryk, A. S., Lee, V. E., and Smith, J. B. (1989). High school organization and its effects on teachers and students: An interpretative summary of the research. *Paper Presented at the Invitational Conference on Choice and Control in American Education: Robert M. La Follette of Public Affairs: University of Wisconsin-Madison*.
- Bynner, J. and Parsons, S. (1997). Does numeracy matter? *Evidence from the national child development study on the impact of poor numeracy on adult life*. Social Statistics Research Unit: City University on behalf of The Basic Skills Agency.
- Calderhead, J. (1996). Teachers, beliefs, and knowledge. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 709–725. New York: Simon and Schuster/Macmillan.
- Cameron, J. (2004). Evidence for an early sensitive period for the development of brain systems underlying social affiliative behavior. *Working Paper*.
- Cameron, S. and Heckman, J. J. (1998). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males. *Journal of Political Economy* **106**(2), 262–333.

Further Reading

- Allan, S. D. (1991). Ability-grouping research reviews: What do they say about grouping and the gifted. *Educational Leadership* **48**, 60–74.
- Allen, J. P. (1989). Social impact of age mixing and age segregation in school: A context-sensitive investigation. *Journal of Educational Psychology* **81**, 408–416.
- Altonji, J. and Dunn, T. (1996). The effects of family characteristics on the return to education. *Review of Economics and Statistics* **78**(4), 692–704.
- American Association of University Women (1992). *How Schools Shortchange Girls*. Washington, DC: American Association of University Women.
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology* **84**, 261–271.

- Carnegie Council on Adolescent Development (1989). *Turning Points: Preparing American Youth for the 21st Century*. New York: Carnegie Corporation.
- Cartledge, G. and Johnson, C. T. (1996). Inclusive classroom for students with emotional and behavioural disorders: Critical variables. *Theory into Practice* **35**, 51–57.
- Cohen, P. A., Kulik, J. A., and Kulik, C. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal* **19**, 237–248.
- Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology* **94**, S95–S120.
- Coleman, J. S. (1961). *The Adolescent Society*. New York: Free Press.
- Coleman, J. S. and Hoffer, T. (1987). *Public and Private High Schools: The Impact of Communities*. New York: Basic Books.
- Connell, J. P. and Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In Gunnar, R. and Sroufe, L. A. (eds.) *Minnesota Symposia on Child Psychology*, vol. 23, pp 43–77. Hillsdale, NJ: Erlbaum.
- Cooper, H. M. (1979). Pygmalion grows up: A model for teacher expectation communication and performance influence. *Review of Educational Research* **49**(3), 389–410.
- Council of the Great City Schools (1992). *National Urban Education Goals: Baseline Indicators, 1990–91*. Washington, DC: Council of the Great City Schools.
- Covington, M. V. (1992). *Making the Grade: A Self-Worth Perspective on Motivation and School Reform*. New York: Cambridge University Press.
- Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives* **15**(2), 213–238.
- Damon, W. and Phelps, E. (1989). Critical distinctions among three methods of peer interaction. *International Journal of Educational Research* **13**, 9–19.
- Deci, E. L. and Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Press.
- Dryfoos, J. G. (1990). *Adolescents at Risk: Prevalence and Prevention*. Oxford: Oxford University Press.
- Dunkin, M. and Biddle, B. (1974). *The Study of Teaching*. Holt, Rinehart and Winston.
- Dweck, C. S. (1999). Self-Theories: Their Role in Motivation. Personality and Development. Philadelphia, PA: Psychology Press.
- Eccles, J. S., Midgley, C., Wigfield, A., et al. (1993). Development during adolescence: The impact of stage–environment fit on adolescents experiences in schools and families. *American Psychologist* **48**, 90–101.
- Eccles, J. and Wigfield, A. (1985). Teacher expectations and student motivation. In Dusek, J. B. (ed.) *Teacher Expectations*, pp 185–217. Hillsdale, NJ: Erlbaum.
- Eder, D. and Felmlee, D. (1984). The development of attention norms in ability groups. In Peterson, P. L., Wilkinson, L. C., and Hallinan, M. (eds.) *The Social Context of Instruction*, pp 189–207. New York: Academic Press.
- Elder, G. H. Jr. and Conger, R. D. (2000). *Children of the Land*. Chicago, IL: Chicago University Press.
- Elias, M. J., Zins, J. E., Weissberg, R. P., et al. (1997). *Promoting Social and Emotional Learning: Guidelines for Educators*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Emmer, E. T. and Evertson, C. M. (1982). Synthesis of research in classroom management. *Educational Leadership* **39**, 342–347.
- Entwistle, D. R. and Alexander, K. L. (1993). Entry into school: The beginning school transition and educational stratification in the United States. *Annual Review of Sociology* **19**, 401–423.
- Epstein, J. L. (1983). The influence of friends on achievement and affective outcomes. In Epstein, J. L. and Karweit, N. L. (eds.) *Friends in School*, pp 177–200. New York: McGraw-Hill.
- Erikson, E. H. (1959). Identity and the life cycle. *Psychological Issues* **1**, 18–164.
- Feinstein, L. (2002). Quantitative estimates of the social benefits of learning, 2: Health (depression and obesity). *Wider Benefits of Learning Research Report No. 6*. London: WBL.
- Feinstein, L. (2006). A synthesis of findings from the Centre for Research on the Wider Benefits of Learning. The Centre for Research on the Wider Benefits of Learning. Institute of Education, London.
- Feldlaufer, H., Midgley, C., and Eccles, J. S. (1988). Student, teacher, and observer perceptions of the classroom environment before and after the transition to junior high School. *Journal of Early Adolescence* **8**, 133–156.
- Felmlee, D. and Eder, D. (1983). Contextual effects in the classroom: The impact of ability groups on student attention. *Sociology of Education* **56**, 77–87.
- Finger, J. A. and Silverman, M. (1966). Changes in academic performance in the junior high school. *Personnel and Guidance Journal* **45**, 157–164.
- Fox, L. M., Brody, L., and Tobin, D. (1985). The impact of early intervention programs upon course-taking and attitudes in high school. In Chipman, S. F., Brush, L. R., and Wilson, D. M. (eds.) *Women and Mathematics: Balancing the Equation*, pp 249–274. Hillsdale, NJ: Erlbaum.
- Franca, V. M., Kerr, M. M., Reitz, A. L., and Lambert, D. (1990). Peer tutoring among behaviorally disordered students: Academic and social benefits to tutor and tutee. *Education and Treatment of Children* **13**, 109–128.
- Fraser, B. J. and Fisher, D. L. (1982). Predicting students' outcomes from their perceptions of classroom psychosocial environment. *American Educational Research Journal* **19**, 498–518.
- Fuerst, J. S. and Fuerst, D. (1993). Chicago experience with an early childhood program: The special case of the child parent center program. *Urban Education* **28**(1), 69–96.
- Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education* **59**, 185–198.
- Gersten, R. and Keating, T. (1987). Long term benefits from direct instruction. *Educational Leadership* **44**, 28–31.
- Glancy, M., Willits, F. K., and Farrell, P. (1986). Adolescent activities and adult success and happiness: Twenty-four years later. *Sociology and Social Research* **70**, 242–250.
- Goldenberg, C. (1992). The limits of expectations: A case for case knowledge about teacher expectancy effects. *American Educational Research Journal* **29**, 517–544.
- Good, T. and Stipek, D. (1983). Individual differences in the classroom: A psychological perspective. In Fernstermacher, M. and Goodlad, J. (eds.) *National Society for the Study of Education Yearbook*, pp 9–43. Chicago, IL: University of Chicago.
- Goodenow, C. (1992). Strengthening the links between educational psychology and the study of social contexts. *Educational Psychologist* **27**, 177–196.
- Goodenow, C. (1993). Classroom belonging among early adolescent students: Relationships to motivation and achievement. *Journal of Early Adolescence* **13**(1), 21–43.
- Grant, L. and Rothenberg, J. (1986). The social enhancement of ability differences: Teacher–student interactions in first and second grade reading groups. *Elementary School Journal* **87**, 29–49.
- Gutman, L. M. and Midgley, C. (2000). The role of protective factors in supporting the academic achievement of poor African American students during the middle school transition. *Journal of Youth and Adolescence* **29**, 223–248.
- Heckman, J. J. and Lochner, L. (2000). Rethinking myths about education and training: Understanding the sources of skill formation in a modern economy. In Danziger, S. and Waldfogel, J. (eds.) *Securing the Future: Investing in Children from Birth to College*, pp 47–86. New York: Russell Sage.
- Heckman, J. J., Lochner, L., and Todd, P. (2003). Fifty years of Mincer regressions. *Working Paper*. University of Chicago.
- Heckman, J. J. and Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review* **91**(2), 145–149.
- Heckman, J. and Vytalil, E. (2005). Econometric evaluation of social programs. In Heckman, J. and Leamer, E. (eds.) *Handbook of Econometrics*, vol. 6. Amsterdam: Elsevier.
- Heinz, W. R. (ed.) (1999). *From Education to Work: Cross National Perspectives*. Cambridge: Cambridge University Press.
- HM Government (2003). Every Child Matters. CM 5860. HMSO.
- Hunt, D. E. (1975). Person–environment interaction: A challenge found wanting before it was tried. *Review of Educational Research* **45**, 209–230.
- Jencks, C. L. and Brown, M. (1975). The effects of high schools on their students. *Harvard Educational Review* **45**, 273–324.

- Johnson, A. (1996). *An Evaluation of the Long-Term Impacts of the Sponsor-A-Scholar Program on Student Performance*. Princeton, NJ: Mathematica Policy Research.
- Jones, M. B. and Thompson, D. G. (1981). Classroom misconduct and integration by sex. *Journal of Child Psychology and Psychiatry* **22**, 401–409.
- Judd, K. (2000). Is education as good as gold? A portfolio analysis of human capital investment. *Working Paper*. Stanford University.
- Jussim, L. and Eccles, J. S. (1992). Teacher expectations II: Construction and reflection of student achievement. *Journal of Personality and Social Psychology* **63**, 947–961.
- Kagan, D. M. (1990). How schools alienate students at risk: A model for examining proximal classroom variables. *Educational Psychologist* **25**, 105–125.
- Kavrell, S. M. and Peterson, A. C. (1984). Patterns of achievement in early adolescence. In Maehr, M. L. (ed.) *Advances in Motivation and Achievement*, pp 1–35. Greenwich, CT: JAI Press.
- Kenny, D. A., Archambault, F. X., and Hallmark, B. V. (1995). The effects of group composition on gifted and non-gifted elementary students in cooperative learning groups. (ERIC Document Reproduction Service No. ED402702).
- Lamborn, S. D., Brown, B. B., Mounts, N. S., and Steinberg, L. (1992). Putting school in perspective: The influence of family, peers, extracurricular participation, and part-time work on academic engagement. In Newmann, F. M. (ed.) *Student Engagement and Achievement in American Secondary Schools*, pp 153–181. New York: Teachers College Press.
- Lazerson, D. B., Foster, H. L., Brown, S. I., and Hummel, J. W. (1998). The effectiveness of cross-age tutoring with truant, junior high school students with learning disabilities. *Journal of Learning Disabilities* **4**, 253–255.
- Lee, V. E. and Bryk, A. S. (1986). Effects of single-sex secondary schools on student achievement and attitudes. *Journal of Educational Psychology* **78M**, 381–395.
- Lee, V. E. and Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education* **62**, 172–192.
- Lee, V. E., Bryk, A. S., and Smith, J. B. (1993). The organization of effective secondary schools. In Darling-Hammond, L. (ed.) *Review of Research in Education*, vol. 19, pp 171–267. Washington, DC: American Educational Research Association.
- Lloyd, L. (1999). Multi-age classes and high ability students. *Review of Educational Research* **69**, 187–212.
- Lord, S., Eccles, J. S., and McCarthy, K. (1994). Risk and protective factors in the transition to junior high school. *Journal of Early Adolescence* **14**, 162–199.
- Lou, Y., Abrami, P. C., Spence, J. C., et al. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research* **66**, 423–458.
- Mac Iver, D. (1988). Classroom environments and the stratification of students' ability perceptions. *Journal of Educational Psychology* **80**, 1–40.
- Mac Iver, D. J., Reuman, D. A., and Main, S. R. (1995). Social structuring of school: Studying what is, illuminating what could be. In Rosenzweig, M. R. and Porter, L. W. (eds.) *Annual Review of Psychology*, vol. 46.
- Maehr, M. L. and Midgley, C. (1991). Enhancing student motivation: A school-wide approach. *Educational Psychologist* **26**, 399–427.
- Mael, F. A. (1998). Single-sex and coeducational schooling: Relationships to socio-emotional and academic development. *Review of Educational Research* **68**, 101–129.
- Marsh, H. W., Chessor, D., Craven, R., and Roche, L. (1995). The effects of gifts and talented programs on academic self-concept: The big fish strikes again. *American Educational Research Journal* **32**, 285–319.
- Marshall, H. H. and Weinstein, R. S. (1984). Classroom factors affecting students' self-evaluations: An interactional model. *Review of Educational Research* **54**, 301–325.
- Martin, A. J., Marsh, H. W., and Debus, R. L. (2001). A quadripolar need achievement representation of self-handicapping and defensive pessimism. *American Educational Research Journal* **38**, 583–610.
- McClellan, D. E. and Kinsey, S. (1997). Children's social behavior in relationship to participation in mixed-age or same-sex classrooms. *Paper Presented at the Biennial Meeting of the Society for Research in Child Development*. Washington, DC.
- Midgley, C. (1993). Motivation and middle level schools. In Maehr, M. L. and Pintrich, P. (eds.) *Advances in Motivation and Achievement: Vol. 8: Motivation and Adolescent Development*, pp 217–274. Greenwich, CT: JAI Press.
- Midgley, C. and Edelin, K. C. (1998). Middle school reform and early adolescent well-being: The good news and the bad. *Educational Psychologist* **33**, 195–206.
- Midgley, C., Feldlaufer, H., and Eccles, J. S. (1988). The transition to junior high school: Beliefs of pre- and post-transition teachers. *Journal of Youth and Adolescence* **17**, 543–562.
- Midgley, C., Kaplan, A., and Middleton, M. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, and at what cost? *Journal of Educational Psychology* **93**, 77–86.
- Mortimer, J. T. and Johnson, M. K. (1999). Adolescent part-time work and postsecondary transition pathways in the United States. In Heinz, W. R. (ed.) *From Education to Work: Cross National Perspectives*, pp 111–148. Cambridge: Cambridge University Press.
- Newmann, F. M., Wehlage, G. G., and Lamborn, S. D. (1992). The significance and sources of student engagement. In Newmann, F. M. (ed.) *Student Engagement and Achievement in American Secondary Schools*, pp 11–39. New York: Teachers College Press.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review* **91**, 328–346.
- Nicholls, J. G. (1990). What is ability and why are we mindful of it? A developmental perspective. In Sternberg, R. J. and Kolligian, J. (eds.) *Competence Considered*, pp 11–40. New Haven, CT: Yale University Press.
- Parsons, J. E. and Ruble, D. N. (1977). The development of achievement-related expectancies. *Child Development* **48**, 1075–1079.
- Passow, A. H. (1988). Issues of access to knowledge: Grouping and tracking. In Tanner, L. N. (ed.) *Critical Issues in Curriculum. Eight-Seventh Yearbook of the National Society for the Study of Education*, pp 205–229. Chicago, IL: University of Chicago Press.
- Perry, N. E. (1998). Young children's self-regulated learning and contexts that support it. *Journal of Educational Psychology* **90**, 715–729.
- Perry, N. E. and Weinstein, R. S. (1998). The social context of early schooling and children's school adjustment. *Educational Psychologist* **33**, 177–194.
- Pintrich, P. and De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology* **82**, 33–40.
- Pratt, D. (1986). On the merits of multi-age classrooms: Their work life. *Research in Rural Education* **3**, 111–116.
- Putnam, R. (1999). *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.
- Reynolds, A. L. (1999). Working with children and adolescents in the schools: Multicultural counselling implications. In Sheets, R. H. and Hollins, E. R. (eds.) *Racial and Ethnic Identity in School Practices: Aspects of Human Development*, pp 213–230. Mahwah, NJ: Erlbaum.
- Reynolds, A. J., Temple, J. A., Robertson, D. L., and Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest – a 15-year follow-up of low-income children in public schools. *Journal of American Medical Association* **285**(18), 2339–2346.
- Rist, R. C. (1970). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review* **40**, 411–451.
- Roderick, M. (1993). *The Path to Dropping out: Evidence for Intervention*. Westport, CT: Auburn House.
- Roeser, R. W. and Midgley, C. M. (1997). Teachers' views of aspects of student mental health. *Elementary School Journal* **98**(2), 115–133.
- Roeser, R. W., Midgley, C. M., and Urdan, T. C. (1996). Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning in school: The mediating role of goals and belonging. *Journal of Educational Psychology* **88**, 408–422.

- Rosenbaum, J. E. (1980). Social implications of educational grouping. *Review of Research in Education* **7**, 361–401.
- Rosenholtz, S. R. and Rosenholtz, S. J. (1981). Classroom organization and the perception of ability. *Sociology of Education* **54**, 132–140.
- Ruble, D. (1983). The development of social comparison processes and their role in achievement-related self-socialization. In Higgins, E. T., Ruble, D. N., and Hartup, W. W. (eds.) *Social Cognition and Social Development: A Sociocultural Perspective*, pp 134–157. New York: Cambridge University Press.
- Rutter, M., Maughan, B., Mortimore, P., and Ouston, J. (1979). *Fifteen Thousand Hours: Secondary Schools and Their Effects on Children*. Cambridge: Harvard University Press.
- Schuller, T., Bynner, J., Green, A., et al. (2001). *Modelling and Measuring the Wider Benefits of Learning, Wider Benefits of Learning Papers No. 1*. London: Institute of Education.
- Schweinhart, L., Barnes, H., and Weikart, D. (1993). *Significant Benefits: The High-Scope Perry Pre-School Study through Age 27*. Ypsilanti, MI: High Scope Press.
- Skinner, E. A. and Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology* **85**, 571–581.
- Sternberg, R. (1985). *Beyond IQ: A Triarchic Theory of Human Intelligence*. Cambridge: Cambridge University Press.
- Stipek, D. J. (1996). Motivation and instruction. In Calfee, R. C. and Berliner, D. C. (eds.) *Handbook of Educational Psychology*. New York: Macmillan.
- Strobel, K. and Roeser, R. W. (1998). Patterns of motivation and mental health in middle school: Relation to academic and emotional regulation strategies. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Diego.
- Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., and Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science* **14**(6), 623–628.
- Wang, M. C. and Gordon, E. W. (eds.) (1994). *Educational Resilience in Inner-City American*. Hillsdale, NJ: Erlbaum.
- Way, J. W. (1981). Achievement and self-concept in multi-age classrooms. *Educational Research Quarterly* **6**, 69–75.
- Wehlage, G. (1989). Dropping out: Can schools be expected to prevent it? In Weis, L., Farrar, E., and Petrie, H. (eds.) *Dropouts from School*, pp 1–23. Albany, NY: SUNY Press.
- Wehlage, G., Rutter, R., Smith, G., Lesko, N., and Fernandez, R. (1989). *Reducing the Risk: Schools as Communities of Support*. Philadelphia, PA: Falmer.
- Weiner, B. (1990). History of motivation research in education. *Journal of Educational Psychology* **82**, 616–622.
- Wigfield, A., Eccles, J. S., and Pintrich, P. R. (1996). Development between the ages of eleven and twenty-five. In Berliner, D. C. and Calfee, R. C. (eds.) *The Handbook of Educational Psychology*. New York: MacMillan.
- Wilczenski, F. L. (1993). Changes in attitudes toward mainstreaming among undergraduate education students. *Educational Research Quarterly* **17**, 5–17.
- Winn, W. and Wilson, A. P. (1983). The affect and effect of ability grouping. *Contemporary Education* **54**, 119–125.
- Youniss, J., McLellan, J. A., and Yates, M. (1997). What we know about engendering civic identity. *American Behavioral Scientist* **40**, 619–630.
- Youniss, J., Yates, M., and Su, Y. (1997). Social integration: Community service and marijuana use in high school seniors. *Journal of Adolescent Research* **12**, 245–262.

Social Aspects of Collaborative Learning

M Arvaja and P Häkkinen, University of Jyväskylä, Jyväskylä, Finland

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Collaboration, Collaborative learning – In this article, mostly referred as construction of shared understanding, coordination, co-construction of knowledge.

Content and relational space of communication – Content space refers to cognitive aspect of collaborative communication, for example, how the subject in hand is reasoned, relational space refers to social aspect of collaborative communication, for example, how students orientate towards each others in interaction (collaborative, individualistic or competitive orientation).

Social aspects – Social dimension in collaborative interaction, social relationships (friendship, status).

Status – (1) Academic status: perceived academic ability or knowledge level; (2) Peer status: perceived attractiveness or popularity.

The ideas presented in this article are based on the recent research on collaborative learning. This research demonstrates the dual nature of collaboration, in which cognitive and social processes of learning interact with each other in a dynamic manner. The article first discusses the central concepts and research trends related to the social dimension of collaborative activity. Further, research results on social relationships in collaborative learning are summarized. Particular emphasis is laid on the friendship and status. Finally, the main findings on social aspects of collaborative learning are synthesized, and some critical challenges for this research field are raised.

Social Dimension in Collaborative Activity

The most widely used definition of collaboration describes it as a construction of shared understanding through interaction with others, where the participants are committed to or engaged in shared goals and problem solving (Dillenbourg, 1999; Littleton and Häkkinen, 1999; Roschelle and Teasley, 1995). In addition to the construction of shared understanding, collaboration is commonly referred to as the co-construction of knowledge (e.g., Rafal, 1996; Baker, 2002), building collaborative knowing (Stahl, 2004), co-argumentation (Baker, 2002), negotiating of shared meaning (Pea, 1993), construction of common

knowledge (e.g., Elbers and Streefland, 2000; Crook, 2002), exploratory talk (Mercer, 1996), or coordination (Barron, 2000). However, on closer inspection, the core of all these terms seems to refer to interaction of a similar type – one that calls for specific qualities from cognitive, social, and emotional perspectives.

According to Barron (2000), the extent to which a group works collaboratively depends on the degree of coordination among the group members. In her study comparing the interaction of successful and less-successful student triads, Barron (2000) found three forms of coordination – shared task alignment, mutuality, and jointly focused attention – by which the interaction of these groups differed. The group activity is coordinated in a situation where the students have shared task alignment. In such a situation, the students have collaborative orientation toward problem solving, and the activity is organized around joint problem-solving efforts. In the activities of groups, this is manifested in the co-construction of solutions and referring to and expanding each other's ideas as opposed to individual solution paths and reference to one's own ideas. The success presupposes that participants have achieved a common ground (Clark, 1996) or alignment in the sense that students work in phase at the same conceptual level (Baker, 2002). Barron (2000) adds that there may be some complementary role division in the activity as in a situation where one person is generating ideas while writing and the other is monitoring the documentation. Thus, there is some horizontal division of labor while the first person is acting at the task level and the other, at the meta-(communicative) level (Dillenbourg, 1999).

Mutuality, Barron's second form of coordination, refers to the extent to which there is reciprocity and balance in the interaction with potential of all members to meaningfully contribute and be heard. Mutuality is reflected in the nature of dialog; how the contributions of others are treated in the discussion and how the ideas offered are engaged by others. Barron refers to transactive dialogs, introduced by Berkowitz and Gibbs (1985), where one partner's reasoning operates on another partner's reasoning. In interaction, the content of the other person's turn of speech is taken into account and acknowledged, for example, by acceptance, clarification, or elaboration, instead of rejection without an explanation. Even the conflicts that arise in the situation are productive (Barron, 2000). It is also typical of mutuality that the turn-taking norms are respected. Barron's (2002) mutuality comes close to Baker's (2002) notion of symmetry. According to Baker (2002), interaction is symmetrical if the participants adopt certain roles equally throughout the

interaction, that is, participate equally in problem solving. Even though Baker (2002) does not refer to symmetry of knowledge, a certain degree of knowledge symmetry is essential to enable equal roles (Dillenbourg, 1999). If the knowledge level of the participants is very different, it easily triggers different (permanent) statuses and roles in the learning situation (Cohen, 1994). In addition, Van Boxtel (2000) observes that the essential prerequisite for co-construction is that all participants equally contribute to the elaboration and solution of the problem at hand. In her work, she describes co-construction as an elaborative episode in which both (all) participants make a verbal and propositional contribution to the elaboration.

The third dimension of coordination is the degree to which the attention is jointly focused during solution-critical moments. In the interaction, there is high sensitivity to one another's attentional states. According to Barron (2000), joint attention is closely related to the notion of mutuality. In addition, the notion of joint attention has similarities with Baker's (2002) term, alignment, in the sense that, in the occasions of joint attention, all the students are working in phase with respect to the activities, and there is not even a horizontal division of labor. As a whole, the three dimensions of coordination – shared task alignment, mutuality, and jointly focused attention – can be regarded as different intertwined aspects of collaborative interaction. For example, the co-construction of knowledge is not possible unless group members have joint focus or mutuality in the situation.

The descriptions presented above well illustrate the dual nature of collaborative interactions. According to Barron (2003), in collaborative activity, the participants have to develop and monitor both the content space and the relational space or, as Sfard and Kieran (2001) put it, the object and meta-level of communication. The content space refers to the cognitive aspect of collaboration: how the subject at hand is reasoned, how the ideas are developed in discussion, and how the shared understanding is constructed. Relational space refers more to the social aspect of collaboration and to the way in which participants in dialog (or monolog) orientate toward each other (and how willing they are to engage in interaction) (Barron, 2003). In the joint activity, the students may have collaborative orientation in their working, or the situation can be competitive, individualistic (Barron, 2003; Mercer, 1996; Sfard and Kieran, 2001), or asymmetrical (Baker, 2002). In a competitive situation, individual participants define themselves through their difference from the others, each having his or her own competing interests, which they try to drive through (Wegerif and Mercer, 1997). Thus, the participants easily end up with disputing rather than arguing for their case. In an individualistic orientation, each student follows his or her own paths of reasoning. In this case, the interaction is more monological than dialogical. For example, in Barron's (2003) study, each of the students in failed groups tried to present and

construct his or her own solutions for the problem. In asymmetrical situations, participants have no possibility to make an equal contribution to the discourse. Reasons may relate, for example, to knowledge (Dillenbourg, 1999) or status differences (Cohen, 1994). In their study, Kumpulainen and Mutanen (1999) distinguished between individualistic, dominant, and collaborative modes of interaction according to the perceived degree of co-construction.

The content and relational spaces are negotiated simultaneously and, thus, compete for limited attention. For example, if the relational space is more focused on competitive interaction or self-focused (individualistic) problem solving, it prevents the participants from gaining joint attention and mutual engagement, and from reaching a common ground on the same topic. At the same time, success in the content space requires success in the relational space. The content and relational spaces thus have a reciprocal relationship, being part of the same collaborative process, and are hard to separate.

Social Relationships in Relation to Collaboration

There is not much research about how students' preexisting relationships within their peer group influence the nature and quality of their interaction in collaborative contexts (Miell and MacDonald, 2000). The study of collaboration has focused more on interactions as opposed to relationships (Azmitia, 1998). However, it has been suggested that students' relationships affect their emotional and social responses to the cognitive working conditions, which show in their willingness to participate in collaborative learning contexts (Crook, 1999; Kutnick and Manson, 1998). Thus, processes and outcomes of collaboration are embedded within the larger context of peer relationships (Azmitia, 1998). Azmitia (2000) points out that collaboration-related research relying on problem-solving tasks that have one correct solution or can be solved through systematic hypothesis testing has contributed to a rosy, calm picture of collaboration placing emphasis on cognitive development. However, in more open-ended and ill-defined problems, collaboration is more stressful because there is no clear script on how to proceed. In these kinds of tasks, collaborators' personalities and relationships as well as the affective elements of interaction play an important role in managing collaboration.

Friendship

There are reports about the positive effects of friendship on collaboration (e.g., Hartup, 1996; Azmitia and Montgomery, 1993; MacDonald *et al.*, 2000). Azmitia and Montgomery (1993) found that collaboration between friends leads to more transactive discourse, greater equality in roles, and

larger increases in knowledge than collaborations between acquaintances. In their study, friends justified their proposals, elaborated on their partners' proposals more often, and also engaged in transactive conflicts more frequently than acquaintances did. In addition, MacDonald *et al.* (2000) found out that, in creative tasks, friends engaged in more transactive discussions than nonfriends, and their musical compositions were rated more highly than those of the latter. Hartup (1996) drew a synthesis on studies comparing friends and nonfriends during collaborative problem-solving tasks. According to these studies, friends tend to engage in more extensive discourse and task-oriented talk, offer suggestions more readily, and are more supportive and critical than nonfriends. In addition, mutuality is more evident and interaction is more positive and equally balanced between friends than between nonfriends. Thus, according to Hartup (1996), social interaction between friends is of the type that is considered to facilitate cognitive development.

However, Berndt *et al.* (1988) did not find differences in the processes and outcomes of the collaborations of friends and acquaintances. One possible reason for the controversy between different studies is that the varied nature and difficulty of tasks may show in differences in the processes and outcomes of the collaborations (Azmitia, 1998). In their study reported above, Azmitia and Montgomery (1993) found out that friends outperformed acquaintances in a posttest only on the most difficult tasks because most friend pairs were able to sustain and repair the collaboration in dealing with difficult tasks, whereas the acquaintances were not. In addition, MacDonald *et al.*, (2000) suggested that, in open-ended creative tasks, nonfriends have more difficulties than friends because they lack the advantage of shared knowledge and accustomed patterns of interaction that would help them maintain interaction and construct knowledge.

The relationship between friendship and successful collaboration is suggested to be due to different reasons. According to Azmitia (1998), mutuality, trust, respect, equality, and fairness, often associated with friendship, are qualities that help friends to be attuned to each other's needs, goals, and points of view, and let them expose their views and also challenge each other. These qualities are beneficial in negotiating shared understandings. According to MacDonald *et al.* (2000), friends are used to establish and maintain a shared social reality in their everyday relationship and also to generate and develop ideas together. Thus, friends have a shared history to lean on (Crook, 1999). On the one hand, this shared history shows in automatized interaction routines that help maintain interaction even in highly demanding tasks (Azmitia, 1998); on the other hand, it shows as shared experiences and prior knowledge (Costin and Jones, 1992; Faulkner and Miell, 2004) that facilitate establishing a common ground (Clark, 1996). In addition, it has been suggested that friends are more motivated and willing to work together (Barron, 2003).

There are, however, age-related developmental differences in the friends' ability to collaborate (see Azmitia, 1998) as well as gender-related differences in the friends' quality of collaboration (e.g., Faulkner and Miell, 2004; Kutnick and Kington, 2005). For example, Kutnick and Kington (2005) found that girl friendship pairs outperformed boy friendship pairs in a scientific reasoning task at the primary level of education. The pupils' interviews revealed that the boys excluded school collaboration as a legitimate activity within the context of their friendships, whereas the girls did not. The boys' friendships were action oriented and based on activities with others outside the school classroom. The study demonstrated that male and female friendship pairs had distinct, culturally (gender related) defined approaches in their cognitive problem-solving activity. These approaches were tied to the experiences they shared with friends in the classrooms and outside the school (Kutnick and Kington, 2005). This is in accordance with Faulkner and Miell's (2004) suggestion that effective collaboration between friends occurs only when it is situated in an activity setting that is meaningful in the context of their relationship.

Status

Several authors (e.g., Dillenbourg, 1999; Baker, 2002) have argued that in order to be capable for true collaboration, the participators must play symmetrical roles in the conversation. They must have the same opportunity to participate, and their level of knowledge must broadly be the same (Dillenbourg, 1999). When the knowledge level among the participants is very different, it easily leads to different roles in the learning situation. These different roles, in turn, can have profound effects on the quality of interaction and learning (Basili and Sandford, 1991; Cohen, 1994; Linn and Burbules, 1993; Richmond and Striley, 1996). For example, Richmond and Striley (1996) found several leadership styles, each with different effects on group interaction. An alienating leader had a negative effect on students' interaction because the leader disregarded the input of others and restricted the discussion process. A democratic leader, however, prompted interaction owing to the open communicative atmosphere that he or she created.

Differences in the perceived academic ability or knowledge level show as academic status differences between students. Academic status is a powerful status characteristic in the classroom because of its obvious relevance to the classroom activities (Cohen, 1994). According to Cohen (1994), students with a higher status easily dominate group interaction and are more likely to be perceived as leaders. Knowledge differences do not even have to be real in the actual situation to raise differences in the roles assumed by the participators, as Cohen (1994) has noted. Differences in the persons' general academic status may

affect interaction; therefore, the power order of the group reflects the initial differences in status, even if the participants share the same level of knowledge in the actual situation, or if the task does not require the academic ability in question. Thus, students with academically high status are expected to be more competent than others, and this status may be generalized to a wide range of school tasks.

Peer status – that is, perceived attractiveness or popularity – may also act as a basis for inequalities in participation in the context of collaborative activity (Cohen, 1994). Popular children tend to get their voice better heard in the group work context. However, the study by Murphy and Faulkner (2000) with preschool children also demonstrated that unpopular children benefited from interacting with popular children. The former had more elaborated disagreements when interacting with the latter than when interacting with other unpopular children. Cohen (1994) relates popularity to academic status, but it has been found (Juvonen and Wenzel, 1996) that, at secondary-level education, popularity depends on qualities other than academic ability, whereas, at the primary level, popularity is often associated with academic ability. Social status factors such as gender can also affect interaction in the collaborative situation. For example, some studies have demonstrated that boys tend to dominate the work in science tasks (Petersen *et al.*, 1991) as well as in computer tasks (Underwood and Underwood, 1990) based on their alleged superiority compared to girls (Underwood and Underwood, 1990). There are contradictory results about the affect of gender *per se* on students' collaboration. Some studies indicate that the interaction of same-gender pairs is qualitatively better and more collaborative compared to mixed-gender pairs (Fitzpatrick and Hardman, 2000; Tolmie and Howe, 1993). Howe and Tolmie (1999) suggest that social unease occasioned by gender can have an influence on the qualitatively lower-level interaction of mixed-gender pairs. However, some studies also fail to demonstrate significant differences in interaction according to gender pairing (e.g., Howe and Tolmie, 1999). These mixed findings can be due to different reasons; a difference in tasks and instructions, in the way performance is measured, as well as a difference in wider classroom culture and context (Fitzpatrick and Hardman, 2000). In other words, these findings can be attributable to the difficulty of gaining context-free results.

Barron (2003) further argues that research on status has concentrated excessively on studying status as a static phenomenon. Therefore, such research has failed to reflect the complexity of interaction and account for dynamic shifts in students' power relations during interaction, as was observed in Barron's (2003) study. In addition, Arvaja *et al.* (2002) noticed in their case study on secondary school students' collaborative science task that the knowledge power held by one student was triggered only occasionally. Hence, it was dynamic in a sense that it did not turn into a permanent role, but was negotiated

throughout the students' activity. Knowledge asymmetry was manifested either in tutoring or a leadership role, which was also in relation with the students' friendship status. The student group consisted of four students: two boys and two girls. The girls were best friends, whereas the boys were classroom peers. It seemed that the non-constructive leadership occurred only when the whole group was present. Altogether, it seemed that when the girls were working with the boys, the social tension between group members contributed to hasty, nonreflective discussion and nonengagement with the task itself. The interaction between the girls was more equal when they were not working with the boys. Only the girls, who were friends, were able to really reason, explain, wonder, and argue, and were able to engage in the problem solving and the task at hand. The only asymmetrical interaction between the girls was tutoring, which, however, promoted collaboration and shared knowledge construction.

In addition, the nature of the task may affect the power relations of the group. According to Cohen (1994), complex tasks with ill-defined problems can be featured as group tasks. A group task is one that requires resources (e.g., knowledge and heuristic problem-solving strategies and skills) that no single individual possesses and is, therefore, probably unable to accomplish the task alone (Cohen, 1994). In group tasks, students are interdependent in a reciprocal fashion (Cohen, 1994). Thus, exchanging ideas and information is vital to success. However, problems with one identifiable correct answer could often be accomplished by single individuals and the interaction would be, in its nature, more like helping each other understand concepts without a need for deeper-level discourse. In addition, according to Chizhik (2001), a single-answer task more easily leads to one participant, perhaps a more knowledgeable person (Cohen, 1994), dominating the discussion. Instead, variable-answer tasks, which elicit knowledge from a wide subject domain, increase the possibility of many participants contributing to the discourse. In another study, Chizhik (1999) found out that gender differences in the participating group's working also diminished in variable-answer tasks.

Summary

Collaborative learning is a fashionable phenomenon nowadays; however, collaboration among students in various learning settings (e.g., in classrooms) is a much more complex phenomenon than has often been thought. While aiming to understand the diverse viewpoints to collaborative learning, we have to consider an extremely complex set of variables, that is, cognitive, social, emotional, motivational, and contextual ones, interacting with each other in a systemic and dynamic manner. This article has particularly focused on the social aspects of collaborative learning.

It has become clear that the line between individual and social processes of learning is blurring, and the main message of many researchers is that we should see individual minds in interaction with group understandings (Stahl, 2004). It is important to support not only the content, but also the social level of interaction (see Barron, 2003). It seems that learners face a dual problem space of this kind as they are supposed to work and learn collaboratively. Furthermore, collaborative situations reflect previous social activities, and are transformed by current interactions and projections of the future. In this process, social relationships deriving from friendship or status are also in a crucial role. For example, Azmitia (1998) has stated that qualities such as mutuality and trust, often associated with friendship, are beneficial in negotiating shared understanding.

Most of the results cited in this article come from studies where relationships – such as status or friendship – in relation to collaboration are explored experimentally by manipulating the group composition according to different status or friendship qualities. The gain for learning is measured in posttests at the individual level. These studies fail to reflect the complexity of interaction and dynamics in shifts in students' relations during interaction. There is a great need for studies that focus on social relationships from the sociocultural perspective, illuminating their dynamic, changing, and negotiated nature in the context of collaboration as well as their historical nature in the form of practices and relations (Arvaja, 2007).

See also: Children's Friendship; Peer Learning in the Classroom; Social Interaction and Learning.

Bibliography

- Arvaja, M. (2007). Contextual resources in meaning negotiations of a student pair in a web-based history project. *International Journal of Educational Research* **46**(3/4), 215–228.
- Arvaja, M., Häkkinen, P., Rasku-Puttonen, H., and Eteläpelto, A. (2002). Social processes and knowledge building during small group interaction in a school science project. *Scandinavian Journal of Educational Research* **46**(2), 161–179.
- Azmitia, M. (1998). Peer interactive minds. Developmental, theoretical and methodological issues. In Faulkner, D., Littleton, K., and Woodhead, M. (eds.) *Learning Relationships in the Classroom*, pp 207–233. London: Routledge.
- Azmitia, M. (2000). Taking time out from collaboration: Opportunities for synthesis and emotion regulation. In Joiner, R., Littleton, K., Faulkner, D., and Miell, D. (eds.) *Rethinking Collaborative Learning*, pp 179–195. London: Free Association Books.
- Azmitia, M. and Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning. *Social Development* **2**(3), 202–221.
- Baker, M. (2002). Forms of cooperation in dyadic problem-solving. In Salembier, P. and Bencheikroun, H. (eds.) *Cooperation and Complexity*, vol. 16, pp 587–629. Paris: Hermès.
- Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences* **9**(4), 403–436.
- Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences* **12**(3), 307–359.
- Basili, P. and Sandford, J. (1991). Conceptual change strategies and cooperative group work in chemistry. *Journal of Research in Science Teaching* **28**(4), 293–304.
- Berkowitz, M. and Gibbs, J. (1985). The process of moral conflict resolution and moral development. In Berkowitz, M. (ed.) *Peer Conflict and Psychological Growth*, pp 71–84. San Francisco, CA: Jossey Bass.
- Berndt, T., Perry, T., and Miller, K. (1988). Friends' and classmates' interactions on academic tasks. *Journal of Educational Psychology* **80**, 506–513.
- Chizhik, A. (1999). Can students work together equitably? An analysis of task effects in collaborative group work. *Social Psychology of Education* **3**(1–2), 63–79.
- Chizhik, A. (2001). Equity and status in group collaboration: Learning through explanations depends on task characteristics. *Social Psychology of Education* **5**(2), 179–200.
- Clark, H. (1996). *Using Language*. New York: Cambridge University Press.
- Cohen, E. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research* **64**(1), 1–35.
- Costin, S. and Jones, D. (1992). Friendship as a facilitator of emotional responsiveness and prosocial interventions among young children. *Developmental Psychology* **28**, 941–947.
- Crook, C. (1999). Computers in the community of classrooms. In Littleton, K. and Light, P. (eds.) *Learning with Computers. Analysing Productive Interaction*, pp 102–117. London: Routledge.
- Crook, C. (2002). Deferring to resources: Collaborations around traditional vs. computer-based notes. *Journal of Computer Assisted Learning* **18**(1), 64–76.
- Dillenbourg, P. (1999). Introduction: What do you mean by collaborative learning? In Dillenbourg, P. (ed.) *Collaborative Learning: Cognitive and Computational Approaches*, pp 1–19. Oxford: Pergamon.
- Elbers, E. and Streefland, L. (2000). Collaborative learning and the construction of common knowledge. *European Journal of Psychology of Education* **15**(4), 479–490.
- Faulkner, D. and Miell, D. (2004). Collaborative story telling in friendship and acquaintanceship dyads. In Littleton, K., Miell, D., and Faulkner, D. (eds.) *Learning to Collaborate: Collaborating to Learn*, pp 7–29. New York: Nova Science.
- Fitzpatrick, H. and Hardman, M. (2000). Mediated activity in the primary classroom: Girls, boys and computers. *Learning and Instruction* **10**(5), 431–446.
- Hartup, W. (1996). Cooperation, close relationships and cognitive development. In Bukowski, W., Newcomb, A., and Hartup, W. (eds.) *The Company They Keep: Friendships and Their Developmental Significance*, pp 213–237. New York: Cambridge University Press.
- Howe, C. and Tolmie, A. (1999). Productive interaction in the context of computer-supported collaborative learning in science. In Littleton, K. and Light, P. (eds.) *Learning with Computers*, pp 24–45. London: Routledge.
- Juvonen, J. and Wenzel, K. (1996). *Social Motivation: Understanding Children's School Adjustment*. New York: Cambridge University Press.
- Kumpulainen, K. and Mutanen, M. (1999). The situated dynamics of peer group interaction: An introduction to analytic framework. *Learning and Instruction* **9**(5), 449–473.
- Kutnick, P. and Kington, A. (2005). Children's friendships and learning in school: Cognitive enhancement through social interaction? *British Journal of Educational Psychology* **75**, 521–538.
- Kutnick, P. and Manson, L. (1998). Social life in the primary school: Towards a relational concept of social skills for use in the classroom. In Campbell, A. and Muncer, S. (eds.) *The Social Child*, pp 165–187. UK: Psychology Press.
- Linn, M. and Burbules, N. (1993). Construction of knowledge and group learning. In Tobin, K. (ed.) *The Practice of Constructivism in Science Education*, pp 91–119. Washington, DC: AAAS Press.
- Littleton, K. and Häkkinen, P. (1999). Learning together: Understanding the processes of computer-based collaborative learning. In Dillenbourg, P. (ed.) *Collaborative Learning: Cognitive and Computational Approaches*, pp 20–31. Oxford: Pergamon.

- MacDonald, R., Miell, D., and Morgan, L. (2000). Social processes and creative collaboration in children. *European Journal of Psychology of Education* **15**(4), 405–415.
- Mercer, N. (1996). The quality of talk in children's collaborative activity in classroom. *Learning and Instruction* **6**(4), 359–377.
- Miell, D. and MacDonald, R. (2000). Children's creative collaborations: The importance of friendship when working together on a musical composition. *Social Development* **9**(3), 348–364.
- Murphy, S. and Faulkner, D. (2000). Learning to collaborate: Can young children develop better communication strategies through collaboration with a more popular peer. *European Journal of Psychology of Education* **15**(4), 389–404.
- Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In Salomon, G. (ed.) *Distributed Cognitions. Psychological and Educational Considerations*, pp 47–87. Cambridge: Cambridge University Press.
- Petersen, R., Johnson, D., and Johnson, R. (1991). Effects of co-operative learning on perceived status of male and female pupils. *Journal of Social Psychology* **131**(5), 717–735.
- Rafal, C. (1996). From co-construction to takeovers: Science talk in group of four girls. *Journal of the Learning Sciences* **5**(3), 279–293.
- Richmond, G. and Striley, J. (1996). Making meaning in classrooms: Social processes in small-group and scientific knowledge building. *Journal of Research in Science Teaching* **33**(8), 839–858.
- Roschelle, J. and Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In O'Malley, C. (ed.) *Computer Supported Collaborative Learning, NATO ASO Series F: Computer and System Sciences*, vol. 128, pp 69–97. Berlin: Springer-Verlag.
- Sfard, A. and Kieran, C. (2001). Cognition as communication: Rethinking learning-by-talking through multi-faceted analysis of students' mathematical interactions. *Mind, Culture and Activity* **8**(1), 42–76.
- Stahl, G. (2004). Building collaborative knowing. Elements of a social theory of CSCL. In Dillenbourg, P., Strijbos, J. W., Kirschner, P. A., and Martens, R. L. (eds.) *Computer-Supported Collaborative Learning What We Know About CSCL and Implementing It in Higher Education*, Vol. 3. Boston, MA: Kluwer.
- Tolmie, A. and Howe, C. (1993). Gender and dialogue in secondary school physics. *Gender and Education* **5**(2), 191–209.
- Underwood, J. and Underwood, G. (1990). *Computers and Learning: Helping Children Acquire Thinking Skills*. Oxford: Blackwell.
- Van Boxtel, C. (2000). *Collaborative Concept Learning. Collaborative Learning Tasks, Students Interaction and the Learning of Physics Concepts*. Dissertation: Utrecht University.
- Wegerif, R. and Mercer, N. (1997). A dialogical framework for researching peer talk. In Wegerif, R. and Scrimshaw, P. (eds.) *Computers and Talk in Primary Classroom*, pp 49–61. Clevedon: Multilingual Matters.

Social Development and Schooling

G W Ladd, Arizona State University, Tempe, AZ, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Friendship – A voluntary, dyadic peer relationship that is typically characterized by reciprocal liking, preference for each other's company, frequent association or interaction, and a positive affective tie between the partners.

Peer-group acceptance/rejection – A concept that refers to children's relations with members of their peer group (e.g., classmates) and is defined in terms of group member's sentiments (i.e., liking vs. disliking) toward the child, and the degree to which these sentiments become consensual. Measures of this construct indicate the extent to which a child is liked versus disliked by members of his or her peer group (e.g., classmates).

Peer victimization – A form of peer relationship in which children are frequently targeted by peers for verbal and/or physical aggression, or other forms of harassment or abuse.

Attempts to describe the demands of schooling and children's educational readiness suggest that classroom environments present children with a complex array of scholastic and interpersonal challenges, all of which require some degree of adaptation on the part of the child (see Ladd, 2003). These diverse challenges – depending on the child and the resources or constraints that are operating in the child's environment – may affect children on many levels and have both immediate and long-term effects on their development and performance in the school context.

Schooling Presents Children with Social as well as Scholastic Challenges

Some of the most obvious challenges of schooling are scholastic in nature, and stem from the instructional features of classrooms, such as didactic small- and large-group instruction, teacher-initiated/monitored learning activities, and programmatic curriculum sequences. Less well recognized are the many types of interpersonal challenges that children confront in school.

During the transition into formal schooling, children typically are faced with shifting social ecologies,

relationships, and resources. Studies show that, as children enter school, most are thrust into new peer groups with unfamiliar classmates (see Ladd and Price, 1987). As the adult-to-peer ratio is often smaller in classrooms than it is in family or preschool environments, interactions with peers become more common than those with adults, and competition for adult attention increases. Opportunities for mixed-age companionship are rare because, unlike family or neighborhood environments, grade school classrooms are segregated by age. The result is that, once in school, children are under greater pressure to succeed among equals. They must succeed at forming relationships with age-mates while also competing with peers for resources and recognition.

In addition to the demands of the classroom peer milieu, children are confronted with the challenge of forming relationships with their teachers. During the transition to grade school, children often lose ties that were formed with preschool or day-care teachers and embark on a relationship with their grade-school teachers. Initially, this relationship may be a source of care giving, but over time, the teacher's primary role is to challenge children in ways that promote learning and development.

Many of these same challenges are repeated as children progress through the grades. In each new classroom they must negotiate their needs in dyadic and group interactions and reestablish or form new relationships with classmates and teachers. Moreover, it is likely that these challenges are intensified when children change schools, or cope with school transitions (see Ladd *et al.*, 2006).

Do Social Relations Play a Role in How Children Adjust and Perform in School?

Although considerable evidence has been gathered on the precursors of children's school adjustment, most of what has been discovered points to the importance of children's cognitive and linguistic skills, their physical-motor skills, and their socioeconomic and ethnic backgrounds. Only recently have researchers systematically explored children's classroom social behavior and relationships as predictors of their school adjustment. This is surprising when one considers that education is, in many respects, a social enterprise. Most modes of instruction require that teachers and students communicate and engage in social interaction. Likewise, educators have become increasingly reliant on peer-mediated activities (e.g., peer collaboration and

tutoring and cooperative learning groups) to promote classroom learning and achievement.

Fortunately, recent theory and evidence on the interpersonal foundations of learning and achievement has elevated this topic's importance within the educational community (e.g., see Hamre and Pianta, 2001; Ladd, 2003). As a result, greater investigative attention has been devoted to the hypothesis that social relations in the school context may influence multiple aspects of children's school adjustment. In particular, it has been proposed that classroom social relations may be instrumental in shaping children's: (1) perceptions and appraisals of school and the classroom environment, (2) psychological and emotional reactions in this context, including both internalizing and externalizing problems, (3) involvement and disengagement in classroom activities, and (4) achievement and academic progress (see Ladd, 2003).

Are Children's Behaviors with Classmates and Teachers Associated with Their School Adjustment?

It has been postulated that the way in which children behave toward classmates and teachers has an important bearing on how they will adapt to school, and progress academically in this context (Ladd, 2003). A key premise within most of these perspectives is that many of the processes and conditions that are essential for learning, such as curiosity, interest, attention, motivation, perceived support, and feelings of competence and security, develop from the interactions children have with their classmates and teachers.

Most of the child behaviors that have been examined as correlates or antecedents of school adjustment can be classified into three categories, termed withdrawn behavior (e.g., avoiding others; being unsociable), prosocial behavior (e.g., cooperating with others), and aggressive behavior (e.g., harming others). Each of these forms of behavior has been linked with children's early and later school adjustment.

Children Who Avoid, Withdraw from, or Remain Solitary among Classmates

Although early studies produced mixed findings, recent evidence suggests that withdrawn children have difficulty adapting to school (see Ladd, 2005). Several subtypes of withdrawn children have been studied (e.g., asocial-withdrawn, anxious-withdrawn, active-isolated, depressed-withdrawn, and aggressive-withdrawn children; see Harrist *et al.*, 1997; Ladd and Burgess, 1999), and it appears that some forms of withdrawn behavior impose greater risk for school adjustment problems than others. During the grade school years, children who fit the depressed-withdrawn subtype tend to exhibit elevated rates of peer neglect and rejection (Harrist *et al.*, 1997), and children who manifest

stable patterns of anxious-withdrawal exhibit higher levels of depression (Gazelle and Ladd, 2003). Aggressive-withdrawn children also appear to be at risk for several types of school adjustment problems. Ladd and Burgess (1999) followed children who displayed two forms of withdrawal – asocial-withdrawn and aggressive-withdrawn behavior – from kindergarten to grade 2 and compared them to samples of nonwithdrawn-aggressive children and normative, matched controls. Across these grades, children in the aggressive-withdrawn group were more likely than normative, matched controls to become rejected by classmates and form teacher–child relationships that were high in conflict and dependency and low in closeness. The relationship difficulties of asocial-withdrawn children, in contrast, were found to be more transient. These children exhibited more dependent relationships with their teachers as they began kindergarten, but not thereafter.

Children Who Are Prosocial or Aggressive toward Others

Children's propensity to engage in prosocial behavior in the classroom represents another potential determinant of their school adjustment. Ladd *et al.* (1999) found that kindergartners who often interacted prosocially with classmates during the first 10 weeks of kindergarten tended to develop mutual friends and had higher peer acceptance by week 14. Other data show that as early as preschool and kindergarten, aggression is common in classrooms and is a significant predictor of later school adjustment. Ladd and Price (1987), for example, found that children who were aggressive toward many rather than a few classmates in preschool classrooms were more likely to develop social difficulties after they entered kindergarten. Similarly, Ladd and Burgess (1999) found that aggressive kindergartners were more likely than their counterparts in a matched risk-comparison group to develop and maintain social difficulties with classroom peers and teachers throughout the early primary grades.

Even more compelling were findings reported by Ladd and Burgess (2001). These investigators reported that aggressive kindergartners exhibited gains in thought problems, misconduct, and classroom disengagement, and were prone toward underachievement and the formation of negative school attitudes as they progressed into the primary school years. Furthermore, children who tended to be chronically aggressive (i.e., remained aggressive across more than 1 year of grade school) were even more likely to exhibit serious or severe forms of school maladjustment.

Findings from research on middle- and high school children further substantiate the link between aggressive behavior and school adjustment difficulties. For example, Kupersmidt and Coie (1990) followed children from grades 5–10 and found that aggressiveness among fifth graders forecasted higher levels of delinquency among

10th graders, and that both aggression and absences were associated with later dropout rates. Other findings suggest that aggression is a better predictor of dropping out of school for boys than for girls (see Ladd, 2005).

Children's propensities to engage in withdrawn, prosocial, and aggressive behaviors also have been linked with features of the teacher-child relationship. In a longitudinal study, Birch and Ladd (1998) found that children's aggressive behavior in kindergarten predicted higher levels of conflict and lower levels of closeness in their relationships with kindergarten and first-grade teachers. Although kindergartners' prosocial interaction styles were associated with teacher-child closeness in both kindergarten and grade 1, this link was not as robust as that found between aggressive styles and teacher-child conflict.

Are Children's Relationships with Classmates and Teachers Associated with Their School Adjustment?

Another line of research has been guided by the premise that children's relationships with classmates and teachers immerse them in processes (e.g., participation vs. exclusion, support vs. conflict, and receiving assistance vs. being ignored) that affect their ability to adapt to school challenges which, ultimately, affects their development in this context (amount of learning, level of school engagement; increases or decreases their sense of worth, competence, etc.; see Ladd, 2003, 2005). As relationships bring different processes to bear upon children and confer different provisions, they vary in adaptive significance for school-related demands (Ladd *et al.*, 1997). In the next three sections, the adaptive significance of several types of classroom relationships is considered.

Classroom Peer Acceptance and Rejection

A growing corpus of findings link children's acceptance or rejection by classroom peers with indicators of their school adjustment (see Ladd, 2005). Early peer rejection – at school entry – has been shown to predict problems such as negative school attitudes, school avoidance, and underachievement during the first year of schooling and thereafter (Ladd, 1990; Ladd *et al.*, 1999; Ladd and Burgess, 2001). Later, in the elementary years, peer acceptance has been linked with loneliness (Parker and Asher, 1993), conduct problems (Ladd, 2006), lower emotional well-being (Ladd, 2006), and academic deficits (Ladd *et al.*, 1997).

Researchers have also attempted to distinguish the contributions of peer acceptance and rejection from those of other classroom relationships. Ladd *et al.* (1997, 1999) found that, even after controlling for other forms of peer and teacher-child relationships, peer rejection predicted

children's participation in the classroom which, in turn, was linked to later achievement. In a similar study, Buhs and Ladd (2001) found that children's peer acceptance at school entry predicted changes in classroom participation which, in turn, predicted later academic and emotional adjustment. In general, these results support the premise that peer acceptance promotes social inclusion which, in turn, yields provisions (e.g., sense of belongingness and engagement in learning activities) that enhance children's interpersonal and scholastic adjustment (Buhs *et al.*, 2006).

Classroom Friendships

Investigators have studied several indicators of classroom friendships, including children's participation in a close friendship, the number of mutual friends they have in their classrooms, the duration of these relationships, and features that reflect the quality of a friendship (see Ladd, 2005). There is growing evidence linking one or more of these facets of friendship to children's school adjustment.

As children enter school, those who find prior friends or form new friendships in their classrooms tend to form favorable school perceptions and do better academically than peers with fewer friends (Ladd, 1990). The processes that typify friends' interactions have also been linked to children's school adjustment. For example, Ladd *et al.* (1996) detected variability in the quality of the friendships that children formed as they entered school, and found that children who saw their friendships as offering higher levels of support and aid tended to see their classrooms as supportive interpersonal environments. Conversely, children (especially boys) who reported higher levels of conflict in their friendships exhibited lower levels of classroom participation. Likewise, it appears that friends facilitate adjustment as children progress through grade school. Findings from a study conducted with third through fifth graders showed that children with supportive friends felt less lonely in school (Parker and Asher, 1993).

Although less well understood, there is some evidence to suggest that friendships do not always contribute positively to children's school adjustment. Berndt and Keefe (1995), for example, found that fighting and disruptiveness in school tended to increase if adolescents had stable friendships with peers who exhibited the same problems. While far from being conclusive or exhaustive, these studies suggest that, in addition to peer-group acceptance, the features of children's friendships (support, companionship, stability, etc.) are potential antecedents of school adjustment across a wide range of ages.

Peer Victimization

The probability that children will encounter peer abuse, or become the targets of bullying or other forms of peer

harassment, increases as they enter school and progress through the primary grades, and evidence indicates that victimized children develop a variety of school adjustment problems (see Ladd, 2005). It has been posited that frequent harassment leads children to become so preoccupied with fears, feelings of social alienation, and safety concerns that they have difficulty attending to school tasks and, eventually, develop negative school attitudes or higher levels of school avoidance.

During the early school years, children who are exposed to higher levels of peer victimization display increases in loneliness and school avoidance (Ladd *et al.*, 1997), and these difficulties tend to become more pronounced if children are chronically maltreated. Kochenderfer-Ladd and Wardrop (2001) found that children who were exposed to victimization over longer intervals between kindergarten and third grade were more likely to feel lonely in school and less satisfied with their classroom peer relationships. Children who were victimized during the early grades but not thereafter did not always recover, or show improvements in their adjustment. These longitudinal findings have been corroborated by cross-sectional evidence gathered with diverse age groups around the world (see Ladd, 2005). Data gathered in many different cultures suggest that victims of peer maltreatment are more likely than nonvictims to report negative feelings and attitudes toward school and classroom tasks (Ladd, 2005; Ladd *et al.*, 1997). Although further investigation is needed, especially across age groups, gender and school contexts, the bulk of extant evidence conforms to the hypothesis that victimization contributes to a number of school-related difficulties.

Teacher–Child Relationships

It has been argued that close, rather than conflictual or dependent, teacher–child relationships benefit children and support their progress in school (see Birch and Ladd, 1998; Hamre and Pianta, 2001). Evidence gathered thus far shows that, among kindergarten children who were at risk for retention, those who had positive teacher–child relationships were more likely to be promoted than were those with negative teacher–child relationships (Pianta and Steinberg, 1992). Similarly, in a longitudinal study, Birch and Ladd (1998) found that kindergartners with conflictual or dependent teacher–child relationships were more likely to develop later adjustment problems, such as delayed academic performance, lower classroom participation, and negative school attitudes. Ladd *et al.* (1999) also found that qualities of the teacher–child relationship predicted later classroom participation and, indirectly, academic achievement. While these findings have yet to be replicated with older samples and varying demographic groups, available evidence implicates the teacher–child relationship as a potential antecedent of children's school adjustment.

Contributions of Multiple Classroom Relationships in Children's School Adjustment

In recent years, research on individual classroom relationships has been supplemented by studies in which investigators gather data on multiple relationships, and examine the relative (differential) contributions of these ties to children's school adjustment. Initial efforts to investigate differential relationship contributions were focused on friendship and peer acceptance (e.g., see Parker and Asher, 1993).

Studies of grade-school children suggest that classroom friendships and peer-group acceptance make distinct contributions to the prediction of socioemotional adjustment and academic competence (Parker and Asher, 1993). With young children, Ladd (1990) found that friendship and peer acceptance uniquely predicted changes in kindergartner's school perceptions, avoidance, and performance. In another study (Ladd *et al.*, 1997), the contributions of several types of peer relationships were examined after adjusting for shared predictive linkages, and some relationships were found to be better predictors of children's school adjustment than others. Peer victimization, for example, predicted gains in children's loneliness above and beyond associations that were attributable to friendship and peer-group acceptance. In contrast, peer-group acceptance uniquely predicted improvements in children's achievement. Overall, these findings were consistent with the view that peer relationships are both specialized in the types of resources or constraints they create for children, and also diverse in the sense that some resources may be found in more than one form of relationship.

Are Children's Classroom Behaviors and Relationships Distinct or Interdependent Precursors of School Adjustment?

With the advent of child and environment theories (see Ladd, 2003), researchers began to investigate multiple antecedents of adaptation and, in particular, the interface between children's classroom behavior and relationships as predictors of their school adjustment. Although studies of this type are still few in number, extant evidence tends to corroborate the view that children's behavioral propensities combine with their involvement in specific types of classroom relationships (and relationship experiences) to predict school adjustment (see Ladd, 2003). In two prospective longitudinal studies, Ladd *et al.* (1999) found significant, direct paths from children's behavior to their classroom relationships, and from these relationships to their participation in classroom activities, even after controlling for entry factors that are known to predict school engagement (e.g., family backgrounds, ethnicity, child's gender, intelligence quotient (IQ)). The strongest of these paths indicated that aggressive behaviors anteceded the formation of negative classroom relationships (e.g., peer rejection and

teacher–child conflict) and, in turn, these negative classroom relationships functioned as an impediment to children's engagement in classroom social and scholastic activities. In addition, consistent with past research on the antecedents of scholastic progress, direct, positive pathways were found from classroom participation to achievement.

In another longitudinal study, Ladd and Burgess (2001) sought to determine whether aggressive children's participation in different types of classroom relationships might increase (e.g., exacerbate) or decrease (i.e., compensate for) their probability of developing psychological and school adjustment problems. Results implied that aggressive children's school adjustment difficulties became more or less serious depending on whether they were accepted or rejected by their classmates. On the one hand, peer-group rejection was associated with increases in aggressive children's thought problems and decreases in their classroom participation, positive school attitudes, and achievement. On the other hand, peer-group acceptance was associated with relative declines in aggressive children's attention problems and misconduct, and relative gains in cooperative participation and school liking.

In addition, researchers have examined whether children's school adjustment varies as a function of the persistence of their behavior patterns and the stability of their participation in classroom relationships. Ladd and Burgess (2001) assessed the duration of children's risk status for aggression and stressful peer relationships as they progressed through the first 2 years of grade school. Compared to early-onset indicators, the chronicity of children's aggressive risk status across grades predicted changes in a host of school adjustment criteria, including increases in attention problems, thought problems, and behavioral misconduct, and decreases in cooperative classroom participation and academic achievement. After controlling for initial peer-group rejection and the chronicity of children's aggressive risk status and other relational risks, the chronicity of peer-group rejection also predicted many of the same forms of school maladjustment that were associated with aggression. These findings corroborated the inference that the predictive contributions of aggressive behavior to children's school adjustment are exacerbated by chronic relational risks.

Taken together, these findings strengthen the credibility of models in which it is assumed that both behavioral risks and classroom relational experiences affect emerging patterns of school adjustment. In addition to children's actions in the classroom, the relationships they form in this context appear to have an important bearing on their school adjustment.

Conclusions and Future Directions

During the school years, it appears that a number of connections develop among children's classroom behavior,

relationships, and school adjustment. Several are worthy of comment, and warrant further investigation.

Behavioral Precursors of School Adjustment

Available evidence suggests that the behaviors children utilize to address classroom challenges may elicit specific environmental reactions that make it more or less likely that they will learn, adjust, and progress in this context. Pertinent to this inference are findings indicating that children's propensities to withdraw versus participate socially in classrooms, or engage in prosocial versus aggressive behaviors, may impact the level and form of their involvement in classroom relationships and activities and, thus, contribute to their school adjustment.

Children who withdraw from classroom interactions appear to restrict their opportunities to participate in social and scholastic activities, and isolate themselves from resources that can be found in classroom relationships (support, affirmation, aid, etc.). Among withdrawn children, those who have been identified as also being depressed, aggressive, or anxious, appear to be at greatest risk for subsequent school maladjustment. Aggressive-withdrawn kindergartners, in particular, tend to develop social difficulties with classmates and teachers, and experience a variety of school adjustment problems as they progress through school.

Children who frequently engage in aggressive behaviors appear to elicit reactions from peers (e.g., ignoring and exclusion) and teachers (e.g., conflictual interactions) that reduce their constructive participation in classroom activities. Such environmental restrictions may adversely affect learning and achievement, and promote the development of negative attitudes toward school, classroom activities, peers, and teachers. Whereas children with prosocial behavioral styles exhibit positive social and scholastic outcomes during kindergarten and thereafter (e.g., supportive peer and teacher–child relationships and greater engagement in classroom activities), those with aggressive styles are more likely to develop incipient and persistent forms of school maladjustment (e.g., loneliness, peer rejection, teacher–child conflict, and classroom misconduct).

Relational Precursors of School Adjustment

As we have seen, the adaptive significance of children's peer and teacher relationships was also investigated independently and conjointly with children's classroom behavior, and this corpus of evidence implied that classroom relationships have an important bearing on several aspects of school adjustment. Both peer and teacher–child relationships appear to be important in this regard.

Evidence indicates that classroom peer relationships are specialized in the types of resources or constraints they create for children. For example, in contrast to classroom

friendships or peer-group acceptance, peer victimization appears to be a stronger influence on the development of loneliness and school avoidance. Peer-group acceptance, by comparison, appears to have greater sway over children's participation in classroom activities and subsequent achievement.

It also appears that the contributions of peer relationships to children's adjustment depend not only on the functional properties of particular peer relationships (e.g., the resources or constraints they confer upon children), but also on the duration of children's participation in these relationships (i.e., their history of exposure to specific relationship processes). Extant findings indicate that children's risk for maladjustment is increased by longer exposures to relational adversities and decreased by stable relational supports.

Conclusions and Future Directions

Available evidence suggests that both children's behavior toward classmates and teachers, and the relationships they form with these persons, play an important role in establishing the social, intellectual, and psychological conditions that underlie school adjustment and scholastic progress. Ultimately, it is the combination of these processes that appears to be most prognostic of children's school adjustment. Much remains to be learned, however, about the role of children's classroom interpersonal behavior and relationships as precursors of school adaptation and adjustment. Given the complexity of these phenomena, there is a need to construct models that will focus the search for interpersonal antecedents, and provide a context for understanding how these factors impact children's school adjustment. It will also be important to consider not only how specific interpersonal factors exert an influence on adjustment, but also which aspects of children's school adjustment are affected by these factors.

Acknowledgments

This work was supported by the National Institutes of Health (R01HD-045906).

Bibliography

- Birch, S. H. and Ladd, G. W. (1998). Children's interpersonal behaviors and the teacher-child relationship. *Developmental Psychology* **34**, 934-946.
- Buhs, E. S. and Ladd, G. W. (2001). Peer rejection as an antecedent of young children's school adjustment: An examination of mediating processes. *Developmental Psychology* **37**, 550-560.
- Buhs, E. S., Ladd, G. W., and Herald, S. (2006). Peer exclusion and victimization: Processes that mediate the relation between peer

- group rejection and children's classroom engagement and achievement? *Journal of Educational Psychology* **98**, 1-13.
- Gazelle, H. and Ladd, G. W. (2003). Anxious solitude and peer exclusion: A diathesis-stress model of internalizing trajectories in childhood. *Child Development* **74**, 257-278.
- Hamre, B. K. and Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development* **72**, 625-638.
- Harrist, A. W., Zaia, A. F., Bates, J. E., Dodge, K. A., and Pettit, G. S. (1997). Subtypes of social withdrawal in early childhood: Sociometric status and social-cognitive differences across four years. *Child Development* **68**, 278-294.
- Kochenderfer-Ladd, B. and Wardrop, J. (2001). Chronicity and instability in children's peer victimization experiences as predictors of loneliness and social satisfaction trajectories. *Child Development* **72**, 134-151.
- Kupersmidt, J. B. and Coie, J. D. (1990). Preadolescent peer status, aggression, and school adjustment as predictors of externalizing problems in adolescence. *Child Development* **61**, 1350-1362.
- Ladd, G. W. (1990). Having friends, keeping friends, making friends, and being liked by peers in the classroom: Predictors of children's early school adjustment? *Child Development* **61**, 1081-1100.
- Ladd, G. W. (2003). Probing the adaptive significance of children's behavior and relationships in the school context: A child by environment perspective. In Kail, R. (ed.) *Advances in Child Behavior and Development*, vol. 31, pp 43-104. New York: Wiley.
- Ladd, G. W. (2005). *Children's Peer Relationships and Social Competence: A Century of Progress*. New Haven, CT: Yale University Press.
- Ladd, G. W. (2006). Peer rejection, aggressive or withdrawn behavior, and psychological maladjustment from ages 5 to 12: An examination of four predictive models. *Child Development* **77**, 822-846.
- Ladd, G. W. and Burgess, K. B. (1999). Charting the relationship trajectories of aggressive, withdrawn, and aggressive/withdrawn children during early grade school. *Child Development* **70**, 910-929.
- Ladd, G. W. and Price, J. M. (1987). Predicting children's social and school adjustment following the transition from preschool to kindergarten. *Child Development* **58**, 1168-1189.
- Ladd, G. W., Kochenderfer, B. J., and Coleman, C. C. (1996). Friendship quality as a predictor of young children's early school adjustment. *Child Development* **67**, 1103-1118.
- Ladd, G. W., Kochenderfer, B. J., and Coleman, C. C. (1997). Classroom peer acceptance, friendship, and victimization: Distinct relational systems that contribute uniquely to children's school adjustment? *Child Development* **68**, 1181-1197.
- Ladd, G. W., Birch, S. H., and Buhs, E. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development* **70**, 1373-1400.
- Ladd, G. W., Herald, S. L., and Kochel, K. P. (2006). School readiness: Are there social prerequisites? *Early Education and Development* **17**, 115-150.
- Parker, J. G. and Asher, S. R. (1993). Friendship and friendship quality in middle childhood: Links with peer group acceptance and feelings of loneliness and social dissatisfaction. *Developmental Psychology* **29**, 611-621.
- Pianta, R. C. and Steinberg, M. (1992). Teacher-child relationships and the process of adjusting to school. *New Directions for Child Development* **57**, 61-80.

Further Reading

- Cairns, R. B. and Cairns, B. D. (1994). *Lifelines and Risks: Pathways of Youth in Our Time*. New York: Cambridge University Press.
- Damon, W. and Phelps, E. (1989). Strategic uses of peer learning in children's education. In Berndt, T. J. and Ladd, G. W. (eds.) *Peer Relationships in Child Development*, pp 135-157. New York: Wiley.

- Eccles, J. S., Wigfield, A., and Schiefele, U. (1998). Motivation to succeed. In Damon, W. and Eisenberg, N. (eds.) *Handbook of Child Psychology*, 5th edn., vol. 5, pp 1017–1095. New York: Wiley.
- Perry, K. E. and Weinstein, R. S. (1998). The social context of early schooling and children's school adjustment. *Educational Psychologist* **33**, 177–194.
- Vandell, D. L. and Hembree, S. E. (1994). Peer social status and friendship: Independent contributors to children's social and academic adjustment. *Merrill-Palmer Quarterly* **40**, 461–477.
- Wentzel, K. R. (1991). Social competence at school: Relation between social responsibility and academic achievement. *Review of Educational Research* **61**, 1–24.

Social Interaction and Learning

K Littleton, Open University, Milton Keynes, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This article is concerned with how learning and development are assisted by the interactions which take place in schools. More specifically, the focus is on the nature and significance of synchronous, face-to-face educational dialogs in classroom contexts. Psychologists and educators have gained from Piaget, Vygotsky, and the researchers who have extended their work, an affirmation of the value of social interaction for learning and development, whether it be interaction between peers of similar levels of understanding or between people more in the roles of learner and teacher (for reviews see Mercer and Littleton, 2007; Light and Littleton, 1999). Indeed, it has been suggested that any: “efficacious pedagogy should be a judicious mix of *immersion* in a community of practice and *overt focusing* and scaffolding from ‘masters’ or ‘more advanced peers’ who focus learners on the most fruitful sorts of patterns in their experience” (Gee, 2000: 201–202). Mindful of this, this article considers both how teachers use dialog to help children learn and develop their ability to reason and the processes through which knowledge and understanding can develop when learners talk and work together relatively autonomously in classroom settings.

How Teachers Use Dialog to Help Children Learn and Develop

Education is not simply a matter of an individual accumulating information; it involves the gradual development of new problem-solving skills and new ways of using language for representing knowledge and making sense of experience. It also entails the gradual induction by teachers of students into new perspectives on the world. Over the last 30 or so years much research has sought to understand how teachers use talk to guide learning and construct a shared version of educational knowledge – common knowledge (Edwards and Mercer, 1987) – with their students. Drawing on this body of work, Mercer (1995) suggested that teachers use talk to do three things:

- (a) *elicit knowledge from students*, so that they can see what students already know and understand and so that the knowledge is seen to be ‘owned’ by students as well as teachers;
- (b) *respond to things that students say*, not only so that students get feedback on their attempts but also that

the teacher can incorporate what students say into the flow of discourse and gather students’ contributions together to construct more generalised meanings;

- (c) *describe the classroom experiences that they share with the students*, in such a way that the educational significance of those joint experiences is revealed and emphasised. (p. 25–26)

Knowledge Elicitation and Questioning

When attempting to elicit knowledge from their students, in addition to using direct elicitations, teachers very commonly utilize a technique which Edwards and Mercer (1987) term cued elicitation. Cued elicitation is a way of drawing out from students the information that is being sought by providing strong visual cues and verbal hints as to what answer is required. While cued elicitation is often accomplished through asking questions, there has been considerable controversy in educational research concerning the use of questions more generally as a strategy for guiding the construction of knowledge. Specifically, there has been disagreement concerning the functions and value of this characteristic form of classroom interaction (see, e.g., Norman, 1992; Wells, 1999).

At one time it was common to find researchers criticizing teachers for using questions and for talking too much. It was claimed, for example by Dillon (1988) and Wood (1992), that because most teachers’ questions are designed to elicit just one brief right answer (which often amounts to a reiteration of information provided earlier by the teacher) this unduly limits and suppresses students’ contributions to the dialogic process of teaching-and-learning. There was criticism too of the characteristic three-part I–R–F (initiation–response–feedback) structure of classroom discourse. The IRF exchanges open with an initiation, usually in the form of a question, from the teacher which elicits a response from a student, to which the teacher typically provides an evaluative follow up or feedback. The suggestion was, as Skidmore (2006: 507) explains, that the I–R–F sequence results in a quiz which requires students to do little more than display their recall of knowledge got by rote and produces a pattern of teacher-led recitation which tends to reinforce the teacher’s authority as the transmitter of received wisdom and severely restricts the possibilities open to students to contribute thoughtfully to classroom talk.

However, most classroom researchers would probably now agree that such judgments were too simplistic. One reason is that critics did not fully acknowledge

teachers' professional responsibility for directing and assessing children's learning of a curriculum, and that it would be perverse not to rely on questions and other prompts to do so. Second, critics tended to assume that all question-and-answer exchanges were performing the same function. However, the forms of a language do not have a simple and direct relationship to their functions. In the classroom, teachers' questions can thus have a range of different communicative functions. They can, for example, be used:

1. . . . to test children's factual knowledge or understanding . . .
What is the nearest planet to the sun?
2. . . . for managing classroom activity . . .
Could we have all eyes to the board please?
3. . . . and as a way of finding out more about what pupils are thinking . . .
Why did you decide to have just three characters in your play?

Even the above account is an oversimplification, because any single question can have more than one function (e.g., the third question above could be used to find out what pupils know and to get them to attend). Also, a question takes on a special meaning in the context of ongoing events. Compare, for example, the function of asking for the name of the nearest planet to the sun before beginning a scheme of work on the solar system, with asking the same question after it is completed.

The key point is that there is a need to distinguish between form and function when analyzing and evaluating questions in teacher-pupil dialog: and one can only judge the function of questions, and any other forms of language, in dialogic context. With respect to the I-R-F while it can result in the learners' display of recalled knowledge, it can also be used creatively by the teacher to: help students plan ahead for a task they are about to carry out, or to review and generalize lessons learned from the tasks they have already performed (Skidmore, 2006, 507). The teacher's follow up, for instance, can be put to multiple uses – including clarification, exemplification, explanation, expansion, or justification of a student's response. It could also invite a student to do any of those things (Wells, 1999). So while teachers' questioning certainly can require children to guess what answer is in the teacher's mind, that is merely one possible function. Teachers' questions can also serve other very important functions in the development of children's learning and their own use of language as a tool for reasoning. They can: encourage children to make explicit their thoughts, reasons, and knowledge and share them with the class; model useful ways of using language that children can appropriate for use themselves, in peer group discussions, and other settings (such as asking for relevant information possessed only by others, or asking why questions to elicit reasons), and provide opportunities

for children to make longer contributions in which they express their current state of understanding, articulate ideas, and reveal problems they are encountering.

Responding to What Students Say and Describing Shared Classroom Experience

While unsuitable contributions to a classroom-based discussion may be rejected or ignored, one of the ways teachers engage with their students is to use incorporate their contributions into the ongoing teaching-learning process. This is accomplished through confirmation or repetition of things of educational significance (to underscore their salience to the whole class) and the elaboration of contributions to further explain or highlight their significance (Edwards and Mercer, 1987; Mercer, 1995). From a student's perspective, school work should ideally have a cohesive, cumulative quality in which specific activities and their goals can be seen to form part of greater whole, as part of a purposeful educational journey.

Research has identified a number of ways in which teachers try to create continuities in the experience of learners, for example, by referring to past events and implicating these in the joint construction of knowledge with their students. Teachers commonly use recaps to summarize what they consider to be the most salient features of a past event for the current activity (Edwards and Mercer, 1987; Mercer, 1995). Recaps can be literal, when a teacher simply sums up what happens (Last week, we began reading *Macbeth*) or reconstructive, the latter being where the teacher rewrites history, presenting a modified version of events which fits his/her current pedagogic concerns. Teachers also frequently use elicitations to help students' recall of past events (e.g., Who can tell me what they found out about the moon in the last lesson?). It is common too for them to mark past shared experiences as significant and relevant by using *we* statements (as in Remember when we looked at the map of Finland?). In these ways, teachers invoke common knowledge and highlight the continuities of educational experience, trying to draw students into a shared, cumulative, and progressive understanding of the activities in which they are engaged.

Alexander (2000), Crook (1999), and other educational researchers have argued that coherent knowledge and purposeful understanding will not naturally emerge for students from their continuous immersion in classroom life: it has to be pursued actively as a goal, through the use of appropriate teaching strategies. Talk with a teacher, and with other students, is perhaps the most important means for ensuring that a student's engagement in a series of activities contributes to their developing understanding of the subject matter as a whole. In order to understand how classroom education succeeds and fails as a process for developing students' knowledge and understanding, research is now beginning to focus on exploring the temporal

relationship between the organization of teaching-and-learning as a series of lessons and activities and how it is enacted through talk and joint activity (see for example, Mercer, 2008; Mercer and Littleton, 2007; Rasmussen, 2005; Scott *et al.*, 2006). The importance of cumulative, rather than simply extended, dialog is central to the notion of dialogic teaching (Alexander, 2004).

Dialogic teaching

Dialogic teaching is a concept which enables us to focus more precisely on the role of the teacher in classroom talk. The concept has emerged from the comparative, cross-cultural research of Alexander (2000). In some ways, classroom talk sounds very similar the world over; but Alexander's work suggests that there seem to be some, quite subtle, variations in the ground rules which normally apply. Even within countries, teachers can set up markedly different expectations among members of their class about how they should engage in dialog. The variation Alexander describes is not revealed by comparisons of the extent to which teachers use questions or other kinds of verbal acts: rather, it concerns more nuanced aspects of interaction such as the extent to which teachers elicit children's own ideas about the work they are engaged in, make clear to them the nature and purposes of tasks, encourage them to discuss errors and misunderstandings, and engage them in extended, linked sequences of dialog or chains of enquiry about such matters.

Dialogic teaching is that in which both teachers and pupils make substantial and significant contributions and through which children's thinking on a given idea or theme is helped to develop and move forward. It is intended to highlight ways that teachers can encourage students to participate actively in dialogs which enable the students to articulate, reflect upon, and modify their own understanding – and, conversely, how they may avoid doing so. Alexander suggests that dialogic teaching is indicated by certain features of classroom interaction: questions are structured so as to provoke thoughtful answers; answers provoke further questions and are seen as the building blocks of dialog rather than its terminal point and individual teacher–pupil and pupil–pupil exchanges are chained into cumulative, coherent lines of enquiry rather than left stranded and disconnected (Alexander, 2004: 32).

In terms of what the teacher actually does in classroom interaction, dialogic teaching can be characterized as that in which students are given opportunities and encouragement to question, state points of view, and comment on ideas and issues which arise in lessons; the teacher engages in discussions with students which explore and support the development of their understanding of content; the teacher takes students' contributions into account in developing the subject theme of the lesson and in devising activities which enable students to pursue their understanding themselves, through talk and other activity, and

the teacher uses talk to provide a cumulative, continuing, contextual frame to enable students' involvement with the new knowledge they are encountering.

Dialogic teaching is essentially a specification of good practice, derived from both theory of the nature of dialog (drawn from the work of Bakhtin, Vygotsky, and others) and observations of practice across a range of cultural settings. It has clear links, in both its origins and its nature, with some other concepts devised by educational researchers such as reciprocal teaching (Brown and Palincsar, 1989), contingent tutoring (Wood and Wood, 1999), dialogic enquiry (Wells, 1999), and dialogic spells (Nystrand *et al.*, 2003). As an educational concept dialogic teaching is both descriptive and prescriptive. It represents an approach to classroom teaching which: aims to be more consistently searching and more genuinely reciprocal and cumulative (Alexander, 2004: 1) than is usually observed in classrooms, anywhere in the world. Dialogic teaching is that which is collective, reciprocal, supportive, purposeful, and cumulative. It requires a teacher to orientate to the state of understanding of students, engage them in exchanges which will reveal the changing limits and possibilities of their developing interests and understandings, and adjust their communication strategies accordingly as classroom interaction progresses. It involves students taking an active, engaged role in both their own learning and that of their classmates; becoming explicitly part of a collective endeavor. It requires the creation and maintenance of the kind of dynamic intersubjectivity that Mercer (1995) has called an intermental development zone.

It is important to note that Alexander also suggests that some key indicators of dialogic teaching concern the ways in which children are seen to talk and work together in collaborative group settings and he particularly identifies the following as being important: children listen carefully to each other; they encourage each other to participate and share ideas; they build on their own and each others' contributions; they strive to reach common understanding and agreed conclusions, yet they respect minority viewpoints (Alexander, 2004: 33). This characterization of children engaged with each other and each other's ideas is juxtaposed with the seeming paradox of children being seen to work everywhere in groups, but rarely as groups.

Talking and Learning Together

While the study of children's group-based activity in school has had a relatively brief history, there has been a great deal of research interest in children's collaborative working, learning, and problem-solving in more general terms. It is clear that children's joint activity has been researched in diverse ways – for example, through large-scale surveys of life in classrooms; experiments in which pairs or groups of children work on specially designed

problem-solving tasks; and detailed analyses of talk between pairs or groups of children working on curriculum-based tasks in school.

Surveying Classroom Activity

Perhaps one of the first messages to emerge from work surveying classroom activity is that, at least in British primary schools, truly collaborative activity rarely happens. This was the conclusion of a large-scale research project carried out in the 1970s called ORACLE (Galton *et al.*, 1980). The ORACLE team of researchers, observing everyday practice in a large number of British primary schools, found that just because several children were sitting together at a table (as was common) did not mean that they were collaborating. Typically, children at any table would simply be working, in parallel, on individual tasks. This problem has also been underscored in a number of more recent studies, some of which have shown that even when children are set joint tasks their interactions are rarely productive (Galton *et al.*, 1999; Blatchford and Kutnick, 2003; Alexander, 2004, 2005). This tells us something important about the nature of everyday educational practice and leads to the conclusion that if simply left to their own devices to discuss something or talk together much classroom-based talk among children may be of limited educational value.

Experimental Studies

Much of the early collaborative learning research consisted of experimental studies of peer interaction which were designed to establish whether working and solving problems collaboratively was in fact more effective than working alone. Typically, children would be given the same task, but allocated either to working collaboratively or working alone, and their performance on the task assessed. Reviewing such studies, Slavin (1980) noted that cooperative or collaborative learning was often judged to increase students' academic achievement, self esteem, and motivation. These sorts of investigations gave rise to a related strand of research in which independent variables, such as the size of the group (e.g., Fuchs and Fuchs, 2000), group composition, with respect to, for example, gender and ability (e.g., Barbieri and Light, 1992; Howe, 1997; Webb, 1989; see also Wilkinson and Fung, 2002 for a review of work in this field), and nature of the task (e.g., Cohen, 1994; Light and Littleton, 1999; Underwood and Underwood, 1999) were manipulated and attempts were made to assess their effects. However researchers now tend to focus less on establishing parameters for effective collaboration and more on the ways in which factors such as task design or group composition influence the nature of collaborative interaction (Dillenbourg *et al.*, 1995; Littleton, 1999; Kleine-Staarmann, in press). This shift to a more process-oriented kind of

investigation has brought with it an interest in the talk and joint activity of learners working together on a task, with attempts being made to identify those interactional features which are important for learning and cognitive change.

Many experimental studies of collaborative interaction have focused on how children talk together when they are working on a problem or task. In particular, correlational techniques have been used to establish whether there is evidence of an association between particular features of the learners' talk and on-task success or subsequent learning gain as indexed by individual performance on a posttest. For example, Azmitia and Montgomery (1993) found that the quality of children's dialog is a significant predictor of their successful problem solving. Studying children engaged in joint computer-based problem-solving tasks, Barbieri and Light (1992) found that measures of the amount of talk about planning, negotiation, and the co-construction of knowledge by partners correlated significantly with successful problem solving by pairs, and to successful learning outcomes in subsequent related tasks by individuals. Similar analytic techniques used by Underwood and Underwood (1999) demonstrated that for pairs of children working on a computer-based problem-solving activity those who were most observed to express opinions, analyze the situation in words, and express agreement and understanding achieved the best outcomes. Experimental evidence thus supports the view that focused, sustained discussion among children not only helps them solve problems but promotes the learning of the individuals involved. While this may seem like common sense, if it is so obviously true, one is led back to the question raised in the previous section of why high-quality peer discussion is not seen in many classroom contexts.

Researching Talk between Pupils in the Classroom

In the classic work *Communication and Learning in Small Groups* Barnes and Todd (1977, 1995) show how knowledge can be treated by pupils or students as a negotiable commodity when they are engaged in joint tasks. They suggest that pupils are more likely to engage in open, extended discussion and argument when they are talking and working with their peers outside the visible control of their teacher, enabling them to take more active and independent ownership of knowledge. Based on their in-depth observations, Barnes and Todd suggest that classroom discussion has to meet requirements for explicitness which would not typically be required in everyday conversation. Knowledge should be made publicly accountable – relevant information should be shared effectively, opinions should be clearly explained, and explanations examined critically. They also argue that the successful pursuit of educational activity through group work depends on learners sharing the same ideas about what is relevant to the discussion and having

a joint conception of what is trying to be achieved by it. These points have been supported by other research studies (Bennett and Dunne, 1992; Galton and Williamson, 1992; Kumpulainen and Wray, 2002; Mercer and Littleton, 2007).

What Counts as Learning?

Within the research literature on collaborative learning, there is considerable diversity in what is conceived of as learning and as a learning outcome. As indicated earlier, for Barbieri and Light (1992), Underwood and Underwood (1999) and others (such as Howe and Tolmie, 1999) learning is seen in terms of individual accomplishments, demonstrated through appropriate tests on individual children after group activity. Most of that research recognizes that the quality of talk and social interaction is a significant factor. However, the more radical possibility is that collaborative talk is not just a stimulant for individual thinking, but can itself be considered a social form of thinking. As some researchers have put it:

talk and social interaction are not just the means by which people learn to think, but also how they engage in thinking. . . . [D]iscourse *is* cognition *is* discourse. . . . One is unimaginable without the other. (Resnick *et al.*, 1997: 2)

This is a challenge to traditional, individualistic accounts of the nature of knowledge, and of learning. It implies that talk is not just the mediating means for supporting individual development, rather that ways of thinking are embedded in ways of using language. From this perspective, the accomplishment by children of particular forms of educated discourse is a valuable educational goal in its own right. This raises the possibility that how a learner engages and interacts with other learners may have a profound and enduring impact on their attainment and, indeed, on their intellectual development.

Supporting and Promoting Productive Interaction

Many opportunities for collaborative learning are fortuitous. They simply emerge as a consequence of being part of a particular community of learners. That said, we still need to understand how best to enable learners' joint endeavors, so that we can promote the most effective opportunities for collaborative learning and design strategies for optimizing collaboration. This concern is reflected in recent research, in which three factors have been given particular attention: task design; quality of relationships; and quality of talk.

Task design

When thinking about the issue of how to support productive group work many researchers have emphasized the significance of task design. It is important that group tasks

should be designed such that learners need to work together on them. Therefore tasks should not be too simple, for if each child can easily solve the problem or complete the task alone, then there is no imperative for joint working. Equally, if the task is too complex for the children, then they will struggle to create understanding and meaning. A group task is one which requires resources that no single individual possesses and is one in which students work interdependently and reciprocally – the exchange of ideas and information being vital to success (Cohen, 1994). It is perhaps not surprising, then, that some research suggests that open-ended, challenging tasks are more effective in facilitating productive interaction than more closed tasks focused on finding one right answer (Cohen, 1994; Van Boxtel *et al.*, 2000). This is in part because closed tasks more easily lead to one participant, perhaps, a more knowledgeable person, dominating the discussion (Arvaja, 2005). A clear task structure and provision of feedback is also important and this might be one of the best ways in which computer-technology can resource joint activity (Howe and Tolmie, 1999). That said, it is not simply a case of getting the task right. Of course, good task design helps: but because the meaning of educational tasks is created through interaction – task design is only part of the story.

Quality of relationships

According to Van Oers and Hännikäinen (2001: 105): “The main reason why discourses in collaborative learning processes ever lead to improved understandings is that the participants in the process are willing to share their understandings and keep on doing so *despite* their disagreements and conflicts . . . the fact that they can ever be productive at all relies on the fact that the participants in this process, for the time being, feel obliged to each other, stay with each other and maintain togetherness.” This claim draws attention to the importance of the relationship between partners as they interact and work together. Researchers investigating how friendships mediate joint activity (e.g., Azmitia and Montgomery, 1993; Hartup, 1998; Youniss, 1999; Vass, 2003) have found that relational closeness is associated with the sharing of ideas, exchanging points of view, and a collective approach to challenging tasks. It seems that the development of close relationships, characterized by a sense of trust and mutuality, enhances learning (Howes and Ritchie, 2002; Underwood and Underwood, 1999). Findings such as these have led some researchers to argue that what is needed is a relational approach to group working, which properly recognizes that classroom learning is a social activity (Blatchford *et al.*, 2003). The suggestion is that training should be given to promote the development of close relationships between classmates, through among other things, developing interpersonal trust between the children – something which is often stressed in work

investigating collaborative activity in the creative arts (see Miell and Littleton, 2004). To this end, Blatchford and colleagues have developed an educational intervention program which they characterize as using a relational approach to the development of group working. Drawing on influences from attachment theory and studies of parent-child interactions, the program engages the participating children in activities designed to foster trust and mutual support and develop communication skills and joint problem solving. Evaluations of the program involving comparisons between experimental and control classes have indicated that this relational approach is not only successful in motivating children to participate in group activity and value it, but that it has a significant impact on their reading and mathematics attainment (Kutnick, 2005). Work by researchers such as Swann (1992) which highlights that some peer-based interactions are highly gendered and are characterized by dominance and asymmetry also add weight to the claim that for group activity to be effective children need to be taught to relate in positive ways.

Quality of talk

Other researchers, such as Mercer and Littleton (2007), suggest that children have to do more than engage with each other in a positive and supportive way; they also should become able to build constructively and critically on each others' ideas. It is their assertion that it is imperative to teach children how to use language to reason together. In collaboration with colleagues, they have developed Thinking Together, a classroom-based approach which places a special emphasis on the role of the teacher as a guide and model for language use, who fosters an inclusive climate for discussion while also enabling children to understand better how language can be used as a tool for thinking. It supports children in learning to talk in groups as well as providing them with opportunities for talking to learn. Through the systematic integration of both teacher-led interaction and group-based discussion children are helped to understand that aims for group activity and the use of spoken language are as much to do with high quality educationally effective talk and joint reasoning through exploratory talk, in which reasoning is accountable and visible, as with curriculum learning. The processes by which children learn how to learn are thus directly addressed, rather than being left to chance. The approach is more than delivering a particular form of communication skills training. It does encourage children to engage in particular ways of talking and working together, and they are explicitly guided in how to use language as a tool for reasoning together. They are encouraged to give reasons, seek clarification, ask questions, listen to each others' ideas, and so on. But children learn much more than a model set of talk strategies, and the goal is not that they will simply adhere to the ground rules for such talk. The main goal is children's active appropriation of

a particular educated way of talking and thinking, one that they understand and appreciate, so that in time they are able to apply, adapt, and develop their use of language flexibly and creatively in their discussions. Evaluations of the approach undertaken with children across a diverse age spectrum shows that teachers' encouragement of children's use of certain ways of using language leads to better learning and conceptual understanding (see Mercer and Littleton, 2007).

Concluding Remarks

There is certainly much more to discover about the ways that language experience in the classroom can contribute to the development of children's abilities to communicate, learn, and reason, but what is known now provides a well-informed basis for the creation of a more dialogic, and more effective, educational practice. It is an uncontroversial claim that through social interaction, children learn how language can be used to describe the world, to make sense of life's experience, and to get things done. However, what children learn from talk in the classroom, and how significant it is for their psychological development and educational progress, will depend a great deal on the range and quality of the dialogs in which they engage.

Acknowledgment

The author wishes to thank Neil Mercer for his comments on an earlier draft and for his permission to draw upon material in their book (*Dialogue and Development*) in the preparation of this article.

See also: Peer Learning in the Classroom; Social Aspects of Collaborative Learning.

Bibliography

- Alexander, R. (2000). *Culture and Pedagogy: International Comparisons in Primary Education*. Oxford: Blackwell.
- Alexander, R. (2004). *Towards Dialogic Teaching: Rethinking Classroom Talk*. Cambridge: Dialogos.
- Alexander, R. (2005). Culture, dialogue and learning: Notes on an emerging pedagogy, paper presented at the *International Association for Cognitive Education and Psychology (IACEP)*, 10th International Conference. University of Durham, UK, July.
- Arvaja, M. (2005). *Collaborative Knowledge Construction in Authentic School Contexts*. Doctoral Thesis, Institute of Educational Research, University of Jyväskylä, Jyväskylä: University Printing House.
- Azmitia, M. and Montgomery, R. (1993). Friendship, transactive dialogues and the development of scientific reasoning. *Social Development* 2(3), 202–221.
- Barbieri, M. and Light, P. (1992). Interaction, gender and performance on a computer-based task. *Learning and Instruction* 2(1), 199–213.
- Barnes, D. and Todd, F. (1977). *Communication and Learning in Small Groups*. London: Routledge and Kegan Paul.
- Barnes, D. and Todd, F. (1995). *Communication and Learning Revisited*. Portsmouth, NH: Heinemann.

- Bennett, N. and Dunne, E. (1992). *Managing Classroom Groups*. London: Simon and Schuster.
- Blatchford, P. and Kutnick, P. (2003). *Special Issue: Developing Groupwork in Everyday Classrooms*. *International Journal of Educational Research* **39**(1–2).
- Blatchford, P., Kutnick, P., Baines, E., and Galton, M. (2003). Towards a social pedagogy of classroom groupwork. *International Journal of Educational Research* **39**, 153–172.
- Brown, A. L. and Palincsar, A. S. (1989). Guided, co-operative learning and individual knowledge acquisition. In Resnick, L. B. (ed.) *Knowing, Learning and Instruction*, pp 393–452. Hillsdale, NJ: Erlbaum.
- Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research* **64**(1), 1–35.
- Crook, C. (1999). Computers in the community of classrooms. In Littleton, K. and Light, P. (eds.) *Learning with Computers: Analysing Productive Interaction*, pp 102–117. London: Routledge.
- Dillenbourg, P., Baker, M., Blaye, A., and O'Malley, C. (1995). The evolution of research on collaborative learning. In Spada, H. and Reiman, P. (eds.) *Learning in Humans and Machines: Towards an Interdisciplinary Learning Science*, pp 189–211. Oxford: Elsevier.
- Dillon, J. J. (ed.) (1988). *Questioning and Discussion: A Multidisciplinary Study*. London: Croom Helm.
- Edwards, D. and Mercer, N. (1987). *Common Knowledge: The Development of Understanding in the Classroom*. London: Methuen/Routledge.
- Fuchs, L. S. and Fuchs, D. (2000). Effects of workgroup structure and size on student productivity during collaborative work on complex tasks. *Elementary School Journal* **100**(3), 183.
- Galton, M., Hargreaves, L., Comber, C., Wall, D., and Pell, A. (1999). *Inside the Primary Classroom: 20 Years On*. London: Routledge.
- Galton, M., Simon, B., and Croll, P. (1980). *Inside the Primary Classroom (the ORACLE Project)*. London: Routledge and Kegan Paul.
- Galton, M. and Williamson, J. (1992). *Group Work in the Primary Classroom*. London: Routledge.
- Gee, J. (2000). Discourse and socio-cultural studies in reading. In Kamil, M., Mosenthal, B., Pearson, P., and Barr, R. (eds.) *Handbook of Reading Research*, vol. III, pp 195–208. London: Erlbaum.
- Hartup, W. W. (1998). The company they keep: Friendships and their developmental significance. In Campbell, A. and Muncer, S. (eds.) *The Social Child*, pp 143–164. Hove: Psychology Press.
- Howe, C. (1997). *Gender and Classroom Interaction: A Research Review*. Edinburgh: The Scottish Council for Research in Education.
- Howe, C. and Tolmie, A. (1999). Productive interaction in the context of computer supported collaborative learning in science. In Littleton, K. and Light, P. (eds.) *Learning with Computers: Analysing Productive Interaction*, pp 24–45. London: Routledge.
- Howes, C. and Ritchie, S. (2002). *A Matter of Trust: Connecting Teachers and Learners in the Early Childhood Classroom*. New York: Teachers College Press.
- Kleine-Staarman, J. (in press). *Collaboration in CSCL: Social Interaction in Primary School Computer-Supported Collaborative Learning Environments*. Doctoral Thesis, submitted to Radboud University Nijmegen, The Netherlands, Institute of Behavioural Sciences.
- Kumpulainen, K. and Wray, D. (eds.) (2002). *Classroom Interaction and Social Learning: From Theory to Practice*. London: Routledge-Falmer.
- Kutnick, P. (2005). Relational training for group working in classrooms: Experimental and action research perspectives. *Paper Presented as Part of the Educational Dialogue Research Unit Seminar Series*. The Open University, Milton Keynes, June.
- Light, P. and Littleton, K. (1999). *Social Processes in Children's Learning*. Cambridge: Cambridge University Press.
- Littleton, K. (1999). Productivity through interaction: An overview. In Littleton, K. and Light, P. (eds.) *Learning with Computers: Analysing Productive Interaction*, pp 179–194. London: Routledge.
- Mercer, N. (1995). *The Guided Construction of Knowledge: Talk Amongst Teachers and Learners*. Clevedon: Multilingual Matters.
- Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *Journal of the Learning Sciences* **17**, 33–59.
- Mercer, N. and Littleton, K. (2007). *Dialogue and the Development of Children's Thinking*. London: Routledge.
- Miell, D. and Littleton, K. (eds.) (2004). *Collaborative Creativity: Contemporary Perspectives*. London: Free Association Books.
- Norman, K. (ed.) (1992). *Thinking Voices: The Work of the National Oracy Project*. London: Hodder and Stoughton.
- Nystrand, M., Wu, L., Gamorgan, A., Zeiser, S., and Long, D. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes* **35**(2), 135–198.
- Rasmussen, I. (2005). *Project Work and ICT: Studying Learning as Participation Trajectories*. Doctoral Thesis, Faculty of Education, University of Oslo, Norway.
- Resnick, L., Pontecorvo, C., and Säljö, R. (1997). Discourse, tools and reasoning. In Resnick, L., Säljö, R., Pontecorvo, C., and Burge, B. (eds.) *Discourse, Tools and Reasoning: Essays on Situated Cognition*, pp 1–22. Berlin: Springer.
- Scott, P., Mortimer, E., and Aguiar, O. (2006). The tension between authoritative and dialogic discourse: A fundamental characteristic of meaning making interactions in high school science lessons. *Science Education* **90**, 605–631.
- Skidmore, D. (2006). Pedagogy and dialogue. *Cambridge Journal of Education* **36**(4), 503–514.
- Slavin, R. E. (1980). Co-operative learning. *Review of Educational Research* **50**(2), 315–342.
- Swann, J. (1992). *Girls, Boys and Language*. London: Blackwell.
- Underwood, J. and Underwood, G. (1999). Task effects in co-operative and collaborative learning with computers. In Littleton, K. and Light, P. (eds.) *Learning with Computers: Analysing Productive Interaction*, pp 10–23. London: Routledge.
- Van Boxtel, C., Van der Linden, J., and Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction* **10**(4), 311–330.
- Van Oers, B. and Hännikäinen, M. (2001). Some thoughts on togetherness: An introduction. *International Journal of Early Years Education* **9**(2), 101–108.
- Vass, E. (2003). *Understanding Collaborative Creativity: An Observational Study of the Effects of the Social and Educational Context on the Processes of Young Children's Joint Creative Writing*. Doctoral Thesis, Milton Keynes: The Open University.
- Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research* **13**(1), 21–39.
- Wells, G. (1999). *Dialogic Enquiry: Toward a Sociocultural Practice and Theory of Education*. Cambridge: Cambridge University Press.
- Wilkinson, I. and Fung, I. (2002). Small group composition and peer effects. *International Journal of Educational Research* **37**, 425–447.
- Wood, D. (1992). Teaching talk. In Norman, K. (ed.) *Thinking Voices: The Work of the National Oracy Project*, pp 203–214. London: Hodder and Stoughton.
- Wood, H. and Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers and Education* **33**, 153–169.
- Youniss, J. (1999). Children's friendship and peer culture. In Woodhead, M., Faulkner, D., and Littleton, K. (eds.) *Making Sense of Social Development*, pp 13–26. London: Routledge.

Social Networks and the Education of Children and Youth

B Schneider and T G Ford, Michigan State University, East Lansing, MI, USA
L Perez-Felkner, University of Chicago, Chicago, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Centrality – Refers to the number of ties one has to others in a network. Those with more ties may have more access to resources contained within the network.

Closure – Refers to the interconnections members have with one another, that is, the social ties among the network members that are historical and bridging.

Density – Refers to the number and strength of ties among individuals within a social network. In formal network theory, density refers to the number of ties observed divided by the number of total possible ties in a bounded group.

Homogeneity – Being in a group with similar demographic and other characteristics.

Homophily – The concept that friendships tend to form among people who perceive themselves as similar to one another.

Proximity – A basis of interpersonal attraction characterized by physical or psychological proximity to another.

Social networks – The social ties among a group of individuals, which can involve as small as two “also known as a dyad” or much larger numbers of individuals.

The social networks of children and youth are formed in the context of their families, peer groups, schools, and neighborhood communities. Researchers studying social networks of children and adolescents have primarily been interested in the formation of friendship relations and the impact of social location on academic achievement and social development. Social networks are commonly defined as the social ties among a group of individuals, which can involve as small as two – also known as a dyad – or much larger numbers of individuals. Through these social ties, social norms can form which constitute social capital, the resources within social groups that increase the potential for members to achieve their interests and goals (Coleman, 1990). Scholars have identified several characteristics of social capital that have direct implications for the formation and function of social networks. These properties include density, closure, and trustworthiness (Coleman, 1988).

Density generally refers to the number and strength of ties among individuals within a social network. In formal network theory, density refers to the number of ties observed divided by the number of total possible ties in a bounded group. The degree to which multiple network members share reciprocal ties with one another determines the cohesiveness of the network (Wasserman and Faust, 1994). Dense social networks can have positive or negative influences on their members depending on the social context of the network and its goals. In children's social networks where academic performance is valued and children have strong relationships with each other, these values are likely to be transmitted and upheld by the network. However, in situations where networks are dense and the goals are negative, such dense ties can encourage negative behavior patterns. For example, Haynie (2001), examining friendship networks using Add Health data, finds a general association with an adolescent's delinquency and that of his or her friends. Results indicate that network density appears to be a critical component of the delinquency–peer association.

Closure refers to the interconnections members have with one another, that is, the social ties among the network members that are historical and bridging in multiple contexts. In formal network theory, closure refers to how all of an ego's alters (e.g., a student's teacher, parent, and sibling) communicate, thereby closing the circle. Coleman (1990) argues that intergenerational social closure, the reciprocal social ties that connect children with their families, help to create functional communities where educational goals are strengthened and acted upon. Parent social networks that include the parents of their children's friends can build social capital that is reinforced through their children's peer networks, their own adult networks, and the intergenerational family and community networks that the parents and their children share. This point is illustrated by Offer and Schneider (2007), who show in their analysis of data from a study of parents and children in 500 families that adolescents' friendships with peers generate friendships among the parents of their friends, suggesting that information and social resources flow from children to adults as well as from adults to children.

Bridging social networks can be especially beneficial for members in low-resource networks. Distinguishing between weak and strong social ties, Granovetter (1973) finds in his study of job searches how individuals in

low-resource networks learn about employment opportunities through weak ties outside their immediate social circle. A similar argument has been made by Kim and Schneider (2005) with respect to schools, where they show how immigrant parents who broker and maintain ties to individuals outside their school communities gain access to more resources for their children's education.

Dense social networks with high degrees of closure promote trustworthiness (Coleman, 1988). This concept of trust has been further developed by Bryk and Schneider (2002) in their study of the effectiveness of urban school reform in Chicago elementary schools, where they show that relational trust which is formed through shared expectations and fulfillment of mutual obligations increases the likelihood of school change and higher academic performance. When relational trust is strengthened in social networks among various role sets, including teachers and parents, teachers and students, and parents and administrators, and the academic and social welfare of the students becomes the top priority, such relationships create a more productive learning environment for the entire school community.

The Study of Social Networks

The social networks perspective distinguishes itself from other research approaches through its emphasis on the importance of the relationships between units in a study. In contrast to many non-network research studies, studies of social networks begin with the assumption that individuals and their actions are interdependent, that is, the behaviors and/or actions of one individual influence the behaviors and/or actions of others within a group. Social network analysis also understands the relational ties which exist between individuals within a network to be pathways, which enable the flow of resources among members. From the perspective of individuals within a group, network analyses also endeavor to ascertain how network structure facilitates or constrains individual action in a network (Wasserman and Faust, 1994).

The study of social networks could be understood as focused on two distinct methodologies: (1) formal network analysis and (2) observational network analysis. Formal network methodology employs mathematical models to describe the relationships within a specific, bounded population. To obtain this information, researchers first identify all members of the population of interest and then ask each subject to report on his/her relations with every other member of the group. In this instance, researchers try to obtain information from all members of a fully enumerated social group, which is often called complete network data (Marsden, 1990).

The methods of describing network structure are being modified and tested; now researchers can examine small

and large networks, with ten to hundreds of thousands of nodes sites of network connections (Moody, 2001a; Wasserman and Faust, 1994). Technological advances over the past 50 years have enhanced the development of these models from those that were static to ones that are more fluid. New techniques, such as dynamic network visualization, capture motion and change within networks, allowing researchers to study how networks develop and change through static flip books and/or dynamic movies (Moody *et al.*, 2005). These new methods are designed to visualize relational change beyond more conventional techniques that use one- or two-dimensional pictures with points, lines, and arrows showing directionality.

Egocentric network studies also use formal methodology to focus on how networks operate around individuals. In these studies, subjects typically report affiliations and ties through surveys or structured interviews. Although these data tend to be less comprehensive than that of the complete network in which they are embedded, they are generally useful in studies where the research question concerns how individuals and small groups evaluate their position and affiliations (e.g., centrality – one who has ties to most network members; popularity – more friendship nominations than one would expect based on network composition) in relation to others in the network. Studies of this type generally generate data through peer nominations where subjects are asked to name their three closest friends. These questions have been asked in the major national longitudinal studies conducted by the National Center of Education Statistics (NCES) over the past 40 years beginning with *High School and Beyond*, followed by the *National Educational Longitudinal Study of 1988 (NELS: 88)*, and, more recently, in the *Educational Longitudinal Survey of 2002 (ELS: 2002)*.

Observation field-based studies provide extensive descriptive information on how members of a group interact with one another in multiple situations, often over time. One of the most recent examples of this type of study was conducted by McFarland (2001) where he used student and teacher surveys, interviews, and school records to examine how social networks contribute to active resistance in 36 classrooms in two Midwestern high schools. Measuring individual and clique-level status, density, and academic standing, McFarland comprehensively describes the characteristics and processes by which student networks create opportunities for student resistance to instructional activities, showing that resistance is related to the structure of student relations and not simply a result of individual responses to teachers' instruction.

Social Networks and Young Children

Friendships among children are generally understood to be significant in determining future developmental outcomes.

The friendships that children form with one another are resources that they can draw upon to cope with the psychological and social stresses of developmental transitions, such as that from childhood to adolescence. In understanding how development is impacted by childhood friendships, Hartup (1996) maintains that it is not enough to know that a child has friends; we should also know something about the identities – including attitudes and behavioral characteristics – of that child's friends, as well as the nature of their relationship. In other words, more comprehensive assessments of children's friendships are needed in order to bring more predictive power to bear on anticipating a child's future disposition and social competence.

Given the salience of friendship networks in the study of child development, it is not surprising that many studies of social networks among children have focused on the formation of peer groups. One of the most frequently cited studies was conducted by Hallinan and Tuma (1978), where they examined the friendship networks of fourth, fifth, and sixth graders over time. Asking children to name their best friends, friends, and nonfriends over time, the researchers measured friendship stability formed through personal relationships and those that were formed through learning-related tasks directed by the teacher. Contrary to expectations, they found that within-classroom student network groupings that were determined by the children's choices, best nominated friends did not emerge as more stable than those friendships that developed through teacher-assigned groups. Their findings suggest that teacher grouping based on classroom tasks strongly affected children's friendship formations.

More recent studies of social networks have also pursued the influence of teacher instructional practices on children's friendship choices. Plank (2000) examined how teachers' task and reward structures influence the academic and social hierarchy in the social networks of Hmong and white students in ten classrooms in five Midwestern elementary schools. Focusing on the effects of social hierarchy on racial and ethnic integration, he found that students from higher social classes tended to be at the center of classroom social networks. Plank suggests that norm-based task and reward structures seem to produce social groups that align across class and racial lines, concluding that the pedagogical style of the teacher directly influences the social network structure of children in the classroom.

Another study by Kubitschek and Hallinan (1998) also focused on teacher activities and friendship patterns. Using social network data, they establish a link between teacher-tracking practices and student friendship choices, demonstrating that these linkages cohere around three bases of interpersonal attraction: propinquity, similarity, and status. They argue that the nature and effects of tracking practices determine friendship choices due to their propensity to encourage intra-track communication

(propinquity), to create greater similarity among students within tracks (similarity), as well as to reflect stratification trends in greater society (status).

Cairns *et al.* (1995) sought to examine the relative stability of friendships and social networks in childhood and adolescence among 131 fourth- and seventh-grade students in two suburban schools over a 3-week period. The researchers used respondent interviews administered at the beginning and end of the observational period to determine social group membership through the social-cognitive map (SCM) procedure. For both children and early adolescents in this study, Cairns *et al.* demonstrate that friendships and social group membership are generally more fluid than has previously been recognized. This fluidity of friendship ties has also been found in adolescent peer and friendship ties, with youth repositioning their social ties throughout the high school experience (Steinberg *et al.*, 1996). However, friendships developed through school-sponsored activities appear to be more stable (Schneider and Stevenson, 1999).

Social Networks and Adolescents

Just as peer interactions play an important role in the development of young children, so also are peer groups influential in adolescent identity development. Adolescence marks the time when young people seek to establish an independent identity from their families and seek acceptance and a sense of belonging through peer groups. As this occurs, similar shifts in the locus of peer relations occur, from dyadic or small-group relationships to peer groups or crowds (Brown and Lohr, 1987). The importance of crowd affiliation on identity development was studied early on by Brown *et al.* (1986). Based on subjective responses of adolescents regarding the importance of crowd affiliation and why it was important, the researchers found that importance of crowd affiliation was negatively associated with age. Thus, younger adolescents tended to value crowd associations, while older adolescents relied on them less heavily, due to the strength of established friendship networks. Furthermore, respondents' sense of identity was not related to the importance they placed on crowd affiliation, but was related to the centrality of their position in the peer network.

Crowd affiliations have also been demonstrated to predict future behaviors, educational attainment, and general psychological adjustment in adolescents. Using widely recognized identity categories such as the jocks, brains, and the princesses, Barber *et al.* (2001) found that jocks and brains had the most positive adjustment in later years, primarily due to their involvement in school-related activities in tenth grade. Future adolescent adjustment was also found by Fuligni *et al.* (2001) to be related to adolescent peer

dependence. The more adolescents reported being willing to sacrifice their talents and school performance for being in a particular group, the poorer was their academic performance and overall adjustment. Taken together, these findings suggest that peer-group orientation plays an influential role in adolescent's identity development.

Due to its relationship to peer groups, identity development in adolescence has been a topic of particular interest to social network researchers. Collecting data on nearly 6000 high school students in California, McFarland and Pals (2005) explore how social networks affect the identity development of adolescents. They found that, while category memberships are highly influential in identity development, the network characteristics of prominence, homogeneity, and bridging lead to higher salience of identity imbalance, which in turn leads to an increased incidence of identity change. Homogeneity, that is, being in a group with similar demographic and other characteristics, exerted the greatest influence on identity change, revealing that, over time, social conformity inhibits identity instability and inconsistency.

The concept that friendships tend to form among people who perceive themselves as similar to one another is termed homophily. These social affiliations tend to be aligned around traits on which people share values (value homophily) or social status (status homophily). McPherson *et al.* (2001), in their review of homophily, argue that the most persistent traits which determine network homogeneity are race and race-like ethnicity.

The tendency toward network homophily and homogeneity is evidenced in several studies. Jackson *et al.* (2006), in their study of 1268 fifth graders' peer and teacher nominations of classroom social network relationships, asked students to rate who is most like them (Like Most), most not like them (Like Least), a leader (Leader), and who is aggressive (Fights). They found that classroom racial composition and the race of the teacher are directly related to the nominations of students into each of these groups. In classes that are majority white, black students are significantly less likely to be nominated by both peers and teachers as a leader, more likely to be categorized as aggressive, and less likely to be nominated in friendship networks. However, they also demonstrate that, as black students are increasingly represented in the classroom, black children's nominations to these categories also improve. Based on these findings, the researchers conclude that white children tend to be more protected in majority black environments – a phenomenon they attribute to their status in the broader social community and to the history of discrimination and bias against blacks.

Examining the substantive integration of friendship networks in varied school contexts, Moody (2001b), using data from the Add Health Study, draws on contact theory to explore the relationship between friendship segregation and school organization and diversity. His findings suggest

a curvilinear relationship between heterogeneity and friendship segregation, finding that once a particular threshold of race salience is reached in the school, integration peaks and then falls. Moody maintains that schools have the greatest effect on racial friendships when they can structure racial mixing through the racial integration of extracurricular activities.

Virtual networks have emerged as sites for establishing socialities, although early studies suggested that ties created through the Internet were weak. Today, adolescents and young adults frequent sites such as MySpace, Facebook, and online dating websites to form new relationships. The relationships formed within these groups may or may not exist additionally outside of virtual space, but they are nonetheless real, to varying degrees. Youth participate in these networks for generally similar purposes as their traditional social groups: to forge new social relationships, find and interact with people who share their interests, and find people to date. Online relationships are now considered a part of the social world of most adolescents. These relationships are becoming increasingly significant to research on social networks not only because they are more prevalent, but also because they refine and reshape understanding of the motivations underlying adolescent friendship formation as well as the possible avenues in which those relationships can be forged. In other words, adolescents who may have traditionally been understood as social isolates, due to their difficulties forming interpersonal relationships with face-to-face friends, now have other outlets for forming relationships which need to be brought to bear for understandings of peer networks in schools.

Previous research into adolescent friendship formation has primarily been analyzed through social needs and social compensation perspectives. The social needs perspective attributes the motivations behind adolescent development to personal needs for intimacy, self-validation, and companionship, whereas the social compensation perspective focuses more on the relationships that adolescents have with their parents to understand motivations behind friendship formation. Research supporting the social compensation approach is exemplified in the findings of Mesch and Talmud (2006). In a survey of a nationally representative sample of adolescent households in Israel (1000 in total), this study examined differences between adolescents who formed online friendships from those who did not, as well as adolescents' perceived strength of social ties in terms of the nature of initial contact (either online or face-to-face). The study found that adolescents reporting conflicts with parents turned to online friendships rather than face-to-face relationships, in part due to the anonymity of online communication. Further, this study's findings challenge previous research of the strength of social ties in online relationships, asserting that it is not technology which affects friendship formation, but rather the social embeddedness of the ties.

Social Networks, Educational Expectations, and Academic Performance

Social networks in schools have been demonstrated to significantly affect students' academic performance. Friendship ties with academically oriented peers have been found to produce academic advantages, particularly for youth in low-performing schools, suggesting that youth in a social environment with lower resources can be protected by the social capital generated in their academically oriented peer relationships (Crosnoe *et al.*, 2003). Further, in a study of friendship networks among 929 fifth-through seventh-grade children, Altermatt and Pomerantz (2005) found that respondents' grades were highly predicted by their friends' report card grades for that academic year – a positive educational achievement effect of their social networks. Nesting their findings within social comparison theory, the authors conclude that friendships with high-achieving peers benefit both high achievers and low achievers. They find moderate evidence however that low achievers' self-esteem decreases as result of friendships with peers who outperform them academically.

In several studies, hierarchy and clique development among students within classrooms have been linked to homogeneity of academic achievement and academic track placement. In a longitudinal study of 1477 pre-adolescents from fourth through seventh grade, Hallinan and Smith (1989) found that classrooms with low degrees of academic variance – especially with regard to high-ability students – also tend to have a low incidence of social clique formation. This finding suggests that academic tracking may have negative consequences for student social development, especially among gifted students. However, in classrooms with mixed-ability students in which the teacher stressed the importance of high academic achievement, the researchers found that cliques tended to form around homogeneity of achievement. This particular finding suggests that teachers – especially those with mixed-ability classes – should be mindful of how they organize their classroom with regard to maximizing academic achievement for all students.

Fuligni *et al.* (1995) maintain that the selection of adolescents into academic tracks should be seen as an important environmental change with impacts on their developmental transition from childhood to adulthood. Measuring adolescent's math grouping status over time in sixth, seventh, and tenth grade, they find that middle- and upper-level students benefit in both their math-related self-concept and academic performance; for low-ability-level students however, their self-concept initially increases, but decreases by tenth grade. Thus, low-ability grouped peers emerge with lower self-concepts than their nontracked peers.

Examples of studies which combine academic data with social network data serve as significant contributions to the study of how networks affect academic outcomes in

children and youth. Supplementing academic information in the Add Health study, the Academic Achievement supplement to Add Health (AHAA) facilitates the measurement of the effect of social networks on academic achievement, controlling for different family, school, and classroom contexts. Early research using this data has found strong relationships between peer networks and advanced mathematics course taking, a factor that weighs considerably in youth's academic careers (McFarland, 2006). A similar study using this data set also shows that female friendships boost advanced mathematics course taking and counter the traditional drop-off in female adolescent participation in advanced math and sciences (Riegle-Crumb *et al.*, 2006).

Social networks need not involve students to promote student performance. Morgan and Sørensen (1999) argue that intergenerational social closure – in this case, dense network connections between the parents of students – has both the potential for negative as well as positive effects on academic achievement, depending on the resources available within the community. They suggest that horizon-expanding schools and organizations shift the roles of monitoring norms and disseminating expectations away from the network actors and toward professionals and others outside of the primary network. Using parent and student-level data from the National Educational Longitudinal Study of 1988, the authors compare math test score gains in horizon-expanding schools and norm-enforcing schools, finding that horizon-expanding network configurations produce greater academic benefits than do norm-enforcing schools in the public school context.

Furthermore, social trust generated within school institutions has been found to generate greater efficiency and effectiveness in educational interventions, with teachers as members activating capital transferred from administrators, reform experts, and/or other peers. Similar to the concept of functional specificity, Frank *et al.* (2004) find that taxing or diffusing social capital reduces its effectiveness. In other words, the less pressures administrators place on their teachers at one time, the more efficiently they are able to tap available resources in the service of implementing educational innovations of benefit to their students.

Social Networks, Deviant Behaviors, and Health

The influence of peer groups on children and youth's development yields both positive and negative outcomes. Peer networks have been demonstrated to promote prosocial behaviors, such as extracurricular participation and leadership (Brown, 1990; Elder, Jr. and Conger, 2000). However, peer groups can also contribute to negative outcomes for youth, such as increased participation in

antisocial behaviors, often to the detriment of their educational futures. Studies examining the operation of delinquent behavior among adolescent peer groups have found significant network effects – in particular, characteristic differences between antisocial and prosocial peer networks (Giordano *et al.*, 1986). In a study of how the peer networks of aggressive children function to promote bullying and other deviant behaviors, Cairns *et al.* (1988) examined the role of aggressive children within their social networks in school, investigating how network structure relates to antisocial behavior. Using cluster analyses and best-friend nominations, they find that aggressive youth do not differ from control subjects in the degree of social cluster membership – in fact, many are often solid, central members of peer groups. These findings demonstrate that aggressive youth are not more likely than other students to become social isolates, and do have networks of peer support, despite being disliked for their behavior.

The development of increasingly sophisticated sociometric data has enabled researchers to examine the etiology of peer socialization into behavior norms, such as the process by which socially isolated youth are initialized into gangs and commit crimes against property. In analyzing these particular delinquent behaviors, Kreager (2004), using Add Health data to identify social isolates, found that isolation alone does not predict future delinquency, which replicates the findings of previous studies (see Haynie, 2001). However, when isolation is combined with peer conflict, or otherwise negative peer encounters, significant increases in delinquency and delinquent peer associations were measured.

In a related study, van Lier *et al.* (2005) support the significance of peer rejection as a predictor of deviance in their study of antisocial behavior among French-Canadian and Dutch boys and girls. Using peer nominations to classify antisocial behavior developmentally from childhood to early adolescence, the researchers employ a network analysis and find that peer rejection most greatly correlates with antisocial behavior and occurs more readily in youth involved in high-delinquency behavior patterns. This suggests that the process by which homophily occurs is a consequence of pre-existing preferences rather than a result of the socializing norms of the peer group.

Additionally, health-related behaviors of youth have been closely linked to social network affiliations. Early substance use has been tied to social networks, in particular, adolescent drug use (Kandel, 1978), drinking (Stattin *et al.*, 1989), and cigarette smoking (Alexander *et al.*, 2001). Moreover, early and risky sexual behavior has been linked with peer-group membership (Bearman *et al.*, 2004). Advances in social network theory and design offer increasingly rigorous and nuanced evidence of how children's behavior and educational futures are affected by the social networks in which they are embedded.

See also: Children's Friendship; Early Social Development and Schooling; Peer Interaction and Learning; Peer Learning in the Classroom; Peer Relations and Socialization of Children and Adolescents with Special Needs; Perspectives on Schooling in the Middle Years; Social Capital, Educational Institutions and Leadership; Social Development and Schooling.

Bibliography

- Alexander, C., Piazza, M., Mekos, D., and Valente, T. (2001). Peers, schools, and adolescent cigarette smoking. *Journal of Adolescent Health* **29**(1), 22–30.
- Altermatt, E. R. and Pomerantz, E. M. (2005). The implications of having high-achieving versus low-achieving friends: A longitudinal analysis. *Social Development* **14**(1), 61–81.
- Barber, B. L., Eccles, J. S., and Stone, M. R. (2001). Whatever happened to the jock, the brain, and the princess? Young adult pathways linked to adolescent activity involvement and social identity. *Journal of Adolescent Research* **16**, 429–455.
- Bearman, P., Moody, J., and Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology* **110**, 44–91.
- Brown, B. B. (1990). Peer groups and peer culture. In Feldman, S. S. and Elliott, G. R. (eds.) *At the Threshold: The Developing Adolescent*, pp 171–196. Cambridge, MA: Harvard University Press.
- Brown, B. B., Eicher, S. A., and Petrie, S. (1986). The importance of peer group ("crowd") affiliation in adolescence. *Journal of Adolescence* **9**, 73–96.
- Brown, B. B. and Lohr, M. J. (1987). Peer group affiliation and adolescent self-esteem: An integration of ego identity and symbolic interaction theories. *Journal of Personality and Social Psychology* **52**, 47–55.
- Bryk, A. S. and Schneider, B. (2002). *Trust in Schools: A Core Resource for Improvement*. New York: Russell Sage Foundation.
- Cairns, R. B., Cairns, B. D., Neckerman, H. J., Gest, S. D., and Gariépy, J. (1988). Social networks and aggressive behavior: Peer support or peer rejection? *Developmental Psychology* **24**, 815–823.
- Cairns, R. B., Leung, M. C., Buchanan, L., and Cairns, B. D. (1995). Friendships and social networks in childhood and adolescence: Fluidity, reliability, and interrelations. *Child Development* **66**, 1330–1345.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology* **94**(S), S95–S120.
- Coleman, J. S. (1990). *Foundations of Social Theory*. Cambridge: Belknap Press of Harvard University Press.
- Crosnoe, R., Cavanagh, S., and Elder, G. H. Jr. (2003). Adolescent friendships as academic resources: The intersection of friendship, race, and school disadvantage. *Sociological Perspectives* **46**(3), 331–352.
- Elder, G. H. Jr. and Conger, R. D. (2000). *Children of the Land: Adversity and Success in Rural America*. Chicago, IL: University of Chicago Press.
- Frank, K. A., Zhao, Y., and Borman, K. (2004). Social capital and the diffusion of innovations within organizations: The case of computer technology in schools. *Sociology of Education* **77**(2), 148–171.
- Fuligni, A. J., Eccles, J. S., and Barber, B. L. (1995). The long-term effects of seventh-grade ability grouping in mathematics. *Journal of Early Adolescence* **15**(1), 58–89.
- Fuligni, A. J., Eccles, J. S., Barber, B. L., and Clements, P. (2001). Early adolescent peer orientation and adjustment during high school. *Developmental Psychology* **37**, 28–36.
- Giordano, P. C., Cernkovich, S. A., and Pugh, M. D. (1986). Friendships and delinquency. *American Journal of Sociology* **91**, 1170–1202.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology* **78**, 1360–1380.
- Hallinan, M. T. and Smith, S. S. (1989). Classroom characteristics and student friendship cliques. *Social Forces* **67**, 898–919.

- Hallinan, M. T. and Tuma, N. B. (1978). Classroom effects on change in children's friendships. *Sociology of Education* **51**(4), 270–282.
- Hartup, W. W. (1996). The company they keep: Friendships and their developmental significance. *Child Development* **67**(1), 1–13.
- Haynie, D. L. (2001). Delinquent peers revisited: Does network structure matter? *American Journal of Sociology* **106**(4), 1013–1057.
- Jackson, M. F., Barth, J. M., Powell, N., and Lochman, J. E. (2006). Classroom contextual effects of race on children's peer nominations. *Child Development* **77**, 1325–1337.
- Kandel, D. B. (1978). Homophily, selection, and socialization in adolescent friendships. *American Journal of Sociology* **84**(2), 427–436.
- Kim, D. H. and Schneider, B. (2005). Social capital in action: Alignment of parental support in adolescents' transition to post-secondary education. *Social Forces* **84**(2), 1181–1206.
- Kreager, D. A. (2004). Strangers in the halls: Isolation and delinquency in school networks. *Social Forces* **83**(1), 351–390.
- Kubitschek, W. N. and Hallinan, M. T. (1998). Tracking and students' friendships. *Social Psychology Quarterly* **61**(1), 1–15.
- Marsden, P. V. (1990). Network data and measurement. *Annual Review of Sociology* **16**, 435–463.
- McFarland, D. A. (2001). Student resistance: How the formal and informal organization of classrooms facilitate everyday forms of student defiance. *American Journal of Sociology* **107**(3), 612–678.
- McFarland, D. A. (2006). Curricular flows: Trajectories, turning points, and assignment criteria in high school math careers. *Sociology of Education* **79**(3), 177–205.
- McFarland, D. A. and Pals, H. (2005). Motives and contexts of identity change: A case for network effects. *Social Psychology Quarterly* **68**(4), 289–315.
- McPherson, J. M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology* **27**, 415–444.
- Mesch, G. S. and Talmud, I. (2006). Online friendship formation, communication channels, and social closeness. *International Journal of Internet Science* **1**, 29–44.
- Moody, J. (2001a). Peer influence groups: Identifying dense clusters in large networks. *Social Networks* **23**, 261–283.
- Moody, J. (2001b). Race, school integration, and friendship segregation in America. *American Journal of Sociology* **107**, 679–716.
- Moody, J., McFarland, D. A., and Bender-deMoll, S. (2005). Dynamic network visualization. *American Journal of Sociology* **110**, 1206–1241.
- Morgan, S. L. and Sørensen, A. (1999). Parental networks, social closure, and mathematics learning: A test of Coleman's social capital explanation of school effects. *American Sociological Review* **64**(5), 661–681.
- Offer, S. and Schneider, B. (2007). Children's role in generating social capital. *Social Forces* **85**(3), 1–18.
- Plank, S. (2000). *Finding One's Place: Teaching Styles and Peer Relations in Diverse Classrooms*. New York: Teachers College Press.
- Riegle-Crumb, C., Farkas, G., and Muller, C. (2006). The role of gender and friendship in advanced course taking. *Sociology of Education* **79**(3), 206–228.
- Schneider, B. and Stevenson, D. (1999). *The Ambitious Generation: America's Teenagers, Motivated but Directionless*. New Haven, CT: Yale University Press.
- Stattin, H., Gustafson, S. B., and Magnusson, D. (1989). Peer influences on adolescent drinking: A social transition perspective. *Journal of Early Adolescence* **9**, 227–246.
- Steinberg, L., Brown, B. B., and Dombusch, S. M. (1996). *Beyond the Classroom: Why School Reform Has Failed and What Parents Need to Do*. New York: Simon and Schuster.
- van Lier, P. A., Vitaro, F., Wanner, B., Vuijk, P., and Crijnen, A. A. (2005). Gender differences in developmental links among antisocial behavior, friends' antisocial behavior, and peer rejection in childhood: Results from two cultures. *Child Development* **76**(4), 841–855.
- Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

Further Reading

- Burt, R. (1992). *Structural Holes*. Cambridge: Harvard University Press.
- Coleman, J. S. (1961). *The Adolescent Society. The Social Life of the Teenager and Its Impact on Education*. New York: Glencoe Free Press.
- Epstein, J. L. and Karweit, N. (eds.) (1983). *Friends in School: Patterns of Selection and Influence in Secondary Schools*. San Diego, CA: Academic Press.
- Friedkin, N. E. and Thomas, S. L. (1997). Social positions in schooling. *Sociology of Education* **70**, 239–255.
- Hallinan, M. T. (1980). Patterns of cliquing among youth. In Foot, H., Chapman, T., and Smith, J. (eds.) *Friendship and Childhood Relationships*, pp 321–342. New York: Wiley.
- Hallinan, M. T. and Sørensen, A. (1985). Ability grouping and student friendships. *American Educational Research Journal* **22**(4), 485–499.

Sociocultural Issues in Motivation

R A Walker, University of Sydney, Sydney, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Until recently, motivation has been considered to be an individual phenomenon. Motivational theorists have conceptualized key constructs in individualistic terms and emphasized the individual origins and nature of motivation, although they have also long recognized that contextual or social factors have a significant influence on these individual processes. Recently, theorists have suggested, after Vygotsky, that motivation is social in nature. Sivan (1986) first suggested this idea 20 years ago but it received a major impetus with the publication of an article by Hickey (1997) 11 years later. Since then, interest in the social nature of motivation has grown as a small number of book chapters and journal articles have been published, and conference papers have been presented on the topic. Although some motivational theorists (Winne, 2004) remain skeptical of this theoretical development, this article suggests that this perspective is worthy of consideration.

Although Vygotsky considered that intellectual and affective aspects of learning were interdependent, his own work foregrounded the intellectual aspects of learning. Sivan's (1986) seminal article provided the first sociocultural analysis of motivation and was based on conventional Vygotskian notions, such as the zone of proximal development (ZPD), assisted learning, peer learning, internalization, and intersubjectivity. Hickey's (1997) article, on the other hand, was primarily concerned with examining conventional motivation research in relation to instructional and educational reform practices derived from Vygotskian sociocultural ideas. Hickey's (1997) article coincided with a renewed interest in the situated and contextual nature of motivation and these ideas have tended to be seen as synonymous with sociocultural theory. While concern with context is a legitimate aspect of Vygotskian sociocultural theory, it is not the defining aspect of such a perspective. The key commitment of sociocultural motivational theorists is their recognition of the social nature and origins of motivation. Accordingly, sociocultural researchers attempt to explain how motivational goals, values, standards, and interests are socially constructed, and how they emerge and develop from social interactions and are manifested in collaborative and individual action.

As Walker *et al.* (2004) have pointed out, sociocultural theorists of motivation require reconceptualization of motivational concepts in social terms. With only a relatively small number of exceptions however, most motivational researchers (e.g., Perry *et al.*, 2006) have maintained

their commitment to motivation as an individual phenomenon while acknowledging the role of social factors, or context, as influences on motivation. To differentiate these theorists and researchers from sociocultural motivation researchers, Walker *et al.* (2004) have referred to them as social influence theorists, applying to the motivational domain a distinction that Rogoff (1998) made in the domain of learning and thinking. This distinction has been used in the current article to frame a discussion of theoretical and meta-theoretical issues and as a basis for selection of empirical research discussed in the article.

Sociocultural Theories of Motivation

The understanding of motivation as social in nature involves complex issues concerning the relationship between the social world and the world of the individual. Sociocultural theories endorse a social epistemology and accord analytical or theoretical primacy to the social world over the individual world, while recognizing that these worlds are closely interlinked and interdependent. According primacy to the social origins of motivation does not, however, mean that explanations of individual motivation can be reduced to social explanations or that social processes determine individual motivation. Sociocultural reductionism (Martin, 2006) and social determinism are avoided in sociocultural theories of motivation through theoretical notions which explain how the social world is internalized and externalized by individuals, and which assert that while there is a dynamic interdependence between the social and individual worlds, they are distinguishable and qualitatively different from each other. Valsiner (1997) has called this latter idea inclusive separation and uses it to explain how bidirectional exchanges occur between the social and individual worlds. Bidirectional exchanges are also explained through Vygotskian concepts such as the ZPD and the notions of transformative internalization and externalization. Transformative internalization and externalization explain how aspects of the social world are selectively internalized by the individual and then externalized in subsequent social interactions; these processes, as explained below, are active and transformative so that the social world is not imprinted on, or transmitted, to the individual, and the individual, in turn, has a unique impact on the social world through transformative externalizations. Taken together, these sociocultural theoretical ideas avoid the reduction of

personal phenomena to social interaction or social processes at large and recognize the agency of the individual. They explain how individual motivation can have social origins yet the individual's intrapsychological motivational functioning is still relatively autonomous from the social world.

These meta-theoretical notions have been endorsed in the work of some, but not all, sociocultural motivation theorists (Walker, 2006; Walker *et al.*, 2004; Pressick-Kilborn *et al.*, 2005). Hickey (2003) asserts that the social world, or context, and the individual are tightly bound or fused with each other, and appears to suggest that they are not distinguishable. This view raises important issues and problems concerning the motivational agency of the individual and the reductionist position of the individual in relation to the social world. Similarly, the common categorization (Hickey, 2003) of the sociocultural motivation worldview as contextualist is inadequate as it makes context the theoretical focus rather than the social origins of motivation.

Sociocultural motivation theorization has mostly focused on well-known and central aspects of Vygotsky's writing: culture and cultural practices, the ZPD, transformative internalization and externalization, and interpersonal relations and intersubjectivity. Taken together, these sociocultural ideas contribute to an explanation of the way that motivation, conceptualized as social in nature, is internalized to become an individual process. Although there is general agreement among sociocultural motivation theorists (e.g., Sivan, 1986; McCaslin, 2004; Walker *et al.*, 2004) on this view, it is not accepted by Hickey and Granade (2004) who consider that the goals and values that support and motivate learners to engage "reside in the practices of knowledgeable communities rather than the hearts and minds of individuals" (p. 224). Theorists such as Walker and associates (Walker, 2006; Walker *et al.*, 2004; Pressick-Kilborn and Walker, 2002), on the other hand, consider that while academic practices have motivational properties, motivation is an internalized attribute of human beings. For instance, aspects of the academic practices of reading and writing may be motivating to learners, but the motivation to engage in these practices is internalized as learners work collaboratively on these activities.

Culture and Cultural Practices

Culture and cultural practices are considered, from a sociocultural perspective, to play a critical role in the construction and emergence of motivation. Cultural practices (Miller and Goodnow, 1995) are recurrent actions or activities that may be maintained, changed, or challenged. They are valued by the communities that engage in them and are associated with a sense of belonging or identity and with particular forms of discourse. They help to

structure learning and thinking activities and have motivational and affective properties and consequences. The academic practices of the school and the classroom, such as those associated with reading and writing for instance, are one type of valued cultural practice. Sociocultural theorists (Walker, 2006; Walker *et al.*, 2004; Nolen, 2001, 2007) consider that motivation is socially constructed as learners engage in academic practices, and so can be considered to emerge from these practices. Sivan (1986) has presented classroom motivation as a socially constructed norm, which, like other classroom norms, forms the basis on which both teacher and student expectations and judgments are made. These expectations include how motivation is displayed in the classroom in the form of motivated behavior that demonstrates interest and willingness to engage in learning activities.

Zone of Proximal Development

The ZPD refers to the learner's ability to successfully complete tasks with the assistance of more capable other people, and for this reason it is often discussed in relation to assisted or scaffolded learning. The creation of ZPDs involves assistance with the cognitive structuring of learning tasks and sensitivity to the learner's current capabilities. Sociocultural (e.g., Sivan, 1986) and mainstream motivational theorists (e.g., Brophy, 1999) have observed that these aspects of the ZPD make it an inherently motivational zone; the ZPD is optimally challenging (Sivan, 1986) because tasks are calibrated to the learner's level, while appropriate support and scaffolding ensure that tasks can be completed successfully. Assistance from others also helps the learner to learn how to work on difficult tasks and how to control or manage anxiety and frustration in the process. Additionally, working within the ZPD is inherently motivating because it involves the transfer of responsibility, or control, for learning, from the teacher or more capable other to the learner. This transfer of control is motivating for the student as it acknowledges student mastery of the task, and hence the learner's developing efficacy. Interaction within the ZPD is also likely to lead to the recruitment of the learner's interest in the task or knowledge domain as the learner comes to value and appreciate the knowledge which is valued by a respected, more capable other person. Furthermore, as learners come to achieve mastery in a knowledge domain, they are more likely to appreciate the relevance and value of the knowledge domain.

The ZPD can also be considered to be a relational (Goldstein, 1999) or affective zone. Goldstein (1999) has characterized the ZPD as a socially mediated space that is formed through relationships involving sensitivity and trust. In a classroom, this space is created by the interactions between students and between students and their teacher, as they engage in supportive activities that

develop learner confidence and positive emotions. This consideration of the ZPD as a shared affective zone also has important motivational implications; the emotional quality and tone of interaction in the ZPD and the sense of caring engendered can have important implications for students' engagement in learning and willingness to challenge themselves.

The ZPD has been extended by Valsiner (1997) into a system of zones which recognizes not only the importance of assisted learning, but also the factors which may assist or constrain learning, and, as Pressick-Kilborn and Walker (2002) have suggested, motivation. Valsiner's system of zones has been used by Pressick-Kilborn and Walker (2002) to understand the development of interest in a classroom learning community.

Transformative Internalization and Externalization

The reciprocal notions of transformative internalization and externalization play a central role in many, but not all, sociocultural theories because they explain the processes through which aspects of the social world become a part of the world of the learner, and conversely how the learner's actions and behavior impact upon his/her social world. Viewed from the perspective of motivation in the classroom, the process of internalization refers to the way individuals selectively internalize values and standards from their interactions with others in the ZPD as they engage in the academic practices of the classroom. The process of internalization is active, constructive, and transformative (Walker *et al.*, 2004) so that the goals, values, and standards constructed by the learner cannot be considered to be transmitted by others. Rather, goals, standards, and values are actively modified or changed by the learner in the process of internalization. When standards and values have been internalized by a learner, they are subsequently externalized in the form of motivated action, behavior, and language, so that internalization may be inferred from these expressions of classroom engagement. It is important to recognize, however, that externalization is also an active and transformative process, so that standards and values are transformed as learners externalize them through their interaction with peers and others.

While the notion of internalization has been prominent in the work of sociocultural motivation theorists, transformative externalization and its relevance for understanding motivation and engagement have so far only been raised in the work of Walker and associates (Walker *et al.*, 2004; Pressick-Kilborn *et al.*, 2005). Walker and associates have identified the role that transformative externalization plays in motivated action and explained how this impacts on the social world. While most theorists (e.g., Sivan, 1986; McCaslin, 2004; Walker *et al.*, 2004)

consider internalization to play an important role in motivation and engagement, Hickey has argued that it is of little importance (Hickey, 2003; Hickey and Granade, 2004). In part, Hickey's assertion is based on a lack of recognition of the transformative nature of internalization, but, in part, it is based on the view that goals, standards, and values are tightly bound to the context in which they are constructed, thus making internalization irrelevant to the understanding of motivation. This view is problematic in that it does not recognize that these goals, standards, and values may become personal attributes that motivate engagement in other contexts.

Interpersonal Relations, Intersubjectivity, and Co-Regulation

The nature and quality of interpersonal relationships between students and their teachers and peers are therefore important in sociocultural perspectives on motivation as they influence the internalization and externalization of motivational standards and values. Interpersonal relations and intersubjectivity are also important for understanding the way that learners and their peers regulate each other's activities, and their motivated engagement in those activities. The notion of co-regulation has been developed in the work of McCaslin *et al.* (2004) and refers to the regulation that teachers provide students and that students provide for each other as they work on activities in the ZPD.

Sociocultural Research in Motivation

With the exception of McCaslin's work, only a small number of sociocultural motivation studies have been conducted to date. These studies have been conducted in different motivational domains and while they have adopted differing sociocultural theoretical positions, they all have a commitment to investigating the social nature of motivation. While some of the studies are developed from an explicit Vygotskian theoretical framework (McCaslin, 2004; Pressick-Kilborn and Walker, 2002), others (Nolen, 2001, 2007; Middleton and Perks, in review) do not ground their research in this framework. Most of the studies have been conducted in naturalistic classroom contexts.

McCaslin outlined a number of research projects undertaken over a 20-year period from the mid-1980s to the early 2000s, all of which were based on sociocultural ideas, particularly those related to the ZPD and internalization (McCaslin and Hickey, 2001; McCaslin, 2004). The projects focused on interpersonal relationships at home and in school and were developed from a model of co-regulation which was considered to facilitate the internalization of social supports and promote adaptive

learning. Co-regulation involves the joint control of an activity by all participants involved in the activity and is enacted by negotiation between the participants. Adaptive learning (McCaslin, 2004) was defined as involving “the internalization of goals, the motivation to commit, challenge, or reform them; and the competence to enact and evaluate those commitments” (p. 254). As such, adaptive learning research examined the opportunities various home and school environments and tasks allowed for the integration of both affective and intellectual aspects of learning. An important recognition in these projects was the understanding that co-regulation in the ZPD does not always enhance motivation and learning and may actually impede it. That is, while co-regulation may have benefits for learners, sometimes learners influence and regulate each other in ways that lead to diminished learning outcomes and reduced motivation. Taken together, McCaslin’s (2004) projects illustrate the potential of Vygotskian sociocultural ideas for understanding the following aspects of student motivation: “(1) the function of task opportunity and the enactment of self-directed activity and motivational beliefs, (2) the affordances of teacher classroom management for student-learning cultural rules of responsibility and citizenship (3) the influence of parent beliefs and behaviour on children’s emergent identity and the negotiation of personal commitments at school, and (4) the power of peer participation in student-valuing curriculum tasks, learning motivation, and motivation to learn” (p. 250). McCaslin’s (2004) research demonstrates the relevance of Vygotskian sociocultural ideas for understanding the social nature of motivation while also recognizing the place of the individual within the sociocultural realm.

Pressick-Kilborn and Walker (2002) reported an investigation into the emergence and development of interest in a primary school classroom. In the study, interest was conceptualized as emerging from collaborative activities in a fifth-grade science community of learners. This conceptualization emphasized the dynamic and interdependent relationship between the children and their classroom activities for understanding the development of interest and suggested that traditional distinctions between situational and personal interest were problematic. Studying interest in a real and complex classroom setting presented challenges to distinguishing situational and individual interest along traditional lines. While, in some activities, students participated in ways that indicated an ongoing and more personally meaningful form of interest in learning, the situational aspect of interest was always apparent. Features of tasks, such as their hands-on or collaborative nature, contributed to personal experiences of interest in learning in the context of a real classroom. The analysis of the development of interest in the study drew on the integration of several established sociocultural notions: the idea of a community of learners, the ZPD and its extensions (Valsiner, 1997), and Valsiner’s

(1997) notions of canalization and self-canalization. Canalization refers to the way that cultural practices channel the activities of individuals, while self-canalization refers to the way individuals are able to channel and direct their own activities. The canalization of opportunities in the social world thus creates the context in which values and goals are internalized and from which interest may subsequently emerge. It is important to recognize, however, that individuals may resist these canalization processes.

The investigation extended over a period of 6 months and involved ethnographic observations in the classroom community, and student and teacher interviews. Students also created personal-interest trajectories to indicate their perceived levels of interest in specific science-based activities during classroom lessons and over the duration of the study. These trajectories allowed insights into changing patterns of interest both at the individual and class level. Pressick-Kilborn and Walker (2002) developed individual case studies from their research data and used them to show how the classroom community both promoted and constrained the emergence and development of interest. The case studies illustrated not only how classroom science practices canalize interest development, but also how processes of self-canalization may lead students to resist engagement in these activities, and thus limit their interest development.

Nolen (2001, 2007) investigated, in a longitudinal study, children’s emergent motivation to read and write and its relation to their developing sense of the nature of reading and writing. The study investigated teachers and their students across a 4-year period, from kindergarten to the end of third grade, during which time they were observed in class activities and interviewed annually about reading and writing. The study examined teachers’ goals in literacy instruction and the literacy activities that students engaged in, and drew relationships between these activities, the students’ understanding of the nature and purpose of literacy, and their motivation to engage in literate activities. A particular focus of the research was the way in which teachers and students jointly constructed their literacy activities and, where such opportunities were limited, how students collaboratively reconstructed the literacy activities they were required to engage in.

The first year of the study involving kindergarten teachers and their students was reported by Nolen (2001), although the focus of this publication was on a subset of students who were at risk for reading and writing difficulties. Classroom observations showed that while all of the teachers engaged in common literacy activities, they differed considerably in the variety and amount of time spent in various literacy activities. Teachers differed in the extent to which they gave students the opportunity to read and write connected text, to choose their own subject matter and activities, and in how structured classroom activities were. The teachers

also differed in the extent to which they emphasized the motivation for learning to read and write, skill development, small and large group instruction, multiple task structures, and autonomous student work. Student interviews showed how student understanding of the nature and purpose of literacy is related to commonly employed literacy activities, and how literacy practices and activities are related to student engagement in and liking of those activities.

The remaining 3 years of the study, from grades 1 to 3, were reported by Nolen (2007) and demonstrated that as the social meaning of literate activity was co-constructed in classroom communities, specific motivations emerged in relation to reading and writing. As with the earlier study, the most frequent and valued activities in classrooms had the greatest impact on students' perceptions of themselves as readers and writers. Where the emphasis, for instance, was on writing for self-expression and for authentic reasons (an author's tea involving family and friends) students were developing identities as writers. Some instructional activities also provided opportunities for choice and autonomy while others were more controlling and allowed little room for independent learning. Student motivations, such as interest, enjoyment, mastery, and ego concerns, emerged from these differing instructional contexts and activities. Their motivations and actions were also shown to be related to the students' differing developmental skill trajectories.

Middleton and Perks (in review) present a view of mastery goal motivation as socially constructed through the written dialogic interactions of students with their teacher in a writing course. The study examined the writing and motivational goals of 17 students enrolled in an elective creative writing class in a rural high school in the United States. A writing workshop approach in the class involved students in extended periods of writing as well as regular conferencing with their teacher. The students were also required to keep dialog folders which contained samples of their writing and a written dialog addressed to their teacher concerning their writing and how it was progressing. They were also asked to identify kinds of feedback that would support them in their writing. The teacher collected these folders at the end of every lesson and responded in writing to each of the students. These dialog folders, many over 30 pages in length, were subjected to a layered discourse analysis which focused on the situated identities of the students, the discourses they engaged when writing, the terminology and language used, and the cultural models enacted by the students and teacher. The dialog folders were also analyzed for the enactment of mastery goals.

The findings of the discourse analysis were developed into a model of mastery engagement in the writing domain. The model suggests that students need to situate themselves as learners of the craft of writing and engage

in the discourse of the craft, and students will engage in mastery learning when they adopt the cultural models of the craft. Learning the craft of writing requires the valuing of writing skills, learning the discourse of the craft, and the adoption of the cultural norms and ways of thinking inherent within the discourse. As such, the goal of becoming a good writer is intertwined with the adoption of mastery motivational goals.

Conclusion

Sociocultural motivation theorists are concerned with the social origins of motivation but recognize that individual motivational processes are neither determined by nor reducible to social processes. They use ideas from the Vygotskian tradition, such as the ZPD and transformative internalization and externalization, to explain how motivation has social origins yet individual agency and the relative autonomy of individual motivational processes are acknowledged. These notions have been prominent in sociocultural-motivation theoretical writing and in some of the empirical research reported, and this trend is likely to continue in the future with investigations involving areas such as the ZPD, intersubjectivity, and caring interpersonal relationships also becoming more prominent. Motivation theorists are also likely to explore other sociocultural perspectives, not explicitly based on Vygotskian ideas, in their attempts to explain the social origins and social construction of motivation. Investigations of the social origins of classroom motivation may increase in number as the trend toward studying motivation in naturalistic classroom contexts (Jarvela and Volet, 2004; Perry *et al.*, 2006) continues to gather momentum since classrooms are inherently social environments and are therefore best researched through theoretical approaches which recognize the social nature of motivation.

Bibliography

- Brophy, J. (1999). Toward a model of the value aspects of motivation in education: Developing appreciation for particular learning domains and activities. *Educational Psychologist* **14**, 75–85.
- Goldstein, L. S. (1999). The relational zone: The role of caring relationships in the co-construction of mind. *American Educational Research Journal* **36**, 647–673.
- Hickey, D. T. (1997). Motivation and contemporary socio-constructivist instructional perspectives. *Educational Psychologist* **32**, 175–193.
- Hickey, D. T. (2003). Engaged participation versus marginal non-participation: A stridently sociocultural approach to achievement motivation. *Elementary School Journal* **103**, 401–429.
- Hickey, D. T. and Granade, J. B. (2004). The influence of sociocultural theory on our theories of motivation and engagement. In McInerney, D. M. and Van Etten, S. (eds.) *Big Theories Revisited*, pp 223–247. Greenwich, CT: Information Age.
- Jarvela, S. and Volet, S. (2004). Motivation in real-life, dynamic, and interactive learning environments: Stretching constructs and methodologies. *European Psychologist* **9**, 193–197.

- Martin, J. (2006). Social cultural perspectives in educational psychology. In Alexander, P. and Winne, P. (eds.) *Handbook of Educational Psychology*, 2nd edn., pp 595–614. Mahwah, NJ: Erlbaum.
- McCaslin, M. (2004). Coregulation of opportunity, activity, and identity in student motivation: Elaborations on Vygotskian themes. In McInerney, D. M. and Van Etten, S. (eds.) *Big Theories Revisited*, pp 249–272. Greenwich, CT: Information Age.
- McCaslin, M. and Hickey, D. T. (2001). Self-regulated learning and academic achievement: A Vygotskian view. In Zimmerman, B. J. and Schunk, D. H. (eds.) *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*, 2nd edn., pp 227–252. Mahwah, NJ: Erlbaum.
- Middleton, M. and Perks, K. (in review). Motivation for mastery: Mastery motivation as a socially constructed form of engagement (Achievement goals as a negotiated activity: A sociocultural examination of student–teacher dialogue in a writing classroom). *Paper Presented at the 10th International Conference on Motivation*. Landau, Germany, 28–30 September 2006.
- Miller, P. J. and Goodnow, J. J. (1995). Cultural practices: Toward an integration of culture and development. In Goodnow, J. J., Miller, P. J., and Kessel, F. (eds.) *Cultural Practices as Contexts for Development*, pp 5–16. San Francisco, CA: Academic Press.
- Nolen, S. B. (2001). Constructing literacy in the kindergarten: Task structure, collaboration, and motivation. *Cognition and Instruction* **19**, 95–142.
- Nolen, S. B. (2007). Young children's motivation to read and write: Development in social contexts. *Cognition and Instruction* **25**(2), 219–270.
- Perry, N. E., Turner, J. C., and Meyer, D. K. (2006). Classrooms as contexts for motivating learning. In Alexander, P. A. and Winne, P. H. (eds.) *Handbook of Educational Psychology*, 2nd edn., pp 327–348. Mahwah, NJ: Erlbaum.
- Pressick-Kilborn, K. and Walker, R. A. (2002). The social construction of interest in a learning community. In McInerney, D. M. and Van Etten, S. (eds.) *Sociocultural Influences on Motivation and Learning*, vol. 2, pp 153–182. Greenwich, CT: Information Age.
- Pressick-Kilborn, K., Sainsbury, E., and Walker, R. A. (2005). Making sense of theoretical frameworks and methodological approaches: Exploring conceptual change and interest in learning from a sociocultural perspective. *Australian Educational Researcher* **32**(2), 25–47.
- Rogoff, B. (1998). Cognition as a collaborative process. In Damon, W., Kuhn, D., and Siegler, R. (eds.) *Handbook of Child Psychology*, vol. 2, 5th edn., pp 679–744. New York: Wiley.
- Sivan, E. (1986). Motivation in social constructivist theory. *Educational Psychologist* **21**, 209–233.
- Valsiner, J. (1997). *Culture and the Development of Children's Action: A Theory of Human Development*, 2nd edn. New York: Wiley.
- Walker, R. A. (2006). Internalizing motivation. Symposium (Chair: Jarvela, S. and Hickey, D.): Where social and self meet in future conceptualizations of engagement. The concept of motivation and the field of motivation research. *10th International Conference on Motivation*. Landau, Germany, 28–30 September.
- Walker, R. A., Pressick-Kilborn, K., Arnold, L. S., and Sainsbury, E. (2004). Investigating motivation in context: Developing sociocultural perspectives. *European Psychologist* **9**, 245–256.
- Winne, P. H. (2004). Comments on motivation in real-life, dynamic, and interactive learning environments: Theoretical and methodological challenges when researching motivation in context. *European Psychologist* **9**, 257–263.

Further Reading

- Hickey, D. T. and McCaslin, M. (2001). A comparative and sociocultural analysis of context and motivation. In Volet, S. and Jarvela, S. (eds.) *Motivation in Learning Contexts: Theoretical and Methodological Implications*, pp 33–56. Amsterdam: Elsevier.
- Lawrence, J. and Valsiner, J. (1993). Conceptual roots of internalization: From transmission to transformation. *Human Development* **36**, 150–167.
- Nolen, S. B. (2006). The role of social context in the development of motivation to write. In Hidi, S. and Boscolo, P. (eds.) *Motivation to Write*, pp 241–255. Dordrecht: Kluwer.
- Valsiner, J. (1997). Magical phrases, human development and psychological ontology. In Cox, B. D. and Lightfoot, C. (eds.) *Sociogenetic Perspectives on Internalization*, pp 237–255. Mahwah, NJ: Erlbaum.
- Valsiner, J. (1998). Dualisms displaced: From crusades to analytic distinctions. *Human Development* **41**, 350–354.
- Van der Veer, R. and Valsiner, J. (1991). *Understanding Vygotsky: A Quest for Synthesis*. Cambridge, MA: Blackwell.

Socioemotional Aspects of Technology-Supported Learning

S Järvelä, University of Oulu, Finland

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Computer supported collaborative learning

(CSCL) – CSCL is a method of supporting collaborative learning using computers and the internet. The purpose of CSCL is to scaffold or support students in learning together effectively.

Contextual motivation – Motivation is no more a separate variable or a distinct factor, which can be applied in explanation of an individual readiness to act or learn – but reflective of the social and cultural environment.

Self-regulated learning – Self-regulated can be used to describe learning that is guided by metacognition, strategic action, and motivation to learn.

Socioemotional aspects of learning – Study of an interaction of affect and cognition including especially motivation and emotion in social interaction.

Introduction

Over the last few years, information and communication technology has opened up new avenues for designing learning and instruction. Technology-supported environments provide arenas for social, interactive, shared, and collaborative activity, as well as new tools for individual learning – in education and the workplace context. Computers are used to facilitate and redefine learning interactions, including those among students, those between students and teachers, and those between students and others in the broader out-of-school community (Renniger and Shumar, 2002). At the same time, these new and more open environments are characteristic of modern society: for example, they are complex, dynamic, less structured with mixed-motive situations, and contain multiple competing, or even contradictory, adaptive demands. Researchers often assume that technology-supported learning environments will sufficiently motivate and cognitively engage students as they study, interact, and collaborate in variety of virtual, computer supported, and online environments. This assumption may not be well founded. Although (often) constructivist-based learning environments incorporate features that may enhance student motivation to learn and

cognitive engagement far more than traditional classrooms, these environments also introduce difficulties that may interfere with students' willingness to engage.

Technology-Supported Learning Environments

During the past 20 years, technology has played an important role in many attempts to create powerful learning environments for supporting learning for understanding (De Corte *et al.*, 2003). From the beginning, the effects of the various forms of computer use on students' learning and motivation and social behavior have been a source of debate and a topic of continuing research. In the early years, the interest in research on motivation and social aspects was minor since the main interest was dealing with more cognitive-oriented research – essentially, to what extent working with computers affects what one does and how well one does it. One of the characteristics of that time was to use the power of new technology for the so-called solo-learner model and to find opportunities to individualize the learning process (Collins, 1986; Papert, 1980; Taylor, 1980). During the past 10 years, the situation has changed. Most of the recent research on the use of information and communication technology in education is more or less explicitly considering technology's possibilities for facilitating social interaction between teacher and students, and among students (Dillenbourg *et al.*, 2008; Barab *et al.*, 2004). Many of the current studies focus on computer-supported collaborative learning facilitated by different network-based collaboration tools and a learning theory based on pedagogical ideas (Koschmann *et al.*, 2002; Srijbos *et al.*, 2003). There is also increased ongoing discussion about online learning, virtual interaction, and wireless devices as cognitive and interactive tools in academic learning, as well as about learning at work (Pea and Maldonado, 2006; Renniger and Shumar, 2002).

Influence of Pedagogical Features in Technology-Supported Learning Environments on Motivation

In contrast to traditional school settings, which are usually well prepared, organized, and controlled by the teachers, the goals of self-organized learning and true student responsibility characterize these new technology-supported

pedagogical cultures (Hartley and Bendixen, 2001). The shift from teacher centeredness toward students' activity presupposes strong self-regulative efforts from students and, at the same time, offers more space for the individual's activities (Edelson *et al.*, 1999). This kind of meaningful and close relationship toward the learning tasks also helps students increase their intrinsic motivation and interest (Järvelä *et al.*, 2001). In addition, the often-mentioned increase in authenticity or anchored instruction to real-world problems with technological simulation or multimedia may have motivational implications (Cognition and Technology Group at Vanderbilt, 1996).

One recent pedagogical model applied to technology environments deals with problem- and inquiry-based learning (Brown and Campione, 1996; Edelson *et al.*, 1999). These pedagogical ideas' contributions and motivational implications are quite clear. The process of knowledge-seeking inquiry starts from cognitive or epistemic goals that arise from the learner's cognitive needs and that cannot be achieved by relying on available knowledge. The learner has a close and meaningful cognitive relationship with the learning task, which contributes to the intrinsic quality of motivation (Ames, 1992).

A technology-supported learning environment can also change the nature of social interaction (Bliss *et al.*, 1999). It presupposes collaborative activities, which usually lead toward the sharing of cognitive achievements with other students and members of the learning community (Stahl, 2004). Although there are contradictory findings concerning the success of collaboration in learning (Salomon and Globerson, 1989), some aspects of social interaction carry motivational implications. For example, peers provide models of expertise. Observing the progress of other students may increase confidence in one's own ability to succeed (Bandura, 1997). Furthermore, peer models provide a benchmark for the students' self-evaluations, thereby helping them to set proximal or more accurate goals (Schunk, 2001).

Collaborative learning, with and without technology, includes a variety of shared processes where individuals aim to regulate the prerequisites for learning together, and an increasing amount of studies emphasize the meaning of motivation and emotions for successful collaboration (Crook, 2000). Social learning situations, where the individuals' characteristics, goals, and demands meet, can evoke emotions and create novel motivational challenges for them (Järvelä *et al.*, 2000; Thompson and Fine, 1999). In collaborative learning processes, these socioemotional conflicts can emerge due to a variety of reasons originating from, for example, individual differences, cognitive conflicts, or modes of interaction. For instance, collaborative learning models presume that group members create a shared conception of a task and then try to reach this goal by equally sharing the responsibility of the learning process (Roschelle and Teasley, 1995). This requires constant negotiation and argumentation between the students

as well as adjustment of individual conceptions and goals, especially when the context of learning is not a face-to-face situation, that is, virtual or computer mediated. Often, the same situations are also socioemotionally challenging and can act as competitive motives, interruptions, or obstacles to motivated action in different phases of the learning process (Järvenoja and Järvelä, 2005).

Empirical Findings on Motivation and Emotion in Technology-Supported Learning Environments

Although there is evidence that students respond positively to the technology-supported learning environments designed according to the constructivist learning theory principles (e.g., Hickey *et al.*, 2001; Mistler-Jackson and Songer, 2000), it remains unclear whether students who enthusiastically participate do get cognitively involved. At the moment, there are not enough systematic empirical research and findings about specific motivational components, such as goals, self-beliefs, efficacy, or values; however, there is increasing interest among motivation researchers to extend their analyses to the new learning contexts.

The investigation of the role of goals and goal orientation has been an important recent development in the achievement motivation theory and research on learning and motivation (Pintrich, 2000). Goals are generally defined as perceptions about the purposes of achievement behavior, representing the meaning that individuals assign to achievement situations (Dweck, 1986). The findings indicate that computer-supported inquiry learning seems to foster productive task-related interaction and enhance student motivation in general (Blumenfeld *et al.*, 1991; Hakkarainen *et al.*, 2002). Similar findings on more enduring adaptive tendencies have been reported in other studies of student learning in computer-supported environments, applying socioconstructivist pedagogical models (CTGV, 1992; Hickey *et al.*, 2001). For example, the findings of Hug *et al.* (2005) in project-based science learning show that student interactions with technology scaffolded their engagement. The students in the study by Hug *et al.* learned to ask meaningful questions and discuss the ideas behind these questions. Järvelä and Salovaara (2004) compared students' achievement goals between the computer-supported collaborative learning group and the control group and found statistically significant differences in learning orientations. In the computer-supported collaborative learning group, learning orientation was maintained during the 3 years of the study, whereas in the control group, it decreased. The findings indicate that when students work in computer-supported and inquiry-based environments, they are engaged and work on the task because they have more freedom to choose their tasks and apply individual learning goals and strategies (Veermans and Järvelä, 2004).

Self-regulated learning has been defined as an active, constructive process whereby learners set goals for their learning and then attempt to plan, monitor, regulate, and control their cognition, motivation, behavior, and context (Boekaerts *et al.*, 2000). It is generally acknowledged that technology-supported learning environments are facilitators for the acquisition of self-regulatory skills (Boekaerts, 1999). The issues of openness, choice, and control of learning tasks in technology-supported environments compared to traditional learning environments may especially stress the individual students' volitional processes. Salovaara and Järvelä (2003) have shown that computer-supported collaborative environments enhance strategy use and intentional learning. Students reported deep-level cognitive strategies such as monitoring and creating knowledge representations and more collaborative information sharing compared to control participants. Similar findings were received in the study by Shell *et al.* (2005) where students' working with established computer-supported collaborative classes reported more knowledge-building goals and more question asking than fellow students. Whipp and Chiarelli (2004), in turn, investigated students' self-regulation in a web-based course. They studied how higher education students used and adapted traditional self-regulation strategies to complete tasks and cope with challenges in a web-based online course. According to their results, the students not only used many traditional strategies, but also adapted planning, organization, environmental structuring, help-seeking, monitoring, and self-reflection strategies in ways that were unique to the online technology context.

There are also less-adaptive consequences to technology-supported learning in terms of students' socioemotional interpretations. The responsibility of setting up one's own learning goals and monitoring one's own learning activities in various online learning environments and computer-supported inquiries can also be quite demanding for some students (Veermans and Järvelä, 2004). Earlier research in traditional classroom learning contexts has shown that there are individual differences in students' dispositions toward challenge, with some responding positively and others avoiding taking up difficult work (Meyer *et al.*, 1997). Moreover, there are also within-student individual differences in terms of how a student may respond to a given task across an inquiry unit (Patrick and Middleton, 2002).

In general, academic emotions are significantly related to student motivation, learning strategies, cognitive resources, self-regulation, and academic achievement (Pekrun *et al.*, 2002). Nevertheless, not only do the emotions themselves vary, but also the sources that cause emotions. Learning situations are important sources of emotions that instigate variety of self-referenced, task-related, and social emotions (Meyer and Turner, 2002). Concerning technology-supported learning environments and students' emotional expressions, Järvenoja and Järvelä's (2005) study investigated

what kind of explanations the students give to their emotional experiences related on computer-supported collaborative learning. One of the critical features found was that, especially in the beginning of the learning project, the self-driven emotions played an important role in order to inhibit or facilitate task involvement. Volet and Wosnitza's (2004) study examined the origin, direction, and impact of emotions in social online learning. Their analysis of social online-learning situations revealed a range of other-directed emotions, in addition to self-, task-, and technology-directed ones. Emotions generated in social online environments are not different in nature from those generated in face-to-face learning situations. What is different in social online learning is the fact that emotions are expressed through technology, and that the disclosure of emotions is necessarily voluntary. The results highlight the multiple directions that emotions can take and the significance of students' interpretations of their emotions on the learning process.

Theoretical Approaches to Investigate Motivation in Technology-Supported Learning Context

While the aim of research has been to understand socioemotional processes in a technology-supported learning environment, there has also been an interest in understanding the learning process by, with, around, and through the computers. The leading questions have been: How does interaction between student motivation and situational features of the technology-supported environment take shape? How do different students cope with the demands of a situation and what is the potential provided by the features of the learning context to their motivational interpretations?

The overall picture of the learners' adaptation to educational environments has increased the researchers' interest in better understanding the interaction between personal and contextual factors in the motivation and achievement of students (Pintrich, 2000). Through the influence of sociocultural and situated cognition theories, it has been recognized that the motivation of individual learners is also influenced by social values and the context in which the learning takes place. Motivation is no longer a separate variable or a distinct factor, which can be applied in explanation of an individual readiness to act or learn – but reflective of the social and cultural environment (Järvelä and Volet, 2004).

The notion of contextual motivation has been widely debated over the last few years (Anderman and Anderman, 2000; Hickey, 1997; Volet and Järvelä, 2001). Conceptualizing motivation in learning contexts builds upon the situated learning paradigm, which views the process of learning as distributed across the learner and the environment in which knowing occurs, and the activity in which the learner is participating (Anderson *et al.*, 2000). This conceptual

framework has provided a useful foundation for understanding students' goals, intentions, and emotions across situations – in real contexts and real time, and the context–person mutual influences have been highlighted in various technology-supported learning environments (Järvelä and Niemivirta, 2001).

Another tendency in research has been to extend the individual analyses of self-regulated learning and motivation to social settings mainly because social and collaborative aspects are emerging features in technology-supported learning models. Although self-regulation research has traditionally focused on the individual perspective (Boekaerts *et al.*, 2000), there is increasing interest in considering the mental activities that are part of self-regulated learning at the social level, with reference to concepts such as social regulation, co-regulation, and shared regulation (McCaslin and Hickey, 2001; Salonen *et al.*, 2005). With regard to the role of the social aspects of learning, these models have examined how the social context plays a role in the generation of cognitions and pursuit of personal goals and, alternatively, how individuals regulate their social context – among other objects of regulation such as cognitions, motivation, and emotions – in order to achieve their goals.

Methodological Challenges

When considering the methodologies for studying the socioemotional aspects of technology-supported learning, challenges remain. The earlier mainstream research on motivation, which acknowledges the impact of context (e.g., research on perceptions of classroom goal structures, see Ames, 1992), explains it on the level of goals, for example, but context is not made operational in real time with real tasks, the here and now of learning. Time perspective is longitudinal or cross-sectional, but not online. Another methodological limitation deals with empirical evidence with social dimensions of motivation.

The value of qualitative methods has lately received recognition in educational psychology, in general (Patrick and Middleton, 2002; Perry, 2002), and in classroom motivation research, in particular (Dowson and McNerney, 2003; Turner and Meyer, 2000), where it has traditionally been neglected more than in general educational psychology. A reason for this recognition is that the educational research had changed progressively from laboratory settings to real-life classrooms. A growing number of researchers have adopted a person-in-context perspective and a multimethod approach to investigate students' situational adaptations in a classroom learning context (Järvelä *et al.*, 2001). In this perspective, motivation is constructed in the interplay between an individual and a context. Usually, this type of research involves the application of qualitative methods, combining qualitative and quantitative approaches, gathering of process-oriented

data, and the use of various units of analysis (Butler, 2002). Recently, new tools have been developed that collect detailed information about students' studying actions by logging the time and context of every learning event (Winne *et al.*, 2006). These traces recorded in a computer program are artifacts of tactics and strategies in a log of fine-grained, temporally identified data that can advance research about how learners engage in learning.

Conclusions

It is known from the research tradition of many centuries that certain motivational processes, such as goals, emotions, and strategies, are the core elements of the learning process. Technology does not question the relevance of these constructs, but calls for us to focus on different perspectives of learning with these theoretical elements. In the future, we should concentrate on finding the critical elements of context that interact with a person – not the technology itself or other contextual features of the environment which, anyway, continuously change.

See also: Cognition and Emotion; Computer-Supported Collaborative Learning: Basic Concepts, Multiple Perspectives, and Emerging Trends; Emotion in Educational Contexts; Learning Strategies; Motivation Regulation; Situated View of Learning.

Bibliography

- Ames, C. (1992). Classrooms: Goals, structures, and motivation. *Journal of Educational Psychology* **84**, 261–271.
- Anderman, L. H. and Anderman, E. M. (2000). Considering contexts in educational psychology: Introduction to the special issue. *Educational Psychologist* **35**, 67–68.
- Anderson, R., Greeno, J. G., Reder, L. M., and Simon, H. A. (2000). Perspectives on learning, thinking, and activity. *Educational Researcher* **29**(4), 11–13.
- Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. New York: Freeman.
- Barab, S. A., Kling, R., and Gray, J. H. (eds.) (2004). *Designing for Virtual Communities in the Service of Learning*. Cambridge, MA: Cambridge University Press.
- Bliss, J., Sälljö, R., and Light, P. (eds.) (1999). *Learning Sites: Social and Technological Resources for Learning*. Oxford: Pergamon.
- Blumenfeld, P. C., Soloway, E., Marx, R. W., *et al.* (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist* **26**(3–4), 369–398.
- Boekaerts, M. (1999). Motivated learning: The study of student – situation transactional units. *European Journal of Psychology of Education* **14**, 41–55.
- Boekaerts, M., Pintrich, P. R., and Zeidner, M. (eds.) (2000). *Handbook of Self-Regulation*. San Diego, CA: Academic Press.
- Brown, A. L. and Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In Schauble, L. and Glaser, R. (eds.) *Innovations in Learning: New Environments for Education*, pp 289–325. Mahwah, NJ: Erlbaum.

- Butler, D. L. (2002). Qualitative approaches to investigating self-regulated learning: Contributions and challenges. *Educational Psychologist* **37**, 59–63.
- Cognition and Technology Group at Vanderbilt (1992). The Jasper Series as an Example of Anchored Instruction: Theory, Program Description, and Assessment Data. *Educational Psychologist* **27**(3), 291–315.
- Cognition and Technology Group at Vanderbilt (1996). Looking at technology in context: A framework for understanding technology and education research. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 807–840. New York: MacMillan.
- Collins, A. (1986). Teaching reading and writing with personal computers. In Oransanu, J. (ed.) *A Decade of Reading Research: Implications for Practice*, pp 171–187. Hillsdale, NJ: Erlbaum.
- Crook, C. (2000). Motivation and the ecology of collaborative learning. In Joiner, R., Littleton, K., Faulkner, D., and Miell, D. (eds.) *Rethinking Collaborative Learning*, pp 161–178. London: Free Association Books.
- De Corte, E., Verschaffel, L., Entwistle, N., and Van Merriënboer, J. (eds.) (2003). *Advances in Learning and Instruction Series: Powerful Learning Environments: Unraveling Basic Components and Dimensions*. Oxford, UK: Elsevier Science.
- Dillenbourg, P., Järvelä, S., and Fischer, F. (2009). The evolution of research on computer-supported collaborative learning: From design to orchestration. In Balacheff, N., Ludvigsen, S., de Jong, T., Lazonder, A., and Barnes, S. (eds.) *Technology-Enhanced Learning. Principles and Products*, pp 3–19. Berlin: Springer.
- Dowson, M. and McInerney, D. M. (2003). What do students say about their motivational goals? Towards a complex and dynamic perspective on student motivation. *Contemporary Educational Psychology* **28**, 91–113.
- Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist* **41**(10), 1040–1048.
- Edelson, D. C., Gordin, D. N., and Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *Journal of the Learning Sciences* **8**, 391–450.
- Hakkarainen, K., Lipponen, L., and Järvelä, S. (2002). Epistemology of inquiry and computer-supported collaborative learning. In Koschmann, T., Miyake, N., and Hall, R. (eds.) *CSCL2: Carrying Forward the Conversation*, pp 129–156. Mahwah, NJ: Erlbaum.
- Hartley, K. and Bendixen, L. D. (2001). Educational research in the internet age. Examining the role of individual characteristics. *Educational Researcher* **30**(9), 22–26.
- Hickey, D. T. (1997). Motivation and contemporary socio-constructivist instructional perspectives. *Educational Psychologist* **32**, 175–193.
- Hickey, D. T., Moore, A. L., and Pellegrino, J. W. (2001). The motivational and academic consequences of elementary mathematics environments: Do constructivist innovations and reforms make a difference? *American Educational Research Journal* **38**(3), 611–652.
- Hug, B., Krajcik, J., and Marx, R. (2005). Using innovative learning technologies to promote learning and engagement in an urban science classroom. *Urban Education* **40**, 446–472.
- Järvenoja, H. and Järvelä, S. (2005). How the students explain their social, emotional and motivational experiences during their learning processes. *Learning and Instruction* **15**, 465–480.
- Järvelä, S. and Niemivirta, M. (2001). Motivation in context: Challenges and possibilities in studying the role of motivation in new pedagogical cultures. In Volet, S. and Järvelä, S. (eds.) *Motivation in Learning Contexts*, pp 105–127. Amsterdam: Elsevier.
- Järvelä, S. and Salovaara, H. (2004). The interplay of motivational goals and cognitive strategies in a new pedagogical culture – a context oriented and qualitative approach. *European Psychologist* **9**(4), 232–244.
- Järvelä, S. and Volet, S. (2004). Motivation in real-life, dynamic and interactive learning environments: Stretching constructs and methodologies. *European Psychologist* **9**(4), 193–197.
- Järvelä, S., Lehtinen, E., and Salonen, P. (2000). Socio-emotional orientation as a mediating variable in the teaching-learning interaction: Implications for instructional design. *Scandinavian Journal of Educational Research* **44**(3), 293–306.
- Järvelä, S., Salonen, P., and Lepola, J. (2001). Dynamic assessment as a key to understanding student motivation in a classroom context. In Pintrich, P. and Maehr, M. (eds.) *Advances in Research on Motivation: New Directions in Measures and Methods*, vol. 12, pp 217–240. Oxford: Elsevier.
- Koschmann, T., Hall, R., and Miyake, N. (eds.) (2002). *CSCL2: Carrying Forward the Conversation*. Mahwah, NJ: Erlbaum.
- McCaslin, M. and Hickey, D. T. (2001). Self-regulated learning and academic achievement: A Vygotskian view. In Zimmerman, B. and Schunk, D. (eds.) *Self-Regulated Learning and Academic Achievement: Theory, Research, and Practice*, 2nd edn., pp 227–252. Mahwah, NJ: Erlbaum.
- Meyer, D. K. and Turner, J. C. (2002). Discovering emotion in classroom motivation research. *Educational Psychologist* **37**(2), 107–114.
- Meyer, D. K., Turner, J. C., and Spencer, C. (1997). Challenge in a mathematics classroom: Students' motivation and strategies in project-based learning. *Elementary School Journal* **97**, 501–522.
- Mistler-Jackson, M. and Songer, N. B. (2000). Student motivation and internet technology: Are students empowered to learn science? *Journal of Research in Science Teaching* **37**, 459–479.
- Papert, S. (1980). *Mindstorms: Children, Computers and Powerful Ideas*. New York: Basic Books.
- Patrick, H. and Middleton, M. J. (2002). Turning the kaleidoscope: What we see when self-regulated learning is viewed with a qualitative lens. *Educational Psychologist* **37**, 27–39.
- Pea, R. D. and Maldonado, H. (2006). WILD for learning: Interacting through new computing devices, anytime, anywhere. In Sawyer, K. (ed.) *Cambridge University Handbook of the Learning Sciences*, ch. 25, pp 427–442. New York: Cambridge University Press.
- Pekrun, R., Goetz, T., Titz, W., and Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist* **37**(2), 91–105.
- Perry, N. (2002). Using qualitative methods to enrich understanding of self-regulated learning. Introduction to the special issue. *Educational Psychologist* **37**(1), 1–3.
- Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory and research. *Contemporary Educational Psychology* **25**, 92–104.
- Renniger, K. A. and Shumar, W. (eds.) (2002). *Building Virtual Communities. Learning and Change in Cyberspace*. New York: Cambridge University Press.
- Roschelle, J. and Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In O'Malley, C. E. (ed.) *Computer Supported Collaborative Learning*, pp 69–97. Heidelberg: Springer.
- Salomon, G. and Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research* **13**(1), 89–100.
- Salonen, P., Vauras, M., and Efklides, A. (2005). Social interaction: What can it tell us about metacognition and coregulation in learning? *European Psychologist* **10**(3), 199–208.
- Salovaara, H. and Järvelä, S. (2003). Students' knowledge building strategies in computer supported collaborative learning. *Learning Environments Research* **6**(3), 267–284.
- Schunk, D. H. (2001). Social cognitive theory and self-regulated learning. In Zimmerman, B. and Schunk, D. (eds.) *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*, 2nd edn., pp 125–152. Mahwah, NJ: LEA.
- Shell, D. F., Husman, J., Turner, J. E., et al. (2005). The impact of computer supported collaborative learning communities on high school students' knowledge building, strategic learning, and perceptions of the classroom. *Journal of Educational Computing Research* **33**(3), 327–349.
- Stahl, G. (2004). Building collaborative knowing: Elements of a social theory of CSCL. In Strijbos, J. W., Kirschner, P. A., and Martens, R. L. (eds.) *What We Know about CSCL and Implementing It in Higher Education*, pp 53–86. Amsterdam: Kluwer.
- Strijbos, J.-W., Kirschner, P., and Martens, R. L. (eds.) (2003). *What We Know About CSCL in Higher Education*. Amsterdam: Kluwer.
- Taylor, R. P. (1980). *The Computer in the School: Tutor, Tool, Tutee*. New York: Teachers College Press.
- Thompson, L. and Fine, G. (1999). Socially shared cognition, affect, and behaviour: A review and integration. *Personality and Social Psychology Review* **3**(4), 278–302.

- Turner, J. C. and Meyer, D. K. (2000). Studying and understanding the instructional contexts of classrooms: Using our past to forge our future. *Educational Psychologist* **35**, 69–85.
- Veermans, M. and Järvelä, S. (2004). Generalized learning goals and situational coping in inquiry learning. *Instructional Science* **32**(4), 269–291.
- Volet, S. and Järvelä, S. (eds.) (2001). *Motivation in Learning Contexts: Theoretical Advances and Methodological Implications*. London: Pergamon/Elsevier.
- Volet, S. E. and Wosnitza, M. (2004). Social affordances and students' engagement in cross-national online learning: An exploratory study. *Journal of Research in International Education* **3**(1), 5–29.
- Whipp, J. and Chiarelli, S. (2004). Self-regulation in a web-based course: A case study. *Educational Technology Research and Development* **52**(4), 5–22.
- Winne, P. H., Nesbit, J. C., Kumar, V., et al. (2006). Supporting self-regulated learning with gStudy software: The learning kit project. *Technology, Instruction, Cognition and Learning* **3**, 105–113.

Volitional Control of Learning

J Husman, Arizona State University, Tempe, AZ, USA
L Corno, Columbia University, Sherborn, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In studies of learning in educational settings, researchers find that students often have conflicting yet equally valued goals. For example, one student, during an introduction to life science, may want to fit in with their peers, pursue questions that interest them, and perform successfully. Even after having committed to a school task, some students find that obstacles and sidetracks lay between start and completion. Completion of tasks can be supported by conscious self-regulation; but students cannot consciously manage every roadblock, stray thought, or discomfort without dividing their attention. In addition to conscious regulation of cognition as they work, students often have implicit, habitual, or automatized processes in place to maintain their effort. Together, these different aspects of self-regulation are what allow for volitional functioning.

Educational researchers disagree about the utility of the concept of volition. Early theory in psychology equated volition with the idea of willpower and misleadingly considered volitional struggles as a sign of personal weakness. Some theorists question the need for volition as a research target because established concepts of efficacy and agency seem sufficient to explain the successes of effective learners (Zimmerman and Schunk, 2002).

Other theorists find explanations of purposive action incomplete without the concept of volition (Gollwitzer, 1999). Motivation leads to commitment but volition denotes follow-through. Volition, defined in modern theory as purposive striving, entails processes by which students implement goals during learning, sustain motivation, and strategically regulate cognition and affect (Corno, 1994). Through volitional control, students regulate emotion, sustain motivation, and manage their learning. Volition is a target for intervention when learners stop short of perseverance in the face of difficulty and distraction, avoid learning projects that require focus and persistence, or fail to develop effective academic work habits. A full explication of goal pursuits in education requires accounting for both motivation and volition.

Our task in this article is to move beyond debates about construct utility, and keep readers in touch with some basic evidence supporting the importance of volitional control in academic learning and performance. Research on volitional control of learning provides valuable insights into how students who want to learn (i.e., students who are already motivated to learn) manage the learning process.

This includes an understanding of self-referenced processes used to cope with distractions and persist in accomplishing goals. Researchers have studied volitional control at different levels:

- as a relatively stable, self-reported orientation or work style that distinguishes different types of students (see, e.g., Kanfer and Heggstad (1997) who study motivational traits; Higgins (1989) who studies self-guides; and Koole *et al.* (2005), who discuss a temperamental predisposition they call action orientation);
- as a recordable process or event influenced by context (e.g., Winne and Jaimeson-Noel, 2003; Xu and Corno, 1998) and
- as an implementation mindset or state that can be induced experimentally by asking students to state exactly where and when they will do assigned work (Oettingen *et al.*, 2004).

A relatively recent body of research, conducted cross-nationally over the past three decades, has become the basis for successful intervention programs, and resulted in useful guidelines for parents and teachers hoping to build good academic work habits in children.

This article is divided into two substantive sections. First, we illustrate how researchers have studied the psychological construct of volitional control in education at different grade levels, highlighting methodologically valid and reliable procedures for capturing the construct. We describe key relationships between aspects of volitional control and educational accomplishments. In the next section, we provide brief descriptions of educational interventions designed to improve learners' volitional control. These interventions have produced positive outcomes at primary, secondary, and postsecondary levels of education.

Understanding Volitional Control in Education

Emotion control, effort management, and good academic work habits are manifestations of volitional control that contribute to successful performance in educational settings (Boekaerts and Corno, 2005). Evidence from different disciplinary and theoretical frames indicates that even young students can strategically manage their own learning. Recent research focuses on emotional regulation as a key aspect of volitional control, and demonstrates its value to

emotional wellbeing. Emotions can also serve as support functions in learning (e.g., they can signal the need to alter a course of action) but they must be channeled productively (Turner and Schallert, 2001).

Effortful Control

The social psychologist, Walter Mischel devised the classic protocol for studying impulse control in young children (see Mischel and Mischel, 1983). In a series of experiments, children as young as 4 had to resist the temptation to touch a desirable object while they sat alone in a room awaiting the return of a researcher. If they successfully avoided touching the object for a specified period of time, the object would be theirs. Observations of the children as they waited showed that those who delayed gratification sang songs and talked to themselves, looked at something else in the room, or played finger games to refocus their attention and help to fill the time. The authors' interpretation was that these children focused their mental energy and controlled their impulses and behavior in order to reach a goal. They were motivated by the goal, but they called forth volitional resources to accomplish it.

The relevance of this research to education was demonstrated some years later when Mischel and his colleagues reestablished contact with families of many of these same children (Shoda *et al.*, 1990). Using correlational procedures to relate early delay time with meaningful outcomes of education, the authors found that students who had been able to delay gratification and self-regulate at age 4 had significantly higher academic achievement and positive parental assessments of adjustment than students who had demonstrated weaker impulse control as children. Although correlation does not imply causation and the sample size was small, this was the first study to make such a longitudinal connection. Mischel and his colleagues established the value of focused attention in children's impulse control. Their analyses stressed the underlying importance of self-instruction and self-motivation strategies used by children as mediators for goal accomplishment, and provided a link between these variables and educational growth.

Mischel's protocols have been used subsequently in other examinations of delay in children (e.g., Kochanska *et al.*, 2000, 2001; Li-Grining, 2007). Strategies associated with delaying gratification are now considered part of a class of self-regulatory mechanisms broadly referred to as effortful control (Rothbart and Ahadi, 1994). Eisenberg *et al.* (2005) have extended research on impulse and effortful control in children.

These authors studied the development of effortful control throughout infancy and childhood in a controlled setting with a large sample size over more than 20 years. They found that individual differences in effortful control

are established in the first years of life, and that environment and parental support is crucial to the development of self-regulation skills (Zhou *et al.*, 2007). This combination of work from the subdisciplines of social and developmental psychology underscores the need for researchers and educators to consider both dispositional and situational factors in addressing students' self-regulation and effortful control.

Developmental researchers who have focused on the examination of effortful control have frequently conducted their research in controlled laboratory settings or through examining parent-child dyads. Students use volitional strategies to support their effortful control in a variety of complex situations: when learning material is not very interesting but students desire to do well; when others in the classroom are distracting; when the relationship between outcome and performance seems unclear; or when there is a need to prioritize a work goal relative to a competing social goal (Corno, 2004). Volitional resources include, but are not limited to, strategic self-instruction to tone down negative affect (Kuhl, 2000), vigilance or self-monitoring (Posner and Rothbart, 1992), structuring and organizing the environment to maximize focus and minimize intrusion (Xu and Corno, 1994), and seeking aid from peers and teachers to bolster persistence (Newman, 1994).

Emotional Regulation

Distractions and roadblocks to learning not only occur externally, but also internally. To be successful at self-regulation, students need to manage their emotional states, as well as their cognitive and motivational environments when they complete academic tasks. Although students experience a wide range of emotions in educational contexts, performance anxiety is a particularly prominent internal obstacle (Stober and Pekrun, 2004). Academic work provokes anxiety in many students; this has been demonstrated through physiological measures (Spangler *et al.*, 2002), surveys (Pekrun *et al.*, 2004), experimental manipulations (Meinhardt and Pekrun, 2003), case studies (Pekrun *et al.*, 2002), and analyses involving multidimensional scaling (Schutz *et al.*, 2002). Students' ability to control anxiety, keeping it at an optimal level, is critical if they are to accomplish their academic goals. This aspect of volitional functioning is supported by affect-regulation strategies.

Affect may be regulated deliberately or automatically as a function of temperament or predisposition (Koole and Jostmann, 2004). Deliberate regulation strategies include slowing one's breathing, reappraising the importance of the task, reminding oneself of personal agency (Schutz *et al.*, 2004), and clearly framing a task completion goal (Boekaerts, 2002). Individuals' temperament or predisposition determines the degree to which they instinctively regulate surprising or unpleasant stimuli by approaching or avoiding (Langens and Morth, 2003).

To be successful at managing emotions, students who are predisposed toward worry and emotionality will benefit from emotional scaffolding. Teachers can provide support and help anxious students to manage appropriate emotional responses to educational challenges. For example, Rosiek (2003) documented how teachers could profitably engage students' curiosity, interest, and even anger as a way of helping them connect to important content. When teachers engage students' emotions, however, it is critical that they provide the support to channel the emotion toward engagement; scaffolding is about engaging some emotion but not too much.

Managing the Tasks of Learning

Managing the tasks of learning similarly requires both deliberate and automatic sources of regulation. Some aspects of school learning tasks are controlled by teachers and curriculum design – sequencing, for example, and pacing. However, students must manage other aspects of tasks deliberately – setting completion goals, budgeting time, and seeking feedback. At a more habitual level, effective task management by students includes localized self-monitoring and selectively attending to pertinent details – error checking, self-instructing, and avoiding distraction.

An example of automatic or intuitive regulation that benefits a task is provided in a study by Snow and Lohman (1984). These authors examined component processes used by test takers when completing standardized tests. From a variety of indicators, they demonstrated that automatically applied organization and control processes were a critical aspect of self-regulation for scholastic achievement. In addition, this research showed that automaticity came about through exercise and appeared to be transferable to situations with similar affordances and constraints (Stanford Aptitude Seminar, 2002).

Implicit control processes, which students are often unaware of, can be made explicit if they are reframed as strategies for attacking tasks or problems. Corno and Kanfer (1993) asked high school and college students to think deeply about and explain how they manage different forms of academic work (homework, problem-solving tasks, and facing difficulty – situations that require effortful processing). Covert strategies (internal to the student) included meta-cognitive control (e.g., make a plan to follow and jump in), emotional control, (e.g., imagine being good at this and how that feels), and motivation control (e.g., give myself instructions about timelines). Overt strategies included control the task situation (e.g., move away from distractions) and control others in the task setting (e.g., tell the teacher if he/she is not being clear; see also McCann and Turner (2004) who developed an academic volitional strategies inventory).

Conscientious students with a good repertoire of both covert and overt strategies will use them adaptively, as appropriate for different tasks and circumstances. As strategy use can require effort, however, consistent conscious strategy use is a challenge for many students, even when they receive direct strategy instruction (Butler and Cartier, 2004; Dewitte and Lens, 1999; Pressley *et al.*, 1990).

Academic Work Habits

Recently, researchers have combined measures of volition obtained from different sources of data as a means to increase reliability and predictive validity. For example, Duckworth and Seligman (2005) defined self-discipline as an indicator of good work habits in high school students. To operationalize the construct, they obtained two independent samples of students from the same magnet school, and combined subjective ratings by students of how they worked (self-reports), objective observations from teachers and parents who rated students on the same qualities, and a performance-based measure (a delay-of-gratification task) modeled after Mischel and Mischel (1983) but designed for older students.

Regression analyses conducted using final grade point average (GPA) as an outcome measure controlled for two related variables: general ability (the Otis–Lennon School Ability Test) and prior GPA. Results showed that the self-discipline composite accounted for over twice the variance attributable to the general ability measure, and 16–20% of the variance beyond that accounted for by prior GPA. Further study showed that the highly self-disciplined adolescents outperformed their peers on achievement tests, high school attendance, and even admission to a selective high school (p. 939).

Although studies such as this give evidence to support the value of good work habits, less well studied are questions about how good work habits develop, and how ineffective habits can be changed (Corno, 2007). As noted above, some research has shown that students' automatic responses to learning tasks can be brought under greater conscious control through training (Bryan and Burstein, 2004; Maes and Karoly, 2005). Other research has suggested that students with poor work habits need more support in their homes and schools; for example, poor work habits are more common in students who attend impoverished schools (Battistich *et al.*, 1995).

Socio-demographic challenges affect the development of children's cognitive organization and control processes but appear unrelated to indicators of impulsivity, which can be temperamental (Li-Grining, 2007). These findings imply that explicit instruction and scaffolding in self-regulation may be especially important for temperamentally impulsive children who come from high-stress backgrounds. One hypothesis is that these children may have fewer effective

models of volitional functioning in their home environments (i.e., an access problem); alternatively, the stress these children experience may interfere with their ability to engage the cognitive resources needed to attend sufficiently to the volitional models they observe (Perry, 1998). Both hypotheses could also be correct; their delays could be compounded; in any case, explanations such as these deserve close investigation.

Selected Intervention Programs in Self-Regulation

Over the past 20 years, educational psychologists have developed a number of intervention programs designed to help students become more effective and efficient learners (Kauffman, 2004). Many of these programs have been studied empirically, but only a subset focused on helping students to function volitionally. Each program that we subsequently discuss included an evaluation component, and targeted at least one aspect of volitional functioning described above.

Elementary School Settings

Harris *et al.* (2005) developed a self-monitoring intervention to support on-task behavior in students diagnosed with attention-deficit/hyperactivity disorder. Students were instructed on two systems: one to help them self-monitor attention and one to self-monitor performance. To monitor attention, students were instructed to note their attention (checking “yes” or “no” on a chart provided for them) after hearing a tone played at the end of a random interval from 10 to 90 s in length. At the end of each study period, the child would mark on a graph the number of times they answered “yes.” To self-monitor performance, the students were asked to graph the number of times they practiced their weekly spelling words. Both interventions showed statistically significant improvements to students’ on-task and spelling study behavior. Aids and instructional support for self-monitoring exercises such as those used by Harris and her colleagues can also be found in Zimmerman *et al.* (1996).

Elementary school teachers have been the focus of other interventions designed to teach or develop self-regulation in students. Perry *et al.* (2004) developed a mentoring program to support preservice elementary school teachers integrating self-regulation strategies into their curriculum. The teachers in their study addressed all phases of self-regulated learning in their lesson plans, that is, planning, implementation, and reflection. An extensive evaluation across five cohort groups found that the teachers who designed complex tasks (e.g., multiple goals, extended activities over time, a variety of processes, and alternative ways for children to demonstrate competence)

provided significantly more opportunities for students to engage in and successfully develop self-regulated learning skills. For example, students were more likely to have choices, control over challenge, opportunities to evaluate their learning, and collaborate with peers.

Middle School Settings

The strains of adolescence put middle school students in the United States (ages 12–14) at an important period in their academic development; the expectations that students’ set at this point in their lives carry over into their future academic and career choices (Tai *et al.*, 2006). Oyserman *et al.* (2002) developed the Pathways for Youth project, a 9-week, after-school program for urban middle school students. The program was designed to support students’ self-regulation and academic performance. Adult guides worked with students to complete seven steps:

1. envisioning possible futures for themselves,
2. conceptualizing those futures as goals,
3. constructing a path for goal obtainment,
4. making explicit connections between present educational activities and the valued future goals,
5. discussing possible roadblocks and forks in the path,
6. brainstorming strategies for managing imagined future obstacles, and
7. interviewing successful adults from the community about their own strategies for reaching goals.

This process supported students through the full self-regulatory cycle from goal setting or motivation to volition (see Corno, 1995). A controlled intervention showed that relative to students in a comparable group, students who participated in the program “. . . reported more bonding to school, concern about doing well in school, ‘balanced’ possible selves, plausible strategies to attain these possible selves, better school attendance, and for boys, less trouble at school” (Oyserman *et al.*, 2002: 313).

Randi (2004) provided opportunities for preservice teachers to generate ways to develop students’ volition as they transitioned from college student to teacher in their teacher-education program. Most of Randi’s teachers were planning to teach in grades 5–8. Their 13-week course focused on motivation and volition theory and classroom applications of self-regulated learning. Working in small groups, students read the research literature, analyzed teaching cases, and wrote about self-regulation strategies they themselves used to follow through on their commitments to become teachers. Teachers also wrote weekly journal entries reflecting on strategies for self-regulation they would use in their own classrooms, and designed curriculum activities and lessons for their students. The coursework led to many examples of units, incorporating self-regulated learning that these teachers could then take into their own middle school classrooms.

Secondary School Settings

The Interactive Learning Group System (ILGS) innovation program (Boekaerts, 1997; Boekaerts and Minnaert, 2003) is an example of a school-wide self-regulation intervention. The project targeted vocational secondary schools in the Netherlands by providing students with ill-structured problems in a highly structured environment. Instructors were trained in self-regulation-based instructional principles. Examples of these principles were “prepare group assignments at home and write them on the blackboard as soon as you enter the classroom. . . so that teachers come to class with an explicit plan and that students are aware of the amount of work to be done”; and “prepare the students for group assignments by providing prior knowledge.” The teacher “modeled the learning processes, making them more transparent to students” (Rozendaal *et al.*, 2005: 144). Another aspect of this program required students to work in interactive, heterogeneous groups on ill-structured problems. Although the intervention was not fully implemented by teachers as planned, the program found some success in providing teachers with insights into students’ self-regulation and changing psychological needs.

Randi (Randi and Corno, 2000) developed and tested a program for teaching high school humanities students about self-regulated learning using a literary quest theme. Students gained an intellectual understanding of the concept by reading about self-regulating characters in the quest literature, and then applied their knowledge by writing about such strategies in their own lives. Students exposed to this curriculum were able to explain the strategies of self-regulation through their writing, and accurately discussed the utility of volitional control in their own lives.

Undergraduate Settings

Several undergraduate programs to support students’ study habits and strategic learning have appeared over the past 25 years. Most programs provide instruction in cognitive learning strategies, and also help students achieve the goals they set for themselves in their pursuit of higher education. In a 5-year, longitudinal study at the University of Texas, Austin, Weinstein *et al.* (1998, 2000) evaluated a learning and study strategies course. This three-credit, nonmajor, educational psychology course was not only designed to focus on cognitive learning strategies, but also provided students with instruction on explicitly using affect and volitional strategies. The evaluation found that 71% of students who successfully completed the course in study strategies had graduated after 5 years, compared to 55% for the general university population, despite the former having entered with lower Scholastic Aptitude Test (SAT) scores. The effectiveness of acquiring volitional strategies could not be separated from the learning of other strategies in this evaluation, however.

Conclusion and Recommendations for Future Research

Volition, defined as purposeful striving, is a useful and necessary concept that complements motivation. Most research on volitional control in academic settings is based on students’ reports of their thinking and behavior that are quantified on ordinal scales and used to predict performance (McCann and Garcia, 1999; Pintrich and Garcia, 1991). However, some investigators have obtained observation data, used interviews, conducted case studies, and designed experimental tasks in order to elicit evidence of volitional control (see Corno, 2000).

Key components of volition include effortful control and emotional regulation. Effortful control of impulses first appears in toddlers, and is associated with positive social, emotional, and academic outcomes later in life. For students who have difficulty with emotional regulation, learning to use particular strategies can be productively channeled, particularly if they are constrained by predispositions such as stress or performance anxiety. Education practitioners can support students’ emotional regulation through instruction in volitional strategies and emotional scaffolding.

Managing learning tasks similarly calls upon both implicit and explicit processes of volition. Explicit processes include organization and control strategies that can be fostered through instruction and modeling. Implicit processes, such as avoiding distraction through focused attention to pertinent details in tasks, are less easily learned through instruction. Both types of processes underlie good work habits, however. Good academic work habits, those products of volitional functioning that come about from repetition of volitional strategies within supportive environments, are important to success in school. Teachers reinforce good work habits all along the age range, even into and through higher education (Corno, 2007).

Intervention programs for adolescents have been more prevalent than those for children and toddlers. Existing intervention programs focus primarily on supporting explicit self-regulation, study strategy use, and motivation through goal setting. A few programs support students’ affective regulation through collaborative learning and mentorship. Perhaps the best examples of programs that support students through all phases of self-regulation are those conducted by Oyserman, and colleagues.

This article is not the avenue for an elaborate agenda on future research on volitional control of learning. However, we can say from our review that there are few direct hypotheses about the aspects of volitional control that are most likely to promote academic success in students. Researchers have not assessed the added value of volitional strategies over and above the learning of cognitive skills. In addition, we argue that little is known about the optimal conditions for volitional control. Although

educational researchers know something about how to develop volitional capabilities in students, there is no evidence about why students from high-stress environments demonstrate volitional delays. We need to know which techniques for teaching volitional control transfer across contexts, and what obstacles may interfere with the adaptive use of volitional strategies in students who have learned them. As discussed previously, questions remain about how good work habits develop, and how ineffective habits can be changed. Finally, we argue that there is a need for more experiments designed to manipulate the covert and overt aspects of volitional control used by students in a variety of educational settings, classrooms, homework completed outside of school, and studying in online environments.

Bibliography

- Battistich, V., Solomon, D., Kim, D., Watson, M., and Schaps, E. (1995). Schools as communities, poverty levels of student populations, and students' attitudes, motives, and performance: A multilevel analysis. *American Educational Research Journal* **32**, 627–658.
- Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction* **7**(2), 161–186.
- Boekaerts, M. (2002). Intensity of emotions, emotional regulation, and goal framing: How are they related to adolescents' choice of coping strategies? *Anxiety, Stress and Coping: An International Journal* **15**(4), 401–412.
- Boekaerts, M. and Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology: An International Review* **54**, 99–232.
- Boekaerts, M. and Minnaert, A. (2003). Measuring behavioral change processes during an ongoing innovation program: Scope and limits. In DeCorte, E. (ed.) *Powerful Learning Environments: Unraveling Basic Components and Dimensions*, pp 71–87. Amsterdam: Pergamon.
- Bryan, T. and Burstein, K. (2004). Improving homework completion and academic performance: Lessons from special education. *Theory into Practice* **43**, 213–219.
- Butler, D. and Cartier, S. (2004). Promoting effective task interpretation as an important work habit: A key to successful teaching and learning. *Teachers College Record* **106**, 1729–1758.
- Corno, L. (1986). The metacognitive control components of self-regulated learning. *Contemporary Educational Psychology* **11**, 333–346.
- Corno, L. (1994). Student volition and education: Outcomes, influences, and practices. In Zimmerman, B. and Schunk, D. (eds.) *Self-Regulated Learning and Academic Achievement: Educational Applications*, pp 229–254. New York: Springer.
- Corno, L. (1995). Working toward foresight and follow-through. *Midwest Educational Researcher* **8**(1), 2–10.
- Corno, L. (ed.) (2000). Conceptions of volition: Investigating theoretical questions; Studies of practice. (Special double issue). *International Journal of Educational Research* **33**(7), 8.
- Corno, L. (2004). Work habits and work styles: Volition in education. *Teachers College Record* **106**(9), 1669–1694.
- Corno, L. (2007). Work habits and self-regulated learning: Helping students to find a “will” from a “way.” In Schunk, D. H. and Zimmerman, B. J. (eds.) *Motivation and Self-Regulated Learning: Theory, Research, and Applications*, pp 197–222. New York: Erlbaum/Taylor and Francis.
- Corno, L. and Kanfer, R. (1993). The role of volition in learning and performance. In Darling-Hammond, L. (ed.) *Review of Research in Education*, vol. 19, pp 301–341. Washington, DC: American Educational Research Association.
- Dewitte, S. and Lens, W. (1999). Volition: Use with measure. *Learning and Individual Differences* **11**(3), 321–333.
- Duckworth, A. L. and Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science* **16**, 939–944.
- Eisenberg, N., Sadovsky, A., Spinrad, T. L., et al. (2005). The relations of problem behavior status to children's negative emotionality, effortful control, and impulsivity: Concurrent relations and prediction of change. *Developmental Psychology* **41**(1), 193–211.
- Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist* **54**, 493–503.
- Harris, K. R., Friedlander, B. D., Saddler, B., Frizzelle, R., and Graham, S. (2005). Self-monitoring of attention versus self-monitoring of academic performance: Effects among students with ADHD in the general education classroom. *Journal of Special Education* **39**(3), 145–156.
- Higgins, E. T. (1989). Continuities and discontinuities in self-regulatory and self-evaluative processes: A developmental theory relating self and affect. *Journal of Personality* **57**, 407–444.
- Kanfer, R. and Heggestad, E. (1997). Motivational traits and skills: A person-centered approach to work motivation. In Cummings, L. L. and Staw, B. M. (eds.) *Research in organizational behavior*, vol. 19, pp 1–57. Greenwich, CT: JAI Press.
- Kauffman, D. F. (2004). Self-regulated learning in web-based environments: Instructional tools designed to facilitate cognitive strategy use, metacognitive processing, and motivational beliefs. *Journal of Educational Computing Research* **30**(1–2), 139–161.
- Kochanska, G., Coy, K. C., and Murray, K. T. (2001). The development of self-regulation in the first four years of life. *Child Development* **72**, 1091–1111.
- Kochanska, G., Murray, K. T., and Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology* **36**, 220–232.
- Koole, S. L. and Jostmann, N. B. (2004). Getting a grip on your feelings: Effects of action orientation and external demands on intuitive affect regulation. *Journal of Personality and Social Psychology* **87**, 974–990.
- Koole, S. L., Kuhl, J., Jostmann, N. B., and Vohs, K. D. (2005). On the hidden benefits of state orientation: Can people prosper without efficient affect-regulation skills? In Tesser, A., Wood, J. V., and Stapel, D. A. (eds.) *On Building, Defending and Regulating the Self: A Psychological Perspective*, pp 217–243. New York: Psychology Press.
- Kuhl, J. (2000). The volitional basis of personality systems interaction theory: Applications in learning and treatment contexts. *International Journal of Educational Research* **33**, 665–704.
- Langens, T. A. and Morth, S. (2003). Repressive coping and the use of passive and active coping strategies. *Personality and Individual Differences* **35**, 461–473.
- Li-Grining, C. P. (2007). Effortful control among low-income preschoolers in three cities: Stability, change, and individual differences. *Developmental Psychology* **43**, 208–221.
- Maes, S. and Karoly, P. (2005). Self-regulation assessment and intervention in physical health and illness: A review. *Applied Psychology: An International Review* **54**, 267–299.
- McCann, E. J. and Turner, J. E. (2004). Increasing student learning through volitional control. *Teachers College Record* **106**, 1695–1714.
- McCann, E. J. and Garcia, T. (1999). Maintaining motivation and regulating emotion: Measuring individual differences in academic volitional strategies. *Learning and Individual Differences* **11**(3), 259–279.
- Meinhardt, J. R. and Pekrun, R. (2003). Attentional resource allocation to emotional events: An ERP study. *Cognition and Emotion* **17**, 477–500.
- Mischel, H. N. and Mischel, W. (1983). The development of children's knowledge of self-control strategies. *Child Development* **54**, 603–619.
- Newman, R. S. (1994). Adaptive help seeking: A strategy of self-regulated learning. In Schunk, D. and Zimmerman, B. (eds.) *Self-Regulation of Learning and Performance: Issues and Educational Applications*, pp 283–301. Hillsdale, NJ: Erlbaum.

- Oettingen, G., Honig, G., and Gollwitzer, P. M. (2000). Effective self-regulation of goal attainment. *International Journal of Educational Research* **33**, 705–732.
- Oyserman, D., Terry, K., and Bybee, D. (2002). A possible selves intervention to enhance school involvement. *Journal of Adolescence* **25**(3), 313–326.
- Pekrun, R., Goetz, T., Perry, R. P., et al. (2004). Beyond test anxiety: Development and validation of the test emotions questionnaire (TEQ). *Anxiety, Stress and Coping: An International Journal* **17**(3), 287–316.
- Pekrun, R., Goetz, T., Titz, W., and Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist* **37**(2), 91–106.
- Perry, N. (1998). Young children's self-regulated learning and contexts that support it. *Journal of Educational Psychology* **90**, 715–729.
- Perry, N., Phillips, L., and Dowler, J. (2004). Examining features of tasks and their potential to promote self-regulated learning. *Teachers College Record* **106**, 1854–1878.
- Pintrich, P. R. and Garcia, T. (1991). Student goal orientation and self-regulation in the college classroom. In Maehr, M. L. and Pintrich, P. (eds.) *Advances in Motivation and Achievement*, pp 371–402. Greenwich, CT: JAI Press.
- Posner, M. I. and Rothbart, M. K. (1992). Attentional mechanisms and conscious experience. In Milner, A. D. and Rugg, M. D. (eds.) *The Neuropsychology of Consciousness*, pp 91–111. New York: Academic Press.
- Pressley, M., Woloshyn, V., Lysynchuk, L. M., et al. (1990). A primer of research on cognitive strategy instruction: The important issues and how to address them. *Educational Psychology Review* **2**, 1–58.
- Randi, J. (2004). Teachers as self-regulated learners. *Teachers College Record* **106**, 1825–1853.
- Randi, J. and Corno, L. (2000). Teacher innovations in self-regulated learning. In Pintrich, P., Boekaerts, M., and Zeidner, M. (eds.) *Handbook of Self-Regulation*, pp 651–685. San Diego, CA: Academic Press.
- Rosiek, J. (2003). Emotional scaffolding: An exploration of the teacher knowledge at the intersection of student emotion and subject matter. *Journal of Teacher Education* **54**(5), 399–412.
- Rothbart, M. K. and Ahadi, S. A. (1994). Temperament and the development of personality. *Journal of Abnormal Psychology* **103**, 55–66.
- Rozendaal, J. S., Minnaert, A., and Boekaerts, M. (2005). The influence of teacher perceived administration of self-regulated learning on students' motivation and information-processing. *Learning and Instruction* **15**(2), 141–160.
- Schutz, P. A., Davis, H. A., and Schwanenflugel, P. J. (2002). Organization of concepts relevant to emotions and their regulation during test taking. *Journal of Experimental Education* **70**(4), 316–342.
- Schutz, P. A., Distefano, C., Benson, J., and Davis, H. A. (2004). The emotional regulation during test-taking scale. *Anxiety, Stress and Coping: An International Journal* **17**(3), 253–269.
- Shoda, Y., Mischel, W., and Peake, P. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology* **26**, 978–986.
- Snow, R. E. and Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology* **76**, 347–376.
- Spangler, G., Pekrun, R., Kramer, K., and Hofmann, H. (2002). Students' emotions, physiological reactions, and coping in academic exams. *Anxiety, Stress and Coping: An International Journal* **15**(4), 413–432.
- Stanford Aptitude Seminar—Corno, L., Cronbach, L. J., Kupermintz, H. K., et al. (2002). *Remaking the Concept of Aptitude: Extending the Legacy of Richard E. Snow*. Mahwah, NJ: Erlbaum.
- Stober, J. and Pekrun, R. (2004). Advances in test anxiety research. *Anxiety, Stress and Coping: An International Journal* **17**(3), 205–211.
- Tai, R. H., Qi Liu, C., Maltese, A. V., and Fan, X. (2006). Career choice: Planning early for careers in science. *Science* **5777**(312), 1143–1144.
- Turner, J. E. and Schallert, D. L. (2001). Expectancy-value relationships of shame reactions and shame resiliency. *Journal of Educational Psychology* **93**(2), 320–329.
- Weinstein, C. E., Hanson, G. R., Powdrill, L., et al. (1998). The design and evaluation of a course in strategic learning. In Higo, I. J. and Dwindle, P. (eds.) *Developmental Education: Meeting Diverse Student Needs*. Chicago, IL: National Association of Developmental Education.
- Weinstein, C. E., Husman, J., and Dierking, D. R. (2000). Interventions with a focus on learning strategies. In Boekaerts, M., Pintrich, P. R., and Zeidner, M. (eds.) *Handbook of Self-Regulation*, pp 727–747. San Diego, CA: Academic Press.
- Wigfield, A. and Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In Wigfield, A. and Eccles, J. S. (eds.) *Development of Achievement Motivation*, pp 91–120. San Diego, CA: Academic Press.
- Winne, P. H. (2004). Putting volition to work in education. *Teachers College Record*, **106**, 1879–1887.
- Winne, P. H. and Jaimeson-Noel, D. L. (2003). Self-regulating studying by objectives for learning: Students' reports compared to a model. *Contemporary Educational Psychology* **28**, 259–276.
- Xu, J. and Corno, L. (1994). Case studies of families doing third grade homework. *Teachers College Record* **100**, 402–436.
- Zhou, Q., Hofer, C., Eisenberg, N., et al. (2007). The developmental trajectories of attention focusing, attentional and behavioral persistence, and externalizing problems during school-age years. *Developmental Psychology* **43**(2), 369–385.
- Zimmerman, B. J., Bonner, S., and Kovach, R. (1996). *Developing Self-Regulated Learners: Beyond Achievement to Self-Efficacy*. Washington, DC: American Psychological Association.
- Zimmerman, B. J. and Schunk, D. H. (2002). Reflections on theories of self-regulated learning and academic achievement. In Zimmerman, B. J. and Schunk, D. H. (eds.) *Self-Regulated Learning and Academic Achievement: Theoretical Perspectives*, 2nd edn., pp 209–308. Mahwah, NJ: Erlbaum.

Further Reading

- Boekaerts, M., Pintrich, P. R., and Zeidner, M. (eds.) (2000). *Handbook of Self-Regulation*. San Diego, CA: Academic Press.
- Kuhl, J., Koole, S. L., Greenberg, J., Koole, S. L., and Pyszczynski, T. (2004). Workings of the will: A functional approach. In Greenberg, J., Pyszczynski, T., and Koole, S. L. *Handbook of Experimental Existential Psychology*, pp 411–430. Guilford Press, New York.
- Mischel, W., Ayduk, O., Baumeister, R. F., and Vohs, K. D. (2004). Willpower in a cognitive-affective processing system: The dynamics of delay of gratification. In Baumeister, R. F. and Vohs, K. D. (eds.) *Handbook of Self-Regulation: Research, Theory, and Applications*, pp 99–129. Guilford Press, New York.
- Oyserman, D., Bybee, D., Terry, K., and Hart-Johnson, T. (2004). Possible selves as roadmaps. *Journal of Research in Personality* **38**(2), 130–149.
- Weinstein, C. E. and Hume, L. M. (1998). *Study Strategies for Lifelong Learning*. Washington, DC: American Psychological Association.
- Zimmerman, B. J., Bonner, S., and Kovach, R. (1996). *Developing Self-Regulated Learners: Beyond Achievement to Self-Efficacy*. Washington, DC: American Psychological Association.
- Zimmerman, B. and Schunk, D. (2007). *Motivation and Self-Regulated Learning: Theory, Research, and Applications*. Boca Raton, FL: Taylor and Francis.

Relevant Websites

- <http://psychclassics.yorku.ca> – Classics in the History of Psychology: What is an emotion?
- <http://www.ericdigests.org> – ERICDigest.org: Self-Regulation through Goal Setting.

<http://www.ncrel.org> – Northcentral Regional Education Laboratory:
21st Century Skills: Self Direction.
<http://www.nwrel.org> – Northwest Regional Educational Laboratory:
Developing Self Directed Learners.
<http://content.scholastic.com>: Scholastic: Developing Self-Regulation.
<http://www.psych.rochester.edu> – Self Determination Theory Self
Regulation (Theory and Questionnaires).

<http://plato.stanford.edu> – Stanford Encyclopedia of Philosophy:
Personal Autonomy.
<http://www.sitemaker.umich.edu>– UM.SiteMaker: Culture and the Self
Research Laboratory.
<http://www.yale.edu> – Yale University: Yale–New Haven Teachers
Institute Unit by Elisabeth Johnson.

Subjective Well-Being

Since it came into discussion in the early 1970s, the idea of subjective well-being (SWB) has been strongly promoted during the last 20 years as an individual and social value as well as a basic psychological need (Diener and Lucas, 2000; Strack *et al.*, 1991). As such, SWB is an indicator of a quality of life and is important for human health and development. The basic aims of well-being research are to identify harmful influences and to support those which foster good health, and to prevent illness. Nowadays, research on SWB is central to positive psychology aiming to build a good quality of life (Linley *et al.*, 2006).

Approaches to SWB

Different research traditions on well-being resulted in different definitions of SWB (e.g., Diener, 1984; Larsen and Diener, 1987; Veenhoven, 1991). There are three main lines of conceptualization of SWB, which can be classified as follows:

- SWB is a specific emotional quality of feeling well.
- SWB is a supra-term for positive emotions like enjoyment, pride, satisfaction, etc.
- SWB is a multidimensional concept combining cognitive and emotional factors.

Despite their differences, all three definitions agree on enjoyment or happiness as a core element of SWB.

As a psychological term, SWB usually refers to a subject's own feelings and evaluations. For example, if a person is happy or proud and evaluates his/her actual situation positively, he/she is expected to feel well; if a person feels sad or angry and evaluates his/her situation negatively, he/she will have a low level of well-being. Feelings and evaluations depend on a person's previous experiences, expectations, goals, and values, that is, his/her reasons for positive or negative emotions and evaluations may differ. SWB also consists of both mental and physical components, that is, a person's thoughts as well as body sensations have to be considered. Defined in this way, SWB differs significantly from single evaluations like satisfaction, or from singular emotions like enjoyment. According to the WHO definition of health, SWB can be classified into psychological, physical, or social forms of SWB (e.g., WHO, 2000).

Because of the coexistence of positive and negative events in life, emotions and evaluations can be negative as well as positive. Consequently, SWB can be defined by the interaction of positive and negative aspects: The more positively life in general or specific relevant domains of life are experienced, the better the well-being; negative feelings and negative evaluations of important circumstances or events in one's life result in a low level of well-being. Thus, to understand and explain a person's SWB, positive and negative dimensions have to be taken into account simultaneously. A high level of SWB occurs when there is a dominance of positive emotions and evaluations over negative emotions and evaluations. The greater the difference between the positive and negative dimensions, the higher the SWB.

SWB can develop in a short- or long-term perspective (Becker, 1994; Kim-Prieto *et al.*, 2005). Short-term SWB is the expression of current positive emotions and cognitions toward an actual situation, for example, working together with a classmate on a problem during the mathematics class. Thus, it is situationally determined. Long-term, habitual SWB is sustainable and indicates a more general evaluation of life situation and circumstances, for example, feeling well in school on the whole. Habitual SWB develops through frequent experiences of short-term well-being, that is, repeated experiences of enjoyment, mastering of academic challenges, feelings of competence in school, and a low amount of negative occurrences will lead to well-being in school. However, habitual SWB is not a simple or accurate aggregation of actual SWB. Specific experiences of actual SWB can contribute differently to habitual SWB, in that actual SWB in situations of special importance can have a stronger impact on well-being than situations of lower relevance.

In order to avoid theoretical confusion and to enable empirical precision, the following differentiation can be suggested:

- There are general characteristics of SWB such as its duration.
- Indicators of SWB are forms of expressions of SWB, for example, being relaxed.
- Components of SWB determine the elements of SWB such as enjoyment or the absence of worries.
- There are accompanying phenomena to SWB, like good mood.
- As long as their causal influence lacks empirical evidence, variables which show a significant relationship to

SWB (like academic degrees) should be described as correlates of SWB.

- Predictors, causes, or sources are responsible for the development of SWB as is also true for positive incidents. Nevertheless, predictors themselves can, to some degree, be influenced by SWB, for example, a person experiencing a high level of SWB is likely to perceive more positive events.
- Consequences of SWB are factors which are influenced by SWB, for example, information processing.

There are a lot of different quantitative and qualitative instruments for the assessment of SWB, ranging from affect–balance–scales to well-being questionnaires to event sampling methods and happiness interviews (Andrews and Robinson, 1991; Diener, 1984; Diener and Biswas-Diener, 2000; Lucas *et al.*, 1996). Of special importance for the accurate assessment of SWB is a multifaceted approach. Various dimensions of SWB, for example, the emotional and cognitive, the positive and the negative, have to be considered. Single items are not able to measure SWB in its full sense and multimethod assessments which explicitly integrate different perspectives on well-being are recommended.

Predictors of SWB

SWB proves to be sensitive to several factors. For this reason, it is also depicted as a multisource phenomenon. The more generally well-being is defined and measured, the more sources can influence it, from weather to situational circumstances to subjective beliefs (Diener, 1984). As a consequence, it is necessary to specify specific predictors for the different components of SWB. Although researchers so far were not successful in uncovering the most potent variables, social integration plays an essential role for SWB. More generally, there is empirical evidence about the impact of the following factors (Hascher, 2004):

1. First, SWB is influenced by personal circumstances. Situational factors such as an individual's financial status and social integration are important for a positive evaluation and contribute directly to subjective well-being. Other situational factors, such as an individual's socio-economical status, are intertwined with specific resources (e.g., social contacts) which influence the fulfillment of goals and aspirations and as a result SWB (e.g. Diener *et al.*, 1995).

2. Second, SWB can be predicted by personality factors. In addition to gender, the dependence of SWB on introversion versus extraversion (two facets of the so-called Big Five) has been confirmed (e.g., Vittersø and Nilsen, 2002). Extraverted individuals experience more positive situations in life and, thus, feel better than more

introverted persons (the so-called Bottom-Up Approach). They also interpret life more positively, react in a more positive way to life events and circumstances, and feel better than more introverted individuals (the so-called Top-Down Approach).

3. Third, subjective differences exist on how life-events are evaluated and interpreted. A person's evaluations are based on different points of references (e.g., Stein *et al.*, 1997). These are, for example, social comparisons, individual goals and expectancies, processes of adaptation to new life-circumstances, coping with challenging situations, control beliefs, or causal attributions.

Common to all three approaches is the subject–environment paradigm: there is no mono-causal relationship, but it is the interplay of the subjects (e.g., their goals, experiences, and attitudes) and the environment (e.g., living and working conditions) which determines SWB.

Well-Being in School

Principally, evaluations about SWB are relevant for a variety of contexts and different branches. For example, during the last years, psychologists paid special attention to comparisons of the SWB of nations (e.g., Biswas-Diener *et al.*, 2005). Till now, educational research concentrated mainly on cognitive aspects of school or on a few socio-emotional variables like test anxiety or school climate. As a result, studies on SWB of students or teachers are not yet common. Nevertheless, there are important findings which need to be taken into consideration for designing powerful learning environments in school. Some of them result from empirical studies which were initiated or conducted by health education research.

Student well-being

Despite a lack of research, SWB is a pedagogical aim in educational theory and practice. Although nowadays broadly accepted, it has also been a controversial topic (Boekaerts, 1993). Skeptical voices doubt the solely positive function of SWB and ask if student SWB could be harmful. They are afraid of neglecting the attainment of academic goals for the benefits of a positive atmosphere in school.

However, good reasons exist to foster SWB in school. Apart from the fact that SWB in school is a value in itself, there is empirical evidence that SWB affects students' learning processes and learning outcomes positively. Although SWB might not directly enhance student achievement, it is intertwined with various factors that contribute to effective learning and is an important criterion for a positive school climate which is necessary for students' learning – especially in highly structured, achievement-oriented, nonoptional learning contexts such as schools (e.g., Mayring and von Rhöneck, 2003).

Students who feel well in school can develop positive attitudes toward school learning and achievement. SWB can also serve as a resource for coping with negative peer influence or drug abuse. Furthermore, student SWB is a crucial indicator for children's and adolescents' health status and is a mirror of school quality.

Definition and research lines

There are various views as to how one could investigate student SWB. One important aspect is how student SWB is defined. Another differentiation stresses the question of whether SWB is an input or an output indicator of school quality. The different ways of operationalization can be described as follows:

1. Students are asked about aspects of their mood states, mostly negative states like depression (e.g., Benjet and Hernández-Guzmán, 2002; Undheim and Sund, 2005). Although SWB is not directly addressed, the analyses are subsumed under the term SWB. Sometimes direct correlations to school variables (e.g., school adjustment) are investigated. However, SWB itself is not investigated, neither in a general sense nor focused to the specific school situation.

2. Students are asked about their SWB by using a general well-being scale for children or adolescents (e.g., Jin and Moon, 2006; Konu *et al.*, 2002). These questions are directed toward life in general. Both dimensions of SWB, positive (like enjoyment) and negative (like worries), are addressed but the focus remains unspecified. Sometimes, additional instruments on school variables like satisfaction with school or school adjustment, which are related to general SWB are integrated. SWB itself is not yet directed to the specific school situation.

3. SWB in school is operationalized by a selective combination of various emotional, social, and cognitive school-related variables, for example, test anxiety, social integration into the classroom, and satisfaction in school (e.g., Eder, 1995). Nonetheless, there is no clear or explicit definition of SWB, rather an argumentation why the selected variables represent well-being.

4. Psychological well-being as a specific quality in the specific school setting is the central variable of research, described as student SWB, SWB of pupils or school SWB, or SWB in school (e.g., Engels *et al.*, 2002; Hascher, 2003, 2004, 2007). Until recently, only a few instruments existed but new ones are being developed and used and the results for student well-being, its correlates, and predictors are presented.

5. In terms of the WHO definition of health, student SWB is discussed under the broader perspective of health and health promotion (e.g., Konu and Lintonen, 2005, 2006). As school is an important context for children and adolescents, SWB in school is conceptualized as a crucial component of general physical and mental health.

So far, empirical studies of student SWB often used single items only, to ask them if they feel well in a learning situation, in a lesson, or in school in general. However, SWB in school as a psychological concept represents not a singular impression, but a more holistic quality of students' subjective experience in school with cognitive and emotional elements: student SWB in school is a quality of experience characterized by the dominance of positive feelings and cognitions toward school, persons in school, and the school context in comparison to negative feelings and cognitions toward school life. Well-being in school represents positive, emotional, and cognitive evaluations of school reality (Van Petegem *et al.*, 2006). It can be seen as an imbalance of positive and negative aspects in favor of positive aspects.

According to multidimensional concepts of general well-being, student SWB can be assumed to consist of a number of dimensions. Six dimensions prove to be crucial for student SWB (Grob *et al.*, 1996; Hascher, 2004; Ryff and Keyes, 1995):

1. positive attitudes and emotions toward school in general,
2. enjoyment in school,
3. positive academic self-concept,
4. absence of worries about school,
5. absence of physical complaints in school, and
6. absence of social problems in school.

How do students feel in school?

In asking this question one uncovers an amazing incongruity: researchers found that school is a highly aversive context and a central source of stress for some children and for many adolescents. When asked about daily experiences in school, students report a clear dominance of negative situations and emotions. Yet, most students indicate middle, some even high, score when asked about their SWB in school. One possible explanation for this discrepancy is that the results depend on the way students are asked about their school. To report about everyday events in school may lead to a focus on negative aspects because of their stronger need to be communicated. Conversely, a questionnaire on student SWB stresses negative as well as positive components, and thus, is able to take both sides of school into consideration simultaneously. Another explanation is given by the theoretical assumption that student SWB is more than simply lining up daily events. A more holistic view on school leads to different (i.e., more positive) results than a focus on singular experiences.

Independent of the method of measurement, the following interindividual differences in student SWB according to achievement, motivation, gender, and age have been empirically confirmed (e.g., Hascher, 2004, 2007; Konu and Lintonen, 2006):

- High achievers feel better in school than low achievers.
- Students with intrinsic motivation report higher SWB than those with extrinsic motivation.
- Girls report higher scores on SWB but are also likely to describe higher scores on negative components like physical complaints, in comparison to boys.
- Students attending lower classes feel better in school than students in higher grades. Toward the end of compulsory education an increase of SWB can be found.

Differences in student SWB cannot be exclusively attributed to individual characteristics. The following learning conditions in terms of school, instruction, and teaching subjects are of relevance as well (e.g., Opdenakker and van Damme, 2000; Van Petegem *et al.*, 2006):

- Schools differ in their level of student SWB indicating the relevance of school culture variables.
- Students report better SWB scores in schools with a higher academic ranking in comparison to schools with a lower academic ranking.
- Higher levels of SWB are reported in lessons with predominant student orientation in comparison to teacher-centered lessons.
- SWB differs across teaching subjects showing that highly selective subjects are more often associated with lower well-being than subjects with lower achievement pressure.

So far, the results refer to student SWB in general. A more detailed view on the various dimensions of SWB shows interesting differences (e.g., Hascher, 2004, 2007; Van Petegem *et al.*, 2007):

- Students report low degrees of physical complaints.
- Equally, the reported amount of social problems is small.
- More frequent are worries in school.
- A lack of enjoyment in school can be found.
- Students' test anxiety correlates negatively with all dimensions of SWB, even with social problems.

Predictors of student well-being

According to the predictors of general SWB, there are many sources of student SWB (see **Figure 1**). Basically,

the fit of individual goals, values, and expectancies with the school situation is crucial for the development of student SWB, that is, the better the fit, the higher the student SWB.

More systematically, predictors of SWB can be divided into three main groups:

1. Student SWB is influenced by conditions in the environment, primarily by school conditions. Not only dissimilarities of school forms but also school-specific variation in terms of action plans in school, school culture, pedagogical orientation, connections of needs and school characteristics, infrastructure, and facilities have an impact on student SWB. Situation-specific classroom factors like the quality of instruction, facilities, participation of all in the classroom, or the fulfillment of basic needs are of comparable importance. Teachers, as well as classmates, play an essential role for student SWB (Hascher, 2003; Van Petegem *et al.*, 2007): the higher the didactical and social competencies of the teachers, the higher the student SWB. In particular, the more positive their attitudes toward school, the higher their enjoyment in school and the better their academic self-concept. In comparison, the higher the achievement stress, the lower the student SWB, specifically, the higher their worries about school, the more frequent their physical complaints. Social discrimination in the classroom or rejection by peers reduces student SWB, that is, increases their social problems in school. Social support, integration into the classroom, and positive interactions during breaks foster student SWB by promoting positive attitudes toward school and students' academic self-concept, and by preventing social problems in the classroom. However, not only experiences in relationship to the school context but also living conditions, family structure, parental academic orientation, and school orientation of friends can influence student SWB (Martin, 2005).

2. Student SWB can be predicted to some degree by personality variables. In addition to gender, exertion readiness, self-efficacy and control beliefs, goal orientation or tolerance to frustration influence student

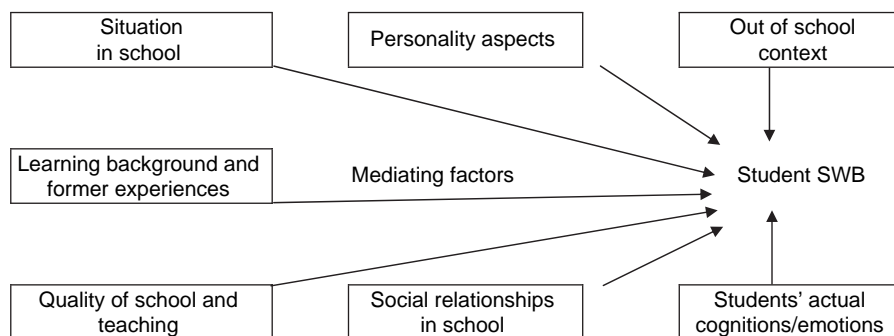


Figure 1 Predictors of student well-being. Hascher, T. (2004). Ein mehrperspektivisches Konzept der Quellen und Bedingungen schulischen Wohlbefindens. In *Wohlbefinden in der Schule (Well-Being in School)*. Münster: Waxmann.

SWB. Gender differences are explained by a gender-specific quality of experiences in that females experience SWB more intensively. There is also a gender-specific way of reporting about emotional experiences, that is, females report about their SWB in a more open way. Exertion readiness, learning orientations, self-efficacy and control beliefs, and tolerance to frustration enable students to manage academic challenges which support success in school. As a consequence, success in school fosters student SWB.

3. Although context and personality variables can influence student SWB directly, most effects result from person–environment interaction. The interaction is represented by subjective differences to how school-related conditions and events are evaluated and interpreted. For instance, different evaluations result from social comparisons, processes of adaptation to the school situation, personal interests, coping with challenging situations, causal attributions, etc. These variables affect how school is perceived, how an individual's situation is related to the context, and they are powerful for students' self-evaluations. They can have a direct impact on student SWB or serve as mediators between school conditions, personality traits, and student SWB.

Teacher well-being

SWB as an indicator of the quality of life is not only important for students but also crucial for teachers' physical functioning and mental health. Till now, however, teacher SWB has not been directly addressed – neither from well-being research nor from research on teachers and teaching. Instead, concepts like life or job satisfaction, intrinsic motivation, flow and emotions at work, or the severe problems of burnout or stress at work are primarily studied as indicators of teacher SWB. Teacher SWB cannot yet be reduced to the existence of job satisfaction or the absence of stress. In future, it has to be studied as a distinct quality.

Job satisfaction is often considered as a valid indicator of SWB at work, although it is frequently defined as a cognitive job attitude, neglecting the emotional dimension of SWB. Moreover, research on job satisfaction shows discrepant results (Bieri, 2006). The majority of teachers are satisfied with their job. However, this is only a limited mirror of their job situation because job satisfaction predicts achievement, burnout, absenteeism, fluctuation, etc., less than what is expected. The clearest differences can be found between teachers experiencing high satisfaction and those experiencing lesser satisfaction: that is, high satisfaction is positively associated with health and negatively with thoughts about job quitting. Low satisfaction goes together with stress and unsuccessful coping strategies.

The subjectively experienced level of stress is high for many teachers despite their positive job satisfaction.

Furthermore, the level of job satisfaction can vary between different areas: for example, teachers are highly satisfied with the content factors (e.g., work with children, instruction, and teaching) but dissatisfied with organizational factors (e.g., classroom size, amount of administrative tasks) and professional support. There are also significant differences in terms of school type, indicating high levels of job satisfaction for nursery and primary school teachers in comparison to secondary school teachers. It must also be noted that relevant sources of job satisfaction (e.g., interaction with students) can at the same time serve as causes for stress. As a result, a more differentiated approach to satisfaction in terms of the quality of experience and the evaluated subject is recommended. For example, different forms of satisfaction can be identified. Some of them result from a negative social comparison or a resigned view on one's own situation. Satisfaction can be experienced, if a person compares his/her situation with others who are in a much worse situation or if they accept the fact that their goals cannot be reached. Thus, individual goals and aspiration levels are crucial variables for the development of job satisfaction.

According to research on general SWB, similar predictors for teacher SWB can be assumed. They can be divided into context factors, personality, and their interplay. Regarding context factors, the workplace is the main context and predictor of SWB. The specific conditions of working in schools and their subjective perception by teachers need to be taken into account, for example, high autonomy of teachers, high responsibility for students' learning outcomes, relatively low support of colleagues, a continuous process of educational top-down reforms, high achievement pressure caused by the curriculum and by school evaluation, the challenging task of handling interruptive student behavior, and so on. Furthermore, a discrimination of intrinsic factors like contribution to children's education and development, and extrinsic factors like salary and social status is helpful.

As a consequence, resources at work are of specific importance (Salanova *et al.*, 2006). Two forms of resources and their interaction can be differentiated: personal resources (e.g., self-efficacy beliefs) and organizational resources (e.g., autonomy and social support at work). Research on personal resources confirms that high levels of self-efficacy beliefs can foster job satisfaction and SWB at work. Similarly, organizational resources have an impact on job satisfaction. Yet, there is no mono-causal relationship; reversed causal relationships are also found, that is, job satisfaction has an influence on future personal and organizational resources.

See also: Affect, Mood and Emotions; Cognition and Emotion; Flexible Learning in Higher Education; School Development for Teacher Learning and Change; School Inspection/External School Evaluation; Teacher Incentives.

Bibliography

- Andrews, F. M. and Robinson, J. P. (1991). Measures of subjective well-being. In Robinson, J. P., Shaver, P. R., and Wrightsman, L. S. (eds.) *Measures of Personality and Social Psychological Attitudes*, pp 61–114. San Diego, CA: Academic Press.
- Becker, P. (1994). Theoretische Grundlagen (Theoretical basics). In Abele, A. and Becker, P. (eds.) *Wohlbefinden (Well-Being)*, pp 13–49. Weinheim: Juventa.
- Benjet, C. and Hernández-Guzmán, L. (2002). A short-term longitudinal study of pubertal change, gender, and psychological well-being of Mexican early adolescents. *Journal of Youth and Adolescence* **31**, 429–442.
- Bieri, T. (2006). *Lehrpersonen: Hochbelastet und trotzdem zufrieden? (Teachers: High Stress Yet Satisfied?)*. Bern: Haupt.
- Biswas-Diener, R., Vittersø, J., and Diener, E. (2005). Most people are pretty happy, but there is a cultural variation: The Inughuit, the Amish, and the Massai. *Journal of Happiness Studies* **6**, 205–226.
- Boekaerts, M. (1993). Being concerned with well-being and with learning. *Educational Psychologist* **28**, 149–167.
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin* **95**, 542–575.
- Diener, E. and Biswas-Diener, R. (2000). *New Directions in Subjective Well-Being Research: The Cutting Edge*. Champaign, IL: University of Illinois, Department of Psychology.
- Diener, E., Diener, M., and Diener, C. (1995). Factors predicting the subjective well-being of nations. *Journal of Personality and Social Psychology* **69**, 851–864.
- Diener, E. and Lucas, R. E. (2000). Subjective emotional well-being. In Lewis, M. and Haviland, J. M. (eds.) *Handbook of Emotions*, pp 325–337. New York: Guilford Press.
- Eder, F. (ed.) (1995). *Das Befinden von Kindern und Jugendlichen in der Schule (Well-Being of Children and Adolescents in School)*. Innsbruck: Studienverlag.
- Engels, N., Aelterman, A., Van Petegem, K., and Schepens, A. (2002). Factors which influence the well-being of pupils in Flemish secondary schools. *Educational Studies* **30**, 127–143.
- Grob, A., Wearing, A. J., Little, T. D., and Wanner, B. EURONET (1996). Adolescents' well-being and perceived control across 14 sociocultural contexts. *Journal of Personality and Social Psychology* **71**, 785–795.
- Hascher, T. (2003). Well-being in school – why students need social support. In Mayring, P. and von Rhöneck, C. (eds.) *Learning Emotions – The Influence of Affective Factors on Classroom Learning*, pp 127–142. Bern u.a.: Lang.
- Hascher, T. (2004). Ein mehrperspektivisches Konzept der Quellen und Bedingungen schulischen Wohlbefindens. In *Wohlbefinden in der Schule (Well-Being in School)*. Münster: Waxmann.
- Hascher, T. (2007). Exploring students' well-being by taking a variety of looks into the classroom. *Hellenic Journal of Psychology* **4**, 331–349.
- Jin, S.-U. and Moon, S. M. (2006). A study of well-being and school satisfaction among academically talented students attending a science high school in Korea. *Gifted Children Quarterly* **50**, 169–184.
- Kim-Prieto, C., Diener, E., Tamir, M., Scollon, C., and Diener, M. (2005). Integrating the diverse definitions of happiness: A time-sequential framework of subjective well-being. *Journal of Happiness Studies* **6**, 261–300.
- Konu, A. and Lintonen, T. P. (2005). Theory-based survey analysis of well-being in secondary schools in Finland. *Health Promotion International Advance Access*, doi:10.1093/heapro/dai028.
- Konu, A. and Lintonen, T. P. (2006). School well-being in grades 4–12. *Health Education Research* **21**, 633–642.
- Konu, A. I., Lintonen, T. P., and Autio, V. J. (2002). Evaluation of well-being in schools – a multilevel analysis of general subjective well-being. *School Effectiveness and School Improvement* **13**, 187–200.
- Larsen, R. J. and Diener, E. (1987). Affect intensity as an individual difference characteristic: A review. *Journal of Research in Personality* **21**, 1–39.
- Linley, P. A., Joseph, S., Harrington, S., and Wood, A. M. (2006). Positive psychology: Past, present, and (possible future). *Journal of Positive Psychology* **1**, 3–16.
- Lucas, R. E., Diener, E., and Suh, E. (1996). Discriminant validity of well-being measures. *Journal of Personality and Social Psychology* **71**, 616–628.
- Martin, P. (2005). *Making People Happy. The Nature of Happiness and Its Origins in Childhood*. London, UK: Fourth Estate, Harper Collins.
- Mayring, P. and von Rhöneck, C. (eds.) (2003). *The Influence of Affective Factors on Classroom Learning*. Frankfurt am Main: Lang.
- Opdenakker, M.-C. and van Damme, J. (2000). Effects of schools, teaching staff and classes on achievement and well-being in secondary education: Similarities and differences between school outcomes. *School Effectiveness and School Improvement* **11**, 165–196.
- Ryff, C. D. and Keyes, C. L. M. (1995). The structure of subjective well-being revisite. *Journal of Personality and Social Psychology* **69**, 719–727.
- Salanova, M., Bakker, A. B., and Llorens, S. (2006). Flow at work: Evidence for an upward spiral of personal and organizational resources. *Journal of Happiness Studies* **7**, 1–22.
- Stein, N. L., Folkman, S., Trabasso, T., and Richards, T. A. (1997). Appraisal and goal processes as predictors of psychological well-being in bereaved caregivers. *Journal of Personality and Social Psychology* **72**, 872–884.
- Strack, F., Argyle, M., and Schwarz, N. (eds.) (1991). *Subjective Well-Being*. Oxford: Pergamon.
- Undheim, A. M. and Sund, A. (2005). School factors and the emergence of depressive symptoms among young Norwegian adolescents. *European Child and Adolescent Psychiatry* **14**, 446–453.
- Van Petegem, K., Aelterman, A., Rosseel, Y., and Creemers, B. (2006). Student perception as moderator for student well-being. *Social Indicators Research*. Springer, doi:10.1007/s11205-006-9055-5.
- Van Petegem, K., Aelterman, A., Van Keer, H., and Rosseel, Y. (2007). The influence of student characteristics and interpersonal teacher behaviour in the classroom on student's well-being. *Social Indicators Research*, doi: 10.1007/s11205-007-9093-7. Springer.
- Veenhoven, R. (1991). Questions on happiness: Classical topics, modern answers, blind spots. In Strack, F., Argyle, I., and Schwarz, N. (eds.) *Subjective Well-Being*, pp 7–26. Oxford: Pergamon.
- Vittersø, J. and Nilsen, F. (2002). The conceptual and relational structure of subjective well-being, neuroticism, and extraversion: Once again, neuroticism is the important predictor of happiness. *Social Indicators Research* **57**, 89–118.
- WHO (2000). *Health and Health Behaviour among Young People. A WHO-Cross-National-Survey*. Copenhagen: World Health Organization.

Further Reading

- Alexandrova, A. (2005). Subjective well-being and Kahneman's 'objective happiness' *Journal of Happiness Studies* **6**, 301–324.
- De Fraine, B., Van Landeghem, G., Van Damme, J., and Onghena, P. (2005). An analysis of well-being in secondary school with multi-level growth curve models and multilevel multivariate models. *Quality and Quantity* **39**, 297–316.
- Efklides, A. and Volet, S. (2005). Emotional experiences during learning: Multiple, situated and dynamic. *Learning and Instruction* **15**, 377–380.
- Hascher, T. (2008). Quantitative and qualitative research approaches to assess student well-being. *International Journal of Educational Research* **47**, 84–96.
- Huebner, E. S., Valois, R. F., Paxton, R. J., and Drane, J. W. (2005). Middle school students' perceptions of quality of life. *Journal of Happiness Studies* **6**, 15–24.
- Pekrun, R. (2005). Progress and open problems in educational emotion research. *Learning and Instruction* **15**, 497–506.

- Ryan, R. M. and Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review Psychology* **52**, 141–166.
- Vedder, P., Boekaerts, M., and Seegers, G. (2005). Perceived social support and well being in school; The role of students' ethnicity. *Journal of Youth and Adolescents* **34**, 269–278.
- Zautra, A. J. and Hempel, A. (1984). Subjective well-being and physical health: A narrative literature review with suggestions for future research. *Aging and Human Development* **19**, 95–110.

Relevant Websites

- www.education.vic.gov.au – Department of Education and Early Childhood Development, Victoria, Australia.
- www.minedu.govt.nz – Ministry of Education, New Zealand.
- www.psych.uiuc.edu – Psychology Department, University of Illinois Urbana-Champaign.
- www.scotland.gov.uk – The Scottish Government.

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

This page intentionally left blank

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

EDITORS-IN-CHIEF

PENELOPE PETERSON
School of Education and Social Policy
Northwestern University
Evanston, IL
USA

EVA BAKER
UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

BARRY MCGAW
McGaw Group Pty Ltd
Mont Albert Nth, VIC
Australia



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
The Boulevard Langford Lane, Kidlington, Oxford OX5 1GB, UK

Copyright © 2010 Elsevier Ltd. All rights reserved

LATVIA
Copyright © 2010 A Joma

STUDENT DATA SYSTEMS AND THEIR USE FOR EDUCATIONAL IMPROVEMENT
Copyright © 2010 J C Wayman, V Cho, and M P Richards

QUALIFICATIONS FRAMEWORKS AND THEIR ROLE IN THE REFORM OF EDUCATION AND TRAINING
Copyright © 2010 M Young

LANGUAGE OF VOCATIONAL EDUCATION AND TRAINING
Copyright © 2010 J Misko, J MacKenzie, and R-A Polvere

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44)(0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at (<http://elsevier.com/locate/permissions>), and selecting *Obtaining permissions to use Elsevier material*

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Catalog Number
A catalogue record for this book is available from the Library of Congress

ISBN: 978-0-08-044893-0

For information on all Elsevier publications
visit our website at books.elsevier.com

PRINTED AND BOUND IN THE UNITED KINGDOM
09 10 11 12 13 10 9 8 7 6 5 4 3 2 1

PREFACE

A preface usually provides a brief introduction to a work, intended to set the stage, provide some background insight, and whet the appetite of the reader. In our case, however, this preface has to address a fundamental question, one that was in our minds at the time we were recruited as Editors-in-Chief for the International Encyclopedia of Education. The question was “Why do we need an Encyclopedia? Its subtext was inspired by the ever-growing, ever-popular Internet. We believe that *this* Encyclopedia is desperately needed and will become a valued resource in education and associated social sciences and arts. The reasons are intellectual and procedural. Anyone with a modicum of knowledge knows that finding and trusting information gleaned from the Internet are two separate actions. The reliance on browsers to help discover references and comments result in resources based on popularity not quality. Pithy titles catch the eye and references rise in the ranks of browser searchers. Related to this is the “editing” in the Internet realm of populist efforts at encyclopedia, references, and other compilations. Once again, after removing offensive material, the accuracy, completeness, lack of bias, and other provenance for entries simply do not exist. Experienced researchers in education can sort through and make intelligent choices. Novices and many journeyman, or practitioners, parents, and policy makers cannot. Contrast how this Encyclopedia was built. Key domains of educational research were identified, and a tentative list of sub-domains or useful applied areas was posited. Then the Editors-in-Chief (apologies for the awkwardness of the term) identified the leading researcher in a particular domain, and with surprisingly little effort, recruited them to participate. They in turn identified the two best researchers in a sub-domain, such as formative assessment or the training of pre-school teachers. The authors of the sections of the Encyclopedia do not represent a collective group of friends and acquaintances, although friendships have been made. Rather they embody a deep and broad scholarly community. The difference from compiled Internet resources is the built-expertise and intellectual engagement of the authors. The summary of the developments and futures in their personal areas of scholarship have been filtered through their years of experience, both as scholars and communicators. Quality, then, is endemic to each piece, developed through this top-down identification of expertise, and made indelible by the bottom-up application of high standards from people leading the sub-domains – the authors, and the domains themselves, the section editors.

On a procedural level, the publishers early committed to the notion that this Encyclopedia would also be an online resource, and access would be available through print, for those with strong bookcases and the persisting love of turning real pages. The Internet version will allow multiple prisms through which the reader may access articles and provide, as it were, an emulation of the Internet in our field, albeit bounded by expertise and high quality.

What must be underscored in the assessment of this effort are the Editors-in-Chief and the publishers’ commitment to find excellence worldwide. We tried very hard to persuade notable scholars from all parts of the world to make contributions. Less than to fulfill the title of “International,” we were on the hunt for perspectives that would enrich the scope and depth of the sections. Our section editors put in enormous time attempting to find the best in the field, wherever they resided. Yet, not everyone is in the volume. Some were overcommitted. Many were not fully confident of their English, and the automated translation software has not yet met standards for technical writing. We believe that such writing and editing tools will make the outreach to an even broader International group of scholars possible in future revisions, or online updates. Furthermore, the birth of the World Educational Research Association (in 2009) will provide a better set of interlocking networks to find and evaluate scholarship from any place on the globe.

Finally, the scope of the effort must be acknowledged: 28 section editors, 926 articles were commissioned, drafted, reviewed, redrafted, edited, and put together in the space of four years. The publishers underwent some internal changes, and alterations in management. We as Editors-in-Chief, changed roles, moved, and also had to keep our own research and development enterprises afloat. Deadlines wobbled; authors dropped from view and had to be replaced.

Yet, at times frustrating as all development is, we find the final product exhilarating. We are enthusiastic not simply because it came into being at all, but because the collective light of the minds that wrote have left a bright resource for the future, one that will impact the way our colleagues understand and experience the educational knowledge, improvement, and impact in the future.

Eva Baker
Penelope Peterson

EDITORS

EDITORS-IN-CHIEF

Penelope Peterson

School of Education and Social Policy
Northwestern University
Evanston, IL
USA

Eva Baker

UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

Barry McGaw

McGaw Group Pty Ltd
Mont Albert Nth
Melbourne, VIC
Australia

SECTION EDITORS

Walter Allen

HEOC Division
University of California, Los Angeles
Los Angeles, CA
USA

Vibeke Grøver Aukrust

Institute for Educational Research
University of Oslo
Oslo
Norway

Dominic J Brewer

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

Kai-Ming Cheng

Faculty of Education
The University of Hong Kong
Hong Kong

Gregory Chung

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Penny Enslin

Department of Educational Studies
University of Glasgow
Glasgow
UK

Noelle Griffin

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Dennis Hocevar

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

David Hopkins

Institute of Education
University of London
London
UK

Mary James

Faculty of Education
University of Cambridge
Cambridge
UK

Sanna Järvelä

Department of Educational Sciences and
Teacher Education
University of Oulu
Oulu
Finland

Thomas Karmel

National Centre for Vocational Education Research
Adelaide, SA
Australia

Harold G Levine

School of Education
University of California, Davis
Davis, CA
USA

Robert L Linn

University of Colorado at Boulder
Ouray, CO
USA

Allan Luke

Faculty of Education
Queensland University of Technology
Brisbane, QLD
Australia

Ann M Mastergeorge

Human Development and Family Studies
University of California, Davis
Davis, CA
USA

Patrick J McEwan

Department of Economics
Wellesley College
Wellesley, MA
USA

Barbara Means

Center for Technology in Learning
SRI International
Menlo Park, CA
USA

Ingrid Moses

University of Canberra
Canberra, ACT
Australia

Bill Mulford

Faculty of Education
University of Tasmania
Hobart, TAS
Australia

Yu Jing Ni

Faculty of Education
The Chinese University of Hong Kong
Hong Kong

Festus Obiakor

Department of Exceptional Education
University of Wisconsin-Milwaukee
Milwaukee, WI
USA

P David Pearson

Graduate School of Education
University of California, Berkeley
Berkeley, CA
USA

Colin Power

Graduate School of Education
University of Queensland
Brisbane, QLD
Australia

Kjell Rubenson

Faculty of Education
The University of British Columbia
Vancouver, BC
Canada

Sandip Sinharay

Education Testing Services
Princeton, NJ
USA

Collette Tayler

Queensland University of Technology
Brisbane, QLD
Australia

Nico Verloop

ICLON, Graduate School of Education
Leiden University
Leiden
Netherlands

HOW TO USE THE ENCYCLOPEDIA

The International Encyclopedia of Education is intended for use by students, research professionals, and interested others. Articles have been chosen to reflect major disciplines in the study of education and common topics of research by academics in this domain. Each article serves as a comprehensive overview of a given area, providing both breadth of coverage for students, and depth of coverage for research professionals. We have designed the encyclopedia with the following features for maximum accessibility for all readers.

The contents of the encyclopedia are arranged alphabetically by section, and within sections, alphabetically by article. The Subject Index is located in Volume 8. Some topics are covered in a multitude of articles from differing perspectives, while other topics may have only one entry. We encourage use of the index for access to a subject area, rather than use of the Contents list alone, so that a reader has a full notion of the coverage of that topic.

The articles include cross-references to other related encyclopedia articles, suggested further readings where applicable, and many contain relevant websites for additional information. We encourage readers to use the cross-references to locate other encyclopedia articles that will provide more detailed information about a subject.

The Further Reading sections include recent secondary sources to aid the reader in locating more detailed or technical information. Review articles and research articles that are considered of primary importance to the understanding of a given subject area are also listed. These suggested further readings are not intended to provide a full reference listing of all material covered in the context of a given article, but are provided as next steps for a reader looking for additional information.

This page intentionally left blank

CONTENTS

Preface	v–vi
Editors	vii–viii
How to use the Encyclopedia	ix

VOLUME 1

A

Adult Education

Adult Education Overview	<i>K Rubenson</i>	1
--------------------------	-------------------	---

Adult Education – Adult Learning, Instruction and Program Planning

Adult Learning	<i>S B Merriam</i>	12
Adult Learning and Instruction: Transformative Learning Perspectives	<i>P Cranton</i>	18
Adult Learning in a Biographic Perspective	<i>L West</i>	25
Adult Learning, Instruction and Programme Planning: Insights from Freire	<i>P Mayo</i>	31
Characteristics of Adult Learning	<i>K Illeris</i>	36
Informal Learning: A Contested Concept	<i>P Hodgkinson</i>	42
Organizational Learning	<i>P-E Ellström</i>	47
Program Planning	<i>A L Wilson and R M Cervero</i>	53
Workplace Learning Frameworks	<i>S Billett</i>	58

Adult Education – as a Field of Study

Class Analysis in Adult Education	<i>T Nesbit</i>	64
Gender Analysis	<i>J Stalker</i>	70
Race and Ethnicity in the Field of Adult Education	<i>J Johnson-Bailey and D Drake-Clark</i>	77
Trends in Workplace Learning Research	<i>T Fenwick</i>	83

Adult Education – Concepts

Lifelong Learning	<i>J Field</i>	89
Rewriting the History of Adult Education: The Search for Narrative Structures	<i>B J Hake</i>	96

Adult Education – Domains and Provision

Adult Basic Education: A Challenge for Vocational Based Learning	<i>R Höghelm</i>	102
Adult Literacy Education	<i>L Tett and R St.Clair</i>	107
Citizenship and Immigrant Education	<i>D Schugurensky</i>	113
Community Based Adult Education	<i>B Connolly</i>	120
Continuing Professional Education: Multiple Stakeholders and Agendas	<i>R M Cervero and B J Daley</i>	127
Financing of Adult and Lifelong Learning	<i>Å Sohlman</i>	133
Labor Education	<i>K Forrester</i>	139
Learning Cities and Regions	<i>C Duke</i>	144
Museums as Sites of Adult Learning	<i>R S Grenier</i>	150
Overview of Lifelong Learning Policies and Systems	<i>D-B Kwon and C H O Daeyeon</i>	156
Popular Adult Education	<i>S Tøsse</i>	162
Provision of Prior Learning Assessment	<i>P Andersson</i>	169
The Age of Learning: Seniors Learning	<i>P Jarvis</i>	175
University Adult Continuing Education: The Extra-Mural Tradition Revisited	<i>M Slowey</i>	183

Adult Education – Economy and Society

Adult Education and Civil Society	<i>J Crowther and I Martin</i>	191
Adult Education and Nation-Building	<i>S Walters</i>	198
Economic Outcomes of Adult Education and Training	<i>A Ferrer and W C Riddell</i>	204
Health and Adult Learning	<i>P Bélanger and M Robitaille</i>	211
Modernization Processes and the Changing Function of Adult Learning	<i>H S Olesen</i>	217
The Political Economy of Adult Education	<i>G Rees</i>	223
Wider Benefits of Adult Education	<i>R Desjardins and T Schuller</i>	229

Adult Education – Participation

Barriers to Participation in Adult Education	<i>K Rubenson</i>	234
Participation in Adult Learning	<i>R Desjardins</i>	240

C**Comparative Education**

Comparative and International Education: English-Language Scholarly Publications	<i>D Post and M Fedri</i>	251
Comparative Education: Philosophical Issues and Concepts	<i>M Mason and C W Evers</i>	257
Comparative Education: Societies and Associations	<i>M Bray</i>	266

Curriculum Development

Curriculum and Religion	<i>C Kapitzke and P Graham</i>	272
Gender and Curriculum	<i>J Marsh</i>	277
Local Teacher-Based Curriculum Development	<i>M D Osborne and H B Barker</i>	283

Curriculum Development – Contexts

Curriculum and Human Rights	<i>C Lohrenscheit</i>	287
Curriculum and the Education of Cultural and Linguistic Minorities	<i>S May</i>	293
Curriculum in Postcolonial Contexts	<i>A Hickling-Hudson</i>	299
Curriculum, Economic and Cultural Development	<i>A Rampal</i>	307
Globalization and Curriculum	<i>R Dale</i>	312
Popular Culture	<i>K Schultz and R Throop</i>	318

Curriculum Development – Evaluation and Research

An Overview of Research in Curriculum Inquiry	<i>F M Connelly and S J Xu</i>	324
Curriculum Evaluation: Approaches and Methodologies	<i>V Klenowski</i>	335
Curriculum Studies, Discourse Analysis, and the Construction of Historical Time	<i>J Laspina</i>	342
Curriculum Use in the Classroom	<i>M Ben-Peretz and B Eilam</i>	348
The Roots and Routes of Teacher-Based Action Research and Curriculum Inquiry: An Historical Perspective	<i>L G Putney and J L Green</i>	355

Curriculum Development – Planning and Development

Curriculum and Syllabus Design	<i>A Woods, A Luke, and K Weir</i>	362
Curriculum and the Publishing Industry	<i>J Larson, A-R Allen, and D Osborn</i>	368
Curriculum and Teacher Change	<i>J G Ladwig</i>	374
Curriculum Governance and Planning	<i>B Levin</i>	379
Curriculum Planning and Systems Change	<i>Z Deng</i>	384
Curriculum, Digital Resources and Delivery	<i>Y Zhao, G Zhang, and C Lai</i>	390
Textbook Development and Selection	<i>P Shannon</i>	397

Curriculum Development – Subject Matter

Curriculum and Globalization: Higher Education	<i>K H Mok</i>	403
Curriculum Development as Subject Matter: Social Studies	<i>J Brophy and J Alleman</i>	411
Curriculum Development in the Area of Reading	<i>C L Madda and T E Raphael</i>	415
Curriculum in the Arts	<i>C M Thompson, L Bresler, and T E Costantino</i>	423
Curriculum: Foreign Language Learning	<i>Angel M Y Lin</i>	428
First Language Learning	<i>T Rogers and L Moffatt</i>	435
Learning a Second Language in First Language Environments	<i>E Bernhardt and D J Tedick</i>	441
Literature	<i>S Hynds</i>	447
Mathematics	<i>B Kaur and C P Vistro-Yu</i>	453
Physical Education and Sports	<i>D Kirk</i>	459
Primary and Elementary/Middle Grades Reading	<i>D E Alvermann and C A Mallozzi</i>	464
Science	<i>Y Kali and M Linn</i>	468

Curriculum Development – Theory

Bilingual Learning (Learning L1 and L2 in an L1 and L2 Environment)	<i>D Schwinge</i>	475
Curriculum and Complex Systems Theory	<i>W-M Roth and J Thom</i>	481

Curriculum and Constructivism	<i>B Davis and D Sumara</i>	488
Curriculum and Critical Theory	<i>L Yates</i>	494
Curriculum and Poststructuralist Theory	<i>J L Miller</i>	499
Curriculum and Structuralist Sociology: The Theory of Codes and Knowledge Structures	<i>J Muller and J Gamble</i>	505
Curriculum Reform	<i>R Gilbert</i>	510
Early Childhood Curriculum and Developmental Theory	<i>S M-Y Lim and C Genishi</i>	514

D

Demography and Social Change

An Evaluation Framework for E-Learning Effectiveness in the Arab World	<i>W AbuSneineh and M Zairi</i>	521
Globalization and Social Justice in Higher Education	<i>E Kimura-Walsh</i>	536
Racial Inequality and the Economic Payoffs of Higher Education	<i>S A Lachtman</i>	542
School Health Education	<i>J S Yang</i>	547

Demography and Social Change – Demography

A Paradigm of Contradictions: Racism and Science Education	<i>B A Brown and J W Mutegi</i>	554
Africana Studies: Past, Present And Future	<i>B R Hare, A Young, J B Stewart, and D P Aldridge</i>	565
Beyond Black Nationalism to Black Internationalism: The Compelling Case for Examining Black Educational Challenges Globally	<i>K Freeman</i>	573
Beyond Racial, Ethnic, and Gender Bias in Education Statistics	<i>J T Lott and C Bennett</i>	581
Caste and Education in India	<i>G G Wankhede</i>	589
Demography, Community, and the Education of Asian American and Pacific Islander Students	<i>R T Teranishi</i>	595
Educational Attainment and Mortality Differentials	<i>H D Horton, L Hixson, and S M Monnat</i>	602
Educational Opportunity and Latino/Chicano College Choice: New Findings, and Theoretical Perspectives	<i>M A Ceja and P A Pérez</i>	610
Ethnic Minority Identity and Educational Outcomes in a Rising China	<i>G A Postiglione</i>	616
Ethnic Studies and Women's Studies: From the Past to the Present, Providing a Space in the Academy for Scholarship Focused on the Histories and Experiences of Marginalized People	<i>C A Brown</i>	623
Gender Equity in Higher Education: Challenges and Celebrations	<i>L Morley</i>	629
Locating Space and Place in the College Access Debate: New Tools for Mapping and Understanding Educational Inequity and Stratification	<i>B Cantwell and J F Milem</i>	636
Public Policy and Inequality in Postsecondary Opportunity: Educational Statistics and the Failure of Education Reform	<i>E P St. John, K Williams, and K Moronski</i>	649
The Education of Indigenous Students	<i>J W Tippeconnic and S C Faircloth</i>	661
The Health Advantages of Educational Attainment	<i>G Y Dinwiddie</i>	667
To Go or Not to Go: Access and Barriers to Ethnic Minority Participation in Higher Education within the United Kingdom	<i>P Brug and T L Oliver</i>	673
Universal Secondary Education and the Two-Tiered School System in the English-Speaking Caribbean	<i>P Anderson and J Devonish</i>	679
Women of Color in Science, Technology, Engineering and Mathematics (STEM): Refining the Concepts, Reframing the Issues	<i>C B Leggon</i>	686

Demography and Social Change – Social Change

Educational Diversity in U.S. Law Schools	<i>C E Daye, A T Panter, W R Allen, M E Deo, and L F Wightman</i>	691
Educational Reform	<i>F Lopez-Segrera</i>	698
Emergence of For-profit Higher Education	<i>X Li</i>	705
Engaged Education: Experiential Learning, Intensive Field Experiences, and Social Change	<i>T R Rone</i>	711
Higher Education and Social Change in Mexico	<i>A Mungaray, G Estrella, P Moctezuma, and M Ramírez-Urquidy</i>	719
The Changing Role of the State in Higher Education	<i>K Gürüz</i>	728
The Role of the State in School Reform: Responding to Increasing Ethnic Diversity in Britain	<i>P Ratcliffe</i>	737
Transforming Mindsets Through Education for Sustainable Development	<i>V M Segovia</i>	746

Demography and Social Change – Sociology

Achievement Motivation in Ethnic Minority Youth	<i>S Graham and E Bryant</i>	753
Affirmative Action and Higher Education in Brazil	<i>J T dos Santos and D M Queiroz</i>	760
Black in White: Black Students at White and Black Colleges	<i>L Evans and J R Feagin</i>	767
Bologna Process: On the way to a Common European Higher Education Area	<i>A V Rezaev</i>	772
Critical Race Theory in Education Research	<i>R Vue and C B Newman</i>	779
Cultural Diversification and Japanese Education: Social Constructions of the New Diversity	<i>R Tsuneyoshi</i>	787
Education in War Countries: A Disempowerment of Society	<i>J L Nkata</i>	793
Educational Attainments of U.S. Black Males and Females: 1971 to 2003	<i>F B Bonner, M-C Jipguep-Akhtar, and R J Harrison</i>	799
Educational Preparation: Fostering the Self-Efficacy and Resilience of Urban Adolescent Youth	<i>M B Spencer and B Tinsley</i>	806
HIV Education for Low-Literate People: Transforming Students and Communities through Paulo Freire's Praxis and the Pedagogy of Action	<i>N Z Haniff</i>	814
How Are Discussions of Interdisciplinary Studies Linked to Diversity Discourse?	<i>C N Spence</i>	823
Identifying Patterns of Doctoral Attrition Across Academic Fields	<i>A James, K J Fosnacht, E B Stolzenberg, and D Thirumalai</i>	827
Male Underachievement in Education Across the Globe: A Shift in Paradigm for Gender Disparities Regarding Academic Achievement	<i>J F L Jackson, J L Moore III, and R A Leon</i>	838
Noncognitive Measures for Higher Education Admissions	<i>W E Sedlacek</i>	845
Peace Education	<i>J Page</i>	850
Remote Control: Africa's Development and Outflow to Developed Nations of Technical Talent	<i>F Haniff</i>	855
Scholar-Baller®: Student Athlete Socialization, Motivation and Academic Performance in American Society	<i>C K Harrison, S Bukstein, J Mottley, E Comeaux, J Boyd, C Parks, and D Heikkinen</i>	860
Service-Learning and Diversity as Mediums to Foster Civic and Community Engagement in Undergraduates	<i>L Oseguera</i>	866
Success and Failure in Education and Criminal Justice: Identifying Common Mechanisms	<i>S Ramirez and A Harris</i>	874
The History of Education: Race and Education	<i>J O Jewell</i>	881

VOLUME 2**E****Early Childhood Care and Education Theory**

- Early Childhood Care and Education: The Family, The Market, and The State *S B Kamerman and M J Neuman* 1
- Gender Issues in Early Childhood Education and Care *B Pocock* 11

Early Childhood Education and Care Governance

- Early Childhood Risk, Protection and Abuse Prevention *E Khoo* 17
- Evaluating Early Childhood Education and Care Programs *M Janus and S Brinkman* 25

Early Childhood Education and Care Investment

- Investing in Early Childhood Education and Care: The Health and Wellbeing Case *T Moore and F Oberklaid* 32
- Cost-effective Early Childhood Programs from Preschool to Third Grade *A J Reynolds, J A Temple, and B White* 38
- Early Childhood in Post-Modern Cultures: Thoughts and Some Concerns *P Gammage* 49
- Investing in Early Childhood Education and Care: Some Policy Implications *J Bennett* 55
- Investing in Early Childhood Education and Care: The Economic Case *G Cleveland and M Krashinsky* 63

Early Childhood Education and Care Programs

- Early Care and Education Programs for Infants and Toddlers *B Nyland* 69
- Literacies in Early Childhood: The Preschool Period *M Fleeer and B Raban* 75
- Participation in Early Childhood Education and Care Programs: Equity, Diversity and Educational Disadvantage *M Vandenbroeck* 81
- Teaching in Early Childhood Centers Instructional Methods and Child Outcomes *I Siraj-Blatchford* 86

Early Childhood Education and Care Provision

- Early Childhood Education and Care in the People's Republic of China *Zeng Xiaodong* 93
- Home-Based and Institutional Early-Childhood Education and Care Services *R Fukkink and H Blok* 98
- Integrated Early Childhood Education and Care Services – Care, Upbringing, Education and Health *S Colley* 104
- Issues of Access and Program Quality *J T Hustedt and W S Barnett* 110
- Kindergarten Transitions and Linkages to Primary School – Readiness Re-conceptualized *A Petriwskyj* 120
- Parent Support in Early Childhood – Approaches and Outcomes *K S Howard, K A Beckmann, and J Brooks-Gunn* 126

Early Childhood Education and Care Research

- Development and Implementation of Early Learning Standards in the United States *C Scott-Little* 132
- Global Status of Early Learning and Development Standards *P R Britto and S L Kagan* 138
- Rethinking Early Childhood Education and Care: Implications for Research and Evaluation *K Thorpe, D Cloney, and C Tayler* 144

Early Childhood Education and Care Theory

- Child Rearing and Early Education: Parents and Professionals: Theoretical and Historical Influences *H Penn* 151

Early Childhood Education and Care Workforce

- Early Childhood Education and Care Professional Profiles in the United Kingdom: Multiagency Teamwork and Transdisciplinary Challenges *A Anning and P Calder* 157
- Early Childhood Education and Care Staff Support: Mentoring and Professional Development in Finland *P Lindberg* 162

Economics of Education

- Education and Health *P Muennig* 169
- Data *M Perez and M Socias* 178
- Empirical Research Methods in the Economics of Education *P J McEwan* 187
- Theoretical Concepts in the Economics of Education *D J Brewer, G C Hentschke, and E R Eide* 193

Economics of Education – Costs and Financing of Schooling

- Cost–Benefit Analysis and Cost–Effectiveness Analysis *C Belfield and H M Levin* 199
- Economic Approaches to Adequacy *J Imazeki* 204
- Economic Approaches to School Efficiency *L L Taylor* 210
- School Finance Reform *T Downes* 216
- School Finance: An Overview *J K Rice, D H Monk, and J Zhang* 222
- The Economics of Tuition and Fees in Higher Education *R G Ehrenberg* 229

Economics of Education – External Benefits of Schooling

- Education and Civic Engagement *T S Dee* 235
- Education and Crime *L Lochner* 239
- Education and Economic Growth *E A Hanushek and L Wößmann* 245
- Parental Socioeconomic Status, Child Health, and Human Capital *J Currie and J Goodman* 253
- The External Benefits of Education *W W McMahon* 260

Economics of Education – Labor Markets, Education and Earnings

- Education and Inequality *J Blanden and S Machin* 272
- Human Capital *E R Eide and M H Showalter* 282
- Race Earnings Differentials *M Carnoy* 288
- Returns to Education in Developed Countries *M Gunderson and P Oreopoulos* 298
- Returns to Education in Developing Countries *H A Patrinos and G Psacharopoulos* 305
- School Quality and Earnings *J R Betts* 313
- Signaling in the Labor Market *M E Page* 321
- The Economics of High School Dropouts *R W Rumberger* 325

Economics of Education – Markets, Incentives and School Choice

- Competition and Student Performance *A A Payne* 331
- Educational Privatization *C Belfield and H M Levin* 337

Student Incentives	<i>E R Eide and M H Showalter</i>	342
The Economic Role of the State in Education	<i>D N Plank and T E Davis</i>	347
The Economics of Catholic Schools	<i>W Sander and D Cohen-Zada</i>	354
The Economics of Charter Schools	<i>R Zimmer and R Buddin</i>	360
The Economics of Parental Choice	<i>R Chakrabarti and J Roy</i>	367
The Economics of School Accountability	<i>D N Figlio and H F Ladd</i>	374
The Efficacy of Educational Vouchers	<i>R Zimmer and E P Bettinger</i>	380
Tiebout Sorting and Competition	<i>T J Nechyba</i>	388

Economics of Education – Production of Schooling

Desegregation, Academic Achievement and Earnings	<i>S Rivkin</i>	394
Education Production Functions: Concepts	<i>D N Harris</i>	402
Education Production Functions: Developed Country Evidence	<i>E A Hanushek</i>	407
Education Production Functions: Evidence from Developing Countries	<i>P Glewwe and S Lambert</i>	412
Family Environment in the Production of Schooling	<i>R Rothstein</i>	423
Neighborhoods and Peers in the Production of Schooling	<i>J Vigdor and J Ludwig</i>	431
Teacher Quality in Education Production	<i>I Pelayo and D J Brewer</i>	438
The Economics of Class Size	<i>D W Schanzenbach</i>	443
The Economics of Early Childhood Interventions	<i>M Nores</i>	450

Economics of Education – Teaching Labor Markets

An Overview of Teacher Labor Markets	<i>H Lankford and J Wyckoff</i>	456
Compensating Differentials in Teacher Labor Markets	<i>J Chambers</i>	465
Economic Approaches to Teacher Recruitment and Retention	<i>S Loeb and J Myung</i>	473
Teacher Incentives	<i>L Santibañez</i>	481
Teacher Supply	<i>P Dolton</i>	489
Teacher Training and Preparation in the United States	<i>M Podgursky, M G Springer, and R Hutton</i>	498
Teachers in Developing Countries	<i>F H Rogers and E Vegas</i>	504
The Economics of Teachers' Unions in the United States	<i>K Strunk</i>	512

Education of Children with Special Needs

Adapted Physical Activity for Students with Special Needs	<i>S R Hodge</i>	519
Assessment Accommodations for Children with Special Needs	<i>R J Kettler and S N Elliott</i>	530
Assistive Technology and Educational Practice	<i>H P Parette and G R Peterson-Karlan</i>	537
Augmentative–Alternative Communication	<i>C Mason and K M Chinn</i>	544
Community Integration and Employment of Youth with Special Needs	<i>B A Ford and A W Daviso III</i>	550
Cooperative Learning for Children with Special Needs	<i>E A Dooley and K R Kossar</i>	555
Co-Teaching	<i>E D Cramer</i>	560
Cultural and Linguistic Diversity and Children with Special Needs	<i>J M Patton, N Day-Vines, and A McSwain</i>	565
Curriculum-Based Assessment and Students with Special Needs	<i>D J O'Shea and J D Katsafanas</i>	571

Diagnosis and Classification	<i>M Sapp</i>	577
Direct Instruction and the Education of Children with Special Needs	<i>T Mehring</i>	584
Disabilities and Gifted Education	<i>V I Daniels and M J McCollin</i>	588
Early Childhood Special Education: Birth to Eight Years Old	<i>T S Lovelace, R Gardner III, J B Axe, and D M Sainato</i>	594
Early Intervention for Students with Special Needs	<i>P A Grant</i>	608
Early Transition of Children with Special Needs	<i>S A Fowler</i>	613
Educating Children Who are Deaf-Blind	<i>J Wood and K M Chinn</i>	621
Educating Children with Deafness and Hearing Impairments	<i>G M Mukuria and C J Eleweke</i>	628
Educating Children with Other Health Impairments (ADHD)	<i>G M Mukuria and F E Obiakor</i>	634
Educating Children with Physical Disabilities	<i>Z P Nkabinde, F E Obiakor, M T Ofor, and D J Smith</i>	640
Educating Children with Specific Learning Disabilities	<i>P A Grant and P B Grant</i>	646
Educating Children with Speech and Language Disorders	<i>S L Green and C D Qualls</i>	654
Educating Students with Autism and Related Disorders	<i>T T Dyches</i>	661
Educating Students with Cognitive Disabilities	<i>J P Bakken</i>	669
Educating Students with Emotional and Behavioral Disorders	<i>S W Smith and G G Taylor</i>	678
Educating Students with Special Needs: An Overview	<i>F E Obiakor, M K Harris, P Harris-Obiakor, and G M Mukuria</i>	688
Educating Students with Traumatic Brain Injury	<i>S Burkhardt and A Rotatori</i>	695
English Language Learners with Special Needs	<i>R Rueda and G Ragusa</i>	701
Functional Behavioral Assessment	<i>J P Bakken</i>	708
History of Special Education	<i>L M Garrick Duhaney and S J Salend</i>	714
Inclusion of Students with Special Needs in General Education Classrooms	<i>M A Prater</i>	721
Instructional Accommodations for Children with Special Needs in Inclusive Settings	<i>A L Yurick, S Musti-Rao, and G Cartledge</i>	727
Interagency Agreements	<i>L A Owens</i>	733
Interdisciplinary Collaboration in Early Childhood	<i>N File</i>	739
Language Development in Children with Special Needs	<i>M S L Stewart</i>	745
Large Scale Assessment and Accountability and Students with Special Needs	<i>M L Thurlow</i>	752
Literacy Instruction for Students with Special Needs	<i>L H Mason</i>	759
Math Instruction for Children with Special Needs	<i>B A Bottge</i>	767
Overrepresentation of Culturally and Linguistically Diverse Students in Special Education	<i>S C Trent</i>	774
Paraprofessionals in Special Education	<i>B Y Ashbaker, S C Dunn, and J Morgan</i>	780
Parent and Family Involvement in the Education of Children with Special Needs	<i>L P Prater</i>	788
Peer Relations and Socialization of Children and Adolescents with Special Needs	<i>L M Bullock, M Wong-Lo, and R A Gable</i>	794
Peer-Tutoring	<i>L J O'Shea and D J O'Shea</i>	802
Postsecondary Participation of Students with Special Needs	<i>R A Stodden and K D Roberts</i>	808
Prereferral Strategies	<i>C G Simpson and J P Bakken</i>	815

Preventing Antisocial Behavior and Delinquency: A Comprehensive School-wide Approach <i>P R Hallam, K R Young, P Caldarella, D G Wall, and L Christensen</i>	820
Programs and Instructional Strategies for Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	829
Related Services for Children with Special Needs <i>S O Obi</i>	834
Risk and Resilience Frameworks in Understanding Special Education <i>W-Y Hsieh and M L Donahue</i>	840
School Dropout Prevention <i>F E Obiakor</i>	845
School-Based Services for Children with Special Needs <i>S O Obi</i>	850
Self-Determination <i>M L Wehmeyer</i>	857
Social Competence <i>L A Owens and S Johnston-Rodriguez</i>	865
The Identification of Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	870
Transition from School to Adult Life <i>M J McCollin and F E Obiakor</i>	876
Visual Impairment <i>W Sapp</i>	880
Workforce Issues in Special Education <i>E Carlson and B Billingsley</i>	886

VOLUME 3

Education of Professionals

An Overview of Accreditation, Certification, and Licensure Processes <i>J J Norcini and M van Zanten</i>	1
Educating America's Military Officers <i>M H Adamshick</i>	7
Education for Architecture in the United States and Canada <i>S Schreiber</i>	13
Education for Medicine <i>M B Anderson</i>	19
Education for the Legal Profession <i>J W Wegner</i>	25
Education of Clergy <i>D O Aleshire</i>	31
Framing Professions Education <i>P S O'Sullivan</i>	37
Issues in Accounting/Business Education <i>W H Gijsselaers and R G Milter</i>	39
Professions – Engineering <i>J Heywood</i>	45
Professions – Health-Related <i>G M Jensen and C B Royeen</i>	51
Professions – Pharmacy <i>D E Beck</i>	58
Teaching for the Complex Practice of Nursing <i>P Benner and M Sutphen</i>	65
Veterinary Education <i>K M Edmondson</i>	71

Education Research Methodology: Quantitative Methods and Research

Abductive Research Methods <i>B D Haig</i>	77
Field Experimentation in Education <i>M A Constan</i>	83
Internal Validity <i>M H Clark and S C Middleton</i>	90
Methods for Approximating Random Assignment <i>B Schneider and S-K McDonald</i>	97
Methods for Approximating Random Assignment: Regression Discontinuity and Propensity Scores <i>L M Scheier</i>	104

Multitrait–Multimethod Designs	<i>M Bong and S Hong</i>	111
Multivariate Longitudinal Data Analysis	<i>H Liu and L Chang</i>	120
Quasi-Experimentation: Two Group Design	<i>E Hong</i>	128
Regression Discontinuity Designs	<i>A Kelava, S Rohrmann, and V Hodapp</i>	134
Sampling	<i>R S Brown</i>	142
Statistical Conclusion Validity	<i>P L Busk</i>	147
Survey Research Methods	<i>A Fink</i>	152

Educational Assessment

An Overview of Educational Assessment	<i>M James</i>	161
Assessment and the Evaluation of Institutional Effectiveness	<i>S M Thomas</i>	172
Assessment in Schools – Dispositions	<i>R D Crick</i>	181
Assessment in Schools Related To Literacy: Reading	<i>K J Topping</i>	189
Impact of Assessment on Students' Learning Strategies and Implications for Judging Assessment Quality	<i>M Segers, F Dochy, and D Gijbels</i>	196

Educational Assessment – Assessment Activities

Alternative Assessment	<i>G Stobart and C Gipps</i>	202
Classroom Assessment Tasks and Tests	<i>J Phelan and J Phelan</i>	209
Dynamic Assessment	<i>J G Elliott, E L Grigorenko, and W C M Resing</i>	220
Instructional System Provided Feedback	<i>H F O'Neil, S Chuang, and T Huang</i>	226
Peer and Self-assessment	<i>C Harrison</i>	231
Portfolio Assessment	<i>V Klenowski</i>	236
Records of Achievement: Beyond Traditional Tests	<i>P Broadfoot</i>	243

Educational Assessment – Assessment in Domains

Assessment in Higher Education	<i>D R Sadler</i>	249
Assessment in Schools – Affective Domain	<i>M Ventura, R D Roberts, and M Zeidner</i>	256
Assessment in Schools – Creative Subjects	<i>T E Costantino and L Bresler</i>	262
Assessment in Schools – Learning to Learn	<i>A Hautamäki, J Hautamäki, and S Kupiainen</i>	268
Assessment in Schools – Literacy Writing (Extended)	<i>J M Parr and L Earl</i>	273
Assessment in Schools – Mathematics	<i>M Brown and J Hodgen</i>	279
Assessment in Schools – Oracy	<i>A L Bailey</i>	285
Assessment in Schools – Primary Science	<i>W Harlen</i>	293
Assessment in Schools – Secondary Science	<i>J Shemwell, A Fu, M Figueroa, R Davis, and R Shavelson</i>	300
Assessment in Schools – Technology Education and ICT	<i>A Jones, B Cowie, and J Moreland</i>	311
Assessment in the Early Years	<i>M J Drummond</i>	316
Assessment in the History, Civics and Social Studies Domains	<i>P G Avery</i>	323
Assessment in the Workplace of Performance, Developing Expertise and Competence	<i>A J Brown</i>	330
Assessment in Vocational Education	<i>K Ecclestone</i>	337

Educational Assessment – Concepts and Issues

Assessing Group Work	<i>M Galton</i>	342
Assessment and the Regulation of Learning	<i>L Allal</i>	348
Challenges of Developing and Implementing Formative Assessment Practices in Schools	<i>R Sutton</i>	353
Formative Assessment	<i>P Black</i>	359
Formative Assessment and Instructional Planning	<i>B Marshall</i>	365
Progression and Assessment: Developmental Assessment	<i>M Forster and G Masters</i>	369
Students' Rights in Assessment and Evaluation	<i>J Gardner and B Holmes</i>	378
Summative Assessment by Teachers	<i>R Daugherty</i>	384
The Multiple Purposes of Assessment	<i>P E Newton</i>	392
The Relationship between Assessment and the Organization and Practice of Teaching	<i>L B Resnick, D William, R Apodaca, and E S Rangel</i>	397

Educational Assessment – Enhancing Quality and Use

Assessment Conversations: Reading and Writing Conferences with Students and with Parents	<i>M Barrs</i>	403
Assessment: Pre-service and In-service Teacher Education	<i>A Davies</i>	410
Classroom Assessment in Policy Context (Australia)	<i>J J Cumming</i>	417
Classroom Assessment in Policy Context (England and France)	<i>M Osborn and M Raveaud</i>	425
Classroom Assessment in Policy Context (French Sources)	<i>D Laveault</i>	432
Classroom Assessment in Policy Context (Hong Kong)	<i>D Carless</i>	438
Classroom Assessment in Policy Context (New Zealand)	<i>T J Crooks</i>	443
Classroom Assessment in Policy Context (Sub-Saharan Africa)	<i>S J Howie, L Zimmerman, and K Draper</i>	449
Moderation of Student Work by Teachers	<i>G S Maxwell</i>	457
School Policies and Practices to Support Effective Classroom Assessment for Learning	<i>D Pedder</i>	464
Student Assessment: Policy and Practice in Eastern Europe	<i>G Bethell</i>	472

Educational Assessment – Key Relationships

Assessment Practice in Policy Context: Latin American Countries	<i>F M Rizo</i>	479
Impact of Assessment on Learner Groups (Boys/Girls)	<i>J Elwood</i>	486
Impact of Assessment on Learner Groups (Disabilities)	<i>M Rouse</i>	492
Impact of Assessment on Students' Test Anxiety	<i>P Buchwald and C Schwarzer</i>	498
Impact of Assessments on Classroom Practice	<i>J L Herman</i>	506
Social Practices in School Assessment and their Impact on Learner Identities	<i>A Pollard and A Filer</i>	512

Educational Evaluation – Concepts, Practice, and Future Directions

Educational Evaluation: Concepts, Practice, and Future Directions	<i>Y J Ni</i>	518
---	---------------	-----

Educational Evaluation – Contemporary Themes in Evaluation

Evaluating E-Learning	<i>D D Williams and C R Graham</i>	530
Evaluating Schools as Learning Communities	<i>S M Thomas</i>	539
Evaluation of Adult Education and Training Programs	<i>G McNamara, P Joyce, and J O'Hara</i>	548
Evaluation of Integrated Health Programs in School	<i>T M Franke</i>	555

Formative Assessment in Teacher Education and Teacher Professional Development	<i>H Tillema</i>	563
Informal Education and Evaluation	<i>T Jeffs</i>	572
Methods to Evaluate Technology	<i>W L Bewley</i>	578

Educational Evaluation – Contexts of Educational Evaluation

Cost Analysis in Evaluation Studies	<i>J D Fletcher</i>	585
Cultural Issues that Can Affect the Validity of Educational Evaluations	<i>R P Durán</i>	592
Evaluation and Accountability	<i>E Sato and S N Rabinowitz</i>	600
Evaluation Governance and Planning	<i>R Picciotto</i>	607
Evaluation Use	<i>V G Thomas</i>	614
Moral and Ethical Issues in Evaluation	<i>M Morris</i>	622
The Role of Stakeholders in Educational Evaluation	<i>S Taut and M C Alkin</i>	629

Educational Evaluation – Evaluation Domains

Curriculum Evaluation	<i>T Levin</i>	636
Evaluating Education in Three Policy Eras	<i>M A Peters, R Britez, and S M Weber</i>	645
Evaluation of Mathematics Education Programs	<i>J Cai</i>	653
Evaluation of Science Education Programs	<i>G Berberoğlu</i>	660
Evaluation of Social Studies and Civic Education Programs	<i>S G Grant</i>	667
Evaluation of Teacher Quality and Practice	<i>W Robinson and J Campbell</i>	674
International Evaluations	<i>A Schleicher</i>	681
National Assessment Programs: The Example of the U.S. National Assessment of Educational Progress	<i>M Schneider</i>	690
Principal Evaluation: Concepts, Needs and Realities	<i>N S Glasman and L D Glasman</i>	698
Program Evaluation	<i>C A Christie and L A Fierro</i>	706
School Based Evaluation: Purposes, Protocols and Processes	<i>J MacBeath</i>	713
School Inspection/External School Evaluation	<i>C Chapman and P Earley</i>	719
The International System for Teacher Observation and Feedback: A Theoretical Framework for Developing International Instruments	<i>L Kyriakides, B P M Creemers, C Teddlie, and D Muijs</i>	726

Educational Evaluation – Evaluation Methods

Evaluation Methodology	<i>M Chatterji</i>	735
Evaluation Reporting and Communicating	<i>R T Torres</i>	746
Longitudinal Evaluation Designs	<i>X Ma</i>	757
Meta-evaluation: Purpose, Prescription, and Practice	<i>L A Wingate</i>	765

Educational Evaluation – Purposes and Models of Educational Evaluation

Defining Quality in Evaluation	<i>P Dahler-Larsen</i>	775
Internal and External Evaluation	<i>D Nevo</i>	781
Needs Assessment in Education	<i>J W Altschuld and T L Lepicki</i>	786
The Purpose of Educational Evaluation	<i>S Mathison</i>	792
Understanding Approaches to Evaluation	<i>A J Love</i>	798

VOLUME 4**Educational Measurement**

Ability Testing	<i>P B Tymms</i>	1
Admissions Testing	<i>R Zwick</i>	7
Appropriateness Measurement: Person Fit	<i>R R Meijer and A M Weekers</i>	15
Automated Essay Scoring: Writing Assessment and Instruction	<i>M D Shermis, J Burstein, D Higgins, and K Zechner</i>	20
Classical Test Theory Reliability	<i>M D Miller</i>	27
Criterion-Referenced Measurement	<i>S A Shrock and W C Coscarelli</i>	31
Differential Item Functioning	<i>M D Hidalgo and J Gómez-Benito</i>	36
Educational Measurement: Overview	<i>R L Linn</i>	45
Equating and Scaling	<i>A A von Davier</i>	50
Fairness	<i>C Gipps and G Stobart</i>	56
Generalizability Theory	<i>R L Brennan</i>	61
Impact of Assessment on Classroom Practice	<i>J L Herman</i>	69
Intelligent Systems	<i>V J Shute and D Zapata-Rivera</i>	75
Item Response Theory	<i>W J van der Linden</i>	81
National Assessments	<i>T J Crooks</i>	89
Norm-Referenced Measurement	<i>H D Hoover</i>	95
Progress Monitoring	<i>L S Fuchs and D Fuchs</i>	102
Rasch Models	<i>D Andrich</i>	111
Scoring Rubrics	<i>J A Arter</i>	123
Second Language Assessment	<i>L F Bachman</i>	140
Self Adaptive Testing	<i>J Revuelta</i>	148
Student Test Results in School Accountability	<i>L S Hamilton</i>	153
Test Development	<i>S M Downing</i>	159
Test Translation and Adaptation	<i>A Allalouf and P Hanani</i>	166
Testing Creativity	<i>M A Runco</i>	170
Testing in History	<i>S Wineburg and B Fogo</i>	175
Validity	<i>R L Linn</i>	181
Validity of Achievement Gains on High-Stakes Tests	<i>D Koretz</i>	186
Validity of Educational Indicators	<i>B P M Creemers and L Kyriakides</i>	193

G**Globalization of Universities and Institutions**

Globalization, Universities and Medium of Instruction	<i>J Lo Bianco</i>	201
The Washington Accord on Engineering Curriculum	<i>E Ko</i>	209

H

Higher Education

Higher Education: An Overview	<i>P Scott</i>	217
-------------------------------	----------------	-----

Higher Education – Management, Leadership and Governance

Academic Managers – Heads of Departments/Schools	<i>S Marshall</i>	229
Leading and Managing the University – Presidents and their Senior Management Team	<i>R Middlehurst, T Kennie, and S Woodfield</i>	238
Management of and in Higher Education Institutions	<i>B Sporn</i>	245
Managerialism and Collegialism in Higher Education Institutions	<i>M Shattock</i>	251
Strategic Planning in Higher Education	<i>L Zechlin</i>	256
The Changing Role of University Presidents, Vice-Chancellors and Rectors	<i>M F Green and P D Eckel</i>	264
The Changing Roles of University Governing Boards and Councils	<i>J Fielden</i>	273

Higher Education – Research and Scholarship

Funding of University Research	<i>G Harman</i>	279
Mobility of PhD Students and Scientists	<i>S Avveduto</i>	286
National Science Policy and Universities	<i>M Henkel and U Kogan</i>	294
Research Commercialization	<i>S Slaughter</i>	300
Research Quality Assessment	<i>I McNay</i>	307
The Management of University Research	<i>J Kirkland</i>	316
The Training of New Scientists	<i>R Neumann</i>	322

Higher Education – Society

Academic Freedom	<i>S F Macintyre</i>	328
Access and Equity in Higher Education	<i>G Baldwin and R James</i>	334
Boundaries of Institutional Autonomy and their Impact on Higher Education	<i>V L Meek</i>	341
Diversity of Higher Education	<i>U Teichler</i>	347
Employability of University Graduates and Graduate Outcomes	<i>J Brennan</i>	354
Higher Education and the Knowledge Society	<i>J Välimaa</i>	360
Higher Education and the Labor Market	<i>J Enders</i>	365
Higher Education and the Transformation of Society	<i>P Scott</i>	370
Quality Assurance in Higher Education – Practices and Issues	<i>D Dill</i>	377
The Role of Accreditation of Higher Education Institutions	<i>J S Eaton</i>	384
Universities and Regional Development	<i>J Goddard and J Puukka</i>	390
Universities' Engagement with Society	<i>D Watson</i>	398

Higher Education – Teaching and Learning in Higher Education

Distance Education and Open Universities	<i>A Kanwar and J Daniel</i>	404
Education Development in Higher Education	<i>Y Ryan and K Fraser</i>	411
Effective Use of Technology in Teaching and Learning in HE	<i>D Laurillard</i>	419

Facilitating Students' Success – Student Support and Learning Resources in Universities	<i>E Woodberry and C Barnes</i>	427
Faculty Values and Expectations and Their Implications for Teaching, Learning and the Organization of Higher Education Institutions	<i>K Harman</i>	433
Flexible Learning in Higher Education	<i>S Alexander</i>	441
Innovation in Teaching and Curriculum Design	<i>J M Dearn</i>	448
Productivity of University Faculty Staff	<i>S Kyvik</i>	455
Promoting Effective Student Learning in Higher Education	<i>K Trigwell</i>	461
Student Attrition from Higher Education Institutions	<i>M Hayden</i>	467
University Students' Experiences of Higher Education	<i>C McInnis</i>	473
What Do University Students Learn?	<i>G D Kuh and B K Bridges</i>	479

Higher Education – Tertiary Education in a Globalized World

Associations of Universities	<i>G F Moodie</i>	485
Collective Bargaining in Higher Education	<i>C Allport, K McAlpine, and S Roberts</i>	491
Financing of Higher Education	<i>B Chapman and P Tulip</i>	499
Higher Education Crossing Borders	<i>J Knight</i>	507
Higher Education in Post-Conflict Societies	<i>M Daxner</i>	514
Higher Education in the New Economies	<i>J Fielden</i>	520
Human Resource Issues in Higher Education	<i>E El-Khawas</i>	527
International Accreditation in Higher Education	<i>A Stella and D Woodhouse</i>	533
Internationalization of Higher Education	<i>M van der Wende</i>	540
National and International Rankings of Higher Education	<i>S Marginson</i>	546
Steering of Higher Education Systems – The Role of the State	<i>K Yokoyama and V L Meek</i>	554
Student and Faculty Transnational Mobility in Higher Education	<i>U Lanzendorf and B M Kehm</i>	559
The Bologna Process in European Higher Education	<i>D Rich</i>	566
The Economics and Finance of Higher Education	<i>A Wolf</i>	573
The Role of the OECD in the Development of Higher Education in a Globalized World	<i>R Yelland</i>	584
Transforming Higher Education in Developing Countries: The Role of the World Bank	<i>J Salmi and R M Bassett</i>	590
UNESCO's Role in the Development of Higher Education in a Globalized World	<i>S Uvalić-Trumbić</i>	597

Higher Education – The Organization and Financing of Higher Education

Higher Education in Federal Systems	<i>M A M Cortés and U Teichler</i>	603
Pathways and Articulation into Higher Education	<i>G F Moodie</i>	609
Unitary, Binary and other National Systems of Higher Education	<i>L Goedegebuure</i>	616
Viewing Private Higher Education: How Much, Where, Why, and What?	<i>D C Levy</i>	622

I

Immigration

Children of Migrant Populations	<i>M M Suárez-Orozco and C Suárez-Orozco</i>	629
Third Culture Kids	<i>R E Van Reken</i>	636

International Comparisons

IEA Studies in ICT	<i>N Law</i>	644
IEA Studies in Mathematics and Science	<i>F K S Leung</i>	650
IEA Study in Civic Education	<i>J Torney-Purta, J-A Amadeo, and J Schwillie</i>	656
IEA: Globalization and Assessment	<i>H Wagemaker</i>	663

International Organizations in Education

Education and the European Union	<i>S Hegarty</i>	669
National University on Educational Planning and Administration, New Delhi	<i>J B G Tilak</i>	676
On School Management – International Council on School Effectiveness and Improvement, Commonwealth Council on Educational Administration and Management	<i>A Walker and T Townsend</i>	681
UNESCO's International Institute for Educational Planning	<i>S J Hite and M Bray</i>	688

L**Leadership and Management**

Leadership and Management Overview	<i>B Mulford</i>	695
------------------------------------	------------------	-----

Leadership and Management – as a Field of Study

Contexts for Leadership at the Beginning of the 21st Century	<i>M Bottery</i>	704
Educational Leadership: Philosophical Issues	<i>C W Evers</i>	710
Ethical/Moral Issues in Educational Leadership	<i>R J Starratt, L Langlois, and P Duignan</i>	717
History of Educational Leadership/Management	<i>R Bates</i>	724
Educational Leadership as a Field of Study: Professional Journals 1994–2006	<i>A R Thomas</i>	731
Research on Educational Leadership – Approaches/Promising Directions	<i>W A Firestone and V M J Robinson</i>	740

Leadership and Management – Leader Preparation

Leaders for Productive Schools	<i>J Murphy, S N Elliott, E B Goldring, and A C Porter</i>	746
New Approaches in Preparing School Leaders	<i>S G Huber</i>	752
Research for Leaders: The National College for School Leadership	<i>G Southworth</i>	762
School Leadership Standards	<i>L Ingvarson</i>	768

Leadership and Management – Leadership Issues

Effective Leadership in Challenging Schools	<i>A Harris</i>	777
Esthetics, Educational Leadership and Management	<i>R Bates</i>	783
Leader Mistreatment of Teachers	<i>J Blase and J Blase</i>	790
Leadership and Gender	<i>J Blackmore</i>	797
Leadership and Student Voice	<i>C Lanskey and J Rudduck</i>	803
Leadership and Urban Education	<i>K A Riley</i>	809
Leadership for Learning	<i>J MacBeath</i>	817
Leadership in Diverse Cultures	<i>A Walker and P K Yu-kwong</i>	824

Leading Diverse Schools	<i>J Ryan</i>	832
Supervision and Personnel Administration	<i>P D Tucker and D G Pounder</i>	839
The Big Challenge: Leadership for Inclusion	<i>M Ainscow and A Sandill</i>	846
The Educational Leader's New Work	<i>P Gronn</i>	852

VOLUME 5

Leadership and Management – Leadership Types

A Distributed Perspective on School Leadership and Management	<i>J P Spillane</i>	1
Leadership: Authentic	<i>P T Begley</i>	7
Leadership: Democratic	<i>J Møller</i>	12
Leadership: Instructional	<i>P Hallinger, K Leithwood, and R H Heck</i>	18
Leadership: Transformative	<i>C M Shields</i>	26
Strategic Leadership	<i>B Davies and B J Davies</i>	34
Teacher Leadership and Organizational Development	<i>A Harris</i>	40
Understanding How Leadership Influences Student Learning	<i>K Leithwood and B Levin</i>	45

Leadership and Management – Politics and Governance

Equity and Educational Effectiveness	<i>P Sammons</i>	51
Leading/Managing Schools in Communities Made Poor	<i>P Thomson</i>	58
Principals Role in Restructuring Schools	<i>G M Crow and K D Peterson</i>	66
School Reform and Restructuring: Self Managing School	<i>B J Caldwell</i>	72
System Leadership	<i>D Hopkins and R Higham</i>	78

Leadership and Management – School and Community

Educational Leadership and Social Capital	<i>J West-Burnham</i>	85
Networked Learning Communities School-to-School Collaboration as an Essential Component of a System Reform Strategy	<i>D Jackson</i>	92
Networks and Communities of Knowledge	<i>T Schuller and H Theisens</i>	100
School Community Relations	<i>R L Crowson and E B Goldring</i>	106
Social Capital, Educational Institutions and Leadership	<i>S Kilpatrick, S Johns, and B Mulford</i>	113

Leadership and Management – School Effectiveness and Improvement

Leadership and Technology	<i>C F Webber</i>	120
Leadership in the Implementation of Innovations	<i>I E Bogotch, T Townsend, and M Acker-Hocevar</i>	128
Leadership: School Improvement	<i>R H Heck and P Hallinger</i>	135
Organizational Learning in Schools	<i>B Mulford and H Silins</i>	143
Professional Learning Community	<i>L Stoll</i>	151
Transformational School Leadership	<i>K Leithwood</i>	158

Learning and Cognition – Issues and Concepts

Attention in Cognition and Early Learning	<i>M Heimann, T Tjus, and K Strid</i>	165
Concept Learning	<i>S J Unsworth and D L Medin</i>	172
Cooperative Learning	<i>R E Slavin</i>	177
Intelligence	<i>R J Sternberg</i>	184
Learning from Multiple Information Sources	<i>H I Strømsø and I Bråten</i>	191
Memory	<i>S Magnussen and T Brennen</i>	197
Metacognition	<i>L Baker</i>	204
Personal Epistemology in Education: Concepts, Issues, and Implications	<i>I Bråten</i>	211
Piaget: Recent Work	<i>L Smith</i>	218
The Adult Development of Cognition and Learning	<i>E A L Stine-Morrow and J M Parisi</i>	225
Vygotsky and Recent Developments	<i>J V Wertsch</i>	231

Learning and Cognition – Issues and Concepts – Focus on Cognition

Cognition and Emotion	<i>F Pons, M de Rosnay, and F Cuisinier</i>	237
Cognition: Overview and Recent Trends	<i>J Wiley and B D Jee</i>	245
Development of Creativity	<i>M A Runco</i>	251
Knowledge Domains and Domain Learning	<i>L Maggioni and P A Alexander</i>	255
Problem Solving and Human Expertise	<i>T J Nokes, C D Schunn, and M T H Chi</i>	265
Problem Solving and Reasoning	<i>R E Mayer</i>	273
Reasoning, Moral and Social	<i>N S Jampol, C B Richardson, and M Killen</i>	279
Remembering as Social Activity	<i>D Middleton</i>	285

Learning and Cognition – Issues, Concepts, Types

Computer-Supported Collaborative Learning: Basic Concepts, Multiple Perspectives, and Emerging Trends	<i>S R Ludvigsen and A I Mørch</i>	290
Learning as Inquiry	<i>S R Goldman, J Radinsky, S Tozer, and D Wink</i>	297
Learning Through Narrative Socialization	<i>A Fasulo</i>	303
Neuroscience Bases of Learning	<i>M H Immordino-Yang and K W Fischer</i>	310
The Neuroscience of Reading	<i>U Goswami</i>	317

Learning and Cognition – Issues, Concepts, Types – Focus on Learning

Learning Strategies	<i>C E Weinstein, J Jung, and T W Acee</i>	323
Learning Through Play	<i>K Aronsson</i>	330
Situated View of Learning	<i>A Collins and J G Greeno</i>	335
The Neuroscience of Aging and Cognition	<i>W J Hoyer and B Bessette-Symons</i>	340

Learning and Cognition – Language, Literacy and Subject-Related

An Overview of Language and Literacy in Educational Settings	<i>V G Aukrust</i>	345
Bilingualism and Learning	<i>L S Verplaetse and E Schmitt</i>	355
Classroom Discourse and Student Learning	<i>L Hemphill</i>	361
Early Writing	<i>B E Hagtvet</i>	367

First Language Acquisition	<i>N B Ratner</i>	375
Foreign Language Learning	<i>C Kinginger</i>	382
Learning Science	<i>J Leach and P Scott</i>	387
Learning to Read	<i>M S Burns and J K Kidd</i>	394
Mathematics Learning	<i>L Verschaffel, B Greer, and E De Corte</i>	401
Music Learning	<i>S Hallam</i>	407
Reading Comprehension: Reading for Learning	<i>C E Snow</i>	413
Second Language Learning	<i>J S Arnfast, J N Jørgensen, and A Holmen</i>	419
Writing: Advanced	<i>R Andrews</i>	426

Learning and Cognition – Other Issues

Gender and Learning	<i>L Yates and J McLeod</i>	432
Parent–Child Relationships in Early Learning	<i>C P Edwards, S M Sheridan, and L L Knoche</i>	438
Peer Interaction and Learning	<i>S Blum-Kulka and S Dvir-Gvirsman</i>	444

Learning and Cognition – Places of Learning

Learning in a Cross-Cultural Perspective	<i>L-F Zhang and R J Sternberg</i>	450
Learning Outside of School	<i>T Nunes</i>	457
Reasoning and Explanatory Talk: Learning in Everyday Settings	<i>D E Beals</i>	464

Learning and Cognition – Theoretical Perspectives – Learning

Apprenticeship Approach to Learning	<i>K Nielsen</i>	469
Connectionism and Learning	<i>T R Shultz</i>	476
Constructivism and Learning	<i>S Sjøberg</i>	485
Cultural–Historical Activity Theory	<i>A Havnes</i>	491
Learning in a Sociocultural Perspective	<i>R Säljö</i>	498
Self-Regulated Learning and Socio-Cognitive Theory	<i>P H Winne and A F Hadwin</i>	503

N

National Systems of Education

Australia	<i>C N Power</i>	510
Austria	<i>M J Gaertner</i>	517
Bangladesh	<i>R Z Dewan</i>	524
Barbados	<i>R Rudder</i>	534
Bulgaria	<i>A Chavdarova and D Vesselinov</i>	540
England	<i>C Whatford</i>	550
Estonia	<i>M Keränen and J Kõrgesaar</i>	555
Fiji	<i>I Senibulu</i>	563
Finland	<i>J E Tuovinen</i>	569
France	<i>M Sellier</i>	576

Germany	<i>A Hüfner and K Hüfner</i>	582
Ghana	<i>R Palmer</i>	590
Greece	<i>K Toura and A Karagiannis</i>	599
Iceland	<i>M Hardardottir and M Gunnlaugsdóttir</i>	605
Indonesia	<i>A C Alwasilah and Furqon</i>	611
Israel	<i>A Volansky</i>	618
Jamaica	<i>E Hannam</i>	627
Latvia	<i>A Joma</i>	630
Lebanon	<i>N A Mneimneh</i>	636
Lithuania	<i>J Navickaitė, L Paurienė, R Zablackė, A Tijūnelytė, and V Kožemiakina</i>	656
Malawi	<i>F R Mkandawire and D Mulera</i>	666
Malta	<i>C Farrugia</i>	673
Mongolia	<i>B Yembuu</i>	681
Namibia	<i>J Ellis</i>	687
New Zealand	<i>R Baker and C Wylie</i>	692
Niger	<i>A A I Daouda</i>	699
Norway	<i>H Farstad</i>	709
Oman	<i>S A Issan</i>	718
Pakistan	<i>N Mahmood and R Malik</i>	727
Poland	<i>A Janowski</i>	737
Qatar	<i>S R Salman</i>	744
Russia	<i>V M Filippov and G F Tkach</i>	750
Rwanda	<i>E Bahizi</i>	758
Scotland	<i>D Raffe</i>	770
Singapore	<i>P T Ng</i>	776
Slovenia	<i>T Plevnik and A B Lakota</i>	782
Solomon Islands	<i>D P Malasa</i>	790
South Africa	<i>P Kgobe and J Pampallis</i>	799
Spain	<i>A Tiana</i>	806
Sri Lanka	<i>B M J Balasooriya, M M Wehella, and S U Wijeratne</i>	813
Suriname	<i>H R Ori</i>	826
Taiwan	<i>C-M Lee</i>	834
Tanzania	<i>R Wedgwood</i>	841
Thailand	<i>C Rukspollmuang</i>	847
The Gambia	<i>Y A-M Jobe</i>	858
The Philippines	<i>E A P Valenzuela</i>	865
Uruguay	<i>B Guinovart and C Troncoso</i>	873
USA	<i>B Pusser and M E R Lloyd</i>	888
Zanzibar	<i>I M Hija</i>	896

VOLUME 6**P****Philosophy of Education**

Philosophy of Education: Overview	<i>P Enslin</i>	1
-----------------------------------	-----------------	---

Philosophy of Education – Contemporary Issues

Neoliberalism, the Market and Performativity	<i>M A Peters</i>	11
Parents and Children's Rights	<i>M Clayton</i>	17
Peace Education	<i>I Gur-Ze'ev</i>	22
Philosophy and Educational Research	<i>D Bridges</i>	29
Race, Critical Race Theory and Whiteness	<i>B Applebaum</i>	36
School Choice and the Common School	<i>G Haydon</i>	44

Philosophy of Education – Philosophical Perspective

African Philosophy of Education	<i>K Horsthemke</i>	50
Feminism	<i>L Stone</i>	56
Hermeneutics	<i>C Bingham</i>	63
Islamic Education	<i>Y Waghid</i>	69
Liberalism and Education	<i>W Feinberg</i>	74
Phenomenology and Existentialism	<i>B Levering</i>	80
Pragmatism	<i>B Thayer-Bacon</i>	86
Psychoanalysis and Education	<i>P Smeyers</i>	92
Social and Cultural Capital in Education	<i>M Print</i>	97
The Aims of Education	<i>R Marples</i>	103
The Analytical Tradition	<i>D Carr</i>	108

Philosophy of Education – Philosophical Themes

Adult Learning: Philosophical Issues	<i>D Beckett</i>	114
Autonomy	<i>C A Wringe</i>	121
Bildung	<i>J Masschelein and N Ricken</i>	127
Critical Theory and Pedagogy	<i>M Mason</i>	133
Critical Thinking	<i>H Siegel</i>	141
Environmental Education	<i>M Bonnett</i>	146
Identity	<i>M S Merry</i>	152
Inclusion	<i>R Cigman</i>	158
Justice and Care	<i>M Tjiattas</i>	164
Lifelong Learning	<i>J Halliday</i>	170
Post-Structuralism and Education	<i>M Mason and M Clarke</i>	175
Reason and Rationality	<i>P Hogan</i>	183

Religious and Spiritual Education	<i>J C Conroy and S J McKinney</i>	189
The Capabilities Approach	<i>L Terzi</i>	194

Primary and Secondary Education – Learning and Teaching in School Age Education

Class Size – Arguments and Evidence	<i>P Blatchford and K C Lai</i>	200
Continuing Professional Development of Teachers	<i>P Earley</i>	207
Homeschooling	<i>P Farenga</i>	214
Innovations and Early Intervention in the Teaching of Literacy	<i>R E Slavin</i>	221
Personalized Learning in School Age Education	<i>D Hopkins</i>	227
Perspectives on Schooling in the Middle Years	<i>V J Russell</i>	233
Reform of the School Workforce	<i>H M Gunter and G Forrester</i>	242
Secondary Education in Developing Countries	<i>P Hughes and A Walker</i>	249

Primary and Secondary Education – Social Justice Issues in School Age Education

Community Focused Schools	<i>A Dyson</i>	255
---------------------------	----------------	-----

Primary and Secondary Education – The Effectiveness of School Systems

Reforming the American School System	<i>M Barber</i>	261
School Effectiveness in Developed Societies	<i>D Reynolds</i>	269
School to School Collaboration: Innovation and Improvement	<i>R Higham</i>	274
Studies of School Improvement in Developing Countries	<i>S E Anderson</i>	281
Sustainable Educational Reform	<i>A Hargreaves</i>	289
The Role of the District in Tri Level Reform	<i>M Fullan</i>	295
Whole School Designs for Enhancing Student Achievement	<i>S Stringfield and J Nunnery</i>	303

Q

Qualitative Research

Action Research in Education	<i>E Stringer</i>	311
Classroom Ethnography	<i>F Erickson</i>	320
Computer Assisted Qualitative Data Analysis	<i>C Silver and A Lewins</i>	326
Conversational Analysis	<i>H G Levine</i>	335
Critical Ethnography	<i>D Foley</i>	337
Critical Race Theory	<i>D Gillborn and G Ladson-Billings</i>	341
Critical Theory	<i>J J Ryoo and P McLaren</i>	348
Cultural Studies	<i>R Gaztambide-Fernández and K Cairns</i>	354
Cultural–Historical Activity Theory	<i>Laboratory for Comparative Human Cognition</i>	360
Design Experiments	<i>A Collins</i>	367
Discourse Analysis	<i>P Nikander and S Vehviläinen</i>	373
Document Analysis	<i>T Rapley and K N Jenkins</i>	380
Ethnography	<i>M Hammersley</i>	386

Ethnomethodology in Education Research	<i>D Macbeth</i>	392
Focus Groups	<i>D L Morgan and K Hoffman</i>	401
Grounded Theory	<i>K Charmaz and A Bryant</i>	406
Hermeneutics	<i>J Dyer</i>	413
Interpretive Research	<i>E Rauscher and M E Graue</i>	419
Interviews and Interviewing	<i>A Marvasti</i>	424
Life History	<i>J A Hatch and S K Newsom</i>	430
Narrative Inquiry	<i>D J Clandinin and J Huber</i>	436
Participant Observation	<i>J Schostak</i>	442
Phenomenology	<i>M van Manen and C A Adams</i>	449
Qualitative Case Studies	<i>S B Merriam</i>	456
Qualitative Data Management	<i>C Grbich</i>	463
Semiotics	<i>M Danesi</i>	475
Symbolic Interactionism, Naturalistic Inquiry, and Education	<i>J A Forte</i>	481
The Quality of Evidence in Qualitative Research	<i>S K Biklen</i>	488
Visual Data in Education Research	<i>J Wagner</i>	498

Qualitative Research – Standards of Proof

Validity: Mapping Diverse Perspectives	<i>G B Rossman, S F Rallis, and A M Kuntz</i>	505
--	---	-----

Quantitative and Qualitative Approaches to Research – Integration

Mixed Methods	<i>A Tashakkori and I Newman</i>	514
---------------	----------------------------------	-----

Quantitative Methods and Research Design

External Validity	<i>G E Matt, A Brewer, and M Sklar</i>	521
-------------------	--	-----

R

Role of NGOs in Globalization

NGOs and Globalization of Education	<i>M Ahmed</i>	529
-------------------------------------	----------------	-----

S

Social and Emotional Aspects of Learning

Achievement Goal Theory: Definitions, Correlates, and Unresolved Questions	<i>T Urdan</i>	537
Affect, Mood and Emotions	<i>S Schnall</i>	544
Anxiety	<i>M Zeidner</i>	549
Attribution Theory	<i>B Weiner</i>	558
Children's Friendship	<i>S R Asher and K L McDonald</i>	564
Coping with Stressful Situations: An Important Aspect of Self-Regulation	<i>M Boekaerts</i>	570
Culture in Motivation Research: A Challenging and Enriching Contribution	<i>K Kimmel and S Volet</i>	576
Early Social Development and Schooling	<i>G M van der Aalsvoort</i>	585

Emotion in Educational Contexts	<i>P A Schutz, P D Quijada, S de Vries, and M Lynde</i>	591
Empathy	<i>R Arnold</i>	597
Flow in Education	<i>J A Schmidt</i>	605
Interest	<i>M Ainley</i>	612
Intrinsic and Extrinsic Motivation	<i>C Levesque, K J Copeland, M D Pattie, and E L Deci</i>	618
Motivating Students in Classrooms	<i>J Brophy</i>	624
Motivation Regulation	<i>C A Wolters and S A Mueller</i>	631
Peer Learning in the Classroom	<i>N M Webb</i>	636
Poverty, Effects of on Social and Emotional Development	<i>N Philipsen Hetzner, A D Johnson, and J Brooks-Gunn</i>	643
Seeking Help as an Adaptive Response to Learning Difficulties: Person, Situation, and Developmental Influences	<i>S A Karabenick and R S Newman</i>	653
Self-Concept in Learning: Reciprocal effects model between academic self-concept and academic achievement	<i>H W Marsh and L F Scalas</i>	660
Self-Efficacy Beliefs	<i>D H Schunk and F Pajares</i>	668
Social and Emotional Outcomes of Learning	<i>J Vorhaus</i>	673
Social Aspects of Collaborative Learning	<i>M Arvaja and P Häkkinen</i>	685
Social Development and Schooling	<i>G W Ladd</i>	691
Social Interaction and Learning	<i>K Littleton</i>	698
Social Networks and the Education of Children and Youth	<i>B Schneider, T G Ford, and L Perez-Felkner</i>	705
Sociocultural Issues in Motivation	<i>R A Walker</i>	712
Socioemotional Aspects of Technology-Supported Learning	<i>S Järvelä</i>	718
Volitional Control of Learning	<i>J Husman and L Corno</i>	724
Wellbeing	<i>T Hascher</i>	732

VOLUME 7

Statistics

An Overview of Statistics in Education	<i>S Sinharay</i>	2
Analysis and Interpretation of Multivariate Data	<i>D J Bartholomew</i>	12
Analysis of Covariance	<i>N T Longford</i>	18
Analysis of Extreme Values in Education	<i>J Beirlant, G Dierckx, K Luwel, and P Onghena</i>	25
Analysis of Variance	<i>B M King</i>	32
Bayesian Statistical Analysis	<i>G Kokolakis</i>	37
Bootstrap Method	<i>K Singh and M Xie</i>	46
Canonical Correlation	<i>A Basu and A Mandal</i>	52
Categorical Data Analysis	<i>S E Fienberg and B W Junker</i>	58
Causal Inference	<i>D B Rubin</i>	66
Cluster Analysis: Overview	<i>S Landau and I Chis Ster</i>	72

Cognitive Psychology and Educational Statistics	<i>H S Kim and S Embretson</i>	84
Computational Statistics	<i>J E Gentle</i>	93
Continuous Probability Distributions	<i>S Sinharay</i>	98
Correspondence Analysis	<i>M Greenacre</i>	103
Data Mining	<i>R S J d Baker</i>	112
Decision Theory	<i>M Mendoza and E Gutiérrez-Peña</i>	119
Design of Experiments	<i>C A Albers and T R Kratochwill</i>	125
Discrete Probability Distributions	<i>S Sinharay</i>	132
Discrimination and Classification	<i>R C Neath and M S Johnson</i>	135
Empirical Bayes Methods	<i>E Ip and G Molenberghs</i>	142
Evaluation Research	<i>T Kellaghan</i>	150
Exploratory Data Analysis	<i>A Unwin</i>	156
Factor Analysis: An Overview and Some Contemporary Advances	<i>F A N Alhija</i>	162
Generalized Linear Mixed Models	<i>S Rabe-Hesketh and A Skrondal</i>	171
Generalized Linear Models	<i>J J Faraway</i>	178
Generating Random Numbers	<i>P J Pashley and A Amodeo</i>	184
Goodness-of-Fit Testing	<i>A Maydeu-Olivares and C García-Forero</i>	190
Graphical Models	<i>R G Almond</i>	197
Growth Modeling	<i>I Plewis</i>	203
Hierarchical Linear Models	<i>D Rindskopf</i>	210
Hypothesis Testing and Confidence Intervals	<i>T R Konold and X Fan</i>	216
Instrumental Variables	<i>E P Bettinger</i>	223
Jackknife Methods	<i>S Sinharay</i>	229
Large-sample Statistical Methods	<i>J M Singer and D F Andrade</i>	232
Latent Class Models	<i>J K Vermunt</i>	238
Markov Chain Monte Carlo	<i>A M Johansen</i>	245
Matrix Algebra	<i>H Yanai and H Ishii</i>	253
Measure of Association	<i>G-H Huang</i>	260
Measures of Central Tendency	<i>D Clark-Carter</i>	264
Measures of Dispersion, Skewness and Kurtosis	<i>D Rindskopf and M Shiyko</i>	267
Meta Analysis	<i>J Sánchez-Meca and F Marín-Martínez</i>	274
Missing Data	<i>C A W Glas</i>	283
Model Selection	<i>S J Haberman</i>	289
Monte Carlo Methods	<i>A M Johansen</i>	296
Multidimensional Scaling	<i>Z Zhang and Y Takane</i>	304
Multiple Comparisons	<i>C Lewis</i>	312
Multivariate Analysis of Variance	<i>N T Longford</i>	319
Multivariate Linear Regression	<i>S Ganesh</i>	324
Multivariate Normal Distribution	<i>J Tacq</i>	332

Nonlinear Regression Analysis	<i>H-H Huang, C K Hsiao, and S-Y Huang</i>	339
Nonparametric Statistical Methods	<i>K Sijtsma and W H M Emons</i>	347
Observational Studies	<i>P R Rosenbaum</i>	354
Order Statistics	<i>I Bairamov</i>	360
Point Estimation Methods with Applications to Item Response Theory Models	<i>F Bartolucci and L Scrucca</i>	366
Principal Components Analysis	<i>D J Bartholomew</i>	374
Probability Theory	<i>T Rudas</i>	378
Recursive Partitioning	<i>S Landau and S Barthel</i>	383
Sampling	<i>J Qian</i>	390
Sequential Probability Ratio Test	<i>T J H M Eggen</i>	396
Sequential Testing	<i>H J Vos</i>	401
Signal Detection Theory	<i>H Abdi</i>	407
Small Area Estimation	<i>A Noble</i>	411
Statistical Analysis of Functional Data	<i>F Ferraty, P Sarda, and P Vieu</i>	417
Statistical Inequalities	<i>H M Mahmoud and J L Gastwirth</i>	426
Statistical Paradoxes	<i>Y-K Tu</i>	429
Statistical Power Analysis	<i>L V Hedges and C Rhoads</i>	436
Statistical Significance Versus Effect Size	<i>X Fan and T R Konold</i>	444
Stochastic Processes	<i>F Baudoin</i>	451
Structural Equation Models	<i>S-Y Lee and X-Y Song</i>	453
Survival Data Analysis	<i>A F De Champlain</i>	459
The Normal Distribution and its Applications	<i>J Tacq</i>	467
Time Series Analysis	<i>P J Brockwell</i>	474
Univariate Linear Regression	<i>M L Hazelton</i>	482
Value-Added Models	<i>D F McCaffrey</i>	489

T

Teacher Education – Evaluation in the Teacher Education Program

Accreditation and Standards in Teacher Education	<i>R J Yinger and K L Daniel</i>	495
Determining Long Term Effects of Teacher Education	<i>C N Brouwer</i>	503

Teacher Education – Evaluation of the Prospective Teacher

The Evaluation of Prospective Teachers in Teacher Education	<i>D E H Tigelaar and J van Tartwijk</i>	511
---	--	-----

Teacher Education – Macro Factors

Diversity in Teacher Education	<i>T Wubbels</i>	518
Social and Cultural Influences on Teacher Education	<i>M Ben-Peretz and R Lotan</i>	525
The Role of Governance in Teacher Education	<i>G E Karlsen</i>	532

Teacher Education – Post-Initial Professional Development

Action Research as a Tool for Teachers' Professional Development	<i>P Ponte</i>	540
Contemporary Approaches to Teacher Professional Development	<i>H Borko, J Jacobs, and K Koellner</i>	548
School Development for Teacher Learning and Change	<i>P Sleegers and K Leithwood</i>	557
Teacher Induction	<i>D Beijaard, J Buitink, and C Kessels</i>	563
Teacher Learning as Workplace Learning	<i>J Imants and K van Veen</i>	569
Teacher Learning with Lesson Study	<i>A Murata</i>	575
Teachers Career Stages and Professional Development	<i>R Fessler and E Rice</i>	582

Teacher Education – Preservice Teacher Education

A Pedagogy of Teacher Education	<i>J Loughran</i>	587
Case Methods in Teacher Education	<i>M Sato and C Rogers</i>	592
Conceptions of Teacher Education	<i>J A Whitcomb</i>	598
Inquiry-Oriented Teacher Education	<i>R Schulz</i>	604
Narratives and Biography in Teacher Education	<i>G Kelchtermans</i>	610
Supervision in Teacher Education	<i>J McCarthy and L F Quinn</i>	615
Taking Prospective Teachers' Beliefs into Account in Teacher Education	<i>M Valcke, G Sang, I Rots, and R Hermans</i>	622
Teacher Education and Models of Teacher Reflection	<i>V K LaBoskey</i>	629
Teacher Education as Teaching for Understanding with New Technologies	<i>M S Wiske and D E Spicer</i>	635

Teacher Education – Preservice: The Knowledge Base

Experienced Teachers' Craft Knowledge	<i>P C Meijer</i>	642
Moral Values in Teacher Education	<i>W Veugelers</i>	650
Pedagogical Content Knowledge	<i>J H Van Driel and A Berry</i>	656
Teacher Education and the Educational Foundations Knowledge Base	<i>F B Murray</i>	662
The Relationship Between Theory and Practice in Teacher Education	<i>F A J Korthagen</i>	669

Teacher Education – Teacher Educators

Characteristics, Scholarship and Research of Teacher Educators	<i>M Lunenberg</i>	676
Professional Development of Teacher Educators	<i>P K Smith</i>	681
Self-Study by Teacher Educators	<i>T Russell</i>	689

Teacher Education – Teaching Specific Domains

Arts Teacher Education	<i>L Beudert</i>	695
Mathematics Teacher Education	<i>K Krainer and S Llinares</i>	702
Music Teacher Education	<i>M W-y Wong</i>	706
Science Teacher Education	<i>J H Van Driel and S K Abell</i>	712
Social Studies Teacher Education	<i>A-L Halvorsen and S M Wilson</i>	719
Teacher Education for Elementary Education	<i>K D Brown and G N Carnes</i>	725

Teacher Education for Teaching the Gifted	<i>A Y Baldwin</i>	732
Teacher Preparation in Mother Tongue Education	<i>M Kooy and H Hulshof</i>	738
Vocational Teacher Education	<i>S Johansson</i>	744

Teacher Education – The Cooperating School

Mentoring in Teacher Education	<i>P D Tomlinson, A J Hobson, and A Malderez</i>	749
Partnerships Between Schools and Higher Education	<i>C Day and L Smethem</i>	757

VOLUME 8

Technology and Learning

An Overview of Technology and Learning	<i>B Means and J Roschelle</i>	1
Educational Data Modeling	<i>S Iksal</i>	11
Student Data Systems and Their Use for Educational Improvement	<i>J C Wayman, V Cho, and M P Richards</i>	14
Technology and Learning: Access in Schools Around the World	<i>R Anderson and J Ainley</i>	21

Technology and Learning – as a Subject of Instruction

Conceptions of Technology Literacy and Fluency	<i>O Erstad</i>	34
--	-----------------	----

Technology and Learning – Assessment

Technology and Formative Assessment	<i>J W Pellegrino</i>	42
Technology for Large-Scale Assessment	<i>R E Bennett</i>	48
Technology Supports for Assessment Design	<i>R J Mislevy, F I Winters, I I Bejar, R E Bennett, and G D Haertel</i>	56

Technology and Learning – Education Reform and Economic Development

National Strategies to Build a Technology Workforce	<i>H-M Cheah</i>	66
Public–Private Partnerships for Educational Reform	<i>V Quah, C P Lim, and C Brook</i>	73
Relating Technology, Education Reform and Economic Development	<i>R B Kozma</i>	81

Technology and Learning – Evaluating Technology

Examining the Effects of Educational Technology Programs: Challenges and Strategies	<i>M Russell</i>	88
---	------------------	----

Technology and Learning – Global Trends

Equity in Technology Access and Opportunities	<i>T Matuchniak and M Warschauer</i>	95
Internet-Based Education	<i>M Bakia</i>	102

Technology and Learning – Issues

Media Use and School Achievement	<i>K Roe</i>	109
Technology and Physical and Social Health	<i>K Subrahmanyam</i>	112

Technology and Learning – Meeting Special Students’ Needs

Universal Design for Learning *D H Rose and J W Gravel* 119

Technology and Learning – School Leadership for Technology Integration

Technology as a Support for School–Community Connections *K M Culp* 125

Technology and Learning – Supports for Managing Instruction

Classroom Uses of Technology to Manage Instruction *W R Penuel* 133

Technology and Learning – Supports for Skill Learning

New Media, Learning from *W Schnotz and H Horz* 140

Reading and Technology *N DeFrance, D Khasnabis, and A S Palincsar* 150

Technological Supports for Acquiring Twenty-First-Century Skills *C Dede* 158

Technology Supports for Acquiring Inquiry Skills *T de Jong* 167

Technology Supports for Acquiring Mathematics *M J Nathan* 172

Technology Supports for Lifelong Learning *A Collins and R Halverson* 184

Technology and Learning – Supports for Student Collaboration

Knowledge Communities in the Classroom *J D Slotta and H Najafi* 189

Technology and Learning – Supports for Subject Matter Learning

Technology Supports for Science Learning *C Delgado and J Krajcik* 197

Technology Supports for Second Language Learning *D Nunan* 204

Technology and Learning – Teacher Learning

Teacher Skills and Knowledge for Technology Integration *N Law* 211

Technology in Preservice Teacher Education *N Davis* 217

Technology Resources for Teacher Learning *J E Hinojosa, C Labbé, and L López* 222

V

Vocational Education and Training

An Overview of Vocational Education and Training *T Karmel* 229

Informal and Non-Formal Learning in Vocational Education and Training *F Beddie and S Halliday-Wynes* 240

School-Based Vocational Education and Training *B Knight and R Sweet* 247

The Economics of Vocational Education and Training *P Carneiro, L Dearden, and A Vignoles* 255

Vocational Education and Training – Individuals and Students

Age Equality in Education and Training *P Taylor* 262

Career Guidance in Vocational Education and Training and in the Workplace *P Plant* 269

Displaced Workers, Unemployed and Vocational Education and Training *B Kriechel* 275

Immigrant Investment in Host Country Vocational Education and Training *I Kogan* 280

Indigenous People and Vocational Education and Training	<i>R G Schwab</i>	289
Persons with a Disability and Vocational Education and Training	<i>F Ferrier and C S Smith</i>	294
Prisoner Education and Training	<i>S Dawe, S H Wynes, and J Misko</i>	300
Vocational Education and Training and the School-to-Work Transition	<i>G Hayward</i>	306

Vocational Education and Training – Industry and Employers

Apprenticeships	<i>E Smith</i>	312
Skills Shortages: Concepts, Measurement and Policy Responses	<i>C Shah and G Burke</i>	320
Training and Learning in the Workplace	<i>R L Jacobs</i>	328
Training in Small Businesses	<i>H Matlay</i>	337

Vocational Education and Training – Research on VET

Language of Vocational Education and Training	<i>J Misko, J MacKenzie, and R-A Polvere</i>	343
Research on Vocational Education and Training	<i>J Robson and B Bailey</i>	352
Vocational Education and Training Information Resources	<i>L-A Harris, R Naidu, and M Saunders</i>	358

Vocational Education and Training – Teaching and Learning

Employability Skills	<i>E Smith</i>	368
Learning Styles in Vocational Education and Training	<i>J Vorhaus</i>	376
Modularization in Vocational Education and Training	<i>H Ertl and G Hayward</i>	383
Personalized Learning and Vocational Education and Training	<i>I Duckett</i>	391
Recognition of Prior Learning and Vocational Education and Training	<i>J Keating</i>	397
The Competency Model	<i>A Gonczi and P Hager</i>	403

Vocational Education and Training – VET in Context

Globalization and Vocational Education and Training	<i>P Loveder</i>	411
Learning to Work for the Future	<i>J Fien, R Maclean, and D N Wilson</i>	416
The Relationship between Technical and Vocational Education and Training and Innovation	<i>K Ruth and L Deitmer</i>	423
The Status of Vocational Education and Training	<i>J Lasonen</i>	429
Vocational Education and Training in Post-Conflict Countries	<i>E Lyby</i>	441

Vocational Education and Training – VET System

Dual System	<i>T Deissinger</i>	448
Industry Involvement in the Vocational Education and Training System	<i>M Hughes and C Hughes</i>	455
Non-Completion in Vocational Education and Training: Issues, Patterns and Possibilities	<i>S Lamb and A Walstab</i>	463
Planning and Policy Development for Technical Vocational Education and Training Systems	<i>R Maclean and M Pavlova</i>	469
Qualifications Frameworks and their Role in the Reform of Education and Training	<i>M Young</i>	476
Quality Assurance in Vocational Education and Training	<i>D Gunning</i>	482

Training Finance	<i>A Ziderman</i>	489
Training Markets	<i>P Ryan</i>	496
Vocational Education and Training Teacher Training	<i>S Nielsen</i>	503
Vocational Education and Training Workforce	<i>D L Wallin</i>	513
Subject Index		519
Contributors		747

STATISTICS

Contents

An Overview of Statistics in Education
Analysis and Interpretation of Multivariate Data
Analysis of Covariance
Analysis of Extreme Values in Education
Analysis of Variance
Bayesian Statistical Analysis
Bootstrap Method
Canonical Correlation
Categorical Data Analysis
Causal Inference
Cluster Analysis: Overview
Cognitive Psychology and Educational Statistics
Computational Statistics
Continuous Probability Distributions
Correspondence Analysis
Data Mining
Decision Theory
Design of Experiments
Discrete Probability Distributions
Discrimination and Classification
Empirical Bayes Methods
Evaluation Research
Exploratory Data Analysis
Factor Analysis: An Overview and Some Contemporary Advances
Generalized Linear Mixed Models
Generalized Linear Models
Generating Random Numbers
Goodness-of-Fit Testing
Graphical Models
Growth Modeling
Hierarchical Linear Models
Hypothesis Testing and Confidence Intervals
Instrumental Variables
Jackknife Methods
Large-sample Statistical Methods
Latent Class Models
Markov Chain Monte Carlo
Matrix Algebra
Measure of Association
Measures of Central Tendency
Measures of Dispersion, Skewness and Kurtosis
Meta Analysis
Missing Data
Model Selection
Monte Carlo Methods
Multidimensional Scaling
Multiple Comparisons
Multivariate Analysis of Variance
Multivariate Linear Regression

Multivariate Normal Distribution
Nonlinear Regression Analysis
Nonparametric Statistical Methods
Observational Studies
Order Statistics
Point Estimation Methods with Applications to Item Response Theory Models
Principal Components Analysis
Probability Theory
Recursive Partitioning
Sampling
Sequential Probability Ratio Test
Sequential Testing
Signal Detection Theory
Small Area Estimation
Statistical Analysis of Functional Data
Statistical Inequalities
Statistical Paradoxes
Statistical Power Analysis
Statistical Significance Versus Effect Size
Stochastic Processes
Structural Equation Models
Survival Data Analysis
The Normal Distribution and its Applications
Time Series Analysis
Univariate Linear Regression
Value-Added Models

An Overview of Statistics in Education

S Sinharay, ETS, Princeton, NJ, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This article intends to provide an overview of the application of statistics to the field of education. Statistics is a vast field, with new topics such as proteomics, ensemble sampling, and statistics in ophthalmology being introduced every now and then. The Statistics section of the encyclopedia did not attempt to provide a summary of all possible topics under the subject. Instead, this section focuses on the topics in statistics that have found applications in the field of education. Most applications of statistics to education are found in the area of educational measurement for the simple reason that statistics, the science that deals with quantitative analysis of data, inherently is related to measurement. (Several such applications – such as automated scoring, differential item functioning, generalizability theory, and item response theory (IRT) – are covered in the educational measurement section of the encyclopedia and will not be repeated here.) Hence, one could argue that some of

the topics included in the statistics section, or, that several examples included in the articles in this section, deal mostly with educational measurement and should have belonged in the Educational measurement section of this encyclopedia. In addition, some of the topics may overlap to a small extent with other topics in this section (e.g., generalized linear models are tools in categorical data analysis, but this section has two separate articles on the two topics). In any case, the number of applications of statistics to education is on the rise. This is because of (1) increases in computing power that have led people to ask questions that could not have been answered 20 years ago in a timely manner, and, (2) increases in the number of educational tests, partially due to the No Child Left Behind (NCLB) Act of 2001 in the USA that requires annual testing in the schools and produces a lot of data. Practitioners in education should find this section, together with the Educational measurement section, helpful as these two provide a comprehensive overview of the statistical methods used in education.

Next, several topics in statistics that are relevant to the field of education are described in brief along with, for some topics, examples of actual applications to the field of education. All of these topics and many more topics are covered in one or more articles in the statistics section.

Exploratory Data Analysis

After data are collected, often the first step in analyzing the data is exploratory data analysis (EDA), which consists of looking at data to see what they seem to say (Tukey, 1977) while relying on simple arithmetic and easy-to-draw pictures or plots. The techniques used in EDA include the following:

- Plotting the data in bar charts, pie charts, histograms, Youden plots, etc.,
- Plotting simple statistics in plots such as mean plots, standard deviation (SD) plots, box plots, etc., and
- Positioning such plots so as to extract the maximum information possible from them.

Consider **Figure 1**, which shows responses of 325 examinees to 15 items regarding mixed-number subtraction (Tatsuoka, 1984). An example item is $4\frac{5}{7} - 1\frac{4}{7}$. The items, sorted according to decreasing proportion correct (i.e., increasing difficulty), are shown along the x -axis in the figure; the examinees, sorted according to increasing raw scores, are shown along the y -axis. A short black horizontal line for an examinee and an item indicates a correct response. Some patterns are immediately visible from the figure. For example, several examinees (24 out of 325) at the top answer all items correctly (clear from the top of the plot being completely black). The examinees with the lowest scores could answer only two items correctly. Interestingly, these items ($\frac{3}{4} - \frac{3}{4}$ and $3\frac{7}{8} - 2$) are not the two easiest items and can be solved without any knowledge of mixed-number subtraction. Further, the lower half of the examinees rarely answered any difficult items correctly.

Wainer (2000, 2005) provided several applications of EDA to education.

Simple Summary Measures

It is often important to examine simple summary measures such as the mean and standard deviation (SD) of numerical information. For example, the *Digest of Education Statistics*, an annual publication of the National Center for Education Statistics (NCES) in the United States of America, includes the number of schools and colleges, teachers, enrolments, and graduates, in addition to other information on education. Further, the 2007 *Digest* reports that the average salary for teachers in 2005–06 was US\$49 109, about 1% higher than in 1995–96, after adjustment for inflation, and that the average reading scores at ages

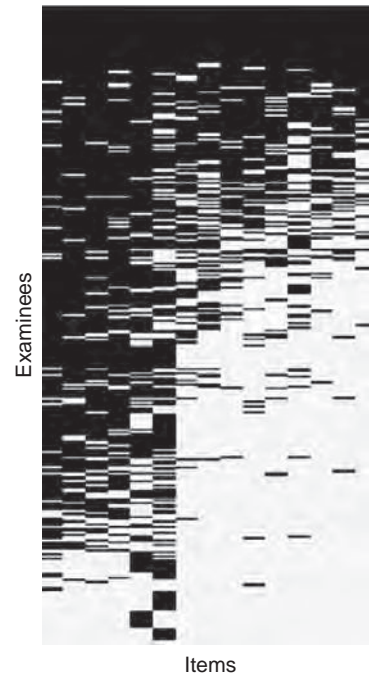


Figure 1 A plot of the responses of 325 examinees to 15 mixed number subtraction items.

9 and 13 were higher in 2004 than in 1971 and the average score for 17-year-olds in 2004 was similar to that in 1971. Measures of central tendency and measures of dispersion, skewness, and kurtosis are discussed in the statistics section of the encyclopedia.

Measures of Association

In education, it is often of interest to examine the amount of association between a group of variables. For example, test administrators administering several tests simultaneously to students will like that scores on different tests (e.g., in reading, writing, mathematics, and science) do not correlate highly with each other. High correlations between such scores may raise questions about the need of so many tests. Choice of the appropriate measures of association depends on whether the variables of interest are continuous, discrete ordinal, or discrete nominal. For example, if the scores on the abovementioned tests are given on a scale of 1 to 100 in 1-point increments, a correlation coefficient may be the appropriate measure of association. On the other hand, if scores are given as 0 or 1 (where a score of 1 means that a student is good on a subject and 0 otherwise), an odds ratio or the Kendall's tau (see, e.g., Agresti, 2002) may be the appropriate measure. It is often of interest to examine the association between two groups of variables, $\mathbf{X} = (X_1, X_2, \dots, X_k)$ and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_l)$. Canonical correlation analysis attempts to find linear combinations of the two groups with high correlations.

Probability Theory

Providing a mathematical description of our beliefs about the systematic properties of a random phenomenon is the first step in several statistical analyses. Usually this is accomplished with the help of probability theory, the branch of mathematics that describes the pattern of chance outcomes. A key idea in probability theory is that of a random variable, which is a variable whose value is a numerical outcome of a random phenomenon, and its distribution. There are several types of random variables, and the articles in the statistics section, on discrete and continuous probability distributions, provide detailed descriptions of them. A knowledge of random variables that follow a normal distribution is essential in several applications of statistics to education (e.g., the distribution plays a key role in topics such as factor analysis, analysis of variance, linear regression, and multilevel models) – hence two articles in the statistics section are devoted to these variables. Stochastic processes are sequences of random variables and are often of interest in probability theory (e.g., the path traced by a molecule as it travels in a liquid or a gas can be modeled using a stochastic process). Order statistics is another area of some importance in probability theory. For example, the ordered scores on a test of a sample of students represent the order statistics for the test score variable for the sample.

Statistical Inference

Statistical inference consists in the use of statistics to draw conclusions about some unknown aspect of a population based on a random sample from that population. Some preliminary conclusions may be drawn by the use of EDA or by the computation of summary statistics as well, but formal statistical inference uses calculations based on probability theory to substantiate those conclusions. Statistical inference can be divided into two areas: estimation and hypothesis testing. In estimation, the goal is to describe an unknown aspect of a population, for example, the average scholastic aptitude test (SAT) writing score of all examinees in the State of California in the USA. Estimation can be of two types, point estimation and interval estimation, depending on the goal of the application. The goal of hypothesis testing is to decide which of two complementary statements about a population is true. Two such complementary statements may be: (1) the students of California score higher on an average on SAT writing than the students of Texas, and (2) the students of California score lower on an average on SAT writing than the students of Texas. Point estimation is discussed in the statistics section of the encyclopedia. Details on interval estimation and hypothesis testing, and power analysis, which play a key role in hypothesis testing are also discussed in the statistics section of the encyclopedia. Often,

an investigator has to perform several hypothesis tests simultaneously. For example, one may want to compare the SAT critical reading scores of several pairs of schools belonging to a geographical region. The article on multiple comparison in the statistics section of the encyclopedia, discusses how to handle such a situation in an appropriate manner.

Principal-Components Analysis

Principal-component analysis (PCA) is often used to reduce multidimensional data sets to a lower number of dimensions for analysis. PCA retains those characteristics of the data set that contribute most to its variance, by keeping lower-order principal components (the ones that explain a large part of the variance present in the data) and ignoring higher-order ones (that do not explain much of the variance present in the data). Such low-order components often contain the most important aspects of the data. For example, in the National Assessment of Educational Progress (NAEP), a large-scale educational survey conducted in the USA (see, e.g., von Davier *et al.*, 2007), values of several hundred background variables are collected for each examinee. As it is important to use the information contained in these background variables in the statistical model used in NAEP, several principal components are computed from the original background variables and are used, instead of the original variable values, in the final NAEP statistical model (see, e.g., Jenkins *et al.*, 2001: 377–378).

Factor Analysis

The beginning of factor analysis lies in the early attempts of Karl Pearson, Charles Spearman, and others to define and measure intelligence – hence factor analysis has been a popular tool in education. When several tests are administered to a group of examinees, one aspect of validation may involve determining whether there are a few underlying abilities or skill variables that govern the examinees' performances on the tests. Factor analysis, which is often used in such situations, attempts to describe the covariance among several variables in terms of a few underlying, but unobservable, random variables called factors. For example, correlations from the group of test scores in classics, French, English, mathematics, and music collected by Spearman suggested an underlying intelligence factor (Johnson and Wichern, 1998). Factor analysis can be considered to be an extension of PCA – both of these techniques approximate the covariance matrix among the variables. A factor analysis is exploratory if the investigator does not have a hypothesis about the number of factors measured by the tests, and confirmatory if the investigator has such hypotheses and conducts statistical tests of them.

Structural-Equation Modeling

Structural-equation modeling is an extension of factor analysis and is a methodology designed primarily to test substantive theory from empirical data. For example, a theory may suggest that certain mental traits do not affect other traits and that certain variables do not load on certain factors, and that structural equation modeling can be used to test the theory. (A mental trait is a habitual pattern of behavior, thought and emotion.) A structural-equation model (SEM) is a system of linear equations among several unobservable variables (constructs) and observed variables. An SEM is composed of two parts: a structural part, linking the constructs to each other (usually, this part expresses the endogenous or dependant constructs as linear functions of the exogenous or independent constructs), and a measurement part, linking the constructs to observed measurements. The second part resembles a confirmatory factor analysis model. The SEMs can be displayed in visual form – these displays are called path diagrams. The full model is then estimated from a data set and inferences drawn.

Classification and Discriminant Analysis

The goal of classification and discriminant analysis are to describe, either graphically or algebraically, the different features of observations from several known groups, and to sort new objects (whose group membership is unknown) into the groups. For example, consider a medical school applications office that has the test scores and other college records of several students who became MDs and several others who did not. When a new application comes, the office may want to classify the applicant into likely to become MD and unlikely based on the test scores and college records.

Cluster Analysis

Cluster analysis is a technique to group similar observations into a number of clusters based on the observed values of several variables for each individual. Cluster analysis is similar in concept to discriminant analysis. The group membership of a sample of observations is known upfront in the latter while it is not known for any observation in the former. As an application of cluster analysis to education, Everitt (1990) describes a data set that has achievement test scores on reading and arithmetic for children in the fourth and sixth grades of 25 schools and the interest is in identifying different levels of performance and assessing similarities and differences in the patterns of change from fourth to sixth grade – cluster analysis is the most appropriate technique for the example.

Multidimensional Scaling

Multidimensional scaling is related to cluster analysis and assigns a location of each sample observation in a low-dimensional space so that their distances are close to their actual distances in multiple dimensions. This idea can be illustrated by reference to map construction. Suppose that one is given a table with the distances between several American cities (and not given the exact locations of them on a map). One could attempt to place the cities on a map so that the distances between the cities on the map are as close as possible to the distances on the table. Once this step is done, the map is used to draw conclusions on the sample; often, a cluster analysis is performed on these low-dimensional representations of the observations.

Categorical Data Analysis

A categorical variable has a measurement scale that consists of a set of categories. For example, the response of a student to a test question is often measured as correct or incorrect.

Categorical variables have two types of scales – nominal (when the categories do not have a natural ordering; e.g., gender and religion) and ordinal (when the categories have a natural ordering; e.g., patient condition being measured as good, fair, and serious). Several statistical methods and concepts have been devised for categorical data. For example, distributions such as the binomial and Poisson distributions, models such as generalized linear models, generalized linear mixed models, logistic regression model, and log-linear models, and methods such as analysis of contingency tables were devised to handle categorical data.

Analysis of Variance, Analysis of Covariance, and Multivariate Analysis of Variance

Analysis of variance (ANOVA) is the statistical procedure of comparing the means of a variable across several groups of individuals. For example, ANOVA may be used to compare the average SAT critical reading scores of several schools. The name of the technique arises from the fact that the first step in an ANOVA is to partition the variance present in the observations into several components. The ANOVA method was the second most frequently used data-analysis procedure in a survey of articles published between 1971 and 1998 in three reputed educational-research journals (Hsu, 2005). Generalizability theory (Cronbach *et al.*, 1963), which is a competitor to the classical theory of reliability of tests, usually applies ANOVA procedures to test scores.

Analysis of covariance (ANCOVA) is used when, like in ANOVA, the interest is in comparing several means, but the investigator also has the values of an additional variable that influences the variable of interest. For example,

ANCOVA may be used to compare the average SAT critical reading scores of several schools where the preliminary scholastic aptitude test/national merit scholarship qualifying test (PSAT/NMSQT) critical reading score of each examinee is available in addition to the SAT critical reading score. (The PSAT/NMSQT is supposed to provide firsthand practice for the SAT.)

Multivariate analysis of variance (MANOVA) is used to compare means of several variables simultaneously across several groups of individuals. For example, one could apply MANOVA to simultaneously compare the average scores on several subjects across several schools. Longford (1990) provides such an example.

Design of Experiments

An experiment is a test in which purposeful changes are made to the input variables of a process or system so that one may observe and identify the reasons for changes that may be observed in the output response. Design of experiments is the science of planning and conducting experiments and analyzing the resulting data so that valid and objective conclusions can be drawn. For example, the education ministry of a country may be interested in conducting an experiment to find out if a particular style of teaching mathematics helps children of fourth grade to learn the subject better than the existing style. In designing an experiment, the ministry has to make sure that any difference that they might observe in the outcome for students who were taught using the new style and those who were taught using the existing style cannot be attributed to a factor other than the teaching style (e.g., if they assign all students from rural areas to the new style and all students from urban areas to the existing style, then a difference can be attributed to the rural vs. urban difference). There are three basic principles in design of experiments – randomization (which means that the assignment of the experimental material and the order in which the individuals receive the experimental material are randomly determined), replication (which refers to repeats of each experimental condition), and blocking (which is the grouping of individuals to create several homogeneous groups before assigning the experimental material).

Observational Studies

In some situations, it is not possible (for reasons such as budget constraints and ethical issues) to design an experiment to answer a question or to test a hypothesis. For example, consider that the interest is in finding whether watching television is affecting the class grades of students. As watching television may have adverse effects, it will be unethical for one to design an experiment and randomly assign students to watch television for different number of hours per day. In such situations, often the only way is to conduct an observational study, that is, to

collect data on a sample of individuals, and apply an appropriate method to draw conclusions. In the example above, one could record the number of hours of television watched by the students and examine its association with their grades. Only limited number of conclusions can be drawn from an observational study. Any observed difference or association has several reasonable alternative explanations. For example, an observed lower score of students watching television longer can be caused by such students having less appropriate home and school inputs, such as fewer books at home or parents who read less to them. Huang and Lee (2009) investigated whether television watching at ages 6–7 and 8–9 affects cognitive development measured by math and reading scores at age 8–9 using a rich childhood longitudinal sample.

Causal Inference and Instrumental Variables

Suppose that an investigator is interested in testing a hypothesis, for example, about the comparison of a new educational program versus the existing program. Whether the investigator performs a randomized experiment or an observational study, he is faced with the question of how to draw inferences about the causal effects of the new program. In other words, if there is a performance difference between the students who were administered the new educational program and those who were administered the existing program, the investigator would like an answer to the question “Is the difference in performance of the two groups caused by the difference in the educational program?” An article in the statistics section discusses how causal inferences can be made. Instrumental variables (IVs) are used to estimate causal relationships when controlled experiments are not feasible. An overview of instrumental variables and of their possible applications to education is discussed in the statistics section of the encyclopedia.

Sampling

There is a growing importance of survey information on individuals, households, institutions, businesses, and environmental resources. Typically, one wants to gather information on a large group of individuals. However, time and cost usually does not allow obtaining information from each individual in the group. In such cases, one usually gathers information on only a sample, which is a small part of the large group. Sampling plays an essential role in drawing conclusions about the large group (which is called the population) from the information contained in the sample. An example of an application of sampling is the National Assessment of Educational Progress (NAEP), an educational sampling survey (Allen *et al.*, 2001). NAEP is the only ongoing measure of what students in the USA know and can do in a variety of subject areas and it reports scores for different demographic groups based on

gender, ethnicity, school type, school location, etc. NAEP draws a sample of students that is representative of the whole student population, applies several statistical techniques, and draws conclusions (an example of a conclusion is that between 1992 and 2000, the percentage of fourth-graders at or above the proficient-achievement level in reading increased by a small, but statistically significant amount) on the whole population based on information contained in the sample.

Bayesian and Empirical Bayes Methods

In a traditional or frequentist statistical analysis, the parameter of a probability model is considered an unknown but nonrandom quantity and only the information contained in the observed data is relevant for any inference. On the contrary, a Bayesian analysis (see, for e.g., Gelman *et al.*, 2003) assumes that the parameter is a random variable with a certain probability distribution, referred to as the prior distribution. The prior distribution quantifies the experimenter's beliefs about the parameter before observing the data. The next step in a Bayesian approach is to update the prior distribution on the basis of the likelihood function of the observed data through Bayes' theorem (Bayes, 1763). The resulting distribution is referred to as the posterior distribution of the parameter and summarizes the information in both the prior distribution and in the data. The influence of the prior distribution on the posterior distribution becomes weaker as the size of the observed data sample increases. The variation of the Bayesian methods in which the parameters of the prior distribution are estimated from the observed data is called empirical Bayes methods. Sinharay (2006) provided a review of the applications of Bayesian methods to educational measurement. Novick and Jackson (1974) included several applications of Bayesian methods to educational measurement. Other examples of applications of Bayesian methods to education are Rubin (1983), who applied Bayesian methods to three problems in educational measurement, Zwick *et al.* (1999), who applied an empirical Bayes method to differential item functioning, and Sinharay (2005), who applied Bayesian model-checking methods to assess the goodness of fit of IRT models. Further details on Bayesian methods and empirical Bayes methods (see, for example, Carlin and Louis, 1996) are discussed in the statistics section of the encyclopedia. Decision theory, which is a Bayesian approach, is concerned with identifying the values, uncertainties, and other issues relevant in a given decision and the resulting optimal decision. An article in the statistics section provides more details on decision theory.

Resampling Methods

Resampling methods (see, e.g., Efron, 1982) draw samples from the observed data to draw certain conclusions about

the population of interest. Two of the most popular resampling methods are the jackknife and bootstrap. Both of these are examples of nonparametric statistical methods.

Jackknife is used in statistical inference to estimate the bias and standard error of a test statistic. The basic idea behind jackknife lies in systematically recomputing the statistic a large number of times, leaving out one observation or a group of observations at a time from the sample. Estimates of the bias and variance of the statistic can be calculated from this set of jackknife replications of the statistic. The jackknife finds several applications in complex sampling schemes, such as multistage sampling with varying sampling weights – an example of such application is NAEP, where the jackknife method is employed to compute standard errors of estimates.

Bootstrap is a statistical method for estimating the sampling distribution of an estimator by sampling with replacement from the original sample, most often with the purpose of deriving robust estimates of standard errors and confidence intervals of a population parameter like a mean, median, and correlation coefficient. It is often used as a robust alternative to procedures based on parametric assumptions, especially when those assumptions are in doubt, or where parametric inference is impossible or requires very complicated formulas for the calculation of standard errors. See, for example, Hanson *et al.* (1993), who applied the bootstrap method to compute the standard error of an equating method.

Nonparametric Inference

Nonparametric methods, or distribution-free methods, are statistical methods that do not rely on assumptions that the data are drawn from a given probability distribution. Nonparametric methods are often applied when less is known about the data (so that a probability distribution cannot be assumed). Due to the reliance on fewer assumptions, nonparametric methods are more robust (i.e., less vulnerable to violations of assumptions). They are also often applied because of their simplicity. Examples of nonparametric methods are Pearson's χ^2 test for assessing independence in a contingency table, jackknife and bootstrap methods for estimating the bias and variance of an estimator, the Wilcoxon Mann–Whitney rank-sum test, the permutation test, the Kolmogorov–Smirnov test for assessing whether two distributions are the same, and spline regression for estimating regression curves of a dependent variable on several independent variables.

Multiple Linear Regression Models

Multiple linear regression models have been extensively used in education (see, e.g., Hsu, 2005). Interestingly, the name regression, borrowed from the title of the first article on this subject (Galton, 1885), does not reflect

either the importance or breadth of application of this method. Multiple regression is the statistical procedure to predict the values of a response (dependent) variable from a collection of predictor (independent) variable values. For example, if scores on multiple predictors and one criterion are available, multiple regression may be used to develop a single equation to predict criterion performance from the set of predictors. Several applications of multiple regression models can be found in the prediction of first-year grade-point average in college from the SAT scores and high school grade-point average (see, e.g., Kobrin *et al.*, 2008). Multiple regression and multivariate multiple regression, the case when there are more than one dependent variables of interest and the interest is in predicting them simultaneously from a set of predictor variables, are discussed in the statistics section of the encyclopedia.

Hierarchical Linear Models and Growth Models

In an application of linear regression, the observations are assumed to be independent. When the assumption of independence is likely to be violated, for example, in an application in which one has data on several students who belong to a few schools (so that the responses of the students within each school are dependent), a popular option is to employ hierarchical linear models (HLMs). These models are also referred to as multilevel models and random-effects regression models. The students constitute the lower level while the schools constitute the higher level in the example. Note that HLMs can also be applied to repeated measures design or longitudinal studies, where individuals are followed and their responses recorded several times over a certain period of time; the repeated measures constitute the lower level and the individuals constitute the higher level; examples of such models are growth models, where, for example, the investigator measures the cognitive growth of students by giving them several tests over a certain period of time. Growth models are increasingly popular in the US due to the NCLB Act of 2001 that puts special emphasis on the cognitive growth of students. For example, in December 2007, the US Secretary of Education Margaret Spellings invited all eligible US states to submit a growth model proposal for the 2007–08 school year. Growth models and HLMs are discussed in the statistics section of the encyclopedia.

Value-Added Models

The NCLB Act of 2001 in the US requires states to ensure that there are quality teachers in every classroom, with quality defined in terms of traditional criteria such as academic training and fully meeting the state's licensure

requirements (Braun, 2005). Interestingly, in this respect, some states have taken the lead by seeking a quantitative evaluation of teachers based on an analysis of the test-score gains of their students. Such evaluations employ a class of models called value-added models (VAMs). These models require data that track individual students' academic growth over several years in different subjects in order to estimate the contributions that teachers make to that growth. Thus, VAMs can be viewed as a special case of growth models and, hence, of HLMs. Given their current state of development, VAMs can be used to identify a group of teachers who may reasonably be assumed to require targeted professional development. These are the teachers with the lowest estimates of relative effectiveness. Despite the enthusiasm these models have generated among many policymakers, several technical reviews of VAMs have revealed a number of serious concerns and it is important that such concerns be properly addressed before VAMs are used to make important decisions.

Generalized Linear Models and Generalized Linear Mixed Models

Linear regression models apply when the response variable can be assumed to be a continuous variable or to be normally distributed. However, in several applications in education, the response does not belong to either of those types. Suppose the interest is in finding out how the socioeconomic status and average parents' education for a class of students affects their performance on a test. If we have the scores on the test for each student, we can employ a linear regression model regressing the test scores on the socioeconomic status and average parent education. However, if we do not have the scores, but only know who passed the test and who did not (which is a binary response), we cannot employ linear regression. Generalized linear models (GLMs) can be used in situations like this. GLMs are extensions of the linear regression model to a wider class of response type such as binary or count data. A GLM requires the specification of two defining characteristics – the distribution of the response and the link function that describes how the mean of the response is linked to a linear combination of the predictors. Generalized linear mixed models (GLMM) are extensions of GLMs to the case when the individuals are clustered (e.g., students belonging to different schools). The statistics section of the encyclopedia includes two articles, one each on GLM and GLMM.

Nonlinear Regression Methods

One is often interested in studying how a set of independent variables affect a dependent variable, but the relationship between them cannot be assumed linear. So

the abovementioned models, all of which assume a linear relationship, cannot be applied. Nonlinear regression methods, which may be applicable in such situations to predict the dependent variable from the independent variables and recursive partitioning, or, classification and regression trees method, which is another method that may be applicable in such situations, are discussed in the statistics section of the encyclopedia.

IRT Models

These models, with numerous applications in education, are discussed in an article in the educational measurement section and are not covered here.

Latent Class Models

A latent class model (LCM) relates a set of observed discrete multivariate variables to a set of latent variables (latent variables are not directly observed but are rather inferred, mostly through a mathematical model, from other variables that are observed; e.g., quality of life or intelligence of a person is a latent variable). It is called an LCM because the latent variable is discrete and divides the population into several classes. A class is characterized by a pattern of conditional probabilities that indicate the chance that variables take certain values. For example, Dayton and Macready (2006) discuss an application in which the observed variables are the responses to ten questions on matrix algebra on a test, the latent variable refers to the knowledge of matrix algebra of students, and the latent classes refer to masters and nonmasters on matrix algebra. Given class membership, the conditional probabilities specify the chance certain answers are chosen. Within each latent class, the observed variables are statistically independent (this is often called local independence). This is an important aspect of LCMs. Usually, the observed variables are statistically dependant. By introducing the latent variable, independence is restored in the sense that variables are independent within classes. The association between the observed variables is thus explained by the latent classes.

Time-Series Analysis

A time series is a sequence of data points, measured typically at successive time points. Time series analysis comprises methods that attempt to understand such time series, often either to understand the underlying context of the data points, or to make forecasts (predictions). Forecasting using a time-series analysis consists of the use of a model to forecast future events based on known past events. An example in education is the prediction of

the number of students who will take a test (e.g., SAT) at an administration based on the numbers from the previous administrations of the test. A time-series model generally reflects the fact that observations close together in time will be more closely related than observations further apart. Three broad classes of time-series models of practical importance are the autoregressive (AR) models, the integrated (I) models, and the moving average (MA) models. There are models, such as the autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) that are combinations of the above three.

Model Fit and Model Selection

Model fit analysis refers to an examination of whether the statistical model employed in an application adequately explains the important features of the data set at hand. Model selection refers to the choice of the statistical model that describes the data best among several competing models. Model fit and model selection analysis for the linear models employed in education do not pose any problems and proceed in a similar manner as in any other statistics field, for example, by using residual analysis, Akaike information criterion (AIC) and Bayesian information criterion (BIC) (see, e.g., Draper and Smith, 1998). However, model fit and model selection analysis for the nonlinear models, especially for the IRT models, are not trivial, primarily because the computations are not straightforward with these models, the response variable is discrete so that normality of the response cannot be assumed, and the number of possible responses is huge so that there is sparseness in the data. Fortunately, with the advent of faster computers, there has been substantial work in these areas. Swaminathan *et al.* (2006) provided a detailed review of the literature on model fit of IRT models. Several model fit statistics have been suggested for testing different aspects of an IRT model: statistics for testing the unidimensionality assumption of the IRT model, item fit statistics, person fit statistics, and overall model fit statistics. In applications of IRT models, it is important to employ the appropriate model fit statistics depending on the intended use of the model, and to evaluate not only statistical significance, but also practical significance. For example, it may happen that the value of a fit statistic is statistically significant so that the model does not predict an aspect of the data, but the misfit has negligible consequences operationally.

Kang and Cohen (2007) provided a detailed review of model selection methods for IRT models. More techniques such as the penalty criteria and the theoretical background behind several techniques are discussed in the statistics section of the encyclopedia.

Other Topics in Statistics

Large-sample methods are essential in the application of several statistical methods as it is often important to know the distribution of a test statistic for a large sample. An overview of how to handle multivariate data, which is common in the field of education, is discussed in the statistics section of the encyclopedia. There are several statistics topics that are of increasing interest to the field of education and will most likely become more popular in the coming years. Examples of such topics are analysis of extreme values, association mining, the relationship between cognitive psychology and educational statistics, correspondence analysis, data mining, functional data analysis, graphical models, instrumental variables, latent class analysis, missing data analysis, Monte Carlo methods, Markov chain Monte Carlo methods, sequential probability ratio test, small-area estimation, survival analysis, and signal detection theory. Hence, articles on these topics are included in this section. Computational methods play a key role in statistics in optimizing the time required to run computationally intensive statistical analyses on data. A knowledge of matrix algebra is important in several topics in statistics (such as linear regression, GLM, and ANOVA). Meta analysis, a technique for combining information from several studies, statistical inequalities and statistical paradoxes, which are three topics that any user of statistical methods should be aware of, are discussed in the statistics section of the encyclopedia. Random number generation is often an important issue in applications of statistics to education, especially those involving Monte Carlo or Markov chain Monte Carlo methods. Evaluation research, which is a form of disciplined and systematic inquiry that is carried out to arrive at an assessment or appraisal of an object, program, practice, activity, or system, is discussed in the statistics section of the encyclopedia.

Conclusions

The list of topics in statistics that is provided above, along with the examples of applications of these topics to education, demonstrate how statistics can be useful in education. As more students are tested, access to data becomes easier, computers become faster, and statistical software packages become more accessible, statistics is sure to find more applications in education. However, this situation also suggests the use of caution on several grounds.

Unless the investigator is careful, the resulting inference from a statistical analysis may be inappropriate. See, for example, Kramer and Gigerenzer (2005), of how statistics can be confusing and has been misused, Haller and Kraus (2002) to see how the concept of statistical significance is

misunderstood by university students and their instructors alike, and Stigler (2005) to see how correlation is often misinterpreted as causation. These works show that there is a need to apply appropriate statistical methods to educational applications. The articles in the statistics and educational measurement sections of this encyclopedia (along with the references and the further readings therein) will help practitioners in the field of education to learn more about statistics and to do a better job in choosing appropriate statistical methods in their applications.

There is often a tendency to fit complicated statistical models to real data. However, before using such a model, a researcher should always make sure that the model performs considerably better than a simpler alternative, both substantively and statistically, and that the model is well identified. It is important to remember the following comment in Rubin (1983)

William Cochran once told me that he was relatively unimpressed with statistical work that produced methods for solving non-existent problems or produced complicated methods which were at best only imperceptibly superior to simple methods already available. Cochran went on to say that he wanted to see statistical methods developed to help solve existing problems which were without currently acceptable solutions.

See also: Analysis and Interpretation of Multivariate Data; Analysis of Covariance; Bayesian Statistical Analysis; Bootstrap Method; Canonical Correlation; Categorical Data Analysis; Cluster Analysis: Overview; Computational Statistics; Continuous Probability Distributions; Decision Theory; Design of Experiments; Discrete Probability Distributions; Discrimination and Classification; Empirical Bayes Methods; Evaluation Research; Exploratory Data Analysis; Factor Analysis: An Overview and Some Contemporary Advances; Generalizability Theory; Generalized Linear Mixed Models; Generalized Linear Models; Generating Random Numbers; Goodness-of-Fit Testing; Growth Modeling; Hierarchical Linear Models; Hypothesis Testing and Confidence Intervals; Instrumental Variables; Jackknife Methods; Large-sample Statistical Methods; Latent Class Models; Matrix Algebra; Measure of Association; Measures of Central Tendency; Measures of Dispersion, Skewness and Kurtosis; Meta Analysis; Model Selection; Multidimensional Scaling; Multiple Comparisons; Multivariate Analysis of Variance; Multivariate Linear Regression; Nonlinear Regression Analysis; Nonparametric Statistical Methods; Observational Studies; Point Estimation Methods with Applications to Item Response Theory Models; Probability Theory; Recursive Partitioning; Sampling; Statistical Inequalities; Statistical Paradoxes; Statistical Power Analysis; Statistical Significance Versus Effect Size; Structural Equation Models; Time Series Analysis; Univariate Linear Regression; Value-Added Models.

Bibliography

- Agresti, A. (2002). *Categorical Data Analysis*, 2nd edn. New York: Wiley.
- Allen, N. L., Donoghue, J. R., and Schoeps, T. L. (2001). *The NAEP 1998 Technical Report (NCES 2001-509)*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* **61.53**, 370–418. [Reprinted with biographical note by Barnard, G. A. (1958) *Biometrika* **45**, 293–315.]
- Braun, H. (2005). *Using Student Progress to Evaluate Teaching: A Primer on Value-Added Models*. Princeton, NJ: Policy Information Center, Educational Testing Service.
- Carlin, B. P. and Louis, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. London: Chapman and Hall.
- Cronbach, L. J., Nageswari, R., and Gleser, G. C. (1963). Theory of generalizability: A liberation of reliability theory. *British Journal of Statistical Psychology* **16**, 137–163.
- Dayton, C. M. and Macready, G. B. (2006). Latent class analysis in psychometrics. In Rao, C. R. and Sinharay, S. (eds.) *Handbook of Statistics*, vol. 26, pp 421–446. Amsterdam: North-Holland/Elsevier.
- Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis*, 3rd edn. New York: Wiley.
- Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Everitt, B. S. (1990). Cluster analysis. In Husen, T. and Postlethwaite, N. (eds.) *International Encyclopedia of Education*, 2nd edn., pp 825–831. Oxford: Pergamon.
- Galton, F. (1885). Regression towards mediocrity in heredity stature. *Journal of the Anthropological Institute* **15**, 246–263.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*. New York: Chapman and Hall.
- Haller, H. and Kraus, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research* **7**, 1–20.
- Hanson, B. A., Zeng, L., and Kolen, M. J. (1993). Standard errors of Levine linear equating. *Applied Psychological Measurement* **17**, 225–237.
- Hsu, T. (2005). Research methods and data analysis procedures used by educational researchers. *International Journal of Research and Method in Education* **28**(2), 109–133.
- Huang, F. and Lee, M. (2009). Dynamic treatment effect analysis of TV effects on child cognitive development. No 0906, Discussion Paper Series, Institute of Economic Research, Korea University. <http://econpapers.repec.org/RePEc:iek:wpaper:0906>.
- Jenkins, F., Kaplan, B., and Lim, Y. (2001). Data analysis for the national writing samples. In Allen, N. (ed.) *The NAEP 1998 Technology Report*, 359–370. Washington, DC: National Center for Education Statistics.
- Johnson, R. A. and Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.
- Kang, T. and Cohen, A. S. (2007). IRT model selection methods for dichotomous items. *Applied Psychological Measurement* **31**(4), 331–358.
- Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., and Barbuti, S. M. (2008). Validity of the SAT for predicting first-year college grade point average. *College Board Research Report No. 2008-5*. New York: College Board.
- Kramer, W. and Gigerenzer, G. (2005). How to confuse with statistics or: The use and misuse of conditional probabilities. *Statistical Science* **20**, 223–230.
- Longford, N. T. (1990). Multivariate variance component analysis: An application in test development. *Journal of Educational Statistics* **15**(2), 91–112.
- Novick, M. R. and Jackson, P. H. (1974). *Statistical Methods for Educational and Psychological Research*. New York: McGraw-Hill.
- Rubin, D. B. (1983). Some applications of Bayesian statistics to educational data. *Statistician* **32**, 55–68.
- Sinharay, S. (2005). Assessing fit of unidimensional item response theory models using a Bayesian approach. *Journal of Educational Measurement* **42**, 375–394.
- Sinharay, S. (2006). Bayesian methods in educational measurement. In Upadhyay, S. K., Singh, U., and Dey, D. K. (eds.) *Bayesian Statistics and Its Applications*, pp 422–437. New Delhi: Anamaya.
- Stigler, S. M. (2005). Correlation and causation: A comment. *Perspective in Biology and Medicine* **48**(1), 88–94.
- Swaminathan, H., Hambleton, R. K., and Rogers, H. J. (2006). Assessing the fit of item response theory models. In Rao, C. R. and Sinharay, S. (eds.) *Handbook of Statistics*, vol. 26, pp 683–718. Amsterdam: North-Holland/Elsevier.
- Tatsuoka, K. K. (1984). Caution indices based on item response theory. *Psychometrika* **49**, 95–110.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- von Davier, M., Sinharay, S., Oranje, A., and Beaton, A. (2007). The statistical procedures used in national assessment of educational progress: Recent developments and future directions. In Rao, C. R. and Sinharay, S. (eds.) *Handbook of Statistics*, vol. 26, pp 1039–1055. Amsterdam: Elsevier.
- Wainer, H. (2000). *Visual Revelations: Graphical Tales of Fate and Deception from Napoleon Bonaparte to Ross Perot*, 2nd edn. Hillsdale, NJ: Erlbaum.
- Wainer, H. (2005). *Graphic Discovery: A Trout in the Milk and Other Visual Adventures*. Princeton, NJ: Princeton University Press.
- Zwick, R., Thayer, D. T., and Lewis, C. (1999). An empirical Bayes approach to Mantel–Haenszel DIF analysis. *Journal of Educational Measurement* **36**, 1–28.

Further Reading

- Crocker, L. and Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Harcourt Brace Jovanovich College Publishers.

Relevant Website

- <http://www.ed.gov> – US Department of Education.

Analysis and Interpretation of Multivariate Data

D J Bartholomew, The London School of Economics and Political Science, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Binary response – A variable which can take only one of two values, such as a yes/no answer to a question.

Categorical variable – A variable for which the observed values fall into categories, such as country of birth or socioeconomic class.

Correlation coefficient – A measure of the strength of the relationship between two variables.

Conventionally, 0 denotes no correlation, +1 a perfect correlation, and –1 a perfect negative relation.

Correlation matrix – A square table giving the correlation coefficients between all pairs of variables.

Dendrogram – A graphical way of presenting the results of cluster analysis; also called a tree diagram – hence the dendro- part of the name.

Latent variable – A variable whose value cannot be observed.

Manifest variable – A variable whose value can be observed.

Product moment correlation coefficient – A correlation coefficient measuring the closeness of a relationship to a straight line.

Introduction

Elementary statistical analysis is concerned with the case where we have just one observation per individual. In such cases our interest is usually in answering questions about how large the measurements are or how variable. The answers to such questions are provided by measures of location and dispersion. For further information, we often look at the frequency distribution to see the pattern of variation which lies behind these summary measures.

With multivariate data we may still be interested in looking at the variables individually, but there is now the possibility of investigating the relationships between variables. When we do this, we are doing multivariate analysis.

In education, for example, there may be measures of performance in a variety of school subjects. One might be

interested in how these were related to one another as, for example, if we wanted to know whether performance in mathematics was related to that in history. Or we might wish to bring demographic, or other, variables into the picture by asking whether overall performance depended on birth order in the family or on the time spent doing homework. When only two variables are involved, we speak of a bivariate analysis but this, of course, is only a special case of multivariate analysis where more than two variables are involved.

Multivariate data, therefore, arises whenever we make more than one observation on an individual. In education, such measures might be of performance in tests, demographic features, type of school attended, and so forth.

The Data Matrix

All multivariate data can be set out in what is called a data matrix and it is convenient to think about it in the following way. The conventional approach is to make the rows of the matrix correspond to persons, or whatever is the basic observational unit; the columns correspond to variables. It will be convenient to use x to denote any variable. In practice, such variables will have names, like score in French or school attended, but all such possibilities are represented by our x . A typical data matrix with n individuals and p variables will then appear as follows:

$$\begin{array}{ccccccc} x_{11} & x_{12} & \dots & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & \dots & x_{2p} \\ \dots & & & & \\ \dots & & & & \\ x_{n1} & x_{n2} & \dots & \dots & x_{np} \end{array}$$

Two subscripts are needed to identify an entry in the table. The first tells us which row the individual is in and the second which column: x_{37} is thus the value for the third individual of the seventh variable. Such data might result from a survey in which a sample of n individuals gives answers to p questions or it might record the score of individuals on a set of tests. In practice, some of the values may be missing, because they were never obtained, have been lost, or have been withheld for some reason. This may materially affect what we can learn from the data.

We have spoken of the entries in the data matrix as values of some variable, like a test score but that is not always the case. Any piece of information can be entered

in a data matrix. It might, for example, be the place of residence or nationality. Again, this will determine what kind of analysis we can do but it does not affect the interpretation of the layout.

The Correlation Matrix

Multivariate analysis is about relationships and, in particular, relationships between variables, taken in pairs. The strength of these relationships is measured by correlation coefficients. The first step in any summarization of multivariate data is thus to obtain the pair-wise correlations between variables. In the case of metrical variables, we normally use the product moment correlations. These are measures of linear relationship and, if the relationship is not linear, it may be sensible therefore to first transform the variables in order to make the relationships approximately linear. If we denote the product moment correlation between variables i and j , by r_{ij} we may set the result out in the correlation matrix as follows:

$$\begin{matrix} 1, & r_{12}, & r_{13}, & \dots & r_{1n} \\ r_{21}, & 1, & r_{23}, & \dots & r_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ r_{n1}, & r_{n2}, & \dots & \dots & 1 \end{matrix}$$

Since p will usually be much smaller than n , this matrix is much smaller than the data matrix and thus already represents a considerable reduction in the amount of information we have to take in. The diagonal elements are unity because each variable is perfectly correlated with itself. Almost all analyses of the data matrix start from the correlation matrix and it is always important to inspect this matrix before going further.

The correlation between two variables is, of course, the covariance between the standardized variables. By starting with the correlation coefficients, we ignore the scales of the original variables. This is often desirable because, in much of social science, the scales are themselves arbitrary. There are occasions, however, where the scales of measurement are relevant and should be taken account of in the analysis but we shall not pursue this possibility here.

What we have said relates to the case where all the variables are metrical. If some or all of them are not, the same kind of information can be conveyed by other types of correlation coefficient. Whether or not this will be a first step in the appropriate form of analysis will depend on what analysis is being carried out. With pairs of binary variables, for example, there are many possible coefficients of which the so-called tetrachoric coefficient is the closest to the product moment coefficient. There are other coefficients appropriate for pairs formed from

pairs of categorical (ordered or unordered) variables and for pairs where one variable is metrical and the other categorical.

Types of Data

Data come in many forms. The result of an examination may be recorded simply as pass or fail. For any individual this is often described as a binary response; collectively such responses are expressed as proportions or percentages. Sometimes an outcome may consist of one of a series of ordered categories as, for example, when the categories are: poor, fair, good, very good, and excellent. Otherwise, we may measure ability by a numerical score or a continuous scale as we would with something like weight. For simplicity, all of these levels of measurement may be grouped into two classes which we shall call categorical and metrical. The latter term is used to cover both discrete and continuous measured variables since we do not usually distinguish between them in multivariate analysis and it is convenient to have a single term to cover both. In many practical multivariate problems, we shall have a mixture of types as commonly happens with data derived from sample surveys. For example, some questions require categorical answers, such as place of residence and nationality. Others may be metrical as when we ask for income or age. The form which the analysis takes will depend very much on whether the variables are metrical or categorical but, conceptually, the logic underlying the methods will be similar.

Whether or not the data are obtained by experiment or observation, is another factor which affects the way they are interpreted. In an experimental situation, which is rather rare in the educational context, the conditions under which the variable values are obtained are controlled by the experimenter. One may, for example, test pupils under a variety of environmental conditions, where the individuals placed into each group can be chosen at random. More usually, we simply observe what is going on in which case we have no control over which factors, or combination of factors, are present. In such cases, it is much more difficult to infer causation.

An entirely different distinction has to be made between variables which are observed, called manifest, and those which are not, called latent. At first sight, this distinction may seem to be absurd because latent variables yield no data and hence would appear to be redundant. However, there are contexts in which there is reason to believe that there are other variables which may be relevant, but which cannot be observed. If these latent variables are in causal relationships with observable variables, we may wish to determine their influence and obtain other information about them indirectly through their effects, which we can observe.

Finally, we need to remind ourselves of the fact that individual values of a manifest variable may be unobserved not because they are latent, in the above sense, but because they have been lost or someone refused to answer a question or, for example, when a candidate in an examination did not have the time to complete all the questions in a test.

Summarization and Inference

The forms of multivariate analysis, which we shall review below, fall broadly into two categories. First, there are descriptive methods which aim to summarize multivariate data in a form where its message is more easily grasped. The calculation of a correlation matrix is an example of a descriptive method which is often a first step on the way to a more complete summarization. Second, there are inferential, or model-based, methods. In such cases, our analysis is of a sample from some population and the purpose is to learn something about the population from which the sample has been drawn, rather than about the sample itself. Thus, we might take a sample of schools and wish to infer something about the population of all schools. The step from the sample to the population can only be made, of course, if we know how the sample has been drawn. In practice, the method of sampling needs to be probabilistic. Probability theory then provides the link that we need to pass from sample to population.

An alternative, but equivalent, way of expressing much the same idea is by supposing that the data have been generated by a probability model. The purpose of the analysis is then expressed by saying that we wish to infer what the model is from what we observe in the values which it yields. For example, a set of test scores obtained by a group of pupils on a particular occasion may be thought of as just one sample among those that might have been obtained on other occasions.

There is a conceptual problem in moving from the univariate to the multivariate case. In univariate statistics, it is common to assume that the sample has come from a normal distribution. Even if this is not so, it is often possible to transform the variable to make it approximately normal. In the multivariate case, things are much more complicated because there are many multivariate distributions which have something in common with the normal distribution to choose from. The obvious candidate is the multivariate normal distribution which not only has the important property that the marginal distributions are normal, but also that the regressions are linear. For this reason, most of the theory of multivariate analysis is based on the assumption that the population distribution is multivariate normal. In the social sciences, in general, and the educational sciences, in particular, this assumption is highly suspect because many distributions of educational quantities are

far from normal (test scores are sometimes an exception). Although we shall mention several methods below, which were originally derived on the assumption of normality, they are less applicable and less generally used in the social sciences. Many of the textbooks of multivariate analysis are, however, based on the multivariate normal distribution and are thus less relevant to educational practice.

Dependence and Interdependence

The correlation matrix treats all variables on the same basis in which case any analysis may be described as interdependence analysis. Sometimes, however, the variables do not have the same status. In such cases, our interest may be in how some variables depend upon others; this arises particularly when there is a temporal ordering of the variables. We may then be interested in knowing how the later ones in the sequence depend on those which come earlier. For example, this might enable us to make predictions about the values of the later variables. The conceptually simplest, and best-known, example of dependence analysis is regression analysis where we have a set of variables whose values we wish to use to predict some other variable. For example, this might be from scores obtained in a job selection exercise where we may wish to predict the performance of a candidate on the job. Such predictions are essential in judging the aptitude of potential pilots, for example, when it is too costly or dangerous to test their ability in actual flight.

Methods

We shall now briefly review some of the multivariate methods which are available. These reviews give no technical material because they are designed to explain the main idea behind each method in such a way as to distinguish it from others. Almost all of the methods are described in more detail, and illustrated, in Bartholomew *et al.* (2008), and in other articles in this encyclopedia. We begin with descriptive methods of the interdependence variety.

Cluster Analysis

The object is to put sample members who are similar to one another into the same group, or cluster. Looked at in terms of the data matrix, we put together sample members (rows of the table) which are similar. This could be done crudely by eye but we need a method that is more objective and which can be programmed for a computer. To do this, we construct a measure of the similarity between any

pair of rows, or groups of rows. The simplest methods, called hierarchical, proceed by first grouping the closest pair of individuals, then, at each stage, grouping the closest pair of clusters already formed, and so on. In this way a tree, or dendrogram, may be constructed. The decision on when to terminate the clustering procedure is, essentially, subjective and usually depends on the meaningfulness of the clusters.

Multidimensional Scaling

This takes cluster analysis a stage further by locating sample members in space according to their distance apart. Similarity and distance are equivalent concepts, one being the inverse of the other: similar objects are close together and dissimilar objects are far apart. As the eye can only easily take in information displayed in two or, at most, three dimensions, the main interest lies in finding plots in two dimensions. The problem is then to locate sample members in a space of small dimension so that their distances apart are as near as possible to the distances calculated from the data matrix. The idea is often illustrated by reference to map construction. Atlases and books of reference often give tables showing the distances apart of towns or cities. If we had the distances only, we could attempt to reconstruct the map. This is a simpler problem in that we know in advance that the map is two dimensional. If we do not know the dimensionality, we have to try different numbers, in turn, to try to get a good fit.

The problem as we have described it is, more exactly, metrical scaling because the distances are treated as real distances. Nonmetric scaling is the term used when all we are entitled to assume is that the distances are ordered. This arises because, in social applications, we may feel confident in ranking the distances without wishing to go as far as assigning them precise values.

Principal Components Analysis

This method is applicable only when the elements of the data matrix are metrical variables. Its prime purpose is to reduce the dimensionality of the data. If there were only two columns in our data matrix, we could plot the rows as points in the plane. Inspection of the plot would then show the salient features of the data. For example, if the points fell close to a line, this fact would be clearly evident and we would recognize that the position of each point could be described by the single number defining its position on the line. If the data matrix has more than two columns, it will not be so easy to visualize the position, but we can still use the same geometrical ideas. Dimension reduction of this kind has been a requirement of educational testing for many years. Candidates may

obtain scores on a variety of subjects, or on repeated tests on the same subject, and we then wish to summarize them in, say, a single number. The traditional way of doing this has been to add them up, or average them, and use the result as a measure of performance. Principal components analysis (PCA) is a way of determining whether or not this is a reasonable process and whether one number can provide an adequate summary.

The correlation matrix is the starting point of PCA and it produces linear functions of the variables which have the property that they are uncorrelated with one another and that they are ordered according to the amount of the total variation for which they account.

Latent Variable Methods

Here we come to a family of methods which are often treated as quite distinct methods but which, in reality, are essentially the same. As in PCA, the key idea is to look for a much reduced number of variables in terms of which to describe the relationships in the data. Their novel feature is that their description is in terms of what we called latent (or unobserved) variables. One way of expressing their purpose is to say that they ask whether the correlation matrix can be explained by supposing that it arises from their dependence on a small set of unobserved variables. The oldest and best known of these methods is factor analysis. This applies when all of the variables involved, observed and unobserved, are metrical. Other methods include latent class analysis, latent profile analysis, and latent trait analysis. The relationship among these methods can be conveniently expressed in a fourfold table as set out below. This table is not exhaustive, as many hybrid models are possible, but it conveys the essential structure of the situation.

Classification of Latent Variable Models

In practice, categorical variables will often be binary, especially in educational applications where the observed variables may be right/wrong answers to test items (Table 1). There are extensive and self-contained treatments of these topics, especially latent trait analysis, but, in this introductory overview it is important to emphasize

Table 1 Observed variables and latent variables

<i>Latent variables</i>	<i>Observed variables (x)</i>	
	<i>Metrical</i>	<i>Categorical</i>
Metrical	Factor analysis	Latent trait analysis
Categorical	Latent profile analysis	Latent class analysis

the essential similarity rather than the differences. In practice, it is important to obtain a parsimonious summary of the data in terms of the latent variables. This often involves trying to identify them with substantive entities. For example, the early work on factor analysis, which was invented by Spearman in 1904, was concerned with his attempt to establish the existence of a common factor which he called general ability, or *g*. Latent trait analysis has similar objects when the observed variables are binary.

Regression Analysis

Regression analysis is the oldest, and probably, most widely used multivariate technique in the social sciences. Unlike the preceding methods, regression is an example of dependence analysis in which the variables are not treated symmetrically. In regression analysis, the object is to obtain a prediction of one variable, given the values of the others. To accommodate this change of viewpoint, a different terminology and notation are used. The variable being predicted is usually denoted by y and the predictor variables by x with subscripts added to distinguish one from another. In linear multiple regression, we look for a linear combination of the predictors (often called regressor variables). For example, in educational research, we might be interested in the extent to which school performance could be predicted by home circumstances, age, or performance on a previous occasion. In practice, regression models are estimated by least squares using appropriate software. Important practical matters concern the best selection of the best regressor variables, testing the significance of their coefficients, and setting confidence limits to the predictions.

Discriminant Analysis

Discriminant analysis is most simply thought of as regression analysis when the variable to be predicted is binary. Suppose that individuals belong to one of two categories to which we may assign the values 0 and 1. On each individual, we measure a number of quantities to help us determine to which category the individual belongs. How do we use these variables to help us decide the category to which the individual is to be allocated? One way to do this is to ask what linear combination of the variables best discriminates. This is essentially the same as asking for a way of predicting the category to which the individual belongs. To determine what is called the discriminant function, we need a sample for which the correct category is known. From this, we determine the regression equation treating the dependent variable as binary.

Path Analysis

Path analysis uses regression analysis to elucidate causal relations among a set of variables. For it to be applicable, the regressor variables must be ordered in time so that it is meaningful to speak of one variable being caused by those which precede it in time. A set of regression equations is then estimated so that each variable is regressed on those which precede it in time and on which it is presumed to depend.

Correspondence Analysis

Correspondence analysis originated in an attempt to elucidate the bivariate relationship between pairs of categorical variables. In that sense it may be regarded as a kind of correlation analysis. Multiple correspondence analysis extends the method to more than two categorical variables and in that form it may be regarded as having similar objects to PCA for metrical variables.

Multilevel Modeling

Multilevel modeling has found many applications in educational research; the models are also known under the names of hierarchical linear models, mixed models, and random effects models. It is used because it matches the structure so often found in educational systems where there is a hierarchy of levels. For example, at the lowest level, we have pupils who are grouped into classes which are part of schools and these, in turn, may be part of a region which, itself is embedded in a country. Sampling may take place at some or at all of these levels. Thus, suppose that there is a linear relationship between performance in an examination and parental income within a class, then that relationship may vary from class to class and school to school. The slope and intercept of any estimated regression equation may vary as we move between different levels of the hierarchy. Formally, we may allow the parameters of any regression model we fit to vary by treating them as random variables. It then makes more sense to estimate the parameters of the populations from which the sample is supposed to be drawn. From this brief description, it is easy to see how the various names for multilevel modeling are derived. There are several levels in the population being studied: the models are linear, the levels are arranged in a hierarchy, the models for each level are mixed together, and their coefficients are regarded as random.

Structural Equations Modeling

Factor analysis aims to explain the correlations among a set of observed variables in terms of a smaller set of latent

variables, or factors. Structural equation modeling takes this one step further by supposing that the latent variables themselves are interrelated. In particular, it is usually supposed that there are linear relations among the variables. It is therefore similar to path analysis except that the paths are between latent variables. This means that the regression equations have to be estimated indirectly from what we can observe – namely, the correlations (or covariances) among the observed variables. The starting point is thus the same as for factor analysis but more is asked of the method. In essence, we have to work out what we would expect the covariances to be if the model – which specifies the linear relations among the latent variables – were true. The parameters of the model are then estimated by choosing their values to make them as close as possible to the expected values. In practice, this is a difficult numerical problem which can only be solved efficiently by using appropriate computer software. The method is thus often known by the acronyms designating the many software packages which are now available, for example, LISREL, AMOS, EQS, Mplus, and GLAMM. As with any complex method of analysis, it is important to check the conclusions carefully as it is often the case that different models can lead to very similar covariance structures.

See *also*: Cluster Analysis: Overview; Correspondence Analysis; Discrimination and Classification; Latent Class Models; Multidimensional Scaling; Principal Components Analysis; Structural Equation Models.

Bibliography

Bartholomew, D. J., Steele, F., Moustaki, I., and Galbraith, J. I. (2008). *Analysis of Multivariate Social Science Data*, 2nd edn. London: Chapman and Hall/CRC.

Further Reading

- Bartholomew, D. J. and Knott, M. (1999). *Latent Variable Models and Factor Analysis, Vol. 7: Kendall Library of Statistics*, 2nd edn. London: Arnold.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.
- Cox, T. F. and Cox, M. A. A. (2001). *Multidimensional Scaling*, 2nd edn. London: Chapman and Hall/CRC.
- Everitt, B. S. and Dunn, G. (2001). *Applied Multivariate Data Analysis*, 2nd edn. London: Arnold.
- Goldstein, H. (2003). *Multilevel Statistical Models*, 3rd edn. London: Arnold.
- Gordon, A. D. (1999). *Classification*, 2nd edn. London: Chapman and Hall/CRC.
- Greenacre, M. and Blasius, J. (eds.) (1994). *Correspondence Analysis in the Social Sciences*. San Diego, CA: Academic Press.
- Heinen, T. (1996). *Latent Class and Discrete Latent Trait Models*. Thousand Oaks, CA: Sage.
- Jolliffe, I. T. (1986). *Principal Components Analysis*. New York: Springer.
- Krzanowski, W. J. and Marriott, F. H. C. (1995). *Multivariate Analysis, Part 1: Distributions, Ordination and Inference*. London: Arnold.
- Krzanowski, W. J. and Marriott, F. H. C. (1995). *Multivariate Analysis Part 2: Classification, Covariance Structures, and Repeated Measurements*. London: Arnold.
- Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman and Hall/CRC.

Analysis of Covariance

N T Longford, Universitat Pompeu Fabra, Barcelona, Spain

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cause – A factor that can be manipulated; that is, its value could be set by design, in complete autonomy (not affecting any of the covariates).

Effect – Difference in the outcomes that can be attributed without any ambiguity solely to a specified cause.

Fixed effect – A factor that is constant across replications.

Hypothesis testing – A data-based procedure for deciding whether to contradict or not to contradict a specific conjecture (hypothesis).

Missing data – Data items that were meant to be collected (as per protocol), but were not.

Potential outcome – Outcome associated with the value of a factor (treatment), constant across the replications in which that value is assigned.

Random effect – A factor (a set of categories) that differs from one replication to the next.

Randomization – Assignment of the value of a factor made independently of the outcome, any covariates, or any attributes of the subjects.

Replication – A repeated application of a data-generating process that is independent of all the previous applications.

Variance – A measure of the dispersion of the values of a variable.

Introduction and Essentials

Analysis of covariance (ANCOVA) is used for studying the differences among a set of K groups. In each group, the outcome variable Y is related to one or several covariates (X) by ordinary regression. These regressions may have some features in common. As a simple example, suppose the outcome variable Y is a score on an educational test at time point 2, following a treatment, such as a course of study, and X is the score on a similar test, measuring the same trait, administered at time point 1, prior to the treatment. Suppose there are K treatments, each subject (student) is assigned to one of them, and this assignment is made completely at random.

In this example, the treatments $1, \dots, K$ form the groups, and the differences among the treatments are reflected

in the different regressions; the conditional expectation $E(Y | X = x; k)$ depends on k . We assume that the conditional distributions of Y given the value of X and the treatment k are independent; students do not influence one another's performances (e.g., by conferring during the exam, or by distracting one another). The simplest model of ANCOVA occurs when X is univariate and X and Y are jointly (bivariate) normally distributed, and

$$Y = \alpha_k + \beta X + \epsilon \quad [1]$$

where ϵ is a random sample from $N(0, \sigma^2)$, β (the regression slope on X) is a constant common to all the groups, and α_k are intercepts, specific to the groups $k = 1, \dots, K$. Thus, the regressions in [1] are parallel (see the left-hand panel of **Figure 1** for an illustration). The treatments have a straightforward ordering according to the values of α_k . Although the intercepts α_k depend on the origin and scale of X , the ordering of the treatments does not. Suppose we use $X^* = cX + d$ instead of X ; $c \neq 0$ and d are some constants. Then the model in [1] is equivalent to

$$Y = \alpha_k - \beta \frac{d}{c} + \frac{\beta}{c} X^* + \epsilon \quad [2]$$

and each intercept is altered by the same quantity $\beta d/c$.

The model in [1] reduces to analysis of variance (ANOVA) when β is set to zero or when there is no covariate X :

$$Y = \mu_k + \epsilon \quad [3]$$

ANCOVA is a regression model with one continuous and one categorical covariate, X and the group or treatment, respectively. It is expressed in the standard form by defining the indicators of the treatments:

$$I_i^{(k)} = 1$$

if subject i is assigned to treatment k , and $I_i^{(k)} = 0$ otherwise. To adhere to the standard form of the ordinary regression, we declare one of the treatments as the reference; no generality is lost by using treatment 1 for this purpose. Now

$$Y = \beta_0 + \sum_{k=2}^K \gamma_k I_i^{(k)} + \beta X + \epsilon \quad [4]$$

or, if the indices of the subjects and treatments are introduced,

$$Y_{ik} = \beta_0 + \gamma_k + \beta X_{ik} + \epsilon_{ik}$$

with the understanding that $\gamma_1 = 0$.

The coefficients γ_k are related to α_k in [1] by the identity

$$\gamma_k = \alpha_k - \alpha_1$$

The hypothesis that $\gamma_k = 0$ for a given treatment $k \neq 1$ is equivalent to $\alpha_k = \alpha_1$, and has therefore a meaning specific to the choice of the reference category. When the choice of this reference is arbitrary, such a hypothesis is not well motivated. The hypothesis that the constants α_k are all identical corresponds to $\gamma_2 = \dots = \gamma_K = 0$, that is, the K groups have a common regression

$$Y = \beta_0 + \beta X + \epsilon$$

This hypothesis is often well motivated and of practical interest.

The assumption of parallel regressions may be too restrictive. It is relaxed by introducing the interactions of the covariate X with the treatment:

$$\begin{aligned} Y &= \beta_0 + \sum_{b=2}^K \gamma_b I^{(b)} + \beta X + \sum_{b=2}^K \delta_b I^{(b)} X + \epsilon \\ &= \alpha_k + \beta_k X + \epsilon \end{aligned} \quad [5]$$

where k denotes the treatment that was applied, regarded as a variable (see the right-hand panel of **Figure 1** for an illustration). The treatments can now be compared by the values of the expectations $\alpha_k + \beta_k x$ for a range of (realistic) values x of X . Of course, the expectation $\alpha_k + \beta_k x$ for one treatment may exceed the expectation for another for some values of x and may fall short for other values. **Figure 2** gives several examples. In all panels except the bottom left-hand one, the treatments are well ordered, because the comparisons between the expected outcomes for pairs of treatments do not depend on x .

If a model, such as [5], is appropriate, then it is also appropriate when we replace X with $c + dX$ for some constants c and $d \neq 0$. We refer to this property as invariance with respect to linear transformations. It is a highly desirable property of any linear model when

instead of a particular version of the covariate X we could equally well have chosen one of its linear transformations.

The combination of invariance with respect to linear transformations of X and arbitrariness of the reference category yields the following rules for how the model in [5] can be simplified:

- The covariate X and all the indicators $I^{(k)}$ (the main effects) should be retained on the right-hand side (in the model) whenever the interactions $I^{(k)}X$ are included;
- all the $K - 1$ interactions $I^{(k)}X$ should be included in or excluded from the model *en masse*;
- when all the interactions $I^{(k)}X$ are excluded, the $K - 1$ indicator variables $I^{(k)}$, $k = 2, \dots, K$ should be retained in or excluded from the model *en masse*.

Corresponding to these rules are guidelines for which hypotheses are meaningful. Thus, it is meaningful to formulate and test only hypotheses that relate to the collection of the $K - 1$ interactions $I^{(k)}X$. Only when these are not present in the model is it meaningful to formulate and test hypotheses about the collection of the $K - 1$ indicators (treatment-contrasts) $I^{(k)}$ or about the (common) slope β . The hypothesis that a single or a subset of the interactions $I^{(k)}X$ have zero coefficients has a meaning contingent on the choice of the reference category.

Extensions to several covariates X are obvious. For each covariate we can define interactions with the treatment indicators $I^{(k)}$. In fact, interactions can be defined also among the covariates; then the invariance rules have to be extended. In principle, both main effects of an interaction have to be retained in the model, even though the values of the parameters associated with these main effects have no interpretation.

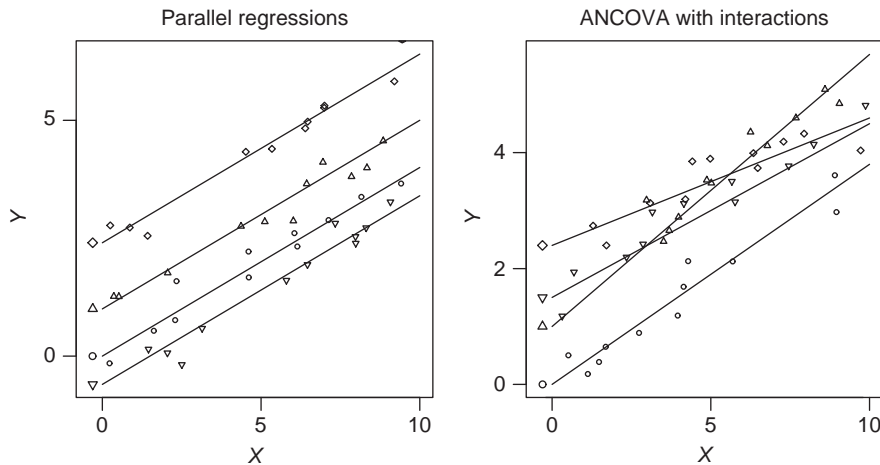


Figure 1 Analysis of covariance. Parallel regressions (no interactions) and covariate-by-group interactions. The symbols indicate association with the group marked at the left-hand margin of each panel.

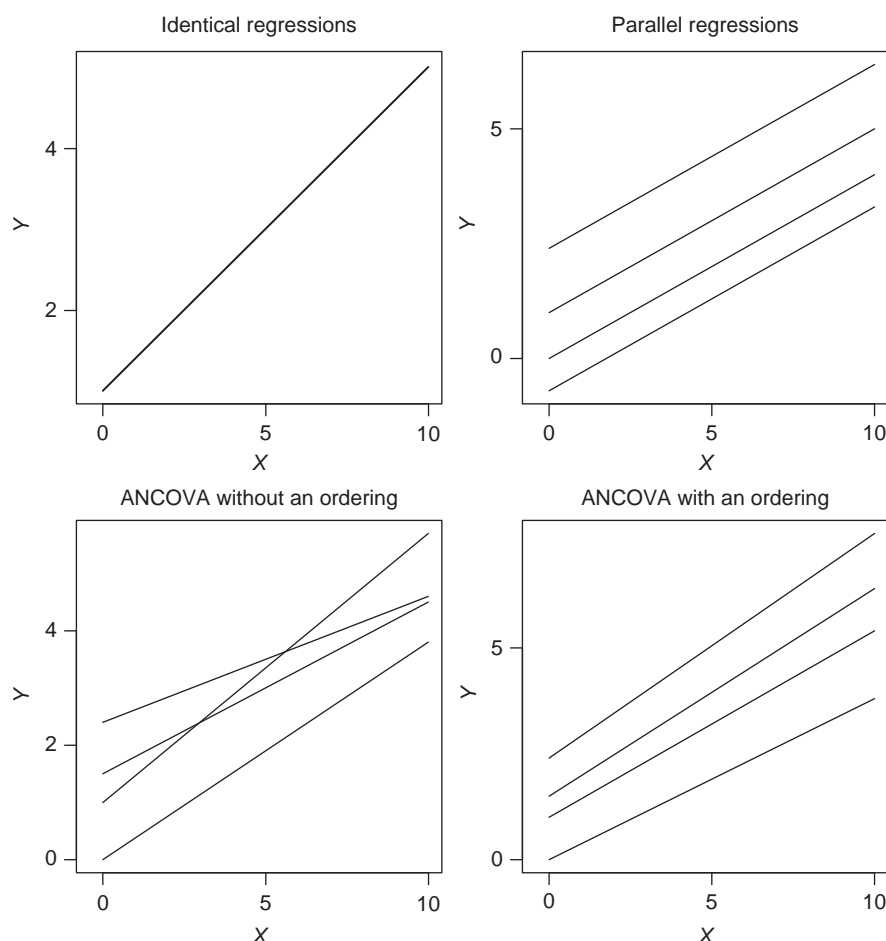


Figure 2 Patterns of the regressions in ANCOVA.

Further Details and Extensions

Random or Fixed

The groups are said to be fixed if they have the same identity in replications; that is, the subjects in the subsample for every group k are drawn from the same subpopulation. The subsample sizes n_k may vary (for a given k) across the replications, as may the values of X . The groups are said to be random, if there is a (large) population of groups, and the groups to be sampled from are themselves selected first, by a sampling process, before subjects are selected from each group to the sample. Thus, when groups are random, their identification (labeling) by the indices $k = 1, \dots, K$ is temporary because they are included in the study by chance. Summaries of the groups, such as the variance (matrix) of their coefficients β_k , may be relevant.

The Potential Outcomes Framework

Randomization of the subjects to treatments is a key design feature that enables a straightforward interpretation of

the model parameters in ANCOVA, as it is in ANOVA. It ensures that the treatment assigned does not depend on the vector of the so-called potential outcomes $(Y^{(1)}, \dots, Y^{(K)})$, which would be realized with the treatments $k = 1, \dots, K$. In practice, only one treatment is assigned to a subject, but in (hypothetical) replications a subject is not assigned the same treatment every time. In replications of the regression or ANCOVA, we have different sets of subjects but the same set of treatments $1, \dots, K$; the subjects are random and the treatments are fixed. If the subjects were also fixed, their outcomes would differ from replication to replication even if they were (or happened to be) assigned the same treatment, because a fresh (independent) ϵ is generated every time.

A framework alternative to ANCOVA assumes that each subject has a fixed vector of the potential outcomes $Y = (Y^{(1)}, \dots, Y^{(K)})$. In the population of subjects (students), these vectors may have a particular distribution, such as a K -variate normal, but for any one subject i who is included in two replications of the study and is assigned the same treatment k in both, the same outcome $Y_i^{(k)}$ is realized.

The problem of estimating any population summaries of Y , such as certain contrasts, can be formulated as a missing data problem. Suppose the entire population participates in the study. If the entire vector Y were observed for every subject, the inference would be straightforward. The observed values of the outcomes are regarded as the incomplete information and the values of Y for all the subjects as the complete information. The missing information is estimated, and the assignment process, which describes how treatments are assigned to subjects, is key. If the assignment is by randomization, the analysis is straightforward; if it depends on the values of the outcomes (e.g., when subjects select their treatment in pursuit of a specific agenda, such as the speediest and most complete recovery), complications arise. A general approach matches each subject which received a particular treatment with a similar subject which received a different treatment. Comparisons are then based on such matched pairs. An obvious advantage of this approach is that some of the assumptions specific to regression (normality, linearity, and homoscedasticity) can be dispensed with.

Multiway ANCOVA

For a small number of groups, no useful structures are usually defined. When there are many groups, they may themselves be clustered, so that cluster A contains a few groups, cluster B some others, and so on.

A group is formed by setting the category (level) of A, to k_A , and category of B, to k_B . For example, the two classifications may be two types of treatment, such as teaching style (say, four categories) and classroom size (two categories); then there are eight possible groups defined by crossing these two factors. Such settings are referred to as two-way designs and their analysis as two-way ANCOVA.

Factors may be nested within clusters. For example, the regions of a country may be the clusters, and its districts the groups. Districts are nested within regions if each district belongs to a single region, that is, when no district straddles regions. Groups are said to be crossed when two (or more) classifications are defined for them, such as classification A with K_A categories and classification B with K_B categories, and there is a category of either classification that does not determine the classification of the other. For example, the year of study and country of origin of students in a typical setting of a college are crossed classifications because some countries are represented in several years and in each year there are students from several countries.

Multiway ANOVA is defined by introducing further factors in the crossed design and further clustering in the nested design. Further, the features of clustering and nesting can be combined.

Other Generalizations of ANOVA

ANCOVA is a generalization of ANOVA in the following sense. In ANOVA, each group is associated with distribution, or model, $N(\mu_k, \sigma^2)$. In ANCOVA, groups are associated with conditional distributions

$$N(x\beta_k, \sigma^2)$$

where x is a vector of regressors (covariates) which include the intercept. Suppose the intercept is the first component of x , $x_1 = 1$. The regression parameter vector β_k has some of its components specific to each group k , and others common to all the groups k :

$$\beta_k = \begin{pmatrix} \beta_{k,\text{spe}} \\ \beta_{\text{com}} \end{pmatrix}$$

either part, $\beta_{k,\text{spe}}$ and β_{com} , may be empty.

More generally, the groups can be associated with any models, that is, conditional distributions, with sets of parameters θ_k governing them. These parameters may have some of their components specific and others common to the groups. The residual variance, if defined as one of the parameters in θ_k , should be common to the groups, and the conditional distributions should be normal, so that these models would be generalizations of ANOVA.

The models need not be related to ordinary regression. Generalized linear models have their ANCOVA versions, as do factor analysis and cluster analysis, stating that each group is associated with the same model (class of distributions), but the groups have different distributions, even if the values of some of their parameters are identical across the groups.

Hypothesis Testing in ANCOVA

In hypothesis testing in general, we specify a hypothesis as a submodel, often by specific values of some of the model parameters, and assess whether the data contain evidence against the hypothesis. For example, in a one-way ANCOVA with a single covariate X and parallel regressions (no interactions), we may specify as the (null-) hypothesis the submodel with identical regressions. We reject the null hypothesis if the data are not compatible with it. Hypothesis testing is conducted by defining a suitable statistic ξ and subset of its values, C , usually an interval or its complement. Then the hypothesis is rejected if the realized value of ξ is outside C .

A standard approach to deriving hypothesis tests is by the likelihood ratio method. Under some general conditions, such tests have good asymptotic (large-sample) properties. Likelihood ratio tests in ANCOVA have a particularly simple description in terms of the fitted (estimated) residual variances $\hat{\sigma}^2$. By way of an example we

give details of the test of the hypothesis that the regressions are identical, against the alternative that they are parallel:

$$H_0 : \gamma_2 = \dots = \gamma_K = 0$$

$$H_1 : \sum_{k=2}^{K-1} \gamma_k^2 > 0$$

Assuming H_0 , the estimator of the residual variance σ^2 has a scaled χ^2 distribution with $n - p$ degrees of freedom:

$$(n - p) \frac{\hat{\sigma}_0^2}{\sigma^2} \sim \chi_{n-p}^2$$

(n is the number of observations and p the number of covariates in the model, including the intercept). The estimator of the residual variance obtained when the alternative model is fitted has $K - 1$ fewer degrees of freedom, one for each additional parameter γ_k that is estimated:

$$(n - p - K + 1) \frac{\hat{\sigma}_1^2}{\sigma^2} \sim \chi_{n-p-K+1}^2$$

The subscript of $\hat{\sigma}^2$ indicates the model considered. Further,

$$(n - p) \frac{\hat{\sigma}_0^2}{\sigma^2} - (n - p - K + 1) \frac{\hat{\sigma}_1^2}{\sigma^2} \sim \chi_{K-1}^2$$

and this statistic is independent of $\hat{\sigma}_1^2$. The F statistic is defined as the ratio

$$F = \frac{(n - p - K + 1) \hat{\sigma}_1^2 - (n - p) \hat{\sigma}_0^2}{(K - 1) \hat{\sigma}_1^2}$$

under the hypothesis it has the F distribution with $K - 1$ and $n - p - K + 1$ degrees of freedom. The hypothesis is rejected for large values of F , so $C = [0, C^*)$ for suitably chosen value of C^* . As a convention, the size of the test, equal to the probability of (inappropriate) rejection when the hypothesis holds, is set to 0.05, and so the critical value C^* is set to the 95th percentile of the F distribution with $K - 1$ and $n - p - K + 1$ degrees of freedom.

The likelihood ratio test for a general hypothesis in ANCOVA proceeds as follows:

1. Specify the general model (B), and the hypothesis (A) as a special case of B, obtained by constraining the values of q parameters in B to given constants. Suppose B involves p model parameters.
2. Estimate the residual variance in both models, obtaining $\hat{\sigma}_A^2$ and $\hat{\sigma}_B^2$. Associate these estimators with their respective degrees of freedom, $n - q$ and $n - p$.
3. Evaluate the F -test statistic

$$F = \frac{\hat{\sigma}_B^2 - \hat{\sigma}_A^2}{\hat{\sigma}_A^2} \frac{p - q}{n - p}$$

4. Reject A if F exceeds the 95th percentile of the F distribution with $p - q$ and $n - p$ degrees of freedom.

This procedure is applicable only for null hypotheses specified by constraining one or several parameters to specific (default) values. Adaptations for some one-sided hypotheses, such as $\beta < 0$, are straightforward. When the hypothesis is that a parameter is in an interval, or a more complex set, several complications arise, the treatment of which, is beyond this article.

Example

In an educational study, a class of 95 students, split into four tutorial groups, were given an exam with an extended problem in probability theory, comprising 15 questions. One group was given the problem on its own, another group was also given hints on how to proceed, and the third group was provided the key definitions and explanation of the terms used. The fourth group was given both hints and the key definitions. Each student was graded on the scale 0–15, by adding up the separate scores for the questions, which themselves can be fractional (0–1, with 0.5, 0.25 and 0.75 quite common). As a covariate, each student's grade-point average from the previous year of study is available. It is recorded on the scale 5–10.

We want to assess the gain that can be attributed to the conditions of the exam, the provision of hints and the key definitions. Relevant to this is how the students are assigned to the tutorial groups. This is done in an administrative fashion, although students can switch from one class to another, so long as no class has too many students. This does not suggest any systematic *a priori* differences among the groups of students. However, the teaching assistants may differ in their qualities relevant to the instruction process, and so the four groups cannot be regarded as 'equivalent'. Furthermore, the treatment applied is confounded with the teaching assistant as a factor. That provides a rationale for adjustment by ANCOVA, and the student's grade in the previous year is regarded as a suitable proxy for it. Some core courses were taken by all students, but students had several optional courses. For simplicity and better illustration of ANCOVA, we disregard their differential difficulty, or strictness in their grading.

The data are plotted in **Figure 3** in four panels, with the observations for a group highlighted in each panel. The solid line is the regression fit to the data for the group, with the details given in the subtitle. The fit for the entire dataset, with the grouping ignored, is drawn by dashes; this fit is $\hat{y} = -3.35 + 1.67X$, with estimated residual variance $\hat{\sigma}^2 = 2.64$. The panels have identical scales, to make an informal comparison of the four groups easier.

The fit for the ANCOVA with no interactions is

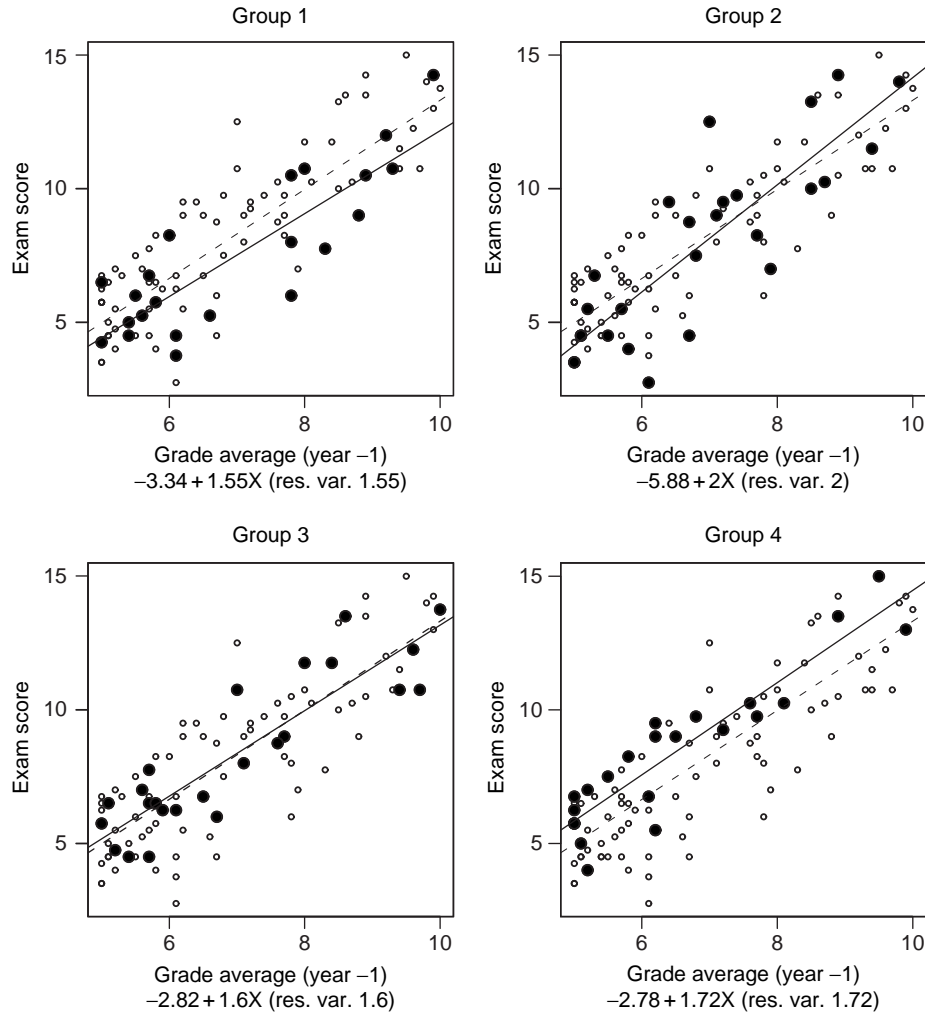


Figure 3 The exam scores and grade averages of the students in four experimental groups. Observations from the indicated group are highlighted (●) and the other observations are drawn by small open circles (o). The regression fitted to the observations in the group is marked by solid line, and the overall regression fit (ignoring the groups) by dashes.

Parameter	Intercept	Grade (Yr - 1)	Group (contrast)		
			2 - 1	3 - 1	4 - 1
Estimate	-4.50	1.72	0.59	0.86	1.78
Standard error		(0.10)	(0.44)	(0.45)	(0.46)

with estimated residual variance $\hat{\sigma}^2 = 2.32$. The display of the fit might invite the conclusion that there is no difference between groups 2 and 1 (judging by the comparison of the estimate 0.59 with the estimated standard error 0.44). Such a conclusion would be flawed, as would be the conclusion that the difference between groups 2 and 1 is small, unless we come to a consensus for what 'small' means. An informed student might reasonably argue that, for example, the difference of one point is far from small,

yet such a difference (between groups 2 and 1) is quite plausible.

To test the hypothesis that there are no differences among the groups (after adjusting for the grade in the previous year), we evaluate the F statistic:

$$F = \frac{2.64 \times (95 - 2) - 2.32 \times (95 - 5)}{2.32} \times \frac{1}{3} = 12.24$$

and compare it with the 95th percentile of the F distribution with 3 and 90 degrees of freedom, which is 2.70. As $12.24 > 2.70$, we conclude that we have evidence that the average performances of the four groups on the exam differ even after adjusting for grade average. With this conclusion, it is not appropriate to make it more specific and state that group 4 has the highest average (adjusted) performance, because the hypothesis tested was solely about differences, without their ordering.

Conclusion

ANCOVA is concerned with studying regressions in a set of groups. Models for ANCOVA cater for a wide range of patterns of these regressions and include procedures for selecting among them. As hypothesis testing is the principal method for this, its general limitations, especially in the context of several hypotheses, have to be carefully considered. Extensions of ANCOVA include structures for the groups, such as crossing, nesting, and their combinations, within-group models that are more complex than ordinary regression (factor analysis and generalized linear models), and the groups can be associated with random effects.

Acknowledgments

This work was partly supported by the Grant SEJ2006-13537 from the Spanish Ministry of Science and Technology.

See *also*: Analysis of Variance; Causal Inference; Hypothesis Testing and Confidence Intervals; Model Selection; Observational Studies; Value-Added Models.

Further Reading

- Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd edn. New York: Wiley.
- Doncaster, C. P. and Davey, A. J. H. (2007). *Analysis of Variance and Covariance*. Cambridge: Cambridge University Press.

- Draper, N. and Smith, H. (1981). *Applied Regression*, 2nd edn., ch. 5. New York: Wiley.
- Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. New York: Springer.
- Holland, P. W. (1986). *Statistics and causal inference*. *Journal of the American Statistical Association* **81**, 945–970.
- Huitema, B. (1980). *The Analysis of Covariance and Alternatives*. New York: Wiley.
- Kutner, M. H., Neter, J., Nachtsheim, C. J., and Li, W. (2004). *Applied Linear Statistical Models*. New York: McGraw Hill.
- Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd edn. New York: Springer.
- Longford, N. T. (2005). Editorial: Model selection and efficiency: Is 'Which model . . .?' the right question? *Journal of the Royal Statistical Society A* **168**, 469–472.
- Longford, N. T. (2005). *Missing Data and Small-Area Estimation. Modern Analytical Equipment for the Survey Statistician*, ch. 11. New York: Springer.
- Longford, N. T. (2007). *Studying Human Populations. An Advanced Course in Statistics*, ch. 1. New York: Springer.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall.
- Rosenbaum, P. R. (2003). *Observational Studies*, 2nd edn. New York: Springer.
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modelling, decisions. 2004 Fisher Lecture. *Journal of the American Statistical Association* **100**, 322–331.
- Searle, S. R. (1987). *Linear Models for Unbalanced Data*. New York: Wiley.

Analysis of Extreme Values in Education

J Beirlant, Katholieke Universiteit Leuven, Heverlee, Belgium

G Dierckx, Hogeschool-Universiteit Brussel, Brussel, Belgium

K Luwel and P Onghena, Katholieke Universiteit Leuven, Heverlee, Belgium

© 2010 Elsevier Ltd. All rights reserved.

Goals of Extreme Value Theory

Extreme value methodology has been constructed with the specific aim to estimate extreme quantiles and/or small exceedance probabilities corresponding to rare events. Classical applications are the calculation of return periods of floods or other catastrophic events, or the prediction of large claims in insurance portfolios. In educational studies applications are less prominent but also here the statistical modeling of rare events can be of interest. We present an application to the modeling of reaction times in a numerosity test when applying the choice/no-choice method. This method involves testing participants under two types of conditions: (1) a choice condition in which participants can freely choose which strategy to use and (2) several no-choice conditions in which participants are required to apply a given strategy on all problems. Asking all participants to use a given strategy on all trials in the no-choice conditions excludes possible selection efforts and/or individual preferences, resulting in unbiased estimates of the speed and accuracy of the strategies under consideration. It also becomes possible to investigate participants' strategy adaptiveness by comparing their actual strategy choices under the choice condition with their ideal strategy choices based on the no-choice performance data.

The application is taken from Luwel *et al.* (2003) where the authors wanted to extend the applicability of the choice/no-choice method toward the domain of numerosity judgment. They aimed at developing a reliable and valid method for measuring the adaptiveness of numerosity judgment strategies, based on the procedure for strategy identification in the individual response-time patterns. Participants were instructed to determine, as quickly and accurately as possible, different numerosities of colored blocks that were presented in a 7 by 7 grid on a computer screen. This task allows for two main strategies: an addition strategy in which the different colored blocks in the grid are added and a subtraction strategy in which the number of empty squares is determined and then subtracted from the total number of squares in the grid. In contrast to the addition strategy which only involves one solution step, the subtraction strategy is a two-step strategy and therefore, cognitively more demanding. The experiment was performed under three conditions:

one choice (C) condition and two different no-choice (NC) conditions: (1) In the C-condition participants were free to use either the addition or the subtraction strategy, (2) in one NC-condition (i.e., the forced addition or FA-condition) all trials had to be solved by means of the addition strategy, and (3) in the other NC-condition (the forced subtraction or FS-condition) the use of the subtraction strategy was required on all trials. Thirty-seven students at the University of Leuven (Belgium) participated in this study. Their mean age was 21 years with ages ranging from 18 to 26 years. Each participant was tested individually and ran three different sessions. The presentation order of the different conditions was counterbalanced over participants with the important restriction that C-condition was always presented first. As a consequence all participants were randomly divided over one of both presentation orders: C/FA/FS or C/FS/FA.

In **Figure 1** the reaction times were plotted as a function of the number of blocks (from 1 to 49 blocks) for each of the three conditions: (1) FA-condition; (2) FS-condition and (3) C-condition. Some participants show extreme long reaction times. Whereas some statisticians would qualify these as outliers, the right skewness of the reaction times distribution (as a function of the block size) is always apparent and appears stronger for larger block sizes in the FA- and C-conditions, and for smaller block sizes in the FS-condition. Extreme value methodology provides models that allow one to model this right-skewness and the largest reaction times. More specifically, it tries to answer questions such as: what is, for a given number of blocks, the extreme reaction time level that occurs only once in 100 trials.

The Extreme Value Modeling Approach

The Basic Model

Here we present the extreme value theory (EVT) approach for pure random samples X_1, X_2, \dots, X_n . The population distribution is assumed to satisfy the domain of attraction condition which expresses that for large sample sizes n the statistical behavior of the maximum of such samples can be well approximated by a nondegenerate distribution which then is shown to be necessarily an extreme value distribution with cumulative distribution function

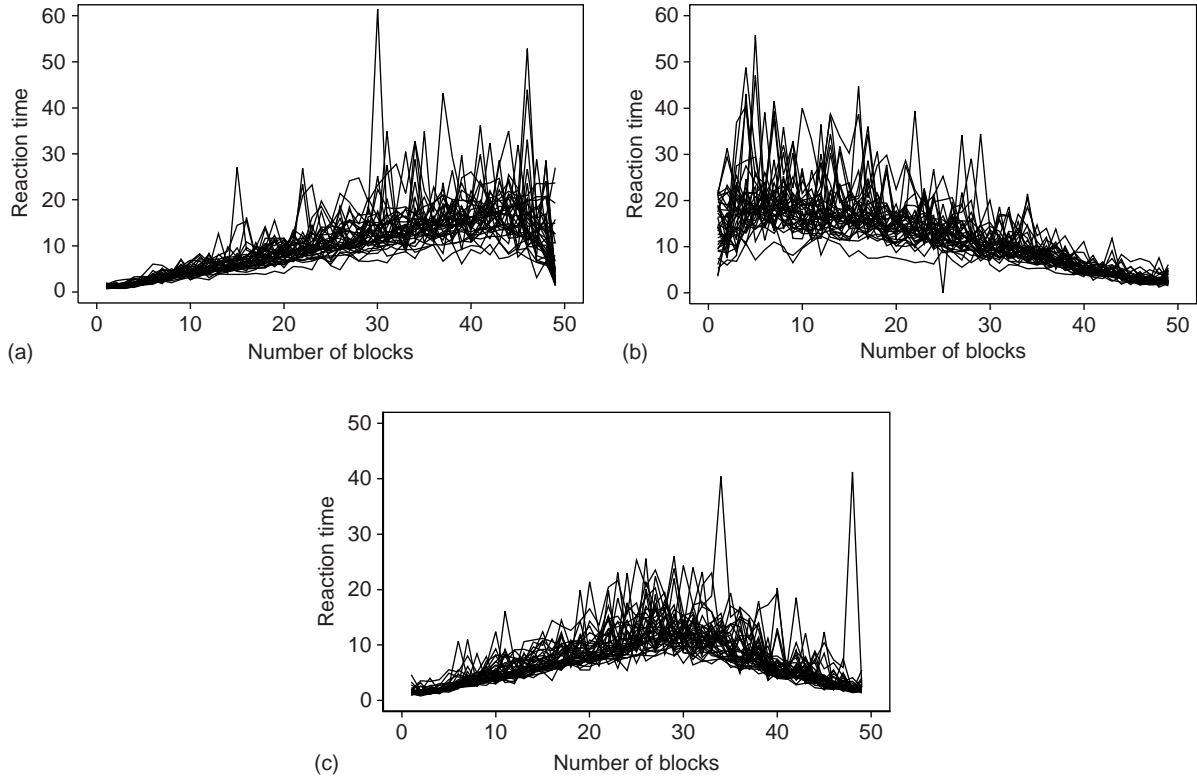


Figure 1 The reaction time (RT) as a function of the number of blocks for 37 participants: (a) addition; (b) subtraction, and (c) choice.

$$\exp[-(1 + \gamma y)^{-1/\gamma}] \quad [1]$$

The parameter γ , called the extreme value index (EVI), characterizes the tail decay of the underlying distribution F : a positive value of γ corresponds to heavy-tailed or Pareto-type distributions for which the density decays as a power law, while $\gamma = 0$ typically indicates an exponentially decreasing tail. Finally, $\gamma < 0$ indicates a distribution F with a finite right endpoint. This general condition is quite broad so that it does not really lead to restrictions in practice.

It can be shown that this framework is equivalent to assuming the following:

The POT-GPD assumption

For large enough threshold values t , the excesses, or peaks over threshold (POTs) $X - t$ (with $X > t$) approximately follow a generalized Pareto distribution (GPD) with cumulative distribution function $1 - (1 + \frac{\gamma}{\sigma} y)^{-1/\gamma}$.

In case $\gamma = 0$ the GPD is to be interpreted as the exponential distribution with cumulative distribution function $1 - e^{-y/\sigma}$ for $y > 0$. This discussion makes it clear that the goal of tail estimation is carried out under a nonstandard model which involves the choice of a threshold t . Of course, the estimation of the EVI parameter γ is an important step toward the goal of tail estimation.

Tail Estimation

The POT approach to estimate γ and $\sigma > 0$ simply consists of using for instance the maximum likelihood approach when fitting the GPD to the excesses $X - t$ for those data X which are larger than t . Quite often, the maximum likelihood estimation procedure leads to appropriate estimates $\hat{\gamma}_k$ and $\hat{\sigma}_k$ which show a stable behavior as a function of the threshold t . We refer to this method as the GP-POT approach.

Most commonly, t is chosen as one of data points itself, say as the $(k + 1)$ -largest observation, and then one lets k vary between 2 and the full sample size n . In **Figure 2** the POT estimators for the extreme value index of the RT values in our case study are shown for each of the three conditions as a function of $k = 2, \dots, 37$ with the number of blocks equal to 30. Remark that in case of the addition strategy the estimated values are clearly positive, while for the other two conditions values around zero are predominant. In **Figure 3** the POT estimators of γ obtained with $k = 17$ are shown for each condition as a function of the number of blocks. Comparing with **Figure 1** we remark that the appearance of high outliers in reaction times leads to predominant positive EVI estimates. **Figure 4** shows the corresponding maximum likelihood estimates of the scale parameter σ . While the POT estimates of the extreme value index do not seem to show trends with the number of blocks, the σ estimates grow linearly for the addition strategy, while

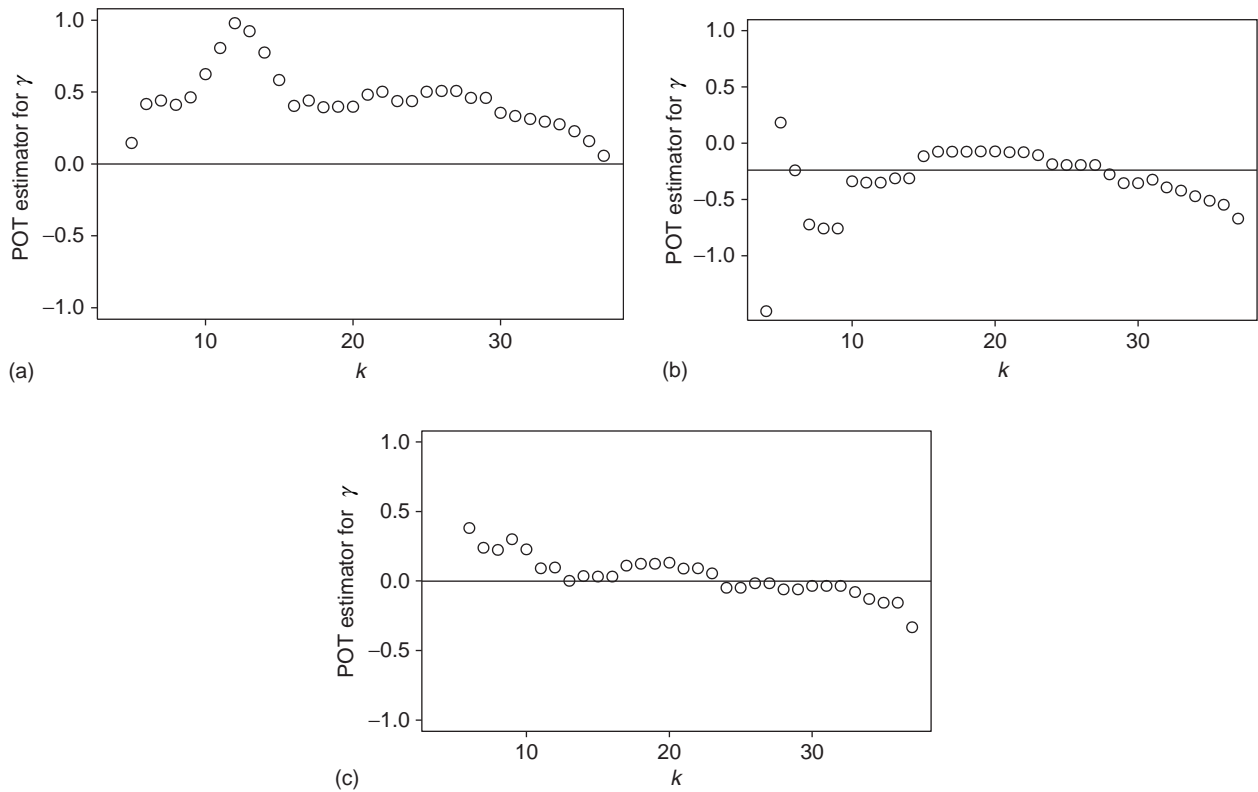


Figure 2 POT estimator for the extreme value index of the reaction times as a function of $k = 2.37$ (37 participants) for 30 blocks for: (a) addition, (b) subtraction, and (c) choice.

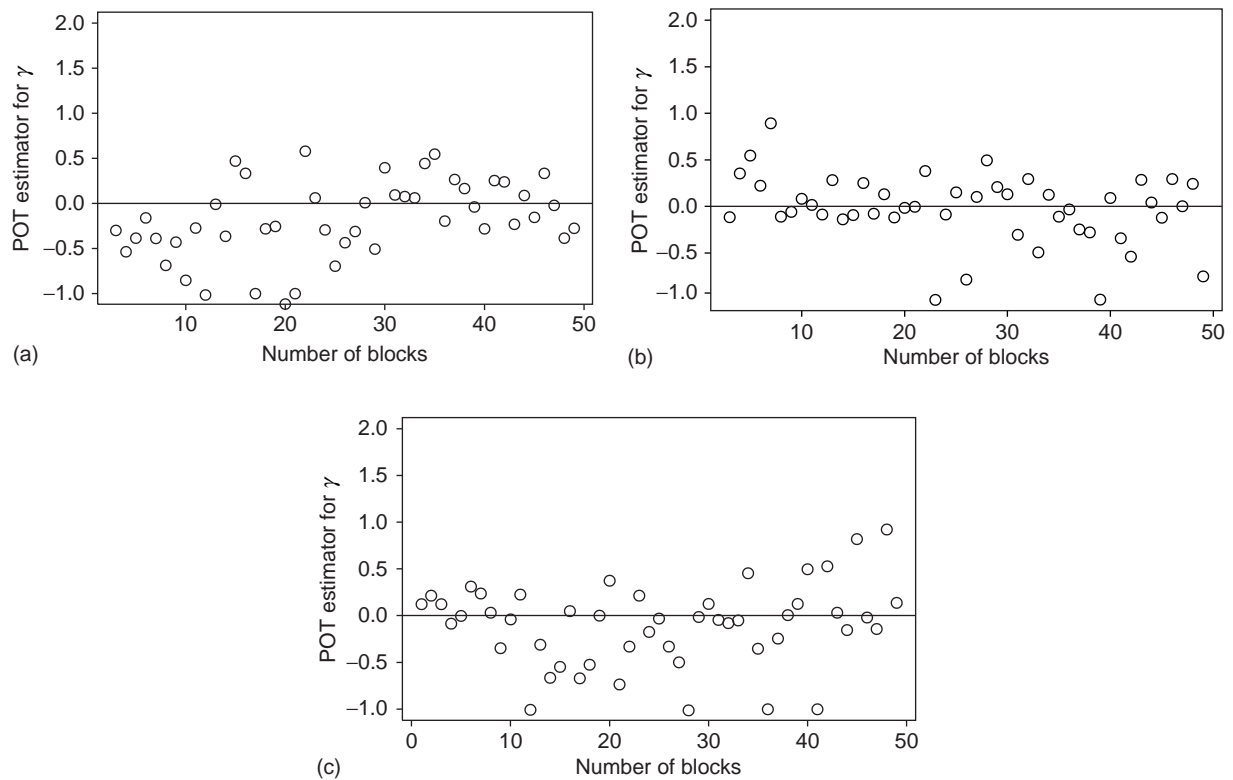


Figure 3 POT estimator of γ using the $k = 17$ largest RTs out of 37 observations for the extreme value index of the reaction times as a function of the number of blocks for: (a) addition, (b) subtraction, and (c) choice.

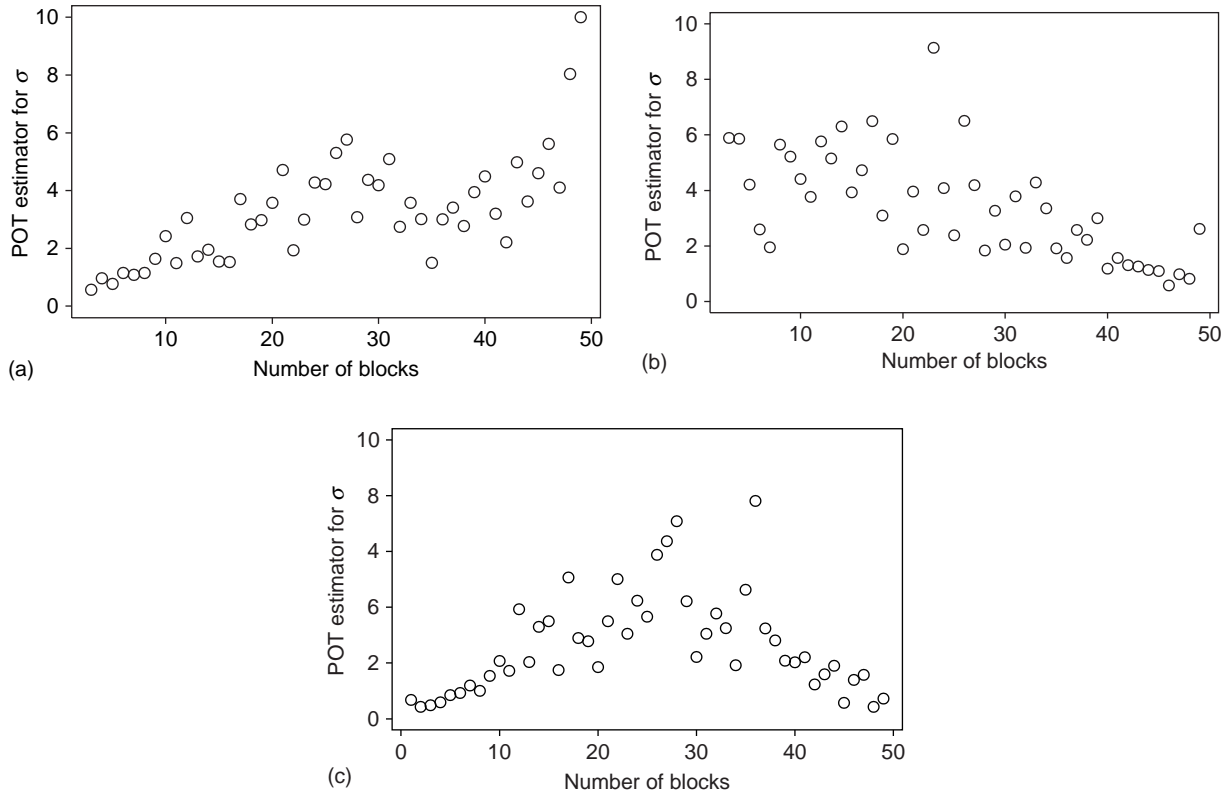


Figure 4 POT estimator of σ using the $k = 17$ largest RTs out of 37 observations for the extreme value index of the reaction times as a function of the number of blocks for: (a) addition, (b) subtraction, and (c) choice.

they decline linearly for the subtraction strategy. For the C-condition the scale estimates again follow the pattern of the original data plot as discussed in **Figure 1** with a change point at about 25 blocks.

Given estimators $\hat{\gamma}_k$ and $\hat{\sigma}_k$, how can one now estimate the exceedance probability p_x of obtaining an observation larger than a given large value x ? To this end the conditional probability of an excess being larger than x (where we restrict attention to the excess data and which is computed along the postulated GPD excess model) is to be multiplied with the probability that an observation is larger than the threshold. For the latter we assume that the threshold is situated deep enough in the sample so that the sample proportion can be used here as an estimate. This fraction equals k/n if the threshold t equals the $(k + 1)$ th largest observation. This leads to

$$\hat{p}_x = \frac{k}{n} \left(1 + \hat{\gamma} \frac{x - t}{\hat{\sigma}} \right)^{-1/\hat{\gamma}} \quad [2]$$

Moreover, an extreme quantile q_p which denotes the outcome level which is exceeded with a chosen probability p (typically smaller than $1/n$) can be estimated by equating [2] to p and solving for x :

$$\hat{q}_p = t + \hat{\sigma} \frac{(k/(np))^{\hat{\gamma}} - 1}{\hat{\gamma}} \quad [3]$$

Pareto-Type Distributions

In the special case $\gamma > 0$, the underlying distribution is member of the class of Pareto-type distributions. In this case the POT-GPD assumption is equivalent to the assumption that the excesses X/t (rather than $X - t$, $X > t$) for large enough t approximately follow a simple Pareto distribution (PD) with cumulative distribution function $1 - \gamma^{-1/\gamma}$, $\gamma > 1$. Again, using the $(k + 1)$ -largest observation as the threshold t , this leads to the following tail estimators that can be used in case $\gamma > 0$:

$$\hat{p}_x^+ = \frac{k}{n} \left(\frac{x}{t} \right)^{-1/\hat{\gamma}} \quad [4]$$

Remark that this is completely similar to [2] with the GPD replaced by the simple PD and the excess $x - t$ replaced by x/t . Also

$$\hat{q}_p^+ = t \left(\frac{k}{np} \right)^{\hat{\gamma}} \quad [5]$$

Here maximum likelihood estimation fitting the strict PD based on the excesses X/t ($X > t$) leads to the Hill (1975) estimator which equals the average of the logarithms of the excesses.

The use of the GPD in the general case $\gamma \in R$, respectively the PD in case $\gamma > 0$, as a consequence of theoretical probabilistic results concerning the distribution of maxima, should of course be validated in practice. One can validate the use of the fitted parametric models through goodness-of-fit methods. For this purpose for instance quantile plotting techniques can be used.

Quantile Plotting

An alternative view to the above material, which allows one to support an extreme value analysis graphically, consists of plotting the ordered data against the corresponding theoretical quantiles of the postulated model. For instance in the specific case of Pareto-type distributions ($\gamma > 0$), the model can be evaluated through the ultimate linearity of a Pareto quantile plot where the logarithm of the j th smallest observation is plotted against the standard exponential theoretical quantile $-\log\left(1 - \frac{j}{n+1}\right)$ ($1 \leq j \leq n$).

In **Figure 5** the Pareto quantile plots for the RTs to count 30 blocks at each of the three conditions are shown. These plots confirm the Pareto-type hypothesis for the corresponding reaction times as indeed above a certain level linearity becomes apparent. In **Figure 6** for

each condition the Hill estimators for $\gamma > 0$ at $k = 17$ are given as a function of the block size.

From **Figures 1–6** the problems arising with application of extreme value methods become apparent. First the choice of k is quite important. The higher the value k (or, the lower the threshold t) the higher the bias of the estimator since the models are only valid for threshold values approaching the endpoint of the underlying distribution. The smaller the k the larger the variance of the estimators because then these are based on fewer data. Also theoretically one can show that optimal k values are different for different estimators (such as the GP-POT and the Hill estimator). Also when only few data are available, such as is the case here with only 37 observations per number of blocks and per condition, the methods can be quite sensitive and yield different conclusions. In fact, the GP-POT approach here indicates a more nuanced picture concerning the signs and values of the EVI estimates, while the positive Hill estimates taken at $k = 17$ indicate EVI values around 0.2 for all numbers of blocks and all conditions. Also the GP-POT approach offers a more flexible framework offering two parameters in order to model the exceedances. In **Figure 7** we offer the GP quantile plots for the RTs corresponding to the counts of 30 blocks based on $k = 17$ exceedances for each of the three conditions. From

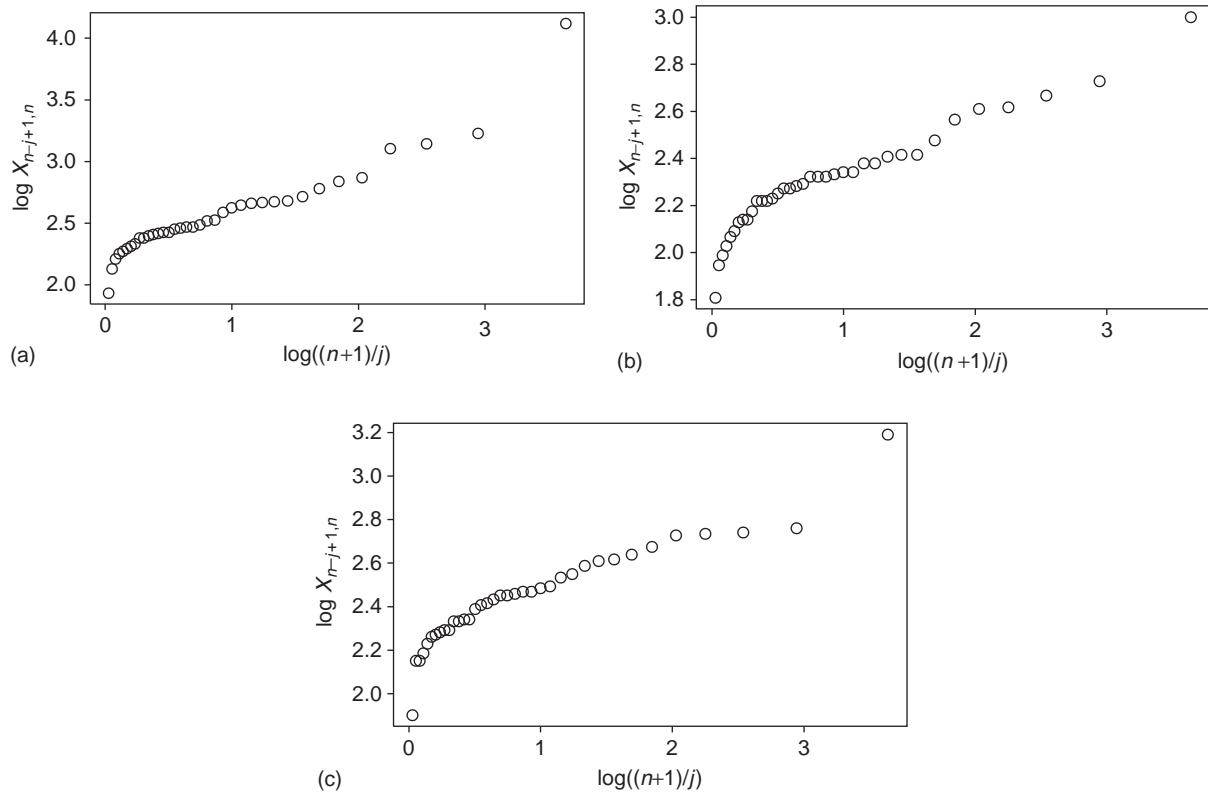


Figure 5 Pareto QQ plots for the reaction times to count 30 blocks. The data are based on results for 37 participants: (a) addition, (b) subtraction, and (c) choice.

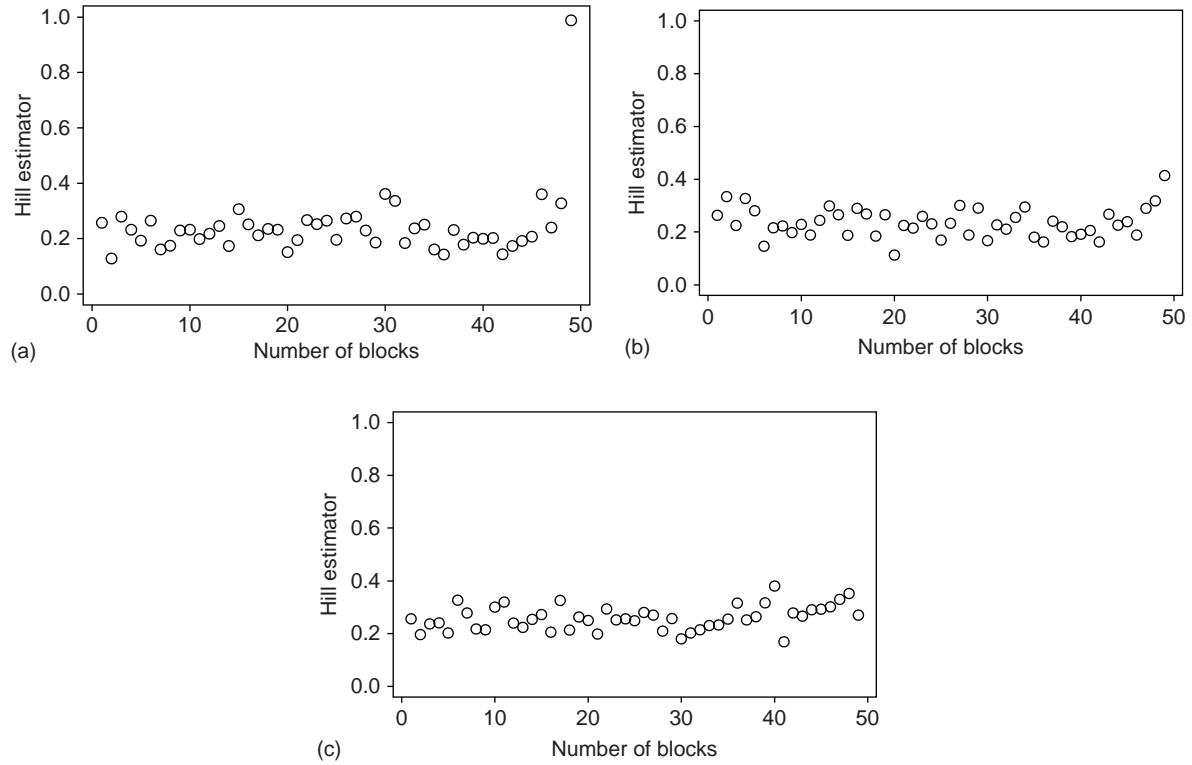


Figure 6 Hill estimator values when using the $k = 17$ largest RTs out of 37 observations for the extreme value index of the reaction times as a function of the number of blocks: (a) addition, (b) subtraction, and (c) choice.

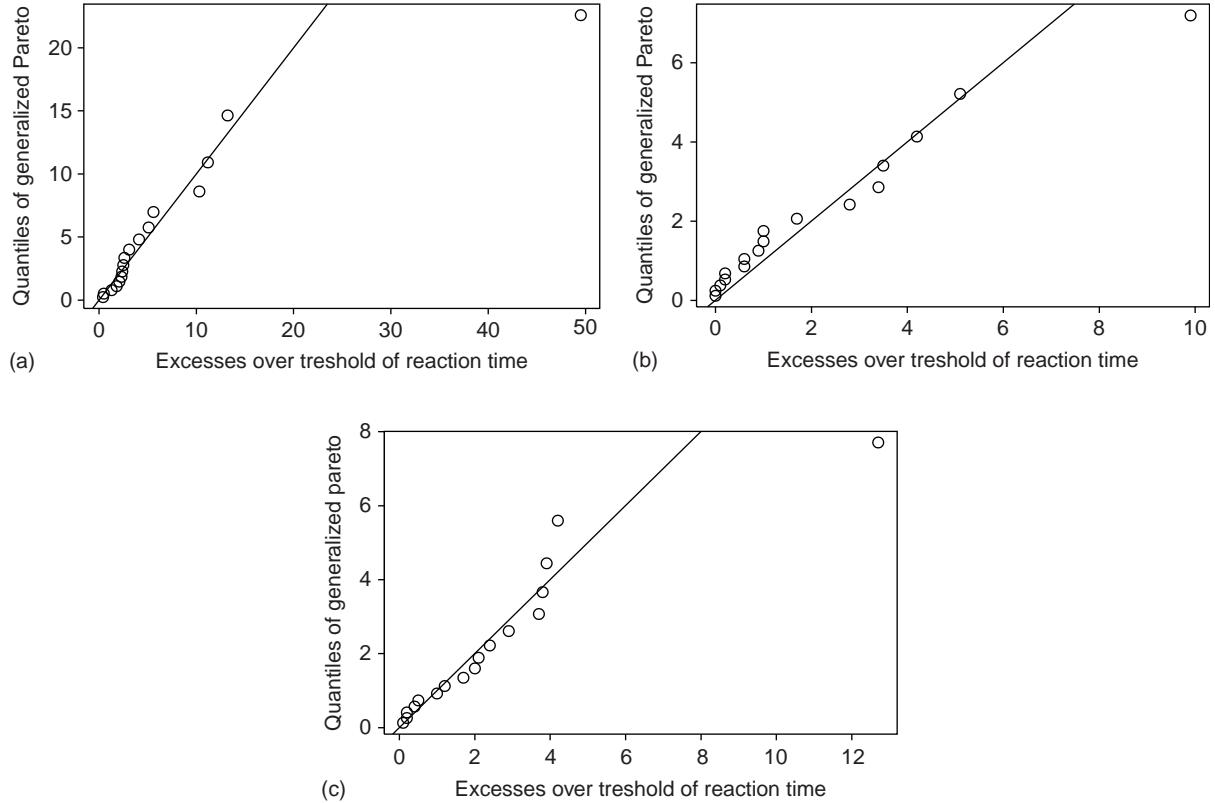


Figure 7 Generalized Pareto QQ plots for the excesses of the reaction times to count 30 blocks over the threshold corresponding to $k = 17$. The estimated γ and σ values for $k = 17$ are used: (a) addition ($\hat{\gamma} = 0.395, \hat{\sigma} = 4.186$), (b) subtraction ($\hat{\gamma} = 0.131, \hat{\sigma} = 2.048$), and (c) choice ($\hat{\gamma} = 0.125, \hat{\sigma} = 2.214$).

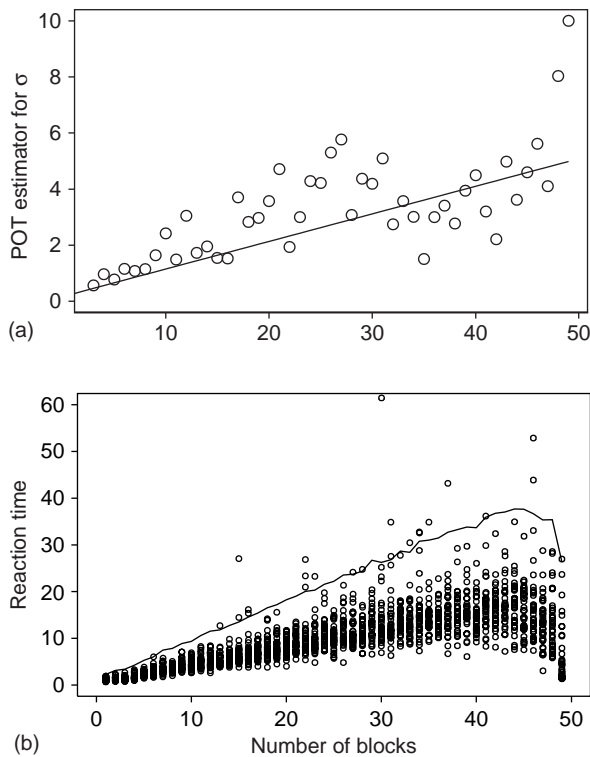


Figure 8 (a) The estimated linear σ -curve and (b) the 0.99 quantile curve based on regression modeling with the original data for the addition strategy data.

this we infer that the GP model appears accurate apart from the largest observation for each of the cases.

Extensions

Our case study motivates the use of regression models which allow one to retain the flexible modeling of the underlying distribution tails as proposed by the maximum domain of attraction framework. This then allows one to use more data points in order to estimate only a limited number of parameters, and hence keep a decent number of degrees of freedom. Davison and Smith (1990) provide an interesting regression modeling approach within the GP-POT modeling framework.

In the case of the addition strategy, we used maximum likelihood estimation to fit the GP distribution with parameters γ and $\sigma = \sigma_0 + \sigma_1 \times (\text{the number of blocks})$. For each number of blocks the 17 largest data points were used so that we used 49×17 excess data. This leads to the estimates $\hat{\gamma} = 0.096$, $\hat{\sigma}_0 = 0.167$, and $\hat{\sigma}_1 = 0.098$. Then

using the GPD as an excess model with σ substituted by the estimated regression function allows one to estimate a high quantile curve as a function of the number of blocks for the addition condition. In Figure 8 the estimated linear σ function is plotted with the original simple σ estimates as shown in Figure 4, next to the extreme quantile curve [3] corresponding with $p = 0.01$ with the original data under the addition strategy.

A Literature Review

In the last few decades extreme value methods in regression, multivariate, and time series settings have been developed and studied. More refined models designed to account for temporal dependence and multivariate data in extreme value analysis include Poisson process models, Markov chain models, and extreme value copulas. Another important issue is the choice of the threshold defining the excesses to which to fit the (generalized) PD. Recently developed techniques offer a solution through stabilizing the estimates over a large range of thresholds.

See also: Exploratory Data Analysis; Goodness-of-Fit Testing; Nonparametric Statistical Methods; Order Statistics.

Bibliography

- Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with comments). *Journal of the Royal Statistical Society, Series B* **52**, 393–442.
- Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3**, 1163–1174.
- Luwel, K., Verschaffel, L., Onghena, P., and De Corte, E. (2003). Analysing the adaptiveness of strategy choices using the choice/no-choice method: The case of numerosity judgement. *European Journal of Cognitive Psychology* **15**, 511–537.

Further Reading

- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004). *Statistics of Extremes – Theory and Applications*. Wiley Series in Probability and Statistics. New York: Wiley.
- Coles, S. G. (2001). *An Introduction to Statistical Modelling of Extreme Values*. London: Springer.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. New York: Springer.
- Reiss, R. D. and Thomas, M. (2001). *Statistical Analysis of Extreme Values*. Basel: Birkhäuser Verlag.
- Weissman, I. (1978). Estimation of parameters and large quantiles based on the k -largest observations. *Journal of the American Statistical Association* **73**, 812–815.

Analysis of Variance

B M King, Clemson University, Clemson, SC, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Alpha – The probability of a type I error (rejection of a true null hypothesis).

Analysis of variance (ANOVA) – A statistical test used to compare the means of three or more populations.

Factor – An independent (treatment) variable.

F ratio – $s^2_{\text{bet}}/s^2_{\text{wi}}$; between-groups variance estimate divided by the within-groups variance estimate.

Grand mean – The mean of all scores in all the treatment conditions.

Interaction effect – In analysis of variance, the joint effect of two or more independent variables that cannot be predicted from the main effects.

Level(s) – Different values of an independent variable (factor).

Main effect – In analysis of variance, the mean differences among the levels of one factor averaged across the levels of the other factor.

Omega-squared – $\hat{\omega}^2_{\text{bet}}$; a measure of effect size for analysis of variance.

Orthogonal comparisons – Comparisons that are unique and do not overlap.

Planned comparisons – Statistical tests used in place of ANOVA when the researcher knows in advance which comparisons are of interest.

Post hoc comparisons – Statistical tests performed after obtaining a significant value of F that indicates which means are significantly different.

Sum of squares (SS) – Sum of the squared deviations from the mean.

Type I error – Rejection of a true null hypothesis.

Imagine that you are superintendent of a school district and you wish to compare the mathematics competency of students enrolled in seven high schools. You randomly select 30 seniors from each school and give them a mathematics achievement test. Your null hypothesis is that the mean mathematics achievement scores in the seven populations (the seven high schools) are equal, that is, there is no difference in scores among the schools. How should you analyze the results? If you had decided to test for differences between all pairs of means with use of a t test, you would have to conduct 21 tests. Suppose that you had

decided in advance to set the probability of a type I error (rejection of a true null hypothesis) at 0.05. This means that when testing for a difference between two population means, in 20 replications of the same study (drawing independent random samples of the same size each time) you would expect the results of one of the t tests to be statistically significant even if there is no difference between the population means. Although for any one test, alpha (probability of a type I error) is 0.05, with 21 independent tests the probability of making at least one type I error is $p = 1 - (0.95)^{21} = 1 - 0.34 = 0.66$. Thus, if you found a statistically significant difference for one or two of your 21 tests, there is a high probability that the result(s) was due to a type I error rather than to a true difference between groups.

You need a test that allows you to make multiple comparisons among means while holding alpha steady at a preselected level. To this end, Ronald Fisher (1890–1962) developed analysis of variance (ANOVA) in the 1920s. An interesting feature of ANOVA is that although it is a test for differences among population means, it can be conducted without calculating the value of any of the sample means.

The Logic of ANOVA

When testing for differences among population means, variation in scores can be due to two things: chance and treatment effect. As the scores in any one sample are drawn randomly from the same population, the scores can differ from one another, and from the sample mean, only due to chance. Scores in different samples, as well as sample means, can differ from one another due to chance, but also because the experimental treatment has an effect. If there is no treatment effect, scores in different groups will differ only because of chance. Thus, when the null hypothesis (there is no treatment effect) is true, the variation of scores around the sample means within each sample ($X - \bar{X}$) and the variation of sample means from the grand mean ($\bar{X} - \bar{X}$; the grand mean, \bar{X} , is the mean of all scores) will both be a reflection of variation in scores due to chance factors.

For any study of two or more samples, the amount by which a score differs from the grand mean ($X - \bar{X}$) can be partitioned into two parts: the amount by which the score differs from its own sample mean ($X - \bar{X}$) and the amount by which the sample mean differs from the grand

mean $(\bar{X} - \bar{\bar{X}})$. The partition of $X - \bar{\bar{X}}$ is represented by the following equation:

$$X - \bar{\bar{X}} = (X - \bar{X}) + (\bar{X} - \bar{\bar{X}}) \quad [1]$$

As the name suggests, in ANOVA we are going to work with variance estimates. Variance estimates are calculated by dividing the sum of the squared deviations from the mean (often called sum of squares, and abbreviated SS) by the degrees of freedom (df). We can calculate the sum of squares total (SS_{total}) for two or more samples by calculating the deviation from the grand mean for every score, squaring each deviation score, and then summing (\sum) all the squared deviations from the grand mean:

$$SS_{\text{total}} = \sum_{\text{all scores}} (X - \bar{\bar{X}})^2$$

Look again at eqn [1]. You can see that we could also calculate SS_{total} by squaring $(X - \bar{X}) + (\bar{X} - \bar{\bar{X}})$:

$$\sum_{\text{all scores}} (X - \bar{\bar{X}})^2 = \sum_{\text{all scores}} (X - \bar{X})^2 + \sum_{\text{all scores}} n(\bar{X} - \bar{\bar{X}})^2 \quad [2]$$

where:

n = the number of scores in a particular group
 k = the number of samples

When we square $(a + b)$ there is a middle term, here $2(\bar{X} - \bar{\bar{X}})(X - \bar{X})$, but in this case it is always zero. What has been accomplished is the partitioning of the total sum of squares into two parts: (1) $\sum_{\text{all scores}} (X - \bar{X})^2$, called SS_{within} (or SS_{error}), a reflection only of chance factors when drawing samples, and (2) $\sum_{\text{all scores}} n(\bar{X} - \bar{\bar{X}})^2$, called SS_{between} , a reflection of both chance factors and treatment effect (when the population means are different). Thus,

$$SS_{\text{total}} = SS_{\text{within}} + SS_{\text{between}}$$

We can now divide the separate sums of squares by their respective degrees of freedom to obtain variance estimates. The degrees of freedom for SS_{total} is the total number of scores minus one: $n_{\text{total}} - 1$. For SS_{within} (abbreviated SS_{w}), the degrees of freedom is calculated by adding $n-1$ for each sample: $n_{\text{total}} - k$. For SS_{between} (abbreviated SS_{bet}), degrees of freedom is the number of samples minus one: $k - 1$.

The end result is two variance estimates. One is the within-groups variance estimate, $s_{\text{w}}^2 (= SS_{\text{w}}/df_{\text{w}})$, which is an estimate of the variance due to chance. It is sometimes called mean square within or mean square error. The other is the between-groups variance estimate, $s_{\text{bet}}^2 (= SS_{\text{bet}}/df_{\text{bet}})$, which is an estimate of the variance due to chance plus any treatment effect. When there is no treatment effect, both are estimates of the same thing. Thus, when the null hypothesis is true (i.e., there is no treatment effect), the ratio $s_{\text{bet}}^2/s_{\text{w}}^2$, called the F ratio,

should be about 1.0. When the null hypothesis is false, s_{bet}^2 will exceed s_{w}^2 , and the greater the treatment effect, the more is the amount by which s_{bet}^2 will exceed s_{w}^2 . Due to sampling variation, we expect the value of the F ratio to be different with every replication of a study (employing new randomly drawn samples of the same size). However, if we set $\alpha = 0.05$ and obtain a value of F that would have occurred by chance less than 5% of the time (if the null hypothesis were true) we reject the null hypothesis and conclude that there is a treatment effect (i.e., the treatment-effect reliably exceeds effects due to chance).

Computation of F : An Example

Let us look at an example. Suppose that we wish to compare two new methods for teaching mathematics with the standard method. We randomly draw a sample of ten subjects for each method and use each method for 1 year, at the end of which we administer a mathematics achievement test to each subject. The hypothetical results are as follows:

Standard method (X)	Method 2 (Y)	Method 3 (Z)
68	85	62
70	90	76
75	80	74
65	78	70
55	82	58
80	86	78
72	92	74
78	76	80
60	94	65
74	84	75
$\sum X = 697$	$\sum Y = 847$	$\sum Z = 712$
$\bar{X} = 69.7$	$\bar{Y} = 84.7$	$\bar{Z} = 71.2$
$\bar{\bar{X}} = 2256/30 = 75.2$		

We now calculate SS_{total} , SS_{w} , and SS_{bet} :

$$\begin{aligned}
 SS_{\text{total}} &= (68 - 75.2)^2 + (70 - 75.2)^2 + \cdots + (65 - 75.2)^2 \\
 &\quad + (75 - 75.2)^2 = 2722.8 \\
 SS_{\text{w}} &= (68 - 69.7)^2 + (85 - 84.7)^2 + (62 - 71.2)^2 \\
 &\quad + (70 - 69.7)^2 + (90 - 84.7)^2 + (76 - 71.2)^2 \\
 &\quad + \cdots + (74 - 69.7)^2 + (84 - 84.7)^2 + (75 - 71.2)^2 \\
 &= 1357.8 \\
 SS_{\text{bet}} &= 10(69.7 - 75.2)^2 \\
 &\quad + 10(84.7 - 75.2)^2 \\
 &\quad + 10(71.2 - 75.2)^2 \\
 &= 1365
 \end{aligned}$$

There are formulas to calculate SS_{total} , SS_{bet} , and SS_{w} that are easier to use than using deviations from means,

but they do not provide you with a basic understanding of ANOVA (see King and Minium, 2008). (It is these formulas that allow you to compute F without ever having calculated a single mean.) It is customary to place the results in a summary table, often called an ANOVA table:

Source	SS	df	s^2	F
Between groups	1365	2	682.5	13.571
Within groups	1357.8	27	50.29	
Total	2722.8	29		

Be sure to place between groups first, above within groups, to set up the proper ratio. As a check on your work, SS_{bet} plus SS_{w} should equal SS_{total} , and df_{bet} plus df_{w} should equal df_{total} (s^2_{w} plus s^2_{bet} do not equal s^2_{total}).

In order to determine if our obtained value of F is statistically significant, we must use a table for the F distribution, found in any introductory statistics textbook. The table gives critical values of F . When our obtained value of F is equal to or greater than the critical value, it indicates that our result is statistically significant (i.e., has a low probability of occurring by chance). The F distribution varies as a function of df_{bet} and df_{w} . Critical values of F are found at the intersection of df for the numerator (2 in our example) and df for the denominator (27 in our example), and in this case the critical values are 3.35 for $\alpha = 0.05$ and 5.49 for $\alpha = 0.01$. Thus, in our example, the probability of a type I error was less than 0.01, that is, the result is statistically significant. In addition to your F statistic and p -value, you should also provide a measure of effect size when reporting results. One commonly used measure for independent-groups ANOVA is omega-squared ($\hat{\omega}^2_{\text{bet}}$) (see King and Minium, 2008). It gives us a population-based, rather than a sample-based, estimate of how much of the proportion of the variance in the dependent variable was attributed to the different levels of treatment.

Interpreting a Significant F Value

Independent-groups ANOVA can be used with two samples, in which case F is the square of the t -statistic that compares the two sample means. A statistically significant result indicates that one population mean is either less than or greater than the other. What does it mean when we obtain a statistically significant value of F for three or more samples? In this case it tells us only that there is a difference among the populations. It does not tell us the manner in which they differ. For three groups, all three population means could be different from one another, or one could be greater than the other two, etc. To determine which means are significantly different from others, we normally use *post hoc* (*a posteriori*) comparisons. Some of the most commonly used tests are Duncan's multiple-range test, the Newman-Keuls test, Tukey's HSD test,

and the Scheffé test. Duncan's test is the least conservative with regard to type I error and the Scheffé test is the most conservative. An explanation of these tests is beyond the scope of this article, but most textbooks will provide a full explanation of one or more of them. However, before you can use any of them you must first have obtained a significant value of F . In our example, all four *post hoc* tests would reveal that teaching method 2 is superior to the other two methods, which did not significantly differ from one another.

There are some underlying assumptions associated with the use of ANOVA. The first is that the populations from which the samples are drawn are normally distributed. Moderate departure from the normal bell-shaped curve does not greatly affect the outcome, especially with large-sized samples (Glass *et al.*, 1972). However, results are much less accurate when populations of scores are very skewed or multimodal (Tomarken and Serlin, 1986), which is frequently the case in the behavioral sciences (Micceri, 1989). In this case, you should consider using the Kruskal-Wallis test, an assumption-freer (nonparametric) test for the independent-groups design (see King and Minium, 2008). This is especially true when using small samples. A second assumption is that of homogeneity of variance, that is, the variances in the populations from which samples are drawn are the same. However, this is a major problem only when variances differ considerably, and is less of a problem if you use samples that are of the same size (Milligan *et al.*, 1987; Tomarken and Serlin, 1986).

The Repeated-Measures Design

ANOVA can also be used with the repeated-measures design (e.g., testing the same subjects under two or more conditions). The major difference from the independent-groups design is that SS_{w} is partitioned into two parts: (1) SS_{subjects} , which is a reflection of the variability in scores resulting from individual differences, and (2) SS_{residual} , a reflection of the variability in scores due to chance. Although three variance estimates can be calculated (s^2_{subjects} , s^2_{bet} , and s^2_{resid}), normally we are interested only in the F ratio $s^2_{\text{bet}}/s^2_{\text{resid}}$ to determine if there is a statistically significant difference among groups. (We are usually not interested in whether or not individuals differ.)

Two-Way ANOVA

In the previous example, we examined the effects of three different teaching methods on mathematics achievement scores using an independent-groups design. There was only one treatment variable – teaching method. However, ANOVA allows us to study two or more treatment variables (called factors) simultaneously. Suppose, for example, that

we were also interested in whether the amount of time devoted to teaching mathematics affected achievement scores. Rather than conducting two separate studies, we can study both factors (teaching method and time spent teaching) simultaneously using a two-factor ANOVA design. If we study all three teaching methods (referred to as levels of the factor teaching method) and two different durations of time spent teaching (two levels; e.g., 45 min per day vs. 90 min), we have a 3×2 design. Let us suppose that we select eight students randomly for each of the six independent conditions (standard teaching method for 45 min per day, standard teaching method for 90 min per day, etc.) and after 1 year we obtain the following mathematics achievement scores:

Teaching method

Teaching time per day	Standard method	Method 2	Method 3	
45 min	60	73	62	$\bar{X}_{45} = 72.67$
	80	87	74	
	68	76	63	
	72	84	73	
	64	83	64	
	76	77	72	
	69	81	67	
	71	79	69	
90 min	$\bar{X} = 70$	$\bar{X} = 80$	$\bar{X} = 68$	$\bar{X}_{90} = 80.67$
	65	92	63	
	81	106	77	
	68	103	70	
	78	95	68	
	70	97	70	
	76	101	72	
	72	99	67	
	74	99	73	$\bar{X} = 76.67$
	$\bar{X} = 73$	$\bar{X} = 99$	$\bar{X} = 70$	
	$\bar{X}_{\text{std}} = 71.5$	$\bar{X}_2 = 89.5$	$\bar{X}_3 = 69$	

The six combinations of rows and columns are called cells. The six cell means, two row means, three column means, and the grand mean are provided. In two-way ANOVA we want to know if there is a main effect for any of the factors, or, in other words, if there are differences in the means of the levels of one factor averaged across the levels of the other factor. In our example, there are two possible main effects and the essential questions can be phrased as: (1) Is there an overall influence of teaching method (for both the 45-min and 90-min conditions), and (2) is there an overall influence of time devoted to teaching (across all three teaching methods)?

There is an equally, if not more, important question that we can examine with a factorial design: Is there an interaction effect – are the differences among the levels of one factor the same for all levels of the other factor? To

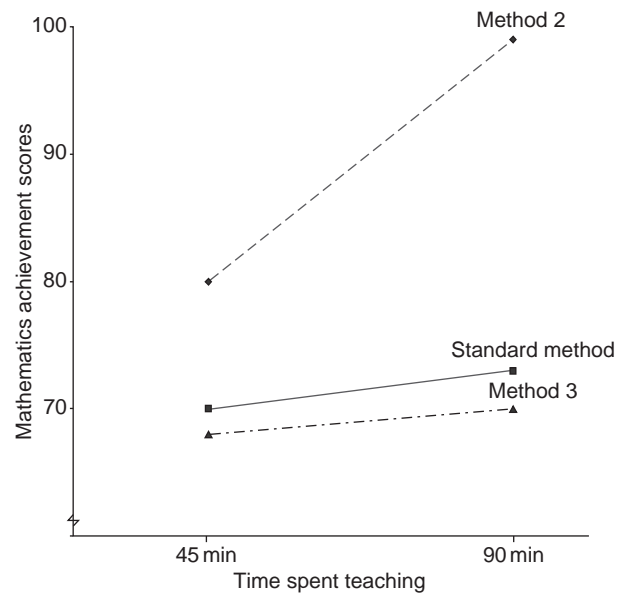


Figure 1 Hypothetical scores in a 3×2 factorial study of the effects of teaching method and time spent teaching on mathematics achievement scores.

better understand this, let us first look at the results of our study in graph form (Figure 1).

In our hypothetical study, if an increase in time spent teaching mathematics had an equal effect for all three teaching methods, the three lines would be parallel. A sizable gap between the lines indicates a possible main effect for teaching method. In our graph, we see evidence for a main effect, but the difference between method 2 and the other two teaching methods depends on the amount of time it is used in the classroom. Method 2 results in moderately better mathematics achievement scores when used for 45 min per day, but dramatically better scores when students are taught mathematics by this method for 90 min per day. If there had been a main effect for teaching method (method 2 better than the other two) without an interaction effect, the differences between method 2 and the others would have been the same for both levels of the other factor (teaching time per day). A significant interaction effect tells us that before we can interpret the effects of a factor, we must examine that factor at each level of the other factor. This will, of course, involve examining cell means. Three-factor experimental designs (e.g., teaching method \times time spent teaching \times gender) and even higher-order designs are possible, but interpretation of interaction effects becomes more difficult.

When calculating two-way ANOVA, SS_{total} is partitioned into four parts. SS_{wc} (within cell) is the equivalent of SS_{w} (within groups) in the one-factor design. SS_{bet} is split into three parts: SS_{rows} , SS_{columns} , and $SS_{\text{rows} \times \text{columns}}$. To see how to calculate two-way ANOVA, see King and Minium (2008). In our example, we obtain a significant

effect for rows ($F = 30.78$, $df = 1/42$, $p < 0.001$), columns ($F = 80.21$, $df = 2/42$, $p < 0.001$), and rows \times columns interaction ($F = 14.59$, $df = 2/42$, $p < 0.001$). As at least one of our main effects was statistically significant, we may now conduct a *post hoc* test to see which levels differed from the others (e.g., method 2 vs. the other two).

Planned Comparisons

When conducting any statistical test one wants the test to have considerable power, that is, a high probability of rejecting a false null hypothesis. *Post hoc* tests allow us to make all possible pair-wise comparisons, but to protect us from making type I errors, the differences between means must be large enough to be declared statistically significant. If one's study is exploratory in nature, then ANOVA is a very good statistical test. However, if one knows in advance of conducting a study which comparisons are important to him or her, one should use planned (*a priori*) comparisons in place of ANOVA and *post hoc* comparisons (Rosnow and Rosenthal, 1989; Wilkinson *et al.*, 1999; Winer *et al.*, 1991). To do so, one should conduct independent-groups *t* tests using s^2_w (calculated from all groups), but to protect against type I errors, all the comparisons must be orthogonal (unique, with no overlap). When comparing k groups, there are only $k-1$ orthogonal comparisons. In our example, we might be interested in the two comparisons standard method versus (method 2 + method 3) and method 2 versus method 3. Unlike *post hoc* comparisons, one does not need a significant overall F in order to use planned comparisons. Even Ronald Fisher recognized the value of planned comparisons:

When the [F] test does not demonstrate significant differentiation, much caution should be used before claiming

significance for special comparisons. Comparisons, which the experiment was designed to make, may, of course, be made without hesitation (Fisher, 1949: 57).

Bibliography

- Fisher, R. A. (1949). *The Design of Experiments*, 5th edn. London: Oliver and Boyd.
- Glass, G. V., Peckham, P. D., and Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. *Review of Research in Education* **42**, 237–288.
- King, B. M. and Minium, E. W. (2008). *Statistical Reasoning in the Behavioral Sciences*, 5th edn. Hoboken, NJ: Wiley.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin* **105**, 156–166.
- Milligan, G. W., Wong, D. S., and Thompson, P. A. (1987). Robustness properties of nonorthogonal analysis of variance. *Psychological Bulletin* **101**, 464–470.
- Rosnow, R. L. and Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist* **44**, 1276–1284.
- Tomarken, A. J. and Serlin, R. C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. *Psychological Bulletin* **99**, 90–99.
- Wilkinson, L. and The Task Force on Statistical Inference (1999). Statistical methods in psychology journals. Guidelines and explanations. *American Psychologist* **54**, 594–604.
- Winer, B. J., Brown, D. R., and Michels, K. M. (1991). *Statistical Principles in Experimental Design*, 3rd edn. New York: McGraw Hill.

Further Reading

- Aron, A., Aron, E. N., and Coups, E. J. (2009). *Statistics for Psychology*, 5th edn. Upper Saddle River, NJ: Prentice Hall.
- Gravetter, F. J. and Wallnau, L. B. (2008). *Essentials of Statistics for the Behavioral Sciences*, 6th edn. Belmont, CA: Thomson Wadsworth.
- King, B. M. and Minium, E. W. (2008). *Statistical Reasoning in the Behavioral Sciences*, 5th edn. Hoboken, NJ: Wiley.

Bayesian Statistical Analysis

G Kokolakis, National Technical University of Athens, Athens, Greece

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Bayesian statistics has been considered, for quite a long time, as a branch of statistics; however, its role in the development of statistical inference is much more profound than that. Its philosophical base traces back to the very initial and rather subjective interpretation of the notion of probability and its use in everyday human activities. It is related with the notion of ‘pithanon,’ as used by Carneades (214–129/8 BC), meaning persuasive, convincing, which allows human beings to take decisions and act without necessarily having perfect knowledge of the situation and conditions they are facing. For a detailed account on Carneades’ philosophical position, the reader is referred to Bett (1989).

According to the Bayesian approach to statistics, statistical inference is not just about collecting, analyzing, and interpreting data that have an endogenous variability, which is the frequentist view of statistics. Rephrasing de Finetti (1974, 1975), it is about developing a plausible way of studying uncertain events through the notion of probability. It also provides a coherent methodology for inductive mathematical reasoning when new information is available. In short, statistics is the study of uncertainty.

Savage (1954) and Lindley (1965, 2000) state that the only self-consistent (coherent) way to deal with uncertainty is by expressing it through the notion of probability. When dealing with a statistical problem, uncertainty refers to data, parameters, and models. Having expressed the uncertainty about these, the rest concerns taking into account the information provided by the data and updating the uncertainty according to the rules of probability calculus. The basic mathematical tool to do it is provided by the Bayes’ theorem. Thus, the Bayesian approach provides a unified and coherent inferential procedure.

It must be mentioned here that the rules of probability calculus, namely: (a) the addition rule (finite or countable), (b) the product rule, and (c) the convexity rule (that the probability lies in the convex unit interval taking the value 1 for the sure event), have to be demonstrated on the basis of axioms about the logical and, at the same time, subjective human behavior in an uncertain environment. There exist several axiomatic systems resulting in the above three probability rules. For a discussion of such axiomatic systems, see Savage (1954), DeGroot (1970), and Fishburn (1986).

The Bayesian approach, although quite simple in principle, is often quite demanding, especially when dealing

with complicated and highly structured statistical models. There are two main tasks to be accomplished: (a) modeling uncertainty about the probabilistic behavior of the statistical data at hand and the parameters involved, including models, and (b) applying very efficient computational methods in order to estimate the parameters and confirm the models. Regarding the first task, much of the statistical work is based on the notions of coherence and exchangeability (both due to de Finetti (1937)), and it is greatly simplified by using the so-called objective priors (diffuse, flat, reference or noninformative priors). Regarding the second, the development of modern and quite efficient computational techniques, such as Markov-chain Monte Carlo (MCMC), has made this task a rather straightforward procedure. Their versatility had an enormous impact on the Bayesian approach and its spread in many disciplines during the last two decades.

Bayesian Inference

We express the uncertainty about getting a data value x related with a population parameter θ , both x and θ being in general multivariate. This is achieved by a critical choice of a joint distribution for x and θ , conditional on the state of knowledge we have at this moment about the population under study. An expert opinion, familiar with the specific population, might be very helpful in making such a choice. Specifically, the task is to choose a probability density $p(x, \theta|K)$ (with respect to Lebesgue or counting measure), where K represents the present state of knowledge. In a more general setup, K might also include a set of modeling assumptions and random hyperparameters. In general, given the state of knowledge K , it is more convenient to choose separately the densities $p(x|\theta, K)$ and $p(\theta|K)$, and then apply the product rule to get

$$p(x, \theta|K) = p(x|\theta, K)p(\theta|K) \quad [1]$$

The product rule again allows us to write the density $p(x, \theta|K)$ in the form

$$p(x, \theta|K) = p(\theta|x, K)p(x|K) \quad [2]$$

and thus, combining the above two results, we have

$$p(\theta|x, K) = p(x|\theta, K)p(\theta|K)/p(x|K) \quad [3]$$

If we apply now the addition rule on both sides of [1], with x fixed, we conclude that

$$p(x|K) = \int p(x|\theta, K)p(\theta|K)d\theta$$

and, thus, introducing the latter result into [3], we get the Bayes' formula (Bayes, 1763)

$$p(\theta|x, K) = \frac{p(x|\theta, K)p(\theta|K)}{\int p(x|\theta, K)p(\theta|K)d\theta} \quad [4]$$

The denominator, either as $p(x|K)$ or under the integral form $\int p(x|\theta, K)p(\theta|K)d\theta$, is called the predictive distribution for x . It does not convey any information about the parameter θ (as far as a data value x does not raise doubts about the state of knowledge K itself) and, thus, it acts as a normalizing constant. For simplicity in the notation, the reference to the state of knowledge K is also omitted. We may then write the Bayes' formula [4] as follows:

$$p(\theta|x) \propto p(\theta)p(x|\theta) \quad [5]$$

Formula [5] allows us to go from $p(x|\theta)$ to $p(\theta|x)$; thus, it is also known as the inverse probability law.

The first factor on the right hand side (RHS) of [5] is called a prior, since it expresses the information we have about the parameter θ prior to receiving the data value x . The conditional density $p(\theta|x)$ expresses the updated, according to the Bayes' theorem, distribution of θ after receiving the data value x . Thus, it is the posterior, with reference to x , distribution of θ .

In practice, instead of a single data value x , we have a random sample $\{x_i, i = 1, \dots, n\}$, denoted for simplicity by $x^{(n)}$. Then, $p(x^{(n)}|\theta) = \prod_{i=1}^n p(x_i|\theta)$, and the Bayes' formula takes the form

$$p(\theta|x^{(n)}) \propto p(\theta) \prod_{i=1}^n p(x_i|\theta)$$

The product $\prod_{i=1}^n p(x_i|\theta)$, considered as a function of the parameter θ , is called the likelihood and is denoted by $L(\theta, x^{(n)})$.

The above posterior can be used as the prior for a second dataset $x^{(n^*)} = \{x_{n+i}, i = 1, \dots, n^*\}$. Indeed, the posterior distribution of parameter θ , given the whole dataset $x^{(n+n^*)} = \{x_{n+i}, i = 1, \dots, n+n^*\}$, is

$$p(\theta|x^{(n+n^*)}) \propto p(\theta) \prod_{i=1}^{n+n^*} p(x_i|\theta) \propto p(\theta|x^{(n)}) \prod_{i=1}^{n^*} p(x_{n+i}|\theta)$$

Thus, the posterior of θ , given the first dataset $x^{(n)} = \{x_i, i = 1, \dots, n\}$, is the prior for the next dataset $x^{(n^*)} = \{x_{n+i}, i = 1, \dots, n^*\}$, and so on.

Quite often, the involved parameter θ consists of two parts, namely, $\theta = (\zeta, \eta)$, where the parameter ζ is the one of interest, whereas the parameter η is a nuisance parameter. In such a case, we only have to integrate out the nuisance parameter η from the joint posterior of $\theta = (\zeta, \eta)$. Specifically, the posterior of the parameter of interest ζ is

$$p(\zeta|x^{(n)}) = \int p(\zeta, \eta|x^{(n)})d\eta$$

usually called marginal posterior.

Exchangeability

It is essential to give an explanation about how the prior for a parameter θ emerges. The answer has been given by de Finetti's theorem (de Finetti, 1937), which constitutes the backbone of the Bayesian statistics. The basic notion here is that of exchangeability.

The random variables X and Y are called exchangeable if

$$\begin{aligned} F(x, y) &= P[X \leq x, Y \leq y] = P[Y \leq x, X \leq y] \\ &= F(y, x) \text{ for all } (x, y) \in \mathbb{R}^2 \end{aligned}$$

This means that the pair (X, Y) is distributed as the pair (Y, X) and, thus, the order does not matter. Similarly, a finite sequence of random variables, $\{X_i, i = 1, 2, \dots, n\}$ is called exchangeable if any permutation, $(X_{i_1}, \dots, X_{i_n})$ say, is distributed as (X_1, \dots, X_n) . The notion of exchangeability is then extended to an infinite sequence of random variables $\{X_n, n = 1, 2, \dots\}$, say, by requiring that every finite subsequence of it is exchangeable.

Let now $\{X_n, n = 1, 2, \dots\}$ be an infinite sequence of binary random variables, each one of which is taking the value 1 or 0 according to the realization of an event (a success) or not (a failure). Suppose we make the statement that for any $n \in \mathbb{N}$, the probability to have k successes in any subsequence of length n of X 's, $\{X_{i_1}, \dots, X_{i_n}\}$, say, depends only on k and n . In other words, we assume that the random variables X_n ($n = 1, 2, \dots$) are exchangeable. Let $P_{k,n}$ denote the above probability, that is,

$$P_{k,n} = P\left[\sum_{j=1}^n X_{i_j} = k\right], k = 0, 1, \dots, n \quad (n = 1, 2, \dots)$$

Then, the above statement of exchangeability of X 's is tantamount to the following:

There exists a probability measure Π over the interval $[0, 1]$ such that

$$\begin{aligned} P_{k,n} &= \binom{n}{k} \int_0^1 \theta^k (1-\theta)^{n-k} \Pi(d\theta) \\ &= \binom{n}{k} \int_0^1 \theta^k (1-\theta)^{n-k} \pi(\theta) d\theta, \\ k &= 0, 1, \dots, n \quad (n = 1, 2, \dots) \end{aligned}$$

where $\pi(\theta)$ denotes the density of the parameter θ , which represents the unknown probability of success (here, π is the derivative of Π with respect to Lebesgue or counting measure over the interval $[0, 1]$).

Thus, the existence of a prior distribution $\pi(\theta)$, for the probability of success θ , does not arise out of nothing, but it is a consequence of our general statement about X 's, that is, that of exchangeability. In de Finetti (1937), a generalization of the above exchangeability result for random variables taking values in \mathbb{R}^d has also been derived. Parametric, together with nonparametric, versions of de Finetti's theorem can be found in Diaconis and Freedman (1980) and in Bernardo and Smith (1994). Specifically, under additional assumptions of minor importance, when the random variables X_m ($m = 1, 2, \dots$) are exchangeable, there exists a parametric function $F(x|\theta)$ with $\theta \in \Theta$ such that the joint distribution Q of X_i ($i = 1, 2, \dots, m$) has the following form:

$$Q(x_1, \dots, x_m) = \int_{\Theta} \left\{ \prod_{i=1}^m F(x_i|\theta) \right\} \pi(\theta) d\theta, m = 1, 2, \dots$$

It is obvious that the converse statement is true in all the above cases.

Exponential Family and Conjugate Priors

A density $p(x|\theta)$ belongs to the exponential family if it can be written as follows:

$$\begin{aligned} p(x|\theta) &= f(x)g(\theta)\exp\left\{\sum_{j=1}^k \varphi_j(\theta)b_j(x)\right\} \\ &= f(x)g(\theta)\exp\left\{\varphi(\theta)^T b(x)\right\}, \\ x &\in \mathcal{X} \subset \mathbb{R}^d \text{ and } \theta \in \Theta \subset \mathbb{R}^m \end{aligned} \quad [6]$$

Then, the conjugate prior for the parameter θ is of the form

$$\begin{aligned} p(\theta) &\propto \{g(\theta)\}^\eta \exp\left\{\sum_{j=1}^k v_j \varphi_j(\theta)\right\} \\ &= \{g(\theta)\}^\eta \exp\{\mathbf{v}^T \varphi(\theta)\}, \theta \in \Theta \end{aligned} \quad [7]$$

with parameters (hyperparameters) η and $\mathbf{v} = (v_1, \dots, v_k)^T$.

We see, therefore, that when $p(x|\theta)$ belongs to the exponential family, so does its conjugate prior $p(\theta)$. Conjugate priors, although restrictive sometimes, have the advantage of great mathematical simplicity in deriving posteriors in closed form. An additional advantage is that limiting versions of conjugate priors correspond to the so-called objective priors. Objective priors are most often improper, since their integral over the whole parameter space is not equal to one. They are, therefore, incoherent, meaning that they do not satisfy the probability axioms. Despite this fact, their posteriors can be proper, and, thus, coherent. Objective priors play a very important role in objective Bayesian statistics.

Let now $\mathbf{x}^{(n)} = \{\mathbf{x}_i, i = 1, \dots, n\}$, be a random sample from [6]. The likelihood for the parameter θ is then

$$\begin{aligned} L(\theta, \mathbf{x}^{(n)}) &= \prod_{i=1}^n p(x_i|\theta) \propto \{g(\theta)\}^n \exp\left\{\sum_{j=1}^k \varphi_j(\theta) \left[\sum_{i=1}^n b_j(x_i)\right]\right\} \\ &= \{g(\theta)\}^n \exp\left\{\varphi(\theta)^T \mathbf{t}(\mathbf{x}^{(n)})\right\} \end{aligned}$$

where $\mathbf{t}(\mathbf{x}^{(n)}) = (t_1(\mathbf{x}^{(n)}), \dots, t_k(\mathbf{x}^{(n)}))^T$ with

$$t_j(\mathbf{x}^{(n)}) = \sum_{i=1}^n b_j(x_i), j = 1, \dots, k$$

The quantity $\mathbf{t} = \mathbf{t}(\mathbf{x}^{(n)})$ is the sufficient statistic for θ , in other words, it conveys the same information about θ as the whole dataset $\mathbf{x}^{(n)}$.

Applying Bayes' theorem now, the posterior has the same form as the prior. Specifically:

$$\begin{aligned} p(\theta|\mathbf{x}^{(n)}) &\propto \{g(\theta)\}^{\eta+n} \exp\left\{\sum_{j=1}^k \varphi_j(\theta) \left[v_j + \sum_{i=1}^n b_j(x_i)\right]\right\} \\ &= \{g(\theta)\}^{\eta+n} \exp\left\{\varphi(\theta)^T [\mathbf{v} + \mathbf{t}(\mathbf{x}^{(n)})]\right\} \\ &= \{g(\theta)\}^{\eta_n} \exp\left\{\varphi(\theta)^T \mathbf{v}_n\right\}, \theta \in \Theta \end{aligned}$$

with updated parameters, namely, $\eta_n = \eta + n$ and $\mathbf{v}_n = \mathbf{v} + \mathbf{t}(\mathbf{x}^{(n)})$.

In almost all textbooks on Bayesian statistics, there are extensive tables and results for priors within the exponential family together with their marginal posteriors, predictive distributions, and their limiting objective forms, cf. DeGroot (1970), Aitchison and Dunsmore (1975), Robert (2001), Gelman et al. (2003) and Press (2003).

Bayesian Statistical Methods

Bayesian Decision-Making

Having derived the posterior for a parameter θ , all information we have about it is included in $p(\theta|\mathbf{x}^{(n)})$. Thus, if we have to take a decision, the consequences of which are related with the true value of θ , it is logical to make best use of the information we have about it, that is, to make use of $p(\theta|\mathbf{x}^{(n)})$. Within the decision-making context, the parameter θ represents the state of nature, whereas the decision to be taken is called the action and denoted by α . The set of all actions under consideration is denoted by \mathcal{A} . The loss implied by taking an action α , when the true state of nature is θ , is defined by a loss function $L(\alpha, \theta)$, usually bounded from below. Since the true state of nature θ is not known, associated with an action α is the Bayes' expected loss

$$R(\alpha) = E[L(\alpha, \theta)|\mathbf{x}^{(n)}] = \int_{\Theta} L(\alpha, \theta) p(\theta|\mathbf{x}^{(n)}) d\theta \quad [8]$$

It must be mentioned that when the action α affects the distribution of the parameter θ , the above expectation has to be with respect to $p(\theta|\alpha, \mathbf{x}^{(n)})$. According to the above setup,

the best action is that associated with the lowest Bayes' expected loss. Thus, a decision α^* is optimal if and only if

$$R(\alpha^*) = \inf\{R(\alpha) : \alpha \in \mathcal{A}\}$$

Such an action α^* is then called the Bayes' action. In general, α^* constitutes the Bayes' rule. For an extensive presentation of statistical decision theory, see DeGroot (1970), Berger (1985), and Lindley (1985).

Parameter Estimation

When in place of the action space \mathcal{A} we have the parameter space Θ , then an action is a choice of a particular value, say $\tilde{\theta}$, for the parameter θ . In other words, the Bayesian estimation problem is a decision problem, where the Bayes' rule gives the optimal estimate for θ , under the considered loss function $L(\tilde{\theta}, \theta)$. According to [8], the Bayes' estimate for θ is the quantity minimizing the posterior expected risk:

$$R(\tilde{\theta}) = E[L(\tilde{\theta}, \theta) | x^{(n)}] = \int_{\Theta} L(\tilde{\theta}, \theta) p(\theta | x^{(n)}) d\theta \quad [9]$$

It can be easily shown that with the quadratic loss function $L(\tilde{\theta}, \theta) = \|\tilde{\theta} - \theta\|^2$, the optimal estimate for θ is its posterior mean, namely, $E(\theta | x^{(n)})$. Thus, we have

$$\begin{aligned} E[\theta | x^{(n)}] &= \underset{(\theta \in \Theta)}{\operatorname{argmin}} R(\tilde{\theta}) \\ &= \underset{(\theta \in \Theta)}{\operatorname{argmin}} \int_{\Theta} \|\tilde{\theta} - \theta\|^2 p(\theta | x^{(n)}) d\theta \end{aligned}$$

It is well known that the class of loss functions $L(\tilde{\theta}, \theta)$ for which the Bayes' estimate of a parameter θ coincides with the posterior mean, is very large. In fact, any loss function that belongs to the class of Bregman divergences has this property. The converse is also true. For the definition and applications of Bregman Divergences in statistics, see Kokolakis *et al.* (2006).

Estimation of the multinomial parameter

Let Y be a discrete random variable taking values from the set $\{1, 2, \dots, m\}$ with probabilities θ_j ($j = 1, \dots, m$), where $\sum_{j=1}^m \theta_j = 1$. In what follows, for notational convenience, we introduce an m -variate random vector $X = (X_1, \dots, X_m)$ such that $X_j = 1$ when $Y = j$ and $X_j = 0$ otherwise ($j = 1, \dots, m$). Then, the probability distribution of X , given the parameter θ , can be written as follows:

$$\begin{aligned} p(x | \theta) &= \prod_{j=1}^m \theta_j^{x_j} = \exp \left\{ \sum_{j=1}^m x_j \log \theta_j \right\}, x_j \in \{0, 1\} \quad [10] \\ &\text{with } \sum_{j=1}^m x_j = 1 \text{ and} \\ \theta \in S^m &= \left\{ \theta \in \mathbb{R}^m : \theta_j \geq 0 \ (j = 1, \dots, m) \text{ with } \sum_{j=1}^m \theta_j = 1 \right\} \end{aligned}$$

The set S^m above is known as the $(m-1)$ -simplex.

Thus, the multinomial distribution is apparently within the exponential family and its conjugate prior is of the form

$$p(\theta) = k \cdot \prod_{j=1}^m \theta_j^{\alpha_j - 1}, \theta \in S^m \text{ with } \alpha_j > 0 \ (j = 1, \dots, m) \quad [11]$$

known as the Dirichlet distribution and denoted by $D(\alpha)$. The normalizing constant k above is given by

$$\begin{aligned} k &\equiv k(\alpha) = \left\{ \int_{S^m} \prod_{j=1}^m \theta_j^{\alpha_j - 1} d\theta \right\}^{-1} \\ &= \frac{\Gamma(\alpha)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_m)} \text{ with } \alpha = \sum_{j=1}^m \alpha_j \end{aligned} \quad [12]$$

where

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt, \alpha > 0$$

is the gamma function, satisfying the following recursive relation:

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha) \text{ for } \alpha > 0 \text{ with } \Gamma(1) = 1$$

Note that the flat noninformative prior corresponds to taking: $\alpha_1 = \alpha_2 = \dots = \alpha_m = 1$.

It can be easily seen that in order to find the mean of the above distribution, a renormalization procedure has to be applied and, thus, the mean will be expressed as the ratio of two normalizing constants. The same is true for any moment of the above distribution. Specifically, with $\nu = (\nu_1, \nu_2, \dots, \nu_m)^T$, we have

$$E \left[\prod_{j=1}^m \theta_j^{\nu_j} \right] = \frac{k(\alpha)}{k(\alpha + \nu)}, \nu_j = 0, 1, 2, \dots \ (j = 1, \dots, m)$$

and, thus,

$$\begin{aligned} E[\theta_j] &= \frac{\alpha_j}{\alpha}, \operatorname{Var}[\theta_j] = \frac{\alpha_j (\alpha - \alpha_j)}{\alpha^2 (\alpha + 1)}, \text{ and} \\ \operatorname{Cov}[\theta_j, \theta_k] &= -\frac{\alpha_j \alpha_k}{\alpha^2 (\alpha + 1)}, \ (j \neq k) \end{aligned} \quad [13]$$

Let $x^{(n)} = \{x_i, i = 1, \dots, n\}$ be a random sample from the multinomial distribution [10]. Then, the likelihood function is

$$L(\theta, x^{(n)}) = p(x^{(n)} | \theta) = \prod_{i=1}^n p(x_i | \theta) = \prod_{j=1}^m \theta_j^{n_j}$$

where $\mathbf{n} = (n_1, \dots, n_m)^T$ with $n_j = \sum_{i=1}^n x_{ij}$, $j = 1, \dots, m$, the sufficient statistics for the multinomial parameter θ , and $\sum_{j=1}^m n_j = n$. Multiplying by the prior [11], we get

$$p(\theta | x^{(n)}) = k(\alpha^{(n)}) \prod_{j=1}^m \theta_j^{\alpha_j^{(n)} - 1}, \theta \in S^m$$

where $k(\alpha^{(n)})$ is as in [12] with updated parameters $\alpha^{(n)} = \alpha + \mathbf{n}$ and $\alpha^{(n)} = \alpha + n$.

Using the quadratic loss function, the optimal estimate for θ is its posterior mean. With the updated parameter $\alpha^{(n)}$, the first result in [13] implies that

$$E[\theta_j | \mathbf{x}^{(n)}] = \alpha_j^{(n)} / \alpha^{(n)} = \frac{\alpha_j + n_j}{\alpha + n}, j = 1, \dots, m \quad [14]$$

It is interesting to write the posterior mean in the following form:

$$E[\theta | \mathbf{x}^{(n)}] = \frac{\alpha}{\alpha + n} \times E[\theta] + \frac{n}{\alpha + n} \times \bar{X}_n \quad [15]$$

that is, the posterior mean of θ is a convex combination of the prior mean $E[\theta]$ and its sample mean \bar{X}_n .

The mean squared errors of the above estimates are given by the associated risks, which, with the quadratic loss function used here, coincide with the posterior variances. Thus, using the second result in [13], we have

$$\begin{aligned} \text{Var}[\theta_j | \mathbf{x}^{(n)}] &= \frac{\alpha_j^{(n)} (\alpha_j^{(n)} - \alpha_j^{(n)})}{\alpha_n^2 (\alpha_n + 1)} \\ &= \frac{(\alpha_j + n_j)(\alpha - \alpha_j + n - n_j)}{(\alpha + n)^2 (\alpha + n + 1)} \\ &(\rightarrow 0 \text{ as } n \rightarrow \infty) (j = 1, \dots, m) \end{aligned} \quad [16]$$

It is obvious, therefore, that when the quantity $\alpha = \sum_{j=1}^m \alpha_j$ is very small compared to the sample size $n = \sum_{j=1}^m n_j$, the above estimates of the components of the parameter vector θ are almost equal to their frequentist estimates and they agree completely when $\alpha_j = 0$ ($j = 1, \dots, m$). Such a choice of hyperparameter values implies an improper prior that nevertheless has a proper posterior provided that $n_j \geq 1$ ($j = 1, \dots, m$).

In addition, from the limiting behavior of the above estimates (cf. [14]–[16]), we conclude that, under the assumption of the exchangeability of X 's, as $n \rightarrow \infty$, the posterior mean of the parameter θ will coincide with its actual value, whatever it is.

A remarkable property of the Dirichlet family of priors is the following.

Let θ be Dirichlet distributed with parameter vector α . With $l < m$, let $I = \{I_1, \dots, I_l\}$ be a partition of the set $\{1, \dots, m\}$ and $\varphi_i = \sum_{j \in I_i} \theta_j$ ($i = 1, \dots, l$). Then, the random vector $\varphi = (\varphi_1, \dots, \varphi_l)^T$ is Dirichlet distributed with parameter vector $\beta = (\beta_1, \dots, \beta_l)^T$, where $\beta_i = \sum_{j \in I_i} \alpha_j$ ($i = 1, \dots, l$).

We conclude, therefore, from the above result that the requirement for coherence does not allow the use of objective flat $D(1, \dots, 1)$ priors for both the parameters θ and φ .

From the above result, it is obvious that the marginal distribution of $\theta \equiv \theta_1$ is the univariate Dirichlet with parameters $\alpha = \alpha_1$ and $\beta = \sum_{j=2}^m \alpha_j$, known as the $Beta(\alpha, \beta)$ distribution. When dealing with binomial data, a reasonable choice for the hyperparameter values α and β

could be $\alpha = \beta = 1$, resulting in the flat noninformative prior

$$p(\theta) = 1, 0 \leq \theta \leq 1$$

The above prior had been used by both Bayes (1763) and Laplace (1785) in their demonstration of the inverse probability law, of which Laplace was aware independently of Bayes.

Application

For demonstration purposes, we present the following application. We analyze the data in Table 1 which refer to the grades (response variable) of a random sample of first-year successful university students in relation to the parents' educational level (explanatory variable).

We have here three cell probability vectors, namely, $\theta_i = (\theta_{i1}, \theta_{i2}, \theta_{i3})$ ($i = 1, 2, 3$), corresponding to the three categories of parents' educational level. We assume that the parameters θ_1 , θ_2 , and θ_3 are independent, Dirichlet distributed with a common hyperparameter vector $\alpha = (\alpha_1, \alpha_2, \alpha_3)$.

With N the 3×3 matrix representing the above contingency table and θ the 3×3 matrix representing the corresponding cell probabilities, the likelihood function is

$$L(\theta, N) \propto \prod_{i=1}^3 \prod_{j=1}^3 \theta_{ij}^{n_{ij}}$$

Thus,

$$p(\theta | N) \propto p(\theta) p(N | \theta) \propto \prod_{i=1}^3 \prod_{j=1}^3 \theta_{ij}^{\alpha_i + n_{ij} - 1}$$

and, therefore,

$$p(\theta | N) = \prod_{i=1}^3 D(\theta_i | \alpha_i + n_i) \quad [17]$$

We conclude, therefore, that the cell probability vectors θ_i ($i = 1, 2, 3$), conditional on N , are again independent, Dirichlet distributed with updated hyperparameter vectors $\alpha_i^{(n)} = \alpha + n_i$ ($i = 1, 2, 3$), respectively.

For the common hyperparameter vector α , we assume that its components α_1 , α_2 , and α_3 are all equal to 1,

Table 1 Numbers of first-year successful university students according to grades received and parents' educational level

Parents' educational level	Grades		
	A	B	C
Elementary	3	32	50
Secondary	7	59	63
University	8	93	134

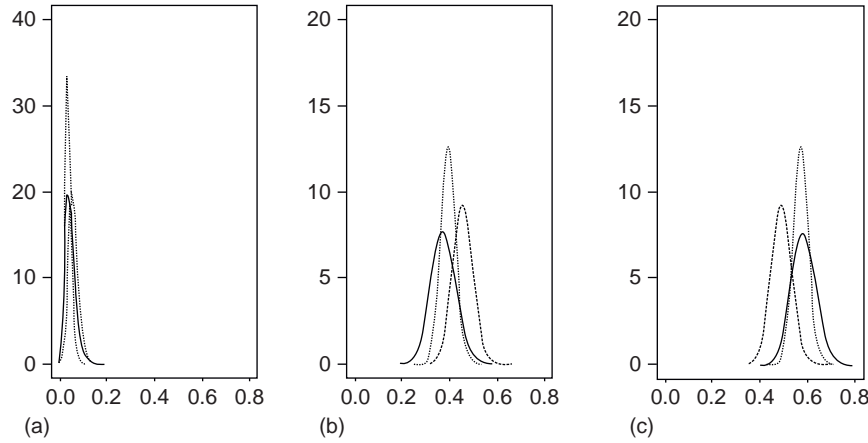


Figure 1 Posterior distributions for cell probabilities θ_{ij} ($i, j = 1, 2, 3$). Continuous lines for elementary parents' educational level ($i = 1$), dashed lines for secondary level ($i = 2$), and dot lines for university level ($i = 3$).

implying an objective flat prior for each one of the parameter vectors θ_i ($i = 1, 2, 3$). Their posterior distributions, as defined by [17], are presented in **Figure 1**. The three cases (a), (b), and (c) correspond to the grade categories A, B, and C, respectively. Specifically, in case (a), we have the posteriors of the cell probabilities θ_{i1} ($i = 1, 2, 3$); in case (b), those of the cell probabilities θ_{i2} ($i = 1, 2, 3$); and in case (c), those of the cell probabilities θ_{i3} ($i = 1, 2, 3$). Comparing the three graphs within each grade category, we may realize that there is no serious discrepancy between the posterior distributions, apart from the fact that the parents' category secondary education, presented with dashed lines, shows a slightly better response (higher located posteriors for A and B grades combined with a lower located posterior for C grades). For sensitivity analysis, several values for the components of the hyperparameter α have been considered in the range (0.5, 5.0), with step 0.5, and all were found to be in agreement with the above conclusion.

Alternative methods for analyzing contingency tables are based on Hierarchical models and generalized linear model (cf. Gelman *et al.*, 2003). For a rather extensive presentation of Bayesian models in educational research, see Novick and Jackson (1974) and Bock (1989).

Interval Estimation

For a scalar parameter θ whose posterior $p(\theta | x^{(n)})$ can be found in closed form, it is a rather easy task to find an interval (a, b) such that

$$F(a) = 1 - F(b) = \frac{1}{2}(1 - \beta) \quad [18]$$

where

$$F(a) = P[\theta \leq a | x^{(n)}] = \int_{-\infty}^a p(\theta | x^{(n)}) d\theta$$

and β a prespecified number in the interval (0, 1). Such a region is called the credibility interval for the parameter θ

with credibility level β . This credibility interval is not necessarily the shortest one. In order to be the shortest, it has to contain the highest posterior density values. Then, it is called the highest posterior density credibility interval. It is obvious that, when the posterior is unimodal and symmetric, the credibility interval defined by [18] coincides with the shortest one. Usually, we simulate data from the posterior $p(\theta | x^{(n)})$ using MCMC methods. Then, we only have to discard the most extreme values at appropriate equal proportions, and the region which is covered by the remaining data, is the required credibility interval.

Hypothesis Testing

According to the Bayesian approach, the hypothesis-testing problem, either for simple (sharp) hypotheses or for composite hypotheses, can also be treated within a decision-theoretic context. It must be mentioned here that the hypothesis-testing problem is not of prime interest within the Bayesian approach, mainly because it oversimplifies the problem of inference by requiring an answer of type 'yes' or 'no'.

Suppose that the parameter space Θ is partitioned into two subsets, say Θ_0 and Θ_1 . Specifically, we assume: $\Theta = \Theta_0 \cup \Theta_1$ with $\Theta_0 \cap \Theta_1 = \emptyset$. Let the null hypothesis H_0 states that $\theta \in \Theta_0$ while the alternative H_1 states that $\theta \in \Theta_1$. Suppose, in addition, that the loss is zero when accepting the right hypothesis, while the loss implied by making a type-I error (rejecting the null hypothesis H_0 when it is true) is a multiple α of the loss implied by making a type-II error (rejecting the alternative hypothesis H_1 when it is true). Then, the Bayes' rule states the following:

$$\begin{aligned} \text{Reject } H_0 \text{ iff } & \frac{P[\Theta_1 | x^{(n)}]}{P[\Theta_0 | x^{(n)}]} \\ & = \int_{\Theta_1} p(\theta | x^{(n)}) d\theta / \int_{\Theta_0} p(\theta | x^{(n)}) d\theta > \alpha \end{aligned} \quad [19]$$

The above ratio is the posterior odds ratio against H_0 (and in favor of H_1).

Consider now two special cases.

Case a: Suppose that both the sets Θ_0 and Θ_1 are single-point sets, that is, $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$, with prior probabilities π_0 and $\pi_1 = 1 - \pi_0$, respectively. Then, the Bayes' rule [19] becomes

$$\text{Reject } H_0 \text{ iff } \frac{\pi_1 p(\mathbf{x}^{(n)}|\theta_1)}{\pi_0 p(\mathbf{x}^{(n)}|\theta_0)} > \alpha$$

The ratio π_1/π_0 represents the prior odds ratio against H_0 , while the quantity

$$B(\mathbf{x}^{(n)}) = p(\mathbf{x}^{(n)}|\theta_1)/p(\mathbf{x}^{(n)}|\theta_0)$$

is called the Bayes' factor against H_0 . Thus, the Bayes' rule becomes:

$$\text{Reject } H_0 \text{ iff } \{\text{prior odds against } H_0\} \times \{\text{Bayes factor against } H_0\} > \alpha \quad [20]$$

Case b: Suppose that the null hypothesis is a simple one, say $H_0: \theta = \theta_0$, while the alternative is the composite one $H_1: \theta \neq \theta_0$. Suppose, as before, that the null hypothesis has a prior probability π_0 . Then, the alternative has a prior probability $\pi_1 = 1 - \pi_0$ and the Bayes' rule is

$$\text{Reject } H_0 \text{ iff } \pi_1 \int_{\Theta_1} p(\mathbf{x}^{(n)}|\theta) d\theta / \pi_0 p(\mathbf{x}^{(n)}|\theta_0) > \alpha$$

with $\Theta_1 = \Theta - \{\theta_0\}$.

It must be mentioned here that using positive probabilities for sharp hypotheses as the above was suggested by Jeffreys (1961).

Extensions

Mixtures of Models and Model Selection

Suppose that, according to our present state of knowledge K , there exists a set $\mathcal{M} = \{1, 2, \dots, M\}$ of possible models from one of which the random sample $\mathbf{x}^{(n)} = \{x_i, i = 1, \dots, n\}$ has been derived. Let M be the discrete random variable defining the actual model. Then, we have

$$p(\mathbf{x}^{(n)}) = \sum_{m=1}^M p(\mathbf{x}^{(n)}|m) p(m)$$

with

$$p(\mathbf{x}^{(n)}|m) = \int_{\Theta_m} p(\mathbf{x}^{(n)}|\theta_m, m) p(\theta_m|m) d\theta_m \quad (m = 1, \dots, M) \quad [21]$$

the predictive distribution of $\mathbf{x}^{(n)}$ when it comes from model m , and θ_m ($m = 1, \dots, M$), the model-specific parameters.

Applying Bayes' theorem, the model posterior probabilities will be

$$p(m|\mathbf{x}^{(n)}) \propto p(\mathbf{x}^{(n)}|m) p(m), \quad m = 1, \dots, M \quad [22]$$

with proportionality constant k defined by

$$k^{-1} = \sum_{m=1}^M p(\mathbf{x}^{(n)}|m) p(m)$$

The posterior odds for the model m against the model m' , say, are as follows:

$$\frac{p(m|\mathbf{x}^{(n)})}{p(m'|\mathbf{x}^{(n)})} = \frac{\pi_m}{\pi_{m'}} \times B_{m,m'}(\mathbf{x}^{(n)}), \quad m \neq m' = 1, \dots, M$$

where $B_{m,m'}(\mathbf{x}^{(n)})$ is the Bayes' factor for the model m against the model m' , that is,

$$B_{m,m'}(\mathbf{x}^{(n)}) = p(\mathbf{x}^{(n)}|m) / p(\mathbf{x}^{(n)}|m')$$

When using Bayes' factors, it is interesting to see that the posterior probability of the model m can be written under the following form:

$$p(m|\mathbf{x}^{(n)}) = \left\{ \sum_{m'=1}^M \frac{\pi_{m'}}{\pi_m} B_{m',m}(\mathbf{x}^{(n)}) \right\}^{-1}, \quad m = 1, \dots, M \quad [23]$$

Thus, the model selection will be based on the result [22] or on its equivalent version [23].

The model selection procedures usually tend to favor models of higher complexity, that is, those with the larger number of parameters. This is a consequence of the fact that the average amount of information provided by an experiment is non-negative (cf. Lindley, 1956; Kokolakis, 1981; Kokolakis, 1985). Although proper priors protect from overfitting, a problem which is very common within the frequentist approach, it might not be suitable to estimate a very large number of parameters using a rather small sample size. In such a case, a cross-validation based on the predictive distributions (cf. Draper, 1995), or the, rather conservative, Bayesian information criterion (BIC), due to Schwarz (1978), might be applied. The latter penalizes the model for the number of parameters it has. Specifically, the model to be chosen is that which maximizes the quantity

$$\text{BIC} = \log \max L(\theta_m, \mathbf{x}^{(n)}) - \frac{1}{2} k_m \log k_m$$

where k_m is the number of parameters within model m , that is, the dimension of θ_m .

Random Hyperparameters and Hierarchical Structures

Suppose that, according to our present state of knowledge K , the hyperparameter α , say, in the prior distribution of θ , is also a random quantity with its own prior distribution. Let \mathcal{A} be the set from which the hyperparameter α takes values. Then, we have the following results:

$$p(\mathbf{x}^{(n)}) = \int_{\mathcal{A}} p(\mathbf{x}^{(n)}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})d\boldsymbol{\alpha}$$

with

$$p(\mathbf{x}^{(n)}|\boldsymbol{\alpha}) = \int_{\Theta} p(\mathbf{x}^{(n)}, \boldsymbol{\theta}|\boldsymbol{\alpha})d\boldsymbol{\theta} = \int_{\Theta} p(\mathbf{x}^{(n)}|\boldsymbol{\alpha}, \boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\alpha})d\boldsymbol{\theta} \quad [24]$$

the, conditional on $\boldsymbol{\alpha}$, predictive distribution of $\mathbf{x}^{(n)}$. The required posteriors are then given by

$$p(\boldsymbol{\alpha}|\mathbf{x}^{(n)}) \propto p(\boldsymbol{\alpha})p(\mathbf{x}^{(n)}|\boldsymbol{\alpha}) = p(\boldsymbol{\alpha}) \int_{\Theta} p(\mathbf{x}^{(n)}|\boldsymbol{\alpha}, \boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{\alpha})d\boldsymbol{\theta} \quad [25]$$

and

$$p(\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{x}^{(n)}) \propto p(\boldsymbol{\theta}|\boldsymbol{\alpha})p(\mathbf{x}^{(n)}|\boldsymbol{\theta}, \boldsymbol{\alpha}) \quad [26]$$

Suppose now that the dataset $\mathbf{x}^{(n)}$ consists of k subsets, namely, $\mathbf{x}^{(n_i)} = \{\mathbf{x}_{ij}, j = 1, \dots, n_i\} (i = 1, \dots, k)$, coming from k populations, each one with its own parameter vector $\boldsymbol{\theta}_i$, a sub-vector of the collective parameter vector $\boldsymbol{\theta}$ above. We have, therefore, $\mathbf{x}^{(n)} = (\mathbf{x}^{(n_1)}, \dots, \mathbf{x}^{(n_k)})$ with $n = \sum_{i=1}^k n_i$. We may think of $\mathbf{x}^{(n_i)} (i = 1, \dots, k)$ as random samples of the same product that come from k production lines.

Now, consider the following hierarchical structure. The random samples $\mathbf{x}^{(n_i)} (i = 1, \dots, k)$, conditional on $\boldsymbol{\theta}_i i = 1, \dots, k$, respectively, are independent, and the parameters $\boldsymbol{\theta}_i i = 1, \dots, k$, conditional on the hyperparameter $\boldsymbol{\alpha}$, are independent and identically distributed according to $p(\cdot|\boldsymbol{\alpha})$. In other words, the random samples $\mathbf{x}^{(n_i)} (i = 1, \dots, k)$ are conditionally independent and the parameters $\boldsymbol{\theta}_i (i = 1, \dots, k)$, are exchangeable. Then, in place of [24]–[26] above, we have

$$p(\mathbf{x}^{(n)}|\boldsymbol{\alpha}) = \prod_{i=1}^k \left\{ \int_{\Theta} p(\mathbf{x}^{(n_i)}|\boldsymbol{\alpha}, \boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i|\boldsymbol{\alpha})d\boldsymbol{\theta}_i \right\} \quad [27]$$

$$p(\boldsymbol{\alpha}|\mathbf{x}^{(n)}) \propto p(\boldsymbol{\alpha}) \prod_{i=1}^k \left\{ \int_{\Theta} p(\mathbf{x}^{(n_i)}|\boldsymbol{\alpha}, \boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i|\boldsymbol{\alpha})d\boldsymbol{\theta}_i \right\} \quad [28]$$

$$p(\boldsymbol{\theta}|\boldsymbol{\alpha}, \mathbf{x}^{(n)}) \propto \prod_{i=1}^k p(\boldsymbol{\theta}_i|\boldsymbol{\alpha}, \mathbf{x}^{(n_i)})p(\mathbf{x}^{(n_i)}|\boldsymbol{\alpha}, \boldsymbol{\theta}_i) \quad [29]$$

The above hierarchical structure was introduced by Lindley and Smith (1972) in their seminal work on the Bayesian linear models. For a rather extensive range of applications of hierarchical structures in linear and generalized linear models, see Gelman *et al.* (2003) and Press (2003).

See also: Markov Chain Monte Carlo.

Bibliography

Aitchison, J. and Dunsmore, I. R. (1975). *Statistical Prediction Analysis*. Cambridge, UK: Cambridge University Press.

- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society London* **V61.53**, 370–418. [Reprinted with biographical note by G. A. Barnard in *Biometrika* **45**, 293–315 (1958)].
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd edn. New York: Springer.
- Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. New York: Wiley.
- Bett, R. (1989). Carneades' Pithanon: A reappraisal of its role and status. *Oxford Studies in Ancient Philosophy* **7**, 59–94.
- Bock, R. D. (1989). *Multilevel Analysis of Educational Data*. New York: Academic Press.
- De Finetti, B. (1937). La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut H. Poincaré* **7**, 1–68. [Translated and reprinted in Kyburg, H. and Smokler, H. (eds.) (1964). *Studies in subjective probability*, pp 93–158. New York: Wiley].
- De Finetti, B. (1974, 1975). *Theory of Probability* (2 vols). New York: Wiley.
- DeGroot, M. H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.
- Diaconis, P. and Freedman, D. (1980). De Finetti's generalizations of exchangeability. In Jeffrey, R. (ed.), *Studies in Inductive Logic and Probability* **2**, 233–249.
- Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society B* **57**, 45–97 (with discussion).
- Fishburn, P. C. (1986). The axioms of subjective probability. *Statistical Science* **1**, 335–358.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*, 2nd edn. London: Chapman and Hall.
- Kokolakis, G. E. (1981). On the expected probability of correct classification. *Biometrika* **68**(2), 477–483.
- Kokolakis, G. E. (1985). A martingale approach to the problem of dimensionality and performance of Bayesian classifiers. Asymptotic results. *Communications in Statistics – Theory and Methods* **14**, 927–935.
- Kokolakis, G., Nanopoulos, Ph., and Fouskakis, D. (2006). Bregman divergences in the $(m \times k)$ -partitioning problem. *Computational Statistics and Data Analysis* **51**, 668–678.
- Laplace, P. S. (1785). Mémoire sur les formules qui sont fonctions de très grands nombres. *Mémoires de l'Académie Royal des Sciences* **21**, 1–88.
- Lindley, D. V. (1956). On a measure of the information provided by an experiment. *Annals of Mathematical Statistics* **27**, 986–1005.
- Lindley, D. V. (1965). *Introduction to Probability and Statistics, Part 2: Inference*. Cambridge: Cambridge University Press.
- Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society B* **34**, 1–41 (with discussion).
- Lindley, D. V. (1985). *Making Decisions*, 2nd edn. Chichester: Wiley.
- Lindley, D. V. (2000). The philosophy of statistics. *Statistician* **49**, 293–337 (with discussion).
- Novick, M. R. and Jackson, P. H. (1974). *Statistical Methods for Educational and Psychological Research*. New York: McGraw-Hill.
- Press, S. J. (2003). *Subjective and Objective Bayesian Statistics*. New York: Wiley.
- Robert, C. P. (2001). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd edn. New York: Springer.
- Savage, L. J. (1954). *The Foundations of Statistics*. New York: Wiley.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464.

Further Reading

- Fox, J. P. and Glas, C. A. W. (2001). Bayesian estimation of a multi-level IRT model using Gibbs sampling. *Psychometrika* **66**, 271–288.
- Jeffreys, H. (1961). *Theory of Probability*, 3rd edn. Oxford: Clarendon.
- Lewis, C. and Sheehan, K. M. (1990). Using Bayesian decision theory to design a computerized mastery test. *Applied Psychological Measurement* **14**, 367–386.

- Rubin, D. B. (1983). Some applications of Bayesian statistics to educational data. *Statistician* **32**, 55–68.
- Sinharay, S. (2005). Assessing fit of unidimensional item response theory models using a Bayesian approach. *Journal of Educational Measurement* **42**(4), 375–394.
- Sinharay, S. (2006). Bayesian statistics in educational measurement. In Upadhyay, S. K., Singh, U., and Dey, D. K. (eds.) *Bayesian Statistics and its Applications*, pp 422–437.
- van der Linden, W. J. (1998). Bayesian item-selection criteria for adaptive testing. *Psychometrika* **63**, 201–216.

Bootstrap Method

K Singh and M Xie, Rutgers University, Piscataway, NJ, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bias – Deviation between the true population value and the mean of a sample estimate (average over all possible samples).

Central limit theorem – It is the mathematical phenomenon which states that, as sample size tends to infinity, the sample mean follows normal curve (bell-shaped curve, also known as Gaussian curve) in limit.

Confidence interval – A sample data-based range such that a given population parameter lies within the range with a specified high probability (called confidence level).

Probability distribution – It describes a chance pattern according to which a random variable takes values.

Random variable – A statistical terminology used to describe a measurement or count which depends on chance.

Regression model – A statistical model used to describe the relationship of a variable of interest (called response variable) and a set of variables (called covariates) which are supposed to influence the response variable.

Sample estimator – A mathematical function of sample data used for approximating a population parameter (characteristic).

Standard error – The deviation between a true parameter value and its sample estimate is squared and averaged over all possible samples. Its square root is called standard error.

Time series – A sequence of measurements, taken at successive time points or intervals.

summarize a sample-based study and generalize the finding to the parent population in a scientific manner. A technical term for a sample summary number is (sample) statistic. Some basic sample statistics are sample mean, sample median, sample standard deviation, etc. Of course, a summary statistic like the sample mean will fluctuate from sample to sample and a statistician would like to know the magnitude of these fluctuations around the corresponding population parameter in an overall sense. This is then used in assessing margin of errors. The entire picture of all possible values of a sample statistics presented in the form of a probability distribution is called a sampling distribution. There is a plenty of theoretical knowledge of sampling distributions, which can be found in any text books of mathematical statistics. A general intuitive method applicable to just about any kind of sample statistic that keeps the user away from the technical tedium has got its own special appeal. Bootstrap is such a method.

To understand bootstrap, suppose it were possible to draw repeated samples (of the same size) from the population of interest, a large number of times. Then, one would get a fairly good idea about the sampling distribution of a particular statistic from the collection of its values arising from these repeated samples. But, that does not make sense as it would be too expensive and defeat the purpose of a sample study. The purpose of a sample study is to gather information cheaply in a timely fashion. The idea behind bootstrap is to use the data of a sample study at hand as a surrogate population, for the purpose of approximating the sampling distribution of a statistic; that is, to resample (with replacement) from the sample data at hand and create a large number of phantom samples known as bootstrap samples. The sample summary is then computed on each of the bootstrap samples (usually a few thousand). A histogram of the set of these computed values is referred to as the bootstrap distribution of the statistic.

In bootstrap's most elementary application, one produces a large number of copies of a sample statistic, computed from these phantom bootstrap samples. Then, a small percentage, say $100(\alpha/2)\%$ (usually $\alpha = 0.05$, is trimmed off from the lower as well as from the upper end of these numbers. The range of remaining $100(1-\alpha)\%$ values is declared as the confidence limits of the corresponding unknown population summary number of interest, with level of confidence $100(1-\alpha)\%$. The above method is referred to as bootstrap percentile method. We shall return to it later in the article.

Introduction and the Idea

Efron (1979) introduced the bootstrap method. It spread like brush fire in statistical sciences within a couple of decades. Now if one conducts a Google search for the above title, an astounding 1.86 million records will be mentioned; scanning through even a fraction of these records is a daunting task. We attempt first to explain the idea behind the method and the purpose of it at a rather rudimentary level. The primary task of a statistician is to

The Theoretical Support

Let us develop some mathematical notations for convenience. Suppose a population parameter θ is the target of a study; say, for example, θ is the household median income of a chosen community. A random sample of size n yields the data (X_1, X_2, \dots, X_n) . Suppose the corresponding sample statistic computed from this data set is $\hat{\theta}$ (sample median in the case of the example). For most sample statistics, the sampling distribution of $\hat{\theta}$ for large n ($n \geq 30$ is generally accepted as large sample size) is bell shaped with center θ and standard deviation (a/\sqrt{n}) , where the positive number a depends on the population and the type of statistic $\hat{\theta}$. This phenomenon is the celebrated central limit theorem (CLT). Often, there are serious technical complexities in approximating the required standard deviation from the data. Such is the case when $\hat{\theta}$ is sample median or sample correlation. Then bootstrap offers a bypass. Let $\hat{\theta}_B$ stand for a random quantity which represents the same statistic computed on a bootstrap sample drawn out of (X_1, X_2, \dots, X_n) . What can we say about the sampling distribution of $\hat{\theta}_B$ (w.r.t. all possible bootstrap samples), while the original sample (X_1, X_2, \dots, X_n) is held fixed? The first two articles dealing with the theory of bootstrap – Bickel and Freedman (1981) and Singh (1981) – provided large sample answers for most of the commonly used statistics. In limit, as $(n \rightarrow \infty)$, the sampling distribution of $\hat{\theta}_B$ is also bell shaped with $\hat{\theta}$ as the center and the same standard deviation (a/\sqrt{n}) . Thus, bootstrap distribution of $\hat{\theta}_B - \hat{\theta}$ approximates (fairly well) the sampling distribution of $\hat{\theta} - \theta$. Note that, as we go from one bootstrap sample to another, only $\hat{\theta}_B$ in the expression $\hat{\theta}_B - \hat{\theta}$ changes as $\hat{\theta}$ is computed on the original data (X_1, X_2, \dots, X_n) . This is the bootstrap CLT. For a proof of bootstrap CLT for the mean, reader is referred to Singh (1981).

Furthermore, it has been found that if the limiting sampling distribution of a statistical function does not involve population unknowns, bootstrap distribution offers a better approximation to the sampling distribution than the CLT. Such is the case when the statistical function is of the form $(\hat{\theta}_B - \hat{\theta})/\text{SE}$ where SE stands for true or sample estimate of the standard error of $\hat{\theta}$, in which case the limiting sampling distribution is usually standard normal. This phenomenon is referred to as the second-order correction by bootstrap. A caution is warranted in designing bootstrap, for second-order correction. For illustration, let $\theta = \mu$, the population mean, and $\hat{\theta} = \bar{X}$, the sample mean; σ = population standard deviation; s = sample standard deviation computed from original data; and s_B is the sample standard deviation computed on a bootstrap sample. Then, the sampling distribution of $(\bar{X} - \mu)/\text{SE}$, with $\text{SE} = \sigma/\sqrt{n}$, will be approximated by the bootstrap distribution of $(\bar{X}_B - \bar{X})/\text{SE}$, with \bar{X}_B = bootstrap sample mean and $\text{SE} = s/\sqrt{n}$. Similarly, the sampling distribution of $(\bar{X} - \mu)/\text{SE}$, with $\text{SE} = s/\sqrt{n}$,

will be approximated by the bootstrap distribution of $(\bar{X}_B - \bar{X})/\text{SE}_B$, with $\text{SE} = s_B/\sqrt{n}$. The earliest results on second-order correction were reported in Singh (1981) and Babu and Singh (1983). In the subsequent years, a flood of large sample results on bootstrap with substantially higher depth, followed. A name among the researchers in this area that stands out is Peter Hall of Australian National University.

Primary Applications of Bootstrap

Approximating Standard Error of a Sample Estimate

Let us suppose, information is sought about a population parameter θ . Suppose $\hat{\theta}$ is a sample estimator of θ based on a random sample of size n , that is, $\hat{\theta}$ is a function of the data (X_1, X_2, \dots, X_n) . In order to estimate standard error of $\hat{\theta}$, as the sample varies over the class of all possible samples, one has the following simple bootstrap approach.

Compute $(\theta_1^*, \theta_2^*, \dots, \theta_N^*)$, using the same computing formula as the one used for $\hat{\theta}$, but now base it on N different bootstrap samples (each of size n). A crude recommendation for the size N could be $N = n^2$ (in our judgment), unless n^2 is too large. In that case, it could be reduced to an acceptable size, say $n \log_e n$. One defines

$$\text{SE}_B(\hat{\theta}) = [(1/N) \sum_{i=1}^N (\theta_i^* - \hat{\theta})^2]^{1/2}$$

following the philosophy of bootstrap: replace the population by the empirical population.

An older resampling technique used for this purpose is Jackknife, though bootstrap is more widely applicable. The famous example where Jackknife fails while bootstrap is still useful is that of $\hat{\theta}$ = the sample median.

Bias Correction by Bootstrap

The mean of sampling distribution of $\hat{\theta}$ often differs from θ , usually by an amount $= c/n$ for large n . In statistical language, one writes

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta \approx O(1/n)$$

A bootstrap based approximation to this bias is

$$\frac{1}{N} \sum_{i=1}^N \theta_i^* - \hat{\theta} = \text{Bias}_B(\hat{\theta}) \text{ (say)}$$

where θ_i^* are bootstrap copies of $\hat{\theta}$, as defined in the earlier subsection. Clearly, this construction is also based on the standard bootstrap thinking: replace the population by the empirical population of the sample. The bootstrap bias-corrected estimator is $\hat{\theta}_c = \hat{\theta} - \text{Bias}_B(\hat{\theta})$. It needs to be pointed out that the older resampling technique called Jackknife is more popular with statisticians for the purpose of bias estimation.

Bootstrap Confidence Intervals

Confidence intervals for a given population parameter θ are sample-based range $[\hat{\theta}_1, \hat{\theta}_2]$ given out for the unknown number θ . The range possesses the property that θ would lie within its bounds with a high (specified) probability. The latter is referred to as confidence level. Of course, this probability is with respect to all possible samples, each sample giving rise to a confidence interval which thus depends on the chance mechanism involved in drawing the samples. The two mostly used levels of confidence are 95% and 99%. We limit ourselves to the level 95% for our discussion here. Traditional confidence intervals rely on the knowledge of sampling distribution of $\hat{\theta}$, exact or asymptotic as $n \rightarrow \infty$. Here are some standard brands of confidence intervals constructed using bootstrap.

Bootstrap percentile method

This method was mentioned in the introduction itself, because of its popularity which is primarily due to its simplicity and natural appeal. Suppose one settles for 1000 bootstrap replications of $\hat{\theta}$, denoted by $(\theta_1^*, \theta_2^*, \dots, \theta_{1000}^*)$. After ranking from bottom to top, let us denote these bootstrap values as $(\theta_{(1)}^*, \theta_{(2)}^*, \dots, \theta_{(1000)}^*)$. Then the bootstrap percentile confidence interval at 95% level of confidence would be $[\theta_{(25)}^*, \theta_{(975)}^*]$. Turning to the theoretical aspects of this method, it should be pointed out that the method requires the symmetry of the sampling distribution of $\hat{\theta}$ around θ . The reason is that the method approximates the sampling distribution of $\hat{\theta} - \theta$ by the bootstrap distribution of $\hat{\theta} - \hat{\theta}_B$, which is contrary to the bootstrap thinking that the sampling distribution of $\hat{\theta} - \theta$ could be approximated by the bootstrap distribution of $\hat{\theta}_B - \hat{\theta}$. Interested readers are referred to Hall (1988).

Centered bootstrap percentile method

Suppose the sampling distribution of $\hat{\theta} - \theta$ is approximated by the bootstrap distribution of $\hat{\theta}_B - \hat{\theta}$, which is what the bootstrap prescribes. Denote 100 ϵ -th percentile of $\hat{\theta}_B$ (in bootstrap replications) by B_ϵ . Then, the statement that $\hat{\theta} - \theta$ lies within the range $B_{0.025} - \hat{\theta}, B_{0.975} - \hat{\theta}$ would carry a probability ≈ 0.95 . But, this statement easily translates to the statement that θ lies within the range $(2\hat{\theta} - B_{0.975}, 2\hat{\theta} - B_{0.025})$. The latter range is what is known as centered bootstrap percentile confidence interval (at coverage level 95%). In terms of 1000 bootstrap replications $B_{0.025} = \theta_{(25)}^*$ and $B_{0.975} = \theta_{(975)}^*$.

Bootstrap- t methods

As it was mentioned in section ‘The theoretical support’, bootstrapping a statistical function of the form $T = (\hat{\theta} - \theta)/SE$, where SE is a sample estimate of the standard error of $\hat{\theta}$, brings extra accuracy. This additional accuracy is due to so-called one-term Edgeworth correction by the bootstrap. The reader could find essential details in Hall (1992b). The basic example of T is the standard

t -statistics (from which the name bootstrap- t is derived): $t = (\bar{X} - \mu)/s/\sqrt{n}$, which is a special case with $\theta = \mu$ (the population mean), $\hat{\theta} = \bar{X}$ (the sample mean), and s standing for the sample standard deviation. The bootstrap counterpart of such a function T is $T_B = (\hat{\theta}_B - \theta)/SE_B$ where SE_B is exactly like SE but computed on a bootstrap sample. Denote the 100 ϵ -th bootstrap percentile of T_B by b_ϵ , and consider the statement: T lies within $[b_{0.025}, b_{0.975}]$. After the substitution $T = (\hat{\theta} - \theta)/SE$, the above statement translates to ‘ θ lies within $(\hat{\theta} - SE b_{0.975}, \hat{\theta} - SE b_{0.025})$ ’. This range for θ is called bootstrap- t -based confidence interval for θ at coverage level 95%. Such an interval is known to achieve higher accuracy than the earlier method, which is referred to as second-order accuracy in technical literature.

We conclude the section with a remark that B. Efron proposed correction to the rudimentary percentile method to bring in extra accuracy. These corrections are known as Efron’s ‘bias-correction’ and ‘accelerated bias-correction’. The details could be found in Efron and Tibshirani (1993). The bootstrap- t automatically takes care of such corrections, although the bootstrapper needs to look for a formula for SE which is avoided in the percentile method.

Some Real Data Example

Example 1 (Skewed Univariate Data). In the first example, the data are taken from Hollander and Wolfe (1999: 63), which represent the effect of illumination (difference between counts with and without illumination) on the rate of beak-clapping among chick-embryos. The boxplot suggests lack of normality of the population. We have carried out bootstrap analysis on the median and the mean. A noteworthy finding is the lack of symmetry of bootstrap- t histogram, which differs from limiting normal curve. The 95% level confidence intervals coming from our analysis for both mean and the median (centered bootstrap percentile method) cover the range [10, 30], roughly speaking. This range represents overall difference (increase) in the beak-clapping counts per minute due to illumination.

Example 2 (Bivariate Data). In this example, the data are from Collins *et al.* (1999), which assess body fat in collegiate football players (Devore, 2003: 553). We study correlation between the BOD and HW measurements; see the data at the end of this section. Here, BOD is BOD POD, a whole body air-displacement plethysmograph, and HW refers to hydrostatic weighing. The sample size is modest, but reasonable for bootstrap methods. As bivariate data consist of n pairs of data, say (X_i, Y_i) , for $i = 1, \dots, n$, one draws a pair of data randomly at a time in the bootstrap resampling. For instance, the first draw could be (X_7, Y_7) followed by (X_3, Y_3) , etc. The box plots of original data in **Figure 1(a)** suggest lack of normality of the underlying populations.

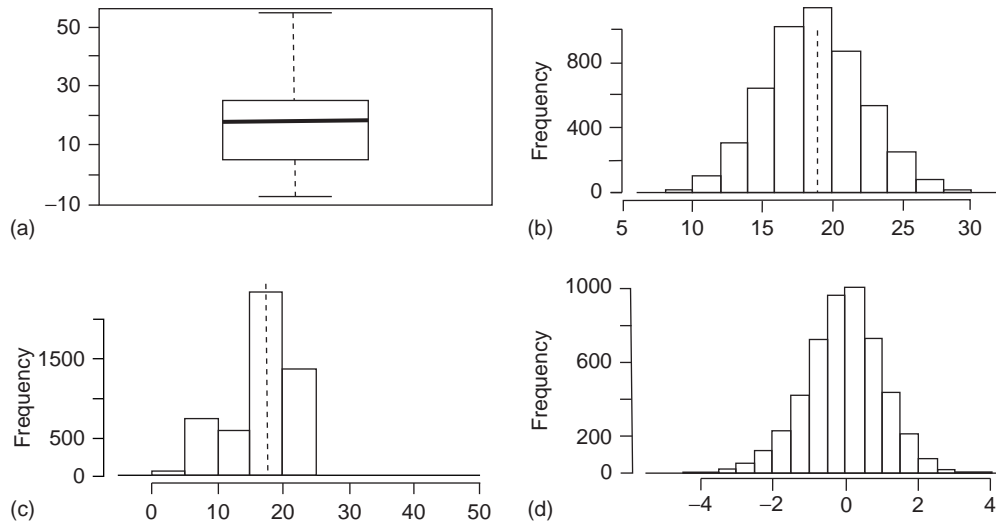


Figure 1 Boxplot of the measurement is presented in (a). Bootstrap distributions of the sample mean, sample median and t^* statistic are plotted in (b)–(d), respectively. The dotted lines in (b) and (c) correspond respectively to the sample mean and sample median. Based the bootstrap distributions, the 95% confidence interval for the population median by the bootstrap percentile method is (4.7000, 24.7000), by the centered bootstrap percentile is (10.5000, 30.5000). The 95% confidence interval for the population mean by the percentile bootstrap method is (10.0960, 28.1200) and by the centered bootstrap method is (9.4880, 27.51200). The bootstrap- t 95% CI for the population mean is (12.9413, 30.8147). Note that the bootstrap- t on the mean shows skewed histogram of the t -distribution.

The histogram for correlations computed on bootstrap bivariate data is plotted in **Figure 2(c)** which is asymmetric (skewed to the left). For this reason, the centered bootstrap percentile confidence interval appears more appropriate. According to our bootstrap analysis, the two measurements have at least a correlation of 0.78 in the population.

Data for Example 1:

-8.5 -4.6 -1.8 -0.8 1.9 3.9 4.7 7.1 7.5 8.5 14.8 16.7 17.6 19.7
20.6 21.9 23.8 24.7 24.7 25.0 40.7 46.9 48.3 52.8 54.0

Data for Example 2:

BOD

2.5 4.0 4.1 6.2 7.1 7.0 8.3 9.2 9.3 12.0 12.2 12.6 14.2 14.4 15.1
15.2 16.3 17.1 17.9 17.9

HW

8.0 6.2 9.2 6.4 8.6 12.2 7.2 12.0 14.9 12.1 15.3 14.8 14.3 16.3
17.9 19.5 17.5 14.3 18.3 16.2

Engineering: A Fitting Bootstrap

A sizable amount of journal literature on the topic is directed toward proposal and study of bootstrap schemes which will produce decent results in various statistical situations. The setup that has been the basis of forgoing discussion is basic and there are many types of departures from it. How to bootstrap in case of two-stage sampling or a stratified sampling? Natural schemes are not hard to think of. Bootstrapping in the case of data with regression models has attracted a lot of attention. There are two schemes which stand out: in one of which the covariate(s) and the

response variable are resampled together (called paired bootstrap), and the other one bootstraps the residuals (=response – fitted model value) and then reconstructs the bootstrap regression data by plugging in the estimated regression parameters (called residual bootstrap). Paired bootstrap remains valid – in the sense of correct outcome in the limit as $n \rightarrow \infty$, even if the error variances in the model are unequal; a property which the residual bootstrap lacks. The shortcoming is compensated by the fact that the latter scheme brings additional accuracy in the estimation of standard error. This is the classic tug of war between efficiency and robustness in statistics (see Liu and Singh, 1992a).

A lot harder to bootstrap are the time series data. Needless to say, time series analysis is of critical importance in several disciplines, especially in econometrics. The sources of difficulty are twofold: (1) Time series data possess serial dependence, that is, X_{T+1} has dependence on X_T , X_{T-1} , etc; (2) the statistical population changes with time, and that is known as nonstationarity. It was noted very early on (see Singh (1981) for m -dependent data) that the classical bootstrap cannot handle dependent data. A fair amount of research has been dedicated to modifying the bootstrap so that it could automatically bring in the dependence structure of the original sampling into bootstrap samples. The scheme of moving-block bootstrap has become quite well known (invented in Kunch (1989) and Liu and Singh (1992). Potitis and Romano are well-known authors on the topic, whose contributions have led to significant advancements on the topic of resampling, in general. In a moving block

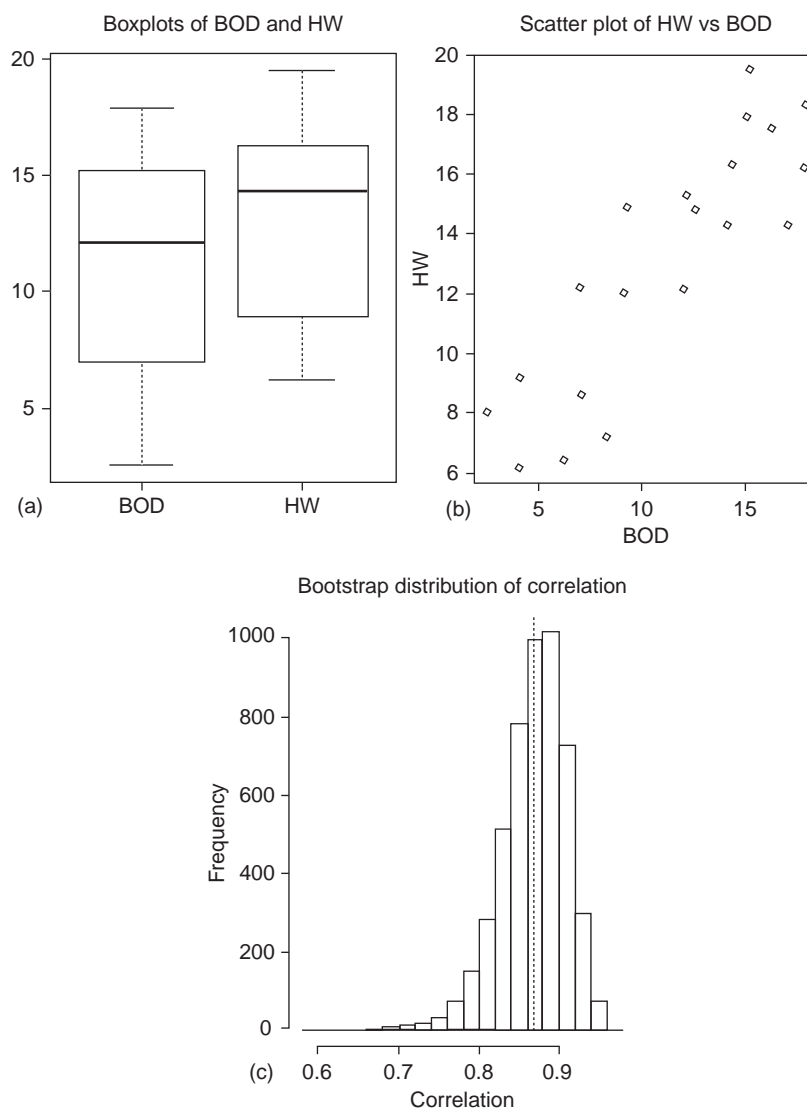


Figure 2 Boxplots of BOD and HW in (a) suggest somewhat non-normal data. Scatter plot in (b) indicates they are highly correlated. Bootstrap inference on the correlation between BOD and HW is presented in (c), which shows the bootstrap distribution (in histogram) of correlation. In particular, sample correlation of BOD and HW is 0.8679 which corresponds to the dotted vertical line in (c). The SE of the correlation is 0.0412 with an estimated bias of 0.0003. The 95% confidence interval of the correlation by the bootstrap percentile method is (0.7222, 0.9490) and the 95% confidence interval by the centered bootstrap percentile method is (0.7868, 1.0136).

bootstrap scheme, one draws a block of data at a time, instead of one of the X_i 's at a time, in order to preserve the underlying serial dependence structure that is present in the sample. There is plenty of ongoing research in the area of bootstrap methodology on econometric data.

The Great m out of n Bootstrap with $(m/n \rightarrow 0)$

There are various types of conditions under which the straightforward bootstrap becomes inconsistent, meaning that the bootstrap estimate of sampling distribution

and the true sampling distribution do not approach to the same limit, as the sample size n tends to ∞ . That means, for large samples, one is bound to end up with an inaccurate statistical inference. The examples include, just to name a few, bootstrapping sample minimum or sample maximum which estimate endpoint of a population distribution (Bickel and Freedman, 1981), the case of sample mean when the population variance is ∞ (Athreya, 1986) and bootstrapping sample eigenvalues when population eigenvalues have multiplicity (Eaton and Tyler, 1991), the case of sample median when the population density is discontinuous at the population median (Huang *et al.*, 1996). Luckily, a general remedy

exists and that is to keep the bootstrap sample size m much lower than the original size. Mathematically speaking, one requires $m \rightarrow \infty$ and $m/n \rightarrow 0$, as $n \rightarrow \infty$. In theory it fixes the problem; however for users, it is somewhat troublesome. How to choose m ? An obvious suggestion would be settle for a fraction of n , say 20% or so. It should be pointed out that in good situations, where the regular bootstrap is fine, such an m is not advisable as it will result in loss of efficiency (see Bickel (2003), for a recent survey on the topic).

Bibliography

- Athreya, K. B. (1986). Bootstrap of the mean in the infinite variance case. *Annals of Statistics* **14**, 724–731.
- Babu, G. J. and Singh, K. (1983). Inference on means using the bootstrap. *Annals of Statistics* **11**, 999–1003.
- Bickel, P. J. (2003). Unorthodox bootstraps (invited papers). *Journal of the Korean Physical Society* **32**, 213–224.
- Bickel, P. J. and Freedman, D. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics* **9**, 1196–1217.
- Collins, M. A., Millard-Stafford, M. L., Sparling, P. B., et al. (1999). Evaluating BOD POD(R) for assessing body fat in collegiate football players. *Medicine and Science in Sports and Exercise* **31**, 1350–1356.
- Davison, A. C. and Hinkley, D. V. (1988). Saddle point approximations in resampling method. *Biometrika* **75**, 417–431.
- Devore, L. J. (2003). *Probability and Statistics for Engineering and the Science*. Boston, MA: Duxbury Press.
- Eaton, M. L. and Tyler, D. E. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *Annals of Statistics* **19**, 260–271.
- Efron, B. (1979). Bootstrap methods: Another look at jackknife. *Annals of Statistics* **7**, 1–26.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals (with discussions). *Annals of Statistics* **16**, 927–953.
- Hall, P. (1992a). Bootstrap confidence intervals in nonparametric regression. *Annals of Statistics* **20**, 695–711.
- Hall, P. (1992b). *The Bootstrap and Edgeworth Expansion*. New York: Springer.
- Hollander, M. and Wolfe, D. A. (1999). *Nonparametric Statistical Methods*, 2nd edn. New York: Wiley.
- Huang, J. S., Sen, P. K., and Shao, J. (1996). Bootstrapping a sample quantile when the density has a jump. *Statistica Sinica* **6**, 299–309.
- Liu, R. Y. and Singh, K. (1992a). Efficiency and robustness in resampling. *Annals of Statistics* **20**, 370–384.
- Liu, R. Y. and Singh, K. (1992b). Moving block jackknife and bootstrap capture weak dependence. In Lepage, R. and Billard, L. (eds.) *Exploring the Limits of Bootstrap*, pp 225–248. New York: Wiley.
- Kunch, H. R. (1989). The jackknife and bootstrap for general stationary observations. *Annals of Statistics* **17**, 1217–1241.
- Singh, K. (1981). On asymptotic accuracy of Efron's bootstrap. *Annals of Statistics* **9**, 1187–1195.
- Beran, R. (1990). Refining bootstrap simultaneous confidence sets. *Journal of the American Statistical Association* **85**, 417–428.
- Bickel, P. J. and Freedman, D. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Annals of Statistics* **12**, 470–482.
- Boos, D. D. and Munahan, J. F. (1986). Bootstrap methods using prior information. *Biometrika* **73**, 77–83.
- Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Annals of Statistics* **16**, 1709–1722.
- DiCiccio, T. J. and Romano, J. P. (1988). A review of bootstrap confidence intervals (with discussions). *Journal of the Royal Statistical Society: Series B* **50**, 538–554.
- Efron, B. (1987). Better bootstrap confidence intervals (with discussions). *Journal of the American Statistical Association* **82**, 171–200.
- Efron, B. (1994). Missing data, imputation and the bootstrap (with discussions). *Journal of the American Statistical Association* **89**, 463–479.
- Freedman, D. A. (1981). Bootstrapping regression models. *Annals of Statistics* **9**, 1218–1228.
- Hinkley, D. V. (1988). Bootstrap methods (with discussions). *Journal of the Royal Statistical Society: Series B* **50**, 321–337.
- Rubin, D. B. (1981). The Bayesian bootstrap. *Annals of Statistics* **9**, 130–134.
- Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer.
- Singh, K. (1998). Breakdown theory for bootstrap quantiles. *Annals of Statistics* **26**, 1719–1732.
- Singh, K. and Xie, M. (2003). Bootlier-plot-Bootstrap based outlier detection plot. *Sankhya* **65**, 532–559.
- Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling procedures (with discussions). *Annals of Statistics* **14**, 1261–1350.
- Young, G. A. (1994). Bootstrap: More than a stab in the dark? (with discussion). *Statistical Science* **9**, 382–415.

Further Reading

Canonical Correlation

A Basu and A Mandal, Indian Statistical Institute, Kolkata, India

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In statistics and data analysis, we are often interested in the relationship and interdependence between two or more variables. The simplest known measure in this respect is the simple correlation coefficient ρ between two variables, usually measured by the product moment estimator, which measures the linear relationship between two variables. Other measures such as the partial correlation coefficient provide refinements to this idea when the correlation between two variables needs to be adjusted for the presence of other variables.

However, sometimes, we are interested in the correlation between one variable X and another set of variables Y , or, more generally, between two different sets X and Y , each having more than one variable. Canonical correlation analysis quantifies the association between two sets of such variables. In the particular case when one of the two groups – say the first group – has only one variable, the linear dependence between this variable X and the variables of the other group Y is quantified through the multiple correlation, the maximum correlation between X and a linear combination of the Y variables. Canonical correlation, which may be called the multiple, multiple correlation, is a further generalization of this idea. It is a dimension reduction technique which attempts to describe the high-dimensional relationship between two groups of variables in terms of a small number of canonical variable pairs. The procedure was initially developed by Hotelling (1935, 1936) in the 1930s.

Canonical correlation analysis focuses on the correlation between a linear combination of the variables in one set and another linear combination of the variables in the other set. The idea is to first determine U_1 and V_1 , linear combinations of the X and Y variables, respectively, such that the correlation between U_1 and V_1 is the highest possible among all such linear combinations. Next, we determine a new pair of linear combinations, say U_2 and V_2 , which has the highest correlation subject to U_2 being uncorrelated with U_1 , and V_2 being uncorrelated with V_1 (the construction actually ensures that U_1 and V_2 are uncorrelated as well, as are U_2 and V_1). The process continues until a subsequent pair of linear combinations no longer produces a significant correlation. The pairs of linear combinations are called the canonical variables, and their correlations are the canonical correlations.

In addition to describing the nature of the relationship between two latent variables, canonical correlation

analysis is useful in determining how many dimensions are needed to account for the relationship. If there are two groups of variables, one containing m variables and the other containing n variables, there could be $m \times n$ different correlations between the different variable pairs of the two sets. Canonical correlation analysis tries to describe this structure in terms of a fewer number of correlations and variable pairs; this number will be no larger than the minimum of m and n .

Canonical Correlations

Let X represent the first set of variables, and let Y denote the second set of variables. Assume that X is m dimensional, and let the dimension of Y be n . Without loss of generality, let $m \leq n$. Let $\text{Cov}(X)$, $\text{Cov}(Y)$ and $\text{Cov}(X, Y)$ be denoted by $\Sigma_{11}^{m \times m}$, $\Sigma_{22}^{n \times n}$, and $\Sigma_{12}^{m \times n}$, respectively, with the superscripts denoting the dimensions of the matrices. We will drop the superscripts when there is no scope for confusion. We will assume that the $(m+n) \times (m+n)$ dimensional square matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

is positive definite, where $\Sigma_{21} = \Sigma_{12}^T$. We will generalize the concept of multiple correlation and describe the correlation structure resulting from the elements of Σ in terms of appropriately chosen pairs of canonical variables.

We will consider linear combinations of the form

$$U_1 = a_1^T X \quad \text{and} \quad V_1 = b_1^T Y \quad [1]$$

where a_1 and b_1 are m and n dimensional coefficient vectors, respectively. From standard statistical theory, we have

$$\text{Var}(U_1) = a_1^T \Sigma_{11} a_1, \quad \text{Var}(V_1) = b_1^T \Sigma_{22} b_1 \quad \text{and}$$

$$\text{Cov}(U_1, V_1) = a_1^T \Sigma_{12} b_1$$

Thus, the correlation between the variables U_1 and V_1 is given by

$$\text{Corr}(U_1, V_1) = \frac{a_1^T \Sigma_{12} b_1}{\sqrt{a_1^T \Sigma_{11} a_1} \sqrt{b_1^T \Sigma_{22} b_1}} \quad [2]$$

Our aim is to choose a_1 and b_1 to maximize the above correlation. The quantity

$$\max_{a_1, b_1} \text{Corr}(U_1, V_1) \quad [3]$$

obtained as a result of the above maximization will be referred to as the first canonical correlation between the

sets X and Y and the corresponding variable pair will be denoted as the first canonical variable pair. Later we will continue the description in terms of the second, third, and subsequent canonical variable pairs.

The development of the canonical correlations will require a few mathematical tools involving eigenvalues and eigenvectors of nonnegative definite matrices, which we state in the following (without proofs). The proofs are available in most standard texts of linear algebra or applied multivariate analysis. Johnson and Wichern (2006), for example, is a typical reference.

Lemma 1: Suppose A is a $p \times p$ nonnegative definite matrix having eigenvalues $\lambda_1 \geq \dots \geq \lambda_j \geq \dots \geq \lambda_p$, with corresponding orthonormal eigenvectors $e_1, \dots, e_j, \dots, e_p$.

- (i) Let s be any p dimensional vector. We are interested in the maximum value of the normalized quadratic form in A over all non-null vectors s . This maximum is given by

$$\max_{s \neq 0} \frac{s^T A s}{s^T s} = \lambda_1 \quad [4]$$

where λ_1 is the largest eigenvalue of A . The maximum is attained when $s = e_1$.

- (ii) Among all nonnull p dimensional vectors s which are orthogonal to e_1, \dots, e_{j-1} , $j = 2, \dots, p$, the same maximization problem leads to

$$\max_{s \perp e_1, \dots, e_{j-1}} \frac{s^T A s}{s^T s} = \lambda_j \quad [5]$$

and this maximum is attained at $s = e_j$.

Define the matrices M and N as

$$\begin{aligned} M &= \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1/2} \quad \text{and} \\ N &= \Sigma_{22}^{-1/2} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1/2} \end{aligned} \quad [6]$$

where $\Sigma_{11}^{-1/2}$ and $\Sigma_{22}^{-1/2}$ are positive definite square roots of Σ_{11}^{-1} and Σ_{22}^{-1} , respectively. Notice that the matrix M can be written in the form BB^T where $B = B_{m \times n} = \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2}$, and hence M is non-negative definite. Similarly, N can be written as $B^T B$, and hence N is non-negative definite as well. This representation also shows that the set of nonzero eigenvalues of M and N are the same (this is a consequence of the eigenvalues of CD and DC being the same, wherever the multiplication of the matrices C and D , and D and C , are defined); in particular if ρ_1 is the largest eigenvalue of M , it is also the largest eigenvalue of N . Also if e is a normalized eigenvector of M corresponding to some nonzero eigenvalue ρ , then $B^T e$ is an eigenvector of N associated with the eigenvalue ρ ; division by $\sqrt{\rho}$ reduces the length of this eigenvector to 1. Thus,

$$f = B^T e / \sqrt{\rho} \quad [7]$$

is a normalized eigenvector of N associated with the eigenvalue ρ .

With this background, we are now ready to state and prove the following theorems.

Theorem 1: The elements of the first canonical variable pair (U_1, V_1) are given by

$$U_1 = e_1^T \Sigma_{11}^{-1/2} X \quad \text{and} \quad V_1 = f_1^T \Sigma_{22}^{-1/2} Y$$

where e_1 and f_1 are the normalized eigenvectors associated with the largest eigenvalue ρ_1 of M and N , respectively. The value of the first canonical correlation is $\sqrt{\rho_1}$.

Proof: We consider linear combinations $U_1 = a_1^T X$ and $V_1 = b_1^T Y$ with the aim of maximizing $\text{Corr}(U_1, V_1)$ over a_1 and b_1 . Let $c_1 = \Sigma_{11}^{-1/2} a_1$ and $d_1 = \Sigma_{22}^{-1/2} b_1$. Using eqn [2], we can represent the correlation coefficient between U_1 and V_1 :

$$\begin{aligned} \text{Corr}(U_1, V_1) &= \frac{a_1^T \Sigma_{12} b_1}{\sqrt{a_1^T \Sigma_{11} a_1} \sqrt{b_1^T \Sigma_{22} b_1}} \\ &= \frac{c_1^T \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} d_1}{\sqrt{c_1^T c_1} \sqrt{d_1^T d_1}} \end{aligned} \quad [8]$$

By a simple application of the Cauchy-Schwarz theorem, we observe that the numerator on the right-hand side of the quantity in eqn [8] is bounded above by

$$\left(c_1^T \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} \Sigma_{21} \Sigma_{11}^{-1/2} c_1 \right)^{1/2} (d_1^T d_1)^{1/2}$$

and the substitution of this in eqn [8] gives

$$\begin{aligned} \text{Corr}(U_1, V_1) &\leq \frac{(c_1^T \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} \Sigma_{21} \Sigma_{11}^{-1/2} c_1)^{1/2}}{(c_1^T c_1)^{1/2}} \\ &= \frac{(c_1^T M c_1)^{1/2}}{(c_1^T c_1)^{1/2}} \leq \sqrt{\rho_1} \end{aligned} \quad [9]$$

The last inequality follows from eqn [4]. Thus, $\text{Corr}(U_1, V_1)$ is bounded by $\sqrt{\rho_1}$, where ρ_1 is the largest eigenvalue of M .

It remains to show that $\text{Corr}(U_1, V_1)$ attains this bound for the particular choices $a_1 = \Sigma_{11}^{-1/2} e_1$ and $b_1 = \Sigma_{22}^{-1/2} f_1$. Since e_1 and f_1 are normalized eigenvectors, direct substitution of the expressions in eqn [8] gives

$$\text{Corr}(U_1, V_1) = \frac{(e_1^T \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} f_1)}{\sqrt{(e_1^T e_1)} (f_1^T f_1)} = e_1^T B f_1 \quad [10]$$

An application of eqn (7) shows

$$\text{Corr}(U_1, V_1) = \frac{e_1^T B B^T e_1}{\sqrt{\rho_1}} = \frac{e_1^T M e_1}{\sqrt{\rho_1}} = \frac{\rho_1 e_1^T e_1}{\sqrt{\rho_1}} = \sqrt{\rho_1} \quad [11]$$

which establishes the desired result.

The last part of the proof can also be shown by checking that these particular choices of a_1 and b_1 lead to equalities in both the inequalities of eqn [9].

Notice that since we are maximizing a correlation which is a scale-independent measure, the value of the correlation would remain the same if the coefficient vectors a_1 and b_1 were multiplied by any scalar constants. The

particular choice given here ensures that the variance of the variables U_1 and V_1 are equal to unity. This is presented more formally later in Theorem 3.

Theorem 2: The j th pair of canonical variables (U_j, V_j) , $j = 2, \dots, m$, are given by

$$U_j = \mathbf{e}_j^T \Sigma_{11}^{-1/2} X \quad \text{and} \quad V_j = \mathbf{f}_j^T \Sigma_{22}^{-1/2} Y$$

where \mathbf{e}_j and \mathbf{f}_j are the orthonormal eigenvectors associated with the j th largest eigenvalue ρ_j of M and N . The j th canonical correlation is given by $\text{Corr}(U_j, V_j) = \sqrt{\rho_j}$.

Proof: We will determine linear combinations $U_2 = \mathbf{a}_2^T X$ and $V_2 = \mathbf{b}_2^T Y$ such that $\text{Corr}(U_2, V_2)$ is maximized subject to $\text{Cov}(U_1, U_2) = \text{Cov}(V_1, V_2) = 0$. By denoting $\mathbf{a}_2 = \Sigma_{11}^{-1/2} \mathbf{c}_2$ and $\mathbf{b}_2 = \Sigma_{22}^{-1/2} \mathbf{d}_2$, we get

$$\begin{aligned} \text{Cov}(U_1, U_2) &= \text{Cov}(\mathbf{e}_1^T \Sigma_{11}^{-1/2} X, \mathbf{c}_2^T \Sigma_{11}^{-1/2} X) \\ &= \mathbf{e}_1^T \Sigma_{11}^{-1/2} \Sigma_{11} \Sigma_{11}^{-1/2} \mathbf{c}_2 \end{aligned}$$

Thus, $\text{Cov}(U_1, U_2) = 0$ implies that \mathbf{c}_2 must belong to the orthogonal space of \mathbf{e}_1 . Similarly, \mathbf{d}_2 can be shown to belong to the orthogonal space of \mathbf{f}_1 .

Therefore, we will have to maximize

$$\text{Corr}(U_2, V_2) = \frac{\mathbf{c}_2^T \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} \mathbf{d}_2}{\sqrt{\mathbf{c}_2^T \mathbf{c}_2} \sqrt{\mathbf{d}_2^T \mathbf{d}_2}} \quad [12]$$

subject to the restrictions on \mathbf{c}_2 and \mathbf{d}_2 . As in Theorem 1, an application of the Cauchy–Schwarz inequality gives

$$\text{Corr}(U_2, V_2) \leq \frac{(\mathbf{c}_2^T M \mathbf{c}_2)^{1/2}}{(\mathbf{c}_2^T \mathbf{c}_2)^{1/2}} \leq \sqrt{\rho_2} \quad [13]$$

where the last inequality follows from Lemma 1 (ii). Direct substitution shows that equality is attained for $\mathbf{c}_2 = \Sigma_{11}^{-1/2} \mathbf{e}_2$ and $\mathbf{d}_2 = \Sigma_{22}^{-1/2} \mathbf{f}_2$.

The result for the successive canonical correlations can be proved in a similar fashion.

Theorem 3: The canonical variable pairs have the properties

$$\text{Var}(U_j) = \text{Var}(V_j) = 1, \quad j = 1, \dots, m \quad [14]$$

$$\text{Cov}(U_j, U_k) = \text{Corr}(U_j, U_k) = 0, \quad j \neq k \quad [15]$$

$$\text{Cov}(V_j, V_k) = \text{Corr}(V_j, V_k) = 0, \quad j \neq k \quad [16]$$

$$\text{Cov}(U_j, V_k) = \text{Corr}(U_j, V_k) = 0, \quad j \neq k \quad [17]$$

Proof: Property [14] is easily verifiable from the definition of the canonical variables, and properties [15] and [16] follow from the construction of the canonical correlation

pairs as outlined in Theorem 2. For property [17], note that

$$\begin{aligned} \text{Cov}(U_j, V_k) &= \mathbf{e}_j^T \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} \mathbf{f}_k \\ &= \mathbf{e}_j^T M \mathbf{e}_k / \sqrt{\rho_k} \\ &= 0 \end{aligned}$$

The last inequality follows from the fact that the vectors \mathbf{e}_j and \mathbf{e}_k are orthonormal eigenvectors of M . The previous equality follows from eqn [7].

In real situations where actual data are generated from the variables of interest, the theoretical quantities in the setup described in Theorems 1–3 are replaced by their sample analogs. Thus, for example, the population covariance matrix Σ will be replaced by the sample covariance matrix S .

Miscellaneous Issues

Canonical correlations are, in general, abstract quantities and one must be careful when interpreting them. At times, some meaningful interpretation can be attached to these by going to lower levels appropriately. By the nature of their construction, canonical correlations are always positive. For two scalar variables, the (single) canonical correlation between them is the absolute value of their ordinary correlation coefficient. The (single) canonical correlation between a scalar variable and a group of variables is – as referred to earlier – quantified by the multiple correlation coefficient. The multiple correlation coefficient is linked to the multiple regression setup and has clear interpretability; it is the correlation between a dependent variable and its best linear predictor.

In the case of two groups each having more than one variable, the j th canonical correlation may be viewed as the multiple correlation between U_j and Y (or the multiple correlation between V_j and X).

An Example with Real Data

In this section, we present a simple example to demonstrate the ideas presented in this article. The data, presented in Waugh (1957), are also available at

<http://lib.stat.cmu.edu/DASL/Datafiles/agecondat.html>

which contain information on the price and consumption per capita of beef and pork in the United States annually from 1925 to 1941 together with other variables relevant to an economic analysis of price and/or consumption of beef and pork over the period (Table 1). The variables that we included in our analysis are

X_1 = Price of beef (cents/lb)

X_2 = Price of pork (cents/lb)

Y_1 = Consumption of beef per capita (lbs)

Y_2 = Consumption of pork per capita (lbs)

The correlation matrix of $(X_1, X_2, Y_1, Y_2)^T$ is given by

$$\begin{pmatrix} 1 & 0.3268 & -0.7524 & -0.3462 \\ 0.3268 & 1 & 0.0834 & -0.7837 \\ -0.7524 & 0.0834 & 1 & -0.2915 \\ -0.3462 & -0.7837 & -0.2915 & 1 \end{pmatrix}$$

The scatter plots of all the pairs of (X, Y) variables are presented in **Figure 1**. As one would expect, X_1 and Y_1 are strongly (negatively) correlated, while X_2 and Y_2 also have a high negative correlation.

The canonical coefficients (a_1 and a_2) for the X variables are given along the columns of the following matrix

Table 1 Consumption and price of beef and pork over the years 1925–41

Year	X_1	X_2	Y_1	Y_2
1925	59.7	60.5	58.6	65.8
1926	59.7	63.3	59.4	63.3
1927	63.0	59.9	53.7	66.8
1928	71.0	56.3	48.1	69.9
1929	71.0	55.0	49.0	68.7
1930	74.2	59.6	48.2	66.1
1931	72.1	57.0	47.9	67.4
1932	79.0	49.5	46.0	69.7
1933	73.1	47.3	50.8	68.7
1934	70.2	56.6	55.2	62.2
1935	82.2	73.9	52.2	47.7
1936	68.4	64.4	57.3	54.4
1937	73.0	62.2	54.4	55.0
1938	70.2	59.9	53.6	57.4
1939	67.8	51.0	53.9	63.9
1940	63.4	41.5	54.2	72.4
1941	56.0	43.9	60.0	67.4

$$\begin{pmatrix} 0.1265 & 0.0859 \\ 0.0341 & -0.1265 \end{pmatrix}$$

and the canonical coefficients for the Y variables are, column-wise,

$$\begin{pmatrix} -0.2066 & -0.1303 \\ -0.1191 & 0.1028 \end{pmatrix}$$

The two canonical correlations for our variables are equal to 0.9728 and 0.7034. The scatter plot of the canonical variables is given in **Figure 2**. The strong linear nature of the relationship among the variables in the first canonical variable pair is clearly observed both from the magnitude of the first canonical correlation, as well as from the corresponding scatter plot. This relationship is much stronger than that among any of the pairs of the original variables. Even the second canonical correlation has a fairly strong linear dependence and the second canonical correlation has a reasonably high value. The scatter plots of the cross canonical variables are presented in **Figure 3** where we plot U_2 against U_1 , V_2 against V_1 , V_2 against U_1 , and V_1 against U_2 .

As expected, the scatter plots in **Figure 3** exhibit a completely random pattern, consistent with the zero correlation these pairs are expected to have. Numerically, the actual computed correlations are

$$\text{Corr}(U_1, U_2) = 9.7145 \times 10^{-17}$$

$$\text{Corr}(U_1, V_2) = 1.1102 \times 10^{-16}$$

$$\text{Corr}(V_1, V_2) = -3.8858 \times 10^{-16}$$

$$\text{Corr}(U_2, V_1) = -1.3878 \times 10^{-16}$$

which are practically equal to zero. On the whole, the canonical correlations give a neat description of the correlation structure of the two groups of variables in this example.

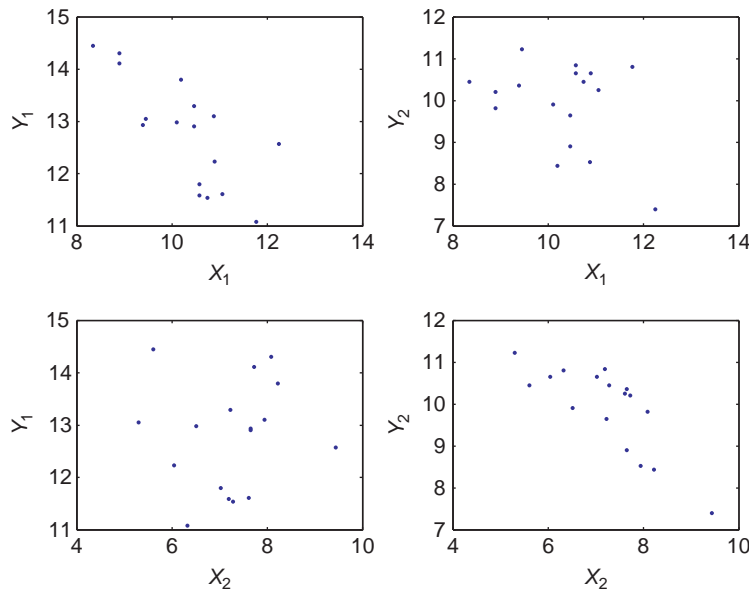


Figure 1 Scatter plots of the Y variables versus the X variables.

Uses of Canonical Correlations

In this section, we provide a very brief outline of the use of canonical correlations in a variety of different settings and some basic references for it for the benefit of the interested reader.

There are many standard textbooks in multivariate analysis which deal adequately with canonical correlations. Useful references, among others, include Anderson (2003) and Johnson and Wichern (2006).

Many useful web-based tutorials of canonical correlation are also available on the Internet (see, e.g., Borga (1999) and Garson (2008)).

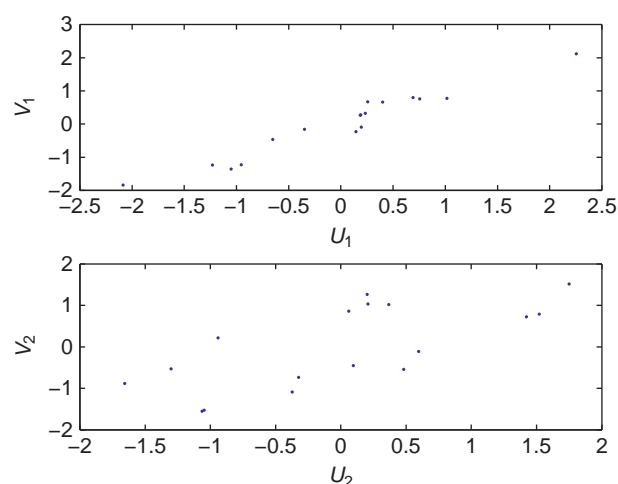


Figure 2 Scatter plots of the canonical variable pairs.

As explained earlier, the interpretation of canonical correlations can be a tricky matter, and there is an expanded literature which deals with this issue. (See, for example, Alpert and Peterson (1972) who discuss possible difficulties of interpreting canonical correlations as measures of the magnitude of relationship between two sets of variables.)

The canonical correlation idea can be extended to define generalized canonical correlation analysis (GCCA) where the classical two set canonical correlation approach is extended to more than two sets. For more details, see Kettenring (1971).

Canonical correlation analysis, in its standard setting, studies the linear relationship between the canonical variables. This approach may be generalized to study the nonlinear relation between two sets of random variables (see Gifi (1990, Chapter 6) for a useful discussion of nonlinear canonical correlation analysis (NCCA)). The method of kernel canonical correlation analysis (KCCA) has also been found to be useful when the embedding of common features is nonlinear or the relation between features is not Gaussian (see, e.g., Akaho (2001)).

Multivariate normality is usually demanded for significance testing in canonical correlation analysis. Choice of an adequate sample size to ensure reasonable power has received some attention in the literature in this context (see Barcikowski and Stevens (1975) and Stevens (1986) for some discussion and recommendations on this issue).

Sometimes, the canonical coefficients are used to determine which original variable(s) a canonical correlation is primarily associated with. This is generally a reasonable approach but can lead to incorrect results when there is high multicollinearity in the data. More discussion

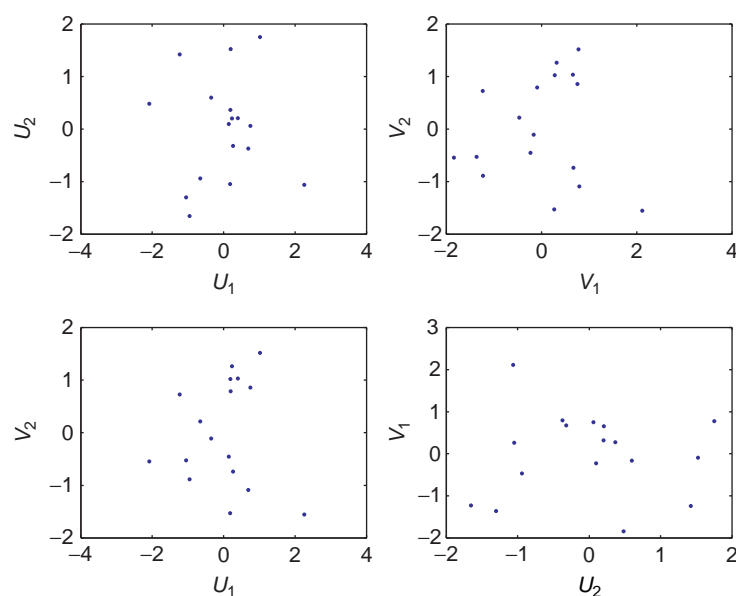


Figure 3 Scatter plots of the cross-canonical variables.

together with alternative recommendations can be found in Levine (1977).

Sometimes, some natural restrictions on the canonical correlations have to be taken into account when performing the analysis. Approaches to perform canonical correlation analysis under such restrictions have been discussed in Das and Sen (1994).

Bibliography

- Akaho, S. (2001). A kernel method for canonical correlation analysis. *Proceedings of the International Meeting of the Psychometric Society (IMPS2001)*, Osaka.
- Alpert, M. I. and Peterson, R. A. (1972). On the interpretation of canonical analysis. *Journal of Marketing Research* **9**(2), 187–192.
- Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd edn. Hoboken, NJ: Wiley.
- Barcikowski, R. and Stevens, J. P. (1975). A Monte Carlo study of the stability of canonical correlations, canonical weights, and canonical variate-variable correlations. *Multivariate Behavioral Research* **10**, 353–364.
- Borga, M. (1999). *Canonical Correlation*. <http://people.imt.liu.se/~magnus/cca> (accessed August 2009).
- Das, S. and Sen, P. K. (1994). Restricted canonical correlations. *Linear Algebra and Its Applications* **210**, 29–47.
- Garson, G. D. (2008). *Canonical Correlation*. <http://www2.chass.ncsu.edu/garson/pa765/canonic.htm> (accessed August 2004).
- Gifi, A. (1990). *Nonlinear Multivariate Analysis*. New York: Wiley.
- Hotelling, H. (1935). The most predictable criterion. *Journal of Educational Psychology* **26**, 139–142.
- Hotelling, H. (1936). Relations between two sets of variables. *Biometrika* **28**, 321–377.
- Johnson, R. A. and Wichern, D. W. (2006). *Applied Multivariate Statistical Analysis*, 6th edn. Englewood Cliffs, NJ: Prentice-Hall.
- Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika* **58**, 433–451.
- Levine, M. S. (1977). *Canonical Analysis and Factor Comparison, Quantitative Applications in the Social Sciences Series, No. 6*. Thousand Oaks, CA: Sage.
- Stevens, J. (1986). *Applied Multivariate Statistics for the Social Sciences*. Hillsdale, NJ: Erlbaum.
- Waugh, F. B. (1957). *Graphic Analysis in Agricultural Economics. Agricultural Handbook No. 128*, Washington, DC: US Department of Agriculture.

Categorical Data Analysis

S E Fienberg and B W Junker, Carnegie Mellon University, Pittsburgh, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Categorical data are ubiquitous and essential in education research. Examples include explanatory variables, such as gender, ethnicity, type of instruction, etc., and response variables such as individual test item scores, total test scores, and achievement levels. Understanding dependence – and independence – relationships among such variables is an essential statistical task in education research. In this article, we survey four key tools for this task: log-linear models for contingency tables, the Cochran–Mantel–Haenszel (CMH) test, logistic regression, and latent variable models.

Contingency Tables

A contingency table consists of counts of units cross-classified according to the values of several categorical (nominal or ordinal) variables. **Table 1** contains counts cross-classified units according to two nominal variables, gender (male/female) and white (yes/no); and two ordinal variables, frequency of math drills (never, seldom, weekly, often, daily) and NAEP achievement level (below basic, basic, proficient, advanced).

Testing for Relationships Among Categorical Variables

What can we learn from **Table 1**? For example, do male students typically perform above, or below, female students? Summing the counts for male and female students separately we arrive at **Table 2**, to which we can apply the Pearson chi-squared test of statistical independence (e.g., Fienberg, 1980), finding $X^2 = 1.930$ on 3 degrees of freedom (df). There is no evidence to support the hypothesis of differential performance on the assessment by gender ($P(X^2 > 1.930) = 0.59$).

Log-linear models (Bishop *et al.*, 1975; Fienberg, 1980) can help extend analyses like this. For **Table 2**, write each count as n_{ij} , indexed by i for gender and j for achievement level, and write the log-linear model as

$$\log E(n_{ij}) = \log(p_{ij}n) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \quad [1]$$

which has the same form as an ANOVA model for the factors u_1 (gender) and u_2 (achievement level). Here, $p_{ij} = P$

(gender = i achievement = j) and $n = \sum_i \sum_j n_{ij}$, with constraints such as

$$\sum_{i=1}^2 u_{1(i)} = \sum_{j=1}^4 u_{2(j)} = \sum_{i=1}^2 u_{12(ij)} = \sum_{j=1}^4 u_{12(ij)} = 0 \quad [2]$$

following Bishop *et al.* (1975). (Any single linear constraint for each subscript in a subscripted u -term, such as setting the first or last value equal to zero is equivalent for the purpose of these expectations.) The above chi-squared test is equivalent to testing whether the interaction terms $u_{12(ij)}$ are all zero in the log-linear model. Model [1] with the $u_{12(ij)}$ all zero is equivalent to statistical independence since, in that case, from eqn [1], $p_{ij}n = \exp(u) \cdot \exp(u_{1(i)}) \cdot \exp(u_{2(j)})$, and this can be used to express

$$p_{ij} = p_i \times p_j \quad [3]$$

Alternatively, the likelihood-ratio statistic comparing the log-additive model ($u_{12(ij)} \equiv 0$) to the saturated model in eqn [1] is $G^2 = 1.929$ on 3 df. The X^2 and G^2 statistics are usually close in value, differing mainly in the presence of small cell counts.

For a 2×2 table such as **Table 3**, another common measure of dependence is the odds ratio

$$\begin{aligned} OR &= \frac{\text{odds}(X_2 = 2 | X_1 = 2)}{\text{odds}(X_2 = 2 | X_1 = 1)} \\ &= \frac{P(X_2 = 2 | X_1 = 2) / (1 - P(X_2 = 2 | X_1 = 2))}{P(X_2 = 2 | X_1 = 1) / (1 - P(X_2 = 2 | X_1 = 1))} \\ &= \frac{p_{11}p_{22}}{p_{12}p_{21}} \end{aligned}$$

which can be estimated as

$$\overline{OR} = \frac{n_{11}n_{22}}{n_{12}n_{21}} \quad [4]$$

From eqns [1] and [2]

$$u_{12(11)} = \frac{1}{4} \log(OR) \quad [5]$$

Hence we can use the log-linear model in eqn [1] to estimate the odds ratio. (Had we used another type of constraint, for example, setting $u_{12(11)} = u_{12(11)} = u_{12(11)} = 0$, then $u_{12(22)} = \log(OR)$. In all cases, we have a scaled function of $\log(OR)$.) When $OR > 1$ ($u_{12(11)} > 0$), the variables in the table are positively associated; when $OR < 1$ ($u_{12(11)} < 0$), they are negatively associated. $OR = 1$ ($u_{12(11)} = 0$) corresponds to statistical independence, as can be seen by substituting [3] into [4].

Table 1 Number of eighth-grade students at each achievement level, cross-classified by gender, ethnicity (white/nonwhite), and frequency of mathematics drills in the classroom

Gender	White	Frequency of Math drill	Mathematics achievement level			
			Below	Basic	Proficient	Advanced
Male	Yes	Never	19	54	54	19
Male	Yes	Seldom	75	118	75	17
Male	Yes	Weekly	352	390	171	29
Male	Yes	Often	198	208	68	10
Male	Yes	Daily	118	95	33	5
Male	No	Never	13	18	9	3
Male	No	Seldom	40	29	10	2
Male	No	Weekly	117	57	13	2
Male	No	Often	68	35	8	1
Male	No	Daily	23	9	1	0
Female	Yes	Never	29	81	63	17
Female	Yes	Seldom	96	177	113	16
Female	Yes	Weekly	315	378	180	27
Female	Yes	Often	253	242	75	12
Female	Yes	Daily	102	73	20	4
Female	No	Never	7	9	4	1
Female	No	Seldom	47	37	11	1
Female	No	Weekly	113	64	18	2
Female	No	Often	86	40	7	0
Female	No	Daily	26	8	1	0

The counts resemble data from the 2007 National Assessment of Educational Progress National Mathematics Assessment.

Table 2 Table 1 collapsed to gender and achievement level only

Gender	Mathematics achievement level			
	Below	Basic	Proficient	Advanced
Male	1023	1013	442	88
Female	1074	1109	492	80

Table 3 Typical 2×2 table

Gender (X_1)	Achievement level (X_2)	
	Basic	Proficient
Male	n_{11}	n_{12}
Female	n_{21}	n_{22}

Hierarchical Log-Linear Models

A shorthand notation helps in specifying interpretable log-linear models: the notation $[1][2]$ refers to the additive model $u + u_{1(i)} + u_{2(j)}$, the notation $[12]$ refers to the model with $u_{12(jk)}$ and all lower-order terms: $u + u_{1(i)} = u_{2(j)} = u_{12(ij)}$ as in eqn [1], and so forth.

These are called hierarchical log-linear models because, by convention, the inclusion of interaction terms like $[12]$ implies inclusion of all lower-order terms; this is known as the hierarchy principle. For example, for a four-way table like **Table 1**, the model $[134][234]$ includes the two three-way interaction terms $u_{134(ikl)}$ and $u_{234(jkl)}$, as well as all two-way interactions except for $u_{12(ij)}$ and all main effects, that is,

$$\log E(n_{ijkl}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + u_{13(ik)} + u_{14(il)} + u_{34(kl)} + u_{23(jk)} + u_{24(jl)} + u_{134(ikl)} + u_{234(jkl)}$$

Holding factors 3 and 4 at fixed levels k_0 and l_0 in this model, most terms in the model are constant and the only terms that vary with factors 1 and 2 are additive,

$$\log E(n_{ijk_0l_0}) = u + u_{3(k_0)} + u_{4(l_0)} + u_{34(k_0l_0)} + u_{1(i)} + u_{13(ik_0)} + u_{14(il_0)} + u_{134(ik_0l_0)}$$

indicating conditional independence of factors 1 and 2, given factors 3 and 4.

We use hierarchical log-linear models to assess dependence among multiple variables cross-classifying counts in a contingency table. Special cases of these models, the log-linear graphical and decomposable models, allow us to search large contingency tables efficiently for independence, conditional independence, and dependence relationships that are suggestive about the causal structure of the underlying data generating process.

When the model of statistical independence $[1][2][3][4]$

$$\log E(n_{ij}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)}$$

is applied to **Table 1**, the likelihood-ratio statistic is $G^2 = 549.18$ on 70 df, strongly suggesting that the four factors are not statistically independent. Further likelihood-ratio testing reveals that the two-way interactions

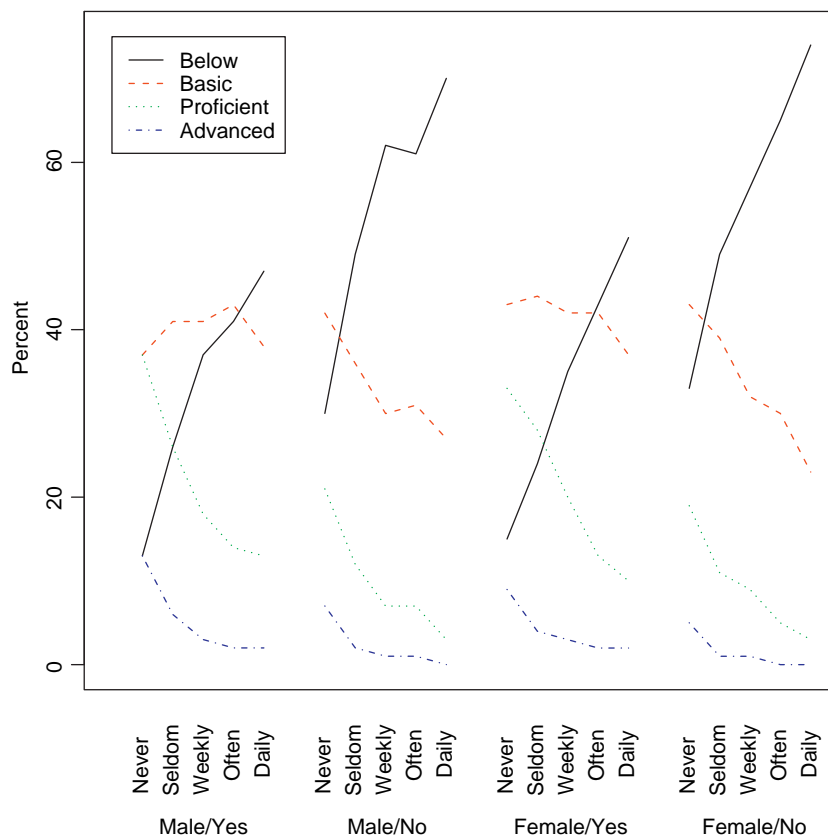


Figure 1 Achievement level percentages within gender, ethnicity, and math drill frequency, computed from **Table 1**.

(also known as first-order interactions) and some three-way interactions (also known as second-order interactions) should be included in the model. But the model [123][423] fits well, $G^2 = 12.75$ on 30 df, implying that gender (factor 1) is conditionally independent of achievement (factor 4) given ethnicity and drill frequency (factors 2 and 3).

Figure 1 shows the percentage of students at each achievement level in each row of **Table 1**. Evidently, the general achievement pattern is the same for males and females, confirming our log-linear analysis, although there are some differences in levels according to ethnicity that might warrant further exploration.

More troubling is that the percentage at the below basic achievement level increases with frequency of math drills, whereas percentages at the higher achievement levels all decrease as math drills become more frequent. A more detailed analysis would show that these trends are statistically significant, not explainable as chance variation alone. Since math drills precede the NAEP assessment, does this suggest that more math drills cause students to perform more poorly? Perhaps, a hidden third variable – students' prior mathematics achievement – influences both the NAEP test score and the decision to expose students to more or less frequent math drills. Thus, it is wise to keep the maxim "correlation (or association) does not mean causation" clearly in mind.

CMH Procedure

Simpson's Paradox and Association Reversal Phenomena

For **Table 1**, we hypothesized that the relationship between frequency of math drill and achievement level could be explained by an unobserved third variable, students' prior mathematics achievement. Suppose, for the sake of argument, that frequency of math drills has no effect on achievement levels among students at the same level of prior mathematics achievement – that is, math drills and NAEP achievement level are conditionally independent, given prior achievement. If, as prior achievement goes down, the frequency of math drills goes up and NAEP achievement level goes down, then when we collapse over prior achievement, math drills and NAEP achievement level will appear to be negatively correlated; this is a version of Simpson's or Yule's paradox (Fienberg, 1980). Even if there is a positive association between math drills and NAEP achievement level within each level of prior achievement, this association can be wiped out or even reversed by collapsing.

For example, consider **Table 4**. Here, students have been cross-classified according to meeting or not meeting an achievement standard on a post-test, experiencing or not experiencing math drills, and total score on a short

Table 4 Cross-classification of math drills (yes/no), pretest score (0–4), and meeting standard (yes/no) for 2694 hypothetical students

	<i>Pretest = 0</i>		<i>Pretest = 1</i>		<i>Pretest = 2</i>		<i>Pretest = 3</i>		<i>Pretest = 4</i>		<i>All students</i>	
	<i>Meet</i>	<i>Standard</i>	<i>Meet</i>	<i>Standard</i>	<i>Meet</i>	<i>Standard</i>	<i>Meet</i>	<i>Standard</i>	<i>Meet</i>	<i>Standard</i>	<i>Meet</i>	<i>Standard</i>
<i>Drill</i>	<i>Yes</i>	<i>No</i>	<i>Yes</i>	<i>No</i>	<i>Yes</i>	<i>No</i>	<i>Yes</i>	<i>No</i>	<i>Yes</i>	<i>No</i>	<i>Yes</i>	<i>No</i>
Yes	93	19	150	88	212	170	153	177	53	180	661	634
No	300	87	197	168	170	199	79	133	11	55	757	642

The far-right subtable collapses (sums) over pretest score groups.

pretest (0–4). In each 2×2 subtable (for pretest = 0, for pretest = 2, etc.) the association between math drill and meet standard is positive ($OR > 1.4$ in each subtable). However, when we collapse (sum) across pretest scores to obtain a single 2×2 table for all students, the association becomes negative ($OR = 0.88$).

Common Odds Ratio and the CMH Procedure

The complete independence log-linear model [1][2][3] provides an unacceptably poor fit to Table 4 ($G^2 = 624$ on 13 df), but the ‘no second-order interaction’ model, [12][13][23], fits quite well ($G^2 = 0.009$ on 4 df). The absence of the three-way interaction means that the odds ratio between math drill and meets standard in each pretest subtable is independent of the pretest score. Therefore, in estimating this common odds ratio and testing independence between math drill and meet standard, we should pool data across pretest score groups to increase estimation precision and statistical power. However, as shown in Table 4, we cannot pool by simply summing cell counts across these groups.

Instead, we apply the CMH procedure, which provides both an estimate of the common odds ratio, and a chi-squared test of conditional independence of math drill and meets standard, given pretest score (i.e., a test of the null hypothesis that the common odds ratio equals 1). The CMH estimate of the common odds ratio pools the numerator and denominator of eqn [4] separately

$$OR_{MH} = \frac{\sum_{k=1}^K n_{11k}n_{22k}/n_{..k}}{\sum_{k=1}^K n_{12k}n_{21k}/n_{..k}}$$

where n_{ijk} is the observed count at level i of math drill, j of meets standard, and k of pretest score, and $n_{..k}$ is the number of students in pretest score group k . We compute confidence intervals for OR_{MH} using standard errors for $\log(OR_{MH})$ and a test for whether $OR_{MH} = 1$ by comparing the square of $\log(OR_{MH})$ divided by its standard error to values of the X^2 distribution on 1 df. (We can accomplish a similar test using a conditional likelihood-ratio test for the model [23][13] given the model [12][23][13], and the difference between the corresponding G^2 values. The CMH test for conditional independence is, however, the uniformly most powerful test (see Holland and Thayer, 1988).

Applying the CMH procedure to Table 4, we find strong evidence against the model of conditional independence of math drill with meeting standard given pretest score) is rejected ($X^2 = 18$ on 1 df), and we estimate the common odds ratio as 1.45, with 95% confidence interval (1.23, 1.72). There is a noticeable positive association between experiencing the math drills and meeting the standard on the post-test, for students in each pretest group.

Application to Differential Item Functioning (DIF) Analysis

The CMH procedure appears regularly in large-scale educational testing to assess whether test items function differently in different social groups. Let i represent the groups of interest, generally called the reference group ($i = 1$) and the focal group ($i = 2$), and let j represent correct ($j = 1$) or incorrect ($j = 2$) response to a test item that is suspected to function differently in the two groups. Finally, let k be the scores on an anchor test that is known or assumed not to function differentially in the two groups. Then, the CMH test of conditional independence is in fact a test of whether group membership (i) affects performance on the item (j) after controlling for students’ achievement level (k). Holland and Thayer (1988) make an important connection between the CMH procedure for DIF detection and item response theory (IRT) models, which we discuss below.

Logistic Regression

The general log-linear model for contingency tables, for example,

$$\log E(n_{ijkl}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(l)} + \dots + u_{12(ij)} + u_{23(jk)} + \dots$$

is an instance of the class of generalized linear models (GLMs; McCullagh and Nelder, 1989), since it expresses the mean of the response variable as a transformation of a linear model

$$E(Y) = g(X\beta) \quad [6]$$

where $Y = n_{ijkl}$ is the cell count, $g(x) = \exp(x)$ is the transformation or link function, X collects together

indicator variables for the levels of the cross-classifying variables, and β collects together the u -parameters. Three distributions – the Poisson, the multinomial, and the product multinomial – are commonly used as sampling distributions for contingency tables; fortunately, maximum-likelihood estimates and standard errors for log-linear models are the same under all three distributions (Bishop *et al.*, 1975; Fienberg, 1980).

Perhaps the most widespread GLM is the logistic regression model. For binary logistic regression, we consider a binary 0–1 response variable Y_i with mean $E(Y_i) = P(Y_i = 1) = p_i$ such that

$$\log\left(\frac{p_i}{1-p_i}\right) = X_i\beta,$$

such that the link function in eqn [6] is $g(x) = \exp(x)/(1 + \exp(x))$, and we use the binomial distribution as the sampling model.

Comparing Log-Linear Models with Logistic Regression

We can apply the logistic regression model to the data of **Table 4**, taking the response variable Y_{ij} to be 1 if a student with pretest score i and math drill level j meets the standard on the post-test, and 0 otherwise. The additive logistic model

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = v_0 + v_{1(i)} + v_{2(j)}$$

fits well, with $G^2 = 0.009$ on 4 df. The perceptive reader will recognize that this is the same likelihood-ratio statistic value as the log-linear model of ‘no three-way interactions’ [12][13][23], which we used to establish that there was a common odds ratio in the section titled ‘Common odds ratio and the CMH procedure.’ This is no accident: a logistic regression model applied in this way to a contingency table is equivalent to the log-linear model with (1) a full interaction for all factors on the right-hand side of the logistic regression model; and (2) an interaction of the response with each term on the right-hand side. Hence, letting factor 3 be the meets standard, the additive logistic regression model above is equivalent to the model [12][13][23] which is exactly the model of no second-order interaction.

At each fixed level i of pretest score, the log-odds-ratio under the additive model for meeting the standard on the post-test between someone who did ($j = 1$) and did not ($j = 2$) experience the math drill condition is

$$\begin{aligned} \log\left(\frac{p_{i1}}{1-p_{i1}}\right) - \log\left(\frac{p_{i2}}{1-p_{i2}}\right) \\ = (v_0 + v_{1(i)} + v_{2(1)}) - (v_0 + v_{1(i)} + v_{2(2)}) \\ = v_{2(1)} - v_{2(2)} \end{aligned}$$

We estimate this to be 0.37 with standard error 0.09, clearly a significant positive association between math drill and meets standard on the post-test. This is, as we would expect, identical to the estimate of the common log-odds-ratio under the log-linear model [12][13][23].

If we were concerned with the possibility of differing odds ratios in different pretest score groups, we could explore this with the interactive logistic regression model

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = v_0 + v_{1(i)} + v_{2(j)} + v_{12(ij)} \quad [8]$$

which is the same as the saturated model [123]. A more parsimonious model would be

$$\begin{aligned} \log\left(\frac{p_i}{1-p_i}\right) = \tau_0 + \tau_1(\text{pretest})_i + \tau_2(\text{mathdrill})_i \\ + \tau_3(\text{pretest})_i \times (\text{mathdrill})_i \end{aligned}$$

where $(\text{pretest})_i$ is student i ’s pretest score (0–4); and $(\text{math drill})_i = 1$ if student i experienced the math drill condition and 0 otherwise. This model, which treats pretest score as an interval variable instead of an ordinal one, does not fit the data in **Table 4** particularly well ($G^2 = 24.78$ on 6 df). In any case, we could estimate odds ratios that vary from one pretest score group to the next from either model [8] or [9] using calculations analogous to eqn [7].

Application to DIF Analysis

The logistic regression approach is also used to assess DIF in large-scale educational testing. The usual approach is to use a model analogous to eqn [9]

$$\begin{aligned} \log\left(\frac{p_i}{1-p_i}\right) = \tau_0 + \tau_1(\text{anchortest})_i + \tau_2(\text{group})_i \\ + \tau_3(\text{anchortest})_i \times (\text{group})_i \end{aligned}$$

where p_i is student i ’s probability of getting the suspect item correct, $(\text{anchor test})_i$ is the score on the anchor test for student i , and $(\text{group})_i$ is student i ’s group membership ($i = 0$ for the reference group, and $i = 1$ for the focal group). The submodel $\tau_2 = \tau_3 = 0$ corresponds to no DIF, $\tau_3 = 0$ corresponds to uniform DIF, with the magnitude of the DIF measured by an effect size based on the log-odds-ratio τ_2 , and finally the full model corresponds to nonuniform DIF, that is, DIF that varies with the anchor test score.

Latent Variable Models of Association

Latent variable models provide an alternative to hierarchical log-linear models for examining dependence structure among categorical random variables. In the typical

latent variable model, we posit an unobserved variable θ_i for each individual i in the study, and assume that observable categorical variables are conditionally independent given θ_i : calling θ factor 0, and the observable variables factors 1, 2, 3, ..., \mathcal{J} , we assume the complete data model (01)(02) ... (0 \mathcal{J}). Conditional independence given the unobservable factor 0 is called local independence in the latent variable literature. We can only observe the categorical variables collapsed over θ , inducing dependence among the observed variables, just as in **Table 4**. This usually leads to a more parsimonious interpretation for the interactions among the \mathcal{J} observable variables than would hierarchical log-linear models.

IRT Models

One of the most widely used and successful classes of latent variable models in educational statistics is the class of IRT models. Here, θ_i is a continuous variable representing student i 's proficiency in an academic area, and $X_{i1}, \dots, X_{i\mathcal{J}}$ are the student's scored responses to a set of \mathcal{J} test items in that area. When the responses are scored 0 for a wrong answer and 1 for a correct answer, we have a simple binary model

$$P(X_{ij} = 1 | \theta_i, \beta_j) = P(\theta_i, \beta_j)$$

where β_j are parameters describing the probability of correct response at a fixed level θ_i . Collapsing the conditional independence model (01)(02) ... (0 \mathcal{J}) over θ (factor 0) leads to a submodel of (123 ... \mathcal{J}),

$$P(x_1, \dots, x_{\mathcal{J}} | \beta_1, \dots, \beta_{\mathcal{J}}) = \int \prod_{j=1}^{\mathcal{J}} P(\theta, \beta_j)^{x_j} (1 - P(\theta, \beta_j))^{1-x_j} dF(\theta) \quad [10]$$

where the product inside the integral represents conditional independence, analogous to independence in the separate 2×2 subtables in **Table 4** (see also eqn [3] above) and the integral itself corresponds to summing across these subtables. Although this model allows for interactions of all orders; it only depends on the parameters $\beta_1, \dots, \beta_{\mathcal{J}}$ and is therefore typically much more parsimonious than the saturated model (123 ... \mathcal{J}), which has $2^{\mathcal{J}}$ parameters.

Because $2^{\mathcal{J}}$ is typically large, relative to N , there will be many zero counts in cells with positive probability under the model, known as sampling zeros. The IRT model effectively smooths over these empty cells, such that it is less sensitive to zero counts than typical hierarchical log-linear models. In case each $P(\theta, \beta_j)$ is nondecreasing in θ , the X_j 's under model (10) will exhibit a particularly strong form of positive dependence, called conditional association (Holland and Rosenbaum, 1986). This can be used to assess fit of the model (10) without knowing the details of the functions $P(\theta, \beta_j)$ or the distribution $dF(\theta)$.

The Rasch Model, Quasi-Symmetry, and DIF Analysis

The special case of the Rasch (1980) model

$$P(X_{ij} = 1 | \theta_i, \beta_j) = \frac{e^{(\theta_i - \beta_j)}}{1 + e^{\theta_i - \beta_j}} \quad [11]$$

provides two important connections between IRT models and log-linear models. For the first connection, Holland (1990) has shown that, for the Rasch model, the logarithm of eqn [10] becomes

$$\log P(x_1, \dots, x_{\mathcal{J}} | b_1, \dots, b_{\mathcal{J}}) = - \sum_{j=1}^{\mathcal{J}} x_j \beta_j + G(x_+; \beta_1, \dots, \beta_{\mathcal{J}})$$

where $x_+ = \sum_{j=1}^{\mathcal{J}} x_j$. This corresponds to the log-linear model

$$\log E(n_{x_1, \dots, x_{\mathcal{J}}}) = \alpha_0 + \sum_{j=1}^{\mathcal{J}} \alpha_j x_j + \sum_{k=0}^K \gamma_k I_k$$

for the counts in the $2^{\mathcal{J}}$ table cross-classifying students' scored responses to the \mathcal{J} test items, where $\alpha_j = -\beta_j$, $\gamma_k = G(k; \beta_1, \dots, \beta_{\mathcal{J}})$, and $I_k = 1$ if $x_+ = k$ and 0 otherwise. This model contains all main effects, and interactions of all orders, but the interactions are constrained to be symmetric: all two-way interactions are equal to γ_2 , all three-way interactions are equal to γ_3 , etc., corresponding to the model of quasi-symmetry. But there are also constraints on the γ_k 's: they are functions of the β_j 's, and obey the order constraints of a sequence of log-moments of the distribution of a non-negative random variable.

For the second connection, suppose we suspect that item \mathcal{J} functions differently in a reference group and a focal group of examinees, but the items 1, 2, ..., $\mathcal{J} - 1$ function the same way in both groups. We could model this in the Rasch model, using one item parameter $\beta_{\mathcal{J}\text{R}}$ for item \mathcal{J} in the reference group and a different parameter $\beta_{\mathcal{J}\text{F}}$ in the focal group, in [11]. The odds ratio – conditional on θ – for correct response to this item between the two groups is

$$OR = \exp(\beta_{\mathcal{J}\text{F}} - \beta_{\mathcal{J}\text{R}})$$

which is invariant across levels of θ . This common odds ratio represents uniform DIF, that is, a constant shift in the location of the item response curve under the Rasch model. Since the total score X_+ is a sufficient statistic for θ in the Rasch model, the MH estimate of the common log-odds ratio, using X_+ as the stratifying variable, is directly interpretable as an estimate of the group difference in difficulty: $\beta_{\mathcal{J}\text{F}} - \beta_{\mathcal{J}\text{R}}$ (Holland and Thayer, 1988; Zwick, 1990).

Other Latent Variable Models

Other latent variable models have become popular for parsimonious and interpretable analysis of contingency table data as well. For example,

- Cognitive diagnosis models (CDMs) are a class of constrained latent class analysis (LCA) models. CDMs begin with a latent variable $\theta_i = (\theta_{i1}, \dots, \theta_{ik})$, where each $\theta_{ik} = 1$ if student i possesses cognitive attribute k , and 0 otherwise. This is equivalent to $C = 2^K$ latent classes, in which each class c is identified with a unique sequence of cognitive attribute indicators $\theta^{(c)} = (\theta_1(c), \dots, \theta_K(c))$. The response probability $P(X_{ij} = 1 | \theta_i, \beta_j) = P(\theta_i, \beta_j)$ is then parametrized in terms of the $\theta_k^{(c)}$'s; for example, a simple conjunctive model (requiring all skills needed for a task to be present, to maximize success on that task) would be

$$P(\theta_i, \beta_j) = \prod_{k=1}^K \left[(1 - s_{jk})^{\theta_k^{(c)}} g_{jk}^{1-\theta_k^{(c)}} \right]^{Q_{jk}} \quad [12]$$

where $Q_{jk} = 1$ if test item j requires cognitive attribute k , and 0 otherwise, and $\beta_j = (g_j, s_j)$. Junker and Sijtsma (2001) provide a brief technical review of some simple CDMs. For a broader overview, see the special issue of *Journal of Educational Measurement* edited by DiBello and Stout (2007).

- The grade of membership (GoM) model can be viewed as a continuous-variables reformulation of the LCA model. The basic LCA model for an individual i in class c , with j th binary response X_{ij} can be written

$$P(X_{ij} = 1 | \theta_i = (0, \dots, 0, 1, 0, \dots, 0)) = \lambda_{jc}$$

where 1 appears in the c th position of the vector $\theta_i = (\theta_{i1}, \dots, \theta_{iC})$, or equivalently

$$P(X_{ij} = 1 | \theta_i) = \sum_{c=1}^C \lambda_{jc} \theta_{ic}$$

In the GoM model, this is a response probability for pure-type c . The GoM model then generalizes this structure by allowing θ_i to be not just a sequence of C 0's and 1's summing to unity, but rather any sequence of C nonnegative weights, summing to unity. Thus, each response probability is a mixture of pure-type probabilities, with mixing proportions equal to the individual's so-called GoM score $\theta_i = (\theta_{i1}, \dots, \theta_{iC})$. The θ_{ic} 's are the individual's grade or degree of membership in each pure type; in this sense the GoM model is like a soft-clustering model. Erosheva *et al.* (2007) show that each GoM model corresponds to an LCA model with a larger number of latent classes but provides a more parsimonious and interpretable representation than the corresponding LCA model.

See also: Dynamic Assessment; Graphical Models; Item Response Theory; Latent Class Models; Statistical Paradoxes.

Bibliography

- Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press. [Reprinted by Springer, 2007.]
- DiBello, L. V. and Stout, W. F. (2007). Guest editors' introduction and overview: IRT-based cognitive diagnostic models and related methods. *Journal of Educational Measurement* **44**, 285–291.
- Erosheva, E. A., Fienberg, S. E., and Joutard, C. (2007). Describing disability through individual-level mixture models for multivariate binary data. *Annals of Applied Statistics* **1**, 502–537.
- Fienberg, S. E. (1980). *Analysis of Cross-Classified Categorical Data*, 2nd edn. Cambridge, MA: MIT Press. [Reprinted by Springer, 2007.]
- Holland, P. W. (1990). The Dutch identity: A new tool for the study of item response models. *Psychometrika* **55**, 5–18.
- Holland, P. W. and Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Annals of Statistics* **14**, 1523–1543.
- Holland, P. W. and Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In Wainer, H. and Braun, H. (eds.) *Differential Item Functioning*. Hillsdale, NJ: Erlbaum.
- Junker, B. W. and Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement* **25**, 258–272.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edn. New York: Chapman and Hall.
- Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests: Expanded Edition with Foreword and Afterword by B.D. Wright*. Chicago, IL: The University of Chicago Press.
- Zwick, R. (1990). When do item response function and Mantel-Haenszel definitions of differential item functioning coincide? *Journal of Educational Statistics* **15**, 185–198.

Further Reading

- Agresti, A. (2002). *Categorical Data Analysis*, 2nd Edn. New York: Wiley.
- Anderson, C. J. and Yu, H. -T. (2007). Log-multiplicative models as association models. *Psychometrika* **72**, 5–23.
- DeBoeck, P. and Wilson, M. (2004). *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York: Springer.
- Christensen, R. (1997). *Log-Linear Models and Logistic Regression*, 2nd edn. New York: Springer.
- Dayton, C. M. and Macready, G. B. (2007). Latent class analysis in psychometrics. In Rao, C. R. and Sinharay, S. (eds.) *Handbook of Statistics, Volume 26: Psychometrics*, Ch. 13, pp 421–446. Amsterdam: North-Holland.
- Edwards, D. (2000). *Introduction to Graphical Modeling*, 2nd edn. New York: Springer.
- Erosheva, E. A. (2005). Comparing latent structures of the grade of membership, Rasch and latent class models. *Psychometrika* **70**, 619–626.
- Fischer, G. H. (1995). The linear logistic test model. In Fischer, G. H. and Molenaar, I. W. (eds.) *Rasch Models: Foundations, Recent Developments and Applications*, Ch. 8, pp 157–180. Berlin: Springer.
- Fox, J. -P. and Glas, C. A. W. (2002). Modeling measurement error in a structural multilevel model. In Marcoulides, G. A. and Moustaki, I. (eds.) *Latent Variable and Latent Structure Models*, pp 245–269. London: Erlbaum.
- Hedeker, D. (2005). Generalized linear mixed models. In Everitt, B. and Howell, D. (eds.) *Encyclopedia of Statistics in Behavioral Science*. New York: Wiley.
- Mariano, L. T. and Junker, B. W. (2007). Covariates of the rating process in hierarchical models for multiple ratings of test items. *Journal of Educational and Behavioral Statistics* **32**, 287–314.
- Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata*, 2nd edn. College Station, TX: Stata Press.

- Schofield, L. S. (2008). Modeling Measurement Error When Using Cognitive Test Scores. In Social Science Research. Ph.D. Dissertation, Department of Statistics. Carnegie Mellon University, Pittsburgh PA.
- van der Linden, W. J. and Hambleton, R. K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer.
- Yen, W. M. and Fitzpatrick, A. R. (2006). Item response theory. In Brennan, R. L. (ed.) *Educational Measurement*. Westport, CT: Praeger.

Relevant Websites

- <http://www.mrc-bsu.com.ac.uk> – MRC Biostatistics Unit, The Bugs Project 2008.
- <http://www.r-project.org> – The R Foundation for Statistical Computing, The R Project 2008.

Causal Inference

D B Rubin, Harvard University, Cambridge, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

A Framework for Causal Inference – Basic Building Blocks

The framework for causal inference that is discussed here is now commonly referred to as the Rubin Causal Model (RCM; Holland, 1986), for a series of articles written in the 1970s (Rubin, 1974, 1976, 1977, 1978, 1980). Other approaches to causal inference, such as graphical ones (e.g., Pearl, 2000), are conceptually less satisfying, for reasons discussed, for instance, in Rubin (2004b, 2005). The presentation here is essentially a brief and relatively nontechnical version of that given in Rubin (2006).

For causal inference, there are several basic building blocks. A unit is a physical object, for example, a person, at a particular point in time. A treatment is an action that can be applied or withheld from that unit. We focus on the case of two treatments, although the extension to more than two treatments is simple in principle but not necessarily so with real data. Associated with each unit are two potential outcomes: the value of an outcome variable Y (e.g., test score) at a point in time t when the active treatment (e.g., new educational program) is used at an earlier time t_0 , and the value of Y at time t when the control educational program is used at t_0 . The objective is to learn about the causal effect of the application of the active treatment relative to the control (treatment) on Y . Formal notation for this meaning of a causal effect first appeared in Neyman (1923) in the context of randomization-based inference in randomized experiments. Let W indicate which treatment the unit received: $W = 1$ the active treatment, $W = 0$ the control treatment. Moreover, let $Y(1)$ be the value of Y if the unit received the active version, and $Y(0)$ the value if the unit received the control version. The causal effect of the active treatment relative to its control version is the comparison of $Y(1)$ and $Y(0)$ – typically the difference, $Y(1) - Y(0)$, or perhaps the difference in logs, $\log[Y(1)] - \log[Y(0)]$, or some other comparison, possibly the ratio. The fundamental problem for causal inference is that, for any individual unit, we can observe only one of $Y(1)$ or $Y(0)$, as indicated by W ; that is, we observe the value of the potential outcome under only one of the possible treatments, namely the treatment actually assigned, and the potential outcome under the other treatment is missing. Thus, inference for causal effects is a missing-data problem – the “other” value is missing. Of importance in educational research, the gain score for a unit, posttest

minus pretest, measures a change in time, and so is not a causal effect.

We learn about causal effects using replication, which involves the use of more than one unit. The way we personally learn from our own experience is replication involving the same physical object (me or you) with more units in time, thereby having some observations of $Y(0)$ and some of $Y(1)$. When we want to generalize to units other than ourselves, we typically use more objects; that is what is done in social science experiments, for example, involving students and possible educational interventions, such as value-added assessment (e.g., Rubin *et al.*, 2004). Replication does not help without additional assumptions. The most straightforward assumption to make is the stable unit treatment value assumption (SUTVA; Rubin, 1980, 1990) under which the potential outcomes for the i th unit are determined by the treatment the i th unit received. That is, there is no interference between units (Cox, 1958) and there are no versions of treatments (Rubin, 1980). Then, all potential outcomes for N units with two possible treatments can be represented by an array with N rows and two columns, the i th unit having a row with two potential outcomes, $Y_i(0)$ and $Y_i(1)$, where each could, in principal, be a vector with many components. Obviously, SUTVA is a major assumption. Good researchers attempt to make such assumptions plausible by the design of their studies. For example, SUTVA becomes more plausible when units are isolated from each other, as when using, for the units, intact schools rather than students in the schools when studying an educational intervention, such as a smoking prevention program (e.g., see Peterson *et al.*, 2000).

In addition to (1) the vector indicator of treatments for each unit in the study, $W = \{W_i\}$, (2) the array of potential outcomes when exposed to the active treatment, $Y(1) = \{Y_i(1)\}$, and (3) the array of potential outcomes when not exposed, $Y(0) = \{Y_i(0)\}$, we have (4) an array of covariates $X = \{X_i\}$, which are, by definition, unaffected by treatment, such as age, race sex, or pretest scores, where the ‘pre’ means prior to the intervention, that is, before t_0 . Covariates can be used to help define causal estimands. All causal estimands involve comparisons of $Y_i(0)$ and $Y_i(1)$ on either all N units, or a common subset of units; for example, the average causal effect across all units that are female as indicated by their X_i , or the median causal effect for units with X_i indicating male and $Y_i(0)$ indicating failure on the posttest under the control treatment.

Under SUTVA, all causal estimands can be calculated from the matrix of scientific values with i th row: $(X_i, Y_i(1), Y_i(0))$. By definition, all relevant information is encoded in $X_i, Y_i(0), Y_i(1)$ and so the labeling of the N rows is a random permutation of $1, \dots, N$, and the matrix is row exchangeable. Covariates play a particularly important role in the analysis of observational studies for causal effects where they are also known as possible confounders or risk factors. In some studies, the units exposed to the active treatment differ in their distribution of covariates in important ways from the units not exposed. To see how this issue influences our formal framework, we must define the assignment mechanism, the probabilistic mechanism that determines which units receive the active version of the treatment and which units receive the control version.

The Assignment Mechanism – Motivating Examples

Even with SUTVA, inference for causal effects requires the specification of an assignment mechanism: a probabilistic model for how some units were selected to receive the active treatment and how other units were selected to receive the control treatment. We first illustrate this model in two trivial artificial examples, and then present formal notation for this model.

Consider a teacher who is considering one of two treatments to apply to each of eight students, a standard and a new one. This teacher is a great teacher and the treatment that is best for each student! When they are equally effective, a fair coin is tossed. **Table 1** gives both the hypothetical potential outcomes in test scores under each treatment for these students, and also their individual causal effects. The column labeled W shows which treatment each student received, $W_i = 0$ or $W_i = 1$ for the i th student.

Notice that the eight individual causal effects indicate that the typical student will do better with the standard

treatment (i.e., the control treatment): the average causal effect is two points in favor of the standard. However, the teacher, who is conducting ideal educational practice for the benefit of the students, reaches the opposite conclusion from an examination of the observed data: the students assigned the new treatment do, on average more than twice as well as the students assigned the control, with absolutely no overlap in their distributions!

What is wrong? The simple comparison of observed outcomes assumes that treatments were randomly assigned, rather than as they were, to provide maximal benefit to the students. More precisely, notice that the teacher, by comparing observed means of the outcome Y , is using the three observed values of $Y_i(1)$ to represent the five missing values of $Y_i(1)$, effectively imputing or filling in \bar{y}_1 , the observed mean of the $Y_i(1)$, for the five $Y_i(1)$ question marks, and analogously effectively filling in \bar{y}_0 for the three $Y_i(0)$ question marks. This process makes sense for point estimation if the three observed values of $Y_i(1)$ were randomly chosen from the eight values of $Y_i(1)$, and the five observed values of $Y_i(0)$ were randomly chosen from the eight values of $Y_i(0)$. However, under the actual assignment mechanism, it does not make sense. It would make much more sense under the actual assignment mechanism to impute the missing potential outcome for each student to be less than or equal to that student's observed potential outcome. The point here is simply that the assignment mechanism is crucial to valid inference about causal effects, and the teacher used a nonignorable assignment mechanism (defined shortly). With a posited assignment mechanism, it is possible to draw causal inferences; without one, it is impossible. It is in this sense that when drawing causal inferences, a model for the assignment mechanism is more fundamental than a model for the potential outcomes.

We next consider a classical paradox in educational research that is easily resolved with the simple ideas we have already presented, despite the controversy that the paradox engendered in some literatures. This example illustrates how important it is to keep this perspective clearly in mind when thinking about causal effects of interventions. Lord (1967) proposed the following example:

A large university is interested in investigating the effects on the students of the diet provided in the university dining halls and any sex differences in these effects. Various types of data are gathered. In particular, the weight of each student at the time of arrival in September and the following June are recorded.

The result of the study for the males is that their average weight is identical at the end of the school year to what it was at the beginning; in fact, the whole distribution of weights is unchanged, although some males lost weight and some males gained weight – the gains and losses exactly balance. The same thing is true for

Table 1 Perfect teacher

Potential outcomes		Observed data			
	$Y(0)$	$Y(1)$	W	$Y(0)$	$Y(1)$
	13	14	1	?	14
	6	0	0	6	?
	4	1	0	4	?
	5	2	0	5	?
	6	3	0	6	?
	6	1	0	6	?
	8	10	1	?	10
	8	9	1	?	0
True averages	7	5	Observed averages	5.4	11

the females. The only difference is that the females started and ended the year lighter on average than the males. On average, there is no weight gain or weight loss for either males or females. From Lord's description of the problem quoted above, the quantity to be estimated, the estimand, is the difference between the causal effect of the university diet on males and the causal effect of the university diet on females. That is, the causal estimand is the difference between the causal effects for males and females, the differential causal effect.

The paradox is generated by considering the contradictory conclusions of two statisticians. Statistician 1 observes that there are no differences between the September and June weight distributions for either males or females. Thus, statistician 1 concludes that

...as far as these data are concerned, there is no evidence of any interesting effect of diet (or of anything else) on student weight. In particular, there is no evidence of any differential effect on the two sexes, since neither group shows any systematic change (Lord, 1967: 305).

Statistician 2 looks at the data in a more sophisticated way. Effectively, he examines males and females with about the same initial weight in September, say a subgroup of overweight females (meaning simply above-average-weight females) and a subgroup of underweight males (analogously defined). He notices that these males tended to gain weight on average and these females tended to lose weight on average. He also notices that this result is true no matter what group of initial weights he focuses on. (Actually, Lord's statistician 2 used a covariance adjustment/regression adjustment.) Therefore the conclusion of statistician 2 is that, after controlling for initial weight, the diet has a differential positive effect on males relative to females because for males and females with the same initial weight, on average, the males gain more than the females.

Let us formulate the problem as advocated here. The units are the students, the time of application of treatment (the university diet) is September, and the time of the recording of the outcome Y is June; accept the stability assumption. The potential outcomes are June weight under the university diet $Y_i(1)$ and under the control diet $Y_i(0)$. The covariates are sex of students, male versus female, and September weight. However, the assignment mechanism has assigned everyone to the new treatment! There is no one, male or female, who is assigned to the control treatment. Hence, there is absolutely no purely empirical basis on which to estimate the causal effects, either raw or differential, of the university diet relative to the control diet. Lord has created partial confusion by making the problem complicated with the introduction of the covariates male/female and initial weight, and by using the observed variable notation, Y_{obs} , where $Y_{\text{obs},i} = W_i Y_i(1) + (1 - W_i) Y_i(0)$, which mixes up the science,

reflected here by $(Y_i(1), Y_i(0))$ and the assignment mechanism, reflected by W_i . For more statistical details of the resolution of this paradox, see Holland and Rubin (1983), and for earlier related discussion, see for example, Lindley and Novick (1981), or Cox and McCullagh (1982). However, the point here is that the paradox is immediately resolved through the explicit use of potential outcomes. In fact, either answer could be correct for causal inference depending on what we are willing to assume about the control diet, as elucidated by Holland and Rubin (1983).

The Assignment Mechanism – Formal Notation

A model for the assignment mechanism is needed for all forms of statistical inference for causal effects. Formally, the assignment mechanism gives the conditional probability of each vector of assignments given the covariates and potential outcomes:

$$\Pr(W|X, Y(0), Y(1)).$$

A specific example of an assignment mechanism is a completely randomized experiment with N units, where $n < N$ are assigned to the active treatment, and $N - n$ to the control treatment.

$$\Pr(W|X, Y(0), Y(1)) = 1/C_n^N \quad \text{if } \sum W_i = n \\ 0 \quad \text{otherwise}$$

An unconfounded assignment mechanism is free of dependence on either $Y(0)$ or $Y(1)$:

$$\Pr(W|X, Y(0), Y(1)) = \Pr(W|X).$$

With an unconfounded assignment mechanism, at each set of values of X_i that has a distinct probability of $W_i = 1$, there is effectively a completely randomized experiment. That is, if X_i indicates sex, with males having probability 0.2 of receiving the active treatment and females having probability 0.5 of receiving the active treatment, then essentially one randomized experiment is prescribed for males and another for females.

The assignment mechanism is probabilistic if each unit has a positive probability of receiving either treatment:

$$0 < \Pr(W_i = 1|X, Y(0), Y(1)) < 1,$$

where the unit level probabilities are known as propensity scores (Rosenbaum and Rubin, 1983). Unconfounded probabilistic assignment mechanisms often allow particularly straightforward estimation of causal effects from all perspectives, and these assignment mechanisms form the basis for inference for causal effects in more complicated situations, such as when assignment probabilities depend on covariates in unknown ways, or when there is noncompliance with the assigned treatment, or even

in observational (nonrandomized) studies. Unconfounded probabilistic assignment mechanisms are essentially generalized randomized experiments, and are called strongly ignorable (Rosenbaum and Rubin, 1983).

A confounded assignment mechanism is one that depends on the potential outcomes. A special class of possibly confounded assignment mechanisms is particularly important to Bayesian inference: ignorable assignment mechanisms (Rubin, 1978). Ignorable assignment mechanisms are defined by their freedom from dependence on any missing potential outcomes:

$$\Pr(W|X, Y(0), Y(1)) = \Pr(W|X, Y_{\text{obs}}).$$

Ignorable but confounded assignment mechanisms arise in practice, most commonly in sequential experiments, where the next (in time) unit's probability of being exposed to the active treatment depends on the success rate of those previously exposed to the active treatment versus the success rate of those exposed to the control treatment, as in play-the-winner designs (e.g., Efron, 1971). All unconfounded assignment mechanisms are ignorable, but not all ignorable assignment mechanisms are unconfounded (e.g., play-the-winner designs).

Modes of Causal Inference

There are two distinct forms of assignment-mechanism-based (or randomization-based) modes of causal inference: one due to Neyman (1923) and the other due to Fisher (1925). There is a third approach (Rubin, 1978), which is posterior predictive (Bayesian).

Fisher's approach is closely related to the mathematical idea of proof by contradiction. The first element in Fisher's mode is the null hypothesis, which is usually $Y_i(1) \equiv Y_i(0)$ for all units: the treatments have absolutely no effect on the potential outcomes. Under this null hypothesis, all potential outcomes are known from the observed values of the potential outcomes, Y_{obs} , because $Y(1) \equiv Y(0) \equiv Y_{\text{obs}}$. It follows that, under this null hypothesis, the value of any statistic, S , such as the difference of the observed averages for units exposed to treatment 1 and units exposed to treatment 0, $\bar{y}_1 - \bar{y}_0$, is known, not only for the observed assignment, but also for all possible assignments W . Suppose we calculate the value of S under each possible assignment (assuming the null hypothesis) and also calculate the probability of each assignment under the randomized assignment mechanism. Knowing the value of S for each W and its probability, we can then calculate the probability (under the assignment mechanism and the null hypothesis) that we would observe a value of S as unusual as, or more unusual than, the observed value of S , S_{obs} . Unusual is defined *a priori*, typically by how discrepant S_{obs} is from the typical values of S . This

probability is the plausibility (p -value or significance level) of the observed value of the statistic S under the null hypothesis: if the null hypothesis were true, the probability of S being as rare, or more rare, than S_{obs} .

Neyman's form of randomization-based inference can be viewed as drawing inferences by evaluating the expectations of statistics over the distribution induced by the assignment mechanism in order to calculate a confidence interval for the typical causal effect. First, an unbiased estimator of the causal estimand (the typical causal effect, e.g., the average) is created, and an unbiased, or upwardly biased, estimator of the sampling variance of that unbiased estimator is found (bias and sampling variance both defined with respect to the randomization distribution). Then, an appeal is made to the central limit theorem for the normality of the estimator over its randomization distribution, whence a confidence interval for the causal estimand is obtained.

With a data set that is not from a randomized study, we try to structure the problem so that we can conceptualize the data as having arisen from an underlying randomized experiment, and then estimate the assignment mechanism via the propensity scores for all the units. A key idea is that, like good experiments, good observational studies are designed, not simply found (Rubin, 2002, 2008). When designing an experiment, we do not have any outcome data, but we plan the collection, organization, and analysis of the data to improve our chances of obtaining valid, reliable, and precise causal answers. The same exercise should be done in an observational study: even if outcome data are available at the design stage, they should be set aside. As observational studies are rarely known to be unconfounded, we are concerned with the sensitivity of answers to unobserved covariates. The methods described by Rosenbaum (2002) are appropriate from the randomization-based perspective.

Posterior predictive (Bayesian) causal inference for causal effects requires a model for the underlying data, $\Pr(X, Y(0), Y(1))$, and this is where science enters. A virtue of the RCM framework is that it separates science – a model for the underlying data, from what we do to learn about science – the assignment mechanism, $\Pr(W|X, Y(0), Y(1))$. This approach directly and explicitly confronts the missing potential outcomes, $Y_{\text{mis}} = \{Y_{\text{mis},i}\}$, where $Y_{\text{mis},i} = W_i Y_i(0) + (1 - W_i) Y_i(1)$. The perspective takes the specification for the assignment mechanism and the specification for the underlying data, and derives the posterior predictive distribution of Y_{mis} , that is, the distribution of Y_{mis} given all observed values:

$$\Pr(Y_{\text{mis}}|X, Y_{\text{obs}}, W).$$

From this distribution and the observed values of the potential outcomes, the observed assignments, and observed covariates, the posterior distribution of any causal effect can,

in principle, be obtained. This conclusion is immediate if we view the above posterior predictive distribution as specifying how to take a random draw of Y_{mis} . Once a value of Y_{mis} is drawn, any causal effect can be directly calculated from the drawn value of Y_{mis} and the observed values of X and Y_{obs} . Repeatedly drawing values of Y_{mis} and calculating the causal effect for each draw generates the posterior distribution of the desired causal effect. Thus, as in Rubin (1978), we can view causal inference entirely as a missing data problem, where we multiply impute (Rubin, 1987, 2004a) the missing potential outcomes to generate a posterior distribution for the causal effects.

With ignorable treatment assignment, all that we need model is the science $\Pr(X, Y(0), Y(1))$, and we can ignore the assignment mechanism. A strength of this model-based approach is that it allows us to conduct causal inference by predicting the missing potential outcomes from observed values. The problem with this approach is the need to specify the distribution $\Pr(X, Y(0), Y(1))$, which sometimes can implicitly involve extrapolations that are extremely unreliable. More details of this approach are beyond the scope of this article, but can be found in, for example, Rubin (2006). With nonignorable treatment assignment, the simplifications previously described do not follow in general, and the analysis typically becomes far more difficult and speculative.

Complications

There are many complications that occur in real-world studies for causal effects, many of which can be handled much more flexibly with the Bayesian approach than with randomization-based methods. Of course, the models involved, including associated prior distributions, can be very demanding to formulate in a practically reliable manner. In addition, Neymanian evaluations are still important.

Most of the field of classical experiment design is devoted to issues that arise with more than two treatment conditions (e.g., Kempthorne, 1952; Cochran and Cox, 1957).

Missing data, due perhaps to unit dropout or machine failure, can complicate analyses more than one would expect based on a cursory examination of the problem. Methods such as multiple imputation (Rubin, 1987, 2004a), the expectation-maximization (EM) algorithm (Dempster *et al.*, 1977), data augmentation (Tanner and Wong, 1987), and the Gibbs sampler (Geman and Geman, 1984) are fully compatible with the Bayesian approach to causal inference outlined in the section titled 'Modes of causal inference'. Gelman *et al.* (2003) provide guidance on many of these issues from the Bayesian perspective.

Another complication, common when the units are people, is noncompliance (e.g., as in Sommer and Zeger, 1991).

Early work related to this issue can be found in economics (e.g., Tinbergen, 1930; Haavelmo, 1944) and elsewhere (e.g., Zelen, 1979; Bloom, 1984). Much progress has been made in recent years on this topic (e.g., Baker, 1998; Baker and Lindeman, 1994; Goetghebeur and Molenberghs, 1996; Angrist *et al.*, 1996; Imbens and Rubin, 1997; Little and Yau, 1998; Hirano *et al.*, 2000).

Further complications include truncation due to death (Rubin, 2000, 2004b; Zhang and Rubin, 2003; Zhang *et al.*, 2007). In the real world, complications typically do not appear simply one at a time. For example, a randomized experiment in education evaluating school choice suffered from missing data in both covariates and longitudinal outcomes; besides, the outcome was multicomponent at each point in time; in addition, it suffered from noncompliance that took several levels because of the years of school, as discussed, for example, Barnard *et al.* (2003) in the context of the school choice example.

Many of the above complications can be viewed as special cases of principal stratification (Frangakis and Rubin, 2002). This appears to be an extremely fertile area for research and application of Bayesian methods for causal inference, especially using modern simulation methods (see, e.g., Gilks *et al.*, 1995).

See also: Analysis of Covariance; Instrumental Variables; Methods for Approximating Random Assignment; Missing Data; Multivariate Linear Regression; Observational Studies; Quasi-Experimentation: Two Group Design.

Bibliography

- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**, 444–472.
- Baker, S. G. (1998). Analysis of survival data from a randomized trial with all-or-none compliance: Estimating the cost-effectiveness of a cancer screening program. *Journal of the American Statistical Association* **93**, 929–934.
- Baker, S. G. and Lindeman, K. S. (1994). The paired availability design: A proposal for evaluating epidural analgesia during labor. *Statistics in Medicine* **13**, 2269–2278.
- Barnard, J., Hill, J., Frangakis, C., and Rubin, D. (2003). School choice in NY city: A Bayesian analysis of an imperfect randomized experiment. In Gatsonis, C., Carlin, B., and Carriquiry, A. (eds.) *Case Studies in Bayesian Statistics*, vol. V, pp 3–97. New York: Springer.
- Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review* **8**, 225–246.
- Cochran, W. G. and Cox, G. M. (1957). *Experimental Designs*, 2nd edn. (reprinted as a "Wiley Classic" (1992)) New York: Wiley.
- Cox, D. R. (1958). *The Planning of Experiments*. New York: Wiley.
- Cox, D. R. and McCullagh, P. (1982). Some aspects of covariance. *Biometrics* **38**, 541–561.
- Dempster, A. P., Laird, N., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* **39**, 1–38.
- Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58**, 403–417.
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*, 1st edn. Edinburgh: Oliver and Boyd.
- Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.

- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2003). *Bayesian Data Analysis*, 2nd edn. New York: CRC Press.
- Geman, S. and Geman, D. (1984). Stochastic relaxation. *Gibbs Distributions, and the Bayesian Restoration of Images, IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1995). *Markov Chain Monte Carlo in Practice*. New York: CRC Press.
- Goetghebuer, E. and Molenberghs, G. (1996). Causal inference in a placebo-controlled clinical trial with binary outcome and ordered compliance. *Journal of the American Statistical Association* **91**, 928–934.
- Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica* **15**, 413–419.
- Hirano, K., Imbens, G., Rubin, D. B., and Zhou, X. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* **1**, 69–88.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association* **81**, 945–970.
- Holland, P. W. and Rubin, D. B. (1983). On Lord's paradox. In Wainer, S. and Messick, H. (eds.) *Principals of Modern Psychological Measurement: A Festschrift for Frederic M. Lord*, pp 3–25. Hillsdale, NJ: Earlbaum.
- Imbens, G. W. and Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics* **25**, 305–327.
- Kemphorne, O. (1952). *The Design and Analysis of Experiments*. New York: Wiley.
- Lindley, D. V. and Novick, M. R. (1981). The role of exchangeability in inference. *Annals of Statistics* **9**, 45–58.
- Little, R. J. and Yau, L. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. *Psychological Methods* **3**, 147–159.
- Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin* **68**, 304–305.
- Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. Translated in *Statistical Science* (1990), 465–472.
- Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge, UK: Cambridge University Press.
- Peterson, A. V., Jr., Kealey, K. A., Mann, S. L., Marek, P. M., and Sarason, I. G. (2000). Hutchinson smoking prevention project: Long-term randomized trial in school-based tobacco use prevention-results on smoking. *Journal of the National Cancer Institute* **92**, 1979–1991.
- Rosenbaum, P. R. (2002). *Observational Studies*, 2nd edn. New York: Springer.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational Studies for causal effects. *Biometrika* **70**, 41–55.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688–701.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
- Rubin, D. B. (1977). Assignment of treatment group on the basis of a covariate. *Journal of Educational Statistics* **2**, 1–26.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* **7**, 34–58.
- Rubin, D. B. (1980). Comment on "randomization analysis of experimental data: The Fisher randomization test" by D. Basu. *Journal of the American Statistical Association* **75**, 591–593.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Rubin, D. B. (1990). Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science* **5**, 472–480.
- Rubin, D. B. (2000). The utility of counterfactuals for causal inference. Comment on A. P. Dawid, "Causal inference without counterfactuals." *Journal of the American Statistical Association* **95**, 435–438.
- Rubin, D. B. (2002). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology* **2**, 169–188.
- Rubin, D. B. (2004a). *Multiple Imputation for Nonresponse in Surveys*. (reprinted with new appendices as a "Wiley Classic.") New York: Wiley.
- Rubin, D. B. (2004b). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics* **31**, 161–170; 195–198.
- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. 2004 Fisher lecture. *Journal of the American Statistical Association* **100**, 322–331.
- Rubin, D. B. (2006). Statistical inference for causal effects, with emphasis on applications in psychometrics and education. In Rao, C. R. and Sinharay, S. (eds.) *Handbook of Statistics, Volume 26: Psychometrics*, pp 769–800. Amsterdam: Elsevier.
- Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *Annals of Applied Statistics* **2**(3), 808–840.
- Rubin, D. B., Stuart, E. A., and Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics* **29**, 103–116.
- Sommer, A. and Zeger, S. (1991). On estimating efficacy from clinical trials. *Statistics in Medicine* **10**, 45–52.
- Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**, 528–550.
- Tinbergen, J. (1930). Determination and interpretation of supply curves: An example. (reprinted in Henry and Morgan (eds.) *The Foundations of Economics*) *Zeitschrift für Nationalökonomie*.
- Zelen, M. (1979). A new design for randomized clinical trials. *New England Journal of Medicine* **300**, 1242–1245.
- Zhang, J. and Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by 'death' *Journal of Educational and Behavioral Statistics* **28**, 353–368.
- Zhang, J., Rubin, D., and Mealli, F. (2007). Evaluating the effects of job training programs on wages through principal stratification. *Advances in Economics* **21**, 93–118.

Cluster Analysis: Overview

S Landau, King's College London, London, UK

I Chis Ster, University College London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

The old idea of sorting similar things into groups underlies all aspects of human activity. For example, languages can be thought of as classification systems where words are used to describe types of events, objects, and people encountered. In addition to being a basic human conceptual activity, classification is fundamental to many branches of science with prominent examples being animal or plant taxonomies informing evolutionary theories in biology, or groupings of elements according to their chemical properties influencing research on the structure of the atom. In the social sciences, including education, it would typically be people that are to be grouped to identify patterns of behavior, achievement, etc.

Numerical methods aimed at discovering groups in data are referred to as cluster analysis. The groups can be sets of objects (individuals, countries, animals, chemical elements, etc.) or sets of variables. It also needs to be emphasized that cluster analysis is aimed at uncovering as-yet-unknown groups of objects; with analogous concepts being unsupervised pattern recognition or numerical taxonomy. In contrast, discriminant analysis or supervised pattern recognition aims to establish rules that classify objects into classes that are known *a priori* based on a set of observable characteristics. Finally, cluster analysis is an exploratory technique. Its primary aim is not to infer anything about population parameters as most statistical methods do – but rather to suggest groupings that might form the basis of future hypotheses to be investigated.

Cluster analysis techniques themselves can be broadly grouped into three classes labeled hierarchical clustering, optimization clustering, and model-based clustering. They operate either directly on a matrix of scores on a number of variables for a set of objects to be classified, or on a matrix of distances or similarities between the objects.

Proximity Measures

Ideally, clusters should be internally cohesive structures that are isolated from each other. To judge this adequacy criteria are needed that encapsulate the concepts of cluster homogeneity (cohesion) and separation (isolation). Such measures can be derived from a matrix of object distances or (dis)similarities; more generally referred to as proximities.

Let \mathbf{X} denote the usual $n \times p$ multivariate data matrix containing the data values describing each object to be clustered,

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & \cdots & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & \cdots & \cdots & x_{np} \end{bmatrix}$$

That is, the entry x_{ij} gives the value for the j th variable of the i th object. This is a two-mode matrix indicating that the rows and columns correspond to different things. For a raw data matrix containing only categorical variables, similarity measures are typically used. The similarity coefficient s_{ik} between objects i and k is one if both objects have identical values for all variables and zero if the two objects differ maximally for all variables. (The corresponding dissimilarity is then simply $d_{ik} = 1 - s_{ik}$.) In the binary case, similarity measures arise from a cross-classification of the counts of matches and mismatches of the p variables. **Table 1** shows some possible similarity measures. A multitude of measures has been proposed largely due to the uncertainty as to how to deal with the count of zero-zero matches (for a comprehensive list see Gower and Legendre (1986)). For categorical variables with more than two levels, similarities are typically calculated by allocating a score $s_{ik,j}$ of zero or one to each variable j , depending on whether the two objects i and k are the same on that variable and then simply averaging to give $s_{ik} = (1/p) \sum_j s_{ik,j}$. For a raw data matrix containing only continuous variables, proximities are typically expressed as distances (dissimilarity measures that fulfill the metric inequality); see **Table 2** for some possible measures. The Euclidean distance and the city block distance are appealing choices because of their geometric interpretation as physical distances between p -dimensional points with the latter traveled in rectilinear configuration.

As an example, consider applying the Euclidean distance measure to the protein consumption data in **Table 3**. The consumption for 25 European countries is measured in grams of protein food per day. Clearly, if we were to calculate distances on the raw data, these would be dominated by the variables relating to products generally eaten in larger quantities, such as cereals. We therefore standardize the variables to unit variance before applying the distance measure, a technique sometimes referred to as autoscaling. **Table 4** shows the resulting distance matrix (a one-mode matrix).

Finally, there are a number of approaches for constructing proximities for mixed-mode data – data in

which some variables are continuous and others are categorical. A simple approach is to construct an appropriate proximity matrix for each variable type submatrix and then to combine the measures.

Returning to our objective of wanting to measure the adequacy of a clustering on the basis of proximities, there are two basic approaches to defining intergroup proximities: (1) define the proximity by a suitable summary of the proximities between individuals from either group, or (2) represent each group by a typical observation and measure the proximity between these centers. Suitable summaries for (1) are the nearest neighbor distance, the furthest neighbor distance, or the average distance. Under approach (2), groups might be represented by, for example, the mean over the objects for each variable (continuous variables only), the so-called centroid, or the object that has the smallest average dissimilarity to all other group members (the medoid).

Hierarchical Clustering

Members of this class of clustering techniques produce a nested sequence of partitions by merging (or dividing)

Table 1 Similarity measures for binary data

Counts of binary outcomes for two individuals Individual <i>i</i>				
Individual <i>j</i>	Outcome	1	0	Total
	1	<i>a</i>	<i>b</i>	<i>a + b</i>
	0	<i>c</i>	<i>d</i>	<i>c + d</i>
	Total	<i>a + c</i>	<i>b + d</i>	<i>p = a + b + c + d</i>
Similarity measures				
Measure	Formula			
Matching coefficient	$s_{ij} = (a + d)/(a + b + c + d)$			
Jaccard coefficient	$s_{ij} = a/(a + b + c)$			
Rogers and Tanimoto	$s_{ij} = (a + d)/[a + 2(b + c) + d]$			
Sokal and Sneath	$s_{ij} = a/[a + 2(b + c)]$			
Gower and Legendre I	$s_{ij} = (a + d)/[a + \frac{1}{2}(b + c) + d]$			
Gower and Legendre II	$s_{ij} = a/[a + \frac{1}{2}(b + c)]$			

Table 2 Distance measures for continuous data

Measure	Formula
Euclidean distance	$d_{ij} = \left(\sum_{k=1}^p (x_{ik} - x_{jk})^2 \right)^{1/2}$
City block distance	$d_{ij} = \sum_{k=1}^p x_{ik} - x_{jk} $
Minkowski distance	$d_{ij} = \left(\sum_{k=1}^p (x_{ik} - x_{jk})^r \right)^{1/r} \quad (r \geq 1)$
Canberra distance	$d_{ij} = \begin{cases} 0 & \text{for } x_{ik} = x_{jk} = 0 \\ \sum_{k=1}^p x_{ik} - x_{jk} / (x_{ik} + x_{jk}) & \text{for } x_{ik} \neq 0 \text{ or } x_{jk} \neq 0 \end{cases}$

clusters. At each stage of the sequence, the new partition is optimally merged (or divided) from the previous partition according to some adequacy criterion. In the end, the sequence of partitions ranges from a single cluster containing all the individuals to *n* clusters containing a single individual. The whole series of partitions is most conveniently described by a tree display called the dendrogram (see later examples). Agglomerative hierarchical clustering proceeds by a series of successive fusions of the *n* objects into groups. In contrast, divisive hierarchical methods divide the *n* individuals successively into finer groups. Divisive methods are not commonly used due to computational problems; for more details see Everitt *et al.* (2001).

A variety of agglomerative techniques exist reflecting the different ways in which intergroup dissimilarities can be defined. For example, merging two clusters when their nearest neighbor distance is minimal leads to an agglomerative procedure called single linkage. Similarly, the use of the furthest neighbor distance or the average distance leads to complete and average linkage, respectively. To describe the agglomerative process, complete linkage will be applied to the first five countries in **Table 3** using the Euclidean distances shown as the shaded area in **Table 4**. Initially, there are five clusters all containing a single country (partition 1). The first step is to combine the closest pair of countries. From **Table 4** this is seen to be Austria and Czechoslovakia (partition 2). The distances between Albania, Bulgaria, and Belgium and the cluster (Austria and Czechoslovakia) need to be evaluated next. Based on the furthest neighbor distance this is found as: $d_{\text{new country, cluster}} = \max(d_{\text{new country, Austria}}, d_{\text{new country, Czechoslovakia}})$. After these new distances are calculated, the smallest value is again used to decide which clusters should be merged; here Belgium and (Austria, Czechoslovakia) are fused (partition 3). The next stage involves evaluating the furthest neighbor distances between clusters [(Austria, Czechoslovakia), Belgium], Albania and Bulgaria. This leads to Albania and Bulgaria being combined (partition 4) and then in the final step [(Austria, Czechoslovakia), Belgium] and (Albania,

Table 3 Protein consumption in 25 European countries for nine food groups (grams per day)

Country	Red meat	White meat	Eggs	Milk	Fish	Cereals	Starch	Nuts	Fruits and vegetables
Albania	10.1	1.4	0.5	8.9	0.2	42.3	0.6	5.5	1.7
Austria	8.9	14	4.3	19.9	2.1	28	3.6	1.3	4.3
Belgium	13.5	9.3	4.1	17.5	4.5	26.6	5.7	2.1	4
Bulgaria	7.8	6	1.6	8.3	1.2	56.7	1.1	3.7	4.2
Czechoslovakia	9.7	11.4	2.8	12.5	2	34.3	5	1.1	4
Denmark	10.6	10.8	3.7	25	9.9	21.9	4.8	0.7	2.4
E Germany	8.4	11.6	3.7	11.1	5.4	24.6	6.5	0.8	3.6
Finland	9.5	4.9	2.7	33.7	5.8	26.3	5.1	1	1.4
France	18	9.9	3.3	19.5	5.7	28.1	4.8	2.4	6.5
Greece	10.2	3	2.8	17.6	5.9	41.7	2.2	7.8	6.5
Hungary	5.3	12.4	2.9	9.7	0.3	40.1	4	5.4	4.2
Ireland	13.9	10	4.7	25.8	2.2	24	6.2	1.6	2.9
Italy	9	5.1	2.9	13.7	3.4	36.8	2.1	4.3	6.7
Netherlands	9.5	13.6	3.6	23.4	2.5	22.4	4.2	1.8	3.7
Norway	9.4	4.7	2.7	23.3	9.7	23	4.6	1.6	2.7
Poland	6.9	10.2	2.7	19.3	3	36.1	5.9	2	6.6
Portugal	6.2	3.7	1.1	4.9	14.2	27	5.9	4.7	7.9
Romania	6.2	6.3	1.5	11.1	1	49.6	3.1	5.3	2.8
Spain	7.1	3.4	3.1	8.6	7	29.2	5.7	5.9	7.2
Sweden	9.9	7.8	3.5	24.7	7.5	19.5	3.7	1.4	2
Switzerland	13.1	10.1	3.1	23.8	2.3	25.6	2.8	2.4	4.9
UK	17.4	5.7	4.7	20.6	4.3	24.3	4.7	3.4	3.3
USSR	9.3	4.6	2.1	16.6	3	43.6	6.4	3.4	2.9
W Germany	11.4	12.5	4.1	18.8	3.4	18.6	5.2	1.5	3.8
Yugoslavia	4.4	5	1.2	9.5	0.6	55.9	3	5.7	3.2

From Hand, D. J., Daly, F., McConway, K., Lunn, D., and Ostrowsky, E. (1994). *A Handbook of Small Data Sets*. London: Chapman and Hall.

Bulgaria) are fused to become a single cluster (partition 5). The fusion process is described graphically by the dendrogram in **Figure 1** with the y -axis showing the distance at which clusters are fused. (Note that the display is not unique; e.g., (Albania, Bulgaria) or [Belgium, (Austria, Czechoslovakia)] can be flipped.) A popular agglomerative method for continuous data which measures cluster adequacy by evaluating distances between cluster centroids is Wards method. The approach merges clusters if the sum of the squared dissimilarities to the cluster centroid (a measure of cohesion) is minimal across all possible merging choices. **Figure 2** shows results of applying single linkage (a), complete linkage (b) and Ward's method (c) to the full distance matrix from **Table 4**. Different distances produce different cluster solutions. A similarity may be observed between complete linkage and Ward's method, but the single linkage solution looks rather different.

Users need to be aware that partitions obtained by hierarchical clustering are irrevocable in the sense that once clusters have been combined in an agglomerative procedure they cannot be split up again. This means that not all possible partitions are evaluated which, while computationally convenient, may mean that an optimal partition is missed. Second, the method may not take account of the cluster structure properly. For example, single linkage is known to be prone to chaining, the

tendency for new points to join the previous cluster in a chain-like fashion. This is what appears to have happened in **Figure 2(a)**. In contrast, complete linkage is known to produce compact clusters which may not always reflect the true data structure either. Efforts have been made to define properties that would be useful for hierarchical cluster methods. For example, point proportionality refers to replication of points not altering the boundaries of partitions. The monotone property, which states that the numerical values of the proximities are unimportant in that only their ranking matters for the clustering, might also be desirable. Single linkage and complete linkage, which otherwise are problematic, possess these properties; see Everitt *et al.* (2001).

Optimization Clustering

In contrast to hierarchical clustering methods optimization clustering methods aim to evaluate the adequacy of all possible partitions of a set of objects into k clusters. For the moment, we consider the number of clusters k to be known. The basic idea behind these methods is that associated with each partition of the n objects into the required number of groups k is an index $c(n, k)$, the value of which is to be optimized. Differences between the clustering methods in this class arise because of the variety of

Table 4 Euclidean distances for protein data after standardisation of variables to unit variance

	<i>Alb</i>	<i>Aus</i>	<i>Belg</i>	<i>Bulg</i>	<i>Czech</i>	<i>Denm</i>	<i>Germ</i>	<i>Finlan</i>	<i>France</i>	<i>Greece</i>	<i>Hungar</i>	<i>Ireland</i>	<i>Italy</i>	<i>Nether</i>	<i>Norwa</i>	<i>Poland</i>	<i>Portug</i>	<i>Roman</i>	<i>Spain</i>	<i>Swed</i>	<i>Switzer</i>	<i>UK</i>	<i>USSR</i>	<i>W Ger</i>	<i>Yugosl</i>
<i>Alb</i>	0																								
<i>Aus</i>	6.14	0																							
<i>Belg</i>	5.94	2.46	0																						
<i>Bulg</i>	2.76	4.9	5.24	0																					
<i>Czech</i>	5.14	2.14	2.22	3.96	0																				
<i>Denm</i>	6.64	3.02	2.54	6.04	3.36	0																			
<i>Germ</i>	6.4	2.58	2.12	5.4	1.88	2.76	0																		
<i>Finlan</i>	5.88	4.08	3.5	5.82	3.98	2.64	4.08	0																	
<i>France</i>	6.3	3.58	2.2	5.56	3.36	3.66	3.8	4.6	0																
<i>Greece</i>	4.26	5.16	4.7	3.76	4.88	5.6	5.62	5.5	4.54	0															
<i>Hungar</i>	4.68	3.28	4	3.34	2.76	5.04	3.68	5.4	4.98	4.12	0														
<i>Ireland</i>	6.76	2.74	1.66	6.22	3.16	2.82	3.04	3.24	3.16	5.7	4.82	0													
<i>Italy</i>	4.02	3.72	3.72	2.86	3.34	4.78	4.32	4.92	3.8	2.16	3.16	4.84	0												
<i>Nether</i>	6	1.12	2.24	5.16	2.2	2.54	2.54	3.38	3.4	5.16	3.5	2.34	3.92	0											
<i>Norwa</i>	5.46	3.88	2.96	5.28	3.52	2	3.28	2.06	3.92	4.62	4.9	3.6	4	3.36	0										
<i>Poland</i>	5.88	2.8	2.94	4.44	2.1	3.84	2.7	4.12	3.6	4.42	3.04	3.74	3.12	2.78	3.7	0									
<i>Portug</i>	6.62	6.52	5.66	6	5.52	5.86	5.26	6.5	5.66	4.78	5.7	7.06	4.66	6.36	4.8	4.82	0								
<i>Roman</i>	2.68	4.66	4.76	1.88	3.56	5.54	4.78	5.06	5.52	3.62	2.48	5.6	3.1	4.64	4.68	3.96	5.62	0							
<i>Spain</i>	5.56	4.88	4	4.84	4.14	5.12	4.08	5.48	4.46	3.1	3.88	5.28	2.88	4.86	4.16	3.4	2.94	4.24	0						
<i>Swed</i>	5.66	2.94	2.58	5.4	3.26	1.38	3.06	2.06	3.82	4.98	4.66	2.86	4.14	2.4	1.5	3.84	5.86	4.86	4.8	0					
<i>Switzer</i>	5.12	2.2	2.34	4.48	2.62	3.18	3.58	3.54	2.42	4.1	3.86	2.82	2.94	1.9	3.34	3.08	6.12	4.36	4.58	2.68	0				
<i>UK</i>	5.94	3.74	1.94	5.8	3.84	3.48	3.92	3.88	2.58	4.62	5.12	2.26	4.18	3.52	3.54	4.5	6.54	5.42	4.72	3.14	2.84	0			
<i>USSR</i>	4.34	4.16	3.16	3.84	2.72	4.16	3.42	3.46	4.24	4.12	3.42	3.9	3.56	3.88	3.26	2.92	5.08	2.76	3.62	3.78	3.8	4	0		
<i>W Ger</i>	6.36	1.64	1.42	5.62	2.18	2.4	1.9	3.66	2.94	5.36	3.9	1.8	4.14	1.28	3.3	3	6.14	5.1	4.6	2.46	2.28	2.9	3.9	0	
<i>Yugosl</i>	2.94	5.44	5.6	2	4.34	6.36	5.52	5.8	6.3	3.94	3.04	6.46	3.58	5.5	5.4	4.5	5.82	0.98	4.56	5.7	5.2	6.26	3.36	5.96	0

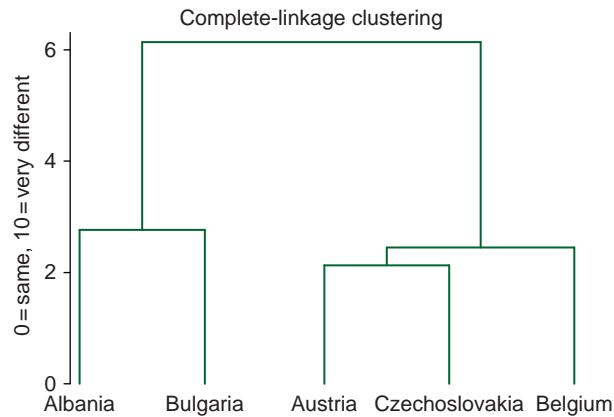


Figure 1 Dendrogram demonstrating complete linkage clustering of first five countries.

criteria that might be optimized. In addition, optimization clustering is a computing-intensive process since the number of necessary evaluations soon becomes very large as the number of clusters and/or the sample size increases. Thus, even with today's computers complete enumeration is not possible and consequently different algorithms have been suggested to optimize the same cluster criterion.

Cluster criteria can be defined on the basis of the proximity matrix or on the basis of the raw data matrix. When using the proximity matrix the criteria tend to define the lack of homogeneity for a single group (e.g., by the maximum dissimilarity between two objects from the group) and then the lack of homogeneity of a partition by suitably aggregating over the groups (e.g., by using a weighted sum over the groups). A number of algorithms for optimizing such cluster criteria were proposed by Kaufman and Rousseeuw (1990), with perhaps the most popular one being the partitioning around medoids (PAM) algorithm which minimizes criterion $c_{\text{PAM}}(n, k) = \sum_{m=1}^k \sum_{i=1}^{n_m} d(\mathbf{x}_{im}, \text{medoid}_m)$, the sum of the dissimilarities between the objects with p -dimensional data vectors \mathbf{x}_{im} , $m \in \{1, \dots, k\}$, $i \in \{1, \dots, n_m\}$ and their cluster medoids. As an example, consider 1993 data from 705 American colleges and universities on student intake, cost, and learning environment published by US News and World Report magazine in 1995. The data set is available at the math forum data collection website. Here, we have selected nine typical variables as shown in Table 5 and use 705 colleges with complete information on these variables. The college sample is presented graphically in Figure 3. This raw data matrix was converted into a proximity matrix by calculating Euclidean distances. To overcome the unit of measurement problem, all variables were transformed so that they appeared to have symmetrical distributions (using ln and logit transformations) and then standardized to have sample mean 0 and standard deviation 1, before calculating distances. Application of the PAM algorithm provided the three clusters indicated in Figure 3. As PAM minimizes distances to the medoids, the latter represent

the best cluster summary. Table 6 shows that PAM cluster 1 consists of universities which enrol moderate numbers of students with a high proportion from the top 10% achievers, have the highest in-state tuition, smallest student-to-faculty ratio, and the highest proportion of graduates. Cluster 2 groups less selective universities that enrol the largest numbers of students, charge the least tuition fees, have a higher student-to-faculty ratio, and produce the smallest proportion of graduates. Cluster 3 contains less selective colleges with small enrolment numbers that charge moderate fees and have an average proportion of graduates.

Most cluster criteria derived from continuous data make use of a decomposition of the total dispersion matrix $\mathbf{T} = \sum_{m=1}^k \sum_{i=1}^{n_m} (\mathbf{x}_{im} - \bar{\mathbf{x}})(\mathbf{x}_{im} - \bar{\mathbf{x}})^T$, where $\bar{\mathbf{x}}$ is the p -dimensional vector of overall sample means for each variable, into

$$\mathbf{T} = \mathbf{W} + \mathbf{B} \quad [1]$$

where $\mathbf{W} = \sum_{m=1}^k \sum_{i=1}^{n_m} (\mathbf{x}_{im} - \text{centroid}_m)(\mathbf{x}_{im} - \text{centroid}_m)^T$ is the within-group dispersion matrix and $\mathbf{B} = \sum_{m=1}^k n_m (\text{centroid}_m - \bar{\mathbf{x}})(\text{centroid}_m - \bar{\mathbf{x}})^T$ is the between-group dispersion matrix. Most popular in this class of cluster criteria is minimization of $c_{\text{trace}}(n, k) = \text{trace}(\mathbf{W})$. This is equivalent to minimizing the sum of the squared Euclidean distances between the objects and their centroids. Popularity is due to most software packages containing a k-means algorithm, which minimizes this criterion by iteratively reallocating an object to another group if the object is nearer (in terms of Euclidean distance to the centroid) to the new group than its own. K-means clustering of the colleges into three groups using the transformed data produced the solution in Table 7, presenting similar substantive results to PAM. Similar to PAM, k -means clustering minimized distances in some sense and is not scale invariant. A scale-invariant alternative cluster criterion is minimization of $c_{\text{det}}(n, k) = \det(\mathbf{W})$. This cluster approach brings the practical benefit of not requiring the user to address the unit of measurement problem by standardization of variables or similar techniques. Both $\text{trace}(\mathbf{W})$ and $\det(\mathbf{W})$ clustering have a tendency to generate clusters of roughly the same size (in terms of number of objects) and of similar shape and volume. Further criteria have been suggested to overcome this; see Everitt *et al.* (2001). Finally, a note of caution: as all the so-called hill-climbing algorithms aim to find a global optimum without having to evaluate all possible partitions, it is important to check convergence against a global optimum, for example, by rerunning the algorithm with a new set of starting values.

Model-Based Clustering

Most cluster analysis methods are essentially heuristic methods in the sense that they do not make explicit

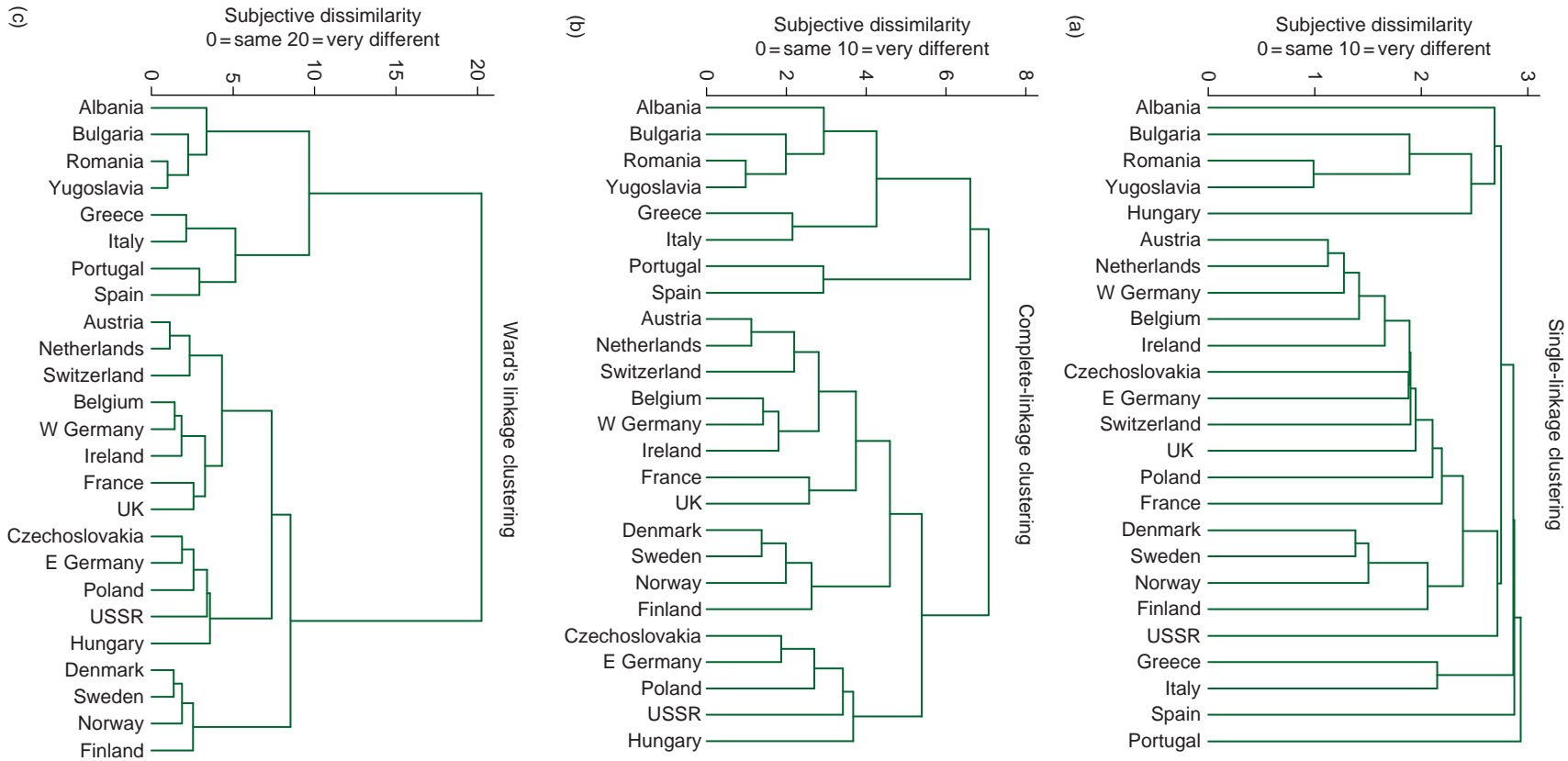


Figure 2 Dendrograms showing partitions resulting from (a) single linkage, (b) complete linkage, and (c) Ward's method applied to protein consumption proximity matrix in **Table 4**.

Table 5 American colleges' data (first ten records only)

<i>College name</i>	<i>Rejection proportion^a</i>	<i>Number enrolled^b</i>	<i>Top 10%^c</i>	<i>In-state tuition^d</i>	<i>Additional fees^e</i>	<i>Personal costs^f</i>	<i>Number PhD faculty^g</i>	<i>Ratio faculty students^h</i>	<i>Percentage graduatedⁱ</i>
Alaska Pacific University	0.24	55	16	7560	130	1500	76	11.9	15
Chicago State University	0.52	777	12	1848	350	2400	47	15.6	18
Brewton-Parker College	0.14	1202	10	4371	130	2000	62	12.6	18
College of the Southwest	0.21	27	7	3120	125	500	24	14.3	20
Northeastern Illinois University	0.28	631	14	1902	236	2178	78	15.1	21
Mount Saint Clare College	0.13	95	16	9900	80	1200	32	13.6	21
Claflin College	0.42	499	21	4412	600	1000	69	16.9	21
Huron University	0.67	124	3	7260	330	1840	31	12.9	21
University of Colorado at Denver	0.53	261	30	1828	240	2138	89	18.1	24
Fayetteville State University	0.27	452	1	740	636	766	75	15.1	24

^aProportion of rejected applications.^bNumber of students eventually enrolled.^cPercentage of new students from top 10% of high school class.^dState-specific tuition in \$ per academic year.^eAdditional costs: books and other materials for study in \$ per academic year.^fLiving and leisure costs in \$ per academic year.^gNumber of PhD staff.^hNo students/no teachers ratio.ⁱPercentage of the students who graduated.

assumptions about the data-generating process. It is therefore impossible to infer from sample to population. Perhaps this presents no real difficulties to investigators involved in an initial exploration of their data where cluster analysis is only used to suggest hypothesis for future investigation. However, attempts have been made to develop a more acceptable statistical approach to the clustering problem, using what are known as finite mixture distributions (McLachlan and Peel, 2000).

Briefly, finite mixture densities are a family of probability density functions of the form

$$f(\mathbf{x}, \mathbf{p}, \boldsymbol{\theta}) = \sum_{m=1}^k p_m g_m(\mathbf{x}, \boldsymbol{\theta}_m) \quad [2]$$

where \mathbf{x} is a p -dimensional random variable and vectors $\mathbf{p}^T = [p_1, p_2, \dots, p_{k-1}]$ and $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_k]$ are parameter vectors. The $p_m \geq 0, m = 1, \dots, k; \sum_{m=1}^k p_m = 1$ are known as mixing proportions and the $g_m(\mathbf{x}, \boldsymbol{\theta}_m), m = 1, \dots, k$ are the component densities being parameterized by $\boldsymbol{\theta}_m$. The number of components forming the mixture is k .

Then, finite mixtures provide statistical models for cluster analysis if we assume that the objects within a cluster arise from one of k subpopulations with different multivariate distributions $g_m(\mathbf{x}, \boldsymbol{\theta}_m)$. The latter distributions may belong to the same family, but differ in the values they have for the parameters of the distributions or come from different families (e.g., Everitt and Bullmore, 1999). The mixing proportions and the parameters of the component densities can be estimated by maximum likelihood. This more formal statistical approach brings the advantage that one can develop cluster criteria whose optimization corresponds to maximizing the log-likelihood under a specified statistical model. This enables specification of the model assumptions under which a cluster criterion is expected to perform well.

Finally, having specified a suitable statistical model and estimated its parameters, the so-called model-based cluster analysis is typically performed by associating an object with a particular subpopulation (cluster) on the basis that this subpopulation maximizes the value of the estimated posterior probability:

$$\Pr(\text{object } i \text{ belongs to cluster } m \mid \mathbf{x}_i) = \frac{\hat{f}_m g_m(\mathbf{x}_i, \hat{\boldsymbol{\theta}}_m)}{f(\mathbf{x}_i, \hat{\mathbf{p}}, \hat{\boldsymbol{\theta}})} \quad [3]$$

over all possible subpopulations $m = 1, \dots, k$.

A common finite mixture model is based on multivariate normal densities with different mean vectors and possibly different covariance matrices. It can be shown that for unconstrained component covariance matrices $\boldsymbol{\Sigma}_m$, $m = 1, \dots, k$ maximization of the finite mixture likelihood is the same as minimizing criterion $c_{\text{unconstrained}}(n, k) = \prod_{m=1}^k [\det(\mathbf{W}_m/n_m^2)]^{n_m}$, where \mathbf{W}_m denotes the (sample) dispersion matrix for the n_m -dimensional m -th subpopulation. If the covariances can be assumed to be the same

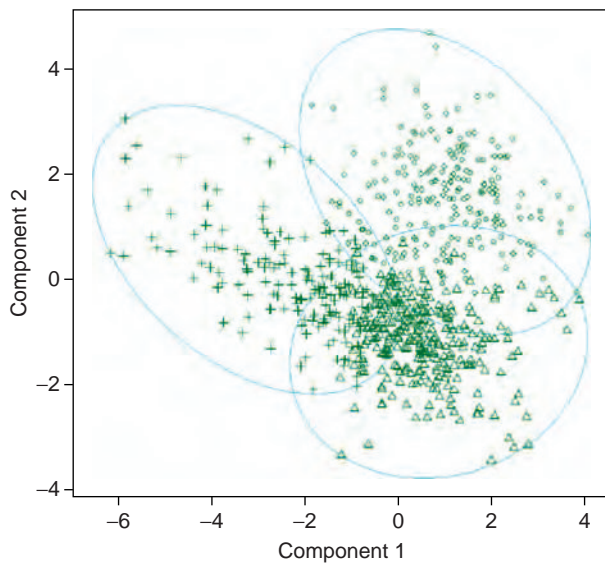


Figure 3 Scatter plot of colleges' data in the space of the first two principal components (symbols and ellipses indicate the three-group PAM solution).

across subpopulations (though remain unknown), that is, $\boldsymbol{\Sigma}_m = \boldsymbol{\Sigma}$ for all m , the corresponding cluster criterion becomes $c_{\text{same cov}}(n, k) = \prod_{m=1}^k [\det(\mathbf{W}/n_m^2)]^{n_m}$. To explore the number of clusters present in the American colleges' data, **Figure 4** shows univariate distributions of the transformed data with overlaid densities. From this it appears that the number of visible subpopulations is 3 (see the three-peak distribution in **Figure 4(d)**). We therefore fitted a three-component multivariate normal mixture with constant (but unknown) component covariance matrices and used posterior probabilities to allocate cluster memberships. This leads to a cluster solution with similar interpretation to those obtained previously.

Finite mixtures with multivariate normal distributions have been widely used because of their computational convenience. However, alternatives have been suggested; for example, multivariate t-distributions for groups of observations with longer tails than normal or atypical observations (McLachlan and Peel, 2000). For binary data multivariate Bernoulli densities which arise by assuming, that within each cluster, the binary variables are independent from each other (the so-called conditional independence assumption), are typically used. The latter multivariate component densities define the classical latent class model (Lazarsfeld and Henry, 1968) and have also been referred to as discrete mixture distributions. An interesting example is reported by Aitkin *et al.* (1981) who fitted latent class models to observations on 38 binary variables describing teaching behavior observations made on 468 teachers. Teachers were allocated to two classes by maximizing their posterior probabilities. **Table 8** summarizes the results by displaying response probabilities for each of the 38 items and each of the two classes. An obvious interpretation is a split into formal and informal teaching styles.

Table 6 Description of PAM three-cluster solution by medoids

Cluster (size)	Student intake			Affordability			Learning environment		
	Rejection proportion	Number enrolled	Top 10%	In-state tuition	Additional fees	Personal costs	Proportion PhD (%)	Faculty/student ratio	Proportion graduated (%)
1 (178)	0.58	−0.38	1.02	1.11	−0.40	−0.83	1.16	0.99	0.92
2 (225)	0.08	1.22	−0.26	−0.17	0.67	0.37	0.27	−0.37	−0.40
3 (302)	−0.44	−0.75	−0.46	0.65	−0.37	−0.41	−0.57	0.19	0.16

Table 7 Description of k -means three-cluster solution by centroids (back-transformed to original scale)

Cluster (size)	Student intake			Affordability			Learning environment		
	Rejection proportion	Number enrolled	Top 10%	In state tuition	Additional fees	Personal costs	Proportion PhD (%)	Faculty/student ratio	Proportion graduated (%)
1 (162)	0.36	786	48	15152	388	1076	87	0.1	82
2 (195)	0.28	1609	20	2497	529	1689	76	0.06	52
3 (347)	0.20	292	27	9910	252	1280	63	0.08	61

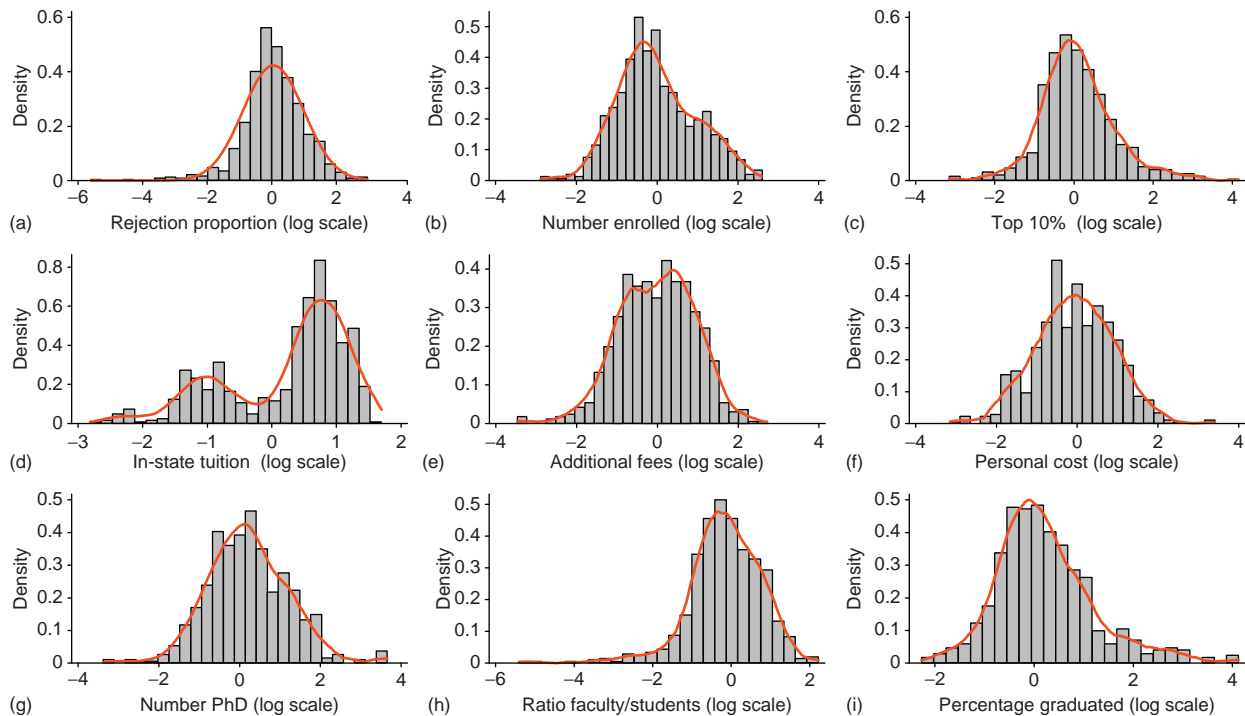


Figure 4 Histograms with overlaid densities for American college data.

Number of Clusters

We have so far studiously avoided the issue of choosing the number of clusters by assuming that an investigator is interested in the whole sequence of nested partitions or that the number of clusters is known *a priori*. However, in practice, one is often forced to decide the number of clusters from the information at hand.

A number of *ad hoc* procedures have been proposed. In the context of hierarchical clustering, selecting a partition is equivalent to cutting the corresponding dendrogram at a given height. This defines a partition such that clusters below that height are distant from each other by at least that amount. The appearance of the dendrogram can therefore informally suggest the number of clusters with large distances between fusion levels, suggesting the “best cut.” For example, applying this rule to the dendrogram in **Figure 2(c)** might suggest five clusters. In the context of optimization, clustering scree plots, which plot the values of the optimized cluster criterion against the number of groups, are most popular. By definition, the criterion improves when the number of groups is increased – so the optimal number of groups is chosen as the level at which large changes in the criterion occur (the elbow in the plot). **Figure 5** shows the scree plot for *k*-means clustering of the college data; confirming that a partition into three clusters was appropriate.

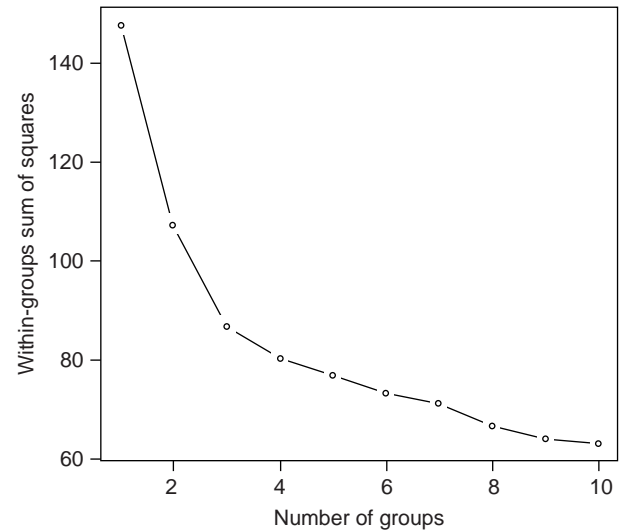
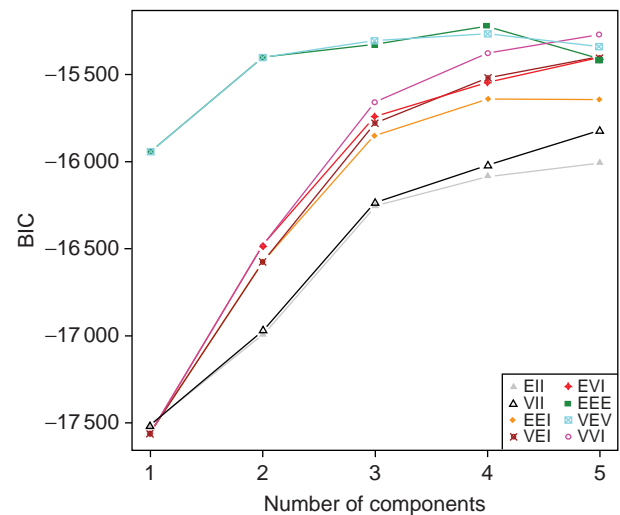
More formal techniques exist which try to overcome the problem of subjectivity. Thirty such methods were reviewed by Milligan and Cooper (1985) in the 1980s and, more recently, 15 indices for high-dimensional binary data were assessed by Dimitriadou *et al.* (2002). Both studies assess the ability of formal/automated methods to detect the correct number of clusters in series of simulated data sets. Based on these simulations techniques introduced by Calinski and Harabasz (1974) and Duda and Hart (1973) are recommended for continuous data, while the index suggested by Ratkowsky and Lance (1978) was the overall best performer for binary data.

More recently, progress on defining formal rules for comparing the quality of different cluster solutions has been made in the context of model-based clustering. As, in this context, cluster memberships are determined by optimizing likelihoods of competing nested models; the latter can be compared to inform the choice of the number of groups. There are problems with the conventional likelihood ratio test as some parameters from the null distribution are on the edge of the parameter space. However, suggestions have been made to overcome this, with perhaps the most practical one being the use of information criteria such as the Bayesian information criterion (BIC) or Akaike’s information criterion (AIC) (Burnham and Anderson, 2002). As an example, we calculated BIC for a number of models for the college data which differed in the numbers of clusters (from 1 to 5) and in the parametrization of the

Table 8 Estimated probabilities (in %) of responding yes to items for two classes of teachers

		Class 1	Class 2
1	Students have choice in where to sit	22	43
2	Students sit in groups of three or more	60	87
3	Students allocated to sitting by ability	35	23
4	Students stay in same seats for most of day	91	63
5	Students not allowed freedom of movement in classroom	97	54
6	Students not allowed to talk freely	89	48
7	Students expect to ask permission to leave room	97	76
8	Students expected to be quiet	82	42
9	Monitors appointed for special jobs	85	67
10	Students taken out of school regularly	32	60
11	Timetable used for organizing work	90	66
12	Use own materials rather than text books	19	49
13	Students expected to know tables by heart	92	76
14	Students asked to find own reference material	29	37
15	Students given homework regularly	35	22
16	Teacher talks to whole class	71	44
17	Students work in groups on teacher class	29	42
18	Students work in groups on work of their own choice	14	46
19	Students work individually on teacher tasks	55	37
20	Students work individually on work of their own choice	28	50
21	Explore concepts in number work	18	55
22	Encourage fluency in English language even if inaccurate	87	94
23	Students work marked or graded	43	14
24	Spelling and grammatical errors corrected	84	68
25	Stars given to students who produce best work	57	29
26	Arithmetic test given at least once a week	59	38
27	Spelling test given at least once a week	73	51
28	End-of-term tests given	66	44
29	Many students who create discipline problems	09	09
30	Verbal reproof sufficient	97	95
31	Discipline: extra work given	70	53
32	Smack	65	42
33	Withdrawal of privileges	86	77
34	Send to headteacher	24	17
35	Send out of room	19	15
36	Emphasis on separate subject teaching	85	50
37	Emphasis on esthetic subject teaching	55	63
38	Emphasis on integrated subject teaching	22	65
	Percentage of teachers attributed to class	54	46

covariance matrix. The latter were chosen to reflect different constraints displaying natural geometric features which can be derived from its spectral decomposition. Clusters exhibiting similar orientation, shape, or volume, satisfying two or all these restrictions, may be desirable within a particular classification context (Bensmail *et al.*, 1997). A range of these restrictions and their results for model-based selection are displayed in Figure 6: for instance, EVI

**Figure 5** Scree plot for *k*-means clustering of colleges' data.

Distributional shape implied by covariance matrix parameterization (see Fraley and Raftery, 2003)

EII: spherical, equal volume
 EVI: diagonal, equal volume, varying shape
 VII: spherical, unequal volume
 EEE: ellipsoidal, equal volume, shape, and orientation
 EEI: diagonal, equal volume and shape
 VEV: ellipsoidal, equal shape
 VEI: diagonal, varying volume, equal shape
 VVI: diagonal, varying volume and shape

Figure 6 Model selection by maximizing BIC. The best model is EEE with four clusters.

indicates that the clusters are constrained to having the same diagonal orientation and equal volume with their shape allowed to vary. Other symbols in **Figure 6** have analogous interpretations. The results show that the EEE criterion (which assumes the same covariance matrix and is equivalent to requiring similar orientation, volume, and shape for all clusters) performs well with three or four clusters.

Finally, it needs to be emphasized that a clustering algorithm will allocate objects into a prescribed number of groups, irrespective of whether there is any true clustering in the data. Many *ad hoc* methods for comparing the number of groups do not allow the consideration of the simplest clustering solution – the one-group solution. However, when empirically determining the number of clusters it is essential that the one-group solution be considered. Rules derived from maximum likelihood theory will allow this. An alternative approach is the GAP statistic (Tibshirani *et al.*, 2001), which compares the quality of cluster solutions for different numbers of groups based on a given (heuristic) cluster criterion.

Conclusion

The general steps involved in a cluster analysis are:

1. definition of the data matrix (including choice, weighting, or standardization of variables);
2. calculation of the proximity matrix;
3. choice of cluster method (to generate a single or a sequence of partitions);
4. decision regarding the number of clusters (for partitions); and
5. validity checks.

We have talked about steps 1–4 in previous sections but not much has yet been said about the final stage – checking the validity of a cluster solution. There are basically two approaches: internal checks and external checks. Internal checks are aimed at establishing cluster isolation and cohesion or demonstrating robustness of the solution under small changes of method (change of proximity measure, optimization criterion, starting values, etc.) or data set used (splitting data into subsamples, adding an error term, etc.). External checks attempt to establish agreement with a gold standard (if one exists) or with some other yet-unused variable/classification that one would theoretically expect to be associated with the solution. For more information on methods for validation checking see Everitt *et al.* (2001).

There are a number of related techniques not yet mentioned. Constrained clustering imposes restrictions on the possible cluster solutions in order to maintain external features; for example, spatial contiguity. Many applications, particularly in psychology and the social sciences, require overlapping clusters; that is, an object is

allowed to be a member of more than one cluster at the same time – for example, a member of several overlapping social networks. Under some circumstances, direct data clustering, which aims to cluster the (two-mode) data matrix into sets of similar objects and variables, can be applied. Finally, neural network techniques, such as Kohonen's self-organizing map, aimed at unsupervised learning can be considered a cluster method.

Nowadays, carrying out cluster analyses is relatively straightforward. Most general-purpose statistical packages contain procedures for hierarchical and optimization clustering. (The analyses presented here were generated in Stata and R.) Routines for model-based clustering are available in some general-purpose packages (e.g., *mclust* in R, Fraley and Raftery, 2003), specialized latent classes and finite mixtures programs (e.g., *LatentGOLD* and *MIXMOD*), or modeling packages such as *Mplus*. In addition, there are a number of packages solely devoted to cluster analyses (e.g., *Clustan*).

In summary, clustering a set of objects can potentially be very useful. However, care needs to be taken to avoid producing misleading results. Researchers do well to remember that cluster analysis is an exploratory technique rather than an inferential method.

See also: Discrimination and Classification; Latent Class Models.

Bibliography

- Aitkin, M., Anderson, D., and Hinde, J. (1981). Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society A* **144**, 419–448.
- Bensmail, H., Celeux, G., Raftery, E. A., and Robert, C. P. (1997). Inference in model based cluster analysis. *Statistics and Computing* **7**, 1–10.
- Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretical Approach*, 2nd edn. New York: Springer.
- Calinski, R. B. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics* **3**, 1–27.
- Dimitriadou, E., Dolničar, S., and Weingessel, A. (2002). An examination of indexes for determining the number of clusters in binary data sets. *Psychometrika* **67**, 137–160.
- Duda, R. O. and Hart, P. E. (1973). *Pattern Analysis and Scene Analysis*. New York: Wiley.
- Everitt, B. S. and Bullmore, E. T. (1999). Mixture model mapping of brain activation in functional magnetic resonance images. *Human Brain Mapping* **7**, 1–14.
- Hand, D. J., Daly, F., McConway, K., Lunn, D., and Ostrowsky, E. (1994). *A Handbook of Small Data Sets*. London: Chapman and Hall.
- Fraley, C. and Raftery, A. E. (2003). Enhanced software for model-based clustering, discriminant analysis, and density estimation: *MCLUST*. *Journal of Classification* **20**, 263–286.
- Gower, J. C. and Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification* **5**, 5–48.
- Kaufman, L. and Rousseeuw, P. J. (1990). *Finding Groups in Data, an Introduction to Cluster Analysis*. New York: Wiley.
- Lazarsfeld, P. L. and Henry, N. W. (1968). *Latent Structure Analysis*. Boston: Houghton Mifflin.

- McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.
- Milligan, G. W. and Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50**, 159–179.
- Ratkowsky, D. A. and Lance, G. N. (1978). A criterion for determining the number of groups in a classification. *Australian Computer Journal* **10**, 115–117.
- Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society, B* **63**, 411–423.

Further Reading

- Everitt, B. S., Landau, S., and Leese, M. (2001). *Cluster Analysis*, 4th edn. London: Arnold.

Relevant Website

<http://mathforum.org> – the Math Forum Data Collection.

Cognitive Psychology and Educational Statistics

H S Kim and S Embretson, Georgia Institute of Technology, Atlanta, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Abstract reasoning ability – The ability to analyze information and solve problems on complex concepts, which is also considered to be the ability to learn new things quickly.

Cognitive psychology – The branch of psychology that examines internal mental processes, including how people think, perceive, remember, and learn.

Item stimulus features – The item content, such as topics or variables that impact cognitive processes applied to an item.

Psychometrics – The field of study concerned with the theories, models, and statistical techniques applied to develop educational and psychological tests, which includes the measurement of achievement, abilities, attitudes, and personality traits.

Spatial ability – The ability to generate, retain, retrieve, and transform well-structured visual images.

Cognitive psychology, particularly the information-processing analysis of problem solving that began in the 1970s, has influenced important developments in psychometrics. According to Mislevy (2006: 262), the focus of the information-processing analysis in psychometrics is on what is happening within people's heads while they are responding to items. Thus, the interpretation of test performance reflects a complex combination of component processing skills, strategies, and knowledge structures (Snow and Lohman, 1993). As noted by Bejar (2008), the synergy between cognitive psychology and psychometrics also led to cognitively based item generation in the 1980s. A well-developed cognitive theory can provide a framework for item design. The cognitive theory elaborates how specific item stimulus features influence the cognitive requirements involved in responding to the item, which in turn allows the test developer to design items for specific levels and sources of cognitive complexity. These design principles may also be embedded in a computer program to automatically generate items with predictable psychometric properties. Item generators have been developed for measuring nonverbal or fluid intelligence, including matrix reasoning items (Embretson, 1998, 1999) and spatial visualization items (Embretson and Gorin, 2001). An item generator has also been developed for mathematical reasoning

items, which can produce new items while the examinee takes the computerized adaptive test (see Bejar *et al.*, 2003).

However, the impact of the cognitive psychology approach is strongest when appropriate statistical models are applied to estimate parameters that reflect the impact of the cognitive test design features on item psychometric properties. If the prediction from the cognitive model is sufficiently strong, it may be possible to use items with little or no empirical tryout (Embretson and Yang, 2007).

In this article, the test design approach for cognitively based item generation is reviewed, and some statistical models that are appropriate for calibrating item psychometric properties are presented. Finally, the test design approach and the corresponding statistical models are illustrated by applications to spatial ability and abstract reasoning.

Test Design for the Cognitively Based Item Generation

Cognitive principles can be incorporated into test design to support construct validity and to test score interpretations. In an earlier article (Embretson, 1998), it was noted that the traditional test development process associated with construct validation has limited implications for test design for two reasons. First, the traditional conceptualization of construct validity emphasizes the establishing of significant relationships of test scores with external variables, such as other measures or criteria. However, this approach confounds the theoretical meaning of the construct with its significance in prediction. Thus, the role of cognitive theory in test design is limited. Second, in the traditional test development process, the evaluation of construct validity is possible only after the test has been developed and data have been collected. If the results do not support the intended construct validity of the test, then test developers must either redefine the construct or redesign the test. Since the latter possibility requires new evidence of construct validity of the test, redefining the construct is often the more practical option. Because of these limitations, the traditional test development approach does not fully establish the theoretical nature of the constructs that are measured during test development phase.

To overcome the limitations of the traditional approach, a principled test design approach with the centralization of cognitive theory is necessary. Mislevy's

evidence-centered design (Mislevy, 1994) and Embretson's cognitive design system (CDS; Embretson, 1994, 1998) are two principled test design approaches. Although the evidence-centered design approach is very flexible and permits diverse applications, the CDS approach is more directly applicable to typical large-scale assessments because items may be banked by estimates of cognitive complexity, which can be useful in equating both linear and adaptive test forms.

The CDS framework interfaces item design principles with test validity by including both a conceptual and a procedural framework. In the conceptual framework, the construct validation concept is divided into two aspects: construct representation and nomothetic span (see Embretson, 1983, 1998). Construct representation concerns the meaning of test scores by understanding the processes, strategies, and knowledge that examinees use to solve items. Nomothetic span concerns the significance of test scores, as determined from external evidence, such as correlations of test scores with other variables. Even though nomothetic span does not define test meaning, it is the consequence of construct representation (Embretson, 1998). The two-part distinction in construct validity can be applied to establish the theoretical nature of the constructs that are measured in the test design phase.

The procedural framework is a series of stages that are shown in **Table 1**. In the first two stages, specification of general goals of measurement is followed by the identification of design features for test items. Although the traditional test development approach also has these stages, cognitive psychology has a central role in CDS because item design features that impact cognitive processing demands are identified. The next four stages in **Table 1** concern item development. This is a research process in which the cognitive modeling results are combined with the target measurement goals. Test developers select or develop cognitive models to represent the item-solving process, and then these models are evaluated empirically on test items. Once a plausible cognitive model has been obtained, the item specifications can be developed based on the variables in the model. That is, the cognitive model variables define stimulus features in the items that can be varied to produce items with specified levels and sources of cognitive complexity. In addition, item stimulus features that prompt irrelevant processes to target measurement goals can be minimized. In the remaining stages, test developers evaluate the psychometric properties of the test using statistical models of item properties. Item parameters are estimated to assess the impact (i.e., magnitude and significance) of the item stimulus features on psychometric properties (e.g., item difficulty). Significant weight parameters of the stimulus features can be interpreted as supporting the cognitive model as well as the construct representation of the test. Another advantage of the CDS approach is the possibility

Table 1 Cognitive design systems

Stage
Specify general goals of measurement <ul style="list-style-type: none"> • Construct representation (meaning) • Nomothetic span (significance)
Identify design features in task domain <ul style="list-style-type: none"> • Task-general features (mode, format, and conditions) • Task-specific features
Develop a cognitive model <ul style="list-style-type: none"> • Review theories • Select or develop model for psychometric domain • Revise model • Test model
Evaluate cognitive model for psychometric potential <ul style="list-style-type: none"> • Evaluate cognitive model plausibility on current test • Evaluate impact of complexity factors on psychometric properties • Anticipate properties of new test
Specify item distributions on cognitive complexity <ul style="list-style-type: none"> • Distribution of item complexity parameters • Distribution of item features
Generate items to fit specifications <ul style="list-style-type: none"> • Artificial intelligence
Evaluate cognitive and psychometric properties for revised test domain <ul style="list-style-type: none"> • Estimate component latent trait model parameters • Evaluate plausibility of cognitive model • Evaluate impact of complexity factors on psychometric properties • Evaluate plausibility of the psychometric model • Calibrate final item parameters and ability distributions
Evaluate psychometric <ul style="list-style-type: none"> • Measure processing abilities • Bank items by cognitive processing demands
Assemble test forms to represent specifications <ul style="list-style-type: none"> • Fixed content test • Adaptive test
Validate

of banking items. As item parameters can be predicted by the stimulus features in the cognitive model prior to tryout, items can be banked by their cognitive demands.

Statistical Models

The many advantages of principled test design involve statistical models to link substantive features of items to psychometric properties. Next, some models that are appropriate for CDS and other design systems are reviewed.

Linear Logistic Test Model

The linear logistic test model (LLTM; Fischer, 1973) was the first psychometric model to effectively bridge cognitive psychology and item design. That is, the model can incorporate item stimulus features into the prediction of

item success. LLTM is a generalization of the Rasch item response theory (IRT) model, which is given as

$$P(X_{is} = 1) = \frac{\exp(\theta_s - \beta_i)}{1 + \exp(\theta_s - \beta_i)}, \quad [1]$$

where $P(X_{is} = 1)$ is the probability that person s answers item i correctly, θ_s is the trait level of person s , and β_i is the difficulty parameter of item i . Because β_i does not include the cognitive variables (item stimulus features) involved in the item, it is replaced with a linear function of cognitive variables in LLTM as follows:

$$\beta_i = \sum_{k=0}^K \eta_k q_{ik} = \eta_0 q_{i0} + \eta_1 q_{i1} + \eta_2 q_{i2} + \cdots + \eta_K q_{iK}, \quad [2]$$

where η_k represents the effect of stimulus feature k , q_{ik} is the score (e.g., 0 = absence; 1 = presence) of stimulus feature k of item i , and $\eta_0 q_{i0}$ is the intercept of the equation. The full LLTM combines eqn [1] with eqn [2]:

$$P(X_{is} = 1) = \frac{\exp(\theta_s - \sum_{k=0}^K \eta_k q_{ik})}{1 + \exp(\theta_s - \sum_{k=0}^K \eta_k q_{ik})}. \quad [3]$$

The \mathbf{Q} matrix that consists of q_{ik} (the indicators of k stimulus features of i items) is usually structured as a (I, K) matrix with rank K , where $K < I$. The method of coding the indicators, q_{ik} , can be either dummy coding or scores for the item on cognitive model variables, depending on the purpose of the application. The example of spatial ability presented below will elaborate how the q_{ik} is coded to represent cognitive model variables.

Two-Parameter-Logistic-Constrained Model

The two-parameter-logistic(2PL)-constrained model (Embretson, 1999) contains parameters to represent the impact of stimulus features on item discrimination as well as on item difficulty. It was first presented with an application to matrix completion items in which item difficulty and item discrimination were both predictable from item stimulus features (Embretson, 1999). Analogous to LLTM, which is a generalization of the Rasch IRT model, the 2PL-constrained model is a generalization of the 2PL model. In the 2PL-constrained model, both item discrimination (α_i) and item difficulty (β_i) parameters of the 2PL model are replaced with the linear combinations of cognitive variables as follows:

2PL model :

$$P(X_{is} = 1) = \frac{\exp[\alpha_i(\theta_s - \beta_i)]}{1 + \exp[\alpha_i(\theta_s - \beta_i)]}; \quad [4]$$

$$\alpha_i = \sum_{k=0}^K \tau_k q_{ik}, \beta_i = \sum_{k=0}^K \eta_k q_{ik}; \quad [5]$$

2PL-constrained model:

$$P(X_{is} = 1) = \frac{\exp[\sum_{k=0}^K \tau_k q_{ik}(\theta_s - \sum_{k=0}^K \eta_k q_{ik})]}{1 + \exp[\sum_{k=0}^K \tau_k q_{ik}(\theta_s - \sum_{k=0}^K \eta_k q_{ik})]}, \quad [6]$$

where $P(X_{is} = 1)$ is the probability that person s passes item i , θ_s is the ability of person s , q_{ik} is the score for stimulus feature k in item i , τ_k and η_k are the weights (or effects) of stimulus feature k in item discrimination and item difficulty, respectively, and $\tau_0 q_{i0}$ and $\eta_0 q_{i0}$ are the intercepts of the equations.

The advantage of the 2PL-constrained model over LLTM is that it includes design features for item discrimination. An item with higher discriminating power provides more information about latent trait (θ) along with less measurement error because the changes in trait level have a greater impact on $P(X_{is} = 1)$ of that item. If we use the 2PL-constrained model, we can identify which item stimulus features affect the discriminating power of the item as well as item difficulty. The application of the 2PL-constrained model is introduced in the example of abstract reasoning ability test below.

Linear Partial Credit Model

The linear partial credit model (LPCM; Fischer and Ponocny, 1994) was developed to assess the effects of treatments based on polytomous ordered response items. Like the LLTM, the LPCM includes the linear function of cognitive variables involved in the item. However, unlike the LLTM that is applicable only to binary response data (e.g., right or wrong), it is designed for analyzing polytomous ordered responses, which are commonly used for attitude or self-rating items. The LPCM is based on the partial credit model, which assigns one independent parameter β_{ib} for each response category of an item as follows:

$$P(X_{ibs} = 1) = \frac{\exp(b\theta_s + \beta_{ib})}{\sum_{b=0}^{H_i} \exp(b\theta_s + \beta_{ib})}, \quad [7]$$

for $i = 1, \dots, I$ items and $b = 0, \dots, H_i$ response categories of item i , where $P(X_{ibs} = 1)$ is the probability that person s chooses category b on item i (all X_{ibs} are assumed to be locally independent), θ_s is the position of person s on the underlying latent trait, and β_{ib} is the easiness parameter of category b of item i (β_{i0} and $\sum_{i=1}^I \sum_{b=1}^{H_i} \beta_{ib}$ are set to zero for model identification and normalization).

In the LPCM, β_{ib} is replaced with a linear function of cognitive variables

$$\beta_{ib} = \sum_{k=0}^K \alpha_k w_{ibk}, \quad [8]$$

where α_k is called the basic parameter, which measures the effect of stimulus feature or experimental treatment k on the response in item i , w_{ibk} is the value of the stimulus

feature or dosage of the treatment, and $\alpha_0 w_{ib0}$ is a normalization constant of the equation. Notice that α_k and w_{ibk} in the LPCM correspond to η_k and q_{ik} in the LLTM (see eqn [2]), respectively. The combination of eqn [7] with eqn [8] is the LPCM.

$$P(X_{ibs} = 1) = \frac{\exp(b\theta_s + \sum_{k=0}^k \alpha_k w_{ibk})}{\sum_{b=0}^{H_i} \exp(b\theta_s + \sum_{k=0}^k \alpha_k w_{ibk})} \quad [9]$$

One of the attractive features of the LPCM is its great flexibility in allowing the w_{ibk} parameter for each response category, thus resulting in a good fit to complex data. However, it should be noted that there is always a trade-off between the flexibility and the simplicity (ease of interpretation) of a model.

Multicomponent Latent Trait Model

Many ability and achievement tests have sources of multidimensionality in the item domain. For example, the Scholastic Aptitude Test (SAT) algebra items require multiple skills or stages to arrive at a correct response, such as prerequisite skills for symbols and conventions, algebraic manipulations, linear functions, and simultaneous equations (Gierl *et al.*, 2007). To better explain why an examinee answers a specific item incorrectly, statistical models that can assess the sources of multidimensionality are needed.

One important aspect of multidimensionality is that the skills or processing stages in the items are often sequentially dependent. Thus, cognitive models for tasks typically postulate a flow of information from one stage to another. Four possible forms of hierarchical structures of processing stages in task performance are presented

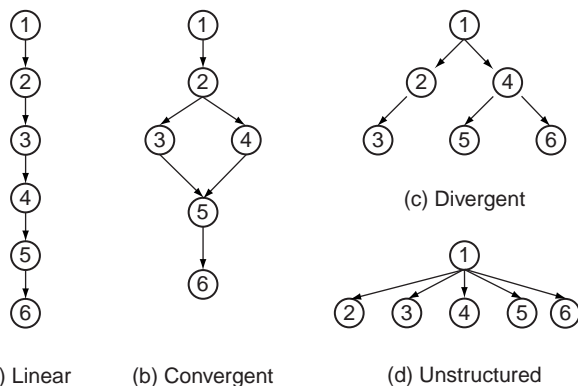


Figure 1 Four different hierarchical structures. Reproduced from Gierl, M. J., Leighton, J. P., and Hunka, S. M. (2007). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills. In Leighton, J. P. and Gierl, M. J. (eds.) *Cognitive Diagnostic Assessment for Education*, pp 242–274. New York: Cambridge University Press.

in **Figure 1** (Gierl *et al.*, 2007). In all the structures, there is a prerequisite skill (stage 1) and orderings among all or some processing stages except in **Figure 1(d)**, which is an unstructured hierarchy.

The multidimensional latent trait model (MLTM; Whitely, 1980) was designed to properly reflect the sequentially dependent stages (called components in this model) based on a product of processing outcome probabilities as follows:

$$P(X_{is} = 1) = \prod_k P(X_{isk}) = \prod_k \frac{\exp(\theta_{sk} - \beta_{ik})}{1 + \exp(\theta_{sk} - \beta_{ik})}, \quad [10]$$

where $P(X_{is} = 1)$ is the probability of success for person s on item i and $\prod_k P(X_{isk})$ is the product of success on each processing component k , given the correct outcome of the preceding component. The right side of the equation contains terms from the Rasch models for the probability of success on each component, where θ_{sk} is the trait level of person s on component k and β_{ik} is the difficulty of item i on component k .

General Component Latent Trait Model

The general component latent trait model (GLTM; Embretson, 1984) is the generalization of the MLTM, which incorporates component stimulus features into the prediction of the component difficulty (β_{ik}) in the item. Therefore, β_{ik} in Equation 10, is replaced with the weighted sum of underlying stimulus features as follows:

$$\beta_{ik} = \sum_{m=0}^m \eta_{km} q_{ikm}, \quad [11]$$

where q_{ikm} is an indicator of the presence (or polynomial values) of stimulus feature m on component k for item i , η_{ikm} is the weight of stimulus feature m on component k , and $\eta_{k0} q_{ik0}$ is an intercept. The full GLTM combines eqn [10] with eqn [11].

$$P(X_{isT} = 1) = \prod_k \left[\frac{\exp(\theta_{sk} - \sum_{m=0}^m \eta_{km} q_{ikm})}{1 + \exp(\theta_{sk} - \sum_{m=0}^m \eta_{km} q_{ikm})} \right]. \quad [12]$$

Like MLTM, GLTM can be estimated readily when both subtask and full task data are available for the same item. The GLTM can be estimated with full task data only under certain circumstances. For example, in Embretson's (1995) study, a working-memory capacity component was separated from the control processes in performance on spatial ability items because a highly predictive model for the difficulty of working memory load was available. Other circumstances in which GLTM can be estimated include setting constraints, data augmentation, and component structures that vary between items.

The GLTM enables us to examine how the underlying stimulus features will impact the difficulty of each component (β_{ik}) based on preestablished cognitive theories. Since GLTM is an extension of the MLTM, it also

estimates individual ability on each component, thus giving more detailed information about an examinee's skill profile.

Applications

Two examples are presented below to illustrate the application of the statistical models to item design. Both examples involve nonverbal measures of ability. Applications of LLTM and the 2PL-Constrained model to items that appear on achievement tests, such as paragraph comprehension (Gorin, 2005) and mathematical problem solving (Embretson and Daniel, 2008), are also available.

Spatial Ability

Embretson (1994) developed the spatial learning ability test (SLAT) by applying the CDS approach. In her study, the practical efficiency of the CDS and the constructive representation of the SLAT were strongly supported. Based on a series of research and cognitive models of the spatial folding task, Embretson (1994) found that the degrees of rotation and the number of surfaces carried provided a good prediction of both item difficulty and response time (see **Figure 2**).

Figure 2 provides some examples to show what the degree of rotation means and what the number of surfaces carried implies in the SLAT item. This figure consists of an unfolded stem on the left side and six cubes (or keys) on the right side, which possibly result from folding down the unfolded stem. The six different keys vary in the degrees of rotation and in the number of surfaces carried. For example,

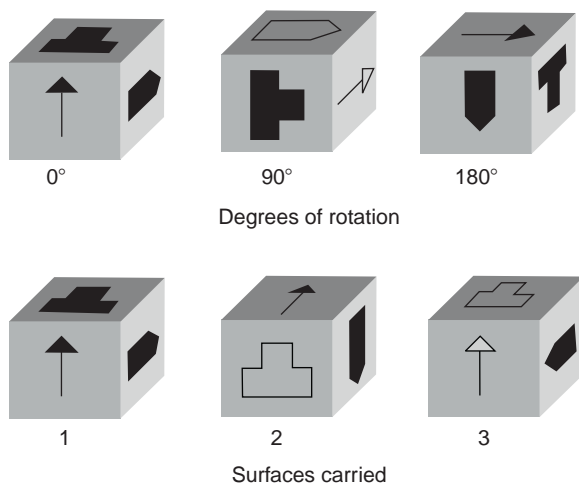


Figure 2 A stem and six correct answers that vary in the degrees of rotation and the surfaces carried. Reproduced from Embretson, S. E. (1994). Application of cognitive design systems to test development. In Reynolds, C. R. (ed.) *Cognitive Assessment: A Multidisciplinary Perspective*, pp 107–135. New York: Plenum Press.

the three keys in the top row of **Figure 2** require different degrees of mental rotation of the unfolded stem to match the stem with the keys 0° , 90° , and 180° , respectively. However, the three keys in the bottom row have 0° of rotation but vary in the number of surfaces required to be carried in folding processes. That is, in the first key of the three, all the three surfaces of \uparrow , \blacksquare , and \blacksquare are adjacent; thus the number of surfaces required to be carried in the mental folding processes is counted as one. In the second key, the surfaces of \uparrow and of \blacksquare are not adjacent; they are two surfaces apart from each other, thus requiring more complex mentally folding process than in the first key. The maximum number of surfaces required to be carried in the mental folding processes is counted as two in this case. Likewise, the maximum number of surfaces required to be carried in the third key is three because the surfaces of \uparrow and of \blacksquare are three surfaces apart from each other. These two sources of cognitive complexity (the degrees of rotation and the number of surfaces carried) were found to have a linear relationship to item difficulty and item response time, thus supporting the theory that mental rotation is analogous to physical rotation (Embretson, 1994).

Distractor types may also impact the folding process (Embretson, 1994). **Figure 3** presents two sample SLAT items. The task is to select one of the four distractors that best matches the unfolded stem on the left side. For the top item of **Figure 3**, each distractor has the same markings (called matched positions), whereas all distractors of the bottom item show different markings (called mixed positions). Mixed positions in the distractors seem to cause greater response time and less accuracy (Embretson, 1994). In the distractors of the top item, the two surfaces of \uparrow and \square are called anchoring points because they are pairs of surfaces with the same markings between the unfolded stem and a distractor. A search process for anchoring points is a prerequisite to the rotation and the folding process in order to solve a SLAT item. Finally, the item on the bottom of **Figure 3** contains two surfaces with undirected markings (i.e., \blacksquare and \square). These two markings

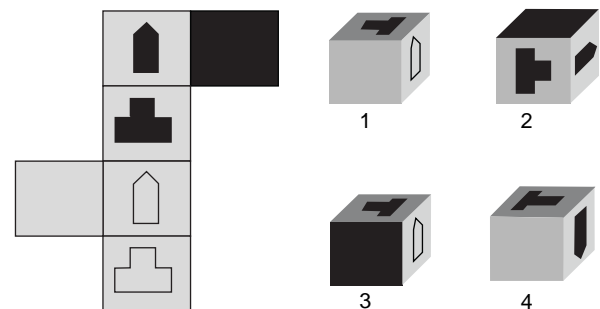


Figure 3 Two SLAT items. From Embretson, S. E. (1994). Application of cognitive design systems to test development. In Reynolds, C. R. (ed.) *Cognitive Assessment: A Multidisciplinary Perspective*, pp 107–135. New York: Plenum Press.

(called undirected markings) appear the same in every orientation, unlike all other markings such as \square and \blacksquare (called directed markings). An item containing undirected markings does not require folding processes and is called a position problem, whereas an item with only directed markings is called a spatial problem. The position problem will be easier than the spatial problem because there is no involvement of the folding process (Embretson, 1994).

In the research on SLAT, eight cognitive variables (stimulus features) were scored in the LLTM based on the cognitive models of the spatial folding task, elaborated above: (1) distractor type, (2) problem type, (3) degrees rotation-linear, (4) degrees rotation-quadratic, (5) surfaces carried-linear, (6) surfaces carried-quadratic, (7) degrees rotation-linear (position items), and (8) degrees rotation-quadratic (position items). **Table 2** shows details on coding each stimulus feature (q_{ik}).

Distractor type and problem type variables are coded by binary numbers (0 and 1). The degrees of rotation and the number of surfaces carried ($q_3, q_4, q_5, q_6, q_7, q_8$) are coded by polytomous numbers to test for both linear and

quadratic trends. For example, if the item has a 90° rotation, the values for the linear contrast and for the quadratic contrast (i.e., q_3 and q_4) will be 0 and -2 , respectively (see **Table 2**). It should be noted that q_3, q_4, q_5 , and q_6 are coded only for spatial items (otherwise, coded as 0), while q_7 and q_8 are scored for position items only (otherwise, coded as 0). Therefore, the degrees of rotation of spatial and position problems are scored separately. For example, if a spatial item has a mixed orientation of distractors, a 180° rotation, and two surfaces carried, then item difficulty (β_i) would be predicted as follows (refer to eqn [2]):

$$\hat{\beta}_i = \eta_0 q_0 + \eta_1(1) + \eta_2(0) + \eta_3(1) + \eta_4(1) + \eta_5(0) + \eta_6(-2) + \eta_7(0) + \eta_8(0).$$

Embretson (1994) also extended mathematical modeling with the cognitive variables to model item response time. The effect parameter (η) of each cognitive variable (q_k) on item difficulty was estimated with LLTM, while the regression coefficient (b) of q_k on item response time was estimated with regression analysis. **Table 3** provides the estimates of the parameters for item difficulty and response time. As shown in **Table 3**, degrees of rotation and the number of surfaces carried contrasts have a significant impact on both item difficulty and item response time in the SLAT. Problem type (spatial vs. position) has a strong impact on item difficulty but not on response time. However, distractor type (matched vs. mixed) is not significant for either item difficulty or item response time.

Abstract Reasoning Ability

The Abstract Reasoning Test (ART; Embretson, 1995) was developed based on the Carpenter *et al.* (1990) processing theory of matrix completion tasks. In the Carpenter *et al.* theory, the major processes involved in solving the matrix completion problems are the generating and evaluating of relationships across the rows and columns. Based on results

Table 2 Cognitive variables and the coding in LLTM

q_k	Cognitive variables	Coding
q_1	Distractor type	0 = matched, 1 = mixed
q_2	Problem type	0 = spatial, 1 = position
q_3	Degrees rotation-linear	$-1 = 0^\circ, 0 = 90^\circ, 1 = 180^\circ$
q_4	Degrees rotation-quadratic	$1 = 0^\circ, -2 = 90^\circ, 1 = 180^\circ$
q_5	Surfaces carried-linear	$-1 = 1 \text{ surface}, 0 = 2 \text{ surfaces}, 1 = 3 \text{ surfaces}$
q_6	Surfaces carried-quadratic	$1 = 1 \text{ surface}, -2 = 2 \text{ surfaces}, 1 = 3 \text{ surfaces}$
q_7	Degrees rotation linear position-items	$-1 = 0^\circ, 0 = 90^\circ, 1 = 180^\circ$
q_8	Degrees rotation quadratic position-items	$1 = 0^\circ, -2 = 90^\circ, 1 = 180^\circ$

Table 3 Cognitive models of item difficulty and response time

Cognitive variable	Item difficulty		Response time	
	η	σ_η	b	σ_b
1. Distractor type	-0.07	0.04	0.78	0.48
2. Problem type	-0.10**	0.05	-0.16	0.55
3. Degrees rotation-linear	0.16**	0.03	1.76**	0.34
4. Degrees rotation-quadratic	0.10**	0.02	0.48*	0.20
5. Surfaces carried-linear	0.83**	0.03	3.65**	0.34
6. Surfaces carried-quadratic	-0.18**	0.02	-1.09**	0.20
7. Degrees rotation linear position-items	-0.01	0.05	-0.07	0.58
8. Degrees rotation quadratic position-items	0.19**	0.03	0.96**	0.34

* $p < 0.05$; ** $p < 0.01$.

Adapted from Embretson, S. E. (1994). Application of cognitive design systems to test development. In Reynolds, C. R. (ed.) *Cognitive Assessment: A Multidisciplinary Perspective*, pp 107–135. New York: Plenum Press.

from a variety of experiments, it was postulated that working-memory capacity was a primary source of individual differences in item solving (Carpenter *et al.*, 1990).

The working-memory load in a matrix problem is influenced by both the number and the level of relationships involved across the rows and columns. First, for the number of relationships of the ART item in **Figure 4**, there is a distribution of three objects (\square , \circ , and Δ), which occur once and only once in each row and column. Moreover, there are two pairwise progressions, that is, the interior object (\times) changes its girth across the columns and its density across the rows. Therefore, a total of three relationships (one distribution of three + two pairwise progressions) are involved in the item in **Figure 4**. Second, for the level of relationships, Carpenter *et al.* (1990) identified five levels, as shown in **Table 4**.

It was observed that people examined one relationship at a time from level one to the higher levels in **Table 4** until they reached the required relationship to solve the matrix item (Carpenter *et al.*, 1990). Thus, they hypothesized that the higher-level relationship requires more processing time as well as extra working-memory capacity to remember the results of earlier processing than the lower-level relationship does. Therefore, greater numbers

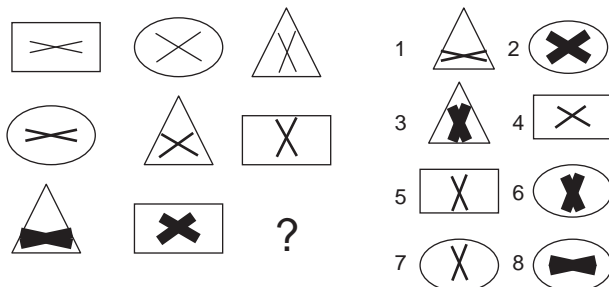


Figure 4 A sample item of the ART. Adapted from Embretson, S. E. (2002). Generating abstract reasoning items with cognitive theory. In Irvine, S. and Kyllonen, P. (eds.) *Generating Items for Cognitive Tests: Theory and Practice*, pp 219–250. Mahwah, NJ: Erlbaum.

Table 4 Order of relationships to apply for matrix problems

Level	Relationship	Definition
1	Constant in a row	An element is the same across rows or columns
2	Pairwise progression	An element changes systematically from entry to entry
3	Figure addition/subtraction	The first two entries in the row or column visually sum to the last entry
4	Distribution of three	An object or attribute appears just once in each row and column
5	Distribution of two	Distribution of three has null values (i.e., one matching element is missing)

and levels of relationships require more working-memory capacity in an examinee. The item in **Figure 4** has two pairwise progressions (level 2, in **Table 4**) and a distribution of three (level 4, in **Table 4**) as we already know. Therefore, the number of the relationships involved in this ART item in **Figure 4** is three, and the levels of the relationships involved in the item are 2 and 4.

Carpenter *et al.* (1990) proposed an additional source of individual difference in solving matrix problems, namely, abstraction capacity. Abstraction capacity is involved in finding abstract correspondence between figures, especially when (1) the correspondence is based on the properties of objects (e.g., position in the entry, texture, orientation), (2) corresponding objects are distorted, or (3) the relationship involves null values. Null values are characteristic of the distribution of two relationships (level 5, in **Table 4**). For example, the sample item in **Figure 5** has a distribution of two objects (\square , Δ) across the rows and columns, while the third object is null. In addition, a figure addition/subtraction relationship (level 3 in **Table 4**) can be found among the interior objects (e.g., $++$, $++$, and $+$). Null values are applicable to the addition/subtraction relationship, if the results of subtractions are zeros.

Carpenter *et al.* (1990) did not introduce any mathematical modeling to predict item difficulty by the sources of individual difference (i.e., working-memory load or abstract correspondence). Embretson (1998) developed ART based on the cognitive variables of Carpenter *et al.* (1990) using the CDS approach. Further, the predictability of item performance from the cognitive model variables was also studied using one of the statistical models described above. To generate items in the CDS approach, the cognitive variables should first be scored in an appropriate statistical model as in the previous example of the SLAT. Embretson (1998) added three perceptual variables to examine if the drawing principles for the item influence item difficulty, in addition to the two cognitive variables (i.e., number of relationships and abstract correspondence). Embretson's perceptual variables are overlay, fusion, and distortion: (1) overlay is scored 1 if independent objects (e.g., \square or \times) are overlaid in an array entry as in **Figure 4**, and 0 if they are placed side by side in the same array location; (2) fusion is scored 1 if two objects placed side by side appear as one object, and 0 otherwise; and (3) distortion is scored 1 if the shape of corresponding objects change perceptually (bending, twisting, stretching, etc.), and 0 otherwise. The 2PL-constrained model was applied because preliminary analyses revealed somewhat unequal item discriminations in the ART. **Table 5** presents two cognitive variables (q_1 , q_2) and three perceptual variables (q_3 , q_4 , and q_5) in the ART and specifies the codes of the sample item of **Figure 4**.

Embretson (1999) estimated the weights (τ_k , η_k) of the stimulus features using a sample of 818 young adults with

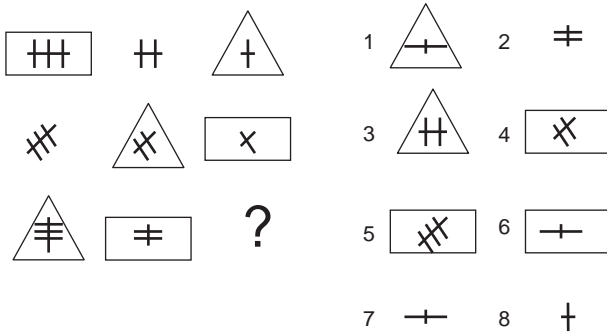


Figure 5 A sample ART item with null values. Adapted from Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods* 3, 300–396.

Table 5 Cognitive/perceptual variables in the ART and the codes for the sample item in Figure 4

q_k	Cognitive/perceptual variables	Codes for the sample item
q_1	Number of relationships	3 (One distribution of three + 2 pairwise progressions)
q_2	Abstract correspondence	0 (no abstract correspondence)
q_3	Overlay	1 (overlaid)
q_4	Fusion	0 (no fusion)
q_5	Distortion	0 (no distortion)

Table 6 Estimates for 2PL-constrained model

Item stimulus feature	Item discrimination		Item difficulty	
	τ	σ_τ	η	σ_η
Intercept	1.710**	0.071	–1.421**	0.069
1. Number of relationships	–0.327**	0.021	0.514**	0.022
2. Abstract correspondence	0.056	0.037	0.844**	0.041
3. Overlay	–0.067	0.038	0.245**	0.039
4. Fusion	–0.107*	0.052	–3.08**	0.047
5. Distortion	0.237**	0.054	–0.064	0.056

* $p < 0.05$; ** $p < 0.01$.

30 item forms of the ART. Table 6 provides the estimated weights for item discrimination and item difficulty and their standard errors. The number of relationships had a significant effect for both item discrimination and item difficulty at the 0.01 level. The number of relationships was negatively associated with item discrimination and positively associated with item difficulty. That is, an item with more relationships across the rows and columns was more difficult but had less discriminating power. Abstract correspondence showed a strong impact on item difficulty but not on item discrimination. For the perceptual features,

fusion showed significant impact on both item discrimination and on item difficulty at the 0.05 level. Therefore, the item discrimination (α_i) and item difficulty (β_i) of the sample item in Figure 4 can be predicted using the 2PL-constrained model as follows:

$$\hat{\alpha}_i = \sum_{k=0}^K \tau_k q_{ik} = 1.710 - 0.327(3) + 0.056(0) - 0.067(1) - 0.107(0) + 0.237(0) = 0.662;$$

$$\hat{\beta}_i = \sum_{k=0}^K \eta_k q_{ik} = -1.421 + 0.514(3) + 0.844(0) + 0.245(1) - 3.08(0) - 0.064(0) = 0.366.$$

Summary

Cognitive psychology has greatly improved the understanding of skills and knowledge structures in task performance. The cognitive theories may also be applied to the development of educational and psychological testing when items have been studied as cognitive tasks. In this article, a principled test design approach, the CDS, and some statistical models that are appropriate to incorporate cognitive theories into the test design system were reviewed. In addition, examples of applications to two different tests were elaborated to show how cognitive variables can be scored on preexisting cognitive theories and how appropriate statistical models can be applied to estimate the effect of the cognitive variable on item parameters. Mathematical modeling with well-supported cognitive theories makes it possible to predict item psychometric properties with little or no empirical tryout, thus bringing the interesting possibility of automatic item generation during the test.

Bibliography

- Bejar, I. I. (2009). Model based item generation: A review of recent research. In Embretson, S. E. (ed.) *New Directions in Measuring Psychological Constructs with Model-Based Approaches*. Washington, DC: American Psychological Association Books.
- Bejar, I. I., Lawless, R. R., Morley, M. E., et al. (2003). A feasibility study of on-the-fly item generation in adaptive testing. *Journal of Technology, Learning, and Assessment* 2. <http://www.jtla.org> (accessed August 2009).
- Carpenter, P. A., Just, M. A., and Shell, P. (1990). What one intelligence test measures: A theoretical account of processing in the Raven's progressive matrices test. *Psychological Review* 97, 404–431.
- Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin* 93, 179–197.
- Embretson, S. E. (1984). A general multicomponent latent trait model for response processes. *Psychometrika* 49, 175–186.
- Embretson, S. E. (1994). Application of cognitive design systems to test development. In Reynolds, C. R. (ed.) *Cognitive Assessment: A Multidisciplinary Perspective*, pp 107–135. New York: Plenum Press.
- Embretson, S. E. (1995). The role of working memory capacity and general control processes in intelligence. *Intelligence* 20, 169–189.

- Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods* **3**, 300–396.
- Embretson, S. E. (1999). Generating items during testing: Psychometric issues and models. *Psychometrika* **64**, 407–433.
- Embretson, S. E. and Daniel, R. C. (2008). Understanding and quantifying cognitive complexity level in mathematical problem solving items. *Psychology Science* **50**, 328–344.
- Embretson, S. E. and Gorin, J. (2001). Improving construct validity with cognitive psychology principles. (Invited article). *Journal of Educational Measurement* **38**, 343–368.
- Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica* **37**, 359–374.
- Fischer, G. H. and Ponocny, I. (1994). An extension of the partial credit model with an application to the measurement of change. *Psychometrika* **59**, 177–192.
- Gierl, M. J., Leighton, J. P., and Hunka, S. M. (2007). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills. In Leighton, J. P. and Gierl, M. J. (eds.) *Cognitive Diagnostic Assessment for Education*, pp 242–274. New York: Cambridge University Press.
- Gorin, J. S. (2005). Manipulation of processing difficulty on reading comprehension test questions: The feasibility of verbal item generation. *Journal of Educational Measurement* **42**, 351–373.
- Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika* **59**, 439–483.
- Mislevy, R. J. (2006). Cognitive psychology and educational assessment. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn, pp 257–305. Westport, CT: American Council on Education/Praeger.
- Snow, R. E. and Lohman, D. F. (1993). Cognitive psychology, new test design, and new test theory: An introduction. In Frederiksen, N., Mislevy, R. J., and Bejar, I. I. (eds.) *Test Theory for a New Generation of Tests*, pp 1–18. Hillsdale, NJ: Erlbaum.
- Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika* **45**, 479–494.

Further Reading

- Embretson, S. E. (2002). Generating abstract reasoning items with cognitive theory. In Irvine, S. and Kyllonen, P. (eds.) *Generating Items for Cognitive Tests: Theory and Practice*, pp 219–250. Mahwah, NJ: Erlbaum.
- Embretson, S. E. and Yang, X. (2007). Automatic item generation and cognitive psychology. In Rao, C. R. and Sinharay, S. (eds.) *Handbook of Statistics: Volume 26. Psychometrics*, pp 747–767. Amsterdam: Elsevier.

Computational Statistics

J E Gentle, George Mason University, Fairfax, VA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Statistical data analysis has always involved computations, and statisticians have traditionally been heavy consumers of CPU cycles. Grier (1999) describes the concurrent development of data analysis, computing machinery, and statistical computing. The development of the personal computer and of user-friendly software for statistical applications has changed many aspects of a statistician's work, but one thing that has not changed is the pressure to develop methods to utilize advanced computing facilities in statistical analyses.

Although all data analysis involves computation, we do not identify all data analysis as 'statistical computing' or as 'computational statistics'. These terms do not refer to straightforward use of a statistical software package to do a standard analysis. We use the term 'statistical computing' to refer to the computational methods that enable statistical methods. The field of statistical computing is grounded in numerical analysis, database methodology, computer graphics, software engineering, and the computer/human interface. The results of research and development in the area of statistical computing have been incorporated in software packages, and for straightforward problems, very little knowledge of this field is necessary for the applied statistician who merely needs access to the software package. To build software to implement a new statistical method, however, the research statistician or software engineer needs a strong background in statistical computing.

We use the term 'computational statistics' somewhat more broadly to include not only the methods of statistical computing, but also a large class of modern statistical methods that are computationally intensive. The field of computational statistics is grounded in mathematical statistics, statistical computing, and applied statistics. The methods of computational statistics would not be possible without the developments in statistical computing.

Emergence of Computational Statistics as a Subdiscipline

Most of the early leaders in the development of both theoretical and applied statistics, such as Fisher and the Pearsons, were masters of the computational machinery of their day. As the machinery became more advanced in the middle third of the twentieth century, some of the computational aspects of the work were passed off to technicians or else was performed by statisticians who did not understand

the nature of numerical computations on the newer hardware. The numerical computations in many statistical analyses were not of high quality; consequently, the analyses often lacked validity. In the face of a growing realization of the numerical inaccuracies, especially for larger and more ill-conditioned problems, there was a spate of activity by both numerical analysts and statisticians with strong backgrounds in numerical mathematics to advance the quality of algorithms and software for statistical computing.

A Working Party on Statistical Computing was formed within the Royal Statistical Society in the late 1960s, the Statistical Computing Section of the American Statistical Association was formed in 1971, and the International Association for Statistical Computing was founded in 1977 as a Section of the International Statistical Institute. All of these associations remain active today in promoting research in statistical computing. Beginning in 1967, the *Symposium on the Interface of Computer Science and Statistics* has been held annually (with only a few exceptions). The biennial COMPSTAT conferences on computational statistics began in 1974. These scholarly associations and conferences promoted research in statistical computing and contributed greatly to the development of the field.

Another important facet of the development of computational statistics has been the large number of peer-reviewed journals devoted to the field. The journals in the mainstream of statistics also have a large proportion of articles in the fields of statistical computing and computational statistics. This is because, to a large extent, recent developments in statistics and in the computational sciences have gone hand in hand.

Some of the major journals in statistical computing and computational statistics are the following.

- *Communications in Statistics – Simulation and Computation*, published quarterly by Marcel Dekker. (Until 1996, it included algorithms in Fortran. Until 1982, this journal was designated as *Series B*.)
- *Computational Statistics* published quarterly by Physica-Verlag (formerly called *Computational Statistics Quarterly*).
- *Computational Statistics. Proceedings of the xxth Symposium on Computational Statistics (COMPSTAT)*, published biennially by Physica-Verlag. (It is not refereed.)
- *Computational Statistics & Data Analysis*, published by North-Holland. The number of issues per year varies. (This is also the official journal of the International Association for Statistical Computing and as such incorporates the *Statistical Software Newsletter*.)

- *Computing Science and Statistics*. This is an annual publication containing papers presented at the Interface Symposium. Publication began in 1967 under the title *Computer Science and Statistics: Proceedings of the xxtb Symposium on the Interface*. Since 1992, the shorter name has been used. It is published by the Interface Foundation of North America. (It is not refereed.)
- *Journal of Computational and Graphical Statistics*, published quarterly by the American Statistical Association.
- *Journal of the Japanese Society of Computational Statistics*, published once a year by the Japanese Society of Computational Statistics.
- *Journal of Statistical Computation and Simulation*, published irregularly in four numbers per volume by Gordon and Breach.
- *Proceedings of the Statistical Computing Section*, published annually by the American Statistical Association. (It is not refereed.)
- *Statistical Computing and Graphics Newsletter*, published quarterly by the Statistical Computing and the Statistical Graphics Sections of the American Statistical Association. (It is not refereed and it is not generally available in libraries.)
- *Statistics and Computing*, published quarterly by Chapman and Hall.

Education in Computational Statistics

Beginning in the late 1960s and early 1970s, most major academic programs in statistics offered one or more courses in statistical computing. The educational programs have promoted the identity of computational statistics as an academic discipline. Instruction in computational techniques has also permeated many of the standard courses in applied statistics.

Developments in computing and the changing role of computations in statistical work have had significant effects on the curricula of statistical education programs both at the graduate and undergraduate levels. Training in statistical computing is a major component in some academic programs in statistics (see Gentle, 2004; Lange, 2004; Monahan, 2004). In all academic programs, some amount of computing instruction is necessary if the student is expected to work as a statistician. The extent and the manner of integration of computing into an academic statistics program, of course, change with the developments in computing hardware and software and in computational statistics.

Numerical Analysis for Statistical Applications: Statistical Computing

The most obvious use of the computer is in support of the existing statistical methods. This does not just involve

some simple computer programming. While anyone can write a program to solve a well-conditioned system of linear equations, to write a robust program that can handle large-scale and ill-conditioned systems efficiently and to implement a smooth user interface, require a considerable amount of training in numerical analysis, as well as an intimate knowledge of the computer hardware and the programming language.

There are some general principles of computer arithmetic that pervade all types of computations. There are also general aspects of all numerical algorithms, but a good algorithm for a specific type of computation may involve some very special steps. An overview of the area of statistical computing is given in part II of the handbook edited by Gentle *et al.* (2004), which includes articles such as ‘random number generation’, ‘Markov chain Monte Carlo technology’, ‘numerical linear algebra’, ‘the grammar of graphics’, ‘object oriented computing’, and so on. Some of these topics are relevant to anyone developing numerical software, some topics are relevant to the careful user of statistical software, and some are important to the developer of software packages.

Three important areas of statistical computing are random number generation, numerical linear algebra, and optimization.

Computations to Support Simulation Methods

The role of Monte Carlo methods and simulation in all of the sciences has increased in importance during the past several years. The growing power of computers and the evolving simulation methodology have led to the recognition of computation as a third approach for advancing the natural sciences, together with theory and traditional experimentation. Monte Carlo is also a fundamental tool of computational statistics. At the kernel of a Monte Carlo or simulation method is random number generation.

Generation of random numbers is also at the heart of many standard statistical methods. The random sampling required in most analyses is usually done by the computer.

The computations required in Bayesian analysis have become viable because of Monte Carlo methods. This has led to much wider applications of Bayesian statistics, which, in turn, has led to the development of new Monte Carlo methods and to the refinement of existing procedures for random number generation.

Various methods for the generation of random numbers have been used. Sometimes, processes that are considered random, are used, but for Monte Carlo methods, which depend on millions of random numbers, a physical process as a source of random numbers is generally cumbersome. Instead of random numbers, most applications

use pseudorandom numbers, which are deterministic but look like they were generated randomly. While many of the earlier random number generators yielded sequences with obvious nonrandom patterns, there are now a number of reliable methods for the generation of sequences of pseudorandom numbers that simulate a uniform distribution over the unit interval (see Gentle, 2003). These are the basic sequences from which are derived pseudorandom numbers from other distributions, pseudorandom samples, and pseudostochastic processes.

Numerical Linear Algebra

Linear algebra is one of the most important mathematical and computational tools in the sciences. One of the most important tools in data analysis is the linear regression model. The main computations required for an analysis using this model are to solve an overdetermined system of linear equations, usually so as to minimize the sum of squares of the deviations. The simplest way these computations can be performed under the least-squares criterion is by first forming the normal equations and then solving them. This method is neither computationally efficient nor numerically stable. Some of the most important early work in statistical computing was the development of good algorithms to solve this problem. These algorithms begin decomposition of the matrix of values of the covariates (see Čížková and Čížek, 2004).

While linear models and linear transformations are important in their own right, the basic role that linear algebra plays in nonlinear models, in optimization, and in other areas of statistics, also makes an understanding of linear methods one of the most fundamental requirements for research in statistics or in the application of statistical methods.

Numerical Solution of Optimization Problems

Optimization problems – maximization or minimization – arise in many areas of statistics. Statistical estimation and modeling both are usually special types of optimization problems. In a common method of statistical estimation, we maximize a likelihood, which is a function proportional to a probability density at the point of the observed data. In another method of estimation and in standard modeling techniques, we minimize a norm of the residuals. The best fit of a model is often defined in terms of a minimum of a norm, such as least squares. The other uses of optimization in statistical applications occur prior to the collection of data, for example, when we design an experiment or a survey so as to minimize experimental or sampling errors.

When a statistical method is based on the solution of an optimization problem, to formulate that problem

unambiguously helps us both to understand the method and to decide whether the method is appropriate to the purposes for which it is applied.

Some of the simpler and more common optimization problems in statistics can be solved easily, often by solving a system of linear equations. Many other problems, however, do not have closed-form solutions, and the solutions must be approximated by iterative methods. There are a variety of such methods (see Ng *et al.*, 2004; Spall, 2004).

Computationally Intensive Methods of Statistics

One of the most important effects of the computer is the introduction of completely new statistical methods. These methods, as all statistical methods, must be well-grounded in the theory of probability and statistics. Because of the large number of numerical computations involved in the methods, the software to implement the methods must make use of the best results from the research in statistical computing. Finally, although the statistician doing the research developing the technique is interested in the method itself, that statistician must be able to develop a computer program to implement the method. Thus, statistical research workers have come a full circle to the tradition of Fisher and the Pearsons, who developed theory and methods but could also make effective use of the current computing technology.

An overview of this area of computational statistics is given in part III of the handbook edited by Gentle *et al.* (2004). In this part of the handbook, there are sixteen articles with titles such as ‘bootstrap and resampling’, ‘Bayesian computational methods’, ‘data and knowledge mining’, and so on.

Monte Carlo Methods in Statistics

Monte Carlo methods are experiments. Monte Carlo experimentation is the use of simulated random numbers to estimate some functions of a probability distribution. A problem that does not have a stochastic component sometimes may also be posed as a problem with a component that can be identified with an expectation of some function of a random variable. The problem is then solved by estimating the expected value by the use of a simulated sample from the distribution of the random variable.

Monte Carlo methods use random numbers, so to implement a Monte Carlo method, it is necessary to have a source of random numbers. As we mentioned above, there are a number of good methods for generating random numbers.

Monte Carlo methods are used in a variety of ways in statistics. They are widely used in the development of statistical methods, very often to compare methods or

modifications of methods. Monte Carlo methods can also be used directly in statistical inference, for example, in Monte Carlo tests, and in parametric bootstrap methods (see Gentle, 2002: Ch. 2).

Resampling Methods

Although subsampling, resampling, or otherwise rearranging a given dataset cannot increase its information content, these procedures can sometimes be useful in extracting information. Randomly rearranging the observed dataset, for example, can give an indication of how unusual the dataset is with respect to a null hypothesis. This idea leads to randomization tests.

There are many useful procedures for data analysis that involve partitioning the original sample. Using subsets of the full sample, we may be able to get an estimate of the bias or the variance of the standard estimator or test statistic without relying too heavily on the assumptions that led to that choice of estimator or test statistic. It is often useful to partition the dataset into two parts and use the data in the training set or estimation set to arrive at a preliminary estimate or fit and then use the data in the validation set or test set to evaluate the fit. This kind of approach is particularly appropriate when we are unsure of our model of the data-generating process. In actual applications, of course, we are always at least somewhat unsure of our model. If the full dataset is used to fit the model, we are very limited in the extent to which we can validate the model.

Subsets of the data can be formed systematically or they can be formed as random samples from the given dataset. Sometimes, the given dataset is viewed as a set of mass points of a finite distribution whose distribution function is the same as the empirical distribution function of the given dataset. In this case, the data partitioning may be done in such a way that observations may occur multiple times in the various partitions. One of the most widely used resampling methods is the bootstrap (see Mammen and Nandi, 2004). An early use of the bootstrap, as cited by Mammen and Nandi, was in educational research. The objective was to set a confidence interval for the correlation coefficient between scores on a standard test and the grade point averages of the same subjects. The bootstrap was an appropriate tool because neither of the two variables could be assumed to follow a normal distribution.

Machine Learning

It is now common to search through datasets and compute summary statistics from various items that may indicate relationships that were not previously recognized. The individual items or the relationships among them may not have been of primary interest when the data were originally

collected. This process is sometimes called data or knowledge mining, or machine learning, particularly if predictive classification is the objective (see Wilhelm, 2004).

The objective is to discover characteristics of the data that may not be expected based on the existing theory. In the language of the database literature, the specific goals of data mining are:

- classification of observations;
- linkage analysis;
- deviation detection; and finally
- predictive modeling.

Of course, the first three of these are the objectives of any exploratory statistical data analysis. Data mining is exploratory data analysis (EDA) applied to large datasets. An objective of an exploratory analysis is often to generate hypotheses, and exploratory analyses are generally followed by more formal confirmatory procedures. The explorations in massive datasets must be performed without much human intervention. Searching algorithms need to have some means of learning and adaptively improving. This will be a major area of research for some time.

Predictive modeling uses inductive reasoning rather than the more common deductive reasoning, which is much easier to automate.

In the statistical classification of observations, the dataset is partitioned recursively. The partitioning results in a classification tree, which is a decision tree, each node of which represents a partition of the dataset. (see Zhang (2004) for a discussion of these methods).

Linkage analysis is often the most important activity of data mining. In linkage analysis, relationships among different variables are discovered and analyzed. This step follows partitioning and is the interpretation of the partitions that were formed.

It is also important to identify data that do not fit the patterns that are discovered. The deviation of some subsets of the data often makes it difficult to develop models for the remainder of the data.

Conclusions

Computational statistics has become a major subdiscipline of statistics. The computational methods of statistical computing remain an important, but somewhat specialized area of statistics. Many data analysts need only a passing knowledge of the details of the computations. Software developers or researchers who must implement a new statistical method, however, need knowledge of the field of statistical computing. A broader use of statistical computing is in designing and performing Monte Carlo studies. A substantial portion of research articles in statistics and in fields that rely on statistical

methodology include Monte Carlo studies that assess the performance of alternative statistical methods in different settings. Such studies require computer programs to generate random numbers and then to loop through multiple analyses of simulated datasets. The authors of such studies need some grounding in the methods of statistical computing.

Computationally intensive statistical methods are widely used in data analysis. Many of these methods are intrinsically nonparametric or are less dependent on distributional assumptions. Resampling methods, for example, are often used in educational research because many of the variables of interest cannot be assumed to have normal distributions. The statistical learning methods of computational statistics have come to play a major role in mining data for unexpected discoveries. Discovery of the unexpected is central to the advancement of knowledge.

Bibliography

- Čížková, L. and Čížek, P. (2004). Numerical linear algebra. In Gentle, J., Härdle, W., and Mori, Y. (eds.) *Handbook of Computational Statistics. Concepts and Methods*, pp 103–136. Berlin: Springer.
- Gentle, J. (2002). *Elements of Computational Statistics*. New York: Springer.
- Gentle, J. (2003). *Random Number Generation and Monte Carlo Methods*, 2nd edn. New York: Springer.
- Gentle, J. E. (2004). Courses in statistical computing and computational statistics. *American Statistician* **58**, 2–5.
- Gentle, J., Härdle, W., and Mori, Y. (2004). *Handbook of Computational Statistics. Concepts and Methods*. Berlin: Springer.
- Grier, D. A. (1999). Statistical laboratories and the origins of statistical computing. *Chance* **4**(2), 14–20.
- Lange, K. (2004). Computational statistics and optimization theory at UCLA. *American Statistician* **58**, 9–11.
- Mammen, E. and Nandi, S. (2004). Bootstrap and resampling. In Gentle, J., Härdle, W., and Mori, Y. (eds.) *Handbook of Computational Statistics. Concepts and Methods*, pp 467–496. Berlin: Springer.
- Monahan, J. (2004). Teaching statistical computing at NC State. *American Statistician* **58**, 6–8.
- Ng, S. K., Krishnan, T., and McLachlan, J. M (2004). The EM algorithm. In Gentle, J., Härdle, W., and Mori, Y. (eds.) *Handbook of Computational Statistics. Concepts and Methods*, pp 137–168. Berlin: Springer.
- Spall, J. C. (2004). Stochastic optimization. In Gentle, J., Härdle, W., and Mori, Y. (eds.) *Handbook of Computational Statistics. Concepts and Methods*, pp 169–198. Berlin: Springer.
- Wilhelm, A. (2004). Data and knowledge mining. In Gentle, J., Härdle, W., and Mori, Y. (eds.) *Handbook of Computational Statistics. Concepts and Methods*, pp 787–812. Berlin: Springer.
- Zhang, H. (2004). Recursive partitioning and tree-based methods. In Gentle, J., Härdle, W., and Mori, Y. (eds.) *Handbook of Computational Statistics. Concepts and Methods*, pp 813–840. Berlin: Springer.

Continuous Probability Distributions

S Sinharay, ETS, Princeton, NJ, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Estimate – A function of sample data used for approximating a population parameter.

Moment – The n th moment about a value c of a random variable that has a probability density function $f(x)$ is

$$\mu'_n = \int_{-\infty}^{\infty} (x - c)^n f(x) dx.$$

Continuous probability distributions, which are distributions of continuous random variables, are amenable to more elegant mathematical treatment than are discrete probability distributions. The values of variables observed in statistical applications are invariably discrete, mainly due to rounding (e.g., though weight of a person is a continuous variable, it is measured in, for example, pounds) – so, technically, continuous distributions should not be applied. However, these distributions are convenient approximations, facilitating mathematical and statistical analysis (Johnson *et al.*, 1994). For example, think of a binomial random variable X with number of trials = 30 and success probability = 0.4. To compute the probability that $X \leq 10$ using binomial probability values, one has to sum several terms of the form $\binom{30}{x} 0.4^x 0.6^{30-x}$. However, if the investigator approximates the binomial distribution (a discrete distribution) by the normal distribution (a continuous distribution), the computation is much simpler as shown as (note that the computation with current fast computers will probably take a very short time for the binomial distribution, but the same computation was not done so quickly even a couple of decades ago)

$$\begin{aligned} P(X \leq 10) &= P(X \leq 10.5) \\ &= P\left(\frac{X - 30 \times 0.4}{\sqrt{30 \times 0.4 \times 0.6}} \leq \frac{10.5 - 30 \times 0.4}{\sqrt{30 \times 0.4 \times 0.6}}\right) = \Phi(-0.56) \end{aligned}$$

where $\Phi(\cdot)$ denotes the cumulative distribution of the standard normal distribution. The value obtained using the approximation of the binomial distribution by the normal distribution is 0.29, which is the same up to 2 decimal places as that obtained without the approximation. A normal approximation helps even more in computing a confidence interval for the binomial success probability (see, as an interesting reading on the issue, Holland (1979), who commented on the tyranny of continuous models in a world full of discrete data).

A key issue with a continuous random variable is that the probability is zero that it will take any specific value. However, there is usually a positive probability that such a variable will fall within an interval. While a discrete random variable has a probability mass function, a continuous random variable has a probability density function (pdf). The integral of the pdf over an interval is the probability that the continuous random variable will fall in that interval.

In the following sections, several continuous distributions are discussed. For each distribution, its genesis, pdf, moments, estimation of the moments, methods to generate values of a random variable following that distribution, and relationships to other distributions (if any) are discussed. The normal distribution, arguably the most frequently used continuous distribution, is covered elsewhere in this encyclopedia and is not covered here.

Beta Distribution

The beta distribution is used to model continuous random variables whose range is between 0 and 1. For example, in Bayesian analyses, the beta distribution is often used as a prior distribution of the parameter p (which is bounded between 0 and 1) of the binomial distribution (see, e.g., Novick and Jackson, 1974).

The pdf of the beta distribution is given by

$$f(x; a, b) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1}$$

where $0 \leq x \leq 1$, $a > 0$, and $b > 0$ are called shape parameters, and $B(a, b)$ is the beta function defined as

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

where $\Gamma(a)$ is the gamma function defined as $\Gamma(a) = \int_{t=0}^{\infty} t^{a-1} e^{-t} dt$.

The expectation of the beta distribution is $\frac{a}{a+b}$ and the variance is $\frac{ab}{(a+b)^2(a+b+1)}$. The parameters a and b are usually estimated by the method of moments, that is, by setting the above-mentioned mean and variance equal to the sample mean \bar{x} and sample variance s^2 and solving them for a and b . The solutions in this case are given by

$$\hat{a} = \bar{x} \left[\frac{\bar{x}(1-\bar{x})}{s^2 - 1} \right], \quad \hat{b} = (1-\bar{x}) \left[\frac{\bar{x}(1-\bar{x})}{s^2 - 1} \right].$$

It is also possible to estimate a and b using maximum likelihood, but the method requires an iterative procedure and hence is more computationally intensive.

Random numbers following the beta distribution can be generated in several ways. If V_1 and V_2 are independent and follow the χ^2 distribution (the section on the χ^2 distribution that follows will discuss generating random numbers following that distribution) with parameters c_1 and c_2 , respectively, then $\frac{V_1}{V_1 + V_2}$ follows a beta distribution with parameters $c_1/2$ and $c_2/2$. Alternatively, if Y_1, Y_2, \dots, Y_n are independent random variables following the $U(0, 1)$ (i.e., the continuous uniform distribution between 0 and 1) distribution, then the s th order statistic Y'_s follows a beta distribution with parameters s and $n - s + 1$.

A well-known application of the beta distribution (actually, that of a more general version of the distribution that has, in addition to the a and b parameters, two more parameters specifying the bounds of the distribution) in education can be found in Lord (1965), where the true test score was modeled using the beta distribution. In addition, in some applications of the Markov chain Monte Carlo (MCMC) algorithm to fit item response theory (IRT) models, the prior distribution for the guessing parameter is sometimes assumed to be the beta distribution.

Cauchy Distribution

The Cauchy distribution is often of mathematical interest due to the absence of any moments (Evans *et al.*, 2000). The pdf of the distribution is given by

$$f(x; a, b) = \left\{ \pi b \left[1 + \left(\frac{x - a}{b} \right)^2 \right] \right\}^{-1}$$

and the cumulative distribution function (cdf) is given by

$$\frac{1}{2} + \frac{1}{\pi} \tan^{-1} \left(\frac{x - a}{b} \right).$$

The mode and the median of the distribution is a , but no moments exist. The Cauchy distribution is unimodal and symmetric but has tails that are much heavier than those of the normal distribution. **Figure 1** compares the pdf of the Cauchy distribution with $a = 0$ and $b = 1$ against that of the standard normal distribution (the plot for the t distribution in the figure will be discussed later).

The ratio of two independent standard normal random variables follows a Cauchy distribution with $a = 0$ and $b = 1$. A Cauchy distribution with $a = 0$ and $b = 1$ is a special case of the Student's t distribution with one degree of freedom (df).

The sample median is a consistent estimate of the median of the Cauchy distribution. A random number following the Cauchy distribution can be generated by generating a random number R from the $U(0,1)$ distribution and then setting $F(x) = R$, where $F(x)$ is the cdf of the Cauchy distribution. The solution is given by $a + b \tan [\pi (R - 0.5)]$.

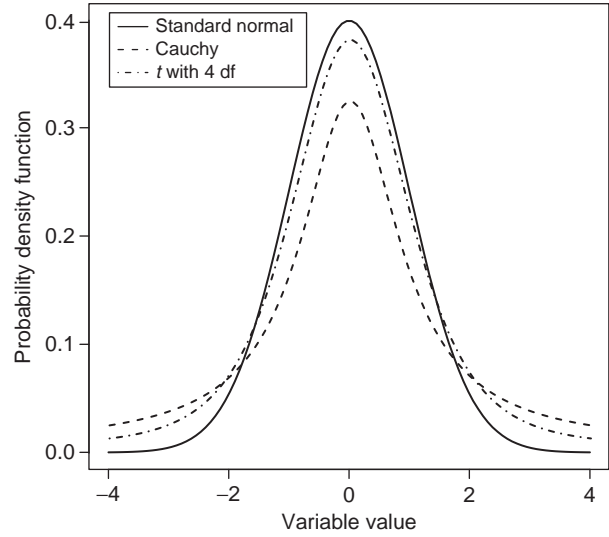


Figure 1 Plot of the pdfs of the standard normal, Cauchy and the Student's t distribution with 4 degree of freedom (df).

χ^2 Distribution

A common occurrence of the χ^2 distribution is that the sum of squares of several independent normal random variables follows a χ^2 distribution. As a consequence, several goodness-of-fit statistics follow the χ^2 distribution when the model fits the data.

A χ^2 random variable with df d and noncentrality parameter $\delta \geq 0$ has the pdf

$$f(x; d, \delta) = \frac{e^{-(x+\delta)/2}}{2^{d/2}} \sum_{j=0}^{\infty} \frac{x^{d/2+j-1} \delta^j}{\Gamma(d/2 + j) 2^{2j} j!}$$

The expectation and variance of the distribution are $d + \delta$ and $2(d + 2\delta)$, respectively. In most common applications of the distribution, the noncentrality parameter δ is 0. The distribution is then called a central χ^2 distribution, and the summation in the pdf reduces to a single term (that for $j = 0$), the mean is d and the variance is $2d$.

If d is an integer, a central χ^2 random variable with df d can be generated by generating d standard normal random numbers and adding the squares of them.

Dirichlet Distribution

The Dirichlet distribution is the multivariate generalization of the beta distribution. In Bayesian analyses, the Dirichlet distribution is often used as a prior distribution of the parameters of the multinomial distribution (see, e.g., Novick and Jackson, 1974: chapter 10–7). Consider a k -dimensional random variable $X = (X_1, X_2, \dots, X_k)$ so that the X_i 's are all positive and the sum of them is 1. If X follows the Dirichlet distribution, its pdf is given by

$$f(x_1, x_2, \dots, x_k; c_1, \dots, c_k) = \frac{\Gamma(\sum_{i=1}^k c_i)}{\prod_{i=1}^k \Gamma(c_i)} \prod_{i=1}^k x_i^{c_i-1}$$

where c_i , $i = 1, 2, \dots, k$ are the parameters of the distribution. The marginal distribution of X_i is the beta distribution with parameters c_i and $\sum_{j=1}^k c_j - c_i$. This property can be used to estimate the parameters c_i using the method of moments as discussed earlier for the beta distribution. For $k = 2$, the Dirichlet distribution becomes the beta distribution with parameters c_1 and c_2 .

To generate a random number $x = (x_1, x_2, \dots, x_k)$ from the Dirichlet distribution with c_i 's as parameters, one can generate a sequence of independent draws y_i , $i = 1, 2, \dots, k$, where y_i is a draw from the gamma distribution (to be discussed later) with parameters $b = 1$ and $c = c_i$ and compute $x_i = \frac{y_i}{\sum_{j=1}^k y_j}$, $i = 1, 2, \dots, k$.

Exponential Distribution

The exponential distribution has been employed to model variables such as the time to decay of radioactive atoms and time to failure of components with constant failure rates. The distribution is a special case of the gamma distribution that is discussed later.

The pdf of the exponential distribution is given by

$$f(x; b) = \frac{1}{b} e^{-x/b}.$$

The mean of the distribution is b and the variance is b^2 . The maximum likelihood estimate of b is the sample mean.

A random number from the exponential distribution can be generated by generating a $U(0, 1)$ random number R and then computing $-b \log(R)$.

Exponential Family of Distributions (Continuous Version)

The exponential family of distributions have pdf of the form

$$\exp \left[\sum_i A_i(\theta) B_i(x) + C(x) + D(\theta) \right]$$

where $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ is the p -dimensional parameter vector of the distribution. Several continuous distributions, such as the beta, gamma, exponential, and normal distributions, are members of the exponential family of distributions. (Uniform and Cauchy distributions are examples of distributions that do not belong to the exponential family.)

The exponential family of distributions have several attractive properties, such as the set of $B_i(x)$'s being jointly

sufficient for the parameters θ . Hence, the family has been applied to several applications in education (see, e.g., Haberman (2008), for an application of the family of distributions to equating).

F Distribution

The pdf of the F distribution is given by

$$f(x; v, w, \delta) = k \frac{e^{-\delta/2} v^{v/2} w^{w/2} x^{(v-2)/2}}{B(v/2, w/2)(w + vx)^{(v+w)/2}},$$

where v and w are df's and positive integers, $\delta \geq 0$ is the noncentrality parameter, and

$$k = 1 + \sum_{j=1}^{\infty} \left(\frac{vwx/2}{w + vx} \right)^j \frac{(v+w)(v+w+2)(v+w+2j-2)}{j!v(v+2)\dots(v+2j-2)}.$$

Often, $\delta = 0$, in which case $k = 1$ and the F distribution is called the central F distribution.

The mean and variance of the central F distribution are $\frac{w}{w-2}$ and $\frac{2w^2(v+w-2)}{v(w-2)^2(w-4)}$.

If two independent χ^2 random variables χ_1^2 and χ_2^2 have df's v and w then $\frac{\chi_1^2/v}{\chi_2^2/w}$ has an F distribution with v and w df. This result can be used to generate random numbers following the F distribution.

The F distribution has often been applied in education in analysis of variance, where the ratio of the mean square of an effect and the mean square of error has an F distribution and a central F distribution under the null hypothesis that the effect is not significant.

Gamma Distribution

The gamma distribution is used to model continuous random variables whose range is between 0 and ∞ . The pdf of the gamma distribution is given by

$$f(x; b, c) = \frac{1}{b\Gamma(c)} \left(\frac{x}{b} \right)^{c-1} e^{-x/b}, \quad b \geq 0, c > 0$$

where b is called the scale parameter and c is the shape parameter. In many applications of the distribution, b is assumed to be equal to 1. The mean of the gamma distribution is bc and the variance is b^2c .

If $2c$ is an integer, then the gamma random variable is half of a χ^2 random variable with $2c$ df. If $c = 1$, then the gamma distribution becomes the exponential distribution. If V_1 and V_2 are gamma random variables, with $b = 1$ and $c = c_1$ and c_2 , respectively, then $\frac{V_1}{V_1 + V_2}$ follows a beta distribution with parameters c_1 and c_2 .

By the method of moments, the estimate of b is $\frac{\bar{x}^2}{\bar{x}}$ and that of c is $\left(\frac{\bar{x}}{\bar{x}^2} \right)^2$. It is possible to estimate the parameters

using the maximum likelihood method, but that requires the use of an iterative algorithm.

If c is an integer, then $-b \prod_{i=1}^c R_i$ follows a gamma distribution with parameters b and c , where the R_i 's are independent $U(0, 1)$ random variables – this result can be used to generate random numbers following the gamma distribution.

The reciprocal of a gamma random variable follows a distribution called the inverse gamma distribution. In several applications of Bayesian methods to IRT models, the prior distribution for the variance parameters is often assumed to be the inverse gamma distribution (see, e.g., Bradlow *et al.*, 1999).

Lognormal Distribution

If a random variable V has a normal distribution with mean μ and variance σ^2 , then e^V has a lognormal distribution with parameters μ and σ^2 . In other words, if a variable has a lognormal distribution, then its logarithm has a normal distribution. The pdf of the distribution is given by

$$f(x; \mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2}$$

where x and σ are both positive. If X follows a lognormal distribution with parameters μ and σ^2 , then $Y = e^a X^b$ follows a lognormal distribution with parameters $a + b\mu$ and $b^2\sigma^2$.

The expectation and variance of the lognormal distribution are given by

$$E(x) = e^{\mu + \frac{\sigma^2}{2}}, \quad V(x) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1).$$

Using its relationships to the normal distribution, the parameters of the lognormal distribution from a sample x_1, x_2, \dots, x_n of draws from the distribution can be estimated as

$$\hat{\mu} = \frac{1}{n} \sum \log(x_i), \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum (\log x_i - \hat{\mu})^2.$$

The easiest way to generate random numbers from a lognormal distribution with parameters μ and σ^2 is to generate random numbers from a normal distribution with mean μ and variance σ^2 and then exponentiate them.

In some applications of Bayesian methods to IRT models, the prior distribution on the slope parameters is sometimes assumed to be a lognormal distribution. For example, the PARSCALE software program, which is used to fit IRT models by several operational testing programs, assumes a lognormal distribution as the prior distribution for the slope parameters (see e.g., du Toit, 2003).

Student's t Distribution

The t distribution is often applied in testing the hypothesis for normally distributed data. For example, for testing the hypothesis that the mean of a normal distribution is 0, the standard test statistic used has a t distribution under the null hypothesis.

The pdf of the distribution is given by

$$\frac{v^{v/2} e^{-\delta^2/2}}{\Gamma(v/2) \pi^{0.5} (v + x^2)^{(v+1)/2}} \sum_{i=0}^{\infty} \Gamma\left(\frac{v+i+1}{2}\right) \frac{(x\delta)^i}{i!} \left(\frac{2}{v+x^2}\right)^{i/2}$$

where v is the df and δ is the noncentrality parameter.

The mean of the distribution is $\frac{\delta(v/2)^{1/2} \Gamma((v-1)/2)}{\Gamma(v/2)}$ and the variance is given by

$$\frac{v}{v-2} (1 + \delta^2) - \frac{v}{2} \delta^2 \left(\frac{\Gamma((v-1)/2)}{\Gamma(v/2)} \right)^2.$$

If the noncentrality parameter is 0, then the distribution is called a central t distribution that has mean 0 and variance $\frac{v}{v-2}$.

The t distribution has a much heavier tail than the normal distribution for small df. **Figure 1** shows a comparison of the pdf of the Cauchy distribution with $a = 0$ and $b = 1$ (that is a special case of the Student's t distribution with 1 df) and that of the Student's t distribution with 4 df against that of the standard normal distribution. The t distribution with 1 df (or the Cauchy distribution) has the heaviest tail followed by that with 4 df.

For large df, the central t distribution becomes close to the standard normal distribution. The square of a central t distribution is the distribution of the ratio of two independent χ^2 random variables, the first having 1 df (note that a χ^2 random variable can be expressed as the square of a standard normal random variable) and the other having v df. It also has an F distribution with df's 1 and v .

To generate a random number following the central t distribution with v df, one can compute the ratio of a standard normal random number and the square root of Y/v , where Y is a random number following the χ^2 distribution with v df and both of these random numbers are generated independently of each other. (Note that Y can be generated as the sum of squares of v independent standard normal random numbers.)

Uniform Distribution

The uniform or rectangular distribution is the distribution that uniform random numbers follow. In addition, this distribution is widely used as the basis for the generation of random numbers from other continuous statistical distributions (Evans *et al.*, 2000).

The uniform distribution with range (a, b) , often denoted as $U(a, b)$, has the pdf $\frac{1}{b-a}$ if $a \leq x \leq b$ and 0 otherwise.

The cdf is 0 if $x \leq a$, $\frac{x-a}{b-a}$ if $a \leq x \leq b$, and 1 if $x \geq b$. When $a = 0$ and $b = 1$, the uniform distribution becomes the standard or unit uniform distribution.

The mean of the distribution is $\frac{a+b}{2}$ and the variance is $\frac{(b-a)^2}{12}$.

For any continuous random variable Y , its cdf $F(y)$ follows a $U(0,1)$ distribution and $F^{-1}(R)$ has the cdf F for R following the $U(0,1)$ distribution. This property is used to generate random numbers from other continuous distributions starting from standard uniform random numbers.

The method of moments estimates of a and b are $\bar{x} - 3^{1/2}s$ and $\bar{x} + 3^{1/2}s$, respectively. To generate a random number x following a $U(a, b)$ distribution, one has to generate a $U(0,1)$ random number R , and then use the transformation $x = a + (b - a)R$.

Wishart Distribution

The Wishart distribution is the multivariate generalization of the χ^2 random variable. It is the probability distribution of the maximum-likelihood estimator (MLE) of the covariance matrix of a multivariate normal distribution. A k -dimensional random variable X following the Wishart distribution has a pdf proportional to

$$e^{\left[-\frac{1}{2}\text{tr}(\Sigma^{-1}X)\right]} |X|^{\frac{n-k-1}{2}}$$

where Σ is a $k \times k$ matrix, n is the df, $\text{tr}(A)$ is the trace of matrix A , and $|A|$ is the determinant of the matrix A . The distribution has a mean of $n\Sigma$.

If Y_1, Y_2, \dots, Y_n are n k -dimensional random numbers independently drawn from a multivariate normal distribution with mean vector μ and variance matrix Σ , then the distribution of $\sum_{i=1}^n Y_i Y_i'$ is k -dimensional Wishart with k df and parameter Σ . This result can be used to generate random matrices following the Wishart distribution.

The inverse of a random variable following a Wishart distribution follows the inverse Wishart distribution. The inverse Wishart distribution is often assumed to be the prior distribution on the variance matrix in a Bayesian approach to IRT modeling (Johnson and Sinharay, 2005).

Conclusions

Among the numerous continuous statistical distributions, this article provides a brief overview of a few distributions

that have found applications in the field of education (see, e.g., Evans *et al.*, 2000; Johnson *et al.*, 1994; and Kotz *et al.*, 2000, for more detailed descriptions of these and many more continuous distributions). Though an observed random variable is never continuous in the true sense of the term (due to rounding), the continuous random variables describe the data in a satisfactory manner in several statistical applications. The choice of the specific distribution to be used in a specific application depends on several issues. If a specific distribution is used in practice for data of the type an investigator has (e.g., a Dirichlet distribution is often applied in practice to model the probabilities from a multinomial distribution), that distribution should be the first choice. If such knowledge is lacking, a sensible approach is to graphically plot the data and try to find a distribution that is expected to describe the data adequately. After fitting a distribution to a data set, methods of goodness-of-fit can be used to examine whether the distribution provides an adequate description of the data set.

See also: Discrete Probability Distributions; Item Response Theory; Probability Theory.

Bibliography

- Bradlow, E. T., Wainer, H., and Wang, X. (1999). A bayesian random effects model for testlets. *Psychometrika* **64**, 153–168.
- du Toit, M. (2003). *Irt from ssi*. Lincoln, IL: Scientific Software International.
- Evans, M., Hastings, N., and Peacock, B. (2000). *Statistical Distributions*. New York: Wiley.
- Haberman, S. J. (2008). Continuous exponential families: An equating tool. *ETS Research Report No. RR-08-05*. Princeton, NJ: ETS.
- Holland, P. W. (1979). The tyranny of continuous models in a world of discrete data. *IHS-Journal* **3**, 29–42.
- Johnson, M. S. and Sinharay, S. (2005). Calibration of polytomous item families using bayesian hierarchical modeling. *Applied Psychological Measurement* **29**, 369–400.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*. New York: Wiley.
- Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). *Continuous Multivariate Distributions*, vol. 1. New York: Wiley.
- Lord, F. M. (1965). A strong true score theory, with applications. *Psychometrika* **30**, 239–270.
- Novick, M. R. and Jackson, P. H. (1974). *Statistical Methods for Educational and Psychological Research*. New York: McGraw-Hill.

Correspondence Analysis

M Greenacre, Universitat Pompeu Fabra, Barcelona, Spain

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Chi-square distance – The weighted Euclidean distance measure between row (or column) profiles of a table, where each squared difference between profile elements is divided by the corresponding element of the average profile.

Contingency table – A cross-tabulation of a set of objects according to two categorical variables; hence, the grand total of the table is the number of objects.

Correspondence analysis – A method of displaying the rows and columns of a table as points in a spatial map, with a specific geometric interpretation of the positions of the points in terms of similarities and differences between rows, the similarities and differences between columns, and the association between rows and columns.

Inertia – The weighted sum of squared distances of a set of points to their weighted average; a measure of variance of a set of multidimensional points, each weighted by its mass.

Mass – The marginal total of a row or a column of a table, divided by the grand total of the table; used as weights in correspondence analysis.

Profile – A row or a column of a table divided by its total; the profiles are the points visualized in correspondence analysis.

Correspondence analysis (CA) is a variant of principal component analysis (PCA) aimed primarily at categorical data, for example, aggregate count data in contingency tables or individual-level responses in questionnaire surveys. Emphasis is placed on visual maps that result from the method, displaying the rows and columns as points in a joint map in which distances and relative positions of points have a specific interpretation. Since most data in the social sciences are categorical by nature, CA is a valuable tool for exploring relationships between variables and comparing individuals or groups of individuals, always in an intuitive graphical format. The theory of CA has a long history (for historical details, see Nishisato, 2006), while the graphical context and interpretation are due to Benzéeri (1973). Benzéeri's contribution was inspired by Karl Pearson's geometric approach to PCA (Pearson, 1901) and Louis Guttman's work on psychometric scaling (Guttman, 1941).

Two forms of CA dominate. First, simple CA is the analysis of a table of frequencies, usually a cross-tabulation of two categorical variables. Second, multiple correspondence analysis (MCA) is the analysis of more than two categorical variables, and can be interpreted as a form of PCA of nominal categorical data. Another way of distinguishing the two approaches is that CA analyzes relationships between two variables, whereas MCA analyzes relationships within a larger set of variables. In fact, both approaches use the same algorithm for their solution, differing only in the coding of the data fed to the algorithm. After an introduction to the basic concepts of simple CA, using a small table, examples of these two forms of the method are given.

Simple CA

The following table summarizes the responses of 315 people in the USA, classified into five education groups, to the question: "How much confidence do you have in your country's education system?" Possible answers were "A great deal," "Quite a lot," "Not much," and "None at all", and in the table below, the last two categories have been combined so that the table has three columns.

	<i>Great deal</i>	<i>Quite a lot</i>	<i>Not much/ none at all</i>	<i>Sums</i>
E1-some primary	3	10	6	19
E2-primary completed	14	47	23	84
E3-some secondary	33	31	20	84
E4-secondary completed	43	44	13	100
E5-some (or completed) tertiary	17	9	2	28
Sums	110	141	64	315

The general idea is to define a set of row or column points in multidimensional space, decide on a distance function between the points and whether the points should be differentially weighted, and then, for interpretation, project the points onto a low-dimensional subspace which comes closest to all the points. In CA, the vectors of relative frequencies, or profiles, of the rows (or of the

columns) define the points: that is, the rows (or columns) divided by their respective sums. The following explanation will be in terms of the row profiles, but exactly the same argument applies to the columns. The profiles of the rows, including the profile of the column sums, labeled average, are given next:

	<i>Great deal</i>	<i>Quite a lot</i>	<i>Not much/none at all</i>
E1-some primary	0.158	0.526	0.316
E2-primary completed	0.167	0.560	0.274
E3-some secondary	0.393	0.369	0.238
E4-secondary completed	0.430	0.440	0.130
E5-some (or completed) tertiary	0.607	0.321	0.071
Average	0.349	0.448	0.203

(e.g., $0.158 = 3/19$ (first cell, first row), and $0.349 = 110/315$ (first cell, last row) – we shall explain below in what sense the last row is an average of the elements in the table.) As the elements of each row profile add up to a constant 1, these three-dimensional profiles are points in a plane; in fact, they are points lying in a two-dimensional triangle joining the unit points for the three categories, namely $[1\ 0\ 0]$, $[0\ 1\ 0]$, and $[0\ 0\ 1]$. This is the triangular, or barycentric, coordinate system, shown in **Figure 1**.

A convenient way of thinking about barycentric coordinates is that each profile is situated at the weighted average, or centroid, of the corners of the triangle, each of which represents the purest or most extreme group of respondents (e.g., corner 1 with values $[1\ 0\ 0]$, corresponds to 100% response in the first category, a great deal). Thus, education group E1, for example, with profile values 0.158, 0.526, and 0.316, lies further from the first corner (a great deal), but closer to the second (quite a lot) because its position is at $(0.158 \times \text{corner 1}) + (0.526 \times \text{corner 2}) + (0.316 \times \text{corner 3})$.

There are two other important concepts in CA: the way profiles are weighted and the measure of distance between profiles. The original row sums of the table (19, 84, etc.) are used as weights, so that each education group is weighted proportional to its subgroup size. In fact, the weights are defined by the relative sums: $0.060 = 19/315$, $0.267 = 84/315$, etc., which are called the (row) masses in CA – the masses, like the profiles, add up to 1. The average profile described above is, thus, the centroid of the row profiles, where each row profile is weighted by its respective mass. As far as the inter-profile distance is concerned, a standardization of the column categories is used which is appropriate for frequency data, with the spin-off that the spatial dispersion of the points will be related to the chi-square statistic for the contingency table

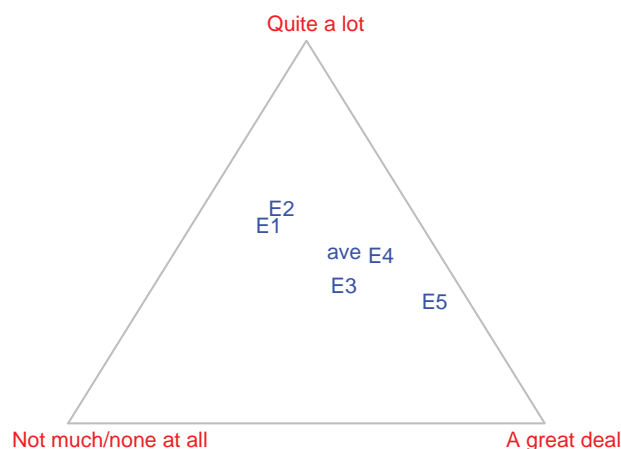


Figure 1 The education profiles and the average profile shown as points in a triangular coordinate system. The vertices of the triangle are the unit points in the underlying three-dimensional Euclidean space.

(see below). In contrast to PCA, where standardization is usually performed by dividing by the square root of the variance of each variable (appropriate for interval-scale data), in CA, the relative frequencies for a category are divided by the square root of the average. For example, in the table above, the values in the first column (0.158, 0.167, etc.) are divided by the square root of 0.349, which is 0.591, the second column is divided by $\sqrt{0.448} = 0.669$, and the last column by $\sqrt{0.203} = 0.451$. With this standardization, the categories of lower average frequency (and usually lower variance) have their profile coordinates divided by smaller values and increase in value more than those associated with categories with higher frequency. Distances between profiles are calculated using the usual Euclidean distance computed on their standardized values – the resulting distance is called the chi-square distance. Geometrically, this changes the equilateral triangle space of **Figure 1** to a triangle with unequal sides – all the corners are stretched out, where the stretching is more for categories of lower frequency.

With this definition of row profiles, row masses, and inter-row distances, the weighted sum of squared distances of the points to their average is exactly the chi-square statistic for the table divided by the grand total of the table, a quantity called the inertia in CA. We, thus, have a multidimensional geometric picture of the row–column association as a cloud of points in space: the further they are away from their average, the higher the inertia, and thus, the chi-square. It is the inertia that is explained in CA, just like the total variance is explained in PCA.

Generalizing this geometry to higher dimensions is straightforward: for four columns, the chi-squared space of the profiles is an irregular tetrahedron in three dimensions (in the application below, we shall give an animated three-dimensional graphical display to show this geometry).

In general, profiles with p elements lie in a $p - 1$ dimensional space. In this initial example, which lies exactly in two dimensions, no dimension reduction is necessary and CA simply rotates the display to principal axes (**Figure 2**). Here, we can see that the triangle formed by the response categories is no longer equilateral, which as described above, is due to the standardization to induce the chi-square distance between profiles.

Computation of the principal axes in order to achieve dimension reduction uses an eigenvalue/eigenvector routine, as in PCA, or more elegantly using the singular-value decomposition for rectangular matrices. These algorithms identify optimal maps, that is, spaces of lower dimensionality, usually two-dimensional planes, which best fit the points. CA and PCA do these in similar ways, by least squares, the only difference being in the definition of distance and the fact that in CA, the points are weighted, which varies the importance of each profile point in the search for a low-dimensional solution – hence, higher mass points are more influential than lower mass points in determining the solution. An equivalent way to think of dimension reduction is reminiscent of the objective of PCA to capture the most variance along the principal axes. In CA, it is the inertia which is being explained optimally along the principal axes. The format of the

results of PCA and CA is essentially the same, for example, there are inertias and percentages of inertia explained on each principal axis, as in **Figure 2**. A special property of CA is that the theory explained above in terms of row profiles is identical for the column profiles, which have the same inertia and the same decomposition along principal axes.

The coordinates of profiles projected onto principal axes are called principal coordinates (e.g., the coordinates of the education groups in **Figure 2**). The coordinates of the unit points, or vertices, projected onto principal axes, are called standard coordinates (e.g., the response category points in **Figure 2**). The map of **Figure 2** is called an asymmetric map of the rows (or row principal map) because it shows the rows in principal coordinates and columns in standard coordinates. In an asymmetric map, the barycentric property again holds: for example, in **Figure 2**, every profile lies at the weighted average of the response categories, where the weights are the profile elements. Thus, education group E5, with profile elements 0.607, 0.321, and 0.071, lies closest to the response category “a great deal” because of its largest weight of response on that category. In examples where the inertia is low, that is, where there are small differences between the profiles, the profiles tend to bunch up around their mean,

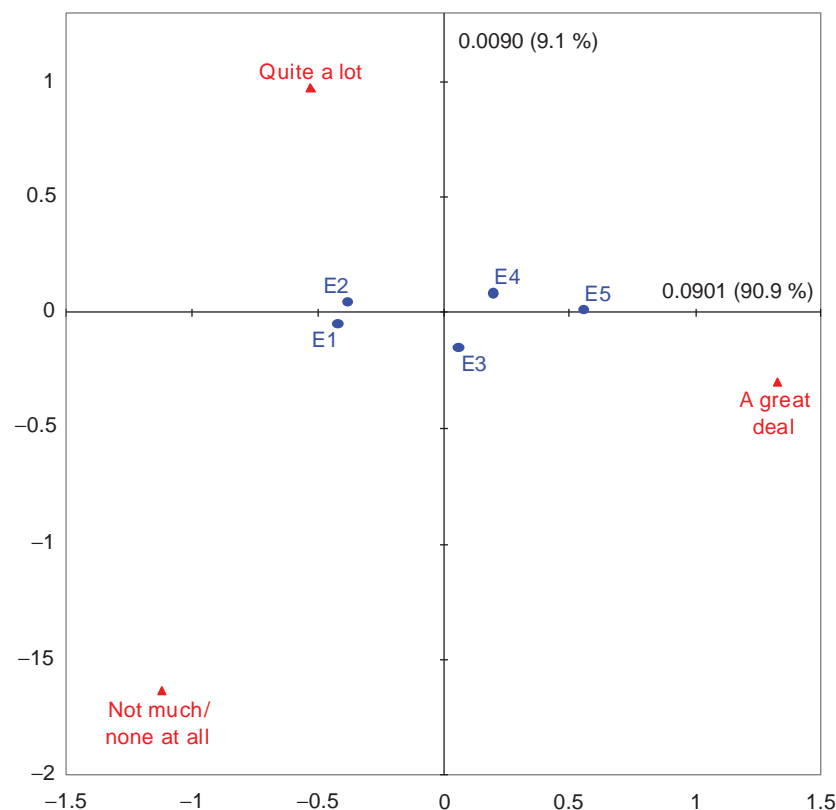


Figure 2 The CA solution corresponding to **Figure 1**, where the profiles have been standardized to give chi-square distances (hence the stretched triangle of the category vertices), and then rotated to principal axes. The inertias along principal axes and their percentages are shown – 100% of the inertia is explained because these data are exactly two-dimensional.

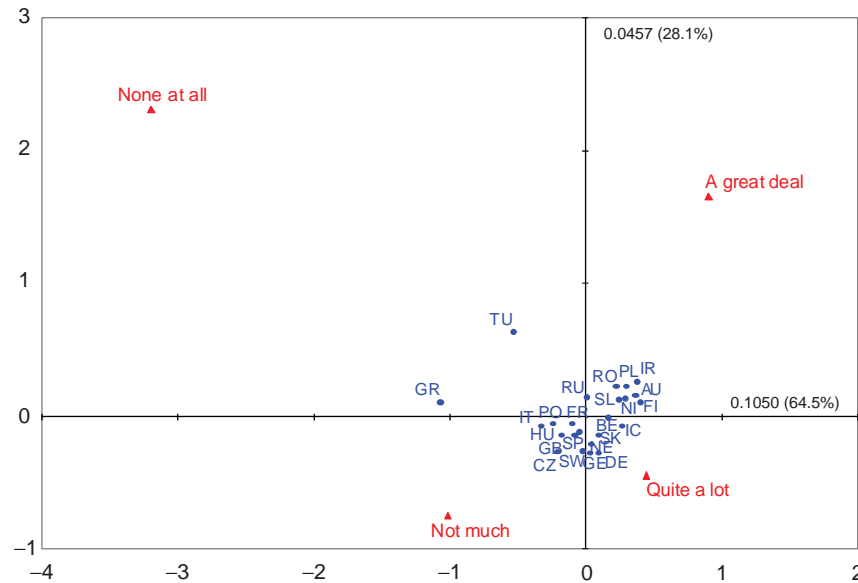


Figure 3 Asymmetric CA map of **Table 1**, showing the country profiles projected onto the plane of the first two principal axes (i.e., principal coordinates), as well as the extreme corners of the profile space (i.e., standard coordinates). The total inertia is 0.1628, of which 92.6% is represented in the map.

and the asymmetric map is impractical (e.g., see the application in **Figure 3**). In this situation, either the profiles need to be magnified in scale or – thanks to the symmetry of the row and column problems – a symmetric map is often reported where both rows and columns are in principal coordinates, thus, having the same dispersions along principal axes.

Application of Simple Correspondence Analysis

Table 1 gives the percentage responses to the question on the education system by samples of respondents from 24 European countries, with the inclusion of Russia and Turkey (data from the World Values Survey). We do not analyze raw frequencies in this case because the original sample sizes in each country do not reflect a valid weighting system for the countries – using percentages means that each country is equally weighted in the analysis, which is preferable.

Two versions of the two-dimensional CA map are shown in **Figures 3** and **4** respectively. In the asymmetric map of **Figure 3**, the four vertices representing the response categories are shown projected onto the optimal map of the countries. The bunching up of most of the country points near the average is clearly seen, and it is easier to distinguish them in the symmetric map of **Figure 4** where the response category points are represented in principal coordinates (these are the profiles of the columns). The response category “none at all” at upper left is well separated from the other categories because of Turkey and Greece, which have very high proportions of this response.

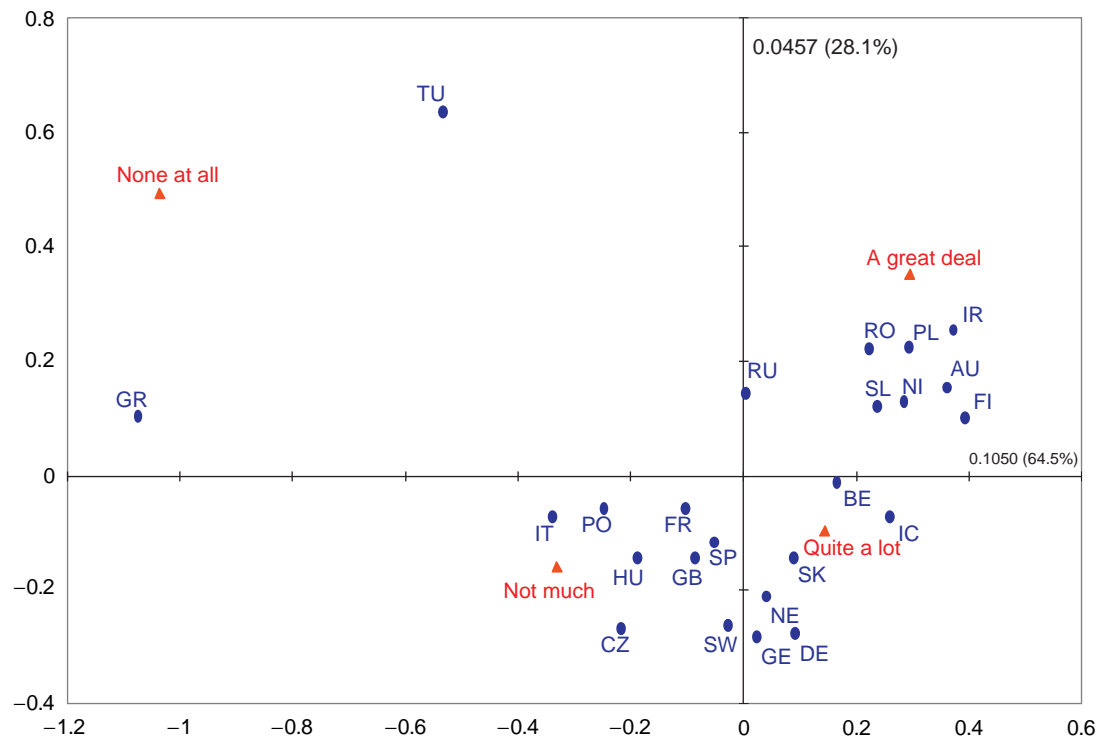
Among the other countries there is a gradient from “not much” (lower left) to “a great deal” (upper right), with Ireland, Poland, and Romania in the most favorable positions and Czech Republic in the least favorable position. Looking at Greece and Turkey again, notice that while both have the highest frequencies of the “none-at-all” response, Turkey, being higher up in the map, must have a relatively high number of the response “a great deal,” indicating a society split on this issue (i.e., more than average on both extremes of the response scale), while Greece is lower down on the left, indicating a society that is consistently lacking in confidence in its education system. These deductions from the map can be verified in the data of **Table 1**.

Multiple Correspondence Analysis

MCA deals with a set of more than two categorical variables, generally variables on similar scales, where interest is focused on inter-variable associations within the set. MCA can be defined in many equivalent ways, and here, we give the definition in terms of the indicator matrix, which is the respondents-by-categories matrix of dummy variables. Each categorical variable generates as many zero/one dummy variables as response categories: these variables form the columns of the indicator matrix. If there are Q variables in a questionnaire, for example, each row of the matrix will consist of zeros, except for Q ones, indicating the set of response categories for that respondent. MCA is the application of CA to this indicator matrix.

Table 1 Percentage responses by citizens of 24 countries to the question: “How much confidence do you have in your education system?”

		<i>Confidence in education</i>			
<i>Country</i>		<i>A Great deal</i>	<i>Quite a lot</i>	<i>Not much</i>	<i>None at all</i>
Austria	AU	29.38	56.90	12.86	0.86
Belgium	BE	18.78	61.27	16.61	3.33
Czech Rep.	CZ	11.87	42.65	43.24	2.24
Denmark	DE	9.76	65.19	23.64	1.41
Finland	FI	26.85	61.96	10.60	0.58
France	FR	13.23	55.24	24.76	6.77
Germany	GE	8.20	64.39	24.90	2.52
Greece	GR	3.62	25.22	48.59	22.57
Hungary	HU	9.34	54.98	28.65	7.03
Iceland	IC	19.35	62.96	16.96	0.73
Ireland	IR	33.50	52.94	12.26	1.30
Italy	IT	13.78	39.42	39.37	7.43
Netherlands	NE	11.35	61.45	24.60	2.61
Poland	PL	32.71	48.55	17.43	1.31
Portugal	PO	13.39	46.44	32.64	7.53
Romania	RO	30.55	48.88	17.70	2.88
Russia	RU	23.65	47.56	22.90	5.88
Slovakia	SK	13.33	62.97	20.59	3.11
Slovenia	SL	26.93	53.37	17.69	2.01
Spain	SP	14.51	53.11	28.32	4.06
Switzerland	SW	10.38	57.39	30.14	2.10
Turkey	TU	26.07	31.29	20.86	21.78
Great Britain	GB	12.83	53.49	29.26	4.41
N. Ireland	NI	27.62	55.13	15.71	1.54
Average		18.37	52.61	24.18	4.83

**Figure 4** Symmetric CA map of Table 1, showing the country profiles as well as the response category profiles projected onto the plane of the first two principal axes (both rows and columns in principal coordinates).

Application of Multiple Correspondence Analysis

To illustrate MCA, we consider the responses in the Spanish sample to three questions: confidence in the education system, as before, confidence in the social security system, and confidence in the healthcare system. The sample consists of 1181 respondents, and each of the three variables has four response categories (the levels of confidence, now labeled 1 – a great deal, to 4 – none at all) as well as a missing value category (labeled 9), thus, five categories for each of the three variables. As the data are analyzed at a nominal level, the missing response can be included simply as an additional category. Thus, in this application, there are 15 categories in total, giving an indicator matrix with 15 columns. Part of the indicator matrix is shown in the upper part of **Table 2**. The best two-dimensional map is given in **Figure 5**, showing the categories in standard coordinates (once again as vertices of an irregular simplex) and the respondents in principal coordinates labeled by their response combination; for example, 439 represents those respondents with confidence none at all for education, not much for social security, and missing response for healthcare). Since each row of the indicator matrix has a profile with elements 0 apart from three positive values of $1/3$ (i.e., in general, Q values of $1/Q$, where there are Q questions), the barycentric property implies that each individual lies at the ordinary average of his or her response categories. For example, the point 439 in **Figure 5** is at the average position of the categories E-4, S-3, and H-9. Of the $5^3 = 125$ possible response combinations, there are 78 observed in this data set, some being much more frequent than others; hence, the labels in **Figure 5** represent varying numbers of respondents piling up at each of the points. For example, the response sets giving the same responses to all questions are 111 – 61 respondents, 222 – 314, 333 – 123, 444 – 16, and 999 – 8. These response sets result in the strong associations between the response categories that have the same wording, clearly visible in **Figure 5**. Otherwise, we notice a typical arch effect inside the simplex, showing the responses in their ordinal sequence, from 1 (upper left) to 2 and 3 (lower down), and 4 (upper right). Individuals inside the arch (e.g., groups with responses such as 114, 141, and 414) have a mixture of the extreme responses; for example, 114 represents respondents with very high confidence in both education and social security, but none at all in healthcare.

Supplementary Points

An important concept in CA and MCA is that of a supplementary point, that is a data vector which contains external explanatory information and which can be represented in

a map to aid the interpretation. Since points are weighted by their masses, the best way to think about supplementary, or passive, points is that they are in the original data set but have zero mass, hence not influencing the solution but projected onto the map determined by the active points that do have positive mass. For example, in **Table 2**, additional rows have been added, which are the cross-tabulations of gender (two categories) and age (six categories) with the three questions. The profiles of these rows are exactly the averages of the respondents corresponding to the respective categories, and can be projected onto the same map, as shown in **Figure 6** (because they lie close to the origin, a magnification of their positions is shown). For example, the point a1 corresponding to the youngest age group lies exactly at the mean position of the subgroup of respondent points in this age group. We can see, for example, that the three first age groups are together toward the side of the map that is less confident, while the trend in the older groups is toward being more confident. Whether these conclusions are significant or not can be judged by regular statistical tests (e.g., univariate or multivariate analysis of variance, or permutations tests).

Contributions

Apart from the visualizations afforded by CA, there are various numerical diagnostics that assist with the interpretation of the results. These are all thanks to the decomposition of the total inertia, which is a weighted sum of squares, into parts due to the rows, or due to the columns, or due to the dimensions of the solution. Each row (or column) makes a part contribution to the principal inertia of each dimension: these contributions, usually expressed as proportions, provide information about which rows (or columns) have played the most important role in determining the dimension. Conversely, each dimension contributes to the inertia of a row (or column): these contributions, again expressed in relative amounts, provide information about the quality of display of each row (or column) in the space of the solution, for example, it can show easily which points are not well represented in the map.

Data Coding

An important feature of the CA approach is the wealth of coding schemes which allow different data types to be transformed so that CA is suitable as a visualization method. Here are two examples of these:

1. *Doubling*. For ratings, rankings, and paired comparisons, each variable engenders two recoded variables that can be thought of as positive and negative poles. For example,

Table 2 Data for MCA of individual response data, as well as aggregated counts for gender and age groups; all categorical data are coded as dummy variables, while the biographical categories are supplementary row points which will be at the averages of the respondents in their respective categories

	<i>E1</i>	<i>E2</i>	<i>E3</i>	<i>E4</i>	<i>E9</i>	<i>S1</i>	<i>S2</i>	<i>S3</i>	<i>S4</i>	<i>S9</i>	<i>H1</i>	<i>H2</i>	<i>H3</i>	<i>H4</i>	<i>H9</i>	<i>m</i>	<i>f</i>	<i>a1</i>	<i>a2</i>	<i>a3</i>	<i>a4</i>	<i>a5</i>	<i>a6</i>
resp#1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0
resp#2	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	1
resp#3	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
<i>m</i>	66	284	176	31	18	56	288	182	40	9	68	281	182	33	11								
<i>f</i>	99	322	146	16	23	90	290	166	39	21	90	319	144	38	15								
<i>a1</i>	27	79	51	10	1	20	73	54	18	3	18	84	46	14	6								
<i>a2</i>	28	131	79	7	2	24	102	94	21	6	28	107	86	21	5								
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮								
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮								

Variables: *E* – Education system; *S* – Social security; *H* – Healthcare.

Levels of confidence: 1 – a great deal; 2 – quite a lot; 3 – not much; 4 – none; 9 – missing.

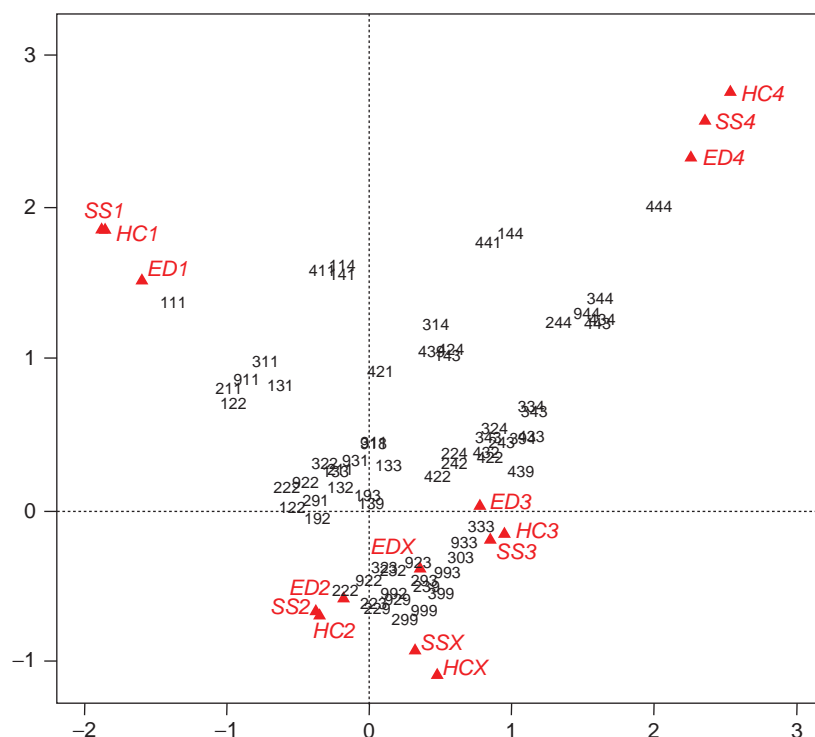


Figure 5 MCA of the individual-level responses to three questions. Percentages of inertia are calculated in a different way in MCA (e.g., see Greenacre, 2007). Here, the percentages are 34.8% and 26.4% respectively on axes 1 (horizontal) and 2 (vertical), that is, 61.2% explained in the solution.

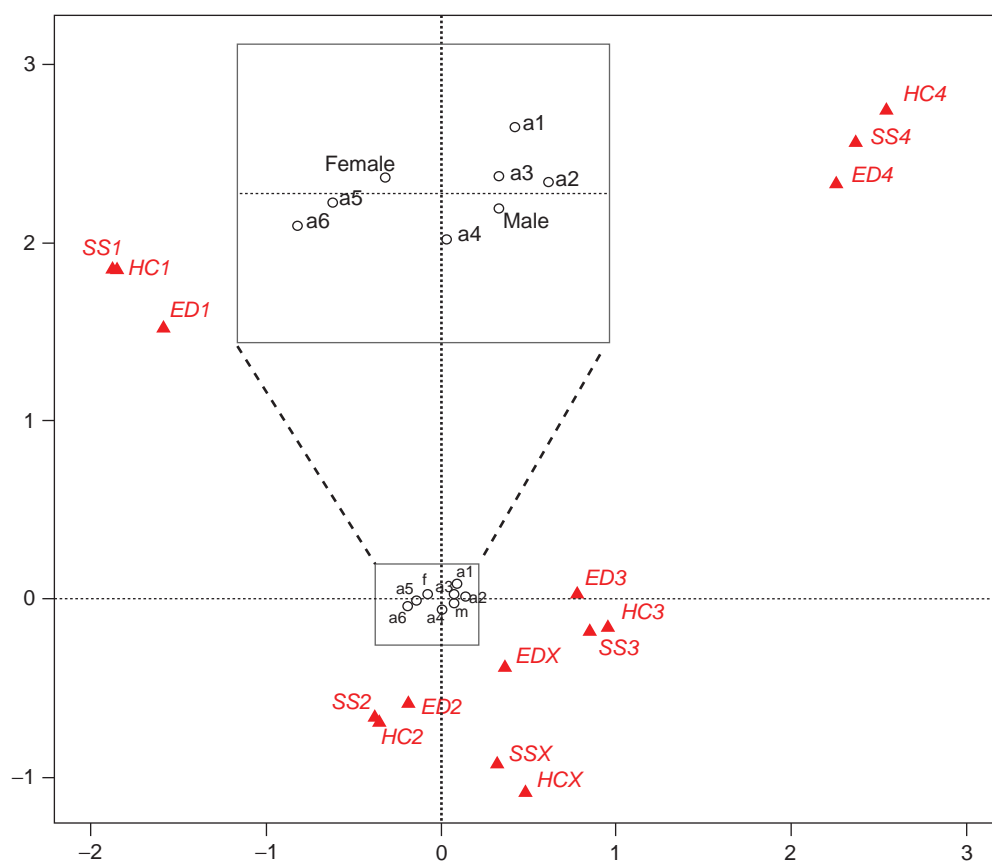


Figure 6 Supplementary points gender and age group in the MCA solution of **Figure 5**.

a value of 2 on a five-point rating scale (negative 1 to positive 5, e.g., disagree to agree) is coded as two values, 1 and 3, respectively, since 2 has one scale point on the rating scale below it and three scale points above it. In a ranking of 10 objects (where 1 is the most preferred, say), a ranking of 3 would be coded $10 - 3 = 7$ and $3 - 1 = 2$, respectively, since there are 7 objects less preferred and 2 more preferred. Notice that the idea is to obtain measures of association between the respondent and the doubled variables, hence, the positive pole in this last example has the higher value of 7 because the object is highly ranked.

2. *Fuzzy coding.* Continuous variables can be cut up into intervals and so converted into categorical data: for example, a variable such as temperature can be divided into three intervals using two cutpoints, so that a value of 2 represents a medium temperature, eventually coded in dummy variable form as $[0 \ 1 \ 0]$. This crisp coding clearly loses much information, and an alternative is fuzzy coding, where the value is coded using so-called membership functions: for example, a recoded value of $[0 \ 0.692 \ 0.308]$ would indicate a temperature somewhat higher up in the medium category toward the high category. The fuzzy coded values also add up to 1 and so may be used in conjunction with other dummy-coded categorical data in a CA.

Special Topics

Several variants of CA and MCA have been developed, especially in the social and environmental sciences for special situations:

1. *Subset CA and MCA.* Especially in survey research, we would like to analyze certain subsets of question responses, for example, just the disagree responses, or all the substantive responses excluding the missings. As CA uses the marginal sums as weights, the simple use of a reduced table leads to problematic situations because the margins change. This can be easily remedied, however, by maintaining the original margins of the full table whenever a subtable is analyzed, resulting in what is called subset CA, or subset MCA as the case may be.
2. *Canonical CA.* Especially in ecological applications, tables of counts are analyzed – typically, abundances of species at sampling locations – and additional variables are observed for each sample. Ecologists are interested in the variance of the abundance data that can be explained by these external variables, so CA is performed in a subspace that is restricted to be related (usually linearly related) to the external variables – this is called canonical CA (CCA). As a special case, if we have one external categorical variable, for example, regions of sampling, CCA is equivalent to aggregating

the counts of each species within each region and then performing CA.

3. *Analysis of square asymmetric tables:* Tables where the rows and columns represent the same sets of objects deserve special treatment, for example, mobility tables, transition matrices between states in behavioral studies, and brand switching in marketing. Usually, the diagonal dominates the data and obscures subtler asymmetrical features off the diagonal. One solution is to split the table as the sum of a symmetric and a skew-symmetric part, and analyze these separately.
4. *Statistical inference.* In simple CA, when the data form a contingency table based on random sampling, there are certain asymptotic results in the literature, based on the multivariate normal approximation to the multinomial distribution, that permit inference to be made about the parts of inertia explained by each dimension, as well as tests whether points differ significantly from the average point at the origin of the map, or differ from one another in pairwise tests. Resampling methods, such as the bootstrap and permutation testing, can estimate the confidence regions of points and test hypotheses in a less formal way. The CA model can also be estimated using maximum likelihood, in which case the interpretation is not based on projective geometry as in CA, but nevertheless permits more rigorous hypothesis testing.

See also: Multidimensional Scaling; Principal Components Analysis.

Bibliography

- Benzécri, J. -P. (1973). *L'Analyse des Données. Tome II: L'Analyse des Correspondances*. Paris: Dunod.
- Greenacre, M. (2007). *Correspondence Analysis in Practice*, 2nd edn. London: Chapman and Hall/CRC.
- Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In Horst, P. (ed.) *The Prediction of Personal Adjustment*, pp 321–348. New York: Social Science Research Council.
- Nishisato, S. (2006). *Multivariate Nonlinear Descriptive Analysis*. London: Chapman and Hall/CRC.
- Pearson, K. (1901). On lines and planes of closest fit to a system of points in space. *Philosophical Magazine* 2, 559–572.

Further Reading

- Blasius, J. and Greenacre, M. (eds.) (1998). *Visualizing Categorical Data*. San Diego, CA: Academic Press.
- Greenacre, M. and Blasius, J. (eds.) (1994). *Correspondence Analysis in the Social Sciences*. London: Academic Press.
- Greenacre, M. and Blasius, J. (eds.) (2006). *Multiple Correspondence Analysis and Related Methods*. London: Chapman and Hall/CRC.
- Murtagh, F. (2005). *Correspondence Analysis and Data Coding with R and Java*. London: Chapman and Hall/CRC.

Data Mining

R S J d Baker, Worcester Polytechnic Institute, Worcester, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Association rule mining – A class of data mining methods where there is automatic discovery of if–then rules of the form that if some set of variable values is found, another variable generally has a specific value.

Automated experimentation – A method used to create data where causal inference is facilitated. In this method, a pedagogical choice is randomly made by learning software in situations where multiple pedagogical choices are feasible.

Causal data mining – A class of data mining methods that attempts to infer causal structure between events within correlational data, using information about timing of events and their covariance, particularly in relation to other variables.

Classification – A class of data mining methods that attempts to predict the value of some variable, where the predicted variable is a binary or categorical variable.

Clustering – A class of data mining methods whose goal is to find data points that are similar to each other, splitting the full data set into a set of groups (such as groups of courses, groups of students, or groups of actions).

Correlation mining – A class of data mining methods whose goal is to find particularly strong linear correlations (positive or negative) between variables.

Data mining – The field of discovering novel and potentially useful information from large amounts of data. (Also see: educational data mining.)

Density estimation – The estimation of the probability density function underlying a data distribution.

Educational data mining – The area of scientific inquiry centered around the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods to better understand students and the settings which they learn in.

Gaming the system – Attempting to succeed in an interactive learning environment by exploiting properties of the system rather than by learning the material, including behaviors such as systematic guessing, hint abuse, spam posting, and point cartels.

Learning decomposition – A type of relationship mining, within the broader rubric of educational data mining. In learning decomposition, exponential learning curves are fit to performance data, relating student success to the amount of each type of pedagogical support a student has received (with an empirically determined weight for each type of support). The weights indicate how effective each type of pedagogical support is for improving learning. Relationships are only viewed as definitively causal if automated experimentation was used.

Q-matrix – A mapping between items and knowledge components (such as skills or concepts).

Regression – A class of data mining methods whose goal is to predict a quantitative dependent variable as a function of other variables.

Relationship mining – A class of data mining methods whose goal is to discover relationships between variables, in a data set with a large number of variables. This may take the form of attempting to find out which variables are most strongly associated with a single variable of particular interest, or may take the form of attempting to discover the relationship between any two variables in a large data set. Broadly, there are three types of relationship mining: association rule mining, correlation mining, and causal data mining.

Slip – A slip occurs when a student obtains an incorrect answer despite knowing the skills or concepts necessary to obtain a correct answer.

Introduction

Data mining, also called knowledge discovery in databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data. Data mining has been applied in a great number of fields, including retail sales, bioinformatics, and counter-terrorism. In recent years, there has been increasing interest in the use of data mining to investigate scientific questions within educational research, an area of inquiry termed educational data mining (EDM). EDM is defined as the area of scientific inquiry centered around the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods

to better understand students and the settings which they learn in.

EDM methods often differ from methods from the broader data mining literature, in explicitly exploiting the multiple levels of meaningful hierarchy in educational data. Methods from the psychometrics literature are often integrated with methods from the machine learning and data mining literatures to achieve this goal.

For example, in mining data about how students choose to use educational software, it may be worthwhile to simultaneously consider data at the keystroke level, answer level, session level, student level, classroom level, and school level. Issues of time, sequence, and context also play important roles in the study of educational data.

EDM has emerged as an independent research area in recent years, culminating in 2008 with the establishment of the annual International Conference on Educational Data Mining, and the *Journal of Educational Data Mining*.

Advantages Relative to Traditional Educational Research Paradigms

EDM offers several advantages, vis-à-vis more traditional educational research paradigms, such as laboratory experiments, in vivo experiments, and design research.

In particular, the advent of public educational data repositories, such as the Pittsburgh Science of Learning Center (PSLC) DataShop and the National Center for Education Statistics (NCES) data sets, has created a base which makes EDM highly feasible. In particular, the data from these repositories are often both ecologically valid (inasmuch as they are data about the performance and learning of genuine students, in genuine educational settings, involved in authentic learning tasks), and increasingly easy to rapidly access and begin research with. Balancing feasibility with ecological validity is often a difficult challenge for researchers in other educational research paradigms. By contrast, researchers who use data from these repositories can dispense with traditionally time-consuming steps such as subject recruitment (e.g., recruitment of schools, teachers, and students), scheduling of studies, and data entry (since data are already online). While the use of previously collected data has the potential to limit analyses to questions involving the types of data collected, in practice data from repositories or prior research have been useful for analyzing research questions far outside the purview of what the data were originally intended to study, particularly given the advent of models that can infer student attributes (such as strategic behavior and motivation) from the type of data in these repositories.

This increase in speed and feasibility has had the benefit of making replication much more feasible. Once a construct of educational interest (such as off-task behavior,

or whether or not a skill is known) has been empirically defined in data, it can be transferred to new data sets. The transfer of constructs is not trivial – often, the same construct can be subtly different at the data level, within data from a different context or system – but transfer learning and rapid labeling methods have been successful in speeding up the process of developing or validating a model for a new context. This has led to many EDM analyses being replicated across data from several learning systems or contexts.

Increasingly, the existence of data from thousands of students, having broadly similar learning experiences (such as using the same learning software), but in very different contexts, gives leverage that was never before possible, for studying the influence of contextual factors on learning and learners. It has historically been difficult to study how much the differences between teachers and classroom cohorts influence specific aspects of the learning experience; this sort of analysis becomes much easier with EDM. Similarly, the concrete impacts of fairly rare individual differences have been difficult to statistically study with traditional methods (leading case studies to be a dominant research method in this area) – EDM has the potential to extend a much wider tool set to the analysis of important questions in individual differences.

Main Approaches

There are a wide variety of current methods popular within EDM. These methods fall into the following general categories: prediction, clustering, relationship mining, discovery with models, and distillation of data for human judgment (Table 1). The first three categories are largely acknowledged to be universal across types of data mining (albeit in some cases with different names). The fourth and fifth categories achieve particular prominence within EDM.

Prediction

In prediction, the goal is to develop a model which can infer a single aspect of the data (predicted variable) from some combination of other aspects of the data (predictor variables). Prediction requires having labels for the output variable for a limited data set, where a label represents some trusted ground-truth information about the output variable's value in specific cases. In some cases, however, it is important to consider the degree to which these labels may in fact be approximate, or incompletely reliable.

Prediction has two key uses within EDM. In some cases, prediction methods can be used to study what features of a model are important for prediction, giving information about the underlying construct. This is a common approach in programs of research that attempt to predict student educational outcomes (Romero *et al.*, 2008), without

Table 1 The primary categories of educational data mining

<i>Category of method</i>	<i>Goal of method</i>	<i>Key applications</i>
Prediction	Develop a model which can infer a single aspect of the data (predicted variable) from some combination of other aspects of the data (predictor variables)	Detecting student behaviors (e.g., gaming the system, off-task behavior, and slipping); developing domain models; predicting and understanding student educational outcomes
Clustering	Find data points that naturally group together, splitting the full data set into a set of categories	Discovery of new student behavior patterns; investigating similarities and differences between schools
Relationship mining	Discover relationships between variables	Discovery of curricular associations in course sequences; discovering which pedagogical strategies lead to more effective/robust learning
Discovery with models	A model of a phenomenon developed with prediction, clustering, or knowledge engineering, is used as a component in further prediction or relationship mining.	Discovery of relationships between student behaviors, and student characteristics or contextual variables; analysis of research question across wide variety of contexts
Distillation of data for human judgment	Data are distilled to enable a human to quickly identify or classify features of the data.	Human identification of patterns in student learning, behavior, or collaboration; labeling data for use in later development of prediction model

predicting intermediate or mediating factors first. In a second type of usage, prediction methods are used in order to predict what the output value would be in contexts where it is not desirable to directly obtain a label for that construct (e.g., in previously collected repository data, where desired labeled data may not be available, or in contexts where obtaining labels could change the behavior

being labeled, such as modeling affective states, where self-report, video, and observational methods all present risks of altering the construct being studied).

For example, consider research attempting to study the relationship between learning and gaming the system, attempting to succeed in an interactive learning environment by exploiting properties of the system rather than by learning the material. If a researcher has the goal of studying this construct across a full year of software usage within multiple schools, it may not be tractable to directly assess, using nondata-mining methods, whether each student is gaming, at each point in time. Baker *et al.* (2008b) developed a prediction model by using observational methods to label a small data set, developing a prediction model using automatically collected data from interactions between students and the software for predictor variables, and then validating the model's accuracy when generalized to additional students and contexts. They were then able to study their research question in the context of the full data set.

Broadly, there are three types of prediction: classification, regression, and density estimation. In classification, the predicted variable is a binary or categorical variable. Some popular classification methods include decision trees, logistic regression (for binary predictions), and support vector machines. In regression, the predicted variable is a continuous variable. Some popular regression methods within EDM include linear regression, neural networks, and support vector machine regression. In density estimation, the predicted variable is a probability density function. Density estimators can be based on a variety of kernel functions, including Gaussian functions. For each type of prediction, the input variables can be either categorical or continuous; different prediction methods are more effective, depending on the type of input variables used.

Popular methods for assessing the goodness of a predictor include linear correlation, Cohen's kappa, and A' (the area under the receiver-operating curve – e.g., Bradley, 1997). Percent accuracy is generally not preferred for classification, as values of accuracy are highly dependent on the base rates of different classes (and hence, a very high accuracy can in some cases be achieved by a classifier that simply always predicts the majority class). When computing the goodness of a predictor, it is important to account for nonindependence of different observations involving the same student – to achieve this goal, EDM researchers often apply meta-analytical methods that can account for partial nonindependence, such as Strube's (1985) Adjusted Z, or select overly conservative estimators that assume complete nonindependence.

Clustering

In clustering, the goal is to find data points that naturally group together, splitting the full data set into a set of

clusters. Clustering is particularly useful in cases where the most common categories within the data set are not known in advance. If a set of clusters is optimal, within a category, each data point in general will be more similar to the other data points in that cluster than data points in other clusters. Clusters can be created at several different possible grain sizes: for example, schools could be clustered together (to investigate similarities and differences between schools), students could be clustered together (to investigate similarities and differences between students), or student actions could be clustered together (to investigate patterns of behavior) (Amershi and Conati, 2006; Beal *et al.*, 2006).

Clustering algorithms can either start with no prior hypotheses about clusters in the data (such as the k-means algorithm with randomized restart), or start from a specific hypothesis, possibly generated in prior research with a different data set (using the expectation-maximization algorithm to iterate toward a cluster hypothesis for the new data set). A clustering algorithm can postulate that each data point must belong to exactly one cluster (such as in the k-means algorithm), or can postulate that some points may belong to more than one cluster or to no clusters (such as in Gaussian mixture models).

The goodness of a set of clusters is usually assessed with reference to how well the set of clusters fits the data, relative to how much fit might be expected solely by chance given the number of clusters, using statistical metrics such as the Bayesian information criterion.

Relationship Mining

In relationship mining, the goal is to discover relationships between variables, in a data set with a large number of variables. This may take the form of attempting to find out which variables are most strongly associated with a single variable of particular interest, or may take the form of attempting to discover which relationships between any two variables are strongest.

Broadly, there are four types of relationship mining: association rule mining, correlation mining, sequential pattern mining, and causal data mining. In association rule mining, the goal is to find if-then rules of the form that if some set of variable values is found, another variable will generally have a specific value. For example, a rule might be found of the form {student is frustrated, student has stronger goal of learning than goal of performance} \rightarrow {student frequently asks for help}. In correlation mining, the goal is to find (positive or negative) linear correlations between variables. In sequential pattern mining, the goal is to find temporal associations between events – for example, to determine what path of student behaviors leads to an eventual learning event of interest. In causal data mining, the goal is to find whether one event (or observed construct) was the cause of another event

(or observed construct), either by analyzing the covariance of the two events (e.g., TETRAD – Scheines *et al.*, 1994) or by using information about how one of the events was triggered. For example, if a pedagogical event is randomly chosen using automated experimentation (Mostow, 2008), and frequently leads to a positive learning outcome, a causal relationship can be inferred.

Relationships found through relationship mining must satisfy two criteria: statistical significance, and interestingness. Statistical significance is generally assessed through standard statistical tests, such as *F*-tests. Because large numbers of tests are conducted, it is necessary to control for finding relationships through chance. One method for doing this is to use *post hoc* statistical methods or adjustments which control for the number of tests conducted, such as the Bonferroni adjustment. This method can increase confidence that an individual relationship found was not likely to be due to chance. An alternate method is to assess the overall probability of the pattern of results found, using Monte Carlo methods. This method assesses how likely it is that the overall pattern of results arose due to chance.

The interestingness of each finding is assessed in order to reduce the set of rules/correlations/causal relationships communicated to the data miner. In very large data sets, hundreds of thousands of significant relationships may be found. Interestingness measures attempt to determine which findings are the most distinctive and well supported by the data, in some cases also attempting to prune overly similar findings. There are a wide variety of interestingness measures, including support, confidence, conviction, lift, leverage, coverage, correlation, and cosine. Some investigations have suggested that lift and cosine may be particularly relevant within educational data (Merceron and Yacef, 2008).

Discovery with Models

In discovery with a model, a model of a phenomenon is developed through prediction, clustering, or, in some cases, knowledge engineering (within knowledge engineering, the model is developed using human reasoning rather than automated methods). This model is then used as a component in another analysis, such as prediction or relationship mining.

In the prediction case, the created model's predictions are used as predictor variables in predicting a new variable. For instance, analyses of complex constructs, such as gaming the system within online learning, have generally depended on assessments of the probability that the student knows the current knowledge component being learned (Baker *et al.*, 2008b; Walonoski and Heffernan, 2006). These assessments of student knowledge have in turn depended on models of the knowledge components in a domain,

generally expressed as a mapping between exercises within the learning software and knowledge components.

In the relationship mining case, the relationships between the created model's predictions and additional variables are studied. This can enable a researcher to study the relationship between a complex latent construct and a wide variety of observable constructs.

Often, discovery with models leverages the validated generalization of a prediction model across contexts. For instance, Baker (2007) used predictions of gaming the system across a full year of educational software data to study whether state or trait factors were better predictors of how much a student would game the system. Generalization in this fashion relies upon appropriate validation that the model accurately generalizes across contexts.

Distillation of Data for Human Judgment

Another area of interest within EDM is the distillation of data for human judgment. In some cases, human beings can make inferences about data, when it is presented appropriately, that are beyond the immediate scope of fully automated data mining methods. The methods in this area of EDM are information visualization methods – however, the visualizations most commonly used within EDM are often different than those most often used for other information visualization problems (Kay *et al.*, 2006; Hershkovitz and Nachmias, 2008), owing to the specific structure, and the meaning embedded within that structure, often present in educational data.

Data are distilled for human judgment in EDM for two key purposes: identification and classification. When data are distilled for identification, data are displayed in ways that enable a human being to easily identify well-known patterns that are nonetheless difficult to formally express. For example, one classic EDM visualization is the learning curve, which displays the number of opportunities to practice a skill on the x -axis, and displays performance (such as percent correct or time taken to respond) on the y -axis. A curve with a smooth downward progression that is steep at first and gentler later indicates a well-specified knowledge component model. A sudden spike upward, by contrast, indicates that more than one knowledge component is included in the model (Corbett and Anderson, 1995).

Alternately, data may be distilled for human labeling, to support the later development of a prediction model. In this case, subsections of a data set are displayed in visual or text format, and labeled by human coders. These labels are then generally used as the basis for the development of a predictor. This approach has been shown to speed the development of prediction models of complex phenomena such as gaming the system by around 40 times, relative to prior approaches for collecting the necessary data (Baker and de Carvalho, 2008).

Main Applications

There have been a wide number of applications of EDM, as reflected throughout this article. In this section, four areas of application that have received particular attention within the field are discussed.

One key area of application is in improving student models, models that provide detailed information about a student's characteristics or states, such as knowledge, motivation, meta-cognition, and attitudes. Modeling the individual differences between students, in order to enable software to respond to those individual differences, is a key theme in educational software research. In the last few years, EDM methods have enabled considerable expansion in the sophistication of student models. In particular, EDM methods have enabled researchers to make higher level inferences about students' behavior, such as when a student is gaming the system, when a student has slipped (making an error despite knowing a skill), and when a student is engaging in self-explanation (Shih *et al.*, 2008). These richer student models have been useful in two fashions. First, these models have increased our ability to predict student knowledge and future performance – incorporating models of guessing and slipping into predictions of student's future performance has increased the accuracy of these predictions by up to 48% (Baker *et al.*, 2008a). Second, these models have enabled researchers to study what factors lead students to make specific choices in a learning setting, a type of scientific discovery with models discussed below.

A second key area of application is in discovering or improving models of the knowledge structure of the domain. In EDM, methods have been created for rapidly discovering accurate domain models directly from data. These methods have generally combined psychometric modeling frameworks with advanced space-searching algorithms, and are generally posed as prediction problems for the purpose of model discovery (e.g., attempting to predict whether individual actions will be correct or incorrect, using different domain models, is one common method for developing these models). Barnes *et al.* (2005) have proposed algorithms for automatically discovering a Q-matrix from data. Cen *et al.* (2006) proposed algorithms for using codified expert knowledge about differences between items to drive automated search for item response theory (IRT) models. Pavlik *et al.* (2008) has proposed algorithms for finding partial order knowledge structure models (Desmarais *et al.*, 1996), by looking at the covariation of individual items.

A third key area of application is in studying the pedagogical support provided by learning software. Modern educational software gives a variety of types of pedagogical support to students. Discovering which pedagogical support is most effective has been a key area of interest for educational data miners. Learning decomposition

(Beck and Mostow, 2008), a type of relationship mining, fits exponential learning curves to performance data, relating student success to the amount of each type of pedagogical support a student has received (with a weight for each type of support). The weights indicate how effective each type of pedagogical support is for improving learning. An illustrative example is given in the next section.

A fourth key area of application of EDM is for scientific discovery about learning and learners. This takes on several forms. Applying EDM to answer questions in any of the three areas previously discussed (e.g., student models, domain models, and pedagogical support) can have broader scientific benefits; for example, the study of pedagogical support may have the long-term potential to enrich theories of scaffolding. Beyond just these three areas, however, there have been many analyses aimed directly toward scientific discovery. Discovery with models is a key method for scientific discovery via EDM. Research on studying whether state or trait factors were better predictors of how much a student would game the system (Baker, 2007) is a prominent example of this approach within EDM research. Learning decomposition methods are another prominent method for conducting scientific discovery about learning and learners.

Illustrative Example

In this section, a brief case study is discussed, as a concrete, best practices example of how the EDM method of learning decomposition (a type of relationship mining) was used to determine the relative efficacy of different types of learning material presented to students.

In Beck and Mostow (2008), data were obtained from 346 American elementary school students reading 6.9 million words, over the course of a year, while using intelligent tutor software that teaches reading. These words were presented in the form of stories, and students and the software took turns choosing stories (the software's choice of stories was based on the student's approximate grade reading level). Beck and Mostow were interested in determining whether rereading a story (a popular option for children) is more or less effective at promoting word learning than encountering the same word in a different story. They were also interested in whether there would be individual differences, such that some students would benefit from a different pattern of practice than others.

Beck and Mostow obtained data for each student's performance in reading each story within the software. Reading time was used as a continuous measure of word knowledge; misreading and help-requests were also taken into account, reading opportunities where these behaviors occurred were assigned a time of 3.0 s (99.9% of word reads were faster than 3.0). An exponential model of practice was set up, relating response time to the function:

$$\text{time} = A \times e^{-b(W - \tau_1 + \tau_2)}.$$

In this equation, parameter A represents student performance on the first opportunity to read a given word, parameter b represents the overall speed of learning, e is 2.718, and τ_1 and τ_2 represent the number of times the word is read, within two different types of practice. In this case, τ_1 was defined as the number of times the word was read when rereading a story and τ_2 was defined as the number of times the word was read when reading a story for the first time. W is the relative speed gain associated with the two types of practice. If W equals 1, the two types of practice are considered to be equally effective; if W is above 1, opportunities of type τ_1 are more effective than opportunities of type τ_2 (and the reverse holds true if W is below 1).

Across the population of students, the median value of W for rereading obtained by Beck and Mostow was 0.49, suggesting that rereading a story leads to approximately half as much learning as reading a new story. Of the 346 students, 95 had a W parameter statistically significantly under 1, whereas only seven students had a W parameter value statistically significantly over 1, a statistically significant result across the entire class.

Beck and Mostow next used the values of W from the model in a subsequent logistic regression analysis (an example of discovery with models). In this analysis, the learning decomposition model was used to split the population into students who benefited from rereading and students who did not benefit from rereading, and a variety of explanatory variables were tested to see if they explained which students benefited from rereading. This analysis determined that students with overall low reading speed who were receiving special needs learning support actually benefited from rereading.

See also: Causal Inference; Cluster Analysis: Overview; Computational Statistics; Discrimination and Classification; Educational Data Modeling; Exploratory Data Analysis; Factor Analysis: An Overview and Some Contemporary Advances; Measure of Association.

Bibliography

- Amershi, S. and Conati, C. (2006). Automatic recognition of learner groups in exploratory learning environments. *Proceedings of ITS 2006, 8th International Conference on Intelligent Tutoring Systems*, pp 463–472.
- Baker, R. S. J. d. (2007). Is gaming the system state-or-trait? Educational data mining through the multi-contextual application of a validated behavioral model. *Complete On-Line Proceedings of the Workshop on Data Mining for User Modeling at the 11th International Conference on User Modeling 2007*, pp 76–80 Corfu, Greece.
- Baker, R. S. J. d., Corbett, A. T., and Aleven, V. (2008a). More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, pp 406–415. Montreal, Canada.

- Baker, R. S. J. d., Corbett, A. T., Roll, I., and Koedinger, K. R. (2008b). Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction* **18**(3), 287–314.
- Baker, R. S. J. d. and de Carvalho, A. M. J. A. (2008). Labeling student behavior faster and more precisely with text replays. *Proceedings of the First International Conference on Educational Data Mining*, pp 38–47. Montreal, Canada.
- Barnes, T., Bitzer, D., and Vouk, M. (2005). Experimental analysis of the Q-matrix method in knowledge discovery. *Lecture Notes in Computer Science 3488: Foundations of Intelligent Systems*, pp 603–611. Saratoga Springs, NY, USA.
- Beal, C. and Lee, H. (2006). Classifying learner engagement through integration of multiple data sources. *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-2007)*. pp 151–156. Boston, MA.
- Beck, J. E. and Mostow, J. (2008). How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, pp 353–362. Montreal, Canada.
- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* **30**, 1145–1159.
- Cen, H., Koedinger, K., and Junker, B. (2006). Learning factors analysis – a general method for cognitive model evaluation and improvement. *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, 12–19. Jhongli, Taiwan.
- Corbett, A. T. and Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction* **4**, 253–278.
- Desmarais, M. C., Maluf, A., and Liu, J. (1996). User-expertise modeling with empirically derived probabilistic implication networks. *User Modeling and User-Adapted Interaction* **5**, (3–4), 283–315.
- HersHKovitz, A. and Nachmias, R. (2008). Developing a log-based motivation measuring tool. *Proceedings of the First International Conference on Educational Data Mining*, pp 226–233. Montreal, Canada.
- Kay, J., Maisonneuve, N., Yacef, K., and Reimann, P. (2006). The big five and visualisations of team work activity. In Ikeda, M., Ashley, K. D. and Chan, T.-W. (eds.) *Proceedings of Intelligent Tutoring Systems (ITS06)*, Taiwan. Lecture Notes in Computer Science 4053, pp 197–206. Berlin: Springer.
- Merceron, A. and Yacef, K. (2008). Interestingness measures for association rules in educational data. *Proceedings of the First International Conference on Educational Data Mining*, pp 57–66. Montreal, Canada.
- Mostow, J. (2008). Experience from a reading tutor that listens: Evaluation purposes, excuses, and methods. In Kinzer, C. K. and Verhoeven, L. (eds.) *Interactive Literacy Education: Facilitating Literacy Environments through Technology*, pp 117–148. New York: Erlbaum/ Taylor and Francis.
- Pavlik, P., Cen, H., Wu, L., and Koedinger, K. (2008). Using item-type performance covariance to improve the skill model of an existing tutor. *Proceedings of the First International Conference on Educational Data Mining*, pp 77–86. Montreal, Canada.
- Romero, C., Ventura, S., Espejo, P. G., and Hervás, C. (2008). Data mining algorithms to classify students. *Proceedings of the First International Conference on Educational Data Mining*, pp 8–17. Montreal, Canada.
- Scheines, R., Sprites, P., Glymour, C., and Meek, C. (1994). *Tetrad II: Tools for Discovery*. Hillsdale, NJ: Erlbaum.
- Shih, B., Koedinger, K. R., and Scheines, R. (2008). A response-time model for bottom-out hints as worked examples. *Proceedings of the First International Conference on Educational Data Mining*, pp 117–126. Montreal, Canada.
- Strube, M. J. (1985). Combining and comparing significance levels from nonindependent hypothesis tests. *Psychological Bulletin* **97**, 334–341.
- Walonoski, J. and Heffernan, N. T. (2006). Detection and analysis of off-task gaming behavior in intelligent tutoring systems. *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, pp 382–391. Jhongli, Taiwan.

Further Reading

- Baker, R. S. J. d., Barnes, T., and Beck, J. E. (eds.) (2008). Educational Data Mining 2008. *Proceedings of the 1st International Conference on Educational Data Mining*. Montreal, Canada, 20–21 June.
- Baker, R. S. J. d., Beck, J. E., Berendt, B., et al. (2007). *Complete On-Line Proceedings of the Workshop on Data Mining for User Modeling, at the 11th International Conference on User Modeling (UM 2007)*. Corfu, Greece.
- Barnes, T. and Beck, J. E. (eds.) (2007). *Proceedings of the Workshop on Educational Data Mining, at the 13th International Conference on Artificial Intelligence in Education*. Los Angeles, CA.
- Moore, A. (2005). *Statistical Data Mining Tutorials*. <http://www.autonlab.org/tutorials> (accessed August 2009).
- Mostow, J. and Beck, J. (2006). Refined micro-analysis of fluency gains in a Reading Tutor that listens. *Thirteenth Annual Meeting of the Society for the Scientific Study of Reading*. Vancouver, BC, Canada.
- Romero, C. and Ventura, S. (2006). *Data Mining in e-Learning*. Southampton, UK: WIT Press.
- Romero, C. and Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications* **33**(1), 135–146.
- Witten, I. H. and Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd edn. San Francisco, CA: Morgan Kaufmann.

Decision Theory

M Mendoza, Instituto Tecnológico Autónomo de México, México D.F., Mexico

E Gutiérrez-Peña, Universidad Nacional Autónoma de México, México D.F., Mexico

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Axiom – A proposition that is not susceptible to proof or disproof; its truth is assumed to be self-evident.

Introduction

The Process of Decision Making

One of the most common activities of human beings is that of decision making. Every person is constantly deciding on a wide variety of different subjects. There are easy as well as difficult decisions; there are important and irrelevant decisions; one must face personal and professional decisions. In the end, we all know, from our particular experience, that there are good and bad decisions, so a natural question arises. Are there any rules or procedures for decision making which guarantee that the final result is a good decision?

A vast amount of effort has been devoted to explore this subject. Psychologists have studied how decision makers work under different conditions. From a philosophical perspective, even the existence of such a thing as a good decision has been questioned. The logical approach has contributed to the understanding of the decision-making process. Mathematics, including statistics, has played a major role in providing a formal structure for the process and defining criteria for optimality.

Under the heading of decision theory, the literature offers an account of the ways people actually make decisions and a discussion on the mechanisms underlying this behavior. This is called a descriptive decision theory. On the other hand, we can also find discussions about the principles to consider when making rational decisions. In this case, we have a normative decision theory.

In this article, we are concerned with a normative decision theory. We discuss the solution of a general class of decision problems and, briefly but not less importantly, comment on the relationship between decision theory and statistical inference.

Decision Problems

A decision problem can be defined as a situation where a person or a group of people (the decision maker) must select one and only one element (an action) from a given

set $A = \{a_1, a_2, \dots, a_k\}$. The idea is to choose the best action, and thus the subjective nature of the solution arises naturally – the best action must be such for the specific decision maker. For this purpose, every action in A is judged in terms of the consequence it produces. Therefore, another component of the problem is the set of consequences $C = \{c_1, c_2, \dots, c_k\}$ where c_i stands for the consequence from action a_i . If, for every action, the corresponding consequence is completely known and occurs every time the action is selected, then we have a decision problem without uncertainty.

Under these circumstances, choosing an action is equivalent to choosing a consequence and the best action will be that leading to the best or most preferable consequence. Thus, the decision maker must define the set of actions A , the set of consequences C , and must also express her personal subjective preferences among the different elements of C . Once this structure is defined, the problem is solved by choosing the action whose consequence is the most preferable. This seems to be a rather simple exercise; however, the importance of a careful definition of the problem cannot be overestimated. The set A must include all available actions, and these must represent actual alternative options.

Concerning the principles we referred to in the section titled ‘The process of decision making’, in the case of a decision without uncertainty, these are implicit and apply to the decision maker’s preferences. They assume all actions to be comparable in terms of preferences and the preference relation to be transitive. If the structure of the problem is not well-defined or the axioms (coherence principles) are not fulfilled, the decision maker will solve (if at all) a wrong problem.

If preferences are expressed in terms of a numerical score u such that, for every pair (i, j) , c_i is less preferable than c_j if and only if $u(c_i) < u(c_j)$, then the problem reduces to maximizing the function u over C . The score u is called the utility function and the solution is any action that maximizes the decision maker’s subjective utility. Sometimes it is easier to use a loss function to represent preferences. If l is a loss function, then c_i is less preferable than c_j if and only if $l(c_i) > l(c_j)$. In this case, the best action minimizes the decision maker’s subjective loss. As long as both functions, loss and utility, describe the same preferences, the optimal decision remains the same.

Decision problems without uncertainty are conceptually simple. Nevertheless, their solution can be difficult in practice. In particular, translation of preferences into a

numerical score is not easy. However, the utility function is just a numerical representation of the preferences and thus these technical problems are usually due to the difficulties associated with the elicitation of preferences.

Recalling the question whether there are any rules or procedures for decision making which guarantee that the final result is a good decision, we can now say that, in the case of decision problems without uncertainty, if the problem is properly structured, the utility function really reflects the decision maker's preferences, and these preferences obey the coherence principles, the answer is yes.

Unfortunately, real-world decision problems are usually more complex than those without uncertainty. It often happens that once an action has been chosen, there is a set of possible consequences associated with it. Among them, one and only one will take place depending upon the occurrence of some uncertain event. The structure of a decision problem under uncertainty is described in the section titled 'Definition of a decision problem'.

Historical Background and Related Fields

Ideas related to decision theory can be traced back to at least the eighteenth century. However, most of the formal developments emerged during the past 80 years and many disciplines have contributed to the methods of decision making. The case without uncertainty belongs to the domain of optimization and operations research. Game theory provided the basic framework for the case under uncertainty. Modern economic theory too deals with interesting decisions, such as the choice of an investment portfolio. In fact, this problem is the origin of an entire field known as financial economics.

The relationship between statistics and decision theory deserves a special mention. Statistics can contribute to the solution of decision problems under uncertainty by means of methods which allow the decision maker to describe her own uncertainty. The result is known as statistical decision theory, and different approaches to statistical inference have led to different statistical decision theories. Conversely, any situation where a statistical inference is required can be seen as a decision problem under uncertainty (an inference is just an assertion regarding the phenomenon of interest and is chosen among several alternatives). From this perspective, a theory of statistical inference can be developed on the grounds of a specific decision theory. The most outstanding example is Bayesian statistics, a statistical theory built upon the axiomatic decision theory described in the section titled 'Axiomatic decision theory'. Bayesian inference, as a theoretical discipline, and Bayesian methods, as a set of tools for inference in practice, have been growing very rapidly in recent years.

Decision theory is an interdisciplinary field with many relevant contributions published in economics, mathematics,

psychology, and statistics, among other areas. The basic concepts in axiomatic decision theory were mainly developed from the 1930s to the 1970s and now appear in a number of Bayesian statistics texts.

Decisions in Education

Decisions are omnipresent in the field of education. A specific syllabus must be chosen, among different alternatives, for each course. The admissions committee must decide whether or not an applicant should be accepted as a student of the program. As part of the grading process, for every exam, a cutoff value must be selected to decide whether a student fails. On the other hand, students face a decision problem when they choose a career; every time they select an elective course they are making a decision; and when a student chooses an answer in a multiple-response test, she is also deciding among a set of alternative options. These are only a few examples of decision problems in education.

For a long time, these problems were faced without any decision-theoretical backing. However, current research in education, specifically on test design, involves decision theory as a basic resource. As discussed by van der Linden (1991), this process is related to a change in the approach to design. Formerly, a test was considered as a measurement tool and research on tests was directed to explore the relation among the observed measurement and the attribute to be measured. More recently, it has been argued that the measurement is relevant insofar as it is useful to make decisions about the examinee. Therefore, an appropriate design must take into account the specific decision problem to be addressed. For examples of this idea we refer the reader to Novick and Lindley (1978), where some particular utility functions are used to identify cutoff scores; Sawyer (1996), who uses decision theory to validate course-placement tests; and Vos (1999), where optimal sequential mastery tests are formulated using decision-theoretical arguments.

Decision Theory

Definition of a Decision Problem

A decision problem under uncertainty is defined by the following elements:

1. $A = \{a_1, a_2, \dots, a_k\}$,
2. $E = \{E_{11}, E_{12}, \dots, E_{1m_1}; E_{22}, E_{21}, \dots, E_{2m_2}; \dots; E_{k1}, E_{k2}, \dots, E_{km_k}\}$, and
3. $C = \{c_{11}, c_{12}, \dots, c_{1m_1}; c_{22}, c_{21}, \dots, c_{2m_2}; \dots; c_{k1}, c_{k2}, \dots, c_{km_k}\}$.

Here, A is an exhaustive and exclusive set of actions; E is the set of uncertain events where, for every action a_i , the collection of events $E_i = \{E_{i1}, E_{i2}, \dots, E_{im_i}\}$ is assumed to be a partition of the certain event Ω . Finally, C is the set of

consequences and is such that, to each pair (a_i, E_{ij}) , there corresponds a consequence c_{ij} . Both the space of actions and the set of uncertain events may contain an infinite number of elements.

In order to solve this problem, the simple optimality principle used in the section titled 'Decision problems', no longer applies. We cannot substitute the choice of an action by the choice of a consequence. In fact, in this new setting, once the decision maker chooses a_i , what she gets is the result of a lottery whose possible prizes are $\{u(c_{i1}), \dots, u(c_{imi})\}$ where $u(c_{ij})$ occurs with probability $p_{ij} = P(E_{ij})$ and, for each i , $p_{i1} + p_{i2} + \dots + p_{imi} = 1$. Here, there is no obvious way to transfer a utility score to the action a_i from the corresponding lottery. In fact, every method devised to solve a problem of this kind relies on a specific proposal of how to carry this out. Moreover, since the event E_{ij} is uncertain for the decision maker, the probability p_{ij} is a measure of her personal beliefs about the occurrence of E_{ij} . Thus, both utilities and probabilities are subjective.

A number of methods have been proposed to solve decision problems under uncertainty; all of them replace the original problem with another which, in a way, does not involve any uncertainty (see the next section).

Example

A graduate studies committee must evaluate every application to a PhD program and decide whether the corresponding candidate can be admitted. In a simplified version, only two actions are considered. Thus $A = \{a_1, a_2\}$, where a_1 = admit the candidate, a_2 = reject the candidate. The consequences of each of these actions are uncertain. Rejection of a candidate would be appropriate if this action prevents admission of a bad student. On the other hand, it would be a mistake to reject a good student. If the set of uncertain events is defined as $E = \{E_1, E_2\}$, with E_1 = candidate's performance is good, E_2 = candidate's performance is poor, then the consequences are given by $C = \{c_{11}, c_{12}, c_{21}, c_{22}\}$ where c_{ij} is the consequence obtained if the action a_i is chosen and the uncertain event E_j occurs. Here, c_{11} and c_{22} represent good outcomes whereas c_{12} and c_{21} are mistakes. A reasonable utility function describing these preferences must be such that $u(c_{ii}) > u(c_{ij})$ for $i, j = 1, 2$. This problem has the structure described by Figure 1.

Intuitive Solutions to a Decision Problem

Here, we review two popular methods used to solve a decision problem under uncertainty. As many others, each of them proceeds in a two-step fashion. First, for every action, a utility score is derived from the corresponding lottery. In the second step, actions are compared as they would be in a problem without uncertainty, so the best action is chosen as that maximizing the derived utility.

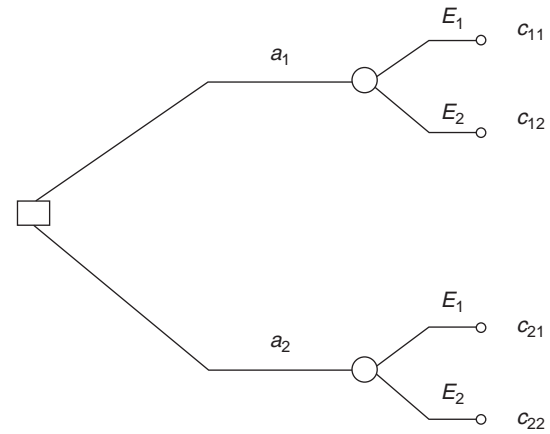


Figure 1 A decision tree: the square represents a decision node while circles represent uncertain nodes.

Minimax

For each action a_i , this method looks at the worst possible consequence and then acts as if this consequence will occur for sure. In other words, Minimax assigns the utility $u_M(a_i) = \min_j u(c_{ij})$ and this new utility is then maximized over the space of actions, so that the best action, denoted by a^* , is such that $a^* = \max_i \min_j u(c_{ij})$. The method is named after this last expression: Maximin, if we work in terms of utilities or Minimax if a loss function is used. This procedure completely ignores the probabilities that the decision maker has assigned to the events in E_i . In some sense, this is a pessimistic approach to decision making and often yields what is known as opportunity loss.

Maximum expected utility

This method assigns, to each action a_i , the weighted average of the utilities assigned to the corresponding consequences, the weights being the respective probabilities, that is, $u_E(a_i) = u(c_{i1}) p_{i1} + u(c_{i2}) p_{i2} + \dots + u(c_{imi}) p_{imi}$. This expected utility is then maximized over the space of actions, so that the best action, denoted by a^* , is such that $a^* = \max_i u_E(a_i)$. This approach to decision making takes into account all the information and is the only one that is consistent with the axiomatic theory described in the next section.

Example (ctd.)

Let $p_1 = P(E_1)$ and $p_2 = P(E_2)$. In this case, the expected utilities are computed as $u_E(a_1) = u(c_{11}) p_1 + u(c_{12}) p_2$ and $u_E(a_2) = u(c_{21}) p_1 + u(c_{22}) p_2$, and the action leading to the larger of these two quantities is chosen. For the sake of simplicity, let us consider the case where the preferences are described by means of a loss function. Moreover, let us assume that consequences associated with successes are assigned a zero loss so that $l(c_{11}) = l(c_{22}) = 0$ (with $l(c_{ij}) > 0$; $i \neq j$). In this setting,

$I_E(a_1) = I(c_{12}) p_2$, $I_E(a_2) = I(c_{21}) p_1$ and then a_1 is chosen as the optimal action if, and only if, $I(c_{12}) p_2 < I(c_{21}) p_1$ or, equivalently, $p_1/p_2 > I(c_{12})/I(c_{21})$. Here, the cutoff value depends on the relative importance of c_{12} and c_{21} . Note that, even if $p_1 < p_2$, a_1 will still be chosen as the best action if $p_1 > (I(c_{12})/I(c_{21})) p_2$.

Axiomatic Decision Theory

Coherence axioms can be formulated in several different ways. One of the most intuitive discussions concerning the axiomatic basis for decision theory appears in Lindley (1972). For a more technical version, the reader is referred to Bernardo and Smith (1994).

Let us represent an action a by $a = \{c_1 | E_1, \dots, c_m | E_m\}$, thus indicating that an action is just a lottery where the decision maker gets the consequence c_j whenever the event E_j occurs. Then, it is clear that any consequence c is a particular case of an action since c is equivalent to $\{c | \Omega\}$, where Ω is the certain event.

1. *Axiom 1* (comparability). For every pair of actions a_1 and a_2 in \mathcal{A} , one and only one of the following conditions holds:

- a_1 is less preferable than a_2 (denoted $a_1 < a_2$)
- a_1 is more preferable than a_2 (denoted $a_1 > a_2$)
- a_1 and a_2 are equally preferable (denoted $a_1 \sim a_2$).

Moreover, it is possible to find two consequences c^* and c_* such that $c^* > c_*$ and $c_* \leq c \leq c^*$ for any consequence c ($a_1 \leq a_2$ means a_1 is not more preferable than a_2 ; $a_1 \geq a_2$ is similarly defined).

Discussion. The comparability of actions implies both the comparability of consequences and the comparability of events. Denote by \bar{E} the complement of the event E and let $a_1 = \{c^* | E_1, c_* | \bar{E}_1\}$ and $a_2 = \{c^* | E_2, c_* | \bar{E}_2\}$. Then comparing a_1 with a_2 is equivalent to comparing the likelihoods of the events E_1 and E_2 : a_1 will be preferable to a_2 if and only if E_1 is more likely than E_2 .

2. *Axiom 2* (transitivity). If $a_1 > a_2$ and $a_2 > a_3$, then $a_1 > a_3$.

Discussion. This axiom assumes that the set of actions can be ordered in such a way that a search for the most preferable element makes sense.

3. *Axiom 3* (substitutability and dominance). If $a_1 > a_2$ when the event E occurs and $a_1 > a_2$ when the event \bar{E} occurs, then $a_1 > a_2$.

Discussion. This axiom implies that, in any given action $a = \{c_1 | E_1, \dots, c_m | E_m\}$, a consequence c_j can be replaced by any action which is equivalent to c_j . This axiom also implies that, given two actions $a_1 = \{c_{11} | E_1, \dots, c_{1m} | E_m\}$ and $a_2 = \{c_{21} | E_1, \dots, c_{2m} | E_m\}$, if $c_{1j} \geq c_{2j}$ for all j , then $a_1 \geq a_2$. Moreover, if $c_{1j} > c_{2j}$ for some j , then $a_1 > a_2$.

4. *Axiom 4* (reference events). The decision maker can conceive of a procedure to generate a random point in the unit square such that, given two regions R_1 and R_2 in the unit square, the event that the random point is contained in R_1 is regarded as more likely than the event that the random point is contained in R_2 if, and only if, the area of R_1 is larger than the area of R_2 .

Discussion. This axiom defines a unit of measurement which can be used to describe the decision maker's uncertainty concerning the occurrence of the various events which are relevant for the decision problem.

Axioms 1–3 are qualitative. They establish the rules that comparison of actions must obey. Axiom 4 is of a quantitative nature and allows the decision maker to measure her degree of belief concerning uncertain events in terms of probability.

The axioms imply a three-step procedure to choose among actions. First, all preferences between consequences must be quantified in terms of a numerical utility (or, alternatively, a loss) function. Second, the uncertainty regarding any event affecting the consequences of an action must be quantified in terms of probability. Finally, the decision maker's preferences about any pair of alternative actions must be described in units of expected utility and hence the best decision will be to choose the action that maximizes such expected utility.

In brief, the main idea behind this formulation is this: if the coherence axioms are acceptable for a decision maker, then she has no other option but to use the maximum expected utility criterion to choose among actions. Any other method leads to the same optimal decisions or conflicts with the axioms. Stating the axioms explicitly allows anyone to evaluate if she is willing to act as a coherent decision maker defined in this particular way. In addition, if the axioms are adopted, the existence of a probability function for the uncertain events, and a utility function for the consequences, is no longer an assumption. It is a fact, following from the axioms themselves.

This is a subjective method since the utility and probability functions must describe the preferences and beliefs of the specific decision maker. Interestingly, the subjective nature of the utility function has not been a matter of controversy. On the other hand, the idea of subjective probabilities led to a lively debate in the early years and even now appears as a relevant issue in some applications (see Schneider (2002), for an example).

When the decision maker is faced with new information concerning the uncertain events, her beliefs must be revised in order to obtain an updated (posterior) probability measure describing all the available knowledge (which includes the original, prior beliefs as well as the new information). Another important consequence of the axioms is that this updating must be consistent with Bayes' rule.

When dealing with decisions under uncertainty, one cannot guarantee that the optimal action is a good decision in the sense that it will necessarily lead to the best possible consequence. Apart from guaranteeing compliance with the coherence axioms, the maximum expected utility method only leads to the action whose expected consequence is the most preferable.

Example (ctd.)

To improve the results of the admission process to the PhD program, the graduate studies committee can ask the candidates to take a selection exam. The corresponding grade of the exam (X) is then used to update the probabilities $P(E_1)$ and $P(E_2)$. If the observed grade is x , then the new probabilities are defined as $P(E_1 | X=x)$ and $P(E_2 | X=x)$ and can be obtained via Bayes' formula as $P(E_i | X=x) = P(X=x | E_i) P(E_i) / P(X=x)$; $i = 1, 2$.

If these updated probabilities are used to make a decision, a_1 must be chosen if, and only if, $P(E_1 | X=x) / P(E_2 | X=x) > \ell(c_{12}) / \ell(c_{21})$. Equivalently, a_1 is the best action if $P(X=x | E_1) / P(X=x | E_2) > (p_2 / p_1)(\ell(c_{12}) / \ell(c_{21}))$. Here, statistical knowledge is useful to propose the models $P(X=x | E_1)$ and $P(X=x | E_2)$ describing how likely a grade is under each of the scenarios E_1 and E_2 . The decision is thus based on the likelihood ratio $P(X=x | E_1) / P(X=x | E_2)$. In this case, the cutoff value depends on both the ratio of the probabilities p_1 and p_2 and the relative importance of the possible mistakes. An interesting issue closely related to this problem is that of the test design. If the classification exam is properly designed, it would be reasonable to expect high grades among those candidates who later prove to be good students and lower grades among bad ones. If the classification exam works well as a screening tool, the ratio $P(X=x | E_1) / P(X=x | E_2)$ should be a nondecreasing function of x and then $P(X=x | E_1) / P(X=x | E_2)$ will be larger than $(p_2 / p_1)(\ell(c_{12}) / \ell(c_{21}))$ only if x is larger than a cutoff given by $b((p_2 / p_1)(\ell(c_{12}) / \ell(c_{21})))$ where the function b depends on how well the exam discriminates between good and bad candidates, a characteristic that can be analyzed through statistical methods. This decision rule leads to the expected result: candidates with a high grade are admitted whereas those with a low grade are rejected. The more subtle issue is related to the cutoff value. How high should a grade be for a candidate to be admitted? Under this decision-theoretical framework, it is clear that the best rule is not necessarily to admit a candidate only if she passes the exam. Depending on the combination of the prior probabilities p_2 and p_1 , losses $\ell(c_{12})$ and $\ell(c_{21})$, as well as on the function b , it might be optimal to admit some candidates who do not pass the exam or to reject some who pass the exam with a relatively low grade. This procedure proposes a solution for the choice of the cutoff value in a straightforward manner which makes it clear

what the components of the problem are and how these should be taken into account.

Some Applications

The literature offers a variety of applications of decision theory in education, mostly in the field of education research. In particular, van der Linden (1991, 1997) provides a comprehensive review of applications to test theory. More recently, Vos (1999) and Vos and Glas (2000) propose a sequential decision procedure to develop adaptive mastery tests, whereas Segall (2004) applies ideas from decision theory to evaluate a new sharing item-response theory model (see also van der Linden (1998)). A common feature of these contributions (as well as many others of this type) is that the role of decision maker is played by the researcher, the instructor, or even the computer system used to administer the test.

From a different perspective, but also in the field of test design, Bernardo (1998) discusses several grading procedures for the case of multiple-choice examinations. There, the student, as the decision maker, must devise a strategy to answer each item of the examination. In this context, however, the corresponding decision problem is defined by the instructor, who determines the set of possible answers A , the set of uncertain events E , the consequences set C , and, very importantly, the utility function (the grading-score mechanism). If the student is a coherent decision maker in the sense of axiomatic decision theory, she must choose the answer maximizing her expected utility. This expectation is computed by combining the grading-score function of the test and the subjective-probability distribution describing the student's knowledge.

Bernardo shows how the utility function can be defined in order to discourage guessing. He also discusses the cases where the aim of the student is either to get the highest mark or just to pass the examination. In addition, and elaborating on an original idea by de Finetti (1965), he explores the results obtained when the student is asked not only to mark the correct option but to report her entire probability distribution over the set of possible answers for each item of the test.

Current Developments

Decision theory is an active field of research. The list of problems under current investigation include: experimental studies of individual and group behavior; use of different methods to understand human judgments and decisions; discussion of alternative normative models; and applications of the theory to a wide variety of subjects.

Specifically, alternative sets of axioms are being explored to produce a structure able to encompass the widest range of real problems; Lipman (1999) and

Dubois *et al.* (2003) are two examples of this type of investigation. Another line of research has to do with elicitation of both utilities and probabilities. Clemen and Reilly (1999) discuss the use of copulas for this purpose whereas Bleichrodt and Pinto (2000) explore a specific procedure in the context of medical decisions. In addition, for some decision problems, maximization of the expected utility can only be accomplished through numerical or simulation methods (see Bielza *et al.* (1999), for instance). Beyond the limits of the theory described in this article, there are researchers who assume that consequences might include several facets, and preferences for each one of them are described through a different utility function. This framework is known as a multi-objective decision problem and there is no generally accepted solution in such a setting (see Dyer *et al.*, 1992). Another problem whose general solution is yet to be formulated is that of group decision. If the decision maker is a committee where each member has a different utility function or a different probability distribution, how should these be combined to choose the action with the largest expected utility? Will this combination be an optimal solution in practice? Weerahandi and Zidek (1981) propose a solution for this problem; Hollenbeck *et al.* (1998) take into account the structure of the group through multilevel theory.

Concerning Bayesian statistics, the statistical ramification of decision theory, current research also includes alternative axiomatic formulations (see Karni, 2007, for a recent example), elicitation techniques (Garthwaite *et al.*, 2005), and applications in an ever-increasing number of fields. There is also a strong research effort directed toward statistical computing. Specifically, simulation techniques for the calculation of posterior distributions using Markov Chain Monte Carlo (MCMC) methods have been successfully employed over the last 15 years (Bhattacharya and Haslett, 2007, provide an example).

See also: Bayesian Statistical Analysis; Classical Test Theory Reliability; Item Response Theory.

Bibliography

- Bernardo, J. M. (1998). A decision analysis approach to multiple-choice examinations. In Girón, J. (ed.) *Applied Decision Analysis*, pp 195–208. Boston: Kluwer.
- Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. Chichester: Wiley.
- Bhattacharya, S. and Haslett, J. (2007). Importance re-sampling MCMC for cross-validation in inverse problems. *Bayesian Analysis* **2**, 385–408.
- Bielza, C., Mueller, P., and Rios Insua, D. (1999). Monte Carlo methods for decision analysis with applications to influence diagrams. *Management Science* **45**, 995–1007.
- Bleichrodt, H. and Pinto, J. L. (2000). A parameter-free elicitation of the probability weighting function in medical decision analysis. *Management Science* **46**, 1485–1496.

- Clemen, R. T. and Reilly, T. (1999). Correlations and copulas for decision and risk analysis. *Management Science* **45**, 208–224.
- de Finetti, B. (1965). Methods for discriminating levels of partial knowledge concerning a test item. *British Journal of Mathematical and Statistical Psychology* **18**, 87–123.
- Dubois, D., Fargier, H., and Perny, P. (2003). Qualitative decision theory with preference relations and comparative uncertainty: An axiomatic approach. *Artificial Intelligence* **148**, 219–260.
- Dyer, J. S., Fishburn, P. C., Steuer, R. E., Wallenius, J., and Zionts, S. (1992). Multiple criteria decision making, multiattribute utility theory: The next ten years. *Management Science* **38**, 645–654.
- Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association* **100**, 680–700.
- Hollenbeck, J. R., Ilgen, D. R., LePine, J. A., Coquitt, J. A., and Hedlund, J. (1998). Extending the multilevel theory of team decision making: Effects of feedback and experience in hierarchical teams. *Academy of Management Journal* **41**, 269–282.
- Karni, E. (2007). Foundations of Bayesian theory. *Journal of Economic Theory* **132**, 167–188.
- Lindley, D. V. (1972). *Bayesian Statistics, A review*. Philadelphia, PA: SIAM.
- Lipman, B. L. (1999). Decision theory without logical omniscience: Toward an axiomatic framework for bounded rationality. *Review of Economic Studies* **66**, 339–361.
- Novick, M. R. and Lindley, D. V. (1978). The use of more realistic utility functions in educational applications. *Journal of Educational Measurement* **15**, 181–191.
- Sawyer, R. (1996). Decision theory models for validating course placement tests. *Journal of Educational Measurement* **33**, 271–290.
- Schneider, S. H. (2002). Can we estimate the likelihood of climatic changes at 2100? *Climatic Change* **52**, 441–451.
- Segall, D. O. (2004). A sharing item response theory model for computerized adaptive testing. *Journal of Educational and Behavioral Statistics* **29**, 439–460.
- van der Linden, W. J. (1991). Applications of decision theory to test-based decision making. In Hambleton, R. K. and Zaal, J. N. (eds.) *Advances in Educational and Psychological Testing*, pp 129–156. Boston, MA: Kluwer.
- van der Linden, W. J. (1997). Decision theory in educational testing. In Keeves, J. P. (ed.) *Educational Research, Methodology, and Measurement: An International Handbook*, 2nd edn., pp 725–736. Oxford: Elsevier Science.
- van der Linden, W. J. (1998). A decision theory model for course placement. *Journal of Educational and Behavioral Statistics* **23**, 18–34.
- Vos, H. J. (1999). Applications of Bayesian decision theory to sequential mastery testing. *Journal of Educational and Behavioral Statistics* **24**, 271–292.
- Vos, H. J. and Glas, C. A. W. (2000). Testlet-based adaptive mastery testing. In van der Linden, W. J. and Glass, C. A. W. (eds.) *Computerized Adaptive Testing: Theory and Practice*, pp 289–310. Boston, MA: Kluwer.
- Weerahandi, S. and Zidek, J. V. (1981). Multi-Bayesian statistical decision theory. *Journal of the Royal Statistical Society A* **144**, 85–93.

Further Reading

- Rao, C. R. and Sinharay, S. (eds.) (2007). *Handbook of Statistics*, Vol. 26: Psychometrics. Amsterdam: North-Holland.

Relevant Websites

- <http://www.eric.ed.gov/> – Education Resources Information Center (ERIC).

Design of Experiments

C A Albers and T R Kratochwill, University of Wisconsin–Madison, Madison, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Causal studies – A type of research study designed to determine whether one or more variables actually cause changes in one or more outcome variables.

Control group – The group that does not receive the intervention within a research study.

Descriptive studies – A type of research study that involves the collection of data to simply describe a phenomenon.

Internal validity – The extent to which a researcher can accurately state that the independent variable caused the changes in the outcome variable(s).

Intervention group – The group that receives the intervention within a research study.

Quasi-experimental design – A research design in which assignment to intervention or control groups is not done by random assignment.

Random assignment – A process or procedure in which participants are assigned to the intervention or control group by a random process.

Randomized experimental design – A research study in which assignment to intervention or control groups is done by random assignment.

Relational studies – A type of research study that attempts to quantify the direction and strength of a relationship between two or more variables.

The topic of research methodology has never been more relevant to education than at the current time, as the shift within education policy is the movement toward evidence-based practice. To establish what actually consists of evidence-based practice, high-quality evidence from research is a necessity. For research to be considered high quality, it is necessary that researchers adhere to certain standards and indicators of quality; these quality indicators and standards ultimately dictate certain features of research designs and methodology. In addition, statistical analyses are implicitly linked to research design, in that these procedures are intended to remove variability (e.g., error) within the system that cannot be addressed by the research design alone. In fact, an appropriate research design and process for data collection allows a researcher to take full advantage of the wide range of new statistical analyses that are available. Consequently, the design of an experiment and the selection of appropriate statistical procedures are critical in establishing evidence-based

practices. In this article, we examine (1) the evolving role of research within education, (2) the legal and professional guidelines for conducting education research, and (3) some available research designs and statistical procedures that are used to conduct education research.

The Evolution of Education Research

Historically, education research has been believed to suffer from a lack of scientific rigor. During the past decade, however, there has been an increased emphasis on developing a scientific foundation for education interventions and practices. The impetus for this increase within education was the passage of the No Child Left Behind Act (NCLB) of 2001, which codified that teachers must use evidence-based teaching practices to ensure that students receive the highest-quality instruction (see Eisenhart and Towne (2003) for a historical perspective of the evolution of the definition of scientifically based education research in NCLB).

Experimental Design and Statistical Procedure Considerations

Consideration of experimental design is critical in education research. When properly designed, an educational experiment will ensure a sufficient amount of data and the right data to properly, clearly, and efficiently answer the research question of interest. In essence, a properly designed experiment will allow the investigator to improve the precision of the answer to the examined research question. Thus, scientifically rigorous experimental designs and statistical analyses enhance the validity of the researcher's conclusions.

Quality Indicators

Numerous governmental and professional organizations have developed documents and guidelines that outline quality indicators of high-quality research to assist practitioners and researchers in developing and identifying evidence-based practices. An astute researcher will be familiar with these quality indicators and proactively consider these indicators when designing an experiment. For example, the What Works Clearinghouse, which evaluates education programs for the United States Department of

Education, developed evidence standards for reviewing group randomized controlled trial studies (Institute for Education Sciences (IES, 2008)).

Similarly, Division 16 (school psychology) of the American Psychological Association, the Society for the Study of School Psychology, and the National Association of School Psychologists sponsored a task force to develop a *Procedural and Coding Manual* for review of evidence-based programs in school psychology (Task Force on Evidence-Based Interventions in School Psychology, 2003, 2009). The manual addresses numerous issues that should be considered *a priori* by researchers, including randomized and nonrandomized designs, statistical treatment and data analysis, measurement issues, comparison groups, and appropriate unit of analysis, among others. Other organizations that developed quality indicators relating to education research include the Council for Exceptional Children (2004) and the National Research Council (2002).

Research Designs

Clearly, an in-depth discussion of the entire range of research designs and the corresponding nuances associated with each research design is beyond the scope of this article. Instead, we provide a broad overview of some common research designs and appropriate statistical procedures associated with these research designs. Research designs can be categorized in various ways; in the broadest sense, research designs can be classified into three general categories, including (1) descriptive, (2) relational, and (3) causal. The selected research design should be congruent with the research question of interest, as well as providing a match with the desired scientific rigor and the practicality of conducting the study.

Descriptive Studies

Descriptive studies are intended to collect data to simply describe a phenomenon. An example of a descriptive study within education would be determining the number of English-language learners attending school within a certain state, which could then have implications for instructional practices. Analysis of descriptive studies generally consists of the use of descriptive statistics, which allows for comparisons across people or groups. Common statistics include measures of distribution, central tendency, and variability (e.g., range and standard deviation).

Relational Studies

Relational studies attempt to quantify the direction and strength of a relationship between two or more variables.

Correlation studies are most often defined as a nonexperimental design as internal validity is weak, thereby indicating that making causal statements between cause and effect is not recommended. An example of a relational study within education would be the examination of the relationship between English-language proficiency (i.e., proficient in English vs. not proficient in English) and academic achievement. Analysis of relational data frequently includes the use of various correlation techniques (e.g., Pearson product moment correlation).

Causal Studies

Causal studies are designed to determine whether one or more variables actually cause changes in one or more outcome variables. Internal validity, which is the extent to which we can accurately state that the independent variable caused the changes in the outcome variable(s), is a critical component within causal studies. To establish internal validity (and thus a causal relationship), three criteria need to be adequately addressed. These criteria consist of:

1. temporal precedence, which is establishing that the cause (i.e., independent variable) occurs before the effect (i.e., outcome);
2. establishing that the cause and effect are related and/or covary; and
3. establishing that there are no plausible alternative explanations.

The most appropriate way to address these criteria is through the use of various research designs; the stronger the research design, the more likely it is that the researcher is able to control sources of error in their methods and results. Specific procedures for controlling this variance include: (a) randomization, (b) accounting for conditions/factors into the design as independent variables, (c) holding conditions/factors constant, and (d) statistical adjustments. Whereas (a), (b), and (c) can be controlled when structuring the research design, (d) is conducted during the analysis stage.

A variety of experimental research designs can be used within causal studies; these designs can be classified as being (1) randomized experimental, (2) quasi-experimental, or (3) single-case designs. The primary and critical difference between randomized experimental and quasi-experimental designs is the presence of random assignment for randomized experimental designs, and the presence of a control group or the use of multiple measures in quasi-experimental designs. Although single-case designs are frequently classified as quasi-experimental designs, we include them as a separate classification within this article as they can serve as an alternative to using large, aggregate group designs. The variations within each type of design are discussed in more detail below.

Randomized Experimental Design

In a randomized experiment, participants are assigned to receive the control or intervention condition by a random process. The critical component within such a design is that the assignment to a control or intervention condition is truly random; this random assignment typically allows for the strongest causal inferences (i.e., high internal validity) to be made free of extraneous assumptions. There are numerous methods of implementing random assignment; Shadish *et al.* (2002: 295) identify seven common different varieties of random assignment, including:

1. simple random assignment,
2. restricted random assignment to force equal sample sizes,
3. restricted random assignment to force unequal sample sizes,
4. batch randomization,
5. trickle process randomization,
6. adaptive randomization strategies, and
7. random assignment from matches or strata.

However, even with randomized designs, various threats to validity (e.g., differential attrition, small sample size, and lack of external validity) can still occur and need to be considered. Some of the more frequently used randomized experimental designs are described in more detail below; however, it should be noted that each of these designs can also be classified as quasi-experimental designs if random assignment is not utilized.

Control group designs

Common control group designs are the posttest-only control group design and the pretest–posttest control group design. However, because the assumption that the groups were equivalent prior to the implementation of the intervention cannot be made as a result of not using a randomized assignment process, the selection and assignment process to groups can pose a threat to the experiment's internal validity. Consequently, causal statements attributing differences in outcomes must be made with extreme caution.

In the posttest-only control group design, pretests are not administered. Instead, the intervention group (or multiple groups) and the control group are only measured on the construct of interest (e.g., test) following the implementation of the intervention. Thus, the posttest scores are used to determine the effects of the intervention. This can be represented as:

$$G_1 \ X \ O_{\text{post}}$$

$$G_2 - O_{\text{post}}$$

where G_1 is group 1 (intervention group), G_2 is group 2 (control group), X is the implemented intervention, O_{post}

is the group 1 outcome measure (i.e., posttest), and O_{post} is the group 2 posttest score.

In the pretest–posttest control group design, the intervention group (two or more groups) and control groups are administered both the pretest and the posttest. Obtaining pretest scores allows the researcher to examine the equivalency of the intervention group and the control group before the intervention was implemented. Having these pretest scores allow for a variety of different analysis procedures to be used, as the pretest scores can be used as a statistical control in the analyses. A pretest–posttest control group design can be represented as:

$$G_1 \ O_{\text{pre}} \ X \ O_{\text{post}}$$

$$G_2 \ O_{\text{pre}} - O_{\text{post}}$$

where G_1 is group 1 (intervention group), G_2 is group 2 (control group), X is the implemented intervention, O_{pre} is the pretest score, and O_{post} is the posttest score.

Factorial designs

Factorial designs are utilized when it is desirable to include two or more independent (i.e., intervention) variables in the design. Within factorial designs, a factor refers to the independent variable. In a study with two independent variables, each of which has two levels, one would have a 2×2 factorial design; altogether, there are four different groups, each of which receives a different combination of levels of factors. A posttest-only 2×2 factorial design can be represented as (with different combinations of levels of factors indicated by subscripts):

$$G_1 \ X_{1,1} \ O_{\text{post}}$$

$$G_2 \ X_{1,2} \ O_{\text{post}}$$

$$G_3 \ X_{2,1} \ O_{\text{post}}$$

$$G_4 \ X_{2,2} \ O_{\text{post}}$$

whereas a pretest–posttest 2×2 factorial design can be represented as:

$$G_1 \ O_{\text{pre}} \ X_{1,1} \ O_{\text{post}}$$

$$G_2 \ O_{\text{pre}} \ X_{1,2} \ O_{\text{post}}$$

$$G_3 \ O_{\text{pre}} \ X_{2,1} \ O_{\text{post}}$$

$$G_4 \ O_{\text{pre}} \ X_{2,2} \ O_{\text{post}}$$

Factorial designs can consist of more than two levels for each of the independent variables. For example, a 3×4 factorial design would require 12 intervention groups. Given that the number of groups can become quite large as the number of levels and factors increase, there are

a variety of designs that can be used to minimize the necessary number of groups. An incomplete factorial design allows for certain intervention combination groups to not be implemented; in this situation, it would be possible to make the excluded group a control group.

Blocked designs

Within blocked designs, blocking is the arrangement of individuals (i.e., units) into similar groups (i.e., blocks). This variable is referred to as a blocking factor, which usually is a source of variability that is not of interest in the study; thus, a blocking factor can also be considered to be a nuisance factor. A simple example of a blocking factor can be gender; a blocked pretest–posttest design in which gender is blocked (thus, the two groups are male and female) can be represented as:

$$B_1 \quad G_1 \quad O_{pre} \quad X \quad O_{post}$$

$$G_2 \quad O_{pre} - O_{post}$$

$$B_2 \quad G_1 \quad O_{pre} \quad X \quad O_{post}$$

$$G_2 \quad O_{pre} - O_{post}$$

where B_1 is the group in which gender is blocked to only include males, B_2 is the group in which gender is blocked to only include females, G_1 is subgroup 1 (intervention group; within both the male and female groups), G_2 is subgroup 2 (control group; within both the male and female groups), X is the implemented intervention, O_{pre} is the pretest score, and O_{post} is the posttest score.

Quasi-Experimental Designs

In many cases, particularly in some educational settings, random assignment is neither feasible nor possible. In this case, quasi-experimental designs can be considered. The core component in a quasi-experimental study is that the participants are not randomly assigned to conditions. Without careful use of statistical controls that account for varying degrees of pre-intervention equivalence, confounds may be present that negatively impact the validity of the results. Consequently, inferring a causal connection is not recommended; at a minimum, the researcher should stress that alternative explanations could account for the differences between the intervention and control groups.

Regression discontinuity designs

IES (2008) defines regression discontinuity designs as “designs in which participants are assigned to the intervention and the control conditions based on a cut-off score on a pre-intervention measure that typically assesses need or merit. This measure should be one that has a known functional relationship with the outcome of interest over

the range relevant for the study sample” (p. 1). The primary difference between a regression discontinuity design and a nonequivalent comparison group design is that in the former, assignment to intervention group is made based on the individual’s score on a pre-program measure. This strategy allows the researcher to target an intervention to a certain individual or group in need of the intervention without compromising internal validity.

Time-series designs

Quasi-experimental time-series designs can be one of the more powerful class of quasi-experimental designs, particularly when the design incorporates various components to strengthen the design elements (e.g., control series). These design elements include consideration of assignment to conditions, measurement considerations, the use of various comparison groups, and the scheduling of interventions (see Shadish *et al.* (2002) for an in-depth discussion of how to strengthen quasi-experimental designs).

Single-Case Research Designs

Although usually labeled a quasi-experimental time-series design, single-case research designs are described in this article as a separate form of research design (formerly termed single-subject or $N = 1$ research) that have a long and influential history in psychology and education (e.g., Kratochwill, 1978; Levin *et al.*, 2003) and can serve as an alternative to using large, aggregate group designs (Shadish and Rindskopf, 2007). Single-case research designs bear similarly to time-series design and have often been regarded as quasi-experimental because they usually do not (but could) include randomization in the experiment. In the single-case design, replication is scheduled to help rule out various threats to validity. Single-case designs can involve a single participant or group as the unit but differ from repeated measures and hierarchical linear modeling (HLM) designs because multiple observations are taken over a long period of time within a design structure of replication and/or randomization of the conditions of the experiment.

There are three major types of single-case design that incorporate replication as a design characteristic and are used in applied and clinical work: the A-B-A-B within-series design, the multiple-baseline design, and the alternating treatment design. The A-B-A-B design represents an attempt to measure a baseline (the first A), a treatment measurement (the first B), the withdrawal of treatment (the second A), and the reintroduction of treatment (the second B) and involves replication of the treatment within one or more series on an individual participant or group. The A-B-A-B format is regarded as a within-series design structure since the intervention is replicated within the repeated assessment across the four phases. Causal inferences are drawn from the replication at three

points in time, going from A to B, from B to A, and from A to B. The multiple-baseline design involves replication across participants, settings, or behaviors in a single participant or groups. In this design, the treatment is staggered and introduced sequentially in a time-lagged fashion usually on at least three separate series, although more add credibility to causal inference. This format involves a combined-series design structure in that a within-series (i.e., A to B) and replication across series/units (participants, settings, and behavior) lagged in time is scheduled. In the alternating treatment design, following a baseline phase, the treatments are alternated in rapid succession (compared to the ABAB design which has more within phase observations or measurements) allowing a comparison of the treatment to baseline or an alternative treatment over repeated observations (e.g., ABABABABAB and ABCBCBCBC). The alternating treatment phases can be counterbalanced or randomized.

In each of these designs the researcher must attend to various features of the data, including mean changes among phases, trend, variability, and autocorrelation in the data. In general, these designs can be improved by structuring various forms of randomization into the design structure, thereby improving the internal validity of the design and

allowing certain forms of statistical analysis, such as randomization tests, to be conducted (see Kratochwill and Levin, in press).

Statistical Analysis Considerations Within Causal Studies

Selecting the appropriate research design, assigning individuals to intervention and control groups, and collecting appropriate data are necessary, but not sufficient, steps to answering a research question. Instead, these procedures are prerequisites for selecting the appropriate method of statistical analysis. There are a tremendous number of possible statistical techniques that can be used to analyze data, including parametric and nonparametric techniques, *t*-tests, analysis of variance (ANOVA), analysis of covariance (ANCOVA), regression analysis, and numerous multivariate methods; **Figures 1** and **2** provide an overview of many of these techniques. Erceg-Hurn and Mirosevich (2008) also provide a concise overview of how to select modern and robust statistical methods.

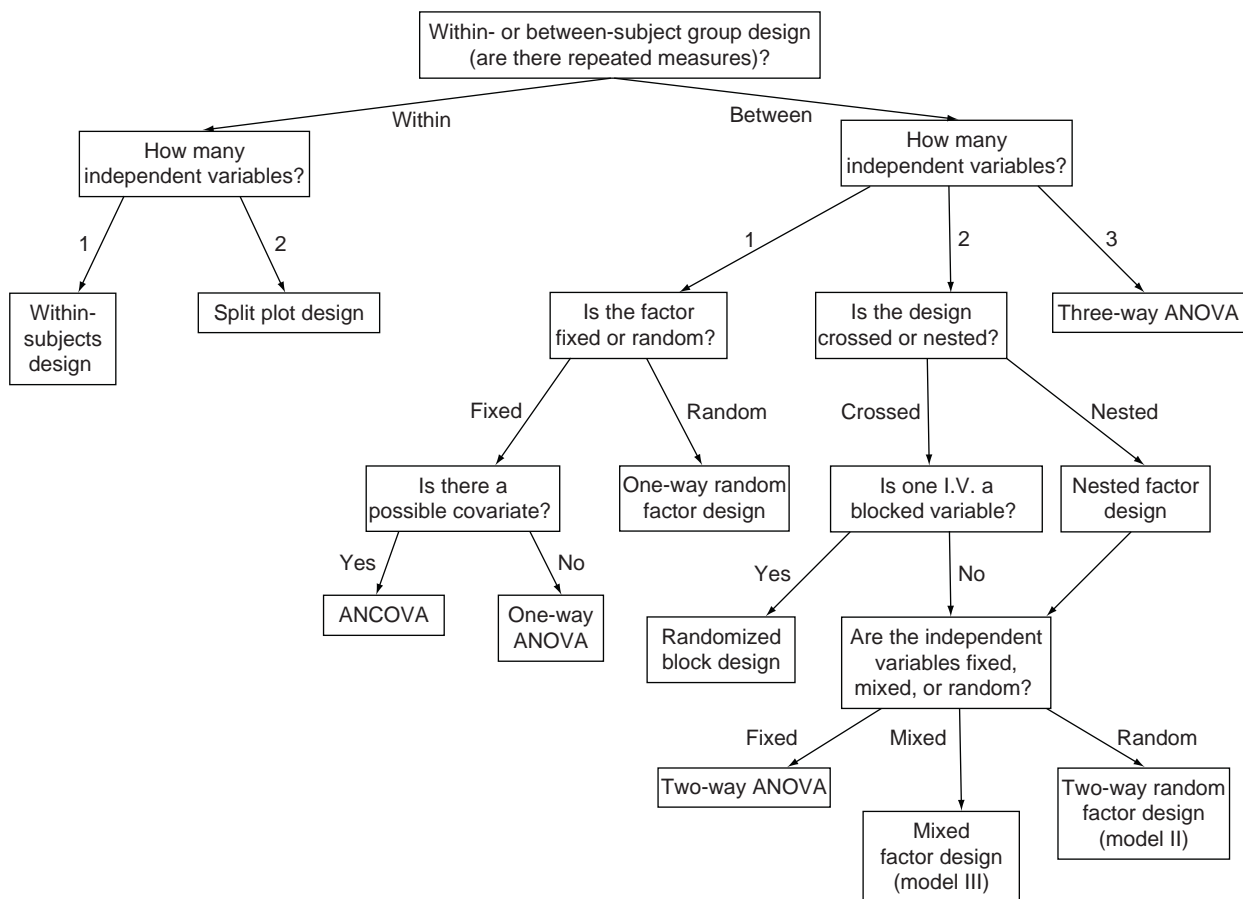


Figure 1 Group design variations and example analysis applications.

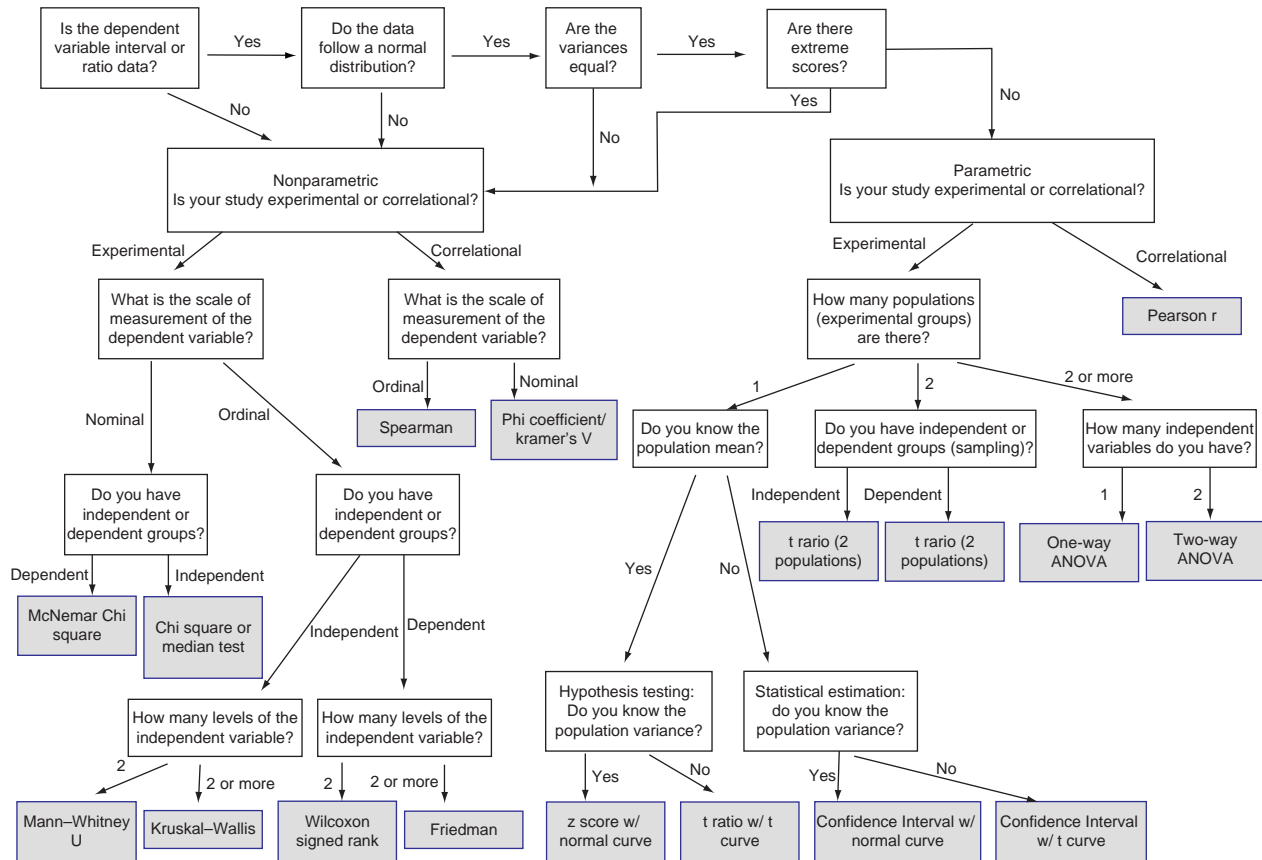


Figure 2 Selection of statistical analysis procedures.

To illustrate the link between research design and statistics, we provide an overview of a research study (Sinclair *et al.*, 1998) that was identified by the What Works Clearinghouse as meeting evidence standards for demonstrating the effectiveness of a high school dropout prevention program. Identified as a randomized controlled trial (using a posttest only group design), participants were randomly assigned using a stratified selection procedure to the intervention group and control group. No pretest measures were obtained, whereas three outcome (i.e., posttest) measures (i.e., participation in school, school performance, and identification with school) were collected. Thus, the design can be represented as:

$$G_1 \times O_{1,2,3}$$

$$G_2 - O_{1,2,3}$$

where G_1 represents the intervention group, G_2 represents the control group, X represents the intervention, O_1 represents posttest measure 1, O_2 represents posttest measure 2, and O_3 represents posttest measure 3.

Analyses, including t -tests and chi-square tests of independence, examined differences between intervention and control groups. Results indicated a significant difference between intervention and control groups on the measure

of participation in school, $\chi^2(1) = 6.87, p < 0.05$. On one of the measures (i.e., class credits) of school performance, the students who received the intervention earned significantly more credits than the students assigned to the control group, $t(90) = 4.01, p < 0.05$. On the third posttest measure (i.e., identification with school), there were no significant differences between intervention and control students.

Summary

The expectations for what constitutes acceptable education practices have undergone a significant and meaningful transformation during the past decade. Whereas past practices were based on a belief in doing what we think should work, current practices and policies are guided by the belief in doing what scientifically sound research has clearly demonstrated to be effective practices. This mandate for using evidence-based practices, which became the expectation as a result of the passage of NCLB in 2001, requires the application of strong research methodology to determine evidence-based practices.

Although the federal government and various professional organizations have clearly articulated a preference

for using randomized experimental designs to determine evidence-based practices, it would be erroneous to automatically presume that such designs are the best of all possible designs. Although randomized experimental designs allow for the strongest causal inferences to be made free of extraneous assumptions, using such a design does not guarantee that various threats to the study's validity are not present. Instead, each of the research designs should be recognized to have relative strengths and weaknesses, with the results being interpreted with these in mind.

To continue establishing evidence-based education practices, there clearly is a need for additional scientifically sound research to be conducted. Critical to this agenda is the use of the appropriate research designs and statistical procedures to answer the question of interest. Although randomized experimental designs have implicitly and explicitly been identified by some to be the gold standard of research designs, there is a growing recognition that the appropriate use of alternate research designs, especially when combined with more advanced statistical techniques, can contribute to the identification of evidence-based practices.

See also: Analysis of Covariance; Analysis of Variance; Causal Inference; Generalized Linear Models; Multivariate Analysis of Variance; Multivariate Linear Regression; Nonparametric Statistical Methods; Univariate Linear Regression.

Bibliography

- Council for Exceptional Children (2004). *Quality Indicators for Research in Special Education and Guidelines for Evidence-Based Practices: Executive Summary*. Arlington, VA: Council for Exceptional Children. http://education.uoregon.edu/grantmatters/pdf/DR/Exec_Summary.pdf (accessed August 2009).
- Eisenhart, M. and Towne, L. (2003). Contestation and change in national policy on "scientifically based" education research. *Educational Researcher* 32(7), 31–38.
- Erceg-Hurn, D. M. and Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist* 63, 591–601.
- Institute of Education Sciences (IES) (2008). *What Works Clearinghouse Evidence Standards for Reviewing Studies*. Washington, DC: United States Department of Education.
- Kratochwill, T. R. (1978). Foundations of time-series research. In Kratochwill, T. R. (ed.) *Single Subject Research: Strategies for Evaluating Change*, pp 1–100. New York: Academic Press.
- Kratochwill, T. R., and Levin, J. R. (in press). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods*.
- Levin, J. R., O'Donnell, A. M., and Kratochwill, T. R. (2003). Educational/psychological intervention research. In Reynolds, W. and Miller, G. (eds.) *Handbook of Psychology: Volume 7. Educational Psychology*, pp 557–581. Hoboken, NJ: Wiley.
- National Research Council (2002). *Scientific Research in Education*. In Shavelson, R. J. and Towne, L. (eds.) Washington, DC: National Academy Press, Committee on Scientific Principles for Education Research, Center for Education/Division of Behavioral and Social Sciences and Education. <http://books.nap.edu/openbook.php?isbn=0309082919> (accessed August 2009).
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.
- Shadish, W. R. and Rindskopf, D. M. (2007). Methods for evidence-based practice: Quantitative synthesis of single-subject designs. *New Directions for Evaluation* 113, 95–109.
- Sinclair, M. F., Christenson, S. L., and Hurley, C. M. (1998). Dropout prevention for youth with disabilities: Efficacy of a sustained school engagement procedure. *Exceptional Children* 65, 7–21.
- Task Force on Evidence-Based Interventions in School Psychology (2003, 2009). *Procedural and Coding Manual*. http://www.indiana.edu/~ebi/documents/_workingfiles/EBImanual1.pdf (accessed August 2009).

Further Reading

- Boruch, R., de Moya, D., and Snyder, B. (2002). The importance of randomized field trials in education and related areas. In Mosteller, F. and Boruch, R. (eds.) *Evidence Matters: Randomized Trials in Education Research*, pp 50–79. Washington, DC: Brookings Institution Press.
- Coalition for Evidence-Based Policy and The What Works Clearinghouse (2005). *Reporting the Results of Your Study: A User-Friendly Guide for Evaluations of Educational Programs and Practices*. Washington, DC: Coalition for Evidence-Based Policy and The What Works Clearinghouse. http://ies.ed.gov/ncee/wwc/pdf/guide_SRF.pdf (accessed August 2009).
- Cone, J. D. (2001). *Evaluating Outcomes: Empirical Tools for Effective Practice*. Washington, DC: American Psychological Association.
- Cook, T. D. and Payne, M. R. (2002). Objecting to the objections to using random assignment in education research. In Mosteller, F. and Boruch, R. (eds.) *Evidence Matters: Randomized Trials in Education Research*, pp 150–178. Washington, DC: Brookings Institution Press.
- Good, P. I. and Hardin, J. W. (2003). *Common Errors in Statistics (and How to Avoid Them)*. Hoboken, NJ: Wiley.
- Hayes, S. C., Barlow, D. H., and Nelson-Gray, R. O. (1999). *The Scientist-Practitioner: Research and Accountability in the Age of Managed Care*, 2nd edn. Needham Heights, MA: Allyn and Bacon.
- Institute of Education Sciences (2005). *Key Items to Get Right When Conducting a Randomized Controlled Trial in Education*. Washington, DC: United States Department of Education. http://ies.ed.gov/ncee/wwc/pdf/guide_RCT.pdf (accessed August 2009).
- Kratochwill, T. R. and Levin, J. R. (eds.) (1992). *Single-Case Research Design and Analysis: New Directions for Psychology and Education*. Hillsdale, NJ: Erlbaum.
- National Institute for Literacy (2005). *What is Scientifically Based Research? A Guide for Teachers*. Jessup, MD: National Institute for Literacy. http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/29/e5/80.pdf (accessed August 2009).
- Phye, G. D., Robinson, D. H., and Levin, J. R. (eds.) (2005). *Empirical Methods for Evaluating Educational Interventions*. San Diego, CA: Elsevier Academic Press.
- Shadish, W. R. and Cook, T. D. (2009). The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology* 60, 607–629.
- Shavelson, R. J. and Towne, L. (2002). *Scientific Research in Education*. Washington, DC: National Academy Press.

Relevant Website

<http://ies.ed.gov> – The What Works Clearinghouse.

Discrete Probability Distributions

S Sinharay, ETS, Princeton, NJ, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Estimate – A function of sample data used for approximating a population parameter.

Moment – The n th moment about a value c of a random variable that has a probability density function $f(x)$ is $\mu'_n = \int_{-\infty}^{\infty} (x - c)^n f(x) dx$.

A discrete random variable (denoted X) is one that can take one of an enumerable or countably infinite set of values x_i , $i = 1, 2, \dots$, each with a specific probability p_i so that the probabilities over all the values add up to 1 (i.e., $\sum_i p_i = 1$). The function that specifies the probabilities for the possible values is defined as the probability mass function (PMF). The cumulative distribution function (CDF) of a discrete distribution is a step function. Discrete random variables occur naturally in statistics. One reason of this is rounding of continuous variables (e.g., height, a continuous variable, is measured in inches). Holland (1979) commented that the world is full of discrete data. Hence, several types of discrete random variables have been suggested in the literature. There also exist several statistical methods specifically designed to handle discrete data – see, for example, Bishop *et al.* (1975).

In the following sections, several discrete distributions are discussed. For each distribution, its genesis, PMF, moments, estimation of the moments, methods to generate values of a random variable following that distribution, and relationships to other distributions (if any) are discussed.

Bernoulli Distribution

The Bernoulli distribution is the most basic discrete distribution. A variable that follows the distribution can take one of two possible values, 1 (usually called a success) or 0 (failure), where the probability of success is p , $0 < p < 1$. An example of a Bernoulli random variable (that is a variable that follows the Bernoulli distribution) is the outcome of a coin toss, where the outcome is either a head (success) or a tail (failure) and the probability of a head is a number between 0 and 1.

The PMF of a Bernoulli distribution is given by $P(X = x) = p^x(1-p)^{1-x}$, where x can be either 0 or 1. The CDF $F(x)$ of the distribution is 0 if $x < 0$, $1-p$ if

$0 \leq x < 1$, and 1 if $x \geq 1$. The mean and the variance of the distribution are p and $p(1-p)$, respectively.

If X_1, X_2, \dots, X_n are independent Bernoulli random variables, all with success probability p , then $\sum_{i=1}^n X_i$, which denotes the number of successes among the X_i 's, is a binomial random variable with number of trials = n and success probability p . The maximum likelihood estimate of p from a sample x_1, x_2, \dots, x_n from the Bernoulli random variable is the sample mean $\frac{1}{n} \sum_i x_i$, which is the proportion of successes in the sample.

To generate a random number X from the Bernoulli distribution, one has to first generate a random number R from the $U(0, 1)$ distribution (the uniform distribution between 0 and 1). If R is less than p , then X is set to 0; otherwise, X is set to 1.

Binomial Distribution

As mentioned above, a binomial distribution is the distribution of the sum of n independent Bernoulli random variables, all of which have the same success probability p . The quantity n is called the number of trials and p the success probability.

The PMF of a binomial distribution is given by

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n.$$

The mean and variance of the distribution are np and $np(1-p)$. For large values of n , the probability that the binomial random variable X takes the value x can be computed by approximating X by a normal variable Y with mean np and variance $np(1-p)$ and computing the probability that Y lies between $x - 0.5$ and $x + 0.5$.

The maximum likelihood estimate of p from a sample x_1, x_2, \dots, x_n from the binomial distribution is the ratio of the sample mean $\frac{1}{n} \sum_i x_i$ and n . To generate a random number from a binomial distribution, one can generate n Bernoulli random variables, all with probability p , and add them up. Alternatively, one can compute the CDF of the binomial distribution, generate $R \sim U(0, 1)$, and then take x as the random number from the binomial distribution if the CDF for $x - 1$ is less than R and the CDF for x is greater than or equal to R (this is often called the CDF-inversion method).

Several applications of the binomial distribution to educational data can be found in Novick and Jackson (1974). In one such application, the number of words

spelled correctly (out of a total of n words) by a student, where each word is equally difficult to spell, is assumed to follow the binomial distribution.

Geometric Distribution

If a sequence of independent Bernoulli random variables is generated with the same success probability p , $0 < p < 1$, the number of trials or failures before the first success follows a geometric distribution.

The PMF of a variable X following the geometric distribution is given by $P(X = x) = p(1 - p)^x$, $x = 0, 1, 2, \dots, \infty$. (Note that the number of possible values of a geometric random variable is countably infinite.)

The mean and variance of a geometric distribution are $\frac{1-p}{p}$ and $\frac{1-p}{p^2}$.

The geometric distribution is a special case of the negative binomial distribution. The sum of several independent geometric random variables with the same success probability is a negative binomial random variable.

The maximum likelihood estimate of p from a sample from the geometric distribution is $\frac{1}{1+\bar{x}}$, where \bar{x} is the sample mean.

The CDF of the geometric distribution for an integer r is given by $P(X \leq r) = 1 - (1-p)^{r+1}$. Using this expression, a random number from a geometric distribution can be generated by first generating $R \sim U(0, 1)$ and then computing the integer part of $\frac{\log(R)}{\log(1-p)}$.

For an example in educational testing where the geometric distribution can arise, see Lewis (2007).

Hypergeometric Distribution

Let an investigator draw a sample of n items without replacement from a population of N elements, of which M has a certain attribute. The number of elements in the sample with the attribute is a hypergeometric random variable. This distribution could arise when, for example, a group of examinees contains M examinees who passed a test and $N-M$ who failed; if one draws a sample of n examinees from the group, the number of examinees who passed in the sample is a hypergeometric random variable.

The PMF of the distribution is given by

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}.$$

The mean and variance of the distribution are $\frac{nM}{N}$ and $\frac{(nM/N)(1-M/N)(N-n)}{N-1}$, respectively.

When $\frac{n}{N}$ is small and N is large, the hypergeometric distribution can be approximated by a binomial random variable with n trials and $p = \frac{M}{N}$.

The parameters of the distribution can be estimated from a sample of size n using the maximum likelihood

method. The estimate of N is the largest integer less than or equal to $\frac{nM}{x}$ and that of M is the largest integer less than or equal to $\frac{(N+1)x}{n}$.

To draw a random number from the distribution, one can construct a routine that simulates an experiment (by drawing elements at random) that will generate a hypergeometric random variable. Alternatively, one can compute the CDF of the distribution and then generate a hypergeometric random number with the help of a single $U(0, 1)$ random number and the CDF-inversion method.

Multinomial Distribution

The multinomial distribution is a multivariate generalization of the binomial distribution. Consider a trial that results in exactly one of some fixed finite number k of possible outcomes, with probabilities p_1, p_2, \dots, p_k (so that $p_i \geq 0$ for $i = 1, \dots, k$ and $\sum_{i=1}^k p_i = 1$), and there are n independent trials. Then let the random variables X_i indicate the number of times outcome number i was observed over the n trials. Then $X = (X_1, X_2, \dots, X_k)$ follows a multinomial distribution with parameters n and \mathbf{p} , where $\mathbf{p} = (p_1, p_2, \dots, p_k)$. An example where a multinomial random variable could occur is during the throw of a dice. Let X_i , $i = 1, 2, \dots, 6$, denote the number of times i is observed in n throws of a dice. Then $X = (X_1, X_2, \dots, X_6)$ has a multinomial distribution. If the dice is fair, then $p_i = \frac{1}{6}$ for all i .

The PMF of the multinomial distribution is given by

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

where $\sum_{i=1}^k x_i = n$. If $k = 2$, the multinomial distribution becomes a binomial distribution with n trials and success probability p_1 . When $X = (X_1, X_2, \dots, X_k)$ follows a multinomial distribution with the PMF given above, X_i follows a binomial distribution with n trials and success probability p_i . Hence, the mean and variance of X_i are np_i and $np_i(1 - p_i)$, respectively. The covariance between X_i and X_j is $-np_i p_j$.

The maximum likelihood estimate of p_i for a multinomial distribution is the ratio of the sample mean of x_i 's and n .

The straightforward way to generate a multinomial random variable is to simulate an experiment (by drawing n uniform random numbers that are assigned to specific bins according to the cumulative value of the \mathbf{p} vector) that will generate a multinomial random variable.

The multinomial distribution has found several applications in educational statistics. For example, this distribution is often applied to model the number of examinees who obtained a specific score (e.g., 70 out of 100) on an examination, or, the number of examinees who have a specific pattern of scores, for example, 0, 1, 1, 2, 1, 0, 0, \dots , in the items of an examination (see, e.g., Holland and Thayer (2000) and Holland (1990), for such applications).

Negative Binomial Distribution

As mentioned earlier, a negative binomial distribution is the distribution of the sum of independent geometric random variables. The number of failures before the n th success in a sequence of draws of Bernoulli random variables, where the success probability is p in each draw, is a negative binomial random variable. The PMF of the distribution is given by $P(X = x) = \binom{n+x-1}{n-1} p^n (1-p)^x$.

The mean and variance of a negative binomial distribution are $n \frac{1-p}{p}$ and $n \frac{1-p}{p^2}$.

The maximum likelihood estimate of p from a sample from the negative binomial distribution is $\frac{n}{n+\bar{x}}$, where \bar{x} is the sample mean.

If p is small, it is possible to generate a negative binomial random number by adding up n geometric random numbers. Another way is to generate a sequence of $U(0, 1)$ random variable values. When the number of those values less than p first reaches n , the number of values that are greater than p is a draw from the negative binomial distribution.

Poisson Distribution

The Poisson distribution describes the probability to find exactly x events in a given length of time if the events occur independently at a constant rate. In addition, the Poisson distribution can be obtained as an approximation of a binomial distribution when the number of trials n of the latter distribution is large, success probability p is small, and np is a finite number.

The PMF of the Poisson distribution is given by $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$, $x = 0, 1, \dots, \infty$, where λ is a positive number. Both the mean and variance of the Poisson distribution are equal to λ .

The maximum likelihood estimate of λ from a sample from the Poisson distribution is the sample mean.

If λ is large, the probability that a Poisson random variable X takes the value x can be obtained by approximating X by a normal variable Y with mean and variance λ and computing the probability that Y lies between $x - 0.5$ and $x + 0.5$.

It is possible to generate a Poisson random variable by generating a random number following the $U(0, 1)$ distribution and then using the CDF-inversion method.

The Poisson distribution has been applied in education, for example, by Novick and Jackson (1974).

Uniform (Discrete) Distribution

In fields such as survey sampling, the discrete uniform distribution often arises because of the assumption that each individual is equally likely to be chosen in the sample on a given draw. The PMF of a discrete uniform

distribution is given by $p(X = x) = \frac{1}{n+1}$, $x = 0, 1, \dots, n$, which implies that X can take any integer value between 0 and n with equal probability. The mean and variance of the distribution are $\frac{n}{2}$ and $\frac{n(n+2)}{12}$.

To generate a random number from the discrete uniform distribution, one can draw a random number R from the $U(0, 1)$ distribution, calculate $S = (n+1)R$, and take the integer part of S as the draw from the discrete uniform distribution.

Conclusions

Among the numerous discrete probability distributions, this article provides a brief overview of arguably the most popular eight such distribution, several of which have found applications in education (see, e.g., Evans *et al.* (2000), Johnson *et al.* (2005), and Johnson *et al.* (1997) for more details of these and many more discrete distributions). The choice of the specific distribution to be used in a specific application depends on several issues. If a specific distribution is used in practice to data of the type an investigator has (e.g., a multinomial distribution is often applied in practice to model the number of examinees who obtained a specific score on an examination), that distribution should be the first choice. If such knowledge is lacking, a sensible approach is to graphically plot the data and try to find a distribution that is expected to describe the data adequately. After fitting a distribution to a data set, methods of goodness-of-fit can be used to examine whether the distribution provides an adequate description of the data set.

See also: Continuous Probability Distributions; Goodness-of-Fit Testing; Probability Theory.

Bibliography

- Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.
- Evans, M., Hastings, N., and Peacock, B. (2000). *Statistical Distributions*. New York: Wiley.
- Holland, P. W. (1979). The tyranny of continuous models in a world of discrete data. *IHS-Journal* **3**, 29–42.
- Holland, P. W. (1990). The dutch identity: A new tool for the study of item response models. *Psychometrika* **55**, 5–18.
- Holland, P. W. and Thayer, D. T. (2000). Univariate and bivariate loglinear models for discrete test score distributions. *Journal of Educational and Behavioral Statistics* **25**, 133–183.
- Johnson, N. L., Kemp, A. W., and Kotz, S. (2005). *Univariate Discrete Distributions*. New York: Wiley.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. New York: Wiley.
- Lewis, C. (2007). Some thoughts on controlling item exposure in adaptive testing. *Paper Presented at the 2007 GMAC® Conference on Computerized Adaptive Testing*.
- Novick, M. R. and Jackson, P. H. (1974). *Statistical Methods for Educational and Psychological Research*. New York: McGraw-Hill.

Discrimination and Classification

R C Neath, City University of New York, New York, NY, USA
M S Johnson, Columbia University, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

A frequently encountered problem in statistical applications to the educational sciences is that of determining to which of two or more groups an object of unknown group membership belongs. For example, we can think of there being two distinct groups of college students, those who graduate within 4 years and those who do not. A number of factors, such as high school grade point average, scores on college entrance exams, and participation in extracurricular activities, might have predictive power in this setting. What specific factor, or combination of factors, provides the most distinct separation between the two groups? Given the profile of a particular individual, what is our prediction as to his or her graduation status?

Broadly defined, discrimination and classification are multivariate statistical techniques for separating two or more populations, and deriving rules to predict group membership for new observations. The two terms are often used interchangeably, and indeed, the distinction between the two is a fuzzy one. Following Johnson and Wichern (2007) and other authors, we define these terms as follows:

1. *Discrimination*. A class of exploratory statistical methods to identify features or aspects of data that separate groups.
2. *Classification*. A body of statistical procedures that allocate observations into two or more well-defined groups.

It should be clear that the distinction between these two ideas is not sharp, as any method that effectively separates two or more populations will also provide a good rule for classifying future observations, and an optimal classification rule requires finding the clearest possible separation of the data.

Throughout this article we assume that our data consist of n observations, each known to have come from one of K populations, and an indicator of which population each observation is from. Thus, we have n observations of $(x_1, \dots, x_p, y)'$, where $(x_1, \dots, x_p)' \in \Omega \subseteq \mathbb{R}^p$ and $y \in \{1, \dots, K\}$, the population indicator. In the classification setting, these observations are sometimes called the training samples, as it is an often recommended practice to set aside a portion of the data to treat as the validation samples, on which the classifier's performance can be evaluated.

For an illustrative example, suppose $K = 2$ and $p = 2$ and consider the data in **Figure 1**. Thus, our data consist of bivariate observations from two populations, represented in **Figure 1** by triangles and circles. The goals of discriminant analysis and classification analysis for this example would be to find a boundary function and a classification boundary that best separates the circles and triangles. A number of methods for accomplishing these goals are described in the next section.

There are a number of other methods that have been suggested for classification; this article focuses on some of the most common approaches currently utilized for classification. Most textbooks in multivariate statistical methods (e.g., Johnson and Wichern, 2007; Rencher, 2002) provide an introduction to discriminant analysis and classification; multivariate texts usually focus on Fisher's methods and the k -nearest-neighbor approach. For a comprehensive discussion of classification and regression trees, see Breiman *et al.* (1984). A thorough review of most modern approaches to classification is available in Hastie *et al.* (2001).

Linear Methods

Regression Methods

One possible approach to discrimination and classification involves viewing the group indicator y as a (categorical) response variable, and running a regression of y on the predictor variables $\mathbf{x} = (x_1, \dots, x_p)'$.

Least-squares regression

For each $k = 1, \dots, K$ we define $y^{(k)}$ to take the value 1 if $y = k$ and 0 otherwise, so $y^{(k)}$ is an indicator of membership in group k . Note that $\Pr(y = k | \mathbf{x}) = \Pr(y^{(k)} = 1 | \mathbf{x}) = E(y^{(k)} | \mathbf{x})$. We next conduct K linear regressions, of $y^{(k)}$ on \mathbf{x} for each $k = 1, \dots, K$. That is, we fit the models

$$E(y^{(k)} | \mathbf{x}) = \beta_0^{(k)} + \beta_1^{(k)} x_1 + \dots + \beta_p^{(k)} x_p$$

estimating $(\beta_0^{(k)}, \beta_1^{(k)}, \dots, \beta_p^{(k)})'$ by the least-squares method. Denote the least-squares estimate of $\beta_i^{(k)}$ by $\hat{\beta}_i^{(k)}$ for $i = 0, 1, \dots, p$. We can then estimate the conditional group membership probabilities by the fitted values

$$\hat{\Pr}(y = k | \mathbf{x}) = \hat{\beta}_0^{(k)} + \hat{\beta}_1^{(k)} x_1 + \dots + \hat{\beta}_p^{(k)} x_p$$

for $k = 1, \dots, K$. It can be shown that the least-squares estimates satisfy $\sum_{k=1}^K \hat{\beta}_0^{(k)} = 1$ and $\sum_{k=1}^K \hat{\beta}_i^{(k)} = 0$ for

$i = 1, \dots, p$. An immediate consequence of this result is that $\sum_{k=1}^K \hat{\Pr}(y = k|x) = 1$ for any value of \mathbf{x} . On the other hand, it is possible for a fitted probability to be less than 0 or greater than 1. If we are in the classification setting, we would classify a new observation \mathbf{x}_0 as belonging to the group k for which $\hat{\Pr}(y = k|\mathbf{x}_0)$ is maximized.

The least-squares regression approach to discrimination and classification is simple, intuitive, and easy to implement. Unfortunately, in addition to the unattractive property of negative estimated probabilities, this approach is well known to be prone to masking, in which one or more categories is missed entirely by the classification rule (see Hastie *et al.*, 2001, for an example).

Logistic regression

An alternative to least-squares regression that guarantees the fitted probabilities will be between 0 and 1 is the method of multinomial logistic regression. We arbitrarily designate the last group, group K , to serve as the baseline category. In the multinomial logit model,

$$\log \frac{\Pr(y = k|x)}{\Pr(y = K|x)} = \beta_0^{(k)} + \beta_1^{(k)} x_1 + \dots + \beta_p^{(k)} x_p$$

for $k = 1, \dots, K-1$. According to this model, the ratio of any two group membership probabilities is a log-linear function of \mathbf{x} , since we have

$$\log \frac{\Pr(y = j|x)}{\Pr(y = k|x)} = (\beta_0^{(j)} - \beta_0^{(k)}) + (\beta_1^{(j)} - \beta_1^{(k)}) x_1 + \dots + (\beta_p^{(j)} - \beta_p^{(k)}) x_p$$

for any j and k , including the baseline category K if we take $\beta_i^{(K)} = 0$ for $i = 0, 1, \dots, p$, a convenient choice to ensure

model identifiability. The group membership probabilities can be solved for explicitly, and are given by

$$\Pr(y = k|x) = \frac{\exp(\beta_0^{(k)} + \beta_1^{(k)} x_1 + \dots + \beta_p^{(k)} x_p)}{\sum_{j=1}^K \exp(\beta_0^{(j)} + \beta_1^{(j)} x_1 + \dots + \beta_p^{(j)} x_p)}$$

The unknown parameters are estimated by the method of maximum likelihood (Agresti, 2002); maximum likelihood estimates are easily found using any statistical software package (though care must be taken to make sure the model parametrization is what the user understands it to be). In the classification problem, a new observation \mathbf{x}_0 is classified as belonging to the group k for which $\hat{\Pr}(y = k|\mathbf{x}_0)$ is maximized.

Example

Consider the illustrative example introduced above. Figure 2 contains the same scatterplot introduced above with two linear boundaries; the dashed line corresponds to the boundary based on least squares and the solid line corresponds to logistic regression. Any point above (and right) of the line would be classified as a triangle; below (and left) would be classified as a circle.

Fisher Discrimination

The goal of discrimination analysis is to transform the data in such a way that observations drawn from different populations will be well separated on the transformed scale. Fisher discrimination finds the one or more linear combinations of \mathbf{x} that best separate the groups. We first address the problem of discriminating between two groups, and then generalize to the K -population case.

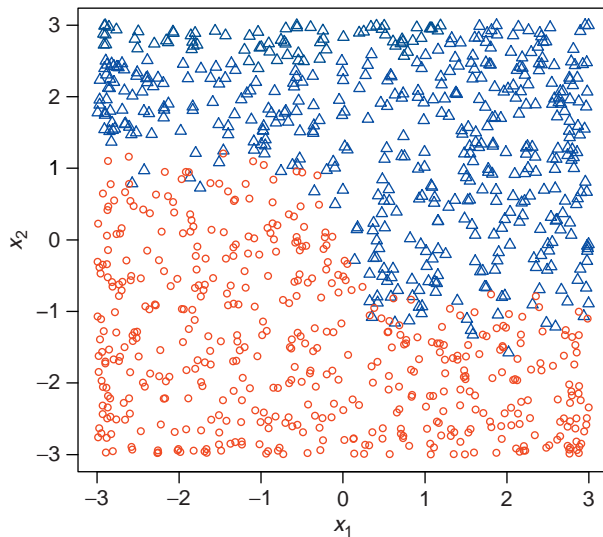


Figure 1 A simple two-group discrimination/classification example with two predictor variables.

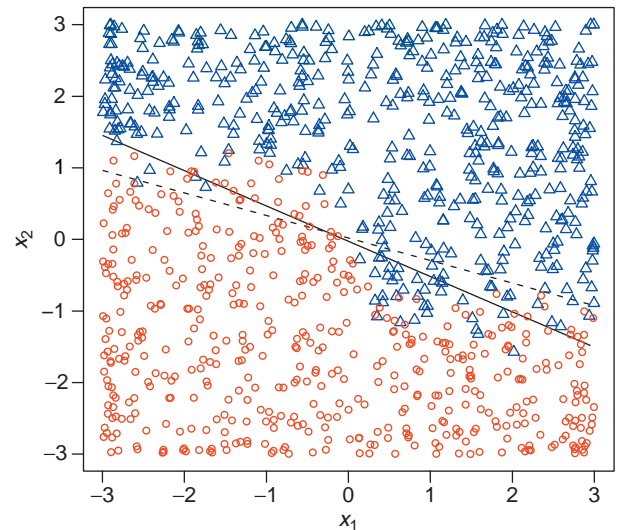


Figure 2 The least-squares (dashed line) and logistic regression (solid line) classification boundaries for the example data set.

Fisher discrimination for two populations

Consider the linear transformation $u = \mathbf{a}'\mathbf{x}$ for $\mathbf{a} \in \mathbb{R}^p$. What choice of \mathbf{a} provides the clearest separation between the two groups? One possible answer is the choice that maximizes the absolute value of the t -statistic for comparing population means. Let \bar{u}_1 and \bar{u}_2 denote the means of the transformed values from populations 1 and 2, respectively, and let s_u denote the pooled sample standard deviation. We wish to make

$$t = \frac{\bar{u}_1 - \bar{u}_2}{s_u \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

as large in magnitude as possible. Let $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ denote the sample mean vectors and \mathbf{S} the pooled sample covariance matrix, so $\bar{u}_k = \mathbf{a}'\bar{\mathbf{x}}_k$ for $k = 1, 2$ and $s_u^2 = \mathbf{a}'\mathbf{S}\mathbf{a}$. We seek the value of \mathbf{a} that maximizes

$$\frac{(\bar{u}_1 - \bar{u}_2)^2}{s_u^2} = \frac{(\mathbf{a}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2))^2}{\mathbf{a}'\mathbf{S}\mathbf{a}}$$

which turns out to be $\mathbf{a} = \mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$. The linear combination $(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}^{-1}\mathbf{x}$ is called Fisher's linear discriminant function. Note that there is no assumption of multivariate normality here. On the other hand, for the notion of maximizing a distance measured in standard deviations to make sense, we must implicitly assume a constant covariance matrix across groups.

Fisher discrimination for K populations

Let x_{kl} denote the l th observation from group k for $k = 1, \dots, K$ and $l = 1, \dots, n_k$. As above, we let $\bar{\mathbf{x}}_k$ denote the k th sample mean, and let $\bar{\mathbf{x}}$ denote the overall mean. We compute the matrices

$$\mathbf{B} = \sum_{k=1}^K n_k (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})'$$

and

$$\mathbf{W} = \sum_{k=1}^K \sum_{l=1}^{n_k} (x_{kl} - \bar{\mathbf{x}}_k)(x_{kl} - \bar{\mathbf{x}}_k)'$$

These are the between groups and within groups sums of squares matrices for a MANOVA of \mathbf{x} by group. If \mathbf{B} is a measure of the variability among the $\bar{\mathbf{x}}_k$'s, then $\mathbf{a}'\mathbf{B}\mathbf{a}$ is a measure of the variability among the $\mathbf{a}'\bar{\mathbf{x}}_k$'s. Similarly, as \mathbf{W} is a measure of the variability among the \mathbf{x} -values within a group, $\mathbf{a}'\mathbf{W}\mathbf{a}$ measures the variability of the $\mathbf{a}'\mathbf{x}$'s. The first linear discriminant is given by the vector \mathbf{a}_1 that maximizes the ratio

$$\frac{\mathbf{a}'\mathbf{B}\mathbf{a}}{\mathbf{a}'\mathbf{W}\mathbf{a}}$$

This turns out to be the eigenvector associated with the largest eigenvalue of $\mathbf{W}^{-1}\mathbf{B}$. The second linear discriminant \mathbf{a}_2 maximizes the above ratio subject to $\mathbf{a}_1'\mathbf{W}\mathbf{a}_2 = 0$.

This turns out to be the eigenvector for the second largest eigenvalue of $\mathbf{W}^{-1}\mathbf{B}$, and so on. There are as many linear discriminants as the rank of the matrix \mathbf{B} .

While Fisher's objective for the linear discriminants was to maximally separate values from different populations, Fisher discrimination can be used as the basis for a classification rule also. Let \mathbf{A} be a matrix whose rows are the first few linear discriminants. Given the value x_0 , classify this case to the population k for which the distance between $\mathbf{A}\mathbf{x}_0$ and $\mathbf{A}\bar{\mathbf{x}}_k$ is smallest.

Fisher's linear discriminant analysis is equivalent (up to a constant of proportionality) to the least-squares regression of the indicator variable on \mathbf{x} .

Support Vector Classifiers

Support vector machines are a relatively modern approach to classification as compared to the other methods described above.

Consider the problem of classifying observations into one of two groups. We say that the observed data are perfectly linearly separable when there exists a linear function $u(\mathbf{x}) = a_0 + \mathbf{a}'\mathbf{x}$ that perfectly separates the two groups in Ω . That is, there exists a vector \mathbf{a} such that $u(\mathbf{x}) > 0$ for all individuals in one group and $u(\mathbf{x}) < 0$ for all observations in the second group. The plane defined by $u(\mathbf{x})$ is called the separating hyperplane.

When the two groups are perfectly separable, there will be infinitely many separating hyperplanes. In order to define a unique solution to the separating hyperplane problem, the method searches for the hyperplane that perfectly separates the two groups and maximizes the distance from the plane to the nearest observed data point; this hyperplane is called the maximum margin hyperplane. Classification is then achieved by the following rule:

$$\text{Classify into } \begin{cases} \text{Group 1} & \text{if } u(x_0) \geq C \\ \text{Group 2} & \text{if } u(x_0) \leq -C \end{cases} \quad [1]$$

The region defined by $|u(\mathbf{x})| < C$ defines the margin, where we are unable to classify observations into either group. Any observed data point on the boundary of the margin is called a support vector.

In most applied problems, the two groups will not be perfectly linearly separable. In such cases, the support vector classifier allows for some observations to be incorrectly classified. The method defines a set of slack variables that allow some of the observations within a group to violate the classification rule in [1]; the size of the slack variable is the distance between the data point and the boundary of group margin. The support vector classifier then maximizes the width of the margin while bounding the size of the slack variables below some predefined constant.

Basis Expansion Methods

In many applications, a linear boundary between the groups will not be adequate for accurate classification. In such cases we can often still rely on linear classifiers, but in a higher dimensional space. In our treatment of the classification problem, we have a p -dimensional vector \mathbf{x} to perform classification. The idea of basis expansion methods is to map this p -dimensional vector into a higher dimensional space, where linear classification performs well.

Consider the example represented in **Figure 3** where the circles and triangles are perfectly separated by the circle defined by $(x_1 - 1)^2 + (x_2 - 1)^2 = 1$. Expanding the squares, we find that classification can be achieved by classifying \mathbf{x}_0 into the circle group if

$$x_1^2 - 2x_1 + 1 + x_2^2 - 2x_2 + 1 < 1$$

and into the triangle group if the inequality is reversed. Therefore, we can get perfect linear separability if we expand from the two-dimensional space (x_1, x_2) to the four-dimensional space (x_1, x_2, x_1^2, x_2^2) . The projection of the linear classifier in the expanded space onto the two-dimensional space creates a nonlinear classifier. Popular expansion methods include polynomial expansion, spline expansion, and kernelization methods.

Basis expansion can also help with the example presented in the introduction (see **Figure 1**). If we expand the logistic regression model to include quadratic and cubic terms, the resulting classification rule fits the data much better as seen in **Figure 4**.

Density-Based Classification Methods

Let f_k denote the probability density function for population k , that is, $x|y = k \sim f_k(\cdot)$, for $k = 1, \dots, K$. The basic

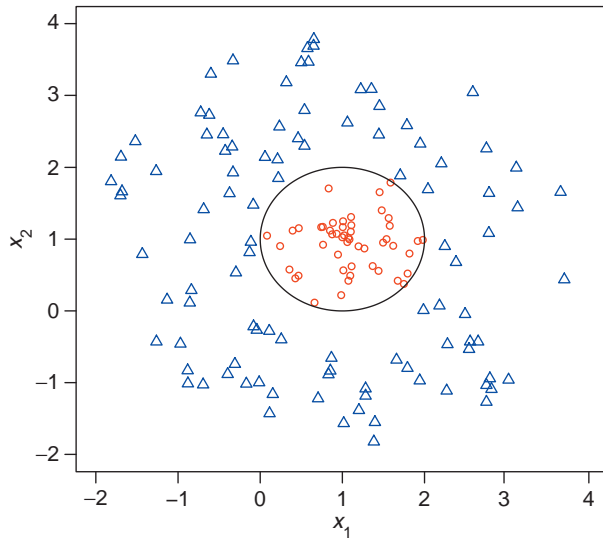


Figure 3 An example where groups are perfectly separated by the circle $x_1^2 + x_2^2 = 1$.

idea in density-based classification is to classify a new observation \mathbf{x}_0 as coming from the population k where the ratio $f_k(\mathbf{x}_0)/f_i(\mathbf{x}_0)$ is high for $i \neq k$. Two more considerations that enter are population prevalence and cost of misclassification. We first illustrate with $K = 2$ populations and then extend to the more general case.

The prevalence of a population is the marginal probability that an object belongs to the population, given no information on the value of \mathbf{x} . Let populations 1 and 2 have prevalence p_1 and p_2 , respectively, so that $p_1 + p_2 = 1$. Misclassification cost accounts for the fact that some mistakes are more severe than others. Let $c(1|2)$ denote the cost of classifying a population 2 object as coming from population 1, and let $c(2|1)$ be the cost of classifying a population 1 object as population 2. We show here that the classification rule to minimize the overall expected cost of misclassification is to classify a new observation \mathbf{x}_0 as coming from population 1 if

$$\frac{f_1(\mathbf{x}_0)}{f_2(\mathbf{x}_0)} > \frac{p_2 c(1|2)}{p_1 c(2|1)} \quad [2]$$

and to classify as population 2 otherwise.

To see why the above rule is optimal, let $EC_1(\mathbf{x})$ denote the expected cost of classifying \mathbf{x} as population 1, and $EC_2(\mathbf{x})$ the expected cost of classifying as population 2. The optimal classification rule will choose population 1 if $EC_1(\mathbf{x}) < EC_2(\mathbf{x})$ and population 2 otherwise.

Now

$$EC_1(\mathbf{x}) = c(1|2)\Pr(y = 2|\mathbf{x}) = c(1|2) \frac{p_2 f_2(\mathbf{x})}{p_1 f_1(\mathbf{x}) + p_2 f_2(\mathbf{x})}$$

and

$$EC_2(\mathbf{x}) = c(2|1)\Pr(y = 1|\mathbf{x}) = c(2|1) \frac{p_1 f_1(\mathbf{x})}{p_1 f_1(\mathbf{x}) + p_2 f_2(\mathbf{x})}$$

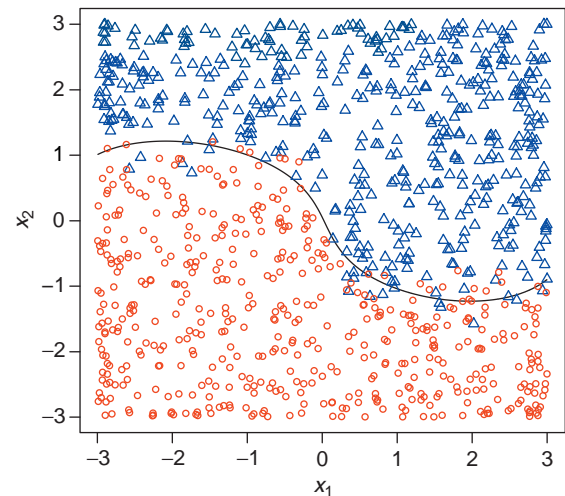


Figure 4 The classification rule based on a logistic regression model that includes both quadratic and cubic terms for the two predictor variables x_1 and x_2 .

so we will classify as population 1 if

$$c(1|2)p_2f_2(\mathbf{x}) < c(2|1)p_1f_1(\mathbf{x})$$

and population 2 otherwise. Thus, [2] describes the classification rule that minimizes the expected cost of misclassification.

We now consider the general K -population case. Again we denote the prevalence of population k by p_k ; this is the marginal probability $p_k = \Pr(y = k)$ for $k = 1, \dots, K$. Let the misclassification cost $c(i|j)$ denote the cost of classifying an object from population j into population i . Assume that $c(i|i) = 0$ for each i . Finally, we again let $EC_k(\mathbf{x})$ denote the expected cost of classifying the value \mathbf{x} as population k . Then

$$EC_k(\mathbf{x}) = \sum_{j \neq k} c(k|j) \Pr(y = j|\mathbf{x}) = \frac{\sum_{j \neq k} c(k|j) p_j f_j(\mathbf{x})}{\sum_{j=1}^K p_j f_j(\mathbf{x})}$$

Given the observed value \mathbf{x}_0 , we should classify to the population k to minimize $EC_k(\mathbf{x}_0)$, that is, to the population k for which

$$\sum_{j \neq k} p_j c(k|j) f_j(\mathbf{x}_0) \text{ is minimized}$$

If all misclassification errors are equally costly, that is, if the $c(i|j)$ are equal for all $i \neq j$, this rule simplifies to choosing the population k to maximize $p_k f_k(\mathbf{x}_0)$.

Normal Populations with Equal Variance

Suppose all K populations are multivariate normal distributions with a common covariance matrix, that is, $\mathbf{x}|y = k \sim N_p(\mu_k, \Sigma)$ for $k = 1, \dots, K$. First suppose $K = 2$. Then

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} = \exp \left\{ -\frac{1}{2}(\mu_1 + \mu_2)' \Sigma^{-1}(\mu_1 - \mu_2) + (\mu_1 - \mu_2)' \Sigma^{-1} \mathbf{x} \right\}$$

and the classification rule [2] reduces to assigning \mathbf{x}_0 to population 1 if

$$-\frac{1}{2}(\mu_1 + \mu_2)' \Sigma^{-1}(\mu_1 - \mu_2) + (\mu_1 - \mu_2)' \Sigma^{-1} \mathbf{x}_0 > \log \left[\frac{p_2 c(1|2)}{p_1 c(2|1)} \right]$$

and population 2 otherwise.

For $K \geq 3$ there is not much simplification for the case of general misclassification costs. If all misclassification costs are equal, we assign \mathbf{x}_0 to the population k , which maximizes

$$\log p_k - \frac{1}{2}(\mathbf{x}_0 - \mu_k)' \Sigma^{-1}(\mathbf{x}_0 - \mu_k) \quad [3]$$

If the K populations are of equal prevalence, this rule reduces to classifying to the population whose mean is closest to \mathbf{x}_0 in Mahalanobis distance. Also, since the quadratic term in [3] is the same for all k , an equivalent classification rule is to choose k to maximize

$$\log p_k - \frac{1}{2} \mu_k' \Sigma^{-1} \mu_k + \mu_k' \Sigma^{-1} \mathbf{x}_0$$

called the linear discriminant score of \mathbf{x}_0 for the k th population.

Normal Populations with Unequal Variances

Now suppose that $\mathbf{x}|y = k \sim N_p(\mu_k, \Sigma_k)$ for $k = 1, \dots, K$. Again, we first consider $K = 2$. Here

$$\begin{aligned} \log \left(\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \right) &= -\frac{1}{2} \mathbf{x}' (\Sigma_1^{-1} - \Sigma_2^{-1}) \mathbf{x} + (\mu_1' \Sigma_1^{-1} - \mu_2' \Sigma_2^{-1}) \mathbf{x} \\ &\quad - \frac{1}{2} \log \left(\frac{|\Sigma_1|}{|\Sigma_2|} \right) - \frac{1}{2} (\mu_1' \Sigma_1^{-1} \mu_1 - \mu_2' \Sigma_2^{-1} \mu_2) \end{aligned}$$

and thus linear discriminant in the classification rule for the case of equal covariance matrices becomes a quadratic discriminant in the general case.

For $K \geq 3$, assuming equal misclassification costs, we would assign \mathbf{x}_0 to population k that maximizes

$$\log p_k - \frac{1}{2} \log |\Sigma_k| - \frac{1}{2} (\mathbf{x}_0 - \mu_k)' \Sigma_k^{-1} (\mathbf{x}_0 - \mu_k)$$

the quadratic discriminant score of \mathbf{x}_0 for population k .

Density-Based Classification in Practice

In practical applications, even where the assumption of multivariate normality is reasonable, parameter values will almost never be known. The usual practice is simply to replace them with their estimators, thus using the above formulas with the sample means $\bar{\mathbf{x}}_k$ in place of the μ_k , sample covariance matrices \mathbf{S}_k for the Σ_k , and the pooled sample covariance matrix $\mathbf{S}_{\text{pooled}}$ for Σ . Even in cases where the normality assumption is violated, it is often possible to transform the data to a scale on which the data are approximately normal. Finally, in problems of moderate dimension, and with very large data sets, it may be feasible to do density-based classification using non-parametric estimates of the unknown density functions using kernel or spline methods.

Classification Trees

The classification-tree approach to a classification problem is based on the notion of decision trees, an example of which is given here. Suppose you have to get to a meeting across town, to which you can walk, take a bus or subway, or hire a taxi cab. You would rather not go by taxi cab, but a bus or subway will take at least 10 min. If your meeting starts in less time than that, you have no choice but to take a taxi. Your preference would be to walk, but only if the weather is nice, and that takes 30 min. You generally prefer the subway to a bus, but the subway is uncomfortable if it is too hot. **Figure 5** gives a decision tree to help you choose your mode of transportation. The endpoints of the decision tree form a partition of the space Ω . This is

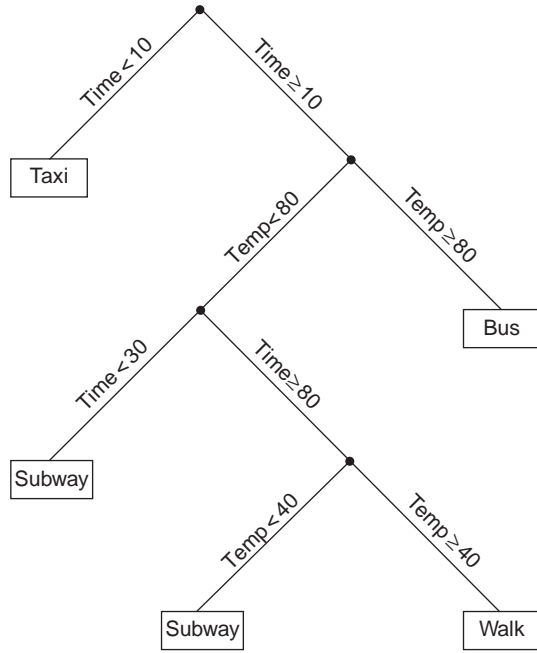


Figure 5 Decision tree for mode of transportation: Take a taxi if time to meeting is less than 10 min, otherwise take a bus if it's too hot ($\text{temp} > 80$), and a subway if it's too cold ($\text{temp} < 40$); walk if the temperature is moderate (40–80) and you have time ($\text{time} > 30$).

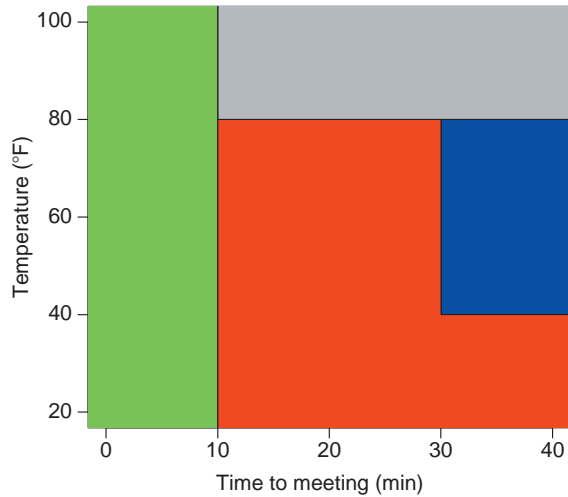


Figure 6 In this plot green indicates a taxi, gray indicates the bus, red indicates the subway, and blue indicates walking.

illustrated in **Figure 6** where we plot time until the meeting on the x -axis and temperature on the y -axis, and use four colors to indicate appropriate mode of transportation given the conditions.

A classification tree is a decision tree for the decision to which population to classify an object \mathbf{x} . There are a number of algorithms for constructing classification trees, the best known of which is the CART algorithm of Breiman *et al.* (1984). The following discussion is sufficiently general as to apply to most algorithms used in practice.

A classification tree is built from top to bottom. The top node contains the entire space Ω and thus the entire training data set. At this and every subsequent node, the algorithm must determine whether the tree should be further split, and if so, how. We address the second point first. Consider an algorithm that constructs splits based on a single variable (it is also possible to split based on linear combinations of the variables). Assuming all our predictors are quantitative, we must consider all splits of the form $\{x_i < s\}$ versus $\{x_i \geq s\}$. At the first node (the top of the tree), there are at most $p(n-1)$ such allowable splits. The algorithm selects the split that maximizes the reduction in node impurity, a measure of the heterogeneity among the training cases classified to that node. Two common measures are the Gini index

$$1 - \sum_{k=1}^K \left(\frac{m_k}{m} \right)^2$$

and the deviance

$$-2 \sum_{k=1}^K m_k \log \left(\frac{m_k}{m} \right)$$

where m is the number of training cases classified to the node being considered for splitting, and m_k is the number of those that are of population k . Unless node impurity is zero, that is, unless $m_k = m$ for some k , a split can be found for which the combined impurity of the two new nodes will be less than that of the parent node. However, this does not necessarily mean that extending the tree is a good idea. Most algorithms use a stopping rule that considers both the number of training cases remaining at the node under consideration and the degree of heterogeneity among those cases, stopping if either of these quantities is too small. One possible strategy is to set highly stringent stopping criteria, to intentionally overgrow the tree and then prune it back at the end. We discuss the method of cost-complexity pruning below.

Let T_0 be a tree grown with a sufficiently stringent stopping rule that the tree is guaranteed to provide a good fit to the training data. For every $\alpha \geq 0$ we define T_α as the subtree of T_0 (i.e., every split in T_α is also in T_0) that minimizes the cost-complexity criterion

$$C_\alpha(T) = Q(T) + \alpha|T|$$

where $Q(T)$ is some measure of the misclassification error for the tree T and $|T|$ is the number of terminal nodes. The former quantity represents the cost of misclassification and the latter penalizes for tree complexity. The family of trees $\{T_\alpha\}$ is found by weakest link pruning, wherein the terminal node that produces the smallest reduction in misclassification error is pruned away; this process is repeated all the way back up to the single node tree T_∞ in which every case is classified to the population with the greatest prevalence in the training sample. From

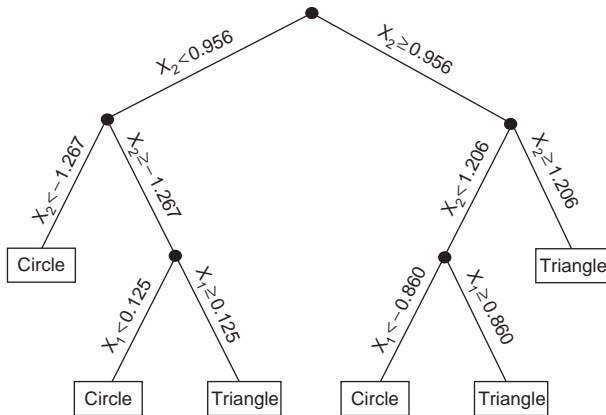


Figure 7 Decision tree for the circles and triangles example.

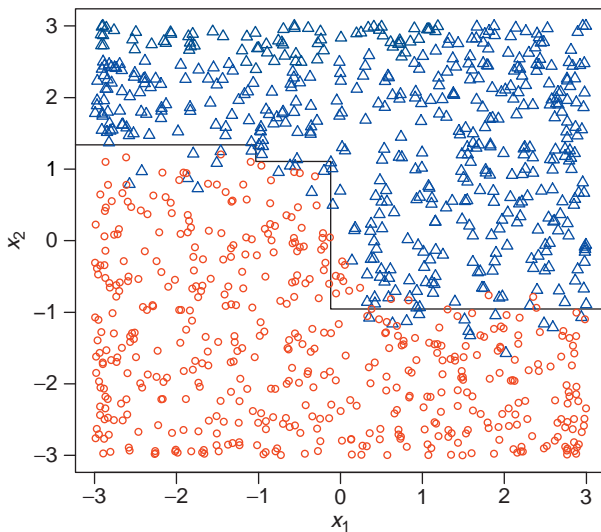


Figure 8 The classification tree rule for the circles and triangles example.

the family of trees $\{T_\alpha\}$, the appropriate value of α , and hence the final classification tree, is found by cross-validation.

Among the strengths of this approach to classification are: (1) the absence of any assumptions about the distributions of $x|y$; (2) the ease with which categorical and ordinal predictors are incorporated into the algorithm (a categorical predictor with q levels requires that $2^{q-1} - 1$ splits be considered, and ordinal variables are treated no differently than continuous, with $q - 1$ splits to consider); and (3) the ability of tree-based methods to make the most

of incomplete observations (a partial observation is dropped only if it is missing the variable being split on).

The classification tree for the circles and triangles example introduced earlier is depicted in **Figures 7 and 8**. The resulting classification rule is very similar to the one derived by cubic logistic regression.

k-Nearest-Neighbor Classification

The k -nearest-neighbor approach to classification is a relatively simple approach to classification that is completely nonparametric. Given a point \mathbf{x}_0 that we wish to classify into one of the K groups, we find the k observed data points that are nearest to \mathbf{x}_0 . The classification rule is to assign \mathbf{x}_0 to the population that has the most observed data points out of the k -nearest neighbors. Points for which there is no majority are either classified to one of the majority populations at random, or left unclassified.

The advantage of nearest-neighbor classification is its simplicity. There are only two choices a user must make: (1) the number of neighbors, k and (2) the distance metric to be used. Common choices of distance metrics include Euclidean distance, Mahalanobis distance, and city-block distance. The number of neighbors is usually selected by either cross-validation or testing the quality of the classifier on a second, test data set.

See also: Analysis and Interpretation of Multivariate Data; Categorical Data Analysis; Cluster Analysis: Overview; Multivariate Analysis of Variance; Multivariate Normal Distribution; Recursive Partitioning.

Bibliography

- Agresti, A. (2002). *Categorical Data Analysis*, 2nd edn. New York: Wiley.
- Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Monterey, CA: Wadsworth.
- Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning Theory*. New York: Springer.
- Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*, ch. 11, 6th edn. New York: Wiley.
- Rencher, A. C. (2002). *Methods of Multivariate Analysis*, ch. 8, 2nd edn. New York: Wiley.

Further Reading

- Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*, ch. 12., 4th edn. New York: Springer.

Empirical Bayes Methods

E Ip, Wake Forest University, Winston-Salem, NC, USA

G Molenberghs, Universiteit Hasselt, Diepenbeek, Belgium

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bayesian inference – A kind of statistical inference that makes extensive use of the Bayes theorem to update probability that a hypothesis may be true as evidence accumulates.

Item response theory – A mathematical theory that uses statistical models to describe the interactions between an individual and items from a questionnaire under a stimulus-response framework.

Maximum likelihood method – A statistical method that fits a model to data. The method is based upon the likelihood of observing the collection of data points under the specified model.

Introduction

The empirical Bayes (EB) approach can be viewed from at least two perspectives: first, as a technical tool for borrowing information across individual cases for performing statistical inference; and, second, as a compromise between two statistical paradigms – Bayesian and frequentist. While the second topic is important for understanding EB's role within the inferential paradigms, it is too vast to do full justice to in this article.

In education, the EB approach is particularly useful for enhancing the quality of educational measurements (student proficiency, teacher performance, school achievement, etc.) when multiple statistical units (schools, teachers, students, etc.) are measured.

Let us illustrate how the EB procedure allows information to be borrowed across individual units for improving the quality of educational measurement. Suppose that a state is interested in using a new portfolio-assessment method for evaluating student performance on language ability. The goal is to evaluate the feasibility of replacing the traditional test with the new method. One is also interested in obtaining an accurate estimate of each school's performance so that they can design enhancement programs and allocate resources. The state administers the new assessment method to a sample of 100 students from each of ten randomly selected schools. The minimum and maximum portfolio scores are 0 and 800, respectively. Table 1 displays the corresponding mean scores.

The question is: Are the mean scores good estimates for reporting school performance?

Because of intra-school variability, a school's score depends upon the sample drawn: if a school's student population is heterogeneous, then the sample-to-sample variation may be large. Due to haphazard chance, a school may therefore appear not to be performing up to its true potential; the converse could also be true.

The primary idea behind EB is to use both local (sample mean score) and global (embedded in the distribution of the mean scores across schools) information to enhance the quality of the estimates. One can think of the ten scores in Table 1 as a sample from the universe of the state's schools. The performance distribution of the state's schools can be used to inform the estimates for individual schools. It can be mathematically proven that – when both local and global data are used – the precision of the estimates for individual schools is better than when only local information is used, in terms of root mean square error.

EB Normal–Normal Model

The EB setup is easily illustrated when both the global and local models are normally distributed: the global model states that the true mean scores θ_i are sampled independently from a common underlying distribution, and are not directly observed; and that they follow a normal distribution with mean μ and variance σ_θ^2 :

$$\theta_i \sim N(\mu, \sigma_\theta^2). \quad [1]$$

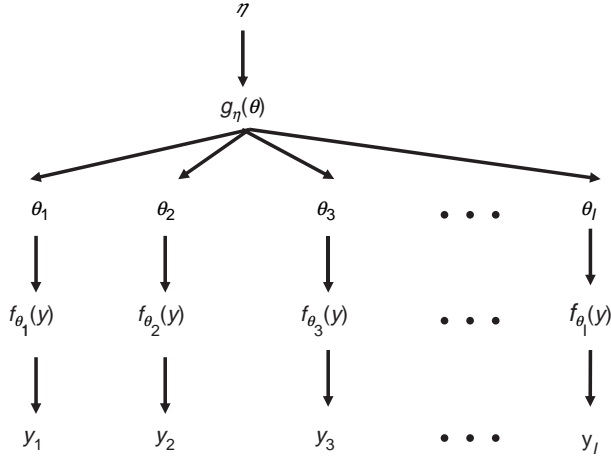
The above distribution is the so-called prior distribution. The local model allows for noise in measuring the score of the individual member and that the observed values of individual score y_i represent a realization of another normal distribution centered at θ_i :

$$y_i | \theta_i \sim N(\theta_i, \sigma_y^2). \quad [2]$$

This hierarchy is depicted in Figure 1. The normal prior distribution defined in [1] provides a model for global information, whereas the distribution $f_{\theta_i}(y)$ (the normal distribution in [2]) contains local information about the individual unit. This structure is rather general and can be applied to a variety of situations. For example, in the school-performance example, the observed values are collected at the student level. Suppose that each school

Table 1 School performance mean scores

School	1	2	3	4	5	6	7	8	9	10
Mean score	504.0	511.3	495.7	505.9	472.3	513.8	505.4	482.7	460.8	505.7


Figure 1 Data generating mechanism in EB approach.

contains a sample of \mathcal{J} students; that the score of the j th student from the i th school is denoted by x_{ij} , $i = 1, \dots, I$, $j = 1, \dots, \mathcal{J}$; and that the student scores follow a distribution with mean θ_i and variance σ_x^2 . Then, the mean score of each school is given by $y_i = (1/\mathcal{J}) \sum_{j=1}^{\mathcal{J}} x_{ij}$, and so:

$$y_i | \theta_i \sim N\left(\theta_i, \frac{\sigma_x^2}{\mathcal{J}}\right) \quad [3]$$

The structure in **Figure 1** is, of course, not confined to normality. Some other examples are presented later in this article.

The EB approach involves the following steps:

(S1) Estimate the global structure, using the marginal distribution of the observed data.

(S2) Estimate the value for each individual unit using global information from step S1, together with local information collected from each individual unit. This specific step of inference is through the posterior distribution of individual true (unobserved) measurement of each unit given the data.

To elaborate on (S1) and (S2), use notation of [1] and [2]. First, consider the case in which σ_θ^2 and σ_y^2 are both known, with only μ unknown and estimated in (S1). Because y_i are independently normally distributed with mean μ and variance $\sigma_\theta^2 + \sigma_y^2$, the maximum-likelihood estimate (MLE) for μ is the sample mean:

$$\hat{\mu} = \bar{y} = \frac{1}{I} \sum_{i=1}^I y_i \quad [4]$$

(S2) involves estimating the true but unobserved mean score for each school. The typical EB estimate for a two-level model of this kind would be the posterior mean of the unknown true mean score given the observed data and the estimated global structure:

$$\hat{\theta}_i^{EB} = E(\theta_i | y_i, \mu) \quad [5]$$

From normal theory, the conditional distributions of θ_i given y_i are independently normally distributed:

$$\theta_i | y_i \sim N(B\mu + (1-B)y_i, (1-\beta)\sigma_y^2) \quad [6]$$

where $B = \frac{\sigma_y^2}{\sigma_y^2 + \sigma_\theta^2}$. Therefore, the EB estimate $\hat{\theta}_i^{EB}$ is given by

$$\hat{\theta}_i^{EB} = B\bar{y} + (1-B)y_i \quad [7]$$

from plugging in the estimated parameter for the global structure – that is, replacing μ with $\hat{\mu} = \bar{y}$. The uncertainty associated with the EB estimate $\hat{\theta}_i^{EB}$ is based on the variance $V(\theta_i | y_i, \mu)$ of the posterior distribution, with 95% confidence interval:

$$(\hat{\theta}_i^{EB} - 1.96 \times \sqrt{(1-B)\sigma_y^2}, \hat{\theta}_i^{EB} + 1.96 \times \sqrt{(1-B)\sigma_y^2}) \quad [8]$$

Note that the EB approach – in contrast with general Bayesian approaches – always directly estimates the global structure from empirical data and then uses plug-in estimates of the global structure for estimating individual locality. The E in EB refers to this plug-in principle. To further illustrate the plug-in principle, retain the normal-normal case with the parameter σ_θ^2 also unknown. The marginal MLE for the pair (μ, σ_θ^2) is given by

$$\hat{\mu} = \bar{y} = \frac{1}{I} \sum_{i=1}^I y_i \quad [9a]$$

$$\hat{\sigma}_\theta^2 = \max(s^2 - \sigma_y^2, 0) \quad [9b]$$

where $s^2 = (1/I) \sum_{i=1}^I (y_i - \bar{y})^2$ is the sample variance of the observed y_i . Following the plug-in principle, the EB estimate $\hat{\theta}_i^{EB}$ for an individual school is now

$$\hat{\theta}_i^{EB} = \hat{B}\bar{y} + (1-\hat{B})y_i \quad [10]$$

where $\hat{B} = \sigma_y^2 / (\sigma_y^2 + \hat{\sigma}_\theta^2)$, in which $\hat{\sigma}_\theta^2$ is plugged in from [9b]. The confidence intervals take form [8], although now using \hat{B} rather than B . In general, the EB-based confidence interval would be too narrow because it does not account for the uncertainty associated with the estimation of \hat{B} . There are different correction methods, including jack-knife and bootstrap. The reading list includes articles that

cover methods not requiring a fully Bayesian specification. An alternative approach is described in the section on EB application to the National Assessment of Educational Progress (NAEP).

Example

The values in **Table 1** were simulated from a two-level process with $\mu = 500$, $\sigma_\theta = 20$, $\sigma_x = 100$, $I = 10$, and $J = 100$. If we assume σ_x to be known, then the variance of the mean score $\sigma_j^2 = 100^2/100 = 100$. Thus, the estimates for (μ, σ_θ^2) are given by (9) and applied to **Table 1**, which results in $\hat{\mu} = 495.8$, $\hat{\sigma}_\theta = 15.3$, and $\hat{B} = 0.3$. **Figure 2** compares the EB estimates and the values of the mean scores, which are MLEs based only on data from one specific school.

The EB estimate in [10] is a weighted linear combination of two sources of information: globally from \bar{y} and locally from y_i . Effectively, the EB approach amounts to a shrinkage effect that pulls all the MLEs toward the overall mean. The strength of the shrinkage depends on the relative heterogeneity of data at the global and local levels. If the true mean scores are relatively homogeneous (small σ_θ^2 or large σ_j^2), then the factor B should be close to 1 – implying that the shrinkage effect would be large, and vice versa. The shrinkage factor B in this case is analogous to the intraclass correlation in classical test theory – the ratio of the between-student variation to the sum of the between-student variation and the between-school variation.

The EB estimate satisfies many desirable properties within the decision-theoretic framework. The square loss function is often used to quantify the risk for choosing one decision (estimator) over another. If the true value of the target parameter is θ – estimated by $\hat{\theta}$ – then the squared loss function is $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$. In the normal-normal setting, if our interest is in the value of θ_i for each individual unit, it can be proven that – under squared loss – the EB estimator is superior to any other estimator, including those that are based on the observed data y_i alone. When the prior distribution for θ is known, then the EB estimator is ideal for any symmetric loss function. When the global

structure is not known or the setup is more general, various versions of EB estimators have been proposed. Many authors have searched for optimality under various scenarios (Stein, Robbins, Maritz, Morris, Efron, Louis, etc.).

The structure in **Figure 1** can be further generalized in several directions. Examples with different parametric choices for local and global levels are presented below.

EB gamma-Poisson Model

Consider data collected over spelling errors made by school children. In a dictation test, school children are given a fixed number of J words. The number of errors that the i th child made can be modeled as a Poisson process with rate parameter θ_i – assumed to vary across individual children (Van Duijn and Bockenholt, 1995). It is mathematically convenient to assume that the θ_i follows a gamma distribution, that is, the global structure for spelling error rate parameters can be specified through the following two-parameter prior distribution:

$$p_{\text{Gamma}}(\theta_i | \alpha, \beta) = \frac{\theta_i^{\alpha-1} \beta^\alpha e^{-\beta\theta_i}}{\Gamma(\alpha)}, \quad \theta_i, \alpha, \beta > 0 \quad [11]$$

in which α and β are, respectively, the shape and rate (inverse scale) parameters, and Γ is the gamma function. If α is a positive integer, then $\Gamma(\alpha) = (\alpha-1)! = (\alpha-1)(\alpha-2) \dots 3 \times 2 \times 1$. Conditional on θ_i , the total number of spelling errors for child i is given by the Poisson count model:

$$P(Y_i = y_i | \theta_i) = \frac{e^{-\theta_i} \theta_i^{y_i}}{y_i!} \quad [12]$$

where $y_i = 0, 1, 2, \dots$. The marginal distribution for Y_i then is a negative binomial distribution:

$$P(Y_i = y_i) = \left(\frac{\beta}{\beta + 1} \right)^\alpha \frac{\Gamma(\alpha + y_i)}{\Gamma(\alpha)} \frac{1}{y_i!} \frac{1}{(\beta + 1)^{y_i}} \quad [13]$$

The standard method for estimating (α, β) is maximum likelihood, with the likelihood function assembled from contributions [13]. It is convenient to reparameterize by the mean of the gamma distribution $\mu = \alpha/\beta$. The

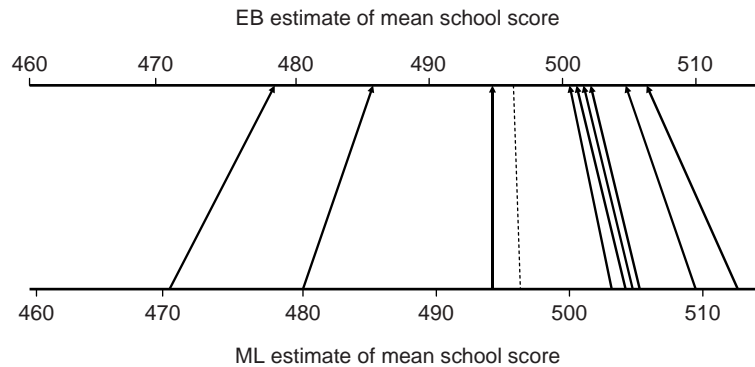


Figure 2 Shrinkage effect from the EB approach in school-performance example.

MLE of μ is $\bar{y} = (1/I) \sum_{i=1}^I y_i$, but there is no closed-form solution for α . One can also use the method of moments for the global structure, that is, matching the moments of the marginal distributions with empirical moments and then solving for the required parameters. This produces (Maritz, 1969):

$$\hat{\mu} = \bar{y} \quad [14a]$$

$$\hat{\beta} = \frac{\bar{y}^2}{(s^2 - \bar{y})}, \text{ if } s^2 > \bar{y} \quad [14b]$$

where $s^2 = \frac{1}{I} \sum_{i=1}^I (y_i - \bar{y})^2$. If $s^2 < \bar{y}$, then the estimated prior distribution is taken to be degenerated at $\theta = \bar{y}$.

The EB estimate for θ_i is the posterior mean, with plug-in parameters given by [14]. The posterior distribution for the gamma-Poisson model is also a gamma distribution with shape and rate parameters $(\alpha + y_i)$ and $(1 + \beta)$. Because this posterior gamma distribution has mean $(\alpha + y_i)/(1 + \beta)$, the EB estimate follows as:

$$\hat{\theta}_i^{EB} = \hat{B}\hat{\mu} + (1 - \hat{B})y_i = \hat{B}\bar{y} + (1 - \hat{B})y_i \quad [15]$$

where $\hat{B} = \hat{\beta}/(1 + \hat{\beta})$. The confidence interval for $\hat{\theta}_i^{EB}$ is based on the posterior variance

$$V(\theta_i | y_i, \mu, \beta) = (\mu\beta + y_i)(1 + \beta)^2 \quad [16]$$

Again, a plug-in estimate for $(\hat{\mu}, \hat{\beta})$ is used.

EB Beta-Binomial Model

The beta-binomial model has been used in the study of criterion-referenced testing for various purposes such as the determination of test length (Novick and Lewis, 1974) and cutoff scores (Huynh, 1977). A criterion-referenced test consists of a sample of n items drawn from a domain of items, developed from a framework of learning objectives. The number correct for the i th student is used to infer the student's proficiency, assumed unobservable. Denote the proficiency of the i th student by θ_i . Then, (1) states that θ_i , $i = 1, \dots, I$, varies across students, following a beta distribution, and (2) states that, given proficiency θ_i , the number of correctly answered questions y_i follows a binomial distribution with parameters (n, θ_i) , where n is the total number of questions on the test. The beta-binomial model can be expressed by the following two equations:

$$p_{Beta}(\theta_i | \alpha, \beta) = \frac{\theta_i^{\alpha-1}(1-\theta_i)^{\beta-1}}{\text{Beta}(\alpha, \beta)}, \quad 0 \leq \theta_i \leq 1 \quad [17]$$

where $\text{Beta}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the beta function; and

$$P(Y_i = y_i | \theta_i) = \binom{n}{y_i} \theta_i^{y_i} (1 - \theta_i)^{n-y_i}, y_i = 0, 1, \dots, n \quad [18]$$

The marginal distribution of y_i is the so-called beta-binomial density:

$$P(Y_i = y_i) = \frac{1}{\text{Beta}(\alpha, \beta)} \binom{n}{y_i} \frac{\Gamma(\alpha + y_i)\Gamma(\beta + I - y_i)}{\Gamma(\alpha + \beta + I)} \quad [19]$$

To estimate the global-structure parameters, we resort to the method of moments, because ML does not generally lead to a closed form. Much as in the case of gamma-Poisson, we benefit from a reparameterization from (α, β) to (μ, λ) , where $\mu = \alpha/(\alpha + \beta)$, and $\lambda = \alpha + \beta$. The parameter μ is the mean of the beta distribution, whereas the parameter λ suggests an effective sample size for the prior distribution, decreasing with variance. If V denotes the variance of the beta distribution, then $V = \mu(1-\mu)/(\lambda+1)$. Note that, if a binary variable Y has mean μ , then with a sample size n the standard error for the MLE for μ is $V = \mu(1-\mu)/n$. Hence, $\lambda + 1$ plays the role of sample size. The moment estimates for (μ, λ) are:

$$\hat{\mu} = \frac{\sum_{i=1}^I y_i}{nI} \text{ and} \quad [20a]$$

$$\hat{\lambda} = \frac{\hat{\mu}(1 - \hat{\mu}) - s^2}{s^2 - \frac{\hat{\mu}(1 - \hat{\mu})}{I}} \quad [20b]$$

where $s^2 = \frac{1}{I} \sum_{i=1}^I (\frac{y_i}{n} - \hat{\mu})^2$.

The posterior distribution $\theta | y_i, \mu, \lambda$ is also a beta distribution with α, β replaced by $\alpha + y_i$ and $\beta + n - y_i$, respectively. Therefore, the mean of the posterior distribution is $\alpha + y_i/(\alpha + \beta + n)$, and after plugging in the estimates from above the EB estimate for individual θ_i is estimated as:

$$\hat{\theta}_i^{EB} = \hat{B}\hat{\mu} + (1 - \hat{B})\frac{y_i}{n} \quad [21]$$

where $\hat{B} = \hat{\lambda}/(\hat{\lambda} + n)$. The shrinkage factor \hat{B} is a function of the relative strength of the sample sizes from global and local sources, with the global source providing an effective sample size $\hat{\lambda}$. When sample sizes vary across units, so does shrinkage. The confidence interval for $\hat{\theta}_i^{EB}$ is based on the estimated posterior variance:

$$V(\theta_i | y_i, \hat{\mu}, \hat{\lambda}) = \frac{(y_i - \hat{\mu}\hat{\lambda})(n - y_i + \hat{\lambda} - \hat{\mu}\hat{\lambda})}{(n + \hat{\lambda})2(n + \hat{\lambda} + 1)} \quad [22]$$

Generalization and Extensions

These instances of the EB method are special cases in which the prior and conditional distributions form conjugate pairs, leading to mathematically tractable marginal distributions. Furthermore, these are examples of the parametric EB approach. **Figure 1** is examined for extension of the approach. Starting from the top, the prior density $g_\eta(\theta)$ that provides the global information assumes a specific parametric form with hyperparameter η . In the normal-normal case, $g_\eta(\theta)$ is the normal distribution, with η assumed known in the first and $\eta = \mu$ in the second case. The conditional distribution for y_i is often assumed to

follow a separate parametric function $f(y | \theta)$. Typically, EB-based inferences target the unobserved individual θ_i , to which end there are the following steps:

1. Form the marginal distribution of the observed data

$$p(y) = \int g_\eta(\theta) f(y | \theta) d\theta, \quad [23]$$

where $g_\eta(\theta) = g_\eta(\theta_1) \times \dots \times g_\eta(\theta_I)$, and $f(y | \theta) = f(y_1 | \theta_1) \times \dots \times f(y_I | \theta_I)$.

2. Using [23], estimate the hyperparameter η using, for example, maximum likelihood; leading to $\hat{\eta}$.
3. Form the posterior distribution $p(\theta_i | y_i, \eta)$ using Bayes' theorem, and plug $\hat{\eta}$ into the posterior:

$$p(\theta_i | y_i, \eta) = \frac{p(y_i | \theta_i) g_\eta(\theta_i)}{\int p(y_i | \theta_i) g_\eta(\theta_i) d\theta_i} \quad [24]$$

4. Form the EB estimate using step (3): $\hat{\theta}_i^{EB} = E(\theta_i | y_i, \hat{\eta})$.
5. Form confidence intervals of $\hat{\theta}_i^{EB}$, using variance function $V(\theta_i | y_i, \hat{\eta})$.

Several observations can be made. First, if the distributions $g_\eta(\theta)$ and $f(y | \theta)$ do not form a conjugate pair, the solutions for $\hat{\eta}$ and $\hat{\theta}_i^{EB}$ may require numerical procedures, as will their corresponding precision measures. The expectation–maximization (EM) algorithm (Dempster *et al.*, 1977) is often used for estimating the hyperparameters. Second, from a Bayesian perspective, the EB approach has stopped short of accommodating uncertainty associated with hyperparameter estimation. This issue is discussed within the next example. Third, the parametric EB approach with regard to $g_\eta(\theta)$ and $f(y | \theta)$ can be extended by approached $g_\eta(\theta)$ nonparametrically instead. The pioneering work of Robbins (1955) established a theoretical foundation. In educational measurement, however, a more common extension is by incorporating regression models into either the prior $g_\eta(\theta)$, the conditional $f(y | \theta)$, or both. This is exemplified in the next section.

EB analysis for the NAEP

This example is chosen for two reasons. First, NAEP is a highly visible educational assessment tool in the United States, and reports are of great public interest because they are often cited to support specific educational and political agendas. Second and more technical, the statistical models used (Beaton and Johnson, 1992) exhibit rich features, useful for illustrating both the power and limitations of the EB approach (*cf.* Scott and Ip, 2002).

The NAEP survey measures the academic performance of US students cross sectionally and over time. Mandated by the US Congress and funded by the federal government, NAEP reports academic achievements and identifies differences in performance between student-population subgroups categorized by demographic and

other contextual variables (e.g., time spent watching television). Unlike individual achievement tests (e.g., Scholastic Aptitude Test (SAT)), NAEP reports on overall performance of subgroups or aggregates of students. While NAEP calculates individual proficiency scores and associated student sampling weights using draws from students' estimated ability distributions, it does not report individual student performance – this being prohibited by law. The NAEP approach for analyzing student data consists of two interlinked statistical models. The first captures local information with regard to a student's academic performance based on his/her responses to cognitive items on a given subject matter. The second model creates a global structure, allowing sharing of information across students and schools. Denote by $\theta_i = (\theta_{i1}, \dots, \theta_{is})$ the subscale proficiency level vector for student i , by x_i a vector of contextual and background variables, and by y_{ij} the response to the j th item (correct = 1, incorrect = 0). For reading assessment, there are three subscales ($S = 3$): reading for literary experience, for information, and to perform a task. The global model states that:

$$\theta_i \sim \text{MVN}(\beta^T x_i, \Sigma) \quad [25]$$

where β groups regression coefficients and Σ is a variance–covariance matrix. Such a specification allows the different subscales to be correlated. The local model for individual students is based on the item-response model (IRT; Lord, 1980). For dichotomous items, the three-parameter logistic IRT model given below is used for scaling student responses:

$$\begin{aligned} f(y_{ij} | \theta_{is(j)}) &= P(Y_{ij} = 1 | \theta_{is(j)}) \\ &= c_j + \frac{(1 - c_j)}{1 + \exp(-1.7a_j(\theta_{is(j)} - b_j))} \end{aligned} \quad [26]$$

where $\theta_{is(j)}$ is the proficiency for subscale s to which item j belongs, $j = 1, \dots, J_s$ and a_j, b_j, c_j are the item discrimination, difficulty, and guessing parameters, respectively. Factor 1.7 in [26] is a historical artifact relating the logit and probit links. To focus on the EB structure, the following simplifications are made: all items are dichotomized; their item parameters are *a priori* determined through a standard IRT-based calibration procedure and will be treated as known; and the student sampling weights are uniform.

The EB approach is based on [25] and [26], and hyperparameter η contains β and Σ . Function $f(y_i | \theta_i)$ is now multivariate because y_i is the vector of the binary responses J_{is} for student i at the items of subscale s . Assuming local (conditional) independence, that is, that the item responses for each subscale are independent given proficiency level θ_{is} , then $f(y_i | \theta_{is}) = f(y_{i1} | \theta_{is}) \times \dots \times f(y_{iJ_{is}} | \theta_{is})$. Further assuming that the responses between subscales are independent, the overall response function, therefore, is the product over all subscales. Because of multiple test

forms, the number of items answered per subscale is student dependent. The likelihood is based on these observed factors only. This approach is valid, assuming that the missing responses are missing at random (Little and Rubin, 1987).

The parameters β and Σ are estimated by marginal maximum likelihood, the core of which is an EM algorithm (Mislevy *et al.*, 1992). A standard EB approach would use the plug-in estimate for the regression coefficients β on students' background variables and Σ , and for the variance-covariance matrix for the subscales, to produce the expected value of student subscale proficiency θ_{is} given the observed responses and the plug-in estimates.

Bayesian shrinkage for θ_{is} tends to pull the estimate of proficiency purely based on an individual's responses toward the mean of the student's subgroup – defined by the levels of x_i . Beaton and Johnson (1990) and Mislevy (1991) found asymptotic bias in statistics involving background variables that are not conditioned on, the magnitude of which relates to the extent to which responses may account for the unobserved proficiency θ_{is} and the degree to which the unconditioned background variables are explained by their counterparts in the model. Thus, the EB estimate with conditioning variables tends to mitigate potential bias by about 10% across many NAEP analyses using the unconditional model (Mislevy, 1991).

Figure 3 depicts the effect of Bayesian shrinkage. Here, standard IRT procedures are applied to the 1996 Long Term Trend Reading Data, based on responses to 22 items on a single scale for $N = 918$ students and assuming that the prior distribution is standard normal. The S-shaped curve shows that the absolute values of the EB estimates are smaller than their ML counterparts and that the shrinkage effect is larger for values that are farther away from the overall mean (zero).

While the EB estimate using the conditioning variables limits bias and, generally, outperforms ML, the measures for uncertainty associated with it (e.g., standard error), are

underestimated – especially when based on the assumption that the plug-in estimates contain no measurement errors. There are many correction procedures – at individual and at population levels. NAEP adopted a methodology – called plausible value – based on multiple imputation (Rubin, 1987). The plausible-value method is implemented through several steps (Johnson *et al.*, 1996):

(P1) Draw a value of (β, Σ) from a normal approximation to the posterior $p(\beta, \Sigma | \mathbf{y}, \mathbf{x})$, denoted by $(\tilde{\beta}^{(1)}, \tilde{\Sigma}^{(1)})$ [the operational NAEP procedure keeps Σ fixed and draws only for β];

(P2) Based on this, compute the mean and variance-covariance matrix for the posterior distribution of the proficiencies vector $p(\theta_i | \mathbf{y}, \mathbf{x}, \tilde{\beta}^{(1)}, \tilde{\Sigma}^{(1)})$;

(P3) Draw a value for $\theta_i(1)$ using a multivariate normal approximation of $p(\theta_i | \mathbf{y}, \mathbf{x}, \tilde{\beta}^{(1)}, \tilde{\Sigma}^{(1)})$, with mean and variance-covariance calculated from (P2); and

(P4) Repeat steps (P1) through (P3) M times.

In NAEP, $M = 5$. The total sampling variance for the proficiency estimate – or of any statistic based on the posterior $p(\theta_i | \mathbf{y}, \mathbf{x}, \beta, \Sigma)$ – is given by the sum of the average sampling variance over the M sets of plausible values $\theta_i^{(m)}$, $m = 1, \dots, M$, and the variance among the M estimates. The former – the so-called within-imputation variance – is meant to be an approximation to the posterior variance $V(\theta_i | \mathbf{y}, \mathbf{x}, \beta, \Sigma)$. The latter – the between-imputation variance – is designed as a correction for the uncertainty due to not directly observing the θ_{is} and is:

$$G = \sum_{m=1}^M \frac{(\theta_{is}^{(m)} - \bar{\theta}_{is})^2}{M-1} \quad [27]$$

where $\bar{\theta}_{is}$ is the mean of $\theta_{is}^{(m)}$. The final estimate of the variance of $\hat{\theta}_{is}^{EB}$ is:

$$V(\hat{\theta}_{is}^{EB}) = \hat{V}(\theta_i | \mathbf{y}, \mathbf{x}, \beta, \Sigma) + (1 + M^{-1})G \quad [28]$$

Hence, this approach aims at conducting proper inference for student proficiency both at the individual and population

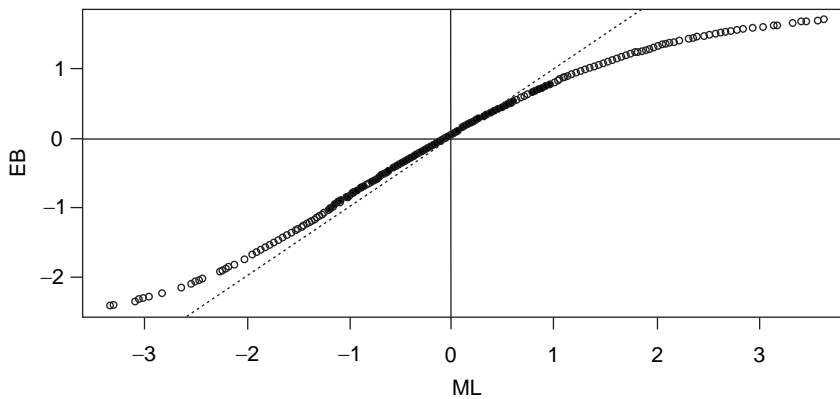


Figure 3 Maximum likelihood estimates and empirical Bayes estimates of student proficiencies in NAEP Long Term Trend Reading Assessment.

levels by enhancing the EB approach so that it approximates a fully Bayesian approach. It is argued that the fully Bayesian approach takes into account the various sources of uncertainty – including those derived from using a plug-in estimate for (β, Σ) .

For illustration, the 1998 eighth-grade national reading assessment ($N = 11\,051$) is analyzed and **Table 2** provides the mean and the within- and between-imputation variances – $M = 100$. For more details and ramifications, see Scott and Ip (2002). Here, the between- and within-imputation variances are comparable – implying that the second source of variability increases the standard error by about a factor $\sqrt{2} = 1.4$, relative to EB.

Other Educational Applications and Conclusion

The EB approach relies on a two-level formulation of how data across statistical units arise, and offers a rigorous theoretical framework for using global information gleaned across individual units for informing local estimates – which are only based on responses from a specific individual. In contrast to a fully Bayesian approach, EB emphasizes estimating prior distributions from the data. From a modeling perspective, the EB approach is amenable for (non)linear multilevel modeling – especially when inference targets unobservables such as proficiencies or teaching skills.

The range of EB applications is wide. Braun (2006) applied the approach to analyze important performance predictors on the Graduate Record Examination (GRE) across various college departments. He used a normal-normal model in which the student's first-year grade point average (GPA) is treated as the response y , the student's GRE scores as predictors, and regression coefficients as unobservables of interest. The global structure is a regression model incorporating department-level covariates.

Meta-analyses for educational studies also make use of EB; information is combined across studies, and the effect size of a specific classroom intervention from study i can be conceptualized as an unobserved variable θ_i . Information is borrowed across studies through the specification of a global data-generating structure (Hedges, 1987; Raudenbush and Bryk, 1985).

For evaluating teacher and school effectiveness, EB was applied to value-added models – a collection of models that attempts to delineate, from a return-on-investment perspective, the effectiveness of school systems and personnel based on the complex interactions among student characteristics, school effects, community characteristics, school district policies, etc. Information is borrowed from teacher assessment results for inferences on a school system or employee, and regression models are incorporated at global and local levels to control for interaction effects. McCaffrey *et al.* (2004) review value-added models and related EB methodologies. Multi-level growth-curve modeling for assessing educational-intervention effects (Plewis, 2000; Pituch, 2001) is another versatile application area. Typically, a growth curve from an individual student contains only few data points, but strength can be borrowed across students to stabilize individual growth curve estimates. This conventionally results in individual curves shrunk toward a smoother, population-based growth version. Another classic EB example is in the study of school effects through hierarchical linear modeling (Raudenbush, 1986). The author sets up a multi-level hierarchical model with covariates at individual and school levels.

The EB approach capitalizes on using information from higher-level statistical units to enhance estimates at individual level, without subjective input on prior distributions. From a Bayesian perspective, the EB approach is regarded as an approximation to a fully Bayesian approach. The general notion of borrowing information

Table 2 EB mean, within-, and between-imputation variance for 1998 NAEP reading proficiency by subgroup

	<i>EB estimate of Mean</i>	<i>Within-imputation variance</i>	<i>Between-imputation variance</i>
Overall	0.506	0.00015	0.00006
Male	0.302	0.00024	0.00013
Female	0.713	0.00023	0.00012
White	0.750	0.00031	0.00011
Black	−0.141	0.00048	0.00044
Hispanic	−0.076	0.00092	0.00034
Asian	0.772	0.00342	0.00141
Native American	0.070	0.00952	0.00663
<HS	−0.133	0.00094	0.00047
HS	0.210	0.00045	0.00031
>HS	0.644	0.00052	0.00034
College	0.838	0.00021	0.00017

$M = 100$ plausible values are used in this calculation.

The last four subgroups are categories from parent education: less than high school, high school, post high school, and college.

across statistical units is powerful and extends readily to many innovative applications. A recent example is in simultaneous multiple hypothesis testing (Efron, 2004). Unsurprisingly, EB has been regarded as one of the most important advances in the field of statistics since World War II (Efron 2007).

See also: Beyond Black Nationalism to Black Internationalism: The Compelling Case for Examining Black Educational Challenges Globally; Generalized Linear Mixed Models; Point Estimation Methods with Applications to Item Response Theory Models; Value-Added Models.

Bibliography

- Beaton, A. E. and Johnson, E. G. (1992). Overview of the scaling methodology used in the National Assessment. *Journal of Educational Measurement* **26**, 163–175.
- Braun, H. (2006). Empirical Bayes methods. In Green, J., Camilli, G., and Elmore, P. (eds.) *Complementary Methods for Research in Education*. Washington, DC: American Educational Research Association.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Association, Series B* **39**, 1–38.
- Efron, B. (2004). Large-scale simultaneous hypothesis testing: The choice of a null hypothesis. *Journal of the American Statistical Association* **99**, 96–104.
- Efron, B. (2007). The future of Statistics, *Amstat News*, September Issue. Virginia: The American Statistical Association.
- Hedges, L. V. (1987). The meta-analysis of test validity studies: Some new approaches. In Wainer, H. and Braun, H. (eds.) *Test Validity*. Hillsdale, NJ: Lawrence Erlbaum Inc.
- Huynh, H. (1977). Two simple classes of mastery scores based on the beta binomial model. *Psychometrika* **42**, 601–608.
- Johnson, E., Mislevy, R. J., and Thomas, N. (1996). Scaling procedures. In Johnson, E. and Carlson, J. (eds.) *The NAEP 1992 Technical Report* (1996). Washington, DC: US Department of Education.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Mahwah, NJ: Erlbaum.
- Maritz, J. S. (1969). Empirical Bayes estimation for the Poisson distribution. *Biometrika* **6**, 349–359.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., and Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics* **29**, 67–101.
- Mislevy, R. J. (1991). Randomization-based inferences about latent variables from complex samples. *Psychometrika* **56**, 177–196.
- Mislevy, R. J., Johnson, E. G., and Muraki, E. (1992). Scaling procedures in NAEP. *Journal of Educational Statistics* **17**, 131–154.
- Novick, M. R. and Lewis, C. (1974). Prescribing test length for criterion referenced measurement. In Harris, C. W., Alkin, M. C., and Popham, W. J. (eds.) *Problems in Criterion Referenced Measurement*. Los Angeles: Center for the study of evaluation, University of California.
- Plewis, I. (2000). Evaluation of educational interventions using multilevel growth curve: The case of reading recovering. *Educational Research and Evaluation* **6**, 83–101.
- Pituch, K. (2001). Using multilevel modeling in large-scale planned variation educational experiments: Improving understanding of intervention effects. *Journal of Experimental Education* **69**, 347–372.
- Raudenbush, S. W. (1986). A hierarchical model for studying school effects. *Sociology of Education* **59**, 1–17.
- Raudenbush, S. W. and Bryk, A. S. (1985). Empirical Bayes meta-analysis. *Journal of Educational Statistics* **10**, 75–98.
- Robbin, H. (1955). An empirical Bayes approach to statistics. *Proceedings of the Third Berkeley Symposium* **1**, 157–163.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.
- Scott, S. and Ip, E. H. (2002). Empirical Bayes and item clustering effects in a latent variable hierarchical model: A case study from the National Assessment of Educational Progress. *Journal of the American Statistical Association* **97**, 409–419.
- Van Duijn, M. A. J. and Bockenholt, U. (1995). Mixture models for the analysis of repeated count data. *Applied Statistics* **4**, 473–485.

Further Reading

- Braun, H. I., Jones, D. H., Rubin, D. B., and Thayer, D. T. (1983). Empirical Bayes estimation of coefficients in general linear model for data of deficient rank. *Psychometrika* **48**, 171–181.
- Braun, H. I. and Zwick, R. (1993). Empirical Bayes analysis of survival curves: Applications to the analysis of degree attainment. *Journal of Educational Statistics* **18**, 285–303.
- Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd edn. Boca Raton, FL: Chapman and Hall/CRC.
- Carlin, B. P. and Louis, T. A. (2000). Empirical Bayes: Past, present, and future. *Journal of the American Statistical Association* **95**, 1286–1289.
- Casella, G. (1985). An introduction to empirical Bayes data analysis. *American Statistician* **39**, 83–87.
- Dagne, G. A., Howe, G. W., Brown, C. H., and Muthen, B. O. (2002). Hierarchical modeling of sequential behavioral data: An empirical Bayesian approach. *Psychological Methods* **7**, 262–280.
- Efron, B. (2005). Bayesians, frequentists, and scientists. *Journal of the American Statistical Association* **100**, 1–5.
- Rubin, D. (1980). Using empirical Bayes techniques in the law school validity studies (with discussion). *Journal of the American Statistical Association* **75**, 801–827.
- Van der Linden, W. J. and Eggen, T. J. H. M. (1986). An empirical Bayes approach to item banking. *Applied Psychological Measurement* **10**, 345–354.
- Zwick, R., Thayer, D. T., and Lewis, C. (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement* **36**, 1–28.

Evaluation Research

T Kellaghan, St. Patrick's College, Dublin, Ireland

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Constructivist evaluation – Seeks information based on stakeholders' "constructions" of an evaluation question or issue in a process in which evaluator and stakeholders communicate with each other, engaging in dialectical process aimed at achieving as much consensus as possible among the varying perspectives of stakeholders.

Deliberative democratic evaluation – Is inclusive (all relevant interests are represented), dialogical (stakeholders and evaluators engage in dialogue), and deliberative (conclusions emerge from reasoning, reflection, and debate).

Empowerment evaluation – Is a collaborative group activity which employs evaluation concepts, techniques, and findings, to help people help themselves in providing a program.

Evaluation model – Is a term used to describe the main concepts, structure, and procedures that are used to guide an inquiry designed to assess an object, program, practice, activity, or system to arrive at defensible descriptions and judgments.

Ideographic inquiry – Focuses on the uniqueness of individuals or events, describing them in terms of their individual attributes and interrelationships between attributes.

Interpretative assumptions – Underlie approaches to evaluation that incorporate the perspectives of participants and in which the evaluator recognizes the influence that his/her role can have on the evaluation process.

Nomothetic inquiry – Has as its aim the discovery of general laws or principles applicable to all individuals (or individuals in specified categories).

Responsive evaluation – Is oriented toward program activities rather than program intents, responds to audience requirements for information, is based on observation and reaction rather than on formal measurement procedures, and takes account of different value perspectives in reaching conclusions.

Scientific management movement – Is the term applied to efforts early in the twentieth century to apply scientific principles to the management of industry, the division of labor, and the training and

supervision of employees (sometimes called Taylorism, named for its main exponent Frederick Taylor).

The term evaluation research has appeared in the educational literature from time to time, at least since the 1970s. There is, however, lack of agreement on its definition. This may be because the activities represented by the term have become more disparate over the years to reveal major differences in underlying assumptions, purposes, procedures, and intended uses. Or it may arise from problems in distinguishing research that can be described as evaluative from other types of educational research.

A definition of evaluation research must take account of the fact that the term represents the intersection of two activities: evaluation and research (see **Figure 1**). Not all evaluation has a research component, and not all research involves evaluation. With this in mind, together with a consideration of the activities to which the term is usually applied, evaluation research may be defined as a form of disciplined and systematic inquiry that is carried out to arrive at an assessment or appraisal of an object, program, practice, activity, or system, with the purpose of providing information that will be of use in decision making.

Basing evaluation research on disciplined and systematic enquiry, which usually involves the methodology of the behavioral or social sciences, is an attempt to control bias and to reach conclusions that will be defensible and open to scrutiny. This distinguishes evaluation research from the evaluation activity of everyday life in which objects or events are judged as good or bad, adequate or inadequate, of good quality or of poor quality, frequently in an intuitive, informal, impressionistic, and possibly biased way. Evaluation research also differs from activities that involve educational evaluation but not research, such as the application of the technology of testing (e.g., to diagnose student learning difficulties, to select students for further education).

The fact that evaluation research is carried out to arrive at an assessment or appraisal is usually interpreted as implying that, compared to basic, pure, or fundamental research (which has as its objective the creation of knowledge), it has a practical orientation. More specifically, it is designed to obtain information that can be acted on in policy formation or management decisions.

This interpretation is reinforced by the fact that evaluation research is a highly politicized activity. The decision to

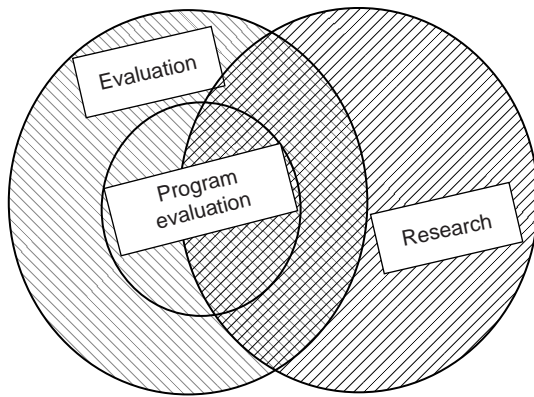


Figure 1 The intersection of evaluation and research.

carry out an evaluation is itself political, and its findings are likely to enter an immensely complex web of social decision making to which many stakeholders may have an input. Furthermore, when the object of an evaluation is an intervention program, results can threaten the conceptual basis of the program, undermine the reputation and careers of those involved, and affect jobs and staff morale (Weiss, 1991). Add to these conditions the fact that research and evaluation findings may challenge traditional views, vested interests, or current fashions, and the reason why findings have not had the same impact on practice as research in medicine or agriculture becomes clearer.

Program Evaluation

Much early evaluation research focused on aspects of a service or ongoing cyclical activities, such as teaching methods and teacher education, or a single variable, such as class size. Since the 1960s, the evaluation of programs (a rather loose term to describe a coordinated well-specified set of activities directed at achieving goals, often involving the design and implementation of new approaches to educational provision or service configuration) achieved considerable prominence. A program may stand alone, it may be implemented in a single site, or it may be large scale, complex, and implemented in multiple sites.

The components of a program evaluation are encapsulated in Stufflebeam's (2003) CIPP model, which is presented as a comprehensive framework for conducting and reporting evaluations. Corresponding to the letters in the acronym CIPP, the model's core concepts are context evaluation (the assessment of needs, problems, and opportunities as bases for defining goals and priorities, and for judging the significance of outcomes); input evaluation (identifying resources and determining if they are adequate); process evaluation (an ongoing check on the implementation of planned activities); and product evaluation (the identification of intended and unintended outcomes, both to help keep the process on track and to determine effectiveness).

A distinction is frequently drawn between the formative and summative evaluation of programs. Formative evaluation is designed to support program improvement and is normally requested by, or carried out by, an agent that is in a position to effect improvement. Summative evaluation, on the other hand, is carried out for, or by, decision makers who need the conclusions of an evaluation for reasons other than development.

Program evaluation is not synonymous with evaluation research. In **Figure 1**, it is depicted as a subset of both evaluation research and nonresearch evaluation activity. While some program evaluations involve a research methodology, others, such as educational connoisseurship, are based on expert judgment. Furthermore, much activity that would be considered evaluation research involves variables (e.g., class size) rather than programs.

A program evaluation will have as its primary objective a description of the functioning and effectiveness of a specific program. However, there is also the expectation that a synthesis of the findings of a number of evaluations could provide evidence that would justify more widespread application, providing guidance for decisions and reforms (see, e.g., What Works Clearinghouse at the Institute of Education Sciences, US Department of Education; Best Evidence Encyclopedia at the Center for Data-Driven Reform in Education at the Johns Hopkins University). It has been argued that support for generalization will be enhanced if evaluations are based on a framework that provides a theoretical rationale for a program (why treatments and implementation processes were chosen) and a description of the mechanisms (often complex) through which program activities are understood to have contributed to intended outcomes (Chen, 1990).

Methodology of Evaluation Research

The methodology of evaluation research is usually categorized as quantitative or qualitative. Quantitative methods emulate as far as possible the logic and methodology of the physical sciences, while qualitative approaches are more attuned to the complexities of the human condition.

Quantitative Methods

Quantitative measures include standardized tests, questionnaires, and a variety of rating scales. The classic quantitative approach also employs an experimental design in which participants are randomly assigned to treatment and control groups. Following the implementation of a treatment, criterion measures are administered and, because participants have been randomly allocated, observed differences in outcome between groups can reasonably be attributed to the differential effect of the treatment. Internal validity is established if evidence supports

the conclusion that experimental treatments were effective in a specific experimental instance. Further evidence, however, is required to establish if the effect can be generalized to other situations (external validity). A frequently cited experimental study was carried out in Tennessee, in which more than 6000 kindergarten students in 329 classrooms were randomly assigned to one of three conditions: small classes (13–17 students), regular classes (22–26 students), and regular classes with a full-time teacher aide. Students remained in their classes from kindergarten to the end of third grade, at which point a comparison of the groups on tests of reading and mathematics revealed a favorable effect of small class size (Finn and Achilles, 1999).

Large-scale experimental studies are rare because of practical and ethical difficulties in randomly assigning participants to treatment and control groups in real-life situations. In their place, a range of quasi-experimental designs, in which treatment is manipulated and outcomes are observed, but which do not involve random assignment to conditions, may be used. For example, if a comparison group is involved, it will be chosen to be as similar as possible to the treatment group (in particular, attending to the possibility of selection bias), information on variables of interest is obtained before and after treatment (pretest and posttest), and covariance analysis or regression may be used to adjust posttest measures, as long as initial differences are not large.

Surveys, a further form of quantitative inquiry, are frequently employed to assess the achievements of students in an education system using standardized tests. While they do not involve a comparison group, it is usually possible to compare the achievements of subpopulations (e.g., males and females; students in urban and rural locations). The best-known example of a survey was commissioned by the US government in the mid-1960s to describe lack of equal educational opportunities in the public school system. The study found that, with some exceptions, minority group students through the grades had lower achievement scores than white students. While the schools attended by minority students also had poorer resources, variation in facilities and curricula accounted for relatively little variation in student achievement (Coleman *et al.*, 1966). The then commissioner of education, Harold Howe II, noted that the US Office of Education would determine how it could use the results to enhance educational opportunities.

Qualitative Methods

It has been argued that, while quantitative studies attempt to match the standards and objectivity of the physical sciences, they fail to take account of the uniqueness and complexities of human behavior. Qualitative studies address this issue when they adopt a naturalistic posture, which recognizes the phenomenological character of the evaluation context,

attempts to represent a variety of perspectives reflecting multiple realities, and allows understanding to emerge.

Qualitative methods are not easily defined, partly because they are derived from several research traditions, which include ethnography, document analysis, interviews, participant observation, and deliberations of focus groups. Data are frequently collected in the context of a case study, which has a number of features: the holistic in-depth examination of a program and of its internal workings; the collection of information from a multiplicity of sources; and the development, testing, and replication of theoretical propositions to arrive at generalizations.

Mixed Methods

On the assumption that methods can be complementary, many commentators encourage flexibility in deciding on methodology, and propose the use of a mixture of quantitative and qualitative methods. It is argued that examining a phenomenon from a variety of perspectives (involving, e.g., multiple operationalization of key constructs, multiple measures, and multiple investigators) facilitates a process known as triangulation (a term borrowed from surveying), in which combining a variety of types of data serves to illuminate, confirm, or challenge conclusions. Quantitative methods will usually be chosen to provide standardized replicable findings on large data sets, and qualitative methods to elucidate the cultural context, themes, dynamics, and internal relationships of programs. Creswell (2003) provides guidelines for the use of mixed methods relating to priority (which approach will be dominant), implementation (in which a clear sequence of activities is prepared; e.g., a survey followed by ethnographic exploration of identified trends), integration (how different types of data will be combined), and theory (the philosophical assumptions underlying methods). While a consideration of philosophical assumptions must raise doubts about the appropriateness of using methods in a study which are based on radically different conceptions of reality, of the nature of knowledge, and of the relationship between evaluator and what is evaluated, the use of mixed methods is increasing.

Early Development of the Field of Evaluation Research

It was not until toward the end of the nineteenth century that developments in the emerging behavioral sciences, emulating the logic and methodology of the physical sciences, provided models for empirical investigation that could form a basis for educational research. At the time, a distinction was drawn between a nomothetic approach, with the aim of establishing general laws governing human and social behavior (leading to explanation of an event as an instance of a law) and an ideographic approach which focused on the

unique elements of individual phenomena (leading to understanding). The distinction resonates to this day in the views of proponents of quantitative experimental procedures and proponents of qualitative naturalistic inquiry.

Evaluation research and its development have been predominantly American-based. In light of this, it is not without interest that the origins of empirical research from which evaluation research later developed were in Europe – in Germany (where Wundt established his laboratory), in France (where Binet devised a method of mental measurement), and in Britain (where Galton, Fisher, and Spearman laid the foundations of statistical analysis). However, what is generally regarded as the first formal evaluation (though it was not called that at the time) was carried out toward the end of the nineteenth century in the United States when Joseph Rice set the scene for the paradigm that was to dominate evaluation research for much of the twentieth century in his study of the effects of instructional time on students' spelling ability. The study was marked by two features: the use of quantitative measures of students' achievements, and a comparative research design (in which the amount of time devoted to spelling instruction was varied).

The early decades of the twentieth century saw the quantitative approach flourish, mainly in the form of surveys of student achievement and school conditions. Under the influence of the scientific management movement, studies were carried out to identify school and teacher efficiency, to diagnose specific system weaknesses, to standardize curriculum practice, to evaluate innovations, and to assess the overall performance of a local education system.

Ralph Tyler, who is generally regarded as laying the foundations of evaluation research as we now know it, signaled two important developments in his 8-year study (Smith and Tyler, 1942). First, evaluation was conceptualized as involving a comparison of intended and actual student achievement, obviating the need for costly and disruptive comparisons between experimental and control groups. Second, objectives of programs were defined in behavioral terms, and formed the basis of instrument development.

The role and extent of evaluation research changed radically in the late 1950s and 1960s as a result of two initiatives of the United States federal government. First, in the wake of the launch of Sputnik in the Soviet Union (in 1957), government responded to lobbying for reform by providing funds to support the introduction of new curricula to schools with the objective of improving achievement, particularly in mathematics, science, and foreign languages. Second, in response to concern about equality of educational opportunity, a broad array of initiatives designed to raise student achievement in socially and economically disadvantaged communities was launched in the war on poverty. The fact that the initiatives had associated requirements for evaluation gave a major boost to evaluation

research. The response was an increase in research activity which favored the use of experimental/randomized or quasi-experimental designs as these were perceived to have the greatest potential to determine if programs were meeting goals, to study cause-effect relationships, and to provide policy guidance. There are many examples of short- and long-term evaluations of school curricula and of intervention programs to serve children in disadvantaged areas in the United States.

Evaluation activity throughout the world, though not as extensive as in the United States, paralleled or was influenced by American practice. During the 1960s and 1970s, for example, as in the USA, evaluations of social and educational interventions and programs were carried out in European countries. Compared to practice in the USA, however, qualitative methods were preferred to quantitative methods. From the 1980s onward, under pressure from the European Economic Community (later the European Union), the focus in evaluation shifted to results and issues relating to quality, accountability, and effectiveness. In many countries, the methodology of evaluation research came to be seen as having a role to play in the management of public affairs, serving to identify performance targets, indicators, and outcomes.

The Professionalization of Evaluation Research

As demand for evaluation by government agencies grew through the 1960s and 1970s, the term evaluation came into common use. Professional organizations with a focus on program evaluation were formed, publications (books and journals) dealing with the topic appeared, and courses in evaluation were established in universities. Evaluation was developing as a field distinct from, but related to, the older disciplines of research and testing.

A variety of evaluation models emerged which reflected greater appreciation of the complexity of evaluation activity, together with greater understanding of how it fitted into organizational and bureaucratic decision making. Increased recognition was accorded the role of political factors and the competing interests of stakeholders, and how they might be accommodated. New conceptualizations of evaluation began to emerge as a range of issues were addressed: the evaluation of program goals; the identification of unintended as well as intended outcomes; descriptions of program inputs and program strategies, and how they might change over time; monitoring of variation in implementation and service delivery; and descriptions of the characteristics of participants and how they responded to their experiences.

Development, however, faltered as social science research in general suffered a set-back in funding in the 1980s following a change in political climate (particularly

in the USA and the UK) which signaled increased interest in measuring the outcomes of programs or systems, often in the context of accountability, quality assurance, and the management of resources. The most outstanding examples of this development are to be found in assessments of student achievement at state level in the USA (e.g., the mandating of academic learning standards and the assessment of student achievement with respect to the standards), at national level (e.g., the US National Assessment of Educational Progress), and at international level (e.g., the Organization for Economic Cooperation and Development (OECD) Program for International Student Assessment (PISA)). Such government-sponsored activities, which are a feature of education systems throughout the world since the 1990s, now account for a considerable amount of the financial resources available for research in education. They resemble activity in the early decades of the twentieth century, except that they are carried out at state, national, and international, not local, level.

Quis Custodiet?

The question posed by the first/second-century Roman satirist Juvenal, *quis custodiet ipsos custodes?* (who will watch over the guardians?) has frequently been asked in the context of the exercise of power, and so is appropriate in the case of evaluation. Given the power that evaluators possess, and the potential for their findings to impact on a variety of individuals, we may ask what procedures are available to monitor or influence the conduct of an evaluation. Four such procedures can be identified. Their identification, however, should not be taken to imply that they are widely used or necessarily effective.

Conformity to Standards

A number of sets of principles or standards that should govern evaluation practice (in particular, program evaluation) are available. Principles established by the American Evaluation Association (AEA) specify that (1) inquiry should be systematic; (2) evaluators should be competent; (3) the integrity of the process should be ensured; (4) the security and dignity of all involved should be respected; and (5) account should be taken of the public interest (Shadish *et al.*, 1995).

While the AEA principles were directed at evaluations across a variety of government and social service sectors, the focus of the Joint Committee on Standards for Educational Evaluation (1994) was on education. The joint committee specified 30 standards categorized under four headings: utility (evaluations should be informative, timely, and influential), feasibility (evaluations should be realistic, prudent, diplomatic, and cost effective), propriety (evaluations should be lawful and ethical), and

accuracy (evaluations should provide accurate information about the merits of a program).

Meta-evaluation

A meta-evaluation is an evaluation of an evaluation designed to assess its quality and to help users decide whether or not to accept and act on its conclusions. While a variety of procedures have been used for this purpose, including consultants' evaluations, secondary analyses of data, and multiple replications, the most systematic description of the procedures required for meta-evaluation has been provided by Stufflebeam (2000). These specify interaction with the meta-evaluation stakeholders, identification of a qualified meta-evaluation team, a review of available information, the collection of new information if required (e.g., in interviews or surveys), a judgment of the evaluation's adherence to standards (e.g., the standards of the Joint Committee on Standards for Educational Evaluation), and the preparation and submission of reports.

Involvement of Stakeholders in Designing and Monitoring an Evaluation

The involvement of stakeholders in planning and designing an evaluation and in monitoring its progress (e.g., in a steering committee) has been proposed as a way of ensuring that some control of an evaluation is assigned to those who will be affected by its findings. It has the further advantage of providing the opportunity for the perspectives and values of stakeholders to be addressed and reflected, while increasing the likelihood that the findings will be acted on. Whether only a select representative group of users or all relevant interests should be represented is a matter of debate.

Involvement of Stakeholders in the Conduct of an Evaluation

A step beyond involving stakeholders in determining the main features of an evaluation and monitoring its implementation is to consider their perceptions, values, and interests as central, and to respond to them in reporting findings, as proposed, for example, in the deliberative democratic evaluation of House and Howe (1999) and the responsive evaluation of Stake (1976). More extreme views regarding power in an evaluation have been expressed by Guba and Lincoln (1989) who proposed that an evaluation report should comprise an interpretation of the meanings that stakeholders construct in the course of an evaluation (constructivist evaluation) and by Fetterman (1994) who saw the evaluator as having a merely technical role in a process in which stakeholders are empowered to carry out their own evaluation, shaping its direction, suggesting solutions to problems, and ultimately taking an active role in effecting social change (empowerment evaluation).

Conclusion

Developments in evaluation research during the twentieth century have resulted in a much richer view of its complexity, and issues that might have been considered relatively straightforward even in the middle of the century no longer seem so. The field also has fractured, however, as conflicts over theory and methodology emerged, relationships between researcher and object of research were reconstructed, and realist assumptions were challenged by interpretative ones.

Meanwhile, in response to the needs of bureaucrats, the focus of evaluation research was being narrowed to address managerial concerns, to help define performance, to set targets and performance indicators, to use student testing as a policy lever, and to base resource-allocation decisions on performance information. From this perspective, there is little interest in monitoring program implementation or in determining why some practices or institutions are effective, and others not. All one needs is to be able to identify the satisfactory and unsatisfactory, often cloaked in terms of quality and lack of quality, with statistical data to support the judgment. A competitive market and threats or promises of sanctions should lead to raised expectations for students, improved teaching, and an increase in students' test scores. In this situation, many of the activities and issues associated with evaluation research described in this article will be redundant. As for the future, the wide range of conflicting interests and ideologies, both within and outside the evaluation community, suggest that major efforts will be required if evaluation research is to develop to the satisfaction of evaluators working in the tradition of social science research, while at the same time fulfilling the role always envisaged for it – the provision of information that will serve the needs of decision makers.

See also: Meta Analysis.

Bibliography

Chen, H.-T. (1990). *Theory-Driven Evaluations*. Newbury Park, CA: Sage.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., et al. (1966). *Equality of Educational Opportunity*. Washington, DC: National Center for Educational Statistics.

Creswell, J. W. (2003). *Research Design. Qualitative, Quantitative, and Mixed Methods Approaches*, 2nd edn. Thousand Oaks, CA: Sage.

Fetterman, D. M. (1994). Empowerment evaluation. *Evaluation Practice* 15(1), 1–15.

Finn, J. D. and Achilles, C. M. (1999). Tennessee's class size study: Findings, implication, misconceptions. *Educational Evaluation and Policy Analysis* 21, 97–109.

Guba, E. G. and Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Newbury Park, CA: Sage.

House, E. R. and Howe, K. R. (1999). *Values in Evaluation and Social Research*. Thousand Oaks, CA: Sage.

Joint Committee on Standards for Educational Evaluation (1994). *The Program Evaluation Standards. How to Assess Evaluations of Educational Programs*. Thousand Oaks, CA: Sage. <http://www.wmich.edu/evalctr/JC-Home.htm> (accessed August 2009).

Shadish, W. R., Newman, D. L., Scheirer, M. A., and Wye, C. (eds.) (1995). *Guiding Principles for Evaluators. New Directions for Program Evaluation*, vol. 66. San Francisco, CA: Jossey-Bass. <http://www.eval.org/Publications/GuidingPrinciples.asp> (accessed August 2009).

Smith, E. R. and Tyler, R. W. (eds.) (1942). *Appraising and Recording Student Progress*. New York: Harper and Row.

Stake, R. E. (1976). A theoretical statement of responsive evaluation. *Studies in Educational Evaluation* 2, 19–22.

Stufflebeam, D. L. (2000). The methodology of metaevaluation. In Stufflebeam, D. L., Madaus, G. F., and Kellaghan, T. (eds.) *Evaluation Models*, 2nd edn., pp 457–471. Boston, MA: Kluwer Academic.

Stufflebeam, D. L. (2003). The CIPP model for evaluation. In Kellaghan, T. and Stufflebeam, D. L. (eds.) *International Handbook of Educational Evaluation*, pp 31–62. Boston, MA: Kluwer Academic.

Weiss, C. H. (1991). Evaluation research in the political context: Sixteen years and four administrations later. In McLaughlin, M. W. and Phillips, D. C. (eds.) *Evaluation and Education: At Quarter Century. Ninetieth Yearbook of the National Society for the Study of Education Part II*, pp 211–231. Chicago, IL: NSSE.

Further Reading

Alkin, M. C. (ed.) (2004). *Evaluation Roots. Tracing Theorists' Views and Influences*. Thousand Oaks, CA: Sage.

Barnett, W. S. and Boocock, S. S. (eds.) (1998). *Early Care and Education for Children in Poverty: Promises, Programs, and Long-Term Results*. Albany, NY: State University of New York Press.

Blalock, A. B. (1999). Evaluation research and the performance management movement. From estrangement to useful integration? *Evaluation* 5, 117–149.

Clarke, A. (1999). *Evaluation Research. An Introduction to Principles, Methods and Practice*. Thousand Oaks, CA: Sage.

Denzin, N. K. (1978). *Sociological Methods: A Sourcebook*. New York: McGraw Hill.

Greaney, V. and Kellaghan, T. (2008). *Assessing National Achievement Levels in Education*. Washington, DC: World Bank.

Julnes, G. and Rog, D. (eds.) (2007). *Informing Federal Policies on Evaluation Methodology: Building the Evidence Base for Method Choice in Government Sponsored Evaluation. New Directions for Evaluation* 113, San Francisco, CA: Jossey-Bass.

Kellaghan, T. and Stufflebeam, D. L. (eds.) (2003). *International Handbook of Educational Evaluation*, 2 vols. Boston, MA: Kluwer Academic.

Madaus, G. F. and Stufflebeam, D. L. (eds.) (1989). *Educational Evaluation: Classic Works of Ralph W. Tyler*. Boston, MA: Kluwer Academic.

McLaughlin, M. W. and Phillips, D. C. (eds.) (1991). *Evaluation and Education: At Quarter Century. Ninetieth Yearbook of the National Society for the Study of Education, Part II*. Chicago, IL: NSSE.

Patton, M. Q. (1997). *Utilization-Focused Evaluation: The New Century Text*, 3rd edn. Thousand Oaks, CA: Sage.

Scriven, M. (1991). *Evaluation Thesaurus*, 4th edn. Newbury Park, CA: Sage.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Shadish, W. R., Cook, T. D., and Leviton, L. C. (1991). *Foundations of Program Evaluation. Theories of Practice*. Newbury Park, CA: Sage.

Slavin, R. E. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher* 37(1), 5–14.

Stufflebeam, D. L., Madaus, G. F., and Kellaghan, T. (eds.) (2000). *Evaluation Models. Viewpoints on Educational and Human Services Evaluation*, 2nd edn. Boston, MA: Kluwer Academic.

Exploratory Data Analysis

A Unwin, University of Augsburg, Augsburg, Germany

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Exploratory data analysis (EDA) is different from classical statistics. It is not about fitting models, parameter estimation, or testing hypotheses, but is about finding information in data and generating ideas. The data being studied may be an independent random sample from a population of interest, but they could just as easily be a collection of data whose provenance is not fully known. EDA is valuable for generating hypotheses, while classical methods are important for testing them.

Researchers and analysts have always been carrying out EDA, even if they did not use that name for it. EDA involves looking at the data in many different ways to gain insights, and incorporating domain knowledge about the data into the analyses.

What is EDA?

EDA is an approach, a methodology rather than a set of techniques. Tukey suggested that carrying out a data analysis meant acting like a detective. You have a general goal in mind, but you should remain open to all kinds of evidence and you cannot be sure in which direction that evidence might lead you. On the other hand, he regarded carrying out classical statistical analyses as acting more like a judge. Hypotheses have to be clearly defined in advance, only certain kinds of evidence are permissible, and there are very specific rules for proceeding. All that this means is that EDA is more of an art than a science. There are general principles that should be followed, but not many formal procedures. It is flexible and data driven rather than regulated and hypothesis driven. Each person will carry out a data analysis in a different way – though they should all discover the same information.

Most introductory statistics textbooks include a discussion of descriptive statistics: usually some standard graphics (e.g., histograms and scatter plots), some tables (frequency and contingency), and summary statistics of location (mean, median, and mode) and of variability (variance, range, and standard deviation). Whether more than this (or less!) is covered depends on the book. The emphasis is on description and not on induction. The graphics and statistics are used in a passive rather than an active way. EDA employs the tools of descriptive statistics, but with the aim of extracting information. The idea is not to print off summary tables,

statistics, and graphics for all variables as some kind of mandatory, although unwelcome task. Instead, the goal is to explore the data in a revealing and insightful way. This demands fast and flexible software, something that is not always available. Large statistical packages tend to be like oil tankers, hugely powerful and difficult to maneuver. The importance for EDA of how tables and graphics are presented and organized should not be underestimated. Analysts need displays that can be quickly assimilated and, if appropriate, immediately investigated in more detail. Being able to discard displays that are currently irrelevant is also valuable. Nothing inhibits clear thinking more effectively than cluttered screens on which you cannot see the wood for the trees (or should that be “on which you cannot see the content for the windows?”). It is quite possible that the need for statisticians to work with the powerful, large packages that are available to carry out standard analyses has discouraged the use of EDA.

In recent years, there have been substantial advances in data mining. It is important to note that data mining is not the same as EDA. EDA can be used to analyze any data set, whereas data mining is used for investigating datasets with many variables, which have probably been collected for purposes other than the current analysis. Data mining applies semi automatic methods to try to uncover information before involving the user. EDA should be carried out with direct user involvement to incorporate domain knowledge. For large data sets with few variables, EDA can generally be performed in the usual way. For large datasets with many variables, some initial screening analyses using the data-mining approach may be very helpful, though user knowledge of what is likely relevant and important may be more helpful still. EDA is concerned with the meaning and interpretation of data, that is, more with understanding than with prediction, to use Breiman's distinction (Breiman, 2001).

Why EDA?

There are many different reasons for carrying out EDA. An obvious initial step in any project is to discover what kind of data you are dealing with. What sort of variables are there? Are the data of high quality or are there erroneous values, missing values, or outliers? Are the data symmetric, skew, or clustered? Are there gaps in the data? Data quality is a particularly important issue and very difficult to handle

with formal methods. Every flawed dataset is flawed in its own way, and EDA is necessary to identify any weaknesses and to assess their impact.

EDA is often associated with initial data analyses, because it is easy to see how much it can help there. In fact, EDA can contribute at all stages of a project. It is useful for checking assumptions and for evaluating models (Gelman, 2004, makes the case strongly for incorporating exploratory graphical checks into the modeling process). It often complements more formal methods (for instance, using interactive mosaic plots to explain the results of log linear modeling of multivariate categorical data). It can be helpful in presenting results to others, showing how the properties of the data are reflected in models and vice versa.

EDA in Education

Quantitative research requires statistical modeling and statistical modeling requires EDA. There are studies on student performance (particularly international comparative studies like the Programme for International Student Assessment (PISA)). There are investigations of the effects of social background and school type. Some educational studies are observational (e.g., the school league tables that are produced in England and elsewhere) and EDA is important for assessing the characteristics of the populations studied and for qualifying how far any results can be generalized. Some studies are experimental, with groups carefully selected for the purposes of comparison. Here too, EDA makes a contribution, in monitoring how well the groups are matched.

EDA Methodology

Given data and a general goal of what the purpose of the analysis is, the first stage is to examine the variables individually, guided by domain experts. Tables, graphics, and basic statistics are all likely to be informative, but it is not sensible to simply draw and print out all possible displays! EDA should be discerning and sensitive, not unresponsive and inflexible. Results of initial EDA analyses will often identify various data-cleaning requirements and the need for further details on particular variables or cases. As always, this involves close cooperation with those who designed the study and collected the data. It may also become apparent that additional data will have to be collected.

Building on the knowledge acquired, one can continue with bivariate analyses, looking for associations between pairs of variables. At this point, data analysis and modeling start to overlap and various simple models may be built to gain understanding. When EDA was first discussed by Tukey, graphs had to be drawn by hand, and even

simple models could only be calculated with difficulty. Nowadays, large numbers of graphics can be drawn quickly and easily, and even complicated models can be fitted to data instantly. Statistical practice is only gradually catching up with the computing power that is already available. What is missing is not so much raw calculating capability as the tools to organize and integrate the resulting mass or mess of results.

Classical modeling is not the only approach for investigating the structure of data sets; there are many multivariate methods that may be applied, such as principal component analysis, multidimensional scaling, correspondence analysis, or cluster analysis. These methods have two features in common: they involve some kind of data reduction and they have little solid statistical grounding. The lack of statistical grounding makes them seem like exploratory techniques, for there are then no fully objective criteria as to which results to accept. However, it is worth emphasizing that the data-reduction aspect of these methods is not always in keeping with the spirit of EDA, because the combinations of variables (e.g., principal components) or of cases (e.g., clusters) may not be easy to interpret.

In general, there are no formal restrictions on which techniques may be used in EDA. Just as it can be sensible to draw a bar chart of continuous data (e.g., to identify frequently occurring values), it can be helpful to use methods that are not strictly valid (e.g., *t*-tests for non-normal data). The aim of EDA is to uncover information in data. Any information brought to light must be tested and properly qualified before being passed on, but the process of discovery should not be hampered by overly restrictive regulations. Any statistical model or technique can be employed in an exploratory way. Just because a data set has not been collected as a random sample from a specified population does not mean tests cannot be carried out as if the data has been collected that way. However, any results obtained have to be treated with caution and used only as approximate guidelines with accompanying qualifications.

The informal blending of a broad range of methods is a distinctive feature of EDA. Another is the extensive use of interactive analyses and graphics. Interaction encourages the incorporation of background information and the ability to respond to results obtained to guide further analyses in the most promising directions.

Techniques for EDA

Although methodology is at the heart of EDA, there are a few techniques which are particularly associated with it. Tukey proposed some graphical tools in connection with EDA and they have remained so in people's minds — sometimes regrettably long after their usefulness has

expired. Stem and leaf plots were a neat idea in the days of pencil and paper, but would rarely be the graphic of choice today, when computers are available to everyone. Box plots, however, have proved very informative and are an efficient way of comparing several distributions in parallel. It is essential to remark that, despite Tukey's original clear definition of a box plot, many alternative definitions are in circulation. You always have to check what definition has been used to ensure that you interpret box plots properly. Given a choice, Tukey's box plot is the one to be recommended. The hinges are drawn at the medians of the upper and lower halves of the data. The inner fences are drawn to the furthest points from the hinges inside $1.5 \times (\text{difference between the hinges})$. Individual points outside the inner fences are marked with circles, and extreme outliers, that is, points outside $3 \times (\text{difference between the hinges})$, are drawn as asterisks. **Figure 1** shows an example of employment data from the 299 German election constituencies in 2005. The skewness of the distribution is unsurprising (though it is perhaps surprising at first sight, that there are no constituencies with a zero value).

Transformations are often mentioned in connection with EDA and Tukey considered them a fundamental component. They are useful for working with skewed data and in other situations, but they suffer from the disadvantage that they may not be easy to interpret on the original scale. If all possible transformations are allowed, then some kind

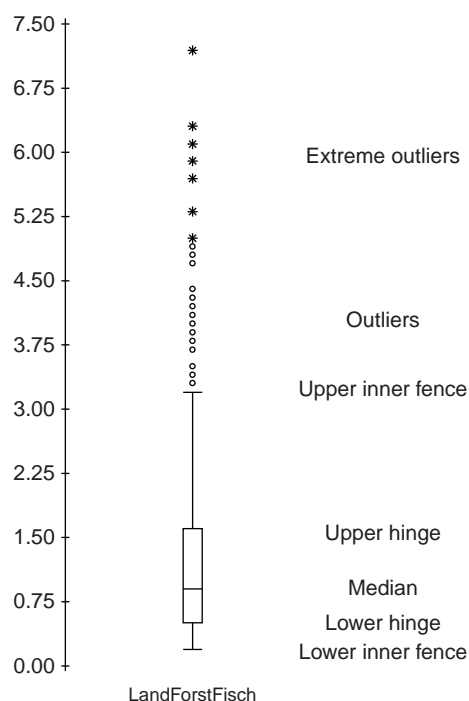


Figure 1 A Tukey box plot of employment rates in agriculture, forestry, and fisheries in the German election constituencies in 2005.

of association among variables will often emerge. Transformations should be made use of judiciously.

Descriptive statistics play a basic role in EDA, though they can be overrated. Calculating correlation coefficients of all possible pairs of variables has a good chance of producing some surprising (and perhaps entertaining) relationships. The reported positive association between risk of obesity and the numbers declaring their religion as Jedi in the 2001 UK census is only one of many examples.

Graphics are essential in EDA. Analytic methods are good for assessing prespecified characteristics (e.g., could these data have come from a normal distribution?), but will not provide any clarification (e.g., because the empirical distribution is skew, there are many outliers, the values are clumped). Graphical displays can often indicate the presence of informative or curious features, which deserve further attention. Birthday distributions in Germany commonly show a sharp peak for the 1st of January. Apparently, immigrants sometimes arrive with no documented day of birth, but just the year. In these cases, they are assigned a birthday of the 1st of January. Another initially surprising feature of birthday distributions can be found among young footballers and other sportsmen: most of them have birthdays early in the year and there is a steady decline from January to December. The explanation is that only players born in a particular year can play in a team and so the ones born earlier in the year are older and therefore bigger and stronger. You do not find features like this without drawing appropriate graphics. Missing value plots, such as shown in **Figure 2**, are elementary tools that can reveal much about raw data. Although missing

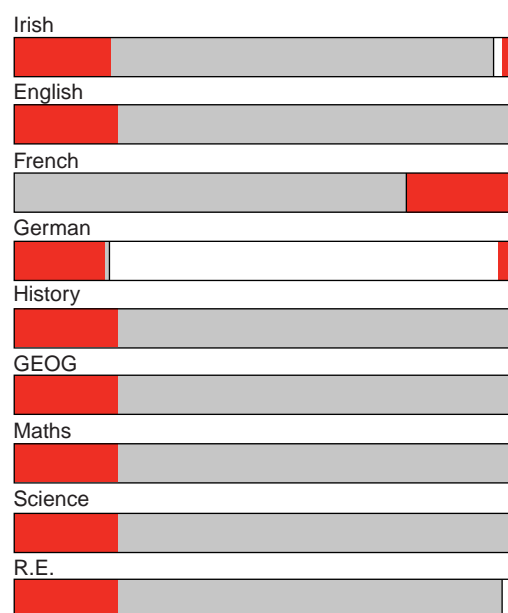


Figure 2 A missing value plot for the exam marks of 126 children taking a maximum of nine school subjects. The children missing a mark in French have been selected and highlighted.

values arise in real data sets quite often, they are rarely to be found in textbook examples. Perhaps this is why they are not used more often. (An example is shown in **Figure 2**.)

Static graphics are very helpful for presenting results (once you know what results you want to present), while interactive graphics are more effective for EDA. To get the most value from multivariate graphic displays, especially parallel coordinate plots for continuous data and mosaic plots for categorical data, they have to be interactive. The abilities to query graphics (what are those outlying cases?), to link them (which are the female students' marks?), to reformat them (sort departments by graduate admission rates), to zoom in and out, to color cases by group, and so on, all support successful analyses. More sophisticated interactive tools such as extended querying, alpha-blending, and selection sequences (Unwin *et al.*, 2006) are also valuable and more developments can be expected.

Graphics alone are not enough and it is important to scrutinize any features found graphically with more formal statistical tools. Graphical displays and statistical models are gradually becoming more integrated in software, which will promote more authoritative applications of EDA in practice.

EDA Literature

In most scientific fields, there is a plethora of books and articles on any subject. EDA is an exception, because it is not specific techniques or models that are most important, but it is the approach. There are similarities with the literature on scientific method. While philosophers have written extensively and theoretically, there are very few books on the practice of science itself. Beveridge's excellent *The Art of Scientific Investigation* (Beveridge, 1950) is a rare example.

There is one classic book on EDA, which was written by John Tukey (Tukey, 1977). This was the culmination of many years work by Tukey in trying to convince other statisticians that data analysis was important. This campaign began with his famous 1962 article in the *Annals of Statistics* (Tukey, 1962). Tukey was a major figure, who contributed many central ideas, and had a major influence on the direction of statistics. Interestingly, his book was written just a little too early. The approach he described had then to be carried out by hand. The first personal computers were only just appearing on the scene. Within a few years they had transformed how data analyses could be carried out. Calculations, which had formerly involved tedious work, could be performed instantaneously and excellent graphics could be drawn without meticulous preparation by hand. Interactive graphics also became possible. They were first illustrated in the picturing, rotation, isolation, and masking (PRIM)-9 system for viewing three-dimension

(3D) rotating plots, the subject of a famous instruction film by Tukey and his collaborators. The only trouble with PRIM-9 (and indeed to a lesser extent with some of the systems that followed) was the enormous cost: one story mentions US\$10 000 an hour in the mid-1970s!

Many statistics textbooks refer briefly to EDA without discussing it in detail. This is a pity, though understandable. It is easier to write about techniques than about methodology (at least in mathematically oriented subjects). Discussions of EDA are more likely to be found in graphically oriented books such as (Unwin *et al.*, 2006), (Cook *et al.*, 2007), and (Theus *et al.*, 2008).

Example

As an example, consider the exam marks of 126 Irish schoolchildren in nine different subjects. The marks in each subject are out of 100. **Figure 2** shows a missing value plot (Unwin *et al.*, 1996). There is a bar for each variable, split into two sections. Each case is represented once in each bar. The proportion of cases with values is marked in gray to the left and the proportion of missings is in white to the right. The white section of the bar for German shows that few children had a mark in that subject. Not surprisingly (this is a real data set), no variable had a complete set of values.

To investigate a potentially interesting subgroup, the children missing a mark in French have been selected by clicking on the white section of the bar for French. These cases are highlighted in all bars and show up darker in **Figure 2**. The two sections of each bar are highlighted separately, for non-missings on the left and for missings on the right. At most there are then four sections in each bar. For instance, the bar for Irish shows from left to right: the proportion with marks in Irish but not in French; the proportion with marks in both; the proportion with marks in French but not in Irish; and the proportion with marks in neither. Almost all those missing a mark in French have a mark in German, which suggests that there must have been a choice between French and German for the children. Close examination of **Figure 2** shows that one child apparently had marks in both subjects and two had marks in neither (these features can more easily be seen by toggling the highlighting or by querying the bars in the plot).

The distributions of individual marks are also revealing. Parallel box plots of the marks in the nine subjects are shown in **Figure 3**. Common scaling has been used and the subjects sorted by the median mark. A number of curious features can be seen. In some subjects, there is a much bigger range than others, with teachers being prepared to give very low or very high marks. In English, all marks were between 27 and 79, while in French the range was

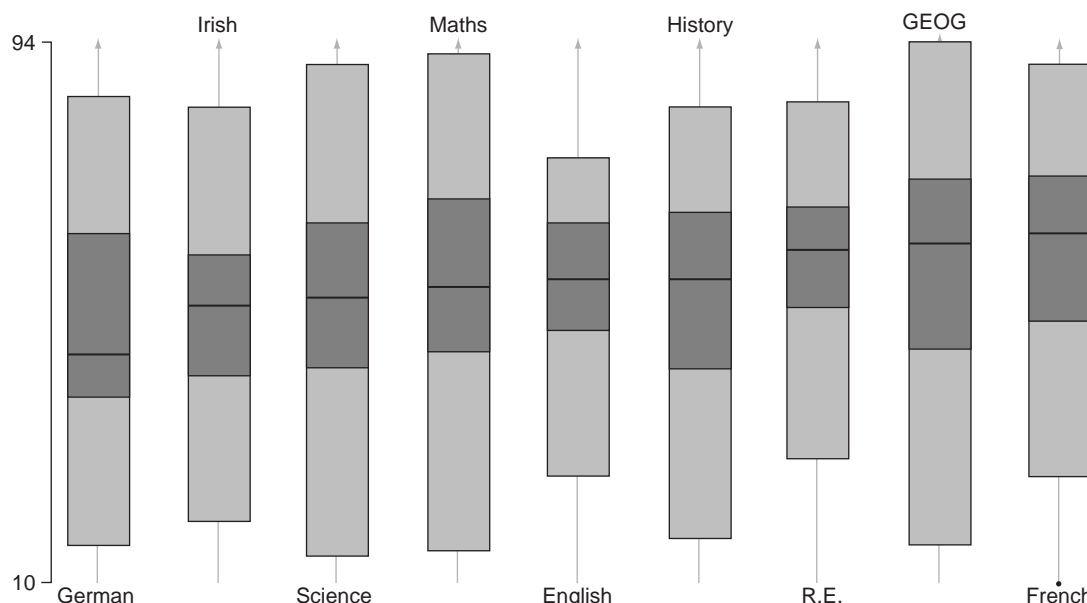


Figure 3 Parallel box plots of the exam results. All box plots are drawn to a common scale and they have been ordered by the median mark.

from 10 to 94. French had the highest median mark but the lowest individual mark.

To get more detail on the individual distributions, histograms may be drawn. Interactively adjusting the binwidth of the histogram for math marks suggests there might be a gap in the data just under the value 40. In **Figure 4**, the binwidths have all been set at exactly 2 with an anchor point of 0. There are gaps in math and history, but not in Irish or science. In the exam system in Ireland, 40 would usually be the pass mark and so a gap would not be surprising: some teachers may raise marks to 40 (a pass) or lower them to 37 (a clear fail) to avoid any discussion of marks with anyone getting 38 or 39. Others may not.

EDA Software

The main requirements for EDA software are a broad range of statistical tools, attractive graphics, and a multi-windows interface. No current package offers all of these capabilities though most provide some of everything and are continually improving in all categories. Interactive capabilities are the weak point of most of the larger packages, as it is difficult to combine many different models into one framework. The R language, originally an imitation of the S language developed for teaching, has gone from strength to strength and is, at the time of writing, the dominant software in statistical research. It can be used for EDA, but is by no means ideal. The same could be said of all the popular packages.

Several impressive researchers were attracted by the idea of interactive graphics early on and some excellent

pieces of commercial software appeared in the 1980s: MacSpin by the Donohoes, which implemented the ideas of PRIM-9, with a much better interface, for the Macintosh; Data Desk by Paul Velleman, which concentrated on providing a consistent implementation of interaction across all graphics types; JMP by John Sall, which was less strong in interactive graphics, but offered much more modeling power. Of these, MacSpin is long gone, partly because it really only provided rotating plots, while JMP and Data Desk are still available in substantially more advanced versions.

There have also been several interesting research software projects providing interactive tools. Peter Huber developed Interactive Statistical Programming (ISP) for interactive statistical analyses in general, but with less emphasis on graphics (Fleischer *et al.*, 1992). Xgobi and its successor Ggobi have extended the idea of rotating plots by incorporating projection pursuit criteria. In recent years, the Gobi group (Cook *et al.*, 2007) have broadened their approach to provide a wider range of graphics tools within their package. The Augsburg group has produced a number of software packages specifically for interactive graphics, including Heike Hofmann's MANET (Unwin *et al.*, 1996) (which introduced interactive mosaic plots and weighted plots) and Martin Theus's Mondrian (Theus, 2002) (which implemented fully interactive parallel coordinate plots and integrated statistical tools from R into graphical displays).

There exist many other packages that offer varying degrees of interactivity for a varying range of graphics, some more effective, some less effective. There are apparently differing views about how interactive should interactive software be to deserve the name.

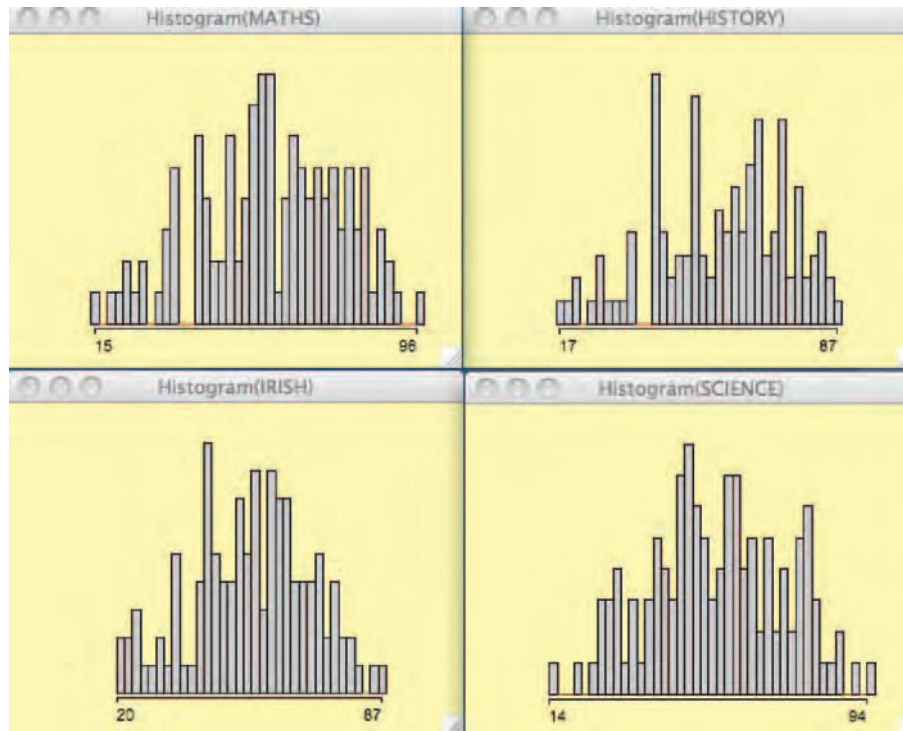


Figure 4 Histograms of exam marks for four subjects. All graphics are on scales of 0–100 with anchor points of 0 and 100. The gaps just under 40 in math and history are clearly visible.

Future Developments in EDA

Increasing computing power and more flexible user interfaces will add to EDA's potential and increase its use. It will be more frequently employed at all stages of a data analysis and not just in the initial stages. Closer integration of statistical modeling capabilities will blur the line between classical statistical analysis and EDA. This reflects what actually happens in practice, as distinct from what is described in theory.

See also: Analysis and Interpretation of Multivariate Data; Data Mining; Missing Data; Visual Data in Education Research.

Bibliography

- Beveridge, W. I. B. (1950). *The Art of Scientific Investigation*. London: Heinemann.
- Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science* **16**, 199–231.

- Cook, D. and Swayne, D. (2007). *Interactive and Dynamic Graphics for Data Analysis*. New York: Springer.
- Fleischer, W., Nagel, M., and Ostermann, R. (eds.) (1992). *Interaktive Datenanalyse with ISP*. Essen: Westarp Wissenschaften.
- Gelman, A. (2004). Exploratory data analysis for complex models. *Journal of Computational and Graphical Statistics* **13**, 755–779.
- Theus, M. (2002). Interactive data visualization using mondrian. *Journal of Statistical Software* **7**, 11.
- Theus, M. and Urbanek, S. (2008). *Interactive Graphics for Data Analysis*. Boca Raton, FL: CRC Press.
- Tukey, J. W. (1962). The future of data analysis. *Ann Math Stat* **33**, 1–67.
- Tukey, J. W. (1977). *Exploratory data analysis*. London: Addison-Wesley.
- Unwin, A. R., Hawkins, G., Hofmann, H., and Siegl, B. (1996). *Interactive graphics for data sets with missing values – MANET*. *Journal of Computational and Graphical Statistics* **5**(2), 113–122.
- Unwin, A. R., Theus, M., and Hofmann, H. (2006). *Graphics of Large Datasets*. New York: Springer.

Relevant Websites

- <http://www.bundeswahlleiter.de> – Federal Returning Officer, Federal Elections.
- <http://stats.math.uni-augsburg.de> – University of Augsburg, Department of Computer-Oriented Statistics and Data Analysis.

Factor Analysis: An Overview and Some Contemporary Advances

F A N Alhija, Tel Aviv University, Tel Aviv, Israel

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Common factor – It is a factor on which at least three observed variables substantially load.

Common factor analysis – It is a statistical technique which uses the correlation matrix between observed variables with communalities on the diagonal to estimate common factors and the loadings of the observed variables on the estimated factors.

Communality – It is the proportion of variance in the observed variable that is explained by a factor structure.

Confirmatory factor analysis – It is a factor analysis conducted for the aim of confirming a hypothesized factor structure.

Eigenvalue – It is the amount of variance in a set of observed variables explained by a common factor or component.

Exploratory factor analysis – It is a factor analysis used to explore the structure underlying a set of observed variables, usually, when there are no *a priori* hypotheses about the factor structure.

Factor analysis – It is a statistical technique used to reduce the dimensionality of a large number of observed variables to a fewer number of factors or components.

Factor rotation – It is a transformation of the principal factors or components in order to approach simple structure.

Loading – It is the correlation between a variable and a factor.

Model identification – An identified model is a model where a specific parameter value uniquely identifies the model, and no other equivalent formulation can be given by a different parameter value.

Model specification – It refers to the delineation of the causal relationships between variables that are thought to be possible (free to be estimated) and those relationships between variables that already have an estimated relationship, which can be gathered from previous studies (fixed).

Multilevel confirmatory factor analysis – It is a confirmatory factor analysis for testing measurement models in which the underlying attribute may vary as a function of levels of observations.

Multilevel data – It is hierarchically structured data whereby individuals are nested within groups. The number of levels can be two or more. Example: students within classes and classes within schools (three-level data).

Simple structure – The factor solution has a simple structure when each observed variable has substantial loading on one factor and each factor is defined by a set of observed variables that load substantially on it and have minor or zero loadings on the other factors.

Overview of Factor Analysis Methods

The term factor analysis refers to a family of statistical procedures used to either identify or validate some hypothesized structure among items/variables that tap a latent construct (Nunnally and Bernstein, 1994). There are two basic types of factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA).

Exploratory Factor Analysis

Typically, EFA is used to explore the underlying dimensions of a construct. The primary considerations inherent in the use of factor analysis include conceptual/theoretical considerations, design considerations, statistical considerations, and reporting considerations.

Conceptual/theoretical considerations

Factor analysts should choose an appropriate factor model, usually component analysis (CA) versus EFA, in accordance with the purpose of the analysis. If the purpose of factor analysis is essentially data reduction, then CA will yield a fewer number of components which represent the original set of variables. The resulting component scores are used in follow-up analyses.

If the goal of the researcher is to interpret the correlations among variables as arising from a smaller set of latent variables/factors, EFA is the method of choice. The latter model recognizes that variables are measured with error and yield coefficients which are less biased.

Design considerations

First, the theoretical and empirical aspects of the construct under inquiry must be thoroughly considered. Second, a sufficient number of items or test questions must be written or obtained to adequately measure the dimension(s) supported by the theory. Factor analysts have suggested a minimum of three to five items per factor. However, the breadth of the construct or its dimension definitions and the expected number of dimensions should be taken into account before deciding on the number of items to include. Third, sufficient sample size is necessary to stabilize estimates and to separate common variance from unique variance. The more exploratory the analysis, the larger the sample needs to be. A rule of thumb is that five to ten subjects per variable are required for a proper factor analysis. However, the number of variables/items, the potential number of factors, the variable to factors (v/f) ratio, the magnitude of factor loadings, and the size of the interfactor correlations need to be considered before an estimate of sample size can be suggested.

Other issues such as restriction in range, linear dependencies in the data, and variables that are not interval should be appropriately taken care of. The more restricted the range of the scores in the observed data, the more problematic the analysis will be. Restriction of range attenuates the intervariable correlations downward which in turn affects the magnitude of loadings, the amount of variance explained by each factor, and the ability to extract the correct number of factors.

Linear dependencies can easily occur when one is factor analyzing a large number of variables (say $v > 30$), due to the increased likelihood that two or more variables may be highly correlated. Linear dependency can result in a zero matrix determinant and an ill-conditioned matrix that cannot be inverted and therefore should be resolved prior to conducting the factor analysis.

The nature of the variables used in factor analysis is often an overlooked issue. Whether the variables are measured on a continuous, ordinal, or dichotomous scale should be treated appropriately. As the number of response categories is reduced, the likelihood increases that the Pearson correlation coefficients will be distorted and more relevant matrices of association should be used as an input.

Statistical considerations

Decisions regarding extraction method, number of factors, and the type of rotation should be made. The most commonly used extraction method within CA is principal component analysis (PCA) while principal axis factoring (PAF) is the commonly used extraction method with EFA. Components or factors are extracted from either the full correlation matrix (with unities on the main diagonal) or the reduced correlation matrix (with communalities – the proportion of each variable's variance explained by a factor

structure – on the main diagonal) one at a time, such that each factor extracts maximum variance from the variables and is independent of the previously extracted factors.

Numerous procedures for determining the number of factors exist. However, the most commonly used are the Cattell's scree plot and Kaiser's eigenvalue greater than one criterion, maybe because these are the default in major computer programs. Researchers are advised to consult several criteria when determining the number of factors. For a more comprehensive overview, see Benson and Nasser (1998).

Through rotation, a transformation of the principal factors or components in order to approach simple structure, the variance extracted is redistributed more evenly across the factors making factor interpretation easier. Two forms of rotation are applicable: orthogonal or oblique. Orthogonal rotation restricts the factors to be independent of one another; whereas oblique rotation allows the factors to be correlated.

The principle applied by most factor analysts to guide factor rotation is simple structure. That is for every pair of factors, there should be several variables whose loadings are high on one factor and low or near zero on the other factor. A loading of at least 0.30 is suggested to define a factor.

Varimax is the most commonly used option for orthogonal rotation. Another popular orthogonal rotation method is quartimax (Thompson, 2004). With orthogonal rotation, three matrices are provided: the unrotated factor matrix, the rotated factor matrix, and the transformation matrix. Usually, only the rotated factor matrix is interpreted.

The most frequently used oblique rotation method is direct oblimin (Benson and Nasser, 1998). Other popular methods for oblique rotation are promax and quartimin. With oblique rotation, four matrices are produced: the unrotated factor matrix, the factor pattern matrix, the factor structure matrix, and the factor correlation matrix. Typically, the factor pattern matrix and the factor correlation matrix are interpreted. A summary and comparison of many of the orthogonal and oblique rotation methods are provided in Gorsuch (1983).

Reporting consideration

Researchers should present sufficient theoretical background for readers to follow the decisions made regarding: the choice and the number of variables, sample size, factor model, input matrix, factor extraction, and factor rotation. The nature of EFA makes it typical to consider several models and rotated solutions in the process of arriving at the final model(s). The models considered could be derived from the literature which may suggest different numbers of factors or relations between factors. The analysis should be rerun after eliminating malfunctioning items. In such cases, eliminating items can result in a sharpening of the factor pattern/structure.

Confirmatory Factor Analysis

The technique of CFA analyzes *a priori* measurement models in which both the number of factors and their correspondence to the indicators are explicitly specified. CFA is then used to assess the fit of a hypothesized factor structure to the data.

Both EFA and CFA are part of the generalized linear model (GLM) and therefore, many of the concepts discussed with regard to EFA apply to CFA as well; however, the two methods differ with regard to several aspects which are summarized in **Table 1**.

Many decisions are inherent in the use of CFA. Some are made prior to the analysis and others are made following the analysis.

Pre-analysis decisions

Rival models

CFA involves testing the fit of models to the data. Other rival models should be tested in order to support the disconfirmability of the model. Disconfirmable and/or plausible models for CFA models are context specific, and some such models are either theoretically or empirically driven.

Table 1 Summary of the primary differences between EFA and CFA

EFA	CFA
The factor model is determined by the analysis	The factor model must be specified <i>a priori</i>
Rotation is performed to reach simple structure	Rotation is irrelevant because the <i>a priori</i> models typically specify simple structure
All parameters should be estimated	The researcher can fix certain parameters to mathematically permissible values, while freeing other model parameters to be estimated
Certain factor loadings may be expected, but cannot be incorporated into the analysis	One or more specific models, each containing some fixed and some freed parameters must be declared as input into the analysis
Error variances cannot be correlated	Error variances are allowed to correlate
Either all pairs of factors are assumed to be correlated or all factors are orthogonal	All pairs of factors can be correlated or only part of them can be correlated
Loading or inter-factor correlation equality constraints cannot be imposed	Equality constraints can be imposed on factor loadings, variances, and inter-factor correlations
Can be performed with SPSS, SAS, BMSBP, STATISTICA, and other statistical programs	Can be performed with LISREL, AMOS, CALIS, and MPLUS

Model identification

A model is said to be identified when, for a given research problem and data set, sufficient constraints are imposed such that a single set of parameter estimates is yielded by the analysis. A necessary, but not sufficient condition, for model identification is that the number of parameters being estimated should not be greater than the available degrees of freedom. The number of degrees of freedom in CFA equals the difference between the number of unique entries in the covariance matrix ($v(v+1)/2$), which is a function of the number of observed variables (v) in the analysis and the number of parameters (p) to be estimated ($df = v(v+1)/2 - p$). Model identification also requires that the measurement scale of each latent variable is specified or constrained in order to set the metric for the estimate.

CFA models can be identified by fixing one of the factor loadings to any number, usually 1. This indicates that the scales of the scores on the latent variable are some multiple of the selected observed variables. Alternatively, the factor variances can be constrained to be any mathematically positive number, usually 1.

Matrix of association to be analyzed

The default choice in most CFA programs is the covariance matrix. The computation of both the Pearson correlation matrix and the covariance matrix assumes data scaled in intervals. In CFA, for statistical reasons, it is preferable to analyze the covariance matrix which yields correct results as long as the data are appropriately scaled and distributed. Whenever there is some doubt regarding either the scaling of the data, or the selection of the input matrix of association, it is advisable practice to apply several reasonable choices such as using polyserial, tetra-choric, or polychoric correlations matrices as input with weighted least squares as an estimation method.

Multivariate normality

Some statistical estimation theories (e.g., maximum likelihood) assume multivariate normality, particularly regarding standard errors of model parameter estimates and some CFA model fit statistics. Two primary ways can be used to test the multivariate normality of the data. First, the graphical technique described by Thompson (1990) can be used with at least 25 observed variables. Second, a statistical test of the multivariate coefficient of kurtosis is available in CFA/SEM software.

If the assumption of multivariate normality is violated, one can use distribution-free parameter estimation theory (e.g., ADF). Nonetheless, this theory requires very large samples compared with the commonly used theories. An alternative is to use item parcels by bundling together observed variables in a way that reduces non-normality. However, this strategy is subject to the general limitations of parceling methods.

Parameter estimation theory

Two alternative parameter estimation theories are fairly commonly used in CFA. Maximum likelihood (ML) theory is the default theory in most statistical packages. ML requires that the data have at least an approximately multivariate normal distribution. Asymptotically distribution-free (ADF) estimation theory yields estimates that tend to be accurate regardless of the distribution shapes as the sample size becomes increasingly large.

Post-analysis decisions

Two major decisions should be made following the analysis: evaluation of model fit and consideration of specific errors in model specification.

Model fit statistics

Numerous indices of fit are provided in the literature and researchers are advised to consult several of them when evaluating model fit. Among the most commonly used are χ^2 goodness-of-fit statistics. A well-fitting model would be expected to have a non-significant χ^2 and a small χ^2/df ratio. Nonetheless, when the sample size is large, substantively correct models also yield significant values for χ^2 . Therefore other fit indices such as the comparative fit index (CFI), the standardized root mean square residual (SRMR), and the root mean square error of approximation (RMSEA) are commonly used. Recommendations for model fit cutoff criteria suggest that CFI should be at least .95, RMSEA < 0.08, SRMR < 0.06. A comprehensive discussion of goodness-of-fit indices is provided in Browne and Cudeck (1993).

Model misspecification statistics

Using irrelevant observed variables, omitting relevant observed variables, using wrong estimation theory, freeing parameters that should be fixed, or vice versa, are examples of model misspecifications. CFA programs do not tell researchers whether they used relevant observed variables but they yield diagnostics regarding the parameters that were freed but possibly should have been fixed and vice versa. In these cases researchers tend to re-specify the model and re-test the re-specified one. Model re-specification is reasonable when each change can be justified theoretically, and when the fit of the re-specified model can be evaluated in a holdout or independent sample.

Factor Analysis of Categorical Data

When developing test or questionnaire scales, the variables subjected to factor analysis are frequently not continuous. Rather, they are often dichotomous or polychotomous Likert scale items. The covariance matrix for these observed categorical variables will not fit the common

factor model, or factor analysis of observations on these variables will yield biased factor loadings. In these cases, besides the traditional parameter estimates, one must also estimate the so-called threshold parameters.

The methods that have been proposed for performing this estimation commonly obtain estimates of polychoric and polyserial correlations at one or more stages of computation. A theoretically appropriate method fits the CFA model to polychoric correlation using either weighted least squares (WLS) or robust WLS. This approach, however, assumes that a continuous, normal latent process determines each observed variable. This assumption cannot really be tested. Therefore some researchers suggest using composites of several items which are often called item parcels so that the resulting scores would be close enough to continuous variables.

Categorical data are likely non-normal; hence, the classic normal theory methods can give test statistics that are seriously biased. Therefore, the ADF approach is suggested as a possible solution. However, this free-distribution approach requires large samples. Another option is the use of rescaled statistics like the Satorra-Bentler scaled chi-square statistics.

Procedures for analyzing categorical variables are implemented in programs such LISCOMP LISTEL, PRELIS, and MPLUS.

Confirmatory Factor Analysis of Multilevel Data

When applied to hierarchical data, which often occurs in practice, CFA is frequently misused. Until recently, the conventional practice of assessing the factor structure of hierarchical data where individuals are nested within groups, such as students nested within classes or employees within organizations, involved either (1) performing the factor analysis at the individual level while ignoring the hierarchical structure of the data or (2) applying the factor analysis only to group-level data. Both approaches yield problematic results (Muthén, 1994) because a non-hierarchical model is inadequate.

In the first case, when the total sample covariance matrix (S_T) is factor analyzed, the assumption of independent and equally distributed observations that underlie factor analysis is violated with clustered observations that are likely to be related. Violation of this assumption biases parameter estimates, standard errors, and the power of statistical significance tests. This can result in either incorrect conclusions regarding the factor structure of the scale under investigation or the appropriateness of the items for measuring the latent construct.

In the second case, the sample between-group covariance matrix (S_B) will be a biased estimate of the population between-group covariance matrix (\sum_B) because it is a function of the population between-covariance matrix

(\sum_B) and the weighted group-size within-group covariance matrix (\sum_W) (Muthén, 1994).

The use of multilevel confirmatory factor analysis (MCFA) is also justified when constructing aggregate-level and multilevel theories are of interest. These theories typically include constructs that may function differently at aggregate levels of analysis (e.g., Heck and Thomas, 2000). Therefore, it is imperative to make sure that a scale measuring a group-level construct exhibits the desired construct validity properties at the aggregate level of analysis.

Currently, it is possible to assess the similarity of the factor-loading patterns across levels of analysis by reformulating the testing of separate covariance matrices across data levels as a test of covariance matrices across different groups. In the case of two levels, one model is formulated for the individual-level variation and another is formulated for the between-group variation in the parameters and the features of a within-group and between-group measurement model are tested simultaneously.

Muthén (1994) demonstrated that current multiple-group SEM software packages – such as LISREL, EQS, or AMOS – can be modified to handle multilevel factor analysis. One available software program that facilitates the specification and testing of two-level structured model is STREAMS that allows relatively easy specification of a variety of structural models. Muthén's Mplus software program specifies multilevel SEM models within which MCFA is executed and creates the necessary within- and between-group matrices in unbalanced designs. EQS version 6.1 features two different methods for performing multilevel SEM analyses.

The MCFA estimation process

Muthén (1994) proposed a five-step procedure to follow when conducting MCFA:

1. conventional CFA on the sample total covariance matrix (S_T);
2. establishment of between-group level variation;
3. confirmatory factor analysis of the sample-pooled covariance matrix (S_{PW});
4. confirmatory factor analysis of the sample between-group covariance matrix (S_B); and
5. multilevel confirmatory factor analysis (MCFA).

The first four steps produce basic information regarding the factor structure of the scale data at different levels of analysis along with information necessary to justify the multilevel analysis. The final step consists of the actual MCFA, and therefore provides the most direct evidence for confirming or refuting the hypothesized multilevel model. Each of the five steps is elaborated through an empirical illustration of the MCFA procedure.

An Illustrative Application of MCFA

Method

Data source and sample

Data used for this illustration were obtained from 3117 students enrolled in 272 courses at a major teacher training college in Israel. The sample size of students from each course varied from 6 to 64, with an average sample size of 16.14. Student ratings of one randomly selected course taught by each instructor constituted the data.

The instrument

Student ratings were collected using an evaluation questionnaire with 13 statements measuring teaching quality. Students responded to each item using a seven-point Likert-type scale designed to assess three dimensions of teaching quality: course content (five items), instructor–student relations (four items), and planning and teaching strategy (four items). Reliability coefficients (Cronbach's alpha) corresponding to the three factors were estimated to be 0.87, 0.86, and 0.84, respectively.

The hypothesized factor model

In earlier research, exploratory factor analysis of the data collected with the same questionnaire yielded three factors as described above (Fresko *et al.*, 2003). Since students are nested within courses, their ratings of college teaching are likely to be nonindependent. Therefore, a three-factor two-level model was formulated and the research hypothesis was that the model should fit the data at both levels (see Figure 1).

Analysis and Results

Preliminary analysis

The individual responses to the 13 items were screened to detect substantial skewness, kurtosis, and/or outliers. Values were acceptable for all 13 items. The total (S_T), the pooled within (S_{PW}), and the sample estimated between-group (S_B) covariance matrices were calculated.

Model estimation and assessment of fit

In this illustration, all the analyses were conducted using EQS version 6.1 with the MUML estimator that produces maximum likelihood parameter estimates, robust standard errors, and a mean-adjusted χ^2 goodness-of-fit statistics for equal group size. For unbalanced data, MUML is a robust limited-information estimator. No specific criteria for evaluating model fit have yet been offered for multilevel modeling. Researchers, therefore apply the same criteria used in the context of SEM as they do to MCFA modeling.

The five-step results are as follows:

Step 1: Conventional CFA on the S_T matrix. The parameter estimates and the fit statistics resulting from this analysis

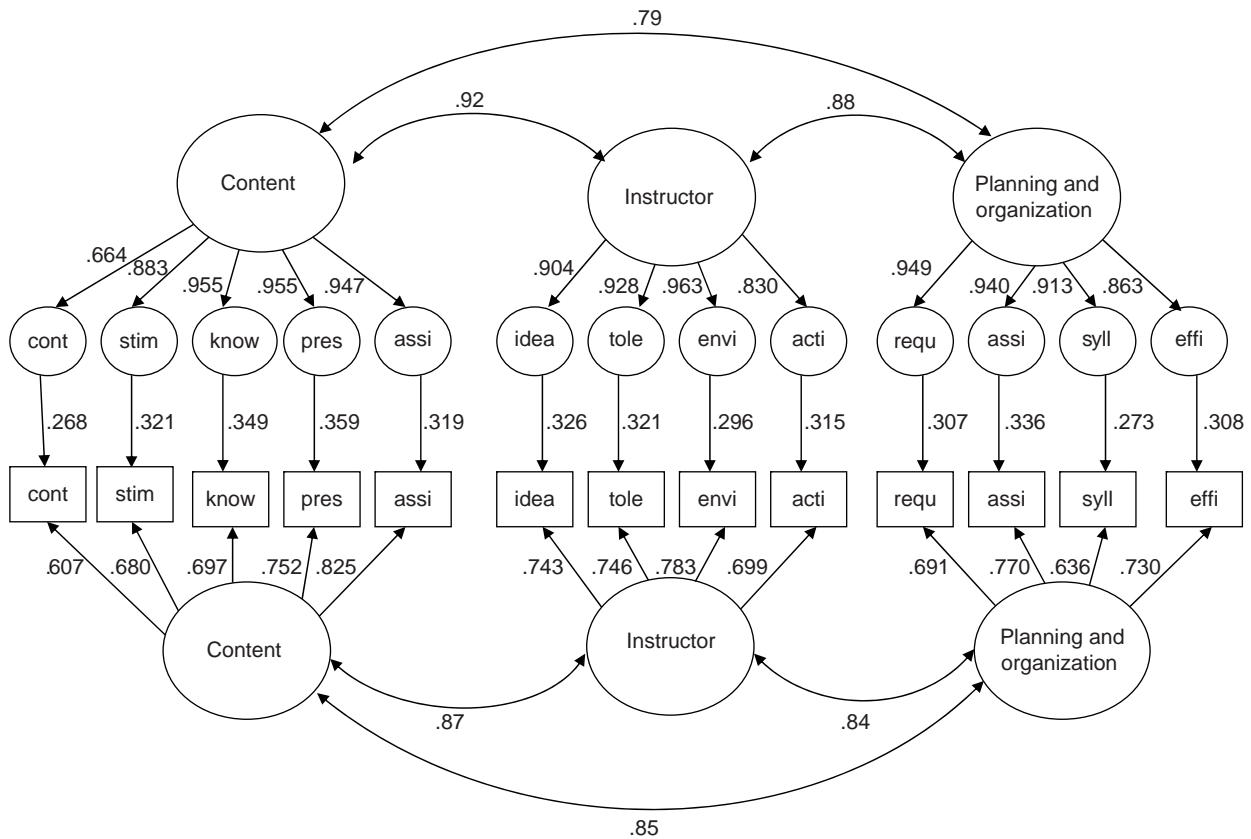


Figure 1 The sample three-factor multilevel model. assi = assignments' contribute to understanding, idea = encourage ideas exchange, tole = instructor is tolerant, envi = creates pleasant environment, acti = activate students, requ = clear requirements, assi = clear assignments, syll = follows the syllabus, effi = use class time efficiently. Errors can be calculated as: $e = (1 - \text{loadings}^2)^{1/2}$.

Table 2 Model fit for *a priori* one- and two-level models^a

Models	χ^2	df	CFI	RAMSEA (C.I.)	SRMR
Total	2046.47	62	.93	.101 (.098, .105)	.043
Within	1525.85	62	.93	.087 (.083, .91)	.042
Between	6907.30	62	.84	.197 (.193, .201)	.059
Multilevel (with all equality constraints)	1977.64	124	.94	.098 (.094, .102)	.049

^aAll χ^2 values are statistically significant ($p > .05$).

df – degrees of freedom; CFI – comparative fit index; RAMSEA – root mean square error of approximation; SRMR – standardized root mean square residual.

are biased when the degree of nonindependence is sizable, particularly when group size is large (Muthén, 1994) or when the factor structure at the within- and between-group levels differs.

The fit of the three-factor model was borderline with significant χ^2 [$\chi^2(62) = 2046.47$, $df = 62$, $p > .05$], and values of .93, .043, and .101 for CFI, SRMR, and RMSEA, respectively. Standardized loading for the 13 items ranged from .624 to .868, with all values statistically significant, suggesting that the items adequately represent the latent constructs.

Step 2: Establishing between-group (course) level variation. Application of MCFA to the data at hand is justified if the systematic between-group variation for each observed variable is relatively large, indicating the multilevel nature of the data. Several techniques have been designed to detect the multilevel character of the data including r_{wg} , η^2 , WABA, and the intra-class correlation (ICC) (James *et al.*, 1984; 1993; Dansereau *et al.*, 1984; Muthén, 1994; Kowzlowzki and Klein, 2000). Muthén's ICC was used in this illustration to quantify the proportion of the between-group variation relative to the total variation in the data.

Table 3 Standardized factor loadings and standard errors for the total sample, the within-group, the between-group, and the two-level models

Variable							Multilevel					ICC
	S_T		S_{pw}		S_B		Within		Between			
	Loading	SE	Loading	SE	Loading	SE	Loading	SE	Loading	SE		
<i>Course content</i>												
This course can contribute to my professional future ^a	0.624		0.605		0.674		0.607		0.664		0.274	
Course content is stimulating	0.733	0.035	0.680	0.038	0.855	0.030	0.608	0.040	0.883	0.124	0.235	
In this course, I am given the opportunity to demonstrate my knowledge and competence	0.749	0.033	0.695	0.036	0.874	0.027	0.697	0.038	0.917	0.110	0.215	
The content of the course is presented in an interesting way	0.801	0.033	0.751	0.036	0.908	0.027	0.753	0.037	0.955	0.113	0.230	
Course assignments contribute to my understanding of the course materials	0.868	0.036	0.826	0.038	0.924	0.035	0.825	0.039	0.947	0.146	0.332	
<i>Student–instructor relation</i>												
Instructor encourages the exchange of ideas in class ^a	0.788		0.742		0.880		0.743		0.904		0.273	
The instructor is tolerant of students’ views that differ from his/her own	0.799	0.022	0.746	0.026	0.886	0.015	0.746	0.027	0.928	0.058	0.278	
The instructor creates a pleasant learning environment	0.845	0.022	0.784	0.026	0.939	0.016	0.783	0.027	0.963	0.064	0.243	
This course provides opportunities for students to be active	0.737	0.024	0.697	0.028	0.776	0.019	0.699	0.030	0.830	0.067	0.263	
<i>Planning and organization</i>												
The instructor defines course requirements clearly ^a	0.766		0.692		0.898		0.691		0.949		0.261	
The instructor provides clear instructions for all assignments	0.822	0.025	0.769	0.032	0.898	0.015	0.770	0.034	0.940	0.053	0.273	
The instructor teaches the course according to the syllabus	0.720	0.022	0.636	0.027	0.864	0.013	0.636	0.028	0.913	0.050	0.280	
Class time is used efficiently	0.770	0.027	0.730	0.034	0.831	0.018	0.730	0.035	0.863	0.065	0.285	

^aLoading fixed to zero to set the metric

The values of ICC range from 0 to 1, with higher values indicating a greater proportion of between-group variance and potential bias if the multilevel nature of the data is ignored. The values for the 13 observed indicators ranged from .230 to .343 (**Table 2**). The ICC values indicated that the between-course variation is relatively sizable, justifying the conduct of an MCFA ($ICC \geq .05$ justifies multilevel analysis).

Step 3: CFA of the S_{PW} matrix. The S_{PW} based on individual-level scores and adjusted for respective group means is used for the analysis of the disaggregate factor structure.

For this illustration, S_{PW} was calculated using EQS version 6.1. The number of observations for the pooled within-group covariance matrix (N-G) was $3117 - 272 = 2845$, whereas for the course level it (G) was 272. The three-factor model based on S_{PW} produced a better fit to the data than the one based on S_T ($\chi^2 = 1525.852$, $df = 62$, $p < .001$, CFA = .93, SRMR = .042, and RMSEA = .088). This indicates that either the construct-relevant variance is substantial at the student level or that the factor structure for the course level differs from that of the student level.

Step 4: CFA of the S_B matrix. The structure obtained for the within-group level may not hold for between-group level data. The appropriateness of the between-group factor structure is therefore examined. Either an estimate of the population between-group covariance matrix or the sample between-group covariance matrix S_B adjusted for the weighted group-size within-group effect (Muthén, 1994) can be used.

The estimated between-group covariance matrix obtained by EQS version 6.1 was used here. The fit results (**Table 2**) reveal that the course level model shows considerably poorer fit to the data. In particular, the RMSEA (.197) and the CFI (.84) reflect poor fit. Alternative two-factor and one-factor models produced an even worse fit. Therefore, for the sake of illustration, it was decided to proceed with the multilevel step with the three-factor model at the student and the course levels.

Step 5: MCFA. The fit results pertaining to the MCFA with all equality constraints across the two levels indicated acceptable fit [$\chi^2 = 1977.640$, $df = 124$, $p < .001$, CFA = .94, SRMR = .049, and RMSEA = .098(.094, 102)]. **Table 2** summarizes the fit data for the conventional CFA model, the within- and between-course models and the two-level model. **Table 3** summarizes the estimated factor loading values and the standard errors for the 13 items in the total, student-level, course-level, and multilevel models together with their ICC values. **Figure 1** depicts the two-level factor model of the student ratings data along with the student-level and course-level factor loadings and inter-factor correlations. The factor loadings at the course-level were stronger than those at the student-level. While the factor loadings obtained from the student-level and the course-level separately were similar in value to the corresponding estimates obtained from the multilevel analysis, the standard errors of these estimates

were larger than those associated with the student and course-level models. This indicates that taking the multilevel structure of the data into account results in different standard errors from those obtained in the disaggregated and aggregated data models separately. Although the fit of the between-course model was poor when analyzed separately, simultaneous analysis of the two-level data produced a better fit.

Summary and Implications

Factor analysis is one of the most commonly used methods in educational and psychological research. Still, it is frequently misused and critical issues are overlooked. Researchers employing factor analysis should carefully review the theoretical and empirical literature in the substantive areas of interest, select a sufficient number of variables that are as representative as possible of the content domain, be aware of the metric used to measure variables, avoid including linearly dependent variables, select a sufficient sample size which is as representative or as random as possible, and use theory to guide the factor analytic decisions which need to be made in the analysis.

Categorical and hierarchical data require special attention and careful selection of appropriate procedures that take the special characteristics of these data into account and therefore yield more accurate factor model estimates. With categorical data, factor analysts should use the appropriate matrix of association as an input for the analysis (e.g., polychoric and polyserial correlation matrices), and consider distribution-free estimation theory or item parcels to avoid inaccuracies resulting from a violation of the normality assumption.

As illustrated in this article, MCFA allows the researcher to investigate the stability of a proposed factor model across organizations such as schools, colleges, universities, and hospitals. Through this approach, the researcher can develop a more refined analysis of construct validity and obtain more accurate estimates at the various levels of analysis. It is hoped that the MCFA framework will provide researchers with an analytic tool that is both powerful and easy to comprehend in order to facilitate addressing the inherent nature of the structure of the data at hand.

Bibliography

- Benson, J. and Nasser, F. (1998). On the use of factor analysis as a research tool. *Journal of Vocational Education Research* **23**(1), 13-33.
- Browne, M. W. and Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K. A. and Long, S. J. (eds.) *Testing Structural Equation Models*, pp 136-162. Thousand Oaks, CA: Sage.

- Fresko, B., Ashknazi, T., and Nasser, F. (2003). *Feedback on Instruction: The Relationship between Student Verbal Comments and Student Ratings, Student, Instructor and Course Characteristics*. Kfar Saba: Beit Berl College and Mofet Institute. (Hebrew).
- Gorsuch, R. L. (1983). *Factor Analysis*, 2nd edn. Hillsdale, NJ: Erlbaum.
- Heck, H. R. and Thomas, L. S. (2000). *An Introduction to Multilevel Modeling Techniques*. Mahwah, NJ: Erlbaum.
- Kozlowski, S. W. J. and Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In Klein, K. J. and Kozlowski, S. W. J. (eds.) *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions*, pp 3–90. San Francisco, CA: Jossey-Bass.
- Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research* **22**, 376–398.
- Nunnally, J. C. and Bernstein, I. H. (1994). *Psychometric Theory*. New York: McGraw-Hill.
- Thompson, B. (1990). MULTINOR: A FORTRAN program that assists in evaluating multivariate normality. *Educational and Psychological Measurement* **50**, 845–848.
- Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*. Washington, DC: American Psychological Association.

Further Reading

- Bentler, P. M. (2003). *EQS 6 Structural Equations Program Manual*. Encino, CA: Multivariate Software.
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: Guilford.
- Dyer, N. G., Hanges, P. J., and Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *Leadership Quarterly* **16**(1), 149–167.
- Goldstein, H. and McDonald, R. (1988). A general model for the analysis of multilevel data. *Psychometrika* **53**, 455–467.
- Flora, D. B. and Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods* **9**(4), 466–491.
- Gustafsson, J. E. and Stahl, P. A. (2000). *STREAMS User's Guide, Version 2.5 for Windows*. Mölndal, Sweden: Multivariate Ware.

- Kenny, D. A. and Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin* **99**, 422–431.
- Kim, J.-O. and Mueller, C. W. (1978). *Introduction to Factor Analysis, What It Is and How to Do It, Series No13: Quantitative Applications in the Social Science*. Newbury Park, CA: Sage.
- Longford, N. T. and Muthén, B. O. (1992). Factor analysis for clustered observations. *Psychometrika* **57**, 581–597.
- Muthén, B. O. and Muthén, L. K. (1998). *Mplus: User's Guide*. Los Angeles: Muthén and Muthén.
- Pedhazur, E. and Pedhazur-Schmelkin, L. (1991). *Measurement, Design, and Analysis. An Integrated Approach*. Hillsdale, NJ: Erlbaum.

Relevant Websites

- <http://bama.ua.edu> – Archives of Structural Equation Modeling Discussion Group.
- <http://www.bls.gov> – Bureau of Labor Statistics. A detailed analysis of the factors entering the projection process covers the period 2006–2016.
- <http://www.let.rug.nl> – Exploratory Factor Analysis, Theory and Application.
- <http://psico.fcep.urv.es> – Manual of the program factor Windows 95/98/NT/2000/XP. Factor is a program developed to fit the Exploratory Factor Analysis model.
- <http://www.statmodel.com> – Mplus is a statistical modeling program that provides researchers with a flexible tool to analyze their data.
- <http://www.mvsoft.com> – Multivariate Software. EQS – Structural Equation Modeling Software.
- <http://www.mwstreams.com> – Multivariate Ware. STREAMS (Structural Equation Modeling Made Simple) has been designed with two main purposes in mind: to be a tool for teaching and learning structural equation modeling (SEM); and to be a productivity tool for modelers.
- <http://www2.gsu.edu> – SEMNET – The Structural Equation Modeling Discussion Network.
- <http://preprints.stat.ucla.edu> – Software for multilevel analysis.
- <http://davidakenny.net> – Structural Equation Modeling (SEM). This page serves as a gate to a tutorial on structural equation modeling or SEM.
- <http://www.sussex.ac.uk> – University of Sussex website.

Generalized Linear Mixed Models

S Rabe-Hesketh, University of California, Berkeley, CA, USA; University of London, London, UK
A Skrandal, Norwegian Institute of Public Health and University of Oslo, Oslo, Norway

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Linear predictor – A linear combination of explanatory variables that is part of a regression model or generalized linear mixed model.

Link function – A function applied to the conditional expectation of the response variable before this is equated to the linear predictor (in a generalized linear model). Examples are the identity, log, and logit link functions.

Random coefficient – Cluster-specific coefficient of a covariate in a regression model for clustered data that varies randomly between clusters.

Random intercept – Cluster-specific intercept in a regression model for clustered data that varies randomly between clusters.

Logistic Random-Intercept Model

To introduce the idea of generalized linear mixed models, we consider the following example. The Program for International Student Assessment (PISA) is an international educational survey funded by the Organisation for Economic Co-operation and Development (OECD) that measures attainment in reading, mathematics, and science among 15-year-old students. Using the United States sample of PISA 2000, we estimate the relationship between reading proficiency (Prof_{ij}) of student i in school j and the following covariates:

- Fem_{ij} : indicator for student being female;
- SES_{ij} : international socioeconomic index (continuous), grand-mean centered and divided by 20;
- HS_{ij} : indicator for highest education level of either parent being high school graduate;
- Coll_{ij} : indicator for highest education level of either parent being college graduate; and
- Eng_{ij} : indicator for test language (English) spoken at home.

Prof_{ij} is a binary variable (1 = Yes, 0 = No); therefore, an appropriate level-1 model is a logistic regression model. Specifically, the log of the odds that the student is proficient versus not proficient is specified as a linear function of the covariates,

$$\log \left\{ \frac{P(\text{Prof}_{ij} = 1)}{P(\text{Prof}_{ij} = 0)} \right\} = \beta_{0j} + \beta_1 \text{Fem}_{ij} + \beta_2 \text{SES}_{ij} + \beta_3 \text{HS}_{ij} + \beta_4 \text{Coll}_{ij} + \beta_5 \text{Eng}_{ij}$$

(Here $P(\cdot)$ refers to the conditional probability, given β_{0j} and the covariates.) The school-specific intercept β_{0j} represents the log-odds of being proficient for students in school j when all the covariates take the value zero, that is for males with average socioeconomic status (SES) whose parents' highest education level is less than high school and who do not speak English at home. The regression coefficient of a covariate represents the increase in log-odds per unit increase in the corresponding covariate when the other covariates and β_{0j} remain constant. These coefficients are sometimes referred to as conditional or school-specific effects because β_{0j} is held constant. Exponentiating the coefficients yields school-specific odds ratios.

The level-2 model for school j is

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \overline{\text{SES}}_j + \zeta_{0j}$$

where $\overline{\text{SES}}_j$ is the mean SES for school j , and ζ_{0j} is a school-specific random intercept assumed to be independent of the covariates and independent across schools with $\zeta_{0j} \sim N(0, \psi)$.

Substituting the school-level model into the student-level model and writing η_{ij} for the log-odds, we obtain the so-called reduced form

$$\eta_{ij} = \gamma_{00} + \gamma_{01} \overline{\text{SES}}_j + \beta_1 \text{Fem}_{ij} + \beta_2 \text{SES}_{ij} + \beta_3 \text{HS}_{ij} + \beta_4 \text{Coll}_{ij} + \beta_5 \text{Eng}_{ij} + \zeta_{0j}$$

a logistic regression model with both student-level (or level 1) and school-level (or level 2) covariates and both a fixed intercept γ_{00} and a random intercept ζ_{0j} .

Maximum likelihood estimates for the logistic random intercept model are presented in **Table 1**. Higher school mean SES, being female, higher individual SES, having at least one parent with a college education, and speaking English at home are all associated with a significantly greater odds of reading proficiency at the 5% level. For an indicator variable, the odds ratio represents the ratio of the odds for two groups, holding constant the other covariates. For instance, the estimated odds ratio of 1.9 for Eng_{ij} means that students who speak English at home have 1.9 times the odds of being proficient in comparison to students who do not speak English at home for a given

school mean SES, gender, parents' education, and random intercept ζ_{0j} .

For a continuous covariate, the odds ratio can be interpreted as the effect of increasing the covariate by a unit; therefore, it is important to understand the scale of the covariate. Individual SES has a standard deviation of 0.88 and school mean SES has a standard deviation of 0.45. Each unit increase in school mean SES is associated with a quadrupling in the estimated odds of being proficient, controlling for individual SES, the other covariates, and the random intercept. Such an additional effect of the school mean of a variable after controlling for the individual-level variable is sometimes referred to as a contextual effect. Here it could be due to the effect of the peers' SES, as well as any omitted variables that are correlated with school SES, such as school resources, parent involvement, and teacher qualifications.

If we had omitted school mean SES from the model, the estimated coefficient of individual SES would have been greater (0.41) due to absorbing some of the contextual effect. Using econometric terminology, we would have had an endogeneity problem because the random intercept, representing the combined effects of all omitted covariates at the cluster level (including school mean SES), would be correlated with individual SES. Another way of thinking about the problem is by remembering that the random intercept is assumed to be uncorrelated with the covariates; therefore, the effects of included covariates are not controlled for any possible school-level confounders. Such control can be achieved for an individual level-1 covariate by mean centering it and/or including its cluster mean as an additional covariate. Alternatively, we can eliminate cluster-level confounding for all level-1 covariates by specifying fixed effects for schools. In linear models, this could be accomplished by including indicator variables for schools. In logistic regression models, an appropriate approach is conditional maximum likelihood estimation. Using this method gives similar estimated coefficients except for a lower estimate of 0.46 for Eng_{ij} , suggesting that there might be an endogeneity problem or contextual effect for this variable.

It is useful to visualize the magnitude of the effects of some variables on the probability of reading proficiency,

Table 1 Maximum likelihood estimates for logistic random intercept model

Parameter	Est	(SE)	OR	(95% CI)
γ_{00}	-2.02	(0.29)		
$\gamma_{01}[\text{SES}_j]$	1.38	(0.18)	4.0	(2.8,5.7)
$\beta_1[\text{Fem}_{ij}]$	0.56	(0.10)	1.7	(1.4,2.1)
$\beta_2[\text{SES}_{ij}]$	0.29	(0.07)	1.3	(1.2,1.5)
$\beta_3[\text{HS}_{ij}]$	0.39	(0.25)	1.5	(0.9,2.4)
$\beta_4[\text{Coll}_{ij}]$	0.71	(0.23)	2.0	(1.2,3.3)
$\beta_5[\text{Eng}_{ij}]$	0.62	(0.29)	1.9	(1.2,2.9)
ψ	0.27	(0.09)		

holding other variables constant. For instance, we can consider boys whose parents' highest level of education is high school and who have SES equal to the overall mean, and obtain predicted probabilities as a function of school mean SES and the indicator variable for speaking English at home. A graph of these predicted probabilities (for the range of school mean SES in the data) is shown in **Figure 1**, where the dashed curves are for students who speak English at home and the solid curves are for students who do not speak English at home. For each line pattern, the curve that is closer to 0.5 represents the population averaged or marginal probability, after integrating or averaging out the random intercept, whereas the other curve represents the conditional or school-specific probability, with the random intercept set to 0, which is also the median probability.

We can see that school mean SES has a large effect, with predicted probabilities differing by more than 0.4 between the lowest and highest SES schools. Speaking English at home increases the probability by about 0.1 for average and above average SES schools. In the very low SES schools, boys with average SES, but whose parents' highest level of education is high school, have very low probabilities of reading proficiency.

Random Coefficient Models

The random intercept model for the log-odds considered in the previous section had the form

$$\eta_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \cdots + \beta_p x_{pij}$$

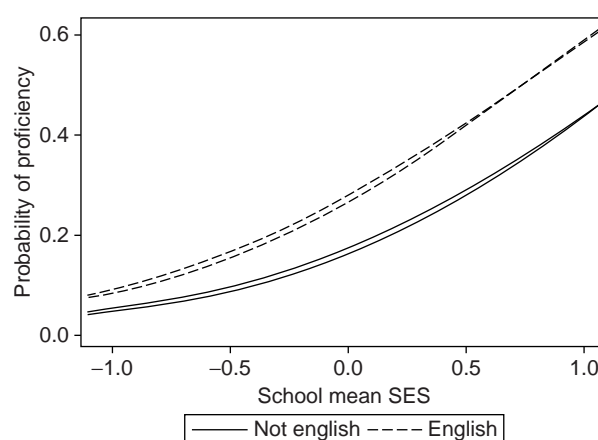


Figure 1 Predicted probability of reading proficiency as a function of school mean SES for students who do (dashed curves) and do not (solid curves) speak English at home (with $\text{Fem}_{ij} = 0$, $\text{SES}_{ij} = 0$, $\text{HS}_{ij} = 1$, and $\text{Coll}_{ij} = 0$). For each line pattern, the curves closer to 0.5 are marginal probabilities and the other curves, median probabilities. Reproduced by permission of the publisher from Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata*, 2nd edn, Figure 6.3. College Station, TX: Stata Press. Copyright 2008 by StataCorp LP.

where β_{0j} is a school-specific random intercept and x_{pij} ($p = 1, \dots, P$) are level-1 covariates with fixed regression coefficients β_p .

In random coefficient models, not only is the intercept school specific, but the effects of at least some level-1 covariates also vary between schools. The simplest example is a model with a random intercept and a random slope for a covariate x_{1ij} . The level-1 model for the log-odds then has the form

$$\eta_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_2x_{2ij} + \dots + \beta_px_{pij}$$

where x_{1ij} has a school-specific regression coefficient or slope β_{1j} , whereas the other level-1 covariates have fixed regression coefficients β_2, \dots, β_p .

The school-specific intercept and slope each have their own level-2 model of the form

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}w_{1j} + \dots + \gamma_{0Q}w_{Qj} + \zeta_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}w_{1j} + \dots + \gamma_{1Q}w_{Qj} + \zeta_{1j}\end{aligned}$$

where w_{1j}, \dots, w_{Qj} are school-level covariates. The school-specific random intercept ζ_{0j} and random slope ζ_{1j} are typically assumed to have a bivariate normal distribution with zero as mean for given covariate values.

Substituting the level-2 models into the level-1 model produces a reduced form with cross-level interaction terms. For instance, the model for β_{1j} includes the term $\gamma_{11}w_{1j}$, and, since β_{1j} multiplies x_{1ij} , we obtain the cross-level interaction $\gamma_{11}w_{1j}x_{1ij}$. Random coefficient models could, of course, also include random effects of more than one covariate varying at level-1.

Different Response Types

All generalized linear mixed models have a so-called linear predictor η_{ij} of the form shown in the previous section. The relationship between the linear predictor and the observed response y_{ij} can be specified in a number of different ways depending on the type of response variable. First, a conditional distribution for y_{ij} is specified as a function of the conditional expectation μ_{ij} , given the covariates and random effects. Second, a link function $g(\cdot)$ is specified so that

$$g(\mu_{ij}) = \eta_{ij}$$

The conditional response distribution is from the exponential family and is characterized by the conditional expectation μ_{ij} as well as a dispersion parameter ϕ that affects the conditional variance

$$\text{Var}(y_{ij}|\mu_{ij}) = \phi V(\mu_{ij})$$

where the variance function $V(\mu_{ij})$ is determined by the chosen distribution.

We briefly discuss the most common generalized linear mixed models for continuous responses, binary responses, and counts.

Continuous Responses

In the continuous case, a linear mixed model is typically assumed, in which the conditional response distribution is normal,

$$y_{ij}|\mu_{ij} \sim N(\mu_{ij}, \sigma_e^2)$$

and the link function is the identity link

$$\mu_{ij} = \eta_{ij}$$

The variance function is $V(\mu_{ij}) = 1$ and the dispersion parameter is $\phi = \sigma^2$.

Binary Responses

Binary responses are assumed to be independently Bernoulli distributed as

$$y_{ij}|\mu_{ij} \sim \text{Bernoulli}(\mu_{ij})$$

for given μ_{ij} . Here the conditional expectation μ_{ij} is also the conditional probability $P(y_{ij} = 1)$ for given values of the covariates and random effects. The most common link function is the logit link

$$\eta_{ij} = \text{logit}(\mu_{ij}) \equiv \log\left\{\frac{\mu_{ij}}{1 - \mu_{ij}}\right\} = \log\left\{\frac{P(y_{ij} = 1)}{P(y_{ij} = 0)}\right\}$$

which can be interpreted as the log of the odds that y_{ij} is 1. An alternative link function is the probit link $\Phi^{-1}(\mu_{ij})$, the inverse standard normal cumulative distribution function. The variance function in either case is

$$V(\mu_{ij}) = \mu_{ij}(1 - \mu_{ij})$$

and the dispersion parameter is equal to 1.

If there are several ($n_{ij} > 1$) independent binary responses per unit ij , the conditional distribution of the number of responses that are 1 is binomial with probability of success μ_{ij} and number of trials n_{ij} . In this case, the variance function is $n_{ij}\mu_{ij}(1 - \mu_{ij})$ and $\phi = 1$. However, if $n_{ij} > 1$, it is possible that the empirical variances are larger or smaller than the model-implied variances. Such overdispersion or underdispersion can be accommodated by estimating ϕ as a free parameter in a so-called quasi-likelihood approach. In likelihood or Bayesian methods, overdispersion is instead induced by including additional random effects varying over level-1 units. Note that there is no such thing as overdispersion or underdispersion if $n_{ij} = 1$, although researchers sometimes attempt to model it.

Latent response formulation

In econometrics and psychometrics, models for binary responses are often specified by imaging an underlying or latent continuous response y_{ij}^* such that the observed response y_{ij} is 1 if the latent response exceeds 0 and y_{ij} is 0 otherwise. An obvious interpretation of y_{ij}^* in the proficiency example is as a continuous measure of reading achievement that must exceed a threshold for the student to be proficient. A linear mixed model is then specified for the latent response

$$y_{ij}^* = \eta_{ij} + e_{ij}$$

If a logistic distribution is assumed for e_{ij} we obtain a logistic regression model, and, if a standard normal distribution is assumed, we obtain a probit regression model.

A simple latent response model with $\eta_{ij} = \beta_0 + \beta_1 x_{ij}$ is shown in the lower portion of **Figure 2**. The density curves represent the distribution of y_{ij}^* for different values of x_{ij} , with means falling on the upward sloping regression line. The threshold 0 is indicated by the dashed horizontal line, and the areas under the density curves exceeding this threshold are shaded. These probabilities that y_{ij}^* is greater than 0 represent the corresponding probabilities that $y_{ij} = 1$. The upper portion of the figure is a graph of these probabilities as a function of x_{ij} and we can recognize the familiar logistic curve. The logistic regression model can thus either be specified through a logit link or using a latent response formulation.

The latent response formulation can be useful for interpreting the random part of the model. For a random intercept model, we can express the within-cluster dependence as the residual intraclass correlation among the latent responses, given the covariates

$$\text{cor}(y_{ij}^*, y_{ij}^* | \mu_{ij}, \mu_{ij}) = \frac{\psi}{\psi + \text{Var}(e_{ij})}$$

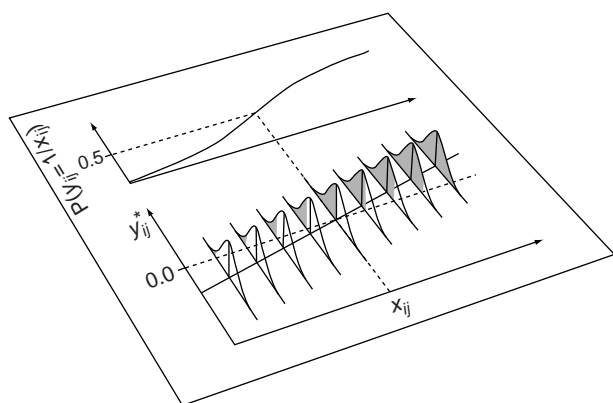


Figure 2 Illustration of the relationship between latent-response model and probability of observed response being 1. Reproduced by permission of the publisher from Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata*, 2nd edn. College Station, TX: Stata Press. Copyright 2008 by StataCorp LP.

where $\text{Var}(e_{ij})$ is $\pi^2/3$ in logistic models and 1 in probit models. For the estimates in **Table 1**, the residual intraclass correlation of the latent responses is estimated as 0.08. Hence, 8% of the residual variance in underlying reading achievement is due to schools. Although this intraclass correlation relies on the concept of a latent response, it is preferable to the correlation among the observed responses which depends on the covariate values.

Counts

Sometimes a response variable is a count of some event such as the number of days a student is absent from school in a year or the number of times a teacher shouts in an hour. The conditional response distribution is typically specified as Poisson,

$$y_{ij} | \mu_{ij} \sim \text{Poisson}(\mu_{ij})$$

and a log link is used

$$\log(\mu_{ij}) = \eta_{ij}$$

The log link ensures that the expected count is nonnegative and produces a multiplicative model for the expected count.

The variance function is $V(\mu_{ij}) = \mu_{ij}$ and $\phi = 1$. Overdispersion or underdispersion can be accommodated by estimating ϕ as a free parameter, and overdispersion can be modeled by including additional random effects varying over level-1 units. The Poisson model cannot be specified through a latent response formulation.

Other Response Types

Ordinal responses include Likert scales for agreement with attitude statements (e.g., disagree, neither agree nor disagree, and agree) and reported frequencies of doing something such as helping children with homework (e.g., daily, several times per week, occasionally, and never). The most common models are so-called cumulative logit or probit models, and these can be specified as logit or probit models for the probabilities of exceeding each of the ordered categories (except the last). A parallel regression assumption is usually made by allowing only the intercept to take different values for different categories. The models can alternatively be specified using a latent response formulation with several, freely estimated thresholds.

Discrete time durations are often of interest in education. Examples include number of semesters before dropout from college, age at which children first enter preschool, and number of years teachers teach at a school. A popular model for analyzing discrete time durations is the continuation ratio logit model. After expanding the data appropriately, this model can be estimated using logistic regression for binary responses.

Unordered categorical data arise when people choose among different alternatives, such as community college, 4-year college, or no higher education after graduation from high school. Multinomial logit models are typically used for such data.

Conditional and Marginal Relationships

The regression coefficients in generalized linear mixed models represent conditional effects in the sense that they express comparisons holding the cluster-specific random effects (and covariates) constant. For this reason, conditional effects are sometimes referred to as cluster-specific effects. In contrast, marginal effects can be obtained by averaging the conditional expectation μ_{ij} over the random effects distribution. Marginal effects express comparisons of entire sub-population strata defined by covariate values and are sometimes referred to as population-averaged effects.

In linear mixed models (identity link), the regression coefficients can be interpreted as either conditional or marginal effects. However, conditional and marginal effects differ for most other link functions. This can easily be seen for a random intercept logistic regression model with a single covariate in **Figure 3**. The cluster-specific, conditional relationships are shown as dotted curves with horizontal shifts due to different values of the random intercept. The population-averaged, marginal curve is obtained by averaging the conditional curves at each value of x_{ij} . We see that the marginal curve resembles a logistic curve with a smaller regression coefficient. Hence, the marginal effect of x_{ij} is smaller than the conditional effect.

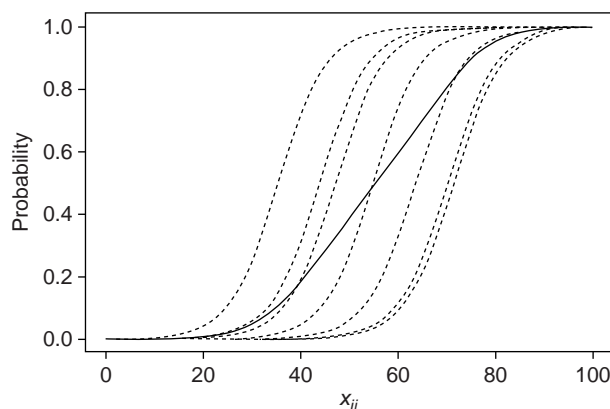


Figure 3 Conditional relationships (dotted curves) and marginal relationship (solid curve) for a random intercept logistic model. From Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, Structural Equation Models*, Boca Raton, FL: Chapman and Hall/CRC.

The difference between conditional and marginal relationships is also visible in **Figure 1**, but it is much less pronounced due to the relatively small estimated random intercept variance.

Estimation and Software

The model parameters to be estimated in most generalized linear mixed models are the fixed regression coefficients and the covariance matrix of the random effects (the variance ψ for a random-intercept model). The most commonly used estimation methods are maximum likelihood, penalized quasi-likelihood (PQL-1), and Markov chain Monte Carlo (MCMC).

Maximum likelihood estimation is not straightforward for generalized linear mixed models because the likelihood involves integrals that cannot be solved analytically. Many software packages therefore use numerical integration, typically adaptive or ordinary Gauss-Hermite quadrature. Examples include the Stata programs `xtlogit`, `xtpoisson`, `xtmelogit`, `xtmepoisson`, and `gllamm`; the SAS procedure `NLMIXED`; and the R-program `glmer`. The estimates in **Table 1** and the predictions for **Figure 1** were obtained using `gllamm` (see multimedia).

Numerical integration is time consuming, particularly if the model includes several random effects. Approximate methods have therefore been suggested, including PQL-1 which is implemented in the SAS procedure `GLIMMIX`, the S-PLUS or R function `glmmPQL`, and the standalone programs `HLM` and `MLwiN`. Unfortunately, these methods sometimes produce biased estimates, in particular for binary responses, small cluster sizes and large intraclass correlations of the latent responses. More accurate approximations have therefore also been implemented (PQL-2 in `MLwiN` and `Laplace6` in `HLM`). Software implementing MCMC for Bayesian estimation includes the general program `BUGS` or `WINBUGS` and the program `MLwiN` which is custom made for generalized linear mixed models.

Assigning Values to Random Effects

It is sometimes required to assign values to the random effects for individual clusters. For instance, random intercepts for schools or teachers in models for student outcomes can sometimes be viewed as measures of effectiveness since they represent the school- or teacher-specific value added taking into account observed covariates such as the prior achievement of students at intake.

If MCMC is used for estimation, there is no real distinction between parameters and random effects, and estimates of the latter are obtained in a straightforward

manner. If maximum likelihood or approximate methods are used for estimation, the parameters (typically the fixed regression coefficients and the covariance matrix of the random effects) are treated as known and replaced by their estimates for the purpose of assigning values to the random effects. The random effects are then either estimated by maximum likelihood or predicted using empirical Bayes.

Viewing the random effects as the only unknown parameters, they can be estimated by maximizing the joint probability distribution (or likelihood) of the responses, given the random effects and covariates, with respect to the random effects. This maximum likelihood estimation is performed independently for each cluster. In linear mixed models, such estimates are also known as ordinary least squares (OLS) estimates.

In empirical Bayes prediction of the random effects, we exploit the information that we have about the random effects before looking at the data for a cluster, namely the estimated random effects distribution, known as the prior distribution. The posterior distribution of the random effects for a cluster is proportional to the product of the prior distribution and the likelihood for the cluster. The mean of the posterior distribution is called the empirical Bayes predictor. In linear mixed models, these predictions are also known as best linear unbiased predictors (BLUPs).

The empirical Bayes predictions tend to shrink to zero compared with the maximum likelihood estimates. This shrinkage occurs because the prior distribution has its mean at zero. The shrinkage is negligible when the likelihood dominates the prior, for instance due to a large cluster size, but can be pronounced for small cluster sizes. Empirical Bayes predictors have lower mean squared prediction errors than maximum likelihood estimators and are, therefore, usually preferred.

There are generally no closed-form expressions for the empirical Bayes predictor except for linear mixed models. For other generalized linear mixed models, empirical Bayes predictions can be obtained by numerical integration.

Some Extensions

We have considered two-level models in this article which are applicable for typical educational datasets where students are nested in schools or where repeated observations on students over time are nested in students. There are often further levels of nesting. For instance, we may have repeated observations nested within students who are nested in schools. Alternatively, students may be nested in classrooms nested in schools, or schools may be nested in school districts or countries. In these cases, we can include random effects at each of the nested levels. More complex designs require crossed random effects.

For instance, in longitudinal studies, students may be in middle school in the first two waves and, subsequently, high school in the next two waves. Middle schools are typically crossed with high schools in the sense that all students from a given middle school do not attend the same high school (or vice versa).

Estimation of models with nested random effects is not much more complex than estimation of standard generalized linear mixed models. In contrast, estimation of models with crossed random effects remains a challenge, except in linear mixed models.

The random part of the linear predictor in generalized linear mixed models is quite restrictive because each random effect ζ_{pj} merely multiplies an observed variable x_{pij} in the reduced form. Generalized linear latent and mixed models (GLLAMMs) extend generalized linear mixed models by allowing the random effects to be multiplied by different parameters for different responses. Furthermore, the random effects can be regressed on observed covariates and other random effects (at the same or higher hierarchical levels). This is useful for incorporating measurement models, such as IRT models, within generalized linear mixed models.

See *also*: Categorical Data Analysis; Empirical Bayes Methods; Generalized Linear Models; Growth Modeling; Hierarchical Linear Models; Value-Added Models.

Bibliography

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman and Hall/CRC.

Further Reading

- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25.
- Clayton, D. G. (1996). Generalized linear mixed models. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds.) *Markov Chain Monte Carlo in Practice*, pp 275–301. London: Chapman and Hall.
- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Goldstein, H. (2003). *Multilevel Statistical Models*, 3rd edn. London: Arnold.
- Goldstein, H. and Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A* **159**, 505–513.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.
- McCulloch, C. E., Searle, S. R., and Neuhaus, J. M. (2008). *Generalized, Linear and Mixed models*, 2nd edn. New York: Wiley.
- Rabe-Hesketh, S. and Skrondal, A. (2009). Generalized linear mixed effects models. In Fitzmaurice, G. M., Davidian, M., Verbeke, G., and Molenberghs, G. (eds.) *Longitudinal Data Analysis*, pp 79–106. Boca Raton, FL: Chapman and Hall/CRC.

- Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata*, 2nd edn. College Station, TX: Stata Press.
- Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika* **69**, 167–190.
- Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* **128**, 301–323.
- Raudenbush, S. W. (1993). A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. *Journal of Educational and Behavioral Statistics* **18**, 321–349.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models*, 2nd edn. Thousand Oaks, CA: Sage.
- Raudenbush, S. W., Yang, M.-L., and Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics* **9**, 141–157.
- Ritz, J. and Spiegelman, D. (2004). A note about the equivalence of conditional and marginal regression models. *Statistical Methods in Medical Research* **13**, 309–323.
- Singer, J. D. and Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics* **18**, 155–195.
- Skrondal, A. and Rabe-Hesketh, S. (2003). Multilevel logistic regression for polytomous data and rankings. *Psychometrika* **68**, 267–287.
- Skrondal, A. and Rabe-Hesketh, S. (2007). Redundant overdispersion parameters in multilevel models. *Journal of Behavioral and Educational Statistics* **32**, 419–430.
- Skrondal, A. and Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society, Series A* **172**, 659–687.
- Snijders, T. A. B. and Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.

Relevant Websites

- <http://www.gllamm.org> – gllam software, documentation, examples, and other resources.
- <http://www.ssicentral.com/htm> – SSI, Scientific Software International, Hierarchical Linear and Nonlinear Models.
- <http://www.cmm.bristol.ac.uk> – University of Bristol, Centre for Multilevel Modelling.

Generalized Linear Models

J J Faraway, University of Bath, Bath, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Linear models are a way of describing a response variable in terms of a linear combination of predictor variables. The response should be a continuous variable and be at least approximately normally distributed. Such models find wide application, but cannot handle clearly discrete or skewed continuous responses. Generalized linear models (GLMs) allow the extension of linear modeling ideas to a wider class of response types, such as count data or binary responses. Many statistical methods exist for such data types, but the advantage of the GLM approach is that it unites a seemingly disparate collection of response types under a common modeling methodology. Estimation, testing, and diagnostics for this class of models follow a standard path. The ideas behind GLMs were developed over many years by many researchers, but the term itself and its definition were introduced by Nelder and Wedderburn (1972) and further popularized by McCullagh and Nelder (1989).

Definition

A GLM requires the specification of two defining characteristics – the distribution of the response and the link function which describes how the mean of the response is linked to a linear combination of the predictors. The distribution is expected to belong to the exponential family of distributions which has the general form:

$$f(y|\theta, \phi) = \exp \left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right]$$

The θ is called the canonical parameter and represents the location while ϕ is called the dispersion parameter and represents the scale. We may define various members of the family by specifying the functions a , b , and c . The three most commonly used members of this family are the normal, Poisson and binomial distributions. For example, for a Poisson with mean μ , we have $\theta = \log(\mu)$, $\phi \equiv 1$, $a(\phi) = 1$, $b(\theta) = \exp(\theta)$, and $c(y, \phi) = -\log y!$. Other less frequently used family members are the gamma and inverse Gaussian distributions. Some other distributions, such as the negative binomial and the Weibull are not family members, but are sufficiently closely related to be used in a GLM-like manner. Other distributions might be used, but at the cost of some of the common methodology that makes GLMs attractive in the first place.

The exponential family distributions have a mean and variance that can be described in terms of the defining functions a and b :

$$\begin{aligned} EY &= \mu = b'(\theta) \\ \text{var } Y &= b''(\theta)a(\phi) \end{aligned}$$

Notice that the mean is a function of θ and so this parameter describes location while the variance is a product of functions of the location parameter and the scale parameter ϕ . $b''(\theta)$ is called the variance function and describes how the variance relates to the mean. In the normal case, $b''(\theta) = 1$ and so the variance is independent of the mean. For other distributions, the variance changes with the mean.

The effect of the predictors on the response is expressed through a linear predictor:

$$\eta = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p = x^T \beta$$

Note that the predictors x_i might involve transformation or combinations of the original raw predictors. Furthermore, categorical predictors can be represented using indicator variables. This allows great flexibility in the representation of the effect of the predictors on the response.

The link function, g , describes how the mean response, $EY = \mu$, is linked to the covariates through the linear predictor:

$$\eta = g(\mu)$$

Any monotone continuous and differentiable function as a choice of g could suffice, but there are some convenient and common choices for the standard GLMs.

For the normal linear model, the identity link, $\eta = \mu$, is almost always used and the resulting GLM is just a standard linear model. For nonidentity g , $y = g^{-1}(x^T \beta) + \varepsilon$. This is not the same as a transform on the response ($g(y) = x^T \beta + \varepsilon$) or the commonly used transformations on the predictors. This is sometimes called a single index model, when g is estimated, not specified.

For the Poisson GLM, the mean μ must be positive. Using an identity link $\eta = \mu$ is not desirable because η can be negative. Using a log link so that $\mu = e^\eta$ ensures $\mu > 0$. This log link means that additive effects of x lead to multiplicative effects on the response. Other choices of link are very rarely used for the Poisson.

For the binomial GLM, let Y have probability of success μ in n trials. If we take the proportion Y/n as the

response, then $EY/n = \mu$. We need a link function that enforces $0 \leq \mu \leq 1$. There are several commonly used ways to ensure this, the logit link being the most frequently used, where $\eta = \log(\mu/(1 - \mu))$. The probit, $\eta = \Phi^{-1}(\mu)$ where Φ is the normal cumulative distribution function, also enjoys some use.

Every major statistical software package has code to fit GLMs. The user is simply required to specify the response type and the link function, along with the specific form of the linear predictor. Thus, knowledge of the theory of GLMs is not essential to use them in practice, but it is important to make an appropriate selection of distribution and link and understand the consequences of that choice.

Inference

The parameters, β , of the linear predictor of a GLM are usually estimated using maximum likelihood. Assuming independent observations, the log-likelihood is

$$\sum_i \left[\frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi) \right]$$

where $a_i(\phi) = \phi$. For the normal linear model, the maximum likelihood estimate (MLE) is just the usual least-squares solution. For other MLEs numerical optimization is required. Typically, an iteratively reweighted least-squares method is used. In most cases, the convergence is rapid. A failure to converge is often a sign of some problem with the model specification or an unusual feature of the data. Estimates of variance of the regression coefficients may be obtained from

$$\text{var}(\hat{\beta}) = (X^T W X)^{-1} \hat{\phi}$$

which is comparable to the form used in weighted least squares.

Hypothesis testing for GLMs can be expressed in terms of comparisons of nested models – a larger model and a smaller model, which usually consists of a subset of the predictors found in the larger model. The smallest model we might consider uses no predictors at all, meaning that we fit a common mean μ for all responses, that is, one location parameter only. This is called the null model.

The largest model we might consider fits the response exactly. Typically, we would need to use n parameters for n data points. This can be achieved in various ways and is called the saturated or full model. This model tells us no more than the data itself and is thus usually uninformative, but it does provide a good baseline for other models.

Suppose we compare the model of current interest to the full model by using the likelihood ratio statistic which is twice the difference in the log-likelihoods of the two

models. This is also called the (scaled) deviance which takes the form:

$$\sum_i 2(y_i \tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)) / \phi$$

where $\tilde{\theta}$ are the estimates under the full (saturated) model and $\hat{\theta}$ are the estimates under the model of interest. This can be written as $D(y, \hat{\mu}) / \phi$ where the numerator is the deviance and the ϕ is the scaling. Deviances for the common GLMs are shown in **Table 1**.

Pearson's X^2 statistic

$$X^2 = \sum_i \frac{(y_i - \hat{\mu}_i)^2}{\text{var}(\hat{\mu}_i)}$$

is an alternative measure of discrepancy that is sometimes used in place of the deviance.

The scaled deviance or indeed the Pearson's X^2 statistic can be viewed as a goodness-of-fit statistic. Provided the model is correct, the scaled deviance and the Pearson's X^2 statistic are both asymptotically χ^2 with degrees of freedom equal to the number of identifiable parameters. For GLMs that have a dispersion parameter, ϕ , this test statistic is not directly usable because we do not know the value of this parameter and, absent any replication in the data, we cannot estimate in a way that is not dependent on the current model. However, for the two important cases of the Poisson and binomial, $\phi = 1$ and the test can be used. Unfortunately, the accuracy of the asymptotic approximation is dubious for smaller datasets. For a binary, that is a 0–1 response, the approximation is worthless.

For comparing a larger model, L , to a smaller nested model, S the difference in the scaled deviances, $D_S - D_L$ is asymptotically χ^2 with degrees of freedom equal to the difference in the number of identifiable parameters in the two models. For the normal model and other models where the dispersion ϕ is usually not known, this test cannot be directly used. However, if we insert an estimate of ϕ we may compute an F -statistic of the form

$$\frac{(D_S - D_L) / (df_S - df_L)}{\hat{\phi}}$$

where $\hat{\phi} = X^2 / (n - p)$ is a good estimate of the dispersion. For the normal model, $\hat{\phi}$ is the residual sum of

Table 1 For the binomial y_i is distributed $\text{Bin}(m_i, p_i)$ and $\mu_i = m_i p_i$.

GLM	Deviance
Gaussian	$\sum_i (y_i - \hat{\mu}_i)^2$
Poisson	$2 \sum_i [y_i \log(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i)]$
Binomial	$2 \sum_i [y_i \log(y_i / \hat{\mu}_i) + (m_i - y_i) \log((m_i - y_i) / (m_i - \hat{\mu}_i))]$

For the Poisson, the deviance is known as the G-statistic. The second term $\sum_i (y_i - \hat{\mu}_i)$ is usually zero if an intercept term is used in the model.

squares for the smaller model divided by its degrees of freedom, and the resulting F -statistic has an exact F distribution for the null. For other GLMs with free dispersion parameters, the statistic is only approximately F distributed.

For every GLM except the normal, an approximate null distribution must be used whose accuracy may be in doubt particularly for smaller samples. However, the approximation is better when comparing models than for the goodness-of-fit statistic.

We may also take a Wald test approach. For a single parameter β_j , we may use the standard error of the estimate to construct a z -statistic of the form $\hat{\beta}_j/se(\hat{\beta}_j)$. This has an asymptotically normal null distribution. This is only an approximate test except in the special case of the normal GLM where the z -statistic is the t -statistic and has an exact t -distribution. The difference-of-deviances test is preferred to the Wald test.

As with standard linear models, it is important to check the adequacy of the assumptions that support the GLM. The diagnostic methods for GLMs mirror those used for normal linear models. However, some adaptations are necessary and, depending on the type of GLM, not all diagnostic methods will be applicable.

Residuals represent the difference between the data and the model and are essential to explore the adequacy of the model. In the normal case, the residuals are $\hat{e} = y - \hat{\mu}$. These are called response residuals for GLMs, but since the variance of the response is not constant for most GLMs, some modification is necessary. We would like residuals for GLMs to be defined such that they can be used in a similar way as in the normal linear model.

The Pearson residual is comparable to the standardized residuals used for linear models and is defined as

$$r_p = \frac{y - \hat{\mu}}{\sqrt{\text{var}(\hat{\mu})}}$$

These are just a rescaling of $y - \hat{\mu}$. Notice that $r_p^2 = X^2$ and hence the name. Pearson residuals can be skewed for non-normal responses. The deviance residuals are defined by analogy to Pearson residuals. The Pearson residual was r_p such that $\sum r_p^2 = X^2$, so we set the deviance residual as r_D such that $\sum r_D^2 = \text{deviance} = \sum d_i$.

Examples

Consider the (hypothetical) data shown in **Table 2** from six classrooms of 25 students each. We record the numbers of students that pass a test in each classroom along with information about a measure of average socioeconomic status and which of two teaching methods were used. It is natural to consider the response as binomially distributed with 25 trials in each case. This would require the assumptions that each student in a given classroom had

Table 2 Data showing the number (pass) of students out of 25 passing a test, socioeconomic status (ses), and teaching method (mth) used

Pass	Ses	Mth
10	1	0
20	2	1
15	3	0
21	4	1
20	5	0
24	6	1

an equal chance of passing the test and that the outcomes within a classroom were independent. It would be reasonable in practice to suspect the validity of these assumptions, but we shall proceed regardless. Heterogeneity in the probabilities would lead to the phenomenon of dispersion, for which modeling techniques exist. We choose a logit link function for this binomial GLM. It is difficult to verify any choice of link function without substantial homogeneous data. In this example, it is an unverifiable, but most common and reasonable choice. Having now specified the GLM by means of the response distribution and link function, we must select the variables for the linear predictor – we will simply take a linear combination of the two here. So our model states that pass_i is distributed $\text{Bin}(25, p_i)$ where

$$\log(p_i/(1 - p_i)) = \beta_0 + \beta_1 \text{ses}_i + \beta_2 \text{mth}_i$$

We fit this model with R software package (R Development Core Team (2008)) with the edited output shown below. Similar results can be expected with other statistical software.

Coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	−0.811	0.441	−1.84	0.0662
ses	0.421	0.131	3.21	0.0013
mth	1.166	0.432	2.70	0.0069

Null deviance: 26.25244 on 5 degrees of freedom

Residual deviance: 0.73718 on 3 degrees of freedom

AIC: 25.61

We see the estimated coefficients ($\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$) take the values (−0.811, 0.421, 1.1660). We can consider the residual deviance of 0.737 as a goodness-of-fit statistic. This should be compared to a χ^2_3 . Since the statistic is not large on this scale, we would conclude that the model does not lack for fit, and so we would probably not want for a more complex model than this. The null deviance of 26.3 is large for a χ^2_5 and so we would conclude that the null model does not fit here, that is to say children in different classrooms do not have a common probability of success. The group size of 25 is large enough that the accuracy of these tests is sufficient. Anything less than 5 for the group

sizes would make the goodness-of-fit test unreliable. We can also compare the fitted model to the null model by taking the difference in the deviances. This gives a test statistic of $26.25 - 0.74 = 25.51$ which is large for a χ^2_2 (two degrees of freedom difference between the two models). So we would prefer the fitted model to the simpler null model in this case.

We can test the significance of the individual predictors in two ways. For example, consider the teaching method. The z -statistic is 2.70 with a p -value of 0.0069, indicating a clearly significant result. However, we would prefer to test this hypothesis by fitting the appropriate smaller model, that is the one without the method variable. This has a residual deviance of 8.52 on four degrees of freedom. The difference in deviances is then $8.52 - 0.74 = 7.78$ on one (i.e., $4-3$) degree of freedom. Using the χ^2_1 , this yields a p -value of 0.005. Although the results are much the same and the second test requires a little more work, the accuracy of the latter test is significantly better, especially with sparser data. To interpret the size of the effect, it is best to use an odds scale. Thus, we can say that the use of the teaching method coded as one increases the odds of a pass by a factor of $\exp(1.166) = 3.21$ relative to the method coded as zero.

We can compute the deviance residuals as respectively

− 0.039 0.384 − 0.115 − 0.670 0.185 0.304

We do not expect deviance residuals for a binomial GLM to follow a standard normal distribution, but we might remark upon cases that exceed three in absolute value. In this case, there is nothing remotely like an outlier to concern us. For a more interesting data set we might proceed with the diagnostic methods used in standard linear models, but this example is too small to warrant further investigation.

Let us consider another example using the (hypothetical) data shown in **Table 3**. The number of incidents in eight schools are shown along with their size in numbers of students and an average school test score. We would like to know if the number of incidents is related to the test score. However, larger schools will generate more incidents and so this requires some adjustment.

The response is a count and best modeled as a Poisson. One might be tempted to model the number of incidents out of the school size as binomial, but it might not be reasonable to think that the incidents are independent events generated with some probability by each student.

We will use the log link to ensure a positive mean μ with the model taking the form

$$\log \mu = \log \text{size} + \beta_0 + \beta_1 \text{score}$$

This is called a rate model. The log size term will result in multiplying the rate at which the incidents occur by the number of students as we might wish here. Notice that this term has no regression coefficient (or we might say it is fixed at one). Such a fixed term is called an offset. The remaining part of the predictor is linear in the score. We fit this model using R although much the same results would be expected with other software. The summary results show

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	−4.6703	0.5319	−8.78	<2e−16
Score	−0.0139	0.0114	−1.22	0.22

Null deviance: 8.0380 on 7 degrees of freedom

Residual deviance: 6.4944 on 6 degrees of freedom

AIC: 33.97

The negative coefficient for score of -0.0139 shows that the rate of incidents decreases with increasing test scores. We might take the z -statistic of -1.22 to test the significance of this effect giving a p -value of 0.22. However, as before, it is better to use the difference of deviance to make the test. Here the null model is the appropriate smaller model so we may compute the test statistic as $8.04 - 6.49 = 1.55$ which yields a p -value of 0.21 using a χ^2_1 , as there is one degree of freedom difference between the two models. Either way, there is insufficient evidence of an effect of test scores on the rate of incidents.

A GLM approach can also be useful for modeling categorical data as seen in this example taken from Faraway (2006). In **Table 4**, we see subjects from the 1996 US NES study cross-classified by political identification (PID) and level of education.

Suppose we consider the counts as Poisson distributed with mean μ_{ij} and use a GLM with log link taking the form

$$\log \mu_{ij} = \mu_0 + \text{PID}_i + \text{education}_j$$

where the both predictors are considered as nominal (unordered) factors, each with seven levels. Then the rate at which observations fall in each of the cells will be some overall mean rate multiplied by a parameter

Table 3 Number of incidents occurring in schools of given sizes and average test scores

Incidents	7	5	1	5	2	1	5	2
Size	985	850	241	495	838	616	759	842
Score	32	74	44	20	50	73	48	36

Table 4 A subset of subjects drawn from the 1996 US NES study cross-classified by political identification and level of education

	<i>MS</i>	<i>HS dropout</i>	<i>HS</i>	<i>Some college</i>	<i>CC degree</i>	<i>BA degree</i>	<i>MA degree</i>
Strong Dem	5	19	59	38	17	40	22
Weak Dem	4	10	49	36	17	41	23
Ind Dem	1	4	28	15	13	27	20
Ind independent	0	3	12	9	3	6	4
Ind Rep	2	7	23	16	8	22	16
Weak Rep	0	5	35	40	15	38	17
Strong Rep	1	4	42	33	17	53	25

reflecting the level of party identification times another parameter representing the level of education. This is equivalent to the claim that the party identification and level of education are independent. The estimated parameters for the factor levels represent the relative proportions of observations falling in each row or column of the data. These may or may not be of interest, but they say nothing about the relationship between the two factors. In GLMs for categorical data, the interaction terms, not the main effects, are usually the center of interest because they model the dependence between the factors.

Fitting this model resulted in a deviance of 40.7 on 36 degrees of freedom. We might compare this to a model that allows dependence between the two factors. This would require a separate parameter for each cell. Since the number of parameters would equal the number of data points, the model would fit perfectly with a deviance of zero on zero degrees of freedom. Now if we compare this latter saturated model to the independence model above, we get a difference in deviance of 40.7 on 36 degrees of freedom resulting in a p -value of 0.27. Hence, we would prefer the smaller model which represents independence of the two factors.

Traditionally, the test for independence would use a Pearson's χ^2 which we might compute here as the sum of the squared Pearson residuals giving a value of 38.8 again on 36 degrees of freedom. Hence, the outcome of the test is quite similar to the deviance-based test.

The analysis so far has not used the ordinality present in both factors. One way to accommodate these within the model is by using scores. We assign scores u_i and v_j to the rows and columns such that $u_1 \leq u_2 \leq \dots \leq u_I$ and $v_1 \leq v_2 \leq \dots \leq v_J$. The assignment of scores requires some judgment. If one has no particular preference, even spacing allows for the simplest interpretation. The model becomes

$$\log \mu_{ij} = \mu_0 + \text{PID}_i + \text{education}_j + \gamma u_i v_j$$

where γ represents the amount of association and is rather like an (unscaled) correlation coefficient, $\gamma = 0$ means independence. We assigned the scores 1–7 respectively to each of the factors and fit this model. We obtained $\hat{\gamma} = 0.0287$ so the positive value indicates a positive

association although we cannot directly interpret the size of the effect. The deviance was 30.6 which represents a difference of 10.1 on just one degree of freedom from the independence model. This is a strongly significant result indicating that we have found an association here. Notice that it was essential to use the information about the ordinality of the factors to obtain this result.

Summary

Generalized linear models provide a common approach to a broad range of response modeling problems. Normal, Poisson, and binomial responses are the most commonly used, but other distributions can be used as well. Apart from specifying the response, GLMs also need a link function to be set which allows further flexibility in the modeling. The GLM can be fitted using a common procedure and a mechanism for hypothesis testing is available. Diagnostics using deviance residuals provide a way to check that chosen models are adequate.

Use of a GLM is by no means sufficient as there are aspects of analysis of all the different GLMs which are specific to that particular response type. For example, while a logistic regression is a GLM the user still needs to understand the particular interpretation of odds in this type of model. With categorical data, the imposition of a GLM sometimes requires parametrizations that are difficult to interpret and other approaches may be more revealing. Thus while GLMs might provide some convenient structure across seemingly different models, we must recognize that they do not provide a comprehensive solution.

The basic GLM has been extended in many directions. Wedderburn (1974) introduced the idea of quasi-likelihood, which avoids explicitly specifying the distribution of the response, requiring only that we specify the link and variance functions. This idea was extended to the notion of generalized estimating equations with particular application to longitudinal data by Liang and Zeger (1986). Lindsey (1997) gives a wide range of applications of GLMs while Fahrmeir and Tutz (2001) describe GLMs for multivariate responses.

Further reading about GLMs can be found in several books. McCullagh and Nelder (1989) is the classical reference although this can be difficult for readers without a background in statistical theory. Gill (2001), Fox (2008), and Dobson and Barnett (2008) provide more accessible texts. Faraway (2006) has an introduction to GLMs based on R software and Hardin and Hilbe (2007) has a text based on Stata.

See also: Categorical Data Analysis; Multivariate Linear Regression.

Bibliography

- Dobson, A. and Barnett, A. (2008). *An Introduction to Generalized Linear Models*, 3rd edn. Boca Raton, FL: Chapman and Hall.
- Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. New York: Springer.
- Faraway, J. (2006). *Extending the Linear Model with R*. London: Chapman and Hall.
- Fox, J. (2008). *Applied Regression Analysis and Generalized Linear Models*, 2nd edn. Thousand Oaks, CA: Sage.
- Gill, J. (2001). *Generalized Linear Models: A Unified Approach*. Thousand Oaks, CA: Sage.
- Hardin, J. and Hilbe, J. (2007). *Generalized Linear Models and Extensions*, 2nd edn. College Station, TX: Stata Press.
- Liang, K. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**(1), 13–22.
- Lindsey, J. (1997). *Applying Generalized Linear Models*. New York: Springer.
- McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.
- Nelder, J. and Wedderburn, R. (1972). *Generalized linear models*. *Journal of the Royal Statistical Society, Series A* **132**, 370–384.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **61**(3), 439–447.

Generating Random Numbers

P J Pashley and A Amodeo, Law School Admission Council, Newtown, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Educational research has a long history of employing random numbers. Before the advent of computers, the random number sequences needed were typically relatively short, utilized primarily within experimental designs that required random assignment of conditions to subjects (e.g., type of instruction delivered in classrooms) or vice versa. The need for these short random number sequences continues today, and they can be found in many of the same places as in the past, such as in the appendices of books on statistical methods, as well as now on the Internet.

Access to ever-increasing computer processing power and availability has ushered in an era of progressively more sophisticated statistical models, analysis tools, and other empirical procedures. Associated estimation and simulation methods can require very large sequences of random numbers. For example, consider the task of simulating correct and incorrect responses from an administration of the Law School Admission Test (LSAT). This assessment currently consists of about 100 multiple-choice items and is routinely administered to samples of approximately 50 000 test-takers at one time. Given model parameters that specify item characteristics and test-taker abilities, 5 million random numbers are needed to simulate a response for each item–test-taker combination. If these responses are used to form the basis for estimating statistics, such as the standard error of equating, many (i.e., tens or hundreds) of these random sets could be required.

The National Assessment of Educational Progress (NAEP) program provides a good case study for the variety of ways in which random numbers are employed in educational research. This is a nationally representative assessment of student achievement in the United States, conducted periodically within a number of subject areas. To ensure representativeness, probability samples of schools and students are randomly selected. For quality-control purposes, random samples of booklets and corresponding scores are routinely picked and inspected for errors during the assessment cycles. For reporting purposes, instead of calculating individual ability estimates, probability distributions of student proficiencies are estimated first, conditional on item responses and background variables. Then for each student in the assessment, five random draws (called plausible values) from a suitably estimated probability distribution are performed. These plausible values are subsequently used to estimate group score statistics and their related variances. For operational research purposes, large-scale simulations, employing millions of

random numbers, are routinely performed to study current and potential future assessment methodology (e.g., see Bennett, 2008). For further details on how the NAEP program makes use of random numbers, their website can be referenced.

This article presents aspects of the generation of long sequences of random numbers, including various types of generating algorithms and the properties of the sequences they yield. Researchers should be aware that their study results can be tainted by employing an inadequate random number generator (RNG). These substandard generators have been and continue to be made readily available to unsuspecting users through some software packages. Before the good and bad qualities of RNGs are addressed, however, the nature of random numbers is discussed.

True Random Numbers

Readers interested in an extended and complete theoretical and philosophical discussion of what a random sequence is should refer to Knuth (1998). For our purposes, we will simply define a true random number as a single unpredictable outcome from a random experiment. To be useful, all the possible events and their corresponding probabilities that constitute the random experiment are usually assumed to be known. For example, one toss of a fair coin (the random experiment) will yield a head or a tail (the possible events). The specific outcome of one coin toss will be unpredictable, even though we know the chance of a head or tail (the event probabilities) is 50/50. In other words, it is not sufficient for random numbers to be unpredictable: to be useful, they must also conform to a known sampling distribution.

Repeated tossing of a fair coin will yield a sequence of random numbers, each of which will constitute an independent observation from a random experiment. In other words, for any particular toss, the results of previous and subsequent tosses will have no impact on the likelihood of a head or tail outcome. Therefore, if a fair coin is tossed 99 times and 99 heads are observed, the chances of a head appearing on the 100th toss will still be 50/50.

While the tossing of coins to produce large sequences of random numbers is clearly impractical, efficient tools for measuring and recording other (seemingly) random physical events have been developed. For example, the ZRANDOM USB generator samples thermal electronic noise to produce random numbers that can be fed into a

computer by way of a universal serial bus (USB) port. Alternatively, large random sequences can be downloaded from websites such as <http://www.random.org> and <http://www.fourmilab.ch/hotbits>, which also employ physical events to produce random numbers (radio noise and radioactive decay, respectively).

Providers of these types of sequences usually claim that they are providing true random numbers. However, just as no truly fair coins exist, all measures of physical events will exhibit some, although hopefully inconsequential, amounts of bias (i.e., departures from the assumed underlying sampling distribution). Researchers are urged to independently test all sequences they use for randomness (see below for more details), regardless of whether they are labeled as true random numbers or not.

Random Number Distributions

Most basic RNGs simulate uniformly distributed random numbers (or uniform variates) that range between zero and one. A theoretical uniform distribution is defined to be continuous over an interval, with endpoints assumed here to be zero and one. In practical terms, however, random numbers are typically output or stored with a relatively small number of decimal places (e.g., six). Therefore, instead of a continuum, one can think of a finite, although very large, number of discrete points between zero and one that the generated random numbers can assume. Under these practical constraints, there should be an equal probability that a generated random number can assume any of the possible discrete points that span the distribution.

Output from uniform RNGs can be categorized into smaller groupings for various purposes, such as simulating the tossing of a coin (e.g., by categorizing random numbers less than 0.5 as heads and those greater than or equal to 0.5 as tails). Simulations of student behavior can be undertaken in a similar fashion. For instance, returning to the LSAT example given earlier, correct and incorrect responses can be simulated by first estimating the probability that test-takers will answer individual items correctly, usually by employing an item response theory (IRT) model. Then individual responses can be simulated by comparing a uniform variate, say the k th number (denoted by u_k) in the generated sequence, to the probability of test-taker i answering item j (estimated to be p_{ij}) as follows:

set the simulated response to correct if $u_k < p_{ij}$; otherwise,
set the simulated response to incorrect.

If random sequences from other distributions are required, uniformly distributed random number sequences are commonly first generated and then transformed into the desired variants. For example, the Box–Muller

transformation (Box and Muller, 1958) can be applied to two uniform variates (u_1 and u_2) to produce two independent univariate standard normal variates (n_1 and n_2) by assigning

$$n_1 = \sqrt{-2\ln u_1} \cos(2\pi u_2), \text{ and } n_2 = \sqrt{-2\ln u_1} \sin(2\pi u_2).$$

Note that the arguments of the $\cos(\bullet)$ function are in radians.

Gentle (2003) provides a thorough review of methodology available to generate random numbers that follow most commonly used univariate and multivariate distributions, along with more general approaches that are applicable to any type of theoretical distributional assumptions.

Deterministic Generation Methods

A required feature of the random sequences used in many educational research studies is that they be reproducible. This allows a researcher to rerun his or her simulation exactly and allows other independent researchers the ability to replicate the study. Because of the unpredictable process by which they are generated, random sequences yielded by physical events must be stored if they are to be reused. These databases can be quite large, and sharing them can be cumbersome.

As a convenient alternative, methods that generate random number sequences algorithmically, called deterministic methods, can be employed. In order to reproduce the sequences these generators yield, they simply need to be rerun with the same initial values. They can also be easily shared with others by listing the algorithm the generator is based on or by providing related software code.

Because of their deterministic nature, the sequences of numbers that these types of generators produce cannot be thought of as truly random. However, for practical purposes, if the sequences produced by a deterministic method adhere sufficiently to the properties of truly (i.e., theoretical) random sequences, they can legitimately be employed in their stead. These types of deterministic generators are often referred to in the literature as pseudorandom number generators. However, we will simply call them RNGs here, given the fact that the ultimate goal of any RNG, including deterministic ones, is to generate sequences of numbers that are indistinguishable from those we would expect to obtain from a true RNG, if one existed. Unfortunately, some early RNGs fell well short of achieving that goal.

An early and extensively studied type of RNG, called the multiplicative congruential generator (MCG), was suggested by D. H. Lehmer (1951). Every new value (x_{j+1}) it generates is a result of the following simple

mathematical manipulation on the previously generated value (x_j):

$$x_{j+1} = (ax_j) \bmod m,$$

where m is the modulus, with $m > 0$; and a is the multiplier, with $0 \leq a < m$.

The mod function finds remainders as follows:

$$(x) \bmod m = x - m\text{FLOOR}(x/m)$$

where the FLOOR(\bullet) function yields the largest integer less than or equal to the given argument.

To begin the sequence, a starting value or seed is chosen, usually denoted as x_0 (with $0 \leq x_0 < m$). Uniform variates (u_j 's) can be derived from the generated values (x_j 's) by dividing them by the modulus:

$$u_j = x_j / m$$

To illustrate a simple MCG, let $x_0 = 7$, $a = 7$, and $m = 10$. The resultant sequence, including the seed, is

$$7, 9, 3, 1, 7, 9, 3, 1, 7, 9, 3, 1, \dots$$

and the related uniform variates are

$$.7, .9, .3, .1, \dots$$

In this example, note that the MCG starts repeating itself after only four values. The length of the repeating cycle, in this case four, is called the period.

Because of its relative mathematical simplicity, many analytical results are available for MCGs. For example, it is known that an MCG has a period of m if and only if:

1. $a - 1$ is a multiple of p for every prime p dividing m ; and
2. if m is a multiple of 4, then $a - 1$ must also be a multiple of 4.

For details on this and other theoretical MCG properties, readers can refer to Knuth (1998).

One unfortunate property that all MCGs possess is that if their sequences, after being assigned to n -tuples, are viewed as points in an n -dimensional cube, all the points will be found to lie on a relatively few parallel hyperplanes (Marsaglia, 1968). To illustrate this type of phenomenon, consider the case of the (once) widely used IBM SYSTEM/360 (1968) RANDU generator. It was essentially defined by a multiplier and modulus with values 65 539 and 2^{31} , respectively. These specifications were used to generate 6000 uniform variates that were combined sequentially into 2000 independent three-tuples:

$$(u_1, u_2, u_3), (u_4, u_5, u_6), \dots, (u_{5998}, u_{5999}, u_{6000}).$$

Figure 1 provides five views of these 2000 generated three-dimensional points plotted within a unit cube. The five views were created by inducing successive vertical rotations and horizontal tilts. Note that views (a) and (e) appear two-dimensional simply because of the orientation

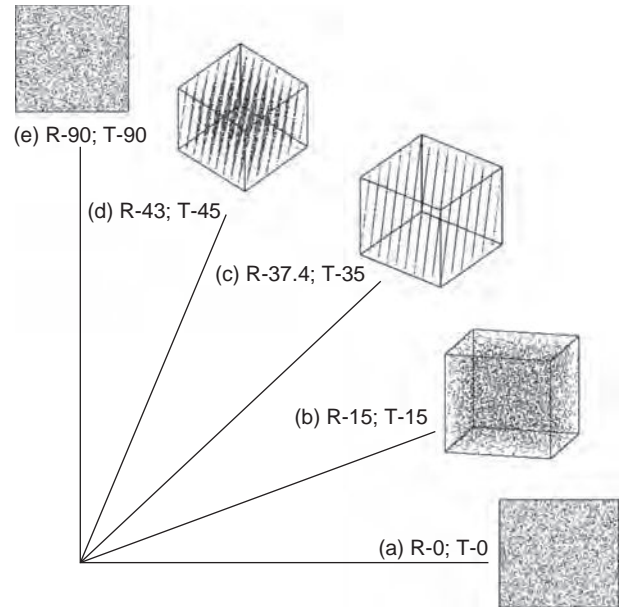


Figure 1 Five views of a unit cube containing 2000 random points, with degree of vertical rotation (R) and horizontal tilt (T) indicated.

of the cube. Ideally, these 2000 points should be randomly scattered throughout the cube, as they appear to be in views (a), (b), and (e). However, views (c) and (d) clearly show that the points actually line up on 15 hyperplanes.

Gentle (2003) provides a good review of other classes of uniform RNGs, including multiple recursive generators such as the lagged Fibonacci congruential generator, matrix-based generators, and nonlinear congruential generators. These have all been investigated analytically and empirically to varying degrees, although none more than the MCG. All have their strengths and weaknesses. A way to address the weaknesses inherent in a single RNG is to combine it with one or more other RNGs into a composite RNG.

One approach to combining two RNGs is called shuffling, whereby one of the generators is used to produce a random sequence and a second is used to reorder the sequence. A different approach entails generating a random sequence of pairs of uniform variates using two separate RNGs, and then choosing one value from each pair using a third RNG to yield the final random sequence. No matter how multiple RNGs are integrated, it appears wise to combine different classes of RNGs into composites. For example, combining two MCGs into a composite might actually exacerbate certain problems rather than solve them.

A type of RNG that has become widely used is the Mersenne twister (Matsumoto and Nishimura, 1998). Along with good empirical randomness results, their type of generator boasts very long periods (e.g., the version MT19937 has a period of $2^{19,937} - 1$). Researchers

who intend to run very extensive simulations should be aware of the period of the RNG they are employing to avoid the possibility of repeat sequencing. Note, however, that long periods do not necessarily guarantee better results in terms of randomness.

The book *Numerical Recipes: The Art of Scientific Computing* (Press *et al.*, 2007) is a good source for descriptions of reasonable RNGs and software for implementing specific versions of them. The authors also discuss approaches to producing variates from nonuniform distributions and provide the software to generate them.

Tests of Randomness

As indicated earlier, problems with many widely used RNGs have been observed in the past. For some applications, these deficiencies may not be critical. However, if random numbers are the basis of an extensive simulation study, reported results could be contaminated with spurious flaws in the RNG being employed.

Researchers testing an RNG for randomness are in the enviable position of knowing the exact underlying distributional properties of the statistical tests they are conducting. This is not the case in almost all empirical educational research situations. For example, consider a simple *t*-test comparing the means from two groups. This test assumes that the means were calculated from two normal distributions whose variances are equal. While these are often reasonable assumptions to make, in real-world research situations, no one knows how closely they represent the true state of affairs.

In contrast, if an RNG is designed to yield uniform variates, for example, we know exactly how the generated sequences should behave. Therefore, we know that if an RNG fails a randomness test, it is not due to departures from underlying distributional assumptions, but rather to the failure of the RNG to conform to them.

As illustrated earlier (see **Figure 1**), graphical methods can be effective in uncovering flaws in random sequences. Besides visual confirmations of basic randomness, plots such as the residual Q-Q plot (e.g., see Chambers *et al.*, 1983) can be used to detect departures from assumed distributional properties. These types of plots are especially effective when paired with related confidence intervals, summary statistics, and significance tests (e.g., see Pashley, 1993).

As there are so many aspects to what constitutes randomness, batteries of tests should be used when assessing the quality of a particular RNG. One such suite of 18 goodness-of-fit tests called DIEHARD was compiled by Marsaglia (1985). A CD containing this suite, along with a selection of RNGs, is available from the author. The National Institute of Standards and Technology (NIST)

also recommends a battery of 16 tests that can be found at their website.

Finally, while running multiple significance tests, one should be cognizant of the overall significance level, not just those corresponding to individual tests. In other words, researchers should be aware that the chance of obtaining a false-positive finding increases with the number of separate significance tests performed.

Additional Features to Consider

While it is important for researchers to conduct a number of statistical tests to ensure that the RNG they are considering will provide adequate results in terms of distributional properties, there are a number of other features that should be evaluated. Two such attributes – the reproducibility and the period of generated random sequences – have already been discussed earlier. Three more features are presented now – portability, efficiency, and prominence – that should also be considered, especially when choosing among a number of deterministic RNGs. With this in mind, we will restrict our attention here to algorithmically produced random numbers.

Portability is an attribute that impacts on how reproducible random numbers are across different computer platforms. In other words, a portable RNG program or software code should yield the same sequences when run on different computers. This is an important feature, not only because programs are often shared among researchers, but also because individual researchers may themselves run their programs on multiple computers that they have access to at work, in a lab, or at home. In addition, programs are sometimes developed and tested on personal computers and then run more extensively on mainframe computers.

Besides affecting the reproducibility of results, lack of portability may have consequences in terms of the expected randomness of the generated sequences. That is, an RNG could pass a battery of tests on one platform, but fail some of them on a different machine. Determining the extent to which an RNG is theoretically portable at the level of machine processing (e.g., determining the effects of employing different data types, computer languages, and compilers) is a complicated undertaking. Fortunately, RNGs can be easily tested for portability across a number of different platforms by comparing generated values at the beginning and near the end of an RNG's useful cycle.

The efficiency of an RNG's algorithm in terms of time to execute should be considered, although this attribute is becoming less of a concern as the processing capabilities of computers continue to increase. That said, if one RNG produces values in half the time of another, this can translate into significant time savings when extensive

simulations are being run. The efficiencies of different RNGs can be easily compared by recording the time each takes to generate very long number sequences. One should be careful to ensure that RNGs are tested in the same manner in which they will eventually be employed in a subsequent simulation (e.g., called as a subroutine). While efficiency can be an important feature to consider, the quality of different RNGs in terms of apparent randomness should almost always trump their efficiency when users are attempting to select among them.

Finally, users choosing among available RNGs would be wise to consider their prominence in terms of how thoroughly they have been studied and reported on in the scientific literature. There is no perfect RNG that will withstand all tests for randomness. However, users of an extensively investigated RNG will know most of its potential limitations, as opposed to a newly proposed RNG that might be an improvement in some sense (e.g., longer periods), but might also contain other unstudied deficiencies.

Summary

Robert R. Coveyou once said, "The generation of random numbers is too important to be left to chance." Even casual users of RNGs should take this quote to heart. While some software packages have upgraded the resident RNGs they provide in recent versions, such as Excel 2003 and above, others have not. Users should certainly be wary of any generic-looking RNG that lacks supporting documentation that specifies the generating methodology used and the randomness tests it has passed.

The required amount of additional investigations into the adequacies of a particular RNG should be proportional to the potential impact that flawed random sequences might have on a research study. The empirical models used in educational research continue to increase in complexity, and these models often characterize large numbers of variables with elaborate interrelationships. Researchers undertaking large-scale simulation studies involving these complex models must ensure that the RNG they employ offers random sequences even when viewed at high levels of dimensionality.

When research results are reported, a description of any RNG used should be included, just as study conditions and subject attributes are normally described. The type of RNG used should be specified, including its components if it is a composite generator, along with how another researcher might obtain or run a copy of the RNG. If the RNG code was embedded in a larger simulation program, the underlying algorithm should be specified in an appendix. The types of tests or test batteries the RNG passed, or is known to have passed, should be documented.

The portability of an RNG employed in simulation studies could also be discussed (e.g., on what types of computers it was used), and reproducibility issues could be addressed by providing values from the beginning (i.e., initial values) and end of a sufficiently long generated sequence. Some indication of the RNG's cycle period would also be useful to other researchers who wish to extend simulation studies.

Regardless of whether one is a casual or high-end user of RNGs, or somewhere in-between, understanding a generator's strengths and weaknesses is important. After all, educational practice should be influenced by valid research findings, not the incidental idiosyncrasies of a flawed RNG.

Bibliography

- Bennett, R. E. (2008). Technology for large-scale assessment. *ETS Report No. RM-08-10*. Princeton, NJ: Educational Testing Service.
- Box, G. E. P. and Muller, M. E. (1958). A note on the generation of random normal deviates. *Annals of Mathematical Statistics* **29**, 610–611.
- Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983). *Graphical Models for Data Analysis*. Belmont, CA: Wadsworth.
- Coveyou, R. R. (1969). Random number generation is too important to be left to chance. In *Studies in Applied Mathematics* 3, pp. 70–111. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Gentle, J. E. (2003). *Random Number Generation and Monte Carlo Methods*, 2nd edn. New York: Springer Science/Business Media.
- Knuth, D. E. (1998). *The Art of Computer Programming: Volume 2. Seminumerical Algorithms*, 3rd edn. Reading, MA: Addison-Wesley.
- Lehmer, D. H. (1951). Mathematical models in large-scale computing units. *Annual Computing Lab Harvard University* **26**, 141–146.
- Marsaglia, G. (1968). Random numbers fall mainly in the planes. *Proceedings of the National Academy of Sciences of the United States of America* **61**, 25–28.
- Marsaglia, G. (1985). *Proceedings of Computer Science and Statistics: 16th Symposium on the Interface*. Amsterdam: Elsevier.
- Matsumoto, M. and Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* **8**(1), 3–30.
- Pashley, P. (1993). On generating random sequences. In Keren, G. and Lewis, C. (eds.) *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues*, pp 395–411. Hillsdale, NJ: Erlbaum.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing*, 3rd edn. New York: Cambridge University Press.

Further Reading

- L'Ecuyer, P. (2004). Random number generation. In Gentle, J. E., Härdle, W., and Mori, Y. (eds.) *Handbook of Computational Statistics*, pp 35–70. Berlin: Springer.
- Marsaglia, G. (1995). *The Marsaglia Random Number Generator CDROM including the Diehard Battery of Tests of Randomness*. <http://www.stat.fsu.edu/pub/diehard> (accessed August 2009).
- Matsumoto, M. and Nishimura, T. (nd). *Mersenne Twister: A Random Number Generator (since 1997/10)*. Hiroshima University Department of Mathematics. <http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/emt.html> (accessed August 2009).

RAND Corporation (1955). *A Million Random Digits with 100,000 Normal Deviates*. Glencoe, IL: Free Press.

Walker, J. (1996). *HotBits: Genuine Random Numbers*. <http://www.fourmilab.ch/hotbits> (accessed August 2009).

<http://www.westphal-electronic.de/zranusbe.htm> – True Random Number Generator ZRANDOM USB, True Random Bits: True Random Numbers

Relevant Websites

<http://nces.ed.gov> – National Center for Education Statistics, The Nation's Report Card: National Assessment of Educational Progress.

<http://www.csrc.nist.gov> – National Institute of Standards and Technology, Computer Security Division: Computer Security Resource Center.

<http://www.random.org> – RANDOM.ORG: True Random Number Service (1998–2006).

<http://www.randomizer.org> – Research Randomizer: Free Random Sampling and Random Assignment.

<http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/ARTICLES/sfmt.pdf> – SIMD-oriented Fast Mersenne Twister: a 128-bit Pseudorandom Number Generator.

Goodness-of-Fit Testing

A Maydeu-Olivares and C García-Forero, University of Barcelona, Barcelona, Spain

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Absolute goodness of fit – The discrepancy between a statistical model and the data at hand.

Goodness-of-fit index – A numerical summary of the discrepancy between the observed values and the values expected under a statistical model.

Goodness-of-fit statistic – A goodness-of-fit index with known sampling distribution that may be used in statistical-hypothesis testing.

Relative goodness of fit – The discrepancy between two statistical models.

Introduction

The goodness of Fit (GOF) of a statistical model describes how well it fits into a set of observations. GOF indices summarize the discrepancy between the observed values and the values expected under a statistical model. GOF statistics are GOF indices with known sampling distributions, usually obtained using asymptotic methods, that are used in statistical hypothesis testing. As large sample approximations may behave poorly in small samples, a great deal of research using simulation studies has been devoted to investigate under which conditions the asymptotic p -values of GOF statistics are accurate (i.e., how large the sample size must be for models of different sizes).

Assessing absolute model fit (i.e., the discrepancy between a model and the data) is critical in applications, as inferences drawn on poorly fitting models may be badly misleading. Applied researchers must examine not only the overall fit of their models, but they should also perform a piecewise assessment. It may well be that a model fits well overall but that it fits poorly some parts of the data, suggesting the use of an alternative model. The piecewise GOF assessment may also reveal the source of misfit in poorly fitting models.

When more than one substantive model is under consideration, researchers are also interested in a relative model fit (i.e., the discrepancy between two models; see Yuan and Bentler, 2004; Maydeu-Olivares and Cai, 2006). Thus, we can classify GOF assessment using two useful dichotomies: GOF indices versus GOF statistics, and absolute fit versus relative fit. In turn, GOF indices and statistics can be classified as overall or piecewise. A third

useful dichotomy to classify GOF assessment is based on the nature of the observed data, discrete versus continuous. Historically, GOF assessment for multivariate discrete data and that for multivariate continuous data have been presented as being completely different. However, new developments in limited information GOF assessment for discrete data reveal that there are strong similarities between the two, and here, we shall highlight the similarities in GOF assessment for discrete and continuous data.

GOF testing with Discrete Observed Data

Consider modeling N observations on n discrete random variables, each with K categories, such as the responses to n test items. The observed responses can then be gathered in an n -dimensional contingency table with $C = K^n$ cells. Within this setting, assessing the GOF of a model involves assessing the discrepancy between the observed proportions and the probabilities expected under the model across all cells $c = 1, \dots, C$ of the contingency table. More formally, let π_c be the probability of one such cell and let p_c be the observed proportion. Let $\boldsymbol{\pi}(\boldsymbol{\theta})$ be the C -dimensional vector of model probabilities expressed as a function of, say, q model parameters to be estimated from the data. Then, the null hypothesis to be tested is $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\theta})$, that is, the model holds, against $H_1 : \boldsymbol{\pi} \neq \boldsymbol{\pi}(\boldsymbol{\theta})$.

GOF Statistics for Assessing Overall Fit

The two standard GOF statistics for discrete data are Pearson's statistic

$$X^2 = N \sum_{c=1}^C (p_c - \hat{\pi}_c)^2 / \hat{\pi}_c, \quad [1]$$

and the likelihood ratio statistic

$$G^2 = 2N \sum_{c=1}^C p_c \ln(p_c / \hat{\pi}_c). \quad [2]$$

where $\hat{\pi}_c = \pi_c(\hat{\boldsymbol{\theta}})$ denotes the probability of cell c under the model.

Asymptotic p -values for both statistics can be obtained using a chi-square distribution with $C - q - 1$ degrees of freedom when maximum likelihood estimation is used. However, these asymptotic p -values are only correct when

all expected frequencies $N\hat{\pi}_c$ are large (>5 is the usual rule of thumb). A practical way to evaluate whether the asymptotic p -values for X^2 and G^2 are valid is to compare them. If the p -values are similar, then both are likely to be correct. If they are very different, it is most likely that both p -values are incorrect.

Unfortunately, as the number of cells in the table increases, the expected frequencies become small (as the sum of all C probabilities must be equal to 1). As a result, in multivariate discrete data analysis, most often, the p -values for these statistics cannot be trusted. In fact, when the number of categories is large (say $k > 4$), the asymptotic p -values almost invariably become inaccurate as soon as $n > 5$. To overcome the problem of the inaccuracy of the asymptotic p -values for these statistics, two general methods have been proposed: resampling methods (e.g., bootstrap), and pooling cells. Unfortunately, existing evidence suggest that resampling methods do not yield accurate p -values for the X^2 and G^2 statistics (Tollenaar and Mooijart, 2003). Pooling cells may be a viable alternative to obtain accurate p -values in some instances. For instance, rating items with five categories can be pooled into three categories to reduce sparseness. However, if the number of variables is large, the resulting table may still yield some small expected frequencies. Moreover, pooling may distort the purpose of the analysis. Finally, pooling must be performed before the analysis is made to obtain a statistic with the appropriate asymptotic reference distribution.

Due to the difficulties posed by small expected probabilities on obtaining accurate p -values for GOF statistics assessing absolute models, some researchers have resorted to examining only the relative fit of the models under consideration, without assessing the absolute model fit. Other researchers simply use GOF indices.

GOF Indices

With L denoting the loglikelihood, two popular GOF indices are Akaike's information criterion (AIC), $AIC = -2L + 2q$ and Schwarz Bayesian information criterion (BIC), $BIC = -2L + q \ln(N)$,

$$AIC = -2L + 2q, BIC = -2L + q \ln(N) \quad [3]$$

The AIC and BIC are not used to test the model in the sense of hypothesis testing, but for model selection. Given a data set, a researcher chooses either the AIC or BIC, and computes it for all models under consideration. Then, the model with the lowest index is selected. Notice that both the AIC and BIC combine absolute fit with model parsimony. That is, they penalize by adding parameters to the model, but they do so differently. Of the two, the BIC penalizes by adding parameters to the model more strongly than the AIC.

GOF Statistics for Piecewise Assessment of Fit

In closing this section, the standard method for assessing the source of misfit is the use of z -scores for cell residuals

$$\frac{p_c - \hat{\pi}_c}{SE(p_c - \hat{\pi}_c)}, \quad [4]$$

where SE denotes standard error. In large samples, their distribution can be approximated using a standard normal distribution. Unfortunately, the use of these residuals even in moderately large contingency tables, is challenging. It is difficult to find trends in inspecting these residuals, and the number of residuals to be inspected is easily too large. Most importantly, for large C , because the cell frequencies are integers and the expected frequencies must be very small, the resulting residuals will be either very small or very large.

New Developments in GOF with Discrete Observed Data: Limited Information Methods

In standard GOF methods for discrete data, contingency tables are characterized using cell probabilities. However, they can be equivalently characterized using marginal probabilities. To see this, consider the following 2×3 contingency table:

	$X_2 = 0$	$X_2 = 1$	$X_2 = 2$
$X_1 = 0$	π_{00}	π_{01}	π_{02}
$X_1 = 1$	π_{11}	π_{11}	π_{12}

This table can be characterized using the cell probabilities $\boldsymbol{\pi} = (\pi_{00}, \dots, \pi_{12})'$. Alternatively, it can be characterized using the univariate $\boldsymbol{\pi}_1 = (\pi_1^{(1)}, \pi_2^{(1)}, \pi_2^{(2)})$ and bivariate $\boldsymbol{\pi}_2 = (\pi_{12}^{(1)(1)}, \pi_{12}^{(1)(2)})$ probabilities, where $\pi_i^{(k)} = \Pr(X_i = k)$, $\pi_{ij}^{(k)(l)} = \Pr(X_i = k, X_j = l)$, and

	$X_2 = 0$	$X_2 = 1$	$X_2 = 2$	
$X_1 = 0$				
$X_1 = 1$		$\pi_{12}^{(1)(1)}$	$\pi_{12}^{(1)(2)}$	$\pi_1^{(1)}$
		$\pi_{(2)}^{(1)}$	$\pi_{(2)}^{(2)}$	

Both characterizations are equivalent, and the equivalence extends to contingency tables of any dimension.

Limited-information GOF methods disregard information contained in the higher-order marginals of the table. Thus, quadratic forms in, say, univariate and bivariate residuals are used instead of using all marginal residuals up to order n .

GOF Statistics for Assessing Overall Fit

Maydeu-Olivares and Joe (2005, 2006) proposed a family of GOF statistics, M_r , that provides a unified framework for limited information and full information GOF statistics. This family can be written as

$$M_r = N \hat{\mathbf{e}}_r' \hat{\mathbf{C}} \hat{\mathbf{e}}_r, \quad [5]$$

where $\hat{\mathbf{e}}_r$ are the residual proportions up to order r and

$$\mathbf{C} = \Gamma_r^{-1} - \Gamma_r^{-1} \Delta_r (\Delta_r' \Gamma_r^{-1} \Delta_r)^{-1} \Delta_r' \Gamma_r^{-1}. \quad [6]$$

Here, Γ_r denotes the asymptotic covariance matrix of the residual proportions up to order r and Δ_r is a matrix of derivatives of the marginal probabilities up to order r with respect to the model parameters. Two members of this family are, for instance, M_2 and M_m . In M_2 only univariate and bivariate residuals are used. In M_m all residuals up to order n , the number of variables, are used. When ML estimation is used, M_m is algebraically equal to Pearson's X^2 .

The asymptotic distribution of any statistic of the M_r family is chi-square with degrees of freedom (df) = number of residuals used $-q$. For the chi-square approximation to M_r to be accurate, the expected frequencies of $\min(2r, n)$ marginals need to be large. Thus, for M_m , expected cell frequencies need to be large, but for M_2 , where $r = 2$, only expected frequencies for sets of $\min(2r, n) = 4$ variables need to be large (provided $n > 4$). As a result, when only low-order margins are used, the asymptotic p -values are accurate even in gigantic models and small samples. Furthermore, often more power is obtained than when all the information available in the data is used. Consequently, Maydeu-Olivares and Joe suggest testing at the highest level of margins for which a model is identified, discarding higher-order margins. Since most models are identified using only univariate and bivariate information (i.e., they can be estimated using only univariate and bivariate information), M_2 should be the statistic of choice. For instance, the two-parameter logistic item-response theory (IRT) model is identified (it can be estimated) using only univariate and bivariate information. As a result, its fit may be tested using M_2 .

GOF Statistics for Piecewise Assessment of Fit

For a piecewise assessment of fit, z -scores for marginal residuals involving one, two, or three variables may be used (i.e., the marginal residuals divided by their SE s). It is simpler to extract valuable information from them than from cell residuals. However, when the number of categories is large, there are often too many marginal residuals. In these cases, Pearson's X^2 may be computed for pairs (or if needed, triplets of variables), provided there are enough degrees of freedom for testing. Here,

df = number of residuals used – number of estimated parameters involved. However, when used for piecewise assessment of fit, X^2 may yield an undue impression of poor fit (Maydeu-Olivares and Joe, 2006). The use of the M_r statistics in place of X^2 for pairs (or triplets) of variables solves this problem.

GOF Testing with Continuous Observed Data

GOF is a very active area of research in structural equation modeling (SEM). In classical SEM applications, multivariate models for continuous data (often involving latent variables) are estimated from some summary statistics (typically means and covariances or correlations). For ease of exposition, here we assume that the model is estimated using covariances, but the results can be easily extended to models estimated from other sets of statistics.

Let $\boldsymbol{\sigma}(\boldsymbol{\theta})$ be the $t = n(n+1)/2$ nonredundant population variances and covariances expressed as a function of q model parameters, and let \mathbf{s} be its sample counterpart. The null hypothesis to be tested is $H_0: \boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\theta})$, that is, that the model holds, against $H_1: \boldsymbol{\sigma} \neq \boldsymbol{\sigma}(\boldsymbol{\theta})$.

GOF Statistics for Assessing Overall Fit

Two procedures can be used to obtain a GOF statistic. The first procedure is based on using the minimum of the fitted function, \hat{F} , multiplied by sample size, N , say $T = N\hat{F}$. This is the usual chi-square test in SEM. The second procedure is based on using a quadratic form in residual summary statistics.

Now, T will only be asymptotically chi-square distributed if the estimator is asymptotically efficient for the distribution of the data. Under multivariate normal assumptions, the efficient estimators are maximum likelihood (ML) and generalized least squares (GLS). Under the asymptotically distribution-free (ADF) distributional assumptions set forth by Browne (1982), the efficient estimator is weighted least squares (WLS). Thus, T is asymptotically distributed as a chi-square with $t-q$ degrees of freedom only when ML or GLS estimation is used under normality, or when WLS estimation is used under ADF assumptions. In all other cases where T is not asymptotically chi-square, its distribution may be approximated by a chi-square if T is scaled by its asymptotic mean or adjusted by its asymptotic mean and variance. These are the so-called Satorra and Bentler (1994) T_s and T_a test statistics, respectively. Thus, if a model is estimated using, for instance, unweighted least squares, T is not asymptotically chi-square, but its distribution can be approximated using a chi-square distribution using Satorra-Bentler corrections.

Regarding the second procedure, Browne (1982) proposed the residual-based statistic

$$T_B = N\hat{\epsilon}'\hat{C}\hat{\epsilon}, \mathbf{C} = \mathbf{\Gamma}^{-1} - \mathbf{\Gamma}^{-1}\mathbf{\Delta}(\mathbf{\Delta}'\mathbf{\Gamma}^{-1}\mathbf{\Delta})^{-1}\mathbf{\Delta}'\mathbf{\Gamma}^{-1}, \quad [7]$$

where $\hat{\epsilon} = \mathbf{s} - \sigma(\hat{\theta})$, $\mathbf{\Gamma}$ is the asymptotic covariance matrix of the residual covariances, and $\mathbf{\Delta}$ is a matrix of derivatives of the population covariances with respect to the model parameters. $\mathbf{\Gamma}$ may be computed under normality assumptions or under ADF assumptions. This statistic is also asymptotically distributed as a chi-square with $t-q$ degrees of freedom.

Unfortunately, when the data are not normally distributed, it has been repeatedly found in simulation studies that the p -values of Browne's statistic are inaccurate unless the number of variables is small and the sample size is very large. Recently, Yuan and Bentler (1997) have proposed a modification of T_B (with the same asymptotic distribution as T_B) whose p -values are more accurate in small samples and nonnormal data than those for T_B . This is

$$T_{YB} = \frac{T_B}{1 + NT_B/(N-1)^2}. \quad [8]$$

GOF Statistics for Piecewise Assessment of Fit

Z -scores for residuals are used for piecewise assessment of fit. The z -score for the residual covariance between variables i and j is

$$\frac{s_{ij} - \hat{\sigma}_{ij}}{\text{SE}(s_{ij} - \hat{\sigma}_{ij})}, \quad [9]$$

where $\hat{\sigma}_{ij} = \sigma_{ij}(\hat{\theta})$. Browne's and Yuan and Bentler's statistics eqns. [7] and [8] are, in fact, a test based on the joint set of residuals eqn [9].

GOF Indices

For continuous data, the AIC and BIC criteria used for model selection are GOF indices. When ML is not used, then the term $-2L$ in eqn [3] is simply replaced by T , the minimum of the estimated fit function multiplied by sample size, that is

$$\text{AIC} = T + 2q, \text{ BIC} = T + q \ln(N). \quad [10]$$

The AIC and BIC indices can be computed for any estimator, as no p -value is computed.

In addition, literally dozens of GOF indices have been proposed. Some may be used to assess the overall fit of the model under consideration, whereas others assess the relative fit of the model. A GOF index that may be used to assess the overall fit of a model is the standardized root mean residual (SRMR),

$$\text{SRMR} = \sqrt{\sum_j \sum_{k < j} \left(\frac{s_{jk}}{\sqrt{s_{jj}}\sqrt{s_{kk}}} - \frac{\hat{\sigma}_{jk}}{\sqrt{\hat{\sigma}_{jj}}\sqrt{\hat{\sigma}_{kk}}} \right)^2} / t. \quad [11]$$

The SRMR may be used to assess the average magnitude of the discrepancies between observed and expected covariances in a correlation metric. Note that there exist slightly different versions of this statistic.

Among GOF indices for relative fit assessment, two popular indices are the Tucker–Lewis index (TLI) and the comparative fit index (CFI), where

$$\text{TLI} = \frac{\frac{T_0}{\text{df}_0} - \frac{T_1}{\text{df}_1}}{\frac{T_0}{\text{df}_0} - 1}, \quad [12]$$

$$\text{CFI} = \frac{(T_0 - \text{df}_0) - (T_1 - \text{df}_1)}{T_0 - \text{df}_0}. \quad [13]$$

Here, M_0 is more restrictive than M_1 , the baseline model. Of the two, CFI is normed to lie between 0 and 1, whereas TLI is approximately normed. Almost invariably, they are used to compare the fit of the fitted model against a model that assumes that variables are uncorrelated. When used in this fashion, TLI and CFI values are very large. When comparing a set of theoretically driven models, it may be more interesting to use as a baseline the simplest theoretically driven model under consideration rather than the substantively uninteresting independence model. When used in this fashion, these statistics express in some sort of percentage how much is gained by each of the models under consideration relative to the most parsimonious model. Note that when using the TLI and CFI indices, M_0 need not be a special case of M_1 .

Discussion

GOF assessment necessarily involves subjective judgment. Models are just approximations to real-life phenomena. Consequently, any model will be rejected if the sample size is sufficiently large. This should not be taken to imply that GOF testing is meaningless. Rather, it is our view that researchers should always assess the overall GOF of their models using a GOF statistic to assess the magnitude of the discrepancy between the data and the model taking into account sampling variability.

If the selected model fits well, researchers should then:

1. Assess the power of the statistic against meaningful deviations from the selected model, as it may well be that the statistic has no power to distinguish between the selected model and substantively meaningful alternative models.
2. Perform a piecewise assessment of the selected model to examine if, although the model fits well overall,

some parts of the data are not well captured by the model.

3. Consider whether models that cannot be distinguished empirically from the selected model (i.e., equivalent models—see MacCallum *et al.*, 1993), exist.

Researchers should always report substantively interesting models equivalent to their selected model when they are aware of them and argue their choice using substantive arguments, since by definition, a choice between two equivalent models can only be made on substantive, not on empirical grounds.

On the other hand, if the model does not fit well, researchers should:

1. Perform a piecewise assessment of the model attempting to determine the source of the misfit. This may be aided by modification indices (i.e., Lagrange multiplier tests), but see MacCallum *et al.* (1992).
2. Assess the magnitude of the discrepancy between the fitted and expected statistics. For covariance structure analysis, the SRMR provided may be used to examine the average magnitude of the discrepancy. In addition, the magnitude of each standardized discrepancy, that is,

$$\frac{s_{jk}}{\sqrt{s_{jj}\sqrt{s_{kk}}}} - \frac{\hat{\sigma}_{jk}}{\sqrt{\hat{\sigma}_{jj}\sqrt{\hat{\sigma}_{kk}}}}, \quad [14]$$

should be inspected.

In any case, when selecting a model among competing alternatives, they should strive for model parsimony. To this end, they may inspect GOF indices that penalize adding parameters by, such as the AIC or BIC indices to the model. Alternatively, they may compute the CFI index, using as a baseline, the most parsimonious substantively meaningful model considered.

Finally, when many variables are modeled, it is unrealistic to expect that any parsimonious model will fit well: the overall test statistics will be large, because some parts of the data will not be well reproduced by any parsimonious model. In this context, Browne and Cudeck's (1993) proposal of assessing whether a model fits the data closely, is an attractive alternative to assessing whether the model fits exactly. Consider the root mean square error of approximation (RMSEA) index,

$$\text{RMSEA} = \sqrt{\max\left(\frac{T - \text{df}}{N \times \text{df}}, 0\right)}. \quad [15]$$

Like other indices, the RMSEA penalizes models with too many parameters. Unlike the AIC or BIC criteria, the RSMEA is bounded below by 0. Furthermore, Browne and Cudeck derived its asymptotic distribution. Testing $H_0 : \text{RMSEA} = 0$ is equivalent to testing $H_0 : \boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\theta})$. For large models, this null hypothesis of exact fit may be

too stringent, and Browne and Cudeck suggested testing instead, $H_0 : \text{RMSEA} \leq 0.05$. This is a test of close fit, where close is arbitrarily defined as $\text{RMSEA} \leq 0.05$. Moreover, model selection may be aided by inspecting the confidence intervals around their respective RMSEAs (Steiger, 2007).

Numerical Examples

Discrete Data

We fitted a two-parameter logistic model with a normally distributed latent trait to the Law school admission test (LSAT) 7 data. A thousand observations are available on five binary variables. The model was estimated by ML. **Table 1** gives some relevant GOF statistics. As these data are not sparse, the p -values for χ^2 and G^2 are accurate, and they are similar to that of the new test M_2 , which yields accurate p -values even for highly sparse data.

Using a 5%-significance level, χ^2 and G^2 suggest that the model be barely accepted, whereas M_2 suggests that the model be barely rejected. Statistics with a higher value to degrees-of-freedom ratio are generally more powerful, and the results of **Table 1** suggest that M_2 has higher power.

The cell residuals are not very helpful for piecewise assessment of fit. Significant residuals (at the 5% level) are obtained for patterns (10000), (01001), and (01000), which might suggest that the fit of the model could improve by dropping Item 2. The inspection of univariate and bivariate residuals offers a different picture. Significant residuals are obtained for (1,3), (1,5), (1,4), and (2,3), which suggest that the model misfits because of item 1. This is indeed the case as reflected in **Table 2**, where we provide Pearson's χ^2 statistics after dropping one item at a time.

Continuous Data

We model the responses of 438 US respondents to the five items of the satisfaction with life scale (SWLS). The items

Table 1 Goodness-of-fit (GOF) results for the LSAT 7 data

Stat	Value	df	p-value	Value/df
χ^2	32.48	21	0.05	1.55
G^2	31.70	21	0.06	1.51
M_2	11.94	5	0.04	2.39

Table 2 GOF results for the LSAT 7 data dropping one item at a time, 7 df

Item dropped	1	2	3	4	5
χ^2	5.01	9.52	8.59	18.68	9.86
p-value	0.66	0.22	0.22	0.01	0.20

Table 3 GOF results for the SWLS data

NT				ADF			
Stat	Value	df	p	Stat	Value	df	p
T	10.38	3	0.02	T_B	8.72	3	0.03
T_B	10.38	3	0.02	T_{YB}	8.55	3	0.04
				T_S	7.17	3	0.04
				T_a	6.90	2.89	0.07

are rating scales with seven response alternatives and will be treated as continuous. A two-factor model where “In most ways my life is close to my ideal,” “The conditions of my life are excellent,” and “I am satisfied with my life,” are taken as indicators of the factor satisfaction with present life, the and “I am satisfied with my life,” “So far I have gotten the important things I want in life,” and “If I could live my life over, I would change almost nothing,” are taken as indicators of the factor of satisfaction with past life. The factors are correlated. **Table 3** lists a number of GOF statistics obtained using the maximum-likelihood fitting function either under normality assumptions (NT) or under asymptotically distribution-free (ADF) assumptions.

Under NT, $T = N\hat{\chi}^2$, the minimum of the ML fit function multiplied by the sample size is provided, along with Browne's test (eqn [7], T_B). Under ADF assumptions, **Table 3** lists Browne's test, Yuan and Bentler's test (eqn [8], T_{YB}), and the Satorra-Bentler mean and mean and variance adjustments to T , T_S , and T_a , respectively. Notice that for T_a , the degrees of freedom are estimated as real numbers.

The model does not fit very well, and there is not much difference between the results under NT or ADF assumptions. All statistics yield similar p -values. Inspection of the z -scores for the residual covariances reveals significant residual covariances among items from different factors. The magnitude of the residuals is not large, however. The average standardized residual is $SRMSR = 0.015$, and the largest standardized residual is 0.04. It appears that the model yields a close enough fit. Indeed, the RMSEA (eqn. [15]) obtained is 0.056, and the p -value for testing whether the population RMSEA is smaller than 0.05, is 0.35.

Concluding Remarks

Our presentation has focused on models where the data can be summarized using some statistics (proportions, covariances, etc.). The fit of many interesting models cannot be assessed using summary statistics. For instance, the fit of a linear regression model cannot be assessed using covariances (i.e., within a SEM framework) because there are zero degrees of freedom. In the context of linear

regression and related models, R^2 is sometimes described as a GOF statistic. However, R^2 is actually a coefficient of determination, the proportion of the dependent variable that can be predicted from the independent variables. A linear regression model can fit the data perfectly, yet, R^2 will be zero if the slope is zero. Pure GOF statistics exist in regression and related models only in the presence of replicates (i.e., repeated observations for the same level of the predictors). In the general linear model, they are generally referred to as lack-of-fit tests. When no replicates exist, then, the observations must be grouped in some way to assess the GOF of the model. A typical example is the Hosmer–Lemeshow GOF statistic for logistic regression (see Hosmer and Lemeshow, 2000).

Clearly, GOF assessment has been more extensively developed in SEM than in other areas. New developments in GOF assessment for multivariate discrete data are strongly related to SEM procedures, and we expect further developments in GOF assessment procedures for multivariate discrete data along the lines of SEM developments. For instance, the notion of testing whether a model fits closely the data (as opposed to exactly), is yet to be brought into the multivariate discrete arena. More research is also needed on GOF assessment when the observed dependent variables are of mixed type (continuous and categorical). Finally, further research is needed on GOF-assessment procedures when data arises from complex sampling schemes, such as those found in multilevel modeling.

In closing, model fit (i.e., absolute GOF) is no guarantee of a model's usefulness. A model may reproduce the data at hand well and yet be useless for the purpose it was developed. On the other hand, a model may fit poorly and yet yield useful predictions. In this context, the fact that a model fits poorly simply means that in principle, a model could be found to reproduce the data better, whose predictions could be very different.

See also: Educational Data Modeling; Item Response Theory; Latent Class Models; Model Selection; Multivariate Linear Regression; Structural Equation Models.

Bibliography

- Browne, M. W. (1982). Covariance structures. In Hawkins, D. M. (ed.) *Topics in Applied Multivariate Analysis*, pp 72–141. Cambridge: Cambridge University Press.
- Browne, M. W. and Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K. A. and Long, J. S. (eds.) *Testing Structural Equation Models*, pp 136–162. Newbury Park, CA: Sage.
- Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression*. New York: Wiley.
- MacCallum, R. C., Roznowski, M., and Necowitz, L. B. (1992). Model modification in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin* **111**, 490–504.
- MacCallum, R. C., Wegener, D. T., Uchino, B. N., and Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin* **114**, 185–199.

- Maydeu-olivares, A. and Cai, L. (2006). A cautionary note on using G^2 (dif) to assess relative model fit in categorical data analysis. *Multivariate Behavioral Research* **41**, 55–64.
- Maydeu-olivares, A. and Joe, H. (2005). Limited and full information estimation and goodness-of-fit testing in 2^I contingency tables: A unified framework. *Journal of the American Statistical Association* **100**, 1009–1020.
- Maydeu-olivares, A. and Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika* **71**, 713–732.
- Satorra, A. and Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In von Eye, A. and Clogg, C. C. (eds.) *Latent Variable Analysis: Applications to Developmental Research*, pp 399–419. Thousand Oaks, CA: Sage.
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modelling. *Personality and Individual Differences* **42**, 893–898.
- Tollenaar, N. and Mooijart, A. (2003). Type I errors and power of the parametric goodness-of-fit test. Full and limited information. *British Journal of Mathematical and Statistical Psychology* **56**, 271–288.
- Yuan, K.-H. and Bentler, P. M. (1997). Mean and covariance structure analysis: Theoretical and practical improvements. *Journal of the American Statistical Association* **92**, 767–774.
- Yuan, K.-H. and Bentler, P. M. (2004). On chi-square difference and z tests in mean and covariance structure analysis when the base model is misspecified. *Educational and Psychological Measurement* **64**, 737–757.
- Bollen, K. A. and Stine, R. (1992). Bootstrapping goodness of fit measures in structural equation models. *Sociological Methods and Research* **21**, 205–229.
- Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology* **44**, 108–132.
- Browne, M. W., MacCallum, R. C., Kim, C.-T., Anderson, B., and Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods* **7**, 403–421.
- Hu, L.-T. and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* **6**, 1–55.
- Joe, H. and Maydeu-Olivares, A. (2006). On the asymptotic distribution of Pearson's X^2 in cross-validation samples. *Psychometrika* **71**, 587–592.
- MacCallum, R. C., Browne, M. W., and Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods* **11**, 19–35.
- MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods* **1**, 130–149.
- Mavridis, D., Moustaki, I., and Knott, M. (2007). Goodness-of-fit measures for latent variable models for binary data. In Lee, S.-Y. (ed.) *Handbook of Latent Variables and Related Models*, pp 135–162. Amsterdam: Elsevier.
- McDonald, R. P. and Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods* **7**, 64–82.
- Satorra, A. and Saris, W. E. (1993). *Power Evaluations in Structural Equation Models*, pp 181–204. Newberry Park, CA: Sage.
- Swaminathan, H., Hambleton, R. K., and Rogers, H. J. (2007). Assessing the fit of item response models. In Rao, C. R. and Sinharay, S. (eds.) *Handbook of Statistics, Vol. 26: Psychometrics*, pp 683–718. Amsterdam: Elsevier.
- Yuan, K. H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research* **40**, 115–148.

Further Reading

- Bartholomew, D. J. and Tzamourani, P. (1999). The goodness of fit of latent trait models in attitude measurement. *Sociological Methods and Research* **27**, 525–546.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin* **107**, 238–246.

Graphical Models

R G Almond, ETS, Princeton, NJ, USA

© 2010 Elsevier Ltd. All rights reserved.

Motivating Example

The idea of representing a complicated statistical model with a graph—a diagram in which icons are used to represent the variables and edges used to represent the relationship between variables—dates back to Wright's pioneering work on path analysis. In 1980, Darroch, Lauritzen, and Speed noted a class of models for contingency tables that had factorization structures and independence properties that could be described with an undirected graph. They called these models graphical models and noted several properties that made them both computationally convenient and readily interpretable. The following example illustrates some of the power of the method.

Table 1 presents the covariance matrix for the scores in five mathematics exams for 88 students. Zero covariances would indicate that scales represented by the five exams are independent. There is no evidence of independence, which is unsurprising as there is likely to be some kind of general mathematical ability factor that will produce correlation among the five scores.

A key insight of graphical models is that it is not the marginal independence of two variables (i.e., their correlation after averaging over the other variables), but their conditional independence (their correlation after conditioning on the other variables in the model) that is important. Dividing by the variance (the main diagonal) produces the correlation matrix. Inverting the correlation matrix and scaling the main diagonal entries to all be 1 produces **Table 2**. In this table, the values which are clearly nonzero have been shaded and the values close to zero are left unshaded. This reveals some conditional independence properties: for example, mechanics is independent of statistics given the other variables.

Figure 1 shows a graphical representation of the partial correlation matrix. In this graph, the nodes (drawn as rounded rectangles with a label) represent the variables in the model. Every place where there is a nonzero (shaded) entry in **Table 2**, there is an edge between the nodes. Note that algebra separates mechanics and vectors from analysis and statistics. This intuition is backed up with a conditional probability relationship that is present in the underlying distribution: mechanics and vectors are conditionally independent of analysis and statistics given algebra. This suggests several possible substantive interpretations, in particular, that algebra is at the intersection of the five different subareas within mathematics.

Graph Theory

A graph $G = \langle \nu, \varepsilon \rangle$ has two sets: the first set, ν , is a set of nodes or vertices (which usually correspond to variables in the model), and the second set, ε , consists of pairs of elements of ν called edges or arcs. Graphs can be either directed or undirected, depending on whether the edge (i, j) is considered to be an ordered or unordered pair. In the directed case, node i is called the parent and node j is called the child and the edge is depicted as an arrow from node i to node j . (The terms ancestor and descendent are also used with directed graphs as the obvious extensions of parent and child.) An undirected edge is equivalent to having both (i, j) and (j, i) as directed edges, and undirected edges are usually depicted as lines with no arrowheads. In either case, if the edge (i, j) is in the graph, then node i and node j are neighbors.

A graph can also be represented by an adjacency matrix: a matrix whose rows and columns correspond to the nodes of the graph and where the entry in cell (i, j) is 1 if node i and node j are neighbors, and zero otherwise. For example, **Table 2** (taking the nonzero elements to be one) is the adjacency matrix of the graph in **Figure 1**. Note that the adjacency matrix of an undirected graph must be symmetric (this is not an issue as the partial correlation matrix must also be symmetric).

A path is a sequence of nodes, X_1, X_2, \dots, X_K for which (X_k, X_{k+1}) is an edge in the graph; for a directed graph, this implies that the path follows the direction of the arrows. A path for which X_1 and X_K are the same is called a cycle, and a graph with no cycles is called acyclic; an acyclic undirected graph is also known as a tree. A set of nodes S separates two nodes X and Y if every path from X to Y or from Y to X passes through a member of S . As shown below, it is this notion of separation that gives graphical models their expressive power because it corresponds to conditional independence statements in the joint probability distribution.

A set of nodes which are all neighbors of each other is said to be complete. A maximal complete set of nodes is called a clique. In **Figure 1**, the set {mechanics, vectors} is complete, but not a clique, as it is contained by {mechanics, vectors, algebra}, which is a clique. An undirected graph can be represented by its set of cliques. For example, $[MeVeAl][AlAnSt]$ are the cliques of **Figure 1** (where each variable is represented by the first two letters of its name). As shown in the next section, the cliques of a

Table 1 Covariance matrix for math grade example (Whittaker, 1990)

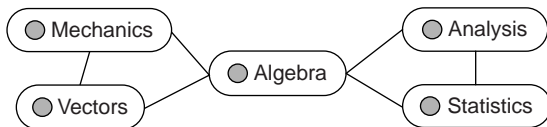
	<i>Mechanics</i>	<i>Vectors</i>	<i>Algebra</i>	<i>Analysis</i>	<i>Statistics</i>
Mechanics	302.29	125.78	100.43	105.07	116.07
Vectors	125.78	170.88	84.19	93.60	97.89
Algebra	100.43	84.19	111.60	110.84	120.49
Analysis	105.07	93.60	110.84	217.88	153.77
Statistics	116.07	97.89	120.49	153.77	294.37

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.

Table 2 Partial correlation matrix for math grade example (adapted from Whittaker, 1990)

	<i>Mechanics</i>	<i>Vectors</i>	<i>Algebra</i>	<i>Analysis</i>	<i>Statistics</i>
Mechanics	1	−0.33	−0.23	0.00	−0.02
Vectors	−0.33	1	−0.28	−0.08	−0.02
Algebra	−0.23	−0.28	1	−0.43	−0.36
Analysis	0.00	−0.08	−0.43	1	−0.25
Statistics	−0.02	−0.02	−0.36	−0.25	1

Adapted from Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Chichester: Wiley.

**Figure 1** Graphical models for five math test scores. Adapted from Whittaker (1990).

graph are important because they correspond to factors in the joint distribution.

Consider any cycle of length greater than three, say X_1 , X_2 , X_3 , and X_4 . An edge between two nonconsecutive nodes in the cycle (say X_2 and X_4) is called a chord. A graph that has no cycles of length greater than 3 without a chord is called triangulated. Triangulated graphs are important because they are particularly simple to work with computationally. For this reason, edges are often filled in to make a nontriangulated graph triangulated. Often, the cliques of the triangulated graph are arranged in a tree (in which the nodes correspond to sets of variables) is such a way that the subset of nodes in which each variable appears is a subtree of the overall structure. This transformation of the original graph is called a tree of cliques, or, if new nodes related to the intersection between variables are added, a junction tree. Generally, the computational complexity of working with a given graphical model is related to the size of the largest clique (after triangulation if necessary). This is called the tree-width. Choosing a fill-in that produces an optimal tree of cliques (i.e., one with the lowest treewidth) is a difficult problem; however, in many cases, simple search heuristics yield reasonable results.

Factorization and Conditional Independence

A graphical model is a joint probability distribution over a collection of variables that can be factored according to the cliques of an undirected graph. Let $\mathcal{G} = \langle \nu, \varepsilon \rangle$ be a graph whose nodes correspond to the variables in the model, and let C be the set of cliques in the graph. Let \mathbf{v} be an instantiation of the values in ν and let \mathbf{v}_C be the corresponding set of values for the variables in $C \subset \nu$. Then we can write the joint probability distribution as:

$$p(\mathbf{v}) = \prod_{C \in \mathcal{C}} \phi_C(\mathbf{v}_C) \quad [1]$$

In this equation, $p(\mathbf{v})$ should be interpreted as a probability mass function if all of the variables in ν are discrete, and a density function if any of the variables are continuous. The factors, $\phi_C(\cdot)$ are called potentials (sometimes Gibbs potentials; the term comes from statistical physics). The potentials can be probability distributions, conditional probability distributions, or products of any of the above. Generally, they are not interpreted directly and often they are left unnormalized.

The original use of the term graphical model was to distinguish a subset of the number of possible log-linear models for contingency tables that could be represented as a graph. Consider the set of possible models over three discrete variables A , B , and C . The models $[A][B][C]$, $[A][BC]$, $[AB][BC]$, and $[ABC]$ (and ones that are equivalent after relabeling the variables) are all graphical – the factors correspond to the cliques. The no-three-way interaction model $[AB][BC][CA]$ is not graphical because

the clique should be $[ABC]$ (the three factors do not correspond to cliques). Difference in the substantive interpretation of the no-three-way interaction model from the three-way interaction model is quite subtle; thus, restricting consideration to graphical models does not greatly limit the interpretability. Although the no-three-way interaction model has fewer parameters, it also is more difficult to work with, requiring iterative solutions to fit the model to data.

Restricting the variables to have a multivariate normal distribution creates a set of models corresponding to covariance selection models. (Note that for normally distributed data, the pairwise correlations define the joint distribution, so all models can be expressed with a graph, although it might be the saturated graph in which all nodes are connected.) In fact, there are three classes of graphical models for which most computations can be done exactly: (1) models in which all variables are discrete (graphical log-linear models), (2) multivariate normal distributions, and (3) finite mixtures of multivariate normal distributions. The graphical notation can be used to express other kinds of models, but operations on those kinds of models frequently involves integrals that must be solved numerically. (For example, the standard unidimensional item response theory model can be drawn as a graphical model with the observable outcome variables conditionally independent given the latent ability variable. However, the model is not a conditional Gaussian distribution; therefore estimating examinee ability using this model requires numeric integration.)

The Markov property of a graphical model states that if X and Y are variables in the graph, and S is a set that separates X from Y in the graph, then X is conditionally independent of Y given S , sometimes written $X \perp\!\!\!\perp Y | S$. This Markov property drives a number of efficient algorithms for computation, as well as aiding in the interpretation of the model. Probability distributions with this property are sometimes called Markov random fields.

Consider the well-known example of Simpson's paradox in which a university has several departments, each of which admits a different proportion of their applicants, and each of which has a different proportion of women in their applicant pool. Further assume that each department is completely gender neutral in their admissions decisions. The graph shown in **Figure 2** illustrates these assumptions. The nodes gender and acceptance are connected in the graph, so they are marginally dependent. Depending on whether women prefer more- or less-selective departments, the university will appear biased against or toward women. The graph tells the story visually:

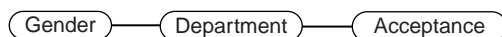


Figure 2 A Graphical model illustrating Simpson's paradox.

it is the preference of women for more (or less) selective departments that causes the marginal dependency.

A fundamental result in the theory of graphical models is that factorization (according to a graph) implies the Markov property (with respect to that graph) and vice versa. This is sometimes known as the Gibbs–Markov equivalence theorem. Going from the factorization to the Markov property requires no special assumptions, but going from the conditional independence statements to the factorization requires that the probability for all configurations of the variables be positive. This can present a problem if there is a deterministic constraint between two or more variables; however, in practical circumstances, it can usually be made to work.

The Gibbs–Markov equivalence property has particular salience when building graphical models from expert opinion. The usual practice is to first elicit the conditional independence properties, then specify the probabilities for the factors defined by the cliques of the graph.

Related Graphical Representations

Although this article describes graphical models used in the narrow sense (models on an undirected graph), the term is often used to describe any model that is factored according to a graph. Two other classes of models, Bayesian networks and chain graph models, use directed or semidirected graphs to expand the expressive power of the graphical models. Path diagrams, while related in intent, have slightly different rules for interpretation.

A Bayesian network is a probability model defined over an acyclic directed graph. It is factored by using one conditional probability distribution for each variable in the model, whose distribution is given conditional on its parents in the graph. Variables which are separated in the graph are still independent, but the simple graph separation used for the undirected graph is replaced with the more complicated d -separation, which takes into account the effect of competing explanations for observed values.

A Bayesian network can be converted into an undirected graphical model by connecting all of the nodes that are involved in each factor. This requires that the parents of each node be joined or married. The process of joining parents is known as moralization and the undirected graph corresponding to a given Bayesian network is called a moral graph. Computing the moral graph is the first step of many computational algorithms for Bayesian networks.

Chain graphs use a mixture of directed and undirected graphical edges to describe independence relationships among the variables. A chain graph is divided into a series of blocks; within a block, all edges are undirected, and between blocks, the edges are directed. The blocks often correspond to stages in an experiment; for example, one

block could be measurements taken before the start of an educational intervention, and another measurements taken after. A third block could represent background variables about the students and schools. The directed edges flow in the temporal direction: from preintervention to postintervention measures and from background variables to either.

Path diagrams bear an obvious similarity to graphical models (especially the directed Bayesian networks) and, indeed, the work on path diagrams was an inspiration for the later work on graphical models. However, there are some subtle and important differences. Some are obvious: path diagrams include explicit nodes for error terms, while these are usually implicit in graphical models and Bayesian networks; and path diagrams allow some kinds of expressions (double-headed arrows and reciprocal relationships) that are not allowed in graphical models. Other differences are more subtle. In particular, the discipline of structural equation modeling concentrates on modeling the covariance matrix, while graphical models concentrate on modeling the inverse covariance matrix. The implication is that the conditional independence assumptions implicit in the structural equation model are not always clearly expressed; hence, the Markov property may not hold for the path diagram. Most path diagrams have an equivalent representation as a graphical model, but there are some exceptions (which typically are models that are difficult to estimate).

Model Fitting, Model Selection, and Analysis of Deviance

The parameters of Gaussian (normal), log-linear, and conditional Gaussian graphical models can be estimated through maximum likelihood estimation. The difficulty of this estimation procedure depends on the topology of the graph. If the graph is triangulated, then the parameters can be estimated separately for each clique. In Gaussian models, the key parameters are based on the covariance matrix of the variables in the clique. For the log-linear models, the model for the clique is just the saturated log-linear model, whose sufficient statistics are just the observed cell counts. Note that this procedure counts the variables in the intersections between the cliques twice; therefore, to avoid this double counting, then divides by the distribution over the variables in the intersection between the cliques. Thus, the final model looks like:

$$p(v) = \frac{\prod_{C \in \mathcal{C}} \phi_C(v_C)}{\prod_{I \in \mathcal{I}} \psi_I(v_I)} \quad [2]$$

where \mathcal{C} is the set of cliques of the graph, \mathcal{I} is the set of intersections; $\phi_C(v_C)$ is the marginal distribution over the

clique \mathcal{C} , and $\psi_I(v_I)$ is the marginal distribution over the intersection \mathcal{I} . This is a fundamental representation of graphical models that is used in many computation algorithms.

If the model graph is not triangulated, then iterative algorithms are required to estimate the model parameters. The most common estimation uses iterative proportional scaling. If prior distributions are specified for model parameters, it is possible to estimate the parameters using Markov chain Monte Carlo estimation.

Many common software packages can be used to fit graphical models to data. The key to specifying a graphical model is noting that the cliques of the graph are specified as multiway interactions using the Wilkinson and Rogers model specification notation, which is used by many common statistics packages such as SAS and S. Using the S `glm` function, the model in **Figure 2** would be written as `'Count ~ Gender*Department + Department*Acceptance'` and the saturated model would be written as `'Count ~ Gender*Department*Acceptance'`. Special graphical model software packages, such as MIM or CoCo, allow models to be specified in the clique notation.

It is usually possible to fit more than one model to a given data set. The saturated model (the one with all possible edges) should always fit, but there may be more parsimonious models that fit as well. A general measure of model fit for graphical models is the deviance which is defined to be twice the difference in log likelihood between the current model and the saturated model. This has an asymptotic χ^2 distribution, with degrees of freedom equal to the difference between the number of parameters in the candidate model and the number of parameters in the saturated model. For Gaussian models, the number of parameters is simply the number of edges. For discrete models, the number of parameters is the sum of the sizes of the cliques, where the size of the clique is defined as the product of the number of states of each of the variables. The size of the saturated model is thus exponential in the number of variables in the model.

Suppose that a researcher is comparing two different candidate models: model A and model B. If model B's graph can be produced from model A's graph by removing edges, model B is nested within Model A. The difference in their deviances is a likelihood ratio test comparing the two models (with the degrees of freedom given by the difference in the number of parameters). If there are a series of nested models (say by successively removing a series of edges), then these can be arranged in a table that is similar to an analysis of variance table. This is sometimes called the analysis of deviance.

Note when the data are sparse with respect to the number of variables (i.e., when there are a large number of empty cells in the contingency table, or because certain

combinations of variables are rare), then the parameter of the saturated model may be difficult to estimate and the sample deviance may have poor approximation to the χ^2 distribution. In such cases, the likelihood ratio test may work well even if the difference of the deviances is not a numerically stable way to compare models.

Stepwise procedures are frequently used to search for the best model. These can either start from the saturated model (all edges present) and proceed to remove edges, or from the complete independence model (no edges present) and proceed to add edges. In the case of log-linear models, model selection through adding or removing edges restricts the search to the class of graphical models; however, the difference in interpretation between the graphical and nongraphical models compatible with the same graph is fairly subtle, and often not relevant for interpreting the results in a wider scientific context. Note that such a stepwise procedure is not guaranteed to converge to the optimum (in the sense of best-fitting the observed data) model.

If the models do not nest, then the usual penalized likelihood measures (e.g., Akaike's information and Bayesian information) can be used to compare models. Searching the space of all possible models is an area of active research, and there are many algorithms proposed for this purpose (especially algorithms that search for directed and semidirected graphs). Most of the algorithms require all data to be at least partially observed (missing values are allowed), but some include latent variables in the search space.

Two cautions are worth making when considering the application of these model discovery algorithms. First, it is possible to over fit models to data, and hence it is important to use cross-validation to evaluate the model fit. Second, many of the model search techniques are motivated by attempts at causal discovery (generally, directed graphical models in which the edges are oriented in the causal direction will have fewer edges). However, the usual limitations of inferring causal relationships from observational data still apply. This is especially important if the work is to be presented to a lay audience that is not familiar with the definitions, assumptions, and limitations of the causality search procedures.

Summary

Graphical models are joint probability distributions that can be factored according to a graph. The key property of graphical models is that separation in the graph implies conditional independence of the variables. This conditional independence property allows computers to use the graphical structure to derive efficiency computation

algorithms. It also helps researchers visualize the properties of the joint distribution.

Bibliography

- Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.
Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. Chichester: Wiley.

Further Reading

- Almond, R. G. (1995). *Graphical Belief Modeling*. London: Chapman and Hall.
Badsberg, J. H. (2001). A guide to CoCo. *Journal of Statistical Software* 6(4). <http://www.jstatsoft.org/v06/i04> (accessed May 2009).
Berge, C. (1962). *A Theory of Graphs and Its Application*. London: Methuen.
Bishop, Y., Fienberg, S., and Holland, P. (1975). *Discrete Multivariate Analysis*. Cambridge, MA: MIT Press.
Cowell, R., Dawid, A., Lauritzen, S., and Spiegelhalter, D. (1999). *Probabilistic Networks and Expert Systems*. New York: Springer.
Cox, D. and Wermuth, N. (1996). *Multivariate Dependencies: Models, Analysis and Interpretation*. London: Chapman and Hall.
Darroch, J., Lauritzen, S., and Speed, T. (1980). Markov fields and log-linear interaction models for contingency tables. *Annals of Statistics* 8, 522–539.
Dempster, A. (1972). Covariance selection. *Biometrics* 28, 157–175.
Dethlefsen, C. and Højsgaard, S. (2005). A common platform for graphical models in R: The gRbase package. *Journal of Statistical Software* 14(17), 12. <http://www.jstatsoft.org/v14/i17> (accessed May 2009).
Edwards, D. (1990). Hierarchical interaction models. *Journal of the Royal Statistical Society, Series B* 52, 3–20.
Edwards, D. (1995). *Introduction to Graphical Modelling*. New York: Springer.
Elidan, G., Lotner, N., Friedman, N., and Koller, D. (2000). *Discovering hidden variables: A structure-based approach*. In NIPS, pp 479–485. Url <http://www-robotics.stanford.edu/~koller/Papers/Elidan+al:NIPS00.pdf> (accessed May 2009).
Glymour, C., Scheines, R., Spirtes, P., and Kelly, K. (2007). *Discovery of Causal Structure*. San Diego, CA: Academic Press.
Golumbic, M. C. (1980). *Algorithmic Graph Theory and Perfect Graphs*. New York: Academic Press.
Jensen, F. (1996). *An Introduction to Bayesian Networks*. New York: Springer.
Jordan, M. I. (ed.) (1998). *Learning in Graphical Models*. Dordrecht: Kluwer Academic Publishers MIT Press.
Lauritzen, S. (1996). *Graphical Models*. Oxford Science Publications.
Martin, J. and VanLehn, K. (1994). Discrete factor analysis: Learning hidden variables in bayesian networks. *Technical Report LRDC-ONR-94-1*. Pittsburgh, PA: LRDC, University of Pittsburgh.
Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika* 12, 341–369.
Moussouris, J. (1974). Gibbs and Markov random systems with constraints. *Journal of Statistical Physics* 10, 11–33.
Neapolitan, R. (2004). *Learning Bayesian Networks*. Upper Saddle River, NJ: Prentice Hall.
Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann.
Speed, T. and Kiiveri, H. (1986). Gaussian Markov distributions over finite graphs. *Annals of Statistics* 14(1), 138–150.
Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics* 5, 161–215.

Relevant Websites

<http://cran.r-project.org> – A number of packages for working with graphical models in R are available in the CRAN Library.

<http://www.hypergraph.dk> – David Edwards' software package MIM for fitting graphical models.

<http://www.cs.ubc.ca/~murphyk/Bayes/bnsoft.html> – Kevin Murphy maintains a list of software packages for graphical models and Bayesian network.

<http://www.r-project.org/gR> – The gR project puts a common interface on many different graphical modeling tools available as R packages.

Growth Modeling

I Plewis, University of Manchester, Manchester, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Growth curve – A graphical or algebraic representation of the change in a variable with age.

Hierarchical linear model – A statistical model for data measured at two or more levels that partitions and accounts for the variances and covariances at different levels, also known as a multilevel model.

Missing at random (MAR) – The condition wherein the data values that are missing depend only on the values that are observed and not, in addition, on the missing values themselves.

Multiple imputation – A method of adjusting for missing data that assumes that the data are missing at random and that takes account of all the information about missingness in the observed data.

Multivariate analysis of variance – A statistical model in which two or more response (or dependent) variables, usually assumed to follow a multivariate normal distribution, are related to one or more categorical explanatory variables.

Random effect – An effect that varies from unit to unit and where the interest is in the distribution (usually assumed to be normal) of this effect and, in particular, in its variance, rather than in the values for each unit.

Introduction

The growth of all animals is both systematic and variable. It is systematic in the sense that it is usually possible to model growth as a smooth function of age (or time) and variable in the sense that there are individual differences: not all members of a species grow at the same rate. Modeling growth is therefore a statistical as well as a biological problem, one that has interested statisticians for well over half a century. Much of the early impetus for the work in this area came from a desire to model weight growth in farm animals (e.g., Wishart, 1938) and height growth in children (e.g., Deming, 1957). Indeed, auxology – the science of human growth – has generated important advances in growth-curve modeling as discussed by, for example, Healy (1989). Like human biologists, since developmental psychologists too focus on changes with age, they also came to appreciate the potential of modeling

age changes in cognitive outcomes such as language development and processing speed. Some (e.g., Bayley, 1949) saw the development of mental abilities as being akin to growth in height. However, the application of growth curves to psychological variables did not proceed in an uninterrupted way. Concerns about measurement error and the comparability of measures over time led to moves away from age-related models – in other words, analyzing difference scores when there are just two repeated measures and fitting growth curves when there are more than two. Instead, conditional or regression models for change became more popular with the measure of a construct at baseline being used as an explanatory or control variable and interest then centering on the correlates of later measures for a fixed baseline. The paper by Cronbach and Furby (1970) was influential in this shift of emphasis, with the move back to growth curves initiated by Rogosa *et al.* (1982) and, particularly, by developments in statistical software which meant that flexible models of growth could be applied to the often messy data generated by longitudinal studies in the behavioral sciences.

Educational researchers also have much in common with developmental psychologists, not least because of their mutual interest in cognitive development and learning. Questions such as how quickly children learn to read and do mathematics, and how progress in subjects taught at school varies by, for example, gender and social class and also by school attended are important areas in educational research. In principle, the development or growth in a construct like mathematical ability can be modeled as a function of age. In practice, however, there are particular difficulties with measures of educational attainment because there are no educational tape measures; in other words, there are no fixed scales that apply across all ages. Instead, researchers use different and age-appropriate measures or tests for pupils at different stages of their school careers, but generating a single-objective scale that can be used across the age range is then problematic. Vertical equating has been used to try to get round this problem but with limited success and then only for rather narrow age ranges.

The absence of fixed scales for educational attainments renders the application of growth-curve models to educational data more difficult than it is to, say, height data but certainly not impossible. Although we are unlikely ever to be able to specify and estimate a single model for reading growth that represents underlying changes with age (e.g., rapid progress in the early years in school, reaching an

asymptote in late adolescence), we can attempt to answer an arguably more important set of questions about the correlates of growth. We can do this by constructing scales that make different assumptions about growth and then testing the sensitivity of estimates of correlates of growth to these different assumptions about scale. This issue is discussed in Plewis (1996), and we return to it in the context of our example. This article sets out growth-curve models for one and more than one attainment variable that is measured (on a continuous scale) on several occasions. First, however, as our example is woven into the development of the models, the study from which the data are taken is introduced.

The Example Data

In our example of an application of growth-curve models to educational attainments, data from the 1970 British Cohort Study (BCS70) are used. This study is the third of what is now a series of four UK birth cohort studies covering cohorts of children born in 1946, 1958, 1970, and 2000/2001. Data were collected from the full cohort on up to five occasions in BCS70 before the age of 21: at birth and then at 5, 10, 16, and 21 years. In addition, a subsample of these children was measured at 22 and 42 months. Consequently, we have longitudinal records (records from the same children) that span the period from early childhood (at 22 months when learning is just beginning) to early adulthood (at age 21 years by which time formal learning has usually ended). Although the initial cohort consisted of over 17 000 births, we restrict ourselves to the subsample of children measured at 22 months ($n = 2457$). We consider two measures of attainment that we call literacy and numeracy while recognizing that the distinction between them is not easily drawn at very young ages. We focus on the relative progress of boys and girls on these two outcomes.

Table 1 gives the number of observations and the observed age range at each measurement occasion. It shows that there was a range of ages at each occasion and also a considerable amount of missing data. The missing

data at occasion 2 are mainly due to item nonresponse, to attrition at occasions 3 and 4, and to a teachers' strike at occasion 5. The sample at age 21 was a 10% subsample of the original cohort, and this resulted in only a small overlap with the 22-month data.

Table 2 gives the pairwise Pearson correlations across occasions for the two underlying variables (with literacy below the diagonal and numeracy above it). These correlations tend to be higher for occasions four, five, and six although the two correlations for occasions five and six are based on rather small samples. For literacy, the correlations decline as the interval between occasions widens, but this decline is not so apparent for numeracy.

We require a scale that covers the age range. There are a number of possibilities but not the space in this article to consider all of them. The approach adopted here is to use z -transformations so that, based on the observed scores, the scale has a mean of zero and a standard deviation of one at each age. This has the advantage of simplicity but does not, of course, incorporate any notion of growth, either in the mean or, importantly, in the variance. We return to scaling issues below.

Growth-Curve Models

Educational researchers are interested in how attainment changes with age and, especially, to what extent there is variation between individuals in their growth or change parameters. In addition, they often want to explain this interindividual variation in terms of both fixed and time-varying characteristics of these individuals and, sometimes, the characteristics of groups to which the individuals are attached, for example, schools and neighborhoods. Moreover, researchers are often interested in analyzing more than just one measure of attainment and want to know how sets of variables change together with age.

A multilevel (ML) or hierarchical linear modeling (HLM) approach to the analysis of these kinds of questions has many advantages. We regard the repeated measures of the outcome as nested within the sampled individuals to form a two-level structure as illustrated for a hypothetical design in **Figure 1**. Each sampled case could be measured on all five occasions as individuals one and two in **Figure 1** were. Both were measured every year for 4 years but then with an interval of 2 years between occasions 4 and 5. They were not, however, both measured at precisely the same ages: for example, at occasion 2, individual one was exactly 1 year old, whereas individual two was age 1.2 years. Individual three is measured just 3 times, missing occasions 2 and 4, whereas individual four is lost from this hypothetical study after occasion two. The specification of the ML models can, as we shall see, be flexibly adapted to a variety of research questions and to a range of data sets. In particular, the models easily handle

Table 1 Sample size and age range by occasion: BCS70

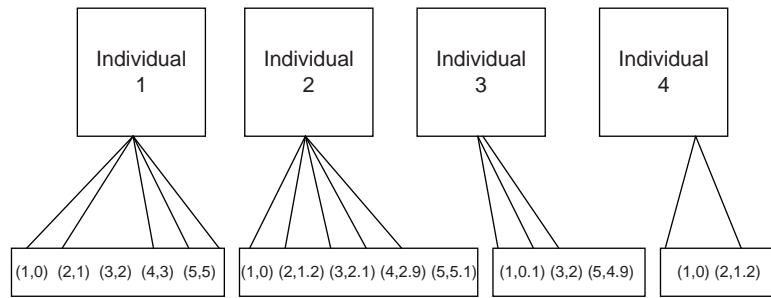
Occasion	Age range (years)	Observed sample sizes	
		Literacy	Numeracy
1	1.8–2.4	2412	2388
2	2.8–4.0	1317	1275
3	4.9–6.0	1636	1889
4	9.7–11.6	1744	1743
5	16.0–16.8	753	487
6	21 ^a	173	124

^aAge at test not collected.

Table 2 Pairwise correlations across occasions for literacy and numeracy and numbers of observations: BCS70^a

Occasion	Occasion					
	1	2	3	4	5	6
1		0.30 (1230)	0.22 (1841)	0.22 (1694)	0.13 (477)	0.20 (122)
2	0.26 (1298)		0.29 (1022)	0.44 (920)	0.43 (269)	0.17 (73)
3	0.26 (1609)	0.39 (882)		0.37 (1471)	0.31 (422)	0.20 (105)
4	0.25 (1718)	0.29 (947)	0.45 (1259)		0.63 (384)	0.55 (96)
5	0.20 (744)	0.16 (436)	0.20 (556)	0.63 (613)		0.64 (32)
6	0.14 (168)	0.05 (110)	0.12 (128)	0.50 (139)	0.57 (60)	

^aLiteracy below the diagonal, numeracy above, and numbers of observations in brackets.

**Figure 1** Hierarchical structure. Bottom boxes: (o,t) refers to occasion and age at measurement (in years).

the difficulties generated by (1) unevenly spaced measurement occasions (e.g., individuals one and two in **Figure 1**), (2) measurements missing for some occasions for all or some variables (individuals three and four), and (3) data collected across a range of ages within any one occasion (at occasion 1 for all four individuals). For many years, analysts used techniques based on multivariate analysis of variance (MANOVA) which could only handle regularly spaced and complete data, but these techniques are now redundant.

Linear Growth

Consider first, a simple model for linear (i.e., straight line) growth for a single continuous response (y) changing with age (a). We write:

$$y_{ij} = b_{0j} + b_{1j}a_{ij} + e_{ij} \quad [1]$$

Model [1] is a two-level regression model with occasions ($i=1..t_j$; level one) nested within individuals ($j=1..n$; level two) where, in **Figure 1**, $\max(t_j) = 5$ and $n = 4$. It is usual to measure age because deviations from its overall mean as the estimates from the model are more readily interpretable then. Here, we center age around 10 years. The model allows both the intercept (b_{0j}) and the growth rate (b_{1j}) to vary between individuals (j) and so we write:

$$b_{0j} = b_{00} + u_{0j} \quad [2]$$

and

$$b_{1j} = b_{10} + u_{1j} \quad [3]$$

where u_{0j} and u_{1j} are level-two random effects; they have zero means, a bivariate normal distribution, and are uncorrelated with the level-one residuals e_{ij} .

Models [2] and [3] are readily extended by including one (or more) explanatory variables z_j which vary across individuals (but not from age to age). In our case, this is gender with $z_j = 0$ for boys and 1 for girls so that we expect both mean differences between boys and girls (b_{01} : model [4]) and also differences in their growth rates (b_{11} : model (5)):

$$b_{0j} = b_{00} + b_{01}z_j + u_{0j} \quad [4]$$

$$b_{1j} = b_{10} + b_{11}z_j + u_{1j} \quad [5]$$

Substituting models [4] and [5] into model [1] gives us:

$$y_{ij} = b_{00} + b_{01}z_j + b_{10}a_{ij} + b_{11}a_{ij}z_j + u_{0j} + a_{ij}u_{1j} + e_{ij} \quad [6]$$

and the effect of gender on the slope or growth rate is represented by coefficient b_{11} for the cross-level interaction $a_{ij}z_j$. Model [6] (equivalent to models [1], [4], and [5] combined) can be fitted rather easily using a standard statistical package or with specialist ML software such as *MLwiN*.

As growth rates vary between individuals, the overall variance of the response y , conditional on (i.e., for fixed values of) the explanatory variables z , is a function of age.

This variance function is obtained from the last three terms – the random effects – in model [6]:

$$\text{var}(\mathcal{Y}_{ij}) = \sigma_{u0}^2 + a_{ij}^2 \sigma_{u1}^2 + 2a_{ij}\sigma_{u0,u1} + \sigma_e^2 \quad [7]$$

The term $\sigma_{u0,u1}$ in model [7] is the covariance between the intercept and the slope from which the corresponding correlation can be estimated. This correlation should, however, be interpreted with caution as it depends both on how age is centered and also on the underlying shape of the individual growth curves. We can extend the variance function in model [7] by allowing the within-person variance σ_e^2 to vary systematically with age rather than remain constant. We do this by allowing the effect of age to vary randomly at level one, replacing b_{1j} by b_{1ij} in model [1]. This enables us to deal with heteroscedasticity.

Nonlinear Growth

The model can be further extended to allow for changes with age that do not follow a straight line. Polynomial functions are flexible and commonly used:

$$\mathcal{Y}_{ij} = \sum_{q=0}^Q b_{qj} a_{ij}^q + e_{ij} \quad [8]$$

In model [8], Q is the order of the polynomial, determined from the data but with the restriction that Q is usually less and often much less than $\max(t_j)$. We now have $Q+1$ random variances and $Q(Q+1)/2$ covariances at level two.

The mean growth curve (i.e., the growth curve for the child that scores at the mean at each age) for our exemplar data from BCS70, measured on the selected (but arbitrary) scale of standardized scores at each age, is just a straight line through the origin with zero slope. Individuals' growth curves can, however, vary about this line as illustrated in **Figure 2**. Moreover, they can, in principle, be nonlinear even though the mean curve is a straight line. For example, some children might make particularly rapid progress in early childhood and then slow down, whereas others might start slowly and then have an adolescent growth spurt. We allow for possible nonlinearity by introducing a quadratic term in age (i.e., age squared) into the model for these data. Our main interest, however, is in the main effect of gender and the cross-level interactions of gender with age and age squared as these estimates of the fixed effects tell us how gender differences in the outcomes vary across the observed age range. **Table 3** gives the estimates, separately for literacy and numeracy, and **Figure 3** shows how gender differences vary by age for the two outcomes. The quadratic term dominates the gender differences for literacy and so girls are ahead of boys at very young ages and again in early adulthood but with only small differences during the

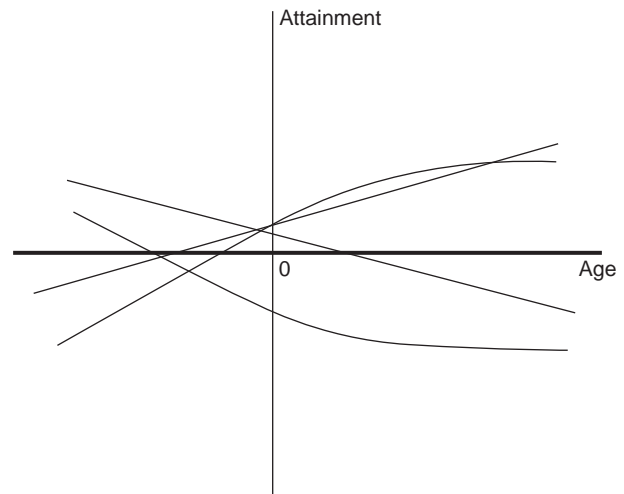


Figure 2 Individual growth curves.

Table 3 Gender estimates and standard errors: Main effect and age interactions from univariate models

<i>Fixed effects</i>	<i>Literacy</i>	<i>Numeracy</i>
Gender	*	0.10 (0.043)
Gender ^a age	*	-0.02 (0.005)
Gender ^a age ²	0.0029 (0.0007)	-0.0025 (0.0007)

^aEstimate less than its standard error.

period in school. The effect for numeracy shows girls slightly ahead of boys until they leave primary school (at age 11) but then falling behind.

Turning to the random effects shown in **Table 4**, we find that, for both outcomes, about 95% of the sample has linear growth rates that vary by about 0.10 units around the mean of zero and quadratic terms that vary by about 0.01 units around zero (rows 2 and 3 in **Table 4** give the variances and so we take the square root of these and multiply by ± 1.96). We also find that the within-person (or level-one) variance is not constant; instead it declines with age (because the covariance in line 9 of **Table 4** is negative), whereas the between-person (or level-two) variance increases with age. We must, however, interpret these changes in variance with age cautiously because the scales are constructed in such a way that the overall variance is constant.

Whenever we fit a statistical model to data, we must pay heed to the assumptions of the model. We can do this by (1) examining the distributions of the estimated residuals at each level (are they approximately normal?), (2) plotting intercept residuals against slope residuals to check for any outliers, and (3) plotting residuals against predicted values from the model and against the explanatory variables to check that, for instance, the assumption of roughly constant residual variance is reasonable.

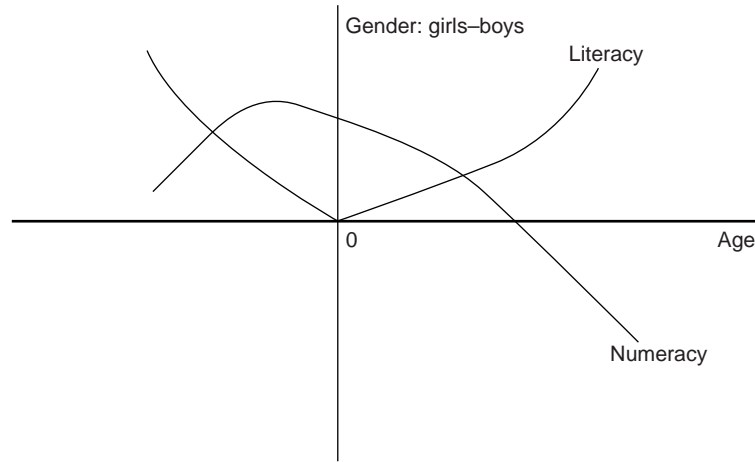


Figure 3 Gender differences by age: literacy and numeracy.

Table 4 Estimates and standard errors: Random effects from univariate models

Random effects	Literacy	Numeracy
L2: Intercept var.	0.63 (0.034)	0.58 (0.034)
L2: Age var.	0.0030 (0.00034)	0.0024 (0.00034)
L2: Age ² var.	0.000022 (0.0000080)	0.000010 (0.0000070)
L2: Intercept, age cov.	0.021 (0.0023)	0.024 (0.0025)
L2: Intercept, age ² cov.	-0.0030 (0.00044)	-0.0021 (0.00043)
L2: Age, age ² cov.	-0.000050 (0.000033)	-0.00010 (0.000032)
L1: Intercept var.	0.39 (0.021)	0.44 (0.022)
L1: Age var.	0.0019 (0.00066)	0.0011 (0.00068)
L1: Intercept, age cov.	-0.016 (0.0017)	-0.017 (0.0018)

L – level; var. – variance; cov. – covariance.

Introducing a Higher Level

If we are fortunate enough to be able to analyze a study designed in such a way that the repeatedly measured individuals are located in a sample of schools, then we can answer questions about differences between schools in growth rates. For simplicity, we assume that a straight line model for growth is adequate and we remain interested in gender differences. Model [1] is now written as:

$$y_{ijk} = b_{0jk} + b_{1jk}a_{ijk} + e_{ijk} \quad [9]$$

Model [9] is a three-level model with occasions ($i=1..t_{jk}$) nested within individuals ($j=1..n_k$) now nested within the third level, schools ($k=1..K$). If we allow each intercept, growth rate, gender difference, and gender difference in growth rates to vary from school to school then model [6], the random part especially, becomes more complex:

$$y_{ijk} = b_{000} + b_{010}z_{jk} + b_{100}a_{ijk} + b_{111}a_{ijk}z_{jk} + v_{00k} + z_{jk}v_{01k} + a_{ijk}v_{10k} + a_{ijk}z_{jk}v_{11k} + u_{0jk} + a_{ijk}u_{1jk} + e_{ijk} \quad [10]$$

There are four random effects at the school level in model [10]: the overall intercept (v_{00k}), the growth rate (v_{10k}), the gender difference (v_{01k}) and the difference between boys' and girls' growth rates (v_{11k}). These random effects generate four variances and six covariances at the school level to add to the two variances and one covariance at the individual level and one variance at the occasion level. We might be able to explain some of the variation at the school level if we have measures – the pupil–teacher ratio, for example – at the school level.

Multivariate Growth-Curve Models

Rather than estimating two separate models, we might exploit the fact that literacy and numeracy are related at each age (the Pearson correlations vary from 0.30 to 0.81) to obtain more efficient estimates from a bivariate model. In addition, we can learn more about how the growth parameters covary. As Plewis (2005) shows, the general multivariate growth curve model (ignoring the school level for simplicity) can be written as:

$$y_{bij} = \sum_{q=0}^{Q_b} b_{bjq}a_{ij}^q + e_{bij} \quad [11]$$

Model [11] is the same as model [8] except for the addition of a third subscript, b ($b=1..H$). Here, H is the number of response variables (two in the example) and we have a three-level model. There is no level-one variation now because level one exists solely to define the multivariate structure. Variation between occasions within individuals is at level two and variation between individuals at level three. Note that Q can vary with b . Not only

does the variance of each response vary with age as in model [7], but the covariances between the responses do as well.

This model will provide us with:

1. Patterns of change that can vary across variables both in terms of their relation with age and with the correlates of growth, z_j
2. Correlations between growth parameters – the b_{bjj} – both within and between variables.
3. Model-based variances, covariances, and correlations at the individual level (level three), both within and between variables.

In fact, in our example, the fixed effects of interest (shown in **Table 5**) are little altered from those of the univariate models. The estimated correlation between the linear growth rates for literacy and numeracy is 0.66 suggesting that individuals' attainments grow at similar rates.

Concluding Remarks

We have seen how to specify and apply growth-curve models to educational attainment data, and our example has illustrated some of the strengths of this approach. First, all the available data on literacy and numeracy have been used, and this has been especially important in the presence of so much missing data: a complete case analysis would have been based on just 22 cases for literacy and 14 cases for numeracy. Second, we have been able to take account of variability in age at measurement within occasion and to handle unequal intervals between occasions because the models use these two measures in different ways: occasion defines a level but age defines one of the explanatory variables. Third, we have been able to show how gender effects vary across the age range. Finally, we have gleaned some extra information by analyzing literacy and numeracy together in a bivariate model.

We can extend the models in a number of ways. Rather than using polynomials to model change with age, we could, for example, use other functions of age (log, reciprocal, etc.) or we could model growth for a series of stages (e.g., before school, primary school, high school) by using different linear models for different stages, known as piecewise linear models or spline functions. We could

also expand the number of explanatory variables. Moreover, we could also use our estimated growth curve to make predictions about adult status, for example, level of educational qualifications, to test whether mean attainment or growth in attainment is a better predictor of later qualifications.

Educational researchers wishing to use growth curve models to learn more about the development of educational inequalities with attainment data should, however, proceed with caution. Plewis (1996) shows that group differences can vary according to the scale chosen. A full analysis of these data requires sensitivity analyses to determine whether and to what extent the conclusions are affected by changes in the age function used to model growth, by changes in the way the outcome measures are constructed at each age, and by varying assumptions about the scale of the underlying or latent variable with age. We might, for example, have generated a literacy age scale with means increasing with age and also allowing variance to increase with age just as height variance does.

The focus in this article has been on continuous outcomes. Sometimes, however, attainments are measured as binary outcomes – whether or not someone has achieved a particular proficiency level – sometimes as counts and sometimes as ordered categories. The latter situation currently holds in England where pupils are assessed in the core subjects of English, mathematics, and science at ages 7, 11, 14, and 16 and allocated to one of a limited number of levels at each age. It is possible to model binary and categorical outcomes as functions of age although the nonlinear nature of the models (logits, probits, etc.) can lead to problems in estimation. Plewis (2005) jointly models three measures of children's behavior using ordered logits.

A rather different approach to modeling proficiency is to model variation in the time taken to reach the required level. This is a survival model rather than a growth-curve model but can be formulated in an ML framework. There are also other ways of modeling change in attainment. For example, as mentioned in the historical background in the introduction, it is often useful to look at the correlates of educational progress: to what extent are variables related to the outcome at occasion t conditional on (or controlling for) earlier or baseline measures. In addition, we might want to relate change or growth in attainment to other time- or age-varying explanatory variables: for example, modeling change in attainment as a function of change in family economic circumstances. These models are widely used in econometrics and examples of both conditional models and models using a time-varying explanatory variable are compared with growth curves in Plewis (2001).

This article has adopted what might be termed a conventional ML or hierarchical modeling approach to growth. A somewhat different way of thinking about these issues has been provided by, for example, Muthén (1997), placing them in a structural-equations modeling framework where

Table 5 Gender estimates and standard errors: Main effect and age interactions from bivariate model

Fixed effects	Literacy	Numeracy
Gender	*	0.082 (0.043)
Gender ^a age	*	−0.02 (0.005)
Gender ^a age ²	0.0027 (0.0007)	−0.0019 (0.0007)

^aEstimate less than its standard error.

the growth parameters (intercepts, slopes, etc.) are regarded as latent variables. This idea has led to a further useful extension: latent class growth analysis where individuals are assumed to belong to a set of underlying or latent classes that reflect different patterns of growth. These classes are derived from the data in such a way that they take account of much of the variability in the growth parameters. This idea has been extended by Muthén (2004) into what he calls general growth mixture modeling which allows for variability between individuals within the latent classes. The software package Mplus can be used to fit these models. These developments have found particular favor in analyses of deviant and delinquent behavior in psychology and criminology where there are both empirical and theoretical grounds for supposing that there are just a small number of classes, each exhibiting a different pattern of growth. There have been fewer applications to educational attainment data where it is perhaps less obvious what the latent classes would be although Muthén (2004) does give an example in terms of engaged and disengaged high school pupils.

A feature of nearly all longitudinal observational studies, and particularly of our example, is the way the observed sample diverges from the target sample over time. Although we have seen how growth-curve models can incorporate all the observed data on the responses, we do have to assume that the data that are missing are missing at random. In other words, providing missingness at occasion t depends only on what has been observed at occasions $t-k$ ($k > 0$) and not on change from $t-1$ to t ; then our estimates are unbiased and efficient. We do, however, lose cases where there are missing data on the explanatory variables unless we use some form of adjustment such as inverse probability weighting or multiple imputation.

To sum up, growth-curve models, applied thoughtfully to longitudinal data, offer educational researchers opportunities for further insights into, for example, learning and into the development of educational inequalities.

See also: Generalized Linear Mixed Models; Hierarchical Linear Models; Multivariate Longitudinal Data Analysis; Observational Studies; Value-Added Models.

Bibliography

- Bayley, N. (1949). Consistency and variability in the growth in intelligence from birth to eighteen years. *Journal of Genetic Psychology* **75**, 165–196.
- Cronbach, L. J. and Furby, L. (1970). How should we measure 'change' – or should we? *Psychological Bulletin* **74**, 68–80.
- Deming, J. (1957). Application of the Gompertz curve to the observed pattern of growth in length of 48 individual boys and girls during the adolescent cycle of growth. *Human Biology* **29**, 83–122.

- Healy, M. (1989). Growth curves and growth standards – the state of the art. In Tanner, J. M. (ed.) *Auxology 88. Perspectives in the Science of Growth and Development*. London: Smith-Gordon.
- Muthén, B. (1997). Latent variable modeling of longitudinal and multilevel data. In Raftery, A. (ed.) *Sociological Methodology*. Boston, MA: Blackwell.
- Muthén, B. (2004). Latent variable analysis. Growth mixture modeling and related techniques for longitudinal data. In Kaplan, D. (ed.) *Handbook of Quantitative Methodology for the Social Sciences*. Newbury Park, CA: Sage.
- Plewis, I. (1996). Statistical methods for understanding cognitive growth: A review, a synthesis and an application. *British Journal of Mathematical and Statistical Psychology* **49**, 25–42.
- Plewis, I. (2001). Explanatory models for relating growth processes. *Multivariate Behavioral Research* **36**, 207–225.
- Plewis, I. (2005). Modelling behaviour with multivariate multilevel growth curves. *Methodology* **1**, 71–80.
- Rogosa, D., Brandt, D., and Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin* **92**, 726–748.
- Wishart, J. (1938). Growth-rate determination in nutrition studies with the bacon pig, and their analysis. *Biometrika* **30**, 16–28.

Further Reading

- Boscardin, C. K., Muthén, B. O., Francis, D. J., and Baker, E. L. (2008). Early identification of reading difficulties using heterogeneous developmental trajectories. *Journal of Educational Psychology* **100**, 192–208.
- Bryk, A. S. and Raudenbush, S. W. (1992). *Hierarchical Linear Models*. Newbury Park, CA: Sage.
- Cudeck, R. and Harring, J. R. (2007). Analysis of nonlinear patterns of change with random coefficient models. *Annual Review of Psychology* **58**, 615–637.
- Harris, C. W. (ed.) (1963). *Problems in Measuring Change*. Madison, WI: University of Wisconsin Press.
- Goldstein, H. (1979). *The Design and Analysis of Longitudinal Studies*. London: Academic Press.
- Goldstein, H. (2003). *Multilevel Statistical Models*, 3rd edn. London: Edward Arnold.
- Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd edn. New York: Wiley.
- Nagin, D. S. (1999). Analyzing developmental trajectories: A semi-parametric, group based approach. *Psychological Methods* **4**, 139–157.
- Plewis, I. (1997). *Statistics in Education*. London: Edward Arnold.
- Plewis, I. (2009). Statistical modelling for structured longitudinal designs. In Lynn, P. (ed.) *Methodology of Longitudinal Surveys*. Chichester: Wiley.
- Singer, J. D. and Willett, J. B. (2003). *Applied Longitudinal Data Analysis*. Oxford: Oxford University Press.
- Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, FL: Chapman and Hall/CRC.
- Snijders, T. A. B. and Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling*. London: Sage.

Relevant Websites

- <http://www.cmm.bristol.ac.uk> – Centre for Multilevel Modelling, Bristol, UK.
- <http://www.statmodel.com> – Mplus.

Hierarchical Linear Models

D Rindskopf, CUNY Graduate Center, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

In many educational contexts, data occur in a nested structure. For example, we may measure the achievement level of a child in a particular subject matter. That child is located (nested) within a particular class of students, and we might expect that the child's performance would have been different had she been in a different class of students. Further, the child's class may be one of many within a particular school, and that school may be one of many within a school system, which is located within a larger educational or political unit (e.g., county, state, and region). Each level of the structure may have its own characteristics that influence the achievement of the child. Hierarchical linear models (and more general models, to be discussed later) are designed to properly model influences on outcome variables at all levels of such a hierarchy, including the proper handling of data dependencies created by the nested structure.

In this article, we describe a variety of models that are designed to accommodate a nested data structure. Some models are linear, but others are nonlinear; some involve nesting of people within higher level units (such as classes), while other models are applicable to repeated measures that are nested within individuals (including longitudinal studies). We begin with a very simple example, and then show how the principles involved can be used to expand the scope and type of models that may be used. These principles are all expansions or generalizations of those used in ordinary least squares (OLS) regression; often, the models can be viewed as a series of regressions for different levels of the data structure. Those unfamiliar with regression should read the article on univariate linear regression in this volume before reading the rest of this one.

Although we will not discuss it in detail, these methods also resolve the problem known as aggregation bias. It has been well known that correlations among variables at a higher level (e.g., measures on states) do not equal (and may be opposite in sign to) correlations among variables at the individual level. (A related problem that arises in categorical data analysis is known as Simpson's paradox.) Multilevel models are constructed so that inferences can be made at the right level, whichever level that happens to be for the problem at hand.

We discuss two examples of multilevel models. The first example is for a situation in which students are nested within colleges, and we wish to predict academic performance from one student-level and one college-level variable. The second example considers longitudinal data,

where a series of observations is made on each person over a period of time. Here, the nesting is of a different sort: Observations (over time) are nested within individuals.

Example 1: Two Levels, One Predictor at Each Level, Linear Model

For a first example, assume that the administration in 10 colleges in a particular country wants to see how well scores on an academic-achievement test given to all students in the last year of secondary school predict performance of students in their first year of college. If we consider each college separately, we may imagine that within each college, researchers fit a regression model to the data; the equation for school j would then be:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$$

In this equation, Y represents the grade of a student after the first year of college, X represents the score of the student on the secondary school-achievement test, i is used to refer to a specific student, and j to refer to a specific school. The coefficients are as in a linear regression model, except that each school has its own intercept (β_{0j}) and slope (β_{1j}).

As we shall be examining the intercept and slope of each school, and modeling them, it is important to be sure that these are quantities of interest. For the slope, there is ordinarily no issue (except perhaps multiplication or division by a constant in order to make a one-unit change more easily interpreted). In order to make the intercept meaningful, we will often have to transform X . One common strategy is centering, which means to create a new variable by subtracting the mean of X from each observed value of X . (For this article, we assume that the mean is computed for the whole data set, not for each college individually; this is called grand mean centering.) In any regression, the intercept represents the predicted value of Y when all predictors have a value of 0; therefore centering X will cause the intercept in each college's regression equation to equal the predicted value of Y for a student who is average on X (based on students from all included colleges). Using the exact sample mean is not necessary; sometimes, it is more important to use a round number for ease of interpretation. For example, if the mean is 497, it might be easier to subtract 500. Or, if a score of 600 is typically required for admissions, one might subtract 600 from all X values instead.

In addition to subtracting some value from X , one might also want to control the scaling further. In the United States, for example, where scholastic aptitude test (SAT) scores have a mean near 500 and a standard deviation near 100, one might divide by 100 so that a one-unit change in the transformed variable is more meaningful than a one-unit change on the original scale. This makes the slope more easily interpreted (and sometimes improves numerical stability.) From this point on, we will assume that X has been suitably scaled.

The term r_{ij} is a residual, indicating that the students' scores on the secondary school test will typically not predict their grades in college perfectly. The variance of the residuals in each college can differ among schools, but often modeling is begun by assuming that it is the same in each; we will denote the within-school residual variance by σ^2 , in which we do not use a subscript for college (j).

The intercepts and slopes will vary from college to college; some of this variation is just sampling variability due to using a sample of students within each college, but the other variability is due to true differences among colleges. The true variability in intercepts will be denoted by τ_{00} , and the true variability in slopes will be denoted by τ_{11} . (There is also a covariance between the intercepts and slopes, denoted τ_{01} , but this is seldom of much interest.)

If τ_{00} and τ_{11} are small, then the intercepts and slopes do not vary much among colleges, and one prediction equation will work equally well in each. The colleges will gain by knowing this, because now their prediction equation is known with greater precision due to the contribution of information from other colleges. Of course, unless the colleges are very similar, it is unlikely that both τ_{00} and τ_{11} will be small, which leads to a consideration of two issues: First, can we explain, based on characteristics of the colleges, why some have larger and others smaller intercepts and/or slopes? Second, even though the colleges differ, can we still use information from other colleges to help improve prediction in each college?

We may believe that characteristics of the colleges will have some relationship to either the intercept or the slope of their prediction equations. For example, it may be that schools differ in the difficulty of their courses, or in their grading standards. These may affect the intercepts or slopes. One might use the college's average value on X as a proxy for selectivity or difficulty; this could be a predictor for intercepts and/or slopes. Another possible measure would be the selection ratio, that is, the proportion of applicants who are admitted to the college.

We will now write the model for variability among colleges out more formally. The equations, which will first be presented without predictors, are:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

The first equation says that the intercept in a particular college (j) is equal to some overall average plus a deviation of that college's intercept from the overall average. The second equation expresses the same idea for the slopes from each college. The variances of the residuals (u_{0j} in the first equation and u_{1j} in the second) are the values τ_{00} and τ_{11} mentioned previously. In this instance, they are called unconditional residual variances, because they are not conditioned (dependent) on any college-level predictors. When we analyze the data we will obtain estimates of both the mean intercept and slope (called fixed effects in most software) and the variances of the residuals (called random effects in most software).

Now suppose that the residual variances are large enough that we believe it worthwhile to try to account for some of their variation by using a college-level variable such as proportion of applicants admitted. In realistic applicants, there are often many potential variables, and variables that help predict differences in intercepts may not be the same variables that help predict differences in slopes. In this simple introduction, however, we will use only one such variable, which we will call W ; individual values of W are denoted W_j , with the subscript j indicating that the value of W will vary across colleges.

The extended model at the college level will now consist of the following equations:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$$

The variances τ_{00} and τ_{11} of the residual terms u_{0j} and u_{1j} now represent conditional variances; that is, the variances remaining after accounting for variation in intercepts and slopes using information provided by the predictor W . These variances should now be smaller than in the unconditional model if W proves to be a good predictor, and measures analogous to R -square in the usual regression context may be calculated.

Multilevel models are very much like ordinary regression models in many ways: the outcome variable may be continuous or categorical. If continuous, it may be normally distributed or have some other distribution (e.g., exponential for survival models). If the dependent variable is categorical, models similar to logistic regression (for binary outcomes) or its extensions to polytomous or ordinal outcomes are possible. The right-hand side of any of the equations may require polynomial coding (e.g., the square and cube) for continuous predictors, dummy variables (or other coding methods) for categorical predictors, and product terms for interactions. The model may be intrinsically nonlinear, so that more complex models must be fit. All of these are possible, but outside the scope of this introductory article.

Example 2: Repeated Measures and Longitudinal Data

In many educational applications, people are measured more than one time. In some cases, these measurements extend over a long period of time; measurements may be daily, weekly, monthly, or yearly, for example. In other cases, the time sequence is not of great importance; for example, most models for test scores do not take into account that some items are administered before others. In either case, we can use the concept of nesting to reconceptualize the model for this type of data: we view observations as being nested within an individual. (In some instances, it might seem more natural to view time of observation as crossed with individuals, but this viewpoint is not usually considered in the multilevel literature, because doing so would reduce the utility of the model in many common situations, as we hope will become apparent shortly.)

As an example, suppose that a researcher wants to examine the growth of children's ability to sight-read words. Each month during the school year, she has every child in a class attempt to read a list of 20 words, and counts the number of words that the child can correctly read. We will treat the number of words read as if it were a continuous outcome variable to simplify the discussion in this section, and will later deal with the complexities that this type of outcome necessitates. The researcher wants to model the number of words read by each child as a function of the number of months into the school year. The time variable may be coded in a number of ways; one might be to start the variable Months at 0 for the first measurement period, and increase by one for each month. Another approach would be to use 0 for the last period of measurement, and number preceding periods as -1 , -2 , etc. Each possible choice determines the meaning of the intercept in the model. Again, for simplicity we will use a linear model, but realize that a straight line is unlikely to be a perfect description of such growth; we write the equation for each student as follows:

$$Y_{ij} = \pi_{0j} + \pi_{1j}M_{ij} + e_{ij}$$

where the subscript j is used to represent the individual student, i is used to represent a measurement occasion, Y_{ij} is the number of words read correctly by student j at occasion i , M_{ij} is the number of months (on whatever scale is chosen; here we will assume 0 is used for the first time of measurement) for measurement of student j at occasion i , and e_{ij} is the residual. The intercept π_{0j} and slope π_{1j} are different for each student, and again we will allow these to be predicted by variables at the higher level, which in this case is the individual student. For example, boys and girls may start at different levels, and they may progress at different rates. We may represent this by the following level-2 equations, in which F is equal to one for females and zero for males:

$$\pi_{0j} = \beta_{00} + \beta_{01}F_j + r_{0j}$$

$$\pi_{1j} = \beta_{10} + \beta_{11}F_j + r_{1j}$$

Of course the structure of all of these equations is exactly the same as the case in the first example where students were nested within colleges; only the context has changed. In these equations, the intercepts β_{00} and β_{10} represent the beginning number of words read (π_0) and the rate of change in words per month (π_1) for an average male (i.e., $F = 0$) in the study. The slopes in these equations, β_{01} and β_{11} , represent the difference in beginning number of words read and rate of change for females as compared to males. Again, the choice of coding makes a difference; gender could be coded in other ways, which would change the interpretation of each of the coefficients in these equations. The residual terms represent the amount by which the true values of π_{0j} and π_{1j} differ from that predicted for the average male or female.

This method of representing the model has several advantages over other methods. Most importantly, from a conceptual point of view, this is a model for each individual's growth pattern and for deviations of individuals from the average pattern. The focus is on the individual, rather than a group. As a consequence of this focus, in principle (and at present, in practice also) missing data are not a problem, as long as the cause for missingness is not related to the outcome variable. Each child could have a different pattern of months during which he or she was measured, and in fact the times of measurement do not even have to follow a common pattern. A child could be measured at months 1.25, 3.75, and so on, rather than exactly at the beginning of each month – the model remains the same. (Of course, in practice one must measure each individual a reasonable number of times, and with those times spaced in a reasonable way, in order to have accurate estimates of that child's growth pattern, but this is another matter.)

Generalized Linear and Nonlinear Models

Two general categories of complications frequently arise with multilevel models. As mentioned previously, not all dependent variables are continuous, and not all continuous variables have a normal distribution. In addition, not all relationships are linear, so nonlinear models must sometimes be considered. This section briefly describes some approaches to each issue.

In the second extended example described, the outcome was the number of words read correctly from a list of 20 words presented to the student. Equivalently, one might calculate the percentage or proportion of the list that was read correctly. If proportions were not extreme (e.g., if most were between 0.2 and 0.8), then using these in a model with a normal distribution might be a good

approximation to the correct analysis. A more accurate approach would be to consider each word as a trial, and the number correct as coming from a binomial distribution with an underlying probability of success for each trial that is constant across words. (An even more accurate model would allow words to have varying difficulties, and possibly to differ in other ways also, as in item-response theory. Often the simplifying assumption of equal difficulty is a good enough approximation to work well in practice.)

The binomial model is closely related to a more widely known model, logistic regression. For these models, the outcome that is modeled is usually the logarithm of the odds of a correct response, commonly called the logit of the probability of a correct response. The right-hand side of the level-1 equation for individual responding still has a linear form in these models; the left-hand side is the transformed value $\ln(\omega/(1-\omega))$, also denoted $\text{logit}(\omega)$ in some sources, where ω is the probability of a student responding correctly. The model for each individual is written in the following form:

$$\ln\left(\frac{\omega_{ij}}{1-\omega_{ij}}\right) = \pi_{0j} + \pi_{1j}M_{ij}$$

where ω_{ij} is the expected (not observed) probability of person j answering correctly during month i . The residual is not explicitly listed in this equation; instead, we state that $Y_{ij} \sim \text{Binomial}(20, \omega_{ij})$; that is, the observed number of words correctly read has a binomial distribution based on 20 trials, with probability ω_{ij} of being correct on each trial.

This is an example of a generalized linear model; these are extensions of linear models that include a linear model, a transform of the dependent variable, and a distributional form that is more general than a normal distribution. In this case, the logit is the transformation and the binomial is the distributional form.

Another example requiring the use of generalized linear models is when the outcome is a count, but the count is not the number of successes in a fixed number of trials, or it is a generally small count out of a very large number of fixed trials (in which case we are using an approximation to the binomial distribution). A researcher might count the number of times a student shows an aggressive behavior toward classmates, or the number of times the student raises his or her hand for the attention of the teacher rather than responding without being called on. If the period of observation is the same for each student and time of observation, then generally the Poisson distribution provides the simplest probability model for counts. A slight extension is required when the time of observation varies, so that the model is for the rate of behavior per period of time (number of hand raises per hour, e.g., when the time of observation may vary between 5 and 20 min). If F represents the expected number of acts and Z represents the amount of time for an observational period,

so that F/Z is a rate per unit time, the level-1 equation can be written in the following form:

$$\ln\left(\frac{F_{ij}}{Z_{ij}}\right) = \pi_{0j} + \pi_{1j}M_{ij}$$

As with the binomial model, there is no explicit error term in the equation because the count variable in the equation is an expected rather than observed frequency, and in this case it has a Poisson distribution.

A related model with the same form is the discrete-time survival model, in which we model the influences on the occurrence of an event. In education, we might wish to determine factors affecting whether, if so when, a student in college will drop out. Modeling whether or not the student drops out involves an ordinary logistic regression (or possibly with students nested within colleges if we observe many colleges). But to model the length of time to dropout is slightly more complicated. We observe each student over a series of semesters, and note whether he or she has dropped out or is still enrolled for that semester. The main aspects of this situation that are different from others with a binary outcome in each time period are that (1) drop out can only occur during one period, and (2) the sequence of observations can be stopped (censored) due to graduation so that the event we are looking for (dropout) may not be observed for that individual. It may not be obvious, but information from students who never drop out is nonetheless useful. Survival models developed in somewhat parallel ways in three areas: medicine (for obvious reasons), engineering (where they are generally known as failure time models), and sociology (where they are known as event history models).

Shrinkage Estimates

In many educational situations (in particular, test administration), one main objective is to estimate the ability (or some other characteristic) of individuals. The most obvious estimate when one merely calculates the total number of items right (or totals points for various nondichotomous items) is to use that total as an estimate. But if the quantity of information available for different individuals is not the same, this obvious estimate has statistical flaws. To take an extreme example, if one individual was administered two items (randomly sampled from an item pool) and got both right, while another was administered a hundred items and got 90 correct, would ability estimates of 1.00 and 0.90 (or monotonic transformations of these) seem reasonable? Would you be willing to believe that the first person has a higher ability than the second, even though he or she answered only two items? If not, what would a reasonable procedure be? It turns out that a reasonable procedure is to shrink all estimates to the mean (much like the phenomenon of regression to

the mean). The more information you have (e.g., the more test items), the less shrinkage there would be for that individual. Hierarchical models provide a general framework within which one can calculate differential shrinkage for individuals depending not only on the amount of information about the individual provided by the statistic of interest (standard error of the estimate), but also other (auxiliary but informative) data (e.g., the amount of training the individual has received in the subject matter). Hierarchical models can also be used to obtain improved estimates of units other than people (classes, schools, or higher level units).

Meta-Analysis

A situation in which it is not obvious that multilevel models can be applied is meta-analysis. In meta-analysis, a large number of studies about a topic are summarized by a measure of effect size, and by other characteristics of the study (age of subjects, whether a randomized design was used, etc.). The objective is to determine what the average effect size is, whether it varies (beyond sampling variability) from study to study, and if so whether study characteristics can be used to predict that variability. If we view subjects as nested within studies, it is easy to see that this falls into the realm of multilevel models, but with the individual level data missing (i.e., we only see summary statistics for the study as a whole). Most programs for multilevel models (and some additional specialized programs) can perform such analyses.

Recommended Reading

Several excellent textbooks are available to learn more about multilevel models. Raudenbush and Bryk (2002), whose notation is used here, and Goldstein (2003) cover a wide range of material on these models. Each of these also is associated with software to fit these models, HLM (Raudenbush *et al.*, 2004) for the former and MLWin (Rasbash *et al.*, 2005) for the latter. Other introductory books include Hox (1995, 2002) and Snijders and Bosker (2000). Gelman and Hill (2007) have a unique viewpoint, and many interesting approaches not found elsewhere. Other references in the reading list include more advanced works, as well as specialized references on various aspects of multilevel models.

Summary and Conclusion

When units are nested within higher level units, multilevel modeling should be used for three major reasons: (1) it is the correct analysis (which ordinarily would be a

sufficient reason), (2) viewing the model in this way leads to insights that are otherwise elusive about relationships, and (3) we avoid certain perplexing problems, such as aggregation bias. (Note that not all nesting produces a multilevel model; only when both levels are random effects is this true. For example, people nested within treatments, as in a typical ANOVA, do not result in the type of model discussed here because levels of treatment are typically fixed in number and are not a random sample from a population of treatments.) Multilevel models can be used not only in obvious situations such as when we have children in classes (and or schools), but also where subjects are nested within studies (meta-analysis), or where observations are nested within people (repeated measures and longitudinal studies). Using these models enables us to estimate effects within each level of a study, as well as between levels. Specialized software is readily available, and several major statistical packages now include the capability to fit multilevel models.

See also: Growth Modeling; Survival Data Analysis; Univariate Linear Regression.

Bibliography

- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.
- Goldstein, H. (2003). *Multilevel Statistical Models*, 3rd edn. London: Edward Arnold.
- Hox, J. J. (1995). *Applied Multi-Level Analysis*, 2nd edn. Amsterdam: TT-Publikaties.
- Hox, J. J. (2002). *Multilevel Analysis: Techniques and Applications*. Mahwah, NJ: Erlbaum.
- Rasbash, J., Steele, F., Browne, W. J., and Prosser, B. (2005). *A User's Guide to MLWin version 2.0*. Bristol: University of Bristol.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edn. Thousand Oaks, CA: Sage.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., and Congdon, R. (2004). *HLM 6: Hierarchical Linear and Nonlinear Modeling*. Chicago, IL: Scientific Software International.
- Snijders, T. and Bosker, R. (2000). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.

Further Reading

- Braun, H. (1989). Empirical Bayes methods: A tool for exploratory analysis. In Bock, R. D. (ed.) *Multilevel Analysis of Educational Data*, pp 19–55. New York, NY: Academic Press.
- Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for repeated measurement Data*. Boca Raton, FL: Chapman and Hall.
- Efron, B. and Morris, C. (1977). Stein's paradox in statistics. *Scientific American* **236**, 119–127.
- Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., and Schabenberger, O. (2006). *SAS for Mixed Models*, 2nd edn. Cary, NC: SAS Institute.
- Longford, N. T. (1993). *Random Coefficient Models*. Oxford: Clarendon Press.

- Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* **10**, 325–337.
- McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*. New York: Wiley-Interscience.
- Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics* **24**, 323–355.
- Singer, J. D. and Willett, J. B. (2003). *Applied Longitudinal Data Analysis*. Oxford, UK: Oxford University Press.
- Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Boca Raton, FL: Chapman and Hall/CRC Press.
- Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester: Wiley.
- Spiegelhalter, D. J., Thomas, A., Best, N. G., and Lunn, D. (2003). *WinBUGS User Manual*. (Version 1.4). Cambridge: MRC Biostatistics Unit. <http://www.mrc-bsu.cam.ac.uk/bugs> (accessed August 2009).

Relevant Websites

- <http://www.cmm.bristol.ac.uk> – Center for Multilevel Modeling.
- <http://www.ssicentral.com> – Scientific Software International (contains many good references to books, articles, and websites).
- <http://www.ats.ucla.edu> – UCLA Statistical Computing (contains useful resources for multilevel models and other advanced statistical procedures).

Hypothesis Testing and Confidence Intervals

T R Konold and X Fan, University of Virginia, Charlottesville, VA, USA

© 2010 Elsevier Ltd. All rights reserved.

Hypothesis testing involves estimating the probability (p) of observing a sample statistic (e.g., sample mean) that is equal to or more extreme than what is obtained from a sample, assuming that the sample is drawn from a population with a hypothesized population parameter value (e.g., population mean, μ). This article provides an overview of hypothesis testing through both point and interval estimation approaches. Although we use a population mean in some examples, the general principles discussed here extend to situations involving other sample statistics (multiple sample means, variance, proportion, correlation coefficient, etc.). Sampling distribution, sampling error, point estimate, interval estimate, statistical errors, and statistical power are among the topics discussed below.

Research Questions and Hypotheses

Research questions and hypotheses are usually based on a researcher's educated guess or intuition grounded in a good understanding of the phenomenon of interest. For example, a teacher might ponder whether a new instructional approach has an effect on student achievement as measured on a state-wide assessment of learning (the research question). Based on his/her pedagogical content knowledge, the teacher may generate a research hypothesis that this new instructional approach does impact student achievement. Although this research hypothesis does not indicate the direction of influence, such qualifications can easily be incorporated to reflect whether the researcher believes that this instructional approach will increase or decrease measured achievement. In general, a research hypothesis reflects a researcher's predictions and it directly shapes the statistical hypotheses to be empirically evaluated.

Statistical Hypothesis

Null Hypothesis

In order to test a research hypothesis, it is necessary to translate it into a statistical hypothesis. A statistical hypothesis is a numerical statement regarding a hypothesized population parameter. For example, knowing that the average state-wide assessment score for the general student population is 72 raw score points, a researcher might state the null hypothesis (H_0) as

$$H_0: \mu = 72$$

The H_0 is often stated to reflect the general state of affairs, or the condition of no difference. In hypothesis testing, assuming that the null hypothesis is true, we evaluate the probability of observing a sample statistic (e.g., sample mean) equal to or more extreme than what was obtained from our sample drawn from the hypothesized population with the population value (population parameter) stated in the H_0 (i.e., $\mu = 72$). In most substantive applications, researchers hope to demonstrate that this probability is small, and the observed sample estimate is unlikely to have arisen from a population as hypothesized in the H_0 . Although this example concerns a hypothesis about a population mean μ , hypotheses concerning any population parameters (e.g., variance, correlation coefficient, and proportion) are essentially the same. So we may express H_0 in a generic form:

$$H_0: \theta = K$$

where θ is the population parameter under investigation, and K is the hypothesized value of that parameter.

Directional and Nondirectional Alternative Hypotheses

Hypothesis testing involves two statistical hypotheses. The first is the null hypothesis (H_0) as described above. For each H_0 , there is an alternative hypothesis (H_a) that will be favored if the null hypothesis is found to be statistically not viable. The H_a can be either nondirectional or directional, as dictated by the research hypothesis. For example, if a researcher only believes the new instructional approach will have an impact on student test scores, but is unsure whether the effect will be positive or negative, the null and alternative hypotheses would be

$$H_0: \mu = 72$$

$$H_a: \mu \neq 72$$

Here, H_a reflects the researcher's uncertainty regarding the directionality, and it allows for a statistical test that considers both possibilities that the new instructional approach could increase test scores or decrease test scores. This is commonly referred to as a nondirectional alternative hypothesis, and is also referred to as a two-tailed test for reasons that are described below.

A directional alternative hypothesis, on the other hand, is useful to accommodate the researcher's prediction that, for example, the new instructional approach will decrease

test scores ($H_a: \mu < 72$ bpm) or will increase test scores ($H_a: \mu > 72$). A directional alternative hypothesis is often referred to as a one-tailed test as described below. It is important to note, however, that for every specified H_0 there will be a single H_a that may assume one of the three forms

$$H_a: \theta \neq K$$

$$H_a: \theta < K$$

$$H_a: \theta > K$$

Sample, Population, and Inferential Statistics

A sample is a subset of a population of interest that is defined by the researcher. For example, one could conceivably collect math achievement data on all fifth graders in the United States. However, given the limited resources, it is more realistic and economical to work with a subset of this group (i.e., sample), with the goal of generalizing the findings to the total population. It is possible, or even very common, that the population of interest is a statistical population that does not really exist. For example, our educator interested in the effectiveness of a new instructional approach may statistically compare two student samples: one sample receiving this new approach, and the other receiving the business-as-usual approach. In this situation, there are two hypothetical statistical populations: one population of students taught under this new approach, and another population of students taught under the business-as-usual instructional approach. In reality, we do not really have these two populations.

A sample statistic $\hat{\theta}$ is a measure of a sample characteristic (sample mean, proportion, correlation, etc.), whereas a measure of the corresponding population characteristic is a population parameter (θ). Researchers are typically interested in understanding the characteristics of a population, but it is often impossible or impractical to obtain data from the whole population (e.g., lack of resources to do so; or a hypothesized statistical population). Inferential statistics are useful for estimating population characteristics (i.e., parameters), and for generalizing the sample findings to the population from which the sample was drawn.

Hypothesis testing is at the center of inferential statistics. Both point estimation and interval estimation are commonly used first steps in the hypothesis testing process. In point estimation, a single population parameter value is hypothesized and used in hypothesis testing. By contrast, interval estimation serves as a useful supplement to point estimation because it provides a likely range of values for the unknown population parameter. The key to

both point and interval estimation is in estimating the sampling variability of a given sample statistic.

Sampling Distribution

In hypothesis testing, sampling variability (i.e., sampling error) of a sample statistic $\hat{\theta}$ is the foundation for estimating the probability of observing a sample statistic equal to or more extreme than what is obtained under the true null hypothesis. A statistic's sampling distribution provides the estimation for sampling error. For example, assume we are interested in the average performance on a math achievement exam (parameter) of all fifth graders in the United States (population). Our best estimate of the unknown population mean parameter (μ) would be the sample average (\bar{X}) based on a random sample of n observations from this population. If this process of drawing random samples of size n from the same population were repeated many times, the means from all the samples would vary, and these means themselves would form a distribution, which is called a sampling distribution of the mean. This sampling distribution concept also extends to other sample statistics (e.g., sample variance, proportion, and correlation).

This sampling distribution captures the sample-to-sample variability of a sample statistic. In the case of \bar{X} , theoretically, the average of all possible sample means, or the expected value of the statistic, $E(\bar{X})$, is equal to the population mean (i.e., μ). Statistics with this property, that is, $E(\hat{\theta}) = \theta$, are referred to as unbiased estimators.

Standard error

Sample-to-sample variability of a statistic, or sampling error, can be quantified by standard error. In the case of sample means as described above, the standard error of the mean quantifies the variability of the sampling distribution of the mean. When samples are randomly selected from a population, the resulting distribution will be approximately normally distributed for reasonable sample size of n (see the subsection titled 'Central limit theorem' below). When the population standard deviation (σ) is known, the standard deviation of the sampling distribution of means, usually referred to as standard error of the mean, is:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

As is obvious, $\sigma_{\bar{X}}$ is a function of the sample size (n). As n increases, $\sigma_{\bar{X}}$ decreases to its lower limit of 0 (i.e., $n \rightarrow \infty, \sigma_{\bar{X}} \rightarrow 0$). Likewise, as sample size decreases, $\sigma_{\bar{X}}$ increases to its upper limit of the population standard deviation σ (i.e., $n \rightarrow 1, \sigma_{\bar{X}} \rightarrow \sigma$). In general, over repeated sampling, sample statistic $\hat{\theta}$'s based on a larger

sample size has less variability (i.e., less sampling error) than those based on a smaller sample size.

Central limit theorem

Several properties of sampling distributions that have important implications for hypothesis testing are captured in the central limit theorem (CLT). In a nutshell, CLT states the following. First, as the size (n) of the samples drawn from a population increases, the standard error of the means decreases, leading to less variability of sample means and smaller sampling error. **Figure 1** presents two sampling distributions of the means, one based on $n = 10$, and the other based on $n = 30$. As shown in **Figure 1**, the distribution based on a smaller sample size ($n = 10$) spreads out more than the one based on a larger sample size ($n = 30$).

Second, the shape of the sampling distribution of the mean becomes increasingly normal as the sample size increases. This is true regardless of the shape of the parent population from which the samples are drawn. As a result, the sampling distribution of \bar{X} can generally be considered as being normally distributed for samples of size $n = 30$ or greater, even though the parent population may not be normally distributed.

Sampling distribution as probability distribution

Hypothesis testing involves estimating the probability of observing a sample statistic ($\hat{\theta}$) equal to or more extreme

than what is obtained from a sample, assuming true H_0 . For this purpose, the sampling distribution of the statistic developed under H_0 serves as the probability distribution. Under this probability distribution, the closer the observed sample statistic (e.g., \bar{X}) is to the population parameter (e.g., $\mu = K$) under H_0 , the more likely that the difference between \bar{X} and μ may have been the result of sampling variation (i.e., sampling error). Sample statistic values near the tails of the sampling distribution represent a greater difference between the sample statistic and the population parameter under H_0 . These values occur less frequently in repeated sampling, and consequently, the probability of observing the sample statistic values near the tails is smaller. Under H_0 , if this probability is judged to be smaller than a predetermined threshold probability level (denoted α), we will conclude that H_0 is likely not true; if it were, it would be very unlikely to observe the sample statistic. Conventionally, $\alpha = 0.05$ is often used as such a threshold, leading to the following decision rules with regard to H_0 :

$$\text{If } p(\hat{\theta}|H_0) \leq \alpha = 0.05 \rightarrow \text{reject } H_0$$

$$\text{If } p(\hat{\theta}|H_0) > \alpha = 0.05 \rightarrow \text{fail to reject } H_0$$

This rule implies that when the observed sample statistic value becomes so rare under the H_0 sampling distribution (e.g., $p < 0.05$), we would be willing to conclude that the value specified in the H_0 is not a plausible population value for the observed sample statistic. It should be noted that this rule does not suggest that such sample statistic values cannot occur under H_0 ; it only states that values so extreme occur less frequently under H_0 . Moreover, there is nothing magical about the conventional $\alpha = 0.05$ decision rule: researchers may choose other values for evaluating H_0 , based on their considerations about type I and type II errors (described below). It is important to note that, given the observed sample statistic, hypothesis testing is unable to address the question of whether H_0 is true. The process only involves the determination of the probability of obtaining sample statistic values equal to or more extreme than the sample result under the assumption that H_0 is true (Cohen, 1994).

Test Statistics and Probability Estimates

In hypothesis testing, to assess the probability of observing values more extreme than a given sample statistic under H_0 , we translate the observed sample statistic value into a test statistic. For example, the test statistic for evaluating a single sample mean (\bar{X}) when the population standard deviation (σ) is known is:

$$Z_{obs} = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

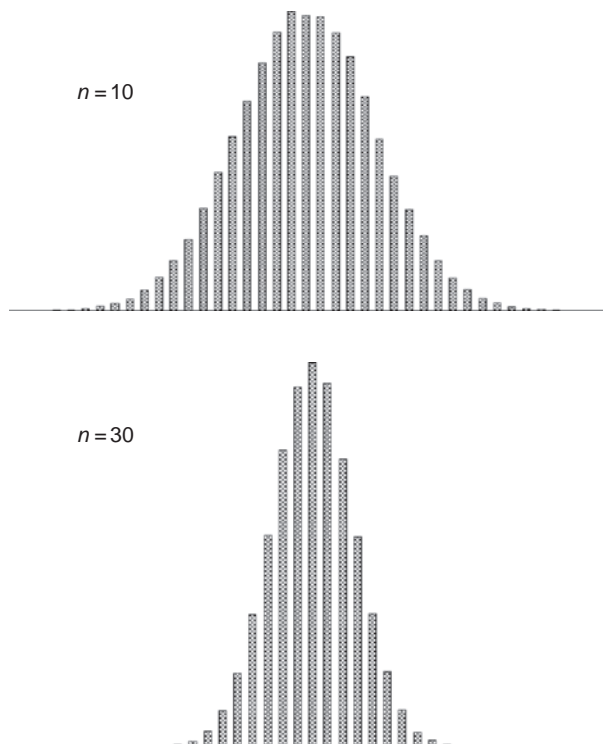


Figure 1 Two sampling distributions of the mean from the same population: $n = 10$ and $n = 30$.

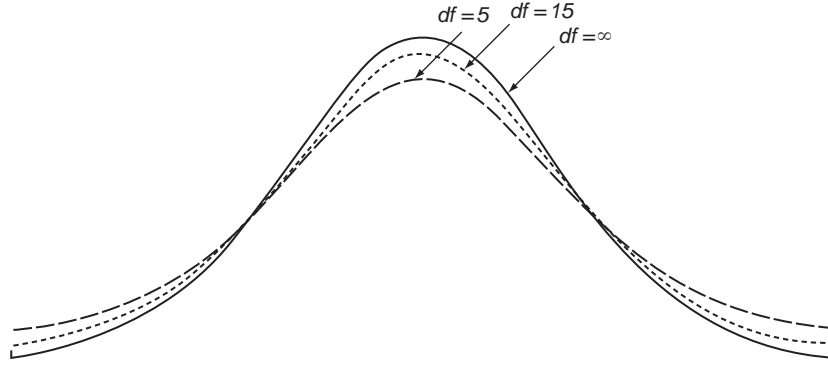


Figure 2 Three t -distributions with $df = 5$, $df = 15$, and $df = \infty$.

This z_{obs} test statistic represents the distance between the sample mean and hypothesized population mean in standard deviation units, where the standard deviation is the standard deviation of the sampling distribution of the mean (i.e., standard error of the mean: (σ/\sqrt{n})).

A larger absolute value of z_{obs} is further away from the hypothesized population mean. The exact probability of observing a value of z_{obs} or larger can be obtained from a normal distribution table.

Application of the z test statistic requires that the population standard deviation σ be known. When σ is unknown and must be estimated with the sample standard deviation (s), the estimated standard error of the mean is $s_{\bar{X}} = \frac{s}{\sqrt{n}}$, and the test statistic becomes:

$$t_{\text{obs}} = \frac{\bar{X} - \mu}{s_{\bar{X}}}$$

In contrast to the single z distribution, there exists a family of t distributions, as defined by degrees of freedom (df) that are based on sample size ($df = n-1$). **Figure 2** presents three t distributions with different df ($df = 5$, 15 , or ∞). Because the t distributions have different shapes, probabilities beyond a given t value depend not only on the value of t , but also on the df on which that statistic is based. As sample size n increases, t distributions converge on the z distribution. Beyond $df = 120$, the difference between z and t distributions is negligible, and a t -test practically becomes a z -test.

Test statistics (e.g., F statistic, χ^2 statistic) other than z and t may be needed as required by a statistical analysis. Regardless of the specific test statistic used in a research situation, the logic and procedure described here for hypothesis testing remain the same.

Decision Rules Based on Test Statistic

The discussion above focused on testing H_0 by comparing the probability of obtaining sample statistic values more extreme than a given sample statistic, assuming the true

H_0 , against a predetermined threshold probability level (i.e., $\alpha = 0.05$). In practice, this is done through a direct comparison of the observed sample test statistic (e.g., z_{obs} , t_{obs} , F_{obs}) against the critical value of the statistic (e.g., z_{cv} , t_{cv} , or F_{cv}) as determined by the predetermined threshold probability level α . For example, the critical value of t (i.e., t_{cv}) associated with a given value of α and a given df is readily available from a table of t distributions. By comparing t_{obs} against t_{cv} , we can establish the following decision rule for testing H_0 :

$$\text{If } |t_{\text{obs}}| \geq |t_{\text{cv}}| \rightarrow \text{reject } H_0$$

$$\text{If } |t_{\text{obs}}| < |t_{\text{cv}}| \rightarrow \text{fail to reject } H_0$$

This decision rule is a direct translation of our previous probability-based decision rule for testing H_0 :

$$\text{If } p(\hat{\theta}|H_0) \leq \alpha \rightarrow \text{reject } H_0$$

$$\text{If } p(\hat{\theta}|H_0) > \alpha \rightarrow \text{fail to reject } H_0$$

For example, our educator researcher previously hypothesized mean student achievement test score of 72 for a population ($H_0: \mu = 72$), and she would like to test this H_0 against the nondirectional alternative hypothesis $H_a: \mu \neq 72$ for $\alpha = 0.05$. From a random sample of 100 participants, the researcher obtained $\bar{X} = 68$, and $s = 8$ (i.e., $s_{\bar{X}} = \frac{s}{\sqrt{n}} = 0.8$). The test statistic is:

$$t_{\text{obs}} = \frac{\bar{X} - \mu}{s_{\bar{X}}} = \frac{68 - 72}{0.8} = -5$$

As $|t_{\text{obs}}| = 5$ exceeds the $t_{\text{cv}} [t_{(.975, df=99)} = 2.28, \alpha = .05]$, the $H_0: \mu = 72$ is rejected, as it is very unlikely ($p < 0.05$) to observe sample mean values equal to or greater than $\bar{X} = 68$ if H_0 were true.

Either a nondirectional or directional alternative hypothesis may be evaluated. When a nondirectional alternative hypothesis is evaluated (i.e., $H_a: \mu \neq K$), the overall α (e.g., $\alpha = 0.05$) is evenly divided into the two tails of the sampling distribution of H_0 , to accommodate the possibility that the direction of difference between \bar{X} and μ can be either

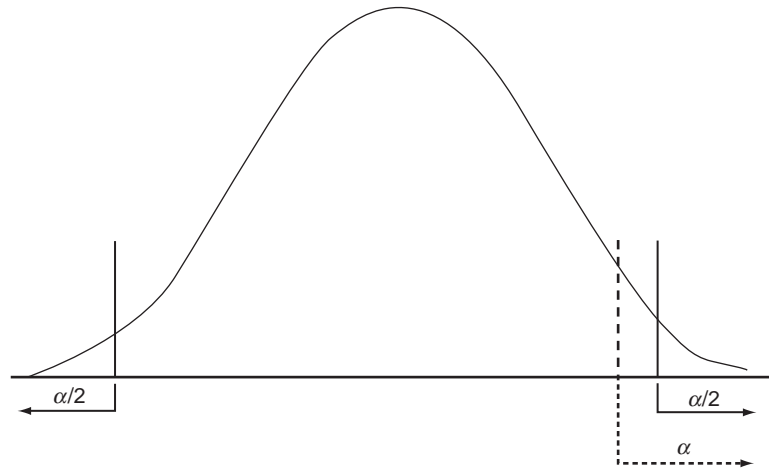


Figure 3 The α at distribution tails for nondirectional (two-tailed) and directional (upper tail) tests.

positive or negative, as shown in **Figure 3** (two solid lines at the tails).

When a directional alternative hypothesis is used (e.g., $H_a: \mu > K$), we only consider the critical value at one tail that corresponds to the hypothesized direction in H_a , and the overall α is located at this tail, as shown in **Figure 3** (dashed line at the right tail for $H_a: \mu > K$). Nondirectional and directional alternative hypotheses are often referred to as two-tailed and one-tailed tests, respectively. For a directional (one-tailed) test, because all α is located at one tail, the critical value of the test statistic will be smaller in absolute value than that in a nondirectional test. For example, for a z -test with $\alpha = 0.05$, the critical values are ± 1.96 for a two-tailed test (nondirectional test), and the critical value for a one-tailed (directional) test ($H_a: \mu > K$) is only 1.65. This makes it easier for a one-tailed test to reject H_0 , as long as the test statistic is in the hypothesized direction. Because of this, if the test statistic is in the hypothesized direction, the one-tailed test has more statistical power (see description of statistical power below) over the two-tailed test. When the test statistic is *not* in the hypothesized direction, however, the test automatically fails to reject the H_0 .

Confidence Intervals in Hypothesis Testing

The hypothesis testing approach described above is a point-estimate approach that results in a dichotomous decision (i.e., reject or fail to reject H_0). On the other hand, interval-estimation approach provides a range of values within which a population parameter is likely to reside. The importance of using confidence intervals (CIs) in educational and psychological research has been emphasized by the American Psychological Association (APA) task force (Wilkinson and Task Force on Statistical Inference, 1999), and such emphasis has been reflected in

the guidelines in the APA publication manual as best practice in reporting empirical results (APA, 2001).

For a hypothesized population value θ , using the sample estimate $\hat{\theta}$, and for a given α , the following CI can be constructed:

$$100(1 - \alpha) C.I. = \hat{\theta} \pm T_{(1-\frac{\alpha}{2})} S_{\hat{\theta}}$$

(This is the classical definition of CI under true H_0 , where the critical values from a central distribution (e.g., central t distribution) are used to construct a symmetrical CI. However, if a true H_0 is not assumed, critical values from a noncentral distribution (e.g., noncentral t distribution) may be used for constructing the CI for the unknown population parameter θ , and this will result in a nonsymmetrical CI. For more details, readers may consult, e.g., Cumming and Finch (2001), Steiger and Fouladi (1997). where $T_{(1-\frac{\alpha}{2})}$ is the appropriate test statistic (e.g., z , t) value at the $100(1 - \frac{\alpha}{2})$ percentile point of its distribution, and $s_{\hat{\theta}}$ is the estimated standard error (i.e., $s_{\hat{\theta}} = \frac{s}{\sqrt{n}}$) of the sampling distribution. Adding and subtracting the quantity of $T_{(1-\frac{\alpha}{2})} s_{\hat{\theta}}$ from $\hat{\theta}$ results in the lower and upper limits of a $CI(\hat{\theta} - T_{(1-\frac{\alpha}{2})} s_{\hat{\theta}}, \hat{\theta} + T_{(1-\frac{\alpha}{2})} s_{\hat{\theta}})$ that has probability of $1 - \alpha$ of containing the unknown population parameter value θ . If the hypothesized population parameter value under H_0 is outside this CI, it leads to the rejection of H_0 . Conversely, if the CI contains the parameter value under H_0 , H_0 will not be rejected.

Using our previous example about a new instructional method, the researcher obtained $n = 100$, $\bar{X} = 68$, and $s = 8$ (i.e., $s_{\bar{X}} = \frac{s}{\sqrt{n}} = 0.8$). For $\alpha = 0.05$, $t_{(975, df = 99)} = 2.28$, the 95% CI would be (66.18, 69.82). (Lower limit is $68 - (2.28 \times 0.8) = 66.18$; upper limit is $68 + (2.28 \times 0.8) = 69.82$.) As this CI does not contain 72 under H_0 , the H_0 would be rejected. With a CI, we are not only

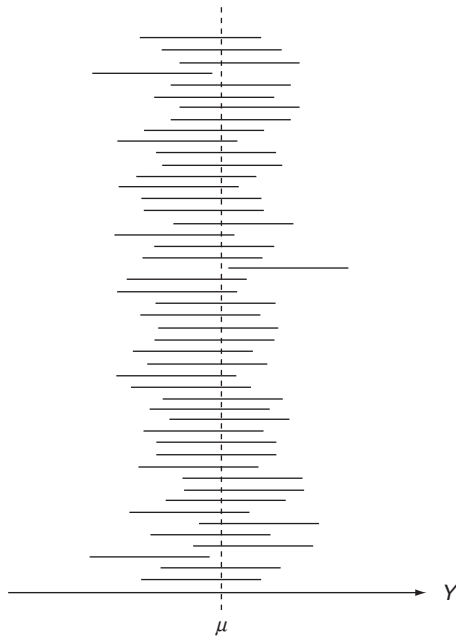


Figure 4 Fifty 95% confidence intervals for population parameter μ .

able to evaluate the H_0 (point estimation), but also able to estimate a range of probable population parameter values (interval estimation). The value of a CI is that, it provides information not only about the sample-based point-estimate, but also about the estimation error associated with the statistic. In the social sciences, the resulting estimation error may be large, and this may explain why such intervals are not often reported (Cohen, 1994).

It is important to note that this 95% CI does not indicate that we can be 95% confident that this interval contains the population parameter (see Thompson, 2007: 427). Rather, this is interpreted to mean that 95% of similarly constructed intervals will contain the population parameter, while 5% of them will not. This is illustrated in **Figure 4** for a hypothetical situation where 50 random samples were drawn from a population with parameter μ , and a 95% confidence interval was constructed from each sample. Out of the 50 CIs, 47 contained the population parameter, and three did not.

Statistical Errors (Type I and Type II) and Power

In hypothesis testing, two types of statistical errors may occur: reject H_0 when it is true (type I error), or fail to reject H_0 when it is false (type II error). On the other hand, a correct decision is made when a true H_0 is not rejected, or when a false H_0 is rejected. These possible outcomes in hypothesis testing are shown in **Figure 5**.

		Decision based on sample data	
		Reject H_0	Fail to reject H_0
True state of nature	H_0 true	Type I error α	Correct decision $1-\alpha$
	H_0 false	Correct decision (power) $1-\beta$	Type II error β

Figure 5 Hypothesis testing errors and correct decisions.

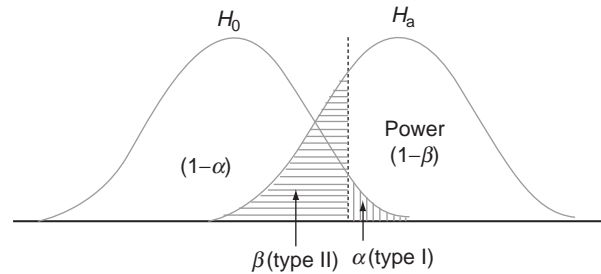


Figure 6 Type I, type II errors and statistical power.

In a hypothesis test, once a decision is made with regard to H_0 , either a type I error (when H_0 is rejected), or a type II error (when H_0 is not rejected), is possible, but not both. We cannot know with certainty whether the decision is actually an error. When H_0 is rejected, only a type I error is possible. The probability for a type I error is known, because it is directly controlled by the researcher through the choice of the α level (e.g., $\alpha = 0.05$), which represents the risk we are willing to take in concluding that the H_0 is false when it may be true. On the other hand, when H_0 is not rejected, only a type II error is possible. Here, the probability for a type II error is unknown, but can be estimated, and is generally not directly controlled by the researcher. In **Figure 5**, when H_0 is true (upper row), the probability for a type I error is α , and the probability of making a correct decision (i.e., not to reject the true H_0) is $1 - \alpha$. When H_0 is false (lower row), the probability for a type II error is β , and the probability of making a correct decision (i.e., to reject the false H_0) is the power of a statistical test ($1 - \beta$).

Figure 6 graphically illustrates two sampling distributions under H_0 and H_a , respectively, as well as the relationships among the probabilities of type I and type II errors, and the statistical power (probability to reject a false H_0). The sampling distribution under H_0 represents all possible values of a statistic, including those in the specified region of α (vertically shaded area under H_0 to the left of the vertical dashed line) where a decision to reject H_0 would be made. This probability is often set at 0.05 to indicate that there is a 5% chance of rejecting a true H_0 (i.e., making a type I error).

Now consider the sampling distribution under H_a , which overlaps with that under H_0 . Sample-based test statistics in the type II error or β region (horizontally shaded area under H_a to the right of the vertical dashed line) would not lead to the rejection of H_0 , even though these sample statistics may in fact belong to the sampling distribution of H_a . As shown in **Figure 6**, statistical power (i.e., the probability to reject a false H_0) denoted by $1 - \beta$, is the area under H_a to the right side of the vertical dashed line defined by α .

Education researchers may desire to have a reasonable level of power (e.g., $\beta \approx 0.80$) for a statistical test. One way to increase statistical power is to increase α (e.g., from 0.05 to 0.10). Imagine the rejection line (vertical dashed line) in **Figure 6** sliding to the left to accommodate a higher α level (i.e., higher probability of making a type I error), and this would result in reducing type II error probability (β) and in increasing statistical power ($1 - \beta$). Conversely, decreasing α (e.g., from 0.05 to 0.01; sliding the rejection line under H_0 distribution to the right) will result in an increase in type II error probability (β) and a decrease in statistical power. Statistical power can also be increased by using larger sample sizes. Other things being equal, larger sample sizes lead to smaller standard errors of the distribution. Consequently, there will be less overlap between the distribution under H_0 and that under H_a , thus a smaller β , and more power (i.e., $1 - \beta$ is larger). Finally, an increase in statistical power can also come by way of greater treatment effects or greater separation between

the H_0 and H_a distributions. Researchers need to understand such interplay among type I error, type II error, and statistical power, and consider the relative cost and consequence of one type of error versus another in a given substantive research application.

See also: Point Estimation Methods with Applications to Item Response Theory Models; Sampling; Statistical Power Analysis; Statistical Significance Versus Effect Size.

Bibliography

- APA (American Psychological Association) (2001). *Publication Manual of the American Psychological Association*, 5th edn. Washington, DC: Author.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist* **49**, 997–1003.
- Cumming, G. and Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and non-central distributions. *Educational and Psychological Measurement* **61**, 532–574.
- Steiger, J. H. and Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In Harlow, L. L., Mulaik, S. A., and Steiger, J. H. (eds.) *What if there were no significance tests?*, pp 221–257. Mahwah, NJ: Erlbaum.
- Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools* **44**, 423–432.
- Wilkinson, L. Task Force on Statistical Inference, American Psychological Association, Science Directorate (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist* **54**, 594–604.

Instrumental Variables

E P Bettinger, Stanford University, Stanford, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The problem of identifying causal impacts is central to educational research and policy analysis. As educators and researchers study educational phenomena or programs, they often are trying to identify specific actions or agents that cause improved outcomes in students, schools, and communities.

Identification of causal effects relies on our ability to estimate unbiased relationships between an outcome (i.e., the dependent variable) and an agent (i.e., the independent variable), and there are many threats to unbiased estimation. For example, selection bias, omitted variables, measurement error, and reverse causality (i.e., the dependent variable actually influencing the independent variable) can lead to biased estimation and faulty inference.

There are many techniques that can help researchers identify causal impacts. Experimental methods, selection-correction models, regression-discontinuity approaches, and instrumental variables (IVs) are just examples of common techniques that researchers can use to estimate unbiased relationships between an outcome and a causal agent.

This article aims to focus on one of these tools, IVs. Our goal is to provide a simple overview of IVs, its key assumptions, and its potential to produce unbiased estimates. Further, the article provides several examples of how IVs estimation has been applied to educational research. The article discusses some of the limitations of IVs and some of the issues and limitations that researchers should be cognizant of when employing IVs.

This article does not intend to produce a technical review of IVs. A number of textbooks (e.g., Wooldridge, 2006; Greene, 2008) provide a basic introduction to IVs and the underlying math. In addition, there are several other reviews that discuss the notion of causality (e.g., Angrist *et al.*, 1996), experimental and quasi-experimental approaches (e.g., Angrist and Krueger, 2001), and parametric approaches to overcoming selection (e.g., Heckman, 1979). The goal is to provide a simple introduction and primer to educational researchers who are beginning to understand and to utilize this promising tool.

Instrumental Variable Estimation

Inconsistent Estimation in Ordinary Least Squares

Suppose that the true relationship between academic achievement and private schooling is given by the following equation:

$$Y_i = a + bP_i + cA_i + u_i \quad [1]$$

where Y_i is the academic achievement of student i , P_i is an indicator for whether student i attends private schooling, A_i is the ability of student i , and u_i is the residual. If an econometrician observes measures of Y , P , and A , he/she could estimate eqn [1] in an unbiased way. However, ability is hard to measure, and an econometrician may not observe it.

The naive econometrician may push forward and estimate the following production function of education:

$$Y_i = a + bP_i + v_i \quad [2]$$

The ordinary least squares (OLS) estimator, in this case, will clearly be inconsistent so long as private schooling and ability are correlated. We can see the inconsistency if we take the probability limit of the OLS estimator as n approaches infinity.

$$\text{plim}[\hat{b}] = b + c^* \text{plim} \left[\frac{\sum_{i=1}^n P_i A_i}{\sum_{i=1}^n P_i^2} \right] \neq b \quad [3]$$

This is a classic and simple example of omitted variable bias. OLS yields inconsistent estimates due to the violation of the classical assumption that the independent variable is independent of the error term, or – in other words – as n approaches infinity, the correlation between P and A does not converge to zero.

How do we then estimate b in lieu of the missing variable? One solution is IVs. We must find another variable that meets two conditions. First, the new variable must influence P , and second, the new variable must be uncorrelated with A (or any other variable that may be part of u_i in eqn [1]). This variable is called an instrument.

Identification

Presented, in this article, is an example of omitted variable because – in educational phenomenon, generally speaking – omitted variable bias is often ignored although likely to be present. Most textbooks (e.g., Greene, 2008) focus on the more general case of simultaneous equation systems. Economists have long recognized the problem of estimating parameters in simultaneous equation systems. Haavelmo (1943), for example, demonstrated that simultaneous equation systems may be unsolvable without additional exogenous variation. This exogenous variation is variation in an independent variable that helps predict one or more of the dependent variables in the model. Suppose that we augmented the

true model in eqn [1] with another equation reflecting the decision to enter private schooling:

$$P_i = f + gZ_i + bA_i + w_i \quad [4]$$

where Z_i represents a variable that influences the likelihood that student i attends private school. Rouse (1998), for example, used the offer of an educational voucher program as the Z variable in an equation similar to eqn [4]. Neal (1997) uses access to Catholic schools as an instrument to find out whether students attended private schooling.

If we consider the system of equations in eqns [1] and [4], there are two endogenous variables – private schooling and student achievement. These variables are determined completely by the system of equations. There are also two exogenous variables – ability and the Z variable. These variables are taken as given and not influenced by the other variables in the model. Together, the system of equations would be referred to – in the academic literature – as the structural model. The system could be solved to make each endogenous variable a function only of the exogenous variables. This system of equations (eqns [5] and [6]) is known as the reduced form.

$$\begin{aligned} Y_i &= (a + bf) + (bg)Z_i + (c + bh)A_i + (u_i + bw_i) \\ &= j + kZ_i + mA_i + e_i \end{aligned} \quad [5]$$

$$P_i = f + gZ_i + bA_i + w_i \quad [6]$$

Equation [6] is identical to eqn [4] since the endogenous variable of private schooling was already a function of all of the exogenous variables in our model. (If either student achievement or another endogenous variable had influenced private schooling in eqn [4], then eqn [6] would have differed from eqn [4].) Equation [1], however, has now been transformed to include Z and to eliminate the other endogenous variable from the model.

In our structural model, eqn [1] shows student achievement as a function of another endogenous variable, private schooling. The coefficient on private schooling (b) is the key coefficient that we seek to identify. One might notice that the ratio of the coefficients on the Z variable in the reduced form would help us measure b . The method of using the ratio of coefficients to identify the structural coefficients is called indirect least squares.

When we refer to the state of identification of eqn [1], there are three possible states. The model could be unidentified, just identified, or over identified. The state of identification varies from equation to equation in the structural model and depends on a comparison of the number of included right-hand-side endogenous variables and the number of excluded exogenous variables. For example, eqn [1] has one right-hand-side endogenous variable (P) and one excluded exogenous variable (Z). When the number of included endogenous variables is equal to the number of excluded exogenous variables, then the

equation is just identified. If Z represented a matrix of excluded exogenous variables, then we would have more exogenous variables (i.e., the rank of Z) than included endogenous variables (P). Equation [1] would then be over identified. If Z did not enter the model, then the model would be unidentified.

IVs Estimation

For a variable to be a good instrument, there are two conditions that must hold. First, Z must be related to the endogenous variable we are trying to predict. For example, in eqn [4], we need the coefficient g not to be equal to zero. Asymptotically, we would say that

$$\text{plim}[Z'P] \neq 0 \quad [7]$$

There has been some discussion in the academic literature concerning the case where the eqn [7] is satisfied, but the covariance between Z and P is close to zero. This case of weak instruments is discussed below.

It can easily be tested whether eqn [7] is satisfied by estimating the first stage of the equation. The first stage of this discussion would be eqn [6]. This could estimate this regression and examine the coefficient g . There are a number of indicators in this first stage that the investigator could report to denote the fit. They could report the t -statistic on the instrument or the R-squared of the equation. The econometrics literature typically relies on the F -statistic testing the joint hypothesis that the coefficients on all of the instruments (i.e., the excluded exogenous variables from eqn [1]) are zero. An F -statistic over ten suggests that the instrument has good explanatory power and that eqn [7] is satisfied.

The second condition required in IV estimation is the exclusion restriction. This restriction is more limiting and generally untestable. The second restriction assumes that

$$\text{plim}[Z'v] = 0 \quad [8]$$

where v derives from eqn [2] and includes any variable that is not included in the regression that affects student achievement. In the present case, there is collapsed ability into the error term since ability is unobserved.

The exclusion restriction, essentially, means that the instrument (Z) does not directly affect the outcome of interest (Y) except through its influence on the right-hand-side endogenous variable (P). Because the exclusion restriction is untestable, it is an obvious source of debate in many papers. Defending an instrument is difficult, and authors, generally, need to think carefully through the instruments' potential correlations with other confounding variables. For example, Rouse (1998) used educational vouchers as an instrument for private school attendance. Angrist *et al.* (2002) were reluctant to do this since most of the families in their sample attended private school anyway. In their

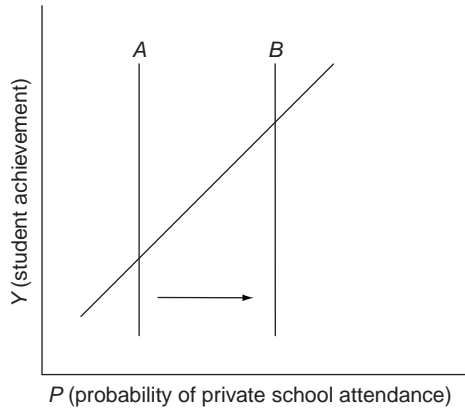


Figure 1 Graphical example of shift in the instrument.

case, the voucher – while correlated with private school attendance – was also a significant boost in the incomes for many families who would have sent their children to private schools irrespective of the voucher. The increase in incomes could have independently affected student achievement.

Concern over the exogeneity of instruments and the exclusion restriction has helped fuel movements toward experiments and natural experiments in economics (Angrist and Krueger, 2001). In the case of experiments, randomization often determines participation. By construction, randomization should be exogenous from the outcomes of interest while influencing the treatment under investigation.

If the instrument is related to the treatment of interest and uncorrelated with our outcome of interest, then the model can be estimated. The model essentially includes two stages. The first stage has been discussed above. In this step, the variation in Z is used to predict P . In the second stage, the predicted value of P is used to estimate the relationship between P and Y .

Figure 1 provides an illustration of the IVs set-up. The upward-sloping line maps the relationship between private school attendance and achievement. The slope is the underlying value of ' β ' in eqn [1]. It is not drawn upward sloping, but it could just as likely be negative sloping. The vertical line ' A ' represents the predicted probability that a person attends private school. Movements in Z cause the probability of private school attendance to increase – leading to a shift from line ' A ' to line ' B '. The outward shift causes variation in student achievement, which allows the measurement of the slope of relationship between achievement and private schooling. The first IV assumption assures that the probability of private schooling depends on the instrument – hence, the outward shift in the schedule. The exclusion restriction assures that only the private school schedule is shifting. If the exclusion restriction was not met and Z influenced student achievement, then the sloped line showing the relationship between P and Y

would have shifted outward as well, and it would have been impossible to measure the slope of this schedule.

In practice, the IVs estimate is defined as

$$\hat{b}_{IV} = (P'Z(Z'Z)^{-1}Z'P)^{-1}P'Z(Z'Z)^{-1}Z'Y \quad [9]$$

In the simple case where Z represents only one variable, **Equation 9** simplifies to

$$\hat{b}_{IV} = (Z'P)^{-1}Z'Y \quad [10]$$

The more general case, in eqn [9], is useful since it could incorporate a more complex system with several instruments.

The standard errors can easily be computed in a single step, assuming independence of the observations and no heteroskedasticity.

$$V(\hat{b}_{IV}) = \hat{\sigma}^2(P'Z(Z'Z)^{-1}Z'P)^{-1}. \quad [11]$$

As in other statistical applications, standard errors can be adjusted for clustering and heteroskedasticity.

Other Considerations in Estimation

Weak Instruments

Angrist and Krueger (1991) wrote one of the most influential papers utilizing IVs. Angrist and Krueger attempted to identify the effects of a high school education on individual's earnings. They recognized that there may be important omitted variables (e.g., ability) that prevent using OLS. Angrist and Krueger recognized that there was variation across states in the compulsory schooling laws and that there was variation in entry dates in school that was correlated with the quarter in which students were born. As a result, students in some states may acquire more years of education before they can drop-out than students in other states. The exact impact on the total number of years of schooling depends on the quarter of birth and the state of birth. Hence, the interaction between quarter of birth and individual state's of birth provides a set of instruments that can predict the number of years of completed schooling. The authors find that an additional year of schooling led to a 9% increase in income. Their use of census data gave them a large sample, and, thus, the estimate was quite precise.

Bound *et al.* (1995) focused on the weak correlation of the instruments in the Angrist and Krueger study with the endogenous regressor of interest (years of schooling). Bound and co-workers replicated Angrist and Krueger's results, but they also generated a random variable based on a similar distribution as the quarter of birth instrument. By construction, the random instruments are unrelated to the number of years of schooling. Nonetheless, the IVs estimates based on this exercise generated estimates that

were, often, statistically significant. The large samples used in these studies were the key. The significance in the first stage relied more heavily on the large-sample properties of the estimators than on the strength of the underlying economic relationships of the variables.

Staiger and Stock (1997) attempted to formalize some empirical rules for identifying the behavior of IV estimates. They argue that the first-stage F -statistic testing that all of the instruments are simultaneously equal to zero is the best indicator of potential weak-instruments problems. Through a number of Monte Carlo simulations, they find that an F -statistic greater than ten indicates that there is no weak-instrument problem. Without an F -statistic greater than ten, the IV estimates are biased toward the OLS estimates. The F -statistic greater than ten has become a rule of thumb for empiricists using IV. However, research on weak instruments and on improving estimation techniques in IV models with weak instruments remains an area of active exploration.

Local Average Treatment Effect (LATE) Interpretation

Another feature of the Angrist and Krueger (1991) paper was that the variation in years of completed schooling was being generated by a set of individuals who, generally, desired to withdraw from school during high school. Imbens and Angrist (1994) examined situations in which causal impacts were being generated by the behavior of a subset of individuals. Their work formalized the local average treatment effect (LATE) interpretation of IVs (see also Angrist *et al.*, 1996).

The key insight in this work is that there could be heterogeneous effects of a treatment on different sets of the population. IVs estimates show the average effect of treatments on individuals for whom the instrument influences individuals' behaviors (i.e., treatment compliers – in the Imbens and Angrist terminology). In the case of the Angrist and Krueger study, less than 2% of the population were compliers (i.e., years of education were influenced by compulsory schooling laws and quarter of birth) (Imbens, 2007). Given the small proportion of the population that generated these estimates, the estimated effects may not generalize to the overall population.

Specification Tests

While we can rarely test the exclusion restriction, researchers often employ Hausman specification tests in IVs models to test certain aspects of the models (Hausman, 1978, 1983). Hausman specification tests generally compare two estimates. Under the null hypothesis, both sets of estimates are consistent, but one is more efficient. Under the alternative hypothesis, only one set is consistent.

First, Hausman tests can be used to compare OLS and IV models. Under the null hypothesis, the OLS

assumptions are not violated. In this case, both OLS and IV yield consistent estimates, but OLS is more efficient. Under the alternative hypothesis of endogeneity, IV is consistent. In our simple example, one could estimate eqn [2] with OLS and IV (using Z as the instrument). The comparison of the coefficients and their corresponding standard errors gives the Hausman test.

Second, Hausman tests are often used to test over-identifying restrictions. Researchers can compare an IV model that is just identified to one where all of the available instruments are used in the estimation. If the IV estimates are similar, then the instruments seem to be valid. If the IV estimates are different, then some of the instruments may be invalid. The test does not shed light on which instruments are problematic and states only that some might be problematic.

Examples of IV Estimation in Education

Returns to Schooling

Economists have, for long, studied the effects of schooling on earnings. In the case of earnings, economists consider a simple model comparing earnings (Y) to years of schooling (S).

$$Y_i = a + bS_i + u_i \quad [12]$$

$$S_i = a + bZ_i + v_i \quad [13]$$

As Griliches (1977) describes, there are a number of reasons why estimating eqn [12] with OLS will lead to biased results. For example, there is an inherent omitted variable as ability is left out of eqn [12] and is, likely, correlated with students' schooling choices. There could also be endogeneity in the schooling and earnings decisions since families are optimizing.

Card (2001) reviews papers published in the 1990s using IVs to estimate the effects of schooling on earnings. To overcome the biases described by Griliches (1977), the studies reviewed by Card utilized characteristics of the schooling system as instruments for the years of schooling. Specific features of schooling systems seemingly affected the overall years of schooling completed and may have been uncorrelated with earnings (see also the discussion in Kling, 1999).

The instruments in the studies reviewed by Card (2001) are as follows:

Interaction of quarter of birth with state and year of birth

As discussed above, this was the instrument used by Angrist and Krueger (1991) and subsequently investigated by Bound *et al.* (1995) and Staiger and Stock (1997). Bound *et al.* (1995) discuss limitations to this

instrument including the case for why this instrument may not satisfy the exclusion restriction.

Distance to school

There have been several permutations of this instrument. Kane and Rouse (1993) used the distance to the nearest college. Card (1995) and Conneely and Uusitalo (1997) used the presence of a university in a specific year. The main idea is that students living closer to schools have more access to them and complete more schooling (see also Maluccio, 1997).

Compulsory school laws

This certainly played a part in Angrist and Krueger (1991). It has also been used by Harmon and Walker (1995). Compulsory schooling laws create variation in the number of years completed (see also Devereux and Hart, 2008).

Military service or benefits

The most prominent paper in this vein Angrist (1992) actually focuses on the effects of military service on earnings, using the Vietnam draft lottery as an instrument. More recent studies (e.g., Angrist and Chen, 2007) have shown that the impact on earnings, likely, went through the impact of military service on educational choices that were influenced by G. I. Bill Benefits. Military experience has also factored into the instruments used in Ichino and Winter-Ebmer (1998) and Lemieux and Card (1998).

Other institutional features

Many authors have identified other shifts in schooling caused by natural experiments. Meghir and Palme (1999) used a change in the schooling system. Duflo (1999) used geographic indicators for regions with targeted school-building programs in Indonesia.

While our list and then number of studies we cite are clearly not all inclusive, the range of instruments should provide some examples of the instruments used to predict schooling choices.

Class Size

Another area where economists have used IVs to identify the impacts of a particular educational intervention is the effects of class size on student outcomes. While class size (C) may affect student achievement (Y), class size and student achievement may be endogenously determined.

$$Y_i = a + bC_i + u_i \quad [14]$$

$$C_i = a + bZ_i + v_i \quad [15]$$

Maimonides' rule

In recent years, the most popular IV for class size has been Maimonides-type rules. Angrist and Lavy (1999) use

Maimonides' rule as an instrument for class size. Maimonides' rule was established in the twelfth century. It stated that class sizes should not exceed 40; hence, in Israel today, schools hire an extra teacher when the class size reaches 40. The rule creates a systematic relationship between enrolment and class size. The Maimonides' rule is likely exogenous – or in other words – it can only impact achievement through its effect on class size. Maimonides-like rules have been used to measure class-size effects in other places such as France (Gary-Bobo and Mahjoub, 2006) and Bolivia (Urquiola, 2001).

Tennessee STAR

A series of papers were written, in the late 1990s, which examined Tennessee's class-size experiment. Schanzenbach (2007) reviews these studies. The Tennessee STAR experiment used randomization to determine the size of classes that individual students attended. While some studies have expressed some worry with regard to either the fidelity of the randomization or the possibility of Hawthorne effects (e.g., Hoxby, 2000), the randomization in STAR is an important instrument being used to measure the effects.

Demographic variation

Hoxby (2000) uses variation in the population. She examines a set of schools where year-to-year variation leads to discrete shifts in the number of classrooms used. Bettinger and Long (2008) use variation in the number of students actually enrolling in college with interest in a specific major as an instrument for class sizes in that major during a particular year.

As in the previous case, the number of studies we cite merely scratches the surface and provides some examples of the instruments used to predict class sizes.

Conclusion

This article is intended to be an introduction to IVs. There are several excellent textbooks (e.g., Greene, 2008) or literature reviews (e.g., Angrist and Krueger, 2001) where the interested reader can learn more. The spatial constraints in this volume as well as the vastness of the technique, its applications, and its interpretation prevent us from a more expansive review.

In education, the surge of papers using IVs has increased dramatically in the last two decades. As Griliches (1977: 2) quipped in a different setting, the number of papers employing IVs has become "a vast river of econometric studies threatening to engulf us all." Caution is always warranted in assuring that instruments truly meet the strict exogeneity assumptions and in the interpretation of these estimates.

Bibliography

- Angrist, J. D. (1992). Estimating the payoff to schooling using the Vietnam-era draft lottery. *NBER Working Paper #4067*.
- Angrist, J. D., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment. *American Economic Review* **92**(5), 1535–1558.
- Angrist, J. D. and Chen, S. (2007). Long-term effects of Vietnam-era conscription: Schooling, experience, and earnings. *NBER Working Paper 13411*.
- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* **91**(434), 444–455.
- Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* **106**(4), 979–1014.
- Angrist, J. D. and Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives* **15**, 69–85.
- Angrist, J. D. and Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* **114**(2), 533–575.
- Bettinger, E. and Long, B. (2008). *Does Size Matter? Class Size in College*, mimeo. Stanford, CA: Stanford University.
- Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* **90**(430), 443–450.
- Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling. In Christofides, L. N., Grant, E. K., and Swidinsky, R. (eds.) *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, pp 201–222. Toronto: University of Toronto Press.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica* **69**(5), 1127–1160.
- Conneely, K. and Uusitalo, R. (1997). *Estimating Heterogeneous Treatment Effects in the Becker Schooling Model*. Princeton University Industrial Relations Section.
- Devereux, P. J. and Hart, R. A. (2008). Forced to be rich? Returns to compulsory schooling in Britain. *CEPR Discussion Paper 6679*.
- Duflo, E. (1999). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *NBER Working Paper Number XXXX*.
- Gary-Bobo, R. J. and Mahjoub, M. B. (2006). Estimation of class-size effects, using 'Maimonides' rule: The case of French junior high schools. *CEPR Discussion Paper 5754*.
- Greene, W. (2008). *Econometric Analysis* (3rd edn, 1997, Prentice Hall), 6th edn. New York: Pearson-Prentice Hall.
- Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica* **45**(1), 1–22.
- Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica* **11**(1), 1–12.
- Harmon, C. and Walker, I. (1995). Estimates of the economic return to schooling for the United Kingdom. *American Economic Review* **85**(5), 1278–1286.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica* **46**(6), 1251–1271.
- Hausman, J. A. (1983). Specification and estimation of simultaneous equation models. In Griliches, Z. and Intriligator, M. (eds.) *Handbook of Econometrics*, vol. 1, pp 391–448. Amsterdam: North-Holland.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* **47**(1), 153–162.
- Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics* **4**, 1239–1285.
- Ichino, A. and Winter-Ebmer, R. (1998). The long-run educational cost of World War II: An example of local average treatment effect estimation. *Center for Economic Policy Research Discussion Paper No. 1895*.
- Imbens, G. W. (2007). Instrumental variables with treatment effect heterogeneity: Local average treatment effects. *NBER Summer Institute Presentation*. http://www.nber.org/~confer/2007/si2007/WNE/lect_5_late_fig.pdf (accessed September 2009).
- Imbens, G. W. and Angrist, J. S. (1994). Identification and estimation of local average treatment effects. *Econometrica* **62**(2), 467–475.
- Kane, T. J. and Rouse, C. E. (1993). Labor market returns to two- and four-year colleges: Is a credit a credit and do degrees matter? *NBER Working Paper Number 4268*.
- Kling, J. (1999). Interpreting instrumental variables estimates of the returns to schooling. *Princeton University Industrial Relations Section Working Paper No. 415*.
- Lemieux, T. and Card, D. (1998). Education, earnings, and the 'Canadian G.I. Bill'. *National Bureau of Economic Research Working Paper Number 6718*.
- Maluccio, J. (1997). Endogeneity of schooling in the wage function. Unpublished Manuscript, Yale University Department of Economics.
- Meghir, C. and Palme, M. (1999). Assessing the effect of schooling on earnings using a social experiment. Unpublished Working Paper, University College London.
- Neal, D. (1997). The effects of catholic secondary school on educational achievement. *Journal of Labor Economics* **15**(1), 98–123.
- Rouse, C. E. (1998). Private school vouchers and student achievement: An evaluation of the Milwaukee Parental Choice Program. *Quarterly Journal of Economics* **113**(2), 553–602.
- Schanzenbach, D. (2007). What have researchers learned from project STAR? *Brookings Papers on Education Policy* **2006/07**, 205–228.
- Staiger, D. S. and Stock, J. H. (1997). Instrumental variables with weak instruments. *Econometrica* **65**(3), 557–586.
- Urquiola, M. (2001). Identifying class size effects in developing countries: Evidence from rural schools in Bolivia. *World Bank, Policy Research Working Paper 2711*.
- Wooldridge, J. (2006). *Introductory Econometrics*. Mason, OH: South-Western Publishing.

Further Reading

- Angrist, J. D. and Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* **90**(430), 431–442.

Jackknife Methods

S Sinharay, ETS, Princeton, NJ, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bias – Deviation between the mean of a sample estimate and the true population value.

Sample estimate – A function of sample data used for approximating a population parameter.

Standard error – The square root of the average squared deviation between a true parameter value and its sample estimate.

The jackknife method is one of the most popular resampling methods. The method was suggested by Quenouille (1949) and further developed by Tukey (1958).

If the statistical model is too complex or not adequately known, it is often difficult to use a theoretical approach to compute the standard error or bias of a statistical estimate $\hat{\theta}$. The jackknife method or jackknifing may be used in such situations to estimate one or both of the bias and standard error of the estimate. The basic idea behind the jackknife method lies in systematically recomputing the estimate leaving out one observation or a group of observations at a time from the sample available to the investigator. The bias and the standard error of $\hat{\theta}$ can be computed from this new set of recomputed estimates (also called jackknife replications).

Description of the Jackknife Method

This section describes the jackknife method for the simple case when the data set contains values of one variable. Denote the available sample (of size n) as $\mathbf{x} = (x_1, x_2, \dots, x_n)$. For example, x_j could be the score of Student j on an educational test. Suppose that the interest is in estimating a parameter θ (e.g., θ could be the mean test score for the student population of interest) and the estimate of θ computed from the sample is denoted as $\hat{\theta}$ (e.g., $\hat{\theta}$ could be the sample mean). Suppose that the interest is in estimating the bias and standard error of $\hat{\theta}$. (Note that the bias and standard error of the sample mean is easy to compute theoretically; we are using this simple example only to illustrate the jackknife method.) The jackknife method creates jackknife samples $\mathbf{x}_{(i)}$ that leave out one observation at a time: $\mathbf{x}_{(i)} = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ for $i = 1, 2, \dots, n$. In other words, the method leaves out the first sample observation to create

the first jackknife sample $\mathbf{x}_{(1)} = (x_2, x_3, \dots, x_n)$, then leaves out the second observation to create the second jackknife sample $\mathbf{x}_{(2)} = (x_1, x_3, x_4, \dots, x_n)$, and so on to finally leave out the last observation to create the n th jackknife sample $\mathbf{x}_{(n)} = (x_1, x_2, \dots, x_{n-1})$. Then the jackknife method computes the value of $\hat{\theta}$ for each of the jackknife samples. That is, it computes $\hat{\theta}_{(i)}$, $i = 1, 2, \dots, n$, where $\hat{\theta}_{(i)}$, which is called the i th jackknife replication, denotes the value of $\hat{\theta}$ computed from the i th jackknife sample $\mathbf{x}_{(i)}$. For example, when $\hat{\theta}$ is the sample mean, $\hat{\theta}_{(i)}$ is the mean of $\mathbf{x}_{(i)}$.

Let us denote

$$\hat{\theta}_{(\cdot)} = \frac{1}{n} \sum_i \hat{\theta}_{(i)}$$

Then, the jackknife estimate of bias of $\hat{\theta}$ is given by

$$\widehat{\text{bias}}_{\text{jack}} = (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta}).$$

The jackknife estimate of the standard error of $\hat{\theta}$ is given by

$$\widehat{\text{se}}_{\text{jack}} = \left[\frac{n-1}{n} \sum_i (\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2 \right]^{1/2}.$$

Thus, the jackknife estimate of the bias is a multiple of the average of the jackknife deviations

$$\hat{\theta}_{(i)} - \hat{\theta}, i = 1, 2, \dots, n,$$

while the jackknife estimate of the standard error is a multiple of the sample standard deviation of the jackknife replications of the estimate $\hat{\theta}$ (see, for example, Efron and Tibshirani (1993), for a theoretical explanation of why the above bias and standard error estimates work well in many situations).

The quantity

$$\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{(i)}$$

is sometimes called the i th pseudo-value (e.g., Efron and Tibshirani, 1993). It can easily be shown that

$$\widehat{\text{se}}_{\text{jack}} = \left[\frac{1}{n(n-1)} \sum_i (\tilde{\theta}_i - \bar{\tilde{\theta}})^2 \right]^{1/2},$$

where $\bar{\tilde{\theta}} = \sum_i \tilde{\theta}_i / n$, so that $\widehat{\text{se}}_{\text{jack}}$ looks like an estimate of the standard error of the mean for data $\tilde{\theta}_i$, $i = 1, 2, \dots, n$. Thus, the pseudo-values act as if they are n independent observations.

Example 1: Population Mean

If θ is the population mean, $\hat{\theta} = \bar{x}$, the sample mean. The i th pseudo-value in this case is the same as x_i , the i th sample observation. In this case, $\hat{\theta}_{(\cdot)} = \bar{x}$ so that $\widehat{\text{bias}}_{\text{jack}} = 0$, which is expected as the sample mean is an unbiased estimator of the population mean. For this case,

$$\widehat{\text{se}}_{\text{jack}} = \left[\frac{1}{n(n-1)} \sum_i (x_i - \bar{x})^2 \right]^{1/2},$$

which is exactly the same as the unbiased estimate of the standard error of the mean.

Example 2: Population Variance

If θ is equal to the population variance $\sigma^2 = \int (x - E(x))^2 dF(x)$ and $\hat{\theta}$ is the sample variance $\sum_i (x_i - \bar{x})^2 / n$, then $\hat{\theta}$ has a bias of $-\sigma^2/n$. In this case,

$$\widehat{\text{bias}}_{\text{jack}} = -\frac{1}{n(n-1)} \sum_i (x_i - \bar{x})^2,$$

which is the unbiased estimate of the population variance multiplied by $-1/n$.

Jackknife Method for Multiple Variables

Often, multiple variables are observed for each individual, for example, in a study of the correlation coefficient among two variables or the regression of one variable on several other variables. In this case, the jackknife method proceeds in a similar manner as described in the case of one variable. The method leaves out all observations from one individual at a time and computes the estimate of interest based on all the information on the other $(n-1)$ individuals. Finally, the bias and standard error are computed as described above.

Example: Correlation and Regression

Table 1 shows the scores of 15 examinees on the Reading and Listening sections of an English test. The score is the number of correct answers in 50 multiple choice questions. The correlation coefficient between the two scores is 0.733. The jackknife estimate of bias of the correlation coefficient is -0.003 , which is somewhat smaller in magnitude than the estimate of the bias (-0.014) from theoretical calculations (Olkin and Pratt, 1958). The jackknife estimate of the standard error of the correlation coefficient is 0.141. In a linear regression of the listening score on the reading score, the regression coefficient for the reading score is 0.52 with a corresponding standard error estimate of 0.13. The jackknife estimate of the bias and standard error of the regression coefficient is 0.02 and 0.18, respectively.

Table 1 Scores of 15 examinees on the Reading and Listening sections of an English test

Examinee	Listening	Reading
1	35	31
2	46	35
3	36	30
4	36	22
5	49	37
6	41	30
7	40	21
8	41	33
9	45	34
10	44	31
11	33	21
12	39	15
13	49	42
14	35	25
15	36	22

Deleted- d and Grouped Jackknife

The jackknife method described above is also called the leave-one-out jackknife or deleted-1 jackknife as it leaves or deletes one observation at a time to obtain the jackknife samples. It is possible to leave out d observations (where $d > 1$) at a time to obtain deleted- d or deleted- d or delete- k jackknife samples. The size of the jackknife samples in this case is $(n-d)$ and there are $\binom{n}{d}$ such jackknife samples. Let $\hat{\theta}_s$ denote the estimate of $\hat{\theta}$ computed from the data set after subset s is removed. The formula for the deleted- d jackknife estimate of standard error is

$$\widehat{\text{se}}_{\text{jack}} = \left[\frac{n-d}{d \binom{n}{d}} \sum_s (\hat{\theta}_s - \hat{\theta}_{(\cdot)})^2 \right]^{1/2},$$

where $\hat{\theta}_{(\cdot)} = \sum_s \hat{\theta}_s / \binom{n}{d}$.

The deleted- d jackknife is often preferred over the deleted-1 jackknife method (Efron and Tibshirani, 1993; Shao and Tu, 1995). One such situation is the application of nonsmooth statistics. (When small changes in the data set cause only small changes in a statistic, it is called a smooth statistic. An example of such a statistic is the sample mean. An example of a nonsmooth statistic is the sample median.) For example, a deleted- d jackknife estimate is better (in terms of achieving consistency) than the deleted-1 jackknife estimate for the sample median.

Usually, d should be more than the square root of n . When n is large, the number of jackknife samples can be large. A practical choice will be not to consider all $\binom{n}{d}$ deleted- d jackknife samples, but divide the observations into roughly $\frac{n}{d}$ groups of size d and leave out one group at a time to obtain roughly $\frac{n}{d}$ jackknife samples. This variation of the deleted- d jackknife is referred to as the grouped jackknife method (e.g., Miller, 1964). The grouped jackknife method is stable, although slightly

less efficient, when compared to the deleted- d jackknife method, but it is much less computationally intensive for large samples (Miller, 1964, 1968). Another option often recommended when n is large is to choose a random sample of the jackknife samples.

Jackknife versus Bootstrap

Both the bootstrap and the jackknife are examples of resampling methods and estimate the variability of a statistic from the variability of that statistic across subsamples, rather than from parametric assumptions. The jackknife is easier to apply to complex sampling schemes, such as multistage sampling with varying sampling weights, than the bootstrap. On the other hand, the bootstrap is usually less computationally intensive for large samples.

The jackknife and bootstrap may often yield similar results. But when used to estimate the standard error of a statistic, the bootstrap gives slightly different results when repeated on the same data, whereas the jackknife gives exactly the same result each time. This is because the bootstrap involves drawing random samples from the sample observations while the jackknife does not involve any randomness.

Application of the Jackknife Method to the National Assessment of Educational Progress

The National Assessment of Educational Progress (NAEP; see, e.g., von Davier *et al.*, 2007) is an ongoing survey of the performance of school students in the U.S. in a number of subject areas, including reading, writing, and mathematics. NAEP fits a complicated item response theory model to the data and computes the estimated proficiencies of several demographic groups (such as Black students and female students) in several subject areas and the corresponding variance estimates. It is not straightforward to compute the variance estimates using standard statistical theory. Jackknife estimates of sampling variance form a portion of the variance estimates.

For example, in NAEP 1998, 62 pairs of primary sampling units (PSUs) were involved. A PSU usually refers to a geographical area in the US. The pairs were composed so that PSU pairs were relatively similar in socioeconomic makeup. In turn, a pair was chosen from the 62 PSU pairs, a single PSU was dropped from the pair, and the estimate (of the mean proficiency of a demographic group on a subject area) was calculated. Let $\hat{g}_G^{(l)}$ denote the estimate obtained when a PSU is dropped from l th PSU pair.

The NAEP jackknife estimate of sampling variance for the group and subject area, denoted V_G , is defined as

$$V_G = \sum_{l=1}^{62} \left(\hat{g}_G^{(l)} - \tilde{g}_G \right)^2$$

where \tilde{g}_G is the estimate of the group mean from the full sample.

For other applications of the jackknife method to educational research, see von Davier *et al.* (2008), Qian (2006), and Longford (1992).

Concluding Remarks

The jackknife method is a popular resampling method that provides estimates of the bias and a standard error of an estimate by recomputing the estimate from subsamples of the available sample. The method has some similarities to the bootstrap method, but may provide different results from bootstrap in real applications. The method is relatively straightforward and hence can be useful in applications where theoretical statistics cannot readily provide estimates of the bias or standard error of an estimate. Such an example, of an application of the jackknife method to NAEP, is briefly described in this article.

See also: Bootstrap Method; Measure of Association; Sampling; Univariate Linear Regression.

Bibliography

- Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Longford, N. T. (1992). Comparison of efficiency of jackknife and variance component estimators of standard errors. *ETS Research Rep. No. RR-92-24*. Princeton, NJ: ETS.
- Miller, R. G. (1964). A trustworthy jackknife. *Annals of Mathematical Statistics* **35**, 1594–1605.
- Miller, R. G. (1968). The jackknife – a review. *Biometrika* **61**, 1–15.
- Olkin, I. and Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *Annals of Mathematical Statistics* **29**, 201–211.
- Qian, J. (2006). Weighting procedures and the cluster forming algorithm for delete-k jackknife variance estimation for institutional surveys. *ETS Research Rep. No. RR-06-15*. Princeton, NJ: ETS.
- Quenouille, M. H. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society, Series B* **11**, 68–84.
- Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer.
- Tukey, J. W. (1958). Bias and confidence in not-quite large sample. *Annals of Mathematical Statistics* **29**, 614.
- von Davier, A. A., Manalo, J. R., and Rijmen, F. (2008). Theoretical and empirical standard errors for two population invariance measures in the linear equating case. *ETS Research Rep. No. RR-08-24*. Princeton, NJ: ETS.
- von Davier, M., Sinharay, S., Oranje, A., and Beaton, A. (2007). The statistical procedures used in national assessment of educational progress: Recent developments and future directions. In Rao, C. R. and Sinharay, S. (eds.) *Handbook of Statistics*, vol. 26, pp 1039–1055. Amsterdam: Elsevier.

Large-sample Statistical Methods

J M Singer, Universidade de São Paulo, São Paulo, Brazil

D F Andrade, Universidade Federal de Santa Catarina, Florianopolis, Brazil

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In a broad sense, the objective of statistical inference is to draw conclusions about some characteristics of a (possibly conceptual) population of interest based on the information obtained from a sample conveniently selected therefrom. Most commonly, these characteristics correspond to parameters, such as the mean or the variance, but they may also involve more general features of the population, such as the associated (frequency) distribution itself. In general, the strategy for inference involves the selection of an appropriate family of stochastic models, an evaluation of their compatibility with the available data (goodness of fit) and the subsequent estimation of or conduction of tests of hypotheses about some components of the chosen family of models. Such models may have different levels of complexity and depend on assumptions with different degrees of restrictiveness, as illustrated by the following examples.

Example 1 Suppose that the proficiency of students in some subject (e.g., mathematics) is measured by the outcome in a given test and that we are interested in comparing the performance of males and females based on data obtained from independent random samples. We may assume that the outcome corresponds to a score Y for which the underlying distribution function F is normal with common variance σ^2 and mean μ_1 for males or μ_2 for females; the data correspond to the scores $Y_{i1}, \dots, Y_{in_i}, i = 1, 2$, obtained from the n_1 male and n_2 female students in the sample. Alternatively, we may consider the outcome to be the classification of the test result for each student as failed, acceptable, good, or excellent so that the data correspond to the frequencies $n_{ik}, i = 1, 2, k = 1, 2, 3, 4$ of male and female students classified in each category. Here, a product multinomial model with parameters π_{ik} denoting the probabilities of classification in the k th response category for students in the i th group, is appropriate.

Example 2 In cases where the test is composed of L dichotomous or polytomous questions or items sampled from a large item population, 'item response theory' models are often appropriate to measure the proficiency of students (Hambleton *et al.*, 1991). In this context, the probability that a student answers an item correctly is modeled as a function of his/her proficiency (a latent trace) and of some characteristics of the item. For example, the one-parameter logistic model is

$$\pi_{lj} = \pi_{lj}(\theta_j, b_l) = P(U_{lj} = 1 | \theta_j, b_l) = \frac{1}{1 + e^{-(\theta_j - b_l)}}$$

where π_{lj} is the conditional probability of a correct answer to item l given by student j , U_{lj} is a Bernoulli random variable with probability of success π_{lj} , θ_j is the proficiency of student j , and b_l is the location (or difficulty) parameter for item l . The data consist of the values of $U_{lj}, j = 1, \dots, n, l = 1, \dots, L$ corresponding to the responses of all sampled students to all sampled items.

In Example 1, it is well known that the Student t -statistic

$$t = (\bar{Y}_1 - \bar{Y}_2) / (S \sqrt{1/n_1 + 1/n_2})$$

where $\bar{Y}_i = n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}, i = 1, 2$, respectively, denote the sample mean score for males and females and $S^2 = [n_1 + n_2 - 2]^{-1} \sum_{i=1}^2 \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$ represents the sample variance, follows an exact t -distribution with $n_1 + n_2 - 2$ degrees of freedom under the hypothesis that $\mu_1 = \mu_2$ and thus, may be employed to construct an exact test of such hypothesis. In the second setup, to test for the equality of the response distributions of males and females, we usually employ Pearson's χ^2 -statistic $Q = \sum_{i=1}^2 \sum_{k=1}^4 (n_{ik} - e_{ik})^2 / e_{ik}$, where $e_{ik} = (n_{i.} n_{.k}) / n, n_{i.} = \sum_{k=1}^4 n_{ik}$ and $n_{.k} = \sum_{i=1}^2 n_{ik}$. Unfortunately, the exact distribution of Q is difficult to specify, especially when the sample sizes are large, and therefore, we must rely on an approximate χ^2 distribution to construct the test.

In Example 2, assuming that the answers of student j to the L items, conditionally on his/her proficiency, are independent, θ_j and b_l can be estimated by maximum likelihood methods, that is, by maximizing the log-likelihood function

$$l(\theta, b) = \sum_{j=1}^n \sum_{l=1}^L \{u_{lj} \log[\pi_{lj}(\theta_j, b_l)] + (1 - u_{lj}) \log[1 - \pi_{lj}(\theta_j, b_l)]\}$$

where θ and b are vectors whose elements are the proficiency and item parameters, respectively, and u_{lj} denotes the realization of U_{lj} . Here, the estimating equations

$$\sum_{j=1}^n [u_{lj} - \pi_{lj}(\hat{\theta}_j, \hat{b}_l)] = 0 \text{ and } \sum_{l=1}^L [u_{lj} - \pi_{lj}(\hat{\theta}_j, \hat{b}_l)] = 0$$

do not have a closed form solution (Baker and Kim, 2004), and the maximum likelihood estimators $\hat{\theta}_j$ and \hat{b}_l must be

obtained by numerical algorithms. Therefore, the quest for approximations to the distributions of such estimators is even more pressing.

With the exception of some simple cases, the exact statistical properties of many useful estimators and test statistics are not obtainable for many models of practical interest. Therefore, the methods for approximating the probability distributions (or some appropriate summary measures) of such statistics are of special concern. The main tools for such approximations are based on statistical properties derived by letting the sample size n be indefinitely large and are investigated under the denomination of asymptotic statistical theory. Since, in practice, the asymptotic results generate methods that must be applied to samples of finite (although large) sizes, they are also known as large-sample methods and, therefore, must be considered as approximations. Nevertheless, they are associated with the following advantages:

- They allow for some flexibility in the selection of models since large-sample properties of estimators and test statistics are usually less dependent on the particular functional form of the underlying distribution than their exact (small-sample) counterparts, and thus, are relatively robust against departures from some assumed model.
- They allow for some relaxation of the assumptions of independence and identical distribution of the sample elements, usually required by exact inferential procedures.
- They generally produce simple and well-studied limiting distributions such as the normal, χ^2 , Weibull, etc.

It is quite natural to expect that as the sample size increases, an estimator should be closer to the parameter it is set to estimate; this property is known as consistency and is intimately related to the concept of stochastic convergence of random variables. Similarly, a test statistic should be able to detect false null hypotheses with increasing confidence when the sample size becomes large. A second natural requirement for large-sample procedures is that the corresponding exact sampling distribution can be adequately approximated by a simpler one, such as the normal or χ^2 distribution, for which tables or computational algorithms are available. This is known as convergence in distribution or weak convergence and constitutes the basis of some important results known as central limit theorems. Third, in a given setup, there are usually many competing statistics satisfying the requirements of consistency and convergence in distribution. In choosing an appropriate one within such a class, we seek for optimality in some sense, such as minimum variance or minimum risk with respect to suitable loss functions in estimation problems, and this motivates the concepts of asymptotic efficiency. Similarly, in testing problems, large sample optimality properties are defined

in terms of alternative hypotheses that become closer to the null hypothesis as the sample size increases.

In this context, we consider the two basic forms of stochastic approximation useful for large samples; the first concerns the approximation of a parameter or possibly a random variable by a sequence of statistics, and the second is related to the approximation of a sequence of distribution functions by another distribution function with known properties.

Stochastic Convergence

We say that a sequence $\{T_n\}$ of real numbers converges to a limit T as n becomes indefinitely large, if for every positive ε , there exists a positive integer n_0 , possibly depending on ε , such that (Apostol, 1974)

$$|T_n - T| < \varepsilon \text{ for } n \geq n_0 \quad [1]$$

This may be equivalently expressed as

$$\sup_{N \geq n} |T_N - T| < \varepsilon \text{ for } n \geq n_0 \quad [2]$$

In dealing with a sequence $\{T_n\}$ of random elements (statistics), we note that no matter how large an n is chosen, the inequality $|T_n - T| < \varepsilon$ may not hold always when n is sufficiently large but may do so with probability arbitrarily close to 1. Furthermore, in the stochastic context, [1] and [2] are not equivalent and the difference in the probabilities of these two sets of events leads to the concepts of convergence in probability, and almost sure convergence, which essentially address different degrees of stochastic approximation. The latter is stronger and, in general, more stringent conditions are required for its validity. With a similar spirit, we may consider the concepts of convergence in the r th mean and complete convergence (Sen *et al.*, 2010). In practice, these concepts are usually employed to show that certain sequences of statistics, such as maximum likelihood or least-squares estimators, are consistent.

A second form of approximation refers to convergence in distribution or weak convergence. Here, we are not concerned with the convergence of the actual sequence of statistics $\{T_n\}$ to some constant or random variable T , but with the convergence of the corresponding distribution functions $\{G_n\}$ to some specific distribution function F . Thus, we say that the sequence of statistics $\{T_n\}$ converges in distribution (or weakly) to T if, for every given $\varepsilon > 0$, there exists an integer n_0 , possibly depending on ε , such that at every point of continuity (x) of F

$$|G_n(x) - F(x)| < \varepsilon, \text{ for } n \geq n_0$$

Although this mode of convergence is the weakest among those mentioned so far, it is very important for statistical applications, since the related limiting distribution function F may generally be employed in the construction of

approximate confidence intervals for and significance tests about the parameters of interest.

Classical applications of convergence results are focused on statistics that may be expressed as sums of functions of independent random variables as the sample means \bar{Y}_i in Example 1. The most important results on the convergence in probability or almost sure convergence of such statistics to the population parameters, are known as laws of large numbers (LLNs). As an example, the Khintchine weak LLN states that the sample mean (\bar{Y}_i) converges in probability to the (finite) population mean (μ_i), provided that the sample elements are independent and identically distributed. Essentially, this means that the probability that \bar{Y}_i differs from μ_i by less than a specified (small) constant can be made arbitrarily closer to 1 by increasing the sample size. The Khintchine strong LLN shows that statistics like \bar{Y}_i converge almost surely to μ_i under the same conditions. Similar results are available even if the underlying random variables are not identically distributed; in general, the price to pay is to require the existence of moments of higher order. This is the case with the Markov weak LLN or with the Kolmogorov strong LLN. By choice of an appropriate norm, we may also extend the results to vector-valued random variables (Sen *et al.*, 2010).

The proofs of such results generally rely on some probability inequalities, the most simple being the well-known Chebyshev inequality: given a sequence $\{T_n\}$ of random variables with mean μ and variance σ_n^2 , it follows that for all $n \geq 1$ and all real numbers t ,

$$P(|T_n - \mu| < t) \geq 1 - \sigma_n^2/t^2$$

Then, we conclude that whenever σ_n^2 converges to 0, T_n converges in probability to μ . To obtain stronger results, we need sharper inequalities such as the Bernstein or Hoeffding inequalities, for example. Although these inequalities were originally developed under the assumption that the underlying random variables are independent, they may be extended to more general cases provided some martingale-type structure (Chow and Teicher, 1978) can be imposed on their dependence pattern. Likelihood ratio test statistics, U -statistics, or empirical distribution functions, have such properties (Sen *et al.*, 2010).

Asymptotic Distributions

The main result on this topic, known as the central limit theorem (CLT), essentially demonstrates that statistics expressed as sums of the underlying random variables, conveniently standardized are asymptotically normally distributed, that is, converge weakly to the normal distribution. The CLT may be proved under different assumptions on the moments and on the dependence structure of

the underlying random variables. The most simple CLT states that the (sampling) distribution of the sample mean of independent and identically distributed random variables with mean μ and finite variance σ^2 may be approximated by a normal distribution with the same mean μ and variance σ^2/n . Although the limiting distribution is continuous, the underlying distribution may even be discrete. An interesting special case occurs when the underlying variable Y has the Bernoulli distribution, with probability of success π . Here, the expected value and the variance of Y are π and $\pi(1 - \pi)$, respectively, and the sample mean p is the proportion of sample elements for which $Y = 1$. It follows that the large-sample distribution of p may be approximated by a $N[\pi, \pi(1 - \pi)/n]$ distribution. This result is known as the De Moivre–Laplace CLT.

Weak convergence results are also available for independent, but not identically distributed (e.g., with different means and variances) underlying random variables, provided some (relatively mild) assumptions hold for their moments. The Liapounov CLT and the Lindeberg–Feller CLT are useful examples. Further extensions cover cases of dependent random underlying variables; in particular, the Hájek–Sidak CLT is extremely useful in regression analysis, where, as the sample size increases, the response variables form a triangular array in which for each row (i.e., for given n), they are independent, but this is not true among rows (i.e., for different values of n). Extensions to cover cases where the underlying random variables have more sophisticated (e.g., martingale-type) dependence structures have been considered by Dvoretzky (1971) among others.

The Slutsky theorem is a handy tool to prove weak convergence of statistics that may be expressed as the sum, product, or ratio of two terms, the first known to converge weakly to some distribution, and the second known to converge in probability to some constant. As an example, consider independent and identically distributed random variables Y_1, \dots, Y_n with mean μ and variance σ^2 . Since the corresponding sample standard deviation S converges in probability to σ and the distribution of \bar{Y} may be approximated by a $N(\mu, \sigma^2/n)$ distribution, we may apply Slutsky's theorem to show that the large-sample distribution of $\sqrt{n}\bar{Y}/S = (\sqrt{n}\bar{Y}/\sigma) \times (\sigma/S)$ may be approximated by a $N(\mu, 1)$ distribution. This allows us to construct approximate confidence intervals for and tests of hypotheses about μ based on the standard normal distribution. A similar approach may be employed to the Bernoulli example by noting that p is a consistent estimator of π .

An important application of Slutsky's theorem relates to statistics that can be decomposed as a sum of terms for which some CLT holds plus a term that converges in probability to 0. Assume, for example, that the variables Y_i have a finite fourth central moment γ and write the sample variance as

$$S^2 = [n/(n-1)] \left\{ n^{-1} \sum_{i=1}^n [(Y_i - \mu)^2 - \sigma^2/n] + \left[\sigma^2 - \sum_{i=1}^n (\bar{Y} - \mu)^2 \right] \right\}.$$

Since the first term within the $\{\}$ brackets is known to converge weakly to a normal distribution by the CLT and the second term converges in probability to 0, we conclude that the distribution of S^2 may be approximated by a $N(\sigma^2, \gamma/n)$ distribution. This is the basis of the projection results suggested by Hoeffding (1948) and extensively explored by Jurečková and Sen (1996) to obtain large-sample properties of U -statistics as well as of more general classes of estimators.

A convenient technique to obtain the asymptotic distributions of many (smooth) functions of asymptotically normal statistics is the Delta-method: if g is a differentiable function of a statistic T_n whose distribution may be approximated (for large samples) by a $N(\mu, \tau^2)$ distribution, then the distribution of the statistic $g(T_n)$ may be approximated by a $N\{g(\mu), [g'(\mu)]^2 \tau^2\}$ distribution, where $g'(\mu)$ denotes the first derivative of g computed at μ . In the context of Example 1, we may be interested in estimating the odds of a failed versus pass response, that is, $\pi_{i1}/(1 - \pi_{i1})$ for the students in the i th group. A straightforward application of the CLT may be used to show that the estimator of π_{i1} , namely, n_{i1}/n_i , follows an approximate $N[\pi_{i1}, \pi_{i1}(1 - \pi_{i1})/n_i]$ distribution. Taking $g(x) = x/(1 - x)$, we may use the Delta-method to show that the distribution of the sample odds $n_{i1}/(n_i - n_{i1})$ may be approximated by a $N\{\pi_{i1}/(1 - \pi_{i1}), \pi_{i1}/[n_i(1 - \pi_{i1})^3]\}$ distribution. This type of result has further applications in variance-stabilizing transformations used in cases (as the above example) where the variance of the original statistic depends on the parameter it is set to estimate.

These concepts and results may also be extended to the multivariate case; in particular, consideration must be given to the Cramér–Wold device, which is a useful process of reduction of multivariate problems to their univariate counterparts. In essence, it states that the asymptotic distribution of the former may be obtained by showing that the asymptotic distribution of every linear combination of the multivariate statistic under investigation may be approximated by a normal distribution.

For some important cases, such as the Pearson χ^2 -statistic, or more general quadratic forms $Q = Q(\boldsymbol{\mu}) = (\mathbf{Y} - \boldsymbol{\mu})' \mathbf{A} (\mathbf{Y} - \boldsymbol{\mu})$ where \mathbf{Y} is a p -dimensional random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix \mathbf{V} and \mathbf{A} is a p -dimensional square matrix of full rank, the (multivariate) Delta-method may not be employed because the derivative of Q computed at $\boldsymbol{\mu}$ is null. If \mathbf{A} converges to an inverse of \mathbf{V} , a useful result known as the Cochran theorem states that the distribution of Q may be approximated by a χ^2 instead of a normal distribution. In fact, the theorem holds even if \mathbf{A} is not of full rank, but converges

to a generalized inverse of \mathbf{V} . This is important for applications in categorical data.

The CLT also does not hold for extreme order statistics such as the sample minimum or maximum; depending on some regularity conditions on the underlying random variables, the distribution of such statistics, conveniently normalized, may be approximated by one of the three types of distributions, namely, the extreme value distributions of the first, second, or third type, which, in this context, are the only possible limiting distributions as shown by Gnedenko (1943).

Given that weak convergence has been established, a question of interest is whether the moments (e.g., mean and variance) of the statistics under investigation converge to the moments of the limiting distribution. Although the answer is negative in general, an important theorem, due to Cramér, indicates conditions under which the result is true (Sen *et al.*, 2010).

For most practical purposes, the specification of an approximate distribution must be complemented with some discussion on the corresponding rates of convergence. In this direction, a bound on the error of approximation of the CLT is provided by the Berry–Esséen theorem; alternatively, such an evaluation may be carried out via Gram–Charlier or Edgeworth expansions (Cramér, 1946). Although this second approach might offer a better insight into the problem than that provided by the former, it requires the knowledge of the moments of the parent distribution and, thus, is less useful in practical applications.

Asymptotic Properties of Estimators and Test Statistics

The convergence of an estimator (in any of the modes described previously) to the parameter being estimated, is certainly desirable and, in general, holds under rather mild regularity conditions often satisfied in practice. A related, but stronger concept, frequently employed in consistency proofs, is that of asymptotic unbiasedness: a sequence $\{T_n\}$ of estimators is asymptotically unbiased if the corresponding expected values converge to the parameter being estimated as n increases indefinitely. In practice, however, such properties are of limited value if not coupled with some form of convergence in distribution. Fortunately, it is possible to obtain such asymptotic distributions for a large class of estimators. In particular, the large-sample distribution of maximum likelihood estimators may be approximated by a normal distribution with variance given by the inverse of the Fisher information, provided a rather mild compactness condition on the second derivative of the log-likelihood is satisfied (Sen *et al.*, 2010). This is generally true for estimators derived

under the exponential family of densities of the underlying random variables. Similar results are also available for other classes of parametric (e.g., robust estimators or more generally, M -estimators) or nonparametric estimators (e.g., those based on U - and V -statistics). Here, however, as shown in Jurečková and Sen (1996), more delicate assumptions are required.

Even if we restrict ourselves to the class of asymptotically normal estimators, we still need some further criteria in order to choose the best candidate; among them lies the concept of asymptotic relative efficiency (ARE), which, in the case of asymptotically normal estimators, corresponds to the ratio of the variances of the approximate distributions. In connection with such ideas, the concept of (absolutely) efficient estimators may be developed with a similar spirit to that of the Cramér–Rao–Fréchet inequality for finite samples; such estimators are termed best asymptotically normal (BAN), and for practical purposes, estimators having this property may be used interchangeably. Note that in general, the asymptotic relative efficiency of two estimators may not coincide with the Fisher asymptotic efficiency, defined as the ratio between the limit of the variance of the estimator (conveniently standardized) and the Cramér–Rao–Fréchet lower bound (Sen *et al.*, 2010).

Large-sample properties of test statistics derived by Wald, score, and likelihood ratio methods may also be obtained along the same lines. Comparing the large-sample efficiency of tests, however, is slightly more complicated, since for fixed alternative hypotheses, for example, $H_A: \mu = \mu_1$ where μ_1 is a constant, the power of such tests converge to 1 as the sample size increases. Therefore, we must restrict ourselves to local Pitman-type alternatives, for example, $H_A: \mu_n = \mu_0 + \Delta/\sqrt{n}$, where μ_0 and Δ are constants, which, in some sense, get closer to the null hypothesis $H_0: \mu = \mu_0$ as the sample size increases; this allows comparisons via the noncentrality parameters of the approximate distributions. Under this approach, we may show that Wald, score, and likelihood ratio methods generate (large-sample) equivalent and locally the most powerful tests.

Extensions to Regression and Categorical Data Models

Linear regression and related models pose special problems, since the underlying random variables are not identically distributed, and in many cases, the exact functional form of their distributions is not completely specified. Least-squares methods are attractive under these conditions, since they may be employed in a rather general setup. In this context, the Hájek–Šidak CLT specifies sufficient conditions on the explanatory variables such that the distributions of the estimators of the regression parameters may be approximated by normal distributions.

As an illustration, consider the simple linear regression model

$$y_i = \alpha + \beta x_i + e_i, \quad i = 1, \dots, n$$

where y_i and x_i represent observations of the response and explanatory variables, respectively, α and β are the parameters of interest, and the e_i correspond to uncorrelated random errors with mean 0 and variance σ^2 . The least-squares estimators of β and α are, respectively, $\hat{\beta} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) / \sum_{i=1}^n (x_i - \bar{x})^2$ and $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$, where \bar{x} and \bar{y} correspond to the sample means of the explanatory and response variables. Irrespective of the form of the underlying distribution of e_i , we may use standard results to show that $\hat{\alpha}$ and $\hat{\beta}$ are unbiased and have variances given by $\sigma^2 [\sum_{i=1}^n x_i^2 / \sum_{i=1}^n (x_i - \bar{x})^2]$ and $\sigma^2 [\sum_{i=1}^n (x_i - \bar{x})^2]^{-1}$, respectively. Furthermore, the covariance between $\hat{\alpha}$ and $\hat{\beta}$ is $-\sigma^2 \bar{x} / \sum_{i=1}^n (x_i - \bar{x})^2$. When the underlying distribution of e_i is normal, we may use standard results to show that $\hat{\alpha}$ and $\hat{\beta}$ follow a bivariate normal distribution. If $\max_{1 \leq i \leq n} (x_i - \bar{x})^2 / \sum_{i=1}^n (x_i - \bar{x})^2$ converges to 0 (Noether's condition) and both \bar{x} and $n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$ converge to finite constants as n increases indefinitely, we may use the Hájek–Šidak CLT to conclude that the same bivariate normal distribution specified above serves as an approximation of the true distribution of $\hat{\alpha}$ and $\hat{\beta}$, whatever the form of the distribution of e_i , provided that n is sufficiently large.

The results may also be generalized to cover alternative estimators obtained by means of generalized and weighted least-squares procedures as well as via robust M -estimation procedures. They may also be extended to generalized linear and nonlinear models.

Applications to categorical data pose additional problems, since many nonlinear models are attractive for their analysis. Although maximum likelihood estimators have optimal large-sample properties, they often require laborious computation because of the natural restrictions involving the parameters of the underlying multinomial distributions. In such cases, they are usually replaced by competitors such as minimum chi-squared, modified minimum chi-squared, or generalized least-squares estimators. Since all these methods generate BAN estimators, their large-sample properties are equivalent (Paulino and Singer, 2006) and the choice among them may rely on computational considerations.

Extensions to Empirical Distribution Functions and Order Statistics

Empirical distribution functions and order statistics have important applications in nonparametric regression models, resampling methods such as the jackknife and bootstrap, sequential testing, as well as in survival and

reliability analyses. In particular, they serve as the basis for the well-known goodness-of-fit Kolmogorov–Smirnov and Cramér–von Mises statistics and for L - and R -estimators such as trimmed or Winsorized means. Given the sample observations Y_1, \dots, Y_n assumed to follow some distribution function F , the empirical distribution function computed at a given real number y is

$$F_n(y) = n^{-1} \sum_{i=1}^n I(Y_i \leq y)$$

where $I(Y_i \leq y)$ is an indicator function assuming the value 1 if $Y_i \leq y$ and 0 otherwise. It is intimately related to the order statistics, $Y_{n:1} \leq Y_{n:2} \leq \dots \leq Y_{n:n}$ where $Y_{n:1}$ is the smallest among Y_1, \dots, Y_n , $Y_{n:2}$ is the second smallest, and so on. For each fixed sample, F_n is a distribution function when considered as a function of y . For every fixed y , when considered as a function of Y_1, \dots, Y_n , $F_n(y)$ is a random variable; in this context, since the $I(Y_i \leq y)$, $i = 1, \dots, n$, are independent and identically distributed zero-one valued random variables, we may apply the CLT to conclude that for each fixed y , the distribution of $F_n(y)$ may be approximated by an $N\{F(y), F(y)[1 - F(y)]/n\}$ distribution, provided that n is sufficiently large. To extend these results to the function F_n computed at all real values y , more sophisticated methods are needed as suggested in Jurečková and Sen (1996) among others.

Conclusion

When the sample size is small, properties of estimators and test statistics required for inferential purposes may be either hard to derive or based on assumptions that are difficult to verify in practical applications. Large-sample methods developed under mild assumptions, usually acceptable for most practical problems, provide approximations for such properties and generally lead to standard inferential procedures, such as confidence intervals based on the normal distribution or χ^2 -distributed test statistics. Although the underlying theory lies on asymptotic (i.e., sample sizes tending to infinity) arguments, the resulting approximate procedures may be satisfactorily employed in cases where the sample size is moderate or even small.

See also: Analysis of Covariance; Categorical Data Analysis; Decision Theory; Educational Data Modeling; Generalized Linear Models; Graphical Models; Multivariate Normal Distribution; Nonparametric Statistical Methods; Order Statistics; Probability Theory; Sequential Testing; Stochastic Processes; Survival Data Analysis; Time Series Analysis.

Bibliography

- Apostol, T. M. (1974). *Mathematical Analysis*. Reading, MA: Addison-Wesley.
- Baker, F. B. and Kim, S. (2004). *Item Response Theory: Parameter Estimation Techniques*, 2nd ed. New York: Marcel Dekker.
- Chow, Y. S. and Teicher, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales*. New York: Springer.
- Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton, NJ: Princeton University Press.
- Dvoretzky, A. (1971). Asymptotic normality for sums of dependent random variables. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 2, pp 513–535. Berkeley: University of California Press.
- Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics* **44**, 423–453.
- Hambleton, R. K., Swaminathan, H., and Roger, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statistics* **19**, 293–325.
- Jurečková, J. and Sen, P. K. (1996). *Robust Statistical Procedures*. New York: Wiley.
- Paulino, C. D. and Singer, J. M. (2006). *Análise de dados categorizados*, (in Portuguese). São Paulo: Blücher.
- Sen, P. K., Singer, J. M., and Pedrosa-de-Lima, A. C. (2010). *From Finite Sample to Asymptotic Methods in Statistics*. Cambridge: Cambridge University Press.

Further Reading

- Ferguson, T. S. (1996). *A Course in Large Sample Theory*. London: Chapman and Hall.
- Lehmann, E. L. (2004). *Elements of Large Sample Theory*. New York: Springer.
- Reiss, R. D. (1989). *Approximate Distributions of Order Statistics. With Applications to Nonparametric Statistics*. New York: Springer.
- Sen, P. K. (1981). *Sequential Nonparametrics*. New York: Wiley.
- Serfling, R. J. (1980). *Approximation theorems of Mathematical Statistics*. New York: Wiley.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. New York: Cambridge University Press.

Latent Class Models

J K Vermunt, Tilburg University, Tilburg, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Introduction

A statistical model can be called a latent class (LC) or mixture model if it assumes that some of its parameters differ across unobserved subgroups, LCs, or mixture components. This rather general idea has several seemingly unrelated applications, the most important of which are clustering, scaling, density estimation, and random-effects modeling. It should be noted that in applied fields, the terms LC model and mixture model are often used interchangeably, which is also what is done here. In the more technical statistical literature on mixture modeling, the term LC analysis is reserved for a specific type of mixture model, that is, a mixture model for a set of categorical items (for the classical LC model).

LC analysis was introduced in 1950 by Lazarsfeld as a tool for building typologies (or clustering) based on dichotomous observed variables (Lazarsfeld, 1950). More than 20 years later, Goodman (1974) made this model applicable in practice by developing an algorithm for obtaining maximum-likelihood estimates of the model parameters, as well as proposed extensions for polytomous manifest variables and did important work on the issue of model identification. Many important extensions of this classical LC model have been proposed since then, such as models containing explanatory variables (Dayton and Macready, 1988), models that relax the local-independence assumption (Hagenaars, 1988), constrained models similar to item response theory (IRT) models (Lindsay *et al.*, 1991; Heinen, 1996), models with multiple latent variables (Magidson and Vermunt, 2001), models for longitudinal data (Van de Pol and Langeheine, 1990), and models for multilevel data (Vermunt, 2003).

Whereas this classical LC model and its extensions are conceived primarily as a clustering and scaling tool for categorical data analysis, LC and finite-mixture models can be useful in several other areas as well. One of these is as a probabilistic cluster-analysis tool for continuous observed variables, an approach that offers many advantages over traditional cluster techniques such as K-means clustering (Wolfe, 1970; McLachlan and Peel, 2000; Vermunt and Magidson, 2002). Another application area is dealing with unobserved heterogeneity, as happens in mixture regression analysis of multilevel or repeated-measurement data (Wedel and DeSarbo, 1994; Vermunt and Van Dijk, 2001).

The remainder of this article is organized as follows. After introducing the simplest type of LC models, various

restricted LC models as well as models with explanatory variables are discussed. Next, an overview of other types of LC and mixture models, which includes various recently proposed extensions, is presented. In the end, attention is paid to parameter estimation, model selection, and software.

Simple LC and Mixture Models

LC analysis is typically used as a tool for analyzing multivariate response data; that is, data consisting of several dependent variables, response variables, or items. We denote the response of subject i on dependent variable j by y_{ij} , and the number of dependent variables by J . The full-response vector of a subject is denoted by \mathbf{y}_i . To make things more concrete, **Table 1** presents a small illustrative data set consisting of three dichotomous responses, y_{i1} , y_{i2} , and y_{i3} (0 = incorrect; 1 = correct). This is a subset of items from a mathematics test administered to 2156 children. The frequency column contains the observed frequency count for each of the eight possible answer patterns.

In addition to the J observed dependent variables, an LC model contains a discrete latent variable. We denote a subject's unobserved score on this latent variable by v_i , the number of LCs by C , and a particular class by c , where $c = 1, 2, \dots, C$. The aim of an LC analysis of the data set in **Table 1** could be to classify pupils into two groups, masters and nonmasters, which differ with respect to the probability of answering the test items correctly. The results obtained with a two-class model will be used to illustrate the various components of an LC model.

LC analysis defines a model for $f(\mathbf{y}_i)$, the probability density of the multivariate response vector \mathbf{y}_i . In the above example, this is the probability of answering the items according to one of the eight possible response patterns, for example, of answering the first two items correctly and the last one incorrectly, which as can be seen in **Table 1** equals 0.161 for the estimated two-class model. The assumption underlying any type of LC or mixture model is that the density $f(\mathbf{y}_i)$ is a weighted average (or mixture) of the C class-specific densities $f(\mathbf{y}_i|v_i = c)$. This is expressed mathematically as follows:

$$f(\mathbf{y}_i) = \sum_{c=1}^C P(v_i = c) f(\mathbf{y}_i|v_i = c). \quad [1]$$

Here, $P(v_i = c)$ denotes the probability that a subject belongs to LC c . For our small empirical example, the estimates of these (prior) class membership probabilities

Table 1 Small data set with three dichotomous responses

y_{i1}	y_{i2}	y_{i3}	Frequency	$f(y_i v_i = 1)$	$f(y_i v_i = 2)$	$f(y_i)$	$f(v_i = 1 y_i)$	$f(v_i = 2 y_i)$	Modal
0	0	0	239	0.004	0.272	0.111	0.020	0.980	2
0	0	1	101	0.010	0.102	0.047	0.128	0.872	2
0	1	0	283	0.038	0.271	0.131	0.175	0.825	2
0	1	1	222	0.104	0.102	0.103	0.605	0.395	1
1	0	0	105	0.020	0.092	0.049	0.248	0.753	2
1	0	1	100	0.054	0.035	0.046	0.703	0.297	1
1	1	0	348	0.208	0.091	0.161	0.774	0.226	1
1	1	1	758	0.562	0.034	0.352	0.961	0.039	1

are 0.601 and 0.399 for classes 1 and 2, respectively (see **Table 2**). The assumed mechanism by eqn [1] is that each individual belongs to one of C exhaustive and mutually exclusive classes with probability $P(v_i = c)$ and that given membership of LC c one provides responses according to the probability density associated to this class. **Table 1** shows the estimated values of $f(y_i|v_i = 1)$ and $f(y_i|v_i = 2)$ for our data sets. As can be seen, LC 1 has higher probabilities for the response patterns with 2 or 3 correctly answered items, whereas class 2 has higher probabilities for the response patterns with 0 or 1 item correct.

The classical LC model combines the assumption of eqn [1] shared by all mixture models with the assumption of local independence. Local independence means that the \mathcal{J} responses are mutually independent given a subject's class membership. It can be expressed as follows:

$$f(y_i|v_i = c) = \prod_{j=1}^{\mathcal{J}} f(y_{ij}|v_i = c). \quad [2]$$

Independence implies that the joint density $f(y_i|v_i = c)$ is obtained as a product of the \mathcal{J} item-specific densities $f(y_{ij}|v_i = c)$. In our example $f(y_{ij} = 1|v_i = c)$ is the class-specific probability of giving a correct answer to item j . As reported in **Table 2**, for a subject belonging to the first LC, these equal 0.844, 0.912, and 0.730 for items 1, 2, and 3, respectively. The local independence assumption implies, for example, that the probability of answering the first two items correctly and the last one incorrectly for someone in LC one equals $0.844 \times 0.912 \times (1 - 0.730) = 0.208$. Note that the local independence assumption is also used in other types of latent variables models, such as in factor analysis and IRT modeling, and is thus not specific for LC analysis.

Combining the two basic eqns [1] and [2] yields the following model for $f(y_i)$:

$$f(y_i) = \sum_{c=1}^C P(v_i = c) \prod_{j=1}^{\mathcal{J}} f(y_{ij}|v_i = c). \quad [3]$$

To complete the model specification, we need to define the form of the conditional densities $f(y_{ij}|v_i = c)$. In the classical LC model for categorical items, these are multinomial probability densities; that is,

Table 2 Parameters (class proportions and probability of a correct answer) obtained with two-class model for data in **Table 1**

	$c = 1$	$c = 2$
$P(v_i = c)$	0.601	0.399
$\pi_{11c} = P(y_{i1} = 1 v_i = c)$	0.844	0.252
$\pi_{21c} = P(y_{i2} = 1 v_i = c)$	0.912	0.499
$\pi_{31c} = P(y_{i3} = 1 v_i = c)$	0.730	0.273

$$f(y_{ij}|v_i = c) = \prod_{r=0}^{R_j-1} \pi_{jrc}^{y_{ijr}^*},$$

where R_j is the number of categories of item j , $0 \leq y_{ij} \leq R_j - 1$, and $y_{ijr}^* = 1$ if $y_{ij} = r$ and 0 otherwise. Note this is a slightly complicated, but mathematically elegant, way to express that someone in LC c has a probability equal to $\pi_{jrc} = P(y_{ij} = r|v_i = c)$ of giving response r to item j . In the special case of a dichotomous response, the multinomial distribution reduces to the Bernoulli distribution with success probability $\pi_{jc} = \pi_{j1c} = P(y_{ij} = 1|v_i = c)$. **Table 2** presents these probabilities for our small example.

It is important to note that LC models cannot only be used with categorical responses, but also with continuous responses and counts. The density $f(y_{ij}|v_i = c)$ could be a binomial, Poisson, or negative binomial distribution for counts, and a normal or gamma distribution for continuous responses. The mixture model for continuous response variables is sometimes referred to as the latent profile model. The parameters of this model are the class proportions and class-specific item means and variances (μ_{jc} and σ_{jc}^2).

By comparing the \mathcal{J} sets of item parameters across classes, one can name the classes. The parameter estimates presented in **Table 2** show that the first class can be named the masters because pupils belonging to that class have much higher probabilities of answering the test items correctly than pupils belonging to the second non-masters class.

Similar to cluster analysis, one of the purposes of LC analysis might be to assign individuals to LCs.

The probability of belonging to LC c given responses \mathbf{y}_i – often referred to as posterior membership probability – can be obtained by the Bayes rule:

$$P(v_i = c | \mathbf{y}_i) = \frac{P(v_i = c) f(\mathbf{y}_i | v_i = c)}{f(\mathbf{y}_i)}. \quad [4]$$

Table 1 reports $P(v_i = c | \mathbf{y}_i)$ for each answer pattern. For example, $P(v_i = 1 | \mathbf{y}_i)$ equals 0.774 for the (1,1,0) pattern, which is obtained as $0.601 \times 0.208 / 0.161$. The most common classification rule is modal assignment, which amounts to assigning each individual to the LC with the highest $f(v_i = c | \mathbf{y}_i)$. The last column of **Table 1** reporting the modal assignments shows that pupils with at least two correct answers are assigned to class 1 and the others to class 2.

In the introduction, we stated that mixture models are statistical models in which parameters are assumed to differ across LCs. But what is the statistical model used in the simple LC models discussed so far? It is the independence model: we assume responses to be independent, with different parameter values for each class. Depending on the scale type of the response variables, these parameters are Bernoulli probabilities, multinomial probabilities, normal means and variances, Poisson rates, etc.

Generalized Linear Models for Item Probabilities/Mean

Haberman (1979) showed that the LC model for categorical response variables can also be specified as a log-linear model for an expanded table, including the latent variable v_i as an additional dimension. Using such a log-linear specification is equivalent to parameterizing the response probability for item j as follows:

$$\log\left(\frac{\pi_{jrc}}{\pi_{j0c}}\right) = \log\left(\frac{P(\mathcal{Y}_{ij} = r | v_i = c)}{P(\mathcal{Y}_{ij} = 0 | v_i = c)}\right) = \alpha_{jr} + \beta_{jcr}, \quad [5]$$

for $1 \leq r \leq R_j - 1$; that is, as a multinomial logistic regression models with intercepts α_{jr} and slopes β_{jcr} (note that we use the first item category, $r = 0$, as baseline). One identification constraint needs to be imposed, for example, $\beta_{j1r} = 0$ (the parameters for class 1 are fixed to 0) or $\alpha_{jr} = 0$ (intercepts are fixed to 0).

For dichotomous responses and binomial counts, the regression model could be a binary logit or probit model, for Poisson counts a log-linear model, and for continuous responses a standard linear model. These are generalized linear models (GLMs) of the form

$$g[E(\mathcal{Y}_{ij} | v_i = c)] = \alpha_j + \beta_{jc}, \quad [6]$$

where $g[\cdot]$ is the link function transforming the expected value of \mathcal{Y}_{ij} to the linear term. For ordinal polytomous variables, one may use an ordinal regression model,

such as an adjacent-category or cumulative logit model. These are models that restrict the item response probabilities π_{jrc} .

Some Restricted Models for Categorical Items

Many interesting types of restricted LC models for categorical items have been proposed, which involve imposing (linear) constraints on either the conditional probabilities π_{jrc} or the logit coefficients of eqn [5]. One of these is the probabilistic Guttman scaling model for dichotomous responses, which is an LC model with $C = \mathcal{J} + 1$ classes, one for each possible total score. The idea is that apart from measurement error, class c should provide a negative answer to the $c - 1$ easiest items and a positive answer to the remaining $\mathcal{J} - (c - 1)$ items. The various types of probabilistic Guttman models differ in the constraints they impose on the measurement error. The simplest and most restricted model is the Proctor (1970) model. **Table 3** presents the parameter estimates obtained when fitting the Proctor model to the data set in **Table 1**. As can be seen, the probability of a correct response is either 0.833 or $0.167 = 1 - 0.833$. The measurement error – or the probability of giving a response which is not in agreement with the class – is estimated to be equal to 0.167. Whereas the Proctor model assumes that the measurement error is constant across items and classes, less-restricted models can be defined which allow the error probabilities to differ across items, classes, or both (see, e.g., Dayton, 1999). Note that these equality constraints on the error probabilities can also be defined using linear constraints on the logit parameters: $\alpha_{j1} = 0$ and $\beta_{j1c} = -\beta^*$ for $c \leq j$ and $\beta_{jc} = \beta^*$ otherwise.

Croon (1990) proposed a restricted LC model that similar to nonparametric IRT (Sijtsma and Molenaar, 2000) assumes monotonic item response functions; that is, $\pi_{j1c} \leq \pi_{j1,c+1}$, or, equivalently, $\beta_{j1c} \leq \beta_{j1,c+1}$. A more restricted version, in which not only classes but also items are ordered, is obtained by imposing the additional set of restriction $\pi_{j+11c} \leq \pi_{j1c}$; that is, by assuming double monotony. Vermunt (2001) has discussed various generalizations of these models.

Table 3 Parameters (class proportions and probability of a correct answer) obtained with proctor model for data in **Table 1**

	$c = 1$	$c = 2$	$c = 3$	$c = 4$
$P(v_i = c)$	0.160	0.155	0.126	0.559
$\pi_{21c} = P(\mathcal{Y}_{12} = 1 v_i = c)$	0.167	0.833	0.833	0.833
$\pi_{11c} = P(\mathcal{Y}_{11} = 1 v_i = c)$	0.167	0.167	0.833	0.833
$\pi_{31c} = P(\mathcal{Y}_{13} = 1 v_i = c)$	0.167	0.167	0.167	0.833

Various authors described the connection between restricted LC analysis and parametric IRT modeling (see, e.g., Heinen, 1996; Lindsay *et al.*, 1991); that is, IRT models with a discrete specification of the distribution of the underlying trait or ability can be defined as LC models with restrictions on the logistic parameters. The key restriction is $\beta_{jrc} = \beta_{jr}^* \cdot \theta_c$ for nominal items and $\beta_{jrc} = \beta_j^* \cdot r \cdot \theta_c$ for ordinal items, where θ_c are LC locations representing the C possible values of the discretized latent trait. These locations may be fixed *a priori*, for example, at $-2, -1, 0, 1$, and 2 in the case of $C = 5$, but may also be treated as free parameters to be estimated. Depending on whether the items are dichotomous, ordinal, or nominal, this yields a 2-parameter logistic, generalized partial credit, or nominal response model. Further restrictions involve equating β_j^* across items, yielding Rasch and partial credit models, and imposing across-category and across-item restrictions on α_{jr} parameters as in rating scale models for ordinal items.

Models with Explanatory Variables

The most important extension of the LC models discussed so far is the possibility to include explanatory variables affecting the responses (Wedel and DeSarbo, 1994) or the class memberships (Dayton and Macready, 1988). Denoting the vector with explanatory variables for subject i by \mathbf{x}_i , the LC model of interest can be formulated as follows:

$$f(\mathbf{y}_i | \mathbf{x}_i) = \sum_{c=1}^C P(v_i = c | \mathbf{x}_i) \prod_{j=1}^{j_i} f(y_{ij} | v_i = c, \mathbf{x}_{ij}). \quad [7]$$

The main difference compared to the model defined in eqn [3] is that now we have a model for $f(\mathbf{y}_i | \mathbf{x}_i)$ – the conditional density of \mathbf{y}_i given \mathbf{x}_i .

Similar to the regression models for the response variables introduced in eqns [5] and [6], one can define a mixture regression model with explanatory variables; that is,

$$g[E(y_{ij} | v_i = c, \mathbf{x}_{ij})] = \alpha_c + \sum_{p=1}^P \beta_{pc} x_{ijp}. \quad [8]$$

As before, y_{ij} may refer to the response on item j by pupil i , in which case the explanatory variables will consist of a design matrix defining the item parameters. However, the model in eqn [8] can also be used for many other purposes. In fact, it is a model for analyzing two-level data sets, where regression parameters are allowed to differ across LCs (of higher-level units). For example, y_{ij} could be the test score of pupil j belonging to school i , and \mathbf{x}_{ij} a set of pupil characteristics (e.g., intelligence quotient (IQ)). A mixture regression model would identify LCs of schools with different intercepts and different effects of child characteristics on the test scores.

Another possible application is in the analysis of longitudinal data, where j is a time point for subject i , and where vector \mathbf{x}_{ij} contains time variables. This yields a LC growth model in which subjects are grouped based on their developmental trajectories (Vermunt, 2007). A last possible application that can be mentioned is in experiments in which subjects are observed in multiple conditions, such as in conjoint studies. The mixture regression model can be used to group subjects based on their reactions on the experimental conditions. In fact, in each of these application types, the LC model is used as a random-coefficient model without parametric assumptions about the distribution of the random effects (Aitkin, 1999; Vermunt and Van Dijk, 2001).

As shown in eqn [7], an individual's class membership may also be predicted using covariates. This is achieved by defining a multinomial logistic regression model for $P(v_i = c | \mathbf{x}_i)$:

$$\log \frac{P(v_i = c | \mathbf{x}_i)}{P(v_i = 1 | \mathbf{x}_i)} = \gamma_{0c} + \sum_{q=1}^Q \gamma_{qc} x_{iq}.$$

Strongly related are multiple-group LC models. These can be defined using the grouping variable as a nominal explanatory in the model.

Extensions

The most common model-fitting strategy in LC analysis is to increase the number of classes until the local-independence assumption holds. This may, however, yield solutions which are difficult to interpret. One alternative approach is to relax the local-independence assumption by allowing for associations between particular item pairs. Hagenaars (1988) showed how to define LC models with local dependencies for categorical responses. With continuous responses, this is easily achieved using multivariate instead of univariate normal distributions for locally dependent items (see, e.g., McLachlan and Peel, 2000; Vermunt and Magidson, 2002).

Another alternative strategy involves increasing the number of discrete latent variables instead of the number of LCs, which is especially useful if the items measure several dimensions. This so-called discrete-factor modeling approach (Magidson and Vermunt, 2001) is a special case of the path-modeling approach for discrete latent variables developed by Hagenaars (1990) and Vermunt (1997). Many other interesting models can be defined within this framework, such as latent Markov models for the analysis of longitudinal data (Van de Pol and Langeheine, 1990) and LC models for cognitive diagnosis (De la Torre and Douglas, 2004).

Recently, models have been developed that contain both discrete and continuous latent variables. Examples of these are mixture factor models (Yung, 1997; McLachlan

and Peel, 2000), mixture structural equation models (Dolan and Van der Maas, 1997), and mixture IRT models (Rost, 1990).

Probably the most recent extension is the multilevel LC model (Vermunt, 2003). One of its variants is a model with discrete latent variables at multiple levels of a hierarchical structure: for example, children belong to LCs with different performances on a set of test items, and schools belong to LCs with different distributions of children across the child-level performance classes. Multilevel LC models can be used for the analysis of two-level multivariate and three-level univariate response data.

Maximum Likelihood Estimation

The parameters of LC models are typically estimated by means of maximum likelihood (ML). The log-likelihood function that is maximized is based on the probability densities defined in eqns [1–3]; that is,

$$\ln L = \sum_{i=1}^N \ln f(\mathbf{y}_i).$$

With categorical responses one will typically group the data and construct a frequency table as we did in **Table 1**. The log-likelihood function for grouped data equals

$$\ln L = \sum_{k=1}^K n_k \ln f(\mathbf{y}_k),$$

where k is a data pattern, K the number of different data patterns, and n_k the cell count corresponding to data pattern k . Notice that only nonzero observed cell entries contribute to the log-likelihood function, a feature that is exploited by several more efficient LC software packages that have been developed within the past few years.

One of the problems in the estimation of LC models for discrete y_{ij} is that model parameters may be nonidentified, even if the number of degrees of freedom – the number of independent cells in the J -way cross-tabulation minus the number of free parameters – is larger or equal to zero. Nonidentification means that different sets of parameter values yield the same maximum of the log-likelihood function or, worded differently, that there is no unique set of parameter estimates. The formal identification check is via the Jacobian matrix (matrix of first derivatives of $f(\mathbf{y}_i)$), which should be column full rank. Another option is to estimate the model of interest with different sets of starting values. Except for local solutions (see below), an identified model gives the same final estimates for each set of the starting values.

Although there are no general rules with respect to the identification of LC models, it is possible to provide certain minimal requirements and point to possible pitfalls. For an unrestricted LC analysis, one needs at least

three responses (y_{ij} 's) per individual, but if these are dichotomous, no more than two LCs can be identified. One has to watch out with four dichotomous response variables, in which case the unrestricted three-class model is not identified, even though it has a positive number of degrees of freedom. With five dichotomous items, however, even a five-class model is identified. Usually, it is possible to achieve identification by constraining certain model parameters.

A second problem associated with the estimation of LC models is the presence of local maxima. The log-likelihood function of an LC model is not always concave, which means that hill-climbing algorithms may converge to a different maximum depending on the starting values. Usually, we are looking for the global maximum. The best way to proceed is, therefore, to estimate the model with different sets of random starting values. Typically, several sets converge to the same highest log-likelihood value, which can then be assumed to be the ML solution. Some software packages have automated the use of multiple sets of random starting values to reduce the probability of getting a local solution.

Another problem in LC modeling is the occurrence of boundary solutions, which are probabilities equal to 0 (or 1) or logit parameters equal to minus (or plus) infinity. These may cause numerical problems in the estimation algorithms, occurrence of local solutions, and complications in the computation of standard errors and number of degrees of freedom of the goodness-of-fit tests. Boundary solutions can be prevented by imposing constraints or by taking into account other kinds of prior information on the model parameters.

The most popular methods for solving the ML estimation problem are the expectation–maximization (EM) and Newton–Raphson (NR) algorithms. EM is a very stable iterative method for ML estimation with incomplete data. NR is a faster procedure that, however, needs good starting values to converge. The latter method makes use of the matrix of second-order derivatives of the log-likelihood function, which is also needed for obtaining standard errors of the model parameters.

Model Selection Issues

The goodness-of-fit of LC models for categorical responses can be tested using Pearson and likelihood-ratio chi-squared tests. The latter is defined as

$$L^2 = 2 \sum_{k=1}^K n_k \ln \frac{n_k}{N \cdot f(\mathbf{y}_k)}.$$

As in log-linear analysis, the number of degrees of freedom (df) equals the number of cells in the frequency table minus 1, minus the number of independent parameters. In an unrestricted LC model,

$$df = \prod_{j=1}^J R_j - C \cdot \left[1 + \sum_{j=1}^J (R_j - 1) \right].$$

Although it is no problem to estimate LC models with 10, 20, or 50 indicators, in such cases, the frequency table may become very sparse and, as a result, asymptotic p -values can no longer be trusted. An elegant but somewhat time-consuming solution to this problem is to estimate the p -values by parametric bootstrapping. Another option is to assess model fit in lower order marginal tables (e.g., in the two-way marginal tables).

Even though models with C and $C + 1$ are nested, one cannot test them against each other using a standard likelihood-ratio test because it does not have an asymptotic chi-squared distribution. A way out to this problem is to approximate its sampling distribution using bootstrapping. But since this method is computationally demanding, usually alternative methods are required for comparing models with different numbers of classes. One popular method is the use of information criteria such as the Bayesian information criterion (BIC) and Akaike information criterion (AIC). Another more descriptive method is a measure for the proportion of total association accounted for by a C -class model, $[L^2(1) - L^2(C)]/L^2(1)$, where the L^2 value of the one-class (independence) model, $L^2(1)$, is used as a measure of total association in the J -way frequency table.

Usually, we are not only interested in goodness-of-fit but also in the performance of the modal classification rule (see eqn [4]). The estimated proportion of classification errors under modal classification equals

$$E = \sum_{i=1}^N \frac{1}{N} \{1 - \max[P(v_i = c | y_i)]\}.$$

This number can be compared to the proportion of classification errors based on the unconditional probabilities $P(v_i = c)$, yielding a reduction of errors measure

$$\lambda = 1 - \frac{E}{1 - \max[P(v_i = c)]}.$$

The closer this nominal R^2 -type measure is to 1, the better the classification performance of a model. Other types of classification error-reduction measures have been proposed based on entropy or qualitative variance.

Software

One of the first LC analysis programs, maximum likelihood latent structure analysis (MLLSA), made available by Clifford Clogg in 1977, was limited to a relatively small number of nominal variables. Today's programs can handle many more variables, as well as other scale types. For example, the LEM program (Vermunt, 1997) provides

a command language that can be used to specify a large variety of models for categorical data, including LC models. Mplus is a command-language-based structural-equation modeling package that implements many types of LC and mixture models. In addition, routines for the estimation of specific types of LC models are available as SAS, R, and Stata macros (see, e.g., Lanza *et al.*, 2007; Skrondal and Rabe-Hesketh, 2004).

Latent GOLD is a program that was especially developed for LC analysis, and which contains both an SPSS-like point and click-user interface and a syntax language. It implements all important types of LC models, such as models for response variables of different scale types, restricted LC models, models with predictors, models with local dependencies, models with multiple discrete latent variables, LC path models, LC Markov models, mixture factor analysis and IRT, and multilevel LC models, as well as features for dealing with partially missing data, for performing bootstrapping, and for dealing with complex samples.

Bibliography

- Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* **55**, 218–234.
- Croon, M. A. (1990). Latent class analysis with ordered latent classes. *British Journal of Mathematical and Statistical Psychology* **43**, 171–192.
- Dayton, C. M. (1999). *Latent Class Scaling Analysis*. Thousand Oaks, CA: Sage.
- Dayton, C. M. and Macready, G. B. (1988). Concomitant-variable latent class models. *Journal of the American Statistical Association* **83**, 173–178.
- De la Torre, J. and Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika* **69**, 333–353.
- Dolan, C. V. and Van der Maas, H. L. J. (1997). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika* **63**, 227–253.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61**, 215–231.
- Haberman, S. J. (1979). *Analysis of Qualitative Data. Vol. 2: New Developments*. New York: Academic Press.
- Hagenaars, J. A. (1988). Latent structure models with direct effects between indicators: Local dependence models. *Sociological Methods and Research* **16**, 379–405.
- Hagenaars, J. A. (1990). *Categorical Longitudinal Data: Loglinear Analysis of Panel, Trend and Cohort Data*. Newbury Park, CA: Sage.
- Heinen, T. (1996). *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. Thousand Oaks, CA: Sage.
- Lanza, S. T., Collins, L. M., Lemmon, D. R., and Schafer, J. L. (2007). PROC LCA: A SAS procedure for latent class analysis. *Structural Equation Modeling* **14**, 671–694.
- Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis and the interpretation and mathematical foundation of latent structure analysis. In Stouffer, S. A., Guttman, L., Suchman, E. A., *et al.* (eds.) *Measurement and Prediction*, pp 362–472. Princeton, NJ: Princeton University Press.
- Lindsay, B., Clogg, C. C., and Grego, J. (1991). Semiparametric estimation in the Rasch model and related models, including a simple latent class model for item analysis. *Journal of the American Statistical Association* **86**, 96–107.

- Magidson, J. and Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays. *Sociological Methodology* **31**, 223–264.
- McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. New York: Wiley.
- Proctor, C. H. (1970). A probabilistic formulation and statistical analysis of Guttman scaling. *Psychometrika* **35**, 73–78.
- Rost, J. (1990). Rasch models in latent classes. An integration of two approaches to item analysis. *Applied Psychological Measurement* **14**, 271–282.
- Sijtsma, K. and Molenaar, I. W. (2002). *Introduction to Nonparametric Item Response theory*. Thousand Oaks, CA: Sage.
- Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. London: Chapman and Hall/CRC.
- Van de Pol, F. and Langeheine, R. (1990). Mixed Markov latent class models. *Sociological Methodology* **20**, 213–247.
- Vermunt, J. K. (1997). *Log-Linear Models for Event Histories*. Thousand Oaks, CA: Sage.
- Vermunt, J. K. (2001). The use restricted latent class models for defining and testing nonparametric and parametric IRT models. *Applied Psychological Measurement* **25**, 283–294.
- Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology* **33**, 213–239.
- Vermunt, J. K. (2007). Growth models for categorical response variables: Standard, latent-class, and hybrid approaches. In van Montfort, K., Oud, H., and Satorra, A. (eds.) *Longitudinal Models in the Behavioral and Related Sciences*, pp 139–158. Mahwah, NJ: Erlbaum.
- Vermunt, J. K. and Magidson, J. (2002). Latent class cluster analysis. In Hagenars, J. and McCutcheon, A. (eds.) *Applied Latent Class Analysis*, pp 89–106. Cambridge University Press.
- Vermunt, J. K. and Van Dijk, L. (2001). A nonparametric random-coefficients approach: The latent class regression model. *Multilevel Modelling Newsletter* **13**, 6–13.
- Wedel, M. and DeSarbo, W. (1994). A review of recent developments in latent class regression models. In Bagozzi, R. P. (ed.) *Advanced Methods of Marketing Research*, pp 352–388. Cambridge, MA: Blackwell.
- Wolfe, J. H. (1970). Pattern clustering by multivariate cluster analysis. *Multivariate Behavioral Research* **5**, 329–350.
- Yung, Y. F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika* **62**, 297–330.

Further Reading

- Hagenars, J. A. and McCutcheon, A. L. (2002). *Applied Latent Class Analysis*. Cambridge, UK: Cambridge University Press.
- Magidson, J. and Vermunt, J. K. (2004). Latent class models. In Kaplan, D. (ed.) *The Sage Handbook of Quantitative Methodology for the Social Sciences*, ch. 10, pp 175–198. Thousand Oaks, CA: Sage.
- Vermunt, J. K. (forthcoming). Mixture models for multilevel data sets. In Hox, J. and Roberts, J. K. (eds.) *The Handbook of Advanced Multilevel Analysis*.
- Vermunt, J. K., Tran, B., and Magidson, J. (2008). Latent class models in longitudinal research. In Menard, S. (ed.) *Handbook of Longitudinal Research: Design, Measurement, and Analysis*, pp 373–385. Burlington, MA: Elsevier.

Markov Chain Monte Carlo

A M Johansen, University of Warwick, Coventry, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bayesian – An approach to statistical inference in which parameters are treated as random variables, allowing Bayes rule $p(y|x) = p(x|y)p(y)/p(x)$ to be used to perform inference. This approach is named after the Reverend Thomas Bayes.

Bias – An estimator is biased if its expectation differs from the true value; the bias of the estimator is the difference between its expectation and the truth.

Consistency – A sequence of estimators is consistent if it converges, as the number of samples approaches infinity, to the true value. Such a sequence is asymptotically unbiased.

Expectation – The expectation (or expected value) of a random variable is, in a sense, the mean value that an infinite ensemble of independent replicates would have. In some sense, it quantifies the value which one would expect a realization to have, but there is no requirement that the expectation is a value which the random variable can take: for example, a Bernoulli random variable with parameter p can take values 0 and 1, but has expectation p .

iid (independent and identically distributed) – A collection of random variables is said to be iid if those variables are independent and share a common distribution. That is, each may be considered to be a single realization of the same probability distribution.

Independent – Random variables are independent if knowledge of one provides no information about the others. Formally, one can require that the conditional distribution of any one of the random variables given the other is the marginal distribution of that random variable, that is, $p_{Y|X}(y|x) = p_Y(y)$ if X and Y are independent.

Markov chain – A temporally ordered collection of random variables with the property that conditional upon the present, the future is independent of the past.

mathematical problems, which make fundamental use of random samples. Markov chain Monte Carlo (MCMC) uses a particular type of (usually discrete time) Markov process to obtain collections of dependent variables. Integration and optimization can both be addressed within this framework; this article concentrates on the former although it also touches on the latter.

The Monte Carlo method as it is understood today was born in Los Alamos in the 1940s, when physicists working on particle transport problems began solving them using something which they termed the Monte Carlo method. (Although earlier examples of Monte Carlo techniques exist, the cost of carrying out the experiments rendered them largely impractical and of only specialist interest.) The precise origin of the name varies from one account to another, but there is general agreement that it stems from its relationship with the games of chance played in the casinos of Monte Carlo. The first revolutionary step introduced at this time was the use of (pseudo)random number generators in place of physical experiments to perform the calculation. The second was the realization that it is not necessary to use collections of independent random variables of known distribution: this is the concept which leads to MCMC. The physics community has continued to contribute to the development of Monte Carlo methodology, especially MCMC, to the present day.

During the 1980s, the increasing use of Bayesian methods and the associated need to evaluate complex, high-dimensional integrals led to a renewed interest in and development of MCMC methods by the statistics community. This was the point at which the use of Monte Carlo methods to approximate general integral expressions, particularly expectations with respect to complicated probability distributions, became widespread. The increased interest also drove – and continues to drive – the development of methodology and theory amongst researchers outside of the traditional application domains of Monte Carlo integration.

In Bayesian statistics, it is extremely common for inferential techniques to require the calculation of expectations with respect to probability distributions which are known only up to a normalising constant and, furthermore, from which it is not possible to obtain samples. Although simple Monte Carlo techniques such as rejection and importance sampling are formally very general, they both require a good proposal distribution in order to work efficiently. These can be difficult to design, particularly in high-dimensional

Background and History

Monte Carlo methods may be thought of as computational techniques for the (usually approximate) solution of

problems. MCMC is a technique which has been developed to provide a method for calculating expectations under high-dimensional (and generally difficult to sample from) distributions.

Markov Chains and MCMC

Markov Chains

A discrete-time Markov chain is, roughly speaking, some collection of random variables with a temporal ordering which have the property that conditional upon the present, the future does not depend upon the past. This concept, which can be viewed as a form of something known as the Markov property, can be formalized by saying that a collection of random variables X_1, X_2, \dots forms a Markov chain if, and only if, the joint pdf of the first n elements of the sequence may be decomposed in the following manner for any value of n :

$$p(x_1, \dots, x_n) = p(x_1)p(x_2|x_1) \dots p(x_n|x_{n-1})$$

Although a class of continuous-time stochastic processes with a similar lack-of-memory property can be defined, these are rarely used in an MCMC context.

The basic idea behind MCMC is that, if it is possible to construct a Markov chain such that a sequence of draws from that chain has similar statistical properties (in some sense) to a collection of draws from a distribution of interest, then it is also possible to estimate expectations with respect to that distribution by using the standard Monte Carlo estimator but using the dependent collection of random variables obtained by simulating a Markov chain rather than an independent collection.

A few concepts are required to understand how a suitable chain can be constructed. The conditional probability densities $p(x_n|x_{n-1})$ are often termed transition kernels, as they can be thought of as the probability density associated with a movement from x_{n-1} to x_n . If $p(x_n|x_{n-1})$ does not depend directly upon the value of n , then the associated Markov chain is termed time homogeneous (as its transitions have the same distribution at all times). A time homogeneous Markov chain with transition kernel k , is said to have a probability density f as an invariant or stationary distribution if

$$\int f(x)k(y|x)dx = f(y)$$

If a Markov chain satisfies a condition known as detailed balance with respect to a distribution, then it is reversible (in the sense that the statistics of the time-reversed process match those of the original process) and hence invariant with respect to that distribution. The detailed balance condition states, simply, that the probability of starting at x and moving to y is equal to the probability of starting at y

and moving to x . Formally, given a distribution f and a kernel k , one requires that $f(x)k(y|x) = f(y)k(x|y)$ and simple integration of both sides with respect to x proves invariance with respect to f under this condition.

The principle of most MCMC algorithms is that, if a Markov chain has an invariant distribution, f , and (in some suitable sense) forgets where it has been, then using its sample path to approximate integrals with respect to f is a reasonable thing to do. This can be formalized under technical conditions to provide an analog of the law of large numbers (often termed the ergodic theorem) and the central limit theorem. The first of these results tells us that we can expect the sample average to converge to the appropriate expectation with probability one as the number of samples becomes large enough; the second tells us that the estimator we obtain is asymptotically normal with a particular variance (which depends upon the covariance of the samples obtained, demonstrating that it is important that the Markov chain forgets where it has been reasonably fast). These conditions are not always easy to verify in practice, but they are important: it is easy to construct examples which violate these conditions and have entirely incorrect behavior.

In order to use this strategy to estimate expectations of interest, it is necessary to construct Markov chains with the correct invariant distribution. There are two common approaches to this problem.

Gibbs Sampling

The first approach is termed Gibbs sampling, and relies on the ability to sample from the conditional distributions of the target distribution. It can be motivated heuristically by the idea that, given a joint distribution one might hope that sampling iteratively from its conditional distributions would ultimately provide a sample from the joint distribution. This can be analyzed by considering the Markov chain that results from the procedure and it can be shown that under suitable conditions this is the case.

More formally, if the target distribution is p over some space which can be factorized into d elements (such as $\mathbb{R}^d = \mathbb{R} \times \dots \times \mathbb{R}$), it is possible to write realizations of the associated random variable as $x^{1:d} = (x^1, \dots, x^d)$ and define $x^{-j} = (x^1, \dots, x^{j-1}, x^{j+1}, \dots, x^d)$ as the random variable excluding its j th component. Gibbs sampling then proposes the use of a collection of d kernels, with $k_j(x_j|x_{-j}) = \delta(x_{j-1}^{-j} - x_{j-1}^{-j})p(x_j^j|x_{-j}^{-j})$ (where δ denotes the Dirac delta measure which places probability one on its argument taking the value zero). That is, the j th kernel keeps all but the j th elements unchanged and samples that component from its conditional distribution given the remaining components. These d kernels can be applied in a deterministic order (systematic scan Gibbs-sampling) or stochastically (random-scan Gibbs sampling).

To verify that Gibbs sampling leads to the correct invariant distribution it suffices for each of the kernels to have the correct invariant distribution independently, and this is trivially so, because

$$\begin{aligned} & \int p(x_{t-1})k_j(x_t|x_{t-1})dx_{t-1} \\ &= \int p(x_{t-1})p(x_t^j|x_{t-1}^{-j})\delta(x_{t-1}^{-j} - x_t^{-j})dx_{t-1} \\ &= p(x_t^{-j})p(x_t^j|x_t^{-j}) = p(x_t) \end{aligned}$$

Gibbs sampling provides a simple algorithm with the properties which are required, but it does require that a suitable collection of conditional distributions are known and can be sampled from and it can perform poorly if the distribution has strongly correlated components (although this can sometimes be addressed by reparameterization). The difficulty with sampling from suitable conditional distributions is a real one for many interesting problems. Even if a closed-form expression for these distributions can be obtained (and this is not often the case), it is often impossible to obtain samples from them.

Figures 1(a) and **1(b)** illustrate the difficulty with correlated random variables. The first shows the trajectory of a Gibbs sampler targeting an uncorrelated bivariate normal distribution; the second shows the trajectory for a highly correlated case. When the variables are highly correlated, conditional upon the knowledge of one the distribution of the other is concentrated and the typical moves are very small leading to a poor exploration of the distribution on a global scale. In **Figure 1(b)** it can be seen that this results in the sampler failing to explore the majority of the distribution. In the Gaussian case, one

could solve this by reprojecting the problem onto the space of eigenvectors of the covariance matrix, and hence eliminating the correlation (in fact, in the Gaussian case this renders the two variables independent), but such reparametrizations can be difficult to find in general.

A technique closely related to Gibbs sampling is based upon the same principle as rejection sampling: obtaining samples of a random variable with density f is equivalent to obtaining samples uniformly from the region beneath the density and discarding the additional dimension. Slice sampling, rather than using rejection sampling to produce samples uniformly from such a region, constructs a Markov chain which has as the uniform distribution over such a region as its stationary distribution. There are a number of difficulties associated with implementing slice samplers in practice.

The Metropolis–Hastings Algorithm

A more general strategy is clearly needed for many interesting problems. The Metropolis–Hastings algorithm is one of a number of algorithms which were proposed to impose detailed balance on a Markov chain using a rejection mechanism: a general proposal kernel is used, but with some probability the chain remains in its present position rather than accepting the new value. It has been shown to be optimal among those algorithms in the sense that it has a smaller associated variance than other options. Intuitively, the algorithm proceeds at each iteration by sampling a proposed value, and this value is accepted with a probability which corrects for the relative probability of the state under the proposal and target distributions.

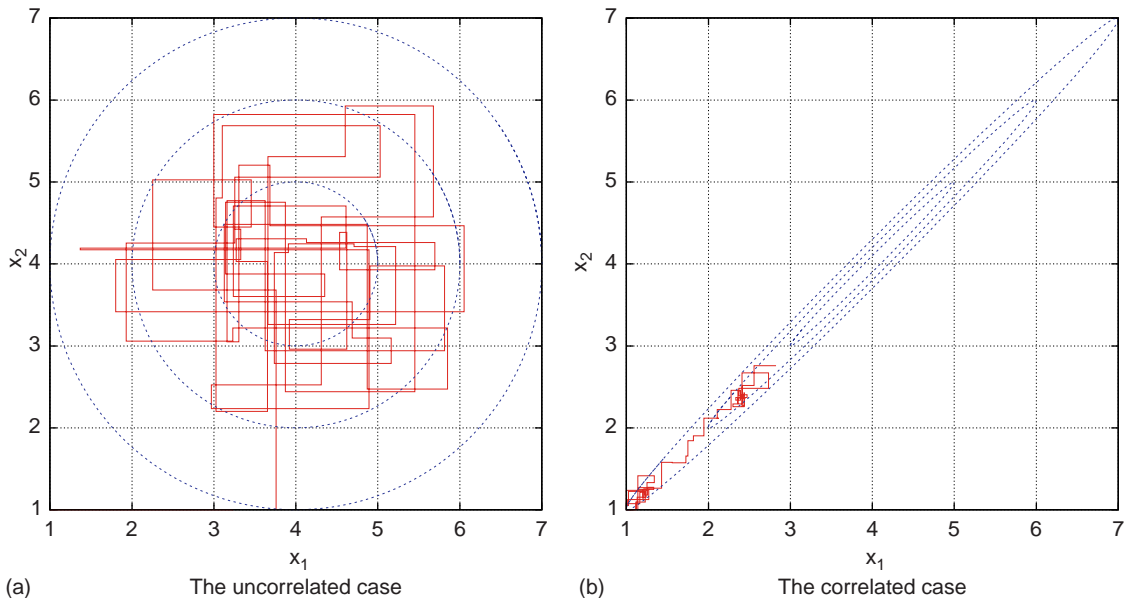


Figure 1 Gibbs sampling: (a) the uncorrelated case and (b) the correlated one. The solid line shows the trajectory of the first hundred samples; dashed lines are contours of the distribution.

If rejection occurs the existing state is replicated — notice this difference from the simple rejection sampling algorithm. More formally, the algorithm for sampling from p using a proposal q proceeds at iteration t as

1. Sample $X' \sim q(\cdot | X_{t-1})$.
2. Calculate

$$\alpha(X_{t-1}, X') = \frac{p(X')q(X_{t-1}|X')}{p(X_{t-1})q(X'|X_{t-1})}.$$

3. Sample $U \sim \mathcal{U}[0, 1]$.
4. If $U \leq \alpha$, set $X_t = X'$.
5. Otherwise, set $X_t = X_{t-1}$.

The acceptance probability α depends upon only the ratio of the target probability density at the proposed point to that at the current point (in addition to the ratio of proposal densities). This has the significant consequence that it is only necessary to evaluate the density up to a normalizing constant. One common choice of q is a symmetric random-walk kernel (such as a Gaussian distribution centered on the previous value). In this case, the acceptance probability α simplifies to $p(X')/p(X_{t-1})$ due to the symmetry of the proposal kernel. This approach is often referred to as the random-walk Metropolis algorithm and corresponds to the first MCMC algorithm to be proposed. Another choice, leading to something termed the independence sampler, is to employ a proposal distribution which is entirely independent of the previous position. Of course, situations between these two extremes exist and often produce better results (e.g., using a proposal comprising a mixture of a diffuse proposal distribution and a local random walk). It is also noteworthy that Gibbs sampling may be interpreted as a Metropolis–Hastings algorithm in which a particular choice of proposal kernel is made, guaranteeing that $\alpha = 1$ at all times (i.e., every proposed value is accepted). Although the Metropolis–Hastings algorithm allows a broad range of models to be explored, care is still required to design algorithms with proposals which allow a thorough exploration of the space and to assess their convergence. In order for a sampler to perform well, it is necessary for it to explore the entire support of the distribution and to move around it quickly. This can only be achieved through the use of good proposal distributions.

By way of an example, consider the random-walk Metropolis–Hastings algorithm. In contrast to the rejection sampling case, it is not desirable to maximize the acceptance probability of proposed moves in such an algorithm. Using a very small proposal variance leads to moves which are almost always accepted (α is always close to one if the density is continuous and proposed moves are small) but movements are all very small, leading to poor

exploration of the distribution on a global scale. In contrast, if a very large proposal variance is used then most moves are proposed in regions of very little probability and many moves are rejected, leading to the sampler sticking with a single value for many iterations. Somewhere in between, good performance can often be obtained. In addition to illustrating these phenomena for a simple example, **Figure 2** illustrates a simple qualitative approach to checking for convergence: looking at traces of the Markov chain – do they move well and do they cover the entire space? Notice that the high variance proposal produces a reasonable looking histogram because of a poor choice of bin width and the low variance proposal would appear to explore the space well if it was not known that there were three modes. A common feature of such diagnostic approaches is that great care is required in order to reach the correct conclusions.

Simulated Annealing

Another common MCMC algorithm, albeit one which is often considered in isolation, is simulated annealing (SA). This is an optimization, rather than integration, algorithm which simulates a time inhomogeneous Markov chain which is intended to provide samples from the regions of parameter space which minimize some objective function. Given a function, $H(x)$, which one wishes to find the minima of (usually the function of interest, but it can be any other function which shares its minimizers), SA provides a numerical method for locating its minima.

Modeled upon the physical annealing processes, SA employs densities proportional to $\exp(-\beta H(x))$ in order to find the value(s), x^* , of its argument which minimize $H(x)$. Corresponding to an inverse temperature in physical annealing, β controls how concentrated these distributions are about the optima: if $\beta \ll 1/H(x^*)$ then the distribution is very flat; as $\beta \rightarrow \infty$ the distribution becomes singular with all of its mass located at the optima.

The simulated annealing algorithm simulates a Markov chain with one such distribution as its invariant distribution at each iteration using the same mechanism as the Metropolis–Hastings algorithm. However, β is gradually increased as the algorithm runs leading to a time-inhomogeneous chain. This allows a single simulation run to be initiated at small β , allowing the chain to explore the space well, and to reach the minimizers which have increasing mass as β increases. Selecting the sequence of values taken by β is nontrivial; if it increases too quickly the simulation will become trapped in a local optimum but the more slowly it increases the more expensive each simulation becomes.

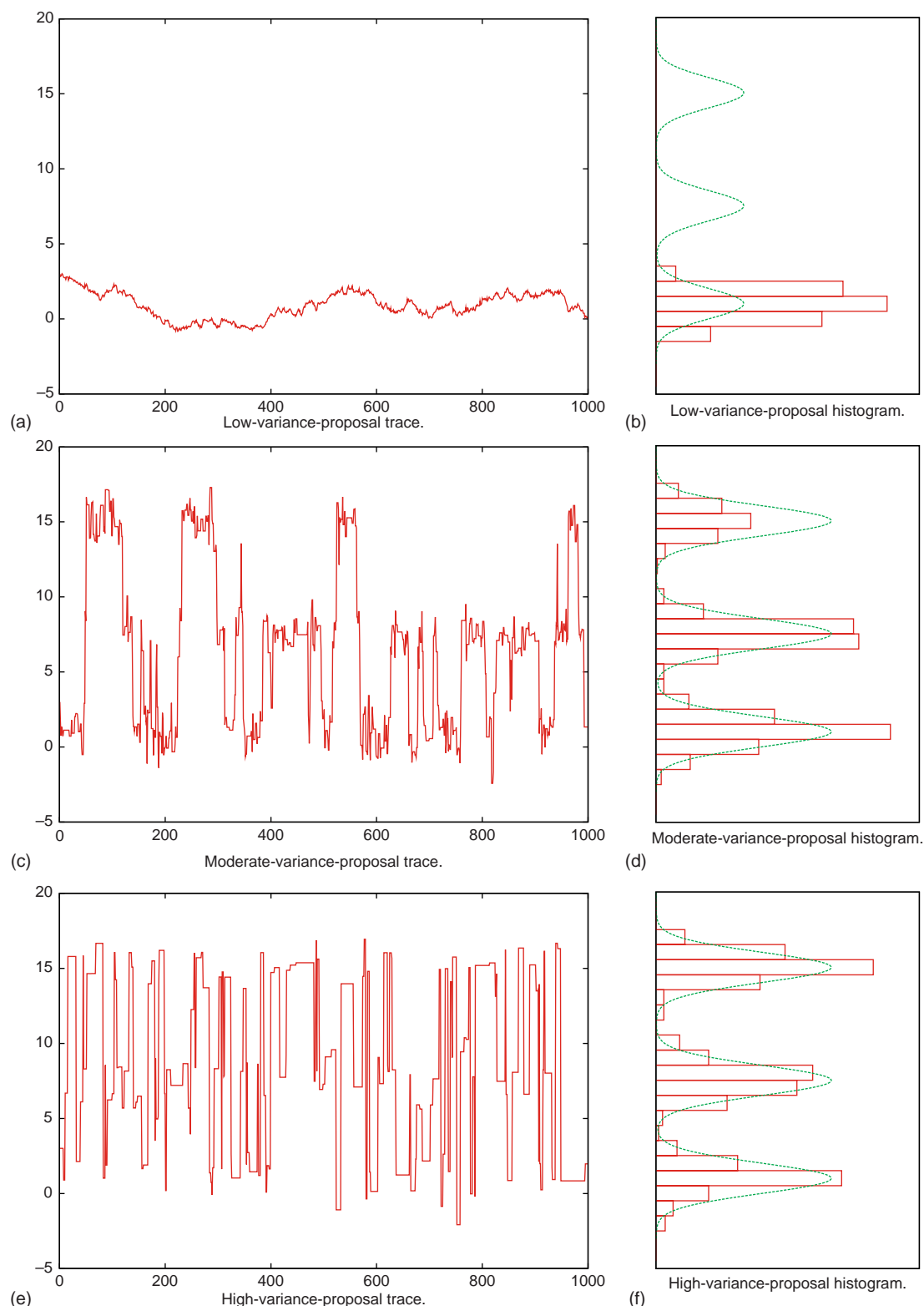


Figure 2 Random-walk Metropolis output: sample sequences and associated histograms (with target density illustrated) for three variances of proposal kernel – (a) low-variance-proposal trace; (b) low-variance-proposal histogram; (c) moderate-variance-proposal trace; (d) moderate-variance-proposal histogram; (e) high-variance-proposal trace; and (f) high-variance-proposal histogram.

Implementation of MCMC Algorithms

Sample Quality

As an MCMC algorithm is rarely initialized from its invariant distribution, there might be some concern that its initial values might bias results even if it does approach this equilibrium distribution later on. To compensate for this, a burn-in period is often implemented: the first N samples being discarded, with N being chosen to be large enough that the chain has reached its stationary regime by this time.

In order to reduce the correlation between samples provided by an MCMC algorithm, it was once popular to thin the chain by recording only every k th value produced by the sampler. In fact, this increases the variance and is undesirable. However, if only limited memory is available then it is possible to improve performance by running a longer chain and keeping every k th value rather than running a shorter chain.

Assessing Convergence

It is also important to assess whether the Markov chain has converged to its stationary distribution. That is, does the chain explore the space fast enough, relative to the number of samples that have been obtained, that the samples produced by it provide a good representation of the stationary distribution of the chain? In general, this is a difficult task. As illustrated in **Figure 2** graphs of sample values against iteration number can provide some information, but are far from infallible. Looking at parameter traces of this form is one common qualitative technique for the assessment of convergence. Unfortunately, it suffers from the problem that it only provides a characterization of the behavior of the chain in the region that it *has* explored. This means that such a technique will fail to diagnose at least one problem: a failure of the chain to reach all parts of the space.

Numerous quantitative methods have been developed. Some of these make use of the output of single chain; these suffer from a similar problem to the simple graphical method. Other employ an ensemble of simulation runs, comparing the output produced by several chains by various techniques – in this case, the usefulness of the resulting analysis can be sensitive to the initialization of these chains. An enormous literature has developed around the assessment of convergence of MCMC methods and it is not possible to adequately summarize that literature here.

Using the Sample

The most direct use that can be made of the sample obtained from an MCMC algorithm is to estimate the sample average of some quantity of interest, and to use this as a proxy for the expectation of that quantity under

the stationary distribution of the Markov chain. However, the sample actually provides an approximate representation of the entire distribution of the quantity of interest and one can use this to obtain considerably more information. For example, given a one-dimensional function, φ , from the space in which samples were obtained to a real number, its full distribution can be summarized using a histogram of its values at the sample points. Summary statistics, such as quantiles, can be obtained in the same manner. For example, if one orders the samples obtained such that $\varphi(x_1) < \varphi(x_2) < \dots < \varphi(x_n)$ then the q th percentile of the distribution of $\varphi(X)$ under the stationary distribution of the Markov chain may be approximated with $\varphi(x_{nq/100})$.

Model Selection

In many settings, often those involving model selection it is useful to have an algorithm which can deal with a collection of densities defined over spaces of different dimension (e.g., consider choosing not just the parameters of a mixture model, but also the number of mixture components). Doing this is technically nontrivial, as the densities are not comparable (just as the mass per unit area of a sheet of steel cannot be compared directly with the density of a steel block). One solution to this problem is termed Reversible Jump Markov chain Monte Carlo and, loosely speaking, it employs a dimension-matching construction in order to produce a Markov chain which explores these spaces of differing dimension whilst retaining reversibility.

Software

Implementing Monte Carlo methods in standard statistical environments is possible, but can be challenging. A number of specialized software packages have been developed to allow nonspecialists to efficiently employ these techniques. Among these are WinBUGS (Bayesian inference Using Gibbs Sampling) a rather general environment for performing MCMC which allows for many algorithms other than Gibbs sampling to be employed and MLWin, a package designed specifically for Bayesian multilevel modeling. Although still not entirely automatic, these programs dramatically simplify the use of these algorithms and provide some amount of support in the assessment of convergence and related issues. However, care is still used to ensure that meaningful results are obtained.

Selected Recent Developments

Although MCMC has driven a revolution in statistics over the past decades, it has limitations. Assessment of convergence is difficult, considerable effort is required to produce efficient samplers and some problems remain

intractable. In order to address these difficulties, particularly the final two, a number of extensions and novel techniques have been developed in recent years.

Perfect simulation is a technique which has been developed to combat the difficulties associated with convergence assessment and the use of a collection of variables which are not iid. Under particular (and, unfortunately, somewhat restrictive) circumstances it is possible to obtain a single sample with the correct distribution by simulating a Markov Chain in such a way that its final value is independent of its beginnings. In order to prevent bias associated with the time at which the sample is produced, considerable care is required in the design of such algorithms.

Two approaches have been developed to reduce the amount of work required to design an efficient sampler. One is to adaptively tune the parameters of the proposal distribution as a sampler is run (ensuring that these lead to the correct distribution introduces nontrivial technical difficulties and this should not be done without thorough consideration of these issues). The other involves using the entire history of the Markov chain as a representation of the distribution of interest and incorporating this into the proposal distribution (again, this is nontrivial to implement in a theoretically justifiable manner).

Finally, much research has been done recently in the area of population-based methods (going under various names: annealed importance sampling, population Monte Carlo, and sequential Monte Carlo – the last of these proposed frameworks encompasses almost all proposed techniques to date). These approaches use a collection of many samples at each iteration and propagate this population around using importance sampling techniques, combined with elements of MCMC. These techniques show great promise and real applications have been developed in the past few years.

Applications

This section provides sketches of two recent areas of educational statistics which have benefited from the use of MCMC methods; these examples show the general usefulness of the techniques but it should be understood that the methods described previously can be used to perform inference in almost any complex statistical model.

Item Response Theory

There has been much recent work on Bayesian inference in item response theory (IRT), employing MCMC algorithms to characterize the posterior distribution of interest. This area is of great importance to test theory, but has been hampered until recently by difficulties associated with performing inference for realistic models.

In a typical IRT model, one has J discrete observations for each of I individuals. Typically, each observation describes performance on a single examination or some other assessment. These results are modeled as a noisy observation of an underlying, unobserved, proficiency influenced by a collection of other factors (some of which vary upon an individual basis; others by assessment criterion). This rapidly leads to a complex model, particularly when factors such as missing observations and variation in the types of assessment are taken into account. Recently developed MCMC algorithms which combine Metropolis–Hastings and Gibbs sampler elements have provided tools which allow the use of models of realistic complexity.

Sensitivity Analysis in Hierarchical Models

Sensitivity analysis is an assessment of the sensitivity of a mathematical model to its modeling assumptions. In statistics, it is often used to determine how sensitive inferences made using a particular model are to the parameters of that model. This is of great importance in real inferential settings, but can be difficult when dealing with complex models.

In addition to inferential tasks, it is possible to use MCMC output to perform sensitivity analysis. A recent application in the field of educational statistics (Seltzer *et al.*, 2002), for example, considered two-level hierarchical models in which the first level corresponded to individual effects and the second to site effects. Particular examples included the efficacy of remedial reading intervention. Using MCMC rather than traditional techniques made it straightforward to employ non-normal distributions in order to ameliorate the effect of outlying observations.

See also: Bayesian Statistical Analysis; Monte Carlo Methods.

Bibliography

- Seltzer, M., Novak, J., Choi, K., and Lim, N. (2002). Sensitivity analysis of hierarchical models employing t level-1 assumptions. *Journal of Education and Behavioural Statistics* 27(2), 181–222.
- Sinharay, S. (2004). Experience with Markov chain Monte Carlo convergence assessment in two psychometric examples. *Journal of Education and Behavioural Statistics* 29(4), 461–488.

Further Reading

- Andrieu, C. (2004). Monte Carlo methods for absolute beginners. In Bousquet, O., *et al.* (eds.) *Advanced Lectures on Machine Learning*, vol. 3176, pp 113–145. Heidelberg: Springer.
- Andrieu, C., de Freitas, N., Doucet, A., and Jordan, M. I. (2003). An introduction to MCMC for machine learning. *Machine Learning* 50, 5–43.

- Bousquet, O., von Luxburg, U., and Rätsch, G. (eds.) (2004). *Advanced Lectures on Machine Learning*, vol. 3176. Heidelberg: Springer.
- Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society B* **63**(3), 411–436.
- Devroye, L. (1986). *Non-uniform Random Variate Generation*. New York: Springer.
- Eckhardi, R. (1987). *Stan Ulam, John von Neumann and the Monte Carlo method*. *Los Alamos Science* **15**, 131–187.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(6), 721–741.
- Geweke, J. (1989). Bayesian inference in econometrics models using Monte Carlo integration. *Econometrica* **57**(6), 1317–1339.
- Green, P. J. (1995). Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds.) (1996). *Markov Chain Monte Carlo In Practice*, 1st edn. London: Chapman and Hall.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika* **52**, 97–109.
- Kirkpatrick, S., Gelatt, C. D. Jr., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science* **270**(4598), 671–680.
- Knuth, D. E. (1998). *The Art of Computer Programming*, vol. 1: Fundamental Algorithms, 3rd edn. Boston: Addison-Wesley.
- Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. *Springer Series in Statistics*. New York: Springer.
- Metropolis, N. (1989). The beginnings of the Monte Carlo method. *Los Alamos Science* **15**, 125–130.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., and Teller, A. H. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092.
- Patz, R. J. and Junker, B. W. (1999). Applications and extensions of MCMC in IRT: Multiple item types, missing data and rated responses. *Journal of Education and Behavioural Statistics* **24**(4), 342–366.
- Patz, R. J. and Junker, B. W. (1999). A straightforward approach to Markov chain monte carlo methods for item response models. *Journal of Education and Behavioural Statistics* **24**(2), 146.
- Peskun, P. H. (1973). Optimum Monte-Carlo sampling using Markov Chains. *Biometrika* **60**, 607–612.
- Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov Chains and applications to statistical mechanics. *Random Structures and Algorithms* **9**, 223–252.
- Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd edn. New York: Springer.
- van Laarhoven, P. J. and Aarts, E. H. (1987). *Simulated Annealing: Theory and Applications*. Norweu, MA: Kluwer Academic Publishers.

Relevant Websites

BUGS Project. <http://www.mrc-bsu.cam.ac.uk/bugs/>
 MCMC Preprint Server. <http://www.statslab.cam.ac.uk/mcmc/>
 MLWin. <http://www.cmm.bristol.ac.uk/MLwiN/>
 Perfect Simulation. <http://dbwilson.com/exact/>
 SMC Preprint Server. <http://www.sigproc.eng.cam.ac.uk/smc/>

Matrix Algebra

H Yanai, St. Luke's College of Nursing, Tokyo, Japan
H Ishii, Nagoya University, Nagoya, Japan

© 2010 Elsevier Ltd. All rights reserved.

Definition, Basic Operations, and Fundamental Properties of Matrix Algebra

Definition and Terminology

A rectangular array of numbers is called a matrix. A matrix with n rows and p columns is referred as an $n \times p$ matrix. If a matrix has n rows and 1 column, it is called an n -dimensional column vector. If a matrix has 1 row and p columns, it is called a p -dimensional row vector. A matrix with 1 row and 1 column, that is, a single numerical value, is called a scalar. A matrix divided into submatrices is called a partitioned matrix.

Boldface capital letters (e.g., \mathbf{A}) are used to represent a matrix. The element in row i and column j of a matrix \mathbf{A} is represented as a_{ij} . Boldface lower letters (e.g., \mathbf{a}) are used to represent a vector.

Some basic types of matrices are listed in Table 1.

Basic Operations

Several matrix operations such as matrix addition and matrix multiplication are given in Table 2.

It should be noted that matrix multiplication is not exchangeable, that is, $\mathbf{AB} \neq \mathbf{BA}$ in general. If $\mathbf{AB} = \mathbf{BA}$, then \mathbf{A} and \mathbf{B} are said to be commutative. If \mathbf{A} is not a null matrix and $\mathbf{A}^2 = \mathbf{O}$, \mathbf{A} is said to be nilpotent. If \mathbf{A} is a square matrix and $\mathbf{A}^2 = \mathbf{A}$, then \mathbf{A} is said to be idempotent.

If \mathbf{A} and \mathbf{B} are partitioned as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$$

and all $\mathbf{A}_{ij} + \mathbf{B}_{ij}$ are defined, then the sum of $\mathbf{A} + \mathbf{B}$ is written as

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} \mathbf{A}_{11} + \mathbf{B}_{11} & \mathbf{A}_{12} + \mathbf{B}_{12} \\ \mathbf{A}_{21} + \mathbf{B}_{21} & \mathbf{A}_{22} + \mathbf{B}_{22} \end{bmatrix}$$

If all $\mathbf{A}_{ik} \mathbf{B}_{kj}$ ($i, j, k = 1, 2$) are present, then the product of \mathbf{AB} is written as

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix}$$

Following are some properties of the transpose. (1) $(\mathbf{A}')' = \mathbf{A}$. (2) $(\mathbf{AB})' = \mathbf{B}' \mathbf{A}'$. (3) $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$. (4) $\mathbf{A}' = \mathbf{A}$ if \mathbf{A} is symmetric. Furthermore,

$$(5) \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}' = \begin{bmatrix} \mathbf{A}_{11}' & \mathbf{A}_{21}' \\ \mathbf{A}_{12}' & \mathbf{A}_{22}' \end{bmatrix}$$

Following are some properties of the trace. (1) $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}')$. (2) $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$. (3) $\text{tr}(k\mathbf{A}) = k \text{tr}(\mathbf{A})$. (4) $\text{tr}(\mathbf{CD}') = \text{tr}(\mathbf{D}'\mathbf{C})$.

For two n -dimensional vectors $\mathbf{a} = (a_1, a_2, \dots, a_n)'$ and $\mathbf{b} = (b_1, b_2, \dots, b_n)'$, their inner product is the sum of products of corresponding components and denoted as

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}'\mathbf{b} = \sum a_i b_i$$

For n -dimensional vector \mathbf{a} , its length or norm is defined by the square root of the sum of its squared components,

$$\|\mathbf{a}\| = (\mathbf{a}'\mathbf{a})^{1/2} = \left(\sum a_i^2 \right)^{1/2}$$

The inner product of two vectors \mathbf{a} and \mathbf{b} equals zero if they are orthogonal. Let θ be the angle between two non-null vectors \mathbf{a} and \mathbf{b} of the same order. Using inner product and norm of vectors, we obtain

$$\cos \theta = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}$$

If $\mathbf{a} \cdot \mathbf{b} = 0$, then $\cos \theta = 0$.

Rank

Vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p$ are called linearly dependent if scalars k_1, k_2, \dots, k_p exist, not all zero, such that $k_1\mathbf{a}_1 + k_2\mathbf{a}_2 + \dots + k_p\mathbf{a}_p = \mathbf{0}$. If no such scalars exist then the p vectors are called linearly independent.

The rank of a matrix \mathbf{A} is defined as the maximum number of linearly independent column or row vectors in \mathbf{A} . The maximum rank of an $n \times p$ matrix \mathbf{A} is $\min(n, p)$. It is defined to have full row rank if $\text{rank}(\mathbf{A}) = n (\leq p)$, and is defined to have full column rank if $\text{rank}(\mathbf{A}) = p (\leq n)$. If a matrix has both row full rank and column full rank, it is called nonsingular. Obviously, a nonsingular matrix is square. An $n \times n$ matrix of rank less than n is called singular.

Following are some properties of rank. (1) $0 \leq \text{rank}(\mathbf{A}) \leq \min(n, p)$, \mathbf{A} is an $n \times p$ matrix. (ii) $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}')$. (iii) $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{AA}') = \text{rank}(\mathbf{A}'\mathbf{A})$. (iv) If \mathbf{A} and \mathbf{B} are $n \times p$ matrices and $(\mathbf{A} : \mathbf{B})$ is an $n \times 2p$ matrix consisting of elements of the matrix \mathbf{B} written to the right of the elements of \mathbf{A} , $\text{rank}(\mathbf{A} + \mathbf{B}) \leq \text{rank}(\mathbf{A} : \mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B})$, and (v) $\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - \text{rank}(\mathbf{A} : \mathbf{B}) \leq \text{rank}(\mathbf{AB}') \leq \min(\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B}))$. (vi) $\text{rank}(\mathbf{BAC}) = \text{rank}(\mathbf{A})$, if \mathbf{B} and \mathbf{C} are nonsingular matrices.

Table 1 Examples of basic type of matrices

Name	Definition	Example
Rectangular	$n \times p$	$\begin{bmatrix} 1 & 2 & 3 \\ -4 & 5 & 6 \end{bmatrix}$
Column vector	$n \times 1$	$\begin{bmatrix} 1 \\ 4 \end{bmatrix}$
Row vector	$1 \times p$	$[1 \ 2 \ 3]$
Scalar	1×1	$[1]$
Square	$n \times n$	$\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}$
Symmetric	Square $a_{ij} = a_{ji}$	$\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$
Diagonal	Square $a_{ij} = 0$ for $i \neq j$	$\begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$
Identity (I)	Square $a_{ii} = 1$ $a_{ij} = 0$ for $i \neq j$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
Upper/lower triangular	Square $a_{ij} = 0$ below/above the diagonal	$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 9 \end{bmatrix}$
Unit vector (1)	$a_i = 1$	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$
Unit matrix (J)	$a_{ij} = 1$	$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$
Null(zero) matrix (0)	$a_{ij} = 0$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Vector Spaces

The set of all n -dimensional vectors is called the n -dimensional Euclidian space and is denoted as E^n . Then a subset S of E^n is said to be the vector subspace (or vector space for short) if any linear combinations of vectors in S are also in the set. The maximum number of linearly independent vectors in a vector space is referred to as the dimension of the vector space. Since rank of a matrix is defined as the maximum number of linearly independent column or row vectors in the matrix, a matrix **X** spans a vector space $S(\mathbf{X})$ whose dimension is equal to $\text{rank}(\mathbf{X})$.

A maximal linearly independent set of vectors in a vector space composes a basis of the vector space. A basis of a vector space is called an orthonormal basis if the vectors in the basis are orthogonal to each other and of unit norms.

In addition, the set of all vectors orthogonal to all the vectors in $S(\mathbf{X})$ is called the orthogonal complement subspace of $S(\mathbf{X})$ in E^n and denoted as $S(\mathbf{X})^\perp$.

Determinants

If **A** is a 2×2 square matrix, the determinant of **A** is calculated as

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

If **A** is a 3×3 square matrix, the determinant of **A** is calculated as

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11}a_{22}a_{33} & -a_{12}a_{21}a_{33} \\ -a_{13}a_{22}a_{31} & -a_{11}a_{23}a_{32} \\ +a_{12}a_{23}a_{31} & +a_{13}a_{21}a_{32} \end{vmatrix}$$

The general definition of the determinant of an $n \times n$ square matrix **A** is

$$|\mathbf{A}| = \sum (-1)^{t(j)} a_{1j_1} a_{2j_2} \dots a_{nj_n}$$

where $t(j)$ is the number of transpositions in a permutation of j_1, j_2, \dots, j_n and the summation is taken over all permutations. $|\mathbf{A}| \neq 0$ if **A** is nonsingular. $|\mathbf{A}| = 0$ if **A** is singular. If **A** is triangular or diagonal, $|\mathbf{A}| = \prod a_{ii}$.

Following are some properties of determinants. (1) $|k\mathbf{A}| = k^n |\mathbf{A}|$. (2) $|\mathbf{A}'| = |\mathbf{A}|$. (3) $|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$. (4) $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}| = |\mathbf{B}||\mathbf{A}| = |\mathbf{BA}|$, matrices of the same order. (5) $|\mathbf{A} + \mathbf{BC}| = |\mathbf{A}| |\mathbf{I}_n + \mathbf{A}^{-1}\mathbf{BC}| = |\mathbf{A}| |\mathbf{I}_p + \mathbf{CA}^{-1}\mathbf{B}|$, **B** and **C'** are $n \times p$ matrices. Furthermore,

$$(6) \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A}||\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}| = |\mathbf{D}||\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}|$$

if **A** and **D** are nonsingular matrices.

Inverse Matrix

The inverse of a square matrix **A** is the unique matrix \mathbf{A}^{-1} which satisfies $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. The inverse exists if and only if **A** is nonsingular, that is, $|\mathbf{A}| \neq 0$.

If **A** is a 2×2 matrix,

$$\mathbf{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

Following are some properties of inverse matrix. (1) $(k\mathbf{A})^{-1} = (1/k)\mathbf{A}^{-1}$. (2) $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$. (3) $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$. (4) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. Furthermore, if **A** is an $n \times n$ nonsingular matrix; **B** as an $n \times p$ matrix; **C** as a $p \times n$ matrix; and **D** as a $p \times p$ nonsingular matrix,

$$(5) (\mathbf{A} + \mathbf{BDC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{D}^{-1} + \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1}$$

$$(6) \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{BECA}^{-1} & -\mathbf{A}^{-1}\mathbf{BE} \\ -\mathbf{ECA}^{-1} & \mathbf{E} \end{bmatrix} = \begin{bmatrix} \mathbf{F} & -\mathbf{FBD}^{-1} \\ -\mathbf{D}^{-1}\mathbf{CF} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{CFBD}^{-1} \end{bmatrix}$$

where $\mathbf{E} = (\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1}$ and $\mathbf{F} = (\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1}$

Table 2 Selected basic matrix operations

Operation	Example
Addition/subtraction $\mathbf{A} \pm \mathbf{B} = [a_{ij} \pm b_{ij}]$ \mathbf{A}, \mathbf{B} of the same order	$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} + \begin{bmatrix} 1 & 2 & 0 \\ 4 & 1 & -2 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 3 \\ 8 & 6 & 4 \end{bmatrix}$
Scalar multiplication $k\mathbf{A} = [k a_{ij}]$	$2 \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{bmatrix}$
Matrix multiplication $\mathbf{AB} = [\sum_k a_{ik} b_{kj}]$ $\mathbf{A}_{n \times p}, \mathbf{B}_{p \times m}$	$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & -4 \\ 2 & 1 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} 5 & -8 \\ 14 & -23 \end{bmatrix}$
Transpose $\mathbf{A}' = [a_{ji}]$	$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}' = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$
Trace $\text{tr } \mathbf{A} = \sum a_{ii}$ \mathbf{A} of square matrix	$\text{tr} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = 1 + 5 + 9 = 15$
Inner product $\mathbf{a} \cdot \mathbf{b} = \mathbf{a}'\mathbf{b} = \sum a_i b_i$ \mathbf{a}, \mathbf{b} of the same order	$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = 4 + 10 + 18 = 32$
Norm of vector $\ \mathbf{a}\ = (\mathbf{a}'\mathbf{a})^{1/2}$ $= (\sum a_i^2)^{1/2}$	$(1^2 + 2^2 + 3^2)^{1/2} = \sqrt{14}$
Norm of matrix $\ \mathbf{A}\ = \text{tr}(\mathbf{A}'\mathbf{A})^{1/2}$ $= (\sum \sum a_{ij}^2)^{1/2}$	$(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2)^{1/2} = \sqrt{91}$

A matrix \mathbf{A} is called an orthogonal matrix, if \mathbf{A} is nonsingular and $\mathbf{A}^{-1} = \mathbf{A}'$, that is, $\mathbf{AA}' = \mathbf{A}'\mathbf{A} = \mathbf{I}$. The inner products between column vectors in an orthogonal matrix are $\mathbf{a}'_i \mathbf{a}_i = \|a_i\|^2 = 1$ and $\mathbf{a}'_i \mathbf{a}_j = 0 (i \neq j)$.

Let us illustrate using an application of an inverse matrix. Consider the simultaneous equation of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2 \end{aligned}$$

with the condition of $a_{11}a_{22} \neq a_{21}a_{12}$. Let $\mathbf{x} = (x_1, x_2)'$, $\mathbf{b} = (b_1, b_2)'$. Then the above equations are represented as $\mathbf{Ax} = \mathbf{b}$, which leads to $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

For example, if

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

then

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \frac{1}{-2} \begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$$

Positive-Definite and Positive-Semidefinite Matrices

Let \mathbf{x} be a column vector of order n and \mathbf{A} be a symmetric matrix of order n . Then, $\mathbf{x}'\mathbf{Ax}$ is called a quadratic form in the vector \mathbf{x} .

If $\mathbf{x}'\mathbf{Ax} > 0$ for all \mathbf{x} except for $\mathbf{x} = 0$, the quadratic form is called positive definite. If $\mathbf{x}'\mathbf{Ax} \geq 0$ for all \mathbf{x} and there exists some non-null \mathbf{x} which satisfies $\mathbf{x}'\mathbf{Ax} = 0$, then the quadratic form is called positive semidefinite. If a quadratic form is positive definite or positive semidefinite, the quadratic form is called nonnegative definite. The matrix \mathbf{A} is correspondingly called a positive-definite/positive-semidefinite/nonnegative-definite matrix. If a matrix \mathbf{A} is positive definite, it is represented as $\mathbf{A} > \mathbf{O}$. Similarly, if a matrix \mathbf{A} is positive semidefinite, it is represented as $\mathbf{A} \geq \mathbf{O}$.

For example, let \mathbf{A} be a correlation matrix of order 2, that is,

$$\mathbf{A} = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$$

The quadratic form $\mathbf{x}'\mathbf{Ax}$ can be written as

$$\begin{aligned} \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= x_1^2 + 2rx_1x_2 + x_2^2 \\ &= (x_1 + rx_2)^2 + (1 - r^2)x_2^2 \end{aligned}$$

Then, \mathbf{A} is positive definite if $-1 < r < 1$. \mathbf{A} is positive semidefinite if $r = \pm 1$.

\mathbf{A} is usually taken as symmetric. This is because $\mathbf{x}'\mathbf{Ax}$ can be rewritten as $\mathbf{x}'\left[\frac{1}{2}(\mathbf{A} + \mathbf{A}')\right]\mathbf{x}$ and because $\frac{1}{2}(\mathbf{A} + \mathbf{A}')$

is symmetric even if \mathbf{A} is not symmetric. This means that $\mathbf{x}'\mathbf{B}\mathbf{x}$ and $\mathbf{x}'\mathbf{C}\mathbf{x}$ are the same whenever $\mathbf{B} + \mathbf{B}' = \mathbf{C} + \mathbf{C}'$.

If $\mathbf{A} > \mathbf{O}$, then \mathbf{A} is nonsingular and full rank, moreover $\mathbf{A}^{-1} > \mathbf{O}$. If $\mathbf{A} \geq \mathbf{O}$, then \mathbf{A} is singular and not full rank.

Eigenvalue, Spectral and Singular value Decompositions

Eigenvalues and Eigenvectors

A scalar λ is called an eigenvalue of an $n \times n$ matrix \mathbf{A} if there exists a non-null n -dimensional vector \mathbf{x} such that $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, or equivalently $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$. A vector \mathbf{x} which satisfies this equation is called an eigenvector for the eigenvalue λ . An eigenvector is called a normalized eigenvector if its norm is equal to 1.

Since an eigenvector is a non-null vector, $\mathbf{A} - \lambda\mathbf{I}$ should be singular; hence, its determinant is 0. Then, eigen values of a matrix \mathbf{A} can be obtained as the n roots of the equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$. This equation is called the characteristic equation of a matrix \mathbf{A} .

For example, the eigenvalues of the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}$$

are obtained by solving

$$\begin{vmatrix} 1 - \lambda & 3 \\ 4 & 2 - \lambda \end{vmatrix} = 0$$

that is,

$$\begin{aligned} (1 - \lambda)(2 - \lambda) - 3 \cdot 4 &= \lambda^2 - 3\lambda - 10 \\ &= (\lambda + 2)(\lambda - 5) \\ &= 0 \end{aligned}$$

That is, $\lambda = -2$ or 5 . Then it follows that $[-3, 3]'$ is an eigenvector for the eigenvalue -2 , and $[-3, -4]'$ for 5 . For example,

$$\begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} -3 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \end{bmatrix} = -2 \begin{bmatrix} -3 \\ 3 \end{bmatrix}$$

Following are some properties of eigenvalues and eigenvectors. (1) $\mathbf{A}(c\mathbf{x}) = \lambda(c\mathbf{x})$. (2) $\mathbf{A}^k\mathbf{x} = \lambda^k\mathbf{x}$. (3) $\sum \lambda_i = \text{tr}(\mathbf{A})$, λ_i are the i th eigenvalues of a matrix \mathbf{A} . (4) $\prod \lambda_i = |\mathbf{A}|$. (v) If $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ and \mathbf{B} is a nonsingular matrix, \mathbf{A} and $\mathbf{B}\mathbf{A}\mathbf{B}^{-1}$ have the same eigenvalues and an eigenvector for $\mathbf{B}\mathbf{A}\mathbf{B}^{-1}$ is $\mathbf{B}\mathbf{x}$.

Spectral Decomposition

Any symmetric matrix \mathbf{A} can be written as

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}' = \sum \lambda_i \mathbf{v}_i \mathbf{v}_i'$$

where $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues of \mathbf{A} and \mathbf{V} is an orthogonal matrix whose column vectors are normalized

eigenvectors. This decomposition is called as spectral decomposition.

For example, eigenvalues of a symmetric matrix

$$\mathbf{A} = \begin{bmatrix} 41 & 12 \\ 12 & 34 \end{bmatrix}$$

are 50 and 25. The corresponding eigenvectors are $(4/5, 3/5)'$ and $(-3/5, 4/5)'$. Then, \mathbf{A} can be written as

$$\begin{aligned} \begin{bmatrix} 41 & 12 \\ 12 & 34 \end{bmatrix} &= \begin{bmatrix} 4/5 & -3/5 \\ 3/5 & 4/5 \end{bmatrix} \begin{bmatrix} 50 & 0 \\ 0 & 25 \end{bmatrix} \begin{bmatrix} 4/5 & 3/5 \\ -3/5 & 4/5 \end{bmatrix} \\ &= 50 \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix} \begin{bmatrix} 4/5 & 3/5 \end{bmatrix} + 25 \begin{bmatrix} -3/5 \\ 4/5 \end{bmatrix} \begin{bmatrix} -3/5 & 4/5 \end{bmatrix} \end{aligned}$$

If \mathbf{A} is a nonsingular symmetric matrix, $\mathbf{A}^r = \mathbf{V}\mathbf{\Lambda}^r\mathbf{V}'$. If \mathbf{A} is a nonsingular symmetric idempotent matrix, eigenvalues of \mathbf{A} should be 0 or 1 since $\mathbf{A}^2 = \mathbf{A}$ leads to $\mathbf{\Lambda}^2 = \mathbf{\Lambda}$.

Singular-Value Decomposition

Let \mathbf{A} be an $n \times p$ matrix. Then $\mathbf{A}'\mathbf{A}$ is a $p \times p$ symmetric matrix and is decomposed as $\mathbf{A}'\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'$, where \mathbf{V} is an orthogonal matrix whose columns are eigenvectors of $\mathbf{A}'\mathbf{A}$ and $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal elements are eigenvalues of $\mathbf{A}'\mathbf{A}$.

Similarly, $\mathbf{A}\mathbf{A}'$ is an $n \times n$ symmetric matrix and is decomposed as $\mathbf{A}\mathbf{A}' = \mathbf{U}\mathbf{\Lambda}\mathbf{U}'$, where \mathbf{U} is an orthogonal matrix whose columns are eigenvectors of $\mathbf{A}\mathbf{A}'$ and $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal elements are eigenvalues of $\mathbf{A}\mathbf{A}'$. Nonzero eigenvalues of $\mathbf{A}'\mathbf{A}$ and those of $\mathbf{A}\mathbf{A}'$ are the same.

Using \mathbf{U} , \mathbf{V} and $\mathbf{\Lambda}$, \mathbf{A} can be written as

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{V}'$$

This is called as singular-value decomposition of \mathbf{A} .

Projection Matrix

Orthogonal Projector

A square matrix \mathbf{P} is called an orthogonal projector (or projection matrix) if it is both idempotent and symmetric, that is, $\mathbf{P}^2 = \mathbf{P}$ and $\mathbf{P}' = \mathbf{P}$ (Rao and Yanai, 1979).

For a given matrix \mathbf{X} of order $n \times p$ ($n \geq p$) where $\mathbf{X}'\mathbf{X}$ is nonsingular, let $\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{Q}_X = \mathbf{I} - \mathbf{P}_X$. Then it follows that $\mathbf{P}_X^2 = \mathbf{P}_X$, $\mathbf{Q}_X^2 = \mathbf{Q}_X$ and $\mathbf{P}_X' = \mathbf{P}_X$, $\mathbf{Q}_X' = \mathbf{Q}_X$. Furthermore, it also follows that $\mathbf{P}_X\mathbf{Q}_X = \mathbf{Q}_X\mathbf{P}_X = \mathbf{O}$. Thus, the square matrices \mathbf{P}_X and \mathbf{Q}_X are called orthogonal projectors onto the range spaces $S(\mathbf{X})$ and $S(\mathbf{X})^\perp$.

For example, let $\mathbf{x} = (1, 1, 1)'$. Then,

$$\mathbf{P}_x = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \text{ and } \mathbf{Q}_x = \mathbf{I} - \mathbf{P}_x = \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

Obviously, $\mathbf{P}_x^2 = \mathbf{P}_x$, $\mathbf{Q}_x^2 = \mathbf{Q}_x$ and $\mathbf{P}_x \mathbf{Q}_x = \mathbf{Q}_x \mathbf{P}_x = \mathbf{O}$ holds true. Let $\mathbf{y} = (y_1, y_2, y_3)'$. It is interesting to note that

$$\mathbf{Q}_x \mathbf{y} = (\mathbf{I} - \mathbf{P}_x) \mathbf{y} = \begin{bmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ y_3 - \bar{y} \end{bmatrix}$$

that is called the deviation score vector from the mean.

An orthogonal projector has following properties:

1. Eigenvalues of \mathbf{P}_X are 1 or 0.
2. $\text{tr}(\mathbf{P}_X) = \text{rank}(\mathbf{P}_X)$.
3. In addition to \mathbf{X} , let \mathbf{Y} be a matrix of order $n \times q$ satisfying $S(\mathbf{X}) = S(\mathbf{Y})$. Then $\mathbf{P}_X = \mathbf{P}_Y$. Thus it follows that an orthogonal projector is uniquely defined onto a given range space $S(\mathbf{X})$ for any choice of \mathbf{X} spanning $V = S(\mathbf{X})$.
4. Let \mathbf{y} be an r -dimensional vector and let \mathbf{b} be a p -dimensional vector. Then, minimizing $\|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2 = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} - \mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$, we obtain $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$ which leads to $\|\mathbf{y} - \mathbf{P}_X \mathbf{y}\|^2$. $\mathbf{P}_X \mathbf{y}$ is called the orthogonal projection of \mathbf{y} onto $S(\mathbf{X})$, and thus it follows that,
5. $\|\mathbf{P}_X \mathbf{y}\| \leq \|\mathbf{y}\|$. Equality holds when $\mathbf{y} \in S(\mathbf{X})$.
6. Let \mathbf{X} and \mathbf{Y} be matrices of orders $n \times p$ and $n \times q$ and let \mathbf{P}_X and \mathbf{P}_Y be orthogonal projectors onto $S(\mathbf{X})$ and $S(\mathbf{Y})$, respectively. Then $\mathbf{P}_X \mathbf{P}_Y$ is the orthogonal projector onto $S(\mathbf{X}) \cap S(\mathbf{Y})$, if and only if $\mathbf{P}_X \mathbf{P}_Y = \mathbf{P}_Y \mathbf{P}_X$. Further, $\text{tr}(\mathbf{P}_X \mathbf{P}_Y) \leq \min(\text{rank}(\mathbf{X}), \text{rank}(\mathbf{Y}))$. If $S(\mathbf{Y})$ is a subspace of $S(\mathbf{X})$, then $\mathbf{P}_X \mathbf{P}_Y = \mathbf{P}_Y \mathbf{P}_X = \mathbf{P}_Y$, $(\mathbf{P}_X - \mathbf{P}_Y)^2 = \mathbf{P}_X - \mathbf{P}_Y$ and $\mathbf{P}_X - \mathbf{P}_Y$ is the orthogonal projection matrix onto $S(\mathbf{X}) \cap S(\mathbf{Y})^\perp$.

Generalized Inverse Matrix

Weaker g-Inverse

When \mathbf{A} is singular or rectangular, \mathbf{A}^{-1} does not exist. In this case, a generalized inverse (g-inverse) \mathbf{A}^- can be defined as a matrix satisfying

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$$

(Rao, 1962). In general, \mathbf{A}^- is not unique as shown in the following example. Let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \text{ and } \mathbf{A}^- = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Then, $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$ implies $a + b + c + d = 1$, which leads to

$$\mathbf{A}^- = \begin{bmatrix} a & b \\ c & 1 - a - b - c \end{bmatrix}$$

and \mathbf{A}^- is not unique.

Following are some properties of the weaker g-inverse.

- (1) $(\mathbf{A}\mathbf{A}^-)^2 = \mathbf{A}\mathbf{A}^-$, $(\mathbf{A}^-\mathbf{A})^2 = \mathbf{A}^-\mathbf{A}$.
- (2) $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}\mathbf{A}^-) = \text{tr}(\mathbf{A}\mathbf{A}^-) = \text{rank}(\mathbf{A}^-\mathbf{A}) = \text{tr}(\mathbf{A}^-\mathbf{A})$.
- (3) $\text{rank}(\mathbf{A}^-) \geq \text{rank}(\mathbf{A})$.

- (4) If $\mathbf{A}\mathbf{x} = \mathbf{0}$ then $\mathbf{x} = (\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{z}$ for some \mathbf{z} .
- (5) $\mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'\mathbf{A} = \mathbf{A}$.
- (6) $(\mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'\mathbf{A})' = \mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'\mathbf{A}$.

Using weaker g-inverse, an orthogonal projector when $\mathbf{X}'\mathbf{X}$ is singular can be defined as $\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$. Property (5) and (6) lead to $\mathbf{P}_X^2 = \mathbf{P}_X$ and $\mathbf{P}_X' = \mathbf{P}_X$ for any choice of $(\mathbf{X}'\mathbf{X})^-$.

Reflexive g-Inverse

If a matrix \mathbf{A}_r^- satisfies both

$$\mathbf{A}\mathbf{A}_r^-\mathbf{A} = \mathbf{A} \text{ and } \mathbf{A}_r^-\mathbf{A}\mathbf{A}_r^- = \mathbf{A}_r^-$$

then \mathbf{A}_r^- is called a reflexive g-inverse of \mathbf{A} . A reflexive g-inverse has a property of $\text{rank}(\mathbf{A}_r^-) = \text{rank}(\mathbf{A})$. For example, if

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

then

$$\mathbf{A}_r^- = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

where $a + b + c + d = 1$ and $ad = bc$.

Least-Squares g-Inverse

Given a system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{b} \notin S(\mathbf{A})$, then the system is not consistent, that is, this solution does not exist. In such a case, we need a least-squares solution such that $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$ is minimized. The solution is written as $\mathbf{x} = \mathbf{A}_l^-\mathbf{b}$ where \mathbf{A}_l^- satisfies

$$\mathbf{A}\mathbf{A}_l^-\mathbf{A} = \mathbf{A} \text{ and } (\mathbf{A}\mathbf{A}_l^-)' = \mathbf{A}\mathbf{A}_l^-$$

\mathbf{A}_l^- can be written as $(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$ and is called a least-squares g-inverse of \mathbf{A} . It should be noted that $\mathbf{A}\mathbf{A}_l^-$ is the orthogonal projector onto $S(\mathbf{A})$ and thus is unique for any choice of \mathbf{A}_l^- . If \mathbf{A} has column full rank, then $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ is one choice of \mathbf{A}_l^- .

For example, if

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

then

$$\mathbf{A}_l^- = \begin{bmatrix} a & b \\ 1/2 - a & 1/2 - b \end{bmatrix}$$

Minimum Norm g-Inverse

If $\mathbf{A}\mathbf{x} = \mathbf{b}$ has nonunique solutions, we need a solution with a minimum value of $\|\mathbf{x}\|$. The solution is $\mathbf{x} = \mathbf{A}_m^-\mathbf{b}$, where \mathbf{A}_m^- satisfies

$$\mathbf{A}\mathbf{A}_m^-\mathbf{A} = \mathbf{A} \text{ and } (\mathbf{A}_m^-\mathbf{A})' = \mathbf{A}_m^-\mathbf{A}$$

\mathbf{A}_m^- can be written as $\mathbf{A}'(\mathbf{A}\mathbf{A}')^-$ and is called a minimum norm g-inverse of \mathbf{A} . It should be noted that $\mathbf{A}_m^-\mathbf{A}$ is the

orthogonal projector onto $S(\mathbf{A})$ and thus is unique for any choice of \mathbf{A}_m^- . If \mathbf{A} has row full rank, then $\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ is one choice of \mathbf{A}_m^- .

For example, if

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

then

$$\mathbf{A}_m^- = \begin{bmatrix} a & 1/2 - a \\ c & 1/2 - c \end{bmatrix}$$

Moore–Penrose Inverse

If $\mathbf{A}\mathbf{x} = \mathbf{b}$ has no solution, but has nonunique least-squares solutions, we need a least-squares solution with minimum value of $\|\mathbf{x}\|$. The solution is $\mathbf{x} = \mathbf{A}^+ \mathbf{b}$, where \mathbf{A}^+ satisfies $\mathbf{A}\mathbf{A}^+ \mathbf{A} = \mathbf{A}$, $(\mathbf{A}\mathbf{A}^+)' = \mathbf{A}\mathbf{A}^+$, $(\mathbf{A}^+ \mathbf{A})' = \mathbf{A}^+ \mathbf{A}$, and $\mathbf{A}^+ \mathbf{A} \mathbf{A}^+ = \mathbf{A}^+$.

\mathbf{A}^+ is unique and is called Moore–Penrose inverse of \mathbf{A} (Moore, 1920; Penrose, 1955).

Moore–Penrose inverse of

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \text{ is } \mathbf{A}^+ = \begin{bmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{bmatrix}$$

Then, for example, the Moore–Penrose solution of

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ is } \mathbf{x} = \begin{bmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3/4 \\ 3/4 \end{bmatrix}$$

Following are some properties of the Moore–Penrose inverse matrix. (1) $\mathbf{A}^+ = \mathbf{A}^{-1}$ if \mathbf{A} is nonsingular. (2) $(\mathbf{A}^+)^+ = \mathbf{A}$. (3) $(\mathbf{A}')^+ = (\mathbf{A}^+)'$. (4) $\mathbf{A}^+ = (\mathbf{A}'\mathbf{A})^{-1} \mathbf{A}'$ if \mathbf{A} is full column rank. (5) $\mathbf{A}^+ = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ if \mathbf{A} is full row rank. (6) $\mathbf{A}^+ \mathbf{B} = \mathbf{O} \Leftrightarrow \mathbf{A}'\mathbf{B} = \mathbf{O}$. (7) $\mathbf{A}^+ = \mathbf{A}$ if \mathbf{A} is symmetric and idempotent.

If \mathbf{A} is symmetric and positive semidefinite, the spectral decomposition of \mathbf{A} can be written as $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'$, and the Moore–Penrose inverse matrix can be obtained as

$$\mathbf{A}^+ = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}'$$

Kronecker Product and Vec Operator

Kronecker Product

Let $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{kl}]$ be $n \times m$ and $p \times q$ matrices, respectively. Then the following matrix of orders $np \times mq$ is called the Kronecker product (or direct product) of \mathbf{A} and \mathbf{B} :

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1m}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2m}\mathbf{B} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \cdots & a_{nm}\mathbf{B} \end{bmatrix}$$

Following properties hold for Kronecker products of matrices: (1) $(a\mathbf{A}) \otimes (b\mathbf{B}) = ab(\mathbf{A} \otimes \mathbf{B})$. (2) $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$. (3) $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}\mathbf{C} \otimes \mathbf{B}\mathbf{D})$. (4) $\text{rank}(\mathbf{A} \otimes \mathbf{B}) = \text{rank}(\mathbf{A})\text{rank}(\mathbf{B})$.

Vec Operator

Let $\mathbf{A} = [a_1 a_2 \cdots a_m]$ be an $n \times m$ matrix where a_j is the j th column vector of \mathbf{A} . Then $\text{Vec}(\mathbf{A}) = (a_1', a_2', \dots, a_m')'$ is the $nm \times 1$ vector which transforms matrix \mathbf{A} into a vector by stacking all the column vectors in \mathbf{A} one underneath the other. The basic relationship between the Vec operator and the Kronecker product is $\text{Vec}(\mathbf{a}\mathbf{b}') = \mathbf{b} \otimes \mathbf{a}$. Following are some properties of the Vec operator; (1) $\text{tr}(\mathbf{A}\mathbf{B}) = \text{Vec}(\mathbf{A}')' \text{Vec}(\mathbf{B})$. (2) $\text{Vec}(\mathbf{A}\mathbf{B}\mathbf{C}') = (\mathbf{C} \otimes \mathbf{A}) \text{Vec}(\mathbf{B})$. (3) $\text{tr}(\mathbf{A}\mathbf{B}\mathbf{C}\mathbf{D}) = (\text{Vec}(\mathbf{D}))'(\mathbf{A} \otimes \mathbf{C}') \text{Vec}(\mathbf{B}')$.

Matrix Methods in Multivariate Statistical Analysis

Correlation Matrix

Let \mathbf{A} be a raw data matrix of order $n \times p$ where n is the sample size and p is the number of variables. Let $\mathbf{Q} = \mathbf{I} - (1/n) \mathbf{1}\mathbf{1}'$. Then $\mathbf{X} = \mathbf{Q}\mathbf{A}$ is the deviation score matrix from the mean for each of p variables. The covariance matrix of p variables, which is denoted by \mathbf{S} , can be written as $\mathbf{S} = (1/n) \mathbf{X}'\mathbf{X}$. The i th diagonal elements is the variance of i th variable and ij nondiagonal elements are covariances between i th and j th variables.

Let \mathbf{D} be a diagonal matrix whose diagonal elements are equal to \mathbf{S} and let \mathbf{Z} be $\mathbf{X}\mathbf{D}^{-1/2}$. The diagonal elements of $\mathbf{D}^{-1/2}$ are the reciprocal values of standard deviations. Then $\mathbf{R} = (1/n) \mathbf{Z}'\mathbf{Z}$ is the correlation matrix of the p variables. A correlation matrix has following properties: (1) \mathbf{R} is nonnegative definite. (2) $0 \leq |\mathbf{R}| \leq 1$. (3) If $|\mathbf{R}| = 1$ then $\mathbf{R} = \mathbf{I}$.

Principal-component analysis

Principal-component analysis proposed by Hotelling (1933) is one of the most familiar methods of multivariate analysis which uses the spectral decomposition of a correlation coefficient or covariance matrix.

We now show an example of principal-component analysis. Table 3 is the correlation coefficient matrix \mathbf{R} among

Table 3 Correlation matrix among achievement tests

	<i>Native</i>	<i>Social</i>	<i>Math</i>	<i>Natural</i>	<i>English</i>
Native language	1				
Social science	0.537	1			
Mathematics	0.402	0.504	1		
Natural science	0.478	0.624	0.644	1	
English	0.562	0.587	0.563	0.576	1

Table 4 Principal-component loadings

	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>	<i>V</i>
Native language	0.736	0.573	−0.182	−0.311	−0.026
Social science	0.817	0.105	0.498	0.127	−0.240
Mathematics	0.779	−0.470	−0.269	−0.136	−0.286
Natural science	0.837	−0.271	0.178	−0.193	0.397
English	0.826	0.103	−0.257	0.475	0.127
Eigen value	3.197	0.644	0.451	0.394	0.313
Contribution (%)	63.39	12.88	9.02	7.88	6.26

achievement tests for a university entrance examination (National Center for University Entrance Examinations of Japan, 1982). Eigenvalues of the correlation matrix are 3.197, 0.644, 0.451, 0.394, and 0.313. The sum of these five eigenvalues attains five that is equal to the number of the tests. In the principal-component analysis using a correlation coefficient matrix, eigenvalues are the variances of corresponding principal-component scores. The eigenvectors are computed and multiplied by the square root of the corresponding eigenvalues ($\mathbf{F} = \mathbf{V}\mathbf{\Lambda}^{1/2}$) and are given in **Table 4**. These values are called the principal-component loadings. It can be observed that $\mathbf{R} = \mathbf{F}\mathbf{F}' = \mathbf{V}\mathbf{\Lambda}\mathbf{V}'$.

It follows that the number of principal components with eigenvalues greater than 1 is 1. Thus, most of the variability in the data are explained by the first-principal component. The first principal component seems to be the general achievement abilities measured by means of five common achievement tests, since the principal-component loadings of the five tests range from 0.736 to 0.839, which are considerably high. Other applications of matrix algebra to statistics are factor analysis, regression analysis.

See also: Principal Components Analysis.

Bibliography

- Hotelling, H. (1933). Analysis of a complex of statistical variables into, principal components. *Journal of Educational Psychology* **24**(417–441); 498–520.
- Moore, E. H. (1920). On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society* **26**, 394–395.
- National Center for University Entrance Examinations of Japan (1982). *University Entrance Examinations Forum* **12**.
- Penrose, R. (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society* **51**, 406–413.
- Rao, C. R. (1962). A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. *Journal of Royal Statistical Society B* **24**, 152–158.
- Rao, C. R. and Yanai, H. (1979). General definition of a projector, its decomposition and application to statistical problems. *Journal of Statistical Planning and Inference* **3**, 1–17.

Further Reading

- Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. New York: Springer.
- Rao, C. R. and Sinharay, S. (eds.) (2006). *Handbook of Statistics, Volume 26: Psychometrics*. Amsterdam: Elsevier.
- Searle, S. R. (1982). *Matrix Algebra Useful for Statistics (Wiley Series in Probability and Statistics)*. New York: Wiley-Interscience.
- Yanai, H. and Takeuchi, K. (1983). *Projection Matrix, Generalized Inverse Matrix and Singular Value Decomposition*. Tokyo: University of Tokyo Press.

Measure of Association

G-H Huang, National Chiao Tung University, Hsinchu, Taiwan

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Agreement – The measure of agreement intends to quantify the reproducibility of the same variable measured more than once.

Cochran–Mantel–Haenszel procedure – When the association between two discrete variables is affected by a third factor, one should look at the odds ratio of the two variables in separate strata by the third factor. The Cochran–Mantel–Haenszel procedure is used to combine the odds ratios for separate strata into an overall summary estimate.

Continuous variables – A variable is called a continuous variable if it can take on any values within a given interval. There are no gaps in the possible values of the variable.

Discrete ordinal/nominal variables – A discrete variable can take values from a discrete set of numbers. When the numbers reflect only the relative order but not specific numerical values, the discrete variable is called a discrete ordinal variable. If the numbers are used merely to identify categories and the categories have no specific ordering, the discrete variable is called a discrete nominal variable.

Kappa statistic – The Kappa statistic is used to measure the degree of nonrandom agreement between two discrete variables. Here, the nonrandom agreement means that two variables agree more often than expected by chance.

Kendall's tau – The Kendall's tau statistic is a nonparametric measure of association based on the number of concordances and discordances in paired observations. The statistic can be applied to the cases where the variables are continuous and/or discrete ordinal.

Odds ratio – The odds ratio is a basic measure of association in a 2x2 table formed from nominal variables. The odds ratio is defined as the ratio of the odds of an event occurring in one group to the odds of it occurring in another group.

Partial correlation – The partial correlation is a measure that assesses the degree of association between two variables after controlling for other variables. It can be calculated as the Pearson correlation coefficient for the residuals from linear regressions of interested variables on controlled variables.

Rank correlation – The rank correlation coefficient uses the ranks of the data, instead of the actual observed values, to compute a correlation coefficient. It is very useful when there exist extreme values in one or both variables, where the Pearson correlation coefficient will be greatly affected.

Researchers often wish to measure the strength of relationship or association between two variables. A high degree of association indicates that changes in one variable tend to be accompanied by changes in the other, and a low level of association would indicate two variables to be almost independent of each other. There are many indices that characterize the association between two variables. This article divides the indices according to the type of variable we are measuring – namely, whether variables are continuous, discrete ordinal, or discrete nominal. In the following discussion, we assume that a set of n bivariate observations $(x_1, y_1), \dots, (x_n, y_n)$ is measured, and x_i 's and y_i 's can be one of the three variable types. We are interested in assessing the association between X and Y .

Association for Continuous Variables

Consider two continuous variables; for example, the score of a math test versus the score of a science test. The Pearson product–moment correlation coefficient (Pearson, 1896; Fisher, 1915) can be used to measure the degree to which two variables are linearly related, and can be calculated as

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

where \bar{x} and \bar{y} are the sample means of x_i 's and y_i 's, respectively. The value of r lies between -1 and $+1$. If $r > 0$, then the variables are positively correlated – as x increases (decreases), so does y . If $r < 0$, then the variables are negatively correlated – as x increases (decreases), y tends to decrease (increase). The values of r close to zero mean that there is no linear relationship between the variables; possible reasons can be that (1) the two variables are independent (e.g., knowledge of the math score in no way improves the prediction of the science score), or (2) the two variables have a nonlinear

relationship (e.g., for students with math scores less than 70/100, the higher the math score, the lower the science score; however, for students with math scores greater than 70/100, the higher the math score, the higher the science score).

If we have outliers (i.e., extreme values) in one or both variables, the Pearson correlation coefficient will be greatly affected. For example, consider the situation that a vast majority of students have math scores within the range of 50/100 and 90/100, but there are a few individuals whose scores are extremely low (e.g., 5/100). In that case, the mean math score becomes very small and the value of r becomes quite small, too. There is another alternative. The Spearman rank correlation coefficient (Spearman, 1904) does not use the actual observed data, but the ranks of the data $(r_{x_1}, r_{y_1}), \dots, (r_{x_n}, r_{y_n})$ to compute a correlation coefficient. Here, r_{x_i} is the rank of x_i among x_1, \dots, x_n , and r_{y_i} is the rank of y_i among y_1, \dots, y_n . If tied values exist (e.g., some students have the same score), one can assign the same rank, the average of ranks that originally correspond to these values, to each of the equal values.

In some instances, it is important to assess the degree of association between two variables after controlling for other variables. Researchers may be interested in the relationship between SAT math and science scores after removing the contributions from age and sex. The partial correlation coefficient (Yule, 1897) is a measure that accomplishes this goal. Suppose we are interested in the association between X and Y , but wish to control for other variables Z_1, \dots, Z_p . The partial correlation coefficient is defined to be the Pearson correlation coefficient for the set of n paired derived variables $(e_{x_1}, e_{y_1}), \dots, (e_{x_n}, e_{y_n})$, where e_{x_i} is the residual for the i th individual from the linear regression of X on Z_1, \dots, Z_p , and e_{y_i} is the residual from the linear regression of Y on Z_1, \dots, Z_p . When there is only one Z variable, the partial correlation coefficient is as

$$r_{XY.Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}$$

where r_{XZ} , r_{XZ} , and r_{YZ} are the Pearson correlation coefficients between X and Y , between X and Z , and between Y and Z , respectively. The correlation between X and Y can be very high, but the partial correlation is low. Students who take the SAT test in their senior year might score higher in both math and science than those who take the test in junior year; therefore, the correlation between SAT math and science scores is high. When we remove this age effect by using the partial correlation, the association might disappear.

Association for Discrete Ordinal Variables

Let x_i 's and y_i 's take values on $1, \dots, I$ and $1, \dots, J$, respectively. Here, the numbers reflect only the relative order, but not specific numerical values. For example, we only

observe students' ranks in math and science tests, but not actual scores. Other than the Spearman rank correlation coefficient we just described, the Kendall's tau rank correlation coefficient (Kendall, 1938) can be used to evaluate the association between X and Y . The Kendall's tau is a nonparametric measure of association based on the number of concordances and discordances in paired observations. Let (x_i, y_i) and (x_j, y_j) be a pair of (bivariate) observations. If $x_j - x_i$ and $y_j - y_i$ have the same sign, we say that the pair is concordant; if they have opposite signs, the pair is discordant. Let C stand for the number of concordant pairs and D stand for the number of discordant pairs. The formula for the Kendall's tau- a is

$$\tau_a = \frac{C - D}{T_0}$$

where $T_0 = n(n-1)/2$ is the total number of pairs of observations. If $\tau_a > 0$, two sets of observations are more likely to be concordant than discordant, and there is a positive relationship between X and Y . $\tau_a < 0$ indicates a negative relationship between X and Y . If there are no tied observations, then $-1 \leq \tau_a \leq +1$. If there are ties, $|\tau_a|$ will be smaller than 1.

When variables have only a few possible values (i.e., I and J are small), the data are typically summarized as an $I \times J$ contingency table with n_{ij} representing the frequency for the i th level of the X variable and the j th level of the Y variable. In our example, this can occur when students are only categorized as excellent, good, average, or poor based on their math and science test results. It should be noted that, now, τ_a no longer lies between -1 and $+1$. Let $n_{i+} = \sum_j n_{ij}$ and $n_{+j} = \sum_i n_{ij}$. A modified version of the Kendall's tau is

$$\tau_b = \frac{C - D}{\sqrt{(T_0 - T_1)(T_0 - T_2)}}$$

where $T_1 = \sum_{i=1}^I n_{i+}(n_{i+} - 1)/2$ is the number of pairs tied on the X variable, and $T_2 = \sum_{j=1}^J n_{+j}(n_{+j} - 1)/2$ is the number of pairs tied on the Y variable. For an $I \times J$ contingency table, C and D can be calculated as

$$C = \sum_{i < i'} \sum_{j < j'} n_{ij} n_{i'j'} \quad \text{and} \quad D = \sum_{i < i'} \sum_{j > j'} n_{ij} n_{i'j'}$$

Association for Discrete Nominal Variables

Here, X and Y are nominal variables, that is, the values of X and Y are used merely to identify categories and the categories have no specific ordering. Contingency tables are usually employed to represent such data. A basic measure of association in a 2×2 table formed from nominal variables (i.e., X and Y both have two categories)

is the odds ratio. The odds ratio is defined as the ratio of the odds of an event occurring in one group to the odds of it occurring in another group. The odds of an event occurring is $p/(1-p)$, where p is the probability of the event. For example, consider the make-up contingency table shown in **Table 1**, with sex as the X variable and classroom participation as the Y variable. The odds ratio of actively participating in classroom for females compared with males is

$$a = \frac{(158/210)/(52/210)}{(176/240)/(64/240)} = \frac{158 \times 64}{176 \times 52} = 1.1$$

Note that the odds ratio can be expressed as the ratio of cross-products of the cell frequencies of the 2×2 table.

The value of the odds ratio falls in $[0, \infty)$. An odds ratio of 1 indicates that the event under study (e.g., active classroom participation) is equally likely in both groups (e.g., females vs males). An odds ratio greater than 1 indicates that the event is more likely in the first group (e.g., females), and an odds ratio less than 1 indicates that the event is less likely in the first group. The odds ratio is invariant under interchange of X and Y variables; that is, the odds ratio of being a female for active classroom participation students compared with occasional classroom participation students is the same as the odds ratio of actively participating in classroom for females compared with males. This property is particularly useful for studies with various sampling schemes. For **Table 1**, one can always obtain the same odds ratio no matter how the study sample is selected. The study can first select 210 females and 240 males and then compare the probability of actively participating in classroom between two genders. Or the study first identifies 334 active classroom participation students and 116 occasional classroom participation students and then determine what proportion of active

classroom participation students are females and what proportion of occasional classroom participation students are females. Or the study can survey 450 students and for each participant we determine his/her gender and classroom participation.

The odds ratio can be transformed to a -1 to 1 scale by converting it to Yule's Q (Yule, 1912):

$$Q = \frac{a-1}{a+1} = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}}$$

In a situation when the association between X and Y is affected by a third factor, one obvious way is to look at the 2×2 table of X versus Y in separate strata by the third factor. For example, if we are concerned that the class attendance affects the association between sex and the classroom participation, we can calculate the odds ratio for the association between sex and classroom participation in each class attendance group (regular versus irregular). To combine the odds ratios for separate strata into an overall summary estimate, we can use the Cochran–Mantel–Haenszel procedure (Cochran, 1954; Mantel and Haenszel, 1959) that provides a weighted average of the separate odds ratios. The Cochran–Mantel–Haenszel odds ratio is estimated as

$$a_{MH} = \left(\sum_k \frac{n_{11}^{(k)} n_{22}^{(k)}}{n^{(k)}} \right) / \left(\sum_k \frac{n_{12}^{(k)} n_{21}^{(k)}}{n^{(k)}} \right)$$

where k represents the k th 2×2 table (stratum). In our example, suppose the 2×2 tables for the regular class attendance group and the irregular class attendance group are in **Table 2**. Then, the Cochran–Mantel–Haenszel odds ratio can be obtained as

$$a_{MH} = \left(\frac{131 \times 4}{200} + \frac{27 \times 60}{250} \right) / \left(\frac{36 \times 29}{200} + \frac{140 \times 23}{250} \right) = 0.5$$

In some instances, we are interested in quantifying the reproducibility of the same variable measured more than once. This refers to the measure of agreement. As such, agreement is a special case of association. If agreement exists between two occasions, association also will definitely exist, but there can be strong association without strong agreement. One can compute a measure of agreement, when, for example, two raters classify a given

Table 1 Sex vs classroom participation

		Classroom participation		
		Active	Occasional	Total
Sex	Female	158	52	210
	Male	176	64	240
	Total	334	116	450

Table 2 Sex versus classroom participation for each class attendance group

		Classroom participation ^a			Classroom participation ^b		
		Active	Occasional	Total	Active	Occasional	Total
Sex	Female	131	29	160	27	23	50
	Male	36	4	40	140	60	200
	Total	167	33	200	167	83	250

^aRegular class attendance.

^bIrregular class attendance.

student's performance into categories of a discrete variable. Let the X variable represent the results of rater 1 and the Y variable the results of rater 2. The cell frequency n_{ij} is the number of subjects that are classified as the i th category by rater 1 and the j th category by rater 2. One obvious way to quantify the agreement between two raters in the 2×2 case is the observed proportion of agreement

$$p_0 = \frac{n_{11} + n_{22}}{n}$$

The observed agreement proportion is significantly affected by the fact that even if two raters use completely different criteria for calling subjects positive or negative, we would expect agreement about certain subjects solely by chance. For example, if the students are very good, the agreement may be high even if the raters do not agree with each other. One approach to correct for the problem of agreement by chance is to calculate the Kappa statistic (Cohen, 1960)

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

where p_e is the expected proportion of agreement purely by chance between two raters. p_e can be calculated under the condition assuming that two raters' results are independent. For example, **Table 3** shows the data for the agreement between two raters. Here, $p_0 = (34 + 248)/318 = 0.89$, and

$$\begin{aligned} p_e &= Pr(\text{rater1} = \text{positive and rater2} = \text{positive}) \\ &\quad + Pr(\text{rater1} = \text{negative and rater2} = \text{negative}) \\ &= \{Pr(\text{rater1} = \text{positive}) \times Pr(\text{rater2} = \text{positive}) \\ &\quad + Pr(\text{rater1} = \text{negative}) \times Pr(\text{rater2} = \text{negative})\} \\ &= \left(\frac{56}{318} \times \frac{48}{318}\right) + \left(\frac{262}{318} \times \frac{270}{318}\right) = 0.73. \end{aligned}$$

Thus, $\kappa = (0.89 - 0.73)/(1 - 0.73) = 0.59$.

The above measures for discrete nominal variables can be generalized to $I \times J$ tables, with $I > 2$ and/or $J > 2$. If one is using odds ratios to describe the associations in an $I \times J$ table, one can start with the number of values, say I and J , for X and Y variables, and then a set of $(I - 1) \times (J - 1)$ odds ratios can be calculated as

$$a_{ij} = \frac{n_{ij}n_{I\bar{j}}}{n_{i\bar{j}}n_{Ij}}$$

where $i = 1, \dots, (I - 1)$ and $j = 1, \dots, (J - 1)$. The Cochran-Mantel-Haenszel procedure can be applied to each of the $(I - 1) \times (J - 1)$ odds ratios. For measuring the reproducibility of a variable with $I > 2$ categories, the Kappa statistic can be obtained by quantifying the agreement on all I categories.

Conclusion

There are different techniques for measuring association for different types of variables. To choose a measure for

Table 3 Rater agreement

		Rater 2		
		Positive	Negative	Total
Rater 1	Positive	34	14	48
	Negative	22	248	270
	Total	56	262	318

two given variables, one can first identify the type of variable for each and follow the recommended measures provided in this article. Caution and appropriate adjustment should be made when there exist factors that can confound the association between variables of interest and when there are several ties as discrete variables involved. It is also important to notice that establishment of a strong association does not imply a causal relationship. Two variables are strongly associated when they tend to change together, whereas causality happens when one variable causes a change in another variable.

See also: Causal Inference; Univariate Linear Regression.

Bibliography

- Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics* **10**, 417–451.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20**, 37–46.
- Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* **10**, 507–521.
- Kendall, M. (1938). A new measure of rank correlation. *Biometrika* **30**, 81–89.
- Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* **22**, 719–748.
- Pearson, K. (1896). Mathematical contributions to the theory of evolution, III. Regression, heredity and panmixia. *Philosophical Transactions Royal Society of London, Series A* **187**, 253–318.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology* **15**, 72–101.
- Yule, G. U. (1897). On the theory of correlation. *Journal of the Royal Statistical Society* **60**, 812–854.
- Yule, G. U. (1912). On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society* **75**, 579–652.

Further Reading

- Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. New York: Wiley.
- Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2002). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd edn. Hillsdale, NJ: Erlbaum.
- Fliss, J. L., Levin, B., and Paik, M. C. (2003). *Statistical Methods for Rates and Proportions*, 3rd edn. New York: Wiley.
- Goodman, L. A. and Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association* **49**, 732–764.

Measures of Central Tendency

D Clark-Carter, Staffordshire University, Stoke-on-Trent, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Effect size – A measure of the magnitude of an effect in a study.

Meta-analysis – A quantitative summary of the results from a number of studies.

Statistical power – The probability of avoiding rejecting a true hypothesis.

When trying to summarize a set of data, for example, the reading ability of a class of children, it can be useful to have a measure that tries to represent the middle of the distribution of scores. Such a statistic is described as a measure of central tendency. With such a statistic, it is possible to make comparisons either between groups or over time for the same group, for example, to judge whether there are differences in children's ability to read (i.e., reading ages) when taught by different methods or to judge whether the class has improved over time or after the introduction of a new reading scheme.

The Basic Types

There are three basic measures of central tendency – the mean, the median, and the mode – but, confusingly, the term average is often used as a synonym for all three, despite their definitions being quite different and thus what they refer to being possibly very different. Each measure is described and the relative advantages and disadvantages of each are discussed.

Mean

The mean is the value most normally thought of when the term average is used. It is calculated by summing all the values in the data set, for example, the reading ages, and then dividing that sum by the number of data points which contributed to the sum, for example, the number of children in the class. Parametric statistical tests, such as the independent samples *t*-test, which are designed to look at the difference between central tendencies, for example, in reading age between two classes of children, will compare the means of the different data sets. The mean, or, more precisely, the arithmetic mean, is described below.

Formally, the mean is calculated using the following equation:

$$\text{Mean} = \frac{\sum_{i=1}^n x_i}{n}$$

where x_i is the i th data point and n is the number of data points in the set.

Median

The median is the mid-point, or its equivalent, when the set of data has been put in order of magnitude. When all the data points are different and there is an odd number of them, then the calculation of the median is particularly straightforward. Imagine you have 25 children in a class. You put their reading ages in order, from lowest to highest. The median value is the reading age which has as many below it as above it – in this case, the 13th reading age. When there is an even number of data points, then the median is the arithmetic mean between the highest number in the bottom-half of the data set and the smallest number in the top-half of the data set. Thus, the median of the set of numbers 9, 11, 12, 13 would be 11.5. Nonparametric statistics, such as the Mann–Whitney U test, which are designed to compare central tendencies, often assess whether medians are the same.

Mode

The mode is the most frequently occurring number in the data set.

Distributions and Measures of Dispersion

Before discussing the relative merits of the three measures, it is necessary to mention the distribution of the data and the need for measures to put a measure of central tendency in context. Most well-known parametric tests, such as *t*-tests or analysis of variance (ANOVA), assume that the data in the sample come from a population in which the data are normally distributed. This distribution is sometimes referred to as the bell curve because like a bell it is symmetrical with the most frequent values in the center and the least frequent values at the edges of the distribution. If the data are normally distributed, then the three measures of central tendency will have the same value. However, if the histogram is asymmetrical and has a long tail at one or the other side of the distribution, that is, the data are skewed, then the three measures of central tendency will differ from each other.

An additional feature of a distribution can be how similar the data points are to each other. It is possible to have two sets of data with exactly the same values for a measure of central tendency even though their distributions are very different. As an extreme example, one class of children might all have exactly the same reading age while another might have a wide range of reading ages but have the same mean reading age as the other class. Therefore, a measure of central tendency is not that useful a statistic on its own; it needs to be presented in the context of a measure of dispersion such as the standard deviation.

Characteristics of the Three Measures of Central Tendency

The calculation of the mean entails using the value of every data point. As a consequence, if one or more values, which are extremely high or low, are added to the data set (or they replace existing members of the set), then the value of the mean will move substantially toward such extreme values. For example, if in a class of children, a child with a particularly low reading age relative to the others was added to the class, then the mean reading age for the class would be reduced, possibly markedly. Take a sample of five children with reading ages ranging from 14 years 0 months to 15 years 0 months, and a mean reading age of 14.5 years. If a child with a reading age of 6 years 0 months joins the class and a child with a reading age of 14 years 0 months leaves the class, then the new mean drops to 12.9 years.

The calculation of the median only relies on the order of the data points. Thus, an extreme value in the set would just be treated as the next value in order and its presence merely contributes to finding the position in which the median value lies. Thus, if the lowest reading age in a class was 14 years and that child was replaced by another with a much lower reading age (say 6 years 0 months), then the median reading age for the class would not change.

The mode has a number of disadvantages. First, if all the data points have different values, then there is no unique mode. Second, if two or more values occur with the same frequency, then there may be no single mode; when there are two modes, the term bimodal is used to describe the distribution and when there are more than two modes, the distribution can be described as multimodal. A third disadvantage of the mode is that if all but two data points in the set are different, then the mode is hardly representative of the group. Finally, the most frequently occurring value might be at one extreme of the distribution and once again not represent the whole set well.

Alternative Versions of the Mean

A number of solutions have been proposed to cope with problems which have been identified with the mean.

Two of them – the geometric and harmonic means – involve a different equation to calculate the mean, while two – the trimmed mean and Winsorized mean – involve either removing data or replacing them with other values. Finally, there may be instances where we want particular cases to contribute more to the calculation of the mean. In such instances, we could find a weighted mean.

Geometric mean

The geometric mean entails finding the product of the numbers and then raising that value by the reciprocal of the number of data points which contributed to the product.

Formally, the geometric mean is calculated using the following equation:

$$\text{Geometric Mean} = \sqrt[n]{\left(\prod_{i=1}^n x_i \right)}$$

where x_i is the i th data point and n is the number of data points in the set.

In the set of data 7, 9, 11, 25, the geometric mean = $(7 \times 9 \times 11 \times 25)^{\frac{1}{4}} = 11.47$.

The geometric mean has an advantage over the arithmetic mean in that it is less affected by extreme values in a skewed distribution; in the above example, the arithmetic mean of the four numbers is 13, larger than the geometric mean. In addition, there are instances where we may want to work out a value which is intermediate between two others (to interpolate), and we have good reason to believe that the figures are increasing geometrically. An example would be if we knew the income level in 1 year and the income level for 2 years later and we had good reason to believe that income had risen by a fixed percentage each year. In this instance, the geometric mean between the two income levels would be an accurate estimate of the income in the intervening year. If in the first year, the salary was 10 000 and it went up by 5% per year, then by year 3 it would be 11 025. The geometric mean of these two figures is 10 500, the true value of the salary in the second year, whereas, if we used linear interpolation, which is equivalent to taking the arithmetic mean of the two, we would overestimate the salary in year 2 as 10 512.5.

Harmonic mean

The harmonic mean involves taking the reciprocal of each number. The inverses are summed and the sum is divided into the number of data points which contributed to the sum. (Alternatively, one could sum the inverses, divide the sum by the number of data points and then find the inverse of the result.)

Formally, the harmonic mean can be calculated using the following equation:

$$\text{Harmonic mean} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

where x_i is the i th data point and n is the number of data points in the set.

The harmonic mean is used in a number of statistical procedures. An example occurs while finding the statistical power for an independent samples t -test when the samples in the two groups are unequal. The harmonic mean of the two sample sizes is a better estimate of the sample size than the arithmetic mean when calculating the level of power which could be achieved. An example demonstrates that unequal sample sizes in such tests can be inefficient in that they yield a level of power below what the same sample would have if the sample sizes were equal (a balanced design). If the total sample for a study were 60 participants and the two groups being compared had 20 and 40 participants, respectively, then the harmonic mean would be just under 27. Thus, this unbalanced design would only achieve an equivalent level of power to that of a balanced design with a total sample size of 54.

Trimmed mean

Trimming involves placing the data points in order of size and then removing a given percentage from the top and bottom tails of the distribution. The trimmed mean is found by calculating the arithmetic mean of the remaining data. Typical figures for trimming are 5% or 10%, which leads to removing the top and bottom 5% or 10% of values. Clearly this procedure removes extreme values.

Winsorized mean

Winsorizing involves placing the data points in order of size and then rather than deleting the values in each tail, replacing them with the next value in the distribution in that tail. In the data set 1, 3, 7, 9, 11, 25, 36 we could replace the two lowest values with the value next to them and the two highest values with the value next to them, producing the set 7, 7, 7, 9, 11, 11, 11. The Winsorized mean is found by calculating the arithmetic mean of the set of data which remains after Winsorizing.

Weighted mean

The weighted mean involves multiplying each data point in a set by a value which is determined by some characteristic of whatever contributed to the data point. An example should help make that rather vague definition clearer. In meta-analysis, a researcher has a set of effect sizes from a number of studies and wishes to combine them to find an overall effect size to summarize the general trend. The larger the sample which was used in a study, the more accurate the effect size found in that study will be as an estimate of the effect size in the population. Presented with the set of effect sizes, the researcher could weight each one by the sample size for that study. In this way, larger studies would be making a

greater contribution to the mean effect size. To do this, one could multiply each effect size by the sample size for that study, sum each of these results, and then divide the sum by the sum of all the sample sizes.

Formally, the weighted mean is found from the following equation:

$$\text{Weighted mean} = \frac{\sum_{i=1}^n (w_i \times x_i)}{\sum_{i=1}^n w_i}$$

where w_i is the weighting given to the i th data point, x_i is the value of the i th data point and n is the number of data points in the set. Returning to the example of a meta-analysis, w_i could be the sample size in the i th study and x_i the effect size for that study.

Robust estimation

Measures of central tendency such as trimmed or Winsorized means are referred to as robust estimators. Some researchers (e.g., Keselman *et al.*, 2007) argue that inferential statistics based on robust estimators can be preferable, when data are not normally distributed, than either traditional parametric and nonparametric tests, or even permutation tests. They suggest that robust estimators are less likely to lead to the wrong decision over whether to reject the Null Hypothesis. For more detail on robust methods, see Wilcox (2001), Wilcox and Keselman (2003), and Maxwell and Delaney (2004).

Conclusions

It is hoped that among the things which the reader has gleaned from this article is the need to question what is entailed when a value based on a measure of central tendency is presented. If it is stated that the figure is an average, then one would want to know which one.

See also: Analysis of Variance; Measures of Dispersion, Skewness and Kurtosis.

Bibliography

- Keselman, H. J., Wilcox, R. R., Lix, L. M., Algina, J., and Fradette, K. (2007). Adaptive robust estimation and testing. *British Journal of Mathematical and Statistical Psychology* **60**(2), 267–293.
- Maxwell, S. E. and Delaney, H. D. (2004). *Designing Experiments and Analysing Data: A Model Comparison Approach*, 2nd edn. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wilcox, R. R. (2001). *Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy*. New York: Springer.
- Wilcox, R. R. and Keselman, H. J. (2003). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods* **8**(3), 254–274.

Measures of Dispersion, Skewness and Kurtosis

D Rindskopf and M Shiyko, CUNY Graduate Center, New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Overview

By far, the most common measurements in education and educational research are test scores of some type. These scores are usually considered to be (close enough to) continuous variables, so a major task is to describe the distribution of these scores, and to make inferences concerning them. Researchers use techniques that fall into two broad categories: visual (qualitative) displays and summary (quantitative) statistics. Each has benefits, so a typical analysis will involve both types of techniques.

In many cases, one will want to know whether the distribution of scores can be adequately characterized as having arisen from a particular type of probability distribution, such as the normal – which is the most well-known such distribution. If so, one can easily summarize all of the relevant information with regard to the distribution in the form of a small number of parameters (unknown constants), which can then be estimated from the data. Both visual displays and summary statistics can be used to make decisions about whether the observed data can be assumed to come from (at least approximately) such a known distributional form.

It is common to describe a distribution of scores in terms of its moments. The concept of moments originates from physics and was borrowed to describe mathematical functions known as probability distributions. Measures of dispersion, skewness, and kurtosis are derived from the central moments. The mean is the first central moment – which describes the center of a set of scores. However, to completely describe a distribution of observations, it is insufficient to know just its center. In addition, information is needed to know how the scores are distributed around the center by computing the standard deviation or variance (the second moment about the mean), data symmetry around the center as measured by skewness (the third moment), and whether data are peaked or flat (or even bimodal), indicated by kurtosis (the fourth moment).

Dispersion is the most central concept in statistics. If a set of numbers were all the same, there would be no variability to describe, explain, or relate to other measures; statistical analysis would be irrelevant (unless the results were due to the unlikely event that the sampling units are all the same). Even the term variable is meant to signify that not all observations will be the same; if they are all the same, it is not a variable but a constant. If we had a measure of mathematical reasoning on which we

assigned everyone a score of 82 (and if there were no reason to believe that anyone else existed to whom we would assign a different score), then the measure is not worthy of statistical analysis. We measure people (or other units) in order to discover, describe, and explain variability: How variable are people on some measure, and why are not they all the same?

However, dispersion has related interpretations that are more interesting. Groups with small dispersion are more homogeneous than groups with large dispersion. Thus, if we predicted an (as yet unobserved) person's score using the average of the people we have observed, we expect to be closer to the right answer when the dispersion is small (the group is homogeneous) than when it is large. Thus, dispersion is also a measure of accuracy of prediction (technically, the mean minimizes the sum of squared errors of prediction and is optimal according to some standard criteria).

A related way to think with regard to dispersion is in the two inverse concepts of entropy and information. Entropy is a technical term for a measure of disorder; more variable results indicate greater disorder among a set of numbers. The opposite concept is information: smaller variability indicates that there is more information concerning the likely value of a variable. These concepts can be applied to more than just continuous measures, so they generalize the concept of dispersion to include categorical variables. For example, if people are spread equally between two categories (e.g., male and female), then prediction of the sex of a new person has maximum uncertainty (entropy) and there is minimal information. On the other hand, if 90% of people enjoy eating pasta, and only 10% do not, then information is greater (and uncertainty is less) with regard to whether a person will like pasta. (Note the close relationship to predictability for a new person.) For the remainder of this article, the emphasis is on measures for continuous variables, which is the typical application in applied statistics.

Skewness and kurtosis are not commonly reported or interpreted by educational researchers. They are informative when describing the shape of a distribution, but this information is mostly of interest when comparing a particular set of scores to the normal distribution. These measures are considered prior to performing statistical analyses relying on the data normality assumption. They can be considered diagnostic measures, which identify problematic distribution shapes that require a further intervention – such as a transformation. Although there

are direct statistical tests of normality, they do not indicate how to proceed when non-normality is detected; skewness and kurtosis (as well as graphical methods) often help correct as well as detect non-normality.

In the following sections, a description of commonly used measures of dispersion is presented, focusing on their advantages and limitations. Further, a brief overview of the normal distribution and how skewness and kurtosis statistics are useful descriptors of the shape of educational data that depart from the normal curve is provided. Throughout the article, there are references to two data sets. The US News and World Report College data set (from their 1995 report on American colleges and universities) summarizes information from 1303 institutions on student standardized tests' performance, college tuition and other expenses, acceptance and graduation rates, and faculty characteristics. The Irish data set provides scores of verbal reasoning test as well as parental education and type of school on a sample of 500 Irish schoolchildren who were 11 years old in 1967.

Dispersion

Range and Interquartile Range

Most commonly, the dispersion of a variable is summarized quantitatively in statistics such as range, interquartile range (IQR), variance, and standard deviation. The range is the simplest way to describe a set of test scores: Subtract the smallest (X_{\min}) from the largest (X_{\max}) value. For example, in the following distribution of math scores ordered from the smallest to the highest (71, 75, 79, 86, 90, 94, 96, and 98), the range would be $X_{\max} - X_{\min} = 98 - 71 = 27$. In comparison, in a different distribution of scores (71, 88, 89, 90, 91, 92, 93, and 98), the range would still be the same (27) – as the minimum and the maximum do not change – although scores in between are much closer to each other. While the range of a particular distribution may be useful for learning about the distance between the smallest and the largest values, the drawbacks are evident. Relying on two extreme observations, the range does not take into account the rest of the scores (see **Figure 1**). It is also sensitive to atypical values (outliers) and to the total sample size, increasing with increasing number of observations. Worst of all, the range in a sample of data does not estimate any property of the population, because for most distributions

(including the normal), the population range is infinite. These characteristics make the range nearly useless for inferential purposes.

One way to improve the range statistic is to omit a certain percentage of the tails in a distribution and calculate the difference on the remaining data. It is a common practice to omit 25% of the data at both ends of a distribution and to compute the range for the middle 50% (from 25% to 75%) of the remaining scores. This dispersion-summary statistic is known as the IQR, because it shows how far the first (Q_1) and the third (Q_3) quartiles are from each other (**Figure 1**): $IQR = Q_3 - Q_1$. Q_1 is the median of the lower half of observations, or the 25th percentile, and Q_3 is the median for the top half, or the 75th percentile. For the first set of math scores, the IQR can be computed in the following way:

$$71, \underbrace{75, 79, 86}_{Q_1 = 77}, | \underbrace{90, 94, 96, 98}_{Q_3 = 95} \quad IQR = 95 - 77 = 18$$

Half of the IQR is referred to as the semi-interquartile range or Q . Both Q and IQR have a simple relationship to the standard deviation (σ , described later in this article) in normally distributed data: $Q = 0.6754 \sigma$ or $\sigma \approx 1.5Q$.

Both IQR and Q do a better job than the range in describing the dispersion of data. The IQR of 2.0 from the second distribution of math scores (from a 100-point scale) indicates that there is not much variation in math performance for the middle 50% of students. In comparison, an IQR of 18 describes much more heterogeneous student performance. The drawback of various range statistics, however, is that they rely on a subset of data points to describe variability rather than taking into account all scores.

Standard Deviation and Variance

One approach to using all of the data points would be to compute the average distance between each pair of data points. The further away the scores are from each other, the greater the total dispersion. This method never became popular, partly because of the great amount of computation involved with many data points, and partly because it did not describe a parameter in the population. Other approaches involved finding a center of the set of points (usually the mean or the median), and finding how

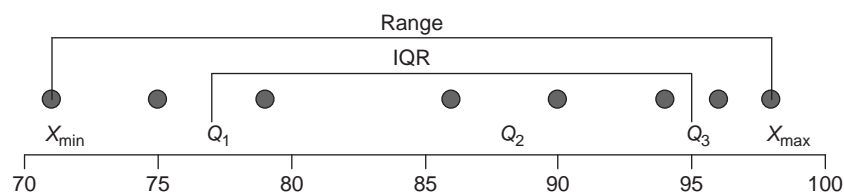


Figure 1 Range and interquartile range for the distribution of math scores (71, 75, 79, 86, 90, 94, 96, and 98).

far each of the data points was from the center. Intuitively – if the dispersion was small – most points would be close to the center, and if the dispersion was large, many points would be far from the center. Using the sample mean as the center, for example, one might calculate what are called deviation scores: $x_i = X_i - \bar{X}$. One might think that deriving just the average of these deviation scores could get a measure of dispersion; however, their sum will always equal zero, and their average would also equal zero. There are two obvious ways to overcome this problem: (1) by taking the absolute value of the deviation scores or (2) by squaring them. The second method is more commonly used because of its better computational and inferential properties.

The formula for computing a sample variance is $s^2 = \frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{(n-1)} \sum_{i=1}^n X_i^2$. In this formula, n refers to the sample size. The division by $n - 1$ instead of n is done so that the resulting quantity will be an unbiased estimate of the population variance; $n - 1$ is called the degrees of freedom. The square root of the sample variance – called the sample standard deviation s – is calculated to convert squared deviations back to the original measurement metric.

For the sample of math scores (71, 75, 79, 86, 90, 94, 96, and 98), with $n = 8$ and $\bar{X} = 86.125$, the standard deviation is computed as follows:

$$s = \sqrt{\frac{(71 - 86.125)^2 + (75 - 86.125)^2 + \cdots + (98 - 86.125)^2}{(8 - 1)}} \\ = \sqrt{\frac{718.875}{7}} \approx 10.13$$

The interpretation of the standard deviation's magnitude is relative to the measurement scale as well as the shape of a distribution. A standard deviation of ten might be small or large, depending on how big a 1-unit step is on that measurement scale, and for a particular type of study. A sample of students that has a standard deviation of ten points on an intelligence quotient (IQ) scale is more variable than a sample with a standard deviation of ten points on the Scholastic Aptitude Test (SAT) commonly used in the United States. Note that a sample standard deviation is usually judged to be large or small compared to some hypothetical population value. In the examples just mentioned, IQ scales commonly have population standard deviations of 15, 16, or 20, while SAT scores have a population standard deviation of 100 points. Some authors attempt to compare the standard deviation to the mean, but this only works in special instances where scores have a meaningful zero point.

Knowing the standard deviation of approximately normally distributed data is very informative because a normal distribution is completely determined by its mean

and standard deviation. For instance, for the scores on the Drumcondra Verbal Reasoning Test (DVRT) given to Irish schoolchildren, which approximately follows the normal curve with a mean of 100 and a standard deviation of 15.5 (Figure 2), it can be deduced that about 68% of the DVRT scores fall within the interval $\bar{X} \pm s$, or between 84.5 and 115.5. Further, about 95% of the scores fall within the interval $\bar{X} \pm 2s$, or between 69 and 131. Knowing that a student scored 110 on the DVRT, it can be determined that the score is about $(110 - 100)/15.5 \approx 0.65$ standard deviations above the mean, corresponding to the 74th percentile, so about three-quarters of the students in the sample would score below that student.

Median Absolute Deviation

Because the standard deviation and variance rely on the mean value to compute the average distance of scores away from the center and since the squared differences are used, these measures are sensitive to extreme values – or outliers. An alternative way to summarize variance of data is to use the statistic known as the median absolute deviation (MAD). MAD is a resistant measure of variability as it relies on the median as the estimate of the center of the distribution, and on the absolute difference rather than the squared difference. Initially, one would need to find an absolute value of individual score differences from the median $|X_i - \text{median}|$, order the obtained absolute distances from the lowest to highest, and find the median of the ordered distances. The complete formula for the MAD is $MAD = \text{median}(|X_i - \text{median}|)$. Because the MAD is the median deviation of scores from the overall median, not all observations are equally weighted in this measure of dispersion. The clear advantage of MAD is the avoidance of influence by outliers. However,

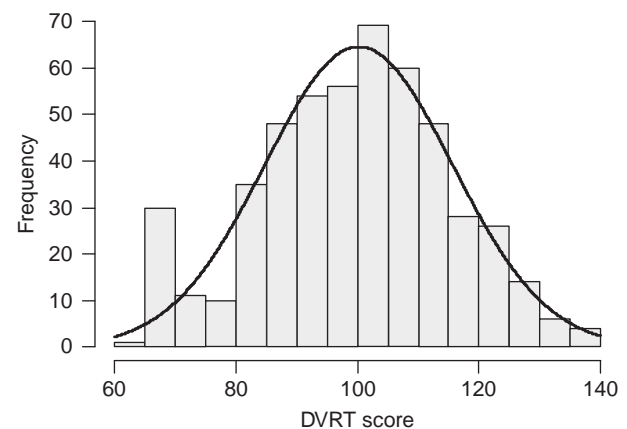


Figure 2 Histogram of the verbal reasoning test scores in a sample of 500 Irish schoolchildren (mean ≈ 100 , $s \approx 15.5$), skewness = -0.08 (SE = 0.11), kurtosis = -0.40 (SE = 0.22) with the normal curve superimposed.

it has its own problems: if the distribution is actually normal, there is a loss of efficiency in that it does not make as much use of the information as is available in the data. Therefore, just as the mean is the measure of choice for the center of a distribution that is (at least nearly) normal, so is the standard deviation the measure of choice for the dispersion in that case. Furthermore, for a normal distribution, the mean and standard deviation completely describe the distribution, so that a data set whose underlying distribution is normal can be described in two numbers instead of requiring the complete set of raw data.

For most educational measures, standard deviation or variance are advantageous and sufficient descriptors of data variability. They are easily computed in any statistical package for any number of variables and sample size.

Shape of a Distribution

Visual techniques are especially valuable for learning about a distribution's shape. Generally, one would be interested in whether a distribution of scores resembles the normal curve, which occupies the central niche in statistical theory. Standardized test scores, measures of intelligence (IQ), and depression and anxiety scores (among other constructs educators rely on in their research) are, often, normally distributed because the central limit theorem, which, roughly, states that the sum of independent random measures is approximately normally distributed, especially as the number of items used in constructing a scale increases. Intelligence, aptitude, or psychological measures are commonly represented as a sum of individual items that may be binary (dichotomous) or ordered. While individual items of a test are not normally distributed – based on the central limit theorem – their sum is often well approximated by the normal curve. Due to well-studied mathematical properties of the normal distribution, data normality eases the inferential process and allows applications of various statistical methods relying on the normality assumption (e.g., *t*-tests, analysis of variance (ANOVA)).

Visually, normally distributed data resemble a smooth bell curve rising in the middle and gradually falling at the tails (**Figure 3**). It has a single center (unimodal) – where the mean, median, and mode are located – and it is symmetric. The tails of the distribution are asymptotic, never reaching the value of zero; theoretically, any real number is a possible outcome. Approximately 68% of data are within the interval $\mu \pm \sigma$, and about 95% of data are within the interval $\mu \pm 2\sigma$ (more precisely, 95% are within the interval $\mu \pm 1.96\sigma$, but, in many situations, applied statisticians round 1.96 to 2 when such approximations are acceptable). As mentioned previously, knowing the mean and variance (or standard deviation) is sufficient to completely describe a normal distribution. The sample mean \bar{X} and variance s^2 are unbiased estimators of the population mean μ and variance σ^2 .

The normal curve is theoretically derived, and distributions of real educational data are not as perfect as in their characteristics. The approximation, however, is close enough to assume that data are normally distributed if they possess certain distributional properties. There are several ways to evaluate whether a particular distribution resembles a normal curve. Graphically, data are often summarized in a histogram with an overlying normal curve (as in **Figure 2**) to allow comparisons between the two. The number of modes is the first property to be noticed. In the example of the DVRT scores, there is only a single peak around the value of 100. In comparison, the distribution of the in-state tuition variable (**Figure 4**) is bimodal; there are two peaks occurring around the values of \$2000 and \$10 000. Based on the histogram, there appear to be two clusters of colleges – one with low, and the other with higher, tuition rates. Upon closer examination, the lower cluster corresponds to public colleges and the higher cluster represents private institutions. In that case, a normal distribution is not a good approximation of such data, and it may be advantageous to analyze each college type separately.

Next, the examination proceeds to two properties of data – skewness and kurtosis – that can be used for two

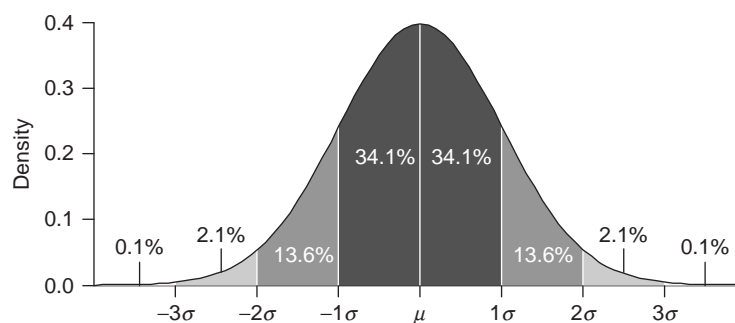


Figure 3 The normal distribution with corresponding areas under the curve.

purposes. Each is a valid descriptive statistic in its own right, helping to define the shape of the distribution of data. However, most often, these statistics are used for another purpose: to determine whether or not a data set is (close enough to) normally distributed so that common methods of analysis can be used.

Skewness

Symmetry is another property of normally distributed data that should be examined. From a graph, one can evaluate whether approximately the same shape and the same data frequency are observed on both sides of a distribution. In the DVRT scores example (Figure 2), the histogram of the scores is symmetric, while the in-state college-tuition histogram (Figure 4) is not. There is a mathematical measure of symmetry known as skewness. The theoretical normal distribution is perfectly symmetrical, and its corresponding skewness statistic is zero. When a distribution has an elongated right tail, it is positively skewed (i.e., it has a positive skewness statistic; e.g., percentage of alumni who donate to their colleges in Figure 5). Conversely, when the left tail is long, the

distribution is negatively skewed (e.g., percentage of applicants accepted by 1291 colleges in Figure 6).

The skewness is computed as follows:

$$skewness = \frac{\sum_1^n (X_i - \bar{X})^3}{(n-1)s^3}$$

Conceptually, the skewness statistic measures cubed standardized differences between individual scores and the mean, keeping the sign of the difference (unlike the case with the standard deviation). If there is a particular number of observations a certain distance below the mean and an equal number is the same distance above the mean, their contributions to skewness will be canceled out, resulting in the skewness of zero. An imbalance of scores around the mean will lead to either large positive or large negative skewness. To decide whether a particular distribution of scores is strongly (significantly) skewed in comparison to the normal, one can perform a hypothesis test. The null hypothesis states that the distribution under consideration has a skewness of zero. To test the hypothesis, the obtained value of the skew statistic is divided by its standard error:

$$SE(skewness) = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}$$

and the obtained ratio is compared to the critical value of 1.96. Note that for large sample size n , the standard error is approximately the square root of $6/n$. If an absolute value of the ratio is smaller than 1.96, the null hypothesis is retained. If it is larger than 1.96, it can be concluded that the skewness is significantly different from zero and the data may require a transformation. Because the standard error largely depends on the sample size, it is harder to detect skewness in small samples, while, too often, large samples will produce statistically significant results even when the departure from normality is small. Therefore, the skewness statistic should be supported by graphical evidence of nonsymmetry.

Although a symmetric distribution has a skewness of zero, the converse is not true – that is, a distribution can

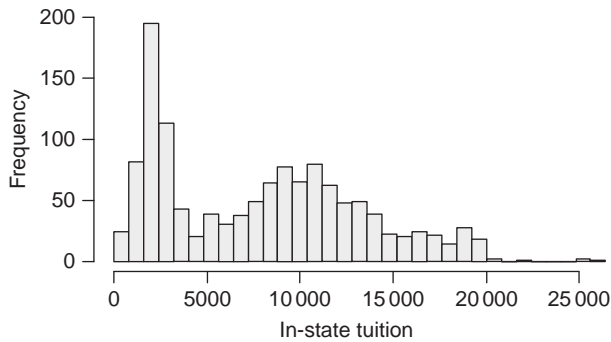


Figure 4 Histogram of the in-state tuition variable measured across 1272 colleges, skewness = 0.43 (SE = 0.07), kurtosis = -0.75 (SE = 0.14).

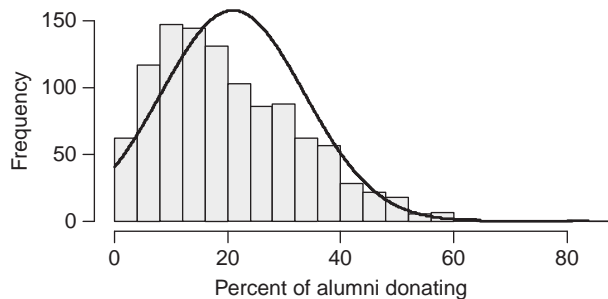


Figure 5 Histogram of the percentage of alumni who donate to their colleges for 1088 institutions, skewness = 0.77 (SE = 0.07), kurtosis = 0.32 (SE = 0.15) with the superimposed normal curve that has the same mean and standard deviation.

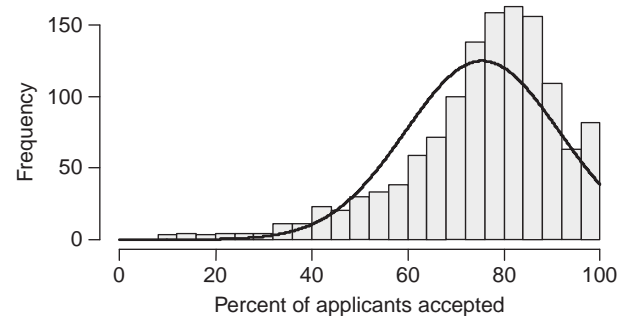


Figure 6 Histogram of the percentage of accepted applicants variable from 1291 colleges, skewness = -1.11 (SE = 0.07), kurtosis = 1.61 (SE = 0.14) with the superimposed normal curve with the same mean and standard deviation.

have a skewness of zero and not be symmetric. This would not be common, but one should remember that checking skewness is a shortcut and not a perfect test of symmetry. A symmetric distribution will have zero moments around the mean for all odd powers, not just the third (i.e., skewness).

The most common transformations of variables with a skewed distribution involve moving the stretched tail closer to the center by some mathematical manipulations, such as taking the square root, inverse, logarithm, or taking scores to a power. While some skewed distributions can be normalized using such mathematical transformations, others cannot or should not be. For instance, nothing can be done to multimodal or J-shaped distributions, or variables with pronounced floor or ceiling effects. There is a breadth of statistical methods available for working with such data. Such solutions are preferable to using methods relying on the normality assumption when it cannot be met.

Kurtosis

The last property of data that is to be discussed to assess whether they come from a normal distribution is kurtosis, which evaluates (approximately) how peaked or how flat a distribution is. The fourth moment of the normal curve has a value of 3, and the general formula is

$$\text{kurtosis} = \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{(n-1)s^4} - 3$$

A kurtosis of zero is obtained for scores from a normal distribution (since we subtract the value of 3 in the kurtosis formula). A negative kurtosis corresponds to a platykurtic, or wide, distribution with more extreme scores than expected in the normal. Such a distribution would be wider and thicker in the tails. The t distribution, for example, has negative kurtosis. A positive kurtosis represents a leptokurtic – or narrow – distribution, with fewer scores in the tails than in the normal (e.g., to some degree in the distribution of the percentage of accepted applicants in **Figure 5**; to a greater degree in the percentage of applicants who donate to their colleges in **Figure 6**).

Even more extreme values of kurtosis are found in multimodal distributions and, in particular, in bimodal distributions (**Figure 4**). Bimodal distributions have a very large proportion of their observations a large distance from the middle of the distribution, even more so than the flat distributions often used to illustrate high values of kurtosis, and have more negative values of kurtosis than other distributions with heavy tails such as the t .

While positive or negative values of the kurtosis statistic may indicate the direction of nonnormality, they have

to be formally tested to see if the difference between the observed kurtosis and that expected in the normal curve is statistically significant. Similar to the test of skewness, one would divide the kurtosis by its standard error:

$$SE(\text{kurtosis}) = 2 \times SE(\text{skewness}) \sqrt{\frac{n^2 - 1}{(n-3)(n+5)}}.$$

Note that for large sample size n , the quantity under the radical is near 1, and the standard error of the kurtosis is approximately twice the size of the standard error of the skewness. An absolute value of the ratio that is larger than 1.96 would indicate a shape of a distribution statistically different from normal. Again – as with any test of a null hypothesis – one must consider sample size when interpreting the result: a large sample size might detect very small departures from normality.

There are no simple transformations available for normalizing a distribution with extreme values of kurtosis, but this is not of much concern to most researchers. The skewness properties of a distribution are much more important for selecting appropriate statistical tests than its kurtosis. Due to the central limit theorem, repeated sampling from a highly kurtotic distribution (e.g. uniform or bimodal) will approximate the normal with sample sizes as low as five or ten. However, to achieve the same results with a skewed distribution, much larger samples are needed.

While the third and the fourth central moments (skewness and kurtosis) are informative in comparing distributional properties to those of the normal curve, there are other approaches to evaluate normality. The most commonly used graphical tests summarize data in forms of quantile–quantile (Q–Q) plots or the normal probability plots. There are also a number of formal statistical tests that measure how closely a given data set resembles the normal (e.g., Kolmogorov–Smirnov and Shapiro–Wilk tests, among others), but these do not provide information about why normality is violated or how to correct it.

Conclusion

Dispersion is the most basic concept in statistics; in fact, the field of statistics is often defined as the study of variation in data. A normal distribution, the most commonly encountered distribution in many fields of statistics, can be completely described by its location (mean) and dispersion (standard deviation or variance). Higher-order moments about the mean – such as skewness and kurtosis – can be used in their own right to help describe a distribution, or as the means to an end: detecting departures from normality.

See also: Analysis of Variance; Generalized Linear Mixed Models; Hypothesis Testing and Confidence Intervals; Measures of Central Tendency; Nonparametric Statistical Methods; The Normal Distribution and its Applications.

Further Reading

- Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco, CA: Holden-Day.
- Cramer, D. (1997). *Basic Statistics for Social Research: Step-by-Step Calculations and Computer Techniques Using Minitab*. London: Routledge.
- Glass, G. V. and Hopkins, K. D. (1995). *Statistical Methods in Education and Psychology*, 3rd edn. Needham Heights, MA: Allyn and Bacon.
- Greaney, V. and Kelleghan, T. (1984). Irish educational transitions data. <http://lib.stat.cmu.edu/datasets/irish.ed> (accessed September 2009).

- Guilford, J. P. (1977). *Fundamental Statistics in Psychology and Education*, 6th edn. New York: McGraw-Hill.
- Lock, R. (2009). *U.S. News & World Report College Data*. http://www.amstat.org/publications/jse/jse_data_archive.html (accessed September 2009).
- Mood, A. M., Graybill, F. A., and Boes, D. C. (1986). *Introduction to the Theory of Statistics*. New York: McGraw-Hill.
- Stuart, A. and Ord, J. K. (1994). *Kendall's Advanced Theory of Statistics, Vol. 1: Distribution Theory*, 6th edn. London: Edward Arnold.

Relevant Websites

- <http://lib.stat.cmu.edu> – Irish Educational Transitions Data.
- <http://www.amstat.org> – Journal of Statistics Education: Data Archive.

Meta Analysis

J Sánchez-Meca and F Marín-Martínez, University of Murcia, Murcia, Spain

© 2010 Elsevier Ltd. All rights reserved.

Glossary

***d* family** – This is a set of effect-size indices which calculates the difference between the means of two groups (e.g., experimental vs. control, pre-test vs. post-test) standardized or not by an estimate of the within-study standard deviation. These effect indices are specially indicated for primary studies where subjects are assigned to different groups or conditions.

Effect size – This quantifies the extent to which the phenomenon investigated is present in the study results, regardless of the sample size and the result of the statistical tests.

Fixed-effects model – In meta-analysis, a fixed-effects model is a statistical model in which it is assumed that the effect sizes calculated from a set of primary studies estimate the same population effect size. The only error source is that produced by sampling error.

Meta-analysis – This is a research methodology that aims to quantitatively integrate the results of a set of primary studies about a given topic in order to determine the state of the art on that topic.

Mixed-effects model – When a meta-analysis aims to search for moderator variables that can explain the variability in the effect-size estimates, a mixed-effects model is applied where the studies compose a random variable and the moderator variables are fixed variables.

Moderator variable – This is a differential characteristic of the primary studies that is coded in a meta-analysis with the purpose of examining its possible relationship with the study results. The moderator variables can be substantive, methodological, or extrinsic.

Publication bias – This occurs when, in a given research field, the publication of research results depends on its direction. Meta-analyses that have included only published studies may suffer a bias in their effect estimates.

***r* family** – This is a set of effect-size indices that groups all of the correlation coefficients between two variables (e.g., Pearson correlation coefficient, biserial correlation, and point-biserial correlation). These effect indices are specially indicated for correlational studies.

Random-effects model – In meta-analysis, a random-effects model is a statistical model in which the effect size obtained from each primary study estimates a different population effect size and, as a consequence, the effect estimates have two error sources: sampling error and variability among the population effect sizes.

Standardized mean difference – This is an effect-size index from the *d* family that reflects the magnitude of the differences between the means of two groups (e.g., experimental vs. control) divided by an estimate of the within-study standard deviation.

Introduction

From the second-half of the last century, research in education has grown exponentially. From this explosion of evidence has arisen a need to develop systematic and objective methods for accumulating the scientific knowledge obtained from the primary studies. With this aim, meta-analysis has emerged as a methodology that is able to quantitatively combine the results of a set of primary investigations on a given topic in order to figure out the state of the art on that topic. Unlike traditional reviews, which are qualitative or narrative and characterized by subjectivity, meta-analysis considers the review of the past research to be a scientific enterprise that, like primary research, should be guided by the principles of objectivity, systematization, and replicability. The main characteristic of meta-analysis is the use of statistical methods to quantitatively integrate the results of the studies.

Although its origins go back to the decade of the 1930s, it was not until the end of the 1970s that G. V. Glass (1976) coined the term meta-analysis to refer to this research methodology. Meta-analysis began to gain popularity in psychology with the pioneer meta-analyses of M. L. Smith and G. V. Glass on the effectiveness of psychotherapy, R. Rosenthal and D. B. Rubin on the effects of interpersonal expectancy on the research results, and with the meta-analytic validity generalization approach of F. L. Schmidt and J. E. Hunter. In education, the pioneer meta-analysis was that carried out by M. L. Smith and G. V. Glass on the effects of class size on the attitudes and academic performance of students. Since then, meta-analysis has been

applied in education in many different research areas such as the effects of instructional programs on student performance, gender differences in academic performance, predictors of academic performance, or the validity of screening and diagnostic instruments in detecting psychoeducational problems.

The phases in developing a meta-analysis, an overview of the effect-size indices, the statistical models most usually applied in meta-analysis, an example to illustrate the meta-analytic calculations, and some final remarks are presented in this article. Throughout the article, the meta-analysis of Ginns (2006) on the instructional effects of spatial and temporal contiguity of learning materials is used as an example.

Phases in a Meta-Analysis

To carry out a meta-analysis, the researcher must follow several steps: (1) formulating the problem, (2) searching for the literature, (3) coding the studies, (4) statistical analysis and interpretation, and (5) publication of the meta-analysis.

Given below are the details of the steps to be followed:

1. *Formulating the problem.* As in any primary research, the first step in a meta-analysis consists of defining its purpose, which generally will be to examine the relationship between two or more (psychological and/or educational) constructs. In this phase, the researcher must define the constructs implied, both theoretically and empirically, as well as review the existing theoretical models and formulate the concrete objectives of the review.

In Ginns' (2006) meta-analysis, the purpose was to review the experimental studies about the effects of instructional learning by manipulating the spatial or temporal contiguity of disparate but related elements of information. In this meta-analysis, Ginns (2006) tried to determine the extent to which the effects of instructional learning could be generalized across different students, learning materials, and testing contexts.

2. *Searching for the literature.* Once the objectives are formulated, the next step involves defining the selection criteria that the empirical studies must fulfill and carrying out as complete a literature search as possible. The selection criteria will depend on the purpose of the meta-analysis, but there are several criteria that should be present in any meta-analysis, such as specifying: (1) the time period of the studies, (2) the design type in the empirical studies, and (3) the language in which the study was written. In order to search for the studies that fulfill the selection criteria, several searching strategies should be used, combining both formal

and informal search procedures. Formal procedures consist of consulting electronic bibliographic databases (e.g., ERIC, PsycINFO), relevant journals, bibliographic reference volumes, and references in studies. Informal sources enable us to find fugitive literature, that is, papers that have not been published (e.g., dissertations, technical reports, papers presented at congresses) or papers published in journals or books that cannot be found through formal sources.

In Ginns' (2006) example, in order to be included in the meta-analysis the studies had to apply a between-group design with random assignment of the subjects to the groups and to report statistical data about the effects of manipulating spatial or temporal contiguity of instructional materials. The search strategy included consulting the ERIC and PsycINFO databases, Science Citation Index, and the references of the papers located.

3. *Coding of studies.* The main purpose of a meta-analysis is to explain the variability found in the study results on a given topic by examining the influence of differential characteristics among them. To accomplish this objective, the studies are subjected to a coding process in which relevant moderator variables of the results are identified. Moderator variables are study characteristics that can influence the outcome. Although the moderator variables to be coded depend on the purpose of the meta-analysis, it is usual to classify them into three main clusters: methodological, substantive, and extrinsic characteristics.

Methodological characteristics refer to those related with the methodology of the study, such as the design type or the amount of attrition in the groups. Substantive characteristics are those directly related with the research topic investigated in the meta-analysis, as well as with the sociodemographic characteristics of the sample. In Ginns' (2006) example, some of the substantive characteristics coded were the type of split-attention effect (spatial or temporal contiguity), the field of study (science, or engineering/technical), and the educational level of the sample. Finally, extrinsic variables are those characteristics that should not be related with the study results because they have nothing to do with the research enterprise. Examples of this type of variables are the publication year of the study, the publication source (published or unpublished), and characteristics of the researchers (e.g., gender, affiliation). In addition to the moderator variables, the main result of each study is summarized by calculating an effect-size estimate.

4. *Statistical analysis and interpretation.* In a meta-analysis a mean effect is calculated from the set of effect sizes, as well as a confidence interval and an assessment of the extent of heterogeneity exhibited by them around the mean effect. If there is more heterogeneity than sampling error can explain, then additional analyses

are done to search for moderator variables of the effect estimates.

5. *Publication.* The final step in a meta-analysis is its publication. The sections that a meta-analytic report should include are basically the same as those in a primary study (introduction, method, results, and discussion), but in the method section, the epigraphs are not the same as those in primary research: literature search, coding of studies, effect-size index, and statistical analysis. A meta-analysis that is to be published must include all the information needed for its potential replication by another researcher.

Effect-Size Indices

An essential requisite to carry out a meta-analysis is to obtain a quantitative index that summarizes the results of each study. Moreover, the quantitative indices obtained from the studies must be in the same metric for their quantitative integration to be possible. The best strategy for accomplishing this requisite is to calculate an effect-size index, which represents the extent to which the phenomenon of interest is manifested in the study results. There are many different effect-sizes indices, the choice depending on the design type of the studies and on the nature of the variables involved in the studies (continuous, dichotomous, etc.). In education, the most frequently applied effect sizes in meta-analysis are those grouped into two families: the *d* family, for experimental and quasi-experimental designs, and the *r* family, for correlational studies (Hedges and Olkin, 1985).

The *d* Family

When the study design involves assigning subjects to different groups (e.g., experimental vs. control) and the dependent variable is continuous, the most appropriate family of effect sizes is the *d* family, which includes a set of effect sizes defined as the difference between two means divided by a within-group standard deviation. When the study comprises a two-group design, the effect size most usually applied from the *d* family is the standardized mean difference, which is calculated by

$$d = c(m) \frac{\bar{y}_E - \bar{y}_C}{S} \quad [1]$$

\bar{y}_E and \bar{y}_C being the estimated means of the experimental and control groups, respectively, S being the estimated pooled within-study standard deviation obtained by

$$S = \sqrt{\frac{(n_E - 1)S_E^2 + (n_C - 1)S_C^2}{n_E + n_C - 2}} \quad [2]$$

with n_E and n_C being the sample sizes of both groups, and S_E^2 and S_C^2 the respective estimated variances. When the study design includes a control group, Glass *et al.* (1981)

proposed dividing the mean difference by the standard deviation of the control group, S_C , instead of S , because sometimes applying a treatment can alter the variability in the scores of the dependent variable. But provided homoscedasticity is met, S is a more efficient estimate of population standard deviation. The $c(m)$ factor is needed to correct a positive bias of the *d* index for small sample sizes and is obtained by

$$c(m) = 1 - \frac{3}{4(n_E + n_C) - 9} \quad [3]$$

The *d* index is an estimate of the population standardized mean difference, $\delta = (\mu_E - \mu_C)/\sigma$, with μ_E and μ_C being the population means of the experimental and control groups, and σ being the common population standard deviation. As an estimate of δ , the *d* index is approximately normally distributed with mean δ and sampling variance σ_d^2 , which is estimated by

$$\hat{\sigma}_d^2 = \frac{n_E + n_C}{n_E n_C} + \frac{d^2}{2(n_E + n_C)} \quad [4]$$

Dividing the difference in mean by the standard deviation of the groups in [1] makes it possible to homogenize the metric of studies that have used different dependent variables to measure the outcome.

Several adaptations of the *d* index have been devised for studies that include pre-test and post-test measures for only one group or for a two-group design (Morris, 2008). Other effect-size indices that can be applied when the design is composed of two groups and the dependent variable is dichotomous, or has been dichotomized, are the difference between the success (or failure) proportions for the experimental and the control groups (or risk difference), the rate between both proportions (or risk ratio), and the odds ratio (Deeks and Altman, 2001).

The *r* Family

When the study results are obtained by applying a correlational design, that is, studies where the researcher does not manipulate independent variables, the most appropriate effect size consists of calculating a correlation coefficient. Depending on the type of variables involved in the relationship (continuous, dichotomous, dichotomized, ordinal variables, etc.), different correlation coefficients are used as the effect size: Pearson correlation coefficient, point-biserial correlation, Spearman's rank order correlation, etc. All of these correlation coefficients make up the *r* family of effect-size indices.

As the correlation coefficients do not always follow a normal distribution, their transformation into the Fisher's *Z* has been proposed to normalize the sampling distribution and to stabilize the variance. Thus, it is very usual in meta-analysis to transform the correlation coefficients, *r*,

into Fisher's Z by means of

$$Z_r = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) \quad [5]$$

where \ln is the natural logarithm. However, there is no clear consensus on this topic, and some authors advise against transforming correlation coefficients into Fisher's Z (e.g., Hunter and Schmidt, 2004). The sampling variance of Fisher's Z is a function of the sample size, N :

$$\sigma_{Z_r}^2 = \frac{1}{N-3} \quad [6]$$

Sometimes, a meta-analysis integrates experimental and correlational studies, so that the meta-analyst has been able to obtain a d index from the experimental studies and an r index from the correlational ones. In order to put all of the effect sizes into the same metric, formulas have been devised to transform them. Thus, to transform a d index into an r index, or vice versa, we can apply the equation

$$r = \frac{d}{\sqrt{d^2 + 4}} \quad [7]$$

There are also formulas for transforming an odds ratio into a d index (Sánchez-Meca *et al.*, 2003). Thus, most of the effect-size indices usually applied in meta-analysis can be transformed into one another.

Statistical Procedures in Meta-Analysis

Once we have coded a set of moderator variables in the studies and we have calculated an effect-size estimate from each of them, the statistical analyses carried out in a meta-analysis aim to answer several questions: (1) What is the average effect magnitude throughout the studies? (2) Is the average effect size statistically significant? (3) Is there heterogeneity among the effect-size estimates? and (4) If the effect estimates are not homogeneous, which of the moderator variables can explain the variability?

There is currently a consensus on the convenience of applying weighting procedures in meta-analysis in order to give more weight to the effect sizes obtained from the studies with a lesser sampling variance, that is, with a larger sample size. Depending on the assumptions made by the meta-analyst, the weighting procedures can be approached from different statistical models, basically, fixed-, random-, and mixed-effects models.

Calculating an Average Effect Size

The statistical analyses in a meta-analysis are guided by a statistical model that must be previously assumed. The main task of the statistical model is to establish the properties of

the effect-size population from which the individual effect-size estimates have been selected. To accomplish the first purpose in a meta-analysis, that is, to calculate an average effect size, two statistical models can be assumed: the fixed- and the random-effects models.

Suppose there are k independent empirical studies about a given topic and T_i is the effect-size estimate obtained in the i th study (here T_i refers to any of the different effect-size indices presented above, both from the d and the r families.) In a fixed-effects model, it is assumed that all of the effect-size estimates come from a population with a common parametric effect size, θ , and as a consequence the only error source is that produced by sampling error, e_i . Thus, the model can be formulated as $T_i = \theta + e_i$, the sampling errors, e_i , being normally distributed with mean 0 and sampling variance σ_i^2 , $e_i \sim N(0, \sigma_i^2)$. Therefore, the effect-size estimates, T_i , are also normally distributed with mean θ and sampling variance σ_i^2 , $T_i \sim N(\theta, \sigma_i^2)$.

In a random-effects model, it is assumed that the effect-size estimates, T_i , estimate different population effect sizes, θ_i , that is, $T_i = \theta_i + e_i$, and θ_i pertains to a distribution of parametric effect sizes with mean μ and variance τ^2 , usually called between-studies variance. The parametric effect sizes can be modeled as $\theta_i = \mu + \varepsilon_i$, ε_i being the errors of the parameters around its mean, μ . Therefore, the random-effects model is formulated as $T_i = \mu + \varepsilon_i + e_i$. Assuming normality, T_i has as mean μ and variance $\tau^2 + \sigma_i^2$, $T_i \sim N(\mu, \tau^2 + \sigma_i^2)$. Thus, the fixed-effects model can be considered a particular case of the random-effects model when the between-studies variance is zero ($\tau^2 = 0$) and, as a consequence, all the parametric effect sizes are equal ($\theta_1 = \theta_2 = \dots = \theta_{i..} = \theta = \mu$).

To calculate an average effect size from a set of studies, each effect-size estimate must be weighted by its precision. Both in a fixed- and a random-effects model, the uniformly minimum variance unbiased estimator (UMVUE) of the average effect size, μ , is that obtained by weighting each effect-size estimate by its inverse variance:

$$T_{\text{UMVUE}} = \frac{\sum_{i=1}^k w_i T_i}{\sum_{i=1}^k w_i} \quad [8]$$

where w_i is the optimal weight for the i th study and, depending on the statistical model assumed, it is defined as $w_i^{\text{FE}} = 1/\sigma_i^2$ or as $w_i^{\text{RE}} = 1/(\tau^2 + \sigma_i^2)$, for the fixed- and the random-effects models, respectively.

In practice, the optimal weights cannot be used, because the within-study sampling variances, σ_i^2 , and the between-studies variance, τ^2 , are unknown. For each effect-size index, formulas have been devised to estimate σ_i^2 and τ^2 . Thus, the estimated weights are defined as $\hat{w}_i^{\text{FE}} = 1/\hat{\sigma}_i^2$ and $\hat{w}_i^{\text{RE}} = 1/(\hat{\tau}^2 + \hat{\sigma}_i^2)$ for fixed- and random-effects models, respectively. (Another option

consists of simply weighting each effect-size estimate by its sample size (Hunter and Schmidt, 2004).) Here $\hat{\sigma}_i^2$ is the estimated within-study sampling variance for the i th study (e.g., eqn [4] for the d index), and $\hat{\tau}^2$ can be calculated by applying one of the different estimators of the between-studies variance proposed in the meta-analytic literature. The τ^2 estimator most usually applied in meta-analysis is that based on the moments method:

$$\hat{\tau}^2 = \frac{Q - (k - 1)}{c} \quad [9]$$

where k is the number of studies, Q is a heterogeneity statistic defined in eqn [17], and c is obtained by:

$$c = \sum_{i=1}^k \hat{w}_i^{\text{FE}} - \frac{\sum_{i=1}^k (\hat{w}_i^{\text{FE}})^2}{\sum_{i=1}^k \hat{w}_i^{\text{FE}}} \quad [10]$$

When $Q < (k - 1)$, then $\hat{\tau}^2$ is negative and must be truncated to zero. Other τ^2 estimators can be consulted in Viechtbauer (2005).

With the respective estimated variances, the population effect size, μ , is then estimated by:

$$T_{\text{FE}} = \frac{\sum_{i=1}^k \hat{w}_i^{\text{FE}} T_i}{\sum_{i=1}^k \hat{w}_i^{\text{FE}}} \quad [11]$$

$$T_{\text{RE}} = \frac{\sum_{i=1}^k \hat{w}_i^{\text{RE}} T_i}{\sum_{i=1}^k \hat{w}_i^{\text{RE}}} \quad [12]$$

for fixed- and random-effects models, respectively. When a fixed-effects model is assumed, T_{FE} is approximately normally distributed and its sampling variance defined as:

$$V(T_{\text{FE}}) = 1 / \sum_{i=1}^k \hat{w}_i^{\text{FE}} \quad [13]$$

Thus, a confidence interval for the average effect size can be obtained by (cf. e.g., Cooper *et al.* 2009):

$$T_{\text{FE}} \pm z_{\alpha/2} \sqrt{V(T_{\text{FE}})} \quad [14]$$

where $z_{\alpha/2}$ is the $100(\alpha/2)$ percentile of the standard normal distribution and α is a significance level.

Under a random-effects model, a better approach for obtaining a confidence interval for the overall effect size consists of assuming a Student t reference distribution with $k - 1$ degrees of freedom, instead of the standard normal distribution:

$$T_{\text{RE}} \pm t_{k-1, \alpha/2} \sqrt{V(T_{\text{RE}})} \quad [15]$$

where $t_{k-1, \alpha/2}$ is the $100(\alpha/2)$ percentile of the Student t distribution with $k - 1$ degrees of freedom, and $V(T_{\text{RE}})$ is an estimate of the sampling variance for T_{RE} , which is obtained by (cf. Sánchez-Meca and Marín-Martínez, 2008)

$$V(T_{\text{RE}}) = \frac{\sum_{i=1}^k \hat{w}_i^{\text{RE}} (T_i - T_{\text{RE}})^2}{(k - 1) \sum_{i=1}^k \hat{w}_i^{\text{RE}}} \quad [16]$$

Alternative approaches to those presented in eqns [14] and [15] have been proposed (cf. e.g., Sánchez-Meca and Marín-Martínez, 2008).

Assessing Heterogeneity in the Effect Sizes

One of the main purposes of meta-analysis is to examine the variability among the effect-size estimates. If the heterogeneity exhibited by the effect estimates can be explained by random sampling alone, then the average effect size represents a good summary of the general trend in the study results. If, on the contrary, the effect estimates reflect true heterogeneity, that is, variability among the effect sizes due to differential characteristics of the studies, then the average effect size does not represent the effect estimates, and the meta-analyst should search for moderator variables that can explain (at least part of) the heterogeneity.

To assess whether a set k of independent effect estimates show true heterogeneity, the most usual strategy in meta-analysis is to apply the heterogeneity Q statistic which, under the null hypothesis of homogeneity ($H_0: \theta_1 = \theta_2 = \dots = \theta_k = \mu$), follows a chi-square distribution with $k - 1$ degrees of freedom, and is obtained through

$$Q = \sum_{i=1}^k \hat{w}_i^{\text{FE}} (T_i - T_{\text{FE}})^2 \quad [17]$$

By assuming a given significance level, α , it is possible to make a decision about whether the effect-size estimates are homogeneous or not. However, due to the low statistical power of the Q statistic when the number of studies is small (about $k < 30$), it is advisable to complement this statistical test with the I^2 index, which quantifies the extent of true heterogeneity present in the studies (Higgins and Thompson, 2002):

$$I^2 = \frac{Q - (k - 1)}{Q} \times 100 \quad [18]$$

When $Q < (k - 1)$, then I^2 is truncated to zero. Values of $I^2 = 25\%$, 50% , and 75% can be interpreted as reflecting small, medium, and high heterogeneity among the effect sizes, respectively.

Analyzing the Influence of Moderator Variables

If the effect-size estimates exhibit more heterogeneity than random sampling can explain, then the meta-analyst has to search for moderator variables that can account for that heterogeneity. In order to accomplish this objective, linear models are assumed, where the effect-size estimates are taken as the dependent variable, and moderator variables coded in the studies act as predictor variables. The most appropriate statistical model for dealing with this task is the mixed-effects model, in which the studies (or the effect-size estimates) are taken as random-effects variables, and study-level moderator variables are taken as fixed-effects variables. When the moderator variable is categorical (e.g., type of population), an analysis of variance by weighted least squares is applied, whereas regression models are used to assess the influence of continuous moderator variables (e.g., duration of the program, mean age in the sample).

In matrix notation the regression coefficient vector of the model is estimated by $\hat{\beta} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{T}$, where \mathbf{X} is a $k \times p$ design matrix that includes p predictors, \mathbf{T} is the effect-sizes vector, with range k , and \mathbf{W} is a $k \times k$ diagonal weighting matrix, with the element in the diagonal defined as the estimated weights: $\hat{w}_i^{\text{ME}} = 1/(\hat{\tau}_{\text{ME}}^2 + \hat{\sigma}_i^2)$. The between-studies variance, $\hat{\tau}_{\text{ME}}^2$, is estimated by:

$$\hat{\tau}_{\text{ME}}^2 = \frac{Q_E - (k - p)}{\text{tr}\mathbf{W} - \text{tr}[\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}]} \quad [19]$$

where tr represents the trace of a matrix and Q_E is the weighted residual sum of squares of the regression model, that is defined as:

$$Q_E = \mathbf{T}'\mathbf{W}\mathbf{T} \quad [20]$$

Model misspecification can be tested by using the Q_E statistic. Under the null hypothesis that the regression model is not misspecified ($H_0: \mathbf{T} = \mathbf{X}\beta$), the Q_E statistic follows a chi-square distribution with $k - p - 1$ degrees of freedom.

Once the model misspecification has been examined, the significance of the full regression model is tested with the weighted regression sum of squares, Q_R , which is obtained by

$$Q_R = \hat{\beta}'\hat{\Sigma}_{\beta}\hat{\beta} \quad [21]$$

where $\hat{\Sigma}_{\beta}$ is the estimated variance–covariance matrix of the coefficient vector. Under the null hypothesis of no relationship between the predictors and the effect estimates ($H_0: \beta = \mathbf{0}$), the Q_R statistic follows a chi-square distribution with p degrees of freedom. Finally, the partialized effect of each predictor can be assessed by applying a Z test for its regression coefficient, $\hat{\beta}_j$:

$$Z = \hat{\beta}_j / \hat{\sigma}_{\hat{\beta}_j} \quad [22]$$

where $\hat{\sigma}_{\hat{\beta}_j}$ is the estimated standard error of the regression coefficient, taken from the j th diagonal element of $\hat{\Sigma}_{\beta}$. Under the null hypothesis of no effect for the predictor, Z follows a standard normal distribution.

An Example

To illustrate the meta-analytic calculations presented in the previous epigraph, we have selected part of the data reported in Ginns' (2006) meta-analysis about the instructional effects of manipulating the temporal contiguity of learning materials. The effect-size index was the standardized mean difference, d_b , defined as the difference between the means of the experimental (temporally integrated materials) and control (non-integrated condition) groups divided by the pooled standard deviation (see eqn [1]). **Table 1** shows the meta-analytic database for this example, which includes information relative to 13 studies. (As Ginns' (2006) meta-analysis did not report the sample size for each study, we have invented them in order to carry out the statistical analyses. Therefore, our results have only illustrative purposes and do not coincide with those presented in Ginns (2006).) The moderator variables reported in **Table 1** were the publication year, the sample size, n_i (here we assumed that the sample sizes of the experimental and control groups were equal in each study, i.e., $n_{Ei} = n_{Ci} = n_i$), the type of testing (group vs. individual), and the field of the study (science vs. engineering/technical). **Table 1** also shows the estimated within-study variance, $\hat{\sigma}_i^2$, for each study, obtained by eqn [4], and the estimated fixed- and random-effects weights: $\hat{w}_i^{\text{FE}} = 1/\hat{\sigma}_i^2$ and $\hat{w}_i^{\text{RE}} = 1/(\hat{\tau}^2 + \hat{\sigma}_i^2)$, respectively (**Figure 1**).

A graphical representation typical in meta-analysis is the forest plot, which consists of showing each individual estimated effect size and its 95% confidence interval and, at the bottom of the graph, the average effect size for all of the studies. This graph helps to capture the general trend and the variability of the effect sizes.

Table 2 presents the average effect size obtained with the 13 effect estimates and the 95% confidence interval for the fixed- and the random-effects models. Thus, assuming a fixed-effects model, the average effect size was $d_+ = 0.768$ (by applying eqn [11]), with confidence limits, 0.637 and 0.899 (by applying eqn [14]). Assuming a random-effects model, the average effect size was $d_+ = 0.888$ (with eqn [12]), with confidence limits, 0.574 and 1.202 (with eqn [15]). In both cases, the average effect size was statistically significant, as the null effect was not contained in the confidence intervals. The confidence interval for the random-effects model was larger (interval amplitude: 0.628) than that of the fixed-effects model (interval amplitude: 0.262), due to the fact that the random-effects model assumes that the effect sizes of the studies estimate different population effect sizes.

Table 1 Meta-analytic dataset of the effects of the temporal contiguity of instructional materials on student performance

Study	Year	Testing	Broad field of the study	n_i	d_i	$\hat{\sigma}_i^2$	\hat{w}_i^{FE}	\hat{w}_i^{RE}
1	1999	Group	Science	12	2.06	0.2551	3.9204	2.4352
2	1983	Group	Science	22	1.40	0.1132	8.8353	3.7209
3	1983	Group	Science	45	0.27	0.0448	22.2968	4.9896
4	1991	Individual	Science	40	1.10	0.0576	17.3724	4.6920
5	1994	Individual	Science	65	0.58	0.0321	31.1885	5.3296
6	1999	Group	Engineering/Technical	22	1.31	0.1104	9.0571	3.7597
7	1984	Group	Engineering/Technical	52	0.65	0.0405	24.6958	5.1005
8	1984	Group	Engineering/Technical	33	0.50	0.0625	16.0000	4.5857
9	1992	Individual	Engineering/Technical	30	1.64	0.0891	11.2259	4.0875
10	1992	Individual	Engineering/Technical	35	1.40	0.0711	14.0562	4.4109
11	1991	Individual	Engineering/Technical	35	0.87	0.0625	15.9874	4.5847
12	1994	Individual	Engineering/Technical	52	0.49	0.0396	25.2424	5.1234
13	2000	Individual	Engineering/Technical	48	0.33	0.0422	23.6777	5.0556

n_i – sample size for each group of the i th study, the total sample size of the study being $N_i = 2xn_i$; d_i – standardized mean difference for the i th study; $\hat{\sigma}_i^2$ – estimated within-study variance for the i th study (calculated by eqn [4]); $\hat{w}_i^{FE} = 1/\hat{\sigma}_i^2$ is the estimated fixed-effects weight for the i th study; and $\hat{w}_i^{RE} = 1/(\hat{\tau}^2 + \hat{\sigma}_i^2)$ is the estimated random-effects weight for the i th study.

From Ginns P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction* 16, 511–525.

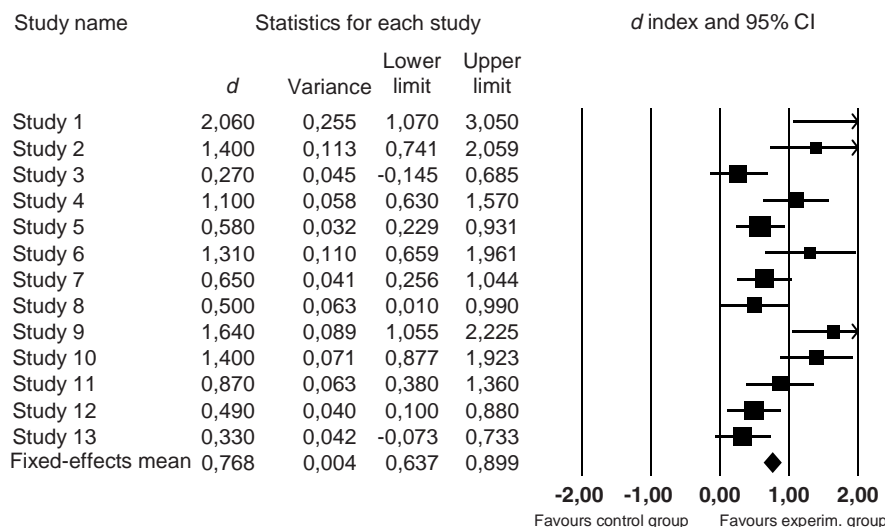


Figure 1 Forest plot of the d indices for 13 studies on the temporal contiguity effects. The variances were calculated by eqn [4], the 95% confidence intervals by eqn [11], and the mean effect size from a fixed-effects model by eqn [13].

Table 2 Summary results of calculating an average effect size and a 95% confidence interval from the fixed- and random-effects models

Statistical model	k	d_+	$V(d_+)$	95% C. I.	
				d_l	d_u
Fixed-effects	13	0.768	0.0045	0.637	0.899
Random-effects	13	0.888	0.0208	0.574	1.202

K – number of studies; d_+ – average effect size calculated by eqns [11] and [12] for the fixed- and random-effects models, respectively; $V(d_+)$ – sampling variance of d_+ , calculated by eqns [13] and [16], respectively; d_l and d_u – lower and upper confidence limits of the 95% confidence interval around d_+ , calculated by eqns [14] and [15], respectively.

The heterogeneity Q statistic, calculated by eqn [17], was statistically significant [$Q(12) = 43.584$, $p < .001$]; the I^2 index, by eqn [18], was of high magnitude: $I^2 = 72.5\%$; and the between-studies variance was $\hat{\tau}^2 = 0.155$ (by eqn [9]). Therefore, the set of effect sizes exhibited more variability than random sampling can explain and, as a consequence, the search for moderator variables is justified. In order to do this, analyses of variance and regression models by weighted least squares and assuming a mixed-effects model can be applied.

Table 3 presents an example of how to analyze the influence of a qualitative moderator variable on the effect size by applying an analysis of variance. The moderator

Table 3 Analysis of variance by weighted least squares and assuming a mixed-effects model of the d indices as a function of the field of the study

Field of study	k	d_+	95% C. I.		Q_{wj}	DF	p
			d_l	d_u			
Science	5	0.947	0.499	1.395	6.336	4	.175
Engineering/ technology	8	0.523	0.523	1.205	6.762	7	.454
$Q_B(1) = 0.084, p = .772$							
$Q_W(11) = 13.098, p = .287$							

K – number of studies for each category; d_+ – average effect size for each category; d_l and d_u – lower and upper confidence limits of the 95% confidence interval around d_+ for each category; Q_{wj} – within-category homogeneity Q statistic which, under the null hypothesis of within-category homogeneity, follows a chi-square distribution with $k - 1$ degrees of freedom (DF); Q_W – global within-category homogeneity Q statistic which, under the null hypothesis of global within-category homogeneity, follows a chi-square distribution with $k - c$ degrees of freedom, c being the number of categories; Q_B – between-categories Q statistic which, under the null hypothesis of equal average effect sizes, follows a chi-square distribution with $c - 1$ degrees of freedom; and p – probability level of the corresponding Q statistic.

Table 4 Simple regression models by weighted least squares and assuming a mixed-effects model of the d indices as a function of the publication year and sample size

Predictor variable	k	$\hat{\beta}_j$	$SE(\hat{\beta}_j)$	$Q_R(DF)$	p	$Q_E(DF)$	p	R^2
Year	13	0.018	0.024	0.604(1)	.437	12.771(11)	.308	.045
Sample size	13	-0.027	0.008	11.886(1)	< .001	10.630(11)	.475	.528

K – number of studies for each category; $\hat{\beta}_j$ – unstandardized regression coefficient for each predictor variable; $SE(\hat{\beta}_j)$ – standard error of $\hat{\beta}_j$; Q_R – weighted regression sum of squares (by eqn [21]); DF – degrees of freedom; Q_E – weighted residual sum of squares (by eqn [20]); p – probability level of the corresponding Q statistic; and R^2 – proportion of variance accounted for.

variable is the field of study (science vs. engineering/technical). The average effect sizes for both categories were very similar: $d_+ = 0.947$ for science and $d_+ = 0.864$ for engineering/technical, leading to a nonstatistically significant result for the Q_B statistic, which assesses whether the mean effect sizes are equal [$Q_B(1) = 0.084, p = .772$]. The Q_W statistic reveals that the model is not misspecified [$Q_W(11) = 13.098, p = .287$]. Therefore, the field of study does not seem to be a moderator variable of the effect sizes.

Finally, **Table 4** presents the results of two simple regression analyses, one for the publication year and the other for the sample size. The sample size exhibited a statistically significant relationship with the effect sizes, as the Q_R statistic reveals [$Q_R(1) = 11.886, p < 0.001$], whereas publication year did not reach statistical significance [$Q_R(1) = 0.604, p = 0.437$]. Moreover, the regression model for sample size was not misspecified [$Q_E(11) = 10.630, p = 0.475$]. Hence, it seems that the effect sizes were associated with sample sizes.

Concluding Remarks

In this article, we have focused on the meta-analytic approach most usually applied in education, which

consists of calculating an effect-size index from each study and putting it into relation with study-level moderator variables that can explain the variability usually exhibited by the effect sizes. In addition to the statistical methods presented here, other analytic approaches have been devised in the meta-analytic literature. Thus, Hunter and Schmidt (2004) proposed the validity generalization approach with the purpose of integrating a set of validity coefficients of a given measurement instrument obtained from applying the instrument to different samples and contexts. In this meta-analytic approach, also named psychometric meta-analysis, different statistical and measurement artifacts that affect the empirical validity coefficients can be corrected, such as the measurement error, the range restriction, or the effects of dichotomizing variables. In the same vein, Vacha-Haase (1998) has proposed the meta-analytic reliability generalization approach, with the aim of integrating a set of reliability coefficients obtained in successive applications of the same measurement instrument, and assessing whether the reliability of the test scores can be generalized across different samples, contexts, and applications. Other specific approaches devised in meta-analysis enable us to apply multivariate models to integrate a set of dependent effect-size estimates obtained from each study (cf. e.g., Becker, 2000), or to integrate correlation matrices

obtained from the studies in order to meta-analyze factor analyses (e.g., Becker, 1996).

Without a doubt, meta-analysis is contributing in education to better accumulate the scientific evidence disseminated across studies by applying the same scientific rigor demanded of primary research. Meta-analysis has become a necessary link between the past research and the future research efforts in any field. The main characteristic of this research methodology is the use of statistical methods to integrate study results but, as it is a relatively new field, more methodological work is needed in order to adapt the current meta-analytic techniques to address more complex problems.

See also: Measure of Association; Statistical Significance Versus Effect Size.

Bibliography

- Becker, B. J. (2000). Multivariate meta-analysis. In Tinsley, H. E. A. and Brown, S. D. (eds.) *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, pp 499–525. New York: Academic Press.
- Becker, G. (1996). The meta-analysis of factor analyses: An illustration based on the cumulation of correlation matrices. *Psychological Methods* **1**, 341–353.
- Cooper, H., Hedges, L. V., and Valentine, J. C. (eds.) (2009). *The Handbook of Research Synthesis and Meta-analysis*, 2nd edn. New York: Russell Sage Foundation.
- Deeks, J. J. and Altman, D. G. (2001). Effect measures for meta-analysis of trials with binary outcomes. In Egger, M., Smith, G. D., and Altman, D. (eds.) *Systematic Reviews in Health Care*, 2nd edn., pp 313–335. London: BMJ Publication Group.
- Giinns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction* **16**, 511–525.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher* **5**, 3–8.
- Glass, G. V., McGaw, B., and Smith, M. L. (1981). *Meta-analysis in Social Research*. Beverly Hills, CA: Sage.
- Hedges, L. V. and Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.
- Higgins, T. P. T. and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**, 1539–1558.
- Hunter, J. E. and Schmidt, G. L. (2004). *Methods of Meta-Analysis: Correcting Error and Bias in Research Synthesis*, 2nd edn. Newbury Park, CA: Sage.
- Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-group designs. *Organizational Research Methods* **11**, 364–386.
- Sánchez-Meca, J. and Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods* **13**, 31–48.
- Sánchez-Meca, J., Marín-Martínez, F., and Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods* **8**, 448–467.
- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement* **58**, 6–20.
- Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics* **30**, 261–293.

Further Reading

- Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2009). *Introduction to Meta-analysis*. Chichester, UK: Wiley.
- Cooper, H. (2009). *Research Synthesis and Meta-analysis: A Step-by-Step Approach*, 3rd edn. Thousand Oaks, CA: Sage.
- Hedges, L. V. and Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods* **3**, 486–504.
- Kalaian, H. A. and Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods* **1**, 227–235.
- Lipsey, M. W. and Wilson, D. B. (2001). *Practical Meta-Analysis*. Thousand Oaks, CA: Sage.
- Petticrew, M. and Roberts, H. (2006). *Systematic Reviews in the Social Sciences: A Practical Guide*. Malden, MA: Blackwell.
- Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research*, revised edition. Newbury Park, CA: Sage.
- Rothstein, H. R., Sutton, A. J., and Borenstein, M. (eds.) (2005). *Publication Bias in Meta-Analysis: Prevention, Assessment, and Adjustments*. Chichester, UK: Wiley.
- Sánchez-Meca, J. and Marín-Martínez, F. (1998). Testing continuous moderators in meta-analysis: A comparison of procedures. *British Journal of Mathematical and Statistical Psychology* **51**, 311–326.
- Schmidt, F. L. (2008). Meta-analysis: A constantly evolving research integration tool. *Organizational Research Methods* **11**, 96–113.
- Schmidt, F. L., Oh, I.-S., and Hayes, T. L. (2009). Fixed versus random effects models in meta-analysis: Model properties and an empirical comparison of difference in results. *British Journal of Mathematical and Statistical Psychology* **62**, 97–128.
- Schulze, R. (2004). *Meta-Analysis: A Comparison of Approaches*. Cambridge, MA: Hogrefe and Huber.
- Sutton, A. J. and Higgins, J. P. T. (2008). Recent developments in meta-analysis. *Statistics in Medicine* **27**, 625–650.
- Viechtbauer, W. (2007). Accounting for heterogeneity via random-effects models and moderator analyses in meta-analysis. *Zeitschrift für Psychologie/Journal of Psychology* **215**, 104–121.
- Whitehead, A. (2002). *Meta-Analysis of Controlled Clinical Trials*. Chichester, UK: Wiley.

Relevant Websites

- <http://www.vanderbilt.edu> – Center for Evaluation Research and Methodology, Vanderbilt University (Nashville, TN).
- www.campbellcollaboration.org – The Campbell Collaboration.
- www.duke.edu – The Campbell Methods Group.
- www.cochrane.org – The Cochrane Collaboration.
- <http://www.um.es> – The Meta-Analysis Unit, University of Murcia, Murcia, Spain.
- <http://mason.gmu.edu> – Website of Prof. David B. Wilson, George Mason University.
- <http://faculty.ucmerced.edu> – Website of Prof. William R. Shadish, University of California, Merced.

Missing Data

C A W Glas, University of Twente, Enschede, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

In educational research, missing data create various problems. On the most elementary level, missing data reduce the sample size which decreases the precision of estimates and the power of statistical tests. From a practical perspective, missing data are often a nuisance because many computational procedures prefer complete data sets. Therefore, statistical packages either only accept complete data sets, or provide tools for creating a complete data set through the imputation of missing values. In both cases, action is required. Finally, missing data may cause a threat to the validity of the inferences. If, for a random part of a sample, some pages in a questionnaire are missing due to an error of the printer, it may not pose a serious threat to validity. However, if teachers are inclined to exclude low-ability students from an educational survey because their assessment results are too embarrassing, inferences will obviously be biased. Rubin (1976) described and named a number of types of mechanisms causing missing data and investigated their consequences for statistical analyses. This topic is addressed first; subsequently, a number of ways of dealing with missing data are discussed.

Missing Data Mechanisms

The mechanism that caused the missing data is mostly unknown to the statistician analyzing the data. To shed some light on the mechanism, the statistician can compare the outcomes obtained from analyses under various assumptions. An important aid for making such assumptions is the following classification of the possible mechanisms causing the missing data. The mechanisms causing the incompleteness are characterized according to their degree of randomness.

Missing Completely at Random

The least problematic is the case where the missingness is unrelated to both observed data and unobserved data. An example is the above-mentioned problem of missing pages in a questionnaire due to an error of the printer. This missingness clearly does not depend on any observed or unobserved responses. In such a situation, the missing data are called missing completely at random (MCAR). To make this more formal, consider variables (say socioeconomic status, intelligence quotient, item responses, and gender) labeled $k = 1, \dots, M$ and cases (students,

teachers, etc.) labeled $i = 1, \dots, N$. Further, for the potential observation, y_{ik} , we define a missing data indicator

$$d_{ik} = \begin{cases} 0 & \text{if } y_{ik} \text{ was not observed} \\ 1 & \text{if } y_{ik} \text{ was observed.} \end{cases}$$

MCAR entails that d_{ik} and y_{ik} are independent, and the methods for handling missing data discussed below can be directly applied without further assumptions.

Missing at Random

A slightly more complicated case arises when the missing data indicators d_{ik} are independent of the missing data, but depend on the observed data. Therefore, if \mathbf{D} is a matrix of all missing data indicators d_{ik} , and \mathbf{Y}_{obs} and \mathbf{Y}_{miss} stand for the observed and missing part of the data, respectively, then

$$p(\mathbf{D} | \mathbf{Y}_{obs}, \mathbf{Y}_{miss}, \phi, \mathbf{X}) = p(\mathbf{D} | \mathbf{Y}_{obs}, \phi, \mathbf{X}),$$

where \mathbf{X} stands for a set of covariates that might be observed and ϕ is a vector of the parameters of the distribution of \mathbf{D} .

As an example, consider the well-known case of missing at random (MAR) that arises in educational assessment in two-stage testing (Lord, 1980). In this testing, students are first presented a so-called routing test. If the number-correct score on the routing test is above some cutoff point, a student is administered a difficult follow-up test, and if the number-correct score is below the cutoff point, an easy follow-up test is presented. The missing data are the responses of the students above the cutoff point on the easy follow-up test and those below the cutoff point on the difficult follow-up test. These missing data are MAR because the missing data indicators are completely determined by the observations, that is, by the number-correct scores on the routing test. The consequences are discussed in the next section.

Ignorable Missingness

Suppose that we want to estimate a parameter or a vector of parameters θ . It seems plausible to do this using the data that are actually available. For instance, one may use maximum likelihood (ML) estimation and maximize the likelihood

$$P(\mathbf{Y}_{obs} | \theta, \mathbf{X}) = \prod_{i=1}^N \prod_{k=1}^M p(y_{ik} | \theta, \mathbf{x}_i)^{d_{ik}} \quad [1]$$

as a function θ . The density $p(y_{ik} | \theta, \mathbf{x}_i)$ may, for instance, be a normal density associated with the regression of y_{ik} on \mathbf{x}_i ,

and θ contains regression coefficients and variance parameters. The density can also be a probability associated with a discrete item response y_{ik} and θ contains the item and person parameters of an item response theory (IRT) model (Lord, 1980). Rubin (1976) shows that estimation procedures that maximize the likelihood given by eqn [1] are justified if ignorability holds. Rubin's so-called ignorability principle entails that ignoring the missing data mechanism in frequentist inferences based on the likelihood function and Bayesian inferences based on the posterior distribution does not create bias if two conditions are met. The first condition is that the missing data should be MAR and the second is that the parameters of interest, say θ , should be distinct from the parameters of the missing data process ϕ . In a frequentist framework, distinctness means that there are no functional dependencies in the parameter space of θ and ϕ (e.g., θ and ϕ should not be identical). In a Bayesian framework, distinctness means that the prior distributions of θ and ϕ are independent. If the conditions MAR and distinctness are satisfied, the missing data process is ignorable in statistical inferences, which means that using the distribution $P(Y_{obs}|\theta, X)$ instead of $P(Y_{obs}, Y_{mis}|\theta, \phi, X)$ and thus ignoring the process that caused the missing data does not produce bias in the parameter estimates.

Applied to the two-stage testing design presented above, this leads to the interesting conclusion that the estimates of θ obtained from the items administered to the examinees are not influenced by the fact that the high-ability students are presented a difficult test and the low-ability students an easy test.

Nonignorable Missingness

When ignorability is violated, missingness should not be ignored. It may be that the MAR assumption is violated because the probability of missingness is related to values that are missing themselves. An example is that respondents in a survey may fail to answer questions of a sensitive nature (e.g., about income, sexual habits, or weight) because their values on these variables are too extreme. Alternatively, distinctness may be violated because the parameters of interest θ and the parameters governing the missing data process ϕ are highly correlated or identical, for instance, when students have the opportunity to choose the items they respond to and do so based on an ability parameter θ that is the target of estimation. If ignorability does not hold, the process causing the missing data must be taken into account in the estimation procedure. This is further clarified below.

Missing Data Techniques

Many taxonomies of methods for dealing with missing data are available (see, e.g., Little and Rubin, 1987 or Buhi *et al.*, 2008); however, none of them is completely

satisfactory. In the present article, we use an imperfect grouping of methods for dealing with missing data into three categories: deletion procedures, single imputation procedures, and model-based estimation and multiple imputations procedures.

Deletion Procedures

The most used deletion procedures are listwise and pairwise deletion. Listwise deletion (or casewise deletion) entails excluding cases with missing values for any variable. For example, all students that fail to respond to one or more items in an assessment are removed from the analysis. In pairwise deletion (or available case analysis), all available data for each variable and each pair of variables are used to compute such statistics as means, variances, and covariances. Though deletion procedures are readily available in multipurpose statistical packages (such as Self-Propelled Semi-Submersible (SPSS), STATA, and SAS), such procedures are not very recommendable. The first problem with this approach is that it reduces the sample size, which results in a decrease of precision of estimates and power of tests. Further, both methods are based on the assumption of MCAR. If the assumption is violated, the inferences are probably misleading. General purpose statistical software packages, such as SPSS, support testing this assumption using Little's chi-square test (Little and Rubin, 1987). A special problem arises with pairwise deletion when covariances are estimated. Since the estimates are based on differing numbers of observations, the resulting covariance matrix may not be proper, that is, it may not be positive definite. In that case, techniques based on the covariance matrix, such as factor analysis and structural equation modeling, become infeasible, or at least highly problematic. Further, correlation matrices derived from covariance matrices estimated using pairwise deletion can be outside the range of -1.0 to 1.0 .

Single Imputation Procedures

Imputation methods are those where the missing data are filled in to create a complete data matrix that can be analyzed using standard methods. Single imputation procedures are those where one value for a missing data element is filled in without defining an explicit model for the partially missing data. A method that is often used is mean substitution where missing values are replaced with the variable mean score. However, an observation y_{ik} can both depend on an effect of variable k and a case-specific effect i . Therefore, more advanced versions of the method take both the effects of the variables and the case-specific effects on the observations y_{ik} into account. An example in educational assessment is Kelly's method for filling in missing student grades. This method estimates standardization weights for subject grades such that the difficulty of subjects and the strictness of possible raters are corrected for.

These standardization weights are then used to approximate a student's grade in a subject which would be obtained in the ideal situation when all students took all subjects, and all subjects were marked by the same examiners (Kelly, 1976).

In hot-deck imputation, missing values are replaced with observed values from a respondent with the same or a very close response pattern on a group of matching variables. Finally, in regression imputation, the missing value for a targeted variable is estimated using the regression of the target variable on all other variables or a subset of all other variables.

One of the problems with all single imputation methods is that the filled-in observations are treated as actual observations in subsequent analysis. However, the filled-in values are estimates, which have standard errors. Multiple imputations can be used to take the uncertainty about the estimates into account. This is one of the topics of the next section.

Model-Based Estimation and Multiple Imputations

It has already been discussed above that if ignorability holds, a model can be estimated only using the observed data. ML estimation (maximization of $P(Y_{obs}|\theta, X)$ as a function of θ) and Bayesian estimation (estimation of the posterior distribution $P(\theta|Y_{obs}, X)$), both with an observed data likelihood as in eqn [1], are especially suited for this approach. For instance, most software for the analysis of educational assessments with IRT modeling uses this approach; examples are Bilog-MG (Zimowski *et al.*, 1996) and Conquest (Wu *et al.*, 1997). However, in many cases, ML and Bayesian methods based on eqn [1] may become quite complicated. Solutions in these cases are found in the Expectation-Maximization (EM) algorithm and the Markov chain Monte Carlo (MCMC) method with data augmentation.

For advanced practitioners, these are valuable tools for computing estimates for complex models with missing data. However, they are also the basis for the generally used procedure of multiple imputation. In multiple imputation (Rubin, 1987, 1996), a number of draws of the missing data (say, between three and six) are made from the posterior distribution of the missing data, given the observed data $p(Y_{mis}|Y_{obs}, \theta)$. This distribution is available through the EM algorithm or the MCMC algorithm, which are both outlined below. Subsequently, for each of these draws, estimates are computed using standard software for complete data, and, finally, the results are averaged over the draws. Accordingly, uncertainty about missing values is taken into account. Multiple imputation for variables following the multivariate normal distribution is supported by programs as NORM (Schafer, 1999), S-plus 6 for Windows (2006), and SAS 8.1 (Yuan, 2000). Multiple imputations of categorical variables can be

created using the log-linear model (Schafer 1997), which is implemented in the missing data library of S-plus (2006). For data sets containing a combination of categorical and continuous variables, Schafer (1997) proposed a general location model, which is a combination of a log-linear and a multivariate normal model. This model is also implemented in the missing data library of S-plus.

In a generalized linear model (such as linear regression, logistic regression, or hierarchical linear modeling), missing data can both appear in dependent and independent variables. When creating imputations for missing values of the independent variables, all information, including the available values for the dependent variables, should be used. (see, e.g., Allison, 2002). If the dependent variables are disregarded, the estimated regression coefficients are biased toward zero. Imputation of missing values for the dependent variable is somewhat more complicated. Von Hippel (2007) proposes a procedure wherein all cases are used to determine the imputations, but where the cases with imputed values on the dependent variable are excluded for the subsequent analyses.

The analysis of categorical variables using log-linear models is based on data organized in frequency tables. An important limitation of imputation of missing values for categorical variables using log-linear models is that the number of variables that can be handled is limited by the size and dimensionality of these frequency tables. Therefore, Vermunt *et al.* (2008) proposed to use latent class modeling as a tool for density estimation in the presence of missing data. In this application, the number of latent classes is chosen to be large because the interest is not in coming up with an interpretable set of latent classes underlying the data, but in accurately estimating the higher-order moments of the data. The procedure is implemented in the software program Latent GOLD.

Multiple imputations can be generated by many procedures (e.g., Latent GOLD uses a nonparametric bootstrap procedure). However, the EM algorithm and the MCMC algorithm are the two most common procedures. They are outlined as follows.

The EM algorithm (Dempster *et al.*, 1977) is a general iterative algorithm for ML estimation in incomplete data problems. It handles missing data as follows. First, given some initial estimate, say θ^* , the posterior distribution of the missing values $p(Y_{mis}|Y_{obs}, \theta^*)$ is computed. In many instances, it also suffices to compute the expectation of the sufficient statistic for θ ; therefore, in these cases, this implies computing an actual value for the missing data. Second, we compute ML estimates of θ using the likelihood $p(Y_{obs}, Y_{mis}|\theta)$ averaged over the distribution of the missing data $p(Y_{mis}, Y_{obs}|\theta^*)$ computed in the previous step. Again, in many instances, it suffices to fill in the expectation of the sufficient statistics for θ , after which averaging is not necessary. Finally, these two steps (the E-step and the M-step) are repeated until the process converges.

An alternative approach to the ML framework with the EM algorithm is a Bayesian approach using the MCMC algorithm with data augmentation. In a Bayesian approach, data and parameters have the same status: they are both seen as random variables. Data augmentation (Tanner and Wong, 1987) entails adding random variables to the model to simplify computations. A much-used version of the MCMC algorithm is the Gibbs sampler (Gelfand *et al.*, 1990; Gelman *et al.*, 1995; Geman and Geman, 1984). For the Gibbs sampler, the variables are partitioned into a number of subsets and samples are drawn from the posterior distribution of each subset, given previous draws. Applied to the present problem, in its simplest versions, the algorithm creates a sample from $p(Y_{mis}, \theta | Y_{obs})$ by alternately sampling from $p(\theta | Y_{mis}, Y_{obs})$ and $p(Y_{mis} | Y_{obs}, \theta)$. After choosing starting values for θ , the algorithm must first make a number of so-called burn-in iterations to move θ into the region where the posterior distribution has its main probability mass. Subsequently, every draw comes from the targeted posterior.

Modeling the Missing Data Process

If the missing data process is not ignorable, the process causing the missing data must be taken into account. This is done by considering both the observed data Y_{obs} and the missing data indicator D as random variables, with distributions with parameters θ and φ , respectively. The likelihood can be written as

$$p(Y_{obs}, D | \theta, \phi, X) = p(Y_{obs} | D, \theta, X) p(D | \phi, X), \quad [2]$$

where X are observed covariates. Note that these covariates influence missingness through the distribution $p(D | \phi, X)$. Well-known examples are found in the general class of stochastic censoring and self-selection models by Heckman (1976, 1979). Heckman initially developed the procedure for continuous data. Applications to categorical data are, for instance, found in Fay (1986), O'Muircheartaigh and Moustaki (1999), Moustaki and O'Muircheartaigh (2000), and Moustaki and Knott (2000). Usually, the covariates X provide the information needed to model the missing data indicators D . However, the missing data indicators D may also depend on latent variables which are not directly observed. In the next section, an example is given to get a flavor of the approach.

Example: Modeling Item Nonresponse in Educational Assessments

Educational assessments are often based on a test with dichotomously scored items. Data from such tests are analyzed using a so-called IRT model (see, e.g., Lord, 1980). An example is the one-parameter logistic model

(1PLM; Rasch, 1960). In the 1PLM, the probability of a correct response is given by

$$p(y_{ik} = 1 | \zeta_i, \beta_k) = \frac{\exp(\zeta_i - \beta_k)}{1 + \exp(\zeta_i - \beta_k)}$$

where ζ_i is the so-called latent ability parameter for student i , and β_k is the so-called difficulty parameter of item k . The likelihood function is obtained using eqn [1]. In most practical situations, the slightly more complex two- and three-parameter logistic models are used (Lord, 1980); however, the 1PLM suffices for the example presented here. In IRT, it is mostly assumed that the ability parameters ζ_i are random effects and that the item parameters β_k are fixed effects. For technical reasons (which are not important here), the likelihood function is integrated over the random effects and the fixed effect parameters are estimated by marginal ML (MML; see Bock and Aitkin, 1981).

In most applications of educational assessment, it can be assumed that the missing data are ignorable. As argued, this also holds for the two-stage testing design discussed above. However, there are situations where ignorability may be violated. For instance, when students skip items they find too difficult and this skipping depends on a latent variable that is highly correlated with the latent ability variable. For these situations, Holman and Glas (2005) formulate a model where the missingness indicators also follow the 1PLM, that is,

$$p(d_{ik} = 1 | \eta_i, \delta_k) = \frac{\exp(\eta_i - \delta_k)}{1 + \exp(\eta_i - \delta_k)}.$$

Note that η_i is a student parameter that is positively related to endorsing the item and δ_k is an item parameter that is negatively related to endorsing the item. Holman and Glas (2005) introduce the assumption that ζ_i and η_i have a bivariate normal distribution. The mean and variance parameters of this distribution are set equal to zero and one, respectively, to identify the location and scale of the latent variables. The correlation between ζ_i and η_i , say ρ , is a free parameter to be estimated. The complete model is given by

$$p(Y_{obs}, D | \theta, \phi) = p(Y_{obs} | D, \zeta, \beta) p(D | \eta, \delta) p(\zeta, \eta | \rho)$$

If $\rho \neq 0$, ignorability is violated because the priors of the parameters of the distribution of the observations $\theta = (\zeta, \beta)$ and those of the missing data indicators $\phi = (\eta, \delta)$ are dependent. The assumption of ignorability only holds if $\rho = 0$.

To show the impact of violation of ignorability on the estimates of the item parameters, a small part of a larger study by Holman and Glas (2005) is shown. Using the model, data were generated for 1000 simulated students, responding to ten items. To keep the presentation simple, an example will be given where all item parameters are

equal to zero, that is, $\beta_k = \delta_k = 0$ for all k . The latent variables ζ_i and η_i were drawn from a bivariate normal distribution with mean and variance equal to one and zero, respectively, and a correlation ρ . Using these parameters, missing data indicators d_{ik} and item responses y_{ik} were generated according to the model. The correlation ρ was varied from 0.0 to 0.9 in steps of 0.1. For every value of ρ , 100 replications were made, that is, 100 data matrices with 1000 simulees and ten items were generated and analyzed. MML estimates were computed for two models: a model labeled MAR where only the item parameters β were estimated and a model labeled NONMAR where the item parameters β , δ , and ρ were estimated. The results are shown in **Table 1**.

The column labeled MAR gives the mean absolute error (MAE) of the estimates of the item parameters under the MAR model averaged over all items and all replications. The true values of the item parameters are all zero; therefore, it can be seen that the MAE in the estimates of the item parameters under the MAR model increases with the magnitude of the violation of ignorability. That is, if there is no violation ($\rho = 0.0$), then the MAE is 0.091, and the severest violation ($\rho = 0.9$) produces an MAE of 0.273. The columns under the label NONMAR give the MAE of the estimates of parameters under the NONMAR model. Note that the MAE of the estimates of the item parameters does not increase with the magnitude of the violation of ignorability. The MAE of the estimate of the correlation between the two latent dimensions does increase. However, a breakdown of this MAE into bias and random estimation error showed that this increase was due to the higher standard errors for lower correlations. Therefore, the estimates of the correlations were not biased. The conclusion is that an analysis under an unjustified assumption of ignorability leads to biased estimates, but the bias is removed by modeling the process that caused the missing data.

Table 1 Mean absolute error of MML parameter estimates under MAR and NONMAR models using 1000 simulees, 10 items, and 100 replications for each value of ρ

ρ	Model				
	MAR		NONMAR		
	β	β	δ	ρ	
0.0	0.091	0.092	0.060	0.055	
0.1	0.075	0.075	0.066	0.048	
0.2	0.086	0.084	0.057	0.053	
0.3	0.121	0.087	0.065	0.058	
0.4	0.131	0.095	0.062	0.054	
0.5	0.152	0.084	0.057	0.037	
0.6	0.147	0.089	0.068	0.033	
0.7	0.186	0.089	0.068	0.034	
0.8	0.200	0.090	0.058	0.026	
0.9	0.273	0.088	0.054	0.010	

An application of an analogous approach is found in the work by Glas and Pimentel (2008). The application concerns an analysis of data from a subscale of the *Nederlandse Intelligentietest voor Onderwijsniveau* (NIO), which is an intelligence test for students in primary education in the Netherlands (van Dijk and Tellegen, 2004). The IRT model for the observed data was the 2PLM. The model for the missing data indicators was an IRT model by Tutz (1997). This model is a generalization of the 1PLM. MML estimates of the parameters were obtained both for the MAR model and the NONMAR model. The correlation between ζ and η was estimated as 0.429, which gives an indication that the missing data process cannot be ignored. The differences between the estimates obtained using the MAR and NONMAR models are plotted in **Figure 1**. The first panel shows how the number of students responding to an item decreases from 3145 to 508. The second panel gives the differences in the estimates of the difficulty parameters β under the two models. Note that the difference increases toward the end of the test. Therefore, the bias in the parameter estimates

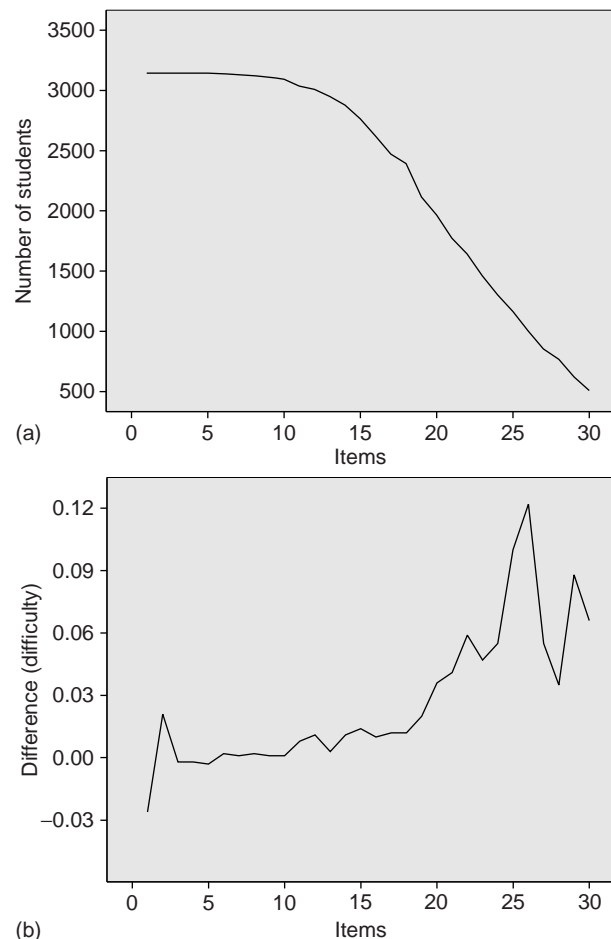


Figure 1 (a) Numbers of students reaching items and (b) the differences of item difficulty estimates between MAR and NONMAR models.

increases as the sample of students responding to an item becomes more and more select. Thus, here also, the conclusion is that the violation of ignorability resulted in biased parameter estimates.

See also: Analysis and Interpretation of Multivariate Data; Bootstrap Method; Educational Data Modeling; Evaluation Research; Exploratory Data Analysis; Jackknife Methods; Large-sample Statistical Methods; Markov Chain Monte Carlo; Monte Carlo Methods; Point Estimation Methods with Applications to Item Response Theory Models.

Bibliography

- Allison, P. D. (2000). Multiple imputation for missing data: A cautionary tale. *Sociological Methods Research* **28**, 301–309.
- Bock, R. D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **46**, 443–459.
- Buhi, E. R., Goodson, P., and Neilands, T. B. (2008). Out of sight, not out of mind: Strategies for handling missing data. *American Journal of Health Behavior* **32**, 83–92.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* **39**, 1–38.
- Fay, R. E. (1986). Causal models for patterns of nonresponse. *Journal of the American Statistical Association* **81**, 354–365.
- Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association* **85**, 972–985.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. London: Chapman and Hall.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.
- Glas, C. A. W. and Pimentel, J. L. (2008). Modeling nonignorable missing data in speeded tests. *Educational and Psychological Measurement* **68**, 907–922.
- Heckman, J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimator for such models. *Annals of Economic and Social Measurement* **5**, 475–492.
- Heckman, J. (1979). Sample selection bias as a specification error. *Econometrika* **46**, 931–961.
- Holman, R. and Glas, C. A. W. (2005). Modeling non-ignorable missing data mechanism with item response theory models. *British Journal of Mathematical and Statistical Psychology* **58**, 1–18.
- Kelly, A. (1976). A study on the comparability of external examinations in different subjects. *Research in Education* **1**, 37–63.
- Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum.
- Moustaki, I. and Knott, M. (2000). Weighting for item non-response in attitude scales by using latent variable models with covariates. *Journal of the Royal Statistical Society, A* **163**, 445–459.
- Moustaki, I. and O'Muircheartaigh, C. (2000). A one dimensional latent trait model to infer attitude from nonresponse for nominal data. *Statistica* **10**, 259–276.
- O'Muircheartaigh, C. and Moustaki, I. (1999). Symmetric pattern models: A latent variable approach to item non-response in attitude scales. *Journal of the Royal Statistical Society, Series A* **162**, 177–194.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
- Rubin, D. (1987). *Multiple Imputations for Nonresponse in Surveys*. New York: Wiley.
- Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association* **91**, 473–489.
- Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.
- Schafer, J. L. (1999). *NORM: Multiple Imputation of Incomplete Multivariate Data under a Normal Model*. University Park, PA: Pennsylvania State University, Department of Statistics.
- S-Plus 6 for Windows (computer software) Seattle, WA: Insightful Corporation.
- Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**, 528–540.
- Tutz, G. (1997). Sequential models for ordered responses. In van der Linden, W. J. and Hambleton, R. K. (eds.) *Handbook of Modern Item Response Theory*, pp 139–152. New York: Springer.
- van Dijk, H. and Tellegen, P. (2004). *NIO, Nederlandse Intelligentietest voor Onderwijsniveau*. (NIO, Dutch Intelligence Test for Educational Level. Amsterdam: Boom Test Uitgevers.
- Vermunt, J. K., van Ginkel, J. R., van der Ark, A. L., and Sijtsma, K. (2008). Multiple imputation of incomplete categorical data using latent class analysis. *Sociological Methodology* **38**(1), 369–397.
- Von Hippel, P. T. (2007). Regression with missing Ys: An improved strategy for analyzing multiple imputed data. *Sociological Methodology* **37**, 83–117.
- Wu, M. L., Adams, R. J., and Wilson, M. R. (1997). *ConQuest: Generalized Item Response Modelling Software*. Melbourne, VIC: Australian Council for Educational Research.
- Yuan, Y. C. (2000). Multiple imputation for missing data: Concepts and new development. *Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference (Paper No. 267)*, Cary, NC: SAS Institute. Retrieved 26 April 2006 from <http://www.ats.ucla.edu/stat/sas/library/multipleimputation.pdf>.
- Zimowski, M. F., Muraki, E., Mislevy, R. J., and Bock, R. D. (1996). *Bilog MG: Multiple-Group IRT Analysis and Test Maintenance for Binary Items*. Chicago, IL: Scientific Software International, Inc.

Further Reading

- Allison, P. D. (2001). *Missing Data*. Thousand Oaks, CA: Sage.
- McKnight, P. E., McKnight, K. M., Sidani, S., and Figueredo, A. J. (2007). *Missing Data: A Gentle Introduction*. New York: Guilford.

Relevant Websites

- <http://gking.harvard.edu> – Amelia II: A program for missing data.
- <http://www.lshrm.ac.uk> – Annotated bibliography.
- <http://www.stata.com> – Data analysis and statistical software.
- <http://www.isr.umich.edu> – IVEware: Imputation and variance estimation software.
- <http://www.vcu.edu> – Mx is a matrix algebra interpreter and numerical optimizer for structural equation modeling and other types of statistical modeling of data.
- <http://www.stat.psu.edu> – Software for multiple imputation.
- <http://www.multiple-imputation.com> – The Department of Statistics at TNO Quality of Life has been applying multiple imputation since 1990. The department specializes in imputation of multivariate social-medical and epidemiological data.
- <http://methcenter.psu.edu> – The Methodology Center is an interdisciplinary center that comprises faculty, research associates, post-docs, and students from several academic disciplines, including human development, psychology, statistics, and public health.
- www.missingdata.org.uk – This site exists to support researchers from the social and medical sciences with the analysis of incomplete datasets, and as a focus of statistical research.

Model Selection

S J Haberman, Educational Testing Service, Princeton, NJ, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bayesian inference – A method of statistical inference in which conclusions about populations are made by combining sample data with prior knowledge concerning population measures.

Continuous variable – A variable such as height or weight which can assume any value in an interval of positive length.

Discrete variable – A variable which assumes a finite or countable number of values.

Hypothesis test – A method of statistical inference in which a statistic based on observed data is used to assess consistency of observed data with a hypothesis concerning a population measure.

Item-response model – A model employed to describe the probability that an examinee in an assessment has any specific pattern of responses.

Likelihood-ratio test – A test of a hypothesis which compares the compatibility of data with model parameters consistent with a specified hypothesis to the compatibility of data with model parameters consistent with an alternative hypothesis.

Linear model – A model which uses a linear function of one or more explanatory variables to predict a response variable.

Multiple-choice test – An assessment in which examinees select an answer to each test item from a list of alternatives.

Pearson chi-square test – A hypothesis test used with discrete data to assess the compatibility of a probability model with the data.

Penalty function – A function used to measure prediction error.

Model selection is a common activity in statistical analysis of data. Two somewhat different approaches can be found. In the first approach, a model is regarded as appropriate for the data if, according to hypothesis tests, the model is consistent with the data. Among models consistent with the data, comparisons may be made in terms of parsimony or in terms of relative consistency with the data. In the second approach, models are compared in terms of their ability to predict data. It is not regarded as likely that any model under study is consistent with data. Instead, prediction measures are estimated and compared for the models under consideration. The second approach becomes

increasingly important as sample sizes become large, for inconsistency of models with data becomes an increasingly serious issue in such cases (Gilula and Haberman, 2001). In this article, the sections on likelihood-ratio tests, Pearson chi-square tests, and Bayesian inference emphasize the approach based on hypothesis tests. The sections on penalty criteria and linear models emphasize prediction measures.

To illustrate some common issues in model selection, consider selection of a model for the responses on a multiple-choice test. For illustrative purposes, data will be used from a writing test with $q = 45$ items numbered from 1 to q . Data are available for $n = 8686$ examinees i numbered from 1 to n (Haberman, 2005, 2007b). For each examinee i and item j , the observation $X_{ij} = 1$ if the response is correct and 0 if the response is not correct. In psychometrics, several common item-response models are often applied to data of this kind. In each item-response model, it is assumed that there are independently distributed unobserved random variables θ_i which represent the abilities of the examinees i , $1 \leq i \leq n$. As is typically the case in applications of these models, the unobserved variables are assumed to have a standard normal distribution. It is assumed that, given the θ_i , the X_{ij} are conditionally independent and, for a particular examinee i and item j , the conditional probability that $X_{ij} = 1$ (the response is correct) only depends on the ability variable θ_i associated with the examinee. The conditional probability that $X_{ij} = 1$ given $\theta_i = \omega$ is $P_j(\omega)$ for some unknown function P_j defined for all real numbers. The models considered are models for P_j . In the one-parameter logistic (1PL) model, it is assumed that, for some unknown constants a and γ_j

$$P_j(\omega) = \frac{\exp(a\omega - \gamma_j)}{1 + \exp(a\omega - \gamma_j)}$$

In the special case of $a = 0$, one obtains the independence model (Ind) that the X_{ij} , $1 \leq j \leq q$, are mutually independent and that the probability is positive that $X_{ij} = k$ for k equal 0 or 1. In the two-parameter logistic (2PL) model, it is assumed that, for some unknown slope parameter a_j and intercept parameter γ_j

$$P_j(\omega) = \frac{\exp(a_j\omega - \gamma_j)}{1 + \exp(a_j\omega - \gamma_j)}$$

In the three-parameter logistic (3PL) model, it is assumed that, for some unknown slope parameter a_j , some

unknown intercept parameter γ_j and some unknown lower asymptote c_j $0 \leq c_j < 1$,

$$P_j(\omega) = c_j + (1 - c_j) \frac{\exp(a_j\omega - \gamma_j)}{1 + \exp(a_j\omega - \gamma_j)}$$

The 1PL model is a special case of the 2PL model in which the slope parameters a_j have a common value a , and the 2PL model is a special case of the 3PL model in which the lower asymptotes c_j are all 0. The basic question to consider is how to decide which of these models is most appropriate for the data under study. One approach considered will be based on tests of hypotheses. The other approach will be based on the ability of the models to describe the joint distribution of the item responses X_{ij} $1 \leq j \leq q$, associated with an examinee i (Hambleton *et al.*, 1991).

Likelihood-Ratio Tests

The models under study may be compared by the use of likelihood-ratio tests (Wilks, 1938). For any vector \mathbf{x} of possible responses x_j equal 0 or 1, $1 \leq j \leq q$, let $p(\mathbf{x})$ be the joint probability that, for some examinee i , the response X_{ij} to item j is x_j for $1 \leq j \leq q$. The distribution of the vector \mathbf{X}_i of responses X_{ij} $1 \leq j \leq q$, is determined by the q -dimensional array \mathbf{p} of probabilities $p(\mathbf{x})$, $0 \leq x_j \leq 1$, $1 \leq j \leq q$. To each model corresponds a set S of possible arrays of probabilities, so that the model holds if, and only if, \mathbf{p} is in S . For example, the independence model holds if, and only if, \mathbf{p} is in the set $S(0)$ of arrays such that

$$p(\mathbf{x}) = \prod_{j=1}^q \frac{\exp(x_j \gamma_j)}{1 + \exp(\gamma_j)}$$

for some real γ_j $1 \leq j \leq q$. In like fashion, if θ is a random variable with a standard normal distribution, then the 1PL model holds if, and only if, \mathbf{p} is in the set $S(1)$ of arrays such that $p(\mathbf{x})$ is the expected value of

$$\prod_{j=1}^q \frac{\exp[x_j(a\theta - \gamma_j)]}{1 + \exp(a\theta - \gamma_j)}$$

for some real a and real γ_j $1 \leq j \leq q$. In like manner, a set $S(2)$ may be defined to correspond to the 2PL model, and a set $S(3)$ may be defined to correspond to the 3PL model. The saturated, or unrestricted, model holds for \mathbf{p} in the set T of arrays with $p(\mathbf{x})$ positive and the sum of all $p(\mathbf{x})$ equal to 1.

The log likelihood function

$$\ell(\mathbf{p}) = \sum_{i=1}^n \log p(\mathbf{X}_i)$$

is the logarithm of the probability that the observed vectors \mathbf{X}_i are obtained for $1 \leq i \leq n$. For a model that \mathbf{p} is in S , likelihood-ratio tests are based on the maximum

$\ell(S)$ of the log likelihood for probability arrays in S . For a subset S' of S , the model that \mathbf{p} is in S' may be compared to a model that \mathbf{p} is in S by the likelihood-ratio chi-square statistic

$$L^2(S, S') = 2[\ell(S) - \ell(S')]$$

In typical cases, if \mathbf{p} is in S' , then $L^2(S, S')$ has an approximate chi-square distribution with $k(S) - k(S')$ degrees of freedom, where $k(S)$ is the dimension of the space S and $k(S')$ is the dimension of S' . Thus, the approximate significance level is the probability $P(S, S')$ that a chi-square variable with $k(S) - k(S')$ degrees of freedom exceeds the observed value of $L^2(S, S')$. The approximation improves as the sample size n increases. If the model that \mathbf{p} is in S' does not hold, then the typical result is that $P(S, S')$ converges to 0 with probability 1 as the sample size increases.

In the comparison of the 1PL to the 2PL model for the data under study, the set $S(1)$ corresponding to the 1PL model has dimension $q + 1$, for there are q unrestricted parameters γ_j and one unrestricted parameter a . The set $S(2)$ corresponding to the 2PL model has dimension $2q$, for there are q unrestricted parameters γ_j and q unrestricted parameters a_j . The resulting chi-square approximation has $2q - (q + 1) = q - 1 = 44$ degrees of freedom in the case under study. The statistic $L^2(S(2), S(1)) = 3772$, so that the approximate significance level is extremely close to 0. In other words, either the 1PL model does not hold, or a very rare event has taken place. In like fashion, one may readily compare the independence model to the 1PL model, for $S(0)$ has q unrestricted parameters. Thus, $k(S(1)) - k(S(0))$ is $q + 1 - q = 1$, and $L^2(S(1), S(0)) = 22109$, so that the approximate significance level is extremely small.

One might expect that the comparison of the 2PL model to the 3PL model would proceed just as the comparison of the 1PL model to the 2PL model and the comparison of the independence model to the 1PL model, but this expectation is not met. The likelihood-ratio chi-square statistic $L^2(S(3), S(2)) = 646$ is readily found, but the chi-square approximation does not hold due to the constraints on the lower asymptote parameters c_j . Under the 2PL model, c_j is at the boundary value 0 of its range. This constraint typically results in a large-sample distribution that is less likely to exceed any given value $x > 0$ than is the chi-square distribution with $k(S(3)) - k(S(2)) = q$ degrees of freedom. Because q is 45, it remains evident that the 2PL model does not hold.

In principle, one could compare the model that \mathbf{p} is in $S(m)$, m equal 0, 1, 2, or 3 to the saturated model that \mathbf{p} is in T , but the chi-square approximation again fails. The problem arises because T has dimension $2^q - 1 = 3.51 \times 10^{13}$, a very large number relative to the sample size n . For use of $L^2(T, S(m))$, it is commonly recommended that the expected number $np(\mathbf{x})$ of \mathbf{X}_i equal \mathbf{x} be at least 1 for all possible values \mathbf{x} of \mathbf{X}_i and that $np(\mathbf{x})$ should be at least 5

for at least 80% of the possible \mathbf{x} (Cochran, 1954). Use of T can be illustrated if one confines attention to the first four responses X_{ij} $1 \leq j \leq 4$, for in this case, the conditions for the chi-square approximation hold. For the independence model, $L^2(T, S(0))$ is 410.7, and there are $2^4 - 1 - 4 = 11$ degrees of freedom, so that $P(T, S(0))$ is very small. Similarly, $L^2(T, S(1)) = 289.1$, and there are $2^4 - 4 = 12$ degrees of freedom for the chi-square approximation, so that $P(T, S(1))$ is very close to 0. Thus the independence and 1PL model for the first four responses do not appear to be compatible with the data. In the case of the 2PL model, $L^2(T, S(2))$ is 20.6, and $2^4 - 2(4) - 1 = 7$ degrees of freedom are associated with the chi-square approximation, so that $P(T, S(2))$ is about 0.0044. This probability is sufficiently low that evidence exists that the 2PL model does not hold for the first four responses; however, the evidence is much less substantial than in the case of the independence or 1PL model.

A convenient feature of likelihood-ratio chi-square statistics is the existence of simple decompositions. For example,

$$L^2(S, S') = L^2(T, S') - L^2(T, S)$$

For instance, in the example with the four initial responses, $L^2(S(2), S(1))$, the comparison statistic for the 1PL model against the 2PL model, is $L^2(T, S(1)) - L^2(T, S(2)) = 268.5$, and the chi-square approximation has $12 - 7 = 5$ degrees of freedom. In this particular example, this comparison provides very strong evidence against the 1PL model. This result is consistent with the previous finding for the test statistic $L^2(T, S(1))$ comparing the 1PL model to the saturated model.

Likelihood-ratio tests may be employed for comparisons of more than two models. Let S be a family of subsets of T . For each S in this family, one may consider the model that \mathbf{p} is in S . In the simplest case, consider a base model that \mathbf{p} is in U , where U includes every set S in the family S but U itself is not a member of the family of models to be compared. One may then consider comparison of the model that \mathbf{p} is in S to the model that \mathbf{p} is in U by means of the likelihood-ratio chi-square $L^2(U, S)$ and the corresponding approximate significance level $P(U, S)$. For instance, in the four-response case, let U be T , so that the base model is the saturated model, and let S contain $S(m)$ for m from 0 to 2, so that the independence model, the 1PL model, and the 2PL model are compared to the saturated model. Preference is given to 2PL model that \mathbf{p} is in $S(2)$, for $P(U, S(2))$ is larger than $P(U, S(0))$ or $P(U, S(1))$. A related criterion based on ratios of likelihood-ratio chi-square statistics to approximate expected values is the F statistic $F(U, S) = L^2(U, S)/[k(U) - k(S)]$. The smallest $F(U, S)$ is sought for S in S (Goodman, 1971; Haberman, 1974). For the four-response example, the best result is again for the 2PL model, for $F(T, S(2)) = 2.9$ is much smaller than $F(T, S(0))$ or $F(T, S(1))$.

Pearson Chi-Square Tests

Although less convenient in the application under study, it should be noted that the Pearson chi-square statistic can be used instead of the likelihood-ratio chi-square (Pearson, 1900; Fisher, 1924). This option typically uses maximum-likelihood estimates to compare the model that \mathbf{p} is in S' to the model that \mathbf{p} is in S under the condition that S' is included in S . The array $\hat{\mathbf{p}}_S$ of probabilities $\hat{p}_S(\mathbf{x})$ is the maximum-likelihood estimate of \mathbf{p} under the model that \mathbf{p} is in S if $\hat{\mathbf{p}}$ is in S and $\ell(\hat{\mathbf{p}}_S) = \ell(S)$. In the case of the unrestricted model that \mathbf{p} is in T , $\hat{\mathbf{p}}_T = \bar{\mathbf{p}}$, the array of fractions $\bar{p}(\mathbf{x})$ of examinees i with $\mathbf{X}_i = \mathbf{x}$. One has

$$X^2(S, S') = n \sum_{\mathbf{x}} [\hat{p}_S(\mathbf{x}) - \hat{p}_{S'}(\mathbf{x})]^2 / \hat{p}_{S'}(\mathbf{x})$$

(Haberman, 1977). The large-sample approximations for $X^2(S, S')$ and $L^2(S, S')$ are the same. The most common applications use $S = T$. For instance, in the four-question case, to test the independence model that \mathbf{p} is in $S(0)$, one can use $X^2(T, S(0)) = 409.0$ in place of $L^2(T, S(0)) = 410.7$ without appreciable effect.

Whether Pearson or likelihood-ratio chi-square statistics are employed, a typical problem in model selection is that no model of any interest is compatible with the data if the sample size n is large (Gilula and Haberman, 2001).

Continuous Data

Although the working example involves discrete variables, it should be emphasized that likelihood-ratio chi-square statistics are also commonly applied to continuous data to examine compatibility of models with data. The basic difference is that in the log likelihood, probabilities are replaced by probability densities. The likelihood function continues to be maximized for each model compared, chi-square approximations remain, and computations of degrees of freedom are unchanged.

Penalty Criteria

An alternative approach to model selection compares models in terms of predictive power rather than in terms of fitting data. This approach often employs information theory (Gilula and Haberman, 2001). One may illustrate arguments by using the results of the 45-item multiple-choice test previously employed to illustrate use of tests of significance. In the approach under study, probability prediction is employed. A probability array \mathbf{u} is used to predict the observation \mathbf{X}_i . The value of the array as a predictor is judged by use of the penalty $-\log u(\mathbf{X}_i)$ (Savage, 1971). This penalty only depends on the probability assigned to the observed value of the responses \mathbf{X}_i .

and the penalty is non-negative. The penalty is 0 if, and only if, the observed responses \mathbf{X}_i have assigned probability 1. The expected penalty is then

$$G(\mathbf{u}) = E(-\log u(X_i))$$

The expected penalty is never less than the entropy

$$H = G(\mathbf{p}) = E(-\log p_{X_i})$$

and $G(\mathbf{u}) = H(\mathbf{p})$ if, and only if, the probability prediction \mathbf{u} is the actual probability array \mathbf{p} .

The penalty criterion may be employed to assess the quality of a model that \mathbf{p} is in S . If the probability array \mathbf{p} is known, then model error may be evaluated by use of the minimum $G(S)$ of G for probability arrays in S . The difference $G(S) - H$ is a measure of model error. The measure is non-negative, and, if \mathbf{p} is in S , then $G(S) = H$ and the error $G(S) - H = 0$. To compare models that \mathbf{p} is in S for S in a family \mathcal{S} of subsets of T , the smallest $G(S)$ is sought for S in \mathcal{S} .

In practice, \mathbf{p} is not known, so that $G(S)$ and H must be estimated. Estimation of $G(S)$ is generally feasible given the log-likelihood maximum $\ell(S)$, for $G(S)$ may be estimated by $\hat{G}(S) = -\ell(S)/n$. In principle, H has estimate $\hat{H} = -\ell(T)/n$; however, this estimate is not satisfactory unless conditions hold for the chi-square approximation to the likelihood-ratio chi-square statistic $L^2(T, S)$. For S' contained in S , $\hat{G}(S') - \hat{G}(S)$ is $L^2(S, S')/(2n)$, and $\hat{G}(S') - \hat{G}(S)$ can be employed to measure the improvement in the model that \mathbf{p} is in S relative to the model that \mathbf{p} is in S' . For example, for the case of the data with 45 items, $\hat{G}(S(0))$ is 28.110, $\hat{G}(S(1))$ is 26.838, $\hat{G}(S(2))$ is 26.621, and $\hat{G}(S(3))$ is 26.583. Despite the very strong evidence that the 1PL model does not hold, the measured difference $\hat{G}(S(1)) - \hat{G}(S(2))$ is relatively small, only 0.217. In contrast, $\hat{G}(S(0)) - \hat{G}(S(1))$ is 1.273. The relative reduction in estimated expected penalty

$$\frac{\hat{G}(S(1)) - \hat{G}(S(2))}{\hat{G}(S(1))} = 0.0081$$

for the 2PL model relative to the 1PL model is quite small compared to the relative reduction

$$\frac{\hat{G}(S(0)) - \hat{G}(S(1))}{\hat{G}(S(0))} = 0.0453$$

for the 1PL model relative to the independence model. The contrast between the 2PL and the 3PL model demonstrates even less difference, for $\hat{G}(S(2)) - \hat{G}(S(3))$ is only 0.037, and

$$\frac{\hat{G}(S(2)) - \hat{G}(S(3))}{\hat{G}(S(2))} = 0.0014$$

is very modest in size.

The estimate $\hat{G}(S)$ is biased. In addition, if one wishes to consider prediction of a new observation \mathbf{X}_{n+1} with the same distribution as the \mathbf{X}_i , $1 \leq i \leq n$, and one wishes to

base the prediction on estimates computed from \mathbf{X}_i , $1 \leq i \leq n$, then further correction is needed. One measure for prediction of \mathbf{X}_{n+1} is the Akaike estimate

$$G_a(S) = \hat{G}(S) + k(S)/n$$

This correction is only satisfactory if the model that \mathbf{p} is in S is valid. An alternative approach (Gilula and Haberman, 2001) uses a criterion $G_g(S)$ similar to the Akaike criterion that does not assume that the model holds. It is also possible to consider $G_c(S) = \hat{G}(S) + bk(S)$ for a positive real constant b , so that a penalty for model complexity is imposed without regard to sample size.

The Bayesian analysis in the next section leads to the Schwarz criterion

$$G_s(S) = \hat{G}(S) + (2n)^{-1}k(S) \log n$$

In large samples, $\hat{G}(S)$ and the criteria of Akaike, Schwarz, and Gilula and Haberman are very similar. For instance, for the 2PL model for 45 items, $G_a(S(2))$ and $G_g(S(2))$ are 26.631 to five significant figures, while $G_s(S(2))$ is 26.668. The data suggest that the 3PL model is slightly preferable to the 2PL model unless $G_c(S(2))$ is compared to $G_c(S(3))$ with the coefficient b at least 0.00083. The advantage of the 2PL model relative to the 1PL model is larger under all criteria, and $G_c(S(1))$ is greater than $G_c(S(2))$ unless b is at least 0.00482.

In the case of continuous data, similar arguments can often be used to those which have been employed for discrete data (Gilula and Haberman, 2000); however, as in the case of likelihood-ratio tests, arguments are based on density and functions rather than on probabilities.

Bayesian Inference

Bayesian inference can be employed to select models (Kass and Raftery, 1995). For example, consider the data with 45 items and S equal to the set of $S(m)$ for m from 0 to 3. Assume that the 3PL model does hold and that comparison is between the independence model, the 1PL model, the 2PL model, and the 3PL model. One may divide the set $S(3)$ corresponding to the 3PL model into four disjoint sets $U(0)$, $U(1)$, $U(2)$, and $U(3)$. Here $U(0)$ is $S(0)$ and, for k equal 1, 2, or 3, $U(k)$ consists of members of $S(k)$ that are not in $S(k-1)$. Thus, $U(0)$ corresponds to the independence model, while $U(1)$ corresponds to case in which the 1PL model holds but the independence model holds. Let the prior probability that \mathbf{p} is in $U(k)$ be $\pi_k > 0$ for $0 \leq k \leq 3$, and let Π_k be the prior distribution of \mathbf{p} given that \mathbf{p} is in $U(k)$. In typical cases, each Π_k is associated with some positive probability densities on the set of model parameters which correspond to \mathbf{p} in $U(k)$. For \mathbf{p} in $U(0)$, a positive density is assigned to the q -dimensional vector with coordinates γ_j for $1 \leq j \leq q$. Similarly, for \mathbf{p} in

$U(1)$, a positive density is assigned to the $(q+1)$ -dimensional vector with coordinates γ_j , $1 \leq j \leq q$, and $a, a \neq 0$. Similar requirements are imposed for \mathbf{p} in $U(2)$ or for \mathbf{p} in $U(3)$. Given these prior distributions, Bayes's theorem can be employed to compute the conditional probability p_k that \mathbf{p} is in $U(k)$ given the observed \mathbf{X}_i , $1 \leq i \leq n$. This conditional probability depends on the log-likelihood function $\ell(\mathbf{u})$ for \mathbf{u} in $S(3)$. In this case, large values of p_k suggest use of the model that \mathbf{p} is in $S(k)$.

Two practical obstacles exist. The value p_k depends on prior distributions that may be difficult to specify in a realistic manner. Computation of p_k is very difficult, although numerical quadrature, simulation methods, and approaches based on Laplace transforms can be considered. A very crude approximation can often be based on the Schwarz criterion $G_s(S(k))$ for $0 \leq k \leq 3$. Under relatively general conditions

$$\frac{n[G_s(S(g)) - G_s(S(k))] - \log(p_k/p_g)}{\log(p_k/p_g)}$$

converges in probability to 0 as the sample size n increases. This result may be used to show that p_k converges in probability to 1 if \mathbf{p} is in $U(k)$. In the example under study, the result favors the 3PL model to the extent that $G_s(S(k))$ is smallest for $k = 3$.

The Bayesian approach described here does not have any obvious basis if \mathbf{p} is not contained in $S(3)$. This restriction is severe if one takes the position that nontrivial models are rarely valid.

Linear Models

Model selection need not be based on complete probability models. A common case involves linear models. For example, consider a sample of n essays, for each of which both holistic scores generated by two human raters and computer-generated essay features are available (Haberman, 2007a). For essay i , let Y_i be the average of the two holistic scores, and let X_{ij} be the numerical value of essay feature j , $1 \leq j \leq k$. For example, X_{i1} might be the square root of the number of characters in essay i . For each essay i , the computer-generated features X_{ij} , $1 \leq j \leq k$, are to be employed to predict the average holistic score Y_i . A simple approach considers a linear predictor

$$\tilde{Y}_i = \alpha + \sum_{j=1}^k \beta_j X_{ij}$$

for some real constants α and β_j , $1 \leq j \leq k$. The prediction error is $Y_i - \tilde{Y}_i$. If \mathbf{X}_i is the vector with coordinates X_{ij} , $1 \leq j \leq k$, then one may consider the case of pairs (Y_i, \mathbf{X}_i) that are independently and identically distributed. Let Y_i have finite and positive variance $V(Y)$, and let \mathbf{X}_i have finite and positive-definite covariance matrix $\text{Cov}(\mathbf{X})$.

The expected mean-square error is then the expected value $E([Y_i - \tilde{Y}_i]^2)$. If the mean $E(Y)$ of the Y_i , the mean $E(\mathbf{X})$ of the \mathbf{X}_i , the variance $V(Y)$ of the Y_i , the covariance matrix $\text{Cov}(\mathbf{X})$, and the covariance vector $\text{Cov}(\mathbf{X}, Y)$ with elements $\text{Cov}(X_{ij}, Y_i)$, $1 \leq j \leq k$, are all known, then the expected mean-squared error $E([Y_i - \tilde{Y}_i]^2)$ is minimized if

$$\tilde{Y}_i = E(Y) + \beta'[\mathbf{X}_i - E(\mathbf{X})]$$

where β , the vector of β_j , $1 \leq j \leq k$, is $[\text{Cov}(\mathbf{X})]^{-1} \text{Cov}(\mathbf{X}, Y)$. The resulting mean square error, denoted by $V(Y|\mathbf{X})$, is the residual variance of Y_i given \mathbf{X}_i . In principle, $V(Y|\mathbf{X})$ can form the basis for evaluation of the value of \mathbf{X}_i in prediction of Y_i .

In practice, means, variances, and covariances must be estimated. Let

$$\bar{X} = n^{-1} \sum_{i=1}^n X_i$$

denote the sample mean of the X_i , and let

$$\bar{Y} = n^{-1} \sum_{i=1}^n Y_i$$

denote the sample mean of the Y_i . The least-squares estimates a of α and \mathbf{b} of β satisfy the equations

$$\begin{aligned} \hat{Y}_i &= a + \sum_{j=1}^k b_j X_{ij} \\ a &= \bar{Y} - \mathbf{b}'\bar{\mathbf{X}} \end{aligned}$$

and

$$\sum_{i=1}^n (Y_i - \hat{Y}_i) X_i = 0$$

Here \mathbf{b} has coordinates b_j , $1 \leq j \leq k$, and $\mathbf{0}$ is the k -dimensional vector with all coordinates 0. As the sample size n becomes large,

$$\hat{V}(Y|\mathbf{X}) = n^{-1} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

approaches $V(Y|\mathbf{X})$ with probability 1, so that $\hat{V}(Y|\mathbf{X})$ may be employed to evaluate the effectiveness of \mathbf{X}_i in linear prediction of Y_i . Thus, $\hat{V}(Y|\mathbf{X})$ can provide a basis for selection of different predictors \mathbf{X} of Y .

Corrections are commonly made to treat biases due to the finite sample size (Draper and Smith, 1998). In standard regression analysis, the assumption is added that, given \mathbf{X}_i , the conditional mean of Y_i is \tilde{Y}_i and the conditional variance of Y_i is $V(Y|\mathbf{X})$. In this case, if \mathbf{X} is continuous and $n > k + 1$, then $V(Y|\mathbf{X})$ has the unbiased estimate

$$\bar{V}(Y|\mathbf{X}) = \frac{n}{n-k-1} \hat{V}(Y|\mathbf{X})$$

These assumptions do not hold exactly in the essay example, for the X_{ij} are discrete variables and the traditional

assumptions of regression are not likely to hold exactly given that holistic scores have a limited range, which in the data to be examined is from 1 to 6.

An alternative approach with fewer assumptions uses PRESS residuals (Draper and Smith, 1981; 325–327). The idea is to predict a new average holistic score Y_{n+1} by using the observed (Y_i, \mathbf{X}_i) , $1 \leq i \leq n$, and the new computer-generated features $X_{(n+1)j}$, $1 \leq j \leq n$. Under traditional regression analysis, the mean squared error from prediction of Y_{n+1} by $\hat{Y}_{n+1} = a + \mathbf{b}'\mathbf{X}_i$ is $V(Y|\mathbf{X})^{\frac{n+k+1}{n}}$, and this mean square error has unbiased estimate $\bar{V}(Y|\mathbf{X})^{\frac{n+k+1}{n}}$. Without traditional assumptions, one may employ PRESS residuals. For each essay i , least-squares estimates a_i for α and b_{ji} , $1 \leq j \leq k$, for β_j are obtained by minimizing the sum of the $(Y_m - \hat{Y}_{mi})^2$ for $1 \leq m \leq n$, $m \neq i$, where

$$\hat{Y}_{mi} = a_i + \sum_{j=1}^k b_{ji} X_{mj}$$

This approach estimates this mean squared error for prediction of Y_{n+1} by

$$\hat{V}_p(Y|\mathbf{X}) = n^{-1} \sum_{i=1}^n (Y_i - \hat{Y}_{ii})^2$$

where

$$\hat{Y}_{ii} = a_i + \sum_{j=1}^k b_{ji} X_{ij}$$

This approach might appear tedious; however,

$$Y_i - \hat{Y}_{ii} = (Y_i - \hat{Y}_i) / (1 - \bar{H}_{ii})$$

for

$$\hat{H}_{ii} = n^{-1} + (\mathbf{X}_i - \bar{\mathbf{X}})' \left[\sum_{m=1}^n (\mathbf{X}_m - \bar{\mathbf{X}})(\mathbf{X}_m - \bar{\mathbf{X}})' \right]^{-1} (\mathbf{X}_i - \bar{\mathbf{X}})$$

(Draper and Smith, 1998; 207–208) It is clearly desirable to have $\hat{V}_p(Y|\mathbf{X})$ smaller.

For one case with 4895 essays and $k = 6$ predictors, one has $\hat{V}_p(Y|\mathbf{X}) = 0.238$. An alternative predictor was considered with $k = 56$ predictors. This case had $\hat{V}_p(Y|\mathbf{X}) = 0.235$, so that a very modest gain was achieved with a much larger set of predictors. A further very modest gain was achieved with 206 predictors, for $\hat{V}_p(Y|\mathbf{X})$ went to 0.232. Whether the modest gain is really worth the cost in terms of predictors needed is a matter of judgment.

Concluding Remarks

Model selection issues are much more varied than can be indicated in a brief survey. The emphasis in this article has

been on simple random sampling. Modifications for more complex sampling are readily accomplished. Although the emphasis here has been on independent observations, model selection is also an issue in the analysis of time series in which observations are dependent. Except for the discussion of linear models, ability of independent variables to predict dependent variables has not been explored. Use of models which involve such prediction arises for polytomous dependent variables as well as for linear models (Gilula and Haberman, 2001). The emphasis in this article has involved situations in which the sample sizes are relatively large and the number of models to consider is relatively small. When sample sizes are not very large and when the number of models is large, then issues of selection bias must be considered. In these cases, a model is much more likely to appear to be the most appropriate due to random variation rather than due to actual superiority of the model. This issue is frequently encountered in regression analysis (Draper and Smith, 1998). In general, model selection depends on the nature of the data, the sample size, and the intended application of the results. The approach must be adapted to the problem at hand.

See also: Categorical Data Analysis; Educational Data Modeling; Goodness-of-Fit Testing; Univariate Linear Regression.

Bibliography

- Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics* **10**, 417–451.
- Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis*, 2nd edn. New York: Wiley.
- Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis*, 3rd edn. New York: Wiley.
- Fisher, R. A. (1924). The conditions under which χ^2 measures the discrepancy between observations and hypothesis. *Journal of the Royal Statistical Society* **87**, 442–450.
- Gilula, Z. and Haberman, S. J. (2000). Density approximation by summary statistics: An information-theoretic approach. *Scandinavian Journal of Statistics* **27**, 521–534.
- Gilula, Z. and Haberman, S. J. (2001). Analysis of categorical response profiles by informative summaries. *Sociological Methodology* **31**, 129–187.
- Goodman, L. A. (1971). The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics* **13**, 33–61.
- Haberman, S. J. (1974). Log-linear models for frequency tables with ordered classifications. *Biometrics* **30**, 589–600.
- Haberman, S. J. (1977). Log-linear models and frequency tables with small expected cell counts. *Annals of Statistics* **5**, 1148–1169.
- Haberman, S. J. (2005). *Latent-Class Item Response Models*. (Research Report RR-05-28). Princeton, NJ: Educational Testing Service.
- Haberman, S. J. (2007a). Electronic essay grading. In Rao, C. R. and Sinharay, S. (eds.) *Handbook of Statistics*, vol. 26, pp 205–233. Amsterdam: North-Holland.

- Haberman, S. J. (2007b). *The Information a Test Provides on an Ability Parameter*. (Research Report RR-07-18). Princeton, NJ: Educational Testing Service.
- Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Kass, R. and Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773–795.
- Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine* **50**, 157–175.
- Savage, L. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association* **66**, 783–801.
- Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics* **9**, 60–62.

Monte Carlo Methods

A M Johansen, University of Warwick, Coventry, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bayesian – An approach to statistical inference in which parameters are treated as random variables, allowing Bayes rule $p(y|x) = p(x|y)p(y)/p(x)$ to be used to perform inference. It is named after Reverend Thomas Bayes.

Bias – An estimator is biased if its expectation differs from the true value; the bias of the estimator is the difference between its expectation and the truth.

Consistency – A sequence of estimators is consistent if it converges, as the number of samples approaches infinity, to the true value. Such a sequence is asymptotically unbiased.

Expectation – The expectation (or expected value) of a random variable is, in a sense, the mean value that an infinite ensemble of independent replicates would have. In some sense, it quantifies the value which one would expect a realization to have, but there is no requirement that the expectation is a value which the random variable can take: for example, a Bernoulli random variable with parameter p can take values 0 and 1, but has expectation p .

iid (independent and identically distributed) – A collection of random variables is said to be iid if those variables are independent and share a common distribution. That is, each may be considered to be a single realization of the same probability distribution.

Independent – Random variables are independent if knowledge of one provides no information about the others. Formally, one can require that the conditional distribution of any one of the random variables given the other is the marginal distribution of that random variable, that is, $p_{Y|X}(y|x) = p_Y(y)$ if X and Y are independent.

Markov Chain – A temporally ordered collection of random variables with the property that conditional upon the present, the future is independent of the past.

Background and History

Monte Carlo methods may be thought of as a collection of computational techniques for the (usually approximate)

solution of mathematical problems, which make fundamental use of random samples. Two classes of statistical problems are most commonly addressed within this framework: integration and optimization. This article concentrates on the former: it is the (approximate) calculation of integrals using collections of random samples that people usually think of when they refer to the Monte Carlo Method. Monte Carlo methodology is also widely used in the simulation of physical, chemical and biological systems.

In the field of education, Monte Carlo methods are most interesting as a computational device for performing statistical inference. Many interesting models have extremely complex structures and cannot easily be dealt with using traditional techniques. Within the Bayesian paradigm, all information upon which inference can be based is encoded within the posterior probability distribution. Using Monte Carlo methods, we are able to characterize these distributions, and calculate expectations under them: the primary inferential technique.

A Simple Example

Monte Carlo methods invert the usual problem of statistics: rather than estimating random quantities in a deterministic manner, random quantities are employed to provide estimates of deterministic quantities. For example, one simple Monte Carlo experiment considers rain which falls uniformly at random (i.e., the location of any raindrop may be interpreted as a realization of a uniformly distributed random variable) over some square region of space, and a circle inscribed within that square. Without reference to any formal probability theory, it is intuitive that the probability of a uniform raindrop falling in any region within the square must be proportional to the area of that region and independent of its location. Consequently, the probability, p , that a raindrop lies within the inscribed circle may be expressed in terms of their areas. If the square has sides of length $2r$, the circle must be of radius r and

$$p = \frac{\pi r^2}{(2r)^2} = \frac{\pi}{4}$$

In itself, this may not seem particularly interesting. However, having expressed π as a function of this probability, its estimators can be used to approximate π . Proceeding analytically is not possible: obtaining the probability requires knowledge of π . Intuitively, it is possible to estimate this probability by counting the proportion of

raindrops which lie within the circle: if n raindrops are observed and m of those lie within the circle, then one may estimate p , using $\hat{p} = m/n$. **Figure 1** shows a computer simulation in which 500 raindrops were distributed uniformly over a square – in this case, $\hat{p} = 383/500$ and, defining $\hat{\pi}$ through the relationship between p and π , $\hat{\pi} = 383/125 = 3.06$. A poor estimate of π , considering the computational effort used, but it is an estimate nonetheless. In fact m is a realization of a binomial (n, p) random variable.

History

While there is some debate about the nature of the first Monte Carlo computations ever carried out (with some authors arguing that they date back as far as the times of ancient Babylon), it is generally agreed that the first modern Monte Carlo experiments were carried out in the later decades of the nineteenth century and, similar to the above example, were concerned with the estimation of π .

Consider dropping a needle of length l uniformly onto an array of parallel lines separated from one another by a distance, $d > l$. **Figure 2** illustrates this. Originally, the question was: What is the probability that the needle intersects a line? This question is relatively easy to answer. Assume that the array is large enough that edge effects are negligible, and assume that the x and y coordinates of the needles center are uniform over the array and the orientation of the needle is uniformly distributed over the interval $[0, \pi)$.

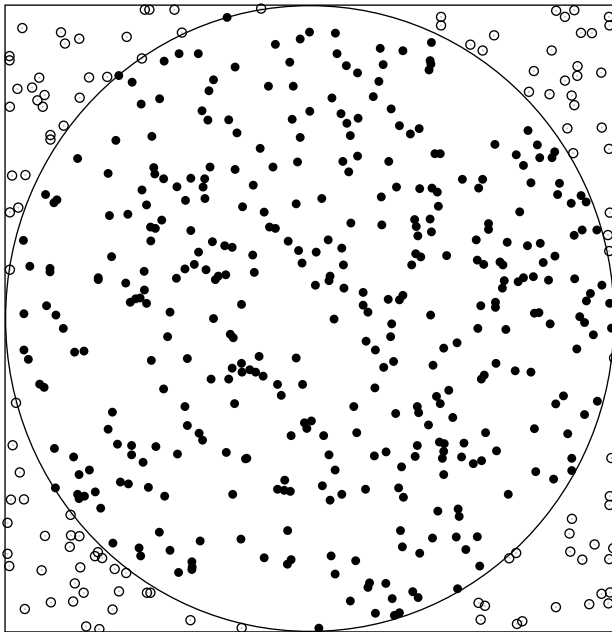


Figure 1 Estimating π : 383 of the 500 samples lie within the circle.

As **Figure 2** illustrates, the needle intersecting a line is equivalent to a rectangle of width $l \sin \theta$ intersecting that line. This happens, for given θ , with probability

$$\mathbb{P}(\text{intersection}|\theta) = \frac{l \sin \theta}{d}$$

The angle θ is uniform on $[0, \pi)$; hence,

$$\begin{aligned} \mathbb{P}(\text{intersection}) &= \int_0^\pi \mathbb{P}(\text{intersection}|\theta) \frac{1}{\pi} d\theta \\ &= \frac{l}{\pi d} \int_0^\pi \sin \theta d\theta = \frac{2l}{\pi d} \end{aligned}$$

Consequently, if n needles are dropped, then the average number which would intersect a line if the experiment were carried out many times is $2nl/d\pi$.

In a Monte Carlo setting, this means that π may be estimated using:

$$\hat{\pi} = \frac{2nl}{Md}$$

where M is the number of needles crossing a line when n are dropped. In 1901, Mario Lazzarini reported carrying out this experiment using a 2.5-cm-long needle and lines separated by $d = 3$ cm. Of his 3408 needles, 1808 crossed a line, suggesting

$$\pi \approx \frac{2 \times 3408 \times 2.5}{1808 \times 3} = \frac{355}{113}$$

which, remarkably, corresponds to the best rational approximation to π with denominator below 16 000, leading to suggestions that this was an early example of fabricated data.

Although these early examples of Monte Carlo techniques were, in some sense, successful, the cost of carrying out the experiments rendered them largely impractical and of only specialist interest. This began to change in Los Alamos in the 1940s, when physicists working on particle transport problems began solving them using something

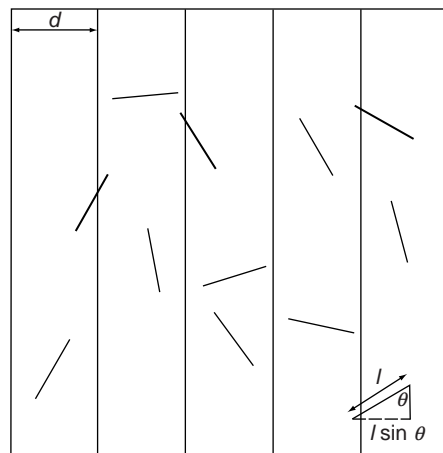


Figure 2 Buffon's needles: thicker needles intersect the grid lines.

which they termed the Monte Carlo method. The precise origin of the name varies from one account to another, but there is general agreement that it stems from its relationship with the games of chance played in the casinos of Monte Carlo. The revolutionary step introduced at this time was the use of random number generators (and, latterly, digital computers producing pseudorandom numbers) in place of physical experiments to perform the calculation.* The physics community has continued to contribute to the development of Monte Carlo methodology to the present day.

Perhaps the final revolution in Monte Carlo methodology occurred during the 1980s, during which the increasing use of Bayesian methods and the associated need to evaluate complex high-dimensional integrals led to interest in and development of the methods by the statistics community. This was the point at which the use of Monte Carlo methods to approximate general integral expressions became widespread. The increased interest also drove – and continues to drive – the development of methodology and theory among researchers outside of the traditional application domains of Monte Carlo integration.

Elementary Methods

This section introduces the basic elements of the Monte Carlo method and places them upon a theoretically justified footing.

The Monte Carlo Method

The expected value $\mathbb{E}(X)$ of a random variable, X , which has a probability density function (pdf), f , may be thought of as the average value it would take if the experiment that produced X were carried out infinitely many times. It may be written in the form of an integral:

$$\mathbb{E}(X) = \int xf(x)dx$$

where integration is carried out over the full range of values which X may assume. In the case of discrete random variables with a probability mass function f , the integral is replaced by a summation, but nothing fundamental changes. In this article, random variables will be assumed to be continuous for notational simplicity, but no significant difficulties arise when applying these techniques in a discrete domain.

Underlying the Monte Carlo method is the (strong) law of large numbers which formalizes this concept. If X_1, X_2, \dots

are a collection of independent and identically distributed (iid) random variables with common distribution f , and $\varphi : E \rightarrow \mathbb{R}$ is a function with finite expectation, then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \varphi(X_i) = \mathbb{E}(\varphi(X))$$

where convergence holds with probability 1.

A refinement of this result is provided by the central limit theorem. If we also have that $\mathbb{E}(\varphi(X)^2) < \infty$, then

$$\lim_{n \rightarrow \infty} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \varphi(X_i) - \mathbb{E}(\varphi(X)) \right) \Rightarrow \mathcal{N}(0, \sigma^2)$$

Where ‘ \Rightarrow ’ denotes convergence in distribution. That is, \sqrt{n} times the sample mean converges in distribution to a normal random variable with variance σ^2 .

The most elementary Monte Carlo method may be justified directly using these two theorems. If we wish to estimate the expectation, $\mathbb{E}(\varphi(X))$, where X is a random variable with density function f , from which we can obtain samples, then we may use the empirical mean of φ obtained from a large sample to estimate the expectation of interest. Given a collection of iid random variables X_1, \dots, X_m which have common distribution f , the simple Monte Carlo estimator of $\mathbb{E}(\varphi(X))$ is

$$\frac{1}{n} \sum_{i=1}^n \varphi(X_i)$$

This may be connected with the rainfall experiment presented earlier by considering a function which takes a value of 1 within the circle and 0 outside and calculating its expectation under the distribution of the co-ordinates which is uniform over the square. To see this explicitly, writing this function as

$$\mathbb{I}_{\text{circle}}(x, y) = \begin{cases} 1 & \text{if } x^2 + y^2 \leq r \\ 0 & \text{otherwise} \end{cases}$$

makes it clear that the area of the circle of interest is equal to the integral of this function over any region which includes the circle, such as the square in which it is inscribed. Consequently, our representation for π may be written in the form

$$\begin{aligned} \pi/4 = p &= \frac{\int_{-r}^r \int_{-r}^r \mathbb{I}_{\text{circle}}(x, y) dx dy}{(2r)^2} \\ &= \int_{-r}^r \int_{-r}^r \frac{1}{(2r)^2} \mathbb{I}_{\text{circle}}(x, y) dx dy \end{aligned}$$

Approximating the final line in this expression through the simple Monte Carlo method would give exactly the same estimator as that described above. The term $1/(2r)^2$ corresponds to the density of the uniform distribution over the square $[-r, r] \times [-r, r]$ and, therefore, the integral may be approximated by sampling pairs of points (X, Y) from this distribution and calculating the mean value of the function $\mathbb{I}_{\text{circle}}$ at the sampled points.

* Pseudorandom numbers are generated deterministically but share certain statistical properties with random numbers. A detailed discussion is beyond the scope of this article.

In order to employ simple Monte Carlo, it is necessary to be able to sample from the distribution of interest. This requirement is not necessarily easily satisfied: Monte Carlo methods are generally used to deal with complicated distributions which do not admit tractable analytic solutions and it can be very difficult to obtain samples from such distributions. Two generally-applicable methods for obtaining samples from known distributions are introduced below, and this is followed by a method for estimating expectations under a distribution using samples from a different distribution. That is, the following two methods which are techniques for sampling from a distribution of interest: the samples obtained by these methods can be used in the simple Monte Carlo method described above. The third technique is a Monte Carlo method in its own right: it provides estimates of expectations directly without using samples from the distribution of interest.

Inversion Sampling

The distribution function, F , associated with a pdf, f , is defined as $F(x) = \int_{-\infty}^x f(y)dy$. If X is distributed according to f , then $F(x) = \mathbb{P}(X \leq x)$. The generalized inverse of a distribution function, F , may be defined as:

$$F^-(p) = \inf\{x : F(x) \geq p\}$$

That is, it is the smallest value of x such that $F(x)$ is at least p . Note that this differs from the inverse only if F is discontinuous (a situation which arises whenever there is a finite probability of the associated random variable taking a particular value).

Inversion sampling transforms a random variable, U , with a uniform distribution on the interval $[0, 1]$ (i.e., a random variable with density 1 over $[0, 1]$ and 0 elsewhere), by setting $X = F^-(U)$. If F^- is the generalized inverse of the distribution function of interest, then X has the distribution of interest:

$$\begin{aligned}\mathbb{P}(X \leq x) &= \mathbb{P}(F^-(U) \leq x) \\ &= \mathbb{P}(U \leq F(x)) \\ &= F(x)\end{aligned}$$

Figure 3 illustrates the method. The hollow circles on the vertical axis correspond to three realizations of U and the filled circles on the horizontal axis are the corresponding realizations of X . This method can be used to sample realizations of any one-dimensional real-valued random variable, provided that the inverse of the distribution function is known.

The Fundamental Theorem and Rejection Sampling

A more general approach is motivated by the fact that sampling a collection of random variables according to a given density is equivalent to sampling uniformly in the

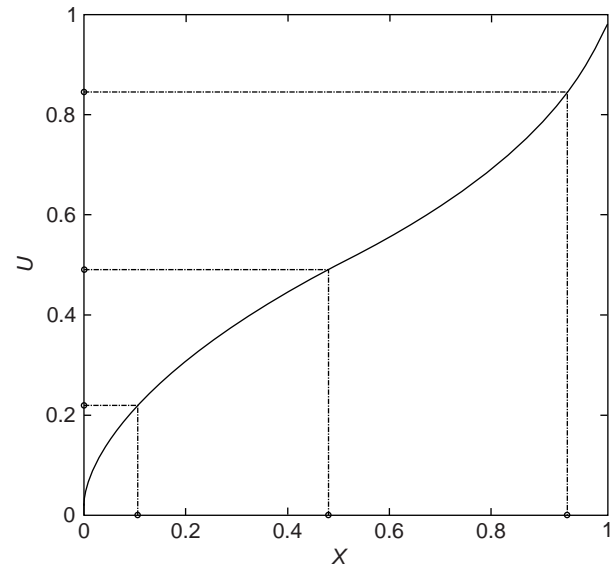


Figure 3 Inversion sampling. Filled circles illustrate samples from the illustrated distribution function.

area under the density graph and discarding the additional dimension. In the case of a real-valued random variable, X , for example, saying that it has density f means that the probability that X lies between x and $x + dx$ is $f(x)dx$ for infinitesimal dx . If we divide the region between $f(x)$ and the x -axis into infinitesimal squares of uniform area and subsequently select one uniformly at random, then it will lead to a sample for X with the same distribution (the number of boxes in the strip of interest is, of course, proportional to $f(x)$) (see **Figure 4** for an illustration). More formally, sampling uniformly from the region $\{(x, u) : 0 \leq u < f(x)\}$ and retaining only x is equivalent to sampling x according to $f(x)$. This is sometimes known as the fundamental theorem of sampling.

Of course, one cannot usually sample uniformly from this region by any direct means – if anything, it is more difficult than sampling from the distribution of interest. Rejection sampling provides samples uniformly under the graph of the density of interest by sampling from a larger area and rejecting those samples which fall outside the region of interest. Formally, given a density, g , from which it is possible to obtain samples and some known constant, $M \geq \sup_x \frac{f(x)}{g(x)}$, one can consider the following algorithm to obtain a sample from f :

- (1) Sample X from g .
- (2) Sample U from $\mathcal{U}(0, M)$.
- (3) If $U > f(X)/g(X)$ reject X and go to 1.
- (4) Accept X as a sample from f .

Figure 5 illustrates the principle: the filled circles indicate accepted samples (which are distributed uniformly beneath the bimodal density) and the hollow circles were rejected (samples were generated uniformly in the area under the rescaled unimodal density).

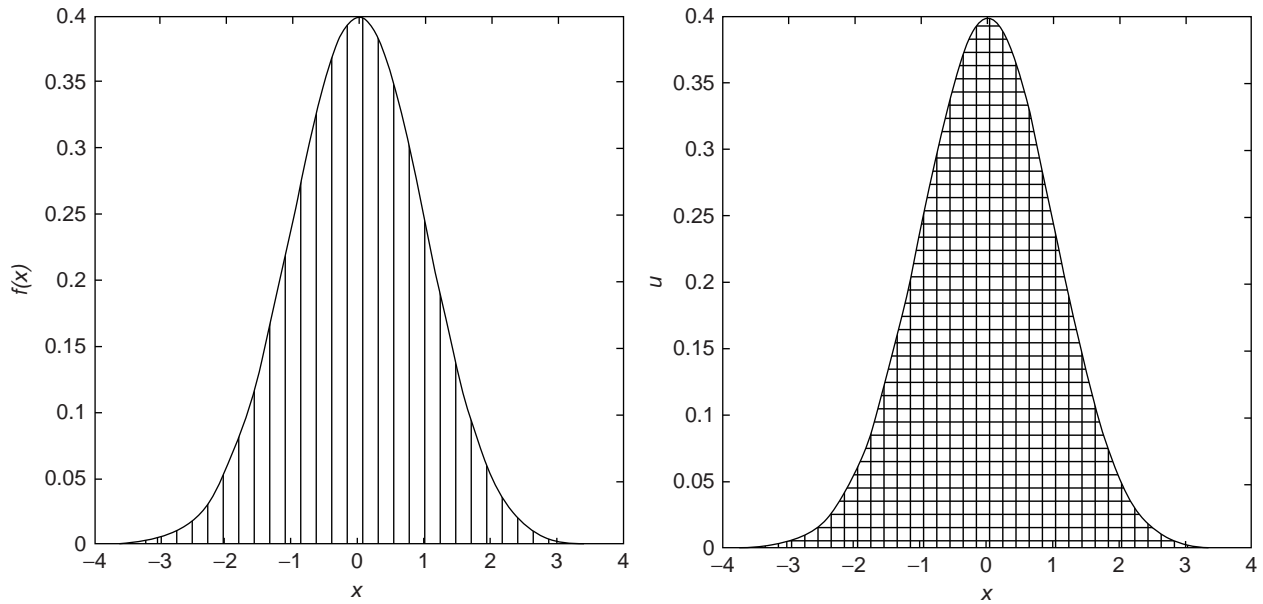


Figure 4 The fundamental theorem of sampling: loosely, sampling from a strip in the left panel with a probability proportional to its height produces a sample with the same distribution over X as would be obtained by choosing a box in the right panel uniformly at random.

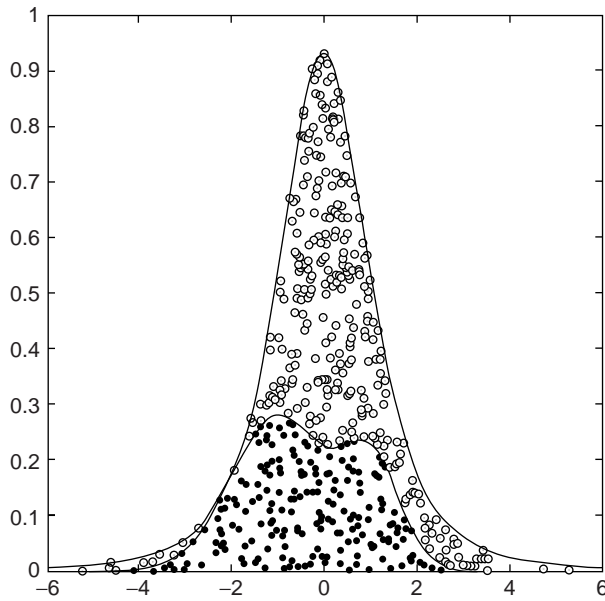


Figure 5 Rejection sampling. Filled circles indicate accepted samples.

Notice that, on average, one sample in M is accepted:

$$\begin{aligned}\mathbb{P}(X \text{ Accepted}) &= \mathbb{P}(U \leq f(X)/g(X)) \\ &= \frac{1}{M} \mathbb{E}[f(X)/g(X)] = \frac{1}{M}\end{aligned}$$

Consequently, for this to be an efficient strategy, it must be possible to sample from a distribution for which a small value of the constant M can be found.

Verifying that the distribution of the accepted samples is, indeed, f , is a simple matter of considering the

probability that a sample lies in an infinitesimal neighborhood, dx , of a point x given that it is accepted:

$$\begin{aligned}\mathbb{P}\left(X \in dx \mid U \leq \frac{f(X)}{g(X)}\right) &= \mathbb{P}\left(X \in dx, U \leq \frac{f(X)}{g(X)}\right) / \mathbb{P}\left(U \leq \frac{f(X)}{g(X)}\right) \\ &= \mathbb{P}(X \in dx) \mathbb{P}\left(U \leq \frac{f(X)}{g(X)} \mid X \in dx\right) / \frac{1}{M} \\ &= g(x)dx \times \frac{1}{M} \frac{f(x)}{g(x)} \times M = f(x)dx\end{aligned}$$

The envelope or squeeze method is a technique for reducing the computational cost of samples obtained by rejection sampling. If cheap-to-evaluate upper and lower bounds for the ratio $f(x)/g(x)$ exist, then it is possible to automatically accept those samples for which U is less than the lower bound and reject those for which it exceeds the upper bound. This means that it is only necessary to evaluate the ratio itself (an operation which may be relatively slow) if the sampled value of U lies in the interval between these two bounds. This can produce substantial computational savings if the bounds are well chosen.

Importance Sampling

Inversion and rejection sampling make it possible to obtain samples from a number of distributions. Unfortunately, many distributions of interest to statisticians are too complicated to allow the efficient generation of samples by these methods. Importance sampling is one

strategy which makes it possible to deal with distributions from which one cannot sample efficiently.

Importance sampling makes use of a distribution other than that of interest, but is used specifically for calculating expectations as it attaches a weight to each sample based on its significance to the integral of interest. If f is the density of interest, we wish to calculate the expectation of φ under that distribution:

$$\begin{aligned} E(\varphi(X)) &= \int \varphi(x)f(x)dx \\ &= \int \varphi(x)\frac{f(x)}{g(x)}g(x)dx \end{aligned}$$

provided that $f(x)/g(x) < \infty$. An importance sampling estimate of the expectation of φ with respect to f is obtained by calculating the simple Monte Carlo estimate of $\varphi f/g$ with samples, X_1, \dots, X_m from g :

$$\int_E \varphi(x)f(x)dx \approx \frac{1}{n} \sum_{i=1}^n \frac{f(X_i)}{g(X_i)} \varphi(X_i)$$

This is illustrated in **Figure 6**.

Importance sampling has a number of advantages over other methods. It does not require that one can simulate from the distribution of interest (in common with rejection sampling) and it does not require that one can obtain a constant M which bounds the ratio $f(x)/g(x)$ (unlike rejection sampling). Even if it is possible to sample directly from f , it can be possible to obtain lower variance estimates if a good choice of proposal distribution, g , is made. Indeed, if φ is a positive function, then choosing $g(x) \propto f(x)\varphi(x)$ will lead to zero variance estimates:

a single sample will yield the exact answer. Unfortunately, this optimal proposal distribution can rarely be used in practice. However, approximations to it, from which it is possible to sample, can often be employed. A widely accepted guideline for the selection of a proposal distribution is that it should lead to bounded importance weights. This is not always necessary to obtain estimates of finite variance, but it is sufficient. In particular, this guideline requires that the tails of the proposal distribution should be at least as heavy as those of the target (and its support must be at least as large as that of the target – that is a general requirement of importance sampling).

As presented above, it is necessary to know $f(x)/g(x)$ pointwise. A simple modification to the algorithm (which leads to an estimator which is biased for finite samples, but consistent, and which typically reduces the variance) allows for situations in which $f(x)/g(x)$ is only known up to a normalizing constant. If $w(x) = Cf(x)/g(x)$, then $\int w(x)\varphi(x)g(x)dx = C\mathbb{E}(\varphi)$ and therefore, if we let $\mathbf{1}$ denote the unit function (the function which maps everything to 1), then we know that:

$$\frac{\int w(x)\varphi(x)g(x)dx}{\int w(x)\mathbf{1}(x)g(x)dx} = \frac{C\mathbb{E}(\varphi)}{C\mathbb{E}(\mathbf{1})} = \mathbb{E}(\varphi)$$

By approximating both the numerator and denominator using a collection of samples from g , an estimator for $\mathbb{E}(\varphi)$ is given by

$$\frac{\sum_{i=1}^n w(X_i)\varphi(X_i)}{\sum_{i=1}^n w(X_i)}$$

This is often referred to as the self-normalized importance sampling estimator.

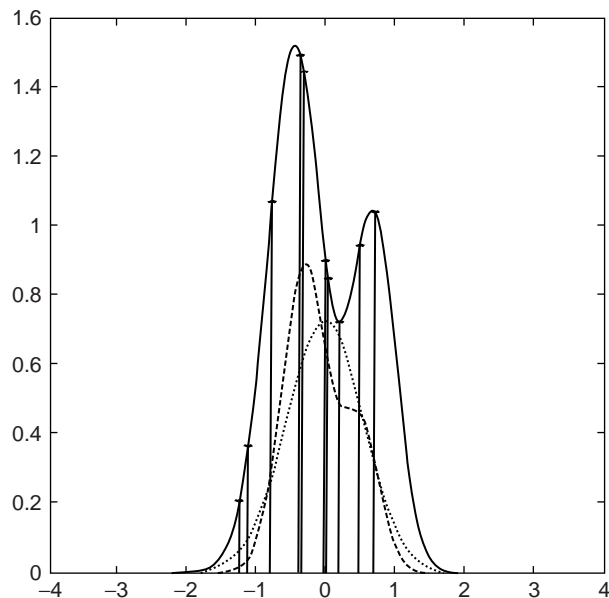


Figure 6 Importance sampling: ten samples and their weights (vertical lines) from a proposal distribution (dotted line) weighted according to the weight function (solid line) appropriate for the target distribution (dashed line).

Variance Reduction Techniques

One criterion which can be used to assess the performance of a Monte Carlo technique is the variance of the estimators which it produces. Reducing this variance is one of the principal aims of the designer of a Monte Carlo simulation. The intuition behind three widely applicable techniques for doing this is presented in this section.

Rao-Blackwellization

If the random variable of interest X may be decomposed as $X = (Y, Z)$ and it is possible to calculate the expectation of the function of interest, φ , given a particular value of Z , then the estimator obtained by performing this integration (the expectation of $\varphi(X)$ conditional on a value of Z is, of course, the integral of that function with respect to the conditional distribution of X given Z) is never worse than that defined on the joint space (in that it has small variance and, by the tower property of expectation, the

same mean $\mathbb{E}[\mathbb{E}(\varphi|Z)] = \mathbb{E}[\varphi]$. This can be seen by considering the straightforward variance decomposition:

$$\begin{aligned}\text{Var}(\varphi(Y, Z)) &= \text{Var}(\mathbb{E}[\varphi|Z]) + \mathbb{E}[\text{Var}(\varphi|Z)] \\ &\Rightarrow \text{Var}(\mathbb{E}[\varphi|Z]) \leq \text{Var}(\varphi(Y, Z))\end{aligned}$$

That is, if it is possible to integrate out some subset of the random variables, then it is always preferable to do so. This is sometimes referred to as Rao-Blackwellization because of its connection to the Rao-Blackwell theorem.

Control Variates

If one knows the expectation of some function ψ under the distribution of interest, then one may use its Monte Carlo estimate (using the same sample as that used to estimate the integral of the function of interest, φ) to create an estimator of 0. When suitably weighted, this estimator will, under certain conditions, have a negative correlation with the estimator of ψ and may be used to reduce its variance. Calculation of the weight required to obtain a variance reduction may not be straightforward.

Antithetic Variables

If one can obtain two estimators of the integral of interest based upon a random sample (in the case of a symmetric distribution, e.g., it is possible to use the reflection of the sample set to provide a second estimator) and which are negatively correlated then the variance of the average of these estimators will be less than that of either one and will certainly outperform the poorer of the two estimators. This is the principle behind the technique of antithetic variables.

More Advanced Approaches

Inevitably, more advanced techniques than those described above are required to deal with complex problems. This section provides a necessarily brief overview of three more sophisticated techniques.

Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) is a technique which is widely used to deal with complex distributions for which the methods described above prove inadequate. They employ suitable sequences of dependent random variables to approximate integrals of interest rather than attempting to obtain independent samples. It is possible to obtain results showing the consistency of such estimates, and that they obey a central limit theorem under certain regularity conditions.

Quasi-Monte Carlo

Quasi-Monte Carlo (QMC) is not really a Monte Carlo technique at all. In QMC, integrals are approximated using deterministic (rather than random) sequences of variables which are chosen to have a property termed low discrepancy. This approach can be interpreted as making a virtue of the inability of computers to provide truly random numbers by selecting points according to an analytic criterion, rather than attempting to mimic randomness. Such methods can exhibit theoretically faster convergence than Monte Carlo approaches, especially when used for the integration of smooth functions in low-dimensional spaces. Numerical simulations have verified this for simple problems in spaces of small dimension. In practice, it has proved difficult to obtain the promised performance improvements in more interesting problems and some care is required when implementing these methods.

Sequential Monte Carlo

Sequential Monte Carlo (SMC) is a term used to refer to algorithms which obtain samples from a sequence of distributions by using an iterated importance sampling strategy together with some refinements. Historically, these algorithms have been widely used for online estimation in time-series problems; more recently, it has been shown that they provide an effective way to obtain samples from general distributions.

See also: Bayesian Statistical Analysis; Markov Chain Monte Carlo; Probability Theory.

Further Reading

- Badger, L. (1994). Lazarrini's lucky approximation of π . *Mathematics Magazine* **67**, 83–91.
- Bousquet, O., von Luxburg, U., and Rätsch, G. (eds.) (2004). *Advanced Lectures on Machine Learning*, Vol. 3176. Heidelberg: Springer.
- Christophe, A. (2000). Monte Carlo methods for absolute beginners. In Bousquet, O., Von Luxburg, U., and Rätsch, G. (eds.) *Advanced Lectures on Machine Learning*, Vol. 3176, pp 113–145. Heidelberg: Springer.
- Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society B* **63**(3), 411–436.
- Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer.
- Doucet, A., de Freitas, N., and Gordon, N. (eds.) (2001). *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. New York: Springer.
- Eckhardt, R. (1987). Stan Ulam, John von Neumann and the Monte Carlo method. *Los Alamos Science* **15**, 131.
- Geweke, J. (1989). Bayesian inference in econometrics models using Monte Carlo integration. *Econometrica* **57**(6), 1317–1339.
- Knuth, D. E. (1998). *The Art of Computer Programming*, Volume 1. *Fundamental Algorithms*, 3rd edn. Boston, MA: Addison-Wesley.
- Liu, J. S. (1987). *Monte Carlo Strategies in Scientific Computing*. *Springer Series in Statistics*. New York: Springer.

- Marsaglia, G. (1977). The squeeze method for generating gamma variates. *Computers and Mathematics with Applications* **3**, 121–325.
- Metropolis, N. (1987). The beginnings of the Monte Carlo method. *Los Alamos Science* **15**, 125–130.
- Morokoff, W. J. and Caflisch, R. E. (1995). Quasi-Monte Carlo integration. *Journal of Comparative Physiology* **122**, 218–230.
- Ripley, B. D. (1987). *Stochastic Simulation*. Wiley.
- Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Method*, 2nd edn. New York: Springer.

Relevant Websites

<http://math.iit.edu> – Quasi Monte Carlo and Monte Carlo Methods.
<http://www-sigproc.eng.cam.ac.uk> – SMC Preprint Server.

Multidimensional Scaling

Z Zhang and Y Takane, McGill University, Montreal, QC, Canada

© 2010 Elsevier Ltd. All rights reserved.

Glossary

(Dis)similarity – The term similarity is used to indicate the degree of likeness between two objects, while dissimilarity indicates the degree of unlikeness. For example, red and pink are more similar (less dissimilar) to each other than red and green. (Red and green are more dissimilar (less similar) than red and pink.) In the similarity data, a larger value indicates more similar objects, while in the dissimilarity data, a larger value indicates more dissimilar objects.

Distance Models – The models used to represent (dis)similarity data in MDS. Although there are other distance models, the Euclidean distance model is the most popular one used in MDS.

Multidimensional scaling (MDS) – A set of data analysis techniques for analysis of similarity or dissimilarity data. It is used to represent (dis)similarity data between objects by a variety of distance models.

Unfolding analysis – One way of representing individual differences in preference judgments. In unfolding analysis, subjects' ideal objects and actual objects are represented as points in a joint multidimensional space in such a way that the distances between them are as inversely related to the observed preferences as possible.

The notion of similarity plays a fundamental role in human cognition (Takane *et al.*, 2009). It serves as an organizing principle by which people categorize, generalize, and classify objects. Multidimensional scaling (MDS) is a set of data analysis techniques for representing (dis)similarity data (similarity or dissimilarity data) by spatial distance models (Takane, 2007). In this article, we explicate the purposes, the mechanism, and the variety of uses of MDS.

This article consists of the following sections: (1) "Introduction," through an example, to illustrate the basic roles and the uses of MDS, (2) "Distance models, fitting criteria, and the data collection methods," (3) "Scale levels and data transformations," (4) "Dimensionality selection," (5) "Individual differences MDS," (6) "Unfolding analysis," and (7) A summary and software for MDS.

Introduction

Some objects are more similar (or dissimilar) to each other than others. For example, red and pink are more similar than red and green. MDS represents the similarity or dissimilarity data among the objects by mapping the points (representing the objects) into a multidimensional space in such a way that the distances between them best accord with the observed (dis)similarity data between the objects. In the above example, the points representing red and pink are located closer in the space than the points representing red and green. By virtue of MDS, we can visually inspect the (dis)similarity data among the objects and investigate the principle underlying the organization of the (dis)similarity data.

To further illustrate the role of MDS, let us take a look at the data in **Table 1**. This table shows dissimilarity data among eight different sports. The names of two sports were presented each time to the subjects, who were asked to indicate the degree of dissimilarity between them on a 11-point rating scale. Entries in the table indicate average dissimilarities between the sports across ten subjects. The eight sports are: (1) baseball, (2) basketball, (3) rugby, (4) soccer, (5) softball, (6) table tennis, (7) tennis, and (8) volleyball. MDS was applied to the table, and the derived object configuration is presented in **Figure 1**.

By inspection, it can readily be seen that MDS indeed located the points corresponding to similar objects close together, while locating those corresponding to dissimilar objects far apart. **Figure 1** shows that the eight sports are roughly classified into four groups, the first consisting of rugby and soccer, the second consisting of volleyball and basketball, the third consisting of baseball and softball, and the fourth consisting of tennis and table tennis. This is consistent with our intuition that the sports within the groups have much in common. The four groups of sports may further be combined in various ways to form larger clusters. For example, baseball, softball, rugby, and soccer might be grouped into one, and the remaining sports (basketball, volleyball, tennis, and table tennis) into the other. Since the first group occupies the upper right portions of the configuration, and the second group lower left portions, we may interpret the direction from upper right to lower left contrasting sports that use a big outdoor field with those that require only a medium- to small-size court. We may also group baseball, softball, tennis, and table tennis into one group, and the remaining sports (rugby,

Table 1 Mean dissimilarity ratings among eight sports

St	1	2	3	4	5	6	7
2	8.0						
3	8.7	8.3					
4	8.6	7.6	4.7				
5	1.3	9.3	9.9	9.4			
6	8.7	8.8	9.6	9.8	8.7		
7	8.0	9.1	9.4	9.6	7.9	2.1	
8	8.4	4.6	8.3	7.1	8.4	7.2	5.6

The eight sports are: 1. baseball, 2. basketball, 3. rugby, 4. soccer, 5. softball, 6. table tennis, 7. tennis, and 8. volleyball.

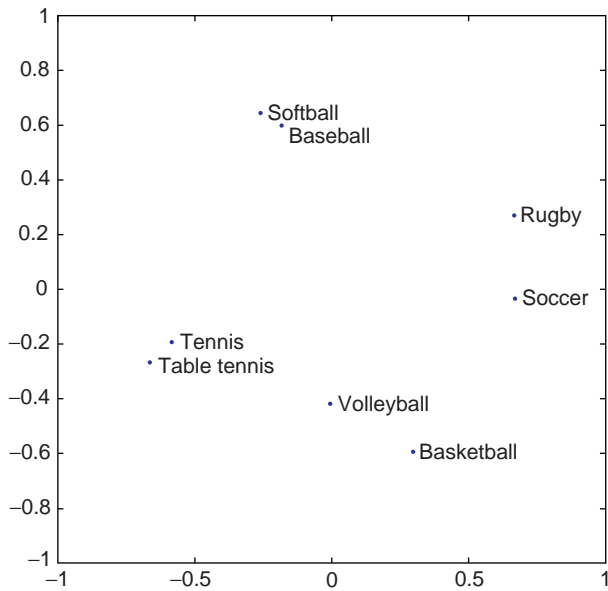


Figure 1 The two-dimensional object configuration of the eight sports from average dissimilarity ratings in **Table 1** (a PROXSCAL solution obtained under the assumption of interval-scaled dissimilarity data and with 100 random initial starts; normalized raw stress = 0.020).

soccer, volleyball, and basketball) into the other. The former use a relatively small ball, while the latter use a big ball. We may call the direction from upper left to lower right the ball size dimension. MDS, simply stated, is a sort of gadget that draws a map similar to the one presented in **Figure 1** based on a set of distance-like quantities (similarity or dissimilarity data) similar to the ones given in **Table 1**. The map facilitates our intuitive understanding of the relationships among the objects represented in the map.

Distance Models, Fitting Criteria, and the Data Collection Methods

As noted above, MDS represents interobject (dis)similarities by interpoint distances. While there are a variety of distance models that may be used in MDS, the one most

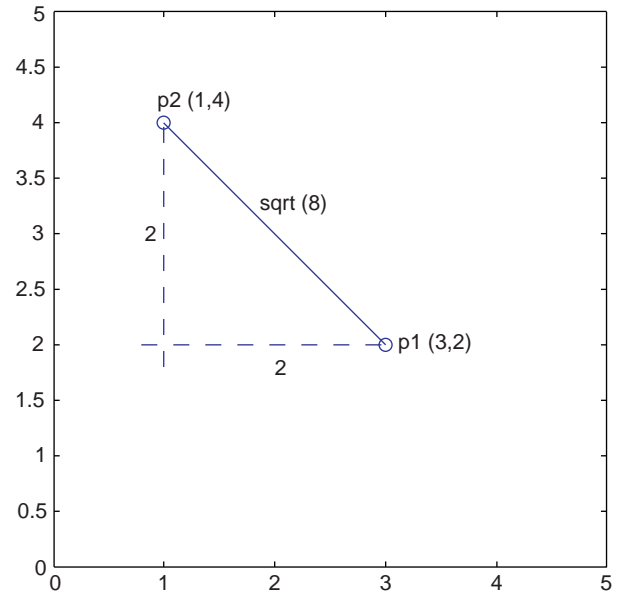


Figure 2 Calculating the Euclidean distance between two points in the two-dimensional space.

frequently used is the Euclidean distance model. Let x_{ir} denote the coordinate of point i (object i) on dimension r . Then, the Euclidean distance between points i and j is calculated by

$$d_{ij} = \left\{ \sum_{r=1}^R (x_{ir} - x_{jr})^2 \right\}^{1/2} \quad [1]$$

where R indicates the dimensionality of the space. Once x_{ir} 's are given, we can locate the points in the space using a Cartesian coordinate system, and we can calculate the distance between them using the above formula. Suppose that $R = 2$, and the coordinates of point 1 on the two dimensions are 3 and 2 ($x_{11} = 3$ and $x_{12} = 2$), and the coordinates of point 2 are 1 and 4 ($x_{21} = 1$ and $x_{22} = 4$). Then, these two points can be located as indicated in **Figure 2**. The Euclidean distance between them can be calculated by $d_{12} = \{(3 - 1)^2 + (2 - 4)^2\}^{1/2} = \sqrt{8} \approx 2.828$. MDS locates the points (i.e., finds their coordinates), representing the objects in such a way that the set of distances calculated from the coordinates best agree with the observed (dis)similarities between the objects.

One important feature of the Euclidean distance is that it is invariant over the choice of origin and orientation of coordinate axes. In MDS, we typically place the origin at the centroid of the object configuration, and rotate the configuration in such a way that the coordinate axes represent substantively meaningful attributes. (Note, however, that some distance models used in MDS, e.g., the weighted Euclidean model, do not allow rotation of axes without changing the interpoint distances. See the section entitled "Individual differences MDS.")

Observed data typically contain a sizable amount of measurement errors, and an exact representation of the data is usually impossible. Rather, we look for the best approximation of the observed (dis)similarity data. To make this notion more rigorous, we need to introduce an index that measure's the goodness (or badness) of agreement between the observed data and the distance model. This index also serves as a criterion to be optimized in MDS. That is, an MDS procedure systematically looks for the object configuration that maximizes the goodness (or minimizes the badness) of fit of the distance model to the observed (dis)similarity data.

Two broad classes of goodness of fit criteria have been used in MDS. One is the least squares (LS) criterion (Kruskal, 1964a, 1964b), and the other is the maximum likelihood (ML) criterion (Ramsay, 1977, 1982). Although the latter has some appeal for its statistical inference capabilities, the former has been far more predominantly used in MDS for its simplicity and flexibility. Let o_{ij} denote the observed dissimilarity between objects i and j (temporarily assumed to have been measured on a ratio scale; see the next section for scale levels of measurement), and let d_{ij} denote the corresponding distance between points i and j in the Euclidean space. Then, the LS criterion is defined by:

$$\phi(\{x_{ir}\}) = \sum_{i < j} (o_{ij} - d_{ij})^2 \quad [2]$$

where $\{x_{ir}\}$ is a collection of object coordinates. (The LS criterion of the above form is often called raw stress in the MDS literature.) This is a badness of fit criterion, meaning that a larger value indicates a larger discrepancy between the distance model and the observed dissimilarity data. The LS MDS attempts to find the set of object coordinates $\{x_{ir}\}$ so as to minimize the discrepancy between the observed o_{ij} and the predicted d_{ij} .

A minimization of the LS criterion generally involves a very complicated process because the distance model is not a simple linear function of its parameters (object coordinates). The equations to be satisfied at the minimum of ϕ usually cannot be solved in closed form, and some kind of iterative methods have to be used to solve them. In the iterative methods, successive approximations to the final solution are obtained by gradually improving the goodness of fit of the solution, starting from an initial guess, until a sufficiently close approximation is found. See the work by Borg and Groenen (2005) for more details of the algorithms used in MDS.

One potential danger of this kind of iterative optimization procedures is known as the problem of convergence to nonglobal minima. The LS criterion used in MDS may have multiple local minima, and iterative optimization procedures may be caught up by one of them that is not the true minimum of the criteria we wish to find. Fortunately, computers nowadays are so powerful that it is not

at all unrealistic to obtain multiple solutions starting from many different initial estimates. Multiple solutions may be compared in terms of their goodness of fit, and the best solution can be chosen, which is more likely to be the globally optimal solution we want.

A variety of (dis)similarity measures may be used as input data to MDS, including (dis)similarity ratings, sorting data, confusion data, frequency of co-occurrences, response latency (reaction time data), frequency of social interactions, profile similarity, and so on. In the (dis)similarity rating methods, objects are presented in pairs to the subjects, who are asked to rate the degree of (dis)similarity between them on a rating scale. (The data from the example presented earlier on dissimilarity among eight sports were collected by this method.) In the sorting method, subjects are given a set of objects and are asked to group them into several groups in terms of their similarity. The number of times two objects are put into same groups is counted over a group of subjects and used as a similarity measure between the objects. See the work by Takane *et al.* (2009) for more systematic descriptions of the data collection methods used in MDS.

Scale Levels and Data Transformations

Scale levels refer to approximate relationships that may be assumed to hold between observed dis(similarity) data and distances. The distinction between different scale levels is important because certain MDS procedures only apply to dis(similarity) data measured on certain scale levels.

There are four scale levels considered in MDS: ratio, interval, log interval, and ordinal. Let o_{ij} denote the observed dissimilarity between objects i and j , and let d_{ij} denote the corresponding distance. In the ratio scale level, it is assumed that $o_{ij} \approx d_{ij}$, where \approx means approximately equal. In this case, d_{ij} can be directly fitted to o_{ij} so as to minimize criterion [2]. No data transformation is necessary. However, it is rare to find the ratio-scaled measurement in social science research.

In the interval-scaled measurement, it is assumed that $o_{ij} \approx ad_{ij} + b$, where a is $+1$ if the data are dissimilarity data, or -1 if they are similarity data (o_{ij} can be either similarity or dissimilarity data), and b is an additive constant. In the case of interval-scaled data, the fitting criterion is generalized into

$$\phi(\{x_{ir}\}, b) = \sum_{i < j} (\pm o_{ij} - d_{ij} - b)^2 \quad [3]$$

and both the object configuration and an optimal value of b have to be estimated that jointly minimize the criterion.

In a log-interval scale, it is assumed that $o_{ij} \approx bd_{ij}^a$, that is, the observed (dis)similarity data are related to underlying distances by a power transformation. When the log is

taken on both sides of the above relationship, we obtain $\ln o_{ij} \approx a \ln d_{ij} + \ln b$, which is a linear relationship between $\log o_{ij}$ and $\log d_{ij}$ similar to the interval scale level, hence the name log-interval scale. Not many MDS procedures recognize this scale level as such, and the (dis)similarity data at this scale level are often analyzed as mere ordinal-scaled data.

In the ordinal scale level, o_{ij} and d_{ij} are assumed to be only monotonically related. That is, $o_{ij} > o_{i'j'}$ implies $d_{ij} \geq d_{i'j'}$ if the data are dissimilarity data, whereas $o_{ij} > o_{i'j'}$ implies $d_{ij} \leq d_{i'j'}$ if the data are similarity data. MDS procedures that are capable of handling ordinal (dis)similarity data are called nonmetric MDS (Shepard, 1962; Kruskal, 1964a, 1964b) and enjoy widest applications. The fitting criterion in this case is modified into:

$$\phi(\{x_{ir}\}, m) = \sum_{i < j} (m(o_{ij}) - d_{ij})^2 \quad [4]$$

where m denotes a monotonic (or an inversely monotonic) transformation. MDS procedures in this case have to find the best monotonic transformation of the ordinal data as well as the object configuration $\{x_{ir}\}$ that jointly minimize the above criterion.

We usually do not know *a priori* exact scale levels that the observed (dis)similarity data satisfy. As a practical strategy, we may start with a weaker assumption, but as soon as we find, as a result of the analysis, that a stronger measurement assumption can be justified, we switch to the stronger assumption. In this way, we can get more reliable results while avoiding unaffordable scale level assumptions.

Dimensionality Selection

One important decision that has to be made in MDS concerns the dimensionality of the solution space. The dimensionality refers to the number of coordinates needed to locate a point in the spatial representation of objects. There are several considerations that should be taken into account in determining the number of dimensions. The derived object configuration has to fit to the data at hand reasonably well, but should not fit too well. A better fit to the data at hand can generally be achieved by merely increasing the dimensionality of the solution space, and too good a fit may compromise the predictability of the model for future observations. One practical strategy for determining the adequate number of dimensions is to analyze the data under varied dimensionalities, say from 1 to 4, plot the fit value against the dimensionality (this is called a scree plot), and identify the point where the improvement in fit flattens out. Such a point is called an elbow in the scree plot.

Another important consideration in determining the dimensionality of the solution space is the interpretability of derived dimensions. Uninterpretable dimensions are

useless and should not be retained (even if they are necessary to account for the observed (dis)similarity data sufficiently well).

In the example of sports data, the raw stress values were 0.186 for the unidimensional solution, 0.020 for the two-dimensional solution, 0.003 for the three-dimensional solution, and 0.001 for the four-dimensional solution. The two-dimensional solution was easily interpretable, whereas the third-dimensional one was not. Thus, the two-dimensional solution was selected as the best one.

Individual Differences MDS

So far, we have assumed that there is only one set of (dis)similarity data. In many applications of MDS, however, (dis)similarity data are collected from a group of subjects. If no systematic individual differences exist, a single common Euclidean distance model may be fitted to all of them simultaneously, or a single Euclidean distance model is fitted to average (dis)similarity data, as has been done in the example presented earlier. In many situations, however, the assumption of no systematic individual differences is unrealistic. In such a case, each (dis)similarity matrix may be analyzed separately, yielding as many object configurations as there are (dis)similarity matrices. A natural question is how they are related. In most cases, there are both common and unique aspects in (dis)similarity judgments obtained from different individuals. If so, we need a methodology that captures both aspects.

The individual differences (ID) MDS model captures both commonality and individual differences in (dis)similarity judgments (Carroll and Chang, 1970). More specifically, it postulates a common object configuration that applies to all individuals; however, dimensions in the common configuration are differentially weighted by different individuals to give rise to differences in (dis)similarity judgments. The idea of differential weighting of dimensions can be captured by the weighted Euclidean distance model:

$$d_{ijk} = \left\{ \sum_{r=1}^R w_{kr} (x_{ir} - x_{jr})^2 \right\}^{1/2} \quad [5]$$

where d_{ijk} is the distance between points (objects) i and j for individual k , x_{ir} is, as before, the coordinate of object i on dimension r in the common object configuration, and w_{kr} is the weight attached to dimension r by subject k . To eliminate the size indeterminacy between the object configuration and the individual difference weights, the former is typically constrained to satisfy $\sum_{i=1}^n x_{ir}^2 / n = 1$ for $r = 1, \dots, R$. In contrast to the simple Euclidean distance model [1], the orientation of the coordinate axes is uniquely determined in the weighted Euclidean distance model. ID MDS estimates both the object coordinates $\{x_{ir}\}$ and the individual differences weights $\{w_{kr}\}$ in such a way

Table 2 Dissimilarity ratings among eight sports by ten subjects

	1	2	3	4	5	6	7
Sub. 1	7						
	8	8					
	9	9	3				
	2	9	10	10			
	9	9	10	9	8		
	10	10	10	10	10	5	
	10	4	9	4	9	9	8
	8						
Sub. 2	9	7					
	11	4	4				
	1	11	11	11			
	11	11	11	10	11		
	10	10	11	11	11	1	
	9	2	8	3	10	11	4
	8						
	10	10					
Sub. 3	8	9	4				
	1	9	10	9			
	10	10	10	10	10		
	10	10	10	10	9	1	
	9	6	10	10	9	10	6
	10						
	11	8					
	9	9	7				
Sub. 4	2	8	11	10			
	7	7	11	10	8		
	8	9	9	10	6	2	
	9	6	10	10	8	6	6
	7						
	8	8					
	10	5	3				
	1	8	9	8			
Sub. 5	4	7	9	6	4		
	4	7	7	8	4	2	
	6	5	7	5	6	5	4
	5						
	8	6					
	7	7	4				
	1	8	9	9			
	7	9	9	11	6		
Sub. 6	5	10	9	8	7	3	
	9	6	5	7	9	7	5
	8						
	9	7					
	9	10	8				
	1	9	9	9			
	9	10	9	10	10		
	8	10	10	10	7	2	
Sub. 7	6	6	8	8	8	7	4
	9						
	10	10					
	9	10	5				
	2	11	9	9			
	11	10	11	11	10		
	10	9	10	10	11	3	
	9	4	10	10	9	10	11

Continued

Table 2 Continued

	1	2	3	4	5	6	7
Sub. 9	10						
	8	9					
	10	7	8				
	1	10	11	11			
	10	10	11	10	11		
	6	9	10	11	10	1	
	8	5	8	8	10	4	5
	8						
Sub. 10	6	10					
	4	6	1				
	1	10	10	8			
	9	5	5	11	9		
	9	7	8	8	4	1	
	9	2	8	6	6	3	3

The eight sports are: 1. baseball, 2. basketball, 3. rugby, 4. soccer, 5. softball, 6. table tennis, 7. tennis, and 8. volleyball.

that d_{ijk} calculated from them best agree with the observed dissimilarity between objects i and j by subject k .

As an example of ID MDS, let us look at the data in **Table 2** obtained from ten subjects. This is the original data from which the data in **Table 1** were calculated by averaging them over the ten subjects. Individual differences MDS was applied to the ten dissimilarity matrices. **Figure 3** displays the derived two-dimensional common object configuration. In the figure, the sports that use a big ball are located toward the right, and those that use a small ball toward the left, so that the horizontal axis can be interpreted as representing the ball size dimension. The vertical axis, on the other hand, places the sports that use a big field at the top, and those that use a small court at the bottom, thus contrasting between the two. (As has been alluded to above, coordinate axes are unrotatable in the weighted Euclidean model, so that we don't have to search for meaningful directions in the space. We simply try to interpret the directions of the coordinate axes.) **Figure 4** depicts the individual differences weights attached to the two dimensions by different subjects. In the figure, two extreme subjects are identified by subject numbers. Subject 6 puts the most emphasis on dimension 1 among all the subjects, whereas subject 10 does so on dimension 2, although, in this particular example, the weights are relatively homogeneous, indicating that there are not many differences in the way the two dimensions are evaluated by the different subjects. In some cases, the weights may show interesting patterns of differences that may be related to subjects' background information such as gender, age, level of education, etc., but, unfortunately, such information is unavailable in the present case (see the work by Takane (2007) for more examples of interesting applications of the ID MDS).

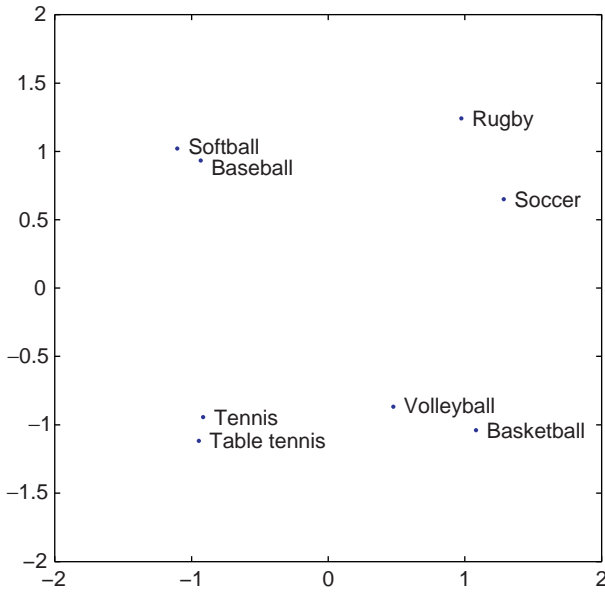


Figure 3 The two-dimensional common object configuration of the eight sports obtained by individual differences MDS of dissimilarity data in Table 2 (a PROXSCAL solution under the assumption of ordinal dissimilarity data and matrix conditional; normalized stress = 0.052).

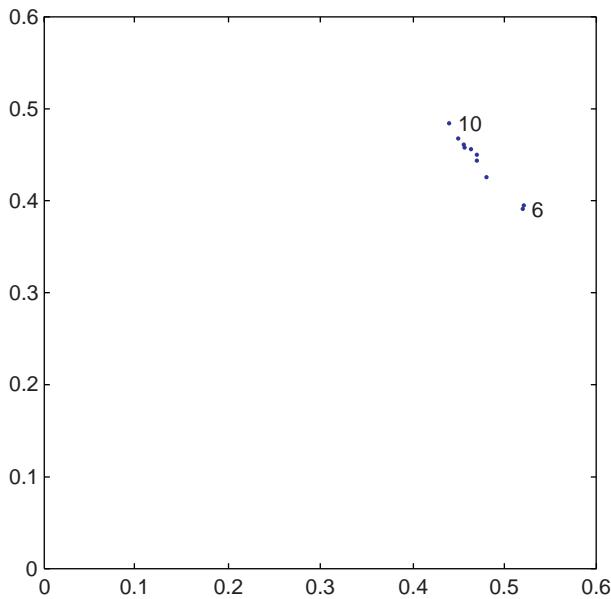


Figure 4 The plot of individual differences weights attached to the two dimensions by the ten subjects for the data in Table 2.

Unfolding Analysis

Individual differences are far more prevalent in preference judgments. Preference data are often analyzed by a variant of MDS called unfolding analysis (Coombs, 1964).

In unfolding analysis, each subject is assumed to have an ideal object represented as the subject's ideal point in the same space as actual objects are represented. The distances between the ideal point and the object points are assumed to be inversely related to the subject's preferences on the objects. Let x_{ir} denote the coordinate of object i on dimension r , and y_{jr} the coordinate of subject j 's ideal point on dimension r . The Euclidean distance between object point i and ideal point j is calculated by

$$d_{ij} = \left\{ \sum_{r=1}^R (x_{ir} - y_{jr})^2 \right\}^{1/2} \quad [6]$$

The coordinates of the ideal and object points are determined in such a way that the preference values of the objects for a particular subject are a decreasing function of the distances between them. This implies that the closer an object point is to his ideal, the more preferred the object is by that subject. The preference relations are thus regarded as representing similarity relations between the subjects' ideal objects and actual objects. In unfolding analysis, we are given an N by n data matrix obtained from N subjects making preference judgments on n objects. By subjecting the data matrix to unfolding analysis, we obtain two coordinate matrices, one for object points and the other for subjects' ideal points.

As an example of unfolding analysis, let us look at Table 3. Thirty-one subjects rank-ordered six different colors from the least preferred to the most preferred. The data are thus similarity data with greater numbers indicating more preferred colors and greater similarities between subjects' ideal color and actual colors. Ties were allowed, and were given the average of ranks they would have received if they were not exactly tied. The six colors used in the study are: orange (o), blue (b), grass color (g1), green (g2), red (r), and purple (p). PREFSCAL (Busing, Groenen, and Heiser, 2005) was used to analyze the data.

Figure 5 depicts the joint MDS configuration of the subject's ideal points and the six colors. The first five subjects in the data set are identified by the integers. By inspection, colors most preferred by these subjects are located close to these subjects' ideal points. For example, subject 1 prefers blue and grass color, while subject 2 red and purple. Subject 5's ideal point is somewhat outlying (from the rest), indicating that another form of preference model called the vector preference model may be more appropriate for the subject (Busing *et al.*, 2005). The analysis was run under the interval scale level assumption rather than the ordinal scale assumption, which was probably more realistic. Nonetheless, the stronger assumption was deemed preferable because of the small number of objects to avoid partially degenerate solutions. (Degenerate solutions are those that

Table 3 Preference rankings among six colors by 31 subjects

Sub/St	o	b	g1	g2	r	p
1	1.0	6.0	5.0	2.5	4.0	2.5
2	1.0	4.0	3.0	2.0	6.0	5.0
3	4.5	4.5	6.0	2.0	3.0	1.0
4	1.0	5.0	4.0	6.0	3.0	2.0
5	6.0	1.0	5.0	2.0	4.0	3.0
6	1.0	4.5	4.5	3.0	6.0	2.0
7	1.0	4.0	6.0	3.0	5.0	3.0
8	1.0	5.5	3.0	2.0	5.5	4.0
9	5.0	4.0	2.0	6.0	1.0	3.0
10	2.5	5.5	2.5	1.0	5.5	4.0
11	1.0	2.0	5.0	3.0	6.0	4.0
12	1.5	5.0	4.0	6.0	1.5	3.0
13	2.0	5.0	3.0	1.0	6.0	4.0
14	5.0	6.0	1.0	3.0	4.0	2.0
15	1.0	5.0	2.0	6.0	3.0	4.0
16	4.5	6.0	3.0	4.5	1.0	2.0
17	1.0	6.0	2.0	5.0	4.0	3.0
18	3.5	5.5	1.0	5.5	3.5	2.0
19	3.0	6.0	2.0	1.0	5.0	4.0
20	3.0	2.0	5.0	1.0	4.0	6.0
21	1.0	2.0	5.0	5.0	3.0	5.0
22	3.0	5.0	4.0	6.0	1.0	2.0
23	6.0	4.5	3.0	1.0	4.5	2.0
24	5.0	2.5	4.0	6.0	2.5	1.0
25	1.0	4.0	5.5	3.0	5.5	2.0
26	6.0	4.0	3.0	5.0	2.0	1.0
27	2.0	5.0	3.0	6.0	1.0	4.0
28	1.0	3.0	6.0	2.0	5.0	4.0
29	1.0	4.0	3.0	2.0	6.0	5.0
30	1.0	4.5	3.0	2.0	6.0	4.5
31	5.0	4.0	3.0	6.0	1.0	2.0

Stimulus labels are: o-orange, b-blue, g1-grass color, g2-green, r-red, p-purple.

exhibit an excellent fit, but are substantively meaningless, e.g., all object points collapsing into one, and all ideal points into another.)

Unfolding analysis is a very useful technique in marketing research. It allows us to understand patterns of individual differences in preference judgments, and their relationships to product features and subjects' background information. This kind of analysis may eventually help marketing analysts to develop practical marketing strategies. Interested readers are referred to the work by Takane (2007) for more examples of application of unfolding analysis.

A Summary and Software for MDS

MDS is designed for visualization of observed (dis)similarity data by distance models. In this article, we discussed essential ingredients for practical uses of MDS, such as the distance models, fitting criteria, the data collection methods, and levels of measurement scales. We also discussed several variants of MDS (simple MDS, individual differences MDS, and unfolding analysis) with concrete

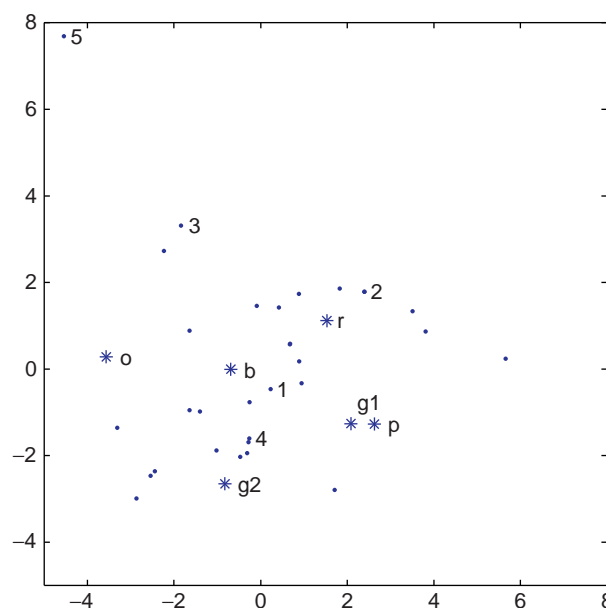


Figure 5 The joint plot of object points (asterisks) and subjects' ideal points (dots) for the preference data on the six colors presented in Table 3 (a PREFSCAL solution obtained under the interval-scaled similarity data). The objects are labeled, and the first five subjects' ideal points are numbered while the remaining ideal points are merely indicated by dots. (Some subjects' ideal points coincide.)

examples of application. For more detailed discussions on these topics, see the work by Takane *et al.* (2009).

ALSCAL (Takane *et al.*, 1977) has been in statistical packages for social sciences (SPSS) for quite a long time now. However, it is slowly being superseded by a newer program PROXSCAL (Busing *et al.*, 1997). The latter directly fits the distance model (rather than the squared distance model), allows multiple random starts (rather than a single rational start), and has better graphing features in the output.

Unfolding analysis has been a difficult one to undertake because of many instances of degenerate solutions. PREFSCAL (Busing *et al.*, 2005) seems to have largely overcome the problem by incorporating a penalty term in the optimization criterion. PREFSCAL has recently been incorporated into SPSS.

MULTISCALE, an ML MDS program, can be downloaded free online along with the program manual.

See also: Analysis and Interpretation of Multivariate Data; Cluster Analysis: Overview; Data Mining.

Bibliography

- Borg, I. and Groenen, P. J. F. (2005). *Modern Multidimensional Scaling: Theory and Applications*. New York: Springer.
- Busing, F. M. T. A., Commandeur, J. J. F., and Heiser, W. J. (1997). PROXSCAL: A multidimensional scaling program for individual differences

- scaling with constraints. In Bandilla, W. and Faulbaum, F. (eds.) *Softstat '97: Advances in Statistical Software*, pp 237–258. Stuttgart: Lucius.
- Busing, F. M. T. A., Groenen, P. J. K., and Heiser, W. J. (2005). Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika* **70**(1), 71–98.
- Carroll, J. D. and Chang, J. J. (1970). Individual differences and multidimensional scaling via an N -way generalization of Eckart-Young decomposition. *Psychometrika* **35**(3), 282–319.
- Coombs, C. H. (1964). *A Theory of Data*. New York: Wiley.
- Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**(1), 1–27.
- Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **29**(2), 115–129.
- Ramsay, J. O. (1977). Maximum likelihood estimation in multidimensional scaling. *Psychometrika* **42**(2), 241–266.
- Ramsay, J. O. (1982). Some statistical approaches to multidimensional scaling data. *Journal of the Royal Statistical Society, Series A (General)* **145**(3), 285–312.
- Shepard, R. N. (1962). Analysis of proximities: Multidimensional scaling with an unknown distance function, I and II. *Psychometrika* **27**(2,3), 125–140 and 219–246.
- Takane, Y. (2007). Applications of multidimensional scaling in psychometrics. In Rao, C. R. and Sinhar, S. (eds.) *Handbook of Statistics*, vol. 26, Psychometrics, pp 359–400. Amsterdam: Elsevier B. V.
- Takane, Y., Young, F. W., and de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika* **42**(2), 7–67.
- Takane, Y., Jung, S., and Oshima-Takane, Y. (2009). Multidimensional scaling. In Millsap, R. E. and Maydeu-Olivares, A. (eds.) *Handbook of Quantitative Methods in Psychology*, pp 219–242. London: Sage.

Further Reading

- Carroll, J. D. and Arabie, P. (1998). Multidimensional scaling. In Birnbaum, M. H. (ed.) *Handbook of Perception and Cognition*, vol 3: *Measurement, Judgment and Decision Making*, pp 179–250. San Diego, CA: Academic Press.
- Cox, T. F. and Cox, M. A. A. (2001). *Multidimensional Scaling*. Boca Raton, FL: Chapman and Hall.
- Everitt, B. S. and Rabe-Hesketh, S. (1997). *The Analysis of Proximity Data*. New York: Wiley.
- Kruskal, J. B. and Wish, M. (1978). *Multidimensional Scaling*. Beverly Hills, CA: Sage.
- Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology* **43**, 25–53.
- Schiffman, S. S., Reynolds, M. L., and Young, F. W. (1981). *Introduction to Multidimensional Scaling*. New York: Academic Press.
- Young, F. W. and Hamer, R. M. (1994). *Theory and Applications of Multidimensional Scaling*. Hillsdale, NJ: Erlbaum.

Relevant Website

<http://ego.psych.mcgill.ca/> – MULTISCALE: The MS-DOS program.

Multiple Comparisons

C Lewis, Fordham University, New York City, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bayes rule – A decision procedure that minimizes the posterior expected loss.

Bonferroni inequality – The probability of at least one event in a family occurring is less than or equal to the sum of the separate probabilities that the individual events will occur.

Contrast – A weighted linear combination of parameters (usually group means), with the sum of the weights constrained to sum to zero.

Coverage rates – The measures of how often a family of confidence intervals includes their intended parameters or contrasts among parameters. Per comparison, per family, and familywise coverage rates may be defined analogously to the corresponding error rates.

Directional error rates – Any of the standard multiple comparisons error rates (per comparison error rate (PCE), per family error rate (PFE), familywise error rate (FWE), or false detection rate (FDR)) with Type I errors replaced by Type III errors.

False detection rate (FDR) – The average of the ratio of the number of Type I errors for a family of comparisons to the number of rejected null hypotheses for that family. (For replications with no rejections, this ratio is defined to be zero.)

Familywise error rate (FWE) – The probability of making at least one Type I error for a family of comparisons.

Per comparison error rate (PCE) – The average number of Type I errors for a family of comparisons, divided by the number of comparisons in the family.

Per family error rate (PFE) – The average number of Type I errors for a family of comparisons.

Type III error – The incorrect identification of the sign of a parameter or of a contrast among parameters.

It is a common situation in educational research to be making statistical inferences based on models that have more than a single parameter. The researcher's interest may be in trying to answer questions about the values of these parameters or to make comparisons among them. By its nature, statistical inference involves the possibility of error, and it is an inescapable fact that, when more inferences are made, more errors are likely to be made. This unsettling state of affairs is known as the multiple

comparisons problem. Although it was published more than 20 years ago, the best general source on multiple comparisons remains the book by Hochberg and Tamhane (1987). Other very useful resources on the subject include Hsu (1996), Miller (1981), Shaffer (1995), and Westfall *et al.* (1999).

Early studies of multiple comparisons were made by Duncan (1955), Dunnett (1955), Fisher (1935), Lehmann (1957a, 1957b), Roy and Bose (1953), Scheffé (1953), and Tukey (1953). These authors and others (see Harter (1980), for the early history of the subject) focused on the problem of multiple comparisons in the context of the analysis of variance (ANOVA). Specifically, in this setting, the comparisons are made among several group or treatment means (denoted by μ_i) for a response variable of interest. These comparisons typically take the form of statistical tests of multiple null hypotheses. For instance, it is common to consider testing $H_0: \mu_i = \mu_j$ for all pairs of group response means.

After an introduction to different measures of error rates for multiple comparisons, a review of some of the better-known procedures for carrying out multiple hypothesis tests comparing means is provided. Next, an example is introduced to illustrate the use of these procedures. Subsequent sections address the issue of directional inferences, the use of Bayesian methods for multiple comparisons, and the topic of simultaneous confidence intervals. Finally, the need for application of multiple comparisons procedures in settings beyond the analysis of variance is illustrated with an example making inferences about the elements of a correlation matrix.

Error Rates

A number of different measures of Type I errors have been considered for use with multiple hypothesis tests. The simplest of these is the per comparison error rate (PCE). It is defined as the average number (over replications) of Type I errors, divided by the total number of comparisons being made in some family of interest. If each null hypothesis for the family of comparisons is tested at some fixed level α (such as 0.05), then the PCE will be at most α (when all null hypotheses are true).

A second measure of Type I errors to be used with multiple comparisons is known as the per family error rate (PFE). For a specified family of comparisons, this is defined as the average number of Type I errors. If each of

k comparisons is tested at a fixed level α , then the PFE will be at most $k\alpha$ (again, when all tested null hypotheses are true). It must always be the case that $PCE \leq PFE$, regardless of how the comparisons are tested.

The most widely used measure of Type I errors for multiple comparisons is called the familywise error rate (FWE). It is defined as the probability of making at least one Type I error for a family of comparisons. If the tests for k comparisons are stochastically independent, and each is carried out at a fixed level α , then the FWE will be at most $1-(1-\alpha)^k$, a bound once again achieved when all the tested null hypotheses are true. If the multiple tests are not independent, there is no simple expression for the FWE. However, the Bonferroni inequality guarantees that, for any family of comparisons, the FWE can never exceed the PFE: $FWE \leq PFE$. (The Bonferroni inequality says that the probability of at least one event in a family occurring is less than or equal to the sum of the separate probabilities that the individual events will occur. Since this latter sum equals the average number of occurrences, the stated result follows immediately.)

The three error rates considered so far were all discussed by Tukey (1953). More recently, a fourth measure, known as the false detection rate (FDR), was introduced by Benjamini and Hochberg (1995). The FDR is equal to the average (over replications) of the ratio of the number of Type I errors to the number of rejected null hypotheses. (For replications with no rejected null hypotheses, this ratio is defined to be zero.) One may think of the FDR as a compromise between the PFE and the PCE. It approaches the PCE when most hypotheses are rejected on most replications. On the other hand, when few hypotheses are rejected on most replications, it approaches the PFE. It will always be the case that $PCE \leq FDR \leq PFE$.

When there is only one comparison, the corresponding test is carried out at level α , and the null hypothesis being tested is true, then $PCE = PFE = FWE = FDR = \alpha$. In this sense, all four multiple comparison error rates are generalizations of the single comparison Type I error rate.

When a family of tests is carried out, each at level α , the PCE is controlled at α for all possible configurations of true and false null hypotheses. This is not generally true for any of the other error rates. The procedures that generally control the PFE at α are those that test comparison i at level α_i with $\sum \alpha_i = \alpha$. The simplest version of this approach for a family of k comparisons sets all $\alpha_i = \alpha/k$. With m group means and a family of all pairwise comparisons, $k = m(m-1)/2$. This is often referred to as the Bonferroni procedure. Note that to demonstrate control of the PFE for all configurations of true and false null hypotheses with these procedures, no reference to the Bonferroni inequality is needed. Using the Bonferroni inequality, $FWE \leq PFE \leq \alpha$, this procedure also controls the FWE at α . The Bonferroni procedure was first proposed by Fisher (1935).

Some Multiple Comparison Procedures

Many multiple comparison procedures (MCPs) have been proposed to control the FWE at a given level α for all configurations of true and false null hypotheses for a given family of comparisons. One method that provides this control for some but not all such configurations is known as the least significant difference (LSD) procedure. It was first proposed by Fisher (1935).

The LSD MCP is a two-step procedure. In the first step, all null hypotheses in the family are tested with a single, omnibus test, such as the F -test in the one-way ANOVA setup. If this test does not reject the overall null hypothesis at level α , then stop. If it does reject, then proceed to the second step: Carry out individual tests for the family of comparisons, each at level α .

If all null hypotheses are true, the omnibus test rejects with probability α . Therefore, the probability of at least one Type I error for an individual null hypothesis in this case is at most α . In other words, when all null hypotheses are true, the LSD procedure controls the FWE at α .

Now consider an example from the one-way ANOVA. Suppose there are ten groups with 50 observations per group. Let nine of the means be identical, with one very different from the rest. Thus, the overall F -test, using 9 and 40 degrees of freedom, (almost) always rejects its null hypothesis when used at level $\alpha = 0.05$. Applying the LSD procedure, this means that $10(9)/2 = 45$ t -tests are always carried out. Of these, nine will always correctly reject their null hypotheses. The remaining 36 tests each reject their (true) null hypothesis with probability 0.05, based on a critical t -value of $t_{0.975}(490) = 1.965$. (The 490 degrees of freedom used here assumes that the pooled within groups variance estimate is used for all tests.) The probability that the largest t -statistic in absolute value will exceed this quantity is given by the studentized range (for nine means and 490 degrees of freedom) tail probability associated with $t_{0.975}(490)\sqrt{2} = 2.779$, namely $p = 0.569 = FWE$. (For more on the LSD, see Hayter, 1986.)

The failure of the LSD procedure to control the FWE for all configurations of means has led to the general advice that should not be used to make multiple comparisons. Other early proposals that display a similar failure to control the FWE for all configurations of means include Duncan's multiple range and the Newman-Keuls procedures. (See Hochberg and Tamhane, 1987.)

Two early proposals (in addition to the Bonferroni procedure) that do provide general control of the FWE are those of Scheffé and Tukey. For the one-way ANOVA, Scheffé's procedure controls the FWE for the (infinitely large) family of linear contrasts. If μ_i , $i = 1, \dots, m$, denote the group means and c_i , $i = 1, \dots, m$, with $\sum_{i=1}^m c_i = 0$ are

specified constants, then $\psi = \sum_{i=1}^m c_i \mu_i$ is a contrast. Note that pairwise differences $\mu_i - \mu_j$ are special cases of contrasts, with the constants $c_i = 1$, $c_j = -1$, and all the other constants equal to 0.

To apply Scheffé's procedure to a contrast ψ of interest, the following F -statistic is computed:

$$F_\psi = \frac{\left(\sum_{i=1}^m c_i \bar{y}_i \right)^2}{\left(\sum_{i=1}^m \frac{c_i^2}{n_i} \right) \hat{\sigma}_w^2}$$

where \bar{y}_i is the sample mean and n_i is the sample size for group i , and $\hat{\sigma}_w^2$ is the pooled within group variance estimate. This statistic is then compared with $(m-1)F_{1-\alpha} \left(m-1, \sum_{i=1}^m n_i - m \right)$. If the statistic exceeds this critical value, then the null hypothesis that $\psi = \sum_{i=1}^m c_i \mu_i = 0$ is rejected. Scheffé's procedure has the attractive feature of compatibility with the overall F -test, in the sense that the overall test leads to rejection of its null hypothesis if and only if Scheffé's procedure finds at least one significant contrast.

Tukey's multiple comparison procedure makes use of the studentized range distribution to control the FWE for making comparisons among means in a one-way ANOVA setup. Best known in the form used to make all pairwise comparisons with equal group sample sizes, Tukey's MCP may also be used for general comparisons in the equal sample size case and for all pairwise comparisons with unequal sample sizes (in which form it is called the Tukey–Kramer procedure). Here, only the Tukey–Kramer MCP is described. It has been shown (Hayter, 1984) that this procedure conservatively controls the FWE at α for all configurations of means in the one-way ANOVA setup.

The test statistic used with the Tukey–Kramer procedure to test all hypotheses of the form $H_0: \mu_i = \mu_j$ is given by

$$Q_{ij} = \frac{|\bar{y}_i - \bar{y}_j|}{\sqrt{\frac{\hat{\sigma}_w^2}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}}$$

This statistic is compared with the critical value

$$Q_{1-\alpha} \left(m, \sum_{i=1}^m n_i - m \right),$$

namely the upper 100α percentile of the studentized range distribution for m means and

$$\sum_{i=1}^m n_i - m$$

degrees of freedom. Note that the test statistic Q_{ij} is equal to the absolute value of the usual t -statistic for comparing the means of groups i and j , multiplied by $\sqrt{2}$.

Many other MCPs have been proposed in the more than 50 years since Scheffé and Tukey introduced their methods. The interested reader should consult Hochberg and Tamhane (1987) for more details. Here, only one additional procedure is described, which was proposed by Benjamini and Hochberg (1995) to control the FDR.

The Benjamini–Hochberg procedure begins by obtaining p -values for all the statistics used to test a family of k hypotheses. (Denote the p -value for the i th hypothesis by p_i .) Next, these p_i 's are ordered with $p_{(1)} \leq \dots \leq p_{(k)}$. Define i_{crit} to be the largest value of $i = 1, \dots, k$, such that $p_{(i)} \leq i\alpha/k$, with $i_{\text{crit}} = 0$ if no such value exists. The null hypotheses associated with the i_{crit} smallest p -values should be rejected. Formally, this MCP has only been shown to control the FDR at α under special circumstances (such as when used with independent test statistics), but numerous simulations have supported the conjecture that it controls the FDR in more general settings.

An Example Applying MCPs

To compare the different MCPs that have been introduced, consider the example mentioned earlier: a one-way ANOVA with 10 groups and 50 observations per group. The family of interest consists of all 45 pairwise comparisons among the group means. To compare the procedures discussed here, they will all be translated into critical values for the usual t -tests (using the pooled within-groups variance from the overall analysis).

First, there is the PCE-controlling procedure, discussed earlier in connection with the LSD. (In what follows, it will be referred to as the LSD, even though the initial F -test is not used.) Here, the critical value needed to control the PCE at $\alpha = 0.05$ is just the 100(1- α /2) percentile of the Student's t distribution with 490 degrees of freedom, namely $t_{0.975}(490) = 1.965$. For the Bonferroni procedure, to control the PFE at $\alpha = 0.05$, the critical value needed for the t statistics is the 100(1- α /90) percentile of the same distribution, or $t_{0.99944}(490) = 3.475$.

Next, consider using Scheffé's procedure to control the FWE at $\alpha = 0.05$. This MCP uses an F -statistic for each pairwise comparison that is just the square of the t -statistic used by the LSD and Bonferroni procedures. In this example, the F -statistics are compared to $9F_{0.95}(9,490) = 9(1.899) = 17.091$. This is equivalent to comparing a t -statistic to the critical value $\sqrt{17.091} = 4.134$.

When Tukey's procedure is applied to control the FWE for all pairwise comparisons at $\alpha = 0.05$, the following critical value from the studentized range distribution

Table 1 A comparison of four MCPs applied to the problem of making all pairwise comparisons in a one-way ANOVA setup with 10 groups and 50 observations per group

Procedure	Critical t -value	Critical p -value	$\alpha = 0.05$ Control	Power for $d = 0.5$	Power for $d = 0.8$
LSD	1.965	0.05000	PCE	0.7037	0.9789
Tukey	3.178	0.00158	FWE	0.2507	0.7939
Bonferroni	3.475	0.00056	PFE (FWE)	0.1667	0.6997
Scheffé	4.134	0.00004	FWE	0.0528	0.4479

is needed: $Q_{0.95}(10, 490) = 4.495$. As noted earlier, this may be translated into comparing the pairwise t -statistics to $Q_{0.95}(10, 490)/\sqrt{2} = 4.495/1.414 = 3.178$.

Table 1 summarizes the results for this example. In addition to the calculations already given above, the p -value associated with each critical t -value is included. The powers of each procedure to reject a null hypothesis with effect sizes of $d = 0.5$ (Cohen's medium effect) and $d = 0.8$ (Cohen's large effect) are also included in the table (Cohen, 1988). Note the substantial drop in power that is the price paid for controlling the familywise Type I error rate for the family of pairwise comparisons, rather than simply controlling the PCE. Also note the modest power advantage enjoyed by the Tukey procedure over the Bonferroni and by the Bonferroni over the Scheffé procedure.

Finally, to apply the Benjamini–Hochberg FDR-controlling procedure (with $\alpha = 0.05$) in this setting, the pairwise difference t -statistic with the smallest absolute value would be compared to the LSD critical value of 1.965. (Equivalently, the corresponding largest p -value would be compared to 0.05.) This process would continue, comparing the ordered absolute values of the t -statistics to successively larger critical values (or the ordered p -values to successively smaller critical p -values), until the first significant result is obtained. At this point, that and all the remaining null hypotheses would be rejected. If the end of the list is reached with none of the t -statistics exceeding their respective critical values (including the largest one not exceeding the Bonferroni critical value of 3.475), then none of the pairwise null hypotheses would be rejected.

Directional Error Rates

In critical discussions of statistical hypothesis testing, many researchers have expressed dissatisfaction with the focus on rejecting a point null hypothesis. One relatively recent example is found in the paper by Jones and Tukey (2000). These authors insist that a point null hypothesis for a contrast is unrealistic in the sense that it can never be exactly true, and that incorrectly identifying the sign of the contrast, sometimes called a Type III error (Kaiser, 1960), should be of much greater concern than making a Type I error. An early treatment of directional errors in a general hypothesis testing framework is given by Lehmann (1950).

Lehmann (1957a, 1957b) and Bohrer (1979) provide useful early discussions of directional error rates for multiple comparisons.

Williams *et al.* (1999) apply this focus on directional errors to propose a modification of the FDR and the Benjamini–Hochberg procedure. Specifically, they redefine the FDR to be the average over replications of the proportion of sign declarations that are incorrect. They show that controlling the usual FDR at α is equivalent to controlling the directional FDR at $\alpha/2$ and advocate this as a general multiple comparisons strategy. Shaffer (2002) provides a general discussion of the issue of directional error rates and multiple comparisons.

Bayesian MCPs

So far, all the discussion of multiple comparisons has taken place in the framework of sampling theory inference and, specifically, using hypothesis testing within that framework. Duncan (1961, 1965) and Waller and Duncan (1969) provided an early treatment of multiple comparisons using Bayesian decision theory. Their focus was on directional errors for pairwise comparisons in what sampling theory would identify as a random-effects ANOVA setup.

Shaffer (1999) showed through a simulation study that Waller and Duncan's (1969) procedure provides control over a random-effects version of the directional FDR. (For a random-effects FDR, the FDR is averaged over sampling of true means as well as sampling of observed data.) This result is supported and extended in a paper by Lewis and Thayer (2004) where a Bayes rule based on a 0–1 loss function (rather than the linear one used in Waller and Duncan, 1969) is proved to control the random effects, directional FDR for the case of all pairwise comparisons. (A Bayes rule is a decision procedure that minimizes the posterior expected loss.) Lewis and Thayer (2009) continued this line of research, introducing a loss function closely related to the directional FDR and deriving a Bayes rule for this function. They showed through the use of simulations that, in the case of a fixed-effects ANOVA setup, their rule functions very similarly to the sampling theory procedure of Benjamini and Hochberg (1995) as modified for directional decisions by Williams *et al.* (1999).

Before leaving the topic of Bayesian multiple comparisons, it may be worth noting that some of the traditional MCPs have a natural translation in Bayesian terms once a directional approach to hypothesis testing is adopted. For example, consider the directional version of the Bonferroni procedure applied to posterior probabilities. The posterior expectation of the number of incorrect sign identifications for this procedure will be at most $\alpha/2$. By the Bonferroni inequality, the corresponding posterior probability of at least one incorrect sign identification will also be controlled at $\alpha/2$ by this MCP.

Multiple Confidence Intervals

Although the discussion of multiple comparisons so far has focused on hypothesis testing, several of the earliest writers on the subject (including Roy and Bose, 1953; Scheffé, 1953; and Tukey, 1953) also consider the problem of constructing multiple confidence intervals. Corresponding to the different error rates discussed for multiple hypothesis tests, there are different coverage rates that may be considered for multiple intervals. Thus, the per comparison coverage rate is defined as the average number of intervals containing the true value of the associated contrast, divided by the number of contrasts in the family. The per family coverage rate is simply the average number of intervals containing their contrast. A more easily interpretable quantity is the average number of intervals not containing their contrast. Finally, the familywise coverage rate is defined as the probability that all intervals in a family contain their contrasts.

In the case of a one-way ANOVA, the form of a confidence interval for a contrast $\psi = \mu_i - \mu_j$ is typically

$$\left(\bar{y}_i - \bar{y}_j - t_{\text{crit}} \sqrt{\hat{\sigma}_w^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}, \bar{y}_i - \bar{y}_j + t_{\text{crit}} \sqrt{\hat{\sigma}_w^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \right).$$

Returning to the specific example discussed earlier (10 groups, 50 observations per group, family of all 45 pairwise comparisons), we can choose values for t_{crit} from **Table 1**. If we choose the value associated with the LSD procedure, namely $t_{\text{crit}} = 1.965$, the resulting intervals would control the per comparison coverage at the 0.95 level. If we take $t_{\text{crit}} = 3.178$, using the value for Tukey's MCP, the intervals would have a 0.95 familywise coverage. In other words, 95% of all replications would have all 45 Tukey intervals covering their contrasts.

The same coverage result holds (conservatively) for the wider intervals based on $t_{\text{crit}} = 3.475$ (the Bonferroni value) or $t_{\text{crit}} = 4.134$ (the Scheffé value). For the Bonferroni intervals, it is also true that the average number of intervals not containing their contrasts is controlled at 0.05. For the Scheffé intervals, the family may also be extended to all contrasts (not just pairwise differences) without losing its familywise coverage of 0.95.

Generalizations to Other Settings

Although almost all the work in multiple comparisons has taken place in the context of comparing means in the traditional ANOVA, it is clear that the issue has much broader implications for statistical inference. For instance, Hochberg and Tamhane (1987) include a chapter on distribution-free and robust procedures, and have sections on categorical data analysis as well as the comparison of variances. To conclude this treatment of the subject, consider the problem of making inferences about the correlations among a set of variables.

Table 2 presents a correlation matrix from an unpublished study, including all bivariate Pearson product-moment correlations among a set of 12 variables (labeled X11 through X62), based on $N = 113$ complete data records. (Actually, these are six variables, each observed at two time points.) The contents of the table are slightly edited output from a standard package of statistical programs. In particular, the output assigns stars to significant correlations: one star if significant at the 0.05 level, and two stars if significant at the 0.01 level. It should be noted that these significance levels refer to PCEs – per comparison Type I error rates.

With 12 variables, there are 66 distinct correlations. If all 66 population correlations were actually equal to zero and a PCE of 0.05 were adopted, then the PFE would be 3.3. In other words, an average of more than three correlations would be erroneously identified as significantly different from zero. Using a PCE of 0.01, the PFE would be 0.66.

The simplest MCP to apply to a set of correlations is the Bonferroni. To control the PFE (and, thus, the FWE) at 0.05, each correlation should only be judged to be significantly different from zero if the corresponding p -value is less than $0.05/66 = 0.0007576$. Based on a t -test with 111 degrees of freedom, this corresponds to declaring as significant only correlations with $|r| > 0.312$. These correlations are indicated with bold print in **Table 2**. Note that, while 31 of the 66 correlations in **Table 2** are significant using a PCE of 0.05, only 20 remain significant using a PFE of 0.05.

Finally, it may be instructive to consider the application of the Benjamini–Hochberg MCP to this problem. (It should be noted that there is no formal guarantee that this procedure will control the FDR for a family of correlations.) To carry out this procedure, it may be easiest to copy the correlations into a single column of a spreadsheet, compute the corresponding p -values (to greater precision than standard software provides), order these p -values, and, starting with the largest, compare them to the critical values $i(0.05/66)$, where i gives the rank (with $i = 66$ being the largest and $i = 1$ the smallest). When this is done, the first p -value that is less than its corresponding critical value is $p = 0.01978$, corresponding to $|r| = 0.219$. This is the 28th p -value, so the corresponding critical value is $28(0.05/66) = 0.02121$.

Table 2 Sample correlation matrix for 12 variables used to illustrate the generality of the multiple comparisons issue

	<i>X11</i>	<i>X12</i>	<i>X21</i>	<i>X22</i>	<i>X31</i>	<i>X32</i>	<i>X41</i>	<i>X42</i>	<i>X51</i>	<i>X52</i>	<i>X61</i>	<i>X62</i>
X11	1.000	0.486**	0.062	0.097	−0.071	0.047	−0.081	−0.026	−0.027	0.185	−0.176	−0.199*
X12	0.486**	1.000	0.079	0.084	0.049	0.080	−.042	0.039	−0.064	0.179	−0.117	−0.165
X21	0.062	0.079	1.000	0.599**	−0.398**	−0.377**	0.735**	0.612**	0.213*	0.220*	−0.161	−0.102
X22	0.097	0.084	0.599**	1.000	−0.349**	−0.518**	0.683**	0.843**	0.110	0.302**	−0.105	−0.053
X31	−0.071	0.049	−0.398**	−0.349**	1.000	0.700**	−0.497**	−0.402**	−0.264**	0.233*	0.149	0.074
X32	0.047	0.080	−0.377**	−0.518**	0.700**	1.000	−0.507**	−0.546**	−0.219*	−0.308**	0.002	−0.014
X41	−0.081	−0.042	0.735**	0.683**	−0.497**	−0.507**	1.000	0.836**	0.206*	0.250**	−0.054	−0.025
X42	0.026	0.039	0.612**	0.843**	−0.402**	−0.546**	0.836**	1.000	0.130	0.341**	−0.024	0.036
X51	−0.027	−0.064	0.213*	0.110	−0.264**	−0.219*	0.206*	0.130	1.000	0.526**	−0.327**	−0.051
X52	0.185	0.179	0.220*	0.302**	−0.233*	−0.308**	0.250**	0.341**	0.526**	1.000	−0.227*	−0.097
X61	−0.176	−0.117	−0.161	−0.105	0.149	0.002	−0.054	−0.024	−0.327**	−0.227*	1.000	0.375**
X62	−0.199*	−0.165	−0.102	−0.053	0.074	−0.014	−0.025	0.036	−0.051	−0.097	0.375**	1.000

**Correlation is significant at the 0.01 level (two-tailed).

*Correlation is significant at the 0.05 level (two-tailed).

Bold correlations significant at the PFE 0.05 level (two-tailed) – Bonferroni.

Italic and bold correlations significant at the FDR 0.05 level (2-tailed) – B-H.

Listwise *N* = 113.

Consequently, all correlations having p -values smaller than 0.02121 are declared significant by the Benjamini–Hochberg procedure. There are an additional eight correlations (beyond those identified by the Bonferroni procedure) that are so declared. They are shown in italics in Table 2. Here, the additional significant results obtained when trying to control the FDR rather than the PFE is clearly illustrated. The Benjamini–Hochberg procedure takes advantage of the fact that there are a number of highly significant correlations in this family to declare still more correlations significant. (Remember that the FDR is the average proportion of Type I errors among the significant results.)

To make a connection with the other topics treated in this article, it should be noted that controlling the PFE with the Bonferroni procedure at the 0.05 level has the effect of controlling the directional PFE at the 0.025 level. Similarly, controlling the Type I FDR at 0.05 has the effect of controlling the directional FDR at 0.025. Working with Bayesian posterior tail probabilities for the correlations instead of sampling theory p -values in this example, the same Bonferroni and Benjamini–Hochberg operations could be applied and would result in the control of the corresponding posterior directional error rates. Finally, a set of 66 simultaneous 95% confidence intervals (or Bayesian posterior intervals) could be computed for all the correlations in this example, based on the Bonferroni procedure, by producing individual $100(1-0.05/66) = 99.924\%$ intervals for each correlation.

See *also*: Analysis of Variance; Bayesian Statistical Analysis; Decision Theory; Hypothesis Testing and Confidence Intervals.

Bibliography

- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* **57**, 289–300.
- Bohrer, R. (1979). Multiple three-decision rules for parametric signs. *Journal of the American Statistical Association* **74**, 432–437.
- Cohen, J. (1988). *Statistical Power Analysis for the Social Sciences*, 2nd edn. Hillsdale, NJ: Erlbaum.
- Duncan, D. B. (1955). Multiple range and multiple F tests. *Biometrics* **11**, 1–42.
- Duncan, D. B. (1961). Bayes rules for a common multiple comparisons problem and related Student- t problems. *Annals of Mathematical Statistics* **32**, 1013–1033.
- Duncan, D. B. (1965). A Bayesian approach to multiple comparisons. *Technometrics* **7**, 171–222.
- Dunnnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association* **50**, 1096–1121.
- Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.
- Harter, H. L. (1980). History of multiple comparisons. In Krishnaiah, P. R. (ed.) *Handbook of Statistics*, vol. 1, pp 617–622. Amsterdam: North-Holland.
- Hayter, A. J. (1984). A proof of the conjecture that the Tukey–Kramer multiple comparisons procedure is conservative. *Annals of Statistics* **12**, 61–75.
- Hayter, A. J. (1986). The maximum familywise error rate of Fisher's least significant difference test. *Journal of the American Statistical Association* **81**, 1000–1004.
- Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures*. New York: Wiley.
- Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. Boca Raton, FL: CRC Press.
- Jones, L. V. and Tukey, J. W. (2000). A sensible formulation of the significance test. *Psychological Methods* **5**, 411–414.
- Kaiser, H. F. (1960). Directional statistical decisions. *Psychological Review* **67**, 160–167.
- Lehmann, E. L. (1950). Some principles of the theory of testing hypotheses. *Annals of Mathematical Statistics* **21**, 1–26.
- Lehmann, E. L. (1957a). A theory of some multiple decision problems. I. *Annals of Mathematical Statistics* **28**, 1–25.
- Lehmann, E. L. (1957b). A theory of some multiple decision problems. II. *Annals of Mathematical Statistics* **28**, 547–572.
- Lewis, C. and Thayer, D. T. (2004). A loss function related to the FDR for random effects multiple comparisons. *Journal of Statistical Planning and Inference* **125**, 49–58.
- Lewis, C. and Thayer, D. T. (2009). Bayesian decision theory for multiple comparisons. *Institute of Mathematical Statistics Lecture Notes – Monograph Series. Optimality: The Third Erich L. Lehmann Symposium* **57**, 326–332.
- Miller, R. G. Jr. (1981). *Simultaneous Statistical Inference*, 2nd edn. New York: Springer.
- Roy, S. N. and Bose, R. C. (1953). Simultaneous confidence interval estimation. *Annals of Mathematical Statistics* **24**, 513–536.
- Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika* **40**, 87–104.
- Shaffer, J. P. (1995). Multiple hypothesis testing: A review. *Annual Review of Psychology* **46**, 561–584.
- Shaffer, J. P. (1999). A semi-Bayesian study of Duncan's Bayesian multiple comparison procedure. *Journal of Statistical Planning and Inference* **82**, 197–213.
- Shaffer, J. P. (2002). Multiplicity, directional (Type III) errors and the null hypothesis. *Psychological Methods* **7**, 356–369.
- Tukey, J. W. (1953). The problem of multiple comparisons. In Braun, H. I. (ed.) *The Collected Works of John W. Tukey: Volume VIII. Multiple Comparisons: 1948–1983*, pp 1–300. New York: Chapman and Hall.
- Waller, R. A. and Duncan, D. B. (1969). A Bayes rule for symmetric multiple comparisons problems. *Journal of the American Statistical Association* **64**, 1484–1503.
- Westfall, P. H., Tobias, R. D., Rom, D., Wolfinger, R. D., and Hochberg, Y. (1999). *Multiple Comparisons and Multiple Tests Using SAS®*. Cary, NC: SAS Institute.
- Williams, V. S. L., Jones, L. V., and Tukey, J. W. (1999). Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of Educational and Behavioral Statistics* **24**, 42–69.

Multivariate Analysis of Variance

N T Longford, UPF, Barcelona, Spain

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Correlation – A scale-free measure of the dependence (association) of two variables.

Covariance – A measure of the dependence (association) of two variables.

Factor – A variable the values of which are (a small number of) categories.

Heteroscedastic – Adjective for ‘unequal variance.’

Homoscedastic – Adjective for ‘equal variance.’

Hypothesis testing – A data-based procedure for deciding whether to contradict or not to contradict a specific conjecture (hypothesis).

Multivariate – Involving several variables.

Normal distribution – A statistical distribution with a particular shape of the frequencies (density) characterized by the bell curve.

Replication – A repeated application of a data-generating process that is independent of all the previous application.

Variance – A measure of the dispersion of the values of a variable.

Introduction and Background

Many educational and behavioral phenomena have several interrelated aspects and cannot be satisfactorily observed by a single variable. When the outcome variables are correlated, studying them by analyzing one variable at a time is highly unsatisfactory; a single (simultaneous) analysis is preferred. Multivariate analysis of variance (MANOVA) is concerned with multivariate outcomes observed on subjects in groups.

MANOVA can be introduced as the obvious generalization of the analysis of variance (ANOVA) from a single to several outcome variables. In MANOVA, the vectors of outcome variables are assumed to have (possibly distinct) H -variate normal distributions in the categories or groups $k = 1, \dots, K$. The vectors have expectations $\boldsymbol{\mu}_k$, but identical variance matrices $\boldsymbol{\Sigma}$. In a typical data set, planned or realized, each of the *a priori* specified groups k is represented by a random sample from the corresponding

distribution $N_H(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$. The principal research questions relate to the expectations $\boldsymbol{\mu}_k$ and the differences $\boldsymbol{\mu}_{k_1} - \boldsymbol{\mu}_{k_2}$ ($k_1 \neq k_2$) or, more generally, to contrasts

$$c_1\boldsymbol{\mu}_1 + c_2\boldsymbol{\mu}_2 + \dots + c_K\boldsymbol{\mu}_K \quad [1]$$

where the scalars c_k are constants that add up to zero: $c_1 + \dots + c_K = 0$. A multivariate contrast is defined as

$$\mathbf{c}_1^T\boldsymbol{\mu}_1 + \mathbf{c}_2^T\boldsymbol{\mu}_2 + \dots + \mathbf{c}_K^T\boldsymbol{\mu}_K \quad [2]$$

where the $H \times 1$ vectors of constants $\mathbf{c}_1, \dots, \mathbf{c}_K$ are such that $\mathbf{c}_1 + \dots + \mathbf{c}_K = \mathbf{0}_H$, the $H \times 1$ vector of identical zeros. The (univariate) ANOVA is a special case of MANOVA, with $H = 1$.

MANOVA is much more complicated than the H ANOVAs applied separately to each outcome variable. First, unlike the scalars μ_1, \dots, μ_K in ANOVA, the vectors $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K$ in MANOVA need not have a well-defined ordering, say, from the lowest to the highest. It is meaningful to ask which group k has the smallest expectation in ANOVA, whereas in MANOVA there need not be a group k such that each component of $\boldsymbol{\mu}_k$ is smaller than the corresponding component of $\boldsymbol{\mu}_{k'}$ for any other group k' . Further, MANOVA assumes not only that each variable is homoscedastic with respect to the groups (has equal within-group variances); the within-group covariances are also identical. One approach to dealing with these complications is to apply ANOVA to the combination $\mathbf{d}^T\mathbf{Y}$ of the outcome vector \mathbf{Y} , for a range of $H \times 1$ vectors of constants \mathbf{d} . An unattractive feature of this proposal is that a continuum of vectors \mathbf{d} has to be considered.

Hypothesis Testing

In MANOVA, one is usually interested whether or not the K expectations $\boldsymbol{\mu}_k$ are identical. Hypothesis testing can be interpreted as regarding equality (homogeneity) of the expectations as the default, and rejecting it only when the data contain information sufficient to contradict homogeneity. Therefore, the equality cannot be confirmed. An alternative viewpoint rejects homogeneity outright, without any data analysis, on the grounds that equality is a special case of a continuum of possible differences among the expectations $\boldsymbol{\mu}_k$; in brief, it is a safe bet that some of the expectations differ.

MANOVA considers two models. In the first, general (unrelated) within-group expectations are assumed:

$$Y = \boldsymbol{\mu} + \boldsymbol{\Delta}_k + \epsilon \quad [3]$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Delta}_k$, $k = 1, \dots, K$, are (unknown) vectors of parameters. The $\boldsymbol{\Delta}_k$ represent the differences among the within-group mean vectors $\boldsymbol{\mu}_k = \boldsymbol{\mu} + \boldsymbol{\Delta}_k$. The second model is a submodel (special case) of the first, in which all vectors $\boldsymbol{\Delta}_k$ vanish. There is an indeterminacy in [3], as a change in $\boldsymbol{\mu}$, to $\boldsymbol{\mu} + \mathbf{a}$, can be compensated by the cancelling change in each $\boldsymbol{\Delta}_k$, to $\boldsymbol{\Delta}_k - \mathbf{a}$. This problem is resolved by declaring group 1 as the reference and setting $\boldsymbol{\Delta}_1 = \mathbf{0}$. Of course, another group can be selected as the reference (or the labels $1, \dots, K$ permuted accordingly). As an alternative, we may insist on a linear constraint, such as

$$\mathbf{D}_1 \boldsymbol{\Delta}_1 + \dots + \mathbf{D}_K \boldsymbol{\Delta}_K = \mathbf{d}$$

where \mathbf{D}_k are matrices of constants and \mathbf{d} a given vector.

The test of homogeneity (equality of the expectations $\boldsymbol{\mu}_k = \boldsymbol{\mu} + \boldsymbol{\Delta}_k$) is derived from the likelihood ratio statistic and is related to the within and between-group sums of squares. Let the respective estimators based on these statistics be

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_0 &= \frac{1}{n-1} \sum_{k=1}^K \sum_{i=1}^{n_k} (\mathbf{y}_{ik} - \hat{\boldsymbol{\mu}}) (\mathbf{y}_{ik} - \hat{\boldsymbol{\mu}})^T \\ \hat{\boldsymbol{\Sigma}}_1 &= \frac{1}{n-K} \sum_{k=1}^K \sum_{i=1}^{n_k} (\mathbf{y}_{ik} - \hat{\boldsymbol{\mu}}_k) (\mathbf{y}_{ik} - \hat{\boldsymbol{\mu}}_k)^T \end{aligned}$$

Assuming that the null-hypothesis is valid,

$$\begin{aligned} (n-1)\hat{\boldsymbol{\Sigma}}_0 &\sim W_H(\boldsymbol{\Sigma}, n-1) \\ (n-K)\hat{\boldsymbol{\Sigma}}_1 &\sim W_H(\boldsymbol{\Sigma}, n-K) \end{aligned} \quad [4]$$

where $W_H(\boldsymbol{\Sigma}, m)$ denotes the H -variate Wishart distribution with variance matrix $\boldsymbol{\Sigma}$ and m degrees of freedom. Note that this distribution is for $H \times H$ symmetric (positive-definite) matrices.

The two estimators in [4] are related by the decomposition

$$(n-1)\hat{\boldsymbol{\Sigma}}_0 = (n-K)\hat{\boldsymbol{\Sigma}}_1 + \sum_{k=1}^K n_k (\hat{\boldsymbol{\mu}}_k - \hat{\boldsymbol{\mu}})(\hat{\boldsymbol{\mu}}_k - \hat{\boldsymbol{\mu}})^T \quad [5]$$

The two terms on the right-hand side are independent; we say that this decomposition is orthogonal. Under the null hypothesis, the summation in [5] has Wishart distribution with variance matrix $\boldsymbol{\Sigma}$ and $K-1$ degrees of freedom. Operating with matrix-distributions is awkward; for example, it is very difficult to define a suitable rejection or confidence region. For the determinants of the sum-of-squares matrices $T = (n-1)\hat{\boldsymbol{\Sigma}}_0$ and $W = (n-K)\hat{\boldsymbol{\Sigma}}_1$, we have the following result:

$$\Lambda = \frac{\det(\mathbf{W})}{\det(\mathbf{T})}$$

has the Wilks' Λ distribution with parameters H , $n-K$, and $K-1$. The matrices \mathbf{W} and \mathbf{T} and $\mathbf{T} - \mathbf{W}$ have Wishart distributions with variance matrix $\boldsymbol{\Sigma}$ and respective degrees of freedom $n-K$, $n-1$, and $K-1$. Provided that $n \gg K$, the transformation

$$- \left\{ n - K - \frac{1}{2}(H - K + 2) \right\} \log(X)$$

of a random variable X with Wilks' Λ distribution has approximately χ^2 distribution with $H(K-1)$ degrees of freedom. A test of homogeneity rejects the null-hypothesis for large values of this statistic.

Fixed and Random Effects

Often a given set of groups $k = 1, \dots, K$ is considered in an analysis, and each group k has a well-identified label that transcends the data set that happens to have been collected. That is, in a (hypothetical) replication of the study, each $k = 1, \dots, K$ would have the same meaning as in the realized study, and the value of $\boldsymbol{\Delta}_k$ would be unchanged. We say that the vectors $\boldsymbol{\Delta}_k$ are fixed. Replications of the study would use these unchanging values of $\boldsymbol{\Delta}_k$.

When there are only a few groups, K , their labels are usually well recognized, and so it is natural to regard $\boldsymbol{\Delta}_k$ as fixed. When there are many of them, it is natural to regard them as a realization from a superpopulation of groups, and regard $\boldsymbol{\Delta}_k$ as random. Then we have to consider a design for sampling from this superpopulation; simple random sampling is the simplest choice. This implies that $\boldsymbol{\Delta}_1$ in one replication refers in general to different groups in two distinct replications, or that the labels $k = 1, \dots, K$ are assigned arbitrarily, with no relation to the original labels (names) that the groups might have. A group that is represented in one replication need not be represented in another replication at all. Therefore, any need to make inferences about $\boldsymbol{\Delta}_1$, or another group with an *a priori* set index, is specific to the replication, and could not have arisen prior to data collection. Descriptions of the collection (superpopulation) of the groups might then be more relevant. To aid this, we declare $\boldsymbol{\Delta}_k$ as random vectors with a multivariate normal distribution

$$\boldsymbol{\Delta}_k \sim N_H(0_H, \boldsymbol{\Sigma}_B) \quad [6]$$

independently both across k and from the subject-level deviations ϵ_{ik} . Setting the expectation in [6] to zero entails no loss of generality. The matrix $\boldsymbol{\Sigma}_B$ describes the pattern of variation of the group-level deviations.

Hypotheses of interest about $\boldsymbol{\Sigma}_B$ relate to its pattern; for example, that $\boldsymbol{\Sigma}_B = 0$, that $\boldsymbol{\Sigma}_B$ is of rank 1, or that $\boldsymbol{\Sigma}_B$ is diagonal. Note that these hypotheses entail some abuse of notation because multivariate normal distribution with a singular variance matrix is not well defined. However, the

hypotheses can be reformulated in terms of normal distributions with fewer dimensions, such as $\Delta_k = D\delta_k$, where D is a $H \times H'$ matrix and $H' < H$, or these singular (degenerate) distributions could be added to the definition of the multivariate normal distribution.

An apparent advantage of the random-effects MANOVA is that each Δ_k can be estimated with greater efficiency than by the sample mean $\hat{\boldsymbol{\mu}}_k$. The conditional distribution of Δ_k given the model parameters $\boldsymbol{\mu}$, Σ , and Σ_B and the data y is

$$(\Delta_k | y; \boldsymbol{\mu}, \Sigma, \Sigma_B) \\ \sim N_H \{ \boldsymbol{\mu} + \Sigma_B(\Sigma + n_k \Sigma_B)^{-1}(\hat{\boldsymbol{\mu}}_k - \boldsymbol{\mu}), n_k \Sigma_B(\Sigma + n_k \Sigma_B)^{-1} \Sigma_B \} \quad [7]$$

The expectation of this distribution can be estimated by replacing the parameters it involves with their estimates. The estimated (conditional) expectation is a suitable estimator of Δ_k ; we denote it by $\tilde{\Delta}_k$. Then $\hat{\boldsymbol{\mu}}_k = \hat{\boldsymbol{\mu}} + \tilde{\Delta}_k$ is a suitable estimator of $\boldsymbol{\mu}_k$.

The estimators $\tilde{\Delta}_k$ and $\hat{\boldsymbol{\mu}}_k$ can be interpreted as (multivariate) compositions of $\hat{\boldsymbol{\mu}}_k$ and $\hat{\boldsymbol{\mu}}$, with weights assigned to $\hat{\boldsymbol{\mu}}_k$ and $\hat{\boldsymbol{\mu}}$ that reflect their precisions. For example, if Σ_B (or $\hat{\Sigma}_B$) is very small, then $\tilde{\Delta}_k = \hat{\boldsymbol{\mu}}$. This should come as no surprise; since Σ_B is small, the vectors of expectations differ very little from one another, and so they differ very little from $\boldsymbol{\mu}$. Therefore, $\hat{\boldsymbol{\mu}}$ is a good estimator of $\boldsymbol{\mu}_k$. When Σ_B is large, in the sense of large variances and eigenvalues, the vectors $\boldsymbol{\mu}_k$ tend to differ a lot from one another, and so $\hat{\boldsymbol{\mu}}$ is a poor estimator of $\boldsymbol{\mu}_k$, and we cannot improve on $\hat{\boldsymbol{\mu}}_k$. In intermediate settings, choosing between $\hat{\boldsymbol{\mu}}_k$ and $\hat{\boldsymbol{\mu}}$ is not a good strategy; combining these two estimators, as done in $\hat{\boldsymbol{\mu}}_k$, is much more efficient. The univariate version of this approach is commonly referred to as borrowing strength across the groups. It amounts to exploiting the similarity of the groups.

Estimating Σ

The between-group variance Σ_B is estimated by moment matching. We form the average sum of squares of deviations

$$S_B = \frac{1}{K-1} \sum_{k=1}^K (\hat{\boldsymbol{\mu}}_k - \hat{\boldsymbol{\mu}})(\hat{\boldsymbol{\mu}}_k - \hat{\boldsymbol{\mu}})^T$$

and adjust it to make it unbiased for Σ_B . Standard manipulations yield the identity

$$E(S_B) = \Sigma_B + \frac{1}{K-1} \Sigma \sum_{k=1}^K g_k$$

where $g_k = 1/n_k - 1/n$. Hence, the unbiased estimator

$$\hat{\Sigma}_B = S_B - \frac{1}{K-1} \hat{\Sigma} \sum_{k=1}^K g_k$$

This estimator is problematic when the sample sizes n_k are very uneven. An alternative estimator

$$S_B^\dagger = \frac{1}{n} \sum_{k=1}^K n_k (\hat{\boldsymbol{\mu}}_k - \hat{\boldsymbol{\mu}})(\hat{\boldsymbol{\mu}}_k - \hat{\boldsymbol{\mu}})^T$$

takes the sample sizes into account, but the estimator it yields,

$$\hat{\Sigma}_B^\dagger = S_B^\dagger - \frac{D-1}{n} \hat{\Sigma}$$

is biased. The bias depends on the association of the sample sizes with the deviations $\boldsymbol{\mu}_k - \boldsymbol{\mu}$. The advantages and drawbacks of $\hat{\Sigma}_B$ and $\hat{\Sigma}_B^\dagger$ depend on the level of imbalance (variation of n_k), but are also highly contingent on the assumption of normality. A composition of the two estimators may be a good compromise, but the weights have to be chosen with care.

Multiway MANOVA

The variable that indicates the group to which a subject is assigned is called a factor. In a study, there may be several factors, such as age groups and the two sexes in a study of a human population. The consistent (systematic) differences among the levels of either of these two factors can be described by the model

$$y_{ikl} = \boldsymbol{\mu} + \Delta_k^{(1)} + \Delta_l^{(2)} + \varepsilon_{ikl} \quad [8]$$

where y_{ikl} is the vector of outcomes for subject $i = 1, \dots, n_{kl}$ in group $k = 1, \dots, K$ of the first factor and group $l = 1, \dots, L$ of the second factor, and ε_{ikl} are a random sample from $N(0_H, \Sigma)$. The groups (levels) k and l of the two factors are associated with respective parameter vectors $\Delta_k^{(1)}$ and $\Delta_l^{(2)}$. Lack of identifiability is dealt with by setting $\Delta_1^{(1)}$ and $\Delta_1^{(2)}$ to 0_H . The model in [8] assumes that the average difference between two groups of one factor, conditional on the value of the other factor (i.e., with the other factor fixed), does not depend on the latter. This assumption can be relaxed by introducing interactions of the two factors. This is equivalent to introducing a single factor with the KL categories formed by pairing every category of one factor with each category of the other factor.

Two factors A and B are said to be crossed if there is at least one category of A that contains units from at least two categories of B and there is at least one category of B that contains units from at least two categories of A. For example, in a typical setting, age groups and sex are crossed factors, because there are subjects from several (or all) categories of age group among both males and females and there are males and females in most (or all) age categories. Factor A is said to be nested within factor B, if each category of A is contained within a category of B. For example, in a typical country, district as a factor is

nested within region as a factor, because each district belongs to a region.

The design of a one-way MANOVA is said to be balanced if each group k has the same sample size n_k . The design of a two-way MANOVA with crossed factors is said to be balanced if each combination of categories k and l of the two factors has the same sample size n_{kl} . For a nested design, with factor A nested within factor B, balance amounts to the same number of categories of A within each category of B and identical sample sizes n_k for each category of A. Balanced designs are associated with efficiency and yield simple estimators.

Inference About the Pattern in Σ

Apart from inference about the within-group means μ_k and their counterparts in multi-way MANOVA, for which a lot of theory carries over from the univariate ANOVA, the covariance structure of within-group variance matrix Σ may also be of interest. For example, Σ may have the compound symmetry form

$$\Sigma = \sigma^2(\mathbf{I}_H + \rho \mathbf{1}_H \mathbf{1}_H^T)$$

where $\sigma^2 > 0$ and ρ are unknown parameters and \mathbf{I} is the identity matrix, so that all variances in Σ are equal to $(1 + \rho)\sigma^2$ and all covariances to $\rho\sigma^2$.

When the vector Y is a time series, its Markovian nature, the conditional dependence of any two components given the intervening observations,

$$Y_{b_1} \perp Y_{b_2} \mid Y_b,$$

for any indices $b_1 < b < b_2$, may be of interest. This pattern of dependence corresponds to tri-diagonal matrix Σ^{-1} .

The universal approach to testing the hypotheses related to these submodels is by the likelihood ratio test, which compares the values of the likelihood at their maxima for the hypothesis and the alternative. However, the counting of the degrees of freedom is problematic when an estimated variance matrix is singular, because the solution is on the boundary of the parameter space.

Example

The following problem has been studied in educational assessment. An educational test, comprising a small number of distinct sections (subtests), is administered in US institutions. These institutions are subscribers (volunteers), and each institution arranges for (some of) its students to take part. The purpose of the test is to assess the relative strengths and weaknesses of each institution in the subject areas represented by the subtests.

MANOVA models can be applied to this setting, with the institution-level vectors of mean subscores, and their contrasts, as the targets of inference. If these vectors are well ordered or their components highly correlated, an institution with a high mean score on one subtest has a high mean score also on the other. Reporting the (sample) subscore means then adds little or no information to reporting the overall mean. In two studies dealing with different testing programs, it was observed that the heterogeneity of the students is substantial, and therefore a large number of students from an institution is required to find any evidence that institutions have different patterns of the subscore means. Since these numbers exceed the numbers of enrolled students in most institutions, the tests are of doubtful value for the intended purpose. The voluntary nature of participation of students within institutions, which may lead to poor representation of the institution by the test-takers, raises further statistical challenges.

Other applications of MANOVA include analysis of growth and development, repeated measurements (as a way of combatting measurement error), and analysis of all characteristics and attributes that cannot be satisfactorily captured by a single variable.

Acknowledgment

This work was partly supported by the Grant SEJ2006-13537 from the Spanish Ministry of Science and Technology.

See also: Analysis of Variance; Assessment and the Evaluation of Institutional Effectiveness; Hypothesis Testing and Confidence Intervals; Matrix Algebra; Multivariate Normal Distribution; Validity of Educational Indicators.

Further Reading

- Anderson, T. W. (1984). Estimating linear statistical relationships. *Annals of Statistics* **12**, 1–45.
- Barker, H. R. and Barker, B. M. (1984). *Multivariate Analysis of Variance (MANOVA): A Practical Guide to Its Use in Scientific Decision-Making*. Tuscaloosa: University of Alabama Press.
- Boik, R. J. (1997). Analysis of repeated measures under second-stage sphericity: An empirical Bayes approach. *Journal of Educational and Behavioral Statistics* **22**, 155–192.
- Efron, B. and Morris, C. N. (1973). Stein's estimation rule and its competitors – an empirical Bayes approach. *Journal of the American Statistical Association* **68**, 117–130.
- Fay, R. E. (1987). Application of multivariate regression to small domain estimation. In Platek, R., Rao, J. N. K., Särndal, C. -E., and Singh, M. P. (eds.) *Small Area Statistics. An International Symposium* pp 91–102. 91–102. New York.
- Ghosh, M. (1997). Wishart distribution via induction. *American Statistician* **50**, 243–246.

- Haberman, S., Sinharay, S., and Puhon, G. (2008). Reporting subscores for institutions. *British Journal of Mathematical and Statistical Psychology* **61**, 79–95.
- Hand, D. J. and Taylor, C. C. (1987). *Multivariate Analysis of Variance and Repeated Measures*. London: Chapman and Hall.
- Kent, J. T., Mardia, K. V., and Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.
- Khatree, R. and Naik, D. N. (1990). A note on testing for interaction in a MANOVA model. *Biometrical Journal* **32**, 713–716.
- Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*, 3rd edn. Ch. 1–8. New York: Springer.
- Longford, N. T. (1990). Multivariate variance component analysis: An application in test development. *Journal of Educational Statistics* **15**, 91–112.
- Longford, N. T. (2005). Editorial: Model selection and efficiency: Is 'Which model . . .?' the right question? *Journal of the Royal Statistical Society A* **168**, 469–472.
- Longford, N. T. (2007). *Studying Human Populations. An Advanced Course in Statistics*, ch. 1. Springer, New York.
- Longford, N. T. (2008). An alternative analysis of variance. *SORT, Journal of the Catalan Statistical Institute* **32**, 77–92.
- Morrison, D. F. (1967). *Multivariate Statistical Methods*. New York: McGraw-Hill.
- Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A* **231**, 140–185.
- Novak, T. (1995). MANOVAMAP: Graphical presentation of MANOVA in marketing research. *Journal of Marketing Research* **32**, 357–374.
- O'Brien, R. G. and Kaiser, M. K. (1985). MANOVA method for analysing repeated measure designs. An extensive primer. *Psychological Bulletin* **97**, 316–333.
- Speed, T. P. (1987). What is analysis of variance? *Annals of Statistics* **15**, 885–910.
- Wilks, S. S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *Annals of Mathematical Statistics* **17**, 257–281.

Multivariate Linear Regression

S Ganesh, Massey University, Palmerston North, New Zealand

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Canonical correlation analysis – A technique to explain or summarize the relationship between two sets of variables by finding a linear combination of each set of variables that yields the highest possible correlation between the composite variable for set A and the composite variable for set B .

Dependent and independent variables –

Statistical models normally specify how one set of variables, called dependent variables, functionally depend on another set of variables, called independent variables. The term '(in)dependent' reflects only the functional relationship between variables within a model.

MANOVA – Multivariate analysis of variance, a generalized form of analysis of variance (ANOVA), is a technique which determines the effects of independent variables on multiple continuous dependent variables, and is generally used to compare several groups with respect to multiple continuous variables.

Method of least squares – A well-founded technique for estimating the unknown parameters in a linear regression model, based on minimization of the sum of squared differences between observed and modeled responses.

Multivariate regression – Multivariate (linear) regression is a technique aimed at finding a linear relationship between many response variables and multiple explanatory variables.

Regression analysis – It provides a best-fit mathematical equation for the relationship between the response variable(s) and regressor variable(s) or covariates.

Wilks' lambda – A general test statistic used in multivariate tests (e.g., MANOVA) based on a combination of dependent variables. Several other statistics are special cases of Wilks' lambda.

In many scientific studies, numerous variables (or measurements) are obtained for each individual or unit studied. Regression analysis is a statistical methodology for predicting the value of one (or more variables) from a collection of other variables. Regression models have been used extensively in the education literature (see, e.g., Hsu, 2005).

In the simplest case, we may be interested in studying the relationship between two sets of variables, say, Y_1, Y_2, \dots, Y_q and X_1, X_2, \dots, X_p , in particular, a linear relationship between them. This approach is the well-known linear-regression modeling, where we assume that the variables Y 's are dependent on the variables X 's. Although in a general linear modeling framework, both the Y 's and the X 's could be quantitative (or metric) and qualitative (or nonmetric), the most practical and commonly used case of linear-regression modeling assumes both the Y 's and the X 's to be quantitative in nature.

In statistical terminology, Y 's are called the response (or dependent or target or output) variables and X 's are known as the regressor (or explanatory or independent or predictor or input) variables.

We may consider three distinguished cases, also indicated by Rencher (2002), according to the number of response and regressor variables:

1. *Simple linear regression.* Here, the interest is on one response Y and one regressor X ; for example, predicting college freshman grade-point average (GPA) based on the student's SAT score.

Usually, a model is postulated by relating the response variable to the regressor variable with unknown parameters. A model that is linear in these parameters is called the simple linear regression model. This model, also referred to as a straight line model, is fitted via a technique known as the least squares and takes the form $Y = \beta_0 + \beta_1 X$, where β_0 and β_1 are the so-called intercept and slope parameters. The reader may read any intermediate-level statistics textbook for further details (e.g., Ott and Longnecker, 2001).

2. *Multiple linear regression.* Multiple regression analysis is used whenever we wish to model the relationship between one response variable and more than one regressor variable. In the preceding example, we could attempt to improve our prediction of college GPA by using, for example, high school GPA, Scholastic Aptitude Test (SAT) scores, and rating of school.

Many different forms of relationship are possible, but the overwhelming emphasis in practical applications is on the linear relationship $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$, known as the multiple linear regression model. The regression coefficients $\beta_0, \beta_1, \dots, \beta_p$ are model parameters whose values need to be estimated from the given data.

3. *Multivariate (multiple) linear regression.* Multivariate multiple regression analysis arises when we have more than one response variable, and we wish to model the relationship between these variables and a set of regressor variables. Attention is again focused almost exclusively on linear relationships. In the preceding example, we may wish to predict freshman college GPA (in the sciences, arts, and humanities) as well as the number of years of college the student will complete based on high school GPA, SAT, and rating of school.

Note that, in multivariate multiple linear regression, multivariate refers to the response variables and multiple represents the independent variables so that (2) above can be regarded as univariate linear regression. The focus of this article is on multivariate linear regression and the main aim is to give a detailed exploration of the topic including examples.

A mathematical formulation of the multiple linear regression model is provided in this article, with relationship to univariate (multiple) linear regression and the method of least squares. The underlying assumptions are discussed and model diagnostics carried out. This article also covers testing of hypotheses, ANOVA partitioning, and simultaneous confidence intervals. The application of the fitted models is illustrated via an example and suitable computer software.

Even though most of the basic theory behind the topic of interest, multivariate linear regression, is generally available in many textbooks and research articles, the following account is based on the seminal work of Khattree and Naik (1995) and Krzanowski (1998).

Formulation of the Multivariate Linear Regression Model

Suppose that we observe, say, q response variables Y_1, \dots, Y_q and p regressor variables X_1, \dots, X_p on, say, n sample individuals or units. Here, we shall focus on the behavior of response variables in terms of the regressor variables. The regressor variables are usually assumed to be measured without error, but the response variables are subject to measurement errors.

Furthermore, the regressor variables may be fixed or random. In the above examples, all X 's are random variables in that a student is selected at random and all the Y 's and X 's are measured for that individual (so, the variables are not under the control of the researcher). In some cases, especially in experimental situations, the X 's are fixed so that they are under the control of the researcher. For example, we may wish to relate, say, yield and the nutritional value of corn to various levels of a certain fertilizer and several varieties. Here, the amounts of fertilizer to be applied and the choices of variety are controlled by the

researcher who then observes the changes in the yield and nutritional responses. We shall interpret the regression model conditional on the type of regressor variables. A detailed account of modeling when the regressors are fixed and when they are random, is given by Rencher (2002).

In the univariate case, the linear regression model is defined by the following equation:

$$y_k = \beta_0 + \beta_1 X_{k1} + \beta_2 X_{k2} + \dots + \beta_p X_{kp} + \varepsilon_k \quad [1]$$

Here, y_k is the response for the k th individual in the given sample, X_{kj} is the value of the j th regressor variable for the k th individual, and ε_k is the random departure term for that individual representing the measurement errors. Since random sampling is assumed, the ε_k 's are taken to be independent and identically distributed random variables, each having mean zero and constant variance σ^2 .

If we collect together all the y_k , β_j , and ε_k into the vectors (with superscript T denoting matrix transposition) $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$, and write the regressor variable values in the $n \times (p+1)$ regressor matrix

$$\mathbf{X} = \begin{pmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{np} \end{pmatrix}$$

then, the above model [1] can be written in the compact form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad [2]$$

In this formulation, $\boldsymbol{\varepsilon}$ is a random vector whose mean is $\mathbf{0}$ ($n \times 1$ vector of zeroes) and whose dispersion (or covariance matrix) is $\sigma^2 \mathbf{I}_n$, \mathbf{I}_n being the $n \times n$ identity matrix, implying that the ε_k 's are identically and independently distributed with mean 0 and variance σ^2 .

In the multivariate situation, that is, when there are more than one response variable, we have to allow each Y_i to have its own linear relationship with all the X_j 's. Thus, we have to specify q different linear models, one between each Y_i and the set of X_j . By analogy with the above eqn [2], we can write these models as

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i \quad [3]$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in})^T$, $\boldsymbol{\beta}_i = (\beta_{i0}, \beta_{i1}, \dots, \beta_{ip})^T$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{in})^T$, for $i = 1, \dots, q$.

Here, y_{ik} is the value observed for the i th dependent variable on the k th sample unit, ε_{ik} is the error term corresponding to the i th dependent variable and k th individual, $(p+1) \times 1$ vector $\boldsymbol{\beta}_i$ are the parameters corresponding to the i th dependent variable, and \mathbf{X} is the same matrix as before.

In fact, the q equations in [3] can be represented compactly via the following matrix notation:

$$Y = XB + \xi \quad [4]$$

where the q vectors y_i are put side by side as columns into the $n \times q$ matrix Y , say, the response matrix, the q vectors β_i similarly into the $(p+1) \times q$ matrix B as columns, say, the parameter matrix, and the q column vectors ϵ_i into the $n \times q$ matrix ξ , say, the error matrix.

It is possible to treat each of the q equations in [3] independently of each other as a univariate linear regression. However, since the response variables may themselves be correlated with each other, this dependence should also be taken into account when fitting the multivariate linear regression model. This, indeed, suggests the need for using the model in eqn [4], where all the dependent variables are analyzed together. Even though the regression parameter estimates and their standard errors from the multivariate approach should be identical to those obtained from the set of univariate models, the added advantage with the multivariate regression model is that it allows for joint inferences involving two or more of the univariate models. For example, one may test the overall significance of the regression parameters (across all fitted univariate models) using a single statistical test. Furthermore, one may also test whether the coefficients of a particular regressor are the same across all fitted (univariate) models. Such hypotheses tests are explained later in this article.

Once again, the n sample individuals are assumed to be independent, but now there is an association among the error terms $\epsilon_{k1}, \dots, \epsilon_{kq}$ ($k = 1, \dots, n$) corresponding to the same individual. Consequently, we can treat the rows of the error matrix ξ as independent observations from a distribution with mean vector zero and dispersion matrix, say, Σ , and for most practical purposes, this distribution is assumed to be multivariate normal. In other words, while the i th column of ξ has the covariance matrix, say, $\sigma_{ii}^2 I_n$, I_n being the $n \times n$ identity matrix, a typical row of ξ has that as, say, $\sum = [\sigma_{ij}]$, a symmetric $q \times q$ matrix that is assumed to be positive definite, where given X , $\text{cov}(Y_i, Y_j) = \text{cov}(\epsilon_i, \epsilon_j) = \sigma_{ij} I_n$, for $i, j = 1, \dots, q$. The assumption of multivariate normality for error vectors is needed for hypothesis testing and construction of confidence regions, even though it is not needed for estimating parameters in the multivariate regression modeling.

Fitting the Multivariate Linear Regression Model via Least-Squares Estimation

Maximizing the likelihood for the model in eqn [4] under the normality condition produces an estimator of B given by

$$\hat{B} = (X^T X)^{-1} X^T Y \quad [5]$$

The same estimator results from the simple process of minimizing the criterion error sum of squares

(and products) via the well-known least-squares approach. To demonstrate, we wish to minimize,

$$\sum_{i=1}^q \epsilon_i^T \epsilon_i = \sum_{i=1}^q (y_i - X\beta_i)^T (y_i - X\beta_i)$$

where y_i , β_i , and ϵ_i are as defined before. This is equivalent to minimizing the trace (i.e., sum of diagonal elements) of the $p \times p$ matrix $(Y - XB)^T (Y - XB)$, which results in a set of normal equations,

$$(X^T X)B = X^T Y$$

Solving these equations yield the estimator \hat{B} , given by eqn [5], as the least squares estimator of parameter matrix B .

Note that, as in the univariate linear regression case, we assume that the regressor matrix X is of full rank, that is, $\text{rank}(X) = p+1$. This implies the existence of $(X^T X)^{-1}$ as $\text{rank}(X^T X) = \text{rank}(X)$. Although a generalized inverse of $(X^T X)$ may be used when X is not of full rank, extra care is needed when interpreting the regression results in such situations.

Moreover, since we may write $\hat{B} = [\hat{\beta}_0 \hat{\beta}_1 \dots \hat{\beta}_q]$, using appropriate columns from eqn [5],

$$\hat{\beta}_i = (X^T X)^{-1} X^T y_i \text{ for } i = 1, \dots, q \quad [6]$$

It is easy to observe that \hat{B} is an unbiased estimator of B , that is, $E(\hat{B}) = B$. Furthermore, we can show

$$\text{cov}(\hat{\beta}_i, \hat{\beta}_j) = \sigma_{ij} (X^T X)^{-1} \text{ for } i, j = 1, \dots, q \quad [7]$$

Equation [6] appears to imply that the regression coefficients in the multivariate linear regression model have the same estimates as they would if each response variable was regressed separately on the entire set of regressor variables, that is, when fitting univariate linear regression models. However, this may not be as simple as it sounds due to the fact that all the individual $\hat{\beta}_{ij}$ in \hat{B} are inter-correlated in the case of multivariate regression modeling. In other words, those $\hat{\beta}_{ij}$ within a column of \hat{B} are correlated because of possible correlations among the regressor variables, and those in different columns of \hat{B} because of the correlations among the response variables.

Hence, instead of performing statistical inference (hypothesis tests, etc.) separately on each β_i , that is, in a univariate regression scenario for response variable Y_i , we should perform multivariate tests of hypotheses, etc., about the regressor matrix B . This, in fact, would allow testing of hypotheses with respect to parameters within univariate models as well as between such models.

Statistical Inference

The hypothesis tests associated with the regression parameters in B are the analogs in multivariate analysis of

variance (MANOVA) of the analysis of variance (ANOVA) F -tests for univariate linear (multiple) regression. The general theory behind MANOVA is beyond the scope of this article, but can be found elsewhere in the encyclopedia and in many textbooks on multivariate analysis, for example, Krzanowski (2000).

Here, we shall simply specify the relevant matrices from which the test statistics are obtained, and by analogy with univariate linear regression. The total variation associated with the response variables is given by the $p \times p$ matrix of (mean-corrected) total sums of squares and products (SSP) defined as

$$\mathbf{T} = \mathbf{Y}^T \mathbf{Y} - n \bar{\mathbf{y}} \bar{\mathbf{y}}^T \quad [8]$$

where $\bar{\mathbf{y}} = [\bar{y}_1, \bar{y}_2 \dots \bar{y}_q]^T$ is the matrix of sample mean vectors of the $\mathbf{y}_i = (y_{i1} \ y_{i2} \dots y_{in})^T$, $i = 1, \dots, q$. Note here that the ij th element of \mathbf{T} corresponds to the sample covariance between the i th and j th response variables (i.e., the mean-corrected SSP divided by appropriate degrees of freedom).

Assuming that the regressor matrix \mathbf{X} is of full-rank $p + 1$, the total matrix \mathbf{T} can be partitioned into the hypothesis matrix, say, \mathbf{H} for testing the overall significance of regression and the error or residual matrix, say, \mathbf{E} . The \mathbf{H} and \mathbf{E} matrices, in the expression $\mathbf{T} = \mathbf{H} + \mathbf{E}$, are given by

$$\mathbf{H} = \hat{\mathbf{B}}^T \mathbf{X}^T \mathbf{Y} - n \bar{\mathbf{y}} \bar{\mathbf{y}}^T \quad [9]$$

and

$$\mathbf{E} = \mathbf{Y}^T \mathbf{Y} - \hat{\mathbf{B}}^T \mathbf{X}^T \mathbf{Y} = (\mathbf{Y} - \mathbf{X} \hat{\mathbf{B}})^T (\mathbf{Y} - \mathbf{X} \hat{\mathbf{B}}) \quad [10]$$

Note that, while $\hat{\mathbf{Y}} = \hat{\mathbf{B}} \mathbf{X}$ is the matrix of the predicted values, the diagonal elements of the matrices \mathbf{T} , \mathbf{H} , and \mathbf{E} , respectively, represent the usual total, regression and error sums of squares for the corresponding dependent variables in the univariate linear regression scenario. Furthermore, eqn [10] implies that an unbiased estimate of Σ is given by $\mathbf{E}/(n-p-1)$.

Testing of (Linear) Hypotheses: Multivariate Tests

In the context of multivariate linear regression models, the hypotheses examined may be, in general, associated with the functions of the regressor matrix \mathbf{B} . Here, we aim to test the redundancy of certain regressor variables, or use the tests in connection with model-reduction schemes as in the process of selecting regressor variables, or when a comparison of two or more populations is needed (where the regressor variables are regarded as experimental factors).

The most popular general form of such (linear) hypotheses is

$$H_0: \mathbf{LBM} = \mathbf{0} \text{ vs } H_1: \mathbf{LBM} \neq \mathbf{0}$$

where $\mathbf{0}$ denotes the matrix of zeroes, and, say, \mathbf{L} is a matrix of size $r \times (p + 1)$, and \mathbf{M} is a $q \times s$ matrix. \mathbf{L} may be regarded as defining a linear function on the regressor side and \mathbf{M} as defining a linear function on the response side.

The \mathbf{L} and \mathbf{M} matrices play different roles. So, they need to be chosen carefully: \mathbf{L} is used to obtain a linear function of the regression parameters within the models, thus, leading to, for example, testing the significance of a single regressor or difference between two regressors associated with each of the response variables; \mathbf{M} , on the other hand, is used for testing the significance of a linear function of regression parameters associated with different response variables but corresponding to the same set of regressors, for example, testing the difference between the coefficients of, say, X_1 in the models for Y_1 and Y_2 . Hence, using both \mathbf{L} and \mathbf{M} would lead to interesting inferences.

When we wish to test only the hypotheses associated with regression parameters, \mathbf{M} is set as the $q \times q$ identity matrix. Similarly, when the significance of the entire multivariate regression model fitted is tested, matrix \mathbf{L} is also set as an identity matrix of size $(p + 1) \times (p + 1)$.

The corresponding hypothesis and error matrices are then given by

$$\mathbf{H} = \mathbf{M}^T \hat{\mathbf{B}}^T \mathbf{L}^T [\mathbf{L}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{L}^T]^{-1} \hat{\mathbf{L}} \mathbf{B} \mathbf{M} \quad [11]$$

and

$$\mathbf{E} = \mathbf{M}^T (\mathbf{Y}^T \mathbf{Y} - \hat{\mathbf{B}}^T \mathbf{X}^T \mathbf{Y}) \mathbf{M} \quad [12]$$

There are many test statistics available to test the hypotheses stated above, most being based on the functions of eigenvalues, say, $\lambda_1, \lambda_2, \dots, \lambda_q$, of the matrix $\mathbf{H} \mathbf{E}^{-1}$, assuming \mathbf{E}^{-1} exists. These include:

1. Roy's maximum root: The largest eigenvalue, say, λ_{\max}
2. Hotelling-Lawley trace:

$$\sum_{i=1}^q \lambda_i = \text{Trace}(\mathbf{H} \mathbf{E}^{-1})$$

3. Pillai's trace:

$$\sum_{i=1}^q \frac{\lambda_i}{1 + \lambda_i} = \text{Trace}(\mathbf{H}(\mathbf{H} + \mathbf{E})^{-1})$$

4. Wilks' lambda:

$$\prod_{i=1}^q \frac{1}{1 + \lambda_i} = \frac{|\mathbf{E}|}{|\mathbf{H} + \mathbf{E}|}.$$

(see Krzanowski (2000) or Khattree and Naik (1995) for further details.)

Note that, Pillai's trace may be regarded as a measure of response variable variation explained by the fitted multivariate linear regression model, and as an analog to the well-known coefficient of determination (R^2) in the context of univariate linear regression.

Although many comparisons of these statistics have been considered in the literature, the results are still indecisive. However, the Wilks' lambda statistic (usually denoted by Λ) has been chosen by many researchers, partly because of its ease of computation, but mainly for the existence of distributional approximations which enable critical values to be readily found. Hence, the remainder of this article considers only the Wilks' lambda test statistics.

Although tables of critical values for this and other test statistics are available, we shall explore the popular F -statistic approximation for Wilks' lambda given by

$$F = \left(\frac{1 - \Lambda^{1/t}}{\Lambda^{1/t}} \right) \left(\frac{st - 2u}{ab} \right) \quad [13]$$

where

$$s = (n - p - 1) - \frac{b - a + 1}{2}, a = \text{rank}(\mathbf{L}), b = \text{rank}(\mathbf{H} + \mathbf{E}), \\ u = (ab - 2)/4$$

and

$$t = \begin{cases} \sqrt{(a^2 b^2 - 4)/(a^2 + b^2 - 5)} & \text{if } (a^2 + b^2 - 5) > 0 \\ 1 & \text{otherwise} \end{cases}$$

The F -statistics in eqn [13] is said to approximately follow an F -distribution with degrees of freedom ab and $(st - 2u)$. The distribution is exact if $\min(a, b) \leq 2$ (see Rao (1973) for details).

Furthermore, if in particular $\mathbf{H}_0: \mathbf{LB} = \mathbf{0}$ (with $\mathbf{M} = \mathbf{I}_q$) is rejected, then we may attempt to provide further inference on the regression parameters. For example, we may compute confidence intervals for the individual components of \mathbf{LB} or just \mathbf{B} (if $\mathbf{L} = \mathbf{I}_{p+1}$). Here, we use Roy's largest eigenvalue statistic in conjunction with the corresponding (approximate) F -distribution (see Khattree and Naik (1995) for further details).

Example and Software

To illustrate the basic theory and ideas above, we shall consider the following example (hereafter referred to as the SAT data):

DASL (see Web resources) describes data from 1994–95 for each of the 50 US states, for the variables current expenditure per pupil in thousands of dollars (COST), the average pupil/teacher ratio (RATIO), the estimated average annual salary of teachers in thousands of dollars (SALARY), the percentage of all eligible students taking the SAT (ELIGIBLE), then average verbal SAT score (VERBAL), the average math SAT score (MATH), and the average total score on the SAT (TOTAL). Additional information on this data can be

found in Guber (1999). It is quite natural to think of linear modeling to predict verbal and math SAT scores for a set of values of COST, RATIO, SALARY, and ELIGIBLE, even though these regressor variables were not fixed by the investigator.

The main aim here is to fit a multivariate linear regression model expressing response variables VERBAL and MATH as linear functions of regressor variables COST, RATIO, SALARY, and ELIGIBLE, and to carry out some hypothesis testing related to the regression parameters as well as the response variables.

Recall that, in matrix notation, the model is given by $\mathbf{Y} = \mathbf{XB} + \boldsymbol{\xi}$, where we may define matrices \mathbf{Y} , \mathbf{X} , \mathbf{B} , and $\boldsymbol{\xi}$ as $\mathbf{Y} = [\mathbf{y}_{\text{VERBAL}} \ \mathbf{y}_{\text{MATH}}]_{50 \times 2}$, $\mathbf{X} = [\mathbf{1} \ \mathbf{x}_{\text{COST}} \ \mathbf{x}_{\text{RATIO}} \ \mathbf{x}_{\text{SALARY}} \ \mathbf{x}_{\text{ELIGIBLE}}]_{50 \times 5}$, and $\boldsymbol{\xi} = [\boldsymbol{\varepsilon}_{\text{VERBAL}} \ \boldsymbol{\varepsilon}_{\text{MATH}}]_{50 \times 2}$, with $\mathbf{y}_{\text{VERBAL}}$, \mathbf{x}_{COST} etc. being the response and regressor vectors of size 50×1 (50 representing the US states), $\mathbf{1}$ being a 50×1 vector of 1s and the 50×1 vectors $\boldsymbol{\varepsilon}_{\text{VERBAL}}$ and $\boldsymbol{\varepsilon}_{\text{MATH}}$ representing the error respective errors; and

$\mathbf{B} = [\boldsymbol{\beta}_{\text{VERBAL}} \ \boldsymbol{\beta}_{\text{MATH}}]_{5 \times 2}$ with, say, $\boldsymbol{\beta}_{\text{VERBAL}} = (\beta_{V_{\text{Int}}} \ \beta_{V_{\text{COST}}} \ \beta_{V_{\text{RATIO}}} \ \beta_{V_{\text{SALARY}}} \ \beta_{V_{\text{ELIGIBLE}}})^T$ and $\boldsymbol{\beta}_{\text{MATH}} = (\beta_{M_{\text{Int}}} \ \beta_{M_{\text{COST}}} \ \beta_{M_{\text{RATIO}}} \ \beta_{M_{\text{SALARY}}} \ \beta_{M_{\text{ELIGIBLE}}})^T$ representing the regression parameters.

Suppose we wish to test the following hypotheses:

- \mathbf{H}_{01} : significance (overall) of the entire model;
- \mathbf{H}_{02} : significance of each of the regressor variables;
- \mathbf{H}_{03} : coefficients or parameter estimates associated with regressors COST and SALARY are the same for each of the response variables VERBAL and MATH (or are the same in the two equations given by eqn [3]).

We may rewrite the above hypotheses in the form $\mathbf{H}_0: \mathbf{LBM} = \mathbf{0}$ with $\mathbf{H}_{01}: \mathbf{B} = \mathbf{0}$, where $\mathbf{M} = \mathbf{I}_{2 \times 2}$ and $\mathbf{L} = \mathbf{I}_{5 \times 5}$ $\mathbf{H}_{02}: \mathbf{LBM} = \mathbf{0}$, where $\mathbf{M} = \mathbf{I}_{2 \times 2}$ and $\mathbf{L} = (0 \ 1 \ 0 \ 0 \ 0)_{1 \times 5}$ for testing the significance of regressor variable COST, $\mathbf{L} = (0 \ 0 \ 1 \ 0 \ 0)_{1 \times 5}$ for testing the significance RATIO, etc.

$\mathbf{H}_{03}: \mathbf{LBM} = \mathbf{0}$, where $\mathbf{M} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}_{2 \times 1}$ and $\mathbf{L} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}_{2 \times 5}$

or equivalently, test $\beta_{V_{\text{COST}}} - \beta_{M_{\text{COST}}} = 0$ and $\beta_{V_{\text{SALARY}}} - \beta_{M_{\text{SALARY}}} = 0$ simultaneously.

Statistical computing software packages, such as SAS (2002–03) and R (2008), may be used readily to carry out a multivariate linear regression analysis and test specific hypotheses such as those mentioned above. We shall utilize the SAS system via coding to perform the required analyses. For a basic introduction to using SAS for statistical analyses, see Der and Everitt (2002).

The annotated SAS codes are presented in Table 1. The well-known regression procedure PROC REG of SAS has been used here with additional statement MTEST for carrying out multivariate tests (e.g., Wilks' lambda with F -approximation). Though PROC REG also produces the regression parameter estimates automatically, for those interested in using matrix manipulations for the various

Table 1 SAS codes for multiple linear regression of SAT data

```

/* read SATdata (in space-delimited text format)
from external file SASdata.dat */
data SATdata;
format STATE $13.;
infile 'D:\Research\MultRegress\SATdata.dat'
firstobs=2;
input SID STATE $ COST RATIO SALARY ELIGIBLE VERBAL
MATH TOTAL;
run;
/* use matrix operations via PROC IML to compute reg
coefs */
%let var1 = VERBAL MATH;
%let var2 = COST RATIO SALARY ELIGIBLE;
proc iml; /* start of IML */
use SATdata;
read all var {&var1} into Y[colname=vars];
read all var {&var2} into X[colname=vars];
n=nrow(Y); q=ncol(Y);
ones=J(n,1,1); X=ones||X; p=ncol(X);
B = (inv(t(X)*X))*t(X)*Y;
print B;
quit;
/* use PROC REG procedure with MTEST statement */
proc reg data=SATdata;
model VERBAL MATH = COST RATIO SALARY ELIGIBLE;
/* test significance of overall model */
OverallModel: mtest intercept, COST, RATIO, SALARY,
ELIGIBLE / print;
/* error and hypothesis matrices printed */
/* test significance of each regression parameter */
COST: mtest COST; /* error and hypothesis matrices
printed */
RATIO: mtest RATIO;
SALARY: mtest SALARY;
ELIGIBLE: mtest ELIGIBLE;
/* test coeffs of COST and SALARY are the same in
VERBAL and MATH */
SameCoef: mtest VERBAL-MATH, COST, SALARY;
run;
quit;

```

(Further details are available from the author, in particular, on the use of PROC IML.)

computations, SAS provides the powerful PROC IML procedure. As an example, PROC IML codes for computing the regression parameter estimates given by $\hat{\mathbf{B}}$ in eqn [5], is also shown in **Table 1**. The estimates are given by $\hat{\mathbf{B}} = [\hat{\boldsymbol{\beta}}_{\text{VERBAL}} \hat{\boldsymbol{\beta}}_{\text{MATH}}]_{5 \times 2}$, with $\boldsymbol{\beta}_{\text{VERBAL}} = (509.699 \ 1.307 \ -2.081 \ 0.630 \ -1.337)^T$ and $\boldsymbol{\beta}_{\text{MATH}} = (536.272 \ 3.156 \ -1.543 \ 1.008 \ -1.567)^T$.

Table 2 shows the relevant portions of the edited output obtained by running the PROC REG codes in **Table 1**. We noted earlier that the regression coefficients in the multivariate linear regression model have the same estimates as they would if each response variable was regressed separately on the entire set of regressor variables (in univariate linear regression models), as supported by eqn [6]. The first part of **Table 2** shows these univariate regression results indicating that with no

restrictions (i.e., just estimating the regression parameters using eqn [5]), the estimated regression relationships to be $\text{VERBAL} = 509.699 + 1.307 \times \text{COST} - 2.081 \times \text{RATIO} + 0.627 \times \text{SALARY} - 1.337 \times \text{ELIGIBLE}$.

$\text{MATH} = 536.272 + 3.156 \times \text{COST} - 1.543 \times \text{RATIO} + 1.008 \times \text{SALARY} - 1.567 \times \text{ELIGIBLE}$.

We can see that the regression coefficients obtained here are exactly the same as those obtained via matrix manipulations shown above. The significance of these regression coefficients, given by the corresponding p values in **Table 2**, are based on the univariate regression approach. So, they need to be interpreted with caution! Ideally, the significance of each of the regression parameters should be examined in a multivariate sense, that is, overall with respect to the two response variables, and these results, obtained via PROC REG codes, are shown below.

Table 2 also shows the estimated error SSP matrix \mathbf{E} and the regression or hypothesis SSP matrix \mathbf{H} (given by eqns [9] and [10]) associated with the null hypothesis H_{01} , that is, for testing the significance of the entire fitted model. Interpretation of these matrices is left to the reader's discretion. The corresponding Wilks' lambda-based multivariate test (labeled overall model) indicates that the approximate F-statistic with 10 and 88 degrees of freedom has a very large value of 291.14 and a very small p value of <0.0001 , thus, providing very strong evidence for significance of the entire multivariate linear regression model fitted.

The multivariate tests (labeled COST, RATIO, SALARY, and ELIGIBLE) for testing the significance of each regressor variable simultaneously among all the response variables indicate that, only the variable ELIGIBLE has a very significant effect on the variables VERBAL and MATH. This is indicated by the F-statistic (78.20) for the corresponding Wilks' lambda with very small p -value (<0.0001). The F-statistics for each regressor has the same degrees of freedom, and the other regressors COST, SALARY, and RATIO all appear to not significantly contribute to VERBAL and MATH.

Finally, consider testing of the hypothesis that the parameter estimates associated with regressors COST and SALARY are the same for each of the response variables VERBAL and MATH. The corresponding F-statistic with 2 and 45 degrees of freedom derived from Wilks' lambda has a value 3.47 and a p -value of 0.0348 (see multivariate test labeled SameCoef). Thus, at best, the hypothesis could only be rejected with weak evidence, indicating that COST and SALARY have approximately the same coefficients in the linear regression models for VERBAL and MATH.

Note here that, the advantage of fitting a multivariate regression model (or bivariate regression model in the above example) over fitting separate (or two, in the example) univariate regression models is that, in general, the former takes into account the correlation(s) between the response variables while the latter ignores such interdependency. To illustrate, consider a hypothesis testing similar

Table 2 Selected and edited SAS output for multiple linear regression of SAT data

Univariate Tests:					
Dependent Variable: VERBAL					
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	509.69910	24.55395	20.76	<.0001
COST	1	1.30662	4.89805	0.27	0.7909
RATIO	1	-2.08140	1.49332	-1.39	0.1702
SALARY	1	0.62996	1.10869	0.57	0.5727
ELIGIBLE	1	-1.33733	0.10740	-12.45	<.0001
Dependent Variable: MATH					
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	536.27243	30.22140	17.74	<.0001
COST	1	3.15598	6.02860	0.52	0.6032
RATIO	1	-1.54283	1.83800	-0.84	0.4057
SALARY	1	1.00796	1.36460	0.74	0.4640
ELIGIBLE	1	-1.56715	0.13219	-11.86	<.0001
Multivariate Test:					
Error Matrix (E)		Hypothesis Matrix (H)			
10379.79661	11009.830615	10499099.203		11685411.169	
11009.830615	15724.438311	11685411.169		13006334.562	
Multivariate Test: Overall Model		Multivariate Statistics and F Approximations			
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.0008608	291.14	10	88	<.0001
Multivariate Test: COST		Multivariate Statistics and F Approximations			
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.99105479	0.20	2	44	0.8206
Multivariate Test: RATIO		Multivariate Statistics and F Approximations			
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.94834102	1.20	2	44	0.3113
Multivariate Test: SALARY		Multivariate Statistics and F Approximations			
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.98762687	0.28	2	44	0.7604
Multivariate Test: ELIGIBLE		Multivariate Statistics and F Approximations			
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.21956036	78.20	2	44	<.0001
Multivariate Test: SameCoef		Multivariate Statistics and Exact F Statistics			
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.86648426	3.47	2	45	0.0398

to the last one above. Here, suppose we wish to compare the coefficients of regressor variable COST for the two regression models fitted, one for response VERBAL and the other for MATH. We may construct a test statistic synonymous with a two-sample t -test when using two separate univariate models, assuming independence between the response variables. On the other hand, using a bivariate/multivariate regression model would yield us a testing process synonymous with the so-called paired t -test. However, we know that, when the assumption of independence is in question, the paired t -test is more powerful than the two-sample t -test. Hence, it is logical to conclude that multivariate regression modeling is more powerful than fitting separate univariate regression models when the correlation structure between the response variables are considered to be an important aspect of the modeling process.

Summary

Linear regression models, in general, are among the most commonly used statistical methods, while multivariate regression models extend the basic idea to many response variables. The theory behind multivariate linear regression modeling is highly developed and easily applied to real problems. Implementation or fitting of multivariate linear regression models is more difficult than fitting a univariate linear regression model, but procedures are readily available in statistics software packages such as SAS and R.

This article provided an easily digestible theory on multivariate linear regression modeling illustrated by an easily understood example. It also focused on modeling with quantitative response and regressor variables. The analysis of models with qualitative regressor variables

with quantitative responses is usually covered in a MANOVA scenario with experimental factors. The case with qualitative responses and quantitative regressors could be tackled via the typical discrimination and classification scenarios, but the case where both the response and regressor variables are qualitative, may be handled by techniques such as canonical correspondence analysis. Many of the above scenarios are also covered in the area of generalized linear models. Another technique that has a very close relationship with multivariate linear regression is known as canonical correlation analysis (CCA). CCA provides a symmetric relationship between the response and regressor variables and attempts to extract most of the association between the two sets of variables via the linear functions of the two sets (Krzanowski, 2000).

The reader is referred to Kharttree and Naik (1995) and Rencher (2002) for an extended coverage on multivariate linear regression modeling, including model diagnostics, response surface modeling, simultaneous confidence intervals, handling fixed and random regressor variables, and dealing with nonsingular matrices leading to the use of generalized inverses.

Bibliography

- Der, G. and Everitt, B. S. (2002). *A Handbook of Statistical Analyses Using SAS*, 2nd edn. London, New York: Chapman & Hall/CRC.
- Guber, D. L. (1999). Getting what you pay for: The debate over equity in public school expenditures. *Journal of Statistics Education* **7**(2), http://www.amstat.org/publications/jse/secure/v7n2/datasets_guber.cfm (accessed October 2009).
- Hsu, T. -C (2005). Research methods and data analysis procedures used by educational researchers. *International Journal of Research & Methods in Education* **28**(2), 109–133.
- Kharttree, K. and Naik, D. N. (1995). *Applied Multivariate Statistics with SAS Software*. Cary, NC: SAS Institute.
- Krzanowski, W. J. (1998). Multivariate multiple regression. In Armitage, P. and Colton, T. (eds.) *Encyclopedia of Biostatistics*, pp 8. Chichester: Wiley.
- Krzanowski, W. J. (2000). *Principles of Multivariate Analysis: A User's Perspective*. Oxford (Oxfordshire), New York: Oxford University Press.
- Ott, R. L. and Longnecker, M. T. (2001). *Introduction to Statistical Methods and Data Analysis*, 5th edn. United States: Thomson Learning, Duxbury Press.
- R (2008). The R foundation for statistical computing. Version 2.7.1. <http://www.r-project.org> (accessed October 2009).
- Rao, C. R (1973). *Linear Statistical Inference*. New York: John Wiley & Sons, Inc.
- Rencher, A. C. (2002). *Methods of Multivariate Analysis*, 2nd edn. New York: Wiley.
- SAS (2002–2003). *SAS Institute Inc. Version 9.1*. Cary, NC, USA. <http://www.sas.com> (accessed October 2009).
- Altunkaynak, B. and Ekni, M. (2002). Detection of influential observation vectors for multivariate linear regression. *Haceteppe Journal of Mathematics and Statistics* **31**, 139–151.
- Goldwasser, M. A. and Fitzmaurice, G. M. (2001). Multivariate linear regression analysis of childhood psychopathology using multiple informant data. *International Journal of Methods in Psychiatric Research* **10**, 1–10.
- Jobson, J. D. (1992). *Applied Multivariate Data Analysis: Volume II: Categorical and Multivariate Methods*. New York: Springer.
- Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*, 6th edn. Upper Saddle River, NJ: Pearson Prentice Hall.
- Kim, K. and Timm, N. (2006). *Univariate and Multivariate General Linear Models Theory and Applications with SAS*, 2nd edn. New York: CRC Press.
- Lutz, J. G. and Eckert, T. L (1994). The relationship between canonical correlation analysis and multivariate multiple regression. *Educational and Psychological Measurement* **54**(3), 666–675.
- Weisberg, S. (2005). *Applied Linear Regression*, 3rd edn. Hoboken, NJ: Wiley-Interscience.
- Wikipedia contributors (2008). Multivariate linear regression. In *Wikipedia, The Free Encyclopedia* (see topics such as General Linear Model and Bayesian Multivariate Linear regression). <http://en.wikipedia.org/wiki/Special:Search?search=Multivariate+Linear+Regression> (accessed October 2009).

Further Reading

- Altunkaynak, B. and Ekni, M. (2002). Detection of influential observation vectors for multivariate linear regression. *Haceteppe Journal of Mathematics and Statistics* **31**, 139–151.
- Goldwasser, M. A. and Fitzmaurice, G. M. (2001). Multivariate linear regression analysis of childhood psychopathology using multiple informant data. *International Journal of Methods in Psychiatric Research* **10**, 1–10.
- Jobson, J. D. (1992). *Applied Multivariate Data Analysis: Volume II: Categorical and Multivariate Methods*. New York: Springer.
- Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*, 6th edn. Upper Saddle River, NJ: Pearson Prentice Hall.
- Kim, K. and Timm, N. (2006). *Univariate and Multivariate General Linear Models Theory and Applications with SAS*, 2nd edn. New York: CRC Press.
- Lutz, J. G. and Eckert, T. L (1994). The relationship between canonical correlation analysis and multivariate multiple regression. *Educational and Psychological Measurement* **54**(3), 666–675.
- Weisberg, S. (2005). *Applied Linear Regression*, 3rd edn. Hoboken, NJ: Wiley-Interscience.
- Wikipedia contributors (2008). Multivariate linear regression. In *Wikipedia, The Free Encyclopedia* (see topics such as General Linear Model and Bayesian Multivariate Linear regression). <http://en.wikipedia.org/wiki/Special:Search?search=Multivariate+Linear+Regression> (accessed October 2009).
- <http://lib.stat.cmu.edu/DASL/DataArchive.html> – DASL, The Data and Story Library.
- <http://www.sas.com/> – SAS, superior software that gives you the power to know.
- <http://www.supershareware.com> – Software download: multivariate regression Popular Software.
- <http://www.r-project.org/> – The R Project for Statistical Computing.
- <http://www.stat.ucla.edu> – UCLA Department of Statistics.
- <http://www.ats.ucla.edu> – UCLA SAS Library.

Multivariate Normal Distribution

J Tacq, Catholic University of Brussels, Brussels, Belgium

© 2010 Elsevier Ltd. All rights reserved.

Elsewhere in the encyclopedia, there is a discussion about univariate normal distribution. This article describes multivariate normal distribution, the case with multiple variables. Univariate normal distribution is defined by two parameters, mean and variance. When multiple variables are involved, we must cope with a vector of means (i.e., a centroid) rather than deal with a single mean, and with a covariance matrix which contains variances (on the principal diagonal) and also covariances (off-diagonal) in place of a single variance. Consequently, the problem becomes considerably more complex, although the basic principles remain the same.

Bivariate Normal Distribution

A special case of the multivariate normal distribution is the bivariate normal distribution with only two variables, so that we can show many of its aspects geometrically. (For more than two variables it becomes impossible to draw figures.) The probability density function of the univariate normal distribution contained two parameters: μ and σ . With two variables, say X_1 and X_2 , the function will contain five parameters: two means μ_1 and μ_2 , two standard deviations σ_1 and σ_2 and the product moment correlation between the two variables, ρ . The probability density function (pdf) of the bivariate normal distribution is given by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right].$$

The constant term can be written in a more compact notation, if we notice that the determinant of the covariance matrix, $|\boldsymbol{\Sigma}|$, simplifies to $\sigma_1^2 \sigma_2^2 (1 - \rho^2)$. Indeed, $\boldsymbol{\Sigma}$ contains on its principal diagonal the variances σ_1^2 and σ_2^2 and on its off-diagonal the covariance $\rho\sigma_1 \sigma_2$, where its determinant is equal to $\sigma_1^2 \sigma_2^2 - \rho^2 \sigma_1^2 \sigma_2^2 = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, |\boldsymbol{\Sigma}| = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$$

Thus, the pdf of the bivariate normal distribution can also be expressed as

$$f(x_1, x_2) = (2\pi)^{-1} |\boldsymbol{\Sigma}|^{-1/2} \exp \left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right].$$

We can see that this pdf displays a general bell-shaped appearance. It looks like a mountain of normal

distribution curves. The surface is centered at the point (μ_1, μ_2) , that is, the centroid. For each point on the bottom X_1, X_2 plane, we have a point $f(X_1, X_2)$ lying on the surface of the bell-shaped mountain (Figure 1).

Conditional Distribution and Marginal Distributions

The conditional distribution $f(X_1 | X_2)$, which is the distribution of X_1 given X_2 , is usually defined as the joint distribution divided by the marginal: $f(X_1 \text{ and } X_2)/f(X_2)$. The latter, the marginal distribution of X_2 , is just the probability distribution of X_2 ignoring information about X_1 (by either summing over or integrating out X_1). For the marginal distribution of X_1 it is just vice versa: it is the probability distribution of X_1 ignoring information about X_2 (by either summing over or integrating out X_2). We can give an idea of the graphical representation of the marginal distribution of X_2 by making the orthogonal projection of the joint distribution onto the plane $X_1 = 0$ at the edge. For the marginal distribution of X_1 it is again vice versa: we make the orthogonal projection of the joint distribution onto the plane $X_2 = 0$ at the edge.

The joint density function is shown in Figure 1. For the marginal density function in Figure 2, the orthogonal projection of the joint function onto the edge plane is shown to give the reader an idea. To give a graphical representation of the conditional distribution as the division of the joint and the marginal, Figure 3 shows the joint distribution for a particular value $X_2 = 5$, divided by the marginal.

Equal-Density Contours

The place, shape, and orientation of the mountain of normal distribution curves are governed by the five parameters. The center of the mountain is determined by the centroid, that is, the vector of means. Depending on the standard deviations and the correlation, the mountain will have the shape of a circle, an ellipse, or a tilted ellipse. This can be seen if we slice horizontally through the bivariate normal surface with a plane parallel to the bottom X_1, X_2 plane. By raising and lowering this cutting plane, we obtain a series of concentric ellipses, which are places with equal height, that is, equal probability, which is why they are called equal-density contours. (An equal

density contour is a set of points for which the square of the distance to the centroid is equal to c^2 , where c is a constant.) If $\sigma_1 = \sigma_2$ and $\rho = 0$, then these equal-density contours become circles. If the standard deviations are unequal, they become ellipses whose major and minor axes are parallel to the bottom X_1, X_2 axes. If ρ is not equal to zero, the ellipses become tilted. The ellipses as equal-density contours are shown in the **Figure 4(a)–(h)**,

first as a mountain with horizontal slices and next from top, once with $\sigma_1 = \sigma_2$ and $\rho = 0$ (circles), once with $\sigma_1 \neq \sigma_2$ and $\rho = 0$ (ellipses) and once with $\sigma_1 \neq \sigma_2$ and $\rho \neq 0$ (tilted ellipses).

It is of interest to note that all kinds of calculations can be made when making use of these equal-density contours. They represent an ellipse so that the square of the distance from a point x on it to the centroid μ is constant,

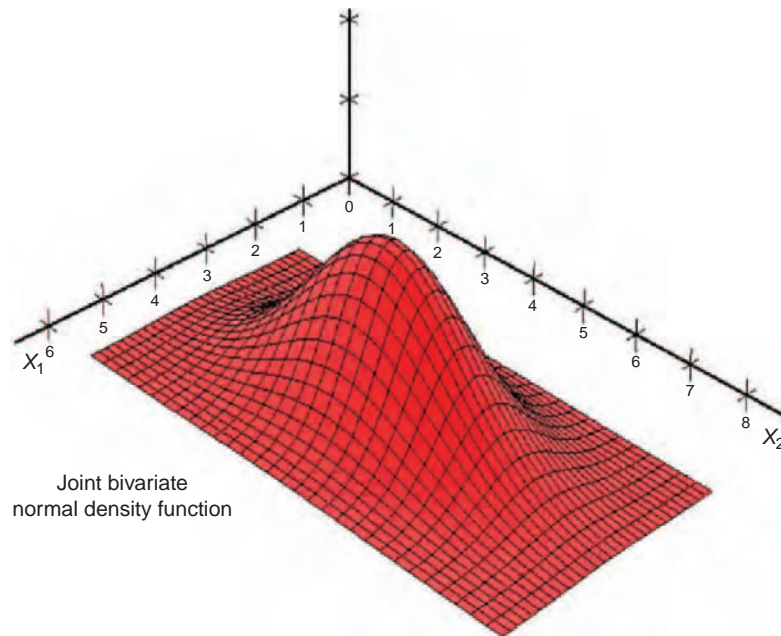


Figure 1 Joint bivariate normal density function.

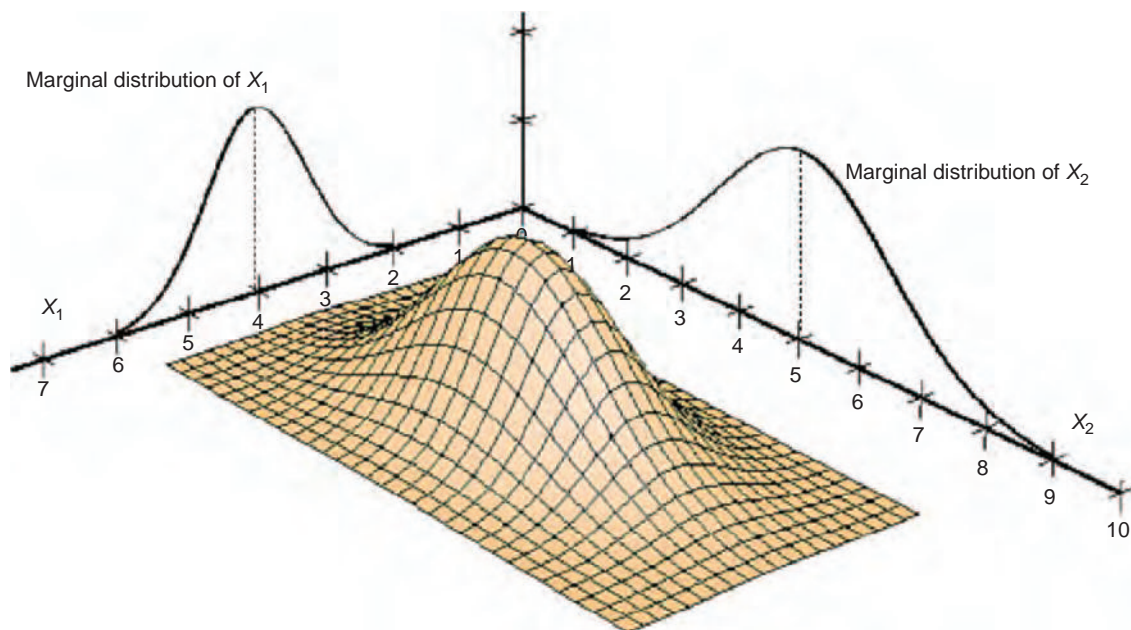


Figure 2 Marginal distributions (as projections).

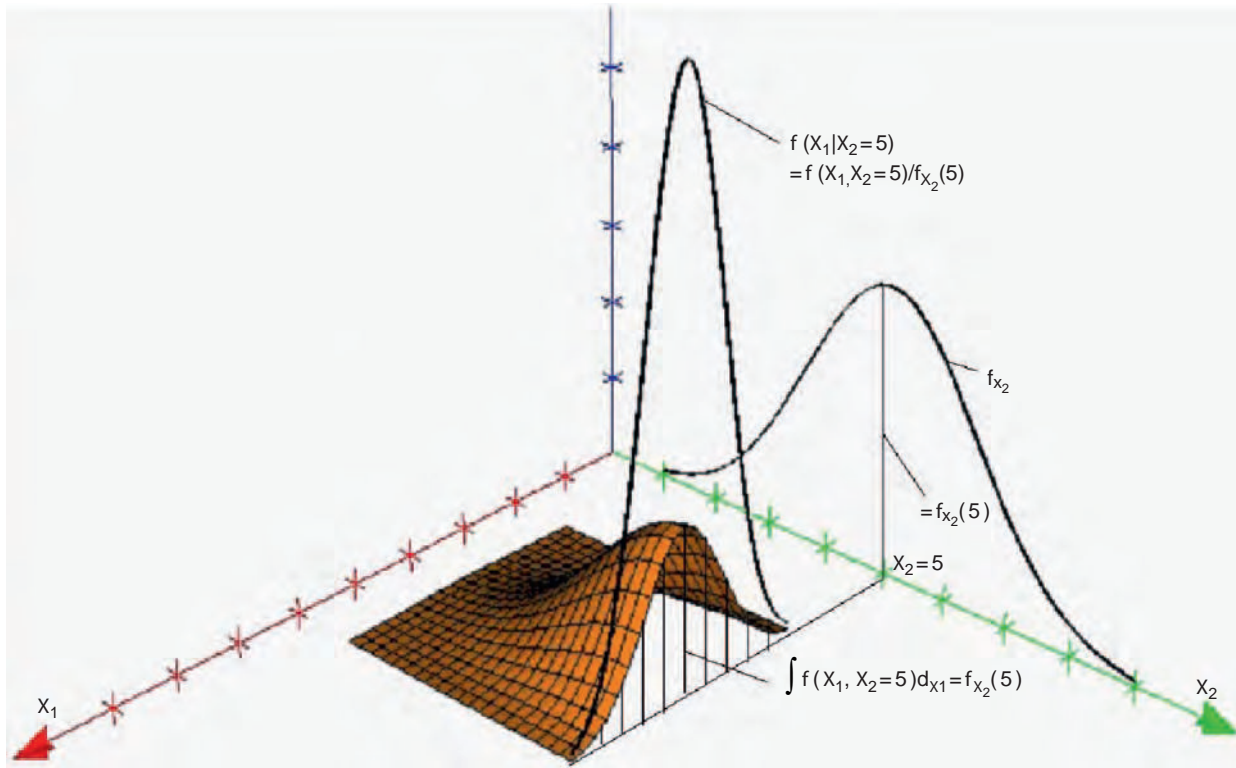


Figure 3 Conditional distribution (joint divided by marginal).

that is, $(x - \mu)' \Sigma^{-1} (x - \mu) = c^2$. It can be shown that $(x - \mu)' \Sigma^{-1} (x - \mu)$ is distributed as chi-square with p degrees of freedom, where p is the number of variables (here, $p = 2$). So, the choice $c^2 = \chi^2_{\alpha}$, that is, the upper $(100\alpha)\%$ th percentile of the chi-square distribution, leads to a contour that contains $(1 - \alpha) \times 100\%$ of the probability. **Figure 4(a)–(h)** below shows eight such equal-density contours.

Higher values of the correlations between X_1 and X_2 are represented by thinner ellipses, whereas lower values would be represented by fatter ellipses that enclose a larger proportion of the population. Next to the case above with $\rho = 0.6$, we show an extra case below with $\rho = 0.92$.

Important Properties

Some properties of the multivariate normal distribution are often used when dealing with statistical models and methods. These properties make it possible to manipulate multivariate normal distributions easily.

Notice that all these properties hold when random vector \mathbf{X} has a multivariate normal distribution, but that the reverse does not hold, for if we have a vector that satisfies one or more of these properties (such as a marginal distribution which is univariate normal), then it is possible that it does not have a multivariate normal distribution.

The key properties of a random variable \mathbf{X} having a multivariate normal distribution are:

- Linear combinations of x -variables from vector \mathbf{X} , that is, $\mathbf{a}'\mathbf{X}$, are normally distributed with mean $\mathbf{a}'\mu$ and variance $\mathbf{a}'\Sigma\mathbf{a}$. This includes the property that the marginal distributions of x -variables from vector \mathbf{X} is normal (see exercise below).
- All subsets of x -variables from vector \mathbf{X} have a multivariate normal distribution. This also includes the property of normal marginals.
- Zero covariance between x -variables from vector \mathbf{X} implies that they are independently distributed.
- The conditional distributions of x -variables from vector \mathbf{X} are multivariate normal. This includes the special case of vector \mathbf{X} being bivariate normal, from which follows that the conditional distribution of X_1 for a fixed value of X_2 is univariate normal. The formulas of mean and variance of this conditional density are $\mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2)$ and $\sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}$, respectively (see exercise below).

Central Limit Theorem

We recall from the article on univariate normal distribution, which was defined by two parameters, mean and variance, that the central limit theorem was very important. This theorem tells us that the sampling distribution of the sample

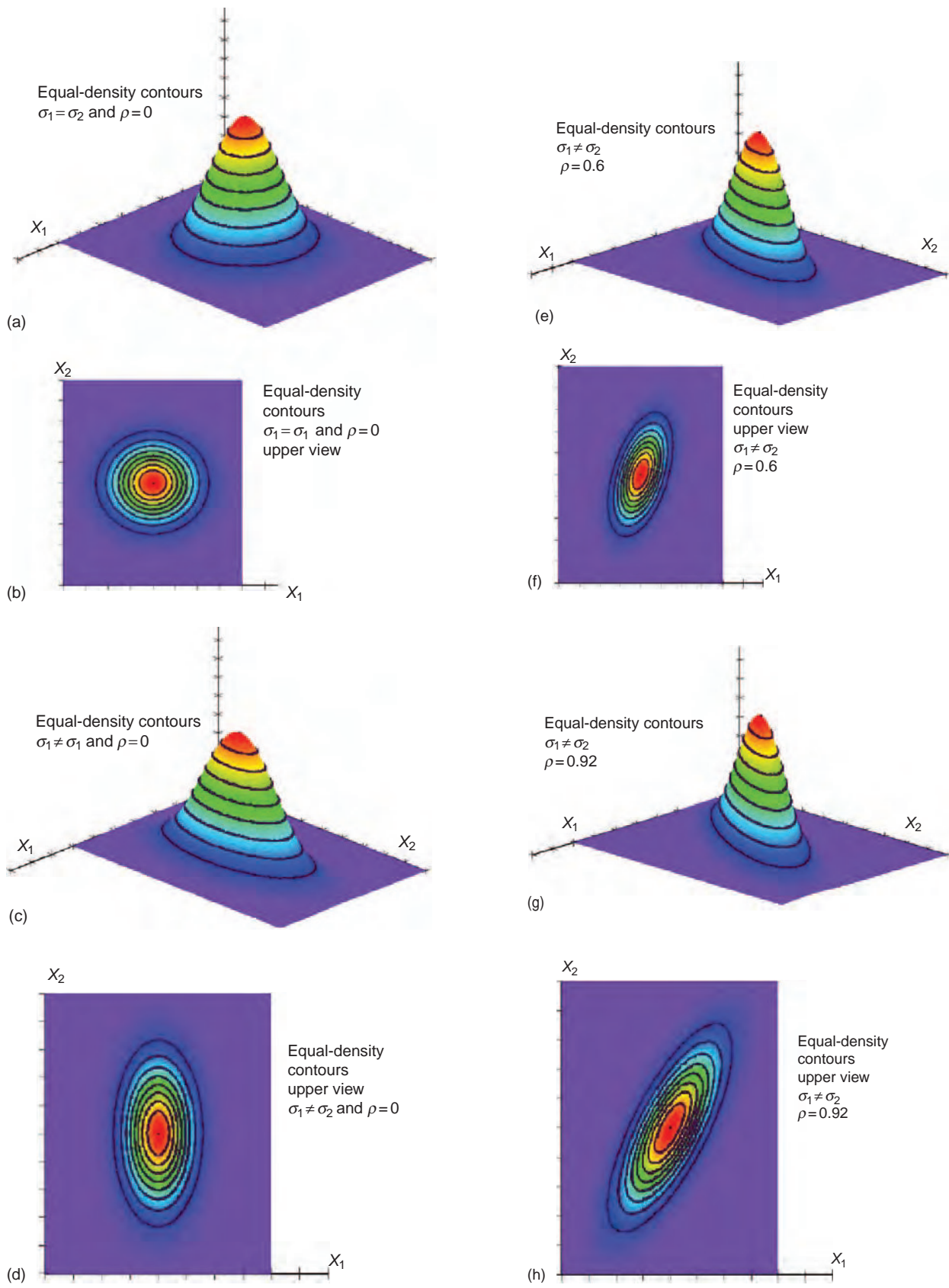


Figure 4 Equal-density contours for the bivariate normal distribution with different σ 's and ρ .

mean, \bar{X} , for a large sample size, is nearly normal, whatever the form of the underlying population distribution (with mean equal to μ , which is the mean of the population, and squared standard error equal to $(1/n)(\sigma^2)$, which is the variance of the population divided by n). It turns out that this theorem can be generalized for the multivariate case: as the sample size is increased, the sampling distribution of the centroid (vector of means) will be multivariate normal, irrespective of the form of the parent population (with centroid equal to μ , which is the centroid of the population, and covariance matrix equal to $(1/n)\Sigma$, which is the population covariance matrix divided by n). As we already mentioned, this central limit is the main cause of the immense popularity of the normal distribution and its multivariate counterpart.

Standardized Distributions and Calculation Examples

If X_1 and X_2 are standardized to z -scores with zero mean and unit standard deviation, we have, analogous to the univariate case, the standardized bivariate normal distribution. Thus

$$f(z_1, z_2) = (2\pi)^{-1} |\mathbf{R}|^{-1/2} \exp \left[-\frac{1}{2} \mathbf{z}' \mathbf{R}^{-1} \mathbf{z} \right],$$

where \mathbf{R} is the correlation matrix of $\mathbf{X} = (X_1, X_2)$. Just like in univariate normal distribution, where we can find areas under the standard normal curve, we shall be able to do the same type of thing with the bivariate normal density. Instead of areas we will now calculate volumes enclosed by the bivariate normal surface. In so doing, we can make probability statements about observations drawn randomly from a bivariate normal distribution with certain parameters.

An Example

Statistician Karl Pearson carried out a study on the resemblances between parents and children. He measured the heights of 1078 fathers and sons, and found that the fathers and sons joint heights approximately followed a bivariate normal distribution with the mean of the fathers' heights = 5 feet, 9 inches ($\mu_1 = 172.5$ cm); mean of sons' heights = 5 feet, 10 inches ($\mu_2 = 175$ cm); standard deviation of fathers' heights = 2 inches ($\sigma_1 = 5$ cm); standard deviation of sons' heights = 2 inches ($\sigma_2 = 5$ cm); correlation between fathers and sons' heights $\rho = 0.5$. (1) Predict the height of the son of a father who is 6'2" (185 cm) tall. (2) What is the probability of obtaining an observation in which the height of the son is more than the height of the father?

The solution is as follows:

1. In this numerical example we consider a bivariate normal density function with the following parameters: $\mu_1 = 172.5$ cm, $\mu_2 = 175$ cm, $\sigma_1 = 5$ cm, $\sigma_2 = 5$ cm,

$\rho = 0.5$. To predict the height of the son of a father who is 6'2" (185 cm) tall, we have to look at the conditional bivariate normal density function $f(X_2 | X_1 = 185)$, which is a univariate normal pdf of X_2 conditioned upon X_1 being held constant at 185 (see properties). From probability calculus we know that for two events A and B, the probability of B given A is obtained by dividing the joint by the marginal: $p(B|A) = p(A \text{ and } B)/p(A)$. Analogously we have $f(X_2 | X_1) = f(X_1, X_2)/f(X_1)$. The formula of the joint bivariate normal density function can be written algebraically as follows:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} \left\{ \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right\} \right]$$

This pdf has to be divided by $f(X_1) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp \left(-\frac{(x_1 - \mu_1)^2}{2\sigma_1^2} \right)$. A little bit of algebra shows that this conditional distribution $f(X_2 | X_1)$ is univariate normal with expected value (mean) $= \mu_2 + (\sigma_{12}/\sigma_1^2)(X_1 - \mu_1) = 175 + ((0.5)(5)(5)/25)(185 - 172.5) = 181.25$ cm. Hence, 181.25 cm is the best prediction of the height of the son (of a father of 185 cm tall).

2. In order to prepare the second exercise we first look at the property that linear combinations of x -variables from vector \mathbf{X} , that is, $\mathbf{a}'\mathbf{X}$, are normally distributed with mean $\mathbf{a}'\mu$ and variance $\mathbf{a}'\Sigma\mathbf{a}$. An example of such a linear combination is the difference between the two variables of our bivariate normal distribution, X_1 and X_2 , where \mathbf{a}' is equal to $[1 \ -1]$. It follows that this difference $X_1 - X_2$ has a univariate normal distribution with mean $\mathbf{a}'\mu = [1 \ -1] \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = 172.5 - 175 = -2.5$ and variance $\mathbf{a}'\Sigma\mathbf{a} = [1 \ -1] \begin{bmatrix} 25 & 12.5 \\ 12.5 & 25 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 25$.

Now we can easily solve the question of finding the probability of obtaining an observation in which the height of the son is more than the height of the father, i.e., $X_2 > X_1$, for this is equal to the probability that the difference $X_1 - X_2$ is smaller than zero. We know that the difference $X_1 - X_2$ has a univariate normal distribution with mean -2.5 and variance 25 , whence we have reduced the problem to a univariate case. The standard deviation of $X_1 - X_2$ is $\sigma = \sqrt{25} = 5$. Its mean is -2.5 . We want the probability $p(X_1 - X_2 < 0)$. The standardized score is $(0 - (-2.5))/5 = 0.5$. The probability $p(X_1 - X_2 < 0)$ is equal to $p(z < 0.5) = 0.69$. So, there is a probability of 69% of obtaining an observation in which the height of the son is more than the height of the father.

Although a computation involving a multivariate normal distribution can often be reduced to one involving

a simpler univariate normal distribution (as we have seen in the exercises above), it may not always be possible. Some computations with the multivariate normal distribution would invariably involve integral calculus, because volumes under the mountain of normal distributions would have to be determined. For example, if one is interested in the probability of father's height being smaller than a and son's height smaller than b , it is given by $p(X_1 < a, X_2 < b) = \int_{-\infty}^a \int_{-\infty}^b f(X_1, X_2) dx_1 dx_2$, and bivariate integral calculus has to be used in the computation. Fortunately, several statistical software packages perform such computations.

An Application: Hotelling's T^2 -Test and Mahalanobis' Distance D^2

In statistics, especially in multivariate analysis, there are many applications in which multivariate normal distribution plays an important role. Of course, linear regression analysis and its extension, structural equation models, are the example, because normality lies at the heart of these techniques. Other examples are discriminant analysis, multivariate analysis of variance, and canonical correlation analysis. Take for example discriminant analysis. For a number of groups, two or more than two, we will examine whether there is a significant difference between the centroids of the groups. If there are two groups, this is an extension of Student's T -test; if there are more than two groups, it is an extension of Fisher's F -test. We concentrate on the case of two groups (in which Hotelling's T^2 will be developed as a special case of Wilks' Λ) and we emphasize testing (rather than predicting group membership).

In Student's T -test there are two groups and only one variable. In such a univariate case, we know that the sampling distribution of the sample mean is normal if the distribution of the population from which the sample is taken is normal, and even if the distribution of the population is not normal, on the condition that the sample is sufficiently large.

In Student's T -test, the t statistic is calculated as $t = |\bar{X}_{(0)} - \bar{X}_{(1)}| / (\frac{\sigma_w^2}{n_0} + \frac{\sigma_w^2}{n_1})^{1/2}$, in which n_0 and n_1 are the group sizes and σ_w^2 (which has to be estimated!) is the pooled average of the two variances (the variance $\sigma_{(0)}^2$ of X in group 0 and the variance $\sigma_{(1)}^2$ of X in group 1) and $(\frac{\sigma_w^2}{n_0} + \frac{\sigma_w^2}{n_1})^{1/2}$ is the standard error of the sampling distribution of differences of means.

The formula of its square t^2 can be written as follows:

$$\begin{aligned} t^2 &= (\bar{X}_{(0)} - \bar{X}_{(1)})^2 / (\frac{\sigma_w^2}{n_0} + \frac{\sigma_w^2}{n_1}) \\ &= \frac{n_0 n_1}{n_0 + n_1} (\bar{X}_{(0)} - \bar{X}_{(1)}) \frac{1}{\sigma_w^2} (\bar{X}_{(0)} - \bar{X}_{(1)}) \end{aligned}$$

Hotelling constructed a statistic, called Hotelling's T^2 , in which multiple discriminating variables are included, in

which a difference of group means now becomes a difference of group centroids, because there are several variables, and in which division by the estimated variance is replaced by multiplication by the inverse of a covariance matrix containing not only the dispersions but also the mutual associations between the discriminating variables:

$$T^2 = \frac{n_0 n_1}{n_0 + n_1} \mathbf{d}' \Sigma_w^{-1} \mathbf{d}$$

Here n_0 and n_1 are again the group sizes and Σ_w is the pooled average of the two covariance matrices Σ_0 and Σ_1 . The vector \mathbf{d} is the difference vector between group centroids. The part $\mathbf{d}' \Sigma_w^{-1} \mathbf{d}$ in the formula of Hotelling's T^2 is Mahalanobis' distance D^2 .

In this multivariate case it holds, analogously, that the sampling distribution of the sample centroid is multivariate normal if the population is multivariate normal and/or if the sample is sufficiently large (which is just an extension of the central limit theorem). Under this assumption of multivariate normality, Hotelling has proven that the value $[(n - p - 1)/p(n - 2)] T^2$ is distributed as F with p and $n - p - 1$ degrees of freedom. So it becomes straightforward to test whether there is a significant difference between the centroids of two groups, that is, between the means of the many variables, taken together and taking their variances and correlation coefficients into consideration. The assumption of multivariate normality can be guaranteed by a large sample. For small samples it has to be tested.

Testing for Multivariate Normality

One should keep in mind that the multivariate normal distribution does not belong to reality, but is an idealtype. In the nineteenth century, Max Weber defined an idealtype as an exaggeration of the mind, which does not occur in the world, but with which events of the world are confronted and compared, some resembling this idealtype and others deviating from it.

Multivariate normality tests check a given set of empirical data for similarity to the idealtype multivariate normal distribution, the null hypothesis being that the data set is similar to the multivariate normal distribution; therefore a sufficiently small p -value indicates deviation from multivariate normality. Multivariate normality tests are often extensions of the well-regarded Kolmogorov–Smirnov and Shapiro–Wilk tests for univariate normality. They include the Cox–Small test and Smith and Jain's adaptation of the Friedman–Rafsky test. In a recent overview, Mecklin and Mundfrom (2004) mention that at least 50 procedures for testing multivariate normality exist, but that the state of the art is not very refined and that little work has been done in evaluating the quality and power of the procedures. They conclude that no single method is sufficient, but that a mixture of methods, that is, graphical

approaches, measures of skewness and kurtosis, and more mathematically sophisticated procedures, are likely to be the most useful.

Unfortunately, tests of multivariate normality are not really multivariate in the literal sense of the word. This is because the comparison of an empirical mountain with an idealtypic mountain is not possible. We have to rely on the properties of normal distributions, such as the property that all linear combinations of normal variables are normal and the knowledge that the contours of the multivariate normal density are ellipsoids. So, instead of comparing mountains – the empirical one and the idealtypic one – we will check whether the marginal distributions and linear combinations of x -variables are normal and whether observations of pairs of x -variables show the elliptical appearance of the equal-density contours. Another possibility is that we look for outliers in the empirical data. As Johnson and Wichern (1992: 177) state, we must pay a price for concentrating on univariate and bivariate examinations of normality. We can never be sure that we have not missed some feature that is revealed only in higher dimensions. It is possible, for example, to construct a non-normal bivariate distribution with normal marginals. But they add, first, that many types of non-normality are often reflected in the marginal distributions and scatter plots, and moreover, that for most practical work, one-dimensional and two-dimensional investigations are ordinarily sufficient.

An example of univariate test is the Q–Q-plot. Such a plot can be made for the marginal distributions of the sample observations on each variable. They are in effect plots of the sample quantile versus the quantile one would expect to observe if the observations actually were normally distributed. The plot of these pairs of points should then lie very nearly along a straight line.

One can also calculate the correlation coefficient r_Q between the sample quantiles and the quantiles expected under normality and test this correlation between empirical Q and idealtypic Q for significance. An improved version of this r_Q approach is given by Shapiro and Wilk, who replace the quantiles expected under normality by a function of the expected value of standard normal-order statistics and their covariances.

For the investigation of more than one characteristic, statisticians mostly suggest plotting the x -variables against the eigenvectors resulting from an evaluation of the eigenstructure of the sample covariance matrix.

As mentioned above, it is also possible to check whether observations of pairs of x -variables show the elliptical appearance of the equal density contours. The idea is that for observations which were generated from a multivariate normal distribution, each bivariate distribution would also

be normal (see properties) and that the contours of constant density would be ellipses. Therefore, a scatterplot should exhibit an overall pattern that is approximately elliptical. Moreover, we should expect approximately the same percentage P of sample observations to lie in the ellipse given by $(x - \mu)' \Sigma^{-1} (x - \mu) \leq \chi^2(P)$, where we have to replace μ and Σ by their estimates. As this is a rather rough procedure, it has been replaced by a more formal method based on the squared generalized distances (see Johnson and Wichern, 1992: 184 ff). This is also the case for another approach, the detection of outliers. Hawkins (1980) gives an extensive treatment of the subject of outliers.

Acknowledgments

The author would like to thank colleagues Ignace Van de Woestijne and Theo Moons of Catholic University of Brussels for their help with the pictures, making use of VisuStat and VisuMath.

See also: The Normal Distribution and its Applications.

Bibliography

- Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*. Englewood Cliffs, NY: Prentice Hall.
- Hawkins, D. M. (1980). *Identification of Outliers*. London: Chapman and Hall.
- Mecklin, C. J. and Mundform, D. J. (2004). An appraisal and bibliography of tests for multivariate normality. *International Statistical Review* 72(1), 123–138.

Further Reading

- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.
- Cox, D. R. and Small, N. J. H. (1978). Testing multivariate normality. *Biometrika* 65(2), 263–272.
- Green, P. (1978). *Analyzing Multivariate Data*. Hinsdale, IL: Dryden.
- Kendall, M. and Stuart, A. (1969). *The Advanced Theory of Statistics. Vol. I. Distribution Theory. Vol. II. Classical Inference and the Linear Model*. New York: McMillan.
- Rose, C. and Smith, M. D. (2002). *Mathematical Statistics with Mathematica*. Berlin: Springer.
- Smith, S. P. and Jain, A. K. (1988). A test to determine the multivariate normality of a dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10(5), 757–761.
- Tacq, J. (1997). *Multivariate Analysis Techniques in Social Science Research*, pp 411. London: Sage.
- Tong, Y. L. (1990). *The Multivariate Normal Distribution*. 271p. Berlin: Springer.

Nonlinear Regression Analysis

H-H Huang, Academia Sinica, Taipei, Taiwan

C K Hsiao, National Taiwan University, Taipei, Taiwan

S-Y Huang, Academia Sinica, Taipei, Taiwan

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Curvature – The curvature measures how a geometric object deviates from being flat. There are various definitions of curvature depending on the context. In nonlinear regression analysis, the relative curvature array developed by Bates and Watts is a widely used approach for assessing curvature and nonlinearity.

Gauss–Newton method – The Gauss–Newton algorithm as a modification of Newton’s method is used to solve nonlinear least-squares problems. It is based on iterative local linear approximation to the regression function and does not require to evaluate the second derivatives.

Intrinsic nonlinearity – Intrinsic nonlinearity refers to a quantity which makes the prediction biased in nonlinear models. The degree of intrinsic nonlinearity is determined by the model and data and is invariant to reparametrization.

Nonlinear regression analysis – Nonlinear regression models refer to models that are nonlinear in parameters. Nonlinear regression analysis mainly concerns the prediction of responses, statistical inferences of parameters estimates, and the goodness of fit of the nonlinear model.

Parameter-effects nonlinearity – Parameter-effects nonlinearity refers to a quantity which affects the rate of convergence of the estimates from Gauss–Newton method. It varies under different reparametrization.

Solution locus – The collection of points of regression function in the sample space evaluated at all feasible parameter values constitutes the solution locus. It is also known as expectation surface.

Introduction

Regression analysis refers to the statistical inferences for a model

$$y = f(x; \theta) + \epsilon \quad [1]$$

where $y \in \mathbb{R}$ is the response variable, $x = (x_1, \dots, x_k) \in \mathbb{R}^k$ are explanatory variables, and $\theta = (\theta_1, \dots, \theta_p) \in \mathbb{R}^p$ are

parameters. We name f the regression function, whose functional form is known up to some unknown parameters θ ; in addition, ϵ is an error term with zero mean and variance σ^2 . When the regression function f is linear in the parameters θ , it leads to the very popular and widely used statistical inferential techniques known as linear regression analysis. However, linear models are not always appropriate; therefore, one often needs to apply a nonlinear regression model, where f is nonlinear in θ . The statistical inferential problem concerning [1] is called a nonlinear regression analysis, when f is nonlinear in θ .

In the following sections, we present a set of topics in nonlinear regression. Among various estimation methods of nonlinear regression, the least-squares method is the most popular. The Gauss–Newton procedure is used to estimate the unknown parameters without calculating the Hessian matrix. Underlying the normality assumption of error terms, the maximum likelihood estimator is equivalent to the least-squares estimator.

In fields such as biochemistry, ecology, economics, and social sciences, there are plenty non-linear regression models being applied for a long time. Yet if the model needs to be scrutinized carefully, the quality of inferences from parameter estimates relies on the magnitude of model nonlinearity and parameter effects. We present a graphical example to illustrate this phenomenon.

Vectors and matrices are denoted, respectively, by bold-faced lower and upper case letters, such as $x \in \mathbb{R}^k$ and $X \in \mathbb{R}^{n \times k}$, and scalars are denoted by roman letters.

Estimation Methods

For a given data set $\{(x_i, y_i) : i = 1, \dots, n\}$ with size n , an empirical model based on the data can be written as

$$y = f(X; \theta) + \epsilon \quad [2]$$

where

$$X = \begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix} = [x^{(1)} \dots x^{(k)}]$$

and

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

In the data design matrix X , the i th row $\mathbf{x}_i' = (x_{i1}, x_{i2}, \dots, x_{in})'$ represents the i th observation, and the j th column $\mathbf{x}^{(j)} = (x_{1j}, x_{2j}, \dots, x_{nj})'$ represents the j th variable. We will use $f(\boldsymbol{\theta})$ to denote $f(X; \boldsymbol{\theta}) = (f(\mathbf{x}_1; \boldsymbol{\theta}), \dots, f(\mathbf{x}_n; \boldsymbol{\theta}))'$ for short. Let $\Theta \subset \mathbb{R}^p$ be the parameter space. The collection of points of regression function in the sample space evaluated at all feasible parameter values denoted by

$$f(\Theta) = \{f(\boldsymbol{\theta}) = f(X; \boldsymbol{\theta}) \in \mathbb{R}^n : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\}$$

is called the solution locus or expectation surface. A linear regression model has the solution locus as a p -dimensional hyperplane lying in the n -dimensional sample space. While a nonlinear regression model has the solution locus as a curved hypersurface in the sample space.

Least-Squares Estimation

The least-squares estimator of $\boldsymbol{\theta}$, denoted by $\hat{\boldsymbol{\theta}}$, is the point in the parameter space such that $f(\hat{\boldsymbol{\theta}})$ is closest to \mathbf{y} in the sample space among all feasible $f(\boldsymbol{\theta})$ in the solution locus. The least-squares estimator is derived from minimization of the residual sum of squares

$$S(\boldsymbol{\theta}) = \sum_{i=1}^n \{y_i - f(x_i; \boldsymbol{\theta})\}^2, \quad \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p \quad [3]$$

When f is differentiable with respect to $\boldsymbol{\theta}$, we solve for the least-squares solution $\hat{\boldsymbol{\theta}}$ in the following system of equations:

$$\left. \frac{\partial S(\boldsymbol{\theta})}{\partial \theta_\ell} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0, \quad \ell = 1, \dots, p$$

The system of equations (called normal equations) are given by

$$\sum_{i=1}^n \frac{\partial f(x_i; \boldsymbol{\theta})}{\partial \theta_\ell} \{y_i - f(x_i; \boldsymbol{\theta})\} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0 \quad [4]$$

for $\ell = 1, \dots, p$. Or, in matrix form as

$$V(\hat{\boldsymbol{\theta}})' \hat{\boldsymbol{\epsilon}} = 0 \quad [5]$$

where $V_{i,\ell}(\boldsymbol{\theta}) = \partial f(x_i; \boldsymbol{\theta}) / \partial \theta_\ell$, and $\hat{\boldsymbol{\epsilon}} = \mathbf{y} - f(\hat{\boldsymbol{\theta}})$. The matrix $V(\boldsymbol{\theta})$, of size $n \times p$, is a velocity matrix with the ℓ th column denoting the instantaneous speed at $\boldsymbol{\theta}$ when moving on the solution locus along the ℓ th coordinate θ_ℓ . In a linear regression model, this velocity matrix is simply the data design matrix X and does not depend on the parameter values $\boldsymbol{\theta}$.

Often, normal equations do not have an analytic solution for $\boldsymbol{\theta}$ and numerical iterative procedures are needed. Below, we introduce the Gauss–Newton method for solving nonlinear least-squares problems based on iterative local linear approximations to the solution locus. Consider a small neighborhood of $\boldsymbol{\theta}^*$. The linear Taylor expansion is given by

$$f(x; \boldsymbol{\theta}) \approx f(x; \boldsymbol{\theta}^*) + \nabla_{\boldsymbol{\theta}} f(x; \boldsymbol{\theta}^*)' (\boldsymbol{\theta} - \boldsymbol{\theta}^*) \quad [6]$$

where

$$\nabla_{\boldsymbol{\theta}} f(x; \boldsymbol{\theta}^*)' = \left(\frac{\partial f(x; \boldsymbol{\theta}^*)}{\partial \theta_1}, \dots, \frac{\partial f(x; \boldsymbol{\theta}^*)}{\partial \theta_p} \right)$$

Therefore, the linear approximation to $f(x; \boldsymbol{\theta})$ around the neighborhood of $\boldsymbol{\theta}^*$ leads to an approximate residual sum of squares:

$$S(\boldsymbol{\theta}) \approx \sum_{i=1}^n \left\{ y_i - f(x_i; \boldsymbol{\theta}^*) - \sum_{\ell=1}^p V_{i,\ell}(\boldsymbol{\theta}^*) (\theta_\ell - \theta_\ell^*) \right\}^2$$

Using the linear Taylor expansion in the iterative update, the corresponding normal equations at the t th iteration are given by

$$V(\boldsymbol{\theta}_t)' V(\boldsymbol{\theta}_t) (\boldsymbol{\theta} - \boldsymbol{\theta}_t) = V(\boldsymbol{\theta}_t)' (\mathbf{y} - f(\boldsymbol{\theta}_t)) \quad [7]$$

The increment

$$\boldsymbol{\delta} = (V(\boldsymbol{\theta}_t)' V(\boldsymbol{\theta}_t))^{-1} V(\boldsymbol{\theta}_t)' (\mathbf{y} - f(\boldsymbol{\theta}_t))$$

serves as an update direction for next iteration. The Gauss–Newton method, an iterative approach, is then given by

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \lambda_t \boldsymbol{\delta}_t$$

where λ_t is the step size. The update direction $\boldsymbol{\delta}$ is derived from the tangent plane approximation to the solution locus. This approximation is only valid in a local neighborhood of the current parameter estimator $\boldsymbol{\theta}_t$. The size of this neighborhood, where the linear approximation is valid, depends on the curvedness of the nonlinear model and its parametrization. It is possible that taking a full step, that is, $\lambda = 1$, will produce an increase in the residual sum of squares, especially when the Gauss–Newton update increment $\boldsymbol{\delta}$ has extended beyond the validity region of linear approximation. Thus, a smaller step size is required to ensure a decrease in residual sum of squares. Practically, λ_t can be taken as 1/2 step size of the previous step, that is, consecutively the step size 0.5, 0.25, 0.125, etc., till a decrease in residual sum of squares (Nocedal and Wright, 2006). Then, the procedure moves on to the next iteration. In addition, it is necessary to set up a requirement to stop the Gauss–Newton procedure. Equation [5] implies that the residual vector is orthogonal to the tangent plane of the solution locus, and can be used as stopping criterion for convergence. This concept leads to the relative offset orthogonality convergence criterion by Bates and Watts (1981).

The estimate of the asymptotic covariance of the least-squares estimate is given by

$$\text{Cov}(\hat{\boldsymbol{\theta}}) = \sigma^2 \left(V(\hat{\boldsymbol{\theta}})' V(\hat{\boldsymbol{\theta}}) \right)^{-1} \quad [8]$$

In a linear model, $V(\hat{\boldsymbol{\theta}}) = X$, and does not depend on the parameter estimates. Furthermore, the covariance expression [8] for a linear model does not require an asymptotic character, that is, it is valid regardless of the sample size.

Maximum Likelihood Estimation

Consider a normal distribution for the error term:

$$y - f(x; \boldsymbol{\theta}) = \epsilon \sim N(0, \sigma^2)$$

Then, the log-likelihood function is given by

$$\ell(\boldsymbol{\theta}, \sigma^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{S(\boldsymbol{\theta})}{2\sigma^2}$$

When σ^2 is known, maximizing the log-likelihood with respect to $\boldsymbol{\theta}$ is the same as minimizing the error sum of squares $S(\boldsymbol{\theta})$. Thus, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_{\text{mle}}$ is the same as the least-squares estimator $\hat{\boldsymbol{\theta}}$. On the other hand, when σ^2 is not known, $\partial \ell / \partial \sigma^2 = 0$ has the solution

$$\hat{\sigma}^2(\boldsymbol{\theta}) = \frac{S(\boldsymbol{\theta})}{n}$$

Substituting $\hat{\sigma}^2(\boldsymbol{\theta})$ into the log-likelihood expression results in

$$\ell(\boldsymbol{\theta}, \hat{\sigma}^2(\boldsymbol{\theta})) = \kappa - \frac{n}{2} \log(S(\boldsymbol{\theta}))$$

where κ is some constant. Maximizing $\ell(\boldsymbol{\theta}, \hat{\sigma}^2(\boldsymbol{\theta}))$ with respect to $\boldsymbol{\theta}_1$ we have

$$\hat{\boldsymbol{\theta}}_{\text{mle}} = \hat{\boldsymbol{\theta}} \quad \text{and} \quad \hat{\sigma}_{\text{mle}}^2 = \frac{S(\hat{\boldsymbol{\theta}})}{n}$$

In summary, when the noise structure follows a normal distribution, the maximum likelihood estimates equal the least-squares estimates.

Good Initials

For a linear model, the Gauss–Newton method will find the minimum in one single iteration from any initial parameter estimates. For a model which is close to being linear, the convergence for Gauss–Newton method will be rapid and not depend heavily on the initial parameter estimates. However, as the magnitude of model nonlinearity becomes more and more prominent, convergence will be slow or even may not occur, and the resulting parameter estimates may not be reliable. In that case, good initials are important.

One approach to find initial values is through transformation, so that the linear regression analysis under the assumption of additive error terms can be utilized (Bates and Watts, 1988; Ryan, 1997). For instance, the reciprocal of the Michaelis–Menten regression function

$$f(x; \boldsymbol{\theta}) = \frac{\theta_1 x}{\theta_2 + x} \quad [9]$$

leads to the model

$$y^{-1} = \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1} \left(\frac{1}{x} \right) + \varepsilon$$

Indeed, this is a linear regression model denoted as

$$\tilde{y} = \beta_0 + \beta_1 \tilde{x} + \varepsilon$$

with $\tilde{y} = y^{-1}$, $\beta_0 = \theta_1^{-1}$, $\beta_1 = \theta_2 \theta_1^{-1}$, and $\tilde{x} = x^{-1}$. The least-squares estimates of β_1 and β_2 can therefore be transformed to provide initial values for the parameters in the Michaelis–Menten model. Alternatively, we can adopt a grid search, which evaluates the residual sum of squares at these parameter grid points and then finds the minimizing point among these grid points to provide initial values. In many occasions, the grid search helps to avoid improper initial values that lead to local, but not global, minimum of the residual sum of squares. Another alternative is to use uniform design points to replace the lattice grid points. Uniform design points are so designed to be as uniform and as space filling as possible and provide a more efficient search scheme (Fang *et al.*, 2000).

Assessing Nonlinearity

Given that the error terms are i.i.d. normally distributed random variables with zero mean, the least-squares estimator in linear models is unbiased, normally distributed and having minimum variance among all linear unbiased estimators (BLUEs) (i.i.d – independent and identically distributed). However, the least-squares estimator for nonlinear models does not have the same properties. The nonlinear least-squares estimator approaches BLUE only asymptotically. In addition, when the sample size increases, the Gauss–Newton iterative estimator becomes asymptotically numerically stable (Jennrich, 1969). However, if the model is highly curved and the first derivative $V(\boldsymbol{\theta})$ changes violently in the iteration, the least-squares estimator can be numerically unstable.

The parametrization is another important issue. Although the shape of the solution locus is fixed, the performance of Gauss–Newton interactive estimator varies with respect to different parameterizations. This depends on the nonlinearity in the model. There are two kinds of nonlinearity: intrinsic nonlinearity and parameter-effects nonlinearity (Bates and Watts, 1980; Hamilton *et al.*, 1982). The intrinsic nonlinearity is associated with the modeling and is invariant under reparametrization. The parameter-effects nonlinearity, however, can be lessened through a proper reparametrization. If either component of the nonlinearity is large, the least-squares estimate is hard to converge, or even does not converge. Furthermore, the asymptotic covariance of the least square estimate, given by $\sigma^2(V(\hat{\boldsymbol{\theta}})' V(\hat{\boldsymbol{\theta}}))^{-1}$, would change greatly in each step of the iteration, and the statistical inference based on the asymptotic normality becomes unreliable. In other words, the least-squares estimator in nonlinear regression depends on the curvedness of the underlying model and on the parametrization adopted.

Intrinsic Nonlinearity and Parameter Effects Nonlinearity

Of various measures for curvedness, the relative curvature array is widely used in nonlinear regression analysis and is presented below. The nonlinear least-squares fit is based on the iterative update of local linear approximations to the solution locus

$$f(x_i; \boldsymbol{\theta}) \approx f(x_i; \boldsymbol{\theta}^*) + \nabla_{\boldsymbol{\theta}} f(x_i; \boldsymbol{\theta}^*)' (\boldsymbol{\theta} - \boldsymbol{\theta}^*) \quad [10]$$

Consider a further extension to include the quadratic term in the Taylor expansion

$$f(x_i; \boldsymbol{\theta}) \approx f(x_i; \boldsymbol{\theta}^*) + \nabla_{\boldsymbol{\theta}} f(x_i; \boldsymbol{\theta}^*)' (\boldsymbol{\theta} - \boldsymbol{\theta}^*) + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^*)' \nabla_{\boldsymbol{\theta}}^2 f(x_i; \boldsymbol{\theta}^*) (\boldsymbol{\theta} - \boldsymbol{\theta}^*)$$

where $\nabla_{\boldsymbol{\theta}}^2 f(x_i; \boldsymbol{\theta}^*) = \partial^2 f(x_i; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}^2$ is a $p \times p$ matrix of second derivatives for each $i = 1, \dots, n$. The magnitude of the quadratic term relative to the linear term determines the difference between the tangent plane [10] and the corresponding solution locus. Display $\nabla_{\boldsymbol{\theta}}^2 f(x_i; \boldsymbol{\theta}^*)$, $i = 1, \dots, n$, into a three-dimensional array $A(\boldsymbol{\theta}^*)$, which is of size $n \times p \times p$. This array $A(\boldsymbol{\theta}^*)$ is called the acceleration array at $\boldsymbol{\theta}^*$. The acceleration array A can be decomposed into two components, one parallel and the other normal to the tangent plane. The intrinsic nonlinearity is based on the normal component, and is invariant corresponding to any reparametrization. Factorization of the acceleration array into the tangent and normal components along with other technical details are placed in 'Appendix'.

In the linear regression model, the solution locus is a hyperplane. Uniform and equispaced grid lines on the p -dimensional parameter space also appear as uniform and equispaced grid lines on this hyperplane. In the nonlinear model, the tangent plane to the solution locus is used to serve as a local approximation to the locus. When these p -dimensional grid lines by $\boldsymbol{\theta}$ are mapped onto this tangent plane, we would prefer to see the resulting mappings as uniform and as equispaced as possible. The parameter-effects curvature measure [18] given in 'Appendix' can assess the deviation of the parameter curves from the uniform coordinate system. **Figure 1** presents plots of parameter curves using simulated data from Michaelis–Menten model.

Confidence Regions

In the iterative updates using local linear approximation, V plays the role as the design matrix X in the linear model. Under some suitable conditions on the error distribution and the regularity of the regression function f , the following holds asymptotically

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim N(0, \sigma^2(V(\boldsymbol{\theta})' V(\boldsymbol{\theta})/n)^{-1}) \quad [11]$$

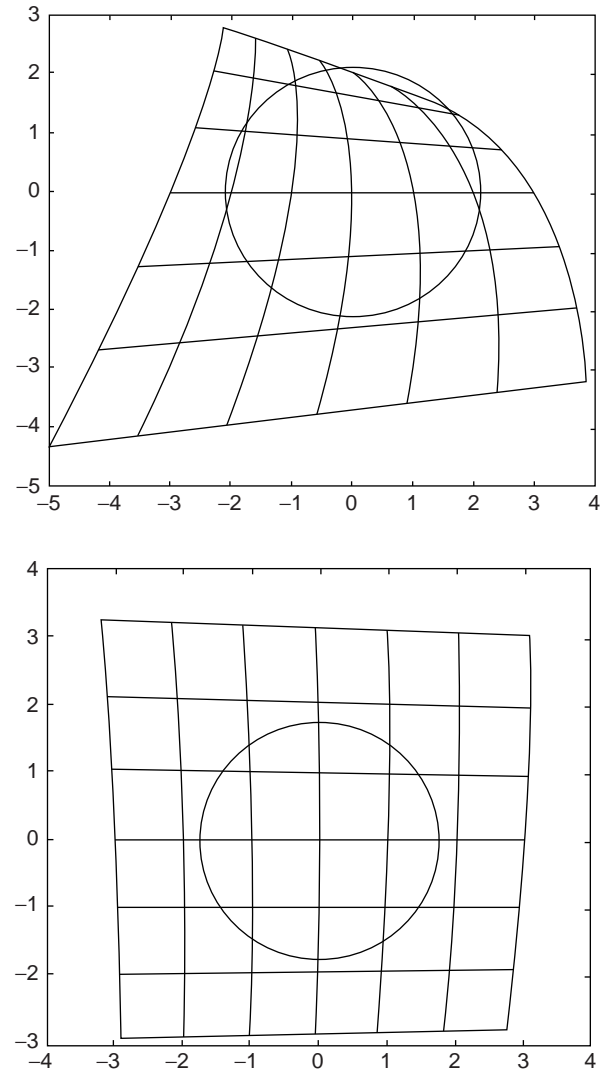


Figure 1 Parameter curves for the orthogonal parameters $\phi = R_{11}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$ using simulated data from the Michaelis–Menten model. The upper figure is with sample size $n = 10$; the lower figure is with sample size $n = 100$.

Using the linear approximation [10] and the asymptotic normality property [11], we can construct a confidence region and make statistical inferences, such as hypothesis testing, concerning $\boldsymbol{\theta}$. The performance of parameter confidence region and inferences depend on how well the linear approximation is, or in another aspect, on the curvedness of the solution locus at the point of least-squares estimate $\hat{\boldsymbol{\theta}}$.

Obtaining the observed information matrix at the convergence, that is, the information matrix evaluated at $\hat{\boldsymbol{\theta}}$, each diagonal element of the inverse of the information matrix provides a minimum variance bound for the corresponding parameter. An approximate $1 - \alpha$ joint confidence region for $\boldsymbol{\theta}$ is

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' V'(\hat{\boldsymbol{\theta}}) V(\hat{\boldsymbol{\theta}}) (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq \rho^2 F(p, n - p; \alpha)$$

or equivalently

$$\phi' \phi \leq \rho^2 F(p, n-p; \alpha)$$

where ϕ is a reparametrization [14] discussed in 'Appendix', and the boundary of this inference region is

$$\{\theta = \hat{\theta} + \rho \sqrt{F(p, n-p; \alpha)} R_{11}^{-1} u \text{ with } \|u\| = 1\}$$

Note that the degree of bias, nonnormality, and excessive variance depend on the model and data. For a given model with a fixed parametrization, data sample size helps to level off the curvedness. As the sample size increases, the sample space dimensionality increases. Consequently, the solution locus becomes flatter and flatter and gets closer to being linear in the higher dimensional sample space (Beale, 1960).

Coordinate System

The coordinate grids on the parameter space, named parameter curves, do not necessarily stay equispaced, straight, and parallel, when first lifted to the solution locus and then projected down to the tangent plane, especially when the locus is highly curved. When projecting the equispaced parameter curves in the parameter space onto the tangent plane, the degree of unequal spacing and the lack of parallelism of the mapped parameter curves lead to a measure of the parameter-effects nonlinearity.

The curvedness of a coordinate system will affect the quality of statistical inferences based on the asymptotic normality. When the mapped parameter curves onto the tangent plane are not uniform grid lines, the resulting confidence region may not be reliable. We use the well-known Michaelis–Menten model [9] to serve as an example for showing the parameter curves on the tangent plane. This model is commonly used for population dynamics in ecology studies, and for pharmacokinetics such as velocity of enzyme reaction. The parameter θ_1 indicates the maximal growth rate or maximal enzyme concentration level, and θ_2 is the half saturation coefficient or called Michaelis–Menten constant. Statistical inferences of these parameters help to understand the growth or increase pattern of a certain species.

The following simulation study illustrates that the parameter curves become more like grid lines in the Michaelis–Menten model, where $\sigma = 40$, $\theta_1 = 100$, $\theta_2 = 0.05$, and x 's come from the absolute values of $N(0, 2)$ random variables. We take two different sample sizes, $n = 10$ and $n = 100$. The mapped parameter curves displayed in **Figure 1** are colored in blue (ϕ_1) and black (ϕ_2). The circle in red prescribes an ideal 95% confidence region for a zero-curvature model and data. We can see in **Figure 1** that the ideal confidence region has gone beyond the square $[-3, 3]^2$ on the tangent plane when $n = 10$, while it stays inside the square for the case when $n = 100$. The radius of the circle is given by $\sqrt{F(2, n-2; 0.05)}$.

Diagnostics and Practical Considerations

The diagnostic study in nonlinear regression models is important for data analysis. The key factors are the nonlinearity of the model and the characteristics of data. After obtaining the least-squares estimates, some diagnostic checks of the magnitude of the intrinsic and parameter-effects nonlinearity are necessary to assess the performance of the estimates and the fitted model. In addition, some practical considerations are discussed.

- *Multicollinearity*: This occurs when some columns in the velocity matrix V are highly correlated and leads to ill-conditioned system of normal equations. It is an indication that the model may be overparametrized, and a simpler model, or a transformation of the regressors or parameters, may be considered. Sometimes, the degree of multicollinearity can be reduced by centering and scaling the regressors (explanatory variables) data (Bates and Watts, 1988).
- *Convergence status of the Gauss–Newton algorithm*: A slow speed of convergence or divergence of the Gauss–Newton algorithm serves as an indication that the model and data combination is not close to linear and the asymptotic normality property at the convergence point is not trustful.
- *Curvature measures at convergence and reparametrization*: Bates and Watts (1981) set bounds on the maximal values of intrinsic and parameter-effects nonlinearity measures in order to determine whether the estimator attains the global minimum of the residual sum of squares; furthermore, they offered a way to evaluate the parameter-effects array under a reparametrization (Bates and Watts, 1988). Nonetheless, little general guidance is provided to choose parameters attaining the minimum of parameter-effects curvatures.
- *Model adequacy*: Although R^2 is a common tool for model adequacy in linear regression analysis, it might be unreasonable to apply it in nonlinear models, as the number of parameters is not related to the number of explanatory variables. Alternatively, other lack-of-fit tests and residual plots are useful in the non-linear case.

Model Building

Model building aims at finding more realistic ways to describe the stochastic behavior observed in data. Models used in data analysis are approximations to the unknown regression function $f(x; \theta)$. It is desired to find nonlinear models that behave close to linear models in estimation and in inferences. In addition to the Michaelis–Menten model [9], we listed a few widely used models that have provided a great deal of useful applications of real data analysis in various scientific fields (Ratkowsky, 1983).

Yield-density model. The following three-parameter models are widely used for modeling relations between yield of a crop and the density of planting.

$$\begin{aligned} f(x; \theta) &= (\theta_1 + \theta_2 x)^{-1/\theta_3} \\ f(x; \theta) &= (\theta_1 + \theta_2 x + \theta_3 x^2)^{-1} \\ f(x; \theta) &= (\theta_1 + \theta_2 x^{\theta_3})^{-1} \end{aligned}$$

The analysis of growth data is important in many fields of study. The following models are practically used for modeling growth data:

1. Weibull model

$$f(x; \theta) = \theta_1 - \theta_2 \exp(-\theta_3 x^{\theta_4})$$

2. Logistic model

$$f(x; \theta) = \frac{\theta_1}{1 + \theta_2 \exp(-\theta_3 x)}$$

3. Richards growth model

$$f(x; \theta) = \frac{\theta_1}{(1 + \exp(\theta_2 - \theta_3 x))^{\theta_4}}$$

There are also other forms of parameterizations.

4. Monomolecular growth model

$$f(x; \theta) = \theta_1 (1 - \exp(-\theta_2 (x - \theta_3)))$$

If we reparameterize by replacing $-\theta_1 e^{\theta_2 \theta_3}$ with θ_2 and $e^{-\theta_2}$ by θ_3 , we have

$$f(x; \theta) = \theta_1 + \theta_2 \theta_3^x \quad (0 < \theta_3 < 1)$$

Some nonlinear regression models are constructed under theoretical consideration, for instance, the Michaelis–Menten model [9]. In physical and chemical fields, there are many models constructed from differential equations (Seber and Wild, 1989). The structural relationships between random variables and the realizations lead to different models; therefore, the assessment of models is very crucial. According to the principle of Occam's razor (Jaynes, 2003), simpler models are preferred than complicated models.

Further Reading

Multiresponse model. The models that we have discussed so far are for single response. When there are more than one response, and when the errors have a joint normal distribution, the least-squares method is still useful. If the normal assumption is not justified, the maximum likelihood or a Bayesian approach (Box and Tiao, 1992) are candidate alternatives.

Levenberg–Marquardt algorithm. The use of the Gauss–Newton algorithm with Levenberg–Marquardt

modifications (Marquardt, 1963) is to speed up the convergence as well as to stabilize the computation for near-singular $V'V$ in the least-squares normal equations. The Levenberg–Marquardt modifications are a compromise between the Gauss–Newton method and the steepest descent method.

Differential geometric view. The measures of curvature described above are based on the geometric properties of the solution locus $f(\theta)$ relative to the parametrization θ . Therefore, a different approach to assessing the degree of nonlinearity is through studying the geometric structures of the distributions. When a normal error assumption is assumed, we are imposing a Euclidean metric on the underlying class of distributions. Changing the error distribution leads to a different metric, thus changing the geometry and concepts of curvature. The study of probability and information by way of differential geometry is known as information geometry. We refer the reader to the work by Amari (1982) and by Amari and Nagaoka (2000) for detailed accounts of differential geometric viewpoints and approaches.

Generalized linear model. The generalized linear models (GLMs) extend the linear regression to include the response variables that are no longer normally distributed, such as binary or count data. If the distribution of response variables belongs to the exponential family, a link function can be applied on the expectation of y to connect to the linear predictors $X\theta$. This makes the expectation surface nonlinear. However, the expectation surface is determined by quantities linear in parameters with a predetermined nonlinear link function. It is thus easier to perform statistical inference under GLMs. Many commercial softwares have been developed for estimation.

Semiparametric and nonparametric regression. If the aim of data analysis is to obtain a good fit to the response curve on the explanatory variables, a nonparametric regression (also known as curve fitting) is probably a better alternative than a nonlinear regression model. The latter is a parametric approach, and its aim is to explore and predict the response at given values of explanatory variables as well as to make statistical inferences based upon interpretation of parameter estimates. The semiparametric regression, which adopts a model that has parametric and nonparametric components, is another alternative to non-linear regression modelling.

Appendix

Factorization of D and A

Due to the symmetric structure of the second derivatives, that is $\partial^2 f(\theta^*) / \partial \theta_\ell \partial \theta_\ell = \partial^2 f(\theta^*) / \partial \theta_\ell \partial \theta_\ell$, there are $p(p-1)/2$ redundant columns in A . Remove these redundant columns, and denote such a resulting matrix by W , which is of size $n \times p(p+1)/2$. Let $D = [V \ W]$, whose

columns are formed by velocity and acceleration vectors. QR-factorize the matrix D as

$$D = \left[\begin{array}{c|c} n \times p & n \times p' \\ \hline \mathbf{Q}_T & \mathbf{Q}_N \end{array} \right] \left[\begin{array}{c} R \\ - \\ 0 \end{array} \right]$$

with

$$R_{(p+p') \times (p+p')} = \begin{bmatrix} p \times p & p \times p' \\ p' \times p & p' \times p' \\ 0 & R_{22} \end{bmatrix}$$

where \mathbf{Q}_T consists of p orthonormal n -vectors and spans the tangent space of the solution locus at θ^* , \mathbf{Q}_N consists of $p' = p(p+1)/2 - p$ orthonormal n -vectors, which are normal to \mathbf{Q}_T and together with \mathbf{Q}_T have the same column span as \mathbf{W} , and \mathbf{R} is a $(p+p')$ square matrix with zero entries in the lower-triangular part. The \mathbf{Q}_0 matrix is composed of the remaining orthonormal column basis. These columns are orthogonal to D and, together with \mathbf{Q}_T and \mathbf{Q}_N , they form a complete orthonormal basis for \mathbb{R}^n . Note that \mathbf{Q}_0 has no contribution to either the intrinsic nor the parameter-effects non-linearity. In fact, we can write an economic QR-factorization:

$$D = [\mathbf{Q}_T \ \mathbf{Q}_N] R \quad \text{with} \quad R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}$$

Note that all the above-mentioned matrices and arrays depend implicitly on θ^* , the current update of the least-squares estimate. We omit θ^* from all these matrices and arrays for notation simplicity when there is no ambiguity.

In the classical geometry, the curvature corresponding to a direction is the ratio of the length of the second derivative to the squared length of the first derivative. In the context of residual sum of squares $S(\theta)$, such a curvature measure combines the first derivatives and the second derivatives in such a way that the nonlinearity due to model and due to parametrization are mixed together. This curvature measure can be decomposed into two components: the intrinsic and the parameter-effects nonlinearity. The second derivative array A can be decomposed accordingly into two components reflecting the model effects and the parameter effects. By left multiplying \mathbf{Q}_T' and \mathbf{Q}_N' to A , we have $A_T = \mathbf{Q}_T' A$ and $A_N = \mathbf{Q}_N' A$ representing, respectively, the component parallel and normal to the tangent plane. Consider a line in the parameter space Θ and its corresponding trajectory along the direction b , or called lifted line, on the solution locus:

$$\text{line in } \Theta : \theta(t) = \theta^* + tb, \text{ lifted line : } f(\theta^* + tb)$$

The second derivative of the lifted line with respect to parameters along the direction b is given by $b' \odot [\mathbf{Q}_T \ \mathbf{Q}_N]' A \odot b$ with decomposition:

$$\overbrace{b' \odot A_T \odot b}^{p \times 1} \oplus \overbrace{b' \odot A_N \odot b}^{p' \times 1},$$

where the direct sum is not a matrix sum but rather is used to denote two components and \odot denotes the slicewise multiplication. For instance, A_N is a $p' \times p \times p$ array with the i th slice a $p \times p$ matrix consisting of entries $A_{N,ijk}$, $j, k = 1, \dots, p$. When the p -vector b and its transpose are right and left multiplied to each slice of A_N , it gives a scalar for each slice. Collectively, $b' \odot A_N \odot b$ forms a p' -vector.

Intrinsic Nonlinearity

The intrinsic curvature corresponding to the direction b at θ^* is given by

$$\frac{\|b' \odot A_N(\theta^*) \odot b\|}{\|V(\theta^*)b\|^2} \quad [12]$$

This quantity is invariant under any reparametrization. Often, we consider a scaled curvature instead:

$$\kappa_b^N = \rho \frac{\|b' \odot A_N(\theta^*) \odot b\|}{\|V(\theta^*)b\|^2} \quad [13]$$

where $\rho = s\sqrt{p}$ with $s^2 = S(\theta^*)/(n-p)$. The quantity ρ is called the standard radius so that $\rho\sqrt{F_{p,n-p;\alpha}}$ is the radius for the $100(1-\alpha)\%$ confidence sphere of a reparametrization ϕ in a neighborhood of θ^* :

$$\phi = R_{11}(\theta - \theta^*) \quad [14]$$

With such a reparametrization, the velocity matrix $\nabla_\phi f(X; \phi)$ will have orthonormal columns, and the second derivative array projected onto the space spanned by \mathbf{Q}_T and \mathbf{Q}_N will become

$$(\mathbf{R}_{11}^{-1})' \odot [\mathbf{Q}_T \ \mathbf{Q}_N]' A \odot \mathbf{R}_{11}^{-1}$$

Then, we have the definition of the relative curvature array as follows:

$$C = \rho (\mathbf{R}_{11}^{-1})' \odot [\mathbf{Q}_T \ \mathbf{Q}_N]' A \odot \mathbf{R}_{11}^{-1}$$

Note that C is a scaled second derivative array with parametrization ϕ [14] and is of the same size as $[\mathbf{Q}_T \ \mathbf{Q}_N]' A$, which is a $(p+p') \times p \times p$ array. Its corresponding decomposition is given by $C = C_T \oplus C_N$ with

$$C_T = \overbrace{\rho (\mathbf{R}_{11}^{-1})' \odot A_T \odot \mathbf{R}_{11}^{-1}}^{p \times p \times p} \quad [15]$$

and

$$C_N = \overbrace{\rho (\mathbf{R}_{11}^{-1})' \odot A_N \odot \mathbf{R}_{11}^{-1}}^{p' \times p \times p} \quad [16]$$

The intrinsic curvature measure can then be expressed as

$$\kappa_d^N = \|d' \odot C_N \odot d\| \quad [17]$$

where d has unit length and the corresponding direction in θ coordinates is $b = \mathbf{R}_{11}^{-1} d$. We can use a maximal value $\max_d \kappa_d^N$ or an average quantity $\int \kappa_d^N$ to account for the intrinsic curvature measure.

Parameter-Effects Nonlinearity

To assess the deviation of parameter curves from the uniform coordinate system, the parameter-effect curvature measure is introduced below

$$\kappa_d^T = \|d' \odot C_T \odot d\| \quad [18]$$

The parameter-effects curvature is the projection of the second derivative array of the solution locus onto the tangent plane, where the derivatives are taken over the parameters of interest. The parameter-effects curvature [15] and the intrinsic curvature [16] are collectively called the relative curvature array. Unlike the invariance of intrinsic curvatures, the parameter-effects curvatures change when a different reparametrization is carried out in the nonlinear model.

Bibliography

- Amari, S. (1982). Differential geometry of the curved exponential family – curvatures and in formation loss. *Annals of Statistics* **10**, 357–368.
- Amari, S. and Nagaoka, H. (2000). *Methods of Information Geometry*. *Transactions of Mathematical Monographs*, Vol. 191. Providence, RI: American Mathematical Society.
- Bates, D. M. and Watts, D. G. (1980). Relative curvature measures of nonlinearity. *Journal of Royal Statistical Society B* **42**, 1–25.
- Bates, D. M. and Watts, D. G. (1981). A relative offset orthogonality convergence criterion for nonlinear least squares. *Technometrics* **23**, 179–183.
- Bates, D. M. and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*. New York: Wiley.
- Beale, E. M. L. (1960). Confidence regions in nonlinear estimation (with discussion). *Journal of Royal Statistical Society, B* **22**, 41–88.
- Box, G. E. P. and Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis*. New York: Wiley.
- Fang, K. T., Lin, D. K. J., Winker, P., and Zhang, Y. (2000). Uniform design: Theory and application. *Technometrics* **42**, 237–248.
- Hamilton, D. C., Watts, D. G., and Bates, D. M. (1982). Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions. *Annals of Statistics* **10**, 386–393.
- Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. New York: Cambridge University Press.
- Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics* **40**, 633–643.
- Marquardt, D. W. (1963). An algorithm for the estimation of non-linear parameters. *SIAM Journal on Applied Mathematics* **11**, 431–441.
- Nocedal, J. and Wright, S. (2006). *Numerical Optimization*, 2nd edn. New York: Springer.
- Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling: A Unified Practical Approach*. New York: Marcel Dekker.
- Ryan, T. (1997). *Modern Regression Methods. Series in Probability and Statistics*. New York: Wiley.
- Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*. New York: Wiley.
- Allen, D. M. (1983). Parameter estimation for nonlinear models with emphasis on compartmental models. *Biometrics* **39**, 629–637.
- Bard, Y. (1974). *Nonlinear Parameter Estimation*. New York: Academic Press.
- Box, G. E. P. and Draper, N. R. (1965). The Bayesian estimation of common parameters from several responses. *Biometrika* **52**, 355–365.
- Clarke, G. P. Y. (1980). Moments of the least squares estimators in a non-linear regression model. *Journal of the Royal Statistical Society, B* **42**, 227–237.
- Cook, R. D. and Tsai, C. L. (1985). Residuals in nonlinear regression. *Biometrika* **72**, 23–29.
- Hamilton, D. C. and Watts, D. G. (1985). A quadratic design criterion for precise estimation in nonlinear regression models. *Technometrics* **27**, 241–250.
- Kass, R. E. (1984). Canonical parametrizations and zero parameter-effects curvature. *Journal of Royal Statistical Society, B* **46**, 86–92.
- Wu, C. F. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics* **9**, 501–513.

Further Reading

Nonparametric Statistical Methods

K Sijtsma and W H M Emons, Tilburg University, Tilburg, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Nonparametric statistical methods – The statistical methods that do not assume a particular distribution, such as the normal, from which the data are sampled.

Introduction

Nonparametric statistical methods are used in situations in which it is unreasonable to assume that the sample was drawn from a population distribution with a particular parametric shape, such as the normal distribution. For example, when the sample size is small and the distribution of the observations is skewed, statistical methods based on the assumption of a normal distribution may be inappropriate. Nonparametric statistical methods do not have *a priori* assumptions about the population distribution and, therefore, are sometimes referred to as distribution-free methods.

Nonparametric methods have several properties that are worth mentioning. First, they are suitable for categorical data, which are common in educational research. Categorical variables have either nominal or ordinal measurement level, but may also represent counts of particular events. Examples of nominal variables include the kinds of arithmetic errors made by students, and the assignment of each student to one of three remedial teaching methods of interest. In statistical analysis, such groups may be identified by means of numbers 1, 2, and 3, which only serve to distinguish them. An example of an ordinal variable is the teacher's rank ordering of students with respect to perceived need of remedial teaching. An example of a count is the frequency of students in each remedial teaching program who receive help.

Nonparametric methods may also be in order for interval-level variables, such as students' motivation, for which the population distribution is unknown. The consequence of using nonparametric methods is that the test scores are treated as ordinal rather than interval variables. By doing this, information is lost about the distances between scores and the power of statistical tests is reduced. Researchers often consider this unacceptable, and analyze

their data as if the sampled scores on a variable stem from a known distribution.

Second, nonparametric methods are appreciated for their weak assumptions. Thus, instead of assuming distributions such as the normal, exact distributions are derived for particular statistics of interest. For example, a nominal variable may follow a multinomial distribution of which the category probabilities are estimated from the data. Other assumptions may be relaxed or completely dropped. For example, instead of assuming a linear relationship between numerical variables as in linear regression, relationships may be estimated from the data, as in kernel smoothing (Fox, 2000a; Ramsay and Silverman, 1997). Weak statistical models are important because they do not make assumptions beyond the level of knowledge expressed in many theories in educational and other social and behavioral sciences.

Third, it is sometimes said that nonparametric statistical models are convenient because they are easy to use. Some restraint may be in order here, because nonparametric methods are often based on complex mathematical considerations that may not be as easy to grasp as those underlying parametric methods. For example, several nonparametric methods are based on combinatorial math, which is notorious for running into badly manageable computations as sample size increases or the numbers of variables increases, and all possible patterns of scores have to be taken into account.

Fourth, the results of nonparametric statistical methods are sometimes not much different from those obtained by means of parametric methods that have been applied even when the assumptions on which they are based were violated. Thus, sometimes parametric statistical methods are robust against the violation of the normality assumption and the use of nonparametric counterparts, which may be more appropriate from a mathematical point of view, becomes less salient. This is true, for example, for the parametric Student's *t*-test. However, for other methods, such as regression analysis, it has been shown repeatedly that the estimation of a relationship from the data may lead to different and interesting results that would have been obscured when a particular parametric function, such as the linear or the logistic, had been fitted. Thus, there is room for nonparametric statistical methods, which are also illustrated by some examples provided in the article.

Examples of Nonparametric Statistical Methods

The goal of statistical analysis is to estimate properties of interest from the data, such as the distribution of a variable or the difference between groups with respect to the means on a variable, and to test hypotheses about interesting research questions. First, we review a few well-known nonparametric statistical tests from the myriad of test procedures (e.g., Siegel, 1956; Siegel and Castellan, 1988; Wasserman, 2004, 2006) based on known, exact sampling distributions, and then we discuss methods for obtaining sampling distributions and probabilities of exceedence when exact sampling distributions are unknown.

Estimating Distributions and Drawing Inferences

An important step in data analysis is to inspect the distribution of the observations. Rather than assuming a normal distribution and estimating the mean and variance, one can also estimate the complete distribution from the data. An example of a nonparametric method is the simple histogram that estimates the population distribution directly from the data (**Figure 1**). As the sample has a limited number of observations, the histogram is discrete and jagged. Thus, sometimes it may be convenient to smooth the histogram. Kernel smoothing produces the result shown by the solid curve in **Figure 1**. Smoothing has the effect of bringing out the salient features of the distribution at the expense of irregularities that probably are due to sampling error. For example, the solid smooth

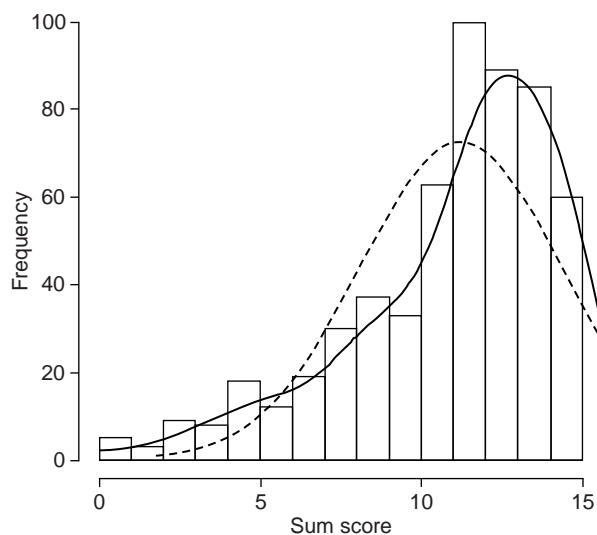


Figure 1 Example of an estimated distribution of sum scores (e.g., the number-correct score on an educational test) by means of a histogram, kernel smoothing (solid curve), and a normal approximation (dashed curve).

curve suggests that the distribution is skewed to the left and that the frequency of score 12 may be too large due to sampling error.

The smooth dashed curve in **Figure 1** shows the normal approximation to the histogram. A glance at the graph shows that the normal seems to overestimate the lower-score frequencies and underestimate the higher-score frequencies. The Shapiro–Wilk test, which is a well-known nonparametric test for evaluating whether the observations deviate from the normal curve, yields a value equal to 0.894 ($P < 0.000$); thus, the hypothesis of normality is rejected. The Kolmogorov–Smirnov test is a more general, often-used nonparametric method that can be used to test whether the data come from a hypothesized distribution, such as the normal. Often, it has less power than the Shapiro–Wilk test to detect violations of normality, but for the data in **Figure 1** the value of the test statistic is 0.183 ($P < 0.000$). Again, normality is rejected. Neither test makes assumptions about the population distribution of the data.

Researchers may wish to know whether the distributions of an outcome variable are the same across groups, for example, as in a control-group study on the effect of a teaching program. When the variable is ordinal, it does not make sense to assume that it follows a normal distribution and compare the means using a parametric t -test. Such a test is also inappropriate when an interval variable does not have a normal distribution. Alternatively, the nonparametric Mann–Whitney U test, also known as the Wilcoxon rank-sum test, is a good candidate for testing the null hypothesis that two independent samples come from the same population against the alternative that the samples come from two different population distributions which are identical in shape but different in location.

The Mann–Whitney U test is based on the common rank ordering according to ascending magnitude of all observations from two samples, say A with size N_1 and B with size N_2 , and counts for each observation from B by how many observations from A it is preceded; test statistic U equals the sum of these counts (for simplicity, we ignore the possibility of ties). When the two distributions are completely separated, each observation from A either precedes each observation from B , which results in maximum $U = N_1 N_2$, or is preceded by each observation from B , which results in minimum $U = 0$. Thus, high and low values of U indicate different distributions, and intermediate values indicate largely overlapping distributions. For small sample sizes (i.e., the size of the larger sample does not exceed 20), the probabilities of exceedence can be read from tables especially prepared for this purpose. When the larger sample exceeds size 20, U has been shown to approach a normal distribution with

$$\mu_U = \frac{N_1 N_2}{2}, \text{ and } \sigma_U = \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12}}.$$

The standard normal statistic $z = (U - \mu_U)/\sigma_U$ may be used to test the null hypothesis using probabilities from tables for the normal distribution. Several other tests for testing differences in location exist for a variety of research designs, such as the Wilcoxon signed rank test for paired samples, and the Kruskal–Wallis test for comparing differences in location between k ($k \geq 3$) independent samples.

Determining Degree of Association

Suppose, two numerical variables X and Y have been sampled from the same population and one wishes to know the degree to which they are linearly related. An extremely well-known statistic for this purpose is Spearman's product-moment correlation, denoted ρ in the population and r in the sample. Assuming bivariate normality for X and Y and a sample of size N , the null hypothesis that $\rho = 0$ can be tested using a t statistic, which can be shown to depend only on r and N .

Again, what if variables X and Y are ordinal or one is reluctant to assume bivariate normality? An alternative to the product-moment correlation is Spearman's rank correlation r_s . Here, the scores on both variables are replaced by the corresponding ranks, which for brevity we also denote X and Y . For example, for a respondent indexed i , the pair of ranks ($X_i = 7, Y_i = 15$) show that (s)he was ranked 7 on variable X and 15 on variable Y . The difference between paired ranks equals $d_i = 7 - 15$, and so on for other respondents. Tied scores are replaced by the average of the ranks assigned, had no ties been observed. The product-moment correlation between the two rankings X and Y provide the values of r_s . Ignoring ties for simplicity, and using differences d between paired ranks, r_s equals

$$r_s = 1 - \frac{6 \sum d_i^2}{N(N^2 - 1)}.$$

For small samples, the null hypothesis that $\rho_s = 0$ can be evaluated by considering all possible rank orderings or permutations of Y given a particular rank ordering of X , and determining the value of r_s for each pair of ranks X and Y . As under the null hypothesis, each permutation of Y has the same probability, $(N!)^{-1}$, the probability of a particular value for r_s simply is a multiple of $(N!)^{-1}$, and a table based on the distribution of r_s given N is readily prepared. This table can be used for hypothesis testing. For, say, $N \leq 30$, this is feasible but for larger N one runs into combinatorial problems. For $N \geq 10$ (Hays, 1981: 598), a t -test may be used for testing the null hypothesis. Other coefficients exist for expressing degree of association, such as Kendall's τ_b , which corrects for tied observations (Liebetrau, 1983: 51–53).

When X and Y are nominal, a two-way contingency table is set up with cells for all combinations (x, y) and

frequency counts in each cell, which reflect how often a particular combination is observed in the sample. The null hypothesis of no association may be tested using a chi-squared statistic. The strength of the association may be expressed by several coefficients, for example, the phi-coefficient when both X and Y have two categories, and Cramér's V when the number of categories is greater and not necessarily equal for X and Y (Liebetrau, 1983).

Bootstrap and Permutation Tests

When the population distribution of the observations is unknown, it may happen that the sampling distribution of a statistic of interest, say T , derived from the data also is unknown. In this case, a computer-intensive way of approximating the sampling distribution of T is to draw a large number of samples of size N with replacement from the sample, and then compute statistic T for each so-called bootstrap sample. The distribution of T across the bootstrap samples can be used to determine a confidence interval for parameter τ of which T is the estimate, for example, by identifying the 2.5th and the 97.5th percentile and using these as lower and upper bounds, respectively. This procedure is known as the nonparametric bootstrap (Efron and Tibshirani, 1993). Statistic T can be any quantity of interest, such as the median or the range of a distribution, a difference between group medians, or an association measure between scores.

An example of a related procedure is known as the permutation test (also, see the discussion of the test of the null hypothesis that Spearman's $\rho_s = 0$). In general, for null hypothesis $\tau = 0$, the statistic T is determined for all possible arrangements of the sampled data, which are also known as permutations. The distribution of T across these permutations is the sampling distribution of T under the null hypothesis. If one wishes testing the null hypothesis $\tau = 0$ at the 5% level against the alternative that $\tau > 0$, then the 95th percentile of the sampling distribution defines T_{crit} . If the observed value of T in the original sample, denoted T_{obs} , exceeds T_{crit} , the null hypothesis is rejected. The number of possible permutations can become excessively large, even for high-powered computers, and then a large random sample of permutations may be used instead. For some statistics T , the distribution of T approaches a known distribution in larger samples, as with Spearman's r_s , which approaches a Student's t -distribution.

Specific Nonparametric Methods: Nonparametric Regression

As regression analysis is regularly used in educational research, this section presents nonparametric regression analyses as an illustration of a more advanced nonparametric

statistical method. First, the parametric multiple regression analysis is discussed. This method relates a response variable, denoted Y , linearly to one or more explanatory variables, denoted $X_j (j = 1, \dots, m)$, such that

$$Y = \alpha + \sum \beta_j X_j + \varepsilon.$$

In the regression equation, the regression intercept is denoted by α , the regression coefficients by $\beta_j (j = 1, \dots, m)$, and the residual error by ε , which is assumed to have 0 mean for fixed values of the explanatory variables. The regression parameters are estimated fitting the regression equation to the data of N observations (indexed by i) so as to minimize $\sum \varepsilon_i^2$. Hypotheses of interest are whether $\beta_j = 0 (j = 1, \dots, m)$ against a one-sided or a two-sided alternative, and whether the amount of variance explained by the model, denoted R^2 , equals 0 (i.e., $R^2 = 0$), against the alternative that it is positive. For testing these hypotheses, it is assumed (among others) that the conditional distribution of Y is normal, with mean $\alpha + \sum \beta_j X_j$ and constant variance σ^2 . This amounts to assuming that $\varepsilon \sim N(0, \sigma^2)$.

An Overview of Nonparametric Regression Methods

Suppose, the researcher has insufficient evidence to support linearity, normality, and equal variance, or his/her substantive theory does not imply this degree of structure. Then, (s)he could try nonparametric regression methods to study the relationships between variables. We illustrate that the exact shape as estimated directly from the data may provide interesting information about relationships.

For simplicity, we consider only models with one explanatory variable, X (see Fox, 1997, 2000b, for many generalizations). Then, when one expects Y and X not to be linearly related, the simplest option is to compute the mean \bar{Y} conditional on separate values of X , and take the mean as the regression estimate; that is, $\hat{Y}|x = \bar{Y}|x$ is computed and a graph is drawn that connects adjacent conditional estimates $\hat{Y}|x$. Problems may arise when X is continuous and many different values are observed so that only few values of Y are tied to one value of X , when the number of different values of X is manageable but the total sample size is relatively small, which leads to the same problem, or when the combination of both occurs.

The problem of too few observations per value of X may be accommodated by the use of so-called bins. Order all observations according to X from small to large, and define a subset of smallest adjacent X values to form the first bin, then a subset of the next adjacent values of X to form the second bin, and so on. Bins are indexed $k = 1, \dots, K$. Then, \bar{Y} is computed based on all observations in the k th bin ($k = 1, \dots, K$). The resulting values $\hat{Y}|\text{bin}(k)$ are plotted in a graph, which often appears jagged. Important decisions in binning concern the

width of a bin and the minimum number of observations in a bin. These decisions affect both the bias and the variance of the estimate of mean Y , and smaller bias often implies greater variance and vice versa.

Bins may also be defined as overlapping windows, where each next window has moved further to the right across the scale of X , so that values of X enter the window from the right and exit the window from the left. As for bins, windows may have either a fixed width defined by values of X or they contain a fixed number of $n + 1$ observations. In local averaging, for each window \bar{Y} is computed on the basis of all the observations on X that are in the window. When the number of observed values on X is small, the resulting graph is jagged, as with binning. Windows that contain a fixed number of $n + 1$ observations can be moved across all N observations on X after these observations have been ordered from small to large, such that each observation can be the central focal value x_0 of a window once. The other observations in the window can be defined to be the n nearest neighbors of x_0 . Then, N conditional estimates $\hat{Y}|x_0$ (the conditioning is on x_0 plus its n nearest neighbors) lead to a relatively smooth curve. As with binning, either the width of the window or the size of n must be determined. Another problem is that with fixed n , x values near the endpoints have more-similar neighborhoods than other x values, so that the regression curve tends to flatten near the endpoints.

Kernel smoothing takes local averaging a step further by differentially weighing the y values corresponding to neighbors of focal point x_0 , such that y values of neighbors close to x_0 receive more weight than y values of neighbors further away in the window. Denoting weights by w_i the local average is obtained by means of

$$\hat{Y}|x_0 = \frac{\sum_{i=1}^N w_i y_i}{\sum_{i=1}^N w_i}.$$

Weight w_i is the kernel function, defined as $w_i = K[(x_i - x_0)/b]$ (b is explained shortly). Several choices are possible but a convenient choice is the Gaussian kernel,

$$w_i = K\left(\frac{x_i - x_0}{b}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - x_0/b)^2},$$

in which b is called the bandwidth. Small bandwidth values produce jittery curves showing too much random detail (too much variance) and large values produce smooth curves that erase salient features of the regression (too much bias). Thus, finding a value of b that counterbalances bias and precision is an important topic.

The data within a window may be distributed approximately equally dense but they may also pile up on one side of the window. In the latter case, estimate $\hat{Y}|x_0$ may be heavily biased. Local polynomial regression provides a

better estimate by fitting a polynomial regression model to the data in the window,

$$Y = \alpha + \beta_1(X - x_0) + \beta_2(X - x_0)^2 + \cdots + \beta_p(X - x_0)^p + \varepsilon,$$

using weighted least squares, such that $\sum_{i=1}^n w_i^2 \varepsilon_i^2$ is minimized so as to obtain estimated regression parameters. Then, for this window $\hat{Y}|x_0 = \hat{\alpha}$ is the point on the regression curve, and an estimated curve is obtained as the window moves along. As for higher p , multicollinearity sets in, $p = 1, 2$, and 3 are convenient practical choices.

With spline regression (Marsh and Cormier, 2002; Ramsay and Silverman, 1997), we return to binning but instead of taking an unweighted average we now fit a polynomial regression model,

$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \cdots + \beta_p X^p + \varepsilon,$$

to the data in the bin such that the regression curves in adjacent bins connect smoothly to constitute one smooth curve across the bins. Not only should curves from adjacent bins connect the points on the boundaries of the bins, also called knots, but in passing a knot, the slope of the curve should not change abruptly, meaning that the first derivative is smooth at the knot, and the curvature of the curve, for example, should not to change abruptly from smooth to rough, meaning that the second derivative also is smooth at the knot. This requires a polynomial to have at least degree $p = 3$, because linear spline functions ($p = 1$) only connect straight lines between bins, and quadratic spline functions ($p = 2$) in addition only guarantee a smooth change of the slope but do not prevent sudden breaks in this change. See Green and Silverman (1994) for methods that control the balance between bias and variance and prevent curves from becoming jumpy.

Both local polynomial regression and spline regression are extremely flexible and overcome many of the weaknesses of other nonparametric regression methods, such as the flattening of curves near the endpoints. Both methods use parametric functions to adequately describe interesting features of relationships but not to hypothesize that relationships are naturally linear, quadratic, and so on. Inference with respect to nonparametric regression is based on estimated confidence bands around the regression curve. Several possibilities exist, the nonparametric bootstrap being one of them. In addition, nested models may be tested against one another.

An Application of Kernel Smoothing to Educational Test Data

Nonparametric regression methods find their application, for example, in a large and important class of models for educational measurement, known as item response models, discussed elsewhere in this encyclopedia. These models use the scores (binary, nominal, and

ordered) of a large sample of students (often, $N \gg 500$) on several items that measure the same ability, for constructing an ability scale on which students can be located. Each item may present, for example, an arithmetic problem to the student and the scores on each of the items may be driven by a student's ability to perform well on such problems.

The key feature of item response models is the nonlinear regression of each separate item score on the ability. Let the random variable for an item score be denoted X_j . It is further assumed that the ability is represented by a latent variable, denoted by θ , the scores on which are inferred from the student's observed item scores through the estimation of the item response model. For $X_j = 0, 1$ (e.g., incorrect/correct solutions), a typical example of a regression function is the 2-parameter logistic function (e.g., Van der Linden and Hambleton, 1997),

$$P(X_j = 1|\theta) = \frac{e^{a_j(\theta - b_j)}}{1 + e^{a_j(\theta - b_j)}};$$

see **Figure 2** for the graphical representation of two logistic functions. Parameter b_j gives the value of θ for which the probability of a correct solution equals 0.5, and parameter a_j is monotonically related to the function's steepest slope, which is in the inflexion point $(b_j, 0.5)$. For parameter estimation, marginal maximum likelihood (MML) may be used, which assumes that θ is normal when the item parameters are estimated, and then assumes that the estimated item parameters are fixed when for each individual an ability score θ is estimated.

An important question is whether a 2-parameter logistic regression function and a normal θ are not unduly restrictive for the problem of interest (Junker and Sijtsma, 2001).

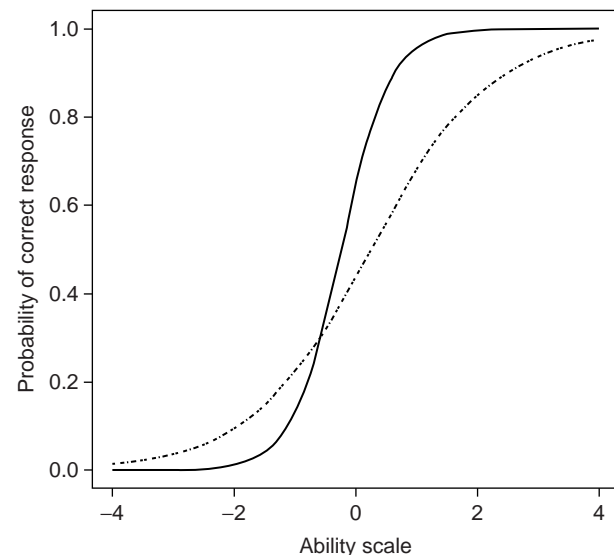


Figure 2 Two 2-parameter logistic functions: solid curve has parameters $b_j = -0.25$ and $a_j = 2$, and dashed curve has parameters $b_j = 0.25$ and $a_j = 1.0$.

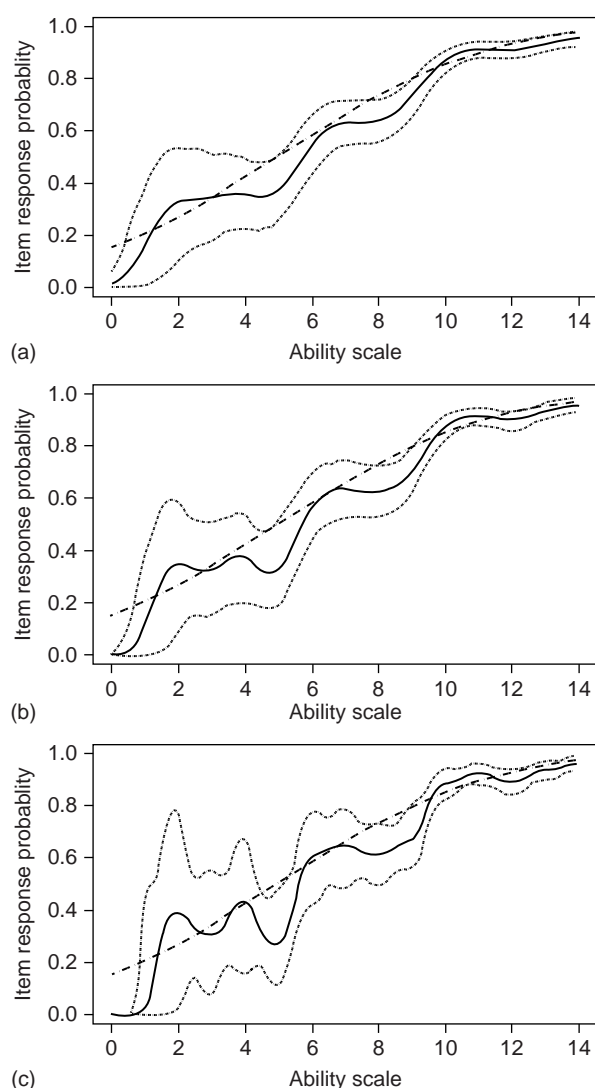


Figure 3 Estimated item response functions (solid curves) with 95% confidence envelopes (dotted curves): bandwidth parameters (a) $h = 2$; (b) $h = 1.5$, and (c) $h = 1.0$. The dashed curve in each figure is the estimated 2-parameter logistic function.

Indeed, in the context of item response theory, binning, kernel smoothing, and spline regression each have been proposed and used as alternatives for logistic and other parametric functions. **Figure 3** shows the estimated kernel regression functions of an item response function from a 15-item arithmetic test, for three levels of the bandwidth parameter: $h = 2$ (**Figure 3(a)**), $h = 1.5$ (**Figure 3(b)**), and $h = 1$ (**Figure 3(c)**). The dashed curve is a 2-parameter logistic curve based on MML estimates of the two item parameters, a_j and b_j . The solid curve is the estimate based on kernel smoothing, and the dotted curves are the 90% confidence envelopes. As h becomes smaller, the regression curve becomes jumpier. Each nonparametric estimate suggests that the logistic estimate often seems

to overestimate the true-response probabilities, and does not adequately capture the slope of the regression function at different ranges of the ability scale.

The results illustrate the usefulness of nonparametric data analysis for revealing systematic trends in the data that might have gone unnoticed had parametric methods been used. For example, the nonparametric regression estimate shows that the item does not discriminate well in the ability range from score 2 to score 5 whereas the slope parameter of the parametric regression function cannot reveal this kind of detail. Thus, the nonparametric estimate provided a more realistic impression of how the item functions for measuring ability differences between students.

Further Reading

Introductory textbooks on nonparametric statistics are, for example, Hollander and Wolfe (1999), Sheskin (2007), and Siegel (1956; Siegel and Castellan, 1988). An example of a multivariate nonparametric method that may be of special interest for educational researchers is the nonparametric approach to multiple analysis of variance (MANOVA) (Puri and Sen, 1971) to test group differences on multiple outcomes (e.g., several cognitive skills). Finch (2005) showed that the nonparametric MANOVA has good statistical properties in situations in which, for example, the normality assumptions underlying the parametric MANOVA are violated. Another example includes the work by Linting *et al.* (2007), who discuss nonlinear principal components analysis based on the monotone transformation of ordinal variables such that the relationship with other variables is optimized. For nonparametric regression techniques, the availability of high-speed computers has led to a rapid development of computer-intensive methods, such as bootstrapping and resampling, linear and nonlinear smoothing, and graphical methods, which can handle complex multivariate data (e.g., Akritas and Politis, 2003; Wasserman, 2006).

Bibliography

- Akritas, M. C. and Politis, D. N. (2003). *Recent Advances and Trends in Nonparametric Statistics*. Amsterdam: Elsevier.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Finch, H. (2005). Comparison of the performance of nonparametric and parametric MANOVA test statistics when assumptions are violated. *Methodology* 1, 27–38.
- Fox, J. (1997). *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage.
- Fox, J. (2000a). *Nonparametric Simple Regression. Smoothing Scatterplots*. Thousand Oaks, CA: Sage.
- Fox, J. (2000b). *Multiple and Generalized Nonparametric Regression*. Thousand Oaks, CA: Sage.
- Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. London: Chapman and Hall.

- Hays, W. L. (1981). *Statistics*, 3rd edn. New York: Holt, Rinehart, and Winston.
- Hays, W. L. (1994). *Statistics*, 5th edn. Fort Worth, TX: Harcourt Brace College Publishers.
- Hollander, M. and Wolfe, D. A. (1999). *Nonparametric Statistical Methods*, 2nd edn. New York: Wiley.
- Junker, B. W. and Sijtsma, K. (2001). Nonparametric item response theory in action: An overview of the special issue. *Applied Psychological Measurement* **25**, 211–220.
- Liebetrau, A. M. (1983). *Measures of Association*. Beverly Hills, CA: Sage.
- Linting, M., Meulman, J. J., Groenen, P. J. F., and Van der Kooij, A. J. (2007). Nonlinear principal components analysis: Introduction and application. *Psychological Methods* **12**, 336–358.
- Marsh, L. C. and Cormier, D. R. (2002). *Spline Regression Models*. Thousand Oaks, CA: Sage.
- Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. New York: Wiley.
- Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. New York: Springer.
- Sheskin, D. J. (2007). *Parametric and Nonparametric Statistical Procedures*, 4th edn. Boca Raton, FL: Chapman and Hall/CRC Press.
- Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. Tokyo: McGraw-Hill Kogakusha.
- Siegel, S. and Castellan, N. J., Jr. (1988). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.
- Van der Linden, W. J. and Hambleton, R. K. (1997). Item response theory: Brief history, common models, and extensions. In van der Linden, W. J. and Hambleton, R. K. (eds.) *Handbook of Modern Item Response Theory*, pp 1–28. New York: Springer.
- Wasserman, L. A. (2004). *All of Statistics. A Concise Course in Statistical Inference*. New York: Springer.
- Wasserman, L. A. (2006). *All of Nonparametric Statistics*. New York: Springer.

Observational Studies*

P R Rosenbaum, University of Pennsylvania, Philadelphia, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Propensity score – The conditional probability of treatment given observed covariates.

Sensitivity analysis – An analysis that shows how relaxing an assumption might alter a conclusion.

Introduction

Definition

William G. Cochran (1965) defined an observational study as a comparison of treated and control groups in which:

the objective is to elucidate cause-and-effect relationships [... in which it] is not feasible to use controlled experimentation, in the sense of being able to impose the procedures or treatments whose effects it is desired to discover, or to assign subjects at random to different procedures.

Treatments, policies, and interventions that are expected to benefit recipients can be studied in experiments that randomly allocate experimental subjects to treatment groups; therefore, comparable subjects receive competing treatments (Boruch, 1997). Observational studies are common, nonetheless, in most fields that study the effects of treatments on people, because harmful or unwanted treatments cannot be imposed on human subjects. When treatments are not randomized, subjects assigned to competing treatments may not be comparable; therefore, differences in outcomes may or may not be effects caused by the treatments (Rubin, 1974).

Examples

Two examples of observational studies are described. Later sections refer to these examples.

Long-term psychological effects of the death of a close relative

Lehman *et al.* (1987) attempted to estimate the long-term psychological effects of the sudden death of a spouse or a child in a car crash. Starting with 80 bereaved spouses and parents, they paired 80 similar individuals who were not bereaved – that is, 80 matched controls – drawn from 7581 individuals who came to renew a drivers license, matching for gender, age, family income before the crash, education level, and number and ages of children. Contrasting their findings with the views of Bowlby and Freud, who had claimed bereavement has brief effects, they concluded:

Contrary to what some early writers have suggested about the duration of the major symptoms of bereavement ... both spouses and parents in our study showed clear evidence of depression and lack of resolution at the time of the interview, which was 5 to 7 years after the loss occurred.

Effects on children of occupational exposures to lead

Morton *et al.* (1982) asked whether children were harmed by lead brought home in the clothes and hair of parents who were exposed to lead at work. They matched children whose parents worked in a battery factory to unexposed control children of the same age and neighborhood, comparing the level of lead found in the children's blood, finding elevated levels of lead in exposed children.

In addition, they compared exposed children whose parents had varied levels of exposure to lead at the factory, finding that parents who had higher exposures on the job, in turn, had children with more lead in their blood. Finally, they compared exposed children whose parents had varied hygiene upon leaving the factory at the end of the day, finding that poor hygiene of the parent predicted higher levels of lead in the blood of the child.

Central Issues

Without random assignment of treatments, treated and control groups may not have been comparable prior to treatment; therefore, differing outcomes may reflect either biases from these pretreatment differences or effects actually caused by the treatment. The difficulty in distinguishing treatment effects from biases is the key consequence of the absence of randomization.

*With permission of the publisher, this article uses material from Rosenbaum, P. R. (2001). Observational Studies: Overview. In Smelser, N. and Baltes, P. (eds.) *International Encyclopedia of Social and Behavioral Sciences*, pp 10810–10815. New York: Elsevier.

A variable measured prior to treatment is not affected by the treatment and is called a covariate. A variable measured after treatment may have been affected by the treatment and is called an outcome. An analysis that does not carefully distinguish covariates and outcomes can introduce biases into the analysis where none existed previously (Rosenbaum, 1984).

A pretreatment difference between treated and control groups is called an overt bias if it is accurately measured in the data at hand, and it called a hidden bias if it is not measured. For instance, if treated subjects are observed to be somewhat older than controls, and if age is recorded, then this is an overt bias. If treated subjects consumed more illegal narcotics than controls, and if accurate records of this are not available, then this is a hidden bias. Typically, overt biases are immediately visible in the data at hand, while hidden biases are a matter of concerned speculation and investigation.

In most observational studies, an effort is made to remove overt biases using one or more analytical methods, such as matched sampling, stratification, or model-based adjustments such as covariance adjustment. Adjustments are discussed in the section titled 'Adjusting for overt bias'. Once this is accomplished, attention turns to addressing possible hidden biases, including efforts to detect hidden biases and to study the sensitivity of conclusions to biases of plausible magnitude. Hidden biases are discussed in the section title 'Detecting hidden bias', 'Sensitivity to hidden bias', and 'Reducing Sensitivity to hidden bias'.

Adjusting for Overt Biases

Matched Sampling

Selecting from a reservoir of potential controls

Matching is the most direct and intuitive method of adjustment, in which an effort is made to compare each treated individual to one or more comparable controls. Matched sampling is most common when a small treated group is available together with a large reservoir of potential controls (Rubin, 1979). For instance, in the study of bereavement by Lehman *et al.* (1987), there were 80 bereaved spouses and parents and 7581 potential controls, from whom 80 matched controls were selected.

The structure of the bereavement study is typical: computerized administrative records aid in identifying and matching treated and control subjects, but additional information is required from matched subjects for research purposes; in this case, psychiatric outcomes are required. Often, it is not practical to obtain the needed research data from all controls in the reservoir and, instead, a sample of controls is used. Matched sampling selects controls who appear comparable to treated subjects.

Goals of matching

Matching attempts to do three things: (1) produce matched pairs or sets that appear similar in terms of observed covariates, (2) produce treated and control groups with similar distributions of observed covariates, and (3) remove that part of the bias in estimated treatment effects that is due to overt biases or imbalances in observed covariates. For instance, with pairs matched for age, goal (1) would be accomplished if each treated subject were matched to a control of the same age, whereas goal (2) would be accomplished if the entire treated group had an age distribution that closely resembled the age distribution in the matched control group, for instance, the same mean age, the same quartiles of the age distribution, and so on.

It is possible to demonstrate (Rosenbaum and Rubin, 1983, theorem 4) that goal (1) suffices, but is not necessary for goal (2), and that goal (2) suffices for goal (3). Pairs that are individually matched for age will balance age; however, it is also possible to balance age with matched pairs that are not individually matched for age, and any matching that balances the distribution of age removes the part of the bias in estimated treatment effects due to imbalances in age. This is fortunate because it is difficult to match individually on many covariates at once, but it is not so difficult to balance many covariates at once. Modern matching algorithms focus on balancing covariates, while comparability within pairs plays a secondary role. A tool in balanced matching is the propensity score.

Propensity score

The propensity score is the conditional probability of receiving the treatment rather than the control, given the observed covariates (Rosenbaum and Rubin, 1983). Note carefully that the propensity score is defined in terms of the observed covariates even if there may be hidden biases due to unobserved covariates. In the simplest randomized experiment, treatment or control is assigned by the flip of a fair coin, and the propensity score equals 1/2 for all subjects no matter what covariates are observed. In an observational study, subjects with certain observed characteristics may be more likely to receive either treatment or control; therefore, the propensity score varies with these observed characteristics.

Matching on one variable, namely the propensity score, is often practical and tends to balance all of the observed covariates. Matching on the propensity score together with a few other observed covariates – say, matching on the propensity score together with age and gender – also tends to balance all of the observed covariates. If it suffices to adjust for the observed covariates – that is, if there is no hidden bias due to unobserved covariates – then it also suffices to adjust for the propensity score alone. These results are theorems 1 through 4 of Rosenbaum and Rubin (1983). In practice, the propensity

score is unknown and must be estimated, perhaps using a logit regression of treatment on covariates.

Structure of matched sets

Although pair matching is common, matching with multiple controls may reduce the standard error of the estimated treatment effect (Smith, 1997). Haviland *et al.* (2008) match with two controls in a study of the effects on teens of joining a gang.

It is not necessary to match every treated subject to the same fixed number of controls. In terms of bias reduction, the optimal form is a full matching in which each matched set may have one treated subject and several controls or one control and several treated subjects (Rosenbaum, 2002; § 10). In this way, when a certain pattern of observed covariates is typical of controls, then a treated subject with this pattern will have several controls; however, if, instead, the pattern is typical of treated subjects, several treated subjects may share one control. Hansen (2004) used full matching in an observational study of the effects of coaching on Scholastic Aptitude Test (SAT) scores.

Matching algorithms

Matching algorithms often express the difference in covariate values between a treated subject and a potential control in terms of a distance. Optimal matching minimizes the total distance within matched sets by solving what is known as a minimum cost flow problem, that is, a mathematical problem which originated in the field of operations research. The use of optimal matching in observational studies is discussed in Rosenbaum (2002, §10) and Hansen's (2007) implementation, called *optmatch*, is freely available in the statistical package R (R Development Core Team, 2007).

Matching and thick description

Since matching focuses attention on comparable pairs or sets of individuals under alternative treatments, it is straightforward to combine a quantitative study of many matched pairs with a qualitative study of a small subset of the matched pairs (Rosenbaum and Silber, 2001). For instance, quantitative results about all pairs might be reported along with narrative or thick descriptions of a few pairs.

Stratification

An alternative to matching is stratification, in which subjects are grouped into strata. Rosenbaum and Rubin (1983) balanced 74 covariates using five strata formed from an estimated propensity score. The optimal stratification – that is, the stratification that makes treated and control subjects as similar as possible within strata – is a full matching as described in the section entitled 'Structure of matched sets'. See Rosenbaum (2002, §10) for proof and

specifics, and Hansen (2007) for implementation of full matching in R.

Model-Based Adjustments

Matched sampling and stratification compare treated subjects directly to actual controls who appear comparable in terms of observed covariates. For instance, as noted earlier, Lehman *et al.* (1987) compared bereaved individuals to matched controls who were comparable in terms of age, gender, income, education, and family structure. In contrast, model-based adjustments, such as covariance adjustment, use data on treated and control subjects without regard to their comparability, relying on a model, such as a linear regression model, to predict how subjects would have responded under treatments they did not receive.

In simulation studies, Rubin (1979) compared matching, covariance adjustment, and combinations of these techniques, concluding that the use of covariance adjustment within matched pairs was superior to either method alone, being both robust and efficient. Dehejia and Wahba (1999) illustrate the hazards of model-based adjustments in the absence of matching.

Detecting Hidden Bias

In an observational study, the investigator takes active steps to detect hidden biases – to collect data recording visible traces of unobserved pretreatment differences, if they exist. In this, the investigator is aided by an elaborate theory, defined by Sir Ronald Fisher in the following discussion from Cochran (1965, 85).

About 20 years ago, when asked in a meeting what can be done in observational studies to clarify the step from association to causation, Sir Ronald Fisher replied: "Make your theories elaborate." The reply puzzled me at first, since by Occam's razor, the advice usually given is to make theories as simple as is consistent with known data. What Sir Ronald meant, as subsequent discussion showed, was that when constructing a causal hypothesis one should envisage as many different consequences of its truth as possible, and plan observational studies to discover whether each of these consequences is found to hold.

Consider the study of lead exposures by Morton *et al.* (1982), discussed in the section titled 'Effects on children of occupational exposures to lead'. Their elaborate theory made three predictions: (1) higher lead levels in the blood of exposed children than in matched control children, (2) higher lead levels in exposed children whose parents had higher exposure on the job, and (3) higher lead levels in exposed children whose parents practiced poorer hygiene upon leaving the factory. Since each of these predictions was consistent with observed data, to attribute

the observed associations to hidden bias, rather than an actual effect of lead exposure, one would need to postulate biases that could produce all three associations. Additionally, these observations are consistent with other studies of lead exposure. The use of elaborate theories and further examples are discussed in Shadish *et al.* (2002) and Rosenbaum (2002, 2004).

The simplest and most common design to detect hidden bias uses two control groups. Since neither control group received the treatment, systematic differences between the control groups cannot be effects of the treatment and must instead be some form of hidden bias. One seeks two control groups that would differ in their outcomes if a specific unobserved bias were present. In this way, hidden biases should produce a difference between the two control groups, whereas an actual treatment effect should produce relatively similar results in the two control groups and a different result in the treated group; see Campbell (1969), Meyer (1995), and Rosenbaum (2002, § 8).

Sensitivity to Hidden Bias

The analytical adjustments discussed in the section entitled ‘Adjusting for overt biases’ can often remove overt biases accurately recorded in the data at hand, but there is typically a legitimate concern that treated and control groups differed prior to treatment in ways that were not recorded. A sensitivity analysis asks how such hidden biases might alter the conclusions of the study.

The first formal method of sensitivity analysis was developed to aid in appraising the effects of cigarette smoking on human health. Responding to objections that smoking might not cause lung cancer, but rather that there might be a genetic predisposition both to smoke and to develop lung cancer, Cornfield *et al.* (1959: 40) wrote:

... if cigarette smokers have 9 times the risk of nonsmokers for developing lung cancer, and this is not because cigarette smoke is a causal agent, but only because cigarette smokers produce hormone X, then the proportion of hormone X-producers among cigarette smokers must be at least 9 times greater than among nonsmokers.

This statement was an important conceptual advance. It is, of course, commonly understood that association does not imply causation – that any observed association can be explained by a hypothetical argument, perhaps an implausible hypothetical argument, postulating an unobserved variable. Although this common understanding is always correct, it contributes little to our appraisal of an observational study because it makes no reference to what was actually observed in this empirical investigation, and no reference to scientific knowledge about the phenomenon under study (Bross, 1960). A sensitivity analysis replaces

the statement – association does not imply causation – by a specific statement about the magnitude of hidden bias that would need to be present to explain the associations actually observed. Strong associations in large studies can only be explained by large, perhaps implausible, biases. Aspects of the method of Cornfield *et al.* (1959) are discussed by Greenhouse (1982) and Gastwirth (1992). Although important as a conceptual advance, this method of sensitivity analysis is limited in that it ignores sampling error produced by having a finite sample rather than the population as a whole, and it is applicable only to binary outcomes. Rosenbaum (2002, §4) discusses a related method of sensitivity analysis that addresses sampling error and may be used with outcome measures of all kinds. See Imbens (2003) and Diprete and Gangl (2004) also.

Reducing Sensitivity to Hidden Bias

The Goal: Sharper Effects, Smaller Biases

Given the choice, one would prefer conclusions insensitive to hidden biases, so that there is less ambiguity about the effects caused by the treatment. A general quantitative tool for investigating the impact of the design of an observational study on its sensitivity to unobserved biases is the design sensitivity (Rosenbaum, 2004). This section discusses three strategies, namely, choice of circumstances, coherent hypotheses, and reducing unit heterogeneity.

Choice of circumstances

In the earliest stages of planning an observational study, the investigator chooses the circumstances in which the study will be conducted. These choices often affect both the size of the treatment effect and the magnitude of hidden bias that is plausible. As a result, these choices affect the sensitivity of the study to hidden bias. Some illustrations of this use of choice will be briefly mentioned. See Rosenbaum (2002, §11) for detailed discussion of these cases and others with examples, and see Meyer (1995) and Angrist and Kreuger (2001) for related discussion.

The investigator examines a broad research hypothesis, one that makes predictions about innumerable circumstances, but examines that hypothesis in particular circumstances where dramatic effects are anticipated, free of distortions from other factors that affect these same outcomes. The investigator seeks a situation in which there is a genuine control group completely shielded from the treatment, and a treated group that received the treatment at an intense dose. One prefers a treatment that is imposed suddenly, not gradually, and haphazardly, not in direct response to characteristics of the individuals under study.

For instance, in the study in the section entitled ‘Long-term psychological effects of the death of a close relative’

by Lehman *et al.* (1987), the effects of bereavement were studied following sudden deaths of close relatives from car crashes. Moreover, employing standardized criteria, they used only those car crashes for which the victim's car was not responsible, reasoning that alcohol or drug abuse or certain forms of psychopathology might increase the risk of a car crash and also be associated with unfavorable psychological outcomes in the family.

As another example, Angrist and Lavy (1999) studied the effects of class size on academic achievement in carefully chosen circumstances. In Israel, a rule dating back to Maimonides requires classrooms with no more than 40 students; therefore, a class with 41 students is divided in two classes of about 20 students. Whereas in the United States, a class of size 40 is likely to be found in a different economic environment from a class of size 20, in Israel, a comparatively haphazard event – the enrollment of an additional child – often separates these two situations.

Coherent hypotheses

A coherent hypothesis makes many highly specific predictions about the effect of a treatment on several outcome variables, perhaps in several treatment groups, probably involving one or more dose–response relationships. Sir Austin Bradford Hill (1965) claimed coherence as an important criterion in appraising evidence that a treatment is the cause of the outcomes with which it is associated. Coherence is related to pattern matching as discussed by Trochim (1985) and Shadish *et al.* (2002).

Informally, one feels that if numerous predictions of a causal theory are simultaneously confirmed, it should be more difficult to attribute this to hidden bias – the study should be less sensitive to bias. This is sometimes true, sometimes false, but when it is true, the reduction in sensitivity to bias can, in principle, be substantial. See Rosenbaum (2002, §9, 2004) and the references there for discussion.

Heterogeneity and causality

Consider two observational studies, one with a larger sample size but more heterogeneous units or individuals, the other with a smaller sample size but less heterogeneous units. Here, heterogeneity refers to the variability in outcomes after adjustment for observed covariates, say the variability in treated-minus-control differences in pairs matched for observed covariates. In a randomized experiment, the larger sample size in the first study might offset its greater variability, so that the two studies are essentially equivalent, say producing confidence intervals of the same length or hypothesis tests with the same power. In contrast, in observational studies, the smaller, less heterogeneous study will typically be better, in the sense that a treatment effect of the same magnitude in the two studies will be less sensitive to unobserved biases in this study (Rosenbaum, 2005). For instance, Ashenfelter

and Rouse (1998) compared the earnings of identical twins with different education, and Card and Krueger (1994) compared Burger Kings affected by a change in minimum wage laws to Burger Kings in an adjacent state that were unaffected. Other examples are discussed in Rosenbaum (2005).

Conclusion

An observational study examines the effects of a treatment in a context in which random assignment of subjects to treatment or control is infeasible for ethical or practical reasons. Unlike a randomized experiment, in an observational study, treated and control groups may not be comparable prior to treatment; therefore, differing outcomes after treatment may not be effects caused by the treatment. If the treated and control groups differ in ways that have been accurately measured, then adjustments are often possible, for instance, by matching treated subjects to controls who are similar in terms of measured pretreatment covariates. Invariably, there is the concern that some important covariate has not been measured; therefore, adjustments for measured covariates may fail to render the groups comparable. Addressing this concern is a central task in the design and analysis of an observational study. One tactic tries to detect bias from unmeasured covariates, for instance, with the aid of two control groups. A second tactic uses sensitivity analysis to appraise how biases of various magnitudes might alter the conclusions. A third tactic tries to design the study to reduce its sensitivity to unobserved biases. The most convincing studies use these tactics together so that each one removes some of the ambiguities left by the others.

Acknowledgment

This work was supported by the US National Science Foundation.

Bibliography

- Angrist, J. and Krueger, A. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives* **15**, 69–85.
- Angrist, J. D. and Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* **114**, 533–575.
- Ashenfelter, O. and Rouse, C. (1998). Income, schooling and ability: Evidence from a new sample of identical twins. *Quarterly Journal of Economics* **113**, 253–284.
- Boruch, R. (1997). *Randomized Experiments for Planning and Evaluation*. Thousand Oaks, CA: Sage.
- Bross, I. D. J. (1960). Statistical criticism. *Cancer* **13**, 394–400.
- Campbell, D. T. (1969). Prospective: Artifact and control. In Rosenthal, R. and Rasnow, R. (eds.) *Artifact in Behavioral Research*, pp 351–382. New York: Academic Press.

- Card, D. and Krueger, A. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review* **84**, 772–793.
- Cochran, W. G. (1965). The planning of observational studies of human populations (with discussion). *Journal of the Royal Statistical Society A* **128**, 134–155.
- Cornfield, J., Haenszel, W., Hammond, E., et al. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute* **22**, 173–203.
- Dehejia, R. and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association* **94**, 1053–1062.
- Diprete, T. A. and Gangl, M. (2004). Assessing bias in the estimation of causal effects. *Sociological Methodology* **34**, 271–310.
- Gastwirth, J. L. (1992). Methods for assessing the sensitivity of statistical comparisons used in Title VII cases to omitted variables. *Jurimetrics* **33**, 19–34.
- Greenhouse, S. (1982). Jerome Cornfield's contributions to epidemiology. *Biometrics* **22**(Supplement).
- Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association* **99**, 609–618.
- Hansen, B. B. (2007). Optmatch: Flexible, optimal matching for observational studies. *R News* **7**, 18–24.
- Haviland, A., Nagin, D. S., Rosenbaum, P. R., and Tremblay, R. (2008). Combining group-based trajectory modeling and propensity score matching for causal inferences in nonexperimental longitudinal data. *Developmental Psychology* **44**, 422–436.
- Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine* **58**, 295–300.
- Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review* **93**, 126–132.
- Lehman, D., Wortman, C., and Williams, A. (1987). Long-term effects of losing a spouse or a child in a motor vehicle crash. *Journal of Personality and Social Psychology* **52**, 218–231.
- Meyer, B. D. (1995). Natural and quasi-experiments in economics. *Journal of Business and Economic Statistics* **13**, 151–161.
- Morton, D., Saah, A., Silberg, S., et al. (1982). Lead absorption in children of employees in a lead-related industry. *American Journal of Epidemiology* **115**, 549–555.
- R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation, <http://www.R-project.org>.
- Rosenbaum, P. R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society, Series A* **147**, 656–666.
- Rosenbaum, P. R. (2002). *Observational Studies*, 2nd edn. New York: Springer.
- Rosenbaum, P. R. (2004). Design sensitivity in observational studies. *Biometrika* **91**, 153–164.
- Rosenbaum, P. R. (2005). Heterogeneity and causality. *American Statistician* **59**, 147–152.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.
- Rosenbaum, P. R. and Silber, J. H. (2001). Matching and thick description in an observational study of mortality after surgery. *Biostatistics* **2**, 217–232.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688–701.
- Rubin, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association* **74**, 318–328.
- Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton-Mifflin.
- Smith, H. L. (1997). Matching with multiple controls to estimate treatment effects in observational studies. *Sociological Methodology* **27**, 325–353.
- Trochim, W. M. K. (1985). Pattern matching, validity and conceptualization in program evaluation. *Evaluation Review* **9**, 575–604.

Further Reading

- Angrist, J. D., Imbens, G., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association* **91**, 444–469.
- Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation*. Boston, MA: Houghton Mifflin.
- Platt, J. R. (1964). Strong inference. *Science* **146**, 347–353.

Order Statistics

I Bairamov, Izmir University of Economics, Izmir, Turkey

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Censored sample – A sample appearing in the life-testing experiments when n items are kept under observation and only part of this sample can be observed.

Counting process – A stochastic process representing the total number of certain event occurring up to time.

Distribution free statistics for a class of distribution function – A statistic whose distribution is the same for all distribution function in the class.

Empirical distribution function – A natural estimation of the cumulative distribution function constructed on the base of the sample.

Extreme order statistics – The maximal and minimal order statistics.

Markov property – This property states that to make predictions of the behavior of a system in the future, it suffices to consider only the present state of the system and not past history.

Order Statistics – Original random sample arranged in order of magnitude.

Probability integral transformation – A transformation that transforms an arbitrary distribution to the uniform distribution.

Robust estimators – Estimators that are efficient in presence of outliers in the sample.

Two sample problem – A hypothesis testing problem verifying whether two sample are identically distributed or not.

The Subject of Order Statistics

The independent and identically distributed random variables, which can be interpreted as results of an experiment measuring values of a certain random variable arranged in order of magnitude, are called order statistics. In the statistical model of many experiments, for instance, in reliability analysis, life time studies, the analysis of time to graduation of students, and testing of strength of materials, the realizations arise in nondecreasing order; therefore, the use of order statistics is necessary. Order statistics are extensively used in statistical inferences: in estimation theory and hypothesis testing.

Let X_1, X_2, \dots, X_n denote a random sample from a population with cumulative distribution function (cdf) $F(x)$. Suppose that the elements of this sample are arranged in order of magnitude and $X_{(1)}$ denotes the smallest; $X_{(2)}$ denotes the second smallest; etc., and $X_{(n)}$ denotes the largest of the set X_1, X_2, \dots, X_n . Then $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ denotes the original random sample arranged in increasing order of magnitude, and these are called the order statistics of the sample X_1, X_2, \dots, X_n . We call $X_{(i)}$, for $1 \leq i \leq n$ the i th order statistic. The subject of order statistics deals with the distributional properties of $X_{(i)}$ itself, and some functions of the subset of the n order statistics and their applications. If, for example, the scores of n students in the exam are X_1, X_2, \dots, X_n then $X_{(n)}$ represents the score of the best student; $X_{(1)}$ is the score of the weakest; the sample range $W = X_{(n)} - X_{(1)}$ is a measure of dispersion; the sample median, defined as $(X_{(n/2)} + X_{(n/2 + 1)})/2$ for an n even and as $X_{[(n+1)/2]}$, for n odd, is a measure of location and estimate the central tendency of scores. Here, $[a]$ is the integer part of the number a . The sample midrange, defined as $(X_{(1)} + X_{(n)})/2$, is also a measure of central tendency.

Order statistics have wide applications in many areas where the use of an ordered sample is important. Order statistics are among the most fundamental tools in non-parametric statistics, because the transformation $U_{(i)} = F(X_{(i)})$ produces a random variable which is the i th order statistics from the uniform population on the interval $(0, 1)$, and therefore $U_{(i)}$ is distribution free, that is, its distribution function is independent of the distribution function F of the original sample. This transformation is called the probability integral transformation.

It is well known from classical statistical theory that the natural estimate of an unknown distribution function is the empirical distribution function, which is a function of order statistics. Therefore, many important statistics in estimation theory and hypothesis testing appear to be an integral functional of the empirical distribution function, and can be expressed in terms of order statistics. Order statistics do not change their order under probability integral transformation, namely if $U_{(i)} = F(X_{(i)})$, $i = 1, 2, \dots, n$, then $U_{(1)} \leq U_{(2)} \leq \dots \leq U_{(n)}$. Due to unique distribution free properties, they are widely used in nonparametric interval estimation and hypothesis testing.

Order statistics and their properties have been extensively studied since the early part of the last century, and recent years have seen a particularly rapid growth of studies. The multiauthored book *Contributions to Order*

Statistics, edited by A. H. Sarhan and B. G. Greenberg, appeared in the Wiley series in probability and statistics in 1962. The first monograph, *Order Statistics* by H. David appeared in 1970 in the same Wiley series and has served as a text, a survey of growth, and a general introduction. The second edition appeared in 1981 and the third, coauthored with H. Nagaraja, in 2003. For further reading the reader is referred to Arnold *et al.* (1992) and Balakrishnan (2007).

Basic Distribution Theory

The elements of the sample X_1, X_2, \dots, X_n are independent and identically distributed (iid), but the order statistics $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are dependent random variables. The distribution of the r th order statistics can be derived using the independence of the random variables X_1, X_2, \dots, X_n and observing that the event $\{X_{(r)} \leq t\}$ occurs if and only if at least r of the observations X_1, X_2, \dots, X_n falls below t . Therefore, taking into account the fact that the probability of occurrence of exactly i of events $\{X_k \leq x\}$ in n independent Bernoulli trials is $\binom{n}{i} F^i(x)(1-F(x))^{n-i}$, the cdf of $X_{(r)}$ can be written as

$$F_r(x) = P\{X_{(r)} \leq x\} = \sum_{i=r}^n \binom{n}{i} F^i(x)(1-F(x))^{n-i} \quad [1]$$

If F is absolutely continuous with probability density function (pdf) f , then [1] can also be rewritten as follows:

$$F_r(x) = \frac{n!}{(r-1)!(n-r)!} \int_0^{F(x)} u^{r-1}(1-u)^{n-r} du \quad [2]$$

Formula [1] holds true for both discrete and continuous distribution functions. Formula [2] is true only for absolutely continuous distributions. Given the realizations of the n order statistics to be $x_{(1)} < x_{(2)} < \dots < x_{(n)}$, the original random variables X_i are restrained to take on the values $x_{(i)}$ ($i = 1, 2, \dots, n$) which by symmetry assigns equal probability to each of the $n!$ permutations of $(1, 2, \dots, n)$. Therefore, the joint pdf of all n order statistics is

$$f_{1,2,\dots,n}(x_1, x_2, \dots, x_n) = n! \prod_{i=1}^n f(x_i) \text{ for } x_2 < x_1 < \dots < x_n$$

Since $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are dependent random variables, then their joint distributions are important. The expressions for the joint pdf's of two or more order statistics can be found in David (1981).

Order statistics from uniform distribution on $[0, 1]$ are important when one needs to generate the order statistics from any distribution using Monte Carlo simulation. If $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are order statistics from the population with cdf F and $F^{-1}(u) = \inf\{x: F(x) \geq u\}$ is the inverse of F , then $F^{-1}(U_{(i)}) = X_{(i)}$, the equality here is in distribution. There are various methods of generating uniform random

variables. Using computer simulation we generate sample U_1, U_2, \dots, U_n from the uniform distribution in $[0, 1]$ and then order the sample. The $X_{(i)}$ value then can be calculated as $X_{(i)} = F^{-1}(U_{(i)})$. For various methods of generating order statistics, see Tadikamalla and Balakrishnan (1998).

The pdf of $W_{rs} = X_{(s)} - X_{(r)}$ when the parent population is uniform in $[0, 1]$ depends only on $s-r$ and not on r and s individually. In addition, the pdf of the sample range $W = X_{(n)} - X_{(1)}$ is $fW_{1n}(x) = n(n-1)x^{n-2}(1-x), 0 \leq x \leq 1$.

Let $X_{(1)}, \dots, X_{(n)}$ be the order statistics based on the sample X_1, X_2, \dots, X_n with cdf $F(x) = 1 - \exp(-\lambda x), x \geq 0$. Then the spacings $Y_1 = X_{(1)}, Y_2 = X_{(2)} - X_{(1)}, \dots, Y_n = X_{(n)} - X_{(n-1)}$ are independent; furthermore, the random variables $Z_r = (n-r+1)\lambda(X_{(r)} - X_{(r-1)}), r = 1, 2, \dots, n$ are iid with cdf $F(x) = 1 - \exp(-x), x \geq 0$, where $X_0 = 0$. If n units are placed under strength test and X_1, X_2, \dots, X_n are independent random variables with exponential distribution and represent the life lengths of these units, then the lengths of time intervals $X_{(r)} - X_{(r-1)}, r = 1, 2, \dots, n$ between two failures are independent and identically distributed random variables. Then $X_{(r)} = \sum_{i=1}^r Z_i / \lambda(n-i+1)$ that is, $X_{(r)}$ can be represented as a sum of i random variables. Then the conditional distribution of $X_{(r+1)}$ given $X_{(1)} = x_1, X_{(2)} = x_2, \dots, X_{(r)} = x_r$ is the same with the conditional distribution of $X_{(r+1)}$ given $X_{(r)} = x_r$. This means that $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ satisfy Markov property and form an additive Markov chain. The Markov property states that to make predictions of the behavior of a system in the future, it suffices to consider only the present state of the system and not the past history. The sequence of dependent random variables satisfying the Markov property is called the Markov chain. This property helps in establishing the Markovian dependence structure of order statistics from the sample with any continuous distribution. It follows that the order statistics $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ from a population with continuous cdf form a Markov chain.

There are some interesting properties of order statistics connected with truncation of these ordered observations. For instance, let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be order statistics of the sample X_1, X_2, \dots, X_n with absolutely continuous cdf F and pdf f . Then given $X_{(r)} = x, X_{(s)} = y$ the joint pdf of $(X_{(r+1)}, X_{(r+2)}, \dots, X_{(s-1)})$ is the same with the joint pdf of the order statistics $(Y_{(1)}, Y_{(2)}, \dots, Y_{(s-r)})$ from the sample Y_1, Y_2, \dots, Y_{s-r} size $s-r$, where Y_i has pdf of the random variable X_i given $x < X_i < y$.

A counting process $\{N(t), t \geq 0\}$ representing the total number of event A occurring up to time t is called a Poisson process if it has stationary and independent increments. Note that, a stochastic process $\{N(t), t \geq 0\}$ is said to have independent increments if, for all $t_0 < t_1 < t_2 < \dots < t_m$, the random variables $N(t_1) - N(t_0), N(t_2) - N(t_1), \dots, N(t_m) - N(t_{m-1})$ are independent. It possesses stationary increments if $N(t+b) - N(t)$ has the same distribution for all t . There is an interesting connection

between the interarrival times of the occurrence of event A and the order statistics. Let $N(t)$ be a Poisson process with rate λ , $\lambda > 0$, then $P\{N(t) = k\} = e^{-\lambda t}(\lambda t)^k/k!$ ($k = 0, 1, 2, \dots$). Denote by X_1 the time of the first event, X_2 the time between the first and the second event, X_n the time between $(n-1)$ th and n th event. Then the sequence $\{X_n\}_{n \geq 1}$ is called the sequence of interarrival times. It is well known that X_1, X_2, \dots, X_n are iid exponential random variables having a mean of $1/\lambda$. Another quantity is $S_n = \sum_{i=1}^n X_i$, $n \geq 1$, the arrival time of, or waiting time until, the n th event. Then given that $N(s) = n$, the n arrival times S_1, \dots, S_n have the same distribution as the order statistics corresponding to the independent random variables uniformly distributed on the interval $[0, s]$, i.e.

$$\begin{aligned} P\{S_1 \leq t_1, S_2 \leq t_2, \dots, S_n \leq t_n | N(s) = n\} \\ = P\{U_{(1)} \leq t_1, \dots, U_{(n)} \leq t_n\} \\ = \frac{n!}{s^n}, 0 < t_1 < t_2 < \dots < t_n, \end{aligned}$$

where $U_{(i)}$ is the i th order statistic from uniform in $[0, s]$ distribution.

Order Statistics in Statistical Inference

Order statistics are essential in several optimal inference procedures and hypothesis testing problems. In many cases when the underlying distribution has finite support, the order statistics themselves become sufficient statistics and, thus, provide minimum variance unbiased estimators and the most powerful test procedures for the unknown parameters. Note that, if the random variable X takes values from the interval $[a, b]$, then we say that this random variable has support $[a, b]$, $-\infty \leq a < b \leq \infty$. If the random variable has finite support and if the support involves the parameters of distribution of this random variable, then many of estimates of parameters involve order statistics. For example, let X be uniform in $[0, \theta]$ random variable, that is, the pdf of X is $f(x; \theta) = 1/\theta$ if $0 \leq x \leq \theta$. It is well known that $X_{(n)}$ is a sufficient, and complete statistic for θ and $\frac{n+1}{n}X_{(n)}$ is an unbiased estimator of θ .

The order statistics appear in a natural way in the inference procedures when the sample is censored and only part of the sample values are available. The censored samples appear in the life-testing experiments when n items are kept under observation until failure. These items could be technical systems or their components, patients under certain drug or clinical conditions, or candidates undergoing exams in complex conditions or under time pressure.

In measuring performance on examinations many teachers use average $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and sample variance $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, where X_1, X_2, \dots, X_n are test scores of students. It is well known that poor lecture

attendance is associated with lower test scores (see, e.g., Myles and Henderson, 2002, Williams *et al.*, 2002; Balch, 1992). The presence of a student with poor attendance will probably lead to a lower test score for the class, and this outlier will affect the sample mean \bar{X} and variance S^2 . Therefore, the existence of outliers will result in misestimation of class performance. To avoid this kind of problem in applications, we need to find estimators only minimally affected by the presence of outliers. In statistical literature such estimators are called robust estimators. One popular robust estimator of the center of a symmetric distribution is the symmetric trimmed mean

$$\hat{\theta}_r = \frac{1}{n-2r} \sum_{i=r+1}^{n-r} X_{(i)}, \quad 0 \leq r \leq \left[\frac{n-1}{2} \right]$$

where we have trimmed the top r and the bottom r order statistics. The trimmed means give less weight to the sample extremes and are suggested as robust estimators. They are robust against the presence of a small number of outliers and highly efficient in their complete absence. Barnett and Lewis (1994) describe an estimator:

$$\hat{\theta} = \begin{cases} \hat{\theta}_0, & \text{if } \max\{(\bar{X} - X_{(n)}), (X_{(n)} - \bar{X})\} < c \\ \hat{\theta}_1 & \text{otherwise} \end{cases}$$

for the mean μ of a sample from an $N(\mu, 1)$, where *distribution suspect one of the values is from $N(\mu + \delta, 1)$* , and $\hat{\theta}_0$ and $\hat{\theta}_1$ are trimmed means and c can be the specified percentile of the distribution of $\max |X_i - \bar{X}|$. This statistic is used for testing for a single extreme outlier. Balakrishnan (2007) computed the bias and mean square error of robust estimators constructed in the base of order statistics and presented tables of numerical values for $n = 10$. For a further reading on robust statistics, see Huber (1981), Andrews *et al.* (1972), and David and Ghosh (1985).

Order Statistics and Education

Order statistics play an important role in educational statistics. In many statistical analyses, the information from a random sample is utilized through the ordered values of the sample. At the beginning of the course 'nonparametric statistics' the students face the nontrivial operation of ordering of a random sample. On the one hand, theoretically the elements of the random sample are random variables, that is, measurable functions $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$ ($\omega \in \Omega$) given in the probability space $\{\Omega, F, P\}$, where Ω is a sample space, F is a σ -algebra of subsets of the sample space, and P is a probability measure. On the other hand, the sample values X_1, X_2, \dots, X_n are considered as realizations of the experiment measuring values of the random variable X , namely, they are numbers. In deriving the cdf of r th order statistics for better

understanding of the idea of proof, we consider the random sample as numbers, when they are actually random variables. Understanding the structure of order statistics presents difficulties when considering the original sample as functions. For example, the distribution function of the maximal order statistic $X_{(n)}$ is $F_n(x) = P\{X_{(n)} \leq x\}$ and if we consider X_1, X_2, \dots, X_n as numbers, then the maximum is less than or equal to x if all numbers are less than or equal to x . Therefore, the event $\{X_{(n)} \leq x\}$ occurs if and only if all the members of original sample are less than or equal to x , namely $\{X_1 \leq x, X_2 \leq x, \dots, X_n \leq x\}$. Now, since the original X 's are independent and all have the same cdf F we can state

$$F_n(x) = P\{X_{(n)} \leq x\}$$

$$P\{X_1 \leq x, X_2 \leq x, \dots, X_n \leq x\} = F^n(x)$$

In general, the cdf of r th order statistic, $P\{X_{(r)} \leq x\} =$

$$P\{\text{at least } j \text{ } X\text{'s are } \leq x\} = \sum_{i=j}^n \binom{n}{i} P\{\text{exactly } i \text{ } X\text{'s are } \leq x\},$$

here again we use the independence of the original sample. This approach is useful for understanding the structure of order statistics that are actually measurable functions of the elements of sample space.

As pointed out above the probability integral transformation $U = F(X)$ transforms the order statistics $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ to uniform order statistics $U_{(1)}, U_{(2)}, \dots, U_{(n)}$ and preserves the order. In hypothesis testing and confidence intervals, the field of nonparametric statistics relies on the concept known as distribution-free property. For the distribution-free hypothesis test, the significance level remains constant over a class of underlying distributional assumptions. The distribution-free or invariant confidence interval has a constant confidence level holding over a class of distribution functions. The test statistic $S = S(X_1, X_2, \dots, X_n)$ is designated distribution free over some class of distributions, say \mathcal{F} if the distribution of S is the same for every distribution in \mathcal{F} . The view that many students have difficulties understanding these concepts seems to be a common one among my colleagues teaching nonparametric statistics. However, this problem can be overcome using order statistics to construct effective examples. For example, let $\mathcal{F} = \mathcal{F}_\theta$ be a scale parameter class; this means that if $F_\theta(x) \in \mathcal{F}_\theta$ then $F_\theta(x) = F(x/\theta)$, for some distribution function F and parameter θ . Then the distribution of statistic defined as midrange divided by the range, that is,

$$T = \frac{1X_{(n)} + X_{(1)}}{2X_{(n)} - X_{(1)}}$$

does not depend on parameter θ ; in other words, this statistic is distribution free for the class \mathcal{F}_θ . As another example, let $\mathcal{F} = \mathcal{F}_c$ be the class of all continuous distribution functions and X_1, X_2, \dots, X_n be a sample from the

population with distribution function $F \in \mathcal{F}_c$, and let X_{n+1} be the $(n+1)$ th observation from the same population independent of X_1, X_2, \dots, X_n , then the probability that X_{n+1} falls into interval $(X_{(r)}, X_{(s)})$ is $(s-r)/(n+1)$. This probability is the same for all distribution functions F from the class \mathcal{F}_c , which means that the interval $(X_{(r)}, X_{(s)})$ constructed by the order statistics is the distribution-free confidence interval for the future observation X_{n+1} . It is interesting to note that if the distribution is continuous under some regularity conditions, the interval $(X_{(r)}, X_{(s)})$ is the only distribution-free interval for the future observation X_{n+1} (see Bairamov and Petunin, 1990).

The importance of order statistics can also be seen in teaching the theory of the ranking statistics used in two-sample problem with unknown shift parameters. The sample observation X_i is said to have rank R_i among X_1, X_2, \dots, X_n if $X_i = X_{(R_i)}$, where $X_{(R_i)}$ is the R_i th order statistic. Let X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_m be independent random samples from continuous distributions with distribution functions $F(x)$ and $G(x) = F(x - \theta)$, respectively, where $-\infty < \theta < \infty$ is an unknown shift parameter. The Mann-Whitney-Wilcoxon nonparametric test for verifying the null hypothesis $H_0: \theta = 0$ against alternative $H_1: \theta > 0, \theta < 0$ or $\theta \neq 0$ is constructed based on the distribution-free property of the rank statistic $W = \sum_{i=1}^n R_i^*$ under hypothesis H_0 , where R_j is the rank of R_j among the m X 's and n Y 's combined and treated as a single set of observations. In a general two-sample problem, when $H_0: F = G$ is to be tested against a general class of alternatives $H_1: F \neq G$, the Kolmogorov-Smirnov test based on the distance $D_{n,m} = \sup_{-\infty < x < \infty} |F_n(x) - G_m(x)|$ of two empirical distribution functions $F_n(x)$ and $G_m(x)$ of the samples X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_m , respectively, is one of the most important consistent hypothesis tests. The empirical distribution function $F_n(x)$, defined as the number of observations X_1, X_2, \dots, X_n less than or equal to x divided by n , has the following expression in terms of order statistics:

$$F_n(x) = \begin{cases} 0, & \text{if } x < X_{(1)} \\ \frac{k}{n} & \text{if } X_{(k)} \leq x < X_{(k+1)} \\ 1 & \text{if } x \geq X_{(n)} \end{cases} \quad \text{for } k = 1, 2, \dots, n-1$$

Example

The selection of students from different schools for a scholarship shortlist has been an issue of public interest. In many countries there is a general consensus on the existence of bias in selection of candidates. Fairness has been defined in a variety of ways; Torndike (1971) proposed a definition of fairness; Cole (1973) made a fundamental assumption in the reviewed models that the applicants are independently and identically distributed

random variables. The model proposed by Olkin and Stephens (1993) used ordered scores of students, namely, order statistics. More specifically, let X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_m be test scores for two groups of students representing two different high schools A and B , respectively. We assume that X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_m are random samples from continuous distributions F_1 and F_2 , respectively. The $m + n$ scores are pooled and jointly ranked, and the top k students are shortlisted for scholarships. Let R be the number of students from the first group that enter the shortlist. We are interested in the probability $P\{R = r\}$ that exactly r students from high school A appear in the shortlist. Olkin and Stephens (1993) provide an elegant solution to this problem. If $F_1 = F_2$, in other words, we consider two different groups of students from identical high schools, then this probability depends only on n, m, r , and k and can therefore be easily calculated. For the special case of choosing a single candidate, that is, $r = k = 1$, the probability that exactly one student from high school A will appear in the shortlist is equal to $n/(n + m)$, which is the probability that the maximal score $X_{(n)}$ of the first group is greater than the maximal score of the second group $Y_{(m)}$, where $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ and $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(m)}$ are the ordered scores of two groups. A numerical analysis shows that when n is low relative to m , this probability is low, which may mean that a student from a small but prestigious school has considerably less chance of being shortlisted when in competition with a larger school. Some numerical values of probability that exactly r students from the high school A enter the shortlist for $r = 1$ and $r = 2$, and $k = 1, 2, \dots, 10$ are given below:

1. $n = 9, m = 30$

k	1	2	3	4	5	6	7	8	9	10
$P\{1, k; 9, 30\}$	0.11	0.12	0.23	0.34	0.37	0.42	0.45	0.46	0.47	0.47

2. $n = 3, m = 25$

k	1	2	3	4	5	6	7	8	9	10
$P\{1, k; 3, 25\}$	0.	0.01	0.02	0.04	0.07	0.10	0.14	0.17	0.21	0.25

Let \mathbf{P} denote the probability of the event that the best candidate from school A with score $X_{(n)}$ is included in the top group of k candidates; thus, at least one student from A is in the list. We provide a numerical example of $n = 9$ and $m = 30$:

k	1	2	3	4	5	6	7	8	9	10
\mathbf{P}	0.23	0.41	0.56	0.67	0.75	0.82	0.878	0.91	0.93	0.95

From the table it can be seen that a shortlist of size at least $k = 4$ is needed in order to guarantee that the probability of at least one student from the group of size $n = 9$ being included in the shortlist is at least 0.905.

In the general case, when $F_1(\cdot) \neq F_2(\cdot)$, the probability that exactly r students from high school A appear in the shortlist is equal to $P\{X_{(n+1-r)} > Y_{(m+r-k)}\} - P\{X_{(n-r)} > Y_{(m+r-k+1)}\}$. In particular, when $F_1(x)$ is $N(\mu, 1)$ distribution and $F_2(y)$ is the $N(0, 1)$ distribution, selected values of the probability that r out of n students from the second high school are chosen on a shortlist of length k are presented in Olkin and Stephens (1993).

Summary

The theory of order statistics is essential in statistical analysis and its applications. Order statistics play an important role in inferential problems including estimation of unknown parameters of distributions in considered statistical models and in hypothesis testing. Before 1970s, most studies were on cases where order statistics originated from independent and identically distributed random variables. In the early 1970s however the robustness issues motivated the study of order statistics from outliers models. Recent years have seen the appearance of a number of studies on both single- and multiple-outlier models and more generally on order statistics from independent and nonidentically distributed random variables. The theory of order statistics from independent but nonidentically distributed random variables involves permanents which is similar to that of the determinants but without the alternating sign. Barnett and Lewis (1994) mainly discuss the single-outlier models. In an excellent review article, Balakrishnan (2007) describes more general model of order statistics from independent and nonidentically distributed random variables, including many important issues such as distributional properties, characterizations, estimation, outliers, robustness. Continuing from the International Conference on Order Statistics and Extreme Values, Theory and Applications 18–20 December 2000 in Mysore, India) organized by N. R. Mohan and H. N. Nagaraja, a series of international conferences devoted to order statistics in Warsaw, Poland (2002–04); Izmir, Turkey (2005); Mashad, Iran (2006); Amman, Jordan (2007) and Aachen, Germany (2008) provided international forums for presentation and discussion of topics related to ordered statistical data. In these conferences, both reviews of previously existing results and new results involving order statistics were presented in the context of topics such as approximations, characterizations, distribution theory and probability models, stochastic ordering, inequalities, censoring, statistical inference, applications of ordered data, information and entropies, nonparametric methods, ranked set sampling,

and asymptotic theory. The special issue on ordered statistical data, approximations bounds and characterizations of Taylor and Francis' journal *Communications in Statistics-Theory and Methods*, vol. 36, no. 7 edited by I. Bairamov consists of selected articles presented at the international conference OSD-2005, Izmir, Turkey.

The elegant theory of order statistics and general models of ordered statistical data is likely to arouse the interest of many scientists working in the area of statistical theory and applications.

See also: Analysis of Extreme Values in Education; Hypothesis Testing and Confidence Intervals; Markov Chain Monte Carlo; Nonparametric Statistical Methods; Stochastic Processes; Survival Data Analysis.

Bibliography

- Andrews, D. F., Bickel, P. J., and Hampel, F. R. *et al.* (1972). *Robust Estimates of Location: Survey and Advances*. Princeton, NJ: Princeton University Press.
- Arnold, B., Balakrishnan, N., and Nagaraja, H. N. (1992). *A First Course in Order Statistics*. New York: Wiley.
- Bairamov, I. and Petunin, Y. I. (1990). Structure of invariant confidence intervals containing the main distributed mass. (*English translation in Probability and Its Applications*, 1991, **35**(1), 15–26.) *Teoriya Veroyatnostey i ee Primeneniya* **35**(1), 15–26.
- Balakrishnan, N. (2007). Permanents, order statistics, outliers and robustness. *Revista Matematica Complutense* **20**(1), 7–107.
- Balch, W. (1992). Effect of class standing on students' predictions of their final exam scores. *Teaching of Psychology* **19**(3), 136–141.
- Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*, 3rd edn. Chichester: John Wiley & Sons Ltd.
- Cole, N. S. (1973). Bias in selection. *Journal of Educational Measurement* **10**, 237–255.
- David, H. A. (1981). *Order Statistics*, 2nd edn. New York: Wiley.
- David, H. A. and Ghosh, J. K. (1985). The effect of an outlier on L-estimates of location and symmetric distributions. *Biometrika* **72**(1), 216–218.
- David, H. A. and Nagaraja, H. N. (2003). *Order Statistics*, 3rd edn. New York: Wiley.
- Huber, P. J. (1981). *Robust Statistics*. New York: Wiley.
- Myles, T. and Henderson, R. C. (2002). Medical licensure examination scores: Relationship to obstetrics and gynecology examination scores. *Obstetrics & Gynecology* Part 1, **100**(5), 955–958.
- Olkin, I. and Stephens, M. A. (1993). On making the shortlist for the selection of candidates. *International Statistical Review* **61**(3), 477–486.
- Sarhan, A. E. and Greenberg, B. G. (1962). *Contributions to Order Statistics*. New York: Wiley.
- Tadikamalla, P. R. and Balakrishnan, N. (1998). Computer simulation of order statistics. In Balakrishnan, N. (ed.) *Order Statistics: Theory and Methods*, pp 65–72. Amsterdam: North Holland Elsevier.
- Tomdike, R. L. (1971). Concept of culture fairness. *Journal of Education Measurement* **8**, 63–70.
- Williams, L., Wiebe, E., Yang, K., Ferzli, M., and Miller, C. (2002). In support of pair programming in the introductory computer science course. *Computer Science Education* **12**(3), 197–212.
- Ahsanullah, M. and Nevzorov, V. B. (2001). *Ordered Random Variables*. Huntington, NY: Nova Science Publishers.
- Ahsanullah, M. and Nevzorov, V. B. (2005). *Order Statistics. Examples and Exercises*. Huntington, NY: Nova Science Publishers.
- Arnold, B. C. and Balakrishnan, N. (1989). *Relations, Bounds and Approximations for Order Statistics. Lecture Notes in Statistics* **53**. New York: Springer.
- Arnold, B. C., Becker, A., Gather, U., and Zahedi, H. (1984). On the Markov property of order statistics. *Journal of Statistical Planning and Inference* **9**, 147–154.
- Bairamov, I. (2007). Advances in exceedance statistics based on ordered random variables. In Ahsanullah, M. and Raqab, M. Z. (eds.) *Recent Developments in Ordered Random Variables*, pp 97–119. New York: Nova Science.
- Balakrishnan, N. (1986). Order statistics from discrete distributions. *Communications in Statistics-Theory and Methods* **15**, 657–675.
- Balakrishnan, N. and Cohen, A. C. (1991). *Order Statistics and Inference. Estimation Methods*. Academic Press.
- Barlow, R. E. and Proschan, F. (1981). *Statistical Theory of Reliability and Life Testing. To Begin With*. Silver Springs.
- Barnett, V. (1976). The ordering of multivariate data (with discussion). *Journal of the Royal Statistical Society A* **139**, 318–354.
- Davis, C. S. and Stephens, M. A. (1983). Approximate percentage points using Pearson curves. *Algorithm AS192. Applied Statistics* **32**, 322–327.
- Eryilmaz, S. (2005). Concomitants in a sequence of independent nonidentically distributed random vectors. *Communication in Statistics-Theory and Methods* **34**, 1925–1933.
- Nagaraja, N. (1992). Order statistics from discrete distributions (with discussion). *Statistics* **23**, 189–216.

Further Reading

Relevant Websites

<http://en.wikipedia.org> – Order statistic.
<http://planetmath.org> – PlanetMath. org.
<http://mathworld.wolfram.com> – Wolfram Mathworld, order statistic.

Point Estimation Methods with Applications to Item Response Theory Models

F Bartolucci and L Scrucca, Università degli Studi di Perugia, Perugia, Italy

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bayesian inference – A method of statistical inference which exploits both the information arising from the data, summarized by the likelihood function, and the experimenter's belief before observing the data, represented by the prior distribution. The name Bayesian comes from the use of the Bayes' theorem, from the work of Reverend Thomas Bayes, in the inferential process.

Estimator/estimate – An estimator is any sample statistic which is used to estimate an unknown population parameter; the sample mean is an example of estimator of the population mean. An estimate is the value of the estimator that is obtained for an observed sample.

Item response theory (IRT) – A body of theory concerning models for the analysis of data collected by the administration of test items aimed at measuring a certain latent trait of interest, such as the ability of a student in a certain field. IRT models are based on the assumption that the probability that a subject responds correctly to an item is a function of the latent trait.

Likelihood function – The likelihood function is the probability or the density of the observed sample expressed as a function of the parameters of interest. It plays a central role in statistical inference and, in particular, for the maximum likelihood estimation method.

Parameter – A parameter is an index used to represent a certain population characteristic. The mean of a variable, such as the income of every resident in a certain region, is an example of parameter. Typically, the parameter is unknown and, then, it has to be estimated. In the classical approach to inference, the parameter is a fixed unknown constant. On the contrary, in Bayesian inference, the parameter is a random variable with a probability distribution representing the uncertainty on its knowledge.

Point estimation – A statistical inference method consisting of assigning a single value, called point estimate, to each unknown parameter.

Random sample – A subset of the population units which is selected by a random mechanism. It is the basis of any method of statistical inference.

Statistic – Any mathematical function used to summarize the information in the sample data.

Statistical inference – The process of using the information in a sample to draw conclusions about the population from which the sample is drawn.

Introduction

The general goal of statistical inference is to provide some conclusions about the distribution of a variable of interest (X) in a certain population on the basis of a random sample, that is, a subset of the population units randomly chosen. Often, the interest is on a particular statistical index describing the distribution of X , which is called parameter (θ); examples are the average income of the population living in a certain region or the variance of the number of employees in the population (in the sense of set) of firms operating in a certain field. In a parametric setting, we assume that the distribution of X is described by a certain statistical model, that is, a family of distributions indexed by the parameter θ , $\{f(x; \theta), \theta \in \Theta\}$, where $f(x; \theta)$ denotes a probability mass function (pmf) or a probability density function (pdf) and Θ is the parameter space. Note that θ may also be a vector of parameters of suitable dimension; however, we will not use a special notation for this case. Provided that the statistical model holds, knowledge of the true value of θ is equivalent to the full knowledge of the distribution of interest.

In this setting, a random sample of size n may be defined as the random vector $X = (X_1, \dots, X_n)$, whose components are mutually independent and identically distributed with marginal distribution $f(x; \theta_0)$, where θ_0 denotes the true value of the parameter of interest. Note that, in the presence of a finite population, sample units have to be drawn with replacement in order to have independence between the random variables X_i , $i = 1, \dots, n$. Another important concept for what follows is that of sample space, which is the set of all the possible samples of size n . This set may be discrete or continuous

according to the nature of the support of X . An observed sample is denoted by $x = (x_1, \dots, x_n)$. Because of the independence of the sample units, its pmf or pdf is simply given by $f(x) = \prod_{i=1}^n f(x_i)$. The information on θ contained in the sample is typically summarized by a sample statistic, which is defined as a mathematical function of the random sample, that is, $T = t(X)$. Examples of sample statistics are the sample mean, $\bar{X} = \sum_{i=1}^n X_i/n$, and the sample variance, $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$. Since X is a random vector, any sample statistic is a random variable with a distribution which is referred to as the sampling distribution.

Different inferential methods are available in the statistical literature. For a unitary description of these methods and related approaches, see Barnett (1999) and Casella and Berger (2002). In the following, we deal, in particular, with point estimation, which consists of assigning to θ a single value out from the parameter space.

Point Estimation Methods

An estimator of the unknown parameter θ may be defined as any sample statistic $T = t(X)$ used to estimate θ . An estimate $t = t(x)$ is the value of the estimator for an observed sample. The statistical theory on point estimation mainly focuses on methods of evaluating estimators and methods of finding estimators.

Evaluation of an Estimator

Methods of evaluating an estimator are necessary because, in principle, different estimators can be used for the same parameter.

The most well-known criterion to compare different estimators is based on the mean squared error (MSE) which, for an estimator T of θ , is defined as the expected value of its (squared) error, that is, $\text{MSE}_\theta(T) = E_\theta[(T - \theta)^2]$. According to this criterion, the estimator T is better than another estimator T^* if the MSE of the first is uniformly smaller than that of the second, that is, $\text{MSE}_\theta(T) \leq \text{MSE}_\theta(T^*)$ for all $\theta \in \Theta$, with at least one value of θ for which the inequality strictly holds. In this case, we say that T is more efficient than T^* . Note that other measures of error, such as $E_\theta(|T - \theta|)$, are possible to evaluate the performance of an estimator. However, MSE is usually preferred because it is easier to treat analytically. Moreover, it may be decomposed as follows:

$$\text{MSE}_\theta(T) = \text{Var}_\theta(T) + \text{B}_\theta(T)^2$$

where $\text{Var}_\theta(T) = E_\theta\{[T - E_\theta(T)]^2\}$ is the variance of T and $\text{B}_\theta(T) = E_\theta(T) - \theta$ is its bias. The first is a measure of dispersion of the distribution of the estimator around its

mean and the second is a measure of its systematic error. Both quantities need to be small in magnitude in order to have a small MSE.

It has to be clear that comparing two estimators on the basis of the MSE is not always possible since it may happen that, between T and T^* , the first is better for certain values of θ , whereas the second is better for other values of θ . Then, in real problems, we cannot expect to find the most efficient estimator, that is, the estimator which has the minimum MSE, for all θ in Θ , among all the possible estimators of the same parameter.

The indeterminacy problem described above is usually dealt with by restricting the class of possible estimators to that of the unbiased estimators. An estimator T of θ is said to be unbiased if its expected value is always equal to the parameter value or, equivalently, it always has null bias; in symbols,

$$E_\theta(T) = \theta \text{ or } \text{B}_\theta(T) = 0, \forall \theta \in \Theta$$

Examples of unbiased estimators are \bar{X} , when the parameter of interest is the population mean (μ), and S^2 , when the parameter of interest is the population variance (σ^2).

For any unbiased estimator, the MSE is equal to its variance and then, within the class of the unbiased estimators of the same parameter, that with minimum MSE corresponds to the uniformly minimum variance unbiased estimator (UMVUE), that is, the unbiased estimator T , such that

$$\text{Var}_\theta(T) \leq \text{Var}_\theta(T^*), \forall \theta \in \Theta$$

for any other unbiased estimator T^* of θ . Different methods are available in the statistical literature to check when an estimator of a certain parameter is the UMVUE. Fundamental results on this topic are represented by the Cramer–Rao inequality, the Rao–Blackwell theorem, and the Lehman–Scheffé theorem. For a detailed illustration of these results and the related theory, see Lehmann and Casella (1998).

Statistical theory on point estimation is also focused on the asymptotic properties of an estimator, which concern its performance when the sample size grows to infinity. The most important of these properties is known as consistency. In particular, an estimator T of θ is consistent if

$$\lim_{n \rightarrow \infty} P_\theta(|T_n - \theta| < \epsilon) = 1, \quad \forall \epsilon > 0$$

where the subscript n has been added to T in order to recall that the distribution of the estimator depends on the sample size. Intuitively, this property says that as the sample size grows and, then, the available amount of information increases, we expect the estimator to attain values closer and closer to the true parameter value. Related properties are those of asymptotic unbiasedness and consistency in MSE. We refer to Cox and Hinkley (1974) and Lehmann (1999) for a detailed description.

Methods of Finding Estimators

Often, intuition can lead us to find an estimator for the parameter of interest. However, more methodical ways of finding estimators are needed. In the following, we discuss three general methods: method of moments, maximum likelihood method, and, under a different inferential approach, Bayesian estimation. The first two methods are formulated following a frequentist approach, which only exploits the information contained in the sample. This is in contrast with the Bayesian approach, which also exploits *a priori* information, that is, the knowledge or subjective opinions about the parameter of interest before any data have been observed.

Method of Moments

One of the oldest and simplest methods of finding estimators is the method of moments, which dates back to K. Pearson in the late 1800s. Suppose that the random sample is drawn from a statistical model depending on a parameter vector $\theta = (\theta_1, \dots, \theta_k)$ of dimension k . This method consists of equating the first k population moments to the corresponding sample moments and solving the resulting system of equations with respect to θ . The population and the sample moments of order r are defined, respectively, as

$$\mu_r(\theta) = \int_{-\infty}^{+\infty} x^r f(x; \theta) dx \text{ and } M_r = \frac{1}{n} \sum_{i=1}^n X_i^r$$

Thus, the system of equations to be solved with respect to θ is

$$\begin{aligned} \mu_1(\theta) &= M_1 \\ &\vdots \\ \mu_k(\theta) &= M_k \end{aligned}$$

In general, the estimators obtained by the method of moments are consistent, asymptotically unbiased, and have asymptotic normal distribution. However, their efficiency can usually be improved upon. For a detailed description of this method see Bowman and Shenton (1985).

Maximum Likelihood Method

The maximum likelihood method is the most popular technique for deriving estimators. It is based on the likelihood function which, for an observed sample \mathbf{x} , is defined as the probability (or density) of \mathbf{x} expressed as a function of θ ; in symbols

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta)$$

This function provides a measure of plausibility of each possible value of θ on the basis of the observed data. Then,

the method at issue consists of estimating θ through the value of θ which maximizes $L(\theta)$ since this corresponds to the parameter value for which the observed sample is most likely. The estimate found in this way, that is,

$$\hat{\theta} = \hat{\theta}(\mathbf{x}) \text{ such that } L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta)$$

is the maximum likelihood estimate (mle) of θ . When expressed as a function of the random sample \mathbf{X} , we have the maximum likelihood estimator (MLE) $\hat{\theta}(\mathbf{X})$. Obviously, this method of finding estimators is in agreement with the likelihood principle, which says that the inferential conclusions on θ based on two different sampling schemes must be the same when these schemes give rise to proportional likelihood functions; see Casella and Berger (2002) for a formal definition of this principle.

In order to find the mle of θ , we have to solve an optimization problem. It is usually simpler to maximize the log-likelihood

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(x_i; \theta)$$

instead of $L(\theta)$. Since the logarithmic function is monotonic increasing, the two maximization problems are equivalent. In the single-parameter case ($k = 1$), the usual procedure to maximize $\ell(\theta)$ is based on solving the likelihood equation

$$\frac{\partial \ell(\theta)}{\partial \theta} = \sum_{i=1}^n \frac{f'(x_i; \theta)}{f(x_i; \theta)} = 0$$

In order to be sure that the found root is the global maximum of $\ell(\theta)$, we also have to verify that

$$\frac{\partial^2 \ell(\theta)}{\partial \theta^2} = \sum_{i=1}^n \frac{f''(x_i; \theta)f(x_i; \theta) - f'(x_i; \theta)f'(x_i; \theta)}{f(x_i; \theta)^2} < 0$$

and evaluate $\ell(\theta)$ at the boundary of the parameter space.

In the multiparametric case, in which $\theta = (\theta_1, \dots, \theta_k)$ with $k > 1$, the problem is usually more complex as it is based on the solving the system of linear equations

$$\frac{\partial \ell(\theta)}{\partial \theta_i} = 0, \quad i = 1, \dots, k$$

An explicit solution for this system is seldom available and, therefore, we need to use iterative algorithms to maximize the log-likelihood. Starting from an initial guess, these algorithms update θ in an appropriate way until convergence, that is, until a stationary point of $\ell(\theta)$ is found. The most popular of these algorithms is known as the Newton–Raphson algorithm. At the $(t + 1)$ -th iteration, it updates θ as

$$\theta^{(t+1)} = \theta^{(t)} + \mathcal{J}(\theta^{(t)})^{-1} s(\theta^{(t)})$$

where $\theta^{(t)}$ is the current value of the parameter, $s(\theta)$ is the score vector, that is, the first derivative vector of $\ell(\theta)$,

whereas $\mathcal{J}(\theta)$ is the observed information matrix, that is, the second derivative matrix of $\ell(\theta)$ with negative sign. A variant of the Newton–Raphson algorithm is the Fisher-scoring algorithm which, at each iteration, updates θ as

$$\theta^{(t+1)} = \theta^{(t)} + I(\theta^{(t)})^{-1} s(\theta^{(t)})$$

where $I(\theta)$ is the Fisher information, that is, the expected value of the observed information matrix. Using the observed or the Fisher information matrix is equivalent if $f(x; \theta)$ belongs to the regular exponential family. Through these algorithms it is usually possible to find a local maximum of the log-likelihood; however, in general, it is not guaranteed that this also corresponds to the global maximum. When there are more than one local maxima, a crucial point is that of the choice of the starting values for the algorithm; often, the method of moments is used to obtain reasonable starting values.

One of the main reasons of the great popularity of the maximum likelihood method is that the resulting estimator has many interesting properties. These properties hold under certain regularity conditions which are not very restrictive, albeit quite technical. In the following, we provide a brief summary of these properties; for a detailed description we refer to Cox and Hinkley (1974), Azzalini (1996), Lehmann and Casella (1998), Lehmann (1999), and Severini (2000).

One of the most important properties is the invariance property, according to which if $\hat{\theta}$ is the MLE of θ , then $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$. Note that this property also holds when $\tau(\theta)$ is not a one-to-one function of θ . Moreover, the maximum likelihood method is in agreement with the so-called sufficiency principle. This is because the estimator depends on the sample only through a sufficient statistic for θ , which represents the only relevant information on the parameter; thus, if θ is the only parameter of interest, further information contained in the sample can be discarded. Another property is that if an UMVUE exists and suitable regularity conditions hold, then the MLE coincides with this estimator.

The previous properties are finite-sample properties. Estimators formulated with the maximum likelihood method also have interesting asymptotic properties. First, $\hat{\theta}_n$ is a consistent estimator of θ . The consistency of the MLE also holds for the estimator $\tau(\hat{\theta})$ of $\tau(\theta)$. Moreover, the MLE is an asymptotically efficient estimator, in the sense that its variance tends to the lower bound of the Cramér–Rao inequality, and, for large samples, the standard deviation of the MLE may be approximated by the square root of the diagonal elements of the inverse of the information matrix. Finally, a very important property is that the estimator has asymptotic normal distribution. This allows us to easily construct confidence intervals and to test hypotheses on the parameter of interest.

Bayesian Estimators

In the frequentist approach, the parameter θ is considered an unknown, but fixed, quantity and only the information coming from the sampling data is relevant for inference. Thus, we do not take into account the prior belief on the parameter. On the contrary, in the Bayesian approach, the parameter θ is considered as a random variable with a certain probability distribution, referred to as the prior distribution, whose role is that of representing the experimenter's belief before observing the data. Within this approach, the prior distribution is updated on the basis of the likelihood function through the Bayes' theorem. The resulting distribution is referred to as the posterior distribution and summarizes the information in both the prior distribution and in the data.

Let $\pi(\theta)$ denote the prior distribution and let $f(x|\theta)$ denote the conditional distribution of the sample given θ . Note that $f(x|\theta)$ is equivalent to the likelihood function discussed earlier. According to the Bayes' theorem, we can derive the posterior distribution of θ on the basis of the observed sample x as

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)}$$

where $m(x)$ is the marginal pdf (or pmf) of the data, which is given by

$$m(x) = \int_{-\infty}^{+\infty} f(x|\theta)\pi(\theta)d\theta$$

The mean, which is often used to summarize the posterior distribution of the parameter, is a Bayesian point estimator of θ . Sometimes, the mode or the median of the posterior distribution is used instead.

A great variety of textbooks on Bayesian inference now exist. For an accessible introduction we suggest Ghosh *et al.* (2006), whereas for more advanced treatment, we suggest Gelman *et al.* (2003) and Robert and Casella (2005) for related computational methods.

Item Response Theory Models

Item response theory (IRT) models are tailored to the analysis of data arising from the administration of a questionnaire made of a series of items which measure a common latent trait, that is, a characteristic of the human being which is not directly observable; for a review see the work by Hambleton and Swaminathan (1996). The main application of these models is in educational assessment, where the latent trait corresponds to a certain type of ability of an examinee.

In the following, after a brief summary of the main assumptions of IRT models for dichotomously scored items, we focus on maximum likelihood methods for

their estimation. For an updated and detailed description of these methods, see Baker and Kim (2004).

Suppose that a questionnaire made of r items is administered to a sample of n subjects, and let y_{ij} , $i = 1, \dots, n$, $j = 1, \dots, r$, be a categorical response variable for the j th item administered to the i th subject. Depending on the context, y_{ij} will denote a random variable or one of its realizations. In educational assessment, the response variable is often binary; in particular, y_{ij} is equal to 1 if subject i responds correctly to item j and to 0 otherwise.

With reference to the binary case, the main assumptions of IRT models may be summarized as follows:

- *Unidimensionality*. For each subject i , the responses to the r items depend on the same latent parameter ϕ_i , which is unidimensional.
- *Local independence*. For each subject i , the responses to the r items are independent given ϕ_i .
- *Monotonicity*. The probability of a correct response of the i th subject to the j th item, in symbols $p_j(\phi_i) = p(y_{ij} = 1 | \phi_i)$, is a monotonic increasing function of ϕ_i .

The probabilities $p_j(\phi_i)$ are modeled via a function which is usually referred to as the item characteristic curve (ICC) or item response function. Different models arise according to the assumed parametrization of the ICC. The most well known is the one parameter logistic (1PL) parametrization, according to which

$$p_j(\phi_i; \beta_j) = \frac{e^{\phi_i - \beta_j}}{1 + e^{\phi_i - \beta_j}}$$

where β_j is the difficulty level of item j . The resulting model is usually referred to as the Rasch model (Rasch, 1960). More general parametrizations were proposed by Birnbaum (1968) and rely on further parameters to describe the item characteristics.

Maximum Likelihood Estimation

In the following, we describe three different methods based on the maximum likelihood paradigm for the estimation of the above IRT models for dichotomously scored items.

Joint maximum likelihood method

The joint maximum likelihood (JML) method consists of maximizing the likelihood of the model, corresponding to the probability of the observed data matrix, with respect to the ability and item parameters jointly.

Under the assumption of local independence, the joint distribution of the response vector $\mathbf{y}_i = (y_{i1}, \dots, y_{ir})$ may be expressed as

$$p(\mathbf{y}_i | \phi_i, \psi) = \prod_j p_j(\phi_i; \psi_j)^{y_{ij}} [1 - p_j(\phi_i; \psi_j)]^{1-y_{ij}}$$

where ψ is the vector of the item parameters, which depend on the IRT model of interest. Under the 1PL model, it only contains the parameters β_j .

Assuming that the response vectors for the subjects in the sample are independent each other, the conditional probability of observing the response matrix Y with elements y_{ij} , $i = 1, \dots, n$, $j = 1, \dots, r$, is equal to

$$p(Y | \phi, \psi) = \prod_i \prod_j p_j(\phi_i; \psi_j)^{y_{ij}} [1 - p_j(\phi_i; \psi_j)]^{1-y_{ij}}$$

where ϕ is the vector of the ability parameters. Then, for an observed matrix of responses Y , the JML method consists of estimating the model parameters by maximizing the joint likelihood

$$L_{\mathcal{J}}(\phi, \psi) = p(Y | \phi, \psi)$$

with respect to ϕ and ψ .

For the Rasch model, the joint likelihood becomes

$$L_{\mathcal{J}}(\phi, \psi) = \frac{e^{\sum_i t_i \phi_i - \sum_j s_j \beta_j}}{\prod_i \prod_j (1 + e^{\phi_i - \beta_j})}$$

where $t_i = \sum_j y_{ij}$ is the number of correct responses provided by subject i and $s_j = \sum_i y_{ij}$ is the number of subjects who responded correctly to item j . The corresponding log-likelihood is equal to

$$\ell_{\mathcal{J}}(\phi, \psi) = \sum_i t_i \phi_i - \sum_j s_j \beta_j - \sum_i \sum_j \log(1 + e^{\phi_i - \beta_j})$$

For each IRT model, the joint likelihood is invariant with respect to certain transformations of the item and ability parameters. Then, suitable constraints have to be put on these parameters in order to make the model identifiable. The following constraints may alternatively be used for the Rasch model:

- $\beta_1 = 0$: in this way, the first item is taken as a reference item and, then, β_j , $j = 2, \dots, r$, has to be interpreted as the difficulty of item j with respect to the first one.
- $\sum_i \phi_i = 0$: in this way, the average ability of the subjects is fixed at 0 and ϕ_i , $i = 1, \dots, n$, is interpreted as the ability of subject i with respect to the average of the group.

These constraints are equivalent in the sense that the maximum value of the likelihood that may be reached under each of them is the same, which is also equal to the unconstrained maximum of the likelihood. Further constraints need to be used for more sophisticated models.

The algorithm to maximize $\ell_{\mathcal{J}}(\phi, \psi)$ that is commonly used is based on a series of Newton–Raphson steps to be performed until convergence. Explicit expressions for the derivatives used in these steps are available; see Hambleton and Swaminathan (1996) and Baker and Kim (2004). We note that the point at convergence does not depend on the starting values chosen for the parameters since, provided that the JML estimate exists, the log-likelihood is a strictly concave function over the full parameter space.

It has to be recalled that the JML estimate is not ensured to exist. A set of conditions on the matrix Y which ensures this existence was derived by Fischer (1981). A necessary condition is that there exist neither subjects who respond correctly or incorrectly to all items nor items to which all subjects respond correctly or incorrectly. These subjects and items must be removed from the dataset when they exist. Moreover, the JML estimator is not consistent as n grows to infinity with r fixed. This does not mean that the estimates obtained by using this method are unreliable. Indeed, if the number of subjects and that of items are large enough, the bias of the JML estimator is expected to be low.

Conditional maximum likelihood

The conditional maximum likelihood (CML) method may be only applied to the Rasch model and it is typically used to estimate its difficulty parameters. The method is based on the maximization of the conditional likelihood of these parameters given a set of minimal sufficient statistics for the ability parameters.

Let $\beta = (\beta_1, \dots, \beta_r)$ denote the vector of item parameters under the Rasch model. First of all, consider that the conditional probability that subject i attains score t_i , given the ability parameter ϕ_i , is equal to

$$p(t_i | \phi_i, \beta) = \sum_{\mathbf{y} \in \mathcal{Y}(t_i)} p(\mathbf{y}_i = \mathbf{y} | \phi_i) = \frac{e^{\phi_i t_i}}{\prod_j (1 + e^{\phi_i - \beta_j})} q_i$$

where $q_i = \sum_{\mathbf{y} \in \mathcal{Y}(t_i)} \exp\{-\sum_j y_j \beta_j\}$, with $\mathcal{Y}(t)$ denoting the set of all the binary vectors $\mathbf{y} = (y_1, \dots, y_r)$ with elements having sum t . These vectors have t elements equal to 1 and $r - t$ elements equal to 0. Consequently, for each subject i , the conditional probability of the response configuration \mathbf{y}_i given the sufficient statistics t_i is equal to

$$p(\mathbf{y}_i | t_i, \phi_i, \beta) = \frac{p(\mathbf{y}_i | \phi_i, \beta)}{p(t_i | \phi_i, \beta)} = \frac{e^{-\sum_j y_{ij} \beta_j}}{q_i} \quad [1]$$

which does not depend on ϕ_i and, then, may be denoted by $p(\mathbf{y}_i | t_i, \beta)$.

The conditional likelihood exploited within the CML method is

$$L_C(\beta) = \prod_i p(\mathbf{y}_i | t_i, \beta)$$

Note that the subjects who respond correctly or incorrectly to all the items do not contribute to this likelihood. Then, the corresponding log-likelihood may be expressed as

$$\ell_C(\beta) = -\sum_j s_j^* \beta_j - \sum_i d_i \log q_i \quad [2]$$

where $d_i = 1\{0 < t_i < r\}$ is a dummy variable equal to 1 if subject i contributes to the likelihood and to 0 otherwise and $s_j^* = \sum_i d_i y_{ij}$. An identifiability problem also arises in this case; for this reason, we must use the constraint $\beta_1 = 0$ or, alternatively, $\sum_j \beta_j = 0$.

In order to maximize $\ell_C(\beta)$, we may use a Newton–Raphson algorithm. Efficient methods to compute the conditional log-likelihood and its derivatives, in a way which is viable even when r is large, are available in the statistical and psychometric literatures (Formann, 1986; Gustafson, 1980; Liou, 1994).

Certain conditions need to be fulfilled for the existence of the CML estimate (Fischer, 1981). In most situations, a condition which ensures this existence is that there do not exist items to which no subjects or all subjects respond correctly. Finally, we have to stress that the main advantage of the CML method is that the resulting estimator is consistent for r fixed as n grows to infinity. For a detailed description of the asymptotic properties of this estimator, see Andersen (1970); see also Andersen (1972).

Marginal maximum likelihood

The marginal maximum likelihood (MML) method is tailored to IRT models formulated under the assumption that the ability is a random parameter. The method consists of maximizing the likelihood corresponding to the manifest probability of the observed responses, that is, the marginal probabilities of these responses once the ability parameters have been integrated out.

Let $f(\phi_i)$ denote the distribution for the ability parameters. The manifest distribution of \mathbf{y}_i is given by

$$p(\mathbf{y}_i | \psi) = \int_{\mathbb{R}} p(\mathbf{y}_i | \phi_i, \psi) f(\phi_i) d\phi_i$$

Moreover, the manifest probability of Y is $p(Y | \psi) = \prod_i p(\mathbf{y}_i | \psi)$ and, for an observed matrix of responses, it corresponds to the marginal likelihood on which the MML method is based. This is denoted by $L_M(\psi, \eta)$, where η is the vector of parameters on which $f(\phi_i)$ possibly depends. The corresponding log-likelihood is denoted by $\ell_M(\psi, \eta)$. The latter is maximized by a version of the EM algorithm of Dempster *et al.* (1977) which is based on the complete likelihood, that is, the likelihood that we could compute if we knew the ability level of each subject. The MML estimator of the item parameters is denoted by $\hat{\psi}_M$ and that of the parameters for the distribution of the ability is denoted by $\hat{\eta}_M$.

A crucial assumption for the implementation of the MML method concerns the distribution of the ability. The most common assumption is that the ability has a normal distribution with unknown mean and variance, which are estimated together with the item parameters. More advanced solutions consist of assuming that the ability has a discrete distribution with a suitable number of support points and weights. The first paper discussing this kind of approach is that of Bock and Lieberman (1970). Further developments are due to Bock and Aitkin (1981), Thissen (1982), Lindsay *et al.* (1991), Pfanzagl (1993), and Schilling and Bock (2005).

Example

In order to provide an illustration of the maximum likelihood estimation methods outlined above, we describe an application based on a dataset which was provided by the Educational Testing Service. The dataset concerns the responses of 1510 students to 12 publicly released items on mathematics collected in 1996 within a project called National Assessment of Educational Progress (NAEP); for details, see Bartolucci and Forcina (2005).

By using the JML method, for the Rasch model we obtained the estimates of the item and ability parameters which are displayed in **Table 1**. Note that we adopted the identifiability constraint $\beta_1 = 0$ and, consequently, we took the first item as a reference item.

The estimates of the item parameters allow us to evaluate the difficulty level of each item. In particular, we can conclude that the last item is the most difficult, whereas the fourth item is the easiest. Moreover, the standard errors and the confidence intervals may be used to assess whether the difficulty level of each item is significantly different from that of the first item. We observe that the second item is not significantly different from the first in terms of difficulty since the corresponding confidence interval includes zero. The same cannot be said about, for example, the third item. Finally, the estimates of the ability parameters can be used to assess the students to whom the questionnaire was administered. We conclude, for instance, that the second and third students have the same ability level as they provided the same number of correct responses, and this ability level is higher than that of the first student.

As an illustration of the CML method, we show in **Table 2** the estimates of the item parameters obtained with this method.

Table 1 JML estimates of the item and ability parameters of the Rasch model for the NAEP dataset

	Estimate	SE	95% Conf.int.	
β_1	0.000	—	—	—
β_2	-0.051	0.097	-0.241	0.138
β_3	0.755	0.093	0.574	0.936
β_4	-1.140	0.111	-1.357	-0.923
β_5	1.672	0.092	1.491	1.853
β_6	0.014	0.096	-0.175	0.202
β_7	0.724	0.093	0.542	0.905
β_8	1.305	0.092	1.125	1.485
β_9	0.365	0.094	0.181	0.549
β_{10}	0.574	0.093	0.391	0.756
β_{11}	2.697	0.098	2.505	2.888
β_{12}	2.751	0.098	2.558	2.944
ϕ_1	-0.080	0.674	-1.400	1.241
ϕ_2	1.193	0.662	-0.104	2.491
ϕ_3	1.193	0.662	-0.104	2.491
ϕ_4	0.770	0.649	-0.501	2.041
ϕ_5	-0.080	0.674	-1.400	1.241
\vdots	\vdots	\vdots	\vdots	\vdots
ϕ_{1510}	2.158	0.750	0.689	3.626

The results are similar to those previously obtained, except that only the estimates of the item parameters are now available. This is because the ability parameters are removed by conditioning on the corresponding sufficient statistics. By comparing the results in **Table 2** with those in **Table 1**, we note that the CML estimates are very similar to the JML estimates and, therefore, the items are equivalently ranked. This is because, with 12 items, we can also obtain reliable estimates with the JML method. To this regard we recall that, in contrast to the JML estimator, the CML estimator is in general consistent.

Finally, we applied the MML method based on a discrete latent distribution for the ability. In this case, we have to choose the number of support points of the latent distribution, which is equivalent to the number of latent classes in the population. Following a standard practice in this field, we make this choice by using the Bayesian Information Criterion of Schwarz (1978) which is based on the minimization of the index:

$$\text{BIC} = -2\ell_M(\hat{\psi}, \hat{\eta}) + \log(n)\#\text{parameters}$$

In practice, we fit the model with an increasing number of classes until this index does not start to decrease. Then, we choose the number of classes of the model with the smallest BIC. For the NAEP data, this procedure leads to the results shown in **Table 3**.

We clearly choose three latent classes. A probability and an ability level are associated to each of these classes. The estimates of these parameters are shown in **Table 4**.

Table 2 CML estimates of the item parameters of the Rasch model for the NAEP dataset

	Estimate	SE	95% Conf.int.	
β_1	0.000	—	—	—
β_2	-0.047	0.092	-0.229	0.134
β_3	0.691	0.088	0.517	0.864
β_4	-1.040	0.106	-1.247	-0.833
β_5	1.521	0.088	1.349	1.693
β_6	0.013	0.092	-0.168	0.193
β_7	0.662	0.089	0.489	0.836
β_8	1.191	0.088	1.019	1.363
β_9	0.334	0.090	0.158	0.511
β_{10}	0.525	0.089	0.351	0.700
β_{11}	2.427	0.092	2.246	2.607
β_{12}	2.474	0.093	2.292	2.655

Table 3 Selection of the number of classes for the latent class Rasch model applied to the NAEP dataset

#Classes	$\ell_M(\hat{\psi}, \hat{\eta})$	#Parameters	BIC
1	-11009	12	22106
2	-10242	14	20586
3	-10166	16	20450
4	-10163	18	20458

Table 4 MML estimates of the ability parameters of the Rasch model for the NAEP dataset (three latent classes)

Class	Ability	Probability
1	−0.645	0.165
2	0.970	0.457
3	2.432	0.378

We can observe that the classes are well separated in terms of ability level. The first class, which includes subjects with the lowest ability level, is also the smallest in the population (it has the smallest probability), whereas the other two classes have a comparable size. The corresponding estimates of the item parameters are, in practice, equal to those obtained with the CML method which are displayed in **Table 2**. This is in agreement with the theory of Lindsay *et al.* (1991) about the equivalence between CML and MML estimation methods for the Rasch model which holds under certain regularity conditions.

See also: Ability Testing; Generalized Linear Mixed Models; Item Response Theory; Latent Class Models; Model Selection; Rasch Models.

Bibliography

- Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of Royal Statistical Society, B* **32**, 283–301.
- Andersen, E. B. (1972). The numerical solution of a set of conditional estimation equations. *Journal of the Royal Statistical Society, B* **34**, 42–54.
- Azzalini, A. (1996). *Statistical Inference Based on the Likelihood*. London: Chapman and Hall.
- Baker, F. B. and Kim, S.-H. (2004). *Item Response Theory: Parameter Estimation Techniques*, 2nd edn. New York: Marcel Dekker.
- Barnett, V. (1999). *Comparative Statistical Inference*. New York: Wiley.
- Bartolucci, F. and Forcina, A. (2005). Likelihood inference on the underlying structure of IRT models. *Psychometrika* **70**, 31–43.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord, F. M. and Novick, M. R. (eds.) *Statistical Theories of Mental Test Scores*, pp 397–479. Reading, MA: Addison-Wesley.
- Bock, R. D. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **35**, 179–197.
- Bock, R. D. and Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika* **35**, 179–197.
- Bowman, K. O. and Shenton, L. R. (1985). Method of Moments. In Kotz, S. and Johnson, M. L. (eds.) *Encyclopedia of Statistical Sciences*, vol. 5, pp 467–473. New York: Wiley.
- Casella, G. and Berger, R. L. (2002). *Statistical Inference*, 2nd edn. Pacific Grove, CA: New York.
- Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Duxbury: Chapman and Hall.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, B* **39**, 1–38.
- Fischer, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika* **46**, 59–77.
- Formann, A. K. (1986). A note on the computation of the second-order derivatives of the elementary symmetric functions in the Rasch model. *Psychometrika* **51**, 335–339.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*, 2nd edn. New York: Chapman and Hall/CRC.
- Ghosh, J. K., Delampady, M., and Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*. New York: Springer.
- Gustafson, J. E. (1980). A solution of the conditional estimation problem for long tests in the Rasch model for dichotomous items. *Educational and Psychological Measurement* **40**, 377–385.
- Hambleton, R. K. and Swaminathan, H. (1996). *Item Response Theory: Principles and Applications*. Boston, MA: Kluwer Nijhoff.
- Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. New York: Springer.
- Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd edn. New York: Springer.
- Lindsay, B., Clogg, C., and Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association* **86**, 96–107.
- Liou, M. (1994). More on the computation of higher-order derivatives of the elementary symmetric functions in the Rasch model. *Applied Psychological Measurement* **18**, 53–62.
- Pfanzagl, J. (1993). A case of asymptotic equivalence between conditional and marginal maximum likelihood estimators. *Journal of Statistical Planning and Inference* **35**, 301–307.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research.
- Robert, C. P. and Casella, G. (2005). *Monte Carlo Statistical Methods*. New York: Springer.
- Schilling, S. and Bock, R. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika* **70**, 533–555.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464.
- Severini, T. A. (2000). *Likelihood Methods in Statistics*. New York: Oxford University Press.
- Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika* **47**, 175–186.

Principal Components Analysis

D J Bartholomew, The London School of Economics and Political Science, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Eigenvalue – It is a number calculated from a square matrix, such as a correlation matrix. There are as many eigen values as columns (or rows) in the matrix and their relative values tell us something about the structure of the matrix.

Eigenvector – It is a set of numbers, having the same dimension as the matrix, each of which is associated with one of the eigenvalues.

Metrical variable – It is a variable measured on a continuous or discrete scale such as length or number of siblings.

Scree diagram (or plot) – It is a diagram showing the variances of principal components plotted in decreasing order of magnitude.

Introduction

The origin of principal components analysis (PCA), as we now know it, is in a paper by Hotelling (1933) published in the *Journal of Educational Psychology*. The place of publication indicates the intended field of application where it has been used for many years alongside factor analysis. Since then it has been used in many other fields, including the biological, physical, and engineering sciences. Its prime purpose is as a means of reducing the dimensionality of a multivariate data set and, also, of illuminating its interpretation by identifying a smaller number of variables which, in a certain sense, summarize the larger set.

The starting point of PCA is the matrix of correlation coefficients derived from the original data set. Strictly speaking, the rationale behind the method requires that the correlations be obtained from variables measured on some continuous scale. In practice, these variables may be precisely continuous, as with something like length or weight, or they may be discrete as when the variable is something we have counted. It is convenient to treat the two in the same way and we may conveniently refer to them collectively, as being metrical (i.e., measured). The method is sometimes used when the correlation coefficients have not been calculated from metrical variables. Such an analysis is not valid but, in spite of this, the method may give a useful, if rough, summary of the structure of the data.

A Preview

Although PCA has many applications, we shall use a single type of example in this exposition. This is because it is one of the most important and typical applications. The simplest way of conveying the basic idea is to take the case where there are only two variables. This we shall do in the following section. First, however, we shall look at a more realistic example to show what we can expect to achieve and to make some preliminary points.

It frequently happens that we have a set of test or examination scores, on a set of pupils. Such a dataset might appear as shown in Table 1.

We have left the number of pupils unspecified and given only the first four rows and the last two rows of the table; similarly, we have given the first four columns and the last. With a class of, say, 100 pupils and 20 subjects, or variables, there is a large amount of information to assimilate and PCA provides one way of summarizing that data in a meaningful way. A common way of making a summary, used by teachers for many years, is to sum or average the elements in each row. If these were marks obtained in the same subject over a period of time this might seem a reasonable thing to do. However, we would still need to consider the fact that the figures in the third column are larger than those in the others. Do we wish to give them greater weight in the final total? Perhaps they have resulted from the conflation of the marks obtained on two occasions, in which case giving them double weight seems just. But if they were marks awarded by a marker who, quite arbitrarily, decided to mark each paper out of 200, instead of 100 like the others, there is no reason for giving them extra weight. This simple illustration makes a point which occurs widely in the social sciences and which needs to be taken account of when carrying out PCA. The question is: Does the scale in which the marks are measured matter? That is, does it convey any relevant information? The answer is important because PCA depends upon it. As a general rule in social science, the units in which variables are measured is arbitrary and it is therefore customary to express them in units of their standard deviations. Equally, we need to consider whether the origin from which the marks are measured is arbitrary. If not, they should all be measured from the same common origin. If our table of data consisted only of marks scored on a scale 0–100 say, these questions do not arise because there is a common origin and scale at the outset, but in general we may impose these features by standardizing the

variables in the manner described. In practice, therefore, PCA is almost always carried out on standardized variables, which means that the product moment correlation coefficients are the starting point.

Before proceeding further with the analysis, it is highly desirable to inspect the correlation matrix carefully. All of the information in the original set of data is contained in the correlation matrix and nothing we do subsequently can alter that – it merely displays it in a different way. Such a preliminary inspection may reveal anomalies in the data and a practiced analyst will also be able to see the broad lines of the structure which PCA will reveal. We shall illustrate this procedure below.

The Case of Two Variables

In order to introduce the main ideas of PCA, we use an example with two variables and a sample size of five. Such a trivial example as this needs no analysis because the whole position can be taken in at a glance. It is, however, sufficient to lay the foundations for a fuller treatment. The data are given in **Figure 1**.

The two variables called x_1 and x_2 are plotted on the two axes of the left-hand diagram. They show roughly equal scatter in the horizontal and vertical directions but there is a marked tendency for x_2 to increase with x_1 . In the right-hand diagram, the points are shown in relation to a line drawn along a direction at, roughly, 45° to the horizontal. This line is, in fact, the first principal component. It is constructed so that it is nearest to the points in

the sense that the sum of squares of the distances of the points from the line is a minimum. If we now describe the positions of the points in relation to the first principal component, we shall note that the variation along the direction of the line is greater than the variation at right angles to the line. We have thus replaced one specification of the points, given by the left-hand diagram, by another as given on the right. Why might this be a useful thing to do? If we were told that the positions of the points had to be described by only one number, instead of two, their locations along the first principal component would be more informative than either of the values of x_1 or x_2 . By looking at the first principal component alone, we have made a useful reduction in dimensionality at the price of losing the variation at right angles to the line. Furthermore, if we could attach some meaning to the first principal component, it would be all the more useful. To make these points more general, we move on to a more realistic, but still simple, example.

A Simple Educational Example

The following example appears to have been first given by Lawley and Maxwell (1963) but the treatment here is based on Bartholomew *et al.* (2008). We start with the correlation matrix which is given in **Table 2**.

As recommended above, we begin with an inspection of the correlations. Their most obvious feature is that they are all positive though none is particularly large. This means that pupils who do well in one subject tend to do well in all. This suggests that there may be some skill or ability underlying all these subjects. A second point to notice is that there is variation in the values of the coefficients; some pairs of subjects are more highly correlated than others. There tend to be relatively high correlations among the three mathematical subjects as there are among the three humanities subjects. However, the inter-correlations among pairs formed by taking one from each group tend to be lower. This suggests that the subjects may be meaningfully grouped in this way. We would expect these two features to be evident in the formal PCA.

In the simple example with only two variables used earlier, we were able to picture the situation because the data could be plotted in two dimensions. With the six variables we now have, the geometrical approach fails, though we can still use the terminology in a suggestive way. Thus, if we imagine the ability scores plotted in six dimensions, we would be looking for the direction of greatest variability to find the first principal component. In practice, we now have to resort to algebra. A direction in the geometrical case becomes a linear combination in algebra. A component will thus be of the following form:

$$y = a_1x_1 + a_2x_2 + \dots + a_6x_6$$

Table 1 Typical layout of marks for pupils (rows) in a number of columns

53	30	80	61	...	37
49	20	116	55	...	32
62	24	150	53	...	44
43	76	98	71	...	76
...
40	57	158	76	...	39
62	69	146	18	...	25

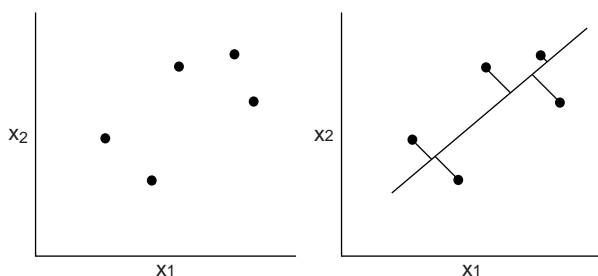


Figure 1 Illustration showing first principal component.

Table 2 Pairwise correlation coefficients between subject marks

	<i>Gaelic</i>	<i>English</i>	<i>History</i>	<i>Arithmetic</i>	<i>Algebra</i>	<i>Geometry</i>
Gaelic	1.00					
English	0.44	1.00				
History	0.41	0.35	1.00			
Arithmetic	0.29	0.35	0.16	1.00		
Algebra	0.33	0.32	0.19	0.59	1.00	
Geometry	0.25	0.33	0.18	0.47	0.46	1.00

The first principal component is determined so that the variance of y is a maximum subject to the total variance remaining fixed. The second principal component is calculated to be uncorrelated with the first and to have the maximum variance and so on as far as the sixth principal component, in this case. In practice, the components are determined very easily by computer software which are available in virtually every package intended for multi-variate analysis. The successive principal components are distinguished by subscripts and their corresponding coefficients (the a s) by inserting a second subscript.

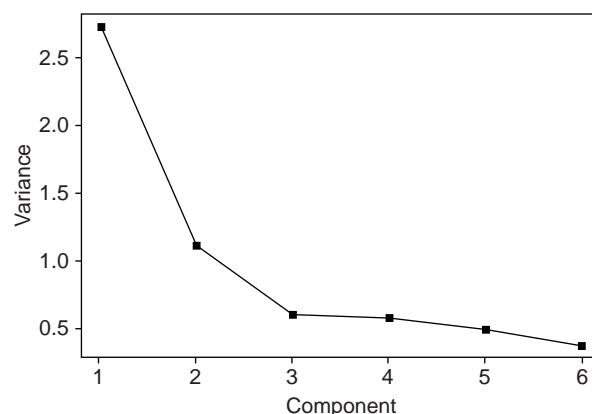
Since the prime object of the analysis is to effect a reduction in dimensionality, an important part of the analysis is to decide whether a small number of components, can, effectively replace the original variables. In the case of two variables, we argued that there was much more variation along the line than at right angles to it. In algebraic terms, we must look at the relative contributions made by the components. We then discard those whose contribution is judged to be too small.

We now illustrate how this is done using the standard computer output for the example of this section. **Table 3** gives the variance of each component expressed on a standard scale and also as a percentage of the total. **Table 3** also includes the cumulative percentages which help in selecting the number of components to discard. The same information can be presented graphically as in **Figure 2**. The latter is sometimes known as a scree plot.

It is clear that the first principal component accounts for almost half the variation and the first and second together for about two-thirds. Thereafter, the succeeding components account for less and less. The scree plot shows, for example, that beyond the third there is little variation left to be accounted for. There is no absolute rule for deciding when to stop including variables and it is here that practical considerations of interpretation come into play. One recommendation is to look for an elbow in the scree plot – that is, where it begins to flatten out. This occurs here at the third component. Another is to stop when the variance falls below 1. The rationale for this is that, on the standard scale we are using, each variable has unit variance and so there is little point in including a component which accounts for less variation than a single observation.

Table 3 Variance explained by each principal component, subject marks data

<i>Component</i>	<i>Variance explained</i>		
	<i>Variance</i>	<i>%</i>	<i>Cumulative %</i>
1	2.73	45.48	45.48
2	1.13	18.81	64.29
3	0.62	10.26	74.55
4	0.60	10.05	84.59
5	0.52	8.71	93.30
6	0.40	6.70	100.00

**Figure 2** Scree plot showing variance plotted against number of principal components, subject marks data.

The second aspect of interpretation concerns the interpretation which can be put on the components. For this we look at the coefficients of the principal components. Those for the first two principal components are given in **Table 4**.

From the first column, we notice that the coefficients are all positive and roughly equal. In effect, this component is not very different from a total or average. In looking for an interpretation, we are asking what such a function measures. It is clearly something to which all the test scores contribute roughly equally and it may thus be reasonably called general academic ability as it indicates something which makes a similar contribution, whatever the subject.

In the second column, the first three coefficients are positive and roughly equal; similarly, the second three are

Table 4 Loadings for the first two principal components, subject marks data

Subject	Loadings	
	First component	Second component
Gaelic	0.66	0.44
English	0.69	0.29
History	0.52	0.64
Arithmetic	0.74	– 0.42
Algebra	0.74	– 0.37
Geometry	0.68	– 0.35

roughly equal but this time they are negative. What the first three contribute is therefore an ability to do well in humanities subjects and the second does the same for the mathematical subjects. The second principal component, however, measures the difference between these two abilities. It, therefore, indicates the extent to which a candidate's abilities favor mathematics rather than humanities.

Although this is a fairly simple set of data, the example illustrates the main elements of a typical PCA.

The General Case

There is little to add to the example we have given for six variables because it sets the pattern for the general case. In mathematical language, we have to calculate the eigenvalues and eigenvectors of the correlation matrix. The eigenvalues correspond to the variances of the principal components as given in **Table 3**. The corresponding eigenvectors provide the coefficients of the variables in the components. Both of these operations are carried out in a routine fashion by the standard software. How many components are to be retained can be judged in the same way as above by means of a scree plot or by identifying the point at which the eigenvalues become less than 1. With larger problems, the question of the interpretability of the resulting components assumes greater prominence. In a correlation matrix where most correlations are positive, the first principal component often turns out to be a general factor measuring the magnitude of something which all individuals possess. Bipolar components, like the second one above, are also common but beyond that it may be more difficult to provide ready interpretations.

Relationship with Factor Analysis

PCA and factor analysis are very similar and it is not uncommon to find PCA regarded as one method of factor analysis – as in some computer packages. The reason for this has its roots in the fact that in the early days, until the arrival of computers, there was very little difference in the

methods of calculation used for both methods. These involved numerical operations on the correlation matrix, or a matrix derived from it, by replacing the diagonal elements by positive numbers between 0 and 1. PCA differed from other methods of factor analysis only in that the diagonal elements were all made equal to unity. In a contemporary treatment, factor analysis starts from a probability model and its parameters are estimated by maximum likelihood or some near-equivalent efficient method. A detailed account of the relationship between the two methods is given in Chapter 7 of Bartholomew *et al.* (2008).

Sampling Issues

We have treated PCA as a purely descriptive method for studying the dimensionality of a multivariate data set. If the data were obtained as a random sample from some population, we might be interested in inferring something about the structure of the population. For this we would need to know the form of the population from which our sample had been drawn and from that we would need to deduce the sampling distributions of the eigenvalues and the elements of the eigenvectors. Some results are available if the population sampled is multivariate normal but in most educational research, these results have been little used and probably cannot be justified.

See also: Analysis and Interpretation of Multivariate Data.

Bibliography

- Bartholomew, D. J., Steele, F., Moustaki, I., and Galbraith, J. I. (2008). *Analysis of Multivariate Social Science Data*, 2nd edn. London: Chapman and Hall/CRC Press.
- Hotelling, H. (1933). The analysis of complex statistical variables into principal components. *Journal of Educational Psychology* **24**, 417–441; 498–520.
- Lawley, D. N. and Maxwell, A. E. (1963). *Factor Analysis as a Statistical Method*. London: Butterworths.

Further Reading

- Dunteman, G. H. (1989). *Principal Components Analysis (Quantitative Applications in the Social Sciences)*. Thousand Oaks, CA: Sage.
- Jolliffe, I. T. (2002). *Springer Series in Statistics: Principal Components Analysis*, 2nd edn. Berlin: Springer.

Relevant Websites

- <http://www.cmm.bristol.ac.uk> Centre for Multilevel Modelling, University of Bristol.
- <http://www.crcpress.com> CRC Press, Taylor & Francis Group.

Probability Theory

T Rudas, Eötvös Loránd University, Budapest, Hungary

© 2010 Elsevier Ltd. All rights reserved.

What Is Probability?

Probabilistic thinking plays an important role in most fields of scientific research. This role is central in disciplines engaged in large-scale data collection and interpretation. A probabilistic model formulates relationships among the observables – relationships that are not supposed to hold exactly for each observation but still give a description of the fundamental tendencies governing their behavior. Probabilistic models allow the researchers to incorporate uncertainty into the fundamental laws they use to describe their findings. In educational research and assessment, uncertainty occurs for two main reasons. First, in the case of individual assessment, the ability or knowledge to be measured usually can be observed, not directly, but through the performance with respect to a certain battery of test items. Uncertainty naturally arises with respect to how the person tested would have performed, if confronted with other similar test items or if the testing had taken place under different conditions. Second, when the aim is to assess the status, achievement, motivation, knowledge, and so on, of a larger population, usually, only a sample from this population is observed, and uncertainty arises whether or not the observed performance of those not sampled would have been the same as that of those in the sample. The first type of uncertainty is called epistemic and is fundamentally related to the difficulty of learning some of the characteristics of interest. The second type of uncertainty is called aleatoric and may be reduced by the application of appropriate sampling methods.

Probability theory is often considered to be a mathematical subject, with a well-developed and involved literature concerning the probabilistic behavior of various systems (see Feller, 1968), but it is also a philosophical subject – where the focus is the exact meaning of the concept of probability and the ways in which it relates to the fundamental aspects of our reasoning (see Kopylov, 2008; Shackel, 2008). There is also a considerable body of psychological research available on the perception of probability and economists have also made important contributions in modeling and understanding human behavior in probabilistic settings (Kopylov, 2008; Shackel, 2008). Given the large number of different approaches, it may not be a surprise that even today – nearly 500 years after the concept of probability was first used – there are competing ways of defining its exact meaning. These differences do have consequences with respect to some of

the statistical analyses one performs but, fortunately, it is usually the actual problem at hand that determines which view one can adopt.

The most generally adopted view of probability is that it is a numerical characteristic of observations or experiments that may be performed repeatedly. This numerical value influences the relative frequencies of the possible outcomes. The larger the number of repetitions, the closer one may expect the observed relative frequency of an outcome to be to its probability. This is the frequentist concept of probability. Probabilities are assumed to be existing independently of the observations made – that is, the frequentist view considers probability to be objective. The goal of a statistical analysis in the frequentist tradition is to reveal the probabilities or some of their relevant properties using the available data. Because a relative frequency always lies between zero and one, probabilities are also between zero and one and some other properties of probability are also implied. There are several applications where the frequentist view seems convincing. For example, when a random sampling procedure chooses a sample from a population, the probability of having an observation with a certain characteristic may be identified with the population fraction of those possessing that characteristic (or a known function thereof, depending on the sampling procedure).

Another approach to the concept of probability considers it to be the degree of belief an individual associates with the occurrence of certain observations. This is called subjective probability. Subjective probability also applies to observations that may not be repeated. For example, a student may associate a probability with his/her passing a certain exam even though that particular exam is not something that could be repeated (a retake is a different exam, taking all its conditions into account).

The goal of the statistical analysis in a probabilistic model, in the subjective sense, is to update the subjective expectations based on the available data. This usually leads to the application of Bayesian methods of statistics. Fortunately, the mathematical properties of probability in the subjective interpretation are essentially the same as in the frequentist interpretation.

Although there is a fundamental difference between the frequentist and subjective interpretations of probability, many of the probabilistic models that are used often in statistics have meanings from both perspectives and so do many of the common statistical procedures.

The split between the frequentist and subjective interpretations of probability developed during the twentieth

century. The first steps of probabilistic thinking (in the sixteenth century) concentrated around the determination of rules of winning strategies in simple games of chance. Later, the study of demography led to the discovery of laws that were probabilistic in nature and, finally, the analysis of measurement errors led to deep and useful results related to probability (Stigler, 1986). The development of mathematical tools to describe probability culminated during the first third of the twentieth century when Kolmogorov (1956) formulated the axioms of probability, with the expectation that all properties of (mathematical) probability may be derived from the axioms. The axioms have been criticized by various researchers but they remain the fundamental framework of the mathematical theory of probability.

The Mathematical Theory of Probability: Axioms

The mathematical theory of probability is based on three fundamental assumptions or axioms. These are not derived or proved based on other considerations but are posited to capture the essence of probability. The axioms refer to the probabilities associated with events that may be observed. The observable events reflect the substantive problem at hand to which the probabilistic model is applied. For example, when 20 students take an exam, the relevant events may be associated with the number who pass it – that is, 0 or 1 or 2 or ... or 20 if it is irrelevant, who passes and who does not. As the example illustrates – when the events of interest are selected – the irrelevant aspects of the problem at hand are disregarded. The fundamental requirement is that following the experiment, one has to be able to determine for every event whether or not that even occurred. If, for example, only Mary, John, and Jim passed the test, the event stating that three students passed has occurred and the others have not. Usually, the impossible event (that never occurs) and the sure event (that always occurs) are also included. There are certain operations that may be applied to events. The two most important operations are addition (union) and product (intersection). If A and B are two events, their sum

$$A + B$$

is also an event and it occurred if and only at least one of A and B occurred. The product

$$AB$$

is also an event and it occurred if and only if both of A and B occurred. With repeated application of these operations, one can generate events such as ‘at least 5 passed’. When the intersection of two events is impossible,

$$AB = \emptyset$$

we say that the events exclude each other.

The above operations may be – and often have to be – applied infinitely many times, one after the other. For example, if one tosses a coin and the question of interest is how many times the coin needs to be tossed before the first tail occurs, this number does not have an upper bound in the sense that no larger value may be observed than the bound. Rudas (2004) gives an intuitive description of the precise meaning of applying the operations infinitely many times. One may think that such infinitely long experiments are not possible in practice, but many procedures and arguments in statistics are asymptotic. This implies that they give a good approximation for large samples but, strictly speaking, are only true for infinitely large samples.

The axioms relate these operations to probabilities.

First, the probability of any event is a number between 0 and 1

$$0 \leq P(A) \leq 1.$$

Second, the probability of the impossible event is

$$P(\emptyset) = 0.$$

Third, if two events exclude each other, then the probability of their sum is the sum of their probabilities

$$P(A + B) = P(A) + P(B), \text{ if } AB = \emptyset.$$

This is, of course, the axiom of the greatest interest. It means that if, for example, no student can go to two classes during the same period, then the probability that a randomly selected (i.e., each student has the same chance of being selected) student in the school goes to the science class or the physical education (PE) class, is the sum of the probabilities that the student goes to the science class and that the student goes to the PE class. This is the additivity of probability, which is its fundamental property.

Note that all the assumed properties hold true for relative frequencies. This is because the axioms are selected to be the basis of a mathematical construction that is intuitively related to the relative frequencies. There are several properties that may be derived from the axioms. For example, that the probability of the certain event is 1 or that the property in the third axiom is also true for more than two events. For example, it is true for all subjects taught in school. In fact, usually one postulates a more general variant of the third axiom, namely that it is true for infinite sequences of events:

$$P(A_1 + A_2 + \dots) = P(A_1) + P(A_2) + \dots, \text{ if } A_i A_j = \emptyset \text{ for all } i \text{ and } j$$

where the meaning of the union of an infinite sequence of events is that at least one of them occurs and the meaning of the sum of an infinite sequence of probabilities is the limit of the finite partial sums (for details see Rudas, 2004).

A question of fundamental importance is whether the mathematical construction of probability that is based on the above axioms possesses the fundamental property

which characterized the frequentist interpretation of probability. The question is of lesser importance in the case of subjective probability that is not necessarily supposed to follow any rules (Kopylov, 2008). It is possible to construct a model with an event that has a fixed probability and imagine an infinitely long sequence of independent observations of the experiment with the relative frequency of the event recalculated following each observation. One obtains a sequence of numbers indicating the relative frequency of the event during the first 1, 2, 3, 4, . . . repetitions of the experiment. Then, mathematical methods – based on the axioms – can be used to show that this sequence, indeed, converges (i.e., gets closer and closer as the number of repetitions increases) to the probability of the event. In other words, the basic conceptual assumption behind the frequentist view of probability is implied by the axioms.

Conditional Probability

It, very often, happens to be the case that the probability of an event is of interest knowing that another even occurred. For example, in educational testing if A and B are two items and one knows that 99.5% of those who got item A correctly will also get item B correctly and only 0.1% of those who did not get item A correctly will be able to answer item B correctly, then it makes very little sense to ask item B following item A , because knowing the result of asking item A will almost certainly determine whether the answer to item B is correct. A formal approach to the intuitive concept used in this argument is that of conditional probability. The conditional probability that B is correct, given that A was correct is 0.995 and the conditional probability that B is correct if A was incorrect is 0.001 in the previous example. It turns out that the conditional probability of B given A has to be defined as

$$P(B|A) = P(AB)/P(A).$$

When $P(B|A)=P(B)$, the two events are independence, because knowing that A occurred does not influence the probability of B . In this case,

$$P(AB) = P(A)P(B).$$

Finding independent structures is of central importance in statistical analysis of data. In most of the social sciences – education not being an exception – most of the observables are related to each other and our understanding of their structure is greatly enhanced if certain independencies or conditional independencies may be assumed. Conditional independence of events A and B , given event C is defined as

$$P(AB|C) = P(A|C)P(B|C).$$

This concept may be extended to more general observables than events, called random variables.

Random Variables

Random variables are among the most important concepts used in probability and statistics. Intuitively, these are models of random quantities and are used in all cases when the observations that are possible may be characterized by a number. For example, there is uncertainty associated with the grade point average (GPA) values students in a class will have by the end of the academic year. These quantities may be modeled with random variables. For an unknown quantity to be a random variable, certain events have to possess probabilities in the sense described earlier. These events are defined by constructing a subset of the real numbers and the events are that the actual value of the random variable belongs to this subset. Interestingly, all such probabilities may be obtained by computing the probabilities of simple events, namely those that the value of the random variable does not exceed a given real number, a

$$P(X \leq a) = F_X(a)$$

for all real a values, where X denotes the random variable. The notation $F_X(a)$ refers to the fact that these probabilities – for a fixed random variable X – may be considered as a function of a . This is the (cumulative) distribution function of the random variable X . Distribution functions have certain properties, including that they cannot decrease if a increases; if a goes to minus infinity, then the value of the function converges to zero; and if a goes to infinity, the value of the function converges to one.

Although distribution functions are very useful in the mathematical study of random variables, another kind of functions – called density functions – are used often to characterize the distribution of random variables. Density functions – usually denoted as $f_X(a)$ – are non-negative functions and their value for a real number a is related to the change in the value of the distribution function around a . More precisely, the change in $F_X(a)$ around a may be measured by

$$\frac{F_X(a + \varepsilon) - F_X(a - \varepsilon)}{2\varepsilon}$$

where ε is a small positive number. This quantity may converge to a number when ε goes to zero. This number is the value of the density function of the random variable X at a . The name of the density function comes from the fact that, in areas where the density function has a higher value, the observations from the random variable occur more frequently – with larger density – than in areas where the function takes on smaller values.

For statistical modeling of observed data, it is important to distinguish between continuous and discrete random variables. The distribution function of continuous random variables is, in fact, a continuous function. Such a random variable takes on all its values with zero probability. Therefore, the concept of probability in this case is not

appropriate to describe the difference between possible observations. The density function, however, distinguishes among the different values also in this case. The value of the density function at a is called the likelihood of value a . Some values have a higher, and others lower, likelihoods. Random variables with a discrete distribution have values that may occur with positive probability and their distribution functions have jumps at such values and are horizontal between two such neighboring values. In this case, there is no density function because the above ratio does not converge to a number for all values of a . Here, the individual probabilities are identified with the likelihoods. The concept of a discrete random variable may be applied to the case when the values are not numbers, rather arbitrary categories. For example, a random variable indicating the educational level of an individual may take on the values elementary, high school, and college. In this case, one does not have a distribution function but each value is observed with a certain probability and these constitute a probability distribution – that is, the values are nonnegative numbers that sum to 1.

Independence of Random Variables

Generalizations of the concept of independence applied to random variables are among the most important ways in which simple models are formulated and tested in statistical analysis. If the random variable A takes on the values a_1, a_2, \dots, a_I and another random variable, B , takes on the values b_1, b_2, \dots, b_J , then observing a_i or observing b_j are events – just like observing both of them, $a_i b_j$. Then the random variables A and B are independent if

$$P(A = a_i, B = b_j) = P(A_i B_j) = P(a_i)P(b_j)$$

for all $i=1, 2, \dots, I$ and $j=1, 2, \dots, J$. The probabilities of the events $A_i B_j$ are the joint distribution of the two random variables. This joint distribution is greatly simplified when the two variables are independent. In this case, the joint distribution if A and B may be described by the marginal distributions $P(A_i)$ and $P(B_j)$. If, for example, $I=5$ and $J=3$, then – instead of the $IJ-1=14$ probabilities in the joint distribution – one has to know only $I-1+J-1=6$ probabilities in the marginal distributions. Perhaps more important is the simplification of the joint distribution as a product of the two marginal distributions. If the categories of variable A denote five different countries and the categories of B three different educational levels, then the appropriate fractions may be interpreted as probabilities and independence would mean that the numbers of inhabitants with a certain level of education within a country are proportional to the population sizes of the countries.

A similar generalization of the concept of conditional independence of two events, given a third one, to random

variables, is also possible. While independence of variables rarely occurs in education or, more generally, in the social and behavioral sciences, conditional independence may be observed surprisingly often – in particular, when conditioning is not on one variable, rather on a combination of variables, which is also a random variable. Sometimes, conditioning on an unobserved (latent) variable is useful. For example, in educational measurement, it is often observed – and sometimes assumed as part of a model (see, e.g., Bond and Fox, 2007) – that the responses of people at the same level of ability to two items are independent from each other. In this case, one variable is the response to one item (right or wrong), the other is the response to the other test item and the variable upon which conditioning occurs is the level of ability.

The concepts of independence and conditional independence may be extended to continuous random variables. In this case, the events are those defining the cumulative distribution function. Two random variables X and Y are independent if

$$F_{XY}(ab) = F(X \leq a, Y \leq b) = F(X \leq a)P(Y \leq b) = F_X(a)F_Y(b)$$

for all possible values of a and b . The function on the left-hand side is known as the joint cumulative distribution function of the two variables. The extension to conditional independence uses the conditional distribution functions of the variables, given a particular value of the conditioning variable.

Normality

In many physical systems, uncertainty is introduced by measurement error and a model may be appropriate in which the measured value Y is assumed to be the sum of the true value X and of a measurement error E

$$Y = X + E$$

where the behavior of the measurement error is a characteristic of the measuring instrument itself. Such an error may be described probabilistically and E may be thought of as a random variable. For example, the usual method of measuring blood pressure gives a value which may not be equal to the real blood pressure – rather, it is the sum of the true value (Y) and some measurement error (E). Very often in practice, the density of E is characterized by having a specific shape – that of a normal distribution. The family of normal distributions consists of distributions with different expected values (averages) and different variances (dispersions). If an instrument's error has a negative expectation then, on the average, it will report smaller values than the truth and if it has positive expectation, then, on the average, it reports larger than true values. When the expectation is zero, the measurement

(process using the instrument) is unbiased. The individual values in this case are likely to deviate from the true value and the typical size of the deviation is about the square root of the variance, the standard deviation. It is customary to use unbiased measurements and instruments with smaller dispersion are preferable to instruments with larger dispersion.

The family of normal densities is characterized by symmetry around the expected value (if it is zero, positive and negative errors are equally likely) and by large errors being less likely than small errors. In fact, with a normal distribution, it is quite unlikely to observe a value more than three standard deviations away from the expected value.

In many cases, statistical analysis is greatly simplified if the variables observed are supposed to have a normal distribution. Note that while the simplification means that the analysis is easier and more reliable for small sample size than it would be without the normality assumption, it has the disadvantage that the normality assumption may exclude potentially important characteristics of the data (such as multimodality or the presence of higher order interactions) from the analysis. An even stronger assumption is joint or multivariate normality of the variables. Such an assumption, however, needs to be justified based on substantive knowledge.

In addition to the measurement-error model, there is another – perhaps more important – way in which normality occurs in probability theory and statistics and this is related to the asymptotic behavior of sequences of random variables.

Asymptotic Theory

The asymptotic behavior of sequences of random variables – that is, the behavior of infinitely long sequences of random variables – is an involved mathematical concept but it has important implications for the statistical analysis of data from large samples. The link between random variables and samples is that each observation in a sample may be regarded as an observation of a random variable – where randomness is induced by sampling and /or measurement error. We have already seen that the sequence of recalculated relative frequencies of an event converges to the probability of the event. This is implied by the so-called law of large numbers that states that the sample mean converges to the population mean.

It turns out, however, that if the sum of the observations is divided by \sqrt{n} , instead of n , then this quantity does

not converge to a number, but remains a random variable asymptotically. To put it differently, the sample mean itself may also be considered a random variable: it is a function (the average of) other random variables (the observations). If the sample mean is multiplied by \sqrt{n} (i.e., the sum is divided by \sqrt{n}), then it will continue to have a distribution asymptotically (as opposed to being concentrated to one value). Quite surprisingly, the asymptotic distribution, often, does not really depend on the original distribution of the variables the averages of which are taken. For the asymptotic result to apply, it is sufficient (disregarding some technical conditions usually fulfilled in practice) that they are independent. In all such cases, the asymptotic distribution will be normal.

This result is called the central limit theorem and it explains the dominant role played by normal distributions. The average of arbitrary independent random variables – if appropriately scaled – tends to have a normal distribution, disregarding what was the shape of the original densities.

See also: Bayesian Statistical Analysis; Categorical Data Analysis; Educational Data Modeling; Large-sample Statistical Methods; Measures of Central Tendency; Measures of Dispersion, Skewness and Kurtosis; Multivariate Normal Distribution.

Bibliography

- Bond, T. G. and Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 2nd edn. Mahwah, NJ: Erlbaum.
- Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, 3rd edn, vol. 1. New York: Wiley.
- Kolmogorov, A. N. (1956). *Foundations of Probability*, 2nd English edn. New York: Chelsea Publishing.
- Kopylov, I. (2008). Subjective probability. In: Rudas (ed.) *Handbook of Probability*, pp 35–48. Thousand Oaks, CA: Sage.
- Rudas, T. (2004). *Probability Theory: A Primer*. Thousand Oaks, CA: Sage.
- Shackel, N. (2008). Paradoxes in probability theory. In Rudas (ed.) *Handbook of Probability*, pp 49–66. Thousand Oaks, CA: Sage.
- Stigler, S. M. (1986). *The History of Statistics*. Cambridge, MA: Belknap Press.

Further Reading

- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, 2nd edn, vol. 2. New York: Wiley.
- Gordon, H. (1997). *Discrete Probability*. New York: Springer.
- Rudas, T. (ed.) (2008). *Handbook of Probability*. Thousand Oaks, CA: Sage.
- Von Mises, R. (1981). *Probability, Statistics and Truth*. New York: Dover.

Recursive Partitioning

S Landau, King's College London, De Crespigny Park, London
S Barthel, GlaxoSmithKline, London

© 2010 Elsevier Ltd. All rights reserved.

Regression-type models, for example, multiple linear regression, logistic regression, generalized linear models, linear mixed models, or generalized linear mixed models, can be used to predict a future object or individual's value of the response variable from its explanatory variable values. However, as with all statistical modeling, the resulting predictions might not be valid if the underlying assumptions are not met. Departures from model assumptions might arise due to a mis-specified relationship between response and explanatory variables; for example, by incorrectly assuming linear relationships with covariates or omitting higher-order interactions between some of the explanatory variables.

An alternative prediction approach has evolved to overcome some of these potential problems. Recursive partitioning techniques provide a prediction rule by forming subgroups of objects (individuals) within which the response variable is relatively homogeneous, on the basis of their values on a set of predictor variables, and assigning each subgroup a predicted response. Future objects can then be allocated to one of these subgroups and the value of their response predicted. The allocation of an object to a subgroup operates recursively, in that the final subgroup is derived at a succession of 'yes/no' decisions. Parent groups or nodes are split into two subgroups or daughter nodes on the basis of the value of a single explanatory/predictor variable at a time. For continuous or ordinal predictor variables x a group can be split on the basis of high and low values; for example, allocate object i to left daughter if $x_i < 10$ and to right daughter if $x_i \geq 10$. For nominal variables, splits that separate the categorical variables into disjoint subsets are allowed; for example, for a nominal outcome z with possible categories A, B, C, or D allocate object i to left daughter if $z_i \in \{A, C\}$ and to right daughter if $z_i \in \{B, D\}$.

The resulting collection of partitioning rules can be conveniently summarized by means of a decision tree. We present some examples below. In the context of predicting the value of a continuous response, the resulting tree is called a *regression tree*; and, in the context of predicting the category of a nominal variable or classification, this is known as a *classification tree*. The recursive partitioning technique itself is also frequently called *classification and regression trees* (CARTs) in reference to the original program for fitting trees to data (Breiman *et al.*, 1984). Note that recursive partitioning for predicting the value of a categorical variable is an example of a supervised learning

with a training sample, which also comprises techniques such as discriminant analysis. In contrast, unsupervised learning to detect as-yet unknown groupings is the domain of cluster analysis.

Classification Trees

We begin by detailing the recursive partitioning method for classification. As an example, consider the data in **Table 1** from the Drumcondra study of educational achievement (Greaney and Kelleghan, 1984) previously analyzed by (Raftery and Hout, 1985). Educational transitions were determined for a sample of 500 Irish school-children aged 11 in 1967 when tuition fees for secondary education were removed. Here, the focus is on predicting taking a leaving certificate (binary variable Y) for the 463 pupils who entered secondary education. The following sociodemographic variables were considered as predictors: gender of pupil, Drumcondra Verbal Reasoning Test score (DVRT), a prestige score of father's occupation (higher = more prestige), and type of school (secondary or specialized vocational).

To derive the best classification rule, the following general steps are usually carried out:

- growing a large initial tree;
- pruning the tree; and
- evaluation of its predictive performance.

The tree is grown – or in other words, the sequence of “if this then that” decision rules generated – by using all or part of the sample as a training sample. The decision tree starts with an initial or root node that comprises all objects. To split the sample into two subgroups (1) every allowable split on each potential predictor variable is examined and (2) the best of these splits is determined on the basis of some criterion and executed to create left and right daughter nodes. Steps (1) and (2) are then re-applied to each of the daughter nodes to determine the node on which to perform the next split. The process is repeated until some stopping rule is met (more on this below). To determine the best split of a node G into left and right daughter nodes, G_L and G_R , a numerical *split function*, $\phi(G, G_L, G_R)$, which measures the improvement in node purity due to splitting, is evaluated for each possible (G_L, G_R) . The best split then corresponds to the node G_0 and daughters $(G_{0,L}, G_{0,R})$ for which

Table 1 Irish education data (first seven pupils)

ID	Sex	DVRT* (70–140)	Leaving certificate from secondary education	Prestige score for father's occupation (18–75)**	School
1	Male	113	Not taken	28	Secondary
3	Male	110	Taken	69	Secondary
4	Male	121	Not taken	57	Secondary
5	Male	82	Not taken	18	Vocational
6	Male	85	Not taken	28	Vocational
8	Male	98	Not taken	43	Vocational
9	Male	92	Not taken	33	Vocational

*Drumcondra Verbal Reasoning Test.

**Score constructed by Raftery, A. E. and Hout, M. (1985). Does Irish education approach the meritocratic ideal? A logistic analysis. *Economic and Social Review* 16, 115–140.

$$\phi(G_0, G_{0,L}, G_{0,R}) = \min_{G, G_L, G_R} [\phi(G, G_L, G_R)] \quad [1]$$

For categorical response variables, a number of split criteria have been suggested. They tend to be based on an *impurity function*, which is a function of the proportions of the training sample belonging to the different possible categories of the response variable, here denoted $p_{G,C_1}, p_{G,C_2}, \dots, p_{G,C_k}$ for node G and k possible response categories C_1, C_2, \dots, C_k . Suitable measures of impurity should take their maximum when a node in the tree contains an equal number of objects from each of the categories (as it is then impossible to associate the node with a response variable category). The function should take its minimum when the node is completely pure; that is, all objects in the node belong to the same response variable category. Two candidates for impurity functions are the *Gini index of diversity*,

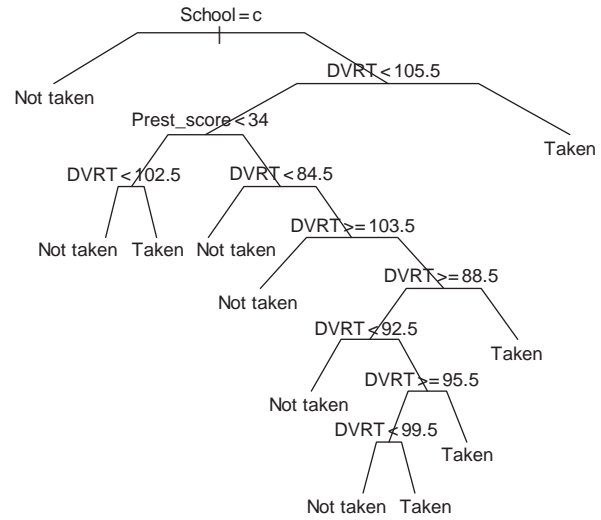
$$\begin{aligned} I_1(p_{G,C_1}, p_{G,C_2}, \dots, p_{G,C_k}) &= \sum_{j=1}^k \sum_{j' \neq j} p_{G,C_j} p_{G,C_{j'}} \\ &= 1 - \sum_{j=1}^k p_{G,C_j}^2 \end{aligned} \quad [2]$$

and the *information index*

$$I_2(p_{G,C_1}, p_{G,C_2}, \dots, p_{G,C_k}) = - \sum_{j=1}^k p_{G,C_j} \ln(p_{G,C_j}) \quad [3]$$

(where $0 \ln(0)$ is taken to be the limit of this function and set to 0). Split functions can then be defined by measuring the improvement in impurity after splitting node G into daughters, G_L and G_R , that is,

$$\begin{aligned} \phi(G, G_L, G_R) &= I_m(p_{G,C_1}, p_{G,C_2}, \dots, p_{G,C_k}) \\ &\quad - \left[\frac{n_{G_L}}{n_G} I_m(p_{G_L,C_1}, p_{G_L,C_2}, \dots, p_{G_L,C_k}) \right. \\ &\quad \left. + \frac{n_{G_R}}{n_G} I_m(p_{G_R,C_1}, p_{G_R,C_2}, \dots, p_{G_R,C_k}) \right], \quad m=1,2 \end{aligned} \quad [4]$$

**Figure 1** Large initial classification tree for Irish Education data based on the Gini index.

where n_G, n_{G_L}, n_{G_R} denote the sizes of the respective nodes.

Figure 1 shows a classification tree for the Irish education data resulting from applying the splitting function derived from the Gini criterion [2]. The first split is made on the basis of the school variable. The left daughter node consists of pupils attending a vocational school and they are predicted not to take the leaving certificate ($Y = 0$). The right daughter node consists of pupils attending a secondary school. They are further split in the next step according to verbal reasoning scores with those achieving higher scores (above 105.5) predicted to take the leaving certificate ($Y = 1$). Those in secondary school with lower verbal reasoning scores are further separated according to their father's prestige rating and following this split into increasingly smaller subgroups according to the DVRT variable. Recursive partitioning on the basis of the information criterion [3] was also carried out. In this instance, it made no difference in terms of decision rules nor final node labels, suggesting that the tree is not sensitive to the choice of splitting criterion.

When growing trees by splitting nodes, to maximize a homogeneity criterion, smaller and smaller nodes of progressively increased homogeneity are generated. Simple stopping rules exist for declaring terminal nodes. Commonly, a node is declared terminal when its size drops below a threshold value – as shown in **Figure 1**. The size of the tree was governed by setting the minimum size of a node to be considered for further splitting to 30 pupils and by restricting terminal nodes to contain at least ten pupils. An alternative rule is to stop when the node is homogeneous enough, for example, when the splitting criterion in [4] falls below some threshold. A more flexible approach to determining the number of branches of a tree is to use

a pruning algorithm. This involves growing a very large decision tree initially and then collapsing later branches using what is referred to as cost–complexity pruning to generate a nested sequence of decision trees. The latter operates by minimizing a cost–complexity function. This function is made up out of the cost of the final tree as measured by the average terminal node impurity

$$\text{Cost}(G_1, G_2, \dots, G_q) = 1/n \sum_b n_{G_b} I_m(p_{G_b, C_1}, p_{G_b, C_2}, \dots, p_{G_b, C_k}) \quad [5]$$

where G_1, G_2, \dots, G_q are the terminal nodes, and the number of terminal nodes q by defining a cost–complexity function as:

$$\text{CC}_\alpha(G_1, G_2, \dots, G_q) = \text{Cost}(G_1, G_2, \dots, G_q) + \alpha q \quad [6]$$

The number $\alpha \geq 0$ is called the complexity parameter. The aim is to simultaneously minimize both the cost and the complexity of the tree, with α determining the trade-off between the two. Under the extreme case $\alpha = 0$ the function is allowed splitting to a point where each terminal node contains only a single observation and so is 100% pure, which clearly bears the danger of overfitting. The larger the value of the complexity parameter the shorter the pruned tree with terminal nodes G_1, G_2, \dots, G_q as the increase in complexity increasingly outweighs the reduction in cost. (The extreme case is given by $\alpha > 0$ that leads to a single terminal node.)

Having said that, selecting a final decision tree still requires the user to specify an arbitrary threshold. This might be in the form of a simple stopping rule or by means of the complexity parameter when carrying out cost–complexity pruning. Therefore, how does one determine the best tree in an objective manner? In practice, the best pruned decision tree is chosen to be the one that best fulfills the goal of predicting the value of the response variable. A nested sequence of competitor trees is constructed, for example, by using a series of values for the complexity parameter. Predictive performance of each competitor tree is evaluated by means of some index and the tree that maximizes the index is chosen. In the context of predicting the values of categorical variables, terminal nodes are assigned to the most frequent category and the tree's predictive performance is typically measured by the percentage of cases that have been correctly classified.

However, as with all statistical model fitting, when (future) predictive performance is evaluated from the same sample that was used to develop the model/decision rule estimates tend to be too optimistic. (This is akin to R^2

in regression being upwardly biased.) To avoid such bias, separate training and test samples can be used to develop and measure the predictive performance of the decision tree, respectively. However, samples are often too small to provide sufficient numbers for reliable tree construction and independent testing. In prediction, an approach that provides the best of both worlds, that is, uses most of the available information to develop the tree and provides unbiased estimates of its predictive performance, is cross-validation. The latter operates by randomly dividing the sample into V parts of equal size, n_v . Each of the V groups is in turn set aside to temporarily serve as a test sample. A decision tree is developed on the basis of the remaining $n - n_v$ objects and predictive performance evaluated on the set-aside cases. If this process is repeated for each of the V subsamples, then the average performance index provides an unbiased estimate of the tree's predictive power. It has been suggested that little is gained by more than ten partitions, so V is often taken to be in the 5–10 range. (Note that some software packages provide the leave-one-out rule for prediction where $V = n$ and $n_v = 1$. This means each subject is set aside exactly once and requires the whole recursive partitioning process to be performed n times.)

Table 2 shows the predictive performance of a nested sequence of decision trees to predict Y generated by cost–complexity pruning. Increasingly larger choices of the complexity parameter α lead to increasingly pruned trees. The tree shown in **Figure 1**, which had been derived by using a simple stopping rule, corresponds to cost–complexity pruning with $\alpha = 0$ and is very complex. Increasing the value of the complexity parameter eventually leads to the best pruned tree shown in **Figure 2** which corresponds to $\alpha = 0.0135$. This value for the complexity parameter and its associated pruned tree maximizes the predictive performance as measured by the cross-validated percentage of cases correctly classified (**Table 2**). (Note that when cross-classification is used to select the complexity parameters, the computer algorithm needs to ensure that nested sequences of decision trees are being evaluated. Most software for recursive partitioning, e.g., the R package *rpart* (Therneau and Atkinson, 1997) used here, ensure this by deriving all cost–complexity pruned trees within each cross-validation step.) Thus, the tree in **Figure 2** represents our best collection of partitioning rules to predict Y . We now conclude that pupils with lower DVRT and attending secondary school scores should be further separated only on father's prestige

Table 2 Cross-validated predictive performance of cost complexity pruning for Irish education data

Number of splits on original sample	0	1	3	4	5	10
Corresponding complexity parameter (α)	0.4730	0.0293	0.0135	0.0090	0.0045	0.0000
Percentage correctly classified by cost complexity function (estimated using $V = 50$ -fold cross-validation)	52.1%	74.7%	76.5%	74.1%	71.9%	74.1%

Values in bold indicate optimal pruning.

score with pupils with higher prestige scores (>34) predicted to take the leaving examination. We estimate that this set of decision rules classifies 76.5% of pupils correctly (compared to 52.16% correctly classified if we simply suggested that every pupil takes the final examination). The graph uses a mosaic plot to illustrate the distribution of the two certificate classes within each terminal node which helps to visualize the uncertainty of assigning the terminal nodes to their most frequent class.

Regression Trees

The recursive partitioning method for predicting the value of a continuous variable, that is, the construction of a regression tree follows the same basic steps as that of a classification tree. The tree is grown by successively splitting nodes into right and left daughter nodes. The heterogeneity of a node with regard to a continuous outcome is typically measured by the sums of squares criterion

$$SS(G) = \sum_{i \in G} (y_i - \bar{y}_G)^2 \quad [7]$$

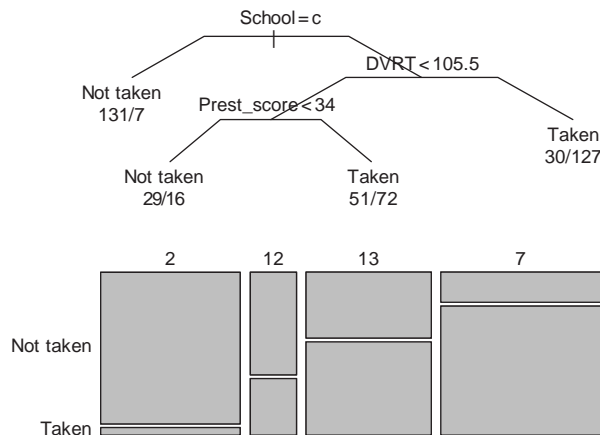


Figure 2 Pruned classification tree ($\alpha = 0.0135$) for Irish Education with class distribution in the terminal nodes depicted by mosaic plot.

where y_i denotes value of the response variable for object i in group G and \bar{y}_G the group average. The split function to be optimized across nodes and daughter nodes G_L and G_R can then be derived as:

$$\phi_{SS}(G, G_L, G_R) = SS(G) - [SS(G_L) + SS(G_R)] \quad [8]$$

As for classification trees, various simple or complex rules can be used to prune the tree. In the context of cost-complexity pruning, the cost of the tree with terminal nodes G_1, G_2, \dots, G_q is measured by the total terminal node heterogeneity:

$$\text{Cost}(G_1, G_2, \dots, G_q) = \sum_{v=1}^q SS(G_v) \quad [9]$$

and the cost-complexity function is simply defined as before in [6] using this cost definition. To determine the best regression tree, the predictive performance of the tree can again be evaluated by means of a test sample or by cross-validation. Terminal nodes are assigned a location summary, typically the mean, and the predictive power of the tree can simply be measured by the cost of its terminal nodes. Alternatively, if a scaled version is preferred, the proportion of variability in the response variable that can be explained by the predictions, that is,

$$R^2(G_1, G_2, \dots, G_q) = \sum_{v=1}^q SS(G_v) / SS(G_0) \quad [10]$$

where G_0 denotes the root node, which is familiar from regression, can be used.

As an example, consider the Forbes list of 2000 leading companies, collected by *Forbes Magazine* for the year 2004 (available in the Handbook of Statistical Analysis Using R (HSAUR) package in R). Available company information includes its name, world ranking, the country the company is situated in, a category describing what the company produces, amount of sales of the company, and its profits, assets, and market value (Table 3). Here, we want to develop a decision tree for predicting profits from sales, assets, and market value. Initial growing of a regression tree based on the sums of squares splitting criterion resulted in the tree displayed in Figure 3. (Here

Table 3 Forbes 2000 data (top 6 ranked companies)

Rank	Name	Country	Category	Sales (billion US\$)	Profits (billion US\$)	Assets (billion US\$)	Market-value (billion US\$)
1	Citigroup	United States	Banking	94.71	17.85	1264.03	255.30
2	General Electric	United States	Conglomerates	134.19	15.59	626.93	328.54
3	American Intl. Group	United States	Insurance	76.66	6.46	647.66	194.87
4	Exxon Mobil	United States	Oil and gas operations	222.88	20.96	166.99	277.02
5	BP	United Kingdom	Oil and gas operations	232.57	10.27	177.57	173.54
6	Bank of America	United States	Banking	49.01	10.81	736.45	117.55

a simple stopping rule based on a low cost–complexity parameter of $\alpha = 0.1$ was used.) Since we are now predicting the value of a continuous variable, the predicted profit value is depicted for each terminal node of the tree. **Table 4** shows the predictive performance for a series of cost–complexity pruned trees. The decision tree corresponding to the complexity parameter with the best R^2 is shown in **Figure 4**. The initial eight terminal nodes have been merged back to four terminal nodes. According to our cross-validation estimate, the regression tree can explain 23.3% of the variability in profits. The boxplots underneath the decision tree in **Figure 4** illustrate the unexplained variability in profits. The final tree suggests that market value is a predictor of profits.

Extensions

An alternative approach to growing trees and then pruning them back to avoid overfitting, is the use of p -values, possibly adjusted for multiple comparisons, for evaluating the quality of a split. To construct a so-called conditional *inference tree*, a significance test of the null hypothesis, that the response variable and a two-level group factor are not associated, is performed at each node of the tree, and the split chosen for which the test statistic most exceeds a prespecified nominal level. For example, to carry out a likelihood ratio test of association with a nominal response, we would use the deviance test statistic which is simply the splitting criterion resulting from the information index multiplied by $2n_G$, and compare it against a nominal level derived from a chi-squared distribution

with 1 degree of freedom. (The latter also illustrates that the p -value depends on both the reduction in impurity as well as the node size.) Alternatively, to stay within a strictly nonparametric framework and avoid assumptions regarding the sampling process, (conditional) permutations tests can be employed. This connects recursive partitioning to classical statistical test problems and so addresses the issue of overfitting; for more information see Hothorn *et al.* (2006). **Figure 5** shows the conditional inference tree for the Irish Education data (created in R package party, Hothorn *et al.* (2005)) obtained by setting the nominal significance level to 5%. Like the best pruned classification tree in **Figure 2**, this approach results in four terminal nodes. However, the recursive splits performed to allocate pupils to one of the terminal nodes differ. While **Figure 2** splits the secondary school pupils with lower DVRT score on father's prestige score in the last partition, the conditional regression tree splits the secondary school pupils with higher DVRT score.

There are further extensions to the basic recursive partitioning principle. For example, parent nodes can be split into more than two daughter nodes and it is also possible to consider splits on a linear combination of two or more continuous or ordinal predictor variables (Breiman *et al.*, 1984). Furthermore, various weighting schemes can be used. Object weights can be incorporated into the split functions. Such weights may reflect predefined varying importances of different objects. Other weights might also be used like inverse probability weights to counteract the effects of nonrandom sampling. Alternatively, predictor variables can be assigned different weights to reflect differential costs occurring when employing them. Finally, for classification trees, prior class probabilities or a loss matrix describing the cost incurred by misclassifying objects for each possible class, can be used. The ability to specify prior probabilities might be particularly useful in case control studies to achieve unbiased estimation of performance indices such as percentage correctly classified, positive predictive value (PPV, proportion of cases in those predicted to be cases), and negative predictive value (NPV, proportion of noncases in those predicted not to be cases) which depend on prevalence; for more on evaluating prediction rules see Hand (1997). (All these weights are, implemented in part (Therneau and Atkinson, 1997)).

Predictions might be subject to substantial uncertainty. Such uncertainty may, for example, manifest itself by

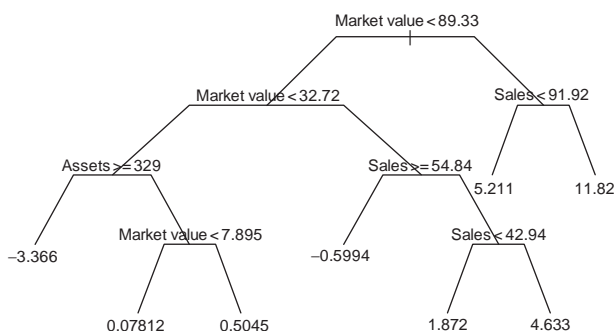


Figure 3 Large initial regression tree for Forbes 2000 data based on the sums of squares criteria.

Table 4 Cross-validated predictive performance of cost–complexity pruning for Forbes 2000 data

Number of splits on original sample	0	1	2	3	4	5	6	7
Corresponding complexity parameter (α)	0.2375	0.04600	0.0426	0.0203	0.0185	0.0110	0.0108	0.0100
R^2 achieved by cost complexity function (estimated using $V = 10$ -fold cross-validation)	0.00%	15.85%	13.82%	23.34%	21.50%	21.25%	19.24%	22.24%

Values in bold indicate optimal pruning.

different decision trees generated during cross-validation. For example, during the cross-validation of the Forbes 2000 data set we noticed that while the mode number of splits suggested was 3, split numbers as low as 1 or as high as 7 also occurred. *Bagging or bootstrap aggregation* is a technique that can be used with many classification or regression methods to reduce the variance associated with predictions and so improve predictive performance. The basic procedure involves sampling objects in the training sample with replacement, generating predictions for each such bootstrap sample, and obtaining an overall prediction by averaging across bootstrap samples. Since regression trees predict the value of the response variable by its mean in a local sample, the variance of the bagged prediction is reduced compared to that of the initial prediction due to averaging over multiple samples. A similar argument can be made for classification trees or more generally for local prediction methods. However, note that bagging is a black

box machine learning approach – we can no longer follow a single succession of yes/no decisions to arrive at our prediction. As an example, we bootstrapped the Forbes 2000 data 1000 times, generated a regression tree (using a stopping rule based on cost-complexity pruning with $\alpha = 0.01$) and associated profit predictions each time and averaged these predictions over only the predictions of profits which had not been included in the respective bootstrap sample. This provided a cross-validated estimate of predictive performance of $R^2 = 25.6\%$, which is somewhat better than the performance achieved by simple CART with optimal cost-complexity pruning (cf. **Table 4**). The bagging procedure is a special case of a more general approach called random forest (Breiman, 2001) which is available, for example, in the R package randomForest (Breiman *et al.*, 2005).

Conclusion

Recursive partitioning are essentially fairly simple non-parametric techniques for prediction and classification. When used in the standard way, they provide trees which display the succession of rules that need to be followed to derive a predicted value or class. This simple visual display explains the appeal to decision makers from many disciplines.

The prediction techniques are nonparametric in that they do not rely on any particular assumption about the type of dependence of the dependent variable on the

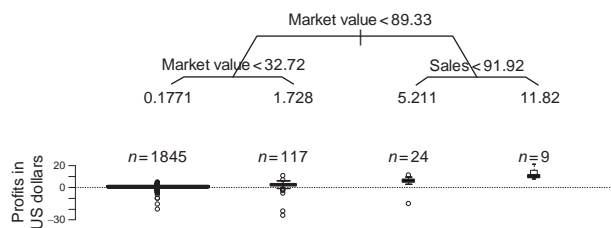


Figure 4 Pruned regression tree ($\alpha = 0.0203$) for Forbes 2000 data with the distribution of profits in the terminal nodes depicted by boxplots.

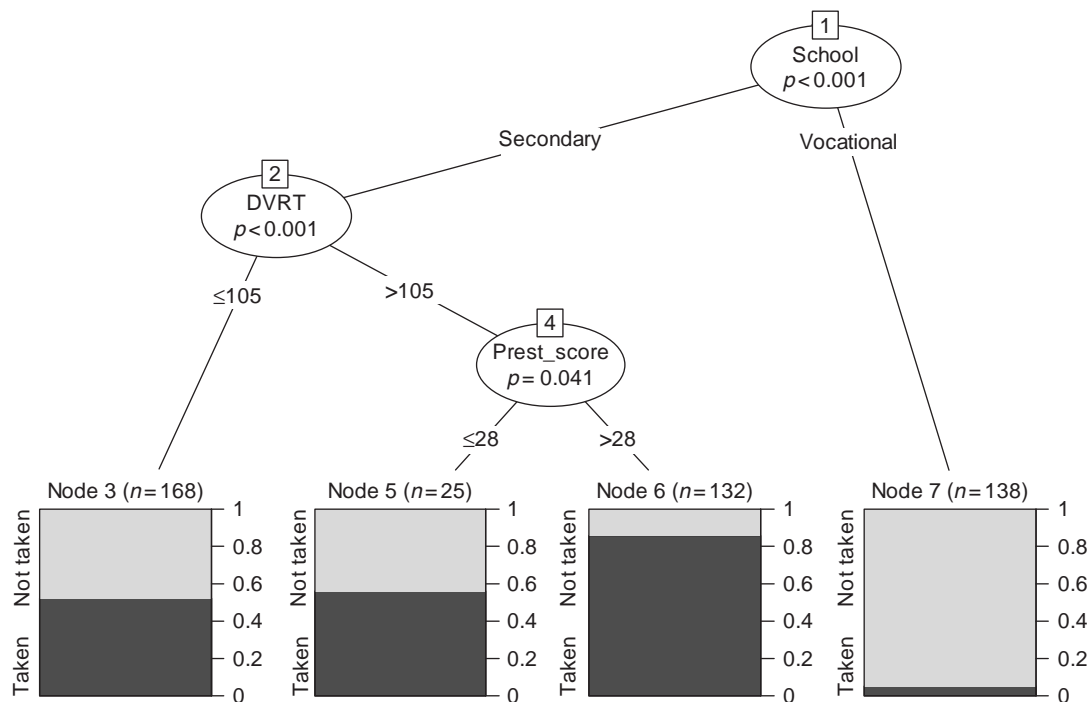


Figure 5 Conditional inference tree for the Irish Education data.

predictors. This distinguishes them from regression models where the relationship between response and predictor variables is parameterized *a priori*, for example, by assuming linear relationships. (However, note that the final tree model can be written as a regression model containing suitably defined indicator variables as explanatory variables.) This nonparametric property is seen as an essential practical advantage of prediction by recursive partitioning over prediction by regression when *a priori* information about the variables is limited.

However, it needs to be emphasized that recursive partitioning is an exploratory rather than an inferential technique. Regression or classification trees can provide a visualization of the main structures of a regression relationship. But users need to be reminded that the fine graining of identifiable subgroups is limited by the available predictors and their strength of relationship with the outcome. Tree information might be useful for developing a model of the nature of the relationship between response and explanatory variables. Once an appropriate population model is identified, inferential techniques can then be applied to a new data sample to make statements about its parameters.

See also: Cluster Analysis: Overview; Generalized Linear Mixed Models; Generalized Linear Models; Multivariate Linear Regression.

Bibliography

- Breiman, L. (2001). Random forests. *Machine Learning* **45**, 5–32.
- Breiman, L., Cutler, A., Liaw, A., and Wiener, M. (2005). Random Forest: Breiman and Cutler's random forests for classification and regression. *R package version 4*, 5–25. <http://cran.r-project.org/web/packages/randomForest/index.html> (accessed September 2009).
- Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Boca Raton, FL: Chapman and Hall/CRC.
- Greaney, V. and Kelleghan, T. (1984). *Equality of Opportunity in Irish Schools*. Dublin: Educational Company.
- Hand, D. J. (1997). *Construction and Assessment of Classification Rules*. Chichester: Wiley.
- Hothorn, T., Hornik, K., and Zeileis, A. (2005). party: A laboratory for recursive part(y)itioning. *R package version 0.2-8*. <http://CRAN.R-project.org> (accessed September 2009).
- Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* **15**, 651–674.
- Raftery, A. E. and Hout, M. (1985). Does Irish education approach the meritocratic ideal? A logistic analysis. *Economic and Social Review* **16**, 115–140.
- Therneau, T. M and Atkinson, E. J (1997). An introduction to recursive partitioning using the rpart routine. *Technical Report 61 Mayo Foundation*. <http://www.mayo.edu/hsr/techrpt/61.pdf> (accessed September 2009).

Relevant Website

<http://cran.r-project.org> – CRAN: Package HSAUR.

Sampling

J Qian, Educational Testing Service, Princeton, NJ, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Cluster sampling – It is a sample design that selects one or several clusters of sampling elements.

Design effect – It is the ratio of the variance of a statistic from the complex design over the variance from the SRS with the same sample size.

Equal probability of selection method (EPSEM) – It is a sample design that samples all units with equal probability.

Horvitz–Thompson estimator – It is a weighted unbiased estimator of population total for the complex design with unequal selection probabilities.

Population – It is the full specified set of elements being studied.

Primary sampling units (PSUs) – They are the clusters on the sampling frame pertaining to the first stage of multiple-stage sampling, such as the public schools in the state.

Probability samples – It is a sample design that uses the methods based on probability theory to minimize subjectivity in the choice of units to survey.

Sample – It is a subset of units/elements drawn from a population.

Sample design – It specifies the method for selecting every possible sample from a sampling frame.

Sample weights – They are values that are used to correct for unequal selection probabilities due to complex sample designs and to adjust for nonresponse or nonparticipation in surveys. The base weight for a sample element usually equals the inverse of its probability of selection.

Sampling error – It is the variation between the estimates based on samples and the true characteristics of a population.

Sampling frame – It is a list of the units or the groups of units in the population.

Sampling unit – It is an element or an aggregate of elements about which information is measured at different stages in a sampling process.

Simple random sampling (SRS) – It is a sample design in which all units are selected with equal probability with no application of stratification and clustering techniques.

Stratified sampling – It is a sample design in which a population is partitioned into strata based on a certain characteristic that is known for every

sampling unit in the population, and then selecting samples independently from each stratum.

Two-stage sampling – It is a sample design that selects primary sampling units (PSUs) for the first stage of selection and then selects the elements in the selected PSUs for the second stage.

Introduction

With the growing importance of sample survey in understanding the world today, sampling plays an essential role in the gathering of information on individuals, households, institutions, businesses, and environmental resources.

The sampling techniques used in gathering information reflect the scope and purposes of a survey. Currently, many modern sample surveys are conducted for a variety of different purposes. One example is the National Assessment of Educational Progress (NAEP), an educational sampling survey (Allen *et al.*, 2001). NAEP, the only ongoing measure of what students in the United States know and can do in a variety of subject areas, reports scores for different demographic groups such as gender, ethnicity, school type, school location, and individuals making use of testing accommodations. It provides objective information on student performance to educators and policymakers at the national, state, and local levels.

Basic Concepts in Sampling

The population (the fully specified set of elements being studied) and the sample (a subset of units/elements drawn from a population) are two basic concepts in sampling. A sampling unit refers to an element or an aggregate of elements about which information is measured at different stages in a sampling process. The size of the sample depends on the accuracy of estimation required given the purpose of the study. A parameter refers to a numerical value, such as the mean or standard deviation, which characterizes the population. Survey sampling creates statistics, which are numerical values derived from the sample, designed to estimate one or more parameters of a particular population. The fundamental idea of sampling is to estimate the parameters of an entire population or to make inferences about a population based on statistical

models and data derived from only a sample of that population. Thus, the intent of sampling is not to describe the particular units that, by chance, are in a sample but instead to obtain a composite profile of a population. In fact, reporting protocols for survey results at times require that individual respondents never be identified. For instance, NAEP, under the provisions of federal laws governing the protection of information, must ensure the protection of individually identifiable data from unauthorized disclosure (National Center for Education Statistics, 2003).

Sample Design and Probability Sampling

Conducting a sampling process entails the successful coordination of several different steps: (1) defining the population of study and specifying the sampling frame; (2) creating a sample design, specifying a sampling method, and determining the necessary sample size; and (3) implementing a sampling plan, which includes sample selection, data collection, and sometimes a weighting process.

Sampling frame is a list of the units or the groups of units in the population. Sample design specifies the method for selecting every possible sample from a sampling frame. Trained statisticians usually specify a sample design using methods based on probability theory to minimize subjectivity in the choice of units to be surveyed. Samples drawn using these methods are called probability samples. The estimates of the characteristics of a population will generally vary from sample to sample. Variation between the estimates based on samples and the true characteristics of a population is referred to as sampling error, which is often expressed in terms of the standard error and bias of estimates. In statistical theory, the sampling errors will converge to zero as the sample size increases. Probability samples provide the basis for calculating the sampling error; the sample design will assign a nonzero probability for each sample to be selected. A special sample design is the equal probability of selection method (EPSEM), which samples all units with equal probability. Occasionally, the sample design only defines the rules for including sample units instead of specifying the explicit inclusion probabilities. Then the inclusion probabilities have to be derived from statistical models (Qian, 2008). The sample design determines the procedure for sample selection and the choice of an estimator of a population parameter, such as a sample average or a Horvitz–Thompson estimator (Cochran, 1977), the latter of which is discussed later.

Sometimes, nonrandom sampling approaches are also used in data collection. Typical nonrandom sampling approaches are convenience sampling and quota sampling. A convenience sample is composed of units that

are available, rather than units that are selected from a sampling frame. An example is a group of students who respond to a flyer posted on a bulletin board on a university campus that seeks participants in a survey. A quota sample sets quotas, using convenience sampling, to obtain the samples. For instance, a quota sample might be filled by some number of customers at shopping malls in different locations who are willing to be interviewed. In practice, the cost of nonrandom sampling is low, but the level of representation of the sample is not under control. Often, a sample of nonrepresentative respondents will yield a poor estimate that severely deviates from the true parameter; in the example of the campus flyer above, only students who are interested in the topic of the study or who are financially strained might respond. Such deviation, or bias, persists even if several samples under the analogous sampling plan are collected. These sampling approaches will cause selection bias because the approaches tend to under/overrepresent certain kinds of units from the sample, and cause response bias because they tend to include those who are willing to be interviewed. These kinds of errors caused by nonrepresentative sampling, which are a type of nonsampling errors, are much harder to quantify than sampling errors. Other types of nonsampling errors can be caused by an inadequately specified sampling frame, a poorly designed questionnaire, or coding and measurement errors. In particular, a well-specified sampling frame is essential to the success of a sampling plan. Possible defects in specifying a frame include noncoverage, undercoverage, overcoverage, duplication, or misclassification (Kish, 1965).

In general, a good sample design for survey research should be characterized by good representation, randomization, and realism (Kish, 1987). Here realism refers to the practical aspects involved with a survey; in particular, realism implies that the sampling plan is implementable in the field and that the cost of the study is under control.

Five Basic Probability Sampling Methods

The five basic probability sampling methods are simple random sampling (SRS), stratified sampling, cluster sampling, multistage sampling, and systematic sampling, all detailed below. More complex sample designs are generally developed based on these five fundamental approaches.

Simple Random Sampling

SRS is the most basic type of probability sampling. For a sample of size n selected by SRS, each sampling unit in the sample has the same chance of selection, and all the possible subsets in the population of size n have the same chance of being chosen as a sample. Note that the SRS design is a type of EPSEM. To draw an SRS sample, the units in the sampling frame may be selected based on

numbers from a table of random numbers or from a random number generator on a computer. In practice, sampling without replacement is generally preferable to sampling with replacement because the former avoids including any sampling unit more than once and thus yields more efficient samples. To gain the same accuracy in estimation, an SRS sample without replacement requires a smaller sample size than the one with replacement, but the units sampled without replacement from a finite population will not be independent because each draw after the first is dependent on the prior draws (Horvitz and Thompson, 1952). This is undesirable because dependence is not a typical assumption and causes complexity in analysis, particularly in variance estimation. SRS is conceptually simple but requires a complete sampling frame, which again may not be an option in practice. Even if a complete frame is specifiable, large-scale surveys, such as NAEP, rarely find it operationally or economically feasible to select an SRS sample of units across a wide area, such as a country. Instead, many practical designs make use of more refined techniques such as stratification, clustering, multistage sampling, and sampling with unequal probabilities of selection.

Stratified Sampling

Stratified sampling designs involve partitioning a population into strata based on a certain characteristic that is known for every sampling unit in the population, and then selecting samples independently from each stratum. This design offers flexibility of sampling methods in different strata and gains improved precision of estimates of target parameters when each stratum is composed of units that are relatively homogenous. In practice, strata are often determined by field conditions, such as geographic regions or administrative structures. Stratified sampling design also enables specific segments of the population to be easily targeted. For example, in the combined NAEP national and state samples (Rust, 2004), the design treated each jurisdiction in the United States as a stratum; therefore, state results can be separately reported and be compared to each other.

Stratified sampling is often made with disproportionate sample allocation across strata, meaning that the stratum proportions in the sample do not represent the corresponding proportions in the population. To remedy this problem, stratum weights, defined for each stratum as the proportion of the size of that stratum to the total population, need to be applied in estimation. When this is done, the mean or proportion estimators for each stratum are unbiased and the estimator for the population, which equals the sum across strata of the stratum weights multiplied together with the corresponding stratum estimator, is also unbiased. Existence of an unbiased estimator for each stratum implies that the average of the estimates

over all possible samples in the stratum for the same design is equal to the true parameter being estimated. If stratum weights are not properly computed, it introduces bias in estimation. In general, the weighted mean of a proper stratified sample has less variability than the arithmetic mean of an SRS sample of the population. The stratum weights should also be employed in variance estimation for stratified data.

Cluster Sampling and Multistage Sampling

Instead of selecting sampling elements, cluster sampling design selects clusters, which are naturally occurring groups of elements. Examples of clusters are schools or residential blocks; a sample will then include all the elements in the selected clusters. Compared with SRS, this is a cost-efficient design because it reduces the number of schools and cities that need to be visited when constructing a sample. Sometimes selecting clusters is also necessary for practical reasons; an English listening test might need to be administered to a room of examinees all at the same time. The problem with the cluster sample design is that it causes an increase in the standard error of estimation when compared to SRS, which is called clustering effect. The clustering effect is mainly due to two factors: homogeneity of the elements within clusters and the average cluster size. The homogeneity within clusters can be measured by the intraclass correlation (Cochran, 1977). In practice, to reduce the homogeneity of the students in each sample of a listening test, students can be randomly selected from across all of a school's classrooms to form a cluster with better representation than a design in which all students in the cluster are taken from an intact classroom. Lowering the average cluster size will yield smaller standard errors.

In general, cluster sampling and stratified sampling reflect different statistical strategies. Cluster sampling focuses on operational feasibility, while stratified sampling stresses targeting specific segments of the population. Cluster sampling includes only elements in the clusters selected, whereas stratified sampling draws random samples from all its predefined strata. For better estimation, the stratified sample design should involve partitioning the population into groups with relatively homogeneous elements, whereas the cluster sample design should involve the forming of clusters with heterogeneous elements.

Despite its practical advantages, cluster sampling could still be costly due to the design requiring that the sample includes all the elements in large selected clusters. An improvement over cluster sampling is the multistage cluster sampling design. For instance, a combined NAEP national and state assessment draws a state sample by using two-stage cluster sampling. In selecting an NAEP state sample, the clusters in the sampling frame pertaining

to the first stage of sampling, called primary sampling units (PSUs), consist of all the public schools in the state; the elements in the selected clusters for the second stage, called secondary sampling units (SSUs), would then consist of the students within each PSU (Rust *et al.*, 2001). The first stage selects schools with probability proportionate to size (PPS) from the sampling frame. This method selects larger schools with a greater probability and smaller schools with a lower probability. The second stage randomly selects students within each sampled school. The total number of students selected within each sampled school is approximately equal. Although it is not an SRS, this design yields an EPSEM sample, which includes students in the sample with approximately equal probability. An example of a three-stage cluster sampling is the General Social Survey (GSS), which is based on the National Opinion Research Center's (NORC) 1990 National Sample (Tourangeau *et al.*, 1993).

Systematic Sampling

Systematic sampling, often applied in selecting samples in the EPSEM design, yields results that are equivalent to those derived from SRS. Moreover, systematic sampling can be used in implementing multistage cluster sampling designs, especially for those using the PPS method in the first or second stage selection process.

A systematic sampling with a random start involves three steps: (1) computing the sampling interval, say p , which equals the population size divided by the desired sample size; (2) randomly selecting an element from the sampling frame between 1 and p , denote this as k ; and (3) including all elements $k+p$, $k+2p$, ... in the sample. To illustrate, let population size N be 90 and sample size n be 6. The sampling interval is 15. Let the first randomly selected number be 8. Then the six elements included in the sample are {8, 23, 38, 53, 68, 83}. In case p is not an integer, p needs to be rounded to the nearest integer. The elements on the frame, instead of being lined up as a queue, needs to be arranged as a circle. Then, an element between 1 and N is randomly selected, and every p th element is selected thereafter until n elements have been selected.

The systematic sampling design is an EPSEM design; the method, in this sense, provides results equivalent to those from random sampling. However, if the elements on the frame have a cyclic pattern, the sampling interval must avoid periodicity in its sampling within the elements or risk having estimates with serious bias. In case of periodicity, one way to avoid it is to sort the elements in a different order.

For a multistage cluster sampling design that employs PPS sampling in its first stage of selection, such as the NAEP state assessment, systematic PPS sampling with a random start can be used to select PSUs. This method of

sampling gives each PSU a selection probability proportional to its size. From sampling frame, a cumulative record of counts is first created that tallies counts of units for all PSUs on the list. The sampling interval is calculated by dividing the total number of units in the file by the number of PSUs to be selected. Then a random start, a number within the range of sampling interval, is picked. The first PSU is selected when the random start hits the range of the PSUs in the cumulative record. Subsequent selection of PSUs is determined by the prior number plus the sampling interval. This process continues until the end of the cumulative record is reached.

A systematic sample can be more efficient and represent the population better if the units on the frame are sorted by certain demographic variables before selection. Sorting in this manner places similar PSUs next to each other on the frame and guarantees that the PSU sample will include a mix that is representative along the dimensions used to sort the list. For example, in the sample design of the General Social Survey (GSS), PSUs were sorted by region (Northeast, Midwest, South, and West), by state within region, by minority quartile within state, and by per capita income within minority quartile.

Sample Size

Sample size is mainly determined by the sample design, required accuracy of estimates, and resource constraints. For a particular design, sample size can be determined by the level of accuracy required, or confidence interval at a given confidence level desired (Cochran, 1977: 75–78). A pilot study can yield estimates of the parameters used in computation by providing useful information for determining sample design and sample size. An increase in sample size will reduce sampling error, although the relationship is not linear. On the other hand, a small sample is easier to manage and will introduce less nonsampling error. In statistical theory, it is sample size and not population size that determines the accuracy of estimates, unless the sample size is relatively large, say greater than 5% of the population. The resource constraints of a study enforce an upper limit on the sample size.

The HT Estimator

The Horvitz–Thompson estimator estimates the population total by its weighted sample total, which equals the sum of all the observed values times the corresponding case weights,

$$\hat{Y} = \sum_{i \in R} \frac{y_i}{\pi_i}$$

where $1/\pi_i$ is the case weight, y_i is the measured value of case i , and R is the set of sampled cases of size n .

This estimator is unbiased with respect to the total population (Horvitz and Thompson, 1952).

Sample weights are used to correct for unequal selection probabilities due to complex sample designs and to adjust for nonresponse or nonparticipation in surveys. Otherwise, the estimator based on observed cases in the sample will be biased. The base weight for a case equals the inverse of its probability of selection. When the target parameter is a mean or proportion, the HT estimator can be estimated by a ratio estimator, which equals the Horvitz–Thompson estimator of the population total divided by the sum of the weights.

The base weights are subject to adjustment for nonresponse, the causes of which could be noncontact or noncooperation. Both types of nonresponse could lead to bias in estimates. In the adjustment, each base weight is multiplied by a nonresponse factor, which equals the inverse of the response rate in each adjustment class (a class formed by the demographic variables of interest, such as gender and ethnicity (Kish, 1992)). So the weighted counts, the sum of weights in the adjustment classes, are equal to the sampled counts of units by design. This adjustment is based on the assumption that the responses missing within each adjustment class are random.

After a nonresponse adjustment, some variables, such as gender proportions, could show considerable differences between a weighted sample distribution and its corresponding population distribution. Poststratification and raking can be used to correct for these differences, to improve the precision of the survey estimates by reducing their mean squared error, and to enhance the comparability of the survey data under study with data from other surveys. Poststratification adjustment matches the weighted sample cell counts to the population cell counts by applying a proportional adjustment to the weights for certain groups (Kish, 1965). The population distributions are often obtainable from census data or other resources. A raking procedure, also called the Deming–Stephan algorithm, iteratively adjusts the case weights in the sample to make the weighted marginal distributions of the sample agree with the marginal distributions of the population on specified demographic variables (Deming and Stephan, 1940; Haberman, 1979).

Variance Estimation

In reporting survey results, one should always provide a summary of sampling error associated with particular results. Nonsampling errors are generally not provided due to the difficulty in measuring them solely from a sample. The sampling error for an estimate is often quantified by a standard error or a confidence interval that is associated with each of the results. The sampling error is mainly determined by three factors: sample size, sample

design, and the weighting effect. Although population size also plays a role, compared with sample size, its effect on sampling error is minor. The weighting effect is measured by the coefficient of variation of weights, which equals standard deviation of weights divided by the mean of the weights.

For SRS with replacement, many textbooks provide the formula for estimating the variance of a mean or proportion estimator. For more complex sampling designs, it is important to ensure that a method appropriate for the design be used for variance estimation; these estimates must reflect the features resulting from the complex design. Variance estimation often requires more complex replication methods, rather than the exclusive use of formulas. Even though a variance formula is available for a Horvitz–Thompson estimator for the cluster sampling design without replacement, there will always be information needed for computation that is missing or unknown (namely, joint probabilities of selection, weights, stratum, and PSU codes). Popular replication methods for a complex sampling design include bootstrap repeated replication, jackknife repeated replication (JRR), and balanced-repeated replication (BRR); the Taylor-series method can also be used in variance estimation, although it is based on linear approximation. In an application, the replicate samples are also subject to the adjustment for the impact of nonresponse, poststratification, and raking. When replication methods are used, the number of replicates should be large enough to enable stable variance estimation (e.g., >30) but small enough for efficient calculation.

The JRR approach, used in NAEP, provides a good example of how replication methods work in practice. Ordinarily, the JRR approach drops one case in each replicate, but in the analysis of data in large-scale assessments such as NAEP, the JRR approach often drops a group of k cases in each replicate (Shao and Wu, 1989). The delete- k JRR approach provides a balance between the number of replicates and the sizes of clusters. This flexibility enables testing programs to comply with diverse requirements from varied studies and variations in sample designs. The computation of the delete- k JRR approach mainly consists of two steps: calculating replicate estimates and estimating the jackknifed variance. The replicate estimates, also known as the pseudo-values of the estimate, are calculated from the replicate samples that are formed by dropping one group of k cases each time from the sample. Let \mathcal{J} be the number of replicates. Let \bar{y} be the mean estimate and $\bar{y}_{(j)}$ be the j th replicate estimate. Then, the jackknife variance of \bar{y} is estimated by $\mathcal{J} - 1$ times the sample variance of the replicate estimates as (Wolter, 2003)

$$\nu_{\mathcal{J}}(\bar{y}) = \frac{\mathcal{J} - 1}{\mathcal{J}} \sum_{j=1}^{\mathcal{J}} (\bar{y}_{(j)} - \bar{y})^2$$

The preferred way to derive appropriate variance estimates is to use a software package that uses statistical techniques such as the Taylor-series extension or the replicate methods mentioned above that do not assume SRS and independently distributed errors. The accessibility information of these packages, such as SUDAAN, WesVar, DAS, and Stata, can be found on the Web.

An estimator based on a complex design will often have a larger variance than the corresponding estimator of an SRS sample of the same size; the ratio of the variance from the complex design over variance from the SRS is defined as the design effect (Kish, 1965). Because of the clustering effect and the variation in weights, the design effect is usually larger than one.

In some situations, advanced models, such as multi-level models, can be used to produce variance estimates that take into account some aspects of complex survey design (Raudenbush *et al.*, 2003). Any design variables and weights, such as those associated with strata or measures of size, should also be taken into account during the modeling process.

An Application of Sampling Techniques in Computerized Adaptive Testing

The application of sampling techniques in item-selection procedures for computerized adaptive testing (CAT; Chang and Ying, 1999) is a recent development. One basic sample design, *a*-stratified multistage CAT with *b*-blocking, partitions the item pool according to the *a* parameters (item-discrimination parameters) of the three-parameter logistic model and requires *b* parameters (item-difficulty parameters) to be evenly distributed across all strata (Chang *et al.*, 2001). A multi-phase sampling approach is applied in item selection. The items with lower *a* parameters are designed to be selected in the early phases of a test, and items with higher *a* parameters are selected in later phases. The CAT sampling approach allows the implementation of different types of constraints on item selection, such as constraints used to balancing content, constraints on lowering item-exposure rate, and constraints that allow application of the shadow-test approach.

See also: Generating Random Numbers; Jackknife Methods; Missing Data; Probability Theory; Small Area Estimation.

Bibliography

- Allen, N., Donoghue, J., and Schoeps, T. (eds.) (2001). *The NAEP 1998 Technical Report*. (NCES 2001-509). Washington, DC: National Center for Education Statistics.
- Chang, H. H. and Ying, Z. (1999). *a*-Stratified multistage computerized adaptive testing. *Applied Psychological Measurement* **23**, 211–222.
- Chang, H., Qian, J., and Ying, Z. (2001). *a*-Stratified multistage computerized adaptive testing with *b* blocking. *Applied Psychological Measurement* **25**(4), 333–341.
- Cochran, W. G. (1977). *Sampling Techniques*, 3rd edn. New York: Wiley.
- Deming, W. E. and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal tables are known. *Annals of Mathematical Statistics* **11**, 427–444.
- Haberman, S. J. (1979). *Analysis of Qualitative Data*, vol. 2. New York: Academic Press.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663–685.
- Kish, L. (1965). *Survey Sampling*. New York: Wiley.
- Kish, L. (1987). *Statistical Design for Research*. New York: Wiley.
- Kish, L. (1992). Weighting for unequal Pi. *Journal of Official Statistics* **8**, 183–200.
- National Center for Education Statistics (2003). *NCES Statistical Standards*. (NCES 2003-601). Washington, DC: National Center for Education Statistics.
- Qian, J. (2008). *Model-Based Weighting and Comparisons*, ETS Research Rep. No. RR-08-17. Princeton, NJ: ETS.
- Raudenbush, S. W., Johnson, C., and Sampson, R. J. (2003). Multivariate, multilevel Rasch model with application to self-reported criminal behavior. *Journal of Sociological Methodology* **33**(1), 169–211.
- Rust, K. F. (2004). Sampling and field operations at Westat, 1983 to 2001. In Jones, L. V. and Olkin, I. (eds.) *The Nation's Report Card: Evolution and Perspectives*, pp 427–448. Bloomington, IN: Phi Delta Kappa International.
- Rust, K. F., Krenzke, T., Qian, J., and Johnson, E. G. (2001). Sample design for the state assessment. In Allen, N. J., Donoghue, J. R., and Schoeps, T. L. (eds.) *The NAEP 1998 Technical Report*, NCES 2001-509, pp 61–78. Washington, DC: National Center for Education Statistics.
- Shao, J. and Wu, C. F. J. (1989). A general theory for jackknife variance estimation. *The Annals of Statistics* **17**, 1176–1197.
- Tourangeau, R., Johnson, R. A., Qian, J., and Shin, H. C. (1993). *Selection of NORC's 1990 National Sample*. Working Rep. No. 307. Chicago, MI: National Opinion Research Center.
- Wolter, K. (2003). *Introduction to Variance Estimation*. New York: Springer.

Sequential Probability Ratio Test

T J H M Eggen, Cito, Psychometric Research Center, Arnhem, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

The sequential probability ratio test (SPRT) was developed more than 60 years ago by Wald (1947) for quality-control problems. It is a statistical procedure in which a choice is made between two simple hypotheses and was initially used to determine whether majority of products (e.g., 80%) in a production process meet specifications or if this was true in fewer cases (e.g., 50%). The statistical properties of the SPRT have since been well established. Various extensions of the original testing procedure have been proposed in the literature and their statistical optimality properties have been studied extensively (see, e.g., Ghosh and Sen, 1991). In this article we will not go into detail regarding the statistical properties of the SPRT, but will focus on to some successful applications of it in educational testing.

In education, practitioners make use of test results for many different purposes, but from an educational measurement point of view, it generally suffices to distinguish between two main aims of testing: the precise estimation of a person's ability in a certain domain or the classification of a person in one of a limited number of proficiency classes. For the latter purpose, the SPRT has often been very successfully applied. In what is called sequential mastery testing (Lewis and Sheenan, 1990), Ferguson (1969) used a basic application of the SPRT to decide whether a student is a master or nonmaster in a certain domain. Since then, starting with Reckase (1983), many algorithms for computerized adaptive testing (CAT) have been developed (e.g., Eggen, 1999) using the SPRT methodology as their basis.

This article discusses the use of SPRT in CAT. After describing the SPRT and the basic elements of CAT, the way the SPRT is used in CAT will be treated. The situation considered is how, on the basis of a test, can be decided whether or not a certain criterion or standard is met. Lastly, this one cutting-point situation is extended to more than one, together with some other extensions.

The Sequential Probability Ratio Test

In sequential testing it is not only the observations X , that are random variables, but also the number of observations, K . Inspired by the Neyman and Pearson (1933) lemma, which provides a method of constructing a most powerful statistical test for deciding between two simple hypotheses, Wald (1947) proposed the SPRT. In his treatment of the problem, Wald considered random variables X with two

possible values: $x=1$ if a product meets the criteria and $x=0$ otherwise. Next, two statistical hypotheses are formulated: the null hypothesis, $H_0: p = p_0$, and the alternative $H_1: p = p_1$ in which p is the unknown proportion of all products meeting the criteria.

If we denote a series of k observations by $\underline{X}_k = (X_1, \dots, X_k)$ and the probability distribution of X_i with $P(X_i = x_i; p) = p^{x_i}(1-p)^{1-x_i}$, then the probability of these k observations is given by

$$P_{0k} = \prod_{i=1}^k p_0^{x_i} (1-p_0)^{1-x_i} \text{ if } H_0 \text{ is true}$$

and

$$P_{1k} = \prod_{i=1}^k p_1^{x_i} (1-p_1)^{1-x_i} \text{ if } H_1 \text{ is true}$$

The SPRT then chooses two constants A and B with $A < B$ and, after making every observation, computes the ratio of the probabilities P_{0k}/P_{1k} and the decision is made as follows:

1. if $P_{1k}/P_{0k} \geq A$, then reject H_0 ;
2. if $P_{1k}/P_{0k} \leq B$, then accept H_0 ; and
3. if $B < P_{1k}/P_{0k} < A$ (the critical inequality of the procedure) then take another observation.

Intuitively, the procedure is: if the outcomes have much larger probability under H_1 than under H_0 , that is, the likelihood ratio is large, then reject H_0 ; if the ratio is small, accept H_0 ; and if the ratio has values within the critical interval, no decision is taken and sampling is continued. The constants A and B are dependent on the size of the acceptable decision errors.

In practice, the log likelihood ratio is evaluated. This ratio is equal to the sum over i of the terms

$$Z_i = \ln \left[\frac{p_1^{x_i} (1-p_1)^{1-x_i}}{p_0^{x_i} (1-p_0)^{1-x_i}} \right] \quad [1]$$

If the acceptable decision errors are specified by

$$P(\text{reject } H_0 | H_0 \text{ is true}) \leq \alpha \text{ and } P(\text{accept } H_0 | H_1 \text{ is true}) \leq \beta \quad (\alpha, \beta \text{ small constants}), \quad [2]$$

then the SPRT procedure is

1. if $\ln B < \sum_{i=1}^k Z_i < \ln A$: take another observation
2. if $\sum_{i=1}^k Z_i \geq \ln A$: reject H_0 [3]

3. if $\sum_{i=1}^k Z_i \leq \ln B$: accept H_0 .

Wald (1947) has shown that the decision error rates are met if

$$A = \frac{1 - \beta}{\alpha} \text{ and } B = \frac{\beta}{1 - \alpha} \quad [4]$$

Furthermore, this SPRT procedure will then stop, with probability 1, with a decision in a finite number of observations.

The SPRT was initially developed for situations in which there is a random sample of a variable with a discrete or continuous distribution with one parameter variable and two simple hypotheses on the value of that parameter. But the theory is generalized in various directions. For our purposes two more general situations are important:

1. Although the SPRT stops with a finite number of observations with probability 1, in educational measurement, it is absolutely necessary to define a maximum test length k_{\max} . The procedure is then called the truncated SPRT (TSPRT).
2. The observations are not random draws from the same distribution, but are independent variables from not necessarily the same distribution.

In the first application by Ferguson (1969), a maximum length was already specified, but the answers on the items were assumed to come from the same binomial distribution, which implies that all items have the same difficulty. In the CAT application, described next, this is not the case.

Computerized Adaptive Testing

In computerized adaptive tests, the construction and administration of the test is computerized and individualized. A different test is constructed for every testee by selecting items from an item bank tailored to the ability of the testee as demonstrated by the responses given thus far. Computerized adaptive tests assume the availability of an item bank, which is calibrated with an item response model. Confining ourselves to item banks with items which are dichotomously scored, logistic item response models are commonly used. In item response theory (IRT), a relation is specified between the non-observable ability θ that is to be estimated and the probability of correctly answering item i . The exact relationship is determined by the parameters of the items. A commonly used IRT model is the two-parameter logistic model (2PL):

$$p_i(\theta) = P(X_i = 1 | \theta) = \frac{\exp(a_i(\theta - b_i))}{1 + \exp(a_i(\theta - b_i))} \quad [5]$$

in which b_i is the location or difficulty parameter and a_i the discrimination parameter.

In computerized adaptive tests the parameters of the IRT model are always assumed to be estimated with such precision that they can be considered to be known. In a computerized adaptive test, the likelihood function of a student's ability, θ , plays a central role in the inference on the student. Given the scores on k items $x_i, i = 1, \dots, k$, this function is given by

$$L(\theta, \underline{x}_k) = L(\theta; x_1, \dots, x_k) = \prod_{i=1}^k p_i(\theta)^{x_i} (1 - p_i(\theta))^{1-x_i} \quad [6]$$

which states the probability of getting the observed scores on the items as a function of θ .

In computerized adaptive tests where the main aim is the efficient estimation of the ability θ of an examinee, this likelihood function [6] is the basis for estimating the ability of an examinee as well as for the selection of items. The maximum likelihood (ML) estimate of the ability after administering k items follows from the maximization [6] with respect to θ . Because of less bias, the weighted maximum likelihood (WML) method proposed by Warm (1989) is a good alternative for ML. WML follows, in the case of the 2PL model [5], from

$$\hat{\theta} = \max_{\theta} \left[\left(\sum_{i=1}^k I_i(\theta) \right)^{1/2} L(\theta; \underline{x}_k) \right] \quad [7]$$

where $I_i(\theta)$ is the Fisher information function of item i , which is defined as the (statistical) expectation of the squared relative change of the likelihood function

$$I_i(\theta) = E \left(\frac{\partial L(\theta; x_i) / \partial \theta}{L(\theta; x_i)} \right)^2$$

which, in the 2PL model is given by

$$I_i(\theta) = a_i^2 p_i(\theta) (1 - p_i(\theta)) = \frac{a_i^2 \exp(a_i(\theta - b_i))}{(1 + \exp(a_i(\theta - b_i)))^2}$$

This information function is commonly used for item selection in computerized adaptive tests : an item is selected if it gives maximum information at the current ability estimate. This method ensures that each examinee is administered items which fit his ability and, consequently, his ability is estimated efficiently.

The SPRT in CAT

When classification in one of two categories is the purpose of testing in CAT, SPRT can be applied as follows. On the latent ability scale, a decision or cutting point θ_0 is given which distinguishes between, for example, a master and nonmaster, or between an examinee who passes and an examinee who fails an exam. A small region on both sides of this point, a so-called indifference zone, is selected. The widths of these regions are δ_1 and δ_2 .

The indifference interval expresses the fact that, owing to measurement errors, making the right decision about examinees very near the cutting point can never be guaranteed. One could also say that the interval expresses the indifference of an examiner to the classification of the examinees who are very near to the cutting point. Next, the statistical hypotheses are formulated:

$$H_0: \theta \leq \theta_0 - \delta_1 = \theta_1 \text{ against } H_1: \theta \geq \theta_0 + \delta_2 = \theta_2 \quad [8]$$

If the acceptable decision error rates are specified as in [2], the test meeting these decision error rates uses as the test statistic, as mentioned above, the ratio of the likelihood function under H_1 and H_0 :

$$LR_k(\theta_2; \theta_1) = \frac{L(\theta_2; x_1, \dots, x_k)}{L(\theta_1; x_1, \dots, x_k)} \quad [9]$$

and involves the following procedure:

If Decision

$\beta/(1 - \alpha) < LR_k(\theta_2; \theta_1) < (1 - \beta)/\alpha$ administer another item

$LR_k(\theta_2; \theta_1) \leq \beta/(1 - \alpha)$ accept H_0

$LR_k(\theta_2; \theta_1) \geq (1 - \beta)/\alpha$ reject H_0 [10]

It can easily be shown (Eggen and Straetmans, 2000) that if the 2-PL model [5] is used, the critical inequality of this test can be written as

$$\frac{\ln\left(\frac{\beta}{(1-\alpha)}\right) - C}{\theta_2 - \theta_1} < \sum_{i=1}^k a_i x_i < \frac{\ln\left(\frac{(1-\beta)}{\alpha}\right) - C}{\theta_2 - \theta_1} \quad [11]$$

In this

$$\begin{aligned} C &= \sum_{i=1}^k \ln\left(\frac{1 + \exp(a_i(\theta_1 - \beta_i))}{1 + \exp(a_i(\theta_2 - \beta_i))}\right) \\ &= \sum_{i=1}^k \ln\left(\frac{1 - p_i(\theta_2)}{1 - p_i(\theta_1)}\right) \end{aligned} \quad [12]$$

which depends only on the parameters of the items on θ_1 , θ_2 , which are all known constants in the procedure. This makes clear that the application of the SPRT is easy because the observed weighted score is compared to known constants.

Example

The following example of a simulation study with an operational item bank (Eggen, 1999) illustrates the performance

of the SPRT in CAT. This item bank contains 250 mathematics items which are used in adult education. Most of the items have an open-ended short answer format, but all the items are scored dichotomously. The items were shown to fit the one-dimensional 2-PL model. The scale was fixed by restrictions on the item parameters. The mean item difficulty is 0, and the geometric mean of the discrimination parameters is 3.09. On this scale, the distribution of the ability in the population was estimated to be normal with a mean of 0.294 and a standard deviation of 0.522.

The simulations were conducted as follows. An ability of a simulee was randomly drawn from $N(0.294; 0.522)$. Three relatively easy starting items were selected and subsequent items were selected with the criterion of maximum Fisher information at the current ability estimate. The simulee's response to an item was generated according to the IRT model and this procedure was repeated for $N = 5000$ simulees.

For varying acceptable decision error rates and widths of the indifference zone (δ is respectively 0.2, 0.3, and 0.4 times the standard deviation of θ), the performance of the procedure was evaluated with the mean number of items required to make a decision \bar{k} and the classification accuracy expressed in the percentages of correct decisions (%cor).

The cutting point on the ability scale in the simulations was $\theta_0 = 0.1$, and the maximum test length was $K_{\max} = 25$. The SPRT adaptive testing procedures were conducted for three different error rates and three different widths of the indifference zone.

As a benchmark, the SPRT procedure was compared to a CAT procedure based on statistical estimation. The procedure proposed by Weiss and Kingsbury (1984), but using the Warm estimate of the ability [7], was conducted. After each item is administered, an estimate is made of the examinee's ability $\hat{\theta}_k$ and of its standard error $se(\hat{\theta}_k)$. Next, a confidence interval $(\hat{\theta}_k - \gamma \cdot se(\hat{\theta}_k), \hat{\theta}_k + \gamma \cdot se(\hat{\theta}_k))$ for the examinee's true ability is constructed, in which γ is a constant that is determined by the required accuracy. The procedure delivers another item as long as the cutting point $\theta_0 = 0.1$ is within the interval. If not, the appropriate decision is made. In the comparison, the value for γ was chosen such that about the same accuracy is reached as with the acceptable decision errors in the SPRTs. The results are given in **Table 1**.

Table 1 Mean number of required items (\bar{k}) of percentage of correct decisions (%cor) in a problem with one cutting point $\theta_0 = 0.1$

	SPRT						Estimation	
	$\delta = 0.11$		$\delta = 0.16$		$\delta = 0.21$		\bar{k}	%cor
Error rate	\bar{k}	%cor	\bar{k}	%cor	\bar{k}	%cor	\bar{k}	%cor
$\alpha = \beta = 0.05$	23.24	95.76	15.37	95.16	11.37	95.20	15.41	94.58
$\alpha = \beta = 0.075$	20.35	95.32	13.54	95.02	9.89	94.70	13.40	94.56
$\alpha = \beta = 0.1$	18.72	95.60	12.51	94.90	9.10	93.94	12.97	94.46

The results show that applying the SPRT instead of a traditional estimation procedure possibly improves the performance of the CAT. A striking result is that the chosen acceptable decision error rates and also the chosen width of the indifference zone hardly influence the percentages of correct decisions, but have a major influence on the average number of items needed for taking this decision. It is clear that the number of items needed is larger with lower allowed error rates and with wider indifference zones. The small differences in the percentage correct decisions are always in the expected direction.

Some Extensions of the Application of SPRT

The SPRT procedure described above can also be applied to other IRT models and used for polytomously scored items (Lau and Wang, 1998). Next, three other extensions of the SPRT application will be addressed.

The SPRT in a Three-Category Classification Problem

The Cat application of the SPRT is readily generalized to the case in which there are more than two decision categories. Following Eggen (1999), the generalization to three categories will be described. In this case, there are two cutting points, θ_1 and θ_2 , by which three levels of ability are distinguished. An indifference zone is identified around each cutting point. This is sketched schematically in **Figure 1**.

Two pairs of hypotheses are formulated:

$H0_1: \theta \leq \theta_1 - \delta_{11} = \theta_{11}$ against $H1_1: \theta \geq \theta_1 + \delta_{12} = \theta_{12}$
 $H0_2: \theta \leq \theta_2 - \delta_{21} - \delta_{22}$ against $H1_2: \theta \geq \theta_2 + \delta_{22} = \theta_{22}$

The SPRT described in [10] is applied for each pair of hypotheses. The specifications of the acceptable decision errors are α_1, β_1 and α_2, β_2 , as defined in [2]. Next, the two SPRTs are combined in one procedure. The decisions to assign a person to a certain category are given in **Table 2**.

This combination procedure of the SPRTs originates from Sobel and Wald (1949). It can be shown that, by using the 2-PL IRT model [5] or any other model

belonging to the exponential family, the simultaneous acceptance of the null hypothesis $H0_1$, and the rejection of $H0_2$ cannot occur. It is noted that Spray (1993) proposed extensions of the use of the SPRT for classification in three and more categories. Her generalization is based on the combination procedure developed by Armitage (1950) which uses the simultaneous application of three SPRTs for classification into three categories instead of only the two needed in the Sobel and Wald (1949) combination procedure proposed here. It is beyond the scope of this article to discuss in detail the properties of these two combination procedures of SPRTs.

In the practical applications, the combined procedure operates as follows:

if
 Decision

$$\text{Level 1} \quad \sum_{i=1}^k a_i x_i < \frac{\ln\left(\frac{\beta_1}{(1-\alpha_1)}\right) - C}{\theta_{12} - \theta_{11}}$$

$$\text{Level 2} \quad \frac{\ln\left(\frac{\beta_2}{(1-\alpha_2)}\right) - C}{\theta_{22} - \theta_{21}} < \sum_{i=1}^k a_i x_i < \frac{\ln\left(\frac{(1-\beta_1)}{\alpha_1}\right) - C}{\theta_{12} - \theta_{21}}$$

$$\text{Level 3} \quad \sum_{i=1}^k a_i x_i < \frac{\ln\left(\frac{(1-\beta_2)}{\alpha_2}\right) - C}{\theta_{22} - \theta_{21}}$$

In all other cases Continue testing

In this procedure C is as in [12] with the appropriate corresponding constants filled in. From this it can be easily seen that if the width of the indifference intervals, for example, $\theta_{22} - \theta_{21} = \delta_{12} + \delta_{22}$, increases, a shorter test can probably be used to take a decision.

Stochastic Curtailment of the TSPRT

Finkelman (2004, 2008) recently introduced the application of stochastic curtailment to enhance the performance of the TSPRT in educational testing. The idea of this method is to stop testing sooner without losing accuracy. The method of curtailment determines whether further

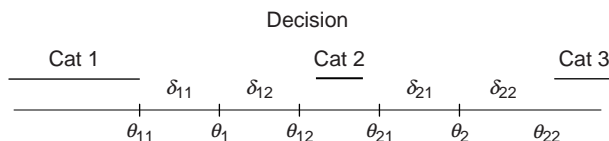


Figure 1 Schematic representation of the classification problem with three categories.

Table 2 Decisions based on combination of two SPRTs

	<i>Decision on test 1 $H0_1: \theta \leq \theta_{11}$ against $H1_1: \theta \geq \theta_{12}$</i>	
<i>Decision on test 2 $H0_2: \theta \leq \theta_{21}$ against $H1_2:$ $\theta \geq \theta_{22}$</i>	Accept $H0_1$	Reject $H0_1$
Accept $H0_2$	Category 1	Category 2
Reject $H0_2$		Category 3

testing will possibly change a classification decision which would be taken if testing were stopped directly. Stochastic curtailment (Lan *et al.*, 1982) also extends the observation to the case in which a change in decision is possible but unlikely. In an example, Finkelman (2004) showed that in the case of one cutting point, a reduction of 20% in the number of items needed, while keeping the same accuracy, can be reached by applying stochastic curtailment.

Conclusion

The SPRT has been shown to be a very useful statistical procedure. In this article the applications in the context of educational testing were explored. If classification in a limited number of categories is the main goal of computerized adaptive testing, the (combination of more) SPRT gives very efficient algorithms.

Bibliography

- Armitage, P. (1950). Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis. *Journal of the Royal Statistical Society, B* **12**, 137–144.
- Eggen, T. J. H. M. (1999). Item selection with adaptive testing with the sequential probability ratio test. *Applied Psychological Measurement* **23**, 249–261.
- Eggen, T. J. H. M. and Straetmans, G. J. J. M. (2000). Computerized adaptive testing for classifying examinees into three categories. *Educational and Psychological Measurement* **66**, 713–734.
- Ferguson, R. L. (1969). *The Development, Implementation, and Evaluation of a Computer-assisted Branched Test for a Program of Individually Prescribed Instruction*. Unpublished doctoral dissertation, University of Pittsburgh: Pittsburgh PA.
- Finkelman, M. (2004). *Statistical Issues in Computerized Adaptive Testing*. Unpublished doctoral dissertation, Stanford University: California.
- Finkelman, M. (2008). On using stochastic curtailment to shorten the SPRT in sequential mastery testing. *Journal of Educational and Behavioral Statistics* **33**, 442–463.
- Ghosh, B. K. and Sen, P. K. (1991). *Handbook of Sequential Analysis*. New York: Marcel Dekker.
- Lan, K. K. G., Simon, R., and Halperin, M. (1982). Stochastically curtailed tests in long-term clinical trials. *Communications in Statistics-Sequential Analysis* **1**, 207–219.
- Lau, C. A. and Wang, T. (1998). Comparing and combining dichotomous and polytomous items with SPRT procedure in computerized classification testing. *Paper AERA, San Diego, April 1998*.
- Lewis, C. and Sheenan, K. (1990). Using Bayesian decision theory to design a computerized mastery test. *Applied Psychological Measurement* **14**, 376–386.
- Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A* **231**, 289–337.
- Reckase, M. D. (1983). A procedure for decision making using tailored testing. In Weiss, D. J. (ed.) *New Horizons in Testing*, pp 237–255. New York: Academic Press.
- Sobel, M. and Wald, A. (1949). A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution. *Annals of Mathematical Statistics* **20**, 502–522.
- Spray, J. A. (1993). Multiple-category classification using a sequential probability ratio test. *Research Report 93-7*. Iowa City, IA: American College Testing.
- Wald, A. (1947). *Sequential Analysis*. New York: Wiley.
- Warm, T. A. (1989). Weighted maximum likelihood estimation of ability in item response theory. *Psychometrika* **54**, 427–450.
- Weiss, D. J. and Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement* **21**, 361–375.

Further Reading

- Eggen, T. J. H. M. (2004). *Contributions to the Theory of Practice of Computerized Adaptive Testing*. Doctoral Thesis. Enschede: University of Twente.
- Finkelman, M. (2004). *Statistical Issues in Computerized Adaptive Testing*. Unpublished Doctoral Dissertation. Stanford University: California.
- Ghosh, B. K. and Sen, P. K. (1991). *Handbook of Sequential Analysis*. New York: Marcel Dekker.
- Van der Linden, W. J. and Glas, C. A. W. (eds.) (2000). *Computerized Adaptive Testing. Theory and Practice*. Dordrecht: Kluwer.
- Wainer, H. (ed.) (2000). *Computerized Adaptive Testing. A Primer*, 2nd edn. Hillsdale, NJ: Erlbaum.
- Wald, A. (1947). *Sequential Analysis*. New York: Wiley.

Sequential Testing

H J Vos, University of Twente, Enschede, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In a sequential mastery test (SMT), the decision of interest is to classify a student as a master or a nonmaster, or to continue testing and administering another random item. The main advantage of SMT is that it provides shorter tests for students who have clearly attained a certain level of mastery (or, clearly, nonmastery) and longer tests for those students for whom the mastery decision is not as clear cut. One of the earliest approaches to SMT dates back to Ferguson (1969) using Wald's well-known sequential probability ratio test (SPRT), originally developed as a statistical quality control for light-bulbs in manufacturing setting. In Ferguson's (1969) approach, the probability of a correct response, given the student's (latent) ability is modeled as a binomial distribution. Modeling response behavior by this distribution function assumes that – given the student's ability – each item has the same probability of being correctly answered, or that items are sampled at random. It should be noted that this assumption is not realistic, because items are, in fact, mostly of different difficulty, implying that the same probability of a correct answer is not realistic as well. Reckase (1983) has proposed an alternative approach to SMT within an SPRT framework. Unlike Ferguson (1969), Reckase (1983) did not assume that items have equal characteristics but allowed them to vary in difficulty and discrimination by using an item response theory (IRT) model instead of a binomial model.

As indicated by Ferguson (1969), three elements must be specified in advance in applying the SPRT framework to sequential mastery testing. First, two values p_0 and p_1 on the proportion-correct metric must be specified, representing points that correspond to lower and upper limits of levels of ability at which a mastery and nonmastery decision will be made, respectively. In addition, these two values mark the boundaries of the small region (i.e., indifference region) where we can never be sure to take the right classification decision, and, thus, in which sampling of another item will continue. Second, two nominal error probabilities α and β must be specified – reflecting the relative costs of the false positive (i.e., type I) and false negative (i.e., type II) error types. Wald (1947) showed that the actual error rates observed in practice, α^* and β^* , are bounded from above by functions of the nominal rates; that is, $\alpha^* \leq \alpha/(1-\beta)$ and $\beta^* \leq \beta/(1-\alpha)$. Intervals can be derived as functions of these two error rates for which mastery or nonmastery is declared, respectively, and for

which testing is continued (Wald, 1947) – that is, mastery is declared if the likelihood ratio $L(\text{response pattern} | p_1)/L(\text{response pattern} | p_0)$ is larger than $(1-\beta)/\alpha$, nonmastery if this likelihood ratio is smaller than $\beta/(1-\alpha)$; otherwise sampling is continued. Third, a maximum test length must be specified in order to classify, within a reasonable period of time, those students for whom the decision of declaring mastery or nonmastery is not as clear cut. It should be emphasized, however, that the SPRT framework might also be applied to SMT if no maximum test length is specified.

A major drawback of the SPRT framework is that it does not take costs of testing explicitly into account. The purpose of this article, therefore, is to derive optimal rules for the SMT that takes costs of testing explicitly into account. Decision rules are hereby prescribed for each possible observed response pattern, that is, what action (i.e., declaring mastery/nonmastery or continuing testing and administering another random item) has to be taken. The framework of Bayesian sequential decision theory will be used (e.g., DeGroot, 1970; Lehmann, 1959).

Bayesian Sequential Principle Applied to SMT

This section indicates how the framework of Bayesian sequential decision theory – in combination with the binomial distribution for modeling response behavior (i.e., the measurement model) and adopting threshold loss for the loss function involved – is applied to SMT.

Framework of Bayesian Sequential Decision Theory

Three basic elements can be identified in Bayesian sequential decision theory. In addition to a measurement model and a loss function, costs of administering one additional item must be explicitly specified in this approach. By doing so, posterior expected losses corresponding to the mastery and nonmastery decisions can now be calculated directly at each stage of testing. As far as the posterior expected loss corresponding to continuous testing is concerned, this quantity is determined by weighting the posterior expected losses corresponding to each of the possible future classification outcomes with the probability of observing those outcomes (i.e., the posterior predictive distributions).

Optimal rules (i.e., Bayesian sequential rules) are then obtained by minimizing the posterior expected losses associated with all possible decision rules at each stage of testing, using techniques of backward induction (i.e., dynamic programming). This technique starts by considering the final stage of testing (where the option to continue testing is not available) and then works backward to the first stage of testing.

Notation

Within the framework of Bayesian sequential decision theory, optimal rules can be obtained without specifying a maximum number of items to be administered. In order to classify, within a reasonable period of time, those students for whom the decision of declaring mastery or nonmastery is not as clear cut, however, an SMT is supposed to have a maximum length of n ($n \geq 1$). Let the observed item response at each stage of testing k ($1 \leq k \leq n$) for a randomly sampled student be denoted by a discrete random variable X_k with realization x_k . The observed response variables X_1, \dots, X_k are assumed to be independent and identically distributed. Further, let the values 0 and 1, respectively, denote the incorrect and correct responses to an item. Furthermore, let the observed number-correct score following the k items that have been administered be denoted by a discrete random variable $S_k = X_1 + \dots + X_k$ with realization $s_k = x_1 + \dots + x_k$ ($0 \leq s_k \leq k$).

The student's true level of functioning is unknown due to measurement and sampling error. All that is known is his/her observed number-correct score s_k . In other words, the mastery test is not a perfect indicator of the student's true performance. Therefore, let the student's (unknown) true level of functioning be denoted by a continuous random variable T on the latent proportion-correct metric, with realization t ($0 \leq t \leq 1$).

Finally, a criterion level t_c ($0 \leq t_c \leq 1$) on T must be specified in advance by the decision-maker using methods of standard-setting. A student is considered a true non-master and true master if his/her true level of functioning t is smaller or larger than t_c , respectively.

Threshold Loss and Costs of Testing

Generally speaking – as noted before – a loss function evaluates the total costs and benefits of all possible decision outcomes for a student whose true level of functioning is t . These costs may concern all relevant psychological, social, and economic consequences (e.g., remedial instruction) which the decision brings along. The Bayesian sequential approach allows the decision-maker to incorporate, into the decision process, the costs of misclassification. In this case, the well-known threshold

Table 1 Table for threshold loss function at stage k ($1 \leq k \leq n$) of testing

True level of functioning	Decision	$T \leq t_c$	$T > t_c$
Declaring nonmastery		ke	$l_{01} + ke$
Declaring mastery		$l_{10} + ke$	ke

loss function is adopted as the loss structure involved. The choice of this loss function implies that the total costs and benefits of all possible consequences of the decisions can be summarized by possibly different constants – one for each of the possible classification outcomes.

An obvious disadvantage of the threshold loss function is that, as can be seen from **Table 1**, it assumes the same constant loss for students to the left or to the right of the criterion level t_c – no matter how large their distance from t_c . For instance, a student who is declared a non-master with a true level of functioning just above t_c gives the same loss as a misclassified true nonmaster with true level of functioning far above t_c . It seems more realistic to suppose that for the misclassified true nonmasters the loss is a strictly increasing function of t . Moreover, the threshold loss is discontinuous; at point $T = t_c$ this function jumps from one constant value to another. This sudden change seems unrealistic in many decision-making situations. In the neighborhood of this point, the losses for correct and incorrect classification decisions should change smoothly rather than abruptly. To overcome these shortcomings, van der Linden and Mellenbergh (1977) proposed a continuous loss function which is a linear function of the true level of functioning t . Although a linear loss function is probably more appropriate for the sequential mastery problem, following Lewis and Sheehan (1990), in the present article, a threshold loss function is adopted for reasons of simplicity and computational efficiency. Another reason for using threshold rather than linear loss is that a linear loss function may be more appropriate in the neighborhood of t_c , indeed, but that the further away from t_c , however, the losses can be assumed to take, more and more, the same constant values again for the possible classification outcomes.

For the sequential mastery problem – following Lewis and Sheehan (1990) – a threshold loss function can be formulated as a natural extension of the one for the fixed-length mastery problem at each stage of testing k as follows:

The value e represents the costs of administering one random item. For the sake of simplicity, these costs are assumed to be equal for each classification outcome as well as for each testing occasion. It has been shown in the literature (e.g., Luce and Raiffa, 1957) that positive linear transformations are allowed to loss functions, that is, multiplying with a positive constant and adding a constant to the loss function are so-called admissible transformations. Applying such an admissible positive linear transformation,

and assuming the losses l_{00} and l_{11} associated with the correct classification outcomes are equal and take the smallest values, the threshold loss function in **Table 1** was rescaled in such a way that l_{00} and l_{11} were equal to zero. Hence, the losses l_{01} and l_{10} must take positive values.

The ratio l_{10}/l_{01} is denoted as the loss ratio R , and refers to the relative losses for declaring mastery to a student whose true level of functioning is below t_c (i.e., false positive) and declaring nonmastery to a student whose true level of functioning exceeds t_c (i.e., false negative).

The loss parameters l_{ij} ($i, j = 0, 1$; $i \neq j$) associated with the incorrect decisions have to be empirically assessed; several methods have been proposed for this in the literature. Most texts on decision theory, however, propose lottery methods (e.g., Luce and Raiffa, 1957) for assessing loss functions empirically. In general, the consequences of each pair of actions and true level of functioning are scaled in these methods by looking at the most and least preferred outcomes. However, in principle, any psychological scaling method can be used.

Measurement Model

Following Ferguson (1969), in the present article the well-known binomial model will be adopted for the probability that, after k items have been administered, s_k of them have been answered correctly. Its distribution at stage k of testing for a given student's true level of functioning t , $P(s_k | t)$, can be written as follows:

$$P(s_k | t) = \binom{k}{s_k} t^{s_k} (1 - t)^{k-s_k} \quad [1]$$

If each response is independent of the other, and if the student's probability of a correct answer remains constant, the distribution function of s_k – given true level of functioning t – is given by [1]. The binomial model assumes that the test given to each student is a random sample of items drawn from a large (real or imaginary) item pool (e.g., Wilcox, 1981). Therefore, for each student, a new random sample of items must be drawn in practical applications of the sequential mastery problem.

Optimizing Rules for the Sequential Mastery Problem

This section shows how optimal rules for SMT can be derived using the framework of Bayesian sequential decision theory. By doing so – given an observed item response vector (x_1, \dots, x_k) – first, the Bayesian principle will be applied to the fixed-length mastery problem by determining which of the posterior expected losses associated with the two classification decisions is the smallest. Next – applying the Bayesian principle again – optimal rules for the sequential mastery problem are derived at

each stage of testing k by comparing this quantity with the posterior expected loss associated with the option to continue testing.

Applying the Bayesian Principle to the Fixed-Length Mastery Problem

As noted above, the Bayesian decision rule for the fixed-length mastery problem can be found by minimizing the posterior expected losses associated with the two classification decisions of declaring mastery or nonmastery. In doing so, the posterior expected loss is taken with respect to the posterior distribution of T . It can easily be verified from **Table 1** and [1] that mastery is declared when the posterior expected loss corresponding to declaring mastery is smaller than the posterior expected loss corresponding to declaring nonmastery, or, equivalently, when s_k is such that

$$(l_{10} + ke)P(T \leq t_c | s_k) + (ke)P(T > t_c | s_k) < (ke)P(T \leq t_c | s_k) + (l_{01} + ke)P(T > t_c | s_k) \quad [2]$$

and that nonmastery is declared otherwise. Rearranging terms, it can easily be verified from [2] that mastery is declared when s_k is such that

$$P(T \leq t_c | s_k) < 1/(1 + R) \quad [3]$$

and that nonmastery is declared otherwise.

$P(T \leq t_c | s_k)$ denotes the probability of the student's true level of functioning being equal to or below the criterion level t_c – given his/her number-correct score s_k . In fact, this probability is one minus the cumulative posterior distribution of T (i.e., the decumulative posterior distribution of T). Assuming a beta prior for T , it follows from an application of Bayes' theorem that, under the assumed binomial model from [1], the posterior distribution of T will be a member of the beta family again (the conjugacy property, see, e.g., Lehmann, 1959). In fact, if the beta function $B(\alpha, \beta)$ with parameters α and β ($\alpha, \beta > 0$) is chosen as prior distribution (i.e., the natural conjugate of the binomial distribution) and the student's observed number-correct score is s_k from a test of length k , then the posterior distribution of T is $B(\alpha + s_k, k - s_k + \beta)$. Hence, assuming a beta prior for T , it follows from [3] that mastery is declared when s_k is such that

$$P[B(\alpha + s_k, k - s_k + \beta)] < 1/(1 + R) \quad [4]$$

and that nonmastery is declared otherwise.

The assumption, in this article, is that uniform distribution on the standard interval $[0, 1]$ is a noninformative prior, which results as a special case of the beta distribution $B(\alpha, \beta)$ for $\alpha = \beta = 1$. In other words, prior true level of functioning can take on all values between 0 and 1 with equal probability. It then follows immediately from [4] that mastery is declared when s_k is such that

$$P[B(1 + s_k, k - s_k + 1)] < 1/(1 + R) \quad [5]$$

and that nonmastery is declared otherwise. The beta distribution has been extensively tabulated (e.g., Pearson, 1930). Normal approximations are also available (Johnson and Kotz, 1970). In general, if t has a beta distribution with parameters (α, β) , then this distribution can be approximated by a normal distribution with mean $\alpha/(\alpha+\beta)$ and variance $\alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$.

Derivation of Bayesian Sequential Rules

Let $d_k(x_1, \dots, x_k)$ denote the decision rule yielding the minimum of the posterior expected losses associated with the two classification decisions at stage k of testing, and let the posterior expected loss corresponding to this minimum be denoted as $V_k(x_1, \dots, x_k)$. Bayesian sequential rules can now be found by using the following backward-induction computational scheme: First, the Bayesian sequential rule at the final stage n of testing is computed. Since the option to continue testing is not available at this stage of testing, it follows immediately that the Bayesian sequential rule is given by $d_n(x_1, \dots, x_n)$, and its corresponding posterior expected loss by $V_n(x_1, \dots, x_n)$.

To compute the posterior expected loss associated with the option to continue testing at stage $(n-1)$ until stage 0, the risk $R_k(x_1, \dots, x_k)$ will be introduced at each stage k ($1 \leq k \leq n$) of testing. Let the risk at stage n be defined as $V_n(x_1, \dots, x_n)$. Generally, given response pattern (x_1, \dots, x_k) , the risk at stage $(k-1)$ is then computed inductively as a function of the risk at stage k as:

$$R_{k-1}(x_1, \dots, x_{k-1}) = \min V_{k-1}(x_1, \dots, x_{k-1}), \quad E[R_k(x_1, \dots, x_{k-1}, X_k | x_1, \dots, x_{k-1})] \quad [6]$$

The posterior expected loss corresponding to administering one more random item following $(k-1)$ items have been administered, $E[R_k(x_1, \dots, x_{k-1}, X_k) | x_1, \dots, x_{k-1}]$, can then be computed as the expected risk at stage k of testing as

$$E[R_k(x_1, \dots, x_{k-1}, X_k | x_1, \dots, x_{k-1})] = \sum_{X_k=0}^{X_k=1} R_k(x_1, \dots, x_k) P(X_k = x_k | x_1, \dots, x_{k-1}) \quad [7]$$

where $P(X_k = x_k | x_1, \dots, x_{k-1})$ denotes the posterior predictive distribution of X_k at stage $(k-1)$ of testing. Computation of this conditional distribution is deferred until the next section. Note that [7] averages the posterior expected losses associated with each of the possible future classification outcomes with weights corresponding to the probabilities of observing those outcomes.

The Bayesian sequential rule at stage $(k-1)$ is now given by: Administer one more random item if $E[R_k(x_1, \dots, x_{k-1}, X_k) | x_1, \dots, x_{k-1}]$ is smaller than $V_{k-1}(x_1, \dots, x_{k-1})$; otherwise, decision $d_{k-1}(x_1, \dots, x_{k-1})$

is taken. The Bayesian sequential rule at stage 0 denotes the decision whether or not to administer at least one random item.

Computation of Posterior Predictive Distributions

As is clear from [7] the posterior predictive distribution $P(X_k = x_k | x_1, \dots, x_{k-1})$ is needed for computing the posterior expected loss corresponding to administering one more random item at stage $(k-1)$ of testing. Assuming the binomial distribution as measurement model and the uniform distribution $B(1,1)$ as prior, it was shown (e.g., Lehmann, 1959) that $P(X_k = 1 | x_1, \dots, x_{k-1}) = (1 + s_{k-1})/(k+1)$, and, thus, that $P(X_k = 0 | x_1, \dots, x_{k-1}) = [1 - (1 + s_{k-1})/(k+1)] = (k - s_{k-1})/(k+1)$.

Determination of Appropriate Action for Different Number-Correct Scores

Using the general backward-induction scheme discussed earlier, for a given maximum number n ($n \geq 1$) of items to be administered, a computer program was developed to determine the appropriate action (i.e., nonmastery, continuation, or mastery) at each stage k of testing for different number-correct score s_k .

As an example, the appropriate action is depicted in **Table 2** as a closed interval for a maximum of 20 items (i.e., $n = 20$). Students were considered true masters if they knew at least 60% of the subject matter. Therefore, t_e was fixed at 0.60. Furthermore – on a scale in which one unit corresponded to the constant costs of administering one random item (i.e., $e = 1$) – the loss parameters l_{01} and l_{10} were both assumed to be equal to 100 (i.e., $R = 1$). These numerical values reflected the assumption that the losses corresponding to taking incorrect classification decisions were rather large relative to the costs of administering one random item.

As can be seen from **Table 2**, at least four random items need to be administered before mastery can be declared. However, in principle, nonmastery can be declared after administering only two random items. Of course, in practice, it can be decided to start making classification decisions only after a certain number of random items have been administered to the student.

Simulation of Different Mastery Testing Strategies

In a Monte Carlo simulation, the Bayesian sequential strategy is compared with other existing approaches to

Table 2 Appropriate action calculated by stage of testing and raw score

Stage of testing	Appropriate action by raw score		
	Nonmastery	Continuation	Mastery
0		0	
1		[0,1]	
2	0	[1,2]	
3	0	[1,3]	
4	[0,1]	[2,4]	
5	[0,1]	[2,4]	5
6	[0,2]	[3,5]	6
7	[0,2]	[3,5]	[6,7]
8	[0,3]	[4,6]	[7,8]
9	[0,4]	[5,7]	[8,9]
10	[0,4]	[5,7]	[8,10]
11	[0,5]	[6,8]	[9,11]
12	[0,5]	[6,8]	[9,12]
13	[0,6]	[7,9]	[10,13]
14	[0,7]	[8,9]	[10,14]
15	[0,7]	[8,10]	[11,15]
16	[0,8]	[9,10]	[11,16]
17	[0,9]	[10,11]	[12,17]
18	[0,10]	11	[12,18]
19	[0,11]		[12,19]
20	[0,12]		[13,20]

mastery testing in terms of average test length (i.e., the number of items that must be administered before a mastery/nonmastery decision is made), classification accuracy, and average loss.

Description of the Testing Strategies Used for Comparison

The first comparison is made with a conventional fixed-length mastery test (CT) in which student performance was recorded as the proportion of correct answers. The student was declared a master for answering 60% or more items correctly after completion of the test, whereas nonmastery was declared otherwise.

The second comparison is made with Wald's SPRT procedure. The limits of the indifference region in which testing will continue were set symmetrically around 0.6 (i.e., the fixed value of t_c) with a bandwidth of 0.1. Hence, the proportion-correct values p_0 and p_1 were set at 0.5 and 0.7, respectively. Furthermore, the error rates α and β were each set equal to 0.1. These values reflected the fact that the losses associated with the false positive decision (i.e., l_{10}) and false negative decision (i.e., l_{01}) in the Bayesian sequential strategy were assumed to be both equal to 100.

According to the SPRT procedure, after k items have been administered with s_k of them being answered correctly, mastery was now declared if the likelihood ratio

$L(x_1, \dots, x_k | p_1) / L(x_1, \dots, x_k | p_0) = [(0.7)^{s_k} (0.3)^{k-s_k} / (0.5)^{s_k} (0.5)^{k-s_k}]$ was larger than $(1-\beta)/\alpha = (1-0.1)/0.1 = 9$, nonmastery if this likelihood ratio was smaller than $\beta/(1-\alpha) = 0.1/(1-0.1) = 0.11$, and otherwise sampling was continued. For those students who could not be classified as either a master or nonmaster before the item pool was exhausted, a classification decision was made in the same way as in the CT procedure, using a mastery proportion-correct value of 0.6.

Type of Test and Item Response Generation

The simulation study was conducted using a set of items that were perfect replications of each other. More specifically, it was assumed that each item could be described by a one-parameter logistic model (Rasch, 1960) with discrimination parameter a of 1 and equal difficulty parameter b of 0. This item pool reflected the choice of the binomial distribution for modeling response behavior in the Bayesian sequential procedure. Maximum test length (MTL) varied from ten to 50 items.

Item responses for 1000 simulated students – drawn from a $N(0,1)$ distribution – were generated for each item in the pool. For known ability of the simulated student and given item difficulty, first the probability of a correct answer was calculated using the one-parameter logistic model. Next, this probability was compared with a random number drawn from the uniform distribution in the range from 0 to 1. The item administered to the simulated student was scored correct or incorrect if this random number was either less or greater than the probability of a correct answer, respectively.

Furthermore, a simulated student was supposed to be a true master if his/her ability used to generate the item responses was higher than a prespecified cut-off point on the $N(0,1)$ -ability metric. Since a value of 0.6 on the proportion-correct metric corresponded following conversion with a value of 0 on the $N(0,1)$ -ability metric, the cut-off point on the $N(0,1)$ -ability metric was set equal to 0.

Results of the Simulation Study

Table 3 indicates that the Bayesian sequential strategy yielded shorter average test lengths than the other two strategies at each MTL level, whereas the phi correlations between true and estimated mastery status (i.e., classification accuracy) were generally somewhat lower at each MTL level. To examine which strategy is most efficient, therefore, we computed – for the other two strategies – the average test length at each MTL level for achieving the same phi correlation as under the Bayesian sequential strategy. In addition, as demonstrated by **Table 3**, it turned out that the Bayesian sequential strategy was

Table 3 Average number of items to be administered and classification accuracy

Strategy	Maximum test length			Maximum test length		
	10	25	50	10	25	50
	Average number of items			Classification accuracy		
CT	10	25	50	0.652	0.787	0.718
SPRT	8.64	14.47	17.42	0.584	0.723	0.718
BAYES	5.84	9.46	11.15	0.610	0.620	0.692

most efficient (i.e., combination of highest phi correlation and shortest average test length) at each MTL level.

Conclusions and Discussion

This article describes the deriving of optimal rules for the sequential mastery problem (non-mastery, mastery, or to continue testing), in the context of education, using the framework of Bayesian sequential decision theory. It should be emphasized, however, that the Bayesian sequential principle is especially appropriate when costs of testing can be assumed to be quite large. For instance, when testlets (i.e., blocks of parallel items) rather than single items are considered. In addition, the proposed strategy might be appropriate in the context of sequential testing problems in psychodiagnostics. Suppose that a new treatment (e.g., cognitive-analytic therapy) is to be tested on patients suffering from some mental health problem (e.g., anorexia nervosa). Following each exposure of a patient to the new treatment, it is desirous to make a decision concerning the effectiveness/ineffectiveness of the new treatment or to continue testing and exposing the new treatment to another randomly selected patient suffering from the same mental health problem. In such

clinical situations, costs of testing generally are quite large and the Bayesian sequential approach might be considered an alternative to fixed-length mastery tests (Vos, 2001).

See also: Computational Statistics; Decision Theory; Educational Data Modeling; Educational Measurement: Overview; Empirical Bayes Methods; Intelligent Systems.

Bibliography

- DeGroot, M. H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.
- Ferguson, R. L. (1969). *The Development, Implementation, and Evaluation of a Computer-Assisted Branched Test for a Program of Individually Prescribed Instruction*. Unpublished doctoral dissertation, Pittsburgh, PA: University of Pittsburgh.
- Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions*. Boston, MA: Houghton Mifflin.
- Lehmann, E. L. (1959). *Testing Statistical Hypotheses*, 3rd edn. New York: Macmillan.
- Lewis, C. and Sheehan, K. (1990). Using Bayesian decision theory to design a computerized mastery test. *Applied Psychological Measurement* **14**, 367–386.
- Luce, R. D. and Raiffa, H. (1957). *Games and Decisions*. New York: Wiley.
- Pearson, K. (1930). *Tables for Statisticians and Biometricians*. London: Cambridge University Press.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Educational Research.
- Reckase, M. D. (1983). A procedure for decision making using tailored testing. In Weiss, D. J. (ed.) *New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing*, pp 237–257. New York: Academic Press.
- van der Linden, W. J. and Mellenbergh, G. J. (1977). Optimal cutting scores using a linear loss function. *Applied Psychological Measurement* **1**, 593–599.
- Vos, H. J. (2001). A minimax procedure in the context of sequential testing problems in psychodiagnostics. *British Journal of Mathematical and Statistical Psychology* **54**, 139–159.
- Wald, A. (1947). *Sequential Analysis*. New York: Wiley.
- Wilcox, R. R. (1981). A review of the beta-binomial model and its extensions. *Journal of Educational Statistics* **6**, 3–32.

Signal Detection Theory

H Abdi, The University of Texas at Dallas, Richardson, TX, USA

© 2010 Elsevier Ltd. All rights reserved.

Overview

Signal detection theory (often abridged as SDT) is used to analyze data coming from experiments where the task is to categorize ambiguous inputs which can be generated either by a known process (called the signal) or be obtained by chance (called the noise in the SDT framework). For example, a radar operator must decide if what she sees on the radar screen indicates the presence of a plane (the signal) or the presence of parasites (the noise). This type of applications was the original framework of SDT (see the founding work of Green and Swets, 1966). But the notion of signal and noise can be somewhat metaphorical in some experimental contexts. For example, in a memory recognition experiment, participants have to decide if the input they currently see was presented before. Here the signal corresponds to a familiarity feeling generated by a memorized input, whereas the noise corresponds to a familiarity feeling generated by a new stimulus.

The goal of signal detection theory is to estimate two main parameters from the experimental data. The first parameter, called d' , indicates the strength of the signal (relative to the noise). The second parameter called C (a variant of it is called β) reflects the strategy of response of the participant of being more willing to say, for example, yes rather than no. SDT is used in very different domains from psychology (psychophysics, perception, memory), medical diagnostics (do the symptoms match a known diagnostic or can they be dismissed as irrelevant), to statistical decision (do the data indicate that the experiment has an effect or not).

The classic work on SDT is Green and Swets (1995), a basic introduction is McNicol (1972), two recent comprehensive references are Macmillan and Creelman (2005) and Wickens (2002).

A few examples of SDT to education can be found in McDermott *et al.* (1992), more recent examples can be found in McFall and Treat (1999), DeCarlo (2005), and DeCarlo and Luthar (2000).

The Model

It is easier to introduce the model with an example, so suppose that we have designed a face memory experiment. In the first part of the experiment, a participant was asked to memorize a list of faces. Then the participant is presented with a set of faces one at a time. Some of these

faces were seen before (these are old faces) and some were not seen before (these are new faces). The task is to decide for each face whether this face was seen (response Yes) or not (response No) in the first part of the experiment.

What are the different types of responses? A Yes response given to an old input is a correct response, it is called a Hit; but a Yes response to a new input is a mistake, it is called a False Alarm (abbreviated as FA). A No response given to a new input is a correct response, it is called a Correct Rejection; but a No response to an old input is a mistake, it is called a Miss. These four types of response (and their frequency) can be organized as shown in Table 1.

The relative frequency of these four types of response is not all independent. For example, when the signal is present (first row of Table 1) the proportion of Hits and the proportion of Misses add up to one (because when the signal is present, the subject can say either Yes or No). Likewise when the signal is absent, the proportion of FA and the proportion of Correct Rejection add up to one. Therefore, all the information in the table is given by the proportion of Hits and FAs.

Even though the proportions of Hits and FAs provide all the information in the data, these values are hard to interpret because they crucially depend upon two parameters. The first parameter is the *difficulty* of the task: the easier the task, the larger the proportion of Hits and the smaller the proportion of FAs. When the task is easy, we say that the signal and the noise are well separated, or that there is a large distance between the signal and the noise (conversely, for a hard task, the signal and the noise are close and the distance between them is small). The second parameter is the *strategy* of the participant: a participant who always says No will never commit a FA; on the other hand, a participant who always says Yes is guaranteed all Hits. A participant who tends to give the response Yes is called *liberal* and a participant who tends to give the response No is called *conservative*.

The SDT Model

So, the proportions of Hits and FAs reflect the effect of two underlying parameters: the first one reflects the separation between the signal and the noise and the second one the strategy of the participant. The goal of SDT is to estimate the value of these two parameters from the experimental data. In order to do so, SDT creates a

model of the participant's response. Basically, the SDT model assumes that the participant's response depends upon the intensity of a hidden variable (e.g., familiarity of a face) and that the participant responds Yes when the value of this variable for the input is larger than a pre-defined threshold.

SDT also assumes that the stimuli generated by the noise condition vary naturally for that hidden variable. As is often the case elsewhere, SDT, in addition, assumes that the hidden variable values for the noise follow a normal distribution. Recall, at this point, that when a variable x follows a Gaussian (aka Normal) distribution, this distribution depends upon two parameters: the mean (denoted μ) and the variance (denoted σ^2). It is defined as

$$\mathcal{G}(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \quad [1]$$

In general, within the SDT framework, the values of μ and σ are arbitrary and therefore we choose the simpler values of $\mu = 0$ and $\sigma = 1$ (other values will give the same results but with more cumbersome procedures). In this case, eqn [1] reduces to

$$\mathcal{N}(x) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}x^2 \right\} \quad [2]$$

Finally, SDT assumes that the signal is *added* to the noise. In other words, the distribution of the values generated by the signal condition has the same shape (and therefore the same variance) as the noise distribution.

Table 1 The four possible types of response in SDT

Reality	Decision (participant's response)	
	Yes	No
Signal present	Hit	Miss
Signal absent	False Alarm (FA)	Correct Rejection

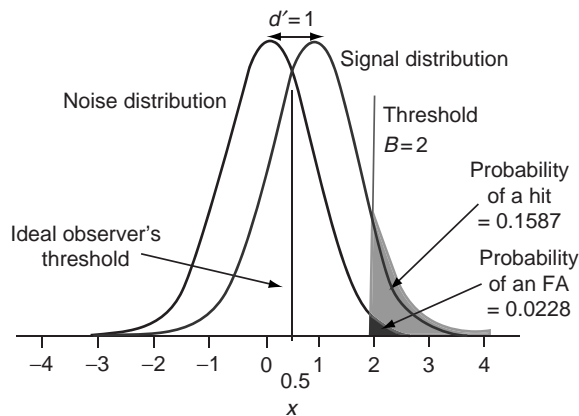


Figure 1 The SDT model.

Figure 1 illustrates the SDT model. The x -axis shows the intensity of underlying hidden variable (e.g., familiarity for the face example). As indicated above, the distribution of the noise is centered at zero (i.e., mean of the noise is equal to zero), with a standard deviation of 1. So, the standard deviation of the noise is equivalent to the unit of measurement of x . The distribution of the signal is identical to the noise distribution, but it is moved to the right of the noise distribution. The distance between the signal and the noise distributions corresponds to the effect of the signal (this is the quantity that is added to the noise distribution in order to get the signal distribution): this distance is called d' . Because the mean of the noise distribution is zero, d' is equal to the mean of the signal distribution.

The strategy of the participant is expressed via the choice of the threshold. There are several ways of expressing the position of this threshold, among the possible candidates we will mention four of them denoted B , D , C , and β . The first quantity B (sometimes called ϑ) gives the position of the threshold on the x -axis. In the example illustrated in **Figure 1**, this value is equal to 2, and so the participant corresponding to the figure has decided that any input with a value of x larger than 2 comes from the signal distribution and is given the response Yes. The position of the threshold can also be given relative to the signal distribution (because the noise has zero mean, B is the distance of the threshold relative to the noise distribution), as the mean of the signal is equal to d' we can compute D as $D = d' - B$ (a value equal to -1 in our example, see **Table 2**).

The most popular way of expressing the location of the threshold, however, is neither from the distribution of the noise nor from the distribution of the signal but relative to what is called the *ideal observer*. The ideal observer minimizes conjointly the probability of a Miss and of an FA. When each type of errors has the same cost, the criterion of the ideal observer is positioned on the average of the means of the signal and the noise distribution. In our example, the threshold of the ideal observer would be equal to $\frac{1}{2}d' = \frac{1}{2} = 0.5$. The value of C is the distance from the actual threshold to the ideal observer; it can be computed as $C = B - \frac{d'}{2} = 2 - 0.5 = 1.5$. The sign

Table 2 The probability or the four possible types of response according to **Figure 1**

Reality	Decision (participant's response)		
	Yes	No	Total
Signal present	Hit	Miss	
Signal absent	False Alarm (FA)	Correct Rejection	
	Pr {Hit} = 0.1587	Pr {Miss} = 0.8413	1
	Pr {FA} = 0.0228	Pr {Correct Rejection} = 0.9772	1

of C reveals the participant's strategy: when $C = 0$, we have the ideal observer; when C is negative the participant is liberal (i.e., responds Yes more often than the ideal observer); when C is positive the participant is conservative (i.e., responds No more often than the ideal observer).

An alternative way of expressing the position of the participant's criterion is given by the quantity called β . It corresponds to the ratio of the height of the signal distribution to the noise distribution for the value of the threshold. Because the distributions of the noise and the signal are normal with variance equal 1, we can compute β from eqn [2] as

$$\beta = \frac{\mathcal{N}(D)}{\mathcal{N}(B)} = \frac{\mathcal{N}(1)}{\mathcal{N}(2)} = \frac{0.2420}{0.0540} \approx 4.4817 \quad [3]$$

Some rewriting can show that eqn [3] can be rewritten as

$$\beta = \exp \{d' \times C\} \quad [4]$$

The quantity β has the advantage of being a likelihood ratio and can be used to interpret SDT within a statistical framework. For practical reasons, it is often easier to compute the logarithm of β ; for example, from eqn [4], we get

$$\ln \beta = d' \times C = 1 \times 1.5 = 1.5 \quad [5]$$

The model illustrated by **Figure 1** generates a specific pattern of response probabilities which can be computed from integrating the normal distribution. So, for example, the probability of a FA is obtained as the probability (i.e., area under the normal distribution) of finding a value larger than 2 with a normal distribution of mean 0 and variance 1 (this can be computed with most statistical packages or from tables such as the ones given in Abdi *et al.*, 2009). This quantity is also called the probability associated to the value 2, in our example it is equal to 0.0228. Along the same lines, the probability of a Hit is obtained as the probability (i.e., area under the normal distribution) of finding a value larger than 2 with a normal distribution of mean 1 (i.e., the mean of the signal) and variance 1, this is equivalent of finding the probability (i.e., area under the normal distribution) of finding a value larger than $2 - 1 = 1$ with a normal distribution of mean $1 - 1 = 0$ and variance 1. This value is equal to 0.1587.

SDT in Practice

The previous example was describing the performance of a participant who behaved according to the SDT model. However, in practice we do not know the values of the parameters of SDT, but we want to estimate them from the performance of the participants. In an experimental paradigm, the only observable quantities are the

Table 3 The performance of a wine taster trying to identify Gamay in a Pinot Noir wine

Reality	Decision (taster's response)		Σ
	Yes (Gamay)	No (Pure Pinot)	
Signal present (Gamay)	Hit	Miss	
	# {Hit} = 9 Pr {Hit} = 0.9	# {Miss} = 1 # {Miss} = 0.1	10 1
Signal absent (Pure Pinot)	False Alarm (FA)	Correct Rejection	
	# {FA} = 2 Pr {FA} = 0.2	# {Correct Rejection} = 8 Pr {Correct Rejection} = 0.8	10 1

participant's responses from which we can derive the number of Hits and FAs.

To illustrate this problem suppose that we want to evaluate the performance of a wine taster whose task is to detect if a wine labeled as made from Pinot Noir has been tempered by the addition of some Gamay (generally considered an inferior grape). Here, the signal corresponds to presence of Gamay. Our wine taster tasted (blindfolded) 20 glasses of Pinot (half of them tempered with some Gamay and half without). The results are reported in **Table 3**, and show that the proportion of Hits and FAs are respectively 0.9 and 0.2. In order to find the values of d' and the criterion, we need to inverse the formulas given above (i.e., eqns [3]–[5]). We need one new notation: for a normal distribution with zero mean, we denote by Z_P the value of the normal distribution whose associated probability is equal to P (e.g., $Z_{0.25} = 1.96$). We denote as Z_H and Z_{FA} the values corresponding to the proportions of Hits and FAs. With these new notations and after some (minor) algebraic manipulations, we find the following set of formulas. The estimation of d' is obtained as

$$d' = Z_H - Z_{FA} = Z_{0.9} - Z_{0.2} = 1.28 - (-0.84) = 2.12 \quad [6]$$

The estimation of C is obtained as

$$\begin{aligned} C &= -\frac{1}{2} [Z_H + Z_{FA}] = -[Z_{0.9} + Z_{0.2}] \\ &= -\frac{1}{2} [1.28 + 0.84] = -0.22 \end{aligned} \quad [7]$$

and $\ln \beta$ is obtained as

$$\ln \beta = d' \times C = 2.12 \times -0.22 = -0.47 \quad [8]$$

(β is obtained as $\exp\{\ln \beta\} = 0.63$).

How to interpret these results? The taster is clearly (but not perfectly) discriminating between Pinots and tempered Pinots (as indicated by a d' of 2.12), this taster is also liberal (in case of doubt the taster will rather say that the wine has been tempered rather than not).

See also: Program Evaluation.

Bibliography

- Abdi, H., Edelman, B., Valentin, D., and Dowling, W. J. (2009). *Experimental Design and Analysis for Psychology*. Oxford: Oxford University Press.
- DeCarlo, L. T. (2005). A model of rater behavior in essay grading based on signal detection theory. *Journal of Educational Measurement* **42**, 53–76.
- DeCarlo, L. T. and Luthar, S. S. (2000). Analysis and class validation of a measure of parental values perceived by early adolescents: An application of a latent class model for rankings. *Educational and Psychological Measurement* **60**, 578–591.
- Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.
- Macmillan, N. A. and Creelman, C. D. (2005). *Detection Theory: A User's Guide*, 2nd edn. Mahwah, NJ: Erlbaum.
- McDermott, P. A., Fantuzzo, J. W., Glutting, J. J., Watkins, M. W., and Baggaley, R. A. (1992). Illusions of ipsative assessment of children's ability. *Journal of Special Education* **25**, 504–526.
- McFall, R. M. and Treat, T. A. (1999). Quantifying the information value of clinical assessment with signal decision theory. *Annual Review of Psychology* **50**, 215–241.
- McNicol, D. (1972). *A Primer of Signal Detection Theory*. London: George Allen and Unwin.
- Wickens, T. D. (2002). *Elementary Signal Detection Theory*. Oxford: Oxford University Press.

Small Area Estimation

A Noble, Massey University, Palmerston North, New Zealand

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Sample surveys are commonly used for data collection in many academic areas, particularly the social sciences. The data collected are then used to estimate statistics of interest and the sample sizes are determined so that those statistics are estimated with an acceptable sampling error. Frequently, there is a demand for estimates at smaller geographic areas or finer cross-classifications of categorical variables, as in **Figure 1**. In the literature, there are a number of terms used to denote these smaller groups; small area, local area, and small domain are common. The term small area is used in this article.

Normally, it will not be possible to obtain direct estimates with acceptable errors for small areas because of the small sample sizes in the smaller partitions. Small-area estimation may then be used to improve the estimates for these smaller areas or cross-classifications. Additional data or a model with assumptions, or both, are needed to improve the estimates. A range of techniques have been developed to apply to this problem. These methods borrow strength from related or similar small areas through explicit or implicit statistical models that connect the small areas via supplementary data. In other words, data are available at the small-area level, which can be used – via a model and often requiring auxiliary data – to estimate parameters for the unknown data at the small-area level.

Their historical use has largely been driven by the data that were available in a particular application. In the earlier methods of small-area estimation, census data were generally available from an earlier time and the objective was to update estimates with data collected more recently or by using historical relationships in the variable of interest over time from more than one census – for example, a linear trend.

More recently, sample survey data – collected for other purposes – have been combined with different information with regard to the small areas – from another source – to produce the estimates. These other sources may include a census.

Marker (1999) demonstrated that most of the historical small-area estimation methods could be placed in a framework of the linear regression model expressed in the equation

$$Y = X\beta + \epsilon \quad [1]$$

Most often, Y will be the vector of estimates for the small areas. The X matrix may be a matrix of data or of indicator

variables, to denote membership of groups, or a mixture of the two. ϵ will be an error term which will, generally, have normal distribution with mean zero.

The methods can be loosely grouped into the following:

- Demographic methods.
- Synthetic and related methods.
- Symptomatic regression.
- Structure-preserving estimation (SPREE).
- Models with area-specific effects.
- Empirical best linear unbiased prediction (EBLUP), empirical Bayes, and hierarchical Bayes approaches.
- Spatial statistics.

This article considers each of these methods in more detail.

Demographic Methods

These methods are also described as symptomatic accounting techniques (SAT). All of these were developed for estimating human population sizes and use particular available information to improve estimates from other data, possibly an earlier census. The literature is not consistent in its classification of these older techniques but they are, generally, extensions of a very simple model in which a previously known estimate – usually from a census – is updated by adding or subtracting births, deaths, immigration, and migration counts for each small area. They are commonly used for predicting future rolls of schools for administrative uses.

In this model, there is no stochastic part as it is assumed that the values are known exactly and it is merely an arithmetic problem.

As administrative records do not, usually, detail all arrivals and departures from each area, various methods have been developed to estimate these from symptomatic variables – that is, variables that reflect the arrivals and departures. These methods include the following: administrative records (ARs), housing unit (HU), vital rates, composite method, and composite method II (CMII). References for these are presented in Ghosh and Rao (1994).

Synthetic and Related Methods

Gonzalez (1973) describes synthetic estimates as follows: “An unbiased estimate is obtained from a sample survey

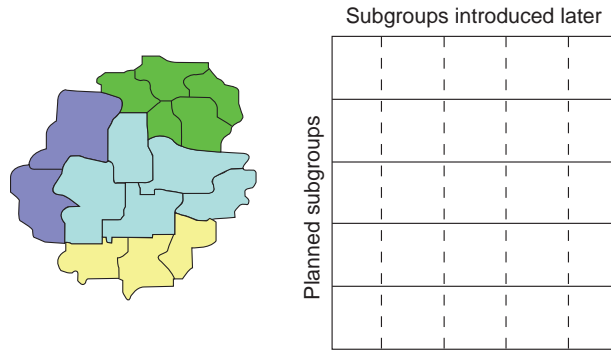


Figure 1 Generation of small areas. Subdivisions of geographic regions or divisions that cut across the divisions used for sampling.

for a large area; when this estimate is used to derive estimates for sub areas under the assumption that the small areas have the same characteristics as the large area, we identify these estimates as synthetic estimates.” They have been used in a number of contexts because:

- they are simple to use,
- they can be applied to general sampling designs, and
- there exists the potential to increase accuracy by borrowing information from similar small areas.

There are a number of related methods which will be described.

Synthetic Estimation

Synthetic estimation uses survey data to estimate proportions of different subgroups – for example, age by sex by ethnicity subgroups, and the estimates for the small areas are derived by taking the appropriate weighted average of these for each small area. The standard formulation for these estimates is

$$\hat{Y}_a = \frac{\sum_j N_{aj} \bar{y}_j}{N_a} \text{ where}$$

\hat{Y}_a is the estimated mean for each small area a , N_{aj} is the size of the population of small area a and subgroup j , N_a is the total population in area a , and \bar{y}_j is the sample average for Y for subgroup j across all small areas.

Symptomatic Regression

Two past censuses are used to define a regression relationship between the variable of interest and ratios of the symptomatic variables at each census. These ratios are then found for the last census and the present time and used to update the variable of interest.

Ratio Correlation

Let $t = \{0, 1, 2\}$ denote two consecutive census years and the current year, note that the 3 years represented may not

be equally spaced. Further, N_{at} and S_{ajt} are the population total and the value for the j th symptomatic variable for the a th local area ($a = 1, \dots, A$) in the year t , respectively.

Further, let: $p_{at} = \frac{N_{at}}{\sum_a N_{at}}$ and $s_{ajt} = \frac{S_{ajt}}{\sum_a S_{ajt}}$ be the corresponding proportions and write $R'_a = \frac{p_{a1}}{p_{a0}}$, $R_a = \frac{p_{a2}}{p_{a1}}$ (which is unknown as p_{at} is what we are trying to estimate) and $r'_{aj} = \frac{s_{aj1}}{s_{aj0}}$, $r_{aj} = \frac{s_{aj2}}{s_{aj1}}$

Using the data ($R'_a, r'_{a1}, \dots, r'_{aj}; a = 1, \dots, A$) and multiple regression we first fit

$$R'_a = \hat{\beta}_0 + \hat{\beta}_1 r'_{a1} + \dots + \hat{\beta}_j r'_{aj}$$

where the $\hat{\beta}$'s are the estimated regression coefficients that link the change, R'_a , in the population proportions between the two census years to the corresponding changes, r'_{aj} , in the proportions for the symptomatic variables. Next, the changes, R_a – in the postcensal period – are predicted as $\hat{R}_a = \hat{\beta}_0 + \hat{\beta}_1 r_{a1} + \dots + \hat{\beta}_j r_{aj}$ using the known changes in the symptomatic variables in the postcensal period and the estimated regression coefficients. Finally, the current population counts are estimated as

$$\tilde{N}_{a2} = \hat{R}_a p_{a1} \left(\sum_a N_{a2} \right)$$

where the total population across all local areas is found from other data.

The difference correlation method uses the same process; however, differences between the proportions at the two time points are used rather than their ratio.

These procedures use regression coefficients in the last inter-censal period and it is assumed that these parameters are applicable to the change from the last census to the present time. Significant changes in the statistical relationship can lead to errors in the current estimates. Such changes are quite likely if the censuses are at intervals of 10 years as the ratios are based on data that will be between 10 and 20 years old.

The methods described, thus far, use only census data. Use of sample survey data can help to reduce the problem of changes in the statistical relationship and give accurate current data for some small areas.

The Sample-Regression Method

The problem of out-of-date data can be avoided by using sample estimates of R_i for those local areas where data exist. The sample regression estimators are then obtained for all areas using the known symptomatic ratios.

The sample regression method has similarities with symptomatic regression, except that, instead of using census data from two past censuses, data are used from a sample of the small areas and the most recent census. Thus, it is assumed that the sample of small areas is

representative of all of them. A version of this method using a logistic model has been used for state-level estimates for the Private School Survey (Causey *et al.*, 1999).

An alternative is to develop a regression model for the variable of interest from the survey data and then make predictions for the whole population using census values for the variables in the regression model. The small-area estimates can then be derived by summing or averaging the variable of interest over the areas. This has the advantage of providing estimates for areas that have no representation in the sample, on the assumption that the model developed from the survey applies in areas not surveyed as well as those surveyed. This methodology has been used in conjunction with the International Adult Literacy Survey data to predict proportions of residents with low literacy levels in small, population wise, administrative areas in New Zealand (Culligan *et al.*, 2004). It is also used by the World Bank for estimating poverty in developing countries (Elbers *et al.*, 2003).

Components of Variance Regression

The methods considered, so far, have considered data at the area level but the components of variance method considers data at the element level, k . The error term has two components – one is random and the other is small-area specific.

The model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$\boldsymbol{\varepsilon}$ is the vector of errors with elements ε_{ak} .

$$\boldsymbol{\varepsilon} = \mathbf{v} + \mathbf{u}$$

\mathbf{v} is the vector of small-area effects; these will be the same for all individuals in each small area, hence \mathbf{v} has elements v_a ; \mathbf{u} is the vector of random errors with elements u_{ak} .

$$\mathbf{v} \sim N(\mathbf{0}, \Sigma_{vv}) \text{ and } \mathbf{u} \sim N(\mathbf{0}, \Sigma_{uu}).$$

This estimate is not optimal unless all of the n_a are equal.

The least square estimates of $\boldsymbol{\beta}$ are then $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$. The average residual for each small area can be calculated and assigned the assumed error structure $E(v_a | \bar{\varepsilon}_a) = \bar{\varepsilon}_a G_a$ where $G_a = (\Sigma_{vv} + n_a^{-1} \Sigma_{uu})^{-1} \Sigma_{vv}$. Using the survey data for that small area and the estimate for average error for each area, an estimate of v_a for each small area can be found.

Structure-Preserving Estimation

Purcell and Kish (1980) propose structure-preserving estimation (SPREE) as a generalization of synthetic estimation in the sense that it makes fuller use of reliable direct estimates. SPREE commonly uses the well-known method of iterative fitting of margins in a multiway table

Original data (census data)			New margins (survey data)		
	Male	Female		Male	Female
Age 1	3	5	8	Age 1	6
Age 2	4	7	11	Age 2	15
	7	12	19	9	12
				21	

Figure 2 Data requirements for SPREE for a very simple table.

(iterative proportional fitting (IPF)), where the margins are direct estimates. Although this method is based on ratios – as in the previous method – the IPF procedure guarantees that the ratio estimates are adjusted to conform to the margins.

In the very simple example in **Figure 2**, the data in the body of the table – on the left – are generated from a census and the margins – on the right – from the survey. The margins are available for the different categories but not for all of the combinations. The data in the body of the table are then adjusted to agree with the margins. This adjustment is achieved by first adjusting the rows then the columns and repeating this until convergence. The interactions between the variables are set by the census data, except for any that may be defined in the new margins. The interactions that are defined by the census data remain unchanged through the iterative process. Only the effects defined by the survey data are changed to agree with the new margins. (The Structure of the table is PREserved in the Estimation, hence SPREE.)

Noble *et al.* (2002) demonstrated that SPREE can be described by two generalized linear models – one for the census data and one for the survey. Terms which are unknown in the sample survey model are substituted from the census model.

This method is commonly used in the estimation of unemployment and for updating census data.

Composite Estimation

The composite estimators cover a wide range of examples – all of which can be characterized by a weighted average of two estimators. The most common simply apply a weighted average to a direct estimator and one of the estimators described above. The weights are based on the variance or mean square error of the two estimators. The composite estimator is intended to offset the bias in the indirect estimator and the variability in the direct estimator. In cases where few areas in the model have direct estimators, there is little advantage in the composite estimator. Synthetic estimators tend to be biased, direct estimators may be unstable. An obvious way to reduce these two effects is to take a weighted average of the two estimates. Such estimators may be written

$$\hat{Y}_a^C = w_a \hat{Y}_{1a} + (1 - w_a) \hat{Y}_{2a} \quad [2]$$

where \hat{Y}_{1a} is a direct estimator and \hat{Y}_{2a} is an indirect estimator, and w_a is a suitably chosen weight ($0 \leq w_a \leq 1$). A number of possible choices of weights have been proposed. Care needs to be taken in their choice as none work well in all situations.

Best Linear Unbiased Prediction

Holt *et al.* (1979) proposed a best linear unbiased prediction (BLUP) estimator of \hat{Y}_i under the following model for the finite population:

$$y_{agk} = \mu_g + e_{agk} \quad k = 1, \dots, N_{ig}; g = 1, \dots, G; \quad k = 1, \dots, m$$

where y_{agk} is the y -value of the k th unit in the cell (a, g) for small area a in domain g , μ_g 's are fixed effects, and the errors e_{agk} are uncorrelated with zero means and variances σ_a^2 .

N_{ag} denotes the number of population elements – in the large domain g – which belong to the small area a . Suppose n_{ag} elements in a sample of size n fall in the cell (a, g) and let \bar{y}_{ag} and $\bar{y}_{\cdot g}$ denote the sample means for (a, g) and g , respectively.

The BLUP estimator of Y_i is given by

$$\hat{Y}_a^B = \sum_g \hat{Y}_{ag}^C = \sum_g \left(\frac{n_{ag}}{N_{ag}} \hat{Y}_{ag} + \left(1 - \frac{n_{ag}}{N_{ag}} \right) \hat{Y}_{\cdot g}^S \right)$$

where $\hat{Y}_{ag} = N_{ag} \bar{y}_{ag}$ and $\hat{Y}_{\cdot g}^S = N_{ag} \bar{y}_{\cdot g}$ or $\hat{Y}_a^B = \sum_g (n_{ag} y_{ag} + (N_{ag} - n_{ag}) \bar{y}_{\cdot g})$ which, in effect, sums the values in the sample from that cell and estimates the total for the rest in that cell by the mean across the g th group multiplied by the number not in the sample. If the sample size for that cell is large the estimate tends to the direct estimator, if it is small it tends to the synthetic estimator. If the sample fraction is small for all of the groups, then no account is made of the relationship between, between-area, and within-area variation. This problem can be reduced by more sophisticated models.

Models with Area-Specific Effects

Two types of models – with area-specific effects – have been proposed.

In the first, area-specific auxiliary data $\mathbf{x}_a = (x_{a1}, \dots, x_{aJ})^T$ is assumed to be known and the parameters of interest, θ_a , are assumed to be related to \mathbf{x}_a .

In particular, we assume that $\theta_a = \mathbf{x}_a^T \boldsymbol{\beta} + v_a$ $a = 1, \dots, A$.

The z_a 's are known positive constants, $\boldsymbol{\beta}$ is the vector of regression parameters, and the v_a 's are iid random variables with $E[v_a] = 0$ and $V(v_a) = \sigma_a^2$.

In the second type, element-specific auxiliary data $\mathbf{x}_{aj} = (x_{aj1}, \dots, x_{ajJ})^T$ are available for the population elements. The variable of interest y_{aj} is assumed to be related to \mathbf{x}_{aj} through a nested error regression model.

$$y_{aj} = \mathbf{x}_{aj}^T \boldsymbol{\beta} + v_a + e_{aj} \quad j = 1, \dots, N_i; a = 1, \dots, A \quad [3]$$

where $e_{aj} = \tilde{e}_{aj} k_{aj}$ and the \tilde{e}_{aj} 's are iid random variables independent of the v_a 's with, $E(\tilde{e}_{aj}) = 0$, and $V(v_a) = \sigma^2$, the k_{aj} 's being known constants and N_a being the number of elements in the a th area.

For making inferences under the first model, some assumptions are required that may be quite restrictive in some applications. Violation of these assumptions will introduce bias to the estimates.

Assuming that direct estimators $\hat{\theta}_a$ are available and that $\hat{\theta}_a = \theta_a + e_a$, we obtain the model $\hat{\theta}_a = \mathbf{x}_a^T \boldsymbol{\beta} + v_a z_a + e_a$, which is a special case of the general mixed linear model. This model is commonly referred to as the Fay–Herriot model as Fay and Herriot (1979) used this model to estimate *per capita* income for small places in the United States.

In the second model, the assumption is made that the sample values also obey the model – while this will be true under simple random sampling it may not be appropriate under more complex sampling designs. The model can, however, be extended to account for such features.

The model in eqn [3] can be written in a matrix form as

$$\mathbf{y}_i^p = \mathbf{X}_i^p \boldsymbol{\beta} + \mathbf{v}_i \mathbf{1}_i^p + \mathbf{e}_i^p$$

The mean of each small area \bar{Y}_a can be written as $\bar{Y}_a = f_a \bar{y}_a + (1 - f_a) \bar{y}_a^*$ where $f_a = \frac{n_a}{N_a}$ and \bar{y}_a, \bar{y}_a^* denote the means for sampled and nonsampled elements, respectively. Estimation of the overall mean can be viewed as prediction of the mean of the nonsampled elements, given the data.

EBLUP, Empirical Bayes, and Hierarchical Bayes Approaches

Lahiri (1996) combines these approaches under the heading of composite estimation, which was discussed above. They are characterized by a weighted average of two estimates – the direct estimate and a synthetic estimate.

$$\hat{\theta}_a^C = w_a y_a + (1 - w_a) \mathbf{x}_a^T \hat{\boldsymbol{\beta}}$$

which is essentially the same as eqn [2].

The choice of weight, w_a , constitutes the difference between the three approaches.

These approaches are useful when direct estimates are available for a reasonable number of the small areas; if they are not, then the weight, w_a , is equal to zero for most areas and the estimate is simply the synthetic estimate.

They are an attempt to combine the zero bias of the direct estimate with the smaller variance of the synthetic estimate.

Spatial Statistics

The basic assumptions made in most simple statistical analysis are that the data are independent and identically distributed. In several circumstances, neither of these assumptions hold true. Spatial data contradict these assumptions. It is intuitively sensible to assume that data collected from some connected geographical regions – defined by arbitrary political boundaries – are likely to be correlated. Cressie (1993) is an authoritative book on spatial statistics.

Data collected in a wide range of situations have long been recognized as having spatial variation. Initially, most research was focused on reducing the effect of spatial dependence in the results. R.A. Fisher – at Rothamstead Experimental Station in England, in the 1920s and 1930s – went to great lengths to remove it, largely, by designing experiments with replication, randomization, and blocking. Since that time, the emphasis has moved to explicitly modeling the spatial variation – both in the types of agricultural experiments that Fisher was involved with and in a wider range of problems. This wider range of problems includes applications in geostatistics, particularly mining, atmospheric science, soil science, epidemiology, public health, and analysis of data from satellites. This range of applications continues to expand rapidly.

Spatial approaches to small-area estimation are many and depend on the available data. In a simple case, direct estimates for each small area may be smoothed by defining a local correlation structure. This will reduce the variance of the estimates by assuming that areas close by are similar. This correlation may be fixed for neighbors or may decay as the distance between sites increases. This distance may be Euclidean or measured on some other metric.

The analysis of such correlated data has seen many advances, particularly in the last 25 years or so of the last century and there are now some fairly standard approaches that can be used. These analyses assume some kind of known structure – generally, a correlation structure – which is defined and parameters are estimated.

Often, there are a number of possible structures that could be investigated. We may consider that areas close to each other or areas with similar demographics could be similar. School attendance in inner-city areas could be similar despite the cities being distant from each other.

These spatial correlations can be included in the analysis by an appropriate covariance matrix and the investigator considers the inclusion of this correlation in the analysis. Many models of spatial dependence could then

be fitted. One which is commonly used is the conditional autoregressive model (CAR).

Gaussian conditional autoregressive processes have been used in spatial statistics to describe the association between random variables observed at fixed sites in some Euclidean space (Cressie, 1993). We can extend the simple model in eqn [1] by adding a spatial term

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z} + \boldsymbol{\varepsilon}$$

The spatial dependence is modeled by the covariance matrix, which includes terms for adjacent areas. This matrix must be symmetric (if i is adjacent to j , then j must also be adjacent to i).

$$\mathbf{z} \sim \text{Gau}(\boldsymbol{\mu}, (\mathbf{I} - \mathbf{C})^{-1}\mathbf{M})$$

where \mathbf{z} is the vector of data values ($z(s_1), \dots, z(s_n)$), $\boldsymbol{\mu}$ is vector of means, \mathbf{C} is the $(n \times n)$ matrix of c_{ij} where $c_{ii} = 0$, $c_{ik} = 0$ where i and k are not adjacent areas and $c_{ij}^2 = c_{ji}^2$ and \mathbf{M} is the diagonal $(n \times n)$ matrix $\text{diag}(\tau_1^2, \dots, \tau_n^2)$.

In effect, the error term is decomposed into two parts – one of which is a latent term explaining variability due to similarities between areas or cells in the table.

In spatial models, there is often a problem of what to do with areas that are on the edge of the lattice. Similar to the surface tension in a liquid, these areas pulled in toward the areas within the lattice with no effect from outside. This may need to be considered in particular examples.

Discussion

Small-area estimation does not seem to be used prevalently in educational research. One explanation of this is that there are copious amounts of data to analyze at very fine levels of detail and direct estimators are available at the small-area level. Most developed countries have highly organized reporting systems that produce information down even to the individual student level. The necessity for small-area methods is not strong.

A new development in the previous 20 years has been a number of international assessments that have been carried out. Examples of these include

- Trends in International Mathematics and Science Study (TIMSS).
- Progress in International Reading Literacy Study (PIRLS).
- Adult Literacy and Lifeskills Survey (ALL).

The first two are conducted by the International Association for the Evaluation of Educational Achievement (IEA) and the Organization for Economic Cooperation and Development (OECD)'s Program for International Student Assessment (PISA) and the last one is conducted by Statistics Canada and the OECD.

These are all in the form of samples surveys from the participating countries and include some auxiliary variables. There exist many possibilities for combining the data from these surveys with census data to make predictions for small areas within the countries surveyed.

Bibliography

- Causey, B. D., Bailey, L., and Kaufman, S. (1999). *Indirect State-Level Estimation for the Private School Survey, Technical Report for the National Center for Education Statistics*, U.S. Department of Education, Office of Educational Research and Improvement. NCES 1999–351.
- Culligan, N., Arnold, G., Noble, A., and Sligo, F. X. (2004). *Analysis of New Zealand Data from the International Adult Literacy Survey*. Wellington: Department of Communication and Journalism, Massey University, N. Z. Report Commissioned by the N. Z. Ministry of Education.
- Elbers, C., Lanjouw, J., and Lanjouw, P. (2003). Micro-Level estimation of poverty and inequality. *Econometrica* **71**(1), 355–364.
- Fay, R. E. and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association* **74**, 269–277.
- Gonzalez, M. E. (1973). Use and evaluation of synthetic estimators. *Proceedings of the Social Statistics Section*, pp 33–36. American Statistical Association: Washington, DC.
- Holt, D., Smith, T. M. F., and Tomberlin, T. J. (1979). A model-based approach to estimation for small subgroups of a population. *Journal of the American Statistical Association* **74**, 405–410.
- Longford, N. T. (2002). Small-area estimation in national surveys. *Short Course at the RSS Conference*, Plymouth.

- Marker, D. (1999). Organization of small area estimators using a generalized linear regression framework. *Journal of Official Statistics* **15**, 1–24.
- Noble, A. D. L., Haslett, S. J., and Arnold, G. (2002). Small area estimation via generalized linear models. *Journal of Official Statistics* **18**(1), 45–60.
- Purcell, N. J. and Kish, L. (1980). Postcensal estimates for local areas (domains). *International Statistical Review* **48**, 3–18.

Further Reading

- Cressie, N. (1993). *Statistics for Spatial Data*, pp 900. New York: Wiley.
- Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal. *Statistical Sciences* **9**, 55–93.
- Goldstein, H. (1995). *Multilevel Statistical Models*. London: Arnold.
- Longford, N. T. (2005). *Missing Data and Small-Area Estimation Modern Analytical Equipment for the Survey Statistician Series: Statistics for Social and Behavioral Sciences XVI*, ISBN: 978-1-85233-760-5. London: Springer.
- Pfeffermann, D. (2002). Small area estimation – new developments and directions. *International Statistical Review* **70**, 125–143.
- Rao, J. N. K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology* **25**, 175–186.
- Rao, J. N. K. (2003). *Small Area Estimation. Wiley Series in Survey Methodology*. New York: Wiley-Interscience/Wiley.

Relevant Websites

- <http://en.wikipedia.org> – Small Area Estimation, Wikipedia.
- <http://www.dipstat.ec.unipi.it> – Small Area Estimation 2007, Dipartimento di Statistica e Matematica Applicata all'Economia.

Statistical Analysis of Functional Data

F Ferraty, P Sarda, and P Vieu, Toulouse University, Toulouse, France

© 2010 Elsevier Ltd. All rights reserved.

In many applications, the object of interest is a function $X = X(t)$ depending continuously on a variable t which may be time, a vector of spatial position, etc. In practice, for each individual i , one observes the function X on a grid t_1, t_2, \dots, t_p and the statistical sample contains the values of $X_i(t_1), \dots, X_i(t_p)$ for $i = 1, \dots, n$ (i.e., n individuals). One encounters such data, for instance, in longitudinal studies (when the measurements t_1, t_2, \dots are dates), in chemistry (i.e., t_1, t_2, \dots could be wavelengths), in image analysis (i.e., t_1, t_2, \dots are a two or three-dimensional vector), etc. Finally, the collected data consist in either a population of discretized curves (spectrometric curves, radar waveforms, electricity consumptions, growth curves, digitized voice recordings, etc.), a collection of surfaces, multidimensional images (digitized pictures, satellite images, multispectral images, hyper-spectral images, times series of images), or a set of any other higher dimensional (discretized) functions. With the development of modern technology, the data are recorded on finer and finer grids, and so one can consider the observed data in a continuous manner. Such data are called functional data, and a new branch of statistics, called functional data analysis (FDA), has emerged with the aim of developing models and methodologies that are able to study such high dimensional data and integrate specific information. The terminology functional variable is also used to indicate a statistical variable whose observations lead to functional data. This recent statistical topic is motivated by numerous domains of applications: biology, biomechanics, chemometrics, computer sciences, ecology, econometrics, medical sciences, industry, etc. Problems encountered in FDA are of the same nature as in general statistics: analyze variability and find structures (factorial analysis, classification, etc.) in a set of functional data, predict a (functional or scalar) variable depending on the variations of a functional predictor (regression), depicting a distribution of functional data, identifying outliers, etc. The reader can find a good overview on FDA methodologies in Ramsay and Silverman (2005) and several case studies in Ramsay and Silverman (2002). A more recent book on FDA from Ferraty and Vieu (2006) deals with general statistical models and discusses the estimation of a distribution of a population of curves and some related features. Functional data can also be studied in the setting of time series (see Bosq, 2000).

Why developing specific statistical strategies for functional data? The starting point of FDA comes from the inability of standard multivariate technics to extract all

the functional richness of such high dimensional data. For instance, a standard multivariate linear regression generally fails when one aims to predict a scalar response from a functional explanatory variable. Thus, our goal is to present the usefulness of FDA as well as its potentialities.

This contribution is organized as follows. We start with a general description of functional data by means of several real examples coming from a wide scope of applied scientific fields (chemometrics, computer sciences, geophysics). Then we discuss why the standard multivariate linear regression (through a spectrometric functional data set) fails when one considers an explanatory functional variable. In a third section, one proposes to give an idea on what one can do with FDA by giving a sample of functional statistical studies (unsupervised classification, regression on functional variable, curves discrimination, and functional prediction). Our main goal is to stress situations where FDA is potentially pertinent. We conclude by opening some motivating perspectives for educational purposes.

A Few Functional Data Sets and Problematics

Let us first introduce some illustrating examples depicting what is called functional data. The first one comes from the food industry and will be the unifying thread throughout this article. It concerns a quality control problem. One has at hand 215 small pieces of meat (i.e., 215 individuals). Each piece of meat is analyzed through a Tecator Infratec Food and Feed Analyzer. This apparatus allows to obtain, for each unit, one spectrometric curve (i.e., near-infrared spectrum in the wavelength range 850–1050 nm). Finally, one collects 215 spectrometric curves (i.e., $\{X_i(t); t \in [850, 1050]\}_{i=1, \dots, 215}$), one per piece of meat. Of course, in reality, one has at hand 215 discretized curves at 100 measurements t_1, \dots, t_{100} (i.e., for the i th unit one gets the discretized curves $X_i(t_1), \dots, X_i(t_{100})$). However, the difference between two consecutive grid points (i.e., $|t_j - t_{j+1}|$) is so small that the discretized curves (see Figure 1) can be considered as continuous ones as in Figure 2.

At this stage, several questions could interest the practitioners: Can we explore this population of curves? Can we extract pertinent informations? Is it possible to build several groups in order to classify such spectrometric curves? Can we depict the distribution of such a population of curves? (median curve?, modal curve?, etc.).

Other interesting questions arise if there are additional variables. Indeed, throughout an analytical chemical process, it is possible to obtain, for each piece of meat, the fat content (i.e., a scalar response Y), and one has to study the sample $\{X_i(\cdot), Y_i\}_{i=1, \dots, 215}$ where $X_i(\cdot)$ refers to the curve $\{X_i(t); t \in [850, 1050]\}$. A natural question could be: can we explain the Y_i 's from the $X_i(\cdot)$'s? Or, in other words, can we predict the unknown fat content for an incoming new curve $X_{216}(\cdot)$? In the chemometrician community, this type of problem is called calibration problem (see Martens and Naes, 1989, for a good overview on this topic). The practical interest of finding a solution to this regression problem resides in the fact that the analytical chemical process to find Y_i 's is much more expensive than the use

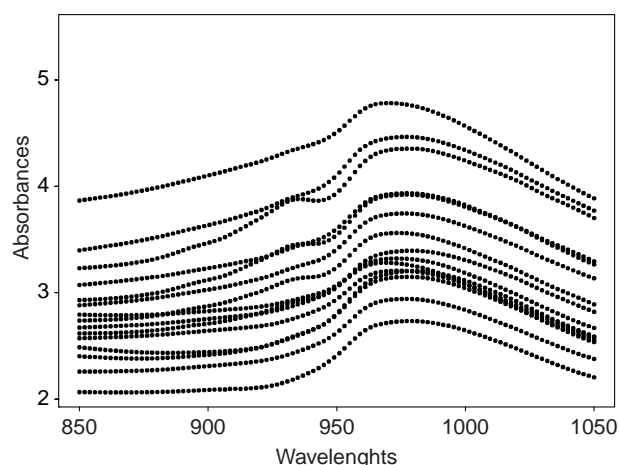


Figure 1 Sample of 15 discretized curves.

of a spectrometer to find $X_i(\cdot)$'s. From a statistical point of view, this is clearly a regression problem where the explanatory variable is the functional one $X(\cdot)$ and the response is the scalar one Y . Another problem could emerge if one splits our population of curves into several groups. For instance, assume that a first group contains the curves with a fat content smaller than 20% and the other with a fat content greater than 20%. So, the class membership for each curve is known. A usual question in such a setting could be the following one: for a new curve $X_{216}(\cdot)$, are we able to assign this curve to a group? Typically, this is a supervised curve classification (or curve discrimination) problem.

One can find numerous studies dealing with functional data sets and devoted to similar or new problematics. For instance, a curve discrimination arises with the following speech recognition data. This data set contains the digitized recordings of speakers of 32 ms duration and concerns five speech frames corresponding to five phonemes transcribed as follows: 'sh' as in 'SHe' (group 1), 'iy' as in 'shE' (group 2), 'dcl' which is the symbol for coding the sound D as in 'Dark' (group 3), 'aa' as the vowel in 'dArk' (group 4) and 'ao' as the first vowel in 'wAter' (group 5). Finally, for each unit, one observes a log-periodogram which is a discretized curve sampled on a fine grid (150 measurements). **Figure 3** displays a sample of 10 log-periodograms for each phoneme (i.e., group).

The question of interest is how to determine the class membership of an incoming logperiodogram, thus leading to a problem of (supervised) classification of a set of curves.

The third example focuses on the monthly time series of sea surface temperature (El Niño) from June 1950 to

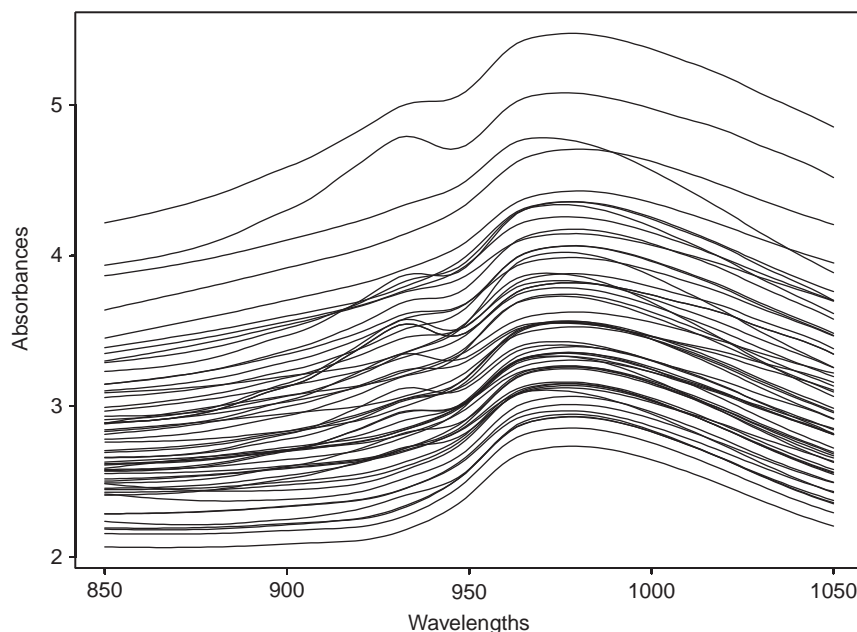


Figure 2 215 spectrometric curves.

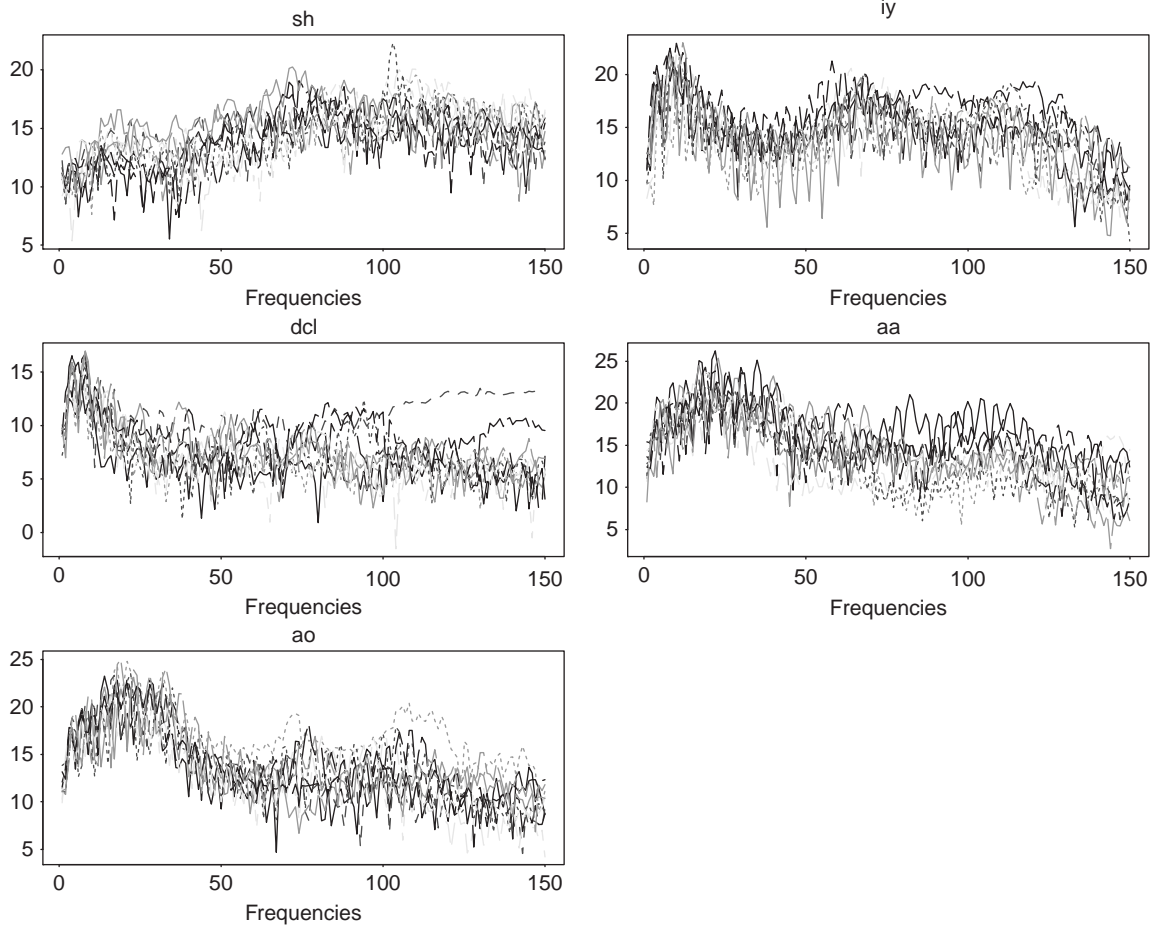


Figure 3 Sample of log-periodograms.

May 2004. The temperatures correspond to an average over the area defined by the coordinates 0–10° South and 90–80° West. **Figure 4** displays this time series, whereas **Figure 5** plots the same data but year by year. In this way, the whole time series is cut into 54 dependent curves (one per year), which can be considered as functional data. Some interesting questions arising in such a situation are the following: Can we forecast future values? Can we predict the maximum temperature for the next year from the previous ones?, etc. This forecasting setting can be viewed as a regression problem with scalar response and functional predictor, but with nonindependent statistical sample.

In all cases, solutions of the presented statistical problems can be found in the recent literature on Food and Drug Administration (FDA).

Why Do Multivariate Methods Fail with Functional Data?

We explain at this point why a standard multivariate method for studying functional data could fail. Let us focus on the spectrometric data presented before. A standard multivariate

point of view would treat each discretized spectrometric curve as one observation of the 100 variables $X(t_1), \dots, X(t_{100})$. Consequently, a standard multivariate linear model for explaining the response Y from the 100 variables is given by $Y = \sum_{j=1}^{100} a_j X(t_j) + \text{error}$. In such a statistical situation, one has ‘to’ estimate the 100 unknown parameters a_1, \dots, a_{100} . Clearly, two crucial problems arise. First high correlation could appear between variables $X(t_j)$ and $X(t_{j'})$ for j and j' close enough (i.e., $\text{corr}(X(t_j), X(t_{j'})) \simeq 1$). To see that, one considers a zoom on the discretized spectrometric curves (**Figure 1**) by focusing on the first 10 measurements (see **Figure 6**). Indeed, **Figure 6** shows that the observations obtained for consecutive measurements are very similar. This results in collinearity in the regression, and hence, to high standard error of estimates.

The second important weakness is that the number of variables can be large in comparison with the sample size. Once again, this leads to poor estimates.

Although some works have been proposed to adapt multivariate methods to this situation, a successful way to overcome the problems described before is to consider the functional data as a functional mathematical object (i.e., $X_i(t_1), \dots, X_i(t_p)$ is the discretization of the random

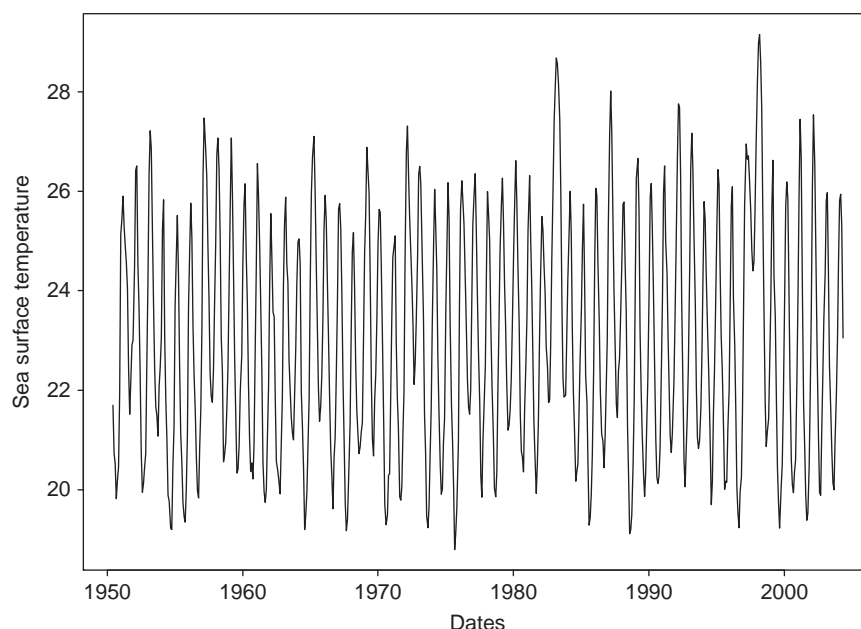


Figure 4 Monthly sea surface temperature.

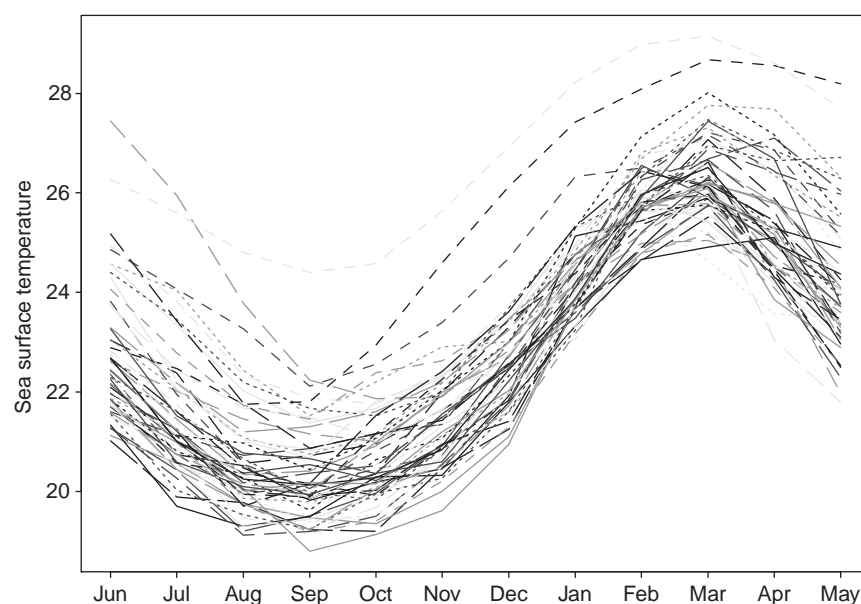


Figure 5 Temperatures year by year.

function $\{X_i(t); t \in T\}$). Unlike multivariate approaches, it is possible to operate on such a functional object some standard mathematical calculus such as differentiation and integration, whereas more sophisticated techniques such as approximation and smoothing can be useful.

Some Modeling Aspects

The main goal of this section is to give some basic ideas on modeling functional data. FDA techniques consist in

developing statistical methods that are able to take into account the richness of the functional features of such data by using functional mathematical tools. For instance, if one considers the sample $\{X_i(\cdot), Y_i\}_{i=1, \dots, m}$ the functional linear regression model is defined by

$$Y_i = \int \rho(t) X_i(t) + \text{error}_i, \quad i = 1, \dots, n$$

(instead of its multivariate version $Y_i = \sum_{j=1}^p a_j X_i(t_j) + \text{error}_i$). A useful mathematical tool consists in expanding the functional parameter $\rho(\cdot)$

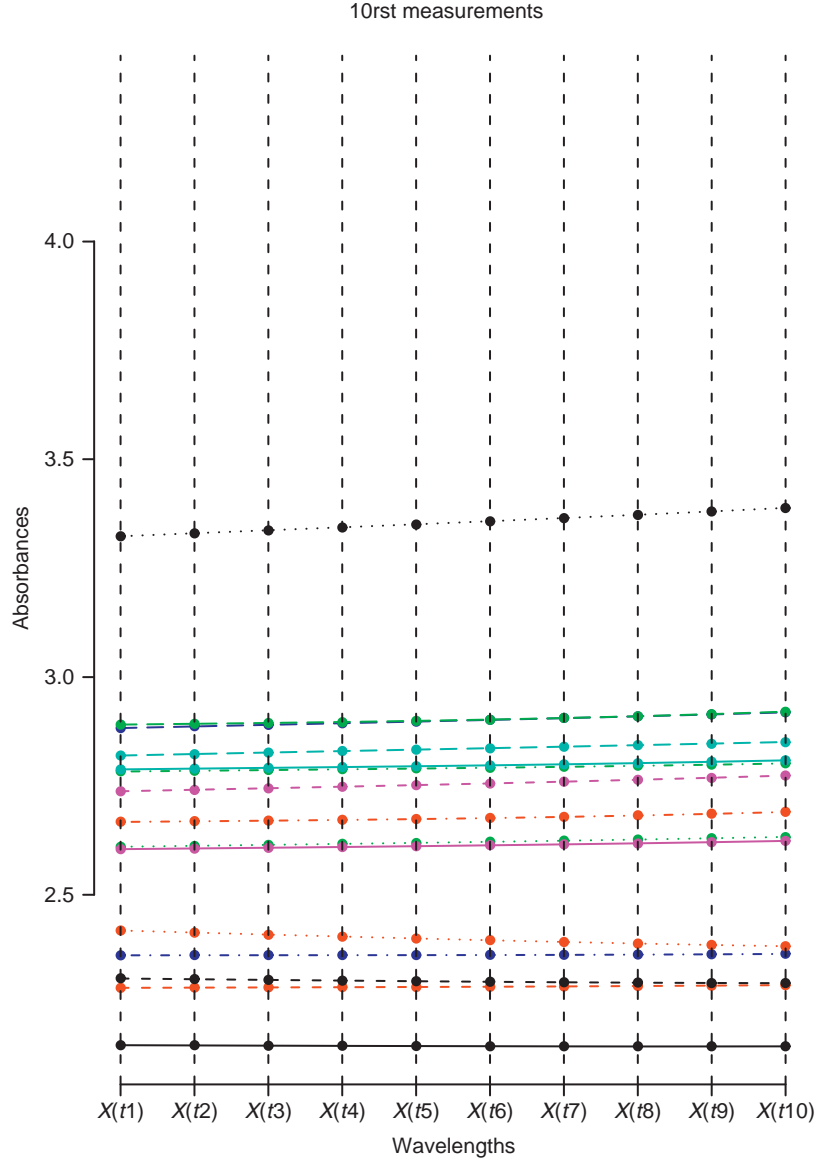


Figure 6 Zoom on spectrometric curves.

onto some fixed function basis (i.e., $\rho(\cdot) = \sum_{k=1}^K b_k B_k(\cdot)$) and also in introducing some smoothness constraints on $\rho(\cdot)$ through a penalized least-squares criterion:

$$C_\rho = \sum_i \left(Y_i - \int \rho(t) X_i(t) dt \right)^2 + \lambda P(\rho)$$

where $P(\rho)$ measures the smoothness of the function ρ (e.g., the L_2 -norm of its second derivative) and λ controls the regularity (the larger the λ , more smooth is the functional parameter). An estimator $\hat{\rho}(\cdot)$ is such that $\hat{\rho}(\cdot) = \min_\rho C_\rho$. If one considers the expansion of $\rho(\cdot)$ in terms of the basis set, minimizing over ρ amounts to minimizing over the vector of coefficients $\mathbf{b} = (b_1, \dots, b_K)$. Let $\hat{\mathbf{b}}$ be the minimizer of C_b over \mathbf{b} : $\hat{\rho}(\cdot) = \sum_k \hat{b}_k B_k(\cdot)$. From a statistical modeling point of view, the functional linear

regression model assumes a linear relationship between the functional variable and the scalar response. Contrary to the multivariate case, one has no graphical tool that is able to display the relationship between a functional variable and a scalar one. In such a context, assuming linearity becomes critical and an alternative consists in proposing more general models, called nonparametric functional models, allowing us to investigate nonlinear relationship. So, a nonparametric functional model is defined by

$$Y_i = R(X_i) + \text{error}_i, i = 1, \dots, n$$

where the regression operator $R(\cdot)$ (which has to be estimated) is assumed to satisfy some mathematical smoothness properties (continuous, lipschitzian, etc.). The linear

regression model can be viewed as a particular case by taking $R(u) = \int \rho(t)u(t)dt$. Concerning the estimation of $R(\cdot)$, kernel-type estimator ideas have been extended to the functional setting:

$$\hat{R}(x) = \sum_{i=1}^n Y_i w_{i,b}(x)$$

with

$$w_{i,b}(x) = K(b^{-1}\delta(X_i, x)) / \left[\sum_{j=1}^n K(b^{-1}\delta(X_j, x)) \right]$$

This estimator is just a weighted average of the responses (i.e., $\sum_i w_{i,b}(x) = 1$). K is a positive decreasing function, $\delta(\cdot, \cdot)$ measures the proximity between two curves (distance, semi-metric, etc.), the smoothing parameter b , called band-width, controls the regularity of $\hat{R}(\cdot)$. So, the principle of such a kernel estimator is very simple: closer x is to X_b , larger is the weight assigned to Y_i (i.e., $w_{i,b}(x)$). Both the bandwidth and the index of proximity $\delta(\cdot, \cdot)$ have to be selected (generally through a criterion based on the quality of the predictions). It is clear that the main feature of the kernel-type estimator is to focus on local properties of the distribution of the functional variable. From a probabilistic point of view, one studies this local behavior through the small balls probabilities defined by $\text{Prob}(\delta(X_i, x) \leq b)$.

It is worth noting that kernel estimator ideas can be applied to settings other than the regression one. For example, the simple kernel-type estimator

$$\hat{f}(x) = \text{const} \times \sum_{i=1}^n K(b^{-1}\delta(X_i, x))$$

is a useful statistical tool for deriving what we call the modal curve (i.e., $x_{\text{mod}} = \sup_x \hat{f}(x)$). One can also define the median curve as the one which minimizes $\sum_{i=1}^n \delta(X_i, x)$ over x . These distribution features are good tools for depicting samples of curves and an interesting by-product concerns the unsupervised classification.

Several other statistical methodologies such as generalized functional linear model, conditional distribution given a functional variable (conditional mode, conditional quantiles, etc.), curves discrimination, survival analysis with functional explanatory variable, testing procedures, etc., have been investigated.

FDA in Action

Spectrometric Data

It is time to give an idea on what one can do with FDA. At first, let us consider the functional data concerning the population of spectrometric curves. We investigate the

local behavior (i.e., small ball probabilities evaluated at each curve $X_i(\cdot)$) of the sample of curves in order to split it into several subgroups. The homogeneity of a subgroup is computed by means of features like median curve, modal curves, etc., and a general homogeneity index is computed for the obtained partition. Finally, an automatic unsupervised classification method for curves is built (the algorithm stops as soon as the gain in terms of homogeneity between two successive partitions is smaller than a threshold). For the spectrometric curves, the proximity index δ is based on their second derivatives and defined by $\delta(X_i, X_j) = \sqrt{\int (X_i''(t) - X_j''(t))^2 dt}$; the derivatives are calculated by using a numerical approximation. This choice of δ is based on the experience of chemometricians and also on a data-driven procedure. The result of this functional classification based on this index leads to three groups. **Figure 7** plots for each obtained group, the original curves, their second derivatives, and the corresponding modal curve (the curve of the sample which maximizes a kernel-type estimator of the distribution function of the population of curves). This example gives a good idea on the usefulness of FDA methods. In particular, modal curves are very interesting features for depicting a distribution of curves.

Now, if one takes into account the responses (i.e., the fat content), the problematic changes as one can try to predict the fat content from the spectrometric curve. This is typically a regression with scalar response and explanatory functional variable. Precursor works as regression on principal components, ridge regression, or partial least-squares extended standard multivariate methods. Recent advances in FDA extended these methodologies by exploiting systematically mathematical background of function spaces. In particular, a nonparametric FDA (with kernel estimator) allows to make fat content predictions as the one plotted in **Figure 8**. The solid line displays a perfect prediction: the proximity of the points with this line indicates a good accuracy of the predictions obtained with FDA. Instead of collecting the accurate fat content, one can have at hand only two groups (G_1 and G_2). For instance, the first one contains the curves corresponding to a fat content smaller than 20% and the second one the other curves. This is typically a curve discrimination (or supervised classification) problem: for an incoming curve X_{n+1} , one has to assign it to one group. Once again, one can find in FDA several methods for discriminating these spectrometric curves into these two groups. Such supervised curves classification methodologies can give very low rates of misclassification. For instance, if one implements the kernel estimator of the posterior probabilities $\text{Prob}(Y_{n+1} \in G_k | X_{n+1})$, one assigns the new curves X_{n+1} to the class of highest estimated posterior probability. This curve discrimination method leads us to a rate of around 2%; one affects (on average) to the right group 98 spectrometric curves over 100.

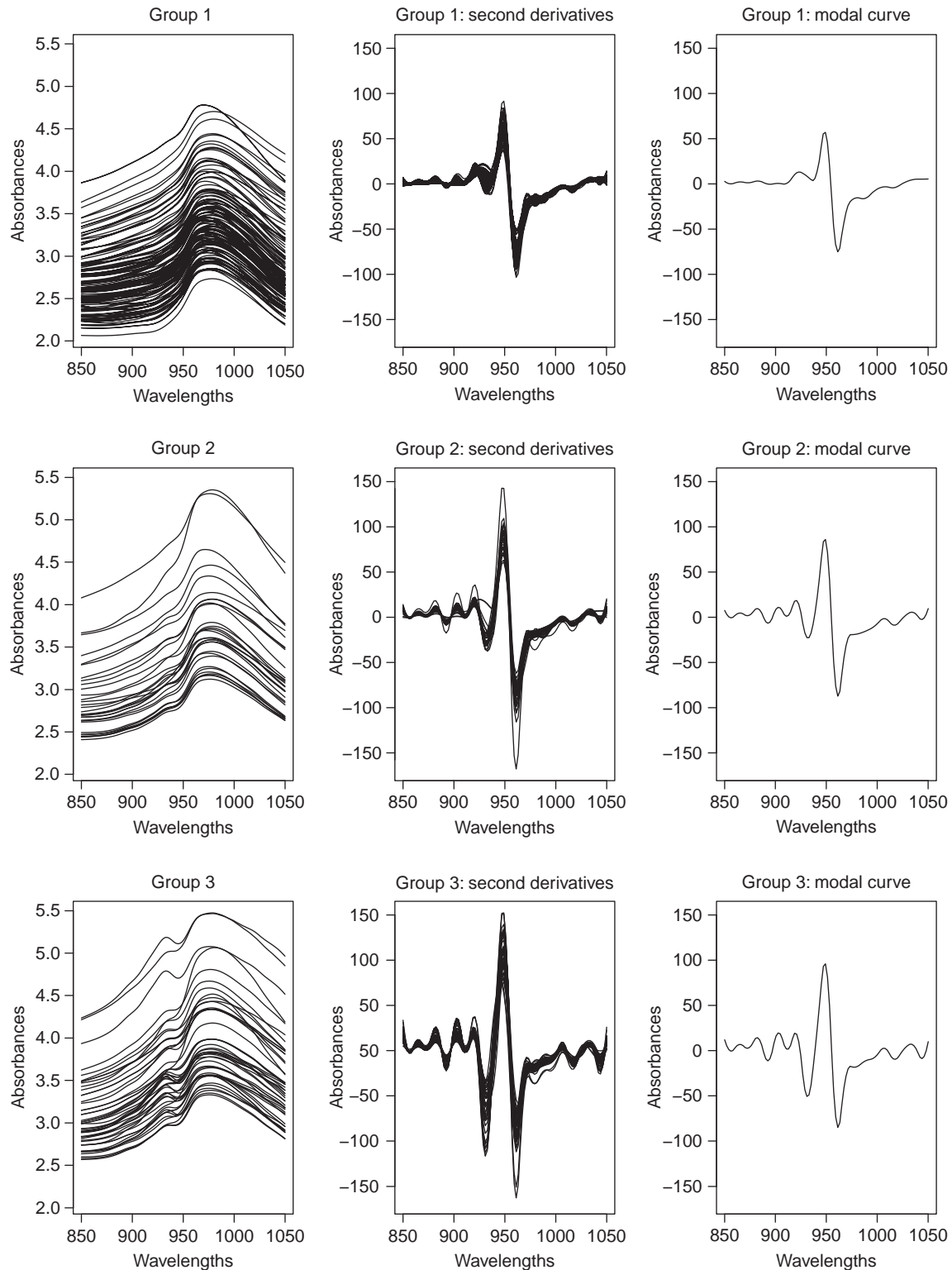


Figure 7 Unsupervised curves classification.

Phoneme Data

If one focuses on the phoneme data set (i.e., log-periodograms) which is also a curves discrimination problem, one can get a rate of misclassification around 8% (according to the functional methodology implemented).

El Niño Time Series

Let us consider now the time series El Niño. In this forecasting situation, FDA allows one to predict future values in an accurate way. **Figure 9** gives an idea of the pertinence of using such statistical methods. The solid

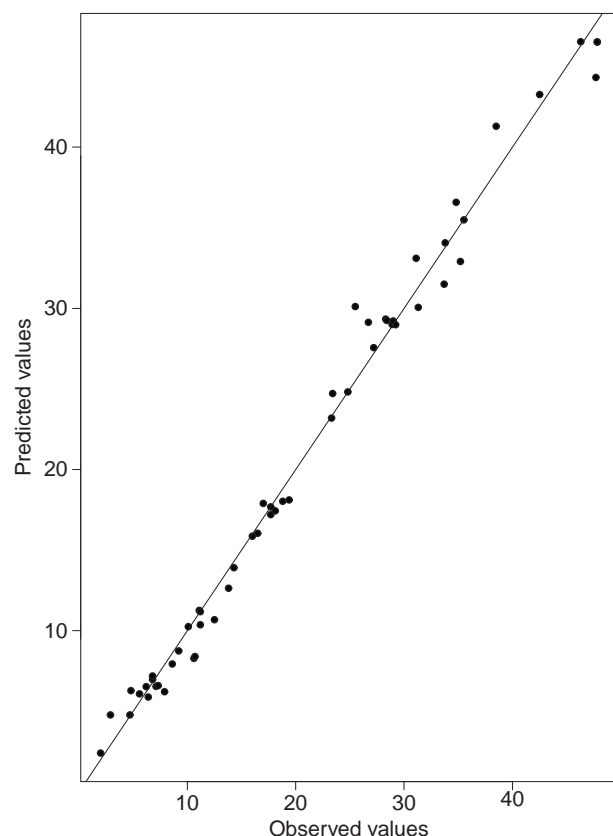


Figure 8 Observed responses vs predicted ones.

line corresponds to the observed values for the last year available in the used El Niño sample and the dashed one displays the forecasted temperatures.

Of course, the previous examples are not an exhaustive presentation on what one can do with FDA. There are several other settings not considered here. An important one occurs when a functional data set proposes a population of curves with measurements changing from one unit to another one (i.e., one has at hand the values $X_i(t_{i_1}), \dots, X_i(t_{i_{p_i}})$, where the measurements $t_{i_1}, t_{i_2}, \dots, t_{i_{p_i}}$ depend on the individual i). This means that, from a standard multivariate point of view, the variables $X_i(t_{i_1}), \dots, X_i(t_{i_{p_i}})$ (even their number) are not the same from one individual to another one, which prevents the use of any standard multivariate statistical analysis. Such data sets are called unbalanced functional data. For instance, this situation can occur when the starting point of the measurements is not the same for each unit. In order to analyze such data sets, preprocessing procedures like curves registration methods or time-warping techniques have been developed. Another interesting case of functional data appears when, for each curve, one gets a few measurements. In such a situation, one says that the functional data are sparse. Such functional data sets require specific methodologies. Other useful developments covering a large

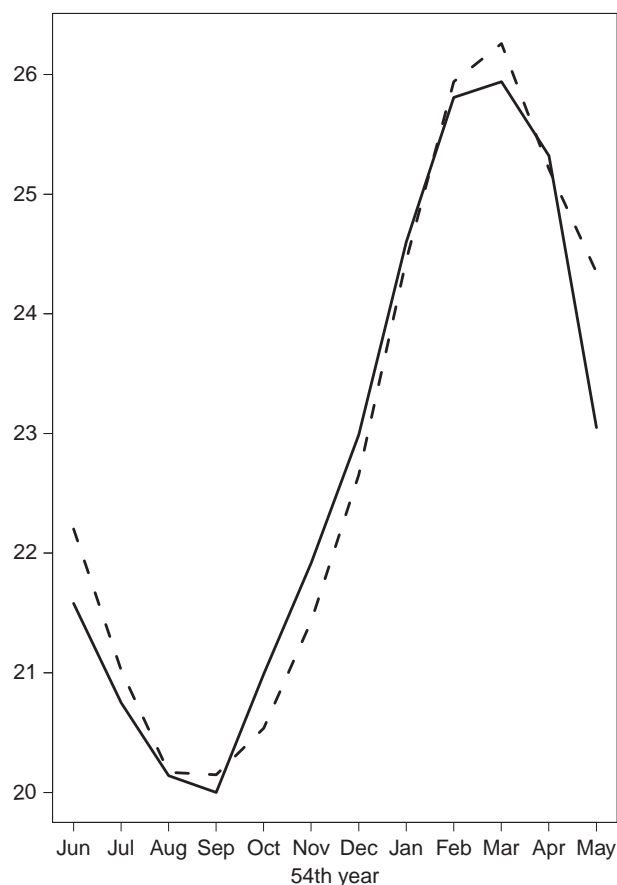


Figure 9 Forecasted/observed values.

number of statistical situations and extending the generalized linear model to the functional data setting have been investigated. Good recent overviews on FDA methodologies as well as potentialities in terms of applications can be found in special issues of various statistical journals entirely devoted to this area: Davidian *et al.* (2004), González-Manteiga and Vieu (2007), and Valderrama (2007) (see also the recent proceedings edited by Dabo-Niang and Ferraty, 2008).

Conclusion

FDA covers a wide area of statistics. Indeed, according to recent works on this topic, one can imagine that several standard statistical methods will be extended in order to take into account functional data. One can also imagine the huge potential of such new statistical methodologies in terms of applications. Indeed, the newness of FDA implies that numerous fields of potential applications are underdeveloped or not investigated up to now. This is the case about educational data. However, phoneme

functional data could be a good example in relation with educational domain. It is clear that one can obtain measurements on children and with FDA techniques, identify some of them with problems such as speech defects. Any study which allows one to get a continuous trajectory for each individual can be analyzed through functional data methods.

It is clear that we have given here a concise overview on FDA in a pedagogical manner. Therefore, only functional data as curves have been presented. In order to complete this curve-based point of view on functional data, it is worth mentioning that FDA can deal with populations of any more high dimensional objects. For instance, with the progress of technologies, it is now possible to collect samples of images (2D-images, multispectral images, hyper-spectral images, etc.). These result in spatio-functional data. This promising recent area in terms of applications is underdeveloped, but certainly future investigations will enlarge our knowledge in this domain.

Bibliography

- Bosq, D. (2000). *Linear Processes in Function Spaces. Theory and Applications. Lecture Notes in Statistics*, 149. New York: Springer.
- Dabo-Niang, S. and Ferraty, F. (2008). *Functional and Operatorial Statistics*. Heidelberg: Physica.
- Davidian, M., Lin, X., and Wang, J. L. (2004). Introduction to the emerging issues in longitudinal and functional data analysis (with discussion). *Statistica Sinica* **14**(3), 613–629.
- Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis. Theory and Practice. Springer Series in Statistics*. New York: Springer.
- González-Manteiga, W. and Vieu, P. (2007). Introduction to the special issue on statistics for functional data. *Computational Statistics and Data Analysis* **51**(10), 4788–4792.
- Martens, H. and Naes, T. (1989). *Multivariate Calibration*. New York: Wiley.
- Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis. Methods and Case Studies. Springer Series in Statistics*. New York: Springer.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis. Springer Series in Statistics*, 2nd edn. New York: Springer.
- Valderrama, M. (2007). Introduction to the special issue on modelling functional data in practice. *Computational Statistics* **22**(3), 331–334.
- Crambes, C., Kneip, A., and Sarda, P. (forth). Smoothing spline estimators for functional linear regression. *Annals of statistics*.
- Ferraty, F. and Vieu, P. (2006). NPFDA R/S+ library. <http://www.lsp.ups-tlse.fr/staph/npfda> (accessed August 2009).
- Frank, I. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools (with discussion). *Technometrics* **35**(2), 109–148.
- Gasser, T. and Kneip, A. (1995). Searching for structure in curve samples. *Journal of the American Statistical Association* **31**, 1179–1188.
- Hastie, T., Buja, A., and Tibshirani, R. (1995). Penalized discriminant analysis. *Annals of statistics* **23**(1), 73–102.
- James, G. M., Hastie, T. J., and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87**(3), 587–602.
- Ramsay, J. O. (2005). FDA Matlab/R/S+ library. <http://www.psych.mcgill.ca/misc/fda> (accessed August 2009).
- Ramsay, J. O. and Li, X. (1998). Curve registration. *Journal of the Royal Statistical Society: Series B* **60**(2), 351–363.
- Yao, F., Müller, H. G., and Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**(470), 577–590.

Further Reading

Relevant Websites

- <http://www.lsp.ups-tlse.fr> – Activities of the STAPH Working Group on Functional and Operatorial Statistics.
- <http://ego.psych.mcgill.ca> – McGill University, Functional Data analysis, a useful, great and pioneer website for studying functional data with lots of complementary methods (see the excellent monographs of J. Ramsay and B. Silverman).
- <http://www.r-project.org> – The R Project for Statistical Computing, website devoted to R (useful language and environment for statistical computing and graphics; available as free software).
- <http://www.lsp.ups-tlse.fr> – website oriented towards practitioners for implementing nonparametric functional (companion website of the book Nonparametric Functional Data Analysis of F. Ferraty and P. Vieu).

Statistical Inequalities

H M Mahmoud and J L Gastwirth, The George Washington University, Washington, DC, USA

© 2010 Elsevier Ltd. All rights reserved.

Markov's Inequality

Discovered around 1905, Markov's inequality bounds a probability for a non-negative random variable by its mean. The inequality is usually stated in the following form.

Let X be a nonnegative random variable, with finite mean μ . Then, for any $a > 0$,

$$P(X \geq a) \leq \frac{\mu}{a} \quad [1]$$

We shall not give many proofs in this article, but shall provide this derivation, which is indicative of the type of proof involved in a class of inequalities related to or derived from this basic one.

The proof proceeds by decomposing the range of possible values for the expression of the mean (expected value) into two parts. For a continuous random variable X with density $f(x)$,

$$\begin{aligned} E[X] &= \mu \\ &= \int_{x=-\infty}^{\infty} xf(x)dx \\ &= \int_{x=0}^a xf(x)dx + \int_{x=a}^{\infty} xf(x)dx \\ &\geq \int_{x=a}^{\infty} xf(x)dx \\ &\geq \int_{x=a}^{\infty} af(x)dx \\ &\geq aP(X \geq a), \end{aligned}$$

which establishes the inequality.

The proof is the same for discrete random variables, with the integrals being replaced with sums. In fact, in the more general theory, the proof is the same, with all integrals being assumed to be Lebesgue integrals (and then, discrete sums and standard integrals follow as special cases).

To illustrate the bound obtained from Markov's inequality, suppose X is a Poisson(3) random variable (with mean 3 and variance 3), following the probability law

$$P(X = k) = \frac{3^k e^{-3}}{k!}, \text{ for } k = 0, 1, 2, \dots$$

The probability that X exceeds its mean, is the probability that X is at least 4. With $\mu = 3$ and $a = 4$, Markov's inequality (1) gives the bound

$$P(X \geq 4) \leq \frac{3}{4} = 0.75$$

In this case, one can calculate the exact probability from the Poisson distribution, namely

$$P(X \geq 4) = \sum_{k=4}^{\infty} \frac{3^k e^{-3}}{k!} \approx 0.32768,$$

well below Markov's bound. In general, one does not expect Markov's inequality to be sharp, as it is general; it is valid for all distributions, and only uses the mean. Two different distributions can have the same mean but quite different tail probabilities.

Chebyshev's Inequality

A special interpretation of Markov's inequality yields an inequality attributed to Chebyshev, which is widely used.

Consider the random variable $|X - \mu|$ (which is, of course, always non-negative), where μ is the mean of X . Then

$$P(|X - \mu| \geq a) = P((X - \mu)^2 \geq a^2) \leq \frac{E[(X - \mu)^2]}{a^2} = \frac{\sigma^2}{a^2} \quad [2]$$

where σ^2 is the variance of X .

The variance form is the most common in application, but it should be noted that it is a special case of a more general inequality, which takes the form

$$P(|X - \mu| \geq a) \leq \frac{E[\phi(X)]}{\phi(a)}$$

where $\phi(u)$ is a symmetric nonnegative function that is increasing on $(0, \infty)$. For example, with $\phi(u) = u^4$, and $\mu = 0$, we have a bound involving the fourth moment:

$$P(|X| \geq a) \leq \frac{E[X^4]}{a^4}.$$

Generally speaking, the inequality is sharper when higher moments are used; that is why the usual Chebyshev's inequality [2] is preferable to Markov's inequality [1].

One does not expect Chebyshev's inequality to always be sharp, since it is valid for all distributions. For most distributions, and values of a , formula [2], gives a conservative upper bounds. For example, suppose Z is the standard normal random variate, with mean 0 and variance 1. According to Chebyshev's inequality, the probability that Z falls within two standard deviations away from 0, is bounded above:

$$P(|Z| \leq 2) \geq 0.75.$$

Using the fourth moment instead, one obtains a closer bound

$$P(|Z| \leq 2) \geq \frac{E[Z^4]}{2^4} = \frac{3}{16} = 0.8125,$$

as $P(|Z| \leq 2) = 0.95$ (approximately).

It should also be pointed out that Chebyshev's inequality provides the sharpest inequality that is valid for all distributions. This can be demonstrated by the following example, where the inequality is attained. Consider the three-valued family of random variables

$$X = \begin{cases} \mu + k\sigma, & \text{with probability } \frac{1}{k^2}; \\ 0, & \text{with probability } 1 - \frac{1}{2k^2}; \\ \mu - k\sigma, & \text{with probability } \frac{1}{k^2}. \end{cases}$$

The mean of this random variable is σ and the variance is σ^2 . Typically, one desires a bound on events of the type $|X| \geq k\sigma$. Chebyshev's inequality gives

$$P(|X| \geq k\sigma) \leq \frac{1}{k^2}.$$

In this specific example,

$$P(|X| \geq k\sigma) = P(|X| = k\sigma) = \frac{1}{k^2}.$$

so, Chebyshev's inequality is attained.

Gauss' Inequality

We mentioned that in the absence of additional information, Chebyshev's inequality is sharpest. Nonetheless, additional knowledge, such as symmetry of the distribution, or unimodality, leads us to craft sharper inequalities. When both properties are known to hold, there is the classic inequality, due to Gauss: for a symmetric unimodal random variable X , with mean μ and variance σ^2 , the inequality

$$P(|X - \mu| \geq t) \leq \frac{4\sigma^2}{9t^2}. \quad [3]$$

holds for $t^2 \geq \frac{4}{3}\sigma^2$. Note that this is sharper than Chebyshev's inequality, because the factor $\frac{4\sigma^2}{9t^2}$ is less than $\frac{\sigma^2}{t^2}$ (the bound in Chebyshev's inequality), but keep in mind that the validity of Gauss inequality is restricted to symmetric unimodal distributions.

Again consider the event $P(|Z| \leq 2)$, where Z is a standard normal variable (which is symmetric and unimodal). Using Gauss' inequality [3] gives us the lower bound

$$P(|Z| \leq 2) \geq \frac{8}{9} \approx 0.8889,$$

which is a noticeable improvement over the lower bound of 0.75 obtained from Chebyshev's inequality and is quite close to the actual value of about 0.95.

Cantelli's Inequality

It should be noted that there is a less-known inequality similar in nature to Chebyshev's: Cantelli's inequality (also known as the one-sided Chebyshev's inequality). Let X be a random variable with mean μ and variance σ^2 . For any $a > 0$,

$$\begin{aligned} P(X - \mu \geq a) &= P(X - \mu + x \geq a + x) \\ &\leq P(|X - \mu + x| \geq a + x) \\ &\leq \frac{E[(X - \mu + x)^2]}{(a + x)^2} \\ &= \frac{\sigma^2 + x^2}{(a + x)^2}, \end{aligned}$$

valid for every $x > 0$. In particular, it is valid for the positive value of x that minimizes the right-hand side. By differentiation, the value $x_0 = \sigma^2/a$ can be shown to minimize the function $(\sigma^2 + x^2)/(x + a)^2$. It follows that

$$P(X - \mu \geq a) \leq \frac{\sigma^2 + x_0^2}{(x_0 + a)^2} = \frac{\sigma^2}{\sigma^2 + a^2}, \text{ for any } a \geq 0. \quad [4]$$

In terms of distances measured in standard deviations, the inequality is

$$P(X - \mu \geq k\sigma) \leq \frac{1}{k^2 + 1}.$$

Cantelli's inequality can be tight. Consider, for example, a Bernoulli ($\frac{1}{2}$) random variable B , which assumes the value 0 with probability $\frac{1}{2}$ and the value 1 with probability $\frac{1}{2}$, and take $a = \frac{1}{2}$. Cantelli's inequality gives

$$P(B - \frac{1}{2} \geq \frac{1}{2}) \leq \frac{1}{2}.$$

However,

$$P(B - \frac{1}{2} \geq \frac{1}{2}) = P(B \geq 1) = P(B = 1) = \frac{1}{2}.$$

so, Cantelli's bound is attained.

Let us revisit the Poisson(3) example we discussed in the section on Markov's inequality [1], and consider the event that it exceeds 5. According to Cantelli's inequality [4], a random variable X with this distribution satisfies

$$P(X \geq 5) = P(X - 3 \geq 2) \leq \frac{3}{3 + 4} \approx 0.42857,$$

while Markov's inequality gives

$$P(X \geq 5) \leq \frac{3}{5} = 0.6,$$

while both exceed the true value 0.185 (approximately), clearly Cantelli's inequality is sharper.

Inequalities for Moments

So far, the common inequalities for the probability of an event have been described and they all involve moments.

It is useful to have inequalities for the moments themselves. We shall very briefly sketch the most salient ones. These inequalities appear more often in advanced topics and measure-theoretic proofs, where a careful argument is usually made about the existence of the moments involved. In the following, the moments are assumed to exist and are finite.

Jensen's Inequality

This inequality states that for any random variable X and convex function ϕ ,

$$\phi(E[X]) \leq E[\phi(X)].$$

A standard use of this inequality assures that the variance

$$\text{Var}[X] = E[X^2] - (E[X])^2$$

is non-negative. This follows from convexity of the function $\phi(u) = u^2$.

Hölder's Inequality

Suppose $p, q > 0$, such that $\frac{1}{p} + \frac{1}{q} = 1$. For two random variables X and Y , we have,

$$E[|XY|] \leq (E[|X|^p])^{1/p} (E[|Y|^q])^{1/q}.$$

The quantity $(E[|X|^p])^{1/p}$ is called the p th norm of X , and is often denoted by $\|X\|_p$ (it has all the properties of distance function in the space $\|X\|_p$ of random variables). Hölder's inequality can then be expressed in the notation

$$E[|XY|] \leq \|X\|_p \|Y\|_q.$$

The special case $p = q = 2$ is known as the Cauchy–Schwarz inequality.

Minkowski's Inequality

Hölder's inequality takes a multiplicative form. It is used in the proof of a moments inequality in additive form, called Minkowski's inequality. This inequality generalizes the triangle inequality. For $p \geq 1$,

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Lyapunov's Inequality

Lyapunov's inequality follows immediately from Hölder's. For any $0 < r < s$,

$$\|X\|_r \leq \|X\|_s.$$

It is usually employed in arguments used to show that specific moment exists. Once it is known that s th moment exists (i.e., it is finite) for some $s > 0$, all lower order moments exist. For instance, if one knows that the fourth moment exists, one is assured of the existence of the first three moments. A typical application is when the second moment exists. In this case, the first moment (the mean) also exists.

Concluding Remarks

We discussed the most common types of inequalities for the probabilities of random variables. There are other families of inequalities that apply to sums of independent random variables. Kolmogorov's inequality stands out as an epitome of the members of the family (see Billingsley, 1995). Another type of inequalities uses the moment-generating function to bound probabilities. Two important inequalities from this class are Chernoff's and Hoeffding's inequalities (see Chernoff, 1952; Hoeffding, 1963).

See also: Measures of Central Tendency; Measures of Dispersion, Skewness and Kurtosis; Probability Theory; The Normal Distribution and its Applications.

Bibliography

- Billingsley, P. (1995). *Probability and Measure*. New York: Wiley.
- Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics* **23**, 493–507.
- Chow, Y. and Teicher, H. (1997). *Probability Theory: Independence, Interchangeability, Martingales*, 3rd edn. New York: Springer.
- Chung, K. (1974). *A Course in Probability Theory*. Orlando: Academic Press.
- Gastwirth, J. (2008). The U.S. Supreme court finds a statute's description of a simple statistical measure of relative disparity 'ambiguous' allowing the secretary of education to interpret the formula. *Zuni Public School District 89 v. U.S. Department of Education II. Law, Probability and Risk* **7**, 225–248.
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58**, 13–30.
- Kendall, M., Stuart, A., and Ord, K. (1987). *Kendall's Advanced Theory of Statistics*. New York: Oxford University Press.
- Ross, S. (2006). *A First Course in Probability*. Upper Saddle River, NJ: Pearson Prentice Hall.

Statistical Paradoxes

Y-K Tu, University of Leeds, Leeds, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

There are many statistical paradoxes in the literature. They gain the reputation as paradoxes, because they often generate confusion or a misleading interpretation of simple data presentation for the unprepared mind. In particular, these paradoxes challenge our ability to decipher causes and effects from nonexperimental observations. This article discusses two well-known statistical paradoxes that have pervaded quantitative research for more than a century. One is the reversal paradox also known as Simpson's paradox in contingency table analysis or Lord's paradox in the analysis of covariance, and the other is regression to the mean. These paradoxes are not only interesting puzzles for intellectual amusement but potentially have a serious impact on the interpretations of evidence from observational educational studies and, more importantly, on public policymaking. In this article, we discuss how these paradoxes occur and how to resolve them.

Simpson's Paradox, Lord's Paradox, and Reversal Paradox

Simpson's paradox, or more appropriately Yule's paradox, is a well-known statistical phenomenon. It is observed when the relationship between two categorical variables is reversed after a third variable is introduced to the analysis of their association, or alternatively where the relationship between two variables within subgroups is different from that of the aggregated data. Although first discussed by Karl Pearson in 1899, it is George Udny Yule, once Pearson's assistant, who provides a detailed treatment of this problem in 1903. A famous example of Simpson's paradox in education is that given by Bickel and co-workers in 1975, regarding real admission rates of male and female students to the University of California, Berkley. Overall admission rates were significantly lower for females than males, yet, females tended to have higher admission rates than males within individual departments. The explanation is that the number of female applicants as a proportion of the total female applicants tended to be high within departments that were also hard to gain entry to and low in those that were easier to enter.

Lord's paradox was named after two short articles published in the psychology literature by Frederick M Lord regarding the use of analysis of covariance (ANCOVA) within nonexperimental studies. Lord's paradox refers

to the relationship between a continuous outcome and a categorical covariate being reversed when an additional continuous covariate is introduced to the analysis. The additional covariate is often a measure made at the start of a study, or at baseline within a longitudinal study, where the outcome is the same variable measured some time later (e.g., following an intervention). The aim is to measure change in the outcome by adjusting for the baseline measurements, and the categorical covariate might be the intervention/control groups – this is the familiar design for ANCOVA. However, Lord's paradox may occur in circumstances not related to the use of ANCOVA, where a covariate is measured at baseline or at the same time as the outcome.

Using Lord's original hypothetical example, consider a group of first-year university students, where body weight is measured on entrance and again 1 year later, to investigate whether or not refectory food has any impact on their weight, and moreover whether or not any such impact differs by gender. Body weight data are then subjected to two types of statistical analysis. The first is to use a paired *t*-test for the difference in weight between measurement occasions. Results reveal no significant difference across all students, and similarly for each gender separately. The independent *t*-test also shows no differences in the changes of weights between genders. One might therefore conclude that refectory food has no significant impact upon body weight among male and female students. The second approach is to use ANCOVA, since it is observed that there are weight differences among males and females at baseline and since ANCOVA adjusts for baseline differences. Results reveal that, after adjustment for baseline weights, males gained weight on average, whereas females lost weight. This is clearly contradictory to the conclusions of the first analysis. The question is which is correct. Specifically, did refectory food have any differential gender effect on student weights? This controversy was first discussed in 1910 between Karl Pearson and Arthur C Pigou when they debated the role of parental alcoholism and its impact on the performance of children; the whole incident is discussed by Stephen Stigler in the first chapter of his book, *Statistics on the Table*.

Both Simpson's paradox and Lord's paradox are the manifestations of the generic reversal paradox, that is, the relationship between two variables is reversed when a third variable is adjusted for. In Simpson's paradox the three variables involved are all categorical variables, whereas in Lord's paradox two are continuous and one

is a categorical variable. Reversal paradox can also occur among three continuous variables. The reverse of the associations between variables after the adjustment of another one is not uncommon in statistical modeling: a problem known as collinearity or multicollinearity. From a purely mathematical viewpoint, the direction of association between two variables (one outcome and one explanatory variable) is determined by their correlations with other explanatory variables in the model, and it should therefore be anticipated that the relationship between two variables in simple regression or correlation may sometimes become different from that in multiple regression or partial correlation in terms of the effect size or even the direction.

So why do we still feel fascinated or confused when the reversal paradox does occur? Although the mathematics behind the reversal paradox is not overly complex, it can be quite cumbersome to make causal interpretation of the results in the context of real life research. For instance, in a recently published study on the effect of pupil movement on school differences in educational achievement, it has been found that the normalized mathematics scores assessed at key stage 2 (KS2, year 6 in primary schools in England) were positively associated with pupils' ages (in months) at this stage of schooling (0.029, standard error = 0.003) in 9226 pupils of 241 schools in Staffordshire, England. The authors stated that this finding is expected as it is known that age is related to attainment at this stage of schooling because older children may develop better cognitive ability. However, this positive association was reversed when the test scores at key stage 1 (KS1, assessed at year 2) were included as a covariate in the so-called traditional value-added analysis. The association between KS2 mathematics scores and ages was reversed from 0.029 to -0.014 ($SE = 0.002$), whereas the association between KS2 and KS1 mathematics was, as expected, positive (0.754, $SE = 0.006$). The interpretation given by the study is that given their KS1 performance younger children do better, indicating that they tend to catch up over this period. A similar finding has been found previously. Statistically, this reverse in the direction of association between KS2 performance and age is not hard to explain: (1) the positive correlation between age and math performance was stronger at KS1 than at KS2 and (2) KS1 and KS2 math scores were also positively correlated. Hence, the positive association between age and KS2 math score in the multiple regression was reversed due to collinearity. The standard interpretation as suggested by textbooks on regression analysis would be that after adjusting for their KS1 math score, there is a difference of 0.014 standard deviation in KS2 math score between pupils with 1-month difference in their age. In other words, the advantage of being older in school in terms of KS2 math score in the simple regression analysis seems to turn into a disadvantage in multiple regression analysis after the adjustment of

KS1 math score. The key question is: are older children doing better or worse than younger ones in KS2 math performance?

To simplify our discussion, let us assume all other confounders, for example, parents' ethnic group, and educational and socioeconomic background, are not correlated with both the math performance and pupils' ages. According to the simple regression analysis, older children, for example, who were born in September, October, or November, were doing better in KS2 math performance than those who were born in August, July, and June of the following year. Therefore, teachers may be expected to give additional attention to younger children. However, according to the results of the multiple regression analysis, the reverse is true, and teachers are expected to give more attention to older children in the class. Which is true?

Let us consider the explanation given by the authors of the original study. As the focus of the original study is not the relationship between pupils' age and math achievement, the authors only gave their interpretation in passing. According to the authors, the reversal of the association between age and KS2 math score after adjusting for KS1 math score seems to suggest that younger children catch up over the period between year 2 and 6, that is, the advantage of being older was more evident in early schooling than later, probably because up to a certain age, all children will on average develop the same level of cognitive ability in mathematics. Therefore, the authors' interpretation seems to be quite plausible. However, this does not directly answer the question: are older children doing better or worse at KS2? And what is the relationship between pupils' age and their math performance at KS2?

The paradox is mainly caused by the difference in the results of unconditional (simple regression) and conditional (on KS1 math score in multiple regression) analysis. Had the regression coefficient not been reversed, it would probably not be a problem at all. The main challenge is to properly interpret the conditional relationship between pupils' age and KS2 math score. If we imagine that pupils' age is a binary variable (i.e., old vs. young), statistically speaking, the unconditional analysis is to compare the sample means of KS2 math scores between different age groups, but the conditional analysis is to compare the adjusted means (i.e., the predicted means for the age groups if they had the same KS1 score) between the age groups. In the latter analysis, we statistically coerce children of different age groups to have the same KS1 math score and then compare their expected means. **Figure 1** illustrates the differences in the comparisons of group means between the unconditional and conditional analyses. The average math test scores of older children at KS1 and KS2 (O_{KS1} and O_{KS2} , respectively) were better than those of younger children at both KS1 and KS2 (Y_{KS1} and Y_{KS2} , respectively).

However, when KS1 was adjusted for as a covariate, the adjusted mean of KS2 for younger children moved up and right along the parallel regression slope (the gray line on the top) and the adjusted mean of KS2 for older children moved down and left along the slope (the other gray line in the figure). The difference in the adjusted means was the difference in the vertical height of the two parallel slopes (the short black line between the ends of the two arrows).

From a causal point of view, did pupils' age cause better or poorer performance in KS2 math? **Figure 2** is a directed acyclic graph which has become popular in statistics and epidemiology in recent years with regard to the discussion of making causal inference in nonexperimental studies. To keep our discussion simple, we again assume that all other confounding variables can be safely ignored. The arrows in **Figure 2** show that Age causes better math scores at both KS1 and KS2. Note that although Age is the pupils' age at key stage 2, older (younger) children in the class were already older (younger) when they first started school. The arrow from KS1 to KS2 means that pupils who achieved better scores at KS1 also performed better at KS2. As KS1 is on the causal pathway from Age to KS2, and therefore should not be adjusted for, if our aim is to assess the overall impact of Age on KS2. The adjustment of KS1 gives rise to a counterfactual interpretation that if all pupils had the same KS1 score, older children did less well than younger ones, whereas older children actually achieved better KS1 score.

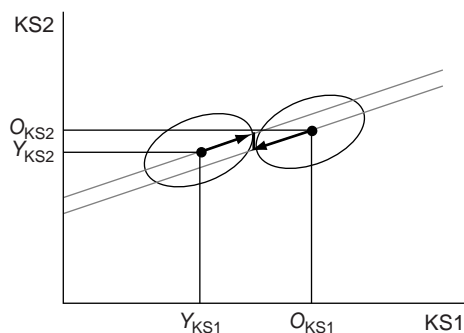


Figure 1 Illustration of reversal paradox – KS1: mathematics scores at key stage 1; KS2: mathematics scores at key stage 2; O_{KS1} : KS1 for older children; Y_{KS1} : KS1 for younger children; O_{KS2} : KS2 for older children; and Y_{KS2} : KS2 for younger children.

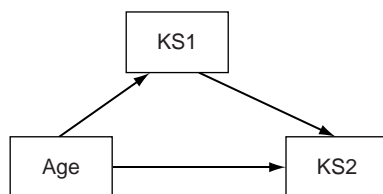


Figure 2 Directed acyclic graph for the causal relationships between KS1, KS2, and Age.

So what do the results of conditional analysis really suggest? The conditional analysis suggests that for younger children who did as well as older children at KS1, they would do better than older children at KS2. This may be plausible, as those young children might be brighter than average, so they could achieve the same score as the average older children at KS1 and then do better at KS2. Therefore, the interpretations of the relationship between pupils' age and their KS2 math score between the conditional and unconditional analysis are no longer contradictory, as the unconditional analysis refers to a general statement on the relationship between pupils' age and their performance at KS2, whereas the conditional analysis refers to a more restricted situation where only a subgroup of young children were selected to compare to average older children.

Regression to the Mean

In December 2007, a report for a charity called the Sutton Trust in the United Kingdom appeared as a headline on the BBC news. By analyzing data from children born in 2000 and 2001 (from the Millennium cohort) in the UK, the study claimed that less bright children from rich families are catching up with bright children from poor families in developmental tests between ages 3 and 5 and are expected to overtake them by the age of 7 (**Figure 3(a)**). Therefore, this study seemed to suggest that the environment in which children grow up has a dominant influence on their test performance in school even at a very early stage of cognitive development. If this conclusion is valid, it would have very important implications for policymaking. Indeed, one major political party leader in the UK referred to these results in an interview with the BBC in April 2008, when he was talking about the vision of his party. Though **Figure 3(a)** seems to suggest that children from poor families who did well in the tests at age 3 did less well at age 5 and children from rich families who did poorly in the tests at age 3 did better at age 5, this converging trend in the test results of cognitive development between these two groups of children may be, however, due to the way in which the data are presented. To help clarify why the observed convergence in the test results does not provide evidence for an association between cognitive development and inequality in family background, it is very useful to revisit how Sir Francis Galton discovered the phenomenon of regression to the mean, more than a century ago, when he was studying the inheritance of body height.

Galton and Regression to the Mean

Galton was interested in the inheritance of human intelligence, as he was an important figure in the eugenic

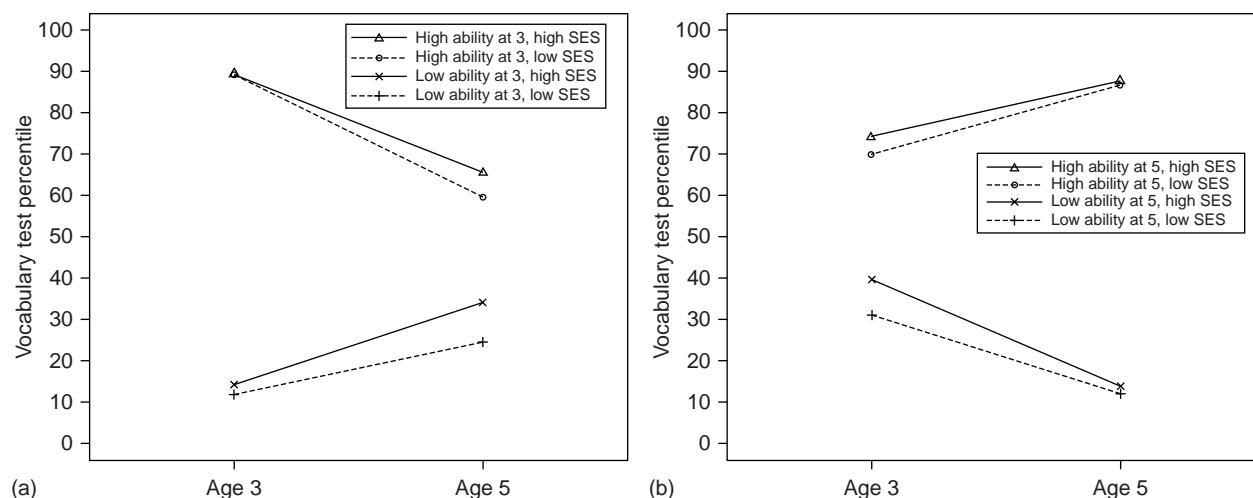


Figure 3 Trends in the children vocabulary test scores between age 3 and 5 when children are grouped by their scores at age (a) 3 or (b) 5.

movement in nineteenth-century Britain. However, as there were no reliable measures for intelligence, he turned to human body measurements. Galton invited families to his laboratory to measure their body heights. As men on average were taller than women, women's heights were multiplied by 1.08. He then plotted the average of both parents' heights (he called it mid-parent height) against their offspring's height. To his great surprise, although adult children of tall parents were still taller than most people, they were on average shorter than their parents. On the other hand, adult children of short parents, while still short, were on average taller than their parents. Galton named this phenomenon regression toward mediocrity, and we know it today as regression to the mean.

Nevertheless, discussions in the literature tended to overlook one important subtlety in Galton's body heights example. Galton's data not only showed that adult children of tall parents were shorter than their parents and adult children of short parents were on average taller than their parents, it also showed that parents of tall children were shorter than their parents and parents of short children were on average taller than their parents. Using the data from Galton's famous study, **Figure 4(a)** shows the trend of body heights across two generations by grouping families according to the parents' heights. Parents with body heights greater or equal to the average height (68 inches) were grouped as tall parents, and their mean body height was 69.51 inches. Parents with body heights less than 68 inches were grouped as short parents, and their mean body height was 66.66 inches. The mean body height of adult children from tall parents was 68.79 inches, and that of adult children from short parents was 67.12 inches. Therefore, human heights looked to be converging across the two generations, suggesting perhaps that, after many generations, there would be fewer very tall or very short people.

Yet, **Figure 4(b)** shows the trend of body heights across the two generations by grouping the families according to adult children's heights. Children with body heights greater or equal to 68 inches were grouped as tall children, and their mean body height was 69.89 inches. Children with body heights less than 68 inches were grouped as short children, and their mean body height was 65.77 inches. The mean body height of parents of tall children was 68.87 inches, and that of parents of short children was 67.58 inches. Therefore, human heights looked to be diverging, and after several generations there might be many more very tall or very short people. Two contradictory conclusions could thus be deduced from the same data, yet neither was true. The reason for two contradictory conclusions is that the correlation between parents and their children's heights was not perfect, and regression to the mean occurs when children's heights were regressed on their parents' heights or vice versa. Both analyses and misinterpretations suffer regression to the mean.

What Is Regression to the Mean?

Suppose Galton could have obtained the records of body heights of grandparents for his families. What would the trend have looked like? If families were grouped according to the children's heights, the trend of human body heights would look like **Figure 4(c)**. If families were grouped according to the grandparents' heights, the trend of human body heights would look like **Figure 4(d)**. The phenomenon of regression to the mean would become more notable between grandparents and children because the correlation between them was smaller than that either between grandparents and parents or between parents and children. In other words, the resemblance in the appearances between grandparents and children would be

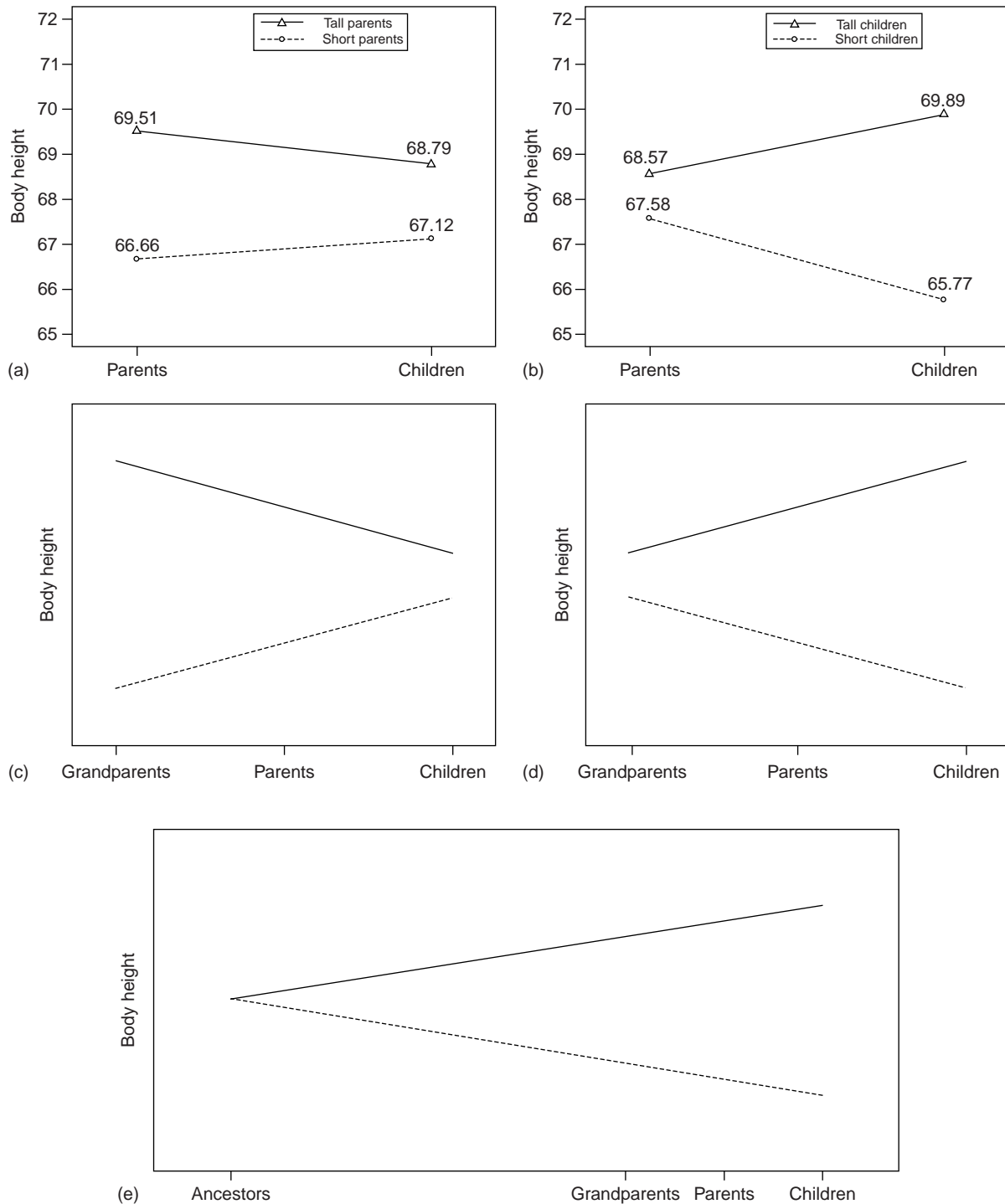


Figure 4 Trends in the body height across generations in Galton's data (a, b) and in hypothetical scenarios (c–e).

less than that either between grandparents and parents or between parents and children.

What if there was a long historical record of body heights for these families; what might the evolution of human body heights throughout several generations look like? If the families were grouped according to the heights

of children (i.e., the last generation), the trend of human body heights would look like **Figure 4(e)**. The resemblance of ancestors and descendants would diminish with time. Our appearances and characteristics have more resemblance with our parents than with our grandparents, and the resemblance between us and our ancestors

will diminish further with each increasing generation. Eventually, the resemblance between us and our distant ancestors would become unrecognizable, and, therefore, our ancestors are no more like us than an average person in their generation. This is why it is called regression to the mean, or as Galton named it regression toward mediocrity. In his seminal paper in 1886, Galton explained that the child inherits partly from his parents, partly from his ancestry. Speaking generally, the further his genealogy goes back, the more numerous and varied will his ancestry become, until they cease to differ from any equally numerous sample taken at haphazard from the race at large. Their mean stature will then be the same as that of the race; in other words, it will be mediocre.

From a statistical viewpoint, positive correlations between body heights of successive generations become smaller the farther back the genealogy goes; as a result, the phenomenon of regression to the mean becomes more remarkable. **Figure 5** is a generalization of regression to the mean between two continuous variables X_1 (ancestors' heights) and X_2 (descendants' heights), when the data are grouped according to the mean value of X_2 (as with Galton's data). When the correlation between X_1 and X_2 (r_{12}) is perfect and positive (i.e. 1), the two lines depicting the trend or trajectories in the evolution of human body heights are parallel (**Figure 5(a)**). When $0 < r_{12} < 1$, body heights appear to be diverging (**Figure 5(b)**). When $r_{12} = 0$, there is no difference in the mean of X_1 between the two groups (**Figure 5(c)**). When $-1 < r_{12} < 0$, body heights not only look to be diverging, but the two lines cross at some point between X_1 and X_2 (**Figure 5(d)**).

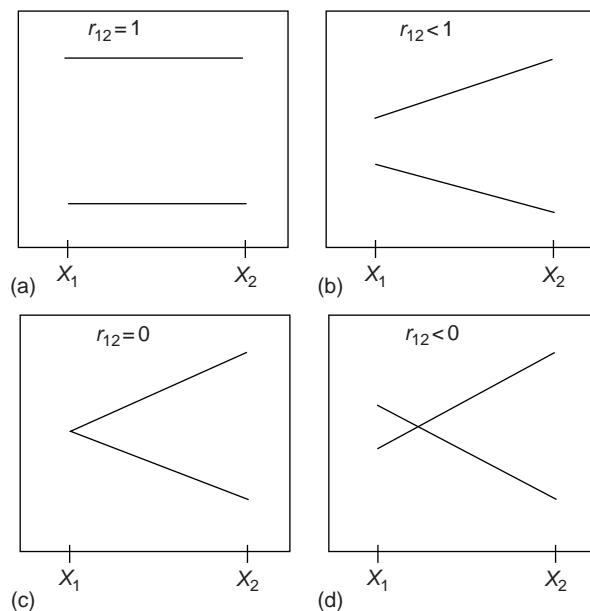


Figure 5 Patterns of trends for two repeated measurements, X_1 and X_2 . r_{12} is the correlation between X_1 and X_2 .

Revisiting the Convergence in Developmental Tests

We use the same data from Millennium Cohort analyzed by the original Sutton Trust report to produce the apparent convergence pattern in developmental tests (**Figure 3**). The variable of interest was Naming Vocabulary ability score assessed at age 3 and 5. The raw scores were then transformed into percentiles. Children were first divided into four groups by their test score in quartiles assessed at age 3. They were then further divided into rich or poor family groups according to their parents' job (professional vs. non-professional). **Figure 3(a)** shows the comparison of high-ability group (i.e., children in the top quarter of test scores) and low-ability group (children in the lowest quarter of test scores) in both rich (parents with professional jobs) and poor (parents with nonprofessional or no jobs) families. Children in the high-ability group from both rich and poor families actually did poorer in the test at age 5, although they on average still did better than those in the low-ability group whose scores were better than previously but still below the average. If this trend continues, the original report suggested that rich children in the low-ability group will overtake poor children in the high-ability group by the age of 7. Even if this convergence trend is genuine (which it is not), it is obvious from **Figure 3(a)** that all children in the high-ability group will be overtaken, irrespective of their being rich or poor.

Following our previous discussion of Galton's data, we now group children according to their test scores assessed at age 5 and their family background at age 3. The results are shown in **Figure 3(b)** in which the test performance appears to be diverging. If we follow the same line of interpretation in the original report, we may conclude that poor children in the high-ability group are doing better and better than rich children in the low-ability group. Yet, neither of the conclusions is supported by the data. Mathematically speaking, this phenomenon is due to the conditional variance of one variable being smaller than its observed variance. Because the variance of both test scores has been constrained to be equal, the variance of test scores at age 5, conditional on the scores at age 3, will become smaller than the variance of age 3. This is why we observe a convergence trend in test scores. The same applies when the variance of test scores at age 3 is conditional on test scores at age 5, and we will therefore observe a diverging trend. Grouping individuals according to their baseline or final values to investigate the trend in the change is in general misleading and should be avoided. Many people, including prominent statisticians and economists, have fallen into this trap.

The investigation of whether inequalities in socioeconomic background have a dominant influence on children's development of cognitive ability is important and will require longitudinal observations. One approach will

be looking at the general trend in the cognitive development for children from different socioeconomic backgrounds. If environment is influential, we may simply observe that rich children in general make greater progress than poor children, or if family background is more important to children with low ability, we may observe a smaller variance of test scores in children from rich families than that of children from poor families, that is, there will be a convergence within rich children.

Conclusion

With the increasing power of personal computers and wide availability of user-friendly statistical software packages, most statistical analyses only take seconds to get the results. However, the reduction in time for computing does not reduce the need for proper understanding and careful thinking. The two statistical paradoxes discussed in this article remind us that there are many traps on the road to our quest for truth, waiting for the unwary to fall into. To infer causality from nonexperimental observations is notoriously difficult, and the two examples in this article illustrate that even simple analysis and data presentation can be seriously misleading. These statistical paradoxes have been discovered a long time ago and repeatedly discussed in the literature, but, as Stigler commented, they never lose their power to surprise.

See also: Analysis and Interpretation of Multivariate Data; Analysis of Covariance; Causal Inference; Growth Modeling; Observational Studies.

Further Reading

- Arah, O. A. (2008). The role of causal reasoning in understanding Simpson's paradox, Lord's paradox, and the suppression effect: Covariate selection in the analysis of observational studies. *Emerging Themes in Epidemiology* **5**, 5.
- BBC News (2007). Bright poor children 'slip back' <http://news.bbc.co.uk/1/hi/education/7141169.stm> (accessed May 2009).
- Bickel, P. J., Hammel, E. A., and O'Connell, J. W. (1975). Sex bias in graduate admissions: Data from Berkeley. *Science* **187**, 398–404.
- Blanden, J. and Machin, S. (2007). Recent changes in intergenerational mobility in Britain. London: The Sutton Trust. <http://www.suttontrust.com/annualreports.asp> (accessed May 2009).
- Cox, D. R. and Wermuth, N. (2003). A general condition for avoiding effect reversal after marginalisation. *Journal of the Royal Statistical Society, Series B* **65**, 937–941.
- David, H. A. and Edwards, A. W. F. (2001). *Annotated Reading in the History of Statistics*. New York: Springer.
- Galton, F. (1886). Regression toward mediocrity in hereditary stature. *Journal of Anthropological Institute of Great Britain and Ireland* **15**, 246–263.
- Goldstein, H., Burgess, S., and McConnell, B. (2007). Modelling the effect of pupil mobility on school difference in educational achievement. *Journal of the Royal Statistical Society, Series A* **170**, 941–954.
- Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin* **68**, 304–305.
- Lord, F. M. (1969). Statistical adjustments when comparing preexisting groups. *Psychological Bulletin* **72**, 337–338.
- Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
- Pearson, K., Lee, A., and Bramley-Moore, L. (1899). Mathematical contributions to the theory of evolution: VI – Genetic (reproductive) selection: Inheritance of fertility in man, and of fecundity in thoroughbred racehorses. *Philosophic Transaction of Royal Society, London, Series A* **192**, 257–330.
- Senn, S. J. (2003). *Dicing with Death*. Cambridge: Cambridge University Press.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Series B* **13**, 238–241.
- Stigler, S. M. (1997). Regression towards the mean, historically considered. *Statistical Methods in Medical Research* **6**, 103–114.
- Stigler, S. M. (1999). *Statistics on the Table*. Cambridge, MA: Harvard University Press.
- Tu, Y. K., Gunnell, D., and Gilthorpe, M. S. (2008). Simpson's paradox, Lord's paradox, and suppression effects are the same phenomenon – the reversal paradox. *Emerging Themes in Epidemiology* **5**, 2.
- Tu, Y. K., West, R., Ellison, G. T. H., and Gilthorpe, M. S. (2005). Why evidence for the fetal origins of adult disease might be a statistical artifact: The "reversal paradox" for the relation between birth weight and blood pressure in later life. *American Journal of Epidemiology* **161**, 27–32.
- Wainer, H. and Brown, L. M. (2004). Two statistical paradoxes in the interpretation of group differences: Illustrated with medical school admission and licensing data. *American Statistician* **58**, 117–123.
- Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika* **2**, 121–134.

Statistical Power Analysis

L V Hedges and C Rhoads, Northwestern University, Evanston, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Intraclass correlation – In two stage samples that first sample intact groups (statistical clusters) and then sample individuals with the groups, the intraclass correlation is the ratio of between-group variance to total variance. It is used to describe the degree of clustering within samples.

Multilevel (multistage) sample – A sample that first obtains a sample of intact groups (statistical clusters) and then samples individuals within the groups. Statistics based on multilevel samples (also called multistage samples) have different properties than those based on simple random (one level) samples.

A critical aspect of planning any research design is ensuring that it is well enough designed to provide definitive evidence for the phenomenon under investigation. Power analysis is used to ensure that a study will have an appropriate chance of yielding a statistically significant result, given the treatment effect that is expected. Power analyses are required as part of proposals for funding research, and it would be almost unthinkable to embark on a large-scale study without conducting a power analysis. This article is an introduction to power analysis for research designs that are most often used in education research.

Statistical Hypothesis Testing

The dominant form of statistical inference in educational research involves hypothesis testing. A null hypothesis that is inconsistent with the research (or alternative) hypothesis of the investigation is formulated. If the observed data would have only a very small chance of occurring if the null hypothesis were true, then the null hypothesis is rejected in favor of the research hypothesis.

Formal hypothesis testing usually requires the explicit specification of how unlikely the test statistic must be to lead to rejection of the null hypothesis. This is typically a small number such as 5% (0.05) or 1% (0.01). This small number is called the significance level (often denoted by the Greek letter α).

Statistical power is the probability of making the correct decision to reject the null hypothesis when it is false. The statistical power of a research design represents the

chance that, if the researcher's substantive hypothesis is correct, the research design will lead to the correct conclusion that the research hypothesis is true.

Statistical power always depends on the research design, the statistical significance level (α), the effect size, and the sample size. The power of the design depends on how that design is organized (e.g., if covariates are used, how the sampling and/or randomization is carried out). All other things being equal, a smaller significance level reduces power. All other things being equal, a design with a larger sample size will have higher statistical power. Finally, statistical power depends on the effect size, which is a quantification of how false the null hypothesis is. All other things being equal, the larger the effect size, the higher the power. In more complex designs, there are also additional factors that affect power.

Only the type of research design, the significance level, and the sample size can, in principle, be changed by the investigator. However, there are often practical constraints on the research designs that can be used. Moreover, a strong scientific convention makes it virtually impossible to utilize significance levels larger than 0.05. The effect size is determined by the phenomenon under investigation, so it cannot be changed. This implies that sample size is usually the only factor that can be easily changed to increase statistical power.

Statistical Power Analysis

Statistical power analysis involves deciding on a specific type of design and a statistical significance level (frequently $\alpha = 0.05$) and then investigating the three inter-related factors of sample size, effect size, and power.

Power analysis might start with a particular sample size and an effect size (often the smallest effect deemed by the investigator to be of scientific interest) and then compute the statistical power. If the statistical power is unacceptably low, the investigator may choose to increase the sample size or change the research design in fundamental ways (such as introducing covariates) to obtain higher power.

Alternatively, power analysis might start with a desired power (e.g., 0.80) and an effect size, and compute the smallest sample size that yields the specified power. Or, power analysis could start with the desired power and the sample size and then compute the smallest effect

size that would yield the specified power, called the minimum detectable effect size.

Retrospective power analyses are sometimes conducted after a study has been completed to better interpret a finding of no statistically significant effect. The question is whether the study had sufficient power to detect the smallest effect size that is scientifically important. If not, failure to reject the null hypothesis does not provide strong confirmation that the effects are too small to be scientifically important.

Procedures for Power Analysis

Regardless of the type of research design involved, the procedure for carrying out statistical power analyses is the same, utilizing tables that relate power to significance level, sample size, and effect size. Different designs have different natural effect size measures that will be used in power analysis. The exact procedure depends on the specific goal of the power analysis, but generally the specific design and a significance level (usually $\alpha = 0.05$) is chosen first. We illustrate the procedure using **Table 1**, a table of power values obtained using the noncentral t -distribution that is appropriate for several designs.

Computing power for fixed sample size and effect size

Finding the power associated with a given sample size and effect size involves entering the table at the row for the appropriate sample size and then moving to the column of that row for the appropriate effect size and reading the entry, which is the power value. This procedure is the same regardless of whether it is conducted prospectively or retrospectively.

Computing sample size for fixed power and effect size

Determining the sample size that will result in a fixed power (often 0.80) for a fixed effect size can also be accomplished with a table such as **Table 1**. First, enter the table in the column corresponding to the appropriate effect size. Then, read down the column until the power value for a row is as large as desired or larger. The sample size corresponding to that row is the desired sample size. Some sources (such as Cohen, 1988) make the process easier by providing different tables giving the minimum sample sizes necessary to obtain a specified power (usually 80%) for various effect sizes.

Computing the minimum detectable effect size

Using a table such as **Table 1**, enter the table on the row corresponding to the appropriate sample size. Then, move across the row until the power is at least the desired value (e.g., 0.80). This value will be approximately the minimum detectable effect size. The value is approximate because

the desired power value may not occur exactly in that row for any tabulated effect size, so that the desired power occurs for an effect size that is in between those defining columns of the table. A more accurate value of the minimum detectable effect size can be obtained by interpolation.

Designs Comparing Independent Groups Using Simple Random Samples

Perhaps the most common research design compares the outcomes in independent groups that differ in a specific way (e.g., receive different treatments). Experiments with completely randomized designs and quasi-experiments using the nonequivalent control group design are examples of these designs. In this section, we consider only experiments that use simple random samples (e.g., subjects all in one school or sampled independently from several schools) and not experiments with more complicated sampling designs (such as those that assign whole schools to the same treatment). Some of these more complicated designs will be considered later.

Effect size

The statistical analyses of these designs may involve the analysis of variance to test a null hypothesis that several groups have identical means. However, such designs are almost always planned by examining the power of the contrast between a particular pair of groups, such as a control group with one of the treatment groups. As a consequence, the most relevant power analysis for planning the design would be the power of that contrast. The effect size for comparing two groups is Cohen's d , defined by

$$d = (\mu_1 - \mu_2) / \sigma$$

where μ_1 and μ_2 are the population means and σ is the within-treatment group standard deviation in the experiment.

How should an effect size be chosen for power analysis?

Choosing the smallest effect size that is scientifically important is a matter of judgment and experience. The effect size d is intuitive to many educational researchers and is widely used in meta-analysis. Consequently, there are many sources of data on d values actually obtained in educational research. Compendia of effect sizes in meta-analyses by Lipsey and Wilson (1993) and other compendia by Bloom *et al.* (2007) may be particularly useful. A somewhat outdated and potentially misleading guideline was offered by Cohen (1988): an effect size of $d = 0.20$ is small, $d = 0.50$ is medium, and $d = 0.80$ is large. These guidelines should be used only if there is no

37	0.07	0.14	0.25	0.40	0.56	0.72	0.84	0.92	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
38	0.07	0.14	0.25	0.41	0.58	0.73	0.85	0.93	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
39	0.07	0.14	0.26	0.41	0.59	0.74	0.86	0.94	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
40	0.07	0.14	0.26	0.42	0.60	0.75	0.87	0.94	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
41	0.07	0.15	0.27	0.43	0.61	0.77	0.88	0.95	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
42	0.07	0.15	0.27	0.44	0.62	0.78	0.89	0.95	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
43	0.07	0.15	0.28	0.45	0.63	0.79	0.89	0.96	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
44	0.07	0.15	0.29	0.46	0.64	0.79	0.90	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
45	0.08	0.16	0.29	0.47	0.65	0.80	0.91	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
46	0.08	0.16	0.30	0.48	0.66	0.81	0.91	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
47	0.08	0.16	0.30	0.48	0.67	0.82	0.92	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
48	0.08	0.16	0.31	0.49	0.68	0.83	0.92	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
49	0.08	0.17	0.31	0.50	0.69	0.84	0.93	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
50	0.08	0.17	0.32	0.51	0.70	0.84	0.93	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
52	0.08	0.17	0.33	0.52	0.71	0.86	0.94	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
54	0.08	0.18	0.34	0.54	0.73	0.87	0.95	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
56	0.08	0.18	0.35	0.55	0.75	0.88	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
58	0.08	0.19	0.36	0.57	0.76	0.89	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
60	0.08	0.19	0.37	0.58	0.78	0.90	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
64	0.09	0.20	0.39	0.61	0.80	0.92	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
68	0.09	0.21	0.41	0.64	0.82	0.93	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
72	0.09	0.22	0.43	0.66	0.85	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
76	0.09	0.23	0.45	0.69	0.86	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
80	0.10	0.24	0.47	0.71	0.88	0.96	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
84	0.10	0.25	0.49	0.73	0.90	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
88	0.10	0.26	0.51	0.75	0.91	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
92	0.10	0.27	0.53	0.77	0.92	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
96	0.11	0.28	0.54	0.79	0.93	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
100	0.11	0.29	0.56	0.80	0.94	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
120	0.12	0.34	0.64	0.87	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
140	0.13	0.39	0.71	0.92	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
160	0.14	0.43	0.76	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
180	0.16	0.47	0.81	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
200	0.17	0.51	0.85	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
250	0.20	0.61	0.92	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
300	0.23	0.69	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
350	0.26	0.75	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
400	0.29	0.81	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
450	0.32	0.85	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
500	0.35	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

other information that could be used to judge the smallest effect size that is likely to be scientifically important.

Example of computing power using Table 1

Suppose that we are interested in determining the statistical power of an experiment that contrasted a treatment group with a control group with $\alpha = 0.05$ level of significance, a sample size of $N = 30$ per group, and an effect size of $d = 0.30$. Entering **Table 1** on the row for $N_0 = 30$ and moving across to the column for an effect size of $d_0 = 0.30$, we read the power value of 0.21. How large would the sample size have to be in order to obtain a power of 0.80? To find it out, we enter the table in the column for the effect size of 0.30 and move down the rows until we reach a power value of 0.80 or larger. Doing so yields a sample size of $N = 180$ per group. Alternatively, we might ask what the minimum detectable effect size is for a power of 0.80 and a sample size of $N = 30$. Entering the table on the row for $N_0 = 30$ per group and moving across the columns of the table, we see that an effect size of $d_0 = 0.70$ yields a power of 0.76 and an effect size of $d_0 = 0.80$ yields a power of 0.86; thus, the minimum detectable effect size is between $d = 0.70$ and $d = 0.80$. Interpolating, we see that a power of 0.80 is 4/10 of the way between 0.76 and 0.86, so the minimum detectable effect size is approximately 4/10 of the way between $d = 0.70$ and $d = 0.80$, or $d = 0.74$.

Designs Utilizing Covariates to Increase Power in Comparing Independent Groups Using Simple Random Samples

The use of covariates can dramatically increase precision in designs comparing independent groups. The covariate is often a pretest of some kind, but it could be any variable that is correlated with the outcome and is measured before treatment begins. In designs that involve covariates, power depends not only on significance level, sample size, and effect size, but also on the effectiveness of the covariate in explaining variance in the outcome variable. Nonetheless, the computations using covariates can be carried out using almost the same procedures as without covariates. The only difference is that instead of using the actual effect size and the sample size to directly compute power, a slightly modified version of each, called the operational effect size and operational sample size, is used.

Operational effect size and operational sample size

Suppose that the actual effect size of interest is d , the actual sample size in each group is N , and the correlation (or multiple correlation if there is more than one covariate) between covariate(s) and the outcome is R . Then, the

operational sample size per group is $N_0 = N - q/2$, where q is the number of covariates used, and the operational effect size d_0 is

$$d_0 = \left(d / \sqrt{1 - R^2} \right) \left[\sqrt{2N / (2N - q)} \right] \quad [1]$$

If the number of covariates is small (e.g., 1), it may be sufficiently accurate to use $N_0 = N - 1$ and

$$d_0 = d / \sqrt{1 - R^2} \quad [2]$$

as the operational sample size and effect size, respectively. The actual effect size d has the same scientific interpretation in this design as in the design without covariates and should be chosen in the same way. The operational effect size is simply a convenient device that allows us to compute statistical power from standard tables.

Example of computing power using Table 1 in a design using covariates

Suppose that, as in the earlier example, we are interested in determining the statistical power of an experiment that contrasts a treatment group with a control group with the $\alpha = 0.05$ level of significance, a sample size of $N = 30$ per group, and an effect size of $d = 0.30$. Now, we introduce a covariate that is correlated ($R = 0.66$) with the outcome variable. The operational sample size is $N_0 = 30 - 1/2 = 29.5$ and the operational effect size is

$$d_0 = \left(0.30 / \sqrt{1 - 0.66^2} \right) \left[\sqrt{(2 \times 30) / [(2 \times 30) - 1]} \right] = 0.403$$

Note that if we had ignored the last term in brackets, we would have computed

$$d_0 = 0.30 / \sqrt{1 - 0.66^2} = 0.399$$

which is the same to two decimal places (namely $d_0 = 0.40$). Because $N_0 = 29.5$ is between 29 and $N_0 = 30$, we will have to interpolate between rows. Entering **Table 1** on the row for $N_0 = 29$ and moving across to the column for an effect size of $d_0 = 0.40$, we read the power value of 0.32, and on the row for $N_0 = 30$ for an effect size of $d_0 = 0.40$, we read the power value of 0.33; thus, the power value is between 0.32 and 0.33 or about 0.325. Note that introducing the covariate has increased the power by approximately 50% (from 0.21 to 0.325) compared to the design with no covariate. How large would the sample size have to be in order to obtain a power of 0.80? To find it out, enter the table in the column for the (operational) effect size of 0.40 and move down the rows until we reach a power value of 0.80 or larger. Doing so yields an (operational) sample size of $N_0 = 100$ per group, or a true sample size of $N = 101$, which is much less than without covariates. Alternatively, we might ask what the minimum effect size that yields a power of 0.80 is for a sample size of $N = 30$ per group when the covariate is included

(the minimum detectable effect size for a power of 0.80 and an operational sample size of $N_0 = 29$). Entering the table on the row for $N_0 = 29$ per group and moving across the columns of the table, we see that an effect size of $d_0 = 0.70$ yields a power of 0.75 and an effect size of $d_0 = 0.80$ yields a power of 0.85; thus, the minimum detectable operational effect size is between $d_0 = 0.70$ and $d_0 = 0.80$. Interpolating, we see that a power of 0.80 is 1/2 of the way between 0.75 and 0.85, so the minimum detectable effect size is approximately 1/2 of the way between 0.70 and 0.80 or about 0.75. Note, however, that this is the minimum detectable operational (not actual) effect size. We can translate this minimum detectable operational effect size into the minimum detectable actual effect size by inverting eqn [2] to obtain

$$d = d_0 \sqrt{1 - R^2} = 0.75 \sqrt{1 - 0.66^2} = 0.56$$

which is much smaller than without covariates.

Designs Comparing Independent Groups Using Multilevel (Clustered) Samples

Many research designs in education compare the outcomes in independent groups but obtain their sample via a multilevel (multi-stage) sampling process. For example, many educational studies obtain the groups of participants by first obtaining schools and then assigning all of the students or teachers in the same school to the same treatment. Cluster-randomized (hierarchical) experiments and many quasi-experiments patterned after them are examples of designs in this category. The term ‘cluster-randomized’ refers to the concept of first sampling intact groups (statistical clusters), then sampling individuals within groups to obtain a two-stage (two-level) cluster sample and then randomly assigning whole clusters to treatments. We discuss here only designs with two levels of sampling (e.g., schools assigned to treatments with students nested within schools).

The use of multilevel samples and cluster randomization can dramatically decrease statistical power in designs comparing independent groups as compared with designs wherein simple random sampling and individual randomization to treatment are used. However, these designs are commonly used either because they are much less costly than the alternatives or because it is politically infeasible to assign individuals within the same school to different treatments. It is also the only alternative to evaluating treatments that must be administered to the entire school because of theoretical considerations (e.g., whole school reforms).

In designs that involve two-level sampling, the power depends not only on significance level, the sample size, and the effect size but also on the amount of clustering in the sample. The amount of clustering is measured by a

quantity called the intraclass correlation, which measures the proportion of the total variance in the outcome that is between (as opposed to within) clusters (e.g., schools). Intraclass correlations range from 0 to 1, but values between 0.05 and 0.30 are most common in educational research. An additional complication is that power in these designs depends not only on the total sample size, but also on the number of clusters in each treatment group, m , and the number of individuals, n , within each cluster (assumed to be the same for each cluster). Choosing a reasonable intraclass correlation value is important for the accuracy of the power computation, so it is desirable to choose on the basis of empirical evidence from populations similar to those in the experiment. One extensive set of intraclass correlation values from representative samples of the United States population was compiled by Hedges and Hedberg (2007).

Power computations for multilevel samples such as cluster randomized experiments can be carried out using almost the same procedures as for experiments with singlelevel samples. The only difference is that, as in the case of singlelevel experiments with covariates, operational effect size and operational sample size are used.

Operational effect size and operational sample size

Suppose that the actual effect size is d , the actual number of clusters assigned to each treatment is m , the sample size in each cluster is n , and the intraclass correlation is ρ . Then, the operational sample size is $N_0 = m$. The operational effect size, d_0 , is

$$d_0 = d \sqrt{\frac{n}{1 + (n-1)\rho}} \quad [3]$$

Note that unless $\rho = 1$, the operational effect size d_0 is always larger than the actual effect size d . However, the operational sample size, $N_0 = m$, is smaller than the actual sample size, mn , so the power in the design with clustering will always be smaller than in the design without clustering.

Example of computing power in a multilevel design using Table 1

Suppose that we are interested in determining the statistical power of a multilevel experiment that contrasted two treatment groups using an $\alpha = 0.05$ level of significance, a sample size in each group of $m = 20$ schools, $n = 25$ students in each school, and an actual effect size of $d = 0.30$. Assume that the intraclass correlation is $\rho = 0.15$. The operational sample size is $N_0 = 20$ and the operational effect size is

$$d_0 = 0.30 \sqrt{\frac{25}{1 + (25-1) \times 0.15}} = 0.700$$

Entering **Table 1** on the row for the operational sample size $N_0 = 20$ and moving to the column for an operational effect size of $d_0 = 0.70$, we see that the power is 0.58. Note that the total sample size is $2 \times 20 \times 25 = 1000$, which is rather large, yet the power is not particularly large. A study using a simple random sample of this size would have had power in excess of 0.99.

How many clusters would be required to obtain a power of 0.80? To find it out, enter the table in the column for the (operational) effect size of 0.70 and move down the rows until we reach a power value of 0.80 or larger. Doing so yields a sample size on $m = 33$ clusters (each with $n = 25$ individuals) per group. We might have wondered how increasing the total sample size by increasing the number of individuals within each cluster (increasing n) and keeping the number of clusters fixed at $m = 20$ would affect the power. It can be shown that the largest operational effect size can possibly be, even if n is infinitely large, $d_0 = 0.78$, which (with $N_0 = m = 20$) corresponds to a power of about 0.67. Thus, increasing sample size by increasing clusters is a much more effective way to increase power in multilevel designs than is increasing the number of individuals within the clusters. This example illustrates that, even an infinite sample size, if it was obtained by increasing the number of individuals within the clusters and not increasing the number of clusters, may not increase the power to 0.80.

Designs Comparing Independent Groups with Multilevel Samples Using Covariates

The use of covariates can dramatically increase precision in designs comparing independent groups using multilevel samples. In designs with two-level sampling and covariates, the power depends not only on the significance level, sample size (the number m of clusters and sample size n within each cluster), effect size, and the amount of clustering in the sample (measured by the intraclass correlation ρ) but also on the effectiveness of the covariate(s) in explaining variance in the outcome variable at the cluster level and the individual (within-cluster) level. Typically, covariate effectiveness is measured by the multiple correlations R_1 and R_2 between the covariate(s) and the outcome at the individual and cluster levels, respectively.

Power computations using covariates can be carried out using a procedure similar to what was used without covariates. The only difference is in how the operational effect size and operational sample size are defined.

Operational effect size and operational sample size

Suppose that the actual effect size is d , the number of clusters assigned to each treatment is m , the sample size

within each cluster is n , the intraclass correlation is ρ , there are q_1 covariates used at the individual level and q_c covariates used at the cluster level, and the multiple correlations between the covariate(s) and outcome at the individual and cluster level are R_1 and R_2 , respectively. Then, the operational sample size is $N_0 = m - q_c/2$. The operational effect size, d_0 , is

$$d_0 = d \sqrt{\frac{n}{1 + (n-1)\rho - R_1^2 - (nR_2^2 - R_1^2)\rho}} \left[\sqrt{\frac{2m}{2m - q_c}} \right] \quad [4]$$

The operational effect size d_0 is always larger than the operational effect size that would have been computed if there were no covariates.

If the number of covariates is small (e.g., 1), it may be sufficiently accurate to use $N_0 = m - 1$ and

$$d_0 = d \sqrt{\frac{n}{1 + (n-1)\rho - R_1^2 - (nR_2^2 - R_1^2)\rho}} \quad [5]$$

as the operational sample size and effect size, respectively.

The covariate–outcome correlations at both the individual and (especially) the cluster level have important effects on the accuracy of the power computation, so they should be chosen on the basis of empirical evidence gathered from populations similar to those in the experiment. One extensive compilation of covariate–outcome correlations at the individual and school level from representative samples of the US population is Hedges and Hedberg (2007).

Example of computing power in a multilevel design with covariates using Table 1

Suppose that we are interested in determining the statistical power of a multilevel experiment that will contrast treatment groups using a $\alpha = 0.05$ level of significance, a sample size in each group of $m = 20$ schools and $n = 25$ students in each school, and an effect size of $d = 0.30$. Assume that the intraclass correlation is $\rho = 0.15$ and that the individual- and cluster-level-squared multiple correlations between the $q_1 = q_c = 1$ covariate and outcome are $R_1^2 = 0.60$ and $R_2^2 = 0.50$ at the individual and cluster levels, respectively. The operational sample size is approximately $N_0 = 20 - 1 = 19$ and the operational effect size is approximately

$$d_0 = 0.30 \sqrt{\frac{25}{1 + (25-1) \times 0.15 - 0.60 - (25 \times 0.50 - 0.60) \times 0.15}} = 1.008$$

Entering **Table 1** on the row for the operational sample size $N_0 = 19$ and moving to the column for an operational effect size of $d_0 = 1.00$, we see that the power is 0.85, so the power of this design is approximately 0.85. Note that if we had used the more complicated expression for the operational effect size given in eqn [4], we would have obtained $d_0 = 1.02$ and essentially the same power.

Conclusion

The procedures described in this article can be used to carry out statistical power analysis with the goal of finding statistical power for a fixed sample size and effect size, finding the sample size that will provide acceptable power for a fixed effect size, and finding the minimum detectable effect size for a fixed sample size and power. One complication that arises is that within treatment group sample sizes or (in designs with clustering) within cluster sample sizes may not be equal. In this case, using the harmonic mean of the sample sizes yields reasonably accurate power computations. Thus, in the case of designs involving simple random samples, we would use $N = 2N_1N_2/(N_1 + N_2)$. In designs wherein the number of clusters in the treatment and control groups are unequal, we would use $m = 2m_1m_2/(m_1 + m_2)$. When the cluster sizes are unequal, we would use $n = m(n_1n_2 \dots n_m)/(n_1 + n_2 + \dots + n_m)$ as the cluster size in formulas involving n .

See also: Analysis of Variance; Design of Experiments; Hypothesis Testing and Confidence Intervals; Statistical Significance Versus Effect Size.

Bibliography

- Bloom, H. S., Richburg-Hayes, L., and Black, A. R. (2007). Using covariates to improve precision: Empirical guidelines for studies that randomize schools to measure the impacts of educational interventions. *Educational Evaluation and Policy Analysis* **29**, 30–59.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. New York: Academic Press.
- Hedges, L. V. and Hedberg, E. C. (2007). Intraclass correlations for planning group-randomized experiments in education. *Educational Evaluation and Policy Analysis* **29**, 60–87.
- Lipsey, M. W. and Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist* **48**(2), 1181–1209.

Further Reading

- Bloom, H. S. (1995). Minimum detectable effects: A simple way to report statistical power of experimental designs. *Evaluation Review* **19**, 547–556.
- Bloom, H. S., Hill, C. J., Black, A. R., and Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement

- effect-size benchmarks for educational interventions. *Journal for Research on Educational Effectiveness* **1**, 289–328.
- Borenstein, M., Rothstein, H., and Cohen, J. (2001). *Power and Precision*. Teaneck, NJ: Biostat.
- Elashoff, J. (2007). *nQuery Advisor*. Saugus, MA: Statistical Solutions.
- Hedges, L. V. and Pigott, T. D. (2001). The power of statistical tests in meta-analysis. *Psychological Methods* **6**, 203–217.
- Hedges, L. V. and Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods* **9**, 426–445.
- Kaplan, D. (1995). Statistical power in structural equation modeling. In Hoyle, R. H. (ed.) *Structural Equation Modeling: Concepts, Issues, and Applications*, pp 100–117. Newbury Park, CA: Sage.
- Konstantopoulos, S. (2008). The power of tests for treatment effects in three-level cluster randomized designs. *Journal of Research on Educational Effectiveness* **1**, 66–88.
- Konstantopoulos, S. (2008). The power of the test for treatment effects in three-level randomized block designs. *Journal of Research on Educational Effectiveness* **1**, 265–288.
- Kraemer, H. C. and Thiernann, S. (1987). *How Many Subjects? Statistical Power Analysis in Research*. Newbury Park, CA: Sage.
- Lipsey, M. W. (1990). *Design Sensitivity: Statistical Power Analysis for Experimental Research*. Newbury Park, CA: Sage.
- Murphy, K. R. and Myers, B. (2004). *Statistical Power Analysis*. Mahwah, NJ: Erlbaum.
- O'Brien, R. G. and Muller, K. E. (1993). Unified power analysis for *t*-tests through multivariate hypotheses. In Edwards, L. K. (ed.) *Applied Analysis of Variance in Behavioral Science*, pp 297–344. New York: Marcel Dekker.
- Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized experiments. *Psychological Methods* **2**, 173–185.
- Raudenbush, S. W. and Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods* **5**(3), 199–213.
- Raudenbush, S. W. and Liu, X. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods* **6**(4), 387–401.
- Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika* **54**, 131–151.
- Satorra, A. and Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika* **50**, 83–90.
- Schochet, P. Z. (2008). Statistical power for random assignment evaluations of educational programs. *Journal of Educational and Behavioral Statistics* **33**, 62–87.
- Spybrook, J., Raudenbush, S. W., Liu, X., Congdon, R., and Martinez, A. (2008). *Optimal Design*. Ann Arbor, MI: University of Michigan.

Relevant Websites

- <http://www.power-analysis.com> – Power Analysis Sample Size Calculation Statistical Software.
- <http://www.statsol.ie> – Software for Sample Size, Multiple Imputation, Hot Decking and Equivalence Testing.

Statistical Significance Versus Effect Size

X Fan and T R Konold, University of Virginia, Charlottesville, VA, USA

© 2010 Elsevier Ltd. All rights reserved.

Statistical Significance

Statistical Significance Testing

There has been a long history of statistical significance testing, and specifically, of null hypothesis significance testing (NHST). The fundamental concept underlying statistical significance test is sampling variation, or sampling error. Assuming a researcher has two treatment conditions, A (a new math instructional approach) and B (the business-as-usual approach currently in use). The researcher is interested in knowing if A is more effective than B (research hypothesis, or RH: A is better than B). To help decide if the RH can be supported, the researcher set up another hypothesis (null hypothesis of no difference, or NH) that A and B are equally effective (i.e., students under A and B will have equal learning outcomes).

Because of sampling variation, even if A and B are equal in terms of instructional effectiveness, it is possible that the method A sample may have higher average score than the method B sample. Therefore, the simple observation that students under method A performed better than those under method B does not constitute sufficient evidence that method A is more effective than method B, because we have not ruled out sampling error as one possible explanation for the observed difference.

If treatments A and B are indeed the same (NH of no difference is true), a small performance difference between A and B samples is more likely to occur by chance than a large performance difference. The question is, how much higher should the method A sample average score be than the method B sample average score before we can say, with reasonable confidence, that the observed difference is not due to sampling variability (i.e., chance)? Once we decide statistically that the sampling variability is unlikely to be a viable explanation for the observed difference, the NH is rejected. The rejection of NH constitutes statistical evidence supporting the RH, because our statistical significance test helps us rule out sampling error, or chance, as a viable explanation for the observed difference between the two samples.

In statistical significance testing, we assess the probability of observing sample statistic values equal to or more extreme than that obtained from a sample data (D) if the NH (H_0) is true, that is, $p(D|H_0)$. If $p(D|H_0)$ is sufficiently small (e.g., $p < 0.05$), the NH can be considered as not viable, and will be rejected. So, statistical significance (i.e., the rejection of the NH) tells us that the random sampling variability is unlikely a viable explanation for

the observed difference between observed sample value and the hypothesized population value, but it does not really inform us about the magnitude or practical importance of our observed sample results. Going back to the example of Methods A and B in teaching math, the rejection of the NH (NH: A and B are equally effective in teaching math) means that it is highly unlikely that sampling error could have been the cause for the observed difference. As a result, we conclude A and B are most likely not equally effective (i.e., rejection of NH). This conclusion, however, does not give us a clear indication about how much more effective method A is than method B in the practical sense. More detailed discussion about hypothesis testing can be found elsewhere in the encyclopedia.

Criticisms of Statistical Significance Test

In the past several decades, NHST has received a considerable amount of criticism. The criticism has centered on the use (or misuse, overuse) of statistical significance in research practice, rather than on the logic of statistical significance test. For example, Kirk (1996) discussed

1. significance testing does not tell researchers what they want to know, but rather, it creates the illusion of probabilistic proof by contradiction;
2. statistical significance testing is often a trivial exercise, as it simply indicates the power of the study design (i.e., sample size); and
3. significance testing “turns a continuum of uncertainty into a dichotomous reject-do-not-reject decision,” and this may lead to a situation in which different researchers obtain identical treatment effects but draw different conclusions, simply due to slight differences in design (e.g., sample size differences).

Kirk (2003) later stated, “it is evident that the current practice of focusing exclusively on a dichotomous reject–non-reject decision strategy of NH testing can actually impede scientific progress . . . The focus of research should be on our scientific hypotheses, what data tell us about the magnitude of effects, the practical significance of effects, and the steady accumulation of knowledge” (p. 100). Vacha-Haase and Thompson (2004) pointed out, statistical significance only evaluates the probability or likelihood of the sample results under the assumption of true NH, but it cannot evaluate the importance of the sample results. Meehl (1978) discussed that the NH, taken literally, is always false. Because of this, statistical significance often

becomes a matter of having sufficiently large sample in order to have enough statistical power to reject the false NH.

At this time, there is a general consensus that, in educational research practice, the importance attributed to statistical significance has exceeded what it deserves, almost to the point that statistical significance becomes the literal equivalent of importance of research findings. Undoubtedly, this situation has been compounded by the unfortunate misnomer of significance in this context.

Effect Size

The criticism of statistical significance has led quantitative researchers to consider other information, such as effect size, to supplement or complement statistical significance test in research practice. As a result, the use of effect size measure has gained popularity in recent years. Effect size is usually a scale-free index that quantifies the difference between the observed sample results and the NH. For example, when the NH states that methods A and B are equally effective in math instruction (i.e., no performance difference between the two approaches), if the observed sample means are 80 and 80, respectively, the effect size is zero. If the observed sample means are not equal (e.g., 80 and 90, respectively), the effect size will be nonzero. The motivation for effect size is to have a scale-free index for the difference, because the actual difference (e.g., a ten-point difference in the example above) is often not interpretable due to arbitrary measurement scaling that is typical in social and behavioral sciences. For example, a 15-point difference on one measure (e.g., Wechsler intelligence quotient (IQ) score scale with $\mu = 100$ and $\sigma = 15$) is drastically different from a 15-point difference on another measure (e.g., the Graduate Record Examination (GRE) score scale with $\mu = 500$ and $\sigma = 100$).

Effect Size Measures

Different effect size measures have been developed for different analytical techniques. In general, most commonly used measures of effect belong to one of two broad categories: (1) measures of effect size, which are based on standardized group mean differences; and (2) measures of association strength, which are based on shared proportion of variance (or variance accounted for).

Standardized Mean Difference as Effect Size

Univariate d

The first category of effect size measure is based on standardized group mean difference. This category is best represented by Cohen's d , or another variation of it (e.g., Glass's Δ). In a generic form, Cohen's d is estimated from a sample as:

$$d = \frac{\bar{X}_{Group1} - \bar{X}_{Group2}}{sd_{pooled}}$$

where \bar{X}_{Group1} and \bar{X}_{Group2} can be sample means from two groups in an experimental design (e.g., $\bar{X}_{Treatment}$ and $\bar{X}_{Control}$), or from any two naturally occurring groups (e.g., male and female group comparison). The sd_{pooled} represents the pooled standard deviation across the two samples.

In an experimental design study, if the researcher believes that the treatment may affect not only the performance level of the treatment, but also the variation of the treatment group, one may consider Glass's Δ of using the control group's standard deviation ($sd_{control}$) to replace sd_{pooled} to calculate d . In a situation where only one group is involved (e.g., a sample is being compared against a hypothesized population value), d can be obtained as:

$$d = \frac{\bar{X}_{sample} - \mu}{sd_{sample}}$$

Multivariate D_M

For multivariate group comparison (e.g., comparison of two groups on multiple outcome variables, as in a Hotelling T^2 test or multivariate analysis of variance (MANOVA)), Mahalanobis distance (D_M) is the multivariate counterpart of d :

$$D_M = \sqrt{(\bar{X}_{G1} - \bar{X}_{G2})' S_{pooled}^{-1} (\bar{X}_{G1} - \bar{X}_{G2})}$$

where \bar{X}_{G1} and \bar{X}_{G2} are the mean vectors of the two groups in the comparison, $(\bar{X}_{G1} - \bar{X}_{G2})'$ is the transposed mean vector difference, and S_{pooled}^{-1} is the inverse of the pooled covariance matrix.

Variance-Accounted-for Measures as Effect Size

The second broad category, variance-accounted-for measures, is based on the relationship strength between variables, and is often represented as an index representing the proportion of variance in the outcome variable that is associated with predictor(s) or independent variable(s). In this category, there are variance-accounted-for measures, and their bias-corrected counterparts.

Variance-accounted-for measures

The most widely known variance-accounted-for measures are η^2 or R^2 . These two are fundamentally the same, but they may be used in slightly different contexts. The general form is:

$$\eta^2 (R^2) = \frac{SS_{source}}{SS_{total}}$$

The numerator is the sum of squares (SS) from a source of research interest, while the denominator represents the

total variance of the outcome variable. In a model that contains only one explanatory factor (predictor), that factor (predictor) is the source of interest. For a model containing multiple factors (predictors), the source of interest may be one variable, a subset of variables, or the full model with all the explanatory variables (predictors). R^2 is usually used in the context of regression analysis, while η^2 is often preferred in the context of analysis of variance (ANOVA). In addition, if the source represents the full model (i.e., all factors in an ANOVA model), R^2 is often used. It should be noted that these contextual nuances in the use of η^2 and R^2 are not really statistically meaningful, as all these analytical models (e.g., t -test, regression, ANOVA) can be considered as part of the general linear model (GLM). In this sense, η^2 and R^2 are fundamentally the same.

In regression analysis, researchers are often interested in evaluating the unique contribution of one or more additional predictors over and above some other predictors already in the model. In this context, R_A^2 researchers often use

$$R_A^2 = R_{k+j}^2 - R_k^2$$

where R_k^2 is the R^2 from the model with k predictors (reduced model), and R_{k+j}^2 is the R^2 for the model with $k+j$ predictors (the full model containing the k predictors, plus some additional j predictors).

In the case of a GLM, when a model contains multiple independent variables, most likely, these variables are non-orthogonal (i.e., correlated; orthogonality may be achieved in an experimental study with balanced design). As a result, the SS attributable to each of the variables is not additive. The standard practice is to use the type III SS (SS_{III} : unique SS), which has been adjusted for the effects of all other variables in the model. As a result, SS_{III} for the effects do not sum up to the total SS . In a model involving multiple variables, their η^2 's do not add up to be more than 100%.

η^2 versus partial η^2

In educational research practice, there appears to be some confusion about η^2 and *partial* η^2 , and we have seen instances where partial η^2 was used and interpreted as if it were η^2 as defined above. For clarification, we consider the definition of partial η^2 :

$$\text{Partial } \eta^2 = \frac{SS_{\text{source}}}{SS_{\text{source}} + SS_{\text{error}}}$$

Obviously, partial η^2 can be quite different from η^2 , because the denominator in the above formula is not the total sum of squares. η^2 and partial η^2 are the same only when the model contains one independent variable (factor). Whenever the model contains more than one independent variable (e.g., categorical factors and/or continuous

predictors), η^2 and partial η^2 will have different (often-times, very different) values. In addition, the sum of partial η^2 's from multiple variables may exceed 100%. The following hypothetical example illustrates the differences between η^2 and partial η^2 :

It is obvious that η^2 and partial η^2 are defined differently. It is also obvious that partial η^2 cannot be interpreted as proportion of variance associated with a factor, and they (partial η^2 's) can add up to be more than 100% in many situations. Instead, partial η^2 represents proportion reduction in residual variance by a factor. A large reduction in residual variance may only represent a very small proportion of the total variance. For this reason, for a model with multiple independent variables, to interpret partial η^2 as if it were η^2 will overestimate (sometimes greatly) the effect of each variable. Partial η^2 is readily available from SPSS package (SPSS GLM procedure), and some SPSS users might have unintentionally misinterpreted partial η^2 as if it were η^2 .

Effect size in HLM

Intraclass correlation coefficient (ICC) is sometimes considered as an effect size measure for random effects (coefficients) model, which subsumes hierarchical linear modeling (HLM) analysis. However, in HLM, proportion reduction in (residual) variance at a given level is probably the most common effect size measure. In a generic form, proportion reduction in variance is (also see Raudenbush and Bryk, 2002: 74) given by

$$\text{Proportion Reduction in Variance} = \frac{\sigma_{\text{residual}}^2(\text{null}) - \sigma_{\text{residual}}^2(\text{full})}{\sigma_{\text{residual}}^2(\text{null})}$$

where $\sigma_{\text{residual}}^2$ is the residual variance at any given level (e.g., level-2 residual variance), and (null) represents a model with no (or fewer) predictors at this level, and (full) represents a model with more predictors at the same level. Basically, this is a measure of proportion of variance explained. But it is not proportion of variance explained out of the total variance, but only the proportion of variance explained out of the variance at a given level. As such, caution is warranted in interpreting this measure, as a large proportion reduction in variance at a given level may only represent a very small proportion of the total variance.

Corrected variance-accounted-for measures

Samples η^2 and R^2 have positive bias, and their bias-corrected counterparts have been proposed. For R^2 , adjusted R^2 in the following form is readily available from any statistical analysis software package (e.g., SAS, SPSS):

$$R_{\text{adjusted}}^2 = 1 - \frac{N - 1}{(N - k - 1)(1 - R^2)}$$

where N is the sample size, and k is the number of predictors in a model. The R_{adjusted}^2 is always smaller than R^2 ,

and the algebraic adjustment can make $R^2_{adjusted}$ negative, even though R^2 is conceptually a positive entity. A negative $R^2_{adjusted}$ suggests that one cannot trust R^2 at all! The adjustment penalizes small sample size and more predictors. With the increase of sample size relative to the number of predictors in the model, the difference between R^2 and $R^2_{adjusted}$ becomes smaller, and eventually negligible.

For fixed-effect ANOVA, omega squared (ω^2) is a widely known bias-corrected counterpart of η^2 :

$$\omega^2 = \frac{SS_{source} - (df_{source})(MS_{error})}{MS_{error} + SS_{total}}$$

ω^2 is always smaller than η^2 , and using the example in **Table 1**, ω^2 for factor A would be:

$$\omega^2 = \frac{20 - (1)(\frac{30}{26})}{(\frac{30}{26}) + 100} = 0.19$$

In addition to the widely known η^2 and R^2 and their bias-corrected counterparts, there are other measures of association strength. In the simplest case, Pearson correlation coefficient itself (or preferably, its squared value) is a variance-accounted-for measure of effect. In many multivariate analytic techniques (e.g., MANOVA), the test statistic Wilk's Λ is related to η^2 ($\eta^2 = 1 - \Lambda$). For canonical correlation analysis, the squared canonical correlation coefficient (R_c^2) is a variance-accounted-for measure of association strength, as it represents the shared amount

of variance between two canonical variates. **Table 2** presents some effect size measures associated with some widely used statistical analysis. More detailed description of many effect size measures and their computational details can be found in other sources (e.g., Cohen, 1988; Grissom and Kim, 2005; Kline, 2004).

Relationship Between Different Measures of Effect Size

Some seemingly different types of effect size measures (e.g., d vs. R^2) may actually be the same statistically. For example, the two major categories of effect size measures (standardized mean difference effect size, e.g., d , and variance-accounted-for effect size, e.g., R^2) are related. As is widely known, many seemingly different analytic approaches (e.g., a two-sample t -test vs. a correlation analysis) are special cases of the GLM. As such, these techniques are statistically related. For example, a two-sample t -test is equivalent to testing that the Pearson correlation coefficient between the outcome variable and dummy-coded group membership is zero. In this case, d and r are related:

$$r = \frac{d}{\sqrt{d^2 + 4}}$$

or

$$d = \frac{2r}{\sqrt{1 - r^2}}$$

Table 1 Illustration of η^2 and partial η^2

Source	df	SS (type III)	η^2	Partial η^2
Model	3	70 ^a		
Factor A	1	20	20/100 = 0.20	20/(20+30) = 0.40
Factor B	1	20	20/100 = 0.20	20/(20+30) = 0.40
A*B	1	10	10/100 = 0.10	10/(10+30) = 0.25
Error	26	30		
Total	29	100		

^aIt is assumed that, to non-orthogonality of the factors (A,B, A*B), their SSs do not sum up to be Model SS of 70.

Table 2 Some common effect size measures

Analysis	Effect size measures
Correlation analysis (continuous, rank-ordered, dichotomous variables)	r (r^2): Pearson r , Spearman r , phi (ϕ) coefficient
Contingency table	Cramer's V , phi (ϕ) coefficient ^a
Independent t -test	d (η^2 , ω^2)
ANOVA	η^2 , ω^2
Regression analysis	R^2 , R^2_A , $R^2_{adjusted}$
Multivariate group comparison (Hotelling T^2 , MANOVA)	D_M , η^2 (1 - Wilk's Λ)
Canonical correlation	R_c^2
Discriminant function analysis	η^2 (1 - Wilk's Λ)
Hierarchical linear modeling (HLM)	Proportion reduction in variance, ICC

^aFor 2×2 contingency table analysis: Cramer's $V = |\text{coefficient}|$.

Therefore, the two types of measures of effect size (standardized mean difference and variance-accounted-for) are not really different statistically.

Use of Effect Size

As statistical significance shows in probabilistic terms how likely to obtain sample statistic values equal to or more extreme than the observed sample statistic if the NH is true, but does not inform about the magnitude of effect or whether the findings are practically meaningful, the use of effect size as an indicator of practical meaningfulness has been widely advocated in recent years (Kirk, 2001; Kline, 2004). The APA Task Force on Statistical Inference recommended to always provide some effect-size estimate when reporting a p value (Wilkinson and The APA Task Force on Statistical Inference, 1999). The most recent *Publication Manual of the American Psychological Association* (5th edition) states:

For the readers to fully understand the importance of your findings, it is almost always necessary to include some index of effect size or strength of relationship in your Results section. . . . to provide the reader not only with information about statistical significance, but also with enough information to assess the magnitude of the observed effect or relationship. (American Psychological Association, 2001: 25, 26)

There is a general consensus that the role of effect size should be enhanced, and the role of statistical significance should be reduced. However, there is disagreement about how much the role of statistical significance should be reduced. On the one hand, statistical significance test has been criticized as representing obstacles to scientific inquiry (e.g., Falk and Greenbaum, 1995; Hunter, 1997; Meehl, 1978). On the other hand, researchers have argued for the legitimate role for the correct use of significance testing (e.g., Cortina and Dunlap, 1997; Hagen, 1997).

Interpreting Effect Size

Based on the observed typical results in social science literature, Cohen (1969, 1988) proposed some tentative benchmarks for what may be considered as small, medium, and large effects:

- small effect: $d \approx |0.2|$; $\eta^2 \approx 1\%$;
- medium effect: $d \approx |0.5|$; $\eta^2 \approx 10\%$; and
- large effect: $d \geq |0.8|$; $\eta^2 \geq 25\%$

This tentative interpretive framework for effect size has been useful for research practitioners, and this is probably a good impetus for the increasing popularity of effect size in research practice (Kirk, 1996). For a better understanding of these tentative benchmarks ($d = 0.20, 0.50$, and



0.80), readers may consult Henson (2006) for easy-to-understand visual graphs of these effect sizes. These interpretive benchmarks, however, should not be considered as a one-size-fits-all rule of thumb, as it is very likely that what can be considered as a small or a large effect will vary across substantive disciplines; for the same effect size (e.g., small effect), the practical importance may be quite different across disciplines. In research practice, the tentative benchmarks for effect size should be considered in light of the typical results in the discipline (Thompson, 2001).

As there are so many different types of effect size measures (Cohen, 1988; Grissom and Kim, 2005; Kline, 2004), and different effect size measures can be very different numerically (e.g., d and η^2), it is critical that the specific effect size measure is reported. It would not be helpful to simply say that an effect size of 0.30 is obtained, because, without knowing what the effect size measure is, this information is not interpretable.

Effect Size and Its Confidence Interval

Sample effect size itself is a random variable. This means that sample effect size is subject to sampling variability, and the degree of its sampling variation is inversely related to sample size. In other words, sample effect size from small samples may deviate farther from the population effect size than that from larger samples. For small sample size conditions, a medium, or even a large, effect size could occur due to sampling error (i.e., by chance) when the true population effect size is zero (Fan, 2001).

The fact that sample effect size is a random variable has been widely known (e.g., Hedges and Olkin, 1985), although some education researchers may have not paid sufficient attention to this. It is not uncommon to hear the discussion to the effect that the outcome of a statistical significance test is influenced by sample size (true!), so attention should be paid to effect size, as if effect size were not influenced by sample size. Undoubtedly, the use of effect size measure makes good quantitative and common sense. However, the use of effect size serves a different purpose from that of statistical significance: while statistical significance test evaluates the probability of obtaining the sample outcome by chance, effect size provides some indication for the magnitude and practical meaningfulness. On the one hand, statistical significance test may detect a trivial effect (sufficient statistical power due to large sample size), or it may fail to detect a meaningful or obvious effect (lack of statistical power due to small sample size). On the other hand, under small sample size conditions, a seemingly meaningful (e.g., medium or even large) effect can occur by chance, thus not trustworthy (Fan, 2001). Because of these considerations, statistical significance and effect size complement each other, but they do not substitute for one another.

		Effect size		
		Small	Medium	Large
Statistical significance				
No		<ol style="list-style-type: none"> 1. It appears that there is neither statistical nor practical effect. 2. Unless future research indicates otherwise, the hypothesis of no or little effect is favored in both statistical and practical sense. 	<ol style="list-style-type: none"> 1. The effect looks promising, but caution is warranted in interpreting the effect size by itself, as medium effect size could have occurred because of sampling error. 2. Type II error (i.e., you fail to statistically detect a true effect) may be a concern, and you may consider the power of your design (e.g., Is your sample too small to statistically detect a meaningful effect?). 	<ol style="list-style-type: none"> 1. You have some evidence for a meaningful effect. However, some caution is still warranted, because a large effect could have been the result of sampling error when the sample size is very small. 2. You may take a critical look at the possible lack of power of your study design.
		<ol style="list-style-type: none"> 1. The statistical significance is not accompanied by practical significance, and it could have been the result of statistical power (i.e., large <i>N</i>). 2. Caution is warranted in interpreting the statistical significant findings, and they may not be interpreted as something practically important. 	<ol style="list-style-type: none"> 1. It is unlikely that the observed effect is due to sampling error. 2. The magnitude of effect is likely to be practically meaningful in many fields of social and behavioral sciences. 3. Conclude that effect is meaningful both statistically and practically. 	<ol style="list-style-type: none"> 1. There is high degree of certainty that the observed effect is not due to chance statistically, and the magnitude of the effect is also practically meaningful. 2. Conclude with confidence that effect is meaningful in both in statistical and practical sense.

The vertical and horizontal arrows indicate the increase of certainty for statistical and practical meaningfulness of the observed effect.

Figure 1 Guidelines for combining statistical significance with effect size. Reproduced with permission from Fan, X. (2001). Statistical significance and effect size: Two sides of a coin. *Journal of Educational Research* 94(5), 275–282. copyright H.W. Wilson company.

Thompson (2002, 2007) proposed the use of confidence intervals for sample effect size. This approach combines the sample size information with effect size measure, and provides variability estimate for a sample effect size. This approach represents a balance between the descriptive use of sample effect size and statistical significance. More importantly, “Confidence intervals for effect sizes are especially valuable because they facilitate *meta-analytic thinking* and the interpretation of intervals via comparison with the effect intervals from related prior studies” (Thompson, 2002: 25, emphasis original). Ultimately, for translating research findings to education practice, a researcher should not only ask if a sample result is likely, but also if an effect is practically noteworthy and replicable.

Conclusion

Statistical significance and effect size are two components that complement each other, but one does not substitute for the other. Statistical significance and effect size serve complementary purposes, and the sole reliance on either can be misleading. Good research practice requires that both should be taken into consideration for making sound quantitative research decisions. **Figure 1** provides some practical guidelines for combining the information of statistical significance and effect size to make sound decisions in research practice.

See also: Hypothesis Testing and Confidence Intervals.

Bibliography

- American Psychological Association (2001). *Publication Manual of the American Psychological Association*, 5th edn. Washington, DC: American Psychological Association.
- Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. New York: Erlbaum.
- Cortina, J. M. and Dunlap, W. P. (1997). On the logic and purpose of significance testing. *Psychological Methods* 2, 161–172.
- Falk, R. and Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory and Psychology* 5, 75–98.
- Fan, X. (2001). Statistical significance and effect size: Two sides of a coin. *Journal of Educational Research* 94, 275–282.
- Grissom, R. J. and Kim, J. J. (2005). *Effect sizes for Research: A Broad Practical Approach*. Mahwah, NJ: Erlbaum.
- Hagen, R. L. (1997). In praise of the null hypothesis statistical test. *American Psychologist* 52, 15–24.
- Hedges, L. V. and Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando, FL: Academic Press.
- Henson, R. K. (2006). Effect-size measures and meta-analytic thinking in counseling psychology research. *Counseling Psychologist* 34, 601–629.
- Hunter, R. L. (1997). Needed: A ban on the significance test. *Psychological Science* 8, 3–7.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement* 56, 746–759.
- Kirk, R. E. (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement* 61, 213–218.
- Kirk, R. E. (2003). The importance of effect magnitude. In Davis, S. F. (ed.) *Handbook of Research Methods in Experimental Psychology*, pp 83–105. Oxford, UK: Blackwell.
- Kline, R. B. (2004). *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research*. Washington, DC: American Psychological Association.

- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology* **46**, 806–834.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edn. Thousand Oaks, CA: Sage.
- Thompson, B. (2001). Significance, effect sizes, stepwise methods, and other issues: Strong arguments move the field. *Journal of Experimental Education* **70**, 80–93.
- Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect size. *Educational Researcher* **31**(3), 25–32.
- Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools* **44**, 423–432.
- Vacha-Haase, T. and Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology* **51**, 473–481.
- Wilkinson, L. and The APA Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist* **54**, 594–604.

Further Reading

- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher* **5**, 3–8.

Stochastic Processes

F Baudoin, Purdue University, West Lafayette, IN, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction and Historical Perspectives

A stochastic process is a mathematical tool that describes a random phenomenon evolving in time.

Let us for instance consider an economist who decides to focus every day on the price of a barrel of oil over the next year. The price at $t = 1, \dots, 365$, say X_t , can not be predicted with absolute certainty, though some prices are more likely than others. This price can then be mathematically modeled by the concept of a random variable: X_t is random. The set, $(X_t)_{t=1, \dots, 365}$ of these random variables is an example of a stochastic process.

The first stochastic process that has been extensively studied is the so-called Brownian motion, named in honor of the botanist Robert Brown (1773–1858), who observed and described in 1828 (see Brown, 1827) the random movement of particles suspended in a liquid or gas. One of the first mathematical studies of this process goes back to the mathematician Louis Bachelier (1870–1946), who presented in 1900, a stochastic modeling of the stock and option markets. However, mainly because of the lack of a rigorous foundation of probability theory at that time, the seminal work of Bachelier was ignored by mathematicians for a long time. However, in his 1905 paper (see Einstein, 1905), Albert Einstein (1879–1955) brought this stochastic process to the attention of physicists by presenting it as a way to indirectly confirm the existence of atoms and molecules. The rigorous mathematical study of stochastic processes really began with the mathematician Andrei Kolmogorov (1903–1987). His monograph (Kolmogorov, 1956), published in Russian in 1933, built up probability theory rigorously from the fundamental axioms in a way comparable to Euclid's treatment of geometry. From this axiomatic approach, Kolmogorov gives a precise definition of stochastic processes. His point of view stresses the fact that a stochastic process is nothing else but a random variable valued in a space of functions (or a space of curves). For instance, if an economist reads a financial newspaper because he/she is interested in the price of a barrel of oil every day over the last year, then he/she will focus on the curve of these prices. According to Kolmogorov's point of view, saying that these prices form a stochastic process is then equivalent to saying that the curve that is seen is the realization of a random variable defined on a suitable probability space. This point of view is mathematically quite profound and provides existence results for stochastic processes (Daniell-Kolmogorov existence result, 1933) as well as pathwise regularity results (Kolmogorov, 1956).

Joseph Doob (1910–2004) writes in the introduction to his famous book *Stochastic Processes* (see Doob, 1953).

[a stochastic process is] any process running along in time and controlled by probability laws ... [more precisely] any family of random variables X_t [where] a random variable ... is simply a measurable function ...

Doob's point of view, which is consistent with Kolmogorov's and built on the work by Paul Lévy (1886–1971), is nowadays commonly given as a definition of a stochastic process. Relying on this point of view that emphasizes the role of time, Doob's work, developed during the 1940s and the 1950s, has quickly become one of the most powerful tools available to study stochastic processes. During the same period in the 1940s, using Kolmogorov's and Doob's results, Kiyosi Itô (1915–) developed the basic tools of the theory of stochastic calculus (see Itô, 1944). This theory makes it possible to define the integrals of stochastic processes as against other stochastic processes and gives sense to the notion of differential equation driven by a random signal (stochastic differential equations).

These different seminal works contributed to give to the theory of stochastic processes a rigorous basis that led to many further developments within and outside mathematics. For further reading on the mathematical theory of stochastic processes, we refer to the book by Revuz and Yor (1999) or the book by Protter (2004). In the next section, we present some examples of stochastic processes and their applications.

Examples and Applications of Stochastic Processes

Markov Chains

Markov chains are among the most important stochastic processes. They are stochastic processes for which the description of the present state fully captures all the information that could influence the future evolution of the process.

Predicting traffic flows, communications networks, genetic issues, and queues are examples where Markov chains can be used to model performance. Devising a physical model for these chaotic systems would be impossibly complicated but doing so using Markov chains is quite simple. Modeling the to-and-fro communications traffic over the Internet is a prime example of what Markov chains can do. Nowadays, the Internet has become exceedingly complex, and traffic moves in a random

way between network nodes. Markov chains are commonly used to describe not only the traffic, which is inherently unpredictable, but also how the network will perform, despite the complexity of its structure. The PageRank of a webpage as used by Google is defined by a Markov chain.

It is also worth mentioning that Markov chains have been recently used in education (see Duys and Headrick, 2004). This study examined the efficacy of an infrequently used statistical analysis in counselor education research. A Markov chain analysis was used to examine hypothesized differences between students' use of counseling skills in an introductory course.

Beyond Markov Chains, Fractional Brownian Motion

In engineering applications of probability, stochastic processes are often used to model the input of a system. Recent studies of communication networks show that real inputs exhibit long-range dependence: the behavior of a real process after a given time t depends not only on the situation at t but also on the whole history of the process up to time t . Moreover, it turns out that this property is far from being negligible because of the effects it induces on the expected behavior of the global system. Therefore, Markov chains cannot be used. Another property that the processes have encountered in communication networks is the self-similarity: their behavior is in distribution of the same, up to a space-scaling, whatever the time-scale is: this is to say that the process $\{X_{at}, t \in [0,1]\}$ has the same distribution as the process $\{a^H X_b, t \in [0,1]\}$, where H is called the Hurst parameter. Several estimations on real data tend to show that H often lies between 0.7 and 0.8, whereas, for instance, the usual Brownian motion has a Hurst parameter equal to 0.5. Processes that satisfy these two properties, long-range dependence and self-similarity, are nowadays being studied extensively. Among them, the so-called fractional Brownian motion introduced by Kolmogorov to model turbulences phenomena is one of the simplest.

Brownian Motion and Financial Mathematics

As pointed out in the section titled 'Introduction,' the Brownian motion is one of the first stochastic processes that has been studied. Since the seminal observation by L. Bachelier

It seems that the market, the aggregate of speculators, can believe in neither a market rise nor a market fall, since, for each quoted price, there are as many buyers as sellers.

and the work by Black and Scholes (see 1973) and Merton (1973), Brownian motion is at the basis of the mathematical modeling of financial markets. It can be used to model the fluctuations of stock prices. This modeling has in particular led to a rational option: the pricing theory. An option is a contract written by a seller that extends to the buyer the right but not the obligation to buy (in the case of a call option) or to sell (in the case of a put option) a particular asset at some time. In return for granting the option, the seller collects a payment (the premium) from the buyer. This premium can be computed by using models involving the Brownian motion (the so-called Black–Scholes formula). For further reading on this topic, we refer the interested reader to the book of Karatzas and Shreve (1998).

Two-Dimensional Brownian Motion and Related Processes

The study of the two-dimensional Brownian motion and of related stochastic processes (SLE) has made it possible to solve many conjectures of physicists concerning critical phenomena in dimension 2. The Fields medal was awarded in 2006 to W. Werner for his work in this domain. For further reading on this topic, we refer readers to the book of Lawler (2005).

Bibliography

- Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy* **81**(3), 637–659.
- Brown, R. (1827). A brief account of microscopical observations made in the months of June, July and August, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosophical Magazine* **4**, 161–173.
- Doob, J. L. (1953). *Stochastic Processes*. New York/London: Wiley/Chapman Hall.
- Duys, D. and Headrick, T. (2004). Using Markov chain analyses in counselor education research. *Counselor Education and Supervision* **44**(2), 108–120.
- Einstein, A. (1905). Über die von der molekularkinetischen. Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik* **17**, 549–560.
- Itô, K. (1944). Stochastic integral. *Proceedings of the Imperial Academy, Tokyo* **20**, 519–524.
- Karatzas, I. and Shreve, S. (1998). *Methods of Mathematical Financial*. New York: Springer.
- Kolmogorov, A. N. (1956). *Foundations of the Theory of Probability*, N. Morrison (trans.), 2nd English edn. New York: Chelsea Publishing.
- Lawler, G. (2005). *Conformally Invariant Processes in the Plane, Mathematical Survey Monographs*. Providence, RI: AMS.
- Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science* **4**, 141–183.
- Protter, P. (2004). *Stochastic Integration and Differential Equations, Vol. 21 of Applications of Mathematics*, 2nd edn. Springer.
- Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd edn. Berlin: Springer.

Structural Equation Models

S-Y Lee and X-Y Song, The Chinese University of Hong Kong, Hong Kong, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In educational science, substantive theory usually involves manifest (observed) and latent variables. Manifest variables are those that can be measured directly, like test scores and grade point average. Very often, it is required to deal with latent variables that cannot be directly measured by a single manifest variable. Examples are intelligence, personality, quantitative ability, and so on. In practice, the characteristics of a latent variable can be partially assessed by a linear combination of some manifest variables. For instance, quantitative ability of secondary school students can be reflected by their test scores in mathematics, physics, and chemistry. In most substantive research in education, it is important to establish an appropriate model to evaluate a series of simultaneous hypotheses about the impacts of latent variables and manifest variables on the other variables, and take the measurement errors into account. Structural equation models (SEMs) are well recognized as the most important statistical method to serve the above purpose. They have been widely applied to many educational problems for example, for measuring the growth of intelligence and its relationship with personality and school environment. At present, there are more than a dozen of commercial software for fitting linear SEMs to data, such as LISREL VIII (Jöreskog and Sörbom, 1996), and EQS6 (Bentler and Wu, 2002), among others. Moreover, due to the strong demand in various fields, the growth of SEM has been very rapid. Many subtle models and statistical methods have been established for better analyses of more complicated data. The objective of this article is to provide a brief description of the standard linear SEMs and some of its generalizations that are useful in educational research, including the nonlinear SEMs, SEMs with ordered categorical variables, and multilevel SEMs. In addition to the maximum likelihood approach, we also introduce the Bayesian approach and the freely available software WinBUGS for statistical inferences, including estimation and model comparison.

Linear Structural Equation Models

In a confirmatory factor analysis (CFA) model, correlations among latent factors can be assessed by their covariance matrix; however, latent variables are never regressed on the other variables. The basic goal of linear SEMs is to

generalize the CFA model for assessing how the latent variables affect each other in various ways. A typical representation of linear SEMs is the LISREL model (Jöreskog and Sörbom, 1996), which consists of two parts, the measurement equation and the structural equation. The structural equation, which specifies relationships among the latent variables, is defined by the following equation:

$$\eta = \Pi\eta + \Gamma\xi + \delta \quad [1]$$

where $\eta(q_1 \times 1)$ is a vector of endogenous (dependent) latent variables and $\xi(q_2 \times 1)$ is a vector of exogenous (independent) latent variables, $\Pi(q_1 \times q_1)$ and $\Gamma(q_1 \times q_2)$ are unknown matrices of regression coefficients that represent the causal effects among η and ξ , and $\delta(q_1 \times 1)$ is a random vector of error measurements or residuals. It is assumed that ξ is distributed as $N[0, \Phi]$, δ is distributed as $N[0, \Phi_\delta]$, and ξ is uncorrelated with δ . The latent vectors η and ξ are related with the manifest random vectors by the measurement model which composes the following CFA models:

$$x_1 = \Lambda_1\eta + \varepsilon_1 \quad [2]$$

$$x_2 = \Lambda_2\xi + \varepsilon_2 \quad [3]$$

where $x_1(r \times 1)$ and $x_2(s \times 1)$ are random vectors of the manifest variables which are the respective indicators for η and ξ , $\Lambda_1(r \times q_1)$ and $\Lambda_2(s \times q_2)$ are loading matrices, and $\varepsilon_1(r \times 1)$ and $\varepsilon_2(s \times 1)$ are random vectors of error measurements. It is assumed that ε_1 and ε_2 are normally distributed with zero means and diagonal covariance matrices. Moreover, ε_1 and ε_2 are uncorrelated, and they are also uncorrelated with η , ξ , and δ . Let $y = (x_1^T, x_2^T)^T$, $\omega = (\eta^T, \xi^T)^T$, and $\varepsilon = (\varepsilon_1^T, \varepsilon_2^T)^T$, eqns [2] and [3] can be expressed as

$$\begin{aligned} y &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix} \begin{pmatrix} \eta \\ \xi \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \\ &= \Lambda\omega + \varepsilon \end{aligned}$$

Hence, linear SEMs are formulated as a CFA model in which the latent factors satisfy a linear structural equation.

Like many statistical models, linear SEMs (and their generalizations discussed in subsequent sections) rely on substantive knowledge in building a model. It is a confirmatory tool rather than an exploratory tool. In applications, we usually have a clear objective of the study, and some basic background about the key structure of the model that is obtained either from the subject knowledge or from preliminary data analysis. This basic background

will be used in model identification, specification, and interpretation. In substantive applications, the basic background of the manifest variables in y suggests a structure of Λ in relating the manifest variables with the latent factors (variables). Often, a nonoverlapping structure of Λ , in which each indicator only measures one of the latent variables, is used to provide clear interpretation of the latent factors. The structure of Λ (e.g., the nonoverlapping structure with fixed loadings at appropriate entries) may be sufficient for identifying the model. Otherwise, more constraints have to be imposed. There are no sufficient and necessary conditions for identification, and the issue has to be considered on a problem-by-problem basis.

Nonlinear Structural Equation Models

As interactive effects of the exogenous latent variables may have substantial impact on the endogenous latent variables, nonlinear SEMs have been developed for establishing more meaningful and correct models (see Kenny and Judd, 1984; Schumacker and Marcoulides, 1998). For the model perspective, nonlinear SEMs are straightforward extensions of linear SEMs, in which the structural equation accommodates nonlinear terms of the exogenous latent variables for predicting the endogenous latent variables. More specifically, the model is defined as follows (see Lee *et al.*, 2003):

$$y_i = A c_i + \Lambda \omega_i + \varepsilon_i \quad [5]$$

$$\eta_i = \Pi \eta_i + B x_i + \Gamma F(\xi_i) + \delta_i \quad [6]$$

where c_i and x_i are vectors of fixed covariates, A and B are unknown matrices of regression coefficients, $\omega^T = (\eta^T, \xi^T)^T$, $F(\xi) = (f_1(\xi), \dots, f_a(\xi))$ is a vector-valued function with linearly independent, nonzero, known, and differentiable functions, $a \geq q_2$, and the definitions of other terms are the same as before.

In this model, more care is needed to interpret the mean vector of y . Let A_k^T and Λ_k^T be the k th row of A and Λ , and let y_k be the k th element of y . For $k = 1, \dots, p$, it follows from [5] that $E(y_{ik}) = A_k^T c_i + \Lambda_k^T E(\omega_i)$. Although $E(\xi_i) = 0$, it follows from [6] that $E(\eta_i) \neq 0$ if $F(\xi)$ is a nonlinear function of ξ . Hence, $E(\omega) \neq 0$ and $E(y_{ik}) \neq A_k^T c_i$. Let $\Lambda_k^T = (\Lambda_{k\eta}^T, \Lambda_{k\xi}^T)$ be a partition of Λ_k^T that corresponds to the partition of $\omega = (\eta^T, \xi^T)^T$. Because $E(\xi) = 0$ and $\eta = (I - \Pi)^{-1} [Bx_i + \Gamma F(\xi)]$, it follows from [6] that

$$\begin{aligned} E(y_{ik}) &= A_k^T c_i + \Lambda_{k\eta}^T E(\eta_i) + \Lambda_{k\xi}^T E(\xi_i) = \\ &= A_k^T c_i + \Lambda_{k\eta}^T (I - \Pi)^{-1} [Bx_i + \Gamma E(F(\xi))] \end{aligned}$$

As $F(\xi)$ is usually not very complicated in most practical applications, $E(F(\xi))$ is not very complex and the computation of $E(y_k)$ is not difficult.

For a linear SEM, the saturated model is defined by $y = \mu + \varepsilon^*$, where ε^* is distributed as $N[0, \Sigma]$, in which Σ is a generic covariance matrix without any structure. The above model cannot be regarded as a saturated model for assessing the goodness of fit of a hypothesized nonlinear SEM, because the distribution of the corresponding ε^* is not normal when the structural equation is nonlinear. Hence, we will consider this issue as a model comparison problem, by comparing a linear model with a nonlinear model having nonlinear terms of exogenous latent variables.

Structural Equation Models with Ordered Categorical Data

Ordered categorical data are frequently encountered in educational research. As the histograms of most ordered categorical variables are either skewed or bimodal, routinely treating them as normal will lead to erroneous conclusions (see Olsson, 1979). A common method for treating ordered categorical variables is thought a threshold approach, in which the categorical data are regarded as manifestation of an underlying normal variable.

Suppose we are interested in a linear SEM or a nonlinear SEM. Now, for each y , we have $y = \{y_o, y_u\}$, where y_o is a subset of variables whose exact continuous measurements are observable, while y_u is the remaining subset of variables such that the corresponding continuous measurements are unobservable. The information associated with y_u is given by an observable ordered categorical vector z as follows. For the k th variable in z ,

$$z_k = m \text{ if } \alpha_{k,m} < y_{uk} < \alpha_{k,m+1} \quad [7]$$

where m is an integral value in $\{0, 1, \dots, b_k\}$, and $\alpha_{k,0} < \alpha_{k,1} < \dots < \alpha_{k,b_k} < \alpha_{k,b_k} + 1$ are the thresholds. In general, we set $\alpha_{k,0} = -\infty$ and $\alpha_{k,b_k+1} = \infty$. For the k th variable in z , there are $b_k + 1$ categories which are defined by the thresholds $\alpha_{k,j}$. Note that an additional indeterminacy is induced by the ordered categorical variables. To tackle this problem, we propose to fix the thresholds at both ends, $\alpha_{k,1}$ and α_{k,b_k} , at preassigned values. This approach implicitly picks measures for the location and the dispersion of y_k . For instance, the range $\alpha_{k,b_k} - \alpha_{k,1}$ provides a standard for measuring the dispersion. For scale invariant models, the choice of the preassigned values for the fixed thresholds only changes the scale of the estimated covariance matrix (see Lee *et al.*, 1990). One common method is to fix $\alpha_{k,1} = \Phi^{*-1}(f_{k,1}^*)$ and $\alpha_{k,b_k} = \Phi^{*-1}(f_{k,b_k}^*)$, where $\Phi^*(\cdot)$ is the distribution function $N[0, 1]$, and $f_{k,1}^*$ and f_{k,b_k}^* are the frequencies of the first category and the cumulative frequencies of categories with $z_k < b_k$, respectively. These restrictions imply that the mean and the variance of the underlying

continuous variable y_k are 0 and 1, respectively. They have been frequently used in Bayesian analyses of SEMs with ordered categorical variables (see Lee, 2007).

Two-Level Structural Equation Models

In educational research, it is common to encounter hierarchical data that are collected from units nested within a large number of clusters. Examples could include random sample of students from within random samples of schools. As individuals within a group share certain common influential factors and hence lead to correlated observations, the assumption of independence among observed observations is violated. Clearly, ignoring the correlated structure of data and analyzing them as independent observations will give erroneous results. Moreover, it is also desirable to establish a meaningful model for the between-groups levels and study the effects of the between-groups latent variables on the within-groups latent variables. Consequently, it is necessary to develop two-level SEMs that take into consideration the correlated structure of data.

Consider a collection of p -variate random vectors y_{gi} , $i = 1, \dots, N_g$, within groups $g = 1, \dots, G$. The sample sizes N_g may differ from group to group so that the data set is unbalanced. At the first level, we assume that, conditional on the group mean v_g , random observations in each group satisfy the following measurement model (see Song and Lee, 2004):

$$y_{gi} = v_g + v_{gi} = v_g + \Lambda_{1g}\omega_{1gi} + \varepsilon_{1gi}, \quad [8]$$

$$g = 1, \dots, G, i = 1, \dots, N_g$$

where Λ_{1g} is a $p \times q_1$ matrix of factor loadings, ω_{1gi} is a $q_1 \times 1$ random vector of latent factors, and ε_{1gi} is a $p \times 1$ random vector of error measurements that is independent of ω_{1gi} and is distributed as $N[0, \Psi_{1g}]$, where Ψ_{1g} is a diagonal matrix. Note that due to the existence of v_g , y_{gi} and y_{gj} are not assumed to be independent. To account for the structure at the between-groups level, we assume that the group mean v_g satisfies the following factor analysis model:

$$v_g = \mu + \Lambda_2\omega_{2g} + \varepsilon_{2g}, \quad g = 1, \dots, G, \quad [9]$$

where μ is the mean vector, Λ_2 is a $p \times q_2$ matrix of factor loadings, ω_{2g} is a $q_2 \times 1$ vector of latent variables, and ε_{2g} is a $p \times 1$ random vector of error measurements that is independent of ω_{2g} and is distributed as $N[0, \Psi_2]$, where Ψ_2 is a diagonal matrix. Moreover, the first- and second-level measurement errors are assumed to be independent. It follows from equations [8] and [9] that

$$y_{gi} = \mu + \Lambda_2\omega_{2g} + \varepsilon_{2g} + \Lambda_{1g}\omega_{1gi} + \varepsilon_{1gi} \quad [10]$$

For assessing the interrelationships among the latent variables, latent vectors ω_{1gi} and ω_{2g} are partitioned as

$\omega_{1gi} = (\eta_{1gi}^T, \xi_{1gi}^T)^T$ and $\omega_{2g} = (\eta_{2g}^T, \xi_{2g}^T)^T$, respectively; where η_{1gi} ($q_{11} \times 1$), ξ_{1gi} ($q_{12} \times 1$), η_{2g} ($q_{21} \times 1$), and ξ_{2g} ($q_{22} \times 1$) are latent vectors, with $q_{j1} + q_{j2} = q_j$, for $j = 1, 2$. The distributions of ξ_{1gi} and ξ_{2g} are $N[0, \Phi_{1g}]$ and $N[0, \Phi_2]$, respectively. The structural equations in the between-groups and within-groups models can be linear or nonlinear. For brevity we consider the following linear structural equations:

$$\eta_{1gi} = \Pi_{1g}\eta_{1gi} + \Gamma_{1g}\xi_{1gi} + \delta_{1gi} \quad [11]$$

$$\eta_{2g} = \Pi_2\eta_{2g} + \Gamma_2\xi_{2g} + \delta_{2g} \quad [12]$$

where Π_{1g} ($q_{11} \times q_{11}$), Π_2 ($q_{21} \times q_{21}$), Γ_{1g} ($q_{11} \times q_{12}$), and Γ_2 ($q_{21} \times q_{22}$) are unknown parameter matrices, δ_{1gi} is a vector of error measurements and is distributed as $N[0, \Psi_{1g\delta}]$, δ_{2g} is a vector of error measurements and is distributed as $N[0, \Psi_{2\delta}]$, and $\Psi_{1g\delta}$ and $\Psi_{2\delta}$ are diagonal matrices. In the within-groups structural equation, we assume as usual that ξ_{1gi} and δ_{1gi} are independent. Similarly, in the between-groups structural equation, we assume that ξ_{2g} and δ_{2g} are independent. Moreover, we assume that the within-groups latent vectors η_{1gi} and ξ_{1gi} are independent of the between-groups latent vectors η_{2g} and ξ_{2g} . Hence, it follows from [11] that η_{1gi} is independent of η_{2g} and ξ_{2g} . The proposed two-level SEM can be identified by the common method of fixing appropriate elements in Λ_{1g} , Π_{1g} , Γ_{1g} , Λ_2 , Π_2 , and Γ_2 at preassigned known values. Similar as before, the proposed model can accommodate ordered categorical variable through the threshold approach as described in the following section (see Lee, 2007).

The Maximum Likelihood Approach For Analysis of SEMs

Let D_o be the observed data set that may contain different kinds of data and let D_u be the latent data set that contains the unobserved continuous measurements that are underlying the corresponding observed data. For example in an SEM with ordered categorical data, $D_o = \{Y_o, Z\}$ and $D_u = \{Y_u\}$, where $Y_o = \{y_{io} : i = 1, \dots, n\}$, $Y_u = \{y_{iu} : i = 1, \dots, n\}$, and $Z = \{z_i : i = 1, \dots, n\}$. Let M be a general SEM that is proposed to fit D_o . This model could be a general SEM which may be a linear model, a nonlinear model, a two-level model, or any combination of the above models. Moreover, let Ω be the collection of all latent vectors in the model, and let θ be the vector of unknown parameters.

For linear SEMs, the traditional maximum likelihood (ML) approach is the most common method in estimating the unknown parameter vector θ on the basis of the covariance structure analysis framework. Several user-friendly packages, such as AMOS, EQS6, and LISREL, can be used to conduct the ML analysis. For more complicated SEMs, for example those presented in the earlier

sections, the Monte Carlo Expectation-Maximization (MCEM) algorithm (Wei and Tanner, 1990) will be utilized to obtain the ML estimates. This algorithm involves the E-step and the M-step. Let $L_c(D_o, D_u, \theta)$ be the logarithm of complete-data likelihood function. At the r th iteration with a current value $\theta^{(r)}$, the E-step is completed by approximating the expectation of $L_c(D_o, D_u, \theta) | D_o, \theta^{(r)}$ with a sufficiently large number of observations simulated from the conditional distribution of D_u given D_o , and $\theta^{(r)}$ through some Markov chain Monte Carlo (MCMC) method, such as the Gibbs sampler (Geman and Geman, 1984). The M-step, which maximizes $Q(\theta | \theta^{(r)})$ to update $\theta^{(r)}$ to $\theta^{(r+1)}$, will be completed through a sequence of conditional maximization steps (see Meng and Rubin, 1993). Some useful applications of MCEM algorithm in the ML analysis of complicated SEMs can be found in Lee and Song (2005, 2007) and Song *et al.* (2008).

A Bayesian Approach for the Analysis of SEMs

In recent years, the Bayesian approach received much attention in SEM (see, for example, Lee (2007) and references therein). This popular statistical approach has the following good features: (1) It directly incorporates genuine prior knowledge in the analysis. In educational studies, good prior information on the parameters in the whole model may be available from various sources, for example, from previous studies or subject experts. (2) The Bayesian methods do not rely on asymptotic theory and give reliable statistical inference even with small sample sizes (see Lee and Song, 2004). (3) The Bayesian methods produce latent variable estimates with good statistical properties, and reliable model comparison statistics, etc.

In a Bayesian approach, θ is considered to be random with a distribution (called prior distribution), and an associated (prior) density function, say, $p(\theta | M) = p(\theta)$ (see Berger, 1985). Bayesian inference is conducted on the basis of the following posterior density function, $p(\theta | D_o, M)$, which is proportional to $p(D_o | \theta, M)p(\theta)$. In large samples, the Bayesian and maximum likelihood approaches are asymptotically equivalent. However, for moderate and small sample, accurate prior information can be incorporated in $p(\theta)$ for achieving better results (see Lee, 2007). Let β_k be a row vector in the regression matrices (e.g. A , Λ , Π , or Γ), ψ_k be a variance in the diagonal matrices of the residuals, and Φ be a covariance matrix of the exogenous latent vector. The following conjugate prior distributions are most common in Bayesian analysis of SEM:

$$\beta_k \stackrel{D}{=} N[\beta_{k0}, H_{k0}], \psi_k \stackrel{D}{=} \text{inverted gamma}(\alpha_{k0}, \gamma_{k0}),$$

$$\text{and } \Phi \stackrel{D}{=} IW(R_0, \rho_0)$$

where β_{k0} , α_{k0} , γ_{k0} , and ρ_0 are given hyperparameter values, H_{k0} and R_0 are given positive-definite matrices, and IW denotes the inverted Wishart distribution (see Lee (2007) for the choices of hyperparameters that are associated with informative and non-informative prior inputs).

The Bayesian estimate of θ , $\hat{\theta}$ is usually defined as the mean of the posterior distribution with density function of cf $p(\theta | D_o, M)$. As this distribution is complicated, $\hat{\theta}$ is obtained as the sample mean of a sufficiently large number of observations of θ that are simulated from the joint posterior distribution $p(\theta, D_u | D_o, M)$ by means of some MCMC methods. The freely available software WinBUGS produces Bayesian estimates of the parameters and the latent variables for various kinds of SEMs, including those mentioned above (see Lee, 2007).

Suppose we need to compare K competing models, M_1, \dots, M_K . A Bayesian model comparison statistic that takes into account the number of unknown parameters in the model is the deviance information criterion (DIC) (see Spiegelhalter *et al.*, 2002). Under a competing model M_k with a vector of unknown parameter θ_k of dimension d_k , we define

$$DIC_k = -\frac{2}{J} \sum_{j=1}^J \log p(Y | \theta_k^{(j)}, M_k) + 2d_k$$

where $\{\theta_k^{(j)}, j = 1, \dots, J\}$ is a sample of observations simulated from the posterior distribution. In model comparison, the model with the smaller DIC value is selected. In analyzing a hypothesized model, WinBUGS (Spiegelhalter *et al.*, 2003) also produces a DIC value for model comparison. As pointed out in the WinBUGS User Manual (Spiegelhalter *et al.*, 2003) in practical application of DIC, it is important to note the following: (1) If the difference in DIC is small, for example, less than 5, and the models make very different inferences, then just reporting the model with the lowest DIC could be misleading. (2) DIC can be applied to nonnested models. (3) Moreover, DIC gives clear conclusion to support the null hypothesis or the alternative hypothesis. Detailed discussion of DIC can be found in Spiegelhalter *et al.* (2002).

An Illustrative Example

We illustrate the Bayesian approach with WinBUGS by an analysis of quality of life for stroke survivors. For stroke survivors, handicap, activities of daily living, and depression have significant impact on health-related quality of life (HRQOL). This example aims to study how these variables relate to HRQOL in the sub-acute phase of stroke recovery. The data set contains 500 Chinese stroke survivors who could complete a QOL questionnaire. Outcome variables include: World Health Organization quality of life measure (total) scores on four domains: physical,

psychological, social, and environment; London handicap scale (LHS) scores, which measures the level of handicap in six dimensions: mobility, independence, occupation, social integration, orientation, and economic self-sufficiency; modified barthel index (MBI) score, which measures the activities of daily living; and geriatric depression scale (GDS) score, which measures the depression of patients. The first four continuous variables form the latent variable of HRQOL, η . The next six variables, which are ordered categorical with a six-point scale, form the latent variable of LHS, ξ . The last two continuous variables are the covariates of MBI, x_1 and GDS, x_2 , respectively. Let y_i denote the $10 \times$ random vector of observable and unobservable continuous variables. A nonlinear SEM with linear covariates (see eqns [5] and [6]) was proposed, in which $A = 0$, $\omega_i = (\eta_i, \xi_i)^T$, $\Pi = 0$, $B = (b_1, b_2)$, $x_i = (x_{i1}, x_{i2})^T$, $\Gamma = (\gamma_1, \gamma_2)$, $F(\xi_i) = (\xi_i, \xi_i^2)^T$, and

$$\Lambda^T = \begin{pmatrix} 1 & \lambda_{21} & \lambda_{31} & \lambda_{41} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \lambda_{62} & \lambda_{72} & \lambda_{82} & \lambda_{92} & \lambda_{10,2} \end{pmatrix} \quad [13]$$

Hence, the nonlinear structural equation is written as

$$\eta_i = b_1 x_{i1} + b_2 x_{i2} + \gamma_1 \xi_i + \gamma_2 \xi_i^2 + \delta_i \quad [14]$$

On the basis of the above model settings, Bayesian estimates of the unknown parameters and their standard error estimates are obtained. The path diagram, together with the estimated results, is depicted in **Figure 1**. The details of less important thresholds are not reported. It is observed that activities of daily living, handicap, and depression are all important predictors of HRQOL. Activities of daily living (MBI) positively impact on HRQOL, which means the more capable the patients are in their daily living, the better the HRQOL. Depression and handicap, however, negatively impact on QOL, which means bad mood and handicap would reduce the HRQOL of the patients. The important finding of this analysis is that the depressive mood in the subacute phase of stroke recovery has the most significant effect on the HRQOL of stroke survivors. Therefore, more attention should be paid to the detection and management of post stroke depression.

To illustrate the use of DIC for model comparison, we let M_1 be the nonlinear SEM defined by [13] and [14]. The two competing models M_2 and M_3 are considered as follows. M_2 is a linear SEM defined by the same measurement equation [13] and a linear structural equation: $v_i = b_1 x_{i1} + b_2 x_{i2} + \gamma_1 \xi_i$. M_3 is a nonlinear SEM

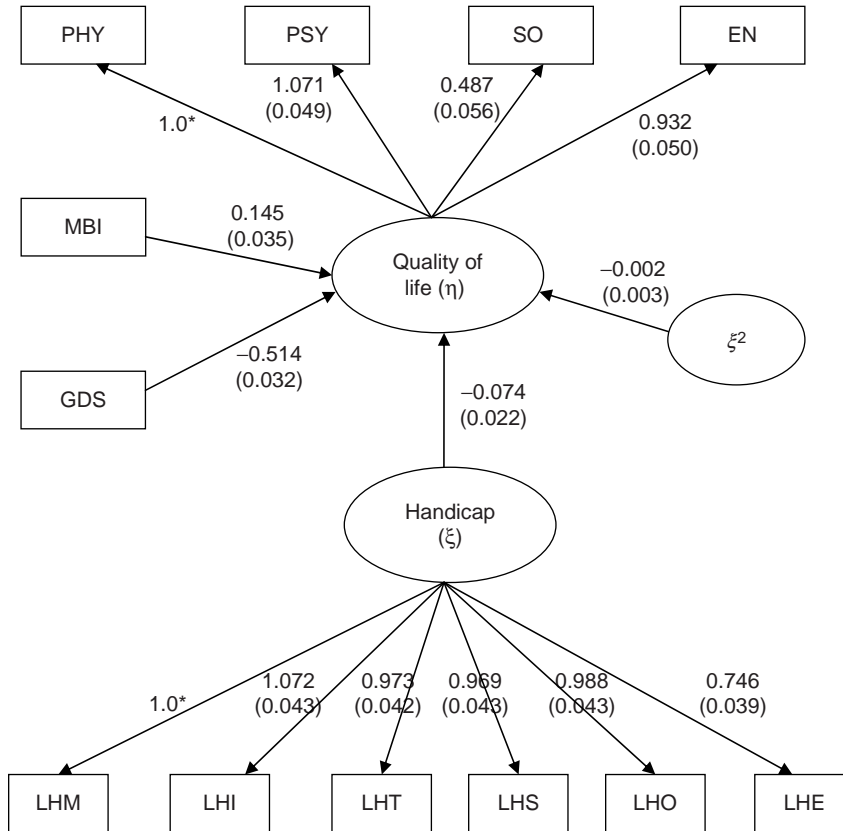


Figure 1 The path diagram and estimated results of the proposed model in analyzing the QOL data, where PHY, PSY, SO, and EN, respectively, denote the HRQOL scores in the four domains; LHM, LHI, LHT, LHS, LHO, and LHE, respectively, denote LHS measured in the six dimensions.

defined by the sample measurement equation [13], and a nonlinear structural equation with the absence of covariates: $v_i = \gamma_1 \xi_i + \gamma_2 \xi_i^2$. Using WinBUGS, we found that the DIC values corresponding to M_1 , M_2 , and M_3 are, respectively, equal to $DIC_1 = 12524.6$, $DIC_2 = 12523.1$, and $DIC_3 = 12575.7$. As DIC_3 is significantly larger than DIC_1 and DIC_2 , the nonlinear SEM without covariates was rejected. Since the estimate values of DIC_1 and DIC_2 are very close, the parsimonious model M_2 should be chosen. Actually, when looking at the estimate of γ_2 and its standard error estimate, we found that the effect of quadratic term is very close to zero and is not statistical significant.

Bibliography

- Bentler, P. M. and Wu, E. J. C. (2002). *EQS6: For Windows User Guide*. Encino, CA: Multivariate Software.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.
- Jöreskog, K. G. and Sörbom, D. (1996). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. London: Scientific Software International.
- Kenny, D. A. and Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin* **96**, 201–210.
- Lee, S. Y. (2007). *Structural Equation Modeling: A Bayesian Approach*. New York: Wiley.
- Lee, S. Y. and Song, X. Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research* **39**, 653–686.
- Lee, S. Y. and Song, X. Y. (2005). Maximum likelihood analysis of a two-level nonlinear structural equation model with fixed covariates. *Journal of Educational and Behavioral Statistics* **30**, 1–26.
- Lee, S. Y. and Song, X. Y. (2007). A unified maximum likelihood approach for analyzing structural equation models with missing nonstandard data. *Sociological Methods and Research* **35**, 352–381.
- Lee, S. Y., Poon, W. Y., and Bentler, P. M. (1990). Full maximum likelihood analysis of structural equation models with polytomous variables. *Statistics and Probability Letters* **9**, 91–97.
- Lee, S. Y., Song, X. Y., and Lee, J. C. K. (2003). Maximum likelihood estimation of nonlinear structural equation models with ignorable missing data. *Journal of Educational and Behavioral Statistics* **28**, 111–134.
- Meng, X. L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80**, 267–278.
- Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* **44**, 443–460.

- Schumacker, R. E. and Marcoulides, G. A. (1998). *Interaction and Nonlinear Effects in Structural Equation Models*. Mahwah, NJ: Erlbaum.
- Song, X. Y. and Lee, S. Y. (2004). Bayesian analysis of two-level nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology* **57**, 131–150.
- Song, X. Y., Lee, S. Y., and Hser, Y. (2008). A two-level structural equation approach for analyzing multivariate longitudinal responses. *Statistics in Medicine* **27**, 3017–3041.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**, 583–639.
- Spiegelhalter, D. J., Best, N. G., and Lunn, D. (2003). *WinBugs Version 1.4 User Manual*. Cambridge: MRC Biostatistics Unit.
- Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithm. *Journal of the American Statistical Association* **85**, 699–704.

Further Reading

- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Mahwah, NJ: Wiley.
- Dolan, C. V. and van der Maas, J. J. L. (1998). Fitting multivariate normal mixtures subject to structural equation modeling. *Psychometrika* **63**, 227–254.
- Dunson, D. B. (2002). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society, Series B* **62**, 355–366.
- Lee, S. Y. and Song, X. Y. (2003). Bayesian analysis of two-level nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical and Statistical Psychology* **57**, 29–52.
- Lee, S. Y. and Song, X. Y. (2003). Model comparison of nonlinear structural equation models with fixed covariates. *Psychometrika* **68**, 27–47.
- Moustaki, I. (2003). A general class of latent variable methods for ordinal manifest variables with covariate effects on the manifest and latent variables. *British Journal of Mathematical and Statistical Psychology* **56**, 337–357.
- Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika* **69**, 167–190.
- Scheines, R., Hoijtink, H., and Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika* **64**, 37–52.
- Song, X. Y. and Lee, S. Y. (2002a). Analysis of structural equation model with ignorable missing continuous and polytomous data. *Psychometrika* **67**, 261–288.
- Song, X. Y. and Lee, S. Y. (2002b). Bayesian estimation and testing for nonlinear factor analysis in several populations. *Structural Equation Modeling* **9**, 523–553.
- Song, X. Y. and Lee, S. Y. (2006). Bayesian analysis of structural equation models with non-linear covariates and latent variables. *Multivariate Behavioral Research* **41**, 337–365.
- Zhu, H. T. and Lee, S. Y. (2001). A Bayesian analysis of finite mixtures in the LISREL model. *Psychometrika* **66**, 133–152.

Survival Data Analysis

A F De Champlain, National Board of Medical Examiners, Philadelphia, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Censored observation – Observations that survived to a certain point in time before dropping out from the study for a reason unrelated to the primary outcome of interest (e.g.: insufficient time for an observation to meet the event).

Cox proportional hazards model – A regression method described by D.R. Cox for modeling survival times, generally as a function of several covariates or predictors of interest. It is referred to as the proportional hazards model because it assumes that the underlying hazard rate (rather than survival time) is solely a function of the independent variables and consistent over time.

Event – A qualitative event that can be localized in time.

Hazard function – The function that describes the probability of failure during a very small time interval (assuming that no failures have occurred prior to that time). Hazard is the slope of the survival curve- a measure of how rapidly subjects are having the event (passing an exam, meeting an outcome of the study, etc).

Kaplan-Meier method – Nonparametric method, also known as the product-limit estimator, that is used to compute survival curves.

Log-rank test – A nonparametric test of significance for the difference between two survival curves.

Survival function – The function that denotes the probability that an examinee survives longer than a defined point in time.

Survival curve – The graphical representation of the survival function.

Introduction

Longitudinal outcomes and related studies are commonplace in educational research. Studies have been undertaken to model time to graduation at the college level (Chimka *et al.*, 2008) – and to estimate teacher attrition and student dropout rates (Willett and Singer, 1991) – as a function of various socioeconomic and curricular variables.

In psychometrics, the advent of computer-based testing (CBT) in the admissions, certification, and licensure

arenas, has significantly altered the testing landscape over the past decade (Mills *et al.*, 2002). By virtue of its continuous delivery nature, CBT does, however, require a shift in the way examinee performance trends are analyzed, interpreted, and reported. Longitudinal data are often not amenable to common methods of analysis.

Limitations of Common Regression Models to Analyze Performance Trends

There are several limitations associated with the use of common regression models for estimating performance trends and other longitudinal problems in education. First, it is difficult to accommodate time-varying explanatory variables in typical regression models. One straightforward approach might entail dummy coding the time-varying explanatory variable such that each of its levels (e.g., day, week, month, etc.) is an indicator of the predictor at a given point. However, the latter strategy rests on the assumption that the indicators are predicting the event, while they might very well be their consequences. For example, studies aimed at modeling recidivism rates as a function of whether someone is employed or not at various points in time is a common study problem in criminology (Finn, 1998). In this case, we could fit a model predicting arrest as a function of employment status on a weekly basis. In a linear regression context, this would require dummy coding employment status on a weekly basis for the duration of the study. However, employment status might be an outcome of the event rather than its predictor – that is, being arrested will undoubtedly impact on whether one is subsequently employed or not. This example illustrates why common regression models are inappropriate in this context.

Another important limitation of conventional regression models for analyzing educational data relates to the difficulty in incorporating censored observations, or data missing due to cases that have not yet experienced the event of interest by the end of the study period. Consider the example of an examinee that has taken a test and still has not passed (and presumably still in the pool of potential future masters). These observations are referred to as being right censored, that is, the event time is greater than some constant (the length of the study period). In educational settings, right censoring is far more common than left censoring (event time is less than some constant) or interval censoring (event time is between two numbers a and b). How could this examinee's outcome be coded at

the end of a study aimed at modeling passing rates using linear regression? Two simple approaches to censoring could be considered with linear regression modeling, although both are potentially problematic. The first approach involves discarding cases that have not yet experienced the event of interest – that is, treating them as missing observations. This solution might be defensible if the number of censored cases is very small, but would be unadvisable with data sets containing a large proportion of examinees that still have not passed the examination of interest by the end of the study period, for example. The second approach entails setting a common arbitrary event value for all censored cases. For example, the researcher might decide to fix event times for all censored cases at a value that corresponds to the length of the study period. Adopting this strategy is ill-advised as event times will be underestimates of their true values.

Irrespective of the model being estimated, the presence of censored data requires special estimation procedures or the resulting parameter estimates could be severely biased – as illustrated in the previous example (Bacik *et al.*, 1998). The survival data models that will be outlined in this article make use of information from both censored and noncensored cases to produce consistent parameter estimates. Additionally, these models can more readily accommodate time-varying explanatory variables.

An Overview of Survival Data Analysis

Brief History

Survival data analysis is attractive in that it overcomes limitations of common regression models and appears to be ideally suited to several educational problems – including longitudinally tracking performance trends for various cohorts of examinees, estimating student and faculty attrition rates, as well as any other problem that centers on modeling time to event data (Cox and Oakes, 1984).

Survival analysis is generally acknowledged to have stemmed from Graunt's publication of the first Weekly Bill of Mortality in the seventeenth-century London as well as the printing of the first life-table by Healy (Lee and Go, 1997). However, the development of survival models that are now routinely used can be traced back to studies undertaken in World War II to assess the durability or the survival of military equipment – a critical issue at that time. In the decades that followed these investigations, survival data models were further devised and applied to address problems in a variety of areas. These models are now routinely used to study a host of issues in the pure and social sciences, including recidivism rates (event-history analysis), survival rates in clinical trials, and longevity of components in engineering (reliability analysis).

Some Basic Terminology

A survival data-analysis set consists of a response variable that reflects the duration of time until a prescribed event occurs – that is, the survival time, T , and possibly a set of covariates thought to be associated with T . The basic goal of survival analysis is to model the underlying distribution of the time to event outcome – either with or without intervening covariates. An event is defined as a qualitative change that can be localized in time (Allison, 1995). A covariate corresponds to a predictor, potentially impacting the time to event outcome. To illustrate, consider the problem of modeling time to graduation in a PhD program (the event) as a function of age, gender, and school (the covariates).

The random survival time variable T can be described by the following three functions:

$$S(t) = 1 - F(t), \quad [1]$$

where $S(t)$ is the survival function or the probability of surviving past time t , while $F(t)$ is the cumulative distribution function of T ,

$$f(t) = \frac{dF(t)}{dt}, \quad [2]$$

where $f(t)$ is the probability density function, also referred to as the unconditional failure rate, and

$$b(t) = \frac{f(t)}{S(t)}, \quad [3]$$

where $b(t)$ is the hazard rate or the conditional failure rate.

The graph of $S(t)$ is generally referred to as the survival curve. The survival curve provides estimates of the probability of still being at risk of experiencing an event at various points in time during the study period. The hazard ratio, $b(t)$, can loosely be interpreted as the instantaneous probability of event occurrence at time t . However, it is not really a probability in that it has no upper bound and can, therefore, exceed unity. In the presence of covariates, it is also interpreted as the relative risk of experiencing the event, given covariate values. Survival data-analysis models are, generally, categorized as being either nonparametric/semiparametric or parametric in nature (Lee, 1992). This article aims to provide a broad overview of these models to educational researchers who may be interested in the general topic of survival data analysis.

Nonparametric and Semiparametric Survival Data-Analysis Models

The Kaplan–Meier Product-Limited Method

For data sets with exact times of censoring and events, the Kaplan–Meier (K–M) Product-limited Method is useful for obtaining an initial indication of the shape of the time to event curve prior to fitting more complex models (Kaplan and Meier, 1958).

The K–M method is essentially akin to a nonparametric maximum likelihood estimation technique. The survival estimate obtained by the K–M method is the product of a series of estimated conditional probabilities. The probability of surviving past time period z , for example, is given by:

$$\hat{P}(T > z) = \hat{S}(z) = p_1 \cdot p_2 \cdot p_3 \cdot \dots \cdot p_z$$

where, p_1 is the proportion of examinees surviving to at least time period 1, p_2 is the proportion of examinees surviving to the second time period (after they have survived the first time period), p_3 is the proportion of examinees surviving to the third time period (after they have survived the second time period), and p_z is the proportion of examinees surviving to time period z , (after they have survived to time period $z - 1$).

The probability of surviving past time t , that is, the probability of still being at risk of experiencing the event past time t , is given by

$$\hat{S}(t) = \prod_{j: t_{(j)} \leq t} \left(1 - \frac{d_j}{r_j}\right) \quad [4]$$

where

$\hat{S}(t)$ = the survival function;

r_j = the number at risk just prior to time t_j ; and

d_j = the number of events at time t_j .

Conversely, the estimate of failure, that is, the cumulative probability of experiencing an event at time t is simply given by

$$\hat{F}(t) = 1 - \hat{S}(t) \quad [5]$$

The use of the K–M method is illustrated with a sample of 80 computer programmers completing a 50-item exam aimed at assessing knowledge of a recently completed technical module. The sample is evenly divided between male and female examinees. Additionally, half of the sample completed the module in a half-day in-house training session whereas the remaining group learned the content online. Passing the examination constitutes the event, within a 30-day test-administration window. Surviving, in this instance, is synonymous with not passing the exam and still being at risk of experiencing the event at various points in time. Survival rates $\hat{S}(t)$ are plotted and shown in **Figure 1** for this set of 80 examinees.

The plot shows a steep drop in the curve within the first week or so of the testing window, suggesting that a large proportion of the cohort met the event (passed the test) early on. In fact, the median survival time for this curve was 10 days, further supporting the claim that a large proportion of the group successfully completed the examination in the earlier stage of test administration. Censored observations are symbolized by the + sign. For example, one examinee attempted to pass the test on day 4, but failed to meet the event. This examinee subsequently

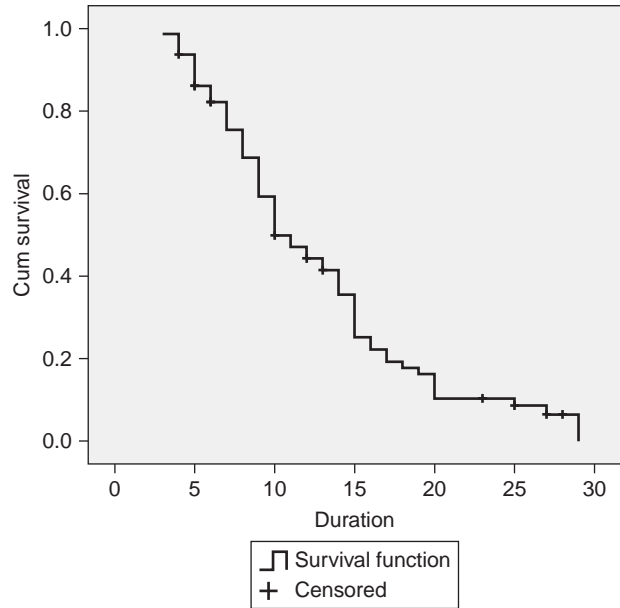


Figure 1 Kaplan–Meier curve for total group-training example.

did not re-attempt the exam (within the 30-day window) and this observation is right censored at day 4.

It may also be of interest to compare two survival curves to see whether event times differ across two treatment groups – for example, as a function of instructional modality (in-house vs. online in the previous example). One useful test to assess whether survival times do differ significantly across strata is the log-rank test. This test is based on the computation of a series of 2×2 contingency tables – one for each time point – of the general form given below,

	<i>Failed</i>	<i>Surviving</i>	<i>At risk</i>
Group A	d_A	$n_A - d_A$	n_A
Group B	d_B	$n_B - d_B$	n_B
Combined	d	$n - d$	n

If there is insufficient evidence to reject the null hypothesis (equal median survival times in both groups), we would expect the total number of failures, d , to be distributed between the two groups A and B in the ratio of the group sizes.

The expected number of failures in group A at time t is given by

$$E_{At} = d \cdot (n_A/n),$$

while the expected number of failures in group B at time t can be obtained by

$$E_{Bt} = d - E_{At}$$

The log-rank test is then computed as,

$$\chi^2 = \frac{(d_A - E_A)^2}{E_A} + \frac{(d_B - E_B)^2}{E_B} \quad [6]$$

where,

d_A = the total number of observed failures in group A;
 E_A = the expected total number of failures in group A;
 d_B = the total number of observed failures in group B; and
 E_B = the expected total number of failures in group B;

It is important to point out that the log-rank test weighs all points of the survival curve equally, that is, it assumes that hazard functions in two groups are proportional over time (they do not cross). Based on our previous example, we were interested in comparing whether the survival functions differed significantly for those examinees that completed the instruction in a traditional fashion (in-house training) versus online. Survival plots for both groups are provided in **Figure 2**. This figure clearly shows that those examinees who completed the in-house training session had a significantly shorter survival time (median survival = 8 days) than those who did the session online (median survival = 15 days), χ^2 (log-rank) = 65.37, $p < 0.0001$.

Cox's Proportional-Hazards Regression Model

In educational research, it is often of interest to assess whether several variables are related to the occurrence of a given event. For example, do sociodemographic and educational variables impact time to graduation, dropout rates, etc? Nonparametric models cannot provide a measure

of the effect of each covariate (adjusted for other predictors in the model) nor can they help the researcher determine which covariate, in a set, is significantly associated with the probability of experiencing an event. Nonetheless, they do constitute useful screening tools to guide subsequent analyses.

Among the many survival data models that have been proposed in the literature, Cox's proportional hazards (PH) regression model (Cox, 1972) is especially useful and popular due to its flexibility in analyzing time-to-event data. The general hazard function for the PH model can be expressed as follows,

$$b(t) = b_0(t) * e^{\beta x} \quad [7]$$

where,

$b_0(t)$ = An arbitrary, non-negative baseline hazard function which does not involve predictor variables.

β_x = Linear combination of p regression coefficients $\{b_1x_{1i} + b_2x_{2i} + \dots + b_px_{pi}\}$.

It is referred to as the PH model because the ratio of the hazards for any two examinees with covariates $x_1 = (x_{11}, x_{21}, \dots, x_{p1})$ and $x_2 = (x_{12}, x_{22}, \dots, x_{p2})$ does not vary with time t , that is,

$$\frac{b_1(t)}{b_2(t)} = \frac{b_0(t)e^{\beta x_1}}{b_0(t)e^{\beta x_2}} = \frac{e^{\beta x_1}}{e^{\beta x_2}}$$

In practice, this function might allow us to model the risk of failing an examination at time t , given the function of time t and the impact of explanatory covariates measured at time t (if applicable). One straightforward way of assessing whether the proportionality assumption holds is to produce a graph plotting the $\log(-\log(\text{survival}))$ versus the log of survival time for the predictor(s) of interest. For illustrative purposes, the latter plot is provided in **Figure 3** for the instruction modality predictor from our previous example. The assumption of proportionality holds for a given predictor when the curves for each level are parallel. As shown in **Figure 3**, the $\log(-\log(\text{survival}))$ versus log of survival time curves are parallel for both groups and support the proportionality assumption.

Cox's PH model is deemed semiparametric in nature as no distribution of survival times is specified. Cox's unique contribution was to devise a means by which to estimate the covariate terms in the model without any consideration of $b_0(t)$, the hazard for an examinee with the value 0 for all the covariates, or the baseline hazard at time t . Since the baseline hazard $b_0(t)$ is not completely specified, it is difficult to use the full likelihood (joint distribution of failure times) to estimate the regression coefficients. Cox (1972) proposed an estimation method which entails maximizing a partial likelihood, as opposed to the full likelihood. The likelihood function for data in a PH model can be factored into the following two parts,

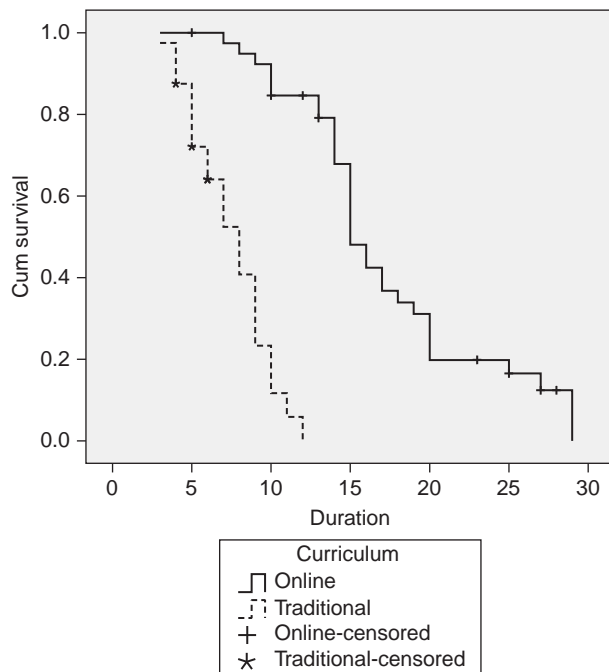


Figure 2 Kaplan-Meier curves by training modality.

$$L = A(\beta) * B(\beta, b_0(t)) \quad [8]$$

where part A contains information relating to the covariates only while part B contains information pertaining to both β s and $b_0(t)$. During the estimation, part B is disregarded and part A is treated as an ordinary likelihood function. Note that part A depends solely on the order that events occurred in, not the exact times. The partial likelihood is then given for all observed events by,

$$L(\beta) = \prod_{k=1}^K \left[\frac{\exp\left(\sum_{j=1}^p \beta_j x_{jk}\right)}{\sum_{l \in R(t_k)} \exp\left(\sum_{j=1}^p \beta_j x_{jl}\right)} \right] \quad [9]$$

where K is the total number of events in the sample and $R(t_k)$ is composed of all examinees whose survival times are at least t_k . Additional computational details on how the partial likelihood is estimated are beyond the scope of this article and can be found in Cox (1972).

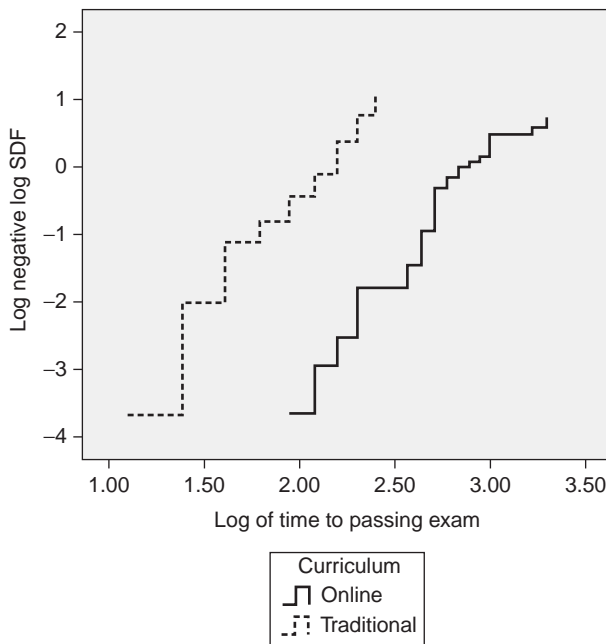


Figure 3 Testing the assumption of proportionality in the Cox PH Model for the training example: Plot of the $\log(-\log(S(t)))$ vs. $\log(t)$.

Cox's PH model has been used in educational research to model a variety of time to event research issues, including passing rates in computer-delivered examinations as a function of educational and background variables (De Champlain *et al.*, 2004, 2006), graduation rates (Chimka *et al.*, 2008; Visser *et al.*, 2007), and response times in CBT (Schnipke and Pashley, 1997).

The fit of the following Cox's PH model was assessed with our previous example, where we attempted to model the time to passing the module examination as a function of gender and instructional method,

$$b(t) = b_0(t)e^{\alpha(t) + \beta_1 x_1 + \beta_2 x_2} \quad [10]$$

where

$b(t)$ = the hazard (risk) of passing the examination at time t ;
 $b_0(t)$ = the hazard for an examinee with the value 0 for all the covariates, i.e., the *baseline hazard* at time t ;

$\alpha(t)$ = function of time t ;

$\beta_1 x_1$ = regression coefficient for gender covariate; and
 $\beta_2 x_2$ = regression coefficient for mode of instruction covariate (traditional or online).

There were no time-dependent covariates in this model – that is, values were assumed to be fixed at all levels of time t .

Initial results suggest that there is sufficient evidence to reject the null hypothesis that all covariate regression coefficients are equal to zero, $LR\chi^2(2) = 58.61, p < 0.0001$. Sixty-eight examinees (85%) had experienced the event by the end of our study period – that is, passed the exam – whereas 12 (15%) were classified as censored observations and presumed to still be at risk of passing in the future. **Table 1** provides the results from Cox's PH model. Regression parameter estimates and their associated standard errors, Wald chi-square tests of significance, and associated type I error rates for each term in the model are shown. It is important to point out that the impact of each covariate is assessed while controlling for other variables in the model. Results show that gender was not significantly related to meeting the outcome – that is, passing the examination, $\chi^2(1) = 0.78, p = 0.38$. Instructional modality was, however, significantly related to meeting the event, $\chi^2(1) = 16.55, p < 0.0001$. Exponentiating the regression parameter value for a given covariate provides one with an estimate of the adjusted hazard ratio, or the relative risk of experiencing an event. Examinees who completed the in-house training (median time to

Table 1 Cox's proportional hazards model: Parameter estimates, standard errors, tests of significance, and type I error rates for training-examination example

Variable	DF	$\hat{\beta}$	$SE(\hat{\beta})$	$\hat{\beta}(95\% CI)$	χ^2	p-value
Gender	1	0.22	0.25	-0.27–0.71	0.78	0.38
Training modality	1	2.81	0.46	1.91–3.71	37.30	<0.0001

event = 8 days) were nearly 17 times more likely ($e^{2.80647} = 16.55$) to experience the event (pass the test) than those who completed the training online (median time to event = 15 days).

Parametric Models in Survival Data Analysis

Parametric approaches for estimating survival curves are appealing – in that, the properties of these models can be used to compute the probability of experiencing an event at different points in time. Unlike Cox's PH model, the latter models are fully parametric – that is, the probability distribution needs to be specified. They also include a scale parameter and are thus sometimes referred to as location-scale models. Popular distributions that have been proposed for this purpose include the Exponential, Weibull, and Log-logistic distributions, to name a few (Lee, 1992).

The exponential model is the oldest and can be viewed as a special case of the more complex Weibull distribution. The hazard function for the exponential model is given by,

$$b(t) = \lambda \quad [11]$$

where λ is a constant hazard function ≥ 0 . This model assumes that T – the function denoting the time of event occurrence – follows an exponential distribution with a constant hazard rate. The hazard function for the Weibull distribution is given by,

$$b(t) = \lambda\gamma(\lambda t)^{\gamma-1} \quad [12]$$

where γ is a parameter reflecting the shape of the hazard function and λ is the scale parameter. Note that the exponential case corresponds to a Weibull distribution with a shape parameter of 1 – that is, a constant hazard rate as time increases. The Weibull model provides greater flexibility in modeling the hazard rate and, as such, has broader application. The hazard rate increases when $\gamma > 1$ and conversely decreases when $\gamma < 1$, as values of time increase. Therefore, the Weibull model is useful for modeling survival rates that display either decreasing, constant, or increasing risk.

Finally, the log-logistic model is appropriate if T follows a log-logistic distribution. The hazard function for this model is given by,

$$b(t) = \frac{\lambda\gamma(\lambda t)^{\gamma-1}}{1 + (\lambda t)^\gamma} \quad [13]$$

where λ and γ , respectively, correspond to the scale and shape parameters of the log-logistic distribution. Unlike the previous parametric models, the log-logistic function has a nonmonotonic hazard function. It is, consequently, ideally suited to model time-to-event data when the researcher suspects that the hazard rate initially increases,

then declines as time values approach infinity. When $\gamma < 1$, the hazard rate for a log-logistic distribution decreases from infinity toward 0, while $\gamma = 1$ reflects a decrease from λ to 0. Finally, the hazard rate increases from 0 to a maximum and subsequently decreases toward 0, when $\gamma > 1$.

Following selection of the survival model, parameters are usually estimated using maximum likelihood. The maximum likelihood time-to-event function estimated in parametric models is given by

$$L = \prod_{i=1}^n f_i(t_i)^{\delta_i} S_i(t_i)^{1-\delta_i} \quad [14]$$

where,

$f_i(t_i)$ = the probability that the event occurs at time t and, $S_i(t_i)$ = the probability that an event occurs after time t (i.e., survival probability).

δ_i = a censoring variable that takes a value of 1 if t is uncensored and 0, if t is censored.

Based on this equation, it can be seen that the information from uncensored observations (those who experience the event) is factored into $f_i(t_i)$, while censored observations contribute to $S_i(t_i)$.

Fit of competing models is typically assessed by comparing the log likelihood values as well as by examining various plots of survival curves. In regard to the first approach, consider L_1 to be the maximum likelihood value for the first parametric model of interest and L_2 , the same value for the second model. A likelihood-ratio chi-square statistic can be computed to compare both models. This test assumes that model 2 is obtained by imposing k restrictions on the parameters of model 1.

For example, one could compare the fit of an exponential model versus a Weibull model, as the former is a restrictive case of the latter function. It is also, generally, recommended to assess fit via plots of various $\hat{S}(t)$ curves. Certain transformations of $\hat{S}(t)$ should be linear if the parametric model fits the data set. Specifically, for an exponential model, the plot of $-\log \hat{S}(t)$ versus time should be linear and pass through the origin.

For the Weibull distribution, the plot of $\log[-\log \hat{S}(t)]$ versus $\log(\text{time})$ should be linear. Finally, for the log-logistic model, the plot of $\log[(1 - \hat{S}(t)/\hat{S}(t))]$ versus $\log(\text{time})$ should be linear. In order to illustrate the use of some parametric survival models, the fit of the exponential and Weibull models to the previous data set was examined.

Results suggest that the Weibull model provided a significantly better fit of the data set than the exponential model, $LR\chi^2(1) = 91.53$, $p < 0.0001$. A plot of $\log[-\log \hat{S}(t)]$ versus $\log(\text{time})$, shown in **Figure 4**, further supports this finding (the trend is approximately linear). Parameters estimated in the Weibull model are shown in **Table 2**. Results are similar to those presented with Cox's PH model and suggest that training modality is significantly

associated with experiencing the event, $\chi^2(1) = 125.84$, $p < 0.0001$; whereas gender is not, $\chi^2(1) = 0.37$, $p = 0.544$. We can estimate the percentage of change in the event time for a unit increase in the predictor by calculating, $100(e^{\hat{\beta}_j} - 1)$, where $\hat{\beta}_j$ is the parameter estimate for predictor j . With regard to training modality, $100(e^{-0.8596} - 1) = -57.67$. Examinees who completed the in-house training program had survival times that were about 58% shorter than those who completed the course online – that is, in-house training examinees experienced the event (passed the exam) in a significantly shorter amount of time than those who chose the other instructional modality.

The usefulness of parametric models is, of course, directly related to the extent to which they fit the observed data set. Consequently, it is critical to assess the fit of each theoretical distribution of interest to the data set, using methods similar to those previously outlined. In practice, however, these models are somewhat restrictive in regard to the possible shape of the hazard function and may not adequately reflect the time to event curve. Another limitation of parametric models in the estimation of time-to-event curves is that they cannot readily accommodate time-dependent covariates in the model (Allison, 1995). Nonetheless, parametric models can be useful approaches to model time-to-event data in instances where the user has some hypothesis as to what the shape of the hazard function might be. Parametric models also allow the

researcher to test the fit of alternative hazard functions – something that is not possible with Cox's PH model.

Modeling Discrete Time Events: The Logit Model

The survival models that have been presented throughout this article were devised under the assumption that an examinee could experience the event at any given point within the study period. However, there are some instances where the units of time are quite large (months, years, etc.) relative to the rate of event occurrence (e.g., instance where a student can graduate at only one point in a yearly academic calendar). In the presence of discrete event-time data, the first step generally entails parsing down an examinee's survival history into discrete events. Once the latter events have been pooled, a logit model can be estimated to predict whether the event of interest did or did not occur at each (discrete) time point,

$$\log\left(\frac{P_{it}}{1 - P_{it}}\right) = \alpha_t + \beta_1 x_{it1} + \dots + \beta_k x_{itk} \quad [15]$$

where, P_{it} is an estimate of the conditional probability of experiencing the event for examinee i at time t given that the event has not yet been met by the examinee, where α_t is the baseline hazard at time t , and where β_k is the regression coefficient for explanatory x_k at time t for examinee i . Discrete survival models have been successfully applied to study a host of educational research issues pertaining to graduation rates (Zwick and Sklar, 2005), as well as student attainment and teacher/faculty attrition patterns (Singer and Willett, 1991, 1993).

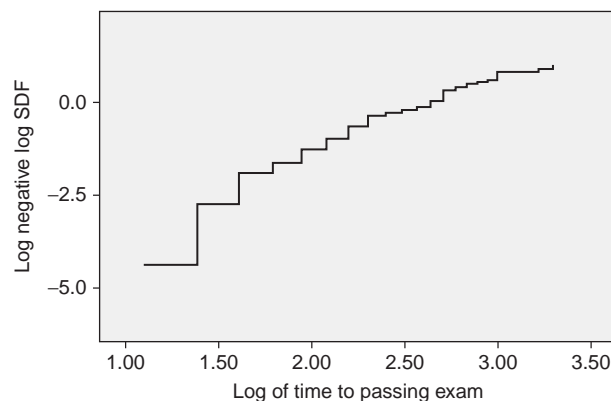


Figure 4 Testing the fit of the Weibull Model: Plot of the log $(-\log(S(t)))$ vs. $\log(t)$.

Software for Survival Data Analysis

At present, the estimation of common nonparametric, semi-parametric, and parametric survival data-analysis models can be done using common statistical packages – including SPSS and SAS. Life-tables, K–M survival curves, and log-rank tests can be computed using the LIFETEST procedure in SAS while they are estimated using the SURVIVAL and KM analyses in SPSS. Parametric survival data analyses can be fit to data sets in SAS using the LIFEREG procedure.

Table 2 Weibull regression model: Parameter estimates, standard errors, tests of significance, and type I error rates for training-examination example

Variable	DF	$\hat{\beta}$	SE($\hat{\beta}$)	$\hat{\beta}$ (95% CI)	χ^2	p-value
Intercept	1	3.91	0.16	3.59–4.23	578.44	<0.0001
Gender	1	−0.05	0.08	−0.20–0.10	0.37	0.54
Training modality	1	−0.86	0.08	−1.01–0.71	125.84	<0.0001
Scale	1	0.31	0.03	0.26–0.38		
Weibull shape	1	3.18	0.30	2.65–3.82		

Although there is no dedicated procedure to estimate parametric survival data-analysis models in SPSS, common functions can be setup through either the constrained nonlinear regression (CNLR) or nonlinear regression (NLR) methods. The PH model can be run in SAS using the PHREG procedure and can be estimated in SPSS using the Cox Regression (COXREG) method. Finally, the logit model can be estimated in SAS using either the CATMOD, GENMOD, or LOGISTIC procedures. In SPSS, similar models can be estimated with the LOGLINEAR procedure.

See also: Educational Data Modeling; Multivariate Longitudinal Data Analysis; Sequential Testing; Time Series Analysis.

Bibliography

- Allison, P. (1995). *Survival Analysis Using the SAS System: A Practical Guide*. Durham, NC: BBU Press.
- Bacik, J. M., Murphy, S. A., and Anthony, J. C. (1998). Drug use prevention data, missed assessments and survival analysis. *Multivariate Behavioral Research* **33**, 573–588.
- Chimka, J. R., Reed-Rhoads, T., and Barker, K. (2008). Proportional hazards models of graduation. *Journal of College Student Retention: Research, Theory and Practice* **9**, 221–232.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society* **B34**, 187–220.
- Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.
- De Champlain, A. F., Sample, L., Dillon, G. F., and Boulet, J. R. (2006). Modeling longitudinal performances on the USMLE and the impact of socio-demographic covariates: An application of survival data analysis. *Academic Medicine* **81**, s108–s111.
- De Champlain, A. F., Winward, M. L., Dillon, G. F., and De Champlain, J. E. (2004). Modeling passing rates on a computer-based medical licensing examination: An application of survival data analysis. *Educational Measurement: Issues and Practice* **23**, 15–22.
- Finn, P. (1998). Job placement for offenders in relation to recidivism. *Journal of Offender Rehabilitation* **28**, 89–106.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457–481.
- Lee, E. T. (1992). *Statistical Methods for Survival Data Analysis*. New York: Wiley-Interscience.
- Lee, E. T. and Go, O. T. (1997). Survival analysis in public health research. *Annual Review of Public Health* **18**, 105–134.
- Mills, C. N., Potenza, M. T., Fremer, J. J., and Ward, W. C. (2002). *Computer-Based Testing: Building the Foundation for Future Assessments*. Mahwah, NJ: Erlbaum.
- Schnipke, D. L. and Pashley, P. J. (1997, March). *Assessing subgroup differences in item response times*. Paper presented at the meeting of the National Council on Measurement in Education, Chicago, IL.
- Singer, J. D. and Willett, J. B. (1991). Modeling the days of our lives: Using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. *Psychological Bulletin* **110**, 268–290.
- Singer, J. D. and Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics* **18**, 155–195.
- Visser, M. S., Luwel, M., and Moed, H. F. (2007). The attainment of doctoral degrees at Flemish universities: A survival analysis. *Higher Education: The International Journal of Higher Education and Educational Planning* **54**, 741–757.
- Willett, J. B. and Singer, J. D. (1991). From whether to when: New methods for studying student dropout and teacher attrition. *Review of Educational Research* **61**, 407–450.
- Zwick, R. and Sklar, J. C. (2005). A note on standard errors for survival curves in discrete-time survival analysis. *Journal of Educational and Behavioral Statistics* **30**, 75–92.

The Normal Distribution and its Applications

J Tacq, Catholic University of Brussels, Brussels, Belgium

© 2010 Elsevier Ltd. All rights reserved.

Normal distribution is also called de Moivre distribution, Gaussian distribution, and Gauss–Laplace distribution, due to the work of Abraham de Moivre, Johann Carl Friedrich Gauss, and Pierre-Simon Laplace in the eighteenth century.

It is often called the bell curve, a name going back to Jouffret in 1872, because the graph of its probability density resembles a bell (Figure 1).

The name normal distribution was coined independently by Charles S. Peirce, Francis Galton, and Wilhelm Lexis around 1875.

Forerunners

Carl Friedrich Gauss became associated with normal distribution when he analyzed astronomical data. He defined the equation of its probability density function (PDF). But it was in fact Abraham de Moivre who first introduced normal distribution (a term used only since the second-half of the nineteenth century) in an article in 1733, which was reprinted in the second edition of his *The Doctrine of Chances*, 1738, as an approximation to the binomial distribution for large values of n . His result was extended by Laplace in his book *Analytical Theory of Probabilities* (1812).

Characterization and Probability Density Function

Normal distribution represents a family of continuous probability distributions, applicable in many fields. It may be defined by two parameters, location and scale: the mean (average, μ) and variance (variability, σ^2), respectively. To indicate that a real-valued random variable X is normally distributed with mean μ and variance σ^2 (≥ 0), we write: $X \sim N(\mu, \sigma^2)$ (Case study 1 and Figure 2).

The continuous PDF (Figure 3) of normal distribution is given by the following function:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}$$

This density function is symmetric about its mean value μ . The mean μ is also its mode and median.

The points of inflection of the curve (i.e., the points where the curve changes from slowly increasing to rapidly increasing, or from rapidly decreasing to slowly decreasing) occur at one standard deviation away from the mean, that is, at $\mu - \sigma$ and $\mu + \sigma$.

About 68% of values drawn from a normal distribution are within one standard deviation away from the mean, about 95% of the values are within two standard deviations (95.4 to be more precise), and about 99.7% lie within three standard deviations. This is known as the 68–95–99.7 rule or the empirical rule (Case study 2 and Figure 4).

Case study 1

An example of the combined action of location (central tendency) and scale (dispersion) is the sleeping behavior of students. In an inquiry among students at Leuven University, Belgium, 3 months before exams, 1 month before exams, and during exams, the number of sleeping hours were normally distributed. The means and variances appeared to be, respectively: (a) 8 and 9, (b) 7.5 and 1.56, and (c) 7 and 0.25. We see that not only the arithmetic means (8, 7.5, 7) but also the variances (9, 1.56, 0.25) are decreasing, the former meaning that students sleep less and less on the average and the latter indicating that the sleeping behavior becomes more and more homogeneous as the exams come nearer.

Case study 2

An example of practical use of the empirical rule is the length of first names given to babies. In a sociolinguistic inquiry of babies in the North of Belgium, a comparison of the number of letters used in first names of baby-boys and baby-girls was planned and a difference of means test was prepared. In order to perform this test, the researchers had to estimate the standard deviation of the population (σ) and – as they were in the dark about this – they used the empirical rule as follows. They assumed a normal distribution. The largest name, such as Marginus-Christianus-Reginus for boys or Cornelia-Elisabeth-Anastasia for girls, contained 26 letters. The shortest name, such as Jo for boys or An for girls, consisted of two letters. So, the range (highest minus lowest length) was $26 - 2 = 24$ letters. Now, we know that in a normal distribution almost 100% of the population (95.4%) is situated between $\mu - 2\sigma$ and $\mu + 2\sigma$ and that an extra standard deviation left and right does not add much (only 2.3% extra according to the empirical rule). It follows that the range is approximately equal to 4 standard deviations. Therefore, the standard deviation σ was (roughly) estimated as $24/4 = 6$.



Figure 1

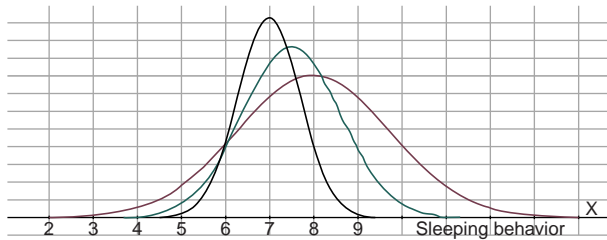


Figure 2

Cumulative Distribution Function

The cumulative distribution function (CDF) of a probability distribution contains the probabilities that a random variable X is less than or equal to X . The cumulative distribution function of the normal distribution (Figure 5) is expressed as follows:

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx, \quad x \in \mathbb{R}$$

Standard Normal Distribution

In many applications in which some random variable X is normally distributed with mean μ and variance σ^2 , we will standardize X to obtain z -scores ($z = (X - \mu)/\sigma$). The distribution of the z -scores is the standard normal distribution, that is, the normal distribution with a mean of zero and a variance of one. Hence, if $X \sim N(\mu, \sigma^2)$, then $z \sim N(0, 1)$.

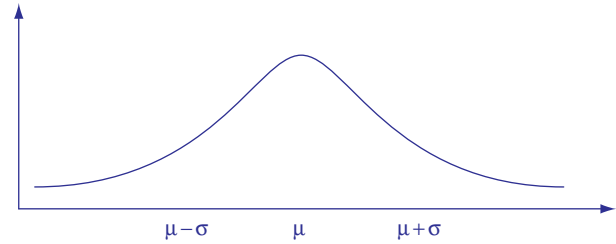


Figure 3

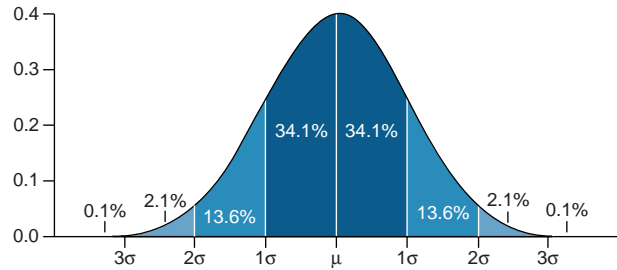


Figure 4

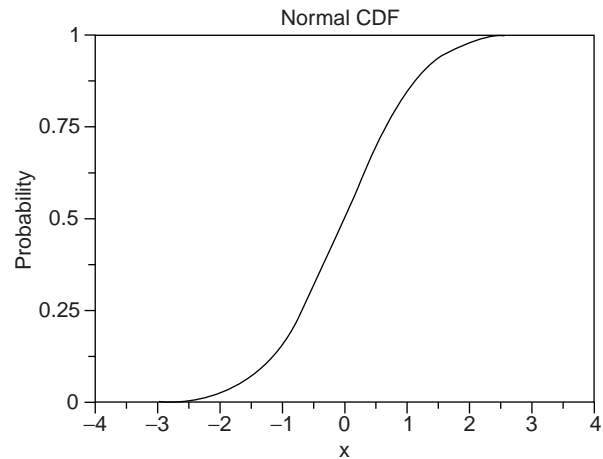


Figure 5

The probability density function of the standard normal distribution is as follows:

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2), \quad z \in \mathbb{R}, \quad \text{where } z = (X - \mu)/\sigma$$

Standard normal distribution is tabulated, usually in the form of values of the CDF, Φ . As normal distributions are simple transformations of the standard one, the tabulated values of the CDF of the standard normal distribution can be used to find the values of the CDF of the normal distribution in many practical applications (Figure 6).

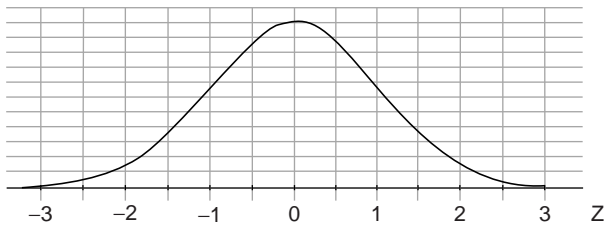


Figure 6

The standard normal distribution is very useful to calculate probabilities. Take the above-mentioned example of sleeping behavior of students at Leuven University. Three months before exams, the mean was 8 h per night and the variance was 9 h. If a student is selected at random, what is the probability of sleeping (a) less than 4 h per night, (b) more than 12 h, and (c) between 4 and 12 h? To find these probabilities, we have to first convert the values of variable X to values of the standardized variable z . The probabilities are then given by tables for the normal function. The solutions are as follows: (a) For $X=4$ we have a z -score of $(4-8)/3 = -1.33$. The area under the normal curve left of $z = -1.33$ is the same as the area right of $z = 1.33$ (because of symmetry), which is $1 - 0.9082 = 0.0918$ (tables). So, a proportion of 9% of the students sleep less than 4 h. (b) For $X=12$ we have $z = (12-8)/3 = 1.33$. This results in the same probability of 0.0918. So 9% of the students sleep more than 12 h. (c) For $X=4$ and $X=12$, we have z -scores of -1.33 and $+1.33$, respectively. The probability left of $z = -1.33$ was calculated to be 0.0918. The probability left of $z = 1.33$ is 0.9082. The difference is $0.9082 - 0.0918 = 0.8164$. So, there is a probability of more than 81% for a student to sleep between 4 and 12 h.

Properties of Normal Distribution

Some properties of normal distribution are applicable to a variety of statistical procedures:

1. If random variable X is normally distributed ($X \sim N(\mu, \sigma^2)$) and a and b are real numbers, then $aX + b$ is normally distributed with mean $a\mu + b$ and variance $(a\sigma)^2$, that is, $aX + b \sim N(a\mu + b, (a\sigma)^2)$.
2. Addition and subtraction of two independent normal random variables result in a normal distribution, but not so for multiplication and division (e.g., the ratio follows a Cauchy distribution, not a normal distribution). So, if $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$, then $X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$ and $X - Y \sim N(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2)$.
3. If X_1, X_2, \dots, X_n are independent standard normal variables, then the sum of their squares $X_1^2 + X_2^2 + \dots + X_n^2$ has a chi-square distribution with n degrees of freedom.

Central Limit Theorem

The importance of normal distribution as a model of quantitative phenomena in the natural and behavioral sciences is due to the central limit theorem. Under certain conditions (such as being independent and identically distributed with finite variance) the sum of a large number of random variables is approximately normally distributed – this is the central limit theorem. Many psychological measurements and physical phenomena (like noise) can be approximated well by the normal distribution. While the mechanisms underlying these phenomena are often unknown, the use of the normal model can be theoretically justified by assuming that many small, independent effects additively contribute to each observation.

From Binomial to Normal

The practical importance of the central limit theorem is that the normal cumulative distribution function can be used as an approximation to some other cumulative distribution functions, for example: a binomial distribution with parameters n and p is approximately normal for large n and for p not too close to 1 or 0. The approximating normal distribution has parameters $\mu = np$, $\sigma^2 = np(1-p)$, where an event (e.g., throwing heads with a coin) can occur or not occur and where p = the probability of occurrence of the event (heads) and $q = 1 - p$ = the probability of nonoccurrence of the event, that is, occurrence of the alternative event (tails). We say that the random variable X has a binomial distribution with parameters n and p and we write $X \sim B(n, p)$. In the example of throwing a fair coin, p and q are both equal to $\frac{1}{2}$. In that case, the binomial distribution is symmetrical. An example in which p and q are unequal would be random variable X = the number of left-handed children. In a school of 500 children with a probability of 10% of meeting a left-handed child, X will show a binomial distribution $X \sim B(500, 0.1)$. Such a distribution would be much skewed if p is extremely small or large, but when n is large, symmetry is restored. The binomial distribution almost corresponds to the normal distribution if p is close to 50% and/or the sample size is large. In the example of left-handed children p is only 10%, but as n is rather large (500) normal approximation holds.

The latter, the tendency of the binomial toward the normal distribution, is nicely illustrated by Galton's board. At the top is a funnel in which balls are dropped. In the middle we see metal bars at an equal distance from each other. At the bottom are receptacles. Each time a ball hits a bar it has just as much probability of falling to the right as to the left ($p = 50\%$). In order for a ball to fall into the receptacle at the extreme left, it must fall to the left each time it hits a bar. The probability that this will occur is, of course, small. The same holds for the extreme right.

The probability of landing in a centrally located receptacle is very great, because a ball may fall first to the right then to the left, and so on. The result with a large number of balls is the Gauss curve of normal distribution, as shown in **Figure 7**.

As the value of n increases and p does not deviate too much from $1/2$, the PDF of the binomial distribution will be obtained by the following form:

$$f(X) = \frac{1}{\sqrt{npq2\pi}} \exp\left(-\frac{(X - np)^2}{2npq}\right)$$

in which p and q have the same meaning as before and $\pi = 3.14$ and $e = 2.72$. This is the normal distribution with $\mu = np$ and $\sigma^2 = npq$. We can now illustrate the normal approximation of the binomial distribution by means of the example of left-handed children. The school has $n = 500$ children. The probability of meeting a left-handed child is $p = 0.1$. Suppose we want to calculate the probability of there being more than 60 left-handed children in the school. Making use of the formulas of binomial distribution, we would have to calculate p (more than 60) = $p(61) + p(62) + p(63) + \dots + p(500)$. This would be an enormous calculation task. Using the normal approximation makes the task rather easy. $X \sim B(500, 0.1)$ will be approximated by $N(50, 45)$, for we have $\mu = np = 500(0.1) = 50$ and $\sigma^2 = npq = 500(0.1)(0.9) = 45$. The standard deviation σ is $\sqrt{45} = 6.7$. Transformation to the standard normal distribution of a number of $X = 60$ left-handed children gives a standardized score of $z = (60 - 50)/(6.7) = 1.5$. From the table of areas under the (cumulative standard) normal curve we have $p(X \leq 60) = p(z \leq 1.5) = 0.93$. It follows that $p(X > 60) = 1 - 0.93 = 0.07$,

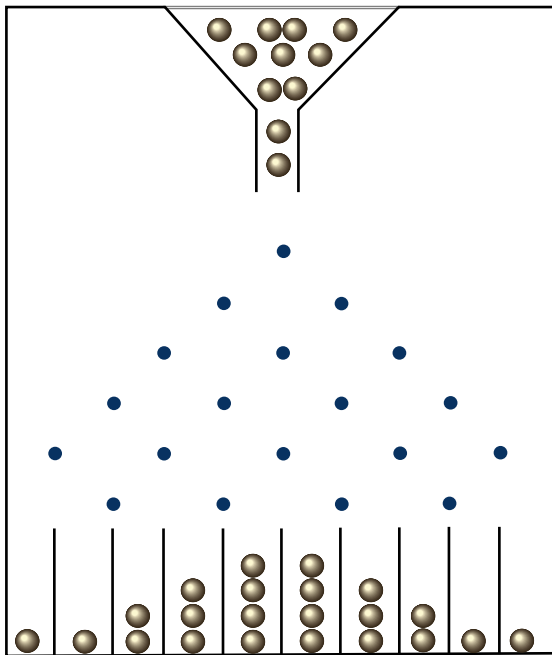


Figure 7

which means that the probability of there being more than 60 left-handed children is 7% (**Figure 8**).

When $p = q = \frac{1}{2}$, we can make a similar calculation to show what is the probability of obtaining between 470 and 530 heads, when 1000 fair coins are thrown. The binomial distribution $B(1000, 0.5)$ is approximated by the normal distribution $N(500, 250)$, for the mean is $\mu = np = 1000(\frac{1}{2}) = 500$ and the variance is $\sigma^2 = npq = 1000(\frac{1}{2})(\frac{1}{2}) = 250$. The standard deviation is $\sqrt{250} = 15.8$. The values $X = 470$ and $X = 530$ have standardized scores of $z = (470 - 500)/(15.8) = -1.9$ and $z = (530 - 500)/(15.8) = 1.9$, respectively. From the table of areas under the (cumulative standard) normal curve we obtain the following probabilities: $p(X \leq 470) = p(z \leq -1.9) = 0.03$ and $p(X \leq 530) = p(z \leq 1.9) = 0.97$. The difference 0.94 is the requested probability of obtaining heads numbering between 470 and 530 heads (**Figure 9**).

Examples, Occurrence

It has been repeated again and again, in a great number of textbooks that (approximate) normal distributions occur in many situations, as a result of the central limit theorem, for, when there is reason to suspect the presence of a large number of small effects acting additively and independently, it is reasonable to assume that observations will be normal.

We would like to call in question this overrating of the bell curve, for effects can also act in a multiplicative (rather than additive) fashion. In that case, the assumption of normality is not justified and it is rather the logarithm of the variable of interest that is normally distributed. The distribution of the directly observed variable is then called log-normal. Financial variables are a good example. Financial indicators such as stock values and commodity prices exhibit multiplicative (rather than additive) behavior because of the exponential nature of inflation, and normal distribution does often not apply to them.

Another example of non-normality is the truncated distribution, which seems to occur in a great many cases of social science investigations (**Case study 3** and **Figure 10**).

Looking at positive examples of normality, the most-quoted applications occur in the area of measurements, for example, measures of length, height, skin area, (for weight sometimes the logarithm thereof), and the length of appendages like hair, claws, nails, and teeth. Another

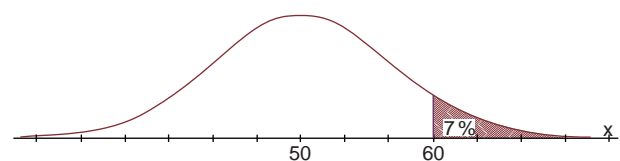


Figure 8

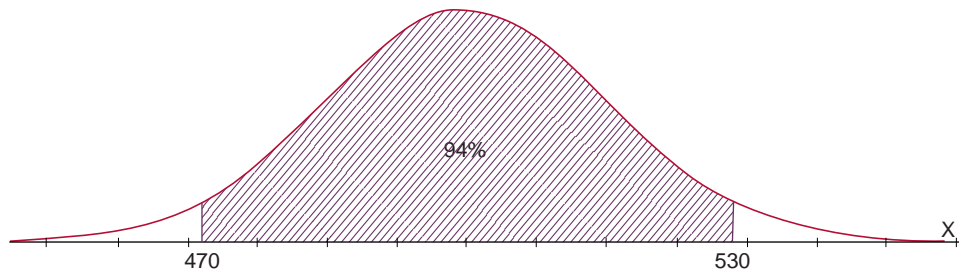


Figure 9

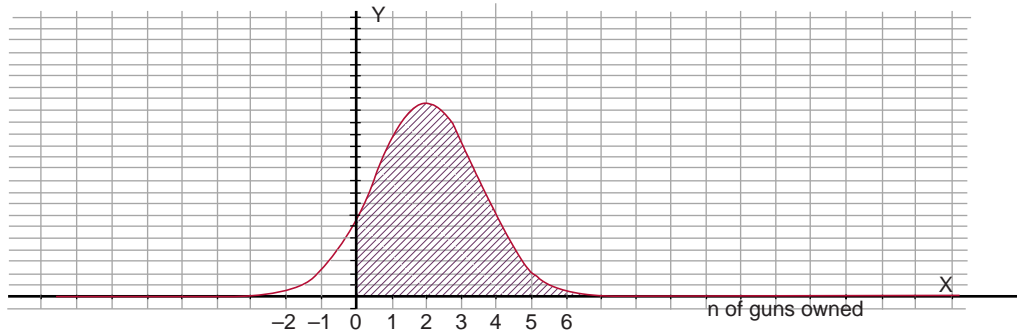


Figure 10

Case study 3

A truncated distribution is a distribution in which the top or the bottom has been removed. An example could be the number of weapons owned. In an inquiry among legislators supporting or opposing a bill on gun control, one of the variables was the number of weapons they owned. The distribution had almost the shape of a normal curve, but was truncated at zero, because a negative number of weapons is impossible. The mean was equal to 2, the standard deviation was approximately equal to 2, the range was between 0 and 6, and one had the impression that, if a negative number of weapons would have been possible, a beautiful bell curve would have emerged ranging from -2 to 6 with the mean of two weapons owned in the middle. The truncation at zero seemed to cause a deviation from normality.

In many examples in social science there seems to be such a left or right truncation.

example of measurement is the blood pressure of adult humans, which is supposed to be normally distributed for males and females separately.

Normality is more particularly the central assumption of the mathematical theory of errors. For example, repeated measurements of the same quantity are expected to yield results which are clustered around some value. If all major sources of errors have been taken into account, it is assumed that the remaining error must be the result of a large number of very small additive effects, and hence normal (because of the central limit theorem).

Similarly, in statistical model-fitting, an indicator of goodness of fit is that the residuals (=errors) be independent and normally distributed.

Many intelligence quotient (IQ)-tests lead to IQ-scores being normally distributed for a variety of populations in society. However, the question whether intelligence itself is normally distributed is more involved, among other reasons because intelligence is a latent variable; therefore, its distribution cannot be observed directly. This question whether or not intelligence represents a bell curve is the topic of the next section.

The Myth of the Bell Curve

Variation in intelligence became the subject of scientific study in the nineteenth century, stimulated by Charles Darwin's theory of evolution. Francis Galton, Darwin's young cousin, tried to devise an intelligence test as we understand the term today. It was his most influential successor, a French psychologist, Alfred Binet, who developed the technique to measure IQ which is still used today. He determined which problems could be solved at which age and then analyzed whether a person had a mental age above or under his real age. By dividing mental age by real age, he obtained a quotient, called IQ. Then, in 1904, a former British Army officer named Charles Spearman made a statistical breakthrough in using factor analysis as a technique for finding what different mental tests have in common, and a conceptual breakthrough in

hypothesizing that all different forms of intelligence can be reduced to one general trait, which he named g for general intelligence.

In line with the central limit theorem, the idea became widespread that a person's intelligence might be the sum of many small random variations in genetic and environmental factors, and that, consequently, IQ-scores are normally distributed. It was for a long time generally accepted that the mean of this distribution is 100 with a standard deviation of 15.

It was common to consider an IQ-score of 70 or less, that is, two or more standard deviations below the mean, as the borderline for 2.5% of the people needing special care and training, who have been labeled moron, imbecile, idiot, cretin, feeble-minded, backward, and the like. As sensitivity, kindness, and political correctness became an issue, these terms were replaced by educationally subnormal, in special needs, educable mentally retarded, trainable mentally retarded, etc. Severity of mental retardation is sometimes broken down into four levels: 50–70 – mild mental retardation, 35–50 – moderate mental retardation, 20–35 – severe mental retardation, and <20 – profound mental retardation.

At the other side of the distribution, an IQ-score of 130 or more, that is, two or more standard deviations above the mean, represents the top 2.5%, who are labeled bright, highly gifted, brilliant, genius, etc. Genius IQ is generally considered to begin around three standard deviations above the mean, that is, 145, representing 0.15% of the population (see the empirical rule explained above: $(1 - 99.7)/2 = 0.15$). A rough guide for levels of brightness is as follows: 135–144 – highly gifted (intellectuals), 145–154 – genius (professors), 155–164 – genius (Nobel prize winners), 165–179 – high genius, 180–200 – highest genius, and >200 – unmeasurable genius. Cyril Burt (1963) names John Stuart Mill, Sir Hamilton, Lord Macaulay, Lord Kelvin, and Sir Francis Galton, whose IQ must have been approximately 200. Einstein was considered to have an IQ of only about 160.

The middle group between one standard deviation below and above the mean, representing 68% of the population (see the empirical rule), is considered normal (**Figure 11**).

In the 1960s, a new controversy about intelligence tests emerged that continues to this day. Psychometricians had long-time debated whether intelligence is almost entirely produced by genes or whether the environment also plays a role, the so-called nature–nurture debate. By the 1960s, the point of contention had shifted dramatically. It became controversial to claim that genes had any effect at all on intelligence. When Arthur Jensen wrote – in 1969 – that education program had yielded disappointing results, and concluded that success in school depended to a considerable degree on IQ, that IQ had a large heritable component, and also that historically blacks had exhibited average IQs substantially below those of whites, the reaction was immediate and violent. Dozens of books and hundreds of articles appeared arguing that mental abilities are determined by environment, with genes playing a minor role and race none at all. A best seller was *The Mismeasure of Man*, published by Harvard paleobiologist Stephen Jay Gould, who argued that intelligence is a bankrupt concept, is culturally biased (they measure white IQ, not black IQ), IQ scores are not constant over an individual's life span and are unrelated to earnings, occupation, productivity, and other important measures of success. There is not just one primary ability, g for – so Gould argued – other scientists have hypothesized many more, two according to Raymond Cattell, half-dozen in Louis Thurstone's theory, a hierarchy in Philip Vernon's, and 120 or more in Joy Guilford's, and moreover, a simple rotation of the factor analysis solution shows that the general factor g disappears.

In this context of discussion, a controversial best-selling book *The Bell Curve* appeared in 1994, which was written by Harvard professor Richard Herrnstein and political scientist Charles Murray, and which ran against the current. Its central point is that intelligence – which is 60–80% determined by nature and 20–40% by nurture – is a better predictor of many factors including financial

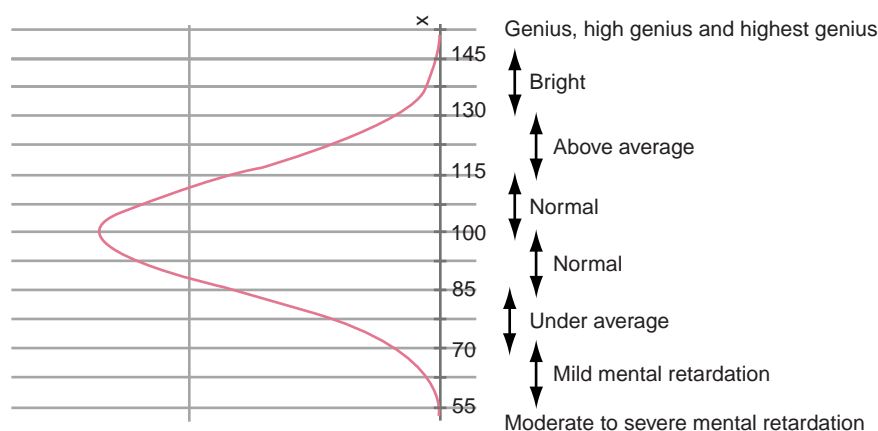


Figure 11

income, job performance, unwed pregnancy, and crime, than parents' socioeconomic status or educational level. The authors write that the debate about whether and how much genes and environment have to do with ethnic differences remains unresolved, and yet the media and many scholars condemned the book, notwithstanding their subtle and balanced text. The discussion has dominated the 1990s and thereafter and one can really fill a whole library with the debate about fueling racism. The short sentence of the long debate is that all the discussions boil down to the nature–nurture debate, so that we really seem to end where we started, albeit on a higher level of mental power. Recommended to the reader!

Applications in Statistics

Normal distribution arises in many areas of statistics. For example, the sampling distribution of the sample mean is approximately normal, even if the distribution of the population from which the sample is taken is not normal, on the condition that the sample is sufficiently large. This principle, which is an application of the central limit theorem, is widely used in statistical estimation and testing.

Many scores are derived from normal distribution, including percentiles, normal curve equivalents, stanines, z -scores, and T -scores.

Normal distribution is also most widely used in statistical tests, because many tests, such as T -tests and ANOVAs, are based on the assumption of normality.

In many cases, for example, when tests are based on the assumption of normality, this assumption has to be tested by means of a normality test, that is, a check whether a given set of empirical data show similarity to normal distribution. The null hypothesis is that the data set is similar to normal distribution; therefore, a sufficiently small p -value indicates deviation from normality. Existing tests of normality are:

- Kolmogorov–Smirnov test,
- Lilliefors test,

- Anderson–Darling test,
- Ryan–Joiner test,
- Shapiro–Wilk test,
- Normal probability plot, and
- Jarque–Bera test.

Special applications of normal distribution can also be found in Q -factor analysis. A generalization is the multivariate normal distribution, which is discussed elsewhere in this encyclopedia.

See also: Multivariate Normal Distribution.

Bibliography

Burt, C. (1963). Is intelligence distributed normally? *British Journal of Statistical Psychology* **XVI**(2), 175–190.

Further Reading

- Belsom, C. (1992). *The Normal Distribution. The School Mathematics Project 16–19*. Cambridge: Cambridge University Press.
- Binet, A. and Simon, Th (1911). *A Method of Measuring the Development of the Intelligence of Young Children*. Lincoln: Courier.
- Blalock, H. (1960). *Social Statistics*. New York: McGraw-Hill.
- de Moivre, A. (1738). *The Doctrine of Chances*. London: Woodfall.
- Goossens, R. (2005). *De Controverse rond 'The Bell Curve' (Controversy around 'The Bell Curve')*. Thesis, Leuven, Faculty of Social Sciences, Department of Sociology.
- Gould, S. J. (1981). *The Mismeasure of Man*. New York: W.W. Norton.
- Green, P. (1978). *Analyzing Multivariate Data*. Hinsdale, IL: Dryden.
- Herrnstein, R. J. and Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.
- Laplace, P.-S. (1812). *Analytical Theory of Probabilities*. Paris: V. Courcier.
- Tacq, J. (1997). *Multivariate Analysis Techniques in Social Science Research*. 411p. London: Sage.
- Tong, Y. L. (1990). *The Multivariate Normal Distribution*. 271p. Berlin: Springer.
- Vanhoutvinck, J. (1975). *Statistiek met wiskundige inleiding (Statistics with Mathematical Introduction)*. Part I, 123p and Part II, 72p Leuven: ACCO.

Time Series Analysis

P J Brockwell, Colorado State University, Fort Collins, Co, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

AIC – Akaike's information criterion, eqn [18], used to select the orders p and q of an appropriate ARMA model to represent observations of a stationary time series.

AICC – A version of the AIC criterion, corrected to improve model selection for short time series.

ARIMA process – A time series which, after differencing one or more times, becomes an ARMA process.

ARMA process – A stationary time series satisfying difference equations of the form [12].

Autocorrelation function – For a stationary time series $\{Y_t\}$, the correlation $\rho(h)$ (eqn [5]) between Y_{t+h} and Y_t is independent of t for every h , and the function $\rho(h)$, $h = 0, \pm 1, \pm 2, \dots$, is called the autocorrelation function of the time series $\{Y_t\}$.

Sample autocorrelation function – From observed values y_1, \dots, y_n , of a stationary time series, the autocorrelation $\rho(h)$ is estimated by the sample autocorrelation function $\hat{\rho}(h)$, defined in terms of the observed data by eqn [8].

Stationary time series – A Sequence of random variables $\{Y_t\}$ satisfying eqn [4]. (The term '2nd-order stationary time series' is often used to distinguish this definition of stationary from that of 'strict stationarity'.)

observations which are made at uniformly spaced times. For this reason, the present article focuses on this case.

Example 1

Figure 1 shows the annual percentages of 18- and 19-year-olds enrolled in schools, colleges, universities, or professional colleges for the years 1947 through 2006. (US Department of Commerce, Census Bureau, Current Population Survey (CPS), October, 1970–2006). The graph shows a clear upward trend with apparently random fluctuations from year to year. For planning purposes, we would like not only to forecast this series, but at the same time produce estimates of the possible errors to be expected in the forecasts. This is a typical problem of time series analysis.

Example 2

Figure 2 shows the monthly birth rate (births per year per 1000 inhabitants) of New York City from January 1947 through December 1959 (Hyndman, R. J. Time Series Data Library). As is often the case with monthly data, the series strongly suggests the presence of a component which repeats itself every 12 months, in this case with a peak around July and a trough in February, corresponding to the seasons of the year. In addition, there appears to be a smooth upward trend and a random component accounting for irregular deviations from the sum of the trend and seasonal components.

Time Series Data

A time series is any sequence of observations recorded at specified times and usually displayed as a time-series plot. This is a graph in which the observations are plotted as a function of time. Examples are shown in **Figures 1** and **2**. Time series abound in all branches of science, engineering, sociology, and economics, and in fact in every field in which observations are recorded over a period of time. The set of times T at which observations are recorded may be a discrete set, as is the case when the observations are recorded at uniformly spaced times (daily rainfall, hourly temperature, annual income, etc.) or it may be a continuous interval, as when the data are recorded continuously (e.g., by a seismograph or electrocardiograph). A very large number of practical problems involve

Objectives

Time series analysis has many different objectives, depending on the field of application. These include forecasting future values of the series, extracting a signal hidden in noisy data, discovering the mechanism by which the data are generated, simulating independent realizations of the series to see how it might behave in the future (and hence, for example, to estimate the probability of extreme events like floods), and eliminating the seasonal component from data sets like the one in example 2 in order to reveal more clearly the underlying trend. In studying monthly economic statistics, for example, it is important to carry out this so-called deseasonalization so as not to confuse the long-term trend with the month-to-month seasonal variation. For all such applications, time

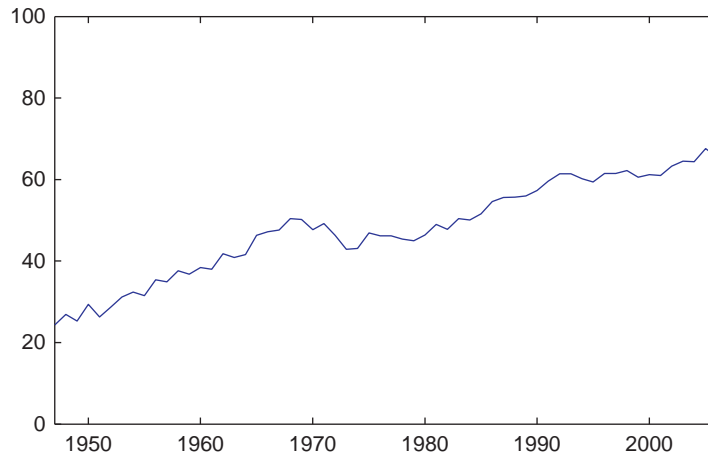


Figure 1 Annual percentages of 18- and 19-year-olds enrolled in educational institutions from 1947 through 2006 (Example 1).

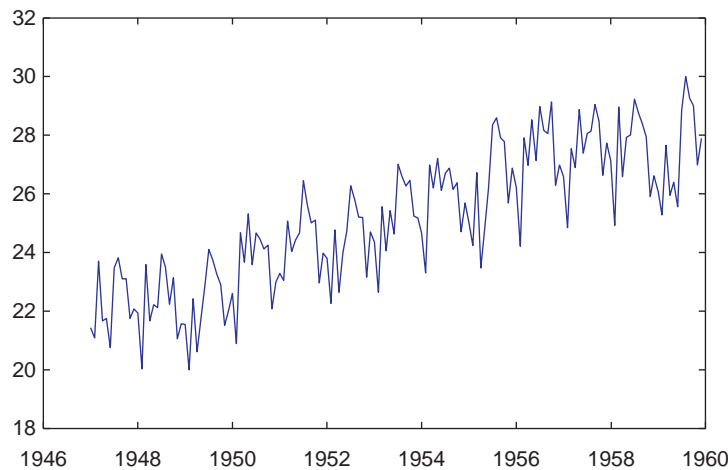


Figure 2 Monthly birth rates in New York City, January 1947 through December 1959 (Example 2).

series analysis usually begins with an attempt to find a mathematical model which provides a good representation of the observed data.

Although it is rarely the case that there is a true mathematical model underlying empirical data, systematic procedures have been developed for selecting a good model, according to clearly specified criteria. Once this is achieved, the selected model can be used to address the questions posed in the preceding paragraph.

Time Series Models

As indicated above, the graph of the monthly birth rates in **Figure 2** suggests representing the rate x_t at time t as the sum of a slowly varying trend component, a period-12 seasonal component and a random component which accounts for the irregular deviations from the sum of the other two. In order to take account of the randomness, we suppose that for each t , the observation x_t is just one of

many possible values of a random variable X_t which we might have observed. This leads to the following classical decomposition model for the birth-rate series,

$$X_t = m_t + s_t + Y_t, t = 1, 2, 3, \dots, \quad [1]$$

where the sequence $\{m_t\}$ is the trend component describing the long-term movement of the series, $\{s_t\}$ is a seasonal component with known period (in this case 12) and $\{Y_t\}$ is a sequence of random variables with mean zero, referred to as the random component. If we can characterize m_t , s_t , and Y_t in simple terms and in such a way that the model [1] provides a good representation of the data, then we can proceed to use the model to make forecasts or to address other questions related to the series. Completing the specification of the model by estimating the trend and seasonal components and characterizing the random component is a major part of time series analysis. The model [1] is sometimes referred to as an additive decomposition model. Provided the observations are all positive, the multiplicative model,

$$X_t = m_t s_t Y_t, \quad [2]$$

can be reduced to an additive model by taking logarithms of each side to get an additive model for the logarithms of the data.

The general form of the additive model [1] supposes that the seasonal component s_t has known period d (12 for monthly data, 4 for quarterly data, etc.) and satisfies the conditions

$$s_{t+d} = s_t \text{ and } \sum_{t=1}^d s_t = 0, \quad [3]$$

while $\{Y_t\}$ is a stationary time series, that is, a sequence of random variables satisfying the conditions,

$$E(Y_t) = \mu, E(Y_t^2) < \infty \text{ and } \text{Cov}(Y_{t+b}, Y_t) = \gamma(b) \text{ for all } t, \quad [4]$$

with the additional property that $\mu = 0$. The function γ is called the autocovariance function of the sequence $\{Y_t\}$ and the value $\gamma(b)$ is the autocovariance at lag b . In the special case when the random variables Y_t are independent and identically distributed, the model [1] is a classical regression model and $\gamma(b) = 0$ for all $b \neq 0$. However, in time series analysis, it is the dependence between Y_{t+b} and Y_t which is of special interest and which allows the possibility of using past observations to obtain forecasts of future values which have a smaller expected squared error than would be obtained if we used the naive forecast, μ . A simple and useful measure of the dependence between X_{t+b} and X_t is provided by the autocorrelation function,

$$\rho(b) = \frac{\gamma(b)}{\gamma(0)}. \quad [5]$$

It measures dependence in the sense that the expected squared difference between Y_{t+b} and Y_t is maximum when $\rho(b) = -1$ and decreases to zero as $\rho(b)$ increases to 1 according to the relation, $E(Y_{t+b} - Y_t)^2 = 2\gamma(0)(1 - \rho(b))$.

From observed values y_1, \dots, y_n of a stationary sequence of random variables $\{Y_t\}$, good estimators of the mean $\mu = E(Y_t)$ and the autocovariance function $\gamma(b)$ are the sample mean and sample autocovariance function,

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i \quad [6]$$

and

$$\hat{\gamma}(b) = \frac{1}{n} \sum_{i=1}^{n-|b|} (y_{i+|b|} - \hat{\mu})(y_i - \hat{\mu}), -n < b < n, \quad [7]$$

respectively. The autocorrelation function of $\{Y_t\}$ is estimated by the sample autocorrelation function,

$$\hat{\rho}(b) = \frac{\hat{\gamma}(b)}{\hat{\gamma}(0)}. \quad [8]$$

Elementary techniques for estimating the trend and seasonal components, m_t and s_t , can be found in many texts on time series analysis (e.g., Brockwell and Davis, 2002).

More sophisticated techniques are employed in the packages X-11 and the updated version X-12 described in Findley *et al.* (1998), and used by the US Census Bureau. Once estimators \hat{m}_t of m_t and \hat{s}_t of s_t have been obtained, they can be subtracted from the observations to yield the residuals,

$$y_t = x_t - \hat{m}_t - \hat{s}_t. \quad [9]$$

A stationary time series model can then be fitted to the residual series to complete the specification of the model. The model is usually chosen from the class of autoregressive moving average (or ARMA) processes defined in the section titled ARMA and ARIMA processes.

Instead of estimating and subtracting off the trend and seasonal components to generate a sequence of residuals, an alternative approach, developed by Box *et al.* (1994), is to apply difference operators to the original series to remove trend and seasonality. The backward shift operator, B is an operator which when applied to X_t gives X_{t-1} . Thus

$$BX_t = X_{t-1}, B^j X_t = X_{t-j}, j = 2, 3, \dots$$

The lag-1 difference operator is the operator $\nabla = (1 - B)$. Thus

$$\nabla X_t = (1 - B)X_t = X_t - X_{t-1}. \quad [10]$$

When applied to a polynomial trend of degree d , the operator ∇ reduces it to a polynomial of degree $d - 1$. The operator ∇^d , denoting d successive applications of ∇ , therefore reduces any polynomial trend of degree d to a constant. Usually a small number of applications of ∇ is sufficient to eliminate trends encountered in practice. Application of the lag- r difference operator, $\nabla_r = (1 - B^r)$ (not to be confused with ∇^r), to X_t gives

$$\nabla_r X_t = (1 - B^r)X_t = X_t - X_{t-r}, \quad [11]$$

eliminating any seasonal component with period r . In the Box-Jenkins approach to time series modeling, the operators ∇ and ∇_r are applied as many times as is necessary to eliminate trend and seasonality and the sample mean of the differenced data subtracted to generate a sequence of residuals y_t which are then modeled with a suitably chosen ARMA process in the same way as the residuals [9].

Figure 3 shows the effect of applying the operator ∇ to the enrolment series of **Figure 1**. There is no longer any apparent trend and the differenced series is a good candidate for representation by a stationary time series model.

In the case of the New York birth-rate series, our first goal is to remove the seasonal component with period 12. This is achieved by application of the lag-12 difference operator $(1 - B^{12})$. **Figure 4** shows the resulting series which, like the series in **Figure 3**, is a good candidate for representation by a stationary time series model. Notice that the operator $(1 - B^{12})$ not only removed the seasonality but also the upward trend in the original series.

Had this not been the case, then further applications of the operator $(1 - B)$ would have been necessary to remove any remaining trend.

In cases where the variability of the observed data appears to change with the level of the data, a preliminary transformation, prior to detrending and deseasonalizing, may be required to stabilize the variability. For this purpose, a member of the family of Box–Cox transformations (see e.g., Brockwell and Davis, 2002) is frequently used.

ARMA and ARIMA Processes

For modeling the residuals $\{y_t\}$ (found as described above), a very useful parametric family of zero-mean stationary sequences is furnished by the autoregressive moving average (or ARMA) processes. The ARMA(p, q) process $\{Y_t\}$ with autoregressive coefficients, ϕ_1, \dots, ϕ_p , moving average coefficients, $\theta_1, \dots, \theta_q$ and white-noise variance σ^2 , is defined as a stationary solution of the difference equations,

$$(1 - \phi_1 B - \dots - \phi_p B^p) Y_t = (1 + \theta_1 + \dots + \theta_q B^q) Z_t, \quad t = 0, \pm 1, \pm 2, \dots \quad [12]$$

where B is the backward shift operator, the polynomials, $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ and $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$, have no common factors and $\{Z_t\}$ is a sequence of uncorrelated random variables with mean zero and variance σ^2 . Such a sequence $\{Z_t\}$ is said to be white noise with mean 0 and variance σ^2 , indicated more concisely by writing $\{Z_t\} \sim \text{WN}(0, \sigma^2)$.

The eqn [12] has a unique stationary solution if and only if the equation $\phi(z) = 0$ has no real or complex root with $|z| = 1$; however, the possible values of ϕ_1, \dots, ϕ_p are usually assumed to satisfy the stronger restriction,

$$\phi(z) \neq 0 \text{ for all complex } z \text{ such that } |z| \leq 1. \quad [13]$$

The unique stationary solution of eqn [12] is then

$$Y_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad [14]$$

where ψ_j is the coefficient of z^j in the power-series expansion,

$$\frac{\theta(z)}{\phi(z)} = \sum_{j=0}^{\infty} \psi_j z^j, \quad |z| \leq 1.$$

Since Y_t in eqn [14] is a function only of Z_s , $s \leq t$, the series $\{Y_t\}$ is said to be a causal function of $\{Z_t\}$ and the condition [13] is called the causality condition for the eqn [12]. (Condition [13] is also frequently referred to as a stability condition.) The autocovariance function of the ARMA process [14] is

$$\gamma(b) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|b|}.$$

Under the condition [13], simple recursions are available for the numerical calculation of the sequence $\{\psi_j\}$ and of the autocovariance function $\gamma(b)$. Details can be

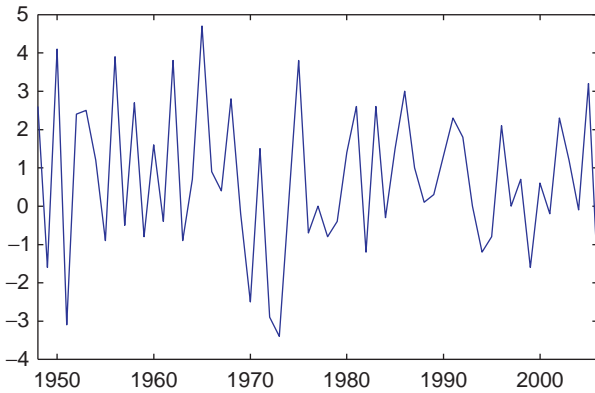


Figure 3 The differenced series $\{\nabla x_t\}$ derived from the annual percentage enrolment series shown in **Figure 1**.

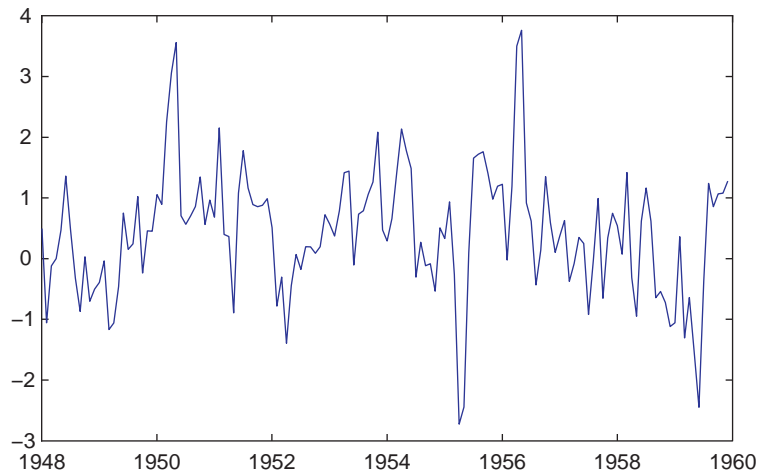


Figure 4 The differenced series $\{(1 - B^{12})x_t\}$ derived from the monthly New York birth-rate series shown in **Figure 2**.

found in Brockwell and Davis (2002). A key feature of ARMA processes for modeling dependence in a sequence of observations is the extraordinarily large range of autocorrelation functions exhibited by ARMA processes of different orders (p, q) as the coefficients ϕ_1, \dots, ϕ_p , and $\theta_1, \dots, \theta_q$, are varied.

Example 3

The causal ARMA(1,0) or AR(1) process is a stationary solution of the equations

$$Y_t = \phi Y_{t-1} + Z_t, \quad |\phi| < 1, \{Z_t\} \sim \text{WN}(0, \sigma^2).$$

The coefficients ψ_j for this process are $\psi_j = \phi^j$ and the autocovariance function of $\{Y_t\}$ is $\gamma(b) = \sigma^2 \phi^{|b|} / (1 - \phi^2)$.

ARMA processes as defined above are mathematical models for zero-mean stationary time series. However, many of the time series encountered in practice, like examples 1 and 2, cannot be well represented by a stationary model (i.e., one which satisfies the conditions [4]) since such a model does not account for trend and seasonality. However we saw in the section titled ‘Time Series models’ how the operators $(1 - B)$ and $(1 - B^r)$ can be used to eliminate trend and seasonality respectively. If a sequence of zero-mean residuals $\{y_t\}$ is generated by applying the operator $(1 - B)^d$ to the original series $\{x_t\}$ and we find that a model of the form [12] is appropriate for $\{y_t\}$, then we are effectively saying that the original series can be represented by the model,

$$\phi(B)(1 - B)^d X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2). \quad [15]$$

These equations are said to represent the original nonstationary series $\{x_t\}$ as an ARIMA(p, d, q) process. More generally, if the original series contains a trend and a seasonal component of period r , and if a sequence of zero-mean residuals $\{y_t\}$, obtained by applying the operators $(1 - B^r)$ and $(1 - B)^d$ to $\{x_t\}$, is found to be well-represented by equations of the form (12), then $\{x_t\}$ is well represented by the model,

$$\phi(B)(1 - B)^d(1 - B^r)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2) \quad [16]$$

and $\{X_t\}$ is said to be a seasonal ARIMA process. Notice that the fitting of ARIMA processes and seasonal ARIMA processes reduces, after the required differencing operations have been applied to the original data, to the fitting of ARMA processes to the residuals $\{y_t\}$. We consider this problem in the next section.

Model Selection and Estimation

Suppose for the moment that we know the orders p and q of the ARMA process [12] which is to be fitted to the residuals

y_1, \dots, y_n and suppose that $\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)$, is the vector of parameters to be estimated. The Gaussian likelihood $L(\beta; y_1, \dots, y_n)$, is the likelihood computed under the assumption that the joint distribution from which y_1, \dots, y_n are drawn is multivariate normal. Thus

$$L(\beta; y_1, \dots, y_n) = (2\pi)^{-n/2} (\det \Gamma_n)^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}_n' \Gamma_n^{-1} \mathbf{y}_n\right) \quad [17]$$

where $\mathbf{y}_n = (y_1, \dots, y_n)'$, Γ_n is the matrix of autocovariances $[\gamma(i - j)]_{i,j=1}^n$ and $\gamma(b)$ is the autocovariance function of the model defined by [12]. Although direct calculation of L is a daunting task, L can be calculated recursively using a transformation of variables as described in Brockwell and Davis (2002). For given values of y_1, \dots, y_n , the likelihood L is then maximized numerically with respect to β , in order to obtain the maximum likelihood estimate $\hat{\beta}$ of the ARMA parameters.

At first glance, maximization of Gaussian likelihood when the observations appear to be non-Gaussian may seem strange. However if the noise sequence $\{Z_t\}$ in the model [12] is any independent identically distributed sequence (with finite variance), the large-sample joint distribution of the estimators (assuming the true orders are p and q) is the same as in the Gaussian case (see Hannan, 1973; Brockwell and Davis, 1991). This large-sample distribution has a relatively simple Gaussian form which can be used to specify large-sample confidence intervals for the parameters (under the assumption that the observations are generated by the fitted model).

The previous discussion assumes that the orders p and q are known. However, this is rarely if ever the case and they must be chosen on the basis of the observations. The choice of p and q is referred to as the problem of order selection. A systematic approach to the problem was suggested by Akaike (1973) when he introduced the Akaike Information Criterion (AIC). He proposed that p and q be chosen by minimizing $\text{AIC}(\hat{\beta}(p, q))$, where $\hat{\beta}(p, q)$ is the maximum likelihood estimator of β for fixed p and q and

$$\text{AIC}(\beta) = -2 \ln(L(\beta)) + 2(p + q + 1). \quad [18]$$

The term $2(p + q + 1)$ can be regarded as a penalty factor which prevents the selection of excessive values for p and q and the accumulation of additional parameter estimation errors. It has been shown in Shibata (1980) that although the AIC criterion does not give consistent estimation of p and q , it is optimal in a certain sense with respect to prediction of future values of the series. A refined small-sample version of AIC, known as AICC, has been developed in Hurvich and Tsai (1989) and a comprehensive account of model selection can be found in Burnham and Anderson (2002).

Having arrived at a potential ARMA model for Y_1, \dots, Y_n , the model should be checked for goodness of fit.

On the basis of the fitted model, the minimum mean squared error linear predictors \hat{Y}_t of each Y_t in terms of Y_s , $s < t$, and the corresponding mean squared errors s_t can be computed (see section titled 'Prediction'). In practice they are computed in the course of evaluating the Gaussian likelihood. If the fitted model is valid, the properties of the rescaled one-step prediction errors $(Y_t - \hat{Y}_t)/\sqrt{s_t}$ should be similar to those of the sequence Z_t/σ in the model (12) and can therefore be used to check the assumed white-noise properties of $\{Z_t\}$ and whether or not the assumption of independence and/or normality is justified. A number of such tests are available (see e.g., Brockwell and Davis, 2002).

Prediction

If $\{Y_t\}$ is a stationary process with mean, $E(Y_t) = \mu$, and autocovariance function, $\text{Cov}(Y_{t+b}, Y_t) = \gamma(b)$, it is well-known that the function $f(Y_1, \dots, Y_n)$ of the observations which is the best predictor of Y_{n+b} , $b > 0$, in the sense of minimizing the expected squared error, $E(Y_{n+b} - f(Y_1, \dots, Y_n))^2$, is the conditional expectation $E(Y_{n+b} | Y_1, \dots, Y_n)$. However this depends in a complicated way on the joint distributions of the random variables Y_t which are virtually impossible to estimate on the basis of a single series of observations y_1, \dots, y_n . However, if the sequence $\{Y_t\}$ is Gaussian, that is, if the joint distribution of (Y_1, \dots, Y_n) is multivariate normal for every n , then the best predictor of Y_{n+b} in terms of Y_1, \dots, Y_n is a linear function and can be calculated as described below.

The best *linear* predictor of Y_{n+b} in terms of $\{1, Y_1, \dots, Y_n\}$, that is, the linear combination $P_n Y_{n+b} = a_0 + a_1 Y_1 + \dots + a_n Y_n$ which minimizes the mean squared error, $E[(Y_{n+b} - a_0 - \dots - a_n Y_n)^2]$, is given by

$$P_n Y_{n+b} = \mu + \sum_{i=1}^n a_i (Y_{n+1-i} - \mu), \quad [19]$$

where the vector of coefficients $\mathbf{a} = (a_1, \dots, a_n)'$ satisfies the linear equation,

$$\Gamma_n \mathbf{a} = \boldsymbol{\gamma}_n(b), \quad [20]$$

with $\boldsymbol{\gamma}_n(b) = (\gamma(b), \gamma(b+1), \dots, \gamma(b+n-1))'$ and $\Gamma_n = [\gamma(i-j)]_{i,j=1}^n$. The mean squared error of the best linear predictor is

$$E(Y_{n+b} - P_n Y_{n+b})^2 = \gamma(0) - \mathbf{a}' \boldsymbol{\gamma}_n(b). \quad [21]$$

If an ARMA model is fitted to the data, the special linear structure of the ARMA process arising from the defining equations can be used to greatly simplify the calculation of the best linear b -step predictor $P_n Y_{n+b}$ and its mean squared error, $\sigma_n^2(b)$. If the fitted model is Gaussian, we can also compute 95% prediction bounds, $P_n Y_{n+b} \pm 1.96\sigma_n(b)$ (see e.g., Brockwell and Davis, 2002).

Example 4

In order to predict future values of the causal AR[1] process defined in example 3, we can make use of the fact that linear prediction is a linear operation and that $P_n Z_t = 0$ for $t > n$ to deduce that

$$P_n Y_{n+b} = \phi P_n Y_{n+b-1} = \phi^2 P_n Y_{n+b-2} = \dots = \phi^b Y_n, b \geq 1.$$

with mean squared error,

$$E(Y_{n+b} - P_n Y_{n+b})^2 = \sigma^2 \frac{1 - \phi^b}{1 - \phi^2}.$$

In order to obtain forecasts and prediction bounds for the original series which were transformed to generate the residuals, we simply apply the inverse transformations to the forecasts and prediction bounds for the residuals.

Example 5

Using the model selection procedure described in the earlier section, the minimum AICC model for the first 49 residuals shown in **Figure 3** (with the sample mean subtracted) is found to be the AR[1] process,

$$Y_t = -0.2519 Y_{t-1} + Z_t, \{Z_t\} \sim \text{WN}(0, 3.530).$$

Using this model and the results of example 4 we obtain the forecasts and prediction bounds for the last ten values (1997–2006) of the original series shown, with the actual observed values, in **Figure 5**.

The Frequency Viewpoint

The methods described so far are referred to as time-domain methods since they focus on the evolution in time of the sequence of random variables X_1, X_2, \dots representing the observed data. If however, we regard the sequence

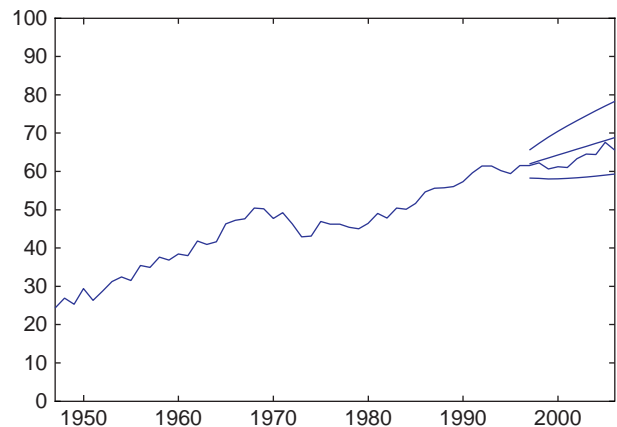


Figure 5 Forecasts and 95% prediction bounds for the last 10 values of the series in **Figure 1** based on the model fitted to the first 50 values.

as a random function defined on the integers, then an alternative approach is to consider the decomposition of that function into sinusoidal components. The observations x_1, \dots, x_n have a Fourier representation

$$x_t = \sum_j a_j e^{i\omega_j t}, t = 1, \dots, n.$$

where $i = \sqrt{-1}$, $\omega_j = 2\pi j/n$ and the sum is over all integers j such that $-\pi < \omega_j \leq \pi$. Analogously, if X_1, X_2, \dots is a stationary time series, then it has a so-called spectral representation,

$$X_t = \int_{-\pi}^{\pi} e^{i\omega t} dZ(t), \quad [22]$$

where $\{Z(t), -\pi \leq t \leq \pi\}$ is a process with uncorrelated increments. A detailed discussion of this approach can be found, for example, in Bloomfield (1976), Brockwell and Davis (1991) and Priestley (1981), but is outside the scope of this article.

Multivariate Time Series

Many time series arising in practice are best analyzed as components of some vector-valued (multivariate) time series $\{X_t\} = (X_{t1}, \dots, X_{tm})'$ in which each of the component series $\{X_{tb}, t = 1, 2, 3, \dots\}$ is a univariate time series of the type already discussed. In multivariate time series modeling, the goal is to account for and take advantage of the dependence not only between observations of a single component at different times, but also between the different component series. For example, if X_{t1} is the daily percentage change in the closing value of the Dow-Jones Industrial Average in New York on trading day t , and if X_{t2} is the analog for the Australian All-Ordinaries Index, then $\{X_t\} = (X_{t1}, X_{t2})'$ is a bivariate time series in which there is very little evidence of autocorrelation in either of the two component series. However, there is strong evidence of correlation between X_{t1} and $X_{(t+1)2}$, indicating that the Dow-Jones percentage change on day t is of value in predicting the All-Ordinaries percentage change on day $t + 1$. Such dependencies in the multivariate case can be assessed by means of the covariance matrices,

$$\Gamma(t + b, t) = [\text{cov}(X_{(t+b),i}, X_{t,j})]_{i,j=1}^m, \quad [23]$$

where the number of components, m , is 2 in this particular example. Stationarity of the multivariate series $\{X_t\}$ is defined as in the univariate case to mean that all components have finite second moments and that the mean vectors $E(X_t)$ and covariance matrices $\Gamma(t + b, t)$ are independent of t . Multivariate white noise, $WN(0, R)$, is defined to be a sequence of zero-mean uncorrelated random vectors $\{Z_t\}$, each with mean vector 0 and covariance matrix R . Multivariate ARMA processes are defined in terms of multivariate white noise by equations analogous to those defining

univariate ARMA processes (for details see Brockwell and Davis, 1991).

State-Space Models

The study of state-space models has had a profound impact on time series analysis. A linear state-space model for a (possibly multivariate) time series $\{Y_t, t = 1, 2, \dots\}$ consists of two equations. The first, known as the observation equation, expresses the w -dimensional observation vector Y_t as a linear function of a v -dimensional state vector X_t plus noise. Thus

$$Y_t = G_t X_t + W_t, t = 1, 2, \dots, \quad [24]$$

where W_t has mean 0 and covariance matrix R_t , the sequence $\{W_t\}$ is uncorrelated and $\{G_t\}$ is a sequence of $w \times v$ matrices. The second equation, called the state equation, determines the state X_{t+1} at time $t + 1$ in terms of the previous state X_t and a noise term. The state equation is

$$X_{t+1} = F_t X_t + V_t, t = 1, 2, \dots, \quad [25]$$

where V_t has mean 0 and covariance matrix Q_t , the sequence $\{V_t\}$ is uncorrelated, $\{F_t\}$ is a sequence of $v \times v$ matrices and $\{V_t\}$ is uncorrelated with $\{W_t\}$ (i.e., $E(W_t V_s') = 0$ for all s and t). To complete the specification, it is assumed that the initial state X_1 is uncorrelated with all of the noise terms $\{V_t\}$ and $\{W_t\}$.

An extremely rich class of models for time series, including and going well beyond the ARIMA models described earlier, can be formulated within this framework (see Hannan and Deistler, 1988). In econometrics, the structural time series models developed in Harvey (1990), in which trend and seasonal components are allowed to evolve randomly, also fall into this framework. The power of the state-space formulation of time series models depends heavily on the Kalman recursions which allow best linear predictors and best linear estimates of various model-related variables to be computed in a routine way. Time series with missing values are also readily handled in the state-space framework.

More general state-space models, in which the linear relationships [24] and [25] are replaced by the specification of conditional distributions, are also widely used to generate an even broader class of models, including models for time series of counts such as the numbers of reported new cases of a particular disease (see e.g., Chan and Ledolter, 1995).

Additional Topics

Although linear models for time series data have found broad applications in many areas of the physical, biological, and behavioral sciences, there are areas where they have been

found inadequate. Consequently, a great deal of attention has been devoted to the development of nonlinear models for such applications. These include threshold models, Tong, 1990, bilinear models, Subba-Rao and Gabr, 1984, random coefficient autoregressive models, Nicholls and Quinn, 1982, Markov switching models, Hamilton, 1994 and many others. For financial time series, the ARCH (autoregressive conditionally heteroscedastic) model of Engle (1982) and its generalized version, the GARCH model of Bollerslev (1986) have been particularly successful. For a recent account of time series models specifically related to financial applications (see Tsay, 2001).

Acknowledgment

This work was partially supported by NSF Grant DMS 0744058.

See also: Model Selection; Multivariate Longitudinal Data Analysis; Signal Detection Theory; Stochastic Processes; Structural Equation Models.

Bibliography

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, Petrov, B. N. and Saki, F. (eds.), pp 267–281. Budapest: Akademia Kiado.
- Bloomfield, P. (1976). *Fourier Analysis of Time Series: An Introduction*. New York: Wiley.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*, 3rd edn. Englewood Cliffs, NJ: Prentice-Hall.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd edn. New York: Springer.
- Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting*, 2nd edn. New York: Springer.
- Burnham, K. P. and Anderson, D. (2002). *Model Selection and Multi-Model Inference*, 2nd edn. New York: Springer.
- Chan, K. S. and Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association* **90**, 242–252.
- Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica* **50**, 987–1087.
- Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., and Chen, B. -C. (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program (with discussion and reply). *Journal of Business and Economic Statistics* **16**, 127–177.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.
- Hannan, E. J. (1973). The asymptotic theory of linear time series models. *Journal of Applied Probability* **10**, 130–145.
- Hannan, E. J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. New York: Wiley.
- Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Hurvich, C. M. and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297–307.
- Nicholls, D. F. and Quinn, B. G. (1982). *Random coefficient autoregressive models: An introduction*. Springer Lecture Notes in Statistics **11**.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series*, vols. I and II. New York: Academic Press.
- Shibata, R. (1980). Asymptotically efficient selection of the order for estimating parameters of a linear process. *Annals of Statistics* **8**, 147–164.
- Subba-Rao, T. and Gabr, M. M. (1984). *An introduction to bispectral analysis and bilinear time series models*. Springer Lecture Notes in Statistics **24**.
- Tong, H. (1990). *Non-linear Time Series: A Dynamical Systems Approach*. Oxford: Oxford University Press.
- Tsay (2001). *Analysis of Financial Time Series*. New York: Wiley.

Univariate Linear Regression

M L Hazelton, Massey University, Palmerston North, New Zealand

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Akaike information criterion – A measure of the overall quality of a model which takes account of both the complexity of the model and how well it fits the data.

Heteroscedasticity – Refers to the presence of different variances between two or more random variables. The complementary concept is homoscedasticity, which is usually assumed to apply to the error terms in a linear regression model.

Method of least squares – A well-founded technique for estimating the unknown parameters in a linear regression model, based on minimization of the sum of squared differences between observed and modeled responses.

Ordinary least squares – The method of least squares is also sometimes referred to as ordinary least squares to distinguish it from related techniques such as weighted least squares.

Raw residual – The value of an observed response minus the corresponding predicted response based on the regression model.

Residual analysis – Refers to the examination of residuals from a regression model in order to assess the validity of the assumptions underlying the model.

Standardized residual – A scaled version of a raw residual.

The problem of modeling the behavior of one random response variable in terms of one or more other explanatory variables is one of the most important aspects of modeling in statistics. When the variables are quantitative (as we shall assume for the most part in this article), regression modeling can be used. Regression models are wide ranging and have been used extensively in the education literature (see, e.g., Hsu, 2005).

In principle, we could attempt to describe multiple features of the distribution of the response in terms of the explanatory variables. However, in many applications it is sufficient to assume that only the mean (or expected) value of the response varies with the predictors, and that it does so in a linear fashion. Under these assumptions, we obtain the classical (multiple) linear regression model. One of the reasons for the widespread use of linear regression is that such models have a highly tractable

mathematical structure (e.g., Jørgensen, 1993). As a result, it has been possible to develop an extensive body of methods and theory for linear regression. This in turn provides the statistical practitioner with a comprehensive and well-understood set of tools for fitting, examining, comparing, and interpreting regression models.

In this article, a mathematical formulation of the linear regression model is provided. Model fitting by the method of least squares is described, and the application of regression models illustrated through an example. The underlying assumptions for regression modeling are discussed, and methods for examining their validity are examined. This article also covers the interpretation of regression models; selection of explanatory variables and comparison of models; and connections with other methods including analysis of variance (ANOVA) and hierarchical linear models.

Formulation of the Linear Regression Model

Suppose we observe a response variable Y and p explanatory (or predictor or regressor) variables x_1, \dots, x_p on n individuals or entities. In classifying one of them as the response, we are not treating the variables in a symmetric manner. Rather, we are focusing on the (stochastic) behavior of the response in terms of the other variables. Therefore, we will consider the explanatory variables to be fixed for the purposes of the regression model. In some applications, the explanatory variables will be fixed by an experimenter; for example, the length of time allocated to a student to complete some task for which we wish to measure a response. In other cases, the predictor variables will have arisen as the result of some random process, but the regression model will represent the behavior of the response conditional on the values of the predictor variables.

As a concrete example (to which we shall refer repeatedly), Guber (1999) describes data from 1994–95 for each of the $n = 50$ US states including the following variables: mean total SAT score (SAT); mean expenditure per pupil in thousands of dollars (EXP); mean pupil/teacher ratio (PTR); estimated mean annual salary of teachers in thousands of dollars (SAL); and percentage of all eligible students taking the SAT (PER). It is quite natural to think of modeling SAT as the response variable so as to understand its relationship with the other variables. While these

explanatory variables were not fixed by an experimenter, it is equally natural to think of modeling SAT conditional on their values so that one may use the model to predict SAT for any given set of values EXP, PTR, PER, and SAL.

The linear regression model is defined by the following equation:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \quad [1]$$

where Y_i is the response for the i th individual (e.g., SAT for the i th state); x_{ij} is the value of the j th explanatory variable for the i th individual (e.g., x_{i1} is the value of EXP in the i th state); and ε_i is the error term for that individual. The regression coefficients $\beta_0, \beta_1, \dots, \beta_p$ are model parameters whose true values are unknown and hence must be estimated from the data. When there is just one explanatory variable, the (multiple) linear regression model in eqn [1] reduces to the simple linear regression model defined by

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad [2]$$

where x_i now denotes the value of the single explanatory variable for the i th individual.

It is usual to make the following four assumptions about a linear regression model, and more specifically, about the error terms in the model:

(A1) The mean error is zero, that is, $E[\varepsilon_i] = 0$. This assumption is equivalent to writing that

$$E[Y_i] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \quad [3]$$

which is a statement that we have correctly specified the form of the mean value of the response variable.

(A2) The error terms $\varepsilon_1, \dots, \varepsilon_n$ are statistically independent.

(A3) $\text{Var}(\varepsilon_i) = \sigma^2$ is constant for all observations, $i = 1, \dots, n$.

(A4) The error terms $\varepsilon_1, \dots, \varepsilon_n$ are normally distributed.

We examine the consequences of failure of these assumptions, and methods for checking their validity, when we discuss model diagnostics.

The linear regression model can be expressed compactly and conveniently using matrix notation. Model [1] becomes simply

$$y = X\beta + \varepsilon \quad [4]$$

where $y = (Y_1, \dots, Y_n)^T$ is the vector of responses (with superscript T denoting matrix transposition); $\beta = (\beta_0, \dots, \beta_p)^T$ is the vector of regression parameters; and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$ the vector of random error terms. In [4]

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \quad [5]$$

is the design matrix for the regression. The rows of X correspond to given individuals while the columns correspond to given variables (with the initial column of ones corresponding to the intercept term β_0).

Fitting Regression Models

Application of a linear regression model to any given data set requires that the model parameters (including the regression coefficients $\beta_0, \beta_1, \dots, \beta_p$ and the error variance, σ^2) be estimated. The regression coefficients can be estimated using the method of least squares (sometimes also referred to as ordinary least squares (OLS) to distinguish it from related techniques such as weighted least squares). If the observed responses are y_1, \dots, y_n , then the least-squares estimates $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ of the regression coefficients are derived by minimizing the sum of squared discrepancies

$$SS(\beta_0, \dots, \beta_p) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_p x_{ip})^2. \quad [6]$$

Using matrix notation, the vector of least-square estimates can be expressed explicitly as

$$\hat{\beta} = (X^T X)^{-1} X^T y. \quad [7]$$

In practice, least-squares estimates are almost invariably calculated using statistical software packages. These use stable numerical algorithms (based on matrix decomposition techniques) to compute $\hat{\beta}$ rather than use a direct implementation of eqn [7].

The least-squares method is not only an intuitively plausible approach to computing estimates of the regression parameters, but is also well supported from a theoretical perspective. It can be shown under assumptions (A1)–(A4) that the least-squares estimates are also maximum likelihood estimates, and hence are optimal in a number of important statistical senses. Even if assumption (A4) fails, least-squares estimation still produces the so-called best linear unbiased estimates (BLUEs), courtesy of Gauss–Markov theory (see, e.g., Plackett, 1950).

Having obtained the least-squares estimates of the regression parameters, we may compute the fitted values,

$$\hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip} \quad [8]$$

which estimate the mean responses. We may also calculate the residual sum of squares,

$$RSS = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_p x_{ip})^2 \quad [9]$$

which provides an overall measure of the discrepancy between the observed and fitted responses, and consequently forms the basis for estimating the error variance σ^2 .

Specifically, an unbiased estimator of the error variance is given by

$$s^2 = \frac{1}{n - p - 1} RSS. \quad [10]$$

Inference

The fitted regression model (i.e., the model with the regression parameters replaced by the least-squares estimates thereof) can be used for a number of purposes. These include estimation of the effect of each of the explanatory variables on the response; testing of hypotheses regarding the relationship between the response and one or more of the explanatory variables; and prediction of the response for a given set of explanatory variables.

To illustrate these ideas **Table 1** provides estimated regression coefficients (typical output from any standard statistics software package) from fitting the SAT data that were introduced earlier. The fitted model equation is hence

$$\widehat{SAT} = 1045.97 + 4.46EXP - 3.62PTR + 1.64SAL - 2.90PER. \quad [11]$$

As an example interpretation of the parameter estimates, the model indicates a 4.46 point increase in expected SAT for a unit increase in EXP (i.e., for a \$1000 increase in mean expenditure per pupil), but it is important to recognize that this assumes that all other variables are held constant (cf. Courville and Thompson, 2001).

The standard error associated with each variable is a measure of the precision with which the coefficient of that variable is estimated. Standard errors for regression coefficients are given by the square roots of the diagonal elements of the (estimated) variance-covariance matrix, $\widehat{Var}(\hat{\beta}) = s^2(X^T X)^{-1}$. The t value for the i th variable (which appears in standard computer output for a regression analysis, like **Table 1**) is the ratio of the variable's estimated coefficient to its standard error: $t = \hat{\beta}_i / SE(\hat{\beta}_i)$. This is an appropriate statistic for testing whether or not the true underlying parameter value is zero. A large t statistic provides evidence against the null hypothesis that the true value of the parameter is zero, and will produce

a small p -value (derived from the t -distribution with $n - p - 1$ degrees of freedom). For example, the p -value for EXP of $P = 0.674$ (from **Table 1**) provides no evidence to suggest that the coefficient of EXP is different from zero, and hence no statistically significant evidence that expenditure has an effect of SAT, having adjusted for the other variables in the model. Nonetheless, a simple linear regression of SAT on just EXP alone returns an estimated coefficient for EXP of $\hat{\beta}_1 = -20.9$ with corresponding p -value of $P = 0.006$, indicating that SAT score is related to expenditure but in a manner that is interrelated with the variables PTR, SAL, and PER.

Confidence intervals for regression parameters provide guidance on the size of the effect of an explanatory variable on the response. A $100(1 - \alpha)\%$ confidence interval for the coefficient of the i th variable is given by:

$$(\hat{\beta}_i - t_{\alpha/2} SE(\hat{\beta}_i), \hat{\beta}_i + t_{\alpha/2} SE(\hat{\beta}_i)), \quad [12]$$

where $t_{\alpha/2}$ is the appropriate critical point (i.e., the $(1 - \alpha/2)$ quantile) of the t -distribution on $n - p - 1$ degrees of freedom. For example, the 95% confidence interval for the coefficient of SAL is given by $(1.638 \pm 2.014 \times 2.387) = (-3.17, 6.45)$ using $t_{0.025} = 2.014$ for a t -distribution on 45 degrees of freedom. This interval includes zero, indicating that we cannot be sure (based on this analysis) that higher teacher salaries are associated with higher SAT scores (having adjusted for the other explanatory variables).

Point prediction from a regression model proceeds by substituting the requisite values of the explanatory variables into the fitted model equation. For example, if we wished to predict SAT results for $EXP = 9$, $PTR = 14$, $SAL = 50$, and $PER = 80$, then by substitution into eqn [11] we get $\widehat{SAT} = 885$ (to the nearest whole number) which is reasonably similar to the observed mean SAT score of 908 in Connecticut for which the values of the explanatory variables are similar to those chosen above. A prediction interval (which should be supplied as a matter of course) to accompany the point prediction \hat{y} can be computed by

$$(\hat{y} - t_{\alpha/2} PE(\hat{y}), \hat{y} + t_{\alpha/2} PE(\hat{y})) \quad [13]$$

where $PE(\hat{Y})$ is the prediction error. For the point prediction above the corresponding 95% prediction error is $PE(\widehat{SAT}) = 36.1$ (and $t_{0.025} = 2.014$ as earlier) so that the prediction interval is (812, 958).

Regardless of the other purposes to which a regression model is developed, it is almost always of interest to characterize the extent to which the explanatory variables can describe the behavior of the response variable. This can be quantified using the coefficient of determination, R^2 , which is equal to the square of the correlation coefficient between the observed and fitted values. It can be interpreted as the proportion of the variation in the

Table 1 Table of regression coefficient for the SAT data

	<i>Estimate</i>	<i>Std. error</i>	<i>t value</i>	<i>Pr(> t)</i>
(Intercept)	1045.97	52.87	19.78	0.000
EXP	4.46	10.55	0.42	0.674
PTR	-3.62	3.21	-1.13	0.266
SAL	1.64	2.39	0.69	0.496
PER	-2.90	0.23	-12.56	0.000

response which is attributable to the explanatory variables. For instance, for the SAT data $R^2 = 0.825$ indicating that over 80% of the variability in SAT scores can be explained by the variables EXP, PTR, SAL, and PER. It should be noted that all these comments on R^2 assume that the regression model in question includes an intercept term. For models without an intercept term (i.e., regression through the origin), R^2 does not have a particularly useful interpretation.

Model Diagnostics

The validity of any conclusions that we draw from a fitted regression model will be contingent on the correctness of the assumptions (A1–A4) listed earlier. Failure of A1 is particularly serious, since this indicates that the form of the regression function (as given in eqn [3]) is misspecified. In such circumstances, estimates of regression parameters will be biased and the model will be unreliable.

Failure of A2 occurs when the responses are dependent (e.g., because of clustering effects which may occur because of spatial proximity, or because the data include repeated observations on a given set of individuals). In this case, the least-squares parameter estimates will be unbiased (though somewhat inefficient) but standard errors computed in the usual way will be incorrect, leading to unreliable test statistics and confidence intervals. The same type of problems will occur if A3 is violated, that is, the errors are heteroscedastic. Failure of A4, that is errors which are not normally distributed, is of less concern. While in theory this invalidates t and F tests based on normality, versions of the central limit theorem and related asymptotic theory imply that the results of such standard tests will be quite reliable unless the sample size is very small.

A variety of diagnostic tools are available for assessing the validity of the model assumptions. These are typically based on the model residuals, and hence the application and interpretation of such diagnostics is often referred to as residual analysis. The raw residuals for the linear regression model are defined by:

$$e_i = y_i - \hat{\mu}_i \quad [14]$$

while the standardized residuals are given by

$$r_i = \frac{e_i}{s\sqrt{1 - b_{ii}}} \quad [15]$$

where b_{ii} is the i th diagonal element of the hat matrix, $H = X(X^T X)^{-1} X^T$. The raw residuals act as substitutes for the (unobserved) error terms, $\varepsilon_1, \dots, \varepsilon_n$, but these residuals have different variances as a result of the model fitting process. The standardized residuals are scaled to have unit variance and are therefore often preferred for diagnostic purposes.

One of the most commonly applied diagnostic tools is a plot of the residuals (either raw or standardized) against the fitted values. Such a plot can be used to spot a number of problems with the fitted model, including detection of:

- outlying observations, which will appear as points with extreme residuals;
- misspecification of the regression function (i.e., failure of A1), which may manifest itself through a discernible trend (e.g., curvature) in the plot; and
- heteroscedasticity in the error terms (i.e., failure of A3), which will typically lead to systematic variation (e.g., showing a funnel-shaped residual plot) in the vertical spread of the residuals from left to right in the plot.

A plot of the raw residuals against fitted values for the SAT data is given in **Figure 1**. It appears from this that West Virginia is an outlier, since the residual for this state is of a markedly larger size than the residuals for any other state. There is also a hint of curvature in the plot (with a downward trend at the left-hand side and an upward trend at the right-hand side of the plot) suggesting that the relationship between the SAT score and each of the explanatory variables may not be linear in every case. A plot of the residuals against each explanatory variable in turn can provide further insight. For instance, plots of the residuals for the SAT model against each of EXP, PER, PTR, and SAL (not shown) reveal clear curvature only for PER, indicating that the assumption of a linear relationship between SAT and this variable is questionable.

Assessment of the independence assumption (A2) can be difficult. A relatively common reason for failure of this assumption is the existence of serial correlation between observations with a clear time dimension (e.g., data collected on a monthly basis). In such cases, a plot of the residuals against the order in which the data were collected can sometimes be revealing although it may be more enlightening to apply standard time series methods to the residuals.

The standard tool for assessing normality of the error terms (A4) is a normal Q - Q (or normal probability) plot of the standardized residuals. This should appear as an

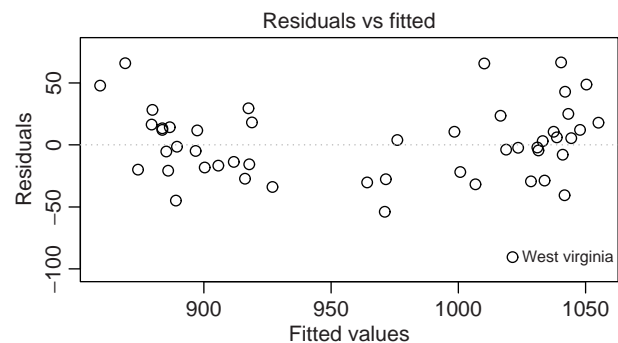


Figure 1 Plot of residuals vs. fitted values for SAT data.

approximately straight line if the assumption is reasonable. The Q - Q plot for the standardized residuals for the SAT data is given in **Figure 2**. Leaving aside the previously identified outlier (West Virginia), this plot does not suggest any major departure from normality in the distribution of the error terms.

As we have seen, the analysis of the SAT data may be complicated because of the existence of an apparent outlying observation – West Virginia. The extent to which this is likely to have a major impact on the model fit (and hence on inferences drawn from the model) depends on the influence that this data point exerts. At a general level, a data point with extreme values for one or more of its explanatory variables has the potential (depending on the value of the response) to have a more pronounced impact on the fit of the model than a data point for which the predictors sit near the center of the data. This potential can be measured by leverage, which is given by h_{ii} (as defined above) for the i th observation. A plot of the standardized residuals against leverage can therefore be illuminating. In addition, Cook's distance provides a measure of the influence for each data point which takes account of both the values of the predictors and the response variable. In essence, Cook's distance works by examining how the model fit would change were the data point under consideration excluded when estimating the regression parameters. Data points with a Cook's distance larger than 1 can be considered very influential and warrant close attention, while those with a Cook's distance between 0.5 and 1 might be described as moderately influential. A number of alternative measures of influence exist, with DFFITS (Belsley *et al.*, 1980) being arguably the most important.

To illustrate these methods, we plot the standardized residuals against leverage for the SAT data in **Figure 3**. It can be seen that West Virginia has rather low leverage and so has only a modest influence on the fitted model despite its extreme value for SAT (relative to its explanatory variables). This is reflected in the value of 0.11 for the Cook's distance for that state. The observation from

Utah appears of greater concern, but its Cook's distance of 0.47 (although the largest of all the observations) is not excessive.

When problems are found with a regression model, then it is necessary to take remedial action. When misspecification of the regression function is suspected, then one approach is to transform the response variable (e.g., by a logarithmic transformation), while another is to consider inclusion of polynomial terms in the explanatory variables. For example, for the SAT data it transpires that addition of the quadratic term in PER leads to a statistically significant improvement in the model fit ($P = 0.000$ when testing the coefficient of PER^2). An alternative is to employ nonparametric or semiparametric regression methods (e.g., Takezawa, 2006; Ruppert *et al.*, 2003). When there is evidence of heteroscedasticity, then the method of weighted least squares should be preferred to ordinary least squares, while failure of the independence assumption can be countered by explicit modeling of a more general covariance structure for the errors. These extensions to the standard linear regression model can both be described using the unifying theory of generalized least squares (e.g., Kariya and Kurata, 2004).

The handling of outliers can be a delicate matter. The first step is to check for transcription errors in the data and correct as necessary. One can consider discarding any remaining outliers (especially those with high influence) and then refit the model, but it is important to document that this has been done. Mindless and automated outlier deletion is to be avoided, since outliers are often of considerable interest. For example, outliers provided an early indication of the depletion of the ozone layer (Christie, 2004). An alternative to deletion is the use of robust statistical methods.

Model Comparison

When there are multiple explanatory variables available, then one can define a variety of alternative regression

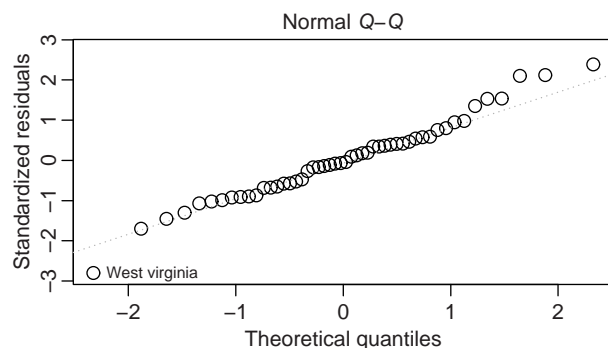


Figure 2 Normal Q - Q plot for standardized residuals for SAT data.

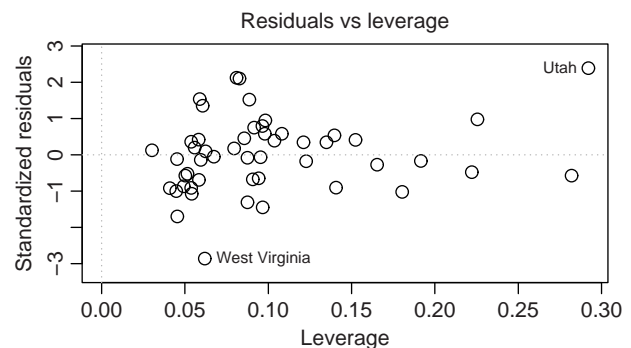


Figure 3 Plot of standardized residuals against leverage for SAT data.

models based on different subsets of these predictors. The question then arises as to which of these models should be preferred, and hence we have need for methods for comparing models. A classical approach to the comparison of two nested models (where a small model contains a subset of the variables in a large model) is to employ F tests based on ANOVA. This reduces to t -testing for a coefficient (as described above) when the models differ by only a single variable. For example, comparison of a large model containing all four explanatory variables (EXP, PTR, SAL, and PER) with a small model containing just the three variables EXP, SAL, and PER can be achieved using the t -statistic for PTR. The value $t = -1.13$ from **Table 1** gives a p -value 0.266, and hence provides no significant evidence against the null hypothesis that the small model is adequate.

An alternative approach is to define some measure of overall quality for each model, and then choose the model for which this is optimum. In quantifying model quality, one should keep in mind two competing goals in model selection. First, we have the principle of parsimony, which in essence states that simple models are preferable to more complex models (other things being equal). Second, we want the model to fit as closely to the data as possible. For linear regression models, complexity can be measured by the number of regression parameters or explanatory variables, while the closeness of the model fit can be measured by residual sum of squares. This motivates measures of quality, Q , of the following form:

$$Q = \log(RSS)/\lambda + \gamma p + c \quad [16]$$

where RSS is the model residual sum of squares, p the number of explanatory variables, and λ , γ are tuning constants and c is an arbitrary constant. When conducting such a comparison, a smaller value for Q indicates a better model.

A number of so-called information criteria can be written in the form of eqn [16] with suitable values for the tuning parameters. The Akaike Information Criterion (AIC) is perhaps best known of these and is obtained by setting $\lambda = 1/n$ and $\gamma = 2$, while the Bayes (or Schwarz) Information Criterion (BIC) is obtained when $\lambda = 1/n$ and $\gamma = \log(n)$ (see Akaike (1974) and Schwarz (1978)). A related quantity is Mallows C_p (Mallows, 1973) in which $\gamma = 2$ and the first term on the right-hand side of eqn [16] is replaced by $RSS/\hat{\sigma}^2$, where $\hat{\sigma}^2$ is an estimate of variance which is assumed fixed for all models under comparison.

As an illustration of the application of the AIC, **Table 2** lists the values of AIC for five models for the SAT data, based on different subsets of explanatory variables. The best model (i.e., with lowest AIC) of those listed is M2, which regresses SAT score on the three variables PTR, SAL, and PER. However, models M1 and M3 are only a few units worse on the AIC scale, so there is little to choose between these and M2.

Table 2 Values of AIC for five regression models for the SAT data

<i>Model</i>	<i>Explanatory variables in model</i>	<i>AIC</i>
M1	EXP + PTR + SAL + PER	353.5
M2	PTR + SAL + PER	351.7
M3	SAL + PER	354.6
M4	PTR + PER	356.2
M5	SAL + PTR	425.3

When a large number of predictor variables are available, a direct comparison of all possible models will be highly computer intensive. A computationally cheaper methodology is then desirable. One approach is to employ a stepwise variable selection algorithm, where we consider systematically adding and/or removing variables in an iterative manner in order to improve the model quality. More specifically, at each step of this type of algorithm we compare the current best model with all alternative models which can be obtained by removing single terms from this model, or adding in single variables which are currently not part of the model. These comparisons may be conducted with F tests, or using AIC scores.

To illustrate the the AIC approach, we conduct variable selection starting with M1 (from **Table 2**) as the initial model. At the first step, we compare M1 with each of the four models defined by deleting one of the variables in M1. It transpires that out of these models and M1 itself, the lowest AIC belongs to model M2 (obtained by removing EXP from M1). Hence M2 becomes the current best model. At the next step, we consider all models that can be obtained by adding or deleting a single variable from M2. These models are M1 (obtained by adding EXP) and M3, M4, and M5 (obtained by deleting the variables PTR, SAL, and PER respectively). All these models have higher AIC values than M2, indicating that this model cannot be improved by a single addition or deletion. The algorithm then terminates, returning M2 as the model of choice.

Summary

Linear regression models are among the most commonly used statistical methods. They combine widespread applicability with a highly developed theoretical basis. Functionality for fitting linear regression models exists in every significant statistics software package (including Minitab, SAS, SPSS, and R) and also in some spreadsheet packages (e.g., Excel).

This article has focused on regression modeling with quantitative explanatory variables. The analysis of models with categorical explanatory variables is covered in the article on ANOVA in this volume, although both regression and ANOVA type models can be unified with the

general framework of the linear model (e.g., Cohen, 1968). Regression models for clustered data and other hierarchical data structures are described in the article on hierarchical linear model in this volume. Regression models for highly non-normal response variables (e.g., binary responses) are covered in the article on generalized linear models in this volume.

See also: Analysis of Variance; Generalized Linear Models; Hierarchical Linear Models; Time Series Analysis.

Bibliography

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**(6), 716–723.
- Belsley, D. A., Edwin, K., and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.
- Christie, M. (2004). Data collection and the ozone hole: Too much of a good thing? *Proceedings of the International Commission on History of Meteorology* **1.1**, 99–105.
- Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin* **70**, 426–443.
- Courville, T. and Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement* **61**(2), 229–248.
- Guber, D. L. (1999). Getting what you pay for: The debate over equity in public school expenditures. *Journal of Statistics Education* **7**(2). <http://www.amstat.org/publications/jse/secure/v7n2/datasets.guber.cfm> (accessed May 2009).
- Hsu, T.-C. (2005). Research methods and data analysis procedures used by educational researchers. *International Journal of Research and Methods in Education* **28**(2), 109–133.
- Jørgensen, B. (1993). *The Theory of Linear Models*. London: Chapman and Hall.
- Kariya, T. and Kurata, H. (2004). *Generalized Least Squares*. Chichester: Wiley.
- Mallows, C. L. (1973). Some comments on Cp. *Technometrics* **15**, 661–675.
- Plackett, R. L. (1950). Some theorems in least squares. *Biometrika* **37**, 149–157.
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge, UK: Cambridge University Press.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**(2), 461–464.
- Takezawa, K. (2006). *Introduction to Nonparametric Regression*. Hoboken, NJ: Wiley-InterScience.

Further Reading

- Chatterjee, S., Hadi, A., and Price, B. (2000). *Regression Analysis by Example*. New York: Wiley.
- Chatterjee, S. and Yilmaz, M. (1992). A review of regression diagnostics for behavioral research. *Applied Psychological Measurement* **16**(3), 209–227.
- Chen, X., Ender, P., Mitchell, M., and Wells, C. (2003). *Regression with SPSS*. <http://www.ats.ucla.edu/stat/spss/webbooks/reg/default.htm> (accessed May 2009).
- Copas, J. B. (1983). Regression, prediction and shrinkage. *Journal of the Royal Statistical Society Series B* **45**, 311–354.
- George, E. I. (2000). The variable selection problem. *Journal of the American Statistical Association* **95**, 1304–1308.
- Kim, K. and Timm, N. (2006). *Univariate and Multivariate General Linear Models Theory and Applications with SAS*, 2nd edn. New York: CRC Press.
- Miller, A. (2002). *Subset Selection in Regression*, 2nd edn. London: CRC Press.
- Montgomery, D. C., Peck, E. A., and Vining, G. G. (2001). *Introduction to Linear Regression Analysis*, 3rd edn. Hoboken, NJ: Wiley-Interscience.
- Pedhazur, E. J. (1997). *Multiple Regression in Behavioral Research*, 3rd edn. Ft. Worth, TX: Harcourt Brace.
- Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. New York: Wiley.
- StatSoft Inc. (2007). General linear models (GLM). *Electronic Statistics Textbook*. <http://www.statsoft.com/textbook/stglm.html> (accessed May 2009).
- Weisberg, S. (2005). *Applied Linear Regression*, 3rd edn. Hoboken, NJ: Wiley-Interscience.
- Wikipedia Contributors (2005). Ordinary least squares. *Wikipedia, the Free Encyclopedia*. http://en.wikipedia.org/w/index.php?title=Ordinary_least_squares&oldid=17317645 (accessed May 2009).
- Wikipedia Contributors (2008). Linear regression. *Wikipedia, the Free Encyclopedia*. http://en.wikipedia.org/w/index.php?title=Linear_regression&oldid=216828588 (accessed May 2009).

Value-Added Models

D F McCaffrey, The RAND Corporation, Pittsburgh, PA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bias – It is the expected value of the difference between a statistical estimate and the quantity that is being estimated.

Fixed effects – Terms included as explicit explanatory variables in linear models in which model parameters are estimated via least squares or maximum likelihood methods. They are typically used to refer to models that include indicator variables for individual units such as schools, teachers, or students.

Longitudinal data – They are data from multiple units with repeated measurement on one or more units over time.

Precision – It is the reciprocal of the sampling variance of a statistical estimator.

Random effects – It is the term included in statistical models that are latent and assumed to be random variables from a common distribution.

General Background on Value-Added Modeling

The term value-added modeling (VAM) comes from the economics literature on the contribution of current inputs to outputs, after accounting for prior inputs. In the education literature, VAM refers to measuring the effects on student achievement of students' current teachers and schools, separate from other inputs. More broadly, the phrase often describes any analysis using longitudinal student test score data to study the effects of educational inputs on achievement.

The origins of VAM of teacher effects date back over 30 years; however, interest in these methods among researchers, policymakers, and educators grew precipitously following the publication of a technical report by William Sanders and June Rivers, in 1996, which found that teacher effects estimated using student test score gains predict student outcomes at least 2 years into the future, suggesting that teachers have persistent effects on their students' achievement and the accumulation of these effects could be substantial. The following year, Sanders and his colleagues published another paper claiming that teachers are the most important source of variance in

student achievement. Interest in VAM was further stoked by replication of the Sanders and Rivers results and several other papers finding that variability among teachers was large and that value-added estimates of teacher effects predict outcomes of teachers' future students. In the United States, the growing availability of student-achievement data due to increased emphasis on test-based accountability and the passage of the No Child Left Behind Act in 2001 led to explosive growth in empirical research on the effects of teachers on student outcomes.

Value-added models have two distinct uses: to assess programs, interventions, educator attributes, and educational practice; and to make inferences about the performance of individual teachers or schools for evaluation and accountability purposes. This article focuses primarily on issues related to accountability and evaluation of individual teachers, although the issues generally apply to inferences about schools and using the model to assess programs, interventions, and the like.

Causal Effects

The goal of value-added modeling is to understand how an educational input (e.g., a particular teacher) changes student achievement relative to other levels of the input (e.g., other teachers). Such an effect is commonly referred to as a causal effect. Statisticians generally define causal effects as differences between counterfactual or potential outcomes. For instance, a student has one potential outcome for the teacher she/he was assigned and a different one for every teacher she/he might be assigned. The difference between the student's potential outcome with his or her assigned teachers and the potential outcome for a different teacher is the causal effect of the assigned teacher relative to this specific alternative. Given that the goal of VAM is to compare the performance of many teachers, a relevant causal effect is the difference between a student's potential outcome for his or her assigned teacher and the average of his or her potential outcomes for all teachers. The average of the causal effects for individual students equals the teacher's overall causal effect. This average can be over all students in the population or over only students in or likely to be in a teacher's class.

To attribute this causal effect to the teacher and not to other factors, the only factor that can change with each potential outcome is the teacher; all other factors must remain the same, including the school and the other

students in the class. Teachers teach within the context of a school and a neighborhood and therefore two types of causal effects might be defined: the effect of a particular educational input (e.g., a teacher) independent of context or a causal effect of the input and the educational context that is likely to vary with that input (a teacher, the school, and the peers). The causal effect of an individual education input distinct from the context might be nearly impossible to estimate, whereas the combined effect of teacher and context might be more relevant to some stakeholders, and its estimation more feasible. The alternative value-added approaches described below make different attempts at separating the context from the teacher inputs. However, given the limitations of the available data, it is likely that all estimates will have some amount of interference from the changing context associated with different teacher assignments.

Advantages of the Value-Added Approach

The primary advantage of VAM compared with other measures used for accountability is its explicit accounting for competing inputs when estimating teacher or school effects. Other accountability metrics – such as comparing schools on average student scores or the percentage of students who surpass an achievement threshold (e.g., the percentage proficient) – typically make no distinctions between causal effects of schools and teachers and differences in other inputs that might exist between students in different schools or classrooms. Some alternatives to VAM use demographic and other broad descriptors of students rather than individual students' prior achievement in an attempt to account for differences in the populations served by different schools or teachers. However, these background characteristics often fail to control for many of the potential differences among schools and classrooms that can be controlled with prior achievement. Hence, compared with alternative approaches to accountability, value-added models place greater emphasis on measuring the inputs of schools and teachers and do so more effectively. However, VAM has its limitations as discussed in the remainder of this article.

Leading Modeling Approaches

The main approaches to VAM have been developed from three distinct perspectives: econometric modeling, statistical modeling, and *ad hoc* procedures. This section describes representative examples of the approaches and their essential features.

Econometric Models

The econometric modeling approach to VAM emerged from the cumulative achievement model. The model

starts with a general education production function for student achievement:

$$A_{it} = A_i[\mathbf{F}_i(t), \mathbf{E}_i(t), \mu_{i0}, \varepsilon_{it}] \quad [1]$$

where A_{it} , individual i 's achievement level at the end of the t th year of schooling, depends on the cumulative achievement function A_i and the entire input histories of family (including neighborhood), $\mathbf{F}_i(t)$, and school-based educational inputs, $\mathbf{E}_i(t)$, a composite variable representing time-invariant characteristics an individual is endowed with at birth (such as innate ability), μ_{i0} , and a normally distributed, mean-zero error, ε_{it} .

Economists make a series of assumptions to specify model [1] in a form that supports estimation of teacher effects using established estimators. In particular, they assume: the cumulative achievement function A_i does not depend on age and is additively separable (i.e., cumulative achievement equals the sum of the effects of the component inputs and the effect of any one input does not depend on the levels of the others); family inputs are time invariant and can be combined into a single term with student inputs; educational inputs can be separated into additive annual contributions of schools, teachers, peers, and other classroom inputs. Without further assumptions, estimation of the value added of the current teacher would require data on the school, teacher, peer, and classroom inputs for each student's current and entire prior schooling experience. Economists typically assume that all educational inputs decay geometrically at a rate λ because these data do not exist, and that the family inputs change at that same rate, yielding the following model:

$$A_{it} = \gamma_s S_{it} + \gamma_T T_{it} + \gamma_P P_{it} + \gamma_C Z_{it} + \varphi_i + \lambda A_{it-1} + \xi_{it}, \quad [2]$$

where T_{it} denotes teacher inputs and can be a single factor for a teacher effect; S_{it} , P_{it} , and Z_{it} denote school, peer, and other classroom inputs; φ_i is the fixed contribution of student and family; and the error term is $\xi_{it} = \varepsilon_{it} - \lambda \varepsilon_{it-1}$. A final assumption that $\lambda = 1$ leads to circumstances where a linear model for the first differences in achievement (gain scores) yields consistent estimators of certain effects, including those of current teachers. The parameters of the model are estimated using least squares with models that include indicators for students and teachers (i.e., student and teacher fixed effects) along with the other variables.

Numerous applications use variants of model [2] to study teachers. Some applications assume $\lambda = 1$; however, others have used $\lambda = 0$ which results in fixed-effects analysis on level scores rather than gain scores. In other applications, λ is estimated by including the lagged scores as a predictor in a linear regression model. Some of these applications treat φ_i as zero and some attempt to estimate both λ and the student coefficients while accounting for the complexities this creates for estimation.

If the students' potential outcomes for a teacher and the associated context are described by a model analogous to [2], then model [2] is a structural model and causal effects of teachers are defined by the contrasts of the coefficients in γ_T . Causal estimates can be obtained using linear regression techniques under the assumption that students assigned to any teacher's class differ from students assigned to other teachers' classes in terms of their general levels of family inputs (φ_i) and their prior achievement scores (A_{it-1}), but not in terms of unobserved factors related directly to their current level of achievement or to their achievement with their assigned teacher. That is, for students assigned to any given teacher, the distribution of the ζ_{it} must not differ from the distribution for other students, in particular it must have its mean as zero.

Statistical Models

Statistical models evolved from the hierarchical linear model (HLM) framework which specifies the sources of variance at each level of a hierarchical data structure linking repeated scores to students and students to classrooms, teachers, and schools. One example of this class of models is the layered model originally used by the Tennessee value-added assessment system. The HLM models are all variants of the following model, sometimes referred to as the variable persistence model, presented here for 3 years of scores on a single test:

$$\begin{aligned} A_{i1} &= \eta_1 + \gamma_{s1}S_{i1} + \gamma_{T1}T_{i1} + \gamma_{X1}X_{i1} + \zeta_{i1} \\ A_{i2} &= \eta_2 + \gamma_{s2}S_{i2} + \gamma_{T2}T_{i2} + \gamma_{X2}X_{i2} + \omega_{21}\gamma_{s1}S_{i1} + \\ &\quad \alpha_{21}\gamma_{T1}T_{i1} + \zeta_{i2} \\ A_{i3} &= \eta_3 + \gamma_{s3}S_{i3} + \gamma_{T3}T_{i3} + \gamma_{X3}X_{i3} + \omega_{32}\gamma_{s2}S_{i2} + \alpha_{32}\gamma_{T2}T_{i2} \\ &\quad + \omega_{31}\gamma_{s1}S_{i1} + \alpha_{31}\gamma_{T1}T_{i1} + \zeta_{i3} \end{aligned} \quad [3]$$

where η_t denotes overall means for the population each year, S_{it} is a vector of indicators identifying each school's share of the instruction for student i in school year t , and T_{it} is an analogous vector for teachers. The elements of γ_{st} and γ_{Tt} are normally distributed random variables representing school and teacher effects that are assumed to be independent across schools, teachers, and years. X_{it} denotes vectors of student variables. Finally, $\zeta_i = (\zeta_{i1}, \zeta_{i2}, \zeta_{i3})'$ are multivariate normal variables with mean zero and an unspecified covariance matrix Σ , and are independent of teacher, school, and other student effects. The α and ω parameters measure the fadeout of teacher and school effects and can be estimated from the data. Extensions to more years of testing or multiple tests are direct.

Restrictions to this model recover most of the other commonly used statistical VAM. For example, the layered model is a special case of model [3] with no school effects (i.e., all the elements of all the γ_{st} are set to zero), no explicit controls for students' characteristics (i.e., all the elements of all the γ_{Xt} are set to zero), and all the α

are set to one so that there is no fadeout of teacher effects. Cross-classified HLMs are a special case of model [3] where again the α (and ω) are all set to one and $\zeta_{it} = a_i + b_it$, with (a_i, b_i) assumed to be jointly normally distributed variables. Traditional HLMs with students nested within classrooms are special cases with α (and ω) set to zero and correlation among ζ_{it} treated as zero.

The model parameters, including the α (and ω) parameters, can be obtained via maximum likelihood or Bayesian methods. Bayesian methods have computational advantages and provide a principled method for making many desired inferences about teachers that account for uncertainty in all the model parameters.

There is little research directly linking model [3] to potential outcomes for students and causal modeling. Given that achievement in year $t > 1$ depends on teacher effects for prior years, the causal model will need to consider potential outcomes for the vector of teacher assignments, so that a student has a set of potential vectors of achievement scores for each set of teachers s/he might have in years 1–3. For the observed data to provide unbiased estimates of the structural model requires complex assumptions about the relationship of the vectors of assignments and a student's potential outcomes to hold true. Often analysts exclude school effects from model [3], so that the model provides no accounting for the school context and the teacher effects are a combination of the contributions of the teacher and the school context.

Ad hoc Models

In practice, the analysis of longitudinal data for evaluating teachers is sometimes less principled than the models above. For example, applications sometimes estimate teacher effects with simple averages of their students' gain scores or their residuals from a linear regression model predicting current achievement scores with a single prior score. Although these simple models might be special cases of the models listed above, they are applied in practice with the simple notion of measuring growth as a means of estimating value added. Research has shown that such simple approaches to VAM can provide poor estimates which confound student background variables with the estimated teacher effects.

In other applications, analysts create elaborate estimation procedures to create value-added measures of teacher effects that have face-validity to teachers and other stakeholders. The system devised for the Dallas Independent School District is an example of such a procedure. This method combines elements of statistical models into a multi-step procedure that was vetted by teachers and stakeholders as it was designed by school district analysts. Although this approach might be appropriate for developing an accountability system that is acceptable to educators, the statistical properties of the

resulting estimators are rarely evaluated and might be difficult to ascertain. If the estimated effects have undesirable statistical properties, these *ad hoc* methods might fail to achieve their goals. Therefore, methods that evaluate the properties of such *ad hoc* procedures are needed, as are tactics that emphasize the value of both face-validity and good statistical properties to educators and policymakers.

Concerns about Bias

Much of the research on value-added modeling has centered on how effectively the models produce estimates of individual teacher effects that accurately reflect their contributions to student learning rather than other factors. The three sources of bias that have consistently been raised as most problematic are: (1) nonrandom assignment of students to schools and teachers within schools; (2) improper assumptions about persistence/fadeout of different factors influencing student achievement; and (3) improper assumptions about test score scales.

Nonrandom Assignment

The fundamental problem of inferring teacher effects from observational data is that students are not randomly assigned to schools or teachers. The students taught by different teachers have systematically different characteristics, and unless properly controlled through statistical adjustment, these differences can introduce bias into estimated teacher effects. How well different modeling approaches correct for classroom selection is a hotly debated topic in VAM research.

The econometric approach accounts for nonrandom assignment through the assumptions that yield model [2] and the assumption that assignment depends only on observed data or time-invariant student and family inputs captured in φ_i . If these assumptions are correct, then potential bias due to differences among classes can be eliminated through the use of student fixed effects. Whether or not the assumptions hold in applications remains an issue of debate. Empirical results on achievement data from one state suggest that the assumptions were inappropriate for those data. However, alternative analyses comparing teacher effect estimates with historic data and the schools' regular assignment practices yielded teacher effects that were comparable to estimates for teachers estimated on randomly assigned classes, suggesting the VAM estimates were not biased. As the disparity in the two sets of empirical results suggests, the evidence on the validity of the assumptions of model [2] is equivocal and might depend on the specific application and the nature of the school and class assignments.

Many applications of the statistical models make no explicit allowances for nonrandom assignment of students.

Rather, the models do not include student covariates and assume that residual student error terms are independent among students and of teacher effects. Consequently, critics of this approach argue that all sources of variation in the data at the classroom level, including clustering of student achievement due to nonrandom assignment, are conflated with teacher effects. However, because the model allows correlation among the repeated measures from individual students, analytic results find this criticism is not necessarily correct: statistical modeling can effectively mitigate biases resulting from heterogeneity in average student achievement among classes, provided enough test scores are used in the modeling. With a fixed number of tests, the estimates cannot remove all the bias, and the models require at least five or more test scores per student to function most effectively at mitigating bias. Moreover, these theoretical assumptions do not apply if classroom assignments are dynamic in that they rely on annual data (e.g., test scores or classroom grades). In addition, as noted above, commonly used statistical models often do not include school factors, which limit inferences to the combined causal effects of teachers and their school contexts.

Some applications of statistical models have been criticized because they do not include student-level covariates. The results noted above suggest that under the right conditions and sufficiently many prior tests, covariates are unnecessary to control for heterogeneity among classrooms. In addition, because statistical models treat teacher effects as random and adjust for prior tests through correlation in the error term, controlling for covariates in these models when teacher effects are correlated with the covariates can overcorrect so that a portion of the true teacher effect is attributed to the covariates, biasing the estimated teacher effects. Methods exist for including student-level covariates without overcorrecting, but when applied to data they do not alter the estimated teacher effect relative to estimates from the layered model without student-level covariates; and the practice of excluding student-level covariates continues in some applications.

Both statistical and econometric models can be biased if there are classroom-level factors that affect achievement or if the student population is stratified such that there are groups of students and teachers in which students in some groups have no teachers in common with any of the students in other groups. Classroom-level variables cannot be included in models with fixed effects and are not accounted for by individual student test scores, so neither econometric nor statistical models could account for these variables. Similarly, fixed-effects estimates can be estimated only within strata and statistical models conflate difference in strata with differences in teachers. To date, there have been few studies on the existence and nature of strata in student-achievement data but in one study of four large countywide school systems in Florida,

stratification did not appear to be a significant problem because nearly all students and teachers were in a common stratum within each county.

Persistence of Inputs to Achievement

It is generally accepted that achievement is a cumulative process and that achievement measured by standardized tests reflects both past and current inputs of educational and other factors as shown in model [1]. As noted above, models make numerous assumptions about the persistence of past inputs to allow for estimation of the effect of current teachers. For instance, many applications, both statistical and econometric, either implicitly or explicitly assume that teacher effects persist undiminished in all future years of student achievement.

However, there is growing evidence that teacher effects fade on standardized assessments and this can result in biased estimates. Studies that estimate α s in model [3] find past teacher inputs into future test scores decay rapidly, declining to about 10–20% of their initial magnitude after only 1 year. Studies using alternative methods also find that effects fade rapidly and never find persistence close to 100%. Limited empirical results demonstrate that errors in the assumptions about persistence can lead to biased estimation of the variability of teachers and systematic errors in teacher effect estimates.

Scaling of Achievement Measures

All VAM estimates of teacher effects depend on achievement test scores, but different approaches place more or less stringent requirements on these tests. In all cases, value-added estimates can only exist in grades and subject areas in which testing occurs and prior achievement data exist. In addition, teachers can only be evaluated on the constructs measured by the achievement tests. Critics sometimes fault the standardized tests that are the foundation of VAM for being too narrow and not measuring higher-order thinking skills, and question the utility of VAM estimates derived from these test scores.

In addition to potential limitations in the range of content of standardized tests, psychometricians have raised concerns about comparability of scores from different grade-level tests even when tests are vertically linked to a single scale. The primary concern is that tests at each grade level measure multiple constructs, and shifts in the mix of constructs across grades can distort test score gains, invalidate assumptions of perfect persistence of teacher effects, and bias VAM estimates.

Model [3] makes less-stringent assumptions about scaling than models that assume complete persistence and this model or variants of it might be better suited for some test score data. However, limitations in the scale, such as floor and ceiling effects and measurement error,

that is, greater at the top and bottom of the scale might invalidate model [3]'s assumptions of normality and implicit linearity in the relationships between scores from pairs of years. Errors about linearity in the relationships among scores from multiple years can be particularly problematic when using these scores to control for differences among widely disparate classes, as is done implicitly in the statistical approaches to VAM. Linearity of scores can be explored empirically and models or scales should be adjusted if the model does not conform to the data.

Concerns about Precision

Bias in estimated effects is of particular concern because the goal of VAM is to evaluate teacher performance on student achievement without conflating teachers' effects with their students' backgrounds, but stakeholders remain skeptical that any evaluation system can meet this goal. However, bias is not the only threat to the utility of VAM estimates. Estimates might also lack sufficient precision to support desired inference. Studies that report precision of individual effects generally show that annual estimates have very large errors. For example, standard errors are typically sufficiently large that about two-thirds of estimated teacher effects are not significantly different from average. Confidence intervals are wide and rankings contain substantial errors, so that only about 30–35% of teachers ranked in either the top or bottom quintile in 1 year remain there the next year. Similarly, studies consistently find that the correlation of estimated effects from pairs of adjacent years is low, ranging from less than 0.2 to typically no higher than 0.5.

Summary

Value-added models are likely to play an increasing role in education research and practice. In particular, there is strong interest in these methods for evaluating teacher performance. Although there is no consensus about the best VAM, there is a general consensus that estimating effects with complex models like models [2] and [3] with multiple years of test scores provides more accurate estimates of teacher effects than simple approaches using just 2 years of scores (a pre- and a post-score). Estimating teacher effects with the average residuals from linear regression models predicting the current score with a single prior achievement score seems particularly prone to bias that favors teachers with classes of higher-achieving students.

However, none of the current VAM methods is robust to all three potential sources of bias: nonrandom assignment of students to classrooms, misspecification of the persistence of prior educational inputs, and the limitations of

achievement tests. Econometric models need to be expanded to allow for more flexible modeling of the persistence of effects and statistical models need to take advantage of flexible modeling of persistence and more fully account for classroom assignment and school inputs. Other problems that need to be addressed include the potential bias from the large number of students with incomplete test score data, general issues of data quality such as verifying student–teacher links, and accounting for teachers who provide only a portion of a student’s instruction. Future research will need to identify the most productive methods for pooling data to increase the precision of estimates. It will also need to explore appropriate ways to use these estimates to support decisions about teachers and train educators to make effective use of the data while minimizing unintended negative consequences that can arise in accountability systems based on limited performance measures. Regardless of the shortcomings of the current models and the need for significant additional research, value-added assessments are clearly an improvement over other metrics used in accountability and provide a valuable attempt to use sophisticated methods to draw the most from increasingly available longitudinal test score data.

See also: Equating and Scaling; Student Test Results in School Accountability.

Further Reading

- Ballou, D. (2008). Test scaling and value-added measurement. *Paper Presented at the Wisconsin Center for Educational Research National Conference of Value-Added Modeling*. April 2008, Madison, WI.
- Ballou, D., Sanders, W., and Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics* **29**, 37–65.
- Braun, H. (2005). *Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models*. Princeton, NJ: Educational Testing Center.
- Goldschmidt, P., Roschewski, P., Choi, K., et al. (2005). *Policymakers’ Guide to Growth Models for School Accountability: How Do Accountability Models Differ?* Washington, DC: The Council of Chief State School Officers.
- Harris, D. and Sass, T. (2006). *Value-added models and the measurement of teacher quality*.
- Kane, T. J. and Staiger, D. O. (2008). Are teacher-level value-added estimates biased? An experimental validation of non-experimental estimates. *Paper Presented at the Wisconsin Center for Educational Research National Conference of Value-Added Modeling*. April 2008, Madison, WI.
- Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee value added assessment system. *Educational Evaluation and Policy Analysis* **25**, 287–298.
- Lissitz, R. (ed.) (2005). *Value-Added Models in Education: Theory and Applications*. Maple Grove, MN: JAM Press.
- Lockwood, J. R. and McCaffrey, D. F. (2007). Controlling for individual heterogeneity in longitudinal models, with applications to student achievement. *Electronic Journal of Statistics* **1**, 223–252.
- Martineau, J. (2006). Distorting value-added: The use of longitudinal, vertically scaled student achievement data for value-added accountability. *Journal of Educational and Behavioral Statistics* **31**, 35–62.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., and Hamilton, L. S. (2003). *Evaluating Value-Added Models for Teacher Accountability, MG-158*. Santa Monica, CA: RAND Corporation.
- McCaffrey, D. F., Sass, T. R., and Lockwood, J. R. (2008). The intertemporal stability of teacher effect estimates. *Paper Presented at the Wisconsin Center for Educational Research National Conference of Value-Added Modeling*. April 2008, Madison, WI.
- Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edn. Thousand Oaks, CA: Sage.
- Raudenbush, S. W. and Wilms, J. D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics* **20**, 307–335.
- Rothstein, J. (2008). “Student sorting and bias in value added estimation: Selection on observables and unobservables.” *Working Paper*. Princeton University and NBER.
- Sanders, W. L. and Rivers, J. C. (1996). *Cumulative and Residual Effects of Teachers on Future Academic Achievement*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.
- Sanders, W., Saxton, A., and Horn, B. (1997). The Tennessee value-added assessment system: A quantitative outcomes-based approach to educational assessment. In Millman, J. (ed.) *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* pp 137–162. Thousand Oaks, CA: Corwin Press.
- Schafer, W. D. (2006). Growth scales as an alternative to vertical scales. *Practical Assessment, Research and Evaluation* **11**. <http://pareonline.net/getvn.asp?v=11&n=4> (accessed May 2009).
- Todd, P. and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal* **113**, F3–33.
- Wainer, H. (ed.) (2004). *Special Issue: Value-Added Assessment*. *Journal of Educational and Behavioral Statistics* **29**, 1.
- Webster, W. and Mendro, R. L. (1997). The Dallas value-added accountability system. In Millman, J. (ed.) *Grading Teachers, Grading Schools*, pp 81–99. Thousand Oaks, CA: Corwin Press.

Relevant Websites

- <http://www.caldercenter.org>– National Center for Analysis of Longitudinal Data in Education Research.
- <http://www.performanceincentives.org>– National Center on Performance Incentives.

T

TEACHER EDUCATION – EVALUATION IN THE TEACHER EDUCATION PROGRAM

Contents

Accreditation and Standards in Teacher Education Determining Long Term Effects of Teacher Education

Accreditation and Standards in Teacher Education

R J Yinger and K L Daniel, University of Cincinnati, Cincinnati, OH, USA

© 2010 Elsevier Ltd. All rights reserved.

Accreditation is a process by which an educational institution demonstrates its capacity to adequately educate and prepare its students. Accreditation most always includes institutional self-examination and external review processes designed to assure the government, citizenry, and other stakeholders that institutions and programs meet particular standards for educational quality. In the United States and in some provinces in Canada, accreditation review is conducted as a nongovernmental and peer-review process. In other countries, accreditation is seen as a government ministerial function or is conducted by public authorities in partnership with independent professional/occupational bodies.

This article considers the role and process of specialized accreditation as a mechanism for the profession of teaching to both demonstrate and improve the quality and effectiveness of teacher education programs. Central to determining the future of professional teacher education is a debate about the control and governance of accreditation processes. This debate involves two different conceptions of teaching. On the one hand, is the view that teaching is a skilled occupation that should either be regulated by local or national governments in the interests of the state or be deregulated such that it is governed by market forces. On the other hand, is the view that teaching is a profession, and as such, should be governed and regulated by teaching professionals themselves.

The Governance of Teacher Education

Gideonse (1993) describes three modes of governance for teacher education. A political mode of governance places primary responsibility on elected officials or government-appointed bodies to set teacher education policy and regulate its implementation. An institutional mode of governance places primary responsibility with the higher education or training institutions for design and implementation decisions. Occupational independence and self-regulation is the hallmark of the professional mode of governance. This usually takes the form of independent professional boards comprised largely of educational practitioners.

Historically, all three of these forms of governance for teacher education have existed in more or less pure forms or in combination with others. As education has become a significant national political issue in recent years, remaining instances of institutional forms of governance have declined. Worldwide, political modes of governance still dominate, but there has been increasing interest within teaching and teacher education to assert more influence through independent professional bodies. Young *et al.* (2007) provide three detailed cases that illustrate the kinds of debates and political contests playing out in the governance of teacher education recently in England and Canada.

In Europe, teacher education is currently dominated by political forms of governance. Among 30 countries recently surveyed with support from the European Commission's Directorate-General for Education and Culture, all but one have a regulated system for evaluating initial teacher education (Eurydice, 2006). In the majority of these countries, quality assurance evaluations are conducted by an agency, committee, or independent body acting on behalf of the public authorities. In the remainder, teacher education evaluation is the direct responsibility of the ministry of education, an inspectorate, or a combination of the two.

The Eurydice survey indicates that in 2006, about one-half of the countries used standards for teacher education to guide the quality-assurance process. Those countries who did not specify standards often used evaluation criteria to shape the evaluation process. Only six countries had no published criteria related to evaluation. Virtually all countries with teacher education regulations evaluate the content of teacher education, teaching methods, and student assessment. All but four countries reported taking into account teacher education student performance.

Though independent professional bodies were established as early as 1954 in the United States (National Council for the Accreditation of Teacher Education) and in 1966 in Scotland (General Teaching Council of Scotland), professional efforts by teachers and teacher educators gained maximum ground since the 1980s. Two Canadian provinces established colleges for teachers (British Columbia in 1987 and Ontario in 1997), which have sought to define professional standards for teachers and to influence government standards and regulations for teacher education.

In 2004, the Australian federal government funded the Australian Institute for Teaching and School Leadership (usually referred to as Teaching Australia) as an independent body to investigate the creation of national standards for teaching and a national accreditation process for teacher education. Currently, only the states of New South Wales, Victoria, and Queensland have legislation requiring formal accreditation or approval of teacher education programs. In June 2007, Teaching Australia released a proposal for a national system of teacher education accreditation including principles, standards, processes, and governance. Trials of this new system were proposed to begin in 2008.

The recent efforts to promote independent accreditation as the best means to improve the effectiveness of teacher education are grounded in notions of professionalism that emphasize the value of specialized knowledge and training for solving client problems and meeting social needs. Implementing this professional model requires a claim by the profession and the granting by society of a jurisdiction for the definition and conduct of this work and for the preparation of effective practitioners.

Jurisdictional Functions of Accreditation and Standards

Accreditation and standards play a central role in particular conceptions of professional work by scholars who focus on social jurisdiction and authority and who argue that professionalization involves establishing claims to specialized knowledge, expertise, and to a particular set of work tasks (Abbott, 1988). This conception is in contrast to theories that portray professionalism mainly as a project of collective mobility where professions struggle as occupations for market shelters and protection of economic interests (Krause, 1996). It also contrasts with an emerging body of theory that characterizes professionalism as a cultural and political development focusing on social responsibility and the public good (Sullivan, 2005).

According to the jurisdictional conceptions of professions, the distinctive work of professionals is the ability to effectively solve human problems using specialized, expert knowledge. Professional jurisdiction refers to claims of control and authority for dealing with particular kinds of problems, and these claims are based primarily on control of particular knowledge and skill. According to Abbott (1988), jurisdictional claims have three components: (1) claims to classify problems, (2) claims to reason about problems, and (3) claims to act on problems in particular ways. The essential ingredient to effectively accomplish these tasks, according to this theory, is the control of abstract knowledge.

The successful modern professions such as medicine, law, and engineering have tied their work to a system of knowledge that formalizes the skills by which professional deliberation and action proceed. In the last century, the formalization of this abstract knowledge has taken place in the university context, and scholars like Abbott assert that the ability of a profession to establish and sustain its jurisdictional authority lies partly in the power and prestige of its abstract, academic knowledge. The ability to control this knowledge and to set the standards for its transmission and application in university-based professional preparation programs is a central rationale for peer-regulated professional accreditation.

Efforts to professionalize teaching and teacher education have been ongoing in recent years, with most examples found in countries such as England, Canada, Australia, and the United States – all of which have a social history of independent professions in medicine, law, engineering, and the like. Because the United States has a long track record of professionalization efforts, including numerous national initiatives seeking to accomplish this goal, the history and current context of teacher education accreditation in the US will be used as a case study to raise major issues and aspects of professional accreditation processes.

Teacher Education Accreditation and Standards: A Case History

Accreditation and standards have played a central role in the professionalization strategy employed by educators in the US since the middle of the twentieth century. Modeled after strategies successfully used in the first-half of the twentieth century by medicine and law, the goal of this internal control strategy is to gather and consolidate control and jurisdiction for teachers' work by means of scientific and scholarly knowledge, university preparation, and internally generated standards for preparation and licensure (Yinger, 2005b).

Professionalism Ascending

The beginning of this professionalization process can be marked formally by the creation of the National Council for Accreditation of Teacher Education (NCATE) in 1954 and the development of consensus-based standards for what new teachers should know and be able to do and what university-based programs should do to produce this desired knowledge and skill (Wise, 2005). The 1980s and 1990s were times of ambitious reform initiatives for the teaching profession in the United States. There was growing recognition that professional standing for educators would require a concerted effort to establish professional standards, a scientific knowledge base, and a new status in relationship to education policy and policymakers. The National Board for Professional Teaching Standards to develop standards and assessments for the board certification of accomplished teaching, the Interstate New Teachers Assessment and Support Consortium to develop performance-based standards and assessments for initial teacher licensure, and teacher education reform organizations such as the Holmes Group and the National Network for Educational Renewal were established.

Central to these initiatives was an increasing sense of convergence and consensus on standards for teaching and teacher education and on the conception of teaching as a career-development process spanning initial teacher education, new teacher induction and support, and continuing professional development activities. These ideas found a crescendo in the 1996 report of the National Commission on Teaching and America's Future, *What matters most: teaching for America's future* (NCTAF, 1996). NCTAF proposed the goal: "within a decade – by the year 2006 – we will provide every student in America with what should be his or her educational birthright: access to a competent, caring, qualified teacher in schools organized for success." Two major recommendations of the commission to achieve this goal were (1) get serious about standards, for both students and teachers and (2) reinvent teacher preparation and professional development. These were to be enabled by a tripartite strategy of recommending that

every teacher in America would graduate from an accredited teacher education program, be licensed using performance-based standards, and work toward the accomplished teacher certification standards established by the National Board for Professional Teaching Standards. By the end of the century, there was growing optimism about the success of this strategy, but growing challenges as well (Yinger, 1999).

Challenges to the Professional Establishment

While the groups above sought to improve teaching and learning by means of consolidating professional jurisdiction for teaching and teacher education, critics outside of education were focusing on other aspects of professionalization. Their criticisms and recommendations for educational reform were based conceptually in the portrayal of the professional agenda as that of pursuing economic and political interests through exclusiveness, elitism, and protectionism. Their goal was to break up the public education establishment's market-shelters through deregulation, competition, privatization, and alternative choices. This applied not only to schooling, but also to teacher education and licensure (Hess *et al.*, 2004).

Alternative organizations such as the National Council on Teacher Quality and the Education Leaders Council were created to advocate opening up public education in the US to the invisible hand of the free-market system, which would allow student and parent choice to correct the problems of the system. With the change of the US Presidency in 2000, deregulation and choice, enforced by a national system of accountability for student and school academic performance did, in fact, become the backbone of federal educational policy. This policy has taken its most influential form in the 2002 reauthorization of the federal government's Elementary and Secondary Education Act referred to as the No Child Left Behind Act. At the state level, multiple programs were established to allow and promote alternative routes for teacher preparation and for teacher licensure for similar reasons.

Within teacher education, two further developments slowed the movement of the internal control professionalization agenda. In 2003, the federal agency that oversees accrediting organizations approved a second accreditor for teacher education, the Teacher Education Accreditation Council (TEAC). This new organization was formed as an alternative to what a number of university administrators and teacher educators regarded as a narrowly controlled, overly cumbersome, and expensive NCATE accreditation process. (It can also be described as an effort to assert the institutional form of governance for teacher education described earlier.) This event set in motion heated and continuing debates about the need for more than one national accreditor in teacher education, the most effective processes for accreditation review, and who should be the ultimate arbiter for teacher

education program standards and quality (Murray, 2005; Tamir and Wilson, 2005; Wise, 2005).

The second development to challenge the internal control strategy in recent years has been the publication of several high-profile reviews of the research literature related to teacher education. A key component of the notion of profession jurisdiction is to demonstrate a scientific knowledge base on which professional practice and preparation is founded. In spite of numerous studies of teaching and teacher education over the past 30 years, research reviews have been unable to substantiate a scientific research base for teacher education (Cochran-Smith and Zeichner, 2005). This situation has reinforced external criticisms that other factors, such as general intelligence and teacher commitments, have greater impact on effective teaching than learned knowledge, skill, and expertise (Walsh, 2001). In turn, this has also ignited calls for more rigorous, scientifically based research in education (Shavelson and Towne, 2002).

Standards-Based Accreditation: Current Challenges

Advocates of teacher education accreditation as a quality control and quality-improvement strategy assert that the strength of the current approaches is that these approaches are standards-based and focused on outcomes (Murray, 2005; Wise 2005). Yet, at same time, there are stakeholders outside of teacher education, as well as within, who continue to question the legitimacy and value of accreditation. Three underlying reasons can be attributed to this condition.

Weak Professional Consensus

The NCATE, the oldest and largest teacher education accreditor in the United States, prides itself on a standards-based approach built on a broad professional consensus built over many years. The difficulty with this stance is that to obtain consensus, NCATE has had to settle for broad, generalized standards that enable them to be met by institutions with differing missions, goals, and capacities. For instance, Standard 1 of the NCATE Unit Standards reads:

Candidates preparing to work in schools as teachers or other professional school personnel know and demonstrate the content, pedagogical, and professional knowledge, skills, and dispositions necessary to help all students learn. Assessments indicate that candidates meet professional, state, and institutional standards.

Though standards-assessment rubrics are provided for institutions and reviewers, there is little operationalization by way of behavioral specificity as to adequately demonstrate evidence of compliance. Underlying the need

to allow broad interpretation of standards is the felt need to accommodate the multiple goals, opinions, and ideologies held by faculty of education and others involved in professional coalitions. The approach can be presented as a professional consensus, but unlike other professional fields, the result is a divergent rather than convergent notion of professional education.

Weak Conceptions of Evidence

Both NCATE and the newer US accreditor, TEAC, state that they rely on evidence to make their judgments about quality. This is true, but both accrediting agencies have weak logics of evidence and therefore weak standards of what constitutes evidence. Again, the solution both have chosen to address the problem of institutional and ideological diversity in teacher education is to deliberately keep expectations broad and inclusive and allow institutions to determine and argue for what constitutes evidence. The TEAC self-study process is built upon institutional determination of what they choose to offer as evidence. Contributing to the difficulty of offering stronger conceptions of evidence is the paucity of valid measures of teacher candidate learning and performance. Also, the predominance of qualitative research as the preferred method for much of the research in teacher education creates difficulties of systematically aggregating, synthesizing, and interpreting research findings beyond local contexts (Cochran-Smith and Zeichner, 2005). Altogether, the lack of a professional logic of evidence prevents current accreditation frameworks from setting and holding professionals responsible for standards necessary to ground an evidence-based profession.

Mixed Professional Support for Accreditation

Teacher education is unique among professional education programs in the US in that nationally recognized accreditation is not universally accepted and supported. National accreditation is a requirement for professional schools in the established professions such as medicine, law, and engineering, and graduation from an accredited professional school is a prerequisite for taking licensure examinations in these fields.

To date, fewer than half of the 1300 education schools in the United States are currently accredited by either NCATE or TEAC. Studies have shown that those institutions who do not participate in national accreditation are not what the field would consider having the weakest programs. A number of highly regarded programs in prestigious universities choose not to participate. Nonparticipating education schools do so for a variety of reasons, but surveys by the American Association for Colleges of Teacher Education have indicated that these institutions believe that the accreditation process does not add value to

their own quality assurance and improvement efforts and is not essential to their demonstration of the effectiveness of their programs.

Additionally, research on the value of teacher education is limited and offers mixed conclusions (Tamir and Wilson, 2005). This fact combined with weak professional consensus and a weak logic of evidence to legitimize effective professional education practices have not surprisingly created a situation where there is limited regard for the impact of teacher education from many university administrators who make decisions about encouraging and financing accreditation processes. Mixed messages on the value added to teacher education accreditation within universities get communicated to students and parents, so that national accreditation is not regarded as a public symbol for quality in the minds of individuals making application decisions to teacher education programs. These same messages are heard by politicians at the state and federal levels, so that national accreditation has not become a required component of education policy, in spite of its advocacy by the professional teacher education organizations and accreditors.

A Future for Teacher Education Accreditation

Given the challenges to accreditation practices in teacher education cited above and the lack of success with the internal control strategies in other professions (Yinger, 2005a), it is unlikely that a continuation of current accreditation practices will reach the professionalization goals of teacher educators or increase the confidence and trust of teacher education stakeholders. Accomplishing these goals will require a systems intelligence strategy.

This strategy requires a two-pronged approach involving strengthening the jurisdictional claims for teacher education combined with a new professional orientation by teacher educators. This strategy has been successfully demonstrated in health-related fields, such as medicine and nursing, but does not currently exist in education. Realizing this strategy in teacher education will require additional institutional expertise and capacity and willingness to reexamine the professional orientation of teacher education faculty.

Systemic Intelligence for Teacher Education

Accreditation is designed to assure the public and other stakeholders that institutions and programs meet particular standards for educational quality.

Demonstrating this involves carefully making two prior decisions: (1) What are the standards for program effectiveness? (2) How should program effectiveness be operationalized?

As described above, current US teacher-education accreditation approaches address these decisions in order to maximize institutional and program freedom and flexibility. This approach reveals a professional logic that both decision makers and scholars employ: that the education system is so unique and its system characteristics so intractable that there is limited opportunity for continuous, substantive systemic improvements. These intractable characteristics include system complexity, unique organization, fragmentation, autonomy, lack of accountability, and the inability to agree on the ultimate goal of education. As a result of this logic, decisions about educational priorities are often being made *ad hoc*—from intuition, opinion, local preferences, and most dangerously, out of political self-interest.

The field of healthcare, facing similar dilemmas nearly two decades ago, pioneered a science of evidence which combined the key areas of systematic evidence synthesis of clinical research, cost analytics, and multicriteria decision making. This introduced an unprecedented form of systemic intelligence, which enabled leadership to enlarge their work to systemic problem solving (Elstein, 2004). Like leaders in healthcare, in marketing, and in environmental and space sciences before, education leaders must embrace these proven techniques to allow them to analyze multiple streams of disparate information and make decisions that are proven to have optimal impact.

These other fields have utilized a decision science that enables continuously getting systemically more intelligent: systemic intelligence. Proven methods have been demonstrated for transparently solving intractable system issues such as: resolving multiple system goals; determining the best data to measure system effect; handling community and political preferences; responding to a lack of rigorous professional evidence; and including expert judgment. These similarly complex systems overcame intractable system characteristics by executing a particular science which formalized the criteria for systemic effectiveness (Parmigiani, 2002). This enabled leaders of the profession to execute a deliberate, transparent, and rigorous evidence logic utilizing the tools of systematic evidence synthesis, multicriteria decision analysis, and enterprise technology infrastructure.

The ability, capacity, and will to accomplish this kind of system level of work in teacher education has recently been realized by one research group in the United States—the Teacher Quality Partnership, a teacher education research consortium in the state of Ohio (Yinger *et al.*, 2007). This science has not yet been applied to teacher education accreditation. To do so will require a commitment to evidence-based decision making and to capacity building in professional organizations and across teacher education programs.

Determining evidence-based program quality standards for teacher education first requires a professional evidence

logic that can identify and claim a professional expertise. To substantiate this particular expertise, research work must focus on: (1) the functional relation of student-learning outcomes and teaching practices and (2) the functional relation of teaching practices and professional learning practices (i.e., professional development and teacher education). To build an evidence-based accreditation system requires an immediate and long-term commitment to simultaneous research working on both these linkages, which fundamentally define our professional expertise. Necessary steps to be taken include:

- Developing a professional consensus on what constitutes evidence. Education researchers are often divided on ideological grounds, but other fields are successfully resolving these controversies.
- Identifying professional effectiveness criteria. There is growing consensus in the United States that the contribution to student academic achievement growth is an important indicator of teacher effectiveness.
- Developing a research knowledge base that demonstrates evidence-of-effect for particular teaching practices and particular professional learning activities. This will require more research designs and analyses that enable causal inference, such as controlled experimental trials.
- Establishing teacher education standards based on strong research evidence demonstrating causal relationships to teacher effectiveness, and where strong evidence is not yet available, incorporate rigorously applied expert judgment.

These are ambitious goals, but success in other complex professional systems demonstrates the possibilities and the means to achieve them. However, merely applying this science will not be sufficient. True evidence-based professional practice requires combining research evidence and professional judgment with client and public goals and preferences. This requires a new professional stance.

Teacher Education as a Public Profession

There is a growing body of theory that approaches professionalism as a cultural and political development focusing on social responsibility and the public good. Sullivan (2005) has proposed the reconstruction of professionalism from the dominance of scientific management practices shaped by corporate bureaucratic and technocratic ideals to a goal of responsive practices shaped by a commitment to public service. Yinger (2005a) has applied these ideas in a proposal to understand teachers as public professionals who have specialized knowledge and moral obligations to the common good. Professional capital depends on society for its legitimization, and part of this legitimacy (jurisdiction) is based on society's expectations that professional

capital will be used to address human problems and promote public purposes. Perkin (1989) described this professional social ideal as prizing mutual service, efficient use of resources, and the responsible use of knowledge for the larger good.

The stance of teachers as public professionals requires a new model of professional relationships with those they serve. Rather than seeing teaching as a knowledge transaction primarily controlled by educators, public professionals view interactions with clients and the community as work driven by mutual responsibility grounded in shared values and goals. May (2000) portrays this relationship as a professional covenant, where the focus of the professional relationship moves beyond expert service to incorporate the promissory integrity and trust necessary to make common work serve the common good.

Framing the work of teachers as public professionals creates a new form of professional accountability. Currently, educational accountability in the US is most often framed as a rating or grading scheme relying upon narrowly conceived performance outcomes. Many educators approach this process by seeking to limit public incursions on their work and by agreeing to fairly general and easily met accountability measures such as average passing rates on standardized academic tests. Public professional relations differ in that the commitment to mutual goals and shared responsibility yields increased involvement in goal setting by noneducators and increased transparency by professionals about their ability to successfully deliver results according to agreed-upon outcomes.

In the context of teacher education accreditation, a public professions commitment and stance by university teacher educators eschews the internal control jurisdictional agenda in favor of a more open and interactive relationship with teacher education's stakeholders – prospective and current teachers, school administrators, and parents and children. Accreditation would continue to be a peer-based activity, but the notion of qualified individuals would be expanded further to include teachers, parents, and the public.

Worries that this transparency and involvement would undermine and erode professional standards and quality can be countered by examples in medicine where patients and the public have been invited in to improve both clinical practice and to revise professional standards and accountability frameworks.

There is a growing movement in the United States to improve hospital practices and outcomes by public involvement strategies directed at evaluating specific clinical outcome results. The Institute for Healthcare Improvement has partnered with numerous American hospitals to create evidence-based medicine initiatives. These initiatives rely upon the application of research-based clinical practices combined with patient and community

engagement activities aimed at jointly setting health-outcome priorities and monitoring detailed data on clinical success rates.

In the United Kingdom, the General Council of Medicine and the National Health Service have for over a decade been involved in public engagement strategies to reform the standards and accountability frameworks for physician and hospital practice. This work has been built on asserting a new professionalism with the patient at the center of medical practice. This has manifested itself in professional preparation focusing on patient-centered medicine and formal public involvement in standard setting and the judgment of quality (Irvine, 2003). Both these examples demonstrate that established professions can create institution-level discourse with public stakeholders, which can produce a re-making of the profession and a re-building of public trust.

Summary

This article has located issues of teacher education standards and accreditation within the framework of professional goals, work, and societal relationships. Efforts for teacher professionalism in the past half-century have been described as a movement by educators to seek professional control of the work of teaching, its standards for practice and preparation, and its social jurisdiction. Political and social counterforces have been shown to have frustrated this work, but the root problems preventing a strong case for professional oversight and control of teaching and schooling processes have been described as weak professional consensus and weak evidence.

Establishing cognitive jurisdiction for teaching requires a professional knowledge base established on strong research evidence. Similarly, social jurisdiction for teaching requires renewed public trust established on evidence of effectiveness and shared education purposes. A professional strategy, shown to be successful in healthcare fields, built upon systemic intelligence and public engagement, has been described as a means to ground the future of teacher education on evidence-based standards and to enable the establishment of universally respected, evidence-based accreditation.

Bibliography

- Abbott, A. (1988). *The System of Professions: An Essay on the Division of Expert Labor*. Chicago, IL: University of Chicago Press.
- Cochran-Smith, M. and Zeichner, K. (eds.) (2005). *Studying Teacher Education: The Report of the American Educational Research Association Panel on Research and Teacher Education*. Mahwah, NJ: Lawrence Erlbaum.

- Eurydice (2006). *Quality Assurance in Teacher Education in Europe*. Brussels: Eurydice European Unit.
- Elstein, A. S. (2004). On the origins and development of evidence-based medicine and medical decision making. *Inflammation Research* **53** (supplement 2), 184–189.
- Gideonse, H. D. (1993). The governance of teacher education and systemic reform. *Educational Policy* **7**(4), 395–426.
- Hess, F. M., Rotherham, A. J., and Walsh, K. (2004). *A Qualified Teacher in Every Classroom? Appraising Old Answers and New Ideas*. Cambridge, MA: Harvard University Press.
- Irvine, D. (2003). *The Doctors' Tale: Professionalism and Public Trust*. Abingdon: Radcliffe Medical Press.
- Krause, E. A. (1996). *Death of the Guilds: Professions, States, and the Advance of Capitalism: 1930 to the Present*. New Haven, CT: Yale University Press.
- May, W. F. (2000). *The Physician's Covenant: Images of the Healer in Medical Ethics*, 2nd edn. Louisville, KY: Westminster John Knox Press.
- Murray, F. B. (2005). On building a unified system of accreditation in teacher education. *Journal of Teacher Education* **56**, 307–317.
- National Commission on Teaching and America's Future (1996). *What Matters Most: Teaching for America's Future*. New York: National Commission on Teaching and America's Future.
- Parmigiani, G. (2002). *Modeling in Medical Decision Making*. Chichester: Wiley.
- Perkin, H. (1989). *The Rise of Professional Society: England since 1880*. New York: Routledge.
- Shavelson, R. J. and Towne, L. (eds.) (2002). *Scientific Research in Education*. Washington, DC: National Academy Press.
- Sullivan, W. (2005). *Work and Integrity: The Crisis and Promise of Professionalism in North America*, 2nd edn. San Francisco, CA: Jossey-Bass.
- Tamir, E. and Wilson, S. M. (2005). Who should guard the gates? Evidentiary and professional warrants for claiming jurisdiction. *Journal of Teacher Education* **56**, 332–342.
- Walsh, K. (2001). *Teacher Certification Reconsidered: Stumbling for Quality*. Baltimore, MD: Abell Foundation.
- Wise, A. E. (2005). Establishing teaching as a professional: The essential role of professional accreditation. *Journal of Teacher Education* **56**, 318–331.
- Yinger, R. J. (1999). The role of standards in teaching and teacher education. In Griffin, G. A. (ed.) *The Education of Teachers: 98th Yearbook of the National Society for the Study of Education*, pp 1–18. Chicago, IL: University of Chicago Press.
- Yinger, R. J. (2005a). A public politics for a public profession. *Journal of Teacher Education* **56**, 285–290.
- Yinger, R. J. (2005). The promise of education. *Journal of Education for Teaching* **31**, 307–310.
- Yinger, R. J., Daniel, K. L., and Lawton, M. (2007). The teacher quality partnership research enterprise: Enabling systemic understanding and improvement. *Paper presented at the biennial meeting of the European Association for Research on Learning and Instruction*. Hungary: Budapest, 28 August to 1 September.
- Young, J., Hall, C., and Clarke, T. (2007). Challenges to university autonomy in initial teacher education programmes: The case of England, Manitoba, and British Columbia. *Teaching and Teacher Education* **23**, 81–93.

Further Reading

- Cochran-Smith, M. (2005). The new teacher education: for better or worse? *Educational Researcher* **34**(7), 3–17.
- Darling-Hammond, L. (2001). Standard setting in teaching: Changes in licensing, certification, and assessment. In Richardson, V. (ed.) *Handbook of Research on Teaching*, 4th edn., pp 905–947. Washington, DC: American Educational Research Association.
- Murray, F. (2000). The role of accreditation reform in teacher education. *Educational Policy* **14**(1), 40–59.

Relevant Websites

<http://www.aacte.org/index.php?/accreditation-issues/> – American Association of Colleges for Teacher Education.
<http://www.teachingaustralia.edu> – Australian Institute for Teaching and School Leadership Limited.
<http://www.ihl.org> – Institute for Healthcare Improvement.
<http://www.ncsl.org> – National Conference of State Legislatures.
<http://ncate.org> – National Council for Accreditation of Teacher Education.
<http://www.bcct.ca> – The British Columbia College of Teachers.

<http://www.holmespartnership.org> – The Holmes Partnership.
<http://www.nbpts.org> – The National Board for Professional Teaching Standards.
<http://www.nctq.org> – The National Council on Teacher Quality.
<http://www.oct.ca> – The Ontario College of teachers.
<http://www.ed.gov> – US Department of Education.
<http://nnerpartnership.org>.
http://www.aacte.org/Programs/Accreditation_Issues/.
http://ccsso.org/projects/Interstate_New_Teacher_Assessment_and_Support_Consortium/.

Determining Long Term Effects of Teacher Education

C N Brouwer, Radboud University Nijmegen, Nijmegen, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The existence of programs for preservice teacher education (TE) is not self-evident. The life and sustainability of such programs ultimately depend on the effects they have on teachers' in-service behavior. Effects of TE are defined here as durable consequences of the preservice acquisition of knowledge and skills and development of motives, as manifested in graduates' professional behavior. A stricter definition pertains to long-term effects of TE programs. Here, such effects are defined as durable consequences of participation in preservice programs which manifest themselves in professional behavior more than 1 year after the completion of preservice TE. It is the responsibility of educational research to demonstrate, describe, explain, and help optimize the relationship between preservice TE and the professional behavior which teachers exhibit and develop during their careers. The description and explanation of long-term effects of preservice programs lend greater credence to the utility of TE.

Relevance of TE Effect Research

Studying TE effects and especially long-term effects empirically is a relevant endeavor in several respects. Quantitative arguments in favor of TE effect studies deal with issues of recruitment, selection, and retention. Any country that wishes to sustain its educational system must know what contribution TE programs make to the availability and employability of its teacher workforce. Like no other component of a country's education system, TE possesses a multiplier function for the quality of schooling and learning throughout the system as a whole. TE programs deliver generations of teachers, who in turn educate generations of pupils and students, who on their part carry with them and transform the knowledge and skills acquired in school, college, or university during the rest of their lives and work. Indeed, the contributions that TE can make to the productivity of a country's educational system eminently justify research into the influence of TE programs on the teaching competence of their graduates.

Aim of TE Effect Research

The aim of TE effect research should be to explain the influence that specific features of preservice programs exert on teaching competence, especially in the longer run. If predictors of success in teaching can be found before

and during TE as well as during beginning teaching and the further career, this evidence could underpin guidelines and strategies for improving TE learning environments. Empirical evidence about long-term TE effects could not only help increase retention in the profession, but also lay a basis for developing teachers' professional practice and well-being during all stages of their careers. The relevance of long-term effect studies for policymaking in the domain of TE is clearly illustrated by the debate about what is an adequate or even the best architecture of TE programs, be they regular 4-year programs or alternative certification programs (e.g., Darling-Hammond and Youngs, 2002). Teacher shortages worldwide lend urgency to such questions. As yet, the empirical knowledge base for designing learning environments for teachers is growing, but seriously incomplete. There is a clear need for research which explains specifically how TE learning effects come about.

TE as a research field characterizes itself by an enormous complexity. Program effects are not exclusively influenced by program features, but these interact with the personal backgrounds of candidates and the varied contexts in which their schooling takes place. To explain the complex interplay between conditions, processes, and outcomes involved in teacher learning, an adequate research base is not yet available, the American Educational Research Association (AERA) Panel on Research and Teacher Education has concluded. However, improvements in research design are underway (Cochran-Smith and Zeichner, 2005, p. 704). This conclusion is even more applicable when long-term effects of TE are taken as criteria.

This article focuses on what we know about TE effects as well as how we know it. Coming to grips with the complexity of teacher learning is crucial to scientific progress in the field of TE. The empirical evidence about effects of TE programs are therefore referenced with special attention to the strengths and weaknesses of TE effect studies. On this basis, validity requirements for TE effect research are explicated and productive research strategies are presented. For this purpose, a number of pioneering and exemplary studies serve as illustrations.

Contested Ground. What We Know about TE Effects and How We Know Them

Empirical evidence about effects of TE programs has begun to emerge during the last decades of the twentieth century, both in Europe and in the English-speaking

world. Toward the end of the century, growing teacher shortages in the United States spurred survey research on a larger scale into the outcomes of alternative certification programs. These programs, developed initially in the US since the 1980s, were intended to attract new candidates to the teaching profession (Dill, 1996; Zeichner and Schulte, 2001) and often consisted of abbreviated or condensed versions of regular TE programs. The need to evaluate the adequacy of these programs has put effect research into TE programs, both alternative and regular programs, high on the agenda. However, long before that time, in fact since the formalization and institutionalization of TE programs began in the 1920s, qualitative research has been carried out representing the experiences of teacher educators in developing programs and of candidates while learning to teach.

The State of the Art

The state of the art in TE effect research becomes apparent from seminal studies and research reviews in this domain.

Seminal studies

Seminal research exploring how teachers become and develop as teachers was reported by Lortie in his well-known book *School Teacher*. In two case studies using survey questionnaires and retrospective interviews, Lortie vividly showed and persuasively argued how a web of social and psychological forces conditions the ways in which teachers are socialized by the existing school system (Lortie, 1975). Lortie's sociological angle reminds the reader of the risks of unreflective reproduction of craft traditions of teaching. In his wake, Lacey explored attitude change in aspiring teachers as they enter the school system and traced the various strategies of occupational socialization open to them. On the basis of five case studies, in which observations and questionnaires were used to follow prospective teachers as they moved from preservice TE through their first year of beginning teaching, Lacey (1977, 1995, p. 619) distinguishes among internalized adjustment, strategic compliance, and strategic redefinition as possible ways for beginning teachers to deal with existing traditions in schools.

Until the 1970s, large-scale empirical evidence about long-term TE effects was virtually nonexistent; in fact, even now, shortly after the turn of the century, it is scantily available. The project Teacher Attitudes, carried out at Konstanz University in the south of Germany, was the first in Europe to demonstrate by means of path analysis on national longitudinal survey data the influence of specific features of preservice programs on outcomes in graduates (Dann *et al.*, 1978, 1981). This project showed that program features encouraging the integration of theory and practice can delay and/or mitigate the occurrence of practice shock and the associated attitudinal adjustment to

existing teaching practices. Building on Kelman's (1974) work on types of attitude change, Dann *et al.* (1978, pp. 96–104) and Müller-Fohrbrodt *et al.* (1978) have identified as crucial in beginning teachers' socialization what they call discrepancy experiences; they experience a rift between idealistic notions developed during TE programs, on the one hand, and pressure from schools as institutions to rely on traditional patterns of behavior, on the other. These discrepancy experiences are strongest in situations where practical action is required (Dann *et al.*, 1981).

This phenomenon of conservative attitude shift was demonstrated repeatedly in the US by the Pupil Control Studies initiated by Hoy (Packard, 1988). It was also documented in early qualitative research by Corcoran (1981), who described entry into teaching in terms of transition shock.

Perhaps the most important contribution of the Konstanz research group is that program characteristics were shown to differ in how fast and in what respects they influenced the U-curve in prospective and beginning teachers' attitude development (Dann *et al.*, 1978, 1981). This evidence supports the conclusion that integrative approaches in TE, in which student teachers' practical experiences are closely linked to theoretical input, can strengthen graduates' innovative teaching competence (Dann *et al.*, 1978, 1981). Similar research and findings were reported from Hamburg by Hinsch (1979) and from the Netherlands by Brouwer and Korthagen (2005).

Research reviews

Both the European research referenced above and the comprehensive review undertaken by the AERA Panel on Research and Teacher Education in the US confirm that TE programs do influence graduates' learning outcomes. The AERA review contains indications that coursework can promote graduates' subject-matter knowledge and that methods courses combined with field work can have an impact on their professional behavior (Cochran-Smith and Zeichner, 2005). The beneficial influence of integrating theory and practice during teacher preparation on teachers' in-service performance was confirmed in the US by reviews and secondary analyses of partly cross-sectional, partly longitudinal survey data carried out by Darling-Hammond (1999, 2000). This synthetic work has yielded two important conclusions: (1) the amount and quality of teacher preparation are positively related to teachers' retention in the profession as well as their success in promoting student achievement and (2) teacher quality is the single most influential factor determining student achievement.

Taken together, these conclusions mean that preservice TE programs have a discernible and relevant influence on teaching and learning in schools, especially if they integrate practical and theoretical components. However, precisely how these effects come about and which specific

program features are responsible for which kinds of effects are questions which require further study of a sophisticated kind.

How to Research Long-Term Effects of Educating Teachers?

So far, much TE research lacks the qualities needed to refine and elaborate upon the available evidence. Wilson *et al.* (2002) reviewed 57 studies about the impact of TE programs published in peer-reviewed journals mostly during the 1990s. They concluded that this research was often small scale, allowed researcher bias, employed few direct measures, and often lacked in detail in reporting about methodology. Similar criticisms were voiced by Wideen *et al.* (1998) in their review of the research about learning to teach. In their judgment, the theoretical basis of such studies should be made more explicit. They also note a lack of longitudinal research and advocate a systemic and ecological approach in TE effect research. Considering this state of affairs, the AERA Panel on Research and Teacher Education calls for research taking into account the contexts of TE programs, participants' backgrounds and characteristics, and how these relate to the processes and outcomes of TE programs. Such research requires employing longitudinal designs more frequently and a theory-driven conceptualization of outcome measures, whose operationalization should go beyond indirect verbal data.

These reviews, with quite a some consensus, point in the same direction. The outcomes of TE programs probably never result from a limited number of social conditions and personal characteristics. They had better be conceived of as resulting from a complex interplay of factors which simultaneously create and constrain opportunities for learning. To understand how this complex interplay affects teacher learning, a type of relational knowledge is needed, as described by Apple (1979); in other words, a systems perspective which pays attention to the manifold relationships between the components of TE as a social system. If TE effect research is to answer its questions, it would benefit not only from stronger conceptualization generally, but also, especially from using ecological and developmental concepts to theorize about its research object.

Modeling the object of research

In order to explain how all the factors operating in TE programs work together to generate teacher learning, TE effect research should rest on more systematic theorizing. A suitable vehicle for achieving this is the decision-oriented evaluation model introduced by Stufflebeam and Webster (1988). This model distinguishes between different types of variables, designated as context, input, process, and

product (CIPP). When applied to preservice TE, context factors include the institutional and cultural constraints in schools, colleges, and universities within which TE programs are offered. Input factors include personal background characteristics of teacher candidates and the resources for training and guiding them on their way to beginning teaching. Process factors are the ways in which programs are implemented, including the degree to which and the ways in which practical exercise and theoretical study are adjusted to each other. These process factors are decisive mediators determining the nature of learning effects in graduates. These effects fall under the CIPP model component product. Models for TE effect research should take into account candidate, program, and context features and preferably include factors from each model component relevant for the study at hand. This kind of modeling encourages the researcher to examine the relationships and the balance of forces between influences emanating from TE programs and from school contexts in particular.

Ecological conceptions of the objects of social research go rather against the grain of mainstream educational psychology, but if the TE research community wants to develop a methodology that is congruent with theorizing about the complexity of teacher learning, it could consider a paradigm shift from a mechanistic toward an organicist worldview (cf. Baltes *et al.*, 1977). In developmental psychology, ecological conceptions have long been available, notably in the work of Bronfenbrenner (1979), who situates the individual as one of the driving forces in a nested structure of systems, that is, the micro-system, in which he or she directly participates, the mesosystem, which constitutes the institutional constraints within which human activity unfolds, and the macrosystem on national and world levels determining the material and cultural conditions of human development. Researching the individual as an entity developing in its social context, warns Bronfenbrenner, is seeking to understand his or her exosystem, that is, how the relationships between the three subsystems distinguished mediate personal development.

Methodology of exemplary studies

What we know about TE effects is determined to an important extent by how we know them. Empirical studies are inevitably shaped by the way in which researchers relate their theoretical perspectives to their choice and use of methods. It is from this angle that we now examine an illustrative selection of exemplary TE effect studies in order to try and cull from the researchers' experience suggestions for conducting ecological inquiry into the influence of TE programs on the teacher workforce and its professional competence. The studies presented below are all examples of longitudinal research.

Phenomenographic studies

Since the 1970s, especially, there has been a host of exploratory, phenomenographic case studies into prospective and beginning teachers' experiences in becoming a teacher (see the review by Wideen *et al.*, 1998). Most of this research is purely qualitative and involves small numbers of respondents, if only because the researchers lacked the resources to process large amounts of qualitative data. The result has been to describe in interpretive ways prospective teachers' experiences as they learned to teach. An advantage of this approach is that it can illuminate the processes and mechanisms involved in teacher learning, especially if the "inner workings" of the TE programs concerned are carefully described (Cochran-Smith and Zeichner, 2005: p. 700). However, studies which restrict themselves to using only qualitative methods in small samples may fail to identify and explain which program features significantly encourage later performance in teaching practice. Without larger numbers of respondents and strategic sampling, TE effect research cannot produce generalizable knowledge about productive learning environments for teachers.

Large-scale quantitative studies

At the other extreme is purely quantitative survey research capitalizing on the law of great numbers, which uncovers patterns and relationships, notably concerning issues of recruitment (cf. the pertinent entry), selection, and retention.

A persuasive example is Ingersoll's (2001) demonstration of a revolving-door effect in teacher turnover. By analyzing with regression techniques the national longitudinal data sets in the US spanning 7 years, Ingersoll showed that contrary to public and political opinion, there are sufficient numbers of teachers available who are willing to fulfill schools' staffing needs. However, too many of them grow dissatisfied too soon with the work conditions in education and therefore move to different schools or leave teaching profession altogether. Chin and Young (2007) reported how personal backgrounds of alternative-route candidates such as ethnic status, age, gender, family obligations, and career history interact with local and regional employment opportunities as well as attractive characteristics of teaching as a profession in determining recruitment patterns.

In the area of teacher selection and self-selection, various longitudinal studies were carried out tracing the backgrounds, employment, and work experience of graduates of alternative certification programs. A review of these studies makes it clear that rather than general measures, such as grade point averages based on standardized achievement tests, candidates' cognitive skills, and levels of cognitive complexity, their self-efficacy and previous experience in instructing and coaching people in group settings have some merit as predictors of retention or at

least success in teaching as rated by teacher educators and employers (Brouwer, *et al.* 2007).

Time-dependent patterns in retention have been uncovered by Shen (2003), using large-scale longitudinal data. Dropout of beginning teachers was shown to increase with time, often in leaps at the beginning of a new school year or semester. In the same research, a positive relation was found between the length of training and the length of time teachers remained in the profession. In addition, support for teachers after graduation can influence their commitment to schools. Various induction packages consisting of different types of support for teachers who were new to a school were shown to differ significantly in their influence on percentages of turnover after the first year of employment. Much of the turnover occurred among teachers without any type of support during induction. Turnover was reduced among teachers who had supportive communication with school leaders, and even more so when they were given opportunities to collaborate with colleagues and had access to teacher networks and extra resources (Smith and Ingersoll, 2004).

In the area of attitude development, Watzke (2007) analyzed the development of work-related concerns in 79 beginning teachers using six repeated measures with a concerns inventory. Fuller (1969) and others have posited a stage chronology, which has tended to portray beginning teachers' development as a series of subsequent steps proceeding from concerns about maintaining a desirable presentation of self through concerns about performing teaching tasks proficiently to concerns about the adequacy of one's own teaching in promoting student learning. In contrast to this self-task-impact-stage chronology, concerns about impact on learners were shown in Watzke's sample to rank consistently highest across time. Concerns about impact, whether academic or personal in orientation, were shown to recur over time instead of characterizing one specific phase of beginning teacher development.

Mixed-methods studies

A study based on ecological assumptions and relating processes and outcomes is the reconstruction of competence development in 357 teachers from the beginning of preservice training until the third year of in-service practice reported by Brouwer and Korthagen (2005). Building on the work of Dann *et al.* (1978) and using a conceptual model derived from the literature on teacher socialization, this study assessed the relative influences on competence development of preservice TE programs on the one hand, and school contexts before and after graduation on the other. The methods used in this study were questionnaires, document analysis, observations, and retrospective interviews with student teachers as they developed into beginning teachers, cooperating teachers, and teacher educators. Among the main findings were that a carefully planned alternation of student-teaching and college-based

periods, a gradual increase in the complexity of student-teaching activities, and the use of a clinical supervision model in cooperative small-group settings encouraged graduates' professional competence, notably their ability to activate learners.

Longitudinal mixed-methods approaches capable of addressing the complexity of teacher learning are also used on a larger scale in the Pathways into Teaching Project in New York City schools. In this project, comprehensive quantitative surveys are combined with detailed qualitative descriptions of an array of different TE programs and the learning experiences of the participants in them (Boyd *et al.*, 2006).

Validity requirements

Against this backdrop of examples and with due caution, the methodological strengths and weaknesses of studies of long-term TE effects can be assessed as follows. Explaining relationships between TE program features and outcomes in graduates is facilitated by combining quantitative and qualitative methods. Using quantitative data from sizeable samples enables the researcher to establish significant relationships and patterns and begin to generalize findings. Using qualitative data, preferably from direct observation, enables the researcher to reconstruct and infer from thick description (Geertz, 1973) in which settings and in which ways teachers acquire and transform professional teaching behaviors. TE effect research with explanatory power is further characterized by longitudinal rather than cross-sectional designs, repeated rather than one-shot measures, and multivariate correlational rather than experimental or quasi-experimental methods of analysis.

Given these experiences and trends in the conduct of TE effect research, it is appropriate to explicate what should count as valid causal explanations in an ecological conception of TE effect research. Therefore, two validity requirements are now presented for the benefit of explaining the complex whole of factors determining effects of TE programs. Conclusions about cause-effect relationships in TE effect research should be considered valid only, when:

- relationships are demonstrated on the basis of a research model incorporating context, input, process, and product factors and
- the processes responsible for these relationships are reconstructed.

The latter demand implies that explanations of human learning are incomplete, if only an experimentalist, black-box approach is taken. Such an approach would underestimate the role that subjective representations, such as perceptions, experiences, opinions, and attributions, play in human learning (cf. Maxwell, 2004). These two validity requirements embody a causal-genetic conception of

scientific explanation in educational research and share important assumptions about the object of research as put forward in Bronfenbrenner's ecological conception of human development.

Suggestions for Conducting TE Effect Studies

The validity requirements proposed above lay a basis for arriving at decisions about research design, instrumentation, data collection, and data analysis (cf. the criteria for research programs advocated by Bronfenbrenner and Lüscher, 1976).

Design

An ecological conception of TE programs and effects as an object of causal-genetic explanation would entail including in the design of an empirical study the main settings of the college-based curriculum, student teaching, and beginning teaching in schools. Prospective teachers should be followed longitudinally from the beginning of training until the second year of in-service teaching, preferably longer. The instruments and measures used should allow recording and understanding the backgrounds, actions, experiences, and motives of the persons studied.

Combining qualitative and quantitative methods helps solve the breadth-depth problem in investigating TE programs and effects. If one type of method is used exclusively, either many situations and persons are studied superficially, or a few are studied thoroughly (Berger, 1974). A solution to this problem is to select, from all the programs and respondents in a study, a smaller number to form a representative subset. These may be termed the whole sample and the subsample, respectively. The whole sample can then be studied using quantitative methods, so that generalizable relationships and patterns can be discovered. The subsample can be studied additionally by means of qualitative methods, so that mediating processes can be reconstructed and explained in sufficient detail.

Instrumentation

In order to arrive at a multifaceted representation of both the outcomes and processes of long-term teacher learning, it is advisable to use multiple instruments for data collection (cf. for the original idea of a multitrait-multimethod matrix in social research, Campbell and Fiske, 1959). Not only self-report methods should be used, but also direct observation, so that attributions by any respondent group can be recognized for what they are and reliable accounts of actual professional performance can be achieved.

The operationalization of observations and measures is crucial for validity. Three epistemological perspectives, that is, assumptions about the nature of TE programs and effects can serve as a guide for operationalization: the ecological, genetic, and activity perspectives. The ecological perspective assumes that the research object in the social sciences always consists of a social system, which is at the same time internally structured and embedded within a wider, often institutional context (cf. Maschewsky, 1979; Tabachnick, 1981). The genetic perspective recognizes all learning as a set of processes, whose unfolding influences learning. This perspective assumes that understanding learning outcomes requires understanding the processes which produce them (cf. Davydov, 1977). The activity perspective entails a focus on the actions of people as these express the continual tension between their personal motives on the one hand and contextual constraints on the other (cf. Zeichner and Tabachnick, 1985).

In quantitative inquiry, low-inference measures should be employed wherever possible. In qualitative inquiry, a balance should be sought between self-report and direct observation data. On-site visits, video recording of teaching, and retrospective and stimulated-recall interviews are therefore suitable methods.

Data Collection

In an ecological inquiry of TE programs and effects, minimally, the main respondent groups should be involved, that is, prospective, beginning and/or experienced teachers, cooperating teachers, teacher educators, and/or professional development facilitators (cf. Cochran-Smith and Zeichner, 2005, p. 331–334).

An important consideration in designing TE effect research is which time spans should be covered in longitudinal data collection. Decisions on this point are obviously made depending on the research questions and available resources, but it should be borne in mind that learning effects need time to come to fruition and to manifest themselves. To determine long-term effects of TE, it is advisable to cover time spans of at least 2 years. When following respondents over time, it is of course necessary to identify them, secure their consent with continued participation, inform them of the study's progress, and take all other measures useful for guarding against attrition generally (Mednick *et al.*, 1984).

Analysis Methods

Having based research design on ecological assumptions, TE researchers face strategic choices in analyzing pluriform data sets. Formulating and accounting for these choices with the help of a predetermined analysis plan can make or break a study's success. Evaluating any specific TE program in an ecological fashion confronts the researcher with social

constellations and processes of unique complexity. Tracing distal causes and mediating factors will be prominent challenges (cf. Baltes *et al.*, 1977). An analysis plan should therefore serve careful hypothesis generation and explicate the reasons why certain analysis techniques are selected and how their use is sequenced.

Given the validity requirement that the explanation of learning outcomes should rest on a reconstruction of the learning processes involved, it is logical to make qualitative analyses preceding quantitative ones, so that hypotheses can be generated and explored in a controlled way. If multivariate analyses yield clear indications of testable relations and/or differences, then experimental or quasi-experimental analyses can follow. This sequence conforms to the plea by Chatterji (2004, 2006) for extended-term mixed-method evaluation (ETMM) designs.

Finally, in analyzing quantitative longitudinal data, caution is needed in handling missing values. Attrition of respondents can easily cause individual data pairs to drop out of specific correlations and regressions. This risk increases to the degree that data pairs are collected further apart in time. To contain this risk, techniques can be used for estimating missing values (Peugh and Enders, 2004).

Conclusion

Qualifying and retaining teachers requires building an empirical knowledge base which can underpin program design in TE. Answering questions relevant for explaining and optimizing long-term effects of preservice TE programs can be achieved by means of carefully planned longitudinal studies combining quantitative and qualitative methods. TE researchers are well advised to conceptualize their questions using explicit, ecological models, which help them to theorize about the crucial role of mediating processes in generating learning effects in teachers. Research design, instrumentation, data collection, and analysis should fulfill validity requirements which are congruent with research models aiming at causal-genetic explanation.

If TE effect research were to succeed in mastering its challenges, evidence-based decision making could become possible in policymaking for TE. At the beginning of the twenty-first century, the greatest challenge for research into the long-term effects of TE is to help solve the dilemma of countering teacher shortages while simultaneously maintaining, enhancing, and assuring the quality of the teacher workforce worldwide.

Bibliography

- Apple, M. (1979). *Ideology and Curriculum*. London: Routledge and Kegan Paul.
- Baltes, P. B., Reese, H. W., and Nesselroade, J. R. (1977). *Life-Span Developmental Psychology: Introduction to Research Methods*. Monterey, CA: Brooks/Cole.

- Berger, H. (1974). *Untersuchungsmethode und soziale Wirklichkeit (Research Method and Social Reality)*. Frankfurt: Suhrkamp.
- Boyd, D. J., Grossman, P., Lankford, H., et al. (2006). Complex by design: Investigating pathways into teaching in New York City schools. *Journal of Teacher Education* **57**(2), 155–166.
- Bronfenbrenner, U. (1979). *The Ecology of Human Development*. Cambridge, MA: Harvard University Press.
- Bronfenbrenner, U. and Lüscher, K. (eds.) (1976). 'Ökologische Sozialisationsforschung' (*Ecological Socialization Research*). Stuttgart: Ernst Klett Verlag.
- Brouwer, C. N. and Korthagen, F. A. J. (2005). Can teacher education make a difference? *American Educational Research Journal* **42**, 153–224.
- Brouwer, C. N., Ackley, B. C., and Fallon, M. A. (2007). Intake assessments for alternative teacher education: Moving from legitimation towards predictive validity. *Assessment and Evaluation in Higher Education* **32**(6), 657–665.
- Campbell, D. T. and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* **56**, 81–105.
- Chatterji, M. (2004). Evidence on "What works". An argument for extended-term, mixed-methods (ETMM) evaluation designs. *Educational Researcher* **33**, 3–13.
- Chatterji, M. (2006). Grades of evidence. *Paper presented at the AERA Annual Meeting 2006*. San Francisco, CA.
- Chin, E. and Young, J. W. (2007). A person-oriented approach to characterizing beginning teachers in alternative certification programs. *Educational Researcher* **36**, 74–83.
- Cochran-Smith, M. and Zeichner, K. (eds.) (2005). *Studying Teacher Education. The Report of the AERA Panel on Research and Teacher Education*. Mahwah, NJ: Erlbaum.
- Corcoran, E. (1981). Transition shock: The beginning teacher's paradox. *Journal of Teacher Education* **32**, 19–23.
- Dann, H.-D., Cloetta, B., Müller-Fohrbrodt, G., and Helmreich, R. (1978). *Umweltbedingungen innovativer Kompetenz. Eine Längsschnittuntersuchung zur Sozialisation von Lehrern in Ausbildung und Beruf (Contextual Conditions of Innovative Competence. A Longitudinal Study of Teacher Socialization During Preservice Training and Beginning Teaching)*. Stuttgart: Klett-Cotta.
- Dann, H.-D., Müller-Fohrbrodt, G., and Cloetta, B. (1981). 'Sozialisation junger Lehrer im Beruf. Praxisschock drei Jahre später' (Occupational socialization in young teachers. Practice shock three years later). *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* **13**, 251–262.
- Darling-Hammond, L. (1999). *Teacher Quality and Student Achievement*. Seattle, WA: University of Washington, Center for the Study of Teaching and Policy.
- Darling-Hammond, L. (2000). How teacher education matters. *Journal of Teacher Education* **51**(3), 166–173.
- Darling-Hammond, L. and Youngs, P. (2002). Defining "highly qualified teachers": What does "scientifically-based research" actually tell us? *Educational Researcher* **31**, 13–25.
- Davydov, V. (1977). *Arten der Verallgemeinerung im Unterricht (Types of Generalization in Education)*. Berlin: Volk und Wissen.
- Dill, V. S. (1996). Alternative teacher certification. In Sikula, J., Buttery, T., and Guyton, E. (eds.) *Handbook of Research on Teacher Education*. New York: Macmillan.
- Fuller, F. F. (1969). Concerns of teachers: A developmental conceptualization. *American Educational Research Journal* **6**, 207–226.
- Geertz, C. (1973). Thick description: Toward an interpretative theory of culture. In Geertz, C. (ed.) *The Interpretation of Cultures*, pp 3–30. New York, NY: Basic Books.
- Hinsch, R. (1979). 'Einstellungswandel und Praxisschock bei jungen Lehrern, eine empirische Längsschnittuntersuchung' (*Attitude Change and Practice Shock in Beginning Teachers, An Empirical Longitudinal Study*). Weinheim: Beltz.
- Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal* **38**, 499–534.
- Lacey, C. (1977). *The Socialization of Teachers*. London: Methuen.
- Lacey, C. (1995). Professional socialization of teachers. In Anderson, W. (ed.) *International Encyclopedia of Teaching and Teacher Education*, pp 616–620. Oxford: Elsevier-Science/Pergamon.
- Lortie, D. C. (1975). *Schoolteacher, A Sociological Study*. Chicago, IL: University of Chicago Press.
- Maschewsky, W. (1979). 'Zur Bedingungskontrolle in der psychologischen Forschung: Rationale, Möglichkeit, Notwendigkeit' (Experimental control in psychological research: Its rationale, possibility, and necessity). In Jäger, M., Leiser, E., Maschewsky, W., and Schneider, U. (eds.) *Subjektivität als Methodenproblem, Beiträge zur Kritik der Wissenschaftstheorie und Methodologie der bürgerlichen Psychologie*, pp 142–174. Cologne: Pahl-Rugenstein.
- Maxwell, J. A. (2004). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher* **33**, 3–12.
- Mednick, S. A., Harway, M., and Finello, K. M. (1984). *Handbook of Longitudinal Research*. New York: Praeger.
- Müller-Fohrbrodt, G., Cloetta, B., and Dann, H. D. (1978). 'Der Praxisschock bei jungen Lehrern. Formen-Ursachen-Folgerungen. Eine zusammenfassende Bewertung der theoretischen und empirischen Erkenntnisse' (*Practice Shock in Young Teachers: Forms, Causes, and Consequences. An Evaluative Summary of Theoretical Insights and Empirical Evidence*). Stuttgart: Klett-Cotta.
- Packard, J. S. (1988). The pupil control studies. In Boyan, N. J. (ed.) *Handbook of Research on Educational Administration*, pp 185–209. New York: Longman.
- Peugh, J. L. and Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Research in Education* **74**, 525–557.
- Shen, J. (2003). Characteristics of those entering teaching via traditional and alternative certification: A national, longitudinal study of fresh college graduates. *Paper Presented at the 2003 Annual Meeting of the American Educational Research Association*. Kalamazoo, MI: Western Michigan University.
- Smith, T. M. and Ingersoll, R. M. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal* **41**, 681–714.
- Stufflebeam, D. and Webster, W. J. (1988). Evaluation as an administrative function. In Boyan, N. J. (ed.) *Handbook of Research on Educational Administration*, pp 569–603. New York: Longman.
- Tabachnick, B. R. (1981). Teacher education as a set of dynamic social events. In Tabachnick, B. R., Popkewitz, T., and Szekely, B. B. (eds.) *Studying Teaching and Learning: Trends in Soviet and American Research*, pp 76–86. New York, NY: Praeger.
- Watzke, J. L. (2007). Longitudinal research on beginning teacher development: Complexity as a challenge to concerns-based stage theory. *Teaching and Teacher Education* **23**, 106–122.
- Wideen, M., Mayer-Smith, J., and Moon, B. (1998). A critical analysis of the research on learning to teach: Making the case for an ecological perspective on inquiry. *Review of Educational Research* **68**, 130–178.
- Wilson, S. M., Floden, R. E., and Ferrini-Mundy, J. (2002). Teacher preparation research: An insiders' view from the outside. *Journal of Teacher Education* **53**, 190–204.
- Zeichner, K. M. and Schulte, A. K. (2001). What we know and don't know from peer-reviewed research about alternative teacher certification programmes. *Journal of Teacher Education* **52**, 266–282.
- Zeichner, K. M. and Tabachnick, B. R. (1985). Social strategies and institutional control in the socialization of beginning teachers. *Journal of Education for Teaching* **11**, 1–25.

Further Reading

- Bourdieu, P. and Passeron, J.-C. (1977). *Reproduction in Education, Society and Culture*. London: Sage.
- Brophy, J. E. and Good, T. L. (1986). Teacher behavior and student achievement. In Wittrock, C. M. (ed.) *Handbook of Research on Teaching*, 3rd edn., pp 328–376. New York: MacMillan.
- Cole, A. L. and Knowles, J. G. (1993). Shattered images: Understanding expectations and realities of field experiences. *Teaching and Teacher Education* **9**, 457–471.

- Dangel, J. R. and Guyton, E. M. (eds.) (2005). *Research on Alternative and Non-traditional Teacher Education: Teacher Education Yearbook XIII*. Lanham, MD: Scarecrow Education/Rowman and Littlefield.
- Frech, H. W. and Reichwein, R. (1977). Der vergessene Teil der Lehrerbildung: Institutionelle Bedingungen und inhaltliche Tendenzen im Referendariat der Gymnasiallehrer (*The Forgotten Part of Teacher Education: Institutional Conditions and Substantive Trends in the Practical Preparation of Gymnasium Teachers*). Stuttgart: Klett-Cotta.
- Fullan, M. G. and Stiegelbauer, S. (1991). *The New Meaning of Educational Change*. London: Cassell.
- Korthagen, F. A. J., Kessels, J., Lagerwerf, B., and Wubbels, T. (2001). *Linking Practice and Theory. The Pedagogy of Realistic Teacher Education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Sarason, S. B. (1996). *Revisiting The Culture of the School and the Problem of Change*. New York: Teachers College Press.
- Veenman, S. A. M. (1984). Perceived Problems of Beginning Teachers. *Review of Educational Research* **54**, 143–178.
- Waller, W. (1967). *The Sociology of Teaching*. New York: Wiley.

Relevant Websites

- www.ctpweb.org – Centre for the Study of Teaching and Policy (CTP).
- www.teacherpolicyresearch.org – Teacher Policy Research (TPR).

TEACHER EDUCATION – EVALUATION OF THE PROSPECTIVE TEACHER

The Evaluation of Prospective Teachers in Teacher Education

D E H Tigelaar and J van Tartwijk, ICLON Leiden University School of Training, Leiden, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Teacher evaluation has been a major focus in literature on teacher education in the last few decades. Methodology for teacher evaluation has been described in special handbooks and many conceptual and research papers have been published about this subject.

In this article on the evaluation of prospective teachers, several aspects and issues are addressed that play an important role in evaluation in teacher education. We revisit the way that changing views on good teaching have influenced teacher evaluation and how evaluation has evolved into assessment, and we give a description of various assessment methods. We also focus on current trends in the assessment of prospective teachers. Finally, the quality of teacher assessments is discussed. Before that, there is a brief introduction to the different purposes of the evaluation of prospective teachers.

Purposes of the Evaluation of Prospective Teachers

The evaluation of prospective teachers can serve various purposes. First, it can be used as a selection method to determine whether individuals are qualified enough to be admitted into teacher education programs. Tests of basic skills in reading, writing, and mathematics; subject matter knowledge; and pedagogical knowledge can be part of such evaluations. Second, it can diagnose and influence the performance and professional growth of student teachers while they are involved in teacher education programs. Third, it can be used to determine whether student teachers should be given their initial teaching licensure. Finally, the results of the evaluation of prospective teachers can be used to monitor the effectiveness of teacher education programs and to account for the outcomes of these programs. These four purposes: selection, diagnosis, licensure, and accountability can be categorized into two main types of evaluation: (1) formative evaluation which is focused on diagnosing and monitoring progress and (2) summative evaluation which is focused on selection, licensure, and accountability.

Views on Good Teaching and Implications for Teacher Evaluation

From the 1960s to today, views on good teaching have evolved. Each of these views has influenced teacher evaluation and the influence of each view can be recognized in different traditions of research on teaching. At first, the ideal good teacher was supposed to be a warm and stable person. Research on teaching in this tradition mainly focused on analyzing teachers' personality characteristics and, from this, predicting the classroom climate that teachers created. In this research, teacher personality traits, such as warmth and indirectness, were related to students' responses in the classroom. However, it proved difficult to define and identify an effective teacher personality.

The most prominent foci in research from the mid-1960s till date have been teacher behavior and teacher cognitions and decision making. These research traditions and their inherent views on good teaching are described below, along with their implications for teacher evaluation.

Teacher Behavior

From the mid-1960s until the early 1970s, in line with the dominance of the behavioral psychology of learning at the time, educational researchers focused on teachers' instructional behaviors and how these related to student-learning outcomes. This implied that a good teacher was expected to possess a repertoire of effective behaviors for optimal student learning. In most studies, a process-product approach was applied in which systematic observation schedules were used to measure and evaluate a teacher's behavior or classroom interaction (the process) and to use these variables to predict student-learning outcomes (the product). Research in the teacher behavior tradition gradually resulted in long lists of effective teacher behaviors and observation schemes to score and evaluate these behaviors. When the implementation of these research findings did not actually lead to improvements in teaching practice, attention shifted to the teacher-thinking processes.

Teacher Cognitions and Decision Making

In the 1970s, influenced by the cognitive revolution in psychology, the focus of research on teaching shifted from teacher behavior to what actually goes on in the mind of the teacher. Examples of topics that were investigated were teachers' expectations, thought processes, decision making and planning, beliefs, practical knowledge, conceptions of subject matter and how subject matter should be taught to students, and differences in the cognitive orientations of novice and expert teachers.

From research on teacher cognitions and decision-making processes, it was increasingly recognized that for good teaching, thought processes are as equally important as behaviors, and that teaching is a complex activity which is heavily bound to the context (e.g., the subject taught, the conditions in the school, and the students' background). It was also acknowledged that teachers actively construct their own knowledge by interpreting events on the basis of existing knowledge, beliefs, and dispositions, and that teachers need a continuing capacity to learn from their experiences (Feiman-Nemser and Remillard, 1996). Although studies on teacher cognitions were not usually related to student outcomes, insights from this research have had implications on the way teaching is evaluated (Haertel, 1991; Darling-Hammond and Snyder, 2000; Uhlenbeck *et al.*, 2002). Teacher evaluation has developed from testing knowledge with paper-and-pencil tests, or using standardized observations, into a process aimed at gaining a comprehensive picture of the development of a teacher's knowledge and performance. Examples of this new type of evaluation, known as assessment, include: peer assessment, self-assessment, assessment in a simulated teaching environment, and the use of portfolios. These methods are described below.

Peer assessment

In peer assessment, colleagues or student teachers assess each other and provide each other with feedback (Boud *et al.*, 1999). Training in using peer-assessment procedures and providing effective feedback is necessary (Dierick and Dochy, 2001; Sluijsmans *et al.*, 2002). Usually, in peer assessment, groups of individuals are involved in defining criteria and procedures for their own assessments. In this way, they internalize assessment criteria and learn to provide useful feedback. Peer ratings tend to be fairer and more accurate in comparison to assessments by supervisors. However, peers often find it difficult to criticize their friends, and thus guidelines for evaluation need to be very clear.

Self-assessment

During self-assessment, student teachers make judgments about their own work and the extent to which they have met criteria and standards (Airasian *et al.*, 1995; Ross and Bruce, 2007). Mostly, student teachers are involved in

identifying the criteria and standards they have to apply to their own performance. Self-assessment may lead to new insights or an awareness of areas that need to be improved and may help prospective teachers to exert more control and responsibility for their own professional development. However, self-assessment needs to be combined with other assessment strategies, for example, peer assessment or observation by supervisors, to validate student teachers' own judgments.

Simulations

The use of simulations has its origin in the wish to assess the complex performances that many work situations require. In assessment centers, student teachers complete simulations – a coherent set of tasks and activities which represent the most important aspects of the work situation, such as planning a lesson or evaluating student work, and explaining and justifying their approach and decisions (Uhlenbeck *et al.*, 2002). The level of task performance and the student teachers' rationale for their approach are the basis for making judgments about a student teacher's knowledge and skills (Haertel, 1991; Shulman *et al.*, 1988). However, developing and carrying out simulations is expensive and time consuming (Uhlenbeck *et al.*, 2002).

Portfolios

When portfolios were first introduced, teachers gathered artifacts (i.e., work samples) that were illustrative of the quality of their own classroom teaching, and wrote captions in which they described what these artifacts were evidence of. Since their introduction in the early 1990s, increasing emphasis has been put on including materials from others, such as feedback from students and colleagues. In this way, a teaching portfolio integrates aspects of peer and self-assessment. In portfolio assessment, the burden of proof is the primary responsibility of the assessed student teachers, who need to convince assessors with the evidence in their portfolios that their level of competence meets the criteria. Since their introduction, more emphasis has also been put on the quality of reflections in the commentaries on the evidence in portfolios (van Tartwijk *et al.*, 2007). In these reflective commentaries, student teachers need to show that they have learned from their experiences and gained improved understanding of their own practice through critical analysis and the use of theory, and by building and reframing their own practical knowledge. However, composing portfolios is very time consuming and the quality of reflections in portfolios is often disappointing (Mansvelder-Longayroux *et al.*, 2007). Nevertheless, portfolios have become widely used in many teacher education programs.

Learning-Oriented Assessment

In the last few decades, the terms learning-centered or learning-oriented, with reference to teacher assessment,

have become common in teacher education (i.e., Carless, 2007). Learning-oriented assessment places more emphasis on the learning elements of assessment processes in teacher education: the purpose of the assessment being to monitor and evaluate student teachers' progress. An optimal integration of assessment and the learning process is important. This means that assessment programs need to be developed that are consistent with the goals for learning in teacher education curricula. In these assessment programs, various methods need to be used and many sources of information should be collected over a long period of time in order to measure the different aspects of student teachers' progress. In this way, a comprehensive picture of the student teacher's development and growth can be gained. The collected evidence can be related to standards and criteria which provide insight into the goals for student teacher learning and performance within the curriculum. Performance levels, which indicate levels of both performance and growth, can also be used to monitor a student teacher's progress.

Quality feedback is an essential ingredient in optimizing assessment for learning. This feedback not only should be focused on nonevaluative descriptions of the characteristics of a student teacher's performance, but should also involve evaluative comments linked to criteria that indicate the quality of a student teacher's work. Finally, it should promote setting goals for improvement (Sadler, 1989). Ultimately, monitoring progress on the basis of assessment information and providing feedback results in a summative judgment about the student teacher when finishing teacher education and gaining licensure. This means that formative and summative goals are related and the latter can be seen as a continuation of the former. The question of who is responsible for assessing different parts of a student teacher's progress and providing support and feedback until finishing teacher education and gaining licensure is a dilemma in this respect. There is a potential role conflict when the supervisor or

mentor takes on the double role of guide and judge of the learning process. This dilemma may be resolved by considering different assessment scenarios. In Table 1, examples of assessment scenarios in the context of portfolio assessment for both formative and summative evaluation of student teachers are outlined.

Standards-Based Teacher Assessments

From the mid-1980s, many efforts have been put into defining a knowledge base that describes what teachers should know and be able to do. Criteria and standards have been developed that specify what should be taught in teacher education programs and programs for teachers' further professional development. A well-known example of such criteria and standards are those that were introduced in 1987 by the Interstate New Teacher Assessment and Support Consortium (INTASC) (see section 'Relevant Websites'). These standards were developed over a long period by a committee consisting of members of the teaching profession, and they have been very influential. Based on these standards, which are discipline-specific, portfolio assessments for ongoing licensure were developed, which required teachers to include several portfolio entries, such as lesson plans, videotapes of and comments about two lessons, and samples of students' work. The INTASC standards are now being used by many teacher education institutes in the United States to define their curricula and to set their own standards for admission, licensure, and further professional development. In recent years, many other countries have also developed standards for teaching and teacher education, for example, the United Kingdom, Australia, New Zealand, and the Netherlands.

Primarily, teaching standards were developed to recognize teachers' professional status, improve their performance, and contribute to their professional learning. However, due to political debate about the presumed

Table 1 Scenarios for the role of the supervisor in portfolio assessment for both formative and summative purposes

The "job-application" scenario

Just as in a job application, student teachers themselves are responsible for a portfolio which proves to an assessment committee that they meet the criteria. The supervisor is not part of this committee. One of the criteria can be that the student teachers demonstrate growth in their portfolio.

The "driver exam" scenario

In this scenario, just like a driving instructor, supervisors help student teachers to prepare themselves for an assessment. Subsequently, an assessor uses information in the portfolio, such as video-recordings and student evaluations, to assess the level of competence, without consulting the supervisor. The situation resembles that of an assessment center, but the evidence is collected in "real" situations.

The "PhD supervisor" scenario

In this scenario, the supervisor helps the student teacher to reach the required level of competence. Once the supervisor decides that the portfolio shows that the student teacher meets the criteria, he or she can decide to ask an assessment committee to look at the portfolio (just like a professor does with the PhD thesis). Most of the time this committee will confirm the supervisor's decision. But if the supervisor does not do a proper job and is too lenient, the assessment committee might reach a negative conclusion.

Source: Van Tartwijk, J., Driessen, E. W., Hoeberigs, B., Kösters, J., Ritzen, M., Stokking, K., and Van der Vleuten, C. P. M. (2003). *Werken met een elektronisch portfolio* [Working with an Electronic Portfolio]. Groningen: Wolters Noordhoff.

decline in schools' and students' performances and failures in previous reform efforts to improve these performances, policymakers have placed more emphasis on researching teacher quality and the value of teacher education programs. As a result, the accountability function of standards has become more prominent, in the United States as well as in other countries. This has led to evidence-based studies on teacher quality in relation to student outcomes, and studies on the quality of teacher education programs.

The Quality of Assessments in Teacher Education

The introduction of new forms of teacher assessment, aimed at evaluating how teachers perform in their own practice, has led to discussion of what determines the quality of these evaluations (see, for example, Andrews and Barnes, 1990). Basically, this discussion comes down to the validity and reliability of assessments. We use the example of classroom management in this section to clarify some technical aspects of the quality of assessments.

Validity

The traditional definition of validity is the extent to which a measurement or test measures what it was designed to measure. From this perspective, three types of validity, that is, content, criterion, and construct validity, can be distinguished as intrinsic properties of a test or assessment. Construct validity refers to the test score as a measure of the assessed characteristic, which should be defined in a conceptual framework. This means that when a test is designed to assess classroom management, for instance, a clear and detailed definition of the construct of classroom management must be offered and the test scores must be related to that construct. Content validity focuses on the degree to which the content of the test and the response properties are representative of a defined domain. An example of such a domain is the extent to which the test of classroom management covers the skills required for good performance in classroom management. Criterion validity refers to the degree to which test scores predict future performance and correlate with the results on other tests measuring the same construct. This implies that when a student teacher is being assessed on his/her classroom management, the scores on the test are expected to depict how the teacher will perform in classroom management in the future, and the findings on the test should be comparable to those on other tests that are intended to measure the student teacher's classroom-management skills.

An expanded notion of validity

Cronbach and Messick use an extended notion of validity. As an alternative to regarding validity as a property of the test as such, they relate validity to the meaning of the test

scores. According to Cronbach and Messick, it is the meaning or the interpretation of the score that must be valid along with any implications of this meaning. The interpretations by assessors, as well as the consequences for the assessed student teacher that follow from these interpretations, are all part of what needs to be valid. This implies that both assessors' scores on a test of classroom management and assessors' suggestions of how to improve classroom management need to be valid. This concept of validity results in a proposal by Messick (1989) for an integrated validity framework, which combines issues related to content, criterion, and construct validity with considerations of value implications and social consequences. The acknowledgment that validity goes further than the properties of a test and is related to interpretations, value implications, and consequences of the use that is made of a test, is generally accepted today in teacher assessments.

Defining constructs and assessment tasks and criteria

As outlined above, in order to judge a student teacher's performance, that is, on classroom management, a well-defined conceptual framework is needed as a basis for the description of the construct to be assessed. Assessment tasks and scoring criteria which are used to evaluate a student teacher's performance should be relevant to and representative of the construct that is to be measured. Quality criteria for assessments should include content validity and content coverage, as introduced by Linn *et al.* (1991), and the extent to which results are representative, as suggested by Frederiksen and Collins (1989). Defining assessment frameworks can be done using a realist approach, in which the concepts in the assessment framework are supposed to reflect reality, or using a more constructivist approach, in which multiple perspectives are regarded as possible (Moss, 1992). In a realist approach, when an assessment framework is developed, the concepts and the assessment framework that are used to define the construct (i.e., classroom management) are classified and researched until a detailed assessment framework has been constructed. The assessment framework is then used as a basis for hypotheses about a student teachers' performance (for instance, in relation to pupils' outcomes), which can be tested in the assessment, and the findings from the assessment are then expected to be a true score of a student teacher's performance. By contrast, in more constructivist approaches to assessment, concepts are seen as aspects to consider in the assessment, which are instrumental and meaningful to their users, but do not necessarily reflect reality. The constructivist approach implies room for multiple perspectives on what is important in a certain teaching area, such as classroom management. The methodological advice for developing assessment frameworks is to focus on the opinions of different stakeholders and to gain an understanding of the cultural context in order to agree on

the construct to be assessed. This means that when an assessment framework is being developed for classroom management, various people who might have interest in the assessment, for example, pupils, teacher educators, school leaders, and the assessed student teachers themselves, are consulted on their opinions with regard to what is important for a teacher's classroom management.

In the literature on teacher assessment, we find elements of both realist and constructivist approaches to the definition of assessment frameworks. The Educational Testing Service in the United States (see section 'Relevant Websites'), for example, adopted a more realist approach by conducting job analyses, research interviews, and analyses of state licensing requirements in order to define an assessment framework. On the other hand, the INTASC assessment frameworks were developed in committees of different stakeholders in education, which is more representative of the constructivist approach.

Consequential validity

Assessments are not only useful in determining the level of teaching competence, but are also diagnostic and can therefore provide learning opportunities for the student teachers involved. When classroom management is assessed, the assessment should provide student teachers with suggestions for further development and improvement in this teaching area. With reference to the benefits of assessment for teaching and learning, educators often speak of the consequential validity of assessments in terms of their potential influence on teaching and learning. Messick included these consequential concerns in his characterization of validity. Messick's recommendation in this respect is to focus on evidence and rationales in evaluating the intended and unintended consequences of the interpretation and use of assessment scores, particularly those associated with scoring bias or unfairness in test use. Any negative impact on student teachers or groups of student teachers should not be attributable to construct underrepresentation or construct-irrelevant variance. When a student teacher has low scores on a classroom-management assessment, this should not be caused by exclusion from the assessment of an aspect of classroom management which is relevant to the assessment and which, if present, would have allowed the assessed student teacher to display competence. Nor should low scores on classroom management be caused by something irrelevant in the measurement which interferes with the assessed student teacher's demonstration of competence.

Additional validity criteria, such as fairness, meaningfulness, and transparency, have been proposed by Linn *et al.* (1991) and Frederiksen and Collins (1989) to enhance consequential validity. Fairness implies that the consequences of assessment should be similar for every student teacher and that there should be no bias toward any group of student teachers. Meaningfulness refers to the view that assessment tasks should be recognizable and considered

important by the student teachers who are assessed. Transparency implies that it should be clearly communicated to student teachers what they are expected to do regarding the problems and tasks included in the assessment. This means that student teachers should be told not only about which knowledge and skills will be assessed when their classroom-management performance is being evaluated, but also by which criteria and standards their performance will be judged and what steps they can take to improve their performance with regard to classroom management.

Studies into fairness have been part of the validity research agenda of high-stakes teacher assessments for many years. Studies into other aspects of consequential validity, such as the relevance and learning potential of teacher assessment, are rather new and limited because validity research is very time consuming, and therefore, many opt out of research into consequential validity (see Lustick and Sykes, 2006).

Reliability

Reliability refers to the extent to which assessment scores are variable when measurements are repeated under the same conditions or conducted by different independent assessors. Since multiple methods are involved in assessments, which are often collected over time, this form of measurement tends to be highly valid. However, studies into the reliability of teacher assessments reveal that there are serious problems with the consistency of assessment scores (Delandshere and Petrosky, 1994; Moss, 1994). This is due to the fact that teacher assessments contain a lot of qualitative, nonstandardized information, originating from a variety of contexts, which inevitably necessitates judges interpreting the meaning of the information as well as considering the context. Standardization of assessment sources, scoring procedures, and assessment criteria, as well as assessor training in using these criteria independently, are often applied to teacher assessments in order to improve consistency. However, these measures might compromise the validity of the measurement as well as the intended consequences from assessments (Moss, 1994). Nevertheless, in the assessment of prospective teachers, it is important that assessment results are both valid and reliable. This is especially important not only in assessments for summative purposes (i.e., for licensure) but it is also vital for formative purposes (i.e., to guarantee the quality of feedback in relation to assessment standards). In order to improve the validity and reliability of assessment results, it has been argued that assessors should follow an interpretist approach to assessment (Moss *et al.*, 1998). This means that assessors should discuss emerging hypotheses with a peer assessor and that hypotheses about an assessed student teacher should be challenged by the presentation of counter examples. When interpretations are being developed, it is recommended that they are continuously tested, challenged, and revised until they have accounted for all the available

evidence about a student teacher. It is also seen as important to carefully document the interpretation process in order to improve transparency. These principles have been researched in studies on INTASC portfolio ratings by Moss *et al.* (1998) and Schutz and Moss (2004). Their investigations reveal that assessors tend to seek confirmation of their initial interpretations, both when they work individually and when challenged by peers, and that they find it difficult to weigh all available evidence about a candidate into an overall judgment. Following these principles, however, places a heavy burden on assessors in terms of time and effort.

Conclusion

In this article, we have discussed several factors relating to the evaluation of student teachers. Standards have been developed for defining assessment criteria to evaluate prospective teachers, along with various methods and procedures for carrying out these assessments and guaranteeing their quality. It is a well-known fact that assessment determines students' learning activities, and therefore, programs for student teacher assessment, which are well embedded within the pedagogy in teacher education curricula, need to be developed. Procedures for establishing the quality of teacher assessments are quite complex and time consuming, which make them difficult to implement in teacher education institutes. Techniques to promote quality in teacher assessments need, therefore, to be developed further and converted into feasible strategies for all those involved in assessment in teacher education.

See also: Experienced Teachers' Craft Knowledge; Science Teacher Education.

Bibliography

- Airasian, P., Gullickson, A., Hahn, L., and Farland, D. (1995). *Teacher Self-Evaluation: The Literature in Perspective*. Kalamazoo, MI: Center for Research of Educational Accountability and Teacher Evaluation.
- Andrews, T. E. and Barnes, S. (1990). Assessment of teaching. In Houston, W. R. (ed.) *Handbook of Research on Teacher Education*, pp 569–598. New York: Macmillan.
- Boud, D., Cohen, R., and Sampson, J. (1999). Peer learning and assessment. *Assessment and Evaluation in Higher Education* **24**(4), 413–426.
- Carless, D. (2007). Learning-oriented assessment: Conceptual basis and practical implications. *Innovations in Education and Teaching International* **44**(1), 57–66.
- Darling-Hammond, L. and Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education* **16**, 523–545.
- Delandshere, G. and Petrosky, A. R. (1994). Capturing teachers' knowledge: Performance assessment and post-structuralism. *Educational Researcher* **23**(5), 11–18.
- Dierick, S. and Dochy, F. (2001). New lines in edumetrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation* **27**, 307–329.
- Feiman-Nemser, S. and Remillard, J. (1996). Perspectives on learning to teach. In Murray, F. B. (ed.) *The Teacher Educator's Handbook: Building a Knowledge Base for the Preparation of Teachers*, pp 63–91. San Francisco, CA: Jossey-Bass.
- Frederiksen, J. R. and Collins, A. (1989). A systems approach to educational testing. *Educational Researcher* **18**(9), 27–32.
- Haertel, E. H. (1991). New forms of teacher assessment. In Grant, G. E. (ed.) *Review of Research in Education*, vol. 17, pp 3–29. Washington, DC: American Research Association.
- Linn, R. L., Baker, E., and Dunbar, S. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher* **23**(16), 1–21.
- Lustick, D. and Sykes, G. (2006). National board certification as professional development: What are teachers learning? *Education Policy Analysis Archives* **14**(5). Retrieved August, 10, 2006 from <http://epaa.asu.edu/epaa/v14n5/>.
- Mansvelder-Longayroux, D. D., Beijjaard, D., and Verloop, N. (2007). The portfolio as a tool for stimulating reflection by student teachers. *Teaching and Teacher Education* **23**(1), 47–62.
- Messick, S. (1989). Validity. In Linn, R. (ed.) *Educational Measurement*, 3rd edn. American Council on Education. Washington, DC: Macmillan.
- Moss, P. M. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research* **62**(3), 229–258.
- Moss, P. M. (1994). Can there be validity without reliability? *Educational Researcher* **23**(2), 5–12.
- Moss, P. M., Schutz, A. M., and Collins, K. M. (1998). An integrative approach to portfolio evaluation for teacher licensure. *Journal of Personnel Evaluation in Education* **12**(2), 139–161.
- Ross, J. A. and Bruce, C. D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and Teacher Education* **23**(2), 146–159.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* **18**, 119–144.
- Schutz, A. and Moss, P. A. (2004). Reasonable decisions in portfolio assessment: Evaluating complex evidence of teaching. *Education Policy Analysis Archives* **12**(33). Retrieved August 10, 2006 from <http://epaa.asu.edu/epaa/v12n33/>.
- Shulman, L. S., Haertel, E., and Bird, T. (1988). *Toward alternative assessments of teaching. A report of work in progress* (report 4-20-88). Teacher Assessment Project, School of Education. Palo Alto, CA: Stanford University.
- Sluijsmans, D. M. A., Brand-Gruwel, S., and Merriënboer, J. J. G. (2002). Peer assessment training in teacher education. *Assessment and Evaluation in Higher Education* **27**(5), 443–454.
- Uhlenbeck, A. M., Verloop, N., and Beijjaard, D. (2002). Requirements for an assessment procedure for beginning teachers: Implications from recent theories on teaching and assessment. *The Teachers College Record* **104**(2), 242–272.
- van Tartwijk, J., van Driessen, E., van der Vleuten, C., and Stokking, K. M. (2007). Factors influencing the successful introduction of portfolios. *Quality in Higher Education* **13**(1), 69–79.

Further Reading

- Anderson, J. R., Reder, L. M., and Simon, H. A. (1996). Situated learning and education. *Educational Researcher* **25**(4), 5–11.
- Bird, T. (1990). The schoolteacher's portfolio: An essay on possibilities. In Millman, J. and Darling-Hammond, L. E. (eds.) *The New Handbook of Teacher Evaluation. Assessing Elementary and Secondary School teachers*, pp 241–256. Newbury Park, CA: Corwin Press Inc.
- Cronbach, L. J. (1989). Construct validation after thirty years. In Linn, R. J. (ed.) *Intelligence: Measurement, Theory and Public Policy. (Proceedings of a symposium in honor of Lloyd G. Humphreys)*, pp 147–171. Urbana, IL: University of Illinois Press.
- Peterson, K. D. (1995). *Teacher Evaluation: A Comprehensive Guide to New Directions and Practices*. Thousand Oaks, CA: Corwin Press.
- Porter, A. C., Youngs, P., and Odden, A. (2001). Advances in teacher assessments and their uses. In Richardson, V. (ed.) *Handbook of Research on Teaching*, pp 259–297. Washington, DC: American Educational Research Association.

Soled, S. W. (ed.) (1995). *Assessment, Testing and Evaluation in Teacher Education*. Norwood, NJ: Ablex Publishing.

www.pacttpa.org – Performance Assessment for California Teachers (PACT).
www.ccsso.org – The Interstate New Teacher Assessment and Support Consortium (INTASC).
www.nbpts.org – The National Board for Professional Teaching Standards in the United States (NBPTS).

Relevant Websites

www.ets.org – Educational Testing Service (ETS).

TEACHER EDUCATION – MACRO FACTORS

Contents

Diversity in Teacher Education

Social and Cultural Influences on Teacher Education

The Role of Governance in Teacher Education

Diversity in Teacher Education

T Wubbels, Utrecht University, Utrecht, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In their review for the American Educational Research Association (AERA) panel on research and teacher education, Hollins and Guzman (2006), based on research in the US and Canada, cite an earlier conclusion on diversity in teacher education “. . . there has been remarkable consistency in the conclusions of previous reviews: Basic changes in teacher education for diversity are necessary, but have not occurred despite 25 years of attention.” This article looks at the validity of this conclusion reviewing international research available on diversity issues in teacher education, including the composition of the teaching workforce, student teachers’ attitudes and beliefs, teaching competence needed for multicultural classrooms, and the preparation in teacher education programs for this competence. Before turning to these topics we first define diversity.

Diversity

Diversity, generally, is a term stemming from biology and meaning the number and variety of different organisms in the environment in which they naturally occur. In education, Echols and Stader (2002) define diversity as “different racial and ethnic groups, cultures, traditions, and belief systems.” This is a somewhat limited definition because many more characteristics to describe different groups can be used. Menter *et al.* (2006), for example, discussing dimensions for inclusion of teachers in a representative teaching workforce, mention eight dimensions of diversity: social class/socioeconomic status, gender, ethnicity (including refugees, asylum seekers, and travelers), bilingualism, religion, disability, sexuality and sexual orientation, and age. Looking at grounds for discrimination shows that this list of eight characteristics is rather complete. Clarke and Drudy (2006), based on laws prohibiting discrimination in Ireland, mention two additional features: marital and family status,

but we feel it is not necessary to add this to the following list of diversity dimension that often are found in the educational literature: cultural, linguistic, ethnic, racial, social class, socioeconomic status, religion, gender, sexual orientation, age, and disability.

Searches since 2004 with Google Scholar for the adjunct presence of the terms teacher education and several types of diversity, produced the results in March 2009 as presented in **Table 1**.

From **Table 1**, it can be seen that the vast majority of hits concerns cultural, linguistic, and ethnic diversity in teacher education. Looking at the number of empirical studies, a similar distribution is found and often these studies combine issues of cultural, linguistic, ethnic, racial, social class, and socioeconomic diversity. The terms ethnic and cultural diversity have even been used frequently interchangeably and sometimes include even linguistic and racial diversity. The term culture refers to the perspectives (values, worldviews, etc.), practices, and products of a social group, that define how this group interprets and interacts with others. An ethnic group, therefore, is a group of people who identify with one another, or are so identified by others. Thus, ethnicity refers to social groups with a shared history, sense of identity, geography, and cultural roots. The narrowest way to think of ethnicity is by referring to people with origins from different parts of the world. The origin of the group identification, however, may take any of these forms: racial, cultural, linguistic, economic, religious, or political, making clear the overlap between ethnic, cultural, linguistic, social class, and socioeconomic diversity.

Some of the diversity features usually are left untreated in reports and studies on diversity of the (student) teacher candidate population, with the least attention being paid to age and disability. The main part of this article treats the combined racial, ethnic, socioeconomic and social class, cultural, and linguistic diversity. Whenever applicable, attention is paid to gender issues. Before that we discuss

Table 1 Number of hits in Google Scholar in March 2009 for various types of diversity in teacher education since 2004

Type of diversity	Number of hits
Cultural	2930
Linguistic	888
Ethnic	757
Racial	297
Religious	173
Gender	93
Sexual	80
Socio economic	21
Social class	12
Age	8
Disability	5

some aspects of the features of religion, sexual orientation, and disability. For age, no relevant studies have been found.

Disability

Research has shown that persons with disabilities can be successful educators who bring a unique perspective to the construction of pedagogies. Teachers with disabilities can also be valuable role models for students; however, they often face barriers in pursuing professional-level occupations. Gabel (2001), in a study of three student teachers with disabilities, concluded that their experiences with disabilities were an untapped resource with rich potential for the construction of pedagogical knowledge.

Religion

Religious diversity has hardly been studied in teacher education and, generally, it is not considered a problematic issue. Only eight out of 61 student teachers in the study by van Hook (2002) mentioned religion as a highly difficult topic to discuss in classrooms. In an Irish study by Clarke and Drudy (2006), of student teachers' beliefs on diversity, 96% showed openness to non-Christian religions. Another Irish study (Leavy, 2005) showed high levels of tolerance and support by student teachers toward religious diversity.

Sexual Diversity

A study by Sherwin and Jennings (2006) examined the coverage of sexual-orientation topics within 77 public university secondary teacher preparation programs across seven US states. They concluded that 40% of programs did not address sexual orientation as a diversity topic. Most programs that did address sexual orientation did not pay attention to the use of knowledge on sexual diversity in the student teaching practicum.

Composition of (Student) Teacher Body

Several reasons are mentioned in the literature for the need to draw the teaching workforce proportionately from all sections of society. The main reason is what Menter *et al.* (2006) call a commitment for universality in the society at large and in particular with politicians. Education therefore has to follow a public service ethic including the need to have a teaching workforce composed from the same broad range of cultural and social experiences as the students and families served by schools. Thus, schools can draw for the curriculum from the resources offered by all cultural groups. Further, many countries are facing a teacher shortage and therefore recruiting teachers from groups that until now did not contribute teachers is a way to get adequate teacher supply. Finally, Menter *et al.* (2006) claim that for countering the problem of lower school success of minority students than of mainstream students, it is necessary to have teachers from the same minorities in order to offer students adequate role models. This claim, however, is not empirically supported. Zumwalt and Craig (2006) conclude, from a review of studies in Northern America, that the evidence on what impact the match between teacher and student race and ethnicity has on student learning, is mixed. Similarly, the sometimes claimed negative influence of feminization in primary education on development of boys is hardly or not at all empirically justified.

Reviews on the composition of the teaching body usually reveal a mismatch between students' cultural and ethnic background in schools and the background of their teachers (Hollins and Guzman, 2006; Zumwalt and Craig, 2006). All over the world growing diversity of the student population is expected whereas the teaching force is expected to lag behind and stay longer predominantly white, middle class, and female. This, however, appears to be a Western stereotype because in many countries, the teaching force looks very different. In many African countries, for example, the teaching force is black and in some Islamic countries, teachers are predominantly male. **Table 2** gives an overview of the gender distribution in primary education in some countries to show how big the differences across countries are.

What probably is the case all over the world is that teachers come from a limited part of the countries' population and that changes in the student population are not immediately reflected in changes in the teacher population. To illustrate the selectiveness of the teaching force, we present data on the student teacher population in comparison to the country's population, for various countries.

Santoro and Reid (2006) report that while 25% of the Australians have a language background other than English, only 2% of the teachers in the state of Victoria were born outside Australia and for whom English was

Table 2 Percentage of female teachers in primary education in several countries from OECD, 2002 see www.oecd.org/dataoecd/36/56/33703261.XLS retrieved 28 July 2007; OECD, 2005 see www.oecd.org/dataoecd/36/36/35324994.pdf retrieved 28 July 2007; www.moedu.gov.bd/edu_statistics.php, retrieved 28 July 2007; www.statpak.gov.pk/depts/fbs/publications/nec2005/nec01.pdf, retrieved 28 July 2007; www.mceetya.edu.au/mceetya/demand_and_supply_2003,11940.html, retrieved 28 July 2007

Country	Percentage of female teachers
OECD average	80
Unites States	88
Bangladesh	36
Tunisia	50
Pakistan	47
India	37
The Netherlands	80
United Kingdom	82
Australia	78

a second language. Further, the ratio of indigenous to nonindigenous teachers is very imbalanced. While for every 73 indigenous students there is one indigenous teacher, there is, for every 13 nonindigenous students, one nonindigenous teacher.

In the Netherlands, the proportion of minority students in primary education is estimated to be about 19%, whereas the proportion of teachers from ethnic minorities is estimated to be between 1.5% and 3.0%. Based on a report on integration of ethnic minorities in the Dutch society, (SCP, 2005) it is clear this will change: in 2003, about 5% of mainstream Dutch secondary students entered teacher education whereas of ethnic minorities 3% did, thus approaching closer to the proportion in the population.

In Scotland, (Menter *et al.*, 2006) the proportion of teachers from ethnic minorities was estimated to be 0.7%, much lower than in the proportion of students in the population (4.6%).

In the United States, the situation is different. Although Zumwalt and Craig (2006) conclude that in schools in the United States teachers are predominantly white (84%), this is not very different from the country's population in 2000 consisting of about 75% whites, 12% blacks, and 12% others (Hobbs and Stoops, 2002).

We conclude that if one wants the teaching force to be a representation of the composition of the population, then in Western countries the feminization in education is a problem. Further, it appears that diversity in the teaching workforce in several countries lags behind the diversity in the society at large. We emphasize, however, that the wish for a representative teaching force cannot be based on solid evidence about the positive effects of such representativeness on student learning.

Student Teachers

For the match between students and teachers, not only are the formal background characteristics such as gender and race important, but also the attitudes and knowledge of (student) teachers about, for example, culture and diversity. Smith *et al.* (1997) identified four factors that help shape attitudes of entrants in teacher education: exposure to different cultural backgrounds (e.g., friendships, dating, sports), education (e.g., influences of teachers and colleges), travel (e.g., moving, vacationing, and military experience), and personal experience with discrimination as a child or as an adult. We now review what is known about diversity-related attitudes and knowledge of student teachers.

From their international review of the literature, Clarke and Drudy (2006) conclude that many student teachers enter teacher preparation programs with a thin base of knowledge relative to their own and other cultural histories and values systems. Many enter teacher education programs believing strongly in an optimistic individualism, the inevitability of triumph over any obstacle through hard work and individual efforts. The review by Hollins and Guzman (2006) concludes that most American teacher education candidates are white, female, middle class from suburbs or small towns and have limited experience with those from cultures different from their own. Many preservice teachers operate from a limited knowledge base about culture and identity, and have been to white schools in predominantly white neighborhoods. Many American candidates hold negative attitudes and beliefs about people different from themselves. Candidates do not feel adequately prepared for teaching in inner-city schools. These findings about attitudes generally are confirmed in other Western countries such as Ireland (Leavy, 2005; Clarke and Drudy, 2006), the United Kingdom (Holden, 2003), New Zealand (Waghorn and Stevens, 1996), and Australia (Santoro and Allard, 2005). Hagan and McGlynn (2004) found, in Northern Ireland, that 96% of student teachers in their study believed that the ability to address the question of diversity was an important dimension of a teacher's work, but that only 39% felt comfortable in diverse situations.

Lesar *et al.* (2006) report on Slovenian teachers' views on the Romanis and migrant children from former Yugoslavia. Teachers in their study seemed to assume that it is up to the child to adjust to the school system rather than vice versa, while the child's parents are required to prepare the child for the system as well as they can. These teachers did not see themselves as responsible either for the children's school results or their social inclusion. They were not searching for constructive solutions to help children overcome quite serious obstacles to knowledge acquisition. Teachers' views showed few signs of multicultural ideas and very little awareness of how important it is to develop

one's mother tongue and one's own culture. Lesar *et al.* (2006) also concluded that inclusive practices are not only nonexistent in teachers' minds, but also in their work.

One of the consequences of the monocultural experiences and the privileged racial and class status of many white college student teachers is their tendency to see themselves as noncultural and nonethnic beings who are colorblind (do not see or treat people of different color differently) and raceless. Research has indicated, however, that teachers who believe that they are colorblind and treat all students equally, actually privilege mainstream students (Schofield 2001).

Teacher Competence for Teaching Diverse Student Populations

After reviewing attitudes and knowledge of student teachers, we now turn to the teachers' competence to cater to the needs of a diverse student population. Most studies on teacher education candidates mention a lack of confidence in student teachers and even experienced teachers do not feel sufficiently competent in this respect (e.g., Clarke and Drudy, 2006; Holden, 2003). The often-mentioned barriers of student teachers toward a multicultural curriculum (e.g., van Hook 2002) could be a result of this lack of confidence in being equipped with the necessary skills.

Haberman (2004), based on his experiences in inner-city American primary and secondary schools, lists what characteristics of star teachers, in these schools, set them apart: their persistence, physical and emotional stamina, caring relationships with students, commitment to acknowledging and appreciating student effort, willingness to admit mistakes, focus on deep learning, commitment to inclusion, and organization skills. They also protect student learning, translate theory and research into practice, cope with the bureaucracy, create student ownership, engage parents and caregivers as partners in student learning, and support accountability for at-risk students. Undergirding these attributes is an ideology of being neither judgmental nor moralistic, not being easily shocked, recognizing their own weaknesses and compensating for it, viewing themselves as successful professionals rescuing students, and deriving energy and well-being from their interactions with students. These teachers combine attention toward children with attention toward content, recognize the imperative need of student success, and are learners themselves. These attributes predict the effectiveness and staying power of teachers serving diverse students in American low-income urban schools.

Many studies mention competencies teachers need for teaching diverse student populations. Humphrey *et al.* (2006), in a study in primary education in seven European countries, concluded that essential for responding to

diversity in the classroom are caring and inclusive attitudes, striving for inclusive values and solidarity in students, building networks with students and between students, and responsive teaching. Allard (2006) in Australia from a case study on one secondary math and physical education teacher concludes that relationships are critical: students who have experienced the trauma of being a refugee, or the bewilderment of being a migrant, or those whose first language differs from the mainstream, are deeply in need of teachers who will take the time to get to know them, appreciate their strengths, and find ways to instill a sense of hope and confidence through their relationships. Further, being able to critically assess the needs of diverse students – not as problems or as lacking in necessary skills, but as learners who might require different approaches to engage with new knowledge, means that new teachers need to have developed a wide repertoire of critical thinking and teaching skills to call upon.

With regard to promoting student learning, constructivist points of views led several researchers to claim that instruction should be organized so as to build on what students bring from home to school while stretching them beyond the familiar, and to stress the need to examine students' backgrounds. Another attribute very often mentioned, and that seems logically needed but has not been empirically corroborated, is a sense of student teacher's own identity (e.g., Banks, 1991; Zeichner *et al.*, 1998). Banks (2001) adds to this the importance of a balanced view of national and global identities.

Wubbels *et al.* (2006) reviewed literature on the interpersonal competence that teachers need for multicultural classes. They conclude that the following aspects of competence are found frequently:

- having high expectations of students, being interested in students perceptions, and being aware of own cognitions, and behavior;
- creating and maintaining a positive relationship and a safe environment through rewards, compliments, showing respect and confidence, and empathy and providing some freedom and responsibility for students; and
- promoting on-task behavior and creating and maintaining a productive, engaging, learning environment through clear structuring, strong leadership, setting, communicating, and enforcing rules, and probing for student problems and understanding.

They also conclude that many of these competencies are generic aspects of good teaching rather than being specific for multicultural classes. Considering the competencies mentioned earlier in this section, the conclusion of Wubbels *et al.* (2006) probably applies broadly to teacher competence for teaching diverse student populations.

Teacher Education Programs

Hollins and Guzman (2006) summarize results of evaluation studies on the preparation of teachers for diversity. They conclude that most universities in the United States have hardly begun to address issues of diversity. However, a few universities offer carefully crafted programs to prepare candidates to teach students from diverse populations. Studies show that when candidates deal with multicultural issues in the teacher education programs this does not always translate to changes in school practices once they leave these programs. After introducing commonly agreed-upon aims of teacher education for diverse student populations, we discuss several pedagogical approaches advocated.

Aims

In the literature, the aims of teacher education for diverse populations seem undisputed and these are consistent with the competencies mentioned in the previous section. By and large, aims that are often mentioned include (e.g., Hollins and Guzman, 2006; Zeichner *et al.*, 1998):

- reducing prejudice, that is, developing positive attitudes in candidates toward groups and individuals different from themselves;
- strengthening cultural sensitivity (including linguistic issues);
- strengthening multicultural competence including building relationships;
- promoting an attitude of openness to different backgrounds;
- equipping for pedagogy for equity and social justice, that is, developing and using curriculum materials and methods that support the achievement of students from diverse and minority groups;
- equipping for responsive teaching and adaptation to student backgrounds; and
- producing intercultural educational material.

It is also not disputed that these aims include a balanced mixture of knowledge, skills, and attitudes and it is considered important not just to focus on approaches and methodologies but also to establish how teachers' beliefs, attitudes, and dispositions are interwoven with their knowledge, skills, and behaviors in the classroom context.

Pedagogy

Several elements for pedagogical approaches in teacher education for diverse populations have been proposed. First, whatever is done, from a constructivist perspective and based on several empirical studies, it is required to take into account teacher candidates' prior beliefs upon

entry into teacher education programs. Prior beliefs are said to represent interpretive lenses through which student teachers attempt to focus, see, visualize, perceive, characterize, understand, and ultimately resolve their teaching concerns. We now list these approaches and provide data on its effectiveness when available.

Field experiences

Placing student teachers for student teaching in multicultural settings or settings different from the student teachers' background is an often-advocated approach (e.g., Zeichner *et al.*, 1998). It takes different forms depending on the student teachers' background: for some, getting acquainted with other cultures might imply going to a rural setting and for others, to an urban school. More extended field experiences can be created through community-based activities, intensive contacts with parents and pupils from other ethnic groups, and getting access to minority community resources. Hollins and Guzman (2006) conclude that there is some evidence for (short-time) positive effects, for example, on student teachers' beliefs. Combining these placements with community-based cross-cultural immersion experiences and having a structured reflection on experiences seem to be essential for effectiveness.

Critical self-reflection

As mentioned in the previous section, teachers need a clear sense of their own ethnic and cultural identities in order to be able to understand those of their students and their families. For teacher education, this implies offering opportunities to reflect on one's personal history and identity, and one's biases. Instruments used are, among others, writing autobiographies, family histories, and having discussions and confrontations with students from other backgrounds.

Analyses of diversity issues in society

Discourse on diversity issues in society can be analyzed with help of film, poetry, novels, scholarly literature, cases, discussions, role play, writing assignments, and e-mail exchanges. Focus of much of these activities is the reduction of prejudice and promoting awareness of diversity and inequality in society. Hollins and Guzman (2006), from studies on prejudice reduction, found more studies with positive than with no or negative effects; the latter, having in common, that they were all of a short time period or had an interrupted time interval.

Courses

A variety of specific courses are being offered in teacher education programs, for example, on diversity, on bilingual and multicultural education, and on equity pedagogy. Garmon (2004) reviewed empirical studies on the impact of diversity and multicultural courses on prospective teachers' attitudes toward, and beliefs about, different racial

groups. He concluded that the impact of such courses has been mixed. Some researchers have reported that student teachers' racial attitudes and beliefs have been changed in a positive direction by a course on diversity; however, others have reported little or no change in student teachers' attitudes and beliefs. Courses on equity pedagogy aim to equip student teachers for using pedagogies for equity and social justice. Hollins and Guzman (2006) report that generally the result of these courses is not impressive.

Modeling good practice

Providing teacher education in a way that caters to the needs of a diverse student teacher population, that is, showing good practice has been proposed as a powerful intervention. Such practice includes, among others, applying constructivist teaching approaches, being interested in student teachers' backgrounds and perceptions and taking these into account, using mixed student teacher groups, having high expectations of student teachers, creating and maintaining positive relationships and a safe environment, showing respect, confidence, and empathy. Most importantly, teacher educators should be aware of their own cognitions, values, prejudices, and behavior. In sum, the adage practice what you preach should be put into practice.

Methodological Remarks

Looking at the research methods used in the studies reviewed, we must conclude that several methodological limitations hinder general conclusions. Many studies use small samples, outcomes measures are not well developed (often only self-reported outcome data are used), there are few longitudinal studies, whereas there are many short-term and small-scale studies that have little general application, no random samples are used, no (randomized) controlled experiments are applied, and too little information on contexts is provided.

Conclusion and Prospects

The conclusion of Hollins and Guzman (2006) mentioned in the introduction still seems to hold: the student body in teacher education is not a reflection of the student body in schools. The plea for a diverse teaching force, however, is rather a political statement than a claim justified by empirical evidence for the usefulness of a diverse teaching force for improving student learning.

Further, student teachers are not well enough prepared for teaching in multicultural classrooms. No overwhelming success of the advocated pedagogic strategies in teacher education has been shown. Results of studies show mixed results at best. Although some studies suggest

a positive impact of teacher preparation approaches, the findings about preparing teachers for diversity are generally inconsistent and inconclusive. This state of affairs could originate from the interaction between program and student teacher characteristics. Effects may not only depend on program content and pedagogy, but also on types of student teachers participating and characteristics of schools in which they are placed for student teaching practice. For every student teacher type, different interventions might be needed. It is well known that student teachers' beliefs and attitudes serve as filters for subsequent learning. Garmon (2004) mentions a study that showed that student teachers with strong negative biases and stereotypes about diverse groups were less likely to develop beliefs and practices helpful for multicultural sensitivity and responsiveness. Garmon (2004) lists six key factors in student teachers for success in changing their attitudes on diversity: their openness, self-awareness/self-reflectiveness, commitment to social justice, intercultural, educational experiences, and experiences in their support group.

Just elaborating aims or pedagogies of teacher education probably will not be sufficient to strengthen the effects of teacher education programs. Using teacher education programs as tools for improving education for students from different cultural groups needs an institutional approach, including such elements as a clear institutional mission and policies reflecting the values of diversity, applying multicultural criteria for selection of entrants, personnel policies, and an entirely multicultural program content and pedagogy (see Zeichner *et al.*, 1998).

Bibliography

- Allard, A. C. (2006). A bit of a chameleon act: A case study of one teacher's understandings of diversity. *European Journal of Teacher Education* **29**, 319–340.
- Banks, J. (1991). Teaching multicultural literacy to teachers. *Teaching Education* **4**(1), 135–144.
- Banks, J. A. (2001). Citizenship education and diversity implications for teacher education. *Journal of Teacher Education* **52**(1), 5–16.
- Clarke, M. and Drudy, S. (2006). Teaching for diversity, social justice and global awareness. *European Journal of Teacher Education* **29**, 371–386.
- Echols, C. and Stader, D. (2002). Education majors' attitudes about diversity. *Education Leadership Review* **3**(2), 1–7.
- Gabel, S. L. (2001). I wash my face with dirty water: Narratives of disability and pedagogy. *Journal of Teacher Education* **52**(1), 31–47.
- Garmon, M. A. (2004). Changing preservice teachers' attitudes/beliefs about diversity: What are the critical factors. *Journal of Teacher Education* **55**(3), 201–213.
- Haberman, M. (2004). Can star teachers create learning communities. *Educational Leadership* **61**(8), 52–56.
- Hagan, M. and McGlynn, C. (2004). Moving barriers: Promoting learning for diversity in initial teacher education. *Intercultural Education* **15**, 243–252.
- Hobbs, F. and Stoops, N. (2002). Demographic trends in the 20th century. *Census 2000 Special Reports, Series CENSR-4*. Washington, DC: Census Bureau, US Government Printing Office.

- Holden, C. (2003). Education for global citizenship: The knowledge, understanding and motivation of trainee teachers. In Ross, A. (ed.) *A Europe of Many Cultures: Proceedings of the Fifth Conference of the Children's Identity and Citizenship in Europe Thematic Network*, pp 361–367. London: CiCe.
- Hollins, E. R. and Guzman, M. T. (2006). Research on preparing teachers for diverse populations. In Cochran-Smith, M. and Zeichner, K. M. (eds.) *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher education*, pp 477–548. Mahwah, NJ: Lawrence Erlbaum Associates.
- Humphrey, N., Bartolo, P., Ale, P., et al. (2006). Understanding and responding to diversity in the primary classroom: An international study. *European Journal of Teacher Education* **29**, 305–318.
- Leavy, A. (2005). 'When I meet them I talk to them': The challenges of diversity for preservice teacher education. *Irish Educational Studies* **24**, 159–177.
- Lesar, I., Čuk, I., and Peček, M. (2006). How to improve the inclusive orientation of Slovenian primary school – The case of Romani and migrant children from former Yugoslavia. *European Journal of Teacher Education* **29**, 387–399.
- Menter, I., Hartshorn, B., Hextall, I., Howell, I., and Smith, G. (2006). Widening access to the teaching profession: Perspectives from Scotland. *European Journal of Teacher Education* **29**, 271–285.
- Santoro, N. and Allard, A. (2005). (Re)Examining identities: Working with diversity in the pre-service teaching experience. *Teaching and Teacher Education* **21**, 863–873.
- Santoro, N. and Reid, J. (2006). 'All things to all people': Indigenous teachers in the Australian teaching profession. *European Journal of Teacher Education* **29**, 287–303.
- Schofield, J. W. (2001). The colorblind perspective in school: Causes and consequences. In Banks, J. A. and Banks, C. A. M. (eds.) *Multicultural Education: Issues and Perspectives*, 4th edn., pp 327–352. New York: JohnWiley.
- SCP (2005). 'Jaarrapport Integratie 2005. Den Haag: Sociaal en Cultureel Planbureau/Wetenschappelijk Onderzoek- en Documentatiecentrum/Centraal Bureau voor de Statistiek'. (The Hague: Social and Cultural Planning/Research and Documentation/ Central Bureau for Statistics).
- Sherwin, G. and Jennings, T. (2006). Feared, forgotten, or forbidden: Sexual orientation topics in secondary teacher preparation programs in the USA. *Teaching Education* **17**, 207–223.
- Smith, R., Moallem, M., and Sherrill, D. (1997). How preservice teachers think about cultural diversity: A closer look at factors which influence their beliefs towards equality. *Educational Foundations* **11**(2), 41–61.
- Van Hook, C. W. (2002). Preservice teachers' perceived barriers to the implementation of a multicultural curriculum. *Journal of Instructional Psychology* **29**, 254–264.
- Waghorn, A. and Stevens, K. (1996). Communication between theory and practice: How student teachers develop theories of teaching. *Australian Journal of Teacher Education* **21**, 48–55.
- Wubbels, Th., Brok, P. J., den, Veldman, I., and Tartwijk, J. (2006). Teacher interpersonal competence for Dutch secondary multicultural classrooms. *Teachers and Teaching: Theory and Practice* **12**, 407–433.
- Zeichner, K., Grant, C., Gay, G., Gillette, M., Valli, L., et al. (1998). A research informed vision of good practice in multicultural teacher education: Design principles. *Theory Into Practice* **37**(2), 163–171.
- Zumwalt, K. and Craig, E. (2006). Teachers' characteristics: Research on the indicators of quality. In Cochran-Smith, M. and Zeichner, K. M. (eds.) *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education*, pp 157–260. Mahwah, NJ: Lawrence Erlbaum Associates.

Further Reading

- Banks, J. A. (2003). *Handbook of Research on Multicultural Education*, 2nd edn. San Francisco: Josey-Bass.
- Banks, J. A. (2008). *An Introduction to Multicultural Education*, 4th edn. Boston, MA: Allyn and Bacon.

Social and Cultural Influences on Teacher Education

M Ben-Peretz, University of Haifa, Haifa, Israel

R Lotan, Stanford University, Stanford, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

From admission requirements to granting certification and professional licensures and in-service professional development, teacher education is a complex and multi-dimensional process (Wang *et al.*, 2003). Local social and cultural elements, and global policies and economic fluctuations influence and shape the various stages of this process. To describe the relative contributions of these elements to pre-service teacher education programs, the article is organized as follows: program entry requirements and selectivity, description and analysis of some of the major components of teacher education programs, exit requirements, and the attractiveness of teaching as a career in different societies. It ends with a clarification of the tensions between global and local aspects of teacher education.

Conceptual Framework

Globalization, with the integration of multinational economic systems, the spread of technology and electronic networks, and the dissemination of information and knowledge across previously impenetrable national borders and iron curtains, has led to various degrees and forms of institutional and organizational isomorphism. However, in this increasingly global environment, many individuals and groups strive toward the preservation of traditional cultures, the maintenance of locally established social norms, and the continuation of accepted communal arrangements. The inherent tension between globalization and regionalism plays out in different arenas of social and cultural life, including education.

A depiction of the precarious balance and often outright conflict between globalization, on the one hand, and the traditional cultural and social forces of local contexts, on the other, guides our discussion of cultural and social influences on teacher education programs. Increasingly in many countries, the aims, the structures, and the processes of teacher education programs, like much of higher education, reflect mimetic institutional arrangements and comparable curricular content. These are brought about by widespread demand for, and consistent efforts toward, the professionalization of teaching. Common wisdom and tentative empirical evidence assert that the quality of teaching is a key predictor of student learning. In turn,

quality of teaching depends on initial preparation, continued professional development, and the structural conditions of teachers' work.

Since elementary and secondary schooling are becoming mandatory in almost all nations, teachers' work and careers are also subject to public concern worldwide. In response, many governmental agencies and educational programs struggle to find ways to recruit qualified candidates (often in short supply), to provide incentives for improved performance, to build attractive career ladders, and to persuade qualified teachers to stay in remote and poor areas, or in other difficult-to-staff locations. These are some of the global concerns related to teacher education that impact this process in diverse contexts.

Yet, teacher education programs also mirror the local cultural and social order and historical developments as they respond to the way a society defines the purpose of education and consequently the role of the teacher. Programs are influenced by contextual needs and constraints as well as political realities and pressures. They reflect ingrained relationships between community values and educational institutions, and dominant beliefs about the school's role in educating children. The following brief overviews serve to illustrate how historically conditioned social and cultural factors have influenced teacher education in two countries: the United States and Chile.

In the United States, the role of schools has changed with concurrent changes in the economic, social, and cultural circumstances of the nation. During the common school era, teachers (i.e., any individuals with minimal level of literacy and education) were primarily expected to be moral guides whose task was to instill values and civic responsibility in their students. A teacher's comportment and appearance were just as important as her knowledge of content and pedagogy. At that time, the public expressed a strong preference for female teachers who were perceived as better nurturers by nature and therefore better suited for working with children.

As industrialization and urbanization moved more and more workers into factories and masses of immigrants arrived in the country, employers and families realized that different skills were needed. Business leaders expressed their dissatisfaction with the poor quality of workers entering the workforce and began applying pressure on national and local government. Then, as now, corporate leaders became concerned that if children were not well prepared for work, the United States would be at a serious

disadvantage when competing with European or Asian manufacturers. Thus, schools were asked to educate workers as they became places that sorted individuals into future occupations and societal ranks (Cuban, 2003; Labaree, 2004). Schools and thus education became sites where people attended for social mobility and economic opportunity. These developments meant that teachers were valued less for their morals and more for the academic knowledge and vocational skills they could impart.

Historical and political upheavals and major societal transformations in Chile had profound impact on the educational system and its institutions, on teaching as a profession, and thus on the preparation of teachers. Cox describes this in part as follows.

As an authoritarian regime with sweeping powers, Chile's military government (1973–1990) profoundly transformed the administrative and financial model applied to the school system. The military regime decentralized school administration, introduced financing instruments based on subsidized demand, eliminated teachers' status as public employees, and used legal instruments and market incentives to stimulate the creation and growth of state-funded private schools – a “big-bang,” “regime-changing” sort of reform. (Cox, 2006: 3)

With the establishment of three successive democratic governments, a new reform-oriented educational agenda was introduced with “policies focused on the *quality and equity* of the school system, *processes*, and *learning outputs*. It also provided a new national-level statute for teachers” (Cox, 2006, p. 3). From the early 1990s, major ideological differences among majority and opposition parties play out over fundamental topics such as the “regulation of the teaching profession, autonomy versus state regulation of the private sector in education (expressed historically in the Chilean tradition as the opposition between the principles of *libertad de enseñanza*, freedom to educate, versus, *derecho a la educación*, or the right to be educated), the appropriate role of state and market mechanisms, and sexual education” (Cox, 2006, p. 4). As a result, in-service and pre-service teacher education in Chile have captured the attention of all major governmental agencies as well as that of the public. Remarkable monetary and human resources are dedicated to a sweeping reform agenda designed to bring about fundamental changes in the organization of schools and in classroom instruction.

The central argument of this article contends that, presently, globalization is reflected in two major aspects of teacher education programs in many national systems: in the characteristics of the student population that teacher candidates are required to support and to serve, and in the rhetoric of the proclaimed goals of these programs. The process of the realization of the proclaimed goals is a tenuous and at times contentious negotiation

influenced by historically and politically established social and cultural factors.

The greater than ever mobility of populations within and across national boundaries has produced significantly changed student demographics worldwide, causing ethnically, racially, often socioeconomically and linguistically, and therefore academically, heterogeneous classrooms and schools. Long-standing global issues, such as poverty, environment, conflict, and social justice, constantly take on new forms, whether as “war on terror,” or “going green.” Holden and Hicks (2007) state that “(t)hese issues affect the society that students live in now and in the future” (p. 13) and that “(t)he world of the early 21st century is complex and fast changing. Local and national issues, events and trends can only be understood if set in the wider global context” (p. 23).

The impact of global influences on teacher education in Scotland, for example, is treated by Hartley (2002). Hartley argues that globalization has three competing influences on teacher education: “those derived from economic globalization; those derived from the requirements of a new knowledge economy; and third, those emanating from the culture of difference (and inequality) which now typifies many developed societies” (p. 251). According to Hartley, the issues of cultural differences and inequality have to be treated in the local context.

Increasingly, in many globally connected countries, established rhetoric regarding the outcomes of teacher education converges around the view that teacher preparation institutions and programs need to produce professionals who can effectively educate heterogeneous student populations in ever-changing contexts while upholding high academic standards and rigorous curricula. As a direct outcome of the quality of teaching, students are expected to become global citizens, who value diversity across and within national borders. They are expected to become productive contributors to the twenty-first-century knowledge economy, successfully mastering vast bodies of information as well as being fluid, critical, and creative thinkers. Today, “teachers need to view themselves as public intellectuals who combine conception and implementation, thinking and practice, with a political project grounded in the struggle for a culture of liberation and justice” (Niemi, 2000: 16).

Entry Requirements and Standards

Globally, post-secondary education has been assuming an increased role in the preparation of teachers. Thus, a 4-year bachelor's or undergraduate degree is typically the standard entry requirement for most systems. In the United States, all teachers are required to hold a bachelor's degree. A 5-year program for both elementary

and secondary teachers is the standard in Thailand (Ingersoll, 2007) and Chile (Cox, 2004), among many other countries. Since 1971, all teacher education in Finland takes place in universities where programs last 4–5 years leading to an MA degree. Completion of these programs is considered the basic qualification for both elementary and secondary teachers in Finland, in order “to unify elementary and secondary school teacher education and to develop an academically high standard education. It was also necessary to answer the needs of a new comprehensive school system” (Niemi, 2000, p. 1).

Other systems, mainly in developing countries, vary both within and across the level and years of education required for teaching at the elementary or at the secondary level. In China and Singapore, the education required of elementary teachers is lower than that of secondary teachers, although in both these systems there is movement to bring elementary teachers on par with secondary teachers. For instance, in China a high school diploma is the minimum level necessary to enter elementary teaching, while upper secondary teachers must have a 4-year college degree. In Singapore, as in Hong Kong, an elementary-level teacher can gain entry with the equivalent of a 2-year sublevel or associate level, while at the secondary level, Singapore requires teachers to complete a 5-year program that includes a bachelor’s degree plus a year of further coursework (Ingersoll, 2007). Local cultural conditions may also determine the length of study required for a teaching certificate. For example, in Israel’s ultra-orthodox community, 2 years of study are sufficient for gaining a teaching certificate, compared to 3 or 4 years for other student teachers who are required to earn at least a BA or BEd. Teachers who do not have these qualifications are prevented from teaching in the secular school system. This is one of the ways in which the ultra-orthodox community upholds social and cultural boundaries closing itself off from mainstream Israeli society.

The mandate to higher education to contribute to and ultimately take responsibility for the preparation of the teaching force reflects progressive societal aspirations and demands for the professionalization of teaching and its elevation to parallel the preparation for the practice of law, medicine, or engineering. This relatively new development reflects a global tendency related to teacher education, as teachers used to lack university degrees in most countries.

The differences between teachers’ entry requirements for elementary and high schools may be related to local factors. In a vast country like China, requirements might be less stringent for elementary teachers because of the very high numbers of children attending elementary schools or lack of access to higher education institutions in geographically remote areas. The fact that Singapore, a small country, also requires less years of university studies in order to teach at the elementary level may be due to

how Singapore views the educational needs of very young children.

In the US, the requirements to become an elementary or secondary level teacher have never been considered especially rigorous. Those entering teaching in the US tend to be in the average to low range academically compared with other college graduates. To facilitate entry, the United States early in the last century created large numbers of low-cost, dispersed, and noncompetitive teacher-training institutions and colleges.

In Singapore and Korea, as in many other Asian countries, teacher-education students are of high academic achievement. These differences may be due to both social and cultural factors and are related to the status of teachers and teaching in the different contexts.

Structural Organization of Teacher Education Programs

Within national systems, there is variability in the sequencing and organization of professional preparation programs. In some programs, educational and professional preparations occur concurrently within a bachelor degree program. In other programs the degree is completed first, with subsequent professional preparation and certification. In the United States, the latter is often called a “fifth-year program” – a post-baccalaureate 1-year teacher preparation program leading to a teaching certificate and sometimes, a master’s degree. Several systems, including in the United States, also have alternative pathways into teaching. In these cases, candidates are not required to have completed professional preparation prior to employment and uncertified teachers can begin teaching before or during their preparation and training. Local shortages of teachers are one cause for such alternative routes into the profession; uncertainty concerning the added value of existing programs of teacher education might be another.

In England, recent changes in teacher education reflect the strong impact of society on the educational process

(The) changing view of education is clearly reflected in the changing name of the government department responsible: in 1992 the DES (Department of Education and Science) became the DFE (Department for Education); in 1995 that was transmitted into the DfEE (Department for Education and Skills). In these movements there is a significant shift, ideologically, from an intrinsic valuing of education as a good in itself to the extrinsic and instrument role it is now seen as playing. Its link, from 1995, with employment has inevitably meant a tighter focus on the training and skills essential for the world of work. (Halstead, 2003: 69).

The Curriculum of Teacher Education Programs

The curriculum of many teacher education programs includes well-established strands: method courses (that often assume mastery of subject matter content by the candidates), educational psychology that includes components of learning theory and child and adolescent development, philosophy of educational thought, history of educational systems and institutions, and sociology of education, evaluation, and policy studies. A separate strand deals with the clinical experiences of candidates in the field. Increasingly, however, teacher education courses aim at helping prospective teachers learn about local social and cultural conditions that impact students' learning and therefore need to be taken into account in teaching. An analysis of six teacher education programs in the USA and four in Israel found that in addition to obtaining a strong theoretical framework that incorporates knowledge of subject matter, educational philosophy, sociology, pedagogy, method courses and practicum, students of these programs are also introduced to the conditions in today's urban education with its diversity and dynamic change (Ben-Peretz and Landler-Pardo, 2007). For example, at Mills College in Oakland, California and at the University of Haifa in Israel (both cities with highly diverse populations), there is a strong emphasis on preparing teachers to understand the implications of social, cultural, and economic diversity and to teach toward equity. Courses, such as "Building structures for equity, excellence, and access," and dealing with the question of "how can we teach each child well in a racist society?" are being offered at Mills College. At the University of Haifa, awareness of the challenges of the rapidly changing conditions of modern societies is expressed in topics such as the "Multicultural curriculum" or "Acclimation problems." Other examples of courses dealing with the difficulties and challenges within teaching are: "Teaching linguistics and cultural minority students" at the University of California in Berkeley, California, "Principles of teaching exceptional individuals," at the University of Los Angeles, California, "Elements of diversity: English language learners" at Harvard University, Massachusetts, and "Teaching heterogeneous classes: adolescence – a challenge for the teacher" or "Management and discipline in heterogeneous classes" both at the Hebrew University, in Jerusalem, Israel. As a reflection of the program's mission and in response to state and national standards for teacher education programs, the Stanford Teacher Education Program, at Stanford University, in California offers courses entitled: "Teaching and learning in heterogeneous classrooms," "Language policies and practices," and "Supporting students with special needs." Syllabi of

some of these courses reveal that the local cultural and social conditions of children's and adolescents' lives are considered important factors in their academic development.

In England, as in South Africa, after apartheid, teacher education attempts to fulfill a role in relation to education for democratic citizenship (Harber and Serf, 2006). Several venues may lead to that goal. Student teachers may be provided with experiences of operating democratically in their courses. Harber and Serf (2006) claim that students in initial training "need to engage with key forms of inequality that challenge the development of a more democratic society – racism, sexism, homophobia, disability discrimination, etc" (p. 991). Moletsane (1998) found that in South Africa "racial and cultural tolerance was promoted on the course and teaching and learning methods encouraged and affirmed a plurality of opinions from the diverse group of students in the class. Classes provided an opportunity for self-expression and open debate by students that helped them to find their own voice in a safe and flexible classroom environment" (in Harber and Serf, 2006, p. 992).

In Germany a conference of ministers of education of the different German states published a special paper calling for an intercultural emphasis in teacher education. In order to prepare future teachers for cultural diversity, the paper recommends that issues of diversity be treated in all courses. These recommendations reflect the impact of global concerns in a local context (Beschluss der Kultusministerkonferenz, Germany, 1996).

At times, preparing student teachers for the social and cultural context of schools reflects elements of local religion, for instance, in the University of Madrid in Spain. At Acatlan University in Mexico, a number of courses in the teacher education program focus on social and cultural issues. Such is a course on problems in education in Mexico in relation to the economic, social, political, and cultural contexts, which is entitled "Problemas actuales de la educacion en Mexico."

A further example of the impact of social and cultural factors on teacher education programs concerns the power of Marxist ideology and implementation problems to inhibit the spread of a "new education movement" in the teacher education programs in the Soviet Union. The new education movement was "part of more general education reformism sweeping Europe and America in the first two decades of the twentieth century" (Partlett, 2007, p. 455). The leader of this movement in the first years of the communist regime in the Soviet Union was Shatskii, who relied on Western scholars like Dewey. Although Shatskii tried to integrate communist ideology with this movement, he was later marginalized within the Soviet educational system. When the Communist party asserted control over education in 1932, many of the new ideas

were rejected. This case reflects the tension between global ideas and local contexts.

Attractiveness of Careers

Compared with other professions, teaching has been considered a less attractive and less desirable line of work in the United States. In surveys of occupational prestige, teachers rank in the middle range. As other historically female-dominated occupations, teaching tends to have lower status and lower pay.

Many Asian nations, however, have a tradition of respect for teachers (Ingersoll, 2007). Relatively high salaries, comprehensive training, and full pay while undergoing training all make teaching an attractive career option in Singapore. In Hong Kong, teaching is ranked relatively high in occupational stature by senior-secondary school students – above accountants, engineers, doctors, and artists. Still, the quality of new entrants to teaching has been a matter of concern in Hong Kong.

In China too, teachers rank relatively high in surveys of occupational prestige. Yet, because teachers' salaries are low in China relative to other occupations, and especially low in rural areas, the occupation is not as attractive as some others (Gang and Meilu, 2007). In Korea, teaching is a relatively sought-after occupation because of its job security and its high social-status standing (Kim, 2007). The rate of teacher turnover is very low because most teachers remain in teaching until the point of retirement. In Japan, teaching profession is an attractive option for college students, is relatively well paid, enjoys respect, job autonomy, and a collaborative community with colleagues and affords chances to grow and develop as educators. Not surprisingly, turnover and quitting rates have traditionally been low (Fujita and Dawson, 2007). Thus, cultural and social differences in the career attractiveness of teaching are significantly related to entry conditions and teacher turnover.

Certification and Licensing

Most educational systems require candidates to obtain a government-issued certificate of license signifying that a candidate has completed required professional preparation and training. In most systems, professional preparation also includes a period of varying length in the field through supervised clinical practice or student teaching. Many systems require some kind of performance exam or paper-and-pencil test for prospective teachers. In some systems, these tests are administered as part of a teacher-preparation program; in others,

they are administered by the school upon employment or by the state. These commonalities reflect the global impact on the increased trend toward the professionalization of teaching.

Conclusion: The Tension between Global Concerns and Local Beliefs

Our analysis has shown the existence of several salient commonalities among teacher education programs in different countries, societies, and cultures. These commonalities include: the growing demand for academic degrees for teachers; the tendency to expand years of preparation; and the emphasis on professional competencies to be fostered in special institutions as well as through the practicum – the growing tendency to determine standards for qualifications of teachers as well as for the requirements to gain a teaching certificate. These commonalities may be interpreted as reflecting the effect of globalization, with its emphasis on the role of teachers in preparing citizens for the future workforce, and for living in a global world. Another factor leading to these common features is the growing view of teaching as one of the professions having a common core knowledge base, and modes of instruction. On the other hand, it seems that teacher education programs are also adapted to local contexts, be they political, economic, or cultural. Thus, teacher education in England reflects the pressure of greater central control, while teacher education in Spain, Mexico, or ultra-orthodox communities in Israel adapts itself to the religious context. It is important to note, however, that even global ideas and frameworks are not passively employed in local contexts. An editorial of *Teaching and Teacher Education* (vol. 15, 1999) deals specifically with global and local dimensions of reforms in teacher education (Elliott, 1999). This special edition brings accounts from developing countries in Africa, South America, and the Far East. Elliott argues that:

Teacher education reforms are not like machines which, providing operators follow instructions, can be made to work in exactly the same way regardless of context. Of course, the imperatives which stem from a world of markets will encourage the borrowing of concepts and models of teaching and teacher development, from those countries who appear to succeed in meeting such imperatives. Ideas are now globally mobilized and their trajectory tends to be from the north and the west to the south and the east. In the process they may be resisted, critiqued or reinterpreted/reconstructed as a basis for conceptualizing reforms in particular local/national contexts. (Elliott, 1999: 134).

The Townsend and Bates (2007) *Handbook of Teacher Education* devotes a whole section to “Globalization and Diversity: Promise or Problem?”. The section includes several chapters that focus on learning in a globalized world and the interplay between culture and teachers’ professional development. Certainly, the diversity of most communities in many parts of the world has made teaching and educating teachers much more difficult than it has ever before. Clearly, it is difficult to “establish a strong all-inclusive education system, based on the best ideas from other parts of the world but still maintaining the cultural integrity of the people.” (p. 8)

There is a growing awareness in developing countries that though it is important to introduce new ideas into teaching and teacher education programs, these have to be integrated with local beliefs and traditions. The fear that converging customs, norms, and ideas, will finally overrun diversity, and will serve to create a homogenous world society has brought about the formation of societies devoted to the preservation of local cultures, such as “Cultural Education and Cultural Sustainability: An International Network of Diaspora, Indigenous and Minority Education.” Moreover, it is not self-evident that the flow of educational ideas is necessarily only from west to east and from north to south. It may well be that in the future innovations in education will arise in diverse cultural contexts, and will spread from these in other countries like the strategy of “lesson study” from Japan to the United States. The potential tensions between global and local educational orientations may also yield new insights into global ideas implemented in local contexts.

Bibliography

- Ben-Peretz and Landler-Pardo, G. (2007). Teachers’ thoughts about rewards: A neglected component in teacher education programs. In Craig, C. J. and Deretchin, L. F. (eds.) *American Teachers Educators’ Yearbook XVI: Imagining a Renaissance in Teacher Education*. MD: Scarecrow Education.
- Beschluss der Kultusministerkonferenz (1996). *Empfehlung “Interkulturelle Bildung und Erziehung in der Schule”* Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland, Germany.
- Cox, C. (2006). Policy formation and implementation in secondary education reform: The case of Chile at the turn of the century. *Education Working Paper Series. Number 3*, Washington, DC: World Bank, Education Advisory Service.
- Cuban, L. (2003). *Why Is It so Hard to Get Good Schools?* New York: Teachers College Press.
- Elliott, J. (1999). Introduction: Global and local dimensions of reforms in teacher education. *Teaching and Teacher Education* **15**, 133–141.
- Fujita, H. and Dawson, W. D. (2007). The qualifications of the teaching force in Japan. In Ingersoll, R. (ed.) *A Comparative Study of Teacher Preparation and Qualifications in Six Nations*, pp 41–54. Philadelphia, PA: Consortium for Policy Research in Education.

- Gang, D. and Meilu, S. (2007). The qualifications of the teaching force in China. In Ingersoll, R. (ed.) *A Comparative Study of Teacher Preparation and Qualifications in Six Nations*, pp 19–27. Philadelphia, PA: Consortium for Policy Research in Education.
- Halstead, V. (2003). Teacher education in England: Analyzing change through scenario thinking. *European Journal of Teacher Education* **26**(1), 63–75.
- Harber, C. and Serf, J. (2006). Teacher education for a democratic society in England and South Africa. *Teaching and Teacher Education* **22**, 986–997.
- Hartley, D. (2002). Global influences on teacher education in Scotland. *Journal of Education for Teaching* **28**(3), 251–255.
- Holden, C. and Hicks, D. (2007). Making global connections: The knowledge, understanding and motivation of trainee teachers. *Teaching and Teacher Education* **23**, 13–23.
- Ingersoll, R. (ed.) (2007). *A Comparative Study of Teacher Preparation and Qualifications in Six Nations*. Philadelphia, PA: Consortium for Policy Research in Education.
- Kim, E. (2007). The quality and qualifications of the teaching force in the Republic of Korea. In Ingersoll, R. (ed.) *A Comparative Study of Teacher Preparation and Qualifications in Six Nations*, pp 55–70. Philadelphia, PA: Consortium for Policy Research in Education.
- Labaree, D. (2004). *The trouble with Ed schools*. New Haven: Yale University Press.
- Moletsane, R. (1998). Towards democratic teacher education in South Africa: An exploratory case study. In Harber, C. (ed.) *Voices for Democracy*. Nottingham: Education Now in association with the British Council.
- Niemi, H. (2000). Teacher education in Finland: Current trends and future scenarios. In *Teacher Education Policies in the European Union: Proceeding of the Conference on Teacher Education Policies in the European Union, and Quality of Lifelong Learning*. Loule, 22 and 23 May 2000. Portuguese Presidency of the Council of the European Union – Ministry of Education. Lisbon. The status and standard of Finnish science, 1997, Study of a culture and society.
- Partlett, W. (2006). Bourgeois ideas in communist construction: The development of Stanislav Shatskii’s teacher training methods. *History of Education* **35**(4–5), 453–474.
- Townsend, T. and Bates, R. (2007). Teacher education in a new millennium: Pressures and possibilities. In Townsend, T. and Bates, R. (eds.) *Handbook of Teacher Education: Globalization, Standards and Professionalism in Times of Change*, pp 3–22. Netherlands: Springer.
- Wang, A., Coleman, A., Coley, R., and Phelps, R. (2003). *Preparing Teachers around the World*. Princeton, NJ: Educational Testing Services.

Further Reading

- Al-Hinai, A. (2007). The interplay between culture, teacher professionalism and teachers’ professional development at times of change. In Townsend, T. and Bates, R. (eds.) *Handbook of Teacher Education: Globalization, Standards and Professionalism in Times of Change*, pp 41–52. Netherlands: Springer.
- Carnoy, M. and Castells, M. (2001). Globalization, the knowledge society, and the network state: Poulantzas at the millennium. *Global Networks* **1**(1), 1–18.
- Cochran-Smith, M. (2004). Taking stock in 2004: Teacher education in dangerous times. *Journal of Teacher Education* **55**, 3–7.
- Day, C. (1997). In-service teacher education in Europe: Conditions and themes for development in the 21st century. *British Journal of In-Service Education* **23**(1), 39–54.
- Gutek, G. L. (2006). *Education in a Global Society: International and Comparative Perspectives*. Long Grove: Waveland Press.
- Irvine, J. J. (2003). *Educating Teachers for Diversity: Seeing with a Cultural Eye*. New York: Teachers College Press.
- Jansen, J. (2007). Learning and leading in a globalized world: The lessons of South Africa. In Townsend, T. and Bates, R. (eds.)

- Handbook of Teacher Education: Globalization, Standards and Professionalism in Times of Change*, pp 25–40. Netherlands: Springer.
- OECD (2003). Attracting, developing and retaining effective teachers. OECD Activity: Country Background Report for Chile. Government of Chile, Ministry of Education, Planning and Budget Division, Department of Research and Statistics.
- Robbins, M., Francis, L., and Elliott, E. (2003). Attitudes toward education for global citizenship among trainee teachers. *Research in Education* **69**, 93–98.
- Sato, N. and McLaughlin, M. (1998). Context matters: Teaching elementary school in Japan and the United States. In Rohlen, T. and Bjork, C. (eds.) *Education and Training in Japan Volume II*, pp 150–164. London: Routledge.
- Sleeter, C. (2001). Preparing teachers for culturally diverse schools: Research and the overwhelming presence of whiteness. *Journal of Teacher Education* **52**, 94–106.

The Role of Governance in Teacher Education

G E Karlsen, South Trondelag University College, Trondheim, Norway

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The phenomenon of governance and government is normally linked to the execution of power with a legal and legitimate basis. In social science literature, government is the traditional term for steering and control. Typically, the government refers to the state, understood as the political and administrative system at the national level and in particular the executive function of the state (Ball, 1990; Dale, 1989; Kuhlne *et al.*, 1999).

The Foucauldian concept of governmentality (Foucault, 2003) has been a guiding concept for recent education policy research with focus on mentalities and rationality of governing of the state and changes over time. Governance is a more versatile term in social sciences and research corresponding to the postmodern form of economics and political actors. In particular, there is a great deal of research about European governance related to the European Union (EU; Olsen, 2002; Kohler-Koch, 2005). In the EU's white paper *European Governance*, the concept of governance is defined as "rules, processes and behaviour that affect the way in which powers are exercised." Openness, participation, accountability, effectiveness, and coherence are considered as principles that should underpin what is called "good governance" (European Commission, 2001: 11, 17).

In this article, governance is used in a neutral meaning. Governance includes macro-actors and processes and relationships to and between states, regional, and local actors and markets in education policy (Lawn and Lingard, 2002). The governance in relation to teacher education also includes informal steering from nongovernmental bodies in the civil society and influence from the institutional culture in teacher education institutions (Lindblad and Popkewitz, 2001).

Teacher education has a high degree of complexity and a great variety in time length, content, organization, and the degree of specialization. In fact, there are more structures than countries if we consider, for instance, Europe (Eurydice, 2002; Garm and Karlsen, 2004). If teacher education is considered to be one educational system – as a unit distinct from others – with common, distinctive traits, it is a nonexistent phenomenon in the real world. What do exist, however, are the continuous political and professional discourses regarding knowledge, attitudes, behaviors, and skills necessary for the teacher profession and the institutions for teacher education.

Historically, the structure, content, and ideology of teacher education institutions mirror the present political

conditions. The governance of teacher education has normally been a national issue. National teacher education and the compulsory school system are normally interlinked (Karlsen and Kvalbein, 2003). They have been and still are key institutions for socialization into the modern state. The impressive research report of the AERA Panel on Research and Teacher Education (Cochran-Smith and Zeichner, 2005) summarize: "It is now widely agreed that teachers are among the most, if not the most, significant factors in children's learning and the linchpins in educational reforms of all kinds." Therefore, the importance of teacher education is obvious in many ways.

The education system and, in particular, teacher education has historically been closely connected with nation building (Vaage, 2001). Education is a part of rights and obligations in a complex relationship between the state and the individual citizen. This is still the normal situation; however, the traditional role of national governance is being challenged by new procedures and new global actors. The territorial representative parliamentary democracy, when limited to national state agents with agreements about rules within the structure of the national state, is being challenged by globalization. Bauman (1998) claims that this threatens to make national democracy irrelevant and powerless. With this background in mind, a simple model will function as an analytical framework for the further discussion of the role of governance in teacher education.

An Analytical Model – Levels and Systems

Although governance of teacher education primarily has a national mission, teacher education and its reforms are embedded in an increasingly global context. Therefore, the role of governance in teacher education has to be analyzed as an interactive process between supranational, national, and organizational (local) levels.

The model is a simplification of the complexity of educational governance. Teacher education includes both the professional education system at the national level and institutions at the local level. The model shows interrelations of main actors and forces affecting governance of teacher education at and within the national level and the globalization pressure from actors at the supranational level (Figure 1).

Traditionally, the nation-state has been associated with a specific geographical area, political entity, and, in addition, a cultural and ethnical unit. Political science

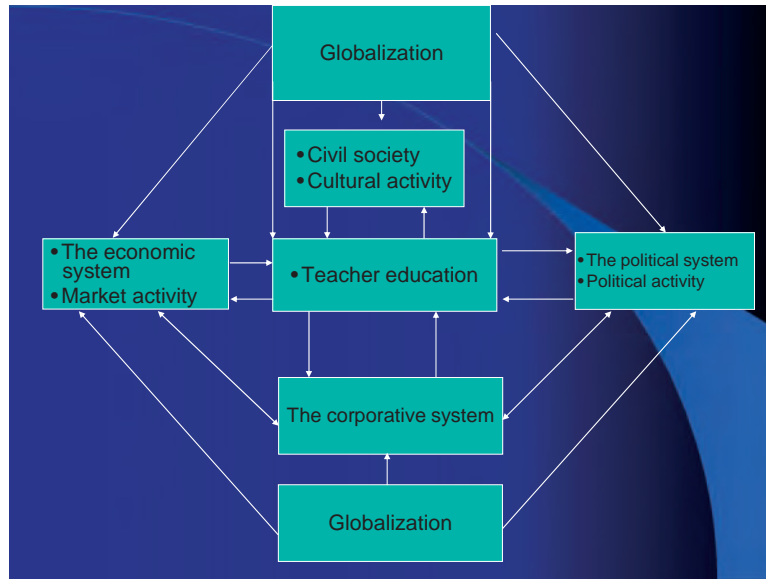


Figure 1 Governance of teacher education—actors and forces.

normally emphasizes the existence of a political and administrative system having internal and external sovereignty over a definite territory (Jones *et al.*, 2004). The central base for the national state is the legal framework accompanied by a set of related political, judicial, and administrative state institutions such as the parliament, state bureaucracy, courts, and police. The school system, and in particular the teacher education, is a key institution in the state.

The market, understood as rational decision making, is basically a social arrangement or a marketplace that allows buyers and sellers to discover information and carry out voluntary exchange of goods and services. Market activity within the economic system is primarily the arena of capital connected to economic transactions. The relation between the political and the civil system and the market will vary, but the move toward the market economy has been obvious (Lavigne, 1995). Although the free-market economy is prospering, the mixed economy is still common in most of the countries, with degrees of government influence in combination of state-owned and private enterprises.

The concept of civil society has had a renaissance both in political and comparative theory. Civil society normally refers to the arena of collective company and actions linked to shared interests, purposes, and values (Bron, 2001). The boundaries between state, civil society, and market are often complex and blurred. The civil society is composed of the totality of voluntary civic and social organizations and institutions that to a high degree form the basis of a functioning society (Hann and Dunn, 2006). The civil society provides rights connected to democracy, political consultation, and the development of life, within culture and religion, in or outside

institutional and organizational frames. The civil society can be understood as the glue of society. In this light, the formation of state and citizenship provides a framework for civil rights.

The growth of the organizational society, particularly in the nineteenth century, contributed to the development of new cooperative structures, creating new channels for influence. Labor unions promoted right to collective bargaining and tariff agreement between their members and the employers. Within the Nordic states system, for instance, negotiations between representatives of unions and employers' organizations gradually became a formalized decision-making channel in addition and, to some extent, as an alternative to the representative democracy (Dølvik and Engelstad, 2003). The political scientist Stein Rokkan (1987) has called this regime of governance "corporate pluralism" (p. 61). In relation to education, this became important because teachers created strong unions. The unions provided corporative power connected to participation in public committees, elected groups, public hearings, and, not the least, contracts about working conditions negotiated between the state and the unions. At the European level, the European Trade Union Confederation (ETUCE) is an influential teacher union with a total of 60 million members. The ETUCE is the only representative cross-sectoral trade union organization at a European level recognized by the EU (ETUCE, 2006).

The overall political neoliberal governance turn in the last 20 years has put both the political system and the civil society under pressure. New global ideologies, new strategies and new macro-actors are brought into force. Therefore, the governance in teacher education must be analyzed in this context starting with globalization and the driving forces behind it.

Globalization and Global Forces

Globalization is normally used as an umbrella term which refers to a complexity of economic, technological, cultural, social, and political interrelationships. The notion is frequently used both as a slogan in public debate and in scientific literature (Spybey, 1996; Held and McGrew, 2000). The term globalization term can be seen as a perception about a pluralistic, worldwide postmodern period after industrialism (Giddens, 1990). Although globalization is not completely a new phenomenon, it has become the dominating contemporary diagnosis in the last twentieth and the early twenty-first centuries.

There is considerable agreement among researchers that the economy is a driving force (Coulby and Zambeta, 2005). Globalization may primarily refer to a process through which national economies became increasingly open to supranational influence (Lundahl, 2005). At the core of globalization lays the new market economy with its market capitalism and worldwide economic activity. Globalization is further interpreted as a consequence of modern science and new technology, particularly in connection with information technology.

Globalization has led to a flood of influences economically, environmentally, politically, socially, and culturally, which are almost impossible to protect oneself against (Mishra, 1999; Martin and Schumann, 1996). The main hypothesis is that we are forced toward standardization not only economically, but also socially and culturally. A UNESCO report, led by the former EU President Jacques Delors, proclaimed a development for this century “towards a globalization of human activity” (Delors, 1996: 41).

Globalization is a phenomenon with consequences and influences – good and bad – also for teacher education. In particular, the themes of central control, the role of the state and teacher education as a cultural reproductive institution, create dilemmas and prospects (Thomas, 2002). The discussion surrounding globalization has essentially been linked to whether this is a desired or undesired development. On the one hand, it is claimed that globalization is necessary, inevitable, and the only possible way to get a larger degree of political and economic stability and predictability (Giddens, 2000). On the other hand, globalization is seen as a serious threat toward established order and common values. The development of antiglobalization forces has grown (Hill, 2006). The political left wing is the most obvious in alliance with environmentalists and labor unions, included teachers unions. The Education International (EI) is the world's largest, and probably the most important, teacher union with its 345 member organizations in 165 countries and territories. The union is critical of the impact of globalization on education and perceives the neoliberal globalization as a significant risk affecting public education, restricting the available policy space for national governments (Hill, 2006).

The acronym TINA (there is no alternative) refers to the slogan attributed to the former prime minister of Britain, Margaret Thatcher, who argued that globalization with its free trade and global economy was the only way for modern states. Others have argued that the global market economy, which is adopted in most of the states worldwide, is a neoliberal political agenda rather than an inevitable choice (Negreponi-Delivanis, 2001). Therefore, we need to examine this agenda in a governance perspective.

A Neoliberal Ideology and New Governance Strategy

The overall neoliberal turn can be perceived as a renewed ideology and a new governance strategy also for governance in teacher education. It appeared on the horizon in the 1980s and 1990s with its belief in the market as an agent for economic growth. The perception was that public sector, including education, was unproductive and unwilling to change. The new idea was that societies guided by the market would stimulate production, revitalize the civic society, and even strengthen democracy by individual choices and consumer power (OECD, 2000).

The governance turn was ideologically rooted in liberalization and, to some degree, conservatism as political ideologies with focus on less government regulations and greater participation of private actors (Beck, 2005). New public management (NPM) became the broad term used to describe the wave of public sector reforms. The main hypothesis was that institutions guided by market rules would reduce costs, be more efficient, and even lead to higher quality, without having negative side effects (Pollitt and Bouckaert, 2000). NPM, with its market orientation and its blithe acceptance of competition as a driving force, put public sector under pressure.

NPM emphasized individualism and freedom of choice. It also assumed that decentralization, deregulation, and restructuring would improve both utilization of resources and higher quality of service. In this perspective, education, and also teacher education, was perceived as a service sector among others. New governance reforms were needed. The Management by Objectives (MbO) model, known from legendary Peter F. Drucker (1964, 1977), and originally developed as a strategy for more cost-efficient production and leadership in industry and business, became a suitable answer also for governance of public education, including teacher education.

Decentralized Centralism – A New Dynamic in Educational Governance

The NPM ideology and the MbO governance strategy involved a new governance dynamic and changed use of

political tools. The relation between the market and the political system changed in favor of the market, however, at the same time, we can observe variation. States do not always passively adapt to the external, international constraints (Sassen, 1998). The earlier works have used the term decentralized centralism trying to catch the governance dynamic between centralization and decentralization processes (Karlsen, 2000).

As a governance strategy, the MbO combined centralization and decentralization in dynamic interplay at the same time. Setting central goals and standards for outcomes were tasks for the macro-level and therefore assigned to centralization, while choice of tools and the responsibility for implementation were duties at the microlevel and seen as decentralization. The MbO strategy was goal and outcome oriented, in which decentralization was fitted into a centralized strategy. The model gave local freedom and probably more acceptance to work under harder pressure. In fact, this was management by results, and the rise of what Fägerlind and Strömquist (2004) have called the evaluative state.

In teacher education, the trend to decentralize responsibility and deregulate state schooling was tied to a rhetoric of enhancing teacher autonomy and professionalism and stressed the importance of the teachers' expert knowledge (Karlsen and Kvalbein, 2003). An opposite centralized trend was the state's intensification of instruments for output control. Thus, increased output control might instead restrict teacher autonomy and contribute to deprofessionalization of teachers' work (Helgøy and Homme, 2004). Delandshere and Arens (2001) interpret new teacher education reforms in the US as decreasing teachers' degree of freedom to shape their own role as teachers. Teachers and teacher educators are perceived more as the object of policy rather than professional participants. However, good teacher education and good teachers have been politically seen as the backbone in any national education system (OECD, 2005). From the state perspective, teacher education is under double governance at the same time. Teacher education is perceived as a governance instrument for desirable educational change in the compulsory school system (Garm and Karlsen, 2004). At the same time, the authorities can use policy tools as governmental instruments for regulation of teacher education.

Changes in Use of Policy Tools

The globalization process and the new neoliberal governance strategy have affected the choice of policy tools and the balance between the state, market, and civil society. The frameworks of legislation and regulation are still important state authority tools, but they are weakened. National laws and regulations are adapted into an increasing supranational legislation (Shapiro and Sweet, 2002).

One example is the General Agreement on Trade in Services (GATS) agreement from 1995 inside the World Trade Organization (WTO), which created an extended multilateral trading merchandise system to services, inclusive of education (Jawara and Kwa, 2003). Knowledge has more and more emerged as a major trading commodity in the globalized economy (Coulby, 2005).

The economy is another tool and of high importance at the state level and is closely connected to the framing of funding systems. The MbO strategy with its focus on measurable outcomes has contributed to a change in allocation systems of state funding from earmarking to lump-sum systems. In public finance, earmarking is a specification of public expenditure, while lump sum is a total amount and one-time payment. Together with other resources, the economy is used as an incentive tool based on the assumption that institutional and individual behavior is motivated by maximizing utility (Eurydice, 1999).

School programs and curriculum as a state governance tool seem to be less important. Normally, the local teacher institution with the staff will have the obligation to elaborate local programs and curricula within a central framework (Klafki, 1999). However, the professional freedom is under pressure from the needs of the knowledge economy giving priority to basic skills and subject knowledge, particularly in science and technology.

While the importance of the above-mentioned tools all seem weakened in teacher education, the importance of evaluation and control have increased more or less worldwide (Wilson and Youngs, 2006). Agencies for quality assurance at the national level through evaluation, accreditation, and recognition have been established. Politically controversial standardized tests for pupils such as the Organization for Economic Cooperation and Development (OECD)-initiated Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS) also effect teacher education even more indirectly. National and international agents for quality assurance, accreditation, and high standards are playing important roles within the new liberal turn serving powerful new actors.

New Supranational Actors – The New Institutionalism

The traditional studies of governance in political science have been based upon study of institutions, actors, and procedure, within the national state. However, over the last 20 years, studies of the new institutionalism have flourished (Rhodes, 1995; March and Olsen, 2005). New actors are playing new roles. The new institutionalism has roots tracing back to the early period after the Second World War. Supranational institutions such as UN (United Nations), WB (World Bank), WTO, and OECD

(Organisation for Economic Cooperation and Development) became increasingly involved in educational policy. In addition, regional institutions such as the EU in Europe, NAFTA (The North American Free Trade Area) in North America, and APEC (Asia–Pacific Economic Cooperation) in Southeast Asia are increasingly active in shaping education policy (Karlsen, 2002, 2006). Although the main agenda for these intergovernmental institutions was to develop policy for economic progress, they gradually included education in their policy actions. Institutions such as the WTO, WB, and OECD are mainly dominated and controlled by the world's richest countries. These institutions are backed by strong capital forces which have considerable influence, for instance, lobbying toward the political system. Molnar (2005) argues that education now is being swept in by global forces and actors and reoriented toward an integrated relationship with commercial interests.

A common ideological basis for the new international actors was, and is, liberalization emphasizing the importance of free trade and free capital movements. While classical liberalization, with roots tracing back to Adam Smith and John Stuart Mill as influential liberal thinkers, emphasized the value of human rationality and the protection of civil liberties, the new liberalism is closely connected to liberal market theory of economics. This new economic liberalism has occurred as a guarantee for efficiency and higher quality at the same time and implied a shift away from education as a universal citizenship rights to education as tradable commodities for the individual consumer. Alexiadou (2005) claims that the economic function of education has prioritized the creating of consumers and workers in the competitive capitalist economies, while the cultural and social role of education has been reduced.

What is relatively obvious is that concepts from the economic realm are being uncritically applied also to the discourse of teacher education. When knowledge and socialization are referred to as products and educational services, when students are users and the school has to adapt to the market and be competitive, it is not only a question of words, but an expression of thinking. When students are transformed into human capital and are perceived as increased advantage in a global competition economy, it affects educational values and reasoning of governance in a fundamental manner (Hill, 2006). The result is likely to be a more instrumental teacher education with diminished cultural and social relevance and weaker ties to the civil society.

The new global economy needs students with basic skills and high academic standards in central subjects, but at the same time flexible and innovative, motivated for quick and lifelong learning, and able to adapt to changes and the use of new technology (Rikowski, 2002). There is already a “global auction for jobs” (Brown *et al.*, 2007, p. 4), where the main

bidders are today's advanced economies. Greater workforce mobility has led to several governance actions for harmonization and a more standardized teacher education.

A More Standardized Teacher Education

One of the most influential policy actions is known as the Bologna process. The Bologna declaration was signed in 1999 by the ministers of education from 29 countries (Bologna declaration, 1999). The main objectives were the idea of establishing a European area of higher education by 2010, thus making academic degrees and quality assurance more comparable and compatible throughout Europe.

Although the Bologna process is not a formal EU process and now consists of 45 member-countries, the process can be perceived as a national implementation of central elements in EU's policy (Karlsen, 2005). The Bologna process is closely connected to the strategic goal in the EU's Lisbon strategy from 2000: making the EU the “the most competitive and dynamic knowledge-based economy in the world” (European Commission, 2002, p. 7). The aim of creating the European education area corresponds to the successful action of the internal market in the late 1980s and early 1990s. Higher education and knowledge are looked upon and treated more like economic commodities.

The Bologna process is in line with the neoliberal way of governing by setting central goals and encouraging the national states to find tools for implementation and, therefore receive feedback through regular stocktaking (Bergen Communiqué, 2005). It can be basically seen as a governance strategy for standardization of educational systems (Fejes, 2005). The reasoning is that the standardizing of higher education will create transnational employability and mobility within a system with quality assurance. The Berlin Communiqué (2003) called for an overarching framework and led to the EQF (Framework for Qualifications of the European Higher Education Area) (Bologna Working Group, 2005). The EQF consist of eight reference levels relating to learners outcomes, to which the national state has to relate the national qualifications frameworks (European Commission, 2004).

National and international agents for quality assurance, accreditation, and high standards are playing an important role, but the dependence of the policymakers varies. As an example, National Council for Accreditation of Teacher Education (NCATE) is recognized by the US Department of Education as an accrediting institution, while National Board for Professional Teachers Standards (NBPTS) is an independent, nonprofit, and nongovernmental organization formed in 1987 to improve quality by developing professional standards. There is also an ongoing work in OECD to create surveys for teachers and principals called Teacher and Learning International Survey (TALIS). Data from different countries regarding

teachers' competences and standards as defined in the survey tend to have a standardizing effect. In addition, professional efforts may have an unintended effect toward standardization, such as the impressive work edited by Linda Darling-Hammond and John Bransford (2005) with the subtitle 'What teachers should learn and be able to do'.

The European Commission initiative and work for common principles for teacher competences and qualifications demonstrate the importance of governance in teacher education. Teachers are called key players in the implementation of educational reforms to make EU the highest performing knowledge-driven economy in the world by 2010. Teachers should be well qualified; a profession based on partnership; important in the lifelong learning context; and a mobile profession (European Commission, 2005: 2,3). The Lisbon strategy and its related work programs, such as the Common Principles for Teachers, are all implemented according to the open method of coordination (OMC). The OMC includes identifying common objectives, agreeing on benchmarks indicators, exchange of good practice, and peer review (Gornitzka, 2006). Alexiadou (2005) describes OMC as a "new governance tool in the sphere of education" suitable to deal with sensitive areas (p. 13). Teacher education, traditionally closely linked to national sovereignty and issues such as national culture, nation building, and national identity, is a sensitive area in this OMC process.

Concluding Remarks

Teacher education has been and still is highly controversial, criticized, and politicized (Cochran-Smith and Zeichner, 2006). Social problems in the society and standards in schools are, in the political debate, seen as being directly correlated to the assumed quality of teacher education. In contrast to other higher education institutions with tradition for autonomy and self-governing, teacher education institutions historically have been used as important tools for state government purposes. Although teacher education still has a low academic status in traditional meaning, it has become more important politically.

Historically, teacher education has played an important role in maintaining a national culture and the vitality in the civil society. Teacher education has been an important political instrument for building a democratic system at the national level. Both at the national and the institutional level, there is increased tension between individual/institutional freedom and institutional/national control. The tension between individual freedom and institutional control can be related to the accountability outcomes ideology and use of policy tools.

Through the close connection to the compulsory school system, teacher education still forms the basis for

all higher education. However, the market is playing a more visible role in governance in teacher education. Globalization forces, new global actors, and a neoliberal governance strategy emphasize teacher education more instrumentally as a key institution in the new global knowledge economy. Now more than ever, teacher education is seen as a vital component in economic competition and growth. The governance discourse about teacher education is less cultural and social and more market oriented. However, the last years have presaged signs of de-globalization and the return of a new nationalism (Bello, 2002). The national state has played and still plays a key role in the governance of teacher education, but more on the demands from the market.

Bibliography

- Alexiadou, N. (2005). Europeanization and education policy. In Coulby, D. and Zambeta, E. (eds.) *Globalization and Nationalism in Education*, pp 128–146. London: Routledge-Falmer.
- Ball, S. J. (1990). *Politics and Policy Making in Education: Explorations in Policy Sociology*. London: Routledge.
- Bauman, Z. (1998). *Globalization: The Human Consequences*. New York: Columbia University Press.
- Beck, U. (2005). *Power in the Global Age: A New Global Economy*. Cambridge: Polity.
- Bello, W. (2002). *Deglobalization; Ideas for a New World Economy*. Dhaka: University Press.
- Bergen Communiqué (2005). *Communiqué of the conference of European ministers responsible for higher education*, Bergen: 19–20 May.
- Berlin Communiqué (2003). *Communiqué of the conference of ministers responsible for higher education*, Berlin: 18–19 September.
- Bologna Declaration (1999). *Joint declaration of the European ministers of education*, Bologna, 19 June.
- Bologna Working Group on Qualification Frameworks (2005). *Framework for Qualifications of the European Higher Education Area*, Copenhagen. Denmark: Ministry of Science, Technology and Innovation.
- Brown, P., Lauder, H., and Ashton, D. (2007). *Towards a High Skills Economy: Higher Education and the New Realities of Global Capitalism. Paper, ISA conference*, Cyprus: 25–27 May.
- Bron, A. (2001). Civil society and biographical learning. *Key Note Presentation*. NERA Congress, Stockholm, 15–18 March.
- Cochran-Smith, M. and Zeichner, K. M. (eds.) (2005). *Studying teacher education. The Report of the AERA Panel on Research and Teacher Education*. Washington, DC: AERA.
- Coulby, D. (2005). The knowledge economy: Institutions. In Coulby, D. and Zambeta, E. (eds.) *Globalization and Nationalism. World Yearbook of Education 2005*, pp 37–56. London: Routledge-Falmer.
- Coulby, D. and Zambeta, E. (2005). Introduction. Trends in globalization. In Coulby, D. and Zambeta, E. (eds.) (2005). *Globalization and Nationalism in Education. World Yearbook of Education 2005*, pp 23–36. London: Routledge-Falmer.
- Dale, R. (1989). *The State and Education Policy*. Philadelphia, PA: Open University Press.
- Darling-Hammond, L. and Bransford, J. (eds.) (2005). *Preparing Teachers for a Changing World*. San Francisco, CA: Josey-Bass.
- Delandshere, G. and Arens, S. A. (2001). Representations of teaching and standards-based reform: Are we closing the debate about teacher education? *Teaching and Teacher Education* 17, 547–566.
- Delors, J. (1996). *Learning: The Treasure Within. Report to UNESCO of the International Commission on Education for the Twenty-First Century*. Paris: UNESCO.
- Drucker, P. F. (1964). *Management for Results: Economic Tasks and Risk-Taking Decisions*. London: Heinemann.

- Drucker, P. F. (1977). *An Introductory View of Management*. New York: Harper's College Press.
- Dolvik, J. E. and Engelstad, F. (2003). *National Regimes of Collective Bargaining in Transformation: Nordic Trends in Comparative Perspective*. Oslo: Makt og demokratiutredningen.
- ETUCE (2006). Education and training 2010: Competences in the knowledge society-developing policies on European and national implementation. *Report ETUCE Council and Conference February 2006*, Belgium: ETUCE.
- European Commission (2001). *European Governance – A white paper*. Luxembourg: Office for Official Publication of the European Communities.
- European Commission (2002). *Communication from the commission: Making a European Area of Lifelong Learning a Reality*. Luxembourg: Office for Official Publication of the European Communities.
- European Commission (2004). *Implementation of "Education and Training 2010" work Program Key Competences for Lifelong Learning. A European Reference Framework*. Luxembourg: Office for Official Publication of the European Communities.
- European Commission (2005). *European Testing Conference on Common European Principles for Teacher Competences and Qualification*, Brussels: 20–21 June.
- Eurydice (1999). *Financing Education: Thematic Bibliography*. Belgium: Eurydice.
- Eurydice (2002). Key topics in education in Europe. Volume 3: The teaching profession in Europe: Profile, trends and concerns. Report II: Supply and Demand – General Lower Secondary Education.
- Fägerlind, I. and Strömquist, G. (eds.) (2004). *Reforming Higher Education in the Nordic Countries*. Paris: International Institute for Educational Planning.
- Fejes, A. (2005). The Bologna-process - governing higher education in Europe through standardisation. In Halvorsen, T. and Nyhagen, A. (eds.) *The Bologna Process and the Shaping of the Future Knowledge Societies*, pp 219–232. Bergen: University of Bergen.
- Foucault, M. (2003). Governmentality. In Rabinow, R. and Rose, N. (eds.) *The Essential Foucault; Selections from the Essential Works of Foucault 1954–1984*, pp 229–245. New York: The New Press.
- Garm, N. and Karlsen, G. E. (2004). Teacher education in Europe: The case of Norway; Trends and tensions in a global perspective. *Teaching and Teacher Education* **20**, 731–744.
- Giddens, A. (1990). *The Consequences of Modernity*. Cambridge: Polity Press.
- Giddens, A. (2000). *The Third Way and Its Critics*. Cambridge: Polity Press.
- Gornitzka, Å. (2006). *The Open Method of Coordination as Practice - A Watershed in European education policy?* Working paper No. 16 2006, Oslo: ARENA .
- Hann, C. and Dunn, E. (eds.) (2006). *Civil Society: Challenging Western Models*. London: Routledge.
- Held, D. and McGrew, T. (eds.) (2000). *The Global Transformations Reader: An Introduction to the Globalization Debate*. Cambridge: Polity Press.
- Helgøy, I. and Homme, A. (2004). *The Relation between Policy Reforms and Governing Tools in Education. Comparing Teacher Professionalization in Norway, Sweden and England. Paper to be presented at the European Conference on Educational Research ECER, Crete: 22–25 September*.
- Hill, D. (2006). Education Services Liberalization. In Roskam, E. (ed.) *Winners or Losers? Liberalizing Public Services*. Geneva: International Labour Office.
- Jawara, F. and Kwa, A. (2003). *Behind the Scenes at the WTO. The Real World of International Trade Negotiations*. London: Zed Books.
- Jones, M., Jones, R., and Woods, M. (2004). *An Introduction to Political Geography: Space, Place and Politics*. London: Routledge.
- Karlsen, G. E. (2000). Decentralized centralism: Framework for a better understanding of governance in the field of education. *Journal of Education Policy* **15**(5), 525–538.
- Karlsen, G. E. (2002). Educational policy and educational programmes in the European Union: A tool for political integration and economic competition? In Ibanez-Martin, J. and Jover, G. (eds.) *Education in Europe: Policies and Politics*, pp 23–49. Dordrecht: Kluwer Academic Publishers, 23–49.
- Karlsen, G. E. and Kvalbein, I. A. (eds.) (2003). *Norsk Lærerutdanning (Norwegian Teacher Education)*. Oslo: Universitetsforlaget.
- Karlsen, G. E. (2005). The Bologna process - A judicial confirmation of EUs policy of education? In Halvorsen, T. and Nyhagen, A. (eds.) *The Bologna Process and the Shaping of the Future Knowledge Societies*, pp 214–218. Bergen: University of Bergen.
- Karlsen, G. E. (2006). *Utdanning, Styring og Marked (Education, Governance and Market)*. Oslo: Universitetsforlaget.
- Klafki, W. (1999). Didaktik analysis as the core of preparation of instruction. In Westbury, I., Hopmann, S., and Riquas, K. (eds.) *Teaching as a Reflective Practice. The German Didaktik Tradition*, pp 139–159. Mahwah, NJ: L. Erlbaum Associates.
- Kohler-Koch, B. (2005). European governance and system integration. *European Governance Papers. No. C-05-01*.
- Lavigne, M. (1995). *The Economics of Transition: From Socialist Economy to Market Economy*. Basingstoke: Macmillan.
- Lawn, M. and Lingard, B. (2002). Constructing a European policy space in educational governance: The role of transnational policy actors. *European Educational Research Journal* **1**(2), 290–307.
- Lindblad, S. and Popkewitz, T. S. (eds.) (2001). *Listening to Education Actors on Governance and Social Integration and Exclusion*. Uppsala: University of Uppsala, Department of Education, Uppsala Reports on Education 37.
- Lundahl, L. (2005). Swedish, European, Global. In Coulby, D. and Zambeta, E. (eds.) *Globalization and Nationalism in Education. World Yearbook of Education 2005*, pp 147–164. London: Routledge-Falmer.
- March, J. G. and Olsen, J. P. (2005). *Elaborating the 'New Institutionalism'* Working paper No 11. Oslo: Centre for European Studies.
- Martin, H.-P. and Schumann, H. (1996). *Die Globalisierungsfalle: der Angriff auf Demokratie und Wohlstand (The Globalization Trap: The Assault on Democracy and Prosperity)*. Reinbek: Rowohlt.
- Mishra, R. (1999). *Globalization and the Welfare State*. Cheltenham: Edward Elgar.
- Molnar, A. (2005). *School Commercialism. From Democratic Ideal to Market Commodity*. New York: Routledge.
- Negreponi-Delivanis, M. (2001). *Conspiring Globalisation*. Athens: Papazisi.
- Olsen, J. P. (2002). Reforming european institutions of governance. *Journal of Common Market Studies* **40**(5), 581–602.
- OECD (2000). *Governance for the Future*. Paris: OECD.
- OECD (2005). *Teachers Matters. Attracting, Developing and Retaining Effective Teachers*. Paris: OECD.
- Pollitt, C. and Bouckaert, G. (2000). *Public Management Reform - A Comparative Analysis*. Oxford: Oxford University Press.
- Rhodes, R. (1995). The institutional approach. In March, D. and Stoker, G. (eds.) *Theory and Methods in Political Science*, pp 42–57. London: Macmillan.
- Rikowski, G. (2002). Education, Capital and the Transhuman. In Hill, D., McLaren, P., Cole, M., and Rikowski, G. (eds.) *Marxism against Postmodernism in Educational Theory*, pp 111–143. Lanham, MD: Lexington Books.
- Rokkan, S. (1987). *Stat, Nasjon, Klasse (State, Nation, Class)*. Oslo: Universitetsforlaget.
- Shapiro, M. and Sweet, A. (2002). *On Law, Politics, and Judicialization*. Oxford: Oxford University Press.
- Sassen, S. (1998). *Globalisation and its Discontents*. New York: The New Press.
- Spybey, T. (1996). *Globalization and World Society*. Cambridge: Polity Press.
- Thomas, E. (ed.) (2002). *Teacher Education. Dilemmas and prospects. World Yearbook of Education 2002*. London: Kogan Page.
- Vaage, S. (ed.) (2001). *Education and nationbuilding: Conference report 14–15 June 2001*. Volda: Volda University College.
- Wilson, S. M. and Youngs, P. (2006). Research on accountability processes in teacher education. In Cochran-Smith, M. and Zeichner, K. M. (eds.) *Studying Teacher Education*, pp 591–644. Washington, DC: AERA.

Further Reading

- Carnoy, M. (1999). *Globalization and Educational Reform: What Planners Need to Know*. Paris: UNESCO.
- Clarke, A. (2001). The recent landscape of teacher education: Critical points and possible conjectures. *Teaching and Teacher Education* **17**, 599–611.
- Cochran-Smith, M. (2005). The new teacher education: For better or worse. *Educational Research* **34**, 3–17.
- Dale, R. (2000). Globalization: A new world for comparative education? In Schriwer, J. (ed.) *Discourse Formation in Comparative Education*, pp 87–110. Frankfurt: Peter Land Verlag.
- Daun, H. (1998). *Restructuring Education in Europe*. Stockholm: Institute of International Education.
- Hargreaves, A. (1994). *Changing Teachers, Changing Times*. London: Cassell.
- Held, D. (ed.) (1998). *A Globalizing World?: Culture, Economics, Politics*. London: Routledge.
- Held, D. and McGrew, A. (2002). *Globalisation/Anti-Globalisation*. Cambridge: Polity Press.
- Ibanez-Martin, J. and Jover, G. (eds.) *Education in Europe: Policies and Politics*. Dordrecht: Kluwer Academic Publishers.
- Karlsen, G. E. (1993). *Desentralisert Skoleutvikling (Decentralized School Development)*. Oslo: AdNotam Gyldendal.
- Kuhnle, S., Flora, P., Urwin, D., and Rokkan, S. (eds.) (1999). *State Formation, Nation-Building, and Mass Politics in Europe: The Theory of Stein Rokkan: Based on His Collected Works*. Oxford: Oxford University Press.
- Popkewitz, T. S. (ed.) (1987). *Critical Studies in Teacher Education. Its Folklore, Theory and Practice*. London: The Falmer Press.
- Rizvi, F. (2004). Theorizing the global governance of educational restructuring. In Lindblad, S. and Popkewitz, T. S. (eds.) *Educational Restructuring. International Perspectives on Traveling Policies*, pp 21–41. Greenwich: IAP.
- Tudor, I. (2006). Teacher training and 'quality' in higher education language teaching. Strategies and options. *European Journal of Teacher Education* **29**, 519–532.

TEACHER EDUCATION – POST-INITIAL PROFESSIONAL DEVELOPMENT

Contents

Action Research as a Tool for Teachers' Professional Development

Contemporary Approaches to Teacher Professional Development

School Development for Teacher Learning and Change

Teacher Induction

Teacher Learning as Workplace Learning

Teacher Learning with Lesson Study

Teachers Career Stages and Professional Development

Action Research as a Tool for Teachers' Professional Development

P Ponte, Leiden University, Leiden, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Action research by teachers is often justified by the argument that the reality of educational practice is complex, changeable, and cannot always be accurately predicted in advance by general academic knowledge. Teachers constantly have to face the question: What in the given circumstances is the best way to act, in order to achieve what is important at this moment? Moreover, teachers who want to change their practice have to deal with issues related to how they or their pupils, colleagues, and others want to organize their lives and their work and so they always touch upon normative and ethical issues. That is why solving a problem or changing a situation is not only linked to empirical evidence but also to normative criteria. In social science this argument led to a plea for interpretative research focused on *verstehen*, that is, meaning construction (Gadamar, 1976). Interpretative research includes research from an insiders' perspective (action research with or by teachers) in addition to research from an outsiders' perspective (academic research on teachers) (Cochran-Smith and Lytle, 1993).

Ideas about the insiders' and outsiders' perspectives in social research parallel a changed view on teachers' learning in schools and in teacher education programs. With regard to teacher education, authors seem to agree about the limitations of the traditional transfer model, because general academic knowledge cannot simply be transferred in the expectation that teachers can apply this knowledge as intended by researchers. Empirical research has shown that teachers hardly ever use the knowledge they had to acquire on their teacher education

courses in their own practice (Gore and Gitlin, 2004). Similar doubts are being expressed about large-scale school reforms, where the professional development of the teachers who ultimately have to deliver the education tends to be neglected. Although brief one-shot workshops and similar activities are often offered, research has shown that these activities are not very effective (Cordingley *et al.*, 2007). Alternative approaches are often proposed which allow teachers to learn in and from their own practice through self-directed, research-based activities. These activities are based on ideas of action research as a tool for professional development, which is the focus of this article.

Action Research as a Tool for Professional Development

A great deal has been written about research by teachers as a tool for professional development and this has produced a range of interpretations. An idea that is common to the various interpretations of action research is that research strategies of the social sciences can be used to change social practice. Elliott (1991), for instance, defined action research as "the study of a social situation with a view to improving the quality of action within it" (p. 69). Carr and Kemmis (1986/1997) gave a more elaborate definition: "Action research is a form of self-reflective inquiry undertaken by participants in social situations in order to improve the rationality and justice of their own social or educational practices, their understanding of these practices, and the situation in which the practices

are carried out” (p. 162). Action research can be carried out by individual teachers, by a group of teachers, or by teachers in collaboration with academic researchers. Five key features can be identified from the literature (Ponte, 2002; Somekh, 2006):

- Action research is carried out by teachers (or other practitioners in the school) themselves. They study questions that emerge from their day-to-day experiences in their own practice and the situation in which they are practicing. The questions are supposed to arise from an experienced gap between what is intended and what actually takes place.
- In action research, teachers engage in reflection and change based on information they have systematically gathered and analyzed themselves. Teachers not only collect and analyze data about their pupils – or others who are involved in the issue under study – but also about their own teaching and learning. As in other forms of interpretative research triangulation, thick descriptions, peer debriefing, and member checks are among the main validation strategies to be used.
- Action research is carried out through dialog and collaboration with colleagues and others within and outside the school. Those others may function as a source of information, as co-researchers, and/or as critical friends who help each other to reflect critically on the design, execution, and results of the action research.
- In action research, those who are involved in the situation that needs to be changed are seen as partners in the research. In teaching, pupils are often, if not always, part of the situation that needs to be changed, and the pupils' voice, therefore, is seen as an essential source of learning for the teachers.

- The results of action research are published and are open to criticism from the wider community.

Teachers usually work with a model, traditionally consisting of complicated spirals and cycles of consecutive steps, such as: the formulation of a general idea; exploration of the general idea; drawing up of a general plan; planning, implementation, monitoring, and evaluation of concrete actions for improvement; and writing up of a case study or report on the teachers' own action research. It has gradually become clear, however, that teachers often struggle with these complex plans (Ebbutt, 1985), and the importance of a strict sequence of steps is now played down. Kemmis and Wilkinson (1998, p. 21) asserted that action research is a “fluid, open and responsive” process. In modern handbooks on action research (see, for instance, Altrichter *et al.*, 1993; Burnaford *et al.*, 2001; Somekh, 2006), research models are increasingly seen as heuristic tools instead of fixed sets of steps.

Box 1 contains a description of a school-based action research project carried out by McLaughlin *et al.* (1996).

History of Action Research

Action research has a long tradition, both in education and in other fields. As early as the beginning of the last century, action-research initiatives in education represented various attempts to bridge the gap between theory (academic knowledge) and teachers' practice. Action research started in the US and later spread to Australia, England, and other countries. A common feature of these early attempts was that they aimed to use research strategies to initiate and realize social (i.e., educational) change. Four important

Box 1

In the period before we started on the project, team members were more often engaged in one-to-one counseling and guidance of pupils. We had the idea that if pupils were better able to put their problems into words, the school would be able to respond more effectively. We assumed that problems related to confidentiality would play a part here. In terms of action research, this gave us our general idea.

This general idea led us to form a group which included members of the school management team, teachers, and school governors. We discussed the problem and invited Colleen – as the external researcher – to help us develop our ideas further. She gave a presentation on what can be learned about our problem from the professional literature. In terms of action research, this can be seen as a form of exploration.

We decided to extend our exploration by gathering factual data. To gather evidence in support of our assumptions on confidentiality, Pam was asked to interview pupils to get their views. Colleen interviewed teachers and Meryl explored school development by keeping a logbook and analyzing relevant documents. The insights gained were presented to the team. Members of the team were given the opportunity to respond by completing questionnaires and taking part in a team discussion. We saw this as additional evidence in support of our assumptions.

We discussed the findings and decided what we wanted to do. In action research terms, we drew up a general plan. Training the team was the first concrete step. This training was systematically analyzed and discussed (in action research terms, monitoring and evaluation). The feedback we obtained from this first step led us to reformulate our general idea and carry out a new explorative phase. The general plan was adapted to tackle problems to do with confidentiality that could not be resolved through training. This was our first action research cycle.

Subsequent cycles helped us to develop school policy on confidentiality. Guidelines were drawn up that were accepted by the whole school. Finally, we published our experiences.

(Summarized by McLaughlin, Clark and Chisholm in James and Ponte 1997, p. 25.)

theories deserve to be mentioned in this context: (1) the pragmatism of Dewey, (2) Lewin's theory on research in social change, (3) Corey's theory on curriculum development and research, and (4) the teacher as researcher approach of Stenhouse. These theories are discussed below in chronological order, leading on to some current interpretations of action research as a tool for professional development.

The foundations of action research were laid in the first place by the democratic reform movement and Dewey's pragmatism in the US at the beginning of the twentieth century. The interpretation of educational situations, according to Dewey (1929/1984), began with reflection on pupils' motives and problems – as they themselves experienced them. Teachers, therefore, were expected to think reflectively. Reflective thinking requires the continual evaluation of beliefs, assumptions, and hypotheses against existing data and against other plausible interpretations of data. The knowledge that teachers construct in this way helps them to change their teaching.

After Dewey's pragmatism, the social-psychological debate about the place of research in social change played a major role. It was Lewin (1947) who argued that communities that aspired to social change would have to study the impact of their own actions, and this meant that they would also have to explore their own norms and values. He was the first person to use the term action research, based on the assumption that methods of social-science research could be applied directly to solve social problems. The basic assumption was that each group that is involved in the situation to be changed would also be involved in planning and carrying out the research and in evaluating the research findings.

In the 1950s a debate emerged on the contribution of research to curriculum development. It was Corey who developed a collective form of action research, in which teachers and school principals worked together on curriculum development with external researchers. The purpose of this collaboration, according to Noffke (1997), was "(1) an increased commitment of teachers to change; (2) an increased probability that the actions proposed to teachers would be possible; (3) a greater range and variety of talent; (4) a reduction of individual risk; and (5) the prevention of feelings of manipulation" (p. 10).

In the 1960s ideas associated with action research faded into the background under the influence of positivist views on the relationship between educational research and educational development. The dominant curriculum model assumed a strategy in which research and development are separated: academics do research, experts design and disseminate, and teachers apply. The results from this approach proved to be disappointing and, partly because of this, action research experienced a revival in the 1970s. Stenhouse (1975) then introduced the teacher as researcher approach in England as an aspect of teacher-based curriculum reform. Like Corey, Stenhouse rejected the product approach to

the curriculum that was dominant at that time in favor of a process approach. In the process approach, the concept of curriculum is not conceived as planning but as realized teaching. The emphasis is no longer on the product but on the process of curriculum development (see, for instance, Grundy, 1987/1995). Like Dewey and Lewin, he associated action research with democratic views on education. These views were bound up with a commitment to human freedom, equality of educational opportunity, and freedom of belief. While Dewey, Lewin, and Corey were discussing action research mainly as a scientific method for social change, Stenhouse (1983) tended to also approach action research as a concept for the teachers' learning. He defined action research as: "a pattern of learning by a thoughtful study of problems; such study becomes research when it is made public by being published, at which point the student (read 'the teacher' for the purposes of this article) makes a claim intended to evoke a critical response" (p. 185).

Approaches That Emerged

The initiatives of Dewey, Lewin, Corey, and Stenhouse led to many different types of approach, for which the term action research is often used. Despite the many and varied interpretations of this concept and the overlaps between them, the following broad classification of action research can be made (see also, McKernan, 1996):

- approaches termed interactive research, development, and Dissemination;
- approaches termed teacher research; and
- approaches termed educational action research.

These approaches differ in a number of respects, in particular with respect to views on (1) the role of the different actors in action research, (2) the purpose of action research, and (3) the relationship between the development and application of insights by teachers. These differences are explained below.

Many interactive research, development, and dissemination projects involve field research that also aims to improve education in schools. This does not always necessarily progress to the actual implementation of methodically planned and evaluated actions for improvement by teachers. The basic assumption, according to McKernan (1996), is that: "each team shares in planning, implementation, analysis and reporting of the research and that team members contribute unique skills and expertise in a collective process. Often teams are made up of university faculty members, teachers, administrators, educational laboratory research and development personnel and funding body staff" (p. 12). Nowadays, this approach has evolved into many forms of collaborative action research in which teachers and researchers work together on a shared issue (McLaughlin *et al.*, 2005; McLaughlin, 2006).

Teacher research is defined by Lytle and Cochran-Smith (1994) as “systematic and intentional inquiry carried out by teachers in their own schools and classrooms, which has the potential to make accessible some of the expertise of teachers and to provide the university and the school communities with unique perspectives on teaching and learning” (p. 24). Various forms of narrative inquiry are also seen as teacher research. In this type of research the main emphasis is on teachers developing theories themselves and, as a logical progression from that, the professional development of teachers. Teacher research does not necessarily include methodically planned and evaluated actions for change (Zeichner and Noffke, 2001). Especially in the initial and post-initial teacher education courses in the US, teacher research is also often linked to Schön's (1987) ideas on reflection, which assume teachers' reflecting on their practices and integrating their observations into their emerging theories of teaching and learning. These forms of teacher research place more emphasis on the personal professional development of teachers than on theory development by teachers.

Both interactive research, development, and dissemination and teacher research seem to assume that teachers automatically change the way they act – or the situation in which they are acting – based on knowledge they develop. Experimenting with the insights obtained is not always seen as a necessary part of the research, whereas it is in the approaches termed educational action research. Stenhouse (1975), for instance, agreed with Lewin, in seeing real actions for change as an essential component of learning through action research. This was one of the reasons why Stenhouse's ideas have been further developed into a concept for professional development of teachers (see, for instance, Elliott, 1991; Tabachnick and Zeichner, 1991; Carr and Kemmis, 1986/1997).

Teachers' Learning through Action Research in Practice

It will be clear from the descriptions above that action research has endless variations. Teachers' learning depends not only on the differences discussed above, but also on the contextual and local conditions in which the research is conducted. Few comprehensive studies have been carried out on the effects of different forms of action research on teachers' learning, so it is difficult to generalize about teachers' learning through action research. Nevertheless, it is possible to offer some answers to the following questions:

- Who is carrying out action research?
- How do teachers learn through action research?
- What do teachers learn from action research?
- Under what conditions do teachers learn through action research?

Box 2

P was finding that his lessons were disorganized, unfocused, and not enjoyable. The pupils were not making much progress either. He asked himself the question: “How can I, in consultation with my pupils and colleagues, change the way I teach so that the class atmosphere improves and the pupils' performance improves?” He asked pupils to write down examples of what they considered to be good and bad teaching. From the long list of good teacher behavior he compiled a short list of 11 good qualities. He listed these in order of importance based on the pupils' opinions. Important good qualities included: “(1) explaining things well, (2) being patient, (3) explaining it until we understand it, ten times if necessary, (4) having a friendly chat at the beginning of a lesson, not starting on the work immediately, and (5) only giving an amount of homework that you can finish in the lesson.” P then asked the pupils to observe him and to score his lessons on the qualities that they themselves had identified as important. He then selected a number of areas where the pupils had given him rather low scores and started to experiment with them. This approach helped the pupils to feel that they were being taken seriously, as a result of which the atmosphere in the class improved (Ponte, 2002).

Box 3

Five experienced teachers in Ontario, Canada, joined a school-university partnership to study the student evaluation practices of 13 exemplary teachers selected for their expertise in using cooperative learning (Phase 1). Data were collected in two semi-structured interviews. Participation in phase 1 of the action research left the teacher-researchers in a state of positive dissonance: dissatisfied with their evaluation methods, aware of attractive new strategies, and confident of their ability to change their practice. In phase 2, the teacher-researchers conducted inquiries in which they developed and enacted strategies for teaching students how to self-evaluate. Data sources included student surveys of attitudes toward evaluation, student focus-group interviews, individual interviews with teacher-researchers, observations of team deliberations, action research reports, and storyboards created by the teacher-researchers. Each teacher-researcher used the phase 1 findings in different ways, with every teacher ignoring some results, modestly adapting others, and, most frequently, reconstructing the intent of the exemplary teachers without adopting their specific methods. The two-phase approach to action research contributed not only to the knowledge base of teaching (creation of a framework for teaching self-evaluation) but also to the personal practical knowledge of the teacher-researchers (how to enact the framework in their classroom) and enhanced teachers' expectations about their ability to bring about learning. These benefits might not accrue in action research projects without the involvement of teachers with average- or high-efficacy beliefs, collaboration in the design and analysis of the projects, or supportive university professors sharing research skills.

(Abstract by Ross *et al.*, 1999.)

Who Is Carrying Out Action Research?

Educational action research is carried out individually (by teachers in a school or by students on undergraduate and postgraduate courses) or collaboratively (by a team of teachers in a school or by university–school partnerships). The example in **Box 2** illustrates that an individual action research project may become a way of being for the teachers. However, its systematic and intentional nature, according to Zeichner (2007), is “what distinguishes it from the everyday thoughtful behaviour of good teachers. The evolving nature of action research is what distinguishes it from mere problem-solving” (p. 2). The example in **Box 3** is one of the many examples to be found on the Internet of collaborative action research in a school–university partnership in which a contribution to the knowledge base of teaching is claimed.

How Do Teachers Learn through Action Research?

Based on a variety of sources, including a review of action research reports in the *Educational Action Research Journal*, Ponte (2007) distinguished five strategies by which teachers learn through action research. These are briefly explained below.

Interaction between the application and construction of professional knowledge

Learning through action research can be characterized by the simultaneous construction and application of knowledge. By simultaneous it is meant that the development and application of knowledge are part of one cyclical process: teachers apply knowledge; they gather information on it; they interpret that information and, based on their interpretations, develop new knowledge, which they then apply again; and so on.

Interaction between academic and professional knowledge

Learning through action research takes place when teachers link their own professional knowledge to the general knowledge of academics. Without theory, without distance, and without abstractions, the knowledge of professionals can get stuck at the level of uncritical experience of everyday events, without consequences for future action. Not linking professional knowledge to academic knowledge may even lead to action research becoming counterproductive.

Interaction between educational knowledge and methodological knowledge

Learning through action research takes place at two levels. The first level concerns educational knowledge about course content, teaching strategy, the pupils, and

so on. To construct this educational knowledge, teachers also develop methodological knowledge or knowledge at the second level; in concrete terms this means knowledge about how to study their own practice.

Interaction between individual and collective knowledge

Teachers learn by constantly making connections between their personal knowledge and the collective knowledge of their peers. The first thing to note about collective knowledge is that it is knowledge which is described in such a way that it can be shared with others; for instance, by presenting reality in the form of a model, or by ordering and naming aspects of reality in a certain way. Shared knowledge is therefore necessarily abstracted knowledge which is open to debate.

Interaction between ideological, instrumental, and empirical knowledge

Teachers learn by linking the instrumental knowledge (What strategies do we normally have at our disposal and how can we apply them?) to ideological knowledge (What goals do we essentially want to achieve with our strategies and what are the moral–ethical pros and cons involved?). Both types of knowledge are still essentially concerned with plans. To fathom out the practical significance of this knowledge, teachers must constantly make a connection with empirical knowledge, that is, with knowledge about the actual teaching situations in which they are engaged in their daily practice.

What Do Teachers Learn from Action Research?

Many (mainly small-scale) studies have demonstrated the positive effects of action research on teachers’ learning (see Campbell *et al.*, 2007; Cordingley *et al.*, 2007; Dadds, 1995; McLaughlin *et al.*, 2005; McLaughlin, 2006; Zeichner and Noffke, 2001; see also, for instance, the *Educational Action Research Journal*, vols. 1–15). The main conclusion that can be drawn from these studies is that there are two different ways in which teachers find that doing action research helps them to develop a sense of empowerment. First, they report being able to influence their own practice and the situation in which they are practicing; and second, they report that they have more control over the process and that leads to even greater influence (knowing how they can realize that influence) (see also, Harris and Muijs, 2006). The latter is associated with the fact that teachers are acquiring habits and skills of inquiry that they continue to use after their first successful experience with action research. Research by, for instance, Caro-Bruce *et al.* (2007) into the effects of various action research projects in the US presents a similar picture.

These studies found that teachers feel encouraged to talk and collaborate with other teachers, and they concluded that teachers' learning through action research led to significant changes in their practice. Teachers become more learner-centered by talking and listening to their students more carefully and they develop higher expectations of what their students know and can do.

Finally, Zeichner (2007) claimed that "although there is very little evidence in the literature of improved student learning associated with action research as measured by standardised test scores, teachers who do action research have reported improvements in pupil attitudes, involvement in classroom activities, the analysis of students' work samples, and teacher-designed assessments." The findings presented here are consistent with studies such as these (Campbell *et al.* 2007; Cordingley *et al.*, 2007; McLaughlin *et al.*, 2005; McLaughlin, 2006; Zeichner and Noffke, 2001).

Under What Conditions Do Teachers Learn through Action Research?

Research findings also show that learning to do action research is more difficult for some teachers than for others, even when the participants are experienced teachers (see, for instance, the reviews of Campbell *et al.*, 2007; Cordingley *et al.*, 2007; Zeichner and Noffke, 2001; see also, Ponte, 2002). This is first to do with the fact that action research involves working on understanding and changing teachers' own practice over the long term, while the hectic everyday working life of teachers often invites them to do just the opposite, that is to seek short-term solutions to problems in an *ad hoc* way. Second, doing action research can force teachers to face up to hard facts, because it constantly requires them to reveal why they did what they did and to actually carry out what they planned. All this means that action research is sometimes difficult to sustain. Even teachers who are experienced with action research turn out to need colleagues to challenge and support them. Kemmis and McTaggart (1988) made the following recommendations based on their experience: "Start small, with a small group, negotiate meeting times, articulate a thematic concern and establish agreement that this concern is shared, establish a time-line for the action research cycle, arrange supportive work-in-progress discussions, work with 'critical friends', register progress not only with the participant group but also with the staff as a whole and other interested parties, make time to write, and be explicit about what has been achieved by reporting progress." (p. 25–26). In line with other studies (Campbell *et al.*, 2007; Cordingley *et al.*, 2007; Ponte, 2002; McLaughlin *et al.*, 2005; McLaughlin, 2006), Zeichner (2007) concluded that these conditions can only be realized when:

- Teachers work in a culture that respects their opinions and the knowledge that they bring to the research experience.
- Teachers have control over most aspects of the research process including defining a research focus, data collection, analysis strategies, and so on.
- Teachers are enabled to collaborate over time in a safe and supportive environment.
- The environment provides teachers with intellectual challenges and stimulation.

Final Remarks: Action Research in Initial Teacher Education

The findings outlined above present a general picture of action research as a tool for teachers' professional development. However, it is necessary to make allowance for differences between action research carried out by student teachers and the action research of experienced teachers. These differences can be made clear by looking at how the five strategies by which teachers learn when doing action research, as set out earlier, need to be modified when they are applied to student teachers.

Regarding interaction between the application and construction of professional knowledge, students are not blank canvasses when they embark on a course; also, what they bring with them is neither professional nor systematized knowledge based on experience in professional practice. Initial teacher education programs cannot therefore start with the application of professional knowledge as a basis from where to go on to develop new knowledge. The students do not have this knowledge when they start the course; they have to build it up gradually. This can be done by making connections with what they do have, namely general ideas about what teaching involves based on their own experiences. In this sense there appears to be a fundamental difference between action research by student teachers and experienced teachers. This also applies to the interaction between academic and professional knowledge. For both groups of teachers this is concerned with the question of how the knowledge component relates to the action component; however, student teachers will ask different kinds of questions from experienced teachers.

With regard to the interaction between educational knowledge and methodological knowledge, students come to the course with the expectation that they will learn how to teach and what their lessons should contain (educational knowledge). Only during the course will they learn to see and use action research as a way to continually develop their practice (methodological knowledge, see above). This focus on lifelong learning will be less obvious among student teachers than among experienced teachers.

As far as the interaction between individual and collective knowledge is concerned, students cannot be expected to have experience of teaching in an institutional, collective environment, let alone of sharing knowledge in such an environment, as envisaged by the ideal action research model. The challenge for the course, therefore, is to teach students to connect the action research they learn about on the course with the situation at the school where they carry out that research.

Finally, with regard to the interaction between ideological, instrumental, and empirical knowledge, student teachers can be expected to concentrate on the instrumental area of knowledge (What do I have to do?) and not yet on the ideological area of knowledge (Why do I want to do that?), or the empirical area of knowledge (Does what I am doing match the reasons why I am doing it?). The challenge for the course here is to stimulate students to use their research to develop knowledge in the different areas and to connect the different areas together.

Teacher education courses must, therefore, make allowance for the specific needs of student teachers, and research findings (Ponte *et al.*, 2004) suggest that courses are most successful in this when they slowly build up to grooming the students to carry out a fully fledged action research project. This means that reflective research-based activities should run right through all components of the curriculum from the start, but that a full-fledged action research project will only be appropriate at the end of the course. More and more action research is being done in projects in which teacher educators, mentors, or teachers in the school and the student teachers collaborate (Campbell *et al.*, 2007). Teachers who take a 1-year teacher education course after first graduating in a particular subject will only get the opportunity to get involved in a full-fledged action research project as part of in-service training activities after completing their initial training. Close cooperation between initial and post-initial activities is therefore essential. Finally, it is possible to conclude that learning to do action research can be classed as a form of professional socialization, that is, the gradual internalization of a set of professional norms and values. This socialization will not be achieved by carrying out an isolated block of action research within a curriculum based on traditional principles. Rather it is a total concept for the professionalism and professional development of teachers that pervades the whole curriculum.

Bibliography

- Altrichter, H., Posch, P., and Somekh, B. (1993). *Teachers Investigate Their Work: An Introduction to the Methods of Action Research*. London: Routledge.
- Burnaford, G., Fisher, J., and Hobson, D. (2001). *Teachers Doing Research*. London: Lawrence Erlbaum Associates.
- Campbell, A., McNamara, O., Furlong, J., Lewis, S. F., and Howson, J. (2007). The evaluation of the National Partnership Project in England: Processes, issues and dilemmas in commissioned evaluation research. *Journal of Education for Teaching* **33**(4), 471–483.
- Caro-Bruce, C., Flessner, R., Klehr, M., and Zeichner, K. (2007). *Equitable Classrooms through Action Research*. Thousand Oaks, CA: Sage Publications.
- Carr, W. and Kemmis, S. (1986/1997). *Becoming Critical*. London: Falmer Press.
- Cochran-Smith, M. and Lytle, S. L. (eds.) (1993). *Inside/Outside: Teacher Research and Knowledge*. New York: Teacher College Press.
- Cordingley, P., Bell, M., Isham, C., Evans, D., and Firth, A. (2007). What do specialists do in CPD programmes for which there is evidence of positive outcomes for pupils and teachers? Report. In *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Dadds, M. (1995). *Passionate Enquiry and School Development: A Story about Teacher Action Research*. London: Falmer Press.
- Dewey, J. (1929/1984). The sources of a science of education. In Boydston, J. A. (ed.) *The Later Works: Vol. 5, 1929–1930*, pp 3–40. Carbondale, IL: Southern Illinois University Press.
- Ebbutt, D. (1985). Educational action research: Some general concerns and specific quibbles. In Burgess, R. (ed.) *Issues in Educational Research*, pp 152–174. London: Falmer Press.
- Elliott, J. (1991). *Action Research for Educational Change*. Buckingham: Open University Press.
- Gore, J. and Gitlin, A. (2004). Re-visioning the academic-teacher divide. *Teachers and Teaching: Theory and Practice* **10**(1), 35–58.
- Gadamar, H. (1976). *Vernunft im Zeitalter der Wissenschaft: Aussätze (Rationality in Century of Science)*. Frankfurt: Suhrkamp.
- Grundy, S. (1987/1995). *Curriculum: Product or Praxis?* London: Falmer Press.
- Harris, A. and Muijs, D. (2006). Teacher-led school improvement. Teacher leadership in the UK. *Teaching and Teacher Education* **22**, 961–972.
- James, M. and Ponte, P. (1997). *Het onderzoeken en verbeteren van de eigen praktijk (Researching and improving ones own practice)*. In Ponte, P. and Zwaal, P. (eds.) *Over Grenzen. Een nieuwe manier van werken aan leerlingbegeleiding (Overboarders: A New Way of Working on Pupil Guidance)*, pp 25–41. Antwerpen: Garant.
- Kemmis, S. and Wilkinson, M. (1998). Participatory action research and the study of practice. In Atweh, B., Kemmis, S., and Weeks, P. (eds.) *Action Research in Practice: Partnerships for Social Justice in Education*, pp 21–47. London: Routledge.
- Lewin, K. (1947). Frontiers in group dynamics II: Channels of group life; social planning and action research. *Human Relations* **1**, 143–153.
- Lytle, S. L. and Cochran-Smith, M. (1994). Inquiry, knowledge and practice. In Hollingsworth, S. and Sockett, H. (eds.) *Teacher Research and Educational Reform*, pp 22–51. Chicago, IL: University of Chicago Press.
- McKernan, J. (1996). *Curriculum Action Research: A Handbook of Methods and Resources for the Reflective Practitioner*. London: Kogan Page.
- McLaughlin, C. (2006). *Researching Schools. Stories from a Schools–University Partnership for Educational Research*. London: Routledge.
- McLaughlin, C., Clark, C. P., and Chisholm, M. (1996). *Counselling and Guidance in Schools: Development Policy and Practice*. London: David Fulton.
- McLaughlin, C., Hawkins, B. K., and Townsend, A. (2005). Practitioner research and enquiry in networked learning communities. *Research Report*. Cambridge: University of Cambridge Faculty of Education for The National College of School Leadership (Networked Learning Communities).
- Noffke, S. E. (1997). Themes and tensions in US action research: Towards historical analysis. In Hollingsworth, S. (ed.) *International Action Research: A Casebook for Educational Reform*, pp 2–16. London: Falmer Press.
- Ponte, P. (2002). How teachers become action researchers and how teacher educators become their facilitators. *Educational Action Research* **10**(3), 399–423.
- Ponte, P. (2007). Postgraduate education as platform: A conceptualisation. In van Swet, J., Ponte, P., and Smit, B. H. J.

- (eds.) *Postgraduate Programmes as Platform: A Research-Led Approach*, pp 19–36. Rotterdam: Sense Publishers.
- Ponte, P., Ax, J., Beijard, D., and Wubbels, T. (2004). Teachers' development of professional knowledge through action research and the facilitation of this by teacher educators. *Teaching and Teacher Education* **20**, 517–588.
- Ross, J. A., Rolheiser, C., and Hogaboam-Gray, A. (1999). Effects of collaborative action research on the knowledge of five Canadian teacher-researchers. Teachers learning from other teachers. *Elementary School Journal* **99**(3), 255.
- Schön, D. (1987). *Educating the Reflective Practitioner*. San Francisco, CA: Jossey Bass Publishers.
- Somekh, B. (2006). *Action Research: A Methodology for Change and Development*. Buckingham: Open University Press.
- Stenhouse, L. (1975). *An Introduction to Curriculum Research and Development*. London: Heinemann Educational Books.
- Stenhouse, L. (1983). *Authority, Education and Emancipation*. London: Heinemann Educational Books.
- Tabachnick, B. R. and Zeichner, K. L. (eds.) (1991). *Issues and Practices in Inquiry-Oriented Teacher Education*. London: Falmer Press.
- Zeichner, K. (2007). The value of educational action research to building and sustaining a high quality education system. *International Academic Conference of Action Research and Teacher Education*, 5–8 December Taiwan: Taitung.
- Zeichner, K. M. and Noffke, S. E. (2001). Practitioner research. In Richardson, V. (ed.) *Handbook of Research on Teaching*, pp 298–330. Washington, DC: American Educational Research Association.
- Atweh, B., Kemmis, S., and Weeks, P. (1998). *Action Research in Practice. Partnerships for Social Justice*. London: Routledge.
- Campbell, A. and Groundwater-Smith, S. (2007). *An Ethical Approach to Practitioner Research*. London: Routledge.
- Coghlan, D. and Brannick, T. (2005). *Doing Action Research in Your Own Organization*. London: Sage Publications.
- Elliott, J. (1991). *Action Research for Educational Change*. Buckingham: Open University Press.
- Fielding, M. and Bragg, S. (2003). *Students as Researchers: Making a Difference*. Cambridge: Pearson Publishing.
- Kemmis, S. and McTaggart, R. (1988). *The Action Research Planner*, 3rd edn. Victoria: Deakin University.
- Kemmis, S. and McTaggart, R. (1981/1992). *The Action Research Planner*, 3rd edn. Victoria: Deakin University.
- McLaughlin, C., Black-Hawkins, K., McIntyre, D., and Townsend, A. (2008). *Networking Practitioner Research*. London: Routledge.
- Ponte, P. and Smit, B. H. J. (2007). *The Quality of Practitioner Research. Reflections on the Position of the Researcher and the Researched*. Rotterdam: Sense Publishers.
- Somekh, B. and Noffke, S. (2009). *Handbook of Educational Action Research*. London: Sage Publications.
- Zeichner, K. M. and Noffke, S. E. (2001). Practitioner research. In Richardson, V. (ed.) *Handbook of Research on Teaching*, pp 298–330. Washington, DC: American Educational Research Association.

Further Reading

- Anderson, G. and Herr, K. (1999). The new paradigm wars: Is there room for rigorous practitioner knowledge in schools and universities? *Educational Researcher* **28**(5), 12–21 and 40.

Contemporary Approaches to Teacher Professional Development

H Borko, Stanford University, Stanford, CA, USA

J Jacobs, University of Colorado, Boulder, CO, USA

K Koellner, University of Colorado, Denver, CO, USA

© 2010 Elsevier Ltd. All rights reserved.

What Does High-Quality Professional Development Look Like in Theory?

Teacher professional development (PD) has become very tightly connected to many school-improvement efforts around the world. In the past several years, there has been an increasingly urgent, perceived need for more PD opportunities, along with assurance that the programs are high quality and effective. Guskey (2000) argued, "Never before in the history of education has greater importance been attached to the professional development of educators" (p. 3). Indeed, as Feiman-Nemser (2001) and others have claimed, if we want schools to offer more powerful learning opportunities for students, we must offer more powerful learning opportunities for teachers – opportunities that are grounded in a conception of learning to teach as a lifelong endeavor and designed around a continuum of teacher learning. Adopting this expanded view of teacher learning and professional practice, educational reforms in all disciplines and across all grade levels stress the need for educators to take part in PD programs that increase their knowledge, improve their practice, and ultimately foster student learning and achievement gains.

Recent shifts in the prominence of ideas about the nature of cognition, learning, and teaching – from behavioral to cognitive to situative – are, by now, well known. These changes have been accompanied by parallel shifts in ideas about teacher learning and PD. Most experts in the field advocate moving away from an in-service training model, where teachers are expected to learn a clearly defined body of skills through a well-specified process, often delivered in one-shot workshops or courses taught away from the school premises. These traditional approaches are generally viewed as overly fragmented, not connected closely enough to classroom practice, and out of alignment with current theories of learning and school reform. They are being replaced by approaches that are more closely aligned with constructivist and situative theories and reform efforts; specifically they are grounded in classroom practice and involve the formation of professional learning communities.

Stein *et al.* (1999) provided a clear and concise overview of the contrasts between traditional in-service staff development and what they call the new model of PD. Regarding traditional in-service approaches, they wrote,

"These forms of professional development were designed to support a paradigm of teaching and learning in which students' roles consisted of practicing and memorizing straightforward facts and skills, and teachers' roles consisted of demonstrating procedures, assigning tasks, and grading students" (p. 238). By contrast, "The new paradigm for professional development represents a clear departure from the use of workshops to teach 'techniques' toward the use of multiple professional development strategies to build teacher capacity to understand subject matter, pedagogy, and student thinking" (p. 263). **Table 1** presents Stein, Smith, and Silver's summary of the traditional and new approaches along four dimensions and illustrates that there are extensive discontinuities in all of the areas considered.

Hargreaves (2000) provided a historical perspective on the nature of teaching and PD that complements Stein, Smith and Silver's analysis. He described the present state of affairs as follows:

In the still emerging age of the collegial professional, there are increasing efforts to build strong professional cultures of collaboration to develop common purpose, to cope with uncertainty and complexity, to respond effectively to rapid change and reform, to create a climate which values risk-taking and continuous improvement, to develop stronger senses of teacher efficacy, and to create ongoing professional learning cultures for teachers that replace patterns of staff development which are individualized, episodic and weakly connected to the priorities of the school. (Hargreaves, 2000, p. 166)

The focus in most of the current PD literature is on providing a long-term, inquiry or learner-centered structure that supports teachers as they collaboratively develop the professional knowledge they need to use in their own context. This new vision of professional learning communities as a structure for PD is closely connected to calls for instructional reform, both generally and in specific content areas. In fact, there is widespread agreement that standards for student learning, instructional practice, and teacher PD should be seen as interrelated components of broad-based educational reform.

Several educational scholars have produced lists of principles or features of high-quality PD, based on literature reviews and accounts of successful PD programs. A comparison of their lists suggests that there is a growing

Table 1 Characteristics of old versus new paradigms for professional development

<i>Inputs to design process</i>	<i>Traditional in-service staff development</i>	<i>New model of professional development</i>
Strategies	Focus on activities (techniques, ideas, and materials) Dominant formats are workshops, courses, and seminars Short duration with bounded personal commitments	Focus on building capacity to understand subject matter and guide students' development of concepts Uses a variety of formats including the provision of in-class support and scaffolding of teacher participation in practice-related efforts (e.g., grade-level meetings, after-school meetings) Longer duration with more open-ended personal commitments
Knowledge and beliefs	Teacher educator sets the agenda Theories of teacher learning based on the psychology of the individual Translation of new knowledge to classroom is a problem to be solved (usually by the teacher)	Iterative co-construction of agenda by teachers and professional developer over time Theories of learning that include social and organizational factors Challenge is to scaffold learning that is both immediately relevant to practice and builds a more generalized knowledge base
Context	Particularities of context not factored into staff development Takes place away from schools, classrooms, and students	Particularities of context play an important role in shaping professional development Takes place in a variety of locations, at least some of which occur in schools and classrooms
Critical issues	Focus is on developing the teacher (teachers participate as individuals) Leadership training not an issue	Focus is on developing the instructional program and the community in addition to the teacher (teachers participate as an organizationally cohesive unit) Leadership training is a big issue

Reprinted from Stein, M.K., Smith, M.S., and Silver, E.A. (1999). The development of professional developers: Learning to assist teachers in new settings in new ways. *Harvard Educational Review* 69(3), 237–269, with permission from Harvard Education publishing Group.

consensus within the field regarding the central features of PD that are effective in improving teaching practice. Interestingly, this emerging consensus does not mean that the agreed-upon features are broadly defined or few in number. On the contrary, the lists of features are relatively extensive and include specific goals, activities, and practices. As the literature and research bases continue to grow over time, it is reasonable to expect even more specificity in this area. The next section highlights conclusions from several widely cited reports detailing the features of high-quality PD.

Features of High-Quality PD (in Theory and in Practice)

After reviewing the literature, six reports to focus on for a more extensive discussion and for comparative purposes were selected (Darling-Hammond and McLaughlin 1995; Hawley and Valli 2000; Knapp 2003; Putnam and Borko 1997, 2000; and Wilson and Berne 1999). Reports that considered PD in general, rather than in particular subject areas were selected. In addition, reports that collectively span a relatively long period of time, from early leaders in the field to recent publications that capture much of the current thinking were chosen. There are many other excellent reports in the literature that have been left out, either in the interest of space or because they did not meet

the selection criteria. It is also important to note that, for the most part, these reports do not provide evidence of student learning to support claims about effectiveness. This limitation is addressed later.

Table 2 presents a synthesis of the characteristics of high-quality PD that were noted in at least three of the six reports that were selected. The table is organized in two sections. The first section includes characteristics related to the content of high-quality PD; the second includes characteristics related to the process and structure of high-quality PD. These characteristics are discussed and elaborated in the following.

Content Characteristics of High-Quality PD

Two features stand out as content characteristics of high-quality PD: the content should be situated in practice and it should be focused (at least in part) on students' learning. Situating the PD content in the practice of teaching helps to ensure that what teachers are learning is relevant to their classroom practice. Specifically, high-quality PD engages teachers in inquiry about the concrete tasks of teaching, assessment, observation, and reflection, and provides them with the opportunity to make connections between their learning and their classroom instruction.

Three reports also noted the importance of focusing PD on students' thinking and learning. These reports indicate that PD activities should help teachers learn

how to elicit and interpret their students' ideas, examine student work, and use what they learn about students' ideas and work to inform their instructional decisions and actions. PD that involves teachers' close consideration of student reasoning opens up opportunities for collective inquiry about issues related to subject matter and pedagogy.

Process and Structure Characteristics of High-Quality PD

There is agreement among the reports that high-quality PD incorporates processes such as modeling preferred instructional strategies, engaging teachers in active learning, and building a professional learning community. When teacher educators model instructional strategies, PD participants have the opportunity to experience these strategies as learners, and then reflect on their learning and on the effectiveness of the strategies from the perspective of teachers. This type of approach is particularly important in times of reform, when teachers frequently are being asked to teach in ways that are substantially different from how they were taught or how they learned to teach.

All six reports included in **Table 2** reinforce the argument that the opportunity for teachers to participate actively and collaboratively in professional communities is an essential component of high-quality PD. By engaging teachers as active participants, PD providers acknowledge that learning is an active process wherein learners construct new understandings based on what they already know and believe. PD experiences are particularly effective when teachers participate in developing the learning opportunities, and work collaboratively to inquire and reflect on their practice. Respect and trust are important features of community development, enabling teachers to engage in discussions that are both supportive and challenging, and that maintain a balance between respecting individual community members and critically analyzing issues in their teaching.

Most of the reports specify that, to the extent possible, PD should be school based and integrated with other aspects of school change. High-quality PD experiences can be situated in multiple contexts, both in schools and offsite. Artifacts of practice, such as videotapes of lessons and student work, can be used to bring the classroom to the PD setting.

Finally, PD should consist of activities that are ongoing and sustainable over time, and that provide the opportunity for teachers to engage in cycles of experimentation and reflection. PD models with a cyclic approach might, for example, engage teachers in developing and enacting a lesson plan designed to meet the needs of all students, and then encourage them to individually and collaboratively reflect on their experiences. PD models that provide

teachers with the opportunity for such professional inquiry support their on-going learning and encourage change.

Research on Effective PD in Practice

Empirical evidence on the effectiveness of PD with these qualities is quite limited. Indeed, as Knapp (2003) cautioned, "The actual benefits of the asserted attributes of good professional development have rarely been examined systematically, either separately or in combination" (p. 120). Although few would argue with assertion of Hawley and Valli (2000) that evaluations of PD programs should incorporate multiple sources of information, both on the processes involved in implementing ideas learned through PD and on the outcomes for students, to date there are few evaluations of this sort. Despite the contemporary press for accountability for student outcomes, most of the research examines the relationship between PD and changes in teacher knowledge and instructional practices. Few research programs have explored the links between teacher learning, classroom practices, and student achievement.

Furthermore, as is true in the broader field of teacher education research, the research on teacher PD has been criticized for its limited scope and lack of scientific rigor. Scholars have run into difficulty when attempting to quantify the principles outlined in the more theoretically based literature. Most research to date consists either of intensive case studies on single programs or of self-report, survey data. In addition, the fields of mathematics and science are more heavily represented than other subject areas or other aspects of classroom practice. Nevertheless, this body of literature is widely read and cited, and it has been a motivating force as educational practitioners and policymakers push for increased PD opportunities.

At the same time, many educational scholars are committed to conducting research that will move the field forward. Borko (2004) suggested a three-phase research agenda for designing, implementing, and investigating scalable models of PD. This agenda provides a framework for guiding purposeful decisions about PD research. Phase 1 research projects take place at a single site and provide initial evidence that a PD program is feasible and can have positive impact on teacher learning. Typically, these proof-of-concept studies are relatively small; both qualitative and quantitative research methodologies are employed to document the design process and to study the impact on teachers' knowledge and practice. Phase 2 research projects build on and extend Phase 1 projects to determine whether a particular PD model is a scalable model. During Phase 2, researchers might study the implementation of a PD program in different settings or by different PD providers. They also might work with

Table 2 Characteristics of effective professional development^a

<i>Characteristics of effective PD</i>	<i>Darling-Hammond and McLaughlin (1995)</i>	<i>Hawley and Valli (2000)</i>	<i>Knapp (2003)</i>	<i>Putnam and Borko (1997, 2000^b)</i>	<i>Wilson and Berne (1999)</i>
<i>Content</i>					
PD content is situated in practice; addresses problems of practice.	PD engages teachers in concrete tasks of teaching, assessment, observation, and reflection that illuminate the processes of learning and development.		Content builds on teachers' pedagogical content knowledge (PCK).	Mantra: Situate teacher education in classroom practice. Ground teacher learning experiences in their own practice.	
Content of PD is focused on students' learning.	PD is connected to and derived from teachers' work with their students.	Content is focused on what students are to learn and how to support student learning; PD addresses impediments to and facilitators of student learning.	Specifically focused on high standards for students.		
<i>Process / structure</i>					
Preferred instructional practices modeled in PD.	Supported by modeling, coaching, and the collective solving of specific problems of practice.		Model preferred instructional practices (e.g., active learning), both in classrooms and in adult learning situations.	Mantra: Teacher educators should treat teachers as they expect teachers to treat students.	
Active teacher learning; teacher inquiry.	Grounded in inquiry, reflection, and experimentation that are participant-driven.	Teachers identify what they need to learn and, when possible, participate in the development of the learning opportunity and / or process to be used.		Mantra: Treat teachers as active learners who construct their own understandings. Empower teachers and treat them as professionals.	Teacher learning is activated, rather than bound and delivered; engage teachers as learners in areas their students will learn, but at a level more suitable to their learning.
Professional learning communities; collaborative learning environments.	PD is collaborative, involving sharing of knowledge among educators and a focus on communities of practice rather than on individual teachers.	PD provides learning opportunities that relate to individual needs but are, for the most part, organized around collaborative problem solving.	Collaborative and collegial learning environments.	Teachers need opportunities to participate in supportive professional learning / discourse communities in order to be successful in constructing new roles or changing their practice.	Communities of teacher learners who are redefining teaching practice; privilege teachers' interactions with one another, build trust and community while creating professional discourse that includes critique.

Continued

Table 2 Continued

<i>Characteristics of effective PD</i>	<i>Darling-Hammond and McLaughlin (1995)</i>	<i>Hawley and Valli (2000)</i>	<i>Knapp (2003)</i>	<i>Putnam and Borko (1997, 2000^b)</i>	<i>Wilson and Berne (1999)</i>
PD settings are appropriate to goals, often school based.	PD is connected to other aspects of school change.	PD is school based and integral to school-based operations. PD is integrated with a comprehensive change process.	School based when possible.	Mantra: Situate teacher education in classroom practice. PD experiences are situated in multiple contexts based on the goals of the PD.	
PD opportunities or models are ongoing and sustainable	PD is sustained, ongoing and intensive.	PD is continuous and ongoing, with follow-up and support for further learning. PD includes support from external sources to provide resources and outside perspectives	Rigorous cumulative opportunities for learning over time.		

^aWe use general terminology to describe characteristics of effective PD (i.e., cells in the first column), whereas we have attempted to use the language of the specific authors within the other cells to capture their voice and meaning.

^bThe 1997 chapter and 2000 journal article by Putnam and Borko both address teacher learning through professional development, but with somewhat different foci. Entries in this column represent a combination of ideas from the two papers.

nonvolunteer teachers or examine student achievement. Successful results from a Phase 2 program of research thus provide evidence that a full scale-up of the PD program would be useful to the field. Phase 3 research projects involve a comparison of multiple, well-defined PD programs, each enacted at multiple sites. Central research issues in these large-scale studies include comparing the resource requirements for the PD programs and their impact on teacher and student learning.

Below we provide examples of Phase 1 research on the PD of mathematics teachers, and of Phase 2 research on the PD of writing teachers. These examples illustrate key features of each research phase. To the best of our knowledge, no Phase 3, full-scale comparative studies of PD models have been conducted to date. However, there is another category of research that entails larger-scale analyses of multiple approaches to PD (but without detailed information about program characteristics or resource requirements); the literature in that area is discussed as well. Some emerging uses of technology in PD are also addressed.

An Example of Phase 1 Research on the PD of Mathematics Teachers

Early phases of development and research on PD models often use a design experiment approach to document the processes of the PD and the impacts on teacher learning. In general, these types of projects have focused

on implementation of a PD program by the designer at a single site, typically with motivated volunteers (Fishman *et al.*, 2003). Our own research – involving the development and implementation of a model of mathematics PD called the problem-solving cycle (PSC) – provides an example of Phase 1 activities.

The PSC model is an iterative, long-term approach to PD, with the goals of increasing teachers' knowledge of mathematics for teaching, improving their instructional practices, and fostering student achievement gains (Koellner *et al.*, 2007). Each iteration of the PSC consists of three interconnected workshops organized around a rich mathematical task. This common experience provides a structure within which the teachers can build a supportive community that encourages reflection on selected mathematical concepts, student thinking, and instructional practices. Central activities for participants include collaboratively solving the mathematical task; teaching it to their students; and exploring the teachers' instructional strategies and their students' mathematical reasoning. Video from the lessons and other artifacts of practice are used to situate the teachers' learning opportunities in the context of their work. The PSC is intended to be implemented by a knowledgeable facilitator, who helps to promote a professional community that supports the individual and collective growth of participating teachers.

Our research utilized a design experiment approach, with a group of eight to ten middle school teachers over

2.5 years. We documented the processes involved in developing and implementing the PSC model, as well as its substantial impact on teachers' knowledge and instructional practices. Our analyses revealed changes in patterns of participation in the workshops over time and suggested connections between teachers' experiences in the workshops and their professional learning. For example, analysis of workshop discussions revealed that over time, teachers talked in an increasingly focused, in-depth, and analytical manner about specific issues related to teaching and learning. On an assessment of algebra knowledge, significant increases in both the number of correct answers and the number of solution strategies used by each teacher were found. In addition, changes in classroom mathematics instruction over the duration of the project were observed. These changes reflected pedagogical topics, such as improving group dynamics and using more effective questioning strategies, emphasized in the PSC workshops. Thus, the findings indicate that, under ideal conditions (i.e., when implemented by the designer, with motivated volunteers), the PD model is feasible and can have a positive impact on teacher learning.

An Example of Phase 2 Research on the PD of Writing Teachers

After initial design research on a PD program many additional questions about the intervention must be answered before it is reasonable to allocate resources to a full scale-up evaluation. Such questions include: (1) What is the impact on student achievement? (2) Is it feasible on a larger scale (e.g., an entire school district) and with nonvolunteer teachers? (3) Can it be adapted to meet the needs and conditions of the local sites? (4) Are the PD materials and resources sufficient to ensure that multiple facilitators can implement the model while maintaining integrity with the designers' intentions? There is not one correct next-step in this phase of program development and research; rather there are multiple paths that can be explored in the ongoing study of a PD program.

Over the past three decades, various investigations of the National Writing Project (NWP) have explored several of these paths. Lieberman and Wood (2003) described the NWP as "arguably the most successful K-12 professional development project ever in the United States" (p. 187). The NWP has had a far-reaching influence on teachers and schools across the United States, where there is a steadily growing NWP network of nearly 200 university- and college-based sites. Initiated in the mid-1970s as the Bay Area Writing Project, as of 2003 the NWP had served approximately 3.5 million teachers. In partnership with schools or school districts, NWP sites host a 5-week summer institute for teachers in all grade levels and disciplines. During the following academic year, many of

these teachers become teacher consultants and facilitate workshops for their colleagues.

A core set of practices undergirds the NWP summer institutes. Designed to situate the teachers' learning in their own writing and classroom practices, these tasks guide teachers in demonstrating their writing instructional activities, studying theory and research about writing instruction, and immersing themselves in writing. However, by design, the focus of the institutes changes over time to meet the evolving needs of teachers and their students. Thus, a number of early institutes focused on prewriting; later ones addressed the use of rubrics; and more recent topics have included standardized assessments, writing with special populations, and technology. In addition, although the NWP sites share a national program model, their design and delivery vary according to local needs, reform priorities, school conditions, and research contexts. Thus, while there has been an ongoing, wide-ranging, and comprehensive evaluation effort, this effort is constrained by the intentionally flexible nature of the NWP.

A number of studies have investigated the impact of the NWP on teachers and students. In their review of this research, Pritchard and Honeycutt (2005) concluded that the positive effects on teachers' instructional practices have been well established. For example, several studies involving classroom observations of NWP-trained teachers indicate that the teachers gained new teaching strategies, used a greater variety of strategies, and devoted more time to writing instruction. Self-report data from teachers who participated in the NWP institutes suggest that they developed a valuable professional network, changed their philosophies about teaching writing, and acquired knowledge and skills related to teaching writing effectively. Of the published studies examining student work, central findings also favor the NWP approach over traditional approaches.

Larger-Scale Analyses of Effective PD

Although we are not aware of any studies that fit Borko's (2004) description of a Phase 3 project, a few researchers have looked across multiple PD programs to ascertain their effectiveness. As one example, Garet *et al.* (2001) surveyed a nationally representative sample of teachers who attended mathematics and science PD programs that received funding from the Eisenhower Professional Development Program. Although their survey covered a relatively large number and wide variety of PD programs, only a few teachers were sampled from each program; also, information about program effectiveness was based solely on teachers' self-reported increases in knowledge and skills and changes in classroom practice. Analyses examined three structural and three core features of the PD programs – features that the researchers identified through a literature review as best practices in teacher PD. The structural

features were form (e.g., study group, workshop), duration (e.g., number of hours, time span), and participation (e.g., collectively from the same school or grade, or individually). The core features were degree of content focus, active learning opportunities, and promotion of coherence (e.g., alignment with state standards).

The researchers found that all six features were related to teachers' self-reported changes in knowledge, skills, and classroom practices. These results support the notion that to be effective, PD models should take into account the array of features identified in the literature as best practices. However, the data revealed that many of the PD programs studied did not incorporate features of high-quality PD. Other analyses of the Eisenhower program have produced complementary findings.

Uses of New Technology in PD

Recent technological trends are having a strong impact on the field of teacher PD. Many new PD models are incorporating various technology-related components, including digital libraries, web-based virtual learning environments, and online and electronic conferencing features. Various forms of online discussion forums are particularly popular. In synchronous conversations, such as chat rooms, participants are present at the same time and respond electronically in real-time to one another. Virtual learning environments that feature asynchronous interactive tools, such as bulletin boards, enable teachers to participate in discussions anytime, anyplace.

Particularly as PD programs seek to scale-up or enter Borko's (2004) Phase 2 as described above, developers are increasingly turning to these contemporary, innovative technologies. As Goldman (2001) argued, "Many forms of electronic technology can overcome time and place constraints and provide the means to reach large numbers of individuals, potentially at costs lower than those associated with the physical presence of professional development personnel" (p. 21). PD that incorporates these technologies can draw upon resources not available locally, provide just-in-time work-embedded support, and accommodate individual teachers' busy schedules.

Currently, a broad array of online PD programs is serving large numbers of teachers. These programs are diverse in their purposes and goals, content areas, pedagogical approaches, and delivery methods. Many incorporate technology-related components that support content and process characteristics of high-quality PD—for example, collaborative virtual learning communities, interactive media to facilitate active teacher participation, video cases that address problems of practice, and digitized work samples to foster exploration of student thinking. Hybrid PD models that feature both online communication

(either asynchronous or synchronous) and face-to-face components are also gaining in popularity.

As one example, the PBS TeacherLine site offers over 100 online PD courses to preK-12 teachers in the fields of mathematics, reading/language arts, science, instructional technology, and instructional sciences. These relatively short courses (6–10 weeks) were developed with input from K-12 educational leaders and research experts and include online discussions with trained facilitators who have backgrounds in the subject matter. Designed to tie to classroom practice and link content to local and national standards, many of the courses feature streaming video of classrooms and digitized versions of student work. Research on TeacherLine includes formative and summative evaluations based on self-report data from participants and facilitators, expert-panel review of selected courses, and an assessment of the quality of online discussion board communications. Results indicate that facilitator quality related directly to participants' ability to apply course content to their classroom practice, and to their satisfaction with the course (Ramsdell *et al.*, 2006).

Another, quite different, example is Tapped In, an online environment for educational professionals that is designed to promote a community of practice. Drawing on rapidly expanding technological advances, Tapped In enables a professional online community that extends beyond local sites. Synchronous and asynchronous discussions, coupled with support tools, such as white boards, sticky notes, and web page viewers, allow teachers to work and learn with a diverse group of colleagues and experts. Schlager and Fusco (2003) argued, "Tapped In has been quite successful in achieving its original goal of bringing together and forging new relationships among education practitioners, providers, and researchers from around the world on a daily basis. Thousands of different people log in each month to engage in activities that include course and workshop sessions, group meetings, and public discussions spanning a wide range of K-12 topics" (p. 204). As of July 2006 there were 20 000 Tapped In members, and the monthly log-in rate was approximately 10–20% of the membership. The developers have documented their conceptual framework and design strategy in detail, and have begun to collect and analyze data on the impact of participants' Tapped In experiences on their professional lives.

Dede *et al.* (2006) reviewed nearly 400 empirical studies of online, face-to-face, and hybrid PD programs. Most of the literature reports program evaluations focused on the effects of program design, delivery, and use on the formation of teacher learning communities, discourse patterns, and levels of participation. Few studies examined changes in teachers' knowledge or skills, and even fewer addressed the impact on student learning. This conclusion supports Borko's (2004) premise that the research on contemporary PD (whether involving new technology or not)

falls mostly in the Phase 1 category. Dede and other leaders in the field have identified the need for more well-designed empirical research studies, and we expect to find a good deal of movement in this regard in the near future.

Conclusion

The field of PD is moving forward at a relatively fast pace. A majority of teachers in the United States are engaging in PD, and there is a push for more PD opportunities across the country. Similar increases in the level of interest and support for teacher PD can be seen throughout the world. Research in this field offers a relatively focused and agreed-upon direction for PD, which is consistent with prevailing theories about the nature of learning and cognition. At this point, much of the focus has shifted from creating theories of high-quality PD, to considering issues related to the design of PD programs based on current theories. Presently, a number of PD programs based on the current theories have been implemented, but few have been expanded beyond the proof-of-concept phase. Only a small research base exists on the impact of contemporary approaches to PD on teachers' learning and classroom practices, and the research base on student achievement is even smaller. However, with steady advances in the design, implementation, and evaluation of PD programs, and the expanding use of new technology, the research base in this field is likely to expand dramatically over the next several years.

See also: Professional Development of Teacher Educators.

Bibliography

- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher* **33**(8), 3–15.
- Darling-Hammond, L. and McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan* **76**(8), 597–604.
- Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record* **103**, 1013–1055.
- Fishman, B. J., Marx, R. W., Best, S., and Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education* **19**, 643–658.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B., and Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal* **38**(4), 915–945.
- Goldman, S. R. (2001). Professional development in a digital age: Issues and challenges for standards-based reform. *Interactive Educational Media* **2**, 19–46.
- Guskey, T. R. (2000). *Evaluating Professional Development*. Thousand Oaks, CA: Corwin.
- Hargreaves, A. (2000). Four ages of professionalism and professional learning. *Teachers and Teaching* **6**(2), 151–182.
- Hawley, W. D. and Valli, L. (2000). Learner-centered professional development. *Phi Delta Kappa Center for Evaluation, Development, and Research*. Research Bulletin Number 27.
- Knapp, M. S. (2003). Professional development as policy pathway. *Review of Research in Education* **27**(1), 109–157.
- Koellner, K., Jacobs, J., Borko, H., et al. (2007). The problem-solving cycle: A model to support the development of teachers' professional knowledge. *Mathematical Thinking and Learning* **9**(3), 271–300.
- Lieberman, A. and Wood, D. (2003). *Inside the National Writing Project*. New York: Teachers College Press.
- Pritchard, R. J. and Honeycutt, R. L. (2005). The process approach to writing instruction: Examining its effectiveness. In MacArthur, C. A., Graham, S., and Fitzgerald, J. (eds.) *Handbook of Writing Research*, pp 275–290. New York: Guilford Press.
- Putnam, R. and Borko, H. (1997). Teacher learning: Implications of new views of cognition. In Biddle, B. J., Good, T. L., and Goodson, L. F. (eds.) *The International Handbook of Teachers and Teaching*, pp 1223–1296. Dordrecht: Kluwer.
- Putnam, R. and Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher* **29**(1), 4–15.
- Ramsdell, R., Rose, R., and Kadera, M. (2006). PBS Teacherline and Concord Consortium's Seeing Math Secondary: Scaling up a national professional development program. In Dede, C. (ed.) *Online Professional Development for Teachers: Emerging Models and Methods*, pp 69–88. Cambridge, MA: Harvard Education Press.
- Stein, M. K., Smith, M. S., and Silver, E. A. (1999). The development of professional developers: Learning to assist teachers in new settings in new ways. *Harvard Educational Review* **69**(3), 237–269.
- Wilson, S. M. and Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. In Iran-Nejad, A. and Pearson, P. D. (eds.) *Review of Research in Education*, vol. 24, pp 173–209. Washington, DC: American Educational Research Association.

Further Reading

- Ball, D. L. and Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In Darling-Hammond, L. and Sykes, G. (eds.) *Teaching as the Learning Profession: Handbook of Policy and Practice*, pp 3–32. San Francisco, CA: Jossey-Bass.
- Darling-Hammond, L. and Sykes, G. (eds.) (1999). *Teaching as the Learning Profession: Handbook of Policy and Practice*. San Francisco, CA: Jossey Bass.
- Day, C. and Sachs, J. (2004). *International Handbook on the Continuing Professional Development of Teachers*. Maidenhead: Open University Press.
- Dede, C. (ed.) (2006). *Online Professional Development for Teachers: Emerging Models and Methods*. Cambridge, MA: Harvard Education Press.
- Elmore, R. F. (2002). *Bridging the Gap between Standards and Achievement: The Imperative for Professional Development in Education*. Washington, DC: Albert Shanker Institute.
- Guskey, T. R. and Huberman, M. (eds.) (1995). *Professional Development in Education: New Paradigms and Practices*. New York: Teachers College Press.
- Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis* **15**(2), 129–151.
- OECD (Organization for Economic Cooperation and Development) (1998). *Staying Ahead: In-Service Training and Teacher Professional Development*. Paris: OECD.
- Schlager, M. and Fusco, J. (2004). Teacher professional development, technology, and communities of practice: Are we putting the cart before the horse? In Barab, S., Kling, R., and Gray, J. (eds.) *Designing for Virtual Communities in the Service of Learning*, pp 120–153. Cambridge, UK: Cambridge University Press.
- Sparks, D. (2002). *Designing Powerful Professional Development for Teachers and Principals*. Oxford, OH: National Staff Development Council.

- Tisher, R. P. and Wideen, M. F. (1990). *Research in Teacher Education: International Perspectives*. Bristol, PA: Falmer Press.
- Villegas-Reimers, E. (2003). *Teacher Professional Development: An International Review of the Literature*. UNESCO: International Institute for Educational Planning.
- Yoon, K. S., Duncan, T., Lee, S.W.-Y., Scarloss, B., and Shapley, K. (2007). Reviewing the evidence on how teacher professional development affects student achievement (Issues and Answers Report, REL 2007-033). Washington, DC: U.S. Department of Education, Institute of Education Science, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Relevant Websites

- www.nwp.org – National Writing Project. Improving writing and learning in the Nation's Schools.
- www.teacherline.pbs.org – Professional development for PreK-12 educators. A service of PBS Teachers.
- www.colorado.edu – Supporting the Transition from Arithmetic to Algebraic Reasoning at the University of Colorado–Boulder.
- www.tappedin.org – The online workplace of an international community of education professionals.

School Development for Teacher Learning and Change

P Sleegers, University of Twente, Enschede, The Netherlands

K Leithwood, University of Toronto, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

In the last decades, schools have been confronted with ongoing large-scale educational reform efforts and restructuring movements as strategies to improve the quality of education and to implement educational innovations. The educational reforms are, among other things, based on new forms of learning and aimed at the arrangement of learning environments intended to stimulate self-regulated, reflective, independent, authentic, and social-interactive learning. An important aim of these restructuring changes is to create opportunities for change at the local level and to enhance capacity building for learning in schools.

Research in educational organizations, however, indicates that changing teachers' practices is extremely difficult to accomplish. In their efforts to understand the complex nature of educational change, researchers have reconceptualized teacher change by using perspectives in which teacher learning embedded in the school is considered a key component of successful school improvement. As a consequence, the capacity of schools to enhance the professional learning of teachers and transform large-scale reform into accountable, learner-oriented, teaching practice has become a major focus in recent research in educational change (Fullan, 1999; Hopkins, 2001; Sleegers *et al.*, 2005; Smylie and Hart, 1999; Toole and Louis, 2002).

In this article, we examine research about the relationship between structural and cultural aspects of the school organization and teacher learning and change. We bring together two views of change that dominate the literature about school improvement and educational change. The first view, which we term the inside view, focuses on the capacity of schools to transform themselves into supportive environments for teacher learning and change. The second view concerns the implementation of externally developed reform designs into schools, the outside view. While one literature often informs the other, they view change through opposite ends of the telescope.

Underlying Assumptions

Insider and outsider approaches to change are premised on quite different assumptions. These assumptions are well captured in Chin and Benne's (1969) distinction between planned change, empirical-rational approaches to change, and normative-reeducative approaches. The empirical-rational strategy focuses on research-based models for change which assume that teachers, as rational

human beings, will implement changes in their classrooms which are demonstrated to improve student learning. The normative-reeducative conception of change, focusing on the professional growth of individuals who make up the system and on the problem-solving capacities of the system itself, assume that personal sense-making and collective learning are keys to improvement.

Empirical-rational and normative-reeducative approaches not only differ in their assumptions about change, but also in the direction for change (for a more elaborate description of these strategies, we refer to Richardson and Placier (2001) who have also used this framework for their thorough review of research into teacher change). While empirical-rational strategies view teachers as mere recipients and consumers of new behavior, beliefs, and programs of researchers, and as policymakers and educators outside the school, the assumption in the normative-reeducative approach is that change is part of a larger process of making sense of situations in which teachers work and live through individual and collective reflection on beliefs and practices. We were able to organize a number of studies about school development and teacher learning and change premised on Chin and Benne's (1969) normative-reeducative approach based on their underlying views on change. These included a wide range of studies about organizational learning, learning organization, and professional communities. The emphasis on organizational learning in schools, schools as learning organizations, and schools as professional communities can be regarded as an attempt to find alternatives for the disappointing results of restructuring schools and the school system. These studies often use a system theory on change that links structural, cultural, and political dimensions of school-workplace environments to professional learning. In these studies, organizational conditions (including leadership) are considered as the main levers of a school's capacity to change and as a prerequisite for linking teachers' professional development to school development. In this article we look at learning in schools and the school's capacity to change through the lens of organizational learning. We prefer the concept of organizational learning instead of the concept of learning organization, because the literature on learning organization is prescriptive, practitioner-oriented, and lacks scientific rigor (see Tsang, 1997). Therefore, we restrict our review to organizational learning in schools and conditions fostering this learning. We begin with a summary of this research.

There is also a considerable body of research premised on a rational–empirical conception of change. During the last 10 years, schools in US and other countries have been involved in implementing research-based school-reform designs that are developed by an organization external to the school or district, also known as comprehensive school-reform designs (e.g., Success for All, Comer School Development Program, Coalition for Essential Schools, Accelerated Schools, Modern Red School House, ATLAS Communities, Different Ways of Knowing and Core Knowledge). The use of these empirical–rational strategies for educational change and the scaling up of these designs can be explained by the current and growing pressure to change, including the push for strong terms of accountability and systematic reforms, and beliefs about the effectiveness of evidence-based decision making. Inspired by these developments, researchers have conducted studies of the effects and the transfer of these models to multiple settings (scaling up).

The Inside View: Organizational Learning in Schools

The concept of organizational learning in schools is used by researchers to gain more insights about the internal innovative capacity of schools to adapt to the dynamics of educational change. The idea that organizations can learn is hardly new. As early as the 1970s, organizational theorists were writing about how organizations could renew and improve their processes, products, and services; in other words, how they could learn (Argyris and Schön, 1974). Although interest in the concept of organizational learning originated in the corporate sector, it has attracted attention in the educational world as well (Leithwood *et al.*, 2001).

Schools are confronted with a continuous stream of changes in their environments (e.g., demographic changes, large-scale educational innovations, and sociocultural renewal). In addition, they face multidimensional restructuring demands to which they must respond. Furthermore, schools are under considerable external pressure because of tightened output controls introduced by accountability policies. Simultaneously, schools are expected to improve not only their internal change capacity but also their instruction through their professional development initiatives. Strengthening schools' internal conditions to promote organizational learning is considered an important prerequisite for addressing this onslaught of expected changes. Furthermore, learning is also expected to reduce the alienation and challenges to educators' motivations as the bureaucratic features of schools and their dysfunctional effects are reduced.

A common thread found in the many definitions of organizational learning is the focus on innovation and improvement through the acquisition of better knowledge and understanding about learning. Organizational learning can be defined as the activities through which organizational members construct new knowledge, or reconstruct existing knowledge in order to improve the functioning of individual organizational members and the organization as a whole (Leithwood *et al.*, 2001). Leithwood and Louis (1998) distinguish different levels of learning in schools, including the learning of individuals in organizational contexts; small group or team learning that occurs within subunits of the organization; and collective learning of the organization as a whole (e.g., new policies or procedures). Individual learning is considered to be necessary but insufficient for organizational learning; and organizational learning is more than the sum of all individual learning. Staff members share their knowledge and expertise by cooperating and exchanging information. These collaborative practices, for example, may focus on developing instructional material and curricula during which teachers exchange experiences, investigate each others' practices, develop new approaches, and learn from each other. The learning of staff within schools is considered to be good on the assumption that higher levels of individual and organizational learning contribute to better organizational functioning.

School improvement research indicates that teacher learning is crucial for improving instructional practices. Moreover, cooperation can contribute importantly to improving instructional quality and, hence, student achievement (Clement and Vandenberghe, 2000; Geijsel *et al.*, 2001; Louis *et al.*, 1996). Positive associations also have been found between teachers' learning and the innovative functioning of schools.

While there are good theoretical reasons to expect that individual and organizational learning will improve classroom practices and school performance, there is little on the empirical side to justify this expectation. Available evidence points to important preconditions associated with the individual, the task, and the school organization. With respect to the individual, Mitchell and Sackney (2000) point to the individual capacity of staff to learn, to actively (re)construct, and to apply knowledge. Such capacities seem to be influenced by psychological factors, such as personal teaching efficacy, autonomy or perceived control, and teachers' commitment, emotions, and sense-making (Coburn, 2001; Rosenholtz, 1991; Spillane *et al.*, 2002; Sleegers *et al.*, 2002; Van Veen *et al.*, 2005). The characteristics of the task to be carried out may also play a role in how motivated are the staff to learn (e.g., the degree of task control and the extent of task variation).

Among the organizational conditions that influence learning among staff, the role of school leaders is key,

especially as it is inspired by the concept of transformational leadership (Geijsel *et al.*, 2003; Leithwood *et al.*, 1999; Leithwood and Slegers, 2006). The role and forms of teacher collaboration aimed at improving instruction and education is also an important organizational condition. Such collaboration can be promoted by creating opportunities for it, by developing the collaborative skills of teachers, and by rewarding teacher collaboration. The intensity of cooperation and learning among staff, as well as the development of the school as a whole are dependent on the degree to which schools create opportunities for teachers' professional learning (Clement and Vandenberghe, 2000; Slegers *et al.*, 2002). Moreover, organizational learning also depends on both the availability of relevant data and agreed-upon standards to assist in interpreting such data. Learning is only possible if school staff are provided with information concerning important school issues (e.g., developments in student performance or the extent of parental participation) (Leithwood *et al.*, 2001; Earl and Katz, 2006).

Finally, Marks *et al.* (2000), among many others, have pointed to the importance of freeing one's organization from traditional structures, empowering teachers in decision-making processes, and developing cultures which value shared responsibilities and values.

The attention that organization learning receives in the literature is in contrast to the amount of empirical research that is available. The many explorative studies that have been carried out do not unequivocally verify or falsify the relationships between variables as hypothesized in the literature. There are some indications that schools, with the characteristics mentioned above, do indeed promote teacher learning, and educational change takes place more easily in those schools. Whether schools, in which collective learning takes place, promote higher levels of student performance has not been studied systematically. Nor has much empirical attention been devoted to the factors which trigger learning, although it is acknowledged, in general, that particular events (e.g., a decrease in student enrolments) may lead to learning.

Besides this system-oriented view of collective learning in schools, a key contribution to the study of school culture and change has been the concept of professional learning community (PLC) (Mitchell and Sackney, 2000; Stoll *et al.*, 2006; Toole and Louis, 2002). The term integrates three robust concepts: a school culture that emphasizes professionalism (is client oriented and knowledge based), stresses learning, and places a high value on teachers' inquiry and connection. Although researchers use different key indicators and variables to describe and measure PLCs, the central idea is about the existence of a social architecture in school organizations that helps shape teachers' attitudes toward new pedagogies. The key to learning, from this perspective, is not

adaptation but creation and the free choice of individuals to participate in a social reality called organization and, through such participation, to learn. Learning in PLCs is conceptualized as dynamic and cyclical. The strict distinction between conditions and effects disappears (Imants *et al.*, 2001). As a result, the links between conditions for learning, learning communities, and school improvement are described as recursive. Collaboration, participation, transformational leadership, reflective dialog, and the like can be an input, throughput, or outcome of learning processes. These potential stimulants to learning become productive insofar as they prompt both individual and social sense-making on the part of teachers. For a more detailed description of the scholarly work in this field, we refer to the article on PLC in this encyclopedia.

The Outside View: Implementation of Comprehensive School Designs

Since the No Child Left Behind (NCLB) legislation (US Congress, 2001), educational leaders and policymakers in the US have devoted much of their energy to creating systems and procedures for measuring school performance and improving the achievement of low-performing students. This policy initiative, with a dominant press for external accountability, has prompted the initiation of hundreds of school-reform efforts especially in failing schools. The NCLB-driven policymakers have stressed evidence-based decision making and the need for strong evidence of program effects to be eligible for federal funding. In order to meet these powerful incentives to improve, low-performing schools have been searching for blueprints, tools, and expertise, which make change possible and are able to demonstrate their effects in scientifically reputable ways. Starting with the comprehensive school reform demonstration (CSR) program in 1997, over 380 comprehensive school reform (CSR) models encompassing virtually all aspects of school operations, including instruction, assessment, classroom management, professional development, parental involvement, school management, and curriculum, have been developed in the US and also introduced in other countries. Many of these models are indeed grounded in scientific research and offer schools the solutions to their improvement problems (for an overview, see Borman *et al.*, 2004). To receive funding, a CSR model must:

1. offer proven methods and strategies;
2. integrate comprehensive designs for effective schools with aligned components;
3. provide high-quality and continuous teacher and staff professional development and training;

4. include measurable goals for student performance and benchmarks;
5. be supported by school faculty, administrators, and staff;
6. provide support for staff, school faculty, and administrators;
7. stimulate parental and community involvement;
8. use high-quality external support from a comprehensive school-reform design team with experience and expertise in school-wide reform;
9. plan to evaluate the implementation of school reforms and the students' results achieved;
10. identify resources (federal, state, local, or private) to support and sustain the school reform; and
11. show evidence of having significantly improved the academic achievement of students in the program as compared with students who have not participated in the program (The Coalition for Comprehensive School Improvement, 2005).

All the early funded reform designs were initially implemented in one location (Memphis, TN, USA), but have since transferred to multiple settings. The use of this empirical-rational strategy for educational change and the scaling up of these designs play a central role in the next stage of NCLB. Scaling up is mostly defined in a traditional way as the deliberate extension of an externally designed CSR design that has been successful in one school to more teachers, schools, and districts (Coburn, 2003).

Although the principles, methods, and strategies of the funded CSR models must be evidence-based, research to date indicates that most do not have strong effects on student achievements (Berends *et al.*, 2002; Desimone, 2002). In the first serious review of evidence on this issue, Borman *et al.* (2003) reviewed 232 studies for their meta-analysis of CSR effects. They found modest-to-positive effects on the achievement scores for the CSR models examined. Three models – the Direct Instruction, the Success for All, and the School Development model – were classified by the researchers as having the strongest effects (effects sizes ranging from +0.15 to +0.21), while 17 of the 24 models produced insufficiently reliable or generalizable results. Recently, some studies using defensible numbers of CSR schools, control schools, independent researchers, and extended time periods were conducted (Borman *et al.*, 2005; Sterbinsky *et al.*, 2006). Using a randomized field trial of 41 schools to evaluate the first-year results of Success for All, Borman *et al.* found effects on one (word attack) of four posttest measures. The magnitude of the school-level effect using the Success for All intervention relative to control schools was equal to an effect size of +0.22 (2 months of additional learning). Using a quasi-experimental design, Sterbinsky *et al.* examined the effects of five different CSR models in ten schools. Results showed

that CSR schools had significantly higher gain scores than control students on word identification, passage comprehension, and oral reading (effect sizes vary from +0.31 to +0.44). With regard to yearly cohorts, only effect sizes were found for passage comprehension (+0.27) in year 3.

Although some effects are apparent by now, the evidence remains very thin and most of the studies conducted have been small scale, often conducted by developers themselves and lacking well-selected designs and sophisticated statistical analyses. Additionally, these effect studies do not increase our knowledge of key issues of educational reform, such as how these models are adopted by teachers, the conditions under which these models can be implemented successfully, the possibilities and boundaries of scaling up CSR models, and the sustainability of educational reforms. There is a growing body of work that tries to shed more light on these issues by studying the CSR implementation processes in schools and the possibilities of scaling up these reforms (Borman *et al.*, 2004; Datnow *et al.*, 2002; Desimone, 2002; Klinger *et al.*, 2006; Sterbinsky *et al.*, 2006). Results of these studies show that most of the CSR reforms were actively shaped and reshaped (co-constructed) by teachers, school leaders, and other local educators to accommodate the various goals, materials, and demands of the program to local context. These processes of mutual adaptation were required to meet the practical demands of everyday school life and the needs of students.

Mutual adaptation also occurs when the content of CSR models differs from the pre-existing beliefs and practices of teachers. In this case, teachers focus on surface manifestations rather than underlying pedagogic principles. The way that different local actors interpret and enact these reforms thus plays a very important role in their implementation and influences program fidelity. Such mutual adaptation has mostly caused frictions with highly prescribed models and insensitive design teams. Most developers of CSR models recommend the exact implementation of the model (program fidelity) using a technical-rational approach that sees teachers as passive implementers instead of collaborators. But inattention to the cultural and political dimensions of change seems to hinder successful implementation. Results also show that strong principal leadership, teacher buy-in, sufficient resources and time, and district support for the reform are important for successful implementation of the model. Furthermore, findings also suggest that schools with a high innovative capacity, collaborated cultures, and highly committed teachers are more successful in implementing reform designs.

Although more research is needed to increase our understanding of conditions fostering the implementation of CSR models, the current findings suggest that for successful implementation of CSR models, attention

should be paid to the interrelations between the way teachers make sense of externally developed models, the structural, cultural, and leadership dimensions of the school and the local context in which the school is embedded (Datnow *et al.*, 2002). Surprisingly, these findings also concur with the results from research into conditions fostering organizational learning as described in the first section of this article. Despite the empirical–rational assumptions used in most of the CSR literature and research on educational change, normative–reeducative change strategies seem to be needed to foster real change within a school. Moreover, the findings also seem to suggest that schools which already have a high innovative capacity and are able to transform their organization into a rich learning environment for teachers can integrate externally developed reforms into their current practice far more easily than schools with a low innovative capacity. Therefore, it is far easier to seed new practices successfully in schools which have fertile soils (Slavin, 1998). Given the fact that CSR models are often thought to offer powerful incentives (tools, blueprints, and expertise) for low-performing schools to improve their performances, these findings present challenges for those wanting to scale up these designs.

Conclusions

In this article, we reviewed the growing body of research in the relationship between school development and teacher learning and change. By using Chin and Benne's classic distinction between empirical–rational and normative–reeducative change strategies, we brought two views of change that dominate the literature about school improvement together. From the system or insider view on teacher change, a normative–reeducative conception of the process focuses on the study of the professional growth of individuals who make up the system and on the change capacities of the system itself. From the implementation or outsider view an empirical–rational approach focuses reformers on research-based models for change, which teachers, as rational actors, are expected to implement in their classrooms. Although these two views use different assumptions about the nature and direction of change, results of research show that they inform each other. Successful educational reform needs to be context specific, integrated with the school's capacity for development, and stimulate collective learning in schools. In order to be effective, we agree with Hopkins (2001) that finding a balance between normative–reeducative and empirical–rational approaches seems to be a better way than using the so-called one-size-fits-all approaches. The challenge is to find change strategies and approaches that fit the best with the existing capacities of the school, or at least, capacities it can realistically acquire.

Future research should focus on the interaction between school internal capacities and external reform programs. Special attention should be paid to issues of sustainability and scaling up as fundamental to the successful implementation of educational reforms. To increase our understanding of key conditions in schools and effective change strategies that can support teacher learning and change, researchers should focus on issues such as the extent to which reforms can go beyond superficial indicators of change to alter teachers' beliefs, norms, and enacted practices (depth), how change can be sustained over time (duration), and how it can be extended to more teachers, schools, and districts (breadth) (Coburn, 2003; Hargreaves and Fink, 2006).

See also: Professional Learning Community.

Bibliography

- Argyris, C. and Schön, D. A. (1974). *Theory in Practice: Increasing Professional Effectiveness*. San Francisco, CA: Jossey-Bass.
- Berends, M., Kirby, S. N., Naftel, S., and McKelvey, C. (2002). *Implementation and Performance in New American Schools: Three Years into Scale-Up*. Santa Monica, CA: Rand.
- Borman, G., Hewes, G., Overman, L. T., and Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research* **73**, 125–130.
- Borman, G. F., Slavin, R. E., Cheung, A., *et al.* (2005). Success for all: First-Year results from the national randomized field trial. *Educational Evaluation and Policy Analysis* **27**, 1–23.
- Borman, K. M., Carter, K., Aladjem, D. K., and LeFloch, K. C. (2004). In *Challenges for the Future of Comprehensive School Reform Research*, pp 109–150. Washington, DC: The National Clearinghouse for Comprehensive School Reform.
- Chin, R. and Benne, K. (1969). General strategies for effecting changes in human systems. In Bennis, W., Benne, K., and Chin, R. (eds.) *The Planning of Change*, 2nd edn., pp 32–59. New York: Holt, Rinehart and Winston.
- Clement, M. and Vandenberghe, R. (2000). Teachers' professional development: A solitary or collegial (ad)venture. *Teaching and Teacher Education* **16**, 81–101.
- Coalition for Comprehensive School Improvement (2005). Recent research on comprehensive school reform: What's working, what's needed. Nashville, TN. <http://www.improvingschools.org> (accessed December 2008).
- Coburn, C. E. (2001). Collective sense-making about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis* **23**, 145–170.
- Coburn, C. E. (2003). Rethinking scale: Moving beyond the numbers to deep and lasting change. *Educational Researcher* **32**, 3–12.
- Datnow, A., Hubbard, L., and Mehan, H. (2002). *Extending Educational Reform. From One School to Many*. London/New York: Routledge/Falmer.
- Desimone, L. (2002). How can comprehensive school reform models be successfully implemented. *Review of Educational Research* **72**, 433–480.
- Earl, L. and Katz, S. (2006). *Leading Schools in a Data-Rich World. Harnessing Data for School Improvement*. Thousand Oaks, CA: Corwin Press.
- Fullan, M. (1999). *Change Forces: The Sequel*. London: Falmer Press.
- Geijsel, F., Slegers, P., Leithwood, K., and Jantzi, D. (2003). Transformational leadership effects on teachers' commitment and effort toward school reform. *Journal of Educational Administration* **41**, 229–256.

- Geijsel, F. P., Slegers, P. J. C., van den Berg, R. M., and Kelchtermans, G. (2001). Conditions fostering the implementation of large-scale innovations in schools: Teachers' perspectives. *Educational Administration Quarterly* **37**, 130–166.
- Hargreaves, A. and Fink, D. (2006). *Sustainable Leadership*. San Francisco, CA: Jossey-Bass.
- Hopkins, D. (2001). *School Improvement for Real*. London: Routledge/Falmer.
- Imants, J., Slegers, P., and Witziers, B. (2001). The tension between organisational sub-structures in secondary schools and educational reform. *School Leadership and Management* **21**(3), 289–308.
- Klinger, J., Cramer, E., and Harry, B. (2006). Challenges in the implementation of success for all in four high-need urban schools. *Elementary School Journal* **106**, 333–350.
- Leithwood, K., Jantzi, D., and Steinbach, R. (1999). *Changing Leadership for Changing Times*. Buckingham: Open University Press.
- Leithwood, K. and Louis, K. S. (eds.) (1998). *Organizational Learning in Schools*. Lisse: Swets and Zeitlinger.
- Leithwood, K. and Slegers, P. (eds.) (2006). *Special Issue: Transformational School Leadership. School Effectiveness and School Improvement* **17**(2).
- Leithwood, K. A., Aitken, R., and Jantzi, D. (2001). *Making Schools Smarter*. Thousand Oaks, CA: Corwin Press.
- Louis, K. S., Marks, H. M., and Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal* **33**, 757–800.
- Marks, H. M., Louis, K., and Printy, S. M. (2000). The capacity for organizational learning. Implications for pedagogical quality and student achievement. In Leithwood, K. and Louis, K. S. (eds.) *Understanding Schools as Intelligent Systems*. Greenwood, MS: JAI.
- Mitchell, C. and Sackney, L. (2000). *Profound Improvement. Building Capacity for a Learning Community*. Lisse: Swets and Zeitlinger.
- Richardson, V. and Placier, P. (2001). Teacher change. In Richardson, V. (ed.) *Handbook of Research on Teaching*, pp 905–947. Washington, DC: American Educational Research Association.
- Rosenholtz, S. J. (1991). *Teachers' Workplace. The Social Organization of Schools*. New York: Teachers College Press.
- Slavin, R. E. (1998). Sand, bricks, and seeds: School change strategies and readiness for reform. In Hargreaves, A., Lieberman, A., Fullan, M., and Hopkins, D. (eds.) *International Handbook of Educational Change*, pp 1299–1313. Dordrecht: Kluwer Academic Publishers.
- Slegers, P., Bolhuis, S., and Geijsel, F. (2005). School improvement within a knowledge economy: Fostering professional learning from a multidimensional perspective. In Bascia, N., Cumming, A., Datnow, A., Leithwood, K., and Livingstone, D. (eds.) *International Handbook of Educational Policy*, pp 527–543. Dordrecht: Kluwer Academic Publishers.
- Slegers, P. J. C., Geijsel, F. P., and van den Berg, R. M. (2002). Conditions fostering educational change. In Leithwood, K. and Hallinger, P. (eds.) *Second International Handbook of Educational Leadership and Administration*, Part 1, pp 75–102. Dordrecht: Kluwer Academic Publishers.
- Smylie, M. A. and Hart, A. W. (1999). School leadership for teacher learning and change: A human and social capital development perspective. In Murphy, J. and Louis, K. S. (eds.) *Handbook of Research on Educational Administration*, 2nd edn., pp 421–443. San Francisco, CA: Jossey-Bass.
- Spillane, J. P., Reiser, B. J., and Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research* **72**, 387–431.
- Sterbinsky, A., Ross, S. M., and Redfield, D. (2006). Effects of comprehensive school reform on student achievement and school change: A longitudinal multi-site study. *School Effectiveness and School Improvement* **17**, 367–397.
- Stoll, L., Bolam, R., McMahon, A., Wallace, M., and Thomas, S. (2006). Professional learning communities: A review of the literature. *Journal of Educational Change* **7**, 221–258.
- Toole, J. and Louis, K. S. (2002). The role of professional learning communities in international education. In Leithwood, K. and Hallinger, P. (eds.) *Second International Handbook of Educational Leadership and Administration*, Part 2, pp 245–281. Dordrecht: Kluwer Academic Publishers.
- Tsang, E. W. K. (1997). Organizational learning and the learning organization: A dichotomy between descriptive and prescriptive research. *Human Relations* **50**, 73–89.
- van Veen, K., Slegers, P., and van de Ven, P.-H. (2005). One teacher's identity, emotions, and commitment to change: A case study into the cognitive-affective processes of a secondary school teacher in the context of reforms. *Teaching and Teacher Education* **21**, 917–934.

Teacher Induction

D Beijaard, Eindhoven University of Technology, Eindhoven, The Netherlands

J Buitink, University of Groningen, Groningen, The Netherlands

C Kessels, Leiden University, Leiden, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Beginning teacher – Owing to local policies and cultures, the definition of a beginning teacher can vary, but is usually referred to as a new teacher starting to work as a teacher after having finished a teacher preparation program.

Induction – The process of initiating beginning teachers into their new role as both teacher and a member of the school organization.

Induction program – A more-or-less planned and formalized system to assist and support beginning teachers.

Induction standards – Standards required by beginning teachers for achieving the status as a qualified teacher during a period of induction (in some countries, new teachers can only become qualified teachers when they meet these standards).

Mentor – A more experienced teacher who gives support, assistance, guidance, and challenge to a teacher with little or no experience.

Teacher attrition – A term referring to teachers who voluntarily and prematurely leave the profession.

Teacher competency – Meeting standardized requirements in terms of knowledge, skills, and attitudes necessary for teachers to properly perform their work.

Teacher retention – A term referring to keeping teachers in the profession.

Introduction

The teacher induction period is generally referred to as the beginning teacher's first year or years of teaching after his or her qualification as a teacher. Induction is the process of initiating beginning teachers into their new role as both a teacher and a member of the school organization. Due to differences in local policies and cultures, the definition of a beginning teacher can vary. Nevertheless, it is increasingly acknowledged in many countries that even a very comprehensive teacher-training program cannot fully prepare the teacher for his or her first job; that is, when new teachers first start to work, they still have much to learn (Britton *et al.*, 2003). To support this additional learning, many schools have recently initiated

teacher induction programs and, in some countries, these programs have actually been made obligatory and now constitute part of the teacher qualification trajectory.

A teacher induction program can be defined as a more-or-less planned and formalized system to assist and support beginning teachers in becoming competent and effective professionals. Good teacher induction programs thus enable new teachers to not just survive but also prosper during their first years of teaching and thereby provide an impetus for continuous development (Cole, 1994). The first years of teaching represent a period (1) of rapid discovery with a steep learning curve and often intense emotions (Huberman, 1989) and (2) in which new teachers typically decide to stick with the profession or leave it.

In this article, an overview of the different research perspectives on teacher induction is first presented. As seen, changing times mean changing perspectives. Special attention is then paid to the current goals of teacher induction, the elements of teacher induction programs, and those program elements that appear to be most effective. In closing, some issues that challenge us to treat teacher induction as a phase in the more long-term professional development of teachers are addressed.

A Brief History of Teacher Induction

In the 1980s, researchers began to publish about the problems of beginning teachers, and teacher induction was introduced as a distinct stage in the training of teachers (e.g., Evans and Tribble, 1986; Veenman, 1984). In numerous studies, the sudden and sometimes traumatic transition from student to starting teacher was described. Based on this growing body of knowledge, teacher shortages, and the alarming numbers of beginning teachers leaving the profession, the notion of teacher induction emerged. It is now known, for example, that not only teaching-related problems and salary considerations but also an experienced lack of support, lack of influence within the school system, feelings of isolation, and overly heavy workloads may prompt beginning teachers to leave the field.

From a research perspective, three shifts of attention can be distinguished in the evolution of teacher induction.

1. From the mid-1980s until the mid-1990s, the focus of teacher induction efforts was on the provision of that instruction-related support considered necessary to

successfully function within the classroom (Gold, 1996). The mismatch between the ideals or expectations of beginning teachers and classroom reality was often found to lead to feelings of uncertainty and/or inadequacy, thus having a considerable impact on the self-efficacy and professionalism of beginning teachers (Flores, 2004). The support provided for beginning teachers during this period in the development of teacher induction was largely based on a deficit model of teaching in which the focus was on the organization and management of instruction and development of instructional routines.

2. Starting in the mid-1990s, teacher induction was increasingly shaped by the adoption of a more integrated approach to the provision of support, the professional development of teachers, and teacher assessment. High standards for teaching and learning that were built upon school/university partnerships and a strong mentoring component that involves not only the provision of support but also the offering of challenges became part of teacher induction programs (Feiman-Nemser, 2001). Given the attention paid to the learning of beginning teachers during their first years of teaching, induction was quickly assigned a position along the teaching continuum and came to serve as a bridge between the initial training of teachers and their later professional development.
3. It is currently agreed that a teacher induction program is most meaningful when embedded in the school context and is part of a more general policy regarding the professional development of teachers. From such a perspective, new teachers are also expected to actively contribute to the development of schools, assumed to have much to offer, and thus taken to be in a position to also contribute to both the teaching profession and the field of education in general (Tickle, 2000). Stated differently, the induction of new teachers is increasingly becoming a part of the school as a learning community. It can be argued that in a well-developed learning community, the content and nature of many existing induction programs need to be reconsidered.

The teacher induction period was initially considered the new teacher's first year of teaching following qualification. In line with the aforementioned changes of perspective on teacher induction, the idea of a clearly delimited, 1-year period of induction was abandoned, and the professional development of teachers throughout their entire careers was taken to be critical.

In sum, the research perspective on teacher induction shifted across a period of two decades from a view in which beginning teachers are assumed to be the passive consumers of the knowledge and experiences of others (a deficit model of teaching) to a view in which beginning teachers are assumed to actively contribute to both, their

own professional development and the professional development of others (a growth model of teaching). This shift was undoubtedly influenced by a more general shift in our thinking about the role of teachers and teaching in contemporary society as new developments in society are known to affect the missions and goals of schools and thus the work of teachers. Just as for professionals in other fields, lifelong learning has become increasingly important for teachers and therefore a more explicit part of school policy. Nonetheless, the initial perspective on teacher induction still predominates in many schools today where new teachers are placed in the same teaching situations as experienced teachers and thus treated as experienced teachers but with a number of deficits.

The Goals of Teacher Induction

In teacher training today, considerably more opportunities for student teachers to practice and work as teachers in schools are provided than in the past. Such practice opportunities are assumed to reduce the so-called practice shock experienced by most beginning teachers (Gold, 1996). Following graduation, however, many new teachers still find themselves in a situation in which: (1) knowledge and skills pertaining to subjects that are different than the ones that they grew comfortable with during their teacher-training period are being tested, (2) the new teacher is given sole teaching responsibility, and (3) the new teacher must juggle often contradictory views of teaching practice (views imposed by the specific school culture, views created by his or her own personal expectations, and views arising from general teaching norms) (Beijaard and Papanaoum, 2002; Kelchtermans and Ballet, 2002; Tickle, 2000). For many beginning teachers, this stressful situation can result in a decision to leave the profession altogether or move to another – presumably better – school. Against this background, most school-based induction programs today can be seen to have not only a professional development goal but also a clear policy goal: investment in the retention of teachers who – with some assistance and within a relatively brief time frame – can be made into more effective teachers as opposed to investment in the relatively expensive enterprise of teacher replacement.

The professional development goal underlying teacher induction may be concentrated on the following aspects of teacher induction: improved quality of teaching (e.g., instruction and relations with students), personal-psychological support (e.g., confidence building and handling of stress), and development as a teacher (e.g., enculturation into the profession, meeting of personal needs, and positioning oneself within the school as a teacher and a colleague). Such an encompassing

professional development goal goes far beyond the maintenance of order, which most people find to be the primary concern of beginning teachers (Feiman-Nemser, 2003). The induction mandates in many countries also “do not rest on an understanding of teacher learning, a vision of good teaching or a broad view of the role formal induction can play in new teacher development” (Feiman-Nemser, 2001: p. 1031). In reality, new teachers are often placed in a sink-or-swim position, alone in their classrooms with little or no support from colleagues, little or no decision-making power, and very few opportunities for additional professional development. In other words, most teacher induction programs today still strongly resemble the first teacher induction perspective (see previous section). Given the conditions and physical circumstances of schools today, new teachers are often left to simply hope that they will survive the first years of teaching (Tickle, 2000).

Many of the people responsible for today’s teacher induction mandates consider a policy goal more important than a professional development goal. In many countries, about 30% of beginning high school teachers and about 25% of beginning elementary school teachers leave the profession within a few years of their start; such teacher attrition is considered wasteful. Despite an advanced education and often extended student teaching experience, the initial years following qualification as a teacher can strongly influence the remainder of the individual’s teaching career (i.e., the decision to remain in the field or not). Thus, it is assumed that teacher induction programs should clearly address this issue of teacher retention.

The Elements of Teacher Induction Programs

Teacher induction programs usually vary with regard to the degree of formality, structure, and intensity; that is, the comprehensiveness (how many activities and people involved), the coherence (extent to which program activities are logically connected), and the longevity (duration) of the programs can widely vary (Breaux and Wong, 2003).

Britton *et al.* (2003) further found comprehensive induction programs, in particular, to typically involve one or more of the following elements.

1. Close contact with a more experienced teacher (a mentor). Generally speaking, one can learn from the experiences of others. Assigning a beginning teacher to a mentor can produce more effective teaching during the early years because the new teacher can learn from guided practice and not just trial and error. However, Feiman-Nemser (2003) suggests that the mentoring of new teachers must be trained because

good classroom teachers may find it difficult to formulate the thoughts underlying their teaching, explain the principles underlying their work, and so on. Teacher mentors must respond to the here-and-now concerns of beginning teachers and also create learning opportunities and establish more long-term goals with these teachers, which mean that mentors receive maximum support for their induction activities within the school.

2. Relations with other beginning teachers (peers). Contact with peers allows beginning teachers to share their here-and-now experiences. Regular peer-support sessions further allow teachers to interact, cooperate, and search for solutions to problems within a safe environment (Portner, 2001). Such support sessions have been found to be particularly helpful for beginning teachers. Positive peer relations can also let new teachers feel that they are part of a community of practice and not operating in isolation.
3. Reflection, inquiry, and examination of oneself and others. It is widely acknowledged that a reflective stance, both personally and professionally, is important for the development of a teaching identity. Along these lines, growing attention is being paid to new teachers enquiring their own teaching practices as a means to professionally develop themselves (Gold, 1996; Tickle, 2000). Inquiry-based approaches are also very promising as they require the active construction of knowledge within a learning community and the connection of experienced-based knowledge to knowledge from other sources (e.g., literature, experts).
4. Observation of and by others. Observation of peers and other colleagues can produce new insights. New teachers can learn from the observation of good teaching practices modeled by more experienced colleagues. Conversely, being observed – usually by the mentor – and receiving formative feedback can highlight strengths and weaknesses or areas in need of development. Adequate feedback can strongly influence the professional development of teachers and beginning teachers in particular.
5. Timing and sequencing of opportunities. It is important that the learning opportunities of new teachers be systematically organized in a professional development curriculum. This pertains to topics to be addressed, relevant literature, and learning activities. In such a curriculum, the roles of the mentors, peers, and other colleagues should also be made abundantly clear.

Other elements that are found to characterize teacher induction programs are: an orientation to and meetings devoted to aspects of school functioning, time to release, reduced teaching loads, no assignment to difficult classes, opportunities to interact with colleagues, and a school leader who facilitates and encourages new teacher learning. Beyond the elements mentioned here and

above, which are usually of a more general character, it has recently been argued to pay greater attention to the specific subject-matter needs of new teachers and the achievement of a better balance between general and specific teacher needs.

In several countries, specific criteria and standards have been recently formulated for competent teaching. From a professional growth perspective, these criteria and standards can further stimulate a meaningful dialog between peers and mentors with regard to the essence of good teaching, the development of personal learning plans, self-assessment, and feedback (Darling-Hammond and Snyder, 2000). For example, regular assessment using such criteria and standards provides new teachers and their mentors with feedback on how things are going, what areas need work, and how much progress has been made (Feiman-Nemser, 2003). In some countries, such as the United Kingdom, specific criteria and standards are also being used for additional qualification purposes; they indicate that a new teacher has met the relevant induction standards at the end of what is usually the first year of teaching.

Critical Elements for Effective Teacher Induction

Numerous studies have documented the value of teacher induction programs. It is generally agreed that such programs improve the capabilities of beginning teachers and reduce attrition rates among new teachers (Fideler and Haselkorn, 1999). Beginning teachers seem to prefer informal methods of professional development in combination with teacher networking. The dominant theme with regard to the induction of new teachers, however, is clearly the mentor. Smith and Ingersoll (2004) found packages containing a combination of mentoring and group induction activities to show the most positive effects in terms of new teachers not going to another school or leaving the profession. In addition to this, having a mentor from the same subject field appears to be critical for effective teacher induction, which means that the development of a framework for the role of the mentor in teacher induction programs may be an important direction for further research.

However, there is no agreement on exactly which elements of induction programs are responsible for their effectiveness. Which elements really make a difference and which elements appear to hinder the implementation of these programs are still not clear. This lack of clarity may be just as much due to some limitations on the existing empirical research on effects of induction programs (Ingersoll and Smith, 2004). Among the limitations on the research to date are the following.

1. Absence of control groups. Not working with control groups in research on the effects of teacher induction programs makes it difficult to attribute particular outcomes to the induction program itself. For example, to be sure that the detected effects can be attributed to being observed and the receipt of feedback, it is critical that such beginning teachers are compared to beginning teachers working in a similar context but without observation and feedback.
2. No control for program-external factors. Factors other than those inherent to the induction programs might explain certain effects or their magnitude and should therefore be considered. For example, the teaching culture that already exists within a school can mediate the effects of an induction program. More specifically, teaching culture characterized by structural cooperation may contribute more positive experiences than a teaching culture characterized by individualism (Williams *et al.*, 2001).
3. Small-scale nature of the studies. The small-scale nature of most of the research on the effects of teacher induction programs makes generalization impossible or difficult at best. Most studies document the value of a teacher induction program for the teachers in a particular school or district, while the effects found in one particular situation need not generalize necessarily to another specific situation.

Future Challenges for Teacher Induction

Many current teacher induction programs lack a clear view of what constitutes good teaching and a professional development perspective on teaching, which could provide some guiding principles for the learning of not only new teachers but also all teachers. It is generally agreed that a mentor plays a key role in the learning of a teacher even after the teacher's first year, but continued professional development calls for additional measures and learning opportunities. For example, learning from teacher networks and, more generally, from being part of a community of teachers with shared school responsibilities (Garet *et al.*, 2001). Nowadays, many schools are developing explicit policies for the professional development of their personnel. As part of such policy efforts, schools will have to challenge the prevailing survival perspective with a new perspective on the professional development of beginning teachers – a perspective that emphasizes both personal and professional growth and also an active role of these teachers themselves in their professional development.

To promote a new professional development perspective on teaching in general and the induction of new teachers in particular, it is important that schools and teacher-training institutes cooperate. First, a comprehensive induction

program places a considerable demand on the experienced teachers in a school, while the experienced teachers often have only limited resources and capabilities when it comes to this demand. Teacher educators can then step in to meet the specific needs of the teachers in a school and perhaps the entire school community. A second reason for cooperation between schools and training institutes is to ground teacher induction not only in practice but also in general principles and theories to be provided by teacher educators, including certain learning activities that are hard to organize in a school (Britton *et al.*, 2003). In addition, an ongoing collaboration between teacher-training institutes and schools can promote the continued development of teachers as professionals. An obvious prerequisite for such collaboration is, of course, that both the schools and teacher education institutes agree upon the goals of the professional development process and the means to be used to realize these goals.

The cooperation between schools and teacher-training institutes with regard to the professional development of beginning teachers, in particular, is now and in the near future seen as one of the most important challenges facing teacher education policy. The amount of time and experiences needed to become a fully competent and effective professional are very much open to question. Furthermore, the cooperation between schools and teacher-training institutes will lead to new insights into what can be learned at the workplace and what cannot. At this moment, very little is known about which characteristics of the teaching context promote the professional development of teachers. Beginning teachers are spending increasingly more time in schools, meanwhile it is good to know that working and learning may involve opposing if not conflicting processes!

Conclusion

It is obviously in the interest of students that their teachers ought to be competent professionals. Research shows that teacher competency can influence student learning gains and study results. High teacher turnover constitutes a threat to student learning, because high turnover rates can create an unstable and less effective learning environment. High teacher turnover rates also impose heavy demands on the remaining teachers and staff within a school and can require considerable time and money for the recruitment, hiring, and training of replacement personnel (Halford, 1999). Taken together, all of these suggest that the adoption of high-quality induction programs as part of a more general growth perspective on the professional development of teachers may help turn the current school situation around and that everyone stands to benefit from such an investment in the end.

On the basis of existing empirical research, it is difficult to draw conclusions about the impact of participation

in a teacher induction program. Do beginning teachers really change or develop their teaching behavior as a result of participation in such a program? Do the changes prompted by involvement in a teacher induction program also influence student learning? The search for the answers to these questions calls for carefully controlled experiments and cohort studies of, for example, teachers during their induction period and thereafter. Studies such as these are needed in addition to those studies that have already documented the value of specific teacher induction programs.

Finally and regardless of which induction program, the teachers in a school and school leaders have been found to shape a beginning teacher's decision to stay or not stay in the field of education the most. The key to retaining teachers is what Breaux and Wong (2003) call belonging, which is a basic human need. Most new teachers want more than just a job. As part of a team and part of a culture, beginning teachers also want to contribute and make a difference. Recognition of the potential of new teachers to contribute to the education of students and utilization of this potential are, however, necessary. Under such circumstances, which also involve mutual trust and respect, more academically talented and very capable teachers would decide to stay in the field.

Bibliography

- Beijaard, D. and Papanaoum, Z. (2002). Teacher induction at the crossroads: An attempt to harmonize the chaos. *Teaching and Teacher Education* **18**, 905–911.
- Breaux, A. L. and Wong, H. K. (2003). *New Teacher Induction: How to Train, Support, and Retain New Teachers*. Mountain View, CA: Harry K. Wong Publication, Inc.
- Britton, E., Paine, L., Pimm, D., and Raizen, S. (eds.) (2003). *Comprehensive Teacher Induction*. Dordrecht: Kluwer Academic Publishers.
- Cole, A. L. (1994). Problems and paradoxes in beginning teacher support: Issues concerning school administrators. *The Alberta Journal of Educational Research* **XL**(3), 297–318.
- Darling-Hammond, L. and Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education* **16**, 523–545.
- Evans, E. D. and Tribble, M. (1986). Perceived teaching problems, self-efficacy, and commitment to teaching among preservice teachers. *Journal of Educational Research* **80**(2), 81–85.
- Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record* **103**(6), 1013–1055.
- Feiman-Nemser, S. (2003). What new teachers need to learn. *Educational Leadership* **60**(8), 25–29.
- Fideler, E. F. and Haselkorn, D. (1999). *Learning the Ropes: Urban Teacher Induction Programs and Practices in the United States*. Belmont, CA: Recruiting New Teachers, Inc.
- Flores, M. A. (2004). *The Early Years of Teaching: Issues of Learning, Development, and Change*. Res-Editore: Porto.
- Garet, M., Porter, A., Desmoine, L., Birman, B., and Yoon, K. S. (2001). What makes professional development effective. *American Educational Research Journal* **38**(4), 915–946.
- Gold, Y. (1996). Beginning teacher support: Attrition, mentoring, and induction. In Sikula, J. (ed.) *Handbook of Research on Teacher Education*, pp 548–591. New York: Macmillan.

- Halford, J. M. (1999). Policies to support new teachers. *Educational Leadership* **56**(8), 85.
- Huberman, M. (1989). The professional life cycle of teachers. *Teachers College Record* **91**(1), 31–57.
- Ingersoll, R. M. and Smith, T. M. (2004). Do teacher induction and mentoring matter. *NASSP Bulletin* **88**, 28–40.
- Kelchtermans, G. and Ballet, K. (2002). The micropolitics of teacher induction: A narrative-biographical study on teacher socialization. *Teaching and Teacher Education* **18**, 105–120.
- Portner, H. (2001). *Mentoring is Not Enough: Everything Else Schools and Districts Need to do*. Thousand Oaks, CA: Corwin Press, Inc.
- Smith, T. M. and Ingersoll, R. M. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal* **41**(3), 681–714.
- Tickle, L. (2000). *Teacher Induction: The Way Ahead*. Buckingham, PA: Open University Press.
- Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research* **54**(2), 143–178.
- Williams, A., Prestage, S., and Bedward, J. (2001). Individualism to collaboration: The significance of teacher culture to the induction of newly qualified teachers. *Journal of Teacher Education* **27**(3), 253–267.

Further Reading

- Feiman-Nemser, S. (2001). Helping novices learn to teach: Lessons from an exemplary support teacher. *Journal of Teacher Education* **52**(1), 17–30.
- Ingersoll, R. M. and Smith, T. M. (2003). The wrong solution to the teacher shortage. *Educational Leadership* **60**(8), 30–33.
- Wang, J. and Odell, S. J. (2002). Learning to teach according to standard-based reform: A critical review. *Review of Educational Research* **72**(3), 481–546.
- Wong, H. K., Britton, T., and Ganser, T. (2005). What the world can teach us about new teacher induction. *Phi Delta Kappan* **86**(5), 379–384.
- Wong, H. K. and Wong, R. T. (2004). *The First Days of School: How to Be an Effective Teacher*. Sunnyvale, CA: Harry Wong Publishing.
- Zeichner, K. M. (2003). The adequacies and inadequacies of three current strategies to recruit, prepare, and retain the best teachers for all students. *Teachers College Record* **105**(3), 490–519.

Teacher Learning as Workplace Learning

J Imants, Radboud University, Nijmegen, The Netherlands

K van Veen, Leiden University, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Introduction

A central notion in theories of workplace learning is the intrinsic and mutual relationship between working and learning. In teacher learning as workplace learning, two themes come together: the content and organization of teachers' work, and the content and process of teachers' learning in the context of their classrooms and schools.

During the last two decades, workplace learning has been an important item in discussions on vocational education and professional development in industrialized societies. These discussions cover the broad field of economic and cultural developments, technological and organizational innovations, knowledge economy and knowledge creation, and professional development and vocational learning in schools and work organizations. In discussions on teacher learning, workplace learning is a significant item for at least two reasons. Workplace learning is assumed to be a promising method for the development of productive learning contexts in teacher education and continuing professional development. Professional learning of aspiring and practicing teachers is expected to result in competences for the teaching of their students as productive workers and as productive learners in their future workplace contexts.

In this article, teachers' workplace learning is defined as changes in teaching practices in classrooms and schools that are mediated through individual teacher learning and problem-solving processes in the school (Ellström, 2001). In the field of education, professional development is the generic term under which teacher learning activities and programs are organized and discussed. Teachers' professional development can be defined as those processes and activities designed to enhance the professional knowledge, skills, and attitudes of educators so that they might, in turn, improve the learning of students (Guskey, 2000). Recently, these activities and programs often build on workplace learning insights. This article aims to explore the relatively new field of teachers' workplace learning in classrooms and schools, as it builds on, and adds new insights to teachers' professional development in schools. The question is how and in what directions teacher learning can be promoted in the context of work in classrooms and schools. First, the state of the art in teachers' professional development is briefly characterized. Next, recent general insights in workplace learning in combination with insights in teachers' workplace learning is discussed.

Teachers' Professional Development

Teachers' professional development literature shows a range of insights regarding views of learning (acquisition vs. constructivism), designs (fragmented vs. ongoing and systematic), and opportunities (formal and informal, mandatory and voluntary, serendipitous and planned). The lifelong curriculum of teachers' professional development seems fragmented and incoherent.

The rationales for professional development also diverge considerably. A first rationale argues that most educational reforms require a great deal of learning on the part of teachers. Teachers need support and advice to understand the reform and work according to the reform. A second rationale starts from the nature of teachers' work. Teachers play a crucial role in students' learning, and professional development of teachers is assumed to be in the interest of the students and teaching as a profession. The first rationale assumes that teachers show personal inadequacies in need of repair, while the second rationale assumes that teacher learning contributes to fulfillment as a practitioner in support of students' learning.

Attention is increasing for teacher learning and teacher professional development programs situated in the workplace in contrast to off-site programs. These off-site activities entail workshops, conference sessions, seminars, lectures, and other short-term training events on subject-matter issues and topics such as cooperative learning and classroom management. Design elements and conditions that can enhance the effectiveness of these activities are longer duration, connection with teachers' work context, focus on subject matter, and emphasis on analysis and reflection (Smith and Gillespie, 2007). By identifying these design elements and conditions, it is a small step toward recent on-site professional development methods that are embedded in daily work in classrooms and schools. These work-embedded methods include training within the school or local context, and creating ongoing professional communities, study circles, and inquiry groups. In addition to this, workplace learning has brought about a shift:

- from a focus on individual teacher knowledge, skills, and teaching competencies, including new instructional methods, to a focus on student learning and specific teacher problems;
- from off-site to on-site activities;

- from single sessions or a short series to long-term and ongoing sessions; and
- from a focus on change as something that is done to teachers, and programs that change teachers as passive participants, to change as a complex process that involves learning with teachers as active learners.

As far as research has been conducted regarding the effectiveness of learning activities that are embedded in daily work, it shows that teacher learning is promoted by a focus on subject-matter knowledge for teaching, understanding of students' thinking, and instructional practices. Further, the process of meaningful learning is slow and uncertain, and some elements of teacher knowledge are easier to change than others. It takes time to develop a community. Teachers have little experience engaging in a professional discourse that is public and critical of their work and the work of colleagues. Finally, it should be noted that combining features of off-site and on-site activities in professional development programs seems to be very effective (Little, 2006; Smith and Gillespie, 2007).

In general, several methods for workplace learning have been developed (coaching, supervision, and collegial consultation; mentoring and master–mate learning; development tasks and action research; and instruction and application in the workplace). These methods have been applied to (student) teachers' professional development in schools. Besides, new methods for the evaluation of (student) teachers' competences and learning are developed and applied (assessment instruments and portfolio). A recurring problem in teacher professional development is the gap between practice (school) and theory (university), and a central aim of these methods for workplace learning is to contribute toward bridging this gap. However, according to Eraut (2000), the problem of the theory–practice gap concerns the relationship between the codified professional knowledge (espoused theory) that is developed and taught at universities, and tacit teachers' knowledge (theory-in-use) that is developed and transmitted in classrooms and schools. Bridging the theory–practice gap cannot be reduced to mutual adaptation of these sources of knowledge. Sound application of one or more of the aforementioned methods is not sufficient to link espoused theory and theory-in-use, because the problematic relationship between these diverse bodies of knowledge has deep epistemological and (micro-)political roots. Besides, the effectiveness of these methods depends on the conditions of the workplaces in which they are applied, which is discussed in the next sections.

Despite an increasing body of knowledge on design elements and conditions, there is very scarce empirically tested knowledge about: (1) the effectiveness of teachers' professional development programs, and (2) more in

detail, about what teachers learn, how they learn, and how this knowledge improves their practice. The current state of practice and research of teachers' professional development is inadequate (Borko, 2004), and this constitutes a serious problem for policy, practice, and research. This does not mean that the design, learning principles, and beliefs of many programs are unreasonable or unsound (Wilson and Berne, 1999).

In summary, the literature on teacher professional development in schools shows a shift toward workplace learning, though at the same time a lack of theoretical and empirical understanding is observed. In the next sections the analysis of teachers' professional development is broadened by taking the mutual relationship between teachers' working and learning as the starting point.

Perspectives on Workplace Learning

The identification of professional development in workplace contexts depends on the perspective from which learning is analyzed. Generally, perspectives on workplace learning differ with regard to goals, outcomes, rationales, and managerial views. Workplace learning varies from adaptive to developmental, in terms of tasks, methods, and results, which are given or not given (Ellström, 2001). A broad range of possible learning outcomes for teachers in the workplace can be found, ranging from habitual reaction to reflective practice, and inquiry (Smylie, 1995). Four perspectives on teachers' workplace learning are distinguished (Nieuwenhuis and van Woerkom, 2006):

1. In teacher education settings, workplace learning primarily is regarded as preparation for working, and qualification is the goal. This perspective plays an important role in studies of on-the-job training in initial teacher education.
2. In ongoing work settings in schools, teachers' workplace learning is regarded as contributing to productivity of schools, often as nonformal learning and as an intrinsic quality of professional work. Development of routines is important in this perspective, besides development of improvements during the performance of work activities. This perspective can be helpful to understand the dynamics and outcomes of teacher induction and socialization.
3. In global market and knowledge economy settings, teachers' workplace learning is closely linked to the vitality of the school, school improvement, and to educational reform.
4. In the individual perspective, teachers' workplace learning is regarded as learning for life and individual professional development in community contexts.

Teachers' Workplace

Teacher learning is located in diverging contexts within the school. Three interrelated levels are distinguished:

1. individual or personal learning by teachers or school leaders within the context of the school;
2. social learning in small groups or teams of teachers; and
3. learning that occurs across the school organization as a whole.

According to Mitchell and Sackney (2000), it is the interrelatedness of conditions and processes at these three levels that explains successes and failures in school improvement and shared learning.

Formally, the school as a workplace consists of diverging units that are charged with diverging tasks and responsibilities. Informally, the school contains several spheres of influence and interaction. For teachers, the work with students, inside and outside the classroom, makes up the core of their work. Opportunities to enhance students' learning are the driving force for teachers to participate in professional development. For the large part of their working day, teachers work individually with (a) group(s) of students, rather isolated from their colleagues. The most meaningful work experiences that promote teachers' workplace learning is therefore gained in the work with students, in the student-related work with colleagues, and in the interactions with the workplace conditions that are closely linked to the work with students and colleagues – the community of practice in which teachers participate. Teachers' workplace learning is expected to be primarily located in the teachers' own sphere of influence and interaction, and in the units in which the work with students is organized. School conditions and processes that are located in the administrators' sphere can be assumed to affect teacher learning processes and results in an indirect way, more specifically, as mediated by processes of reinvention and reinterpretation by teachers in their own sphere. For example, an educational policy document of a school or district will not influence teachers' professional development directly. Depending on how teachers interpret the school's educational policy, this document will affect teacher learning indirectly in foreseen and unforeseen directions.

Workplace Conditions in Schools

The educational systems in industrialized countries are characterized by restructuring and standardization of work in schools, aimed at improved student learning and the promotion of a new professionalism in the teachers' profession. Simultaneously, tendencies of

bureaucratization and intensification in schools and teachers' work have been identified (Ballet *et al.*, 2006). It can be assumed that restructuring and standardization, combined with bureaucratization and intensification, will have significant and contradictory impact on the content, processes, and effects of teachers' workplace learning in schools. This actual impact can be opposed to the formal aims of reform. This complex relationship between working and learning is elaborated in this section.

In his review of general workplace learning research, Ellström (2001) distinguishes five groups of factors that can be assumed to be critical for facilitating or constraining the integration of learning and work: (1) the learning potential of the task; (2) opportunities for feedback, evaluation, and reflection on the outcomes of work actions; (3) formalization of work processes; (4) employee participation in handling problems and developing work processes; and (5) learning resources. These general factors reasonably summarize the results of overviews of workplace learning conditions in schools, with the exception of one group of factors. Schooling is a strongly moral and normative enterprise, and in overviews of workplace conditions for teacher learning, shared norms are often mentioned as a critical factor. Shared norms are added as a group of school factors to Ellström's general overview.

Learning Potential of the Task

Generally, task complexity, variety, autonomy, and control are regarded as important determinants of the learning potential of the workplace. Traditionally, teachers' work is viewed as complex, varied, and autonomous. Superficially, the school seems to be a favorable context for workplace learning. However, studies on intensification show tendencies of deprofessionalization in teachers' work, starting from a combination of narrowing and formalizing the traditional teaching tasks, and simultaneously adding new nonteaching tasks to teachers' work. These developments are accompanied by experiences of strong work pressure and negative emotions. Further, work redesign and task differentiation in schools only have weak links to aspired forms of teacher learning and student learning. To be an effective workplace learning strategy, redesign of work should focus attention on crucial problems of curriculum and instruction.

Teachers strongly depend on routines because they make an enormous amount of decisions each day they work (Eraut, 2000). Teachers' actions can be described as routinized when they no longer need to think about what they are doing because they have done it so many times before. In routinization, explicit knowledge is converted to tacit knowledge through repetition. Besides individual schemata, routines also are collective schemata for both understanding and acting: they supply teachers with shared understanding of who is doing what, and this

understanding allows teachers to correctly perform a practice. External and internal pressures can force teachers and schools to change their routines. Routines may restrict school's responses to change, but they can also help the school to survive changes in environmental demands and expectations (Hoeve *et al.*, 2006). Routines in teaching tasks have the potential to be impediments and opportunities for workplace learning.

Opportunities for Feedback, Evaluation, and Reflection on the Outcomes of Work Actions

Feedback on the results of work activities is considered to be necessary for learning to occur. Activities in which teachers reflect on student performance, work, and thinking can promote professional development. The effects of feedback can be cognitive as well as motivational. The function of feedback depends on the existence of clear goals. At the level of classrooms and schools, educational goals often are vague or inconsistent, and they can change during teaching. The problem of getting clear feedback on student performance because of unclear goals can be an obstacle for teacher learning in contexts where instructional tasks, methods, and results are given. However, inconsistent or unclear goals may promote analysis, evaluation, and reflection that are assumed to be important for innovative teacher learning.

In schools, much information is available that could serve as feedback and promote reflection on the quality of teaching and learning, for example, school self-evaluation reports, reports by the inspectorate, files of student assessment outcomes, aggregated student test results, and teachers' accounts of classroom experience (Little, 2006). Systematic use of this data on student performance, work, and thinking as sources for learning by individual teachers and teacher teams hardly occurs in schools.

Formalization of Work Processes

The daily educational and pedagogical practices in the school are important opportunities and constraints for teacher learning. This learning is focused on subject content, pedagogic and instructional methods, and classroom management (Scribner, 1999). Traditional forms of formalization (fixed regulations and procedures) of daily work processes in schools still tend to dominate instruction, classroom management, and organization in schools. For example, a persistent problem in education is the limited flexibility to deal with student diversity in schools and classrooms, notwithstanding inclusive aims. Moreover, images of how schools and classrooms should be structured and should function, as well as daily practices, build on traditional scientific management principles. A gap is existent between the innovative expectations on workplace learning in schools, and the highly

formalized instructional practices of aspiring and practicing teachers as contexts for learning.

However, starting from the observations that work in schools is routinized and that the daily knowledge base of teachers is strongly tacit, a process of formalizing instruction and guidance practices can also contribute to teacher learning. Formalization of instruction and guidance can serve as the starting point for questioning existing instructional routines and making the tacit explicit and open for discussion. In that case, the aim is teachers getting a better understanding of (1) what and how their students learn and think, and (2) how student learning is promoted by teachers' instruction and guidance. This learning by formalization should not be confused with simply replacing an old instructional method by a new one (Smith and Gillespie, 2007).

Teacher Participation in Handling Problems and Developing Work Processes

Learning is facilitated if teachers have access to and are able to successfully participate in problem handling and activities aimed at the development of improvements in instruction and guidance, both formally and informally. Teachers share positive perceptions of participation in decision making in schools with regard to instructional, pedagogical, and curricular topics, as opposed to administrative topics. Teacher participation in decision making can create frustration, anger, and distrust, when the implementation of decisions that were made is blocked or not promoted by school management. Teachers should be able to participate in all the steps of the cycles in which improvements are developed: planning, development, performance, and evaluation. When the majority of teachers is only expected to perform instructional tasks that are planned, developed, and evaluated by teacher leaders and managers, learning will decrease.

Teachers' collegial interactions vary in task interdependence and alignment (synergy in directions of aims and actions of different actors). High alignment and interdependence, for example, in the context of joint problem solving, are associated with richer and more stimulating learning environments for teachers. This collaboration can produce shared understanding and investment, thoughtful development, and a fair, rigorous test of selected ideas. Lower levels of interdependence and alignment occur when teachers exchange stories, tips, or materials. This can result in reinforcement of and emotional support for maintaining routines (Rosenholtz, 1989).

Shared Norms

Shared norms that affect teacher learning in the workplace can be distinguished in two types: norms regarding student learning and development, and norms

regarding teachers' work and relationships with colleagues. Collective focus on and shared responsibility for student learning are identified as important characteristics of teachers' professional communities in schools. The absence of a shared conceptual and technical language or a common technical culture in schools has been assumed to be a serious hindrance for the development of these shared norms. However, precise analysis of teacher professional conversations shows that too deep a skepticism on the opportunities in schools for developing shared norms oversimplifies and underestimates the potential of teachers' collegial practices for teacher learning (Little, 2006).

Isolation from colleagues has a negative effect on teacher learning by creating invisible walls between teachers and diminishing the valuable role of collaborative activities (Scribner, 1999). This individualism, in which personal experience plays a central role, is regarded as a defining feature of the teaching profession. Strong traditional cultures are based on teachers collectively defending this individualism and existing conceptions of appropriate curriculum and instruction, even in the face of student failure. Besides, strong innovative and professional cultures can contribute to norms of collegiality and experimentation.

Learning Resources

Three resources for workplace learning are teachers' prior knowledge, curricular and instructional materials, and time. Teacher learning by experience presupposes conceptual tools (e.g., pedagogical content knowledge) and explicit knowledge about student learning and how student learning is promoted by instruction and pedagogic methods. These conceptual tools and knowledge are used to identify and interpret relevant experiences while teaching, and they play an essential role in workplace learning.

Varied sources of instructional and curricular materials promote reinvention of instructional practices and lower barriers for experimentation. These practical materials provide opportunities for teachers to experience what works in daily classroom practice. However, when implemented superficially, the availability of the same sources bears the risk of counterproductive learning effects, because the anticipated student learning results are not realized in the case of superficial implementation.

Time is an important resource for participation in formal learning activities, as well as for reflection as part of the ongoing work process, especially in combination with related factors, for example, school schedules. Professional development activities should be conducted often and long enough to ensure gains in knowledge, skills, and confidence. Formal availability of time for professional development is necessary but not sufficient

as a resource for teachers to learn. Meaningful and time-consuming collaboration is often squeezed to margins of teachers' worklives (on Sunday afternoons or weekly evenings in teachers' houses). Daily work in schools should be scheduled and organized in such a way that teachers have a fair chance to take advantage of the time available for learning and development. Without these organizational protections, time for learning will disappear under the pressure of chains of incidents and daily routines. A complex trade-off occurs between time for production and time for learning in schools (Ellström, 2001).

Conclusions

Before presenting the conclusions, it should be noted that many of the insights presented in this article lack a solid base of research, although some research had been done on these topics. The field lacks an explicit shared conceptual language, and empirical research often is lacking the rigor that is needed for generalizations. Nevertheless, some tentative conclusions can be drawn from the previous analysis.

Summarizing the evidence, two perspectives on teachers' workplace learning and work-embedded professional development can be distinguished. These perspectives vary in assumptions on manageability and predictability of learning processes and outcomes in workplace contexts. In one view, teacher workplace learning is a specific activity that can be managed in the direction of formalized school outcomes and desired change. In the other view, teacher learning outcomes range from improvement in school performance to counterproductive effects on teaching and student learning. In this latter view, teacher learning is emergent, and steering opportunities for management to intervene effectively in this process are weak. An unanswered question is whether these perspectives are complementary or contradictory toward each other.

Several potentially effective methods for workplace learning have been identified. The effectiveness of these methods should be regarded within the specific workplace contexts in which they are applied, and keeping an eye open for the conflicting epistemological assumptions of the diverging bodies of knowledge (espoused theories and theories-in-use) that they try to unite.

Six factors have been identified for the integration of teachers' working and learning in schools. Usually, these factors are regarded as objective characteristics of the schools in which they emerge. However, it can be assumed that the impact of these factors mainly is located in how they are interpreted and reinvented by teachers, and how teachers make sense of these factors in their daily work. The relationships between these factors and teacher learning are complex and paradoxical. These factors

play a multidimensional role in teacher workplace learning: they potentially promote opposite directions of learning (reinforcing routines or promoting innovations) and they can have opposite side effects on teacher workplace learning (commitment or isolation, empowerment or alienation, etc.).

Traditionally, teachers have been regarded as passive participants in professional development and workplace learning programs, lacking knowledge and skills. Of late, teachers are more often regarded as active learners, and as agents in co-determining their working and learning goals, contents, processes, and outcomes. In this respect, the six factors for the integration of working and learning in schools also should be regarded as co-determined by the teachers themselves. Starting from insights in sense-making in organizations and schools, all aspects of working and learning in schools can be regarded as part of an ongoing process of reinvention and reinterpretation by teachers and other agents in and around the school.

Generally, high-commitment, flexible, and learning-intensive work systems are assumed to promote the integration of learning and working, but scientific management principles still tend to dominate the design of many workplaces (Ellström, 2001). The same phenomenon seems to hold for schools. Rosenholtz (1989) distinguished learning-enriched and learning-impovertished schools. Rosenholtz and other researchers found far more learning-impovertished schools than learning-enriched schools (Little, 2006). Workplace learning in schools has potential advantages, like facilitated transfer of learning, accommodation of different needs and learning styles, ability to tap collective knowledge, and the natural promotion of collaboration and school renewal. But there are also potential problems in workplace learning. Not all learning at work is good, as is shown by negative lessons learned by (student) teachers, reinforcement of existing biases, and refinement of poor practices (Bredeson, 2003). Moving all learning to the workplace might create too heavy a burden on the workplace. Effective professional development in schools implies the search for a productive balance and creative swing between adaptive and innovative varieties of individual and organizational learning, and on-site and off-site learning.

Bibliography

- Ballet, K., Kelchtermans, G., and Loughran, J. (2006). Beyond intensification towards a scholarship of practice: Analysing changes in teachers' work lives. *Teachers and Teaching: Theory and Practice* **12**, 209–229.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher* **33**(8), 3–15.
- Bredeson, P. (2003). *Designs for Learning: A New Architecture for Professional Development in Schools*. Thousand Oaks, CA: Corwin Press.

- Ellström, P. E. (2001). Integrating learning and work: Problems and prospects. *Human Resource Development Quarterly* **12**, 421–435.
- Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology* **70**, 113–136.
- Guskey, T. R. (2000). *Evaluating Professional Development*. Thousand Oaks, CA: Corwin Press.
- Hoeve, A., Jorna, R., and Nieuwenhuis, A. (2006). 'Het leren van routines: van stagnatie naar innovatie' (Learning from routines: From stagnation to innovation). *Pedagogische Studiën* **83**, 397–409.
- Little, J. W. (2006). *Professional Development and Professional Community in the Learning-Centered School*. Arlington, VA: National Education Association.
- Mitchell, C. and Sackney, L. (2000). *Profound Improvement: Building Capacity for a Learning Community*. Lisse: Swets and Zeitlinger.
- Nieuwenhuis, A. and van Woerkom, M. (2006). 'Rationales achter werkplekleren' (Rationalities for work-based learning). *Pedagogische Studiën* **83**, 343–353.
- Rosenholtz, S. (1989). *Teachers' Workplace: The Social Organization of Schools*. New York: Longman.
- Scribner, J. (1999). Professional development: Untangling the influence of work context on teacher learning. *Educational Administration Quarterly* **35**, 238–266.
- Smith, C. and Gillespie, M. (2007). Research on professional development and teacher change: Implications for adult basic education. *Review of Adult Learning and Literacy* **7**, 205–244.
- Smylie, M. (1995). Teacher learning in the workplace: Implications for school reform. In Guskey, T. and Huberman, M. (eds.) *Professional Development in Education: New Paradigms and Practices*, pp 92–113. New York: Teachers College Press.
- Wilson, S. M. and Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. In Iran-Nejad, A. and Pearson, P. D. (eds.) *Review of Research in Education*, **24**, pp 173–209.

Further Reading

- Coburn, C. E. (2005). Shaping teacher sensemaking: School leaders and the enactment of reading policy. *Educational Policy* **19**(3), 476–509.
- Conley, S. and Enomoto, E. (2005). Routines in school organization: Creating stability and change. *Journal of Educational Administration* **43**, 9–21.
- Day, C., Sammons, P., Stobart, G., Kington, A., and Gu, Q. (2007). *Teachers Matter: Connecting Work, Lives and Effectiveness*. Maidenhead: Open University Press, McGraw-Hill.
- Hoekstra, A. (2007). *Experienced Teachers' Informal Learning in the Workplace. Doctoral Dissertation*. Utrecht: IVLOS Institute of Education, Utrecht University.
- Huysman, M. (2000). An organizational learning approach to the learning organization. *Journal of Work and Organizational Psychology* **9**, 133–145.
- Imants, J. (2003). Two basic mechanisms for organisational learning. *European Journal of Teacher Education* **26**, 293–311.
- Little, J. W. and McLaughlin, M. W. (eds.) (1993). *Teachers' Work: Individuals, Colleagues, and Contexts*. New York: Teachers College Press.
- Louis, K., Marks, H., and Kruse, S. (1996). Teachers' professional community in restructuring schools. *American Educational Research Journal* **33**, 757–798.
- Meirink, J. (2007). *Individual Teacher Learning in a Context of Collaboration in Teams*. Doctoral Dissertation, Leiden University: Leiden.
- Reis, S. (2005). *Learning to Teach Reading in English as a Foreign Language*. Doctoral Dissertation, Radboud University, Nijmegen.
- Spillane, J., Reiser, B., and Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research* **72**(3), 387–431.
- Weick, K. (1995). *Sensemaking in Organizations*. Thousand Oaks, CA: Sage.
- Zwart, R. (2007). *Teacher Learning in a Context of Reciprocal Peer Coaching*. Doctoral Dissertation, Radboud University, Nijmegen.

Teacher Learning with Lesson Study

A Murata, Stanford University School of Education, Stanford, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Lesson study is a collaboration-based teacher professional development approach that originated in Japan (Fernandez and Yoshida, 2004; Lewis and Tsuchida, 1998; Stigler and Hiebert, 1999). When it was first introduced in the United States in the late 1990s, it quickly gathered the attention of the US educators as it was thought to be an adoptable and effective innovation (Choksi and Fernandez, 2004; Lesson Study Research Group, 2007; Lewis *et al.*, 2004; National Research Council, 2002; North Regional Educational Laboratory, 2002; Richardson, 2004; Stepanek, 2001, 2003; Wilms, 2003). Lesson study incorporates effective characteristics of other professional development programs identified in prior research. It is site based, practice oriented, focused on student learning, collaboration based, and research oriented (Bell and Gilbert, 2004; Borko, 2004; Cochran-Smith and Lytle, 1999, 2001; Darling-Hammond, 1994; Wang and O'Dell, 2002; Little, 2001; Hawley and Valli, 1999; Wilson and Berne, 1999). Lesson study was also one of the foci for the International Congress on Mathematics Education (ICME) Ninth Conference in 2002 and has spread in many other countries (Fujita *et al.*, 2004; Lo, 2003; National College for School Leadership, 2004; Shimizu *et al.*, 2005). Despite the rapid rate of adoption, this form of professional development is still new in countries other than Japan, and thus schools and teachers in these countries are still at the early stage of adopting the innovation to their existing systems. There is an emerging body of lesson study literature, but we do not yet have a coherent and shared understanding and model of teacher learning with lesson study. The purpose of this article is to give an overview of lesson study with its structure (and variation), its history, and emerging lesson study research literature, in order to explicate teacher learning models in lesson study as well as to identify future research agenda with lesson study.

Lesson Study: History, Structure, and Variation

Lesson study is an instructional improvement approach that places teachers in the center of professional activity, with their interests and desire to better understand student learning. The idea is simple: teachers come together with a shared question regarding their students' learning, plan a lesson to make this student learning visible in

practice, and examine and discuss what they observe. Through the process, teachers have multiple opportunities to discuss student learning and how their teaching affects it. Lesson study typically follows the steps outlined in **Figure 1**, with a research lesson (live lesson observation) as a centerpiece of the study process (Fernandez and Yoshida, 2004; Lewis, 2002; Lewis and Tsuchida, 1998; Murata and Takahashi, 2002; Wang-Iverson and Yoshida, 2005).

After identifying a lesson goal, teachers plan a lesson. They will take a certain teaching approach to make student learning visible, keeping their lesson goal in mind. The main purpose of this step is not to design a perfect lesson but to test out a teaching approach in the lesson context to investigate how students learn. As they anticipate students' possible responses and craft the details of the lesson in a certain way, teachers learn the key aspects of the lesson, how students may possibly respond to these aspects, and the different thinking and reasoning behind the possible responses. Teachers will then focus on the particular student thinking in the lesson, take notes on different student approaches, and, after the lesson, discuss the data they have collected in a debriefing. While there are other professional development programs that incorporate many of the characteristics of lesson study (e.g., action research and teacher research), what sets it apart from others is the live research lesson. As discussed in the section below, research lessons create unique learning opportunities for teachers compared to other forms of professional development.

In Japan, lesson study has been widely used in schools for over a century, and many Japanese educators attribute success in changing their teaching practice to lesson study (Lewis *et al.*, 2006; Shimizu *et al.*, 2005). Any new educational approaches introduced in Japan are examined by teachers using lesson study as a foundational mechanism to support the improvement of teaching and better understand in practice. In many cases, teachers assume the main role in making the new approaches adoptable, and thus the approaches become more practical and understandable. Thus, lesson study works effectively to connect theory and practice in Japan. After the Third International Mathematics and Science Study, Stigler and Hiebert (1999) introduced lesson study to the US audience. Since then, lesson study has spread throughout the United States, and now over 400 schools are currently involved in lesson study (Lesson Study Research Group, 2007). It has also attracted the attention of an international audience, and there have been more than a dozen international

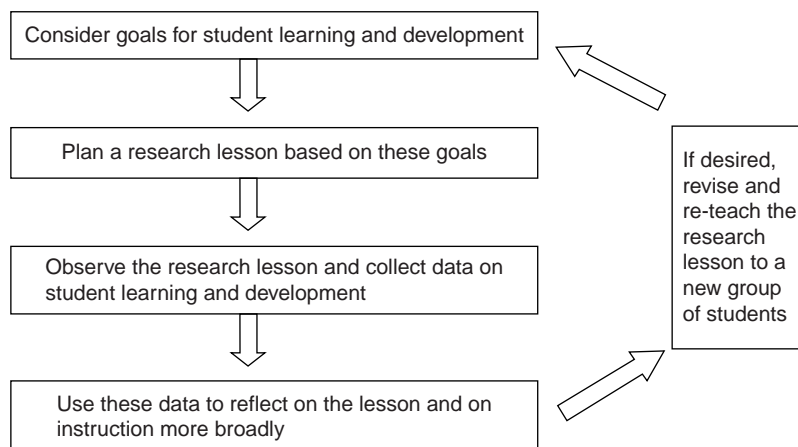


Figure 1 Lesson study cycle.

conferences and workshops around the world for people to share their experiences and progress with lesson study (Fujita *et al.*, 2004; Conference on Learning Study, 2006).

While lesson study is known in the United States (and other parts of the world) for a small, school-based collaboration typically in the subject area of mathematics, lesson study comes in many different shapes and sizes in Japan. There is a small, school-based lesson study as well as a larger-scale national-level lesson study (Murata and Takahashi, 2002; Lewis and Tsuchida, 1998; Shimizu *et al.*, 2005). For a large-scale and national-level lesson study in Japan, teachers often travel long distances to participate, and hundreds of people gather for one event. For a middle-scale and district-level lesson study, teachers may come together for a district's professional development day for which they have a list of choices of lessons with different grade levels, subject areas (Lesson study research conducted outside of Japan has primarily been focused on mathematics so far, therefore the chapter reports findings from these studies on mathematics instruction, however, lesson study may be used for all subject areas.), and topics to attend. These different kinds of lesson study meet different needs and interests of the teachers. While a small-scale and in-school lesson study is effective for teachers to improve their teaching for students of a particular community, where teachers share knowledge of the students and the community, a large-scale lesson study is important when a new educational approach (e.g., problem-based math instruction and collaborative learning) is introduced, and teachers across different schools are trying to make sense of what that means in their classrooms. Lesson study provides an opportunity to present a visible example of a new idea for teachers to discuss, ask questions, and come to construct a shared understanding of the new idea. Different forms of lesson study provide different learning opportunities for teachers.

Teacher Learning Models with Lesson Study

Teacher Learning

For decades, educational researchers have sought to understand what teachers learn and how teachers' learning translates into improvement of their teaching (e.g., Cohen, 1990; Franke *et al.*, 2001; Fuller, 1969; Fuller and Bown, 1975; Sherin, 2002; Shulman, 1987, 1996; Wood *et al.*, 1991). These questions are enormously challenging because of the intricate dynamics among teaching practices, beliefs, and knowledge (Sherin, 2002; Smith, 2000; Thompson, 1984), the many types of knowledge (of content, students, and curriculum, pedagogy) entailed in teaching (Shulman, 1987; Ball and Bass, 2000), and the limited transfer from many traditional forms of teacher education (e.g., lectures on content) to one's own practice (Ball and Bass, 2000; Ball and Cohen, 1996; Remillard, 2000). Teaching is a system, and changing one part of the system may not change the whole in the long run, since other parts will work to maintain the original function of the system (Stigler and Hiebert, 1999). Research and theory suggest that changes in teaching practice may be facilitated by opportunities for teachers to analyze familiar teaching practices in depth, make connections between familiar and new practices, and understand the changes needed in each part of the system of teaching in order to create an overall change (Franke *et al.*, 2001; Sherin, 2002; Smith, 2000). Professional development contexts that enable reexamination and reorganization of existing ideas – rather than simple acquisition of new ideas – may be most effective in changing teaching practice.

In considering the helpful context for learning of mathematics for teaching, Ball and Bass (2000) discuss a big challenge for which teachers are required to integrate across many kinds of knowledge, with this integration occurring in the context of particular teaching situations.

Since situations are classroom specific, the unpredictability makes it difficult for teachers to act on the spot. Focusing on student ideas in lessons and flexibly understanding what they mean, where they come from, and how to guide them to the goals set for the lesson requires the integration of the knowledge of students, content, and pedagogy. Curricular materials themselves do not readily present the richness of the mathematical ideas behind topics, nor do they help teachers gain the sufficient mathematical content knowledge necessary to teach the topic (Ball *et al.*, 2005; Remillard, 2000). Helping teachers learn the mathematics they need to know, and supporting them in understanding the connections between the particular knowledge with student learning and pedagogy, is key to teachers' learning and development. Lesson study allows teachers to focus on a specific topic and classroom context so that they can make sense of the meaning of the experience in their everyday teaching. However, their learning does not end there. What teachers learn in lesson study extends further than their immediate professional development context.

Teacher Learning with Lesson Study

Three-part teacher learning model

In examining the development and adaptation process of lesson study in the United States, Lewis *et al.* (2006) identified critical research needs, one being explication of the innovation mechanism. In order to understand how

lesson study supports instructional improvement, we need to better understand what happens with teachers in its process. Initially in the United States, people were interested in the curricular resources (e.g. lesson plans) teachers produce as results of lesson study to improve instruction. While that was a reasonable expectation, after several years of lesson study effort, we are now in a better position to understand that in the area of supporting instructional improvement, lesson study helps produce a lot more than mere lesson plans. Murata *et al.* (2004) suggest areas that develop and interact to support teacher learning in lesson study, and the modified model is shown in **Figure 2**. The three broad areas are teachers' knowledge, teachers' commitment and community, and learning resources (see also Lewis *et al.*, 2006).

As Fernandez (2005) also mentions in her study, lesson study provides opportunities for teachers to develop their pedagogical content knowledge. As different types of knowledge (e.g., knowledge of content, curricula, and student learning) essentially come together and interact with one another in the lesson study cycle (**Figure 1**), it creates an ideal context for which teachers combine these types of knowledge to make content accessible to their students. While different types of knowledge are typically learned in different settings by teachers in traditional professional development approaches (e.g., attending a lecture on math content and reading a book on classroom management), lesson study requires them to come together and work interdependently to support student learning in

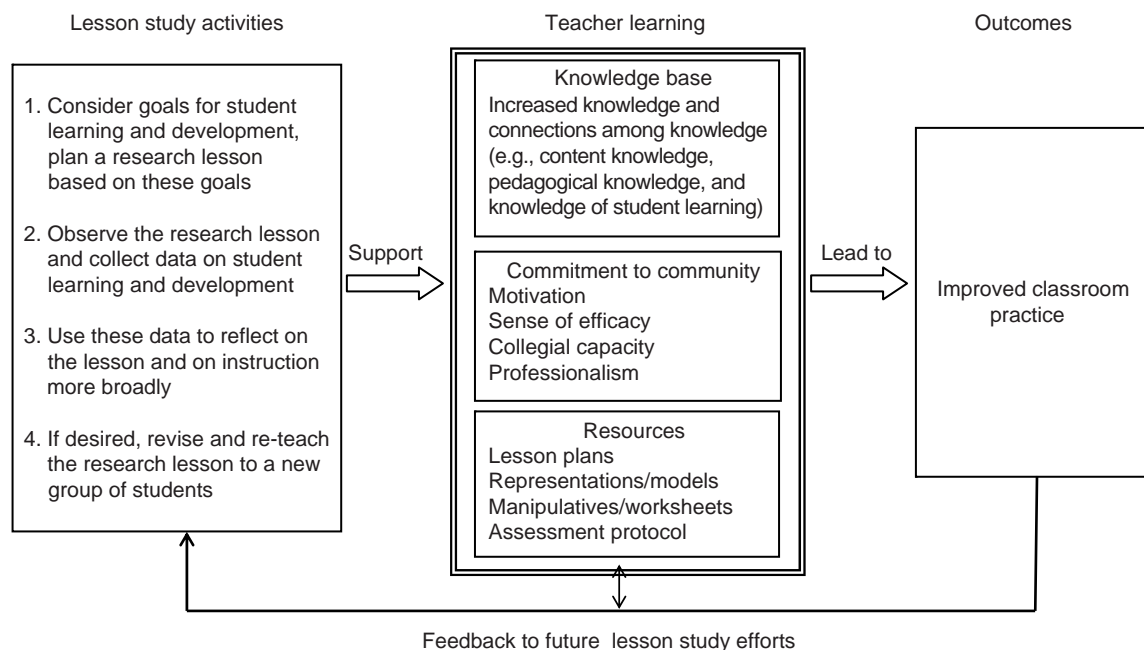


Figure 2 Lesson study activities, teacher learning, and outcomes. Modified from Murata, A., Lewis, C., and Perry, R. (2004). Teacher learning and lesson study: Developing efficacy through experiencing student learning. In McDougall, D. (ed.) *Proceedings of the Twenty-Sixth Annual Meeting of North American Chapter of the International Group of the Psychology of Mathematics Education*, pp. 985–992. Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

the very practice of teaching, thus helping teachers experience different types of knowledge in a coherent and related whole.

Development of knowledge and connections among different types of knowledge support, and are supported by, the communities teachers create through lesson study. While teaching is considered as an independent practice in the United States and elsewhere, lesson study brings teachers who are otherwise isolated in their school structure to come together and work collaboratively. It is likely that teachers who teach similar content to similar students will have similar questions and issues about teaching. When these teachers gather and share their ideas and resources, a meaningful learning community is created, and the sense of belonging and professionalism developed in the community helps teachers to commit to their profession and motivates them to continually improve their practice (Grossman *et al.*, 2001). For the teachers who collaboratively plan a research lesson, the process helps to add purpose to their everyday work. Their everyday life is experienced as a part of the larger professional endeavor among colleagues and a purposeful activity for which different events in the process have reasons for happening.

Most obviously, the development and improvement of learning resources is a part of the lesson study process. Teachers' knowledge development and professional community growth interact with the development of resources (e.g., lesson plans), and as the resources are refined and improved, they provide a meaningful context for teachers to discuss student learning and to focus on the lesson. Just as young students find it helpful to have hands-on manipulatives to develop mathematical concepts, lesson plans become concrete scaffolds for teachers to focus their attention and learn about the specific content area in discussion.

The three areas (knowledge development, community development, and material development; **Figure 2**) are essential for instructional improvement with lesson study to support teacher learning. Many professional development programs aim only to help teachers develop knowledge for teaching. While the single-focus approach is effective in certain ways, when considering the sustainability of professional growth and teacher motivation, the three-part teacher-learning model identifies the interactive relationships among different areas of the model.

Focus on student learning

When it comes to mathematics teaching and learning, the new reform requires teachers, as they practice in classrooms, to balance and juggle existing knowledge of students, content, curriculum, and pedagogy while incorporating new ideas to make the practice more conceptually strong and student centered (National Council of Teachers of Mathematics, 2000). Teaching is viewed as an interactive process for which student learning and mathematics content

come together through effective teacher facilitation. The interactive teaching requires teachers to know how students typically think and express their understanding so that teachers can effectively facilitate their learning by weaving different ideas and providing experiences that encourage connection building among concepts and ideas. What binds different parts of the lesson study cycle is student learning, as teachers identify goals in terms of student learning of a topic, investigate curricular materials that teach the topic, plan a lesson to make student learning visible in the classroom with the topic, gather data in the lesson, and, afterward, discuss the student learning that occurred during the lesson. Teachers typically become increasingly knowledgeable about the particular topic (content) and student learning of the topic in the process, and learn to listen to their students' ideas.

One of the most important aspects of teacher learning that emerges from the lesson study process is the new way to see teaching as a series of activities of inquiry around student learning. It helps cultivate the new attitude toward teaching, that teaching is not a one-way and didactic path, but a two-way integration of student ideas and content in a meaningful manner facilitated by teachers, and that can be challenging. The strong focus on student learning in the lesson study process continually reminds teachers how important it is for them to understand students' ideas in good teaching and helps bring the visions of reform closer to their actual classroom practices.

Overcoming the challenges of adapting lesson study

The Columbia University Teachers College Lesson Study Group presented several US lesson study cases and identified central characteristics of the participation that limited teachers' learning. When working with Japanese colleagues, the US teachers were challenged to find a strong research focus and to stay with the research process with lesson study (Fernandez *et al.*, 2003). The US teachers also struggled with developing a meaningful research hypothesis, developing means to explore the hypothesis, using evidence to make claims, and generalizing the findings. In another study for which the US and Japanese teachers were interviewed about their lesson study experiences, the researchers found that the US teachers were more likely to describe content goals (e.g., learning how to add fractions) in disconnect to other goals (e.g., student disposition) and focused heavily on what teachers do in lessons and not on student discovery and autonomy (Fernandez and Cannon, 2005). Fernandez also investigated how teachers took advantage of learning opportunities that were created by lesson study (Fernandez, 2005), and, in the study, the lack of strong mathematics content knowledge and reasoning skills kept the teachers from taking full advantage of the opportunities to learn. However, the author also describes how the teachers in the

study collaboratively anticipated and discussed their students' thinking, revised and taught a lesson multiple times, and reflected on particular aspects of student thinking of mathematics that supported their learning as teachers.

One of the strengths of lesson study is that it places teachers in the middle of their learning process. In order for teachers to take full advantage of the opportunities, it requires them to have a research-oriented and inquisitive disposition. However, if the teachers do not have the prerequisite, the disposition can be gradually developed through participating in the lesson study process. Opportunities provided through lesson study support teachers to develop knowledge and research skills and engage in their future lesson study in more effective and meaningful ways. While it will take longer for beginning lesson study participants to learn to refine the critical research process, in most cases, these teachers will become familiar with the expectations by their second or third lesson study experiences. In the meantime, the sense of community and new professionalism will keep the teachers motivated. Thus, these challenges found in the case studies mentioned above should be considered the necessary learning steps for teachers who are beginning to see teaching as a research process for the first time.

Research Lesson: Centerpiece of Lesson Study

As mentioned previously, the research lesson is central to the lesson study process. In surveying 125 Japanese teachers, Murata and Takahashi (2002) found that teachers identified the research lesson as the most important element of lesson study that helped their professional development. Through research lessons, teachers could see models of teaching and make sense of how these models affected student learning. The research lesson works to improve classroom practice, spread new content and approaches, connect classroom practice to broader educational goals, and explore conflicting ideas, thus creating demand, shaping national policy, and honoring the role of classroom teaching (Lewis and Tsuchida, 1998).

Research lessons are observed live, and they provide a special learning opportunity that teachers would not find otherwise. Unlike watching a video segment of classroom teaching or reading teaching episodes in books, live lessons are experienced as a whole. Different events unfold as the lesson flows with interactions among people in the classroom. These events cannot be understood by the analysis of separate parts of the whole alone (Davis and Summitt, 2003; Herbst, 2003). Classrooms are complex, and teachers have their own unique expert knowledge to understand this complexity. With their knowledge, they notice aspects of classroom experiences in their own ways and understand them as parts of a complex whole. They

see relationships between small events that may be invisible for people who have not spent time in classrooms before as teachers. When experienced teachers come together and observe a live lesson, their expert knowledge comes to the surface as they interpret the effectiveness of the lesson and discuss it in the debriefing. The novice teachers who experience the lesson with experienced teachers are apprenticed into their knowledge through participation.

In research lessons, outside commentators are typically invited to share their ideas at the end of the debriefing sessions (Watanabe, 2002). These commentators can be university researchers who specialize in the very subject areas that are addressed in the research lessons, or other experienced teachers who are interested in the topics. The main expectation for these commentators is to bring out the characteristics of the observed research lesson, tie them to research or theories of teaching-learning and/or conceptual development of students, and present a bigger picture of what their observations mean in the field of education. Unlike typical school consultants who observe and give feedback on aspects of teaching, focusing on what teachers should do to make the particular lesson better, the research lesson commentators pull together the different ideas and data shared in the debriefing to present a coherent picture of student learning. It requires good knowledge of the addressed topic, experiences in classrooms, and certain personal communication styles to be an effective outside commentator, and with the short history of lesson study in countries other than Japan, this is an area that requires attention and development.

Summary and Future Research on Teacher Learning with Lesson Study

This article attempted to give an overview of lesson study and research conducted in terms of teacher learning with lesson study thus far. As mentioned, we are still at a very early stage of understanding how teachers learn with lesson study, and there are several key research possibilities for the future.

Since the lesson study process, especially the research lesson context, makes the teaching process and teachers' thinking visible, teachers' knowledge and its development can be studied in the settings. It is yet unclear what teachers' expert knowledge consists of and how different parts of this knowledge interact in effective teaching. Lesson study creates an ideal research setting for the investigation. Especially when experienced and novice teachers engage in lesson study together, the characteristics of the expert knowledge may surface in their work and discussions. Since lesson study activities focus primarily on student learning, we suspect that the expert knowledge includes a deep understanding of how students think

about particular topics or subjects at hand. The way teachers understand it and how teachers see their understanding as helpful for the planning and teaching of lessons will be important for educational researchers and teacher educators to know, so that we can effectively support all teachers' learning in lesson study as well as in other settings.

When schools and teachers face challenges in adopting lesson study to the existing school and classroom systems, these challenges should be carefully examined, explained, and understood locally. For any innovation, initial challenges are expected, and they provide a way for us to better understand the existing system as well as the innovation. If a challenge is due to a structural incompatibility (e.g., teachers do not have common planning time in their school day), administrative intervention may be sufficient to overcome the difficulty. If a challenge is due to societal difference (e.g., teachers are not familiar with the inquiry process), the lesson study structure may be modified to support teachers to learn about the research process to participate in meaningful ways for their own professional development and instructional improvement. As Lewis *et al.* (2006) discuss, we do not want lesson study to be considered ineffective after just a few years' implementation with mismatched program evaluations. In order to nurture lesson study in different contexts, we need to take time to understand the emerging challenges and carefully adjust and modify parts of the existing innovation while it continues to support teacher learning. Understanding various challenges with an innovation would illuminate aspects of teacher learning that would never come to light in original contexts alone.

In the twenty-first century, we have virtual and electronic means to connect lesson study groups across a nation or globally. There are virtual lesson study LIST-SERV that connect people across the United States, providing a context for them to ask questions, share progress they make with one another, and invite each other to research lesson events (e.g., Lesson Study Research Group, 2007). Some lesson study groups are connected electronically to collaborate together (e.g., Math Star, 2007). With lesson study being a new innovation, the regular flow of new information is essential for teachers' work and collaboration to be successful. This also provides an ideal research context for which we can investigate the challenges from teachers' points of view. We may also evaluate the effectiveness of electronic community and technology use for adopting innovation and understand the role they play in professional community development and learning. How well a certain virtual community is used and kinds of information exchanged will not only help identify the effectiveness of the virtual means of community but also aspects of lesson study, its challenges, and teacher learning in the process.

As we try to come to a shared model of teacher learning with lesson study, we need research that synergistically investigates different aspects of teacher learning as well as the aspects of lesson study to understand the learning process better. International collaboration may be the next step in establishing a research community, as we need to understand what we know at this point as a community that will inform what we need to know next.

Bibliography

- Ball, D. L. and Bass, H. (2000). Bridging practices: Intertwining content and pedagogy in teaching and learning to teach. In Boaler, J. (ed.) *Multiple Perspectives on Mathematics Teaching and Learning*, pp 83–104. Westport, CT: Ablex Publishing.
- Ball, D. L. and Cohen, D. K. (1996). Reform by the book: What is – or might be – the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher* 25(9), 6–8 and 14.
- Ball, D. L., Hill, H. C., and Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator* (Fall 2005) 14–22 and 43–46.
- Bell, B. and Gilbert, J. (2004). A model for achieving teacher development. In Gilbert, J. (ed.) *The Routledge Falmer Reader in Science Education*, pp 258–278. New York: Routledge Falmer.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher* 33(8), 3–15.
- Cochran-Smith, M. and Lytle, S. (1999). The teacher research movement: A decade later. *Educational Researcher* 28(7), 15–25.
- Cochran-Smith, M. and Lytle, S. (2001). Beyond certainty: Taking an inquiry stance on practice. In Lieberman, A. and Miller, L. (eds.) *Teachers Caught in the Action: Professional Development in Practice*, pp 45–60. New York: Teachers College Press.
- Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis* 12, 311–329.
- Conference on Learning Study (2006). Programme. <http://www.ied.edu.hk/cdspfe/lsconference/programme.htm> (accessed December 2008).
- Choksi, S. and Fernandez, C. (2004). Challenges to improving Japanese lesson study: Concerns, misconceptions, and nuances. *Phi Delta Kappan* 85(7), 520–525.
- Darling-Hammond, L. (1994). Developing professional development schools: Early lessons, challenges, and promise. In Darling-Hammond, L. (ed.) *Professional Development Schools: Schools for Developing a Profession*, pp 1–27. New York: Teachers College Press.
- Davis, B. and Summitt, E. (2003). Understanding learning systems: Mathematics education and complexity science. *Journal for Research in Mathematics Education* 34, 137–167.
- Fernandez, C. (2005). Lesson study: A means for elementary teachers to develop the knowledge of mathematics needed for reform-minded teaching? *Mathematical Thinking and Learning* 7(4), 265–289.
- Fernandez, C. and Cannon, J. (2005). What Japanese and U.S. teachers think about when constructing mathematics lessons: A preliminary investigation. *Elementary School Journal* 105(5), 481–498.
- Fernandez, C., Cannon, J., and Choksi, S. (2003). A U.S. – Japan lessons study collaborative reveals critical lenses for examining practice. *Teaching and Teacher Education* 19, 171–185.
- Fernandez, C. and Yoshida, M. (2004). *Lesson Study: A Case of a Japanese Approach to Improving Instruction through School-Based Teacher Development*. Mahwah, NJ: Lawrence Erlbaum.
- Franke, M., Carpenter, T. P., Levi, L., and Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American Educational Research Journal* 38(3), 653–689.

- Fujita, H., Hashimoto, Y., Hodgson, B. R., et al. (2004). *Proceedings of the Ninth International Congress on Mathematics Education*. Norwell, MA: Kluwer.
- Fuller, F. F. (1969). Concerns of teachers: A developmental characterization. *American Educational Research Journal* **6**, 207–226.
- Fuller, F. F. and Bown, O. H. (1975). Becoming a teacher. In Ryan, K. (ed.) *Teacher Education*, pp 25–52. Chicago, IL: Chicago University Press.
- Grossman, P., Wineburg, S., and Woolworth, S. (2001). Toward a theory of teacher community. *Teachers College Record* **103**(6), 942–1012.
- Hawley, W. D. and Valli, L. (1999). The essentials of effective professional development: A new consensus. In Darling-Hammond, L. and Sykes, G. (eds.) *Teaching as the Learning Profession: Handbook for Policy and Practice*, pp 127–150. San Francisco, CA: Jossey-Bass.
- Herbst, P. (2003). Using novel tasks in teaching mathematics: Three tensions affecting the work of the teacher. *American Educational Research Journal* **40**, 197–238.
- Lesson Study Research Group (2007). *Timeline of U.S. Lesson Study*. <http://www.tc.edu/lessonstudy/lsgroups.html> (accessed December 2008).
- Lewis, C. (2002). *Lesson Study: A Handbook of Teacher-Led Instructional Change*. Philadelphia, PA: Research for Better Schools.
- Lewis, C., Perry, R., and Hurd, J. (2004). Deeper look at lesson study. *Educational Leadership* **61**(5), 18–23.
- Lewis, C., Perry, R., and Murata, A. (2006). How should research contribute to instructional improvement: The case of lesson study. *Educational Researcher* **35**(3), 3–14.
- Lewis, C. and Tsuchida, I. (1998). A lesson is like a swiftly flowing river: Research lessons and the improvement of Japanese education. *American Educator* (Winter) 14–17 and 50–52.
- Little, J. W. (2001). Professional development in pursuit of school reform. In Lieberman, A. and Miller, L. (eds.) *Teachers Caught in the Action: Professional Development that Matters*, pp 28–44. New York: Teachers College Press.
- Lo, M. L. (2003). Lesson study and its impact on teacher development. *ISSP Newsletter* 4–5.
- Math Star New Mexico Lesson Study (2007). Lesson study New Mexico style. http://pt3.nmsu.edu/lesson_study/index.html (accessed December 2008).
- Murata, A., Lewis, C., and Perry, R. (2004). Teacher learning and lesson study: Developing efficacy through experiencing student learning. In McDougall, D. (ed.) *Proceedings of the Twenty-Sixth Annual Meeting of North American Chapter of the International Group of the Psychology of Mathematics Education*, pp 985–992. Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.
- Murata, A. and Takahashi, A. (2002). District-level lesson study. *How Japanese teachers improve their teaching of elementary mathematics*. Paper Presented at a Research Pre-Session of the Annual Meeting of the National Council of Teachers of Mathematics, Las Vegas, NV.
- National College for School Leadership (2004). Research lesson study: Experimenting with collaborative learning. *Networked Learning Group Publication* **6**, 30–31.
- National Council of Teachers of Mathematics (2000). *Principles and Standards of School Mathematics*. Reston, VA: Author.
- National Research Council (2002). *Studying Classroom Teaching as a Medium for Professional Development: Proceedings of a U.S.–Japan Workshop*. Washington, DC: National Academy Press.
- North Regional Educational Laboratory (2002). *Teacher to Teacher: Reshaping Instruction through Lesson Study*. Naperville, IL: North Central Regional Educational Laboratory.
- Remillard, J. T. (2000). Can curriculum materials support teachers' learning? Two fourth-grade teachers' use of a new mathematics text. *Elementary School Journal* **100**(4), 331–350.
- Richardson, J. (2004). Lesson study: Teachers learn how to improve instruction. In *Tools for Schools*, pp 1–3. Oxford, OH: National Staff Development Council.
- Sherin, M. G. (2002). When teaching becomes learning. *Cognition and Instruction* **20**(2), 119–150.
- Shimizu, S., Isoda, M., Okubo, K., and Baba, T. (eds.) (2005). *Mathematics Lesson Study in Japan through Diagrams*. Tokyo: Meiji Tosho.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review* **57**(1), 1–21.
- Shulman, L. S. (1996). Just in case: Reflections on learning from experience. In Colbert, J. A., Desberg, P., and Trimble, K. (eds.) *The Case for Education: Contemporary Approaches for Using Case Methods*, pp 197–217. Boston, MA: Allyn and Bacon.
- Smith, M. S. (2000). Balancing old ad new: An experienced middle school teacher's learning in the context of mathematics instructional reform. *Elementary School Journal* **100**(4), 351–375.
- Stepanek, J. (2001). A new view of professional development. *Northwest Teacher* **2**(2), 2–5.
- Stepanek, J. (2003). A lesson study team steps into the spotlight. *Northwest Teacher* **4**(3), 9–11.
- Stigler, J. W. and Hiebert, J. (1999). *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom*. New York: Summit Books.
- Thompson, A. G. (1985). The relationship of teachers' conceptions of mathematics and mathematics teaching to instructional practice. *Educational Studies in Mathematics* **15**, 105–127.
- Wang, J. and O'Dell, S. J. (2002). Mentored learning to teach according to standards-based reform: A critical review. *Review of Educational Research* **72**(3), 481–546.
- Wang-Iverson, P. and Yoshida, M. (2005). *Building Our Understanding of Lesson Study*. Philadelphia, PA: Research for Better Schools.
- Watanabe, T. (2002). Learning from Japanese lesson study. *Educational Leadership* **59**(6), 36–39.
- Wilms, W. W. (2003). Altering the structure and culture of American public schools. *Phi Delta Kappan* **85**(8), 606–613.
- Wilson, S. M. and Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education* **24**, 173–209.
- Wood, T., Cobb, P., and Yackel, E. (1991). Change in teaching mathematics: A case study. *American Educational Research Journal* **28**, 587–616.

Further Reading

- Fernandez, C. (2002). Learning from Japanese approaches to professional development. *Journal of Teacher Education* **53**, 393–405.

Teachers Career Stages and Professional Development

R Fessler and E Rice, Johns Hopkins University, Baltimore, MD, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The literature on teacher career stages over the years has provided useful frameworks for assessing teacher needs and planning appropriate professional development activities. Foundational work in this area can be traced to developmental psychologists. Teacher career stage conceptualizations that emerged in the 1970s were based primarily on observations and reflections of authors, resulting in models of sequential stages that mirrored the timeline of teachers' experiences. Later elaborations, often based on teacher interviews, provided greater differentiation within stages and acknowledged some individual differences. These models, however, were still fairly static, assuming that teachers progressed through stages in a linear, lock-step fashion. Finally, subsequent approaches have introduced more dynamic models that describe alternate pathways and variations brought on by organizational and personal environmental factors that impact careers.

Recent dramatic environmental changes require a reassessment of the way we view career stages and teacher development. These include organizational factors, such as teacher shortages and the concomitant creation of fast-track routes to teaching, and personal factors, such as the changing values brought to teaching by both career changers and new generations of young people entering the profession. By understanding how the changing terrain of teaching is impacting the stages of teacher career development, it may become possible to better address the needs of an evolving teaching population.

Approaches to Teacher Career Stages

Early Foundations

The knowledge base about teachers' development is rooted in a more general literature addressing the stages of human development. Indeed, the whole notion that teachers go through a set number of stages is traceable to Erikson's (1959) writings on the stages of adult development. Erikson's work on the development of identity also gave rise to seminal works on the development of cognitive complexity and moral development. Frances Fuller was the first to apply a similar model to teachers. Fuller (1969) linked categories of concerns to phases of teachers' careers, from early concerns for self to concerns for mastering the tasks of teaching, to eventual concerns

about the impact of one's practice on students. Fuller's theories led to a model of personalized teacher education designed to match the phases experienced by teachers, introducing the idea that professional development must be differentiated in accordance with teacher needs.

While the earliest models of teacher career stages built upon the work of Fuller, they generally derived their categories from common-sense observations and reflections on the careers of teachers (Unruh and Turner, 1970; Gregorc, 1973; Katz, 1972). These models, while espousing somewhat different sets of stages, all emphasized preservice, induction, and maturity. While Gregorc and Katz each offered a fourth stage, the general model was still one where the beginning teacher moves through one or two stages to become a mature, fully functioning professional. All these models were consistent with more general notions about how people progressed through careers: they are trained, they go through an initial introduction/learning period on the job, and finally they do their job. Little attention was paid to individual differences or variations in teachers' careers, and there was no analysis of environmental factors that might impact career development. Still, these descriptive models did provide an important foundation for further research, and they helped focus attention on the unique professional development needs of preservice and new teachers.

Research-Based Linear Models

In a series of studies at the Ohio State University in the United States, Kevin Ryan, along with Flora, Burden, Newman, and Patterson, began to examine the issue of the careers of teachers as a whole (Ryan *et al.*, 1979). Using personal interviews with teachers, these studies looked at first-year teachers, teachers with 4–20 years experience, teachers with 20–30 years experience, and retired teachers. The researchers documented the changing concerns of teachers as well as their implications for professional development, showing how experienced teachers' needs changed from technical skills to larger issues such as how to teach creatively. However, they still presented teacher development as a static and unidirectional process, with mature teachers treated as one homogeneous group. The later work by Burden (1982), while valuable for its use of a data-driven approach, continued to share many of the characteristics of earlier theorists. Given that his focus in the Ohio State studies was on teachers with 4–20 years of experience, it is interesting

that Burden's proposed stages of survival, adjustment, and maturity were very similar to the schemas of Unruh and Turner or Katz. A major limitation that Burden shared with earlier authors was his concentration on the first few years of teacher development, positing a mature or fully functioning teacher by the fourth or fifth year of a teacher's career. He, like many earlier theorists, had little to say about differentiated teacher development after the first few years.

Taken as a whole, though, the Ohio State studies did bring needed change to conceptualizations of teacher career stages. The focus on research, along with the description of individual differentiation within career phases and a sophisticated analysis of these individual differences, vastly expanded the descriptive knowledge about teacher careers. However, these studies still did not provide a conceptual framework for thinking about teacher careers other than as a linear model, and they paid little attention to either external or personal environmental factors that might differentially impact the careers of individual teachers.

Nonlinear Dynamic Models

Throughout the 1980s and into the 1990s, several researchers began to propose more complex, multidimensional analyses of teacher career stages. The two notable characteristics shared by these approaches were a dynamic, nonlinear view of teacher development and a focus on the role played by various environmental factors in the specific developmental paths taken by a given teacher. Huberman (1989) introduced the notion of career trajectories that describe a series of pathways, or options, that occur during teachers' careers. His themes/phases were derived from interviews with women and a more general population, resulting in the documentation of specific pathways for different groups. In his generic model, teachers begin with concerns about survival and discovery and move into a period of stabilization, followed by multiple pathways that differentiate based on their experiences and responses to their changing environment. He further suggested that professional development must respond to these differentiated experiences and trajectories.

Another nonlinear approach emerged in the early work on the Teacher Career Cycle Model (Burke *et al.*, 1984). These researchers observed teachers' career pathways through direct observations, teacher interviews, case studies, and literature reviews to develop a model that presented a social-systems orientation to teacher career stages. As most fully developed (Fessler and Christensen, 1992), the Teacher Career Cycle Model emphasized eight phases: preservice, induction, competency building, enthusiastic/growing, career frustration, career stability,

career wind-down, and career exit. Individual teachers' paths through these stages, though, were uniquely impacted by both organizational and personal environmental factors. Organizational factors included features such as school regulations, the management style of leaders, the atmosphere of public trust (or a lack of it), societal expectations, and the role of professional organizations and unions. Personal factors included considerations such as the family life of the teacher, positive events or crises in the teacher's life, individual dispositions, outside interests, and the life stage of the teacher. What differentiated this approach from earlier models was that it postulated a dynamic ebb and flow, with teachers moving in and out of stages in response to influences from both the personal and organizational environmental dimensions.

While these various approaches have served as useful methods of analysis for practitioners and researchers for several decades, recent dramatic environmental changes require a reassessment of career stages and their implications for teacher professional development.

Emerging Environmental Changes and Career Stage Theory

Virtually all earlier models of teacher development were based on several assumptions that may no longer hold true. One such assumption is the existence of well-defined, easily differentiated periods of preservice and in-service education. In addition, the more static models of career phases assumed a long-term commitment to a career in teaching, and even the developers of more environmentally sensitive models did not contemplate teaching careers of only 3–5 years. Environmental changes since the early 1990s have challenged these assumptions. These changes flow from both organizational and personal environmental factors.

Organizational and External Environmental Factors

While in 1992 Fessler and Christensen could still argue that teacher turnover was declining, by the end of that decade the teacher shortage had become perhaps the most written about problem in education. Many educational authorities have responded by experimenting with fast-track routes to the classroom. Although many of these efforts have resulted in high-quality programs, some allow new teachers to enter the profession with only a summer of training, and there have been increasing calls to allow individuals with specialized content knowledge to bypass traditional teacher preparation programs and move directly into the classroom based on their demonstrated

content knowledge. In addition, various alternative routes to teaching have attracted an increasing number of career changers to the teaching workforce. The existence of such programs blurs the distinction between pre- and in-service stages, and for many new teachers it may no longer make sense to separate preservice from induction and competency-building stages. This suggests that new models of teacher career stages must allow for the reality that current teachers simultaneously inhabit more than one career phase. Similarly, with programs such as Teach for America in the United States requiring only 2 years of service from their participants, the entire career cycle may be compressed and truncated for a growing number of teachers.

The teacher career cycle is currently impacted by a number of additional changing organizational environmental factors. Numerous calls for accountability in schools have resulted in an increased emphasis on high-stakes testing of students. Often the desire to enforce accountability and content-driven definitions of teacher quality has led to the marketing of curricula that are much more structured than past approaches. In their most extreme forms, these curricula change the role of teachers from that of creative professional decision makers to actors following a script. A concomitant shift to more top-down management styles in some school districts has led to less flexibility for creative teachers. Even those districts now moving back toward site-based management are empowering principals and other school leaders more than teachers. Finally, the challenges presented by urban, urbanizing, and poor rural school districts greatly impact the conditions under which teachers are asked to do their jobs. These teachers often face a dearth of resources, outdated textbooks, constantly changing leadership, and lack of goal clarity.

As described in the section titled 'Personal environmental factors', new generations of teachers may have difficulty functioning in these evolving working conditions. Indeed, it is likely that a combination of these changing environmental factors has resulted in an increase in teachers moving into the career exit stage. Many urban districts in the United States lose 50% of their new teachers within the first 5 years of teaching. For this reason, recent researchers have increasingly focused on the initial phases of teacher development as practitioners strove to deal with the demands for many thousands of new teachers. Post-initial phases of professional development have received much less attention during this era, and some of the more recent work has actually returned to less dynamic models, perhaps because of the stress on induction and mentoring of new teachers.

Some researchers studying the teacher shortage have argued that the problem is not actually one of teacher

recruitment, but of teacher retention (Ingersoll, 2001; NCTAF, 2003). This research has led to a variety of approaches meant to keep teachers in the classroom longer, including re-recruitment models to bring former teachers back to the classroom and providing classroom-embedded teacher leadership opportunities rather than requiring those with leadership aspirations to exit teaching to become administrators.

These changing organizational factors have had enormous impact upon teacher career stages. While there have been some positive developments, the overwhelming trend has been one of a decrease in teacher morale and a corresponding increase in career frustration leading to early career exit. Even those who stay in education for the long term are exiting the teaching profession earlier. Pressures to promote good teacher leaders to administrative positions (where many schools also face shortages) cause more good teachers to leave the classroom prematurely. This further points to a need to redefine the terms used to describe teacher career development.

Solving the twin crises of the teacher shortage and teacher retention will require new strategies. Teachers need professional development that provides them the necessary tools to make sense of the vast amounts of data provided by the accountability movement in ways that help them to improve their own instruction. The goal can be conceptualized as finding ways to continue to return mid- and late-career teachers to the enthusiastic and growing stage, perhaps through offering them the opportunities to build new competencies. This reinforces the need for classroom-embedded teacher leadership opportunities that do not lead to an administrative track. Some examples of this possibility are indeed being tried. Both professional-development schools and mentoring programs provide opportunities for mid- and late-career teachers to develop new competencies. There is at least some evidence that this helps teachers reenter the enthusiastic and growing stage of teacher professional development. For example, an unpublished survey of experienced teachers acting as peer mentors for beginning teachers in Baltimore, Maryland (USA) showed that the mentors felt reinvigorated in their own teaching by the opportunity to share their knowledge with their younger colleagues. There is also evidence that such mentoring helps to keep new teachers teaching longer, thus increasing the number of mid- and late-career teachers available for such leadership tasks (Smith and Ingersoll, 2004). Finally, whatever its other problems or virtues, the charter school movement in the United States often provides a place for teacher-leaders to guide their school's direction. These teachers are building new skills and are often re-engaged with questions of teaching and learning. All of these provide examples of what could be called a reinvigorating stage of teacher professional development.

Personal Environmental Factors

Teacher shortages in many countries have led to a variety of experiments in new ways to bring teachers into the classroom. These include efforts to attract career changers and young people. Each group brings new personal environmental factors with implications for professional development. Career changers bring their own unique traits to the teaching workforce. While their content knowledge is often wide and deep, they sometimes exhibit limited awareness of the characteristics of modern youth, especially in poor urban and rural areas (those areas in greatest need of teachers). This group of new teachers will thus need less professional development about rapidly changing content areas, but correspondingly more around pedagogical issues and approaches to teaching. They often enter teaching after careful reflection, suggesting that they may remain committed to teaching longer than some of their younger counterparts.

There is an emerging literature that addresses the values and characteristics of new generations of young adults that influence the way they look at the world of work and their career pathways (Howe and Strauss, 2000; Lancaster and Stillman, 2002; Raines, 2003). The Generation Xers, who were born between about 1970 and 1981, are generally described as being very skeptical of authority and bureaucracy, yet respectful of competence, and as individuals who work hard to advance their careers. They are oriented toward results, often very efficient and competent, and tend to be independent. Teachers in this category often see teaching as a steppingstone to advancing their careers – either within or outside of education. This group presents challenges and opportunities to school leadership, who are faced with the need to focus these talented individuals, while at the same time recognizing that many do not see teaching as a long-term career.

Millennials, who were born between about 1982 and 2000, are described as a generation that has been raised by parents who have been very nurturing and supportive. They have a great deal of confidence in themselves and high self-esteem. One of their defining characteristics is their heavy reliance on technology for personal and professional communications. As a group, they are described as wanting to make a difference, optimistic and hopeful about the future, but with little patience for bureaucracies that they do not agree with or understand. They expect a workplace that is challenging, collaborative, creative, and fun, with opportunities for both advancement and financial rewards. They demand inclusiveness, diversity in the workplace, and respect for all. New models of professional development may be needed for individuals demonstrating these characteristics, with a focus on their strengths, group work, efficiency, and increased involvement in their own growth plans (Raines, 2003). They are likely to be

receptive to setting goals for students that are fair and challenging and that reflect a sense of social justice.

These generalized characteristics of the emerging workforce seem to be in direct conflict with some of the organizational factors mentioned in the previous section. Further work is needed to create an organizational climate that is responsive to the personal needs of the generation Xers and millennials in order to tap their rich potential as effective teachers. Millennials' idealism and desire to contribute to social justice issues should be strong factors in recruiting them into teaching, but collaborative, nurturing, supportive environments and leadership styles will be needed to keep them in the classroom. Finally, expectations for financial rewards and career advancement might make these new-generation teachers more receptive than their predecessors to ideas such as merit pay and new versions of career ladders.

New Directions for Career Stage Connections to Professional Development

Based on the evolving environmental influences on teachers' careers discussed above, the following suggestions are offered for further development of career cycle research and applications:

1. There is a need to continue to examine changing environmental conditions and their impact on teacher career stages and professional development. The environmental changes outlined in this article reflect developments of the past 15–20 years, but they too will evolve as the profession faces new challenges and opportunities in the years ahead.
2. Researchers need to consider new conceptualizations of career stages and pathways that take into consideration emerging alternative teacher preparation models and subsequent professional development needs. Those prepared through alternative routes that include minimal professional education experiences will need support in building their competencies in pedagogy during their first few years of teaching.
3. Opportunities for teachers to be deeply involved in planning and decision making seem consistent with the needs and expectations of the generation Xers and millennials, but are often lacking in our top-down, accountability-driven, high stakes-testing environment. We will need to reconcile this conflict if we are to attract and keep a new generation of enthusiastic teachers committed to quality, diversity, and high achievement for all students.
4. A closer examination of generational differences (including newly evolving ones) is needed to determine the implications for professional development for early

career teachers, for collaboration among intergenerational school personnel, and for assisting senior leaders in understanding how to work with and tap the idealism of the millennials and the career-driven expectations of the generation Xers. As mentioned earlier, these new-era teachers may be receptive to merit pay and career-ladder models, but this might create conflict with their more senior colleagues.

5. Career state research should be further emphasized as a foundation for addressing teacher shortages as a retention issue by providing appropriate professional development and support to give early- and mid-career teachers more positive experiences and reasons to stay. This should include providing more opportunities for teachers to take on leadership roles that do not require them to leave the classroom, such as through mentoring and professional-development schools.
6. School leaders must address the potential disconnect between the increased emphasis on standardized teaching associated with testing and teaching for tests, on the one hand, and the need for creative teachers to be more involved in instructional planning and decision making, on the other. Can this be reconciled by more sensitivity to the needs of individuals at various stages of their careers? Can leaders keep teachers involved while still emphasizing accountability?

Further research and refinement of career stage models will be useful in addressing these and related issues facing educators in the coming generations and will move us forward in our understanding of post-initial stages of teacher professional development.

Bibliography

- Burden, P. (1982). Implications of teacher career development: New roles for teachers, administrators, and professors. *Paper Presented at the Annual Workshop of the Association of Teacher Educators*, Slippery Rock, PA, USA, August.
- Burke, P., Fessler, R., and Christensen, J. (1984). *Teacher Career Stages: Implications for Staff Development*. Bloomington, IN: Phi Delta Kappa Educational Foundation.
- Erikson, E. H. (1959). *Identity and the Life Cycle: Selected Papers*. New York: International Universities Press.
- Fessler, R. and Christensen, J. (eds.) (1992). *The Teacher Career Cycle: Understanding and Guiding the Professional Development of Teachers*. Boston, MA: Allyn and Bacon.
- Fuller, F. F. (1969). Concerns of teachers: A developmental conceptualization. *American Educational Research Journal* 6(2), 207–226.
- Gregorc, A. F. (1973). Developing plans for professional growth. *NASSP Bulletin* (Dec.), 1–8.
- Howe, N. and Strauss, W. (2000). *Millennials Rising: The Next Great Generation*. New York: Vintage Books.
- Huberman, M. (1989). The professional life cycle of teachers. *Teachers College Record* 91(1), 31–57.
- Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal* 38(3), 499–534.
- Katz, L. G. (1972). Developmental stages of preschool teachers. *Elementary School Journal* 73(1), 50–54.
- Lancaster, L. C. and Stillman, D. (2002). *When Generations Collide: Who They Are, Why They Clash, How to Solve the Generational Puzzle at Work*. New York: Harper Collins.
- National Commission on Teaching and America's Future (NCTAF) (2003). *No Dream Denied: A Pledge to America's Children*. Washington, DC: Author.
- Raines, C. (2003). *Connecting Generations: The Sourcebook for a New Workplace*. Menlo Park, CA: Crisp Publications.
- Ryan, K., Flora, R., Newman, K., et al. (1979). The stages in teaching: New perspectives on staff development for teachers' needs (ASCD audio-tape). *Presentation to the Association for Supervision and Curriculum Development*.
- Smith, T. M. and Ingersoll, R. M. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal* 41(3), 681–714.
- Unruh, A. and Turner, H. E. (1970). *Supervision for Change and Innovation*. Boston, MA: Houghton Mifflin.
- Darling-Hammond, L. (1994). *Professional Development Schools: Schools for Developing a Profession*. New York: Teachers College Press.
- Darling-Hammond, L. and Bransford, J. (2005). *Preparing Teachers for a Changing World*. San Francisco, CA: Jossey-Bass.
- Erikson, E. H. (1982). *The Life Cycle Completed: A Review*. New York: Norton.
- Fessler, R. (1995). Dynamics of teacher career stages. Guskey, T. R. and Huberman, M. (eds.) *Professional Development in Education*, pp 171–192. New York: Teachers College Press.
- Fessler, R. and Ingram, R. L. C. (2003). The teacher career cycle revisited: New realities – new responses. Davies, B., West-Burnham, J., and Collarbone, P. (eds.) *The Handbook of Educational Leadership and Management*, pp 584–590. London: Pearson Education.
- Huberman, A. M., Grounauer, M., and Marti, J. (1993). *The Lives of Teachers (Vie des enseignants)*. New York: Teachers College Press.
- Sprinthal, N. A., Reiman, A. J., and Theis-Sprinthal, L. (1996). Teacher professional development. Sikula, J. (ed.) *Second Handbook of Research on Teacher Education*, pp 667–703. New York: Macmillan.
- Strauss, W. and Howe, N. (1991). *Generations, The History of America's Future, 1584 to 2069*, 1st edn. New York: Morrow.
- Steffy, B. E. and Kappa Delta Pi (2000). *Life Cycle of the Career Teacher*. Indianapolis, IN: Kappa Delta Pi.

Further Reading

TEACHER EDUCATION – PRESERVICE TEACHER EDUCATION

Contents

A Pedagogy of Teacher Education
Case Methods in Teacher Education
Conceptions of Teacher Education
Inquiry-Oriented Teacher Education
Narratives and Biography in Teacher Education
Supervision in Teacher Education
Taking Prospective Teachers' Beliefs into Account in Teacher Education
Teacher Education and Models of Teacher Reflection
Teacher Education as Teaching for Understanding with New Technologies

A Pedagogy of Teacher Education

J Loughran, Monash University, Clayton, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

One way of understanding what a pedagogy of teacher education means is to think of it as comprising the processes and practices of teaching and learning about teaching, that is, a pedagogy of teacher education is an articulation of the knowledge of teaching and learning about teaching. Central to the idea of a pedagogy of teacher education is the use of the term pedagogy. Pedagogy is not meant as a synonym for teaching, rather pedagogy is about the relationship between teaching and learning. Therefore, a major purpose in articulating a pedagogy of teacher education is to ensure teacher educators' professional knowledge is translated into the practices of teacher education more generally.

The roots of a pedagogy of teacher education can be traced back to research in the 1980s where issues about pedagogy, in and for teacher education, initially gained currency. Liston and Zeichner (1988) worked with ideas about critical pedagogy and teacher education in an attempt to begin to pull together a number of the important structures and practices that they considered were at the heart of challenging the nature of teacher education. Berliner (1988) focused considerable attention on the pedagogy for teacher education as he advocated the need to better recognize and value the knowledge and skills essential to expert teaching and how that might impact teacher-education practices. However, a pedagogy of teacher education began to emerge as an area of interest in its own right as a consequence of the prevailing

approaches to researching teacher education and an interest in that which:

... happens behind the doors of those [teacher education] classes, about the ideas of teacher education, about the pedagogy of teacher education. ... discussions on a programmatic level are not fine grained enough to get at those ... What, I might ask, are students thinking, feeling, learning, believing (Wilson, 1990: 2)

The need to know more about teaching and learning about teaching also converged with concerns about the education of teacher educators. Therefore, the teaching and learning experiences of both teacher educators and student teachers became fundamental to a better understanding, that which comprises a pedagogy of teacher education.

A Growing Momentum

The shift in the research agenda, from studying teacher education from afar to something more upfront and personal, led to a conceptualization of teaching and learning about teaching through a pedagogy of teacher education.

Teacher educators must turn their attention to the content and pedagogy of teacher education. This includes considering the prior knowledge and beliefs that prospective teachers bring with them, as well as the models those novice teachers have to reflect upon and learn from.

(Ball and Wilson, 1990: 14)

In the early 1990s the ability of traditional teacher-education programs to positively impact, and therefore change, the nature of school teaching was increasingly questioned. Heaton and Lampert (1993) made clear the need for a “new language for talking about practice and new ways of interacting with learners” (p. 46). They argued that teacher educators themselves needed to be doing the same in their teaching about teaching, as they expected of their student teachers in their practice; so the notion of a pedagogy of teacher education became firmly established:

This means inventing not only a new approach to research but also a new pedagogy of teacher education: one where the actual problems of practice are central, where the question of what it means to know and learn . . . are integrated in ongoing work with students . . . [it] requires a new kind of discourse, new definitions, new rules of argument. (Heaton and Lampert, 1993: 46)

The Construct Takes Hold

Throughout the 1990s, ideas about a pedagogy of teacher education kept resurfacing through studies into teacher-education processes and practices as a stronger focus on teaching and learning about teaching took hold. However, when Korthagen *et al.* (2001) described the way in which a pedagogy of teacher education might be conceptualized, both philosophically and practically, in teacher educators’ practice, the construct resonated in new and powerful ways in the teacher education community.

Hamilton and Pinnegar (2000) viewed the development of a pedagogy of teacher education as being embedded in the work of self-study, or more so, as an outcome of self-study. Cochran-Smith (2000) viewed a pedagogy of teacher education as one way of better understanding the nature of learning in teacher education, such that more powerful professional development of teachers and teacher educators might be realized. Tigchelaar and Korthagen (2004) examined how to better link theory and student teachers’ experiences, in order to shed more light on what a pedagogy of teacher education might entail, while Doyle and Carter (2003) focused on a narrative perspective because they saw the knowledge base of teaching as residing in “the stories of experience as a teacher” (p. 134). Feiman-Nemser (2001) conceptualized a number of central tasks in learning to teach: (1) analyzing beliefs and forming new visions; (2) developing subject-matter knowledge for teaching; (3) developing an understanding of learners and learning; (4) developing a beginning repertoire (of teaching); and (5) developing the tools for studying teaching. These tasks gave a structure for ways of considering concrete aspects on what a pedagogy of teacher education might entail.

It is not difficult to see how this growing range of ideas and implications for better understanding the processes of teaching and learning about teaching and, in part, building on the shift from studying teacher education from afar to near by involving teacher educators themselves, helped to shape views of teacher preparation not as training but rather as a model for teacher learning, building on an approach based on practitioner inquiry. One important implication of a practitioner-inquiry approach is in the way that teaching itself becomes more deeply examined when the basis of research questions and methodological approaches are led by those involved in doing the teaching, and that is further exemplified when the practitioners are teacher educators.

Teaching as a Discipline: The Problematic Nature of Practice

Tillema and Kremer-Hayon (2005) drew attention to the dilemmas faced by teacher educators as they enact a pedagogy of teacher education. Their dilemmas (partnership with colleagues vs. individualism, conservatism vs. constructivism, guidance vs. mentorship, etc.) are not dissimilar to Berry’s (2007) exquisite articulation of tensions in teaching about teaching (the tension between: (1) telling and growth; (2) confidence and uncertainty; (3) action and intent; (4) safety and challenge; (5) planning and being responsive; and (6) valuing and reconstructing experience). These researchers’ knowledge outcomes (dilemmas and tensions) highlight ways in which a pedagogy of teacher education requires an ongoing responsiveness to the problematic nature of teaching about teaching.

By highlighting teaching as being problematic it makes clear the need for teacher educators to articulate and act upon a knowledge of practice which in itself becomes an *entrée* to new ways of seeing into the teaching and learning of both students of teaching and teacher educators. Holzer (2002) illustrated the difficulties created for learners when teaching is not viewed as being problematic and hinted at developing an understanding of teaching as a field of study (or discipline) in its own right. Therefore, if teaching is a discipline, an articulation of a pedagogy of teacher education is crucial for teaching about teaching to be more than the delivery of teaching information and to be embodied in the very act of teaching the practice of teaching:

the question to be raised concerns the pedagogy of learning to teach. . . . it is a significant and different entry point in conceptualizing teacher education. . . . central to this type of learning are the ways the teachers will be engaged in the work of learning to teach . . . teaching now becomes the “subject matter” of teacher education. (Holzer, 2002: 393–395)

A pedagogy of teacher education then requires teacher educators to be involved in investigating teaching and learning in ways that will lead them to learn from practice while being involved in practice. The issues, concerns, and dilemmas of practice become a major focus in explicating the problematic nature of teaching and therefore help to direct teaching of teaching.

Teaching about Teaching

Central to a pedagogy of teacher education is the point that the teaching of teaching is not about the simple delivery of information and ideas of teaching. Rather, it is the teaching about knowledge of practice through the careful and purposeful creation of episodes in which that knowledge is borne out in experiences of practice. Creating meaningful experiences of learning to teach for students of teaching requires the knowledge of teacher educators' practice to be made explicit so that that which is so often tacit might be made accessible, and therefore able to be examined in meaningful ways. To do so requires not only making the tacit explicit, but also developing and using a shared language of practice.

Making the Tacit Explicit

Polanyi (1966) highlighted the tacit nature of teaching and helped to make clear some of the difficulties associated with articulating a knowledge of practice. The ability to teach and to explain why particular approaches to teaching work, do not necessarily go hand in hand. Making the tacit explicit requires sophisticated skills and teacher educators need to be experts in such practice.

Inviting students of teaching to look into the tacit aspects of teaching includes: (1) thinking aloud in order to articulate problems, issues, and concerns of practice in practice; (2) journal writing by teacher educators for/to student teachers in order to share insights into a teacher educator's thinking about teaching episodes and events; and (3) openly challenging one's existing practices whereby taken-for-granted approaches to teaching and behaviors of teaching (of both teacher educators and students of teaching) are confronted, deconstructed, and reconstructed in order to make learning about teaching real and meaningful within a teaching and learning experience.

A Language of Practice

A language of practice is essential for creating new ways of considering and sharing knowledge of teaching and learning so that both, the generalizable and the specific,

are able to be captured, articulated, and portrayed. Many different ways of articulating a language of practice can be found in the literature including: (1) axioms; (2) assertions; (3) dilemmas; (4) paradoxes; (5) principles of practice; (6) program principles; and (7) summary statements.

In these examples of ways of articulating a language of practice, there are two forms of knowledge that are being drawn upon. First is a generalized form of knowledge more in line with the traditional research-knowledge outcomes (i.e., applicable across cases and derived from scientific approaches to research and knowledge claims). Second is knowledge derived from learning about practice through reflection on practice and is somewhat idiosyncratic and case based (i.e., developed by considering something which has been learnt through experience over time).

Understanding what it means to enact a pedagogy of teacher education depends, in some part, on acknowledging that one form of knowledge does not exist without the other. A knowledge of practice, somewhat paradoxically, needs to be conceptualized as being specific and context dependent as well as being abstract and generalizable. In that way, a pedagogy of teacher education is about making the knowledge of practice applicable, meaningful, and useful for teaching in teaching as well as being identifiable or applicable across teaching contexts. Therefore, how students of teaching are able to apprehend and respond to such views of teaching becomes crucial in translating teaching about teaching into meaningful learning about teaching.

Learning about Teaching

Being a student of teaching (Bullough and Gitlin, 2001) requires balancing dual agendas for learning. There is a need to learn about the subject matter of teacher education (i.e., knowledge of such things as questioning, classroom management, mixed ability teaching, etc.) as well as the way in which that subject matter is taught (i.e., the teaching that is employed to examine that subject matter). These two aspects were described by Russell (1997) as the content turn and the pedagogical turn and highlight that it is not only what is taught, but also how it is taught, that influences learning about teaching.

As noted earlier, because pedagogy is about the dynamic relationship between teaching and learning it stands to reason that a pedagogy of teacher education must therefore be overtly responsive to both teaching and learning about teaching. Hence, teacher educators need to be cognizant of the dual learning agenda of students of teaching as they construct and enact their pedagogy of teacher education. This dual learning agenda is also impacted by the way in which students of teaching experience the shift from student to teacher.

From Student to Teacher

Students of teaching enter teacher preparation programs with well-formed images of what they think teaching is and how it is performed. Their apprenticeship of observation (Lortie, 1975) can have a profound effect on what they expect of, and think they should do, in a teacher preparation program (i.e., commonly to be able to be told how to teach).

The transition from student to teacher highlights a variety of feelings, issues, and concerns for students of teaching that can further complicate their learning about teaching. Through a pedagogy of teacher education, this transition should be a time in which “the student perspective is valued and appropriately responded to as new understandings of the world of teachers and teaching expand and develop” (Loughran, 2006: 124).

Fuller highlighted how student teachers’ concerns shift over time (Fuller, 1969; Fuller and Bown, 1975). Although some have argued against such a developmental model, it is nonetheless important that both teacher educators and students of teaching are aware of, and sensitive to, the changes in concerns in learning about teaching and how they impact practice at different times during teacher preparation. This need to recognize shifts in concerns is one reason why a pedagogy of teacher education aligns so closely with notions of developing reflective practitioners through teacher-preparation programs.

Developing Reflective Practitioners

There has been a long tradition of developing student teachers as reflective practitioners (Grimmett and Erickson, 1988). Central to reflection is the need to be able to see a problem (e.g., curious/puzzling/engaging situation), and to do so from different perspectives. In learning about teaching, students of teaching need to learn how to recognize their own problems of practice in order to begin to direct their own professional learning of teaching. Therefore, a pedagogy of teacher education must be conceptualized and constructed in ways that helps students of teaching begin to see more deeply into the nature of practice so that they are able to recognize and respond to the problems of practice in positive ways.

Learning about teaching, like teaching about teaching, requires an inquiry stance in order for individuals to search for, and respond to, their own problems of practice so that their development of knowledge of, and actions in, teaching might develop and be extended in appropriate ways while still being responsive to their needs and concerns. One way of encouraging this to happen is through a student-teacher-as-researcher approach.

Seeing into Teaching through a Student-Teacher-as-Researcher Stance

There is a paucity of research in the student-teacher-as-researcher approach. Fundamentally, what a student-teacher-as-researcher approach requires, and is inherent in a pedagogy of teacher education, is the serious creation by teacher educators of possibilities for student teachers to “have opportunities to confront the abstract and unbounded complexity of teaching” (Cochran-Smith, 1991: 113) that is, to come to be challenged by the problematic nature of practice and to inquire into it.

By embedding learning about teaching in experience and by inquiring into those experiences through a student-teacher-as-researcher stance, more meaningful ways of learning about teaching are encouraged. Embedding learning about teaching in student teachers’ own experiences is important for supporting their attempts to build deeper understandings of knowledge of practice. A pedagogy of teacher education is enacted by supporting student teachers’ learning about teaching by going beyond the technical and inviting them to seek to become truly professional. That is catalyzed by inviting students of teaching to adopt a student-teacher-as-researcher stance in their learning about teaching.

Conclusion

A pedagogy of teacher education comprises principles, expectations, and practices that matter in shaping teaching and learning about teaching. Enacting a pedagogy of teacher education is about teacher educators better aligning their intentions and their practices so that the complex nature of teaching and learning can genuinely be unpacked and critiqued in teaching and learning about teaching. Teacher educators not only need to be able to teach, but they also need to be able to teach teaching, which means their practice should hold up for scrutiny the problematic nature of practice within the experiences of teaching and learning they create for their students of teaching.

A pedagogy of teacher education is, at one level, about the individual practice of teacher educators and how, through inquiry, that specialist knowledge of practice might be captured, articulated, shared, and critiqued. However, at another level, there is also a pressing need for such knowledge to impact teacher-education practices more generally

“first [at] the institutional level whereby . . . program organization and structure reflect the way in which a pedagogy of teacher education is inherently intended to

shape teacher education as a whole. . . . [and also within] the teacher education community [whereby] ideas, issues, concerns and conceptualizations might be developed, debated, articulated and portrayed in ways that will progress the field of teacher education”

(Loughran, 2006: 176).

Bibliography

- Ball, D. L. and Wilson, S. M. (1990). Knowing the subject and learning to teach it: Examining assumptions about becoming a mathematics teacher. *April 1990 at the Annual Meeting of the American Educational Research Association in Boston*. Retrieved 7 November 2006. <http://ncrtl.msu.edu/http/rreports/html/pdf/rr907.pdf> (accessed December 2008).
- Berliner, D. C. (1988). Implications of studies on expertise in pedagogy for teacher education and evaluation. In Pfeleiderer, J. (ed.) *New Directions for Teacher Assessment. Proceedings of the 1988 ETS Invitational Conference*, pp 39–64. Princeton, NJ: Educational Testing Service.
- Berry, A. (2007). *Tensions in Teaching about Teaching: A Self-Study of the Development of Myself as a Teacher Educator*. Dordrecht: Springer.
- Bullough, R. V. J. and Gitlin, A. (2001). *Becoming a Student of Teaching: Linking Knowledge Production and Practice*, 2nd edn. London: Routledge-Falmer.
- Cochran-Smith, M. (1991). Reinventing student teaching. *Journal of Teacher Education* **42**(2), 104–118.
- Cochran-Smith, M. (2000). The future of teacher education: Framing the questions that matter. *Teaching Education* **11**(1), 13–24.
- Doyle, W. and Carter, K. (2003). Narrative and learning to teach: Implications for teacher-education curriculum. *Journal of Curriculum Studies* **35**(2), 129–137.
- Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *The Teachers College Record* **103**(6), 1013–1056.
- Fuller, F. F. (1969). Concerns of teachers: A developmental conceptualization. *American Educational Research Journal* **6**(2), 207–226.
- Fuller, F. F. and Bown, O. H. (1975). Becoming a teacher. In Ryan, K. (ed.) *Teacher Education: The 74th Yearbook of the National Society for the Study of Education, Part 11*, pp 25–52. Chicago: University of Chicago Press.
- Grimmett, P. P. and Erickson, G. (1988). *Reflection in Teacher Education*. NY: Teachers College Press.
- Hamilton, M. L. and Pinnegar, S. (2000). On the threshold of a new century: Trustworthiness, integrity, and self-study in teacher education. *Journal of Teacher Education* **51**(3), 234–240.
- Heaton, R. M. and Lampert, M. (1993). Learning to hear voices: Inventing a new pedagogy of teacher education. In Cohen, D. K., McLaughlin, M. W., and Talbert, J. E. (eds.) *Teaching for Understanding: Challenges for Policy and Practice*, pp 43–83. San Francisco, CA: Jossey-Bass.
- Holzer, E. (2002). Conceptions of the study of Jewish texts in teachers’ professional development. *Religious Education* **97**(4), 377–403.
- Korthagen, F. A. J., Kessels, J., Koster, B., Langerwarf, B., and Wubbels, T. (2001). *Linking Practice and Theory: The Pedagogy of Realistic Teacher Education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Liston, D. P. and Zeichner, K. M. (1988). Critical pedagogy and teacher education. *Paper Presented at the American Educational Research Association*. New Orleans: LA, USA, (April).
- Lortie, D. C. (1975). *Schoolteacher*. Chicago, IL: Chicago University Press.
- Loughran, J. J. (2006). *Developing a Pedagogy of Teacher Education: Understanding Teaching and Learning about Teaching*. London: Routledge.
- Polanyi, M. (1966). *The Tacit Dimension*. Garden City, NY: Doubleday.
- Russell, T. (1997). Teaching teachers: How I teach IS the message. In Loughran, J. and Russell, T. (eds.) *Teaching about Teaching: Purpose, Passion and Pedagogy in Teacher Education*, pp 32–47. London: Falmer Press.
- Tigchelaar, A. and Korthagen, F. (2004). Deepening the exchange of student teaching experiences: Implications for the pedagogy of teacher education of recent insights into teacher behavior. *Teaching and Teacher Education: An International Journal of Research and Studies* **20**(7), 665–679.
- Tillema, H. and Kremer-Hayon, L. (2005). Facing dilemmas: Teacher-educators’ ways of constructing a pedagogy of teacher education. *Teaching in Higher Education* **10**(2), 203–217.
- Wilson, S. M. (1990). The secret garden of teacher education. *Phi Delta Kappan* **72**(Nov), 204–209.

Further Reading

- Berry, A. and Loughran, J. J. (2002). Developing an understanding of learning to teach in teacher education. In Loughran, J. and Russell, T. (eds.) *Improving Teacher Education Practices Through Self-Study*, pp 13–29. London: Routledge Falmer.
- Dinkelmann, T., Margolis, J., and Sikkenga, K. (2001). From teacher to teacher educator: Experiences, expectations, and expatriation. *Paper Presented at the American Educational Research Association*. Seattle, WA, USA, (April).
- Guilfoyle, K., Hamilton, M. L., Pinnegar, S., and Placier, M. (1995). Becoming teachers of teachers: The paths of four beginners. In Russell, T. and Korthagen, F. (eds.) *Teachers Who Teach Teachers: Reflections on Teacher Education*, pp 35–55. London: Falmer Press.
- Hamilton, M. L. (2004). Professional knowledge, teacher education, and self-study. In Loughran, J. J., Hamilton, M. L., LaBoskey, V. K., and Russell, T. (eds.) *International Handbook of Self-Study of Teaching and Teacher Education Practices*, vol. 1, pp 375–419. Dordrecht: Kluwer.
- Hoban, G. F. (ed.) (2005). *The Missing Links in Teacher Education Design: Developing a Multi-Linked Conceptual Framework*. Berlin: Springer.
- Loughran, J. J. (2004). Student teacher as researcher: Accepting greater responsibility for learning about teaching. *Australian Journal of Education* **48**(2), 212–220.
- Pereira, P. (2000). Reconstructing oneself as a learner of mathematics. In Loughran, J. and Russell, T. (eds.) *Exploring Myths and Legends of Teacher Education. Proceedings of the Third International Conference of the Self-Study of Teacher Education Practices*. Herstmonceux Castle, East Sussex, England.
- Pinnegar, S. (1995). (Re)Experiencing student teaching. In Russell, T. and Korthagen, F. (eds.) *Teachers Who Teach Teachers: Reflections on Teacher Education*, pp 56–67. London: Falmer Press.
- Russell, T. and Loughran, J. (eds.) (2007). *Enacting a Pedagogy of Teacher Education*. London: Routledge.
- Trumbull, D. (2004). Factors important for the scholarship of self-study of teaching and teacher education practices. In Loughran, J. J., Hamilton, M. L., LaBoskey, V. K., and Russell, T. (eds.) *International Handbook of Self-Study of Teaching and Teacher Education Practices*, vol. 2, pp 1211–1230. Dordrecht: Kluwer.

Case Methods in Teacher Education

M Sato and C Rogers, University of Minnesota, Twin Cities, MN, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Since the 1980s, case methods have gained traction in teacher education. Typically, cases are narratives drawn from real-life instances and used as teaching tools to catalyze group discussion and individual reflection. Published case books exist on a variety of topics (e.g., teaching in diverse contexts, classroom assessment practices, teaching particular subject matter concepts, and using group work) and some teacher education programs engage their candidates in the writing of cases, which are then selectively compiled into in-house casebooks for future classes to use. Case discussion as a pedagogical method is currently used in teacher education programs to support the development of beginning teachers as well as the mentoring and coaching practices of cooperating and mentor teachers (Shulman and Sato, 2006). Case writing by preservice teachers has also emerged as a pedagogy that has the potential of supporting the development of beginning teachers' understanding of instructional practices, professional development during their clinical experiences, the application of theoretical understandings to practice, and the development of their moral agency. Similarly, Loughran (2006) has argued for the use of anecdotes of difficult or unresolved situations constructed from the experiences of preservice teachers as a means of providing focus for their reflections on practice.

The construction and definition of what constitutes a case in teacher education varies widely. Cases can be constructed as exemplars of practice, as dilemmas faced in practice that can only be resolved through compromise, as historical documents that represent issues from the past that linger in the present, or as representations of everyday practice. In recent years, cases using multimedia representations of practice have become more common. Multimedia formats can provide classroom artifacts such as student work samples, classroom video, teacher narratives and reflections, and other classroom documents. These collections of artifacts are usually centered on a particular unit of instruction, instructional project, or issue of student learning rather than the traditional narrative construction of a written case. Akin to having preservice teachers write their own cases, one group of researchers has demonstrated the benefits of developing preservice teachers' abilities to identify, interpret, and analyze evidence of exemplary teaching by having them construct video cases of their mentor teachers during their student teaching experience (Beck *et al.*, 2002).

In her review of the literature on case methods in teacher education, Merseth (1996) suggested three purposes for the use of case methods. The first purpose describes cases as exemplars or models of best practices that depict theoretical or desired principles in action. The second purpose is to provide opportunities for analysis of a complex situation and to practice decision making vicariously. She states that "in this conception, cases are not used explicitly to exemplify theory but rather to present situations from which theory emerges" (p. 728). The third purpose of case methods is to stimulate reflection and potentially help the preservice teacher develop the habits and skills of reflective practice.

Merseth based this conceptual organization partially on the work of Lee Shulman (1986), who is widely recognized as catalyzing the emergence of case methods in teacher education. In his 1985 American Educational Research Association presidential address, Shulman provided a framework for the knowledge base for teaching and proposed three forms of teacher knowledge – propositional knowledge, case knowledge, and strategic knowledge. Case knowledge was described as "knowledge of specific, well-documented, and richly described events . . . [in which] the knowledge they represent is what makes them cases" (p. 11). Shulman speaks to both the substance of the cases and the pedagogical function that a case can serve by describing three types of cases:

Prototypes exemplify theoretical principles. *Precedents* capture and communicate principles of practice or maxims. *Parables* convey norms or values. Naturally, a given case can accomplish more than a single function. (p. 11, emphasis in original)

McAninch (1995) took a different approach in providing a conceptual framework for organizing the case methods literature in teacher education. She drew on Feiman-Nemser's (1990) five-part classification scheme for the orientation of teacher education programs: (1) academic, (2) practical, (3) technological, (4) personal, and (5) critical/social. In this analysis, McAninch attributed particular case books and some research efforts around case methods to one of these orientations. Using this programmatic or ideological approach to describe the purposes of cases and case methods seems problematic insofar as it assumes that case methods would be used by the instructors within a teacher education program the same way. It also does not account for the potential of using a

particular case toward different or multiple purposes as suggested by Shulman.

The rest of this article addresses both the empirical and theoretical bases for case methods in teacher education. We highlight some of the empirical research from the past decade to illustrate the kinds of research questions being pursued and the theoretical lenses that researchers use in framing their studies and analyzing their data. We point the reader to other compilations of the literature for summaries of earlier research (Merseeth, 1996; Sykes and Bird, 1992).

Empirical Base for Case Methods in Teacher Education

While many proponents of case methods have suggested the benefits of using this pedagogy, the call for more empirical research on case methods has been echoing since the early 1990s (Lundeberg, *et al.*, 1999; McAninch, 1995; Merseeth, 1996; Sykes and Bird, 1992).

Case methods research over the past decade has continued to focus on many of the issues that Merseeth identified in 1996: what preservice teachers learn from engaging in case methods, the influence of cases on how teachers think, and the influence of instructional contextual factors on how teachers learn through cases. As cases are increasingly constructed using multimedia formats, the contextual factors associated with learning from cases now account for differing electronic contexts and processes such as synchronous versus asynchronous participation in case discussions, supporting successful online dialog during case analysis, and differences between text-based and video formats of cases.

Lundeberg *et al.* (1999) compiled a research base for teaching and learning with cases. In their edited volume, they explore the fundamental questions raised in the title of the book: *Who Learns What from Cases and How?* They conclude that comparing case methods to other forms of pedagogy to determine comparative effectiveness would not be a fruitful direction for future research given the multiple issues related to isolating variables and conditions that might contribute to learning outcomes. Rather, they suggest that research should focus on what is it that students actually learn from case methods. For example, pedagogues and researchers often purport that case methods can help develop a teacher's reasoning ability, problem-framing skills, reflective mindsets, subject-matter knowledge, or pedagogical repertoires. Lundeberg *et al.* argue that teacher education as a field has not developed a shared meaning of these concepts and that the constructs are sometimes ill-defined in individual studies.

Researchers use a variety of data collection and analytical methods to determine effects of case methods on

preservice teachers' learning and development: analyzing transcripts of case discussions; interviewing and surveying participants about perceptions of learning and change; examining changes in coursework such as journals and writing samples; and asking participants to review their own writing and reflect on changes in their thinking and learning. Fewer studies have used quasi-experimental designs using pre/post measures of treatment and control groups.

In the past decade, the research on the pedagogical use of cases has examined how to structure case discussions, the role of the facilitator in case discussions, how expectations for case writing are communicated to students, and how case writing is scaffolded. Case books written for instructional purposes have also attended to pedagogical approaches in using cases by including discussion and reflection questions with cases or guidelines for facilitation and analysis of case discussions. Because of the episodic character of cases, some have argued that the pedagogical approach used with cases must make explicit efforts to help students use the particularities of the case to examine broader theoretical issues and principles for learning and teaching that the cases may represent (Shulman, 1992). Determining the quality of the case in relationship to the purpose for using the case is also essential to understanding the effectiveness of case pedagogies.

The long-term effect of learning from case methods and the transfer of what is learned to practice continues to be rare in the published literature. One study followed up with 31 teachers through mailed surveys to determine the impact of multimedia case-based instruction on their practice (Maloch and Kinzer, 2006). Telephone interviews with eight participants were also conducted to explore further the respondents' perceptions of the value of instructional methods courses, differences between traditional methods class procedures and case methods, and retention of case content over time. The authors report that the respondents found the use of multimedia cases to be influential on their learning, especially in helping them remember particular strategies they could use in their teaching and in providing high instructional engagement. Somewhat troubling, however, was the dominant perception among respondents that the cases used in their courses were meant to represent exemplars or models of practice. The instructors' intentions were that the cases represent the complexity of practice and the need for teachers to make contextually based decisions. The view of cases as models of practice seems to have limited the novice teachers' views about the applicability of the cases to their own teaching practice. Rather than seeing the substance of the cases as resting on underlying theoretical principles or as a tool to stimulate reflection, several respondents commented on not seeing direct applicability of the cases to their current teaching circumstances.

The appeal of case methods in teacher education has drawn international attention in the research literature. At the University of Twente in the Netherlands, Van Den Berg *et al.* (2004) examined whether exposure to a multimedia case of a teacher's exemplary science unit would transfer to student teachers' practice during their student teaching experience. After engaging with the online case, the novice teachers were asked to create new curricular experiences. The majority of students mirrored the curriculum presented in the case, following it like a script, while a few students used the case as a springboard for creating novel curriculum both within the subject-matter discipline represented in the case and to other subject-matter areas. Additionally, institutions in the Philippines and Singapore have been developing research around case-based pedagogies as a basis for reform in the teaching and learning of elementary school science and the adoption of constructivist instructional approaches for preservice teachers (Arellano *et al.*, 2001).

Theoretical Base for Case Methods in Teacher Education

While we agree with the need for more systematic development and continued research on case methods, we also emphasize that this research needs to be theoretically grounded in ways that allow the work to be useful within and across the diverse theoretical traditions of the field. Published research demonstrates several different perspectives on learning and teacher development. The degree to which research studies are positioned within a theoretical framework, however, varies from studies lacking identifiable theoretical perspectives to studies that use theories to guide their research questions, research designs, use of case methods, and analysis. In this section, we briefly explore the dominant theoretical perspectives currently used in research on case methods and describe a few studies that demonstrate strong theoretical grounding in pursuing research questions about case methods in teacher education.

Situated Cognition Theory

Situative theorists view learning as not only a process that occurs inside the mind of the individual, but also as being bound up in the physical and social contexts of activity as well as the activity itself (Brown *et al.*, 1989; Lave and Wenger, 1991). The situation in which one learns, including the interactions with others, the environment, and the physical and symbolic tools that one uses are all fundamental parts of the learning that takes place. The knowledge itself is distributed throughout the situation or the system and the learning is emergent and social. When

designing learning activities from the perspective of situated cognition theory, educators immerse learners in experiences and environments that are as-authentic-as-possible contexts in which their new ideas and behaviors will be applied. Putnam and Borko (2000) propose several approaches for engaging teachers in activities that map onto the tenets of situated learning: grounding teachers' learning in their own practice, using artifacts of practice in professional development activities, and using case-based learning experiences. They suggest that "rather than putting teachers *in* particular classroom settings, cases provide vicarious encounters with those settings" (p. 8, emphasis in original).

Using situated learning theory in their analysis, Yoon *et al.* (2006) investigated how case methods particularly influenced novice elementary school teachers' self-efficacy about teaching science. The researchers collected pre-case discussion questionnaires, transcripts of case discussion activities, and post-case online discussion board postings of 12 elementary school teaching candidates at the University of Toronto. The authors designed the case discussions to allow for multiple points of entry for participants with varying background knowledge and experiences to participate in the science-teacher community discourse. The structure of the case activities allowed the novice teachers to experience legitimate peripheral participation (Lave and Wenger, 1991) within the science teaching community by engaging in discussions about real classroom experiences. As the novice teachers gained self-confidence through their participation in the activities embedded in the case methods, they moved closer to the center of the community activity and discourse. The authors argue that the case served as a tool for the participants to make sense of a real classroom situation, test their knowledge, and to begin to align their discourse with the larger science-teaching community.

Through participation in the case activity and engagement with each other, we suggest that students were crossing boundaries by applying terminology and pedagogical activities learned in previous workshops, cross-referencing theoretical constructs from class readings, and relating personal experiences to the case scenario. The case itself represented an important boundary object around which meaning was constructed and shared, while the case method provided the necessary brokering mechanism (Yoon *et al.*, 2006, p. 25).

Cognitive Flexibility Theory

Among the approaches to learning that draw on cognitive theories, cognitive flexibility theory recognizes that some domains of knowledge are complex and ill-structured. In contrast to well-structured domains that may be organized in linear or hierarchical formats that translate into sequential instruction, ill-structured domains are

viewed as irregular and contextual (Spiro *et al.*, 1987). The theory suggests that instructional design for such domains should represent the complexity of the domain by creating webs of information (e.g., through the use of hyper-text), using multiple perspectives, and embedding the knowledge within multiple contexts. Students of ill-structured knowledge domains should engage in critical analysis, perspective taking, and problem framing to support the transfer of knowledge to new contexts. Thus, cognitive flexibility within such complex domains:

...includes the ability to represent knowledge from different conceptual and case perspectives and then, when the knowledge must later be used, the ability to construct from those different conceptual and case representations a knowledge ensemble tailored to the needs of the understanding or problem-solving situation at hand (Spiro *et al.*, 1991, 24).

While some aspects of learning to teach may reside in learning factual information and theoretical principles, many aspects of teaching can be considered a complex and uncertain endeavor. Many illustrations of teaching practice demonstrate the complexity of teachers' knowledge application, decision making, immediacy of response, and the overall management of dilemmas in daily practice. Based on this view of teaching, knowledge for teaching requires a flexibility of application and practiced decision making that would allow beginning teachers to respond flexibly to messy and context-dependent variation.

In a study from the University of Haifa in Israel, Eilam and Poyas (2006) used cognitive flexibility theory to study the efficacy of engaging with multimedia cases. They examined 21 responses of trainees in a university teacher education program using pre/post tests that emphasized implicit cognitive processes as central to teaching before and after a one-semester intervention. To address the development of cognitive flexibility in responding to messy classroom situations, a new teacher-training course was designed for future junior high and high school literature teachers and was offered via the Internet in the last semester of the teacher preparation program. The course comprised five video cases taken from real literature classrooms with accompanying written contextual information about the students, the teacher, the school, and the curriculum, as well as the materials used during the lessons. The participants also had access to the university library database, interview excerpts with the teacher on the video, external experts, and a web forum to engage in dialogue. In analyzing the web-based materials, the novice teachers were asked to describe the wide variety of contextual factors in complex classroom scenarios, interpret interrelations and interactions among those factors, and evaluate the possible effects on teaching and learning in the given situation. Based on content analysis of written responses to particular tasks prior to and after completion

of the course, the authors concluded that the novice teachers demonstrated an increased ability to survey and grasp the complexity of the whole classroom system including the implicit teaching-learning processes.

Social Cognitive Theory

One argument for the use of case methods suggests that cases can be used as a means of presenting exemplars of practice, allowing novices to see complex practices in action, thus providing a model of performance. This argument draws most strongly on social cognitive theory (Bandura, 1986) that posits that behaviors are learned by watching others' modeled behaviors. According to this theory, rather than mimicking the exemplar, the learner abstracts the structure and the underlying principles governing the behavior and then uses those principles to generate new versions of the behavior. It has been shown that the learner can produce innovative behaviors from observing multiple models and combining features from different models. The theory also suggests that cognitive skills can be verbally modeled through think-alouds that communicate reasoning while engaged in cognitive activities such as problem solving.

The Best Practices project at Arizona State University attempts to provide a collection of authentic technology-based teaching models by creating cases of real teachers who have been identified as exemplary by peers and university faculty and staff (Kurz *et al.*, 2005). The developers caution that the term "best practices" is not meant to imply that the teachers included in the database are the best, but the premise of their project rests on providing models of teaching and reflective practice behaviors to be used in teacher preparation.

Reasoning and Reflective Practice

In many ways, conceptions of teacher development grounded in theories of reasoning and reflective practice are distinct from one another; yet, their close association with the practical nature of teachers' work binds them together in this discussion and in much work associated with case methods. Reasoning can be viewed as a rational process of planned action. Practical activities, such as teaching, often demand on-the-spot action without the luxury of forethought. The Aristotelian concept of *phronesis*, often translated as practical reasoning or practical wisdom (Dunne, 1993), incorporates not only the action derived from thought but also suggests that one may come to know what one thinks through taking action. The knowledge of the practitioner is bound up in the action and can be further articulated or made explicit through reflection on that action. Schön (1987) similarly described knowledge-in-action as spontaneous and subtle, a taken-for-granted kind of knowing. Through reflection-in-action, one can

make adjustments and even perform on-the-spot experiments or engage in a “pattern of inquiry” (p. 27) that allows one to adapt action to the particulars of the situation. Reflection-on-practice allows for the postanalysis of the situation and a review of one’s performance, with the potential of evaluating the effectiveness and process of one’s choices.

Using a framework centered on reasoning to examine the potential outcomes of case methods is less concerned with knowledge structures and cognitive processes of transfer and places more emphasis on addressing the complexities of practice through action, including the moral means and ends of one’s choices. Harrington (1995) explored the development of preservice teachers’ reasoning by examining student’s written analyses of dilemma-based cases. She analyzed the responses according to the following framework for making reasoned decisions: problem identification, awareness of alternative perspectives, warranting of solution, consideration of consequences, and reflectiveness. She argued that case methods have the potential for preparing teachers to “make reasoned decisions—decisions incorporating and balancing theoretical, normative, practical, and consequential reasoning—foster[ing] the development of critically reflective practitioners” (p. 212).

Much has been written about the use of case methods to promote reflective practice and increase the reflective capacity of novice teachers. Many, however, have criticized the overused and underdefined concept of reflection in teacher education practice and research. Harrington *et al.* (1996) argued that “preparing teachers who are reflective requires a clear conception of what reflection is and the means to foster its development” (p. 25). In their study of 21 undergraduate students studying to be elementary school teachers, they analyzed a set of case analyses completed by the students to determine if students’ case analyses could be used to understand particular aspects of reflection exhibited by the preservice teachers. Drawing on Dewey’s (1933) definition of reflection, the authors examined the nature of open-mindedness (the ability to understand and take multiple perspectives), wholeheartedness (the ability to identify and address the limitations in one’s assumptions about authority and understand the complexity and ambiguity of issues), and responsibility (the ability to consider the moral and ethical consequences of choices and actions on self, others, and the broader society). The authors concluded that the students’ case analyses demonstrated evidence of these multiple dimensions of reflection.

Bridging Theory and Practice

Using case methods in teacher preparation is time intensive and may place limits on curriculum coverage in

teacher preparation programs. Yet, case methods have been lauded as a means by which the often-criticized gap between theory and practice can be bridged in the education of teachers. Traditional teacher preparation models often maintain distinctions and separations among theoretical knowledge emphasized in college and university classrooms, observation of expert teachers in classrooms, and the first-hand practice of the novices during clinical placements. The use of case methods allows prospective teachers to analyze an authentic situation critically, generate multiple theoretical and practical interpretations of it, and formulate plans of action for a particular situation.

In her reflections on 20 years of working with teachers developing and using case methods, Judith Shulman (2003) wrote about a “vision of teaching as a deeply professional and profoundly complex form of theoretically enriched *practice*. The practice may be informed by theory, but it can ultimately only be learned and understood through the study of practice” (p. 2). Shulman puts forth a theory of teacher learning grounded in the study of practice that is exemplified through activities such as case writing, case analysis, and portfolio analysis of practice. Thus, case methods have the potential to not only situate the learning of the teacher in a particular context but also to provide opportunities to theorize from practice while practicing theory.

See also: Professional Development of Teacher Educators; Teacher Education and Models of Teacher Reflection; Teacher Learning with Lesson Study; The Relationship Between Theory and Practice in Teacher Education.

Bibliography

- Arellano, E., Barcenal, T., Bilbao, P., *et al.* (2001). Using case-based pedagogy in the Philippines: A narrative inquiry. *Research in Science Education* **31**, 211–226.
- Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Beck, R. J., King, A., and Marshall, S. K. (2002). Effects of videocase construction on preservice teachers’ observations of teaching. *Journal of Experimental Education* **70**, 345–361.
- Brown, J. S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* **18**, 32–42.
- Dewey, J. (1933). *How We Think*. Boston: D.C. Heath & Company.
- Dunne, J. (1993). *Back to the Rough Ground: Practical Judgment and the Lure of Technique*. Notre Dame, IN: University of Notre Dame Press.
- Eilam, B. and Poyas, Y. (2006). Promoting awareness of the characteristics of classrooms’ complexity: A course curriculum in teacher education. *Teaching and Teacher Education* **22**, 337–351.
- Feiman-Nemser, S. (1990). Teacher preparation: Structural and conceptual alternatives. In Houston, W. R., Haberman, M., and Sikula, J. (eds.) *Handbook of Research on Teacher Education*, 1st edn, pp 195–211. New York: Macmillan.

- Harrington, H. L. (1995). Fostering reasoned decisions: Case-based pedagogy and the professional development of teachers. *Teaching and Teacher Education* **11**, 203–214.
- Harrington, H. L., Quinn-Leering, K., and Hodson, L. (1996). Written case analysis and critical reflection. *Teaching and Teacher Education* **12**, 25–37.
- Kurz, T. L., Llama, G., and Savenye, W. (2005). Issues and challenges of creating video cases to be used with preservice teachers. *TechTrends* **49**, 67–73.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Loughran, J. (2006). *Developing a Pedagogy of Teacher Education: Understanding Teaching and Learning about Teaching*. New York: Routledge.
- Lundeberg, M. A., Levin, B. B., and Harrington, H. L. (1999). *Who Learns What from Cases and How? The Research Base for Teaching with Cases*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Maloch, B. and Kinzer, C. (2006). The impact of multimedia cases on preservice teachers' learning about literacy teaching: A follow-up study. *The Teacher Educator* **41**, 158–171.
- McAninch, A. R. (1995). Case methods in teacher education. In Husen, T. and Postlethwaite, T. N. (eds.) *The International Encyclopedia of Education*, 2nd edn, pp 5943–5948. Oxford, UK; New York: Pergamon.
- Merseeth, K. K. (1996). Cases and case methods in teacher education. In Sikula, J. (ed.) *Handbook of Research on Teacher Education*, 2nd edn, pp 722–741. New York: Macmillan and the Association of Teacher Educators.
- Putnam, R. T. and Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher* **29**, 4–15.
- Schön, D. A. (1987). *Educating the Reflective Practitioner*. San Francisco: Jossey-Bass Publishers.
- Shulman, J. H. (ed.) (1992). *Case Methods in Teacher Education*. New York: Teachers College Press.
- Shulman, J. H. (2003). *From Practice to Theory and Back Again: Cases and Portfolios as Instruments for Professional Development*. San Francisco: WestEd.
- Shulman, J. and Sato, M. (eds.) (2006). *Mentoring Teachers toward Excellence: Supporting and Developing Highly Qualified Teachers*. San Francisco: Jossey-Bass Publications.
- Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher* **15**, 4–14.
- Spiro, R. J., Feltovich, P. J., Jacobson, M. J., and Coulson, R. L. (1991). Knowledge representation, content specification, and the development of skill in situation-specific knowledge assembly: Some constructivist issues as they relate to cognitive flexibility theory and hypertext. *Educational Technology* **31**, 22–25.
- Spiro, R. J., Vispoel, W. P., Schmitz, J. G., Samarapungavan, A., and Boerger, A. E. (1987). Knowledge acquisition for application: Cognitive flexibility and transfer in complex content domains. In Britton, B. K. and Glynn, S. M. (eds.) *Executive Control Processes in Reading*, pp 177–199. Hillsdale, NJ: Lawrence Erlbaum.
- Sykes, G. and Bird, T. (1992). Teacher education and the case idea. *Review of Research in Education* **18**, 457–521.
- Van Den Berg, E., Jansen, L., and Blijleven, P. (2004). Learning with multimedia cases: An evaluation study. *Journal of Technology and Teacher Education* **12**, 491–509.
- Yoon, S., Pedretti, E., Pedretti, L., et al. (2006). Exploring the use of cases and case methods in influencing elementary preservice science teachers' self-efficacy beliefs. *Journal of Science Teacher Education* **17**, 15–35.

Further Reading

- Colbert, J., Desberg, P., and Trimble, K. (eds.) (1996). *The Case for Education: Contemporary Approaches for Using Case Methods*. Boston: Allyn & Bacon.
- Grossman, P. (2005). Research on pedagogical approaches in teacher education. In Cochran-Smith, M. and Zeichner, K. (eds.) *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education*, pp 425–476. Mahwah, NJ: Lawrence Erlbaum.
- Hammerness, K. M., Darling-Hammond, L., and Shulman, L. (2002). Toward more expert thinking: How curriculum case-writing prompts the development of theory based professional development knowledge in student-teachers. *Teaching Education* **13**, 219–243.
- Harrington, H. L. and Garrison, J. (1992). Cases as shared inquiry: A dialogical model of teacher preparation. *American Educational Research Journal* **29**, 715–735.
- Hatch, T. (2005). *Into the Classroom: Developing the Scholarship of Teaching and Learning*. Indianapolis, IN: Jossey-Bass.
- McAninch, A. R. (1993). *Teacher Thinking and the Case Method*. New York: Teachers College Press.
- Merseeth, K. K. (1991). The early history of case-based instruction: Insights for teacher education today. *Journal of Teacher Education* **42**, 243–249.
- Wassermann, S. (1994). *Introduction to Case Methods Teaching*. New York: Teachers College Press.

Relevant Websites

- <http://ctell.uconn.edu> – Case Technologies to Enhance Literacy Learning, Teachers College at Columbia University, University of Georgia, University of Connecticut, and University of Illinois at Chicago.
- <http://gallery.carnegiefoundation.org> – Inside teaching: A living archive of practice, Carnegie Foundation for the Advancement of Teaching.
- <http://www.wested.org> – Institute for Case Development, WestEd.
- <http://www.intime.uni.edu> – Integrating New Technologies into the Methods of Education (INTIME), University of Northern Iowa.
- <http://quest.carnegiefoundation.org> – Making teaching public: A digital exhibition, Carnegie Foundation for the Advancement of Teaching, Teachers College at Columbia University, and National Center for Restructuring Education, Schools, and Teaching.
- <http://www.eliteracy.org> – Reading Classroom Explorer, Michigan State University and the Center for the Improvement of Early Reading Achievement.

Conceptions of Teacher Education

J A Whitcomb, University of Colorado at Boulder, Boulder, CO, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Most individuals hold vivid memories of teachers who have touched their lives. Popular culture promotes stories of teachers as heroes who nurture the student's passion for learning. Large-scale studies using value-added modeling have demonstrated the teacher's impact on student achievement, as measured by standardized tests. The confluence of personal memories, popular culture images, and research findings places teaching at the center of educational reform efforts. Not surprisingly, interest in teaching draws attention to teacher preparation, a perennially contentious topic. For decades, teacher education has weathered criticism, for example, low entry standards, lack of intellectual rigor, overemphasis on theory, disconnect from the reality of schools, and programmatic incoherence. Both internal and external critics describe teacher education as a weak intervention.

Sustained critique reflects both the politicized nature of teacher education and distinct conceptualizations of good teaching and, by extension, teacher preparation. This article examines four conceptions of teacher education: (1) social justice, (2) teaching for understanding, (3) contemplative, and (4) clinician-professional. Each articulates a particular vision for the aims of teacher preparation that is grounded in an argument for educational purposes and features of good teaching. Each reflects an intellectual rationale that guides curricular goals and program design. The four approaches presented build upon and extend previous discussions of conceptions of teacher education; see Table 1 (Feiman-Nemser, 1990; Liston and Zeichner, 1991; Zeichner, 1987, 1993).

These conceptions are robust. Since the early 1990s, each approach has been elaborated and developed, with some approaches more than others, at the center of innovation and theory development in teacher education. No new conceptual approach has emerged to reframe conversations within the field about the aims of teacher education. This article emphasizes current understandings and instantiations.

Social Justice Approach

A social justice approach to teacher education begins with the idea that a central purpose of education is to redress social, economic, and political inequities. Its intellectual roots lie within a radical progressive tradition. In this

conception, a good teacher understands socioeconomic and political forces that maintain structural inequality and oppression, including how schools as institutions reinforce the *status quo* and further inequitable educational experiences. Given the centrality of race, ethnicity, and class in inequality, a social justice approach to teaching involves an ongoing commitment to grapple with these matters. A teacher for social justice enacts curriculum so that students develop both a critical social consciousness and the intellectual and practical tools to be agents of change. Students study the experiences of those who have been marginalized along with possibilities for liberation. The teacher also ensures that students learn skills and knowledge associated with the most powerful (cultural capital), thereby helping to promote access to all levels of society. Classrooms are democratic communities where the teacher helps construct an ethos of care and respect. Finally, a social justice teacher also embraces an identity as a community activist and sees this work as an extension of her teaching.

Social justice approaches to teacher education have been part of the teacher-education landscape since the 1930s. In the US, they emerged as part of a larger social movement to redress the ills of capitalism made evident in the Great Depression. Liston and Zeichner (1991) and Feiman-Nemser (1990) describe several well-regarded historical examples including the New College at Teachers College, Columbia University in the 1930s, the National Teacher Corps, and the elementary education program at the University of Wisconsin, Madison. Rodgers (2006) adds to our understanding of historical programs with her study of The Putney Graduate School (1950–64) which was known for turning the souls of its participants. The rise of social foundations courses as a core part of teacher preparation programs is also directly attributable to social reconstructionist or social justice conceptions of the purpose of schooling. While many of these early programs were small in scope, they offered compelling visions of this approach.

Since the late 1980s, reform conversations in teacher education highlight social justice. Several programs have demonstrated both intellectual willingness and practical approaches that illustrate the viability and vitality of this approach. These programs begin with the proposition that learning to teach is a transformative process. Though program structures (e.g., length of program, course sequences, and location and duration of field experiences) vary, most social justice programs have similar curricular aims.

Table 1 Comparative summary of conceptions, paradigms, conceptual orientations, and traditions of teacher education

<i>Conceptions of teacher education</i>	<i>Paradigms of teacher education (Zeichner, 1987)</i>	<i>Conceptual orientations (Feiman-Nemser, 1990)</i>	<i>Traditions of practice, traditions of reform (Liston and Zeichner, 1991; Zeichner, 1993)</i>
Social justice		Critical	Social-reconstructionist Teaching for understanding
Contemplative	Academic	Academic	
Clinician-professional	Personalistic	Personal	Developmentalist
	Behaviorist	Technological	Social-efficiency
	Traditional craft	Practical	
	Inquiry		

One aim is to help candidates develop conceptual tools to understand how the *status quo* marginalizes and oppresses some members of society. Teacher educators challenge and disrupt candidate's assumptions about how cultural diversity, race, class, gender, sexual orientation, and disabilities shape learner's experiences and opportunities in school and society. The aim is to foster candidate's critical consciousness. Through critical inquiry and reflection, candidates examine structural features of schools that promote inequality and, at the same time, analyze how their personal background might affect interactions with members of oppressed or marginalized groups. Such analyses guide candidates to grasp the political nature of teaching and inspire them to embrace the role of change agents.

While there is no one specific social justice pedagogical repertoire, most programs emphasize the following practices. Candidates learn to adapt and develop curriculum and assessments that reflect learner's lives and worlds and that foster learner's critical understanding of inequities. To promote access, candidates learn to build bridges from learner's worlds to school knowledge, to scaffold their thinking and learning of intellectually demanding academic tasks, and to provide appropriate linguistic supports for those whose primary language is not the language of instruction. Candidates develop abilities to establish and maintain safe, respectful classroom communities and to facilitate highly charged conversations, particularly ones that address race, class, language, and power. Many of the practical tools just described are consistent with culturally responsive approaches to teaching and thus are not the exclusive purview of a social justice conception of teacher education. One key distinction with social justice teacher education is the attempt to encourage candidates to take an activist role in standing up to unjust schooling practices. In recent years, this has involved taking a public stance on rigid accountability systems, creating places and spaces for parent voices to be heard, working on school or district committees to broaden the curriculum or to challenge tracking policies, and participating in community activist organizations.

Much scholarship on social justice conceptions of teacher education provides conceptual arguments for what capacities candidates need to develop along with

analytical descriptions of specific programs, courses, or learning experiences that foster this development. Rigorous studies examining evidence of effective social justice teacher-education programs are just emerging. McDonald (2005) found that programs explicitly committed to teach for social justice were more effective in helping candidates develop conceptual tools, rather than practical tools. Montano *et al.* (2002) also found that many of the typical structures in teacher education do not adequately prepare candidates to be change agents; thus, they recommend providing candidates with direct experiences in community-based activist organizations that allow them to identify with larger historical, social movements. In sum, since 1990, social justice conceptions of teacher education have been the center of research activity and program development. Much has taken place to give concrete shape to this conception of teacher education.

Teaching for Understanding Approach

Within a teaching for understanding approach, the central purpose of schools is to teach academic content. What comprises school subject matters has changed significantly over the last century, resulting in a rethinking of the teacher's role and the role of content in teacher preparation. In the early part of the twentieth century, academic content entailed mastering basic literacy, numeracy, and a defined body of content knowledge. Recitation was the most common teaching practice. The image of the teacher as a scholar-professional (Sokkett, 2008) prevailed. Essentially, the intellectual work of liberal arts study was the central feature of teacher preparation. Starting in the mid-twentieth century, reformers attempted to transform the content of the school curriculum from one based on factual recall to one where students grappled with substantive disciplinary questions and ideas, where students learned to evaluate claims and how new knowledge was constructed in academic fields, and where problem solving, dialog, and critical inquiry were normative. This notion of the curriculum synthesized and reflected Dewey's understanding that disciplinary content must be connected to learner's experiences, Bruner's argument for

teaching in intellectually honest ways, and Schwab's discussions of the substance and syntax of the disciplines. It is also consistent with constructivist and sociocultural theories of learning. Kennedy (2005) identifies three ideals that capture reform-minded views of teaching and learning academic content: (1) more rigorous and important content, (2) more intellectual engagement, and (3) universal access to knowledge.

Serious efforts to enact this approach to content in schools sparked a rethinking of the teacher's role in the classroom. Reformers often called this approach teaching for understanding. A good teacher in this approach acts as a facilitator. Essential to a teacher's pedagogical repertoire is the ability to structure engaging problems or essential questions that guide learners to construct understanding of disciplinary concepts rather than to receive finished knowledge. Teachers are keen perceivers of learner's thinking, attending to what is sensible in children's naive or novice ideas and guiding them to ways of thinking that are more accepted by scholars within the discipline. They facilitate dialog and manage an intellectually complex flow of information among learners. This is a significant reframing of the scholar-professional of the late nineteenth and early twentieth century, where imparting knowledge was the expectation. The emphasis on fostering the life of the mind and conceptual understanding of content provoked a renewed discussion of the role of content in teacher's preparation.

In 1986, Shulman (1986a) argued content was the missing paradigm in teacher education. He contended teachers must develop a pedagogical understanding of content, or pedagogical content knowledge, defined as

The most regularly taught topics in one's subject area, the most useful forms of representations of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations – in a word, ways of representing and formulating the subject that makes it comprehensible to others. Pedagogical content knowledge also includes an understanding of what makes the learning of specific topics easy or difficult; the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning of those most frequently taught topics and lessons. (Shulman, 1986b: 9, 10)

Shulman's conceptualization of pedagogical content knowledge both reflected and triggered a round of reforms centered on the role of content in teacher preparation. For example, one policy response in the United States was to reduce the number of required education courses and to increase courses in the academic major. However, the typical coursework within an academic major does not necessarily help teacher candidates develop pedagogical content knowledge, as that curriculum aims to prepare future scholars and researchers.

Yet, through content courses, teachers develop firmly held beliefs about the nature of the discipline and what content is most important, and these beliefs shape their classroom practice. As a result, starting in the mid-1980s, addressing teacher's beliefs about content became a primary site for teacher-education-program reform. Much of this activity occurred within content methods courses.

Distilling the research, Grossman *et al.* (2005) outline questions to guide curriculum and pedagogy in content methods courses. In addition to inquiry into the nature of the discipline, teacher candidates should examine how school knowledge is defined and elaborated, consider what student performances of understanding content matter look like, analyze primary curricula for teaching subject matter, investigate how teachers assess student's conceptual understanding, and gain familiarity with and appraise effective pedagogical practices in the content areas. Reformers who value teaching for understanding recognize that this kind of teaching is complex and ambitious and that it more likely to happen in workplace contexts that are organized to support teacher's collaboration. Teacher educators have developed or adopted practices to cultivate the candidate's pedagogical content knowledge and pedagogical repertoire to teach for understanding (e.g., lesson study, video clubs, case methods, and extended studies of student thinking about a particular topic). Overall, the emphasis on teaching for understanding as an approach to teacher education is most evident in content methods courses, rather than programmatic efforts; however, given the dominance of methods courses in many programs, teaching for understanding has come to occupy a featured space in the conceptual frameworks of many programs.

While teacher educators embraced the importance of pedagogical content knowledge and refined methods courses and programs around this aim, a vocal minority has persistently sought to return to a model where a deep exposure to disciplinary content is sufficient preparation to be an effective teacher. Despite challengers, teacher educators have moved substantially to conceptualize a vision of teaching and learning where both students and teachers engage in lively inquiry into the big ideas of the disciplines. In this conception, the teacher's active intellectual life is prominent.

Contemplative Approach

A contemplative approach to teacher preparation articulates with the most enduring of educational aims, the quest to know oneself, to connect with others, and to find one's place in the social and natural world. In this vision of education, learning occurs when individuals take up thoughtful explorations of meaning openly, critically, and even playfully. Together, teachers and students delve

into the mysteries and majesty of the human condition and the natural world. When compared to teaching for understanding, a contemplative approach offers a more transcendent vision of knowledge, one that, at its best, fosters wisdom along with understanding. Its roots are found in both humanist and spiritual traditions. To be a good teacher requires knowing one's subject and one's students, but most crucial is a teacher's knowledge of self. Palmer captures this view:

Teaching, like any truly human activity, emerges from one's inwardness, for better or worse. As I teach, I project the condition of my soul onto my students, my subject, and our way of being together. The entanglements I experience in the classroom are often no more or less than the convolutions of my inner life. Viewed from this angle, teaching holds a mirror to the soul. (Palmer, 1997: 15)

While the contemplative approach does not diminish the importance of craft and content knowledge, exploration of the inner landscape of a teacher's life is what anchors practice. The focus of teacher education thus turns to cultivating self-knowledge, developing an emotional and compassionate presence, and tapping into one's sense of vocation or mission. Teacher education is foremost about the development of identity and integrity rather than initial mastery of certain techniques or practical tools.

The contemplative bears a strong resemblance to approaches that others have described as personalistic, personal, and developmentalist (see **Table 1**), particularly in its commitment to developing and nurturing the self who teaches. Theoretical interest in self and identity that emerged in the 1990s gathers together studies of teacher beliefs and attitudes along with narrative studies of teacher's lives. This work provides new conceptual language and analytical tools to understand how teachers construct a sense of themselves. This theoretical lens on identity draws attention to the relationships between an individual's inner core or private self and one's situational and socially constructed identities. It also gives prime attention to the relational and emotional aspect of integrating one's evolving, and partially unknowable, inner landscape with one's external landscapes. It brings out the paradox that one must focus inward to turn outward. It affirms the central role of storytelling in the learning-to-teach process. Finally, it maps onto a larger developmental process of becoming a mature adult.

A related, but distinct, aspect of contemplative teacher education derives from a broader effort to introduce contemplative practices into schools and to characterize teaching as a contemplative practice. Drawing from contemplative traditions within several major religious traditions (e.g., Buddhism and Quaker), contemplation offers an antidote to the relentlessness and task orientation that seems to characterize contemporary schooling. Educators who turn to contemplation attend to the inner lives of

children and teachers. They speak about creating sacred space and time in the classroom, being mindful, and viewing students from a stance of wonder, receptivity, and acceptance (Miller, 1994). Contemplative teacher educators seek to introduce new teachers to contemplative practices enacted in k-12 settings and to model those practices within their own work.

Since contemplative teacher education is arguably the least mainstream of the four conceptions, programmatic examples are less abundant than the other approaches discussed in this article. Historical examples include the Bank Street College (in the 1930s), the Prospect School Teacher Education Program (1968–90), and the Humanistic Teacher Education Program at the University of Florida (Feiman-Nemser, 1990; Liston and Zeichner, 1991). These efforts reflect a progressive educational vision. Since the late 1990s, professional development, rather than initial teacher education, has been an active site for contemplative approaches, though these programs operate on the margins (see, e.g., the *Courage to Teach* program or the master's in Contemplative Education Program at Naropa University). It is premature to say that there is a pedagogy of contemplative teacher education. Distinctive features of courses and programs that fit within this approach include creating opportunities for candidates to draw meaning from their life histories, introducing candidates to meditative practices, giving attention to social-emotional dimensions of learning, and providing space for candidates to learn to listen and attend to children and youth.

Clinician-Professional Approach

A clinician-professional approach to teacher education aims to prepare teachers whose conceptual understandings and pedagogical repertoire reflect a consensually established professional knowledge base. (The term clinician-professional comes from Sockett's (2008) delineation of four moral and epistemic purposes for teacher education.) This approach implicitly compares teachers to medical doctors prepared in a Western context and views medical education as the model. A turning point in US medical education was the 1910 Flexner report, which resulted in a common curriculum grounded in the scientific basis for medical practice and the establishment of medical preparation in the university. Since the beginning of the twentieth century, teacher education, as a low-status profession, has struggled to gain credibility and respect. Invoking the Flexner report, one response to improve the quality and status of teacher education has been to establish a research-based, scientific approach to teaching and learning that informs teacher-education curriculum. In this sense, unlike the three conceptions discussed above, the clinician-professional

approach has a less distinctive normative vision for the broader aims of education and, by extension, good teaching.

The intellectual roots for a scientific approach to educational research extend back to Thorndike's argument (1910) in the first issue of the *Journal of Educational Psychology* where Thorndike asserted that experimental science and statistical analyses should guide all educational research. Within this approach, throughout the twentieth century, basic research in children's learning and development and scientific analyses of teaching practice shaped teacher-education curricula. Behavioral, cognitive, and sociocultural orientations within psychology played successive roles in defining conceptual content of the knowledge base for teaching, a good teacher's pedagogical repertoire, and how best to teach teachers. In this way, the clinician-professional approach is the latest and most sophisticated instantiation of the behavioral, technical, or social-efficiency traditions described by others (see **Table 1**).

Process-product research in the 1960s to the 1970s exemplifies a behaviorist inquiry into teaching. Process-product studies correlated specific, observable teaching behaviors (e.g., developing explicit objectives, time on task, wait time, or guided practice) with student achievement as measured on standardized tests. Subsequently, these discrete target teaching behaviors formed the basis for competency-based teacher-education programs. In the 1980s, the tight connection between behavioral psychology and teaching loosened as the field of psychology took a cognitive turn, and as educational research began to accept qualitative methods. Shulman's (1986a) critique of process-product research invited scholars to broaden research methods employed to examine teaching practice. His argument also spurred scholarly activity to define a research-based knowledge base for teaching. The cognitive turn gave greater attention to teacher's thinking, beliefs, and reflection. For this reason, teacher-education practices such as case methods or journals focused on fostering sound deliberation, decision making, and reflection.

Consistent in these evolving instantiations of a professional knowledge base is a commitment to social science research as the proper grounding for teacher's knowledge and practice. At the start of the twenty-first century, two distinct approaches to define the clinician-professional knowledge base and pedagogical repertoire are in play. The American Board for Certification of Teacher Excellence (ABCTE), in an effort to circumvent university-based teacher preparation, has developed a certification process based on passing multiple-choice exams that assess content and professional teaching knowledge. Exam content standards for the latter draw heavily from teaching behaviors identified in the process-product studies of the 1960s and 1970s. The multiple-choice exam format reinforces competency-based teacher education's emphasis on discrete, decontextualized practice. In contrast, the National Academy of Education's

Committee for Teaching Education publication *Preparing Teachers for a Changing World* (Darling-Hammond and Bransford, 2005) represents the most comprehensive attempt to codify the knowledge base and curriculum for teacher preparation. Research evidence that informs the National Academy report draws from studies conducted within different conceptual frameworks and methodologies (e.g., experimental, naturalistic observations).

The National Academy report envisions the teaching profession as serving democratic purposes. In this sense, it bears a resemblance to social justice conceptions of teacher education, as the aim is to foster children's learning and to enhance their life chances. The study organizes the knowledge base into three domains: (1) knowledge of learners and how they learn and develop, (2) conceptions of curriculum content and goals, and (3) understanding of teaching practices. While building upon previous articulations of a knowledge base, this report reflects recent scholarly activity emphasizing diversity, the role of language in learning, and the complexity of classroom worlds.

What distinguishes professional teachers is their knowledge and commitment to "practice based on what is known by the profession as a whole, rather than only one's own personal experience" (Darling-Hammond and Bransford, 2005: 15). Given the rapidly changing world of schools, the report positions teachers as adaptive experts, who are able to balance the dual aims of achieving efficiency in practice and engaging in innovation. The clinical-professional's pedagogical repertoire includes many of the research-based practices referenced in the three approaches described previously. Like a physician whose practice evolves as scientific studies refine or transform our understanding of the body and health/disease, the clinical-professional's repertoire evolves as understanding grows about how teachers support diverse students to learn complex subject matter.

Preparing to be a clinician-professional involves both carefully guided practice in evidence-based pedagogies and sustained inquiry into one's own practice. The contexts for teacher learning matter as learning will be most robust if it is both in and of practice. Overall, the clinician-professional approach tends to dominate teacher-education programs, in part because its comprehensiveness embraces aspects of other conceptions described in this article and in part because the knowledge base and pedagogical repertoire subsumed within this conception are codified in many national standards for program accreditation.

Conclusion

In closing, the four conceptions outlined in this article characterize the terrain of teacher education at the start of the twenty-first century. Each conception begins with an argument for a broad educational aim and its attendant

image of good teaching. By tracing back to each conception's intellectual roots and historical examples, current instantiations are located within an ongoing dialog of teacher-education practice. The influence of sociocultural learning theories, as it relates to both children/youth and teacher learning, is evident in all four approaches. Similarly, teacher educators who affiliate with each conception are grappling with how best to prepare teachers for increasingly heterogeneous classrooms. To educate teachers within a particular orientation involves creating opportunities for candidates to learn distinctive conceptual ideas and teaching practices. Yet, in all likelihood, most teacher-education programs reflect some blending of these conceptual approaches. The elaboration of each conception since the early 1990s (see **Table 1**) along with the dynamic interplay among them suggests a certain vitality and intellectual energy within the field of teacher education.

See also: A Pedagogy of Teacher Education; Case Methods in Teacher Education; Diversity in Teacher Education; Experienced Teachers' Craft Knowledge; Inquiry-Oriented Teacher Education; Moral Values in Teacher Education; Narratives and Biography in Teacher Education; Pedagogical Content Knowledge; Taking Prospective Teachers' Beliefs into Account in Teacher Education; Teacher Education and the Educational Foundations Knowledge Base; Teacher Learning with Lesson Study; The Relationship Between Theory and Practice in Teacher Education.

Bibliography

- Darling-Hammond, L. and Bransford, J. (eds.) (2005). *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do*. San Francisco, CA: Jossey-Bass.
- Feiman-Nemser, S. (1990). Teacher preparation: Structural and conceptual alternatives. In Houston, W. R. (ed.) *Handbook of Research on Teacher Education*, 2nd edn., pp 195–211. New York: Macmillan.
- Grossman, P., Schoenfeld, A., and Lee, C. (2005). Teaching subject matter. In Darling-Hammond, L. and Bransford, J. (eds.) *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do*, pp 201–231. San Francisco, CA: Jossey-Bass.
- Kennedy, M. (2005). *Inside Teaching: How Classroom Life Undermines Reform*. Cambridge, MA: Harvard University Press.
- Liston, D. and Zeichner, K. (1991). *Teacher Education and the Social Conditions of Schooling*. New York: Routledge.
- McDonald, M. (2005). The integration of social justice in teacher education: Dimensions of prospective teachers' opportunities to learn. *Journal of Teacher Education* **56**, 418–435.
- Miller, J. P. (1994). *The Contemplative Practitioner: Meditation in Education and the Professions*. Westport, CT: Bergin and Garvey.
- Montano, T., Lopez-Torres, L., DeLissovoy, N., Pacheco, M., and Stillman, J. (2002). Teachers as activists: Teacher development and alternate sites of learning. *Equity and Education* **35**, 265–275.
- Palmer, P. (1997). The heart of a teacher: Identity and integrity in teaching. *Change* **29**(6), 14–21.
- Rodgers, C. (2006). "The turning of one's soul" – learning to teach for social justice: The Putney Graduate School of Teacher Education (1950–1964). *Teachers College Record* **108**, 1266–1295.
- Shulman, L. S. (1986a). Paradigms and research programs in the study of teaching: A contemporary perspective. In Wittrock, M. C. (ed.) *Handbook of Research on Teaching*, 3rd edn., pp 3–36. New York: Macmillan.
- Shulman, L. S. (1986b). Those who understand: Knowledge growth in teaching. *Educational Researcher* **15**(2), 4–14.
- Sokkett, H. (2008). The moral and epistemic purposes of teacher education. In Cochran-Smith, M., Feiman-Nemser, S., and McIntyre, D. J. (eds.) *Handbook of Research on Teacher Education: Enduring Questions in Changing Contexts*, 3rd edn., pp 45–65. New York/London: Routledge, Taylor and Francis Group/The Association of Teacher Educators.
- Zeichner, K. (1987). Alternative paradigms of teacher education. *Journal of Teacher Education* **34**, 3–9.
- Zeichner, K. (1993). Traditions of practice in U.S. preservice teacher education programs. *Teaching and Teacher Education* **9**, 1–13.
- Adams, M., Bell, L., and Griffin, P. (eds.) (2007). *Teaching for Diversity and Social Justice*, 2nd edn. New York: Routledge.
- Cochran-Smith, M. and Zeichner, K. (eds.) (2005). *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education*. Mahwah, NJ: Erlbaum.
- Cochran-Smith, M., Feiman-Nemser, S., and McIntyre, D. (2008). *Handbook of Research on Teacher Education: Enduring Questions in Changing Contexts*, 3rd edn. New York/London: Routledge, Taylor and Francis Group/The Association of Teacher Educators.
- Darling-Hammond, L. (2006). *Powerful Teacher Education Programs: Lessons from Exemplary Programs*. San Francisco, CA: Jossey-Bass.
- Harteley, D. and Whitehead, M. (eds.) (2006). *Teacher Education: Major Themes in Education*. New York: Routledge.
- McInerney, D. and VanEtten, S. (eds.) (2003). *Sociocultural Influences in Teacher Education Programs*. Greenwich, CT: Information Age.
- Murray, F. B. (ed.) (1996). *The Teacher Educator's Handbook: Building a Knowledge Base for the Preparation of Teachers*. San Francisco, CA: Jossey-Bass.
- Reynolds, M. (1989). *Knowledge Base for the Beginning Teacher*. New York: Pergamon.
- Solomon, R. P. and Sekayi, D. (eds.) (2007). *Urban Teacher Education and Teaching: Innovative Practices for Diversity and Social Justice*. New York: Routledge.

Further Reading

Inquiry-Oriented Teacher Education

R Schulz, University of Manitoba, Winnipeg, MB, Canada

© 2010 Elsevier Ltd. All rights reserved.

The new scholarship of teaching and teacher education recognizes the enormous challenges of schooling in the twenty-first century, and emphasizes the importance of preparing teachers who understand the complexities of their role. Cochran-Smith (1999) contends that the most promising way of learning about such a complex role, so fraught with uncertainty, is based on inquiry within communities, rather than training for individuals.

In North America, the notion of inquiry was put forward over a century ago by Dewey (1904/1965), who criticized teacher education programs for placing too much emphasis on training and mastery or delivery of technical skills. He argued that, although first-hand practical experience in the schools is critical for the preparation of new teachers, the experience might well be miseducative if new teachers improved in the mechanics of teaching but failed to go on growing. Therefore, he advocated for teacher education programs which prepared new teachers to become critically thoughtful, ongoing students of education.

Some decades ago, in England, Lawrence Stenhouse (Rudduck and Hopkins, 1985) argued in a similar vein for the preparation of teachers to become extended professionals, proficient as problem solvers in their own schools, and knowledge generators for the profession. Stenhouse maintained that engaging in inquiry was essential to becoming an extended professional, and that systematic, self-critical inquiry was “linked to the strengthening of teacher judgement and consequently to the self-directed improvement of practice” (p. 3).

Inquiry can be defined as the systematic, intentional study of one’s own professional practice. (Cochran-Smith and Lytle, 1999). Inquiry is a stance, a way of knowing about teaching and learning from teaching. Inquiry is not a singular event – it is an integral part of teaching, a process that practitioners engage in consistently throughout their professional career as a way of learning from practice, even as they are engaged in the practice of teaching. Known by many names – teacher research, practitioner research, and insider research – teacher inquiry focuses on the concerns of teachers, as opposed to outside researchers, and involves classroom teachers in the disciplined and rigorous processes of researching their own questions in ways that can be scrutinized and verified by others as reliable and trustworthy. The term, as used in this article, therefore encompasses the various forms of research conducted by practitioners (individually and collaboratively) in their own schools and classrooms, for the purpose of increasing their own learning, enhancing

student learning, and contributing to school change. Key assumptions here are: (1) the knowledge needed to understand and improve teaching, and learning cannot be generated only by those outside; (2) teachers and teacher candidates who work in particular contexts are among those who have significant knowledge of both what the problems and questions are; and (3) through systematic data collection and analysis, how the problems within that context can be solved. Additionally, the knowledge generated by practitioners, and made public, is often also useful beyond the context in which it was generated. In this way inquiry becomes an important aspect of redefining teacher professionalism.

The conceptual scholarship on inquiry-oriented teacher education is extensive and rich, but the empirical research is not as robust. In her synthesis of the research on teacher education pedagogies, Grossman (2005) concluded that there were relatively few empirical studies about the impact of inquiry on teachers’ learning. Nevertheless, as we relinquish historically rooted notions of the teacher as “technician, consumer, receiver, transmitter, and implementer of other people’s knowledge” (Cochran-Smith and Lytle, 1999: 16) and shift to an expanded conception of the teacher as a knower, thinker, knowledge generator, and change agent, the understanding of what is required in teacher education programs has also changed, and an emphasis on inquiry is becoming more evident.

An inquiry orientation in teacher education constitutes a dramatic shift from traditionally organized programs, which disseminate a knowledge base constructed almost exclusively by outsiders. By contrast, inquiry-oriented teacher education programs present learning as a process of jointly creating knowledge. They see inquiry as an integral part of teaching, and focus on the processes of learning for teaching as well as learning from teaching. Teacher candidates learn about teaching by studying, in systematic and intentional ways, their own experiences both in the academic setting and in the field placement, at the same time they become critical consumers of the research of others. In a profession that usually strives for certainty and answers, inquiry-oriented teacher education programs prize questioning. Teaching is understood as a form of experience that is contextual, value-laden, intentional, as well as intuitive. As such, a different relationship between theory and practice is recognized. Methods, courses, and their historical preoccupation with technique are not the backbone of the program, and theoretical suppositions are not necessarily the starting point. Instead, the teacher candidates’ questions might be the

starting point, and as Smits (2007) suggests, active engagement with situations and in action is imperative with theoretical suppositions entering in a more dialogical fashion.

Different Understandings of Inquiry

Because of the widespread use of the term inquiry in teacher education, there is a need to discuss the many different understandings of what it means to be inquiry oriented. Definitions in practice and in the literature overlap, at times they pull in different directions and compete with each other. Various definitions of inquiry exist in science literature, and these may differ again from the way educators in humanities define inquiry and a pedagogy for inquiry. These different definitions and understandings include, among others: (1) inquiry as a process for co-constructing curriculum in the classroom; (2) inquiry as a process of questioning and knowledge building, and a means through which research is accomplished; (3) inquiry framed within concepts of construction and narration, in opposition to research, which comes from a tradition of discovery and revelation; (4) practical inquiry, where the aim is to generate or enhance practical knowledge; and (5) inquiry as a way of knowing in community, central to which is the idea that the work of inquiry communities is both social and political.

The point of inquiry, whether in the humanities or the sciences, is that there is always more to learn. As new knowledge is acquired in any discipline, the old, suddenly upended ideas seem quaintly uninformed. For children to be drawn into learning, educators know that they need to be with teachers who can engage them, immerse them in a current thought, entice them to pursue what is not known, and teach them the appropriate methods for this pursuit. If students are to become inquiry-minded critical thinkers and creative problem solvers, teachers can be no less so. The shared belief therefore, regardless of definition and discipline, is that if inquiry is to be a transformative force, what it is not, is merely technical tinkering.

While genuine inquiry might be conceptualized in the literature as a transformative force for teacher development and school change, it can also be reshaped to support a variety of educational or government agendas. Currently in England, where the standards movement holds sway, with its emphasis on accountability, inspection, regulation, centralized policy making, and nationally driven policy implementation, there is little encouragement for local meaning making, or constructing knowledge within the school community. Opportunities for defining one's own questions are reduced, and the inquiry teacher candidates engage in during their practicum placement often consists of helping the school make sense of new government initiatives or fine-tuning

strategies that make classroom routines or curriculum delivery more efficient (Schulz and Hall, 2004).

Inquiry is trivialized if it is understood only as a project, or if its purpose is primarily to tinker with and perfect certain skill sets.

As inquiry, with its breadth of definitions, becomes institutionalized in a greater variety of contexts, and co-opted by different political interests, there is a danger that it will become anything and everything, eventually devolving into nothing. However, in our efforts to reduce the plethora of definitions that surround inquiry, we want to be careful not to simplify the concept and the term to the extent that it can be contained within a single prescriptive, reductionist, and technical phrase. This would dilute its potency. The multiple definitions of inquiry need not be a liability. They can, in fact, be seen to enlarge and enrich the term, the process, and the possibilities, when those committed to the broader transformative goals of inquiry continue to debate and discuss among themselves their various understandings and practices. An ongoing conversation about what it means to inquire, best engaged in jointly by teacher educators, teacher candidates, and practicing teachers, if incorporated as a central component of a teacher education program, would create a second layer of inquiry, an inquiry into inquiry, whereby the practice of inquiry is modeled for teacher candidates, at the same time that the processes and outcomes of inquiry are being studied.

Inquiry is a stance, a way of being, and a habit of mind. It is a curiosity about one's own and other's learning, a wanting to know, being open to different possibilities and new ways of knowing, believing, that it is in wondering and questioning that learning begins, and then pursuing that learning in systematic and intentional ways. Taking an inquiry stance means becoming, as Dewey suggested, a "thoughtful and alert student of education." Alfred North Whitehead used the term "scholarly ignorance." If we recognize that there is much that we do not know, that there is something important to be gained by the process of inquiry, then it is our scholarly ignorance that can be generative of genuine inquiry.

Inquiry-Oriented Teacher Education Programs

Inquiry-oriented teacher education programs generally expect their candidates to be inquirers/researchers who have a breadth of knowledge and a variety of tools to ask questions about educational environments. Inquiry-oriented programs focus on learning for teaching as well as learning from teaching. They expect that teacher candidates will reflect on and continually evaluate the effects of their choices on students, their families, and other professionals in the learning community. Programs that

raise the awareness of teacher candidates to the larger issues of education, and encourage dispositions of inquiry, must also provide their teacher candidates with the skills and tools to confront the complexities and uncertainties of teaching. Different teacher education programs in different countries have approached this task differently.

For the last 30 years, teacher education in Finland has been research-based. Proceeding from the belief that effective teachers cannot only be consumers of research, that they must also be producers of research, teacher education programs in Finland require their teacher candidates to take both qualitative and quantitative research courses, as a way of preparing them for their extended professional teaching role. Currently, universities in Hungary are restructuring their teacher education programs. Recognizing Finland's success in international comparisons of education, Hungary is drawing on lessons about inquiry-oriented approaches from that country (Csapo, 2007).

Darling-Hammond (2006), in her work, identified a number of features common to seven exemplary US teacher education programs. All the programs she studied required that teacher candidates engage in inquiries or research about teaching, and Herr and Price (2007) estimate that approximately half the teacher education programs in the United States now require their teacher candidates to engage in action research, reflecting the current emphasis in that country of preparing teachers who see themselves as learners and inquirers into their own practice.

From within an inquiry-based literacy teacher education paradigm, teacher candidates learn to teach by observing children closely, being curious about children's wondering, supporting them in making their literacy learning visible, and collaboratively opening up worthwhile inquiry potentials. Curriculum development in the classroom is seen as a shared process, with learners and teachers collaborating together to make decisions based on the experiences and needs of a specific community of learners. There is a sense of negotiation and fine balance between what the children bring to the experience and the teacher's professional understandings of teaching, learning, and young children's development. Choice, decision making, and reflective practice inform and define the curriculum that develops. It is understood that there is not a single official learning path or end point. Teacher candidates develop their professional sense through reflecting in action, engaging in continued professional reading, and dialoguing with trusted others who are positioned differently within a professional inquiry community but who take teacher candidates' ideas seriously.

Another perspective on inquiry-oriented teacher education comes from the work of Korthagen *et al.* (2006). Drawing on their analysis of three teacher education programs in Australia, Canada, and the Netherlands, they outline some fundamental principles that they

believe make a difference. These include a belief about the importance of student teachers engaging in inquiry through researching their own practice and experiences as teacher candidates. They write that when teacher candidates create and share "their understanding of practice through results of their own research, perceived distinctions between theory, practice, transmission of knowledge and socialization into teaching may be confronted and their professional roles may be better recognized, defined and enhanced" (p. 1035). Because the contexts of the programs that these three researchers studied share similarities to many others around the world, they contend that this fundamental principle is also applicable to other contexts.

Consistent with international trends, an emergent interest in inquiry-based teaching and teacher education is also occurring in China, with movement away from traditional didactic teaching practices and a heavy emphasis on rote memorization, to an approach that incorporates inquiry methods. Stewart (2006), Zhou (2006), and Zhang (2005) report that while most teachers and teacher educators are still unfamiliar with inquiry-oriented approaches, many are showing interest in this new direction. At the same time, however, there is resistance to this call for change, and there are numerous systemic and cultural barriers to the immediate enactment of inquiry-oriented education.

The sole teacher-preparation provider in Singapore, the National Institute of Education, has also undertaken a major review of teacher preparation programs. Reconceptualizing the role of the teacher has become the focus of attention, and Tan *et al.* (2007) note that the emphasis is on developing "future-oriented teachers who are adaptable and flexible to meet the uncertain demands of a changing environment. The programs are designed to emphasize inquiry, innovation, reflection, mutual respect, personal connection, collaboration and community . . ." (p. 14).

If we think of inquiry as a way of knowing in communities, then taking an inquiry stance can mean, among other things, careful observation of what is happening in classrooms, documenting and analyzing practices, sharing and critiquing these with colleagues, as a way of learning from and with others, participating in study groups, going beyond discussions of "what works" to understanding "why" it works, and then extending the discussion to examine and question the values and assumptions that underpin our practices. In Japan, the practice of lesson planning (Stigler and Hiebert, 1999) is an example of one form of sustained, focused, and collaborative inquiry. Japanese math teachers work in groups, sometimes for months at a time, to plan a single math lesson, which is then taught, observed by fellow teachers, analyzed, modified, improved and retaught over and over again. The focus of this ongoing collaborative inquiry and refinement of practice is not on individual teachers and their teaching styles, but on

teaching methods, and how these might be improved in ways that support student learning. Japanese teachers who participate in this form of inquiry into their teaching see themselves as contributing to their own professional development, as well as to the more general development and dissemination of knowledge about teaching.

As Britzman *et al.* (1997) suggest, a great deal of the work of teacher education should be to “produce debate, multiple perspectives on events, practices, and effects, to move toward creative dialogue on practices. . .”(p. 20). In addition to a focus on research courses and participation in action research, collaborative lesson planning, and study groups, inquiry-oriented teacher education programs frequently utilize case studies, videotaping, critical incident methodology, and protocols to guide the way in which teacher educators and teacher candidates explore problems of practice together.

Other commonly used inquiry tools are reflective journals, essays, and portfolios. When these are used in very structured ways, their focus tends to be on a rational and logical analysis of experiences. A more comprehensive approach to reflection also pays attention to nonrational processes that might be accessed through visioning exercises, metaphor analysis, or the use of drawings, as Mitchell and Weber (1999) advocate, to help teacher candidates become aware of the often unarticulated beliefs and values that influence their actions in the classroom.

Self-study through various means, such as writing teacher stories, life stories, or autobiographies, provides another opportunity to explore the connections between our personal and professional lives (Loughran and Russell, 2002). Through self-study we can gain a better understanding of why we teach the way we do, how our teaching is connected to both our disciplinary backgrounds as well as to who we are, and how we might teach differently when faced with students who are very different from us. Autobiographical understanding is an important part of inquiring into what it means to practice as a teacher, and as such, it is foundational to change.

The practicum plays a critical role in every teacher education program. It has traditionally been the place where teacher candidates are evaluated on their performance or their delivery of newly learned techniques, a place where the focus is generally on technical knowledge, rather than a place where deep thinking is developed, where good learning – which involves inquiry – happens (Schulz, 2005). Certainly, teacher education programs must address the technical and procedural aspects of teaching, because planning for teaching and managing a classroom are of unarguable importance. However, the attainment of craft skills, while necessary, is not sufficient preparation for the professional role of teaching. If the practicum experience is to be educative, it must also provide opportunities for growth and learning, must help teacher candidates to understand the full scope of the

teacher’s role, and develop in them the capacity to learn from future experiences.

A promising format for inquiry within the practicum could begin with a question that the teacher candidate wants to pursue. The underlying assumption here is that if the inquiry is to lead to an instructional improvement, it needs to reflect the personal interest of the inquirer. After generating a question, the next step would be to gather evidence in the form of student and teacher work. The data could include the teacher candidate’s observational notes, notes from peer observations, students’ assignments, and videotapes of classroom interactions or students commenting on their work. Ideally, multiple perspectives are brought to the issue in order to understand the many dimensions embodied in the evidence from the classroom. To this end, the data are examined and discussed with the collaborating teacher, with other teacher candidates in the practicum setting, and with the university professor, who can also draw the group’s attention to sources in the research literature, relevant to the topic at hand.

Structures, such as protocols for discussion, can facilitate the goals of inquiry. They guide the discussants, deepen the conversation, help to surface thought-provoking questions, and open up other avenues of exploration. A common protocol practice is to have the presenting teacher candidate put forward the evidence he or she has collected in response to the original inquiry question, but without any elaborating commentary. The first opportunity to comment is given to the other group members who, individually, are asked to describe the evidence before them, making no judgments initially, but raising questions about the data, which can often lead to new insights about the issue. Only after the others at the table have completed their initial commentaries on the evidence, does the teacher candidate explain why he or she chose particular pieces of evidence to share with the group and why they seem to have significance. This is followed by a whole group discussion of what the evidence presented by the teacher candidate tells them about the inquiry question under consideration, what other evidence might be helpful in considering this question, suggestions for changes in practice based on new understandings, and what might be the focus for the next meeting.

Many teacher education programs are now also creating spaces and providing supports for their teacher candidates to engage in inquiry outside the more conventional courses and practicum sites. They incorporate into their programs opportunities for teacher candidates to work with children in hospitals, museums, or art galleries, to join activist groups focused on particular issues or school policies, to participate with parents in community initiatives, or to collaborate with faculty members in research projects. These alternative sites of learning become increasingly important as teacher candidates engage with diverse others to make meaning of their experiences, and question the status quo within schools.

In addition, a number of teacher education institutions are continuing their commitment to their teacher candidates beyond graduation, by supporting them in various ways through their early career experiences. This support often takes the form of regularly scheduled workshops for teacher candidates, where the preservice focus on inquiry is taken up again, but now within the context of the new teachers' classroom settings. Online circles of inquiry facilitate communication among new teachers and their former professors and provide additional electronic networking and support for the new teachers.

The importance of dialog to an inquiry stance is evident in the research currently being conducted in nine different countries, with teachers of English as a foreign language. The Brazilian component of this research (Jordao, 2007) involves the construction of open spaces in teacher education programs where teacher candidates and practicing teachers interact dialogically, to question and be questioned in their assumptions about their identities, their social roles, the role of English as a foreign language, their students' identities, the value of education, and the educational value of English as a foreign language in Brazil. The focus within these open spaces is not entirely on a rational and logical analysis of experiences. Equally important is intuitive and affective engagement with alternative perspectives, where multiple literacies in the form of images, film, caricatures, and narratives are used to illustrate to teachers that even our most reasoned and logical positions are ideological and localized in specific contexts. This can lead to an understanding that no position or theoretical stance is inherently better, and teacher candidates as well as teachers then begin to feel more at ease to construct and locate their own interpretations and value systems, and to engage with different perspectives more critically.

When teachers and teacher candidates are encouraged to pursue their personal wonderings, and their practical experiences are structured so that they can work together in an ongoing way with their peers, their professors, collaborating teachers, parents, and other professionals to share ideas, solve problems, and explore issues – not just at the technical level, but also at the deeper interpretive and critical levels – then they are experiencing effective collaborative inquiry.

Issues and Challenges to Inquiry-Oriented Teacher Education

Even as teacher education is moving toward a more complex notion of teaching, the political climate in many jurisdictions is moving toward a more technical stance. Given these tendencies, it becomes all the more important that teacher candidates are thoughtful and alert students of education who understand the political and ideological restructuring that is occurring, and who have the

knowledge, skills, and disposition to deconstruct and respond to the events around them.

Anderson and Herr (1999) believe that teacher research is presently marginalized, at least in part, because it is a potential threat to the hegemony of traditional forms of research, and if teachers wish to counter entrenched epistemologies, their contributions (through their inquiry practices) to a counter-discourse will have to be strong, intellectually rigorous, and useful. However, it is difficult for teacher candidates and teachers to act upon their commitment to inquiry when the culture of the school they find themselves in is not receptive to the concept.

Schulz and Hall (2004) found that genuine inquiry was seldom a priority in the English and Canadian schools they studied. Echoing the findings from many other studies, the teacher candidates in Schulz and Mandzuk's (2005) study expressed dismay at the disconnect between what they had been taught in their university classrooms, and what they were experiencing in their practicum settings. The exhausting dailiness of school demands seemed to suppress inclinations toward contemplation and inquiry, and the teacher candidates were concerned that what they were learning in their university classrooms would not be well received in the school classrooms. When the school learning community values more local and particular forms of learning and knowledge, it can provide the necessary support for teachers' emerging understandings as they inquire. Within a school community that values inquiry, where teachers and teacher candidates have administrative, collegial, and university support, they can find a space where taking risks and theorizing about teaching is encouraged. However, when the school setting does not support teachers' inquiry stance, the work of the university teacher education program is easily eroded.

Therefore, teacher educators have a responsibility to challenge school resistance to genuine inquiry, to work collaboratively with schools to promote ongoing teacher learning, and decrease the positional differences between the worldviews of school and university. A shared school/university commitment to teacher preparation and support holds much promise, but it also brings its own challenges. Working collaboratively across institutions that reflect the cultural divide between schools and universities is a complex undertaking in and of itself. Additionally, the institutional values of universities neither encourage nor adequately reward professors' practical work with teachers in the field, privileging instead, more traditional paths of research and publication. Similarly, state education officials seldom encourage teachers who want to be recognized as knowledge generators. Instead, they are more likely to become anxious when teachers find their voice as educators, question local policies, and propose alternative agendas. In Hungary, for example, educators involved in the reform of teacher education see the example of

Finland as a model to follow, but state restrictions will not allow them to implement all the educative reforms they propose (Csapo, 2007). In China, the cultural barriers to change are even more pronounced. These examples alert us to some of the broader systemic issues related to inquiry-oriented teaching and teacher education.

A large body of work currently exists that theorizes and describes the use of inquiry as a way to enhance both teacher and student learning. The scholarship in teacher education contends that inquiry-oriented approaches to teacher education disrupt the deskilling of teaching and encourage resistance to the implementation of ineffective teaching practices. Inquiry-oriented approaches in teacher education hold the promise of nurturing the intellectual leadership capacity of teacher candidates and supporting a disposition toward critical thoughtfulness and ongoing learning. Over the last 25 years, the growth of this body of literature has been accompanied by an ever-growing number of teacher education programs around the world that describe themselves as inquiry oriented. A next step in this development is for teacher educators to model and exhibit an inquiry stance themselves by producing sound, robust evidence of the impact of inquiry-oriented teacher education on the learning of students. Long-term programs of research, as part of the program of teacher preparation, should become the norm, with ties to student learning established wherever possible.

Bibliography

- Anderson, G. and Herr, K. (1999). The new paradigm wars: Is there room for rigorous practitioner knowledge in schools and universities? *Educational Researcher* 28(5), 12–21.
- Britzman, D., Dipbo, D., Searle, D., and Pitt, A. (1997). Toward an academic framework for thinking about teacher education. *Teaching Education* 9(7), 15–26.
- Cochran-Smith, M. (1999). Learning to teach for social justice. In Griffin, G. (ed.) *Teacher Education: 98th Yearbook of the National Society for the Study of Education*. Chicago, IL: University of Chicago Press.
- Cochran-Smith, M. and Lytle, S. (1999). The teacher research movement: A decade later. *Educational Researcher* 28(7), 15–25.
- Csapo, B. (2007). Research-based teacher education: International trends and national characteristics. *Paper Presented at Borders and Crossings*. Inaugural Partnership Conference, University of Manitoba/University of Szeged. Szeged, Hungary.
- Darling-Hammond, L. (2006). *Powerful Teacher Education*. San Francisco, CA: Jossey-Bass.
- Dewey, J. (1904/1965). The relation of theory to practice in education. Borrowman, M. (ed.) *Teacher Education in America: A Documentary History*. New York: Teachers College Press.
- Grossman, P. L. (2005). Research on pedagogical approaches in teacher education. In Cochran-Smith, M. and Zeichner, K. (eds.) *Review of Research in Teacher Education*. Washington, DC: American Educational Research Association.
- Herr, K. and Price, J. (2007). Centering social justice and equity in action research: Pedagogical challenges for teacher educators. *Paper Presented at the American Educational Research Association*. Chicago.
- Jordao, C. M. (2007). The identity process of EFL teachers in a Brazilian context. *Paper Presented at the Centre for Globalization Studies*. University of Manitoba, Canada.
- Korthagen, F., Loughran, J., and Russell, T. (2006). Developing fundamental principles for teacher education programs and practices. *Teaching and Teacher Education* 22(8), 1020–1041.
- Loughran, J. L. and Russell, T. (eds.) (2002). *Improving Teacher Education Practices through Self-Study*. New York: Routledge Falmer.
- Mitchell, C. and Weber, S. (1999). *Reinventing Ourselves as Teachers: Beyond Nostalgia*. Philadelphia, PA: Falmer.
- Rudduck, J. and Hopkins, D. (eds.) (1985). *Research as a Basis for Teaching. Readings from the Work of Lawrence Stenhouse*. London: Heinemann.
- Schulz, R. (2005). The practicum: More than practice. *Canadian Journal of Education* 28(1–2), 146–167.
- Schulz, R. and Hall, C. (2004). Difficulties in promoting inquiry in teacher education partnerships: English and Canadian perspectives. *Journal of Education for Teaching* 30(3), 255–269.
- Schulz, R. and Mandzuk, D. (2005). Learning to teach, learning to inquire: A 3-year study of teacher candidates' experiences. *Teaching and Teacher Education* 21, 315–331.
- Smits, H. (2007). Myths of inquiry: Inquiring into teaching and learning in teacher education. Paper Presented at the Western Canadian Association for Student Teaching Conference. Canada: Winnipeg.
- Stewart, V. (2006). China's modernization plan. *Education Week* 25(28), 1–3.
- Stigler, J. W. and Hiebert, J. (1999). *The Teaching Gap*. New York: Free Press.
- Tan, C. T., Chong, S., and Wong, A. (2007). Reconceptualizing teacher education in the knowledge age: Balancing change and values development. *Asia Insights*, No. 2.
- Zhang, B. (2005). Opportunities and challenges of China's inquiry-based education reform in middle and high schools: Perspectives of science teachers and teacher educators. *International Journal of Science and Mathematics Education* 1(4), 477–503.
- Zhou, N. (2006). Reflections on curriculum change: Overview of directions, policy issues and capacity building in Asia-Pacific contexts. China National Institute of Educational Research. Seminar/Workshop on the Management of Curriculum Change. June, Quezon City, Philippines.

Further Reading

- Bullough, R. V. and Gitlin, A. (1995). *Becoming a Student of Teaching: Methodologies for Exploring Self and School Context*. New York: Garland.
- Cochran-Smith, M. and Lytle, S. (2004). Practitioner inquiry, knowledge, and university culture. Loughran, J., Hamilton, M. L., LaBoskey, V., and Russell, T. (eds.) *International Handbook of Research of Self-Study of Teaching and Teacher Education Practices*. Amsterdam: Kluwer.
- Cochran-Smith, M. and Zeichner, K. M. (eds.) (2005). *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education*. London: Lawrence.
- Cole, A. and Knowles, G. (2000). *Researching Teaching. Exploring Practice through Reflexive Inquiry*. Needham Heights, MA: Allyn and Bacon.
- Darling-Hammond, L. and Bransford, J. (eds.) (2005). *Preparing Teachers for a Changing World. What Teachers Should Learn and Be Able to Do*. San Francisco, CA: Jossey-Bass.
- Davis, B., Sumara, D., and Luce-Kapler, R. (2008). *Engaging Minds: Changing Teaching in Complex Times*, 2nd edn. New York: Routledge.
- Hoban, G. F. (ed.) (2005). *The Missing Links in Teacher Education Design*. New York: Springer.
- Jakku-Sihvonen, R. and Niemi, H. (eds.) (2006). *Research-Based Teacher Education in Finland: Reflections by Finnish Teacher Educators. Research in Educational Sciences* 25. Turku: Finnish Educational Research Association.
- Weinbaum, A., Allen, D., Blythe, T., et al. (2004). *Teaching as Inquiry: Asking Hard Questions to Improve Practice and Student Achievement*. New York: Teachers College Press.

Narratives and Biography in Teacher Education

G Kelchtermans, Katholieke Universiteit Leuven, Leuven, Belgium

© 2010 Elsevier Ltd. All rights reserved.

The Narrative Turn in Educational Research and Teacher Education

The 1980s showed a growing international interest among researchers in the field of human sciences in narrative and biographical approaches and methods. Since “narrative is the discourse structure in which human action receives its form and through which it is meaningful” (Polkinghorne, 1988, p. 135), narratives were considered as a powerful way to unravel and understand the complex processes of sense-making that constitute teaching.

Educational researchers quickly joined this theoretical, epistemological, and methodological development. The evolution came to be known as the narrative turn and it represented an important shift in the way teachers and their work were conceived of. Teachers’ talk about their professional lives and practices is very often spontaneously framed in narrative form. They use anecdotes, metaphors, images, and other types of storytelling to recall, share, exchange, or account for their experiences in classrooms and schools. Storytelling is the natural way through which people make sense of the events, situations, and encounters they find themselves in. “Humans are storytelling organisms who, individually and socially, lead storied lives. The study of narrative, therefore, is the study of the ways humans experience the world” (Connelly and Clandinin, 1990, p. 2).

Teaching is a complex social process, in which different people are interactively engaged. Teachers and students interact based on their interpretation of the situation, their own actions, as well as those of the others. Sense-making lies at the very heart of the teaching and learning process. Narrative accounts of these experienced interactions reveal the particular meaning they have or had for the people involved.

Using one exemplary illustration of how narrative techniques are being used in teacher education, the narrative turn will be further explained both in its practical applications and its theoretical grounding.

The Critical Incident Vignette: A Storied Example

The seminar room fills up with chatter as about 20 student teachers enter, move around, and find a seat in the circle of tables and chairs. Throughout the chatter, I pick up excited comments on experiences during their

recently ended practical training. After a brief welcome, I announce that the exercise I would want them to engage in will encompass three phases. An individual phase, a group phase, and finally a plenary phase that will take place during the next meeting in the seminar.

As a start I ask all of them to take paper and pen for an individual autobiographical reflection. They are asked to recall the best or worst teacher (the choice is theirs) they have ever experienced during their lives as a pupil or a student. The task is to focus on one typical event or experience with that former teacher and write a short narrative account of the event. A possible alternative assignment is to have them reflect back on their best or worst experience in their school lives as a pupil or a student. The assignment is most often reacted on by the student teachers with surprise, some hesitating smiles or raised eyebrows, but they all soon concentrate on the paper. After only a few minutes everyone is busy writing. Deep sighs, grim looks, and also amused smiles can be observed.

As most of them are about to end their writing, I ask them to find a catchy title for the piece. The title should both trigger the reader’s curiosity and summarize the key idea that is supposed to become clear from the vignette. Finally, I ask all the participants to turn the sheet and use the back to elaborate in a more systematic and explicit way what the experience has meant to them, what constituted its particular significance, and what possible relevance the story does have for their actual work as (student) teachers.

At the end of the meeting, instructions are given for the second phase of the exercise. The participants are being asked to meet in groups of three to perform a triangle discussion. They should themselves pick a time and place to meet and should allow for about one and a half hour for the discussion. During that discussion the participants are required to perform and strictly stick to one of the three prescribed roles: storyteller, soundboard, or listener. The first role is that of the storyteller: he/she starts with sharing the title and the story, followed by his/her reflective analysis of its actual significance. Once the storyteller has finished, the soundboard starts to play his/her role. The person in this role can ask clarifying questions about context, particular details, etc., which the storyteller then answers or explains. No comments or interpretations are to be made. This is the privilege of the listener, the third role. He/she can suggest different interpretations or open up alternative ways of viewing the experiences from the

story. However, all reflections or interpretations have to remain grounded in the story. No noncommittal speculations are allowed. Finally, the storyteller gets a chance to react with final comments.

Once this round is over, everyone changes roles and the cycle is repeated until every participant in the triangle has performed each of the three roles. The strict distinction between the role of soundboard and listener may have an artificial feel, but it is deliberately included to force the participants to suppress their spontaneous tendency to interpret the story against the background of their own experiences. Postponing this interpretation adds to the necessary active listening as well as to a growing awareness about one's temptation to engage in overhasty commenting.

The third and last phase in the exercise is a plenary discussion during the next seminar meeting and concentrates on the insights that were revealed through the structured storytelling and analysis. The discussion is guided and directed in such a way that the focus is on the participants' actual thinking about teaching and about themselves as teachers. Furthermore, the public reflection should also include relevant elements of the professional context and the relationship between that context and the narrator. A climate of trust and safety is established for the discussion and all stories and reflections are, in principle, being accepted and valued. At the same time, however, critical questions are being asked, alternative interpretations provided, different types of arguments juxtaposed and commented on. Throughout the discussion, links are made to relevant theoretical perspectives and concepts to (re)frame issues, or to deepen the reflective analysis.

This brief account of the critical incident vignette, as I developed and used in both preservice and in-service training (Kelchtermans, 2007), is meant as an example of how narratives can be purposefully used in teacher education. It allows the participants to experience how a personal anecdote reveals not only its significance while it happened, but also its relation with the present. The stories from the past reveal a lot about the way the storyteller thinks about him/herself in the present, especially when used with experienced teachers or with student teachers in the final phase of their training (after internship experiences). The context (for e.g., other people in the scene, the organizational conditions, the social and cultural environment in and around the school, etc.) is also included and its particular meaning and impact on the storyteller become evident in the emerging picture. Taking up the different roles contributes to an increased awareness of the social, interactive character of the shared sense-making that takes place. Furthermore, the interactions entail the possibility of breaking up taken-for-granted interpretations, providing alternative understanding of that past experience and – as a consequence – of one's actual situation.

Narrative Practices with Student Teachers

The actual practices of using narratives and biographical accounts in teacher education are quite diverse, yet they very often reflect a common underlying rationale. This can be illustrated by the example of the critical incident vignette.

Both research and common experiences show that life stories most often contain events that appear to have had a very significant meaning for the narrator. Often these events are all but extraordinary or spectacular. Sometimes they may even look trivial to others. But for the storyteller these events, encounters, experiences are their turning points, which made them reflect and reconsider their taken-for-granted views. These so-called critical incidents are "key events in an individual's life, and around which pivotal decisions revolve. They provoke the individual into selecting particular kinds of actions, which lead in particular directions. (...) Critical incidents are a useful area to study, because they reveal, like a flashbulb, the major choice and change times in people's lives" (Sikes *et al.*, 1985, p. 57).

The critical incident vignette is one exercise that takes up this idea. The assignment to reflectively recall, narratively share and analyze a particularly meaningful event can take different forms. Yet it always, more or less, includes the three steps: (1) description of the event, (2) identification of the issue that is raised in or by it, and (3) a statement of the sense the student made of it with respect to his/her process of learning to teach (Carter, 1993).

It thus always aims at triggering a reflective move from the storied event to an analysis of its particular meaning and relevance to the narrator's actual thinking and actions.

Another group of narrative techniques or exercises draws on metaphorical thinking: (student) teachers are invited to come up with a metaphor to typify themselves as teachers or to define what, in their opinion, constitutes the core of good teaching. These personal teaching metaphors (Bullough, 1991) thus represent, in a condensed, narrative way, their key ideas about themselves as (student) teachers. They may also reveal how these ideas can change over time. As such it makes sense to engage (student) teachers in exercises with personal teaching metaphors or accounts of critical events at different moments during the process of learning to teach (e.g., as an element in a reflective portfolio which they compose during their teacher education or as part of a professional diary). Comparing the differences between that actual metaphor and the ones developed at earlier stages helps to explore and understand one's personal learning process as well as the experiences that triggered the changes.

Narrative methods also allow student teachers' reflections to move beyond the technical issues of appropriately

using knowledge and skills to achieve effective teaching performances (Kelchtermans, 2007). Very often the student teachers' reflections as well as the teacher educator's feedback on practical teaching activities concentrate merely on these technical issues of how to improve one's actions. Although the importance of this technical dimension in teacher professionalism goes undisputed, there is more to good teaching than skilful performance and efficiently dealing with the "how to" questions. Teaching and being a teacher inevitably also involves deciding on moral dilemmas, living through emotional experiences, and dealing with political issues of power and influence. Coping with these moral, emotional, and political dimensions of teaching constitutes an equally important agenda for teacher education as do the issues of teaching skills and (content) knowledge. Through narrative accounts of teaching experiences, student teachers can reflectively address their moral puzzlements (e.g., conflicting norms about good teaching), emotional concerns (e.g., the experience of powerlessness or self-doubt when pupils do not show the learning outcomes that one had aimed for), or political actions (e.g., conflicts with colleagues or principals about the most appropriate content or methodology in teaching). Sharing and discussing on these dimensions help student teachers to accept these dimensions as inherent to the teaching profession and thus essential to their identity as teachers.

In order to achieve this, a common underlying pattern composed of two complementary elements can be found in the different narrative techniques in teacher education. First, the techniques constitute a vehicle for helping the narrators express through different narrative forms their deeply held ideas about themselves as a teacher, their values, and beliefs about (good) teaching, motivation, etc. Second, this individual, personal storytelling however, always needs to be complemented by particular interventions and reactions by the audience, that is, peers and/or teacher educators. Their comments and questions on the publicly shared narrative accounts aim at questioning, probing, challenging, and deepening the reflective analysis. The storytelling in itself is not enough. Many authors stress the need to link story-telling with forms of narrative inquiry, in which the stories and the beliefs they reveal are critically analyzed, questioned, re-storied. Interpersonal dialog, narrative accounts, and the critical use of theoretical knowledge thus have to go hand in hand.

The striving for authenticity as well as the critical probing of the accounts constitute two essential and necessary principles for the guidance of narrative practices. The guidance must avoid the practice wherein the sharing of the narratives does not simply affirm or strengthen student teachers' personal ideas and actions. Critical and public examination may result either in a more informed conviction and awareness of one's ideas, or in a thoughtful

reconsideration and rethinking of what was until then taken for granted.

Although an environment that is supportive, trusting, and positive is a necessary condition to allow the sharing of and learning from narratives to take place, it remains a continuous task for the users of narratives to avoid the trap of uncritical self-satisfaction. The way to achieve this is to create and maintain a productive level of discomfort that helps the critical reflection to move beyond the story as such.

In order to properly balance safety and challenge in these activities, the sharing of autobiographical anecdotes by teacher educators with their students is often used to lay the ground and build the necessary trust to engage the student teachers in narrative accounts about themselves and their lives.

Narrative Language, Context, and Voice

Although teachers' professional actions are, to an important degree, determined by the structural conditions of policy regulations, institutional environments, and organizational settings, the narrative turn explicitly reminds that teaching and learning to teach cannot be properly understood or conceived without acknowledging the obvious role of agency and sense-making. Teachers are competent social agents engaged in educating children. Teaching thus involves planning, deliberation, intentional and purposeful intervention, and action. Yet, its actual evolvement in the work with students can never be fully predicted or controlled. Therefore, technical and scientific language is too limited to capture the full experience of teaching and learning to teach.

Narrative language offers possibilities to get beyond these limits, because of a series of particular characteristics. First, a core process in narratives is the emplotment: by situating experiences in a narrative, the narrator organizes them in a particular linguistic frame and order (beginning–middle–end) through which the particular meaning of his/her experiences emerges (plot).

Narrative language further performs at the same time a referential and an evaluative function. The referential function describes events and experiences from the past in a temporal order. The evaluative function links the events with the moment of narrating by revealing what the experiences meant for the people involved in the present (Labov and Waletzky, 1973).

Third, narratives thus always link past experiences to the actual social context. They even constitute that context, because they reflect a need to communicate, to share a meaningful experience with others (Hoepfel, 1983). Through storytelling a particular social setting emerges with different actors and roles. On the one hand, there is the narrator or storyteller who provides the narrative. His/her narrative account allows for the personal, subjective

voice to be spoken and heard. Yet, on the other hand, in the telling of the story there is always an audience already implied. Narratives are told or written by someone, who always by that time has an explicit or implicit idea of possible audiences. As such, the narratives are fundamentally intersubjective or interpersonal in nature. This applies to the direct act of storytelling as well as to different forms of written narrative accounts.

Fourth, the contextualizing characteristic of stories, however, not only includes its spatial dimension, but also the temporal dimension. Stories situate experiences both in space and time. This is most obvious in (auto)biographical stories where experiences from one's personal life are being recalled (life stories) – thus (auto)biographical stories reveal how the narrator's understanding of the present is influenced by experiences from the past as well as his/her expectations about the future.

The use of stories and narratives is finally also valued for its emancipatory power. Narratives allow for the practitioner's voice to be heard and taken seriously in educational research and theory development. Yet, although acknowledging the emancipatory potential, several authors have warned for the trap of subjectivism or a romantic misconception of that emancipation. Creating a space for the voices that are often silenced in educational research is important, but it is equally crucial that these voices are always interpreted and understood as situated in and thus influenced by political, ideological, institutional, and structural contexts (Goodson, 1992).

Narratives, Teacher Thinking and Professional Identity

Within the overall interest in narratives and biography, several particular elements and evolutions have contributed to their appeal and relevance for teacher education.

First, right from 1980s, the developments in cognitive psychology have had an important influence on educational research. In line with these developments, educational researchers became convinced that in studying teaching as well as the training of future teachers, attention needs to be paid not only to performance skills, but also to the beliefs that guide (student) teachers' actions (Clark and Peterson, 1986). This so-called teacher thinking research stresses the importance of studying teachers' cognitions and the particular form, content, and development of the know-how they use in their work. This knowledge can partly be characterized as formal or systematic knowledge. This knowledge is based on research, for example, explicitly organized in, and provided by, the teacher education curricula. But apart from this formal knowledge, teaching involves other forms of knowing or understanding that are more informal, experiential, idiosyncratic in nature, and as such closely linked to the

individual's experiences during his/her life. The acknowledgment of both formal and informal types of professional knowledge, and forms of understanding in teacher professionalism, as well as the need for every student-teacher to integrate them in a personal way, have had far-reaching consequences for the pedagogy of teacher education and, more in particular, stimulated the use of narratives in teacher education.

Second, the acknowledgment of the experiential nature of (student) teachers' informal knowledge inevitably leads to a biographical perspective on learning to teach. Student teachers do not enter teacher education as blank sheets, but rather bring with them about 15 years of personal experiences in schools. They have lived and worked in many classrooms, experienced dozens of different teachers, and thus inevitably have built a very personal set of ideas and beliefs about what it means to be a teacher, to be teaching, and to be a student. This personal interpretative framework (Kelchtermans, 2007) is thus embedded in one's biography as a (student) teacher. At the same time, the framework is influenced by and develops through the learning that takes place during the teacher education program. But in its turn it also influences that learning process. It acts as an interpretative lens, mediates and filters the way student-teachers deal with and make sense of the curriculum of their teacher education program, interpret experiences during practical training, etc. It thus is both an outcome of and a condition for this learning.

Yet, because of its biographical, experiential, and largely idiosyncratic nature, this framework and the understandings it constructs may be one-sided, limited, or simply wrong. As such it may negatively bias student teachers' learning and sense-making of the formal curriculum and especially of the experiences during practical training (internships). Several authors have argued that, the often observed washing out effect of teacher education programs once teachers have started their career, is due to the fact that the implicit personal, interpretative lenses of the student teachers were not deliberately made explicit or discussed. In order to control possible biases and ensure a sustainable, in-depth integration of the curriculum by the student teachers, the curricula of teacher education programs have to aim at making this implicit knowledge explicit and bringing it up for critical scrutiny and discussion. Only this way the validity of student teachers' personal professional knowledge, as the basis for their professional decisions and actions, can become more valid and justified.

Narratives and storytelling can play a part in this process of making the implicit knowledge explicit. In order to do so, however, it is of crucial importance that work with narratives in teacher education goes beyond the telling and sharing of stories as such. Storytelling, therefore, has to be linked with forms of narrative inquiry,

in which the stories and the beliefs they reveal are critically analyzed, questioned, and re-storied (as is illustrated for example by the exercise on the critical incident vignette).

Third, the use of narratives in teacher education *per se* is also in line with and contributes to the emphasis on teachers' reflectivity. Reflection, both as an attitude and skill refers to people's capacity to look back on and think about themselves, their actions, and the situation they find themselves in. Because of the dynamic, highly unpredictable, and complex character of teaching and learning processes, reflectivity is a crucial tool for teachers to properly understand professional situations, choose appropriate actions, and learn from these experiences. Through reflection, teachers thoughtfully examine and integrate new knowledge and practical experiences, with deeply held beliefs and opinions. Narratives and different forms of storytelling have proven to be powerful tools to engage student teachers in reflection. As argued before, narratives are particularly powerful as triggers for reflection on the nontechnical aspects of teaching and learning to teach such as the emotional, ethical, and political aspects.

A final reason for the importance of narrative and biography in teacher education is the recognition that the person or the teacher is a constitutive part of the educational situation. Therefore, teachers' idea of themselves as teachers, their sense of identity is of much greater importance to them as practitioners than in occupations where the person can easily be separated from the craft. When speaking about their job, teachers demonstrate "persistent self-referentialism" (Nias, 1989): their self as a person is always implicated. Developing a sense of professional identity as a teacher constitutes a red thread for student-teachers' learning and development throughout their teacher education. (Auto)biographical reflection or narrative accounts allow for this self to be explicitly addressed.

See also: Self-Study by Teacher Educators.

Bibliography

- Bullough, R. V. (1991). Exploring personal teaching metaphors in preservice teacher education. *Journal of Teacher Education* 42, 43–51.
- Carter, K. (1993). The place of story in the study of teaching and teacher education. *Educational Researcher* 22(1), 5–18.
- Clark, C. and Peterson, P. (1986). Teachers' thought processes. In Wittrock, M. (ed.) *Handbook of Research on Teaching*, 3rd edn., pp 255–296. New York: Macmillan.
- Connelly, F. M. and Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational Researcher* 19(5), 2–14.
- Goodson, I. (ed.) (1992). *Studying Teachers' Lives*. London: Routledge.
- Hoeppel, R. (1983). 'Perspektiven der erziehungswissenschaftlichen Erschliessung autobiographischer Formen der Selbstreflexion' (*Perspectives for the Use of Autobiographical Self-reflection in Educational Research*). *Zeitschrift für Pädagogik* 18. Beiheft, pp 307–312. Weinheim Basel: Beltz.
- Kelchtermans, G. (2007). Professional commitment beyond contract. Teachers' self-understanding, vulnerability and reflection. In Butcher, J. and McDonald, L. (eds.) *Making a Difference: Challenges For Teachers, Teaching, and Teacher Education*, pp 35–53. Rotterdam: Sense Publishers.
- Labov, W. and Waletzky, J. (1973). Erzählanalyse. Mündliche Versionen persönlicher Erfahrung. In Ihwe, J. (ed.) *Literaturwissenschaft und Linguistik*, vol 2, pp 78–126. Frankfurt am Main: Athenäum.
- Nias, J. (1989). *Primary Teachers Talking. A Study of Teaching as Work*. London: Routledge.
- Polkinghorne, D. (1988). *Narrative Knowing and The Human Sciences*. Albany, NY: State University of New York Press.
- Sikes, P., Measor, L., and Woods, P. (1985). *Teacher Careers. Crises and Continuities*. London: Falmer.
- Bruner, J. (1986). *Actual Minds, Possible Worlds*. New York: Plenum.
- Carter, K. and Doyle, W. (1996). Personal narrative and life history in learning to teach. In Sikula, J., Buttery, Th. J., and Guyton, E. (eds.) *Handbook of Research on Teacher Education*, 2nd edn., pp 120–142. New York: Macmillan.
- Casey, K. (1995–1996). The new narrative research in education. *Review of Research in Education* 21, 211–253.
- Clandinin, J. (2007). *Handbook of Narrative Inquiry. Mapping a Methodology*. Thousand Oaks: Sage.
- Doyle, W. and Carter, K. (2003). Narrative and learning to teach: Implications for teacher-education curriculum. *Journal of Curriculum Studies* 35, 129–137.
- Gudmundsdottir, S. (1991). Story-maker, story-teller. Narrative structures in curriculum. *Journal of Curriculum Studies* 23, 207–218.
- Kelchtermans, G. (1993). Getting the story, understanding the lives. From career stories to teachers' professional development. *Teaching and Teacher Education* 9, 443–456.
- Krüger, H.-H. and Marotzki, W. (eds.) (1996). *Erziehungswissenschaftliche Biographieforschung*. Opladen: Leske-Budrich.
- Laboskey, V. K. and Cline, S. (2000). Behind the mirror: Inquiry-based storytelling in teacher education. *Reflective Practice* 1, 359–375.
- Loughran, J. and Russell, T. (eds.) (1997). *Teaching about Teaching: Purpose, Passion and Pedagogy in Teacher Education*. London: Falmer.
- Loughran, J., Hamilton, M. L., LaBoskey, K. V., and Russell, T. (eds.) (2004). *International Handbook of Self-Study of Teaching and Teacher Education Practices*. Dordrecht: Kluwer Academic Publishers.
- Personal Narratives Group (1989). *Interpreting Women's Lives: Feminist Theory and Personal Narratives*. Bloomington, IN: Indiana University Press.
- Schön, D. (1983). *The Reflective Practitioner. How Professionals Think in Action*. London: Temple Smith.
- Sparkes, A. (1994). Life histories and the issue of voice: Reflections on an emerging relationship. *International Journal of Qualitative Studies in Education* 7, 165–183.
- Trotman, J. and Kerr, T. (2001). Making the personal professional: Pre-service teacher education and personal histories. *Teachers and Teaching: Theory and Practice* 7, 157–171.

Supervision in Teacher Education

J McCarthy and L F Quinn, University of Nevada, Las Vegas, NV, USA

© 2010 Elsevier Ltd. All rights reserved.

Supervision in Teacher Education

Supervision in teacher education is meant to assist teacher education candidates preparing to enter the profession for the first time and to assist practicing teachers in the development of the skills necessary for effective teaching. Supervision is usually carried out through three separate processes: (1) observation, (2) analysis, and (3) discussion or sharing information collected and organized by the supervisor with the teacher education candidate or teacher. Teachers in the field often comment that they learned how to teach by practicing the act of teaching in actual classroom situations with a skillful supervisor to guide them and that student teaching is often viewed as the most useful part of the teacher education process (Roth, 1983). John Dewey (1963) speaks of experiences that are educative, and says that the direction in which growth takes place must be specified and indicated. Experience alone is not always a good teacher and teacher education programs around the world have implemented some form of supervision as a part of teacher preparation.

Approaches to supervision have been described through various terms (some philosophical and some practical) as: essentialism, experimentalism, and existentialism (Glickman, 1990); positivism, phenomenology, and critical theory (May and Zimpher, 1986); and learner-centered (Paris and Gespass, 2001). Whatever the name given to supervision, observing, analyzing, and discussing the act of teaching is viewed as an effective way to improve teaching performance. Additionally, government departments of education exert a strong influence on supervision in teacher education by setting requirements for the amount of clinical practice a teacher must complete in order to be licensed.

Historical Perspectives on Supervision in Teacher Education

Supervision in teacher education is not a new component of preparing teachers or providing professional development opportunities for teachers already in the field. While reform efforts to improve the conditions of teacher education have proliferated in the past 40 years, the importance of a well-trained supervisor to facilitate the growth of a novice or an experienced teacher remains as central to the process today as it did nearly a century ago. Bulletin #14, *The professional preparation of teachers for American public schools*, a report issued by the Carnegie

Foundation for the Advancement of Teaching in 1920, recommended that teacher education programs be clinically based, with ongoing opportunities for observation and supervised practice in real world settings (Imig and Imig, 2006).

A little more than 50 years after the Bulletin #14 report, the AACTE Commission on Education for the Profession of Teaching determined that educators must develop teaching skills through practice in a controlled training environment accompanied by effective direction and supervision. The effective use of both laboratory and field-centered approaches has long been viewed as crucial to the development of the complex competencies of teaching.

Early schools were often staffed by low-paid individuals with little formal education. The profession of teaching had yet to develop and supervising teachers was often left to local officials whose task was to visit schools and make recommendations for their improvement (Fuller, 1982). Supervisors may also appear in the form of educated individuals or a visiting committee whose mission was to see that the teaching was competent and the learning adequate. Even in the early stages of supervision in teacher education, it was understood that supervisors would be objective, draw conclusions based on the evidence, and give balanced feedback. An increase in the number of teachers required to meet the needs of the ever-growing number of students and schools resulted in more emphasis being placed on teacher preparation and supervision.

Normal schools of the nineteenth century in the United States (US) gave detailed attention to supervised practice that followed a significant amount of time spent observing in classrooms (Clifford and Guthrie, 1988). Feiman-Nemser (1990) reports that James Earl Russell, dean of Teachers College (1894–1927), described a model laboratory school where exemplary teachers would oversee the training of novices. In the mid-1960s, the Harvard University Masters in Teaching Program initiated a clinical supervision process in which the teacher education candidates and teachers helped determine the focus of their own supervision (Goldhammer, 1969).

We find similar stages of development in teacher preparation programs with regard to supervision practices around the world. There are numerous efforts underway at this time to standardize teacher education programs. The Bologna Accord (1999) in Europe and the Report of the House of Representatives Standing Committee

(www.gmacbolognaproject.com/docs/BolognaCheatSheet-2005) on Education and Vocational Training in Australia (Hartsuyker, 2007), as well as several national committees and commissions in the US (including a commission sponsored by the National Academy of Education) bear testament to this movement. The European Commission (2007) has also recently put forth proposals aimed at improving the quality of teacher education in member states and standardizing the guidelines. All of these reports and commissions have implications for the future supervision of teacher candidates.

Supervision in Constructing a Professional Identity

The practice of teaching and learning to teach are so inextricably connected that attempting to help someone learn to teach without actually critiquing their attempts at teaching seems pointless. Interactions between experts and novices in teacher education that focus on instruction have been the subject of numerous research studies. When the emphasis in supervision is on helping teachers develop skills in delivering instruction, helping the same teachers construct a professional identity may become secondary. Recent research on supervision has focused on teacher participation in their own development as teachers through collaboration with colleagues, peer feedback, reflective practice, and action research (Glickman, 1990; Holland *et al.*, 1992; Yusko, 2004; Shin *et al.*, 2006). One view of supervision as described by Glickman (1990) sees “the focus of supervision as a process of eliciting the teacher’s own thinking and planning.” (p. 550).

Constructing a professional identity cannot occur in a vacuum. This process requires that quality leaders assist novice teachers in their journey towards a professional identity. The complex act of the construction of a professional identity by teacher education candidates during the transition from university setting to a school context is examined through a collaborative partnership utilizing Internet discourse among all participants (Dauvisis *et al.*, 2005; The Association of Teacher Educators’ Commission on Quality Leaders for Novice Teachers, 2006). Dubar’s (1996) description of professional identity development as the simultaneously stable and provisional; individual and collective; subjective and objective; and biographic and structural result of different processes of socialization that build individuals and define institutions encompasses the many variables associated with learning to teach in a social environment.

Research on Supervision

Research on supervision in teacher education has demonstrated that supervisory observations alone are not sufficient to fully interpret and critique the experiences of

teacher education candidates and teachers. With this in mind, institutions of higher education around the world plan and implement strategies for improving supervisory practices in teacher education. In the University of Montreal’s 4-year teacher preparation program, at least 700 hours in a school setting are required. Supervisors and teacher education candidates in this program are taught the processes of practical argument (Fenstermacher, 1987), shared reflective practice (Tochon, 1996), and use of videotapes to gain a greater awareness of the relationship between feedback and action in the supervisory process (Gervais, 2005). The benefits of paired-placements and the power of communication within particular contexts to improve teacher education practice in collaborative settings are examined in detail by Sorensen (2005) at the University of Nottingham.

The use of specific tasks to help preservice teachers in Santiago, Chile identify, interpret, and analyze professional performance within a specific institutional culture is discussed by Rittershaussen *et al.* (2005). This study examines the knowledge constructed by preservice teachers regarding professional performance and the challenges they perceive in becoming members of a professional context through dialog with supervisors. A study by Chan *et al.* (2007) looked at the preconceptions of teaching and learning held by preservice teachers in Singapore. Another study by de Leon-Carillo looked at the preconceptions of teacher roles held by preservice teachers in the Philippines. The preconceptions that students bring with them to the teacher preparation program can serve as guides for the field experiences they need to have in order to become effective professionals. Other studies, including one conducted by Ballard (2002), have shown that student teacher beliefs are greatly influenced by those of the cooperating teachers with whom they work. These findings indicate the need for careful selection and training of cooperating teachers.

Supervision and Induction

Supervision is a natural by-product of induction programs. Induction programs are designed to give new or experienced teachers support and assistance rather than simply assessing their work. As part of induction programs, mentors, that is, teachers with a record of success in their own classrooms, are assigned to guide other teachers in the acquisition of the skills of teaching and the professional tools necessary for a teaching career. Induction programs often include a form of clinical supervision for teachers who face trouble in the classroom.

The school administrator plays an important role in establishing an environment that facilitates the supervision and the work of mentors. Experienced teachers are asked by their administrators to serve as mentors or cooperating teachers. Some teachers assume the mentor role eagerly as another step in their own professional

development. Others may be less than enthusiastic. Whichever way teachers approach the task of supervision, they have firm opinions about what conditions must be present for teachers to be successful. Many perceive the school administrator to be responsible for establishing an environment that supports both supervision and the mentoring process and that assists experienced teachers in their role as a professional guide. The support that cooperating teachers and mentors are able to offer teachers is strongly influenced by physical considerations such as the proximity of classrooms, the schedules, and the teaching assignments, all of which are ultimately controlled by the administrators (Quinn, 1994).

The availability of such induction support and mentoring varies greatly among teacher education contexts, ranging from nonexistent to outstanding. There is a growing body of research to draw on regarding the value of such support for the beginning teacher. Darling-Hammond (2006) cites numerous studies that show that teachers who participate in a supervised student teaching experience are more likely to stay in the profession, as are those who participate in mentoring programs during their first 2 years of teaching (p.339).

The Supervision Triad

The activities and processes of supervision in teacher education normally occur in a triad (teacher education candidate or practicing teacher; cooperating teacher or mentor; and university supervisor or administrator). While variations on the degree of involvement and classification of members of the triad do exist, the teacher education candidate or teacher is the purpose for the supervision, while another, more experienced teacher takes responsibility for observing and critiquing the necessary teaching activities. A university or college supervisor, whose responsibility is to guarantee that requirements of the institution are fulfilled, works closely with all other members of the triad as well as the school administrator to ensure that the supervision is being conducted in a purposeful and supportive manner. This triad formation is common to most, if not all, existing teacher education programs.

It is expected that the supervision triad will create an effective learning environment for the development of practical knowledge, skills, and dispositions relevant to the professional growth of teachers. Through completion of specific assignments such as writing and implementing lessons for small groups or an entire class of students during field-based experiences, the teacher education candidate observes the complexities of classroom life and is able to analyze these observations through discussion with an experienced cooperating teacher and a university supervisor. Supervisory discussions encourage teacher education candidates and teachers seeking

professional growth to reflect upon their instructional practices. Such discussions help them recognize the elements necessary to prepare their own classroom learning environments to accommodate the individual needs of learners.

The purpose of supervision, whether it is to improve teaching behaviors, to increase teacher reflection and critical analysis, to increase skill in using a particular teaching strategy, or to improve general satisfaction with teaching, must be clear among all members of the supervisory triad for its greatest effect to be achieved.

Participants in a supervision triad

Teacher education candidate

The teacher education candidate or teacher seeking professional support is the purpose for supervision. Their role is to demonstrate knowledge, skills, and performance that meet expected standards in the areas of planning; instruction and management of the classroom environment; and management of the activities and behavior of students. In addition, candidates and teachers are expected to comply with all school and district policies.

The cooperating teacher

The cooperating teacher is often viewed as the single most important influence by teacher education candidates in the process of learning to teach. Costa and Garmston (1987) list three contributions that a cooperating teacher or mentor can offer to others. They model best practices, they pass along the tools of the trade, and they encourage development of the intellectual process of teaching in others by articulating their actions and thought processes. Effective cooperating teachers and mentors become what Costa and Garmston (1994) refer to as cognitive coaches.

At some point in the interaction with a practicum student, student teacher, intern, or beginning teacher, the cooperating teacher or mentor will have to give criticism or offer advice. Henry and Beasley (1989) offer suggestions for interactive practices by cooperating teachers and mentors that lead to effective supervisory experiences. Among these suggestions, clear and specific communication and feedback, modeling espoused behaviors, consistency, offering rationale for actions and suggestions, and using problem-solving strategies to guide professional growth are considered as most supportive of supervisory interactions.

The university supervisor

The main role of the university supervisor in the supervision triad is to see that institutional requirements are completed by the teacher education candidate. The university supervisor must often assume roles other than that of observing teaching practices and providing constructive analysis of said practices. The university supervisor may find it necessary to facilitate relationships among

teachers, to serve as a personal confidant when necessary, to provide individualized instruction, and to provide guidance in areas that may have been overlooked by other participants in the triad (Zimpher, 1980). The university supervisor may also serve as a seminar leader and a placement consultant.

All participants in the supervision triad need to possess certain personal characteristics that help to prevent disharmony and discord. The characteristics most relevant to the supervision triad focus on interpersonal and communication skills. Before supervision in teacher education takes place, agreements must be established and guidelines set. It is imperative that all members of the triad become familiar with the supervisory process to be followed and be aware of the potential outcomes of the process. Supervision styles should accommodate differing stages of teacher development. Seminal development of frameworks for classifying supervisory behaviors and conferences were offered by Amidon and Flanders (1967) and Spaulding (1967). Variations on these frameworks have produced a variety of ways to conduct and assess supervision in teacher education.

Changing Perspectives on Supervision in Teacher Education

Supervision for social justice

Diversity in the school-age population is a major concern for educators in many nations. Changing demographics and increasing diversity have presented challenges for the majority of teacher education programs. Many nations have expressed concern that the needs of all students are not being met in schools as they are currently staffed with teachers who are not knowledgeable on working with diverse student populations. Some programs are trying hard to provide preservice teachers with the knowledge, skills, and experiences they need in order to provide all the students with quality education. The challenges are many.

Han and Singh (2007) report on the difficulties in the recruitment and retention of student teachers and teachers from minority and immigrant groups in Australia. There is a mismatch between the type and length of the field experiences needed by immigrant education students who have had no experience with the culture of Australian schools and the generic model of field experiences offered by teacher preparation programs. The minority preservice teachers believe that their supervisors do not understand their needs. As a result, the increasing diversity of Australian students is not being met with an increase in teacher candidates from other ethnic or linguistic groups.

Teacher education programs are currently seeking answers to questions related to preparing high-quality teachers for all student populations. Some of these questions are: (1) How can preservice teachers learn to

work with the strengths that diverse students bring with them to school rather than searching for reasons why they are deficient according to the teacher's values and expectations? (2) How can supervisors facilitate the development of teachers who can be effective with all types of learners in a diverse classroom?, and (3) How can clinical supervision help preservice teachers provide a level playing field for all members of the classroom group? The answers to these dilemmas are critical at this time of increasing globalization.

One elementary school in the southwestern US has more than 80 flags hanging in the lunchroom representing the homelands of the students in the school. The minority has become the majority in the Clark County School District. Yet the teaching force is predominantly Anglo and female. Teacher educators in many nations, such as England, Ireland, and the US to name a few, are grappling with these issues. Given the current political realities, supervisors and teachers alone cannot alter the reality of poverty, class, and language. What they are trying to do is equip preservice teachers with experiences that allow them to work in culturally diverse classrooms and begin to reflect on what skills and knowledge they will need to provide equitable learning experiences for all (Jacobs, 2006).

Alternative routes to licensure

There appear to be differences in the types and models of supervision provided for students in university-based teacher education programs and for those who come in through an alternative program or the normal school route. This seems to hold true globally. Some alternative programs shorten the field experience segment of the program. Some students do not have clinical experience until their last semester, or student teaching experience. The person actually supervising may differ as well. In some alternative teacher education programs there are no university supervisors, only cooperating teachers.

Some programs in the US put college graduates immediately into a classroom before they have any pedagogy coursework. These novice teachers take classes while on the job and are assigned mentor teachers to work with them while they are teaching in their own K-12 classrooms. If the quality of the mentoring is good, many of these new teachers can do quite well.

Other licensure programs that consist mainly of online coursework have only limited fieldwork. Their students have no preliminary practice or observations in classrooms, but hit the ground running during their student teaching experience. Again, the quality of the supervision and mentoring plays a big role in the success or failure of these teachers. We can expect to see more alternative routes to licensure developing around the world as large numbers of existing teachers retire and the populations in many countries continue to grow.

School university partnerships

A unique response to the challenge of university-based teacher education programs working closely with public schools has been that of the professional development school (PDS). In some places, these are called partner schools. In the US, PDSs were initiated in response to a recommendation of the Holmes Group, an organization of College of Education deans (Darling-Hammond, 2006). The purpose of the PDSs was to create partner schools that worked closely with teacher education faculty to create an environment conducive to learning to teach all children. Public school teachers were viewed as teacher educators and provided supervision of clinical experiences. Sometimes, the experience of the teacher education candidates in PDSs is modeled on medical education and preservice teachers actually do rounds similar to the practices of medical students in hospital rotations. Such arrangements require expert supervision. Teachers at PDSs often engage in graduate coursework in supervisory practices.

Variations on a Traditional Model of Supervision in Teacher Education

Teacher education programs around the world are experimenting with new models of supervision. These models are sometimes driven by contextual variables and sometimes by sound pedagogical theory. For example, one program in the US is trying paired student teacher placements similar to other experiments. They have grounded their program in Vygotsky's zone of proximal development and are trying to provide the teacher education candidates with peer support as they progress through the developmental levels of learning to teach (Baker and Milner, 2006). The newly adopted teacher education guidelines in Ethiopia call for five consecutive practicum courses in partner schools from the beginning of the program (Degago, 2007). One program in rural Australia has facilitated the placement of future teachers in remote rural schools by implementing a support system that incorporates face-to-face mentoring with innovative use of communication technology (see more about technology in the next section.). Some programs in European nations require 1 or 2 full years of clinical experience before licensure (Cobb, 1999). Others in some developing countries still rely on the normal school license rather than a degree program. Students in these programs rely on the teachers in the schools for supervision and mentoring.

Supervisory Assessment of Field Experiences

This is an age of teacher accountability for student learning. Thus, many supervisors look for evidence that learning is taking place in the classroom of preservice

teachers. Some programs use formalized data such as pre- and post-test scores. Others have students reflect, in their portfolios, on the effectiveness of the lessons they taught and an analysis of what worked and what did not work and why. Many accreditation bodies are demanding this sort of evidence before licensure of new teachers and many state and national legislatures are also demanding this type of evidence of teacher effectiveness at the preservice level.

The Use of Technology in Supervision

The age of technology has influenced methods of supervision in teacher education. A variety of technological tools exist that can support supervision and student self-assessment in teacher education. Some teacher preparation programs provide preservice teachers with laptops to use for journaling their daily experiences in the classroom. Students can also utilize their laptops to communicate with their supervisors and peers on a regular basis, thus creating a community of learners. Electronic communication is especially helpful for students doing field work in remote or rural areas, but can be used by all preservice teachers. Some forms of supervision make use of videocams and electronic recording devices for students to utilize in the reflection and analysis of their teaching efforts. Technology can be seen as a tool to improve the quality of student practice and teacher mentoring. In the information age driven by technology, supervision in teacher education may become more self-directed than that in the past, as novices and experienced teachers alike turn to the Internet for advice and examples.

Techniques for using technology in reflection/self-analysis

Technology has also provided preservice teachers with greater opportunity than ever before for self-analysis and reflection of their work. Students can even download demonstration teaching classes and watch expert teachers teach in real classrooms. They then can analyze their own teaching based on what they saw. Many teacher preparation programs are teaching their students the evaluation systems that will be used to assess their performance when they are teaching full time. Some programs are even using these systems to assess candidate skill and expertise, so as to better prepare the teacher for the expectations of the workplace. Many of these assessments are online and students can get immediate feedback from their supervisors.

Portfolios have been used in teacher preparation programs for some time. Now, however, students are preparing electronic portfolios of their culminating field experiences and are including videoclips of their teaching. Not only are these portfolios used by supervisors to

mentor future teachers, but they can also be used as part of an application package when the candidate applies for professional employment.

Online supervision in teacher education

Online supervision in teacher education programs does exist, but even with online teacher education programs there is normally a full-time, supervised, in-classroom experience with a clinical supervisor to observe and evaluate performance based on acceptable professional standards. However, the use of webcams and other digital imaging in computers makes it easy to demonstrate teaching practice for someone not present in a specific classroom. Teacher education programs have long used videotape recordings of teaching practice as a form of self-analysis and discourse between supervisors and teachers. While the face-to-face interactions between supervisors and teachers have been the norm, it is possible that technologies to aid online interactions will influence changes in supervision in teacher education.

The Future of Supervision in Teacher Education

It is likely that supervision in teacher education will continue to be a critical element in learning to teach and to improve teaching skills. Presentations at international and national conferences attest the interest of individuals in creating meaningful supervisory practices that result in effective teaching. At the 2007 meeting of the Association of Teacher Educators held in San Diego, California, Dr. Fedotova, Director of the International Information Center in Russia, discussed the challenges of Russian education under the process of globalization and the world educational community. At the 2007 Annual Meeting of the American Educational Research Association, Dr. Ping Deters, from the University of Toronto, presented a paper on internationally educated teachers. As research is shared among members of the global community of teacher educators, knowledge of what works in supervision will traverse international boundaries, become refined and adapted to individualistic contexts, and will continue to expand in concept and practice.

Summary

Supervision in teacher education is an issue that is of major concern to educators all over the world. The articles discussed elsewhere in this encyclopedia will further refine and define the major issues raised in this article. In this age of teacher accountability, teacher educators are searching for effective and efficient ways to provide experiences that will prepare teachers to facilitate equitable opportunities for all students to learn and achieve success. While individual national contexts will

influence practice, there are many commonalities with regard to the challenges and implementation of effective supervision.

Bibliography

- Amidon, E. and Flanders, N. (1967). Interaction analysis as a feedback system. In Amidon, E. and Hough, J. (eds.) *Interaction Analysis: Theory, Research, and Application*, pp 121–140. MA: Addison Wesley.
- Association of Teacher Educators (2006). *Special Panel 132. State of the ATE Commission of Quality Leaders for Novice Teachers*. Philadelphia.
- Baker, R. S. and Milner, J. O. (2006). Complexities of collaboration: Intensity of mentors' responses to paired and single student teachers. *Action in Teacher Education* 28(3), 61–72.
- Ballard, B. (2003). Preservice Teachers' Beliefs about Classroom Management Before and After Student Teaching. Unpublished doctoral dissertation, University of Nevada.
- Chan, K., Tan, J., and Khoo, A. (2007). Pre-service teachers' conceptualizations about teaching and learning: A closer look at Singapore cultural context. *Asia-Pacific Journal of Teacher Education* 35(2), 181–195.
- Clifford, G. and Guthrie, L. (1988). *Ed. School: A Brief for Professional Education*. Chicago, IL: University of Chicago Press.
- Cobb, V. (1999). *An International Comparison of Teacher Education*. (ERIC Digest). Washington, DC: ERIC Clearinghouse on Teaching and Teacher Education.
- Costa, A. L. and Garmston, R. J. (1994). Student teaching: Developing images of a profession. *Action in Teacher Education* 9, 5–11.
- Darling-Hammond, L. (2006). *Powerful Teacher Education*. San Francisco, CA: Jossey-Bass.
- Dauvisis, M., Kaddouri, M., and Vandroz, D. (2005). *A Project with Educational Intention: Methods of Acculturation and Development of Professional Competences*. Paper presented at the American Educational Research Association Annual Meeting, Montreal.
- Degago, A. T. (2007). A first-timer's impressions of engaging in action research: A case study in Ethiopian pre-service teacher education. *Action in Teacher Education* 29(1), 71–80.
- Dubar, C. (1996) *The Formation of Professional Identity*. La Decouverte, Reperes, Bricolage.
- The European Commission Press Release August 6, 2007. Teachers need good education too! The Commission proposes to improve the quality of teacher education in the European Union.
- Europa Press Release. August 6, 2007. Teachers good education too! The Commission proposes to improve the quality of teacher education in the European Union. Retrieved 12/2/07.
- Feiman-Nemser, S. (1990). Teacher preparation: Structural and conceptual alternatives. In Robert, H. W. (ed.) *The Handbook of Research on Teacher Education*, pp 212–233. New York: Macmillan.
- Fenstermacher, G. D. (1987). On understanding the connections between classroom research and teacher change. *Theory into practice* 26(1), 3–7.
- Fuller, W. E. (1982). *The Old Country School: The Story of Rural Education in the Middle West*. Chicago, IL: University of Chicago Press.
- Gervais, C., Desrosiers, P., Lepage, M., and Correa Molina, E. (2005). *Explication of Pedagogical Knowledge of Experienced Teachers to Student Teachers in Preservice Training*. Paper presented at the American Education Research Association Annual Meeting, Montreal.
- Glickman, C. (1990). *Supervision of Instruction, A Developmental Approach*. Newton, MA: Allyn and Bacon.
- Goldhammer, R. (1969). *Clinical Supervision: Special Methods for the Supervision of Teachers*. New York: Holt, Rinehart, and Winston.
- Han, J. and Singh, M. (2007). Getting world English speaking student teachers to the top of the class: Making hope for ethno-cultural diversity in teacher education robust. *Asia-Pacific Journal of Teacher Education* 35(3), 291–309.
- Hartsuyker, L. (House of Representatives Standing Committee on Education and Vocational Training.) (2007). *Top of the class: Report of the inquiry into teacher education*. Canberra, ACT: The Parliament of the Commonwealth of Australia.

- Henry, M. A. and Beasley, W. W. (1989). *Supervising Student Teachers: The Professional Way*, 4th edn. Terre Haute, IN: Sycamore Press, Inc.
- Holland, P. E., Clift, R., Veal, M. L., Johnson, M., and McCarthy, J. (1992). Linking preservice and inservice supervision through professional inquiry. In Glickman, C. D. (ed.) *Supervision in Transition: 1992 Yearbook of the Association for Supervision and Curriculum Development*, pp 169–182. Reston, VA: ASCD Publications.
- Imig, D. G. and Imig, S. R. (2006). The teacher effectiveness movement: How 80 years of essentialist control have shaped the teacher education profession. *Journal of Teacher Education* **57**(2), 167–180.
- Jacobs, J. (2006). Supervision for social justice: Supporting critical reflection. *Teacher Education Quarterly*, pp 23–39 Fall, 2006.
- May, W. T. and Zimpher, N. L. (1986). An examination of three theoretical perspectives on supervision. Perceptions of preservice field supervision. *Journal of Curriculum and Supervision* **1**(2), 83–99.
- Paris, C. and Gespass, S. (2001). Examining the mismatch between learner-centered teaching and teacher-centered supervision. *Journal of Teacher Education* **52**(5), 398–412.
- Quinn, L. F. (1994). The importance of structure in providing uniform quality in mentoring/induction programs. *Mentoring*, **2**(1), 5–12.
- Rittershaussen, S., Suzuki, E., Montecino, C., and Solis, M. C. (2005). *Identifying Professional Performance and Institutional Culture through Specific Tasks*. Paper presented at the American Educational Research Association Annual Meeting, Montreal.
- Roth, R. A. (1983). *The status of the profession: Selected characteristics of teacher education and teaching*. (ERIC Document Reproduction Service No. ED 233 988).
- Shin, E., Wilkins, E. A., and Ainsworth, J. (2006). The nature and effectiveness of peer feedback during an early clinical experience in an elementary education program. *Action in Teacher Education* **28**(4), 40–52.
- Sorenson, P. (2005). *Peer Learning in the Practicum: The Use of Collaboration and Dialogue in Preservice Teacher Development*. Paper presented at the American Educational Research Association Annual Meeting, Montreal.
- Spaulding, R. L. (1967). A coping analysis schedule for educational settings (CASES). In Anita, S. E. and Gil, B. E. (eds.) *Mirrors for behavior*, pp 587–589. Philadelphia, PA: Research for Better Schools.
- Yusko, B. P. (2004). Caring communities as tools for learner-centered supervision. *Teacher Education Quarterly* Summer, 53–72.
- Zimpher, N. L. (1980). A closer look at university student teacher supervision. *Journal of Teacher Education*, 31 (Jul-Aug): 11–15.
- Acheson, K. A. and Gall, M. D. (2002). *Clinical Supervision and Teacher Development: Preservice and Inservice Applications*, 5th edn. Fresno, CA: Wiley.
- Dangel, J. R. (2006). *Research on Teacher Induction: Teacher Education Yearbook XIV*. Lanham, MD: Rowman and Littlefield Education.
- Darling-Hammond, L. and Bransford, J. (eds.) (2005). *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do*. San Francisco, CA: Jossey-Bass.
- de Leon-Carillo, C. (2007). Filipino pre-service education students' preconceptions of teacher roles viewed through a metaphorical lens. *Asia-Pacific Journal of Teacher Education* **35**(2), 197–217.
- Gray, J. (ed.) (2006). *Making Teaching Public: Reforms in Teacher Education. Proceedings of the 2006 Australia Teacher Education Conference*. Perth, WA: ATEA.
- Harlin, R. P. (2000). Developing reflection and teaching through peer coaching. *Focus on Teacher Education* **1**(1), 1–10.
- Henry, M. A., Beasley, W. W., and Brighton, K. L. (2002). *Supervising Student Teachers*, 6th edn. Terre Haute, IN: Sycamore Press, Inc.
- Knowles, J. G., Cole, A. L., and Presswood, C. S. (1994). *Through Preservice Teachers' Eyes: Exploring Field Experiences through Narrative and Inquiry*. New York: Merrill.
- Little, J. W. (1982). Norms of collegiality and experimentation: Workplace conditions of school success. *American Educational Research Journal* **19**(3), 325–340.
- McInerney, P. (2007). From naïve optimism to robust hope: Sustaining a commitment to social justice in schools and teacher education in neoliberal times. *Asia-Pacific Journal of Teacher Education* **35**(3), 257–272.
- McIntyre, D. J. and Byrd, D. M. (eds.) (1996). *Preparing Tomorrow's Teachers: The Field Experience*. Thousand Oaks, CA: Corwin Press.
- Moore, R. (2006). Taking action: Assessing the impact of preservice learning on teaching. *Action in Teacher Education* **28**(3), 53–60.
- Nelligan, P. (2006). Rural professional experiences supporting preservice teachers. In Gray, J. (ed.) *Making Teaching Public: Reforms in Teacher Education. Proceedings of the 2006 Australia Teacher Education Conference*. Perth, WA: ATEA.
- Newhouse, P. (2006). Mobile education devices for preservice teachers. In Gray, J. (ed.) *Making Teaching Public: Reforms in Teacher Education. Proceedings of the 2006 Australia Teacher Education Conference*. Perth, WA: ATEA.
- Oliva, P. F. (1988). *Supervision for Today's Schools*, 3rd edn. New York: Longman.
- Stella, C. S. C., Forlin, C., and Lan, A. M. (2007). The influence of an inclusive education course on attitude change of pre-service secondary teachers in Hong Kong. *Asia-Pacific Journal of Teacher Education* **35**(2), 161–179.
- Yee, A. (1969). Do cooperating teachers influence the attitudes of student teachers? *Journal of Educational Psychology* **60**(4), 327–332.

Further Reading

- Acheson, K. A. and Gall, M. (1996). *Techniques in the Clinical Supervision of Teachers: Preservice and Inservice Applications*, 4th edn. Fresno, CA: Wiley.

Taking Prospective Teachers' Beliefs into Account in Teacher Education

M Valcke, G Sang, I Rots and R Hermans, Ghent University, Ghent, Belgium

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Beliefs – Beliefs are psychological understandings, premises, or propositions that are felt to be true.

Educational beliefs – These refer to beliefs about the role of the teacher and/or student in instructional processes. In the literature, a bipolar distinction is made between teacher-centered traditionalistic and more progressive or student-centered educational beliefs.

Expectancy value theory – According to this motivation theory, beliefs are a central construct that influences what teachers value and expect, and how this influences their actual teaching behavior.

Mathematics beliefs – Beliefs that refer to conceptions about the nature of mathematics, the nature of mathematics teaching, and the process of learning mathematics.

Teacher beliefs – These are psychological understandings, premises, or propositions about processes, variables, and actors that are central to learning and instruction settings, such as educational beliefs, epistemological beliefs, beliefs about inclusive education, etc.

Introduction

Teachers' beliefs have become a crucial variable in studying teaching behavior and teacher education. From the 1980s to the present, researchers have investigated both explicit and implicit knowledge and beliefs of preservice, novice, and experienced teachers to identify these beliefs and to examine how knowledge and beliefs affect learning to teach. Teacher beliefs are central to recent theories about teachers' identity. In their review about teachers' professional identity, Beijaard *et al.* (2004) state that beliefs are conceived as important constituents of teachers' professional identity formation. This implies that beliefs are of vital importance for teacher education since it determines the way the student teachers develop their meaning making and decision making. This suggests that beliefs are part of a complex set of variables and processes that determine teaching

behavior. The complex nature of teachers' beliefs implies that teacher education is critical to consider the development of a consistent belief system in teachers and student teachers.

Defining Teacher Beliefs

The concept of beliefs has a long history and has been defined in a variety of ways. Consequently, a conceptual confusion can be observed since authors easily adopt alternative terms. In addition, "some researchers refer instead to teachers' 'principles of practice', 'personal epistemologies', 'perspectives', 'practical knowledge', or 'orientations'" (Kagan, 1992, p. 66). It is therefore not surprising that Pajares (1992, p. 307) considered teacher beliefs as being a: "messy construct," noting that "the difficulty in studying teachers' beliefs has been caused by definitional problems, poor conceptualizations, and differing understandings of beliefs and belief structures."

Richardson (1996) defines beliefs as psychological understandings, premises, or propositions that are felt to be true. As such, beliefs can be represented as estimates of the likelihood that the knowledge someone has about a proposition or a subjective experience is correct. Alternatively, beliefs also refer to the likelihood that an event or state of affairs will occur (Wyer and Albarracín, 2005). Furthermore, as beliefs are clustered as a set of interrelated beliefs in a broader, general belief structure or system, they can vary in strength. The more a belief is interrelated with others in this structure, the more difficult it is to alter the belief in question (Pajares, 1992).

Since we focus on beliefs in the context of preservice teachers, the question is often raised with reference to the relationship between beliefs, attitudes, and knowledge. With reference to the relationship with attitudes, a distinction is made between an affective, a cognitive, and a conative component (behavioral or action part). Presently, beliefs are considered as representing the cognitive part of attitudes. The difference between beliefs and knowledge is presently also clearer. A differentiation commonly made is that beliefs are based on judgments and evaluations (subjective probability), whereas knowledge refers to objective verifiable facts.

Teacher Beliefs in Interaction with Other Processes and Variables

In the literature, the attention paid to teachers' beliefs cannot be isolated from the attention paid to other and clearly related constructs. From the mid-1980s, educational researchers began to focus on nonbehavioral components of teaching behavior, such as: (1) teachers' beliefs about classroom, students, school, and learning; (2) teachers' decisions for designing and presenting a teaching activity; (3) teachers' perceptions on classroom-teaching affairs; and (4) teachers' roles and their self-images (Kagan, 1995). This observation is important since it positions beliefs in the complex overall setting of teachers and their context. It is helpful to study the central position of beliefs as a motivational construct. In this context, it is most useful to reiterate the expectancy-value perspective on motivation as it has been expanded by Wigfield and Eccles (2000) and to rephrase it in terms of teacher-related processes. We stress the importance of this broader orientation toward beliefs in view of the consequences that can be derived for teacher-education practices. The model implies that beliefs consist of affective components, goal orientations, competency judgments, and perceptions about the teaching tasks to be carried out. In addition, the model points out that beliefs influence what students or teachers value. It helps to understand how beliefs are related to teaching expectations and how this results in choices, persistence, duration, and engagement in teaching activities. In addition, the model also stresses the fact that beliefs are influenced by cognitive processes, such as perceptions of the social context and interpretations of attributions related to earlier teaching experiences and

incidents. Moreover, the model stresses the impact of the external context. The relationships depicted in the model help to understand how the outcomes of teaching performance affect the extent to which future behavior will be valued, how beliefs are affected, and what expectations do result from these experiences. The model helps to understand why teacher's beliefs tend to be associated with a congruent style of teaching. The feedback loop in the model is of critical importance. The feedback loop gives teacher educators and trainers the opportunity to influence beliefs. **Figure 1** illustrates how the beliefs are part of this complex interplay with other variables and processes.

Positioning teachers' beliefs in this wider setting of mechanisms that influence actual teaching behavior also helps to make clear that beliefs and belief systems serve as personal guides in helping individuals to define and understand the world and themselves (Pajares, 1992). Teachers' educational beliefs are understandings, premises, or propositions about education, established through multitudinous experiences (Pajares, 1992). Teachers' beliefs are considered as relatively stable and act as a filter through which new knowledge and experiences are screened for meaning (Kagan, 1992) and that underlie teachers' planning, decision making, and behavior in the classroom.

Beliefs about What?

In addition, the model presented above makes it clear that the focus on the beliefs could be as varied as the teaching profession itself and therefore reflects on the issues

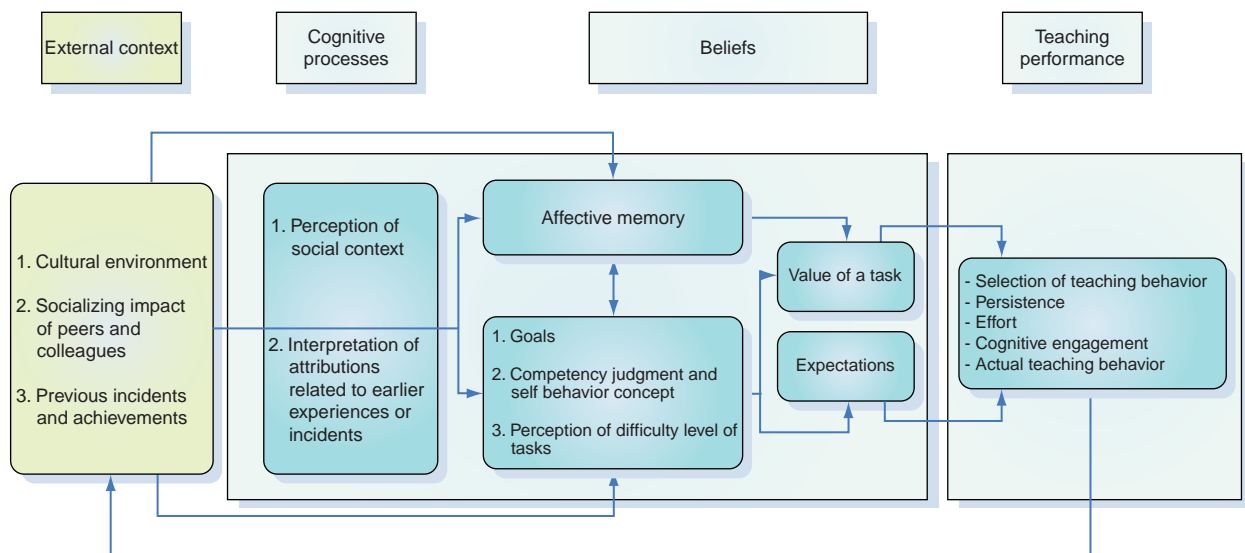


Figure 1 Beliefs in relation to motivational variables and processes. From Wigfield, A. and Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology* 25, 68–81.

related to learners (e.g., beliefs about inclusion and about diversity), knowledge (epistemological beliefs), teaching components (beliefs about the curriculum, beliefs about what learning content is important, beliefs about media (information and communication technology (ICT)), teaching strategies, evaluation, etc.), parents, instructional context, and organizational dimensions. The literature reflects this varied nature of the focus of teachers' beliefs. For example, He and Zhao (2006) state that teachers' beliefs include beliefs about learners, teaching, disciplines, teaching efficiency, and teacher themselves. This is in line with distinctions made by Kagan (1992) who indicates that beliefs are a particularly provocative form of personal knowledge that is generally defined as pre- or in-service teachers' implicit assumptions about student learning, classrooms, and the subject matter to be taught. A major problem in research of teacher beliefs is the wide-ranging nature of beliefs. Often research about educational beliefs is too context-free. Hence, almost 15 years ago, Pajares (1992) considered restricting the object of research study to a particular educational belief or a limited belief system and to clearly define the object under investigation.

Measuring Teachers' Beliefs

The nature of beliefs makes it difficult to measure them straightforwardly. Since beliefs cannot be observed directly, they are measured indirectly. Shavelson *et al.* (1986) presented the first overview of available methods to study beliefs. They refer to analysis of policies, repertory grid technique, and process tracing (e.g., think-aloud, stimulated recall, and journal keeping). Kagan (1992) presents a systematic description of techniques that is still viable. She refers to (1) experimental tasks that ask teachers to think aloud when analyzing classroom vignettes or when viewing their videotaped performances (stimulated recall); (2) semi-structured interviews, during which teachers are asked to recall specific classroom events and decisions; (3) concept maps that teachers draw to depict their understandings of pedagogical issues; (4) a close analysis of the language that teachers use to describe their thoughts and actions; and (5) traditional Likert-type attitude scales. Other approaches embrace case study methodologies, analysis of teachers' concerns, and in-depth interviews.

Beliefs as Entry Qualifications of Prospective Teachers

A result of the growing interest in teachers' beliefs is the conviction that preservice teachers already enter teacher education with a clearly established belief system about

teaching, learning, and themselves as teachers developed over years of life experience and exposure to a wide variety of teaching–learning situations and contexts. It is well documented that these preconceived ideas remain quite stable during teacher training (e.g., Pajares, 1992; Wideen *et al.*, 1998) and have a strong influence on student teachers' understandings of and experiences in their classroom placements (e.g., Kagan, 1992; Richardson, 1996).

Old beliefs must be challenged and proven unsatisfactory in order to assimilate new beliefs into existing conceptions. This is in line with the general findings of Quillen (unpublished PhD thesis) that: "Teacher education researchers have long been aware of the power of beliefs and the resistant-to-change beliefs of many preservice teachers as well as experienced teachers." This puts teacher educators in a difficult situation since the beliefs:

- are well established by the time student teachers enter their preservice training;
- are already formed at an early stage;
- develop into a belief system that is difficult to change; and
- act as a selective filter to develop new knowledge.

This explains, in part, why student teachers tend to reiterate their personal classroom experiences and also explains why some authors stress the predictive value of beliefs for a successful entry into the teaching profession. The latter explains why it might be important to unlearn and discard some beliefs.

The results of the research about the beliefs of incoming student teachers also clearly point out that, next to stability, in the initial beliefs, changes also could be observed, but this depended largely on the beliefs held by the supervising teachers. The former is not only critical in view of developing sound professional teaching competences, but also to reconsider the nature of the teacher education experience itself. Teacher educators should be aware of the fact that student teachers enter the teacher-education setting with a set of beliefs about teacher education. Preservice teachers tend to think of teaching primarily as a task involving affective, interpersonal relationships rather than a profession requiring a skilled and knowledgeable practitioner. Generally, preservice teachers do not expect to receive much from their education classes and appear to be less interested in what they perceive to be theory and more interested in practical approaches. They feel that they would be good teachers without any preparation, because they believe that the majority of their knowledge about teaching would come from practice in the field or through trial and error when they eventually enter the classroom.

It appears from the findings of these studies that preservice teachers underestimate the complexity of teaching. Yet, student teachers who are guided by these

kind of naïve, idealistic, and unrealistic teaching beliefs have been found to feel overwhelmed, shocked, and disillusioned when confronted with the complexities and responsibilities of a classroom (e.g., Wideen *et al.*, 1998). Student teachers who fail to reconstruct their images of self as teachers appropriately may encounter frustrations sufficiently severe to terminate their teaching career before they really had a chance to begin (Kagan, 1992). Therefore, in order to improve professional preparation and teaching practices, teacher educators cannot neglect the role of beliefs and need to develop specific strategies and approaches to cater to belief systems as an integrated part of the teacher education system. Moreover, teacher educators need to be explicit about their own beliefs about teacher education and training.

Changing Teachers' Beliefs in Teacher Education

Given the established nature of beliefs, and the fact that the beliefs work as a selective filter in developing professional teaching competences, teacher-education approaches are challenged to consider the pervasive influence of teachers' beliefs. This could result in the implementation of a series of specific activities or teacher-education strategies. However, the processes by which beliefs can be changed are not well understood. In addition, research overall is rather dour about the prospects of changing candidates' beliefs. The simple adoption of a number of strategies, tricks, etc., is insufficient. Teacher educators themselves should, first of all, adopt an updated conception as to how student teachers evolve and how teacher education should take this into consideration. There is clear empirical evidence that shows how teacher education programs are too focused on transmitting pedagogical knowledge that hardly gives consideration to modifying beliefs. In their review, Wideen *et al.* (1998) conclude that successful teacher education programs not only merely change but also build upon student teachers' beliefs by making use of a systematic and consistent support of teacher educators during institutional meetings, as well as of cooperating teachers during teaching-practice periods. Hence, these teacher education programs are characterized by the integration of practical experiences and theoretical study.

The approach of the pedagogy of realistic teacher education (Korthagen *et al.*, 2001) is exemplary in this context. This approach starts from concrete practical problems and the concerns experienced by student teachers in real contexts. It aims at the promotion of systematic reflection of student teachers on their own and their students' wanting, feeling, thinking, and acting on the role of context, and on the relationships between those aspects. Furthermore, this realistic approach states that

teacher education should respect the three consecutive levels along which professional learning develops (gestalt, schema, and theory). Teacher education should help student teachers become aware of their needs, find useful experiences, and reflect on these experiences. At the gestalt level, actions are typically based on unconsciously triggered needs, values, meanings, feelings, and behavioral inclinations. Gestalts refer to the personal conglomerates of needs, concerns, values, meanings, feelings, and behavioral inclinations united into one inseparable whole, which often unconsciously (or only partly consciously) plays a role in shaping teacher behavior. During the process of professional learning, the student teacher may reflect on his or her gestalts and develop a conscious schema of concepts, characteristics, and principles that are helpful in describing practice. The theory level is reached when a logical order is constructed in and between schemata, resulting in a coherent theory.

The above paragraph not only implies a number of potentially successful strategies to influence teachers' beliefs, but also reconsiders the model presented earlier about how beliefs as part of a complex interplay of variables and processes might be helpful when considering the antecedents and consequences of the adoption of specific beliefs. Considering the importance of the external context, the model suggests that teacher education approaches should:

- Present a richer and more varied cultural context where a multitude of beliefs and related teaching and learning approaches are being adopted. Changes in teachers' beliefs are generally not effected by reading and applying the findings of educational research, but, instead, student teachers appear to obtain most of their ideas from actual practice. Therefore, it seems important to put preservice teachers in a sufficient number of real-life situations and oblige them to reappraise their existing beliefs, attitudes, and knowledge. Select a rich variety of learning situations that provides opportunities to reflect on and reconceptualize beliefs about teaching and learning.
- There is plenty of research evidence that points at the critical role of sharing and discussing instructional practices with peers. Hence, we recommend bringing student teachers in contact with peers and cooperating teachers who have adopted a wide variety of beliefs and related instructional approaches. In the context of educational innovations based on the use of ICT, grass-roots projects and good practices have, for example, proven to be successful teacher education projects. However, mere observation of personal practices and instructional models of peers and cooperating teachers is insufficient. Reflection on these experiences and practices should be fostered in order to result in more sophisticated epistemological beliefs.

- Develop a feedback approach that accepts and stimulates varying, dissimilar, and divergent thinking about and the adoption of teaching and learning approaches; and giving the critical impact of cognitive processes that underlie teachers' beliefs, student teachers should be invited to reconsider their perception of the social context and their attributions related to earlier instructional success or failure.
- Rath (2001) presents a list of concrete and more general approaches that are more or less in line with the former statements about changing (student) teachers' beliefs:
- Confronting student teachers with dissonance. Dissonance theory suggests that if we engage teacher candidates in activities that arouse dissonance, beliefs might change. This implies that we arrange for conflicts between new experiences and past experiences. In this context, it is important to raise conflicting beliefs to the surface, and to foster discussion, comparison, examination, and challenge.
- Apprenticeship experiences. This implies that apprenticeship experiences are critical at the start of a teacher education program. This is in sharp contrast to earlier approaches where internships were rather a part of the later phases in teacher education. Of course, the teacher educator should carefully select settings that are helpful in activating specific beliefs.
- – Values clarification involves asking student teachers to reconsider the beliefs they hold. After examination, reacceptance, considering alternatives, anticipating consequences, and trying their implications, the values are expected to develop.
- Case study. In this instructional strategy, student teachers are invited to study cases of instruction through different lenses: the lens of their own beliefs, belief systems derived from constructivism, or direct instruction or the project method. It is hypothesized that this could invoke changes in the belief systems.

Changing Particular Preservice Teachers' Beliefs: Mathematics and Educational Innovations

To exemplify some of the approaches described earlier to change teacher beliefs, we focus on practices and related research that tried to change teacher beliefs related to mathematics and educational innovations (e.g., ICT).

Based on a literature review in relation to teachers' beliefs on mathematics, Ernest (1989) concludes that three components of beliefs have a significant impact on teachers' instructional practices:

- his/her view or conception of the nature of mathematics,
- his/her model or view of the nature of mathematics teaching, and
- his/her model or view of the process of learning mathematics.

This belief structure can be partly related to the earlier general description of beliefs and points at epistemic issues, beliefs about how peoples learn, and how we teach. Quillen (unpublished PhD thesis) reviewed the literature about mathematics beliefs of preservice teachers. She studied in detail the way the student teachers adopted relational beliefs or instrumental beliefs. Relational beliefs about teaching and learning mathematics refer to the inclination to provide opportunities for students to explore, investigate, use a variety of problem-solving strategies, and use prior knowledge to solve problems involving concepts that have not been previously taught. Instrumental beliefs about teaching and learning mathematics mirrors direct instruction, teaching by telling, and using memorization of rules, formulas, and procedures to solve problems.

There is abundant literature about attempts to change teacher beliefs related to mathematics. Most attempts share the basic characteristic that challenging these beliefs is done in an explicit way in order to transform the beliefs from nonevidential to evidential (Green, 1971) and exposing them to the individual preservice teacher in view of a critical analysis and discussion with student peers. Strategies to make beliefs explicit build on involving preservice teachers in doing mathematics and learning through talking mathematics. Making beliefs explicit is also the basis of intervention studies that research the extent to which preservice teachers adopt the constructivist base of the new mathematics curriculum, introduced by the National Council of Teachers of Mathematics (NCTM) in the USA in 2000. Researchers have pointed out that most teachers' past experiences with mathematics are/were rather in conflict with the NCTM model that embraces constructivist pedagogical beliefs. Most teachers reflect beliefs that build on their experiences with traditional, behaviorist methods of mathematics instruction that mainly build on transmission and absorption. This creates a tension between teachers' beliefs and the ambitions of the mathematics curriculum innovation. The study of Swars *et al.* (2006) is an attempt to influence the related teachers' beliefs. The study clearly shows that it required a long-lasting intervention (2 years) and a wide variety of instructional strategies to successfully change the teachers' beliefs about the teaching of mathematics. Strategies comprised: thorough reading of the basic assumptions of the new curriculum, classroom discussions and learning activities focused on social-constructivist pedagogy, analytical viewing of

classroom videotapes of classrooms, clinical interviews with children (e.g., about the children's understandings of number and operations), report writing about field experiences using the new type of mathematics pedagogy, and analysis of NCTM-based lessons (check of coherence between practices and principles).

Another extensively studied field where attempts to change and influence teacher beliefs is that of educational innovations in general and the implementation of ICTs in particular. In the literature, there is a general agreement that the adoption of an educational innovation can only be explained when the teachers' beliefs are also taken into account. In the research literature, it is, for example, acknowledged that teachers' beliefs tend to be associated with a particular use of ICT in the classroom (e.g., Ertmer, 2005). Studies explain this by hypothesizing that teachers who use computers do so because their conceptions of ICT use fit into their existing teaching belief system. If teachers perceive that computers address important instructional and learning needs, the perceived value will be higher. There is growing evidence that teachers, for example, adopting constructivist beliefs, are more dynamic computer users. Research that focuses on changing teacher beliefs about ICT builds on a variety of strategies. Typical studies stress on the importance of setting up of long-term initiatives. Furthermore, they also build on critical reflection by student teachers on their own video-taped science lessons culminating in a focus group session. This strategy seems to make preservice teachers able to differentiate between their own beliefs and teaching practices that are not or in line with critical assumptions. Research also indicates that changes in beliefs do not imply that earlier beliefs have to disappear. Whereas in the past, authors centered on bipolar and contrasting dimensions in belief structures, for example, the bipolar distinction between teacher-centered traditionalistic and more progressive or student-centered educational beliefs, researchers have shifted their attention toward a multidimensional approach of the structure in belief systems. In newer approaches to influencing and changing preservice teachers' beliefs, teacher educators rather try to promote that student teachers hold both traditionalistic and progressive educational beliefs. Recent research points out that teachers adopting both strong constructivist beliefs and strong traditional beliefs reflect a higher adoption of and more integrated use of ICT.

Conclusions

Wrapping up the conceptual, theoretical, and empirical base in relation to teachers' beliefs results in a complex and especially challenging agenda for teacher educators. Though current conceptions about teachers' belief are

rooted in a strong research tradition, it is not yet clear how teacher educators can cater to teachers' beliefs in an adequate way. Promising directions for future research and educational practice adopt a comprehensive view toward the education of student teachers and in-service teachers. They consider teachers as active agents with a central personal responsibility in the development of their professional competences.

See also: A Pedagogy of Teacher Education; Mathematics Teacher Education.

Bibliography

- Beijaard, D., Meijer, P. C., and Verloop, N. (2004). Reconsidering research on teachers' professional identity. *Teaching and Teacher Education* **20**, 107–128.
- Ernest, P. (1989). The knowledge, beliefs and attitudes of the mathematics teacher: A model. *Journal of Education for Teaching* **15**, 13–33.
- Ertmer, P. A. (2005). Teacher pedagogical beliefs: The final frontier in our quest for technology integration? *Educational Technology Research and Development* **53**, 25–39.
- Green, T. (1971). *The Activities of Teaching*. New York: McGraw-Hill.
- He, Y. and Zhao, H. H. (2006). Conceptual knowledge about teacher beliefs. *Journal of Yibin University* **5**(5), 115–117.
- Kagan, D. M. (1992). Implications of research on teacher beliefs. *Educational Psychologist* **27**, 65–90.
- Kagan, D. M. (1995). Research on teacher cognition. In Ornstein, A. C. (ed.) *Teaching: Theory into Practice*, pp 226–238. Boston, MA: Ally and Bacon.
- Korthagen, F., Kessels, J., Koster, B., Lagerwerf, B., and Wubbels, T. (eds.) (2001). *Linking Practice and Theory: The Pedagogy of Realistic Teacher Education*. London: Lawrence Erlbaum.
- Pajares, F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research* **62**, 307–332.
- Raths, J. (2001). Teachers' beliefs and teaching beliefs. *Early Childhood Research and Practice. An Internet Journal on the Development, Care, and Education of Young Children* **3**. <http://ecrp.uiuc.edu/v3n1/raths.html> (accessed December 2008).
- Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In Sikula, J. (ed.) *Handbook of Research on Teacher Education*, pp 102–119. New York: Macmillan.
- Shavelson, R. J., Webb, N. M., and Burstein, L. (1986). Measurement of teaching. In Wittrock, M. C. (ed.) *Handbook of Research on Teaching*, 3rd. edn., pp 50–91. New York: Macmillan.
- Swars, S., Smith, S., Smith, M., and Hart, L. (2006). Elementary preservice teachers' changing pedagogical and efficacy beliefs during a developmental teacher preparation program. *Proceedings of the PME-NA 2006. 28th Psychology in Mathematics Education Conference*, pp 548–555. Mexico, Mérida: Universidad Pedagógica Nacional.
- Wideen, M., Mayer-Smith, J., and Moon, B. (1998). A critical analysis of the research on learning to teach: Making the case for an ecological perspective on inquiry. *Review of Educational Research* **68**, 130–178.
- Wigfield, A. and Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology* **25**, 68–81.
- Wyer, R. S. and Albarracín, D. (2005). Belief formation, organization and change: Cognitive and motivational influences. In Albarracín, D., Johnson, B. T., and Zanna, M. P. (eds.) *The Handbook of Attitudes*, pp 273–322. Hillsdale, NJ: Erlbaum.

Further Reading

- Cunningham, D. J., Schreiber, J. B., and Moss, C. M. (2005). Belief, doubt and reason: C. S. Peirce on education. *Educational Philosophy and Theory* **37**, 177–189.
- Fang, Z. H. (1996). A review of research on teacher beliefs and practices. *Educational Research* **38**, 47–65.
- Gregoire, G. M., Ashton, T., and Algina, J. (2004). Changing preservice teachers' epistemological beliefs about teaching and learning in mathematics: An intervention study. *Contemporary Educational Psychology* **29**(2), 164–185.
- Hofer, B. K. and Pintrich, P. (2002). *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*. London: Lawrence Erlbaum Associates Publishers.
- Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies* **19**, 317–328.
- Raths, J. and McAninch, A. C. (eds.) (2003). *Teacher Beliefs and Classroom Performance: The Impact of Teacher Education*, vol. 6. Greenwich, CT: Information Age Publishing.
- Richardson, V. (2003). Preservice teachers' beliefs. In Raths, J. and McAninch, A. C. (eds.) *Teacher Beliefs and Classroom Performance: The Impact of Teacher Education*, pp 1–22. Greenwich, CT: Information Age Publishing.
- Rokeach, M. (1976). *Beliefs, Attitudes, and Values. A Theory of Organization and Change*. San Francisco, CA: Jossey-Bass.
- Woolfolk, H. A., Davis, H., and Pape, S. J. (2006). Teacher knowledge and beliefs. In Alexandra, P. A. and Winne, P. (eds.) *Handbook of Educational Psychology*, 2nd edn., pp 715–737. Mahwah, NJ: Lawrence Erlbaum Associates.

Teacher Education and Models of Teacher Reflection

V K LaBoskey, Mills College, Oakland, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Defining Reflection

One would be hard-pressed to locate a teacher education program in the world today that does not somehow express interest in the promotion of teacher reflection. Indeed, the notion has become so deeply embedded in conceptualizations of teaching and learning to teach that many no longer see the need to make such an orientation explicit in the descriptions of their programs. On the one hand, this could be perceived as a good thing – the inherent value of reflective practice is no longer in question. On the other hand, it can mean that educators are not paying enough attention to the meaning or effectiveness of its definition, intent, implementation, or outcome. Furthermore, consistency of use can too often be assumed and significant differences overlooked; this failure to problematize teacher reflection can diminish its potential and undermine progress.

The aim of this article is to draw attention to the similarities and differences in current understandings and applications of reflection in teacher education and offer a conceptual framework for refocusing and reinvigorating the work in this arena. It begins with an articulation of common definitions of teacher reflection, looking first at the roots of this notion in the scholarship of John Dewey and Donald Schön. The aspects of these long-standing formulations that seem to persist, crossing temporal and contextual boundaries, are identified. The fine-tuning that has occurred over time, as well as the most widespread shifts in emphasis, are then portrayed. Employing this generalized definition as a lens, a framework for identifying and differentiating the most prevalent models of reflection now existing in teacher education is presented. The framework consists of three interrelated continua: purpose, content, and means. This discussion concludes with a consideration of the implications of these variations, along with recommendations for how the continua might be utilized to inform and enhance subsequent steps with regard to the research and practice of reflective teacher education.

Historical Roots

Rodgers (2002), one of the many scholars concerned with the role of reflective thinking in both teachers' and students' learning, worries that "reflection has suffered from a loss of meaning. In becoming everything to everybody, it has lost its ability to be seen" (p. 843). In order to

prevent that from happening, she revisits four of the main criteria that characterized Dewey's original conceptualization of the term:

1. Reflection is a meaning-making process that moves a learner from one experience into the next with deeper understanding of its relationships with and connections to other experiences and ideas. It is the thread that makes continuity of learning possible, and ensures the progress of the individual and, ultimately, society. It is a means to essentially moral ends.
2. Reflection is a systematic, rigorous, disciplined way of thinking, with its roots in scientific inquiry.
3. Reflection needs to happen in community, in interaction with others.
4. Reflection requires attitudes that value the personal and intellectual growth of oneself and of others. (Rodgers, 2002: 845).

The result of Rodgers' subsequent analysis is a reminder that, if engaged as Dewey intended, reflection is the means by which human beings learn from their experiences so as to improve themselves and their society. It is the only way for making meaning that is justified – empirically, theoretically, morally – and thus educators ignore its systematic articulation and development at their peril, particularly with regard to teacher education, where the learning of both the novice teachers and their future students is the essence of the endeavor.

The field of teacher education would do well, in other words, to revisit Dewey's (1938) definition of reflection: "Active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which it tends" (p. 9). It needs, therefore, to be both the means of reflective teacher education and the end toward which it aims – the preparation of teachers who both embrace this process of making meaning from experience for themselves and conceptualize and facilitate their students' learning in the same way. Certainly, these two core ideas are fairly universal in the field and seldom contested. Teacher educators agree that reflection is a means for candidates to learn from practice, which can subsequently be employed to improve that practice. But the details of the why, what, and how of this process can vary greatly.

Many in the field concur that one of the main reasons reflection has become so widely embraced in recent years and so well integrated into teacher education is the work

of Donald Schön (1983). The magnitude of his impact is in part due to the fact that he considered reflective practice to be synonymous with professional thinking. Reflection is not merely contemplation or recollection, and not only for the purpose of learning; it is problem based and action directed. It is the means by which professionals can solve their problems of practice in meaningful and ever more effective ways. The terms he used to identify the two ways of engaging the process – reflection-on-action and reflection-in-action – make that connection explicit. By suggesting that the latter is even a possibility, he helped shift attention to the enactment and outcome of reflective deliberation.

Key to successful problem solving, in Schön's view, is the ability to frame – and reframe – a problem of practice appropriately, since that determines and delimits the solutions sought and generated. In general, teacher education embraced these notions and over the last several years has given particular attention to making better connections between theory and practice and to helping novices learn the skills and dispositions necessary to reflection on and in action. But ironically, this widespread acceptance may have contributed to the loss of meaning and the invisibility to which Rodgers (2002) referred. Consequently, those who encourage another resurgence of interest in teacher reflection are doing so by fine-tuning its definition and by emphasizing particular aims and expectations for both the process and its results.

Recent Developments

Many of the characterizations of reflective teacher education that emerged just prior to or during the Schön revitalization included categorizations of various types of reflection. Van Manen (1977), for instance, identified three levels of reflectivity: practical/technical, social/political, and moral/ethical. Though he, and many others, suggested that the practical/technical was of a lower level and thus not the ultimate aim, it was nonetheless viable as a focus or form of reflection. One of the widespread changes occurring in the field in recent years is the elimination of the practical/technical as a stand-alone category, which was actually advocated by both Dewey and Schön. Donahue (2005), for instance, in discussing the meaning of reflection in the teacher education program at Mills College states: "While some policymakers, textbook publishers, parents, and perhaps even teachers think of teaching as merely a technical activity of implementing curriculum according to established criteria, most teachers appreciate that teaching is intellectual work and reflection is a vital means of nurturing the intellectual dimension of teaching" (p. 40). Apparent in this statement is the assumption that reflective teaching is different from technical teaching.

Most teacher education programs engaged with reflective teaching at present accept that, at the very least, they

need to be attentive to the substantive learning and the transformed thinking of their student teachers. Feiman-Nemser (2001), when writing about how to strengthen and sustain teaching for the future, makes the following claim: "After decades of school reform, a consensus is building that the quality of our nation's schools depends on the quality of our nation's teachers. . . . This paper rests on a single premise with far-reaching consequences—if we want schools to produce more powerful learning on the part of students, we have to offer more powerful learning opportunities to teachers" (pp. 1013–1014). In their preservice programs, then, student teachers need to engage in cycles of reflective inquiry in order to "begin forming the habits and skills necessary for the ongoing study of teaching in the company of colleagues" (p. 1019).

Korthagen and Vasalos (2005) articulate a comparable recommendation in greater detail. They advocate for a structured support process aimed at the promotion of core reflection, that which is focused on the deepest levels of a student teacher's personality – mission and identity. The means for facilitating this type of reflection is through their ALACT model, which consists of five phases: action, looking back on the action, awareness of essential aspects, creating alternative methods of action, and trial. It bears much resemblance to the typical action research cycle, but gives greater emphasis to the affective dimension of the learning to teach process and to the need for surfacing and exploring contradictions between a candidate's identity and mission and her beliefs and behaviors.

Several in the field at present, while applauding this insistence on deeper learning and transformed beliefs, consider it to be a necessary but not sufficient effort. Also essential is greater specification of the ends toward which reflective teaching is directed. Those in this camp are reemphasizing the moral imperative inherent in Dewey's original definition: "I will argue that efforts to prepare teachers who are reflective must both foster genuine teacher development and support the realization of greater equity and social justice in schooling and the larger society" (Zeichner, 1996, p. 201). This implies that a certain set of criteria be applied in determining whether the solutions teachers develop for their problems of practice are meaningful and effective – the aim of professional reflection, according to Schön. What is more, the nature of the problems that are detected and the ways in which they are framed are challenged by this orientation, thus calling into question the terminology itself and its accompanying protocols: "This involves framing certain educational issues as dilemmas rather than problems with clear solutions and deliberating thoughtfully about decisions that involve competing claims to justice" (Cochran-Smith, 2004, p. 15).

In essence, these teacher educators are stressing the fact that reflection in teaching and teacher education is a means to other ends and not an end in itself. The two,

however, are necessarily linked; learning to teach for social justice requires taking what Cochran-Smith refers to as an inquiry stance toward practice. The candidates must, therefore, reflect on their ideas and actions in critical inquiry communities where issues of equity are raised and fundamental assumptions, prejudices, and even values are interrogated. Others who stress similar types of political and ethical outcomes employ the idea of reflection in a comparable way. Kumashiro (2002), who engages with what he calls antioppressive teacher education, argues that typical forms of self-reflection are not enough unless they include the exposure and interruption of the harmful repetitions embedded in both intentional and unintentional teaching activity.

What is noticeable in these representative descriptions of the purpose and facilitation of reflection in contemporary teacher education programs is the frequency with which terms like inquiry and teacher research are utilized. In many instances, they are considered to be synonymous with reflective teaching – the expressions are used interchangeably. For many teacher educators, taking an inquiry orientation to practice is the same as reflective teaching. Thus, various forms of teacher inquiry are considered to be the means by which student teachers engage in and learn about reflection on and in practice. In surveying 12 European preservice and in-service programs about their conceptions of reflective practice, Clarke and Chambers (1999) found that many equated it with action research. In fact, several who favored such approaches never used the term reflection with regard to their programs at all, which tends to validate Rodgers' previously mentioned concerns. Others consider inquiry and reflection to be somewhat different endeavors; in this volume, for instance, there are separate articles on models of reflection in teacher education and inquiry-oriented preparation programs. In either case, more explicit exploration of the relationship between the two is needed. The following section presents a framework for differentiating models of teacher reflection that might be useful in that regard.

Models of Reflective Teacher Education

The literature on reflection in teacher education can be categorized according to the foci suggested by Donahue (2005): "In defining reflection, theorists have wrestled with questions of why teachers reflect, what they reflect about, and how they reflect" (p. 39). In part, this selectivity is an artifact of the research and reporting process – the circumscription necessary to a meaningful and manageable investigation or essay. But in many instances, these delimitations also represent a differentiation in the philosophies and priorities guiding the design and implementation of particular reflective teacher education programs. Consequently, these categories should be useful for

making distinctions among the programs. In addition, there is considerable within-category variation that might best be characterized by three nondevelopmental continua – purpose, content, and means. These continua are interrelated, meaning that where a model of teacher reflection placed on one continuum tends to suggest where it would or should be placed on the others. These three continua, therefore, have the potential to be utilized as a framework for identifying, differentiating, and ultimately constructing and evaluating models of reflection in teacher education.

In the next section, each continuum is described in more detail, drawing upon examples from the literature and the generalized definitions previously summarized for illustration and explanation. Then the ways in which the three seem to interrelate in theory and practice are discussed, thereby demonstrating the conceptual integrity of the conglomerate and its potential as an analytical framework for the design, critique, and transformation of reflective teacher education.

Purpose

This continuum is concerned with the aim of the reflective process – why student teachers engage in reflection and what should be gained from that endeavor. On one end of the continuum (let us say the left end to facilitate comparison with the other continua) is a focus on the development of the lifelong learning and problem-solving capacities of individual candidates. On the other end (let us say the right end) the focus is on the transformation of social institutions and systems in ways more consistent with democracy, equity, and social justice. What is important to clarify here, especially in this first discussion, is that this is not a developmental continuum. It is not meant to imply that attention to student teacher learning is at a lower level or a less important area of concern than attention to social transformation. Nor should it suggest that the intention is to move a student teacher or a teacher education program through a series of stages that culminate in an exclusive focus on changing society.

What the continuum does convey is a viable range of options. Any purpose not inclusive of at least one end of the continuum would not be consistent with the current and long-standing definitions of reflection described above. For instance, a technical orientation intended simply to aid novices in applying preexisting curricular packages would not be considered reflective teacher education because it is outside the range; it could not be placed anywhere on this continuum. Furthermore, programs and interventions that do not attend at all to an articulation of its ultimate goals would not qualify as a model of teacher reflection. What the continuum also embodies are the multiple possibilities for combining both aims. Anywhere on the continuum, aside from the extreme ends, represents some combination of the two.

The middle point, therefore, signifies a full and balanced integration of both aims.

In applying this analytical tool to the literature review, several programs and projects that could be located near the left end of the continuum were discovered. They were concerned with the meaningful learning of individual student teachers and with the development of the attitudes and skills of reflection on and in action that would ensure lifelong growth and ever-improving practice. Feiman-Nemser (2001), for instance, argued for the need to place “serious and sustained teacher learning at the center of school reform” (p. 1014). Though this perspective may be closer to the middle of the continuum, she characterizes improvement primarily in terms of more powerful student learning. Though this is a critical component of institutional and social transformation, it does not include the interrogation and undoing of embedded inequitable and oppressive structures so fundamental to the social-justice end of this continuum.

In contrast, programs or practices near the right end of the continuum were difficult to locate. To be sure, the number of reflective teacher education programs that consider more equitable institutions and just societies to be their ultimate aim is increasing. But these also embrace the goal of individual growth, which would situate them closer to the center of the continuum. In fact, most would consider one to be impossible without the other. On the one hand, social transformation is not achievable without individual teacher transformation and, on the other, improvement in teaching is meaningless in the absence of excellent and equitable outcomes for all learners both within school and beyond. Such a finding might suggest a need to refine the continuum in the future, but at any rate, it does not seem to interfere with its analytical potential, or that of the other continua.

Content

This continuum is concerned with the content of reflection – what it is that student teachers are reflecting on and about. It constitutes the target and the source for the issues and dilemmas addressed by the reflective process. At the left end of the continuum, attention is directed toward the student teacher – his/her internal processing and external actions, and often, the relationship between the two. At the right end, novices interrogate the inequities embedded in educational institutions and systems and the societal structures, norms, and discourses they represent and validate.

As in the previous instance, more examples could be found for the left side of the continuum. Many teacher education programs consider the primary impetus for teacher reflection to be the problems of practice that they encounter: “I define reflection as the deliberate and purposeful act of thinking which centers on ways of

responding to problem situations in teaching and learning” (Loughran, 1996, p. 14). Like Loughran, most teacher education programs that direct reflection toward problems of practice insist on attention to the relationship between teaching and learning. A teacher’s beliefs, knowledge, or behavior cannot be understood, evaluated, or improved irrespective of an assessment of student learning. But again, only a few raise questions about whether or not those student outcomes represent greater equity or improved social justice. Those who do refer to social justice seldom incorporate explicit deliberations about the larger context or include challenges to the ways in which typical modes of educational problem solving may reify those structures.

Also as above, few programs seemed to situate themselves at the far right end of the content continuum. Those who insist that student teachers reflect about systemic inequities, also emphasize the need for individuals to interrogate their own beliefs and practices in that regard, which would be at the center of the continuum. Gitlin (2005), for instance, proposes the following: “what needs to occur at a fundamental level to confront oppression and move toward any form of social justice is to make everyday politics the object of inquiry” (p. 15). This requires reflection on oneself and one’s cultural community as well as that of the other, and to determine what traditions of interaction between the two protect the repressive *status quo* so that those conventions can be disrupted. Noteworthy here is that such models of teacher reflection do not simply represent a straightforward combination of the two ends of the continuum. In these exemplars, the nature of self-reflection is different; it reframes the problems of practice using the equity lens. The means for promoting such reflection, though similarly variable, seem to be somewhat less transfigurative.

Means

This continuum is concerned with how student teachers engage in the reflective process and how they learn to do so. At the left end of this continuum are means for facilitating and encouraging the interrogation and improvement of personal assumptions, knowledge, and pedagogy. At the right end are means for fostering the capacities and propensities necessary for political activism. The vast majority of current research into reflective teacher education seems to be concerned with this continuum – figuring out how to teach and support student teacher reflection, as well as ensure that candidates incorporate it into their practice for the long term. In these instances, the purpose and content, though not the direct objects of attention, are often mentioned in the course of process description or implied by the evaluative structures used to interpret the effects. Thus, an approximate placement on all three continua is usually possible.

As with the others, most programmatic efforts for fostering reflection would be placed on the left half of the continuum. They are directed toward helping individual student teachers learn the skills and dispositions necessary for engaging in reflection on and in practice both during the program and throughout their careers. They represent efforts to help student teachers take an inquiry orientation to practice – to frame and reframe their practical problems, design and implement powerful pedagogies, and analyze and evaluate student learning outcomes – so that they can learn from their actions, develop their professional knowledge base, and engage in more effective teaching. These studies explore the potential benefits of a whole range of very specific strategies, for example, journals, seminars, mentoring, coaching, metaphor production, action research, reading groups, autobiography, mural-making, dramatic enactment, and so on. Other investigations consider more holistic interventions, such as Korthagen's ALACT model for reflection (Korthagen and Vasalos, 2005). There is a growing body of literature, referred to as self-study, where the teacher educator researchers are equally concerned with their own learning and development through various forms of investigative research that include similar varieties of reflective deliberation (Loughran *et al.*, 2004).

Some of these studies that address the hows of reflection also examine means for helping student teachers, and often their teacher educators, to interrogate and transform their deeply held beliefs, assumptions, and understandings using the lens of equity and social justice. They endeavor to undo existing prejudices and help develop a critical stance toward the processes of teaching and learning, again through particular interventions or through a more pervasive structure, such as the principled-practice approach utilized at Mills College (Kroll *et al.*, 2005). These models would be located nearer to the center of the continuum, especially if, as in the latter case, they also advocate for instruction that will prepare novice teachers and their students for civic engagement and encourage them to “join with others in larger movements for educational and social equity” (Cochran-Smith, 2004, p. 159). Again, it means that there are no models on the far right end of this continuum. Instead, those who attempt to include strategies for nurturing political activism also engage means for the development of the social-reconstructionist dimension of teaching. The center position is, in that sense, more additive than transformative; it adds large-scale involvement to the already-reframed critical pedagogies.

Implications

As noted above, most of the readings in this domain were targeted at just one of the continua, especially individual articles or chapters, in part due to the limitations of time

and space. But even in those instances, placements on the other dimensions could often be inferred. If a study was investigating the impact of a strategy for promoting critical reflection, for instance, the goal of greater social justice was either stated or implied and the content for reflection tended to incorporate systemic traditions. But, the framework was also useful in detecting potential contradictions. If a model claimed to have the aim of promoting social justice, but did not include structural inequities in their investigatory content, that purpose might be called into question. At any rate, it was clear that the three continua are and should be consistent with one another in a coherent model of reflection in teacher education and thus together can serve as a framework for future analysis and development.

Purpose, content, and means are, therefore, three foci to which all teacher education programs concerned with the promotion of teacher reflection need to give serious attention in both their deliberations and enactments. Even though all research studies cannot focus on all three at once and there can be great benefit from isolated attention to programmatic details, the other dimensions must always be considered in both the framing and interpretation of those investigations.

To be considered a viable model for teacher reflection in teacher education, the purpose, content, and means of these models must be articulated and there needs to be consistency among them with regard to placement on the continua. If the fundamental purpose of education is considered to be greater democracy, equity, and social justice for children and for the world, reflective teacher education programs must be designed and enacted so that they could be situated at the center of all three continua.

See also: Inquiry-Oriented Teacher Education; Self-Study by Teacher Educators.

Bibliography

- Clarke, B. L. and Chambers, P. A. (1999). The promotion of reflective practice in European teacher education: Conceptions, purposes and actions. *Pedagogy, Culture and Society* 7, 291–303.
- Cochran-Smith, M. (2004). *Walking the Road: Race, Diversity, and Social Justice in Teacher Education*. New York: Teachers College Press.
- Dewey, J. (1938). *Experience and Education*. New York: Collier Books-Macmillan.
- Donahue, D. M. (2005). Preparing and supporting the reflective practitioner. In Kroll, L. R., Cossey, R., Donahue, D. M., *et al.* (eds.) *Teaching as Principled Practice: Managing Complexity for Social Justice*, pp 35–56. Thousand Oaks, CA: Sage Publications.
- Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record* 103, 1013–1055.
- Gitlin, A. (2005). Inquiry, imagination, and the search for a deep politic. *Educational Researcher* 34, 15–24.
- Korthagen, F. and Vasalos, A. (2005). Levels in reflection: Core reflection as a means to enhance professional growth. *Teachers and Teaching: Theory and Practice* 11, 47–71.

- Kroll, L. R., Cossey, R., Donahue, D. M., et al. (eds.) (2005). *Teaching as Principled Practice: Managing Complexity for Social Justice*. Thousand Oaks, CA: Sage Publications.
- Kumashiro, K. K. (2002). Against repetition: Addressing resistance to anti-oppressive change in the practices of learning, teaching, supervising, and researching. *Harvard Educational Review* **72**, 67–92.
- Loughran, J. J. (1996). *Developing Reflective Practice: Learning about Teaching and Learning through Modeling*. London: Falmer Press.
- Loughran, J. J., Hamilton, M. L., LaBoskey, V. K., and Russell, T. (2004). *International Handbook of Self-Study of Teaching and Teacher Education Practices*. Dordrecht: Kluwer Academic Publishers.
- Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record* **104**, 842–866.
- Schön, D. A. (1983). *The Reflective Practitioner*. New York: Basic Books.
- Van Manen, M. (1977). Linking ways of knowing with ways of being practical. *Curriculum Inquiry* **6**, 205–228.
- Zeichner, K. (1996). Teachers as reflective practitioners and the democratization of school reform. In Zeichner, K., Melnick, S., and Gomez, M. L. (eds.) *Currents of Reform in Preservice Teacher Education*, pp 199–214. New York: Teachers College Press.
- Dewey, J. (1910). *How We Think*. New York: Heath.
- Ghaye, T. (2000). Into the reflective mode: Bridging the stagnant moat. *Reflective Practice* **1**, 5–9.
- Hamilton, M. L. (ed.) (1998). *Reconceptualizing Teaching Practice: Self-Study in Teacher Education*. London: Falmer Press.
- Kraft, N. P. (2002). Teacher research as a way to engage in critical reflection: A case study. *Reflective Practice* **3**, 175–189.
- LaBoskey, V. K. (1994). *Development of Reflective Practice*. New York: Teachers College Press.
- LaBoskey, V. K. (1997). Teaching to teach with purpose and passion: Pedagogy for reflective practice. In Loughran, J. and Russell, T. (eds.) *Teaching about Teaching: Purpose, Passion and Pedagogy in Teacher Education*, pp 150–163. London: Falmer Press.
- Loughran, J. J. (2002). Effective reflective practice: In search of meaning in learning about teaching. *Journal of Teacher Education* **53**, 33–43.
- Lyons, N. and LaBoskey, V. K. (eds.) (2002). *Narrative Inquiry in Practice: Advancing the Knowledge of Teaching*. New York: Teachers College Press.
- McEntee, G. H., Appleby, J., Dowd, J., et al. (eds.) (2003). *At the Heart of Teaching: A Guide to Reflective Practice*. New York: Teachers College Press.
- Richert, A. E. (ed.) (1993). *Special Edition: Reflective Teaching and Teacher Education*. *Teacher Education Quarterly* **20**, 5–100.
- Schön, D. A. (1987). *Educating the Reflective Practitioner*. San Francisco, CA: Jossey-Bass.

Further Reading

- Bruner, J. (1996). *The Culture of Education*. Cambridge, MA: Harvard University Press.

Relevant Website

- <http://www.reflectivepractices.co.uk> – Reflective Learning, UK.

Teacher Education as Teaching for Understanding with New Technologies

M S Wiske and D E Spicer, Harvard Graduate School of Education, Cambridge, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Activity theory – A developmental conception of ways that technologies, practical action in the world, individuals, and communities shape and are shaped by the process of learning.

Collaborative curriculum design tool – An online environment that supports collaboration on the design of curriculum structured by the Teaching for Understanding framework. It includes examples, supportive prompts and resources, an integrated message board to facilitate communication with colleagues whom the designer adds to a design team, tools for supporting alignment of goals with instruction and assessment, and means for attaching related files and websites, and publishing the final design.

Teaching for Understanding (TfU) – An educational model, based on collaborative research conducted at the Harvard Graduate School of Education, including five elements: generative topics, understanding goals, performances of understanding, ongoing assessment, and collaborative reflective communities.

Understanding as performance – A concept of understanding as a capability to think with what one knows and apply it flexibly in varied situations.

WIDE World (Wide-scale Interactive Development for Educators) – An innovative professional development program developed and based at the Harvard Graduate School of Education. Its mission is to engage and support large numbers of educators across multiple levels of school systems in aligning their efforts to form professional networks that promote excellent instruction and improved performance by students, teachers, coaches, and leaders. It supports the application of research-based frameworks, such as Teaching for Understanding, by providing interactive online professional development courses and face-to-face support services to K-12 educators in the US and around the world.

Introduction

The proliferation of knowledge with new information and communication technologies, coupled with a clear

consensus from recent research on learning, dictate that education goes beyond simply transmitting information and procedures to focus on cultivating continuing learners. This requirement is certainly true for teacher education. A recent AERA Panel on Research and Teacher Education (2005) emphasized that teacher education must be an ongoing process of inquiry that does not end with certification but continues throughout the professional career, and across diverse educational settings. In this article, we examine the synergistic contributions of an explicit educational framework, called Teaching for Understanding (TfU), and networked technologies in helping teachers support the active learning that they and their students require in the twenty-first century. We use activity theory to conceptualize the roles of the TfU framework, networked technologies, and professional communities in continuous teacher education. Three examples illustrate how this approach can be applied to varied, interconnected stages and contexts of teachers' learning.

Mastery Teaching to Promote Understanding

Before analyzing the process of teacher education, we must consider the kind of learning that teachers need to promote. Public schools spread throughout industrializing democracies during the nineteenth century partly to prepare a broad range of citizens as contributors in increasingly hierarchical and mechanized workplaces and pluralistic societies. Teachers aimed to transmit basic skills through drill and rote recitation to groups of pupils who varied widely in age and experience. By the turn of the twentieth century, progressive educators criticized such rote learning and advocated more inquiry-based approaches that took account of students' varied interests. Despite a century of efforts, didactic instruction that Goodlad (1984) calls "frontal teaching," still prevails in most public schools around the world.

Recently, progressive reforms have gained new allies among some policymakers seeking to foster students' capability to think creatively, analyze and critique knowledge, reason systematically, apply knowledge flexibly, formulate and solve new problems, collaborate effectively in teams, and develop skills and habits of mind that promote continuous learning. While the dominant political discourse remains one of accountability, some leaders

in business and government have embraced an expanded conception of accountability that no longer views mastery teaching as primarily transmitting a canon of facts, theories, and accepted procedures. Enlightened leaders have come to view teaching as a process of helping students construct their own understanding and apply their knowledge through interaction using varied means, among diverse people, and within authentic contexts. In many countries, an emerging collective goal for education emphasizes deep and flexible student understanding.

What do we mean by understanding? The TfU project, a 6-year collaborative action research study convened at the Harvard Graduate School of Education during the 1990s, began with this same question. University-based researchers working with teacher researchers in schools came to define understanding as a flexible performance capability – being able to think with one’s knowledge and apply it creatively and appropriately in a range of circumstances. They conceived dimensions of understanding that encompass not only conceptual knowledge in one or more domains, but also methods of inquiry and reasoning, forms for presenting one’s ideas effectively with multiple symbol systems in varied situations, and appreciation of the purposes and limitations of learning in various disciplines.

Developing a flexible performance capability is as important for educators as it is for their students. Conceptual knowledge and familiarity with techniques are of limited use to a professional who also is not able to retrieve, synthesize, generate, and apply such knowledge effectively in practice and in interaction with others. Eraut (1994) whose research focuses on the nature and development of professional knowledge writes, “. . . learning knowledge and using knowledge are not separate processes but the same process.” Therefore, teaching for understanding is both a goal and an appropriate means for educating teachers.

TfU as Goal and Means

Harvard’s TfU project identified features of curriculum, pedagogy, and assessment that were common across examples of lessons that successfully developed learners’ understanding. As described by Blythe (1998) and Wiske (1998) researchers and teachers distilled these features into a coherent framework that crystallizes responses to core educational questions that are explained below.

What Topics Are Worth Teaching for Understanding? (Generative Topics)

Topics relate to important ideas in subjects, domains, or disciplines that are central to the required curriculum, connect both to learners’ experience and the teacher’s passions, can be approached through multiple entry points and/or media, and generate continuous inquiry.

What Should Learners Come to Understand? (Understanding Goals)

Goals focus on understanding key concepts, methods of inquiry and reasoning, purposes for learning, and disciplined forms of expressing understanding. Understanding goals aim beyond remembering information or practicing routine skills to target creative applications of knowledge. Effective teachers publicize explicit understanding goals that define the purpose of learning and serve as a basis for assessing progress.

How do Learners Develop and Demonstrate Understanding? (Performances of Understanding)

Effective teachers require students to stretch their minds as they develop and demonstrate understanding of target goals. Teachers sequence accessible yet challenging learning activities that engage students’ initial ideas and interests. Through guided inquiry, teachers progressively coach students toward more sophisticated, synthetic, and independent levels of performance. A rich range of learning activities allows students to work with varied media, through multiple intelligences, and in different participation structures so that all students are engaged and appropriately challenged.

How can the Students and Teachers Assess Achievements in Ways that Promote Further Learning? (Ongoing Assessment)

Teachers and others frequently analyze student performances in relation to public criteria that align precisely with understanding goals. Students may even help define the criteria for high-quality work. Frequent assessments include generating pertinent suggestions for improving both students’ work and the teacher’s planning. Multiple authentic audiences, including students themselves, parents, subject matter experts, and interested community members, may participate in assessment and feedback.

How can We Work Together Most Effectively within and beyond the Classroom? (Collaborative, Reflective Communities)

Teachers who promote understanding develop a culture of respect, reciprocity, and reflection with learners. Students build capacity to learn from and with one another by identifying common goals, developing a shared language and routines for reflective dialogue, and celebrating diversity and shared accomplishments.

The TfU framework shares many features with various formulations of twenty-first century learning, other syntheses of educational research, and characteristics

that most teachers recognize in their own good ideas and practices. These similarities underscore the generic applicability of this framework as a common, comprehensive language and structure for designing, conducting, and explicitly modeling effective pedagogy. In the examples presented later, we highlight the elements of TfU as mediators of teacher education activities whose goal is mastery of teaching that promotes students' deep and flexible understanding.

Networked Technology as a Mediator of Teacher Education

Education of preservice teachers often occurs far away from classrooms where teachers work, which hampers learning through authentic performances. Academic institutions, where most teacher preparation is based, are usually distant from the schools and other authentic settings where novice teachers might practice teaching. Skilled teacher mentors are scarce and travel is expensive, which limits opportunities for teachers to benefit from expert coaching.

Learning through reflective performance is equally difficult for in-service teachers. Finding time to analyze their work with colleagues and coaches is difficult. In hurried meetings, teachers are hard-pressed to reflect deliberately about lesson plans, teaching, and assessment. Corridor conversations and disorganized bits of paper are not easily captured in a form that promotes ongoing assessments or sustained collegial collaboration among teachers.

Networked digital technologies provide numerous advantages for overcoming these difficulties of teacher education and promoting the elements of effective TfU. Potential contributions of new technologies include:

Providing accessible templates and tools that structure work around common concepts, for example, interactive online environments for developing lesson plans. Enabling teacher learners and mentors to capture, store, and revise documents, videos, audio files, and images digitally in forms that can easily be shared, annotated, reviewed, analyzed, and co-constructed.

Rapidly transmitting written, as well as spoken, communication across distance and time, thereby encouraging more deliberative dialog and easing the limitations of synchronous meetings.

Making the work of students, teachers, and coaches visible to many so that individual insights benefit large groups.

Facilitating ongoing assessment with feedback from multiple sources. For example, teachers post work in the online environment where it can be reviewed, in relation to criteria, by teachers themselves as well as peers/coaches, and then easily revised.

In these ways, online technologies provide support for the use of a shared framework, collaborative exchange of materials and ideas, and sustained coaching, and enable ladders of professional development that stimulate educators to stay in the profession as they seek ways of contributing to and benefiting from contact with a professional community without abandoning direct work with students.

Teacher Education Viewed with Activity Theory

Activity theory offers a developmental view of the ways that conceptual frameworks and technologies, practical actions in the world, individuals, and social institutions shape and are shaped by one another in the learning process. This makes it particularly suitable for analyzing the roles of the TfU framework, networked technologies, and professional communities in teacher learning.

Building on the ideas of Vygotsky (1981) activity theory portrays activities such as teacher preparation and in-service professional development as continually changing, complex, self-organizing systems. It regards activity as a potential generator of both individual and organizational learning.

Engeström (1987) and more recently Roth and Lee (2007) used activity theory to describe how several distinct elements play crucial parts in mediating – literally being in the middle of – individual and group learning processes, as diagramed in the paradigmatic triangle of cultural–historical activity theory shown in **Figure 1**. The activity of developing expertise as a teacher is mediated by the division of labor and social rules that structure interaction within a particular community, such as the institutional setting of a school. The division of labor and rules are established and shift over long periods of time, serving collective motives that constitute the object or goal of activity. Of particular importance to our discussion of teacher learning are the mediating instruments in use within such communities. These instruments may encompass concepts (e.g., the components of a specific educational framework) as well as material tools (e.g., interactive digital tools and websites). Instruments and elements of community shape how the subject(s) or learner(s) orient, think, and perform in and for a particular activity.

Activity systems are in continual flux as participants negotiate tensions among different elements. Learning can be viewed as the resolution of contradictions among elements allowing the subject to take advantage of an expanded range of actions, including access to a wider repertoire of mediational means, taking on new roles, and forging robust ways of working within existing constraints.

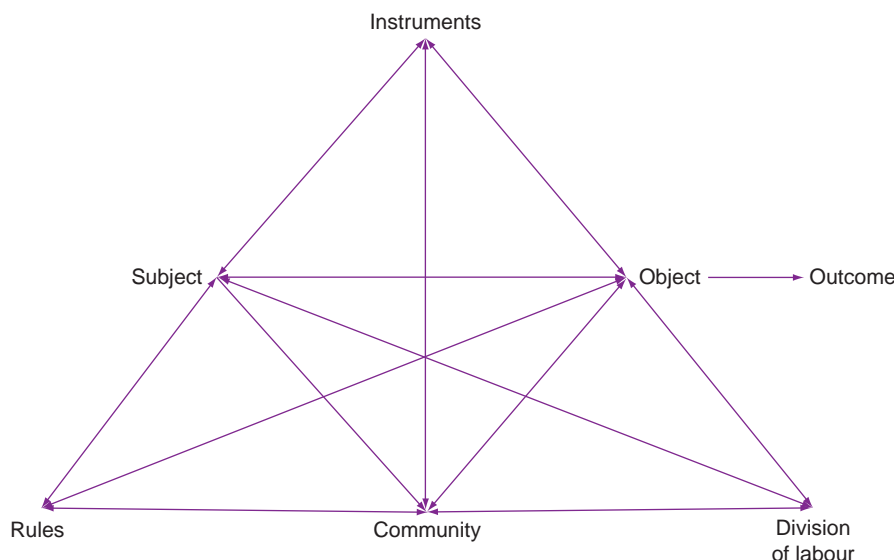


Figure 1 Engeström's model of the socially distributed activity system. Reproduced from Engeström, Y. (1987). *Learning by Expanding: An Activity-Theory Approach to Developmental Research*, p. 78. Helsinki: Orienta-Konsultit.

Consequently, as Engeström (1987) emphasizes, new objects or goals come into view and the system of activity as a whole is capable of evolving.

We use activity theory to examine teacher education within various systems of activity that orient subjects toward a goal or object, broadly characterized as mastery of teaching or pedagogical expertise. We highlight ways that two kinds of instruments – networked digital technologies and the TfU framework – expand the range of possible action for those pursuing mastery of teaching that promotes students' deep and flexible understanding. Activity theory illuminates the influence of these instruments in three different settings of teacher education: a graduate-level university course, an online course for practicing educators, and a school district.

Teacher Education in Varied Contexts

Teacher education involves continuous learning throughout an educator's career, including preservice teacher preparation, in-service professional development, and ongoing learning through inquiry and interaction within and among multiple levels of educational practice (e.g., relationships with administrators, researchers, and professional networks). Consequently, we chose examples that illustrate how TfU and new technologies promote mastery teaching aimed toward students' understanding in various phases and situations of teacher education.

Formal Teacher Education Course

One of the authors uses TfU as an implicit and progressively explicit mediating instrument in her course at the

Harvard Graduate School of Education called Designing Curriculum for Deep Learning. This one-semester course is for graduate students whose prior teaching experience varies considerably. Some are veteran teachers, some are professionals in other fields without prior teaching experience, and some are completing a teacher preparation program. The course applies and models the elements of TfU and the instructor emphasizes these parallels as the course progresses.

The generative topic of the course is designing curriculum that applies research-based principles to promote teaching and learning for understanding. Understanding goals for the course include understanding how to: (1) apply principles of effective teaching, learning, and curriculum design; (2) use technology to improve teaching and learning; and (3) participate effectively in a collaborative, reflective professional community. The core performance of understanding in this course is the design of a small unit of curriculum (1–3 weeks of lessons), developed in collaboration with a practicing educator, which explicitly exemplifies principles learned in the course. Students in the course are helped to identify a teacher partner whose setting, subject matter, and learners align with the student's own interests. Prior graduates of the course often volunteer to serve as a teacher partner, thereby enriching and extending the professional community of course participants. Designing a unit with and for the teacher partner adds to the generativity of the topic by anchoring the project in an authentic context of each student's choice and by enabling students to make a significant contribution in the world while benefiting from the consultation of an experienced teacher.

Students use an online collaborative curriculum design tool which is structured with elements of the TfU

framework to scaffold the design and refinement of a curriculum plan in collaboration with their teacher and other members of the class. This online tool includes specific support for designing each element of the framework and a message board where students exchange ideas with their collaborators. The teaching staff use the message board to provide ongoing assessment and feedback. As members of the class become more familiar with design criteria based on the theories from their reading, including TfU, and more skilled in contributing to a collaborative, reflective community, they are taught how to become effective self- and peer-evaluators.

The course is an activity system in which the concepts of TfU and the online curriculum design tool are mediating instruments that orient subjects (the students) toward the object of expertise in the design of curricula that promote understanding. This system includes a community that encompasses course members and collaborating teachers and the broader institutional setting of a graduate, professional school. The rules in this community include active application of theories to practice through design, reflection, and revision. The division of labor encourages a collaborative, reflective community whose members, including the instructor, contribute knowledge and provide feedback on colleagues' curriculum designs to improve learning.

Online Professional Development Program: Overlapping Systems of Activity

WIDE World at the Harvard Graduate School of Education offers online professional development courses that both model and teach participants how to apply research-based pedagogies. Several of the courses support designing and teaching with the TfU framework; some use the online collaborative curriculum design tool. During six biweekly sessions, participants learn to apply each of the four core elements of the TfU framework while engaging in participation structures that build the fifth element, a reflective learning community, both within the course and in on-site teams. Teachers may enroll in the online courses with two to three colleagues who work on the course assignments together. On-site teams in WIDE World courses are clustered into online study groups organized by subject area with approximately ten other teams or individuals who interact regularly with an online coach.

WIDE World online courses use the TfU language and networked technology to mediate in-service professional education for teachers:

1. The design of the courses models the same educational principles (TfU) that teachers are taught to use in their classrooms with their students.
2. Course assignments focus on trying out TfU ideas with students while the course is underway, thereby

building participants' capacity to think flexibly with what they learn and apply it in a range of contexts.

3. Experienced study group coaches offer tailored support and suggestions to help participants actively apply what they are learning in their own situation. Coaches also promote systematic exchanges and dialog among participants using the language of TfU to encourage development of a collaborative, reflective professional community within and across the online study groups.
4. By recruiting online coaches from graduates of its courses and by offering online courses for school leaders, WIDE World seeks to develop local capacity to sustain and extend the work initiated by participants in WIDE World courses.

In this setting of teacher education, activity in the online course occurs at the overlap of two other activity systems that exist on a longer time horizon – the institutional setting of the course participants' schools, on the one hand, and the institutional setting of the WIDE World project and the graduate school of education that serves as its home, on the other. The transient activity setting of the online course can be viewed as what Senge (1994) terms a practice field where participants work toward the object of mastery teaching. Networked technology provides ways of supporting experimentation with new mediational means (e.g., concepts of TfU and novel participation structures in both online and on-site teams) in ways that expand the possibilities for action within the participants' own classrooms. In this way, the activity setting of the course attempts to bring into view the object of mastery teaching for understanding within participants' schools.

Learning at Work through Continuous Inquiry: Embedded Systems of Activity

Researchers including Guskey (2002) and Little (1999) have shown that instantiating teachers' learning in new teaching practice hinges not just on individual learning but, most important, also on organizational learning. WIDE World's programs aim to stimulate continuous inquiry to strengthen practice in educational systems by providing related online courses for teachers, coaches, and leaders. In terms of activity theory, the subjects of this more-encompassing activity system includes people at all levels of educational practice that constitute schooling. In this example, the object on which we have focused in earlier examples – mastery teaching for student understanding – expands to encompass the practice and learning of all involved in promoting this goal. Such an object might be labeled mastery educational practice for students' deep and flexible understanding as the collective motive encompassing students, teachers, and leaders as subjects.

WIDE World uses TfU as a consistent model and common language across programs for educators working at all levels. Networked technologies provide flexible environments to support reflective, collaborative professional communities within and across role groups and innovating school systems. In this way, the activity system of a particular classroom can be viewed as part of the larger activity system of an entire school, which is in turn part of the more encompassing activity system of a larger organization – a school district or educational authority.

The Binghamton, New York, public schools have participated in a range of WIDE World programs over 3 years to promote ongoing learning for teachers, coaches, and leaders. District-level instructional leaders in Binghamton initially took WIDE World's TfU course. As they experienced learning with and applying the framework principles, they envisioned using online courses as a means of developing a shared language of teaching across the district. They focused especially on the comprehensive high school, which was under pressure to improve student performance. The administrators recruited teacher teams to enroll in WIDE World's online courses. Some teachers took additional online courses and were trained as WIDE World coaches. Administrators worked with two teams in the high school – a Math team and a team of English teachers – to develop a reference cohort that is expected to encourage other teachers to experiment with the framework and move toward greater coherence across the district.

During the past year, the Binghamton administrators have participated in an online action research seminar that WIDE World hosts. The seminar aims to develop local capacity in school systems to assess the impact of professional development as part of evidence-based continuous improvement in teaching and learning. Participants in the seminar collaborate on refining research questions, developing instruments and methods, analyzing the work of leaders, teachers, and students, and sharing results of this research. As the administrators collaborate with teachers to support and assess the impact of online professional development, they strengthen a local community whose language, roles, and interactions are shaped by the TfU framework.

Implications and Future Directions

The importance of schools as learning organizations and the ascendance of distributed professional networks as a mediator of professional learning is a consequence of shifts in work practices accelerated by information and communication technologies. New technologies serve as both stimulants of and means for restructuring schools

as complex, data-rich environments and making increasing demands professional teachers, new entrants, and established veterans alike, for higher quality and greater equity in their practices. The examples we have presented suggest an expansive view of possible settings for teacher education, ones that promote continuous learning aimed toward mastery teaching for students' deep and flexible understanding. The examples include the a traditional classroom in a graduate school of education linked to various school settings, the overlapping activity systems of school and university connected by networked technologies, and the embedded activity systems of organizational learning within a school system supported by online professional development.

One cross-cutting theme throughout this discussion is the importance of viewing professional learning as extending across the professional life span. Another theme is the use of the TfU framework as a specific, yet flexible, mediating instrument that conveys research on effective teaching in a usable form. It also provides a common language for educators as they continue to study, share, and improve practice.

Our experience with WIDE World, since its inception in 1999, suggests several trajectories of development in teacher education in the near future:

Global: Professional networks will become more global, enabling educators to learn from a wider range of colleagues as they develop, debate, and apply conceptual frameworks, technologies, and examples of practice across geographic and cultural distances.

Personal: Educators will be able to blend and extend opportunities for teaching and learning across their professional life span, connecting formal and informal learning through onsite experiences at university, conferences, and the workplace with online learning opportunities. Networked technologies will enable educators to play multiple roles in varied professional communities throughout their careers.

Organizational: Interweaving work across different settings of educational praxis will support more fluid collaboration among teacher educators in universities, schools, research organizations, policymaking agencies, and other places.

Technological: Continued developments in multimedia and online technologies will enable participants in teacher education to capture and share a wider range of artifacts both synchronously and asynchronously across distances, thereby evolving methods and resources for ongoing inquiry and improvement.

Conclusion

We analyzed how the TfU framework, networked technologies, and multiple layers of professional communities

can support connected, continuous teacher education in varied contexts. Activity theory gave us a lens to discern elements in an array of activity systems aiming toward a common goal of mastery teaching and learning to foster students' deep and flexible understanding. We portrayed three settings in which the TfU framework and networked technologies serve as mediating instruments to support this goal for educators. The examples of a graduate-level university course, an online course bridging school and university, and a district-wide initiative aimed at increasing educators' opportunities for learning illustrate ways in which teaching for understanding and new technologies support coherent, sustained teacher education at multiple levels across overlapping systems of activity. Such a panorama requires teacher education to be viewed in an expansive frame, one that entails organizational as well as individual learning, extends across the professional life span, and encompasses the wide variety of settings where the activity of learning occurs.

Bibliography

- AERA Panel on Research and Teacher Education. (2005). Studying teacher education. In Cochran-Smith, M. and Zeichner, K.M. *The Report of the AERA Panel on Research and Teacher Education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Blythe, T. (1998). *The Teaching for Understanding Guide*. San Francisco, CA: Josey-Bass.
- Engeström, Y. (1987). *Learning by Expanding: An Activity-Theory Approach to Developmental Research*. Helsinki: Orienta-Konsultit.
- Eraut, M. (1994). *Developing Professional Knowledge and Competence*. London: Falmer Press.
- Goodlad, J. I. (1984). *A Place Called School: Promise for the Future*. New York: McGraw-Hill.
- Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice* 8(3/4), 381–391.
- Little, J. W. (1999). Organizing schools for teacher learning. In Darling-Hammond, L. and Sykes, G. (eds.) *Teaching as the Learning Profession: Handbook of Policy and Practice*, pp 233–262. San Francisco, CA: Josey Bass.
- Roth, W.-M. and Lee, Y.-J. (2007). Vygotsky's neglected legacy: Cultural-historical activity theory. *Review of Educational Research* 77 (2), 186–232.
- Senge, P. (1994). *The Fifth Discipline Fieldbook: Strategies and Tools for Building a Learning Organization*. New York: Doubleday.
- Vygotsky, L. S. (1981). The genesis of higher mental functions. In Wertsch, J. V. (ed.) *The Concept of Activity in Soviet Psychology*, pp 144–188. Armonk, NY: ME Sharpe.
- Wiske, M. S. (ed.) (1998). *Teaching for Understanding*. San Francisco, CA: Josey-Bass.
- Cohen, D. K., McLaughlin, M. W., and Talbert, J. E. (eds.) (1993). *Teaching for Understanding: Challenges for Policy and Practice*. San Francisco, CA: Josey-Bass.
- Cole, M. (1996). *Cultural Psychology: A once and Future Discipline*. Cambridge, MA: Belknap Press of Harvard University Press.
- Eddy Spicer, D. *Learning by Contradiction: Two Developmental Perspectives on Learning in the Midst of Work*. Unpublished Qualifying Paper, Harvard University, Cambridge, MA.
- Engeström, Y., Miettinen, R., and Punamäki-Gital, R. L. (1999). *Perspectives on Activity Theory*. Cambridge, MA: Cambridge University Press.
- Huberman, M. (2001). Networks that alter teaching: Conceptualisations, exchanges and experiments. In Craft, A., Burgess, H., and Soler, J. (eds.) *Teacher Development : Exploring Our Own Practice*, pp 141–159. London: Paul Chapman Publishing in association with The Open University.
- Leont'ev, A. N. (1978). *Activity, Consciousness, and Personality*. Englewood Cliffs, NJ: Prentice-Hall.
- National Research Council (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academy Press.
- OECD (2001). *Schooling for Tomorrow: What Schools for the Future?* Paris: Organisation for Economic Co-operation and Development.
- OECD (2003). School leaders: Changing roles and impact on teacher and school effectiveness. Commissioned paper by the Education and Training Policy Division, OECD, for the activity attracting, developing and retaining effective teachers. Paris, France: Organisation for Economic Co-operation and Development. <http://www.oecd.org/dataoecd/61/61/2635399.pdf> (accessed December 2008).
- Partnership for 21st Century Skills. (2003). Learning for the 21st century: A report and mile guide for 21st century skills. Washington, DC. http://www.21stcenturyskills.org/downloads/P21_Report.pdf (accessed December 2008).
- Wells, G. and Claxton, G. (2002). *Learning for Life in the 21st century: Sociocultural Perspectives on the Future of Education*. Oxford, UK: Blackwell Publishers.
- Wertsch, J. V. (1981). *The Concept of Activity in Soviet Psychology*. Armonk, NY: ME Sharpe.
- Wiggins, G. and McGighe, J. (2005). *Understanding by Design*, Expanded 2nd edn. Alexandria: Association for Supervision and Curriculum Development.
- Wiske, M. S., Rennebohm Franz, K., and Breit, L. (2005). *Teaching for Understanding with New Technologies*. San Francisco, CA: Josey-Bass.

Relevant Websites

- <http://learnweb.harvard.edu> – The Collaborative Curriculum Design Tool at the Harvard Graduate School of Education.
- <http://wideworld.harvard.edu> – WIDE World (Wide-scale Interactive Development for Educators) at the Harvard Graduate School of Education.

Further Reading

- Chaiklin, S., Hedegaard, M., and Jensen, U. J. (eds.) (1999). *Activity Theory and Social Practice: Cultural-Historical Approaches*. Aarhus: Aarhus University Press.

TEACHER EDUCATION – PRESERVICE: THE KNOWLEDGE BASE

Contents

Experienced Teachers' Craft Knowledge

Moral Values in Teacher Education

Pedagogical Content Knowledge

Teacher Education and the Educational Foundations Knowledge Base

The Relationship Between Theory and Practice in Teacher Education

Experienced Teachers' Craft Knowledge

P C Meijer, Utrecht University, Utrecht, the Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Knowledge base of teaching – It refers to all profession-related insights that are potentially relevant to the teacher's activities.

Stimulated recall – Interview in which teachers explicate what they were thinking in response to a videotape of a lesson they have just given, aimed at retrieving teachers' knowing-in-action.

Teachers' craft knowledge – The accumulated wisdom and know-how teachers construct through experience, and which is the basis for dealing with everyday teaching situations.

Craft Knowledge and the Teacher Education Knowledge Base

Many established theories about teaching do not seem to correspond to what teachers know and do in practice. Past experiments with curricular and educational reforms often failed because they did not match teachers' ideas about what works in practice. These reforms were usually descriptions of what a teacher had to do in order to implement the reform successfully. In other words, they prescribed teacher behavior. There seems to be a lack of correspondence between ideas about what teachers should know and what they do know. In the field of teacher education, teacher educators are regularly confronted with student teachers who indicate that they see little congruence between the theories taught in the teacher education institute – described, for example, in the book by Reynolds (1989) *Knowledge Base for the Beginning*

Teacher – and what experienced teachers do and say in practice, the latter seeming to be of far more use for teaching.

Scholars, who emphasize the importance of what experienced teachers know for use in teacher education (e.g., Grimmer and MacKinnon, 1992; Loewenberg Ball and Cohen, 1999), suppose that experienced teachers are professionals. An important feature of professionals is that they share a common knowledge base (cf. Schön, 1983). Based on this feature, it is assumed that at least a part of experienced teachers' knowledge is shared by many teachers. However, no general agreement exists about the specifics of this type of knowledge. One question addressed in this article is whether there is an agreed-upon knowledge base that can – at least partly – be derived from the content of what experienced teachers know, and which can contribute to the enhancement of beginning teachers' teaching practice. Another question is how this might be done.

The knowledge base of teaching is described as all profession-related insights that are potentially relevant to the teacher's activities. The main function of such a knowledge base is to improve the practical arguments in the teacher's thinking process. Fenstermacher (1986) stated that the criterion of benefit of educational research in general should be "the improvement of practical arguments in the minds of teachers and other practitioners" (p. 44). These practical arguments refer to the beliefs and moral convictions that underlie a teacher's actions and eventually account for student learning. However, teachers' thinking processes and the practical arguments they use are highly complex, reflecting the nature of their work. For beginning teachers, the challenge is to develop their own practical arguments and craft knowledge in a well-considered way. It is important for them to

be acquainted with the knowledge of experienced teachers, how they draw upon this knowledge when teaching, and how this knowledge relates to more theoretical insights about teaching. This might help the development of their own craft knowledge and, as such, the development of their teaching practice.

Teachers' knowledge has been an object of study for several years. In most of these studies, the importance of this type of knowledge – also described as teachers' craft knowledge – for the education of new teachers is emphasized. Feiman-Nemser and Remillard (1996) indicated that research on teachers' craft knowledge "...further reveals the complexities and uncertainties of interactive teaching and the need for considerable thinking in action" (p. 76).

Loewenberg Ball and Cohen (1999) stated that, in order to help student teachers to develop into professional practitioners, teacher education programs should include attention to the kinds of knowledge, skill, and other qualities necessary for professional teaching. According to Lottero-Perdue and Brickhouse (2002), learning a vocation is only possible when learners have access to the expertise of old-timers. This allows beginners to eventually participate fully in the activities of the communities of practice. In their study, an important conclusion was that adequate social access to knowledge is an important factor for acquiring competence in the workplace. Adequate social access to knowledge was linked in their study to an organizational expectation that information can and should be shared, through interaction and direct communication, and through the use of written sources.

The concept of teachers' craft knowledge is to a certain extent trivial in the literature. From the beginning, research into this type of knowledge has been grounded in a high degree of regard for teachers. Research into teachers' personal practical knowledge resulted from dissatisfaction with the way teachers were viewed in most research on teaching that was done until then, in which teachers' knowledge was given little credit. Many researchers followed by taking teachers seriously as holders of (practical) knowledge, and aimed at giving teachers a voice in research on teachers. However, other authors point to the problematic character of investigating this type of knowledge, and sometimes even warn against taking this knowledge seriously. They claim, for example, that teachers' craft knowledge also contains fragmentary and inaccurate opinions. It is also highly subjective in nature, and its relationship to what teachers actually do in practice is doubtful. They prefer to focus on situation-independent claims about teacher knowledge and behavior. Most teacher education programs, however, contain a mixture of general and situation-independent knowledge about teaching, on the one hand, and teachers' craft knowledge, on the other – the latter pertaining to, at minimum, a mentoring program of some kind.

The question of the contents and nature of the knowledge underlying teachers' professional behavior is essential. In this article we elaborate upon these questions: What are the content and nature of this type of knowledge? Why is it essential to incorporate it in the teacher education knowledge base? How might this be done? Research findings show that experienced teachers' craft knowledge provides information about the types or categories of knowledge needed to teach well, and about how experienced teachers use and integrate this knowledge while teaching. More specifically, teachers' craft knowledge provides information about the way research-based knowledge and teaching practice are related and combined.

Teachers' Craft knowledge: Nature and Content

The term craft knowledge refers to the accumulated wisdom and know-how teachers construct through experience, and which is the basis for dealing with everyday teaching situations. Grimmer and MacKinnon (1992) labeled this type of knowledge as a form of professional expertise and, in line with Schön (1983), found it to be embedded in reflection-in-action. Reflection-in-action plays a central role in professional performance and its development. It refers to a mental process in which practitioners combine elements of their own knowledge in order to deal with uncertain, unique, or complex situations. In teaching, such situations are ongoing and require an immediate response. This is also an important reason for the difficulties beginning teachers report when they try to act upon theories they encounter in teacher education. Teaching situations are complex and multifaceted. Dealing with such situations requires multiple kinds of knowledge to be brought together in an integrated way. However, such situations require split second responses. Integration of knowledge in such a short period of time requires routinization that beginning teachers do not have. This explains why beginning teachers can analyze such situations afterward (i.e., reflection-on-action), but find it difficult to do so while teaching (i.e., reflection-in-action, based on craft knowledge). Although student teachers acquire knowledge in teacher education, using that knowledge in action is an old problem in learning to teach.

Teacher education programs often have separate courses about various knowledge bases, such as subject matter or child psychology; however, the task of teaching requires that these knowledge bases, or categories of knowledge, be integrated and simultaneously accessible. Observing how experienced teachers behave, sitting in the back of the classroom, does not automatically help student teachers develop a deeper understanding of how the various knowledge bases are integrated in teaching.

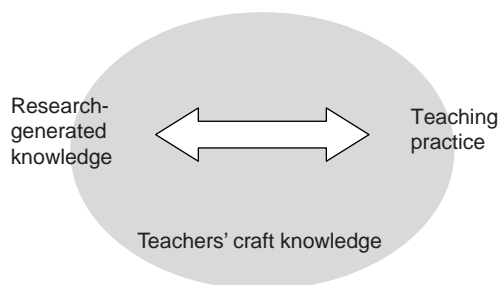


Figure 1 Relationship between teachers' craft knowledge, and research-generated knowledge and teaching practice.

Furthermore, simply copying the behavior of experienced teachers will most probably result in conservatism. An important use of research on teachers' practical knowledge is that it can provide student teachers "with a realistically complex picture of the cognitive aspects of teaching" (Clark and Lampert, 1986) that goes beyond the "how" of teaching and provides insight into the "why" underlying teaching.

An essential feature of teachers' craft knowledge is that it involves research-generated knowledge and specific knowledge about teachers' own teaching practice, and, most interesting, knowledge of how these two are related. Grimmer and MacKinnon (1992) stated that teachers' knowledge is derived "...from a considered response to experience in the practice setting, and, though related to knowledge that can be taught in the lecture hall, it is formed over time in the minds of teachers through reflection" (p. 387).

Figure 1 shows the general view on this issue: teachers' craft knowledge provides information about how experienced teachers relate research-generated knowledge to their teaching practice, and vice versa. The arrow in **Figure 1** represents this relationship.

Teacher education programs were criticized in the past for leaving a gap between theory (i.e., research-generated knowledge) and practice, because they were not able to link the two. Teacher education programs consisted of two parts: an institutional part concerned with research-generated knowledge, and a student practice part. Attempts to link the two were superficial, and mostly the student teachers' responsibility. Experienced teachers' craft knowledge provides information about just that relationship, and is valuable as such. What is known about what teachers know is described through an exploration of the nature and content of teachers' craft knowledge in the next two subsections.

Nature

In the literature on teachers' knowledge, a host of names are used: craft knowledge, case knowledge, situated

knowledge, (personal) practical knowledge, and tacit knowledge, to name but a few. As Fenstermacher (1994) remarked, these names do not necessarily refer to different types of knowledge, but to the fact that the contexts in which they are used differ. A great deal of the variety in conceptualizations of teachers' craft knowledge can be explained when underlying assumptions about the nature of this type of knowledge are identified. Several characteristics of teachers' craft knowledge can be identified that have been emphasized in studies of this concept:

1. it is personal, which means that each teacher's knowledge is to some extent unique;
2. it is contextual, meaning that it is defined in, and adapted to, the classroom situation;
3. it is based on (reflection on) experience, indicating that it originates in, and develops through, experiences in practice;
4. it underlies teaching practice;
5. it is mainly tacit, indicating that many teachers are not used to articulating their knowledge;
6. it is related to teaching practice, on the one hand, and to research-generated knowledge, on the other; and
7. it is content related, meaning that it is connected with the subject taught.

The last aspect, the content-relatedness of teachers' craft knowledge, has a number of serious implications for the discussion about the inclusion of this type of knowledge in the teacher education knowledge base. This becomes even more evident when what is known about the content of teachers' craft knowledge is taken into account.

Content

In many studies of teachers' craft knowledge, researchers have focused on its content, generating insights into categories which make up teachers' knowledge and beliefs. Shulman (1987) was one of the first to investigate the kinds of knowledge that teachers possess and that underlie their actions; he developed a number of domains and categories of teacher knowledge, which other researchers have used, expanded, and refined. These categories are:

- content knowledge;
- general pedagogical knowledge, with special reference to those broad principles and strategies of classroom management and organization that appear to transcend subject matter;
- curriculum knowledge, with a particular emphasis on the materials and programs that serve as tools of the trade for teachers;
- pedagogical content knowledge, that special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding;

- knowledge of learners and their characteristics;
- knowledge of educational contexts, ranging from the workings of the group or classroom, the governance and financing of school districts, to the character of communities and cultures; and
- knowledge of educational ends, purposes, and values, and their philosophical and historical grounds (Shulman, 1987: 8).

Studies in this area of research are usually conducted in a specific subject-matter context, such as English, social studies, mathematics, and science. Particularly in secondary science education, the concept of pedagogical content knowledge remains topical, and seems to refer mostly to the type of knowledge that is directly related to teachers' actual practice when teaching their subject. What is striking in most of these studies is that the content of teachers' craft knowledge is seen as an integration of several categories of knowledge. This is in line with Grimmer and MacKinnon's (1992) definition of craft knowledge, in which it is clear that this kind of knowledge encompasses several elements of the kinds of knowledge distinguished by Shulman and his followers. Craft knowledge in teaching, they stated, is essentially teachers' accumulated wisdom, derived from their understandings of the many dilemmas inherent in teaching. It represents intelligent and sensible know-how in the action setting. It is "...a particular form of morally appropriate and sensible know-how that is constructed by teachers...in the context of their lived experiences and work around issues of content-related and learner-focused pedagogy" (p. 396).

Part of teachers' craft knowledge, therefore, is knowledge about the various knowledge areas. All teachers have knowledge about learners, the content, the curriculum, etc. The differences found in the various studies concern what teachers know about these areas and how they integrate their knowledge about these different areas. Many researchers describe these differences in a classification of types of knowledge, in which teachers' craft knowledge is described along similar categories (often based on the categories that Shulman developed and that were described above). The content of these categories and the way the categories are related is quite uniform within the types distinguished, but shows substantial differences across the types. To illustrate such differences, **Table 1** shows descriptions of three teachers' craft knowledge about a category described as student learning and understanding. The descriptions are from a qualitative study on experienced teachers' craft knowledge about teaching reading comprehension to 16- to 18-year-old students (Meijer, 1999). The right column contains citations of the three teachers that illustrate the differences found. Other qualitative studies show similar patterns in teachers' craft knowledge (or practical knowledge, pedagogical content knowledge, etc.).

In sum, the results of studies on teachers' craft knowledge show not only common features in this type of knowledge, but also substantive differences, reflecting that this type of knowledge is idiosyncratic by nature. This variety and idiosyncrasy may pose problems as well as challenges in the search for craft knowledge underlying teaching, particularly when the aim is to incorporate this type of knowledge in teacher education. There is no one straightforward way to do so when the idiosyncratic nature of teachers' craft knowledge is taken into account. The next section contains some strategies to involve this type of knowledge in the teacher education knowledge base that are described in the literature.

How Can Teachers' Craft Knowledge be Involved in the Teacher Education Knowledge Base? – Eliciting Teachers' Knowing-In-Action

Leinhardt (1990) described craft knowledge as encompassing the wealth of teaching information that very skilled practitioners have about their own practice. She argued that this type of knowledge needs to be merged with theory in order to design a complete, valid assessment of teachers. Also beyond the field of teacher assessment, the need to combine the craft knowledge of (expert) teachers with research-based theories is felt, especially in the field of teacher education. The question is how such knowledge can be accessed and processed in order to aid student teachers' professional development. In a recent attempt to capture what beginning teachers should learn, Hammerness *et al.* (2005) stated first that "wisdom can't be told" (p. 371). Expert teachers cannot simply tell what they know in a way that is ready to be used by beginning teachers, mainly because of the tacit nature of this type of knowledge. Second, Hammerness *et al.* described the issue of knowing but failing to have this knowledge guide one's actions. Teachers do not automatically draw on knowledge acquired during, for example, coursework to help them solve problems in practice. Taking these two issues into account, the focus needs to be not on the knowledge that teachers have about their teaching, but on the knowledge that is embedded in their teaching, and that directly guides their actions. This is a crucial assumption behind research on teachers' craft knowledge: teachers' knowing is expressed in their teaching. Schön (1983) argued that to understand the practice of professionals, we need to understand their knowing-in-action. This is comparable to Leinhardt's (1988) idea that teachers' knowledge should also be examined in use in order to enable understanding of the complex process of teaching. Leinhardt stated that teachers' cognitions and actions should be investigated while they are teaching, because,

Table 1 Descriptions of three teachers' craft knowledge about the category "Student learning and understanding"

Teacher	Description of category "Student learning and understanding"	Some citations from semi-structured interviews and stimulated recall interviews
Subject-oriented teacher (Craft knowledge type A)	The knowledge and skills that are needed in reading comprehension are present in students in rudimentary form. However, they do not see what is important in a text, so they definitely need the teacher to help them. Reading comprehension is very difficult for students; some will never learn it because they do not see the main idea in texts, they just see "words, words, words" and they lack vocabulary. Furthermore, students have difficulties reading texts that do not fit in with their environment. Not all students have the same working style: some use the teacher's own method (as they should), but others, for reasons of their own, have their own way of working. During the lesson, the teacher was often wrong about what the students did or did not understand. Quite often he expressed surprise if students did not show understanding where he had expected it.	<p>...if you let them work for themselves, they often don't understand at all what it [i.e., the text] is about...they pick up a few things, but they don't see what's important and what's not, so they just pick up what amuses them.</p> <p>...I was thinking about the problems you have to deal with as a teacher. You can choose the wrong subject, you can have children in the classroom who don't do their best, are neither motivated nor interested, or children who are too stupid to understand a certain text, or children who are motivated, but lack vocabulary. How can you deal with all this? I thought, well, that's disappointing, that one student thinks it [i.e., the text] is about football, and another thinks something else. But no one was aware of what it was really about, no one, no one understood anything at this moment... [and at a later moment:] They still didn't understand what it was about... and it bothered me, I thought, I don't want to hear the same wrong answers again.</p>
Student-oriented teacher (Craft knowledge type B)	Students barely master the necessary skills for reading comprehension: their reading level is low. Able students are intelligent and gifted, whereas students who have difficulties with reading comprehension are always asking the wrong questions and read incorrectly. The teacher thinks it is lack of talent that causes this. On the other hand, he also referred to the students' environment when assessing their abilities.	<p>[Why a student has difficulties understanding texts]...I don't know, different science, I can't tell, I don't understand. It's like a fiddler, how come not everyone is able to play the violin in the same way?</p> <p>In general, the students' reading level is low. But that's not their fault...one does notice that there is a cultural decrease, I think that societal influences are important here...And that pushes young children not to take an interest in non-material matters...And it leads to the fact that they are not capable of concentrating long, and because of that are not able to read for long periods of time.</p>
Student-learning-oriented teacher (Craft knowledge type C)	Students at this age and level have a reasonable vocabulary and read fairly well, but lack general knowledge. If the level of difficulty of the texts is increased gradually and students use (some of) the strategies the teacher gives them, every student can learn to comprehend texts. However, students who are poor in reading comprehension tend to be chaotic in their approach to a text, which interferes with their acquisition of reading strategies. During the lesson, this teacher appeared to be focused on student learning and understanding, and tried to anticipate problems they might run into.	<p>They are chaotic. This becomes manifest, for example, in a summary that isn't a summary: Too many words, difficulty in capturing the essence in a text, difficulty in finding key sentences. You might say that, for such a student, it's very difficult to separate essentials and side-issues.</p> <p>When a student is reading out loud like this, I just read along with him...and here I pay attention to: does he read well, does he read as if he understands what he is reading? Here, at this moment, I knew when I was preparing this lesson that this would cause them problems, this 'pluralism in linguistic usage'...so here I thought, yes, yes, I knew this was going to happen.</p> <p>At that moment, when they were working for themselves, I had the opportunity to read the text for myself again. And I tried to read it through their eyes, and tried to picture what problems they would run into.</p>

at that moment, knowing and acting are inseparable. Brown *et al.* (1989) argued that all knowledge is produced in activities and situations. Knowledge is, therefore, situated in actions or, as Leinhardt defined it, it is "embedded in the artifacts of a context" (Leinhardt, 1988: 148). This implies, according to Brown *et al.*, that knowledge should be investigated in the situation in which it is being used. They stated that the examination of knowledge can be inappropriate, or even invalid, if this knowledge is stripped from the way it is used.

To involve teachers' craft knowledge needs in teacher education might imply two strategies. First, student teachers can gain access to experienced teachers' craft knowledge that has been examined by others and, for example, written down in (multiple) case studies or typologies, such as illustrated in **Table 1**. Second, student teachers might investigate experienced teachers' craft knowledge themselves. Some examples of how experienced teachers' craft knowledge might be involved in teacher education and the conditions that have to be met in order to do so are described below.

Cases

The findings of several studies indicate that detailed descriptions of the contents of teachers' craft knowledge are relevant to student teachers. Researchers and teacher educators have suggested ways of incorporating such descriptions, often as cases, narratives, or stories, in teacher education. Related to this, Hammerness *et al.* (2005) describe "analysis of teaching" as a crucial factor in teacher education. Cases not only concern written cases, but also include videos of teaching practices. Databases with videotapes of classroom teaching exist at many teacher education institutes, for example, at the University of Michigan, at the Carnegie Foundation's Knowledge Media Lab, and at the New Teacher Center at the University of California, Santa Cruz, and there is a national database of teaching practice in the Netherlands. A related initiative is a website, which has a digital exhibition of the Teachers College Record highlighting websites on which multimedia are used to document teaching and learning in classrooms in California, Philadelphia, and New York City, from elementary and high schools, and in math, language arts/English, and social studies. All these databases include interviews with the teachers, who comment on their videos by giving arguments for choices they made. Finally, links to relevant theories are provided. In this way, student teachers can easily access many teaching practices and are provided with information about the knowledge behind what is shown in the teaching fragments and how this relates to relevant theories (i.e., research-generated knowledge).

Stimulated Recall

Research on teachers' craft knowledge has also resulted in a range of creative methods to elicit this type of knowledge, such as various interview techniques, mind-mapping assignments, and responses to videos. One instrument of particular interest is the stimulated recall interview, in which, teachers explicate what they were thinking in response to a videotape of a lesson they have just given. The information retrieved in this way closely relates to Schön's idea of knowing-in-action, and to Leinhardt's plea for investigating teacher knowledge in use. The instrument has its origin in studies by Bloom (1953) and has been used by several researchers since then.

Using this instrument allows student teachers to analyze the craft knowledge of an experienced teacher. The stimulated recall interview is in fact a substitute for the thinking-aloud technique. However interesting a thinking-aloud technique would be in capturing a teachers' thinking while he or she is teaching, this technique is clearly not suitable as it would interfere with lessons. In a stimulated recall interview, a videotape of a lesson a teacher has just given is used to aid the teacher's recall of interactive thoughts during the lesson and to stimulate him or her to relive the lesson. The stimulated recall interview can be used to make much of teachers' tacit thinking explicit and to elicit cognitions underlying their observable actions. Such cognitions are characterized by the following:

1. split-second thoughts;
2. tied to the specific context (i.e., the lesson);
3. closely connected to teachers' knowledge and beliefs, on the one hand, and
4. closely connected to classroom practice, on the other; and
5. integrative in nature.

These characteristics might be problematic when the aim is to generate objective statements about the cognitions teachers in general have while teaching, but they make such cognitions all the more interesting for student teachers who are in the process of trying to understand what is going on in the classroom.

There are several ways to use the stimulated recall technique in teacher education. For example, Ethell and McMeniman (2000) described a way in which the technique can be systematically planned in a teacher education context. They developed the intervention "making explicit and gaining access to the thinking underlying expert teaching," which involved a series of workshops with student teachers structured in four steps:

- Step 1: observation and inference.
- Step 2: making explicit personal theories.
- Step 3: expert teacher's reflection on action (i.e., responses in the stimulated recall interview).
- Step 4: contrasting novice and expert reflections.

Meijer *et al.* (2002) also described the stimulated recall interview as a tool for student teachers to elicit their experienced mentor teachers' craft knowledge. They described the value of systematic examination of experienced teachers' practical knowledge by student teachers as, among other things, the window it provides on the thoughts behind teachers' observable teaching. In addition, it gives opportunities to relate these underlying thoughts to the theoretical and more abstract notions student teachers are confronted with in teacher education.

Conclusion

Examining experienced teachers' craft knowledge provides student teachers with opportunities to better understand other teachers' teaching and their own teaching by (1) experiencing the relationship between a teacher's craft knowledge and his or her teaching, (2) learning about the relationship between other teachers' craft knowledge and their own (developing) craft knowledge, and (3) exploring the relationship between (experienced) teachers' craft knowledge and more theoretical notions they learn about in teacher education. This understanding can help student teachers develop their own teaching in a more conscious way, as they can underpin the choices they make in their teaching more thoughtfully.

In sum, the conditions for incorporating teachers' craft knowledge in the teacher education knowledge base in a sensible way are the following:

1. the knowledge has to pertain to teachers' knowing-in-action;
2. this knowledge needs to be related to the student teachers' own (developing) craft knowledge;
3. it also needs to be related to research-generated knowledge and theories; and
4. it then needs to be related to student teachers' future actions and practical arguments.

Final Remarks

Teachers' craft knowledge is one of the sources of the teacher education knowledge base. It can provide valuable insights into the complexities of teaching and contribute to the professional development of beginning teachers, as we have described in this article. It is necessary, however, to take into account the principle of the conservatism of practice as described by, for example, Loewenberg Ball and Cohen (1999) and the process described by Zeichner and Tabachnick (1981), who found that liberal and progressive insights student teachers encountered in teacher education were washed out by daily school experiences

and conversations with their experienced colleagues. One of the most essential components of the process of learning to teach is learning how to learn in and from practice. Linking theory and practice is vital. Student teachers need to learn various strategies for linking theory and practice. Experienced teachers' craft knowledge is just one valuable source in this process.

Bibliography

- Bloom, B. S. (1953). Thought processes in lectures and discussions. *Journal of General Education* 7, 160–169.
- Brown, J. S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* 18(1), 32–42.
- Clark, C. and Lampert, M. (1986). The study of teacher thinking: Implications for teacher education. *Journal of Teacher Education* 37(5), 27–31.
- Ethell, R. G. and McMeniman, M. M. (2000). Unlocking the knowledge in action of an expert practitioner. *Journal of Teacher Education* 51(2), 87–101.
- Feiman-Nemser, S. and Remillard, J. (1996). Perspectives on learning to teach. In Murray, F. B. (ed.) *The Teacher Educator's Handbook. Building a Knowledge Base for the Preparation of Teachers*, pp 323–380. San Francisco, CA: Jossey-Bass.
- Fenstermacher (1986). Philosophy of research on teaching: Three aspects. In Wittrock, M. C. (ed.) *Handbook of Research on Teaching*, 3rd edn., pp 37–49. New York: MacMillan.
- Fenstermacher, G. D. (1994). The knower and the known: The nature of knowledge in research on teaching. *Review of Research on Teaching* 20, 1–54.
- Grimmett, P. P. and MacKinnon, A. M. (1992). Craft knowledge and the education of teachers. *Review of Research in Education* 18, 385–456.
- Hammer, K., Darling-Hammond, L., and Bransford, J. (2005). How teachers learn and develop. In Darling-Hammond, L. and Bransford, J. (eds.) *Preparing Teachers for a Changing World. What Teachers Should Learn and Be Able to Do*, pp 358–389. San Francisco, CA: Jossey-Bass.
- Leinhardt, G. (1988). Situated knowledge and expertise in teaching. In Calderhead, J. (ed.) *Teachers' Professional Learning*, pp 146–168. London: The Falmer Press.
- Leinhardt, G. (1990). Capturing craft knowledge in teaching. *Educational Researcher* 19(2), 18–25.
- Lottero-Perdue, P. S. and Brickhouse, N. W. (2002). Learning on the job: The acquisition of scientific competence. *Science Education* 86(6), 756–782.
- Loewenberg Ball, D. and Cohen, D. K. (1999). Developing practice, developing practitioners. Toward a practice-based theory of professional education. In Darling-Hammond, L. and Sykes, G. (eds.) *Teaching as the Learning Profession*, pp 3–32. San Francisco, CA: Jossey-Bass.
- Meijer, P. C. (1999). *Teachers' Practical Knowledge. Teaching Reading Comprehension in Secondary Education*. Leiden The Netherlands: Doctoral dissertation, Leiden University.
- Meijer, P. C., Zanting, A., and Verloop, N. (2002). How can student teachers elicit experienced teachers' practical knowledge? Tools, suggestions and significance. *Journal of Teacher Education* 53(5), 406–419.
- Reynolds, M. C. (1989). *Knowledge Base for the Beginning Teacher*. Oxford: Pergamon.
- Schön, D. A. (1983). *The Reflective Practitioner. How Professionals Think in Action*. New York, Basic Books.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review* 57(1), 1–22.
- Zeichner, K. M. and Tabachnick, B. R. (1981). Are the effects of university teacher education 'washed out' by school experience? *Journal of Teacher Education* 17(3), 7–11.

Further Reading

- Calderhead, J. (1996). Teachers: Beliefs and knowledge. In Berliner, D. C. and Calfee, R. C. (eds.) *Handbook of Educational Psychology*, pp 709–725. New York: Macmillan.
- Cochran-Smith, M. and Lytle, S. L. (1999). Relationships of knowledge and practice: Teacher learning in communities. In Iran-Nejad, A. and Pearson, C. D. (eds.) *Review of Research in Education*, vol. 24, pp 249–305. Washington, DC: AERA.
- Connelly, F. M. and Clandinin, D. J. (1985). Personal practical knowledge and the modes of knowing: Relevance for teaching and learning. In Eisner, E. (ed.) *Learning and Teaching the Ways of Knowing*, pp 174–198. Chicago, IL: University of Chicago Press.
- Hoyle, E. and John, P. D. (1995). *Professional Knowledge and Professional Practice*. London: Cassell.
- Loughran, J., Mulhall, P., and Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching* **41**, 370–391.
- Meijer, P. C., Beijaard, D., and Verloop, N. (2002). Examining teachers' interactive cognitions using insights from research on teachers' practical knowledge. In Sugrue, C. and Day, C. (eds.) *Developing Teachers and Teaching Practice*, pp 162–178. London: Routledge/Falmer.
- Shimahara, N. K. (1998). The Japanese model of professional development: Teaching as craft. *Teaching and Teacher Education* **14**, 451–462.
- Van Driel, J. H., Verloop, N., and De Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching* **35**(6), 673–695.
- Verloop, N., Van Driel, J. H., and Meijer, P. C. (2001). Teacher knowledge and the knowledge base of teaching. *International Journal of Educational Research* **35**, 441–461.

Relevant Websites

- <http://www.tcrecord.org> – Making Teaching Public: A digital exhibition.
- <http://paletvo.rdmc.ou.nl> – *Ruud de Moor Centrum* for the professionalization of teachers (a department within the *Open Universiteit Nederland*).
- <http://www.carnegiefoundation.org> – The Carnegie Foundation for the Advancement of Teaching.
- <http://www.newteachercenter.org> – The New Teacher Center.

Moral Values in Teacher Education

W Veugelers, University for Humanistics, Utrecht, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Values are embedded in the curriculum, the school culture, and the behavior of the teacher. The question is not whether teachers display values in education, but rather which values they display, and how the teachers work with values in their teaching. Values are embedded in educational practices, in curriculum materials, and in discourses in school. Values are the ideological flavor of teaching. Values can be taught explicitly, but values are always interwoven in regular teaching practices. Sometimes these hidden values become conscious when teachers reflect on their educational practice. But generally, the values remain hidden and do their work and influence the identity development of students.

Moral values are values that express ideas about the good life. There is constant debate about which values constitute the good life. Aristotle introduced the idea in his *Nicomachean ethics*, and there is continuous debate in philosophy and social science about which values are important, and about what exactly is meant by a certain value. Based on their ideas about living together in society, Rawls and Kohlberg advocated the value of justice. Gilligan and Noddings criticized the cognitive emphasis in justice, and stressed the affective side of living together and argued for the value of care. From a multicultural perspective, Banks espoused the value of diversity – the appreciation of difference. Moral values are quite abstract but become meaningful in concrete practices and narratives. Stressing the same abstract moral value can still result in different concrete interpretations of the value and in distinct practices. Moral values should therefore be studied in their context.

The development of people's values is conceptualized in different ways in both academic work and in practice, and is embedded in more extended concepts about human development. For example, the development of values is part of personal and social development, moral development, identity development, and citizenship development. What these concepts have in common is that the focus is on the development of the individual. Values are expressed in attitudes that become visible in opinions and concrete behavior. Attitudes based on moral values refer to being in society, being with others, and the reflective view on one's own identity.

Adding value to development implies that human development is seen not as a natural biological process, but as a process of giving meaning to the world. It is a process of growing into cultural practices and positioning oneself by articulating what is valuable; it is a personal reconstruction of cultural practices and narratives. It is making sense of your life.

Different Concepts of Moral Development

People differ on the pedagogical and moral goals of education, on how educational institutes can work with values. In educational policy, both in research and in practice, there are different ideas about the desired outcomes of education in the personal, social, and moral domains; about how to work with values in education; and about the role of schools. As in the character education movement, the focus can be on teaching the right values, or as in the Kohlberg tradition of moral development, the focus can be on developing skills to reflect on values and behavior. Youngsters in the Kohlberg tradition are supported in developing cognitive skills to articulate their own values. In this tradition, analyzing and communicating values are important tools for moral judgment. Despite Kohlberg's emphasis on the value of justice and on the educational concept of just community schools, Kohlberg has been criticized by the character education movement for stimulating value relativism. According to these critics, Kohlberg and his followers were not clear about the moral goals they wanted to attain.

In turn, the character education movement stressed the importance of good behavior and virtues. This was criticized by the Kohlbergians because the character education movement disregards the critical autonomy that people as human beings can have in making moral judgments. A second criticism was that they saw moral values not as situated but as fixed values. For the character education movement, living the good values is more important than reflective and communicative competences in the moral domain. There was strong polarization in the early 1990s between the proponents of a value approach and a more developmental approach. Nowadays, many scholars are aware of the importance of both values and competences to reflect and communicate, and they attempt to integrate both values and competences (Solomon *et al.*, 2001; Veugelers and Vedder, 2003). Values are not seen as purely abstract and fixed, but as situated, and analytical and communicative skills are necessary in order to work reflectively. As an educator, you may express your values because displaying the values you find important is unavoidable. However, what is important is the room you leave for others to articulate their own values.

Moral development is linked with many aspects of the individual, such as the social being and citizenship. Moral development can be seen as more fundamental, as driving the concrete attitudes and behavior. Moral values tend to

be rather abstract, but they become concrete in narratives and practices. Because moral values are so abstract and fundamental, they are difficult to teach and to change. Given the fact that moral values are important for human development, education cannot fail to pay attention to them. Even if education does not wish to give explicit attention to moral values, the values do their work through the hidden curriculum. In, for example, the work of Durkheim, Bourdieu, and Willis, the sociology of education has shown that values always do their work. If a teacher opts to keep the values hidden in educational practice, the teacher is, in fact, also making a moral choice. When values remain hidden and are not the subject of reflection, the emphasis is more on adaptation than on a critical moral judgment, and students are not then challenged toward a more self-regulated personal moral development.

Moral Education in Contemporary Education

The task of moral education has, in recent years, been reinvented in many countries. Currently, the most used concept is that of citizenship development. The concept of citizenship has moved beyond the political level to the social, the interpersonal level, and even the intrapersonal level – it is about how you live your life. Citizenship development is linked to identity development. Processes of nation building and citizens' identification with civil society and political structures are arguments in favor of citizenship education on the political level. Social cohesion and public behavior are arguments on the social level. Self-regulation on the interpersonal and intrapersonal level is a necessary condition for both the social and the political levels. The political, the social, the interpersonal, and the intrapersonal levels are linked by values, narratives, competences, and practices. Education could choose to say that we only work on the political level, and that we leave the social and the interpersonal level to the private world of the students, but this would then mean that the linkage of the four levels remains unreflective. The political level then continues to influence the social and interpersonal levels, but only in a hidden way.

Citizenship development can aim at different social, cultural, and political practices: at adapting to society by emphasizing obedience and social behavior in an unreflective way; on a more individualistic type of citizenship that stresses autonomy and disregards social concern; or on a more critical-democratic type of citizenship that considers both autonomy and social concern to be important (Veugelers, 2007). The idea of educating for democracy has been strongly advocated by Dewey. He spoke of a democratic way of life, democracy as lifestyle. Citizenship development is about knowledge, skills, and attitudes

(values), and the combination of these elements is often referred to as competences. Knowledge and skills are important. For a democratic society to flourish, it is necessary to know about democracy, to have the competences to act democratically. However, what is crucial are the values, the will to behave democratically. Willing to live a democratic way of life is seen as the pedagogical aim of contemporary citizenship education in democratic societies (Banks, 2004). A link between citizenship development and moral development is necessary to give citizenship a moral foundation, in order to make the linkage of the political, the social, and the interpersonal meaningful.

Values are interwoven in all aspects of teaching: in the curriculum, in the school culture, and as moral examples in teachers' behavior. Working with values is an essential part of teaching. Educating students to become teachers requires the teachers to learn how values are embedded in education; how they themselves, as reflective practitioners, can consciously create moral-based practices in education; and what different philosophical, pedagogical, and political theories and religious and cultural traditions say about moral development and the role of education.

The Moral in Teacher Education

Many scholars mention the important role teachers play in moral education, and there is a growing body of work that focuses on the education and professional development of teachers themselves. There are publications about the moral task of education that extends the research to the field of teacher education (see, e.g., Buzzelli and Johnston, 2002), and there is also research on teacher education that now includes a moral perspective (e.g., Russell and Loughran, 2007). In their book *The Moral Dimensions of Teaching*, Buzzelli and Johnston (2002, p. 132) present a moral perspective on teacher education. It includes redefining the teacher–student relationship, examining the personal beliefs and philosophies of students, re-contextualizing course methods and content knowledge, and ensuring authentic field experience and student activities. For them “morality constitutes that set of a person's beliefs and understanding which are evaluative in nature: that is, which distinguish, whether consciously or unconsciously, between what is right or wrong, good and bad.” (Buzzelli and Johnston, 2002, p. 3). Redefining the teacher–student perspective means sharing power and authority and becoming co-learners and inquirers with students. It is moral reflection because it entails the careful observation of, thinking about, and criticism of how we build relationships with our students. This requires moral courage and moral imagination.

Students should, according to Buzzelli and Johnston, be invited into an apprenticeship in these moral sensibilities

and to examine, test, and reformulate their beliefs and personal philosophies. When it comes to methods, students should learn how methods and content influence the growth and development of individuals in a moral sense, in what they learn about themselves and their relationship with the world, and in the differences between learners in this process of meaning making. When students enter classrooms during their field experience or their teaching practice, they become moral agents in the lives of the children in those classrooms. They need to develop an awareness of the moral significance and moral meanings of policies, practices, routines, and of the rituals of the classroom and the school.

More details about the moral dimension in teacher education can be found in the work of Johnson and Reiman (2007). In their research with students in teacher education they developed a matrix of indicators for the moral/ethical domain by using a neo-Kohlbergian framework. The post-conventional schema includes the following indicators:

- realizes the curriculum can be viewed from multiple perspectives;
- considers the benefits and consequences of instructional choices;
- takes into account a variety of learning styles when planning activities;
- holds a humanistic–democratic view of learner discipline;
- views rules as being designed to safeguard certain rights;
- considers rules as alterable and relative;
- is sensitive to students' rights;
- makes decisions based on the context of situations; and
- self-concept is organized around moral principles.

There is no strong empirical base for moral education in teacher education. Most of the work that has been conducted has focused on pedagogical goals and on formulating the consequences of research for educating prospective teachers in teacher education. We think it is necessary to develop a multiperspective view on moral values in teacher education. It can help make both the theory and the practice more suitable to support students in their transition toward becoming teachers.

Different Perspectives on Moral Values in Teacher Education

This section presents an overview of research and practices of different perspectives on working with moral values in teacher education. We speak of perspectives because the various approaches differ in philosophical background, pedagogical goals, and in suggested methodology. It is important to articulate different ways of thinking

about moral education, not to decide what is really moral but to stimulate dialogs about possible pedagogical goals and practices. The distinction made between the different perspectives is inspired by the work on teacher education by Liston and Zeichner, Aloni's philosophical foundations of humanistic education, and on our own work on different ways of teaching values.

The five perspectives are presented in a sequence that can be seen as different stages in the educating of teachers. Each stage adds a new element to its predecessor. The various perspectives are embedded in different ideas about the teaching and learning of moral values, the kind of citizenship society needs, and the very task of education. The perspectives articulate different educational practices that assume different types of methodology and goals of teachers, and therefore of teacher education. When presenting the five perspectives, we articulate what makes this approach unique.

We distinguish the following perspectives:

- value transfer;
- reflective practitioner;
- moral sensitivity;
- participation and dialog; and
- moral politics

In describing and analyzing these perspectives we refer to the way these perspectives work with values, skills, praxis, and social action.

Value Transfer

This approach focuses on the transfer of moral values in education. In this view, morality consists of virtues, of traits that support good behavior. As a teacher, being a good moral example and teaching students about good moral people are important methods in this approach. Student reflectivity is not really strengthened. The value transfer method is part of a pedagogical vision that has well-defined ideas about the good life and about important cultural traditions. The focus is more on getting youngsters involved in existing cultural practices rather than on challenging them to position themselves in an open, multicultural, and changing society. This approach can be situated in a national educational system that is concerned about its cultural heritage, or in a tradition that is based on a religious worldview that perceives its worldview to be more static than dynamic.

The emphasis in teacher education based on this perspective is on creating strong moral characters in student teachers, on learning about good people in the national and one's own cultural history, on telling stories in which the good moral life of concrete historical or contemporary examples are expressed, and on methods for correcting antisocial and antimoral behavior (Sockett, 1993). The main focus is behavior, not reflection or action.

Reflective Practitioner

According to many sociological analyses, modern society needs citizens who are flexible and reflective. Many professions require professionalism based on what Schön referred to as reflection-in-action. In education, this calls for a practice in which youngsters reflect on their behavior, take responsibility for their actions, and try out new behavior in an experimental and reflective manner. The reflective paradigm has had considerable influence on the thinking about teacher education. Prospective teachers in reflection-oriented teacher education practice are continuously challenged to question their beliefs and their teaching behavior. The emphasis in the reflective-practitioner approach to teacher education is on reflection on one's own professional behavior (Korthagen, 2004). The assumption is that one's personal beliefs influence one's pedagogical-didactic behavior.

Teachers should learn to consciously handle their methods and actions. Trying to make teachers and prospective teachers aware of their hidden beliefs, to further develop their pedagogical identity, and to expand their competences are part of this paradigm. Moral values can be part of the reflection process. Research shows that it is not easy to reach the moral level in reflection-oriented teacher education (Lunenberg *et al.*, 2006).

Moral Sensitivity

Moral values are abstract and are normative because they say something about the good life, about good and bad. Moral values are embedded in all narratives and practices, but how does one detect them? Following the reflection perspective, we can ask questions about why one chooses a certain alternative. We can do it critically and ask on which values the statement or behavior is based; we can ask what drives the action. In dialogs, we can challenge students to explain why they chose these ideas and practices. For moral reasoning, you must not only know cognitively that moral values are involved, but you must also have the sensitivity to feel and be aware of when moral values are at stake. The third perspective is therefore moral sensitivity (Campbell, 2003; Tirri, 1999). Knowing about moral values and moral dilemmas is not enough. Teachers should have the awareness to detect when moral values are involved.

Teachers should develop the sensitivity to see when moral values are at stake and how meaning is given to them. Not only is knowledge about teachers' behavior, as in the reflective-practitioner perspective, necessary, but what is even more important is an insight into the identity development of students; and this implies, seeing how students position themselves, give meaning to their experiences and the world around them, and how they work with the values involved. Teachers should incorporate this moral sensitivity in the art of their teaching.

Analyzing processes of meaning making by students and teachers in many situations and comparing different constructions for the same situation can stimulate the development of moral sensitivity. Students should learn to speak in moral language. Engaging theories and practices that focus on moral dilemmas and on moral concerns of participants can enhance the development of moral sensitivity.

Participation and Dialog

More culturally oriented sociological analyses of society argue that youngsters need to develop dialogic competences and an active action-oriented participation in society. Education should therefore pay more attention to dialogical learning, learning by experience, and activity-oriented learning. Dewey's pragmatism and Vygotsky's cultural activity theory have influenced the development of this perspective. From a Deweyan perspective, participation is not enough; experiences should lead to the transformation of knowledge and active processes of knowledge construction and competence development. A dialogic approach, based on the discourse theory of Habermas, has always been an important element in Kohlberg-oriented moral education (Oser, 1994). Students should be involved in communication actions that challenge their ideas and deconstruct the values interwoven in them – this can help them reconstruct their own personal beliefs.

Students in teacher education should learn to use these dialogic and action-oriented activities in a learning environment that supports this kind of professional development. Gaining practical experience is a crucial element in many teacher education programs (Ten Dam and Blom, 2006). Action research and networking should be added to this learning. A perspective aimed at participation and dialog, as mentioned above, can create dialogic learning and change. Teaching, for Dewey, is a moral endeavor because its constituent acts have moral meanings in their own right and genuine intellectual work always implies moral development (Hansen, 2003). Hansen shows how moral assumptions are manifest in rituals, such as classroom beginnings, in teachers' style, and in the curriculum. Hansen argues for more attention toward moral knowledge and moral judgment in education and in teacher education. Participation, action, and reflection should include concern about moral values.

Moral Politics

We refer to the fifth perspective as moral politics. Morality, in this view, is seen as embedded in a political context, and the aim of education is formulated as political action for social change. Teacher education and social action are linked in this perspective. Beyer (1996) speaks of democratic education, Oakes and Rogers (2006) use the concepts of teaching for changing the world, and teaching for social justice. Empowerment of people and working on structural

changes that enhance humanity and change at grassroots level are considered to be pedagogical goals. The work of Dewey and Freire, and critical theory and critical pedagogy have influenced the development of this perspective.

Teacher education in this perspective is linked with social activist groups in and outside schools, and students learn by trying to bring about change in schools and in the community. In their teaching practice, prospective teachers attempt to create democratic classrooms, to include minorities' perspectives in the curriculum, and organize participatory social inquiry in and outside the school. Teaching is seen as social and moral engagement and as creating a better and more just world. Reflection and action are linked and aimed at realizing these moral goals. Teacher educators are clear in their critical emancipatory view, and their particular choice for schools that serve disadvantaged children. They hope prospective teachers opt to enter this kind of education.

We have placed the five perspectives in **Table 1** that shows the differences between the different perspectives.

We can argue that all the five perspectives have their particular strengths depending on the goals they have in education and the corresponding teacher education method. If one wants to focus in education on character building and on making clear the values one finds important, then the value-transfer perspective is the most useful one. When education is seen as very flexible in its content, structure, and teaching methods, and is one in which teachers are autonomous and considered to be professionals, then a more reflective perspective is necessary. If this kind of education considers the development of morality as crucial and important, then the moral-sensitivity perspective is desired. If learning is seen as transforming practices and learning by doing and dialog, then the participation and dialog perspective is needed. And if this transforming practice is seen as part of a struggle for social justice and democracy, then the moral-politics perspective is useful. We are aware of the fact that concrete practices often show a unique combination of elements of these perspectives.

It is easy to say that teacher education should pay more attention to moral values. As we have shown in this article, moral values can be incorporated in numerous different ways in education in general, and in teacher education

in particular. Teacher educators can choose how they include moral values in their teaching education program.

Bibliography

- Banks, J. A. (ed.) (2004). *Diversity and Citizenship Education*. San Francisco: Jossey-Bass.
- Beyer, L. E. (1996). *Creating Democratic Classrooms*. New York: Teachers College Press.
- Buzzelli, C. A. and Johnston, B. (2002). *The Moral Dimensions of Teaching*. New York: Routledge Falmer.
- Campbell, E. (2003). *The Ethical Teacher*. Maidenhead: Open University Press.
- Hansen, D. T. (2003). Teaching as a moral action. In Richardson, V. (ed.) *Handbook of Research on Teaching*, 4th edn., pp 826–857. Washington: AERA.
- Johnson, L. E. and Reiman, A. J. (2007). Beginning teacher disposition: Examining the moral/ethical domain. *Teaching and Teacher Education* **23**, 676–687.
- Korthagen, F. A. J. (2004). *Linking Practice and Theory. The Pedagogy of Realistic Teacher Education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Lunenburg, M., Korthagen, K., and Willemse, M. (2007). Value-based teacher education: The role of knowledge, language and golden moments. In Russell, T. and Loughran, J. (eds.) *Enacting a Pedagogy of Teacher Education. Values, Relationships and Practices*, pp 166–181. London: Routledge.
- Oakes, J. and Rogers, J. (2006). *Learning Power. Organizing for Education and Justice*. New York: Teachers College Press.
- Oser, F. K. (1994). Moral perspectives on teaching. *Review of Research in Education* **20**, 57–127.
- Russell, T. and Loughran, J. (eds.) (2007). *Enacting a Pedagogy of Teacher Education. Values, Relationships and Practices*. London: Routledge.
- Sockett, H. (1993). *The Moral Base for Teacher Professionalism*. New York: Teachers College Press.
- Solomon, D., Watson, M. S., and Battistich, V. A. (2001). Teaching and schooling effects on moral/prosocial development. In Richardson, V. (ed.) *Handbook of Research on Teaching*, 4th edn., pp 566–603. Washington: AERA.
- Ten Dam, G. and Blom, S. (2006). Learning through participation. The potential of school-based teacher education for developing a professional identity. *Teaching and Teacher Education* **22**, 647–660.
- Tirri, K. (1999). Teachers' perceptions of moral dilemmas at school. *Journal of Moral Education* **28**, 31–47.
- Veugelers, W. (2007). Creating critical-democratic citizenship education: Empowering humanity and democracy in Dutch education. *Compare* **37**, 105–119.
- Veugelers, W. and Vedder, P. (2003). Values and teaching. *Teachers and Teaching* **9**(4), 377–390.

Further Reading

- Aloni, N. (2003). *Enhancing Humanity. The Philosophical Foundations of Humanistic Education*. Dordrecht: Kluwer Academic Publishers.
- Cummings, R., Harlow, S., and Maddux, C. D. (2007). Moral reasoning of in-service and pre-service teachers: A review of the research. *Journal of Moral Education* **36**, 67–78.
- Dewey, J. (1923). *Democracy and Education*. New York: Macmillan.
- Haste, H. (2004). Constructing the citizen. *Political Psychology* **25**, 413–440.
- Lickona, T. (1991). *Educating for Character: How Our Schools Can Teach Respect and Responsibility*. New York: Bantam.
- Liston, D. P. and Zeichner, K. M. (1991). *Teacher Education and the Social Conditions of Schooling*. New York: Routledge.

Table 1 Different perspectives on moral values in teacher education

	Values	Skills	Praxis	Social action
Value transfer	x			
Reflective practitioner		x		
Moral sensitivity	x	x		
Participation and dialog	x	x	x	
Moral politics	x	x	x	x

- Oser, W. and Veugelers, W. (eds.) (2008). *Getting Involved: Citizenship Education and Sources of Moral Values*. Rotterdam: Sense publishers.
- Power, F. C., Higgins, A., and Kohlberg, L. (1989). *Lawrence Kohlberg's Approach to Moral Education*. New York: Columbia University Press.
- Schön, D. (1983). *The Reflective Practitioner*. New York: Basic Books.
- Westheimer, J. and Kahne, J. (2004). What kind of citizen? The politics of education of democracy. *American Educational Research Journal* **41**, 237–269.

Pedagogical Content Knowledge

J H Van Driel, ICLON–Leiden University Graduate School of Teaching, Leiden, The Netherlands
A Berry, Monash University, Clayton, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The term pedagogical content knowledge (PCK) was introduced by Lee Shulman in his presidential address to the American Educational Research Association (Shulman, 1986). Shulman argued that, for a long time, research on teaching and teacher education had undeservedly ignored questions dealing with the content of the lessons taught. Shulman presented a strong case for PCK as a specific form of knowledge for teaching which refers to the transformation of subject-matter knowledge in the context of facilitating student understanding. Teachers need this type of knowledge to structure the content of their lessons, to choose or develop specific representations or analogies, to understand and anticipate particular preconceptions or learning difficulties of their students, and so on. Shulman asserted that teachers had a unique way of looking at practice and his intrigue with the manner in which they did so encouraged an examination of teachers' pedagogical thinking in ways that, it was anticipated, would reveal what teachers must know to best teach their content to their students.

Shulman's introduction of PCK, that is, his call for attention to the knowledge that teachers use and need to teach specific content, connects with a well-established tradition in the European literature on teaching and teacher education. The German word *Fachdidaktik* is often used to refer to this tradition, which is more philosophical than empirical. *Fachdidaktik* may be translated as the pedagogy of subject matter.

In preparation programs for secondary teachers within this European tradition, a relatively large amount of time is usually reserved for coursework on *Fachdidaktik*. In this context, preservice teachers are required to address questions concerning the significance of the content for their students' future, the structure and the exemplary value of the content, and the ways in which the content may be represented to make it interesting and comprehensible for students. However, the *Fachdidaktik* tradition has not provided detailed insights into the ways in which teachers transform subject-matter knowledge, and how they relate their transformations to student understanding during classroom communication in order to overcome student misconceptions, thus contributing to meaningful learning. Such issues lie at the heart of the concept of PCK, which explains why empirical research on PCK adds a new perspective to the *Fachdidaktik* tradition.

In the past 20 years, a plethora of publications studying and integrating PCK into teacher education have been published in the research literature. Publications have addressed various subject areas, including English, mathematics, science, physical education, and social studies (Grossman *et al.*, 2005). Although researchers have thus begun to push to find ways of examining what teachers know, still few concrete examples of PCK in subject areas have emerged. Hence, the construct itself has not yet positively impacted the valuing of teachers' professional knowledge and practice.

Van Driel *et al.* (1998) highlighted this concern when they noted that although the research community embraced the notion of PCK, few topic-specific examples existed in the literature to illuminate this important aspect of teachers' professional knowledge. Suggestions to account for this paucity of concrete examples include the tacit nature of teachers' knowledge and the absence of a shared language or structure to adequately discuss that knowledge. Related to this is the problem that teachers are often unaware of the knowledge they possess, which is often contextualized and associated with particular students, events, and classrooms.

In this article, we begin with a brief review of the ways in which PCK has been conceptualized in the research literature, followed by an overview of research that has focused on the way PCK develops, and the factors influencing this development. Concluding this article, we discuss research on PCK in relation to the practice of teaching and teacher education.

Conceptualizing Pedagogical Content Knowledge

One of the preeminent issues in the research literature is the ongoing discussion on the nature of PCK: What is it, and how can it be represented? The key elements in Shulman's original conception of PCK include knowledge of representations of subject matter, on the one hand, and understanding of specific learning difficulties and student conceptions, on the other. Shulman suggested that the more representations and strategies teachers have at their disposal within a certain subject domain, and the better they understand their students' learning processes in the same domain, the more effectively can they teach in that domain.

In a later article, Shulman (1987) included PCK in what he called “the knowledge base for teaching.” This knowledge base consisted of seven categories, three of which were content related (i.e., content knowledge, PCK, and curriculum knowledge). The other four categories referred to general pedagogy, learners and their characteristics, educational contexts, and educational purposes. In the subsequent years, many scholars have used Shulman’s knowledge base for teaching as a starting point, and have conceptualized PCK in other ways. Although almost all scholars adopted the two key elements of PCK mentioned above (i.e., knowledge of comprehensible representations of subject matter and understanding of content-related learning difficulties), they have extended the concept by including other categories of knowledge in PCK from Shulman’s knowledge base for teaching, such as knowledge of curriculum, or assessment.

In an often-cited paper, Magnusson *et al.* (1999) presented PCK as a separate domain of teacher knowledge which exists alongside other domains, such as pedagogical knowledge and beliefs. In their discussion of the nature of PCK, they presented a model in which PCK for science teaching consists of five aspects or components: (1) orientations toward teaching science, (2) knowledge of science curricula, (3) knowledge of students’ understanding of science, (4) knowledge of assessment in science, and (5) knowledge of subject-specific and topic-specific strategies. Acknowledging that these components may interact in very complex ways, these authors claim that effective teachers need to develop expertise in all aspects of PCK, and with respect to all topics they teach. Orientations toward teaching science, in particular, have been identified as a critical component within this PCK model (Friedrichsen and Dana, 2005). Sources that shape teachers’ orientations toward teaching science include: prior work experiences, professional development choices, beliefs about students and about learning, as well as time constraints.

Several scholars have addressed problems associated with investigating PCK in terms of specific components or categories. This may lead to a segmented and simplistic representation of teachers’ knowledge. To acknowledge its constantly evolving nature, Cochran *et al.* (1993) preferred to speak of pedagogical content knowing (PCKg). Similarly, Mason (1999) emphasized the dynamic nature of PCK in research and teaching, respectively. Rather than representing PCK as a fixed or static body of knowledge, Mason perceived PCK as an ability to combine content knowledge of a discipline with the teaching of that discipline. Adding to this, Hashweh (2005) proposed to consider PCK as a collection or repertoire of pedagogical constructions, which teachers acquire when repeatedly teaching a certain topic.

Loughran and colleagues argued that researchers often fail to acknowledge the complex relationships and

interactions existing within a teacher’s personal professional knowledge base (Loughran *et al.*, 2001, 2004). Attempting to portray PCK in a way that is valid and useful from a teacher’s point of view, these authors constructed a series of resource folios for various (science) topics consisting of a content representation (CoRe) in combination with the so-called pedagogical and professional-experience repertoires (PaP-eRs). For example, a resource folio for the topic of force includes a CoRe and PaP-eRs about teaching and learning this topic. A CoRe is structured around questions related to some of the elements of Shulman’s knowledge base, in particular, knowledge of the main content ideas associated with a specific topic, teaching procedures and purposes, and knowledge about students’ thinking.

Each CoRe is connected to a collection of PaP-eRs, which illustrate aspects of PCK in action of the topic under consideration. PaP-eRs are short narratives based on teachers’ accounts of teaching a specific topic and are intended to make explicit teachers’ pedagogical reasoning, that is, the thinking and actions of a capable teacher in teaching a specific aspect of the content. PaP-eRs include a variety of narrative representations, for example: a dialog between two teachers exploring their approach to the teaching of particular content and student responses to it; a teacher’s annotated curriculum document, or a student’s perspective of a teaching/learning situation. The function of PaP-eRs is to elaborate and give insight into the various interacting elements that comprise a teacher’s PCK.

To conclude this section, it is clear that PCK has attracted considerable attention in the research literature, but there is no universally accepted conceptualization of PCK. The general fuzziness around the concept of PCK has meant that that which is searched for, and that which is uncovered is variable indeed. In fact, the literature on PCK reveals several dilemmas:

- Since various scholars have included different elements within their conceptualization, it is unclear whether PCK should be seen as knowledge which is very specific for each topic that teachers teach, or one should think in terms of general or subject-specific PCK? The former implies that teachers need specific PCK for each particular topic they teach, whereas the latter suggests that there are ideas, for example, about teaching strategies, which are useful across topics within a subject area (e.g., algebra, or even mathematics).
- Given the discussion about the complex, dynamic, and holistic nature of PCK, there are obvious challenges for researchers investigating PCK in analytical, empirical studies, to avoid breaking down PCK into segmented and static entities, resulting in a simplistic and hence misleading picture of teachers’ knowledge.
- Related to the previous issue is the problem of how research into PCK can become more relevant for the

practice of teaching and teacher education, so that PCK becomes an important concept, not only for researchers, but also for teachers and educators.

The Development of PCK

Marks (1990) perceived the development of PCK as an integrative process revolving round the interpretation of subject-matter knowledge and the specification of general pedagogical knowledge. Since this process is rooted in classroom practice, the implication is that preservice or beginning teachers usually have little or no PCK at their disposal. Magnusson *et al.* (1999) argued that development of PCK is a complex and nonlinear process. These authors stated that a teacher education program can never completely address all the components of PCK that a teacher needs. Research reported by Van Driel *et al.* (1998) concluded that while subject-matter knowledge is prerequisite, the major source of a teacher's PCK is teaching experience. Hence, the development of PCK should also be an important element of the continuing education of teachers.

How PCK develops is discussed in the following section. This discussion is organized according to the following themes relevant to PCK development: (1) the role of subject-matter knowledge, (2) teaching experience, (3) a focus on student learning, and (4) the design of teacher education.

The Role of Subject-Matter Knowledge

Several authors have pointed at the problematic nature of preservice teachers' content knowledge, both substantive and syntactical, when they enter teacher education (e.g., Halim and Meerah, 2002). When their subject-matter knowledge is very limited and contains many misconceptions, preservice secondary teachers hardly develop PCK during their teacher education program. Working with beginning elementary science teachers, Appleton (2003) tried to compensate for these teachers' lack of content knowledge and PCK by providing them with what he called activities that work that included, for instance, hands-on activities intended to raise interest among young children around particular science content. Using these activities apparently helped these teachers to start teaching science with some success, and consequently stimulated the beginning of the development of their PCK.

From studies on experienced teachers, however, it appears that a strong and well-integrated subject-matter knowledge does not guarantee the smooth development of an individual's PCK. In particular, when teaching unfamiliar topics, it appears that experienced teachers drew from their general pedagogical knowledge, which can

constitute a supporting framework for the development of their PCK (Hashweh, 1987).

Teaching Experience

Preservice teachers' lack of teaching experience explains why they usually express little to no PCK. Added to this, the concerns of many preservice teachers related to self-confidence and an ability to teach may also, at least initially, hinder the development of their PCK. This means, until a teacher has gained sufficient confidence and experience and mastered basic classroom skills, that the development of PCK as a readily accessible and useful translation of subject-matter knowledge into classroom practice may be delayed. However, while teaching experience may promote the development of PCK, the provision of structured opportunities for reflection on the relationship between subject-matter knowledge and classroom practice is also important for facilitating the development of preservice teachers' PCK. Without such opportunities PCK development is at best, haphazard and at worst, barely apparent.

The role of teaching experience was highlighted in a longitudinal study of 10 years of one elementary science teacher (Mulholland and Wallace, 2005). These authors found that all the various knowledge bases that comprise PCK grew over time, although in different ways. In particular, general teaching knowledge bases and interactive knowledge bases, such as knowledge about learners, developed substantially, whereas subject knowledge bases remained relatively stable. Moreover, this study illustrated the development of PCK as a complex pattern of interactions between these various components, which is situated in the teacher's classroom. In this particular study, new curriculum materials appeared to be an important source for this development.

A Focus on Student Learning

Central to the notion of PCK is the idea that instruction can only be effective if it is attuned to the ways in which students learn specific content. Preservice teachers usually have limited ability to relate students' prior knowledge to instruction and to adequately anticipate student conceptions. Consequently, it has appeared to be important to provide preservice elementary teachers with opportunities to observe student learning in practice and, in connection to this, to study the research literature (e.g., on misconceptions regarding specific topics). Applying these ideas, Zembal-Saul *et al.* (2002) found that preservice teachers were able to focus on student learning, while maintaining a strong subject-matter emphasis, thus contributing to the development of their PCK. Similarly, De Jong and Van Driel used a workshop format in which preservice teachers were asked to prepare and teach a

series of lessons on an important theme from the chemistry curriculum. As a result, these preservice teachers demonstrated a growing awareness of specific instructional strategies in relation to a better understanding of student thinking about the topic of their lesson series. Reflecting on and discussing teaching experiences appeared to be crucial for this development (De Jong *et al.*, 2005; Van Driel *et al.*, 2002).

The Design of Teacher Education

Various studies have been conducted on ways of structuring and organizing preservice teacher education programs to promote the development of PCK (De Jong *et al.*, 2005; Loughran *et al.*, 2006; Zembal-Saul *et al.*, 1999). However, the reported impact of such programs on the development of preservice teachers' PCK is varied. It would appear that the preservice teachers experience difficulty merging subject-matter courses and education courses that are not integrated by design. In programs that were not deliberately structured to help promote integration of different types of knowledge, preservice teachers appeared to retain separate views of subject matter and pedagogy as opposed to the integrated knowledge base advocated by PCK. In the context of such nonintegrated programs, it has been demonstrated that knowledge development appears most strongly influenced by individual and contextual factors, resulting, among other outcomes, in the adoption by preservice teachers of conventional instructional strategies, stressing procedures instead of student understanding.

Some significant gains in the growth of PCK through short-term, intensive, skill-oriented workshops or via specific activities have been reported. Justi and Van Driel (2005) incorporated the design of a lesson series in an action research project that was a requirement for preservice secondary teachers as part of their teacher education program. It was found that, in particular, reflective activities (such as writing reports and sharing experiences in collective meetings) stimulated the development of subject-matter knowledge and PCK of these preservice teachers. Sperandio-Mineo *et al.* (2005), working with a group of 28 preservice teachers of physics, found similar results, concluding that the development of PCK is a complex process which is connected with deepening of subject-matter knowledge and improved awareness of pedagogical issues. Loughran *et al.* (2006) employed a modified version of their aforementioned CoRe and PaPeRs framework in a preservice science teacher education program, whereby preservice teachers developed their own CoRes and PaPeRs based on their learning about PCK at university and from their experiences of teaching in schools. In this case, the structured framework together with ongoing opportunities for reflection about the process of their knowledge development supported deeper

subject-matter understandings of these preservice teachers and a growing awareness of the value of PCK for teaching.

Conclusion

We conclude that research on the development of PCK presents an ambiguous picture. Obviously, teaching experience and subject-matter knowledge are important, but contextual and personal factors apparently may lead to quite different processes of knowledge development. The development of PCK is perhaps then best viewed as a complex interplay between knowledge of subject matter, teaching and learning, and context, and the way in which teachers combine and use this knowledge to express their expertise. In the context of preservice teacher education, PCK can be promoted by addressing both preservice teachers' subject-matter knowledge and their educational beliefs, in combination with providing them with opportunities to gain teaching experience, and in particular, to reflect on these experiences. Specific workshops may serve as an intermediate measure, helping to initialize and frame subsequent teaching experiences.

At the level of specific topics, however, there is little research to inform us about the ways in which teachers transform subject-matter knowledge and how they relate their transformations to student understanding. Neither has there been specific input for teacher education in this respect. Instead, most recommendations regarding PCK development for preservice and in-service teacher education are of a rather general nature.

In our view, PCK implies a transformation of subject-matter knowledge so that it can be used effectively and flexibly in the communication process between teachers and learners during classroom practice. From this perspective, the main aim in studying teachers' PCK is to understand how and why teachers teach a certain topic the way they do and, in particular, how their teaching approach is related to, or focused on, student learning. Achieving such an understanding is vital to the development of effective preservice and in-service programs.

However, much of the research on PCK so far has not served this purpose very well. This can be explained partly by the tacit nature of teachers' personal professional knowledge, which has urged scholars to develop instruments and procedures that help to make this knowledge explicit and measure PCK in a valid manner, rather than exploring what the construct might offer to the practice of teaching and teacher education. In addition, it must be noted that PCK is quite sensitive to personal characteristics of teachers and their working contexts. Several studies (e.g., Henze, 2006) have reported substantial differences between the PCK of experienced teachers around the same topic area, even when their

subject-matter knowledge is similar and when they teach the same curriculum. These differences appear to stem from a range of factors including different orientations toward teaching, different purposes and practices.

Nevertheless, it seems possible to capture and portray PCK in such a way that key notions of teaching and learning a specific topic are made explicit (Van Driel *et al.*, 1998). In addition, discussing and sharing such key notions among teachers may contribute to the establishment of a collective PCK, that is, a shared or common form of teachers' professional practical knowledge about teaching certain subject matter. At the same time, there should, of course, be room for individual teachers to adapt or complement this shared knowledge to their own situations (cf. the above-mentioned resource folios of Loughran and colleagues).

By introducing the construct of PCK, examining, analyzing, and modeling it in preservice teacher education practice (both at university and in school-based practicum), it is more likely that teaching will be seen as a specialized and sophisticated practice. In this way, PCK may emerge as a way of thinking about teaching subject matter that encourages learning about teaching that goes beyond the acquisition of instructional strategies and techniques to include an understanding of how learners develop an insight in specific school subjects.

See also: A Pedagogy of Teacher Education; Contemporary Approaches to Teacher Professional Development; Experienced Teachers' Craft Knowledge; Science Teacher Education; Taking Prospective Teachers' Beliefs into Account in Teacher Education; Teacher Education and the Educational Foundations Knowledge Base.

Bibliography

- Appleton, K. (2003). How do beginning primary school teachers cope with science? Toward an understanding of science teaching practice. *Research in Science Education* **33**, 1–25.
- Cochran, K. F., DeRuiter, J. A., and King, R. A. (1993). Pedagogical content knowing: An integrative model for teacher preparation. *Journal of Teacher Education* **44**, 263–272.
- De Jong, O., Van Driel, J. H., and Verloop, N. (2005). Pre-service teachers' pedagogical content knowledge of using particle models in teaching chemistry. *Journal of Research in Science Teaching* **42**, 947–964.
- Friedrichsen, P. F. and Dana, T. M. (2005). Substantive-level theory of highly regarded secondary biology teachers' science teaching orientations. *Journal of Research in Science Teaching* **42**, 218–244.
- Grossman, P., Schoenfeld, A., and Lee, C. (2005). Teaching subject matter. In Darling-Hammond, L. and Bransford, J. (eds.) *Preparing Teachers for a Changing World. What Teachers Learn and Should Be Able to Do*, pp 201–231. San Francisco, CA: Jossey-Bass.
- Halim, L. and Meerah, S. M. (2002). Science trainee teachers' pedagogical content knowledge and its influence on physics teaching. *Research in Science & Technological Education* **20**, 215–225.
- Hashweh, M. Z. (1987). Effects of subject matter knowledge in the teaching of biology and physics. *Teaching and Teacher Education* **3**, 109–120.
- Hashweh, M. Z. (2005). Teacher pedagogical constructions: A reconfiguration of pedagogical content knowledge. *Teachers and Teaching: Theory and Practice* **11**, 273–292.
- Henze, I. (2006). *Science Teachers' Knowledge Development in the Context of Educational Innovation*. PhD Dissertation, ICLON, Universiteit Leiden, Leiden.
- Justi, R. and Van Driel, J. H. (2005). The development of science teachers' knowledge on models and modelling: Promoting, characterizing, and understanding the process. *International Journal of Science Education* **27**, 549–573.
- Loughran, J., Milroy, P., Berry, A., Gunstone, R., and Mulhall, P. (2001). Documenting science teachers' pedagogical content knowledge through PaP-eRs. *Research in Science Education* **31**, 289–307.
- Loughran, J., Mulhall, P., and Berry, A. (2004). In search of pedagogical content knowledge for science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching* **41**, 370–391.
- Loughran, J. J., Berry, A., Mulhall, P., and Woolnough, J. (2006). Understanding and valuing the development of pedagogical knowledge in science teacher education. In Eilks, I. and Ralle, B. (eds.) *Towards Research-Based Science Teacher Education*, pp 65–76. Aachen: Shaker.
- Magnusson, S., Borko, H., and Krajcik, J. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In Gess-Newsome, J. and Lederman, N. G. (eds.) *Examining Pedagogical Content Knowledge*, pp 95–132. Dordrecht: Kluwer Academic Publishers.
- Marks, R. (1990). Pedagogical content knowledge: From a mathematical case to a modified conception. *Journal of Teacher Education* **41**, 3–11.
- Mason, C. L. (1999). The TRIAD approach: A consensus for science teaching and learning. In Gess-Newsome, J. and Lederman, N. G. (eds.) *Examining Pedagogical Content Knowledge*, pp 277–292. Dordrecht: Kluwer Academic Publishers.
- Mulholland, J. and Wallace, J. (2005). Growing the tree of teacher knowledge: Ten years of learning to teach elementary science. *Journal of Research in Science Teaching* **42**, 767–790.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher* **15**(2), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review* **57**, 1–22.
- Sperandio-Mineo, R. M., Fazio, C., and Tarantino, G. (2005). Pedagogical content knowledge development and pre-service physics teacher education: A case study. *Research in Science Education* **36**, 235–268.
- Van Driel, J. H., Verloop, N., and De Vos, W. (1998). Developing science teachers' pedagogical content knowledge. *Journal of Research in Science Teaching* **35**, 673–695.
- Van Driel, J. H., De Jong, O., and Verloop, N. (2002). The development of pre-service chemistry teachers' PCK. *Science Education* **86**, 572–590.
- Zemba-Saul, C., Starr, M., and Krajcik, J. (1999). Constructing a framework for elementary science teaching using pedagogical content knowledge. In Gess-Newsome, J. and Lederman, N. G. (eds.) *Examining Pedagogical Content Knowledge*, pp 237–256. Dordrecht: Kluwer Academic Publishers.
- Zemba-Saul, C., Krajcik, J., and Blumenfeld, P. (2002). Elementary student teachers' science content representations. *Journal of Research in Science Teaching* **39**, 443–463.

Further Reading

- Abell, S. (2007). Research on science teacher knowledge. In Abell, S. and Lederman, N. (eds.) *Handbook of Research on Science*

- Education*, pp 1105–1149. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bullough, R. V. J. (2001). Pedagogical content knowledge circa 1907 and 1987: A study in the history of an idea. *Journal of Teaching and Teacher Education* **17**, 655–666.
- Geddis, A. N. (1993). Transforming subject-matter knowledge: The role of pedagogical content knowledge in learning to reflect on teaching. *International Journal of Science Education* **15**, 673–683.
- Gess-Newsome, J. and Lederman, N. G. (1999). *Examining Pedagogical Content Knowledge*. Dordrecht: Kluwer Academic Publishers.
- Gudmundsdottir, S. (1991). The narrative nature of pedagogical content knowledge. *Paper presented at the Annual Meeting of the American Educational Research Association*. Chicago, IL (ERIC Document Reproduction Service No. ED 341 991).
- Gudmundsdottir, S. and Shulman, L. (1987). Pedagogical content knowledge in social studies. *Scandinavian Journal of Educational Research* **31**, 59–70.
- Lederman, N. G. and Gess-Newsome, J. (1992). Do subject matter knowledge, pedagogical knowledge, and pedagogical content knowledge constitute the Ideal Gas Law of science teaching? *Journal of Science Teacher Education* **3**, 16–20.
- Loughran, J. J., Berry, A., and Mulhall, P. (2006). *Understanding and Developing Science Teachers' Pedagogical Content Knowledge*. Rotterdam: Sense Publishers.
- Rovegno, I. (1995). Theoretical perspectives on knowledge and learning and a student teachers' pedagogical content knowledge of dividing and sequencing subject matter. *Journal of Teaching in Physical Education* **14**, 284–304.
- Veal, W. R. and MaKinster, J. G. (1999). Pedagogical content knowledge taxonomies. *Electronic Journal of Science Education* **3**(4). <http://unr.edu/homepage/crowther/ejse/vealmak.html> (accessed December 2008).
- Westbury, I., Hopmann, S., and Riquarts, K. (2000). *Teaching as a Reflective Practice: The German Didaktik Tradition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Teacher Education and the Educational Foundations Knowledge Base

F B Murray, University of Delaware, Newark, DE, USA

© 2010 Elsevier Ltd. All rights reserved.

Teacher educators are doubly interested in teaching as it is both their subject matter and their method. Like all teachers, it is important for them to know how much of the subject matter their students have already mastered so that they can build on their students' prior knowledge of the field. Long before there were teacher educators and teacher education programs, there was effective teaching because we are after all a teaching species, a species whose young cannot, and do not, survive unless they are taught, invariably and historically, by persons with no formal schooling in teaching, let alone in the disciplines of the foundations of education (psychology, philosophy, history, and sociology).

Some have argued (e.g., Gilbert and Borish, 1997) that teaching is so basic to life that it can be found at the cellular level, but most researchers find that teaching requires (1) an intention to teach, (2) a belief that the pupil does not know something, and (3) the teacher's implicit understanding of the pupil's mind (Strauss and Ziv, 2004; Premack and Premack, 1996). Even if teaching was as simple as requiring the presence of a teacher when an activity was performed repeatedly, there are few documented cases of new learning in nonhumans in these instances (Visalberghi and Frigaszy, 1996, p. 286).

Children as young as 3 years have been shown to teach (Ashley and Tomasello, 1998) and Strauss *et al.* (2002) found that children's style of teaching a new board game (or building something) changed from demonstration and modeling at 3–4 years to predominately verbal explaining at 5 and 6 years (i.e., from show to tell). Seven-year-olds can adapt their teaching on occasion to their perception of what their pupils already know and can do. They sometimes ask their pupils if they have understood and tailor their teaching to the learners' mistakes. The way children teach is also influenced by schooling itself. Maynard (2004) found that older Mayan children (6–11 years) who had been to school were also able to adopt school-like teaching with their younger siblings (didactic teaching at a distance) in place of indigenous teaching practices used in families for cooking and weaving (close-up interactive demonstrations).

Teacher education students come to attend teacher education programs, as Olson and Bruner (1996) also argued in their account of folk pedagogy, with two natural strategies already in place – showing and telling. Before they begin their teacher education programs, teacher

education students have also acquired a naïve, but serviceable, theory of the human mind (see Heider, 1958; Baldwin, 1980 on naïve psychology). In this naïve or common sense psychology, they see the pupil's school achievement as tied to only four factors – the pupil's ability, effort, task difficulty, and luck. With these four factors, they think they can explain completely the pupil's success or failure by attributing the level of the pupil's work to his/her ability, effort, to the difficulty of the school task, or to plain luck. The problem with this naïve account is that it is unduly limiting insofar as ability is concerned, to take only one example, which is not fixed or stable but varies from moment to moment interactively with many other mental factors, not just the few in the naïve theory (Baldwin, 1980; Murray, 1991).

Regrettably, even after the completion of a teacher education program, many experienced teachers' reasoning about teaching is still fairly limited and constrained by views of teaching that the prospective teacher had before beginning a teacher education program. Strauss and Shilony (1994) interviewed both experienced and novice science and humanities' teachers about how they would teach a topic of their choice to children of various ages (7–17 years). Both novice and experienced teachers in both disciplines conceptualized teaching only as the flow of information from their heads to their pupils' heads, acknowledging that their own role was only to devise manageable and interesting ways of entry into the students' mind so that the information could be stored and anchored appropriately. The student is passive, a receptacle waiting to be filled, and if the information fails to flow to its destination, the receptacle was taken to be too small and/or the student was inattentive. The novices and experienced teachers were indistinguishable in their responses and therefore, Strauss and Shilony attributed their responses to their preteacher education training and pre-existing common sense naïve views of teaching. There is also some clear evidence that experienced teachers perform differently from novices (Berliner, 1988). They are able, for example, to read highly contextualized cues in the classroom that allow them to spot trouble and also contribute to the smooth functioning of the class. These skills, however, were seemingly learned on the job and not in their teacher education program.

Astington and Pelletier (1996) cataloged the following tenets of prospective teachers' beliefs about teaching:

(1) children are born with abilities and capacities that unfold linearly in time, (2) instructional sequences should match developmental sequences, (3) learning occurs sequentially within a hierarchy of skills, and (4) student errors are attributable to incomplete learning or inattention. When the pupil is to learn to do something, the teacher need only demonstrate or model it, and when the pupil is to know something, the teacher need only tell the pupils what they need to know.

Teachers' reasoning about their pupil's accomplishments, in fact, proves to be surprisingly indistinguishable from the layperson's explanations. Kuhn (1991) found, for example, when she asked teachers, "What causes children to fail in school?" the teachers' reasoning was that the pupil "shows no superiority over reasoning regarding the other topics. . . performance for the school topic is in fact somewhat inferior to their performance for the other topics. . ." (p. 249). Their reasoning, like about half the other subjects in her study, tends to be absolutist, that is, they believe it is possible with sufficient study to know with certainty the causes of a phenomenon, often through a personal commitment to a theory or assertion.

On the whole, the prospective teacher's prior knowledge and beliefs make it difficult to have teachers learn some of the newer teaching techniques such as dialectic, discovery, invention, cooperative learning, and collaboration, because the prior techniques are often the opposite of the newer strategies. They may, for example, require the prospective teacher to refrain entirely from showing or telling. They are also at variance with the Astington and Pelletier (1996) prior tenets because developmental pathways, for example, are rarely linear, but often show fits and starts, oscillations, and even reversals, particularly when performance is at an optimal level or when a new skill is being developed (Fischer and Bidell, 1998).

The teacher education program, particularly the foundations disciplines, are meant to shore up the mental life of the teacher, especially their states of mind when they explain to themselves their student's behavior, when they insist on certain classroom practices, and establish schoolwork requirements. It gives them a basis for deciding, for example, whether to prohibit handheld calculators in arithmetic lessons, to group students by ability, to require memorization of certain facts and not others, to socially promote some students to the next grade, to allow some students to skip a grade level, or to adopt new teaching methods. In other words, the knowledge base of the educational foundation disciplines refers to that body of opinion, information, and knowledge that the professional teacher relies on to justify her professional decisions.

This knowledge is important if only because teachers attempt only what they believe is possible. The judgment that an aspect of education is impossible, or impossibly difficult, usually flows from the prior misconceptions held by prospective teachers because they tend to be firm and

absolute tenets. The core knowledge of educational foundations typically promotes more flexible forms of pupil accomplishment that go beyond what can be simply given to the pupil by showing and telling.

Having said that, it is important to note the limitations of what can be had from the educational foundations disciplines. On logical grounds alone, sound practices can be derived from both true and false theoretical propositions. The soundness of an educational practice is simply not guaranteed by whether it fits with what is taught in teacher education courses, even when the builders of the theory, who have created the educational technique, think it stems from their theories. Although a proven effective teaching technique, programmed instruction does not tend to lose that effectiveness when the various theoretical requirements Skinner (its inventor) placed on it are violated (e.g., programs that scramble frames randomly, delay or omit reinforcement, and require only covert responding are as effective as their theoretically sanctioned opposites). While the Montessori practices are compatible with Piaget's theory, they were developed independent of it and succeeded quite distinctly from the truth or falsity of Piaget's propositions. Both Piaget and Skinner, for example, have different but adequate explanations for Montessori teaching. This should not be surprising because there are many instances in science and other disciplines where different theoretical positions yield the same practical recommendations. For example, both the Ptolemaic view in astronomy, which holds that the Earth is at the center of the Universe, and the Copernican view of astronomy, which holds that the sun is at the center, yield the same navigational practices for traveling on Earth.

There is still another problem for teacher educators when they try to connect propositions in their educational foundations courses with practices in the schools. What teachers are able to do and what they actually do are often quite different. Even if they have a competence that would be sufficient to accomplish a teaching task, they may accomplish the task in a way that bypasses the competence altogether as can be seen in the following example. Virtually all adults, and most preschoolers, can give the plural of the following nonsense nouns – *wug*, *wot*, and *gutch*; in fact, they could even give the past tense if these same words were presented to them as verbs (e.g., Today A is *wugging* B, but yesterday A *wugged* B, and so on). Yet, almost no one knows the linguistic rule for the assignment of the plural (/z/, /s/, /ez/) or past tense (/d/, /ed/, /t/) allomorphs (*wugs*, *wots*, *gutches*), but everyone acts as if they knew the rule for the pluralizing these nonsense words even though it is obvious that they do not. They base their flawless plural and past tense performance on some factors of language competence apart from their formal knowledge of linguistics even though their formal knowledge of linguistics, if they had it, would have been sufficient for them

to pluralize the nonsense nouns and form the past tense of the nonsense verbs.

Thus, even when teacher educators succeed and the prospective teacher learns the content of the teacher education curriculum, there is still no guarantee that the teacher based his or her teaching practice on some knowledge of educational psychology, history, philosophy, or some body of research literature acquired in the teacher education program and its foundations courses. The fact that they could have, might have, or even should have, is no evidence that they did, or would ever, base their practice on the knowledge they acquired in their teacher education program.

Consequently, any effort to enhance the teacher's knowledge of the educational foundations could ultimately be pointless because the expert teacher, like the fluent speaker, could very well have based his or her performance on some other area of competence (Peterson, 1946). Even if prospective teachers have acquired a competence based on their study of the educational foundations, there is still the question of whether they would be disposed to use that competence in the classroom (Murray, 2007).

Having said this, however, we know that how the teacher does theorize or speculate about school events is a powerful determiner of the events. Prospective teachers, and unfortunately some experienced teachers, make a number of teaching errors that one day may be viewed as malpractice. These seem to stem from a naïve representation of teaching and schooling (Brophy and Good, 1986; Evertson *et al.*, 1985). For example, novices, especially those with the best and most decent intentions, blunder when they teach pupils who are very different from themselves. They treat pupils from whom they have low expectations in the following ways: they seat them outside the zone of frequent interaction; they treat them as a group rather than as individuals; and they look at them and call on them less often. When they do call on them, they ask them lower-level questions and give them less time to respond and fewer hints than they give their more able pupils. They give them less praise, even when they are correct, and more blame when they are wrong than they give pupils from whom they have higher expectations. They may do all this not only out of a mistaken sense of kindness, but also an impoverished view of the nature of teaching and the factors that figure in it (Hawley and Rosenholtz, 1984; Murray, 1996). There is hardly a better example of how the teacher's prior beliefs and implicit theory of classroom events influences practices than this example of how teaching practices flow from the teacher's expectations and predictions about his or her pupils' ability and effort.

Even if the teacher had acquired the information about the diverse groups of pupils in the classroom from the teacher education program, and was disposed to act in

sympathy with it, the teacher, even an experienced teacher, might still fail to act on the information in an appropriate manner (Murray, 2007). Cohen (1990), for example, reports a case of one enthusiastic teacher, Mrs. O, who believed she had adopted an innovative teaching approach only to have it shown that her innovative teaching was stubbornly hobbled with traditional and natural techniques which undermined her good intentions.

Given that so much rests upon teachers' expectations, predictions, and the way they think about schooling, the role of the foundations courses is to insure that teachers can evaluate evidence, can spot a fad or an unsound proposal for innovation, have an educated view of how the pupil's mind develops, have a reasoned and informed position on the major public policy issues that affect schooling, and the intellectual resources to face novel and unexpected classroom problems and opportunities. Amateurs often make bad guesses and predictions and they have few defenses against destructive educational fads because common sense and folk wisdom are often inadequate for the analysis of genuinely new and novel events.

Had the US and the UK teachers and administrators in the 1960–70s known more educational history, for example, they might not have required, as they did, one-quarter of USA and nearly all the UK pupils to learn to read in a special alphabet (*ita*, the initial teaching alphabet) that had a separate letter or symbol for each of the 45 sounds (phonemes) of English (e.g., *wuns* in *ita* would be once in traditional orthography). Had they known the history of spelling-reform efforts, they would have known that *ita* was but a version of *phonataby*, another alphabet used to the same end in the 1860s and the reasons why it was discarded. Had they known more educational psychology (in particular, the Osgood learning transfer surface), they could have predicted that the shift from *ita* to the regular alphabet would harm the pupil's spelling (negative transfer) even though it would actually help reading (positive transfer). Had they known more about educational research, they would have evaluated the claims for the benefits of *ita* more carefully than they did, quite differently from their knowledge of the history of spelling reforms or positive and negative transfer in learning psychology.

The study of the foundations disciplines seeks to give teachers a representation of schooling that allows them to ask for and search out the information that would be needed to evaluate the claims of any innovator. While they may have no information about *phonataby*, *per se*, they would know what kind of information they should have before implementing the innovation.

Prospective, novice, and experienced teachers are constantly presented with educational happenings or phenomena (girls on the whole do better on language problems, boys do better on spatial relations problems; the words in the middle of any list take longer to learn than the words at the beginning or end; pupils can recognize more words than

they can produce; some pupils learn their lessons faster than other pupils; some pupils conform to the rules of the classroom and some resist them; and so on). Teachers will inevitably speculate on the reasons for these and other phenomena they find in their classrooms. The complete account of an educational phenomenon entails the search for answers to a number of questions about a phenomenon that are informed by the foundations disciplines: what kind of phenomenon is it (was it a learning issue, motivational, social class, or genetic?); how did it get that way (was it always that way, only recently, only that way for certain groups, or does it come and go?); how does it work (was it produced through learning, maturation, or culture?); and how can it work better, and what is the point of it (fitting pupils for the economy, increased social justice, self-actualization)? The preponderance of recent scholarship in education has centered on the question, "How does it work," that is, the mechanisms that make schooling work. However, at least four aspects of any educational event – whether at the pupil or school system level – also need to be specified before it can be said that the event is completely intelligible, understood, or explained. The complete account treats:

1. the substance of the event: the relevant biology of the behavior in the case of the pupil or in the case of a classroom, school, or system, the various tangible agents, laws, policies, and structures that make it up.
2. the antecedent conditions that are necessary and sufficient for the event: those which elicit, trigger, and maintain the event. These are the causes of the event, in the usual sense of the term cause. In the case of pupil accomplishment, it is thought to have something to do with what the teacher did and knew.
3. the form or structure of the event or that which makes the event one phenomenon rather than another. If a child finally comes to say $2 + 2$ always equals 4 and never any other number, is this a change in language, logic, learning, teaching effectiveness, personality, memory, or what? What is the change a change of and how could we know? If a school begins to group pupils according to their standardized test scores, what kind of change has occurred in the school and community? When can historical events be said to coalesce as one historical period or era rather than another?
4. the purpose or point of the event, what it leads to, how it fits in with everything else, and how other events in the pupil's life, for example, makes the event intelligible and vice versa. Does the event mean anything? What is the point, significance, or meaning of the event?

Consider the curious fact that some young pupils, even those who know how to count correctly, think that spreading out a row of five marbles makes the row have more than five marbles while other pupils know that the number in the row is still five. How could the teacher

figure out how is it that some children know that the number of objects is independent of their spatial array while other children act as if the array also changed the number of objects (concepts of this sort are found at all levels of the curriculum)? Some teachers see the issue as a matter of a logical deduction, the child's ability to deduce the correct answer, and others see it merely as a matter of memory – being able to recognize and give the right answer. We will return to this example later.

Many issues in education are really debates about whether the explanation for a phenomenon is to be found solely in the antecedent conditions for it or in its form or structure as well. The debate is not so much about the student's accomplishment, but about what it means. For example, in one view there is alarm when a pupil who had used mice as the plural for mouse begins to say *mouses*, while in the other view, the reintroduction of *mouses*, wrong as it is, is a sign of linguistic progress because it shows a newly found sensitivity to a new class of linguistic rule. Along similar lines, in one view the pupil who argues that Churchill was right to firebomb Dresden gives exactly the opposite answer from the pupil who argues Churchill's act was wrong. Yet, the two arguments, while contradictory on the surface, may be structurally identical and signify that the two pupils have achieved precisely the same level of moral and logical development. On a larger scale, some see the function of the schools as the liberating transmission of the culture from one generation to the next, while others see the very same events as the imposition of a class structure that represses individual freedom and reduces opportunities.

The attempt to explain educational change totally in terms of antecedent events, the necessary and sufficient conditions, carries with it a determinism that many think precludes the possibility of spontaneous events, of emergent events, of discontinuities, of stages in the pupil's grasp of a subject matter insofar as spontaneity can exist only when events cannot be reduced to, or linked strictly and solely to, those events that come before them. Thus, it could be said that theories that allow for educational changes that are inherently unpredictable from their antecedents introduce a more sophisticated or advanced view of education.

If all there is to know about an earlier educational period is known, it still may not allow the teacher to predict the nature of the pupil's subsequent educational accomplishments. The point here is that so much depends upon the teacher's view of the issue and the way in which he or she thinks about whether the pupil's behavior is determined or open-ended and whether the object of instruction is the correct response or the underlying structure of the response, and so on.

Consequently, because teachers, like every reflective person, invent theories and explanations, teacher educators need some criteria for evaluating the available

theories in the educational foundations before including them in the teacher education curriculum. There simply are differences in the power and adequacy of various educational theories that must be acknowledged and respected.

What must we have from a good theory? What questions must it answer about the classroom phenomena it attempts to explain? What problems must it resolve or eliminate? All educational phenomena have more or less the same basic form – there is some change, or difference, in the student (or teacher, class, school, district, etc.) that is found over a relatively long period of time (weeks, months, years) in the case of development, or in a relatively short time (seconds, minutes, hours, and days) in the case of learning. What must a good theory tell us about these changes and differences? What should prospective teachers be thinking about, or considering, to shift their theorizing from those of an amateur to those of a professional?

Contemporary scholarship points to ten categories that a good theory should address and that the teacher should consider:

1. the form or pattern of the change – a way to identify, name, or define the phenomenon, a way to distinguish it from other phenomena. There must be a way to measure it and perhaps produce it. The act of naming or defining carries the risk that more may be thought to be known and understood than really is, but it is an essential part of any theory to name the phenomenon and the criteria for the naming. Thus, we call the child's incorrect response to the spatial array of the number of marbles, nonconservation, and thereby mark it off from other phenomena and begin the act of knowing and making sense of it.
2. cause – the causes of the phenomenon, the necessary and sufficient conditions for it – the eliciting conditions. This is not to say that the specifications of the necessary and sufficient conditions constitute the whole account or explanation, but only that they are a part of the story, so to speak, and without their specification, a theory would not be satisfying. Thus, we want to know under what conditions the child will assert that there are as many marbles as there were before the row of marbles was spread out, and under what conditions the child will claim there are more or fewer marbles.
3. mechanisms – what cognitive device produces the phenomenon. How do these devices function, and how, over the time span during which the change takes place, do they actually produce the change? In the number of marbles in the row example, what roles do learning, social interaction, imitation, cognitive dissonance, mediation, maturation, perception, and so on play in the child's exhibition of the

phenomenon? How do they lead the student who was wrong to be right?

4. the educational end point – the educational goal for the changes. Later events cannot cause earlier events, but they can help us make sense of the earlier events by showing what the earlier events lead to – nonconservation leads to conservation and not the other way around. The child's error that the number of marbles changed as they were spread out would make more sense if we had a way to see it as an instance of the child's newly acquired notion of logical necessity and not merely as an episodic peculiarity of the young child's thought (Murray, 1990). It is helpful to know why the earlier periods fail to hold the developing mind at some point or stage for a longer time even without knowing the character of the next stage or period. Because of the open-ended character of education and the potential for inherently unpredictable outcomes of mental functioning and accomplishment, it may be inevitable that theorists will always understand more of where the mind has been, so to speak, than where it is going.

Moreover the good theory may provide a way to think about better educational outcomes. Given that many educational outcomes are possible, and that just as many evolutionary solutions are possible for species development, the good theory could be asked to account for, identify, explain, and clarify the better of the available outcomes – outcomes that maximize what it means to be human. We ask that the good theory critically examines the educational outcomes that appear to be necessary and unalterable to determine whether they are really just one of a range of possible educational outcomes.

5. the meaning of the phenomenon – the attribute of intentionality and the question of what the student meant or intended when he/she said the row now had more than five marbles. The question of the meaning or significance of the phenomenon is a micro-version of the question of the goal of education. It is an inquiry into the purpose and significance of some aspect of mental life.

Just as there is no single interpretation of a text, apart from a framework of interpretation, one would not expect that there could be a single meaning of a behavioral phenomenon that was unrelated to a theory or an interpretative framework that could give meaning to the event.

6. unity of knowledge – the good theory must have the potential for coherence with the other disciplines. Thus, we would expect not to be surprised that the onset of conservation (knowing the number of marbles remained the same), for example, is linked to increases in brain lateralization, brain surface, the completion of myelination, increases in EEG alpha activity, and increases in working memory.

7. deductive formalism – while the logic of discovery and the logic of justification are different, the good theory should eventually have a form such that phenomena are explained by virtue of their being implications of general propositions of a theory. The appeal for a deductive formalism in the good theory does not mean that theory building needs to proceed by that route, but only that eventually there would be a version of the theory that can be expressed in the fashion of the hypothetico-deductive system (e.g., conservation makes sense within a developmental theory that explains many other phenomena like conservation and some others that are very different like seriation, transitivity, class inclusion, and egocentrism, but all derivable from a hypothetical device – concrete operational thought).
8. cohort specification – it is now understood that generational and historical effects compromise the generality of many findings, like the shape of the growth of intelligence function (Baltes *et al.* 1977). Such effects should now be seen as an integral part of the phenomenon. Of course, cohort, or time of the subject's birth, is merely a proxy for some yet-to-be-discovered historical and cultural factors that operated during a particular period, and that these factors would significantly alter research findings that are reported in research paradigms that were not sensitive to generational factors.
9. cultural and social determinants – the identification of factors and mechanisms that operate uniquely in particular historical periods, and not in other periods, has led researchers to consider wider ranges of these contextual and interactive factors (Rogoff and Lave, 1984).
Native Americans in the southwest, are not fooled as much by conservation problems about the clay ball's weight as other children appear to be, and Bedouin children are not fooled as much as others about conservation of liquid amount (see Murray, 1981). In virtually every domain of psychology, substantial effects can be attributed to factors that appear to be features of a particular context, social or cultural group, geographic location, historical time period, and so on. Obviously the successful theory will find a way to make sense of this – at the moment – bewildering array of context-specific influential factors.
10. the theorist – the interdependence of fact and theory or text and interpretation leads to an examination of the theorist as a person and thinker. The interpretive framework that allows events to be facts in a science is shaped presumably by personal features of the theorist that heretofore were considered irrelevant when science was viewed exclusively as objective and self-correcting. At the moment we can only

speculate how Piaget's account of moral development would differ had it been formulated outside Protestant Geneva, or how Skinner's account of learning would differ had it not been formulated by an American and so forth. However, as theories are invariably written and otherwise promulgated, their meaning is also subject to all the hermeneutic issues implicated in the attempt to specify the meaning and significance of an educational change in mental functioning in the first place.

In summary, a good theory of schooling is what will eventually be found in the educational foundations disciplines of the teacher education curriculum, and it will be a key ingredient in what the teacher relies upon in professional decision making.

Bibliography

- Astington, J. W. and Pelletier, J. (1996). The language of mind. In Olson, D. R. and Torrance, N. (eds.) *The Handbook of Education and Human Development*, pp 591–619. Oxford: Blackwell.
- Ashley, J. and Tomasello, M. (1998). Cooperative problem solving and teaching in preschoolers. *Social Development* 7, 143–163.
- Baldwin, A. (1980). *Theories of Child Development*, 2nd edn. New York: John Wiley.
- Baltes, P., Reese, H., and Nesselroade, J. (1977). *Life-Span Developmental Psychology: Introduction to Research Methods*. Monterey, CA: Brooks/Cole.
- Berliner, D. (1988). Implications of studies of expertise in pedagogy for teacher education and evaluation. *New Directions for Teacher Assessment Proceedings of the 1988 ETS Invitational Conference*, pp 39–68. Princeton, NJ: Educational Testing Service.
- Brophy, J. and Good, T. (1986). Teacher behavior and student achievement. In Wittrock, M. (ed.) *Handbook of Research on Teaching*, 3rd edn, pp 328–375. New York: Macmillan.
- Cohen, D. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis* 12, 311–330.
- Evertson, C., Hawley, W., and Zlotnick, M. (1985). Making a difference in educational quality through teacher education. *Journal of Teacher Education* 36(3), 2–12.
- Fischer, K. and Bidell, T. (1998). Dynamic development of psychological structures in action and thought. In Damon, W. (ed.) and Lerner, R. (vol. ed.) *Handbook of Child Psychology*, vol. 1, pp 467–561. New York: John Wiley & Sons.
- Gilbert, S. F. and Borish, S. (1997). How cells learn, how cells teach: Education in the body. In Amsel, E. and Renniger, A. K. (eds.) *Change and Development: Issues in Method, and Application*, pp 61–76. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hawley, W. and Rosenholtz, S. (1984). Good schools: What research says about improving student achievement. *Peabody Journal of Education* 61(4).
- Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.
- Kuhn, D. (1991). *The Skills of Argument*. New York: Cambridge University Press.
- Maynard, A. (2004). Cultures of teaching in childhood: Formal schooling and Maya sibling teaching at home. *Cognitive Development* 19, 517–535.
- Murray, F. (1981). The conservation paradigm: Conservation of conservation research. In Sigel, I., Brodzinsky, D., and Golinkoff, R. (eds.) *New Directions and Applications of Piaget's Theory*, pp 143–175. Hillsdale, NJ: Lawrence Erlbaum.
- Murray, F. (1990). The conversion of truth into necessity. In Overton, W. (ed.) *Reasoning, Necessity, and Logic: Developmental Perspectives*, pp 183–203. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Murray, F. (1991). Questions a satisfying developmental would answer: The scope of a complete explanation of developmental phenomena. In Reese, H. (ed.) *Advances in Child Development and Behavior*, vol. 23, pp 39–47. New York: Academic Press, Inc.
- Murray, F. (1996). Beyond natural teaching: The case for professional education. In Murray, F. B. (ed.) *The Teacher Educator's Handbook*, pp 3–13. San Francisco, CA: Jossey-Bass.
- Murray, F. (2007). Disposition: A superfluous construct in teacher education. *Journal of Teacher Education* **58**, 381–387.
- Olson, D. and Bruner, J. (1996). Folk psychology and folk pedagogy. In Olson, D. R. and Torrance, N. (eds.) *The Handbook of Education and Human Development*, pp 9–27. Oxford: Blackwell.
- Peterson, H. (ed.) (1946). *Great Teachers Portrayed by Those Who Studied under Them*. New York: Vintage Books.
- Premack, D. and Premack, A. J. (1996). Why animals lack pedagogy and some cultures have more of it than others. In Olson, D. R. and Torrance, N. (eds.) *The Handbook of Education and Human Development*, pp 302–323. Oxford: Blackwell.
- Rogoff, B. and Lave, J. (eds.) (1984). *Everyday Cognition: Its Development in Social Context*. Cambridge, MA: Harvard University Press.
- Strauss, S. and Silhoney, T. (1994). Teachers' models of children's mind and learning: Implications for teacher education. In Hirschfeld, L. A. and Gelman, S. A. (eds.) *Mapping the Mind: Domain Specificity in Cognition and Culture*, pp 455–473. New York: Cambridge University Press.
- Strauss, S., Ziv, M., and Stein, A. (2002). Teaching as a natural cognition and its relation to preschoolers' developing theory of mind. *Child Development* **17**, 1473–1487.
- Strauss, S. and Ziv, M. (2004). Teaching: Ontogenesis, culture, and education. *Cognitive Development* **19**, 451–456.
- Visalberghi, E. and Frigaszy, D. M. (1976). Pedagogy and imitation in monkeys. In Olson, D. R. and Torrance, N. (eds.) *The Handbook of Education and Human Development*, pp 277–301. Oxford: Blackwell.

Further Reading

- Cochran-Smith, M. and Zeichner, K. (eds.) (2007). *Studying Teacher Education*. Mahwah, NJ: Lawrence Erlbaum and Associates.
- Cochran-Smith, M., Nemser, S. F., and McIntyre, J. (eds.) (2008). *Handbook of Research on Teacher Education: Enduring Issues in Changing Contexts*, 3rd edn. New York: Routledge/Taylor & Francis Group and the Association of Teacher Educators.
- Draper, P. (1976). Docile and economic constraints on child life among the !Kung. In Lee, B. and Devore, I. (eds.) *Kalahari Hunter-Gatherers*, pp 199–217. Cambridge: Harvard University Press.
- Floden, R. E. and Meniketti, M. (2005). Research on the effects of coursework in the arts and sciences in the foundations of education. In Cochran-Smith, M. and Zeichner, K. (eds.) *Studying Teacher Education*, pp 251–308. Mahwah, NJ: Lawrence Erlbaum & Associates.
- Konner, M. (1976). Maternal care, infant behavior and development among the !Kung. In Lee, B. and Devore, I. (eds.) *Kalahari Hunter-Gatherers*, pp 218–245. Cambridge: Harvard University Press.
- Wilson, S. M., Floden, R. E., and Ferrini-Mundy, J. (2001). *Teacher Preparation Research: Current Knowledge, Gaps, and Recommendations*. University of Washington: Center for the Study of Teaching and Policy.

The Relationship Between Theory and Practice in Teacher Education

F A J Korthagen, VU University, Amsterdam, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

A Perennial Problem of Teacher Education

In 1904, John Dewey noted a gap between theory and practice in teacher education (Dewey, 1904), and he discussed possible approaches to bridging this gap (see also Shulman, 1998). Nevertheless, during the whole of the twentieth century, it has remained the central problem of teacher education worldwide (Lanier and Little, 1986). In various analyses of this problem, the focus has often been on the question of how practice can be better linked to theory. Only relatively recently, has attention focused on the possibility that theory can be better linked to practice.

Below, the theory–practice relationship will be analyzed, as well as the nature of knowledge relevant to practice. We will also discuss different approaches to improve the link between practice and theory, which leads to the conclusion that no single trick will solve the perennial issue.

The Sacred Story about Theory and Practice

During the late nineteenth and early twentieth century as psychological and pedagogical knowledge developed, academics wanted to offer this knowledge to teachers in order to change education and adapt it to scientific insights. This is how the idea of the professionalization of teachers began. Indeed, as Hoyle and John (1995) point out, the availability of a recognized body of knowledge is one of the most important criteria for categorizing an occupational group as professional (see also McCullough, 1987). It seemed only logical to teach important theories to preservice and in-service teachers, who could then apply this knowledge base in their teaching. Clandinin (1995) calls this “the sacred theory–practice story”, and Carlson (1999) speaks about the “theory-to–practice approach”. Wideen, *et al.* (1998) put it like this:

... the implicit theory underlying traditional teacher education was based on a training model in which the university provides the theory, methods and skills; the schools provide the setting in which that knowledge is practiced; and the beginning teacher provides the individual effort to apply such knowledge. In this model, propositional

knowledge has formed the basis of university input. (Wideen, *et al.*, 1998: 167)

In many places, this has led to a system in which experts in certain domains teach this knowledge to prospective teachers. As Barone *et al.* (1996) argue, the result is often that a teacher program consists of a collection of isolated courses in which theory is presented with hardly any connection to practice. Ben-Peretz says:

The hidden curriculum of teacher education tends to communicate a fragmented view of knowledge, both in coursework and in field experiences. Moreover, knowledge is “given” and unproblematic. (Ben-Peretz, 1995)

Schön (1983, p. 21) speaks about the technical-rationality model, which is based on the notion that “professional activity consists in instrumental problem solving made rigorous by the application of scientific theory and technique.” In fact, three basic assumptions are hidden in this view (cf. Hoyle, 1980):

1. theories help teachers perform better in their profession;
2. these theories must be based on scientific research; and
3. teacher educators should make a choice concerning the theories to be included in teacher education programs.

The technical-rationality model has been dominant for many decades (Sprinthall, *et al.*, 1996; Imig and Switzer, 1996, p. 223), although many studies have shown its failure in strongly influencing the practices of the graduates of teacher education programs, which is discussed in the next section.

Meager Impact of Teacher Education

Many researchers have shown that the traditional technical-rationality model does not function well. Zeichner and Tabachnick (1981), for example, discussed that many notions and educational conceptions, developed during preservice teacher education, were washed out during field experiences. Comparable findings were reported by Cole and Knowles (1993) and Veenman (1984). Veenman also points toward the severe problems teachers experience once they have left preservice teacher education, a phenomenon named the transition shock. It is interesting to note that this transition shock is described in studies from many different countries.

For example, at Konstanz University in Germany, large-scale research has been carried out into this phenomenon (Müller-Fohrbrodt, *et al.*, 1978; Dann, *et al.*, 1981; Hinsch, 1979). It showed that teachers pass through a quite distinct attitude shift during their first year of teaching, in general creating an adjustment to practices current in the schools, and not to recent scientific insights into learning and teaching. Building on the work of the Konstanz research group, Brouwer and Korthagen (2005) did an extensive quantitative and qualitative study in the Netherlands among 357 student teachers, 128 cooperating teachers, and 31 teacher educators, which again showed the dominant influence of the school on teacher development (see also an early study by Lortie, 1975).

In their well-known overview of the literature on teacher socialization, Zeichner and Gore (1990) put forward that researchers differ in the degree to which they consider teacher socialization to be a passive or an active process. However, all studies on teacher development emphasize that it is very difficult for an individual to really influence established patterns in schools. Educational change appears to be a beautiful ideal of teacher educators, but generally indeed not much more than an ideal. Bullough (1989) emphasizes that in this respect there is a severe problem in teacher education. As Zeichner and Gore (1990, p. 343) put it:

Studies that have focused on the institutional and cultural levels of analysis have clearly shown, for example, that various ideological and material conditions within teacher education institutions, schools, and societies serve to establish limits on the range of options available to both teacher education students and teacher educators.

As a result, the impact of teacher education on the practice of their students is very limited, as Wideen, *et al.* (1998) conclude in a thorough and extensive review of the international research into the outcomes of teacher education, a conclusion that was also drawn by the AERA Research Panel on Teacher Education in a meta-analysis of North-American research (Cochran-Smith and Zeichner, 2005). Several studies show that beginning teachers struggle for control, and experience feelings of frustration, anger, and bewilderment. The process they go through is more one of survival than of learning from experiences. Novice teachers do not feel sufficiently prepared by their teacher educators, and come to view colleagues in the schools as realistic role models, as the people who do know how one should go about teaching.

Apart from the fact that the traditional technical-rationality approach to teacher education creates little transfer from theory to practice, this approach creates another fundamental problem. Elliot (1991, p. 45) states that teachers who realize that they are not able to use the theory presented to them by experts, often feel threatened by theory. These feelings of threat are further enhanced

by the generalized form in which experts tend to formulate their knowledge and by the ideal views of society or individuals behind their claims. Often the result is a dislike of theoretical deepening in teachers (Cole, 1997).

Causes for the Gap between Theory and Practice

As Robinson (1998, p. 17) reminds us, “narrowing the research-practice gap is not just a matter of disseminating research more effectively or of using more powerful influence strategies.” The causes for the gap lie deeper and a variety of these causes have been put forward in the literature. First, we have already seen that from a sociological perspective one can frame the problem as one of socialization toward patterns existing in the schools.

A second cause often mentioned in the literature is the complexity of teaching (e.g., Hoban, 2002, p. 35–40). Hoban (2005, p. 9) states: “(. . .) what a teacher does in a classroom is influenced by the interaction of many elements such as the curriculum, the context, and how students respond to instruction at one particular time.” Hoban continues by saying that this view of the nature of teaching necessitates holistic judgment (cf. Day, 1999) about what, when, and how to teach in relation to a particular class, and this is something for which it is hard to prepare teachers. Moreover, practice is generally ambiguous and value-laden (Schön, 1983). Robinson (1998), states that in specific educational situations, even experts have different opinions of what is the best way to use theory. Different theories may each have their value in explaining a certain aspect of the situation, and lead to different perspectives.

A third cause for the theory–practice divide often mentioned has to do with the learning process within teacher education itself, even before the stage in which theory can be applied to practice. According to many researchers, teachers’ prior knowledge plays a powerful role in their learning (Wubbels, 1992). Their preconceptions show a remarkable resistance to change (Joram and Gabriele, 1998). This can in part be explained by their firm roots in the many years of experiences that student teachers themselves have had as students within the educational system (Lortie, 1975). Preconceptions also shape the way new knowledge is being understood. Stofflett and Stoddart (1994), for example, argue that teachers’ conceptions of how subject matter should be taught are strongly influenced by how they themselves learned the subject content. These authors showed that student teachers who themselves experienced learning in an active way, are more inclined to plan lessons that facilitate students’ active knowledge construction. Huibregtse *et al.* (1994), showed that, even with experienced teachers, there is a strong relation between their preferred way of teaching

and the way they themselves are used to learning: they have a limited view of the learning styles of their students, and tend to project their own way of learning onto the learning of their students.

A fourth cause has been named as the feed-forward problem: “resistance from the student teacher at the time of exposure to given learnings and, later, protestations that the same learning had not been provided in stronger doses” (Katz *et al.*, 1981, p. 21; see also Bullough *et al.*, 1991, p. 79). This problem can also be stated as follows: in order to learn anything during teacher education, student teachers must have personal concerns about teaching or they must have encountered concrete problems. Otherwise, the usefulness of the theory is not clear to them, and they are not motivated to study it. Later, when they do come across problems, there is often no opportunity to acquire the relevant theoretical insights.

Other authors add that teacher development cannot be fully understood if it is considered merely from a cognitive perspective. Teaching is a profession in which feelings and emotions play an essential role (Nias, 1996; Hargreaves, 1998a), but “the more unpredictable passionate aspects of learning, teaching and leading (...) are usually left out of the change picture” (Hargreaves, 1998b, p. 558). The problem of promoting fundamental professional change is first of all a problem of dealing with the natural emotional reactions of human beings to the threat of losing certainty, predictability, or stability. This affective dimension is neglected very often in the technical-rationality approach, which is another cause for the transfer problem.

The human aspect can be elaborated even further. Many authors emphasize that learning to teach cannot be separated from the person of the learner (see for an overview Korthagen, 2004). Bullough (1997, p. 21) states that “teacher identity-what beginning teachers believe about teaching and learning and self-as-a-teacher-is of vital concern to teacher education; it is the basis for meaning making and decision making.” He adds that the most important learning outcomes will thus be personal, idiosyncratic, and probably immeasurable. This concurs with Day (1999, p. 94) who talks about the “unpredictable, personalized nature of teaching.”

Finally, there is a cause for the transfer problem that has drawn so much scholarly attention that the entire next section is devoted to it.

Practical versus Formal Knowledge

Clark and Lampert (1986, p. 28) argue that once inside school, teachers “are expected to accomplish complex and even conflicting goals. Under these circumstances, a priori knowledge identified by researchers about the relationship between particular decisions or actions and their

outcomes, is of limited worth.” Teachers often have little time to think and thus need prompt and concrete answers to situations (Eraut, 1995). Action-guiding knowledge is rather different from the more abstract, systematized, and general expert knowledge teacher educators often present to student teachers (Tom, 1997). Various terms are used to name this difference, but generally used are the concepts of practical knowledge and formal knowledge. According to Fenstermacher (1994), the former type of knowledge develops in teachers by participating in and reflecting on their own actions and experiences; it is situated knowledge (Brown, *et al.*, 1989). This practical knowledge enables teachers to deal effectively with practical problems. The validity of this knowledge is confined to the type of contexts or situations within which the events occur, and teachers often have difficulty in putting it into words. This lack of explicitness may also limit student teachers’ learning from the practical knowledge of their mentor teachers if no additional measures are taken (Zanting, *et al.*, 2001).

Formal knowledge, or propositional knowledge, is the knowledge produced by conventional research in order to answer a question such as: What is known about effective education? It meets criteria for reliability, validity, and has the potential for generalization. Fenstermacher refers to Aristotle, who already made a similar distinction between two types of knowledge, which he called *phronesis* and *episteme*.

Kessels and Korthagen (1996) emphasize that formal knowledge or *episteme* (which they also refer to as theory with a capital T) should certainly not be absent from teacher education programs, as, now and then, student teachers should be helped to see the larger picture of educational knowledge. On the other hand, they add that:

More often, however, they need knowledge that is situation-specific and related to the context in which they meet a problem or develop a need or concern, knowledge that brings their already existing, subjective perception of personally relevant classroom situations one step further. This type of knowledge is called *phronesis*. We could also call it ‘theory with a small t’.

The character of *phronesis* is more perceptual than conceptual: it focuses the attention of the actor in the situation on certain situational characteristics important to the question as to how to act in that situation. To put it concisely, *episteme* aims most of all at knowing more about many situations, while the emphasis of *phronesis* is on perceiving more in a particular situation and finding a helpful course of action on the basis of strengthened awareness. The best translation of *phronesis* may be practical wisdom.

An important cause for the limited impact of many teacher education programs may be that they focus too much on formal knowledge (*episteme*) and do not support

their students sufficiently in developing their perceptual awareness or *phronesis* (Loughran, 2006, p. 8–9). This could mean that teacher educators themselves create the gap between theory and practice. Of course, the conditions under which teacher education takes place are generally not very supportive of a change in old habits: large enrolments and limited time for teacher educators to visit student teachers during their teaching practice are some of the most significant inhibiting factors (Barone *et al.*, 1996, p. 1117).

Directions for Bridging the Gap between Theory and Practice

During the recent decades, several strategies have been introduced into teacher education with the aim of bridging the gap between theory and practice. It is difficult to present a complete overview of all of them, but in this section some productive strategies will be described.

First, several attempts have been made to improve the theory-into-practice approach through pedagogical strategies, such as the promotion of reflection (e.g., Clift *et al.*, 1990; Schön, 1987), or through tools like (video) cases, portfolios, etc. Each of these has its merits in helping teachers develop useful action-guiding theory with a small *t*, or in supporting them to make connections between existing theory and their own implicit thinking about education. However, a fundamental solution to the theory–practice divide may require more radical changes in program structure, in which “greater continuity exists between teacher preparation and the schools where beginning teachers begin their teaching careers” (Wideen *et al.*, 1998, p. 159). Before discussing this promising development, a warning has to be given regarding an extreme elaboration of this idea. In many programs, the traditional approach of theory first, practice later has been replaced by the adage practice first, theory later (Sandlin, *et al.*, 1992). Alternative certification programs have been created in which novice teachers sometimes receive very little theoretical background, and teacher education becomes more of a process of guided induction into the tricks of the trade. In many places in the world, this trend is also influenced by the need to solve the problem of teacher shortages. Although this development may satisfy teachers, politicians, and parents, there is a great risk involved. The balance seems to shift completely from an emphasis on theory to reliance on practical experience. Such an approach to teacher education does, however, not guarantee success. Long ago, Dewey (1938, p. 25) already stated that “the belief that all genuine education comes about through experience does not mean that all experiences are genuinely or equally educative.” (cf. Loughran, 2006, p. 22) As already discussed above, teaching experience can be a process of socialization rather than an

opportunity for professional development (cf. Wideen, *et al.*, 1998). Hence, there is a risk that in a practice-first approach, the basic question, namely how to integrate theory and practice, is still not solved.

Nevertheless, promising developments are visible under the umbrella of professional development schools (PDSs) (Bullough and Kauchak, 1997; Darling-Hammond, 1994). The idea is to develop collaborative partnerships between institute-based teacher educators and school-based teachers, sharing the responsibility for the preparation of prospective teachers. In this context, there is a focus on developing new teaching methods, and an emphasis on an ongoing professional development for all involved in such projects (Abdal-Haqq, 1997).

Both in the PDS movement and in the more general trend to move teacher education to the schools, there are two significant aspects: on the one hand, these developments mirror a wish to ground teacher education more strongly within practical contexts, but on the other hand, teacher education faculty involved tries to avoid the risk of early socialization toward traditional educational patterns. That is why these attempts at strengthening the practical relevance of teacher education are characterized by an emphasis on critical reflection on current practices, and attempts to adapt formal knowledge to practice. This requires strong and long-term partnerships between university faculty and teachers in the schools, which in certain places have been very effective (see e.g., the Oxford Internship Model described by McIntyre and Hagger, 1992).

Referring to publications by authors supporting collaboration structures between schools and universities, Furlong *et al.* (1996) conclude that the PDS approach may finally allow for a real integration of theory and practice. However, case studies also illustrate many problems associated with the PDS approach (see Darling-Hammond, 1994). Castle (1997, p. 221) concludes that “many of the problems stem from the reality that change of this nature involves individuals and relationships.” An important problem is that when a leading person moves to another job, the whole collaboration structure may collapse.

Other strategies of which their advocates claim that they really bridge the gap between theory and practice are those putting teacher research in the forefront of professional development. (Cochran-Smith and Lytle, 1993; Stenhouse, 1975). Ponte (2005) observes that the idea that teachers themselves can do research, goes back as far as the beginning of the twentieth century, and has led to a variety of slightly different approaches to teacher research. Lewin (1946), for example, introduced the term action research and emphasized its role in promoting social change. Other common terms are reflective inquiry, practitioner research, and self-study research (see for an overview McKernan, 1996). One of the challenges of teacher research is that “it is in danger of becoming

anything and everything" (Cochran-Smith and Lytle, 1999, p. 21). This has led to attempts to develop explicit methodologies for this kind of research (e.g., Bullough and Pinnegar, 2001), and pleas to put more effort into connecting the outcomes of individual studies to the broader research community (Zeichner, 2007).

Another fruitful strategy for linking theory and practice focuses on the concerns students develop in practice, and tries to build teacher education on these concerns and student teachers' preconceptions that these concerns bring to light. Korthagen *et al.* (2001) describe how this can be done in a so-called realistic approach, and present evidence that this does indeed bridge the theory–practice gap. Others focus on the importance of making student teachers' preconceptions explicit through their narratives about practice (e.g., Kelchtermans, 1993). However, enacting such approaches requires a strong investment in staff development, as they build on specific pedagogical interventions.

More recently, social constructivist views and the notion of communities of practice (Wenger, 1999) have led to approaches that no longer focus on the individual practitioner, but consider professional development as being strongly intertwined with professional cultures. Although the strength of such social constructivist approaches is that they start from the reality of professional development, which is always embedded in a certain social context, one does have to find ways to circumvent narrow-mindedness and undesirable socialization. In other words, the fundamental tension between what happens in a community of practice and the wish to deepen practice with existing theory is still there to be taken care of. According to Husu (2002) a key element is not to strive for unanimous agreement, but to focus on discourse and the testing of plural meanings.

Darling-Hammond (1999, 2006) identified a number of additional components of teacher education curricula that appear to be beneficial, such as a shared, clear vision of good teaching apparent in all coursework and clinical experiences; well-defined standards of practice; a rigorous core curriculum taught in the context of practice; and intensively supervised, extended clinical experiences (at least 30 weeks), carefully chosen to support what students have learned in their courses. The importance of this latter component is emphasized in a vast amount of literature claiming that the most important factor in promoting the relationship between theory and practice is individual coaching or supervision (e.g., Korthagen *et al.*, 2001; Showers and Joyce, 1996). The idea is that a strong supervisor may be able to effectively connect the student's personal experiences in educational settings and his/her present concerns to theory – both theory with a small t and theory with a big T. However, Franke and Dahlgren (1996) show that not all supervisors do so.

Toward an Integrated View

As Lanier and Little (1986) argue, all teacher education programs have to confront the problem of the gap between theory and practice. It is now more than a century since Dewey expressed his concern about this gap. Has the solution been found? Each of the approaches discussed in the previous section has its own merits and seems to solve part of the problem. Similar to teaching, learning to teach is also too multifaceted to be dealt with in a simple way. As Wideen *et al.* (1998) suggest, what is needed is an integrated view, in which all the aspects influencing teacher development are taken into account, and which combines several of the approaches mentioned above (compare Hoban, 2005).

The Research Needed

Finally, an additional issue has to be mentioned explaining why it has remained so difficult to bridge the theory–practice gap. For many decades, there has also been a gap between the research carried out on teacher education and the work of teacher educators in teacher education programs. Research on teacher education did reveal the meager impact of programs, but what actually goes on inside teacher education all over the world has remained obscure (Zeichner, 1999). As a result, research has generally been of little practical help to those teacher educators wishing to promote the integration of theory and practice, as Cochran-Smith and Zeichner (2005) observe. Only relatively of late has research into teacher education been carried out from an insider perspective (cf. Anderson and Herr, 1999). Especially, the growth of so-called self-study research by teacher educators (see e.g., Loughran *et al.*, 2004) has helped us learn more about the details of approaches that teacher educators use – which of these are effective and which are not. The educational community can benefit from studies describing what is really happening in preservice and in-service programs in different countries, especially from studies linking program goals with careful analyses of the behavior of teacher educators or cooperative teachers, and the effects on student teachers' learning processes. In such studies, it is important to pay special attention to the contextual influences of practicum schools. Day (1999) is an excellent example of this kind of perspective in in-service contexts. Such studies may clarify that it is possible to link theory and practice in teacher education in such a way that positive effects on graduates can be demonstrated, as Brouwer and Korthagen (2005) have shown. Through such research, chances may increase that teacher education can make a difference, and therefore that in the long run its success may counterbalance the criticism from politicians, school principals, parents, and even the teachers themselves.

See also: Action Research as a Tool for Teachers' Professional Development; Experienced Teachers' Craft Knowledge; Taking Prospective Teachers' Beliefs into Account in Teacher Education; Teacher Education and Models of Teacher Reflection; Teacher Learning as Workplace Learning.

Bibliography

- Abdal-Haq, I. (1997). *Professional Development Schools: Weighing the Evidence*. Washington DC: AACTE.
- Anderson, G. L. and Herr, K. (1999). The new paradigm wars: Is there room for rigorous practitioner knowledge in schools and universities? *Educational Researcher* 28(5), 12–21, 40.
- Barone, T., Berliner, D. C., Blanchard, J., Casanova, U., and McGowan, T. (1996). A future for teacher education. In Siluka, J. (ed.) *Handbook of Research on Teacher Education*, 2nd edn, pp 1108–1149. New York: Macmillan.
- Ben-Peretz, M. (1995). Curriculum of teacher education programs. In Anderson, L. W. (ed.) *International Encyclopedia of Teaching and Teacher Education*, pp 543–547. Oxford/New York/Tokyo: Elsevier Science/Pergamon.
- Brouwer, N. and Korthagen, F. (2005). Can teacher education make a difference? *American Educational Research Journal* 42(1), 153–224.
- Brown, J. S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* 18(1), 32–42.
- Bullough, R. V. J. (1989). *First Year Teacher: A Case Study*. New York: Teachers College Press.
- Bullough, R. V. (1997). Practicing theory and theorizing practice in teacher education. In J. Loughran, J. and Russell, T. (eds.) *Purpose, Passion and Pedagogy in Teacher Education*, pp 13–31. London/Washington DC: Falmer Press.
- Bullough, R. V. and Kauchak, D. (1997). Partnerships between higher education and secondary schools: Some problems. *Journal of Education for Teaching* 23(3), 215–233.
- Bullough, Jr. R. V., Knowles, J. G., and Crow, N. A. (1991). *Emerging as a Teacher*. London: Routledge.
- Bullough, R. V. and Pinnegar, S. (2001). Guidelines for quality in autobiographical forms of self-study research. *Educational Researcher* 30(3), 13–21.
- Carlson, H. L. (1999). From practice to theory: A social constructivist approach to teacher education. *Teachers and Teaching: Theory and Practice* 5(2), 203–218.
- Castle, J. B. (1997). Toward understanding professional development: Exploring views across a professional development school. *Teachers and Teaching: Theory and Practice* 3(2), 221–242.
- Clandinin, D. J. (1995). Still learning to teach. In Russell, T. and Korthagen, F. (eds.) *Teachers Who Teach Teachers*, pp 25–31. London: Falmer Press.
- Clark, C. and Lampert, M. (1986). The study of teacher thinking: Implications for teacher education. *Journal of Teacher Education* 37(5), 27–31.
- Clift, R., Houston, W. R., and Pugach, M. C. (eds.) (1990). *Encouraging Reflective Practice in Education: An Analysis of Issues and Programs*. New York: Teachers College Press.
- Cochran-Smith, M. and Lytle, S. (1993). *Inside/Outside: Teacher Research and Knowledge*. New York: Teachers College Press.
- Cochran-Smith, M. and Lytle, S. (1999). The teacher research movement: A decade later. *Educational Researcher* 28(7), 15–26.
- Cochran-Smith, M. and Zeichner, K. M. (eds.) (2005). *Studying Teacher Education: The Report of the Panel on Research and Teacher Education*. Washington DC: American Educational Research Association/Mahwah: Erlbaum.
- Cole, A. L. (1997). Impediments to reflective practice. *Teachers and Teaching: Theory and Practice* 3(1), 7–27.
- Cole, A. L. and Knowles, J. G. (1993). Teacher development partnership research: A focus on methods and issues. *American Educational Research Journal* 30(3), 473–495.
- Dann, H. D., Müller-Förhbrodt, G., and Cloetta, B. (1981). Sozialization junger Lehrer in Beruf: Praxisschock drei Jahre später. [Professional socialization of young teachers: The transition shock three years later.] *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 13, 251–262.
- Darling-Hammond, L. (1994). *Professional Development Schools: Schools for Developing a Profession*. New York: Teachers College Press.
- Darling-Hammond, L. (1999). America's future: Educating teachers. *The Education Digest* 64(9), 18–23.
- Darling-Hammond, L. (2006). *Powerful Teacher Education: Lessons from Exemplary Programs*. San Francisco: Jossey-Bass.
- Day, C. (1999). *Developing Teachers: The Challenges of Lifelong Learning*. London: Falmer Press.
- Dewey, J. (1904). The relation of theory to practice in education. In McMurry, C. A. (ed.) *The Relation between Theory and Practice in the Education of Teachers: Third Yearbook of the National Society for the Scientific Study of Education*, pp 9–30. Chicago: University of Chicago Press.
- Dewey, J. (1938). *Experience and Education*. New York: Macmillan.
- Elliot, J. (1991). *Action Research for Educational Change*. Buckingham: Open University Press.
- Eraut, M. (1995). Schön shock: A case for reframing reflection-in-action? *Teachers and Teaching: Theory and Practice* 1(1), 9–22.
- Fenstermacher, G. D. (1994). The knower and the known: The nature of knowledge in research on teaching. *Review of Research in Education* 20, 3–56.
- Franke, A. and Dahlgren, L. O. (1996). Conceptions of mentoring: An empirical study of conceptions of mentoring during the school-based teacher education. *Teaching and Teacher Education* 12(6), 627–641.
- Furlong, J., Whitty, G., Whiting, C., et al. (1996). Re-defining partnership: Revolution or reform in initial teacher education? *Journal of Education for Teaching* 22(1), 39–55.
- Hargreaves, A. (1998a). The emotional practice of teaching. *Teaching and Teacher Education* 14(8), 835–854.
- Hargreaves, A. (1998b). The emotions of teaching and educational change. In Hargreaves, A., Lieberman, A., Fullan, M., and Hopkins, D. (eds.) *International Handbook of Educational Change*, pp 558–575. Dordrecht/Boston/London: Kluwer.
- Hinsch, R. (1979). 'Einstellungswandel und Praxisschock bei jungen Lehrern: Eine empirische Längsschnittuntersuchung' (*Attitudinal Change and Transition Shock in Beginning Teachers: An Empirical Longitudinal Study*). Weinheim: Beltz.
- Hoban, G. F. (2002). *Teacher Learning for Educational Change*. Buckingham/Philadelphia: Open University Press.
- Hoban, G. F. (2005). *The Missing Links in Teacher Education Design: Developing a Multi-Linked Conceptual Framework*. Dordrecht: Springer.
- Hoyle, E. (1980). Professionalization and deprofessionalization in education. In Hoyle, E. and Megarry, J. (eds.) *World Yearbook of Education 1980: Professional Development of Teachers*, pp 42–56. London: Kogan Page.
- Hoyle, E. and John, P. D. (1995). *Professional Knowledge and Professional Practice*. London: Cassell.
- Huibregtse, I., Korthagen, F., and Wubbels, T. (1994). Physics teachers' conceptions of learning, teaching and professional development. *International Journal of Science Education* 16(5), 539–561.
- Husu, J. (2002). *Representing the Practice of Teachers' Pedagogical Knowing*. Turku, Finland: Finnish Educational Research Association.
- Imig, D. G. and Switzer, T. J. (1996). Changing teacher education programs. In Sikula, J. (ed.) *Handbook of Research on Teacher Education*, 2nd edn, pp 213–226. New York: Macmillan.
- Joram, E. and Gabriele, A. (1998). Preservice teacher's prior beliefs: Transforming obstacles into opportunities. *Teaching and Teacher Education* 14(2), 175–191.
- Katz, L., Rath, J., Mohanty, C., Kurachi, A., and Irving, J. (1981). Follow-up studies: Are they worth the trouble? *Journal of Teacher Education* 32(2), 18–24.
- Kelchtermans, G. (1993). Getting the story and understanding the lives: From career stories to teachers' professional development. *Teaching and Teacher Education* 9, 443–456.

- Kessels, J. P. A. M. and Korthagen, F. A. (1996). The relationship between theory and practice: Back to the classics. *Educational Researcher* 25(3), 17–22.
- Korthagen, F. A. J. (2004). In search of the essence of a good teacher: Towards a more holistic approach in teacher education. *Teaching and Teacher Education* 20(1), 77–97.
- Korthagen, F. A. J., Kessels, J., Koster, B., Lagerwerf, B., and Wubbels, T. (2001). *Linking Practice and Theory: The Pedagogy of Realistic Teacher Education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Lanier, J. and Little, J. W. (1986). Research in teacher education. In Wittrock, M. C. (ed.) *Handbook of Research on Teaching*, 3rd edn., pp 527–560. New York: Macmillan.
- Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues* 2(4), 34–46.
- Lortie, S. (1975). *Schoolteacher: A Sociological Study*. Chicago: University of Chicago Press.
- Loughran, J. (2006). *Developing a Pedagogy of Teacher Education*. London: Routledge.
- Loughran, J. J., Hamilton, M. L., LaBoskey, V. K., and Russell, T. (2004). *International Handbook of Self-Study of Teaching and Teacher Education Practices*. Dordrecht: Kluwer.
- McCullough, R. C. (1987). Professional development. In Craig, R. L. (ed.) *Training and Development Handbook*, pp 35–65. New York: McGraw Hill.
- McIntyre, D. and Hagger, H. (1992). Professional development through the Oxford Internship Model. *British Journal of Educational Studies* 40(3), 264–283.
- McKernan, J. (1996). *Curriculum Action Research: A Handbook of Methods and Resources for the Reflective Practitioner*. London: Kogan Page.
- Müller-Fohrbrodt, G., Cloetta, B., and Dann, H. D. (1978). 'Der Praxischock bei jungen Lehrern' (*The Transition Shock in Beginning Teachers*). Stuttgart: Klett.
- Nias, J. (1996). Thinking about feeling: The emotions in teaching. *Cambridge Journal of Education* 26(3), 293–306.
- Robinson, V. M. J. (1998). Methodology and the research-practice gap. *Educational Researcher* 27(1), 17–26.
- Sandlin, R. A., Young, B. L., and Karge, B. D. (1992). Regularly and alternatively credentialed beginning teachers: Comparison and contrast of their development. *Action in Teacher Education* 14(4), 16–23.
- Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.
- Schön, D. A. (1987). *Educating the Reflective Practitioner*. San Francisco, CA: Jossey-Bass.
- Showers, B. and Joyce, B. (1996). The evolution of peer coaching. *Educational Leadership* 53(6), 12–16.
- Shulman, L. S. (1998). Theory, practice, and the education of professionals. *Elementary School Journal* 98(5), 511–526.
- Sprinthall, N. A., Reiman, A. J., and Thies-Sprinthall, L. (1996). Teacher professional development. In Sikula, J. (ed.) *Handbook of Research on Teacher Education*, 2nd edn., pp 666–703. New York: Macmillan.
- Stenhouse, L. (1975). *An Introduction to Curriculum Research and Development*. London: Heinemann.
- Stofflett, R. and Stoddart, T. (1994). The ability to understand and use conceptual change pedagogy as a function of prior content learning experience. *Journal of Research in Science Teaching* 31(1), 31–51.
- Tom, A. (1997). *Redesigning Teacher Education*. Albany, NY: State University of New York.
- Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research* 54(2), 143–178.
- Wenger, E. (1999). *Communities of Practice: Learning, Meaning and Identity*. Cambridge: Cambridge University Press.
- Wideen, M., Mayer-Smith, J., and Moon, B. (1998). A critical analysis of the research on learning to teach: Making the case for an ecological perspective on inquiry. *Review of Educational Research* 68, 130–178.
- Wubbels, T. (1992). Taking account of student teachers' preconceptions. *Teaching and Teacher Education* 8(2), 137–149.
- Zanting, A., Verloop, N., and Vermunt, J. D. (2001). Student teachers eliciting mentors' practical knowledge and comparing it to their own beliefs. *Teaching and Teacher Education* 17, 725–740.
- Zeichner, K. (1999). The new scholarship in teacher education. *Educational Researcher* 28(9), 4–15.
- Zeichner, K. (2007). Accumulating knowledge across self-studies in teacher education. *Journal of Teacher Education* 58(1), 36–46.
- Zeichner, K. M. and Gore, J. M. (1990). Teacher socialization. In Houston, W. R. (ed.) *Handbook of Research on Teacher Education*, pp 329–348. New York: Macmillan.
- Zeichner, K. and Tabachnik, B. R. (1981). Are the effects of university teacher education washed out by school experiences? *Journal of Teacher Education* 32, 7–11.

Further Reading

- Dann, H. D., Cloetta, B., Müller-Fohrbrodt, G., and Helmreich, R. (1978). *Umweltbedingungen innovativer Kompetenz*. [Contextual Conditions of Innovation Competence] Stuttgart: Klett-Cotta.
- Day, C., Sammons, P., Stobart, G., Kington, A., and Gu, Q. (2007). *Teachers matter: Connecting work, lives and effectiveness*. Maidenhead, UK: Open University Press.
- Marton, F. and Booth, S. (1997). *Learning and Awareness*. Mahwah, NJ: Lawrence Erlbaum.

TEACHER EDUCATION – TEACHER EDUCATORS

Contents

Characteristics, Scholarship and Research of Teacher Educators

Professional Development of Teacher Educators

Self-Study by Teacher Educators

Characteristics, Scholarship and Research of Teacher Educators

M Lunenberg, VU University, Amsterdam, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Characteristics of Teacher Educators

Teacher educators have diverse backgrounds and working circumstances.

Many teacher educators have been primary or secondary school teachers before becoming teacher educators. They have a bachelor's or master's degree in education or in the subject they teach. There are, however, also subject specialists and specialists in social studies with a master's or PhD degree, without previous teaching experiences, who become teacher educators.

Teacher educators work in schools (school-based teacher educators), colleges and universities. For some it is their primary occupation, while others combine teaching at schools with being a teacher educator, or are professors in a subject area who hold an additional job in teacher education.

The scope of tasks and responsibilities of teacher educators is broad. Teacher educators teach a variety of subjects. They are responsible for the education of future teachers at both the undergraduate and the graduate level. However, that is not their only responsibility. Those who teach the teachers are the linchpins in educational reforms of all kinds, which include taking diversity into account, ensuring that students pass high-stakes tests, developing the documentation required by professional accrediting agencies, complying with stringent new regulations, building genuine partnerships with schools or universities, parents, and other stakeholders, and developing curricula (see Cochran-Smith (2003)). In other words, new developments in education almost always affect teacher education and teacher educators. Beyond that, teacher educators are expected to set a good example for future teachers with regard to implementing these new developments in practice.

Moreover, the shortage of teachers in many countries (Moon, 2007) has led to a variety of teacher education programs that offer shorter and/or more intensive routes

to a diploma. In combination with the growing amount of public attention paid to the quality of teaching and teachers this affects the work of teacher educators and increases the pressure put on them.

In spite of the considerable amount of attention that has been paid to the preparation of teachers in the last decade, scant attention has been directed to the preparation of teacher educators (Cochran-Smith, 2003). Mueller (2003) states that one is simply stamped teacher educator upon accepting a job with that title. The development of expertise and an identity as a teacher educator is a process most teacher educators go through alone (Lunenberg and Hamilton, 2008). There is hardly any educational support for the transition to what is, for almost all novice teacher educators, a second profession. Whether formal education for teacher educators is the solution is the question. Cochran-Smith (2003) states that although a curriculum for becoming a teacher educator is conceivable, it is not likely that it would constitute the major entry point or means of role definition for a majority of teacher educators. Thus, she continues, we need to think of the education of teacher educators as a continuous process, one that is forever. Working together in communities, teacher educators should explore and reconsider assumptions, try to understand the values and practices of families and cultures that are different from their own, and construct a pedagogy that take these differences into account in locally appropriate and culturally sensitive ways. Smith's contribution to this encyclopedia describes the state of the professional development of teacher educators.

With regard to the content of the professional development of teacher educators, Murray (2005) concludes that the priorities for teacher educator induction are: acquiring the pedagogical knowledge and experience appropriate to teaching teachers, enhancing the scholarship teachers acquire through school teaching and developing it through published research outcomes, and becoming research-active. As far as acquiring the pedagogical knowledge and

experience for teaching teachers is concerned, Murray's research shows that most former teachers have to discover in practice what it means to teach adults who are preparing to become teachers. In the following, we focus on the other priorities of becoming a teacher educator: enhancing scholarship and becoming research-active.

Scholarship of Teacher Educators

Traditionally, teacher educators' research has had a low status within the educational research community, a reflection of the low status of teacher educators in colleges and universities around the world. This is due in part to the diversity of teacher educators, their knowledge of and skills for research, and their level of involvement in teacher education. Most teachers do not carry out research as part of their job. Therefore, their research knowledge and skills are often either rusty or at the level of an undergraduate student. A substantial number of them become teacher educators without research being part of their job description. Professors of teacher education often situate their teaching in teacher education and their research in another (subject) field, not in teacher education. Knowles and Cole (1995) described a vivid example, explaining how they almost developed split personalities. In their teacher education practices, they worked with views of knowledge as being context-bound, personal, and dynamic. In their research work, however, they had to show their faith in the traditional means of knowledge growth, which almost always implied a certain degree of distance between the author and the object of research, a distance that could be well accounted for with the aid of the then-dominant research methodology. Beyond this, Cochran-Smith (2003) warns that many teacher educators (like other teachers) simply do not have sufficient time to act as researchers in their own practice because they are often part-time, temporary, or school-based personnel, and so on.

This situation, however, is changing. Katz and Coleman (2005) found that there is a global trend toward academization in teacher education. More often, teacher educators are expected to be both excellent teachers of teachers and respected researchers. As a consequence, more and more novice teacher educators are expected to conduct and publish research to get tenure.

The growing emphasis on teacher educators' research is in line with the growing emphasis on having students and student teachers perform research in schools and in teacher education institutions, respectively. Thus, teacher educators must not only become experts in conducting research, but are also expected to guide their student teachers' research and to teach their student teachers how, in turn, they can guide research by their pupils. This is no small challenge, as the support and coaching the novice teacher

educators receive as they strive to become researchers is, generally speaking, still minimal to nonexistent.

Nevertheless, there is a significant growth in teacher educators' research. The birth of self-study research at the beginning of the 1990s was an enormous stimulant. Zeichner (1999) states that it was probably the single most significant development ever in the field of teacher education research. The self-studies of teacher educators' own practices offer a way to address the contradictions between teaching and research that professors experience (Loughran *et al.*, 2004). As previously mentioned, traditionally the teacher educator's main task was teaching and if research was involved, the focus was almost always outside teacher education. Without question, the idea that teacher educators' own practices could be the object of research has been an important factor in promoting self-study research. Self-study research carried out by teacher educators focuses on their own teaching and academic practices.

Zeichner (1999) emphasizes that self-study appears to be a productive way for teacher educators to connect the academic task of conducting research with their own professional development and the development of the profession as a whole. In this way, self-study research contributes to further development of the scholarship of teacher educators and fits into the trend of the academization of teacher education. Self-study research also offers novice teacher educators the opportunity to start with small-scale research, focused on questions and problems they encounter in their new teaching education practice. A detailed information about self-study research is discussed elsewhere in this encyclopedia.

Research by Teacher Educators

The emergence of teacher educators' research is embedded in three trends, which in turn are connected with the shortage of teachers and the pressure to improve teacher quality. The first trend is academization, as described above.

The second trend concerns changes in ideas about teaching and teacher education, such as the paradigm shift to the constructivist approach and the increasing school-based education of teachers. The constructivist approach gave rise to the view that knowledge cannot simply be transmitted, but that people construct their own knowledge on the basis of their experiences and that novice teachers, therefore, should be treated as active learners who construct their own understandings. Consequently, it means that the role of the teacher educator had to change congruently with the role of the teacher. In practice this change gave rise to questions and problems to be studied. An interesting example is given by MacKinnon and Scarff-Seater (1997) in their self-study.

The authors agreed that teacher education students needed opportunities for testing, discussing, and comparing various perspectives on teaching. However, in studying their practice they discovered that this point of departure could lead to misunderstandings about theory among students. For example, one of their students wrote that constructivism had taught him that he need not require any knowledge in science in order to teach it, because he could simply allow his students to figure things out for themselves. Based on this study of their own practice, MacKinnon and Scarff-Seater concluded that attention must be paid to the manner in which students learn about constructivism and the way a constructivist perspective is used in teacher education programs.

Because of the shortage of teachers, among other reasons, the amount of learning time student teachers spend at school than at teacher education institutes is increasing. The increasing school-based nature of teacher education has also evoked new research questions; for example, the division of tasks between teacher educators in teacher education institutes and teacher educators who are school-based, and the consequences of this division of tasks for the relationship between theory and practice.

The third trend is the political pressure for standardization in teaching and teacher education. Moon (2007) even states that more policy attention has been given to teacher education in the 1990s than in the 100 years that preceded it. Policymakers want research that provides clear and final answers (Cochran-Smith and Zeichner, 2005). Teacher educators' research counterbalances this trend. Teacher educators realize that a main characteristic of many of their studies is that the problem under investigation develops, shifts, and changes in response to the continual shifts in education. Berry (2004) concludes that teacher educators' efforts to tackle problems of practice rarely result in tidy solutions. Knowledge that is developed through investigations of practice reflects the indeterminate swampy zone of practice described by Schon (1983). In line with Bullough and Pinnegar (2003) one can state that the aim of teacher educators' research is "to provoke, challenge and illuminate rather than confirm and settle" (p. 20).

The currently available research by teacher educators is rooted mainly in North America, Australia/New Zealand, and Western Europe. Israel too has a sound history of research by teacher educators. In Asian and Latin American countries there is a growing recognition for teacher education; however, academic studies by teacher educators are still an exception rather than a rule. In Africa there is still a general scarcity of publications on teacher education and of teacher educators (UNESCO, 2005).

Within the broad range of subjects studied by teacher educators as well as by academics, governmental institutes and educational organizations, some common themes on

which teacher educators' research seems to focus can be identified. These are:

- Learning by students and teachers, such as studies on the perceptions, beliefs, and perspectives of learners, their learning strategies and learning styles, and the role of reflection. Collaborative learning and learning in communities are also studied within this theme. In many studies on new manners of assessment, such as the use of portfolios, teacher educators also focus on students' learning. The studies within this theme are often related to the constructivist approach, as mentioned above.
- The relationship between theory and practice, and between schools and teacher education institutes. Studies on theory and practice focus on the question of how to link theory and practice, and on the role of reflection in this process. Studies on professional development schools and on the role of mentors can also be situated within this theme. With the shift toward school-based teacher education this theme has had even more attention from teacher educators than before.
- Subject-related research. Subject studies are a classic research theme for teacher educators, because these studies can be linked relatively easily with research in other departments. In mathematics especially there is a solid, still-growing knowledge base on learning and teaching mathematics. For this growth, Ball *et al.* (2001) give credit primarily to studies on pedagogical content knowledge that have been carried out since the 1990s. Recently there has been an increase in ethnographic and qualitative studies on learning and teaching languages (Hancock, 2001). This increase seems to be connected, at least in part, with internationalization and the migration of people all over the world. Teachers at all levels confront these changes, and consequently teacher educators have to prepare teachers for them.
- Diversity and ethics. Studies on the influence of gender, race, and socioeconomic background on learning and teaching are on the increase especially in Western countries. Differences in gender, race, and socioeconomic background have an impact on the canonical knowledge about multicultural education (among other topics) that exists and is used in teacher education and teaching. They also affect relationships in education and its hierarchy. There is an increase in teacher educators' autobiographical research on the diversity theme. In contrast, studies by teacher educators on moral issues in their own teacher education classes are still scarce; most studies focus on the role of values in primary or secondary education.
- Pedagogy of teacher education and the role of the teacher educator. These studies focus on how and why to use knowledge in a specific context, on being able to adapt and change, that is, to be responsive to the teaching and learning environment. Self-study research by teacher educators has also increased the number of studies on the

relationship between the teacher educator and the novice teacher, both in terms of personal histories and backgrounds and the consequences for learning and teaching. These studies focus on the crucial role of teacher educators in teacher education and its influence on multicultural education, among other areas. Studies in this field are often based on the European *Bildungs* tradition, in which the link between teaching and learning is the focus point (Van Manen, 1999).

- With the increase in research by teacher educators, the attention given to methods and methodology used in teachers' and teacher educators' research has also grown, especially with regard to inquiry, case studies, action research, narrative studies, and self-study research. One can conclude that the main focus of these studies is on qualitative methods.

Since the past decade, information and communication technology seems no longer to be a theme *an sich* in teacher educators' research. Computers have become a tool, used for example in collaboration, and are studied in this context. In the same sense, the use of video has become a part of studies on pedagogical approaches.

More information about the findings of studies on the fields mentioned here can be found in the diverse entries on teacher education.

Challenges for the Future

As described above, western countries have the lead in teacher educators' research. An overview of teacher educators' research, however, is not available. A systematic overview of research in teacher education is available only in North America. The American Educational Research Association (AERA) panel for studying teacher education, which wrote the overview, explored the contents and boundaries of research on teacher education. In its official summary it stated that education and teacher education pose many kinds of questions, including some that are grounded in moral, ethical, social, philosophical, and ideological concerns. The panel continued by saying that, although these questions can be shaped and understood more fully on the basis of evidence, they cannot be settled by empirical evidence alone (Cochran-Smith and Zeichner, 2005). Therefore, research on teacher education and teacher educators' research is, or should be, aimed at gathering empirical evidence, not only to increase the knowledge base for teaching and teacher education, but also to stimulate the growth of practical wisdom by novice teachers and by teacher educators themselves.

Research on teacher education and especially research by teacher educators is a relatively young field. Most studies conducted by teacher educators are small-scale and predominantly based on qualitative data, such as

narrative data from course assignments, field notes, and transcriptions of classroom observations (Cochran-Smith and Zeichner, 2005). As a result, the growing number of available studies is often not looking at the same thing, or when they are, different methodologies are used. More coherent and longitudinal research on all above-mentioned themes is needed.

There are exceptions, however: the longitudinal study by Furlong *et al.* (2006), for example, on the development of teacher education in the United Kingdom from the early 1990s onward. In the UK, teacher education institutes within higher education are obliged to cooperate with schools, unlike elsewhere in the world. Schools offer teacher education programs without involving teacher education institutes. Furlong and his colleagues concluded that after a period in which schools and teacher education institutes looked for creative ways to work together, national policies and the pressure to solve the shortage of teachers led to a technical, rationalistic approach that flattens complexity and suppresses the epistemological and pedagogical dimensions of teacher education.

Another important finding by the AERA panel is that, until recently, the three layers involved in teacher education (teaching of teacher educators, student teachers learning to teach, and learning by the students of novice teachers) have hardly been studied in connection with each other. However, the first initiatives to do so have now been taken by teacher education institutes in the US. The Teacher Pathways Project, carried out by Stanford University and the University of Albany, is a multiyear study of teachers and teacher preparation programs that follows New York City teachers through their teacher education and careers. It studies – Who they are?, Whether or not they stay in teaching?, and What factors affect the learning of the students they teach? (Boyd *et al.*, 2007). This study is a multidisciplinary project in which well-known teacher educators as well as other specialists have participated. The project explicitly states that one of its important goals is to evaluate current education politics and to provide education policymakers with current, useful data to inform their policy decisions.

The Teachers for a New Era Project is another example of a collaborative study on the three layers of teacher educator impact. It is carried out by 11 partners in the US. The project's starting point is that the quality of the teacher has a profound influence on pupil learning and might even be considered the single-most important factor in student achievement in schools. The focus is on the development of evidence-based programs; the evidence includes attention to gains in pupil learning accomplished under the tutelage of teachers who are graduates of the programs conducted by the participating projects. Like the Pathways Project, this project also aims to have political influence.

An important aspect of both these projects is the attempt to carry out large-scale, multimethod studies without losing

sight of the shifts and changes in the daily practice of education, and by avoiding unproductive standardization.

On a smaller scale, comparable initiatives have been undertaken elsewhere. For example, the program council for educational research of the Netherlands Organization for Scientific Research (NWO/PROO) provides grants for top-level research. It has made one of its main objectives in which the learning and teaching of teaching, and learning by students will be integrated. This offers teacher educators the opportunity to apply for research grants that could make it possible to include their own practice in large-scale studies.

In summary, the development of coherent research programs by teacher educators on diverse themes is the most important challenge for future teacher educators' research. Small-scale studies can progress if theoretical backgrounds and methodological approaches are shared from the beginning of these studies. In addition, larger and more experimental studies are also needed. As Zeichner (2007) states, in order to meet the challenge of developing coherent research programs, collaboration between teacher educators from different institutes and from different countries is necessary. Fortunately, there is a growing international community of teacher educators, one that is building a tradition of working together.

See also: Professional Development of Teacher Educators; Self-Study by Teacher Educators.

Bibliography

- Berry, A. (2004). Confidence and uncertainty in teaching about teaching. *Australian Journal of Teacher Education* 48(2), 149–166.
- Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J. (2007). Examining teacher preparation: Does the Pathway make a difference? <http://www.teacherspolicyresearch.org/TeacherPathwaysProject>.
- Bullough, R. and Pinnegar, S. (2003). Guidelines for quality in autobiographical forms of self-study research. *Educational Researcher* 30(3), 13–21.
- Cochran-Smith, M. (2003). Learning and unlearning: The education of teacher educators. *Teaching and Teacher Education* 19(1), 5–28.
- Cochran-Smith, M. and Zeichner, K. M. (eds.) (2005). *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education*. Mahwah: Lawrence Erlbaum Associates Inc.
- Furlong, J., Campbell, A., Howson, J., Lewis, S., and McNamara, O. (2006). Partnership in English initial teacher education: Changing times, changing definitions: Evidence from the Teacher Training Agency National Partnership Project, Special Edition (Teacher Education). *Scottish Educational Review* 37, 32–45.
- Hancock, C. (2001). The teaching of second languages research trends. In Richardson, V. (ed.) *Handbook of Research on Teaching*, pp. 358–369. Washington, DC: AERA.
- Katz, E. and Coleman, M. (2005). Autonomy and accountability of teacher-educator researchers at a college of education in Israel. *Innovations in Education and Teaching International* 42(1), 5–13.
- Knowles, J. G. and Cole, A. L. (1995). Teacher educators reflecting on writing in practice. In Russell, T. and Korthagen, F. (eds.) *Teachers Who Teach Teachers*, pp. 71–94. London: Palmer Press.
- Ball, L., Lubienski, T., and Mewborn, S. (2001). Mathematics. In Richardson, V. (ed.) *Handbook of Research on Teaching*, pp. 433–456. Washington DC: AERA.
- Loughran, J. J., Hamilton, M. L., Laboskey, V. K., and Russel, T. (2004). *International Handbook of Self-study of Teaching and Teacher Education Practice*. Dordrecht: Kluwer Academic Publishers.
- Lunenberg, M. and Hamilton, M.-L. (2008). Who and where are we as teacher educators? Finding our identities as teacher educators and threading a golden chain. *Teacher Education Quarterly* 35(1), 185–205.
- Mackinnon, A. and Scarff-Seater, C. (1997). Constructivism: Contradictions and confusions in teacher education. In Richardson, V. (ed.) *Constructivist Teacher Education*, pp. 38–56. London: Falmer Press.
- Moon, B. (2007). Research Analysis: Attracting, Developing and Retaining Teachers: A Global Overview of Current Policies and practices. UNESCO Working Paper. http://portal.unesco.org/education/en/ev.php-URL_ID=40218&URL_DO=DO_TOPIC&URL_SECTION=201.html (accessed 24 December 2008).
- Mueller, A. (2003). Looking back and looking forward: Always becoming a teacher educator through self-study. *Reflective Practice* 4(1), 67–84.
- Murray, J. (2005). New teacher educators and induction into higher education. *European Journal of Teacher Education* 28(1), 67–85.
- Schon, D. (1983). *The Reflective Practitioner. How Professionals Think in Action*. London: Temple Smith.
- UNESCO (2005). Capacity-building of teacher training institutions in sub-Saharan Africa. UNESCO. http://portal.unesco.org/education/en/ev.php-URL_ID=40218&URL_DO=DO_TOPIC&URL_SECTION=201.html (accessed 24 December 2008).
- Van Manen (1999). The language of pedagogy and primacy of student experience. In Loughran, J. (ed.) *Researching Teaching: Methodologies and Practices for Understanding Pedagogy*, pp. 13–27. London: Palmer Press.
- Zeichner, K. (1999). The new scholarship in teacher education. *Educational Researcher* 28(9), 4–15.
- Zeichner, K. (2007). Accumulating knowledge across self-studies in teacher education. *Journal of Teacher Education* 58(1), 36–46.
- Brown, E. (2004). The significance of race and social class for self-study and the professional knowledge base of teacher education. In Loughran, J. J., Hamilton, M. L., Laboskey, V. K., and Russell, T. (eds.) *International Handbook of Self-study of Teaching and Teacher Education Practice*, pp. 517–574. Dordrecht: Kluwer Academic Publishers.
- Korthagen, F. A. J., Kessels, J., Koster, B., Lagerwerf, B., and Wubbels, T. (2001). *Linking Practice and Theory: The Pedagogy of Realistic Teacher Education*. Mahwah: Lawrence Erlbaum Associates.
- Korthagen, F. A. J., Loughran, J., and Lunenberg, M. (eds.) (2005). *Special Issue: Teaching teachers: Studies into the Expertise of Teacher Educators. Teaching and Teacher Education* 21(2).
- Loughran, J. (2006). *Developing a Pedagogy of Teacher Education*. New York: Routledge.
- Lunenberg, M., Ponte, P., and van der Ven, P.-H. (2007). Why shouldn't teachers and teacher educators conduct research on their own practices? *European Educational Research Journal* 6(1), 13–24.
- Munby, H., Russell, T., and Martin, A. (2001). Teachers' knowledge and how it develops. In Richardson, V. (ed.) *Handbook of Research on Teaching*, pp. 877–904. Washington, DC: AERA.
- Wilson, S. M. (2006). Finding a canon and core: Meditations on the preparation of teacher educator-researchers. *Journal of Teacher Education* 57(3), 315–325.

Further Reading

Relevant Website

- <http://www.teachersforanewera.org> – Teachers for a New Era Project, About teachers for a new era.
- <http://www.teacherspolicyresearch.org/TeacherPathwaysProject/tabid81/Default.aspx>.

Professional Development of Teacher Educators

P K Smith, University of Bergen, Bergen, Norway

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The purpose of this article is to provide an overview of issues related to professional development of teacher educators. Professional development, as defined in this article, is an internal process in which professionals engage within a formal or informal framework. The process is rooted in critical self-analysis of professional practice, and steps taken to improve current practice are called professional development activities.

Professional development occurs when beliefs and assumptions about the profession change, and as a result, professional practice changes. The latter is related to professional decisions taken prior to action.

This model of action (Figure 1), which is contextually dependent, presents the idea that professional decisions take place not only based on theoretical and practical knowledge, but also based on analysis of the immediate context within a space of professional autonomy, a core value of higher education (Altbach and Lewis, 1995). The volume of space varies with the context, with the personality of the teacher, stage in career, and most of all, with teacher educators' professional confidence. Professional growth takes place when autonomous actions within the autonomous space are improved as a product of learning processes, theoretical as well as experiential.

Koster *et al.* (1996) made the distinction between school- and institution-based teacher educators. School-based teacher educators are experienced teachers who accept students of teaching into their classes during the practicum, whereas institution-based teacher educators are teacher educators teaching in higher education institutions. This article is limited to discussing professional development of institution-based teacher educators.

The first part of the article deals with purposes for teacher educators' professional development activities, the context in which activities take place, including level of formality, if engagement is on an individual basis or within communities of practice and learning. Next, examples of professional development activities are briefly discussed in light of expected constraints.

Purposes of Professional Development

Improving Teacher Education

From the profession's point of view, the main purpose of teacher educators' development is to improve teacher

education at a local and at a national and international level. Loughran (2006) advocates the urgency of developing a pedagogy of teacher education, a shared understanding of teaching teachers from a situational and broader perspective. Creating a pedagogy of teacher education depends on teacher educators being "conscious not only of what they are teaching, but also the manner in which teaching is conducted" (p. 11). Critical reflection is a prerequisite for professional development; however, it is not sufficient if it is carried out in isolation. Professional knowledge needs to be articulated to make it accessible to students of teaching and to other teacher educators. In this manner, a pedagogy of teacher education is developed.

The urgency of relevant and high-quality teacher education is highlighted in the OECD (2005) report titled 'Teachers matter'. The report informs about an aging teaching force in many countries, and teacher education needs to attract competent candidates who, as qualified teachers, will know how to meet the multiple challenges faced in schools undergoing constant change. Teacher educators not only need to stay updated, but they are also expected to be involved in forming educational change. Current and future school teachers take on different roles than the teaching roles they were exposed to as pupils. Currently, they are asked to support pupils in becoming independent and self-directed lifelong learners (OECD, 2005). Students of teaching need to personally experience self-directed learning during teacher education, and teacher educators serve as models (Lunenberg, *et al.*, 2007) for ongoing learning and professional development for students of teaching.

Meeting External Standards

External standards for teacher educators are being developed in various countries, often as a reaction to the wide criticism of teacher education (Apple, 2001; Berliner, 2000; Turley and Nakai, 2000; Nieme, 2002; Koster and Korthagen, 2001). Engaging in and documentation of professional development are among the several standards for teacher educators. The American Association for Teacher Educators (ATE) has developed a set of seven standards, and standard 3 reads as follows: "Master teacher educators inquire into, and reflect on, their own practice and demonstrate commitment to lifelong professional development."

In the list of indicators which elaborate on this standard, teacher educators are required to systematically

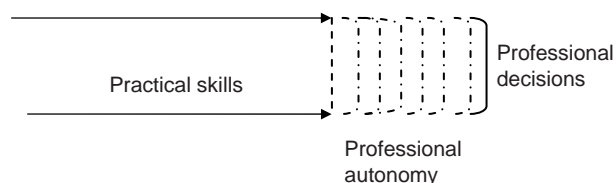


Figure 1 Professional decision making. From Smith, K. (2007). Empowering school and university-based teacher educators in their role as assessors of student teachers' teaching practice – a school–university cooperation. *Educational Research and Evaluation* 13(3), 279–293.

plan their professional development, stay updated in terms of theory and practical experiences, and document the process of professional growth for self- and external assessment purposes.

Five competencies are listed as professional standards for Dutch teacher educators, of which the fifth is called “development and personal growth competence” (Koster and Dengerink, 2001, p. 350). As with the American example, Dutch teacher educators are also required to “make their own learning process visible to colleagues and students” (Koster and Dengerink, 2001, p. 350). Smith (2005), in her study of how novice teachers and teacher educators define a “good teacher educator,” found that novice teachers value teacher educators practicing a meta-cognitive approach to teaching by explaining why and how they teach as they do, and teacher educators themselves were found to highly value self-awareness and engagement in professional development as characteristics of “the good teacher educator” (p. 184).

Internal Drive to Learn and Improve

Koster *et al.* (2005) examined what teacher educators themselves thought should be included in the professional profile for teacher educators in various teacher-education settings in the Netherlands. Out of six task areas that teacher educators are expected to master, three are considered necessary for all, the first one reads: “the teacher educator working on his/her own development and that of colleagues (professionalism and well-being)” (p. 165). In terms of competencies, teacher educators are expected to have a “very necessary” competence which was found to be the “communicative and reflective competence” which was broken down into subcompetences; one of which is worded: “being able to evaluate one’s own teaching and make changes accordingly” (Koster *et al.*, 2005, p. 167).

Smith (2003) argues that engaging in professional-development activities and updating professional knowledge is essential not only to ensure the quality of teacher education, but also to maintain enthusiasm and interest in the profession. She refers to the work of Day (1999) who found that teachers, who are professionally challenged, are less likely to leave the profession as result of professional

fatigue, and they seem to be happier in their jobs. Smith claims that there is little reason to believe that teacher educators feel differently.

Strengthening the Professional Status within Higher Education

In the previous section, personal intrinsic motivation for professional development has been briefly discussed; however, there is, to a large extent, also personal extrinsic motivation for professional development of teacher educators.

Documentation of engagement in professional-development activities is frequently required by institutions for tenure, promotion, and to gain a stronger professional status. Nixon (2001) points at a global trend toward academization of teacher education of which formal professional learning outcomes are required in the form of advanced academic degrees and publication of research. In Israel, new teacher education faculty are required to hold a PhD to be given tenure, and experienced teacher educators are requested to obtain a PhD or to go for early retirement (Katz and Coleman, 2005). In the current trend of standardization and academization, there is external pressure on teacher educators who want to strengthen their academic position within the institution by becoming visible researchers, program designers, and holders of PhD degrees to be seen as being actively involved with professional-development activities. The higher the academic rank aspired for, the heavier the focus on research and publication, which seems to lessen the importance of excellence in teaching (Katz and Coleman, 2002; Smith, 2005).

Novice Teacher Educators

Recent research documents the frustrating transition the novice teacher educators go through in the process of shifting from being expert teachers to being novice teacher educators and inexperienced researchers in a context which values research more than teaching (Murray and Male, 2005; Zeichner, 2005; Dinkelman, *et al.*, 2006; Ritter, 2007). Most novice teacher educators lack formal training as teacher educators, they are chosen for the position because they are successful school teachers and school-based teacher educators. The shift from being what Murray and Male (2005) calls first-order practitioners in school to second-order practitioner in an academic institution is often done by accepting the position as teacher educators (Dinkelman *et al.*, 2006) without any special preparation (Zeichner, 2005). The new academic context causes a transformation of professional identity, from being a teacher relating to children to being a teacher educator relating to adults and becoming involved with unfamiliar academic activities (Dinkelman *et al.*, 2006; Ritter, 2007), research

being perceived as most demanding (Murray and Male, 2005; Dinkelman *et al.*, 2006; Zeichner, 2005).

Zeichner (2005) discusses four main professional activities that novice teacher educators are recommended to be involved with: self-study of own practice, active involvement in developing curricula and courses, learning about new school reforms, and constantly keeping updated with professional literature. Engaging in research and publication is a new requirement which causes anxiety and stress to many (Murray and Male, 2005). Systematic self-study and action research are modes of research which support new teacher educators in the unavoidable professional development processes (Zeichner, 2005).

Contexts for Professional Development

Professional development of teacher educators takes place in different contexts, which also dictates the level of formality. Eraut (1994) introduced the terms off-job and on-job learning in professional development processes.

Off-Job Learning

Off-job learning is formal learning which takes place in contexts other than the regular workplace, for example, enrolment in formal studies for advanced degrees.

Another type of off-job professional development activities are formal in-service courses for teacher educators; however, these are, in most settings, rare: "there is often little or no professional development provided for faculty to help them learn how to continually improve their work with novice teachers" (Zeichner, 2005, p. 119).

An exception to this is the Mofet Institute in Israel (Korthagen, 2001; Smith, 2003, 2005), a government-funded independent institution which serves as the center for development activities for the country's teacher educators, novice as well as experienced teacher educators. The Mofet Institute has five main focuses:

- In-service courses of various length and degrees of formality, the longest of which goes over 2 years and certifies teacher educators.
- One-day courses and seminars with well-known professionals in the field, national and international conferences, face-to-face as well as virtual events.
- Research support unit which provides financial support, research guidance, statistical support, and initiates research projects.
- Publication center which is the major publication house for teacher education material in Israel. The center supports authors of books financially and with expert counseling. The only peer-reviewed journal in Hebrew in the field of teacher education is published by Mofet's publishing house.

- Resource and information communication technology (ICT) center which holds a wide range of international literature in paper as well as in electronic form. Many publications are translated into Hebrew.

The Mofet institute provides opportunities for off-job learning which cover a wide range of professional-development activities for teacher educators.

On-Job Learning

On-job learning is learning from practice through reflection, self-study, informal meetings, and networks with colleagues, and in more formal communities of learning. An example of the latter can be found at the University of Bergen, Norway, where all staff members at the Faculty of Psychology are members of research groups, and those involved with teacher education meet regularly to share research ideas, getting feedback on projects with which they are involved, supporting each other when preparing for conference presentations, and reviewing each others' papers before they are sent to professional journals. Teacher educators meet regularly to discuss the university's teacher education program, and recently members of staff have started video-recording their teaching and present the recordings to the staff inviting suggestions for improvement (Smith and Krumsvik, 2007). Time is set off for working and supporting each other research-wise as well as teaching-wise. This is how these teacher educators have interpreted Lave and Wenger's (1991) concept of communities of learning and Wenger's (1998) idea of communities of practice.

A more recent development of on-job learning is work-based learning (WBL). WBL is based on a socioconstructivist view of learning. Learners construct meanings from experiences in a particular context in dialogs with others (Brodie and Irving, 2007). Van Eekelen *et al.* (2005), in their study of teacher learning, revealed that learning in interaction and learning by doing was reported to be the most important factor in professional development. Billett (2001) defines three key factors in workplace learning: (1) engagement in everyday work tasks, (2) direct guidance from co-workers, and (3) indirect guidance provided by the workplace itself and others in the workplace.

Currently, professional learning is not viewed as an isolated individual journey, but an ongoing process which involves other members of the professional community. Personal growth is still perceived as an individual process; it is the personal beliefs, assumptions, and performances that change. The process, however, is embedded in professional learning within a community of practice.

Professional Development Activities

In this section, some common professional activities for teacher educators are discussed.

Formal Studies for Advanced Academic Degrees

A number of countries have initiated steps to improve the quality of teacher education, example of which is the requirement that teacher educators hold advanced academic degrees, at least master degrees, and in many contexts, also PhDs. Teacher educators also undertake formal studies because they want to do a better job as professionals. However, in many settings, formal studies are traditional and not directly related to teacher educators' work. The studies are theoretically based within a specific discipline or in psychology and do not directly link to the work of teacher educators as practitioners (Zeichner, 2005). Alternative PhD programs or EdD programs with a practitioner focus are introduced in the UK (e.g., University of Bath) as well as in the US (e.g., University of Arizona).

Universities need to offer more programs of a similar kind if teacher educators are required to hold advanced academic degrees relevant to the profession. Lunenberg and Willemse (2006) claim that teacher educators enter the profession for two reasons: "1) they have been good teachers or 2) they were experts in some particular area" (p. 82). The first group is large, and the motivation to engage in further studies is enhanced by offering practically oriented study programs.

Short Courses and Sporadic Lectures

A frequent professional development opportunity is attending lectures of well-known experts. Even though invited experts might provide new insights and trigger reflective processes, the impact of such activities in terms of changing beliefs, attitudes, and practice have been found to be meager (Darling-Hammond, 2006). Dochy (2004) carried out a study with graduate students, which looked at what effect various teaching approaches had on the students' long-term memory, and the results are not surprising:

- lectures: 4.5%,
- reading: 11%,
- audiovisual media: 22%,
- demonstration: 32%,
- group discussion: 56%,
- practical: 75%, and
- teaching the subject to others (peers): 82%.

One might assume that the impact of teacher educators does not differ much to that of graduate students.

Participation in Professional and Conferences

Part of conference participation is indeed listening to lectures, thus getting new insights into current trends within the professional field; however, conference participation

offers additional professional-development opportunities. If teacher educators present papers, a major part of professional learning precedes the conference itself. The research, preparing the presentation, and writing the paper is an intense professional-development activity whose product is presented to a wide audience for comments and feedback. The viability of the outcome of the learning process, the personal knowledge, is tested, and the received feedback feeds into the professional development process and takes it forward. Moreover, the opportunity to create professional networks beyond one's own institution is an important professional-development activity in itself.

Research

Research and scholarship of teacher educators is discussed in depth in another article in this volume; it is therefore discussed in brief in the current article. Research is a basic requirement of higher education faculty. It is also viewed as one of the most stressful professional-development activities that teacher educators undertake (Murray and Male, 2005; Yogeve and Yogeve, 2006).

Cochran-Smith (2005) argues that research is part of teacher educators' responsibility, and it complements practical aspects of the job. Being consumers of research is essential to being updated with new development in the field. As producers of research by systematically enquiring into the local context, teacher educators are involved with improving practice and contributing to an emerging pedagogy of teacher education (Cochran-Smith, 2005; Loughran, 2006). There is a close link between teacher educators' research activities and practice to the extent that "the boundaries between research and practice often blur, creating opportunities for reflection on and improvement of the practice of teacher education" (Borko *et al.*, 2007, p. 6). Yogeve and Yogeve (2006) examined differences between the research carried out in Israeli teacher education colleges and that carried out in education departments at the universities. They noticed that there is an emerging distinction of practitioners' research content-wise and that college-based teacher educators conduct more practically oriented research, yet there were no significant differences in terms of research methodology applied. There is a trend "toward action research and other qualitative studies, it nonetheless does not largely depart from the traditional types and methods of research on which university schools of education focus" (Yogeve and Yogeve, 2006, p. 40). The authors explain this by claiming that college teacher educators depend on university supervision, and that there is not yet a solid research foundation and culture within colleges of teacher education.

The positive professional-development opportunities related to research enjoy a wide consensus in the literature on teacher education (Yogeve and Yogeve, 2003;

Zeichner, 2007; Cochran-Smith, 2005; Borko *et al.*, 2007); however, there is also sharp criticism of the quality of research carried out by teacher educators. Cochran-Smith (2005) argues that teacher-education research needs to go beyond local practice and link to general research issues in teacher education (Zeichner, 2007) if the impact of teacher-educator research is to be noticed.

Self-Study

Self-study has become a frequent type of research for teacher educators. "Self-study as a formal approach to research seeks to increase understanding of oneself; teaching; learning; and the development of knowledge about these" (Loughran, 2004, p. 9). It is closely linked to critical self-analysis of professional practice, which is the core of professional development as defined in this article.

Recent publication of self-studies is wide and much of it is published in the journal *Studying Teacher Education* edited by John Loughran and Tom Russell. Self-study has been identified as an individual research journey into personal practice of teacher education today, however, with an increasing pressure of going beyond the self (Cochran-Smith, 2005; Zeichner, 2007; Borko *et al.*, 2007). Loughran, in a recent paper, states that, "the reality is that self-studies are dramatically strengthened by drawing on alternative perspectives and reframing situations, thus data, ideas, and input that necessitate moving beyond oneself" (Loughran, 2007, p. 12). Various egocentric models of reflection were met with similar criticism by supporters of sociocultural learning theories, and the reflective dialog (Kolb, 1998) with a wider audience is today seen as essential for professional development.

Feedback on Teaching

Another dialectical form for professional development essential to teacher educators is feedback on teaching from superiors, colleagues, and students.

Feedback from superiors

Feedback from superiors is usually based on observation followed by a feedback session or an annual appraisal meeting. An additional source of information about the quality of teaching is student responses to formal feedback questionnaires. Feedback meetings with superiors have the potentiality to be forceful triggers for professional development; however, this is not always the case. Poyas and Smith (2007) found that teacher educators are unhappy with the way their work is being assessed. There is a lack of transparency and documentation of how superiors are informed about the quality of the teacher educators' work, a fact to which deans of faculties and heads of departments admitted. If this is the case, there is little substance in superior feedback, and it is not

very useful for professional development purposes, and even less for assessment purposes.

Feedback from colleagues

Peer feedback is found to have a positive impact on learning, and especially research by Filip Dochy and his colleagues on peer feedback in higher education is encouraging (Dochy *et al.*, 1999; Struyven, *et al.*, 2003; among others). The approach of peer feedback as a professional-development activity for teacher educators is less common, yet Smith (2003) strongly advocates it, and it was found to be conducive to professional learning in a self-study of two teacher educators at the University of Bergen reported by Smith and Krumsvik (2007). The main obstacle to collegial feedback is that it is not common and requires a supportive and safe atmosphere within the staff, which removes reluctance to invite colleagues to observe one's teaching and to share critical incidents (Tripp, 1993). Reflective dialogs with peers who share knowledge about the local context provide valuable opportunities for teacher educators' professional development.

Feedback from students

A major problem in evaluating teaching in teacher education is the lack of clarity regarding what tools to use when evaluating quality teaching.

The most common evaluation instrument is student feedback in the form of standardized questionnaires, and whereas such questionnaires collect opinions on certain aspects of teaching, other, not less important, aspects of teacher educators' responsibilities, such as creating knowledge, involvement in educational issues in general, and pursuing professional development, are rarely included. It has been claimed that student questionnaires lack validity due to narrow documentation of teacher educators' responsibilities and work (McKeachie, 1997). Many teacher educators are doubtful as to whether students are capable of evaluating teaching in teacher education and worry that students misuse the trust and the power they have been given (Seldin, 1993).

Thus, formal student feedback based on a standardized questionnaire, according to Smith and Welitzker-Polak (2007), is not found to support teacher educators' professional development. The lack of trust put in formal standardized feedback questionnaires is not only related to the tool itself, but also just as much to the way the collected information is handled. If student feedback is the main source on which managerial decisions are made, the feedback is perceived as stressful and does not encourage professional development. Smith and Welitzker-Polak's study (2007) suggests that teacher educators are not against formal student feedback in itself as long as they are the owners of the information and draw upon it in discussions with colleagues and superiors in a nonthreatening and

supportive atmosphere. Informal student feedback collected by the individual teacher educators has proven to be more conducive to professional growth, mainly because it reduces a commonly held defensive attitude to standardized student feedback (Smith and Welitzker-Polak, 2007).

Portfolios and Professional Development

Portfolio has recently become a widely used tool for documenting professional development (Klenowski *et al.*, 2006; Smith and Tillema, 2006; among others), and there is evidence that it enhances professional development mainly by enabling the professional to take responsibility of and monitor personal professional development, and not leaving it to others. Professional development cannot be forced; the professional has to be open to engage in development activities and to change (Day, 1999). The portfolio is useful in documenting professional growth reflected in learning activities and personal critical analysis (reflection). It is found to be contributing to professional development and be of sustained use when the portfolio is compiled voluntarily and not viewed as compulsory for assessment purposes. However, advantages inherited in voluntary portfolio use are, to a certain extent, reduced by the fact that portfolios are less likely to be maintained when not made compulsory within a formal framework (Smith and Tillema, 2001). Compiling a professional portfolio is time consuming, and time is a major constraint when teacher educators' professional development is concerned.

Constraints

The necessity of professional development of teacher educators is, today, not disputed, yet there are several constraints that reduce the range of development and its impact on teacher educators' professional growth.

Time and Resources

Time availability is, perhaps, the main constraint on professional development. Even though some research reports that new teacher educators feel they have more time for reflection as teacher educators than as school teachers (Dinkelman *et al.*, 2006), other research reveals that a frequent complaint by teacher educators is being short of time (Smith and Tillema, 2006). Korthagen (2000) reports that some American teacher educators work 45 h per week, and the lower the academic rank of the teacher educators, the higher the teaching load becomes, a fact which reduces professional-development opportunities. Professional development time needs to be integrated in the job responsibilities including meeting time for staff development. This is a question of resources, as more

people are needed to share the total workload of teaching, administration, and professional development (research, networking, community learning, etc.).

Support

Lack of support from the immediate workplace environment has also been found to be a constraint to professional development (Smith and Polak-Welitzker, 2007). In a competitive environment, people keep quiet about professional questions and challenges, trying to manage alone for the fear of losing the job, or for missing out on tenure and promotion opportunities. Professional growth lies in critical analysis of successful as well as less successful practice, and it is hampered if the environment is not perceived as supportive and opportunities for reflective dialog are few (Day, 2007).

Confidence and Motivation

Confidence in one's ability to learn, grow, and achieve is closely linked to support from significant others as well as to the motivation to pay the price in terms of time, effort, and frustration inherited in learning (Smith, 2006). There are, as discussed previously in this article, personal and external reasons to engage in professional development, and similarly the motivation is of intrinsic and extrinsic nature. Motivation lies in the personal need to grow and learn; however, external incentives found in a goal-oriented, self-chosen, professional-development plan with clear milestones and increased appreciation by the professional community is likely to affect the process positively. Thus, teacher educators as a profession move from a level of professional expertise to a level of professional scholarship. "To be scholarly is to engage in intellectual inquiry and to bring it to scrutiny for the public arena" (Trigwell, 2007).

Conclusions

The view adopted in this article is that professional development is an integrated part of teacher educators' responsibilities as is made apparent in emerging standards for teacher educators. Reasons for engaging in professional development vary; there are intrinsic, personal drives for improvement and continuous learning, as well as external demands by institutions and by policy- and decision-making bodies.

Professional development causes changes in personal beliefs, assumptions, and performance within an autonomous professional space where decisions are taken. It is the individual person that grows. However, the activities are fertilized by dialogs with the immediate and wider professional community, the underlying theory is socio-cultural in which communities of learning and of practice

enhance the individual as well as the collective learning process and contribute to improved outcomes of learning.

Strategies for professional development reflect the personality of the individual, the context, and external goals. They all require, however, resources such as time, effort, and support by colleagues and management, all of which also form constraints on teacher educators' professional learning. Developing regional and national resourceful professional development centers for teacher educators might be a future way to challenge some of the constraints.

This topic – professional development of teacher educators – is wide and enjoys an increasing attention of researchers and educators. The intention of the current article has been to discuss the most salient characteristics of the topic in brief with a humble acknowledgment that relevant issues and informative research have not been included.

Bibliography

- Altbach, P. G. and Lewis, L. S. (1995). Professional attitudes – an international survey. *Change* 27(6), 50–57.
- Apple, M. W. (2001). Markets, standards, teaching, and teacher education. *Journal of Teacher Education* 52(2), 182–196.
- Berliner, D. C. (2000). A personal response to those who bash teacher education. *Journal of Teacher Education* 51(5), 358–371.
- Billett, S. (2001). Learning through work: Workplace affordances and individual engagement. *Journal of Workplace Learning* 13(5), 209–214.
- Borko, H., Liston, D., and Whitcomb, J. A. (2007). Genres of empirical research in teacher education. *Journal of Teacher Education* 58(1), 3–11.
- Brodie, P. and Irving, K. (2007). Assessment in work-based learning: Investigating a pedagogical approach to enhance student learning. *Assessment and Evaluation in Higher Education* 32(1), 11–19.
- Cochran-Smith, M. (2005). Teacher educators as researchers: Multiple perspectives. *Teaching and Teacher Education* 21, 219–225.
- Darling-Hammond, L. (2006). *Powerful Teacher Education*. San Francisco, CA: Jossey-Bass.
- Day, C. (1999). *Developing Teachers: The Challenges of Lifelong Learning*. London: Falmer Press.
- Day, C. (2007). Identity, well-being and effectiveness: The emotional contexts of teaching. *Paper Presented at the Bi-Annual ISATT Conference at Brock University, Canada*, July, 5–9.
- Dinkelmann, T., Margolis, J., and Sikkenga, K. (2006). From teacher to teacher educator: Experiences, expectations, and expatriation. *Studying Teacher Education* 2(1), 5–23.
- Dochy, F., Segers, M., and Sluijsmans, D. (1999). The use of self-, and peer-, and co-assessment in higher education: A review. *Studies in Higher Education* 24(3), 310–350.
- Dochy, F. (2004). Assessment engineering: Aligning assessment, learning and instruction. *Keynote Presented at the Second Biannual Northumbria/Earli SIG Assessment Conference, University of Bergen* June, 23–25.
- Eraut, M. (1994). *Developing Professional Knowledge*. London: Falmer Press.
- Katz, E. and Coleman, M. (2005). Autonomy and accountability of teacher-educators at a college of education in Israel. *Innovations in Education and Teaching International* 42(1), 5–13.
- Klenowski, V., Askew, S., and Carnell, E. (2006). Portfolios for learning, assessment and professional development in higher education. *Assessment and Evaluation in Higher Education* 31(3), 267–286.
- Kolb, D. A. (1998). Experiential learning: From discourse model to conversation. Interview. *Lifelong Learning in Europe* 3(3), 148–153.
- Korthagen, F. A. J. (2000). From neglected group to spearhead in the development of education. In Willems, G. M., Stakenborg, J. J. H., and Veugelaers, W. (eds.) *Trends in Dutch Teacher Education*, pp 35–48. Leuven-Apeldoorn: Garant.
- Korthagen, F. A. J. (2001). Teacher education: A problematic enterprise. In Korthagen, F. A. J., Kessels, J., Koster, B., Lagerwerf, B., and Wubbels, T. (eds.) *Linking Practice and Theory – The Pedagogy of Teacher Education*, pp 1–19. Mahwah, NJ: Lawrence Erlbaum Associates.
- Koster, B. and Dengerink, J. (2001). Towards a professional standard for Dutch teacher educators. *European Journal of Teacher Education* 24(3), 343–354.
- Koster, B. and Korthagen, F. (2001). Training teachers for the realistic approach. In Kessels, J., Koster, B., Lagerwerf, B., and Wubbels, T. (eds.) *Linking Practice and Theory – The pedagogy of Teacher Education*, pp 239–254. Mahwah, NJ: Lawrence Erlbaum Associates.
- Koster, B., Brekelmans, M., Korthagen, F., and Wubbels, T. (2005). Quality requirements for teacher educators. *Teaching and Teacher Education* 21, 157–176.
- Koster, B., Korthagen, F. A. J., Wubbels, T., and Hoornweg, J. (1996). Roles, competencies and training of teacher educators, a new challenge. In Befring, E. (ed.) *Teacher Education for Quality*, pp 397–411. ATEE, Oslo: Lobo Grafisk.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Loughran, J. (2004). A history and context of self-study of teaching and teacher education practices. In Loughran, J. J., Hamilton, M. L., LaBoskey, V. B., and Russell, T. (eds.) *International Handbook of Self-study of Teaching and Teacher Education Practices*, pp 1169–1184. Dordrecht: Kluwer.
- Loughran, J. (2006). *Developing a Pedagogy of Teacher Education*. London: Routledge.
- Lunenberg, M. and Willemse, M. (2006). Research and professional development of teacher educators. *European Journal of Teacher Education* 29(1), 81–98.
- Lunenberg, M., Korthagen, F., and Swennen, A. (2007). The teacher educator as a role model. *Teaching and Teacher Education* 23(5), 586–601.
- McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist* 52, 1218–1225.
- Murray, J. and Male, T. (2005). Becoming a teacher educator: Evidence from the field. *Teaching and Teacher Education* 21, 125–142.
- Nieme, H. (2002). Active learning – a cultural change needed in teacher education and schools. *Teaching and Teacher Education* 18(7), 763–780.
- Nixon, J. (2001). Towards a new academic professionalism: A manifesto of hope. *British Journal of Sociology of Education* 22(2), 227–245.
- OECD. (2005). *Teachers matter: Attracting, Developing and Retaining Effective Teachers*. Paris: OECD.
- Poyas, Y. and Smith, K. (2007). Becoming a community of practice, the blurred identity of clinical faculty teacher educators. *Teacher Development* 11(3), 313–334.
- Ritter, J. K. (2007). Forging a pedagogy of teacher education: The challenges of moving from classroom teacher to teacher educator. *Studying Teacher Education* 3(1), 5–22.
- Seldin, P. (1993). The use and abuse of student ratings of professors. *The Chronicle of Higher Education* 46, A40.
- Smith, K. and Krumsvik, R. (2007). Video papers – a means for documenting practitioners' reflections on practical experiences: The story of two teacher educators. *Paper Presented at the Bi-Annual ISATT Conference at Brock University, Canada*, July 5–9, 2007. Research in Comparative and International Education 2(4), (online).
- Smith, K. and Tillema, H. (2006). *Portfolios for Professional Development: A Research Journey*. New York: Nova Science Publishers.
- Smith, K. and Welitzker-Polak, M. (2007). What can they say about my teaching? Teacher Educators' Attitudes to Students Evaluating Their Work. *Paper Presented at the ECER Conference (European Conference on Educational Research)*, University of Ghent September 19–21.

- Smith, K. (2003). So, what about the professional development of teacher educators. *European Journal of Teacher Education* **26**(2), 201–215.
- Smith, K. (2005). Teacher educators' expertise: What do novice teachers and teacher educators say? *Teaching and Teacher Education* **21**, 177–192.
- Smith, K. (ed.) (2006). Motivating students through assessment. In *Bridging the Assessment Gap in English-Medium Higher Education*, pp 109–121. Bocum: AKs-Verlag.
- Smith, K. and Tillema, H. (2001). Long term influences of portfolios on professional development. *Scandinavian Journal of Educational Research* **45**(2), 183–203.
- Struyven, K., Dochy, F., and Janssens, S. (2003). Students' perceptions about new modes in higher education. In Segers, M., Dochy, F., and Cascallar, E. (eds.) *Optimising New Modes of Assessment: In Search of Qualities and Standards*, pp 171–223. Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Trigwell, K. (2007). *Quality teaching: Some insights from research in higher education. Keynote Presented at the Bi-Annual ISATT Conference at Brock University, Canada, July 5–9.*
- Tripp, D. (1993). *Critical Incidents in Teaching. Developing Professional Judgement*. London: Routledge.
- Turley, S. and Nakai, K. (2000). Two routes to certification: What do student teachers think. *Journal of Teacher Education* **51**(2), 122–134.
- Van Eekelen, I. M., Boshuizen, H. P. A., and Vermunt, J. D. (2005). Self regulation in higher education teacher learning. *Higher Education* **50**, 447–471.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. Cambridge, MA: Cambridge University Press.
- Yogev, S. and Yogev, A. (2006). Teacher educators as researchers: A profile of research in Israeli teacher colleges versus university departments of education. *Teaching and Teacher Education* **22**, 32–41.
- Zeichner, K. (2005). Becoming a teacher educator: A personal perspective. *Teaching and Teacher Education* **21**, 117–124.
- Zeichner, K. (2007). Accumulating knowledge across self-studies in teacher education. *Journal of Teacher Education* **58**(1), 36–46.

Further Reading

- Brunstad, P. O. (2007). Military professional actions – more than – knowledge and skills. *PACEM* **10**(2), 59–70.
- Campbell, D. M., Cignetti, P. B., Melenzyer, B. J., Nettles, D. H., and Wyman, R. M. (1997). *How to Develop a Professional Portfolio – A Manual for Teachers*. Needham Heights, MA: Allyn and Bacon.
- Shulman, L. (1998). Teacher portfolio – a theoretical activity. In Lyons, N. (ed.) *With Portfolio in Hand. Validating the New Professionalism*, pp 23–37. New York: Teacher College Press.
- Smith, K. (2007). Empowering school and university-based teacher educators in their role as assessors of student teachers' teaching practice – a school–university cooperation. *Educational Research and Evaluation* **13**(3), 279–293.

Relevant Websites

- <http://www.ate1.org> – Association of Teacher Educators (ATE).
- <http://www.mofet.macam.ac.il> – The MOFET Institute: Research, Curriculum and Program Development for Teacher Educators.
- <http://www.west.asu.edu> – University of Arizona.
- <http://www.bath.ac.uk> – University of Bath.

Self-Study by Teacher Educators

T Russell, Queen's University, Kingston, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Self-study by teacher educators is a relatively recent research activity, becoming a named area within the field of teacher education research in 1993 when a special interest group was formed within the American Educational Research Association (AERA). The group's name, Self-Study of Teacher Education Practices (S-STEP), does little to reveal what self-study is, and the goal of this article is to describe self-study, explain why it is significant within teacher education research, illustrate how self-study is carried out, and summarize some of the influence that self-study can generate. A growing self-study literature of conference proceedings, edited book collections, and refereed journal papers led to publication, in 2004, of the *International Handbook of Self-Study of Teaching and Teacher Education Practices* (Loughran *et al.*, 2004), which is now the most definitive reference on self-study by teacher educators. To introduce the context in which self-study research is situated, it is helpful to begin with a recent perspective on the fundamental problems faced by those who work in teacher education.

Self-Study as a Response to the Challenges of Learning to Teach

The significance of self-study as a relatively recent development in teaching and teacher education becomes clearer against the backdrop of the broad context in which teacher educators work. Darling-Hammond (2006) describes three fundamental problems that represent significant challenges to the work done by teacher educators in preservice programs of teacher preparation. These problems have been faced by teacher educators for decades, implicitly if not explicitly.

1. The problem of the apprenticeship of observation: "Learning to teach requires new teachers to understand teaching in ways quite different from their own experience as students" (p. 35).
2. The problem of enactment: "Learning to teach requires that new teachers not only learn to 'think like a teacher' but also to 'act like a teacher'" (p. 35).
3. The problem of complexity: "Learning to teach requires new teachers to understand and respond to the dense and multifaceted nature of the classroom" (p. 35).

The first problem, concerning the apprenticeship of observation, is not just intriguing; it is rarely addressed explicitly in planning and enacting a preservice teacher education program. "A significant challenge teachers face is that they enter teaching having already had years of experience in schools" (p. 35). Darling-Hammond quotes from a classic sociological analysis of teaching in which Lortie observes that all students, including those in preservice teacher education programs, typically lack a great deal of information about why teachers act as they do:

They are not privy to the teacher's private intentions and personal reflections on classroom events. Students rarely participate in selecting goals, making preparations or postmortem analysis. Thus they are not pressed to place the teacher's actions in a pedagogically oriented framework. (Lortie, 1975, p. 62)

Even when pressed, many teacher candidates find it difficult to describe how and why they went about selecting goals and making preparations for their teaching, and then analyzing the effects of their teaching. As a case in point, preservice candidates may seem to understand what we refer to as reflection, but they are often challenged when trying to analyze and develop new frames for what they are doing in the practicum classroom.

The innovative teaching approaches that teacher educators urge prospective teachers to consider and that they believe could improve what happens in schools have emerged from and are supported by the extensive research over the last 40 years on how people learn. Research tells us that people learn best when they are active, challenged, and engaged. Research reminds us that all students come to us with prior knowledge in the subject area. The largely invisible apprenticeship of observation generates the prior knowledge and beliefs that our candidates bring with them, but teacher educators are not necessarily adept at recognizing, exposing, and responding to those beliefs and adapting their teaching accordingly.

If prospective teachers consider themselves to be blank slates or if they are treated as such, they are unlikely to see teaching in new ways that help them to understand and cope with the first years of teaching. They may also be less likely to go on to become the teachers that they and their teacher educators hope they will become. The problem of the apprenticeship of observation spills over into the problem of enactment. "Learning how to think and act in

ways that achieve one's intentions is difficult, particularly if knowledge is embedded in the practice itself." Much information "best emerges in the actual work of teaching—and guides the planning and instruction that follows" (Darling-Hammond, 2006, p. 37). "Novices bring their own frames of reference to the ideas they encounter in teacher education; these may be incompatible with the approaches they are learning about in their coursework and clinical work" (p. 38).

To illustrate the problem of complexity, Darling-Hammond cites the work of Lampert (2001) and extracts these four elements of the problem (Darling-Hammond, 2006, p. 39, emphasis in original):

1. *Teaching is never routine.*
2. *Teaching has multiple goals* that must be addressed simultaneously.
3. *Teaching is done in relationship to diverse groups of students.*
4. *Teaching requires multiple kinds of knowledge to be integrated.*

Although teacher educators might feel that they understand the meaning of these propositions, it is quite another issue to describe how teacher education courses can help candidates to appreciate the problem of complexity as their teacher educators work to help them prepare for practicum experiences and the first year of teaching. Teacher candidates often expect that they will wear the teaching mantle with ease. Unwittingly, preservice programs may suggest that putting on the teaching mantle is relatively straightforward. Until candidates acknowledge and confront their educational histories as well as the diversity of learners in every classroom and the challenges of creating contexts of productive learning, teacher education will continue to be the poor cousin of university disciplines.

The problems of preservice preparation, induction, and professional development have been documented. The charge of fragmentation and conceptual impoverishment applies across the board. There is no connective tissue holding things together within or across the different phases of learning to teach.

The typical preservice program is a collection of unrelated courses and field experiences. Most induction programs have no curriculum, and mentoring is a highly individualistic process. Professional development consists of discrete and disconnected events. Nor do we have anything that resembles a coordinated system. Universities regard preservice preparation as their purview. Schools take responsibility for new teacher induction. Professional development is everybody's and nobody's responsibility. (Feiman-Nemser, 2001, p. 1049)

One broad conclusion that can be drawn from these researchers is that teacher education is far more complex than it has commonly been taken to be, if we consider the

short period of professional preparation relative to other professions and if we consider the lack of explicit attention to the problems of apprenticeship of observation, enactment, and complexity. When self-study of teacher education practices arrived on the landscape of teacher education, it was with a strong sense that teacher educators themselves must look within. We learned from the apprenticeship of observation: "we teach as we were taught" tends to be just as true for teacher educators as it is for teachers. We also must confront the problems of enactment and complexity. One of the very first observations that teacher candidates may make in their education courses involves the teacher educator who uses one teaching method (often transmission-based) to advocate another (often inquiry-based). If teacher educators cannot practice what they preach and enact their own values and understandings, what incentive is present for teacher candidates to attempt to do so? Similarly, Darling-Hammond's four statements about the problem of complexity are just as true for teaching done in preservice programs as they are for teaching done in primary and secondary schools. If teacher educators wish to address the problems of apprenticeship of observation, enactment, and complexity with those they wish to help to learn to teach, they are very unlikely to succeed if they have not considered their own development and current practices from the perspectives provided by those same problems.

What is Self-Study and How Does One Begin?

In the context of teacher education, self-study is research in one's own setting of practice to understand and reduce the gap between the good intentions of teacher educators and the actual learning of preservice teacher candidates. Two related steps are essential:

1. Self-study requires making what Schön (1992) described as a reflective turn, a fundamental change in one's perspective on personal teaching practices.

When we attend to what we know already, appreciating the artistry and wisdom implicit in competent practice, believing that by reflection on that practice we can make some of our tacit knowledge explicit, we take a "reflective turn" that leads us to see students and teachers . . . as participants in a kind of reflective practice, a communicative and self-reflective practice of reciprocal inquiry. (Schön, 1992, p. 123)

2. Self-study requires viewing those learning to teach not as empty vessels to be filled with knowledge they lack but as experienced students with extensive but incomplete and potentially misleading knowledge of

teaching that must be explored and reinterpreted as part of becoming a teacher.

Broadly speaking, self-study begins with discomfort and dissatisfaction with the view that learning to teach is a straightforward matter of learning how to teach by being told and then moving to practicum settings to practice what one has already learned. Self-study becomes a powerful research strategy when a teacher educator is willing to explore the complexities of learning by those who wish to become teachers in the context of exploring simultaneously the complexities of one's own learning to teach. Self-study begins by attending to events, experiences, and interactions that had previously been ignored, not because they were irrelevant but because they were taken for granted and accepted as familiar and unproblematic.

To illustrate the nature of self-study, consider the following personal accounts provided by four teacher educators who contribute regularly to the literature of self-study of teacher education practices (Hamilton *et al.*, 1998). Five years into their self-study research, these individuals each responded to the question of whether self-study research is changing teacher education.

John Loughran focused his early self-study on identifying principles of his own practice as a teacher educator:

I believe that one of the most important shifts in my teaching practice is bound up in the ability . . . to articulate the principles of practice that underpin my teaching. Student-teachers do not learn about teaching by being told about it, nor do they learn about teaching simply by experiencing it. They learn about teaching by being involved in a range of activities and possibilities, through reflection on experience and by being able to question that which they experience and those with whom that experience is associated. (p. 1)

He lists six elements that he has identified by attending closely to the learning of the student teachers he teaches: relationships, trust, independence, purpose, engagement/challenge, and modeling (pp. 1–2).

Mary Lynn Hamilton began her report with the event that triggered her self-study:

Within the first few minutes of my new career . . . , I overheard a young woman say quite passionately that her professors never enacted their beliefs. From her perspective, the university classroom was riddled with inconsistencies. That unknown student inadvertently triggered my deliberate inquiry into my own teaching—in what ways do my actions in the classroom conflict with [my] beliefs? . . . Early in my career as a teacher educator I recognized the importance of aligning beliefs and actions in the classroom. As I did that I saw that teacher evaluations and my own comfort level in the classroom improved. . . . I noticed shifts in my language about teaching and in my students' language about teaching and learning. (p. 2)

Tom Russell's entry into self-study was driven in part by returning to the secondary school classroom after an absence of more than 20 years:

Self-study has enabled me to understand my teaching and my students as never before. When the term "self-study" was gaining attention and interest, I had taken myself back to the secondary school physics classroom to better understand the work I was helping people learn. Now I had a better conceptual framework for viewing both my own development as a teacher and the development of those learning to teach. I also began teaching an M.Ed. course in action research, and through successes and failures, self-study became the key to better understanding. . . .

I can now credit self-study with enabling me to develop the following pedagogical principles that were present in my teaching in 1997–98 but very absent in 1992–93.

- Build on and trust in their extensive early teaching experiences.
- Build a community in my classroom.
- Every class includes a significant experience, with implications for how we teach.
- Gather backtalk [students' comments about the quality of their learning] often and distribute it freely.
- Identify a few central themes, post visual cues to each, and link to them as often as possible.
- Replace lectures with notes on a website; use classroom time for experiences and discussions.
- Encourage and reward self-directed learning.
- At the end of the course, ask students what they think you learned from them. (p. 3)

Vicki LaBoskey wrote that virtually all her research since beginning self-study "has been designed to help me learn about my own teaching. The driving force behind my work has been an effort to improve the ways in which I teach about teaching to prospective and inservice teachers" (p. 4). She gives four examples, and the third illustrates how self-study can influence one's work with colleagues:

The third instance is on a larger scale; it involves a change in program philosophy that resulted from self-studies my colleagues and I did of our individual and combined efforts with regard to our guiding principles. During the 1996–97 academic year, our teacher credential program had been structured around five fundamental principles. Throughout that year we engaged our student teachers in a number of evaluative activities in order to help us determine how well we were doing with regard to these goals. We learned that, despite our emphasis on constructivist learning theory, our student teachers were not as strong as we wanted them to be in the teaching of powerful subject matter knowledge to their students. As a result, we added a sixth guiding principle this year

which had a direct impact on both what we taught and how we taught it. (p. 4)

Additionally, even though self-study is the study of oneself, it is not an activity done on one's own. Loughran and Northfield (1998) explained the rationale for collaboration during self-study:

The learning in self-study is intensely personal, but self-study itself requires collaboration. . . . In this approach, the intensely personal aspects of the study that might otherwise be simply accepted without challenge or scrutiny are able to be professionally and constructively challenged from within the study itself. (p. 14)

There is always a danger that individuals will interpret situations in ways that reinforce existing perceptions. Genuine study of classrooms is associated with a willingness to consider alternate frames of reference, and colleagues are an important source of ideas and support as the teaching and learning are reviewed. (p. 16)

Thus, it is typically both important and highly productive to identify a collaborator for ongoing analysis of one's questions, perspectives, and data during self-study research.

Focal Points of Self-Study Research

The preceding examples illustrate the early efforts of four teacher educators who hold strong commitments to self-study research. In each instance, the trigger was an experience that required better understanding of personal teaching practices in order to improve the quality of learning by those preparing to become teachers. Identifying principles of pedagogical practice is a significant theme in much self-study inquiry. Two of the main focal points for self-study researchers are the improvement of teacher education programs (e.g., Loughran and Russell, 2002) and the fostering of social justice (e.g., Tidwell and Fitzgerald, 2006).

As previously noted, the enterprise of preservice teacher education is unique in several significant respects. No other profession works with individuals who already know so much, strictly by observation, about what experienced teachers do. No other profession works with individuals who have spent so much time talking, usually superficially, about the quality of teaching and their own school learning experiences. Because this knowledge of what teachers do is inferred but never checked, preservice teachers know very little about how teachers think about their work. Perhaps more significantly, teacher education is the only profession that must practice the profession of teaching as it tries to help others develop propositional and experiential knowledge of the profession. When a teacher educator teaches preservice candidates, it is only

too easy to fall into contradictions such as lecturing about the importance of not lecturing. In teacher education, it is impossible to separate how one teaches from what one teaches. It is the high probability of unintended but significant gaps between what we say and what we do as teacher educators that drives many teacher educators to self-study.

Just as when teaching is thought of as transmission of curriculum content to students, when preservice teacher education is conceptualized as transmission of knowledge about teaching from an experienced individual to a beginner, there is little need to listen to what students have to say about their learning experiences. Self-study by teacher educators challenges this perspective directly, and self-study often includes finding ways to listen to and include the voices of those being taught. Cook-Sather (2002) argued well for the importance of listening to students, and breaking out of old assumptions and patterns is an important element of self-study:

Most power relationships have no place for listening and actively do not tolerate it because it is very inconvenient: to really listen means to have to respond. Listening does not always mean doing exactly what we are told, but it does mean being open to the possibility of revision, both of thought and action. At a minimum, it means being willing to negotiate. Old assumptions and patterns of interaction are so well established that even those trying to break out of them must continue to struggle. And understanding that is part of what it means to listen. (p. 8)

Methodology of Self-Study Research

From the perspective of self-study methodology, two documents are particularly significant. The first is Bullough and Pinnegar's (2001) paper on 'Guidelines for quality in autobiographical forms of self-study research'. The prominence of this journal and the arguments set forth by Bullough and Pinnegar added important legitimacy to self-study methodology. Most self-study research is qualitative rather than quantitative; statistical information can be relevant and helpful, but self-study tends to focus on improving the quality of teaching and that of students' learning, both in individual teacher education classrooms and in teacher education programs.

The second document is LaBoskey's (2004) chapter in the *International Handbook of Self-study of Teaching and Teacher Education Practices* (Loughran *et al.*, 2004). Her chapter summarizes "the epistemological, pedagogical, and moral/ethical/political underpinnings of self-study, which serve as the conceptual framework for the field" (LaBoskey, 2004, p. 817). As such, it is an essential reference and excellent entry point for anyone new to self-study. LaBoskey identifies four central features of self-study research methodology:

“improvement-aimed,” “interactive,” employing “multiple, primarily qualitative methods,” and demanding “that we formalize our work and make it available to our professional community for deliberation, further testing, and judgment” (pp 859–860). LaBoskey’s conclusion offers insights that link the methodology of self-study to its purposes:

The methodology of self-study is well conceptualized, well grounded in epistemological and pedagogical theory, and well justified by interconnected moral, ethical, and political values and ideals. It has clear features that . . . will allow us to proceed with integrity and evaluate with confidence. . . . Those of us in the field need to continue the process by incorporating into our teaching and research practice the understandings and procedures we deem trustworthy enough to risk trying, with appropriate adaptation and assessment, in our own programs with our own students. (p. 860)

Zeichner (2007, p. 37) has argued that self-study has found its place in teacher education research by adding “dignity to the important yet universally undervalued work of those faculty and staff who educate a nation’s teachers.” He has also issued an important challenge to those who engage in self-study research, urging them to focus on how their learning from self-study can contribute to the questions asked generally by teacher educators and policymakers:

The political realities . . . in many countries . . . require that self-study researchers become more engaged with the mainstream of research in teacher education and insert the perspectives and voices of practicing teacher educators more centrally into the policy debates that frame teacher education practice at the local level. (p. 43)

Now that self-study research by teacher educators has developed a significant body of literature and a generally accepted research methodology, it is certainly appropriate for self-study researchers to add their research findings to the broader body of educational research.

Summary

Self-study by teacher educators addresses the long-standing gap between what preservice teacher education programs set out to accomplish and the professional learning actually perceived by those who complete these programs and begin to teach. Those who engage in self-study recognize that learning to teach is more complex than those who enter teacher education programs might imagine. Self-study begins with a willingness to address the familiar gap between goals and realities by gathering data that can challenge and promote rethinking of long-standing assumptions. Principles of pedagogical practice

are often an important issue in self-study research. Improving teacher education practices, at both personal and program levels, and fostering social justice have been significant focal points in much self-study research by teacher educators. The methodology of self-study is now well documented and the field is ready to more directly make contributions to our knowledge and understanding of major issues in preservice teacher education

Bibliography

- Bullough, R. V., Jr. and Pinnegar, S. (2001). Guidelines for quality in autobiographical forms of self-study research. *Educational Researcher* 30(3), 13–21.
- Cook-Sather, A. (2002). Authorizing students’ perspectives: Toward trust, dialogue and change in education. *Educational Researcher* 31(4), 3–14.
- Darling-Hammond, L. (ed.) (2006). *Powerful Teacher Education*. San Francisco, CA: Jossey-Bass.
- Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record* 103, 1013–1055.
- Hamilton, M. L., LaBoskey, V. K., Loughran, J., and Russell, T. (1998). Have five years of self-study changed teacher education? Artifacts of our personal development as teacher educators. In Cole, A. L. and Finley, S. (eds.) *Conversations in Community: Proceedings of the Second International Conference of Self-Study of Teacher Education Practices, 1998, Herstrmonceux Castle*. Kingston, ON: Queen’s University Faculty of Education, pp 1–5.
- LaBoskey, V. K. (2004). The methodology of self-study and its methodological underpinnings. In Loughran, J. J., Hamilton, M. L., LaBoskey, V. K., and Russell, T. (eds.) *International Handbook on Self-Study of Teaching and Teacher Education Practices*, pp 817–869. Dordrecht: Kluwer.
- Lampert, M. (2001). *Teaching Problems and the Problems of Teaching*. New Haven, CT: Yale University Press.
- Lortie, D. (1975). *Schoolteacher: A Sociological Study*. Chicago, IL: University of Chicago Press.
- Loughran, J. J., Hamilton, M. L., LaBoskey, V. K., and Russell, T. (eds.) (2004). *International Handbook on Self-Study of Teaching and Teacher Education Practices*. Dordrecht: Kluwer.
- Loughran, J. and Northfield, J. (1998). A framework for the development of self-study practice. In Hamilton, M. L. (ed.) *Reconceptualizing Teaching practice: Self-Study in Teacher Education*, pp 7–18. London: Falmer.
- Loughran, J. and Russell, T. (2002). *Improving Teacher Education Practices through Self-Study*. London: Routledge Falmer.
- Schön, D. A. (1992). The theory of inquiry: Dewey’s legacy to education. *Journal of Curriculum Studies* 22, 119–139.
- Tidwell, D. and Fitzgerald, L. (eds.) (2006). *Self-Study and Diversity*. Rotterdam: Sense Publishers.
- Zeichner, K. (2007). Accumulating knowledge across self-studies in teacher education. *Journal of Teacher Education* 58(1), 36–46.

Further Reading

- Aubusson, P. and Schuck, S. (2006). *Teacher Learning and Development: The Mirror Maze*. Dordrecht: Springer.
- Berry, A. (2007). *Tensions in Teaching about Teaching: Understanding Practice as a Teacher Educator*. Dordrecht: Springer.
- Hamilton, M. L. (ed.) (1998). *Reconceptualizing Teaching Practice: Self-Study in Teacher Education*. London: Falmer.

- Hoban, G. F. (ed.) (2005). *The Missing Links in Teacher Education Design: Developing a Multi-Linked Conceptual Framework*. Dordrecht: Springer.
- Kosnik, C., Beck, C., Freese, A. R., and Samaras, A. P. (eds.) (2006). *Making a Difference in Teacher Education through Self-Study: Studies of Personal, Professional and Program Renewal*. Dordrecht: Springer.
- Kroll, L., Cossey, R., Donahue, D. M., et al. (2004). *Teaching as Principled Practice: Managing Complexity for Social Justice*. Thousand Oaks, CA: Sage Publications.
- Loughran, J. (2006). *Developing a Pedagogy of Teacher Education: Understanding Teaching and Learning about Teaching*. London: Routledge.
- Loughran, J. and Russell, T. (2007). Beginning to understand teaching as a discipline. *Studying Teacher Education* 3(2), 201–211.
- Russell, T. and Loughran, J. (eds.) (2007). *Enacting a Pedagogy of Teacher Education: Values, Relationships, and Practices*. London: Routledge.
- Samaras, A. P. (2002). *Self-Study for Teacher Educators: Crafting a Pedagogy for Educational Change*. New York: Peter Lang.

TEACHER EDUCATION – TEACHING SPECIFIC DOMAINS

Contents

Arts Teacher Education
Mathematics Teacher Education
Music Teacher Education
Science Teacher Education
Social Studies Teacher Education
Teacher Education for Elementary Education
Teacher Education for Teaching the Gifted
Teacher Preparation in Mother Tongue Education
Vocational Teacher Education

Arts Teacher Education

L Beudert, University of Arizona, Tucson, AZ, USA

© 2010 Elsevier Ltd. All rights reserved.

This article examines arts teacher education in relation to the following arts disciplines: dance education, drama/theater education (drama and theater are used interchangeably herein), music education, and visual art education, given that it is these art forms that are mostly taught within schools. Our focus is primarily on the preparation of preservice teachers who will eventually work as specialists in a specific arts discipline in primary and secondary schools; however, the preparation of nonspecialist primary teachers, who are required to teach one, or more, of the arts disciplines, are also addressed.

An International Arts Education Landscape

The arts are culturally diverse and relate to humans all over the world. They are fundamental forms of understanding and are part of our lived experiences. Educating about and within the arts involves introducing learners to a range of traditional and innovative forms, such as classical and popular music, film, dance, plays, sculpture, crafts, singing, photography, digital arts, and folk art and dances. The arts not only sustain and reflect cultural traditions and customs, but they also engage and confront sociocultural issues (e.g., sexuality, gender, race, ethnicity, politics, economics, oppression, social justice, the environment, and spirituality) within contemporary societies. In addition, educators, artists, musicians, performers, actors, and dancers respond to the rapid development of the Internet,

and new media and technologies within their artworks, dances, plays, musical compositions, and performances. Furthermore, traditional borders between the various arts are becoming blurred and less distinct, as artists pursue interdisciplinary connections and associations across and within the arts. Alternatively, however, trends toward globalization can lead to the trivialization of the cultural aspects of local arts and customs (e.g., in India, Africa, and Latin America), since many art forms are passed on via cultural traditions and apprenticeships within many indigenous groups.

An education in the arts is a vital component within our lives, and it is central to the human and equal rights of all (Mason, 1999). Nevertheless, how diverse peoples, cultures, societies, and governments view the arts vary, and these viewpoints, in turn, affect the nature and availability of arts education and arts teacher education worldwide.

Unfortunately, the arts in schools do not fare as well as other school subject disciplines. The arts are often perceived as nonacademic, or are considered too costly in terms of resources and time within school curricula; thus, they frequently reside on the peripheries of formal primary, secondary, and even tertiary schooling, despite appeals, worldwide, for their inclusion in education. In some countries (e.g., the Netherlands, Germany, Israel, England, Australia, Canada, New Zealand, Japan, and the United States), children and youth participate in at least one of the arts disciplines, at some point, during their primary and secondary school education. Regrettably

though, in parts of Europe, Africa, the Middle East, Latin America, and Asia, fewer possibilities exist, if at all, for children to be formally educated within the arts.

In many countries (an exception is the United States), specialist arts teachers tend to teach only in secondary schools as primary classroom teachers, with little or no background in the arts, are assigned to teach arts education. In some nations (e.g., the United States) where music and visual art specialists teach at the primary level, foci on testing in the so-called academic subjects (e.g., mathematics, language arts, and science) are a factor in the gradual decline in the number of arts courses taught in schools; and in some countries (e.g., Australia and Norway), interdisciplinary or integrated arts education courses are taught in primary schools, rather than separate courses in one or more of the arts disciplines. Besides, there are fewer dance and drama/theater education programs within primary and secondary schools and tertiary education compared to arts educational programs in music and visual art.

Given the aforesaid, the task of the arts educator within schools is both broad and multifaceted. Thus, arts teacher educators are charged with training arts teachers, primary- and secondary-level arts specialists, and nonspecialist primary teachers, who can (1) understand, negotiate, and manage the complexities of teaching the arts within today's changing and diverse societies; (2) develop curricular, pedagogical, and inquiry approaches and strategies that will allow them to teach their respective arts disciplines authentically, ethically, and meaningfully within schools; (3) maintain and advocate for the inclusion of the arts within curricula and schools; and (4) confront and tackle societal and global indifference to the arts themselves. In addition, arts teacher educators themselves must strive by necessity to be knowledgeable about, and up-to-date within, their particular arts specialty.

Navigating the Arts Teacher Education Landscape

In order to find our way around this vast arts teacher education landscape more clearly, this article focuses on four specific themes within arts teacher education; each of which has its own set of challenges, dilemmas, and strengths. These themes are: (1) the changing nature of arts content knowledge, (2) emerging pedagogical practices within arts teacher education, (3) professional and institutional issues, and (4) what teacher education can learn from arts teacher education. Examples from each of the individual arts education disciplines are interwoven within each theme.

The Changing Nature of Arts Education Content Knowledge

Content knowledge within arts education

In most countries, arts specialists are required to have a strong background (at least a baccalaureate degree) in an arts discipline and/or related subject matter; nonetheless, not all countries (e.g., Greece, Italy, Spain, and Belgium) provide opportunities for preservice professional training. Obviously, the nature of these degrees and the coursework options within them will vary; there is no one single degree or set of mandated courses that future arts teachers are required to complete in their subject disciplines. Some may have studied within traditional universities and colleges that offer distinct arts education degrees in dance, drama/theater, music, or visual art education. Others may have earned a degree primarily in dance, theater, music, or visual art from either a traditional university or college or from a specialized, professional arts institution, such as a college of art or conservatory of music or drama. This latter group of prospective teachers will enter some form of postbaccalaureate or graduate-level teacher education program in order to become qualified to teach. In some nations, preservice arts teachers do hold baccalaureate degrees, but they must take competitive exams before being accepted into teaching (e.g., France) or they must have earned master's degrees before taking teacher education coursework (e.g., Finland).

Given the aforementioned, prospective arts teachers bring to their teacher education coursework a range of differing dance, theatrical, musical or artistic abilities, understandings, and skills. Nonetheless, merely being knowledgeable within an art form or being competent as a musician or dancer does not automatically mean that all artists or performers have the requisite knowledge base to teach others about the arts. For example, prospective dance teachers must not only be experts at dancing, but they must also develop an understanding of choreography, improvisation, dance history, dance criticism, dance forms, esthetics, and issues related to cultural diversity within dance, if they are to be successful dance educators in primary and secondary classrooms. Moreover, in some countries (e.g., in Hong Kong, Finland, and Taiwan), preservice dance teachers are prepared within physical education teacher education programs rather than in programs that specialize in dance education.

In contrast, drama/theater education preservice teachers, who have been introduced to the academic aspects of theater (e.g., plays, playwrights, and theater history) in their tertiary degree studies, will enter preservice drama/theater education programs in which they will be introduced to aspects of creative dramatics, drama in education, educational drama, process drama, or theater in education. Drama/theater education in many primary

and secondary schools is viewed as child-centered and as a means of helping learners develop social skills, such as empathy and caring, although some contemporary drama/theater teacher educators are beginning to argue that both practical and theoretical frameworks must be tackled within drama/theater education programs.

The dichotomy within drama/theater education implies that, at times, the art forms that are taught within primary and secondary schools may not resemble those that are promoted within tertiary degree programs or reflected in contemporary arts practices. In visual art education, for example, many teacher educators are concerned that preservice teachers are able to develop primary and secondary school curricula that enable children and youth to explore contemporary and multicultural issues, themes, and contexts, such as those related to gender, identity, social justice, design, contemporary crafts, community, the environment, and today's media-infused and visual world. All too often, however, the artworks created in schools, particularly primary schools, are devoid of these qualities, and resemble what can be termed child art or school art.

Arts teacher educators must ensure that today's preservice teachers, once they start their teaching careers, do not always replicate the types of arts education practices that they themselves experienced in their own primary and secondary schooling, and so fail to implement contemporary ideas within the field. Furthermore, primary classroom teachers who teach visual art, and the other arts disciplines, do not often have the necessary content background to teach the arts successfully; thus, they must be helped to develop and teach arts curricula that move learners' understandings beyond simplistic concepts and practices.

Balancing local and global knowledge in arts education

Today's arts teacher educators must successfully prepare teachers to work within their local communities, honor the differences of their students, and acknowledge the traditions of local artists and, simultaneously, recognize the sociocultural aspects of the arts in a changing world. Immediate quandaries surface when making decisions about what content knowledge should preservice teachers know within their respective arts disciplines, whose knowledge should they teach, and what roles should new media and technologies play in preservice arts teacher education. These three major tensions are briefly highlighted further:

First, there is the concern that preservice arts teacher education programs in non-Western nations are often associated with Eurocentric practices and colonial models of education. Preservice teachers leave these programs with little preparation and expertise in arts subject matter that reflects the ethnicities, cultures, and identities of their future students. Dance teacher education practices

in Brazil serve as an example. Dance has only become a serious area of study since the 1990s, despite Brazil's long history of dance traditions, which originate from the cultural practices of indigenous peoples, black African slaves, and Portuguese colonizers. Nevertheless, dance education is still mostly taught from Eurocentric, classical, and academic perspectives. Such foci on Western practices are also evident in the other arts education disciplines; thus, many Brazilian teacher educators strongly advocate that preservice arts teachers must be taught about local arts traditions, coupled with knowledge about the informal arts cultures of Brazilian children and youth.

Second, local communities are changing in many parts of the world, especially in the West, partly due to immigration patterns and shifting demographics. For example, within some European countries, arts teacher educators are faced with the challenge of preparing teachers to teach students from families of recent immigrants and refugees. In many European countries, this challenge is exacerbated further when national educational policies and school curricula frequently emphasize teaching about a country's own specific culture, rather than curricula that relate to the multicultural backgrounds of these displaced children and youth.

Third, in today's world of mP3 players, iPods, mobile phones, podcasting, and Internet sites, such as YouTube, the retrieval of music, videos, performances, and visual data from the Internet is accessible in many continents. Thus, arts teacher educators must help preservice teachers understand the curricular and pedagogical benefits and problems associated with these technologies within arts education. In visual art education, for instance, inquiries into digital technology, advertising, cyberspace, and mass media are becoming prominent, because some teacher educators recognize the potential of these technologies for promoting intercultural understanding, along with exploring contemporary and ethical issues related to analyzing, exploring, and deconstructing media and visual culture.

The issues outlined above suggest that arts teacher education faculty members face serious challenges in terms of what they should teach and advocate within their courses and research. Fundamentally, arts teacher educators cannot help future arts teachers learn all there is to know, for as African music teacher educator, Oponelo (2000) argues, teachers cannot adequately be prepared in both Western and local African musical traditions. Mason (1999) points out that the traditional arts should not succumb to the trends, unpredictability, and uniformity of globalization, whereas other arts teacher educators believe that the integration of new technologies are inevitable within primary, secondary, and tertiary arts education. Tensions clearly exist in balancing local and global knowledge within arts teacher education.

Emerging Pedagogical Practices and Inquiries in Arts Teacher Education

Arts teacher education: Pedagogy and practice

Preservice arts teachers must develop an understanding of the interrelationships of arts-based content with successful and authentic pedagogical methods that can transform this content meaningfully within classroom contexts. For example, performance is central to music, yet preservice music teachers must not just perform music, instead they are obliged to identify with appropriate pedagogical methods intrinsic to helping their students learn to play, understand, and appreciate music.

Preservice teacher education coursework and field and practicum experiences within school settings introduce teacher candidates to various pedagogical methods within their respective arts disciplines. An array of teaching and learning practices is threaded throughout arts teacher education. Most teacher educators adopt approaches such as traditional lecturing and demonstrations, cooperative and group learning, and one-on-one tutoring. Others teach (1) using various technologies, such as television and online distance learning; and (2) employing feminist pedagogical approaches, which they believe serve as models for facilitating collaborative learning and student construction of knowledge within the preservice classroom. Still others advocate for the development and modeling of pedagogies intrinsic to the arts that they teach.

Refashioning preservice arts teachers' beliefs

Preservice teachers must not only perceive themselves as change-agents for the arts within schools, thereby transforming, enriching, and expanding the lives of their students through the arts, but they must also acknowledge that, as life educators (Nieto, 2007), they must seek change within themselves and so question their own assumptions and beliefs about teaching the arts and their students. How then do arts teacher educators help prospective teachers acknowledge the many ways in which learners learn and so teach for deeper understanding, yet at the same time, confront, question, and reflect upon their own pedagogical beliefs? For example, how are preservice teachers encouraged to examine their beliefs about arts content, pedagogy, and learners, together with their assumptions about the arts and others in terms of gender, race, class, culture, sexual orientation, and disabilities? How can they reflect on their own arts background and education, so as not to merely imitate experiences and curricula grounded in their own primary and secondary schooling? For example, preservice teachers' beliefs about the nature of visual art education are important indicators of how capable they will be as future visual art teachers; merely having a strong visual art background does not necessarily mean that preservice teachers will be successful teachers. The following

pedagogical strategies are examples of how some arts teacher educators address these issues in their courses.

In drama/theater teacher education, teacher educators use performance methods that enable preservice teachers to enact their pedagogical, social, and political beliefs in forms of role-playing. As a case in point, Greenwood (2002) drew upon interactive drama/theater education as a cross-cultural tool with preservice Maori educators in remote northern New Zealand. A participatory drama workshop encouraged preservice teachers and instructors to inquire into their educational goals and philosophies, as well as to question their deep-rooted assumptions about the Maori culture. Future teachers experimented with role-playing and acted out scenes. Interactive drama/theater techniques are successful pedagogical strategies that allow for the examination of preservice teachers' beliefs and self-knowledge.

Within dance, drama/theater, and visual art education, the body is viewed as a performing, sensory, and pedagogical tool for exploring cultural constructions of identity, gender, race, social justice, and sexual orientation. Such interactive strategies facilitate preservice teachers' own interrogations of their assumptions about others, which in turn, allow them to take steps toward acknowledging and dispelling issues related to gender-bias, sexism, homophobia, and racism, as well as to reflect upon their future roles within classrooms.

Visual journals (comprising, e.g., drawings, paintings, text, and collages) are constructed in visual art education preservice classrooms. The images within the journals serve as reflective commentaries on preservice teachers' professional development. For instance, the creation of visual or performance metaphors (e.g., works of art, texts, installations, musical compositions, and performance pieces) engages preservice teachers in conceptualizing and representing their teaching beliefs. Preservice teachers in the arts construct traditional portfolios of their experiences within student teaching and courses, or create electronic portfolios (comprising artworks, photographs, videos, performances, curriculum vitae, and teaching statements) that are often placed on Internet websites.

Arts teacher educators encourage the use of autobiographical life histories and the writing of educational memoirs with preservice arts teachers. For instance, teacher educators are responsible for not only helping nonspecialist primary teachers become more confident in teaching the arts, but they must also provide ways in which they can contemplate their reluctance to teach the arts. If primary classroom teachers, for instance, write and reflect on stories about their own prior arts experiences and their own art teachers, they are often more likely to acknowledge the value of teaching and integrating the arts in their classrooms in authentic ways, even if they find teaching them difficult.

Some arts teacher educators utilize reflective journaling practices in order to foster the growth of preservice teachers as caring, ethical, and committed professionals. Reflective journaling allows preservice teachers to intentionally evaluate their beliefs about the arts, as well their attitudes and beliefs about learners. Additionally, arts teacher educators ask preservice teachers to examine written case studies about arts teaching. By reading and analyzing these case studies, preservice teachers can explore and recognize sociocultural issues related to teaching within the arts. In addition, preservice teachers write their own narratives and case studies about teaching, especially after observing and working within actual primary- and secondary-school settings. Online blogs and Internet chat rooms can also facilitate ongoing discussions about classroom theories and practices and the nature of arts education.

Professional and Institutional Concerns

An extensive literature review conducted for this article found that arts teacher education programs and practices unfortunately remain underresearched, although several comprehensive syntheses of international arts education research exist. Teacher education research is not central to the research carried out within each of the distinctive arts education fields. Large-scale and worldwide studies have not been completed on this topic. In particular, demographic information and data on arts teacher education programs are notably absent. The following professional and institutional concerns are now briefly examined.

Developing a familiarity with arts teacher educators

As the arts education disciplines shift philosophically and are continually being redefined, it is important to understand whether arts teacher educators themselves are able to accept and adapt to such changes. Today's arts teacher educators face many pressures to prepare preservice arts teachers for a lifetime of work in a very different world to that in which they themselves were educated. Surprisingly, few details exist about their professional and working lives, as they are silent and missing persons in the arts education literature.

In the United States, for example, many arts teacher educators work in small teaching-oriented tertiary institutions, where, because of heavy teaching loads and institutional pressures and commitments, they are not required to, or have insufficient time to conduct research; thus, they contribute little to the published scholarship within arts education. Their intellectual work frequently comprises designing courses, presenting in-service workshops, or working within their own arts disciplines as studio

artists, dancers, performers, and musicians. Those faculty members, who do conduct research, are less inclined to study teacher-education issues, even if they are involved in teacher training, and many seem reluctant to self-identify as arts teacher educators. Moreover, many faculty members responsible for teaching teachers are either employed part-time or are graduate students. In the United Kingdom, teachers in schools take on a large part of arts teacher education.

This lack of research on faculty teacher educators and their programs is worrisome and perplexing because it is difficult to ascertain how current and knowledgeable teacher educators are within their particular disciplines. For instance, what do arts teacher educators teach, and does this content reflect the changes that are taking place within their disciplines? Are they knowledgeable about and/or willing to inquire into and tackle such issues as balancing local and global knowledge, or incorporating advances in new media and technology within their courses? What are their beliefs about the arts and teaching? And how are their beliefs reflected in their pedagogical and intellectual practices? To add to these queries and uncertainties is the notion that a number of arts teacher educators are raising questions about the professional concerns of faculty members (and preservice teachers) who are culturally and/or racially diverse, gay or lesbian, and/or experience disabilities.

Recruiting preservice arts teachers

There is a need to recruit and prepare diverse arts teachers to meet the needs of the changing composition of schools. Teaching within the individual arts disciplines is obviously not an attractive profession to many within the younger generation. In the United States, for instance, there is not only a shortage of arts teachers of color, especially African-American teachers, but there is also a shortage of arts teachers, in general. What is more, there is an increase in the number of teachers who are retiring, and newer teachers are leaving the profession after teaching for only a few years. This state of affairs is exacerbated by the shortage of tertiary arts teacher education programs, especially in Latin America, parts of Europe, and on the African continent.

Given diminishing resources and changing priorities in education, some primary and secondary schools, especially within Canada and the United States, employ visual artists, musicians, dancers, and actors to teach children and youth, rather than qualified arts teachers. These artists usually have little or no pedagogical background or expertise in teaching their art forms, and are often called artists-in-residence or teaching artists. Basically, these artists have not taken any tertiary teacher-education coursework. It is clear that arts teacher education must become involved in preparing these artists in the future.

International alliances

At the 2006 UNESCO World Conference on Arts Education in Portugal, the International Drama/Theater and Education Association (IDEA), the International Society for Music Education (ISME), and the International Society for Education through Art (InSEA) created an alliance within which, they were determined to enlist support for the inclusion of arts education in schools and communities across the globe. This alliance draws on members from 90 countries, and illustrates how international organizations are pivotal forces in advocating for, as well as promoting, arts education and arts teacher education worldwide.

Learning from Arts Teacher Education

Teacher education is about the sharing of a collective body of knowledge about teaching and learning. What then can non-arts teacher educators learn from the arts and arts education? Dance, drama/theater, music, and visual art are fundamental forms of understanding (Eisner, 2002), since they embrace aural, visual, and kinesthetic traditions. Furthermore, knowledge within the arts contributes to the intellectual, emotional, imaginative, moral, sensory, and spiritual development of persons. Critical to a person's development and education as a dancer, musician, performer, or artist is an ability and willingness to take risks, improvise, solve pictorial and spatial problems, communicate, make judgments, collaborate, and work independently.

Exemplars from teaching and learning within arts teacher education can inspire teaching and learning practices within teacher education, in general. Bresler (2005) observes how aspects of musicianship, such as empathy, perceiving, interpreting, and communicating, are present in the intricacy, nuances, and unpredictability of teaching-learning relationships. Marques (2004) suggests that teaching is like choreographing a dance as arts teachers spend time researching and selecting content; these actions require rehearsals, discipline, and patience as an ongoing dialog between theory and practice takes place. Schonman (2005) proposes that teaching is about enactment and is a performing art; as a result, the theatrical provides an alternative means of reconceptualizing and analyzing teaching. Furthermore, arts-based pedagogical and research strategies such as visual portfolios, photographic and video documentation, performances, and visual and autobiographical narratives can assist in revealing the contextual nature of primary, secondary, and tertiary classrooms more fully.

Goodson (2007) notes that increasing globalization and knowledge-based economies place additional demands on schools, as well as adversely affect the "hearts and minds" (p. 137) of teachers as human beings. The qualities truly

needed in today's teachers resemble those qualities embraced, shared, and communicated within and across the arts (Eisner, 2002; Nieto, 2007). Future teachers need to possess many skills in order to be prepared for their multidimensional roles in today's multicultural and complex schools; therefore, preservice arts teachers, on account of their knowledge of the arts and their arts education preparation, may be at a distinct advantage compared to peers in other subject areas.

Conclusion

After reviewing sources for this article, it is clear that challenges and questions linger within the textures and sounds that lace together the fabric of international teacher education within dance, music, drama/theater, and visual art education. This extraordinary fabric is interconnected with copious threads and is simultaneously crafted with diverse multicultural lineages, which, at times, remain invisible within the arts education literature and teacher education as a whole. International arts teacher education must be more vividly portrayed and distinctly heard as teacher educators, practitioners, and preservice teachers advocate for, research, and grapple with the complexities of teaching and learning the arts within and across national borders.

See also: Curriculum in the Arts; Music Teacher Education.

Bibliography

- Bresler, L. (2005). What musicianship can teach educational research. *Music Education Research* 7, 169–183.
- Eisner, E. W. (2002). *The Arts and the Creation of Mind*. New Haven, CT: Yale University Press.
- Goodson, I. (2007). All the lonely people: The struggle for private meaning and public purpose in education. *Critical Studies in Education* 48, 131–148.
- Greenwood, J. (2002). Marking the unmarked: Using drama cross-culturally in teacher education in New Zealand. *Systems Research and Behavioral Science* 19, 323–329.
- Marques, I. (2004). Giving meaning to dance education in Brazil. In Mason, R. and O'Farrell, L. (eds.) *Issues in Arts Education in Latin America*, pp 75–86. Kingston, ON: Queens University Press.
- Mason, R. (1999). Multicultural arts education and global reform. In Boughton, D. and Mason, R. (eds.) *Beyond Multicultural Art Education: International Perspectives*, pp 3–18. New York, NY: Waxman Münster and Berlin München.
- Nieto, S. (2007). Solidarity, courage and heart: What teacher educators can learn from a new generation of teachers. *Intercultural Education* 17, 457–473.
- Oponelo, P. (2000). Cultural policies in Kenya. *Arts Education Policy Review* 101, 18–24.
- Schonman, S. (2005). Theatrical representation of teaching as performance. In Brophy, G. and Pinnegar, S. (eds.) *Learning from Research on Teaching: Perspective, Methodology, and Representation*, pp 28–331. Amsterdam: Elsevier.

Further Reading

- Barbosa, A. M. (1999). Cultural identity in a dependent country: The case of Brazil. *Arts Education Policy Review* **94**, 35–39.
- Beudert, L. (2006). *Work, Pedagogy and Change: Foundations for the Art Teacher Educator*. Reston, VA: National Art Education Association.
- Bonbright, J. and Faber, R. (eds.) (2002). *Research Priorities for Dance Education: A Report to the Nation*. Bethesda, MD: National Dance Education Organization.
- Boughton, D. and Mason, R. (eds.) (1999). *Beyond Multicultural Art Education: International Perspectives*. New York, NY: Waxman Münster, Berlin: München.
- Bresler, L. (ed.) (2007). *International Handbook of Research in Arts Education* (2 vols.). Dordrecht: Springer.
- Colwell, R. and Richardson, C. (eds.) (2002). *The New Handbook of Research on Music Teaching and Learning*. New York, NY: Oxford University Press.
- Eisner, E. W. and Day, M. D. (eds.) (2004). *Handbook of Research and Policy in Art Education*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Freedman, K. and Hernandez, F. (eds.) (1998). *Curriculum, Culture, and Art Education: Comparative Perspectives*. Albany, NY: State University of New York Press.
- Leong, S. (1997). *Music in Schools and Teacher Education: A Global Perspective*. University of Western Australia: ISME Commission for Music in Schools & Teacher Education and Callaway International Resource Centre for Music Education.
- Mason, R. and O'Farrell, L. (eds.) (2004). *Issues in Arts Education in Latin America*. Kingston, ON: Queens University Press.
- Rasmussen, B. and Østern, A. (eds.) (2002). *Playing Betwixt and Between: The IDEA Dialogues*. Bergen: IDEA Publications.

Mathematics Teacher Education

K Krainer, Universität Klagenfurt, Klagenfurt, Austria

S Llinares, Universidad de Alicante, Alicante, Spain

© 2010 Elsevier Ltd. All rights reserved.

Student teachers, teachers, and teacher educators are all regarded as teachers and as lifelong learners. Teachers are seen as active constructors of their knowledge, embedded in a variety of social environments that influence and shape them and, at the same time, are influenced and shaped by them. Therefore, teachers are expected to continuously reflect on their practice and to change it where appropriate. Teacher education is understood as a goal-directed intervention in order to promote teachers' learning, including all formal kinds of teacher preparation and professional development as well as informal (self-organized) activities. Mathematics teacher education (MTE) aims at improving teachers' beliefs, knowledge, and practice and contributes to affective and cognitive growth of the students. It is a challenge to find answers to the questions of where, under what conditions, how, and why mathematics teachers learn, and how important the domain-specific character of mathematics is. The major sources of this article are recent analyses on research in the field of MTE.

Mathematics Teacher Education: A Practical and Emerging Field of Research

In the last 20 years, researchers in mathematics education have increased attention on mathematics teacher education (MTE). This shift is reflected in the emergence of international handbooks. In 1996, the first *International Handbook of Mathematics Education* (Bishop *et al.*, 1996) was published. The first *International Handbook of Mathematics Teacher Education* (Wood *et al.*, 2008) has been published in 2008. Till about 1990, MTE was mainly a field of practice (mostly presenting success stories, more or less grounded on the basis of theory and evidence.). Since then it also increasingly became a field of research. In 1998, the *Journal of Mathematics Teacher Education* (JMTE) was launched. This journal marked the beginning of a new era since this was the first international journal focusing genuinely on research in MTE. It was only 4 years earlier that the founding editor of JMTE stressed that "as a profession we had just begun to recognize the significance of conceptual orientations for guiding research on teacher education (Cooney, 1994)." Simultaneously, numerous books and articles on the problems and progress in describing and interpreting learning processes by mathematics

teachers were also published (see e.g., Jaworski *et al.*, 1999; Krainer *et al.*, 1999; Lin and Cooney, 2001; Peter-Koop *et al.*, 2003; Sowder, 2007).

Common Trends in MTE and Related Research

There are three common trends that can be sifted out of the literature on learning in mathematics of student teachers, teachers, and teacher educators over the last 20 years (see Llinares and Krainer, 2006). These three trends refer to the intervention, goals and designs of teacher-education activities, as well as to the research focus on the processes and results of these activities. With regard to two trends, namely (1) teacher educators' and researchers' increasing attention to the social dimension and (2) attention to teachers' reflections, there is sufficient evidence including examples that indicate the variety of corresponding activities. The third trend, namely the increasing attention to the general conditions of teacher education (e.g., time, structure, institutional settings, and human resources), is newer and can be seen as an influence of work done on the practice and research in MTE in other fields, for example, organizational development.

Social Dimension

For a long time, MTE has been seen as the training of individual student teachers and teachers. They were basically regarded as receivers of mathematical and other kinds of knowledge. Teacher education was seen as an intervention process predominated by teacher educators, the (student) teachers were expected to listen and to put the learned into practice. Research was mainly investigating individual teachers' beliefs, knowledge, and practice, in many cases assessing whether these beliefs, knowledge, and practices were regarded as sufficient. When research on mathematics teaching increasingly showed the complexity of this process, at first the role ascribed to mathematics teachers changed. They were, for example, expected to put more emphasis on supporting students' joint construction of meaning and to reflect the didactic contract between the teacher and the students. Teacher educators' reflections on their own practice and voices of experienced teachers increased the demand that

teacher educators themselves should represent a new role model for teachers. Indicators for that movement were the emergence and usage of new theories that go beyond cognitive views on learning. In MTE, several steps toward explaining teachers' learning in theoretical models have been taken. Since teachers' learning is a complex process in which multiple factors intervene, a diversity of models focusing on particular aspects of teachers' learning is more likely to achieve than an overarching theory. Early models were mainly taken or modified from psychology, and later from sociology (e.g., Wenger's communities of practice). On the other hand, various approaches from mathematics educators have been developed (e.g., teaching triad; four dimensions of teachers' professional practice; four analytic domains and mediating processes; epistemic approach to teaching-learning process; and three-layer model). All these models take advantage of approaches and theories developed in other contexts and disciplines, like constructivist theories, organizational theory, or system theory.

Overall, this caused a qualitative shift of attention in practice and research toward putting a greater emphasis on the social dimension in teacher education. Increasingly, papers in teacher education began to refer to school development and some sorts of communities, for example, teacher-inquiry groups, study groups, communities of practice, or networks of critical friends. All these social systems are intended to be a good basis for facilitating joint reflection and to lead to an improvement of practice. For example, a study shows that the formation of site-based professional communities worked well in one school but poorly in the other. The decisive factors, among others, were teachers' (pedagogical content) knowledge and their expectations about students' learning.

The shift to the social dimension is mirrored in research literature. In the conference proceedings of the *Psychology of Mathematics Education* (PME) between the periods 1990–95 and 1996–2001 the percentage of studies that refer to the sociological and sociocultural theories increased from 3% to 10% (see Lerman and Tsatsaroni, 2004). This trend is similar in mathematics education journals. Notions like collaborative learning, organizational learning, sharing knowledge, designing didactic contracts, negotiating norms, systemic thinking, etc., influenced the thinking and writing of mathematics (teacher) educators. To some extent, the focus on collaborative learning, sharing knowledge, etc., has been built both on a certain trust in the ability of teachers to reflect, observe, and argue, and on the conviction that these competencies are central to the development of teachers. Based on these principles, the view of teacher education cannot be confined to activities that aim at individual teacher's growth, it has also be linked to a kind of learning in communities, including more formal kinds of

school development as well as participation in informal networks.

In this context, new information and communication technologies are being used to support both the social interaction among mathematics (student) teachers and the emergence of communities. New technological tools (e.g., video papers and blogs) make it possible to extend the boundaries of the class and can remove the boundaries between preservice education and professional development. New questions arise, for both teacher education as a field of practice and research: How are new forms of discourse and new forms of participation operated? How do they mediate the construction of meaning? How can communities be created and sustained? And how can socialcultural perspectives in teacher education be combined with other approaches?

Teachers' Reflections

Partially connected with the social shift mentioned above, we can see a stronger emphasis on the reflection of teachers as a second trend in MTE. Reflection is increasingly considered a key element in the development of processes required for ongoing learning, since it is a means by which teachers continue learning about teaching and about themselves as professionals. Often teachers are involved in teacher-education activities where they reflect on the mathematical thinking processes of students or on the reflections of their own learning (e.g., in contexts where they do mathematical activities themselves or in discussions about their practice and beliefs). Here, it is assumed that the attitudes, skills, and knowledge (student) teachers need are generated when they treat the knowledge and theory produced by others as generative material (points of reference) for interpretation of mathematics teaching, students' mathematical thinking, and the contextual conditions of teaching.

In addition, teachers' reflection can be used as a key element in investigating the growth of the teachers. For example, a focus in this type of research examines what teachers notice when observing their classes, the interpretations they give of events and the changes they propose in their practice. The findings from these studies reveal that differences in the beginning teachers' reflection were linked to changes in practice.

Other studies investigate the changing views of teachers on the students' learning processes and outcomes. For example, a researcher studied teachers' expectations on how secondary students verify conjectures in geometry. The mathematics teachers tended to underestimate students' reasoning on a four-level scale. However, when the same teachers were demonstrated typical examples of the reasoning of students (thus getting a better feeling for students' reasoning levels), their assessment improved. This shows that theory-based models (here for levels of

students' thinking) can not only be used as research instruments but also as a means to involve teachers in meaningful reflections and discussions.

In general, research reports show that teachers' learning is not only promoted by meaningful activities, but also by reflections on these activities. In many cases, teachers' reflections play a double role: they aim at increasing the teachers' understanding (of mathematics, of students' mathematical thinking, of institutional constraints on teaching, etc.), and at the same time they are used by researchers as means to describe and interpret teachers' learning.

Since about 10 years, increasingly interactive multimedia learning environments were used to support (student) teachers' learning about teaching mathematics and to study their growth when analyzing teaching events. Research instruments influence practice, and practice itself is the place where new instruments are generated. In particular, supporting teachers to write down experiences in a systematic and self-critical way is an intervention toward an investigative attitude which is the entrance card to participation in a research community and culture. Nowadays, the new forms of discourse generated by the currently available communication tools such as video paper, blogs, bulletin boards, and so on, use writing as an instrument for collaborative reflection and as a tool for inquiry. The new type of text generated by these communication tools can be used as an improvable object that favors the generation of a progressive discourse and as one that acts as the focus of collaborative knowledge building. For example, the process of creating a video paper – as a multimedia document that integrates and synchronizes different forms of representation, including text, video, and images, in one single nonlinear cohesive document – supports teacher reflection. The same is true for writing in order to participate in online discussions as a means to notice mathematics teaching. These processes of writing are qualitatively different from simply watching a video from a whole lesson to reflect on one's practice: when mathematics (student) teachers' have to select relevant clips or particular events from a videotaped lesson to discuss, their reflection is fostered by initiating a more analytical process that might not necessarily happen if they had just watched a whole lesson without needing to produce a written text. In this sense, for example, the multimodal character of video papers (combining video and text) enriches reflection and thus also the process of meaning-making and knowledge construction.

In combination with the assumption that teachers are key persons of educational change, it makes sense to regard them as reflective practitioners, researchers, and experts. Recently papers on teacher educators' growth have been written indicating that mathematics teacher educators' learning and their self-reflection is an important issue as well.

General Conditions

In particular, with regard to practicing teachers' learning, a third trend in the research on MTE can be recently observed, namely a stronger awareness of the general conditions on which teachers work (and thus also teacher educators' interventions) is based. Presumably, influenced by the closer look at the social context of teachers' learning and the particular focus on their reflections, researchers get more evidence and thus also look more deeply at the resources and the support that teachers get (or do not get) when they participate in teacher-education activities. Some studies indicate factors concerning mathematics teachers' working conditions that supported or constrained teachers' work. In particular, relationships of mathematics teachers with the school administration and other teachers or the presence of a teacher leader were decisive factors that influenced the development of teachers' professional communities at the school sites. At best, general conditions at the school and at the district level support the formation of teacher communities, and expert teachers out of these initiatives (communities) are respected as teacher leaders and eventually earn official roles in the district. However, in many cases, informal teacher communities are not interconnected with formally appointed groups, committees, and task forces: there exists a gap between a school or district as a designed organization (consisting of formally designated roles and divisions of labor together with official policies, procedures, routines, management systems, organizational units, etc.) and as a lived organization (comprising the groups within which work is actually accomplished together with the interconnections between them).

Consequences for the Field

It is worth taking the three trends seriously and regarding them as the challenges for the future. First, looking at MTE as a field of practice, it makes sense to critically reflect the balances of the individual and the social (dimensions of teaching and learning), of action and reflection (in and on practice), and of internal and external (resources and support); it is equally important to concentrate on the specific challenges of mathematics education, for example, the insufficient image of mathematics and mathematics teaching or the change that new means (e.g., technology) bring to it.

Second, more research is needed in MTE. Many studies on these phenomena so far used qualitative research methods, and it makes sense to go further there in order to generate new explanations and assumptions. It is desirable to use the synergy of teachers' expertise and therefore to engage them in research activities and to support research action, among others, with the goal that some of them might develop deeper interest in research and will enlarge

the scientific community. Most teachers do not read research papers; however, being involved in such projects as mentioned above, bridges the gaps for some teachers. This might be the starting point for cooperative intervention research balancing the development interest (achieving an improvement through the intervention) and the research interest (gaining understanding with regard to the specific case, eventually generating more general scientific knowledge). However, also more external and quantitative research are needed, in particular, looking at the outcomes of different types of teacher education or at longitudinal studies of mathematics teachers' learning and career. In all these cases, large populations are necessary to test relevant hypotheses. International comparative studies such as IEA teacher education and development study in mathematics (TEDS-M; see e.g., Tatto *et al.*, 2007) will certainly bring interesting results and new challenges for teacher education. Similar to studies on student assessment, researchers began to create competence models in order to describe different kinds of knowledge of teachers and to use it in research in mathematics (and science) teacher education. Overall, there is a future challenge to combine qualitative and quantitative research methods and to integrate systematic reflections of teachers into research projects (see Adler *et al.*, 2005).

Third, the fact that mathematics is a major subject at schools and in comparative studies can be taken as a chance to see MTE as a collaborative endeavor of researchers, practitioners, and educational policymakers. For example, as a reaction to weak results in comparative studies like TIMSS and PISA, some countries started nationwide teacher-education initiatives in order to improve the situation. Since teachers and teacher educators are important change-agents of this kind of reform, these initiatives are well advised to take into account the ideas, needs, and wishes of the scientific community and the educational policy, and equally those of teachers and teacher educators, which are nearest to the place where the steps of change are taken and given its concrete form.

Teacher education is best done in safe environments where it is possible for teachers to speak out their fears and problems. Laying blame on teachers' wrong beliefs, insufficient knowledge, and bad practice is not promising until teacher educators and researchers know the general solutions to all particular problems. This is an unreachable and undesirable vision since all sciences including mathematics are socially constructed and principally fallible. However, researchers (in mathematics education) can and should make suggestions that are grounded in theoretical considerations and based on evidence. They should aim at presenting (written, oral, video-based, etc.) starting points for discussions which serve not only new scientific insights, but also new steering knowledge for educational policies and new practice-relevant proposals for teachers and teacher educators.

Bibliography

- Adler, J., Ball, D., Krainer, K., Lin, F.-L., and Novotna, J. (2005). Mirror images of an emerging field: Researching mathematics teacher education. *Educational Studies in Mathematics* **60**(3), 359–381.
- Bishop, A., Clements, K., Keitel, C., Kilpatrick, J., and Laborde, C. (eds.) (1996). *International Handbook of Mathematics Education*. Dordrecht: Kluwer.
- Cooney, T. (1994). Research and teacher education In search of common ground developed. *Journal for Research in Mathematics Education* **25**, 608–636.
- Even, R. and Ball, D. L. (eds.) (2008). *The Professional Education and Development of Teachers of Mathematics. The Fifteenth ICMI Study*. New York: Springer.
- Jaworski, B., Wood, T., and Dawson, S. (eds.) (1999). *Mathematics Teacher Education. Critical International Perspectives*. London: Falmer Press.
- Krainer, K., Goffree, F., and Berger, P. (eds.) (1999). *On Research in Mathematics Teacher Education. European Research in Mathematics Education I. III*. Osnabrück: Forschungsinstitut für Mathematikdidaktik.
- Lerman, S. and Tsatsaroni, A. (2004). *Surveying the Field of Mathematics Education Research. Regular Lecture at ICME-10*. Denmark: Copenhagen. <http://www.icme-organisers.dk/dg10/Leemanpaper.pdf>.
- Lin, F.-L. and Cooney, T. (eds.) (2001). *Making Sense of Mathematics Teacher Education*. Dordrecht: Kluwer.
- Llinares, S. and Krainer, K. (2006). Mathematics (student) teachers and teacher educators as learners. Gutiérrez, A. and Boero, P. (eds.) *Handbook of Research on the Psychology of Mathematics Education. Past, Present and Future*, pp 429–459. Rotterdam: Sense Publishers.
- Peter-Koop, A., Begg, A., Breen, C., and Santos-Wagner, V. (eds.) (2003). *Collaboration in Teacher Education. Examples from the Context of Mathematics Education*. Dordrecht: Kluwer.
- Sowder, J. (2007). The mathematical education and development of teachers. Lester, F. (ed.) *Second Handbook of Research on Mathematics Teaching and Learning*, pp 157–223. Greenwich, CT: NCTM.
- Tatto, M. T., Schwille, J., Senk, S., et al. (2007). *IEA Teacher Education and Development Study in Mathematics (TEDS-M). Conceptual Framework. Policy, Practice and Readiness to Teach Primary and Secondary Mathematics. Field Trial*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA) in collaboration with Michigan State University and the Australian Council for Educational Research.
- Wood, T., (series ed.) Jaworski, B., Krainer, K., Sullivan, P., and Tirosch, D. (vol. eds.) (2008). *International Handbook of Mathematics Teacher Education*. Rotterdam: Sense Publishers.

Further Reading

- Cooney, T. and Wiegel, H. (2003). Examining the mathematics in mathematics teacher education. Bishop, A. J., Clements, M. A., Keitel, C., Kilpatrick, J., and Leung, F. K. S. (eds.) *Second International Handbook of Mathematics Education*, pp 795–822. Dordrecht: Kluwer.
- Jaworski, B. and Gellert, U. (2003). Educating new mathematics teachers: Integrating theory and practice, and the roles of practicing teachers. Bishop, A. J., Clements, M. A., Keitel, C., Kilpatrick, J., and Leung, F. K. S. (eds.) *Second International Handbook of Mathematics Education*, pp 823–876. Dordrecht: Kluwer.
- Zaslavsky, O., Chapman, O., and Leikin, R. (2003). Professional development in mathematics education: Trends and tasks. Bishop, A. J., Clements, M. A., Keitel, C., Kilpatrick, J., and Leung, F. K. S. (eds.) *Second International Handbook of Mathematics Education*, pp 875–915. Dordrecht: Kluwer.

Music Teacher Education

M W-y Wong, Hong Kong Baptist University, Hong Kong, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

This article introduces various systems of music teacher education in some selected countries and discusses the current challenges for music teacher education.

Music Teachers and Music Teacher Education Systems

There are diverse types of music teachers in the field of music education. Some are employed by schools to teach as specialist music teachers, generalist music teachers, band directors, choir masters, and instrumental teachers; some are employed by music companies or music studios; and some are self-employed as instrumental or music theory teachers for individual students or small groups.

As there are diversities in the types of music teachers, economic development, and social and cultural contexts around the world, there are diverse types of institutionalized music teacher education programs that developed in different parts of the world, ranging from those with no specific preparation at all to postgraduate education at universities. Music teachers who teach in private music studios or music companies usually are proficient musicians, but they are not required to have formal music teacher education or pedagogical training. There are usually some government regulations on the basic requirements of the qualifications of school music teachers. In some countries, basic education for all tends to include completion of high school, and teachers are required to have a university education. In some countries, only senior high school teachers are required to hold university degrees, while primary and junior high school teachers usually hold subdegree qualifications. The following sections describe the formal music teacher education for school music teachers available in selected countries. Further information about music teacher education of these countries can be found in the websites of teacher education institutions and government organizations provided at the end of this article.

Africa

In Africa, structures of music teacher education systems vary from country to country. In South Africa, music teacher education programs are offered by universities and colleges of education. Various kinds of qualifications, such as diplomas, bachelor degrees, and postgraduate degrees are granted to graduates of teacher education

programs with duration ranging from 1–3 years. In the Republic of Botswana, a music teacher education course is only offered at subdegree diploma level by one of the six government-operated colleges of education that trains teachers for schools. Information about music teacher education is lacking in the official websites of governments and universities of other African countries.

Asia

In Asia, music teacher education programs are provided by various education institutions and universities. In Malaysia and some places of the Peoples' Republic of China, music teacher education programs are offered as 3-year nondegree programs for secondary school graduates by colleges of education, normal schools, and conservatories. In Japan, Korea, Philippines, Singapore, and some places of the People's Republic of China, music teacher education programs are offered as 4-year degree programs and 1-year postdegree diplomas by colleges of education, normal universities, and comprehensive universities with music education as specialization. Prospective teachers of these programs have to complete practice teaching in schools. The duration of practice teaching differs between countries, and lasts from 1 month to 1 year.

In some countries, such as Korea, Malaysia, and Singapore, music teacher candidates are selected by their government. The music teacher trainees of Malaysia and Singapore are sponsored by their governments and receive a monthly salary throughout their period of music teacher training, and are assigned to serve in government schools upon completion of the music teacher education programs. In the Philippines, graduates of professional education diploma programs specializing in music education are eligible to take the Licensure Examination for Teachers, which qualifies them to teach in schools.

Europe

According to the European Forum for Music Education and Training (2005), there are various institutions providing music teacher education in the countries of the European Union. These institutions include conservatoires, *musikbochschulen*, *académies de musique*, music departments and education departments in general universities, and teacher training colleges that train music teachers for schools and instrumental teachers and vocal

teachers at undergraduate and postgraduate level. Due to the reforms in European higher education, music teacher education programs have been undergoing changes since 2000. The duration of these programs ranges from 3 to 6 years. Pedagogical subjects are offered to performing music students as well as music teacher education students because most music graduates may teach. Teacher candidates have to complete teaching practice in school. Continuing professional development programs, with the duration ranging from 4 months to 3 years for music teachers, are offered by higher education institutions. At the postgraduate level, higher degrees in music teacher education are also available.

North America

In Canada, music teacher education courses are offered as part of initial teacher education programs which are offered by universities. Professional development programs for in-service teachers are offered by universities as graduate studies, or sponsored by provincial and territorial governments, teacher associations, and school boards when new curriculum or policies are implemented (Connelly and Clandinin, 2004). Teacher re-certification programs are also available for teachers who are new to the province or have left the profession for a period of time and do not have recent teaching experience. The required qualification for music teachers in different provinces varies.

In the United States, colleges and universities offer different models of music teacher education programs. There are 4-year undergraduate degree programs, 5-year integrated degree programs, and 1-year postgraduate degree programs that lead to teacher certification. The requirements of teacher certification for music teachers vary from state to state. In-service professional development programs and higher degrees in music teacher education are also available in some universities (Nierman *et al.*, 2002).

Southern Hemisphere

In Australia, music teacher education programs are offered by universities and colleges at undergraduate and postgraduate level. Teacher candidates have to complete 45–80 days of teaching practice in schools. Three-year and 4-year full-time undergraduate degree programs, 4-year and 5-year full-time double degree programs, and 1-year full-time postgraduate degree programs are available. Graduates of these courses are qualified for registration as specialist music teachers in schools. In-service professional development courses are offered by universities and private institutions.

In New Zealand, music teacher education programs are offered by universities at undergraduate level. Graduates

who study a 4-year full-time bachelor degree in music education with teaching practice in schools are qualified to teach music in schools.

Music Teacher Education Curricula

The components of music teacher education curriculum can be categorized under the headings of methodology and pedagogy of music teaching. According to traditional trends of music teacher training, some music educators regard subject-matter knowledge of music and practical skills of instrumental playing as the most important knowledge and skills that music teachers should possess; while some educators believe that pedagogical skills are more important. However, there is an increasing emphasis on inquiry-based learning and expecting music teachers to develop the ability to conduct classroom research and practice reflective instruction (Leglar and Collay, 2002). Research methods and microteaching have become essential elements in music teacher education programs.

Challenges of Music Teacher Education

Education Reforms

Following the trends of education reforms around the world, the development of teacher education has become a concern in terms of policy and practice. There is a need to restructure teacher education systems, arrange provision required for facilitating teachers to implement change in curriculum, and provide opportunities for professional development of teachers. These concerns apply to music teacher education as well.

Connecting music education with arts education is one of the current trends in curriculum reform. In some countries, music has been submerged in the arts learning area. Music teachers may need to broaden their knowledge in other art forms. Therefore, preservice and in-service music education students should be nurtured with cross-curricular attitude to help them succeed in the music curriculum of the new century (Kimpton, 2005). Music teacher education programs may have to provide more general knowledge on arts subjects to cater to the changes in school music curriculum. Some music teacher educators indicate that music teachers may lack sufficient musical knowledge and skills to teach music if the time allocated for music education is reduced because of adding more arts subjects into music teacher programs (Pascoe *et al.*, 2005). In order to support cross-disciplinary studies and curriculum integration, music teacher education programs have to take the lead to support collaborative works with teachers and teacher educators of other disciplines in widening the worldview of music teacher education candidates, and provide opportunities for students to

experience holistic learning beyond the traditional framework of subject-based curriculum.

Reform in assessment has been coupled with education reform in many countries of the world in recent decades. Many educators have advocated the use of outcome-based assessment which measures learners' success by their individual achievement (Boschee and Baron, 1994), alternative assessment, formative assessment, or assessment for learning that could better reflect students' learning process (Wiggins, 1998; Black *et al.*, 2004). Music educators also advocate the use of diverse types of assessment, and hold positive views about using nontraditional assessment, such as teacher-created assessment rubrics, student-created assessment rubrics, self-assessment, peer assessment, portfolio assessment, and assessment of listening activities to facilitate learning to help students learn better and to assess students' learning processes and learning outcomes (Asmus, 1999; Cantwell and Jeanneret, 2004; Cavner and Gould, 2003; Chiodo, 2001; Eppink, 2002). These changes in music assessment rely more on teacher-constructed assessment. However, the concept of formative assessment or assessment for learning may be rather new to music educators as well as music teacher educators. Therefore, equipping music teacher education students with the knowledge and ability to design and implement assessment reform in music education is going to be one of the major challenges for music teacher education programs (Colwell, 2002).

To foster student's creativity through creative music making in classroom instruction is another common advocacy in curriculum reform. Many music teachers are performers instead of composers. They may not be familiar with creative work in classroom and express hesitation in implementing music-composing activities (Orman, 2002). Therefore, music teacher education programs may need to provide knowledge and skills to help music education students and in-service music teachers to develop their interests, enthusiasm, and confidence in implementing creative music work in the classroom.

Curriculum reform has caused great frustrations to music teachers because of the heavy workload required and the enormous change of curriculum and assessment (Cox, 1999; Fallis, 1999). Since most music teachers are trained musicians before they enter the teaching profession, they may not be aware of the issues in school education and music education. In addition to developing music teachers with an open and innovative mindset to cope with the challenge of dealing with educational changes, it is necessary for music teacher education programs to provide music teacher education students with the opportunities to familiarize and socialize themselves with the reality and expectations of the workplace, and enhance their understanding of the position of music among various disciplines in school education, as well as

strike a healthy balance between their identities as musicians and music teachers.

Professional Partnership

Collaborative professional development partnership is considered to bring about benefit to all parties that are involved: it provides valuable experience for (1) collaborating music teachers to learn about new techniques and materials; (2) music teacher educators to have the opportunity to keep up with the current practice in music classrooms; and (3) music teacher candidates to relate theory and practice in classroom music teaching (Morin, 2000; Conkling and Henry, 2002).

There are various forms of partnerships, such as integrating academic and professional preparation for beginning teachers with professional development schools and partnership programs (Henry, 2001). Collaboration among teachers to offer beginning-teacher induction and mentor programs could provide more support for preservice teachers and beginning teachers so that they feel more confident and prepared (Conway, 2001; Burton, 2004). Building partnership networks for in-service teachers could benefit novice and expert teachers as well as music teacher educators through professional-development workshops and site-based collaborative research (Robbins and Stein, 2005). Artists in residence programs have been a common strategy of partnership program for schools and artistic agencies. Such programs could help music teachers provide a wide range of musical experiences for students because they may not be expert in all kinds of music. Interdisciplinary curriculum is another common strategy for building professional partnerships. Planning and implementing of interdisciplinary curriculum requires music teachers and teachers of other subjects to work collaboratively. However, music teachers may not have the knowledge and experience to administer these collaborative learning programs or work with experts in other disciplines. Therefore, it would be another challenge for music teacher education programs to equip music teacher education students with the necessary experiences in developing collaborative professional partnerships.

Diverse Musical Cultures and World Music

Due to colonialism in the nineteenth century, music of Western cultures has been dominating the school music curriculum in many countries. The advocacies of teaching music with a multicultural approach, a pan-cultural approach, and global perspectives have been discussed among music educators of all levels since the mid-twentieth century (Campbell, 2004; Walker, 1990; Leith-Phillipp, 1995; Reimer, 2002). In the ever-increasingly global society with music of Western culture as the major stream in music

programs and music teacher education programs, music teachers are facing the challenge of broadening students' musical and cultural perspectives through appreciation of non-Western musical styles on the one hand, and the lack of relevant knowledge and resources about non-Western musical styles on the other. Since appreciating music of a culture includes the experiences of listening to the sound and understanding music as one of the ingredients that contribute to culture, it is necessary for music teacher education programs to provide music teacher education students with the experience, knowledge, and skills in guiding students to appreciate different musical cultures around the world.

Mass Media

The younger generations around the world are brought up in an environment with much stimulation from mass media. Developing students' competencies with new media forms could be seen as a way for them to express themselves (Buckingham, 2003) and also a challenge for music educators. Students are more easily attracted by computerized audio-visual equipments such as mobile phones with all kinds of audio-visual recording functions, Wi-fi, iPod, Karaoke, etc. that play popular music promoted by the mass media. Making use of the music from mass media may attract more attention from students and fill up certain culture and generation gaps between music teachers and their students. Music teachers who are educated through traditional music teacher education programs may need to equip themselves with the necessary knowledge and skills so as to guide students to apply musical knowledge and skills in producing creative works in new media forms. It would be a challenge for music teacher education programs to develop music teacher education students' sense of lifelong learning and provide professional development programs for in-service teachers to refresh their knowledge and skills in making use of mass media in music teaching.

Computer Technology

Computer technology has become an indispensable part of peoples' lives in the twenty-first century. The younger generations are increasingly familiar with the application of computer technology, which has been advocated in the school music curriculum under the current education reforms around the world. Computer technology has become a valuable resource for music teachers who wish to broaden the musical experience of students.

The Internet offers a library of references for students to experience a wide variety of music on many websites. Music teachers may use the Internet as an additional method of enhancing and supplementing classroom

music activities (Bauer and Daugherty, 2001). The musical functions of computer technology include the opportunities for exploring and creating music. Music technology could be a way to lower barriers for students to notate musical compositions with traditional notation skills and the need for performing skills (Reese, 1995), and computer-assisted music composition could facilitate students develop their music creativity (Reese and Hickey, 1999). With the use of computer technology, music teachers could apply computer technology to facilitate lesson planning and daily classroom instruction, such as music printing, music sequencing, multimedia authoring, and creating personal homepages with sound files (Walls, 2000), as well as producing interesting audio-visual aids to facilitate students in appreciating music with holistic experiences of human senses. However, research studies found that music teachers might encounter difficulties and change in pedagogical practice when using computer technology in classroom teaching (John, 2005). As computer technology is developing at a speedy pace, incorporating technology training as a part of teacher education curriculum is becoming essential. In order to help music teachers better equip themselves with adequate knowledge and skills in the application of computer technology in music teaching and learning, music teacher education programs should consider providing preservice and in-service music teachers with the knowledge and skills for processing and editing computerized audio and visual files for promoting music appreciation; performing as well as composing activities; developing music teacher education students' interests and ability to explore music resources that are available from the Internet; and applying computer technology in all phases of music production, such as, music creating, recording, editing, mixing, and sharing.

Supply of Music Teachers

Music education in school relies on quantity and quality of teacher supply. Though parents usually have high expectations of music education for their children, the *status quo* of music teacher education may not satisfy their expectation. Comparatively speaking, the monetary returns and the workplace conditions of school music teachers are not as attractive as that of teachers who run private music studios; the poorer living conditions of school music teachers can hardly attract able and talented students to pursue music teacher education and this results in a shortage of music teachers (Asmus, 1999; Wang, 2004; Lee, 2004; Kim, 2001). Many new music teachers never enter the profession after their student-teaching experience and many of them drop out of the profession during their first few years of teaching (Kimpton, 2005). There are many reasons why capable

music students do not enter the teaching profession. Difficult working conditions, low salary, low priority of music education within the school curriculum, and the conflict of dual identities of being a musician and a music teacher could cause job dissatisfaction and occupational stress that result in music teacher attrition (Scheib, 2004; Scheib, 2006). In order to help music teachers lessen their job dissatisfaction and occupational stress, there is a need for preservice music teachers to foster their identity of being a teacher during their university-level education and to convince school administrators to support the professional development of music teachers as musicians as well.

Implications for Music Teacher Education

In response to the challenges mentioned above, music teacher education should develop in the direction of facilitating music education students pursue lifelong learning in the following areas:

- apply computer technology in classroom music instruction;
- reflect on the structures of music education in schools and socialize according to the reality and expectation of the workplace, so as to balance the identities between musician and music teacher;
- develop reform-mindedness to cope with the challenge of educational change in the movement of school curriculum reform;
- design a well-balanced curriculum to develop the musical abilities of students and cater to students' learning needs as well as their holistic growth;
- develop the ability to employ a variety of strategies to evaluate students' learning processes and learning outcomes so as to find directions for future development of a music curriculum; and
- support innovative and collaborative endeavors, such as, collaborating with teachers of other disciplines in providing opportunities for students to experience holistic learning beyond the traditional framework of subject-based curriculum.

Bibliography

- Asmus, E. (1999). The increasing demand for music teachers. *Journal of Music Teacher Education* 8(2), 5–6.
- Bauer, W. I. and Daugherty, J. F. (2001). Using the Internet to enhance music teacher education. *Journal of Music Teacher Education* 11(1), 27–33.
- Black, P., Harrison, C., Lee, C., Marshall, B., and William, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan* 86(1), 9–21.
- Boschee, F. and Baron, M. A. (1994). Outcome-based education: Some answers for the uninitiated. *Clearing House* 67(4), 193–196.
- Buckingham, D. (2003). *Media Education: Literacy, Learning and Contemporary Culture*. Cambridge: Polity.
- Burton, S. L. (2004). Where do we begin with inquiry-based degree programs? *Journal of Music Teacher Education* 14(1), 27–33.
- Campbell, P. S. (2004). *Teaching Music Globally: Experiencing Music, Expressing Culture*. New York: Oxford University Press.
- Cantwell, R. H. and Jeanneret, N. (2004). Developing a framework for the assessment of musical learning: Resolving the dilemma of the “parts” and the “whole”. *Research Studies in Music Education* 22, 2–13.
- Cavner, D. and Gould, E. (2003). Whole language and music-listening instruction. *Music Educators Journal* 89(5), 19–24.
- Chiodo, P. (2001). Assessing a cast of thousands. *Music Educators Journal* 87(6), 17–23.
- Colwell, R. (2002). Assessment's potential in music education. In Colwell, R. and Richardson, C. (eds.) *The New Handbook of Research on Music Teaching and Learning*, pp 1128–1158. Oxford, UK: Oxford University Press.
- Conkling, S. W. and Henry, W. (2002). The impact of professional development partnerships: Our part of the story. *Journal of Music Teacher Education* 11(2), 7–13.
- Connelly, F. M. and Clandinin, D. J. (2004). Canadian teacher education in transform. In Cheng, Y. C., Chow, K. W., and Mok, M. C. M. (eds.) *Reform of Teacher Education in the Asia-Pacific in the New Millennium*, pp 35–43. Dordrecht: Kluwer Academic Publishers.
- Conway, C. (2001). What has research told us about the beginning music teacher? *Journal of Music Teacher Education* 10(2), 14–22.
- Cox, G. (1999). Secondary school music teachers talking. *Music Education Research* 1(1), 37–45.
- Eppink, J. A. (2002). Student-created rubrics. *Teaching Music* 9(4), 28–32.
- European Forum for Music Education and Training (2005). Research study on music teacher training programmes in Europe. *Report of the European Forum for Music Education and Training*. Bonn, Germany: European Music Council.
- Fallis, T. L. (1999). Standards-based instruction in rehearsal. *Music Educators Journal* 85(4), 18–24.
- Henry, W. (2001). Music teacher education and the professional development school. *Journal of Music Teacher Education* 10(2), 23–29.
- John, P. (2005). The sacred and the profane: Subject sub-culture, pedagogical practice and teachers' perceptions of the classroom uses of ICT. *Educational Review* 57(4), 472–481.
- Kim, H. S. (2001). Korean vision of teacher reform: An attempt for quality control and empowerment. *A paper presented at the First International Forum on Education Reform*.
- Kimpton, J. (2005). What to do about music teacher education: Our profession at a crossroads. *Journal of Music Teacher Education* 14(2), 8–21.
- Lee, M. N. N. (2004). Malaysian teacher education into the new century. In Cheng, Y. C., Chow, K. W., and Mok, M. C. M. (eds.) *Reform of Teacher Education in the Asia-Pacific in the New Millennium*, pp 81–91. Dordrecht: Kluwer Academic Publishers.
- Leglar, M. and Collay, M. (2002). Research by teachers on teacher education. In Colwell, R. and Richardson, C. (eds.) *The New Handbook of Research on Music Teaching and Learning*, pp 855–873. Oxford, UK: Oxford University Press.
- Leith-Phillipp, M. (1995). *Teaching Musics of the World*. Affalterbach, Germany: Philipp.
- Morin, F. (2000). Developing collaborative partnerships for student teaching in music. *Journal of Music Teacher Education* 10(1), 6–15.
- Nierman, G. E., Ziechner, K., and Hobbel, N. (2002). Changing concepts of teacher education. In Colwell, R. and Richardson, C. (eds.) *The New Handbook of Research on Music Teaching and Learning*, pp 818–839. Oxford, UK: Oxford University Press.
- Orman, E. K. (2002). Comparison of the national standards for music education and elementary music specialists' use of class time. *Journal of Research in Music Education* 50(2), 155–164.
- Pascoe, R., Leong, S., MacCallum, J., et al. (2005). *National Review of School Music Education*. Canberra: Australian Government.
- Reese, S. (1995). MIDI-assisted composing in your classroom. *Music Educators Journal* 81(4), 37–40.

- Reese, S. and Hickey, M. (1999). Internet-based music composition and music teacher education. *Journal of Music Teacher Education* 9(1), 25–32.
- Reimer, B. (2002). *World Musics and Music Education: Facing the Issues*. Reston, VA: MENC.
- Robbins, J. and Stein, R. (2005). What partnerships must we create, build, or reenergize in K-12, higher and professional educational for music teacher education in the future? *Journal of Music Teacher Education* 14(2), 22–29.
- Scheib, J. W. (2004). Why band directors leave: From the mouths of maestros. *Music Educators Journal* 91(1), 53–57.
- Scheib, J. W. (2006). Policy implications for teacher retention: Meeting the needs of the dual identities of arts educators. *Arts Education Policy Review* 107(6), 5–10.
- Walker, R. (1990). *Musical Beliefs*. New York: Teachers College.
- Walls, K. C. (2000). Technology for future music educators. *Journal of Music Teacher Education* 9(2), 14–21.
- Wang, Y. J. (2004). Patterns of development of Chinese teacher education in a reform context. In Cheng, Y. C., Chow, K. W., and Mok, M. C. M. (eds.) *Reform of Teacher Education in the Asia-Pacific in the New Millennium*, pp 63–79. Dordrecht: Kluwer Academic Publishers.
- Wiggins, G. P. (1998). *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco, CA: Jossey Bass.

Relevant Websites

<http://www.education.gov.za> – Department of Education, Republic of South Africa.

<http://www.emc-imc.org> – European Music Council.

<http://www.arts.auckland.ac.nz> – Faculty of Arts, University of Auckland.

<http://www.unimelb.edu.au> – Home page of the University of Melbourne, Victoria, Australia.

<http://www.isme.org> – International Society for Music Education.

<http://www.kunitachi.ac.jp> – Kunitachi College of Music.

<http://www.ipm.edu.mo> – Macau Polytechnic Institute.

<http://www.studymalaysia.com> – Malaysian Education Guide Online.

<http://www.moe.gov.sg> – Ministry of Education, Singapore.

<http://www.menc.org> – National Association for Music Education.

<http://www.music.usyd.edu.au> – Sydney Conservatorium of Music.

<http://www.u-gakugei.ac.jp> – Tokyo Gakugei University.

<http://www.up.ac.za> – University of Pretoria – University of Pretoria.

<http://www.cnu.edu.cn/english/index.asp> – Capital Normal University, Beijing, Peoples Republic of China.

<http://xs.ied.ecnu.edu.cn> – East China Normal University, Shanghai, Peoples Republic of China.

<http://www.fed.cuhk.edu.hk> – Faculty of Education, The Chinese University of Hong Kong, Hong Kong SAR, Peoples Republic of China.

<http://educ.hkbu.edu.hk> – Department of Education Studies, Hong Kong Baptist University, Hong Kong SAR, Peoples Republic of China.

<http://www.ied.edu.hk> – Hong Kong Institute of Education, Hong Kong SAR, Peoples Republic of China.

<http://www.mext.go.jp> – Ministry of Education, Lulhere, Sports, Science and Technology, Japan.

<http://www.teachernet.gov.hk> – Teacher Net, United Kingdom.

Science Teacher Education

J H Van Driel, ICLON-Leiden University Graduate School of Teaching, Leiden, The Netherlands

S K Abell, University of Missouri, Columbia, MO, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Research findings from the last 30 years indicate that students, regardless of nationality, have difficulties learning science (Osborne and Dillon, 2008). They harbor misconceptions about science concepts that are resistant to change by traditional teaching methods. They often fail to develop scientific thinking processes and habits of mind. Even worse, it has often been reported that students have a low appreciation of, and negative attitudes toward science. In many countries, they score poorly on standardized tests and choose not to enter pathways to scientific careers. The development and implementation of science education standards, new curriculum materials, and inquiry-based instructional strategies are often regarded as solutions to the problem of poor learning in science. However, the negative trends outlined above continue despite repeated attempts at science curriculum reform in countries around the globe.

At the same time, numerous studies have shown that the science teacher is the most important factor in improving student achievement (Hattie, 2003). Thus, the preparation of new science teachers is crucial to the future of science education. How to prepare high-quality science teachers who can help achieve the goals of science education reform is the topic of this article. The nature of science education has specific implications for the preparation of prospective science teachers. Science teachers need to learn to address students' preconceptions and misconceptions and to use specific conceptual-change strategies. Also, the experimental nature of science urges the use of laboratory work in science classes, which implies the need to develop specific teacher competences related to instructional strategies (e.g., guided inquiry teaching), as well as in the domain of safety issues.

We start this article by discussing what we know about the future teachers of science who enter science teacher preparation programs. Next, we discuss what the literature says about curriculum, instruction, and assessment in science teacher education. We conclude with implications for the continuing education of teachers of science.

A number of important constructs from the teacher education literature frame our thinking about science teacher preparation. First, Shulman's notion of pedagogical content knowledge (PCK) helps us think about what we want prospective science teachers to learn and what we as science teacher educators need to know to effectively

instruct them (Shulman, 1986). Second, a conceptual-change framework helps us attend to the knowledge and beliefs that students have when they enter science teacher preparation programs, and find ways in which science teacher educators can challenge and help students of teaching modify their ideas (Russell and Martin, 2007). Third, and related to the previous point, experiences in classrooms provide the basis for reflection, which drives the professional development of teachers (Schön, 1983).

Prospective Science Teachers' Knowledge and Beliefs

Teacher education programs, which focus exclusively on the preparation of science teachers, almost always aim at teaching science at the secondary level (students aged 12–18). For primary education, the preparation to teach science is normally a relatively small part within a broad curriculum aimed at learning to teach all subjects in the primary grades. In higher education, little, if any, preparation to teach is included in science doctoral programs. Recent views of science teacher education recognize the importance of diagnosing teachers' prior knowledge and beliefs about science teaching and learning as they enter teacher preparation programs and courses. Therefore, in this section, we summarize primarily what we know from research about students in secondary science teacher preparation programs, focusing on their knowledge of science content and the nature of science, their orientations to teaching science, and their emerging knowledge for teaching science (i.e., PCK).

Subject-Matter Knowledge

Research in science teacher education has identified problems concerning the subject-matter knowledge of future teachers, including their understanding of science concepts as well as their views of the nature of science as a discipline. In particular, it appears that prospective secondary science teachers, even those with a more extensive science background, have serious weaknesses in their science content knowledge (e.g., misconceptions about science topics which in many cases are very similar to those of younger students). Although the relationship between science teachers' content knowledge and their teaching practice is contested, there is ample evidence in the

research literature that science teachers with a weak content background tend to avoid certain topics, or apply mainly didactic approaches when teaching these topics, sticking closely to their textbooks and asking lower-level questions (e.g., Hashweh, 1987).

Science teacher educators believe that prospective teachers need to know something about the epistemology of science, or what is called the nature of science. Abd-El-Khalick *et al.* (1998) proposed five aspects of the nature of science that should be addressed in science teacher preparation, namely, that scientific knowledge is: (1) tentative and subject to change; (2) empirically based on and/or derived from observations of the natural world; (3) subjective and/or theory-laden; (4) partly the product of human inference, imagination, and creativity involving the invention of explanations; and (5) embedded in social, cultural, and historical practices. The research on prospective teachers' knowledge of the nature of science has demonstrated inadequate nature of science conceptions and practices. Many researchers have pointed out that prospective science teachers hold naive and positivist conceptions of what science is, and how scientific knowledge is developed, for instance, believing that the substantive content of science is fixed and unchangeable rather than tentative (e.g., Lederman, 1992). This view seems related to the belief that scientific knowledge can be transmitted to learners in a straightforward and nonproblematic way, thus conflicting with the well-accepted constructivist idea that learners actively construct knowledge themselves, based on their previous ideas and experiences.

Orientations toward Teaching Science

Orientations toward teaching science refer to teachers' knowledge and beliefs about the purposes and goals of science education for a specific target group. According to Magnusson *et al.* (1999), these orientations inform teachers' instructional decisions, and therefore, to a large extent, determine teachers' choices of textbooks, instructional strategies, and ways of evaluation. This is probably also true for teachers of other subjects.

Research has pointed out repeatedly that students enter science teacher preparation with specific ideas about teaching and learning science (e.g., Loughran, 2007). These orientations have been developed over many years of experiencing science teaching as learners, from primary school through higher education. Being successful learners of science themselves, they often tend to teach science in the way they were taught. This starting point implies various problems:

- The belief that the way they were taught science should be used as a model for their own teaching, since it worked quite well for them, may stimulate the use of traditional approaches to teaching science (e.g., didactic

or activity-driven; Magnusson *et al.*, 1999), which are problematic when science for all or science for public understanding is an important goal of the curriculum.

- The belief that teaching science is straightforward, and that learning to teach science merely requires to follow a script, conflicts with the notion of teaching as a complex activity (Loughran, 2007), and may seriously hinder the consideration of alternative perspectives (such as conceptual change or guided inquiry).
- Being based on, and intertwined with, experiences over many years, prospective teachers' beliefs appear to be quite stable and do not change easily (Davis, 2004).

Pedagogical Content Knowledge

Following the model of Magnusson *et al.* (1999), we briefly discuss the following elements of PCK from the perspective of future science teachers: (1) knowledge of student learning of science, (2) knowledge of science instructional strategies, (3) knowledge of science curricula, and (4) knowledge of science assessment.

Knowledge of student learning of science

Studies on science teachers' knowledge of students' alternative meanings of science (misconceptions) and their approaches to learning science demonstrate that, not surprisingly, experienced teachers usually have more knowledge in this area than future and beginning teachers. In particular, studies contrasting the knowledge of expert teachers with that of beginning teachers emphasized that the latter are normally unaware of students' learning difficulties, and ignorant of their alternative conceptions (e.g., Akerson *et al.*, 2000).

Knowledge of science instructional strategies

Science instructional strategies, such as inquiry-based, conceptual change, context-based, and problem-based approaches are not very well known to beginning teachers. Moreover, it appears that beginning teachers do not easily adopt these strategies partly because: they conflict with their orientations toward teaching science; partly because they have rarely experienced good models of these strategies; and partly because they are in conflict with the school culture (Adams and Krockover, 1997). Topic-specific strategies, aimed at promoting students' understanding of certain science concepts (e.g., Ohm's law, chemical equilibrium, and homeostasis) can be used effectively only if a science teacher is knowledgeable about the ways students learn and come to understand the respective topics. Since prospective teachers cannot be expected to have this sort of knowledge at the start of their preparation, they tend to follow the strategies available for them (i.e., in science textbooks) in a strict manner (Hashweh, 1987).

Knowledge of science curricula

Few studies have been published on future and beginning science teachers' curricular knowledge. Southerland and Gess-Newsome (1999) found that student teachers of elementary science took their science curriculum as given, without questioning the topics to be taught. In addition, a study by Geddis *et al.* (1993) showed that student teachers in a secondary science program tended to focus on nonessential concepts because they did not know which topics in their curriculum were critical, and how deeply to cover such topics.

Knowledge of science assessment

There is a dearth of studies on future and beginning teachers' knowledge of science assessment. According to our experience, the challenge of science teacher education is to broaden the knowledge of assessment from traditional assessment for grading (e.g., tests and lab reports) to assessment for learning through formative and summative authentic methods. Such methods aim to monitor and assess students' conceptual understanding; students' laboratory, inquiry and modeling skills; and their epistemological beliefs (cf. nature of science).

The above discussion shows that, in the context of science teacher education, many students have knowledge and beliefs that need to be challenged, extended, and elaborated (Loughran, 2007; Russell and Martin, 2007). How this can be done in the course of a science teacher education program is discussed in the following sections.

The Curriculum for Science Teacher Education

The design of curricula for science teacher preparation (both the overall teacher preparation program and science-specific pedagogy courses) is based in large part on values and purposes for science teaching and science teacher education held by science teacher educators. Most programs for science teacher preparation around the world (see Abell, 2000) recognize the importance of subject-matter knowledge, general pedagogical knowledge, and PCK in preparing high-quality science teachers. Thus, these programs include recognizable and common components to build these knowledge bases: liberal arts courses, science content courses, general pedagogy courses (e.g., educational psychology, classroom management, educational history and philosophy), and subject-specific teaching and learning courses (called methods courses in the US and something similar to subject matter for teaching in Europe, e.g., *Fachdidaktik* in Germany). Most programs also value the authority of learning from experience (Russell and Martin, 2007) and thus include significant supervised field experiences in school classrooms. The programs vary widely in terms of the relative

emphasis placed on these components, on program length, and on academic level. In some countries, for example, science teacher education takes place in 2-year colleges, in some places it requires a 4-year university program, and in other places teacher preparation is a fifth year or master's program. Recently, countries suffering from science teacher shortages, such as the US, the UK, and the Netherlands, have seen the advent of postbaccalaureate alternative certification programs that attract mid-career individuals into the teaching profession.

The enactment of these components is influenced by the philosophical orientation of the teacher preparation program. Researchers have proposed different ways to categorize the purposes and goals for science education. For example, Roberts (1988) defined various curriculum emphases in science educations, including providing a solid foundation in science, preparing students for the next grade level, science skills development, and helping students generate everyday explanations. Such classification schemes can help us think about the curriculum of science teacher education as well. For example, science teacher preparation programs of the past focused on preparing teachers as technicians capable of implementing specific strategies (e.g., wait time, cooperative groups). Recent views of science teacher education recognize the importance of challenging future teachers' prior knowledge and beliefs, and helping them see viable alternatives to transmission-oriented science teaching. Russell and Martin (2007) called this orientation to science teacher education "teaching for conceptual change." Science teacher education curricula that follow a conceptual-change orientation are aimed at helping future teachers to reflect on practice and make decisions that are grounded in student learning. Whereas the technical approach to science teacher education, referred to above, often resulted in teachers who strictly followed their textbooks and focused on teaching for the test, recent curricula for science teacher education often aim at preparing science teachers to design and test curricular materials (e.g., Justi and Van Driel, 2005).

Most programs meant to prepare future science teachers include subject-specific pedagogical coursework focused on the teaching and learning of science at particular grade levels. Dependent on the teacher educator's orientation to teaching science teachers (see Abell and Bryan, 1997), these courses may include topics such as:

- how students learn science, including common misconceptions and conceptual change approaches;
- state and national standards and curriculum for teaching science (including the nature of science, personal and societal goals, science habits of mind, and science concepts and principles);
- science instructional strategies and representations for teaching specific science topics (e.g., use of models in

- science, laboratories and laboratory safety, context-based instruction, demonstrations, computer simulations, analogies, science talk, and science writing); and
- assessment in science classrooms.

We conclude that curriculum design for science teacher preparation programs should be deliberate and based on accepted goals and purposes for science teacher education. To accomplish the goals and purposes of science teacher preparation, science teacher educators must have a repertoire of effective instructional strategies. We address this topic in the next section.

Instructional Strategies in Science Teacher Education Programs

Given the gap between the beliefs about science teaching and learning held by individuals who enter science teacher preparation programs and those who design such programs, science teacher educators need strategies to identify and address prospective science teacher beliefs. Davis (2004) reviewed several instructional strategies that might be effective for identifying science teachers' initial knowledge and beliefs, which include the use of narratives, pictorial representations, descriptions of ideal science lessons, metaphors, and concept maps. Science teacher educators also need strategies to promote changes to these ideas. For this purpose, strategies such as discussing cases, engaging in action research or lesson study, and promoting reflection have been successful (Abell *et al.*, 1998; Russell and Martin, 2007). In any case, it is crucial that science teacher educators monitor the extent to which the strategies they apply promote conceptual change and knowledge integration among prospective teachers. Science teacher educators use different strategies depending on their learning goals for prospective teachers. To develop prospective teachers' knowledge of students' learning difficulties and of their alternative conceptions, researchers have demonstrated that effective strategies include (1) observing student learning in classroom situations (both in real classrooms, as with the use of video-vignettes), (2) interviewing students and experienced teachers, (3) analyzing data related to student learning (e.g., transcriptions of classroom discussions, think-aloud protocols, and student paperwork; Geddis, 1993), and (4) studying and discussing relevant research literature in relation to prospective teachers' observations and practical experiences (Van Driel *et al.*, 2002).

Some studies have shown that a better understanding of student misconceptions (e.g., by studying authentic student responses to test questions) may help prospective teachers understand the importance of specific instructional strategies, such as the use of analogies (Geddis, 1993). Veal advocated the use of content-specific vignettes,

consisting of written descriptions of teaching and learning episodes, as an effective strategy to help prospective teachers develop ideas about teaching a specific topic (Veal, 2002). In the context of elementary science teacher education, Appleton (2003) suggested providing prospective teachers with so-called activities that work as a means to increase both their science content knowledge and their confidence to teach science, and also function as a source for developing science PCK.

However, only a few studies report successful attempts to promote science teachers' knowledge and the use of instructional strategies in practice. All these attempts revolve around gaining practical experiences with the instructional strategies under consideration, in particular, the planning and teaching of science lessons incorporating such strategies, followed by reflection by the prospective teachers to make sense of their experiences (Zemba-Saul *et al.*, 2002). De Jong *et al.* (2005) conducted a project with prospective chemistry teachers on the teaching of particle models. First, the participants analyzed their own experiences as learners, followed by going through some chemistry textbooks. The core of the project consisted of planning and teaching a series of lessons aimed at promoting student understanding of particle models, followed by sharing reflections within the group of prospective teachers and their educators. The researchers found that whereas all participants became more aware of student learning difficulties in this area, about half of the sample also gained a better understanding of the possibilities and limitations of the use of specific teaching strategies, focusing on particle models. In the same domain, Justi and Van Driel (2005) took this approach further by providing prospective teachers with examples of potentially useful instructional strategies, and then asking them to conduct an action-research project on the design and teaching of a lesson series focusing on the use of teaching models. This approach, which involved studying the research literature, collecting and analyzing data from their students, contributed to prospective teachers' understanding of the potential and pitfalls of the instructional strategies they had tried out in their project.

In the domain of laboratory work, Clermont *et al.* (1993) focused on the use of chemical demonstrations, applying a workshop format which enabled the participants (i.e., novice chemical demonstrators) with opportunities to practice demonstrations, both with colleagues who gave feedback and with school students. They found that the participants became more aware of the complexities of several demonstrations, how these complexities could hinder student learning, and how to adapt and simplify demonstrations to avoid these problems. The authors concluded that intensive, short-term, skill-oriented workshops may improve teachers' PCK of demonstrations. In a similar design, combining microteaching and field-based practical experiences, Niess (2005) stimulated prospective

science teachers to integrate technology (i.e., real-time data collection devices) in the design of their lessons, leading to the development of what she labeled technology-enhanced PCK.

Concluding this section, we have seen that approaches which promote prospective science teachers' knowledge and beliefs about the teaching and learning of science combine the following characteristics: (1) explicating initial ideas about teaching and learning; (2) analyzing student learning of science; (3) planning and conducting science lessons, either in workshops, or, preferably, in real classrooms; and (4) reflecting and discussing experiences, thus relating activities (1), (2), and (3).

Assessment of Science Teacher Preparation

Assessment of science teacher preparation can take place at various levels. Program-level assessment of science teacher education programs in some countries is governed by program accreditation processes. In the US, for example, the National Science Teachers Association in cooperation with National Council for the Accreditation of Teacher Education (NCATE) specifies standards and benchmarks for programs that outside reviewers use to judge program quality (Gilbert, 1994). Most recently, in order to comply with outside policy agencies and to provide a summative assessment of prospective teacher learning, science teacher education programs often require graduating teachers to synthesize their learning through portfolios or action research projects, including evidence of student performance.

Assessment of science teachers in specific courses can also be accomplished via multiple measures. A plethora of assessment methods can be found in the science teacher knowledge research literature (Abell, 2007). To assess science teacher subject-matter knowledge, researchers have used concept maps, interviews, diagrams, and tests. Similar types of measurements would be applicable to the college science course to assess prospective teachers. To assess PCK, researchers have devised a number of tasks, such as lesson preparation, reflection on video cases, the development of content representations and pedagogical experience repertoires (Loughran *et al.*, 2004), and the use of rubrics to distinguish among different levels of PCK (Taylor and Gess-Newsome, 2007). Each of these types of assessments is designed to both document teacher knowledge development and to help them learn from the assessment itself.

Programs of science teacher preparation also assess prospective teacher learning during field experiences, including student teaching internships. For example, supervisors use observation rating scales to provide formative feedback to prospective teachers as well as

summative evaluations of their performance. Prospective teachers often track their performance through reflective journals and analysis of classroom videos. In these cases, the goal is to help prospective teachers think about student learning and their own practice, and make changes in their practice based on feedback from the assessment.

Assessment in science teacher preparation thus serves several purposes. It can be used to diagnose prospective teachers' initial ideas, to provide formative feedback to guiding learning and instruction, and to provide summative evaluation for grades, program accreditation, and teacher licensure. Very similar to assessment in science classrooms, the best assessment provides a learning experience as well as data on knowledge and performance.

Conclusions and Further Professional Development

Science education, whether it aims at introducing all learners to the main ideas and principles of science or at the training of future scientists, calls for teachers with specific qualifications. To teach science effectively, that is, in a way that promotes students' understanding and abilities, science teachers need a thorough understanding of the ways their students learn science content and skills, and what sort of learning difficulties may occur, and why. Moreover, it is important that science teachers understand what and how science can be interesting or challenging for their students. Closely connected to this understanding, science teachers need to develop a large repertoire of instructional strategies and representations of science content, which they can use in classroom practice in a flexible way so as to accommodate student learning, stimulate interest in science, and anticipate differences between students. Moreover, similar to teachers of other subjects, they need to know and use a repertoire of formative and summative assessment techniques, which goes beyond the traditional and familiar ways of testing. In this article we have demonstrated how science teacher education can contribute to the initial development of the professional knowledge needed to teach science. For a more lengthy discussion, the reader is referred to the 'Further Reading' section. We conclude this article with some remarks about science teachers' further professional development.

Preparing science teachers during initial teacher education is only the first step in their professional development. As newly qualified and beginning teachers, they continue to work toward gaining expertise. In particular, since science knowledge is dynamic, science education is constantly evolving, thus requiring science teachers to be continuously critical of what and how they teach. The challenge for science teacher educators is to help teachers develop into adaptive experts (Bransford *et al.*, 1999), that

is, professionals who have developed a set of important routines, but at the same time, keep looking for new possibilities to improve and adapt their practice, experimenting with new approaches, and so on. In any case, it is important that science educators help science teachers move beyond copying the models from their own education that worked well for them in the past, because those methods likely will not work with many of their students.

Science teacher education is facing a critical time when the need to prepare high-quality science teachers is at a peak. Although we recognize that learning to teach science is a career-long endeavor, we believe that a strong foundation must be laid during the science teacher preparation program. In order to improve science teacher preparation, the field must attend to the preparation of future science teacher educators. Our doctoral programs must develop, in addition to adept researchers, informed and skilled science teacher educators who understand who their students are, how to develop meaningful curriculum and effective instruction, and how to assess the development of science teacher knowledge.

See also: A Pedagogy of Teacher Education; Contemporary Approaches to Teacher Professional Development; Mathematics Teacher Education; Pedagogical Content Knowledge; Social Studies Teacher Education; Taking Prospective Teachers' Beliefs into Account in Teacher Education; Teacher Induction; Teacher Preparation in Mother Tongue Education.

Bibliography

- Abd-El-Khalick, F., Bell, R. L., and Lederman, N. G. (1998). The nature of science and instructional practice: Making the unnatural natural. *Science Education* **82**, 417–436.
- Abell, S. (2000). *Science Teacher Education: An International Perspective*. Dordrecht: Kluwer Academic Press.
- Abell, S. (2007). Research on science teacher knowledge. In Abell, S. and Lederman, N. (eds.) *Handbook of Research on Science Education*, pp 1105–1149. Mahwah, NJ: Lawrence Erlbaum Associates.
- Abell, S. K. and Bryan, L. S. (1997). Reconceptualizing the elementary science methods course using a reflection orientation. *Journal of Science Teacher Education* **8**, 153–166.
- Abell, S. K., Bryan, L. A., and Anderson, M. A. (1998). Investigating preservice elementary science teacher reflective thinking using integrated media case-based instruction in elementary science teacher preparation. *Science Education* **82**, 491–510.
- Adams, P. E. and Krockover, G. H. (1997). Beginning science teacher cognition and its origins in the preservice secondary science teacher program. *Journal of Research in Science Teaching* **34**, 633–653.
- Akerson, V. L., Flick, L. B., and Lederman, N. G. (2000). The influence of primary children's ideas in science on teaching practice. *Journal of Research in Science Teaching* **37**, 363–385.
- Appleton, K. (2003). How do beginning primary school teachers cope with science? Toward an understanding of science teaching practice. *Research in Science Education* **33**, 1–25.
- Bransford, J. D., Brown, A. L., and Cocking, R. R. (1999). *How People Learn: Brain, Mind, Experience and School*. Washington, DC: National Research Council.
- Clermont, C. P., Krajcik, J. S., and Borko, H. (1993). The influence of an intensive in-service workshop on pedagogical content knowledge growth among novice chemical demonstrators. *Journal of Research in Science Teaching* **30**, 21–43.
- Davis, E. A. (2004). Knowledge integration in science teaching: Analysing teachers' knowledge development. *Research in Science Education* **34**, 21–53.
- De Jong, O., Van Driel, J. H., and Verloop, N. (2005). Pre-service teachers' pedagogical content knowledge of using particle models in teaching chemistry. *Journal of Research in Science Teaching* **42**, 947–964.
- Geddis, A. N. (1993). Transforming subject-matter knowledge: The role of pedagogical content knowledge in learning to reflect on teaching. *International Journal of Science Education* **15**, 673–683.
- Geddis, A. N., Onslow, B., Beynon, C., and Oesch, J. (1993). Transforming content knowledge: Learning to teach about isotopes. *Science Education* **77**, 575–591.
- Gilbert, S. W. (1994). NSTA in NCATE: Some results from the folio reviews and directions for the future. *Journal of Science Teacher Education* **5**, 131–138.
- Hashweh, M. Z. (1987). Effects of subject-matter knowledge in the teaching of biology and physics. *Teaching and Teacher Education* **3**, 109–120.
- Hattie, J. (2003). *Teachers Make a Difference: What is the Research Evidence?* Paper presented at the Australian Council for Educational Research Annual Conference on Building Teacher Quality, Melbourne.
- Justi, R. and Van Driel, J. H. (2005). The development of science teachers' knowledge on models and modelling: Promoting, characterizing, and understanding the process. *International Journal of Science Education* **27**, 549–573.
- Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching* **29**, 331–359.
- Loughran, J. J. (2007). Science teacher as learner. In Abell, S. and Lederman, N. (eds.) *Handbook of Research on Science Education*, pp 1043–1066. Mahwah, NJ: Lawrence Erlbaum Associates.
- Loughran, J., Mulhall, P., and Berry, A. (2004). In search of pedagogical content knowledge in science: Developing ways of articulating and documenting professional practice. *Journal of Research in Science Teaching* **41**, 370–391.
- Magnusson, S., Borko, H., and Krajcik, J. (1999). Nature, sources, and development of pedagogical content knowledge for science teaching. In Gess-Newsome, J. and Lederman, N. G. (eds.) *Examining Pedagogical Content Knowledge*, pp 95–132. Dordrecht: Kluwer Academic Publishers.
- Niess, M. L. (2005). Preparing teachers to teach science and mathematics with technology: developing a technology pedagogical content knowledge. *Teaching and Teacher Education* **21**, 509–523.
- Osborne, J. and Dillon, J. (2008). *Science Education in Europe: Critical Reflections*. London: The Nuffield Foundation.
- Roberts, D. A. (1988). What counts as science education? In Fensham, P. J. (ed.) *Development and Dilemmas in Science Education*, pp 27–54. London: Palmer Press.
- Russell, T. and Martin, A. K. (2007). Learning to teach science. In Abell, S. and Lederman, N. (eds.) *Handbook of Research on Science Education*, pp 1151–1176. Mahwah, NJ: Lawrence Erlbaum Associates.
- Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher* **15**(2), 4–14.
- Southerland, S. A. and Gess-Newsome, J. (1999). Preservice teachers' views of inclusive science teaching as shaped by images of teaching, learning, and knowledge. *Science Education* **83**, 131–150.
- Taylor, J. and Gess-Newsome, J. (2007). *Exploring Tools and Methods for Measuring Pedagogical Content Knowledge*. Paper presented at the annual meeting of the Association for Science Teacher Education, Clearwater, FL.
- Van Driel, J. H., De Jong, O., and Verloop, N. (2002). The development of pre-service chemistry teachers' PCK. *Science Education* **86**, 572–590.

- Veal, W. (2002). Content specific vignettes as tools for research and teaching. *Electronic Journal of Science Education* **6**(4), 37 pp. <http://wolfweb.unr.edu/homepage/crowther/ejse/ejsev6n4.html>.
- Zemba-Saul, C., Krajcik, J., and Blumenfeld, P. (2002). Elementary student teachers' science content representations. *Journal of Research in Science Teaching* **39**, 443–463.

Further Reading

- Appleton, K. (2007). Elementary science teaching. In Abell, S. and Lederman, N. (eds.) *Handbook of Research on Science Education*, pp 493–535. Mahwah, NJ: Lawrence Erlbaum Associates.
- Davis, E. A., Petish, D., and Smithey, J. (2007). Challenges new science teachers face. *Review of Educational Research* **76**, 607–651.
- Gess-Newsome, J. (1999). Secondary teachers' knowledge and beliefs about subject matter and their impact on instruction. In Gess-Newsome, J. and Lederman, N. G. (eds.) *Examining Pedagogical Content Knowledge*, pp 51–94. Dordrecht: Kluwer Academic Publishers.
- Hewson, P. W. (2007). Teacher professional development in science. In Abell, S. and Lederman, N. (eds.) *Handbook of Research on Science Education*, pp 1179–1203. Mahwah, NJ: Lawrence Erlbaum Associates.
- Lederman, N. G. (2007). Nature of science: past, present, and future. In Abell, S. and Lederman, N. (eds.) *Handbook of Research on Science Education*, pp 831–879. Mahwah, NJ: Lawrence Erlbaum Associates.
- Luft, J. (2007). Minding the gap: Needed research on beginning/newly qualified science teachers (Guest editorial). *Journal of Research in Science Teaching* **44**, 532–537.
- Roth, K. J. (2007). Science teachers as researchers. In Abell, S. and Lederman, N. (eds.) *Handbook of Research on Science Education*, pp 1205–1259. Mahwah, NJ: Lawrence Erlbaum Associates.

Social Studies Teacher Education

A-L Halvorsen and S M Wilson, Michigan State University, East Lansing, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Authentic pedagogy – An instructional model that highlights the authenticity of student learning and achievement; it stimulates students to form deep understandings of a topic or discipline and develop new understandings by adopting various inquiry methods, collaborating with others, and addressing real-world public problems.

Citizenship education – A broad field of study with goals to help students respect different cultures, develop the willingness and capability to participate in politics at different levels, cultivate skills to work with others, and examine and address environmental and human right issues in a global context.

Content knowledge – The facts and organizing principles of subject areas.

Culturally relevant pedagogy – An educational perspective with the goal of meeting the academic and social needs of culturally diverse students; it seeks to make school learning meaningful and relevant to children's lives beyond school.

Global citizenship education – A field of study that seeks to balance unity and diversity of peoples and cultures worldwide; it concentrates on the education of students to understand and appreciate different cultures, to think and address problems with global consciousness and compassion, and to collaborate for the common good.

Multicultural education – An area of study with the goals of helping all students to develop knowledge, skills, and attitudes to participate effectively in democratic society and of creating equal educational opportunities for students from diverse racial, ethnic, social-class, and cultural groups.

Pedagogical content knowledge (PCK) – A theoretical perspective for understanding teaching and learning proposed by Lee Shulman and colleagues as part of the Knowledge Growth in Teaching project; PCK is the knowledge formed by subject matter knowledge, pedagogical knowledge, and context knowledge.

General Challenges in Social Studies Teacher Education

As context for understanding the state of current challenges roiling the field, we begin with a very brief tour of some enduring and contemporary problems (e.g., purpose and content of social studies teacher education; competing notions of citizenship; the lack of research in social studies teacher education; and the content versus method debate for preparation courses) faced by social studies educators and teacher educators. Because the problems inherent in social studies teacher education are linked to social studies education in general, our discussion draws upon both fields.

Purpose and Content

First is the issue of purpose. Social studies and history education have historically suffered from multiple goals. Often, social studies is earmarked as the curricular domain responsible for the educating of citizens (although one might argue that educating good citizens is the responsibility of all educators, no matter what is the focus of their subject area). Others have suggested that social studies concerns the teaching of a set of subject areas, which can be narrowly conceived or broad. In the US, for example, there are some who argue for the teaching of history, economics, and geography alone. Others argue for a broader view. For example, the National Council for the Social Studies (NCSS), the leading US organization in the field, lists anthropology, archeology, economics, geography, history, law, philosophy, political science, psychology, religion, and sociology as the core areas of social studies focus, complemented by appropriate content from the humanities, mathematics, and natural sciences. Thus, social studies teacher education similarly faces the challenge of purpose, goals, and objectives. Should the preparation focus on preparing teachers to teach skills for citizenship, disciplinary content, disciplinary skills, an integrated approach, and/or community service? These goals, often competing, complicate the role of the social studies teacher educator.

A related issue is how to teach these disciplines collectively, integrating them as a coherent field of study, while simultaneously preserving the integrity of the various domains. The enormity of this challenge – providing

both breadth and depth in numerous disciplines – can be intimidating and discouraging. The social studies teacher educator faces a related challenge: What liberal and specialized education prepares future teachers to teach a broad sweep of topics that they cannot possibly master in deep ways?

Moreover, the inherently controversial nature of many topics and concepts related to the social studies – for instance, equality, liberty, diversity, war, human rights, racism, or terrorism – can be a lightening rod for often heated debates. Whose history gets told, who benefits from what economic trend, who has what kinds of rights, where boundaries are located, and who has jurisdiction over whom – these are a handful of issues that social studies teachers confront on a regular basis. There are two related special challenges for the social studies teacher educator: First, teacher educators need to recognize that their own students – prospective teachers – will have diverse opinions and views on such topics. Second, those teacher educators need to help prospective teachers become educators who can respect the multiple perspectives of their own students (no matter what the teacher's own perspectives may be), as well as create educational opportunities that might involve surfacing and confronting those differences.

Content Knowledge versus Method/Pedagogy

Second, a contentious area in social studies teacher education is the debate on disciplinary content versus method, or pedagogy. Disciplinary content refers to the facts and organizing principles of subject areas. For the social studies subjects, such as history, geography, and civics, disciplinary content may include significant people, historical events, the Earth and its peoples, and forms of government. Furthermore, disciplinary content also refers to the disciplinary processes or thinking skills. Such processes involve thinking chronologically, interpreting primary documents, creating and reading maps, and making collective decisions. Generally, education students study such content in their prerequisite courses in the individual disciplines prior to their enrollment in education courses. However, research has shown that preservice teachers are often not well prepared in these disciplines. This is not surprising given the fact that most prospective teachers never get too deeply into any one subject area given the breadth of topics they need some familiarity with. As a result, teacher educators in social studies frequently teach content as well. The problem that may arise for the teacher educator is twofold: insufficient course time to teach subject content and insufficient in-depth expertise in all the various disciplines.

Methods courses address the means teachers use to convey knowledge, skills, and values to their students. They may use general strategies, following the guidelines

of authentic pedagogy; complex instruction; controversial public issues; or understanding by design. In their methods courses, preservice teachers are introduced to these general strategies that focus on content standards, assessments, and instructional approaches; or teacher educators may focus on teaching preservice teachers specific instructional strategies, such as inquiry, simulation, concept formation, discussion, and case-based methods. In such methods courses, there is the assumption that preservice teachers have acquired content knowledge in the various disciplines.

Social studies teacher educators in methods courses are directly involved in the content versus method debate when they are confronted with the reality of whether to teach content, method, or a combination of the two. Many social studies teacher educators argue it is impossible to teach one in the absence of the other – that the dichotomy is actually invalid. However, questions remain about the optimal mix of content and method, as well as the proper extent and nature of the content taught. These questions are further complicated by the problem that, as Lee Shulman suggests, in addition to knowing subject-matter content, teacher educators must also know pedagogical content knowledge (PCK), which is the content of competent instruction.

Competing Notions of Citizenship

Third, while many social studies education scholars and organizations consider citizenship training the cornerstone of social studies instruction, they often have competing notions of what citizenship is. One list of the defining characteristics of citizenship was provided by a multinational panel of 182 scholars, practitioners, and policy leaders from 26 countries of the Citizenship Education Policy Study (CEPS). The panel identified what citizenship should mean in the twenty-first century:

- ability to understand, accept, appreciate, and tolerate cultural differences;
- capacity to think in a critical and systematic way;
- willingness and ability to participate in politics at the local, national, and international levels;
- ability to work with others in a cooperative way and to take responsibility for one's roles/duties within society;
- willingness to resolve conflict in a nonviolent manner;
- ability to look at and approach problems as a member of a global society;
- willingness to change one's lifestyle and consumption habits to protect the environment; and
- ability to be sensitive toward and to defend human rights (e.g., rights of women, ethnic minorities).

Although the CEPS panel's list is laudable in intent, the citizenship characteristics it identifies are, nevertheless, very broad and open to multiple (often competing)

interpretations. For example, interpretation of the fourth bullet, the ability to work with others in a cooperative way and to take responsibility for one's roles/duties within society, is vague enough to describe both the passive, law-abiding citizen and the radical, reformer citizen. It is also possible that the CEPS citizenship characteristics may be too idealistic. Depending on geographic, political, and social circumstances, the seventh bullet, the willingness to change one's lifestyle and consumption habits to protect the environment, may be interpreted by many societies as an unrealistic, if not unattainable, virtue.

For teacher educators in social studies, finding an agreed-upon conception of citizenship, at either the national or global level, presents a challenge. Unlike mathematics where the Pythagorean theorem, for example, is the same, context to context, a key goal of social studies education – education for citizenship – varies not only between nations, but also within nations. Some countries refer to the term citizenship as the judicial relationship between citizen and state, whereas in other countries, the term refers to the way citizens interact and coexist with one another.

In recent years, education researchers have focused on citizenship as commitment to community action. Marilynne Boyle-Baise and Carl Grant have traced the evolution of teaching citizenship in the US from the 1970s to the present and have shown how citizenship, which had once been thought of as community participation education, currently is considered service learning. In the earlier understanding, citizenship meant the participation of society in community decision making, but more recently citizenship has adopted the Rawlsian ideals of justice that focus on collective participation in the distribution of community resources as a way of promoting the common good. Both of these notions of citizenship, past and present, are concerned more with societal action than with the acquisition of knowledge.

Other organizations have offered competing interpretations of citizenship that, by contrast, stress knowledge as the key component of citizenship. For example, in 2003, the Fordham Foundation, a Washington, DC research organization influential in educational circles, published *Where Did Social Studies Go Wrong?* This book, in its back-to-the-basics approach to teaching social studies, argues that citizenship studies should focus on the history and civic knowledge and values of Western traditions. Furthermore, the book argues that social studies education (including teacher education) has been hijacked by the so-called forces of political correctness with the result that history study has been watered down, replaced by topics such as multiculturalism, environmental studies, peace studies, and social justice.

However citizenship study in social studies education is defined, research shows that preservice teachers generally have only a vague understanding of the notion of

citizenship. Their struggle with defining and teaching citizenship in their classes may be because of the multiple definitions of citizenship they have learned and because their own experience as participating citizens has been rather limited. For teacher educators in social studies, the challenge to find a responsible and neutral way to instruct preservice teachers in citizenship pedagogy is one of their most difficult challenges.

Moreover, standards concerning the good citizen say nothing about the fact that many citizens of many nations do not want their children to think critically, consider themselves global citizens, protect the environment, or defend human rights. Citizenries are conflicted, and social studies teachers are not well prepared when they enter schools ignorant of the mismatch between the variety of assumptions the public might hold about good citizenship and these normative values. Thus, social studies teacher educators also face the challenge of preparing new teachers to sensitively navigate the idealized norms of schooling and the larger public and community in which that schooling takes place.

Lack of Research on Social Studies Teacher Education

Research on teacher education more generally is limited in scope and method. Research on social studies teacher education suffers from similar problems. A common research model in social studies teacher education at present is the self-study: an investigation of the researcher's own practice undertaken to better understand and improve her/his own teaching. Such case studies on teacher educators' reflective practice are interesting for their description and analysis of the problems, challenges, and successes of the individual teacher educators. However, this research is not easily generalized since it provides no macro-level findings on best practices in teacher education, and there is considerable variability in the quality of that research.

Another area of research on social studies teacher education focuses on the influence of methods courses and field experiences on preservice teachers' beliefs and values, rather than on their acquisition of skills and knowledge. Some studies have found that while there were important connections between students' learning in methods courses and their development of beliefs, preservice teachers still had difficulty implementing the practices learned in the methods courses. Again, this research is similar to much other research concerning teacher education. Samples tend to be modest, instruments tend to be locally developed and not validated, and outcome measures tend to be far removed from measures of student learning.

In summary, contemporary research into social studies teacher education is primarily concerned with

understanding the predispositions of preservice teachers and the effect of methods courses on these beliefs. The research tends to be done by teacher educators who are studying their own programs or practices, and there is very little large-scale or comparative work to draw upon. One argument against this stream of research is that an understanding of preservice teachers' beliefs has little relevance for understanding their teaching approaches; another argument points out that, without evidence of improvement in teachers' practices or student learning, it is difficult to know what insights can be gleaned from the current research.

Trends in Social Studies Teacher Education

Social studies teacher education, like all teacher education, is not a static activity, but changes constantly in response to new trends developed from new research findings and policy mandates concerning schooling and teacher certification or licensure. An awareness and understanding of the current trends in the field is useful in envisioning the development of the field. Three of the most influential trends in social studies teacher education are authentic pedagogy, global citizenship education, and multicultural education/culturally relevant pedagogy (CRP).

Authentic Pedagogy

Authentic pedagogy is an instructional model that sets out three guidelines for curriculum design, instruction, and assessment: (1) construction of knowledge; (2) disciplined inquiry; and (3) value beyond school. In construction of knowledge, students should produce or construct new knowledge (as opposed to reproducing knowledge). In disciplined inquiry, learning should be rooted deeply in a field of knowledge. In value beyond school, learning should have utilitarian or personal value beyond classroom application. Many scholars of social studies teacher education believe strongly in the value of authentic pedagogy as a way to improve learning. Since many disciplines included in social studies are associated with boring pedagogy (e.g., memorizing countries and capitals and dates of wars), proponents of authentic pedagogy work toward teaching social studies in ways that allow students to probe content thoroughly and with application to their lives outside school. This pedagogy is also known as authentic academic achievement.

The NCSS has also taken an active role in advancing more authentic instruction in social studies. The NCSS has identified five key criteria of powerful social studies education: meaningful, integrative, value-based, challenging, and

active. The goal of social studies education that meets these criteria is to help students build a social understanding of the human condition and of their civic responsibility. One professional development program for social studies teachers, produced in partnership with the NCSS, is *Powerful and Authentic Social Studies (PASS)*. This program defines social studies education standards based on both the authentic academic achievement and powerful social studies models. These criteria and standards for social studies teaching and learning reflect educators' commitment in making social studies relevant and intellectually rigorous for students. Many teacher educators in social studies methods courses use these models to evaluate preservice teachers' lesson plans.

However, research shows that generally preservice teachers and in-service teachers do not follow such criteria and standards in their instruction and assessments. Instead, they follow teacher-centered practices that mix lecture with textbook readings and close-ended questioning. Some researchers argue that such traditional social studies teaching, that de-emphasizes standards, denies students instruction that is intellectually rigorous and relevant.

Global Citizenship Education

Global citizenship education (GCE) is the trend that promotes a global dimension to education by seeking to balance unity and diversity of peoples and cultures worldwide. As the world becomes increasingly interconnected through technology, international relations, and human rights projects, educators have begun defining the criteria of a participatory, global citizen. Elizabeth Heilman identifies seven capacities of GCE, including a focus on combining talents to work toward the common good, compassion for others' situations and concern for human rights, and the ability to examine policies and issues critically and ethically. Social studies education, which includes instruction in citizenship, history, and geography, is well positioned to advance this trend.

For the social studies teacher educator, instruction in GCE presents special concerns. Many preservice teachers, as well as in-service teachers (especially those at the elementary level), are often nervous about social and political issues. Some may not feel prepared for, or comfortable with, issues of public concern that are of a controversial or complex nature. They also worry about adverse parental reaction if they make war, religion, race, and so forth the topics of classroom study. In addition, there is a concern that preservice teachers lack the necessary, in-depth knowledge of world geography and history, and of foreign political and economic systems. Furthermore, the global experiences of new social studies teachers are often limited, and that background knowledge and experience might limit their capacity to be the

kinds of teachers GCE needs. Currently, the sensitivity of the issues and the complexity of the instruction make GCE more of an aspiration than an implemented educational trend in social studies teacher education.

Multicultural Education/CRP

Multicultural education is a field of study that generally seeks equality of educational opportunity. CRP seeks teaching strategies that advance this goal. The current trend in multicultural education and CRP is especially significant in social studies education that features culture, human rights, ethnic and social relationships, and issues of race, class, and gender. Thus, social studies teacher educators are concerned with teaching the presentation of multicultural subject content in the classroom and with sensitizing preservice teachers to the needs, interests, and skills of students from diverse backgrounds.

Several scholars of social studies education and multicultural education have set content and practice standards for teachers in this area. Gloria Ladson-Billings, has led critically important research on classroom implementation of relevant teaching strategies based on five criteria of CRP:

1. when students are treated as competent they are likely to demonstrate competence;
2. when teachers provide instructional scaffolding, students can move from what they know to what they need to know;
3. the focus of the classroom must be instructional;
4. real education is about extending students' thinking and abilities; and
5. effective teaching involves in-depth knowledge of both the students and the subject matter.

These criteria not only reflect praiseworthy pedagogical practices, but also reflect the idea that culturally relevant teaching means expecting high-quality work from all students. No distinction is made between the standards set for students from lower socioeconomic backgrounds and those set for students from more privileged backgrounds.

One of the issues of greatest concern in CRP is the importance of recognizing racial and cultural diversity in students. Various scholars and organizations have frequently addressed this issue as it relates to classroom teachers. Teacher educators continually emphasize the importance of connecting school learning to children's lives beyond the school setting, which often means making the curriculum more inclusive of the beliefs and traditions of various cultures and ethnicities. The Multicultural Education Consensus Panel, comprised of eight US scholars of teacher education, proposed 12 essential principles for education and diversity, one of which addresses teacher learning. This principle emphasizes the importance of

professional development programs that instruct teachers in the relationships among ethnic groups and in the ways students' background (race, ethnicity, language, and social class) influence their behavior in schools.

A second issue in multicultural education and CRP is the generally acknowledged cultural divide that often exists between students and their teachers. For example, in the US, many K-12 teachers, especially at the elementary level, are white females from mid- to upper-level socioeconomic backgrounds. Similarly, countries such as France and Sweden, with increasing numbers of ethnic minority immigrants, experience the same separation between teacher and student. Scholars argue that this cultural disjuncture between students and teachers has critical consequences for students' attitudes toward and interest in social studies education. The argument is not that minority teachers are superior as social studies teachers, but that minority children may be more open to learning if taught by people of more similar background.

Summary

Owing, in large part, to the controversy in education around social studies as a course of study that replaces the individual academic disciplines of history and the social sciences, social studies teacher education is also a controversial area. Therefore, the social studies teacher educator faces numerous challenges in preparing preservice teachers for their work in the classroom. Broadly, these challenges relate to the general questions of what and how social studies should be taught. Among the specific challenges that the social studies teacher educator faces are the content versus method debate, the issue of how to define and teach citizenship, and the lack of education research on the pedagogic experiences and problems of teaching social studies.

Social studies teacher education is an evolving field, and today many educational trends influence its development. Standards and criteria for instruction in citizenship and multiculturalism have been proposed; more relevant pedagogy is recommended to replace traditional methods of instruction; and more educational technology in the classroom is available, and its use is encouraged. Each of these trends creates both opportunities and difficulties for social studies teacher education, especially as one considers them together. Meanwhile, there remains a dearth of comprehensive research on social studies teacher education that could inform how and what we teach new teachers in their professional education programs.

With continued experimentation, discussion, and research, social studies teacher education will continue to develop. There are societal and educational issues, challenges, and trends that schools of education and professional development programs will have to deal with as the

practice and theory of teaching teachers to teach evolves. Social studies teacher educators, as much, if not more than other teacher educators, will have to be adaptable, innovative, and committed as they educate preservice teachers for their teaching assignments.

See also: A Pedagogy of Teacher Education; Characteristics, Scholarship and Research of Teacher Educators; Moral Values in Teacher Education; Pedagogical Content Knowledge; Teacher Education as Teaching for Understanding with New Technologies.

Further Reading

- Adler, S. (1991). The education of social studies teachers. In Shaver, J. (ed.) *Handbook of Research on Social Studies Teaching and Learning*, pp 210–221. New York: Macmillan.
- Adler, S. (ed.) (2004). *Critical Issues in Social Studies Teacher Education*. Greenwich, CT: Information Age Publishing.
- Avery, P. G. (2003). Using research about civic education to improve courses in the methods of teaching social studies. In Patrick, J. J., Hamot, G. E., and Leming, R. S. (eds.) *Civic Learning in Teacher Education: International Perspectives on Education for Democracy in The Preparation of Teachers*, pp 45–65. Bloomington, IN: ERIC Clearinghouse for Social Studies/Social Science Education.
- Avery, P. G. (2004). Social studies teacher education in an era of globalization. In Adler, S. (ed.) *Critical Issues in Social Studies Teacher Education*, pp 37–58. Greenwich, CT: Information Age Publishing.
- Banks, J. A., Cookson, P., Gay, G., et al. (2005). Education and diversity. *Social Education* 69(1), 36–40.
- Boyle-Baise, M. and Grant, C. A. (2004). Citizen/community participation in education: Historic change in terms of engagement. In Adler, S. (ed.) *Critical Issues in Social Studies Teacher Education*, pp 145–164. Greenwich, CT: Information Age Publishing.
- Brophy, J. and Allman, J. (1996). *Powerful Social Studies for Elementary Students*. Fort Worth, TX: Harcourt Brace.
- Gay, G. (2000). *Culturally Responsive Teaching*. New York: Teachers College Press.
- Heilman, E. (2008). Including voices from the world through global citizenship education. *Social Studies and the Young Learner*.
- Newmann, F. M., Marks, H. M., and Gamoran, A. (1995). Authentic pedagogy: Standards that boost student performance. *Issues in Restructuring Schools*, 8, pp 1–12. Madison, WI: Center for Education Research, Center on Organization and Restructuring of Schools.
- Stanley, W. B. (ed.) (2001). *Critical Issues in Social Studies Research for the 21st Century*. Greenwich, CT: Information Age Publishing.
- Parker, W. C., Ninomiya, A., and Cogan, J. (1999). Educating world citizens: Toward multinational curriculum development. *American Educational Research Journal* 36(2), 117–145.
- Whitson, J. A. (2004). What social studies teachers need to know: The new urgency of some old disputes. In Adler, S. (ed.) *Critical Issues in Social Studies Teacher Education*, pp 9–35. Greenwich, CT: Information Age Publishing.
- Wilson, S. M. (2001). Review on history teaching. In Richardson, V. (ed.) *Handbook of Research on Teaching*, 4th edn., pp 527–544. Washington, DC: American Educational Research Association.
- Yon, M. and Passe, J. (1990). The relationship between the elementary social studies methods course and student teachers' beliefs and practices. *Journal of Social Studies Research* 14(1), 13–24.

Relevant Websites

- <http://www.citized.info> – Citizenship and teacher education.
- <http://www.ecs.org> – Citizenship Education.
- <http://www.socialstudies.org> – National Council for the Social Studies (NCSS).

Teacher Education for Elementary Education

K D Brown and G N Carnes, University of South Carolina, Columbia, SC, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Accountability – It denotes efforts to cultivate and control for high-quality instruction, including accreditation, high stakes testing, and credentials.

Accreditation – An evaluation process in which a group of evaluators examines an elementary-teacher preparation program to ensure that it is meeting established standards of quality.

Candidate – An individual admitted to, or enrolled in a teacher-education program to earn his/her first credentials to teach.

Education for all (EFA) – Commitment originally made in Jomtien, Thailand (1990) and renewed in Dakar, Senegal (2000) by over one hundred and fifty governments and an equal number of organizations to provide education for all by 2015.

Effective teacher – Teacher impacts to motivate students to learn specified dispositions, skills, and/or knowledge.

In-service teacher – Educator who has earned certification from a teacher preparation program.

Millennium development goals (MDGs) – Eight goals agreed to by the world's governments and major development organizations. Goal #2 is to achieve universal primary education.

Pedagogical knowledge – General concepts, theoretical insights, and research about effective ways in which one teaches.

Preservice teacher – Designation synonymous with candidate.

Primary – Level of education synonymous with elementary education in some countries.

Universal elementary education (UPE) – Access and completion for both boys and girls of a full course of primary schooling.

The international push to introduce qualified teachers into classrooms and the high frequency of teacher shortages around the world shape the way individual countries and international organizations structure and experiment with the preparation of elementary teachers. Many stakeholders view education, particularly the education of young children, as an important avenue for national development. In developing countries in particular, economic growth, development, and improved living standards often result from the provision of quality education to a country's youngest children. While many

factors influence the quality of education, teacher preparation is arguably the central component. Therefore, the preservice preparation of elementary teachers and the ongoing professional development of in-service teachers merit special attention. The commitment of industrialized countries to education for all (EFA) and the approaching millennium development goal (MDG) of universal primary education (UPE) in every country by 2015 brings renewed global interest and concerns about the appropriate training, quality, and availability of elementary-school teachers.

Elementary education varies around the world. Generally, students in elementary education range between the ages of 6 and 14, encompassing grades 1–8. Estonia certifies elementary teachers to teach in grades 1–6, while Kenya certifies its graduates to teach grades 1–8. Within the US, elementary education often spans grades 1 through 5 or 6, overlapping the grade bands for early childhood and middle school. In some countries, the grade band is more limited, focusing on students whose ages may range from 7 to 12. Just as Cobb found in her international comparison of teacher certifications, diversity in certification bands is not limited to national boundaries. Within a country or region, elementary grade levels vary. For example, some teacher preparation programs within the United States offer a Middle Child certification for individuals who wish to teach in grades 4–8. Other US states certify their graduates to teach grades 2–5, K–8, or preschool through 6th grade.

While great diversity in age and grade range continues to define elementary education, the twenty-first century is marked by an emerging global convergence in the definition and government commitment to elementary education. In 2007, the joint expert group (JEG) of the United Nations Educational, Scientific and Cultural Organization (UNESCO) attempted to establish universal parameters that would define the length of elementary education. One possibility linked the duration of compulsory basic education to the minimum age for admission to employment, which was 15 years, under Article 2(3) of ILO Convention No. 138 (1973). Another viewpoint suggested that the first 5 years, subsequent to pre-school education, might be a more suitable minimum requirement. The JEG decided that more conceptual work was necessary to harmonize and clearly define the parameters and nature for elementary and basic education. Although universal parameters have yet to be defined, the global commitment to elementary education has been well-established through the commitment of the vast majority

of the world's governments to EFA and the achievement of the MDG to provide universal elementary education by 2015.

Despite this global commitment to defining and supporting elementary education, the conditions of elementary schools for which teacher education institutes prepare candidates vary tremendously among industrialized and developing countries. UNESCO data indicate that these differences include the pupil–teacher ratio (ranging from 43:1 in sub-Saharan Africa, to 17:1 in Eastern Europe and Central Asia) and the percentage of trained teachers in the classroom (69% in sub-Saharan Africa and 96% in East Asia). Despite the notable increase in elementary-school enrollments from 1999 to 2005 totaling 41 million children, with 36% increase reported in sub-Saharan Africa and 22% in South and West Asia, the Education for All Global Monitoring Report projects that of the 86 countries that have not yet reached universal elementary enrollment, 58 will not be able to achieve this by 2015. Of additional importance to the general context of global elementary education is the significant proportion of uncertified elementary-school teachers common to sub-Saharan Africa and East Asia.

System of Elementary-Education Programs

Teacher candidates for the elementary level have traditionally been educated in stand-alone special teacher training institutes/institutions, also called normal schools. In many parts of the world, the training of preservice elementary teachers has shifted to public or private tertiary-level institutions of higher learning, including universities and nonuniversity, postsecondary institutions. In the United States, majority of elementary-teacher candidates are trained in large state universities, although alternate-route teacher preparation programs based outside the university system have grown in the last decade. In Hong Kong, however, preservice elementary teachers continue to be trained in a specialist institution. In most developing countries and some industrializing countries in Central and South America, prospective elementary-school teachers are given the option of specialized secondary-level programs leading to certification. Conditions for entry into an elementary-teacher education program vary considerably. With the exception of the People's Republic of China, most of the industrialized countries now require applicants to complete secondary education. Individuals who enroll in elementary-teacher education programs in the People's Republic of China typically have completed junior high (secondary) school before enrolling in a 3- or 4-year teacher preparation program.

States play a significant role in the regulation and funding of teacher training programs. Control over the teacher-education institution is typically held by a national-level Ministry of Education (at times with a role for university representatives), a regional or state-level body or by the institution itself. Certification requirements typically are determined by Ministries of Education or, in more decentralized education systems, by a state government. In the US, states are considered gatekeepers since they decide which institutions are permitted to train teachers and determine the standards for a teaching license. The teacher-education system in Germany is similarly designed to the decentralized, state-based system in the US allowing each federal state control over the teacher examination, which is the formal certificate required to teach in a state public school. National regulatory boards have developed in the case of the US, as evidenced by the National Council for the Accreditation of Teacher Education (NCATE) and the Teacher Education Accreditation Council (TEAC), but these bodies lack the authority to regulate and standardize teacher education programs. Despite the absence of these organizations' formal authority over teacher-education programs, most initial teachers in the US will be expected to have graduated from an accredited teacher education program.

General Components of Preservice Teacher Education Programs

Quality teacher preparation provides new teachers with the skills, dispositions, and competence they need to begin their teaching careers and remain in the profession. Darling-Hammond identified several features of exemplary teacher preparation programs, including:

- a common, clear vision of high-quality teaching that is apparent in all coursework and clinical experiences;
- a curriculum grounded in substantial knowledge of child and adolescent development, learning theory, cognition, motivation, and subject-matter pedagogy, taught in the context of practice;
- extended connected clinical experiences (at least 30 weeks) which are carefully chosen to support the ideas and practices presented in simultaneous, closely interwoven coursework;
- well-defined standards of practice and performance that are used to guide and evaluate coursework and clinical work;
- strong relationships, common knowledge, and shared beliefs among school- and university-based faculty; and
- extensive use of case study methods, teacher research, performance assessments, and portfolio evaluation to ensure that learning is applied to real problems of practice.

While the specific aspects of elementary-teacher education programs vary considerably around the world, four components are commonplace: qualifying exams, a tertiary degree requirement, academic and professional coursework, and a teaching practicum. The structure of elementary-teacher education varies greatly and a broad discussion exists around the globe concerning the appropriate balance of these curricular and experiential components of teacher education. Williamson and Morris have found that in countries influenced by the craft-oriented British model of teacher education (e.g., Hong Kong, Australia, and New Zealand), there is a greater emphasis on the practicum component of the program. In countries that have programs with a more academic emphasis (e.g., China and Japan), a greater proportion of time is earmarked for the study of subject matter. An additional model of teacher education common in many African countries is what Avalos identifies as a behavioral skills-training approach, which prepares candidates to act in particular ways rather than exercise judgment.

The twenty-first century is bringing a more systematic restructuring and redesigning of elementary-teacher education programs. For member countries of the European Union, the Bologna Declaration strove to create a uniform bachelor- and master-degree system across Europe by 2009. For countries like Germany, Blömeke has observed that the Declaration has meant a major shift in the way elementary candidates are educated; taking the place of the traditional two-phase system (described below) will be a three-phase system involving a 3-year bachelor degree program, a 1-year master's phase and a 2-year practical training phase. A second common reform is the shift in teacher preparation to the increased role of schools in the initial teacher-education process typically through partnerships. The following section reviews these four components in details and provides specific examples from elementary-teacher education programs.

Exams

While many programs require teacher candidates to demonstrate their ability to apply pedagogical knowledge in an elementary classroom, teacher candidates must complete one or more teacher examinations to show how they are able to synthesize knowledge that resulted from coursework and field experiences. Examinations that are used to determine candidates' academic promise and levels of literacy are increasingly common. In some countries, elementary-teacher applicants must successfully pass a national exam. In Germany, candidates must complete oral and written exams at two points in their preservice education. A two-phase system of education constitutes German elementary-teacher preparation. The first period, an academic focus on subject and pedagogical

courses lasting 3.5 years, concludes with the First State Examination. Successful completion of this exam allows the student to graduate to the second phase of his/her education, which emphasizes learning and developing teaching skills through a mix of student teaching and studying general and subject-specific pedagogy. The Second State Examination concludes the preservice teacher's formal initial education. In the US, the mix of these four components is arranged differently. Many states use one exam, typically the PRAXIS I, as an entry exam into the elementary-teacher education program. After approximately 3 years of coursework, including a practicum, preservice teachers take their final exam, PRAXIS II, as part of their final certification process.

Tertiary Degree Requirement

International organizations such as UNESCO support a global standard requiring tertiary qualifications for teaching at the elementary level. A significant range exists among states, even within a geographic region, to achieve this international ideal. For example, almost all of Jordan's elementary-school teachers have this tertiary qualification, whereas only 14% of Tunisia's elementary teachers meet this requirement. UNESCO data reflect the low percentage of China's elementary teachers with this qualification – 13% – which contrasts sharply with the high percentage – 97% – of Peruvian elementary teachers. Some countries, like Vietnam, do not require a tertiary degree as a requirement for elementary-school teaching. The length of the tertiary qualification also varies by country. Most Organization for Economic Cooperation and Development (OECD) countries require 3–4 years of tertiary training, but Russia only requires 2 years and Germany mandates 5.5 years. The extent of this tertiary requirement has an impact on the academic foundation of the elementary teachers and the future need for in-service education.

Academic and Professional Coursework

Teacher-education programs typically require a general and professional education with an area of specialization. The purpose of the academic core courses in an elementary-teacher education program is to provide the candidate with a sound general education. In most US institutions, elementary-teacher candidates must complete liberal arts courses, increasing their level of literacy in several disciplines. This general course work aligns with common courses required for other university students in humanities, social science, natural science, mathematics, and physical science disciplines. The pedagogical or professional courses that include pedagogical studies, teaching methods, and strategies aim to acquaint preservice teachers with instructional strategies and issues specific to elementary education. In many industrialized countries,

this component of the teacher-education program includes general exposure to the curricular content concentrating on broad frameworks to facilitate the understanding of the individual and social context of learning and teaching (i.e., psychology, social foundations, etc.).

Practice/Practicum

A fourth component of teacher-education programs is the practicum or teaching practice. The practicum, also called practical training, provides opportunities to develop teaching skills and experience on the ground, in elementary classrooms. The education of elementary candidates typically includes some degree of actual classroom experience under the guidance of an experienced teacher, using an apprentice-oriented approach. The practicum component of the preservice program can last from a couple of weeks (i.e., in Taiwan) to almost 4 months (i.e., in New Zealand).

A vigorous, international, and multi-decade debate has been ongoing concerning the design and efficacy of the practicum. Schulz finds critics of the traditionally designed practicum, which emphasizes the mentoring teacher's evaluation of the preservice teacher's demonstrated technical skill in the classroom, argue for a practicum with a broader educative purpose that provides candidates with opportunities for inquiry, for trying and testing new ideas within supportive environments, and for engaging in new conversations about teaching and learning.

In-Service Education

In-service education, also called in-service training (IST), in-service training, education, and practice (ISTEP), teacher professional development (TPD), and continuing education, refers to the professional development of teachers who are already in the classroom. Governments worldwide rely on in-service education to improve and update the skills of current teachers as well as to introduce and implement educational reform. For example, in the 1990s, state governments in countries, including Vietnam and Namibia, have depended on IST to promote child-centered curricular reform. ISTEP sessions for elementary teachers range in length from a couple of hours (e.g., Estonia) to at least a month (e.g., Mongolia before voucher reform) during the academic year or breaks; IST may be mandatory, a requirement for continued certification, or voluntary. The financing and administration of ISTEP varies worldwide. In industrialized countries like New Zealand, groups associated with national or teacher-education programs, such as the Advisory Service, operate in-service teacher education, whereas in many developing countries, organizations based outside of

the country, such as international nongovernmental organizations (NGOs) and bilateral- and multilateral organizations dominate ISTEP.

International organizations (e.g., UNESCO) have found existing in-service education programs to be inadequate and ineffective. Although participation rates in in-service programs may be high, the transfer of the skills learned in these IST sessions and a significant change in teacher action is restricted by, at least, four factors: (1) a disjuncture between the assumed nature and the reality of classroom conditions; (2) the lack of follow-up support by skilled mentors; (3) the continued dominance and demands of exams, whose role and content may be out of line with the proposed reform; and (4) outdated teaching methods used in in-service sessions.

Quality Control of Elementary-Teacher Education Programs: Accreditation

While governments historically had some accountability measures, the increased prevalence of international studies of pupil performance (e.g., PISA and TIMSS) and growth of domestic comparative examinations have given rise to greater accountability measures and mechanisms for teacher preparation programs. Increased interest in and concern about teacher quality worldwide is related to evidence that an effective elementary teacher is the strongest determinant for student achievement. Many countries identify effective teachers as an important factor for student achievement.

Accreditation of elementary-teacher education programs varies considerably, as revealed by the following discussion of the situation in US, Canada, and China. Given the differences in the manner in which elementary/elementary-teacher preparation programs are accredited throughout the world, commonalities are few. The teacher-education programs are judged by a defined set of expectations that serve as a rubric for approval. These guidelines establish coherence to the assessment of the quality of teacher education. In many cases, the evaluation team consists of individuals within and outside of the institution. This configuration adds to the level of credibility to the accreditation process.

In some countries and provinces (i.e., US and Ontario, Canada), the quality of the teacher-education program is based on qualities that teacher candidates demonstrate. In another instance, accreditation is based on opportunities that candidates have to demonstrate competencies. In the US, several elementary-teacher preparation programs submit folios to the National Council for Accreditation of Teacher Education (NCATE), which accredits almost 98% of teacher-education programs, or the Teacher Education Accreditation Council (TEAC). In doing so, they must provide summaries of candidates know and

can do in light of 19 standards that the Association for Childhood Education International (ACEI) and other professional organizations have identified for elementary-teacher candidates through the use of six to eight assessments. In addition to reviewing the submitted report, the evaluation team visits the institution to interview faculties and teacher candidates to gain additional insights into the evidence under consideration. On the other hand, TEAC accredits elementary-teacher education programs on the basis of an internal audit conducted by education faculty in light of TEAC principles that focus on the program's claims regarding candidates' subject-matter and pedagogical knowledge and teaching skills. Regardless of the accreditation body that a teacher-education program chooses, accreditation is voluntary. Canadian teacher-education programs submit their applications for accreditation to evaluation teams within their respective territories, followed by visits from those evaluators.

The quality of teacher education in the People's Republic of China falls directly under the purview of the State Education Commission (SEC). One of 23 of its subdivisions, the Teacher Education Bureau (TEB), formulates policy on teacher education. The SEC outlines the curricular framework for all normal institutions that are directly under the supervision of the SEC. The framework consists of foundation courses (i.e., politics, moral education, a second language, and physical education), professional education courses that include pedagogy, psychology, philosophy, history of education, and sociology; subject-matter specialization, replicating knowledge of a secondary teacher; and optional courses that may include art appreciation and extracurricular activities. The elementary-teacher education programs within the normal schools are accredited on the basis that their curricula are strictly aligned with policies set forth.

Universal Concerns and Problems with the Education of Elementary-School Teachers

Relationship between Educational Reform and Teacher Education

One of the central conversations in the field of elementary-teacher training, and teacher education more broadly, is the relationship between educational reform and teacher education. Scholars in the field of teacher preparation note the general trend of developing reform policies without the input or initial involvement of teachers at the preservice or in-service levels. Zeichner and Tabachnik find this reform approach common in Africa, where teacher education focuses on preparing teachers to exhibit the skills needed to implement reform that others have conceptualized, perpetuating authoritarian classroom patterns that many African teachers experienced

in their own education. In some rare cases, attempts to restructure or revamp teacher education are made before the introduction of a significant educational reform. Countries such as Papua New Guinea lengthened elementary-teacher education in the 1980s and 1990s, before instituting broad reform.

Professionalism

High-quality teacher preparation strongly predicts teacher retention and good teaching practice. Given this association, concern about the quality of teacher-education programs and the graduates of these institutions is common worldwide. In developing countries, especially in sub-Saharan Africa, where it is common for elementary-school teachers to start teaching without formal training, governments seek to increase teacher professionalism as part of the broader EFA reform agenda. In the United States, a major component of the No Child Left Behind (NCLB) legislation is the development and definition of a highly qualified teacher at every level including elementary.

Industrialized and developing countries face key challenges regarding the cultivation of quality primary-school educators, including cost, lack of institutional infrastructure, and the tension between increasing the number of teachers while maintaining (or improving) the quality of their preservice and in-service instruction. Inventive solutions have developed over the last 20 years to improve the quality of teacher education. Distance education or information and communication technology (ICT) has developed as one strategy to assist in preservice and in-service teacher training. India, for example, relies on distance-learning technology via EDUSAT, the world's first education satellite, to enhance the quality of primary teacher education.

Shortages

An additional problem facing countries around the globe is a persistent shortage of teachers at the elementary level; a dilemma that is particularly acute in developing regions experiencing a sharp increase in student enrollment, attempting to train millions of teachers to serve these students and trying to meet commitments to EFA. In many African countries, elementary-school enrollments have swollen beyond structural capacity due to the dropping of past mandatory school fees. For example, in Chad the number of students in school has doubled between 1995 and 2005. Schools located in less-desirable geographic locations, which include both rural and urban areas, or serving less-privileged populations (i.e., ethnic minorities, indigenous populations, and lower-income students) face the greatest teacher shortages, in terms of adequate training and supply, worldwide. In industrialized countries, such as the United States and the United

Kingdom, shortages align with specific subject areas and specializations, especially special education, foreign language, mathematics, and science.

These shortages have shaped the structure of PRETEP and INSTEP teacher-education and certification programs. In the hopes of attracting teachers, governments have shortened the initial teacher training cycle (e.g., Uganda, Tanzania, and Malawi) and lowered the admission requirements for the training (e.g., Mozambique). Lewin and Stuart highlight one example of a program designed to increase the number of elementary-school teachers is the Malawi Integrated In-service Teacher Education Program (MIITEP), which connects college-based education to school-based training and supplemental local and distance education sessions. In the United States, teacher shortages have spurred the development of alternative teacher certification programs and led to the recruitment of international teachers.

Conclusions

One of the central concerns of societies around the world is the availability and quality of their children's first years of school. In some countries, primary education constitutes the maximum schooling for a child. As we see in the goal of universal primary education, access to primary education, especially for girls, in many regions of the world remains a distant goal. In other countries, the elementary level becomes the foundation for future years of secondary and tertiary education. In both cases, the primary school teacher is of critical importance in the child's basic social, academic, and emotional development. This review of teacher education for elementary teachers examines the many ways that societies have agreed to prepare teachers for this central role. While the context of teacher education varies significantly around the world, broad agreement exists that, as future professionals, candidates need some specialized professional training in an institution that more frequently is housed in a university or other tertiary learning program. Another point of global convergence is that some system of quality control both for the candidate and the institute of teacher education should exist. Typically, state-controlled certification and institutional accreditation perform this function. Finally, a global look at teacher education for elementary teacher underscores the common problems and issues facing states concerning the development and improvement of public schools. Among the most pronounced of these problems are the persistent issue of how to best involve teacher education in system-wide school reform, the most effective and cost-effective ways to cultivate high-quality teachers, and the critical shortages of primary school teachers. At the start of twenty-first century,

we find cooperation between states, NGOs, and multinational organizations aimed at improving primary schools and elementary-teacher education around the world.

Further Reading

- Avalos, B. (2000). Policies for teacher education in developing countries. *International Journal of Educational Research* **33**, 457–474.
- Blömeke, S. (2006). Globalization and educational reform in German teacher education. *International Journal of Educational Research* **45**, 315–324.
- Cobb, V. L. (1999). An international comparison of teacher education. *ERIC Digest*.
- Cochran-Smith, M. and Zeichner, K. M. (eds.) (2005). *Studying Teacher Education: The Report of the AERA Panel on Research and Teacher Education*. Mahwah, NJ: Lawrence Erlbaum.
- Darling-Hammond, L. (1997). Doing what matters most: Investing in quality teaching. *ERIC Digest*.
- Darling-Hammond, L. and Bransford, L. (eds.) (2005). *Preparing Teachers for a Changing World: What Teachers Should Learn and be Prepared to Do*. San Francisco, CA: Wiley.
- Lewin, K. M. and Stuart, J. M. (2003). Insights into the policy and practice of teacher education in low-income countries: The multi-site teacher education research project. *British Educational Research Journal* **29**, 691–707.
- Moon, B. (2007). *Research Analysis: Attracting, Developing, and Retaining Effective Teachers: A Global Overview of Current Policies and Practices*.
- Morris, P. and Williamson, J. (eds.) (2000). *Teacher Education in the Asia-Pacific Region: A Comparative Study*. New York: Falmer Press.
- Schulz, R. (2005). The practicum: More than practice. *Canadian Journal of Education* **28**, 147–167.
- Shimahara, N. K. and Holowinsky, I. Z. (eds.) (1995). *Teacher Education in Industrialized Nations: Issues in Changing Social Contexts*. New York: Garland Publishing.
- UNESCO (2001). Teachers for tomorrow's schools: Analysis of the World Education Indicators. http://www.uis.unesco.org/TEMPLATE/pdf/wei_execsum_EN.pdf
- UNESCO Institute for Statistics (2002). Global Education Digest 2003 – Comparing Education Statistics across the World.
- Williamson, J. and Morris, P. (2000). Teacher education in the Asia-Pacific region: A comparative analysis. In Morris, P. and Williamson, J. (eds.) *Teacher Education in the Asia-Pacific Region*, pp 265–284. New York: Falmer Press.
- Zeichner, K. and Dahlström, L. (eds.) (1999). *Democratic Teacher Education Reform in Africa: The Case of Namibia*. Boulder, CO: Westview Press.
- Zeichner, K. and Tabachnik, B. (1999). Participatory development and teacher education reform in Namibia. In Zeichner, K. and Dahlström, L. (eds.) *Democratic Teacher Education Reform in Africa: The Case of Namibia*, pp 207–221. Boulder, CO: Westview Press.

Relevant Websites

- <http://www.capfe.gouv.qc.ca> – Committee Approval of Training Programs for Education. This webpage provides an overview of the accreditation process in Quebec.
- <http://unesdoc.unesco.org> – Education For All. Global Monitoring Report (2008) Summary: Education for All by 2015, Will We Make it?
- <http://www.acei.org> – The Association for Childhood Education International (ACEI). This website includes links to accreditation issues and procedures for elementary-teacher education programs.
- <http://www.ncate.org> – The National Council for Accreditation of Teacher Education (NCATE). This website includes a glossary related to accreditation in the United States and links to accreditation issues related to elementary education.

<http://www.oct.ca/home.aspx> – The Ontario College of Teachers. The links include information for finding a teacher, becoming a teacher, accreditation of teacher-education programs in Ontario, Canada.

<http://www.teac.org> – The Teacher Education Accreditation Council (TEAC). Links contain general information about TEAC and its accreditation processes.

www.unesco.org – The United Nations Educational, Scientific, and Cultural Organization (UNESCO). Website contains news, statistics, and other information that provide an international perspective on elementary education.

<http://portal.unesco.org> – UNESCO's Teacher Education Site.

Teacher Education for Teaching the Gifted

A Y Baldwin, University of Connecticut, Storrs, CT, USA

© 2010 Elsevier Ltd. All rights reserved.

Teacher Education for Teaching the Gifted

The international community has often taken its ideas regarding the gifted from the activities of those universities and persons actively involved in Education of the Gifted. The differences in the philosophies of the different nations have often made the US model the most difficult one to follow. Adjustments, realignments, and specific criteria for the quality or characteristics deemed necessary for teachers of the gifted are made; however, the evaluation of the criteria established is sorely lacking. Do we dare move forward in this century without a full understanding of the importance of the necessary attributes of teachers of the gifted within the context of the global environment? Attention must be paid to the emerging efforts to clarify these attributes. Although the context of this article is designed to focus on the teacher of the gifted, it is important to note that defining what we mean by the use of the term gifted is important.

Although the use of the term gifted and its definition have varied throughout the last three decades, the movement toward a more inclusive use of the term has begun to emerge. Previously, giftedness was inextricably tied to the IQ score and also judged by the academic achievements of the individual. Research and experiential examples have brought to the attention of proponents in this field that gifted characteristic can be exemplified in many different ways (Baldwin, 1984; Baldwin and Vialle, 1999; Renzulli, 1978; Ford, 1994; and Frasier, 1987). The observant teacher is often the person who recognizes this quality and advocates for inclusion of this student in programs for the gifted. However, too many students are not recognized because teachers who have not had training in this field are often tied to the notion that giftedness is synonymous with achievement and good behavior; therefore, they are led to believe that the absence of these qualities means that giftedness does not exist.

The field of Education of the Gifted has entered a new era where the qualities of teachers of the gifted are becoming the focus of attention. Although this change of focus is not limited to teachers of the gifted, it means that the qualities of teachers selected for teaching the gifted are dependent, in large part, on the quality of the training these teachers acquired during their preservice years or in-service and graduate programs.

Evaluation of the quality of training that preservice teachers have received is too often missing. For teachers of

the gifted, adjustments and realignments are often made in specific criteria for the quality or characteristics deemed necessary. Baldwin (1993) outlined the research and findings by Bishop (1968). Although this study was done quite some time ago, it is still referenced in contemporary research studies on teachers of the gifted. He reported on those characteristics which the subjects of his study said were present in successful teachers of the gifted. He determined that these teachers were experienced, of superior intellect, mature individuals, creative, had high personal achievement needs, and tended to be student oriented. This early determination of the characteristics of teachers of the gifted has been repeatedly found in subsequent studies of these characteristics.

Teacher Training

Concern about the issue of teacher training was highlighted in the winter 2007 issue of *Educational Horizons*. Each article was related to the experiences and training of teachers for this generation and contemporary times. Darling-Hammond and Baratz-Snowden (2007) in their summary of the National Academy of Education Committee on Teacher Education's report referred to the inequality of the training of teachers and posed the question of concern which is, "How is it that we permit so many ill-prepared individuals to assume such an important role in society?" (p. 111). Some of the answers to these questions were proposed as follows (p. 112):

- As a society, we do not invest seriously in the lives of children, most especially poor children and children of color who receive the least-prepared teachers.
- The conventional view of teaching is simplistic: teaching is viewed merely as proceeding through a set curriculum in a manner that transmits information from the teacher to the child.
- Many people do not understand what successful teaching requires, and do not see teaching as a difficult job that requires rigorous training.
- Others believe that there is not much more to teaching than knowing the subject matter that children should learn.
- Many state licensing systems reflect these attitudes and have entry requirements that lack demanding standards, especially for teachers who teach poor and minority students.

- Researchers and teacher educators have only recently come to consensus about what is necessary basic knowledge for entering the classroom and how and when such knowledge and skill should be acquired.

As editors of this research, Darling-Hammond and Baratz-Snoden commented, "Thus beginning teachers need consistent opportunities to apply what they are learning, to analyze what happens, and to adjust their efforts accordingly. They need to engage in inquiry and reflection about learning, teaching, and curriculum as well as direct instruction in specific areas of content" (p. 115). For teachers of the gifted, each of these answers is important. The inclusion of students of color in programs for the gifted, in large part, depends upon the type of instruction the student has had; therefore, well-defined criteria for excellence are needed.

Teacher Education Standards for Teaching the Gifted

The field of Education of the Gifted has grown to the point where the preservice training of teachers in understanding the characteristics of gifted students has become as essential as in-service training and advanced degrees in education of the gifted student.

Although standards-based programs have been successful in keeping teachers on track with skill development, in the field of education of the gifted it has denied the process of creative planning and development which is crucial in the effective educational experiences of gifted students.

The National Association for Gifted Children (NAGC) and the Council for Exceptional Children (CEC) cooperatively developed over a period of 3 years, teacher preparation standards for teachers of the gifted. These standards were approved by NCATE. The ten standards are listed under the title *NAGC-CEC Teacher Knowledge & Skill Standards for Gifted and Talented Education*. Each of the standards has subcategories that define what is expected in each standard. These standards are listed below with a short explanation of how each should be reflected in planning and teaching for gifted students (see NAGC report (2006) for listings of each subcategory and procedures for assessing each).

Standard 1: Foundations

Issues that deal with philosophies, theories, diverse and historical points of view, and human issues are the basis for this standard. The field of Education of the Gifted is fraught with divergent thinking about who is gifted and how to best provide for these needs. Each teacher needs to have a good foundation upon which to make decisions about the appropriate approach to teaching the gifted child in his or her class.

Standard 2: Development and Characteristics of Learners

Understanding the differences in the various learning capabilities of children is an expectation of all teachers but a crucial factor with teachers of the gifted. The range of differences within the classroom of gifted students requires knowledge of these learning differences and how they can be accommodated.

Standard 3: Individual Learning Differences

Culture and experiences play an important role in the learning differences of students. The role of nurture versus nature is significant in establishing the basic expectations of potential giftedness in a student.

Standard 4: Instructional Strategies

Having a repertoire of teaching strategies is crucial in differentiating lessons for gifted students. There is a need to engage all students in challenging, multicultural curricula.

Standard 5: Learning Environments and Social Interactions

The encouragement of independence of thought and the ability to live harmoniously with others of different cultures is an important role for teachers of the gifted.

Standard 6: Language and Communication

The teacher of the gifted should provide opportunities for oral and written communication that provide effective ways in which students can interact with their peer groups and show understanding and appreciation of the cultures of various groups.

Standard 7: Instructional Planning

The adaptation of materials that suit the learning requirements of the gifted child is important for the teacher of the gifted. The goals that are provided in the curriculum could be long range as well as short range but must be adapted to meet the needs of the learner.

Standard 8: Assessment

Assessment strategies are important in the general and specific analysis of the needs of the particular student. Identification of abilities is varied and teachers need to know the advantages or disadvantages of the various assessment strategies.

Standard 9: Professional and Ethical Practice

Teachers of the gifted respect the laws and regulations that refer to the teaching of students. They also maintain ethical behaviors when dealing and teaching students.

Standard 10: Collaboration

Working with professionals in the field helps teachers to share experiences and techniques that have been successful. This sharing and collaboration should be extended to parents and communities.

The training (related to methods), education (extension of knowledge), and evaluation (the quality of teachers' effectiveness with students) are all important concerns in a teacher's effectiveness in the classroom. The movement by NAGC and CEC toward designing a set of standards for teachers of the gifted recognizes those qualities that indicate effectiveness.

Among the many things of the legacy of Lee Shulman, until recently the president of the Carnegie Foundation for the advancement of teacher education, is the advanced study of teacher education programs, to bring out the richness, beauty, and complexity of teaching by making it visible as is done in other professions (Carnegie Foundation News, 2007). It is important that programs to train teachers of the gifted design the teacher education programs so that this richness, beauty, and complexity can be recognized.

Research Available on Qualities and Training of Teachers of the Gifted

Questions still remain regarding the characteristics that are specific for teachers of the gifted. Baldwin (1993) expressed at that time that there was a need for answers to the continuing flow of questions regarding the desired abilities of teachers of the gifted. Were these characteristics any different from ones desired for all children? These questions that still need empirical answers are (p. 621):

- Should the teacher of the gifted be highly gifted, if so what should the IQ score be?
- How extensive should the knowledge of the teacher of the gifted be?
- Should the basic characteristics required to teach gifted elementary grade students be the same for gifted students at all levels?
- What type of educational background should the teacher of the gifted have?
- What is the difference between a teacher of the gifted and a good teacher of all the other students?
- What specialized training should teachers of the gifted have?

The data cited by Baldwin (1993) and Baldwin *et al.* (2000) contain nomenclature that provides a continuing

perception of what the teacher of the gifted should be. These perceptions are solid enough to give institutions of higher education a good road map for developing appropriate selection strategies, courses, and evaluation of the students at the conclusion of this coursework.

Research Methods and Findings

The methodologies used to find teacher characteristics have mainly been qualitative. Methods, such as case studies, narrative studies, surveys, correlational studies, and causal/experimental studies, have been used and are still being represented in the most recent studies mentioned here. Gentry (submitted) emphasized in her study that “with qualitative research, the intent is not to generalize but to inform” (p. 28). She has proposed that the research from her study, that included 400 teachers out of which a group of 18 were said to be exemplary teachers, should give future researchers and persons concerned about qualities of teachers of the gifted basic data from which to conduct further study. This would also be a starting point for preservice and postservice training program design. Gentry's research was designed to determine the perceptions of students about those teachers that students considered exemplary.

Four themes emerged from her research as follows:

- Theme 1: These teachers know and take a personal interest in their students.
- Theme 2: These teachers set high expectations for themselves and for their students.
- Theme 3: These teachers make content and learning meaningful and relevant to the future and respect students' choices.
- Theme 4: These teachers have a clear passion for their students, teaching, and for their content.

These themes that reflected student's perceptions of exemplary behaviors of teachers of the gifted mirrored the behaviors that other researchers (e.g., Bishop, 1980; Baldwin *et al.*, 2000; Whitlock and DuCette, 1989) have also indicated in their research.

In Gentry's research, high expectations of teachers and their students were ranked highest among the most desirable traits listed in the themes above.

Findings from this study raise several questions for future research and planning on behalf of Teacher Education Universities. Can teachers be taught how to develop positive relationships with students? What role does humor play in teaching and learning in a broader context? What might Career and Technical Education and teachers who have professional experiences offer education in general? Can the attributes of the exemplary teachers be used to develop better pre-service and in-service

teachers? How can passion for content, students, and teaching be assessed and used to recruit and retain quality teachers? Gentry (submitted)

In an effort to pinpoint qualities that reflect teachers of the gifted, Vialle and Quigley (2002) found in their research ($n = 377$) regarding the perceptions of children toward their teachers, that personal and social characteristics of teachers were highly regarded. The perceptions of students in both of these studies give teacher education faculty information that must be considered in the process of training teachers of the gifted.

Another study was undertaken to determine the effectiveness of two experiences: one of a specific practicum within a Saturday Program for gifted students and the other an online course to provide this knowledge base of giftedness to undergraduate preservice teachers (Bangel *et al.*, 2006). In their study of two different methods of preparing preservice teachers in education of the gifted, they concluded that the standard teacher education practicums do not include the necessary experiences with gifted students. "Teacher education programs strive to prepare novices to become high-quality teachers for our youth. Among those youth will be a number of gifted students who desire to have appropriate education opportunities but have not had them, due to lack of training for their teachers" (p. 356–357).

Desired Traits for Teachers of the Gifted

Over and over again, scholars in the field have listed those characteristics they think are important for teachers of the gifted. Although empirical research is limited in this area, observations and modified studies have indicated similar theories about the needs of the gifted teacher. In **Table 1**, many of these theories are listed under the suggestion of Maker (1982) of three areas that she felt were the important ones to be used in analyzing a teacher's preparedness to teach the gifted as well as a means of addressing training for teachers of the gifted. These areas are:

- *Philosophical* which focuses on the teacher's feelings about education of the gifted. What are the ideals and beliefs about education of the gifted to be considered in this area?
- *Professional* which focuses on the type of training necessary for the teacher of the gifted. What type of courses, professor's knowledge, and practicum experiences are needed?
- *Personal/innate quality/personality* what qualities are present that would make an individual a good prospect for teaching the gifted student?

Listing of themes by Vialle and Quigley (2002) coincided with Maker's ideas closely and described their three

Table 1 Classification of behaviors for teachers of the gifted in three basic areas

Basic area	Behaviors
Philosophical	Positive attitudes toward the concept of giftedness Accurate concept of giftedness Personal beliefs regarding education of the gifted Willing to advocate for the gifted Dedicated to help students reach their potential
Professional	In-depth knowledge of subject matter Knowledge of varying learning styles Works with students to reach their goals Uses creative methods to find solutions to problems Remains alert to educational tools and use of them Able to stimulate students Professional training in education of gifted IQ score of 128 or above Flexible with use of time Provides proper environmental support Well organized in planning Culturally responsive in planning curricula Displays gifted behaviors
Personal/innate quality/personality	Willing to take risks Self confident Perceives emotion and accepts students emotion Warm and friendly with support of students Sense of humor Fair in making decision with students Enthusiastic High verbal skills Introspective

areas as personal–social characteristics; teaching strategies/approach; and intellectual–cognitive characteristics. Their listing of numerous sources indicate that there is great consensus about those characteristics deemed necessary for the teacher of the gifted in each of these three areas. These characteristics are needed for all teachers; however, following model revision and caution for models of teacher training by Dall'Alba and Sanberg (2006), the additional qualities of flexibility, willingness to take risks, acceptance of ambiguity, respect for student ideas, out-of-the-box thinking, acceptance of differences of opinions, and encouragement of students to pursue different trajectories of learning are difficult to define but necessary additional characteristics for teachers of the gifted.

Teacher Education Programs Challenges

In selecting and training teachers, personal practical knowledge is crucial because in spite of the planned

curriculum, he or she will need to bring to the classroom their way “of reconstructing the past and the intentions of the future to deal with the exigencies of the present situation” Connelly and Clandinin (as cited in Craig (2006)).

Meeting the Needs of Gifted Children from Various Ethnic Groups

Giftedness can be expressed in many ways. There are many faces of the gifted and we should lift the masks as suggested by Baldwin and Vialle (1999). Defining giftedness and identification of giftedness have become the Achilles heel of efforts to show that giftedness can be exhibited in many ways. Inclusion in programs of children from different cultures is presently a great concern for professionals in this field. Scott *et al.* (1996) reported that, “in the United States of America, children from culturally different and/or low socioeconomic environments constitute a growing percentage of all students, yet assessment tools that effectively evaluate their academic potential are lacking” (p. 147). Little has changed since that time; therefore, advocacy from teachers plays a large role in recognizing the latent abilities of these children. More than three decades ago, the problem of gifted black students in gifted classes (the term black is now referred to as African-American) was addressed when several students were to be selected for the first class for black gifted students in this segregated school system. The identification process used at that time would have eliminated all but three of the students of this class. Advocacy and a different approach to identifying these students were used. A challenging curriculum, exposure to new ideas, and an opportunity to engage in exciting and stimulating activities showed on subsequent achievement tests, that effective experiences within the classroom had allowed the latent abilities to bloom. These students were diamonds in the rough (Baldwin, 1977).

As Baldwin and Vialle (1999) have stated, giftedness can be evident among students of different cultures, physical impairments, language differences, and socioeconomic circumstances. These imaginary masks have often kept students from being recognized as gifted. Teachers of tomorrow must have the training and experience needed to lift these masks.

Summary and Recommendations

The proper education of the gifted requires that teacher training programs for the preservice teachers become an important goal for teacher education programs in higher education institutions and local educational centers and school districts. The present focus on teacher training mirrors the previous focus that was on the gifted students

and their needs. There has always been a need for gifted students to have the opportunity to explore new ideas and to work with students of similar ability. For teachers, it is also important that they be given the necessary training that will free them to explore new ideas with the understanding of the myriad of differences in the interests and needs of these students. Meeting their needs is not an elitist concept but one that is egalitarian meaning that all children should have the opportunity to develop his or her abilities as far as possible.

The new standards for teachers of the gifted are guidelines within which programs and teachers are assessed; however, these guidelines are only the beginning for the teacher of the gifted. Teaching is an art which mirrors the same color, drama, and excellence of achievement that is felt by the conductor of a symphony, the producer of a play, or the display of a graphic artist. The qualities inherent in these achievements are difficult to quantify but must be there in successful teachers of the gifted.

We are moving toward improving the training of teachers of the gifted but we have miles to go in the process. The tools for this process are available and must be put into action in order to capture the great potential for the future.

Bibliography

- Baldwin, A. Y. (1977). Tests can underpredict: A case study. *Phi Delta Kappan* **58**, 620–621.
- Baldwin, A. Y. (1984). *The Baldwin Identification Matrix 2 for the Identification of the Gifted and Talented: A Handbook for Its Use*. New York: Trillium Press.
- Baldwin, A. Y. (1993). Teachers of the gifted. In Heller, K. A., Monks, F. J., and Passow, A. H. (eds.) *International Handbook of Research and Development of Giftedness and Talent*, pp 621–629. Tarrytown, NY: Pergamon.
- Baldwin, A. Y. and Vialle, W. (eds.) (1999). *The Many Faces of Giftedness: Lifting the Masks*. Belmont, CA: Wadsworth.
- Baldwin, A. Y., Vialle, W., and Clarke, C. (2000). Global professionalism and perceptions of teachers of the gifted. In Heller, K. A., Monks, F. J., Sternberg, R. J., and Subotnik, R. F. (eds.) *International Handbook of Giftedness and Talent*, 2nd edn., pp 565–572. Amsterdam: Elsevier Science.
- Bangel, N. J., Emersen, D., Capobianco, B., and Moon, S. M. (2006). *Journal for the Education of the Gifted* **29**, 139–361.
- Bishop, W. (1968). Successful teachers of the gifted. *Exceptional Children* **34**, 317–325.
- Carnegie Foundation News (2007). Lee, S. Shulman to Retire from Carnegie Presidency. info@carnegiefoundation.org.
- Craig, C. (2006). *American Education Research Journal* **43**, 257–293.
- Dall’Alba, G. and Sandberg, J. (2006). Unveiling professional development: A critical review of stage models. *Review of Educational Research* **76**, 383–412.
- Darling-Hammond, L. and Baratz-Snowden, J. (eds.) (2007). A good teacher in every classroom: Preparing the highly qualified teachers our children deserve. National Academy of Education Committee on Teacher Education. *Educational Horizons: The Official Publication of Pi Lambda Theta* **85**, 11–132.
- Frasier, M. M. (1987). The identification of gifted black students: Developing new perspectives. *Journal for the Education of the Gifted* **10**, 155–180.

- Ford, D. (1994). Research-based Decision Making Series. *The Recruitment and Retention of Black Students in Gifted Programs*. Storrs, CT: The National Research Center on the Gifted and Talented, The University of Connecticut.
- Gentry M. *Student-Identified Exemplary Teachers: What Makes Them Tick?* (out for review).
- Maker, C. J. (1982). *Curriculum Development for the Gifted*. Rockville, MD: Aspen systems Corporation.
- Renzulli, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan* **60**, 180–184.
- Scott, M. S., Duel, L. S., and Urbano, R. C. (1996). Identifying cognitively gifted ethnic minority children. *Gifted Child Quarterly* **40**, 147–153.
- Vialle, W. and Quigley, S. (2002). Does the teacher of the gifted need to be gifted. *Gifted and Talented International* **17**(2), 85–90.
- Whitlock, M. S. and DuCette, J. P. (1989). Outstanding and average teachers of the gifted: A comparative study. *Gifted Child Quarterly* **33**, 15–21.
- Baldwin, A. Y. (ed.) (2004). *Culturally Diverse and Underserved Populations of Gifted Students*. Thousand Oaks, CA: Corwin Press.
- Baldwin, A. Y. (2006). Diversity: Perceptions of pre-service teachers – A cause for concern for gifted minorities. In Belle, W. and Gillian, E. (eds.) *Diversity in Gifted Education: International Perspectives on Global Issues*. Abington: Routledge.
- Colangelo, N., Assouline, S. G., and Gross, U. M. (eds.) (2004). *A Nation Deceived: How Schools Hold Back America's Brightest Students*. vols. I and II. Iowa City, IA: University of Iowa.
- Karnes, F. and Parker, J. (1983). Teacher certification in gifted education: The state of the art and considerations for the future. *Roeper Review* **6**, 18–21.
- Reis, S. (1991). The need for clarification in research designed to examine gender differences in achievement and accomplishment. *Roeper Review* **13**, 193–198.
- Vialle, W. and Tischler, K. (2005). Teachers of the gifted: A comparison of students' perspectives in Australia, Austria, and the United States. *Gifted Education International* **19**(2), 173–181.

Further Reading

- Baldwin, A. Y. (1986). *Teacher Training in Developing Countries. Giftedness a Continuing Worldwide Challenge*, pp 434–439. New York: Trillium Press (now Royal Fireworks Press).

Relevant Website

- <http://www.NAGC.org> – NAGC–CEC Teacher Knowledge and Skill Standards.

Teacher Preparation in Mother Tongue Education

M Kooy, University of Toronto, Toronto, ON, Canada

H Hulshof, ICLON, Leiden University Graduate School of Teaching, Leiden, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

The article offers an overview of the evolving state of teacher education in mother tongue (L1), the home language of nations (e.g., Greek in Greece and English in Australia). Mother tongue education's focus on language, which reflects a nation's culture and identity (as expression and medium of communication), lies at the heart of schooling. Unlike other subject areas, it functions as both the medium and object of study. Not surprisingly, differences that exist on the place and purpose of mother tongue education remains the most debated and contested school subject. Little agreement on the specific subject matter content and definitions makes it persistently difficult to define a shared understanding of the core concepts and skills of the school subject. Understandably, this affects and even shapes how teacher education in mother tongue occurs in the university programs (Luke, 2000). The complex, often parallel, and even competing views and practices of mother tongue education affect the ability to clearly spell out the state of a field that is itself multiple, varied, and rapidly changing.

The conventional theory/practice divide between the university and the school results in differences between academic content of mother tongue in university programs and mother tongue subject content in schools. Hence, what preservice teachers learn before entry into mother tongue subject teaching in their field experiences in schools may not transition smoothly into schools but become a parallel body of knowledge (O'Neill, 2000). The knowledge of linguistics and literature acquired at the university, therefore, may differ significantly from the content knowledge required in the mother tongue as school subject.

Developments in educational theory regarding the relations between the what (subject content) and the how (teaching abilities) matter to mother tongue education in that they provide guidelines for this pedagogical translation. Global efforts in mother tongue teacher education, then, share the goal of approaching academic study of mother tongue in order to socially relevant educational practice (Darling-Hammond and Bransford, 2005).

Not surprisingly, university-based teacher education in mother tongue varies in programs and practices within and between nations. Nevertheless, most programs share three consistent features: (1) subject-matter knowledge background; (2) professional and pedagogical knowledge development; and (3) field-based (student teaching) experience(s) in mother tongue education. The transition from content knowledge to pedagogical knowledge

predominates in teacher education since it strives to bridge academic content knowledge with knowledge of mother tongue as school subject and its pedagogical practices.

Mother Tongue Subject-Matter Knowledge

Content and background knowledge in mother tongue required for admission to teacher education ranges from a few undergraduate university courses in literature and language (such as at the University of Toronto) to a master's degree (such as in the Netherlands). Generally, evidence of mother tongue competence comes in the form of university records (unlike some programs requiring a competency test such as in second-language education). The assumption that mother tongue content knowledge translates into the knowledge required for creating mother tongue teaching strategies in the school subject matter often fails to materialize. Some research is beginning to identify specific, translatable subject-matter content knowledge specifically directed to mother tongue school subject matter. Nevertheless, determining specifically what background content knowledge applies most directly to mother tongue teaching in schools remains underinvestigated and requires further inquiry and research (Lucas and Grinberg, 2008).

The subject mother tongue in schools (and by association, preservice teacher education) remains particularly contentious in many national contexts, and becomes the site for ongoing (often political) debate. The conflict occurs around what constitutes mother tongue subject matter and what competencies and skills students are expected to achieve. These debates affect all stakeholders in the educational process (e.g., from preservice teachers to students and policymakers). Periodic shifts and trends in economies and ruling political parties call for increased competence in mother tongue to bolster the edge of competitiveness in the world. These competencies are often measured in test results (a particularly pointed example of this is the current no-child-left-behind policies in the United States).

Mother tongue as subject remains contested because of its central place in education; that is, expression and reflection of a national culture occurs through language. Since all subject-matter teaching and learning in schools (and elsewhere) occurs through language, mother tongue serves both as the source of communication and as an

object of study. In addition, given the recognition that literacy represents a level of power not available to all, access becomes an issue (e.g., Freire's work in South America and denial of literacy for girls in Afghanistan under the Taliban). The complicated network of variables that influence and shape what counts as subject-matter knowledge remains a major issue for mother tongue subject in preservice teacher education.

Professional and Pedagogical Knowledge Development

Professional education for preservice teachers in mother tongue education explores how background knowledge and content translates into teaching subject matter. Indeed, translating disciplinary content knowledge into pedagogy remains the singularly most important purpose of mother tongue methods courses.

"Knowing the subject matter terrain is the foundation for pedagogical content knowledge" (Grossman and Schoenfield, 2005: 204, 205) and a teacher's "content knowledge organized for teaching" is a better predictor of successful teaching than content knowledge alone (Grossman and Schoenfield, 2005: 206). Grossman (1989) compared the effect of subject-specific coursework in pedagogical content knowledge (PCK) by contrasting six beginning English teachers, only three of whom graduated from teacher education. She demonstrates that subject-specific coursework can be a powerful influence on how teachers think about and teach their subjects. Given this, it remains critical to identify the specific mother tongue content knowledge required for developing sound pedagogical knowledge for mother tongue teaching (Hulshof and Verloop, 2002, 2005), although precisely what knowledge remains to be determined.

Shulman's conception of PCK (in 1986), remains a key feature in understanding the nature of practices in mother tongue teaching. This understanding moves beyond subject knowledge to specific subject knowledge that matters for teaching. Shulman (2004) understands this knowledge as "ways of representing and formulating the subject that make it comprehensible to others" (p. 203). The European tradition of didactics (pedagogy) has always been associated with the why and what questions of teaching. Neither purely an academic discipline nor an educational theory, it represents a conversation or a bridge between the two (Freire calls this praxis). Hence, content in teaching has always played a significant role in thinking about pedagogical reasoning. This undoubtedly influences the European perception and view of PCK related to mother tongue education (Hulshof and Verloop, 1996).

What students are expected to achieve in skill and competence in mother tongue education undergoes

periodic shifts. The common recurring belief (often political) that students lag behind their counterparts in other parts of the world in language and literacy skills brings a surge of remedial new policies, programs, and practices targeted to raise student scores and make them more globally competitive.

Decisions by policymakers reflect conceptions of purposes, practices, and models of mother tongue education. Responding to the call to upgrade and improve student performance frequently leads to imposing a skill-and-drill model of mother tongue education (e.g., traditional, canonical texts such as Shakespeare, the study of grammar, and a focus on language skills). Frequently, high-stakes testing drives the curriculum and the test results provide evidence for improvement (the US no-child-left-behind policy provides an example). University teacher educators and researchers often offer models of mother tongue reflecting a sociocultural, critical literacy perspective (e.g., texts represented in technology, media, and graphic novels; writing for understanding (process rather than product orientation); and language and meaning as socially constructed). The two orientations reflect distinctively different practices and pedagogies. This tends to perpetuate the divide existing between university mother tongue teacher-preparation courses and the realities of practices in the schools.

A current exception to this gap appears to be Finland. Its top international ranking in mother tongue language (Linnaky, 1993; Valijarvi *et al.*, 2002) begs the question: How does the Finnish school system determine subject-matter content and pedagogical practices? Who makes such decisions? How does this affect mother tongue teacher education? Perhaps Finland's decentralized school system and its considerable autonomy given to teachers make collaborative decision making and agreement about shared practices and subject-matter content more realistic and possible. Perhaps the relationship between the schools and universities minimizes the traditional gap. A close examination of its mother tongue preservice teacher programs is warranted. As mother tongue is both medium of communication and object of study, comprehending the implications of PCK for mother tongue education is very complex.

A robust understanding of PCK takes into account the intersection of subject content, cognition, and students and includes, among other things, the ability to anticipate and respond to typical student patterns of understanding and misunderstanding by generating multiple examples and representations of challenging topics that make content accessible to a wide range of learners. For mother tongue language education, this suggests learning multiple and alternative approaches and strategies to literacy and language (primarily reading and writing).

Three significant issues affect how L1 coursework is structured and practiced in teacher education: (1) content

of methods courses; (2) increasing diversity and languages in L1 classrooms; and (3) movement from literacy to literacies.

Mother Tongue Education Methods Courses

Methods courses vary among nations and within mother tongue education methods courses. The inconsistency is particularly noteworthy because mother tongue education is central to schooling; mother tongue education courses often represent the core requirement for every year of secondary schooling (e.g., in Canada, mother tongue English or French represents the only mandatory subject in every secondary grade). The importance of understanding how to create links and transitions to pedagogical practices in preparing teachers for mother tongue education in the schools cannot be understated.

In Europe, mother tongue methods courses tend to use similar content structures: (1) objectives, content, and examination of the school subject; (2) nature and history of the school subject; (3) innovations in subject-specific pedagogy; (4) curriculum development; (5) lesson preparation; (6) educational psychology; (7) textbook analysis; (8) relationship between academic discipline and school subject; and (9) research in the school subject and preliminary research skills.

Decisions on course content and practices for mother tongue education are particularly complicated because of a lack of common and shared standards and a number of versions of mother tongue education that continue to exist. Barnes (1984) summarized these as: (1) basic skills; (2) personal growth (Dixon's (1967) model); (3) disciplinary; and (4) critical literacy. Each model reflects a particular theoretical framework and the resulting values, priorities, practices, and expectations particular to it. Questions arise for mother tongue education courses: What model is privileged? Which one(s) is ignored or undermined? Does the university program aim to prepare preservice mother tongue teachers for the realities of the classroom (reflecting what happens in schools) or is it required to expose students to the four models cited above, or should it be responsible for advancing the most recent research that includes a postmodern perspective and realities of a changing world and cultures (such as the increasing diversity and languages represented in many urban classrooms)? The conundrum captured in the persistent lack of agreement about what constitutes viable and defensible mother tongue subject and practices fuels ongoing debates on the course content and practices in the university program.

School realities and university ideologies often clash and, in doing so, perpetuate the existing chasm separating the two. Teachers in preparation programs are vulnerable, apt to accept (at least in the abstract) models proposed in university methods courses. A lack of applied and

practical experience (e.g., reading about critical literacy as a sufficient condition for understanding and transforming the knowledge into pedagogy) prevents drawing from earlier experience and knowledge to make decisions based on extensive teacher knowledge. Mother tongue methods courses, then, precariously balance the decisions about what to include and exclude in the preparation of teachers.

To make the transformation of content knowledge a reality requires participating in the practices and pedagogies of mother tongue education. When complemented by readings of current perspectives and discussions in the class, the methods course draws on the intellectual and practical elements of learning to teach (Smagorinsky *et al.*, 2003). Through personal engagement and reflection, it becomes possible for preservice teachers in mother tongue to imagine and transform their content knowledge (Mansvelder-Longayroux, 2006).

Mother Tongue Teacher Education in/for a Changing World

The changing demographics in many urban contexts make an already-complex subject area even more unwieldy (Beavis and O'Mara, 2006; Green, 2006). Toronto, the world's most multicultural city, has classrooms with many, even no children born in Canada. This also increasingly becomes the case and even the norm in some (particularly urban) schools in many other nations (Darling-Hammond and Bransford, 2005). Preparing teachers for these new realities is often neglected, even ignored, at the preservice level. Preservice mother tongue education teachers are often struck by how unprepared they are for teaching in diverse classrooms (Kooy, 2006).

The shifting classroom demographics raise critical questions: How are mother tongue teachers being prepared to teach for this ever-changing world? What approaches, perspectives, and practices reflect what "prospective teachers need to [do to] reorganize their subject matter knowledge into knowledge about how to teach subject matter to diverse students" (Bransford *et al.*, 2005: 15; Lucas and Grinberg, 2008)? How will teacher-education programs attract a more diverse teaching population (which remains primarily white and female, at least in Canada and the United States) that more clearly reflects the student populations?

The growth of transnational movements requires remaking mother tongue education courses to meet the changing demographics of classrooms and its implications for teaching and learning. Cochran-Smith (2004) recommends transforming teacher preparation "informed by the large bodies of work that now exist about . . . the preparation of teachers for a diverse society, which come from critical and multicultural perspectives intended to interrupt the norms

of conventional teacher education” (p. 212). Inevitably, this focuses on (mother tongue) language.

From Literacy to Literacies in Mother Tongue Education

Conventionally defined and practiced in mother tongue education, literacy involved reading and writing – that is, developing skills in learning how to read and the study of reading and writing (reception and production) of school texts. Current social and technology trends are redefining and reshaping traditional understanding of school literacy from the singular to plural literacies. Socially, this takes into account transnational movements and immigration that continue to change the demographics of classrooms. Diverse multilingual student populations bring multiple literacies into mother tongue classrooms. Parallel to this reality is a growing sense that *literacy* continues to move into *literacies* with the advent of technology requiring skills, such as computer literacy, and becoming literate in multiple new ways such as newer text forms (e.g., video games, film, and modern graphic novels). Green (2006) argues that recent questions concerning literacy and technology “represent quite critical, crucial challenges . . . reaching to the very heart of English [sic] teaching and English studies in education” (p. 371). Beavis and O’Mara (2006) add that, “literacy and technology are perhaps the central matters of concern for English teaching today” (p. 253). In such changing realities, understanding how to engage students in multiple literacies (Luke, 2000; Comber and Reid, 2006) that more clearly reflect the social and textual realities of student and teacher lives is becoming increasingly important.

Incorporating new knowledge to more adequately prepare mother tongue teachers includes such possibilities as adding a minor or supplemental certificate, a course specifically dealing with multiple languages and cultures of diverse classrooms, or requiring a foreign language credit prior to admission to teacher preparation (particularly mother tongue teacher courses) for diverse student populations (Lucas and Grinberg, 2008: 625). The need to adjust mother tongue education in the universities for effective and meaningful teacher preparation must include the recognition that globalization, changing classrooms, and the accompanying social and technological changes require new ways of approaching mother tongue education in schools.

Field-Based Experiences

Preservice mother tongue education teachers shape their teacher knowledge through the intersection of three lenses: background content knowledge, university methods courses, and field experiences. Field experiences often constitute the highlight of the teacher-preparation

experience since here opportunities to learn in and from practice occur. In the classroom, preservice teachers experience the subject in action. Examples of subject matter used in the interaction of theory and practice include: working with different levels of literacy competences, application of new strategies for reading comprehension instruction, and introducing poetry and linguistic subjects in the classroom. Mother tongue teacher education in the Netherlands, for instance, begins at the university with discussions about the subjects in methods courses. Field experiences in the schools follow where mother tongue education teachers, with guided application, apply the principles and practices of these related and relevant mother tongue subject areas at school. Finally, the preservice teachers use a portfolio to reflect on their performance (Mansvelder-Longayroux, 2006). Since field experiences affect and, to varying degrees, shape perceptions and perspectives of mother tongue as school subject, they play a significant role in mother tongue teacher preparation.

Competing models of mother tongue education (see above) and to the significant gap existing between university programs and the choices of schools and supervising teachers for field experiences. Preservice mother tongue teachers often enter schools and teachers that differ from the idealized versions of mother tongue in the university. School realities often reflect the incredible pressures of mother tongue teachers to demonstrate student competence and improvement in literacy and language (e.g., through test results). The preservice teachers witness mother tongue education practiced with an “ethic more geared toward coverage and control” and a “diminishing value placed on student-centered teaching methods” (Smagorinsky *et al.*, 2003: 22). Preservice teachers can feel caught between competing models of mother tongue education and are challenged to make sense and meaning of the field. Nevertheless, the knowledge of the tensions between the two institutional cultures and ideologies remains; some work on reconciling and collaborating is beginning to emerge (Clark, 2001; Kooy, 2006).

The call is out for more collaborative negotiations to align mother tongue education as a methods course at the university and mother tongue as school subject in the schools. The lack of adequate supervisory teachers for effective field experiences complicates the issue further. Unless such connections occur, the mother tongue teaching preparation can be parallel rather than integrated experiences to build teacher learning.

Concluding Remarks

Mother tongue education is in a state of flux, even turmoil as it struggles to find its way and map out its place. The paradigm wars and the changing social and technological advances seriously challenge conventional ways

of teaching mother tongue. The duel between traditional and postmodern conceptions of mother tongue education persists (Mitchell *et al.*, 2006). This leaves a conundrum that presently has little solution or agreement among those interested in mother tongue education in schools. What effect does this tension have on successful and meaningful mother tongue education? Can we agree on the kind of professional content knowledge in mother tongue education that offers a pedagogical translation into the profession and practices that prepares for authentic teaching and learning?

The tensions, in part, arise from political and neo-conservative movements in many nations and schools that continue to alter the face of mother tongue education (see Doecke *et al.* (2007) for an example in Australia). Public and political calls for increasing competence in literacy (assessed through high-stakes testing) and the pervasive lack of agreement on the nature and purposes of mother tongue education contribute to the uncertainty of the field. Nevertheless, global attention to literacy implies a need for university–school collaborations to cultivate stronger ties and work together to responsibly and effectively prepare preservice teachers of mother tongue for diverse languages and cultures, texts, and contexts, in light of the new and changing social realities of their classrooms (Lucas and Grinberg, 2008).

In many national contexts, preparing mother tongue teaching remains problematic. Since mother tongue/language is central within school curricula – the heart of any given culture – its purposes, practices, and ideologies significantly impact meaningful participation in society. Clearly, this shared issue calls for large-scale comparative, international studies. Currently, the field is characterized by many small-scale, one-off, and often (though not necessarily) qualitative studies. A lack of research funding results in a fragmented research field without the mechanisms for cumulative or comparative building of knowledge, or a strong theory of teacher education. Looking across institutional, multinational contexts offers more possibilities for cohesion and general research conclusions (Wilson *et al.*, 2001). Such research knowledge would provide opportunities to discuss intersections and differences in mother tongue education that, in turn, could affect mother tongue methods courses. Larger-scale, multinational research studies have the potential to build a repertoire of rigorous research that can inform policy debates, government inquiries, and curriculum designers across varied contexts (Mitchell *et al.*, 2006). The critical nature and centrality of mother tongue education requires finding defensible and shared ways to prepare teachers that reflect the changing nature, knowledge, and practices of mother tongue education in which teacher knowledge and skill make translation from teacher knowledge to effective and transformational student learning experiences possible.

Bibliography

- Barnes, D. (1984). *From Communication to Curriculum*. London, UK: Penguin Publishing.
- Beavis, C. and O'Mara, J. (2006). Preparing English teachers for a changing world. In Doecke, B., Howie, M., and Sawyer, W. (eds.) *Only Connect: English Teaching Schooling and Community*, pp 349–355. Adelaide: Australian Association for the Teachers of English/Wakefield Press.
- Bransford, J., Darling-Hammond, L., and LePage, P. (2005). Introduction. *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do*, pp 1–39. San Francisco, CA: Jossey-Bass.
- Clark, C. (2001). *Talking Shop: Authentic Conversation and Teacher Learning*. New York: Teachers College Press.
- Cochran-Smith, M. (2004). *Policy, Practice, and Politics in Teacher Education: Editorials from the Journal of Teacher Education*. Thousand Oaks, CA: Crowin Press/American Association for Colleges of Teacher Education.
- Comber, B. and Reid, J. (2006). Literacy education: What's new and what needs to be new? In Doecke, B., Howie, M., and Sawyer, W. (eds.) *Only Connect: English Teaching Schooling and Community*, pp 335–348. Adelaide: Australian Association for the Teachers of English/Wakefield Press.
- Darling-Hammond, L. and Bransford, J. (eds.) (2005). *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do*. San Francisco, CA: Jossey-Bass.
- Dixon, J. (1967). *Growth Through English: A Report on the 1966 Dartmouth Seminar*. South Yorkshire: National Association for the Teaching of English.
- Doecke, B., Howie, M., and Sawyer, W. (2006). *Only Connect: English Teaching, Schooling and Community*. Kent Town, SA: Wakefield Press in association with the Australian Association for the Teaching of English.
- Green, B. (2006). Afterword. In Doecke, B., Howie, M., and Sawyer, W. (eds.) *Only Connect: English Teaching, Schooling and Community*, pp 370–371. Adelaide: Wakefield Press/AATE.
- Grossman, P. and Schoenfeld, A. (2005). Teaching subject matter. In Darling-Hammond, L. and Bransford, J. (eds.) *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do*, pp 201–231. San Francisco, CA: Jossey-Bass.
- Grossman, P. L. (1989). A study in contrast: Sources of pedagogical content knowledge for secondary English. *Journal of Teacher Education* 40(1), 24–31.
- Hulshof, H. and Verloop, N. (1996). Content in teaching: The relationship between the evolving concept of 'pedagogical content knowledge' and the European 'didactics' tradition. In Ephraty, N. and Lidor, R. (eds.) *Teacher Education: Stability, Evolution and Revolution. Proceedings of the Second International Conference*, pp 445–463. Israel: Wingate Institute.
- Hulshof, H. and Verloop, N. (2002). The use of analogies in language teaching: Representing the content of teachers' practical knowledge. *Journal of Curriculum Studies* 34, 77–90.
- Hulshof, H. and Verloop, N. (2005). Do teachers need instructional analogies to teach mother-tongue (L1)? *L1 – Educational Studies in Language and Literature* 5, 23–37.
- Kooy, M. (2006). *Telling Stories in Book Clubs: Women Teachers and Professional Development*. New York: Springer.
- Linnaky, P. (1993). Exploring the secret of Finnish reading literacy achievement. *Scandinavian Journal of Educational Research* 37(1), 63–74.
- Lucas, T. and Grinberg, J. (2008). Responding to the linguistic reality of mainstream classrooms: Preparing all teachers to teach English language learners. In Cochran-Smith, M., Feiman-Nemser, S., and McIntyre, D. (eds.) *Handbook of Research on Teacher Education: Enduring Questions in Changing Contexts*, 3rd edn, pp 606–636. New York: Routledge.
- Luke, C. (2000). New literacies in teacher education. *Journal of Adolescent and Adult Literacy* 43(5), 424–435.

- Mansvelder-Longayroux, D. (2006). *The Learning Portfolio as a Tool for Stimulating Reflection by Student Teachers*. Thesis, Leiden University, The Netherlands.
- Mitchell, J., Murray, S., and Nuttall, J. (2006). Teacher education: What are its prospects? In Doecke, B., Howie, M., and Sawyer, W. (eds.) *Only Connect: English Teaching, Schooling and Community*, pp 321–334. Adelaide: AATE/Wakefield Press.
- O'Neill, M. (2000). Language, literacy and learning: Modeling integration and reflective practice in preservice teacher education. *Teaching and Teacher Education* **16**, 613–633.
- Shulman, L. (2004). *The Wisdom of Practice: Essays on Teaching, Learning, and Learning to Teach*. San Francisco, CA: Jossey-Bass.
- Smagorinsky, P., Cook, L., and Johnson, T. (2003). The twisting path of concept development in learning to teach. *Report Series 16002*. State University of Albany, Albany, NY: The National Research Center on English Learning and Achievement.
- Valijarvi, J., Linnakyla, P., Kupari, P., Reinikainen, P., and Arffman, I. (2002). The Finnish success in PISA – and some reasons behind it: PISA 2000. *Educational Report*. Institute for Educational Research, 70p.
- Wilson, M., Floden, R., and Ferrini-Mundy, J. (2001). Teacher preparation research: Current knowledge, gaps, and recommendations. A research report prepared for the US Department of Education by the Center for the study of teaching and policy in collaboration with Michigan State University.
- Language and Education: Knowledge About Language*, 2nd edn, vol. 6, pp 385–400. New York: Springer.
- Herrlitz, W. and van de Ven, P.-H. (2007). Comparative research on mother tongue education: IMEN's aims, points of departure, history and methodology. In Herrlitz, W., Ougstad, S., and van de Ven, P.-H. (eds.) *Research on Mother Tongue Education in a Comparative International Perspective: Theoretical and Methodological Issues*, pp 13–42. Utrecht: Utrecht Studies in Language and Communication/Rodopi Publishing.
- Kiefer, S. and Sallamaa, K. (2004). *European Identities in Mother Tongue Education*. Conference Proceedings: EUDORA Intensive Program. Slovenia, 147p.
- Kooy, M. and Doecke, B. (2008). Special issue: Teacher learning and development in mother tongue teaching. *L1-Educational Studies in Language and Literature* **8**(4), 126.
- McCann, T., Johannessen, L., and Ricca, B. (2005). *Supporting Beginning English Teachers: Research and Implications for Teacher Induction*. Urbana, IL: National Council of Teachers of English.
- Organization of Economic Cooperation and Development (2005). *Teachers Matter: Developing and Retaining Effective Teachers*. Paris: OECD Publishing.
- Staunton, J. (2008). *Deranging English/Education: Teacher Inquiry, Literary Studies, and Hybrid Visions of "English" for 21st Century Schools*. Urbana, IL: National Council of Teachers of English.
- Van de Ven, P.-H. (2007). A collaborative dialogue: Research in Dutch language education. *English Teaching: Practice and Critique* **6**(3), 112–132.
- Warshauer, F. S., Simons, E., Kainin, J., and Casereno, A. (1999). *Inside City Schools: Investigating Literacy in Multicultural Classrooms*. Urbana, IL: National Council of Teachers of English.

Further Reading

- Adams, A. (2005). *Teaching the Mother Tongue in a Multilingual Europe*. London: Continuum Publishers.
- Alsup, J. (2005). *Teacher Identity Discourses: Negotiating Personal and Professional Spaces*. Urbana, IL: National Council of Teachers of English.
- Garcia, O. (2008). Multilingual language awareness and teacher education. In Cenoz, J. and Hornberger, N. (eds.) *Encyclopedia of*

Relevant Websites

- <http://www.ilo.uva.nl> – International Association for the Improvement of Mother Tongue Education.
- <http://www.ncte.org> – National Council of Teachers of English.

Vocational Teacher Education

S Johansson, University of Umeå, Umeå, Sweden

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In this article, vocational teacher education is explored and distinguished from other forms of teacher education. The notion of vocational teacher includes a person who teaches youth or adults skills leading to competency in a specific vocation, trade, or profession. Examples would be a teacher preparing students for careers as a mechanical engineer, carpenter, baker, auto mechanic, hairdresser, chef, restaurateur, backhoe operator, store clerk, welder, electrician, nursing assistant, plumber, painter, etc.

Issues considered below are all related to vocational teacher education. However, since vocational education and vocational teacher education are complex activities, necessarily intertwined with each other, one cannot be treated without the other. Furthermore, in order to ascertain what distinguishes vocational teacher education, one must, to some degree, compare it with other forms of teacher education. However, since pedagogy, didactics, sociology, and psychology are common subjects relevant in most teacher education, they are deliberately avoided.

Today vocational education (hence vocational teacher education as well) is becoming increasingly important on the international, national, regional, and local political arenas. A well-educated and well-trained work force is universally considered a key factor in the competitive global marketplaces, an attraction to capital investment and to the development and sustainability of a prosperous society. The large number of recent concepts that have appeared in research on vocational education and work-based learning indicates the rapidly increasing value accorded to the subject. Such concepts include reflective practitioner (Schön, 1991), life-long learning, and tacit knowledge (Polanyi, 1966), situated learning (Lave and Wenger, 1991), activity learning (Engeström, 1987, 1994, 1999), and master-learning (Nielsen and Kvale, 2000). All those concepts indicate that learning is something more than theoretical issues relating to formal education at schools or in universities. Thoughts about the nature of learning influence both vocational education and the training of vocational teacher's education since they introduce new assumptions and alternative ways of thinking what vocational education really is and how it should be carried out. At the same time, old views on what type of knowledge is valuable are being challenged, and this in turn has a strong impact on which issues are to be treated in vocational teacher education and how it should be conducted. New concern can also be noted at

international research conferences where special interest groups devoted to vocational education and vocational teacher training are rapidly growing.

Many Different Models

Societies have developed vocational systems where specific competencies and skills needed by its work force are reproduced and transferred to its youth. Depending on historical and cultural circumstances, these systems have taken many different forms. In some places state authorities regulate education, while in others trade-specific agents, employer organizations, or labor unions struggle for control. The cost of this education may be paid by taxes or may be market based with student fees. There are places where training is customarily offered in evening classes, although the full-time model predominates. Training can take the form of an apprentice program or take place in school workshops, or in a combination of both. In some regions, traditionally handicraft is well established; elsewhere industrial mass production is emphasized. Despite the differences, an issue that appears to be universal is the stability of the traditional division between male and female. Since vocational education is deeply intertwined with historical, cultural, and traditional factors, when changes are contemplated, a number of stakeholders attempt to influence them to suit their own ends. Embedded in this Gordian knot of traditions and stakeholders is the vocational teacher training itself. Because of the wide variety of vocational education models, there are equally many different models of vocational teacher education. Some programs simply ask for college education, while others require a university degree. The requisite vocational competence for such teachers may be in the vocational teacher training. Alternately, teacher training may be limited to pedagogical issues. There is a model in which two types of vocational teachers work in pairs: one teaching theoretical issues in the classroom and the other providing hands-on-training in a workshop setting. In many cases models vary within a national system, depending on the vocation at stake. Thus, there seem to be as many forms of vocational teacher education as there are vocational education models. Nevertheless, although this wide variety of models makes vocational teacher education difficult to characterize, it makes us focus on the least common denominator found in diverse models in order to come up with an adequate description.

Differences between Vocational- and Subject-Oriented Teacher Education

To clarify the issues that distinguish vocational teacher education from other forms of teacher education, regardless of the model or where it is applied, some aspects are explored in what follows. Our aim is to point out what is essential to vocational teacher education and does not affect other categories of teacher education to the same extent.

When introducing the notion of vocational education, a dividing line seems to emerge between it and subject-oriented education (e.g., mathematics, language instruction, natural science, social science, etc.). It is noticeable that subject-oriented education has a knowledge area corresponding to a university scientific discipline. This knowledge area originates in the academy and slowly trickles down into course syllabi at the schools. A teacher wanting to specialize in an academic subject thus looks to a university's departments when deciding what their own education should focus on. Contrary to this, when a prospective vocational teacher plans a course of study, there is no academic base to imitate; instead, one must look at areas of activity in working life. These activity areas require specific competencies for carrying out tasks in a given vocation, trade, or profession. The different points of departure, a knowledge area for subject specialist teachers and an activity area for vocational teachers, distinguish between and shape those two teacher categories.

Steering Systems in Vocational- and Vocational Teacher Education

School systems may be looked upon as containing a number of arenas, in which each has its own task (Lindensjö and Lundgren, 2000). In the policy arena, political agents formulate a comprehensive curriculum, generally on a national level, that besides containing laws and regulations, directs teachers work within the school system. With respect to vocational education, representatives from the various branch organizations, unions, and employer committees establish syllabi for each specific vocation. Often vocational teachers are involved in such work. Course syllabi are also constructed for other competencies judged important on regional and local levels. Here vocational teachers play a major role. In the transformation arena, curricula and syllabi are transformed in a shape that approaches actual training. The job of planning for actual training, taking numerous specific issues in a profession into account, is an essential part of what has to be mastered by all vocational teachers and, therefore, integrated into teacher education. Finally, in the realization arena,

the actual training of students takes place. It forms the essential core of teacher training.

The main guide of the vocational teacher education is the curriculum. Here the teacher educator finds the answers to the question of what education should focus on. A curriculum may be more or less detailed. It may be divided up into a number of smaller parts with fixed content and strict time limits, or it may be more broadly descriptive, with fewer details. The former would result in a course of instruction directed by strict control and discipline. The latter curriculum would be conducive to a higher degree of freedom and allow more problem-oriented and individualized teacher training and give student teachers the opportunity of participating in the design of their own training.

The arenas described above are common to formal education, but in both vocational education and vocational teacher education there are additional arenas formed by branch, employer labor, and trade unions whose opinions have to be considered. Whatever mode of procedure is adopted, insight into the arenas mentioned above is an essential part of all vocational teacher education. However, an institution must not only plan and carry out an existing syllabus, but also possess the ability to construct syllabi suitable for training in competencies important for the local work force.

Admission Requirements

One of the most critical aspects in models of vocational teacher education is the admissions process. Besides its function of determining who may enter the vocational teacher profession, it also expresses prevailing assumptions about the best way to carry out vocational training. In some places academic qualifications are considered a necessity and work experience is given less importance, while elsewhere the opposite is true. Two major standpoints can be found, although a combination would be preferred – either considerable practical work experience is demanded, with knowledge, skills, and competencies acquired during a certain length of time in the profession at stake; or theoretical knowledge, as evidenced by academic credits, is asked for, in the form of a master degree. A third standpoint holds that the ability to motivate students and cooperate with stakeholders is the most important qualification. In practice, what is sought in a potential vocational teacher is closely related to the existing vocational education model.

Laws and Regulations

Vocational teacher education emphasizes on awareness of the rules and regulations that apply to each specific trade

or vocation. These may be safety precautions for electrical installations in bathrooms or environmental regulations concerning the disposal of old carpets. The knowledge of such laws and regulations is essential for a vocational teacher to know, so that they can be conscientiously included in the training of students.

Promoting Aspects of Vocational Education

Vocational education, as with other types of education, has a hidden curriculum behind the official one (Jackson, 1968). Many studies point out that vocational training involves skills and knowledge for a particular vocation; equally important is the invisible transfer of the prevailing work culture with its embedded values (Willis, 1977). This factor distinguishes vocational education from general education. Something hidden is transferred in the way training is carried out, whether it takes place in a school workshop or on the job. Students will unconsciously be induced into accepting hierarchical order and division of work at their future workplace (Johansson, 1999). For example, working in a restaurant in Hanoi is very different from doing the same work in a restaurant in New York or in Monrovia. In some places, work is organized according to a strict hierarchy and with a strong division of work. In other places, work is organized in a more relaxed way, affording a higher degree of freedom to workers. How vocational education is carried out will greatly depend on how work is organized by the local work culture and tradition. The issue of the degree of freedom will, in turn, strongly affect how vocational teacher education needs to be conducted. Vocational teacher education must clarify these unseen aspects in training future teachers. In addition, a vocational teacher often serves as a model for students and may therefore have a strong impact on their values and behavior.

Learning Material

As mentioned earlier, the ability to prepare for vocational training is an important part of vocational teacher education. Textbooks are generally only available for the larger trades. The reason is that the demand is so low and the useful life of such books so short that publishers hesitate to enter the field. Another reason is that equipment and machinery used in vocational training are often of different manufacture and specific teaching material has to be written accordingly. Consequently, vocational teachers must be capable of preparing their own teaching materials, and the necessity of doing so should be stressed as a part of their education.

Place of Training

If training is carried out in a school workshop, the vocational teacher must have the ability to select suitable equipment and maintain all types of machinery. Another responsibility is to keep an adequate supply of consumable material on hand because in most mechanical, electrical, and wood shops, as well as in places such as kitchens, vehicle-repair shops, and hair-dressing salons, the amount of consumable material is overwhelming. A well-developed management system is needed to purchase, store, and inventory all items. The selection, maintenance of equipment, and consumables are specific issues that have to be addressed in vocational teacher education.

When training is carried out in the form of an apprenticeship program or as on-the-job training, a problem emerges of structuring and controlling the student's education. In addition to issues of responsibility decisions about what training should include must be made, depending on what issue is at stake at the moment. No matter what is contained in the syllabus, the training must proceed along the lines of what is currently going on at the job site. However, uncertainty develops as to work quality and grading, since the work is done in the absence of the teacher. This is again a unique issue that distinguishes vocational teachers from other kind of teachers.

Multidisciplinary, Activity- and Problem-Oriented Tasks

Vocational education has a strongly multidisciplinary aspect. For example, a bricklayer must know something about physics when deciding how to build a brick wall, some knowledge of chemistry to choose the right kind of cement, practical psychology when dealing with customers and superiors, and accounting when calculating prices. Thus, in planning training, instruction in diverse areas of knowledge must be included.

Another factor the vocational teacher needs to consider is that all tasks within a vocation are both activity and problem oriented. Every task in every job involves one problem or another to be solved. The objective may be clear, but neither the method nor the means can be taken for granted. Although many problems are routine based, others are complex and require technical equipment and special skills. Operations can be reduced to a number of more or less well-defined tasks, each consisting of a problem to be solved using specific tools. For example, when a baker makes bread, questions arise concerning the type of flour, the amount of spices, the tools that are necessary, and the choice of oven temperature, in addition to where flour, spices, and yeast should be bought. Other decisions must be made with regard to how ingredients

should be transported, where they should be stored, and how much to sell the baked goods for. The task of a vocational teacher is to simulate situations like these in order to conduct training as close to actual conditions as possible.

Grading Students and Quality Control

Grading in vocational training classes is another area that distinguishes vocational teacher education from other types of teacher training. To judge the level of a student's competence to perform a work-related task differs considerably from measuring knowledge in other science areas. Work-related tasks include invisible elements of knowledge and skills. Furthermore, such tasks mostly have to be made more or less independently and in complex social contexts. Judgments involved include such matters as responsibility, creativity, accuracy, speed, sustainability, and the ability to reflect upon the quality of one's work. In technical vocations the work is often concrete and thereby more easy to measure. Nevertheless, awarding grades will be even more complicated when the task to be judged is in a service or care-giving profession where competency in cooperation and communication are very important. Thus, the ability to judge behavior is a competence that has to be cultivated in vocational teacher education.

The work context students will enter should be considered when vocational teacher training is planned and carried out. Vocations are embedded in complex social contexts that include others in that vocation, adjacent vocational areas, and customers. For example, a long-distance truck driver has to cooperate with storage company workers, both when loading and unloading, truck inspectors at weigh stations, customers, superiors, and toll collectors when driving on international routes, not to mention the responsibility of driving a heavy-duty, 64-wheel vehicle.

Vocational Student Teacher

In a situation where the entire course of vocational teacher training is conducted at a college or university and the future teachers will have limited practical experience when hired for their first teaching job, they will suffer from a lack of credibility, both with their students and from local business representatives, making it difficult to arrange on-the-job training for their students in local business.

On the other hand, in a model where the students are admitted to teacher education programs due to their work experience, they will suffer from lack of previous formal education in areas such as reading and writing

skills. Conflicts might also develop when middle-aged workers enter university classrooms. Studies indicate that some of the latter do not fully accept the school's focus on academics, preferring to stress their own vocational competence.

Recent Trends in Vocational Teacher Education

Two major trends are visible in contemporary vocational education. One is an increasing academization, that is, traditional vocations are finding their way into universities. The other is an increasing tendency for basic vocational education to be carried out in an apprentice form. These two trends are seemingly incongruous, although they, in fact, complement each other. In almost every work place, there are those who carry out routine tasks with little competence demanded of them, and those who handle the more complicated tasks. This division of work was formerly negotiated at the workplaces, but with the increase of educational differences, the division will be drawn at the university. Today subordinate status depends mostly on the length of one's experience, seniority principle. In the future, as theoretical education increases, a position might not depend on seniority but on educational achievement. Another aspect to take into account is that simple and routine tasks are disappearing with the increasing automation and computerization. As a result, there are fewer and fewer jobs that demand low competence today.

The first trend has resulted in the greater part of vocational teacher education today being conducted at the university level or it is on its way there. This has put considerable pressure on vocational teacher education, especially where work experience was formerly considered a valuable admission requirement. Thus, those engaged at the vocational teacher education have a double role. On the one hand, vocational teacher education should prepare a student for a teaching career as envisaged; on the other hand, the same student should be qualified to enter a research program at the university level. This equation is not easy to balance and causes problems, especially when admission requirements are based on long-term work.

Bibliography

- Engeström, Y. (1987). *Learning by Expanding*. Helsinki: Orienta-Konsultit.
- Engeström, Y. (1994). *Training for Change: New Approach to Instruction and Learning in Working Life*. Geneva: International Labour Office.
- Engeström, Y. (1999). *Perspectives on Activity Theory*. Cambridge: Cambridge University Press.

- Jackson, P. (1968). *Life in Classrooms*. New York: Rinehart and Winston.
- Johansson, S. (1999). *Transfer of Technical Training Know-How. A Study of Consultancy Services in Aid Practice*. Doctorial Dissertation, Umeå: Education Department, Umeå University.
- Lave, J. and Wenger, E. (1991). *Situated Learning. Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Lindensjö, L. and Lundgren, U.-P. (2000). *Educational Reforms and Political Steering*. Stockholm: HLS.
- Nielsen, K. and Kvale, S. (2000). *Master-Learning, Learning as Social Praxis*. Lund: Studentlitteratur.
- Polanyi, M. (1966). *The Tacit Dimension*. New York: Garden City.
- Schön, D. (1991). *The Reflective Practitioner. How Professionals Think in Action*. London: Arena.
- Willis, P. (1977). *Learning to Labour. How Working Class Kids Get Working Class Jobs*. London: Saxon House.

Further Reading

- Engestrom, Y. and Tuomi-Grohn, T. (eds.) (2003). *Between School and Work: New Perspectives on Transfer and Boundary-Crossing*. Oxford: Elsevier Science.
- Furlong, J., Barton, L., Miles, S., Whitting, G., and Whitty, G. (eds.) (2000). *Teacher Education in Transition: Re-Forming Professionalism*. Buckingham: Open University Press.
- Grollman, P. and Rauner, F. (eds.) (2007). *International Perspectives on Teachers and Lecturers in Technical and Vocational Education (Technical and Vocational Education and Training: Issues, Concerns and Prospects)*. Dordrecht: Springer.
- Houston, R. (ed.) (1990). *Handbook of Research on Teacher Education : A Project of the Association of Teacher Educators*. New York: Macmillan.

- Kamarainen, P., Graham, A., and Brown, A. (2002). *Transformation of Learning in Education and Training: Key Qualifications Revisited*. Luxembourg: CEDEFOP.
- Lauglo, J. and Maclean, T. (eds.) (2005). *Vocationalisation of Secondary Education Revisited*. Dordrech: Springer.
- Mottier, I. and de Vries, M. J. (eds.) (2006). *International Handbook of Technology Education. Reviewing the Past Twenty Years*. Rotterdam: Sence Publishers.

Relevant Websites

- <http://www.cedefop.europa.eu> – Cedefop.
- <http://www.trainingvillage.gr> – European Research Overview (ERO).
- <http://www.bibb.de> – Federal Institute for Vocational Education and Training.
- <http://www.ilo.org> – International Labour Organization (ILO).
- <http://www.iveta.org> – International Vocational Education and Training Association.
- <http://www.nccte.com> – National Centers for Career and Technical Education.
- <http://www.ncver.edu.au> – National Centre for Vocational Education Research.
- <http://www.atee.org> – Teacher Education in Europe.
- <http://portal.unesco.org> – United Nations Educational, Scientific and Cultural Organization, Themes, Education.
- <http://www.unip-net.org> – UNIP, a network on innovation and professional development in TVET.
- <http://www.vet-research.net> – VETNET, a European Research Network in vocational education and training.
- <http://www.voced.edu.au> – VOCED, a research database for technical and vocational education and training.
- <http://www.wfto.org> – World Federation of Technology Organizations (WFTO).

TEACHER EDUCATION - THE COOPERATING SCHOOL

Contents

Mentoring in Teacher Education

Partnerships Between Schools and Higher Education

Mentoring in Teacher Education

P D Tomlinson, University of Leeds, Leeds, UK

A J Hobson, University of Nottingham, Nottingham, UK

A Malderez, University of Leeds, Leeds, UK

© 2010 Elsevier Ltd. All rights reserved.

The mentor phenomenon (Little, 1990) has achieved a meteoric rise in teacher education internationally over the past couple of decades. While mentoring is now a major feature in various phases of teachers' professional development worldwide, the present article focuses on the mentoring of beginning teachers during their initial teacher preparation (ITP) courses and subsequent early induction into teaching.

Teacher mentoring is characterized by multiplicity in a range of aspects, from variations in meaning carried explicitly or implicitly by the term itself, through the emphases and strategies characterizing alternative approaches, to different ways of implementing and organizing mentoring in practice – all of these variously informed in turn by a range of ideas and theories relating directly or indirectly to professional learning. As background to the current research picture regarding beginning teacher mentoring, we briefly sketch the nature of its recent advent, the theories and perspectives potentially informing it, and the varying relations of such theoretical resources to practical mentoring.

Mentoring in Context: Advent, Theory, Versions

The Advent of Teacher Mentoring

The modern rise of mentoring in teacher education can be traced in large part to two relatively distinct origins in the 1980s. On the one hand, mentoring arrangements were introduced by educational policymakers at various levels, particularly in the USA, to contribute to teacher professional development, not least by way of mitigating the reality shock encountered by beginning teachers, but also as a means of rewarding and retaining the capable teachers who might take up the mentoring role (Little, 1990).

Mentoring also featured as part of alternative certification programs designed to increase the supply of teachers at points of need (Feiman-Nemser, 1990). At this point, the approach tended to derive from business management, where mentoring involved a one-to-one relationship with the protégé or mentee in which the more senior and experienced mentor was expected to combine a variety of functions, including teacher, sponsor, role model, and confidant.

On the other hand, the 1980s also saw a more intellectually grounded development by teacher education institutions of new forms of teacher preparation centrally involving field-based mentoring. A prime example among several in the UK was the Oxford University Internship Scheme, a radical attempt to tackle what its developers saw as major problems endemic in the traditionally dominant pattern of teacher education. The new scheme placed central emphasis on the role of the mentor, a subject teacher who took primary responsibility for the professional education in classroom teaching of one or two intern students, within a systematic partnership arrangement negotiated between the university education department and the schools involved. Comparable higher education-based initiatives were also seen in North America during this period (Feiman-Nemser, 1990), with policy groups and educationists calling for a more substantial involvement of schools in teacher preparation.

Since then, there have been moves in many countries toward beginning teacher mentoring. Perhaps the most clear-cut of these occurred in the UK, where in 1992–93 the government mandated that student teachers in England and Wales should henceforth spend at least two-thirds of their courses in schools, where practicing teachers should play a major role in supporting their attainment of government-specified teaching competences (now standards).

Table 1 Recent perspectives on teacher professional knowledge and its acquisition

<i>Perspective</i>	<i>Central features</i>
1. Reflective practice	Emphasis, associated particularly with Dewey, on the importance of conscious problem-solving for effective practical action; Schön's extension to include thinking within the course of action as well as outside it, plus his rejection of the traditional idea that professional capability is simply the application of abstract knowledge
2. Craft knowledge	(Teaching) craft as situated intelligent know-how involving intuitive sensitivities, awarenesses, and action capabilities; anchored in personal experience and educable through reflection, but not reducible to explicit rules, principles, or routines. Associated particularly with Grimmer, who draws on aspects of 1, 7, 11, and 12
3. Personal practical knowledge	Associated particularly with Elbaz; strong overlap with 2, but recognition of explicit theory as well as theory embedded in action. Connelly and Clandinin emphasize the embedding of knowledge in teachers' personal narratives and metaphors
4. Practical reasoning	Emphasis mainly by philosophers on role of explicit critical thinking and argument in deciding on and justifying educational actions
5. Explicit teacher knowledge	Focus on what types and content of declarative knowledge are required as a basis for effective teaching. Associated particularly with L.S.Shulman's emphasis on integration of subject knowledge with pedagogical insight
6. Values and personal engagement	Teaching seen as a craft that is inherently moral by way both of affecting persons and requiring commitment and motivation from persons, which are thus integral aspects of teacher professional capability
7. Constructivism	Originally a psychological theory that human awareness, knowledge, and capability are formed not just passively by the impact of external reality, but also by an interplay between this and what the person brings to the encounter with it and how they actively think about issues. Disagreements among proponents as to how far the influence of such individual preconceptions and processes can be modified by outside/social agents, and as to the epistemological consequences of a constructivist psychology
8. Study of expertise	Recently developed psychological investigation of characteristically complex human forms of capability and their development, typically by means of expert–novice comparisons
9. Cognitive psychology of skill	Post World War II psychological study of skill and skill acquisition, originally using information-processing models and focusing on simple manual activities, later extending to more complex capacities, including social skills
10. Situated cognition and learning	Psychological–anthropological stance arguing that knowledge and capability are gained within and relate to particular social and historical situations, as opposed to dealing to abstract generality. Cognitive capacities are seen as typically distributed across and involving coordination among members of social groups, and inherently involving use of tools, whether physical or conceptual
11. Implicit cognition and learning	Philosophical tradition, mainly associated with Ryle, emphasizing that intelligent action capability or knowing how does not necessarily involve explicit declarative knowledge or knowing that. Recent experimental psychological research confirming not only that cognitive processes involved in action may occur unconsciously, but that acquisition of some kinds of capability may occur without conscious deliberation

Theories and Research on Teacher Professional Knowledge and Learning

Concurrent with the above developments, the same period also saw substantial growth in theorizing and research on the nature and development of teaching capability, under the label of teacher (professional) knowledge. Munby *et al.* (2001) and Tsui (2003) provide useful reviews of this now extensive work, which has yielded a range of perspectives, derived from various analytical, theoretical, and empirical bases. **Table 1** briefly characterizes the major representatives, which are covered more substantially by various other entries in this encyclopedia.

Drawing implications from such sources for teacher mentoring and development requires recognition, on the one hand, that these perspectives are not only distinguishable, but that, as Munby and his co-authors emphasize, there are various kinds of issues and tensions among

several of them. To this can be added that even in their own terms, some positions are themselves subject to critiques and alternative versions. Nevertheless, it is equally important, on the other hand, to make clear that, in various ways and to varying degrees, these perspectives overlap and combine to support a number of general themes of relevance to teacher education, including school-based mentoring. **Table 2** indicates these themes and their degree of support from different sets of perspectives.

The most notable theme concerns the centrality of practical experience for the development of professional know-how (**Table 2**, theme A), which is supported in differing respects by the great majority of the perspectives cited. However, any tendency to replace the traditional privileging of theory in teacher education by exclusive experientialism would, for example, be countered by a range of support for the importance of conscious reasoning

Table 2 Professional learning themes and their supporting perspectives

<i>Theme</i>	<i>Source/perspective</i>
A Professional learning requires practical action and experience	Emphasized by: Reflective practice; craft knowledge; personal practical knowledge; study of expertise; cognitive psychology of skill; implicit cognition and learning; situated cognition and learning. Included in: Values and personal engagement; constructivism; practical reasoning; explicit teacher knowledge.
B Importance of explicit knowledge and conscious reasoning/reflection	Emphasized by: Explicit teacher knowledge; reflective practice; practical reasoning; cognitive psychology of skill. Included in: Constructivism; study of expertise; values and personal engagement; personal practical knowledge.
C Intuitive character of developed know-how	Emphasized by: Craft knowledge; reflective practice; personal practical knowledge; study of expertise; cognitive psychology of skill; implicit cognition and learning; Included in: Situated cognition and learning
D Power of learner-teachers' pre-conceptions	Emphasized by: Constructivism; cognitive psychology of skill Included in: Reflective practice; practical reasoning; explicit teacher knowledge; values and personal engagement; study of expertise
E Developmentally staged nature of professional learning	Emphasized by: Study of expertise; cognitive psychology of skill Included in: Constructivism
F Promotion of responsible autonomy	Emphasized by: Values and personal engagement; Included in: Reflective practice; personal practical knowledge; craft knowledge.

and declarative knowledge (theme B). Purposefulness and explicit problem solving are also important aspects of a somewhat neglected precursor of the modern study of expertise, the cognitive psychology of skill (Table 1, perspective 9), which also provides a corrective to the behaviorist associations still characterizing much usage of this term (cf. Tomlinson, 1999). Thus differential emphases among these well-grounded perspectives may signal the complementary, combined importance of several themes, rather than competition between them as mutually exclusive alternatives.

Tensions do nevertheless remain. Widespread indication that well-established components of skilled action and cognition tend to be deployed intuitively and unconsciously (theme C) does, for example, imply problems for potential communication of and reflection on teachers' professional know-how, even if at the same time it helps explain the strength of the teaching and learning preconceptions (theme D) student teachers typically arrive with from their own lengthy experiences as school pupils. On the other hand, changing relationships between explicit and implicit processes in teaching development (Tomlinson, 1999) are illuminated by a further theme (E), the claim that progress toward relatively full professional capability tends to proceed through a series of developmental stages. While this would imply the need for different forms of learning assistance at these different stages, differences of detail among development models (e.g., Dunne, 1994; Furlong and Maynard, 1995) do raise further issues.

Finally, the value-led personal engagement aspect of teaching (theme F) has received increasingly explicit consideration in recent decades, as well as being part

and parcel of the reflective, craft knowledge and personal practical knowledge approaches. Together with traditional humanistic psychological concerns, these sources make it clear that cognitive, emotional, and motivational support for the development of responsible professional autonomy must feature as a further basic strand in promoting teacher professional development.

Versions of Mentoring

From the now considerable range of literature, it is noteworthy that the term mentoring is used with varied meanings, that is, to convey what are actually a range of somewhat differing concepts, that are typically complex and multifaceted. Such meanings of the term show the family resemblances typical of linguistic usage, in that different sets of mentoring conceptions have different common strands as well as differing contrasts. Important kinds of defining features at stake here include mentoring functions or purposes, strategies or ways of going about achieving such purposes, and contexts in which the activity is to be set.

Thus, on the one hand, it seems clear that in teacher education over the last couple of decades, mentoring has moved on from the career-advancement focus of its early business model and come largely to mean assisting teacher professional development/learning (function aspect) on a one-to-one basis (context/strategy aspect) in the professional activity setting (context aspect). On the other hand, nevertheless, such a claim requires qualification at least in that: (1) some usages involve a more specific concept, for example, adding strategy criteria such as that by definition

mentoring must include promotion of reflection or exclude explicit mentor evaluation of the mentee (e.g., Wang and Odell, 2002) and (2) the above conception is itself relatively general, admitting of considerable further differences of detail.

Given these often subtle differences among mentoring conceptions, it is important not to take the meaning of the term for granted, but to establish the facets specifically defining any particular usage, whether explicitly or implicitly, conceptually or operationally. When one does this, it becomes apparent that existing versions of mentoring show considerable variation and idiosyncrasy (Wang and Odell, 2002).

The relationship of practical developments in mentoring to theoretical ideas and research is difficult to track, but appears to have been both complex and evolving. School-based mentoring could itself be seen as an obvious embodiment of the practical experience theme noted above, and although its arrival in 1980s teacher education appears to have been prompted more substantially by policymakers or practitioner reaction to the traditional separation of intellectual learning and practical experience, Schön's ideas about reflective practice do seem to have acted as a particularly influential catalyst, judging by the speed and extent to which reflective vocabulary entered the discourse and espoused theory of teacher educators (Furlong and Maynard, 1995).

However, the take-up of other perspectives listed in **Table 1** has been less clear-cut and apparently more sporadic. In their extensive 2002 review of mentoring research, Wang and Odell did classify predominant assumptions underlying mentoring programs under three theoretical paradigms. However, this was a *post hoc* analysis and the authors made clear their view that none of these paradigms suffices alone, particularly for the mentoring of progressive reform-minded teaching. When we turn to the growing number of practical books on mentoring, we often find little reference to formal theoretical ideas, even if there is frequently linkage with research, often the authors' own. Those that do make explicit reference to theoretical frameworks tend to do so within a paradigmatic pluralism which seems appropriate to the challenging nature of the practical field. Currently, such explicit usage of theoretical approaches in the context of educational research findings seems to be on the increase, at least as far as academic journal publications are concerned.

Research on Mentoring

The last two or three decades have witnessed an explosion in the number of research studies and publications dealing with various aspects of mentoring. However, this work is decidedly disparate and few comprehensive studies are available. Furthermore, the evidence base relating to the

effectiveness of mentoring remains limited, doubtless partly because of the inevitable difficulties of disentangling the effects of mentoring from other kinds of process and influence. Nevertheless, a number of common and useful findings have begun to emerge with respect to the mentoring of beginning teachers.

Effectiveness and Benefits

From its modern outset, teacher mentoring has been seen as holding the promise of professional motivation and development not only for beginning teacher mentees, but also for teachers providing the mentoring, and thence benefiting schools and educational systems more broadly.

Benefits for mentees

Research suggests that one-to-one mentoring is an important, if not the single most effective, method of supporting and facilitating the professional development of trainee and neophyte teachers. A wide range of benefits of mentoring for beginning teachers have been documented, including reduced feelings of isolation, increased confidence and self-esteem, professional growth, increased self-reflection and problem-solving capacities, and the assimilation of their mentors' practices. The benefits of mentoring featuring most commonly among research findings on this issue relate to mentors' provision of emotional and psychological support, important because mentees' emotional condition is argued to have wide consequences for their progress. A number of studies (e.g., Bullough, 2005; Moor *et al.*, 2005) have reported evidence of mentors boosting the confidence and increasing the morale and job satisfaction of beginning teachers.

Research also points to the impact of mentoring on different aspects of the developing capabilities of beginning teachers, most notably their behavior and classroom management skills, and their ability to manage their time and workloads (e.g., Moor *et al.*, 2005). More generally, mentors have also been found to play an important role in the socialization of beginning teachers, in helping them to learn and adapt to the norms, standards, and expectations associated with being a teacher in a given context (Wang and Odell, 2002).

Benefits for mentors

A wealth of evidence coming predominantly from mentors' own accounts indicates that mentoring beginning teachers tends to have a positive impact on the professional lives of mentors themselves. First, with respect to motivation and commitment, research has found that many mentors derive satisfaction from undertaking the mentor role, especially through seeing their mentees progress and noticing evidence of their own impact on mentees' teaching development. Through the responsibility involved, through mentors feeling reassured when

their ideas are validated by university tutors, feeling less isolated as teachers and enjoying increased collaboration and enhanced professional recognition, mentoring is thus found to consolidate mentors' teacher identity and to increase their sense of self-worth. Studies report mentors claiming increased confidence in their own teaching, improved relationships with pupils and colleagues, and a generally revitalized or reenergized engagement with teaching.

In a study typical of many others in its outcomes, Lopez-Real and Kwan (2005) found that 70% of mentors in a school–university ITP partnership program in Hong Kong claimed to have benefited professionally from their mentoring role. Reported gains included new and improved teaching strategies, enhanced knowledge and use of information and communication technology (ICT), improved communication skills, greater self-reflection, and increased knowledge and support capability with respect to the professional development needs of beginner teachers and others.

There are multiple sources of potential impact upon mentors' own learning via their involvement in mentoring beginning teachers. Perhaps the largest body of research evidence in this area relates to mentors' learning through self-reflection or critical reflection on their own practice. Most mentors in Lopez-Real and Kwan's (2005) study thus suggested that the mentoring process forced them to reflect on their own teaching, for example, because they "felt compelled . . . to account for and explain the reasons for their [teaching] methods" (p. 19), while many of the mentors involved in research conducted by Hagger and McIntyre (2006) were also said to have welcomed the experience because "it made them think about their own teaching". Mentors have also reported learning from their beginning teacher mentees, from their participation in mentor training courses, from university tutors in university–school partnership ITP programs and, more generally, from increased opportunities to talk to others about teaching and learning.

Benefits for schools and educational systems

It might be expected that some of the above-reported benefits of mentoring would produce consequent gains for these teachers' pupils and schools. The evidence on this particular outcome is limited, however, again partly because of the complexity of researching it. However, there is growing indication, largely from the United States, that mentoring programs for teachers in their first years in the profession can under certain conditions increase retention and stability, in that teachers who are mentored are both less likely to leave teaching and less likely to move schools within the profession (e.g., Johnson *et al.*, 2005). It is also possible that schools and educational systems may benefit from the enhanced retention of those teacher-mentors who become more confident, committed, and capable as a result of their participation in

mentoring. There is limited direct evidence of this to date, though in the UK Moor *et al.* (2005) noted a number of additional benefits for schools involved in this kind of mentoring program.

We now turn to what research tells us about the conditions under which the benefits of mentoring are most likely to be realized.

The Conditions for Effective Mentoring

Contextual support for mentoring

The success of mentoring programs in general and of particular mentoring relationships is influenced by a variety of contextual factors. The most consistent finding in this area is that mentoring is more likely to be effective where teacher-mentors are provided with additional release or noncontact time in which to prepare for and undertake the mentoring role, while successful mentoring is further facilitated where timetabling allows mentors and mentees to meet together during the working day. There is some evidence that mentoring is more likely to lead to positive outcomes where mentors receive financial reward and/or some other form of incentive or recognition for their work; where it is carried out in contexts which are relatively free from excessive emphases on externally determined goals and agendas; where mentors are committed to and involved in the design and evaluation of the broader programs of which mentoring is a part; and where such programs are coherently integrated and not characterized by fragmentation among the contributing school- and higher education-based teacher educators.

Mentoring is also more likely to be successful where it takes place within schools which are characterized by collegial and learning cultures and which value learning teachers (Edwards, 1998), where both mentors and mentees have access to support outside of the mentoring relationship and where there are mechanisms in place for both mentee and mentor to initiate without blame the establishment of an alternative pairing.

Mentor selection and pairing with mentees

Research confirms that the success of beginning teacher mentoring is at least a partial function of the ways mentors are selected and paired with mentees: not all good or experienced teachers make good mentors, and not all good mentors make good mentors of all beginning teachers. Mentors should be effective practitioners who can model good professional practice and be professionally respected by their mentees, but perhaps most importantly of all, they should want to do the job and be committed to the work of mentoring. They need to be supportive, non-judgmental and trustworthy, have a positive demeanor, possess good listening skills and the ability to empathize, as well as the willingness and ability to take an interest in beginning teachers' work and lives.

Mentoring has been found more likely to be successful where decisions about mentor–mentee pairings take account of mentees' strengths and limitations, and where the mentor and mentee get along both personally and professionally. Mentoring also tends to be more effective, other things being equal, when mentors teach the same subject specialism as their mentees, but becomes less effective when the mentors are the mentees' head teacher or deputy head teacher, the latter being explained in terms of more senior colleagues finding less time for mentoring and beginning teachers tending to be more inhibited by the high status of the mentor within the school staff structure (Hobson *et al.*, 2007; Johnson *et al.*, 2005).

Mentoring strategies

Like all forms of teaching, mentoring is most effective where it addresses and responds to the needs of the learner/mentee. This means that mentors of beginner teachers should respect the individuality of their adult learner mentees, taking account of mentees' particular learning styles and concerns by way of strategies that are appropriate to the stage of development they are at. However, as Wang and Odell's (2002) review in the context of reform-oriented teaching reminds us, mentoring must keep sight of its particular goals. Thus, there are indications that early in the mentoring relationship, mentors should help mentees identify and engage in critical interrogation of their conceptions of teaching, learning to teach, and mentoring, including discussion of the nature and advantages of different forms of reflection, since these can otherwise present barriers to mentees' professional learning and development. Mentors should also seek to agree with mentees the individual goals of the mentee and the objectives of the mentoring relationship, and should revisit and review these objectives and goals periodically and, where appropriate, revise them.

While the extent to which mentors are able to address mentees' individual needs can be pivotal to the success or otherwise of mentoring, research has also found a number of general approaches, strategies, and tactics to be effective and likely to be valued by beginning teacher mentees.

First, effective mentors provide their mentees with emotional and psychological support, make them feel welcome, accepted and included, and are approachable. Second, effective mentors are those who not only dispose of sufficient time and do spend time with their mentees, but also do so on a regular basis. Third, effective mentors allow their mentees an appropriate degree of autonomy to make decisions and to develop their own teaching styles, while not being too *laissez-faire* or hands off. Fourth, a considerable number of studies have found that one of the most valued aspects of the work undertaken by mentors with beginning teachers is lesson observation (both of and by the mentee) which involves analysis of the processes involved. Mentors' observation of the lessons of their

mentees tends to be valued most where the objectives of the observation are agreed upon in a preobservation conference, and where the postobservation conference focuses on specific aspects of mentees' teaching and includes constructive comments from the mentor. It should be conducted in a sensitive, nonthreatening way, providing an opportunity for genuine dialog in which there is joint explanation and exploration of perceptions leading to agreement on the strengths and weaknesses of the mentee's teaching. The latter deserves some emphasis in the light of indications of teacher mentors failing to challenge their mentees' assumptions sufficiently (e.g., Edwards, 1998). The dialog should also include discussion of the likely impacts of observed teaching actions and leave the mentee with clear ideas on how he/she might overcome any problems and work toward improving weaknesses. Such discussions would also hopefully be located within attempts to scaffold mentees into deeper levels of thinking and reflecting about teaching and learning.

Finally, a number of writers have argued from a variety of perspectives that collaborative teaching, including planning and reflection, by teacher mentors and their mentees offers a mentoring strategy with strong potential for realization of many of the themes listed earlier in **Table 2**. Apart from a small number of exceptions (e.g., Burn, 1997), however, research on such a strategy is so far notable by its absence.

Mentor preparation and support

Many have argued that since even excellent school teachers may not be effective facilitators of adult professional learning and because the mentoring role is a relatively new one that can clash with traditional norms and structures of teaching (Little, 1990), it is important to provide mentors with preparation and support. This should be practical and specific, including strategies for functions such as observation and discussion of mentee teaching, promoting mentee reflection, and discussing pedagogical issues with them. But on the basis of research findings (e.g., Bullough, 2005) and the kinds of modern perspectives on professional knowledge and learning mentioned earlier, it is equally conclusive that such preparation must go beyond the behavioral inculcation traditionally associated with training and include the cognitive resources and supported participatory opportunities needed to develop intelligent mentoring capability and identity. Writers such as Bullough (2005) suggest the utility of mentors engaging in discussion of mentoring with other teacher-mentors and university-based teacher educators. However, while Graham (1997) provides evidence that such forms of collaborative inquiry are effective in overcoming mentor isolation, the evidence base regarding the effects of different kinds of mentor preparation is generally still rather sparse.

The Dark Side of Mentoring

A number of commentators (e.g., Sundli, 2007) have castigated what they see as an uncritical optimism widely characterizing literature on mentoring in education. Warnings concerning what Long (1997) termed “the dark side of mentoring” are nevertheless equally long-standing and in recent years, such possibilities have been considered by several published studies. The overall pattern of evidence indicates that rather than being inevitable consequences inherent in mentoring *per se*, such downsides are functions of failure at various levels to ensure the sorts of conditions for effective mentoring considered above. This is nevertheless not to suggest that all conditions for effective mentoring are always achievable or, conversely, that their nemesis, the obstacles or sources of collateral damage to mentoring (Kilburg and Hancock, 2006), are always possible to overcome.

Disadvantages of mentoring for mentors

Some studies attest failures to realize the sorts of potential benefits to mentors referred to above, but others indicate that mentoring can actually be disadvantageous or harmful to them. Three main problems are documented.

First, many studies (e.g., Hobson *et al.*, 2007; Moor *et al.*, 2005) have reported mentors experiencing difficulties relating to increased and competing demands on their time as a result of their involvement in mentoring in addition to normal teaching roles. As well as impacting on mentors’ work–life balance and stress, this can also contribute to difficulties in accommodating the needs of their mentees. There seems to be no evidence of any related, negative impact on the learning of such mentor teachers’ own pupils, though this may be due to the lack of research or publication.

Second, research has found that mentors sometimes experience feelings of insecurity, nervousness, threat, and even inadequacy at the prospect of their lessons being observed by mentees or by their mentees presenting new and challenging ideas (e.g., Bullough, 2005). Third, some studies have indicated that mentors can feel isolated in the role (e.g., Graham, 1997).

There are indications in the research of such problems being caused or exacerbated by mentors being given insufficient or no additional time in which to carry out their mentoring duties, by a lack of preparation for the role, or by inadequate training that deals substantially only with its administrative aspects.

Disadvantages for beginning teachers and educational systems: Limitations of mentoring in practice

Research on beginning teacher mentoring has also uncovered variation in the nature and quality of mentoring support provided and has documented evidence of poor

mentoring practice, which has negative consequences for mentees, their schools, and wider education systems. There appear to be four main kinds of failings.

First, some studies have found mentors failing to provide sufficient support for beginner teachers’ emotional and psychological well-being, characterized in many instances by general unavailability. In some cases the situation can be worse, with student teachers being overloaded or even feeling bullied (Maguire, 2001) by their school-based mentors.

Second, and as noted, considerable research has suggested that some beginning teachers are not sufficiently challenged by their mentors – in some cases, by not being given sufficient responsibility or autonomy and freedom to innovate. Edwards (1998), for example, argued that partly due to the assessment framework of ITP in England and partly to protect their own pupils and their learning, primary phase teacher-mentors in their study tended to guide their student teacher-mentees into low-risk activities.

Third, numerous studies have shown mentors tending to see their role primarily in terms of the provision of safe sites for trial and error learning, focusing in their interactions with mentees on technicalities or performance issues such as classroom management and subject content, devoting little or insufficient attention to pedagogical issues and in particular to reform-minded teaching, to the promotion of reflection incorporating an examination of principles behind the practice, or to issues of social reform and social justice. Indeed, some studies (e.g., Sundli, 2007) have shown that some teacher-mentors hold a transmission perspective on teaching and learning, have a limited understanding of concepts such as critical reflection, continue to hold dualist notions of theory and practice, and lack the confidence to incorporate theoretical insights into their work with mentees.

One outcome of such failings is that, in spite of the explicit aim of some mentoring programs being to reduce teacher attrition, they have actually contributed to trainee and beginning teachers withdrawing from their courses or dropping out of teaching (e.g., Hobson *et al.*, 2007). A further potential consequence is that this restricted quality of mentoring serves to restrict mentees’ learning and development in a variety of ways. One finds little evidence, for example, of school-based mentoring achieving its often-stated aim of reducing theory–practice dualism among beginning teachers and helping them to recognize and utilize the theoretical work covered in their ITP programs (Bullough, 2005). Not surprisingly, a number of studies have suggested that such restricted kinds of mentoring result in the reproduction of conventional norms and practices, rendering beginning teachers less likely to develop reform-minded and learning-centered approaches to teaching or to advance social reform and social justice agendas (Wang and Odell, 2002).

Conclusion and Prospect

Beginning teacher mentoring continues to establish itself evermore firmly worldwide and can draw on an increasing range of theory informing professional learning and useful findings from a burgeoning research effort in teacher mentoring. Nevertheless, while the multilayered complexity of teaching teachers ensures no fewer challenges to mentoring research than those facing pedagogy and pedagogical research generally, we must note, for example, that very few indeed of the studies informing this article went beyond accounts of mentoring by mentors and mentees themselves, or involved any meaningful comparison between the professional learning outcomes of persons receiving and not receiving mentoring support. It is to be hoped that while firmly retaining their methodological and philosophical criticality, future researchers in this field may also be somewhat more adventurous in their investigation of the effects, including both potential benefits and costs, of particular versions and strategies of beginning teacher mentoring.

See also: Cognitive Psychology and Educational Statistics; Contemporary Approaches to Teacher Professional Development; Experienced Teachers' Craft Knowledge; Memory; Partnerships Between Schools and Higher Education; School Development for Teacher Learning and Change; Situated View of Learning; Teachers Career Stages and Professional Development.

Bibliography

- Bullough, R. V., Jr. (2005). Being and becoming a mentor: School-based teacher educators and teacher educator identity. *Teaching and Teacher Education* **21**, 143–155.
- Burn, K. (1997). Learning to teach: The value of collaborative teaching. In McIntyre, D. (ed.) *Teacher Education Research in a New Context: The Oxford Internship Scheme*, pp 145–161. London: Paul Chapman Publishing.
- Dunne, R. V. (1994). The acquisition of professional activity. In Harvard, G. and Hodgkinson, P. (eds.) *Action and Reflection in Teacher Education*, pp 105–124. Norwood, NJ: Ablex.
- Edwards, A. (1998). Mentoring student teachers in primary schools: Assisting student teachers to become learners. *European Journal of Teacher Education* **21**(1), 47–62.
- Feiman-Nemser, S. (1990). Teacher preparation: Structural and conceptual alternatives. In Houston, W. R. (ed.) *Handbook of Research on Teacher Education*, pp 212–233. New York: Macmillan.
- Furlong, J. and Maynard, T. (1995). *Mentoring Student Teachers: The Growth of Professional Knowledge*. London: Routledge.
- Graham, P. (1997). Tensions in the mentor teacher–student teacher relationship: Creating productive sites for learning within a high school English teacher education program. *Teaching and Teacher Education* **13**(5), 513–527.
- Hagger, H. and McIntyre, D. (2006). *Learning Teaching from Teachers: Realising the Potential of School-based Teacher Education*. Maidenhead, UK: Open University Press.
- Hobson, A. J., Malderez, A., Tracey, L., et al. (2007). *Newly Qualified Teachers' Experiences of their First Year of Teaching: Findings from Phase III of the Becoming a Teacher Project*. Nottingham: Department for Children, Schools and Families (DCSF).

- Johnson, S., Berg, J., and Donaldson, M. (2005). *Who Stays in Teaching and Why: A Review of the Literature on Teacher Retention*. Cambridge, MA: The Project on the Next Generation of Teachers, Harvard Graduate School of Education.
- Kilburg, G. and Hancock, T. E. (2006). Addressing sources of collateral damage in four mentoring programs. *Teachers College Record* **108**(7), 1321–1338.
- Little, J. W. (1990). The mentor phenomenon and the social organization of teaching. In Cazden, C. B. (ed.) *Review of Research in Education*, vol. 16, pp 297–351. Washington, DC: American Educational Research Association.
- Long, J. (1997). The dark side of mentoring. *Australian Educational Researcher* **24**(2), 115–133.
- Lopez-Real, F. and Kwan, T. (2005). Mentors' perceptions of their own professional development during mentoring. *Journal of Education for Teaching* **31**(1), 15–24.
- Maguire, M. (2001). Bullying and the post-graduate secondary school trainee: An English case study. *Journal of Education for Teaching* **27**(1), 95–109.
- Moor, H., Halsey, K., Jones, M., et al. (2005). *Professional Development for Teachers Early in their Careers: An Evaluation of the Early Professional Development Pilot Scheme*. Nottingham: Department for Education and Skills.
- Munby, H., Russell, T., and Martin, A. K. (2001). Teachers' knowledge and how it develops. In Richardson, V. (ed.) *Handbook of Research on Teaching*, 4th edn., pp 877–904. Washington, DC: American Educational Research Association.
- Sundli, L. (2007). Mentoring – a new mantra for education? *Teaching and Teacher Education* **23**, 201–214.
- Tomlinson, P. D. (1999). Conscious reflection and implicit learning in teacher preparation. Part I: Recent light on an old issue. *Oxford Review of Education* **25**(3), 405–424.
- Tsui, A. B. M. (2003). *Understanding Expertise in Teaching: Case Studies of ESL Teachers*. Cambridge: Cambridge University Press.
- Wang, J. and Odell, S. J. (2002). Mentored learning to teach according to standards-based reform: A critical review. *Review of Educational Research* **72**(3), 481–546.

Further Reading

- Beck, C. and Kosnick, C. (2000). Associate teachers in pre-service education: Clarifying and enhancing their role. *Journal of Education for Teaching* **26**(3), 207–224.
- Feiman-Nemser, S. (2001). Helping novices learn to teach: Lessons from an exemplary support teacher. *Journal of Teacher Education* **52**(1), 17–30.
- Franke, A. and Dahlgren, L. O. (1996). Conceptions of mentoring: An empirical study of conceptions of mentoring during the school-based teacher education. *Teaching and Teacher Education* **12**(6), 627–641.
- Hobson, A. J. (2002). Student teachers' perceptions of school-based mentoring in initial teacher training (ITT). *Mentoring and Tutoring* **10**(1), 5–20.
- Lee, J. C. and Feng, S. (2007). Mentoring support and the professional development of beginning teachers: A Chinese perspective. *Mentoring and Tutoring* **15**(3), 243–263.
- Malderez, A., Hobson, A. J., Tracey, L., and Kerr, K. (2007). Becoming a student teacher: Core features of the experience. *European Journal of Teacher Education* **30**(3), 225–248.
- Oberski, I., Ford, K., Higgins, S., and Fisher, P. (1999). The importance of relationships in teacher education. *Journal of Education for Teaching* **25**(2), 135–150.
- Roth, W.-M. and Tobin, K. (2002). *At the Elbow of Another: Learning to Teach by Co-teaching*. New York: Peter Lang Publishing.
- Tomlinson, P. D. (1995). *Understanding Mentoring: Reflective Strategies for School-based Teacher Preparation*. Buckingham, UK: Open University Press.
- Whisnant, E., Elliott, K., and Pyncheon, S. (2005). *A Review of Literature on Beginning Teacher Induction*. Tacoma, WA: Centre for Strengthening the Teaching Profession.

Partnerships Between Schools and Higher Education

C Day and L Smethem, University of Nottingham, Nottingham, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This article examines the nature of partnerships across different education systems, presenting a selection of those which, available research suggests, hold most relevance to the advancement of teaching and learning in all countries in this century. In the international arena, all governments are focusing their attention on high-quality teaching in the drive to raise standards and improve education for all pupils, regardless of their background or circumstances. In these contexts, the importance of partnerships between agencies, which serve the needs of children and young people, has been realized by governments. This is especially the case in education. Pre-service teacher education, for example, has long been criticized in many countries as being too theory driven and dominated by higher education. Thus, many governments in Europe, North America, and Australia have moved away from supporting teacher education programs, which separate the learning about being a teacher function (the province of universities) from the learning to teach function (the province of schools). At the same time, there has been a widespread acknowledgment that the continuing education and training of teachers is essential in a rapidly changing world as governments intervene more frequently in all aspects of the life of schools with new policies designed to improve standards.

This article identifies and discusses key purposes, forms, processes, and outcomes of contemporary pre- and in-service partnerships between schools and higher education. Such partnerships have been acknowledged internationally as being an important strand in strategies to reduce inequalities, raise the quality of education for all teachers, and enhance the quality and relevance of school education in the twenty-first century. Although the concept of partnerships for educational improvement is not new, there is a growing body of literature across initial teacher education, teacher professional development, and educational research, which acknowledges school–higher education institution (HEI) partnerships as being powerful ways for educators to form collaborations that can result in improved practice and results for students. The literature also points to the need for time, trust, and mutual respect as conditions for success in partnerships (e.g., Sachs, 2003).

Partnerships have been defined as consisting of “two or more parties that share common goals that cannot be reached by either party independently” (Barnett *et al.*, 1999).

Several writers highlight mutuality within partnerships, including the importance of mutual respect, common goals, and interests. However, simply sharing the same set of goals does not guarantee that these will be achieved successfully:

for partnerships are not always benevolent and pressure groups outside the school frequently have more than the children’s interests at heart. (Hargreaves and Fullan, 1998: 5)

Success depends upon whether and how the leadership is distributed in the system, how the partnership is conducted, and on what basis. It is important, therefore, to separate out organizational arrangements for partnerships from the individual and organizational learning that may take place within them. School–HEI partnerships may take a number of different forms which include:

- postgraduate programs, generally for midcareer teachers and nonaccredited in-service programs; both may involve classroom-based research;
- several schools working together with an HEI in one district or country (perhaps working with schools in another district or country);
- a school working with consultants in a local authority/school district and/or HEI;
- research consortia (funded, for example, by education ministries or unfunded);
- action research projects (funded or unfunded); and
- collaborative research networks (formal or informal).

These forms of partnership share three assumptions: (1) critical reflection is a necessary part of being a professional in schools and universities; (2) learning only from one’s own experience will ultimately limit learning; and (3) change is more likely to occur if those who lead change in schools and classrooms are active participants in shaping change agendas. At face value, partnerships champion knowledge which is produced by the user community rather than that which is usually produced by knowledge experts who are far removed from the context in which knowledge will be applied. In effect, partnerships create a variety of opportunities for teacher professional development, including learning from colleagues, which some regard as highly influential. However, partnerships are neither always emancipatory nor do they necessarily promote models of professionalism that may appeal to teachers themselves and to those working in HEIs, especially when they become vehicles for the implementation of

centralized or localized policies, which are not seen by teachers to be beneficial to the students with whom they work. We now examine six innovative school–university partnerships from different parts of the world.

School–Higher Education Partnerships Involving Preservice and Experienced Teachers

Preservice Teacher Education

Teacher educators based in the HEI often lack credibility with pre- and in-service teachers as they are perceived as being out of touch with schools and educational practice. Preservice teachers therefore tend to value more highly their interactions with practicing teachers. The OECD's survey (OECD, Organisation for Economic Cooperation and Development) of initial teacher education notes international concern about the "limited cooperation between teacher education institutions and schools" (OECD, 2005: 134) and resultant lack of "coherence and alignment" (OECD, 2005: 108); it highlights a need for greater sharing of expertise that would enable trainee teachers to make connections between theory and practice. The aim is to guard both against the apprenticeship or tips-for-teachers model of learning to teach, which entirely school-based training might produce, and the irrelevance of theory to practice, which university-dominated training models may be perceived to produce. In England and Wales, since 1992, the government has required schools to be full partners in initial teacher education for secondary school teachers, with trainee teachers spending 24 weeks out of a 36-week course in schools working with teachers and pupils (Department for Education, 1992). Yet in this partnership model, although there is shared responsibility for the assessment of practical teaching, in most instances the leadership largely remains in the control of the HEI, with the partnership characterized as an unequal expert–client relationship (Brisard *et al.*, 2005). Thus, in themselves, partnership models do not always represent equality of roles or decision-making processes, or production of shared knowledge.

Professional Development Schools

In many parts of North America, professional development schools (PDSs) have been established which combine preservice, in-service, and collaborative research and development, drawing on the influential work by Holmes Group (1986) in reforming the national teacher education system in the United States. PDSs embody the principles of collaborative partnerships between schools, school districts (local authorities), and university faculties of education, but go beyond those established in preservice teacher

education in engaging university and school staff in research and development directed at improving practice and enhancing student learning. In PDSs, then, collaboration is the norm for teachers, most of whom otherwise spend most of their working lives isolated in their classrooms. Learning Consortia in Canada perform a similar function. Beginning teachers who have been trained in a PDS are, according to Darling-Hammond (2006: 163), generally more positively viewed by employers than new teachers who have been trained elsewhere. To a large extent, he/she attributes this to the powerful connections between theory and practice in both the school and the HEI, which result in a capacity for constant improvement (2006: 185). Working with beginning teachers has also been shown to benefit experienced teachers who, by reflecting on and articulating their own practice, enhance their own teaching. Indeed, it is claimed that PDSs lead to change in practice for preservice teachers, experienced teachers, and university teachers.

School-Level Development: Improving the Quality of Education for All

Hopkins *et al.* (1998) created Improving the Quality of Education for All (IQEA) in the 1990s, primarily in the United Kingdom. Key concepts in the IQEA approach to authentic school improvement are reflection and collaborative action inquiry led by School Improvement Groups whose work is closely allied to whole-school improvement planning and is supported by senior teachers. Partnerships between school, university, and local authority (district) are premised upon a commitment to working with rather than on schools (Hopkins *et al.*, 1998: 5). The model aims to develop an inquiry-based professional learning community through staff development, with whole-school in-service training events focused on teaching and learning, interdepartmental discussion of and workshops on teaching and learning, at times drawing on expertise within the school, at times working with HEI colleagues. The underpinning framework is one of collaborative inquiry as a means of generating shared knowledge and sustaining improvement.

The National Schools Network and Innovative Links Project Between Schools and Universities

The National Schools Network and Innovative Links Project are school-based research partnerships in Australia with HEIs that focus on improving the work practices and conditions of teachers and students and enhancing student learning (Sachs, 2003: 23). Both projects engage teachers and academics in collaborative research projects for which they jointly identify a focus. The projects are based on the formation of learning communities. Like PDS and IQEA, they actively acknowledge that teacher learning is central

to reform and the impact that teachers' engagement in their own learning may have upon student learning outcomes. The Innovative Links Project was a nationwide initiative which represented formal and explicit partnerships between schools and universities, which engaged teachers in learning which challenged their conceptions of schooling, teacher education, and teacher professionalism. Fundamental to the project was teacher inquiry, which was directly focused on enhancing student learning. The success of this project can be judged by the fact that some partnerships continued their work beyond the end of external funding.

Sachs (2000: 85) maintains that both projects are underpinned by core principles of learning which comprise the basis for "rethinking the practice of teacher professionalism – learning, participation, collaboration, cooperation and activism." She asserts that these principles are fundamental values for the socially responsible and active professional (Sachs, 2000: 83). These projects also enable academics to redress the perception that they are out of touch with the real world of teaching. There is, then, an important difference between these and the PDS movement emanating from America and the IQEA project emanating from England, in that the Australian projects more overtly address and interrogate the nature of teacher professionalism, and, within this, the emancipation of teachers as being critical to educational reform.

Networked Learning Communities

Networked learning communities (NLCs) are groups of schools, usually though not always closely connected geographically, which come together for a lengthy period of time (usually more than 2 years) in order to design and conduct classroom and school-centered inquiries together, which they believe will be of collective benefit. They almost always involve, also, partnership(s) with one or more external agencies. As with other partnerships in this article, they are underpinned by principles of user-produced knowledge, learning, collaboration, trust, and practical inquiry. NLCs exist in many countries (e.g., through European Community-funded projects).

The nature of relationships in NLCs is one of the key challenges to the growth in network learning, which involves partnerships between schools and higher education. It seems to have been assumed in these initiatives that the peer pressure exerted implicitly or explicitly by those involved in active participation will somehow, of itself, move the school, teachers, and universities further in the desired direction of change. In all cases of partnership, however, it seems that the quality of leadership is key to successful change. In the case of networks, it is network leaders whose role is to mediate between the ownership of activities at the school level and the broader interests and needs of the network.

The creation of NLCs takes time and is not always easy (there will always be individuals and groups whose individual or collected vested self-interest may not be served by this). In the process, new knowledge, skills, and dispositions will need to be developed and tentative steps to change supported.

Collaborative Action Research

For many years, teachers from higher education in different parts of the world have promoted and supported systematic practical school and classroom inquiry among teachers. Indeed, this form of inquiry, often referred to as participatory action research and teacher research, is a fundamental part of the work of PDSs, IQEA, and NLCs. However, its origins lie in principles of teacher empowerment, emancipation, and change, the belief that it is teachers themselves who should take decisions about classroom improvements and that these should be based upon the collection and analysis of data. Many universities now provide masters-level qualifications in action research, and supporting non-accredited work of this kind and specialist international journals contain accounts of teachers' work (e.g., *The International Educational Action Research Journal*). Unlike the examples used previously in this article, action research may be carried out by individual or groups of teachers at school level with minimal financial support. However, it is usually supported by a critical friend from higher education. This was the case in the Best Practice Research Scholarship Scheme, a UK Department for Education and Skills initiative to support teachers' continuing professional development. Approximately, 1000 small grants were awarded each year (2000–2003) to serving classroom teachers to support school- and classroom-focused research projects, supported by partnership colleagues in HEIs. In their evaluation, Furlong and Salisbury (2005) reported considerable strengths in this initiative, which was popular with schools and HEIs, with significant impact on teachers' learning, on practice, and on pupils' learning. Collaborative action–research partnerships may focus upon investigating aspects of classroom practice by the systematic collection and analysis of data from one's own classroom and the emphasis is always on improvement through reflection and self-confrontation within an ethical framework of justice, critique, and professionalism (Campbell and Groundwater-Smith, 2007). Alongside these, there are also examples of larger-scale collaborations and even systems that have adopted action research as a means to improvement. For example, Walker (2002) reports on the use of this in teacher education reforms in Namibia and Brazil where policymakers do adopt this approach to partnership learning, however, Cochran-Smith and Lytle (1999) warn that their motives can be prioritized over those of the teachers.

Discussion

Sachs (2003) highlights the potential of school–HEI partnerships for improving teacher professionalism and professional development. She also warns that:

In order to develop collaborative partnerships we must not underestimate the high level of trust and mutual respect that this entails. Without trust and respect, partnerships are on very shaky ground. (Sachs, 2003: 75)

The examples of partnerships which are provided in this article are not intended to be comprehensive or representative of the huge range of partnership configurations that exist, often informally, in and between different teachers, schools, schools and districts, universities, and other agencies. They have been selected, however, as examples of systemic and institutional trends which, in themselves, illustrate an acknowledgment that if schools are to serve the increasingly complex needs of economic success and the well-being of societies in the twenty-first century, then learning alone is no longer sufficient. In the knowledge society, interdependence and connectedness are the new watchwords.

In theory, then, partnerships, where each partner contributes equally to knowledge production and improvement in practice, can create important learning opportunities for the

creation of knowledge in, about, and for practice, which has significant impact on learners, individual teachers, professional learning communities, schools, and HEIs. While traditional school–HEI partnerships rest on the maintenance of unequal expert–client relationships, reciprocal, collaborative, and equal relationships assume that both parties engage in professional development and contribute to their partner's professional learning. It is important to differentiate between partnerships that are formed as a result of teachers', schools', and universities' needs and vision and are driven by these, and those which are created by governments primarily as a means of implementing and embedding external reform agendas. Examples of both have been provided in this article. The prerequisites for successful, sustainable partnerships are time, knowledge and expertise, mutually respectful relationships, shared goals, collaboration rooted in joint ownership and accountability, strategies to overcome potential challenges, and supportive leadership. **Table 1** presents the key features of three kinds of partnership: those which are externally designed, equal and internally designed, and the possibilities which each holds for change (Day, 2007).

The success of each type of partnership will depend upon factors related to geography (how dispersed the partners are relates to their ability to sustain necessary interaction); mode of learning (how involved teachers are in choice and decision making about inquiry will determine

Table 1 Features and consequences of different partnership models for change

	<i>Externally-designed partnerships</i>	<i>Equal partnerships</i>	<i>Internally designed partnerships</i>
Content	Externally defined	Externally/internally defined	Internally defined
Planning	System and central initiative led	External/internal agencies collaborate	Internal agencies lead
Process	External agency employs, commissions or provides tutors/mentor/coaches/consultants	External and internal agencies identify need and jointly determine and resource the support required	Internal agencies provide or buy in support
Evaluation	External quality assurance and control against externally defined standards	Agreed mix of external quality assurance and internal quality control	Internal quality assurance and control only
Contractual relationship	External agency holds participants accountable for meeting its agenda/targets, i.e., they are regarded as 'sub-contractors' to be audited	All agencies negotiate and agree outcomes (formative and summative), roles and responsibilities, and reciprocal accountabilities.	Self-defined outcomes
Possible consequences	Dependency model Little or no examination of ethical or pedagogical issues or contextual variables	Collegial/interdependency model. Collaborative development/joint ownership/collegial relationships. Sense of trust/ownership/respect. Examination of ethical/pedagogical purposes.	Independency model Sense of exclusive ownership. No necessary examination of pedagogical purposes
	Change likely to be short-term	Change may be likely	Change may or may not be sustained. The dependent variable is the quality of school leadership and culture

Adapted from Day, C. (2007). Collaborative approaches to (continuing to develop effective teachers). In Whitehead, J. (ed.) *Collaborative Approaches to Preparing and Developing Effective Teachers*, pp 22–28. London: Universities Council for the Education of Teachers, Occasional Paper No.16.

their sense of ownership and thus the successful transfer and/or implementation of new attitudes, knowledge, or skills); the nature of communicative decision making (whether they are shared or independent); the leadership and, professional development support (nature, quality, frequency, and relevance); the ability of the system to adapt to new environments (structures, cultures, and knowledge systems); and the dispositions of those who work within HEIs and schools to work together.

There are five tensions within each of the forms of partnership provided in this article and included in the table above, which must be managed, if not resolved, if success is to be achieved and sustained:

1. leadership;
2. accountability;
3. cooperation to collaboration between schools and higher education;
4. sustainability; and
5. change and improvement.

Leadership

Robust partnerships are those that are based on reflection, inquiry, and open discussion (internal scrutiny and critique) of values, principles, and practice and the personal, social, and policy contexts in which these exist. They demand the continuing, sustained engagement and support of all parties. Paradoxically, contribution to knowledge production in internally designed partnerships which result in a sense of ownership and responsibility by all relies on a strong element of joint leadership and control, relational trust and respect, and is especially dependent upon the quality of school leadership. Thus, attempts to replicate traditional hierarchies and power relations built on expertise and status rather than stature and influence are likely to produce adverse effects. If partnerships are to be successful, leadership trust is essential. It is important because members of the partnership will need to be willing to take risks, rely upon each other to gain in individual and collective self-efficacy (a sense of increased competence), exercise honesty and openness, and be emotionally confident in their relationships. The ingredients of partnership learning systems suggest sustained attention by leaders in the building of collective intra and interorganizational trust. Trust is the glue that holds partnerships together.

Accountability

Externally designed partnerships which are created in order to implement externally imposed reform agendas hold participants accountable and audit their progress. The sense of ownership is lacking in this client-expert

relationship, and change of participants' practice, attitudes, values, and beliefs is likely to be short-lived. Equal, collegial partnerships, which are jointly developed and owned by all partners, are more likely to sustain change. Successful partnerships are underpinned by a contractual relationship of ongoing negotiation and agreement of outcomes that address a common agenda, which takes into account the needs of teachers, schools, and universities. In this way, all partners have clearly defined complementary roles and responsibilities, joint ownership of the project and partnership, and an equal investment in ensuring that the aims and objectives are met. They hold themselves jointly responsible and jointly accountable.

Cooperation and Collaboration Between Schools and Higher Education

There is an inherent tension in striving to establish equal partnerships between teachers and university academics. The traditional theory-practice divide between academics and teachers, for example, still exists in many countries and is a major constraint in achieving success. Teachers in schools and universities also inhabit different professional cultures of power and autonomy. There is, for example, no culture of collaboration between staff in universities. The traditions are that individual achievement is rewarded (through research and publications). Collaboration is, then, not easy. The intensification of teachers' work also means they might have little time for working intensively with a mentor or colleague. Sachs (2003) distinguishes between collaboration and cooperation in school-HEI partnerships. Collaboration is based on trust and joint decision making. Cooperation does not have equal partners, and traditionally experts hold the knowledge. Where teachers actively participate as co-researchers, jointly choose the research, its design, and means of development, and are supported in concomitant change processes by tutors from higher education, the inquiry undertaken is more likely, it seems, to support long-term change. Sutherland *et al.* (2005) highlight the need for teacher educators to continue to maintain links in order to enhance their credibility with practicing teachers and preservice teachers, a commitment which will require support from teacher education departments.

Sustainability

Fullan (2001: 7) argues that deep and lasting change in education, of the type that is required for students to cope with the complex demands of a global and knowledge society, is not achieved through implementation of the latest policy, but rather through fundamental change in "the culture of the classrooms, the schools, the districts."

Success of schools–HEI partnerships depends not only on partners' initial commitment to the partnership and common goals and interests but also upon their willingness and capacity to sustain that commitment and the extent to which they receive the moral and practical support of leadership. Lack of time is generally acknowledged as being the greatest challenge for all partners, especially in reform contexts where teachers face increased workloads, diversification of roles, and changing workplace conditions. The termination of external funding to support collaboration can result in projects being discontinued. Turnover of committed personnel may also relegate initiatives to the charge of individuals without the necessary institution-wide support and thus adversely impact on outcomes.

Alongside leadership, ownership of goals, ongoing negotiation of inquiry focus and of roles and responsibilities, and sense of achievement are key elements in the amelioration of such challenges and the sustainability of partnerships.

Change and Improvement

The partnerships identified in this article may be said to provide illustrations designed primarily to combine internal and external pressures for change. However, while it is clear that they embody change at a system level in relation to the promotion of new forms of learning relationships and new ways of producing learning for change and improvement, little research has been conducted into the relationships between systemic (organizationally controlled) and individual (personally empowered) change. Yet the nature of this relationship is one of the key challenges to the growth of partnerships between schools, local authorities (school districts), and universities.

Partnerships represent an important, innovative strand of the change process, focusing as they do upon the promotion of interdependent as opposed to independent learning, and the freeing up of traditionally boundaried roles, responsibilities, and expertise. The danger that they may be used as a form of prescriptive policy implementation, on the evidence here, seems remote. Rather, they seem to create the space for learning which is perceived to be defined by and significant to the participants, locating responsibilities for decisions about change and improvement with those who are closest to teaching and learning in classrooms but ensuring, also, that they are able to consult with and learn from those outside schools who have similar felt responsibilities and moral purposes for the betterment of children and young people.

Conclusion

This article has drawn on HEI–school partnerships established in different policy contexts and educational cultures

that have been judged by research to be successful in order to identify and explore their key contexts, purposes, processes, characteristics, benefits. Even within accounts of these, however, challenges to relationships and sustainability have been identified. These reside in issues of power, culture, time, ownership, and leadership. Partnerships are not easy ways to engage in improving teachers and teaching; they are still a minority endeavor in all countries. However, it would appear that they do represent desirable productive routes for the further development of teacher professionalism, school improvement and enhanced student learning and achievement which go some way toward meeting the twin challenges of education in the twenty-first century of building social cohesion (rather than allowing by default isolation and fragmentation), and enhancing capacities for learning and change (in an increasingly economically unstable society). We conclude that the evidence, so far, is consistent in pointing to their important contribution to raising the quality of teaching and learning in the ever-shifting landscapes of policy and practice.

Bibliography

- Barnett, B. G., Hall, G. E., Berg, J. H., and Camarena, M. M. (1999). A typology of partnerships for promoting innovation. *Journal of School Leadership* 9(6), 484–510.
- Brisard, E., Menter, I., and Smith, I. (2004). *Models of Partnership in Programmes of Initial Teacher Education. A Systematic Review of the Literature*. Edinburgh: General Teaching Council for Scotland.
- Campbell, A. and Groundwater-Smith, S. (eds.) (2007). *An Ethical Approach to Practitioner Research: Dealing with Issues and Dilemmas in Action Research*. London: Routledge.
- Cochran-Smith, M. and Lytle, S. (1999). Relationship of knowledge and practice: Teacher learning in communities. In Iran-Nejad, A. and Pearson, C. (eds.) *Review of Research in Education*, vol. 24, pp 249–306. Washington: DC: American Educational Research Association.
- Darling-Hammond, L. (2006). *Powerful Teacher Education*. San Francisco, CA: Jossey-Bass.
- Day, C. (2007). Collaborative approaches to (continuing to develop effective teachers). In Whitehead, J. (ed.) *Collaborative Approaches to Preparing and Developing Effective Teachers*, pp 22–28. London: Universities Council for the Education of Teachers, Occasional Paper No.16.
- Department for Education (1992). *Initial Teacher Training: (Secondary Phase) (Circular 9/92)*. London: HMSO.
- Fullan, M. G. (2001). *The New Meaning of Educational Change*, 3rd edn. New York: Teachers College Press.
- Furlong, J. and Salisbury, J. (2005). Best practice research scholarships: An evaluation. *Research Papers in Education* 20(1), 45–83.
- Hargreaves, A. and Fullan, M. (1998). *What's Worth Fighting for Out There?* New York: Teachers College Press.
- Holmes Group (1986). *Tomorrow's Teachers: A Report of Holmes Group*. East Lansing, MI: Holmes Group.
- Hopkins, D., West, M., and Ainscow, M. (1998). *Improving the Quality of Education for All*. London: David Fulton Publishers.
- Organisation for Economic Cooperation and Development (OECD) (2005). *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*. Paris: OECD.
- Sachs, J. (2000). Rethinking teacher professionalism. In Day, C., Fernandez, A., Hauge, T. E., and Møller, J. (eds.) *The Life and Work of Teachers*, pp 76–89. London: Falmer Press.

- Sachs, J. (2003). *The Activist Teaching Profession*. Maidenhead: Open University Press.
- Sutherland, L. M., Scanlon, L. A., and Sperring, A. (2005). New directions in preparing professionals: Examining issues in students in communities of practice through a school–university partnership. *Teaching and Teacher Education*, **21**(1), 79–92.
- Walker, M. (2002). Subaltern professionals: Acting in pursuit of social justice. In Day, C., Elliott, J., Somekh, B., and Winter, R. (eds.) *Theory and Practice in Action Research: Some International Perspectives*. Oxford: Symposium Books.

Further Reading

- International Journal of Educational Management*, **20**(3).
- Lee, C. K. J., Lo, L. N. K., and Walker, A. (eds.) (2004). *Partnership and Change Toward School Development*. Hong Kong: The Chinese University Press.

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

This page intentionally left blank

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

THIRD EDITION

EDITORS-IN-CHIEF

PENELOPE PETERSON
School of Education and Social Policy
Northwestern University
Evanston, IL
USA

EVA BAKER
UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

BARRY MCGAW
McGaw Group Pty Ltd
Mont Albert Nth, VIC
Australia



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
The Boulevard Langford Lane, Kidlington, Oxford OX5 1GB, UK

Copyright © 2010 Elsevier Ltd. All rights reserved

LATVIA
Copyright © 2010 A Joma

STUDENT DATA SYSTEMS AND THEIR USE FOR EDUCATIONAL IMPROVEMENT
Copyright © 2010 J C Wayman, V Cho, and M P Richards

QUALIFICATIONS FRAMEWORKS AND THEIR ROLE IN THE REFORM OF EDUCATION AND TRAINING
Copyright © 2010 M Young

LANGUAGE OF VOCATIONAL EDUCATION AND TRAINING
Copyright © 2010 J Misko, J MacKenzie, and R-A Polvere

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44)(0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at (<http://elsevier.com/locate/permissions>), and selecting *Obtaining permissions to use Elsevier material*

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Catalog Number
A catalogue record for this book is available from the Library of Congress

ISBN: 978-0-08-044893-0

For information on all Elsevier publications
visit our website at books.elsevier.com

PRINTED AND BOUND IN THE UNITED KINGDOM
09 10 11 12 13 10 9 8 7 6 5 4 3 2 1

PREFACE

A preface usually provides a brief introduction to a work, intended to set the stage, provide some background insight, and whet the appetite of the reader. In our case, however, this preface has to address a fundamental question, one that was in our minds at the time we were recruited as Editors-in-Chief for the International Encyclopedia of Education. The question was “Why do we need an Encyclopedia? Its subtext was inspired by the ever-growing, ever-popular Internet. We believe that *this* Encyclopedia is desperately needed and will become a valued resource in education and associated social sciences and arts. The reasons are intellectual and procedural. Anyone with a modicum of knowledge knows that finding and trusting information gleaned from the Internet are two separate actions. The reliance on browsers to help discover references and comments result in resources based on popularity not quality. Pithy titles catch the eye and references rise in the ranks of browser searchers. Related to this is the “editing” in the Internet realm of populist efforts at encyclopedia, references, and other compilations. Once again, after removing offensive material, the accuracy, completeness, lack of bias, and other provenance for entries simply do not exist. Experienced researchers in education can sort through and make intelligent choices. Novices and many journeyman, or practitioners, parents, and policy makers cannot. Contrast how this Encyclopedia was built. Key domains of educational research were identified, and a tentative list of sub-domains or useful applied areas was posited. Then the Editors-in-Chief (apologies for the awkwardness of the term) identified the leading researcher in a particular domain, and with surprisingly little effort, recruited them to participate. They in turn identified the two best researchers in a sub-domain, such as formative assessment or the training of pre-school teachers. The authors of the sections of the Encyclopedia do not represent a collective group of friends and acquaintances, although friendships have been made. Rather they embody a deep and broad scholarly community. The difference from compiled Internet resources is the built-expertise and intellectual engagement of the authors. The summary of the developments and futures in their personal areas of scholarship have been filtered through their years of experience, both as scholars and communicators. Quality, then, is endemic to each piece, developed through this top-down identification of expertise, and made indelible by the bottom-up application of high standards from people leading the sub-domains – the authors, and the domains themselves, the section editors.

On a procedural level, the publishers early committed to the notion that this Encyclopedia would also be an online resource, and access would be available through print, for those with strong bookcases and the persisting love of turning real pages. The Internet version will allow multiple prisms through which the reader may access articles and provide, as it were, an emulation of the Internet in our field, albeit bounded by expertise and high quality.

What must be underscored in the assessment of this effort are the Editors-in-Chief and the publishers’ commitment to find excellence worldwide. We tried very hard to persuade notable scholars from all parts of the world to make contributions. Less than to fulfill the title of “International,” we were on the hunt for perspectives that would enrich the scope and depth of the sections. Our section editors put in enormous time attempting to find the best in the field, wherever they resided. Yet, not everyone is in the volume. Some were overcommitted. Many were not fully confident of their English, and the automated translation software has not yet met standards for technical writing. We believe that such writing and editing tools will make the outreach to an even broader International group of scholars possible in future revisions, or online updates. Furthermore, the birth of the World Educational Research Association (in 2009) will provide a better set of interlocking networks to find and evaluate scholarship from any place on the globe.

Finally, the scope of the effort must be acknowledged: 28 section editors, 926 articles were commissioned, drafted, reviewed, redrafted, edited, and put together in the space of four years. The publishers underwent some internal changes, and alterations in management. We as Editors-in-Chief, changed roles, moved, and also had to keep our own research and development enterprises afloat. Deadlines wobbled; authors dropped from view and had to be replaced.

Yet, at times frustrating as all development is, we find the final product exhilarating. We are enthusiastic not simply because it came into being at all, but because the collective light of the minds that wrote have left a bright resource for the future, one that will impact the way our colleagues understand and experience the educational knowledge, improvement, and impact in the future.

Eva Baker
Penelope Peterson

EDITORS

EDITORS-IN-CHIEF

Penelope Peterson

School of Education and Social Policy
Northwestern University
Evanston, IL
USA

Eva Baker

UCLA Center for the Study of Evaluation
Graduate School of Education and Information Studies
University of California, Los Angeles
Los Angeles, CA
USA

Barry McGaw

McGaw Group Pty Ltd
Mont Albert Nth
Melbourne, VIC
Australia

Gregory Chung

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Penny Enslin

Department of Educational Studies
University of Glasgow
Glasgow
UK

Noelle Griffin

CRESST
UCLA Center for the Study of Evaluation
Los Angeles, CA
USA

Dennis Hocevar

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

SECTION EDITORS

Walter Allen

HEOC Division
University of California, Los Angeles
Los Angeles, CA
USA

Vibeke Grøver Aukrust

Institute for Educational Research
University of Oslo
Oslo
Norway

Dominic J Brewer

Rossier School of Education
University of Southern California
Los Angeles, CA
USA

Kai-Ming Cheng

Faculty of Education
The University of Hong Kong
Hong Kong

David Hopkins

Institute of Education
University of London
London
UK

Mary James

Faculty of Education
University of Cambridge
Cambridge
UK

Sanna Järvelä

Department of Educational Sciences and
Teacher Education
University of Oulu
Oulu
Finland

Thomas Karmel

National Centre for Vocational Education Research
Adelaide, SA
Australia

Harold G Levine

School of Education
University of California, Davis
Davis, CA
USA

Robert L Linn

University of Colorado at Boulder
Ouray, CO
USA

Allan Luke

Faculty of Education
Queensland University of Technology
Brisbane, QLD
Australia

Ann M Mastergeorge

Human Development and Family Studies
University of California, Davis
Davis, CA
USA

Patrick J McEwan

Department of Economics
Wellesley College
Wellesley, MA
USA

Barbara Means

Center for Technology in Learning
SRI International
Menlo Park, CA
USA

Ingrid Moses

University of Canberra
Canberra, ACT
Australia

Bill Mulford

Faculty of Education
University of Tasmania
Hobart, TAS
Australia

Yu Jing Ni

Faculty of Education
The Chinese University of Hong Kong
Hong Kong

Festus Obiakor

Department of Exceptional Education
University of Wisconsin-Milwaukee
Milwaukee, WI
USA

P David Pearson

Graduate School of Education
University of California, Berkeley
Berkeley, CA
USA

Colin Power

Graduate School of Education
University of Queensland
Brisbane, QLD
Australia

Kjell Rubenson

Faculty of Education
The University of British Columbia
Vancouver, BC
Canada

Sandip Sinharay

Education Testing Services
Princeton, NJ
USA

Collette Tayler

Queensland University of Technology
Brisbane, QLD
Australia

Nico Verloop

ICLON, Graduate School of Education
Leiden University
Leiden
Netherlands

HOW TO USE THE ENCYCLOPEDIA

The International Encyclopedia of Education is intended for use by students, research professionals, and interested others. Articles have been chosen to reflect major disciplines in the study of education and common topics of research by academics in this domain. Each article serves as a comprehensive overview of a given area, providing both breadth of coverage for students, and depth of coverage for research professionals. We have designed the encyclopedia with the following features for maximum accessibility for all readers.

The contents of the encyclopedia are arranged alphabetically by section, and within sections, alphabetically by article. The Subject Index is located in Volume 8. Some topics are covered in a multitude of articles from differing perspectives, while other topics may have only one entry. We encourage use of the index for access to a subject area, rather than use of the Contents list alone, so that a reader has a full notion of the coverage of that topic.

The articles include cross-references to other related encyclopedia articles, suggested further readings where applicable, and many contain relevant websites for additional information. We encourage readers to use the cross-references to locate other encyclopedia articles that will provide more detailed information about a subject.

The Further Reading sections include recent secondary sources to aid the reader in locating more detailed or technical information. Review articles and research articles that are considered of primary importance to the understanding of a given subject area are also listed. These suggested further readings are not intended to provide a full reference listing of all material covered in the context of a given article, but are provided as next steps for a reader looking for additional information.

This page intentionally left blank

CONTENTS

Preface	v–vi
Editors	vii–viii
How to use the Encyclopedia	ix

VOLUME 1

A

Adult Education

Adult Education Overview	<i>K Rubenson</i>	1
--------------------------	-------------------	---

Adult Education – Adult Learning, Instruction and Program Planning

Adult Learning	<i>S B Merriam</i>	12
Adult Learning and Instruction: Transformative Learning Perspectives	<i>P Cranton</i>	18
Adult Learning in a Biographic Perspective	<i>L West</i>	25
Adult Learning, Instruction and Programme Planning: Insights from Freire	<i>P Mayo</i>	31
Characteristics of Adult Learning	<i>K Illeris</i>	36
Informal Learning: A Contested Concept	<i>P Hodgkinson</i>	42
Organizational Learning	<i>P-E Ellström</i>	47
Program Planning	<i>A L Wilson and R M Cervero</i>	53
Workplace Learning Frameworks	<i>S Billett</i>	58

Adult Education – as a Field of Study

Class Analysis in Adult Education	<i>T Nesbit</i>	64
Gender Analysis	<i>J Stalker</i>	70
Race and Ethnicity in the Field of Adult Education	<i>J Johnson-Bailey and D Drake-Clark</i>	77
Trends in Workplace Learning Research	<i>T Fenwick</i>	83

Adult Education – Concepts

Lifelong Learning	<i>J Field</i>	89
Rewriting the History of Adult Education: The Search for Narrative Structures	<i>B J Hake</i>	96

Adult Education – Domains and Provision

Adult Basic Education: A Challenge for Vocational Based Learning	<i>R Höghelm</i>	102
Adult Literacy Education	<i>L Tett and R St.Clair</i>	107
Citizenship and Immigrant Education	<i>D Schugurensky</i>	113
Community Based Adult Education	<i>B Connolly</i>	120
Continuing Professional Education: Multiple Stakeholders and Agendas	<i>R M Cervero and B J Daley</i>	127
Financing of Adult and Lifelong Learning	<i>Å Sohlman</i>	133
Labor Education	<i>K Forrester</i>	139
Learning Cities and Regions	<i>C Duke</i>	144
Museums as Sites of Adult Learning	<i>R S Grenier</i>	150
Overview of Lifelong Learning Policies and Systems	<i>D-B Kwon and C H O Daeyeon</i>	156
Popular Adult Education	<i>S Tøsse</i>	162
Provision of Prior Learning Assessment	<i>P Andersson</i>	169
The Age of Learning: Seniors Learning	<i>P Jarvis</i>	175
University Adult Continuing Education: The Extra-Mural Tradition Revisited	<i>M Slowey</i>	183

Adult Education – Economy and Society

Adult Education and Civil Society	<i>J Crowther and I Martin</i>	191
Adult Education and Nation-Building	<i>S Walters</i>	198
Economic Outcomes of Adult Education and Training	<i>A Ferrer and W C Riddell</i>	204
Health and Adult Learning	<i>P Bélanger and M Robitaille</i>	211
Modernization Processes and the Changing Function of Adult Learning	<i>H S Olesen</i>	217
The Political Economy of Adult Education	<i>G Rees</i>	223
Wider Benefits of Adult Education	<i>R Desjardins and T Schuller</i>	229

Adult Education – Participation

Barriers to Participation in Adult Education	<i>K Rubenson</i>	234
Participation in Adult Learning	<i>R Desjardins</i>	240

C**Comparative Education**

Comparative and International Education: English-Language Scholarly Publications	<i>D Post and M Fedri</i>	251
Comparative Education: Philosophical Issues and Concepts	<i>M Mason and C W Evers</i>	257
Comparative Education: Societies and Associations	<i>M Bray</i>	266

Curriculum Development

Curriculum and Religion	<i>C Kapitzke and P Graham</i>	272
Gender and Curriculum	<i>J Marsh</i>	277
Local Teacher-Based Curriculum Development	<i>M D Osborne and H B Barker</i>	283

Curriculum Development – Contexts

Curriculum and Human Rights	<i>C Lohrenscheit</i>	287
Curriculum and the Education of Cultural and Linguistic Minorities	<i>S May</i>	293
Curriculum in Postcolonial Contexts	<i>A Hickling-Hudson</i>	299
Curriculum, Economic and Cultural Development	<i>A Rampal</i>	307
Globalization and Curriculum	<i>R Dale</i>	312
Popular Culture	<i>K Schultz and R Throop</i>	318

Curriculum Development – Evaluation and Research

An Overview of Research in Curriculum Inquiry	<i>F M Connelly and S J Xu</i>	324
Curriculum Evaluation: Approaches and Methodologies	<i>V Klenowski</i>	335
Curriculum Studies, Discourse Analysis, and the Construction of Historical Time	<i>J Laspina</i>	342
Curriculum Use in the Classroom	<i>M Ben-Peretz and B Eilam</i>	348
The Roots and Routes of Teacher-Based Action Research and Curriculum Inquiry: An Historical Perspective	<i>L G Putney and J L Green</i>	355

Curriculum Development – Planning and Development

Curriculum and Syllabus Design	<i>A Woods, A Luke, and K Weir</i>	362
Curriculum and the Publishing Industry	<i>J Larson, A-R Allen, and D Osborn</i>	368
Curriculum and Teacher Change	<i>J G Ladwig</i>	374
Curriculum Governance and Planning	<i>B Levin</i>	379
Curriculum Planning and Systems Change	<i>Z Deng</i>	384
Curriculum, Digital Resources and Delivery	<i>Y Zhao, G Zhang, and C Lai</i>	390
Textbook Development and Selection	<i>P Shannon</i>	397

Curriculum Development – Subject Matter

Curriculum and Globalization: Higher Education	<i>K H Mok</i>	403
Curriculum Development as Subject Matter: Social Studies	<i>J Brophy and J Alleman</i>	411
Curriculum Development in the Area of Reading	<i>C L Madda and T E Raphael</i>	415
Curriculum in the Arts	<i>C M Thompson, L Bresler, and T E Costantino</i>	423
Curriculum: Foreign Language Learning	<i>Angel M Y Lin</i>	428
First Language Learning	<i>T Rogers and L Moffatt</i>	435
Learning a Second Language in First Language Environments	<i>E Bernhardt and D J Tedick</i>	441
Literature	<i>S Hynds</i>	447
Mathematics	<i>B Kaur and C P Vistro-Yu</i>	453
Physical Education and Sports	<i>D Kirk</i>	459
Primary and Elementary/Middle Grades Reading	<i>D E Alvermann and C A Mallozzi</i>	464
Science	<i>Y Kali and M Linn</i>	468

Curriculum Development – Theory

Bilingual Learning (Learning L1 and L2 in an L1 and L2 Environment)	<i>D Schwinge</i>	475
Curriculum and Complex Systems Theory	<i>W-M Roth and J Thom</i>	481

Curriculum and Constructivism	<i>B Davis and D Sumara</i>	488
Curriculum and Critical Theory	<i>L Yates</i>	494
Curriculum and Poststructuralist Theory	<i>J L Miller</i>	499
Curriculum and Structuralist Sociology: The Theory of Codes and Knowledge Structures	<i>J Muller and J Gamble</i>	505
Curriculum Reform	<i>R Gilbert</i>	510
Early Childhood Curriculum and Developmental Theory	<i>S M-Y Lim and C Genishi</i>	514

D

Demography and Social Change

An Evaluation Framework for E-Learning Effectiveness in the Arab World	<i>W AbuSneineh and M Zairi</i>	521
Globalization and Social Justice in Higher Education	<i>E Kimura-Walsh</i>	536
Racial Inequality and the Economic Payoffs of Higher Education	<i>S A Lachtman</i>	542
School Health Education	<i>J S Yang</i>	547

Demography and Social Change – Demography

A Paradigm of Contradictions: Racism and Science Education	<i>B A Brown and J W Mutegi</i>	554
Africana Studies: Past, Present And Future	<i>B R Hare, A Young, J B Stewart, and D P Aldridge</i>	565
Beyond Black Nationalism to Black Internationalism: The Compelling Case for Examining Black Educational Challenges Globally	<i>K Freeman</i>	573
Beyond Racial, Ethnic, and Gender Bias in Education Statistics	<i>J T Lott and C Bennett</i>	581
Caste and Education in India	<i>G G Wankhede</i>	589
Demography, Community, and the Education of Asian American and Pacific Islander Students	<i>R T Teranishi</i>	595
Educational Attainment and Mortality Differentials	<i>H D Horton, L Hixson, and S M Monnat</i>	602
Educational Opportunity and Latino/Chicano College Choice: New Findings, and Theoretical Perspectives	<i>M A Ceja and P A Pérez</i>	610
Ethnic Minority Identity and Educational Outcomes in a Rising China	<i>G A Postiglione</i>	616
Ethnic Studies and Women's Studies: From the Past to the Present, Providing a Space in the Academy for Scholarship Focused on the Histories and Experiences of Marginalized People	<i>C A Brown</i>	623
Gender Equity in Higher Education: Challenges and Celebrations	<i>L Morley</i>	629
Locating Space and Place in the College Access Debate: New Tools for Mapping and Understanding Educational Inequity and Stratification	<i>B Cantwell and J F Milem</i>	636
Public Policy and Inequality in Postsecondary Opportunity: Educational Statistics and the Failure of Education Reform	<i>E P St. John, K Williams, and K Moronski</i>	649
The Education of Indigenous Students	<i>J W Tippeconnic and S C Faircloth</i>	661
The Health Advantages of Educational Attainment	<i>G Y Dinwiddie</i>	667
To Go or Not to Go: Access and Barriers to Ethnic Minority Participation in Higher Education within the United Kingdom	<i>P Brug and T L Oliver</i>	673
Universal Secondary Education and the Two-Tiered School System in the English-Speaking Caribbean	<i>P Anderson and J Devonish</i>	679
Women of Color in Science, Technology, Engineering and Mathematics (STEM): Refining the Concepts, Reframing the Issues	<i>C B Leggon</i>	686

Demography and Social Change – Social Change

Educational Diversity in U.S. Law Schools	<i>C E Daye, A T Panter, W R Allen, M E Deo, and L F Wightman</i>	691
Educational Reform	<i>F Lopez-Segrera</i>	698
Emergence of For-profit Higher Education	<i>X Li</i>	705
Engaged Education: Experiential Learning, Intensive Field Experiences, and Social Change	<i>T R Rone</i>	711
Higher Education and Social Change in Mexico	<i>A Mungaray, G Estrella, P Moctezuma, and M Ramírez-Urquidy</i>	719
The Changing Role of the State in Higher Education	<i>K Gürüz</i>	728
The Role of the State in School Reform: Responding to Increasing Ethnic Diversity in Britain	<i>P Ratcliffe</i>	737
Transforming Mindsets Through Education for Sustainable Development	<i>V M Segovia</i>	746

Demography and Social Change – Sociology

Achievement Motivation in Ethnic Minority Youth	<i>S Graham and E Bryant</i>	753
Affirmative Action and Higher Education in Brazil	<i>J T dos Santos and D M Queiroz</i>	760
Black in White: Black Students at White and Black Colleges	<i>L Evans and J R Feagin</i>	767
Bologna Process: On the way to a Common European Higher Education Area	<i>A V Rezaev</i>	772
Critical Race Theory in Education Research	<i>R Vue and C B Newman</i>	779
Cultural Diversification and Japanese Education: Social Constructions of the New Diversity	<i>R Tsuneyoshi</i>	787
Education in War Countries: A Disempowerment of Society	<i>J L Nkata</i>	793
Educational Attainments of U.S. Black Males and Females: 1971 to 2003	<i>F B Bonner, M-C Jipguep-Akhtar, and R J Harrison</i>	799
Educational Preparation: Fostering the Self-Efficacy and Resilience of Urban Adolescent Youth	<i>M B Spencer and B Tinsley</i>	806
HIV Education for Low-Literate People: Transforming Students and Communities through Paulo Freire's Praxis and the Pedagogy of Action	<i>N Z Haniff</i>	814
How Are Discussions of Interdisciplinary Studies Linked to Diversity Discourse?	<i>C N Spence</i>	823
Identifying Patterns of Doctoral Attrition Across Academic Fields	<i>A James, K J Fosnacht, E B Stolzenberg, and D Thirumalai</i>	827
Male Underachievement in Education Across the Globe: A Shift in Paradigm for Gender Disparities Regarding Academic Achievement	<i>J F L Jackson, J L Moore III, and R A Leon</i>	838
Noncognitive Measures for Higher Education Admissions	<i>W E Sedlacek</i>	845
Peace Education	<i>J Page</i>	850
Remote Control: Africa's Development and Outflow to Developed Nations of Technical Talent	<i>F Haniff</i>	855
Scholar-Baller®: Student Athlete Socialization, Motivation and Academic Performance in American Society	<i>C K Harrison, S Bukstein, J Mottley, E Comeaux, J Boyd, C Parks, and D Heikkinen</i>	860
Service-Learning and Diversity as Mediums to Foster Civic and Community Engagement in Undergraduates	<i>L Oseguera</i>	866
Success and Failure in Education and Criminal Justice: Identifying Common Mechanisms	<i>S Ramirez and A Harris</i>	874
The History of Education: Race and Education	<i>J O Jewell</i>	881

VOLUME 2**E****Early Childhood Care and Education Theory**

- Early Childhood Care and Education: The Family, The Market, and The State *S B Kamerman and M J Neuman* 1
- Gender Issues in Early Childhood Education and Care *B Pocock* 11

Early Childhood Education and Care Governance

- Early Childhood Risk, Protection and Abuse Prevention *E Khoo* 17
- Evaluating Early Childhood Education and Care Programs *M Janus and S Brinkman* 25

Early Childhood Education and Care Investment

- Investing in Early Childhood Education and Care: The Health and Wellbeing Case *T Moore and F Oberklaid* 32
- Cost-effective Early Childhood Programs from Preschool to Third Grade *A J Reynolds, J A Temple, and B White* 38
- Early Childhood in Post-Modern Cultures: Thoughts and Some Concerns *P Gammage* 49
- Investing in Early Childhood Education and Care: Some Policy Implications *J Bennett* 55
- Investing in Early Childhood Education and Care: The Economic Case *G Cleveland and M Krashinsky* 63

Early Childhood Education and Care Programs

- Early Care and Education Programs for Infants and Toddlers *B Nyland* 69
- Literacies in Early Childhood: The Preschool Period *M Flear and B Raban* 75
- Participation in Early Childhood Education and Care Programs: Equity, Diversity and Educational Disadvantage *M Vandenbroeck* 81
- Teaching in Early Childhood Centers Instructional Methods and Child Outcomes *I Siraj-Blatchford* 86

Early Childhood Education and Care Provision

- Early Childhood Education and Care in the People's Republic of China *Zeng Xiaodong* 93
- Home-Based and Institutional Early-Childhood Education and Care Services *R Fukkink and H Blok* 98
- Integrated Early Childhood Education and Care Services – Care, Upbringing, Education and Health *S Colley* 104
- Issues of Access and Program Quality *J T Hustedt and W S Barnett* 110
- Kindergarten Transitions and Linkages to Primary School – Readiness Re-conceptualized *A Petriwskyj* 120
- Parent Support in Early Childhood – Approaches and Outcomes *K S Howard, K A Beckmann, and J Brooks-Gunn* 126

Early Childhood Education and Care Research

- Development and Implementation of Early Learning Standards in the United States *C Scott-Little* 132
- Global Status of Early Learning and Development Standards *P R Britto and S L Kagan* 138
- Rethinking Early Childhood Education and Care: Implications for Research and Evaluation *K Thorpe, D Cloney, and C Tayler* 144

Early Childhood Education and Care Theory

- Child Rearing and Early Education: Parents and Professionals: Theoretical and Historical Influences *H Penn* 151

Early Childhood Education and Care Workforce

- Early Childhood Education and Care Professional Profiles in the United Kingdom: Multiagency Teamwork and Transdisciplinary Challenges *A Anning and P Calder* 157
- Early Childhood Education and Care Staff Support: Mentoring and Professional Development in Finland *P Lindberg* 162

Economics of Education

- Education and Health *P Muennig* 169
- Data *M Perez and M Socias* 178
- Empirical Research Methods in the Economics of Education *P J McEwan* 187
- Theoretical Concepts in the Economics of Education *D J Brewer, G C Hentschke, and E R Eide* 193

Economics of Education – Costs and Financing of Schooling

- Cost–Benefit Analysis and Cost–Effectiveness Analysis *C Belfield and H M Levin* 199
- Economic Approaches to Adequacy *J Imazeki* 204
- Economic Approaches to School Efficiency *L L Taylor* 210
- School Finance Reform *T Downes* 216
- School Finance: An Overview *J K Rice, D H Monk, and J Zhang* 222
- The Economics of Tuition and Fees in Higher Education *R G Ehrenberg* 229

Economics of Education – External Benefits of Schooling

- Education and Civic Engagement *T S Dee* 235
- Education and Crime *L Lochner* 239
- Education and Economic Growth *E A Hanushek and L Wößmann* 245
- Parental Socioeconomic Status, Child Health, and Human Capital *J Currie and J Goodman* 253
- The External Benefits of Education *W W McMahon* 260

Economics of Education – Labor Markets, Education and Earnings

- Education and Inequality *J Blanden and S Machin* 272
- Human Capital *E R Eide and M H Showalter* 282
- Race Earnings Differentials *M Carnoy* 288
- Returns to Education in Developed Countries *M Gunderson and P Oreopoulos* 298
- Returns to Education in Developing Countries *H A Patrinos and G Psacharopoulos* 305
- School Quality and Earnings *J R Betts* 313
- Signaling in the Labor Market *M E Page* 321
- The Economics of High School Dropouts *R W Rumberger* 325

Economics of Education – Markets, Incentives and School Choice

- Competition and Student Performance *A A Payne* 331
- Educational Privatization *C Belfield and H M Levin* 337

Student Incentives	<i>E R Eide and M H Showalter</i>	342
The Economic Role of the State in Education	<i>D N Plank and T E Davis</i>	347
The Economics of Catholic Schools	<i>W Sander and D Cohen-Zada</i>	354
The Economics of Charter Schools	<i>R Zimmer and R Buddin</i>	360
The Economics of Parental Choice	<i>R Chakrabarti and J Roy</i>	367
The Economics of School Accountability	<i>D N Figlio and H F Ladd</i>	374
The Efficacy of Educational Vouchers	<i>R Zimmer and E P Bettinger</i>	380
Tiebout Sorting and Competition	<i>T J Nechyba</i>	388

Economics of Education – Production of Schooling

Desegregation, Academic Achievement and Earnings	<i>S Rivkin</i>	394
Education Production Functions: Concepts	<i>D N Harris</i>	402
Education Production Functions: Developed Country Evidence	<i>E A Hanushek</i>	407
Education Production Functions: Evidence from Developing Countries	<i>P Glewwe and S Lambert</i>	412
Family Environment in the Production of Schooling	<i>R Rothstein</i>	423
Neighborhoods and Peers in the Production of Schooling	<i>J Vigdor and J Ludwig</i>	431
Teacher Quality in Education Production	<i>I Pelayo and D J Brewer</i>	438
The Economics of Class Size	<i>D W Schanzenbach</i>	443
The Economics of Early Childhood Interventions	<i>M Nores</i>	450

Economics of Education – Teaching Labor Markets

An Overview of Teacher Labor Markets	<i>H Lankford and J Wyckoff</i>	456
Compensating Differentials in Teacher Labor Markets	<i>J Chambers</i>	465
Economic Approaches to Teacher Recruitment and Retention	<i>S Loeb and J Myung</i>	473
Teacher Incentives	<i>L Santibañez</i>	481
Teacher Supply	<i>P Dolton</i>	489
Teacher Training and Preparation in the United States	<i>M Podgursky, M G Springer, and R Hutton</i>	498
Teachers in Developing Countries	<i>F H Rogers and E Vegas</i>	504
The Economics of Teachers' Unions in the United States	<i>K Strunk</i>	512

Education of Children with Special Needs

Adapted Physical Activity for Students with Special Needs	<i>S R Hodge</i>	519
Assessment Accommodations for Children with Special Needs	<i>R J Kettler and S N Elliott</i>	530
Assistive Technology and Educational Practice	<i>H P Parette and G R Peterson-Karlan</i>	537
Augmentative–Alternative Communication	<i>C Mason and K M Chinn</i>	544
Community Integration and Employment of Youth with Special Needs	<i>B A Ford and A W Daviso III</i>	550
Cooperative Learning for Children with Special Needs	<i>E A Dooley and K R Kossar</i>	555
Co-Teaching	<i>E D Cramer</i>	560
Cultural and Linguistic Diversity and Children with Special Needs	<i>J M Patton, N Day-Vines, and A McSwain</i>	565
Curriculum-Based Assessment and Students with Special Needs	<i>D J O'Shea and J D Katsafanas</i>	571

Diagnosis and Classification	<i>M Sapp</i>	577
Direct Instruction and the Education of Children with Special Needs	<i>T Mehring</i>	584
Disabilities and Gifted Education	<i>V I Daniels and M J McCollin</i>	588
Early Childhood Special Education: Birth to Eight Years Old	<i>T S Lovelace, R Gardner III, J B Axe, and D M Sainato</i>	594
Early Intervention for Students with Special Needs	<i>P A Grant</i>	608
Early Transition of Children with Special Needs	<i>S A Fowler</i>	613
Educating Children Who are Deaf-Blind	<i>J Wood and K M Chinn</i>	621
Educating Children with Deafness and Hearing Impairments	<i>G M Mukuria and C J Eleweke</i>	628
Educating Children with Other Health Impairments (ADHD)	<i>G M Mukuria and F E Obiakor</i>	634
Educating Children with Physical Disabilities	<i>Z P Nkabinde, F E Obiakor, M T Ofor, and D J Smith</i>	640
Educating Children with Specific Learning Disabilities	<i>P A Grant and P B Grant</i>	646
Educating Children with Speech and Language Disorders	<i>S L Green and C D Qualls</i>	654
Educating Students with Autism and Related Disorders	<i>T T Dyches</i>	661
Educating Students with Cognitive Disabilities	<i>J P Bakken</i>	669
Educating Students with Emotional and Behavioral Disorders	<i>S W Smith and G G Taylor</i>	678
Educating Students with Special Needs: An Overview	<i>F E Obiakor, M K Harris, P Harris-Obiakor, and G M Mukuria</i>	688
Educating Students with Traumatic Brain Injury	<i>S Burkhardt and A Rotatori</i>	695
English Language Learners with Special Needs	<i>R Rueda and G Ragusa</i>	701
Functional Behavioral Assessment	<i>J P Bakken</i>	708
History of Special Education	<i>L M Garrick Duhaney and S J Salend</i>	714
Inclusion of Students with Special Needs in General Education Classrooms	<i>M A Prater</i>	721
Instructional Accommodations for Children with Special Needs in Inclusive Settings	<i>A L Yurick, S Musti-Rao, and G Cartledge</i>	727
Interagency Agreements	<i>L A Owens</i>	733
Interdisciplinary Collaboration in Early Childhood	<i>N File</i>	739
Language Development in Children with Special Needs	<i>M S L Stewart</i>	745
Large Scale Assessment and Accountability and Students with Special Needs	<i>M L Thurlow</i>	752
Literacy Instruction for Students with Special Needs	<i>L H Mason</i>	759
Math Instruction for Children with Special Needs	<i>B A Bottge</i>	767
Overrepresentation of Culturally and Linguistically Diverse Students in Special Education	<i>S C Trent</i>	774
Paraprofessionals in Special Education	<i>B Y Ashbaker, S C Dunn, and J Morgan</i>	780
Parent and Family Involvement in the Education of Children with Special Needs	<i>L P Prater</i>	788
Peer Relations and Socialization of Children and Adolescents with Special Needs	<i>L M Bullock, M Wong-Lo, and R A Gable</i>	794
Peer-Tutoring	<i>L J O'Shea and D J O'Shea</i>	802
Postsecondary Participation of Students with Special Needs	<i>R A Stodden and K D Roberts</i>	808
Prereferral Strategies	<i>C G Simpson and J P Bakken</i>	815

Preventing Antisocial Behavior and Delinquency: A Comprehensive School-wide Approach <i>P R Hallam, K R Young, P Caldarella, D G Wall, and L Christensen</i>	820
Programs and Instructional Strategies for Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	829
Related Services for Children with Special Needs <i>S O Obi</i>	834
Risk and Resilience Frameworks in Understanding Special Education <i>W-Y Hsieh and M L Donahue</i>	840
School Dropout Prevention <i>F E Obiakor</i>	845
School-Based Services for Children with Special Needs <i>S O Obi</i>	850
Self-Determination <i>M L Wehmeyer</i>	857
Social Competence <i>L A Owens and S Johnston-Rodriguez</i>	865
The Identification of Students with Gifts and Talents <i>V I Daniels and M J McCollin</i>	870
Transition from School to Adult Life <i>M J McCollin and F E Obiakor</i>	876
Visual Impairment <i>W Sapp</i>	880
Workforce Issues in Special Education <i>E Carlson and B Billingsley</i>	886

VOLUME 3

Education of Professionals

An Overview of Accreditation, Certification, and Licensure Processes <i>J J Norcini and M van Zanten</i>	1
Educating America's Military Officers <i>M H Adamshick</i>	7
Education for Architecture in the United States and Canada <i>S Schreiber</i>	13
Education for Medicine <i>M B Anderson</i>	19
Education for the Legal Profession <i>J W Wegner</i>	25
Education of Clergy <i>D O Aleshire</i>	31
Framing Professions Education <i>P S O'Sullivan</i>	37
Issues in Accounting/Business Education <i>W H Gijsselaers and R G Milter</i>	39
Professions – Engineering <i>J Heywood</i>	45
Professions – Health-Related <i>G M Jensen and C B Royeen</i>	51
Professions – Pharmacy <i>D E Beck</i>	58
Teaching for the Complex Practice of Nursing <i>P Benner and M Sutphen</i>	65
Veterinary Education <i>K M Edmondson</i>	71

Education Research Methodology: Quantitative Methods and Research

Abductive Research Methods <i>B D Haig</i>	77
Field Experimentation in Education <i>M A Constan</i>	83
Internal Validity <i>M H Clark and S C Middleton</i>	90
Methods for Approximating Random Assignment <i>B Schneider and S-K McDonald</i>	97
Methods for Approximating Random Assignment: Regression Discontinuity and Propensity Scores <i>L M Scheier</i>	104

Multitrait–Multimethod Designs	<i>M Bong and S Hong</i>	111
Multivariate Longitudinal Data Analysis	<i>H Liu and L Chang</i>	120
Quasi-Experimentation: Two Group Design	<i>E Hong</i>	128
Regression Discontinuity Designs	<i>A Kelava, S Rohrmann, and V Hodapp</i>	134
Sampling	<i>R S Brown</i>	142
Statistical Conclusion Validity	<i>P L Busk</i>	147
Survey Research Methods	<i>A Fink</i>	152

Educational Assessment

An Overview of Educational Assessment	<i>M James</i>	161
Assessment and the Evaluation of Institutional Effectiveness	<i>S M Thomas</i>	172
Assessment in Schools – Dispositions	<i>R D Crick</i>	181
Assessment in Schools Related To Literacy: Reading	<i>K J Topping</i>	189
Impact of Assessment on Students' Learning Strategies and Implications for Judging Assessment Quality	<i>M Segers, F Dochy, and D Gijbels</i>	196

Educational Assessment – Assessment Activities

Alternative Assessment	<i>G Stobart and C Gipps</i>	202
Classroom Assessment Tasks and Tests	<i>J Phelan and J Phelan</i>	209
Dynamic Assessment	<i>J G Elliott, E L Grigorenko, and W C M Resing</i>	220
Instructional System Provided Feedback	<i>H F O'Neil, S Chuang, and T Huang</i>	226
Peer and Self-assessment	<i>C Harrison</i>	231
Portfolio Assessment	<i>V Klenowski</i>	236
Records of Achievement: Beyond Traditional Tests	<i>P Broadfoot</i>	243

Educational Assessment – Assessment in Domains

Assessment in Higher Education	<i>D R Sadler</i>	249
Assessment in Schools – Affective Domain	<i>M Ventura, R D Roberts, and M Zeidner</i>	256
Assessment in Schools – Creative Subjects	<i>T E Costantino and L Bresler</i>	262
Assessment in Schools – Learning to Learn	<i>A Hautamäki, J Hautamäki, and S Kupiainen</i>	268
Assessment in Schools – Literacy Writing (Extended)	<i>J M Parr and L Earl</i>	273
Assessment in Schools – Mathematics	<i>M Brown and J Hodgen</i>	279
Assessment in Schools – Oracy	<i>A L Bailey</i>	285
Assessment in Schools – Primary Science	<i>W Harlen</i>	293
Assessment in Schools – Secondary Science	<i>J Shemwell, A Fu, M Figueroa, R Davis, and R Shavelson</i>	300
Assessment in Schools – Technology Education and ICT	<i>A Jones, B Cowie, and J Moreland</i>	311
Assessment in the Early Years	<i>M J Drummond</i>	316
Assessment in the History, Civics and Social Studies Domains	<i>P G Avery</i>	323
Assessment in the Workplace of Performance, Developing Expertise and Competence	<i>A J Brown</i>	330
Assessment in Vocational Education	<i>K Ecclestone</i>	337

Educational Assessment – Concepts and Issues

Assessing Group Work	<i>M Galton</i>	342
Assessment and the Regulation of Learning	<i>L Allal</i>	348
Challenges of Developing and Implementing Formative Assessment Practices in Schools	<i>R Sutton</i>	353
Formative Assessment	<i>P Black</i>	359
Formative Assessment and Instructional Planning	<i>B Marshall</i>	365
Progression and Assessment: Developmental Assessment	<i>M Forster and G Masters</i>	369
Students' Rights in Assessment and Evaluation	<i>J Gardner and B Holmes</i>	378
Summative Assessment by Teachers	<i>R Daugherty</i>	384
The Multiple Purposes of Assessment	<i>P E Newton</i>	392
The Relationship between Assessment and the Organization and Practice of Teaching	<i>L B Resnick, D William, R Apodaca, and E S Rangel</i>	397

Educational Assessment – Enhancing Quality and Use

Assessment Conversations: Reading and Writing Conferences with Students and with Parents	<i>M Barrs</i>	403
Assessment: Pre-service and In-service Teacher Education	<i>A Davies</i>	410
Classroom Assessment in Policy Context (Australia)	<i>J J Cumming</i>	417
Classroom Assessment in Policy Context (England and France)	<i>M Osborn and M Raveaud</i>	425
Classroom Assessment in Policy Context (French Sources)	<i>D Laveault</i>	432
Classroom Assessment in Policy Context (Hong Kong)	<i>D Carless</i>	438
Classroom Assessment in Policy Context (New Zealand)	<i>T J Crooks</i>	443
Classroom Assessment in Policy Context (Sub-Saharan Africa)	<i>S J Howie, L Zimmerman, and K Draper</i>	449
Moderation of Student Work by Teachers	<i>G S Maxwell</i>	457
School Policies and Practices to Support Effective Classroom Assessment for Learning	<i>D Pedder</i>	464
Student Assessment: Policy and Practice in Eastern Europe	<i>G Bethell</i>	472

Educational Assessment – Key Relationships

Assessment Practice in Policy Context: Latin American Countries	<i>F M Rizo</i>	479
Impact of Assessment on Learner Groups (Boys/Girls)	<i>J Elwood</i>	486
Impact of Assessment on Learner Groups (Disabilities)	<i>M Rouse</i>	492
Impact of Assessment on Students' Test Anxiety	<i>P Buchwald and C Schwarzer</i>	498
Impact of Assessments on Classroom Practice	<i>J L Herman</i>	506
Social Practices in School Assessment and their Impact on Learner Identities	<i>A Pollard and A Filer</i>	512

Educational Evaluation – Concepts, Practice, and Future Directions

Educational Evaluation: Concepts, Practice, and Future Directions	<i>Y J Ni</i>	518
---	---------------	-----

Educational Evaluation – Contemporary Themes in Evaluation

Evaluating E-Learning	<i>D D Williams and C R Graham</i>	530
Evaluating Schools as Learning Communities	<i>S M Thomas</i>	539
Evaluation of Adult Education and Training Programs	<i>G McNamara, P Joyce, and J O'Hara</i>	548
Evaluation of Integrated Health Programs in School	<i>T M Franke</i>	555

Formative Assessment in Teacher Education and Teacher Professional Development	<i>H Tillema</i>	563
Informal Education and Evaluation	<i>T Jeffs</i>	572
Methods to Evaluate Technology	<i>W L Bewley</i>	578

Educational Evaluation – Contexts of Educational Evaluation

Cost Analysis in Evaluation Studies	<i>J D Fletcher</i>	585
Cultural Issues that Can Affect the Validity of Educational Evaluations	<i>R P Durán</i>	592
Evaluation and Accountability	<i>E Sato and S N Rabinowitz</i>	600
Evaluation Governance and Planning	<i>R Picciotto</i>	607
Evaluation Use	<i>V G Thomas</i>	614
Moral and Ethical Issues in Evaluation	<i>M Morris</i>	622
The Role of Stakeholders in Educational Evaluation	<i>S Taut and M C Alkin</i>	629

Educational Evaluation – Evaluation Domains

Curriculum Evaluation	<i>T Levin</i>	636
Evaluating Education in Three Policy Eras	<i>M A Peters, R Britez, and S M Weber</i>	645
Evaluation of Mathematics Education Programs	<i>J Cai</i>	653
Evaluation of Science Education Programs	<i>G Berberoğlu</i>	660
Evaluation of Social Studies and Civic Education Programs	<i>S G Grant</i>	667
Evaluation of Teacher Quality and Practice	<i>W Robinson and J Campbell</i>	674
International Evaluations	<i>A Schleicher</i>	681
National Assessment Programs: The Example of the U.S. National Assessment of Educational Progress	<i>M Schneider</i>	690
Principal Evaluation: Concepts, Needs and Realities	<i>N S Glasman and L D Glasman</i>	698
Program Evaluation	<i>C A Christie and L A Fierro</i>	706
School Based Evaluation: Purposes, Protocols and Processes	<i>J MacBeath</i>	713
School Inspection/External School Evaluation	<i>C Chapman and P Earley</i>	719
The International System for Teacher Observation and Feedback: A Theoretical Framework for Developing International Instruments	<i>L Kyriakides, B P M Creemers, C Teddlie, and D Muijs</i>	726

Educational Evaluation – Evaluation Methods

Evaluation Methodology	<i>M Chatterji</i>	735
Evaluation Reporting and Communicating	<i>R T Torres</i>	746
Longitudinal Evaluation Designs	<i>X Ma</i>	757
Meta-evaluation: Purpose, Prescription, and Practice	<i>L A Wingate</i>	765

Educational Evaluation – Purposes and Models of Educational Evaluation

Defining Quality in Evaluation	<i>P Dahler-Larsen</i>	775
Internal and External Evaluation	<i>D Nevo</i>	781
Needs Assessment in Education	<i>J W Altschuld and T L Lepicki</i>	786
The Purpose of Educational Evaluation	<i>S Mathison</i>	792
Understanding Approaches to Evaluation	<i>A J Love</i>	798

VOLUME 4**Educational Measurement**

Ability Testing	<i>P B Tymms</i>	1
Admissions Testing	<i>R Zwick</i>	7
Appropriateness Measurement: Person Fit	<i>R R Meijer and A M Weekers</i>	15
Automated Essay Scoring: Writing Assessment and Instruction	<i>M D Shermis, J Burstein, D Higgins, and K Zechner</i>	20
Classical Test Theory Reliability	<i>M D Miller</i>	27
Criterion-Referenced Measurement	<i>S A Shrock and W C Coscarelli</i>	31
Differential Item Functioning	<i>M D Hidalgo and J Gómez-Benito</i>	36
Educational Measurement: Overview	<i>R L Linn</i>	45
Equating and Scaling	<i>A A von Davier</i>	50
Fairness	<i>C Gipps and G Stobart</i>	56
Generalizability Theory	<i>R L Brennan</i>	61
Impact of Assessment on Classroom Practice	<i>J L Herman</i>	69
Intelligent Systems	<i>V J Shute and D Zapata-Rivera</i>	75
Item Response Theory	<i>W J van der Linden</i>	81
National Assessments	<i>T J Crooks</i>	89
Norm-Referenced Measurement	<i>H D Hoover</i>	95
Progress Monitoring	<i>L S Fuchs and D Fuchs</i>	102
Rasch Models	<i>D Andrich</i>	111
Scoring Rubrics	<i>J A Arter</i>	123
Second Language Assessment	<i>L F Bachman</i>	140
Self Adaptive Testing	<i>J Revuelta</i>	148
Student Test Results in School Accountability	<i>L S Hamilton</i>	153
Test Development	<i>S M Downing</i>	159
Test Translation and Adaptation	<i>A Allalouf and P Hanani</i>	166
Testing Creativity	<i>M A Runco</i>	170
Testing in History	<i>S Wineburg and B Fogo</i>	175
Validity	<i>R L Linn</i>	181
Validity of Achievement Gains on High-Stakes Tests	<i>D Koretz</i>	186
Validity of Educational Indicators	<i>B P M Creemers and L Kyriakides</i>	193

G**Globalization of Universities and Institutions**

Globalization, Universities and Medium of Instruction	<i>J Lo Bianco</i>	201
The Washington Accord on Engineering Curriculum	<i>E Ko</i>	209

H

Higher Education

Higher Education: An Overview	<i>P Scott</i>	217
-------------------------------	----------------	-----

Higher Education – Management, Leadership and Governance

Academic Managers – Heads of Departments/Schools	<i>S Marshall</i>	229
Leading and Managing the University – Presidents and their Senior Management Team	<i>R Middlehurst, T Kennie, and S Woodfield</i>	238
Management of and in Higher Education Institutions	<i>B Sporn</i>	245
Managerialism and Collegialism in Higher Education Institutions	<i>M Shattock</i>	251
Strategic Planning in Higher Education	<i>L Zechlin</i>	256
The Changing Role of University Presidents, Vice-Chancellors and Rectors	<i>M F Green and P D Eckel</i>	264
The Changing Roles of University Governing Boards and Councils	<i>J Fielden</i>	273

Higher Education – Research and Scholarship

Funding of University Research	<i>G Harman</i>	279
Mobility of PhD Students and Scientists	<i>S Avveduto</i>	286
National Science Policy and Universities	<i>M Henkel and U Kogan</i>	294
Research Commercialization	<i>S Slaughter</i>	300
Research Quality Assessment	<i>I McNay</i>	307
The Management of University Research	<i>J Kirkland</i>	316
The Training of New Scientists	<i>R Neumann</i>	322

Higher Education – Society

Academic Freedom	<i>S F Macintyre</i>	328
Access and Equity in Higher Education	<i>G Baldwin and R James</i>	334
Boundaries of Institutional Autonomy and their Impact on Higher Education	<i>V L Meek</i>	341
Diversity of Higher Education	<i>U Teichler</i>	347
Employability of University Graduates and Graduate Outcomes	<i>J Brennan</i>	354
Higher Education and the Knowledge Society	<i>J Välimaa</i>	360
Higher Education and the Labor Market	<i>J Enders</i>	365
Higher Education and the Transformation of Society	<i>P Scott</i>	370
Quality Assurance in Higher Education – Practices and Issues	<i>D Dill</i>	377
The Role of Accreditation of Higher Education Institutions	<i>J S Eaton</i>	384
Universities and Regional Development	<i>J Goddard and J Puukka</i>	390
Universities' Engagement with Society	<i>D Watson</i>	398

Higher Education – Teaching and Learning in Higher Education

Distance Education and Open Universities	<i>A Kanwar and J Daniel</i>	404
Education Development in Higher Education	<i>Y Ryan and K Fraser</i>	411
Effective Use of Technology in Teaching and Learning in HE	<i>D Laurillard</i>	419

Facilitating Students' Success – Student Support and Learning Resources in Universities	<i>E Woodberry and C Barnes</i>	427
Faculty Values and Expectations and Their Implications for Teaching, Learning and the Organization of Higher Education Institutions	<i>K Harman</i>	433
Flexible Learning in Higher Education	<i>S Alexander</i>	441
Innovation in Teaching and Curriculum Design	<i>J M Dearn</i>	448
Productivity of University Faculty Staff	<i>S Kyvik</i>	455
Promoting Effective Student Learning in Higher Education	<i>K Trigwell</i>	461
Student Attrition from Higher Education Institutions	<i>M Hayden</i>	467
University Students' Experiences of Higher Education	<i>C McInnis</i>	473
What Do University Students Learn?	<i>G D Kuh and B K Bridges</i>	479

Higher Education – Tertiary Education in a Globalized World

Associations of Universities	<i>G F Moodie</i>	485
Collective Bargaining in Higher Education	<i>C Allport, K McAlpine, and S Roberts</i>	491
Financing of Higher Education	<i>B Chapman and P Tulip</i>	499
Higher Education Crossing Borders	<i>J Knight</i>	507
Higher Education in Post-Conflict Societies	<i>M Daxner</i>	514
Higher Education in the New Economies	<i>J Fielden</i>	520
Human Resource Issues in Higher Education	<i>E El-Khawas</i>	527
International Accreditation in Higher Education	<i>A Stella and D Woodhouse</i>	533
Internationalization of Higher Education	<i>M van der Wende</i>	540
National and International Rankings of Higher Education	<i>S Marginson</i>	546
Steering of Higher Education Systems – The Role of the State	<i>K Yokoyama and V L Meek</i>	554
Student and Faculty Transnational Mobility in Higher Education	<i>U Lanzendorf and B M Kehm</i>	559
The Bologna Process in European Higher Education	<i>D Rich</i>	566
The Economics and Finance of Higher Education	<i>A Wolf</i>	573
The Role of the OECD in the Development of Higher Education in a Globalized World	<i>R Yelland</i>	584
Transforming Higher Education in Developing Countries: The Role of the World Bank	<i>J Salmi and R M Bassett</i>	590
UNESCO's Role in the Development of Higher Education in a Globalized World	<i>S Uvalić-Trumbić</i>	597

Higher Education – The Organization and Financing of Higher Education

Higher Education in Federal Systems	<i>M A M Cortés and U Teichler</i>	603
Pathways and Articulation into Higher Education	<i>G F Moodie</i>	609
Unitary, Binary and other National Systems of Higher Education	<i>L Goedegebuure</i>	616
Viewing Private Higher Education: How Much, Where, Why, and What?	<i>D C Levy</i>	622

I

Immigration

Children of Migrant Populations	<i>M M Suárez-Orozco and C Suárez-Orozco</i>	629
Third Culture Kids	<i>R E Van Reken</i>	636

International Comparisons

IEA Studies in ICT	<i>N Law</i>	644
IEA Studies in Mathematics and Science	<i>F K S Leung</i>	650
IEA Study in Civic Education	<i>J Torney-Purta, J-A Amadeo, and J Schwillie</i>	656
IEA: Globalization and Assessment	<i>H Wagemaker</i>	663

International Organizations in Education

Education and the European Union	<i>S Hegarty</i>	669
National University on Educational Planning and Administration, New Delhi	<i>J B G Tilak</i>	676
On School Management – International Council on School Effectiveness and Improvement, Commonwealth Council on Educational Administration and Management	<i>A Walker and T Townsend</i>	681
UNESCO's International Institute for Educational Planning	<i>S J Hite and M Bray</i>	688

L**Leadership and Management**

Leadership and Management Overview	<i>B Mulford</i>	695
------------------------------------	------------------	-----

Leadership and Management – as a Field of Study

Contexts for Leadership at the Beginning of the 21st Century	<i>M Bottery</i>	704
Educational Leadership: Philosophical Issues	<i>C W Evers</i>	710
Ethical/Moral Issues in Educational Leadership	<i>R J Starratt, L Langlois, and P Duignan</i>	717
History of Educational Leadership/Management	<i>R Bates</i>	724
Educational Leadership as a Field of Study: Professional Journals 1994–2006	<i>A R Thomas</i>	731
Research on Educational Leadership – Approaches/Promising Directions	<i>W A Firestone and V M J Robinson</i>	740

Leadership and Management – Leader Preparation

Leaders for Productive Schools	<i>J Murphy, S N Elliott, E B Goldring, and A C Porter</i>	746
New Approaches in Preparing School Leaders	<i>S G Huber</i>	752
Research for Leaders: The National College for School Leadership	<i>G Southworth</i>	762
School Leadership Standards	<i>L Ingvarson</i>	768

Leadership and Management – Leadership Issues

Effective Leadership in Challenging Schools	<i>A Harris</i>	777
Esthetics, Educational Leadership and Management	<i>R Bates</i>	783
Leader Mistreatment of Teachers	<i>J Blase and J Blase</i>	790
Leadership and Gender	<i>J Blackmore</i>	797
Leadership and Student Voice	<i>C Lanskey and J Rudduck</i>	803
Leadership and Urban Education	<i>K A Riley</i>	809
Leadership for Learning	<i>J MacBeath</i>	817
Leadership in Diverse Cultures	<i>A Walker and P K Yu-kwong</i>	824

Leading Diverse Schools	<i>J Ryan</i>	832
Supervision and Personnel Administration	<i>P D Tucker and D G Pounder</i>	839
The Big Challenge: Leadership for Inclusion	<i>M Ainscow and A Sandill</i>	846
The Educational Leader's New Work	<i>P Gronn</i>	852

VOLUME 5

Leadership and Management – Leadership Types

A Distributed Perspective on School Leadership and Management	<i>J P Spillane</i>	1
Leadership: Authentic	<i>P T Begley</i>	7
Leadership: Democratic	<i>J Møller</i>	12
Leadership: Instructional	<i>P Hallinger, K Leithwood, and R H Heck</i>	18
Leadership: Transformative	<i>C M Shields</i>	26
Strategic Leadership	<i>B Davies and B J Davies</i>	34
Teacher Leadership and Organizational Development	<i>A Harris</i>	40
Understanding How Leadership Influences Student Learning	<i>K Leithwood and B Levin</i>	45

Leadership and Management – Politics and Governance

Equity and Educational Effectiveness	<i>P Sammons</i>	51
Leading/Managing Schools in Communities Made Poor	<i>P Thomson</i>	58
Principals Role in Restructuring Schools	<i>G M Crow and K D Peterson</i>	66
School Reform and Restructuring: Self Managing School	<i>B J Caldwell</i>	72
System Leadership	<i>D Hopkins and R Higham</i>	78

Leadership and Management – School and Community

Educational Leadership and Social Capital	<i>J West-Burnham</i>	85
Networked Learning Communities School-to-School Collaboration as an Essential Component of a System Reform Strategy	<i>D Jackson</i>	92
Networks and Communities of Knowledge	<i>T Schuller and H Theisens</i>	100
School Community Relations	<i>R L Crowson and E B Goldring</i>	106
Social Capital, Educational Institutions and Leadership	<i>S Kilpatrick, S Johns, and B Mulford</i>	113

Leadership and Management – School Effectiveness and Improvement

Leadership and Technology	<i>C F Webber</i>	120
Leadership in the Implementation of Innovations	<i>I E Bogotch, T Townsend, and M Acker-Hocevar</i>	128
Leadership: School Improvement	<i>R H Heck and P Hallinger</i>	135
Organizational Learning in Schools	<i>B Mulford and H Silins</i>	143
Professional Learning Community	<i>L Stoll</i>	151
Transformational School Leadership	<i>K Leithwood</i>	158

Learning and Cognition – Issues and Concepts

Attention in Cognition and Early Learning	<i>M Heimann, T Tjus, and K Strid</i>	165
Concept Learning	<i>S J Unsworth and D L Medin</i>	172
Cooperative Learning	<i>R E Slavin</i>	177
Intelligence	<i>R J Sternberg</i>	184
Learning from Multiple Information Sources	<i>H I Strømsø and I Bråten</i>	191
Memory	<i>S Magnussen and T Brennen</i>	197
Metacognition	<i>L Baker</i>	204
Personal Epistemology in Education: Concepts, Issues, and Implications	<i>I Bråten</i>	211
Piaget: Recent Work	<i>L Smith</i>	218
The Adult Development of Cognition and Learning	<i>E A L Stine-Morrow and J M Parisi</i>	225
Vygotsky and Recent Developments	<i>J V Wertsch</i>	231

Learning and Cognition – Issues and Concepts – Focus on Cognition

Cognition and Emotion	<i>F Pons, M de Rosnay, and F Cuisinier</i>	237
Cognition: Overview and Recent Trends	<i>J Wiley and B D Jee</i>	245
Development of Creativity	<i>M A Runco</i>	251
Knowledge Domains and Domain Learning	<i>L Maggioni and P A Alexander</i>	255
Problem Solving and Human Expertise	<i>T J Nokes, C D Schunn, and M T H Chi</i>	265
Problem Solving and Reasoning	<i>R E Mayer</i>	273
Reasoning, Moral and Social	<i>N S Jampol, C B Richardson, and M Killen</i>	279
Remembering as Social Activity	<i>D Middleton</i>	285

Learning and Cognition – Issues, Concepts, Types

Computer-Supported Collaborative Learning: Basic Concepts, Multiple Perspectives, and Emerging Trends	<i>S R Ludvigsen and A I Mørch</i>	290
Learning as Inquiry	<i>S R Goldman, J Radinsky, S Tozer, and D Wink</i>	297
Learning Through Narrative Socialization	<i>A Fasulo</i>	303
Neuroscience Bases of Learning	<i>M H Immordino-Yang and K W Fischer</i>	310
The Neuroscience of Reading	<i>U Goswami</i>	317

Learning and Cognition – Issues, Concepts, Types – Focus on Learning

Learning Strategies	<i>C E Weinstein, J Jung, and T W Acee</i>	323
Learning Through Play	<i>K Aronsson</i>	330
Situated View of Learning	<i>A Collins and J G Greeno</i>	335
The Neuroscience of Aging and Cognition	<i>W J Hoyer and B Bessette-Symons</i>	340

Learning and Cognition – Language, Literacy and Subject-Related

An Overview of Language and Literacy in Educational Settings	<i>V G Aukrust</i>	345
Bilingualism and Learning	<i>L S Verplaetse and E Schmitt</i>	355
Classroom Discourse and Student Learning	<i>L Hemphill</i>	361
Early Writing	<i>B E Hagtvet</i>	367

First Language Acquisition	<i>N B Ratner</i>	375
Foreign Language Learning	<i>C Kinginger</i>	382
Learning Science	<i>J Leach and P Scott</i>	387
Learning to Read	<i>M S Burns and J K Kidd</i>	394
Mathematics Learning	<i>L Verschaffel, B Greer, and E De Corte</i>	401
Music Learning	<i>S Hallam</i>	407
Reading Comprehension: Reading for Learning	<i>C E Snow</i>	413
Second Language Learning	<i>J S Arnfast, J N Jørgensen, and A Holmen</i>	419
Writing: Advanced	<i>R Andrews</i>	426

Learning and Cognition – Other Issues

Gender and Learning	<i>L Yates and J McLeod</i>	432
Parent–Child Relationships in Early Learning	<i>C P Edwards, S M Sheridan, and L L Knoche</i>	438
Peer Interaction and Learning	<i>S Blum-Kulka and S Dvir-Gvirsman</i>	444

Learning and Cognition – Places of Learning

Learning in a Cross-Cultural Perspective	<i>L-F Zhang and R J Sternberg</i>	450
Learning Outside of School	<i>T Nunes</i>	457
Reasoning and Explanatory Talk: Learning in Everyday Settings	<i>D E Beals</i>	464

Learning and Cognition – Theoretical Perspectives – Learning

Apprenticeship Approach to Learning	<i>K Nielsen</i>	469
Connectionism and Learning	<i>T R Shultz</i>	476
Constructivism and Learning	<i>S Sjøberg</i>	485
Cultural–Historical Activity Theory	<i>A Havnes</i>	491
Learning in a Sociocultural Perspective	<i>R Säljö</i>	498
Self-Regulated Learning and Socio-Cognitive Theory	<i>P H Winne and A F Hadwin</i>	503

N

National Systems of Education

Australia	<i>C N Power</i>	510
Austria	<i>M J Gaertner</i>	517
Bangladesh	<i>R Z Dewan</i>	524
Barbados	<i>R Rudder</i>	534
Bulgaria	<i>A Chavdarova and D Vesselinov</i>	540
England	<i>C Whatford</i>	550
Estonia	<i>M Keränen and J Kõrgesaar</i>	555
Fiji	<i>I Senibulu</i>	563
Finland	<i>J E Tuovinen</i>	569
France	<i>M Sellier</i>	576

Germany	<i>A Hüfner and K Hüfner</i>	582
Ghana	<i>R Palmer</i>	590
Greece	<i>K Toura and A Karagiannis</i>	599
Iceland	<i>M Hardardottir and M Gunnlaugsdóttir</i>	605
Indonesia	<i>A C Alwasilah and Furqon</i>	611
Israel	<i>A Volansky</i>	618
Jamaica	<i>E Hannam</i>	627
Latvia	<i>A Joma</i>	630
Lebanon	<i>N A Mneimneh</i>	636
Lithuania	<i>J Navickaitė, L Paurienė, R Zablackė, A Tijūnelytė, and V Kožemiakina</i>	656
Malawi	<i>F R Mkandawire and D Mulera</i>	666
Malta	<i>C Farrugia</i>	673
Mongolia	<i>B Yembuu</i>	681
Namibia	<i>J Ellis</i>	687
New Zealand	<i>R Baker and C Wylie</i>	692
Niger	<i>A A I Daouda</i>	699
Norway	<i>H Farstad</i>	709
Oman	<i>S A Issan</i>	718
Pakistan	<i>N Mahmood and R Malik</i>	727
Poland	<i>A Janowski</i>	737
Qatar	<i>S R Salman</i>	744
Russia	<i>V M Filippov and G F Tkach</i>	750
Rwanda	<i>E Bahizi</i>	758
Scotland	<i>D Raffe</i>	770
Singapore	<i>P T Ng</i>	776
Slovenia	<i>T Plevnik and A B Lakota</i>	782
Solomon Islands	<i>D P Malasa</i>	790
South Africa	<i>P Kgobe and J Pampallis</i>	799
Spain	<i>A Tiana</i>	806
Sri Lanka	<i>B M J Balasooriya, M M Wehella, and S U Wijeratne</i>	813
Suriname	<i>H R Ori</i>	826
Taiwan	<i>C-M Lee</i>	834
Tanzania	<i>R Wedgwood</i>	841
Thailand	<i>C Rukspollmuang</i>	847
The Gambia	<i>Y A-M Jobe</i>	858
The Philippines	<i>E A P Valenzuela</i>	865
Uruguay	<i>B Guinovart and C Troncoso</i>	873
USA	<i>B Pusser and M E R Lloyd</i>	888
Zanzibar	<i>I M Hija</i>	896

VOLUME 6**P****Philosophy of Education**

Philosophy of Education: Overview	<i>P Enslin</i>	1
-----------------------------------	-----------------	---

Philosophy of Education – Contemporary Issues

Neoliberalism, the Market and Performativity	<i>M A Peters</i>	11
Parents and Children's Rights	<i>M Clayton</i>	17
Peace Education	<i>I Gur-Ze'ev</i>	22
Philosophy and Educational Research	<i>D Bridges</i>	29
Race, Critical Race Theory and Whiteness	<i>B Applebaum</i>	36
School Choice and the Common School	<i>G Haydon</i>	44

Philosophy of Education – Philosophical Perspective

African Philosophy of Education	<i>K Horsthemke</i>	50
Feminism	<i>L Stone</i>	56
Hermeneutics	<i>C Bingham</i>	63
Islamic Education	<i>Y Waghid</i>	69
Liberalism and Education	<i>W Feinberg</i>	74
Phenomenology and Existentialism	<i>B Levering</i>	80
Pragmatism	<i>B Thayer-Bacon</i>	86
Psychoanalysis and Education	<i>P Smeyers</i>	92
Social and Cultural Capital in Education	<i>M Print</i>	97
The Aims of Education	<i>R Marples</i>	103
The Analytical Tradition	<i>D Carr</i>	108

Philosophy of Education – Philosophical Themes

Adult Learning: Philosophical Issues	<i>D Beckett</i>	114
Autonomy	<i>C A Wringe</i>	121
Bildung	<i>J Masschelein and N Ricken</i>	127
Critical Theory and Pedagogy	<i>M Mason</i>	133
Critical Thinking	<i>H Siegel</i>	141
Environmental Education	<i>M Bonnett</i>	146
Identity	<i>M S Merry</i>	152
Inclusion	<i>R Cigman</i>	158
Justice and Care	<i>M Tjiattas</i>	164
Lifelong Learning	<i>J Halliday</i>	170
Post-Structuralism and Education	<i>M Mason and M Clarke</i>	175
Reason and Rationality	<i>P Hogan</i>	183

Religious and Spiritual Education	<i>J C Conroy and S J McKinney</i>	189
The Capabilities Approach	<i>L Terzi</i>	194

Primary and Secondary Education – Learning and Teaching in School Age Education

Class Size – Arguments and Evidence	<i>P Blatchford and K C Lai</i>	200
Continuing Professional Development of Teachers	<i>P Earley</i>	207
Homeschooling	<i>P Farenga</i>	214
Innovations and Early Intervention in the Teaching of Literacy	<i>R E Slavin</i>	221
Personalized Learning in School Age Education	<i>D Hopkins</i>	227
Perspectives on Schooling in the Middle Years	<i>V J Russell</i>	233
Reform of the School Workforce	<i>H M Gunter and G Forrester</i>	242
Secondary Education in Developing Countries	<i>P Hughes and A Walker</i>	249

Primary and Secondary Education – Social Justice Issues in School Age Education

Community Focused Schools	<i>A Dyson</i>	255
---------------------------	----------------	-----

Primary and Secondary Education – The Effectiveness of School Systems

Reforming the American School System	<i>M Barber</i>	261
School Effectiveness in Developed Societies	<i>D Reynolds</i>	269
School to School Collaboration: Innovation and Improvement	<i>R Higham</i>	274
Studies of School Improvement in Developing Countries	<i>S E Anderson</i>	281
Sustainable Educational Reform	<i>A Hargreaves</i>	289
The Role of the District in Tri Level Reform	<i>M Fullan</i>	295
Whole School Designs for Enhancing Student Achievement	<i>S Stringfield and J Nunnery</i>	303

Q

Qualitative Research

Action Research in Education	<i>E Stringer</i>	311
Classroom Ethnography	<i>F Erickson</i>	320
Computer Assisted Qualitative Data Analysis	<i>C Silver and A Lewins</i>	326
Conversational Analysis	<i>H G Levine</i>	335
Critical Ethnography	<i>D Foley</i>	337
Critical Race Theory	<i>D Gillborn and G Ladson-Billings</i>	341
Critical Theory	<i>J J Ryoo and P McLaren</i>	348
Cultural Studies	<i>R Gaztambide-Fernández and K Cairns</i>	354
Cultural–Historical Activity Theory	<i>Laboratory for Comparative Human Cognition</i>	360
Design Experiments	<i>A Collins</i>	367
Discourse Analysis	<i>P Nikander and S Vehviläinen</i>	373
Document Analysis	<i>T Rapley and K N Jenkins</i>	380
Ethnography	<i>M Hammersley</i>	386

Ethnomethodology in Education Research	<i>D Macbeth</i>	392
Focus Groups	<i>D L Morgan and K Hoffman</i>	401
Grounded Theory	<i>K Charmaz and A Bryant</i>	406
Hermeneutics	<i>J Dyer</i>	413
Interpretive Research	<i>E Rauscher and M E Graue</i>	419
Interviews and Interviewing	<i>A Marvasti</i>	424
Life History	<i>J A Hatch and S K Newsom</i>	430
Narrative Inquiry	<i>D J Clandinin and J Huber</i>	436
Participant Observation	<i>J Schostak</i>	442
Phenomenology	<i>M van Manen and C A Adams</i>	449
Qualitative Case Studies	<i>S B Merriam</i>	456
Qualitative Data Management	<i>C Grbich</i>	463
Semiotics	<i>M Danesi</i>	475
Symbolic Interactionism, Naturalistic Inquiry, and Education	<i>J A Forte</i>	481
The Quality of Evidence in Qualitative Research	<i>S K Biklen</i>	488
Visual Data in Education Research	<i>J Wagner</i>	498

Qualitative Research – Standards of Proof

Validity: Mapping Diverse Perspectives	<i>G B Rossman, S F Rallis, and A M Kuntz</i>	505
--	---	-----

Quantitative and Qualitative Approaches to Research – Integration

Mixed Methods	<i>A Tashakkori and I Newman</i>	514
---------------	----------------------------------	-----

Quantitative Methods and Research Design

External Validity	<i>G E Matt, A Brewer, and M Sklar</i>	521
-------------------	--	-----

R

Role of NGOs in Globalization

NGOs and Globalization of Education	<i>M Ahmed</i>	529
-------------------------------------	----------------	-----

S

Social and Emotional Aspects of Learning

Achievement Goal Theory: Definitions, Correlates, and Unresolved Questions	<i>T Urdan</i>	537
Affect, Mood and Emotions	<i>S Schnall</i>	544
Anxiety	<i>M Zeidner</i>	549
Attribution Theory	<i>B Weiner</i>	558
Children's Friendship	<i>S R Asher and K L McDonald</i>	564
Coping with Stressful Situations: An Important Aspect of Self-Regulation	<i>M Boekaerts</i>	570
Culture in Motivation Research: A Challenging and Enriching Contribution	<i>K Kimmel and S Volet</i>	576
Early Social Development and Schooling	<i>G M van der Aalsvoort</i>	585

Emotion in Educational Contexts	<i>P A Schutz, P D Quijada, S de Vries, and M Lynde</i>	591
Empathy	<i>R Arnold</i>	597
Flow in Education	<i>J A Schmidt</i>	605
Interest	<i>M Ainley</i>	612
Intrinsic and Extrinsic Motivation	<i>C Levesque, K J Copeland, M D Pattie, and E L Deci</i>	618
Motivating Students in Classrooms	<i>J Brophy</i>	624
Motivation Regulation	<i>C A Wolters and S A Mueller</i>	631
Peer Learning in the Classroom	<i>N M Webb</i>	636
Poverty, Effects of on Social and Emotional Development	<i>N Philipsen Hetzner, A D Johnson, and J Brooks-Gunn</i>	643
Seeking Help as an Adaptive Response to Learning Difficulties: Person, Situation, and Developmental Influences	<i>S A Karabenick and R S Newman</i>	653
Self-Concept in Learning: Reciprocal effects model between academic self-concept and academic achievement	<i>H W Marsh and L F Scalas</i>	660
Self-Efficacy Beliefs	<i>D H Schunk and F Pajares</i>	668
Social and Emotional Outcomes of Learning	<i>J Vorhaus</i>	673
Social Aspects of Collaborative Learning	<i>M Arvaja and P Häkkinen</i>	685
Social Development and Schooling	<i>G W Ladd</i>	691
Social Interaction and Learning	<i>K Littleton</i>	698
Social Networks and the Education of Children and Youth	<i>B Schneider, T G Ford, and L Perez-Felkner</i>	705
Sociocultural Issues in Motivation	<i>R A Walker</i>	712
Socioemotional Aspects of Technology-Supported Learning	<i>S Järvelä</i>	718
Volitional Control of Learning	<i>J Husman and L Corno</i>	724
Wellbeing	<i>T Hascher</i>	732

VOLUME 7

Statistics

An Overview of Statistics in Education	<i>S Sinharay</i>	2
Analysis and Interpretation of Multivariate Data	<i>D J Bartholomew</i>	12
Analysis of Covariance	<i>N T Longford</i>	18
Analysis of Extreme Values in Education	<i>J Beirlant, G Dierckx, K Luwel, and P Onghena</i>	25
Analysis of Variance	<i>B M King</i>	32
Bayesian Statistical Analysis	<i>G Kokolakis</i>	37
Bootstrap Method	<i>K Singh and M Xie</i>	46
Canonical Correlation	<i>A Basu and A Mandal</i>	52
Categorical Data Analysis	<i>S E Fienberg and B W Junker</i>	58
Causal Inference	<i>D B Rubin</i>	66
Cluster Analysis: Overview	<i>S Landau and I Chis Ster</i>	72

Cognitive Psychology and Educational Statistics	<i>H S Kim and S Embretson</i>	84
Computational Statistics	<i>J E Gentle</i>	93
Continuous Probability Distributions	<i>S Sinharay</i>	98
Correspondence Analysis	<i>M Greenacre</i>	103
Data Mining	<i>R S J d Baker</i>	112
Decision Theory	<i>M Mendoza and E Gutiérrez-Peña</i>	119
Design of Experiments	<i>C A Albers and T R Kratochwill</i>	125
Discrete Probability Distributions	<i>S Sinharay</i>	132
Discrimination and Classification	<i>R C Neath and M S Johnson</i>	135
Empirical Bayes Methods	<i>E Ip and G Molenberghs</i>	142
Evaluation Research	<i>T Kellaghan</i>	150
Exploratory Data Analysis	<i>A Unwin</i>	156
Factor Analysis: An Overview and Some Contemporary Advances	<i>F A N Alhija</i>	162
Generalized Linear Mixed Models	<i>S Rabe-Hesketh and A Skrondal</i>	171
Generalized Linear Models	<i>J J Faraway</i>	178
Generating Random Numbers	<i>P J Pashley and A Amodeo</i>	184
Goodness-of-Fit Testing	<i>A Maydeu-Olivares and C García-Forero</i>	190
Graphical Models	<i>R G Almond</i>	197
Growth Modeling	<i>I Plewis</i>	203
Hierarchical Linear Models	<i>D Rindskopf</i>	210
Hypothesis Testing and Confidence Intervals	<i>T R Konold and X Fan</i>	216
Instrumental Variables	<i>E P Bettinger</i>	223
Jackknife Methods	<i>S Sinharay</i>	229
Large-sample Statistical Methods	<i>J M Singer and D F Andrade</i>	232
Latent Class Models	<i>J K Vermunt</i>	238
Markov Chain Monte Carlo	<i>A M Johansen</i>	245
Matrix Algebra	<i>H Yanai and H Ishii</i>	253
Measure of Association	<i>G-H Huang</i>	260
Measures of Central Tendency	<i>D Clark-Carter</i>	264
Measures of Dispersion, Skewness and Kurtosis	<i>D Rindskopf and M Shiyko</i>	267
Meta Analysis	<i>J Sánchez-Meca and F Marín-Martínez</i>	274
Missing Data	<i>C A W Glas</i>	283
Model Selection	<i>S J Haberman</i>	289
Monte Carlo Methods	<i>A M Johansen</i>	296
Multidimensional Scaling	<i>Z Zhang and Y Takane</i>	304
Multiple Comparisons	<i>C Lewis</i>	312
Multivariate Analysis of Variance	<i>N T Longford</i>	319
Multivariate Linear Regression	<i>S Ganesh</i>	324
Multivariate Normal Distribution	<i>J Tacq</i>	332

Nonlinear Regression Analysis	<i>H-H Huang, C K Hsiao, and S-Y Huang</i>	339
Nonparametric Statistical Methods	<i>K Sijtsma and W H M Emons</i>	347
Observational Studies	<i>P R Rosenbaum</i>	354
Order Statistics	<i>I Bairamov</i>	360
Point Estimation Methods with Applications to Item Response Theory Models	<i>F Bartolucci and L Scrucca</i>	366
Principal Components Analysis	<i>D J Bartholomew</i>	374
Probability Theory	<i>T Rudas</i>	378
Recursive Partitioning	<i>S Landau and S Barthel</i>	383
Sampling	<i>J Qian</i>	390
Sequential Probability Ratio Test	<i>T J H M Eggen</i>	396
Sequential Testing	<i>H J Vos</i>	401
Signal Detection Theory	<i>H Abdi</i>	407
Small Area Estimation	<i>A Noble</i>	411
Statistical Analysis of Functional Data	<i>F Ferraty, P Sarda, and P Vieu</i>	417
Statistical Inequalities	<i>H M Mahmoud and J L Gastwirth</i>	426
Statistical Paradoxes	<i>Y-K Tu</i>	429
Statistical Power Analysis	<i>L V Hedges and C Rhoads</i>	436
Statistical Significance Versus Effect Size	<i>X Fan and T R Konold</i>	444
Stochastic Processes	<i>F Baudoin</i>	451
Structural Equation Models	<i>S-Y Lee and X-Y Song</i>	453
Survival Data Analysis	<i>A F De Champlain</i>	459
The Normal Distribution and its Applications	<i>J Tacq</i>	467
Time Series Analysis	<i>P J Brockwell</i>	474
Univariate Linear Regression	<i>M L Hazelton</i>	482
Value-Added Models	<i>D F McCaffrey</i>	489

T

Teacher Education – Evaluation in the Teacher Education Program

Accreditation and Standards in Teacher Education	<i>R J Yinger and K L Daniel</i>	495
Determining Long Term Effects of Teacher Education	<i>C N Brouwer</i>	503

Teacher Education – Evaluation of the Prospective Teacher

The Evaluation of Prospective Teachers in Teacher Education	<i>D E H Tigelaar and J van Tartwijk</i>	511
---	--	-----

Teacher Education – Macro Factors

Diversity in Teacher Education	<i>T Wubbels</i>	518
Social and Cultural Influences on Teacher Education	<i>M Ben-Peretz and R Lotan</i>	525
The Role of Governance in Teacher Education	<i>G E Karlsen</i>	532

Teacher Education – Post-Initial Professional Development

Action Research as a Tool for Teachers' Professional Development	<i>P Ponte</i>	540
Contemporary Approaches to Teacher Professional Development	<i>H Borko, J Jacobs, and K Koellner</i>	548
School Development for Teacher Learning and Change	<i>P Sleegers and K Leithwood</i>	557
Teacher Induction	<i>D Beijaard, J Buitink, and C Kessels</i>	563
Teacher Learning as Workplace Learning	<i>J Imants and K van Veen</i>	569
Teacher Learning with Lesson Study	<i>A Murata</i>	575
Teachers Career Stages and Professional Development	<i>R Fessler and E Rice</i>	582

Teacher Education – Preservice Teacher Education

A Pedagogy of Teacher Education	<i>J Loughran</i>	587
Case Methods in Teacher Education	<i>M Sato and C Rogers</i>	592
Conceptions of Teacher Education	<i>J A Whitcomb</i>	598
Inquiry-Oriented Teacher Education	<i>R Schulz</i>	604
Narratives and Biography in Teacher Education	<i>G Kelchtermans</i>	610
Supervision in Teacher Education	<i>J McCarthy and L F Quinn</i>	615
Taking Prospective Teachers' Beliefs into Account in Teacher Education	<i>M Valcke, G Sang, I Rots, and R Hermans</i>	622
Teacher Education and Models of Teacher Reflection	<i>V K LaBoskey</i>	629
Teacher Education as Teaching for Understanding with New Technologies	<i>M S Wiske and D E Spicer</i>	635

Teacher Education – Preservice: The Knowledge Base

Experienced Teachers' Craft Knowledge	<i>P C Meijer</i>	642
Moral Values in Teacher Education	<i>W Veugelers</i>	650
Pedagogical Content Knowledge	<i>J H Van Driel and A Berry</i>	656
Teacher Education and the Educational Foundations Knowledge Base	<i>F B Murray</i>	662
The Relationship Between Theory and Practice in Teacher Education	<i>F A J Korthagen</i>	669

Teacher Education – Teacher Educators

Characteristics, Scholarship and Research of Teacher Educators	<i>M Lunenberg</i>	676
Professional Development of Teacher Educators	<i>P K Smith</i>	681
Self-Study by Teacher Educators	<i>T Russell</i>	689

Teacher Education – Teaching Specific Domains

Arts Teacher Education	<i>L Beudert</i>	695
Mathematics Teacher Education	<i>K Krainer and S Llinares</i>	702
Music Teacher Education	<i>M W-y Wong</i>	706
Science Teacher Education	<i>J H Van Driel and S K Abell</i>	712
Social Studies Teacher Education	<i>A-L Halvorsen and S M Wilson</i>	719
Teacher Education for Elementary Education	<i>K D Brown and G N Carnes</i>	725

Teacher Education for Teaching the Gifted	<i>A Y Baldwin</i>	732
Teacher Preparation in Mother Tongue Education	<i>M Kooy and H Hulshof</i>	738
Vocational Teacher Education	<i>S Johansson</i>	744

Teacher Education – The Cooperating School

Mentoring in Teacher Education	<i>P D Tomlinson, A J Hobson, and A Malderez</i>	749
Partnerships Between Schools and Higher Education	<i>C Day and L Smethem</i>	757

VOLUME 8

Technology and Learning

An Overview of Technology and Learning	<i>B Means and J Roschelle</i>	1
Educational Data Modeling	<i>S Iksal</i>	11
Student Data Systems and Their Use for Educational Improvement	<i>J C Wayman, V Cho, and M P Richards</i>	14
Technology and Learning: Access in Schools Around the World	<i>R Anderson and J Ainley</i>	21

Technology and Learning – as a Subject of Instruction

Conceptions of Technology Literacy and Fluency	<i>O Erstad</i>	34
--	-----------------	----

Technology and Learning – Assessment

Technology and Formative Assessment	<i>J W Pellegrino</i>	42
Technology for Large-Scale Assessment	<i>R E Bennett</i>	48
Technology Supports for Assessment Design	<i>R J Mislevy, F I Winters, I I Bejar, R E Bennett, and G D Haertel</i>	56

Technology and Learning – Education Reform and Economic Development

National Strategies to Build a Technology Workforce	<i>H-M Cheah</i>	66
Public–Private Partnerships for Educational Reform	<i>V Quah, C P Lim, and C Brook</i>	73
Relating Technology, Education Reform and Economic Development	<i>R B Kozma</i>	81

Technology and Learning – Evaluating Technology

Examining the Effects of Educational Technology Programs: Challenges and Strategies	<i>M Russell</i>	88
---	------------------	----

Technology and Learning – Global Trends

Equity in Technology Access and Opportunities	<i>T Matuchniak and M Warschauer</i>	95
Internet-Based Education	<i>M Bakia</i>	102

Technology and Learning – Issues

Media Use and School Achievement	<i>K Roe</i>	109
Technology and Physical and Social Health	<i>K Subrahmanyam</i>	112

Technology and Learning – Meeting Special Students’ Needs

Universal Design for Learning *D H Rose and J W Gravel* 119

Technology and Learning – School Leadership for Technology Integration

Technology as a Support for School–Community Connections *K M Culp* 125

Technology and Learning – Supports for Managing Instruction

Classroom Uses of Technology to Manage Instruction *W R Penuel* 133

Technology and Learning – Supports for Skill Learning

New Media, Learning from *W Schnotz and H Horz* 140

Reading and Technology *N DeFrance, D Khasnabis, and A S Palincsar* 150

Technological Supports for Acquiring Twenty-First-Century Skills *C Dede* 158

Technology Supports for Acquiring Inquiry Skills *T de Jong* 167

Technology Supports for Acquiring Mathematics *M J Nathan* 172

Technology Supports for Lifelong Learning *A Collins and R Halverson* 184

Technology and Learning – Supports for Student Collaboration

Knowledge Communities in the Classroom *J D Slotta and H Najafi* 189

Technology and Learning – Supports for Subject Matter Learning

Technology Supports for Science Learning *C Delgado and J Krajcik* 197

Technology Supports for Second Language Learning *D Nunan* 204

Technology and Learning – Teacher Learning

Teacher Skills and Knowledge for Technology Integration *N Law* 211

Technology in Preservice Teacher Education *N Davis* 217

Technology Resources for Teacher Learning *J E Hinojosa, C Labbé, and L López* 222

V

Vocational Education and Training

An Overview of Vocational Education and Training *T Karmel* 229

Informal and Non-Formal Learning in Vocational Education and Training *F Beddie and S Halliday-Wynes* 240

School-Based Vocational Education and Training *B Knight and R Sweet* 247

The Economics of Vocational Education and Training *P Carneiro, L Dearden, and A Vignoles* 255

Vocational Education and Training – Individuals and Students

Age Equality in Education and Training *P Taylor* 262

Career Guidance in Vocational Education and Training and in the Workplace *P Plant* 269

Displaced Workers, Unemployed and Vocational Education and Training *B Kriechel* 275

Immigrant Investment in Host Country Vocational Education and Training *I Kogan* 280

Indigenous People and Vocational Education and Training	<i>R G Schwab</i>	289
Persons with a Disability and Vocational Education and Training	<i>F Ferrier and C S Smith</i>	294
Prisoner Education and Training	<i>S Dawe, S H Wynes, and J Misko</i>	300
Vocational Education and Training and the School-to-Work Transition	<i>G Hayward</i>	306

Vocational Education and Training – Industry and Employers

Apprenticeships	<i>E Smith</i>	312
Skills Shortages: Concepts, Measurement and Policy Responses	<i>C Shah and G Burke</i>	320
Training and Learning in the Workplace	<i>R L Jacobs</i>	328
Training in Small Businesses	<i>H Matlay</i>	337

Vocational Education and Training – Research on VET

Language of Vocational Education and Training	<i>J Misko, J MacKenzie, and R-A Polvere</i>	343
Research on Vocational Education and Training	<i>J Robson and B Bailey</i>	352
Vocational Education and Training Information Resources	<i>L-A Harris, R Naidu, and M Saunders</i>	358

Vocational Education and Training – Teaching and Learning

Employability Skills	<i>E Smith</i>	368
Learning Styles in Vocational Education and Training	<i>J Vorhaus</i>	376
Modularization in Vocational Education and Training	<i>H Ertl and G Hayward</i>	383
Personalized Learning and Vocational Education and Training	<i>I Duckett</i>	391
Recognition of Prior Learning and Vocational Education and Training	<i>J Keating</i>	397
The Competency Model	<i>A Gonczi and P Hager</i>	403

Vocational Education and Training – VET in Context

Globalization and Vocational Education and Training	<i>P Loveder</i>	411
Learning to Work for the Future	<i>J Fien, R Maclean, and D N Wilson</i>	416
The Relationship between Technical and Vocational Education and Training and Innovation	<i>K Ruth and L Deitmer</i>	423
The Status of Vocational Education and Training	<i>J Lasonen</i>	429
Vocational Education and Training in Post-Conflict Countries	<i>E Lyby</i>	441

Vocational Education and Training – VET System

Dual System	<i>T Deissinger</i>	448
Industry Involvement in the Vocational Education and Training System	<i>M Hughes and C Hughes</i>	455
Non-Completion in Vocational Education and Training: Issues, Patterns and Possibilities	<i>S Lamb and A Walstab</i>	463
Planning and Policy Development for Technical Vocational Education and Training Systems	<i>R Maclean and M Pavlova</i>	469
Qualifications Frameworks and their Role in the Reform of Education and Training	<i>M Young</i>	476
Quality Assurance in Vocational Education and Training	<i>D Gunning</i>	482

Training Finance	<i>A Ziderman</i>	489
Training Markets	<i>P Ryan</i>	496
Vocational Education and Training Teacher Training	<i>S Nielsen</i>	503
Vocational Education and Training Workforce	<i>D L Wallin</i>	513
Subject Index		519
Contributors		747

TECHNOLOGY AND LEARNING

Contents

An Overview of Technology and Learning

Educational Data Modeling

Student Data Systems and Their Use for Educational Improvement

Technology and Learning: Access in Schools Around the World

An Overview of Technology and Learning

B Means and J Roschelle, SRI International, Menlo Park, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Agent – A software program that assists the user in performing some task, taking into account the nature of the user and the situation.

Asynchronous communication technology – Any technology designed to support an exchange of messages with posting and reading occurring at different times; examples include email and blogs.

Avatar – A graphic representation of oneself that a user either creates or chooses to represent his/her identity online; for example in a chat room or multiplayer game.

Clickers – Devices used to register a response as part of a classroom response system.

Immersive environment – A computer-created scene or world that gives the user the sensation of being within the scene rather than being outside of it.

Information and communication technology – A general term encompassing computer hardware and software as well as their application.

Instructional management software – System that manages the gathering, processing, and display of student data from computer-based instruction.

Online learning – Learning that occurs over the Internet.

Open source – A computer program for which the source code is made available for free use and modification by others.

Podcast – Digital media file (either audio or video) distributed over the Internet for download and playback on a mobile device or personal computer.

Recommendation system – Software that uses information about a user, in conjunction with data on the opinions and choices of all users of the system, to identify and present items or products that are likely to interest the user.

Response system – Hardware and software that enable members of a class or audience to respond simultaneously to a question from an instructor or presenter, who receives a tabulation or other representation of the aggregated responses.

Social networking site – A Web site that supports building online communities by having users post a public or semi-public representation of themselves and indicate other users with whom they wish to share information; examples include Facebook, Twitter and MySpace.

Synchronous communication system – Systems in which all users are logged on at the same time; examples include videoconferencing, Internet telephony and instant messaging.

Virtual manipulative – A software representation of a thing or system that the user can interact with.

Advances in information and communication technology (ICT) have enabled its use to support teaching, learning, and assessment both within and outside of schools. The current state of access to ICT within schools and ways in which ICT is used are described below. At the same time, it is noted that the increased use of ICT in schools has not produced the dramatic changes in pedagogy and learning content that some technology proponents had predicted. Formal educational systems impose curriculum standards, high-stakes tests, and criteria for promotion or selection into educational programs at higher grade levels that reinforce continuity in educational approaches. Research on the implementation of ICT in schools and classrooms has found that elements outside the technology – including principal support, teacher training, the design of the learning activity in which ICT is used, and assessment practices – influence the extent to which ICT is used and

its effectiveness as a tool for learning. Thus, those who would like to see ICT transform the nature of education are still waiting for the long-promised radical shift. Reformers can point to current technology trends as potential sources of major change, however. These technology trends include the increasing availability of open-source course content on the Internet; the dramatic rise of collaborative, user-generated content (termed Web 2.0); and the high level of engagement triggered by immersive, multi-user games with complex structures and highly realistic, dynamic graphics.

Spread of Technology Within Schools

Few areas of education have experienced as much change over the last 15 years as the use of ICT to support learning. In 1994, when the prior edition of this encyclopedia was published, the World Wide Web was in its infancy. At that time, only 35% of public schools in the United States had Internet access anywhere in the building and just 3% of instructional classrooms had a computer connected to the Internet. By 2005, in contrast, 94% of instructional rooms had computers with Internet access (Wells and Lewis, 2006). The United States was not alone in making a major investment in technology in schools during this time frame. By 2006, schools in the 22 countries participating in the SITES 2006 international survey had, on average, one computer with Internet access for every seven students (Law *et al.*, 2008).

During the same time period, the cost of computing power dropped dramatically, and smaller, lighter computing devices became available in the form of personal digital assistants, MP3 players, mobile phones, and tablet PCs. Today's iPhone is vastly more powerful than a top-of-the-line 1994 desktop computer.

Further, advances in projection technology have made it feasible to provide bright, lightweight projectors to every classroom. Indeed, many classrooms go beyond projectors to provide an electronic whiteboard on which the teacher can write or draw, interact with computer applications, and access resources on the Internet.

The declining cost of computers and the increase in digital resources for learning have led a number of schools, districts, and countries to launch one-to-one computing initiatives to provide every student with a computer. For example, in 2005 Australia launched a AU\$1.2-billion plan to provide every secondary student in years 9–12 with a computer. At the same time, it must be acknowledged that the playing field with respect to access to ICT that can support learning is still not level. A survey of principals in 56 countries participating in the 2006 PISA assessment of 15-year-olds, for example, found that schools in the 14 countries with the strongest technology infrastructures had an Internet computer for roughly every three or four

students, while the 14 countries with the least technology had an Internet computer for every 25 or more students, and three countries reported no instructional computers with Internet access in their schools.

There is no doubt that technology will continue to become more available to teachers and students. The more difficult challenge has proven to be the design, development, and dissemination of educationally meaningful learning activities supported by technology.

School Uses of Technology for Learning

Technology use within schools has been promoted since the mid-1980s. Early predictions of how personal computers and learning software would transform education, however, have given way to an appreciation of the complexity of technology integration and of educational change. Some critics have argued variously that ICT has had no significant impact on schools (Cuban *et al.*, 2001) or that it has had negative impacts (Healey, 1999). Even so, technology continues to be a part of school instruction, albeit a less common instructional practice than more traditional pedagogies such as teacher lecture or question-and-response cycles.

In a 2005 survey of US teachers, 45% reported requiring their students to use technology once a week or more (Bakia *et al.*, 2008). The most often teacher-reported weekly uses of technology were to practice or review mathematics or reading topics, enrichment activities in mathematics and reading, and taking tests or quizzes online. Student use of technology for their school work does not necessarily occur within school, however. Across the 40 countries with students' participating in the PISA 2006 survey, only 9% of 15-year-olds reported using a computer at school almost every day. Even among countries with the largest number of Internet computers in classrooms, none had as many as a third of its students reporting use of a computer at school on a daily basis. In contrast, 75% or more of students in more than a third of the participating countries reported home use of computers on a daily basis.

Typical Uses

Observers of ICT in schools often contrast uses that seek to make the current system more efficient (e.g., by providing for more practice or feedback in the same amount of time) with those that seek to fundamentally change or transform the nature of teaching and learning (Means *et al.*, 2004). Among the types of applications that most educators would consider enhancements of efficiency are uses of new technologies for transmitting information (such as lectures offered online); skills practice software for reading, mathematics, and second language

learning; and instructional management systems that keep track of the mastery status of every student on a set of skills or content standards. These technology uses can be incorporated into classrooms as they are currently constituted without fundamentally changing the nature of what is taught or interactions between students and teachers. Likewise, many large-scale installations of electronic whiteboards have succeeded only in reproducing existing teaching practice using a new medium. In contrast, technologies that structure and scaffold independent inquiry and design activities on the part of individual students or groups of students and those that are designed to support collaborative knowledge-building activities require changes in the content of learning and in the roles of teachers and students.

While case study research tends to highlight more innovative, transformative uses of technology (some of which will be described below), large-scale surveys provide a more representative picture of what teachers are asking students to do with ICT. In teacher surveys conducted during the first wave of microcomputer use in schools in the mid-1980s, the most commonly reported student uses of technology were as part of computer literacy and programming classes and for drill-and-practice in basic skills (Becker, 1985). Twenty years later, a 2005 national teacher survey in the United States (Bakia *et al.*, 2008) found that teaching technology as a subject was no longer a major emphasis: rather the dominant pattern of school technology use was a combination of general productivity software and continued use for basic skills development. The most frequent uses of technology across grade levels were word processing, Internet research, and reading or mathematics skills practice, with each of these practices occurring once or more a week in the classes of about a third of all teachers. The fourth most common practice – use of software to teach content in a subject other than mathematics or reading – was reported as a weekly occurrence in just 16% of classrooms.

Potential uses of technology within classrooms extend far beyond the typical uses described above. Across a wide range of subject areas, technology has been used to organize and individualize student instruction, provide a motivating, realistic context for learning, give students access to a broad range of resources, and structure learning tasks in ways that reduce the load on memory and stimulate use of analytic strategies and reflection. Some of these uses are described in more detail below.

Subject-Specific Uses of ICT

Reading

Software has been developed to foster emergent literacy in preschool children through word and alphabet games. More broadly, ICT can support creation of an environment that supports young children's literacy skills through the

creation of literacy props such as signs, tickets, and cards that can be used in play-like activities involving literacy skills. In addition, commercial software products aimed at developing phonemic awareness and phonics skills are broadly available. Research on the effectiveness of such products has produced mixed results and suggests that the child's entering skill level, the nature of the non-ICT-based instruction offered, and the phonetic regularity of the child's language may be factors in explaining why some studies find an advantage of software use while others do not.

There is a greater body of supportive research for the use of technology to support struggling readers with the development of greater reading fluency and comprehension. Software can present students with a text to read and provide a digitized spoken presentation of the text synchronized with the child's utterances. When the child has difficulty with a word, these systems typically have mechanisms for helping to analyze the word or presenting it orally so that the child can continue reading. Speech recognition systems are now capable of processing a young reader's spoken word and determining whether or not corrective feedback is in order. Such reading support systems help students acquire needed reading practice, and many students appreciate the patience and privacy of a computer-based system.

Software systems can also support students' vocabulary learning and comprehension through hyperlinks to definitions or illustrations of unfamiliar words or concepts. Some systems provide summaries of main ideas or periodic comprehension checks to help students monitor their level of understanding. Interactive Strategy Training for Active Reading and Thinking (iSTART), for example, is a web-based program that uses animated agents to teach comprehension strategies such as paraphrasing, predicting, and elaboration. Universal Design for Learning (UDL) is an approach in which instructional materials are designed with optional supports so that they can be used effectively by the broadest possible spectrum of learners (Rose and Meyer, 2002), including those with disabilities. These supports include many of the kinds of mechanisms discussed above (text to speech translation, glossaries, prompts for strategic reading strategies) plus an electronic log of the student's work. Research on the application of UDL to learning from science texts has found that having students manipulate electronically presented diagrams of key concepts or having the support of computer agents enhances students' concept learning (Dalton and Strangman, 2006).

Writing

As noted above, the most pervasive use of technology within schools is as a productivity tool, making word processing software the leading educational application. Numerous studies have compared the composition

and editing skills of students with and without access to technology. A 1993 meta-analysis of the early word processing studies (Bangert-Drowns, 1993) found a modest but significant positive effect of word processing on overall writing quality. More recent studies of the effects of giving students a personal laptop have found a positive effect on student writing skills (Kulik, 2003; Penuel, 2006).

Mathematics

Graphing calculators have been the leading technology in mathematics classrooms; most secondary students in the United States and several other countries have access to a graphing calculator. While some critics have warned that the introduction of calculators into mathematics instruction will lead to an inappropriate de-emphasis on calculation skills, mathematics educators recommend that the calculators be used not as a substitute for skill learning but to support students when they are engaged in learning mathematical concepts. Calculators can handle calculations and graph construction, freeing students' cognitive resources to concentrate on solution strategies and understanding concepts. Other software tools supporting students' mathematics problem solving include the Geometer's Sketchpad and Cabri Géomètre, which can be used to support students' exploration and development of geometry proofs, and Tinkerplots and Fathom, which provide tools for organizing, analyzing, and representing data in ways that support understanding of statistics. SimCalc MathWorlds, a research-based tool, similarly provides representations that support students' learning of concepts in Algebra and the transition to Calculus. Advanced calculators incorporate Computer Algebra Systems, which can support students in reasoning about algebra strategically, while allowing the calculator to handle tedious rewriting of symbols. Libraries of virtual manipulatives provide similar capabilities to explore mathematical situations and concepts, but within a narrower focus. Overall, these tools provide support for both calculation and representation of mathematical concepts.

Cognitive tutors represent another advance in the use of ICT in mathematics learning. A cognitive tutor observes the step-by-step process of a students' problem solving and intervenes when the student deviates from an expert solution process. Carnegie Tutors are available for pre-algebra, algebra I and II, geometry, and integrated mathematics. Half a million students in roughly 2600 US schools have used one or more of these tutors.

The commercial market also contains many examples that reflect earlier computer-assisted instruction approaches. These approaches typically present students with tutorials and then offer feedback and hints as students solve problems.

Recent advances in classroom networking are leading to mathematical tools that allow groups of students

to participate in constructing mathematical objects. For example, the commercial TI-Navigator system allows students to simultaneously submit graphs or equations to the teacher and offers the teacher controls to compare and contrast students' work. Research projects have been exploring the potential of collaboration with such specifically mathematical tools.

Second language learning

Technology has played a significant role in many language classrooms since the language labs of the 1960s. While software designed to teach a language and provide practice opportunities remains available and popular in CD-ROM form, the rise of Internet usage within schools has ushered in a new era with a wider range of learning activities in support of second-language learning. In addition to the option of formal language instruction, students can practice and enhance their second language skills by working with original materials such as newspaper articles, blogs, and speeches, developed for the purpose of communication rather than language teaching, available on the Internet in the language they are learning. Opportunities for language classes to interact online with native speakers of the language they are studying have become commonplace. Hybrid language classes, which combine elements of face-to-face instruction and online learning, have become increasingly popular.

Science

Technology tools have become essential to the practice of science, and this change in the nature of the profession has influenced science instruction. Probes and sensors connected to computers, online databases, computer simulations, and models are central to the practice of scientific fields such as physics, biochemistry, and environmental science. Versions with appropriate interfaces and simplifications for education purposes have been developed for many of these, while in other cases students use the same tools as science professionals. In addition to supporting measurement and data storage and analysis, these technology-based tools provide access to phenomena that occur at a size or timescale that would be difficult or impossible for the unaided human to apprehend and, in some cases, involve dangers that make them inappropriate for classroom use. Students can interact with computer-based simulations of viral epidemics, for example, providing a context demonstrating the real-world relevance of science concepts while protecting students from the real world's hazards.

A number of ICT-supported science education programs involve having students interact with scientists working in the field that students are studying. Features such as email, electronic chat, webcasts, and bulletin boards have been used to support student-scientist interactions in programs, including Kids as Global Scientists,

GLOBE, BioKids, Project FeederWatch, and Hands-on Universe. In these programs, students are expected to collect scientific data using scientist-developed protocols and to contribute their findings to a shared database on the World Wide Web. In BioKids, for example, students use personal digital assistants (PDAs) with special software to record the animal life within a study site near their school (Songer, 1996). Because the data are entered into a database shared with other schools, data can be aggregated across schools within the same city, or at a state or regional level. Students can then use BioKids data to address questions about changes in biodiversity over time or in different areas of their state.

Another set of technology-supported tools specific to science learning are systems that build in supports for science inquiry skills. The Knowledge Integration Environment (KIE) and Web-based Inquiry Science Environment (WISE), both developed at the University of California, Berkeley, present students with prompts, reminders, and organizational devices that they use as they explore scientific issues and controversies (Linn, 2006).

Rapidly Growing Uses

Open-source online resources and curriculum materials

One of the biggest transformations wrought by the World Wide Web is the increased availability of the world's cultural, historical, and information resources. The treasures of major museums such as the Louvre and the British Museum are available to anyone with an Internet connection. History students can inspect original documents such as manifests of ships engaging in the slave trade or Leonardo de Vinci's notebooks.

Beyond access to information and artifacts, the Web also offers educational resources organized into units or learning activities and increasingly, entire courses that a school or individual may choose to use. Many educational jurisdictions have organized collections of instructional resources and learning activities keyed to their content standards. The governments of Australia and New Zealand, for example, have collaborated in supporting The Learning Federation, which develops educational content tied to those countries' curricula, and makes it available online to schools at no cost. Curriki is a community web site where educators can post K-12 open-source lesson plans and curriculum materials for review, modification, and use by other teachers. Intended as a global resource, the Curriki site comes in Spanish, French, Hindi, and Bahasa (the native language of Indonesia) as well as English versions. At the tertiary level, the Open Courseware Consortium (OCW) organized by the Massachusetts Institute of Technology (MIT) has made materials from 1800 courses from major universities available online at

no cost. While online collections offer teachers access to impressive resources, teachers often have too little time to search for appropriate resources and too little expertise to assemble a coherent plan of instruction from a library of fragmentary resources.

Online learning and virtual schools

Web-based online learning has been well-established in tertiary education and corporate training since the 1990s. More recently, this use of technology has become a significant part of K-12 education as well. A survey of US school districts commissioned by the Sloan Consortium (Picciano and Seaman, 2007) estimated that 700 000 US public school students were taking online courses in school year 2005–06 – a tripling of K-12 online learners in the 3 years following a National Center for Educational Statistics survey of districts concerning their online course offerings in 2002–03 (Setzer and Louis, 2005). Most of the online courses taken by K-12 public school students at the time of the survey were at the high school level, but the practice now is spreading to lower grades.

In addition to online course offerings through conventional schools, there are also increasing numbers of K-12 school programs that are entirely online. By fall 2007, 28 of the 50 US states had online virtual high school programs (Tucker, 2007). The Florida Virtual School, the first statewide public virtual school in the United States, had 64 000 secondary and middle school students enrolled in online classes in 2008–09.

Online learning has become popular because of its potential for providing more flexible access to content and instruction at any time, from any place. Online learning is often selected with goals such as (1) increasing the availability of learning experiences for learners who cannot or choose not to attend traditional face-to-face offerings, (2) assembling and disseminating instructional content more cost-efficiently, or (3) enabling instructors to handle more students, while maintaining learning outcome quality that is equivalent to that of comparable face-to-face instruction.

Different technology applications are used to support different models of online learning. One class of online learning models uses asynchronous communication tools (e.g., e-mail, threaded discussion boards, and newsgroups) to allow users to contribute at their convenience. Synchronous technologies (e.g., webcasting, chat rooms, and desktop audio/video technology) are used to approximate such face-to-face teaching strategies such as delivering lectures and holding meetings with groups of students. Early online learning applications tended to implement one model or the other, but current online learning applications tend to combine multiple forms of synchronous and asynchronous online, as well as occasional face-to-face, interactions.

Frequent assessment and individualization of learning

Some of the earliest learning software, developed in the 1960s by Patrick Suppes at Stanford University, incorporated the principle of individualizing the material presented to a learner based on that particular learner's state of knowledge. Proponents of mastery learning (an instructional approach organized around specific learning objectives with learning time allowed to vary so that each student works on an objective until he or she has mastered it) quickly appreciated the value of ICT for supporting their instructional approach. While current generations of instructional software are built in new programming languages and have more sophisticated interfaces, much of the mathematics and reading software for elementary and secondary students is a revision of early computer-assisted instruction (CAI) content developed with a mastery learning design. Common software features include reports for teachers showing each student's performance and a summary of the performance of the whole class, organized by curriculum standard or instructional objective.

The increased accountability pressures around mathematics and reading achievement ushered in by the No Child Left Behind Act of 2001 in the United States and similar legislation in other countries has sparked more widespread interest in frequent assessment of students on the content and skill objectives that are part of the accountability system. Even for schools where none of the instruction in reading and mathematics takes place on computer, it has become common for students to take frequent interim or benchmark tests to see where the student is with respect to accountability standards at a point where there is still time to provide extra services or focused instruction on areas of weakness. In many cases, these assessments are provided on computer and scored almost instantaneously. Instructional management software helps teachers and school leaders maintain records and generate reports of interim assessment data. Mislevy and colleagues describe technology's role in supporting the systematic design of assessments and assessment delivery.

Technology is also enabling the development and use of formative assessments. Formative assessments are those that provide feedback to the learner and the instructor that is used to advance learning. Such assessments are most effective when conducted in close proximity to the relevant learning experience and when they have been designed based on research on how people learn the content being assessed. Such assessments reveal both what a student does not understand and why it is not understood. The advantage of technology is in extending the range of situations or tasks that can be presented to the student so that more complex skills and the students' ability to integrate multiple concepts and skills can be assessed. Complex, multi-step problems, data sets, simulations, and models can be incorporated into

technology-based assessments, and not just the student's final answer to the problem but each step taken in addressing the problem can be recorded automatically. Pellegrino describes examples of research-based formative assessments in reading, science, and mathematics. Bennett describes how these same technology capabilities are starting to extend the range of skills addressed by assessments designed for large-scale administration, such as the Keystage 3 ICT Assessment Tasks developed by England's National Assessment Agency and new science assessments being studied for potential incorporation into the National Assessment of Educational Progress (NAEP) in the United States.

Emerging Uses

Technology changes at a much more rapid pace than education, and whenever a new technology emerges, there are those who want to see it applied in schools. Experimentation with mobile phones, podcasts, and social networking sites as educational tools are ongoing. Three major themes in emerging research and development on educational technology are collaborative knowledge building through Web 2.0 applications, immersive environments and games for learning, and interactive classroom communication systems. A common thread among these three themes is increased opportunities for contribution and participation.

Web 2.0

Although without a precise definition, the term Web 2.0 was coined to cover what are considered second-generation Internet sites, distinguished by their dynamic content shaped by multiple users. It includes social networking sites such as MySpace, recommendation systems such as that used by Amazon.com, and community-developed resources such as Wikipedia. Within an education context, Web 2.0 applications engage students in working with others (other students, instructors, or content experts) to explore and build knowledge around concepts or phenomena. The WISE science learning environment described above is an example of a Web 2.0 educational application as is the Knowledge Forum, developed at the Ontario Institute for Studies in Education (Scardamalia and Bereiter, 2006).

The Knowledge Forum software environment is a multimedia database that allows students, teachers, and external mentors to post notes containing information, questions, or interpretations. It was designed to promote knowledge-building discourse – the kind of exchange of information, conjectures, questions, and interpretations one would expect to see in a research laboratory (Scardamalia and Bereiter, 2006). The software provides a set of labels including 'my theory,' 'new information,' 'this theory explains,' and 'this theory does not explain' to scaffold students' efforts to make sense of and build theories about the content of their shared database. The software also

provides ‘views,’ which are alternative organizations of the database contents – for example, categorization of cases on the basis of different features or on the basis of chronology or consistency/inconsistency with a certain theory or school of thought. Knowledge Forum has been used with groups ranging from kindergartners to graduate students and professionals (Scardamalia and Bereiter, 2006) and in a range of countries, including Canada, the United States, Finland, and Hong Kong.

Immersive environments and games

A second technology trend with potential to influence education is the increasing realism and interactivity offered by multi-user virtual environments. While some critics decry the amount of time that young people spend on multi-player online games such as World of Warcraft, others argue that the level of engagement engendered by such games can and should be harnessed for educational purposes (Federation of American Scientists, 2006). A number of so-called serious games have been developed to demonstrate the educational potential of this approach (Barab *et al.*, 2004; Steinkuehler, 2006). River City, for example, is an environment in which students learn middle school science concepts by acting as a character (an avatar) in a simulated eighteenth-century city where the citizens are falling ill at an alarming rate. Students use their knowledge of biology along with the results of tests conducted online with equipment such as virtual microscopes to investigate the mechanisms through which the disease is spreading (air, water, or insects). Students collaborate to write up their research findings as a report to River City’s mayor. River City’s developers believe that the power of such virtual environments lies in their ability to situate student learning in an authentic context while providing scaffolding through system prompts and input from experts that is gradually faded as students acquire stronger inquiry skills.

Interactive classroom communication systems

Another long-standing trend in classroom technology has been increased use of student response systems or clickers. In their simplest form, these devices allow students to simultaneously vote on a multiple-choice question posed by the teacher. The votes can be anonymously collected and displayed as a histogram, indicating the proportion of students that selected each answer. This facility, on its own, has little impact on learning. However, it can support new pedagogical strategies, such as Peer Instruction (Crouch and Mazur, 2001), that increase learning. In Peer Instruction, for example, the instructor uses the feedback to focus student attention on the cognitive contrast between the two most frequent answers to the question. The students are asked to work in pairs to convince each other of the correct answer. This process

of argumentation often results in students’ convergence toward the correct response. In addition, response systems provide feedback to the teacher, who can use this feedback to adjust instruction.

Classroom communications systems now go considerably beyond simple multiple-choice responses. For example, researchers have developed participatory simulations in which all students take the role of an agent in a simulation. Further, using a networked classroom communication system such as the TI-Navigator, students can contribute mathematical expressions, graphs, data sets, and other mathematical objects to a common shared display. In another research prototype, GroupScribbles, students can contribute handwritten notes (like familiar office sticky notes) to a shared group display, enabling many forms of collaborative brainstorming and critique.

Evidence of the Effects of ICT on Teaching and Learning

Parents and teachers report that students’ motivation for learning increases with the use of ICT, and most believe that ICT enhances students’ learning. Rigorous evidence regarding the effects of ICT on objectively measured learning outcomes is harder to come by, however.

Evaluating the effectiveness of technology is complex because there are so many different technologies and even more ways in which technologies can be used. Some have argued for trying to isolate the effects of technology *per se* by providing exactly the same pedagogy and content with and without technology in an experimental design (Clark, 1983). ICT proponents argue that this is a self-defeating strategy, however, because the most important reason to use ICT in instruction is to support kinds of learning that cannot be accomplished without it (making an all-other-things-equal control condition logically impossible).

A recent meta-analysis of controlled studies comparing learning outcomes for online and face-to-face classes, found that students learned more in classes that included online learning activities (Means *et al.*, 2009). The researchers noted, however, that the classes that combined online and face-to-face elements differed from the conventional face-to-face classes not just in terms of the additional instructional medium but also in terms of differing content, instructional strategies, and learning time.

Increasingly, researchers see technology as infrastructural and not itself as an intervention. New computing power, networking, and displays provide the capability to create new instructional activities, but without those new activities, the infrastructure *per se* adds little to teaching and learning. Further, new technology-enabled instructional activities often require additional teacher professional development and new forms of assessment. Hence, a modern perspective emphasizes systems of instruction

which build upon the new infrastructural possibilities offered by ICT.

Systems of instruction include both ICT components and curriculum, teacher training, and other supports. The strongest and best-designed studies have examined such technology-supported systems of instruction in large numbers of classrooms with random assignment of teachers or classes to experimental and control conditions. The effectiveness of the SimCalc MathWorlds software and associated instructional activities and teacher development was evaluated in a randomized controlled trial involving 95 middle school mathematics teachers. Students who used SimCalc attained a better understanding of rate and proportionality than did similar students in control classes.

In contrast, a randomized control trial conducted by the US Department of Education involving 16 different commercial mathematics and reading software applications found no effect of software use on either reading or mathematics learning (Dynarski *et al.*, 2007). A potentially critical feature of the US Department of Education study is that schools and teachers were left to their own devices to determine how they would implement the commercial software; there was no R&D organization helping teachers develop an instructional system to integrate the software with the rest of their instructional program. Moreover, because every school had teachers in both the experimental and the control conditions, a schoolwide program of support for software implementation was not possible.

Effective Implementation of Technology-Supported Learning

A number of large-scale teacher surveys have provided insights into the characteristics of teachers and school environments associated with teacher integration of ICT into instruction. Survey data suggest that teachers who themselves have stronger ICT skills are more likely to have students engage in ICT-supported activities for their classes. The extent to which teachers have their students use ICT is larger also for teachers who have participated in professional development around technology integration and for those whose school leaders promote the use of technology (Zhao *et al.*, 2002). There are also teacher characteristics with no obvious connection to technology use that nevertheless predict the extent of technology integration in a teacher's classroom. Teachers with constructivist or student-centered teaching philosophies have their students use technology more often than do other teachers (O'Dwyer *et al.*, 2005). Teachers who engage in a broader range of professional activities (collaborating with colleagues around instructional issues in their school, participating in professional activities outside the school, and acting as leaders in professional activities) are

significantly more likely than other teachers to have their students use technology weekly or more often (Becker and Riel, 2000). Teacher leaders are particularly likely to have their students use ICT in more sophisticated ways involving knowledge building, problem solving, and collaboration (Riel and Becker, 2008).

Beyond the question of the extent to which ICT is being used in schools and classrooms, there is the broader system-level question of the educational goals its use is addressing. UNESCO (2008) has provided a framework describing different strategies for using ICT to support education reform. The technology literacy strategy aims to harness technology to increase the basic skills of the populace and also to increase students' skill in using ICT, often through taking a specific course taught by a specialized technology teacher. The knowledge deepening strategy seeks to prepare students to use ICT to address complex high-priority problems in the real world, and implementations of this strategy include student use of digital resources and technology tools to address complex problems in a content area (e.g., by exploring a simulation of a watershed in science class). Finally, the knowledge creation strategy involves teaching and using technology in ways that prepare students to become innovators through developing the problem solving, communication, collaboration, and critical thinking skills that are often called twenty-first-century skills.

Law, elsewhere in this encyclopedia, suggests that the skills that teachers need with respect to ICT vary for these different strategies. The technology literacy strategy requires only a few teachers with strong ICT skills to help students acquire technology fluency. The knowledge deepening approach, in contrast, requires all teachers to have not just basic technology competency but also a knowledge of available technology resources in their subject area and an understanding of how these resources can be used to enhance learning in this domain. The knowledge creation strategy requires teachers to be able to move beyond teaching the content conventionally covered in their subject area to develop learning activities that provide the opportunity to learn and refine twenty-first-century skills. Law suggests that execution of this strategy requires a school's staff to be able to use technology to support their communication and collaboration with each other and to manage and analyze student assessment data. Hinostroza, Labbe, and Lopez describe how ICT can be used to support the learning that teachers must do to support such educational reform strategies.

Some researchers have reasoned that acquisition of twenty-first-century skills and the likelihood that these skills will be applied outside of school can be increased through learning experiences that place students in an active role within a collaborative group working with a realistic or situated problem context. By making the context of learning similar to the contexts within which

students will apply their learning (i.e., complex problems that transcend disciplinary boundaries, problem-solving teams with differentiated roles, ICT supports for information finding, reasoning, and communication), educators expect to promote inquiry skills while also making learning activities more motivating and increasing the likelihood that acquired skills and dispositions will transfer to real-world situations.

Use of Technology for Learning Outside of School

As noted above, ICT is becoming increasingly ubiquitous in young peoples' lives and is in fact more common outside of school than inside. Some observers have expressed strong concern about potentially harmful social and physical side effects of excessive technology use, but available empirical research suggests a more balanced view, with some areas of risk and general benefits.

Collins and Halverson suggest that people have many experiences involving searching for and synthesizing information from disparate sources, and using ICT to expand one's cognitive capacity outside of school. Through the Internet, young people engage in communities of individuals with like interests, engage in sophisticated gaming, and acquire new knowledge and skills when they need them (see also Jenkins *et al.*, 2006). Collins and Halverson see the world moving toward a culture of lifelong learning with ICT playing a pivotal role in supporting learning activities, a majority of which will occur outside of brick-and-mortar schools.

It is also possible that in-school and out-of-school learning activities will develop new lines of convergence supported by technology. Signs of this are apparent in the growth in online learning connected to educational institutions. Several universities offering their courses online have reported that the largest single group of online learners consists of regular residential students who prefer the anywhere/anytime quality of the online course to sitting in a lecture hall. In K-12 education, some school districts are offering both classroom-based and online versions of key secondary school courses. In addition to those students opting for the online version from the beginning, there are students taking the classroom-based course who avail themselves of the online learning option for part of the course after becoming ill or needing to travel. Increasingly, students as well as adult learners will expect to be able to access tailored learning resources when and where they want them.

Conclusion

ICT has become increasingly prominent in schools, and many classrooms now have processing power, networking

capabilities, and projected displays that were barely imaginably two decades ago. However, technology is infrastructural and by itself has produced few changes in teaching and learning. Effective use of this new infrastructure depends on systems of instruction that include cognitive tools, pedagogical activities, teacher professional development, and appropriate forms of assessment. Many systems of instruction incorporate the capability of technology to provide increased interactivity and feedback. Further, the most successful systems of instruction have been subject-matter-specific. Reading, mathematics, and science each require specific kinds of software and approaches. The Internet is also making it possible to provide original source content, for example, from museums, directly to the classroom. Emerging approaches to ICT include immersive games, collaborative Web 2.0 applications, and classroom communication systems – all means to increase student participation and collaboration in learning. Still, the evidence that technology improves learning is rather thin. Additional research is needed that tests technology in the context of full instructional systems with adequate implementation support.

See also: Classroom uses of Technology to Manage Instruction; Conceptions of Technology Literacy and Fluency; Internet-based Education; Reading and Technology; Relating Technology, Education Reform and Economic Development; Student Data Systems and Their Use for Educational Improvement; Teacher Skills and Knowledge for Technology Integration; Technological Supports for Acquiring Twenty-First-Century Skills; Technology and Formative Assessment; Technology and Physical and Social Health; Technology for Large-Scale Assessment; Technology Resources for Teacher Learning; Technology Supports for Acquiring Inquiry Skills; Technology Supports for Acquiring Mathematics; Technology Supports for Assessment Design; Technology Supports for Lifelong Learning; Technology Supports for Science Learning; Technology Supports for Second Language Learning.

Bibliography

- Bakia, M., Yang, E., and Mitchell, K. (2008). *National Educational Technology Trends Study: Local-Level Data Summary*. Washington, DC: US Department of Education.
- Barab, S., Thomas, M., Dodge, T., Carteaux, R., and Tuzun, H. (2004). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development* 53(1), 86–108.
- Bangert-Drowns, R. L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research* 63, 69–93.
- Becker, H. J. (1985). How schools use microcomputers: Results from a national survey. In Chen, M. and Paisley, W. (eds.) *Children and Microcomputers: Research on the Newest Medium*, pp 87–107. Beverly Hills, CA: Sage.
- Becker, H. and Riel, M. (2000). *Teacher Professional Engagement and Constructivist-Compatible Computer Use*. Report #7 Teaching, Learning and Computing: 1998: National Survey. Available online at

- the Center for Research on Information Technology and Organizations, University of California, Irvine.
- Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research* 53(4), 445–449.
- Crouch, C. H. and Mazur, E. (2001). Peer instruction: Ten years of experience and results. *Physics Teacher* 69, 970–977.
- Cuban, L., Kirkpatrick, H., and Peck, C. (2001). High access and low use of technologies in high school classrooms: Explaining an apparent paradox. *American Educational Research Journal* 38(4), 813–834.
- Dalton, B. and Strangman, N. (2006). Improving struggling readers' comprehension through scaffolded hypertexts and other computer-based literacy programs. In McKenna, M. C., Labbo, L. D., Kieffer, R. D., and Reinking, D. (eds.) *International Handbook of Literacy and Technology*, vol II, pp 75–92. Mahwah, NJ: Erlbaum.
- Dynarski, M., Agodini, R., Heaviside, S., et al. (2007). Effectiveness of reading and mathematics software products: Findings from the first student cohort. *Report to Congress. NCEE 2007-4006*. Washington, DC: US Department of Education.
- Federation of American Scientists (2006). *Harnessing the Power of Video Games for Learning*. Washington, DC: Federation of American Scientists.
- Healey, J. M. (1999). *Failure to Connect: How Computers Affect Our Children's Minds – and What We Can Do About It*. New York: Simon and Schuster.
- Jenkins, H., Clinton, K., Purushotma, R., Robinson, A. J., and Weigel, M. (2006). *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*. Chicago, IL: The MacArthur Foundation.
- Kulik, J. A. (2003). *Effects of Using Instructional Technology in Elementary and Secondary Schools: What Controlled Evaluation Studies Say*. Arlington, VA: SRI International.
- Law, N., Pelgrum, W. J., and Plomp, T. (2008). *Pedagogy and ICT Use in Schools around the World: Findings from the IEA SITES 2006 Study*. Hong Kong: University of Hong Kong, Comparative Education Research Centre.
- Linn, M. (2006). The knowledge integration perspective on learning and instruction. In Sawyer, R. K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 243–264. Cambridge: Cambridge University Press.
- Means, B., Roschelle, R., Penuel, W., Sabelli, N., and Haertel, G. (2004). Technology's contribution to teaching and policy: Efficiency, standardization, or transformation? In Floden, R. E. (ed.) *Review of Research in Education*, vol. 27, pp 3–56. Washington, DC: American Educational Research Association.
- Means, B., Toyama, Y., Murphy, R., Bakia, M., and Jones, K. (2009). *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online-Learning Studies*. Washington, DC: US Department of Education.
- O'Dwyer, L. M., Russell, M., and Bebell, D. (2005). Identifying teacher, school, and district characteristics associated with middle and high school teachers' use of technology: A multilevel perspective. *Journal of Educational Computing Research* 33(4), 369–393.
- Penuel, W. R. (2006). Implementation and effects of 1:1 computing initiatives: A research synthesis. *Journal of Research on Technology in Education* 38(3), 329–348.
- Picciano, A. G. and Seaman, J. (2007). *K-12 Online Learning: A Survey of U.S. School District Administrators*. Boston, MA: Sloan Consortium.
- Riel, M. and Becker, H. (2008). Characteristics of teacher leaders for information and communication technology. In Voogt, J. and Knezek, G. (eds.) *International Handbook of Information Technology in Primary and Secondary Education*, pp 397–417. New York: Springer.
- Rose, D. and Meyer, A. (2002). *Teaching Every Student in the Digital Age: Universal Design for Learning*. Alexandria, VA: ASCD.
- Scardamalia, M. and Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In Sawyer, R. K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 97–115. Cambridge: Cambridge University Press.
- Setzer, J. C. and Louis, L. (2005). *Distance Education Courses for Public Elementary and Secondary School Students: 2002-03 (NCES 2005-010)*. Washington, DC: National Center for Education Statistics.
- Songer, N. B. (1996). Exploring learning opportunities and coordinated network enhanced classrooms: A case of kids as global scientists. *Journal of the Learning Sciences* 5(4), 297–327.
- Steinkuehler, C. A. (2006). Massively multiplayer online video gaming as participation in a discourse. *Mind, Culture, and Activity* 13(1), 38–52.
- Tucker, B. (2007). *Laboratories of Reform: Virtual High Schools and Innovation in Public Education*. Washington, DC: Education Sector Reports.
- UNESCO (2008). ICT competency standards for teachers: Policy framework. CIO2007/WS/21. <http://portal.unesco.org> (accessed August 2009).
- Wells, J. and Lewis, L. (2006). *Internet Access in U.S. Public Schools and Classrooms: 1994–2005 (NCES 2007-020)*. Washington, DC: National Center for Education Statistics, US Department of Education.
- Zhao, Y., Pugh, K., Sheldon, S., and Byers, J. (2002). Conditions for classroom technology innovations. *Teachers College Record* 104(3), 482–515.

Further Reading

- Becker, H. J., Ravitz, J., and Wong, Y. (1999). Teacher and teacher-directed student use of computers and software. *Technical Report #3: Teaching, Learning, and Computation, 1998 National Survey*. Irvine, CA: University of California at Irvine.
- Condi, R. and Munro, B. (2007). *The Impact of ICT in Schools – A Landscape Review*. Coventry: Becta Research.
- Means, B. (2008). Technology's role in curriculum. In Connelly, F. M., He, M. F., and Phillion, J. (eds.) *Sage Handbook of Curriculum and Instruction*, pp 123–144. Thousand Oaks, CA: Sage.
- Nunnery, J. A., Ross, S. M., and McDonald, A. (2006). A randomized experimental evaluation of the impact of Accelerated Reader/Reading Renaissance implementation on reading achievement in grades 3 to 6. *Journal of Education for Students Placed at Risk* 11(1), 1–18.
- Roschelle, J. M., Pea, R. D., Hoadley, C. M., Gordin, D. N., and Means, B. M. (2000). Changing how and what children learn in school with computer-based technology. *Children and Computer Technology* 10(2), 76–101.
- Smith, M. S. (2009). Opening education. *Science* 323, 89–93.

Educational Data Modeling

S Iksal, LIUM (Computer Sciences Laboratory of the University of Maine), Laval, France

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Educational modeling language – An educational modeling language is a semantic information model and binding, describing the content and process within a unit of learning from a pedagogical perspective for the purpose of support, reuse, and interoperability.

Indicator – A particular observable, because it has a specific meaning in the pedagogical field. It is calculated or simply built with some observed, and reflects the quality of interaction, activity, and learning with the learning environment. If we consider the indicator with the teacher, it defines what is relevant to observe. For the analyst or the engineer, the indicator is an observable, that is to say an observation mean and a set of observables used in the calculus.

Observable and observed – When we are interested in the observation modeling and the observation analysis, each observational data are defined by a specification of the way of getting before being instantiated by the observation. An observable is a variable which has to be calculated through the use's observation of a learning environment. An observed is an observable calculated. A raw data is a particular observable/observed.

Observational means – Observation modeling consists in defining all analysis techniques, automatic or not, that provide an observable. Observation analyzing consists in applying these techniques on tracks in order to get an observed. Hence, these techniques are called observational means.

Pedagogical scenario – The result of the design process of a learning activity. A scenario is composed of learning objectives, learning activities organization, time line, students' tasks description, and the evaluation modalities and their sequencing.

Raw data – By definition, the nature and format of a track are specific from the learning environment. In order to capitalize, share, and reuse the tracks analysis techniques and tools, one of the objectives of modeling is to represent tracks with a format independent from the learning environment used. Moreover, only a part of data is significant for the use analysis. Observation modeling also focuses on the retrieval and the extraction of relevant data from

tracks. Data of this type are called raw data; they are provided by the transformation of the track after the collection, and they have to be represented with a format independent of a learning environment.

Track – Each user of an interactive system may keep some tracks of its use. When the recording of tracks by the system is deliberate, the data structure generated is a sequence of actions which are temporally ordered. Often, tracks are collected inside a log file. Sometimes, they are preprocessed to be integrated inside a database or to be valid to an XML Schema. Some learning environments use videos, polls, or interviews to collect tracks of use. These tracks are all considered as data collected by a learning environment, without considering the ways of collection.

Introduction

Nowadays, numerous universities and enterprises have made the choice of electronic (e)-learning environments to increase individual or specialized learning. Technology-enhanced learning (TEL) has transformed the activities organization in that domain. Teachers are assisted by instructional designers, developers, and sometimes by usage's analysts. All these people need to communicate and to negotiate together during their activities. Therefore, it is necessary to have means and tools to facilitate this dialog. During the last decade, we observed a lot of work around specification, representation, and modeling. In the field of e-learning, these works fall into two categories which seem to merge actually. The first one concerns the learning activity modeling. The idea is to describe the teacher's pedagogical intentions behind the learning activity in order to play that description inside a learning environment. This description has to support the negotiation between learning actors. It is also interesting for the replay of the activity and for its reengineering. This specification is called a scenario. The scenario is one of the two faces of a same coin of educational data modeling. The second one concerns the observation of the learning session. Teachers or instructional designers are interested in a feedback of that new usage: How the environment is used? What exactly is the scenario played by the learner? Have all resources been used by learners? How do learners have appropriate resources? The environments

offer a lot of generic tracking features. Therefore, we now have a huge amount of data that increases all day long. At the same time, we observed that people on the Web want to share and reuse techniques in order to manage these data. Due to the variety of learning environment, all techniques available are specific to an environment and to the learning scenario when it is defined. However, there are two main problems. First, there is a gap between the observational need and the element really observed and analyzed. We observed that usage analysis is often guided by the technical capabilities of the learning environment instead of being teacher oriented. Next, actually, there is no standard for observation specification, everything is built on demand. However, some proposals are made to try a kind of normalization of what have to be observed, why, and how that can be analyzed. In this article, we present what is educational data modeling and how it is used. We conclude by some perspectives of work in that recent domain.

Educational Data Modeling

Modeling becomes very important in e-learning today. We can consider two main areas: first, the scenarization of the learning activity and, next, the scenarization of the observation of the learning activity.

The Learning Scenario

Dealing with the teaching expertise in order to explicit it for embarking it in technology-enhanced systems is one of the main and original aims of the artificial intelligence and education community. Nowadays, the growing interest for the Web technologies applied to distance education has fostered the emergence of standards as IMS Learning Design (LD) and ADL SCORM with, mainly the reusability and the sharing concerns. Koper (Open University of the Netherlands) proposed to describe pedagogical activities with an educational modeling language (EML). The EML model structures pedagogical activities by units of study, that is to say a lesson, a case study, etc. These objects are organized and linked with specific relationships. EML focuses on the concept of activity and considers a unit of study as a set of activities realized by a set of actors in a specific learning environment. These are three main categories of activities, such as learning activities, tutoring activities, and instrumental activities.

Next to this proposal, various languages were developed: PALO, EML, Learning Design Language (LDL), IMS-LD, etc. If we consider LD, which is the most used and also the most criticized, it has been defined as a standard that is pedagogically neutral, which allows different kinds of learning strategies to be supported. Actually, it has the most advanced specifications and several projects of LD editors or LD players are well advanced. The pedagogical

intention as well as the context of use of a scenario written with LD are therefore described in meta-data or diffused in the scenario itself, without any possibility for making them explicit regarding the neutrality of the expressiveness of the model of the language. From the LD specification designer's viewpoint, any pedagogical scenario can be modeled as a method specifying activities to certain actors and in a certain order. A scenario must at least contain a collection of components and a method. Components can be roles (learners, tutor, etc. that can be split in many subgroups), activities, or static scheduling of activities (activity structures), whereas who (which role) does what (which activity) and at which moment is determined by the method (which, strictly speaking, can be considered as the scenario). A method is made up of one or many plays formed by a series of acts. Different plays will represent alternative scenarios, while acts are subsets of the scenario that allow the activity synchronization for the different roles (all actors involved in an act must have finished it before starting the next one). Moreover, the theater metaphor provided by IMS/LD could seem obvious and not natural for some teachers with specific pedagogical expertise and aims in mind. As a consequence, on the one hand, extensions for LD are proposed for better fitting with specific learning activities, and we can also find some authoring tools and design patterns for helping teachers to model their scenarios with this language. On the other hand, we have other languages which are alternatives to LD, better suited for specific purpose as collaborative learning activities for instance, in the case of LDL.

The Observational Scenario

Currently, most of the e-learning systems need some kind of feedback on the usage in order to improve them. In the specific context of distance learning and teaching, the desynchronization between teachers' two major roles – instructional designer and tutor – brings about a lack of uses feedback. The software development process should explicitly integrate a usage analysis phase, which can provide designers with significant information on their systems' uses for a reengineering purpose. Semantic Web aims at facilitating data management on the Web. It brings languages, standards, and corresponding tools that make the sharing and building of automatic and semiautomatic programs easier. Automatic usage analysis is often made by mathematicians or computer engineers. In order to facilitate the appropriation, comprehension, and the interpretation of results by teachers and instructional designers, who are the main actors of an e-learning system development process, some people consider they should be fully integrated in this analysis phase.

This is the reason why educational data modeling also focuses on the observational needs and the tracks' analysis. The teacher has to define his/her observational needs.

He/she has to be able to describe them with questions such as what he/she wants to observe? When (during the session or after the session)? How the result has to be delivered? Also, why it is interesting to observe that specific point? There is no standard actually to specify observational needs. Some people propose to use specific tools and others consider that it is better to use the EML used for the learning scenario. The specification of observational needs is a process of negotiation between the teacher and the analyst and also between the analyst and the developer. In fact, the analyst has to clearly understand the teacher in order to provide the awaited result. The analyst has also to negotiate with the developer in order to know whether it is possible to collect all the data with the learning environment, and if it is not possible, to analyze the cost of the development.

The description of observational needs is necessary for defining prescribed indicators. Therefore, the indicator has a specific meaning in the pedagogical field. It is calculated or simply built with some, and reflects the quality of interaction, activity, and learning with the learning environment. Concerning indicators, some projects of the European Network of Excellence called Kaleidoscope have studied this object and its use (ICALTS, TRAILS, DPULS, IA, etc.). They focused on the observation in the learning life cycle, and proposed definitions, classifications of observational data, and also information on how to support the designer of a TEL system during the modeling phase.

Indicators are described and built from a set of other data. We present now the typology of the DPULS project. The primary data are not calculated or elaborated with the help of other data or knowledge. They could be recorded before, during, or after the learning session using, for instance, a log file recorded by the system, a video tape of the learner recorded during the session, a questionnaire acquired before or after the session, or the sets of posts in a forum. This kind of data is classified as datum. The data qualify data which are linked to the learning situation and could be involved in the usage analysis. Additional data could be – which may be picked from the learning materials, as the meta-data of a learning object – the formal planned scenario for the pedagogical situation, any information which is directly accessible, etc. Predictive data type refers to the outcomes provided by the learning session actors (learners, tutors, and/or teachers). This kind of data is mainly produced by the learners, intended to be assessed, but it also could be, for instance, a tutor's report on the activity of a learner, or on the use of a resource. The data qualify primary data which are *a priori* by an actor of the learning situation (a learner, teacher, or a tutor) or part of the analysis (output by an analyst, designer, or any learning staff member who is involved or concerned with the analysis). The data calculated are inferred from primary data or other data.

Actually, these are various ways to compute indicators, for instance, Tracks-Based System (SBT)) and Usage Tracking Language (UTL).

Conclusion

Educational Data Modeling is a recent domain. There is no real standard, but there are many proposals that have to be evaluated and used by teachers and actors of a learning situation. The development of learning environment has to integrate the concept of learning scenario specified through an external editor and also the concept of observational scenario in order to provide a feedback on the learning session. This feedback not only facilitates the improvement of the quality of the scenario and the learning situation, but it also allows developers to improve their learning environment. Being aware of the need to model education data offers the capabilities of sharing, reusing learning scenario and indicators. The next step concerns the capitalization of these data and their integration inside the learning environments. The main goals behind this work are the qualification of data for information retrieval and the automation of their use.

Further Reading

- CAViCoLA (2006). Computer-based analysis and visualization of collaborative learning activities, *European Network of Excellence Kaleidoscope*. <http://www.noie-kaleidoscope.org> (accessed October 2009).
- Choquet, C. and Iksal, S. (2007). *Special Issue: Usage Analysis in Learning Systems: Existing Approaches and Scientific Issues*. Modeling tracks for the model driven reengineering of a TEL system. *Journal of Interactive Learning Research* **18**(2), 161–184.
- Choquet, C., Iksal, S., Levene, M., and Schoonenboom, J. (2008). Users' data: Trails analysis. In Balacheff, N., Ludvigsen, S., de Jong, T., et al. (eds.) *Technology-Enhanced Learning – Principles and Products*, pp 195–214. New York: Springer.
- IA (2005). Interaction analysis – supporting participants in technology based learning activities, European Network of Excellence Kaleidoscope. <http://www.noie-kaleidoscope.org>.
- ICALTS (2004). Interaction & collaboration analysis' supporting teachers & students' Self-regulation, European Network of Excellence Kaleidoscope. <http://www.noie-kaleidoscope.org> (accessed October 2009).
- Koper, R. (2001). Modeling units of study from a pedagogical perspective – The pedagogical metamodel behind EML. Educational Expertise Technology Centre, Open University of the Netherlands. <http://hdl.handle.net/1820/36> (accessed October 2009).
- LISTEN (2007). The LISTEN project. <http://www.cs.cmu.edu/~listen/> (accessed October 2009).
- Mille, A. (2006). From case-based reasoning to traces-based reasoning. *Annual Reviews in Control* **30**(2), 223–232.
- Rawlings, A., Rosmalen, P. V., Koper, R., Rodríguez-Artacho, M., and Lefrere, P. (2002). *Survey of Educational Modeling Languages (EMLs)*. Technical Report. September 19th 2002.
- Schoonenboom, J., Levene, M., Heller, J., Keenoy, K., and Turcsányi-Szabó, M. (2007). *Trails in Education: Technologies that Support Navigational Learning*. Rotterdam/Taipei: Sense Publishers.
- TRAILS Project (2004). Personalised and collaborative trails of digital and non-digital learning objects, European Network of Excellence Kaleidoscope. <http://www.noie-kaleidoscope.org> (accessed October 2009).

Student Data Systems and Their Use for Educational Improvement

J C Wayman, V Cho, and M P Richards, The University of Texas at Austin, Austin, TX, USA

© 2010 J C Wayman, V Cho, and M P Richards. Published by Elsevier Ltd. All rights reserved.

Glossary

Assessment system – A computer data system that rapidly organizes and analyzes benchmark assessments.

Calibration – A term first used by Wayman *et al.* (2006) to describe a process for establishing an aligned, common vision for education and how data use will support this vision.

Data – Any piece of information that can help educators know more about student learning. Examples include summative achievement tests, interim or benchmark tests, grades, disciplinary information, and teacher judgment.

Data use – The effective examination of educational data for educational improvement.

Data warehousing system – A computer data system that integrates other data systems, providing complete access to historical data of all types.

Student data system – A computer system that enables teachers and other educators to access, aggregate, and examine student data.

Student information system (SIS) – A computer data system that provides automated management of transactional data such as attendance, grade reports, enrollment, and course scheduling.

describe student data systems as an effective use of education dollars under the American Recovery and Reinvestment Act of 2009. Organizations such as the Data Quality Campaign are supportive of data systems, but also are concerned with examining difficulties in implementing these systems.

Technology is a critical component of any educational data use initiative. Historically, however, educational data have been stored in ways that rendered them inaccessible to most educators; as such, use of educational data frequently involved an unrealistic amount of time and effort (Stringfield *et al.*, 2001). Fortunately, the recent advent of more user-friendly computer systems that deliver appropriate data in a timely fashion (Wayman, 2007; Wayman and Cho, 2008; Wayman *et al.*, 2004) has provided a more robust infrastructure for effective data use.

As the quality of data systems available to educators has improved dramatically, so too have the number and type of computer systems being marketed toward educators increased exponentially (Mieles and Foley, 2005; Wayman, 2007; Wayman *et al.*, 2004). Unfortunately, research about the functions and limitations of these different systems is relatively scarce, leaving educators with little information on which to rely in selecting a student data system. Toward that end, we offer a description of the different types of student data systems available to educators in the United States, an overview of the key features of student data systems, and a discussion of the current and future use of student data systems in educational contexts.

Introduction

In the United States, recent federal and state accountability mandates such as the No Child Left Behind Act (NCLB) have increasingly focused the attention of educators on the collection and reporting of student achievement data. Implicit in these policies is the expectation that educators will use these and other data to make decisions that improve educational practice. Largely as a result of this policy shift, phrases such as data use, data-driven decision making, and data-based decision making are increasingly part of the educator's lexicon. Some districts have exceeded the reporting requirements of NCLB and other policies, using student data systems to perform advanced functions such as predicting student achievement (e.g., Datnow *et al.*, 2007).

About the time that this article goes to press (in late 2009), policies from the Obama administration frequently

Types of Student Data Systems

A burgeoning number of student data systems are available to educators, providing a variety of functions designed to efficiently deliver student data. Because of the varied functions that may be performed, no single system provides access to every function a school or district may want. As such, educators must integrate various types of data systems to find one or more that best suit their needs.

Wayman (2005) outlined three common types of data systems that deliver student data to educators: (1) student information systems (SISs), which provide real-time accounting of daily school functions (e.g., attendance and schedules); (2) assessment systems, which rapidly organize and analyze frequent benchmark assessments;

and (3) data warehousing systems, which provide access to historical data of all types and link disparate databases.

In addition to these primary types of data systems, Wayman (2007) identified instructional management systems (some vendors are starting to use this term to generically describe their system, regardless of the functions it performs), which are not designed specifically for student data, but which incorporate student data as part of a broader service. These systems often offer connections between student data and instructional supports such as curriculum resources, learning standards, intra-staff communication, and home-school linkages. Finally, educators have recently begun asking for educational content linked to data, resulting in an increasing number of systems that serve this purpose.

Although commercial vendors are increasingly looking toward offering solutions that integrate the functions of these various types of data systems, such systems are still relatively inchoate. As such, educators must piece together several individual systems to form one comprehensive data solution. Thus, it is critical for educational personnel to understand the various functions and limitations of the various system types. In the following paragraphs, we describe the functions and limitations of the three main systems described by Wayman (2005).

Student information systems. SISs support a vast array of functions that improve the daily work flow of schools and districts. Such systems provide automated management of previously time-intensive and detailed work such as attendance, grade reports, enrolment, and course scheduling. SISs are also increasingly being used to facilitate communication between home and school, providing parents with a portal to access information regarding their children's academic progress.

SISs, however, are rarely designed to facilitate assessment of student learning, support analysis of assessment data, examine longitudinal data beyond the current year, or investigate multiple variables at once.

Assessment systems. Unlike SISs, assessment systems offer powerful tools for facilitating periodic, systemic assessment of student learning. Often used with commercially provided and researched assessments, these systems offer school personnel the ability to administer, grade, and organize assessments, and use their results to illuminate strengths and deficiencies. Assessment systems also support instructional management, often offering assessments aligned and linked to state standards and curriculum supports.

Because of the volume of data generated, assessment systems generally do not provide access to assessment data over time. Further, assessment systems typically fail to link assessment data to other forms of student data such as SIS data, historical student data, demographic information, and other assessments.

Data warehousing systems. Systems that serve to integrate and offer seamless access to disparate forms of data are

loosely termed in education circles as data warehousing systems. (Many such systems are not data warehouses in the strictest sense of the term, instead integrating a fraction of the data that a true data warehouse might.) At the district level, a data warehouse may contain state assessment data, locally administered assessments, student learning histories, teacher information, and demographic information such as absences, free- and reduced-price lunch status, and special education status. Data warehouses are a powerful tool for analyzing a variety of learning phenomena, placing nearly all available student data at the fingertips of the teacher, administrator, or staff member.

Because of the volume of data stored in data warehouses, such systems are not designed to be dynamic; rather, they require periodic uploading or refreshing of the data. As a result, data warehouses are not used to perform the functions of real-time SISs or timely assessment systems. Instead, these functions are performed via other systems and loaded into the data warehouse.

Key Features of Student Data Systems

Although research on student data systems is a relatively young field, an emerging research base permits insight into the key characteristics of student data systems. Wayman *et al.* (2004) surveyed data systems and identified the features central to an effective student data system. They categorized these features into four general factors: (1) user friendliness, (2) user features, (3) information access, and (4) data quality. These factors provide a convenient framework for discussing the extant research on the important features of data systems.

User friendliness. An important aspect of any software, user friendliness is especially crucial in educational settings, where teachers, administrators, and staff have little time or tolerance for difficult or counter-intuitive solutions. Empirical research has demonstrated that absent efficient technology effective data use is both limited and labor intensive (Stringfield *et al.*, 2001; Supovitz and Klein, 2003; Wayman *et al.*, 2007). Moreover, even in the presence of efficient technology, unfriendly interfaces and frustrating response times may constitute a significant barrier to consistent use (Wayman and Stringfield, 2006). Conversely, user-friendly, efficient technology has been demonstrated to facilitate capacity building for independent data use (Chen *et al.*, 2005; Lachat and Smith, 2005; Wayman and Stringfield, 2006).

User features. User features such as query tools, graphic displays, and remote access are important in promoting effective educational data use. Because different educators at different levels of the educational hierarchy access different information for different purposes, studies have demonstrated the importance of powerful query tools that

allow users to extract data appropriate to their role such as teacher, principal, or district superintendent (Chen *et al.*, 2005; Lachat and Smith, 2005; Wayman and Stringfield, 2006). Chen *et al.* (2005) demonstrated the importance of interesting and relevant graphic interfaces, such as histograms, pie charts, scatter plots, and gauges, in facilitating the inquiry of teachers.

Information access. Information access is crucial to the success of any data system. For instance, because many educators work outside school or office hours, it is critical that they have access to data systems when and where they need them (Wayman and Stringfield, 2006). Besides system availability, information access also implies effective methods of presenting information and the availability of a broad range of relevant data. Contrasting examples of these were shown by Brunner *et al.* (2005). In studying data reports available to educators in New York City, Brunner *et al.* (2005) found that a student data system that produced reports in a comprehensive, easy-to-understand fashion was highly useful to both educators and parents. On the other hand, these reports were not populated with a broad range of dynamic data. Ultimately, their utility became short-lived as user advancement necessitated deeper inquiry than was supported by the reports.

Data quality. A data system is only as good as the data it contains. As such, high-quality, accurate data, and the ability to maintain an accurate, quality database is paramount to any data system. Wayman and Conoly (2006) underscored this point, demonstrating that a system with rich, clean data could be rapidly implemented in a large district. The study also described postimplementation sustainability issues faced by the district regarding the expansion and maintenance of their data. Inputting accurate data is an important consideration, and is often found to be more difficult than initially anticipated (McLeod, 2005; Wayman *et al.*, 2004).

The Use of Data Systems in Educational Contexts

While data systems are a necessary component of effective data use, they are not sufficient. To be successful, a data initiative must also address the context in which the data system is implemented. Prior to the last 3 or 4 years, electronic data systems were not frequently employed in schools, thus limiting research on data use in technology-rich educational contexts (e.g., Chrispeels *et al.*, 2000; Massell, 2001; Supovitz and Klein, 2003). As data systems have become more common, however, a growing corpus of research has examined the use of data systems in educational contexts. This work offers a number of interesting insights and reveals a number of contextual factors that are critical to the effective use of educational data systems.

Calibration and alignment with district vision. An effective data initiative involves the alignment of all district organizational levels and personnel around a systemic vision for data use that addresses how data will be used to support teaching and learning (Datnow *et al.*, 2007; Halverson *et al.*, 2005; Mandinach *et al.*, 2008; Wayman *et al.*, 2007). To ensure that student data systems will be effective in supporting district goals, districts should purchase or construct such systems only after a district-wide vision for data use has been outlined (Wayman and Cho, 2008).

To establish an aligned, common vision for education and how data use will support this vision, educators should engage in a formal process of calibration (Wayman *et al.*, 2006, 2007), which involves a productive dialog around questions such as: What do we mean by teaching and learning? How will teaching be conducted under these definitions? How will we assess student learning so we know it when we see it? How will we react to results? To ensure richer discussions, Wayman *et al.* (2007) recommend that districts conduct a thorough, district-wide evaluation of available data, existing data practices, and support capacities, prior to commencing the calibration process.

Principal leadership. Principals are often cited as key players in the data process and they play an integral role in the health of any data initiative (Copland, 2003; Lachat and Smith, 2005; Wayman and Stringfield, 2006; Young, 2006). For example, Wayman and Stringfield (2006) found that principals were the single most important factor contributing to the success of the schools in their study. Principals of successful schools were heavily invested in the use of data systems at their schools and were often avid users of data systems. In addition to modeling with their own data use, principals demonstrated a variety of innovative methods of support for these systems such as requiring data reports at all conferences, granting time and development to allow teachers to hone their skills at using the system, and meshing different types of systems to meet specific situational needs.

Conversely, research suggests that ineffective data initiatives are often characterized by poor principal leadership (Wayman *et al.*, 2007; Wayman *et al.*, 2009; Young, 2006). For example, Wayman *et al.* (2009) described schools where ineffective use strategies muted the extent to which a perfectly effective data system was implemented. The authors surmised that ineffective principal preparation was a leading cause of these inefficiencies.

Faculty involvement. While principals are the key catalysts for data use, teachers are its key users. Wayman (2005) has suggested that full faculty involvement is critical to the ultimate success of a data initiative with computer systems. However, faculty involvement is often predicated not only on principal leadership but on the fit between the data system and teachers' work as educators (Wayman *et al.*, 2007, 2009). Teachers have been

shown to be enthusiastic about the use of a data system when it meets their needs (Brunner *et al.*, 2005; Chen *et al.*, 2005; Lachat and Smith, 2005; Wayman and Stringfield, 2006), but dismissive when they perceive it to be cumbersome or not useful (Brunner *et al.*, 2005; Wayman and Stringfield, 2006). Thus, while educators are often skeptical of data use and data systems because of concerns that such initiatives will add to their already full workdays (Ingram *et al.*, 2004; Valli and Buese, 2007; Wayman *et al.*, 2009), such skepticism may be mitigated by ensuring that student data systems are integrated with the work of teachers.

Educator collaboration. Researchers often tout the advantages of collaboration in promoting inquiry (Massell, 2001; Murphy, 2002; Young, 2006). In practice, however, fostering collaboration among faculty can be difficult. Collaborating around a data system, whether formally or informally, can often help remedy the situation, bringing focus and a common language to collaborative efforts (Wayman *et al.*, 2006). For example, Wayman and Stringfield (2006) observed formally structured collaboration involving data systems, while Chen *et al.* (2005) and Lachat and Smith (2005) described informal collaboration opportunities that arose around data systems. Further, the collaboration–data relationship is reciprocal: While data systems can support collaboration, collaboration can also ensure the success of data initiatives (Chen *et al.*, 2005; Lachat and Smith, 2005; Wayman, 2005; Wayman *et al.*, 2006; Wayman and Stringfield, 2006).

Other institutional supports. A variety of other institutional supports influence the implementation of data systems, including educator training and development, system access, and time for data use.

Although data systems are powerful tools, their success, like that of any tool, is contingent upon adequate user training, knowledge, and support. Thus, it is important to ensure that educators are data literate and proficient in using their data systems in everyday practice via training that is immediately, specifically, and practically relevant to each educator. Although most commercial vendors offer some sort of training as part of their agreement, districts should provide deeper and ongoing training on the use of the system.

Full access to student data, wherein educators are given personal login information that allows them 24-h access to data and system functions they are legally allowed to access, has been shown to enhance and broaden teacher data use (Chen *et al.*, 2005; Wayman and Stringfield, 2006). Moreover, lack of technological access has been cited as a hindrance to effective data use (Halverson *et al.*, 2005; Supovitz and Klein, 2003). However, some leaders are wary of providing access to data for all educators (Young, 2006), often concentrating access to data on certain ‘go-to’ individuals rather than encouraging educators to access data directly (Lachat and Smith, 2005). Wayman and Cho

(2008) have cautioned against over-reliance on such ‘go-to’ individuals, arguing that data systems are most effective when individual educators can and do access data directly.

Another key support for effective use of data systems is structuring educators’ schedules to allow them time to examine data (e.g., Ingram *et al.*, 2004; Lachat and Smith, 2005; Wayman, 2005; Wayman and Stringfield, 2006). While some researchers have cited the effectiveness of providing large portions of time in periodic intervals (Massell, 2001; Young, 2006), others have emphasized the importance of providing time more frequently. Wayman and Cho (2008) have argued that educators should be permitted and encouraged to use the data system on a daily basis, thus conveying that data use is valued, expected, and supported.

Future Directions

As evidenced by the previous discussion, data systems hold great potential for promoting educator inquiry and fostering educational improvement. Accordingly, most extant research paints a rosy picture of the implementation of these systems. However, these findings must generally be interpreted with caution, as most research has been conducted on schools and districts singled out for their exemplary use of such systems. Thus, although knowledge of data systems and their use has increased dramatically over the past years, additional research on nonexemplary schools is necessary to strengthen the research base and determine whether research findings generalize beyond these exemplary contexts (e.g., see Gallagher *et al.*, 2008; Wayman *et al.*, 2007). Further, the rarity of such exemplary schools suggests that student data systems are still far from ubiquitous facilitators of educational improvement. In the following, we discuss five themes shaping the future of educational data systems: acquisition, interoperability, modularity, triangulation, and future technological advancements.

Acquisition. One key issue involves the acquisition of student data systems. In acquiring a student data system, districts must generally choose whether to buy a commercially available system or build their own. Wayman *et al.* (2004) suggested that most districts are better off buying a commercial system than developing their own, noting the difficulty of building a data system. The authors suggested that commercial vendors bring economy of scale, and that many districts lack the time and local talent to build an efficient, first-rate system. Further, although some districts may view commercial data systems as expensive luxuries, Wayman *et al.* (2004) suggested that districts likely will recupereate these monetary costs in terms of efficiency, time, and regained teaching opportunities.

Regardless of the decision to build or buy, districts should take care to adopt systems specifically tailored to

their needs. To this end, districts can be forward looking by engaging in calibration exercises and other formal needs assessments (Wayman *et al.*, 2007). This enhances equally the request for proposal (RFP) for districts buying systems, as well the system design process for districts developing their own.

Interoperability. Another central issue is data system interoperability. As noted previously, no system provides every function a district would want. As such, educational personnel must piece together independent systems to address all their data needs. Exchange of data among disparate systems, however, is often problematic owing to proprietary specifications and restrictions. Districts may partially mitigate this problem by using a commercially available product, as many vendors have partnerships with other vendors providing different services and have thereby ironed out many data exchange problems. However, districts are at the mercy of whatever partnerships their member has forged.

The Schools Interoperability Framework (SIF) is an example of an organization working to solve these compatibility issues. The SIF is a partnership of vendors, educators, researchers, and other stakeholders that has issued definitions and specifications to govern data exchange. Vendors may elect to become SIF-certified by agreeing to exchange data following these protocols and having their systems pass a complex (and expensive) vetting process. Districts and states are increasingly including SIF-certification as a requirement in their RFPs for data system acquisition.

Modularity. Although there is not yet a single universal student data system, many commercially available and locally developed systems are large and serve many functions. Unfortunately, employing such a system may in practice not be ideal for most districts. Some districts are already hamstrung by their current system because buying data systems is like buying a house – once bought, the owner must stay in it for a while to recoup their investment. Moreover, the bigger the investment, the more it costs to change. Ideally, however, schools and districts could buy or build different modules that could plug and play with other modules. These modules would serve a variety of functions currently available in larger systems. For instance, a district or school might require three or four modules that together would perform the functions of a single student data system. Thus, buying data systems would be more like buying a blender than a house – once bought, the owner can easily and affordably upgrade or discard the module. A promising vision in this regard is offered by Midgley (2006), in advocating for open-source solutions in education.

Triangulation. Student data systems and the work of education (e.g., teaching) are continually evolving together. Technological advancements have the potential to improve teacher practice and decision making, thus opening the door to new policies for supporting teaching and learning.

At the same time, changes in policy and pedagogy change the landscape of demands placed upon data systems. At the crux of this co-evolution of teaching and technology is the issue of how educators triangulate using multiple forms of data.

Currently, much of the attention given to data by educators, commercial vendors, and policymakers has focused on standards-based accountability and its assessments. This is only one layer of data. In fact, the most effective practices involve educators individually and collectively triangulating with multiple forms of data (Datnow *et al.*, 2007; Wayman *et al.*, 2006; Wayman and Stringfield, 2006). Besides accountability-style assessments, other important data elements may include benchmark or interim assessments, classroom assessments, and professional judgment. Given better access to a broader range of appropriate data, educators report increased capacity for understanding and attending to the ‘whole student’ (Lachat and Smith, 2005; Supovitz and Klein, 2003; Wayman and Stringfield, 2006).

Future technological advancements. There are a number of new applications of technology already in place outside of education that we believe hold promise for more effective data systems. First, advancements in technology and assessment are leading to more advanced ways to interpret student learning. In many ways, these technology-mediated assessments break the mold of standardized testing, offering more creative, dynamic approaches to measuring students’ abilities. For instance, Silva (2008) describes one assessment where students are asked to develop remedies for traffic congestion caused by population growth and improvise solutions in light of research reports, budgets, and other documents. Also, Tucker (2009) describes one assessment where students are able to plan and conduct experiments with virtual helium balloons, as well as another assessment where students determine the appropriate trajectory and speed for rescuing an injured skier. These examples are different from current assessments in that: (1) both assessments go beyond looking for right or wrong answers, and (2) technology allows a more thorough, complex evaluation of student skills in ways such as applying algorithms that assess students’ sequences of decisions and actions. Whereas some educators may have previously found it difficult to weave standardized data into practice (Ingram *et al.*, 2004), assessments such as the above suggest that the data provided to educators may soon be more complex and tied to real-world skills.

Second, handheld mobile devices have become almost ubiquitous outside of schools. Outside of education, technology is mobile and comes to the user; within education, educators still go to the technology (e.g., accessing a data system through the PC in their office). Although not yet common, companies are currently marketing applications that allow teachers and principals to access data through

their handhelds while walking around the classroom, moving about the school, and interacting with parents, students, or other educators. In schools, such technologies could change how students might interact with content and each other, as well as how teachers might deliver interventions or process data about their students (Schuler, 2009).

Third, intelligent or push technologies could offer automatic recommendations based upon the data at hand and use profiles. For instance, when a user executes a particular query, the system might respond with a message noting that users who performed this query also performed another type of query, or that users who performed this query accessed a particular resource. These technologies are becoming increasingly capable at tapping into practice-based, tacit knowledge or professional intuition (Alavi and Leidner, 2001; Stenmark, 2000/2001), so it is not far-fetched to imagine data systems one day offering teachers automatic support derived from patterns in student learning, attendance, and other focus areas. That support may range from recommendations about assisting particular students to connecting educators to professional development or other colleagues.

Fourth, technology that helps connect professionals to resources and to each other is important to organizational performance (Alavi and Leidner, 2001; Malone *et al.*, 1987). Technologies can help capture and distribute expertise, and we believe this capacity holds promise for data systems. For instance, video and other multimedia are one way to effectively share best practices, often in ways that are more rich in content and craft knowledge than traditional print media (Thomas *et al.*, 2001). Similarly, social networking sites and wiki-style resources have become a popular way to connect and share information. One could imagine these applied to instructional resources and lessons, thus further pooling district knowledge and distributing that expertise collegially. Concerns about the quality of these resources may be allayed through peer commentary – the capacity to review, rate, and comment about items online has been found to effectively affect other users' attitudes and promote shared norms (Forman *et al.*, 2008).

Conclusion

Data systems provide a critical tool for leveraging student data toward educational improvement. Recent technological advancements have fueled the development of data systems facilitating the organization, access, and querying of large volumes of student data. Without the proper technology, even the most capable and interested educators may find it challenging to use data to facilitate student learning. Such technology, while necessary, is not sufficient to ensure effective data use. Rather, data use initiatives must adapt to the complex context in which

data systems are embedded, through alignment with district vision and via the provision of key institutional supports. Although research regarding the use and utility of student data systems has increased substantially over the past years, additional research is required to better understand how data systems may be used to facilitate improvement. Further, effective and efficient systems are still far from ubiquitous. It is argued that the future of student data systems will be contingent upon the ability of districts to acquire interoperable and modular data systems whose capacity can expand in accordance with technological advances.

Bibliography

- Alavi, M. and Leidner, D. E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly* **25**(1), 107–136.
- Brunner, C., Fasca, C., Heinze, J., *et al.* (2005). Linking data and learning: The Grow Network study. *Journal of Education for Students Placed at Risk* **10**(3), 241–267.
- Chen, E., Heritage, M., and Lee, J. (2005). Identifying and monitoring students' learning needs with technology. *Journal of Education for Students Placed at Risk* **10**(3), 309–332.
- Chrispeels, J. H., Brown, J. H., and Castillo, S. (2000). School leadership teams: Factors that influence their development and effectiveness. *Advances in Research and Theories of School Management and Educational Policy* **4**, 39–73.
- Copland, M. A. (2003). Leadership of inquiry: Building and sustaining capacity for school improvement. *Educational Evaluation and Policy Analysis* **25**, 375–395.
- Datnow, A., Park, V., and Wohlstetter, P. (2007). *Achieving with Data: How High-Performing School Systems Use Data to Improve Instruction for Elementary Students*. Los Angeles: University of Southern California, Rossier School of Education, Center on Educational Governance.
- Forman, C., Ghose, A., and Wiesenfeld, B. (2008). Examining the relationship between review and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research* **19**(3), 291–313.
- Gallagher, L., Means, B., and Padilla, C. (2008). *Teachers' Use of Data Systems to Improve Instruction*. Washington, DC: US Department of Education.
- Halverson, R., Grigg, G., Pritchett, R., and Thomas, C. (2005). *The New Instructional Leadership: Creating Data-Driven Instructional Systems in Schools*. http://www.wcer.wisc.edu/publications/workingPapers/Working_Paper_No_2005_9.pdf (accessed October 2009).
- Ingram, D., Louis, K. S., and Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record* **106**, 1258–1287.
- Lachat, M. A. and Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed at Risk* **10**(3), 333–349.
- Malone, T. M., Yates, J., and Benjamin, R. I. (1987). Electronic markets and electronic hierarchies: Effects of information technologies on market structure and corporate strategies. *Communications of the ACM* **30**(6), 484–497.
- Mandinach, E. B., Honey, M., Light, D., and Brunner, C. (2008). A conceptual framework for data-driven decision making. In Mandinach, E. B. and Honey, M. (eds.) *Data-Driven School Improvement*, pp 13–31. New York: Teachers College Press.
- Massell, D. (2001). The theory and practice of using data to build capacity: State and local strategies and their effects. In Fuhrman, S. H. (ed.) *From the Capitol to the Classroom: Standards-Based Reform in the States*, pp 148–169. Chicago, IL: University of Chicago Press.

- McLeod, S. (2005). Technology tools for data-driven teachers. <http://www.microsoft.com/education/ThoughtLeadersDDDM.msp> (accessed October 2009).
- Midgley, S. (2006). The case for open markets in education. http://www.firstmonday.org/issues/issue11_7/midgley/index.html (accessed October 2009).
- Mieles, T. and Foley, E. (2005). *Data Warehousing: Preliminary Findings from a Study of Implementing Districts*. Philadelphia, PA: Annenberg Institute for School Reform.
- Murphy, J. (2002). Reculturing the profession of educational leadership: New blueprints. *Educational Administration Quarterly* **38**(2), 176–191.
- Schuler, C. (2009). *Pockets of Potential: Using Mobile Technologies to Promote Children's Learning*. New York: The Joan Ganz Cooney Center at Sesame Workshop.
- Silva, E. (2008). *Measuring Skills for the 21st Century*. Education Sector Reports. Washington, DC: Education Sector.
- Stenmark, D. (2000/2001). Leveraging tacit organizational knowledge. *Journal of Management Information Systems* **17**(3), 9–24.
- Stringfield, S., Reynolds, D., and Schaffer, E. (2001). Fifth-year results from the High Reliability Schools project. *Symposium Presented at the Meeting of the International Congress for School Effectiveness and Improvement*. Toronto, Canada.
- Supovitz, J. A. and Klein, V. (2003). *Mapping a Course for Improved Student Learning: How Innovative Schools Systematically Use Student Performance Data to Guide Improvement*. Philadelphia, PA: Consortium for Policy Research in Education.
- Thomas, J. B., Sussman, S. W., and Henderson, J. C. (2001). Understanding "strategic learning": Linking organizational learning, knowledge management, and sensemaking. *Organization Science* **12**(3), 331–345.
- Tucker, B. (2009). *Beyond the Bubble: Technology and the Future of Student Assessment*. New York: Education Sector.
- Valli, L. and Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal* **44**(3), 519–558.
- Wayman, J. C. (2005). Involving teachers in data-driven decision-making: Using computer data systems to support teacher inquiry and reflection. *Journal of Education for Students Placed at Risk* **10**(3), 295–308.
- Wayman, J. C. (2007). Student data systems for school improvement: The state of the field. In *TCEA Educational Technology Research Symposium: Vol. 1*, pp 156–162. Lancaster, PA: ProActive.
- Wayman, J. C. and Cho, V. (2008). Preparing educators to effectively use student data systems. In Kowalski, T. J. and Lasley, T. J. (eds.) *Handbook on Data-Based Decision-Making in Education*, pp 89–104. New York: Routledge.
- Wayman, J. C., Cho, V., and Johnston, M. T. (2007). *The Data-Informed District. A District-Wide Evaluation of Data Use in the Natrona County School District*. Austin: The University of Texas.
- Wayman, J. C., Cho, V., and Shaw, S. M. (2009). *First-Year Results from an Efficacy Study of the Acuity Data System*. Austin: The University of Texas.
- Wayman, J. C. and Conoly, K. (2006). Managing curriculum: Rapid implementation and sustainability of a districtwide data initiative. *ERS Spectrum* **24**(2), 4–8.
- Wayman, J. C. and Stringfield, S. (2006). Technology-supported involvement of entire faculties in examination of student data for instructional improvement. *American Journal of Education* **112**(4), 549–571.
- Wayman, J. C., Midgley, S., and Stringfield, S. (2006). Leadership for data-based decisionmaking: Collaborative data teams. In Danzig, A., Borman, K., Jones, B., and Wright, B. (eds.) *New Models of Professional Development for Learner Centered Leadership*, pp 189–206. Mahwah, NJ: Erlbaum.
- Wayman, J. C., Stringfield, S., and Yakimowski, M. (2004). *Software Enabling School Improvement Through Analysis of Student Data (CRESPAR Technical Report No. 67)*. Baltimore, MD: Johns Hopkins University.
- Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education* **112**(4), 521–548.

Relevant Websites

- <http://www.dataqualitycampaign.org> – Data Quality Campaign.
- <http://edadmin.edb.utexas.edu/datause/> – Jeffrey C. Wayman's data use website.
- <http://www.sifinfo.org> – SIF Association.

Technology and Learning: Access in Schools Around the World

R Anderson, University of Minnesota, Wayzata, MN, USA

J Ainley, Australian Council for Educational Research, Camberwell, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Access to ICT – Is the availability of ICT or “technology” for use generally within a school or home setting.

ICT – See Information and communication technology.

IEA (International Association for the Evaluation of Educational Achievement) – Is an international organization that coordinates comparative assessments and other research studies.

Information and communication technology (ICT) – Refers to any technology used for access to information or to communicate information. In particular, it encompasses multi-media generation and consumption. ICT is sometimes called “technology” or “educational technology” in the context of education.

PISA (Programme for International Student Assessment) – Is an international assessment coordinated by the OECD, the Organization for Economic Cooperation and Development.

SITES (Second International Technology in Education Study) – Was an international survey of teachers and schools coordinated by the IEA.

TIMSS (Trends in International Mathematics and Science Study) – Is an international assessment program for math and science and is coordinated by the IEA.

Use of ICT – Consists of some occasional or more frequent “hands-on” activity with a hardware or software technology application.

19 and 26 countries, respectively. In 2006, the Organization for Economic Cooperation and Development Programme for International Student Assessment (OECD PISA) included questions on technology for samples of 15-year-olds in 57 countries. These data, combined with results from two other recent international comparative studies, provide a profile of student access to technology for learning worldwide.

Student access to ICT in learning environments became a political issue in the later part of the twentieth century. Many educational policy makers came to believe that ICT augmented and enhanced learning, and investments in ICT infrastructure spread around the globe (Plomp *et al.*, 2009). This encouraged educational researchers to seek evidence regarding the instructional effectiveness of different types of technological applications to learning (cf. Cuban, 2003; Kozma, 2003; Law *et al.*, 2008). Meanwhile most education decision makers supported the acquisition of ICT because of the intuitively compelling argument that ICT provides a means of receiving and creating new knowledge (Anderson, 2008).

Conceptual Aids

Several theoretical traditions aid understanding of the spread of ICT in schools. The oldest relevant theory is the innovation adoption theory championed by Everett Rogers over 50 years ago (Rogers and Rogers, 2003). This popular theory views technology innovation as a sociological process involving successive stages of susceptibility influenced by precipitating factors. Another important theory of adaptation to technology is the technology acceptance model (TAM) that depends upon individuals' perception of its utility and ease of use (Bagozzi, 2007). In organizational and cross-cultural contexts, theories that focus upon resources and organizational decision makers are needed. For example, resource dependence theory assumes that the principal concern of organizations is to control environmental resources (Pfeffer, 1982). Anderson and Magnan (1996) applied the theory to the adoption of school networking, arguing that adoption would occur when a school perceived pressures from stakeholder groups to obtain networked resources.

Two other conceptual approaches offer insight by suggesting how resources and economics are moderated by social processes. Warschauer (2003) detailed how unique

Introduction

Recent international surveys enable us to compare a large number of countries in terms of students, access to computers and the Internet. Not only does this article report these statistics on information and communication technology (increasingly called ICT for short) access, but it also discusses their implications. In 1992 and 1998, international surveys of computer access in schools were conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA) in

cultural contexts, especially institutional embeddedness, shape how education responds to new digital technology. Institutional embeddedness encompasses the many ways that cultural contexts, social norms, and organizational processes play a part in determining acquisition of ICT, adequacy of training and support, and effective use of ICT in learning (Warschauer, 2003). Using data from 82 countries, Drori and Jang (2003) found that national sophistication in networking was a function more of international cooperation and scientific growth than of economic development. Lack of well-developed institutional links with other countries account for instances where countries had low penetration of technology despite having the resources to acquire it.

One of the difficulties of research on technology access within education is the confusion over seemingly simple terms like access, availability, use, familiarity, and proficiency. In an attempt to be clear, we define technology access as equivalent to the availability of technology within a school or home setting. There are times when a distinction between access and availability is necessary, such as if one is studying how schools may keep certain students from using technology. Technology use consists of some minimal level of hands-on implementation of a hardware or software technology. No international study has yet conducted an assessment of ICT proficiency, but in this report, we infer that a student or teacher has some minimal level of skill or proficiency if he or she responds in the affirmative to a question about ability to perform a specific ICT-based task, for example, pending an e-mail. It is helpful to keep in mind that educational systems in some countries, especially in Eastern Europe and Asia, have long traditions of giving priority to having students learn to use ICT (acquire technology skills) rather than to using ICT to learn other subjects and skills (such as mathematics or inquiry skills).

Data Sources

Large-scale international studies assessing student competence in various areas of study provide a unique opportunity to gain information about students' access to educational resources. The findings from PISA, Trends in International Mathematics and Science Study (TIMSS), and Second International Technology in Education Study (SITES; SITES, 2006) provide an indication of the extent to which participating students from around the world have access to ICT and the ways in which they and their teachers use technology. These studies provide internationally comparable data based on rigorously designed and implemented samples of well-defined populations.

The PISA is an assessment of reading, mathematics, and science literacy among 15-year-old students, which

was developed by the OECD (OECD, 2007). Fifty-seven countries participated in PISA 2006, which encompassed an in-depth assessment of science and a less detailed assessment of reading and mathematics literacy (OECD, 2007). As part of the PISA 2006 survey, schools in all countries provided information about the computer resources available and 40 countries administered a set of questions to students about their familiarity with and confidence in using ICT (OECD, 2007).

The TIMSS is conducted every 4 years by the International Association for the Evaluation of Educational Achievement (IEA); it examines students' achievement in mathematics and science at grades 4 and 8 (Mullis *et al.*, 2008; Martin *et al.*, 2008). Fifty-nine countries participated in the 2006/7 cycle of TIMSS at either or both of grades 4 and 8: 37 at grade 4 and 50 at grade 8. Teachers provided information about the availability and use of ICT in their teaching, and students provided an indication of their access to, and use of, computers at home and at school.

The SITES was initiated by the IEA to investigate the role of ICT in education. Module 1 (SITES-M1) was a school survey of principals and technology coordinators in 26 countries (Pelgrum and Anderson, 1999). Module 2 (SITES-M2) explored the relationship between ICT and teaching through a set of 174 case studies of innovative pedagogical practices using ICT from 28 countries (Kozma, 2003). SITES 2006 was a survey of grade-8 mathematics and science teachers in 22 countries, as well as the principals and technology coordinators in their schools (Australia completed the survey as a benchmarking country in 2007). It investigated the use of ICT in mathematics and science teaching in schools, the factors associated with the use of ICT, and the relationship of ICT to pedagogical practices.

Access to ICT

ICT in Schools

The PISA 2006 school questionnaire asked principals to provide an indication of both the overall number of computers in the school and the number of computers in the school available for instructional use (OECD, 2007). On the basis of this information, the number of computers available per student can be calculated. **Table 1** contains two ratios, one for the total school computers per student and the other for total school computers available for instruction. (For each ratio, the country mean and the associated 95% confidence interval is shown. The precision of an estimate derived from a sample survey can be represented as a confidence interval (typically the 95% confidence interval) that indicates the range around the population parameter within which a survey estimate would fall on 95% occasions. The confidence interval is 1.96 times the standard error.) The countries are sorted

Table 1 Number (ratio) of computers per student by country, PISA 2006^a

	<i>Ratio of computers per student</i>					
	<i>Total computers</i>		<i>Instructional computers^b</i>		<i>Internet computers^c</i>	
	<i>Mean^d</i>	<i>Confidence interval^e</i>	<i>Mean^d</i>	<i>Confidence interval^e</i>	<i>Mean^d</i>	<i>Confidence interval^e</i>
Japan	0.38	0.05	0.29	0.04	0.33	0.05
Liechtenstein	0.37	0.09	0.34	0.08	0.35	0.10
United States	0.35	0.03	0.28	0.03	0.32	0.03
United Kingdom	0.35	0.01	0.28	0.01	0.33	0.01
Norway	0.34	0.02	0.25	0.02	0.31	0.02
Australia	0.34	0.02	0.27	0.02	0.32	0.02
Sweden	0.33	0.04	0.27	0.04	0.31	0.04
Austria	0.30	0.04	0.24	0.03	0.27	0.04
Korea	0.29	0.04	0.18	0.03	0.28	0.04
New Zealand	0.28	0.02	0.20	0.01	0.27	0.02
Canada	0.28	0.01	0.22	0.01	0.26	0.01
Iceland	0.26	0.02	0.18	0.02	0.25	0.02
Luxembourg	0.25	0.07	0.23	0.06	0.25	0.07
Hong Kong-China	0.25	0.01	0.20	0.01	0.24	0.01
Switzerland	0.24	0.02	0.21	0.02	0.20	0.02
Slovenia	0.23	0.03	0.17	0.02	0.21	0.03
Chinese Taipei	0.23	0.02	0.17	0.02	0.22	0.02
Denmark	0.22	0.01	0.19	0.01	0.20	0.01
Colombia	0.21	0.06	0.14	0.04	0.08	0.03
Belgium	0.21	0.02	0.17	0.02	0.18	0.02
Finland	0.20	0.02	0.16	0.01	0.18	0.02
Hungary	0.19	0.03	0.16	0.03	0.16	0.03
Italy	0.19	0.01	0.16	0.01	0.16	0.01
Netherlands	0.19	0.01	0.15	0.01	0.17	0.01
Macao-China	0.18	0.02	0.15	0.02	0.17	0.02
Czech Republic	0.15	0.02	0.12	0.02	0.13	0.02
Estonia	0.15	0.01	0.11	0.01	0.14	0.01
Qatar	0.15	0.04	0.12	0.03	0.06	0.02
Poland	0.15	0.02	0.12	0.02	0.13	0.02
Ireland	0.14	0.01	0.11	0.01	0.12	0.01
Mexico	0.14	0.01	0.11	0.01	0.07	0.01
Spain	0.13	0.01	0.11	0.01	0.12	0.01
Israel	0.13	0.01	0.11	0.01	0.09	0.02
Germany	0.13	0.01	0.11	0.01	0.11	0.01
Greece	0.13	0.02	0.10	0.01	0.11	0.01
Portugal	0.11	0.01	0.08	0.01	0.10	0.01
Latvia	0.11	0.01	0.08	0.01	0.08	0.01
Lithuania	0.10	0.01	0.07	0.01	0.08	0.01
Slovak Republic	0.09	0.01	0.08	0.01	0.07	0.01
Uruguay	0.08	0.01	0.07	0.01	0.05	0.01
Croatia	0.08	0.01	0.06	0.01	0.06	0.01
Turkey	0.08	0.01	0.06	0.01	0.06	0.01
Bulgaria	0.07	0.01	0.06	0.00	0.04	0.01
Romania	0.07	0.01	0.06	0.01	0.02	0.00
Thailand	0.06	0.01	0.05	0.01	0.04	0.01
Montenegro	0.06	0.03	0.04	0.02	0.02	0.02
Jordan	0.06	0.00	0.05	0.00	0.02	0.00
Chile	0.06	0.01	0.04	0.00	0.04	0.00
Russian Federation	0.05	0.01	0.04	0.01	0.01	0.00
Serbia	0.05	0.01	0.04	0.00	0.02	0.00
Argentina	0.04	0.01	0.03	0.01	0.02	0.01
Indonesia	0.03	0.01	0.03	0.00	0.00	0.00
Tunisia	0.02	0.00	0.02	0.00	0.01	0.00
Brazil	0.02	0.00	0.02	0.00	0.01	0.00
Kyrgyzstan	0.02	0.00	0.01	0.00	0.00	0.00
Azerbaijan	0.01	0.00	0.00	0.00	0.00	0.00

Continued

Table 1 Continued

	<i>Total computers</i>		<i>Ratio of computers per student</i>			
			<i>Instructional computers^b</i>		<i>Internet computers^c</i>	
	<i>Mean^d</i>	<i>Confidence interval^e</i>	<i>Mean^d</i>	<i>Confidence interval^e</i>	<i>Mean^d</i>	<i>Confidence interval^e</i>
Average (all)	0.17		0.13		0.14	
Average (OECD countries)	0.21		0.17		0.20	
Average (non-OECD countries)	0.09		0.07		0.08	

^aStatistics computed from PISA database. From OECD (2007).

^bInstructional computers are computers in the school available for instructional use.

^cInternet computers are all computers in the school connected to the Internet or World Wide Web.

^dMeans are average number of computers per student averaged across schools within each country.

^eConfidence intervals indicate the range around the population parameter within which a survey estimate would fall on 95% occasions. The confidence interval is 1.96 times the standard error.

into descending order within the table by each country's total computer ratio, and then the countries are grouped into quartiles.

The rankings (and quartiles) of countries by total and instructional computer ratios are nearly identical. The principal exceptions are Korea and Iceland in that they rank considerably higher on total computers than instructional computers because such a high proportion of their school computers are used for administration. Whereas the average percent of computers used for administration across all countries was 24%, in Korea and Iceland, it was 38% and 30% respectively. These two countries are in the upper quartile for total number of computers, but in the second quartile for instructional computers. They exchange places with Switzerland and Denmark. Otherwise, the quartile assignments remain essentially the same for total computer ratio and instructional computer ratio.

On average across all the participating countries, the total number of computers per student was 0.17, indicating that there are approximately six students to each computer. However, there is a large difference between the countries with the lowest number of students per computer and those with the highest. In **Table 1** the countries have been grouped into four equal-sized groups.

Group 1 contains the countries with the highest level of computer provision in schools. These countries, which have computer to student ratios from 0.25 to 0.38, include Japan, the United States, the United Kingdom, Norway, Australia, Sweden, Austria, Korea, New Zealand, Canada, Iceland, Luxembourg, and Hong Kong. In broad terms, these countries can be thought of as having three or four students per computer.

Group 2 includes countries with the next highest level of computer provision in schools with computer to student ratios from 0.15 to 0.24. It includes countries such as Switzerland, Slovenia, Chinese Taipei, Denmark, Colombia, Belgium, Finland, Hungary, Italy, Netherlands,

Macao-China, Czech Republic, Estonia, and Qatar. In general terms these countries have typically five students per computer.

Group 3 embraces countries with slightly less abundant provision of computer resources in schools. It includes countries such as Poland, Ireland, Mexico, Spain, Israel, Germany, Greece, Portugal, Latvia, Lithuania, Slovak Republic, Uruguay, Croatia, and Turkey. The computer student ratios in these countries range from 0.08 to 0.15 which equates to more than seven students per computer.

Group 4 includes countries with the least abundant provision of computer resources in schools with computer to student ratios of 0.07 or less and on average being equivalent to more than 20 students per computer. This group includes countries such as Bulgaria, Romania, Thailand, Montenegro, Jordan, Chile, Russia, Serbia, Argentina, Indonesia, Tunisia, Brazil, Kyrgyzstan, and Azerbaijan.

Countries also differ according to the percentage of computers that are connected to the Internet, as shown in **Table 1**. Countries such as Colombia, Qatar, and Mexico have high to moderate computer ratios but relatively low ratios of Internet-connected computers. Internet technology evolved several decades after computer technology so we might expect the gap in Internet-connected computers to be greater. Taking the mean Internet computers per student as a percent of the mean computers per student, we find that in 16 countries more than 90% of school computers are connected to the Internet. In an additional 17 countries, between 80 and 90% of computers are connected to the Internet. There are 12 countries where fewer than half the school computers are connected to the Internet.

Perhaps what is most striking in these distributions in **Table 1** is the difference between computers and Internet computers in the bottom quartile of countries. In these 14 developing countries, most have five or more computers per 100 students. However, most of these countries have only 2 or even fewer Internet computers per 100 students.

ICT at Home

The findings of TIMSS 2006/7 also provide an indication of the opportunity students in grade 4 and grade 8 have to access ICT at home. Among grade 4 students, approximately 70% had access to a home computer and 56% had Internet access. For grade 8 students, the figures were almost the same: 70% and 50%. However, access to ICT resources at home varies considerably across the surveyed nations. **Table 2** provides an indication of the percentage of grade 8 students with an Internet connection at home.

There were 12 countries in which more than 80% of grade 8 students reported having an Internet connection at home: Hong Kong SAR, Norway, Sweden, Korea, England, Scotland, Australia, Chinese Taipei, Singapore, the United States, Slovenia, and Israel. There were a further 10 countries in which more than half (but fewer than 80%) of the grade 8 students had an Internet connection at home. In the remaining 26 countries, fewer than half the students had an Internet connection at home.

Using ICT

ICT Use at School

PISA 2006 provides data on the frequency with which 15-year-old students use ICT at school. The data in **Table 3** indicate that ICT use is not routine in the sense of being used every day, even in countries where there is a high level of availability. There are countries that stand out as having a significantly higher percentage of students reporting the daily use of ICT at school than in any other country: New Zealand, Australia, Denmark, Canada, Austria, Norway, and the Netherlands. These countries all have 16% or more students (i.e., one in six) reporting daily ICT use at school.

There is an association between the percentage of students reporting the daily use of ICT and the availability of ICT as reflected in the mean computer to student ratio for each country. The between-country correlation coefficient is 0.37 and would be 0.45 if Japan is not included (Japan has a high mean computer to student ratio but a low percentage of students reporting daily use of ICT). This association suggests that availability is associated with ICT use at school but that the association is far from perfect.

ICT Use at Home

As shown in **Table 3**, the daily use of ICT by 15-year-old students is much higher at home than at school. In every country, the percentage of students using ICT at home every day is greater than the percentage using it at school by a large margin. On average, less than one in ten

Table 2 Percent of grade 8 students with Internet connection at home, TIMSS 2007^a

	Percent	Confidence interval ^b
Hong Kong SAR	97	1
Norway	97	1
Sweden	97	1
Korea, Republic of	96	1
England	92	1
Scotland	92	1
Australia	89	1
Chinese Taipei	89	1
Singapore	87	1
United States	87	1
Slovenia	86	1
Israel	84	2
Japan	77	2
Czech Republic	76	2
Bahrain	74	2
Qatar	74	1
Kuwait	71	1
Italy	70	2
Lithuania	66	2
Cyprus	65	2
Hungary	62	3
Bulgaria	57	3
Serbia	47	3
Saudi Arabia	41	3
Morocco	37	3
Lebanon	36	3
Oman	35	3
Romania	33	3
Russian Federation	32	3
Bosnia and Herzegovina	31	3
Palestinian, Natl Auth.	31	2
Malaysia	27	3
Egypt	25	2
Iran, Islamic Rep.	25	3
Jordan	24	2
Ukraine	22	2
Thailand	20	3
Turkey	20	2
Syrian Arab Rep.	19	2
Tunisia	18	2
Armenia	17	2
Algeria	15	2
Colombia	15	3
Georgia	14	2
Botswana	13	1
El Salvador	10	2
Ghana	10	1
Indonesia	8	2

^aPercentages obtained from TIMSS 2007 Report. From Mullis *et al.* (2008).

^bEstimation of confidence intervals takes account of clustering in, and other characteristics of, the sample.

students use ICT everyday at school compared to two-thirds of students using ICT every day at home. There is no evidence in these data that school use compensates for lack of home use of ICT. There is a small positive

Table 3 Frequency of computer use at school and home, PISA 2006^a

	<i>Percentage of 15-year-old students who:</i>							
	<i>Use a computer at school</i>				<i>Use a computer at home</i>			
	<i>Almost every day</i>		<i>Once or twice a week</i>		<i>Almost every day</i>		<i>Once or twice a week</i>	
	%	CI ^c	%	CI	%	CI	%	CI
Australia	24	2	49	2	77	1	17	1
Austria	17	2	55	3	69	2	21	2
Belgium	5	1	50	2	80	1	13	1
Bulgaria	10	2	68	3	67	3	5	1
Canada	18	1	29	1	83	1	11	1
Chile	5	1	33	4	51	3	11	1
Colombia	6	1	72	6	29	3	10	1
Croatia	3	1	65	3	71	2	16	1
Czech Republic	10	2	59	3	72	2	13	1
Denmark	21	3	44	2	84	1	11	1
Finland	3	1	48	3	82	1	12	1
Germany	2	1	29	3	74	2	16	1
Greece	5	1	53	3	53	2	19	1
Hungary	10	1	75	2	67	2	17	1
Iceland	8	1	46	2	90	1	7	1
Ireland	8	1	40	3	48	2	29	2
Italy	6	1	44	3	64	1	20	1
Japan	3	1	46	4	25	1	27	1
Jordan	7	1	76	2	56	2	13	1
Korea	4	1	32	5	60	2	33	2
Latvia	8	2	72	3	65	2	11	1
Liechtenstein	5	^b	55	5	83	4	10	3
Lithuania	5	1	60	3	75	2	7	1
Macao-China	2	0	83	1	74	2	17	1
Netherlands	16	2	49	3	91	1	6	1
New Zealand	27	2	23	2	61	2	25	1
Norway	17	3	37	2	89	1	7	1
Poland	3	1	58	5	72	2	9	1
Portugal	10	1	50	2	75	2	12	1
Qatar	10	1	58	1	67	1	19	1
Russian Federation	5	2	68	4	55	3	13	2
Serbia	4	1	81	2	67	2	10	1
Slovak Republic	6	1	59	3	62	2	15	1
Slovenia	3	1	62	2	80	1	14	1
Spain	3	1	39	3	70	2	16	1
Sweden	10	2	37	3	85	1	11	1
Switzerland	6	1	37	2	77	1	16	1
Thailand	12	1	68	3	31	2	10	1
Turkey	9	2	44	3	39	3	14	1
Uruguay	6	1	33	3	46	2	14	1
Average	9		52		67		14	

^aStatistics computed from PISA database. From OECD (2007).^bNumbers in cell too small to provide a reliable estimate.^cEstimation of confidence intervals takes account of clustering in, and other characteristics of, the sample.

association between the percentages of students using ICT on a daily basis at school and at home (the correlation coefficient is 0.21).

Patterns of ICT Use

PISA 2006 provides data on the frequency with which 15-year-old students use various ICT functions. These functions embrace two groups of activities (a structure

supported by an analysis of the relationships among the responses) (OECD, 2007: 342).

The first group of functions is referred to as Internet/entertainment use and is probably close to what is referred to in some literature as social networking. It includes the following items:

- browsing the Internet for information about people, things, or ideas;

- playing games;
- using the Internet to collaborate with a group or team;
- downloading software from the Internet (including games);
- downloading music from the Internet; and
- communicating (e-mail or chat rooms).

The second group of functions is referred to as program/software use and includes the following items:

- writing documents (e.g., with Word or Word Perfect);
- using spreadsheets (e.g., Lotus 123 or MS Excel);
- drawing, painting, or using graphics programs;
- using educational software such as mathematics programs; and
- writing computer programs.

Figure 1 and Table 4 provides an indication of the percentages of 15-year-old students using each of these functions at least once or twice each week. The data are represented as the first quartile, median, and third quartile of the distribution of country percentages. From these data it is evident that the most frequently used functions are concerned with communication (e-mail or chat rooms) or browsing the Internet for information followed by downloading music and playing games.

Of the program/software functions only writing documents and using graphics programs were identified by substantial numbers of students as being used at least weekly.

Learning and ICT

There are two strands to the application of ICT in school education. The first concerns developing the ICT proficiency of students so that they can participate fully in a modern society where ICT plays such an important role in our work and social lives. The second strand refers to the use of ICT to facilitate new approaches to teaching and learning in schools and to more closely align what students do in the classrooms with what they are required to do in real life. In this sense ICT facilitates access to a wider range of information sources, enriches the possibilities of transforming information, and enhances the possibilities for communication and collaboration (Anderson, 2008).

Learning to Use ICT

Findings from PISA 2006 provide an indication of students' perceptions of their ability to carry out various ICT tasks and their confidence in performing these tasks. Students participating in PISA 2006 were asked to indicate, for a number of ICT tasks, how well they could perform that task on a scale from 'I can do this very well by myself,' through to 'I can do this with help from someone' and 'I know what this means but cannot do it' to 'I don't know what this means'. A factor analysis (using an oblique

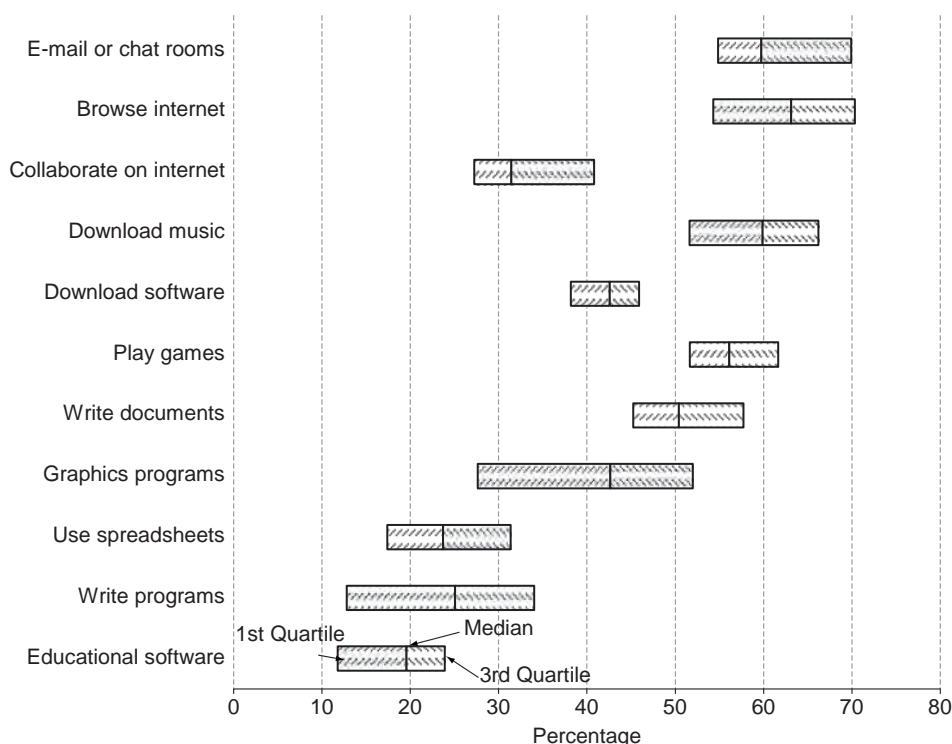


Figure 1 Percentage of 15-year-olds using computer functions at least weekly: (a) Statistics computed from the PISA database. From OECD (2007). (b) the left-hand side of the bar represents the first quartile and the right-hand side represents the third quartile. The line in the middle of the bar is the median.

Table 4 Frequency of computer use at school and home, PISA 2006^a

	<i>Chat</i>	<i>Virus</i>	<i>Edit photos</i>	<i>Data base</i>	<i>Data to CD</i>	<i>Move Files</i>	<i>Search Internet</i>	<i>Down- load files</i>	<i>Attach to e-mail</i>	<i>Word processing</i>	<i>Spread sheets</i>	<i>Presenta- tions</i>	<i>Download music</i>	<i>Multi- media</i>	<i>E-mails</i>	<i>Web Page</i>
Australia	90	48	64	28	83	89	96	86	86	92	60	86	84	57	93	35
Austria	90	50	73	34	83	90	95	79	77	90	61	76	75	53	90	26
Belgium	94	51	70	31	82	90	96	83	87	84	39	64	85	55	93	38
Bulgaria	81	46	52	38	73	77	81	72	65	65	40	38	77	36	75	15
Canada	96	54	65	28	82	88	97	90	90	93	52	73	91	50	96	44
Chile	81	38	65	36	70	74	89	72	64	70	45	66	72	42	75	23
Colombia	56	22	35	31	52	61	75	62	38	47	37	58	51	33	51	27
Croatia	77	36	49	27	81	82	87	67	53	79	60	48	69	40	72	23
Czech Republic	90	54	65	26	80	87	94	83	78	83	64	61	79	48	90	31
Denmark	91	40	59	12	75	87	95	75	84	90	54	69	74	45	94	30
Finland	92	51	60	17	74	83	97	69	82	89	37	44	71	35	96	25
Germany	89	52	70	29	83	84	96	73	71	85	54	57	70	48	88	32
Greece	58	31	40	29	64	69	73	62	42	63	42	43	66	46	53	19
Hungary	74	47	54	28	77	83	87	76	68	80	47	47	74	35	83	26
Iceland	96	49	65	18	78	83	95	70	78	90	42	65	79	36	93	69
Ireland	57	30	45	25	70	74	89	70	47	81	42	47	68	33	70	21
Italy	57	30	48	18	74	78	76	66	48	83	46	55	65	44	58	18
Japan	44	11	27	10	34	42	85	46	51	43	27	23	43	16	63	21
Jordan	49	28	40	48	50	62	51	48	34	57	58	56	44	44	36	33

Korea	93	54	57	10	46	89	98	96	96	89	22	46	96	35	98	20
Latvia	91	37	59	24	75	83	92	75	78	80	60	66	75	38	90	29
Liechtenstein	94	56	75	37	81	92	96	85	93	90	63	81	83	56	96	34
Lithuania	83	42	56	23	82	80	90	67	58	76	63	43	78	31	82	20
Macao-China	80	36	40	23	61	88	89	84	69	57	31	59	86	53	77	29
Netherlands	95	50	70	26	78	88	97	90	90	90	40	68	92	49	97	41
New Zealand	83	42	58	25	79	85	94	80	78	88	59	74	78	44	89	25
Norway	95	57	62	21	85	91	97	88	85	91	59	71	90	53	94	57
Poland	80	48	50	34	74	84	86	72	63	82	69	58	66	44	70	29
Portugal	74	44	68	53	82	85	92	67	67	83	65	77	67	57	81	30
Qatar	69	37	49	39	58	66	70	66	61	57	48	56	75	59	66	38
Russian Fed.	38	33	41	30	63	69	47	43	31	65	44	42	46	35	38	21
Serbia	56	33	48	32	77	78	65	57	37	67	57	44	61	34	45	18
Slovak Republic	65	30	42	19	64	69	86	70	49	69	47	40	61	31	71	26
Slovenia	87	52	58	23	81	89	92	84	74	83	65	63	79	44	83	24
Spain	91	46	56	32	84	89	92	78	62	76	54	59	80	41	84	27
Sweden	96	46	63	14	78	87	97	79	84	84	43	53	79	52	95	24
Switzerland	89	48	68	27	79	87	95	79	80	86	47	56	78	44	91	27
Thailand	34	17	17	18	37	44	56	37	27	36	26	39	36	18	42	15
Turkey	68	27	43	28	61	72	73	60	44	56	36	40	65	45	53	30
Uruguay	79	25	49	31	59	73	85	72	66	69	49	65	66	44	80	22
Average (all)	78	41	54	27	71	79	86	72	66	76	49	57	72	43	77	29

^aData source is the PISA 2006 database (OECD, 2007). <http://pisa2006.acer.edu.au/>

rotation) of the student response data suggested that most of the tasks formed two groups (and could be used to generate two scales). This reflects a pattern in which students who report being able to do one of the things in a group independently are also likely to be able to do the others in the group.

The first group concerned ICT Internet tasks and comprised the following tasks (listed with the average percent correct):

- searching the Internet for information (86%);
- chatting online (78%);
- writing and sending emails (77%);
- downloading files or programs from the Internet (72%);
- downloading music from the Internet (72%); and
- attaching a file to an email message (66%).

The second group concerned ICT product development tasks and comprised the following tasks (listed with the average percent correct):

- using a word processor (e.g. to write an essay for school) (76%);
- creating a presentation (e.g. using MS PowerPoint) (57%);
- editing digital photographs or other graphic images (54%);
- using a spreadsheet to plot a graph (49%);
- creating a multimedia presentation (with sound, pictures and video) (43%);
- using software to find and get rid of computer viruses (41%);
- constructing a web page (29%); and
- creating a database (e.g. using MS Access) (27%).

In addition, there were two items that did not fit either of these groups: moving files (79%) and copying data to a CD (71%).

Using item response theory (IRT) analyses of student responses to these questions, two scales were constructed by the PISA group: confidence in ICT high-level tasks and confidence in ICT Internet tasks. **Table 5** records the mean scores and associated confidence intervals on these scales for the participating countries.

The mean scores in **Table 5** indicate a considerable difference between countries in terms of student's confidence in using ICT. The range in the country means for confidence in ICT product development tasks is 1.2 logits and in the country means for confidence in ICT Internet tasks, the range is 1.9 logits. The highest means for confidence in ICT product development tasks were recorded for Liechtenstein (but with small numbers of students), Portugal, Australia, Norway, Austria, and Canada. The lowest means on this scale were recorded for Thailand and Japan. The highest means for confidence in ICT Internet tasks were recorded for Korea, the Netherlands, and Canada, with the lowest means on this scale being recorded for Thailand, Jordan, and the Russian Federation.

Further research is needed to investigate the relation between student confidence and measured performance and to probe the antecedents of differences in ICT proficiency.

Using ICT in Teaching Mathematics and Science

SITES 2006 surveyed grade-8 mathematics and science teachers in 23 countries to find out whether they had used ICT in any type of teaching activity during that school year. **Table 6** shows the percentages of mathematics and science teachers who reported having used ICT with their classes. The lowest usage levels were reported by mathematics teachers (18%) and science teachers (16%) in South Africa. At the other end of the spectrum, very high percentages (more than 80%) of science teachers in Australia, Singapore, Hong Kong, and Alberta, and of mathematics teachers in Norway reported using ICT in their grade 8 teaching.

In 14 of the 23 countries, the percentage of teachers reporting ICT-use was significantly higher for science teachers than for mathematics teachers within the same country. In the remaining countries, the difference in ICT use between science and mathematics teachers was not statistically significant. In general, the higher the percentage of science teachers in a country using ICT, the higher the percentage of mathematics teachers using ICT in their grade 8 teaching. Furthermore, among both science and mathematics teachers, and within each country, use of ICT is higher when the teachers themselves are confident users of ICT. This suggests that an important element in the use of ICT is building the expertise of teachers. At present, the two activities where ICT was reported to be used by the highest proportions of teachers were looking up ideas and information and short-task projects. The activity in which ICT was reported to be used by the smallest proportions of both science and mathematics teachers was field study activities. The potential of ICT as a tool for data collection, management, and interpretation is not yet being realized on a wide scale.

Conclusion

A glance at the cross-national comparisons contained in our tables leaves us with the impression that national economic wealth determines school ICT infrastructure and use of technology for learning. While there is some truth to that conclusion, it fails to take into account several critical mediating factors, especially institutional embeddedness (Warschauer, 2003) and complexity of linkages within world networks (Drori and Jang, 2003). Without these social conditions in place, an ICT infrastructure may not be built, and if it is built, it may not be used effectively or even used at all.

Table 5 Student confidence in performing ICT product development tasks and ICT internet tasks, PISA 2006^a

	<i>Confidence in ICT product development tasks</i>			<i>Confidence in ICT internet tasks</i>	
	<i>Scale score^b</i>	<i>Confidence interval</i>		<i>Scale score</i>	<i>Confidence interval</i>
Liechtenstein	0.42	0.10	Korea	0.59	0.02
Portugal	0.41	0.04	Netherlands	0.49	0.03
Australia	0.30	0.02	Canada	0.47	0.02
Norway	0.29	0.03	Norway	0.41	0.03
Austria	0.28	0.04	Liechtenstein	0.38	0.08
Canada	0.27	0.03	Australia	0.33	0.02
Netherlands	0.20	0.03	Belgium	0.32	0.03
Germany	0.18	0.03	Sweden	0.27	0.03
Belgium	0.17	0.03	Czech Republic	0.20	0.04
Czech Republic	0.17	0.05	Switzerland	0.16	0.03
Poland	0.13	0.04	Denmark	0.15	0.03
Slovenia	0.12	0.02	Slovenia	0.14	0.02
New Zealand	0.08	0.03	New Zealand	0.14	0.03
Iceland	0.08	0.03	Iceland	0.13	0.03
Switzerland	0.08	0.03	Finland	0.13	0.03
Chile	0.03	0.05	Austria	0.13	0.03
Spain	0.02	0.03	Latvia	0.13	0.04
Denmark	0.02	0.03	Macao-China	0.02	0.03
Latvia	0.01	0.04	Spain	0.02	0.03
Hungary	-0.05	0.04	Germany	0.01	0.03
Sweden	-0.06	0.04	Hungary	-0.09	0.05
Croatia	-0.06	0.04	Lithuania	-0.13	0.05
Qatar	-0.10	0.03	Chile	-0.13	0.09
Finland	-0.11	0.03	Bulgaria	-0.14	0.07
Lithuania	-0.14	0.04	Portugal	-0.16	0.05
Jordan	-0.16	0.05	Poland	-0.18	0.05
Uruguay	-0.19	0.05	Uruguay	-0.21	0.05
Macao-China	-0.20	0.03	Croatia	-0.30	0.04
Serbia	-0.22	0.05	Slovak Republic	-0.44	0.05
Bulgaria	-0.24	0.05	Qatar	-0.49	0.03
Korea	-0.24	0.04	Ireland	-0.50	0.06
Turkey	-0.27	0.05	Turkey	-0.60	0.07
Greece	-0.28	0.04	Italy	-0.63	0.03
Italy	-0.29	0.03	Greece	-0.67	0.05
Colombia	-0.29	0.06	Colombia	-0.72	0.09
Ireland	-0.33	0.04	Japan	-0.82	0.04
Russian Federation	-0.34	0.06	Serbia	-0.85	0.06
Slovak Republic	-0.36	0.05	Thailand	-1.17	0.06
Thailand	-0.63	0.05	Jordan	-1.23	0.07
Japan	-0.80	0.04	Russian Federation	-1.30	0.09
Average	-0.05		Average	-0.15	

^aStatistics computed from PISA database. From OECD (2007).^bScale scores are in logits.

Lack of well-developed institutional links with other countries account for instances where countries, for example, Columbia, had relatively low use of computers in schools, even though the schools had a lot of them. Simply putting ICT into schools does not assure that it will be used. Schools also need commitment from external communities and need to invest in support services and teacher training. Japan is another case in point; they had a lot of school computers, but relatively few students who were able to perform a variety of Internet and ICT tasks. In contrast,

Portugal had few computers but their students were relatively skilled in their ability to use the Internet and ICT.

Merely installing and using ICT in the classroom does not necessarily improve learning. Wenglinsky (1998) correlated computer use in schools with math and science achievement and found that when use included any kind of ICT use, then achievement was actually lower. However, achievement was higher when use was limited to specific applications like advanced simulations. This supports Warschauer's (2003) argument that the contextual meaning

Table 6 Percentages of teachers of science and mathematics at Year 8 who report using ICT in their teaching, 2006^a

	<i>Science</i>			<i>Mathematics</i>	
	<i>Percent</i>	<i>Confidence interval^b</i>		<i>Percent</i>	<i>Confidence interval</i>
Australia	86	3	Norway	80	5
Singapore	84	3	Denmark	77	4
Hong Kong, SAR	82	3	Ontario	75	5
Alberta	79	4	Singapore	73	4
Ontario	75	4	Hong Kong, SAR	70	4
Norway	74	5	Australia	69	5
Denmark	70	5	Lithuania	62	5
Slovenia	68	4	Alberta	62	5
Chile	66	5	Italy	58	4
Lithuania	66	5	Chile	56	4
Finland	61	5	Slovak Republic	51	4
Italy	58	4	France	49	5
Thailand	57	5	Finland	48	5
Moscow	57	4	Moscow	45	5
Slovak Republic	56	4	Thailand	44	5
Spain, Catalonia	56	4	Russian Federation	41	7
France	54	5	Estonia	40	7
Estonia	54	5	Slovenia	40	4
Israel	53	4	Spain (Catalonia)	38	4
Russian Federation	49	5	Chinese Taipei	35	3
Chinese Taipei	48	4	Japan	23	4
Japan	44	5	Israel	22	3
South Africa	16	4	South Africa	18	4

^aData are computed from SITES 2006 database (SITES 2008).

^bEstimation of confidence intervals takes account of clustering in, and other characteristics of the sample.

^cComputation by Louise Wenn is gratefully acknowledged.

of the learning activity plays the critical role in determining whether ICT use in learning will be successful.

These conclusions do not negate concerns for digital divides in learning. Rather, they clarify how the possession of the technology itself does not produce social gaps; how the technology is defined and integrated with social and institutional elements in schools widens or narrows the gaps between advantaged and disadvantaged groups. Van Dijk (2005) argues that the digital gaps are increasingly becoming more problematic as the knowledge-based economies continue to predominate. Servon (2002) examines attempts to bridge the digital divide in the United States and concludes that access clearly is an incomplete approach. She advocates a community-based approach in which community technology centers partner with schools and other community-building organizations.

While this report encompasses well over 50 countries, we must keep in mind that the world has upwards of 200 more countries and many of them cannot afford to put electricity and books into schools, much less ICT. Most of the poorest nations were not included in these studies, so it is not possible to systematically estimate how many students in the world lack technology access. However, given that roughly half the world's population lives on less than \$2 (US) per day, it is probably safe to speculate that less than half of the world's primary and secondary age

children attend a school in which they can access either a computer or the Internet. In recent years, a global digital divide movement has emerged out of many peoples' concern over digital inequity.

One final conclusion to draw from this analysis is that very little is known about the nature of ICT in education worldwide. The statistics are sketchy and the indicators to assessments of other subject added like after thoughts. SITES 2006 was designed to study learning pedagogy and ICT, but teachers were surveyed in only a handful of countries. Presumably, as more and more learning moves online, the need for raising the priority of conducting solid research on ICT access and learning at all levels will be recognized.

See *also*: Equity in Technology Access and Opportunities; Relating Technology, Education Reform and Economic Development; Technology as a Support for School-Community Connections.

Bibliography

Anderson, R. E. (2008). Implications of the information and knowledge society for education. In Voogt, J. and Knezek, G. (eds.) *International Handbook of Information Technology in Primary and Secondary Education*, pp 5–20. New York: Springer.

- Anderson, R. E. and Magnan, S. (1996). The adoption of external networking by American schools. *American Sociologist* 27(1), 55–64.
- Bagozzi, R. P. (2007). The legacy of the technology acceptance model and a proposal for a paradigm shift. *Journal of the Association for Information Systems* 8, 244–254.
- Cuban, L. (2003). *Oversold and Underused: Computers in the Classroom*. Cambridge, MA: Harvard University Press.
- Drori, G. S. and Jang, Y. S. (2003). The global digital divide: A sociological assessment of trends and causes. *Social Science Computer Review* 21(2), 144–162.
- Kozma, R. (ed.) (2003). *Technology, Innovation, and Educational Change: A Global Perspective*. Eugene, OR: ISTE.
- Law, N., Pelgrum, H., and Plomp, T. (eds.) (2008). *Pedagogy and ICT Use in Schools around the World: Findings from the SITES 2006 Study*. Hong Kong: Springer.
- Martin, M. O., Mullis, I. V. S., Foy, P., et al. (2008). *TIMSS 2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College. <http://timss.bc.edu/TIMSS2007/sciencereport.html> (accessed August 2009).
- Mullis, I. V. S., Martin, M. O., Foy, P., et al. (2008). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College. <http://timss.bc.edu/TIMSS2007/mathreport.html> (accessed August 2009).
- OECD (Organisation for Economic Co-operation and Development) (2007). *PISA 2006 Science Competencies for Tomorrow's World*. Paris: OECD.
- Pelgrum, W. J. and Anderson, R. E. (eds.) (1999). *ICT and the Emerging Paradigm for Life Long Learning*. Amsterdam, NT: International Association for the Evaluation of Educational Achievement.
- Pfeffer, J. (1982). *Organizations and Organization Theory*. Marshfield, MA: Pitman.
- Plomp, T., Anderson, R. E., Law, N., and Quale, A. (eds.) (2009). *Cross National Policies and Practices on Information and Communication Technology in Education*, 2nd edn. Greenwich, CT, USA: Information Age.
- Rogers, E. M. and Rogers, E. (2003). *Diffusion of Innovations*, 5th edn. New York: Free Press.
- Servon, L. J. (2002). *Bridging the Digital Divide – Technology, Community and Public Policy*. Malden, MA: Blackwell.
- Van Dijk, J. A. G. M. (2005). *The Deepening Divide – Inequality in the Information Society*. Thousand Oaks, CA: Sage.
- Warschauer, M. (2003). *Technology and Social Inclusion – Rethinking the Digital Divide*. Cambridge, MA: MIT Press.
- Wenglinsky, H. (1998). *Does It Compute?: The Relationship between Educational Technology and Student Achievement in Mathematics*. Princeton, NJ: ETS.

Further Reading

- Anderson, R. E. (2001). Youth and information technology. In Mortimer, J. and Larson, R. (eds.) *The Future of Adolescent Experience: Societal Trends and the Transition to Adulthood* pp 175–207. New York: Cambridge University Press.
- Anderson, R. E. (2008). Large-scale quantitative research on new technology in teaching and learning. In Coiro, J., Knobel, M., Lankshear, C., and Leu, D. J. (eds.) *Handbook of Research in New Literacies*, pp 67–102. Mahwah, NJ: Erlbaum.
- Anderson, R. E. (2008). National contexts. In Plomp, T., Law, N., and Pelgrum, H. (eds.) *Pedagogy and ICT Use in Schools around the World: Findings from the SITES 2006 Study*, pp 37–66. Hong Kong: Springer.
- Bolt, B. and Crawford, R. (2000). *Digital Divide – Computers and Our Children's Future*. New York City: TV Books.
- National Research Council (NRC) (1999). *Being Fluent with Information Technology*. Washington, DC: National Academy Press.
- Spitzer, K. L., Eisenberg, M. B., and Lowe, C. A. (1998). *Information Literacy: Essential Skills for the Information Age*. Syracuse, NY: Eric Clearinghouse on Information and Technology (IR-104).

TECHNOLOGY AND LEARNING – AS A SUBJECT OF INSTRUCTION

Conceptions of Technology Literacy and Fluency

O Erstad, University of Oslo, Oslo, Norway

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Appropriation – Becoming part of one's identity as learner.

Cultural ecology of literacy – The practices involved in reading, writing, and exchanging information in online environments, as well as the values associated with such practices – cultural, social, political, and educational.

Digital competence – The confident and critical use of Information Society Technologies for work, leisure and communication. Consisting of knowledge, skills, and attitudes. These competencies are related to logical and critical thinking, to high-level information management skills, and to well developed communication skills. At the most basic level, ICT skills, comprise the use of multimedia technology to retrieve, assess, store, produce, present, and exchange information, and to communicate and participate in networks via the Internet.

Digital divide – Processes of inclusion and exclusion in the information society.

Fluency – The ability to reformulate knowledge, to express oneself creatively and appropriately, and to produce and generate information (rather than simply to comprehend it).

ICT literacy – The ability of individuals to use ICT appropriately to access, manage, integrate, and evaluate information, develop new understandings (create), and communicate with others in order to participate effectively in society.

Literate lives – What it means to be literate in a culture ('bildung'), and democratic participation.

Media/Internet literacy – The ability to access, understand and create communications in a variety of contexts.

Multiple literacies – The many different kinds of literacies needed to access, interpret, criticise, and participate in the emergent new forms of culture and society.

Technological literacies – Social practices in which texts (i.e., meaningful stretches of language) are constructed, transmitted, received, modified, shared (and otherwise engaged), within processes employing codes which are digitized electronically, primarily, through not exclusively, by means of (micro)computers.

Introduction

Change forces related to the advances of information and communication technologies (ICT) have created an increased interest in the social implications of these developments. Sociological descriptions of key processes in our societies today use terms such as the information society (Mattelart, 2003), to indicate the increased availability of information to the population at large, the knowledge society (UNESCO, 2005), to indicate an escalation in the importance of knowledge building (Bereiter, 2002) for all levels of social development, and the network society (Castells, 1996), to denote the way organizations and people work together in new ways due to the use of new technologies. All these point to the importance of knowing about information technology, which raises important questions about literacy and fluency in how we use and relate to such technologies.

The influential report *Being Fluent with Information Technology* (1999) from the National Research Council Committee on Information Technology Literacy (NRC) (commissioned by the National Science Foundation) in the US gives the following explanation of the difference between the terms literacy and fluency:

Generally, computer literacy has acquired a skills connotation, implying competency with a few of today's computer applications, such as word processing and email. Literacy is too modest a goal in the presence of rapid change, because it lacks the necessary staying power. As the technology changes by leaps and bounds, existing skills become

antiquated and there is no migration path to new skills. . . . To adapt to changes in the technology . . . involves learning sufficient foundational material to enable one to acquire new skills independently after one's formal education is complete. This requirement of a deeper understanding than is implied by the rudimentary term computer literacy motivated the committee to adopt fluency as a term connoting a higher level of competency. (NRC, 1999: 2)

As stated in this quote, it is necessary to distinguish between a more skills orientation and a higher level of competency. Other concepts like ICT literacy, digital competence, and digital literacy have later been used to indicate the same advanced use of technology as that which is denoted by fluency in the aforementioned quote, at the same time taking technological developments into consideration.

An Historical Dimension

A distinction is traditionally made between orality and literacy (Ong, 1982), with the latter being defined as reading and writing. In recent years, there has been an interest in how traditional conceptions of literacy change due to new digital technologies. An important point is that technology literacy and fluency changes over time due to sociocultural processes (Scribner and Cole, 1981). The German literary scientist and media theorist Friedrich Kittler has described this as the development of different cultural techniques over time (Kittler, 1990). Similar perspectives are reflected in sociocultural learning theories where learning is related to the use of specific artefacts and tools (Säljö, 1999).

This also points to the relationship between mastery and appropriation (Wertsch, 1998). One might, for example, master the use of the technology without having appropriated it so that it becomes a part of one's identity as a learner. This can easily be the situation when introducing digital technologies in schools without considering the broader social issues of how children and young people use such technologies outside of schools (Buckingham, 2003).

The link between computer technology and literacy in educational practices can be described as moving through three main phases. The first is the mastery phase (1960s to the mid-1980s), with an emphasis on gaining knowledge and skill on how to master the computer and different programming languages. The second is the application phase (mid-1980s to late 1990s) with a more accessible user interface, whereby the computer is perceived as an everyday tool, which can be applied to a wide range of activities in education, work, leisure, and the home. This is accompanied by the production of training materials and the appearance of mass certification schemes focusing on basic levels of ICT

competence. The third is the reflective phase (late 1990s on). Movement to this phase was stimulated by the recognition that ICT could be a vehicle to accomplish student-centered pedagogies. All students need key skills of communication, numeracy, use of ICT, and learning how to learn. In terms of the evolution of computer/ICT literacy, there is an awareness of the need for more critical, evaluative, and reflective approaches to using technology.

An important break in the conceptual understanding of literacy and technology was brought about by what has been called the new literacy studies in the 1970s and 1980s. Several researchers at that time (see, e.g., Scribner and Cole, 1981; Heath, 1983; Street, 1984; Graff, 1987) were critical about the conception of literacy as a neutral set of skills, or what Street (1984) describes as the autonomous model of literacy, which means literacy as a set of neutral skills can be used in different contexts and for different purposes to complete a set of tasks. The new literacy studies expanded this limited notion of literacy to take into account sociocultural influences. The term literacies emerged to signal the different ways people use language and the different systems of representation in social practices. One consequence was that the concept of literacy was opened up to include interaction with different text forms, which need to be studied in different social practices (Barton, 1994). In addition, this trend has been reinforced by studies of how children and youth use different media and media education in schools, where the term media literacy has been used to emphasize the need to teach children about the social and cultural influences of different media in our society (Tyner, 1998; Buckingham, 2003). The term digital literacy builds on these conceptions and is then linked to the development of digital technologies and media forms.

In the mid-1990s an international group of researchers, the New London Group (Cope and Kalantzis, 2000), gathered to make a statement on the research agenda on literacy and technologies. They argued for the use of multiliteracies as an overall term. Their research perspective implies that literacy is a many faceted thing, from Gee's (2003) work on computer games to Kress's (2003) research on multimodal literacy. In a way, this points back to what Freire and Macedo (1987) described as reading the word and the world, implying that literacy is linked to a critical analysis of the social framework in which men exist and it is embedded in social practices that are important for people. It is not only about how we use specific artefacts, but also about the social implications of such artefacts for knowledge production (Wertsch, 1998; Bereiter, 2002) and learning (DiSessa, 2001).

In her book *Literacy for Sustainable Development in the Age of Information* Rassool (1999) presents an overview of different debates on literacy during the last decades. Her point is that research perspective on technology and literacy needs to reconceptualize power structures within the information society, with an emphasis on communicative

competence in relation to democratic citizenship. Digital technologies create new possibilities for how people relate to each other, how knowledge is defined in negotiation between actors, and how it changes our conception of learning environments in which actors make meaning. Empowerment is related to the active use of different tools, which must be based upon the prerequisite that actors have the competence and critical perspective on how to use them for learning.

In the *Handbook of Literacy and Technology* (1998), with the subtitle ‘Transformations in a post-typographic world’, David Reinking *et al.* present several perspectives on how the development of digital technologies changes conceptions of text, of readers and writers, and ultimately of literacy itself. This implies that technology literacy and fluency relate to changes in traditional cultural techniques such as reading and writing, and yet meanwhile opening up new dimensions to what it means to be a competent reader and writer in our culture.

Definitions and Conceptions

What exactly should be included within the conceptual domain of literacy has become increasingly fuzzy, especially among those educators and researchers whose professional interests emanate from that term (Reinking, 1998: xiv). This is, of course, due to the fact that literacy is not a static term but relates to technological innovations, and cultural and political strategies and developments.

In the report *Digital transformations: A framework for ICT literacy* (ETS, 2002), an international group of scholars, initiated by the Educational Testing Service (ETS) in the US, decided to use the concept of ICT literacy instead of other terms such as competency, ability, or fluency used by others. This indicates that concepts, conceptions, and definitions within this field changes over time. It is, however, possible to define two main conceptions of technology literacy and fluency. One is oriented toward skills needed to operate the technology, and the other toward social practices where technology is embedded.

Lankshear and Knobel (2006) describe, on the one hand, the traditional conception of digital literacy as the in thing, implying that it is about relating to the technology itself, as seen in the way Gilster (1997) argues for digital literacy or Potter (2001) for media literacy. The last decade of research has been on trying to map the interrelationship between cognition and the use of technological tools (see, e.g., Lajoie, 2000; Gärdenfors and Johansson, 2005), which builds an understanding of literacy and fluency as a combination between certain technologies and cognitions.

On the other hand, Lankshear and Knobel (2006) argue for a focus on digital literacies as social practices, as technology used in certain social and cultural settings.

To a larger extent than the former, this conception sees literacy and learning as embedded in social practices inside and outside of schools, for example, in the case of young people playing video games (Gee, 2003). Lankshear (1997) proposes a definition of technological literacies in line with this thinking:

Technological literacies may be defined as social practices in which texts (i.e. meaningful stretches of language) are constructed, transmitted, received, modified, shared (and otherwise engaged), within processes employing codes which are digitized electronically, primarily, though not exclusively, by means of (micro)computers Our main interest . . . is with *extant communications and practices of reading, writing, viewing, manipulating, communicating, etc. digital texts*, and their potential integration into critical forms of literate practice. (1997: 141)

This definition is not bound by certain technologies. It proposes to study literacies in practice (what people do with technologies and digital texts), and not as something predescribed, indicating that we need to understand what people are already practicing concerning technological literacies and what the role of education should be in employing such literacies for new knowledge levels.

In her book *Literacy in a Digital World* (1998), Kathleen Tyner studies some of the elements of a modern interpretation of literacy both related to what she terms tool literacies, to indicate the necessary skills to be able to use the technology, and literacies of representations, to describe the knowledge of how to take advantage of the possibilities that different forms of representation give the users, especially the new information and communication technologies. This indicates a similar division between a tool orientation of literacy and a more reflective social process.

Ofcom, the independent regulator and competition authority for the UK communications industries, initiated several reviews on media and Internet literacy defining it as “the ability to access, understand and create communications in a variety of contexts” (Buckingham, 2004: 2). This definition indicates different dimensions of technology literacy and fluency, with basic access as the first and foremost. Understanding includes both comprehension and critique. Moreover, creation includes both interaction with media and creation of media by the public (Livingstone *et al.*, 2004).

On a broader cultural level, Selfe and Hawisher (2004: 5) discuss what they refer to as cultural ecology of literacy, examining how factors such as age, gender, ethnic and racial group, and geographical background interplay with historical, cultural, economic, and political factors to affect a person’s development of technological literacy. Referring to the conflicting views on defining technological literacies, they propose their definition as “the practices involved in reading, writing, and exchanging information in online environments, as well as the values associated with such

practices – cultural, social, political, and educational” (Self and Hawisher, 2004: 5).

Toward Multiple Literacies

Some terms are oriented toward specific areas of people’s use of technology, while other terms are more general, such as media, ICT, and digital. These broader terms can be said to draw upon other terms, which have gained new or increased relevance with the emergence of digital environments. Some even use the alternative term *electracy* (Ulmer, 2003; Erstad, 2003) to indicate a break from literacy and reading and writing toward using electronic media. The following are a few specific terms that seem to be important at present:

- *Information literacy*. This has developed in the US since the late 1980s as a re-focusing of bibliographic instruction in academic libraries. It indicates both the ability to recognize when information is needed and how to locate, evaluate, and use effectively the needed information (see also Bawden, 2001). With the increasing importance of the World Wide Web as a source of information, the idea of information literacy has gained more centrality.
- *Visual literacy*. This has developed out of art criticism and art education, and was initially concerned with both the physiology and psychology of perception, and the way in which artists and designers have used perspective, ratio, light, color, and other techniques of visual communication (Messaris, 1994). Visual images have of course always been a powerful medium for the interpretation of information and the communication of meaning. Yet today, the wealth and complexity of visual imagery is stronger than ever.
- *Multimodal literacy*. Kress argues that in the posttext language is itself being affected by visual forms. “The screen more than the page is now the dominant site of representation and communication in general, so that even in writing, things cannot be left there The screen is the site of the image, and the logic of the image dominates the semiotic organisation of the screen” (Kress, 2003: 65).
- *Communication literacy*. This emphasizes communication as a key element in the educative process. This perspective builds on the increased interest in the collaborative learning as seen in Computer Supported Collaborative Learning (CSCL). Learners must be able to communicate effectively as individuals and work collaboratively in groups.

Some literacy theorists have sought to hold together the many new literacies under some umbrella concepts stressing the plurality of literacies, such as *multiliteracies* (Cope and Kalantzis, 2000, 2009; Snyder, 2002) and

metamedia literacy (Lemke, 1998). According to Kellner (2002: 163), “The term ‘multiple literacies’ points to the many different kinds of literacies needed to access, interpret, criticise, and participate in the emergent new forms of culture and society.” Kress (2003) however argues against the multiplicity of literacies, suggesting that it leads to serious conceptual confusion. He believes that instead of taking this path, it is necessary to develop a new theoretical framework for literacy, which can use a single set of concepts to address the various aspects of literacy. Although Tyner (1998: 63–68) recognizes the need to refer to *multiliteracies*, she feels that the dangers of oversimplification, and of isolation of literacies that are complementary and should therefore be linked, mean that constellations of linked literacies should be clearly identified and the relationships drawn out, and that the broad heading literacy be retained as an overarching concept.

Developing Frameworks

The different definitions and conceptions of technology literacy and fluency have been related to certain frameworks and the development of standards for educational practices. In January 2001, the ETS assembled a panel to develop a workable framework for ICT literacy. The outcome was the report *Digital Transformation. A Framework for ICT Literacy* (ETS, 2002). Building on this document one might, as the Australian authorities have done (Ainley *et al.*, 2006), define ICT literacy as

the ability of individuals to use ICT appropriately to access, manage, integrate and evaluate information, develop new understandings (create), and communicate with others in order to participate effectively in society.

The concepts mentioned are described in **Figure 1**.

All of these terms are oriented toward information handling. They also relate to the issues of problem solving and self-regulation. This consists of more general competencies that are not connected to specific subjects in school or specific contents in subjects. They can be taught and are not only related to what is learned in school settings, but also to situations outside the school.

Similar points are made in the earlier report *Being Fluent with Information Technology* (NRC, 1999). Fluency in Information Technology (FITness) covers three types of knowledge:

1. Contemporary skills: “the ability to use particular (and contemporary) hardware or software resources to accomplish information processing tasks” (NRC, 1999: 18). Naturally these skills will change over time as hardware and software evolve.
2. Foundational concepts: “the basic principles and ideas of computers, networks, and information” (NRC, 1999: 2–3). These include computer structure, information systems,

<i>Item</i>	<i>Description</i>
Access	Knowing about and knowing how to collect and/or retrieve information
Manage	Applying an existing organizational or classification scheme
Integrate	Interpreting and representing information. It involves summarizing, comparing and contrasting
Evaluate	Making judgments about the quality, relevance, usefulness, or efficiency of information
Create	Generating information by adapting, applying, designing, inventing, or authoring information
Communicate	Processing information in a way that highlights main points and process it to others

Figure 1 Key concepts of ICT literacy. Data from Australian Report, written by Ainley *et al.* (2006), but which a Ministry report by MCEETYA.

networks, modeling, algorithmic thinking and programming, and the limitations of IT and its social impact.

3. Intellectual capabilities, which “integrate knowledge specific to information technology with problem domains of personal interest to individuals” (NRC, 1999: 20). These are general thinking skills which might be recognizable in many disciplines, and include sustained reasoning, managing complexity, testing solutions, evaluating information, collaboration, anticipating change, and expecting the unexpected.

The concept of fluency in this connotes the ability to reformulate knowledge, to express oneself creatively and appropriately, and to produce and generate information (rather than simply to comprehend it) (NRC, 1999: 9).

Other frameworks have used digital competence as an overall term. One example is the working group on key competences of the European Commission Education and Training 2010. This program identifies digital competence as one of the eight domains of key competences, defining it as “the confident and critical use of Information Society Technologies for work, leisure and communication” (European Commission, 2004: 14). Information society technologies (IST) are defined as “offering services based on the use of Information and Communication technologies (ICT), the Internet, digital content, electronic media, etc., via, for example, a personal computer, a mobile telephone, an electronic banking machine, an eBook, digital television, etc.” (European Commission, 2004). Digital competence is regarded as consisting of knowledge, skills, and attitudes, as shown in **Figure 2**.

Competence here refers to a combination of skills, knowledge, aptitudes, and attitudes, and the disposition to learn in addition to know how (European Commission, 2004: 3). Several initiatives for such standards are now being developed around the world. They are defined as important tools for teachers in the way they use technologies in their educational practices. It is, however, important

that such standards do not become static tests, but can relate to technological and cultural change processes.

In addition, it is important to stress that technology literacy and fluency is related to situational embedding, that is, the use of technology within life situations. To understand such processes we have to look at different contexts where literacy is practiced and given meaning. This is especially important when relating it to how children and young persons use digital technologies.

Several studies show that young people gain most of their competence in using digital technologies outside the formal institutions of knowledge building (Livingstone, 2002; Buckingham, 2003; Alvermann, 2002). Thus, digital competence among young people today is of direct relevance to discussions about learning in schools, and it seriously confronts earlier conceptions of literacy and learning.

Building on the above discussion, one might see conceptions of technology, literacy, and fluency as three different aspects that are brought together (**Figure 3**).

How we relate to texts is changing due to the fact that new genres and multimodal expressions are developing. New demands are then put on our competencies, fluency, and literacies in using such texts. Furthermore, such processes are framed within certain sociocultural practices where literacies are conceptualized.

Implications

Conceptions of technology literacy and fluency also open up broader implications.

Literate Lives

This refers to issues of what it means to be literate in a culture. It has a philosophical connotation to the German term *bildung*. Related to digital technologies, *bildung* suggests an integrated, holistic approach toward cultural competence.

Framework for key competences in a knowledge-based society				
Domain	Definition of the competence	The competence consists of the following elements of knowledge, skills, and attitudes as appropriate to the context		
		Knowledge	Skills	Attitudes
Digital competence	<p>Digital competence involves the confident and critical use of Information Society Technologies (IST)¹⁸ for work, leisure and communication.</p> <p>These competences are related to logical and critical thinking, to high-level information management skills, and to well developed communication skills.</p> <p>At the most basic level, ICT skills comprise the use of multimedia technology to retrieve, assess, store, produce, present and exchange information, and to communicate and participate in networks via the Internet.</p>	<p>Sound understanding of the nature, role and opportunities of IST in everyday contexts comprises¹⁹:</p> <ul style="list-style-type: none"> Understanding the main computer applications, including word processing, spreadsheets, databases, information storage and management. Awareness of the opportunities given by the use of Internet and communication via electronic media (e-mail, video conferencing, other network tools); and the differences between the real and virtual world. Understanding the potential of IST to support creativity and innovation for personal fulfilment, social inclusion and employability. Basic understanding of the reliability and validity of the information available (accessibility/acceptability) and awareness of the need to respect ethical principles in the interactive use of IST. 	<p>As IST have many and growing applications in everyday life, such as learning and leisure activities, the required skills comprise:</p> <ul style="list-style-type: none"> Ability to search, collect and process (create, organize, distinguish relevant from irrelevant, subjective from objective, real from virtual) electronic information, data and concepts and to use them in a systematic way. Ability to use appropriate aids (presentations, graphs, charts, and maps) to produce, present or understand complex information. Ability to access and search a website and to use Internet based services such as discussion fora and e-mail. Ability to use IST to support critical thinking, creativity and innovation in different contexts at home, leisure, and work. 	<ul style="list-style-type: none"> Propensity to use IST to work autonomously and in teams; critical and reflective attitude in the assessment of available information. Positive attitude and sensitivity to safe and responsible use of the Internet, including privacy issues and cultural differences. Interest in using IST to broaden horizons by taking part in communities and networks for cultural, social and professional purposes.

Figure 2 Digital competence as defined by the EC key competences working group of the Education and Training 2010 Programme. From *Key Competences for Lifelong Learning: A European Reference Framework*. Directorate-General for Education and Culture, p 14. Copyright European Union, 2004.

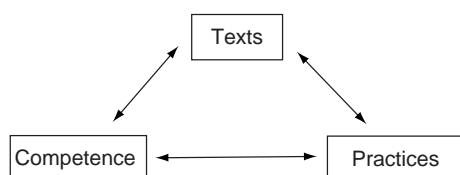


Figure 3 Three aspects that influence conceptions of technology, literacy, and fluency.

Traditionally conceived in an elitist way as a distinction between the literate and the illiterate, the technological development changes conceptions of what it means to be literate and enlightened in a technological culture.

Literate lives also refer to democracy and participation. In a society where digital technologies are having an impact on all aspects of social life, a question remains about how it might influence democratic processes, and what it means to be an informed citizen (e-citizenship).

These sentiments are echoed by the Declaration of Principles of the World Summit on the Information Society (2003):

Each person should have the opportunity to acquire the necessary skills and knowledge in order to understand, participate actively in, and benefit fully from, the Information Society and the knowledge economy. ... Awareness and literacy in ICTs are an essential foundation in this regard. (WSIS, 2003: §29, 31)

Digital Divide

Another important issue is the growing concern about digital divides in the world and within societies. In several of his texts, Warschauer (2007) has raised questions about inclusion and exclusion in the information society. Who are the winners and losers in this society? In this sense, literacy issues are redefined in relation to the challenges of technology-saturated societies, and furthermore, to the role of education and schooling as social processes of overcoming digital divides.

Curriculum Change

In several countries (e.g., Norway, Finland, Singapore, and New Zealand) technology literacy or digital competence has been written into school curricula as a key competence area. Issues of technology and education have thus moved from curriculum matters of access, implementation, and teacher competence, to issues of knowledge building and literacy when using technology. However, within the research field, there is still a need to specify how technology literacy and fluency may impact learning processes in schools.

Assessment

Another implication is the role of ICT and assessment as a driving force for educational change (McFarlane, 2003).

Without changes in the way we assess students and what we assess, technology literacy and fluency can easily become mere good intentions with little real impact on literacy practices in education. Some efforts have been initiated. For instance, the National Educational Technology Standards in the US have developed an online test for students and teachers to measure ICT literacy, which is based on certain standards and definitions mentioned earlier. According to these standards, all students should have opportunities to demonstrate the following performances: basic operations and concepts; social, ethical, and human issues; technology productivity tools; technology communications tools; technology research tools; technology problem-solving and decision-making tools. Other approaches toward performance assessment using ICT are now being tried out in Australia and Hong Kong (Ainley *et al.*, 2005).

Future Perspectives

Although there are many approaches to literacy, there is now a consensus that literacy is a social phenomenon as well as an individual characteristic. Literacy development is to a large extent linked to economic growth and the development of civic consciousness and political maturity. The literate person lives within the literate society:

Literacy is no longer exclusively understood as an individual transformation, but as a contextual and societal one. Increasingly, reference is made to the importance of rich literate environments – public or private milieus with abundant written documents (e.g. books, magazines and newspapers), visual materials (e.g. signs, posters and handbills), or communication and electronic media. (UNESCO, 2005: 159)

An important driving force today is the processes of convergence. This relates to how technologies merge, how production of content changes, how new text formats are developed, and how the users relate to information as part of communication networks in different ways. All this brings about new challenges for how we conceptualize technology literacy and fluency. As this article shows, the different perspectives reflect the complexity of skills, competencies, and literacies for being a citizen in the twenty-first century.

Issues of literacy and fluency in relation to different technologies also raise some key issues about how our education system is tuned to the challenges of the knowledge society. Conceptions of technology literacy and fluency in a future perspective would then be just as much about knowledge development and the educational framework as about using specific technologies.

Bibliography

- Ainley, J., Frailon, J., and Freeman, C. (2005). *National Assessment Program – ICT Literacy Years 6 & 10 Report*. Carlton South, Australia: the Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA).
- Ainley, J., Frailon, J., Freeman, C., and Mendelovits, J. (2006). Assessing information and communication technology literacy in schools. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Francisco, April 2006.
- Alvermann, D. (ed.) (2002). *Adolescents and Literacies in a Digital World, New Literacies and Digital Epistemologies*, vol. 7. New York: Peter Lang.
- Barton, D. (1994). *Literacy. An Introduction to the Ecology of Written Language*. Oxford: Blackwell.
- Bawden, D. (2001). Information and digital literacies: A review of concepts. *Journal of Documentation* 57(2), 218–259.
- Bereiter, C. (2002). *Education and Mind in the Knowledge Age*. Mahwah, NJ: Erlbaum.
- Buckingham, D. (2003). *Media Education. Literacy, Learning and Contemporary Culture*. Cambridge, UK: Polity Press.
- Buckingham, D. (2004). *The Media Literacy of Children and Young People. A Review of the Research Literature*. London: Ofcom.
- Castells, M. (1996). *The Rise of the Network Society*, vol. 1. Oxford: Blackwell.
- Cope, B. and Kalantzis, M. (eds.) (2000). *Multiliteracies: Literacy Learning and the Design of Social Futures*. London: Routledge.
- Cope, B. and Kalantzis, M. (2009). 'Multiliteracies': New literacies, new learning. *Pedagogies - An International Journal* 4(3), 164–195.
- DiSessa, A. (2000). *Changing Minds: Computers, Learning, and Literacy*. Cambridge, MA: MIT Press.
- Erstad, O. (2003). Electracy as empowerment. Students activities in learning environments using technology. *YOUNG. Nordic Journal of Youth Research*, 11(1), 11–28. London: Sage.
- ETS (2002). *Digital Transformation: A Framework for ICT Literacy*. Princeton, NJ: Educational Testing Service.
- European Commission (2004). *Key Competences for Lifelong Learning: A European Reference Framework*. Directorate-General for Education and Culture. ec.europa.eu/education/policies/2010/doc/basicframe.pdf.
- Freire, P. and Macedo, D. (1987). *Literacy: Reading the Word and the World*. London: Routledge and Kegan Paul.
- Gärdenfors, P. and Johansson, P. (2005). *Cognition, Education and Communication Technology*. Mahwah, NJ: Erlbaum.
- Gee, J. P. (2003). *What Video Games Have to Teach Us about Learning and Literacy*. New York: Palgrave Macmillan.
- Gilster, P. (1997). *Digital Literacy*. New York: Wiley Computer.
- Graff, H. J. (1987). *The Labyrinths of Literacy: Reflections on Literacy Past and Present*. London: Falmer Press.
- Heath, S. B. (1983). *Ways with Words*. Cambridge: Cambridge University Press.
- Kellner, D. (2002). Technological revolution, multiple literacies, and the restructuring of education. In Snyder, I. (ed.) *Silicon Literacies. Communication, Innovation and Education in the Electronic Age*, pp 154–169. London: Routledge.
- Kittler, F. (1990). *Discourse Networks 1800/1900*. Stanford, CA: Stanford University Press.
- Kress, G. (2003). *Literacy in the New Media Age*. London: Routledge.
- Lajoie, S. P. (ed.) (2000). *Computers as Cognitive Tools*. Mahwah, NJ: Erlbaum.
- Lankshear, C. (1997). *Changing Literacies*. Buckingham: Open University Press.
- Lankshear, C. and Knobel, M. (2006). Digital literacies: Policy, pedagogy and research considerations for education. *Nordic Journal of Digital Literacy* 1, pp 12–24. Oslo: University Press.
- Lankshear, C. and Knobel, M. (eds.) (2008). *Digital Literacies: Concepts, Policies and Practices*. New York: Peter Lang.
- Lemke, J. L. (1998). Metamedia literacy: Transforming meanings and media. In Reinking, D., McKenna, M., Laboo, L. D., and Kieffer, R. D. (eds.) *Handbook of Literacy and Technology. Transformations in a Post-Typographic World*, pp 283–302. Mahwah, NJ: Erlbaum.
- Livingstone, S. (2002). *Young People and New Media*. London: Sage.

- Livingstone, S., van Couvering, E., and Thumim, N. (2004). *Adult Media Literacy. A Review of the Research Literature*. London: Ofcom.
- Mattelat, A. (2003). *The Information Society*. London: Sage.
- McFarlane, A. (2003). Assessment for the digital age. *Assessment in Education: Principles, Policy and Practice* 10(3), 261–266.
- Messaris, P. (1994). *Visual Literacy. Image, Mind & Reality*. Boulder, CO: Westview.
- NRC (National Research Council Committee on Information Technology Literacy) (1999). *Being Fluent with Information Technology*. Washington, DC: National Academy Press.
- Ong, W. J. (1982). *Orality and Literacy: The Technologizing of the World*. London: Routledge.
- Potter, W. J. (2001). *Media Literacy*, 2nd edn. Thousand Oaks, CA: Sage.
- Rassool, N. (1999). *Literacy for Sustainable Development in the Age of Information*. Clevedon: Multilingual Matters.
- Reinking, D., McKenna, M. C., Labbo, L. D., and Kieffer, R. D. (eds.) (1998). *Handbook of Literacy and Technology. Transformations in a Post-Typographic World*. Mahwah, NJ: Erlbaum.
- Säljö, R. (1999). Learning as the use of tools: A sociocultural perspective on the human-technology link. In Littleton, K. and Light, P. (eds.) *Learning with Computers: Analysing Productive Interaction*, pp 144–161. London: Routledge.
- Scribner, S. and Cole, M. (1981). *The Psychology of Literacy*. Cambridge, MA: Harvard University Press.
- Selfe, C. L. and Hawisher, G. E. (2004). *Literate Lives in the Information Age: Narratives of Literacy from the United States*. Mahwah, NJ: Erlbaum.
- Snyder, I. (ed.) (2002). *Silicon Literacies. Communication, Innovation and Education in the Electronic Age*. London: Routledge.
- Street, B. (1984). *Literacy in Theory and Practice*. Cambridge, UK: Cambridge University Press.
- Tyner, K. (1998). *Literacy in a Digital World. Teaching and Learning in the Age of Information*. Mahwah, NJ: Erlbaum.
- Ulmer, G. L. (2003). *Internet Invention. From Literacy to Electracy*. New York: Longman.
- UNESCO (2005). Education for all: Literacy for life. *Fourth Annual Education for All Global Monitoring Report*. Paris: UNESCO.

- UNESCO World Report (2005). *Towards Knowledge Societies*. http://portal.unesco.org/ci/en/ev.php-URL_ID=20507&URL_DO=DO_TOPIC&URL_SECTION=201.html (accessed August 2009).
- Warschauer, M. (2007). *A Literacy Approach to the Digital Divide*. <http://www.gse.uci.edu/faculty/markw/lit-approach.pdf> (accessed August 2009).
- Wertsch, J. (1998). *Mind as Action*. New York: Oxford University Press.
- World Summit on the Information Society (WSIS) (2003). Draft declaration of principles. *World Summit on the Information Society, Geneva*. http://www.itu.int/wsis/documents/doc_multi.asp?lang=en&id=1154|1155 (accessed August 2009).

Further Reading

- Cope, B. and Kalantzis, M. (eds.) (2000). *Multiliteracies: Literacy Learning and the Design of Social Futures*. London: Routledge.
- National Research Council Committee on Information Technology Literacy (1999). *Being Fluent with Information Technology*. Washington, DC: National Academy Press.
- Reinking, D., McKenna, M. C., Labbo, L. D., and Kieffer, R. D. (eds.) (1998). *Handbook of Literacy and Technology. Transformations in a Post-Typographic World*. Mahwah, NJ: Erlbaum.
- Tyner, K. (1998). *Literacy in a Digital World. Teaching and Learning in the Age of Information*. Mahwah, NJ: Erlbaum.
- Williams, K. (2003). Literacy and computer literacy: Analyzing the NRCs 'being fluent with information technology'. *Journal of Literacy and Technology* 3(1). www.literacyandtechnology.org/volume3/literacy.pdf.

Relevant Website

- <http://www.cnets.iste.org> – National Educational Technology Standards.

TECHNOLOGY AND LEARNING – ASSESSMENT

Contents

Technology and Formative Assessment

Technology for Large-Scale Assessment

Technology Supports for Assessment Design

Technology and Formative Assessment

J W Pellegrino, University of Illinois at Chicago, Chicago, IL, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Effect size – A measure of the strength of the relationship between two variables in a statistical population, or a sample-based estimate of that quantity. The magnitude is often expressed in terms of the difference between two means divided by a standard deviation for the data.

This article describes how technology can be and is being used to support the process of implementing quality assessment, especially formative assessment. First, it considers what is meant by formative assessment, why it is such an important assessment activity, and how technology might support implementation of a range of assessment processes. The discussion includes some of the issues related to the implementation and use of technology-based tools and systems for formative functions. Then it presents examples of technology-based applications, including large-scale, district- or state-level commercial systems that have come into prominence in the last few years as well as small-scale, content- and classroom-based diagnostic systems often derived from cognitive and instructional research. The article concludes with a brief discussion of current opportunities and possibilities for the future.

A Consideration of Definitions, Timescales, Users, and Uses

Wise instructional decision making and differentiated instruction depend on teachers being able to take evidence of what students know and can do, compare it to expected learning outcomes, and create learning opportunities that reflect appropriate next steps for individual students (Donovan and Pellegrino, 2004; NCTM, 1995,

2000; Webb, 1997). Quality classroom-based assessment can support such a process and serve as a critical component of effective teaching, typically leading to enhanced student learning outcomes (Black *et al.*, 2004; Pellegrino, *et al.*, 2001; Shepard, 2000; Stiggins, 2001, 2002). For example, there is now a substantial body of evidence from K-12 classrooms showing that quality formative assessment practices enhance teaching effectiveness and produce student academic achievement gains ranging from 1/2 to a full course grade, reflecting effect sizes of 0.4 to 0.7 or higher (Black and Wiliam, 1998). (Such effect sizes reflect higher performance in conditions with formative assessment relative to control conditions without such assessment. The difference is on the order of 0.4 to 0.7 standard deviations relative to the mean performance of the control conditions.) Similar findings have been obtained in studies conducted within higher education settings (Nyquist, 2003).

Often it is assumed that classroom assessment is synonymous with formative assessment, and that large-scale assessment, such as district-, state-, or national-level tests, is synonymous with summative assessment. In fact, summative and formative functions can be identified for most assessment activities regardless of the level at which they are operating (see, e.g., Pellegrino and Hickey, 2006). As highlighted in influential discussions by Sadler (1989) and Black and Wiliam (1998), assessment is ultimately about feedback. These and many other articles have led to widespread appreciation that assessment is maximally valuable to educators to the extent that it generates useful information for advancing learning and instruction so long as the information is then acted upon. It is the latter feature – the use of assessment to directly improve the outcomes of teaching and learning that makes something formative.

Drawing from the work of Lemke (2000), it is apparent that different assessment practices can be understood as operating at different timescales spanning minutes, days,

weeks, months, and years. Timescale is important because the different competencies that various assessments aim to measure (and therefore the appropriate timing for being impacted by feedback) are timescale specific. What teachers and students say in class constitute verbal exchanges; these exchanges make up the lesson; a sequence of lessons make up the unit; units form a curriculum, and the curricula form an education. Each of these elements operates on a different cycle or timescale: second to second, day to day, month to month, and year to year. All can be potentially impacted by assessment and thus there is a tendency to confuse the timescale of the assessment with the function served – formative or summative. It is how useable the information may be, combined with how the information is used, that defines the function of an assessment as formative or summative. Given that there are formative uses for assessments that operate at different timescales, it must be noted that evidence of the positive impact of formative assessment on student learning outcomes has been most clearly obtained for cases where it is very close and proximal to the timing of instruction (see also Ruiz-Primo *et al.*, 2002). The issue of timescale and proximity to instruction is one that we return to in the sequel when we consider technology-based systems that have been developed commercially for use at school and district levels versus those that are much closer to curricular content and the instructional process.

As indicated above, assessment results allow educators and policymakers to address a variety of questions about how a district, school, program, group, or individual is performing. The goal is to make assessment results available to the right people, in the right form, at the right time so that assessment informs decisions and actions. Before individuals can answer questions, they must have the assessment results to work with. The goal is to collect the results in timely and efficient ways that allow them to answer their most important questions. Finally, it is not enough to answer questions – they must act on those answers. The goal is to help people use assessment results to make decisions about what actions they should take. A technology infrastructure is a collection of tools and processes that help people achieve these goals.

There are a variety of roles that technology might play in this complex process. It is likely that a collection of tools, rather than a single tool, will fulfill these roles. However, information must flow appropriately between tools to support the differing needs of different people. **Table 1** lists some of the roles that technology might play in the assessment process. Part of designing and implementing an assessment system for a school, district, or state is determining the appropriate roles for technology in that system.

Given the range of assessment tasks and data on student performance that are potentially available within the

Table 1 Functions of technology in assisting the assessment process

<i>Collecting</i>	<i>Reporting</i>	<i>Using</i>
Create assessment	Manage assessment data	Identify resources
Administer assessment	Analyze assessment data	Identify possible actions
Collect assessment data	Create reports	Design professional development
Score assessment	Distribute reports	

Source: Reproduced from Table 2.3, p45, in Pellegrino, J. W., & Goldman, S. R. (2007). *Beyond rhetoric: Realities and complexities of integrating assessment into teaching and learning*. In C. Dwyer (Ed). *The future of assessment: Shaping teaching and learning* (pp. 7–52). Mahwah, NJ: Erlbaum.

educational context, it is clear that the processes of collecting, scoring, and interpreting assessment data and then longitudinally tracking student performance are formidable. It is highly unlikely that any teacher, principal, or district or state administrator would be able to be successful in using all the assessment data available to him or her without the support of technology tools to assist in such a process across a variety of assessment levels and timeframes. Thus, an important direction for development and implementation of assessment systems is the design of specific and general technology-based tools that can assist educators in managing the assessment process. Ideally, one would like to have a system with extensive diagnostic assessment capability. That is potentially far off, however, and requires considerable research and development efforts. Even so, some very basic data collection, data entry, and data management tools can be of great assistance to teachers, principals, and others. The design and deployment of even simple technology tools must ultimately rely on a technology infrastructure that connects the classroom to powerful database management and information retrieval systems that operate within and across schools and systems. This is especially true when the classroom assessment data are viewed as part of a coordinated system of assessment data that would potentially include curriculum-embedded assessment information, benchmark assessment data, and state-level test data.

The next section describes two distinct classes of technology-based systems and tools designed to support the formative functions of assessment. The key distinction between them is how close the assessment is to integration with classroom curriculum and instruction – very close and proximal (tied to specific curriculum content) versus distal (tied to state or national standards).

Examples of Technology-Based Assessment Resources and Tools

Systems Driven by Content Standards – Large-Scale, Commercial Products

In the United States the number of summative assessments has been increasing recently with the authorization of No Child Left Behind Act (NCLB) at the federal level. In preparation for statewide exams required under NCLB, school districts are turning to more frequent in-house assessments that can produce data useful for decision making. Schools that effectively collect and utilize data have been shown to improve student achievement. This culture of using data to improve student achievement involves making decisions based on a variety of data such as instructional practice, goal implementation, achievement, and assessment data (Datnow *et al.*, 2007).

Perie *et al.* (in press) discuss the fact that while most assessment systems created commercially are marketed as formative assessment systems, they should properly be classified as interim assessment systems. Typically, these systems are either web- or server-based software with item banks that teachers use to provide periodic checks of student understanding. Those periodic checks can be at any designated time – monthly, tri-monthly, etc. – but rarely provide diagnostic feedback that teachers and students can use to address immediate deficiencies. In reviewing several commercially available assessment systems, we found this to be true with few exceptions.

Often these assessment systems are constructed to include items from the core subject areas and claim to be aligned to state or national standards. Some companies will specify that their items are aligned with the standards of a specific state (such as California) or those of the National Council of Teachers of Mathematics (NCTM, 2000). In either case, school officials need to check alignment claims for themselves and evaluate the assessment items' alignment with their own local standards and purposes. To the extent that assessment systems are aligned with state standards, they have the potential to create a practice summative assessment mirroring the yearly state exam. Most commercial assessment systems are touted as providing diagnostic, formative, and summative assessments, all from the same item bank. The difference in the assessments comes from the timing of the assessments, the frequency, and how the results are used.

A key theme found throughout evaluation of assessment companies' technology products for creating and dealing with assessment data, is the speed and ease of scoring. The process of formative assessment begins when assessment results are available with a very short turnaround time. From a commercial perspective, it seems that it is the turnaround time that qualifies an assessment as formative rather than the data's usefulness in providing feedback to both the teacher and the student. In some cases, the

predictive validity of the assessment also places it within the formative realm, since proficiency on the state NCLB assessments is considered the learning goal. If assessments are given early enough, teachers can use assessment data to diagnose student deficiencies and address them before administration of the state test. District officials like the ability to give an assessment at the beginning of the year for student placement purposes, and in the middle and end of the year for evaluative purposes. Teachers like the automatic scoring, ease of test construction, and reporting aspects.

These technology-based assessment systems rely on speed and an abundance of reporting options, but are mostly tied specifically to state and national standards rather than to an individual school's curriculum. Knowing the standards in which students are having difficulty is just a first step in a process of unpacking what the standards mean, identifying where they are taught within the curriculum, and connecting items with those aspects of the curriculum. As we noted earlier, formative assessment is not effectively assessment for learning if it is not utilized by the classroom teacher immediately to inform instruction and address student deficiencies.

While most assessment systems contain item banks that are aligned to content standards, it is unknown, but highly doubtful, that these item banks are fully developed to the point where an effective assessment can be constructed that can diagnose both what and why a student does not understand. This is the type of actionable information that teachers can use to change their instructional methods and goals so that students can succeed. In order for an assessment to do this, it has to be constructed in such a way that it includes a sufficient set of items with carefully selected distracters related to a single concept to pinpoint student problems and misconceptions. Typically, large-scale standardized assessments contain one to three items related to a target content standard and cannot identify student misconceptions. If the goal is to provide formative assessment, the item banks should be analyzed to verify their capabilities beyond creating small versions of a large scale standardized assessment, as well as their quality.

Table 2 presents a nonexhaustive set of technology-based assessment resources keyed to content standards. Little has been written regarding the use of these systems although some manufacturers have conducted small studies on the effectiveness of the assessments included within their system. These studies involve both the reliability of the assessment and an aspect of validity and are proudly displayed on company websites. Other studies involve the use of assessment in a general sense such as the effect of the frequency of assessments (Bangert-Drowns *et al.*, 1991; Kika *et al.*, 1992; Kling *et al.*, 2005; Martinez and Martinez, 1992). Further investigations are needed, however, in order to understand how best to use these systems, what

Table 2 Nonexhaustive set of technology-based assessment resources

<i>DataDirector</i> www.achievedata.com
<i>NWEA Measures of Academic Progress (MAP) Tests</i> http://www.nwea.org
<i>Odyssey</i> http://compasslearningodyssey.com/odysseyassessment.html
<i>MyAccess, Algebra Readiness Diagnostic Testing Program, Student Progress Monitoring System</i> http://www.vantagelearning.com
<i>Pearson Prosper Assessment System</i> http://formative.pearsonassessments.com/prosper/index.htm
<i>Princeton Review Online</i> http://www.princetonreview.com/educators/instructional/assessment.asp
<i>PLATO Assessment Solutions eduTest</i> http://www.plato.com/District-Solutions/Assessment-and-Data-Management.aspx
<i>Scantron Achievement Series</i> www.scantron.com
<i>Harcourt Assessment Learnia</i> www.harcourtassessment.com
<i>McGraw Hill Yearly Progress Pro</i> www.mhdigitallearning.com
<i>ASCD & Northrop Grumman ASPIRE Assessment System</i> http://www.ascd.org/portal/site/ascd/menuitem.b2283a36d0f2d7e257e54210e3108a0c/
<i>Acuity</i> www.acuityforschool.com

may be missing from them, and what a district should look for if purchasing a system.

Small-Scale, Research-Based (diagnostic) Systems

Among the most intriguing applications of technology are those that extend the nature of the problems that can be presented and the knowledge and cognitive processes that can be assessed. By enriching task environments through the use of multimedia, interactivity, and control over the stimulus display, it is possible to assess a much wider array of cognitive competencies than has heretofore been feasible. New capabilities enabled by technology include directly assessing problem-solving skills, making visible sequences of actions taken by learners in solving problems, and modeling and simulating complex reasoning tasks. Technology also makes possible data collection on concept organization and other aspects of students' knowledge structures, as well as representations of their participation in discussions and group projects. A significant contribution of technology has been to the design of systems for implementing sophisticated classroom-based formative assessment practices. Technology-based systems have been developed to support individualized instruction by extracting key features of learners' responses, analyzing patterns of correct and incorrect reasoning, and providing rapid and informative feedback to both student and teacher.

Given what we have noted above about the affordances of technology, it is relatively easy to differentiate the present class of technology-based assessment systems

Table 3 Research-based assessment resources

<i>DIBELS (Dynamic Indicators of basic Early Literacy Skills)</i> http://dibels.uoregon.edu
<i>Summary Street</i> http://www.pearsonkt.com/prodSSst.shtml
<i>IMMEX (Interactive Multimedia Exercises)</i> www.immex.ucla.edu
<i>DIAGNOSEr</i> http://www.diagnoser.com
<i>Carnegie Tutors</i> http://www.carnegielearning.com
<i>Assistments</i> http://www.assistments.org

relative to the large-scale systems discussed in the prior subsection. What unifies these systems as a group and separates them from the systems described above is that they are each based on an underlying analysis or model of learning and performance in the content domain and they typically make use of a range of task and performance formats, and yield more complex and multifaceted forms of data. These systems often explicitly promote or provide formative interventions and very detailed diagnoses of student understanding. The strong cognitive emphasis underlying the design of these systems results in a much deeper assessment, but this depth often occurs at the expense of breadth. That is, these systems are typically not able to assess as much of a content area as the large-scale systems, but they are often richer in diagnostic and instructionally useful information.

Table 3 presents the names of example research-based assessment resources and systems covering the reading, science, and mathematics content domains. This is by no means a comprehensive list of such systems and resources. We have, however, attempted to represent the variety of options that are becoming available. For several of these examples there is a substantial base of published work related to their implementation in classrooms and on their effectiveness in improving student learning and performance (see, e.g., Good *et al.*, 2001; Goodman *et al.*, 2006; Feng and Heffernan, 2005; Kintsch *et al.*, 2000; Minstrell and Kraus, 2005; Ritter *et al.*, 2007; Stevens and Dexter, 2003; Vendliniski and Stevens, 2002).

Summary and Implications

It is an exciting time in the field of assessment for several reasons. First, individuals have realized that there are multiple roles for assessment to play in the educational process and that one of the most valuable roles is the formative function of assisting student learning. Second, cognitive research and theory have provided us with rich models and representations of how students understand many of the key concepts in the curriculum, how they develop knowledge structures, and how to analyze and understand simple and complex aspects of student performance. Third, technology makes it possible for us to present students with a much wider array of tasks and

situations where students can learn and where they can show us what they know and how they know it. Thus, there is an interesting and powerful confluence among theory, research, technology and practice, especially as regards the integration of curriculum, instruction, and assessment.

In numerous areas of the curriculum, information technologies are changing what is taught, when and how it is taught, and what students are expected to be able to do to demonstrate their knowledge and skill. These changes in turn are stimulating people to rethink what is assessed, how that information is obtained, and how it is fed back into the educational process in a productive and timely way. This situation creates opportunities to center curriculum, instruction, and assessment around cognitive principles. With technology, assessment can become richer, more timely, and more seamlessly interwoven with multiple aspects of curriculum and instruction. As discussed earlier, the most useful kinds of assessment for enhancing student learning often support a process of individualized instruction, allow for student interaction, collect rich diagnostic data, and provide timely feedback. The demands and complexity of these types of assessment can be quite substantial, but technology makes them feasible. In diagnostic assessments of individual learning, for example, significant amounts of information must be collected, interpreted, and reported. No individual, whether a classroom teacher or other user of assessment data, could realistically be expected to handle the information flow, analysis demands, and decision-making burdens involved without technological support. Thus, technology removes some of the constraints that previously made high-quality formative assessment difficult or impractical for a classroom teacher. Several examples discussed earlier illustrate how technology can help infuse ongoing formative assessment into the learning process.

Clearly, we are just beginning to see how to harness technology to support the formative and summative functions of assessment. A great deal needs to be learned about the quality and efficacy of systems operating at both the large-scale level and the small-scale level. Not the least of the concerns facing us is integrating assessment tools and practices into the educational system and the practices of teachers. However, we must also take note of the fact that extremely powerful information technologies are becoming as ubiquitous in educational settings as they are in other aspects of people's daily lives. They are almost certain to continue to provoke fundamental changes in learning environments at all levels of the education system. Many of the implications of technology are beyond people's speculative capacity. Little more than 15 years ago, for example, few could have predicted the sweeping effects of the Internet on education and other segments of society. The range of computational devices and their applications is expanding exponentially, fundamentally

changing how people think about communication, connectivity, information systems, educational practices, and the role of technology in society.

See also: Technology for Large-Scale Assessment; Technology Supports for Assessment Design.

Bibliography

- Bangert-Drowns, R. L., Kulik, J. A., and Kulik, C. C. (1991). Effects of frequent classroom testing. *Journal of Educational Research* **85**(2), 89–99.
- Black, P., Harrison, C., Lee, C., Marshall, B., and William, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan* **86**(1), 8–21.
- Black, P. J. and William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan* **80**(2), 139–148.
- Datnow, A., Park, V., and Wohlstetter, P. (2007). Achieving with data: How high-performing school systems use data to improve instruction for elementary schools. Los Angeles, CA: Center on Educational Governance, University of Southern California. http://www.usc.edu/dept/education/cegov/focus/education_reform/publications.html.
- Donovan, M. S. and Pellegrino, J. W. (eds.) (2004). *Learning and Instruction: A SERP Research Agenda*. Washington, DC: National Academies Press.
- Feng, M. and Heffernan, N. T. (2005). Informing teachers live about student learning: Reporting in the assessment system. *Technology, Instruction, Cognition and Learning* **3**, 1–14.
- Good, R. H., Gruba, J., and Kaminski, R. A. (2001). Best practices in using dynamic indicators of basic early literacy skills (DIBELS) in an outcomes-driven model. In Thomas, A. and Grimes, J. (eds.) *Best Practices in School Psychology IV*, pp 679–700. Washington, DC: National Association of School Psychologists.
- Goodman, K. S., Flurkey, A., Kato, T., et al. (2006). *The Truth about DIBELS: What It Is, What It Does*. Portsmouth, NJ: Heinemann.
- Kika, F. M., McLaughlin, T. F., and Dixon, J. (1992). Effects of frequent testing of secondary algebra students. *Journal of Educational Research* **85**, 159–162.
- Kintsch, E., Steinhart, D., Stahl, G., et al. (2000). Developing summarization skills through the use of LSA-based feedback. *Interactive Learning Environments* **8**, 87–109.
- Kling, N., McCorkle, D., Miller, C., and Reardon, J. (2005). The impact of testing frequency on student performance in a marketing course. *Journal of Education for Business* **81**, 67–72.
- Lemke, J. J. (2000). Across the scale of time: Artifacts, activities, and meaning in ecosocial systems. *Mind, Culture, and Activity* **7**(4), 273–290.
- Martinez, J. and Martinez, N. C. (1992). Re-examining repeated testing and teacher effects in a remedial mathematics course. *British Journal of Educational Psychology* **62**, 356–363.
- Minstrell, J. and Kraus, P. (2005). Guided inquiry in the science classroom. In Donovan, M. S. and Bransford, J. D. (eds.) *How People Learn: History, Mathematics, and Science in the Classroom*, pp 475–512. Washington, DC: National Academies Press.
- NCTM (National Council of Teachers of Mathematics) (1995). *Assessment Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- NCTM (National Council of Teachers of Mathematics) (2000). *Principles and Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Nyquist, J. B. (2003). *The Benefits of Reconstructing Feedback as a Larger System of Formative Assessment: A Meta-Analysis*. Unpublished master's Thesis, Vanderbilt University, Nashville, TN.
- Pellegrino, J., Chudowsky, N., and Glaser, R. (eds.) (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. (National Research Council's Committee on the Foundations of Assessment). Washington, DC: National Academy Press.

- Pellegrino, J. W. and Hickey, D. (2006). Educational assessment: Towards better alignment between theory and practice. In Verschaffel, L., Dochy, F., Boekaerts, M., and Vosniadou, S. (eds.) *Advances in Learning and Instruction Series: Instructional Psychology: Past, Present and Future Trends. Sixteen Essays in Honour of Erik De Corte*, pp 169–189. Oxford: Elsevier.
- Perie, M., Marion, S., and Gong, B. (2009). Moving towards a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, **28**(3), 5–13.
- Quellmalz, E. and Pellegrino, J. W. (2009). Technology and testing. *Science* **323**, 75–79.
- Ritter, S., Anderson, J. R., Koedinger, K. R., and Corbett, A. (2007). Cognitive tutor: Applied research in mathematics education. *Psychonomic Bulletin and Review* **14**(2), 249–255.
- Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., and Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching* **39**(5), 369–393.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science* **18**, 119–144.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher* **29**(7), 4–14.
- Stevens, R. and Dexter, S. (2003). Developing teachers' decision-making strategies for effective technology integration: A simulation design framework. *Paper Presented at American Educational Research Association Conference*.
- Stiggins, R. J. (2001). *Student-Involved Classroom Assessment*, 3rd edn. Upper Saddle River, NJ: Prentice-Hall.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan* **83**(10), 758–765.
- Vendlinski, T. and Stevens, R. (2002). Assessing student problem-solving skills with complex computer-based tasks. *Journal of Technology, Learning, and Assessment* **1**(3). <http://www.jtla.org> (accessed August 2009).
- Webb, N. L. (1997). Determining alignment of expectations and assessments in mathematics and science education. *NISE Brief* **1**(2). <http://facstaff.wcer.wisc.edu/normw/1997alignmentbrief.htm> (accessed August 2009).

Technology for Large-Scale Assessment

R E Bennett, Educational Testing Service, Princeton, NJ, USA

© 2010 Elsevier Ltd. All rights reserved.

This article presents a conceptual framework for understanding the recent evolution of technology for large-scale assessment, gives examples of applications that illustrate that evolution, and identifies issues associated with the use of technology in large-scale assessment. (For purposes of the article, the term, technology, is taken in its narrow sense to refer to computer, network, software, and related capabilities. The term, large-scale assessment, is used to refer to professionally designed tests administered to large numbers of students under standardized conditions for making a variety of educational decisions.) **Table 1** presents a framework adapted from Bennett (1998), which describes three generations of large-scale assessment that differ in the degree to which they capitalize on advances in technology. The first generation is one of infrastructure building in which the hardware, software, and facilities for delivering tests by computer are put into place. Because this is an initial developmental phase, tests of this generation may be more costly than their paper versions. Tests of this generation take relatively limited advantage of technology. They measure traditional skills, use test designs and item formats closely resembling those of paper administration, are given in centers (e.g., commercial or school labs) using stand-alone computers or machines connected via a local area network, are administered as one-time measurement events, and primarily serve institutional (as opposed to individual) needs. Among the few advanced capabilities such tests are likely to offer is adaptive administration, where the measure is dynamically customized to match the skill level of the examinee taking it. (In adaptive testing, the examinee's skill level is re-estimated after each item is administered. The next item to be administered is then chosen based, in part, on that updated skill-level estimate.)

Next-generation tests represent a qualitative change from those of the first generation. Although they continue to serve institutional needs primarily and to be administered in centers, they leverage technology for such item formats as constructed response and for initial forays into the assessment of proficiencies not directly measurable through conventional means (e.g., skill in electronic information search). (A constructed-response item is a question for which the examinee must generate a response instead of selecting it from a short list of options.) Because the infrastructure has been substantially established, next-generation tests also can more effectively use technology to improve efficiency. They may employ technology to generate test items or to score constructed responses

automatically, and utilize the Internet, not only for test delivery, but for other internal processes (e.g., human performance scoring), as well as for customer interaction (e.g., examinee scheduling, score reporting).

The third generation, Generation R, represents reinvention in the sense that the purposes, processes, and results of testing differ in important ways from those of the two previous generations. Through the affordances of technology, Generation R tests are grounded in modern cognitive-scientific research; become more integrated with instruction, sampling performance repeatedly over time; employ complex simulations to better model real-world tasks; permit more natural interaction with computers; allow for the measurement of new skills in more sophisticated ways; and come to serve both institutional and individual needs.

We now turn to a discussion of examples within each of the three generations and of the issues those applications raise. In considering these examples, it should be noted that few fit unambiguously into a single generation because testing programs advance over time in their use of technology and because a program may simultaneously incorporate characteristics of more than one generation.

First-Generation Computer-Based Tests

The first generation of technology-based educational testing is in place to a significant degree for several assessment purposes. These purposes include college course placement, graduate and professional admissions, teacher certification, state accountability, and school achievement monitoring. With the introduction of the College Board's Computerized Placement Tests in 1986 (now ACCUPLACER) and of ACT's COMPASS, college course placement was one of the first purposes served by large-scale computer-based educational assessment. These adaptive tests are given millions of times per year at very low cost because institutional users take full responsibility for scheduling and administration, and because the low-stakes nature of the assessment purpose does not require the frequent creation of new test items.

Computerized testing for teacher certification dates to the early 1990s with the introduction of Praxis I, now the Praxis I: Pre-Professional Skills Assessments (PPST). This nationally administered test is intended for use by colleges and universities to evaluate individuals for entry into US teacher education programs. It is required for

Table 1 Three generations of large-scale educational assessment

Generation	Key characteristics
First-generation computer-based tests (infrastructure building)	<ol style="list-style-type: none"> 1. Primarily serve institutional needs 2. Measure traditional skills and use test designs and item formats closely resembling paper-based tests, with the exception that tests may be given adaptively 3. Administered in dedicated test centers as a one-time measurement 4. Take limited advantage of technology
Next-generation electronic tests (qualitative change and efficiency improvement)	<ol style="list-style-type: none"> 1. Primarily serve institutional needs 2. Use new item formats (including multimedia and constructed response), and may make initial attempts to measure new constructs, to begin to change what is assessed 3. Improve efficiency through automatic item generation, automated scoring, and the use of the Internet for internal processes, as well as for interactions with consumers 4. Administered in dedicated test centers as a one-time measurement
Generation R tests (reinvention)	<ol style="list-style-type: none"> 1. Serve both institutional and individual purposes 2. Integrated with instruction so that performance is sampled repeatedly over time 3. Designed according to cognitive principles 4. Use complex simulations that model real environments and allow more natural interaction with computers 5. Administered at a distance 6. Assess new skills in more sophisticated ways (e.g., through simulation or extended constructed response)

Adapted from Bennett, R. E. (1998). *Reinventing Assessment: Speculations on the Future of Large-Scale Educational Testing*. Princeton, NJ: Policy Information Center, Educational Testing Service.

certification by a majority of US states. The test is administered through dedicated centers, most of which are commercially operated.

At least three major programs use technology delivery for graduate and professional admissions. These programs include the Graduate Record Examinations (GRE) General Test, introduced on computer in 1992, the Graduate Management Admission Test (GMAT), first offered on computer in 1997, and the Test of English as a Foreign Language (TOEFL), introduced electronically in 1998 (A timeline of the history of the MBA and the GMAT, 2004; Kirsch *et al.*, 1998; Schaeffer *et al.*, 1995). Several hundred thousand tests are taken on computer by each of these programs annually, including in US and overseas venues. In their first-generation versions, each of these tests was adaptive (at least in part), and was delivered primarily in dedicated centers via local area networks, with transfer of test items and examinee responses accomplished through specially arranged private data lines.

For state accountability purposes, two programs stand out. Although both programs utilize the Internet for delivery, they are first generation in most other ways. The Virginia e-SOL (Standards of Learning) assessments, initiated in 2001, comprise well over a dozen different tests in English language arts, mathematics, science, and social studies (Department of Education and Virginia Information Technologies Agency, 2005). The tests are administered operationally at the middle- and high-school levels for such high-stakes decisions as promotion and graduation. The

e-SOL assessments were given over 900 000 times in the 2005–06 school year, with the overwhelming majority of the state's high-school students participating (R. Triscari, personal communication, September 5, 2006).

The Oregon Technology Enhanced Student Assessment (TESA) is an adaptive test administered in grades 3–12 in reading, mathematics, science, and social studies. It was administered over 1 million times in the 2004–05 school year (State: Online testing helped raise scores, 2005). Among other things, TESA is employed by the state for federal reporting purposes under the US No Child Left Behind Act.

Finally, for district achievement monitoring, the best example is the Northwest Evaluation Association's Measures of Academic Progress (MAP). This adaptive testing program includes measures in math, reading, language use, and science. The MAP tests are used by more than 2400 US school districts (Measures of Academic Progress, 2006). Because MAP tests are adaptive, they can more accurately measure the achievement of students who are at widely varying skill levels than can a conventional test. Moreover, MAP tests can be administered to the same students more than once a year for purposes of estimating progress.

The examples above illustrate several points. First, computer-based tests are already being administered on a large scale for a variety of educational purposes. In that regard, first-generation computer-based testing is established, even if it is not yet the dominant mechanism for all of those purposes.

A second point is that these first-generation tests are for the most part relatively simple, multiple-choice measures meant to closely resemble the task format of a paper-and-pencil exam. In fact, several of the programs, including Praxis I and the Virginia e-SOL assessments, offer paper forms intended to be comparable to the computerized versions. The most notable distinction from a paper test is in the adaptive nature of delivery and scoring utilized by several of the programs.

A third point is that because they constitute an initial developmental stage, these tests were inevitably expensive. Substantial investments had to be made in the infrastructure for large-scale technology-delivered testing – hardware, software, network capability, physical facilities, methods, and processes had to be purchased or newly created. As a result, in most cases, first-generation tests proved more costly than their paper counterparts (Wainer and Eignor, 2000).

These first-generation tests have helped identify at least two critical issues. One issue relates to infrastructure. For college course placement, teacher licensure, and graduate and professional admissions, the US infrastructure is adequate, especially when testing is offered continuously instead of in a limited number of fixed (and consequently larger) administrations. Outside of the US, the adequacy of the infrastructure varies widely, such that for the internationally offered GRE General Test, paper versions are sometimes the standard. At the school level, adequacy of infrastructure is also a concern. The US National Center for Education Statistics reports that there were 3.8 students for every Internet-connected computer located in public-school instructional rooms in 2005 (Wells and Lewis, 2006), probably enough to support simultaneous testing of grade cohorts. This ratio varied, however, by school size and minority enrollment, suggesting the potential for problems in some locations. Data from the Organization for Economic Cooperation and Development (OECD) suggest that, outside the US, the adequacy of the infrastructure is in general less developed (Schleicher, 2006).

A second critical issue relates to the comparability of scores. One type of comparability concerns the scores from computer versus paper tests. Students may perform differently if they are more comfortable or practiced at taking tests in a particular mode (e.g., in taking writing tests on paper vs. on computer). For tests offered in both modes, or for tests transitioning from paper to computer, it is essential to support claims regarding the comparability of scores empirically.

Such evaluations should, at a minimum, include consideration of the similarity of score distributions across modes and of the similarity of examinee rank orders. With adults, research has generally shown score distributions and rank orders to be highly similar across computer and paper formats for cognitive tests like those used in education (e.g., Gallagher *et al.*, 2000; Mead and Drasgow, 1993).

For school-age students, the existing research base is far less extensive and largely unpublished, so dependable conclusions are difficult to draw. The US National Assessment of Educational Progress (NAEP) has conducted two nationally representative studies, one in mathematics and one in writing, both focused primarily on 8th grade students (Bennett *et al.*, 2008; Horkay *et al.*, 2006). In both cases, the computer and paper tests did not appear to measure the same skills.

Even in those cases where a test is offered only on computer and all students are proficient in taking it that way, infrastructure differences among (and within) schools can affect score comparability and, hence, fairness. For example, under some computer-based test delivery models, questions may display differently from one machine to the next, changing test scores in unwanted ways. Bridgeman, *et al.* (2003) found scores on reading comprehension items taken from the SAT I: Reasoning Test to vary as a function of monitor size and resolution, presumably because of differences in the need to scroll text passages.

Next-Generation Electronic Tests

Next-generation tests are distinguished from first-generation tests by qualitative change achieved through the introduction of new item types, including constructed response, and more generally by attempting to measure new constructs not as easily assessed through conventional means. Further, next-generation tests use technology for efficiency improvement. Qualitative change and efficiency improvement may come about as additions to existing first-generation tests, thereby moving them to the next generation, or alternatively through the introduction of totally new measures.

The first example of a test that has introduced new item types is the new TOEFL, now called TOEFL iBT. The TOEFL iBT includes a speaking section in which the examinee responds via microphone to audio prompts delivered through a headset. Questions ask the examinee to do such things as listen to a short lecture and summarize it; or read a brief passage, listen to a short lecture, and answer a question calling for integration of information from the two sources.

A second example comes from the Indiana Core 40 End-of-Course Assessments, available in selected subjects for middle- and high-school students. These tests were among the first electronic assessments to contain constructed-response item types, a qualitative change from the multiple-choice format that dominates first-generation tests. The English 11 assessment includes short-answer and essay questions for which students type their responses. The algebra assessments include questions that call for the entry of numerals, mathematical expressions, and short-answer text responses.

An initial attempt to measure constructs not directly accessible through conventional means is illustrated by the Educational Testing Service (ETS) iSkills assessment (Educational Testing Service (ETS), 2008). This assessment is intended for students transitioning to college or already enrolled, and targets the ability to use technology as a tool for cognitive purposes. Questions are presented in an applied context and the examinee is asked to use generic versions of such common tools as a word processor, presentation software, e-mail client, web browser, and spreadsheet.

Whereas the infrastructure development needs of first-generation tests often did not allow for cost savings, next-generation tests are applying technology to internal and external processes, including item development, constructed-response scoring, and interaction with customers. An example of a tool employed for operational item development is the Mathematics Test Creation Assistant (Singley and Bennett, 2002). With this tool, the test developer creates an item model, an abstract representation of a question class from which the computer can generate multiple instances. A given model can be tightly structured so as to produce instances that are highly similar in difficulty (Graf *et al.*, 2005), or more loosely formulated to produce instances with a wider range of difficulty. This tool is being used to create questions for the GRE General Test quantitative section, as well as for such state programs as the Virginia Standards of Learning assessments.

Technology is being applied to constructed-response scoring in two distinct ways. Online scoring involves the presentation of an examinee's constructed response to a human judge via computer display. If the test was administered on paper, the scoring vendor must first convert the response to digital form. If the test was administered on computer, no such conversion is needed. Most major testing companies have online scoring systems. In some configurations, human judges are gathered in a central location to conduct scoring. In other configurations, scoring is distributed, with judges working from home or office via the Internet. Online scoring is employed for the constructed-response portions of such tests as the NAEP, Advanced Placement (AP) Program Examinations in Chinese and Japanese, the California High School Exit Exam (CAHSEE), the GRE General Test, Praxis: Professional Assessments for Beginning Teachers, and the TOEFL iBT, including items from the speaking section described above.

In contrast to online scoring, automated scoring entails grading a constructed response entirely by machine. Several such programs exist for scoring essay-length responses, including the Intelligent Essay Assessor (Pearson Knowledge Technologies), Intellimetric (Vantage), Project Essay Grade (PEG) (Measurement, Inc.), and e-rater (ETS). Tests that use automated scoring operationally include the constructed-response portions of the

GMAT, ACT's COMPASS, the College Board's ACCUPLACER, and the Indiana Core 40 examinations in biology, algebra, and English 11. For the GMAT, the machine score is combined with that of a single human rater. For all of the other tests, the machine score is used alone.

Electronic interaction between customers and testing programs is perhaps best illustrated through advances in registration and reporting. Examinees for many educational testing programs may register using the Internet. Such registration is available for the College Board's SAT, the ACT Assessment, and the GMAT, among many others.

In addition to being able to electronically register, score results for many testing programs are also routinely delivered electronically, speeding the process of reporting considerably and making it less costly. For institutional purposes, electronic reporting can also make results more useful. The NAEP Data Explorer is probably the most advanced of such tools, allowing users to generate reports instantaneously with the variables and data representations of their choosing. The resulting reports may involve comparisons of the performance of demographic groups (by gender, race/ethnicity, poverty level, region of the country, school location, school type), or of groups categorized by student factors (e.g., affective disposition), instructional content and practice (e.g., modes of instruction), teacher factors (e.g., credentials and experience), or school factors (e.g., organization). Reports may include tables, graphs, charts, or text.

The examples above illustrate two points. First, these examples illustrate that large-scale testing programs are beginning to use technology broadly. Programs are not just using technology for test delivery and response collection but to improve the efficiency of other internal processes, including item creation and scoring. Further, programs are taking advantage of the Internet as a network to connect more directly with consumers for such purposes as registration and reporting. A second key point is that testing programs are beginning to use technology to change what is being tested through the introduction of constructed-response item types and by considering skill in using the computer as an object of assessment in its own right.

The advances associated with next-generation tests have brought novel issues. One such issue relates to automated scoring. Research suggests that automated essay scoring programs tend to agree with the scores of human judges to about the same degree as two judges agree between themselves (Keith, 2003). Should we care how these computer programs achieve that agreement? For example, should we care what response features are electronically judged and how those features are combined to produce scores? Some companies, claiming proprietary technology, refuse to release enough detail about their scoring mechanisms to permit evaluation of the processes

used. Other companies have released processing details. In such cases, however, it is not always clear that the automated scoring uses response features that account sufficiently for the characteristics of good writing or that combine those features in defensible ways (Bennett, 2006; Ben-Simon and Bennett, 2007). Ben-Simon and Bennett (2007) found that the dominant features used to compute scores automatically were essay length and mechanics, with only minimal weight given to content or organization and development. The concerns raised by such findings are potentially problematic because they may make automated scoring harder to explain, defend, and generalize to new testing purposes and populations.

Generation R (Reinvention)

Among the defining characteristics of Generation R tests are a basis in cognitive principles (as well as content considerations), the use of complex simulation, sampling of student performance repeatedly over time, integration of assessment with instruction, and the measurement of new skills in more sophisticated ways than possible in next-generation tests. Generation R tests produce information that can support summative and also, to a significant degree, formative decisions by repeatedly probing in a theoretically driven manner using instructionally sound tasks. In the Generation R view, more assessment opportunities allow for better tasks and for higher-quality information through aggregation. Although no operational examples exist in large-scale educational assessment, instances are emerging, each with some of the key characteristics.

One example is not an assessment at all but an instructional program that thoroughly integrates assessment. The Cognitive Tutors (Carnegie Learning) present instruction in mathematics designed around adaptive control of thought—rational (ACT-R), an extensive theory of learning, knowledge organization, and human performance (Anderson, 1993). Of particular relevance for Generation R is that the Cognitive Tutors dynamically build a rich representation of a given student's domain knowledge as the student interacts with the tutor by solving problems. The student may interact with the tutor repeatedly over time, updating the tutor's representation of the student's knowledge state and extending that representation to new domain components as problems relevant to those components are encountered. The tutor uses this representation (the student model) to make relatively fine adjustments to the instruction it presents.

A second example illustrates the use of extended, complex, highly contextualized simulations, as well as how such simulations might be employed formatively to measure new skills in sophisticated ways. The Key Stage 3 Information and Communication Technology (ICT) Assessment Tasks are available on demand through what was England's

National Assessment Agency (National Assessment Agency (NAA, 2008), now the Qualifications and Curriculum Development Agency. The purpose of the assessment tasks is to help develop information and communications technology skills and offer students instant feedback on their performance. The assessment tasks call for the integration of skills by requiring the use of multiple ICT tools in concert (e.g., word processor, web browser, and spreadsheet) to solve extended problems, in much the same way such tools are used in real work and academic environments.

Similar to England's Qualifications and Curriculum Development Agency, the US NAEP has been exploring the use of complex simulations for measuring new skills. In the Technology-Rich Environments Study (Bennett *et al.*, 2007), nationally representative student samples were administered measures intended to tap skill in using the computer for problem solving in a science context. Of special note is that the tasks were intended to afford incidental learning opportunities beyond those possible from conventional assessment. One task, shown in **Figure 1**, asked students to use a simulated Internet to find out, and explain why, scientists sometimes use helium gas balloons instead of other mechanisms to explore planetary atmospheric space. A second set of tasks, one of which is shown in **Figure 2**, asked the student to discover through experimentation the relationship between the payload mass that a scientific helium balloon could carry and the altitude to which the balloon could rise in the atmosphere. Performance on these tasks was scored in terms of both the quality of the constructed-response answers and process features (e.g., whether the experiments conducted covered a wide enough range of payload masses to support defensible conclusions about the relationship between mass and altitude). Based on student behavior, scale scores were produced using a psychometric methodology similar to that employed in operational NAEP assessments.

A vision for a system that incorporates many of the characteristics of Generation R has been proposed by Bennett and Gitomer (2009). In their view, first, accountability assessment, formative assessment, and professional support should be grounded in a common conceptual base derived, not only from curriculum objectives, but also from cognitive-scientific research. Second, they envision accountability and formative assessment as comprising primarily extended, integrated tasks worth teaching toward. Finally, they propose that accountability assessment be periodic, instead of a one-time event, and that the results of those periodic measurements be aggregated to form a final accountability result. Periodic assessment allows for at-risk students to be identified while there is still time to take instructional action and for tasks to be more instructionally worthwhile, because there is more time for assessment in the aggregate. Also, the aggregation of results across measurement occasions means that no student, teacher,

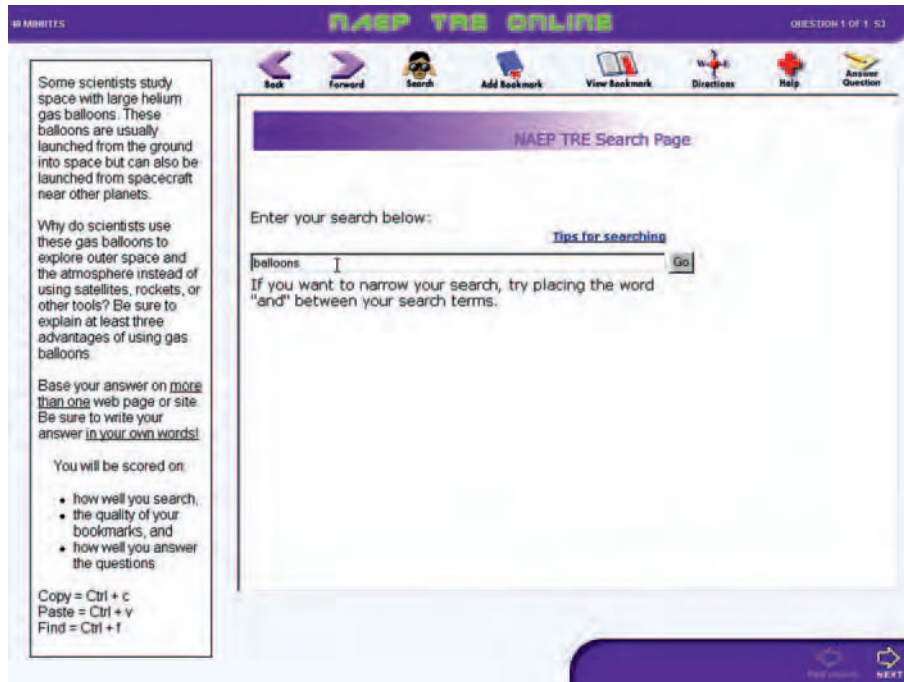


Figure 1 Interface and motivating problem for the NAEP technology-rich environments web-search scenario. From Bennett, R. E., Persky, H., Weiss, A. R., and Jenkins, F. (2007). *Problem Solving in Technology-Rich Environments* (NCES 2007-466). Washington, DC: National Center for Education Statistics, US Department of Education.



Figure 2 Interface and motivating problem for the NAEP technology-rich environments simulation scenario. From Bennett, R. E., Persky, H., Weiss, A. R., and Jenkins, F. (2007). *Problem Solving in Technology-Rich Environments* (NCES 2007-466). Washington, DC: National Center for Education Statistics, US Department of Education.

administrator, or school can suffer significant adverse consequences for one bad performance.

As noted, no operational Generation R tests exist. This situation is the case because many significant

issues remain to be solved. For one, how might cognitive principles be integrated with such content-based domain descriptions as those embodied in state or national curriculum standards? One possibility is to use cognitive

principles to translate the outcomes typically represented by curriculum standards into the processes, strategies, and knowledge structures needed to achieve those outcomes.

A second issue concerns the aggregation of information gathered from repeated samplings over time to produce summary proficiency estimates usable for such consequential decision-making purposes as school evaluation and student promotion. Would a simple average suffice? Should greater weight be given to more recent samplings?

Third, how should performance on complex tasks be scored? Responses to such tasks are, by their very nature, composed of many locally dependent components. For example, a student may misread instructions or make an early mistake that affects his or her responses to all subsequent task components. Similarly, a student with interest in a certain area may by chance encounter a task for which such background knowledge will provide an unfair advantage on all components. Such dependencies can perhaps be accounted for through psychometric modeling or be acceptably dampened through the use of a sufficient number of complex tasks.

Fourth, how might the use of a large-scale assessment comprising complex tasks be made affordable? Complex tasks are costly to produce and to score meaningfully. Affordability will come only through technological advances in the tools for task creation and automated scoring.

Finally, how might teachers and education policy-makers be convinced to spend more time on assessment? Repeated sampling, after all, means a test that is, in the aggregate, potentially longer than the traditional one-time administration. If assessment continues to be regarded as a burdensome activity comprising irrelevant tasks providing little instructionally useful information, no one should want more of it. However, if Generation R assessments can be designed as learning events, can consist of exercises worth teaching toward, and can produce actionable results, the reaction may be different.

Summary

This article has presented a conceptual framework for understanding the recent evolution of technology for large-scale assessment, given examples to illustrate that evolution, and identified issues associated with technology use. The conceptual framework described three generations of tests: (1) infrastructure building, (2) qualitative change and efficiency improvement, and (3) reinvention.

The examples reviewed make clear that a first generation of large-scale, computer-based educational tests is in place, even if such tests are not the dominant mode for all purposes. These first-generation tests essentially duplicate the key characteristics of paper-and-pencil measures, with the most notable distinction being the adaptive

nature of delivery and scoring utilized by several programs. The large-scale administration of first-generation tests for graduate- and professional-school admissions, college placement, teacher certification, state accountability, and school achievement monitoring demonstrates that to a significant degree the required infrastructure exists, though further development is clearly needed in many locations around the world.

The reviewed examples also suggest that next-generation tests are emerging. Such tests are beginning to use technology to change what is being assessed through the introduction of constructed-response item types and by considering skill in using the computer as an object of assessment in its own right. Additionally, next-generation tests are pursuing efficiency improvement by using technology for generating items automatically, scoring constructed responses, and connecting more directly with test consumers.

Although no operational third-generation tests exist, some fundamental characteristics of these envisioned measures have appeared in instructional products and national studies. These characteristics include basing assessment design on cognitive principles, repeated sampling of performance over time, use of complex simulations, and tasks intended to constitute learning events.

Several important issues were noted. Mentioned with respect to first-generation tests was the need to evaluate comparability claims empirically when both paper and electronic versions are offered, variation in the quality of infrastructure for some populations and testing purposes, and the fact that some students may not be able to take tests effectively on computer for reasons of skill or comfort. For next-generation tests, of particular concern was the use of undisclosed processes for automated scoring or of processes that were not designed in keeping with construct considerations. Finally, with respect to the envisioned Generation R, a host of issues was apparent. These issues include integrating cognitive principles with content-based domain descriptions, aggregating information gathered from repeated samplings over time, scoring complex tasks, making large-scale assessment comprising such tasks affordable, and convincing teachers and policy-makers to spend more time on assessment.

Bibliography

- Anderson, J. R. (1993). *Rules of Mind*. Hillsdale, NJ: Erlbaum.
- A timeline of the history of the MBA and the GMAT (2004). In McLean, V. A. (ed.) *Graduate Management Admission Council*. <http://www.gmac.com/selections/2004/timeline.pps> 280,25,Slide25 (accessed August 2009).
- Bennett, R. E. (1998). *Reinventing Assessment: Speculations on the Future of Large-Scale Educational Testing*. Princeton, NJ: Policy Information Center, Educational Testing Service. <ftp://ftp.ets.org/pub/res/reinvent.pdf> (accessed August 2009).
- Bennett, R. E. (2006). Moving the field forward: Some thoughts on validity and automated scoring. In Williamson, D. M., Mislevy, R. J.,

- and Bejar, I. I. (eds.) *Automated Scoring of Complex Tasks in Computer-Based Testing*, pp 403–412. Mahwah, NJ: Erlbaum.
- Bennett, R. E., Braswell, J., Oranje, A., et al. (2008). Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *Journal of Technology, Learning and Assessment* 6(9). <http://escholarship.bc.edu/jtla/vol6/9/> (accessed August 2009).
- Bennett, R. E., and Gitomer, D. H. (2009). Transforming K-12 assessment: Integrating accountability testing, formative assessment, and professional support. In Wyatt-Smith, C. and Cumming, J. (eds.) *Educational Assessment in the 21st Century*, pp 43–61. New York: Springer.
- Bennett, R. E., Persky, H., Weiss, A. R., and Jenkins, F. (2007). *Problem Solving in Technology-Rich Environments*. (NCES 2007-466). Washington, DC: National Center for Education Statistics, US Department of Education. <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007466> (accessed August 2009).
- Ben-Simon, A. and Bennett, R. E. (2007). Toward more substantively meaningful automated essay scoring. *Journal of Technology, Learning and Assessment* 6(1). <http://escholarship.bc.edu/jtla/vol6/1/> (accessed August 2009).
- Bridgeman, B., Lennon, M. L., and Jackenthal, A. (2003). Effects of screen size, screen resolution, and display rate on computer-based test performance. *Applied Measurement in Education* 16, 191–205.
- Department of Education and Virginia Information Technologies Agency (2005). A statewide web-based standards of learning technology initiative. Richmond, V. A.: Author http://www.pen.k12.va.us/VDOE/Technology/soltech/full_report.pdf (accessed August 2009).
- Educational Testing Service (ETS) (2008). *iSkills Overview*. <http://www.ets.org> (accessed August 2009).
- Gallagher, A., Bridgeman, B., and Cahalan, C. (2000). *The Effect of Computer-Based Tests on Racial/Ethnic, Gender, and Language Groups* (ETS RR-00-08). Princeton, NJ: ETS. ftp://ftp.ets.org/pub/gre/gre_96-21p.pdf (accessed August 2009).
- Graf, E. A., Petersen, S., Steffen, M., and Lawless, R. (2005). *Psychometric and Cognitive Analysis as a Basis for the Design and Revision of Quantitative Item Models* (RR-05-25). <http://www.ets.org/Media/Research/pdf/RR-05-25.pdf> (accessed August 2009).
- Horkay, N., Bennett, R. E., Allen, N., Kaplan, B., and Yan, F. (2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *Journal of Technology, Learning and Assessment* 5(2). <http://escholarship.bc.edu/jtla/vol5/2/> (accessed August 2009).
- Keith, T. Z. (2003). Validity of automated essay scoring systems. In Shermis, M. D. and Burstein, J. C. (eds.) *Automated Essay Scoring: A Cross-Disciplinary Perspective*. Mahwah, NJ: Erlbaum.
- Kirsch, I., Jamieson, J., Taylor, C., and Eignor, D. (1998). *Computer Familiarity among TOEFL Examinees* (RR-98-06). Princeton, NJ: Educational Testing Service. <http://www.ets.org/Media/Research/pdf/RR-98-06.pdf> (accessed August 2009).
- Mead, A. D. and Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin* 114, 449–458.
- National Assessment Agency (NAA) (2008). *Report on the 2007 Key Stage 3 ICT Test Pilot*. London: National Assessment Agency. http://testsandexams.qcda.gov.uk/libraryAssets/media/Key_stage_3 ICT Pilot 2007_report.pdf.
- Schaeffer, G. A., Steffen, M., Golub-Smith, M. L., Mills, C. N., and Durso, R. (1995). *The Introduction and Comparability of the Computer Adaptive GRE General Test* (RR-95-20). Princeton, NJ: Educational Testing Service. <http://www.ets.org/Media/Research/pdf/RR-95-20-Schaeffer.pdf> (accessed August 2009).
- Schleicher, A. (2006). *Are Students Ready for a Technology-Rich World? What PISA Studies Tell Us*. Paris: Organization for Economic Co-operation and Development (OECD). <http://www.pisa.oecd.org/dataoecd/28/53/35996337.ppt> (accessed August 2009).
- Singley, M. K. and Bennett, R. E. (2002). Item generation and beyond: Applications of schema theory to mathematics assessment. In Irvine, S. and Kyllonen, P. (eds.) *Item Generation for Test Development*, pp 361–384. Mahwah, NJ: Erlbaum.
- State: Online testing helped raise scores (2005). *eSchool News Online*. <http://www.eschoolnews.com/news/showStoryts.cfm?ArticleID=5826> (accessed August 2009).
- Wainer, H. and Eignor, D. (2000). Caveats, pitfalls, and unexpected consequences of implementing large-scale computerized testing. In Wainer, H. (ed.) *Computerized Adaptive Testing: A Primer*, 2nd edn., pp 271–299. Mahwah, NJ: Erlbaum.
- Wells, J. and Lewis, L. (2006). *Internet Access in U.S. Public Schools and Classrooms: 1994–2005* (NCES 2007-020). Washington, DC: National Center for Education Statistics, US Department of Education. <http://nces.ed.gov/pubs2007/2007020.pdf> (accessed August 2009).

Further Reading

- Bennett, R. E. (2001). How the Internet will help large-scale assessment reinvent itself. *Education Policy Analysis Archives* 9(5). <http://epaa.asu.edu/epaa/v9n5.html> (accessed August 2009).
- Bennett, R. E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *Journal of Technology, Learning, and Assessment* 1(1). <http://escholarship.bc.edu/jtla/vol1/1/>.
- Bennett, R. E., Morley, M., and Quardt, D. (2000). Three response types for broadening the conception of mathematical problem solving in computerized tests. *Applied Psychological Measurement* 24, 294–309.
- Drasgow, F., Luecht, R. M., and Bennett, R. E. (2006). Technology and testing. In Brennan, R. L. (ed.) *Educational Measurement*, 4th edn. pp 471–515. Westport, CT: American Council on Education/Praeger.
- Ericsson, P. F. and Haswell, R. (eds.) (2006). *Machine Scoring of Student Essays: Truth and Consequences*. Logan, UT: Utah State University Press.
- Irvine, S. and Kyllonen, P. (eds.) (2002). *Item Generation for Test Development*. Mahwah, NJ: Erlbaum.
- Measures of Academic Progress (2006). <http://www.nwea.org>.
- Mills, C. N., Potenza, M. T., Fremer, J. J., and Ward, W. C. (eds.) (2002). *Computer-Based Testing: Building the Foundation for Future Assessments*. Mahwah, NJ: Erlbaum.
- Shermis, M. D. and Burstein, J. C. (eds.) (2003). *Automated Essay Scoring: A Cross-Disciplinary Perspective*. Mahwah, NJ: Erlbaum.
- Williamson, D. M., Mislevy, R. J., and Bejar, I. I. (eds.) (2006). *Automated Scoring of Complex Tasks in Computer-Based Testing*. Mahwah, NJ: Erlbaum.
- Zenisky, A. L. and Sireci, S. G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education* 15, 335–362.

Technology Supports for Assessment Design

R J Mislevy and F I Winters, University of Maryland, College Park, MD, USA

I I Bejar and R E Bennett, ETS, Princeton, NJ, USA

G D Haertel, SRI International, Menlo Park, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Adaptive testing – Testing in which tasks are administered based on an examinee's previous responses, in order to improve the efficiency of measurement.

Assessment argument – An explication of the rationale for making inferences about an examinee's knowledge or skill, based on the observation of a sample of performances in an assessment.

Assessment delivery – The processes by which assessment tasks are selected to administer, examinees respond, responses are evaluated, and scores are reported.

Assessment implementation – The processes of constructing the elements of an operational assessment, such as authoring tasks, specifying scoring rubrics, and calibrating the psychometric model.

Classical test theory – A probabilistic model that posits observed test scores to be the sum of a true score and a random measurement error.

Cognitive diagnosis models – The psychometric models that characterize examinees in terms of their proficiencies on multiple aspects of knowledge and skill, that are typically required in combination to carry out complex task performances. These models are used in some applications of diagnostic assessment.

Conceptual assessment framework (CAF) – In the evidence-centered design approach to assessment, the CAF builds specific models for use in a particular assessment, taking into account the specific purposes and requirements of that assessment. The primary models of the CAF are the student model, the task model, and the evidence model.

Constructed response – In contrast to selection tasks such as multiple-choice and true–false, constructed response tasks require an examinee to produce an answer such as an essay, a proof, a diagram, or a performance.

Construct-irrelevant variance – The variance in scores among examinees due to differences in knowledge or skill that are not central to the knowledge and skills that are intended to be

assessed. For example, if it is meant to assess chemistry knowledge with an essay test, skills in composition can influence scores in ways that do not reflect differences in the chemistry knowledge.

Construct-relevant variance – The variance in scores among examinees due to differences in knowledge or skill that are central to the knowledge and skills that are intended to be assessed.

Design pattern – An assessment design pattern systematically provides advice about how to construct tasks, capture performances, and score work in some skill area, such as model-based reasoning or building scientific explanations.

Diagnostic assessment – An assessment that seeks to assess, and report in terms of, multiple specific elements of knowledge or skill required to perform in a specified domain. For example, a cognitive diagnosis analysis of mixed-number subtraction test would report on an examinee's mastery of finding a common denominator, adding fractions with like denominators, converting a mixed number to an improper fraction, etc.

Domain analysis – In the evidence-centered design approach to assessment, domain analysis encompasses the processes of researching and cataloging aspects of the world relevant to an assessment activity. It includes expert–novice studies, cataloging representations, surveying curricular materials, etc.

Domain modeling – In the evidence-centered design approach to assessment, domain modeling encompasses the process of building assessment arguments at a narrative level, in terms of the knowledge or skill to be measured, the kinds of performances required, the features of tasks, and the rationale connecting them.

Evaluative submodel – In the evidence-centered design approach to assessment, the evaluative submodel of the evidence model specifies how examinees' work on the task is scored.

Evidence model – In the evidence-centered design approach to assessment, the evidence model specifies how examinees' performances will be evaluated and combined across tasks. It contains evaluative and statistical submodels.

Instantiation Variables – A class of variables such that the class name itself can be used as if it were the automatic instantiation variable. This provides a mechanism for accessing default instances of the class.

Intelligent tutoring system – A computer system that dynamically adapts learning content to objectives, needs, and preferences of a learner.

Item forms – The frameworks for authoring families of test items, with a common form and slots that are filled in with different content in order to produce distinct items. In common usage, item forms are usually meant for producing relatively simple tasks, such as arithmetic items.

Item models – The frameworks for authoring families of test items, with a common form and slots that are filled in with different content in order to produce distinct items. In common usage, item models are usually meant for producing more complex tasks such as architectural design problems.

Item response theory – A probabilistic model that models observed responses to test items as functions of parameters for examinee characteristics such as ability and item characteristics such as difficulty.

Knowledge representation – A knowledge representation denotes the notation or formalism used to code information. Maps and graphs are examples of widely used knowledge representations. Test specifications and structural equations diagrams are more specialized knowledge representations that are used in test design and analysis.

Layers of assessment – The distinguishable kinds of activities that are carried out in different stages in the design and conduct of assessments. The layers in the evidence-centered design (ECD) assessment framework are domain analysis, domain modeling, CAF, assessment implementation, and assessment delivery.

Objects – In object-oriented computer programming, objects are structures that consist of data, procedures carried out on the data, and messages that can be sent to other objects. In assessment systems, examples of objects are tasks, work products, scoring algorithms, and task design patterns.

Schema – A schema is an organized structure that describes some event or set of relationships at a general level, and has slots that can be filled in to describe particular instances.

Scoring procedures – Scoring procedures are involved in the identification, extraction, summarization, and evaluation of key features of performance in student work. In particular, scoring procedures are used to synthesize nuggets of information in the form of values of observable variables or tasks.

Statistical submodel – In the evidence-centered design approach to assessment, the statistical submodel of the evidence model specifies how scores across tasks are combined to produce scores. This may be a simple total score or a complex algorithm defined through a psychometric model.

Stem – For simple assessment tasks, the stem is the question or prompt to which the examinee responds. In multiple-choice tasks, for example, the stem can be a question, which is followed by a small number of possible answers from which the examinee chooses.

Student model – In the evidence-centered design approach to assessment, the student model contains variables that characterize the knowledge and skills of the examinee.

Task design wizard – A task design wizard is a computer program that guides a task designer through choices, which when made, create an assessment task. A task designer wizard is built around a task model.

Task model – In the evidence-centered design approach to assessment, a task model is a schema that contains descriptions of elements and interactions that define a class of assessment tasks, and variables which, when specified, provide the information needed to implement a specific task.

Task-model variables – In the evidence-centered design approach to assessment, the task model variables indicate systematic ways that tasks following the same task model can differ from one another. Specifying the values of the task-model variables provides the information necessary to implement a specific task.

Web-based adaptive testing – The administration of a test over the Web, where the selection of the next test question to be presented to a student and the decision to end the test are performed dynamically by a computer based on a student profile which is created and updated through interaction with the system. Web delivery enables this efficient testing procedure to be carried out virtually anywhere in the world, and to draw upon the rich array of media and learning environments in creating tasks and valuating performances.

Introduction

Educational assessment involves making inferences about what students know or can do, based on observing what they say, do, or make in a handful of particular circumstances (i.e., assessment tasks). (Task encompasses short self-contained items that comprise familiar educational tests, but performances which may be longer, more, complex, or interactive, such as troubleshooting a computer network or carrying out a scientific investigation. Assessments consisting of tasks such as these also need to be designed and constructed around assessment arguments.) Designing an assessment requires analyzing the domain, building an argument, and creating tasks. These processes require gathering, organizing, and transforming information in a variety of representational forms. At each stage, technology can help organize work, enhance validity, and increase efficiency. This article discusses roles of technology supports for assessment design. It begins by reviewing assessment arguments and layers in the design process.

Arguments, Layers, and Knowledge Representations

An active line of research in assessment is making explicit the assumptions and principles that underlie educational assessment so that tools can be developed to make assessment design both more efficient and more consistent in quality. Examples of such work include Embretson's (1998) cognitive assessment design system and Luecht's (2002) integrated test design, development and delivery. The present article uses the language of a third such framework, Mislevy *et al.* (2003b) evidence-centered design (ECD) approach. ECD is an approach to assessment that uses a principled framework that is applicable to many forms of assessment.

The central ideas in ECD are the assessment argument, layers of assessment, and the role of knowledge representations. Messick (1994: 16) orients assessment designers to the key aspects of an assessment argument by asking, "what complex of knowledge, skills, or other attributes should be assessed. . . Next, what behaviors or performances should reveal those constructs, and what tasks or situations should

elicit those behaviors?" These questions lead to formal specifications in terms of student, evidence, and task models (see **Figure 1**), which will be discussed in detail in the section titled 'Supports for the CAF'.

Design typically begins with a purpose and some ideas about what is important in the domain, and then it moves to the specific materials and processes of an operational assessment. The ECD organizes the design process in terms of the following layers: domain analysis, domain modeling, conceptual assessment framework (CAF), assessment implementation, and assessment delivery (Mislevy and Riconscente, 2006). **Table 1** summarizes these layers in terms of their roles, key entities, and knowledge representations that assist in achieving each layer's purpose. The layering suggests a sequential design process, but cycles of iteration and refinement across layers are the norm.

The fundamental work in assessment design can be viewed as creating, transforming, and using information in the form of knowledge representations within and between such layers.

Supports for Domain Analysis

Domain analysis is the first step in the process of designing and delivering an assessment. It entails gathering information about how people acquire, construct, represent, use, and communicate knowledge within the domain. It lays the foundation for later layers by defining the knowledge, skills, and abilities (KSAs) assessment users want to make inferences about, student behaviors they can base their inferences on, and situations that will elicit those behaviors. This process is relevant to assessments of all kinds, whether formative or summative, large-scale or classroom.

Technology supports can be used in domain analysis to aid in defining and gathering the domain information, then to organize it in ways that will inform or facilitate the design of assessment tasks. One well-studied example of a technology tool that streamlines domain analysis is Shute *et al.* (2000) automated knowledge elicitation tool DNA (for Decompose, Network, Assess). The DNA provides structured, user-friendly web forms to elicit domain experts' input on declarative-, procedural-, and conceptual-knowledge requirements of common tasks in the domain. Other examples include Williams's (2000) Cognitive

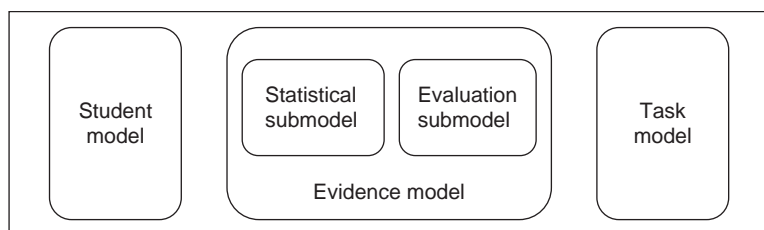


Figure 1 The central models of the conceptual assessment framework.

Table 1 Layers of evidence-centered design for educational assessments

<i>Layer</i>	<i>Role</i>	<i>Key entities</i>	<i>Selected knowledge representations</i>
Domain analysis	Gather substantive information about the domain of interest that has direct implications for assessment: how knowledge is constructed, acquired, used, and communicated.	Domain concepts, terminology, tools, knowledge representations, analyses, situations of use, patterns of interaction.	Content standards, concept maps (e.g., <i>Atlas of Science Literacy</i> , AAAS, 2001); representational forms and symbol systems of domain of interest, such as, algebraic notation, maps, computer interfaces.
Domain modeling	Express assessment argument in narrative form based on information from domain analysis.	Knowledge, skills, and abilities; characteristic and variable task features, potential work products, potential observations.	Assessment argument diagrams, design patterns, content-by-process matrices.
Conceptual assessment framework	Express assessment argument in structures and specifications for tasks and tests, evaluation procedures, measurement models.	Student, evidence, and task models; student model variables, observable variables, and task model variables rubrics; measurement models; test assembly specifications.	Test specifications; algebraic and graphical representations of measurement models; PADI task template; item generation models; generic rubrics; algorithms for automated scoring.
Assessment implementation	Implement assessment, including authoring presentation-ready tasks, scoring guides or automated evaluation procedures, and calibrated measurement models.	Task materials (including all materials, tools, affordances); pilot test data for honing evaluation procedures and fitting measurement models.	Coded algorithms for rendering tasks, interacting with examinees, evaluating work products; tasks as displayed; computer-deliverable representations of materials; American Standard Code for Information Interchange (ASCII) files of item parameters.
Assessment delivery	Coordinate interactions of students and tasks: task- and test-level scoring; reporting.	Tasks as presented; work products as created; scores as evaluated.	Renderings of materials; numerical and graphical summaries for individuals and groups; results files.

PADI, Principled Assessment Designs in Inquiry.

Analysis Tool – Human Computer Interaction (CAT-HCI) software tool for eliciting experts' domain knowledge, and from CHI systems for analyzing and subsequently modeling cognitive tasks (Zachary *et al.*, 2000). More specialized tools can be employed for specific purposes. Technology-based tools can be used to capture information about experts' knowledge that would be difficult to capture with more traditional methods of task analysis. For example, in the domain of satellite-image analysis, eye trackers have been used as part of cognitive task analyses (Kurland *et al.*, 2005).

Supports for Domain Modeling

Domain modeling structures the outcomes of domain analysis in a form that reflects the structure of an assessment argument, in order to ground the more technical student, evidence, and task models that are required in the subsequent, more technical, CAF layer shown in **Figure 1**. One tool for this purpose is design patterns, a concept that originated in architecture (Alexander *et al.*, 1977), and was later adapted in computer science (Gamma *et al.*, 1995). As

defined by Alexander *et al.* (1977), a design pattern provides a “description of a problem which occurs over and over again in our environment, and then describes the core of the solution to that problem, in such a way that you can use this solution a million times over” (p. x). Design patterns for assessment compile knowledge about ways to address assessment challenges that recur across domains or within particular domains (Mislevy *et al.*, 2003a). Examples of design patterns include model revision and building a scientific explanation. Each makes explicit the knowledge and skills that the developer wants to measure, the kinds of observations that could provide evidence about the acquisition of the knowledge and skills, and features of task situations that allow the examinee to provide the requisite evidence. Hively *et al.* (1968) proposed a type of design pattern called item forms for assessing behavioral objectives. Item models (e.g., Bejar *et al.*, 2003) and item structures (Embretson, 1998) were developed to incorporate an accounting of task features from the perspective of their cognitive-processing demands.

Technology supports can facilitate the domain-modeling process by helping designers create, search, or tailor design

patterns. In Principled Assessment Designs in Inquiry (PADI), design patterns help designers bridge the educational goals (in the form of standards, learning objectives or knowledge, skills, and abilities) with the technical design specifications that will be realized in an operational assessment.

Embodying a validity argument in an assessment requires more than authoring individual assessment tasks. The assessment designer must judiciously use certain design patterns in ways and in combinations that best support the purpose of the assessment as a whole (see, e.g., Davidson and Lynch (2001), on test specifications). Many considerations, including resources, operational constraints, content representation, and statistical information, impact on which design patterns and how many can be used to produce an optimal design. Optimization tools that have been developed for test assembly (van der Linden, 2005) can be used to sort through competing sets of design patterns that would lead to an assessment satisfying an array of constraints, including the adequacy of coverage of KSAs or educational standards.

Supports for the CAF

The CAF concerns technical specifications for the nuts and bolts of assessments, that is, the materials and processes that embody assessments. The CAF contains the student, evidence, and task models as shown in **Figure 1**. An assessment argument that was laid out in narrative form in the domain-modeling layer is now expressed in terms of specifications for tasks, measurement models, scoring methods, and delivery requirements. Details about task features, measurement model parameters, stimulus material specifications, and the like are expressed in terms of representations and data structures that will guide their implementation and ensure their coordination. Technology supports can be applied with each of the CAF models to facilitate task designers' work. Moreover, knowing the supports that are available at the level of the CAF helps shape the designer's work in domain analysis and domain modeling. The kinds of supports that are available to implement and carry out assessments thus define the universe of assessments that can be conceived.

Student Model

Scores from assessments help synthesize evidence from students' performances, in terms of variables organized around KSAs. These are student-model variables. Each variable corresponds to some aspect of knowledge, skill, ability, proficiency, etc., presumed to drive probabilities of observable response. Psychometric models such as classical test theory, item response theory, and cognitive-diagnosis

models use probability-based methods to ground inferences about students and characterize their precision. Especially in applications requiring multivariate models, designers will find it useful to use technology-based tools to construct modular components of the required student models and assemble them for diagnostic assessment and intelligent tutoring systems, such as van Lehn's Andes physics tutor (Conati *et al.*, 2002). These reusable student-model pieces are combined with the measurement-model pieces described below, allowing assessment designers to create tasks that support detailed feedback without having to create complicated measurement models anew for each task.

Reusability of measurement models is of particular importance for not only tasks requiring multivariate models, but also those with multiple scores, dependencies, or modeled relationships between task features and measurement parameters (Rupp, 2002). The psychometric implications of these relationships usually lie outside the expertise of domain specialists who create assessment tasks. However, assessment task authors can be provided with a library of preconfigured measurement fragments around which to write any number of unique assessment tasks (Almond *et al.*, 2002). When these kinds of supports are available, designers can link task models to student- and evidence-model fragments, so that the designers are not constrained to assessment arguments that simply address overall proficiency in the kinds of tasks that comprise the assessment.

Evidence Model

The evidence model provides a mechanism for connecting observable student behaviors to student-model variables. The evidence model has two components, herein called the evaluative and statistical submodels but referred to elsewhere by other names, including evidence extraction and evidence synthesis (e.g., Williamson *et al.*, 2006).

The evaluative submodel involves extracting the features that are relevant to a given student-model variable from observable behavior (or some work product), and then judging those features in terms of the evidence they provide about the KSAs of interest. In simple situations, this may just be correctness. In tasks where examinees produce more complex performances, this may be efficiency of problem-solving steps, appropriateness of cohesive devices in an essay, or time-on-target in a training exercise. This extraction-and-judgment operation can be performed by machine, by human rater, or by some combination of the two. In the case of a multiple-choice test, the operation is straightforward because the features are represented by the penciled marks on a scannable answer sheet, and judging the correctness of each mark is governed by the match between the mark's position and an answer key. For a constructed response, the rater (human or automated)

must locate relevant features and then judge the quality of each feature. Again, knowing what technologies will be available to support evaluation determines the range of performances a designer will note in domain analysis, and it determines the range of possibilities for capturing and evaluating evidence that will be available for constructing assessment arguments in domain modeling. Williamson *et al.* (2006) provide in-depth discussion of automated methods for evaluating complex performances, from the perspective of ECD.

Various commercially used technology tools exist for feature extraction and judgment. For example, the Online Scoring Network (OSN; Odendahl, 1999) represents a class of tool that displays the student's constructed response (or a digitized image of it) on a computer screen for the human rater to process. Such tools must be able to aid the human rater in extracting and judging features relevant to the assessment argument and avoiding ones tangential to it. For example, OSN allows the rater to highlight and annotate portions of a digitized response as a means of organizing relevant evidence. A compatible kind of technological support for human scoring provides underlying communication and database capabilities that allow scorers at different locations to evaluate work products, and that allow assessment designers to monitor, analyze, and improve raters' performances (Whalen and Bejar, 1998).

Other technology tools are capable both of automatically extracting relevant features from constructed responses and judging their adequacy. Examples of tools for evaluating essay responses include e-rater (Burststein, 2003), the Intelligent Essay Assessor (Landauer *et al.*, 2000), and Project Essay Grade (Page, 2003). The features such tools extract and judge, however, must also be aligned with the ones assessment designers intended as evidence for informing standing on student-model variables (Bennett, 2006).

The statistical submodel is the second component in the evidence model. The statistical submodel provides a mechanism for combining the judged features in a principled manner, in terms of a score or value for one or more student-model variables. Ideally, this synthesis would include an indication of how much evidence should be associated with each resulting score, as provided by item parameters in item response theory models or weights in summed scores. Dependencies among the several features of a complex performance can be difficult to model. However, as with student models, technology supports make it possible to construct appropriate statistical models from pre-constructed modules in computer-based assessment (Rupp, 2002), so that a task designer can build unique assessment tasks, guided conceptually by design patterns, around configurable evidence structures. In this way, technology supports based on an ECD or similar unified assessment framework can improve the designer's work in terms of both conception and implementation.

The statistical submodel may additionally incorporate multiple features of a complex performance, to produce an intermediate item-level score. This is common in automated essay scoring. Regardless of whether the aggregation is within items, across items, or both, the aggregation should be done in a manner that is consistent with the assessment argument. For example, the use by some automated scoring programs of such brute-empirical methods as step-wise regression undermines the assessment argument because the features selected and the weights assigned may diverge from one task and examinee sample to the next, as well as from the judgments of writing experts (Bennett, 2006; Bennett and Ben-Simon, 2006).

Task Model

During domain analysis the assessment designer identifies potential tasks to elicit the relevant student variables. A task model is a more detailed structure that includes information about how the information it elicits is related to other components of the assessment, and it serves as the blueprint for instantiating actual tasks to be presented to the student. Item forms and item models, mentioned as forms of domain modeling, can serve as the basis for task models by augmenting them with task-model variables. A simple example is an item form for two-digit addition without carrying: $ab + cd$, where a , b , c , and d are digits between 0 and 9, and $b+d < 9$. The task-model variables are a , b , c and d . Alternatively, for tasks that call for complex responses, a task template (Riconscente *et al.*, 2005) is a means of reaching the definition of a task model. In either case, a task model is a schema with variables that are needed to make the task model come alive, so to speak.

Figure 2 shows task-model variables in context. The left portion of **Figure 2** shows the flow of information at assessment time. A student is presented a task and then produces a work product such as a written response, a graphical design, a mathematical formula, or a response log. The work product is then analyzed to extract response variables (evidence extraction) and possibly to aggregate them (evidence synthesis). The result is a set of observable variables that characterizes the work product in a manner consistent with the statistical model that updates the student-model variable(s) via a suitable measurement model. **Figure 2** also shows some possible task-model variables. The variable shell refers to a template that contains details about the task as it appears to the student. The content variables indicate key features of stimulus material, domain focus, knowledge requirements, and the like, which can be used to select tasks or assemble them into tests. **Figure 2** assumes that a task model is the basis for producing instances (i.e., particular tasks) and, therefore a subset of the task-model variables are designated as instantiation variables, the values of which are used to create the specific tasks. If automated

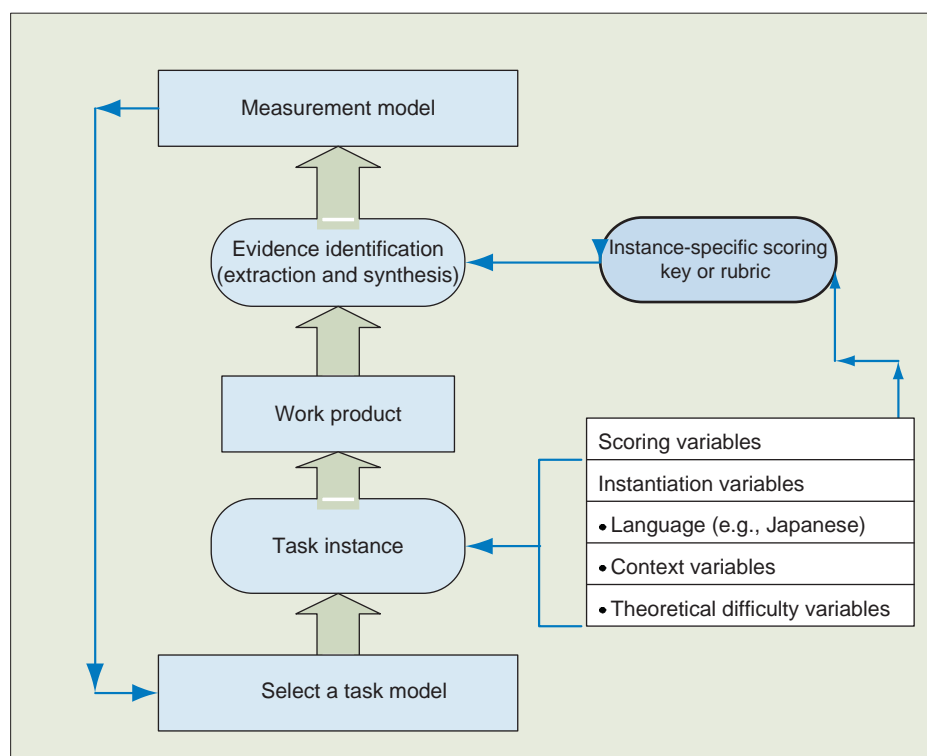


Figure 2 Schematic showing two roles of task-model variables.

scoring is being used, these variables may be also used by the scoring procedure (Braun *et al.*, 2006). Alternatively, if scoring is done by judges, these variables inform the scoring rubric for each specific instance of the task model. **Figure 2** also assumes that the domain analysis and modeling have been sufficiently thorough that a difficulty model is available to estimate or impute the difficulty of specific task-model instances (e.g., Enright and Sheehan, 2002). Finally, **Figure 2** assumes that the language in which a task is rendered is a variable (see Higgins *et al.*, 2005). A relational database can be designed to hold all the information for tasks. The challenge in any particular project is to enable queries that are relevant to assessment design and to display the results in a form amenable to design reasoning, for example, in the form of a task design wizard (e.g., Hamel and Schank, 2006).

Supports for Implementation

Task Authoring Support Tools

Task-authoring tools help developers create tests more efficiently. Many such tools exist, especially for authoring multiple-choice and other selected-response items; banking them so that they can be easily retrieved; and assembling them into tests. Such tools as Questionmark Perception incorporate item templates that facilitate paper as well as on-screen formatting, permitting a full preview of the

layout as the item will appear to the examinee. However, in addition to improving efficiency, task authoring tools ideally should support the assessment argument by channeling the developer toward some types of implementation and away from others. Tools, that in some way incorporate considerations found important in domain analysis can be particularly helpful in this regard.

An instructive example is SourceFinder (Passoneau *et al.*, 2002; Sheehan *et al.*, 2006), which assists test developers with one particular aspect of the item authoring process. SourceFinder searches online repositories (e.g., databases of literary and scientific journals) to help the test developer more quickly locate appropriate passages for graduate admissions-level reading comprehension questions. The tool uses natural-language processing techniques to characterize each text selection in terms of critical features. The critical features were chosen because of their relationship to the assessment argument, in particular, and their potential to influence passage difficulty in construct-relevant and irrelevant ways. For example, level of argumentation features ensure that text selections contain conflict, divergent ideas, uncertainty about outcomes, or other characteristics that can serve as the basis for questions that call upon the targeted verbal reasoning construct. Other features, such as detecting specialized jargon, center on reducing irrelevant variance. Echoing the theme of modularity, SourceFinder allows features to be added or removed as new natural language processing

methods emerge or as the assessment argument changes to accommodate, for example, a different test purpose or population.

Another example of a tool that incorporates considerations found important in domain analysis is the Mathematics Test Creation Assistant (TCA) (Singley and Bennett, 2002). The TCA helps task designers in two ways to make task authoring more efficient and more strongly connected to the research carried out in domain analysis and the argument construction carried out in domain modeling. First, TCA allows the test developer to create item models, abstract descriptions that are more general than specific test items, essentially constrained versions of task models. Working at this more general level of abstraction brings the test developer conceptually closer to the categories that comprise the domain analysis than would be possible by crafting each test question individually. Second task designers can author many instances of tasks from the item models.

Creating individual tasks can further be supported with technology assistance. An example of this type is Katz's (1995) Free-Response Authoring, Delivery, and Scoring System (FRADSS) task-authoring system, which permits test developers to construct computer-based items from objects. Each object brings with it capabilities that enable the item to behave in certain ways (e.g., present an animation), or the examinee to act upon it (e.g., draw a line, shade a portion of a figure, and move figures). The developer can create items from various combinations of objects, interact with them as would the examinee, revise them in real time, assemble a test, and deliver it in pilot form.

Automated Task Generation

Within the framework described above, task generation refers to instantiating an actual assessment task as a function of the variables that comprise the task model. Generation can be fully automated or partially manual depending on the complexity of the stem or stimulus material. In the manual case task specifications are instantiated rather than ready-to-deliver tasks. For example, assessment of architectural expertise by means of design problem solving is an instance requiring task specifications that are then used by subject-matter experts to finalize a task (Bejar, 2002). Skills such as reading comprehension call for a variety of task models, some of which can be instantiated automatically. For example, consider a format that asks the student to summarize a paragraph where difficulty is to a large extent a function of the attributes of the text to be summarized. Automating the instantiations of such a task requires a detailed text model that describes by means of a set of variables the attributes of the text, such as genre, lexical, syntactic, and other relevant textual attributes (Deane *et al.*, 2006). In addition, a statistical

model that predicts the likely difficulty of text supplies parameter values to the evidence model. The predicted difficulty for each potential text can be stored along with the text to facilitate retrieval. Retrieving candidate texts of a given difficulty from a corpus becomes a simple database query (Sheehan *et al.*, 2006).

In domains such as mathematics (e.g., Bejar *et al.*, 2003) or fluid intelligence (Embretson, 1999; Hornke, 2002; Newstead *et al.*, 2006) automated instantiation of task models has reached operational status. Math TCA, mentioned above, has also been used for authoring tasks from task models (see Graf *et al.*, 2005).

Supports for Delivery

Technology-rich delivery environments, such as simulations, are complex and expensive to create and to score. Complexity means that there are many opportunities for assessment designers to inadvertently do a good job measuring unintended KSAs and a bad job measuring the intended ones. The expense associated with creation and scoring means that recovering from bad design decisions will almost certainly be very costly and very possibly infeasible. Thus, the delivery environment should attempt from inception to reinforce the assessment argument by minimizing construct-irrelevant variance and maximizing construct-relevant variance. Information about relevant knowledge and ways to obtain it is best addressed early on, in domain analysis and domain modeling (Luecht, 2002). The System of Intelligent Evaluation using Tests (SIETTE) used in Europe in several projects incorporates this philosophy (Conejo, *et al.*, 2004); in addition to managing web-based adaptive testing SIETTE supports item models and multilingual tests.

The delivery platform used in the National Assessment of Educational Progress (NAEP) technology-rich environments project offers an example of technology support for delivery that addresses considerations from domain analysis and domain modeling (Bennett *et al.*, 2003). In this project, eighth-grade students were asked to use the computer for problem solving in a science context. The task involves conducting simulated experiments to discover the relationship between a set of quantities (the payload mass carried by a helium gas balloon and the altitude to which it can rise in the atmosphere). Student behavior in this environment provides evidence for several student-model variables, including a general problem-solving with technology variable, and two more specific ones, scientific inquiry and computer skills.

To minimize irrelevant variance associated with learning the environment, the tools employed for conducting experiments are explained in a brief tutorial and, also, utilize such common software conventions as dialog boxes and wizards. Further, the tools are organized around

a representation of the experimental process designed to reinforce what students are expected to do substantively (i.e., design experiment, run experiment, interpret results).

Evidence for the student-model variables is gathered from two sources, both of which are intended to contribute to construct-relevant variance. The first key source of evidence is the adequacy and completeness of the written description of the relationship between payload mass and altitude that the student provides. A second source of evidence is how the student uses the tools to arrive at that relationship. Tool use is relevant only because the tools, by design, provide conceptual and data structures to support the assessment argument. For example, the student's use of the design-experiment and run-experiment tools indicates whether the student has executed enough experiments – and covered the range of payload masses sufficiently – to support a defensible conclusion about the relationship between payload mass and altitude. Similarly, his or her use of the interpret-experiment tools indicates whether a table or graph was created that even includes the two variables relevant to solving the problem.

Conclusion

Designing an assessment entails many carefully considered steps, or layers. At each layer, technology supports can aid the thinking and building of the processes and elements that are needed to embody the assessment argument. They can make particular tasks more efficient, as well as open up new possibilities for tasks that were previously too expensive, difficult, or not even considered. However, the focus should not be on technology *per se*, but rather the ways in which it can support the process of designing an assessment so that it fully embodies the assessment argument. Thinking about assessment and technology supports from the perspective of assessment arguments, layers, and knowledge representations provides a method for understanding and evaluating the roles of current technology supports for assessment design not only as it exists today but also as it will evolve in the future.

See also: Technology and Formative Assessment; Technology for Large-Scale Assessment.

Bibliography

- Alexander, C., Ishikawa, S., Silverstein, M., et al. (1977). *A Pattern Language: Towns, Buildings, Construction*. New York: Oxford University Press.
- Almond, R. G., Steinberg, L. S., and Mislevy, R. J. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning, and Assessment* 1(5). <http://www.bc.edu/research/intasc/jtla/journal/v1n5.shtml> (accessed August 2009).
- Bejar, I. I. (2002). Generative testing: From conception to implementation. In Irvine, S. H. and Kyllonen, P. C. (eds.) *Item Generation for Test Development*, pp 199–218. Mahwah, NJ: Erlbaum.
- Bejar, I. I., Lawless, R. R., Morley, M. E., et al. (2003). A feasibility study of on-the-fly adaptive testing. *Journal of Technology, Learning and Assessment* 2(3), 4–29.
- Bennett, R. E. (2006). Moving the field forward: Some thoughts on validity and automated scoring. In Williamson, D. M., Mislevy, R. J., and Bejar, I. I. (eds.) *Automated Scoring of Complex Tasks in Computer-Based Testing*, pp 403–412. Hillsdale, NJ: Erlbaum.
- Bennett, R. E. and Ben-Simon, A. (2006). Toward theoretically meaningful automated essay scoring. Jerusalem, Israel: National Institute for Testing and Evaluation (NITE). <http://clickit.ort.org.il/files/upl/224004906/934404432.pdf> (accessed August 2009).
- Bennett, R. E., Jenkins, F., Persky, H., and Weiss, A. (2003). Assessing complex problem-solving performances. *Assessment in Education* 10, 347–359.
- Braun, H., Bejar, I. I., and Williamson, D. M. (2006). Rule based methods and mental modeling. In Williamson, D. M., Mislevy, R. J., and Bejar, I. I. (eds.) *Automated Scoring of Complex Tasks in Computer Based Testing*, pp 83–122. Mahwah, NJ: Erlbaum.
- Burstein, J. (2003). The e-rater® scoring engine: Automated essay scoring with natural language processing. In Shermis, M. D. and Burstein, J. C. (eds.) *Automated Essay Scoring: A Cross-Disciplinary Perspective*. Mahwah, NJ: Erlbaum.
- Conati, C., Gertner, A., and VanLehn, K. (2002). Using Bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction* 12(4), 371–417.
- Conejo, R., Guzmán, E., Millán, E., et al. (2004). SIETTE: A web-based tool for adaptive testing. *International Journal of Artificial Intelligence in Education* 14, 29–61.
- Davidson, F. and Lynch, B. K. (2001). *Testcraft: A Teacher's Guide to Writing and Using Language Test Specifications*. New Haven, CT: Yale University Press.
- Deane, P., Sheehan, K. M., Sabatini, J., Futagi, Y., and Kostin, I. (2006). Differences in text structure and its implications for assessment of struggling readers. *Scientific Studies of Reading* 10(3), 257–275.
- Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods* 3, 380–396.
- Embretson, S. E. (1999). Generating items during testing: Psychometric issues and models. *Psychometrika* 64, 407–433.
- Enright, M. K. and Sheehan, K. (2002). Modeling the difficulty of quantitative reasoning items: Implications for item generation. In Irvine, S. H. and Kyllonen, P. (eds.) *Item Generation for Test Development*, pp 129–158. Mahwah, NJ: Erlbaum.
- Gamma, E., Helm, R., Johnson, R., and Vlissides, J. (1995). *Design Patterns: Elements of Reusable Object-Oriented Software*. Upper Saddle River, NJ: Pearson.
- Graf, E. A., Peterson, S., Steffen, M., and Lawless, R. (2005). *Psychometric and Cognitive Analysis as a Basis for the Revision of Quantitative Item Models* (No. RR-05-25). Princeton, NJ: Educational Testing Service.
- Hamel, L. and Schank, P. (2006). A wizard for PADI assessment design, *PADI Technical Report 11*. Menlo Park, CA: SRI International.
- Higgins, D., Futagi, Y., and Deane, P. (2005). Multilingual generalization of the Model Creator software for math item generation (RR-05-02), *ETS Research Report*. Princeton, NJ: Educational Testing Service.
- Hively, W., Patterson, H. L., and Page, S. H. (1968). A "universe-defined" system of arithmetic achievement tests. *Journal of Educational Measurement* 5, 275–290.
- Hornke, L. F. (2002). Item generation models for higher cognitive functions. In Irvine, S. H. and Kyllonen, P. C. (eds.) *Item Generation for Test Development*, pp 159–178. Mahwah, NJ: Erlbaum.
- Katz, I. R. (1995). FRADS: A system for facilitating rapid prototyping by end users. In Anzai, Y. and Ogawa, K. (eds.) *Proceedings of the Sixth Annual International Conference on Human-Computer Interaction*. Amsterdam: Elsevier.
- Kurland, L., Gertner, A., Bartee, T., Chisholm, M., and McQuade, S. (2005). Using cognitive task analysis and eye tracking to understand imagery analysis. *Mitre Corporation Technical Report #05-1365*. Bedford, MA: Mitre Corporation.

- Landauer, T. K., Laham, D., and Foltz, P. W. (2000). The intelligent essay assessor. *IEEE Intelligent Systems* **15**, 27–31.
- Leucht, R. M. (2002). From design to delivery: Engineering the mass production of complex performance assessments. *Paper Presented at the Annual Meeting of the National Council on Measurement in Education*, New Orleans, LA.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher* **23** (2), 13–23.
- Mislevy, R. J., Hamel, L., Fried, R. G., et al. (2003a). Design patterns for assessing inquiry in science. *PADI Research Report #1*. Menlo Park, CA: SRI International. http://padi.sri.com/downloads/TR1_Design_Patterns.pdf (accessed August 2009).
- Mislevy, R. J. and Riconscente, M. (2006). Evidence-centered assessment design: Layers, concepts, and terminology. In Downing, S. and Haladyna, T. (eds.) *Handbook of Test Development*, pp 61–90. Mahwah, NJ: Erlbaum.
- Mislevy, R., Steinberg, L. S., and Almond, R. G. (2003b). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives* **1**(1), 3–62.
- Newstead, S. E., Bradon, P., Handley, S. J., Dennis, I., and Evans, J. S. B. T. (2006). Predicting the difficulty of complex logical reasoning problems. *Thinking and Reasoning* **12**(1), 62–90.
- Odendahl, N. (1999). Online delivery and scoring of constructed-response assessments. *Paper Presented at the Annual Meeting of the American Educational Research Association*, Montreal, Quebec.
- Page, E. B. (2003). Project essay grade (PEG). In Shermis, M. D. and Burstein, J. C. (eds.) *Automated Essay Scoring: A Cross-Disciplinary Perspective*, pp 262–273. Mahwah, NJ: Erlbaum.
- Passoneau, R., Hemat, L., Plante, J., and Sheehan, K. M. (2002). *Electronic Sources as Input to GRE Reading Comprehension Item Development: Source Finder Prototype Evaluation* (RR-02-12). Princeton, NJ: Educational Testing Service. <http://www.ets.org> (accessed August 2009).
- Riconscente, M. M., Mislevy, R. J., and Hamel, L. PADI Research Group (2005). An introduction to PADI task templates. *Principled Assessment Designs for Inquiry (PADI) Technical Report 3*. Menlo Park, CA: SRI International.
- Rupp, A. A. (2002). Feature selection for choosing and assembling measurement models: A building-block-based organization. *International Journal of Testing* **2**, 311–360.
- Sheehan, K. M., Kostin, I., Futagi, Y., Hemat, R., and Zuckerman, D. (2006). *Inside Source Finder: Predicting the Acceptability Status of Candidate Reading Comprehension Source Documents* (RR-06-24). Princeton, NJ: Educational Testing Service. <http://www.ets.org> (accessed August 2009).
- Shute, V. J., Torrealano, L. A., and Willis, R. E. (2000). DNA: Toward an automated knowledge elicitation and organization tool. In Lajoie, S. (ed.) *Computers as Cognitive Tools: No More Walls, II*, pp 309–338. Mahwah, NJ: Erlbaum.
- Singley, M. K. and Bennett, R. E. (2002). Item generation and beyond: Applications of schema theory to mathematics assessment. In Irvine, S. and Kyllonen, P. (eds.) *Item Generation for Test Development*, pp 361–384. Hillsdale, NJ: Erlbaum.
- van der Linden, W. J. (2005). *Linear Models for Optimal Test Design*. New York: Springer.
- Whalen, S. J. and Bejar, I. I. (1998). Relational databases in assessment: An application to online scoring. *Journal of Educational Computing Research* **18**, 1–13.
- Williams, K. E. (2000). An automated aid for the conduct of a detailed cognitive task analysis for modeling human computer interaction performance. In Schraagen, R., Chipman, S. F., and Shalin, V. L. (eds.) *Special Series on Cognitive Task Analysis*, pp 165–180. Hillsdale, NJ: Erlbaum.
- Williamson, D. M., Mislevy, R. J., and Bejar, I. I. (2006). (eds.) *Automated Scoring of Complex Tasks in Computer-Based Testing*. Mahwah, NJ: Erlbaum.
- Zachary, W. W., Ryder, J. M., and Hicinbothorn, J. H. (2000). Building cognitive task analyses and models of a decision-making team in a complex real-time environment. In Chipman, S. F., Shalin, V. L., and Schraagen, J. M. (eds.) *Cognitive Task Analysis*, pp 365–384. Mahwah, NJ: Erlbaum.

Further Reading

- American Association for the Advancement of Science (2001). *Atlas of Science Literacy*. Washington DC: AAAS.
- Baker, E. L. (2002). Design of automated authoring systems for tests. *Proceedings of Technology and Assessment: Thinking Ahead Proceedings from a Workshop*, 79–89, Washington, DC: Published of Collection: National Research Council, Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education.
- Bejar, I. I. (2009). Model-based item generation: A review of recent research. In Embretson, S. E. (ed.) *Measuring Psychological Constructs: Advances in Model-Based Approaches*. Washington, DC: American Psychological Association Books.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin* **112**(3), 527–535.
- Messick, S. (1989). Validity. In Linn, R. L. (ed.) *Educational Measurement*, 3rd edn., pp 13–103. New York: American Council on Education/MacMillan.
- Mislevy, R. and Haertel, G. (2006). Implications of Evidence-Centered Design for Educational Testing. Menlo Park, CA: SRI International.
- Mislevy, R. J. and Haertel, G. D. (2006). Implications of evidence-centered design for educational assessment. *Educational Measurement: Issues and Practice* **26**(4), 6–20.
- Simon, H. A. (2001). Complex systems: The interplay of organizations and markets in contemporary society. *Computational and Mathematical Organization Theory* **7**(2), 79–85.
- VanLehn, K. and Niu, Z. (2001). Bayesian student modeling, user interfaces and feedback: A sensitivity analysis. *International Journal of Artificial Intelligence in Education* **12**(2), 154–184.
- Williams, K. E., Hultman, E. G., and Graesser, A. (1998). CAT: A tool for eliciting knowledge on how to perform procedures. *Behavior Research Methods, Instruments and Computers* **30**(4), 565–572.
- Wilson, M. R. (2005). *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Erlbaum.

Relevant Websites

- <http://www.andestutor.org> – ANDES physics tutor.
- <http://www.project2061.org> – Atlas of Science Literacy.
- <http://www.cognitiveagent.com> – Workshop Announcement: Getting Your Distance Learning Used Effectively. (iGEN Cognitive Agent Software).
- <http://www.marte.lcc.uma.es> – System of Intelligent Evaluation using Tests.
- <http://www.questionmark.com> – QuestionMark Perception.

TECHNOLOGY AND LEARNING – EDUCATION REFORM AND ECONOMIC DEVELOPMENT

Contents

National Strategies to Build a Technology Workforce

Public–Private Partnerships for Educational Reform

Relating Technology, Education Reform and Economic Development

National Strategies to Build a Technology Workforce

H-M Cheah, Nanyang Technological University, Singapore

© 2010 Elsevier Ltd. All rights reserved.

Glossary

IDM – This refers to collaborative media, including the digital format, which allow for active participation by the users of the media.

Introduction

Given the lack of any natural resources, Singapore recognized very early on that its national well-being is dependent on successfully building its human capital. The focus on education, both within the formal educational system as well as adult learning, has a clear link to economic perspective. This emphasis has been reiterated continually, and is likely to remain a key aspect of Singapore's development for the present and foreseeable future (Ng, 2002).

Increasingly, the development of Singapore's human capital needs to be seen in the larger international and globalized context. Singapore's past economic success has depended largely on its ability to provide excellent infrastructure and to establish conditions that attracted multinational corporations (MNCs) to build industrial capacity in the country. While these are likely to remain important for the foreseeable future, the knowledge-based and globalized environment will require the development of human capital that can not only effectively add value to these processes, but also lead in innovation and value creation.

Apart from strengthening local capability and capacity, building a successful education system that can produce top talent for a technology workforce can potentially act as a magnet for internationally mobile talents. It does this in three important ways. First, it provides a strong educational environment for talented children overseas, and this can be a key consideration for relocating to Singapore.

Second, foreigners educated in Singapore will be more likely to form important ties to the country, providing a basis for continual links and contributions to the economy. Third, the diversity injected into the education system through foreign participation in local schools can further strengthen the system by providing a platform for mutual learning and greater interactions among students from diverse backgrounds. These are important to the country, as the ability to attract talent that can contribute to the economic activities of the country is a key contributor to sustained economic growth.

In developing and shaping the education system to build a technology workforce, it is also useful to be conscious of the need to retain talent. Given the nature of the globalized environment, where structures and participation within a flat world (Friedman, 2007) is such that talents can move easily across national boundaries, it is important to maintain strong ties with these mobile talents. Thus, graduates from an education system should ideally be locally anchored and globally relevant. In other words, while the system develops the student holistically and equips him/her with important twenty-first-century skills and disposition, it is equally important to also develop a strong sense of belonging. This is not an easy outcome to achieve.

Singapore's Education and Economic Developments

To appreciate the development of Singapore's education from a largely developing world system to a developed world system in a short span of time, a review of the parallel educational developments, information and communication technology (ICT) plans and economic demands is useful, highlighting the strong link between education and economic imperatives. **Figure 1** places key education,

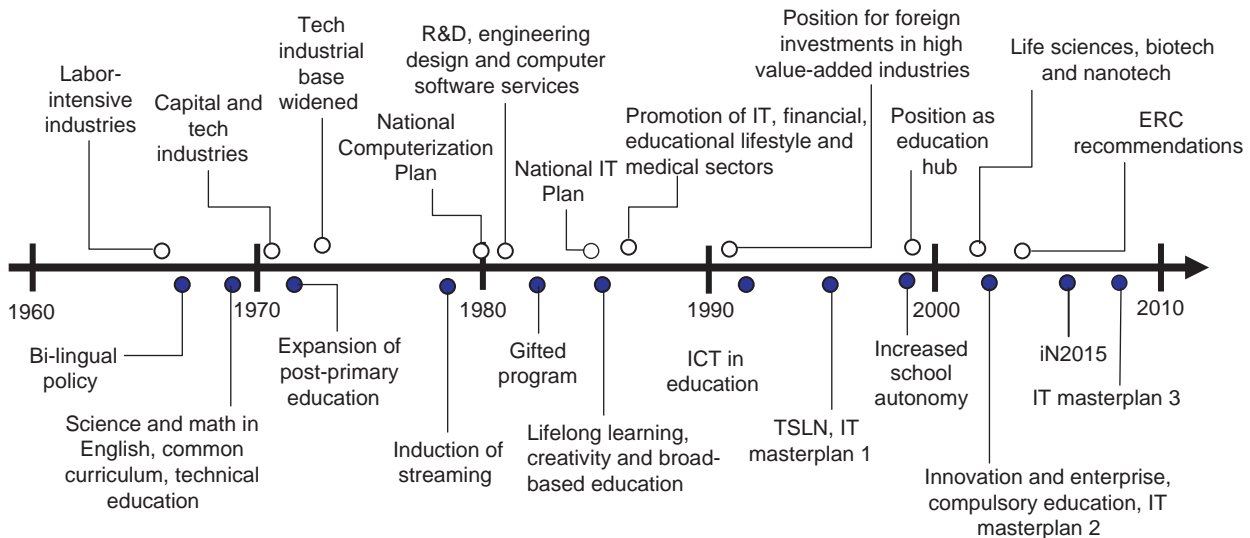


Figure 1 Parallel developments in education, manpower, and economic policies.

ICT, and economic initiatives along a chronological timeline from around the time when Singapore became independent in 1965 to the present. It charts the progress and shift in emphasis in these three arenas and how they interact and move in tandem.

The parallel developments may be broadly grouped into four stages with increasing alignment among educational development, manpower needs, science and technology policies, and global forces as the economy matures.

From early 1960s through to the early 1970s, Singapore, having become independent in 1965, was acutely aware of the need for economic survival. This was a period where there was a high dependency on the transfer of technology from foreign MNCs, and the focus was on exploiting available technologies rather than developing new ones. Although most children had access to primary education during this period, the adult literacy rate was only about 50% (Aldcroft, 2000). These factors prompted a concerted push to enhance the quality of education. To cater to the export-oriented manufacturing, there was a rapid expansion of primary and secondary education aiming to foster national identity and to produce a workforce with the necessary skills. It was also during this period that a bilingual policy was introduced, which gradually shaped up to require each student to learn English and one other official language. (Apart from English, the other three official languages are Chinese, Malay, and Tamil.) A common curriculum was introduced such that mathematics and science subjects were required to be taught in English. These two policies have enabled subsequent generations of Singaporeans to stay connected to the world economy as it grew and evolved; and to keep abreast of technological developments which facilitated their addition to the local capacity as the country gained traction in its economic push.

In the second stage of development, roughly from early 1970s to early/mid-1980s, Singapore oversaw a period of rapid industrialization where local manufacturing capabilities were developed through adapting and innovating acquired practices. The nature of the manufactured products also moved significantly up the value chain, becoming more complex and high-tech. To support such activities, and to lay the infrastructure to attract investment as well as to further develop capabilities, the National Computerization Plan was launched in 1980 to (1) computerize the major functions of every government ministry; (2) develop and facilitate the growth of a local information technology (IT) industry; and (3) develop the necessary IT manpower to meet the future needs of industry. After this plan had established a reasonable foundation, the second plan, entitled National IT Plan, was introduced in 1986 to shift the focus from the public to the private sector, specifically to integrate computing and communications via networking technologies to provide one-stop, fast, and efficient services. This stage of development, spanning about one-and-a-half IT plans, coincided with the consolidation of various education policies to develop greater coherence which served to prepare the workforce to later take advantage of the effects of globalization (Yip and Sim, 1990). For instance, a focus on mathematics and science was reflected in the expansion of higher education in local polytechnics and universities to produce the number of engineers and technicians who can support the buildup of high-tech manufacturing capabilities. The need to maximize the individual's potential led to the introduction of streaming within the education system based largely on academic ability. The gifted program for high-ability students was also established a few years later. The primary aim of these education policies was to provide a differentiated curriculum that can best

match the capability of students. Although these policies attracted a fair amount of discomfort in later years and probably increased the stress level of students, the positive outcome was a significant reduction in student attrition and a sharpened system that could cater better to individual needs.

The third stage, from about mid-1980s to mid-/end-1990s, represented further shift in the balance of industrial focus. This stage was sandwiched between two recessions, one in 1986 and the Asian financial crisis in 1998; and in both cases, it prompted important adjustments to the economic structure which helped the country pull out of these difficult periods relatively intact. Although the economy continued to put in place infrastructure that made it attractive for foreign investments, there was a shift toward high value-added industries and the development of an external economy that could be effectively linked to the domestic economy. Perhaps the most significant development was the realization that the economy cannot just rely on exploiting technologies developed elsewhere and transferring such capabilities to the resident population. While this remained a key aspect of Singapore's industrial engagements, the need for producing new ideas and technologies initiated a concerted phase of large-scale injection of funds for research and development (R&D), especially toward the mid-1990s. The National Science and Technology Board (NSTB) was set up in 1991 to oversee the development and implementation of integrated science and technology policies. This helped drive the first national technology plan which had a strong applied industrial focus. Important parallel structures were set up during this period to attract R&D activities into Singapore. These included the establishment of the Science Park next to the National University of Singapore (NUS) to stimulate R&D activities; and the Institute of Molecular and Cell Biology that signaled an important area of development.

The first ICT master plan for education (MP1) was launched in 1997 to tap the affordances of ICT for teaching and learning. This led to a systemic leveling up of ICT skills for all teachers, as well as putting in place a strong ICT infrastructure for schools. This was in line with the articulated wish to develop "Thinking Schools, Learning Nation" (Goh, 1997) and a culture of continual learning beyond formal education. On the education landscape was also the further provision of vocational and technical training at the secondary and postsecondary levels. Over and above all these, promotion of various identified R&D areas was supported by the encouragement given to students to undertake relevant courses of study at the university level. Greater refinements to the education system were also brought into being during this period, especially the increased porosity between different education tracks, allowing students to have greater flexibility and choice over subjects and academic levels in the different streams.

The fourth stage began from around the late 1990s with an economic plan that emphasized knowledge-based industry and technopreneurship akin to the spirit exemplified by Silicon Valley (Wong, 2001). At this point, the earlier focus on R&D was beginning to bear fruit as spin-offs from R&D institutions and venture capital increased noticeably. However, the increased uncertainties of the world economy prompted the need for constant innovation and the necessary speed to exploit the economic potential of any ensuing products. For this to be the case, it is important to increase the connectivity both within and outside national boundaries to facilitate collaborative and efficient development of ideas. To engage in the latest frontiers in various fields, a number of key research institutes were established, focusing on ICT, chemical and engineering sciences, electronics manufacturing, biomedical, and new media. These efforts were further consolidated through the setting up of the National Research Foundation (NRF) in 2006 to coordinate the research of various institutions situate within a larger national framework, and drive the national R&D agenda (NRF).

In line with these developments and to cultivate a mindset of lifelong learning, the Workforce Development Agency was formed in 2003 with an initial focus on continually upgrading low-skilled or unemployed workers. It also made available some S\$50 m for the training and development of private-sector professional upgrading in science and technology. On the education front, several initiatives were also introduced to lend greater coherence to the concerted effort. These include introducing the Teach Less, Learn More (MOE, 2007) initiative that focused on transforming teaching and learning practices so as to achieve engaged learning. Project work was introduced in grades 11 and 12 as assessable components that would count toward local university admission. These efforts were aimed at creating a more holistic schooling experience for pupils. At the same time, research institutions and centers continued to promote the learning of science and technology. For instance, there was a significant increase in the number of scholarships offered by agencies such as the Agency for Science, Technology and Research (A*STAR). As for teachers, various professional-development support and resources were made available, which included the setting up of a DNA center to cater to teachers teaching life sciences in schools.

Overview of Singapore's Education System

Singapore has largely followed a 6–4–2 structure (i.e., 6 years of primary, 4 years of secondary, and 2 years of preuniversity education) for grades 1–12 until recent years. For a better catering to the diversity in talents, and for providing more time for learning, greater flexibility was

injected into the system to equip it better to meet the diverse needs of the education and economic landscape. One important change was the introduction of integrated programs (IPs) that allow selected students to skip one high-stake national examination at grade 10. (These programs are available for students who have demonstrated sufficiently high academic ability such that they are most likely to gain entry to university education.) The time freed up from exam preparation can then be used by the schools to engage the students in a broader range of pursuits. Two variations of IPs are currently available, one that takes students in at grade 7 through to grade 12; and the other for students from grade 9 to grade 12.

Some degree of streaming into three different tracks is also done at the start of secondary education for a better matching of the capability of the students with the orientation and delivery of the programs. Successful completion of these programs allows students to either move on to preuniversity education or start postsecondary education. On the whole, about 66% of the primary-one cohort of students gain admission into institutes of higher education (~24% gets into universities and ~42% gets into polytechnics) (Ed Stats, 2008).

About 21% of the primary-one cohort, after grade 10, continue their studies in the Institute of Technical Education (ITE). This is a postsecondary institution that offers pre-employment training in various areas of specialization, including courses on automotive, electronics, hairdressing, and printing. It currently has ten campuses which will be consolidated into three regional campuses catering to 21 000 students by 2015.

There are five polytechnics in Singapore to provide vocational and professional preparation for students through its various diploma programs. Examples of these 3-year diploma courses include nursing, engineering, and financial studies. The polytechnics have established very strong links with industries, enabling them to adjust their programs to respond quickly to the changing skills demands. To provide greater options and variety within the education system, polytechnics could also explore providing degree programs in selected areas that are not traditionally provided by the local universities, such as in hospitality studies.

Three government-supported universities, namely, the Nanyang Technological University (NTU), the NUS, and one privately funded university, called the SIM University, cater to the bulk of local university education needs. In April 2006, NTU and NUS were granted autonomous status (SMU already had autonomous status when it was founded in January 2000) as not-for-profit educational corporations. Although the government continues to subsidize the fees of local students studying in the three universities, the autonomous status means that each university can set its own practices and operation guidelines and have control over the courses and programs it offers. They will also have the autonomy to shape their

respective research directions and develop the strategies to meet their goals. For instance, the universities have been able to offer specialized joint programs with other distinguished universities, such as the Massachusetts Institute of Technology (MIT), Stanford University, and Chicago School of Business, usually at the postgraduate level. The privately funded university, the SIM University, caters to adult learners.

In addition to the above universities, there are a number of foreign universities offering undergraduate and postgraduate degrees. They are normally affiliated with a local partner with a variety of delivery models, such as co-teaching between local and overseas academics. Most of the affiliated universities are from Australia, the UK, and the USA and the degrees are awarded by these foreign universities.

From IT Plans to Interactive and Digital Media Engagements

The development of both an enabling infrastructure and a workforce with the mindset and skills to undertake high-complexity jobs are central to Singapore's ability to compete effectively in the global environment. A key set of initiatives spanning almost three decades has provided the backbone for such developments.

In recognizing the importance of using ICT to enhance major aspects of business, educational, and administrative transactions, five national IT plans had been launched and implemented to support the needs of the country as its economy grew. After the completion of the first two national IT plans mentioned earlier in this article, the third, entitled IT200 (NCB, 1992), was started in 1992 to transform Singapore into an intelligent and wired island. This was envisaged to provide the sort of telecommunication infrastructure needed to attract cutting-edge economic activities to the country. It provided Singapore ONE, the first nationwide broadband network (NCB, 1997) to reach most households, all schools, and many public and private institutions. The fourth national IT plan, called Infocomm21, was Singapore's strategic plan for information and communication technologies (Infocomm) in the new economy (Infocomm21, 2000a). The vision of Infocomm21 was to "develop Singapore into a vibrant and dynamic global Infocomm Capital with a thriving and prosperous e-Economy and a pervasive and infocomm-savvy e-Society." To drive this plan, the Infocomm Development Authority (iDA) of Singapore was formed by merging the National Computer Board (NCB) and Telecommunications Authority of Singapore (TAS). In Infocomm21, it was recognized that the active participation of the private sector in taking the lead to come up with projects and enterprises was of crucial importance. As such, the intentions of the Singapore government were to play the facilitator and catalyst roles,

for instance, by reviewing policy or regulatory impediments with a view toward making them, where possible, compatible with a knowledge-based economy. At the same time, there would be concerted efforts to attract new international players onto the infocomm scene of Singapore, and to help local enterprises to regionalize and globalize (Yeo, 2000a, b).

Recognizing the increasing need to have greater coherence across various sectors in Singapore for any future infocomm efforts, the iDA led an effort in 2005 which brought the public, private, and people sectors together to develop a 10-year master plan to realize the potential of infocomm over the next decade. This resulted in the Intelligent Nation 2015 (iN2015) report which was published in June 2006 (IDA, 2006), and provided pointers to the way ahead for infocomm development. It laid out the three key themes of innovation, integration, and internationalization which support the vision of establishing an intelligent nation, a global city, powered by infocomm. This is perhaps the most comprehensive plan to date in terms of coverage and the tight alignment of key sectors, which are: digital media and entertainment; education and learning; financial services; government; healthcare and biomedical sciences; manufacturing and logistics; tourism, hospitality, and retail; infocomm infrastructure; services and technology development; enterprise development for Singapore-based companies; and infocomm manpower development. The report also exemplified the complex and interconnected

nature of the economy, and further emphasized the need to take holistic approaches to develop a technology-enabled workforce.

The environment characterized through the increasing presence and use of interactive and digital media (IDM) calls for the development of twenty-first-century skills, in particular, that of social participatory skills, where each person will need to be able to communicate and actualize effective collaboration. Where the IDM environment interacts with the formal learning sphere, one of the resulting effects would be that the learning outcomes, as desired by traditional instructions, would become uncertain. This is because the learner can easily connect to, and interact with, a host of other learning resources, environments, and information. It will thus be challenging for educators to evolve a system that can cater to the learning needs of learners operating within the IDM arena.

IT in Education Master Plans

The introduction of ICT into teaching and learning is seen to be an important element in developing a spirit of creativity and innovativeness among learners (Thaman, 2002). The integration of ICT into education began in the late 1970s when computers were introduced into the schools to raise the computer literacy of students and to support the teaching of computer science (Wong, 1989). Since then,

Table 1 Goals, programs, and achievements of the first two ICT master plans for education

<i>Goals</i>	<i>Main programs</i>	<i>Key achievements</i>
<i>1st ICT master plan in education (MP1)</i>		
<ul style="list-style-type: none"> ● Enhance linkages between schools and the world around them ● Generate innovative processes in education ● Enhance creative thinking, lifelong learning, and social responsibility ● Promote administration and management excellence in the education system 	<ul style="list-style-type: none"> ● Curriculum and assessment ● Content and learning resources ● Physical and technological infrastructure ● Human resource development 	<ul style="list-style-type: none"> ● All schools have basic IT provisions, including broadband ● Pupils performed well in international IT-based activities ● Preservice teachers received IT skills and pedagogies courses ● In-service teachers acquired basic IT skills and pedagogies ● Resource portal – Edumall
<i>2nd ICT masterplan in education (MP2)</i>		
<ul style="list-style-type: none"> ● Pupils use IT effectively for active learning ● Connections between curriculum, instruction, and assessment are enhanced using IT ● Teachers use IT effectively for professional and personal growth ● Schools have the capacity and capability in using IT for school improvement ● Active research in IT in education ● Infrastructure that supports widespread and effective use of IT 	<ul style="list-style-type: none"> IT in curriculum and assessment ● Professional development ● Capability and capacity building ● Research and development ● Infrastructure and support 	<ul style="list-style-type: none"> ● Established platforms for ICT in education innovation (Future Schools, Lead@ICT schools) ● Baseline ICT standards for pupils ● Pervasive use of ICT for teaching and learning ● Strengthen R&D in ICT in education (Learning Sciences Lab in NIE)

several initiatives and projects were implemented to further enhance the use of ICT in education. Although these were systematic efforts, the advent of the Internet brought about greater possibilities and demanded greater coherence in transforming teaching and learning practices. The first IT master plan (MP1) for education was launched in 1997, followed by the second master plan (MP2) in 2002, and, recently, the third in 2008. The key details of the first two IT master plans for education are summarized in **Table 1**.

The key success of MP1 was its ability to systemically put in place a strong infrastructure to support the use of ICT for teaching and learning for both the schools and institutes of higher learning (IHL). It has also ensured that all teachers were exposed to and being trained on key ICT skills and using ICT for teaching and learning. This laid the foundation upon which MP2 was built. During MP2,

which represented a period for seeding innovations for meaningful ICT use, schools were encouraged to undertake research activities and exploration to develop transformative ICT use in the classroom. These efforts were channeled through structures such as the Future Schools and LEAD@ICT schools initiatives. The third ICT master plan for education (MP3), which was announced in August 2008, will be building on the work of MP1 and MP2.

Looking Ahead: Human Capital Development

Similar to other developed nations, Singapore is facing the challenge of having to undergo a critical transformation of its economy in response to the demands of changing

Table 2 Selected recommendations of the Economic Review Committee (ERC) on enhancing human capital

<i>Empowering the individual</i>	School sector
	<ul style="list-style-type: none"> • Greater customization of learning through allowing secondary school students wider access to subjects beyond respective base stream • Broaden school curriculum at school level • More active and autonomous school governance through building more links with industry • Set up steering committee tasked with promoting successful arts and sports education • Present school info to represent holistic development
	Higher education sector
	<ul style="list-style-type: none"> • Increase university cohort participation rate to 30% in the long term • Expand postgraduate intake to facilitate knowledge creation • All polytechnics and universities to combine certain National Manpower Council (NMC) targets to enable them to respond to changing industry needs
<i>Enhancing the system at corporate level</i>	Lifelong learning through continuing education and training (CET)
	<ul style="list-style-type: none"> • Provide financial incentives to encourage greater participation in CET • Strong performance assessment system to encourage training provider to enhance quality and relevance of training. • Set up Centre for Adult Learning (CAL) to ensure that trainers are trained in adult pedagogy • Create one-stop system of decentralized centers for access to training programs/grants
	A global mindset
	<ul style="list-style-type: none"> • Develop Singapore as education hub • Increase exchange programs for more international exposure for Singapore students • Increase international exposure for teachers • More postings between civil service and offshore public service/private sector • Scale up presence of Singapore international schools
<i>Enriching the environment</i>	Developing human capital management (HCM)
	<ul style="list-style-type: none"> • Provide continuing education and training to equip human resource (HR) managers with business skills • Raise HR capabilities of business enterprises
	Upgrading manpower industry to raise HCM competencies
	<ul style="list-style-type: none"> • Establish center or network of research centers for HCM • Partner professional and trade associations for the training and development of professionals, managers, executives, and technicians
	Enhance worker employability
	Framework to attract, retain, and engage talent
	Mechanism to build collaborations between arts, sciences, and technologies
	Governmental commitment and support for arts, culture, sports, and recreation (ACSR)
	<ul style="list-style-type: none"> • Civil service to take lead to give greater emphasis and recognition to sports and cultural participation and achievements • Encourage public and private partnerships to develop and implement a package of programs to encourage public/private sponsorships
	Increased opportunities and public awareness to participate in ACSR activities

global conditions brought about, to a large extent, by the pace and nature of technological innovations. These changes will impact the job market, with the likely characteristics of an ever-shortening life-span for learned skills. The requirements of the country's human capital will thus have to be constantly reviewed and aligned to changing needs.

In the past, Singapore's growth was largely based on strong infrastructure support and MNC-led industrial development of local capabilities. Moving ahead, it is likely that future growth will depend on the workforce's ability to value-add to these processes in ways that focus on value-creation, innovation, and international cooperation. The economic review subcommittee on enhancing human capital (ERC, 2003) has identified three key areas on which to focus developmental efforts. Selected recommendations are summarized in **Table 2**.

It is clear that in attempting to develop a technological workforce, there is a need to adopt a whole-country approach that represents a concerted, cross-agencies effort. The work of the Economic Review Committee (ERC) was a wide-ranging review taking into consideration the constantly changing demands of the global economy and it attempted to make sufficient sense of it so as to shape up a series of recommendations that aimed to provide the Singapore workforce a cutting edge in competing globally. The focus on human capital is an inevitable consequence of the country's lack of practically any resources. The strong alignment among the parallel development of science and technology, education, economy, and manpower is a testament to a nation's efforts to forge ahead for continuing economic viability.

Bibliography

- Aldcroft, D. (2000). Education and development: The experience of the four little tigers. In Latham, A. J. H. (ed.) *Asia-Pacific Dynamism, 1550–2000*, pp 169–184. London: Routledge.
- Ed Stats (2008). *Education Statistics Digest 2008*. <http://www.moe.gov.sg/education/education-statistics-digest/online-interactive> (accessed September 2009).
- ERC (2003). *Economic Review Sub-Committee Report on Enhancing Human Capital*. <http://app.mti.gov.sg/default.asp?id=507#5> (accessed September 2009).

- Friedman, T. L. (2007). *The World is Flat 3.0: A Brief History of the Twenty-First Century*. Picador.
- Goh, C. T. (1997). *Speech by PM Goh Chok Tong at the Opening of the 7th International Conference on Thinking*, 2nd June 1997, Singapore.
- IDA (2006). *Reports of iN2015*. <http://www.ida.gov.sg/About%20us/20071005103551.aspx> (accessed September 2009).
- Infocomm21 (2000a). *Infocomm21 Report, Infocomm Development Authority of Singapore*. <http://www.ida.gov.sg> (accessed September 2009).
- MOE (2007). Ministry of Education's information on the teach less learn more initiative. <http://www3.moe.edu.sg/bluesky/tllm.htm> (accessed September 2009).
- NCB (1992). National Computer Board. *A Vision of an Intelligent Island: IT2000 Report, 29 October 2002*. <http://www.egov.gov.sg> (accessed September 2009).
- NCB (1997). National Computer Board. *Singapore One*. http://www.etw.org/2003/case_studies/reg_dev_singapore-one.htm.
- Ng, E. H. (2002). *Executive Summary of the Report of the Economic Review Sub-Committee on Enhancing Human Capital*. 28 October 2002. <http://app.mti.gov.sg/default.asp?id=507#5> (accessed September 2009).
- Thaman, S. (2002). IT in learning: Preparing for a different future. *Speech Delivered at iTopia 2002 on 24th July 2002 at Suntec City*, Singapore.
- Wong, P. K. (2001). From leveraging Multinational Corporations to fostering technopreneurship: The changing role of S&T policy in Singapore. In Low, L. and Johnston, D. M. (eds.) *Singapore Inc: Public Policy Options in the Third Millennium*, pp 35–84. Singapore: Asia Pacific Press.
- Yeo, C. T. (2000a). Developing Singapore into a Trusted Global Hub in the net economy. *Speech at the Infocomm21 Leadership Dialogue*, Singapore, 1st August 2000.
- Yeo, C. T. (2000b). Strategies to develop Singapore into an Infocomm Hub. *Speech Given at the Opening of COMDEX ASIA 2000*, Singapore, 5th April 2000.
- Yip, S. K. J. and Sim, W. K. (eds.) (1990). *Evolution of Educational Excellence: 25 Years of Education in the Republic of Singapore*. Singapore: Longman.

Further Reading

- Asia-Pacific Economic Cooperation (APEC) (2003). *The Drivers of New Economy in APEC: Innovation and Organizational Practices*. Singapore: APEC Secretariat.

Relevant Websites

- <http://www.ite.edu.sg> – Institute of Technical Education, Singapore.
- <http://app.mti.gov.sg> – Ministry of Trade and Industry, Government of Singapore.
- <http://www.nrf.gov.sg> – NRF: National Research Foundation's R&D Agenda.

Public–Private Partnerships for Educational Reform

V Quah, Singapore

C P Lim, Edith Cowan University, Perth, WA, Australia

C Brook, Curtin University of Technology, Perth, WA, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

There is growing belief that no single sector can effectively bring together the resources and capability to address or resolve the social and development issues we are facing (Googins and Rochlin, 2000; Tennyson, 2003). Traditionally, a key role of government has been the provision of the infrastructure required to sustain the economic activities of the people they govern. Governments fulfill this role through the collection of various taxes to develop and maintain public infrastructure and services. However, as populations grow, the cost of providing and maintaining these infrastructure and services increases; and, in many instances, the complexities associated with managing large public projects and programs become burdensome. Public–private partnerships (PPPs) – as possible mechanisms for developing and sustaining public infrastructure and services – have created growing interest from governments around the world (Moore, 2005). Such partnerships were identified at the 1992 United Nations Conference on Environment and Development – the Rio Earth Summit – as central to achieving global sustainable development.

Prior to engaging in any exploration of PPP, it is important to establish a common understanding of what is meant by the term. Public is used to refer to any service or infrastructure provided by the state whereas private refers to any endeavors that are associated with the private sector or nonstate sector. This includes both civil not-for-profit organizations and those more commonly associated with profit-related objectives (Latham, 2006). For the latter, they are usually corporations that have the resources and desire to undertake social responsibility. Partnership is used to refer to activities that require shared objectives, responsibilities, decision making, and risk. PPP then focuses on the shared responsibility of both the public and private sectors toward the provision and maintenance of infrastructure and services for the general population.

Despite the growing interest in PPP, there have been few peer-reviewed articles with regard to specific case studies of PPP and how they provide or fail to provide for infrastructure and services for the country. Such case studies are, especially, lacking in the area of education reform. Based on a collective case study of Microsoft's Partners-In-Learning (PiL) initiative, this article aims to first contribute to a better understanding of PPP in the

context of education reform from the perspective of a private sector partner, and then draw guidelines for PPP to bring about education reform.

Under this initiative, Microsoft establishes partnerships with ministries of education, national and local government bodies, and other public and private organizations to empower students and teachers to realize their full potential in learning and teaching, mediated by information and communication technologies (ICTs). The PiL initiative consists of three programs: the Learning Grants program, Fresh Start for donated computers, and School Agreement subscription licensing program. While the implementation of the PiL initiative is relatively recent (since 2003), it has made significant impacts on teaching and learning in schools. The impacts include the curriculum, professional development of school staff, school ICT culture, teacher use of ICT for teaching, and student use of ICT for learning. Although the approaches toward the implementation of the PiL initiative vary among the countries – and their levels of ICT-use in teaching and learning are different – there are many similarities in their experiences of implementation.

Models and Guidelines for PPP

There has been growing interest in PPP with conflicting opinions about their potential benefits and pitfalls. While there have been many successful projects made possible by PPP in the past, several academics caution against the pitfalls and highlight the importance of ensuring value for money of the projects, meeting the intended outcomes of the projects (Grout, 2005), and protecting the interests of the public (Moore, 2005). Despite these words of caution, the Academy for Educational Development (AED) describes PPP as untapped resources and suggests exploitation of such resources. In light of such growing interest in PPP among both the public and private sectors – and the considerable debate of the virtues of PPP – this section examines models and guidelines for PPP that may deliver the desired outcomes.

There are several ways to categorize the guidelines for PPP – those associated with cost–benefit analysis and those associated with project management. For the former, Hart, (2003) posits a mathematical model used to calculate the benefits of PPP versus totally public ownership.

For the latter, the National Association of State Chief Information Officers (NASCIO) suggests a model for project development that explores how projects may operate rather than to measure the benefits (NASCIO, 2006). This model is made up of eight key building blocks for effective PPP that include: (1) commitment from executive leadership; (2) a statutory foundation for partnering; (3) direct public sector involvement; (4) a well-crafted plan; (5) effective communication with stakeholders; (6) the right opportunity; (7) the right partner; and (8) well-defined management processes. These building blocks are consistent with the elements of effective project management outlined by Luca (1994) suggesting that models of effective PPP may resemble frameworks of project management. It has also been suggested that the signing of the memorandums of understanding (MOUs) is an important first step toward a public-private partnership (Inamdar, 2004).

In contrast, Davies and Hentschke, (2006) describe an exploration of two educational PPPs that yield an insightful four-stage model:

1. preconditions for partnerships where preconditions refer to the conditions that are required for the partnership to exist;
2. change dimensions that emerge as a result of partnerships where change dimensions refer to the types of changes that are pursued through the partnership;
3. partnering mechanisms that serve as means or vehicles binding the organizations to each other; and
4. success indicators of the partnerships that are the products and byproducts of the partnership which provide value to the partnering relationship.

Drew, (2004) describes a set of guidelines for projects attempting to bridge the digital divide. It includes the completion of a comprehensive needs analysis, completion of pilot studies, keeping costs down, development of multipurpose projects, and seeking funding assistance.

While each of these models and set of guidelines for PPP provides valuable insight, further exploration is required to ascertain the educational benefits associated with PPP, the conditions required for effective PPP, and the roles and responsibilities of each party in the PPP.

PPP, ICT, and Education

The primary motivation for using ICT in schools is that it supports students in their own constructive thinking, allowing them to transcend their cognitive limitations, and engages them in cognitive operations they may not have been capable of otherwise (Lim, 2007). Although master plans, national policies, and strategies of ICT in education exist in many countries, access to – and use of – ICT in classrooms within some countries and communities

are still very low (Plomp *et al.*, 2003). Such digital divides may widen the gap across communities in the areas of quality of life, competitiveness, and economic development. It is in this context of narrowing the digital gap that many governments – especially those in developing countries – have signed MOUs with private corporations that have the resources and the desire to undertake social responsibility (Mitra, 2003). Dangwal *et al.* (2005) provide an account of how a PPP seeks to bridge this gap and provide educational opportunity to those individuals typically associated with the disadvantaged. In this project, the government of a developing nation establishes a PPP with a private sector technology company to provide technical solutions to support the education of students who were irregular or nonattendees in traditional classrooms. While the paper does not describe the structure, operational procedures or the project management of the PPP, it outlines a project that has clear educational benefit for disadvantaged children.

This single case study provides insight into a PPP that has brought about significant educational advantage for those who are involved, and it also highlights the need for further projects to bridge the digital divide in ways that bring about significant educational reform. At the same time, there have been PPP that harness upon ICT to bring about break-the-mold school reforms. One such PPP is the Co-NECT model of school reform that is funded by New American Schools (NAS) – working together with private and public sector partners. The Co-NECT project emphasizes the development of students' higher-order skills through their work on extended, multidisciplinary projects. To provide a conducive environment for project-based learning, Co-NECT restructures the organization of students by forming them into clusters composed of students across grade levels. Co-NECT also provides professional development for the teachers and works closely with the principals to support the teachers. Although ICT is a pivotal tool in all curriculum areas, the Co-NECT model does not position ICT to be a reform in and of itself. In this model, ICT serves as a hub that holds together several aspects of a school reform effort (Russell and Haney, 1997).

Overview of the PiL Initiative

For about two decades, Microsoft has worked with schools and ministries of education to foster greater understanding of – and access to – ICT tools and training. This work has given the company a unique understanding of the challenges faced in different countries and cultures as they struggle to develop and fund ICT in educational programs. Consistent with the present data and assessments with regard to the growing digital divide within countries and communities, Microsoft has identified the three key issues:

1. *Under-utilized resources.* Although donated computers are economic necessity in many countries and communities, they contribute to accessibility problems due to outdated hardware, compatibility problems, and licensing constraints.
2. *Affordability.* The ultimate measure of ICT affordability is the cost of innovation and access, and the basic cost of ICT relative to *per capita* income.
3. *Know-how.* Teacher training and curriculum deployment are pivotal in addressing the economic, cultural, geographic, and physical barriers to ICT in education.

Microsoft launched the PiL in 2003 – a 5-year global initiative that targeted both developed and developing countries. The implementation of PiL was unique to the participating country and its culture. For developed countries, strategic investments promoted pedagogical advantage by training teachers to become fluent or elegantly confident with ICT in order to move less affluent students from basic to more sophisticated competencies. Developing countries – on the other hand – placed strategic emphasis on achieving significant technical competence among high-performing students as a priority for boosting their economies. Regardless of the strategic focus, government and schools require new partners and new resources for developing more competitive ICT skill-sets and capabilities in their countries.

The initiative comprises of three distinct – yet integrated – programs.

1. *Partners in learning grants.* The program seeks to partner with governments and education leaders to deliver local tools and resources for comprehensive ICT-skills training and curriculum leadership. With the goal of empowering schools and communities to improve student learning through school and teacher leadership, the program intends to deliver support in the form of teacher training and ongoing coaching/mentoring, student and teacher ICT-skills-assessment tools, high-quality courseware for teacher development, access to digital content, student certification and technical support, and research and measurement of success. Funding supports the establishment of local Microsoft Information Technology (IT) Academy Centers through joint partnerships with local community organizations, educational institutions, or local training providers. The Centers provide school leadership training, ICT-skills development and curriculum integration.
2. *Partners in learning fresh start for donated computers program.* The program seeks to remove licensing barriers that prevent primary and secondary schools from utilizing donated computers. It targets at donated machines received by schools (Pentium II or older) that are originally licensed with Windows. These schools can request

a copy of Windows 98 and/or Windows 2000 media from Microsoft. This ensures that schools can leverage their donated personal computers regardless of whether or not proof-of-license was provided with the donation.

3. *Partners in learning school agreement subscription.* This program provides extra support for the neediest schools where their ICT-acquisition budgets are constrained by other pressing and necessary social and economic investments. It attempts to reduce the initial cost of education desktop, reduce the cost of upgrading and maintaining the schools' current installed base and promote an easy-to-administer program for managing software investments over time. An MOU is signed with the national or local government where eligible schools can receive a free upgrade to Windows XP Pro for both new and current installed base PCs and/or to acquire Office XP Pro for both new and current installed base for a low price per license per year.

In 2003, the PiL initiative was initially launched in a mix of developed and developing countries including India, Brazil, China, Russia, Thailand, Canada, the United Kingdom, France, Germany, Taiwan, and Japan. For the following 3 years, the initiative has been rolled out to more countries. Access and affordability are central to the initiative where it facilitates ICT in education and ICT access based on community and country needs and priorities. The following section showcases a range of projects in both developing and developed countries – India, Brazil, and United States. As there is a greater variety of projects under the Learning Grants Program as compared to the other two programs, all the case studies discussed fall under this program.

The Collective Case Study

Case Study One: A Partnership with the Government of Maharashtra in Pre-Service Teacher Education

India is a large country with a population of more than a billion. She has the third largest economy in purchasing power and is the second fastest-growing economy in the world. Hidden beneath all the economic growth, India has been – and is still – battling the challenge of poverty and providing quality education for all. Bringing ICT into schools is one of the major challenges. In places where there is access to ICT infrastructure, the availability of trained teachers is a challenge. Under the Learning Grants Program or *Shiksha* as it is termed in India, Microsoft has partnered with ten state governments to train government schoolteachers in the use of ICT for teaching and learning. Since its launch 3 years ago, over 1 00 000 teachers have been trained and at least 5 million students

have been taught by these teachers. One of the state governments that has been involved in this program is the Government of Maharashtra. The partnership commenced, in 2005, with the setting up of three IT academies in Pune, Nagpur, and Aurangabad to train government schoolteachers.

Pre-service teacher education project in the state of Maharashtra

Equipping new teachers with a set of ICT competencies was a challenge identified by the Government of Maharashtra – there was a large number of pre-service teachers and an absence of ICT being integrated into the curriculum in the pre-service teacher-education program. To address these challenges, it was planned that at least 1 05 000 pre-service teachers and over 1000 teacher educators would be trained under a co-constructed process and system that have been set up by Microsoft, the Government of Maharashtra, and the teacher-education institutions. The focus of the project for teacher educators is on soft skills, such as peer-to-peer coaching and pedagogy strategies of incorporating ICT into lesson plans. As the teacher educators become more competent in the use of ICT in their teaching and learning, they serve as role models for the pre-service teachers in adopting ICT in their own classrooms. Such a process ensures the sustainability of the project; where the project is part of the Maharashtra Government's effort to draw upon the teaching and learning opportunities of ICT as an essential driver in the state educational system.

Expected impacts of pre-service teacher education project

This partnership will not only impact the present pre-service teachers and future teachers involved in the project but it will also sustain and scale-up the project to ensure that future generations of pre-service teachers in Maharashtra will undergo the pre-service teacher-education program with an ICT component. As a result, the pre-service teacher-education program in the state of Maharashtra may be used as a prototype to be customized to other teacher-education institutions in India and the rest of the world. Moreover, this project has the potential of changing the teaching and learning processes of the classrooms in the state of Maharashtra. It provides the pre-service teachers with the opportunities for ICT-mediated learning – and more important – equipped them with a set of competencies to integrate ICT in their lesson plans. With this set of ICT competencies among the new teachers, the pressure on the state government for providing in-service teacher education in ICT is eased.

Case Study Two: A Partnership with the State School Isaac Schraiber in a Reading and Writing Project

The State School Isaac Schraiber is located in an under-developed area of the city of São Paulo in Brazil. The literacy level of the students in the school has been low and the learning engagement of the students needs to be enhanced. The school principal has attended a series of ICT for School Leadership professional development workshops under the Learning Grants Program. With his newly acquired understanding of ICT and authentic learning experiences, he crafted a project on the Revitalization of the Cipoaba Stream. The idea was to empower the students to be actively involved in community issues and – as a result of such participation – engage them in reading and writing.

Reading and writing project in State School Isaac Schraiber

The project commenced, in 2004, and was developed by the principal, together with the staff of the school. It aimed at enhancing reading and writing competencies among the students by situating the learning tasks within the context of initiating the revitalization of the Cipoaba Stream – a stream that was located near the school. ICT was pivotal in the implementation of the project as an information and construction tool. To support the project, Microsoft provided both hardware and software necessary for the implementation of the project. It also provided ICT training for both the teachers and students. Based on the environmental problems created by the Cipoaba Stream, the students – at all levels in the school – researched on the Internet for newspaper articles, news clips, songs, and films concerning the related environmental problems. They debated on different environmental topics in the classrooms and created their own books under the supervision of their Portuguese teachers. Teams of students met in the library once a month to brainstorm and plan for actions to engage the community to address the environmental problems facing it.

Field studies to the stream were planned and carried out by the geography, history, mathematics, and physical education teacher to collect data and promote a better understanding of the Stream and the problems associated to it. Photos were taken during these field trips to make records of the area and create models to represent the stream in the past, present, and future. Window Movie Maker was then used to create movies by the students to raise awareness of the initiative to the community and encourage more proactive actions from all levels of the community.

Impact of the reading and writing project

As a whole-school approach was adopted, the teachers reported a significant enhancement of learning engagement

among the students who were participating in the project. The students were more task oriented during the lessons and were engaged not only in reading and writing, but more important, in digital literacy and ICT competencies. They could, effectively, use digital cameras, manipulate digital pictures, use Window Movie Maker, and represent their understanding through hypermedia narratives. The students were also more involved with social issues and showed deeper understanding about the environment. At the same time, the project managed to promote community actions that included the construction of a collector tube, the creation of an environmental park in the neighborhood, and the re-urbanization of irregular dwellings.

Case Study Three: A Partnership with the School District of Philadelphia in the School of the Future Project

The partnership between Microsoft and the School District of Philadelphia was agreed in 2003 to investigate, determine, and deliver a high school equipped to prepare learners for the twenty-first century. There was a meeting of two goals between the two partners: (1) Microsoft was considering investing in and creating a School of the Future (SOF) resource similar to the Home of the Future; and (2) Philadelphia has not built a new school in the last 40 years and the chief executive officer (CEO) was hired on a platform of capital expansion. In addition, hence, the decision was made by both parties to enter into a partnership to build a high school in Philadelphia.

School of the future project in Philadelphia

The goal of this project was to build and redefine the norm for 9–12 urban education. The result is a sustainable and replicable model that addresses functions of business and administration processes and educational practices. Innovation was achieved in five areas: technical architecture, instruction, community engagement, building design, and business management. The Microsoft project manager led the development of each of these innovations. The innovations developed were categorized into three areas: process, people, and environment. The first step for building the SOF was to create a process that would act as a framework and guide for the project. That framework was a 6i development process: introspection, investigation, inclusion, innovation, implementation, and introspection. Based on this process, a set of strengths – weaknesses – opportunities – threats (SWOT)-analysis tools and resources was developed due to the district's lack of strategic planning competencies. More important, training was provided and professional development resources to guide education planning were developed. At the same time, the partnership leveraged on Microsoft's approach to hiring and the start-up team of the SOF

went through training on behavioral interviewing and the loop to hire the principal and staff for the school.

The next area of innovation is people. The education competencies describe the full range of characteristics necessary to help a school district achieve its organizational goals and vision. They were developed in partnership between Microsoft, Lominger International, and school leaders from around the world. To ensure that the impact of this project extends beyond Philadelphia, numerous capacity-building strategies were in place to reach educational leaders around the world. They included the United States Quarterly SOF Briefings, where about 100 education leaders/teams from around the world participated in a 2-day workshop that introduced them to SOF resources and acquainted them with best practices that they might bring back to their organizations. The other professional development event was the SOF World Summit that brought together education elites to discuss and debate schooling for the future.

The third area is environmental innovations. These included the learning space matrix and the technical architecture design. Schools are constantly trying to identify ways that space can support instruction. Through research of appropriate characteristics of learning spaces, six characteristics have been identified for the spaces in the SOF building. For each space, documentation of how these characteristics have been reflected in the design was created in the form of the Learning Space Matrix. At the same time, utilizing the majority of the Microsoft Product suites, the SOF technical architecture provided a complete Learning Gateway Framework with role-specific access for students, parents, teachers, and administrators.

The SOF is a public high school for grades nine through 12. It is a neighborhood high school with no special admits. It has approximately 750 students and a comprehensive curriculum. Approximately US\$60 million was funded by the School District of Philadelphia while Microsoft contributed primarily human capital and development support.

Expected impact of the school of future project

The expected major impact of the project is to develop a school that will serve as a laboratory and model for the school of the future. In doing so, it aims to be replicable – not only throughout the School District of Philadelphia, but also districts nationwide, both large and small, and possibly internationally as well. It will also serve as a planning model of how the private and public sector work together to reform education and – in the long term – transform neighborhoods. The end result of this partnership will be a state-of-the-art educational institution incorporating the latest technology that will advance students' academic possibilities and be a conduit for career development. Some of the processes used in the setting up of the SOF – such as the hiring of staff process – have been

adopted by the district to hire all new principals. The District has been bombarded with inquiries about this project since its inception. Universities, technology corporations, pharmaceutical companies, and other large and small businesses have expressed their interest in this project. Universities and colleges are primed to assist the District with professional development, best practices, and educational research. The business sectors' resources are limitless and the project can tap on their expertise and insight in management, leadership development, and job training. In short, the SOF will be technology rich and a beacon of educational innovation.

Making Sense of the Case Studies: Guidelines for Effective PPP

Adoption of a Holistic Approach

Despite the investment to bring ICT into the classrooms, little impact has been observed on the roles of students and teachers, curriculum, and mode of assessment (Cuban *et al.*, 2001). Much too often, ICT has been introduced to replace traditional learning and teaching media or traditional method of delivery without any changes to the mode of assessment or the organization of the classroom or the running of the school. The learning media may have changed, but the practices and policies to integrate these new innovations remain constant. If nothing significant changes in the learning environment – save the introduction of ICT – few if any important effects can be expected. Therefore, a holistic approach toward the use of ICT in education needs to be adopted (Lim, 2007).

Based on the overview of the initiative and the case studies, it is clear that a holistic approach toward the use of ICT in education has been adopted by the PiL initiative. Besides providing access to working hardware and affordable software, the initiative worked on providing the facilitating factors for teachers and students to use ICT in teaching and learning. These facilitators included school-leadership and peer-coaching training, professional development of teachers, ICT curriculum for students, ICT competency standards for teachers, assessment of students, and sharing of expertise and experiences with partners. These factors were, often, interdependent. For example – without the school leadership training – professional development of teachers might not bring about a transformation of classroom practices as the teachers lacked a supportive school environment to put into practice their new set of competencies. Therefore, for PPP to be effective, the various parties in the partnership adopt a holistic perspective toward the development and implementation of the project, program, or initiative. However, it should be noted that such an approach will only work if all the partners have a good understanding of the sociocultural contexts of the school district or

classrooms they are working in; that is, a situation and needs analysis of the context that the project or program is situated in is absolutely crucial.

Situation and Needs Analysis

This phase consists of analyzing the teachers and students and their needs, the curriculum and mode of assessment, the resources, the school and educational system and their needs, and the community and its needs. The situation-and-needs-analysis phase was present in the three different programs of the PiL initiative and the three case studies. This was most obvious in the formulation of the initiative where the analysis phase was based on the vast experiences and data that Microsoft has due to its work with the education sector in the last two decades. This phase was also supported by literature review of ICT in educational research studies and data collection and analysis of existing educational practices and policies in the country involved in the initiative. It was with the identification of the three problem areas of under-utilized resources, affordability, and know-how that the three integrated programs of Learning Grants, Fresh Start, and School Agreement are developed. Therefore – for PPP to be effective – the situation and needs-analysis phase is carried out to provide the foundation for the project or program.

Piloting and Scaling-Up

Starting with a small diverse sample allows the consolidation of resources and efforts by the partners. This may serve as a piloting phase where best practices and lessons learnt are documented and barriers and potential problems are addressed. The PiL initiative commenced, in 2003, with a few countries – both developed and developing. The initiatives run in these countries may be likened to pilot testing the various programs in the initiative. The three case studies are also good examples. They serve as research sites where a prototype of the innovation is being developed, implemented, and evaluated. The prototype is not restricted to ICT and ICT-mediated practices, but it includes the conditions in which ICT can be used effectively. With the prototype revised and refined, the next phase is the scaling up of the innovation. Scalability is the ability of the innovation to achieve its expected outcomes as it grows; it depends on the possibility to implement the innovation at diminishing marginal cost where replicating the innovation requires less resources and yet produces similar or better outcomes. It has the additional benefits of accumulating knowledge, experiences, and shared resources (Hart, 2003). Therefore – for PPP to be effective – the piloting of the project or program provides the stepping stones for the scaling up of the innovation.

Ownership by Stakeholders

A common thread that emerges from both the review of contemporary literature referring to PPP and the exploration of the three PiL projects included in this article is the need of shared project ownership. In the context of educational systems, shared ownerships of PPP need to extend beyond the higher-level leadership of both the public and private sector partner to include the multiple stakeholders often associated with educational systems (Googins and Rochlin, 2000). Typically, this would include ownership by key stakeholders from centralized educational systems, education districts, schools, and the school communities. For example, the success of the project in the State School Isaac Schraiber was due to the whole-school approach adopted; it gave the teachers and students ownership of the project by empowering them to make decisions and act upon these decisions. Therefore – for PPP to be effective – the ownership of the project or program is shared among the partners and their major stakeholders.

Accountability and Sustainability

Accountability of schools has, often, been very narrowly defined in many countries where it is being associated to grades that are determined by students' performance on high-stakes tests. Although this may be the most no-fuss and economical approach, accountability works in different ways and at different levels. Based on the case studies, it is clear that accountability based on standardized test scores is different from accountability for providing the learning environment to cultivate social responsibilities, empower students to be autonomous learners, and involve communities in and around the school. In all projects and programs, the partners are accountable to the major stakeholders with respect to the expected outcomes. Although improvement in grades may be a pragmatic outcome for many schools, there is an urgent need to broadly define accountability to reflect the essence of education as the holistic development of a child (Darling-Hammond, 2004). In order to avoid misunderstandings, the partners in the PPP have to negotiate and define the accountability of the project or program to its major stakeholders.

Another important dimension of PPP is the sustainability of the innovation – that is, the ability of the innovation to persist over time. The processes and culture within the school or school district are pertinent to the sustainability of project or program. More often than not, there is a need for the partners to build the processes and shift the culture of the school or district during the intervention phase. It is only then that the innovation is most likely to become a permanent and integral part of the school. At the same time, sustainability encapsulates the spatial milieu as well, and it includes transplanting the innovation into spaces such as networks, virtual spaces, and social movement (Hargreaves, 2002). Therefore – for

PPP to be effective – accountability and sustainability are two issues to be negotiated and addressed among the partners. For the latter, it also involves the sustainability of the partnership.

Conclusion and Future Work on PPP

At present, efforts by many countries – especially developing and emerging nations – in education and educational reforms are progressing very slowly. This is compounded by rapid technological advancement that has resulted in a widening digital divide, and further disadvantaging for children and youths from the lower socioeconomic groups. Governments have been exploring PPP in an attempt to stem the flow of this negative cycle and to optimize the use of available resources. While there is conflicting evidence with regard to the past successes of the PPP approach to provision of public services and infrastructure, there is evidence that – under certain conditions – PPPs are effective.

The PiL initiative and its associated projects and programs that have been examined in this article have demonstrated how PPPs support educational reforms. There was evidence that the partnership with the State School Isaac Schraiber resulted in a significant enhancement of learning engagement among participating students. The partnership with the state of Maharashtra has the potential of changing the teaching and learning processes of classrooms teachers and the partnership with the School District of Philadelphia resulted in a model for the school of the future and served as a catalyst for increased community participation and the mobilization of resources. These case studies provide evidence of the benefits of PPP when a set of conditions is present to support the partnership in achieving its intended outcomes.

The review of the literature has revealed two PPP models – one that focuses on cost-benefit analysis and the other focusing on project management that includes roles and responsibilities. Based on the discussion of the three case studies and the initiative, there are several emerging issues that have commonalities with the two models. These commonalities suggest that the models are not mutually exclusive, but – in some ways – complementary. It suggests that a PPP has an increased likelihood of being effective when certain conditions are met. These conditions include:

- adoption of a holistic approach by the project or initiative;
- situation and needs analysis of the context of implementation;
- piloting and scaling up of project or initiative;
- shared ownership by all stakeholders of the project or initiative; and
- accountability and sustainability of the project or initiative.

The evidence presented in this article suggests that the present interest of public and private sectors in PPP as

a mechanism for educational reform is well founded. However, if these projects are to achieve the intended outcomes, there is a need to further explore the conditions required for successful implementation.

Bibliography

- Cuban, L., Kirkpatrick, H., and Peck, C. (2001). High access and low use of technology in high school classrooms: Explaining an apparent paradox. *American Educational Research Journal* **38**(4), 813–834.
- Dangwal, R., Jha, S., Chatterjee, S., and Mitra, S. (2005). A model of how children acquire computing skills from hole-in-the-wall computers in public places. *Information Technologies and International Development* **2**(4), 41–60.
- Darling-Hammond, L. (2004). Standards, accountability and school reform. *Teachers College Record* **106**(6), 1047–1085.
- Davies, B. and Hentschke, G. (2006). Public–private partnerships in education: Insights from the field. *School Leadership & Management* **26**(3), 205.
- Drew, T. (2004). Bridging the digital divide in the schools of developing countries. *International Journal of Instructional Media* **31**(1), 89.
- Googins, B. and Rochlin, S. (2000). Creating the partnership society: Understanding the rhetoric and reality of cross-sectoral partnerships. *Business and Society Review* **105**(1), 127–144.
- Grout, P. (2005). Value for money measurements in public private partnerships. *EIB Papers* **10**(2), 32–56.
- Hargreaves, A. (2002). Sustainability of educational change: The role of social geographies. *Journal of Educational Change* **3**(3–4), 189–214.
- Hart, O. E. (2003). Incomplete contracts and public ownership: Remarks, and an application to public–private partnerships. *Economic Journal* **113**, 69–76.
- Inamdar, P. (2004). Computer skills development by children using ‘hole in the wall’ facilities in rural India. *Australasian Journal of Educational Technology* **20**(3), 337–350.
- Latham, M. (2006). Private, public partnerships: Fact of fiction – what is the score? (Electronic Version). <http://www.cfbt.com/pdf/Michael%20Latham%20Paper%20PPPs%20Fact%20or%20Fiction.pdf> (accessed October 2009).
- Lim, C. P. (2007). Effective integration of ICT in Singapore schools: Pedagogical and policy implications. *Education Technology Research and Development* **55**(1), 83–116.
- Luca, J. (1994). *Project Management for New Media*. Perth: New Media Management.
- Mitra, S. (2003). Minimally invasive education: A progress report on the “hole-in-the-wall” experiments. *British Journal of Educational Technology* **34**(3), 367–371.
- Moore, M. H. (2005). Creating public value through private/public partnerships. <http://www.unpan1.un.org/intradoc/groups/public/documents/CLAD/clad0052203.pdf> (accessed October 2009).
- NASCIO (National Association of State Chief Information Officers) (2006). *Keys to Collaboration: Building Effective Public Private Partnerships*.
- Plomp, T., Anderson, R. E., Law, N., and Quale, A. (2003). *Cross-National Policies and Practices on Information and Communication Technology in Education*. Greenwich, CT: Information Age.
- Russell, M. and Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education Policy Analysis* **5**(3).
- Tennyson, R. (2003). *The Partnering Toolbook*. London: The International Business Leaders Forum (IBLF).

Relating Technology, Education Reform and Economic Development

R B Kozma, SRI International, California, San Francisco, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

The Economic Rationale for Educational ICT

Often policymakers employ an economic rationale in calling for increased technology in schools by claiming that the use of information and communications technologies (ICTs), such as computers, the Internet, and multimedia, will prepare students for the global economy and increase the nation's economic competitiveness. For example, in the 2001–03 technology master plan for Ireland, the minister for education and science claimed that the plan "... is vital to ensure that we acquire important skills needed to secure our future economic wellbeing" (n.d., p. 3). At the launch of Singapore's second ICT master plan, the minister of education claimed it would help students prepare for a more competitive future and a new, innovation-driven economic era (Ministry of Education, Singapore, 2002). In describing its technology-based program of educational reform, the Ministry of Planning and International Cooperation in Jordan (2004) claimed that "... the education sector must be responsive to employment market demands in key industries and develop critical 'Knowledge Economy skills' at all levels of the education system" (p. 1).

However, the logical chain that connects the introduction of computers in schools to resulting economic prosperity is rarely articulated by policymakers. This article examines the relationships among ICT, education, and economic development and explores the potential causal connections between the use of technology and economic development in the context of other corollary changes that may be needed in education if the use of technology is to contribute to economic improvement.

ICT, Education, and Economic Returns

There is some empirical support in economic studies for the policymaker's faith that ICT and education can contribute to economic growth. For example, ICT by itself has been an important contributor to economic growth over the past decade and a half (OECD, 2004a; Jorgenson and Vu, 2005). An increase of 2.5–3.0% in gross domestic product (GDP) during a year is usually considered a good rate of growth in developed countries. From 1995

to 2001, ICT by itself accounted for an increase of about 0.5% in GDP in the Organization for Economic Cooperation and Development (OECD) countries, with particularly strong contributions in the United States, Canada, the Netherlands, and Australia, amounting to about one-fourth of GDP growth (OECD, 2004a). Overall in OECD countries, ICT contributed about 16% of the growth in GDP during the first-half of the 1990s and increased to about 20% during the second-half of the decade.

Economists have also established that education is an important contributor to economic growth in both microeconomic studies – those that look at personal returns on investment – and macroeconomic studies – those that look at public returns on investment, including resulting increases in economic growth. Microeconomic data from 42 countries found that an average rate of return for an additional year of schooling was a 9.7% increase in personal income (Psacharopoulos and Patrinos, 2002). In a macroeconomic study, Barro (2000) found that there was an additional 0.44% growth in a country's per capita GDP for each additional average year of schooling attained, a return on investment of 7%. A review by Sianesi and Van Reenen (2002) found a return of 3–6% and a review by Stevens and Weale (2003) found returns that ranged from 6% to 12%.

Of particular interest is the finding of the Barro (2000) study that the quality of education had a stronger relationship to economic growth than did levels of education attainment; that is, the amount learned was more important than the number of years of schooling. Barro found that on assessments of science and math, national scores that were higher by one standard deviation equated to 1% growth in per capita GDP.

While macroeconomic studies show a positive relationships between ICT and economic development and between education and economic development, there is a lack of research analyzing the different components of the educational system – such as curriculum, assessment, teacher quality, or the use of ICT – in terms of their economic impact. Research of this sort is needed before a strong empirical argument can be made for economic benefits of educational ICT investments. But short of this research, the foundations for such an argument must be built upon economic theory and national case studies in which economic growth has corresponded to the implementation of educational ICT policies and programs.

Sources of Economic Growth

According to macroeconomists (see e.g., Stiglitz and Walsh, 2002; Krugman and Wells, 2006), growth in economic output can occur through one of two mechanisms. First, growth can occur as a result of increased input factors: the economy grows as more equipment is purchased and more workers enter the labor force – what economists call capital accumulation. However, at some point, growth based on capital accumulation is subject to diminishing returns – additional increases in input result in smaller and smaller increases in output. Sustained growth results from an increase in the economic output per worker, that is, an increase in productivity.

Economic theory describes three sources of increased productivity: capital deepening (the use of equipment that is more productive than earlier versions), higher-quality labor (a more knowledgeable workforce that is more productive), and technological innovation – the creation, distribution, and use of new knowledge.

Knowledge has some special qualities that are important to economists. Unlike raw material it can be used multiple times without depreciated value and unlike equipment, it can be used by many people at the same time. Knowledge can also be shared widely at little cost. These facts open up the possibility of a productivity factor with compounding rather than diminishing returns. That is, additional investments in knowledge result in new products that, in turn, facilitate the production of new knowledge that leads to continuous growth. These unique qualities of knowledge serve as the basis of what has come to be called the knowledge economy (Stiglitz, 1999).

Implications of Economic Theory for Education Reform and ICT Use

The three factors that lead to productivity-based growth (capital deepening, higher-quality labor, and new knowledge) serve as the basis for three complementary, somewhat overlapping models that connect education with economic development by:

- increasing the basic skills and technological uptake of the workforce (Birdsall *et al.*, 1995; Lall, 2000) – or the knowledge-acquisition model;
- increasing the quality of the workforce and their ability to use knowledge to add value to economic output (Hanusek and Kimbo, 2000; UNIDO, 2003) – or the knowledge-deepening model; and
- increasing the ability of the workforce to produce new knowledge and innovate (OECD, 2004b; Schliecher, 2006) – or the knowledge-creation model.

These three models have somewhat different implications for changes in each of the components of the educational system: policy goals, curriculum and assessment,

pedagogy, school organization and administration, and teacher professional development, as well as the use of ICT.

Knowledge-Acquisition Model

The policy rationale of this approach is to improve economic productivity by increasing the basic skills of the workforce and its capacity for taking up new technologies, thus enabling the deepening of physical capital. Corollary policy goals might include increased school enrolments, increased technology-literacy skills, and improved reading- and mathematics-literacy skills. The strategy might be implemented through policies such as increasing the level of compulsory schooling, mandatory testing, and school accountability, or the requirement of a larger number of courses in science and mathematics. Countries may also try to support these goals by integrating technology into the traditional educational system and making few changes otherwise (Law *et al.*, 2000; Schofield and Davidson, 2002; Schofield, 2006). Technologies involved in this approach may include the use of computers along with productivity software; drill and practice, tutorial, and web content (Kozma, 2003). Under the knowledge-acquisition model, changes in the curriculum may be limited to adding ICT as a subject or including time in the curricula of other subjects for the incorporation of ICT. The curriculum may otherwise continue to be divided by traditional subject areas and assessments that emphasize basic skills. Changes in pedagogical practice may involve the use of various technologies, tools, and e-content as part of whole-class, group, and individual student activities. Teachers may increase their digital-literacy skills and use technology for classroom presentations, for management tasks, and to acquire additional skills for their own professional development. Pedagogy may otherwise consist of the teacher delivering information that is received by students. Little change in social structure is required by this approach other than, perhaps, the spatial placement of hardware and technology resources in the classroom.

Knowledge-Deepening Model

As they entail more significant changes in the education system, the other two approaches to educational reform may have a more profound effect on economic growth and the standard of living. The policy rationale for the knowledge-deepening approach is to increase the ability of the workforce to add value to the economic output by applying a deep knowledge of school subjects, particularly science and mathematics, to solve complex problems encountered in real-world situations of work and life (Tucker, 1996). The traditional curriculum is bounded by school experience, and the focus is often on factual recall. But knowledge of school subjects may remain inert

and not be applied outside the class or contribute to productivity or innovation. To combat this problem, the knowledge-deepening approach supports economic productivity and social development by enhancing students' understanding of school subjects and making these subjects more relevant to the problems and challenges of the community and the workplace. Rather than the superficial coverage of a large number of topics, the curriculum focuses on a smaller number of key concepts, principles, and procedures and on how these ideas are organized and interconnected within and across subject areas to form complex knowledge systems (Bransford *et al.*, 2000). Classroom activities and projects that collaboratively engage students in the solution of extended, open-ended, real-world problems are an important component of the knowledge-deepening approach (Krajcik and Blumenfeld, 2006). Technology can also play an important role as students use visualizations and simulations to explore, understand, and apply complex knowledge (Edleson and Reiser, 2006; Lehrer and Schauble, 2006). Networking can help teachers and students connect classroom activities and learning to the outside world (Bruckman, 2006; Kozma, 2003). Aided by technology, extended assessments consisting of several parts can parallel the complex tasks that students will encounter in the real world (Pellegrino *et al.*, 2001; Mislevy, *et al.*, 2003; Means, 2006). As this type of learning is more complex, teachers need to understand both their subject area and the processes that students employ, and the problems that they face when they engage in deeper learning of subject matter (Darling-Hammond, 1997; Fishman and Davis, 2006). Teachers should be able to integrate open-ended software tools and subject-specific applications with student-centered teaching methods and collaborative projects in support of students' deep understanding of key concepts and the application of this knowledge to solve complex, real-world problems (Means and Olson, 1995; Means *et al.*, 2001; Schofield, 2006).

Knowledge-Creation Model

The policy rationale for the third approach – knowledge creation – is to build a workforce and citizenry that are continually engaged in and benefit from knowledge creation and innovation. The argument goes that if students are to participate in an economy and society in which the creation, sharing, and use of new knowledge are the basis for sustained development, their educational preparation must go beyond the learning of established knowledge (Bereiter, 2002). Knowledge creation does not conflict with knowledge deepening; rather, it can build on a base of deep understanding of school subjects – not only science and mathematics but also social sciences, humanities, and the creative arts. Beyond the learning of key concepts and principles and their use to solve

complex problems, students engage in the sustained, collaborative process of building on current knowledge to create and share new knowledge and cultural artifacts. Knowledge-creation skills include the ability to use a range of technology tools; to search for, organize, and analyze information; to communicate effectively in a variety of forms; to collaborate with others of diverse skills and backgrounds; and to think critically, innovatively, and creatively (Partnership for the 21st Century, 2003; New Commission on Skills of the American Workforce, 2006). But paramount among the knowledge-creation skills are those that allow students to continue their learning throughout their lifetimes – their ability to set their own goals, determine what they already know, assess their strengths and weaknesses, design a learning plan, stay on task, track their own progress, and build on successes and adjust to failures. These skills will enable students to sustain their own personal development and contribute to that of the economy and society in a constantly changing world. With this approach, teachers design a learning community in the classroom in which students are continuously engaged in building their own and each others' learning skills (Scardmalia and Bereiter, 2006). Teachers and students use a variety of networked devices, digital resources, and electronic environments to create knowledge communities for students and colleagues which support the process of continuous, reflective learning and the development of knowledge creation and critical-thinking skills (Scardmalia and Bereiter, 2006; Stahl *et al.*, 2006). The goal is to transform schools into learning organizations in which all actors are involved in the learning process (Elmore, 2004; Fullan, 2001a, b; Senge *et al.*, 2000) and teachers are master learners who are constantly engaged in educational experimentation and innovation in collaboration with their colleagues and outside experts to produce new knowledge about learning and teaching practice (Bransford *et al.*, 2005; McLaughlin and Talbert, 2001).

Examples of Technology, Education Reform, and Economic Development

Singapore and Finland provide two examples of how economic development, education reform and ICT policies can be connected in mutually supportive ways.

The Case of Singapore

Singapore has demonstrated impressive growth over the past 40 years, due in large part to its economic development policies (Anwar and Zheng, 2004; Blomstrom *et al.*, 2002; Castells and Himanen, 2002; Rajan, 2003). Singapore is an island city state of 4.3 million people in Southeast Asia. While Singapore was considered a poor country in

1960 with a per capita GNP of less than US\$5000 (in real US dollars), its per capita GNP in 2000 was US\$35 000, placing it among the countries with the highest standard of living in the world (Rajan, 2003). The government's initial economic development strategy focused on the development of physical and labor capital, that is, a strategy of growth based on capital accumulation. With a small population and land mass, Singapore had few competitive advantages. The government provided incentives to transnational corporations in targeted industrial clusters – such as consumer electronics – who located production facilities in their country. Singapore could thus tap into global value chains of these industries and with it their imported technology and knowledge. Using this approach, Singapore was able to grow its work force, its physical capital, and its economy; but indigenous companies did not participate in this growth. To sustain growth, the government shifted to productivity-based development approaches.

From the beginning of Singapore's modern history in the mid-1960s, there has been a strong connection between the country's economic development plans and its education plans (Ashton *et al.*, 2002; Brown and Lauder, 2000). Officials from the Ministry of Trade and Industry chair the Economic Development Board, a cross-ministry agency that sets directions for policies in other relevant ministries, including education. Initial plans in the 1960s moved from a low-skill, low-wage economic strategy (i.e., capital accumulation) to one of increased productivity based on universal primary education and an increase in literacy, numeracy, and English-language fluency that would provide the resulting workforce with the skills needed to take up the new technology offered by multinational companies (i.e., capital deepening). The government refined its strategy over time, leveraging initial gains in the economy to move from low value-added to high value-added manufacturing and services that required a better-trained workforce (i.e., high-quality labor). Concurrently, Singapore's education policy required secondary schools to produce higher levels of skills in science, mathematics, and language and tertiary institutions to produce more engineers and scientists. The improved quality of Singapore's education system resulted in top scores among all countries in both mathematics and science in both the fourth and eighth grades in the 2003 Trends in International Mathematics and Science Study (TIMSS), an international assessment of student achievement (Mullis *et al.*, 2004a, b).

In its current form, Singaporean economic policy aims at developing a knowledge-based economy that upgrades existing industrial clusters (such as consumer electronics) and promotes the development of new industrial clusters (such as microelectromechanical systems and nanotechnology) and new exportable services in areas such as healthcare, education, and creative industries. In coordination with this shift, the education ministry has instituted a

number of reforms under the title Learning to Think, Thinking to Learn: Towards Thinking Schools, Learning Nation (Ministry of Education, Singapore, 2000). An important component of this reform is to create a better balance between the acquisition of knowledge, the application of concepts, and the development of individual creativity and enterprise. Thus, Singapore's curriculum has been broadened to include information skills, thinking skills, and creativity, self-management skills, and character development. Cross-discipline project work has been introduced into classrooms and assessments revised to include measurement of students' ability to use information, think creatively, and communicate.

ICT has been an important component of Singapore's educational reform. In 1997, Singapore initiated a 5-year ICT plan, called Master Plan for IT in Education, to incorporate technology into the school system (Mui *et al.*, 2004). This US\$1.2 billion project aimed to create an ICT-enriched school environment for every child. This first master plan focused primarily on installing computers and high-bandwidth Internet in schools and classrooms and training teachers on the use of computers. In 2002, Singapore's education ministry launched its Master Plan 2, in coordination with new reform policies. The new master plan adopted a more systemic, holistic approach in which all the key components of the system – ICT, curriculum, assessment, instruction, professional development, and school culture – were integrated. For example, the curriculum was reduced by 10–30% to allow for the integration of technology into the subject areas and university admission criteria were modified to include the submission of an electronic portfolio of student work, in addition to exam scores.

The Case of Finland

Finland's trajectory was different from that of Singapore, emphasizing knowledge production early on in its development. Finland is a country of 5.2 million people. In 1960, Finland too was relatively poor, at least by European standards. In that year, the per capita GNP was less than US\$7400 (in constant dollars), according to the Center for International Comparisons at the University of Pennsylvania. But by 2000, Finland's per capita GNP was US\$24 000, placing it among the countries with the highest standard of living in the world. Between 1990 and 2000, there was a fundamental structural transformation of Finland's economy, as it moved from a raw materials-based manufacturing economy to one with a high concentration in high-tech products, particularly in the area of telecommunications. In responding to a recession in the early 1990s, the Government of Finland adopted an economic policy that emphasized new knowledge as a productivity factor. Finland instituted a series of policy changes that shifted resources from the subsidization of large but uncompetitive industries to investments in

infrastructure, education, and research and development (Blomstrom *et al.*, 2002; Castells and Himanen, 2002). Both public and private research and development (R&D) investments grew rapidly in the 1990s. Public investments encouraged the development of small and medium enterprises (SMEs), through the establishment of incubators for start ups, and cross-sector, private–public collaborations through research grants for groups of SMEs, large corporations, and universities. Networking among these organizations improved knowledge flows and increased productive interactions.

These policies and programs were coordinated by a common vision of how all sectors of society would benefit from economic growth. In the early 1990s, the Ministry of Finance appointed a broad-based Information Science Advisory Board (2000) to construct a national information society strategy and articulate a vision for what Finland would be like as a country enriched by ICT “. . . in which knowledge and expertise form part of the culture and also the key factor in production” (p. 5). The Finnish Information Society Program emerged from this process.

Education was an important part of Finland’s economic shift toward an information society. The school system in Finland is highly decentralized and decision making is distributed across sectors (Ministry of Education, Finland, 1999, 2004). The ministry conceptualizes learning as an individual and community process of knowledge creation and collective problem solving. Each school writes its own curriculum based on very general guidelines from the National Board of Education and developed through discussions among teachers and parents. As a result, school curricula may be quite diverse across the country. Schools and teachers are given the authority to select teaching materials that correspond to the curriculum. Businesses also work closely with schools through apprenticeships and on-the-job training. Business leaders participate in school decision making.

The quality of Finland’s education system is evidenced by the fact that Finnish students scored second to (and not statistically lower than) students in Hong Kong, among 40 countries participating in the mathematics portion of the Program for International Student Assessment (PISA) (OECD, 2004c). (Singapore did not participate in the recent PISA nor did Finland participate in the recent TIMSS.) The country also scored first among nations on the science and reading portions of this test of 15-year-olds. In addition, Finland scored first in a special assessment of students’ problem-solving skills that measured students’ ability to analyze problem situations, apply knowledge to solve problems, and evaluate, justify, and communicate results (OECD, 2004d). The ministry attributes Finland’s excellent performance on PISA to free, high-quality education across the country, high-quality teachers with a high degree of autonomy, development-oriented assessment that gives students feedback on their

progress, and a socioconstructivist approach to learning that treats students as autonomous learners who are guided to develop their study skills and plan their life careers.

As part of this Finnish Information Society Program, the Ministry of Education developed the Information Strategy for Research and Education (Ministry of Education, Finland, 1999, 2004; Kankaanranta and Linnakyla, 2004). The goals of the education policy are to develop learning-centered instructional approaches, move from once-and-for-all training to lifelong learning, ensure that all students have information skills and that all teachers achieve a high level of professional skills, and build global education and research networks. Government programs have helped schools purchase computers, link them to the Internet, promote the introduction of ICT as a tool for teaching and learning, and conduct in-service training for teachers to reform the pedagogical practices in their schools, especially with regard to collaborative teaching.

Policy Implications

Causal connections cannot be claimed for either of these cases. It cannot be determined whether educational policies resulted in economic growth or economic growth provided the resources that allowed for more-advanced educational policies. Similarly, the co-occurrence of economic growth and use of ICT in schools does not justify conclusions about one of these factors causing the other. Most likely, the relationship is transactional. In their macroeconomic study of developing East Asian countries, Birdsall *et al.* (1995) found a double virtuous cycle between economic and educational policy. That is, better education policies contributed to economic growth, which enabled better policies, more growth, and so on. This finding argues for a more holistic approach to the formulation of policies and programs, one that emphasizes alignment rather than the search for causal mechanisms.

From this perspective, national ICT, education reform, and economic-development policies will have the greatest impact if they are aligned so that they can reinforce each other. This includes alignment between agencies, alignment within agencies, alignment through layers of the government, and alignment between policies and programs. Both Singapore and Finland established cross-ministerial agencies to coordinate their policies and programs within the government. This coordination can result in the high-level integration of policies that are articulated in a common vision for how technology, education, and economic development can benefit the entire society, as was done in both Singapore and Finland. Within this vision, ICT policies can best contribute to educational reform and economic development if these policies are all

aligned and if teacher professional development includes skills and activities that incorporate ICT into classroom practices and connects them to broader education and economic goals.

See also: Conceptions of Technology Literacy and Fluency; Public–Private Partnerships for Educational Reform.

Bibliography

- Anwar, S. and Zheng, M. (2004). Government spending on research and development and industrial production in Singapore. *International Journal of Asian Management* 3, 53–65.
- Ashton, D., Green, F., Sung, J., and James, D. (2002). The evolution of education and training strategies in Singapore, Taiwan, and S. Korea: A development model of skill formation. *Journal of Education and Work* 15(1), 5–30.
- Barro, R. (2000). *Education and Economic Growth*. Paris: OECD.
- Bereiter, C. (2002). Education in a knowledge society. In Smith, B. (ed.) *Liberal Education in a Knowledge Society*, pp 11–34. Chicago, IL: Open Court.
- Birdsall, N., Ross, D., and Sabot, R. (1995). Growth and inequity reconsidered: Lessons from East Asia. *World Bank Economic Review* 9(3), 477–508.
- Blomstrom, M., Kokko, A., and Sjöholm, F. (2002). Growth and innovation policies for a knowledge economy: Experiences from Finland, Sweden, and Singapore. *Background Paper for the LAC Flagship Report*. Washington, DC: World Bank.
- Bransford, J., Brown, A., and Cocking, R. (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academic Press.
- Bransford, J., Darling-Hammond, L., and LePage, P. (2005). Introduction. In Bransford, J. and Darling-Hammond, L. (eds.) *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do*, pp 1–39. San Francisco, CA: Jossey-Bass.
- Brown, P. and Lauder, H. (2000). *The Future of Skill Formation in Singapore (Working Paper #3)*. Cardiff: School of Social Sciences, Cardiff University.
- Bruckman, A. (2006). Learning in online communities. In Sawyer, R. K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 461–472. Cambridge: Cambridge University Press.
- Castells, M. and Himanen, P. (2002). *The Information Society and the Welfare State: The Finnish Model*. Oxford: Oxford University Press.
- Darling-Hammond, L. (1997). *The Right to Learn: A Blueprint for Creating Schools that Work*. San Francisco, CA: Jossey-Bass.
- Edleson, D. and Reiser, B. (2006). Making authentic practices accessible to learners: Design challenges and strategies. In Sawyer, R. K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 335–354. Cambridge: Cambridge University Press.
- Elmore, R. (2004). *School Reform from the Inside Out: Policy, Practice, and Performance*. Cambridge, MA: Harvard Education Press.
- Fishman, B. and Davis, E. (2006). Teacher learning research and the learning sciences. In Sawyer, R. K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 535–550. Cambridge: Cambridge University Press.
- Fullan, M. (2001a). *Leading in a Culture of Change*. San Francisco, CA: Jossey-Bass.
- Fullan, M. (2001b). *The New Meaning of Educational Change*, 3rd edn. New York: Teachers College Press.
- Hanusek, E. and Kimbo, D. (2000). Schooling, labor-force quality, and the growth of nations. *American Economic Review* 90(5), 1184–1208.
- Jorgenson, D. and Vu, K. (2005). Information technology and the world economy. *Scandinavia Journal of Economics* 107(4), 631–650.
- Kankaanranta, M. and Linnakyla, P. (2004). National policies and practices on ICT in education: Finland. In Plomp, T., Anderson, R., Law, N., and Quale, A. (eds.) *Cross-National Information and Communication Technology Policies and Practices in Education* pp 213–232. Greenwich, CT: IPA.
- Kozma, R. (2003). Technology and classroom practices: An international study. *Journal of Research on Computers in Education* 36(1), 1–14.
- Krajcik, J. and Blumenfeld, P. (2006). Project-based learning. In Sawyer, R. K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 317–334. Cambridge: Cambridge University Press.
- Krugman, P. and Wells, R. (2006). *Macroeconomics*. New York: Worth.
- Lall, S. (2000). *Skills, Competitiveness and Policy in Developing Countries (Working Paper Number 46)*. Oxford: Queen Elizabeth House.
- Law, N., Yuen, H., Ki, W., et al. (2000). *Changing Classrooms: A Study of Good Practices in using ICT in Hong Kong Schools*. Hong Kong: Centre for Information Technology in School and Teacher Education, University of Hong Kong.
- Lehrer, R. and Schauble, L. (2006). Cultivating model-based reasoning in science education. In Sawyer, R. K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 371–388. Cambridge: Cambridge University Press.
- McLaughlin, M. and Talbert, J. (2001). *Professional Communities and the Work of High School Teaching*. Chicago, IL: University of Chicago Press.
- Means, B. (2006). Prospects for transforming schools with technology-supported assessment. In Sawyer, R. K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 505–520. Cambridge: Cambridge University Press.
- Means, B. and Olson, K. (1995). *Technology's Role in Education Reform: Findings from a National Study of Innovating Schools*. Washington, DC: US Department of Education, Office of Educational Research and Improvement.
- Means, B., Penuel, W., and Padilla, C. (2001). *The Connected School: Technology and Learning in High School*. San Francisco, CA: Jossey-Bass.
- Ministry of Education, Finland (1999). *Education, Training and Research in the Information Society: A National Strategy for 2000–2004*. Helsinki: Ministry of Education.
- Ministry of Education, Finland (2004). *Information Society Programme for Education, Training, and Research: 2004–2006*. Helsinki: Ministry of Education.
- Ministry of Education, Singapore (2000). *Mission with a Passion: Making a Difference*. Singapore: Ministry of Education.
- Ministry of Education, Singapore (2002). *Speech by Mr. Tharman Shanmugaratnam, Senior Minister of Trade and Industry and Education*. <http://www.moe.gov.sg/speeches/2002/sp24072002.htm> (accessed September 2009).
- Ministry of Planning and International Cooperation, Jordan (2004). *National Social and Economic Development Plan: 2004–2006*. Amman: Ministry of Planning and International Cooperation.
- Mislevy, R., Steinberg, L., Almond, R., Haertel, G., and Penuel, W. (2003). Improving educational assessment. In Haertel, G. and Means, B. (eds.) *Evaluating Educational Technology: Effective Research Designs for Improving Learning*, pp 149–180. New York: Teachers College Press.
- Mui, Y., Kan, E., and Chun, T. (2004). National policies and practices on ICT in education: Singapore. In Plomp, T., Anderson, R., Law, N., and Quale, A. (eds.) *Cross-National Information and Communication Technology Policies and Practices in Education*, pp 495–508. Greenwich, CT: IPA.
- Mullis, I., Martin, M., Gonzalez, E., and Chrostowski, S. (2004a). *TIMSS 2003 International Mathematics Report*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).
- Mullis, I., Martin, M., Gonzalez, E., and Chrostowski, S. (2004b). *TIMSS 2003 International Science Report*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).
- New Commission on Skills of the American Workforce (2006). *Tough Choices, Tough Times: The Report of the New Commission on the Skills of the American Workforce*. San Francisco, CA: Jossey-Bass.
- Organization for Economic Co-operation and Development (OECD) (2004a). *The Economic Impact of ICT: Measurement, Evidence, and Implications*. Paris: OECD.
- Organization for Economic Co-operation and Development (OECD) (2004b). *Innovation in the Knowledge Economy: Implications for Education and Learning*. Paris: OECD.

- Organization for Economic Co-operation and Development (OECD) (2004c). *Learning for Tomorrow's World: First Results of PISA 2003*. Paris: OECD.
- Organization for Economic Co-operation and Development (OECD) (2004d). *Problem Solving for Tomorrow's World: First Measures of Cross-Curricular Competencies from PISA 2003*. Paris: OECD.
- Partnership for the 21st Century (2003). *Learning for the 21st Century*. Washington, DC: Partnership for the 21st Century.
- Pellegrino, J., Chudowski, N., and Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.
- Psacharopoulos, G. and Patrinos, H. (2002). Returns to investment in education: A further update. *World Bank Policy Research Working Paper 2881*. Washington, DC: World Bank.
- Rajan, R. (2003). Introduction and overview to "Sustaining competitiveness of in the new global economy: A case study of Singapore." In Rajan, R. (ed.) *Sustaining Competitiveness in the New Global Economy: A Case Study of Singapore*, pp 1–11. Cheltenham: Edward Elgar.
- Scaradmalia, M. and Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In Sawyer, R. K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 97–115. Cambridge: Cambridge University Press.
- Schliecher, A. (2006). *The Economics of Knowledge: Why Education is Key for Europe's Success*. Brussels: Lisbon Council.
- Schofield, J. (2006). Internet use in schools. In Sawyer, R. K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 521–534. Cambridge: Cambridge University Press.
- Schofield, J. and Davidson, A. (2002). *Bringing the Internet to School: Lessons from an Urban District*. San Francisco, CA: Jossey-Bass.
- Senge, P., Cambron-McCabe, N., Lucas, T., et al. (2000). *Schools that Learn*. New York: Doubleday.
- Sianesi, R. and Van Reenen, J. (2002). *The Returns to Education: A Review of the Empirical Macroeconomic Literature*. London: Institute for Fiscal Studies.
- Stahl, G., Koschmann, T., and Suthers, D. (2006). Computer-supported collaborative learning. In Sawyer, R. K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 409–426. Cambridge: Cambridge University Press.
- Stevens, P. and Weale, M. (2003). *Education and Economic Growth*. London: National Institute of Economic and Social Research.
- Stiglitz, J. (1999). *Public Policy for a Knowledge Economy*. Washington, DC: World Bank.
- Stiglitz, J. and Walsh, C. (2002). *Principals of Macroeconomics*, 3rd edn. New York: Norton.
- Tucker, M. (1996). Skills standards, qualifications systems, and the American workforce. In Resnick, L. and Wirt, J. (eds.) *Linking School and Work: Roles for Standards and Assessment*, pp 23–52. San Francisco, CA: Jossey-Bass.
- United Nations Industrial Development Organization (UNIDO) (2003). *Industrial Development Report 2002/2003: Competing through Innovation and Learning*. New York: UNIDO.

Further Reading

- Department of Education and Science, Ireland (n.d.). *Statement of Strategy: 2005–2007*. Dublin: Department of Education and Science.
- Information Society Advisory Board, Finland (2000). *Finland as an Information Society: The Report of the Information Society Advisory Board*. Helsinki: Information Society Advisory Board.
- Kozma, R. (in press). Comparative analyses of policies for ICT in education. In Voogt, J. and Knezek, G. (eds.) *International Handbook of Information Technology in Education*. Amsterdam: Kluwer.
- Kozma, R. (2005). National policies that connect ICT-based education reform to economic and social development. *Human Technology* 1(2), 117–156.
- Ministry of Education, Singapore (2006). *Overview of Masterplan II for ICT in Education*. <http://www.moe.gov.sg/edumall/mp2/mp2.htm>.

TECHNOLOGY AND LEARNING – EVALUATING TECHNOLOGY

Examining the Effects of Educational Technology Programs: Challenges and Strategies

M Russell, Boston College, Chestnut Hill, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Access to digital technologies in US schools has increased dramatically over the past 20 years. Today, all US public schools are connected to the Internet. On average, US public schools have one computer for every 3.8 students (Ed Week, 2009). Increasingly, schools, districts, and states are adopting 1:1 computing programs that aim to provide each student and teacher with a mobile computing device. With the recent introduction of mini-laptops that cost less than US\$400 and increased federal expenditures on educational technology, the number of 1:1 programs is likely to increase in the coming years.

In addition to computing devices provided by schools, many students bring their own digital devices to schools. These devices take the forms of MP3 players, cell phones, and devices that bundle multiple functions into a single platform such as iPhones. While schools are struggling to form policies and practices regarding these student-owned devices, there is little doubt that the presence of computer-based technologies has altered the landscape of US public schools over the past 20 years.

Examining the effects that the use of digital technologies has on teaching, learning, and school quality, however, is challenging. This article explores some of the factors that complicate research on the effects of educational technologies and suggests strategies that researchers can employ to acquire more accurate estimates of the effects that digital technologies are having on teaching, learning, and school quality. The factors explored here include: clearly defining goals of technology programs, placing outcomes in the context of implementation, measuring technology use, aligning outcome measures with goals and use, and employing statistical methods that account for the multilevel factors that influence technology use and adjust for intraclass correlations. Each of these issues is examined separately.

Defining Goals

Digital technologies are generally provided to schools in one of three ways. One approach introduces digital technology to help address a specific challenge. For example,

digital science probes attached to a computing device are introduced to help students deepen their understanding of data collection and scientific concepts. As another example, computer-based assessment systems are introduced to help teachers monitor students' acquisition of specific knowledge or skills.

A second way in which digital technologies are introduced in schools is through research conducted by a teacher or an external researcher. Here again, a given digital technology is introduced to address a specific challenge. In these cases, however, the tool has often been newly developed or has not been widely adopted in other schools or classrooms. For this reason, the purpose for introducing the technology is to improve the tool's design to enhance its usability, develop methods for integrating the tool with classroom practices, or collecting preliminary evidence about the tool's effectiveness.

A third approach to introducing digital technologies is large-scale adoption of a hardware device that is capable of running many different applications and can be used to meet a variety of needs. Placing computers in each classroom, acquiring a computer laboratory, and 1:1 laptop programs are all examples of large-scale adoptions.

When digital technologies are introduced to address a specific challenge or as part of a research project, the goals for the use of that technology are usually defined before the technology is delivered. Science probes are introduced to increase student engagement during learning, improve data-collection skills, and improve understanding of specific scientific principles. Computer-based assessment systems are intended to provide teachers with more immediate information to inform instruction and to improve student performance in the areas assessed by the system. When digital technologies are introduced as part of research, researchers have specific questions they are addressing and hypotheses they are testing.

When digital technologies are introduced as part of a large-scale hardware adoption, the goals are often less clearly defined. In contrast to studies that examine the effect that individual software applications have on specific aspects of students' learning (see Goldberg *et al.*, 2003; Kulik, 1994; Sivin-Kachala, 1998; Waxman *et al.*, 2003; for

meta-analyses of nearly 800 studies focusing on specific technology applications), most large-scale technology-adoption programs introduce a digital device to meet a variety of goals. In addition, when implementing large-scale adoption programs, most school districts do not articulate the specific goals for their technology programs prior to launching their adoption programs (Herman, 1994).

Although schools purchase technology to enhance teaching and learning, when hardware is first introduced, schools are often unsure how to use the technology in order to have an impact on teaching and learning. It is not until teachers have had multiple opportunities to work with the new technology that they begin to develop a sense of how to use the technology or how it can have an impact on teaching and learning. As an example, consider the Maine Laptop Program, which provided laptops for every middle school student in the state. When first launched, professional development focused primarily on the mechanics of using laptops, such as logging in, launching programs, creating documents, saving files, etc. It was not until the second and third years of the program that professional-development programs for specific subject areas were introduced and focused on specific instructional uses of the laptops (Silvernail and Buffington, 2009).

The lack of clearly defined goals for large-scale adoption programs challenges researchers and evaluators who aim to examine the effects that newly introduced technology has on teaching, learning, and school quality. To overcome this obstacle, two opposing strategies are available. First, a goals-free approach to program evaluation can be employed (Knott, 1993; Scriven, 1972). In a goals-free approach the effects of a technology program are examined with an open lens that aims to detect and document any changes that affect teachers, students, and school leaders.

A second strategy begins by working closely with program leaders to define program goals and then identify the conditions and actions required to obtain those goals. As part of this process, each intended use of the technology is clearly articulated. For each use, the people impacted by the use are identified and the intended outcomes are defined. Next, the infrastructure, personnel, and training required to support each use are detailed. Finally, a timeline for implementation and expected outcomes is defined. Through this process, a logic model is developed that links each intended outcome with the supports believed to be necessary to achieve the outcomes. Research and evaluation then focus on the articulated logic model to examine the extent to which requisite conditions and actions are put in place, and whether the intended outcomes follow (Russell, 2002).

These two approaches contrast sharply with each other. In the first, the researcher or evaluator acknowledges that the program goals are ill-defined and therefore looks broadly at a program to capture any outcomes (positive or negative) that may emerge. In the second, the researcher or evaluator works closely with program

leaders to clearly define a range of desirable outcomes, develops a logic model that details what is required to reach those outcomes, and then focuses closely on examining the extent to which the supports are put in place and lead to the outcomes.

Considering Implementation

It is important that researchers and evaluators examine the effects that digital technologies have on teaching, learning, and school quality. However, it is equally important to place those outcomes in the context of program implementation (Baker, 1999; Herman, 1994; Russell, 2002). Like all attempts to alter teaching, learning, and school processes, the full effects of a given digital tool will not be realized until teachers, students, and school communities develop comfort, knowledge, and skill with the new tool. Traditionally, developing comfort, knowledge, and skill are aided by professional development and, more recently, by collaborative efforts among teachers to improve practices. In addition to training and time to develop comfort, the use of digital technologies often requires additional support structures. Some of these supports are physical, such as easy access to hardware, reliable networks, server communications, and Internet connections. Other supports have a human element, such as personnel who provide technical support, and leadership that embraces and encourages experimentation with a new technology. The degree to which schools provide these physical and human-support structures varies widely. In turn, this variation influences the degree to which and ways in which teachers adopt and adapt digital technologies (Miranda, 2006). As a result, before outcomes can be attributed to the use of a computer-based tool, it is essential to examine and document the way in which a digital technology has been implemented (Heinecke *et al.*, 1999).

Before examining the outcomes of a technology program, it is important to identify the conditions that must be established before meaningful effects can reasonably occur. Depending on the computer-based tool, these enabling conditions might include: (1) providing access to equipment, software, and other physical infrastructure such as the Internet or servers; (2) professional development; (3) time for teachers to learn how to use the tool; (4) time for teachers to work with other teachers or curriculum technology specialists to develop lessons that incorporate the computer-based tool into instruction; (5) time for students to develop the essential skills to make use of the tool; and (6) sufficient time for teachers to improve these lessons. Unless the conditions believed necessary for successful use of the computer-based tool are established and it is established that the tool is actually being used, it is premature to examine the impact the tool has on the intended outcome.

When examining the impact of a computer-based tool, it is also important to consider the extent to which the tool is readily accessible to students and teachers, and used by a critical mass of students and/or teachers. As an example, in a study that attempted to examine the effect that a program, which placed computers in low-performing schools, had on student learning, the researchers assumed that the installation of computers at a ratio of ten students per computer provided sufficient access to the computer-based tools (Angrist and Lavy, 2001). Moreover, it was assumed that the act of placing computers in the school laboratory would lead to the use of computers by students. As a result of these assumptions, analyses focused on comparing changes in test scores in schools that were provided with the computer labs versus those that were not. The analyses, however, did not consider the extent to which computers were being used to aid learning within specific subject areas in either setting. While the study was interpreted to provide evidence that the use of computers do not improve student learning, in fact the study did not focus on use of computers – only the presence of computers in a computing laboratory.

In addition to placing outcomes in the context of use, it is also important to consider the extent to which students and teachers have easy access to a given digital technology. Recent research provides evidence that the ratio of students to computer-based tools has a large impact on the extent to which the tools are used. Across several studies that compare the amount of use that occurs prior to and following the introduction of 1:1 computing programs, researchers have consistently found sharp increases in use when a given technology is provided for each student (Russell *et al.*, 2003, 2004; Bebell, 2009; Shapley *et al.*, 2008). Much of this research also documents that after technology is available for each student, the teacher's role within the classroom shifts. As an example, one study that focused on the introduction of AlphaSmarts (i.e., portable writing devices) at a 1:1 ratio found that the amount of time teachers spend managing student use of the technology decreased and the amount of time they spend working directly with students on their writing increased (Russell *et al.*, 2003). Given the complex relationship between presence, access, use, and a teacher's role in the classroom, simply establishing that technology is present is not sufficient for measuring how often and in what ways a computer-based tool is used. Moreover, without understanding how and how often a tool is used, it is difficult to assess the impact of that tool on the intended outcome.

Measuring Technology Use

When examining educational technology, it is essential to consider the extent to which and ways in which the tool is used by students and teachers. Given the wide range of

computer-based tools and resources that can be used to support teaching and learning, specific, rather than general, measures of computer use are required. Again, Angrist and Lavy's (2001) study serves as a good example of the challenge posed by measuring computer use. In this study, teachers were presented with the following survey items:

1. Which of the following do you use when teaching?
 - Xeroxed worksheets.
 - Instructional booklet.
 - Games.
 - Computer software or instructional computer programs.
 - TV programs.
 - Other audio-visual materials.
2. For each item, teachers were given the following options:
 - Not at all.
 - Sometimes.
 - Frequently.
 - Almost always.

Each teacher's response to how often they used computer software or instructional computer programs served as the measure of instructional use of computers for mathematics and Hebrew.

It may be reasonable to infer that teachers who report that they frequently use computer software or instructional computer programs employ different instructional practices than do teachers whose response is not at all. However, this general measure does not provide sufficient information about how, when, why, or who is actually using computer-based tools. As a result, a teacher who uses PowerPoint several times a week to present information that otherwise would be presented using an overhead projector will appear to use computer-based tools more than the teacher who has students work for 30 min twice a week on a mathematics tutorial. Similarly, this general measure of computer use makes this second teacher appear the same as a teacher who might also have students use a computer twice a week for 30 min to collect data, create spreadsheets, and produce graphs to explore relationships in data. Finally, a teacher who uses computers daily outside of class to create worksheets or record grades might appear to be the most frequent user of computer-based tools. In other words, simply asking teachers to report the frequency with which they use computers does not provide sufficient information to document how computer-based tools are being used or what impact these uses may have on learning. To answer these research questions, computer-use in the context of a given study must be clearly defined and measured.

When measuring technology use by teachers and students, some researchers have developed several items each of which focuses on a specific use of technology (Bakia *et al.*, 2007; Becker, 1999; Russell *et al.*, 2003). Often, these data are collapsed into a single measure of technology use. This aggregated measure is useful for

providing a general sense of whether or not teachers are using technology for some aspect of their job. However, like the single survey item employed by Angrist and Lavy that focuses on computer use in general, these aggregate measures are inadequate for understanding the extent to which technology is being used by teachers, the purposes of these uses, and the impact these uses may have on learning outcomes.

Instead of aggregating several survey items into one measure that represents the general use of computers, a richer understanding of computer use results when these individual items are used to create multiple measures of computer uses. As an example, Russell *et al.* (2003) used data from several survey items, each of which focused on a specific use of technology, to develop seven separate scales that measure teachers' technology use. These seven scales include:

- teachers' use of technology for class preparation (preparation);
- teachers' professional e-mail use (professional e-mail);
- teachers' use of technology for delivering instruction (delivering instruction);
- teachers' use of technology for accommodation (accommodation);
- teacher-directed student use of technology during class time (student use);
- teacher-directed student use of technology to create products (student products); and
- teachers' use of technology for grading (grading).

Analyses of the seven teacher technology-use scales formed from the items showed that each of the individual scales exhibited widely divergent frequency distributions (Bebell *et al.*, 2004). For example, teachers' use of technology for preparation was negatively skewed while the use of technology for instruction was strongly positively skewed. Like instructional use, the distributions for assigning student products and providing accommodations were positively skewed. Using technology for grading had a weak positive skew while teacher-directed student use was approximately normally distributed. Use of e-mail however, presented a bi-modal distribution, with a large percentage of teachers reporting frequent use and a large portion of the sample reporting no use. In short, the distribution of responses differed greatly across the seven scales. However, when all of the survey items comprising these scales were summed to create a single generic composite measure of technology use, the distribution closely approximated a normal distribution revealing none of the patterns observed in the specific technology-use scales.

When compared to a single generic measure of technology use, multiple measures of specific technology use offer a more nuanced understanding of how teachers are using technology and how these uses vary

among teachers. Studies that have utilized a multifaceted approach to measuring technology use have revealed many interesting patterns that were obscured when only general measures of use were employed (Mathews, 1996; Ravitz *et al.*, 1999; Bebell *et al.*, 2004). For example the analysis of the USEIT teacher data indicated that the frequency of teachers' technology use for instruction and accommodating lessons was unrelated to the frequency with which they asked students to use technology during class time. Similarly teachers' use of technology for grading operated independently of teachers' use of technology for preparation of their lessons (Bebell *et al.*, 2004). Given these and other findings that demonstrate the value of discrete measures of technology use in understanding how teachers are using technology and what effects these uses may have on student learning, it is imperative that research on educational technology abandon single, generic measures of technology use and instead develop nuanced measures that reflect specific, discrete uses of educational technology.

Aligning Outcome Measures

A decade ago, McNabb *et al.* (1999: 4) wrote, "Standardized tests scores have become the accepted measure with which policymakers and the public gauge the benefits of educational investments." Given the requirements of No Child Left Behind Act to test every student in grades 3–8 and at least one grade in high school, the importance of standardized tests as a valued outcome measure remains strong today.

Given the availability of standardized and state-mandated tests, it is cost effective and efficient to employ scores from these tests to assess the effect of a computer-based tool on student learning. This strategy has been used for several studies including that of Angrist and Lavy (2001), who used Israel's national Hebrew and mathematics standardized test, Wenglinsky (1998), who used scores from the National Assessment of Educational Progress (NAEP), O'Dwyer *et al.* (2005, 2008), who used scores from the Massachusetts Comprehensive Assessment System (MCAS), and Shapley *et al.* (2008), who used scores from Texas state tests.

While standardized test scores can be useful for comparing performance or changes in performance across a large sample of students, it is essential to consider whether the domain tested by the standardized or state-mandated test is too broad to represent the type of learning believed to be impacted by the use of a given technology. As an example, the NAEP eighth-grade mathematics test includes items from several subdomains that include number sense, probability and statistics, algebra, and geometry. Many software and instructional uses of technology, however, focus on developing students'

skills in only one or two of these areas. For example, an evaluation of a Massachusetts school district revealed that most third- and fourth-grade teachers in this district used computers as part of their mathematics instruction to help students develop spatial reasoning skills (Russell, 2000). However, on the state's fourth-grade mathematics test, only two of the 39 items related to spatial reasoning. Given that only two items measured the skills that may be impacted by this use of computers, the total state test score did not provide a measure that was sensitive to the potential effects of computer use. While analyses could focus on the subset of items that measure spatial reasoning skills, the small number of items was insufficient to provide a reliable estimate of student achievement in this area.

Rather than employing state test results as an outcome measure, an alternate strategy is to develop custom measures that contain a larger number of items specific to the skills and knowledge believed to be affected by computer use. Although it can be difficult to convince teachers and/or schools to administer an additional test, aligned measures of student learning will result in more reliable scores and will increase the validity of inferences about the impacts of specific computer uses on student learning. When developing custom measures that are aligned with the types of skills and knowledge targeted by a given technology or technology program, it is essential to consider the technical quality, including the reliability and validity, of the data-collection tools.

When collecting measures of student learning, it is also important to consider the method used to administer and collect outcome measures. Several studies conducted over the past decade provide evidence that paper-based measures of student learning underestimate the achievement of students who are accustomed to working with technology. This effect occurs when students are required to produce work using a tool (i.e., paper and pencil) that differs from what they use regularly to produce work for the classroom (Horkay *et al.*, 2006; Russell, 1999; Russell and Haney, 1997; Russell and Plati, 2001, 2002). Through a series of randomized experiments, Russell and his colleagues provide empirical evidence that students who are accustomed to writing with computers in the classroom perform between 0.4 and 1.1 standard deviations higher when they are allowed to use a computer when performing tests that require them to compose written responses (Russell, 1999; Russell and Haney, 1997; Russell and Plati, 2001, 2002). These findings were replicated in a study conducted as part of the 2004 NAEP writing test (Horkay *et al.*, 2006). These results suggest that researchers studying the effects of educational technology are particularly at risk for underestimating the ability of technology-savvy students when they rely on paper-based assessment instruments as their outcome measures.

Together, the poor alignment between the broad domain measured by a standardized test, the specific skills and knowledge targeted by a given computer-based tool, and the mode of administration effect that results with a paper-based test, relying on standardized tests to examine the effects of technology use on student learning can be problematic. When using standardized tests, or any measure of student achievement, it is important that the constructs measured by the instrument are aligned with the constructs developed through students' uses of technology and that the form of the test (e.g., paper- or computer-based) does not add construct-irrelevant variance.

Employing Appropriate Analytic Methods

The use of technology by teachers and students is influenced by many factors. Some of these factors are specific to the individual using a given technology and include characteristics such as prior experience, comfort using technology, grade level, pedagogical beliefs, and subject area (Becker, 1999; O'Dwyer *et al.*, 2004; Ravitz *et al.*, 1999). Other factors reside at higher levels within an educational setting and include factors such as local leadership and vision, funding, local policies, beliefs about teaching and learning, available infrastructure, and support (Culp *et al.*, 2003). Given the interrelationships among factors that are found at different levels of a school system, student, teacher and school use of technology can be conceptualized in terms of a multilevel and hierarchical framework. Individual students use technology within the context of a classroom; teachers operate within a school and are influenced by its technology policies; and schools, in turn are affected by district level decisions regarding the allocation and use of technology. This natural hierarchy necessitates the use of multilevel modeling techniques (Rumberger, 2000) when examining or comparing the effects of a given technology or technology programs across settings.

The simplest multilevel model that can be specified is a two-level model. Individuals within groups are modeled at the first level, and their group membership is modeled at the second level. In the three-level model, the individual is modeled at level one, and their membership of two groups is modeled at levels two and three. For example, students are nested within classrooms and classrooms are nested within schools. Alternatively, teachers are nested within schools, and schools within districts. The sample design employed by any research study dictates the complexity of the multilevel model that can be employed; more complex sample designs allow more levels to be modeled, thus yielding more useful and accurate information about the contexts that affect the outcomes of interest. Failure to account for the nested structure in which a given technology is employed has

the potential to produce inaccurate and misleading conclusions about the effects of a digital technology on teaching, learning, and school quality.

In recent years, several studies have employed multi-level modeling techniques to examine the use and effect of digital technologies in educational settings. To estimate the effect that several factors have on the use of technology by teachers for instructional purposes, Miranda (2006) developed a series of multilevel structural-equation models. To examine the relationship between specific uses of computers by students and performance on mathematics and language arts tests, O'Dwyer *et al.* (2005, 2008) developed a series of hierarchical linear models which revealed that the relationship between computer use and test performance varied depending on the specific use of technology. Shapley *et al.* (2008) also employed hierarchical linear modeling techniques to examine the effect that a 1:1 laptop program had on student performance in several Texas schools. Finally, Silvernail and Buffington's (2009) study of a professional development program that focused on the use of technology for mathematics instruction also employed multi-level modeling techniques to adjust for the influence of classroom-specific factors. Across each of these examples, more sophisticated analytic techniques were required to separate the effects that occur at different levels within a learning environment and to account for correlations that occur within classrooms that result from commonalities among students who work with a given teacher, are placed in a given educational program, and/or live in a similar environment.

Summary

Given the investment that our society is making in educational technologies, it is important that research examine the effect that these technologies are having on teaching, learning, and school quality. When examining these effects, however, it is important to collect nuanced measures of the ways in which the technologies are being used. It is also important to place these uses in the context of program implementation and to employ outcome measures that are aligned with and sensitive to the types of changes that a given use is expected to have. Finally, given the complex structure of educational systems, it is important to employ statistical methods that account for the multiple levels that influence teaching, learning, and school quality, and which adjust for relationships that occur within these levels. By considering each of these issues, research is likely to provide a more accurate and nuanced understanding of the effects that digital technologies are having on teaching, learning, and school quality.

See also: An Overview of Technology and Learning.

Bibliography

- Angrist, J. and Pischke, V. (2001). *New Evidence on Classroom Computers and Pupil Learning*. Bonn: Institute for the Study of Labor.
- Baker, E. (1999). Technology: How do we know it works? *Secretary's Conference on Educational Technology*. Washington: DC. <http://www.ed.gov/rschstat/eval/tech/techconf99/whitepapers/paper5.html>.
- Bakia, M., Mitchell, K., and Yang, E. (2007). *State Strategies and Practices for Educational Technology: Volume I – Examining the Enhancing Education through Technology Program*. Washington, DC: US Department of Education. <http://www.ed.gov/rschstat/eval/tech/netts/netts-vol1.pdf> (accessed September 2009).
- Bebell, D. (2009). *Berkshire Wireless Learning Initiative 1:1 Laptop Evaluation Study Outcomes*. Chestnut Hill, MA: Technology and Assessment Study Collaborative.
- Bebell, D., Russell, M., and O'Dwyer, L. M. (2004). Measuring teachers' technology uses: Why multiple measures are more revealing. *Journal of Research on Technology in Education* 37(1), 45–63.
- Becker, H. J. (1999). *Internet Use by Teachers: Conditions of Professional Use and Teacher-Directed Student Use*. Irvine, CA: Center for Research on Information Technology and Organizations.
- Culp, K. M., Honey, M., and Spielvogel, B. (2003). Local relevance and generalizability: Linking evaluation to school improvement. In Means, B. and Haertel, G. (eds.) *New Approaches to Evaluating the Impact of Educational Technology*. New York: Teachers College Press.
- Ed Week (2009). Technology Counts: Breaking Away from Tradition. 28(6).
- Goldberg, A., Russell, M., and Cook, A. (2003). The effect of computers on student writing: A metaanalysis of studies from 1992 to 2002. *Journal of Technology, Learning, and Assessment* 2(1). <http://www.jtla.org> (accessed September 2009).
- Heinecke, W., Blasi, L., Milman, N., and Washington, L. (1999). New directions in the evaluation of the effectiveness of educational technology. *A Paper for the Secretary's Conference on Educational Technology*. Washington: DC. <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED452825>.
- Herman, J. L. (1994). Evaluating the effects of technology in school reform. In Means, B. (ed.) *Technology and Education Reform: The Reality Behind the Promise*, pp 133–167. Jossey-Bass: San Francisco, CA.
- Horkay, N., Bennett, R. E., Allen, N., Kaplan, B., and Yan, F. (2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment* 5(2). <http://www.jtla.org> (accessed September 2009).
- Knott, T. The role of evaluation in the educational reformation. In Estes, N. and Thomas, M. (eds.) *Rethinking the Roles of Technology in Education*. Cambridge, MA: Massachusetts Institute of Technology.
- Kulik, J. (1994). Meta-analytic studies of findings on computer-based instruction. In Baker, E. and O'Neil, H. Jr (eds.) *Technology Assessment in Education and Training*, pp 9–33. Erlbaum: Hillsdale, NJ.
- Mathews, J. (1996). Predicting teacher perceived technology use: Needs assessment model for small rural schools. *Paper Presented at the Annual Meeting of the National Rural Education Association*. San Antonio: TX.
- McNabb, M., Hawkes, M., and Rouk, U. (1999). Critical issues in evaluating the effectiveness of technology. *A Summary Report on Secretary's Conference on Educational Technology. Evaluating the Effectiveness of Technology*. Washington, DC. <http://www.ed.gov/Technology/TechConf/1999/confsum.html> (accessed September 2009).
- Miranda, H. (2006). Predictors of technology use for elementary school teachers in Massachusetts: A multilevel sem approach. Chestnut Hill, MA: Boston College. <http://escholarship.bc.edu/cgi/viewcontent.cgi?article=5699&context=dissertations>.
- O'Dwyer, L., Russell, M., and Bebell, D. (2004). Identifying teacher, school and district characteristics associated with teachers' use of technology: A multilevel perspective. *Educational Policy Analysis Archives* 12(48). <http://epaa.asu.edu/epaa/v12n48/v12n48.pdf> (accessed September 2009).
- O'Dwyer, L. M., Russell, M., Bebell, D., and Tucker-Seeley, K. R. (2005). Examining the relationship between home and school computer use

- and students' English/language arts test scores. *Journal of Technology, Learning, and Assessment* 3(3). <http://escholarship.bc.edu/cgi/viewcontent.cgi?article=1053&context=jtla> (accessed September 2009).
- O'Dwyer, L. M., Russell, M., Bebell, D., and Tucker-Seeley, K. R. (2008). Examining the relationship between student's mathematics test scores and computer use at home and at school. *Journal of Technology, Learning and Assessment* 6(5). <http://escholarship.bc.edu/cgi/viewcontent.cgi?article=1121&context=jtla> (accessed September 2009).
- Ravitz, J., Wong, Y., and Becker, H. (1999). *Teacher and Teacher Directed Student Use of Computers and Software*. Irvine, CA: Center for research on information technology and organizations.
- Rumberger, R. (2000). A multi-level, longitudinal approach to evaluating the effectiveness of educational technology. *Paper Presented at the SRI Design Meeting on Design Issues in Evaluating Educational Technologies*. http://homepages.shu.ac.uk/~edsjlc/ict/information_sources/US_office_of_technology/eval%20IT/rumberg5.doc.
- Russell, M. (1999). Testing writing on computers: A follow-up study comparing performance on computer and on paper. *Educational Policy Analysis Archives* 7(20). <http://epaa.asu.edu/epaa/v7n20/> (accessed September 2009).
- Russell, M. (2000). It's time to upgrade: Tests and administration procedures for the new millennium. *A Paper for the Secretary's Conference on Educational Technology*, Washington, DC. http://www.usca.edu/essays/voll2002/time_to_upgrade_revised.pdf (accessed September 2009).
- Russell, M. (2002). Framing technology program evaluations. In Heinecke, W. and Blasi, L. (eds.) *Methods of Evaluation Educational Technology*, pp 145–158. Greenwich, CT: Information Age Publishing.
- Russell, M., Bebell, D., Cowan, J., and Corbelli, M. (2003). An AlphaSmart for each student: Does teaching and learning change with full access to word processors? *Computers and Composition* 20, 51–76.
- Russell, M., Bebell, D., and Higgins, J. (2004). Laptop learning: A comparison of teaching and learning in upper elementary classrooms equipped with shared carts of laptops and permanent 1:1 laptops. *Journal of Educational Computing Research* 30(3), 313–330.
- Russell, M. and Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Educational Policy Analysis Archives* 5(3). <http://epaa.asu.edu/epaa/v5n3.html> (accessed September 2009).
- Russell, M., O'Dwyer, L., Bebell, D., and Miranda, H. (2003). *Technical Report for the USEIT Study*. Boston, MA: Boston College, Technology and Assessment Study Collaborative.
- Russell, M. and Plati, T. (2001). Effects of computer versus paper administration of a state-mandated writing assessment. *Teachers College Record*. <http://www.tcrecord.org/Content.asp?ContentID=10709> (accessed September 2009).
- Russell, M. and Plati, T. (2002). Does it matter with which I write: Comparing performance on paper, computer and portable writing devices. *Current Issues in Education* 5(4). <http://cie.ed.asu.edu/volume5/number4/index.html> (accessed September 2009).
- Scriven, M. (1972). Pros and cons about goal-free evaluation. *Evaluation Comments* 3(4), 1–7.
- Shapley, K., Sheehan, D., Maloney, C., and Caranikas-Walker, F. (2008). *Evaluation of the Texas Technology Immersion Pilot: Outcomes for the Third Year (2006–07)*. Austin, TX: Texas Center for Educational Research. http://www.tcer.org/research/etxtp/documents/y3_etxtp_quant.pdf.
- Silvernail, D. L and Buffington, P. J (2009). *Improving Mathematics Performance Using Laptop Technology: The Importance of Professional Development for Success*. Gorham, ME: Center for Education Policy, Applied Research, and Evaluation. http://www.tcer.org/research/etxtp/documents/y3_etxtp_quant.pdf (accessed September 2009).
- Sivin-Kachala, J. (1998). *Report on the Effectiveness of Technology in Schools, 1990–1997*. Washington, DC: Software Publisher's Association.
- Waxman, H. C., Lin, M. F., and Michko, G. M. (2003). A meta-analysis of the effectiveness of teaching and learning with technology on student outcomes. Naperville, IL: Learning Point Associates. <http://www.ncrel.org/tech/effects2/waxman.pdf> (accessed September 2009).
- Wenglinsky, H. (1998). *Does it compute? The Relationship between Educational Technology and Student Achievement in Mathematics*. Princeton, NJ: Educational Testing Service Policy Information Center.

Further Reading

- Haertel, G. and Means, B. (2003). *Evaluating Educational Technology: Effective Research Designs for Improving Learning*. New York: Teachers College Press.
- Heinecke, W. and Blasi, L. (2001). *Methods of Evaluating Educational Technology*. Charlotte, NC: Information Age Publishing.
- Means, B. and Haertel, G. D. (2004). *Using Technology Evaluation to Enhance Student Learning*. New York: Teachers College Press.
- Russell, M. (2006). *Technology and Assessment: The Tale of Two Interpretations*. Charlotte, NC: Information Age Publishing.

Relevant Websites

- <http://ctl.sri.com> – Center for Technology in Learning.
- <http://www.iste.org> – International Society for Technology in Education (ISTE).
- <http://www.iste.org> – Journal of Research on Technology in Education (JRTE), International Society for Technology in Education (ISTE).
- <http://escholarship.bc.edu/jtla/> – Journal of Technology, Learning, and Assessment, eScholarship@BC.
- <http://www.cse.ucla.edu> – National Center for Research on Evaluation, Standards, and Student Testing (CRESST), UCLA.
- <http://www.setda.org> – State Educational Technology Directors Association (SETDA).
- <http://www.bc.edu/research/intasc/> – Technology Assessment Study Collaborative at Boston College.
- <http://virtualassessment.org> – Virtual Assessment Project.

TECHNOLOGY AND LEARNING – GLOBAL TRENDS

Contents

Equity in Technology Access and Opportunities
Internet-based Education

Equity in Technology Access and Opportunities

T Matuchniak and M Warschauer, University of California, Irvine, Irvine, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Digital divide – The inequality in access to computers and the Internet.

Introduction

Access to information and communications technology (ICT) is vital in the knowledge economy and information society of the twenty-first century. The last 20 years has seen a great proliferation and diffusion of digital media; yet, there remain large inequities in both access to and use of ICT in the United States and around the world.

This article examines the issue of equity and technology as it relates to access, use, and outcomes, both at home and in schools; it surveys the landscape within the United States as well as internationally, looking at the divide both within and across countries; and it discusses current approaches and efforts to promote more equitable access to and use of ICT.

Equity within the United States

Home Access

Although virtually all American children have access to ICT at school, the digital divide is wider in children's home environments. According to the National Telecommunications and Information Administration (2008b), only 61.7% of American households have some type of Internet access (see **Table 1**). Not surprisingly, the greatest gap is seen between low-socioeconomic status (SES) and high-SES households and between households with high and low educational attainment. Similar disparity is seen across racial groups, with Black and Hispanic children less likely than Whites and Asians to have Internet access at home.

There further exists a disparity in the type of Internet connectivity in different households. In 2008, only 59.5% of households with children had broadband connectivity (National Telecommunications and Information Administration, 2008b), and traditionally disadvantaged groups continued to lag behind even further. For example, from 2007 to 2008 broadband adoption slightly declined among the poor (28–25%; Horrigan, 2008). Finally, there is great differential by SES and ethnic group as to the user–computer ratio in home environments with children in White and Asian families much more likely to have their own computer and children in Black and Hispanic families much more likely to share a single computer among four to five family members (Warschauer and Matuchniak, in press).

Home Use

A recent US study of 800 youth and young adults suggests that home use of the Internet is highly bifurcated, with the large majority using the Internet principally for informal interaction with their friends, and a much smaller minority using the Internet for advanced multimedia production, national and international communication, and self-directed learning (Ito *et al.*, 2008). Prior research suggests that the latter group is disproportionately White and from high-income families, with low-income, African-American, and Latino youth discouraged from such sophisticated usages by lack of up-to-date hardware or software, lack of literacy or computer skills, or lack of social support (Attewell and Winston, 2003). Finally, boys have been shown to play computer games substantially more than girls (Lenhart *et al.*, 2008), leading some to be concerned that girls could have fewer opportunities to develop the technological fluency and advanced skills associated with such play.

Academic Outcomes from Home Use

Children's access to ICTs at home appears to have a positive effect on learning outcomes. An analysis of two

Table 1 Percent of US households with Internet access

	Percent of households with Internet access			Broadband as percent of those with access
	Total with access	Broadband	Dial-up	
Total households	61.7	50.8	10.7	82.3
Family income				
Under US\$5000	31.9	26.7	5.3	83.6
US\$150 000+	95.5	90.3	5.0	94.6
Educational attainment of head of household				
Elementary	18.5	13.1	5.4	70.8
BA+	84.1	74.2	9.7	88.2
Race of head of household				
White	67.0	54.9	11.8	82.0
Black	44.9	36.4	8.4	80.9
Native American	41.5	29.8	11.2	71.9
Asian	75.5	69.1	6.1	91.5
Hispanic	43.4	35.2	8.0	81.1

From National Telecommunications and Information Administration (2008b). *Households Using the Internet in and Outside the Home, by Selected Characteristics: Total, Urban, Rural, Principal City, 2007*.

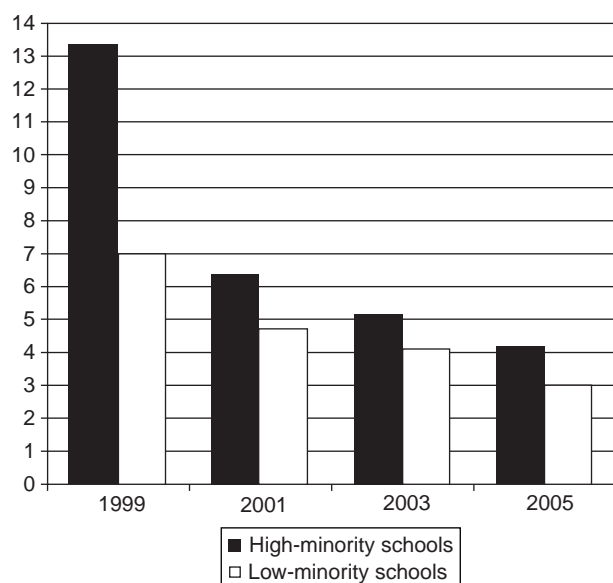


Figure 1 Number of students per Internet-connected computer in high-minority and low-minority US K-12 Schools, 1999–2005. From Wells, J., Lewis, L., and Greene, B. (2006). *Internet Access in U.S. Public Schools and Classrooms: 1994–2005*. Washington, DC: NCES.

national data sets indicated that, when controlling for income, race, parental education, and other factors, teenagers who have access to home computers are 6–8 percentage points more likely to graduate from high school than teenagers who do not have home computers (Beltran *et al.*, in press). However, this effect is highly differentiated, with high-SES students receiving the greatest academic benefit from home computer access and low-SES and minority students either receiving a small benefit or a negative affect from such access, with the differences

attributed to disparities in social support in the home environment that influence the manner in which children use technology in academically productive or unproductive ways (Attewell and Battle, 1999; Clotfelter *et al.*, 2008).

School Access

Much progress has been made over the last 10–15 years in increasing access to computers and the Internet in US schools. In 2005, 94% of instructional rooms had connectivity (compared to only 3% in 1994), and 97% of those have broadband connections (as compared to only 80% in 2000). Similar gains have been seen over the years with regard to the ratio of students to Internet-accessed computers, which in 2005 was 3.8:1 as compared to 12.1:1 in 1998 (Wells *et al.*, 2006). Gaps in school access to computers by race have narrowed but still exist. According to the most recent national US data, the ratio of students to Internet-connected computers in high-minority schools (schools with 50% or more racial or ethnic minorities) was 4.1:1 as compared to only 3.0:1 in low-minority schools (schools with fewer than 6% minorities; Wells *et al.*, 2006) (Figure 1).

School Use

Some 85% of Whites in grades pre-K to 12 report using a computer at school compared to 80% of Hispanics and 82% of Blacks, suggesting that ICTs have become an integral part of school life. The current disparities that remain are less about whether computers are used and more about how they are used. An analysis of national data showed that White and Asian students were more likely than Black or Hispanic students to use computers for cognitively challenging learning tasks such as simulations, and Black

and Hispanic students were more likely to use computers for less-challenging drill-and-practice tasks (Wenglinsky, 1998). A subsequent study (Becker, 2000) confirmed these differences and showed that similar divides existed among high- and low-SES groups.

Differences in how computers are used between high- and low-SES students appear to be related to greater experience and technological expertise among teachers in high-SES schools (Wenglinsky, 1998; Becker, 2000); differential academic expectations and standards in high-versus low-SES schools (Warschauer, 2000); pressure to focus on basic skills and test scores in low-SES schools (Warschauer *et al.*, 2004); and reluctance to assign technology-based homework to low-SES students who may not have computers at home (Warschauer *et al.*, 2004).

Academic Outcomes from School Use

Disparities in academic outcomes associated with technology use in schools mirror those at home. A review of national data sets indicated that SES is the single strongest factor predicting whether technology use is positively (for high-SES students) or negatively (for low-SES students) associated with changes in test-score outcomes in reading, mathematics, and science (Wenglinsky, 2005). Further analysis in the same study showed that the type of drill-and-practice computer activities most frequently used in low-SES schools contribute to a decrease in reading, math, and science test scores, while the more cognitively challenging simulations and applications predominately used in high-SES schools contribute to higher test scores in those areas.

In summary, though there has been some progress in equalizing physical access to computers and the Internet, especially in school environments, there still remains broad inequality in the United States in terms of how diverse groups use new technology and the academic outcomes associated with such use.

Equity within Other Countries

The United States is not alone in its differential access to ICT; other economically developed and traditionally more egalitarian countries such as the United Kingdom, Australia, Sweden, Denmark, and Japan also exhibit stratification in access to ICT. As one might expect, developing nations too face similar if not greater challenges in providing equal access to ICT to all its citizens. We briefly review issues of technology and equity in a number of countries, drawing on general societal data when specific educational data are not available.

Canada

Although more Canadians are now online than in previous years (in 2007, 73% of Canadians aged 16 or older went online as compared to 68% in 2005), the digital divide between the rich and the poor continues to exist – 91% of those in the top economic quintile reported using the Internet as compared to only 47% of those in the lowest economic quintile. Age and level of education are also important factors in the unequal use of ICTs. Ninety-six percent of young Canadian adults between the ages of 16 and 24 went online in 2007 as compared to 29% of those aged 65 and older. A full 84% of Canadians with some postsecondary education used the Internet in 2007 as compared to 58% of those who had less education (Canadian Internet Use Survey, 2008).

The numbers on school use reveal other divides. Whereas 98% of public schools in Canada had Internet-connected computers, less than 25% of the schools had a majority of their computers running on the most current operating system. This aging effect of the hardware no doubt has an impact on use as more and more computers breakdown or become obsolete (Information and Communication, 2004). In some cases, though, school computers are the only access students have to ICT. Students in rural areas of Canada are less likely to have home computers than students in urban areas, resulting in higher school use (29%) for rural students than for urban students (19%; Digital Divide, 2003).

The significance of these divisions is all the more troubling in the light of a study on the link between ICT use and literacy skills which showed that people who do not have access to computers also have significantly lower literacy skills than computer users (Study, 2005).

European Union

In 2004, on average, 54% of the European Union (EU) households had a computer and 43% had Internet access. These numbers vary quite a bit from country to country, with 50% of the households in countries such as Denmark, Norway, Luxembourg, Germany, and the Netherlands having Internet access and only about 16% of households in countries such as Latvia, Lithuania, Hungary, Bulgaria, Romania, and Turkey having Internet access. These geographical divides appear to be related to urbanization, with thinly populated, rural areas across Europe having low Internet penetration (Demunter, 2005). Some countries stand out for their especially high or especially low ICT penetration levels. Sweden ranks as one of the highest with 73% of households having Internet access of which 40% are connected via broadband connections. Greece ranks at the bottom with only 24% of households having Internet access of which a mere 6.84% are connected via

Table 2 World Internet usage

<i>Region</i>	<i>Percent of population using the Internet in 2008</i>	<i>Usage growth, 2000–08</i>
Africa	5.6%	1100.0%
Asia	17.4%	474.9%
Europe	48.9%	274.3%
Middle East	23.3%	1296.2%
North America	74.4%	132.5%
Latin America/ Caribbean	29.9%	860.9%
Oceania/ Australia	60.4%	172.7%
World total	23.8%	342.2%

From <http://www.internetworldstats.com/stats.htm> (July, 2009).

broadband connections (Supporting Policy Development for eInclusion, 2005).

Looking at the demographics of the population, as expected, there is a significant age gap in ICT use, with 75% of those under the age of 24 being Internet users, but only 11% of those between the ages of 64 and 74 being Internet users. Level of education is also a factor in ICT use. Those who had completed tertiary education were 2.5–3 times more likely to use the Internet than those who had lower-secondary education. A third important factor was the presence of children in the home. Some 70% of all homes with children had computers, whereas only 46% of childless homes had computers (Demunter, 2005).

India

People in less-developed nations face similar, if not worse, inequities. In India, for example, a boom in the information technology industry has created a handful of multi-millionaire entrepreneurs and a small middle class consisting of software engineers and other ICT workers, but it has had virtually no trickle-down effect on the vast majority of the poor and illiterate. The population of India is over one billion, 41.6% of whom live below the international poverty line (Poverty in India, 2006), 42% (adults) of whom are illiterate, and 70% of whom live in villages (Rao, 2005). These numbers alone could account for the tremendous disparities that exist in ICT access and use within India. But in addition to the factors of poverty and illiteracy, there exist other factors that further polarize the population.

The first of these is the divide that exists between those who speak English and those who do not. In 2005, approximately 5% (50 million) of the Indian population spoke English fluently, the remaining 95% thus being effectively shut out from access to the vast majority (60–80%) of the world's websites, which are in English. Another factor contributing to the divide is the global economy and the consequent emergence of a new elite in Indian society – the

digerati. Unlike the older elite whose privileges were based on caste and inherited wealth, the new digerati owe their position of privilege to education, entrepreneurial skills, and staying on the cutting edge of technology. Finally, there is the issue of infrastructure and cost. The main obstacle to effective penetration and diffusion of ICT is the lack of infrastructure such as reliable power supply, teledensity, and Internet connectivity. India struggles in all three areas. Furthermore, as with many developing countries, India is faced with assessing the cost–benefit balance of expanding ICT penetration when hundreds of millions of citizens lack basic education, healthcare, and adequate nutrition (Keniston, 2003).

China

The technology policies of the Chinese government over the last 10 years have favored cities and schools/universities, ignoring other, more rural geographic areas and groups (nonschool populations). As a result, those in the urbanized, densely populated Eastern provinces have a much higher level of Internet access and use (30–40% of the population) than those in the rural, sparsely populated Western provinces (5% of the population; Hung, 2003).

In addition to geography, age and education also impact the digital divide in China as they do in other countries. The young and those with higher levels of education use the Internet the most. Sixty-nine percent of Internet users are under the age of 30, and 70% have either a high school or a tertiary degree (Statistical Survey Report on the Internet Development in China, 2008). Interestingly though, unlike in the United States, Internet use in China is not concentrated in the highest-income demographic. This is because students are the heaviest users of the Internet, and typically students have low or no income. Whereas ICT is still the privilege of a small group of people in China, Internet access and use are growing rapidly every year. By June 2008, the number of Internet users (aged 6 and above) had reached 253 million (approximately 20% of the total population) – up 35 million from late 2007 – making China number one in the world when it comes to number of Internet users (Statistical Survey Report on the Internet Development in China, 2008).

Equity Across Countries

Internet access varies widely from region to region, with the greatest access being observed in North America and the least being observed in Africa (see **Table 2**). Not surprisingly, those regions with the least access to the Internet are growing at the fastest rate.

At the country level, the International Telecommunication Union developed the Information and Communication Technology Opportunity Index (ICT OI) to measure

(and compare) access and usage of ICT in countries across the globe. Based on this index, 183 economies were measured and divided into four categories, high, upper, medium, and low. The data revealed that greater economic development is correlated with high levels of access to and use of ICT and lower economic development is correspondingly correlated to low levels of access to and use of ICT.

Yet, some countries have nevertheless managed to overcome their economic disadvantage and score relatively higher on the index, probably due to their small land areas (as in the case of small-island developing states such as Jamaica) and/or their high population density (as in the case of Singapore), both of which make the diffusion of access to ICT easier. When measured over time, the growth rate across all categories has been generally increasing, and more importantly, the gap appears to be narrowing somewhat between the low ICT opportunity group and all other groups, possibly due to the combined effect of mobile-phone saturation in the higher groups and a corresponding catch-up effect in the low group (Measuring the Information Society, 2007).

Approaches to Increasing Equity

Although the digital divide has narrowed somewhat both within and across countries, efforts to further increase equity in access are ongoing. These initiatives span all areas of need from hardware and connectivity to software and content. Often these efforts have involved or stimulated debate about what types of technology access and support are needed to benefit people who are impoverished or otherwise socially marginalized, with many suggesting that hardware access is insufficient without attention to issues of content, language, literacy, and social support (see e.g., Dimaggio *et al.*, 2004; Warschauer, 2003).

An early effort to improve ICT equity in schools was the US E-Rate program, launched in 1996 to provide low-cost Internet connectivity to schools as well as to promote and expand Internet access in out-of-school community technology centers such as libraries. Similar initiatives in countries such as the United Kingdom, Israel, and Germany have had positive results (Balanskat, 2002).

Another ongoing effort has been through the creation of public access points for computers and the Internet, known in much of the world as telecenters and in the United States as community technology centers. These centers often combine public access to technology with formal or informal instruction and support (see discussion in Warschauer, 2003). In highly impoverished rural areas such as parts of India, more common are individual computer kiosks, with a single computer operated by a local employee who finds the computer-based or online information the client desires (Warschauer, 2003).

To enable a higher degree of computer access throughout the world, a number of groups have tried to develop specialized low-cost computers that meet the needs of the poor. In the first few years of the twenty-first century, a multi-language handheld computer for the Indian market called the Simputer received a lot of attention, but in the end failed to sell more than a few thousand units as it could not compete in cost with the falling prices of mass-market computers (Multi-Language, Handheld Computer Yet to Reach the Masses, 2005).

The most prominent current effort in this direction is the XO computer developed and distributed by the One Laptop Per Child (OLPC) program. OLPC's goal was to develop a low-cost, high-functioning laptop designed to run on innovative freeware that would be used by individual children throughout the developing world. Since OLPC's inception in 2005, about 300 000 XO computers have been deployed worldwide (Kraemer *et al.*, 2009). This number, though sizable, is far below the original goals of OLPC. Controversies associated with the program include its large minimum-order requirements (originally at one million units, and then reset to 100 000 units), its reversal of its original decision to run exclusively on open-source software (with eventual inclusion of a Windows option), its target population of schoolchildren in developing and impoverished countries (with many believing that these countries have higher social and educational priorities than individual laptops for children), and its neglect of attention to local educational environments and to teachers' professional development (see discussion in Kramer *et al.*, 2009). Despite any shortcomings of the OLPC project, it is widely viewed as bringing about important innovations in hardware and software design and in helping spawn a new generation of low-cost laptop computers known as netbooks and smartbooks.

Research suggests that the provision of individual computers will not in and of itself erase educational divides, as schools in high-SES communities find it easier to integrate laptops into instruction than do schools in low-SES communities, due to the students' greater literacy skills, English language ability, and prior experience with computers; the support for educational computing found in students' homes and neighborhoods; and the support structures for educational technologies in the schools (Warschauer, 2006). This has been labeled the Sesame Street Effect (see Attewell and Winston, 2003), referring to a more general phenomenon in which technology-based reforms targeted at low-income populations are often best exploited by higher-income populations that can better leverage their preexisting educational, social, and cultural resources. Some initiatives have thus tried to go beyond hardware or connectivity to provide more suitable open-access educational resources. For example, Massachusetts Institute of Technology's (MIT) Open Courseware project has placed the material from some 1800 undergraduate and graduate courses online that

has been accessed by more than 50 million people (Reiddy, 2008). Beyond these initiatives, it is widely believed that professional development, support systems, and curriculum development are all vital means to make technology access meaningful, and are thus required for lessening any divides. Efforts to make improvements in these areas generally take place at the local or state level.

One organization that seeks to promote and coordinate initiatives internationally is the Global Alliance for ICT and Development (GAID), established in 2006 by the United Nations, Department of Economic and Social Affairs. GAID's four main areas of focus for 2008–10 are access, connectivity, content, and education, reflecting the growing recognition that a broad, multifaceted approach is needed to improve educational and social development with technology.

Conclusion

Over the last 10 years, there have been many shifting challenges to providing equal access to ICT. In the United States and other developed countries, the discussion has moved from inequities in physical access and connectivity to the more nuanced issues of meaningful and purposeful use. In developing countries, the majority of people lack even basic access to computers and the Internet, but extending such access for educational purposes has to be balanced with a range of other social and educational priorities.

See also: Technology and Physical and Social Health; Technology as a Support for School–Community Connections.

Bibliography

- Attewell, P. and Battle, J. (1999). Home computers and school performance. *Information Society* 15(1), 1–10.
- Attewell, P. and Winston, H. (2003). Children of the digital divide. In Attewell, P. and Seel, N. M. (eds.) *Disadvantaged Teens and Computer Technologies*, pp 117–136. Münster: Waxmann.
- Balanskat, A (2002). Digital divide. <http://insight.eun.org> (accessed September 2009).
- Becker, H. J. (2000). Who's wired and who's not: Children's access to and use of computer technology. *Future of Children* 10(2), 44–75.
- Beltran, D. O., Das, K. K., and Fairlie, R. W. (in press). Are computers good for children? The effects of home computers on educational outcomes. *Economic Inquiry*.
- Canadian Internet Use Survey (2008). Statistics Canada. <http://www.statcan.gc.ca/daily-quotidien/080612/dq080612b-eng.htm> (accessed September 2009).
- Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2008). Scaling the digital divide: Home computer technology and student achievement. http://www.hks.harvard.edu/pepg/PDF/events/colloquia/Vigdor_ScalingtheDigitalDivide.pdf (accessed September 2009).
- Demunter, C. (2005). The digital divide in Europe. Statistics in focus. Eurostat. <http://epp.eurostat.ec.europa.eu> (accessed September 2009).
- Digital Divide (2003). Digital divide in schools: Student access to and use of computers (2003). Statistics Canada. <http://www.statcan.gc.ca/daily-quotidien/030623/dq030623b-eng.htm> (accessed September 2009).
- Dimaggio, P. J., Hargittai, E., Celeste, C., and Shafer, S. (2004). Digital inequality: From unequal access to differentiated use. In Neckerman, K. (ed.) *Social Inequality*, pp 355–400. New York: Russell Sage Foundation.
- Horrigan, J. (2008). Home broadband adoption. <http://www.pewInternet.org> (accessed September 2009).
- Hung, C. (2003). Income, education, location and the Internet – the digital divide in China. *Paper Presented at the Annual Meeting of the International Communication Association*, San Diego, CA: Marriott Hotel. <http://www.allacademic.com> (accessed September 2009).
- Information and Communication (2004). Information and Communication, technologies in schools survey, Statistics Canada. <http://www.statcan.gc.ca/dli-ild/data-donnees/ftp/ictss-etice-eng.htm> (accessed September 2009).
- Ito, M., Horst, H., Bittanti, M., et al. (2008). Living and learning with new media: Summary of findings from the digital youth project. <http://digitalyouth.ischool.berkeley.edu> (accessed September 2009).
- Keniston, K. (2003). Introduction: The four digital divides. In Keniston, K. and Kumar, D. (eds.) *IT Experience in India: Bridging the Digital Divide*, pp 11–36. New Delhi: Sage.
- Kraemer, K. L., Dedrick, J., and Sharma, P. (2009). One laptop per child: Vision vs. reality. *Communications of the ACM* 52(6), 66–73.
- Lenhart, A., Kahne, J., Middaugh, E., et al. (2008). Teens, video games, and civics. <http://www.pewInternet.org> (accessed September 2009).
- Measuring the Information Society (2007). International Telecommunication Union. <http://www.itu.int/net/home/index.aspx> (accessed September 2009).
- Multi-Language, Handheld Computer Yet to Reach the Masses (2005). Associate Press. <http://www.msnbc.msn.com/id/7385216/> (accessed September 2009).
- National Telecommunications and Information Administration (2008a). *Networked Nation: Broadband in America 2007*. Washington, DC: NTIA.
- National Telecommunications and Information Administration (2008b). *Households Using the Internet in and Outside the Home, by Selected Characteristics: Total, Urban, Rural, Principal City, 2007*. http://www.ntia.doc.gov/reports/2008/Table_HouseholdInternet2007.pdf (accessed September 2009).
- Poverty in India (2006). World Bank. <http://web.worldbank.org> (accessed September 2009).
- Rao, S. S. (2005). Bridging digital divide: Efforts in India. *Telematics and Informatics* 22(4), 361–375.
- Reiddy, C. (2008). MIT open courseware. <http://www.boston.com> (accessed September 2009).
- Statistical Survey Report on the Internet Development in China (2008). China Internet Network Information Center. <http://www.cnnic.cn/uploadfiles/pdf/2008/2/29/104126.pdf> (accessed September 2009).
- Study (2005). Study: The link between information and communication technology use and literacy skills. Statistics Canada. <http://www.statcan.gc.ca/daily-quotidien/051205/dq051205b-eng.htm> (accessed September 2009).
- Supporting Policy Development for elclusion (2005). EU Country Reports. <http://www.ipolicy.eu> (accessed September 2009).
- Warschauer, M. (2000). Technology and school reform: A view from both sides of the track. *Education Policy Analysis Archives* 8(4). <http://epaa.asu.edu/epaa/v8n4.html> (accessed September 2009).
- Warschauer, M. (2003). *Technology and Social Inclusion: Rethinking the Digital Divide*. Cambridge: MIT Press.
- Warschauer, M. (2006). *Laptops and Literacy: Learning in the Wireless Classroom*. New York: Teachers College Press.
- Warschauer, M., Knobel, M., and Stone, M. (2004). Technology and equity in schooling: Deconstructing the digital divide. *Educational Policy* 18(4), 562–588.

- Warschauer, M. and Matuchniak, T. (in press). New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. *Review of Research in Education*.
- Wells, J., Lewis, L., and Greene, B (2006). Internet Access in U.S. Public Schools and Classrooms: 1994–2005. Washington, DC: NCES.
- Wenglinsky, H. (1998). Does it compute? The relationship between educational technology and student achievement in mathematics. *ETS Policy Information Center*.
- Wenglinsky, H. (2005). *Using Technology Wisely: The Keys to Success in Schools*. New York: Teachers College Press.
- World Internet usage. <http://www.Internetworldstats.com> (accessed September 2009).

Relevant Websites

- <http://un-gaid.org> – Global Alliance for ICT and Development (GAID).
- <http://insight.eun.org> – Insight.
- <http://www.itu.int> – International Telecommunication Union.
- <http://www.Internetworldstats.com> – Internet World Stats.
- www.pewInternet.org – Pew Internet and American Life Project.
- <http://www.statcan.gc.ca> – Statistics Canada.
- <http://web.worldbank.org> – The World Bank.
- <http://www.gse.uci.edu/markw> – University of California, Irvine, Department of Education.

Internet-based Education

M Bakia, SRI International, Arlington, VA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Affordances or, specifically, learning affordances – The features or characteristics of curriculum and software that are specifically designed to allow for particular types of action.

Asynchronous – A communication style that does not require participants to engage in dialog simultaneously, that is, participants can engage whether or not a participant is present or available at a particular time. Postal communication and e-mail are two examples of asynchronous communication.

Avatar – An object representing the embodiment of the user on a computer, which could be in the form of a three-dimensional model used in computer games, a two-dimensional picture used on Internet forums and other communities, or a textual representation.

Credit recovery – This refers to a student's need to retake courses either failed earlier or not yet attempted but needed to advance to the next grade or to graduate. Credit-recovery courses and programs are, generally, geared toward students at risk of dropping out of school.

Dual credit – The courses taken by secondary students, often at an institution of higher education, that allows a student to earn credit – both toward high school graduation and for credit at colleges and university – often, giving students advanced standing when they enter college. Dual-credit courses that are not offered by an institution of higher education have, typically, gone through an articulation-and-approval process that certifies the course as equivalent in scope and rigor to related college courses.

Learner control – The ability of students to influence the pace and sequencing of their curriculum. For example, they can engage in text in a self-determined order or pause, rewind, or fast-forward video segments.

Synchronous – A Communication style that requires participants to engage in dialog, more or less, at the same time. Face-to-face and telephone-mediated dialogs are two examples of synchronous communication.

Internet-Based Education Defined

Internet-based education is referred to by multiple terms, including online learning, e-learning, cyber learning, and Web-based learning. (Internationally, and particularly in developing countries, the term open schooling is also used. Although open schooling traditionally relies on radio, television, and correspondence technologies, Internet-based resources are increasingly common as Internet access expands around the world.) More recently, the term open educational resources (OER) is being used to represent content, tools, and software that are freely available on the Web for educational purposes (Smith, 2009). These terms are used somewhat interchangeably both within the education community and in the text that follows. As the diversity of terms might suggest, it is important to note that a great variety of administrative practices and requirements, content and instructional strategies, technology requirements, and costs are associated with Internet-based education. Internet-based education is offered by public, private, and nonprofit entities and may operate locally, regionally, nationally, or internationally.

Higher education – and to a lesser extent the corporate training industry – were principal drivers of online learning, starting in the mid-1990s as Internet access began to take hold in many places around the world (Mandinach, 2005). The example of just one university in the US demonstrates the impact that Internet-based educational resources can have internationally. In 2000, the Massachusetts Institute of Technology (MIT) made core content – such as syllabi, reading lists, and video-based lectures – available freely on the Web. Since then, 90 universities and colleges from 14 countries have adopted or adapted MIT's resources for local use (Smith, 2009).

Nearly 4 million students enrolled in at least one online course in the US in the fall of 2007, representing a little over 20% of the total student population in higher education (Allen and Seaman, 2008). There are many examples of online education in higher education around the globe. Universiti Putra Malaysia launched its first entirely online program in December of 1999 with 200 students, and serves other countries in the region – including Cambodia, Myanmar, Mauritius, Singapore, and Sri Lanka. Similarly, Libya, Morocco, Sudan, and Tunisia all cite the development of virtual or open universities in Africa (Farrell and Isaacs, 2007), following the development of virtual and open universities throughout Europe and Asia.

The practice of online learning is more recently being applied to elementary and secondary education – which is the primary focus of this article. Examples of elementary and secondary uses of Internet-based education may include primary-grade students working on beginning reading skills over the Internet, middle-school students collaborating with practicing scientists in the design and conduct of research, and high school dropouts taking courses online to recover credits and qualify for graduation. In addition, entire secondary school programs provided over the Internet are increasingly available – such as the charter and contract schools affiliated with K12 Inc., Connections Academy, and Insight Schools. The format and media of Internet-based approaches range from conventional didactic lectures or textbook-like information delivered over the Web, to Internet-based collaborative role playing in social simulations and highly interactive strategy games.

There are two fundamental types of Internet-based education: one replaces existing face-to-face courses and the other is intended to enhance face-to-face courses. Internet-based education that is designed to replace face-to-face courses is, typically, seen as an evolution of distance education – an instructional arrangement in which a student and teacher are separated by time and space. Correspondence courses were among the earliest forms of distance education, and distance education has continued to adopt the latest communications technologies available. In the case of online learning, the Internet has replaced earlier technologies – such as the postal service, radio or television – to deliver instruction at a distance.

A recent and increasingly common practice in secondary and tertiary education is to use Internet-based resources to enhance face-to-face courses. This style of instruction is referred to as a blended or hybrid approach. Blended courses are, typically, defined as face-to-face courses that include a significant online component. These activities are similar to earlier uses of computer-based instructional materials that pre-date widespread access to the Internet in secondary schools. As computing and graphical properties of computers in schools have advanced, enhanced digital resources that are retrieved via the Internet are increasingly common in traditional classrooms. A related practice is emerging in which online courses incorporate significant face-to-face elements in order to enhance the relationships and learning among teachers and students. Similarly, blended programs allow students to take both online and face-to-face courses within a given degree program. This variety of blended approaches appears to be more common in secondary education than in higher education.

Researchers have begun to suggest online learning classification systems which account for a full range of practices from entirely face to face to entirely online. Blended or hybrid models – such as those described above – are, often, defined as having anywhere from 25% to 80% of instruction

occurring online, depending on the classification system. The term Web-facilitated instruction has also been introduced to describe courses with less intensive use of Web-based technologies within face-to-face courses. Examples of this type of use comprise including Web pages in course readings, e-mail communication between teachers and students, and course-management systems where syllabi and other resources are available online.

Technologies that allow students and teachers to communicate at the same time (e.g., via telephone) are called synchronous, while those that rely on sequential communication over time (e.g., posted letters or e-mail) are called asynchronous. A variety of new instructional features and capabilities (sometimes referred to as affordances) are now possible with the advent of Internet-based instruction – including a blend of synchronous and asynchronous approaches.

The latest developments in online learning include multimedia materials, allowing for simulations, visualizations, and auditory reinforcement of key concepts. Some programs have also begun to incorporate avatars or virtual tutors – which depict animated characters that mimic human facial expressions to communicate emotional support or concern. For example, WiloStar3D allows enrolled students to experience an immersive graphical environment depicting a virtual campus. Students take courses in Internet-based three-dimensional (3D) classrooms. Although the experience for students is clearly two-dimensional, graphics are rendered to give the illusion of a 3D space. Teachers are represented in this space as human-like avatars. Similarly – at the Florida Virtual School (FLVS) – avatars are being introduced specifically to display empathetic emotions to the learner. They provide audio when students click on the avatar picture and also exhibit body movements and facial expressions (Reeves *et al.*, 2007). Some developers have suggested that this humanizes online learning and may increase student motivation, particularly for younger students who have not developed skills sufficiently for working independently. Similarly, online learning can take advantage of a variety of communication methods – such as online chat, e-mail, and audio- and video-conferences.

Prevalence of Internet-Based Education in Secondary Education

Online Learning in North America

Although many countries have Internet-based education programs, the largest online learning programs for secondary school students appear to be in North American countries – with the United States and Canada both having large numbers of secondary students enrolled in a wide variety of online courses and programs (Cavanaugh and Clark, 2007). The United States appears to be

one of the few countries where multiple, national surveys of online educational have been conducted. One study conducted by the US Department of Education looked more generally at distance education, and not at Internet-based education specifically. Based on district-survey responses, it estimated that distance-education enrolments in elementary and secondary schools in the school year 2004–05 were almost 507 000 – up from 328 000 in the school year 2002–03 (Zandberg and Lewis, 2008). Districts reported that the use of the Internet was on the rise, with about 70% of districts reporting student enrolments in technology-based distance education also reporting that at least some of these students took courses online. A more recent study estimated that 1 030 000 K12 students took one or more online or blended courses during the 2007–08 school year (Picciano and Seaman, 2008). However, separate estimates were not provided for online or blended courses. (These estimates should be viewed cautiously since the study obtained a response rate of less than 10%.) Enrolment patterns in the US suggest that most students who participate in formal Internet-based education take one or two courses online while enrolled in traditional face-to-face, full-time programs (Zandberg and Lewis, 2008).

Another way to gauge online learning activity is to track government policies that support Internet-based education. In the United States, 44 states supported sizeable Internet-based learning opportunities for students, in the form of online courses, in 2008 (Watson *et al.*, 2008). Twenty-one of these states supported full-time, online schools – which were typically either state-sanctioned charter schools or districts within a state that offered courses to students beyond their traditional geographical borders. The remaining states provided supplemental online courses for brick-and-mortar schools within the state. One of the best-known state-funded programs is the FLVS, which became the first statewide, public virtual school in the United States in 1997. In 2008–09, FLVS served an estimated 64 000 middle- and high school students (Young *et al.*, 2009). In addition to state-led programs, ~40% of high schools offered technology-based distance-education courses in 2004–05 – the most recent year for which estimates are available.

Many Canadian provinces and local districts also actively support Internet-based education. For example, Ontario – a province that is particularly active in this arena – has identified high-quality online learning as an official strategy for increasing high school-graduation rates. Ontario supported more than 50 online courses during the school year 2008–09, many of them in core subjects – such as language, arts, and mathematics. In addition to offering full courses, Ontario maintains a database with thousands of learning objects that include lesson plans for teachers and other multimedia artifacts for use in classrooms or as instructional supplements.

International K12 Projects of Note

Many other countries are also supporting Internet-based education (Powell and Patrick, 2006), although K12 student participation does not appear as widespread as in North America. Australia has been a long-time leader in the use of distance education – suggesting fertile ground for Internet-based education. Where online learning is an explicit policy to replace face-to-face courses, the priority is placed on serving rural, remote communities. In addition, the Learning Federation is a large-scale, joint effort between Australia and New Zealand to create digital content for schools. Efforts focus on the creation of Internet-based artifacts and learning modules to be used in traditional settings or as supplemental activities for students.

Another example of international adoption of Internet-based education is European SchoolNet – an international partnership of 31 European ministries of education that coordinates and supports a variety of educational projects that rely on the Internet, including efforts to develop common standards and specifications for European countries. Another Schoolnet project – CALIBRATE – has sought to develop a common platform for Internet-based teaching materials that could be searched by teachers across the continent. Among European countries, England is often seen as a leader, particularly through the British Educational Communications and Technology Agency (BECTA) – the government agency responsible for effectively integrating innovative technology applications throughout the educational system.

Singapore – a long-time international leader in the use of technology in K12 education – uses Internet-based resources, primarily, to enhance face-to-face education. For example, several web-portals are in use to teach young children how to speak second languages (Hawthorne and Tan, 2005). Similarly, Internet-based objects are replacing earlier reliance on compact disks-read only memory (CD-ROMs) and packaged software. In addition, approximately 75% of Singapore's schools had subscribed to a learning management system by 2006 (Powell and Patrick, 2006).

Countries not typically associated with high-tech adoption also use the Internet for educational purposes. For example, Iran supports four national schools which offer courses online. Courses are developed by the government and provided at no cost to students.

Common Purposes for Online Learning in K12 Education

The popularity of Internet-based education is premised on its potential to offer access to content and instructional feedback anytime and anywhere that an Internet connection is available. Others suggest that well-designed Internet-based

education may improve student outcomes by increasing the individualization of curricula to match the strengths and interests of students, as well as by expanding the range of resources available. Although no systematic surveys have included detailed questions with regard to the size and purposes of Internet-based education, conventional wisdom suggests that online learning is most valuable to particular types of students and for particular subject areas. Typical applications vary depending on whether Internet-based education is being used to replace face-to-face courses or to enhance face-to-face instruction.

Online Learning as Educational Replacement

Online courses intended to replace face-to-face courses have tended to be promoted for rural areas, for advanced courses, and to allow at-risk students to meet graduation requirements.

In rural areas

Rural areas have been targeted users of online courses and programs, following a tradition of distance education in these areas. Online learning has been promoted for rural areas where qualified teachers may not be available, especially in areas of national shortage such as mathematics and science. Small, rural schools may not have enough interested or prepared students to warrant the costs of offering some specialized or advanced courses in their curriculum. Research has shown that rural schools do not provide extra help and support needed to enter and succeed in higher-level mathematics courses, for example, and high schools in rural communities are less likely to offer advanced placement (AP) coursework than schools in suburban or urban locations. Individuals and small groups of students can join together across disparate locales to participate in a virtual course; qualified instructors can be drawn from the faculty at multiple high schools; and per-student course-costs can be reduced if they are amortized over larger numbers of students.

For advanced courses

Even in suburban and urban areas, online instruction has been promoted for advanced courses, especially for AP courses. High school students, often, take AP courses in order to demonstrate college-readiness or to gain advanced standing as college freshman. The supply of qualified AP teachers has not been able to keep up with the demand, however, with shortages most acute in urban as well as rural areas. There is evidence that advanced courses do, in fact, account for a sizable proportion of online enrolments. One study estimated that 25% of US districts with students enrolled in technology-based distance education had students enrolled in AP courses offered through distance education in 2004–05, and 40% of districts with technology-based distance education had

secondary students enrolled in dual-credit college-level courses offered through distance education. The combined enrolments in AP and dual-credit college-level technology-based distance-education courses accounted for 15% of all technology-based distance-education enrolments in 2004–05.

For at-risk students

Online learning providers are also beginning to identify students in need of remediation as likely beneficiaries of online learning. Nearly 20% of FLVS students are reported to be either at risk or in need of credit recovery – a term used to refer to the need to retake courses previously failed or to accumulate more course credits to meet graduation requirements (Young *et al.*, 2009). In the past, distance education, in general, has been shown to benefit nontraditional students, especially those with daytime responsibilities, since asynchronous learning allows for flexible scheduling. However, poorly performing students are a relatively new focus of online learning providers. Previously, only academically successful secondary students were thought to have the skills and motivation necessary to work independently. However, online learning providers argue that online curriculum that increases individualization and provides specialized support to non-traditional or poorly performing students can keep some students in the education system who would otherwise be lost. (As mentioned above, serving these students is an explicit policy objective of the Ontario government in Canada.)

Online Learning as Educational Enhancement

While the use of the Internet to enhance education takes myriad forms, including course-management systems that serve as electronic repositories as well as activities that encourage cross-cultural communication and exploration, tools and resources for science seem to dominate online enhancement materials at the K12 level. One example is River City – a science-replacement unit focusing on scientific inquiry. River City provides a 20-day curriculum – in which, students spend 14 days in an immersive, online environment and 6 days working on classroom activities. Approximately 7000 students used River City during school year 2005–06. Similarly, Web-based Information Science Education (WISE) consists of replacement modules that help students visualize difficult scientific concepts. The program targets middle- and high school science courses. Bugscope is a third example of an Internet-based supplement for classroom-based instruction. Bugscope provides science courses with access to an electron scanning microscope. Students interact online with scientists who operate the microscope while dissecting an insect. All of these examples suggest that the Internet can provide resources

and experiences that would not, otherwise, be available in the classroom.

Research on the Effectiveness of Internet-Based Education

This section focuses on two common metrics of the effectiveness of online instruction: student retention and learning outcomes. Data provided primarily address research with regard to Internet-based education that is intended as a replacement for traditional, face-to-face instruction, although the text regarding learning outcomes also includes data that address the effectiveness of blended approaches.

Student Retention

Research has documented that retention rates are lower for online courses than for face-to-face courses – sometimes dramatically so. This is true across educational levels and is consistent with the distance-education literature generally. For online college and university undergraduates, studies show that the retention rate can range from 50% to 80%, with online course administrators reporting that the retention rate for online courses is 10–20% lower than that for traditional classroom environments (Herbert, 2006). A survey of state online course-providers in the United States showed a considerable range of reported retention rates among 32 providers. Retention rates for online courses varied from 50% to 99%, with most in the range from 65% to 85%. Reported retention rates for online programs (as opposed to courses) were similar. Respondents reported program retention rates between 60% and 95%, with most programs reporting retention somewhere between 70% and 85% (Watson and Ryan, 2007). Several reasons for the relatively high dropout rate in online courses and programs have been suggested, including difficulty accessing technology, the relative immaturity of students, and competing demands – particularly for nontraditional students who may be working or have increased family responsibilities (Moore *et al.*, 2002).

Student Learning Outcomes

Reviews of the distance-education literature (Cavanaugh, 2001; Moore, 1994; Russell, 1999), generally, suggest that there is no significant difference between learning outcomes for distance and face-to-face education. Usually, the lack of a significant positive difference is viewed negatively in research on educational interventions, but because distance education has its roots in efforts to supply services where face-to-face instruction is not possible, the lack of an effect is viewed positively. However, it

is important to note that the general finding that there is no significant difference between distance and face-to-face models of instruction masks large differences in findings across specific studies. A recent, large-scale meta-analysis found tremendous variability in effect sizes, which ranged from -1.31 to $+1.41$ across 232 studies included in the review (Bernard *et al.*, 2004).

A recent review of research studies, conducted between 1996 and 2007, that focused on the learning effects of Internet-based education specifically (Means *et al.*, 2009) found few experimental or controlled quasi-experimental research studies involving K12 students. Only five studies involving secondary and middle-school subjects and meeting accepted quality standards for rigorous meta-analysis were identified. Each of these studies compared student learning in a blended condition – one that included both face-to-face and Internet-based components – to a face-to-face only condition. These five studies addressed eighth-grade student achievement in social studies, eighth- and ninth-grade Algebra 1, middle-school students in Spanish language courses, and a study of fifth-grade student in science classes in Taiwan. Across the five studies, there were seven independent learning outcomes measured for both blended learning and control students. Results were mixed: Three of the comparisons produced significant effects favoring a blended learning condition, one had a significant negative effect favoring face-to-face instruction, and three contrasts did not attain statistical significance. The results for the larger group of studies involving learners of all ages were less equivocal: students in online learning conditions (either purely online or a blend of online and face-to-face elements) performed not just as well as, but better than, students receiving conventional face-to-face instruction on average.

The same group of studies offers evidence that particular design elements of Internet-based education influence student achievement. Earlier meta-analyses have found an advantage for asynchronous over synchronous communication, with certain practices being more effective in one communication format over the other (Bernard *et al.*, 2004). For example, face-to-face meetings with instructors and telephone contact with an instructor were associated with higher student achievement in the studies of synchronous distance education but not in those of asynchronous distance learning.

There is also some suggestion that the ability of the learner to control learning media is important. Increased learner control suggests that students can influence the pace and sequencing of the curriculum. For example, if a video is included in a Web-based module, systems that allow learner control let students pause, rewind, or skip to particular segments of the video as opposed to needing to passively view an entire presentation from beginning to end. Other methods for engaging active student participation in

learning – including soliciting student reflection and self-monitoring of comprehension – may also contribute to student achievement. On the other hand, practices that promote superficial or relatively passive learning experiences for students do not appear to be associated with improved student achievement (Zhang, 2005).

Summary

Internet-based education is a fast-growing practice in secondary education. Based on the available research, Internet-based education appears to be growing at exponential rates. It continues to evolve in many different ways and represents a tremendous variety of policies, practices, and learning objectives. Internet-based resources may be used to augment an otherwise traditional face-to-face course or an entire course or program of courses may be offered entirely online.

The use of Internet-based education has far outpaced research with regard to its effectiveness – particularly for younger students in elementary and secondary education. The handful of rigorous experimental or quasi-experimental studies that have been conducted suggest that Internet-based learning is usually at least as effective as traditional classroom instruction. It is not uncommon, in fact, for students to learn more when part of their instruction takes place online. The most promising online learning practices seem to involve blending asynchronous and synchronous styles of interaction and incorporate both face-to-face and Internet-based instructional components.

See also: Equity in Technology Access and Opportunities; Technological Supports for Acquiring Twenty-First-Century Skills; Technology Supports for Lifelong Learning; Technology Supports for Science Learning.

Bibliography

- Allen, I. E. and Seaman, J. (2008). *Staying the Course: Online Education in the United States*. Needham, MA: Sloan Consortium.
- Cavanaugh, C. and Clark, T. (2007). The landscape of K12 online learning. In Cavanaugh, C. and Blomeyer, R. (eds.) *What Works in K12 Online Learning*. Washington, DC: International Society for Technology in Education.
- Farrell, G. and Isaacs, S. (2007). Survey of ICT and education in Africa. *A Summary Report, Based on 53 Country Surveys*. Washington, DC: infoDev/Worldbank.
- Hawthorne, E. and Tan, S. C. (2005). Learning in Singapore primary schools. *Innovate Journal of Online Education* **1**(3), [online].
- Herbert, M. (2006). Staying the course: A study in online student satisfaction and retention. *Online Journal of Distance Learning Administration* **9**(4), [online].
- Mandinach, E. B. (2005). The development of effective evaluation methods for e-learning: A concept paper and action plan. *Teacher College Record* **107**(8), 1814–1835.

- Means, B., Toyama, Y., Murphy, R., Bakia, M., and Jones, K. (2009). *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online-Learning Studies*. Washington, DC: U.S. Department of Education.
- Moore, K., Bartkovich, J., Fetzner, M., and Ison, S. (2002). Success in cyberspace: Student retention in online courses. *Paper Presented at the Annual Forum for the Association for Institutional Research*, 42nd, Toronto, Ontario, Canada.
- Picciano, A. G. and Seaman, J. (2008). *K12 Online Learning: A 2008 Follow-Up of the Survey of U.S. School District Administrators*. Needham, MA: The Sloan Consortium.
- Powell, A. and Patrick, S. (2006). *An International Perspective of K12 Online Learning: A Summary of the 2006 NACOL International E-Learning Survey*. Washington, DC: North American Council for Online Learning.
- Reeves, S., Vangelis, M., Vevera, L., and Jensen, V. (2007). Teaching and learning mathematics online: How Florida virtual school builds community through established practices. In Cavanaugh, C. and Blomeyer, R. (eds.) *What Works in K12 Online Learning*. Washington, DC: International Society for Technology in Education (ISTE).
- Watson, J., Gemin, B., and Ryan, J. (2008). *Keeping Pace with Online Learning: A Review of State-Level Policy and Practice*. Evergreen, CO: Evergreen Consulting Associates.
- Watson, J. and Ryan, J. (2007). *Keeping Pace with Online Learning: A Review of State-Level Policy and Practice*. Evergreen, CO: Evergreen Consulting Associates.
- Young, J., Birtolo, P., and McElman, R. (2009). Virtual success: Transforming education through online learning. *Learning and Leading with Technology* 12–17 [online].
- Zandberg, I. and Lewis, L. (2008). Technology-based distance education courses for public elementary and secondary school students: 2002–03 and 2004–05. (NCES 2008-008). Washington, DC: National Center for Education Statistics, Institute for Education Sciences, U.S. Department of Education.
- Zhang, D. (2005). Interactive multimedia-based e-learning: A study of effectiveness. *American Journal of Distance Education* **19**(3), 149–162.

Further Reading

- Cavanaugh, C. and Blomeyer, R. (eds.) (2007). *What Works in K12 Online Learning*. Washington, DC: International Society for Technology in Education.
- Englert, C. S., Zhao, Y., Dunsmore, K., Collings, N. Y., and Wolbers, K. (2007). Scaffolding the writing of students with disabilities through procedural facilitation: Using an Internet-based technology to improve performance. *Learning Disability Quarterly* **30**(1), 9–29.
- Long, M. and Jennings, H. (2005). *“Does It Work?”: The Impact of Technology and Professional Development on Student Achievement*. Calverton, MD: Macro International.
- North American Council for Online Learning (NACOL) (n.d.) *National Standards of Quality Online Courses*. Vienna, VA: NACOL.
- O'Dwyer, L. M., Carey, R., and Kleiman, G. (2007). A study of the effectiveness of the Louisiana Algebra I online course. *Journal of Research on Technology in Education* **39**(3), 289–306.
- Rockman, et al. (2007). *ED PACE Final Report*. Submitted to the West Virginia Department of Education, San Francisco: Rockman, et al. www.rockman.com/projects/146.ies.edpace/finalreport (accessed August 2009).
- Sun, K., Lin, Y., and Yu, C. (2004). A study on learning effect among different learning styles in a web-based lab of science for elementary school students. *Computers and Education* **50**(4), 1411–1422.
- US Department of Education, Office of Innovation and Improvement (2007). *Connecting Students to Advanced Courses Online: Innovations in Education*. Washington, DC.
- Watson, J. (2008). *Blended Learning: The Convergence of Online and Face-to-Face Education*. Washington, DC: North American Council for Online Learning.

Watson, J. and Gemin, B. (2008). *Using Online Learning For At-Risk Students and Credit Recovery*. Washington, DC: North American Council for Online Learning.

<http://www.inacol.org> – International Association for K12 Online Learning.

<http://www.netc.org> – K12 Online Education.

<http://www.thelearningfederation.edu.au> – The Learning Federation.

<http://www.wilostar3d.com> – WiloStar3D.

Relevant Websites

<http://www.elearningontario.ca> – e-Learning Ontario.

<http://www.europeanschoolnet.org> – European Schoolnet.

TECHNOLOGY AND LEARNING – ISSUES

Contents

Media Use and School Achievement

Technology and Physical and Social Health

Media Use and School Achievement

K Roe, Katholieke Universiteit Leuven, Leuven, Belgium

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Does media use have negative effects on literacy and academic achievement? The question is an old one. Concern among educators that media communication is detrimental to the work of the school, in general – and to the development of children's literacy, in particular – can be traced back to at least to concern over popular literature in the mid-nineteenth century. Not surprisingly, then, the question of the nature of the relationship between educational achievement and the mass media is one that has been around since mass communication became a discernible area of study.

Historical Background

In the 1920s, the question was posed by the Payne Fund Studies into the effects of movies on children. The results suggested that movies provide a special learning format that facilitates an unusually high retention of factual material, but that movie fans did worse in their academic work and were less well-behaved and less popular in school than members of the comparison group.

The next major (and rather less objective) treatment came, in 1953, as part of Wertham's broadside attack against comic books. In this instance, the reader does not need to go further than the inside flaps of the cover to learn that comic books are viewed as an invitation to illiteracy and, "harm the development of reading from the lowest level of the most elementary hygiene to the highest level of appreciation of good literature."

On the whole – except in very broad mass versus high culture terms – radio seems curiously to have escaped the odium of critics. Where found, criticism tended to focus on cultural passivity and vulgarization. During the 1950s, however, it was television (TV)'s turn to be examined and, as early as 1951, empirical studies began to appear. Most early research not only failed to find much evidence

for negative effects on school work, some studies actually stressed TV's positive influence, conclusions subsequently reinforced on both sides of the Atlantic. In 1958, one major British study concluded that, on the whole, TV appeared to have no direct effects on school work, nor did it lead to poor concentration, a reduction of interest in school, or a correlation with reading skills. In 1961, an American study reported that watching TV could contribute to a fast start in learning because it helps broaden children's vocabularies, although the advantage gained (over nonviewers) was only temporary.

Although studies reporting negative relationships between TV and scholastic achievement did appear in the 1960s, the general consensus was that there were no simple, direct negative effects – with the result that interest in the topic began to wane. However, as TV became omnipresent in the developed world and cable, satellite and video developments led to a multiplicity of channels and contents, and the subject, once more, became topical. A good deal of the credit for this revival of interest can be attributed to Neil Postman – whose polemical books and articles set up the school as a vital bastion of traditional values against the (alleged) disastrous effects of TV. Although Postman's rhetoric was impressive and his ideas, at times, seductive, only a minority of studies could still be said to have given unqualified support to his position.

The mid-1980s to the mid-1990s witnessed another period of relative neglect – as research attention was diverted elsewhere. However, the last decade has, once again, seen a renewal of interest – undoubtedly, as a result of developments in digital technology. This time, however, the parameters of the discourse have been opened up to include not only specific forms of communication (such as computer games and the Internet), and specific skills (such as reading and writing) but the whole concept of literacy and what being literate in the digital world of the twenty-first century actually means. One consequence of these developments is an increasing tendency to differentiate the concepts of literacy and achievement into

various subdomains such as computer literacy, media literacy, etc. According to some observers, digital developments may be transforming the very way in which we communicate – a process which is moving us inexorably toward an entirely new form of literacy in which authorship and reading online will not involve anything resembling current literacy practices. As a result, it is argued, two ideologies of literacy are now in mortal competition: one encoded in books and traditional pedagogics, the other located in the media, computer games, and the Internet – with the educational system caught in the crossfire. In other words, the traditional praxis of the school – based on hierarchical access to print – is, increasingly, at odds with the more diverse, interactive, and less linear forms made available by digital technologies.

Theoretical Perspectives

Based predominantly on studies of TV and school achievement, four major theoretical perspectives have been identified.

The first is displacement theory – which basically states that TV-viewing takes time away from reading, schoolwork, and other activities beneficial to children's intellectual development. However, the earliest studies in both the United States and Europe failed to find consistent evidence for such an effect. Rather, it appears that other mass media and play activities are TV's main competitors and children who neglect reading because of TV would neglect reading in favor of something else if TV were not available.

The second perspective is information-processing theory which postulates that TV (and by extrapolation other electronic media) influences the way people think. Commonly, while reading is assumed to activate the mind, TV-viewing is assumed to be mentally pacifying. However, there has been little, if any, convincing empirical support for this hypothesis.

The third is short-term gratifications theory. This perspective argues that TV and computer games have radically altered the expectations which children have with regard to learning, most notably in the demand for novel, fast-paced, and constant stimulation – expectations which are antithetical to the slower pace, and deferred gratifications, of school-based education. There is much anecdotal and observational evidence for this hypothesis but it has seldom been tested rigorously.

The fourth, and only positive, perspective is interest-stimulation theory – which hypothesizes that the media can enhance learning by creating and stimulating interests and augmenting knowledge. However, while some early studies indicated that TV might stimulate reading by awakening interests in new subjects, more recently, there is a dearth of empirical support for this assertion.

Research

Although there has never been consistent, conclusive evidence that TV displaces reading and has a negative effect on general school achievement, by the 1980s, a substantial number of studies had reported negative correlations between reading achievement and TV viewing. However, most of these studies were based on cross-sectional data employing bivariate analyses that failed to control for possible mediating variables and were, thus, unable to address the issue of causality. These shortcomings led to more sophisticated research designs, making it possible to analyze different groups of children, different types of media and their contents, and different areas of academic achievement. The results of such research have considerably refined the nature of the relationship by identifying significant mediating factors such as age (TV does not inhibit reading achievement among young children but can do so among teenagers), cognitive development (among same-age younger children, the cognitively more developed watch more TV than the cognitively less developed, whereas the reverse is true among older children and adolescents), and academic achievement (with curvilinear differences between high, average, and low achievers). Socioeconomic status differences, too, influence both media-use frequency and reading achievement.

In an attempt to bring some coherence to these disparate sets of results, some researchers have created large-scale, longitudinal designs employing multivariate analyses. The results show, for example, that TV viewing can be both a positive and a negative influence on achievement – depending on the type of content viewed, the social context of viewing, and the developmental level of the viewers. Other studies have tested the efficacy of the various approaches by comparing negative models (media use leads to lower school-achievement scores), positive models (more use of the print media leads to higher achievement scores), and school models (which postulate that school-based factors, such as motivation and academic self-concept, better explain variations in achievement scores than do external factors such as media use). While some support accrues to all three models, the negative model has been found to be empirically stronger than the positive model, while the strongest and most parsimonious results were provided by the school model. These studies also reaffirmed the importance of age, gender, and socioeconomic status in mediating the postulated relationships.

Conclusions

There is no common agreement as to the definition or empirical operationalization of the concept of literacy which – both historically and in contemporary society – carries a multiplicity of meanings. Mostly, it seems to be

equated with functional literacy – defined, basically, as the ability to read and write, although broader definitions encompass achievement in all academic areas. More recently, concepts such as digital or computer literacy have appeared.

The predominant theoretical perspective informing most research has been – and to some extent still is – a simple negative-effects model, dominated by TV use. Various academic consequences of excessive TV-use have been postulated, for example, poor concentration, reading and writing difficulties, and poor examination results. Attempts to formulate and test alternative theoretical models have been uncommon.

Empirical research in the field has seldom been sustained and systematic. The great majority of studies has been limited, one-off, and cross-sectional in nature and the fact that they have been conducted by researchers from a variety of different disciplines – using a wide array of measurement and analytical methods – has made rigorous comparison of results difficult and hazardous. Moreover, many studies of the relationship between media use and educational achievement have been based secondarily on data collected primarily for other purposes. As a result of the piecemeal nature of the research, it is still extremely difficult to draw firm conclusions from the available evidence. Consequently, there is little consensus among researchers.

The enormous increase in the amount and types of TV now available to children – not to mention the rapid diffusion of other forms of media, such as computers (and their games) and the Internet (with all its diverse applications) – has radically altered the whole research context and has imposed new theoretical and methodological parameters. Moreover, this has infused the issue of whether or not media use retards academic achievement with a new sense of urgency and has accentuated the already serious concern felt by many educators, parents, and policymakers.

Further Reading

Cool, V. A., Yarbrough, D. B., Patton, J. E., Runde, R., and Keith, T. Z. (2004). Experimental effects of radio and television distractors on children's performance on mathematics and reading assignments'. *Journal of Experimental Education* 62(3), 181–194.

- Ennemoser, M. (2003). Television effects in kindergarten and elementary school Grades, and Differential Effects. *Nervenheilkunde* 22(9), 443–456.
- Jackson, L. A., von Eye, A., Biocca, F. A., et al. (2006). Does home internet use influence the academic performance of low-income children? *Developmental Psychology* 42(3), 429–435.
- Koolstra, C. M. and van der Voort, T. W. A. (1996). Longitudinal effects of television on children's leisure time reading: A test of three explanatory models. *Human Communication Research* 23(1), 4–35.
- Koolstra, C. M., van der Voort, T. H. A., and van der Kamp, L. J. T. (1997). Television's impact on children's reading comprehension and decoding skills: A three year panel Study. *Reading Research Quarterly* 32, 128–152.
- Li, X. and Atkins, M. S. (2004). Early childhood computer experience and cognitive and motor development. *Pediatrics* 113, 1715–1722.
- Neumann, S. (1991). *Literacy in the Television Age*. Norwood, NJ: Ablex.
- Postman, N. (1986). *Amusing Ourselves to Death*. London: Heinemann.
- Quaiser-Pohl, C., Geiser, C., and Lehmann, W. (2006). The relationship between computer-game preference, gender, and mental-rotation ability. *Personality and Individual Differences* 40, 609–619.
- Roberts, D. F., Henriksen, L., and Voelker, D. H. (1993). Television and schooling: displacement and distraction hypotheses. *Australian Journal of Education* 37(2), 198–211.
- Roe, K. (1992). Different destinies – different melodies: School achievement, anticipated status and adolescents' tastes in music. *European Journal of Communication* 7(3), 335–358.
- Roe, K. (1995). Adolescents use of socially disvalued media: Towards a theory of media delinquency. *Journal of Youth and Adolescence* 24(5), 617–631.
- Roe, K., Eggermont, S., and Minnebo, J. (2001). Media use and academic achievement: Which effects? *Communications: The European Journal of Communication Research* 26(1), 39–57.
- Roe, K. and Muijs, D. (1998). Children and computer games: A portrait of the heavy user. *European Journal of Communication* 13(2), 181–200.
- Rosengren, K. E. and Windahl, S. (1990). *"Media Matter" TV use in Childhood and Adolescence*. Norwood, NJ: Ablex.
- Salomon, G. (1984). Television is "easy" and print is "tough": The differential investment of mental effort as a function of perceptions and attributions. *Journal of Educational Psychology* 76, 647–658.
- Shin, N. (2004). Exploring pathways from television viewing to academic achievement in school age children. *Journal of Genetic Psychology* 165(4), 367–381.
- Subrahmanyam, K., Greenfield, P., Kraut, R., and Gross, E. (2001). The impact of computer use on children's and adolescents' development. *Applied Developmental Psychology* 22, 7–30.
- Vandewater, E. A. and Bickham, D. S. (2004). The impact of educational television on young children's reading in the context of the family. *Journal of Applied Developmental Psychology* 25(6), 717–728.
- Willoughby, T. (2008). A short-term longitudinal study of internet and computer game use by adolescent boys and girls: Prevalence, frequency of use, and psychosocial predictors. *Developmental Psychology* 44(1), 195–204.
- Wright, J. C., Huston, A. C., Murphy, K. C., et al. (2001). The relations of early television viewing to school readiness and vocabulary of children from low-income families. *Child Development* 72(5), 1347–1366.

Technology and Physical and Social Health

K Subrahmanyam, California State University, Los Angeles, Los Angeles, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Bulletin boards – Bulletin boards are online public spaces, where users post messages asynchronously (i.e., not in real time) and there may or not be a time lag between messages, which are available for viewing long after they have been posted. Bulletin boards may be organized around a focal topic, such as college admissions, computer use, health, politics, and religion and interactions involve both the exchange of information, advice, as well as emotional support and encouragement.

(Subrahmanyam and Smahel, In press; Adolescents, Internet, and Development. Springer Publishing.)

Displacement hypothesis – Proposed by Nie and Hillygus, it suggests that time spent on the Internet may displace physical activities or social interactions. (Nie, N. H., & Hillygus, D. S. (2002). Where does Internet time come from? A reconnaissance.

IT&Society, 1, 1–20.)

Eating disorders – According to the National Eating Disorders Association “**Eating disorders are illnesses with a biological basis** modified and influenced by emotional and cultural factors.” They include different subtypes such as anorexia nervosa (self-starvation and excessive weight loss) and bulimia nervosa (cycle of binge eating and self-purging). <http://www.nationaleatingdisorders.org/>.

Effect size – A measure of the strength of the relationship between two variables, or the size of the observed relationship in practical terms.

Generalized physiological arousal – Is a bodily response consisting of changes in breathing, heart rate, and blood pressure. Media-induced arousal lingers after exposure, and may have an energizing effect leading to agitation and restlessness in children's behavior (e.g., while playing or interaction with peers) (Valkenburg, 2004) (Valkenburg, P. M. (2004). *Children's responses to the screen: a media psychological approach*. Lawrence Erlbaum Associates.)

Meta analysis – A statistical analysis of other studies' statistical effects.

Nintendinitis – A specific form of tendonitis where severe pain is experienced in the extensor tendon of the right thumb due to the repeated playing of a video game such as the Nintendo.

Online chat rooms – Online spaces, that are public or private, where users interact with each other in real time; they typically involve multiple users participating in several simultaneous conversations occurring in the chat space (Greenfield & Subrahmanyam, 2003) (Greenfield, P. M., & Subrahmanyam, K. (2003). Online discourse in a teen chatroom: New codes and new modes of coherence in a visual medium. *Journal of Applied Developmental Psychology*, 24, 713–738.)

Photosensitive epilepsy – A condition in which epileptic seizures are triggered by visual stimuli such as flashing or flickering lights.

Self-injury – Where the individual deliberately inflicts harm to himself or herself (e.g., by cutting oneself with a razor blade), but not necessarily with suicidal intent.

Social networking utilities – Allow users to create public or private profiles and form networks of “friends” with whom they can interact publicly (e.g., status updates or wall posts) and privately (e.g., private messages). Users can also post user-generated content (e.g., written notes, photos, and videos), which often elicit comments and result in further interaction. Examples include, www.myspace.com, and www.facebook.com (Subrahmanyam and Smahel, In press; Adolescents, Internet, and Development. Springer Publishing.)

Technology has become ubiquitous in today's society and questions abound as to its role in the lives of children and adolescents. Among youth, commonly used technologies include computers, Internet, and electronic devices such as cellphones, Ipods, and sidekicks. Computers include desktop computers and laptop computers as well as electronic games played on computers, stand alone game systems (e.g., X-box or the Wii), and portable, hand-held game systems (e.g., gameboy, Sony play station portable (PSP)). The Internet includes general purpose websites, websites that allow social networking (e.g., MySpace and Facebook), websites that allow online journaling (e.g., www.livejournal.com), websites that allow sharing of videos and music (e.g., YouTube) as well as a wide variety of applications including chat rooms, e-mail, instant messaging, and bulletin boards. Interestingly, the distinctions between the different technologies is blurring – thus, for

instance, electronic games can be played on game systems as well as via the Internet, and hand-held mobile devices such as the iPhone or Blackberry combine in one hand-held device, a mobile phone, as well music, e-mail, web browsing, maps/global positioning system (GPS) and other capabilities.

Three features mark young people's use of these technologies: First, technologies such as computers and the Internet have become so diffused that most youngsters in the United States today have access to them (Jones and Fox, 2009). Second, youth use these technologies for extensive amounts of time (Lenhart *et al.*, 2005; Roberts *et al.*, 2005). Third, it appears that children are using these technologies at a younger and younger age; for example, in the 2003 Kaiser Report, Zero to Six, 64% of children between 4 and 6 years of age knew how to use a computer mouse to point and click (Rideout *et al.*, 2003). Given these trends, concerns have been raised regarding the impact of digital technologies on children's development. The article examines the impact of technology on children's and adolescents' physical and social health.

Impact of Technology on Physical Health

Research on the impact of technology on physical health has examined two issues. The first issue concerns the physical effects of using technology, such as the risk of injuries and seizures as well as obesity. A second issue concerns the health-related uses of technology – for example, the use of the Internet to obtain information about wellness and illnesses as well as use of technology to deliver treatments and interventions.

Physical Effects of Technology

With regard to the physical effects of technology, research has identified both direct and indirect effects. Direct effects include the risks of seizures, hand injuries, and changes in physiological arousal such as heart rate (Subrahmanyam *et al.*, 2001). Based on an extensive review of the literature, Subrahmanyam *et al.* (2000) report that computer and videogames may trigger epileptic seizures in certain users; such seizures appear most likely among those suffering from photosensitive epilepsy, a condition in which epileptic seizures are triggered by visual stimuli such as flashing or flickering lights. Another effect associated with videogame playing is a form of tendinitis or nintendinitis, "which is a sports injury characterized by severe pain in the extensor tendon of the right thumb as a result of the repeated pressing of buttons during game play" (Subrahmanyam *et al.*, 2001). To guard against both risks, users should read and follow guidelines included with videogames and take the same precautions as those that are recommended to prevent any sort of repetitive injury (e.g., frequent breaks).

A meta-analysis by Anderson and Bushman (2001) suggests that exposure to violent videogames increased physiological arousal. In the studies that were considered in the meta-analysis, physiological arousal was measured using systolic blood pressure, diastolic blood pressure, and heart rate. (The association between videogame play and aggressive and violent behavior is discussed later in this article.)

Research on the indirect physical effects of technology have been framed within the context of the displacement hypothesis, which suggests that because time is a finite quantity, time spent on the Internet comes at the expense of other activities, such as participation in physical activities (e.g., organized sports) (Nie and Hillygus, 2002). These concerns have taken on added urgency in the face of reports that the rate of childhood obesity is increasing, in the US. Based on a comprehensive review of research on this topic, the 2004 Kaiser report concluded that youth who used more media were more obese. Interestingly, the report found that media technology use was not displacing physical activities. For instance, people watching less television were often engaging in other sedentary activities, such as talking on the phone, playing board games, and reading books. Instead, it appeared that the content of media, particularly the food advertising, may play a role in increasing rates of obesity among those with more media exposure. At the same time, online advertising may have potential as a medium for antiobesity campaigns targeting youth.

Technology and Health Information

One important health-related use of technology, particularly computers and the Internet, is that it can be used to obtain information about various health-related issues. Suzuki and Calzo (2004) have suggested that because of the option to remain anonymous online, the Internet is particularly attractive as a source of information about sensitive health-related questions that adolescents may not be comfortable asking an adult (such as a parent or physician) in person. Other advantages that they identified included the Internet's availability 24 h a day, 7 days a week, the option to passively obtain information (by looking at other people's questions and the responses they received), and the ability to get advice and suggestions from more individuals than would be possible from one's circle of face-to-face friends. Based on an analysis of a health-related online bulletin board targeting adolescents, Suzuki and Calzo (2004) reported that health-related concerns and questions were among the more frequently posted question categories. Frequently posted health-related questions concerned sexual health, pregnancy and birth control, body image, and grooming of genital areas. The Internet is also more generally used to access health information by adolescents. In a survey study

of tenth graders, Borzekowski and Rickert (2001) reported that 49% reported using the Internet to access health information. The use of the Internet for health information can be potentially very valuable for segments of society that would otherwise not have access to such information. For instance, in another recent study, Borzekowski *et al.* (2006) found that approximately 53% of 15–18-year-olds living in Accra, Ghana used the Internet to search for health information, and that the Internet was a particularly important information source for youth who were out of school. Some of the topics that were most often searched for by the youth in this study were sexually transmitted diseases, sexual abuse, sexual activities, and diet and nutrition, fitness, and exercise. The Internet thus has tremendous potential to reach disadvantaged youth in the area of health education.

Not only are technologies, such as the Internet, used to obtain general health information, they are also used to obtain and disseminate information about specific diseases and illnesses as well as to receive support from peers who have experienced or are experiencing the same illness. Specific Internet technologies that have been studied include web pages and bulletin boards for illnesses such as cancer, eating disorders, and self-injury.

Kyngas *et al.* (2001) found that for adolescents and young adults (between 16 and 22 years) suffering from a chronic illness such as cancer, searching for information on the Internet was a good coping strategy. Suzuki and Beale (2006) studied the personal web pages created by teens with cancer and found three potentially beneficial functions: self-presentation (e.g., essays and poetry), information dissemination (e.g., lists, charts, and hyperlinks), and interpersonal connection (e.g., guestbook entries). They concluded that such web pages were being used by adolescent cancer sufferers for self-expression, information access, and contact with peers. The access to information that technology affords is particularly important given that access to cancer information in young patients is related to lowered anxiety and depression as well as to less anxiety about treatments (Suzuki and Beale, 2006).

The interpersonal connections with peers that technology makes possible may be particularly valuable for isolated youth (e.g., those living in small towns or rural areas), youth suffering from relatively rare illnesses, as well as youth suffering from illnesses that they may not be comfortable talking about to their peers in person (e.g., AIDS, eating disorders, or self-injury behavior). Online bulletin boards and chat rooms are Internet venues that allow youth to form such connections. Whitlock *et al.* (2006) studied 406 online message boards that specifically targeted self-injurious behaviors (e.g., cutting). The authors of these studies report that the boards were most popular among young females between 14 and 20 years of age, the group also most at risk for such behavior.

Similar results have been found in a study of an electronic support group for those suffering from eating disorders, also common among adolescent females. Winzelberg (1997) analyzed 306 messages that were posted on an eating-disorder electronic support group over a period of 3 months. Analysis suggested that most of the posters were in their teens or early twenties. In terms of message content, the most common message types were self-disclosure, requests for information, and direct provision of emotional support. Interestingly most of the messages were posted between 7 p.m. and 7 a.m., when traditional face-to-face sources of support are typically not available.

Although online forums may provide information and support about eating disorders, online images could also have negative consequences. In an experimental study, female undergraduate students ($M = 18.7$ years) were randomly shown one of three kinds of websites: a proanorexic website, a website of female fashion showing average-sized models, or a home-décor site. Pre- and posttreatment questionnaires suggested that among participants who viewed the proanorexic website there was a decrease in self-esteem, appearance self-efficacy, and perceived attractiveness, as well as an increase in negative affect and perceptions of being overweight (Bardone-Cone and Cass, 2006).

The foregoing review suggests that although the Internet may have many positive benefits in the area of illness and disorders, they may also serve to normalize potentially dangerous behaviors such as self-injury, anorexia, and bulimia. This is an important concern considering that the Internet is now home to innumerable sites that provide detailed information and instructions about several dangerous behaviors that are gaining popularity among adolescents (e.g., sniffing glue or choking oneself).

Technology and Treatment Delivery

Technological resources have also been found useful to deliver interventions and treatments. One recent study assessed the effectiveness of an Internet-based lifestyle-behavior-modification program for overweight African-American adolescent girls who also had one obese biological parent (Williamson *et al.*, 2005). Participants were randomly assigned to either an interactive behavioral program (including Internet counseling) or a health-education program delivered over the Internet. The intervention also included face-to-face sessions for all participants. Although only 50% of the participants completed the 6-month trial, measurements taken at baseline and 6 months later showed that adolescents in the interactive behavioral program and their parents lost more body weight; both parents and adolescents showed lowered dietary fat intake. This study suggests that a more passive Internet health-education program is not as effective as an interactive one.

Another study assessed the effectiveness of a psycho-educational intervention using a videogame called

Re-Mission among adolescents (and young adult) cancer patients. A multisite trial was conducted with 375 cancer patients between 13 and 29 years of age; 197 were randomly assigned to receive the Re-Mission game, and 176 received an alternative videogame (Beale *et al.*, 2006). After playing the game, participants filled out a questionnaire about the acceptability of the game as an intervention and its ability to change patients' knowledge, attitudes, and treatment-related self-care behaviors. The authors report that participants expressed a good level of acceptance of the game and that the game had a moderate level of credibility as an intervention. Not surprisingly, acceptability and credibility ratings were related to the amount of time a patient played the videogame (Beale *et al.*, 2006). Playing the game also increased a patient's quality of life, knowledge about cancer information presented in the game, and ability to manage side effects. Importantly, young people who played the game seemed better able to adhere to treatment regimens as evidenced by higher blood levels of chemotherapy and higher rates of antibiotic utilization (Hope Lab, 2007).

Finally, technology has also been used to enhance the physician-patient relationship and to empower chronically ill youth. Rich (2004) describes the video intervention/prevention assessment (VIA) method, which was designed to give children and adolescents with health problems an opportunity to express themselves through video media. Children with chronic asthma were given video recorders, and asked to explain their illness to their clinicians. Four benefits were found for the participants: first, they felt empowered by their ability to express themselves; second, they were able to teach others about their unique individual experiences fighting diseases that millions are afflicted with; third, they learned about the process of creating videos, as did the researchers who worked with them; and finally, they became more self-aware of the risks they were taking and the paths their illnesses were taking.

For the clinicians, the narratives recorded in these videos provided a picture of the actual psychosocial environments in which the patients lived, as well as some of the patients' innermost thoughts. Sometimes, the videos even demonstrated environmental triggers, without the patients even realizing they were doing so. The videos also chronicled depression and the risky decisions that adolescents made because of the stigma they felt from their asthma. Thus, the clinicians benefited because they were able to see information and points of view that were usually not observable in traditional office visits. The video diaries that were created cast light on some of the challenges the patients faced outside of the doctors' offices, including life at home, at school, and interacting with other health-care practitioners. Although the VIA technology was originally created for young asthma patients, it has been used to explore the experiences of youth suffering

from obesity, sickle-cell disease, cystic fibrosis, human immunodeficiency virus (HIV), and spina bifida.

Technology and Social Development

Research on the role of technology in the social development of children and adolescents has focused on two main sets of questions. The first set of questions focus on the effects of interacting with violent content such as videogames; a related issue concerns how technology may be transforming bullying, victimization, and other aggressive behaviors typical among youth. The second set of questions focuses on the effects of computer-mediated communication (CMC) forms on youths' developing social relationships.

Technology and Aggression

One favorite pastime among youth is that of videogames and a recurring concern in the years since their introduction has been their increasingly violent content. In a study of 607 eighth- and ninth-grade students, it was found that on average, participants spent 9 h a week playing videogames and that males played for significantly longer periods of time than females (13 vs. 5 h a week) (Gentile *et al.* 2004). It appears that there is little parental knowledge and control over youths' use of this medium. Less than half of the respondents (31%) think that their parents understand the videogame ratings put out by the Entertainment Software Ratings Board (ESRB), and only 15% reported that their parents consistently monitor the ratings when purchasing or renting videogames. Of note is the finding that 10% of the participants reported having games, whose content would not be approved by their parents. Importantly, the study found that youth who reported greater exposure to videogame violence were more hostile and reported getting into arguments with teachers more frequently, and were more likely to be involved in physical fights.

A meta-analysis by Anderson (2004) reveals that "playing violent videogames was associated with increases in aggressive behavior, aggressive cognition, aggressive affect, and physiological arousal, and with decreases in helping behavior." The methodologically stronger studies in the meta-analysis yielded stronger effects than did the methodologically weaker studies. The author (Anderson, 2004) found the effect size of violent videogame exposure on aggressive behavior to be 0.28. To put this in perspective, it is larger effect than that of condom use on HIV risk or that of passive workplace smoke exposure on lung cancer. Based on a review of the research, Anderson *et al.* (2003) similarly conclude that experimental studies show that violent videogames cause short-term increases in aggressive thoughts, affect, and behavior as well as

decreases in helping behavior. They report that correlational studies of exposure time and aggressive behavior show a link between repeated exposure to violent videogames and real-life aggression. Finally, their review of longitudinal research showed modest but suggestive evidence of long-term effects of the repeated playing of violent videogames on aggressive and violent behavior.

A second major concern with regard to technology and social health are reports that bullying is migrating from school yards to cell phones, instant messages, and the Internet. A survey done in the UK found that 20% of the 770 respondents (11–19) had reported being bullied through technology – they had received a threat via e-mail, Internet, chatroom, or text message (NCH, 2005). Mobile text messaging was the most commonly reported form of bullying. Bullying in Internet chat rooms and via e-mails was reported by 5% and 4% of the sample, respectively. Cell phone cameras are another for harassment. Ten percent of the respondents in the UK survey reported feeling embarrassed, uncomfortable, or threatened by a picture that somebody took of them using a cell phone camera. Ironically, much of the harassment is being carried out by people known to the victims. Researchers in the US have similarly found that about 9% of youth Internet users have experienced online harassment (Wolak *et al.*, 2006). Forty-seven percent reported being harassed via instant messaging, 13% by e-mails, 11% by chatrooms, and 3% by blogs.

Research on the characteristics of cyber bullies and their victims has found that cyber bullies are more likely to report poor parent–child relationships, substance use, and delinquency (Ybarra and Mitchell, 2004). Victims of harassment were also more likely than other youth to report symptoms of depression (Ybarra and Mitchell, 2004). More research is necessary to understand the relation between depression symptomatology and the likelihood of being bullied.

Technology and Social Relationships

Relations with peers and families are an important part of development and technology has become an important means of connecting with others. The popularity of the communication applications of the Internet among youth has raised many concerns. First, there is concern that because CMC lacks important features, such as gestures and eye contact, it is not as rich as face-to-face communication forms. Second, there is concern that online relationships may involve strangers, who are not part of one's offline life. Consequently, online relationships are hypothesized to represent weak ties, which are characterized as relationships that have superficial and easily broken bonds, infrequent contact, and narrow focus (Kraut *et al.*, 1998). Research on this topic has therefore examined the extent of youths' online interactions with strangers

and the impact of their online interactions on their psychological well-being.

To what extent do youth interact online with friends versus strangers? To some extent this is dependent on the popular technology application at any given time. For instance, when chat rooms were popular, contact with strangers was more likely. With the introduction of instant messaging, this changed and youth seemed to be using this newer technology mostly to connect with offline friends (Boneva *et al.*, 2006). The currently popular social networking sites seem to make possible contact both with friends and strangers. For instance, in one national survey conducted in 2006, 40% of 14–22-year-olds who use social networking sites such as MySpace reported that they had been contacted online by a stranger whom they did not know before (Annenberg Public Policy Center, 2006). A subsequent survey conducted in 2007 found that among youth who use social networking sites, the majority do so to keep in touch with friends whom they see frequently (91%) or to keep in touch with friends whom they see rarely in person (82%) (Lenhart and Madden, 2007).

The limited research that has been conducted on the topic suggests that adolescents' online relationships may not be of the same quality as their offline ones. For instance, in one study of Israeli adolescents, it was found that online friendships were shorter in duration and were not as close as measured by the topics discussed (e.g., less personal topics) and frequency of shared activities (Mesch and Talmud, 2006). Research also suggests that relationships from the online realm move to the offline realm only rarely; of the 256 respondents of a national survey ($N = 1501$), who reported close online relationships, only 41% of them reported that they had met their online friend face to face (Wolak *et al.*, 2002). More research is necessary to explore this question, given social networking utilities that seek to enable users to make new friends.

How are online communication modes impacting youth's existing friendships? In a survey study, 80% of the respondents reported using the Internet to maintain existing friendship networks (Valkenburg and Peter, 2007). Participants who communicated more often on the Internet felt closer to existing friends, but only if they were using the Internet to communicate with friends rather than strangers. Interestingly, participants who felt that online communication was more effective for self-disclosure also reported feeling more close to their offline friends compared to adolescents who did not view online communication as allowing for more intimate self-disclosure compared to offline communication.

A related question is whether youths' use of technologies, such as the Internet, displaces activities important for adolescent social development, such as face-to-face interactions with peers. According to the displacement hypothesis, because time is not infinite, time spent on the Internet displaces other daily activities, in particular

face-to-face social interactions (Nie and Hillygus, 2002). Proponents of this view suggest that not only would Internet use displace adolescents' real interactions with peers and family, it also may be substituting weak ties (with online friends) for strong ones (with offline friends). Research suggests that weak ties typically provide less consequential social support than do more intimate ties (Krackhardt, 1994). Consequently, research has examined whether greater Internet use is associated with weaker social ties as well as lowered well-being (e.g., greater depression).

Early research suggested that greater use of the Internet was associated with declines in adolescents' well-being and with weaker social ties (Kraut *et al.*, 1998). Adolescents' perceptions about the quality of their family relationships was also found to be negatively related to frequency of Internet use (Mesch, 2003). However, more recent studies have not found a link between adolescents' time online and psychological well-being, as indicated by dispositional or daily well-being (Gross *et al.*, 2002) or loneliness (Subrahmanyam and Lin, 2007). Finally, time spent online does not appear to be related to aspects of social networks, such as size of local and distant social circles and amount of face-to-face communication (Kraut *et al.*, 2002).

One reason for this conflicting pattern of findings could be that the measures of Internet time use that have been used in the studies did not distinguish between different kinds of Internet use. For instance, e-mailing and chatting with school friends might contribute to well-being, surfing the web for information about sports, music, or movies might have no impact on well-being, and accessing pornographic materials might actually threaten well-being. One study that distinguished among different types of online activity found that more time spent in chatrooms, online browsing, and games was related to higher levels of social anxiety among older adolescent and young adult males, but not females (Mazalin, and Moore, 2004). This correlational relationship does not resolve the question of whether participants are becoming more depressed because of their chat room use or whether they are drawn to chat rooms in the first place because they are depressed and are looking for support, which may be lacking in their offline life. Further complicating the picture is the finding that for extroverted teens, Internet use appears to be associated with declines in loneliness (Kraut *et al.*, 2002). Although no systematic relation between time use and dispositional loneliness has been found, daily social anxiety and loneliness is related to whether a youth communicates via instant messaging with a stranger on that given day (Gross *et al.*, 2002). Adolescents' perceptions regarding the relationship strength may also be important. Subrahmanyam and Lin (2007) found that for a sample of urban teens, loneliness was not related to the total time spent online, nor to the time spent on

e-mail, but was predicted by participants' gender (male adolescents were more lonely) and perceptions regarding their online relationships. Adolescents who felt that their relationship with online partners was one that they could turn to in times of need also reported feeling more lonely.

Conclusion

Overall, extant research presents a mixed picture of the effect of technology on children and adolescents. In the area of physical health, technology poses slight risks such as nitendinitis or epileptic seizures, and physiological arousal; there also seems to be some link between media use and obesity. Although the latter risk is potentially worrisome, education efforts can help minimize the problem. On the positive side, research suggests that technologies such as the Internet and computer games can be enormously helpful for disseminating information about health and illness as well as for delivering treatments and interventions.

In the area of social development, technology has provided new means for youth to bully and victimize their peers; while such behaviors have always existed among youth, more needs to be done to increase awareness in parents, teachers, and others who work with young people. Violent content in videogames contributes to aggression in youth, and parents are often uninformed concerning the nature of the games that their children play. In the area of interpersonal connections, although online relationships may not be of high quality, research has not found systematic links between time online and well-being. As newer forms of online communication are emerging, more research is necessary to understand the relation between youth's online and offline social worlds.

See also: An Overview of Technology and Learning; Examining the Effects of Educational Technology Programs: Challenges and Strategies; New Media, Learning from; Relating Technology, Education Reform and Economic Development.

Bibliography

- Anderson, C. A. (2004). An update on the effects of playing violent video games. *Journal of Adolescence* **27**, 113–122.
- Anderson, C. A., Berkowitz, L., Donnerstein, E., *et al.* (2003). The influence of media violence on youth. *Psychological Science in the Public Interest* **4**, 81–110.
- Anderson, C. A. and Bushman, B. J. (2001). Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: A meta-analytic review of the scientific literature. *Psychological Science* **12**, 353–359.
- Annenberg Public Policy Center. (2006). Stranger contact in adolescent online social networks. <http://www.annenbergpublicpolicycenter>.

- org/Downloads/Releases/Release_HC20060920/Report_HC20060920.pdf (accessed September 2009).
- Bardone-Cone, A. M. and Cass, K. M. (2006). Investigating the impact of pro-anorexia websites: A pilot study. *European Eating Disorders Review* **14**, 256–262.
- Beale, I. L., Marin-Bowling, V. M., Guthrie, N., and Kato, P. M. (2006). Young cancer patient's perceptions of a video game used to promote self care. *International Electronic Journal of Health Education* **9**, 202–212.
- Boneva, B. S., Quinn, A., Kraut, R. E., Kiesler, S., and Shklovski, I. (2006). Teenage communication in the instant messaging era. In Kraut, R., Brynin, M., and Kiesler, S. (eds.) *Information Technology at Home*, pp 612–672. New York: Oxford University Press.
- Borzekowski, D. L. G., Fobil, J. N., and Asante, K. O. (2006). Online access by adolescents in Accra: Ghanaian teens' use of the Internet for health information. *Developmental Psychology* **42**, 450–458.
- Borzekowski, D. L. G. and Rickert, V. I. (2001). Adolescents, the Internet, and health: Issues of access and content. *Journal of Applied Developmental Psychology* **22**, 49–59.
- Gentile, D. A., Lynch, P. J., Linder, J. R., and Walsh, D. A. (2004). The effects of violent video game habits on adolescent hostility, aggressive behaviors, and school performance. *Journal of Adolescence* **27**, 5–22.
- Gross, E. F., Juvonen, J., and Gable, S. E. (2002). Online communication and well-being in early adolescence: The social function of instant messages. *Journal of Social Issues* **58**, 75–90.
- Hope Lab TM (2007). Re-mission. outcomes study: A research trial of a video game shows improvement in health-related outcomes for young people with cancer. <http://www.hopelab.org/docs/HopeLab%20-%20Re-Mission%20Outcomes%20Study.pdf> (accessed September 2009).
- Jones, S. and Fox, S. (2009). Generations online in 2009. <http://pewresearch.org/pubs/1093/generations-online> (accessed September 2009).
- Krackhardt, D. (1994). The strength of strong ties: The importance of Philos in organizations. In Nohria, N. and Eccles, R. (eds.) *Networks and Organizations: Structure, Form, and Action*, pp 216–239. Boston, MA: Harvard Business School Press.
- Kraut, R., Kiesler, S., Boneva, B., et al. (2002). Internet paradox revisited. *Journal of Social Issues* **58**, 49–74.
- Kraut, R., Patterson, M., Lundmark, V., et al. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist* **53**, 1017–1031.
- Kyngas, H., Mikkonen, R., Nousiainen, E. M., et al. (2001). Coping with the onset of cancer: Coping strategies and resources of young people with cancer. *European Journal of Cancer Care* **10**, 6–11.
- Lenhart, A. and Madden, M. (2007). *Social Networking Websites and Teens: An Overview*. Washington, DC: Pew Internet and American Life Project. http://www.pewinternet.org/~media/Files/Reports/2007/PIP_SNS_Data_Memo_Jan_2007.pdf.
- Lenhart, A., Madden, M., and Hitlin, P. (2005). Teens and technology: Youth are leading the nation to a fully wired and mobile nation. *Pew Internet and American Life Project*, Washington, DC. http://www.pewinternet.org/~media/Files/Reports/2005/PIP_Teens_Tech_July2005web.pdf.
- Mazalin, D. and Moore, S. (2004). Internet use, identity development and social anxiety among young adults. *Behavior Change* **21**, 90–102.
- Mesch, G. S. (2003). The family and the Internet: The Israeli case. *Social Science Quarterly* **84**, 1039–1050.
- Mesch, G. S. and Talmud, I. (2006). The quality of online and offline relationships, the role of multiplexity and duration. *Information Society* **22**, 137–149.
- NCH (2005). *Putting U in the Picture: Mobile Bullying Survey 2005*. http://www.filemaker.co.uk/educationcentre/downloads/articles/Mobile_bullying_report.pdf.
- Nie, N. H. and Hillygus, D. S. (2002). Where does Internet time come from? A reconnaissance. *IT & Society* **1**, 1–20.
- Rich, M. (2004). Health literacy via media literacy: Video intervention/prevention assessment. *American Behavioral Scientist* **48**, 1–25.
- Rideout, V. J., Vandewater, E. A., and Wartella, E. A. (2003). *Zero to Six: Electronic Media in the Lives of Infants, Toddlers and Preschoolers*. Menlo Park, CA: A Kaiser Family Foundation Report.
- Roberts, D. F., Foehr, U. G., and Rideout, V. (2005). *Generation M: Media in the Lives of 8-18 Year-olds - Report [Electronic Version]*. Retrieved December 16, 2008 from. <http://www.kff.org/entmedia/7251.cfm>.
- Subrahmanyam, K., Kraut, R. E., Greenfield, P., and Gross, E. (2000). The impact of home computer use on children's activities and development. In Shields, M. K. (ed.) *The Future of Children: Children and Computer Technology*, vol. 10, pp 123–144.
- Subrahmanyam, K. and Lin, G. (2007). Adolescents on the net: Internet use and well-being. *Adolescence* **42**, 659–677.
- Suzuki, L. K. and Beale, I. L. (2006). Personal web home pages of adolescents with cancer: Self-presentation, information dissemination, and interpersonal connection. *Journal of Pediatric Oncology Nursing* **23**, 152–161.
- Suzuki, L. K. and Calzo, J. P. (2004). The search for peer advice in cyberspace: An examination of online teen health bulletin boards. *Journal of Applied Developmental Psychology* **25**, 685–698.
- Valkenburg, P. and Peter, J. (2007). Preadolescents' and adolescents' online communication and their closeness to friends. *Developmental Psychology* **43**, 267–277.
- Whitlock, J. L., Powers, J. L., and Eckenrode, J. (2006). The virtual cutting edge: The Internet and adolescent self-injury. *Developmental Psychology* **42**, 407–417.
- Williamson, D. A., Martin, P. D., and White, M. A. (2005). Efficacy of an Internet-based behavioral weight loss program for overweight adolescent African-American girls. *Eating and Weight Disorders* **10**(3), 193–203.
- Winzelberg, A. (1997). The analysis of an electronic support group for individuals with eating disorders. *Computers in Human Behavior* **13**, 393–407.
- Wolak, J., Mitchell, K. J., and Finkelhor, D. (2002). Close online relationships in a national sample of adolescents. *Adolescence* **37**, 441–455.
- Wolak, J., Mitchell, K., and Finkelhor, D. (2006). Online victimization of youth: Five years later. <http://www.unh.edu/ccrc/pdf/CV138.pdf> (accessed September 2009).
- Ybarra, M. L. and Mitchell, K. J. (2004). Youth engaging in online harassment: Associations with caregiver-child relationships, Internet use, and personal characteristics. *Journal of Adolescence* **27**(3), 319–336.

Further Reading

Kaiser Foundation. (2004). The role of media in childhood obesity. <http://www.kff.org/entmedia/7030.cfm> (accessed September 2009).

Relevant Websites

<http://www.digitalcenter.org> – Center for the Digital Future, USC Annenberg School, Phoebe Schramm.

<http://www.digitalllearning.macfound.org> – MacArthur Foundation: Digital Media and Learning.

<http://www.pewinternet.org> – Pew Internet and American Life Project.

<http://www.futureofchildren.org> – The Future of Children Princeton Brookings: Children and Computer Technology.

<http://www.kff.org> – The Kaiser Family Foundation: Entertainment Media Studies.

TECHNOLOGY AND LEARNING – MEETING SPECIAL STUDENTS' NEEDS

Universal Design for Learning

D H Rose and J W Gravel, Center for Applied Special Technology, Wakefield, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Elementary and Secondary Education Act – The Elementary and Secondary Education Act (ESEA) is the main federal law affecting education from kindergarten through high school (from 2001 to 2009 known as No Child Left Behind) (US Department of Education).

The Higher Education Opportunity Act of 2008 – The Higher Education Opportunity Act (Public Law 110-315) (HEOA) was enacted on 14 August 2008, and reauthorizes the Higher Education Act of 1965, as amended (HEA). This latest reauthorization now includes language around Universal Design for Learning (US Department of Education, www.ed.gov).

Universal design – Universal design is an architectural concept that focuses on the design of products, buildings, or environments so they can be used readily by the widest possible range of users.

Universal design for learning (UDL) – UDL is a framework for teaching and learning that often capitalizes on the power and flexibility of modern technologies to address the needs of the broadest possible range of students.

Universal design, although well established in architecture and other domains, is relatively new to education. However, there are signs that the newness is waning. In the recent higher education bill passed by the US Congress (The Higher Education Opportunity Act of 2008), the application of universal design to education – called ‘universal design for learning’ (UDL) – is prominently defined and recommended for colleges and universities. Furthermore, a task force composed of nearly 30 major national educational organizations has been formed to advocate for the inclusion of UDL in upcoming revisions of ESEA.

In spite of this recent increase in exposure and advocacy, most educators do not really know what UDL means, and still others have heard of it but do not know

whether or how it differs from the architectural meaning of universal design. In this article, we hope to clarify what is meant by UDL and especially to provide the evidentiary basis for its principles and a review of its guidelines.

Universal Design and Universal Design for Learning

Universal design focuses on the design of products, buildings, or environments so they can be used readily by the widest possible range of users. Virtually all architects in the US now create buildings that are designed from the outset to reduce or eliminate architectural barriers through designs that consider the needs of diverse people. This practice is now recognized as more cost-effective and equitable than trying to retrofit buildings later or providing customized accommodations to individuals who are unable to navigate poorly designed structures. Universally designed environments are engineered for flexibility and designed to anticipate the need for alternatives, options, and adaptations to meet the challenge of diversity. While originally conceived to meet the needs of individuals with disabilities, universal designs actually make buildings that are more accessible and functional for everyone.

A good example of universal design in action comes from the history of television captioning. When captioning first became available, it was an expensive add-on purchase intended for people with hearing impairments. Building captioning into every television, rather than retrofitting it later, turned out to be a better, and more universal, design. It now benefits not only those with hearing impairments, but also exercisers in health clubs, travelers in airports, individuals working on their language skills, and couples who go to sleep at different times. The key to universal design is building options into initial designs, making better choices available for everyone.

UDL is one part of the overall movement toward universal design. The term UDL emphasizes the special purpose of learning environments – they are not created to provide access to information (that is the role for libraries and the Internet) but instead to foster the changes in knowledge and

skills that we call learning. While providing access to information is often essential to learning, it is not sufficient. Success also requires that the means for learning – the pedagogical goals, methods, materials and assessments of instruction – are also accessible. UDL is the process by which we attempt to ensure that the means for learning, and their results, are equally accessible to all students.

Part One: UDL – The Basic Framework

The framework and guidelines for UDL are not derived from the principles for architecture. Instead, they are based on research and practice from multiple domains within the learning sciences – education, developmental psychology, cognitive science, and cognitive neuroscience. The research in those fields guides both the scope of the pedagogy that UDL addresses (i.e., the critical elements of teaching and learning) and the range of the individuals that UDL addresses (i.e., the critical elements of individual differences).

At its simplest, the scope of UDL is based entirely on three principles:

1. providing multiple means of representation;
2. providing multiple means of action and expression; and
3. providing multiple means of engagement.

These three principles have been chosen because they address three critical features of any teaching and learning environment: the means by which information is presented to the learner, the means by which the learner is required to express what they know, and the means by which students are engaged in learning.

While there are many ways to articulate the fundamentals of teaching and learning, the choice of these three foundational principles stems from their commonality across many aspects of theory and research in the learning sciences. Consider the field of cognitive neuroscience where it is common to think of three broad divisions of the learning brain: (1) the pattern recognition capabilities in the posterior regions of the cortex, (2) the motor and executive capabilities in the frontal regions of the cortex, and (3) the affective or emotional capabilities in the medial regions of the nervous system. While even this division is an oversimplification, it is an articulation that is common and draws historically on Luria's classic work and has been elaborated and modified by many others. It is by design that the three principles of UDL match up well with this framework from neuroscience – addressing in turn the perceptual learning of the posterior cortex, the strategic and motor learning of the anterior cortex, and the affective or emotional learning of the medial and orbital frontal cortex – in order to be systematic in considering learning differences.

Beyond cognitive neuroscience, however, researchers and theorists in other learning sciences have adopted very

similar frameworks to consider the scope of teaching and learning. Among the most prominent, Lev Vygotsky, the preeminent Russian psychologist, and Benjamin Bloom, the American educational theorist, both adopted a similar three-part framework for their foundations.

From the three principles, nine guidelines have been developed that form the primary foundation of UDL. Those guidelines articulate the principles but their main purpose is to guide educators and curriculum developers in using evidence-based means of addressing the wide range of individual differences that any classroom typically experiences. To those guidelines we now turn.

Part Two: The UDL Guidelines and Their Research Base

A critical foundation of the UDL approach is to identify those evidence-based practices which should be included within an overall universal design. Typically, these practices have already proven effective for individual students in the margins, but they are generally not integrated within the one-size-fits-all curriculum of regular education. By taking advantage of the power and flexibility of modern technology, UDL provides a vehicle for delivering these practices to the individual students for whom they are likely to be most effective. But which practices, and for whom?

The UDL Guidelines help in making informed decisions about what practices are optimal. They articulate the specific practices that have been shown to be effective for one or specific types of learning or learners and that should be considered as important options to ensure that students with a full range of abilities and disabilities can access and progress in the general curriculum. The Guidelines assist curriculum developers (these may include teachers, publishers, and others) in designing flexible curricula that reduce barriers to learning and provide robust learning supports to all learners.

Each of the Guidelines, along with the three to four checkpoints associated with each Guideline, are based on the work of hundreds of researchers in many different institutions. A description of the Guidelines (CAST, 2008) and the associated checkpoints is described in the following. For a more detailed description of the UDL Guidelines and for a listing of the full references, the reader is referred to the National Center on UDL website.

Guidelines for Providing Multiple Means of Representation

Guideline 1: Provide options for perception

To be effective in diverse classrooms, curricula must present information in ways that are perceptible to all students. It is impossible to learn information that is imperceptible to

the learner, and difficult when information is presented in formats that require extraordinary effort or assistance. To reduce barriers to learning, therefore, it is important to ensure that key information is equally perceptible to all students by: (1) providing the same information through different sensory modalities (e.g., through vision, or hearing, or touch); (2) providing information in a format that will allow for adjustability by the user (e.g., text that can be enlarged, sounds that can be amplified). Such multiple representations ensure that information is not only accessible to students with particular sensory and perceptual disabilities, but also easier to access for many others. When the same information, for example, is presented in both speech and text, the complementary representations enhance comprehensibility for most students. Checkpoints suggesting specific strategies for providing options for perception are listed as follows:

- options that customize the display of information;
- options that provide alternatives for auditory information; and
- options that provide alternatives for visual information.

Guideline 2: Provide options for language and symbols

Students vary in their facility with different forms of representation – both linguistic and nonlinguistic. Vocabulary that may sharpen and clarify concepts for one student may be opaque and foreign to another. A graph that illustrates the relationship between two variables may be informative to one student and inaccessible or puzzling to another. A picture or image that carries meaning for some students may carry very different meanings for students from differing cultural or familial backgrounds. As a result, inequalities arise when information is presented to all students through a single form of representation. An important instructional strategy is to ensure that alternative representations are provided not only for accessibility, but for clarity and comprehensibility across all students as well. Checkpoints suggesting specific strategies for providing options for language and symbols are listed as follows:

- options that define vocabulary and symbols;
- options that clarify syntax and structure;
- options for decoding text or mathematical notation;
- options that promote cross-linguistic understanding; and
- options that illustrate key concepts nonlinguistically.

Guideline 3: Provide options for comprehension

The purpose of education is not to make information accessible, but to teach students how to transform accessible information into useable knowledge. Decades of cognitive science research have demonstrated that the capability to transform accessible information into useable knowledge

is not a passive process but an active one. Constructing useable knowledge, knowledge that is accessible for future decision making, depends not upon merely perceiving information but upon active information processing skills like selective attending, integrating new information with prior knowledge, strategic categorization, and active memorization. Individuals differ greatly in their skills in information processing and in their access to prior knowledge through which they can assimilate new information. Proper design and presentation of information – the responsibility of any curriculum or instructional methodology – can provide the cognitive ramps that are necessary to ensure that all students have access to knowledge. Checkpoints suggesting specific strategies for providing options for comprehension are listed as follows:

- options that provide or activate background knowledge;
- options that highlight critical features, big ideas, and relationships;
- options that guide information processing; and
- options that support memory and transfer.

Guidelines for Providing Multiple Means of Action and Expression

Guideline 4: Provide options for physical action

A textbook or workbook in a print format provides limited means of navigation or physical interaction (e.g., by turning pages with fingers, handwriting in spaces provided). Many interactive pieces of educational software similarly provide only limited means of navigation or interaction (e.g., via dexterously manipulating a joystick or keyboard). Navigation and interaction in those limited ways will raise barriers for some students – those who are physically disabled, blind, dysgraphic, or who have various kinds of executive function disorders. It is important to provide materials with which all students can interact. Properly designed curricular materials provide a seamless interface with common assistive technologies through which individuals with motor disabilities can navigate and express what they know – to allow navigation or interaction with a single switch, through voice-activated switches, expanded keyboards, and others. Checkpoints suggesting specific strategies for providing options for physical action are listed as follows:

- options in the mode of physical response;
- options in the means of navigation; and
- options for accessing tools and assistive technologies.

Guideline 5: Provide options for expressive skills and fluency

There is no medium of expression that is equally suited for all students or for all kinds of communication. On the contrary, there are media which seem poorly suited for

some kinds of expression, and for some kinds of students. While a student with dyslexia may excel at story telling in conversation, he may falter drastically when telling that same story in writing. Alternative modalities for expression should be provided both to level the playing field among students, and to introduce all students to the full range of media that are important for communication and literacy in our multimedia culture. Additionally, students vary widely in their familiarity and fluency with the conventions of any one medium. Within media, therefore, alternative supports should be available to scaffold and guide students who are at different levels of their apprenticeships in learning to express themselves competently. Checkpoints suggesting specific strategies for providing options for expressive skills and fluency are listed as follows:

- options in the media for communication;
- options in the tools for composition and problem solving; and
- options in the scaffolds for practice and performance.

Guideline 6: Provide options for executive functions

At the highest level of the human capacity to act skillfully are the so-called executive functions. Associated with the prefrontal cortex in the brain, these capabilities allow humans to overcome impulsive, short-term reactions to their environment and instead to set long-term goals, plan effective strategies for reaching those goals, monitor their progress, and modify strategies as needed. Of critical importance to educators is the fact that executive functions have very limited capacity and are especially vulnerable to disability. This is true because executive capacity is sharply reduced when: (1) executive functioning capacity must be devoted to managing lower-level skills and responses which are not automatic or fluent (due to either disability or inexperience) and thus the capacity for higher-level functions is taken and (2) executive capacity itself is reduced due to some sort of higher-level disability or to lack of fluency with executive strategies. The UDL approach typically involves efforts to expand executive capacity in two ways: (1) by scaffolding lower-level skills so that they require less executive processing and (2) by scaffolding higher-level executive skills and strategies so that they are more effective and developed. Previous guidelines have addressed lower-level scaffolding, this guideline addresses ways to provide scaffolding for executive functions themselves. Checkpoints suggesting specific strategies for providing options for executive functions are listed as follows:

- options that guide effective goal-setting;
- options that support planning and strategy development;

- options that facilitate managing information and resources; and
- options that enhance capacity for monitoring progress.

Guidelines for Providing Multiple Means of Engagement

Guideline 7: Provide options for recruiting interest

Information that is not attended to, that does not engage students' cognition, is in fact inaccessible. It is inaccessible both in the moment – relevant information goes unnoticed and unprocessed – and in the future: relevant information is unlikely to be remembered. As a result, teachers devote considerable effort to recruiting student attention and engagement. However, students differ significantly in what attracts their attention and engages their interest. Even the same student will differ over time and circumstance: their interests change as they develop and gain new knowledge and skills, as their biological environments change, and as they differentiate into self-determined adolescents and adults. It is, therefore, important to have alternative ways to recruit student interest; ways that reflect the important inter- and intra-individual differences among those students. Checkpoints suggesting specific strategies for providing options for recruiting interest are listed as follows:

- options that increase individual choice and autonomy;
- options that enhance relevance, value, and authenticity; and
- options that reduce threats and distractions.

Guideline 8: Provide options for sustaining effort and persistence

Many kinds of learning, particularly the learning of skills and strategies, require sustained attention and effort. When motivated to do so, many students can regulate their attention and affect in order to sustain the effort and concentration that such learning will require. However, students differ considerably in their ability to self-regulate in this way. Their differences reflect disparities in their initial motivation, their capacity and skills for self-regulation, their susceptibility to contextual interference, and so forth. A key instructional goal is to build the individual skills in self-regulation and self-determination that will equalize such learning opportunities (see Guideline 9). In the meantime, however, the external environment must provide options that can equalize accessibility by supporting students who differ in initial motivation, self-regulation skills, etc. Checkpoints suggesting specific strategies for providing options for sustaining effort and persistence are listed as follows:

- options that heighten salience of goals and objectives;

- options that vary levels of challenge and support;
- options that foster collaboration and communication; and
- options that increase mastery-oriented feedback.

Guideline 9: Provide options for self-regulation

While it is important to design the extrinsic environment so that it can support motivation and engagement (see Guidelines 7 and 8), it is also important to develop students' intrinsic abilities to regulate their own emotions and motivations. The ability to self-regulate – to strategically modulate one's emotional reactions or states in order to be more effective at coping and engaging with the environment – is a critical aspect of human development. While many individuals develop self-regulatory skills on their own, either by trial and error or by observing successful adults, many others have significant difficulties in developing these skills. Unfortunately, most classrooms do not address these skills explicitly, leaving them as part of the implicit curriculum that is often inaccessible or invisible to many. Furthermore, those classrooms that address self-regulation explicitly generally assume a single model or method for doing so. As in other kinds of learning, considerable individual differences are much more likely than uniformity. A successful approach requires providing sufficient alternatives to support learners with very different aptitudes and prior experience in learning to effectively manage their own engagement and affect. Checkpoints suggesting specific strategies for providing options for self-regulation are listed as follows:

- options that guide personal goal-setting and expectations;
- options that scaffold coping skills and strategies; and
- options that develop self-assessment and reflection.

Part Three: Future Directions and Research

In the sections above, we have provided the basis for the general framework of UDL as well as a summary of the UDL Guidelines. To realize the benefits of UDL, however – to improve actual instruction – UDL must be actualized within a specific lesson or unit. While it will never be possible to evaluate the application of UDL to every lesson, in every content area, in every grade level, and for every type of student, it is important to provide evidence that such applications can be implemented effectively in real classrooms. This kind of research is only in its infancy, but exemplars are emerging.

In the field of literacy, for example, there is a growing body of research demonstrating the effectiveness of digital reading environments that are designed in a manner consistent with UDL. CAST's research in this area began with a study of struggling adolescent readers who were reading

digital texts with embedded supports and scaffolds based on the reciprocal teaching strategies of Palincsar and Brown. That study showed that students in the UDL reading environments outperformed peers in printed texts when they later took standardized comprehension tests. These experimental results have now been replicated or extended with other populations who typically struggle in literacy, including students who are deaf, students with cognitive disabilities, and English language learners. In light of these positive results, the US Department of Education has initiated a large-scale efficacy study with the commercial version of *Thinking Reader* (see below). Much more research of this type is needed.

Finally, it is important to recognize that one critical type of research is completely absent: research on full district-wide implementation of UDL. It is not enough merely to show that the core practices and instructional elements of UDL, or even a specific application of UDL, are evidence based. It is essential also to show that the UDL approach can be applied effectively, and at scale, in full district-wide implementations. This is challenging (and necessary) because adequate implementation of UDL will require more than adoption of new technologies or techniques; it will require comprehensive reform throughout a whole school system. As of today, there are no schools or districts where UDL is thus fully realized.

Is that kind of systemic implementation research likely to happen? Three directions are promising. First, there are a number of schools, districts, and even states that have begun to systematically implement UDL at various levels of comprehensiveness. There is, for example, a state-wide UDL initiative in Michigan where schools are incented to adopt UDL practices in their teaching, purchasing, and policies. Many individual districts or schools throughout the country, such as Cincinnati Public Schools in Ohio and Bartholomew Consolidated School Corporation in Indiana, are doing the same.

Second, publishers have recognized the market potential of instructional materials in UDL-type formats. Tom Snyder Productions of Scholastic has released nine award-winning middle school trade books (e.g., *The Giver*; *Bud, Not Buddy*) in this format, calling them *Thinking Readers*. In early 2009, Pearson produced its core history textbook in a completely digital format with many UDL features. Other publishers are preparing to follow suit.

Third, and most importantly, the US Department of Education recently endorsed a National Instructional Materials Accessibility Standard throughout the United States. This standard – called NIMAS – requires that print instructional materials (e.g., textbooks) must be available in digital formats so that they are more accessible to students of all kinds. This legislation is remarkably revolutionary – the first of its kind to recognize that printed instructional materials are no longer adequate to

serve the full population of students in our schools. The Higher Education Bill of 2008 confirms the importance of UDL in the future of education.

While UDL is still in the very early stages of development, it seems clear that there is considerable progress throughout the various sectors of public and private education. Turning that momentum into systemic evidence-based reform will take a great deal more work. There is much to gain in doing that work.

Bibliography

CAST (2008). *Universal Design for Learning Guidelines Version 1.0*. Wakefield, MA: Author.

Further Reading

- Barsalou, L. W., Breazeal, C., and Smith, L. B. (2007). Cognition as coordinated non-cognition. *Cognitive Processing* 8(2), 79–91.
- Bloom, B. S. (1984). *Taxonomy of Educational Objectives*. Boston, MA: Allyn and Bacon.
- Cytowic, R. E. (1996). *The Neurological Side of Neuropsychology*. Cambridge, MA: MIT Press.
- Dalton, B., Pisha, B., Eagleton, M., Coyne, P., and Deysher, S. (2002). Engaging the text: Reciprocal teaching and questioning strategies in a scaffolded learning environment. *Final Report to the U.S. Department of Education*. Wakefield, MA: CAST.
- Dalton, B. and Proctor, C. P. (2008). The changing landscape of text and comprehension in the age of new literacies. In Coiro, J., Knobel, M., Lankshear, C., and Leu, D. (eds.) *Handbook of Research on New Literacies*, pp 297–324. Mahwah, NJ: Erlbaum.
- Dalton, B. and Rose, D. (2008). Scaffolding digital reading comprehension. In Block, C. C. and Parris, S. R. (eds.) *Comprehension Instruction: Research-Based Best Practices*, 2nd edn., pp 347–361. New York: Guilford Press.
- Dalton, B. and Strangman, N. (2007). Using technology to support struggling readers' comprehension: A review of the research. In Reinking, D., McKenna, M. C., Labbo, L. D., and Keiffer, R. D. (eds.)

- Handbook of Literacy and Technology*, 2nd edn. Mahwah, NJ: Erlbaum.
- Goldberg, E. (2001). *The Executive Brain: Frontal Lobes and the Civilized Mind*. New York: Oxford.
- Luria, A. R. (1973). *The Working Brain*. New York: Basic Books.
- Moran, J., Ferdig, R. E., Pearson, P. D., Wardrop, J., and Blomeyer, R. L., Jr. (2008). Technology and reading performance in the middle-school grades: A meta-analysis with recommendations for policy and practice. *Journal of Literacy Research* 40(1), 6–58.
- Palincsar, A. S. and Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction* 1(1), 117–175.
- Proctor, C. P., Dalton, B., and Grisham, D. L. (2007). Scaffolding English language learners and struggling readers in a universal literacy environment with embedded strategy instruction and vocabulary support. *Journal of Literacy Research* 39(1), 71–93.
- Proctor, C. P., Uccelli, P., Dalton, B., and Snow, C. E. (2009). Understanding depth of vocabulary and improving comprehension online with bilingual and monolingual children. *Reading and Writing Quarterly* 25(4), 311–333.
- Rose, D. H. and Meyer, A. (2002). *Teaching Every Student in the Digital Age: Universal Design for Learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Rose, D., Meyer, A., and Hitchcock, C. (2005). *The Universally Designed Classroom: Accessible Curriculum and Digital Technologies*. Cambridge, MA: Harvard Education Press.
- Rosenzweig, M. R., Breedlove, S. M., and Watson, N. V. (2005). *Biological Psychology: An Introduction to Behavioral and Cognitive Neuroscience*, 4th edn. Sunderland, MA: Sinauer.
- Sanguinetti, V. R. (2007). *The Rosetta Stone of the Human Mind: Three Languages to Integrate Neurobiology and Psychology*. New York: Springer.
- Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Relevant Websites

- <http://cast.org> – Cast, Universal Design for Learning.
- <http://nimas.cast.org> – NIMAS Development & Technical Assistance Centers.
- <http://www.udlcenter.org/>

TECHNOLOGY AND LEARNING – SCHOOL LEADERSHIP FOR TECHNOLOGY INTEGRATION

Technology as a Support for School–Community Connections

K M Culp, Education Development Center, Inc., New York, NY, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The wide range of activities and relationships covered by the terms home–school connections or school–community connections has been articulated elsewhere (Epstein, 2001; Ferguson *et al.*, 2008; Keyes and Gregg, 2001), and include building parental engagement with the educational process, enhancing educational opportunities for students' family members, and stimulating community investment in the school.

Schools need to establish multiple forms of communication and coordination of activities with families and the community in order to achieve these diverse and ambitious goals. They must find ways not only to communicate basic information, such as school calendars, emergency closings, and school policies, but also to coordinate supports and interventions with parents or guardians, to serve particular students in need of assistance. They must locate relevant expertise and resources within the community and build relationships with those individuals and institutions. Moreover, they must invite input and sustain conversation with parents and community members.

Timely and responsive communication is critical to meeting the challenge of creating strong connections between schools and students' homes, and the Internet has opened up new possibilities for organizing the relationship between schooling and the community. This article focuses on how electronic information and communication technologies (ICTs) have been used to help strengthen these connections. It focuses primarily on the period since 1993, when Internet access began to increase rapidly both in schools and in students' homes, and on how technology is supporting day-to-day community building around schools and the business of schooling. It places relatively less emphasis on the use of technology to support sustained learning experiences in the community beyond the school. A wide range of projects and research addresses this domain, and they are closely linked to a larger discussion of technology's role in supporting student learning (Blumenfeld *et al.*, 1991; Boss and Krauss, 2007; McKay and McGrath, 2006).

How Can ICTs Help?

Schools face a particularly challenging set of circumstances as they seek to communicate, coordinate, and partner with students' families and the local community. First and foremost, their relationship with their target audience is indirect; they have relatively few opportunities for direct contact with parents or community members, and must rely on students' backpacks or the local media to carry messages for them. Schools also often need to communicate over considerable geographic distances, and often in multiple languages.

Technology tools can help to address many of the obstacles to direct, engaged communication among educators, parents, and community members:

1. *Location and time.* Schools are often physically isolated buildings that offer limited opportunities for parents to enter and engage in face-to-face communication with educators. Many school schedules allow little to no time for regular communication or activities outside of instructional interactions between teachers and students. This makes engagement with parents, community members, or other community resources challenging to arrange or to sustain. Technology can support asynchronous communication, whether via voicemail or e-mail, and can help teachers make their students' work visible to others when and how they wish to, via websites, video, and digital pictures.
2. *Complex, multiple cultural differences.* As schools serve changing, diverse populations, the perspectives parents and community members bring to their engagement with the school will vary widely and will change over time. Technology allows educators to tailor their messages to individual parents and families, can support the translation of written documents, and can provide parents with access to richer resources that they can use to familiarize themselves with schools.
3. *Scale.* The scale of modern schooling often pits the general – school-wide messages or initiatives – against the specific – the need for detailed review and

discussion of individual students' needs. Technology can relieve this tension by allowing educators to reach large groups of parents quickly and easily via voice-mail, e-mail, or website announcements.

4. *Complexity.* Information about student achievement is increasingly complex, and schools face significant pressure to make that data meaningful to parents and community members. Using data management and visualization tools to support parental engagement with accountability and achievement data is a relatively new area of work, building on lessons teachers are learning as they develop their own capacity to work through and make sense of achievement test data.

Several features of ICTs would appear to be well suited to these challenges. First and foremost, ICT can remove the child as a point of transmission. E-mail messages, information posted to websites, or voicemail sent out from a central office can reach parents or the community at large without relying on the child acting as an intermediary. ICT supports both automation and customization of information, which can help educators carry out repetitive communication tasks more easily, while tailoring information to individual families or updating regularly changing information quickly and easily. Technology can also allow a broader pipeline of information and content to flow between the home and the school, particularly in schools involved in laptop computer initiatives or that use classroom websites extensively to support or document project work.

The Research Base

As Ferguson *et al.* (2008) discuss, the research base on the impact of home-school-community connections in general on student outcomes remains weak. While much descriptive work has been done, few studies have either used systematic analyses to identify the key features of promising practices or policies, or conducted rigorous research to test the connections between those programmatic features or structures and student outcomes. Even less research has looked specifically at the effective use of technology to facilitate home-school-community connections.

One meta-analysis, however, has specifically examined the literature on the use of technology to support school-community connections (Penuel *et al.*, 2002). Penuel and his colleagues report a similar state of affairs to Ferguson *et al.* (2008), explaining that there has been little substantial research done to investigate either the effectiveness of technology-based home-school supports, or the conditions of implementation that might support success, and very little that meets the necessary standards for meta-analysis. However, the authors were able to review a handful of program evaluations, and found consistent, if small, positive relationships. Note that this meta-analysis

includes projects that provided laptop or desktop computers for students to use at home, because these programs' goals included increasing parental involvement in students' schoolwork and building the technology skills of other family members. Specifically, the analysis found that these programs had, on average, a significant impact on parent-school communication. They also found that the programs had some impact on students' writing and mathematics achievement (Penuel, 2006: 41), but raised several caveats about the technical qualities of the studies under review.

Relatively little research on the impact of ICTs on home-school connections has been published since this review, suggesting that these findings are still the most rigorous ones available on the impact of technology-supported home-school connections on student learning or on levels of parental engagement in children's education.

A small body of descriptive research has also explored the implementation of projects intended to use technology to improve home-school connections (Nudell *et al.*, 2005; Poor Children Benefit, 1995). These projects do not demonstrate program impact in a systematic way. Rather, they focus on the differential impact of the socio-economic context of the use of ICTs, and how those contextual factors shape the nature of the home-school connection. These studies are an important reminder that the persistent cultural influences that shape home-school connections can be addressed but not eliminated through the introduction of ICTs intended to improve those connections.

Using Specific Technologies to Facilitate School-Community Communications

Between 1994 and 2005, The proportion of US classrooms with Internet access rose from 3% to 94% (Wells and Lewis, 2006). The ratio of students to Internet-enabled computers in classrooms also dropped during this period, reaching one computer per 3.8 students in 2005 (NCES, 2006). This period saw many research initiatives, grassroots efforts, and commercially developed projects using this new technology to support improved home-school and school-community connections.

This article briefly reviews a number of ways that schools and districts are using technology to support home-school-community connections. It groups those approaches into two main categories: practices that facilitate communication between schools and their communities, and practices designed to support the reorganization and restructuring of schooling.

E-Mail

The rapid growth in the use of e-mail during this period has raised both opportunities and challenges for teachers

and administrators who sought to improve their communications with parents and community members. As home access to the Internet has grown, schools and districts have been able to share information quickly and with a broad audience. School districts now frequently use e-mail blasts to share time-sensitive information such as announcements of emergency closings, or to solicit community support for initiatives such as upcoming budget votes or fundraising activities.

Like doctors, classroom teachers have been slow to fully embrace e-mail as a primary means of communication (Fishman, 2007). Teachers spend their working time almost entirely engaged in face-to-face work with either students or their colleagues. Their work lives have not typically been organized to accommodate daily communication with any larger audience, such as parents or community members.

Parents, in particular, could benefit from and may be eager for more frequent and detailed communication with teachers about their children's activities and progress in school. However, teacher–parent communications have typically been primarily one way (through flyers, notes, and newsletters to parents). Infrequent opportunities for interaction, such as open houses or parent/teacher conferences, are often time constrained, and more frequent communication often depends on parental initiative and scheduling conversations before or after school hours. Of course, there are many exceptions to this pattern, particularly among dedicated teachers and the parents of children facing significant difficulties in school. But on average, substantive exchanges between the teachers and parents of individual children in public schools are few and far between.

E-mail and other forms of electronic communication would appear to open up significant new opportunities for teachers and parents to align their efforts to support individual children's progress through school. However, using e-mail makes teachers more accessible to parents and community members, and an increase in the frequency of communication without any associated increase in the time available for this work. As a group, teachers have been hesitant to embrace new expectations for e-mail use, mostly likely due concerns that their already-overextended time commitments will be further strained, or that parents will expect an unrealistic level of day-to-day input about their children's experiences.

The 2007–2008 administration of the *National Center for Education Statistics* (NCES) Schools and Staffing Survey, a nationally representative survey conducted by the US Department of Education's National Center for Education Statistics, will be asking for the first time about teachers' use of e-mail to communicate with parents, and these findings will provide an important new benchmark for understanding how prevalent this form of communication really is.

While teachers have been slow to integrate regular e-mail communication into their work, the communities surrounding them have found ways to use e-mail to increase community focus on and awareness of school activities. Homegrown e-mail listservs and message boards have been common for years, bringing interested parents in contact with one another to support the dissemination of both formal and informal information and opinions about local schools. Similarly, local membership groups such as parent–teacher organizations, booster clubs, and parents supporting athletic programs all use e-mail lists commonly now to share information. While increased communication and coordination in groups like these may seem unremarkable at this point in the development of consumer communication technologies, a byproduct of these activities is likely an increased self-identification among participants as members of some coherent, meaningful grouping. This kind of secondary impact among active members of online communities is well documented for other kinds of interest groups (Goodsell and Williamson, 2008; Drentea and Moren-Cross, 2005), and is likely to be just as true for the communities of parents and other supporters surrounding specific schools or school districts.

Websites

When web browsers first began to appear in the mid-1990s (Netscape was first released to the public in 1994), teachers who had already been interested in using technology to support instruction quickly identified websites as an important new tool for sharing classroom activities with parents. A small but enthusiastic subset of teachers learned hypertext markup language (HTML), and classroom websites displaying student book reports, drawings, and science projects multiplied rapidly. Teachers hoped that by making student work more broadly visible, they would give their students the experience of sharing their work with a broader audience, and would provide parents with a richer window into the work their students were doing in the classroom.

Schools and districts with motivated technology coordinators also started creating websites, publishing curriculum standards, school calendars, and monthly lunch menus, putting their library catalogs online, and sharing pictures of school events. The Web provided a powerful new opportunity for schools and districts to represent themselves and to share detailed information with their community. A well-designed website allowed a school district to knit together information from the different administrative layers and physical sites that constitute the district to represent a cohesive, interconnected whole – something that previously was only accessible as a tacit body of knowledge shared by those already deeply engaged with the district. A family new to a school district, for example, might gain both concrete information and a

powerful set of messages about the tone, goals, strengths, and focus of a school district by visiting the district website. Such sites also opened schools up to new connections with other districts, schools, and individual classrooms around the country and the world.

By the late 1990s, commercial software publishers had moved into the school and district website business, and they continue to provide schools and districts with packages that can be customized to support a range of web-based activities, including supporting parental engagement in students' learning and in school activities. These tools tend to focus on information delivery, providing schools with tools to help them provide parents with a view of their students' progress, grades or attendance, or information about school or community activities, but rarely support broadly interactive or highly personalized forms of communication or interaction.

Summary

Schools and districts are slowly but steadily integrating ICTs into traditional forms of communication and information sharing between schools and their communities. However, much of the ongoing work of forging these connections continues to involve little use of technology. This may reflect a persistent focus among advocates and educators on face-to-face communication, and bringing community members together for special events and sustained learning opportunities. An implicit goal in this work is maintaining the brick-and-mortar school as a pivotal physical location in the community – a point of interaction, exchange, and mutual support. This leaves ICTs as important but tangential tools, improving communication among stakeholders but not playing a central role in shaping the work that they do together.

Using ICTs to Facilitate a Reorganization of the School–Community Relationship

Some educators, researchers, and software and infrastructure companies are using technologies to develop other approaches to schooling. These approaches involve a much broader reframing of relationships among students, their families and communities, and the experience of schooling. In various ways, they challenge the assumption that the brick-and-mortar school is at the center of the educational equation, and increasingly frame the individual student and his or her family as a consumer, placing their needs and priorities at the center and repositioning the school as a provider of services.

One-to-One Computing

Since the early 1990s there has been a gradual increase in initiatives to provide every student with a laptop

computer. These efforts have ranged in scope from individual grade levels in a single school to entire states. For an overview of the evolution of these initiatives, see Keefe and Zucker (2003). For a detailed review of the research base on one-to-one computing, see Penuel (2006). Rationales for providing students with laptop computers, and specific goals for such programs, have varied widely. Typically, these programs not only focus on the opportunity to make technology use ubiquitous in the school, but also share goals that involve strengthening home–school connections by ensuring that all students have computer access at home, and that technology access is seamless between the home and school. Perhaps the best-known statewide initiative has been the Maine Learning Technology Initiative, which provided all seventh- and eighth-grade students in the state, and their teachers, with laptop computers. Silvernail and Gritter (2007) provide an overview of the program, and discuss evaluation findings that demonstrate an association between use of the laptop computers and improvement in students' writing skills.

An earlier example was the Buddy Project, which provided students with desktop computers at home, was active in Indiana between 1988 and 2000, and placed a particularly strong emphasis on home–school connections. Policymakers who supported the program hoped that providing laptops to students would reengage parents in their students' learning, and encourage the growth of family members' technology skills Lemke and Martin (2004). (The Buddy Project continues in a somewhat different form, with an increased focus on writing skills. See Lemke and Martin (2004) for a detailed discussion of the project.)

One-to-one computing initiatives are also gaining visibility worldwide, as the One Laptop Per Child (OLPC) project initiative and Intel's marketing of the low-cost ClassMatePC bring one-to-one computing to diverse communities in many countries. Research, following implementation of the OLPC project or Intel's laptop computers, may provide a pathway for looking more broadly at what roles portable computers can play when schools and communities are defined very differently than they are in the United States. In a recent example, Urrea's (2008) dissertation looked closely at one-to-one computer use (using the OLPC computers) in a small, rural Costa Rican classroom, and grounded her interpretation in the relevance of the program to strengthening connections between the one-teacher school and the surrounding community.

Sharing Accountability Data

Perhaps the largest shift of all in US K-12 education during the 1990s has been the investments made in large-scale testing to meet the accountability expectations of the

Elementary and Secondary Education Act of 2001, better known as No Child Left Behind (US Department of Education, 2002). In response to a range of expectations – from parents who want evidence of school effectiveness to local business owners anxious to use test scores as an indicator of the quality of local schools and hence of the health of the local community – schools have been seeking ways to make testing data accessible and meaningful to their communities. Given the complex and often time-sensitive nature of testing data, educators have looked to technology solutions to help them summarize, share, and explain multiple kinds of accountability data both internally and with community members.

Tools now being used for these purposes range from homegrown PowerPoint presentations to complex data management systems developed by large testing companies (see, e.g., Pearson Benchmark). Some of these systems include components for sharing grades, attendance data, and other information about individual students with their parents (see, e.g., the Skyward system). Other external vendors have also stepped in to address parents' and community members' interest in performance data from their schools. Organizations, including SchoolMatters and GreatSchools, are building front ends to testing databases and inviting parents and community members to review evidence about schools' performance and often to add their own reviews describing their perceptions of their children's schools. Just for the Kids, a site supported by the National Center for Educational Achievement, takes a more complex approach, not only allowing interested individuals to investigate a particular school's performance, but also providing comparisons to other, similar schools in the state. None of these systems, however, is specifically investing in helping parents to explore and understand their own child's testing data or to put their child's achievement in the context of larger patterns of performance within a school or district.

While these sites are likely to be valuable to parents, they also mark a significant movement away from a key assumption of the home–school–community connection literature – that the school is a community created in large part by the voluntary contributions and efforts of the parents whose children attend it. Instead, these services speak to parents as consumers – individuals seeking out, selecting and purchasing those services they choose to receive. This is markedly different from the traditional goal of treating parents as potential participants in the educational process. Like many other services and products being developed and delivered on the Web, these sites invite parents to seek out particular services that meet their perceived needs and characteristics, which is at odds with the traditional role of the school as a point of common ground where families and educators interact and influence one another over time.

Distance Learning and Virtual Schools

Since the 1960s, schools have used various technologies to provide their students with access to courses they would not otherwise have the opportunity to take (Reiser, 1987). The technologies used to support distance learning have evolved from closed-circuit television systems to video-conferencing to web-based streaming video or text- and audio-driven systems such as Elluminate. A history of the evolution of these technologies is available in Reiser (1987), and a summary of the research base on the effectiveness of distance learning is available in Allen *et al.* (2006). A summary on the effectiveness of web-based learning is available in Means *et al.* (2009).

In the mid-1990s, as schools built their access to the Internet, educators naturally were eager to use this new resource to supplement their traditional course offerings. In the following years, schools and universities, as well as an increasing number of for-profit vendors, have begun to offer districts access to individual courses and soon comprehensive curricula. However, policymakers also began to consider not only how the Internet might support communication between students, teachers, and parents, but also could make possible entirely new approaches to defining where and how students could learn. By 2000, an influential report from the Web-Based Education Commission squarely focused its description of the power of the Internet for learning on the opportunity to address learners as individuals, moving away from the classroom or even the school as an organizing unit and appealing to the notion of anytime, anywhere learning (Web-Based Education Commission, 2000).

As high-speed networks became almost ubiquitous in the early 2000s, web-based schooling began to be a viable, and potentially profitable option. Schools and districts use many different mechanisms for developing or acquiring and delivering online courses, ranging from using entirely locally created efforts to partnering with local universities to purchasing services from large corporations. For-profit vendors, such as The Connections Academy, Edison School Inc., and Insight Schools, play a significant role in delivering online learning opportunities. These vendors provide districts with a comprehensive school program, delivered entirely or almost entirely online, and attended by students who work entirely outside of the traditional school building. The emergence of this industry has also dovetailed with state and federal support for charter schools, a policy mechanism that makes it possible for school districts to buy into for-profit online schools. In the USA, the National Education Technology Plan of 2004 strongly argued in favor of the further development of this model, presenting it as one of the seven top priorities for education technology spending (US Department of Education, 2004: 41). For a more detailed review of the current state of online learning opportunities, see Tucker (2007).

Considering these types of schools from the traditional framework of home–school connections raises challenges. They can be seen as strengthening the connection of individual students to their communities, as they offer an in-between strategy such that a student who might otherwise be home-schooled can still be served through the public education system. Students with significant health challenges can also continue to receive a public education when they are unable to attend school physically. However, on the other hand, when students work entirely from distant locations, traditional notions of the school as a distinct grouping of students that not only intersects with students' home communities but also develops and sustains its own structures and culture, largely fades away.

Summary

Each of the three kinds of efforts described above (one-to-one computing, virtual schools, and school comparison sites) reflects two changes that have occurred well outside of individual schools or districts. First, technology has continued to evolve rapidly, and consumer electronics have continued to drop in price. While homes and schools are mostly on par with one another in terms of basic computer-based Internet access, schools have not kept pace with domestic investments in cell phones and other portable, Internet-enabled devices (Shuler, 2009). As youth and many of their parents move toward an online-all-the-time way of living, enabled by social networking sites, smart phones, and laptop computers, schools have moved far more slowly to adapt to and accommodate these new tools. Consequently, new cultures of communication, consumerism, cultural participation, and community building have emerged and are fostering enormous amounts of youth participation, but have made few inroads into traditional schools.

At the same time, technology and media industries and the federal government have largely moved their investments away from school-based instructional technologies and toward using new technologies to support other organized learning opportunities that appeal directly to parents and bypass the brick-and-mortar school. For example, institutions such as the Discovery Channel and the Public Broadcasting Service (PBS) have invested heavily in developing online educational media that will appeal to children, and providing supports to parents to help them make effective use of the resources. The German publisher Bertelsmann is developing a product called Scoyo that will offer extensive content-rich online educational resources and will be marketed directly to parents and their children for home use. These same organizations do invest in teachers through other efforts, such as PBS' TeacherLine professional development program. However, these initiatives are distinct from one another

and few connections are made that might encourage teachers to make links between their own training and the experiences that their students might be having at home using similar resources.

These evolutionary forces are also reflected in the emergence of School 2.0, a phrase that describes a loose movement of school technologists, researchers, and policymakers who are re-imagining how schooling might function if educators were able to embrace current Web 2.0 notions like personal broadcasting, two-way communication, group construction of knowledge, and intensive social networking. For examples of current discussions of School 2.0, see SRI's School 2.0 eToolkit, the for-profit Eduvo.com and The wiki site school20.wikispaces.com. These conversations echo early discussions about the potentially transformative power of microcomputers in the 1980s (see the archives of the Apple Classroom of Tomorrow project, or an overview of past proceedings of the National Education Computing Conference, both available online). Like those earlier conversations, School 2.0 discussions make connections between the flexibility and transparency provided by Web 2.0 tools and a shared desire to create schools in which student learning is more learner directed, open ended, and happening largely outside of the school walls. While some functioning schools are explicitly focused on School 2.0 concepts (Lehmann, 2007), many educators have little opportunity to institutionalize such core School 2.0 concepts as open campuses, flexible curricula, and ubiquitous access to portable technology. To date, discussions of School 2.0 have not yet really focused on parental engagement or home experiences, even though the visions of schooling they promote would have significant consequences for parents' relationships to the educational process.

Conclusion

Traditional schools are still adapting to the technological innovations of the 1990s. For educators seeking to expand and improve communication between schools and their communities, many aspects of new communication tools are appealing, and will allow them to be more responsive, detailed and timely in their communications than ever before. However, other potential uses of technology to sustain and deepen these relationships remain largely untapped, in part because the tools they would involve are still challenging and new for many schools.

The Penuel *et al.* (2005) review of the literature on technology and school–community connections provides limited but positive evidence that technology can help improve relationships across school boundaries that can make a difference in student's achievement, parents' level of involvement, and student engagement in schools. While

little research is currently underway in this domain, it seems important to look more closely at how schools that are ahead of the curve in adopting more innovative technologies (such as handheld devices or social networking sites) may be using those resources to connect with students' lives outside of the school building, as well as with families and community members. Similarly, advocacy and research organizations in the fields of parental involvement and school–community partnerships need to be encouraged to consider how technology can be used to extend and enhance their work, and to include it as a factor to be examined in their research efforts.

Penuel's findings also suggest some areas for a new line of research focused on the principles of 'School 2.0.' What kinds of information, we need to ask, do parents most need access to when their students are participating in this kind of school experience? When the boundaries of brick and mortar or the 3:45 afternoon bell are no longer in place, what still needs to be shared among educators and parents to support students effectively? At a time when students' facility with technology is being praised and examined extensively, it may also be important to continue to examine how students' and parents' access to technology can best be leveraged to ensure that parents can be full partners in their children's education.

See *also*: Equity in Technology Access and Opportunities.

Bibliography

- Allen, M., Mabry, E., Mattrey, M., Bourhis, J., Titsworth, S., and Burrell, N. (2006). Evaluating the effectiveness of distance learning: A comparison using meta-analysis. *Journal of Communication* **54**(3), 402–420.
- Blumenfeld, P. C., Soloway, E., Marx, R. W., and Krajcik, J. S. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist* **26**(3 and 4), 369–398.
- Boss, S. and Krauss, J. (2007). *Reinventing Project-Based Learning: Your Field Guide to Real-World Projects in the Digital Age*. Eugene, OR: International Society for Technology in Education. <http://www.iste.org> (accessed August 2009).
- Drentea, P. and Moren-Cross, J. L. (2005). Social capital and social support on the web: The case of an Internet mother site. *Sociology of Health and Illness* **27**(7), 920–943.
- Epstein, J. (2001). *School, Family, and Community Partnerships: Preparing Educators and Improving Schools*. Boulder, CO: Westview Press.
- Ferguson, C., Ramos, M., Rudo, Z., and Wood, L. (2008). The school-family connection: Looking at the larger picture. A review of current literature. *Southwest Education Development Laboratory, National Center for Family and Community Connections with Schools*. Austin, TX: SEDL. <http://www.sedl.org/connections/resources/sfclitrev.pdf> (accessed August 2009).
- Fishman, E. S. (2007). Physicians E-mailing patients. Transcript of a podcast, #17 in a series. Manhattan Research, New York. http://www.manhattanresearch.com/podcast/physician_patient_emailing.aspx (accessed August 2009).
- Goodsell, T. L. and Williamson, O. (2008). The case of the brick huggers: The practice of an online community. *City and Community* **7**(3), 251–271.
- Keefe, D. and Zucker, A. (2003). *Ubiquitous Computing Projects: A Brief History*. Ubiquitous Computing Evaluation Consortium. Menlo Park, CA: SRI Center for Technology and Learning. <http://www.k12one2one.org/Reports.aspx> (accessed August 2009).
- Keyes, M. and Gregg, S. (2001). School–community connections: A literature review. ERIC document #ED459047. http://eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=trueand_ERICExtSearch_SearchValue_0=ED459047and_ERICExtSearch_SearchType_0=noandaccno=ED459047 (accessed August 2009).
- Lehmann, C. (2007). *Designing School 2.0: A study of Philadelphia's Science Leadership Academy*. Design Share. <http://www.designshare.com/index.php/articles/science-leadership-academy/> (accessed August 2009).
- Lemke, C. and Martin, C. (2004). One-to-one computing in Indiana: A state profile. The Metiri Group. http://www.metiri.com/NSF_Study/INProfile.pdf (accessed August 2009).
- McKay, M. and McGrath, B. (2006). Real world problem-solving using real-time data. *International Journal of Engineering Education* **23**(1), 36–41.
- Means, B., Toyama, Y., Murphy, R., Bakia, M., and Jones, K. (2009). Evaluation of evidence-based practices in online learning: A meta-analysis and review of online-learning studies. Washington, DC: U.S. Department of Education, Office of Planning Evaluation, and Policy Development, Policy and Program Studies Service.
- Nudell, H., Gersick, A., Culp, K. M., Ba, H., and Tally, B. (2005). The ecology of children's computing. New York, NY: EDC/Center for Children and Technology. http://cct.edc.org/report_summary.asp?numPublicationId=202 (accessed August 2009).
- Penuel, W. R. (2006). Implementation and effects of one-to-one computing initiatives: A research synthesis. *Journal of Research on Technology in Education* **38**, 329–348.
- Penuel, W. R., Kin, D. Y., Michalchik, V., et al. (2002). *Using Technology to Enhance Connections between Home and School: A Research Synthesis*. Menlo Park, CA: SRI International. ctl.sri.com/publications/downloads/Task1_FinalReport3.pdf (accessed August 2009).
- Poor Children Benefit (1995). Study Claims Poor Children Benefit a Lot From Learning to Use Telecommunications. *CyRev: A Journal of Cybernetic Revolution, Sustainable Socialism and Radical Democracy* (2). <http://www.net4dem.org/cyrev/archive/issue2/articles/TCOMStudy/TCOMStudy1.htm> (accessed August 2009).
- Reiser, R. A. (1987). Instructional technology: A history. In Gagne, R. M. (ed.) *Instructional Technology: Foundations*, pp 11–48. Hillsdale, NJ: Erlbaum.
- Shuler, C. (2009). *Pockets of Potential: Using Mobile Technologies to Promote Children's Learning*. New York: The Joan Ganz Cooney Center at Sesame Workshop. www.joanganzcooneycenter.org (accessed August 2009).
- Silvermail, D. and Gritter, A. K. (2007). Maine's middle school laptop program: Creating better writers. *Technical brief of the Maine Education Policy Research Institute*. Gorham, ME: University of Southern Maine. <http://usm.maine.edu/cepare/> (accessed August 2009).
- Tucker, B. (2007). *Laboratories of Reform: Virtual High Schools and Innovation in Public Education*. Washington, DC: Education Sector.
- Urrea, C. M. (2008). One to one connections: Building a community learning culture. *Dissertation Abstracts International Section A: Humanities and Social Sciences* **68**(10-A), 4267.
- U.S. Department of Education (2002). No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425.
- U.S. Department of Education, Office of Educational Technology (2004). Toward a new golden age in American education: How the internet, the law and today's students are revolutionizing expectations. Washington, DC: U.S. Department of Education. <http://www.ed.gov/about/offices/list/os/technology/plan/2004/site/edlite-default.html> (accessed August 2009).
- Web-based Education Commission (2000). The power of the internet for learning: Moving from promise to practice. Washington, DC: U.S. Government Printing Office. <http://www.gov/offices/AC/WBEC/FinalReport/WBECReport.pdf> (accessed August 2009).
- Wells, J. and Lewis, L. (2006). Internet access in U.S. public schools and classrooms: 1994–2005 (NCES 2007-020). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Relevant Websites

<http://www.apple.com> – Apple Classrooms of Tomorrow.

<http://www.greatschools.net> – GreatSchools.

<http://www.iste.org> – International Society for Technology in Education:
National Education Computing Conference: NECC and the evolution
of technology in education in the U.S.

<http://www.just4kids.org> – Just for the Kids.

<http://www.csos.jhu.edu/P2000> – National Network of Partnership
Schools at Johns Hopkins University.

<http://www.laptop.org> – One Laptop Per Child.

<http://www.pbskids.org> – PBS TeacherLine and Parent resources.

<http://etoolkit.org> – School 2.0 eToolkit.

<http://School20.wikispaces.com> – School 2.0 wiki.

<http://www.schoolmatters.com> – SchoolMatters.

<http://www.scoyo.com> – Scoyo.

<http://www.skyward.com> – Skyward.

<http://www.sedl.org> – Southwest Education Development Laboratory's
National Center for Family and Community Connections with
Schools.

<http://www.cosmeo.com> – The Discovery Channel: Cosmeo.

TECHNOLOGY AND LEARNING – SUPPORTS FOR MANAGING INSTRUCTION

Classroom Uses of Technology to Manage Instruction

W R Penuel, Center for Technology in Learning, CA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Instructional-management systems –

Technologies that support teachers to develop lesson plans, track student progress on assessments, and record students' instructional experiences.

Student-information systems – Systems that enable tracking and reporting of student attendance, grades, course enrolment, test scores, and other data. Access to data from student information systems varies, depending on the data source. Teachers and parents often have access to some but not all data in the system.

Student-response systems – Technologies that rely on handheld devices for student input to answer questions posed by teachers. The systems also allow teachers to display student responses in the aggregate with a projection device, to facilitate awareness of the class distribution of responses.

Introduction

Managing instruction encompasses the range of ways in which educators plan, coordinate, and evaluate how students interact with curricular materials, the teacher, and each other. As students' interactions can never be fully determined ahead of time and are – even within the same classroom – quite varied depending on students' interests and abilities, managing instruction requires continuous monitoring and adjustment. For this reason, management of instruction requires strategies and can be enhanced with tools that support ongoing cycles of planning, enacting, and evaluating instruction.

Within education, instructional management takes place at different levels and different timescales. Policy-makers at the state level manage instruction typically by setting standards and monitoring progress through annual accountability testing. Districts and schools provide teachers with textbooks and curricula for use in planning and

coordinating instruction, and most invest in teacher professional development to support use of these instructional materials. To evaluate instruction, districts and schools typically rely on annual assessment data in the aggregate, but increasingly they also examine disaggregated data for individuals and groups and use benchmark tests administered at multiple points during the year to gauge students' progress while there is still time to take corrective action. For their part, teachers manage instruction by adapting materials and activities they have been given by their school or district, implementing mandated testing systems, and engaging in a range of formal and informal classroom-assessment activities.

Technology exists to support individual teachers in managing instruction for students over both the long and short term. Student-information systems (SIS) support long-term management of instruction; they are technologies that support sharing and reporting of basic data on schedules, test results, and background information on students among teachers, administrators, and parents in a school. Over the medium term, instructional-management systems (IMS) can support teachers in developing lesson plans, tracking student progress on assessments, and recording students' instructional experiences over the course of the school year (Archibald, 2001). Another class of technologies supports teachers in coordinating and evaluating instruction as it takes place, allowing teachers to adjust instruction on the basis of up-to-date information about what students know and can do.

This article focuses on two classes of technologies that have been developed to help coordinate students, the teacher, and the curricular content as teachers manage instruction in their classrooms. These two classes are distinct from SIS in that the primary users of the system are students and teachers, rather than administrators and parents. They are distinct from IMS in that they enable different forms of teaching and interaction with content, and they are intended to provide teachers and students alike with real-time data on the effectiveness of instruction. The article provides a review of the features of these classes of technology and the research evidence related to their effects on teaching and learning. The concluding

section focuses on future directions of technology development and research on these classes of technology designed to help teachers manage instruction.

Student-Response Systems

Student-response systems are a relatively mature class of technologies in use for more than 20 years in elementary, secondary, and higher education. These technologies enable teachers to collect and display student responses to teacher questions and to use this information to evaluate student understanding and the effectiveness of their instruction. Indirectly, the technology also supports teachers in their role as managers of instruction by facilitating more active forms of student participation in class, especially in the midst of whole-class instruction.

Features of Student-Response Systems

The core features and required set ups for student-response systems are relatively few. Student-response systems are comprised of a single teacher computer, a projection system for the teacher, a handheld response device for each student, and software that uses infrared frequencies to transmit information from the student devices to the teacher laptop. In some systems, the handhelds are wirelessly connected to a classroom network or the Internet so that the teacher and student devices can be linked without cumbersome cables.

A primary function of the software on the teacher's computer is to input teacher questions. Most systems support primarily selected-response (multiple-choice) types of questions, but a few now support open-ended or textual responses. Teachers can enter a question ahead of time and display it for students, and they can also pose questions without having to enter them as text into the software. For purposes of planning instruction, teachers can consult banks of questions that instructors at the postsecondary level have found useful for testing students' conceptual understanding or for sparking discussion (see, Mazur, 1997, for examples in physics).

Classroom-communication-system software also tracks student responses as they come in and, with the aid of a projection device, displays the student responses for all to see. The tracking includes a real-time display of whether each handheld device has responded to a particular question. A teacher can register each handheld to a particular user and record those users' responses to questions in an electronic grade book that persists beyond an individual classroom session. When ready, the teacher can display the student responses, and a key feature here is that the default mode of display is to show the percentage or number of students who have given a particular response without associating individual responses with a student's name. The result is that the display is an anonymous display of

results, but each student can judge how many other students gave the same answer that they did. In addition, the display does not necessarily show the right answer, so the teacher can stop and re-teach or ask students to explain their answers before she or he decides to identify the correct answer and give an explanation.

It is possible to support the basic functions of student-response systems without technology, by giving students placards with answer choices that only their teachers can see (Judson and Sawada, 2002; Randolph, 2007). But there are some advantages of student-response systems over placards. For example, the response systems allow teachers to keep records of student responses for later review with little added effort required of the teacher. In addition, response-system technology allows the cycle of question and answer to take place in a very short period of time. It provides students and teachers with more rapid feedback on progress without slowing down the pace of teaching (Roschelle *et al.*, 2004).

Teaching with Student-Response Systems

Instructors have put forth models for how best to teach with student-response systems, and researchers have studied teaching practices that help teachers effectively manage instruction.

Models for teaching with student-response systems

There have been at least two models developed for promoting interactive discussion in conjunction with response-system technologies, peer instruction (Mazur, 1997) and the assessing-to-learn model (Dufresne and Gerace, 2004). Both models include time for discussion, but they emphasize to different degrees the role of individual contributions, peer discussion, and class-wide discussion.

Peer instruction is a model for teaching with student-response systems that Eric Mazur at Harvard University developed for use in his introductory physics class. In peer instruction, lecture periods are broken into a series of 10–15-min periods, and each period focuses on a particular concept. At the end of each period, the instructor asks students to answer three or so questions that require the students to demonstrate their qualitative understanding of the concepts. Students have 1 minute to think and then record their answers using the system, as well as to record their level of confidence in their answers. After recording their answers, students turn to a peer to discuss and explain the reasoning behind their answer. Then, after 1 or 2 more minutes, the instructor asks students to record a revised answer and to judge their confidence in their new answer. If the teacher is satisfied that the students have grasped the concept (based on the percentage of correct answers the

second time around coupled with high levels of confidence), then she or he can move on. If not, the instructor can repeat the cycle with a new question on the same subject.

In assessing to learn, as with peer instruction, teaching with student-response systems is organized around specific items. The items in assessing to learn are intended to help teachers diagnose student difficulties, and the answer choices include both correct answers and common misconceptions. The assessing-to-learn model emphasizes whole-group discussion rather than peer discussion. After the system tabulates responses, the teacher displays the distribution to the class as a whole, in order to ensure that all students know how others in the class have responded. Then, the teacher asks a volunteer to explain his or her answer until each different answer choice has been represented. Class discussion ensues, in which the teacher facilitates discussion so as to elicit students' thinking on the topic before offering a correct answer and explanation for the question.

Research on teaching with student-response systems

Much of the research on teaching with student-response systems confirms the importance of presenting good conceptual questions to students, as encouraged in peer instruction and assessing to learn. To stimulate discussion, for example, researchers have suggested that questions that yield divergent student responses are more effective than those that are easy or that lead all students to a single incorrect answer (Wit, 2003). For teachers using response systems to assess student learning, it appears that questions that elicit students' preconceptions and that help teachers adapt their teaching to the needs of students are most effective (Draper and Brown, 2004; Wit, 2003). The timing of questions further shapes the nature of the information that an instructor gains about student understanding. Questions posed after a lecture or explanation can be used to check understanding (Dufresne *et al.*, 1996a). By contrast, posing questions before a lecture tends to elicit preconceptions in ways that can be used to shape instruction (Dufresne *et al.*, 1996a). Structuring opportunities for peer or whole-class discussion also appears to be a critical aspect of promoting greater classroom interaction with response systems (Dufresne and Gerace, 1994; Judson and Sawada, 2002). Researchers have observed that discussion prompts students to engage in alternative ways of thinking about a concept or problem (Dufresne *et al.*, 1996a) and aids in developing deeper student understanding of the meaning of concepts (Judson and Sawada, 2002). Explaining one's reasoning to a peer has the potential to transform students' misconceptions (Judson and Sawada, 2002). Response systems facilitate discussion by providing an anchor (aggregate responses on a shared display) and set of artifacts to which students can refer in the process of building knowledge (Truong *et al.*, 2002).

Researchers acknowledge that difficulties are often encountered when response systems are introduced to classes. Many teachers find it difficult to adjust their teaching style to incorporate interactive engagement into their lectures (Dufresne *et al.*, 1996a). Further, not all subject matters lend themselves well to the kinds of factual and conceptual questions that response systems are designed to accommodate. Research with response systems has focused primarily on the domains of physics, engineering, and computer science, domains in which the ability to learn target or correct answers to conceptual questions is critical (Draper and Brown, 2004). In other disciplines, teachers have found it useful to use student-response systems to pose different kinds of questions to students, about topics such as their perceived interest or boredom in class or their perspectives on some social or historical issue (Anderson *et al.*, 2003; DiGiano *et al.*, 2003; Piazza, 2002; Sung *et al.*, 2004).

One of the key limitations of this class of technologies is the limited support for posing open-ended questions. Scholars who have developed newer technology systems have tended to criticize student-response systems as limited by their emphasis on social comparisons in data – that is, comparisons of how many students got a particular answer correct – rather than on the subject matter itself (Penuel *et al.*, 2004). Some companies marketing student-response systems have developed capabilities for different kinds of student input, in order to address the need for greater variety in question types (Robinson, 2002).

Effects of Student Response Systems

Some of the best empirical evidence concerning the effectiveness of the classroom networks in improving student-learning comes from university-level physics instruction. A number of studies have found that there is an association between students' understanding of introductory physics concepts and use of peer instruction (Fagen *et al.*, 2002; Mazur, 1997). When teachers use interactive engagement methods similar to peer-instruction methods, students in these teachers' classes make higher gains on both the force-concept inventory and the Halloun-Hestenes mechanics diagnostic test than students in comparison classrooms (Hake, 1998; Sokoloff and Thornton, 1997). In neither of these sets of studies, however, is the specific contribution of technology to students' conceptual gains explored. A study of students in engineering classes has similar results and limitations: students made conceptual gains, but the specific contribution of technology cannot be inferred due to limitations in the experimental design (Roselli and Brophy, 2002).

Several studies have also reported that the use of classroom networks helped improve student engagement and participation in class (Boyle and Nicol, 2002; Burnstein and

Lederman 2001; Crouch and Mazur, 2001; Cue, 1998; Dufresne *et al.*, 1996a; Fagen *et al.*, 2002; Kaput and Hegedus, 2002; Poulis *et al.*, 1998; Truong *et al.*, 2002; Webking, 1998; Woods and Chiu, 2003). In some studies, higher attendance has also been reported as an indicator of increased student interest and enjoyment of the class (Duncan, 2005; Ganger and Jackson, 2003; Wit, 2003). The explanation of these findings may come from a changed perception among students that it is important that they participate actively in the class. For example, students often report that seeing their own response in a group display makes them feel more accountable (Davis, 2003; Draper and Brown, 2004). The fact that all students pick an answer – even when they are uncertain about it – that then becomes part of an aggregated display stimulates students to think about the teachers' questions (Dufresne *et al.*, 1996a; Nicol and Boyle, 2003).

The assessment functions of classroom networks appear to enhance both teachers' and students' awareness of the students' level of understanding of the subject matter being taught. Researchers report that classroom networks help students assess their own understanding (Draper and Brown, 2004; Piazza, 2002). At the same time, the systems help teachers identify areas where individual students may need additional instruction or where the entire class may benefit from more and different encounters with the subject matter (Draper and Brown, 2004). In these accounts, researchers again point to the role of the shared display as critical. The display helps students know where they are in relationship to others (Dufresne *et al.*, 1996b). If they see that others are having difficulty, they may have the confidence to ask for help or for clarification when they do not understand something (Anderson *et al.*, 2003; Dufresne *et al.*, 1996a).

Classroom-Network Technologies

Newer technologies – often called classroom-network technology – provide an expanded set of capabilities for supporting classroom interactions. These newer systems provide expanded options for linking students together with the teacher to support group interaction as well as the generation of data on student thinking. Classroom-network technologies have in common with student-response systems a shared display of student responses, around which students and teachers can focus attention and activity (Penuel *et al.*, 2004). As with student-response systems, classroom-network technologies provide students and teachers with a representation of student thinking, which both provides feedback on student understanding in class and anchors discussion. This newer class of technologies, however, is distinct from student-response systems in its features and in the ways in which teachers and students interact with

technology. As the technology is more recent, however, there is less research on its effectiveness.

Key Features of Classroom Network Technologies

What distinguishes more sophisticated classroom-network technologies from student-response systems are the inputs supported and the ability to allow specific instructional content to structure the activity. Student-response systems largely limit teachers to posing multiple-choice or other closed-ended questions of students in classrooms. Newer classroom-network technology enables students to input a variety of answers to teacher tasks, including open text and graphical images. In addition, this classroom-network technology also supports the teaching of specific concepts by allowing students to identify themselves as, for example, mathematical objects in a Cartesian plane (Kaput and Hegedus, 2002). In this and other examples, technology is not content independent, as with student-response systems; rather, the technology helps structure students' interaction with content in ways that help make difficult concepts visible and more concrete to students.

The components of classroom-network technology include software for teachers and students to use, a teacher computer, student devices, and a projection device. The software capabilities differ, depending upon the subject matter. For example, in a mathematics classroom where students are studying linear functions, the software supports students being able to project a point of their own choosing particular x, y coordinates onto a Cartesian graph displayed using a graphing calculator. By contrast, in a science classroom, students studying how the coordination of traffic lights influences congestion on city streets might use that same calculator to control a particular traffic light. The display runs a simulation showing the effects of their action, and students can change their own timing of their traffic light to examine its impact on congestion (Wilensky and Stroup, 2000).

The above two examples illustrate the potential of a commonly used student-input device within contemporary classroom-networks: graphing calculators. Such calculators allow students not only to carry out computations of complex mathematics but also to input text. Thus, the calculators have all the functionality of student-response systems, plus they are devices that can be used to teach specific algebraic and geometric concepts.

Classroom-network applications of the kind now being developed have uses that cannot be replicated easily without the use of technology. No analog to individual students projecting themselves as points on a graph that displays the aggregate results of points on a single line exists that can be quickly implemented in a classroom. Similarly, there is no easy way without technology for students to develop and test simulations of traffic in order

to understand how local interactions (traffic lights turning red or green) can lead to complex aggregate patterns of behavior such as traffic. As such, new classroom-network technology offers a potentially transformative approach to managing instruction, especially in mathematics and science classrooms where they have been most widely used to date.

Teaching with Classroom Networks

Teaching with classroom-networks provides for distinctive roles for students that teachers must be prepared to support if they are to expect students to benefit from interacting with curriculum content using networks. In particular, advanced network technologies foster active participation in class by positioning students as agents in class-wide activity. In participatory simulations, for example, students do not manipulate a system or process from outside but from within it (Colella, 2000). In these simulations, each person's agreement to participate in a collaborative, class-wide effort at experimentation is critical to enable the class to construct a model of the system being simulated (Wilensky and Stroup, 1999). The system structure and students' roles in it thus exert a pressure to participate which can be amplified by social processes in which students attempt to coordinate each others' activities (Andrews *et al.*, 2003).

The specific subject matter plays a much more central role in enabling and constraining certain kinds of classroom activity with advanced network technology. The classroom display, rather than representing a distribution of students' answers to a question, might represent a Cartesian plane. Each student can manipulate his or her point in this plane with a handheld device such as a graphing calculator (Kaput and Hegedus, 2002). Researchers argue that the fact that the shared display represents a mathematical space is critical because it provides a way for students to develop an embodied, personalized viewpoint on mathematical concepts (Hegedus and Kaput, 2004). Moreover, as more work is represented publicly, researchers argue that learning can increasingly occur in the social space to complement the individual device-interaction space (Ganger and Jackson, 2003).

Within the subject-matter space, students do not have to be guided only toward correct answers. Advanced networks can support a pedagogy that emphasizes the role of generative, space-creating play in fostering learning (Stroup *et al.*, 2002). For the play to be effective, it cannot be not an unstructured exploration of the subject matter, but it must be bound by rules that designers have implemented or that the students themselves have agreed to in class (Stroup *et al.*, 2002). Especially when students are allowed to construct the rules for play, they are able to test different conjectures and hypotheses and potentially come to a deeper understanding of concepts (Wilensky and Stroup, 2000).

Effects of Classroom Networks

To date, little research has been conducted on the effects of classroom networks. However, there is some preliminary evidence that participatory simulations that rely on network technology can be effective in promoting student learning (Hegedus, 2003; Lonsdale *et al.*, 2004; Wilensky and Stroup, 2000). For example, one study examined what students learned when they engaged in a participatory simulation focused on developing an understanding of how climate change might affect the water cycle. Students used mobile phones and a floor display that showed a map of the water cycle. This study found that students in the participatory simulation learned more about the underlying rules of the system than did students who participated in a control condition in which students participated in a workshop on the water cycle (Lonsdale *et al.*, 2004).

Future Directions for Research and Development

Both student-response systems and new classroom-network technology represent attempts to provide teachers with tools to manage and fine-tune their instruction as it is unfolding in the classrooms. Although these systems are less common than conventional IMS (which teachers use for planning in advance of starting the next unit of instruction) student-response systems and classroom networks represent an important line of research and development. This line of research and development foregrounds the teacher as a thinker and decision maker tasked with solving complex problems on the spot. As our understanding of teachers as instructional decision makers and managers deepens, the knowledge of technology requirements for supporting teachers in this role is likely to increase.

One area of possible future development would extend the current capabilities of both student-response systems and classroom networks. In particular, both might be integrated into classroom management and SIS. Such an integration would provide a richer picture of student achievement and documentation of students' instructional experiences to stakeholders in the educational system. Other possible extensions include the use of handheld technology to support pairing and grouping of individuals who are most likely to be able to learn from one another or to succeed in a particular task. The assignment could be based on data available through the system on general abilities, or it could be based upon data particular to the subject at hand. Finally, more research and development into the kinds of displays and feedback that are readily interpretable by teachers as they work and also after class can extend the utility of systems for evaluating instruction and planning for future lessons.

It is important that research aimed at evaluating the effects of currently available technologies is implemented

in parallel with new technologies. Already, researchers at the Physics Education Research Group at the University of Massachusetts are engaged in a study of the effectiveness of the assess-to-learn model. Their research will examine the effectiveness of teacher professional-development activities designed to prepare instructors to use the approach in their classrooms. In addition, researchers at the Ohio State University are currently engaged in a large-scale randomized control trial to study the impact of using graphing calculators in networked classrooms to teach middle-school mathematics. Many more such studies are needed to examine the professional-development requirements, critical-instructional practices, and outcomes of technologies for managing instruction.

See also: Technology and Formative Assessment.

Bibliography

- Anderson, R. J., Anderson, R., VanDeGrift, T., Wolfman, S. A., and Yasuhara, K. (2003). *Promoting Interaction In Large Classes with a Computer-Mediated Feedback System*. Paper presented at the Computer Supported Cooperative Learning '03, Bergen, Norway.
- Andrews, G., Woodruff, E., MacKinnon, K. A., and Yoon, K. S. (2003). Concept development for kindergarten children through a health simulation. *Journal of Computer Assisted Learning* 19(2), 209–219.
- Archibald, D. A. (2001). Information technology and the goals of standards-based instruction: Advances and continuing challenges. *Education Policy Analysis Archives* 9(48).
- Boyle, J. and Nicol, D. J. (2002). *The Interactive Classroom And Studio (Powerpoint document)*. Glasgow, Scotland: Strathclyde University.
- Burnstein, R. and Lederman, L. M. (2001). Using wireless keypads in lecture classes. *Physics Teacher* 39, 8–11.
- Colella, V. (2000). Participatory simulations: Building collaborative understanding through immersive dynamic modeling. *Journal of the Learning Sciences* 9(4), 471–500.
- Crouch, C. H. and Mazur, E. (2001). Peer instruction: Ten years of experience and results. *The Physics Teacher* 69(9), 970–977.
- Cue, N. (1998). *A Universal Learning Tool for Classrooms?* Paper presented at the First Quality in Teaching and Learning Conference, Hong Kong International Trade and Exhibition Center (HITEC), Hong Kong SAR, China.
- Davis, S. (2003). Observations in classrooms using a network of handheld devices. *Journal of Computer Assisted Learning* 19(3), 298–307.
- DiGiano, C., Yarnall, L., Patton, C., et al. (2003). Conceptual tools for planning for the wireless classroom. *Journal of Computer Assisted Learning* 19(3), 284–297.
- Draper, S. W. and Brown, M. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning* 20(2), 81–94.
- Dufresne, R. J. and Gerace, W. J. (1994). Using “extended scenario” to enhance learning during interactive lectures. Retrieved 3/25/03, 2003 from UMass Physics Education Research Group Website.
- Dufresne, R. J. and Gerace, W. J. (2004). Assessing-to-learn: Formative assessment in physics instruction. *The Physics Teacher* 42(7), 428–433.
- Dufresne, R. J., Gerace, W. J., Leonard, W. J., Mestre, J. P., and Wenk, L. (1996a). Classtalk: A classroom communication system for active learning. *Journal of Computing in Higher Education* 7(2), 3–47.
- Dufresne, R. J., Gerace, W. J., Leonard, W. J., Mestre, J. P., and Wenk, L. (1996b). Using the Classtalk classroom communication system for promoting active learning in large lectures. *Journal of Computing in Higher Education* 7(2), 3–47.
- Duncan, D. (2005). *Clickers in the Classroom: How to Enhance Science Teaching Using Classroom Response Systems*. San Francisco, CA: Pearson.
- Fagen, A. P., Crouch, C. H., and Mazur, E. (2002). Peer instruction: Results from a range of classrooms. *The Physics Teacher* 40(4), 206–207.
- Ganger, A. C. and Jackson, M. (2003). Wireless handheld computers in the preclinical undergraduate curriculum. *Medical Education Online* 8(3), 1–7.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods. *American Journal of Physics* 66, 64–74.
- Hegedus, S. (2003). *Improving Algebraic Thinking through A Connected SimCalc Mathworlds Classroom*. Paper presented at the 27th Conference of the International Group for the Psychology of Mathematics Education held jointly with the 25th Conference of the PME-NA, Honolulu, HI.
- Hegedus, S. and Kaput, J. (2004). *An Introduction to the Profound Potential of Connected Algebra Activities: Issues of Representation, Engagement and Pedagogy*. Paper presented at the 28th Conference of the International Group for the Psychology of Mathematics Education, Bergen, Norway.
- Judson, E. and Sawada, D. (2002). Learning from past and present: Electronic response systems in college lecture halls. *Journal of Computers in Mathematics and Science Teaching* 21(2), 167–181.
- Kaput, J. and Hegedus, S. (2002). *Exploiting Classroom Connectivity by Aggregating Student Constructions to Create New Learning Opportunities*. Paper presented at the 26th Conference of the International Group for the Psychology of Mathematics Education, Norwich, UK.
- Lonsdale, P., Baber, C., and Sharples, M. (2004). *Engaging Learners with Everyday Technology: A Participatory Simulation Using Mobile Phones*. Paper presented at the Mobile Human Computer Interaction 2004: 6th International Symposium, Glasgow, UK.
- Mazur, E. (1997). *Peer Instruction: A User's Manual*. Prentice Hall: Upper Saddle River, NJ.
- Nicol, D. J. and Boyle, J. (2003). Peer instruction versus class-wide discussion in large classes: A comparison of two interaction methods in the wired classroom. *Studies in Higher Education* 28(4), 457–473.
- Penuel, W. R., Roschelle, J., Crawford, V., Shechtman, N., and Abrahamson, A. L. (2004). Workshop report: Advancing research on the transformative potential of interactive pedagogies and classroom networks. Menlo Park, CA: SRI International.
- Piazza, S. (2002). *Peer Instruction Using an Electronic Response System in Large Lecture Classes*. Presentation document presented at the Pennsylvania State University Center for Education Technology Services “Teaching with Technology” series: Departments of Kinesiology Mechanical Engineering, and Orthopaedics and Rehabilitation Center for Locomotion Studies.
- Poulis, C., Massen, C., Robens, E., and Gilbert, M. (1998). Physics learning with audience paced feedback. *American Journal of Physics* 66, 439–441.
- Randolph, J. J. (2007). Meta-analysis of the effects of response cards on student achievement, participation, and intervals of off-task behavior. *Journal of Positive Behavior Interventions* 9(2), 113–128.
- Robinson, S. (2002). *Discourse: Decades of Achievement Results*. Princeton, NJ: ETS Technologies, Incorporated.
- Roschelle, J., Penuel, W. R., and Abrahamson, A. L. (2004). The networked classroom. *Educational Leadership* 61(5), 50–54.
- Roselli, R. J. and Brophy, S. (2002). *Exploring an Electronic Polling System for the Assessment of Student Progress in Two Biomedical Engineering Courses*. Paper presented at the American Society for Engineering Education Annual Conference and Exposition: Montreal, Quebec.
- Sokoloff, D. R. and Thornton, R. K. (1997). Using interactive lecture demonstrations to create an active learning environment. In Redish, E. F. and Rigden, J. S. (eds.) *The Changing Role of Physics Departments in Modern Universities: Proceedings of ICUPU*. College Park, MD: The American Institute of Physics.
- Stroup, W. M., Kaput, J., Ares, N., et al. (2002). *The Nature and Future of Classroom Connectivity: The Dialectics of Mathematics in the*

- Social Space*. Paper presented at the Annual Conference of Psychology of Mathematics Education in North America: Athens, GA.
- Sung, M., Gips, J., Eagle, N., DeVaul, R., and Pentland, S. (2004). MIT.EDU: System architecture for real-world distributed multi-user applications in classroom settings. In Roschelle, J., Chan, T. W., Kinshuk, and Yang, S. (eds.) *Proceedings of the 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education*, pp 43–50. Los Alamitos, CA: IEEE Computer Society.
- Truong, T. M., Griswold, W. G., Ratto, M., and Star, L. (2002). *The Active Class project: Experiments in Encouraging Classroom Participation*. San Diego, CA: University of California, San Diego.
- Webking, R. (1998). *Classtalk in Two Distinctly Different Settings*. El Paso, TX: University of Texas-El Paso.
- Wilensky, U. and Stroup, W. M. (1999). *Participatory Simulations: Network-Based Design for Systems Learning In Classrooms*. Paper presented at the Computer Supported Collaborative Learning, Stanford University.
- Wilensky, U. and Stroup, W. M. (2000). *Networked Gridlock: Students Enacting complex dynamic phenomena with the HubNet Architecture*. Paper presented at the The Fourth Annual International Conference of the Learning Sciences, Ann Arbor, MI.
- Wit, E. (2003). Who wants to be . . . The use of a personal response system in statistics teaching. *MSOR Connections* 3, 5–11.
- Woods, H. A. and Chiu, C. (2003). Wireless response technology in college classrooms. Retrieved June 15, 2005, from <http://www.audience-response-rentals.com/classroom-response/systems-1.htm> (accessed October 2009).

Further Reading

- Penuel, W. R., Roschelle, J., and Abrahamson, L. (2005). Research on classroom networks for whole-class activities. In *Proceedings of the IEEE International Workshop on Wireless and Mobile Technologies in Education*, pp 222–229. Los Alamitos, CA: IEEE.

Relevant Websites

- <http://www-perg.phast.umass.edu> – Physics Education Research Group at the University of Massachusetts-Amherst: Includes resources on the assess-to-learn model and findings from research on student-response systems in physics classrooms.
- <http://galileo.harvard.edu> – Project Galileo: A resource for instructors on peer instruction, with examples of questions instructors can use to test students' conceptual understanding.
- <http://ccl.northwestern.edu> – The Center for Connected Learning and Computer-Based Modeling: A research group investigating the use of classroom networks to explore complex systems in science.
- <http://www.simcalc.umassd.edu> – The SimCalc Project: A research group investigating the use of classroom networks to teach the mathematics of change in middle school.

TECHNOLOGY AND LEARNING – SUPPORTS FOR SKILL LEARNING

Contents

New Media, Learning from Reading and Technology
Technological Supports for Acquiring Twenty-First-Century Skills
Technology Supports for Acquiring Inquiry Skills
Technology Supports for Acquiring Mathematics
Technology Supports for Lifelong Learning

New Media, Learning from

W Schnotz, University of Koblenz-Landau, Landau, Germany

H Horz, University of Applied Sciences Northwestern Switzerland, Olten, Switzerland

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Animation – Visual, usually computer-based display of an object or a more abstract scenario either in the form of a dynamic realistic picture or a dynamic logical picture (graph), which changes its structure during the course of time. Animations can be used to display change, but also to focus the observer's attention to specific details of an object or scenario.

Cognitive flexibility hypothesis – The assumption that criss-crossing through a complex network-like hyper-medium will result in a complex mental network in the mind of the learner that allows higher cognitive flexibility.

Coherence principle – Assumption that the multimedia principle, the contiguity principles and the modality principle apply only if the text and picture information can be integrated into a coherent whole.

Comment information – Information within a sentence or paragraph that conveys information about the topic of the sentence or paragraph.

Comprehensibility approach – A pragmatic approach to text designs, which assumes that four factors are important to enhance text comprehension: Linguistic simplicity, structure and organization, conciseness, and affective or motivational stimuli.

General redundancy principle – Assumption that if a text or a picture is fully intelligible without further information, students learn better from only the text or only the picture than from the text combined with the picture.

Gestalt Laws – Rules of the perceptual organization of stimuli, usually applied on visual perception, as studied by the school of Gestalt Psychology. The gestalt laws prescribe, which elements are grouped together into one gestalt. Examples of Gestalt laws are the law of proximity, the law of similarity, the law of good continuation, the law of conciseness, and the law of common fate.

Hypermedia – An electronic medium, which allows computer-based (and usually web-based) access to specific information via a network-like structure, often referred to as information space.

Linear text – Traditional form of written text, which is expected to be read in a linear fashion, sentence by sentence and paragraph by paragraph, according to the inherent sequence of the text.

Logical pictures (graphs) – Visualizations of relations that cannot be perceived in the real world (such as the increase of birth-rates) through a graphical object such as a line graph, bar chart, etc.

Mental models – A (hypothetical) form of mental representation that consists of internal objects, which hold a structural or functional analogy to the represented external object.

Modality principle – Assumption that students learn better from pictures with auditory text than from pictures with visual text.

Multimedia – In the technical sense: the combination of multiple technical media. In the representational sense: the combination of different forms of representation such as text and pictures. In the modality sense: the use of multiple sensory

modalities for conveying information such as the auditory and the visual modality.

Multimedia principle – Assumption that students learn better from text with pictures than from text alone.

New Media – General term for various kinds of digital, usually computer based electronic media.

Non-linear text – Printed text with a flexible access structure due to graphical layout or electronic hypertext, which is expected to be read in a pre-determined linear fashion (i.e. sentence by sentence) only within paragraphs, whereas the paragraphs are not expected to be read in a specific pre-determined linear sequence.

Propositional representations – A (hypothetical) form of mental representation that consists of complex internal symbols (so-called propositions) which describe the represented content in a kind of mental language.

Realistic pictures – Pictures of objects or scenarios, which resemble the depicted object or scenario.

Semantic memory – Memory system that includes (usually general) world knowledge, based on information from others or on one's experiences.

Spatial contiguity principle – Assumption that learning from text and picture is enhanced, when both sources of information are presented contiguous to each other in space.

Specific redundancy principle – Assumption that students learn better from pictures with (only) auditory text than from pictures with auditory and visual text.

Temporal contiguity principle – Assumption that learning from text and picture is enhanced, when both sources of information are presented contiguous to each other in time.

Topic information – Information within a sentence or paragraph that indicates what the sentence or paragraph is about.

Working memory – Memory system for cognitive processing of information, which is currently in the focus of attention. According to the theoretical model of Baddeley, working memory consists of a central executive and (at least) two slave-systems: the phonological loop and the visuo-spatial sketchpad. Recently, an episodic buffer has been added as a further subsystem.

New media allow flexible combinations of different forms of representation, such as text and static or animated pictures, with different sensory modalities, such as the visual modality when – for example, written text and pictures are to be presented – and the auditory modality when – for example, spoken text and sounds are to be presented. The combination of different forms of representation is commonly referred to as multimedia (Mayer, 2001, 2005). As new media are based on computer (i.e., information processing) technology, they also allow a high degree of interactivity and can provide users with quick access to a multiplicity of information. A set of text paragraphs interconnected via a nonlinear, network-like access structure is usually referred to as a hypertext. When not only text paragraphs, but also pictures or sounds are presented via this kind of nonlinear access structure, a hypertext turns into a hypermedium.

A common assumption in the field of instructional design is that multimedia and hypermedia allow better adaptation of instruction to the needs and preferences of learners. Furthermore, new media are often assumed to increase motivation in learners, thus increasing their investment of cognitive effort and resulting in better learning. The history of media in education reveals, however, that a new medium *per se* has never improved learning (Cuban, 1986). Learning with new media can only be successful to the extent that it takes into account the functioning of the human cognitive system.

Text in Multimedia Learning Environments

Texts – irrespective of whether they are written or spoken – are sequences of verbal information that guide an individual's cognitive processing toward the construction of a coherent knowledge structure with regard to the topic described in the text (Kintsch, 1998). Most cognitive psychologists agree that, when reading or listening to a text, a learner constructs multiple forms of mental representations: a mental representation of the text surface, a propositional representation of the text's semantic content, and a mental model of the content described in the text (Graesser *et al.*, 1997). These mental representations are constructed via an interplay of data-driven bottom-up processes and concept-driven top-down processes.

Whereas reading a written text and listening to a spoken text are not very different with regard to higher cognitive processes – including the construction of a propositional representation and a mental model – they are different with regard to lower-order cognitive processing. Written text usually provides a relatively stable form of information, allowing the reader to go back if a previous phrase has not been understood. Reading a written text enables more self-directed processing on

Promises of Learning with New Media

The term new media is used as a synonym for the use of computer technology to transmit and deliver information.

a micro-level than does listening to a spoken text. When listening to a text, information that has not been processed during presentation cannot be processed later because of the transient nature of the information carrier – acoustic sound waves. In multimedia environments, technical features for going back within an audio file do allow repeated presentation, but selecting specific passages to review is not as easy with audio files as it is in written text.

Linear and Nonlinear Text

In traditional printed text, information is presented in a linear order. In hypertext – on the other hand – relatively short text segments are connected via specific links (so-called hyperlinks) that create a network-like access structure and allow presentation of information in a nonlinear manner. Hyperlinks can help avoid redundancy by linking to the same passage from other different passages, whereas conventional linear text would require repetition of the corresponding passage.

It has often been assumed that hypertext (and – as an extension, hypermedia) would result in more effective and more efficient learning than traditional text because the highly flexible access to information in hypertext allows more freedom in learning. The network-like structure of information presentation was also conjectured to be especially appropriate for learning because it would correspond to the associative, network-like knowledge structures in semantic memory. This so-called cognitive plausibility hypothesis was abandoned because a closer analysis revealed that hypertexts still require sequential reading and cognitive processing and, therefore, cannot directly implement network-like knowledge structures.

In contrast to the simple cognitive plausibility hypothesis, the cognitive flexibility hypothesis assumes that the crisscrossing of learners through the network-like structure of hypertext-information spaces allows the processing of the same information in different contexts depending on the previously read passages and, thus, increases cognitive flexibility. However, the assumption that a network-like information representation would foster knowledge construction in long-term memory has not been confirmed empirically, raising the question of whether hypertext should be considered an information-access device rather than a learning device (Rouet and Levonen, 1996).

Designing Text

Designing written or spoken text for multimedia learning environments is, essentially, complex problem solving that involves the satisfaction of multiple constraints. Text design needs to take into account the instructional objectives, the learners' prior knowledge and cognitive skills, and the limitations of the human cognitive system in syntactic and semantic processing of language. A pragmatic approach to

the design of text has been suggested by Langer *et al.* (1974) in their comprehensibility approach. This approach specifies four key elements of text quality: linguistic simplicity, structure and organization, conciseness, and stimulation.

Linguistic simplicity

Sentences of a text should be relatively short and include simple phrases. If possible, familiar, concrete, vivid words should be used. For texts in multimedia learning environments, linguistic simplicity is especially important for texts in multimedia learning environments because learners read texts on the screen more slowly and get tired more quickly than when reading texts on paper.

Structure and organization

Texts should be well structured internally with regard to their content. They should refer to a global topic, and their internal structure should be made obvious by an appropriate external organization – including headlines, different fonts, shaded text boxes, and so on. Graphical overviews and classification schemata are especially important in fostering learning from hypertext.

Conciseness

Texts should focus on essential information and avoid prolixity or redundant information, without increasing semantic density *too much*. A medium degree of conciseness is considered better than maximum conciseness for comprehension. Hypertexts allow avoiding unnecessary redundancy by interlinking text elements.

Additional stimulation

A text should stimulate readers through the use of vivid formulations and by presenting people and examples from real life. This kind of additional stimulation is assumed to motivate individuals and prolong their process of learning.

Because of the limited capacity of working memory, only a restricted amount of information is available for cognitive processing at any one moment. Accordingly, a learner has to know, at every moment of reading or listening, the specific topic that the text is about in order to activate the corresponding prior knowledge. Whenever the topic changes, the change has to be signaled clearly to the learner *to allow him or her the necessary focus tracking*. The necessary signals for shifting the focus of attention can be transmitted through the topic information within the corresponding sentences or text segments.

In every sentence or text segment, two information components can be differentiated: the topic information and the comment information – the topic information indicates what the text is about; the comment information is the specific information about the topic. Readers or listeners identify the topic information in each sentence (Fletcher, 1984) and try to match it with the previous topic. If the match is positive, the previous topic is still valid.

If the match is negative, a mental search process is triggered in order to find (or create) the new specific topic within the learner's mental representation. The topic information has to contain enough information to allow the learner to identify the new topic, meaning that the various search parameters should allow easy topic identification.

A further guideline for text design is thematic continuity. If a topic has been introduced, it should not be unnecessarily interrupted, because when the topic is picked up again, the reader has to reactivate what he or she has learned so far about the topic, and will usually not reactivate all the previous information. Accordingly, what the individual has in mind during the process of reading will be less rich or less elaborated when topic continuity is low than when topic continuity is high (Schnotz, 1993).

Pictures in Multimedia Learning Environments

Multimedia learning environments use static or dynamic pictures (animation) in conjunction with text as additional sources of information. In contrast to texts, pictures possess structural characteristics in common with what they represent. From a cognitive point of view, pictures can illustrate, organize, explain, and improve memory. However, pictures are, also used sometimes, for primarily decorative purposes. From a purely cognitive point of view, decorative pictures can be considered seductive details because they require some working-memory resources without contributing to deeper understanding of the subject matter (Harp and Mayer, 1998). However, the widespread use of pictures for decorative purposes suggests that these pictures increase the attractiveness of the learning material and therefore pictures may have a motivational function (*cf.* Fisch, 2004).

Kinds of Pictures

Pictures can be subdivided into two main categories: realistic pictures and logical pictures (i.e., graphs). Realistic

pictures are surfaces, which can be looked at to create (nearly) the same perception as looking at the corresponding real object. Realistic pictures reveal (to some degree) similarity with what they represent; they represent height by height, width by width, color by color, etc. The degree of realism can differ. Photographs and realistic drawings show a high degree of realism, whereas simplified drawings and illustrations contain less detail, and schematic pictures present only the most essential elements of the depicted object (see **Figure 1**). Whereas realistic pictures have similarity with the content that they depict, graphs possess only an abstract structural commonality with what they represent. For example, a bar graph can represent nonspatial features such as birth rates, import quotas, and so on by spatial features such as the length of bars. In other words, represented features and representing features may be different from each other. Both realistic pictures and logical pictures can further be subdivided into static and dynamic ones according to their temporal stability.

Dynamic pictures (also referred to as animations) are, often, assumed to result in better learning – when the content being portrayed is itself dynamic. A meta-analysis by Höffler and Leutner (2007) found an average superiority of animations in terms of learning compared to static displays, especially if human movements were involved. However, whether and under which conditions animations will lead to better learning outcomes is still under debate (Lowe and Schnotz, 2008) because, on the one hand, animations provide more information (i.e., information about processes), but on the other they, animations provide information that is transient in nature. If learners have to compare different states within an animation, they have to hold previously seen pictures in working memory over a longer period, which can lead to cognitive overload of the working memory. Although learning with movies and animations is quite popular, these media are also, from a motivational point of view, not necessarily the best choice for learning, because viewing movies is usually considered easy and hence may trigger less learner effort than does reading a text, which is considered more difficult (Salomon, 1984). There is a risk that learners do not elaborate mentally the content of animated pictures with

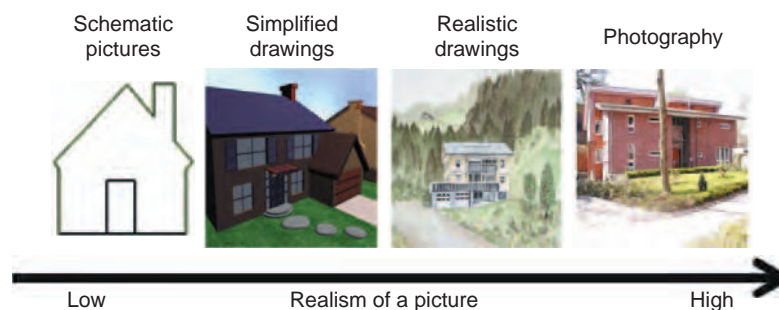


Figure 1 Degrees of realism of a picture.

the same effort that they elaborate the mental model they derive when reading a text.

Understanding Pictures

Pictures provide a relatively direct way to construct mental models, because the external and the internal representation are both analogous in nature. The process of understanding a picture includes both visual perception and higher-order cognitive processing. The basic processes of visual perception are largely pre-attentive. These processes are primarily bottom-up, include automated visual routines, and run in parallel. In order to not only perceive but also understand a picture, higher-order cognitive processes, which are attentive in nature, are required. The learner must read features of the perceptual representation and interpret them semantically in order to understand what the picture represents. These semantic processes are run serially and include both data-driven, bottom-up and concept-driven, top-down components. Accordingly, they are strongly affected by both the learner's prior knowledge and his/her goal orientation.

In contrast to reading text, which suggests a specific linear sequence of cognitive processing, reading pictures provides high degrees of freedom with respect to the sequencing of pictorial information during cognitive processing. However, there are different ways to influence the learner's processing order. For example, certain parts of a picture can be highlighted by increasing their visual contrast to the background or by labeling. Directive signs – such as arrows, frames, or drawn magnifiers – can be used to emphasize certain areas within the picture. Furthermore, processing order can be influenced by numbering pictorial elements. Finally, there are cultural habits – such as reading from the left to the right in the Western world – that can also influence the processing of pictures.

A high degree of realism is not necessarily beneficial for picture comprehension. Instead, the degree of realism should be adapted to the level of the learners' prior knowledge. Dwyer (1978) demonstrated that novices learn better from schematic pictures with limited detail rather than from highly realistic pictures.

Understanding graphs differs from understanding realistic pictures because graphs do not have perceptual similarity to what they represent, which implies that the learner cannot use perceptual schemata based on everyday experience. Sub-semantic pre-attentive processing includes discrimination, identification, and grouping of graphical components, such as points, lines, and surfaces, according to the so-called Gestalt laws. This kind of processing results in the perception of specific graphical configurations. Semantic interpretation of these configurations requires that the individual knows the underlying representational conventions – that is, has acquired a corresponding graphic schema (Pinker, 1990), which allows him or her to map the

graphical configuration onto a corresponding mental model of the learning content. Understanding graphs requires cultural skills, which have to be learned. Empirical findings suggest that experts and learners with higher domain-specific prior knowledge are better at detecting global visual patterns and create more elaborate and more hierarchically structured knowledge structures from these patterns, whereas learners with lower prior knowledge are more bound to local search processes.

Designing Pictures

Pictures should be designed in such a way that their elements can be easily detected and identified. Essentially, this means that the design should take the so-called Gestalt laws into account. According to the Law of Proximity, spatially adjacent elements are more likely to be associated with each other than are distant elements. According to the Law of Similarity, elements with a similar appearance (in terms of form, color, orientation, texture, etc.) are more likely to be associated than are nonsimilar elements. According to the Law of Closure (or the Law of 'Good Gestalt'), visual information is organized as simply as possible, resulting in simple, concise forms. According to the Law of Good Continuation, two crossing lines will be kept separated by the implicit assumption that the crossing lines proceed continuously at the crossing point. Finally, the Law of Common Fate assumes that different lines with the same course are grouped into one unit.

Graphs should be designed in a way that supports the message to be conveyed as clearly as possible. This requires consideration of syntactic, semantic, and pragmatic constraints. Syntactic constraints refer to the relation between pictorial elements. For example, elements that form a set should be grouped together by using the same color, the same form, the same texture, or by using a common frame. Semantic constraints refer to the relationship between the graph and the represented subject matter. Different visual attributes differ in their usefulness for different representational purposes. For example, quantities are usually represented by the size of graphical elements (such as the lengths of bars), whereas different qualities are represented by different colors, textures, or graphical forms. Pragmatic constraints refer to the specific communicative purpose of the graph. Learners should not be tempted to draw incorrect inferences because of interrupted scales, exaggerated size differences stemming from simultaneous variation of multiple attributes (such as length, width, and depth), or due to the use of multiple axes in a graph that suggests a comparison of trends when no such comparisons can be made. Generally, the number of visual attributes that are used to convey information should not be higher than the number of represented attributes.

Learners often assume that one short look is enough to understand a graph. Graphs and diagrams should, therefore, be designed in a way that counteracts this tendency toward superficial processing. Relevant elements can be emphasized by frames, by the magnifying-glass technique, or by directive signs such as arrows, which guide the learner's attention toward key elements. For specific communicative purposes, specific forms of graphs, have been developed, such as pie graphs, bar graphs, line graphs, and scatter plots. As these forms of graphs are frequently used, readers are more likely to possess the corresponding graphic schemata required for understanding these graphs than the schemata needed to comprehend other forms of visualization (Pinker, 1990).

Learning with Multimedia

Theoretical Models of Learning from Multimedia

Numerous studies have found that students learn better from text combined with pictures than from text alone. Whereas early research focused on the mnemonic function of pictures combined with narrative text, recent studies also deal with the explanatory function of pictures. Mayer and his colleagues have found that students understand technical devices or natural phenomena better when they learn from text and pictures. This so-called multimedia effect seems to be especially strong when learners have low domain-specific knowledge (Mayer, 2001; Mayer and Moreno, 1998). In other cases, however, adding pictures to a text has been shown to have detrimental effects on learning (Sweller *et al.*, 1998). So, the question arises as to when multimedia learning is more successful than learning from a single medium, and why it is more successful under those circumstances.

Dual coding theory

The advantage of combining texts with pictures was often explained through Paivio's dual coding theory (DCT). According to this theory, the human cognitive system entails two separate – but related – subsystems: a verbal and an imagery system. Both systems can interact, but can also be activated independently, and both have a limited capacity. Usually, verbal information is processed only in the verbal subsystem, whereas pictorial information is processed in both systems. As the two subsystems are interconnected, cross-referential connections between the representation of a subject matter in the verbal system and a representation of the same subject matter in the imagery system can be established. The overall memory representation is, therefore, more elaborated, which is assumed to be the basis for the positive text–pictures effects.

Cognitive theory of multimedia learning

Building on the DCT, Mayer (2001) developed a cognitive theory of multimedia learning (CTML). This theory

assumes that the human cognitive system includes verbal and pictorial (image) subsystems. Accordingly, individuals can use different representational formats to internally encode and store knowledge. Furthermore, the theory draws on the ideas of Baddeley (1986) about human working memory. Baddeley hypothesizes that humans' working memory is characterized by a rather limited capacity and high decay rates. It includes a central executive and two slave systems: the phonological loop and the visual sketchpad. Similarly, Mayer assumes that there are two sensory subsystems in working memory: a phonological system and a visual system. His first basic assumption with regard to multimedia learning merges the notion of dual coding and the notion of two sensory subsystems. Humans are supposed to process information in their working memory through two channels: an auditory–verbal channel and a visual–pictorial channel. Mayer's second basic assumption is that these two channels have a limited capacity to convey and process information. His third basic assumption is that humans are active sense-makers: They engage in active cognitive processing to construct coherent knowledge structures from both the available external information and their prior knowledge.

According to Mayer's theory, verbal selection processes synthesize a so-called propositional text base, and verbal organization processes lead to a text-based mental model. Similarly, pictorial selection processes synthesize a picture base, and pictorial organization processes lead to a picture-based mental model. Integration processes constitute referential relations between the text-based and picture-based mental models (see **Figure 2**).

Integrated model of text and picture comprehension

The parallelism of the text and picture processing assumed in Mayer's model can be questioned on the grounds that texts and pictures use completely different representation principles. Schnotz (2005) has developed an integrated model of text and picture comprehension (IMTPC), which is depicted schematically in **Figure 3**. This model incorporates the following assumptions: Text and picture comprehension take place in a cognitive architecture that includes a working memory of limited capacity, modality-specific sensory registers as information-input systems, and a long-term memory. Verbal information (i.e., information from written or spoken texts) and pictorial information (i.e., information from visual pictures and from nonverbal auditory sounds) are transmitted to working memory through the visual channel and the auditory channel, respectively. Further information processing in working memory takes place in two representational channels: the verbal channel and the pictorial channel. Information from written or spoken text is processed in the verbal channel. Information from visual pictures or from sounds is processed in the pictorial channel. The channels have

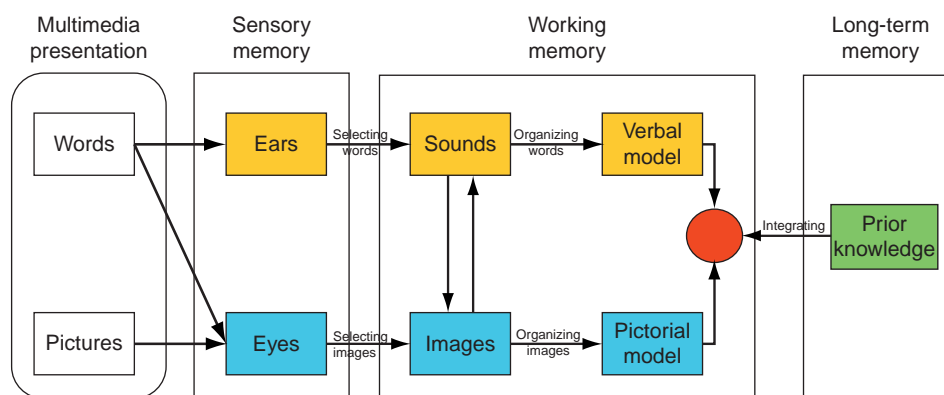


Figure 2 Cognitive Theory of Multimedia learning. From Mayer, R. E. (ed.) (2005). *The Cambridge Handbook of Multimedia Learning*. Cambridge: Cambridge University Press. S. 37.

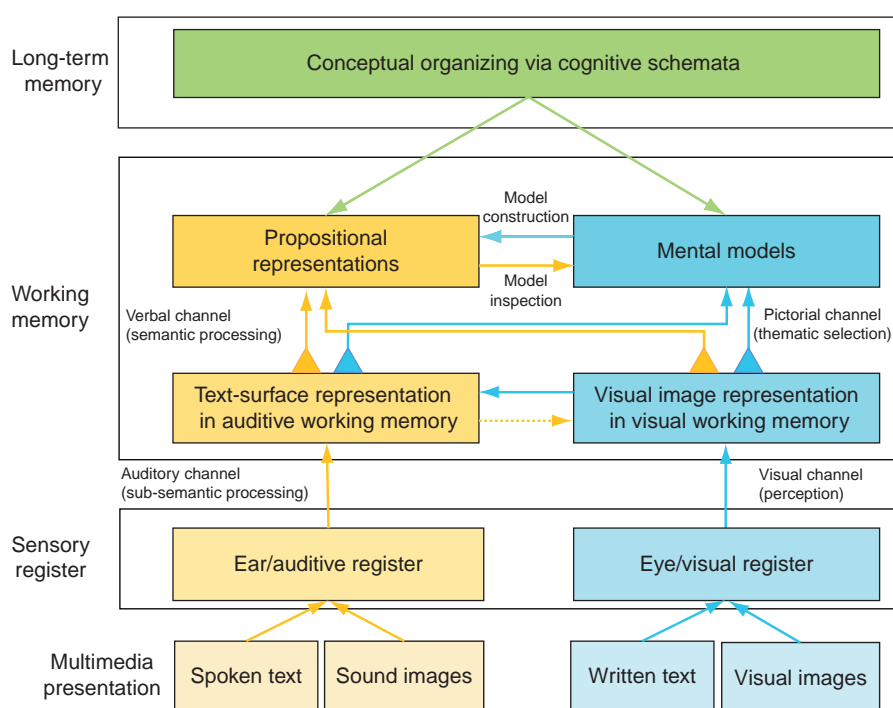


Figure 3 Integrative model of text and picture comprehension. From Schnotz, W. (2005). An integrated model of text and picture comprehension. In Mayer, R. E. (ed.) *Cambridge Handbook of Multimedia Learning*, pp. 49–69. Cambridge: Cambridge University Press.

limited capacity to process and store information. Text and picture comprehension are active processes of coherence formation. In comprehension, individuals engage in building coherent knowledge structures from the available external verbal and pictorial information, and from their prior knowledge.

The CTML and the IMTPC differ in terms of structure as well as in terms of some predictions. With regard to structure, the CTML merges sensory aspects and representational aspects by assuming an auditory–verbal channel and a visual–pictorial channel. The IMTPC, on the contrary, keeps sensory aspects and representational

aspects distinct: the sensory channels are assumed to feed information into working memory, whereas the representational channels (verbal and pictorial) are supposed to process information within working memory. Whereas the CTML assumes the construction of two mental models (a verbal and a pictorial one) in comprehension and learning, the IMTPC assumes that only one mental model is constructed. Whereas the CTML predicts that students generally learn better from text and pictures than from text alone, the IMTPC predicts that adding pictures is not always beneficial, and can also be harmful for learning: First, learners may use the pictures instead of

the text and, therefore, process the text only superficially, resulting in lower learning than when using text alone (Schnotz and Bannert, 2003). Second, learners with sufficient prior knowledge may be able to construct a mental model without pictorial support; they would only waste time and effort when learning from text and pictures rather than from text alone (Sweller *et al.*, 1998). Third, because processing of information in the pictorial channel is assumed to include procedures of structure mapping between picture and mental model, a picture with a non-appropriate structure for the task at hand may interfere with the construction of an appropriately structured mental model, resulting in lower learning performance (Schnotz and Bannert, 2003).

Research Findings

The two models agree that the main reason for the learning-enhancing function of multimedia compared to single media is that multimedia can support integrative processing of verbal and pictorial information in working memory. Integrative processing requires both verbal and pictorial information to be held simultaneously in working memory. Because the capacity of working memory is limited, and decay rates are high, students learn better from words and pictures than from words alone when verbal and pictorial information are simultaneously available in working memory. From this fundamental condition of integrative processing, various principles for the design of multimedia can be derived.

Coherence and contiguity

Students learn better from words and pictures than from words alone if the words and pictures are semantically related to each other (the coherence condition) and if they are presented closely together in space or in time (the contiguity condition). Text and pictures can only contribute to a joint mental model construction if corresponding text information and picture information are simultaneously available in working memory. Because information decays quickly from working memory, this requires the combined presentation of words and pictures as far as possible, then fulfilling the contiguity condition. If a picture is combined with a spoken text, simultaneous availability of related pictorial and verbal information in working memory is enhanced by the simultaneous presentation of the picture and the semantically related part of the text (i.e., temporal contiguity). If a picture is combined with written text, all information has to enter working memory through the visual channel. Since only one kind of information can be processed through the visual channel at a time, the eye has to switch between pictures and words. This so-called split of visual attention implies that unproductive visual search processes from the picture to

the text, and vice versa, have to take place. If pictures and related written words are presented closely to each other (i.e., spatial contiguity), visual search processes are reduced. In short, high spatial contiguity of pictures and written words enhances the simultaneous availability of pictorial and verbal information in working memory.

Modality

Various studies have shown that students learn better when pictures are presented with spoken text instead of written text (Mayer, 2001; Mayer and Moreno, 1998). If a picture is combined with written text, information processing has to be highly selective: The eye has to focus either on the picture or on the text and, thus, information from only one source can be conveyed at a specific moment. Because all visual information has to be processed through the visual channel, which has a limited capacity to convey information, the overall information input into working memory within a limited amount of time is lower than in the case where both the visual and the auditory channel are being used. Thus, split of attention, which is essentially the use of one information channel for different sources of information, results in a reduced amount of information that is put into working memory within a specific amount of time. The negative effects of split attention on learning are especially pronounced when animated pictures are used instead of static pictures because of the fleeting nature of animations (Mayer, 2001). The modality effect results from differences in the amount of information that is processed through sensory channels as well as from differences in the amount of working memory capacity that is involved in further cognitive processing (Mayer and Moreno, 1998).

Redundancy

Contrary to the DCT and the CTML, which assume that adding pictures to texts always leads to better learning because two codes in memory are better than one, the integrated model predicts that the combination of texts and pictures can also have detrimental effects under specific conditions. One of the detrimental effects is called the redundancy effect (Sweller *et al.*, 1998). If learners have high prior knowledge, they frequently do not need text as well as pictures as information sources because one source provides all the required information. In this case, adding a picture to a written text or adding a written text to a picture means adding unneeded information. Although only one of the two information sources is needed, the eye wanders between both of them resulting in a split of attention. Thus, the learner loses time and expends mental effort searching for redundant information. The integrated model predicts that learners with higher expertise will perform better with only one information source (i.e., text or picture) while those with lower expertise will perform better with

two information sources (i.e., text and pictures). This prediction is in line with various research findings (cf. Mayer, 2001; Sweller *et al.*, 1998) and is also referred to as the expertise reversal effect (Kalyuga *et al.*, 2000).

Designing Multimedia

From the empirical findings presented above, the following principles for designing multimedia learning environments can be derived:

- *Multimedia principle.* Use text combined with content-related pictures when learners have low prior knowledge, but sufficient cognitive abilities to process both the text and the pictures.
- *Spatial contiguity principle.* If written text is used, present it in close spatial proximity to the picture.
- *Temporal contiguity principle.* If spoken text is used, present it in close temporal proximity to the picture.
- *Modality principle.* If animation is used, use spoken text instead of written text.
- *Specific redundancy principle.* Do not add written text that duplicates spoken text combined with pictures.
- *General redundancy principle.* Do not combine text and pictures if learners have sufficient prior knowledge and cognitive ability to construct a mental model also from one source of information.
- *Coherence principle.* Do not use extraneous words and pictures. Do not add unnecessary sound or music.

Whereas Universal Design for Learning (Rose and Meyer, 2006) generally suggests presenting, multiple representation to give learners various ways of acquiring information and knowledge, empirical research on multimedia learning suggests that this may be a two-edged sword, because too many sources of information presented at the same time can overwhelm the learner's cognitive capacity. In any case, designers of instructional material should resist the temptation to add irrelevant bells and whistles to multimedia learning environments. Simply speaking, less can be more.

Outlook

Understanding multiple external representations – such as text, static or animated pictures, and graphs – requires processing the different sources of information in an integrative manner. Combined usage of text, pictures, and graphs in multimedia learning environments allows better learning for individuals with low domain-specific prior knowledge – those who are not yet able to construct multiple mental representations from a single external source of information. However, the combined usage of text, pictures, and graphs requires skills of mapping information from one external representation to another external representation. These skills include not only simple

verbal literacy in terms of reading written text, but also skill in reading information from pictures and graphs, which can be referred to as visual or pictorial literacy.

Successful teaching and learning with multiple external representations requires more than following simple rules of thumb in a mechanical way. Rather, it requires scaffolds for orientation based on empirical research, and simultaneous and systematic consideration of multiple constraints. In this way, the design of multimedia learning environments can go beyond pure intuition and make instructional materials look good, through the manipulation of surface variables. Instead, the design of multimedia learning environments should take the functioning of the human cognitive system into account.

See also: Learning from Multiple Information Sources; Technology Supports for Lifelong Learning.

Bibliography

- Baddeley, A. D. (1986). *Working Memory*. Oxford, UK: Clarendon Press.
- Cuban, L. (1986). *Teachers and Machines: The Classroom Use of Technology Since 1920*. New York: Teachers College Press.
- Dwyer, F. M. (1978). *Strategies for Improving Visual Learning*. Pennsylvania, PA: Learning Services.
- Fisch, S. M. (2004). *Children's Learning from Educational Television: Sesame Street and Beyond*. Mahwah, NJ: Erlbaum.
- Fletcher, C. (1984). Markedness and topic continuity in discourse processing. *Journal of Verbal Learning and Verbal Behavior* **23**, 487–493.
- Graesser, A. C., Millis, K. K., and Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology* **48**, 163–189.
- Harp, S. F. and Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology* **90**, 414–434.
- Höfler, T. N. and Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction* **17**, 722–738.
- Kalyuga, S., Chandler, P., and Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology* **92**, 126–136.
- Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. New York: Cambridge University Press.
- Langer, I., Schulz von Thun, F., and Tausch, R. (1974). *Verständlichkeit in Schule, Verwaltung, Politik, Wissenschaft. (Comprehensibility in Schools, Administration, Politics, Science)*. München: Reinhardt.
- Lowe, R. K. and Schnotz, W. (eds.) (2008). *Learning with Animation. Research Implications for Design*. New York: Cambridge University Press.
- Mayer, R. E. (2001). *Multimedia Learning*. New York: Cambridge University Press.
- Mayer, R. E. (ed.) (2005). *The Cambridge Handbook of Multimedia Learning*. Cambridge: Cambridge University Press.
- Mayer, R. E. and Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology* **90**, 312–320.
- Pinker, S. (1990). A theory of graph comprehension. In Freedle, R. (ed.) *Artificial Intelligence and the Future of Testing*, pp 73–126. Hillsdale, NJ: Erlbaum.
- Rose, D. H. and Meyer, A. (2006). *A Practical Eader in Universal Design for Learning*. Cambridge, MA: Harvard Educational Publishing Group.
- Rouet, J. -F. and Levonen, J. J. (1996). Studying and learning with hypertext: Empirical studies and their implications. In Rouet, J. -F., Levonen, J. J., Dillon, A., and Spiro, R. J. (eds.) *Hypertext and Cognition*, pp 9–23. Mahwah, NJ: Erlbaum.

- Salomon, G. (1984). Television is “easy” and print is “tough”. *Journal of Educational Psychology* **76**, 647–658.
- Schnotz, W. (1993). Adaptive construction of mental representations in understanding expository texts. *Contemporary Educational Psychology* **18**, 114–120.
- Schnotz, W. (2005). An integrated model of text and picture comprehension. In Mayer, R. E. (ed.) *Cambridge Handbook of Multimedia Learning*, pp 49–69. Cambridge: Cambridge University Press.
- Schnotz, W. and Bannert, M. (2003). Construction and interference in learning from multiple representations. *Learning and Instruction* **13**, 141–156.
- Sweller, J., van Merriënboer, J. J. G., and Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychological Review* **10**, 251–296.

Further Reading

- Rouet, J. F., Lowe, R., and Schnotz, W. (eds.) (2008). *Understanding Multimedia Documents*. New York: Springer.

Reading and Technology

N DeFrance, Grand Valley State University, Grand Rapids, MI, USA

D Khasnabis and A S Palincsar, University of Michigan, Ann Arbor, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Decoding – Methods used to decipher the printed work.

Digitized speech – Artificial speech compiled from recordings of human speech.

Hypertext – Electronic text that is linked to other chunks of text.

Multimedia – The integration of multiple forms of media, such as graphs, illustrations, photographs.

Phonemic awareness – The ability to isolate and attend to individual sounds.

Reading in Light of Technology

In addition to providing alternative ways of supporting the acquisition and development of traditional print-literacy skills, technology changes the nature of text and reading. Indeed, technology has expanded the notion of text to include a broad array of media objects, including: digitized multimedia (e.g., video, sound, graphics, text, and photographs) that are presented in a linear structure, and hypertext and hypermedia, which are constructed in parallel or hyper-based structures.

While the literacy community rejects the notion that any texts, no matter what their structure, ever have predetermined meaning since the reader is always making connections – reading in and reading between the lines (Purves, 1999) – it is certainly the case that hypertexts make the role of the reader more salient in co-authoring the text, since the reader determines the paths taken and ignored, and can reorder, change, augment, and delete paths that may have initially been forged or proposed by the creator of the hypertext. This characteristic offers new challenges and opportunities to readers, and changes the vista of reading research. However, research on these aspects of reading and technology is emergent; hence, in this article, we focus principally on research that has been conducted regarding the use of technology to support the development of print-literacy knowledge and skills, and to provide compensatory and remedial support for students who require additional assistance to learn to read and to use reading to advance their learning.

We have organized the article around areas of reading instruction deemed crucial by the National Research Council Committee on Preventing Reading Difficulty in Young Children (Snow *et al.*, 1998) and the Report of the National Reading Panel on Teaching Children to Read (National Institute of Child Health and Human Development, 2000). Those key areas are: alphabets (phonemic awareness and phonics), fluency, and comprehension (vocabulary and comprehension strategy) instruction. There is also a discussion of some applications of technology to encourage early reading and writing-related behaviors, or emergent literacy skills, which lay a foundation for conventional literacy. In the final section, we identify research efforts that are underway to exploit the additional affordances of new technologies.

Emergent Literacy

Very young children engage in reading and writing-related activities during cooperative and dramatic play in environments rich in print materials. They discover the symbols – numbers, letters, and icons – used to think about and express ideas; they use symbols to create their own messages in conventional and unconventional ways (Roskos and Neuman, 2001). Computers can support preschoolers' emergent literacy by contributing to dramatic play, offering practice in conventional literacy skills, and enabling children to explore their world and communicate their discoveries, (McGee and Richgels, 2006).

Preschoolers can explore their world via the World Wide Web. Computer use prompts children to collaborate and share their literacy knowledge (Labbo and Ash, 1998). They can communicate messages of songs, celebrations, classroom mascots, seasonal artwork, fall leaves of local trees, information about their pets, and photos of their communities with peers around the world. With computers, teachers and students can easily produce literacy props such as menus, tickets, maps, signs, and shopping catalogs for dramatic play. Inclusion of these props in play centers increases literacy-related play and the opportunity to develop literacy skills (Labbo and Ash, 1998). Alphabet-game software provides practice in the conventional literacy skill of letter-sound association. CD-ROM talking books with animation and sound effects support literacy development by improving children's understanding of stories, sustaining their attention with minimal adult scaffolding, and developing word-reading skills

(but only in those with some grounding in sound–symbol correspondence). Children at risk of school failure have been shown to produce more complex retellings and learn more new words and sentence structures from multimedia storybooks than from storybooks with static pictures (Bus *et al.*, 2006).

The utility of computers in preschool classrooms is constrained by technical limitations: the average preschool classroom computer is 10 years old, limiting the functions it can perform. Software is not yet capable of listening to children to take their dictation or provide corrective feedback on verbal responses (McGee and Richgels, 2006). In addition, there are attitudinal limitations: preschool teachers may not recognize the importance of information and communication technologies (ICT), may be reluctant or inadequately supported to integrate technology meaningfully into the curriculum, or may devalue computer competencies that children bring from home (Turbill and Murray, 2006).

Phonemic Awareness and Phonics

Mastery of phonics and phonemic awareness is prerequisite to developing the fluent word recognition of skilled reading (National Institute of Child Health and Human Development, 2000). Phonemic awareness is the ability to isolate and attend to the individual sounds, or phonemes, of spoken words. Phonemic awareness is necessary to develop an understanding of phonics, the system of letter–sound relationships needed to put spoken words into writing. Direct instruction in phonics and phonemic awareness includes tasks such as: matching sounds to letters, counting sounds in words, identifying beginning, middle, and ending sounds in words, rhyming, segmenting (cat → c-a-t), blending (b-a-t → bat), and manipulating (stop without s' → top) sounds to read and spell real words and pseudowords.

An abundance of commercially available beginning-reading software uses colorful, animated graphics to deliver potentially tedious drills in an entertaining format so that students will expend sufficient time and attention on phonemic awareness and phonics tasks (Burns *et al.*, 1999). Ideally, this software should reflect best practices in phonics and phonemic teaching such as progressing from blending individual phonemes to blending onsets and rimes, and from applying skills with single syllables to using multisyllable words (McKenna, 2002). Software can gate these progressions, ensuring that students demonstrate mastery of each skill in a carefully structured sequence. Software can provide spoken models and corrective feedback. Thus some tasks formerly performed by a teacher may be delegated to a computer, allowing students to work at their own pace, independently.

While generally effective and fun, computer-assisted phonics instruction may not be necessary or cost effective for typically achieving readers who often make comparable

gains in response to systematic phonics and phonemic-awareness instruction with and without computer support (Blok *et al.*, 2002; Dynarski, *et al.*, 2007; Macaruso *et al.*, 2006; Olson and Wise, 2006). There is some evidence that speakers of English may benefit more from computer-assisted instruction in beginning-reading skills than speakers of languages (e.g., Dutch) in which letters and sounds have more predictable relationships than they do in English (Bus *et al.*, 2006).

Direct, computer-assisted instruction in phonics and phonemic awareness has proven to be more effective than such instruction without supplementary computer support, for certain readers. Computer-assisted instruction is particularly effective in improving phoneme awareness, and phonologic decoding, in at-risk and poor readers, who need intensive, extended practice (Barker and Torgesen, 1995; Blok *et al.*, 2002; Macaruso *et al.*, 2006; Mitchell and Fox, 2001; Olson and Wise, 2006; Wise *et al.*, 1999). In some studies, practice in phonemic-awareness tasks also led to improved word identification (MacArthur *et al.*, 2001). Struggling readers, in the second grade and younger, benefit more than older struggling readers, who may improve in phonics and phonemic awareness without significant improvement in word recognition (Olson and Wise, 2006). Software programs that target phonemic awareness, but do not engage the participant in reading words, have not proven effective in improving word reading. Examples of such programs are those that target auditory temporal and speech processing, or perception of the order of a sequence of rapidly presented speech sounds or tones, and articulatory awareness, or attention to the movements of the mouth associated with each speech sound (Olson and Wise, 2006).

Fluency

Once readers have sufficient phonic and phonemic awareness to decode some words, they must practice their decoding skills through wide reading to build a repertoire of automatically recognized words (Kuhn and Stahl, 2006). Fluent reading is characterized by quick, accurate, and automatic word recognition and meaning-laden prosody (rate, phrasing, pitch, and inflection of speech) that facilitate comprehension. An example of computer-assisted support for fluent reading is talking books that offer (1) digitized speech with synchronized highlighting allowing the reader to follow along as the computer reads aloud and (2) prompts for word identification, either phonological word analysis (part-word prompts) or speech feedback (whole-word prompts) that the reader can request when confronted with difficult words. Talking books offer guided, independent reading opportunities that allow the reader to reduce the amount of computer-provided support as reading skill improves (McKenna, 1998).

For struggling readers, digitized speech provides opportunities for repeated readings with an expert, a practice

shown to improve reading fluency for print texts (National Institute of Child Health and Human Development, 2000). Again, a time-intensive task typically guided by an adult is shifted to the computer, allowing readers who need intensive practice to do so independently. Struggling readers who are older or adverse to monitoring and correction by others may find the computer to be a more amenable practice partner (Anderson and Speck, 2001).

There is evidence that extended exposure to computer-provided narration, accompanied by highlighted text, focuses readers on print and improves word reading, though benefits accrue only to those who have some beginning-reading skills. Struggling readers of any age or grade usually need some concepts of print, knowledge of letter-sound relationships, a basic level of phonemic awareness, and at least a small repertoire of sight words before digitized speech accompanying text will produce gains in word recognition and, sometimes indirectly, phonemic awareness (Bus *et al.*, 2006; Das-Smaal *et al.*, 1996; Heimann *et al.*, 1995; Higgins and Raskind, 2000; Olson and Wise, 2006; Raskind and Higgins, 1999).

Audio prompts for phonological analysis of difficult words include supplying: phonemes one at a time until the reader has enough of the word to identify it, breaking words into syllables or onset and rime (ring → r-ing), and phonic analogies (if h-at is hat, then fl-at must be ...). These prompts may reinforce phonologic decoding skills, but they also draw the reader away from the text enough to disrupt comprehension (McKenna, 1998).

Readers can also request speech feedback, that is, to have selected, difficult words read aloud. In general, reading with speech feedback has a positive impact on speed and accuracy of oral reading for struggling, school-aged readers (Dawson *et al.*, 2000; Moseley and Hartas, 1993; Van Daal and Reitsma, 1993; Van Daal and van der Leij, 1992). Simply supplying difficult words by means of digitized speech, rather than eliciting through hints or analogies, may be a better form of speech feedback for struggling readers since it does not require reader prerequisites (e.g., sight words or knowledge of print conventions) (McKenna, 1998). It does, however, require that the reader self-monitor and adopt strategies for accessing speech feedback. Some struggling readers require considerable training in monitoring their own reading before they request speech feedback when their reading is in error (McKenna, 1998).

Several promising technologies are currently in development. Speech-recognition software shows the potential to facilitate fluent reading and, thus, comprehension, without either the reader asking for assistance or an adult monitoring. A speech-recognition system can listen to a student read aloud and supply troublesome words when a student struggles (Adams, 2006). Text-to-speech-translation software reads aloud any text that can be loaded or scanned into the computer. This software offers

readers access to an array of reading materials, including content from the World Wide Web, thus maintaining reader interest when it might begin to flag in the face of frustration.

Vocabulary

When beginning readers decode, they translate words from an unfamiliar form in print to a familiar form in speech. Soon, though, readers encounter words whose printed and spoken forms are both unfamiliar. To ensure comprehension of text, reading instruction often includes vocabulary instruction, including computer-assisted instruction, to develop word meanings (National Reading Panel, 2000). Research shows that computer-assisted instruction can be more effective than traditional methods of direct vocabulary instruction (NRP, 2000). Productive applications of technology are those which address key components of vocabulary instruction including: multiple, varied encounters with target words, opportunities to use target words in novel contexts, and information about how knowledge of a word can be developed (Nagy and Scott, 2000).

Electronic texts can offer multiple, varied encounters to develop knowledge of important words through hypertext links – to glossary, simpler substitutions, examples of the word used in a sentence, additional background information, and explanations – and multimedia illustrations and graphics enhanced with animation, interactive features, narration, and sound effects. Readers learn words best when hypertext-linked information and animation features are not simply called up, but mediated, that is, extended through actions such as questioning, clarifying, or making inferences (Blachowicz and Beyersdorfer, 2006). Mediation opportunities, typically found in discussion with adults or peers, can be built into electronic texts as prompts to apply word knowledge, thus making the reader more independent. Multimedia CD-ROM storybooks have been used successfully to enhance word awareness and word knowledge in second-language learners, readers who have a particular need to acquire many new words in a short period of time. It is important that these animations be congruent with the text (Bus *et al.*, 2006).

Comprehension and Learning from Text

Comprehension is broadly conceived as an active process in which readers simultaneously extract and construct meaning as they interact with written language (Rand Reading Study Group, 2002). Hence, in addition to quick, accurate word recognition and broad, deep word knowledge, readers must also bring, or seek, additional resources to support making sense of text. This activity entails the use of cognitive processes of selecting, organizing, connecting, and evaluating what is being read. Strategies include making predictions while reading, drawing inferences,

self-questioning, paraphrasing, and using text features, including the structure of text, to support meaning making.

There are at least three ways in which technology is thought to have the potential to support text comprehension. The first is through supporting decoding so that readers have ready access to the ideas in text. The second is by offering multiple representations of information (e.g., prose, graphics, and animations), and the third is by using technology to teach students to be more strategic in their reading of text.

The use of text-to-speech (TTS) and digitized speech technologies is thought to facilitate comprehension by bypassing decoding problems so that students can focus on constructing meaning from the text. Indeed, many studies have shown that reading with speech feedback positively affects comprehension (Elbro *et al.*, 1996; Elkind *et al.*, 1993; Lundberg and Oloffson, 1993; Montali and Lewandowski, 1996). However, only one study (Elbro *et al.*, 1996) has found improvement that transferred to reading traditional text without TTS support. This particular study provided grade 2–6 Scandinavian students who had reading and language disabilities with uniquely interactive TTS support for 20 min a day for 40 days. When students clicked on a word, the word was presented broken into visual and auditory segments, and students were required to generate their own pronunciation before hearing the whole word. The interactive nature of the TTS support and the extended time for which students were offered TTS support may have enhanced the effectiveness of the treatment. Another factor that may affect the effectiveness of TTS support is the learner's age. Lundberg and Oloffson (1993) suggest that older students' comprehension may benefit more from TTS support because they read more difficult texts with a higher percentage of complex and multisyllabic words. Another possible explanation is that older students may be able to make better decisions about targeting words for TTS support. It should be noted that several studies have reported that TTS does not affect comprehension (Farmer *et al.*, 1992; Leong, 1995; Wise and Olson, 1995; Wise *et al.*, 1998, 2000). Possible explanations for this lack of influence are that students with learning disabilities do not access supports, even when warranted, and that the effect of TTS may vary by reading ability and type of reading difficulty. It is also possible that the effect of TTS may be mediated by the larger instructional context.

While TTS allows for basic access to text, struggling readers often face additional problems such as a lack of background knowledge or a lack of strategic reading skills that prevent them from learning from text. A few studies have developed hypertexts including various types of representation supports, such as an online glossary, background material, or a summary of main ideas, that are designed to help readers understand content (Boone and Higgins, 1993; Higgins *et al.*, 1996; Leong, 1995; MacArthur and Haynes,

1995; Moore-Hart, 1995). With the exception of Leong's (1995) study, which found no effects for TTS, for TTS with explanations of difficult words, for TTS with explanations and with reading awareness prompts, or for TTS with simplified passages, these studies tend to demonstrate that the hypertexts improve comprehension for struggling readers of varying ages.

Multimedia applications can provide enhanced access to ideas through visual and verbal representations. Pictures, drawings, diagrams, graphs, tables, and video clips (static, animated, or interactive) complement written words. Dual-coding theory states that simultaneously processing both verbal and visual representations of an idea, that is, dually coding the information, will make the information more memorable (Kamil *et al.*, 2000). Multimedia electronic text facilitates comprehension in young readers (Bus *et al.*, 2006), typically achieving and struggling elementary readers (Kinzer and Risko, 1998), and high school (Webb, 2005) and college students (Ainsworth and VanLabeke, 2004). However, the technological capacity to produce multimedia is currently well in advance of research to establish how characteristics of the reader, cognitive demands of the content, and the multimedia features of text should be aligned to produce an optimum learning environment for a particular reader (Hegarty, 2004).

A burgeoning area of inquiry is the use of technology to teach readers to be more strategic in their reading of text. Illustrative is the design and study of Interactive Strategy Training for Active Reading and Thinking (iSTART) (McNamara *et al.*, 2004), a Web-based tutoring program that uses animated agents to teach reading strategies to young, adolescent (grades 8–12), and college-aged students. The program is derived from research on self-explanation reading training (SERT) and teaches metacognitive reading strategies in the context of self-explanation. Specifically, users of iSTART are taught to engage in comprehension monitoring, paraphrasing, predicting, elaborating, and bridging (tying the current sentence with material covered previously in the text) as means of enhancing learning from text. When students are first introduced to the software, an agent (an animated electronic one-on-one trainer) provides self-explanations while the learner watches; but as the learner progresses through the modules, he or she creates self-explanations that are evaluated by the agent. Research to date suggests that the program is most beneficial for students with the least knowledge about the domain at hand, as well as for students who are less strategic in their approach to text. Furthermore, protocol analyses indicated that iSTART improved the quality of students' self-explanations that, in turn, were reflected in improved comprehension scores.

A program of research conducted by Dalton and her colleagues at the Center for Applied Special Technologies has been designed to investigate the use of

technology to enact universal design for learning (UDL), (Rose and Meyer, 2002), an approach through which the needs and interests of the broadest spectrum of learners is considered from the outset. In addition to providing such supports as TTS and online glossaries, students working in this environment have technology supports for being strategic in their reading. As they work their way through the text, they are periodically prompted to stop and apply a strategy. They enter their response in writing or audio-record it and save their response to an electronic work log that can be viewed at any time by the student and teacher. The program offers five levels of support that move students from high support to low support to independent application of the strategies, ending with an open-ended response option that can be used for any purpose (e.g., making a journal entry). Furthermore, the level of support is adjusted to meet the demands of particular strategies. For example, more support is provided for summarization than for predicting (Dalton *et al.*, 2002).

Dalton and Palincsar and their colleagues have explored the use of digital environments to support students' learning of challenging science concepts. In their reading-to-learn research (Dalton *et al.*, 2006), typically achieving and struggling fifth graders readers interacted with one of three versions of digital texts, all of which presented the same prose and graphics. The static version offered TTS and embedded vocabulary support. The interactive diagram version, which contained the same supports as the static version, also offered a prose/diagram-interaction feature (PDI) that would animate the graphic that corresponded with information presented in the prose; furthermore, students in this condition were directed to use a diagram-manipulation (DM) feature to explore ideas that were presented in the prose. The third version was enhanced with the addition of two animated pedagogical agents, who provided procedural and conceptual support. Students in the static group demonstrated significantly less growth in science vocabulary and concept learning than students in either of the interactive conditions. Students in the interactive conditions significantly outperformed their counterparts in the static condition, with struggling readers in both conditions making gains comparable to their typically achieving peers. There was no advantage to the addition of agents in this environment.

Reading on the Internet

The Internet has assumed an increasingly important role in contemporary life, with nearly 75% of all US households reporting that they had Internet access in 2004 (Nielsen/NETRatings, 2004), 93% of US classrooms reported to have at least one computer linked to the Internet (Kleiner and Lewis, 2003). In 2003, the ratio of students to instructional computers with Internet access in

public schools was 4.4 to 1, a decrease from 12.1 to 1 in 1998 (National Center for Education Statistics, 2005).

Literacy researchers argue that the open networked nature of the Internet requires that readers not only construct meaning from text, deploying the kinds of strategies described above, but also that readers must be sufficiently flexible and purposeful in their reading to make thoughtful, efficient, and effective use of the hyperlinks, graphics, and icons that are integral to accessing information in Internet contexts (Leu *et al.*, 2004; Spiro, 2004). The research in this area is lean, but points to the challenges that school-aged students experience when conducting inquiry on the Internet (Lyons *et al.*, 1997) and trying to use information that has been located there (Wallace *et al.*, 2000). Using case study and verbal-protocol analysis methodologies, Coiro and Dobler (2007) investigated how skilled sixth-grade readers searched a number of multilayered websites carrying out an array of activities (e.g., looking for answers to both literal and inferential questions). Their findings illustrated the broad array of processes their participants engaged in, including conventional strategies, such as goal setting, predicting, comprehension monitoring, and evaluation of relevancy, in addition to novel and more complex reading processes such as drawing upon prior content knowledge in hand with knowledge of website structures and Web-based search engines.

Theory Development Regarding Reading in Relationship to Technology

The preponderance of theory development informing our understanding of text comprehension has been conducted with traditional print text. There have been numerous and vivid attempts to capture the differences between traditional (print) text, multimedia, and electronic media and the implications of these differences for defining and studying literacy, broadly writ. Typically, while print literacy is described as linear, ordered, sequential, hierarchical, and logical, hypertext is characterized as fluid, spatial, decentered, bottom up, and playful (Burbulus, 2001; Ryan, 1999).

Reflecting these differences in text, there is a lively discussion in the literacy community about the need for expanded definitions of literacy and new theories that place ICTs at their center. Leu *et al.* (2004) argue that for the past 500 years, literacy has emerged from a variety of social contexts but has primarily been shaped by the book and printing press. But today, both the social context and technologies are rapidly changing; ICTs and the Internet are in the process of becoming the central technologies of literacy. Such a change demands new theory that acknowledges the changing nature of literacy as a function of technology, the multiple literacies called for, and enabled by new technologies, and the new forms of

strategic knowledge necessary for engaging in new literacies. Fruitful theoretical stances that seem well-matched to ICTs include: social semiotic theory (Lemke, 1998), cognitive flexibility theory (Spiro *et al.*, 1991), cultural theory (Cope and Kalantzis, 2000), and social and linguistic theory (Chandler-Olcott and Mahar, 2003). The next generation's encyclopedia entry on reading and technology is likely to reveal the fruits of applying these theoretical perspectives to the study of reading.

Summary

Burgeoning research and development in the area of technology has dramatically affected the landscape of opportunities for the teaching and learning of print literacy. Technology has been used to support learners in the areas of emergent literacy, phonemic awareness and phonics, fluency, vocabulary, and comprehension. While the utility of computers in preschool is constrained by availability, software has been used to support emergent readers' engagement with rich print materials. Software programs that support young learners' mastery of phonics and fluency development have demonstrated effectiveness, largely due to the need for extensive practice and extended exposure in these areas. Vocabulary development is also aided by the multiple varied encounters to important words that electronic texts can offer. Technology has been used to support reading comprehension in at least three ways: supporting decoding, offering multiple representations of texts, and teaching students to be more strategic in their reading of text. The Internet has provided learners with a new medium for developing literacy skills. Limited research in this area has shown that readers must be particularly flexible and purposeful when accessing information via the Internet.

See also: Universal Design for Learning.

Bibliography

- Adams, M. J. (2006). The promise of automatic speech recognition for fostering literacy growth in children and adults. In McKenna, M. C., Labbo, L. D., Kieffer, R. D., and Reinking, D. (eds.) *International Handbook of Literacy and Technology*, vol. II, pp 109–128. Mahwah, NJ: Erlbaum.
- Ainsworth, S. and VanLabeke, N. (2004). Multiple forms of dynamic representation. *Learning and Instruction* **14**, 241–255.
- Anderson, R. S. and Speck, B. W. (2001). *Using Technology in K-8 Literacy Classrooms*. Upper Saddle River, NJ: Prentice-Hall.
- Barker, T. A. and Torgesen, J. K. (1995). An evaluation of computer-assisted instruction in phonological awareness with below average readers. *Journal of Educational Computing Research* **13**(1), 89–103.
- Blachowicz, C. Z. and Beyersdorfer, J. (2006). Vocabulary development and technology: Teaching and transformation. In McKenna, M. C., Labbo, L. D., Kieffer, R. D., and Reinking, D. (eds.) *International Handbook of Literacy and Technology*, vol. II, pp 341–348. Mahwah, NJ: Erlbaum.
- Blok, H., Oostdam, R., Otter, M. E., and Overmatt, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research* **72**(1), 101–130.
- Boone, R. and Higgins, K. (1993). Hypermedia basal readers: Three years of school-based research. *Journal of Special Education Technology* **12**(2), 86–106.
- Burbulus, N. (2001). Paradoxes of the web: The ethical dimensions of credibility. *Library Trends* **49**(3), 441–453.
- Burns, S. M., Griffin, P., and Snow, C. E. (eds.) (1999). *Starting Out Right: A Guide to Promoting Children's Reading Success*. Washington, DC: National Academy Press.
- Bus, A. G., de Jong, M. T., and Verhallen, M. (2006). CD-ROM talking books: A way to enhance literacy? In McKenna, M. C., Labbo, L. D., Kieffer, R. D., and Reinking, D. (eds.) *International Handbook of Literacy and Technology*, vol. II, pp 341–348. Mahwah, NJ: Erlbaum.
- Chandler-Olcott, K. and Mahar, D. (2003). Adolescents' anime-inspired "fanfictions": An exploration of multiliteracies. *Journal of Adolescent and Adult Literacy* **46**(7), 556–566.
- Coiro, J. and Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for an locate information on the Internet. *Reading Research Quarterly* **42**(2), 214–257.
- Cope, B. and Kalantzis, M. (2000). *Multiliteracies: Literacy Learning and the Design of Social Futures*. London: Routledge.
- Dalton, B., Pisha, B., Eagleton, M., Coyne, P., and Deysher, S. (2002). Engaging the text: Reciprocal teaching and questioning strategies in a scaffolded learning environment (*Final Report to the U.S. Department of Education, Office of Special Education Programs*). Peabody, MA: Center for Applied Special Technology.
- Dalton, B., Proctor, P., Palincsar, A., et al. (2006). The role of interactive diagrams and pedagogical agents in supporting struggling and typically achieving readers' comprehension of digital science text. *Paper Presented at the Annual Meeting of the National Reading Conference*, Los Angeles, CA.
- Das-Smaal, E. A., Klapwijk, M. J. G., and van der Leij, A. (1996). Training of perceptual unit processing in children with a reading disability. *Cognition and Instruction* **14**(2), 221–250.
- Dawson, L., Venn, M. L., and Gunter, P. L. (2000). The effects of teacher versus computer reading models. *Behavioral Disorders* **25**(2), 105–113.
- Duke, N. K. and Pearson, P. D. (2002). Effective practices for developing reading comprehension. In Farstrup, A. E. and Samuels, S. J. (eds.) *What Research Has to Say About Reading Instruction*, pp 205–242. Newark, DE: International Reading Association.
- Dynarski, M., Agodini, R., Heavisude, S., et al. (2007). Effectiveness of reading and mathematics software products: Findings from the first student cohort. *Report to Congress*. NCEE 2007-4006. Washington, DC: US Department of Education.
- Elbro, C., Rasmussen, I., and Spelling, B. (1996). Teaching reading to disabled readers with language disorders: A controlled evaluation of synthetic speech feedback. *Scandinavian Journal of Psychology* **37**, 140–155.
- Elkind, J., Cohen, K., and Murray, C. (1993). Using computer-based readers to improve reading comprehension of students with dyslexia. *Annals of Dyslexia* **43**, 238–259.
- Farmer, M. E., Klein, R., and Bryson, S. E. (1992). Computer-assisted reading: Effects of whole-word feedback on fluency and comprehension in readers with severe disabilities. *Remedial and Special Education* **13**(2), 50–60.
- Hegarty, M. (2004). Dynamic visualizations and learning: Getting to the difficult questions. *Cognition and Instruction* **14**, 343–351.
- Heimann, M., Nelson, K. E., Tjss, T., and Gillberg, C. (1995). Increasing reading and communication skills in children with autism through an interactive multimedia computer program. *Journal of Autism and Developmental Disorders* **25**(5), 459–480.
- Higgins, E. L. and Raskind, M. H. (2000). Speaking to read: The effects of continuous vs. discrete speech recognition systems on the reading and spelling of children with learning disabilities. *Journal of Special Education Technology* **15**(1), 19–29.
- Higgins, K., Boone, R., and Lovitt, T. (1996). Hypertext support for remedial students and students with learning disabilities. *Journal of Learning Disabilities* **29**(4), 402–412.

- Kamil, M. L., Intrator, S. M., and Kim, H. S. (2000). The effects of other technologies on literacy and literacy learning. In Kamil, M. L., Mosenthal, P. B., Pearson, P. D., and Barr, R. (eds.) *Handbook of Reading Research*, vol. III, pp 771–788. Mahwah, NJ: Erlbaum.
- Kinzer, C. A. and Risko, V. J. (1998). Multimedia and enhanced learning: Transforming preservice education. In Reinking, D., McKenna, M. C., Labbo, L. D., and Kieffer, R. D. (eds.) *Handbook of Literacy and Technology: Transformations in a Post-Typographical World* pp 185–202. Mahwah, NJ: Erlbaum.
- Kleiner, A. and Lewis, L. (2003). *Internet Access in Public Schools and Classrooms: 1994–2002*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education.
- Kuhn, M. R. and Stahl, S. A. (2006). More than skills and drill: Exploring the potential of computers in decoding and fluency instruction. In McKenna, M. C., Labbo, L. D., Kieffer, R. D., and Reinking, D. (eds.) *International Handbook of Literacy and Technology*, vol. II, pp 295–301. Mahwah, NJ: Erlbaum.
- Labbo, L. D. and Ash, G. E. (1998). What is the role of computer-related technology in early literacy? In Neuman, S. B. and Roskos, K. A. (eds.) *Children Achieving: Best Practices in Early Literacy* pp 180–197. Newark, DE: International Reading Association.
- Lemke, J. L. (1998). Metamedia literacy: Transforming meanings and media. In Reinking, D., McKenna, M. C., Labbo, L. D., and Kieffer, R. D. (eds.) *Handbook of Literacy and Technology: Transformations in a Post-Typographic World*, pp 283–301, Mahwah, NJ: Lawrence Erlbaum Associates.
- Leong, C. K. (1995). Effects of on-line reading and simultaneous DECtalk auding in helping below-average and poor readers comprehend and summarize text. *Learning Disability Quarterly* **18**, 101–116.
- Leu, D. J. Jr., Kinzer, C. K., Coiro, J., and Cammack, S. W. (2004). Toward a theory of new literacies emerging from the Internet and other information and communication technologies. In Ruddell, R. B. and Unrau, N. (eds.) *Theoretical Models and Processes of Reading*, 5th edn, pp 1570–1613. Newark, DE: International Reading Association.
- Lundberg, I. and Olofsson, A. (1993). Can computer speech support reading comprehension? *Computers in Human Behavior* **9**, 283–293.
- Lyons, D., Hoffman, J., Kracjik, J., and Soloway, E. (1997). An investigation of the use of the World Wide Web for online inquiry in a science classroom. *Paper Presented at the Annual Meeting of the National Association for Research in Science Teaching*, Chicago, IL.
- MacArthur, C., Ferretti, R., Okolo, C. M., and Cavalier, A. R. (2001). Technology applications for students with literacy problems: A critical review. *Elementary School Journal* **101**(3), 273–301.
- MacArthur, C. A. and Haynes, J. B. (1995). Student assistant for learning from text (SALT): A hypermedia reading aid. *Journal of Learning Disabilities* **28**(3), 50–59.
- Macaruso, P., Hook, P. E., and McCabe, R. (2006). The efficacy of computer-based supplementary phonics programs for advancing reading skills in at-risk elementary students. *Journal of Research in Reading* **29**(2), 162–172.
- McGee, L. M. and Richgels, D. J. (2006). Can technology support emergent reading and writing? Directions for the future. In McKenna, M. C., Labbo, L. D., Kieffer, R. D., and Reinking, D. (eds.) *International Handbook of Literacy and Technology*, vol. II pp 369–377. Mahwah, NJ: Erlbaum.
- McKenna, M. C. (1998). Electronic texts and the transformation of beginning reading. In Reinking, D., McKenna, M. C., Labbo, L. D., and Kieffer, R. D. (eds.) *Handbook of Literacy and Technology: Transformations in a Post-Typographical World*, pp 45–60. Mahwah, NJ: Erlbaum.
- McKenna, M. C. (2002). Issues in technology: Phonics software for a new millennium. *Reading and Writing Quarterly* **18**, 93–96.
- McNamara, D. S., Levinstein, I. B., and Boonthum, C. (2004). iStart: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, and Computers* **36**(2), 222–233.
- Mitchell, M. J. and Fox, B. J. (2001). The effects of computer software for developing phonological awareness in low-progress readers. *Reading Research and Instruction* **40**(4), 315–332.
- Montali, J. and Lewandowski, L. (1996). Bimodal reading: Benefits of a talking computer for average and less skilled readers. *Journal of Learning Disabilities* **29**(3), 271–279.
- Moore-Hart, M. A. (1995). The effects of multicultural links on reading and writing performance and cultural awareness of fourth and fifth graders. *Computers in Human Behavior* **11**(3–4), 391–410.
- Moseley, D. and Hartas, C. (1993). Say that again please: A scheme to boost reading skills using a computer with digitized speech. *Support for Learning* **8**(1), 16–21.
- Nagy, W. E. and Scott, J. A. (2000). Vocabulary processes. In Kamil, M. L., Mosenthal, P. B., Pearson, P. D., and Barr, R. (eds.) *Handbook of Reading Research*, vol. III, pp 209–228. Mahwah, NJ: Erlbaum.
- National Center for Education Statistics (2005). *Internet Access in U.S. Public Schools and Classrooms: 1994–2003* (Parsad, B. and Jones, J). Washington, DC.
- National Institute of Child Health and Human Development (2000). *Report of the National Reading Panel. Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction* (NIH Publication No. 00-4769). Washington, DC: US Government Printing Office.
- Nielsen/NETRatings (2004). Three out of four Americans have access to the Internet according to Nielsen/NETRatings. <http://www.nielsen-netratings.com> (accessed September 2009).
- Olson, R. K. and Wise, B. (2006). Computer-based remediation for reading and related phonological disabilities. In McKenna, M. C., Labbo, L. D., Kieffer, R. D., and Reinking, D. (eds.) *International Handbook of Literacy and Technology*, vol. II, pp 57–74. Mahwah, NJ: Erlbaum.
- Purves, A. (1999). The web of text and the web of god: An essay on the third information transformation. *Canadian Journal of Communication* **24**(1).
- Rand Reading Study Group (2002). *Reading for Understanding: Toward an R&D Program in Reading Comprehension*. Santa Monica, CA: RAND.
- Raskind, M. H. and Higgins, E. L. (1999). Speaking to read: The effects of speech recognition technology on the reading and spelling performance of children with learning disabilities. *Annals of Dyslexia* **49**, 251–281.
- Rose, D. and Meyer, A. (2002). *Teaching Every Student in the Digital Age: Universal Design for Learning*. Alexandria, VA: ASCD.
- Roskos, K. and Neuman, S. B. (2001). Environment and its influences for early literacy teaching and learning. In Neuman, S. B. and Dickenson, D. K. (eds.) *Handbook of Early Literacy Research* pp 163–172. New York: Guilford.
- Ryan, M. (1999). Cyberspace, virtuality, and the text. In Ryan, M. (ed.) *Cyberspace Textuality: Computer Technology and Literary Theory*, pp 78–107. Bloomington, IN: Indiana University Press.
- Snow, C. E., Burns, S. M., and Griffin, P. (eds.) (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.
- Spiro, R. J. (2004). Principled pluralism for adaptive flexibility in teaching and learning. In Rudell, R. B. and Unrau, N. (eds.) *Theoretical Models and Processes of Reading*, 5th edn, pp 654–659. Newark, DE: International Reading Association.
- Spiro, R. J., Feltovich, P. J., Jacobson, M. J., and Coulson, R. L. (1991). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. *Educational Technology* **11**(5), 24–33.
- Turbill, J. and Murray, J. (2006). Early literacy and new technologies in Australian schools: Policy, research, and practice. In McKenna, M. C., Labbo, L. D., Kieffer, R. D., and Reinking, D. (eds.) *International Handbook of Literacy and Technology*, vol. II, pp 93–108. Mahwah, NJ: Erlbaum.
- Van Daal, V. H. P. and Reitsma, P. (1993). The use of speech feedback by normal and disabled readers in computer-based reading practice. *Reading and Writing: An Interdisciplinary Journal* **5**(3), 243–259.
- Van Daal, V. H. P. and van der Leij, A. (1992). Computer-based reading and spelling practice for children with learning disabilities. *Journal of Learning Disabilities* **23**, 186–195.
- Wallace, R. M., Kupperman, J., Krajcik, J., and Soloway, E. (2000). Science on the web: Students online in a sixth grade classroom. *Journal of the Learning Sciences* **9**, 75–104.

- Webb, M. E. (2005). Affordances of ICT in science learning: Implications for an integrated pedagogy (1995). *International Journal of Science Education* 27(6), 705–735.
- Wise, B. W. and Olson, R. K. (1995). Computer-based phonological awareness and reading instruction. *Annals of Dyslexia* 45, 99–122.
- Wise, B. W., Olson, R. K., Ring, J., and Johnson, M. (1998). Interactive computer support for improving phonological skills. In Metsala, J. and Ehri, L. (eds.) *Word Recognition in Beginning Literacy* pp 189–208. Mahwah, NJ: Erlbaum.
- Wise, B. W., Ring, J., and Olson, R. K. (1999). Training phonological awareness with and without explicit attention to articulation. *Journal of Experimental Child Psychology* 72, 271–304.
- Wise, B. W., Ring, J., and Olson, R. K. (2000). Individual differences in gains from computer-assisted remedial reading. *Journal of Experimental Child Psychology* 77, 197–205.
- Kamil, M. L., Intrator, S. M., and Kim, H. S. (2001). The effects of other technologies on literacy and literacy learning. In Kamil, M. L., Mosenthal, P. B., Pearson, P. D., and Barr, R. (eds.) *Handbook of Reading Research*, vol. III, pp 771–788. Mahwah, NJ: Erlbaum.
- Leu, D. J. (2000). Literacy and technology: Deictic consequences for literacy education in an information age. In Kamil, M. L., Mosenthal, P. B., Pearson, P. D., and Barr, R. (eds.) *Handbook of Reading Research*, vol. III, pp 743–770. Mahwah, NJ: Erlbaum.
- MacArthur, C. A., Ferretti, R. P., Okolo, C. M., and Cavalier, A. R. (2001). Technology applications for students with learning disabilities: A critical review. *Elementary School Journal* 101(3), 272–301.
- McKenna, M. C., Labbo, L. D., Kieffer, R. D., and Reinking, D. (eds.) (2006). *International Handbook of Literacy and Technology* vol. II. Mahwah, NJ: Erlbaum.
- Moreno, R. and Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology* 94(1), 256–163.
- Olson, R. K., Wise, B., Ring, J., and Johnson, M. (1997). Computer-based remedial training in phoneme awareness and phonological decoding: Effects on the post-training development of word recognition. *Scientific Studies of Reading* 1, 235–253.
- Reinking, D., McKenna, M. C., Labbo, L. D., and Kieffer, R. D. (eds.) (1998). *Handbook of Literacy and Technology: Transformations in a Post-Typographical World*. Mahwah, NJ: Erlbaum.
- Verhoeven, L., Segers, E., Bronkhorst, J., and Boves, L. (2006). Toward interactive literacy education in the Netherlands. In McKenna, M. C., Labbo, L. D., Kieffer, R. D., and Reinking, D. (eds.) *International Handbook of Literacy and Technology*, vol. II, pp 41–53. Mahwah, NJ: Erlbaum.
- Wentink, H. W. M. J., van Bon, W. H. J., and Schreuder, R. (1997). Training of poor readers' phonological decoding skills: Evidence for syllable-bound processing. *Reading and Writing: An Interdisciplinary Journal* 9(3), 163–192.

Further Reading

- Anderson-Inman, L. M. and Horney, M. A. (1998). Transforming text for at-risk readers. In Reinking, D., McKenna, M. C., Labbo, L. D., and Kieffer, R. D. (eds.) *Handbook of Literacy and Technology: Transformations in a Post-Typographical World*, pp 15–43. Mahwah, NJ: Erlbaum.
- Blok, H., Oostdam, R., Otter, M. E., and Overmatt, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research* 72(1), 101–130.
- Dalton, B. and Strangman, N. (2006). Improving struggling readers' comprehension through scaffolded hypertexts and other computer-based literacy programs. In McKenna, M. C., Labbo, L. D., Kieffer, R. D., and Reinking, D. (eds.) *International Handbook of Literacy and Technology*, vol. II, pp 75–92. Mahwah, NJ: Erlbaum.
- Frontera (2007). Kinder Project. <http://www.geocities.com/Athens/Acropolis/4616/projects/projects.html> (accessed September 2009).
- Hall, T. E., Hughes, C. A., and Filbert, M. (2000). Computer-assisted instruction in reading for students with learning disabilities: A research synthesis. *Education and Treatment of Children* 23(2), 173–193.

Relevant Website

<http://www.geocities.com> – Kinder Project.

Technological Supports for Acquiring Twenty-First-Century Skills

C Dede, Harvard Graduate School of Education, Cambridge, MA, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Augmented reality – A simulated experience created by interweaving real and digital people, places, and objects.

Avatar – The digital representative of a participant in a virtual world.

Complex communication – The ability to help others master complicated concepts through providing a variety of explanations and examples.

Data mining – The techniques used to find and interpret interesting patterns in very large data sets.

Digital literacy – The types of fluent use of tools, applications, and media to accomplish complex tasks.

Expert decision making – The ability to solve difficult problems for which all standard methods of problem solving fail.

Immersive media – The emerging media, such as multi-user virtual environments and augmented realities, in which the user has the sensation of being present in a digitally created world.

Multi-user virtual environment (MUVE) – A virtual world containing the avatars of users, computer-based agents, and digital artifacts.

Situated learning – Learning through apprenticeships and mentoring on authentic, real-world tasks.

Skills Needed for Twenty-First-Century Work and Citizenship

The twenty-first century seems quite different from the twentieth in the capabilities people that need for work, citizenship, and self-actualization. In response, society's educational systems must transform their objectives, curricula, pedagogies, and assessments to help all students attain the sophisticated outcomes requisite for a prosperous, attractive lifestyle based on effective contributions in work and citizenship. This article describes an innovative strategy by which new pedagogies based on emerging immersive media can aid all students in attaining sophisticated twenty-first-century skills and knowledge.

Thus far, the twenty-first century has seen a dramatic shift in the economic model for industrialized countries

(Dede *et al.*, 2005). Systems of economic development based on geography, trade rules, and tariffs; slow dissemination of scientific and technological discoveries; and long cycles of product life have given way to global trade, rapid product innovation, the lowering of trade barriers, rapid dissemination of scientific and technological discovery, and quick global deployment and movement of capital and the means of production (Chang, 2003). In the early twenty-first century, income and wealth come from applying technology and novel ideas to create new products and processes. Adding value to products and processes is the key to growing jobs and incomes in this new economic environment (Aubert and Reiffers, 2004).

Numerous reports on the global, knowledge-based economy and the flat world document that tomorrow's workers must be prepared to shift jobs and careers more frequently, to be flexible and adaptable in acquiring job skills, and to integrate and focus a changing mix of job-derived and education-based knowledge on business processes and problems (Card and Dinardo, 2002; Business-Higher Education Forum, 2005; National Academy of Science, 2006; American Association of Colleges and Universities, 2007). As the global economy continues to evolve, predictions are that workers will change jobs seven or eight times during their work life. The worker of the twenty-first century must have science and mathematics skills, creativity, fluency in information and communication technologies (ICT), and the ability to solve complex problems (Friedman, 2005); these are also important capabilities for citizenship at the local, national, and global level.

Further, the types of work done by people, as opposed to the kinds of labor done by machines, are continually shifting as computers and telecommunications expand their capabilities to accomplish human tasks. Economists Levy and Murnane (2004) have documented a very important aspect of what constitutes twenty-first-century understandings and performances:

Declining portions of the labor force are engaged in jobs that consist primarily of routine cognitive work and routine manual labor – the types of tasks that are easiest to program computers to do. Growing proportions of the nation's labor force are engaged in jobs that emphasize expert thinking or complex communication – tasks that computers cannot do. (pp. 53–54)

These economists go on to explain that “expert thinking [involves] effective pattern matching based on detailed

knowledge; and metacognition, the set of skills used by the stumped expert to decide when to give up on one strategy and what to try next” (Levy and Murnane, 2004: 75). What a skilled auto-mechanic does when all diagnostic systems show normal functioning, but the car is still malperforming is expert decision making – inventing new problem-solving heuristics when all standard strategies have failed. “Complex communication requires the exchange of vast amounts of verbal and nonverbal information. The information flow is constantly adjusted as the communication evolves unpredictably” (Levy and Murnane, 2004: 94). A skilled teacher is an expert in complex communication, able to improvise answers and facilitate dialog in the unpredictable, chaotic flow of classroom discussion.

Twenty-first-century education needs to prepare students for a world in which almost all types of routine cognitive tasks are done by computers and in which expert thinking and complex communications are the core intellectual capabilities by which people attain prosperity and economic security individually, as a region, and as a nation. These higher-order performances are based on fundamental knowledge about how to do simpler types of work, so the shift needed is not to remove the learning of routine cognitive skills (such as basic arithmetic operations) from the curriculum. Rather, the fundamental change involves deemphasizing fluency in simple procedures as an end-goal of preparation for work and life (e.g., counting bills as a bank teller), and instead using these routine skills as a substrate for mastering complex mental performances valued in the future workplace, such as advising clients about global investment strategies tailored to their individual situations.

Thus, a crucial challenge for education in all countries is to align curriculum and learning to new economic and governance models based both on a global, knowledge-based workplace and on emerging world-level problems such as human-induced climate change (Friedman, 2008). To accomplish this, we must transform children’s learning processes in and out of school and engage student interest in gaining twenty-first-century skills and knowledge for work, citizenship, and a satisfying lifestyle.

Organizing Twenty-First-Century Skills into a Curricular Framework

A variety of groups have developed frameworks that describe the twentieth-century skills that students should master as part of their education. Space does not permit summarizing and comparing all these descriptions. Such litanies of twenty-first-century skills often lack clarity about why a skill is twenty-first-century, as opposed to knowledge that has been generically useful at any point in history. Determining what constitutes a twenty-first-century skill requires stipulating the metrics by which one judges whether a human performance is truly significant in

its projected importance to attaining an attractive, prosperous job and lifestyle – and also is sufficiently different from twenty-first-century skills to merit inclusion. Assessing the degree to which the capability is valuable in work, and citizenship is key to these metrics, as is distinguishing in kind between perennial and contextual performances.

For example, collaboration is a perennial capability, always valued as a trait in workplaces across the centuries; as such, the basic value of this interpersonal performance is not intrinsically special to our emerging economic context. However, the degree of importance for collaborative capacity is growing in an era where work is increasingly done by teams of people with complementary expertise and roles, as opposed to individuals doing manual operations on an assembly line. Thus, even though perennial in nature, collaboration is worthy of inclusion as a twenty-first-century performance because, for the context in which today’s students will function as adults, the importance of cooperative interpersonal capabilities is substantially higher than in the prior industrial era.

In contrast, the ability to rapidly filter huge amounts of incoming data, extracting information valuable for decision making, is a contextual capability. Due to the prevalence of ICT, for the first time in human history people are inundated by enormous amounts of data that they must access, manage, integrate, and evaluate. The ability to separate signal from noise in a potentially overwhelming flood of incoming data is a twenty-first-century performance not in degree but in kind – because this is novel in history as a valuable capability. This distinction is important because, unlike perennial capabilities, new types of human performances are typically not part of the legacy curriculum inherited from twentieth-century educational systems.

Some frameworks for twenty-first-century skills discuss new literacies based on the evolution of ICT (Dede, 2008). For example, Jenkins *et al.* (2006) delineate a set of novel literacies based on new media (**Table 1**).

This list illustrates how twenty-first-century skills, such as collaboration and information filtering, are linked to the capabilities of emerging interactive media.

Dede *et al.* (2005: 5) describes the types of learning strengths, styles, and preferences of what he calls neomillennial students. These students are defined not by the timeframe of their birth but by the way they like to learn – through use of immersive collaborative media, such as multiplayer online games or virtual environments like Second Life:

- fluency in multiple media, valuing each for the types of communication, activities, experiences, and expressions it empowers;
- learning based on collectively seeking, sieving, and synthesizing experiences, rather than individually locating and absorbing information from some single best source;

Table 1 Selected new literacies delineated by Jenkins *et al.* (2006)

Play	The capacity to experiment with one's surroundings as a form of problem solving
Performance	The ability to adopt alternative identities for the purpose of improvisation and discovery
Simulation	The ability to interpret and construct dynamic models of real-world processes
Appropriation	The ability to meaningfully sample and remix media content
Multitasking	The ability to scan one's environment and shift focus as needed to salient details
Distributed cognition	The ability to interact meaningfully with tools that expand mental capacities
Collective intelligence	The ability to pool knowledge and compare notes with others toward a common goal
Judgment	The ability to evaluate the reliability and credibility of different information sources
Transmedia navigation	The ability to follow the flow of stories and information across multiple modalities
Networking	The ability to search for, synthesize, and disseminate information
Negotiation	The ability to travel across diverse communities, discerning and respecting multiple perspectives, and grasping and following alternative norms

From Jenkins, H., Clinton, K., Purushotma, R., Robinson, A. J., and Weigel, M. (2006). *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*. Chicago, IL: The MacArthur Foundation, with permission from Henry Jenkins.

- active learning based on experience (real and simulated) that includes frequent opportunities for reflection;
- expression through nonlinear, associational webs of representations rather than linear stories (e.g., authoring a simulation and a web page to express understanding, rather than a paper); and
- co-design of learning experiences personalized to individual needs and preferences.

As discussed later, using immersive collaborative simulations in classroom settings offers a powerful method to nurture twenty-first-century skills.

Leu *et al.* (2007) describe four characteristics of the new literacies generated by ICT. First, emerging ICT tools, applications, media, and environments require novel skills, strategies, and dispositions for their effective use. Second, new literacies are central to full economic, civic, and personal participation in a globalized society. Third, new literacies constantly evolve as their defining ICT continuously are renewed through innovation. Fourth, new literacies are multiple, multimodal, and multifaceted.

Leu's third point raises important issues about stability: How durable are twenty-first-century performances in their applicability to work, citizenship, and self-actualization? How quickly will additional, important twenty-first-century understandings and behaviors emerge as ICT continue to evolve? Certainly, tools, applications, media,

and environments are changing rapidly, with no end in sight. Typically, despite predictions of paperless offices or the end of the book, this evolution involves adding additional literacies and understandings rather than new performances displacing previously useful behaviors. Examining predictions about the future of ICT helps to illuminate what these additional literacies may be and what knowledge and skills may become less useful over time.

Overall, the skills advocated by various groups as twenty-first century are sophisticated cognitive, affective, and psychosocial performances generally manifested through fluent use of digital tools and media. In the twenty-first century, people frequently use ICT to accomplish objectives never before attainable (e.g., remote collaboration via groupware among a team scattered across the globe). How can we not only prepare today's students for the new jobs generated by these capabilities, but also build their capacity to hold novel careers that will emerge a decade or two from now based on future breakthroughs in ICT and other fields?

An Innovative Strategy for Teaching Illustrative Twenty-First-Century Skills

The remainder of this article describes how innovative, technology-aided pedagogies can help students learn sophisticated cognitive, affective, and psychosocial performances important for twenty-first-century work and citizenship. A cluster of twenty-first-century skills neglected in most current curricula is collective problem resolution via mediated interaction. In twenty-first-century work, knowledge is grounded in a setting and distributed across a community, rather than abstract and isolated within individuals (OECD, 2004). Problem finding (the front end of the inquiry process: making observations and inferences, developing hypotheses, and conducting experiments to test alternative interpretations of the situation) is crucial to reaching a point where the work team can do problem solving. Individual and collective metacognitive strategies for making meaning out of complexity (such as making judgments about the value of alternative problem formulations) are vital.

Each person involved in this interrelated suite of twenty-first-century skills has strong strategies both in effective pattern matching based on detailed knowledge and in judging when to give up on a particular problem-solving strategy and instead try another approach. Individuals on the work team are adept at manipulating sophisticated ICT applications and representations utilized within the complementary perspectives they bring to bear (e.g., using a spreadsheet to examine financial hypotheticals). They also are skilled in expressing core insights from their knowledge to others who have different backgrounds and experiences. Richly interactive

complex communication among team members is not limited to face-to-face dialog, but frequently relies on mediated interaction across distance in which the team co-constructs and negotiates shared interpretive understandings and a problem-resolution strategy.

For example, a school district might task a team of teachers, school administrators, parents, and local business executives to develop a plan for improving students' educational outcomes in mathematics. Potential factors leading to subpar educational performance include individual differences in native language, gender, culture, and socioeconomic status; teachers' experience and preparation in mathematical content, subject-specific pedagogy, classroom management, and student engagement; state and district policies related to educational reform, the curricular materials used in mathematics, and the capacity of the technology infrastructure at local schools, among others. Under these circumstances, individual and collective skills in problem finding, inquiry, metacognition, collaboration, expert decision making, complex communication, and use of ICT tools, communicative media, and representations are vital to the team's success.

In classroom settings, sophisticated ICT capable of supporting the teaching of such collective problem resolution via mediated interaction are now emerging. This is important because, until this point in history, schools have lacked the capacity to inculcate twenty-first-century skills and knowledge best learned through participation in real-world communities of practice. Reports such as the National Research Council's *How People Learn* cite situated theories of learning (mentoring and apprenticeships in communities of authentic, real-world practice) as powerful in life, but very difficult to achieve in school settings. Emerging interactive media now have the capability to redress this deficit, a vital advance in teaching many twenty-first-century skills.

Emerging Media for Situated, Immersive, and Collaborative Simulation

Three complementary technological interfaces are currently shaping how people learn, with multiple implications for K-12 education:

- The familiar world-to-the-desktop interface provides access to distributed knowledge and expertise across space and time through networked media. This interface provides the models for learning that now underlie most tools, applications, and media in K-12 education.
- Emerging multi-user virtual environment (MUVE) interfaces offer students an engaging Alice-in-Wonderland experience in which their digital emissaries in a graphical virtual context actively engage in experiences with the avatars of other participants and with computerized agents. Researchers are exploring

the affordances of such models for learning in K-12 education (Dede, 2009).

- Augmented reality (AR) interfaces enable ubiquitous computing models. Students carrying mobile wireless devices through real-world contexts engage with virtual information superimposed on physical landscapes (such as a tree describing its botanical characteristics or an historic photograph offering a contrast with the present scene). This type of mediated immersion infuses digital resources throughout the real world, augmenting students' experiences and interactions. Scholars are starting to study how these models for learning aid students' engagement and understanding (Klopfer, 2008).

MUVEs empower creating digital contexts inaccessible in the real world, while AR enables the infusion of virtual contexts within physical locations. For reasons of space, only the MUVE interface is used as an example of immersive collaborative simulation to teach twenty-first-century skills.

In a MUVE, knowledge is grounded in a setting and distributed across a community, rather than isolated within individuals. Contrary to conventional K-12 instruction where knowledge is decontextualized and explicit, in MUVEs, the learning is situated and tacit; problem finding is central to problem solving. This parallels the nature of twenty-first-century work, as well as the learning styles and strengths of today's digital-age students.

The Power of Situated Learning for Inculcating Twenty-First-Century Performances

Situated learning requires authentic contexts, activities, and assessment coupled with guidance from expert modeling, mentoring, and legitimate peripheral participation. As an example of legitimate peripheral participation, graduate students work within the laboratories of expert researchers, who model the practice of scholarship. These students interact with experts in research as well as with other members of the research team who understand the complex processes of scholarship to varying degrees. While in these laboratories, students gradually move from novice researchers to more advanced roles, with the skills and expectations for them evolving. Potentially quite powerful, situated learning is infrequently used for classroom instruction because classrooms are by design isolated from the real world in order to encourage reflection and to provide custodial care. However, immersive interfaces can draw on the power of situated learning even in school settings by enabling collaborative simulations with problems and contexts similar to the real world.

Situated learning is important in part because of the crucial issue of transfer. Transfer is defined as the application of knowledge learned in one situation to

another situation and is demonstrated if instruction on a learning task leads to improved performance on a transfer task, typically a skilled performance in a real-world setting. One of the major criticisms of instruction today is the low rate of transfer generated by conventional instruction. Even students who excel in schooling or training settings often are unable to apply what they have learned to similar real-world contexts. Situated learning addresses this challenge by making the setting in which learning takes place similar to the real-world context for performance in work or personal life. Learning in well-designed digital contexts can lead to the replication in the real world of behaviors successful in simulated environments (Schwartz *et al.*, 2005).

Moreover, the evolution of an individual's or group's identity is an important type of learning for which simulated experiences in immersive interfaces are well suited. Reflecting on and refining an individual identity is often a significant issue for students of all ages (Colwell, 2007), and learning to evolve group and organizational identity is a crucial skill in enabling innovation and in adapting to shifting contexts (Engeström and Middleton, 1996). The social sciences see both the self and the organization as often fragmented, with complementary parts, rather than centralized and unitary (Sternberg and Zhang, 2001). Identity play through trying on various representations of the self and the group in virtual environments provides a means for different sides of a person or team to find common ground and the opportunity for synthesis and evolution, vital for the affective and psychosocial aspects of learning twenty-first-century skills.

Immersion is important in this process of identity exploration because virtual identity is potentially unconstrained by physical attributes such as gender, race, and disabilities, even though individuals often infuse these attributes into their avatars (Everett and Watkins, 2008; Pitaru, 2008). Virtual environments based on games and simulations illustrate how participants take advantage of fluidity in the identities they present. Simulations in virtual environments and augmented realities increase the value of these explorations by providing realistic feedback on how the real world responds to various patterns of individual and group behavior.

But what is so special about the situated learning now enabled in classrooms by emerging media? After all, outside of school contexts, students have opportunities for situated learning without using technology. One attribute that makes mediated immersion different and powerful is the ability to access information resources and psychosocial community distributed across distance and time, broadening and deepening experience. A second important attribute is the ability to create interactions and activities in mediated experience not possible in the real world, such as teleporting within a virtual environment, enabling a distant person to see a real-time image of

your local environment, or interacting with a (simulated) chemical spill in a busy public setting. Both of these attributes are actualized in immersive interfaces.

Learning Twenty-First-Century Skills in the River City MUVE

With funding from the National Science Foundation, for almost a decade, the author's research team has used design-based research methods to study an educational MUVE called River City. In the River City MUVE, a participant takes on the identity of an avatar, a virtual personal in the world, and communicates with other participants' avatars via text chat and virtual gestures. In this graphical virtual context (Figure 1), participants also interact with digital artifacts, such as viewing pictures or manipulating tools (e.g., an online microscope), as well as with computer-based agents (Figure 2).

The River City curriculum is a middle school science unit designed around national content standards and assessments in biology, ecology, epidemiology, and scientific inquiry. Students work in teams of three to collaboratively solve the problem of why the residents of River City are falling ill. Students travel back in time – into a series of virtual worlds – to the period in history when scientists were just discovering bacteria.

The curriculum is historically accurate and contains pictures from the Smithsonian Institute that help to portray accurately what the time period was like. The River City virtual world is an industrial eighteenth-century city with a river running through it (see Figure 3). Different forms of terrain influence water runoff in the various neighborhoods (wealthy area, middle-class area, and tenements). Digital artifacts include audio clues of sick residents coughing or mosquitoes buzzing, as well as images from that period in history (Figure 4).

Engaging in inquiry also involves students learning how to collect data and test their hypotheses and use the tools of scientists, such as an online microscope. For example, in River City, students can use the virtual microscope to

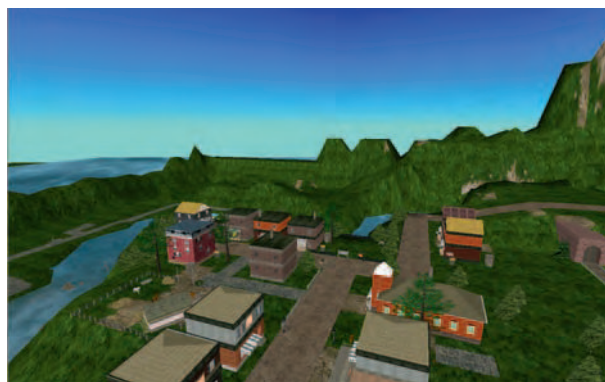


Figure 1 Aerial view of River City.

take water samples from one of the water-sampling stations in the city (Figures 5 and 6).

Students use twenty-first-century knowledge and skills to identify a problem and then develop and test a hypothesis based on one of the three disease strands infecting the city (i.e., insect-borne, air-borne, and water-borne). After testing

their hypotheses, students analyze their data and then write up their research in the form of a report to the mayor of River City. The report, based on the concept of a lab report, describes their experiment, research findings, conclusions, and recommendations for how the mayor can stop the spread of illness in River City.

These reports constitute representations of students' inquiry learning. From a technical standpoint, MUVEs are unique in their ability to keep minutely detailed records of the moment-by-moment movements, actions, and utterances of each participant in the environment. With the data-tracking system, we are able to collect, store, and retrieve information on the activities of each student as s/he explores the MUVE. These data form the basis of a personal MUVE history of each student that follows him or her from session to session, in the form of extensive log files – a feature impossible to replicate in a classroom-based experience. The level of detail in these records is extensive: the logs indicate exactly where students went, with whom they communicated, what they said in these interactions, what virtual artifacts they activated, and how long each of these activities took. Using data-mining techniques on these detailed event logs provides a powerful lens by which we can study individual students' use of twenty-first-century skills in this learning environment and document gains in students' engagement, learning, and self-efficacy.



Figure 2 Avatar and agent.

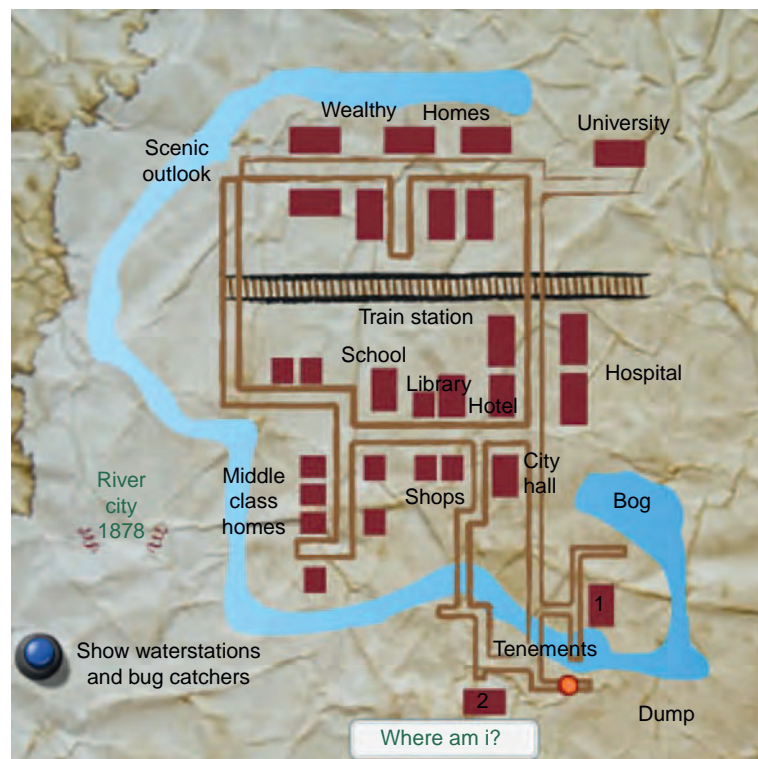


Figure 3 Map of River City.



Figure 4 Noted nineteenth-century woman scientist.

Conclusion

This extended example illustrates how situated, immersive, collaborative simulations based on MUEs can support twenty-first-century understandings and performances, such as collective problem resolution via mediated interaction, while in classroom settings. Other immersive interfaces, such as AR, can accomplish similar engagement and learning. Over time, these interfaces are becoming increasingly prevalent (Dede, 2009), and research is revealing how best to actualize their potential for learning and teaching.

A huge challenge confronts efforts to implement new technologies such as those as described in this article: What do we deemphasize in current instruction and assessment to make room for twenty-first-century understandings? The curriculum is crowded with low-level facts and recipe-like procedures (e.g., In what year did Columbus discover America? What are the seven steps of historical inquiry?), as opposed to nuanced understandings and performances. (i.e., what confluence of technological, economic, and political forces led to the age of exploration around the end of the fifteenth century? By what process of interpreting historical data did you reach this conclusion?)

Downgrading the importance of some material in the current curriculum is much harder than adding content

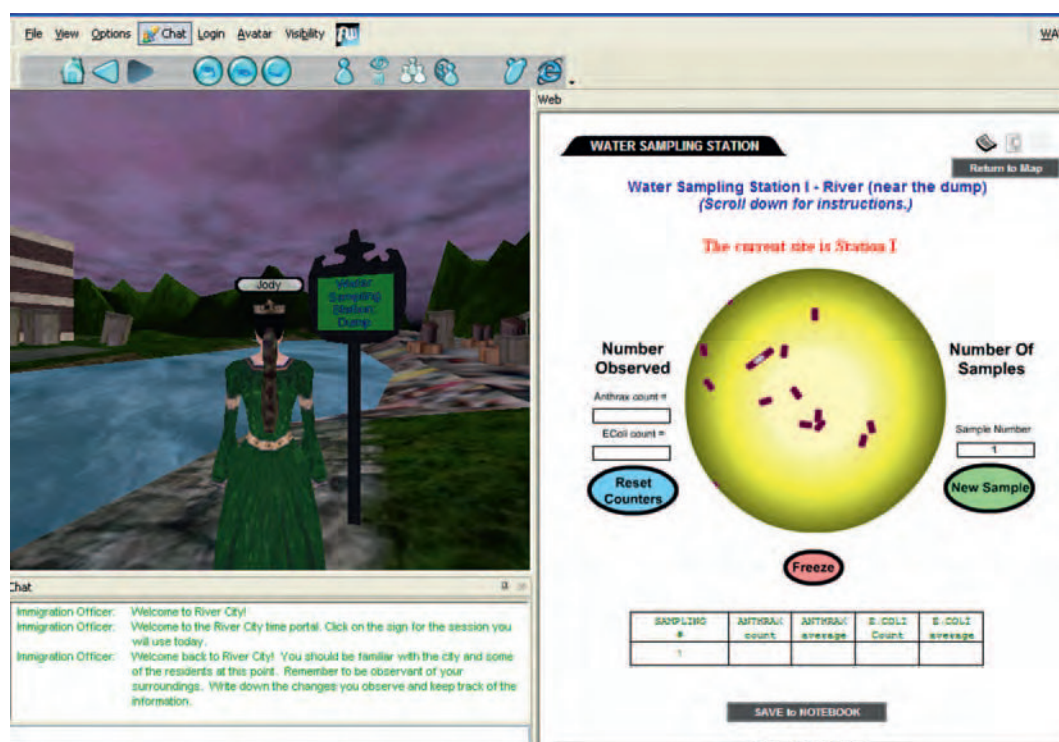


Figure 5 Taking a water sample with the virtual microscope.

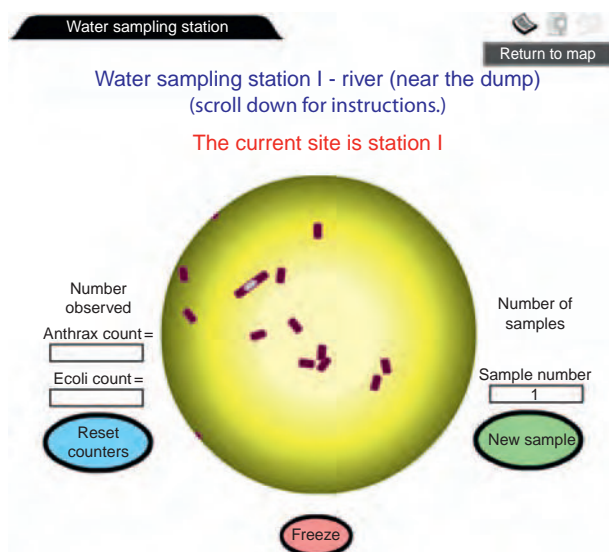


Figure 6 Close-up of microscope. Students click 'Freeze' and count the number of *E. coli* and anthrax in the water.

and skills because omission requires unlearning on the part of adults. A major challenge in professional development is helping teachers, policymakers, and local communities unlearn the beliefs, values, assumptions, and cultures underlying schools' standard-operating practices, such as 45-min class periods that allow insufficient time for all but superficial forms of active learning by students. Altering deeply ingrained and strongly reinforced rituals of schooling takes more than the superficial interchanges typical in make-and-take professional development or school board meetings. Intellectual, emotional, and social support is essential for unlearning and for transformational relearning that can lead to deeper behavioral changes to create next-generation educational practices. Educators, business executives, politicians, and the general public have much to unlearn if twenty-first-century understandings are to assume a central place in schooling.

Acknowledgments

The author's research team's studies of virtual reality and MUEs are funded by the National Science Foundation, and research on augmented reality is funded by the US Department of Education. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation or the US Department of Education.

See also: Technology Supports for Acquiring Inquiry Skills; Technology Supports for Lifelong Learning.

Bibliography

- American Association of Colleges and Universities (2007). *College Learning for the New Global Century*. Washington, DC: AACU.
- Aubert, J. and Reiffers, J. (2004). *Knowledge Economies in the Middle East and North Africa: Toward new Development Strategies*. Washington, DC: World Bank.
- Business-Higher Education Forum (2005). *A Commitment to America's Future: Responding to the Crisis in Mathematics & Science Education*. Washington, DC: Author.
- Card, D. and DiNardo, J. E. (2002). Skill-biased technological change and rising wage inequality: Some problems and puzzles. *Journal of Labor Economics* 20(4), 733–783.
- Chang, H. (2003). Market, the state and institutions in economic development. In Chang, H. (ed.) *Rethinking Development Economics*, pp 41–60. London: Anthem Press.
- Colwell, J. (2007). Needs met through computer game play among adolescents. *Personality and Individual Differences* 43(8), 2072–2082.
- Dede, C. (2009). Immersive interfaces for engagement and learning. *Science* 323(5910), 66–69.
- Dede, C., Korte, S., Nelson, R., Valdez, G., and Ward, D. (2005). *Transforming Education for the 21st Century: An Economic Imperative*. Chicago, IL: Learning Point Associates.
- Engeström, Y. and Middleton, D. (eds.) (1996). *Cognition and Communication at Work*. Cambridge, UK: Cambridge University Press.
- Everett, A. and Watkins, S. C. (2008). The power of play: The portrayal and performance of race in video games. In Salen, K. (ed.) *The Ecology of Games: Connecting Youth, Games, and Learning*, pp 141–166. Cambridge, MA: Massachusetts Institute of Technology Press.
- Friedman, T. L. (2005). *The World is Flat: A Brief History of the Twenty-First Century*. New York: Farrar, Straus, and Giroux.
- Friedman, T. L. (2008). *Hot, Flat, and Crowded: Why We Need a Green Revolution and How It Can Improve America*. New York: Farrar, Strauss, and Giroux.
- Jenkins, H., Clinton, K., Purushotma, R., Robinson, A. J., and Weigel, M. (2006). *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*. Chicago, IL: The MacArthur Foundation.
- Kloper, E. (2008). *Augmented Reality: Research and Design of Mobile Educational Games*. Cambridge, MA: MIT Press.
- Leu, D. J., Zawilinski, L., Castek, J., et al. (2007). What is new about the new literacies of online reading comprehension? In Rush, L., Eakle, J., and Berger, A. (eds.) *Secondary School Literacy: What Research Reveals for Classroom Practices*, pp 37–68. Urbana, IL: National Council of Teachers of English.
- Levy, F. and Murnane, R. J. (2004). *The New Division of Labor: How Computers are Creating the Next Job Market*. Princeton, NJ: Princeton University Press.
- National Academy of Science (2006). *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*. Washington, DC: National Academy Press.
- Organization for Economic Co-operation and Development (OECD) (2004). *Innovation in the Knowledge Economy: Implications for Education and Learning*. Paris: Author.
- Pitaru, A. (2008). E is for everyone: The case for inclusive game design. In Salen, K. (ed.) *The Ecology of Games: Connecting Youth, Games, and Learning*, pp 67–88. Cambridge, MA: Massachusetts Institute of Technology Press.
- Schwartz, D. L., Bransford, J. D., and Sears, D. (2005). Efficiency and innovation in transfer. In Mestre, J. (ed.) *Transfer of Learning from a*

Modern Multidisciplinary Perspective, pp 1–51. Greenwich, CT: Information Age Publishing.

Sternberg, R. J. and Zhang, L. -F. (eds.) (2001). *Perspectives on Thinking, Learning, and Cognitive Styles*. Mahwah, NJ: Erlbaum.

Further Reading

Barab, S., Thomas, M., Dodge, T., Carteaux, R., and Tuzun, H. (2004). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development* **53**(1), 86–108.

Clarke, J. and Dede, C. (2009). Robust designs for scalability. In Moller, L., Huett, J. B., and Harvey, D. M. (eds.) *Learning and Instructional Technologies for the 21st Century: Visions of the Future*, pp 27–48. New York: Springer.

Commission on Behavioral and Social Sciences and Education, National Research Council (2000). In Bransford, J. D., Brown, A. L., and Cocking, R. R. (eds.) *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*, Washington, DC: National Academies Press.

Dede, C. (2008). Theoretical perspectives influencing the use of information technology in teaching and learning. In Voogt, J. and Knezek, G. (eds.) *International Handbook of Information Technology in Education*, pp 43–62. New York: Springer.

Educational Testing Service (2002). *Digital Transformation: A Framework for ICT Literacy*. Princeton, NJ: Author.

Leitch Review of Skills (2006). *Prosperity for All in the Global Economy – World Class Skills*. London: Her Majesty's Treasury.

Neulight, N., Kafai, Y. B., Kao, L., Foley, B., and Galas, C. (2008). Children's learning about infectious disease through participation in a virtual epidemic. *Journal of Science Education and Technology* **16**(1), 47–58.

North Central Regional Educational Laboratory and the Metiri Group (2003). *enGauge 21st Century Skills: Literacy in the Digital Age*. Naperville, IL: NCREL.

Partnership for 21st Century Skills (2006). *A State Leader's Action Guide to 21st Century Skills: A New Vision for Education*. Tucson, AZ: Author.

Steinkuehler, C. A. (2006). Massively multiplayer online video gaming as participation in a discourse. *Mind, Culture, and Activity* **13**(1), 38–52.

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. New York: Cambridge University Press.

Relevant Websites

<http://isites.harvard.edu> – Handheld Augmented Reality Project, Harvard University.

<http://www.21stcenturyskills.org> – Partnership for 21st Century Skills.

<http://muve.gse.harvard.edu/rivercityproject> – The River City Project.

Technology Supports for Acquiring Inquiry Skills

T de Jong, University of Twente, Enschede, The Netherlands

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The idea of inquiry, or scientific discovery, as a learning approach has a long history (Bruner, 1961; Dewey, 1938) and has frequently been promoted as an educational reform. Learning theorists and psychologists such as Piaget and Papert have endorsed the idea of inquiry learning where learners draw on their own experiences and prior knowledge to discover what needs to be learned. Leading organizations have acknowledged inquiry as an important aspect of schooling. For example, in 1989, the American Association for the Advancement of Science (AAAS) stated that: "In learning science, students need time for exploring, for making observations, for taking wrong turns, for testing ideas, for doing things over again; time for building things, calibrating instruments, collecting things, constructing physical and mathematical models for testing ideas; time for learning whatever mathematics, technology, and science they may need to deal with the questions at hand; time for asking around, reading, and arguing; time for wrestling with unfamiliar and counterintuitive ideas and for coming to see the advantage in thinking in a different way" (American Association for the Advancement of Science, 1989). This has led to inquiry becoming an explicit part of the curriculum for many nations, including the United States (National Research Council, 1996).

Inquiry learning has a number of potential learning outcomes. First, inquiry learning can be used to learn about a domain, for example, physics, that is the subject of the inquiry process. Learning a domain through inquiry has been studied extensively (see e.g., de Jong, 2006a; de Jong and van Joolingen, 1998; Linn *et al.*, 2006). Second, engagement in the inquiry process means that inquiry or scientific thinking skills can be acquired (White and Frederiksen, 2005). Third, through inquiry, students can acquire positive attitudes toward the domain or toward science in general (Cognition and Technology Group at Vanderbilt, 1992).

In this article, the focus is on acquiring inquiry skills. As a further focus technological environments are used (i.e., computer simulations) as the contexts for learning. The article starts with an overview of inquiry learning processes. Next, a short overview of technological environments suited for inquiry learning is presented. Then, scaffolds that can be incorporated in these technology-based environments to support acquisition of inquiry skills are described. Finally, a short research agenda is discussed.

The Inquiry Process

Scientific literacy is an important goal in many curricula. While there is not one agreed-upon definition of scientific literacy, it embraces aspects such as, for example, understanding the nature of science, as well as ideas of how the scientific process works (Millar, 2006). Inquiry is the central process of scientific advancement. The National Research Council (NRC), in 1996, defined inquiry as:

Inquiry is a multifaceted activity that involves making observations; posing questions; examining books and other sources of information to see what is already known; planning investigations; reviewing what is already known in light of experimental evidence; using tools to gather, analyze, and interpret data; proposing answers, explanations, and predictions; and communicating the results. Inquiry requires identification of assumptions, use of critical and logical thinking, and consideration of alternative explanations. (National Research Council, 1996)

This description lists a large set of processes that constitute inquiry learning. De Jong (2006b) presented a number of processes that encompass the many facets of inquiry as formulated by the NRC: orientation, hypothesis generation, experimentation (experiment design, prediction, data interpretation), drawing a conclusion, and making an evaluation. Similar processes constitute the inquiry cycle specified by others. White and Frederiksen (2005), for example, distinguish development of a research question, generation of hypotheses, investigation design, analyzing data, model creation, and evaluation. Besides these transformative processes, inquiry also requires regulative processes such as planning and monitoring (de Jong and van Joolingen, 1998). Zimmerman (2007), for example, points to the importance of record keeping as an activity within inquiry learning. Inquiry skills can be seen as a distinct skill set, apart from general metacognitive skills (Veenman *et al.* 2002). In their study of the economics of an intelligent tutoring system, Smithtown *et al.* (1990) concluded that general intelligence is certainly a component of discovery learning, but that specific scientific behaviors account for considerably more variance (p. 71). Schunn and Anderson (1999) showed that discovery skills are distinct from general reasoning ability and that they can be learned and transferred from one research area to another.

What Constitutes Skillful Inquiry?

Possessing adequate inquiry skills means knowing the overall structure of the inquiry process and its constituent processes and also knowing how to execute these processes. In an earlier work, de Jong and van Joolingen (1998) outlined a number of difficulties that students have with inquiry processes; several of these were also recently identified in an overview of inquiry in children (Zimmerman, 2007).

Hypothesis generation is probably one of the most central processes in inquiry learning. Shute and Glaser (1990) found that learners with hypothesis-driven behavior have better learning outcomes. De Jong and Van Joolingen (1998) described several difficulties learners generally encounter in hypothesis generation. These difficulties have two aspects. First, learners may have problems simply with forming hypotheses (Njoo and de Jong, 1993), which may result in conducting experiments that are not guided by a hypothesis. Second, they may be unable to adapt their hypotheses on the basis of experimental evidence (Chinn and Brewer, 2001; Klahr and Dunbar, 1988; Kuhn *et al.*, 1995). Difficulties with coordinating hypotheses and experimental data may be more pronounced for children than they are for adults (Kuhn *et al.*, 1988; Wilhelm *et al.*, 2005).

Experiment design is another crucial aspect in inquiry learning. The most-valued experimental strategy is the so-called vary-one-thing-at-a-time (VOTAT) or control-of-variables (CVS) strategy (Chen and Klahr, 1999). Learners using this strategy focus on the effects of one variable. The alternative of holding one variable constant and varying all other variables leads to confounded experiments. Though there is considerable inter- and also intra-individual variation in the use of strategies (Schauble, 1996), there is evidence that designing proper experiments is a trainable skill (Kuhn *et al.*, 2000) and that there is age-related and experience-related development in the use of controlled experiments (Zimmerman, 2007).

A third important aspect of skillful inquiry is the ability to regulate the learning process. This concerns processes such as planning and monitoring. In inquiry-learning environments, which have by nature a relatively open character, a multitude of routes is open to students, stressing the importance of an overall plan of where to go. Since results from inquiry do not organize themselves, careful monitoring of the learning process is also necessary. Research has shown that active regulation leads to higher learning gains (Azevedo *et al.*, 2004). Surprisingly, however, students do not seem very active in regulating their inquiry process (Manlove and Lazonder, 2004).

A number of studies have identified characteristic differences between successful and unsuccessful students. Lavoie and Good (1988) studied the cognitive processes of 14 high school students learning about water pollution with a simulation. One of the comparisons they made

concerned the cognitive processes of good and poor predictors (students received a prediction test after the simulation). The main differences were that compared to poor predictors, good predictors used more abstract reasoning (e.g., introducing qualitative scales for variables), worked in a more systematic way (e.g., changing only one variable at a time, more often returning variables to baseline conditions, and more frequently looking for worst and best conditions), made notes during the exploration, were able to find more complex relations (bidirectional and ratio relations), and showed a high interest and motivation (e.g., by persisting to complete the learning sequence). Thus, it seems that successful predictors conduct more informative experiments (e.g., by controlling extraneous variation, changing one variable at a time, and returning variables to baseline conditions) and are better able to interpret the data (e.g., inferring regularities in the data, finding more complex relations). In a similar vein, Schauble *et al.* (1991) compared the discovery behavior of learners who were successful in learning with a simulation on electricity with those who were unsuccessful. They found that both groups of students were equally active in the simulation but that good learners performed better in: "... the class of evidence generation (controlling extraneous variation), evidence interpretation (generating and evaluating alternative hypotheses, inferring regularities in the data, producing sufficient evidence to support a hypothesis), data management (systematic data recording), and planning (developing plans that are goal oriented rather than procedure oriented)" (p. 223). Klahr *et al.* (1993) observed the discovery behavior of learners in the Big Trak environment. They identified a number of successful heuristics for hypothesis generation and experiment design, such as designing simple experiments to enable monitoring, designing experiments that give characteristic results, focusing on a single dimension of a hypothesis, and exploiting surprising results.

Traditional environments for inquiry learning include practicum and laboratory experiences (see Minstrell and Kraus, 2005, for a series of examples), but today there are opportunities to use technology-based learning environments that are specifically geared toward inquiry learning and that may also contain specific scaffolds for inquiry-learning processes.

Technology-Enhanced Inquiry-Learning Environments

Technological environments that are specifically suitable for inquiry learning include simulations, games, and adventures, remote labs, data sets, and hypermedia environments. All these environments are based on the fact that the knowledge that needs to be acquired is not presented explicitly to learners but must be inferred from the environment.

Students, for example, have to infer the relations between variables by manipulating them and observing their behavior; this relation is not explicitly given in the environment. An exception to this might be hypermedia environments (e.g., WISE; see Linn *et al.*, 2004b) where information often is partly offered in a more direct way. In all the other environments, students must make inferences themselves to see the information that is in the environment.

A basic form of inquiry learning is the situation in which learners perform experiments by manipulating values of variables and observe outcomes of their experiments. This type of inquiry can take place in physical laboratories, such as the wooden ramp in the work of Klahr and Nigam (2004) or in technology-enhanced learning (TEL) environments. TEL environments may mimic the real laboratory (van Joolingen *et al.*, 2005) or provide learners with virtual environments based on simulations (Blake and Scanlon, 2007; de Jong and van Joolingen, 1998), or, in exceptional cases, include a combination of real and virtual offerings (Zacharia, 2007). The advantages of TEL environments are that they provide students with a safe environment in which experimentation can be done (and redone) quite effortlessly. In addition, technology enables the inclusion of scaffolds for students (see later). In studies, where a comparison between physical and virtual learning environments are made, there is an advantage for students in virtual environments on the acquisition of conceptual knowledge (Zacharia and de Jong, in preparation).

Zimmerman (2007: 217) lists several studies that show that just performing inquiry tasks leads to more skillful execution of inquiry processes. It is, however, now clear that effective learning is only realized through a scaffolded or supported inquiry process (Mayer, 2004). These effects have principally been found for domain knowledge, but effects on inquiry skills are also reported (Linn *et al.*, 2006).

Inquiry Scaffolds and the Acquisition of Inquiry Skills

Technology-enhanced inquiry environments provide the opportunity to include cognitive scaffolds in the software. Overviews of such scaffolds appeared in de Jong and van Joolingen (1998), Quintana *et al.* (2004), Linn *et al.* (2004a), and de Jong (2006b). This section explores which technological scaffolds are available for pivotal inquiry processes.

Hypothesis generation is one of the central processes in inquiry learning, but it also seems to be the most difficult one to support. Many learning environments provide students with assignments or at least issues that need investigation (de Jong, 2006b). This seems a straightforward and effective strategy since research has shown the benefits of providing students with readymade

hypotheses (Njoo and de Jong, 1993). The Inquiry Island environment (e.g., White and Frederiksen, 2005) includes a specific advisor (called Ingrid Inventor) who helps students brainstorm new ideas. The process of creating hypotheses can be supported by giving students a so-called hypothesis scratchpad – a tool that helps students to create testable hypotheses (Shute and Glaser, 1990; van Joolingen and de Jong, 2003). Basically, a hypothesis scratchpad provides students with elements of hypotheses (variables, relations, and conditions) that they can combine to form a testable hypothesis. So far, however, studies have been unable to demonstrate advantages of the use of hypothesis scratchpads (see e.g., Gijlers and de Jong, 2005).

Scaffolding the design of experiments is frequently done by providing students with experimentation hints. These hints are often heuristics for performing experiments (Veermans *et al.*, 2006). The most well known of these heuristics is the CVS or the VOTAT strategy (Klahr, 2005; Klahr and Nigam, 2004; Kuhn, 2005; Kuhn and Dean, 2005). Veermans *et al.* (2006) present a longer list of other heuristics, such as simplified values, design experiments that give characteristic results, and use of extreme values. Kuhn and Dean (2005) found that simply suggesting that students focus on one variable resulted in this behavior even when the hints were no longer present. In addition, students following this strategy were also better in drawing inferences from experimental data. Hulshof and de Jong (2006), however, found an effect of this type of hint only on domain knowledge and not on inquiry skills. Teachers can also give direct instruction on how to design experiments. Klahr and Nigam (2004) found that students who received an explanation of the CVS strategy along with examples of good and poor experiments were more successful in acquiring CVS skills than were students who received no instruction. Keselman (2003) found that a group of students who saw models of how to manipulate variables and make predictions improved most on domain knowledge and experiment design skills. Beishuizen *et al.* (2003) also found that training sixth graders on designing CVS experiments and inferencing skills gave students an advantage in finding interaction effects.

Planning of activities is supported at a higher level by presenting an overview of the inquiry cycle (Manlove *et al.*, 2007; White *et al.*, 2002). Manlove *et al.* (2007) found that students who were supported this way wrote better lab reports but constructed poorer models than students who did not receive this guidance. White *et al.* (2002) report better use of notebook facilities indicating understanding of inquiry by students working with Inquiry Island. White and Frederiksen (2005) report that students who learned with Inquiry Island showed higher gains on an inquiry test than students who received traditional instruction. At a more fine-grained level of planning, students can receive assignments that give

them short-term goals for action. For monitoring, students can be offered software scaffolds that help them save and reorganize experiments (van Joolingen and de Jong, 2003).

Conclusions and Research Issues

Research into inquiry skills reflects the complex nature of the topic. Inquiry as such is a comprehensive set of activities, and studies differ in the aspects of inquiry under investigation. In addition, the subjects' age levels, domains involved, and measuring techniques differ widely across studies. Finally, definitions of key terms such as discovery learning and direct instruction are certainly not unequivocal (Zimmerman, 2007). Despite this, some general conclusions can be drawn.

Most studies of inquiry learning have focused on the acquisition of domain knowledge. Studies that assessed the acquisition of inquiry skills concentrated mainly on experiment design. Combining an inquiry task with direct hints or instruction on how to perform informative experiments is clearly an effective approach for improving students' experiment-design skills. Several scaffolding methods that give students overall inquiry templates to work with exist for the support of regulative (planning and monitoring) skills. Following these inquiry templates leads to better inquiry skills. On the other hand, not many tools are available for the hypothesis-generation aspect of inquiry, and the effectiveness of existing tools is not established.

One development that the field would profit from would be a more widely accepted way of measuring inquiry skills. Currently, this ranges from performing a short inquiry cycle (White and Frederiksen, 2005), experimenting in a domain-free specifically designed experimentation environment (Hulshof *et al.*, 2005), creating the design of an experiment (Klahr and Nigam, 2004), and evaluating existing experiment plans (Keselman, 2003) to general and domain-specific tests containing various types of questions (Linn *et al.*, 2006). The character of the inquiry test may well partially influence the experimental outcomes.

Two research developments seem promising for a more flexible and full scaffolding of the inquiry process. One is more adaptive scaffolding where scaffolds depend on the actual experimentation behavior of students. One of the first attempts toward this goal can be found in Veermans *et al.* (2006). The second direction is collaborative inquiry where collaborating learners help to scaffold and advance each other's inquiry skills (Gijlers and de Jong, 2009).

See also: Conceptions of Technology Literacy and Fluency; Technological Supports for Acquiring Twenty-First-Century Skills.

Bibliography

- American Association for the Advancement of Science (1989). *Science for All Americans*. New York: Oxford University Press.
- Azevedo, R., Guthrie, J. T., and Seibert, D. (2004). The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia. *Journal of Educational Computing Research* **30**, 87–111.
- Beishuizen, J., Wilhelm, P., and Schimmel, M. (2003). Computer-supported inquiry learning: Effects of training and practice. *Computers and Education* **42**, 389–402.
- Blake, C. and Scanlon, E. (2007). Reconsidering simulations in science education at a distance: Features of effective use. *Journal of Computer Assisted Learning* **23**(6), 491–502.
- Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review* **31**, 21–32.
- Chen, Z. and Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development* **70**, 1098–1120.
- Chinn, C. A. and Brewer, W. F. (2001). Models of data: A theory of how people evaluate data. *Cognition and Instruction* **19**, 323–393.
- Cognition and Technology Group at Vanderbilt (1992). The Jasper series as an example of anchored instruction: Theory, program, description, and assessment data. *Educational Psychologist* **27**, 291–315.
- de Jong, T. (2006a). Computer simulations – technological advances in inquiry learning. *Science* **312**, 532–533.
- de Jong, T. (2006b). Scaffolds for computer simulation based scientific discovery learning. In Elen, J. and Clark, R. E. (eds.) *Dealing with Complexity in Learning Environments*, pp 107–128. London: Elsevier Science Publishers.
- de Jong, T. and van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research* **68**, 179–202.
- Dewey, J. (1938). *Logic: The Theory of Inquiry*. New York: Holt.
- Gijlers, H. and de Jong, T. (2005). The relation between prior knowledge and students' collaborative discovery learning processes. *Journal of Research in Science Teaching* **42**, 264–282.
- Gijlers, H. and de Jong, T. (2009). Sharing and confronting propositions in collaborative inquiry learning. *Cognition and Instruction* **27**, 239–268.
- Hulshof, C. D. and de Jong, T. (2006). Using just-in-time information to support scientific discovery learning about geometrical optics in a computer-based simulation. *Interactive Learning Environments* **14**, 79–94.
- Hulshof, C. D., Wilhelm, P., Beishuizen, J. J., and van Rijn, H. (2005). FILE: A tool for the study of inquiry learning. *Computers in Human Behavior* **21**, 945–956.
- Keselman, A. (2003). Supporting inquiry learning by promoting normative understanding of multivariable causality. *Journal of Research in Science Teaching* **40**, 898–921.
- Klahr, D. (2005). Early science instruction. Addressing fundamental issues. *Psychological Science* **16**, 871–872.
- Klahr, D. and Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science* **12**, 1–48.
- Klahr, D., Fay, A. L., and Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology* **25**, 11–146.
- Klahr, D. and Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science* **15**, 661–668.
- Kuhn, D. (2005). What needs to be mastered in mastery of scientific method? *Psychological Science* **16**, 873–874.
- Kuhn, D., Amsel, E., and O'Loughlin, M. (1988). *The Development of Scientific Thinking Skills*. San Diego, CA: Academic Press.
- Kuhn, D., Black, J., Keselman, A., and Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction* **18**, 495–523.
- Kuhn, D. and Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science* **16**, 866–870.
- Kuhn, D., Garcia-Mila, M., Zohar, A., and Andersen, C. (1995). Strategies of Knowledge Acquisition (*Monographs of the Society for*

- Research in Child Development*), vol. **60**(4), pp 1–157. Hillsdale, NJ: Erlbaum.
- Lavoie, D. R. and Good, R. (1988). The nature and use of predictions skills in a biological computer simulation. *Journal of Research in Science Teaching* **25**, 335–360.
- Linn, M. C., Bell, P., and Davis, E. A. (2004a). Specific design principles: Elaborating the scaffolded knowledge integration framework. In Linn, M., Davis, E. A., and Bell, P. (eds.) *Internet Environments for Science Education*, pp 315–341. Mahwah, NJ: Erlbaum.
- Linn, M. C., Davis, E. A., and Bell, P. (eds.) (2004b). *Internet Environments for Science Education*. Mahwah, NJ: Erlbaum.
- Linn, M. C., Lee, H. -S., Tinker, R., Husic, F., and Chiu, J. L. (2006). Teaching and assessing knowledge integration in science. *Science* **313**, 1049–1050.
- Manlove, S. and Lazonder, A. (2004). Self-regulation and collaboration in a discovery learning environment. *Paper Presented at the First Meeting of the EARLI-SIG on Metacognition*. Amsterdam: The Netherlands.
- Manlove, S., Lazonder, A. W., and de Jong, T. (2007). Software scaffolds to promote regulation during scientific inquiry learning. *Metacognition and Learning* **2**, 141–155.
- Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist* **59**, 14–19.
- Millar, R. (2006). Twenty first century science: Insights from the design and implementation of a scientific literacy approach in school science. *International Journal of Science Education* **28**, 1499–1521.
- Minstrell, J. and Kraus, P. (2005). Guided inquiry in the science classroom. In Donovan, M. S. and Bransford, J. D. (eds.) *How Students Learn*, pp 475–513. Washington, DC: The National Academies Press.
- National Research Council (1996). *National Science Education Standards*. Washington, DC: National Academy Press.
- Njoo, M. and de Jong, T. (1993). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *Journal of Research in Science Teaching* **30**, 821–844.
- Quintana, C., Reiser, B. J., Davis, E. A., et al. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences* **13**, 337–387.
- Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology* **32**, 102–119.
- Schauble, L., Glaser, R., Raghavan, K., and Reiser, M. (1991). Causal models and experimentation strategies in scientific reasoning. *Journal of the Learning Sciences* **1**, 201–238.
- Schunn, C. D. and Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science* **23**, 337–370.
- Shute, V. J. and Glaser, R. (1990). A large-scale evaluation of an intelligent discovery world: Smithtown. *Interactive Learning Environments* **1**, 51–77.
- van Joolingen, W. R. and de Jong, T. (2003). SimQuest: Authoring educational simulations. In Murray, T., Blessing, S., and Ainsworth, S. (eds.) *Authoring Tools for Advanced Technology Educational Software: Toward Cost-Effective Production of Adaptive, Interactive, and Intelligent Educational Software*, pp 1–31. Dordrecht: Kluwer Academic Publishers.
- van Joolingen, W. R., de Jong, T., Lazonder, A. W., Savelsbergh, E., and Manlove, S. (2005). Co-Lab: Research and development of an on-line learning environment for collaborative scientific discovery learning. *Computers in Human Behavior* **21**, 671–688.
- Veenman, M. V. J., Prins, F. J., and Elshout, J. J. (2002). Initial inductive learning in a complex computer simulated environment: The role of metacognitive skills and intellectual ability. *Computers in Human Behavior* **18**, 327–341.
- Veermans, K. H., van Joolingen, W. R., and de Jong, T. (2006). Using heuristics to facilitate scientific discovery learning in a simulation learning environment in a physics domain. *International Journal of Science Education* **28**, 341–361.
- White, B. Y. and Frederiksen, J. (2005). A theoretical framework and approach for fostering metacognitive development. *Educational Psychologist* **40**, 211–223.
- White, B. Y., Frederiksen, J., Frederiksen, T., et al. (2002). Inquiry Island: Affordances of a multi-agent environment for scientific inquiry and reflective learning. *Paper Presented at the Fifth International Conference of the Learning Sciences (ICLS)*, Mahwah: NJ.
- Wilhelm, P., Beishuizen, J. J., and Van Rijn, H. (2005). Studying inquiry learning with FILE. *Computers in Human Behavior* **21**, 933–943.
- Zacharia, Z. C. (2007). Comparing and combining real and virtual experimentation: An effort to enhance students' conceptual understanding of electric circuits. *Journal of Computer Assisted Learning* **23**(2), 120–132.
- Zacharia, Z. C. and de Jong, T. (in preparation). Should physical and virtual manipulatives be combined or used alone when understanding of physics concepts is the target goal?
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review* **27**, 172–223.

Technology Supports for Acquiring Mathematics

M J Nathan, University of Wisconsin-Madison, Madison, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Technology has become essential to the practice of mathematics. Technology allows some areas of mathematics to flourish, such as the close relationships between fractal geometry with computer graphics, and statistics with computer programs (Ben-Zvi and Garfield, 2004; Frame and Mandelbrot, 2002). Technology also plays a special role in the learning of mathematics. Technology supports for mathematics education largely reflect the form and shifts of underlying theories of learning and intellectual behavior. Information processing (IP) theory and theories of cognitive skill acquisition and concept learning have been dominant influences in mathematics education for the last half century, and these are presented in the first section titled 'Skills and concept learning.' While these traditional cognitive perspectives on knowledge and learning continue to play a significant role in governing the types of technological resources for mathematics education (as well as shaping curriculum, instruction, and assessment practices more broadly), increasingly, designs of educational technology are being influenced by emerging theories of learning and practice. The second section titled 'Mathematical discovery' focuses on technologies that support learning by discovery. In the third section titled 'Collaborative problem solving,' collaborative learning technologies are presented that often draw on situated, social, and cultural perspectives of learning and constructivist views of knowledge. The fourth section titled 'Embodied cognition' reviews some of the newest technologies in craft and fabrication, inspired by emerging views of embodied cognition. Some systems appear in multiple places within this simple taxonomy. In the final section titled 'Challenges facing technology supports for acquiring mathematics,' some of the persistent challenges facing technologies for mathematical learning are discussed.

Skills and Concept Learning

There is a rich history of tools specifically designed for mathematics instruction, including formal systems of manipulatives such as Dienes' blocks, pattern boards, and Cuisenaire rods. Manipulatives are concrete objects that are designed to help students learn math concepts by representing the quantities, operations, and relations between quantities without requiring that the learner

use or comprehend the written (i.e., formal) representations of the same concepts (Uttal, 2003).

Generally, the use of math manipulatives as educational aids is predicated on a view of learning as the acquisition of skills and abstractions (Chao *et al.*, 2000). The mental tool view frames learning as the acquisition of skills and symbolic structures that parallel the physical states and actions of the objects. Learning from this perspective typically favors highly structured and consistent practice to enable the acquisition and speedup of procedures, which may eventually become automated. In the abstraction view, math learning is facilitated through generalization across a range of varied experiences that all model a common concept. This form of learned behavior tends to exhibit slower and more deliberate responses that are highly accurate, along with more frequent use of rudimentary strategies. Even with these trade-offs in mind, the research literature does not consistently show an advantage for manipulatives for math learning among primary grade students (e.g., Hiebert, 1989; Sowell, 1989; Uttal *et al.*, 1997).

Computer-based manipulatives and other instructional tools have emerged to capture the concrete qualities of materials along with the added control and flexibility of digital media. One of the earliest and most influential was Logo, which can be regarded as the forerunner of computers as microworlds, and objects-to-think-with. In a somewhat parallel fashion, graphing calculators emerged as hand-held tools for math activity and math instruction. These early forms also spawned computer algebra systems and more free-form modeling tools, such as Geometric Supposer and Geometer's Sketchpad.

Hand-Held Graphing Calculators

The first graphing calculator was introduced by Casio in 1985. However, graphing calculators exhibited a much greater influence on classroom learning in the early 1990s, with contributions from Hewlett-Packard and Texas Instruments (TI). The TI-82 (released 1993) and the ubiquitous TI-83 (1996; see **Figure 1**) transformed secondary and tertiary math education by putting into students' hands an affordable, portable, and accessible device that allowed them to analyze, program, and visualize mathematical procedures and structures. By the year 2000, over 80% of high-school mathematics teachers in the US who were surveyed, reported using hand-held graphing calculators in their classrooms (Hudson *et al.*, 2002).



Figure 1 Classroom teacher and students using the TI-83-Plus hand-held graphing calculator by Texas Instruments. Photo courtesy of Texas Instruments Education Technology.

Hand-held graphing technology has the potential to change the nature of mathematics instruction and learning, as well as alter the very content of the mathematics that gets taught in schools (Waits and Demana, 1994, 2000). However, reviews of US textbooks exhibit a fairly simplistic use of the technology (Burrill, 2004; Senk *et al.*, unpublished).

Internationally (in studies of Great Britain, Sweden, France, Australia, New Zealand, South Africa, Israel, Netherlands, and the United States), hand-held graphing-technology use generally facilitates learners' concept development and its use is predictive of higher performance gains and measures of problem-solving skills. Those who use graphing calculators show a better understanding of functions, applied problem solving in algebra, and interpreting graphs (Burrill *et al.*, 2002).

Studies suggest that the increased conceptual learning that accompanies use of graphing calculators does not come at the expense of building facility in procedural skills (Burrill, 2004). Frequent use of graphing calculators tends to accompany more graph use and greater flexibility with representations, solution strategies, and reasoning with real data. More broadly, classrooms with graphing calculator use tend to foster a more constructivist climate, with more conjecturing, more frequent use of multiple solutions, and higher levels of discourse than those classrooms with infrequent calculator use. Using graphing technology for nonroutine activities, such as mathematical discovery and complex problem solving, tends to support increased conceptual understanding and higher achievement, while use of technology for routine calculations does not (Dugdale *et al.*, 2004).

Computer Algebra Systems

The next major breakthrough was the incorporation of computer algebra systems (CASs) into calculators. The CASs allow users to perform mathematical operations on

symbolic expressions in support of problem solving, generalization, and reasoning about functions. The CAS made its way into hand-held calculators in the late 1980s with the HP-28, but it was when both Casio (the FX2.0, released in 1996) and TI released new models (the TI-92, released in 1995, and the TI-89 in 1998) that the technology became more affordable and more prevalent among high-school and then middle-school classes.

The promise is that CAS allows learners and teachers to focus on conceptual aspects of expressions and functions (Heid, 1988; Pierce and Stacey, 2007) rather than getting caught up with the mechanics of symbol manipulation, a benefit that may help low-performing students, in particular (Kuzler, 2000). The research does support this view, generally, though, as with other forms of calculators and technology, more generally, these influences are mediated by the types of lessons and instructional approaches. Scholars have shown that calculus students within conceptually oriented classrooms who used CAS demonstrated more conceptual knowledge than those who had skills-oriented lessons (Heid, 1988; Palmiter, 1991). CAS implemented in conjunction with other standards-based practices such as sense making and group discussions also supports college students' understanding and reasoning with symbolic expressions (Keller and Russell, 1997).

Geometric Supposer

In addition to support for skills in computation and symbol manipulation, there are powerful tools such as Geometric Supposer for supporting skills and concept development in geometric reasoning. Geometric Supposer is designed to support exploration and discovery of properties of Euclidean (plane) geometry by providing primitive operations for drawing, analyzing, measuring, and manipulating diagrams. Generalization across cases (induction) is supported by repeating operations on arbitrary exemplars, an experience which may assist students in formulating deductive proof. Geometric Supposer is one of the clearest examples of a system that instantiates many of the ideals of the constructivist philosophy of mathematics education because it allows direct access and construction to otherwise abstract objects, procedures, and concepts.

Although students typically struggle with using geometric diagrams, year-long use of Geometric Supposer has resulted in improvements with diagrams and students' understanding of the objects to which they refer, as well as the variety of ways that diagrams can be viewed and described (Yerushalmy and Chazan, 1990).

Cognitively-Based Tutoring

Some of the most exciting and well-researched technology-based systems come from the class of intelligent tutoring

systems (ITSs). Drawing from cognitive science, artificial intelligence (AI), and expert systems, ITSs typically institute an expert module that contains, usually in rule-based form, knowledge of the skills involved in successful problem-solving behavior, and previously identified bugs, or common errors, in the rules of learners. In the ITS instructional model, the learner engages in problem solving while the expert module tracks learner behavior. Erroneous steps, recognized through pattern matching by the rule-based expert module, are usually met with immediate feedback, in keeping with the theory of the acquisition of cognitive skill. The expert module can also provide context-specific hints, or help, and offer an expert-level solution to the problem (Anderson, 1988).

While there are many ITSs, those derived from Anderson's (1996) ACT theory – the Cognitive Tutors – are of particular interest because of their strong theoretical and empirical support, and because the essential architecture has proven so resilient over time and across a range of mathematical domains and age groups. Cognitive Tutors (Figure 2) provide differentiated instruction in pre-algebra, algebra I and II, geometry, and integrated mathematics to 500 000 students in around 2600 US middle schools and high

schools. Current systems incorporate classroom curricula (often 3 days per week), technology-based instruction and practice (2 days per week), and teacher professional development. Students engage in explanation-based reasoning, as well as goal-directed problem solving.

A number of empirical studies have established the effectiveness of this approach (see Ritter *et al.*, 2007, for a review). For example, Algebra I students in Pittsburgh, PA and Milwaukee, WI, who used the tutor showed superior gains from their peers overall, particularly on performance-based tests of problem solving and uses of multiple representations (Koedinger *et al.*, 1997).

Mathematical Discovery

As educational researchers more fully embraced expanded and alternative views of knowledge and learning such as constructivism (e.g., Cobb *et al.*, 1992) and situated cognition (e.g., Brown *et al.*, 1989), these have been reflected in the designs of educational technologies. New technologies were designed to be student centered, to draw directly from students' own knowledge of mathematical and physical

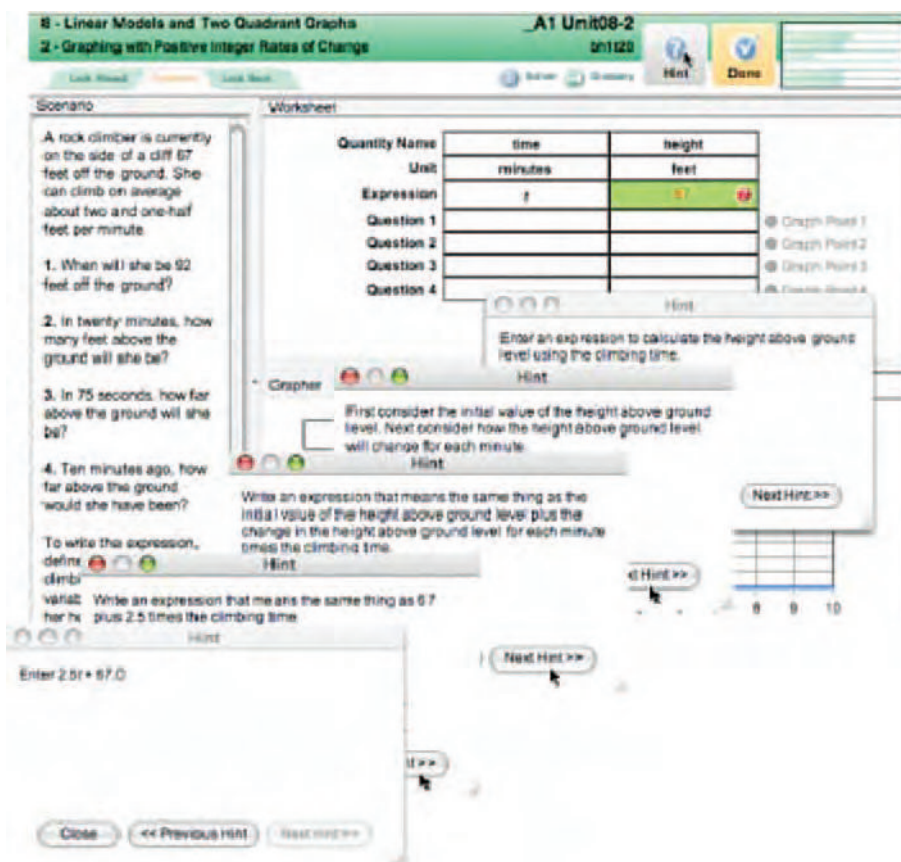


Figure 2 A series of hints, followed by direct instruction provided to a student using the Algebra Cognitive Tutor developed by Carnegie Learning, Inc. Adapted from Koedinger, K. R. and Aleven, V. (in press). Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educational Psychology Review*.

phenomena, and to put more responsibility of discovering and articulating connections between representations and across representation, math concepts, and procedures. Technologies that emphasize the interconnections among representations and the distributed nature of knowledge as it exists between the student and the technology also lead to different types of activities and they reconceptualize the aims of math education as one of learner-centered making meaning.

Logo

Logo (Papert, 1980) is both a computer programming language and a microworld, a designed learning environment to promote mathematical reasoning and problem-solving skills through an innovative process of directing the actions of a mathematical creature called the Logo turtle. The turtle can move forward or backward, stop, and rotate to the left or right, and raise and lower its pen in response to programmed commands. Although the original turtle was a physical robot that ran along the floor or paper, in later versions it was replaced by a graphical turtle on a computer screen.

The Logo environment offers a way for the child to externalize mathematical ideas and procedures and project them onto the actions and properties of the turtle and the Logo programming language (Eisenberg, 2003). Yet, it also becomes an object-to-think-with (Resnick *et al.*, 1996) and has been used to conceptualize many areas of mathematics, including modern algebra and group theory, computer science, cybernetics, as well as Euclidean and non-Euclidean geometry (Abelson and diSessa, 1981).

Logo has long been a tool for doing mathematics and mathematics instruction, and there is a large collection of empirical studies investigating its impact on mathematics learning, teaching, and discovery. For example, fourth graders familiar with Logo programming were better able to apply what they learned and elaborate on their procedural interpretations of geometry concepts than those taught from an inquiry-based approach (Lehrer *et al.*, 1989). In other studies, Logo improved students' use of geometric models in other areas of mathematics, generalization and abstraction of geometric operations, and improved complex reasoning along with more general cognitive skills (Battista and Clements, 1991; Clements and Battista, 1991, 1992; Lehrer and Littlefield, 1993).

As is the case with educational technology, more generally, the effects of Logo have as much to do with the teaching and the engagement of the students, as the technology itself (Kozma, 1991, 1994; though also see Clark, 1983, 1994).

The essential ideas conveyed in Papert's (1980) original work, *Mindstorms*, inspired a broad range of technological designs for learning and instruction, including: StarLogo, which uses concepts of parallel computation to introduce participants to the computational and cognitive aspects of

modeling complex, dynamic systems (e.g., Colella *et al.*, 1999); and the NetLogo Project (reviewed below) at Northwestern University and The University of Texas at Austin (Wilensky, 1999; Wilensky and Stroup, 1999), which supports distributed computing.

Function Probe

Function Probe (Confrey and Mahoney, 1991, 1996) was an early and highly influential technology design for math education to come under the constructivist paradigm. Function Probe integrated a user-friendly calculator with tabular, graphical, and symbolic representations of phenomena to promote the active construction of mathematical understanding for high-school algebra, trigonometry and functions, and high school and college pre-calculus, as well as integrated math and science instruction. Data could be entered in by students or imported using external sensors. Operations on one representation were reflected as changes in other, linked representations, thus promoting representational fluency (Nathan *et al.*, 2002). The system follows one of Kaput's (1989) observations, that mathematical meaning making is actually built upon the ability to translate within and among various representations, and that fundamentally, meaning is based on a "relational semantics" between "linking representations" including internal mental representations and physical systems as well as tables, symbols, and graphs (p. 168). Function Probe also provided a modeling environment that allowed students to articulate and explore their own conceptions of mathematical and physical events as a means toward advanced understanding (Confrey and Doerr, 1994).

ANIMATE

The ANIMATE system presented another alternative to cognitive skill acquisition and the intelligent tutoring system paradigm. It was presented as an unintelligent tutoring system to emphasize that the program contained no expert module and made no attempts at modeling or tracing the knowledge states of the student (Nathan, 1990). Instead, ANIMATE assumed knowledge was distributed among the interactions between the student and the system.

The focus of ANIMATE was on student discovery of quantitative representations for modeling and solving algebra story problems involving systems of equations, including those for distance-rate time (e.g., collision and overtake) problems, combined work, and compound interest. In ANIMATE, a student constructed the algebraic equations that drove an animation of the referent story problem situation (e.g., planes flying at different rates and leaving at different times). Because of the direct causal link between the formal expressions of the algebraic solution (the mathematical symbols and structures to be learned) and the animation (the situation-based meaning

of the mathematical expressions), animated actions that were inconsistent with the student's mental model of the story situation suggested errors in the proposed solution representations, the nature of which were highly constrained by the type of misbehavior. This interlinking of students' thinking and the control of the system is illustrated conceptually in **Figure 3**. Students first proposed and then iteratively debugged their algebraic representations and tested them until an acceptable situation was depicted in the animation. Neither the student nor the system could solve the problems and make meaningful connections on their own. In this way, the intelligence was not localized solely in the computer program or in the learner, but rather from the intelligent interactions needed to bring the mathematics and animation in line with the student's mental model of the story problem.

Students who used ANIMATE learned to reason explicitly about the situations described in typical word problems, and performed better on paper-and-pencil transfer tasks. The ANIMATE users also tended to spontaneously correct their own algebraic errors during problem solving in far greater frequency than control subjects (Nathan *et al.*, 1992; Nathan, 1998).

Dynamic Geometry Systems

Dynamic Geometry computer software such as Geometer's Sketchpad (Finzer and Jackiw, 1998) and Cabri Géomètre (Laborde, 2000), along with Geometric

Supposer mentioned above, offer alternatives to conventional proof-based explorations of geometry by supporting direct manipulation, tracing, and visual forms of thinking without the prior stage of re-representing the intended actions into natural or formal languages.

Classroom observations of Cabri use show that junior high school students in Japan using Cabri are more inclined to explore the propositions and theorems directly through construction and manipulation than under the traditional curriculum, they better visualize the geometric claims, and develop a better sense of what is to be proved (Namura, 1999).

While experimental results are scarce, studies do show that systems such as Cabri and Sketchpad help to promote the proper classroom environment, activities, and forms of interactions that foster deductive reasoning among students (Jones, 2000).

Dynamic Statistics Packages

Several recent dynamic systems allow students to delve into data analysis activities and statistical forms of reasoning without formal knowledge of probability and statistics or the conventions of Cartesian graphs. Data-oriented statistics instruction has moved into mainstream education where it is recognized as a critical methodological tool for reasoning about data with variability, engaging in scientific reasoning, making judgments and evaluating claims, and being an informed citizen. This

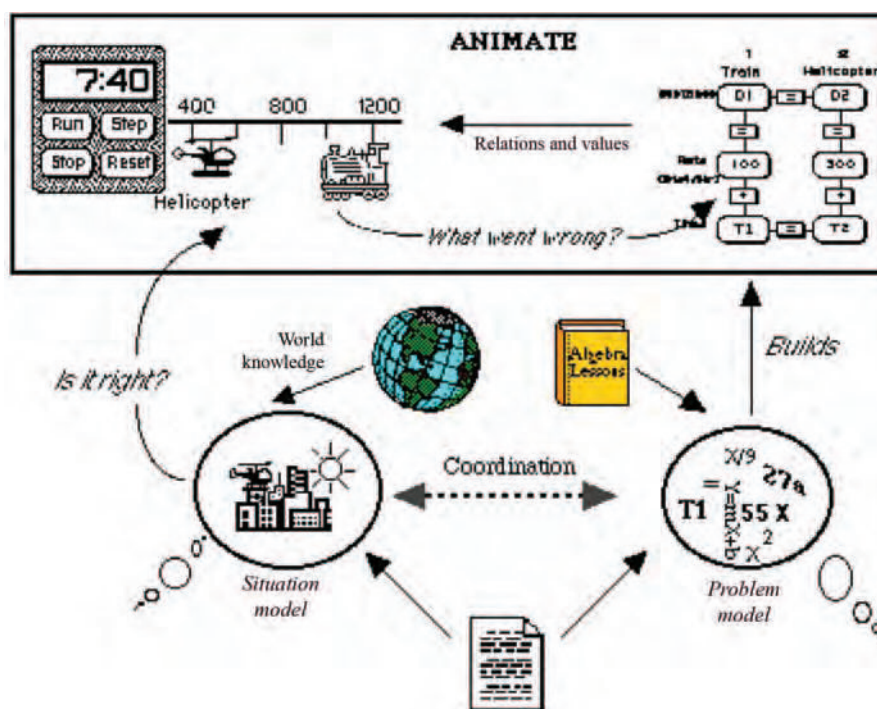


Figure 3 The conceptual underpinnings of learning to model story problems with the ANIMATE system.

reflects the growing appreciation of the need to understand and promote statistical literacy and reasoning more broadly (Ben-Zvi and Garfield, 2004; Cobb, 1993; Rumsey, 2002).

TinkerPlots (Konold, 2002) lets students approach data analysis questions conceptually and in a constructivist manner by providing functionality for organizing and representing data graphically and for progressively constructing and deconstructing graphs in order to develop a more grounded interpretation of the meanings intended by the representations. It also provides several visual methods for displaying variability and co-variation between variables, important ideas for analysis and modeling that are often difficult for students to grasp (Konold and Pollatsek, 2002). In a similar way, Fathom (Erickson, 2000; Finzer and Erickson, 1998) supports discovery of patterns through visually rich exploratory data analysis.

While no formal tests of the effectiveness of these systems compared to other methods have been reported, the approaches they offer are consistent with many of the recent prescriptions for mathematics education offered within current learning theory and standards-based educational reform.

Collaborative Problem Solving

Anchored Instruction uses extended collaborative groups and digital technology to present complex, real-life scenarios – often presented in extended, interactive videos – to situate problems and their solutions in meaningful contexts. The rich and engaging narratives serve as the anchors and give the ensuing mathematical problem-solving activities a meaningful connection to the world. Anchored instruction has been used for a variety of age groups, ability levels, and cultures.

The Adventures of Jasper Woodbury

In math education, the pioneering work on anchored instruction was implemented in *The Adventures of Jasper Woodbury* (CTGV, 1992, 1997). Jasper was developed by the Cognition and Technology Group at Vanderbilt University as a series of 12 open-ended, videodisk-based problem-solving activities that often took groups of elementary-, middle- and high school students several days or more to formulate extended solutions to the series' challenges. Generally, solutions required groups of students to make multiple passes through the information, a variety of plans and computational procedures, and clarifying assumptions. As one might expect, the rich contexts, problem-solving sessions, and clarifying assumptions resulted in complex solutions, no two of which were identical.

Experimental evaluations showed that students using the Jasper program exhibited comparable performance

levels on basic mathematical concepts as matched controls, but superior performance on more complex single- and multistep word problems and multistep planning tasks (CTGV, 1992).

Teaching Enhanced Anchored Mathematics

Enhanced Anchored Mathematics is a form of anchored instruction that situates problems in authentic and meaningful contexts specifically to advance the problem-solving skills of low-achieving and special-education students, adolescents with emotional disabilities, and even re-incarcerated adults (Bottge *et al.*, 2003, 2007; Bottge and Watson, 2002). The anchored activities tie in many of the aspects of traditional industrial arts.

In *Fraction of the Cost* (Bottge *et al.*, 2002; see **Figure 4**), for example, students collaborate to build a life-size skateboard ramp with wood. Problems and information needed to solve the problems, along with irrelevant information, are naturally embedded in the narrative context. Students must use a variety of planning, budgeting, construction, and computational skills. Later, students face transfer problems such as building a working hovercraft.

In empirical research, classroom observers witnessed high and sustained levels of engagement among students of all math abilities, including those with a history of frequent off-task behaviors as they participated in multi-step mathematical reasoning, planning, and problem solving (Bottge *et al.*, 2002, 2003, 2007).

Networked Devices and Participatory Simulations

As technology has progressed to enable networked and distributed interactions, designers have used technology to facilitate students' roles in socially active, life-sized, computational simulations of dynamic systems. Many applications built in earlier decades have recently been reconfigured for networked, interactive uses. Hand-held graphing calculators, discussed earlier in the context of concept learning and skill acquisition, now also serve as one of the central means by which students access distributed networks. Technicians at TI, principally the Classnet Team, are largely responsible for one of the most widely used networking systems, the HubNet hardware, which serves as the central computer architecture through which distributed information is processed and aggregated. Networked systems used by the NetLogo Project at Northwestern University, The University of Texas at Austin, and the SimCalc Project at SRI and the University of Massachusetts at Dartmouth all draw on the TI graphing calculator platform.

Participatory simulations are among the most innovative of these new systems. They provide an individual, first-person perspective from inside the system itself, and

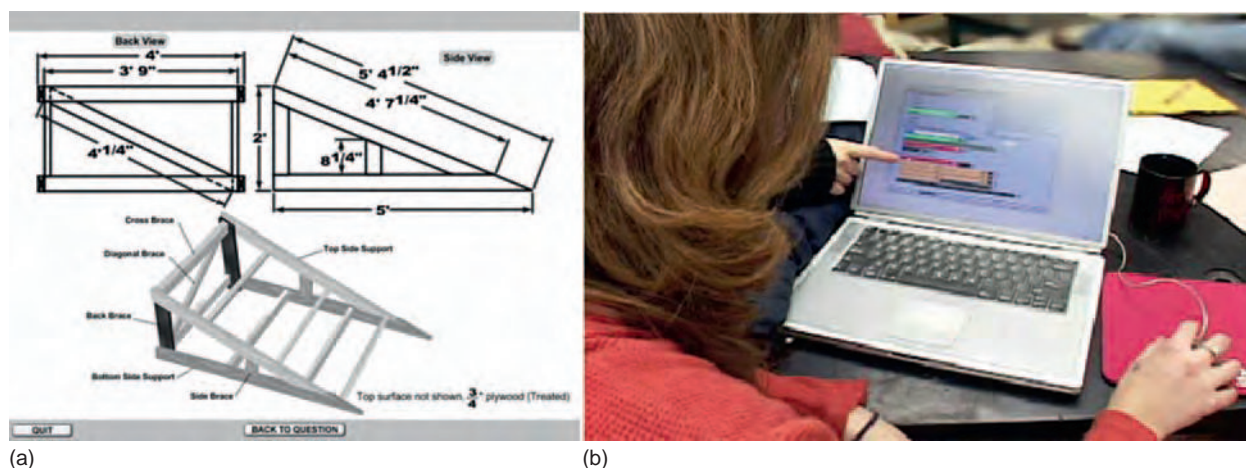


Figure 4 (a) A schematic diagram showing the dimensions needed to build a skateboard ramp for the Fraction of the Cost activity from Teaching Enhanced Anchored Mathematics (TEAM). Adapted from Stephens, A. C., Bottge, B. A., and Rueda, E. (2009). Ramping up on fractions. *Mathematics Teaching in the Middle School*, 14(6), 520–526. (b) A student using the TEAM system. Courtesy of Bottge.

of its emergent behaviors (Colella *et al.*, 1998; Resnick and Wilensky, 1998). The emergent behavior of the system at a macroscopic level and its relation to individual participant's microscopic actions can then become the object of collective discussion and analysis. For example, NetLogo is a programmable modeling environment for simulating complex systems dynamics that occur naturally in biological and social phenomena, such as traffic gridlock and flocking behavior. HubNet allows NetLogo to run participatory simulations in the classroom, where a whole class can enact the behavior of a system even while each student controls only a small part of the system from a networked computer or TI graphing calculator. Thus, participatory simulation activities support new forms of classroom interactions that explore and model complex mathematics typical of biological and social phenomena.

SimCalc

The SimCalc Project (Roschelle and Kaput, 1996; Hegedus, 2005) uses hand-held computer technology to democratize access to the mathematics of change and variation, including ideas underlying calculus (Kaput, 1994, 1997). The MathWorlds software, along with the hand-held technology supports computation, but also can represent mathematical ideas in ways that are important for conceptual understanding.

As with earlier tools, such as Function Probe and ANIMATE, the original versions of MathWorlds allows students to primarily control the motions of animated characters by constructing and modifying mathematical functions represented in graphical, tabular, or algebraic forms. Students can build up their understanding of the mathematics by seeing how changes in formalisms lead to changes in the corresponding animation. As an important

extension from earlier dynamic systems that support proportional and linear functions, students are also asked to model stories that correspond to piecewise linear functions of familiar situations that represent different phases of action (such as resting).

As SimCalc has matured, its capabilities to support classroom connectivity have become more central (Kaput and Hegedus, 2002). SimCalc leverages the revolutionary aspects of wireless networked technology in several ways. It allows teachers or students to collect and display student responses, and thereby support large-scale forms of classroom interaction. As a mobile form of technology, SimCalc flexibly supports new kinds of social and participatory structures than were previously possible with tethered systems and those that require students to gather around a single monitor and keyboard (**Figure 5**). At the core of this connectivity is the use of hub technology that rapidly and wirelessly communicates to the teacher's computer.

SimCalc has been shown to be effective in randomized control-group studies at the middle-school and high-school levels. In a multiyear study of middle-school classrooms in Texas, SimCalc use led to statistically significant gains each year (Roschelle *et al.*, 2008). Detailed analyses showed that SimCalc students exhibited their gains on the most advanced math concepts, while showing no concomitant loss on basic material.

Serious Games

Games have been a long-running source of inspiration for mathematics activities and instruction, and gaming technology stepped easily into this practice as personal computers came on the scene and advances were made in interactivity (e.g., *How the West was Won*; Burton and



Figure 5 A teacher oversees students working with Simcalc.

Brown, 1979) and computer graphics (*Green Globes*; Dugdale, 1982). With the advent and popularity of video games, we have seen emergence of a variety of studies showing learning gains in mathematical reasoning from video game playing (e.g., Okagaki and Frensch, 1994; Subrahmanyam and Greenfield, 1994; Wenglinisky, 1998). While important, these are often unconscious and indirect effects of learning rather than the intended purpose behind the game design.

Some educational scholars have argued that education has much to learn from video game design (Gee, 2003/2007; Shaffer, 2007; Squire, 2006). Serious games exploit many of the compelling attributes of video games that have stoked their popularity. These are advanced computer-based environments with sophisticated graphics and sound that engage the learner/player with a compelling narrative about the video environment and support self-motivated progress, but are specifically designed to support learning rather than entertainment.

While a few, early serious game systems exist (e.g., *Quest Atlantis*, *Civilization III*, *The Triple A Game Show*, *Revolution*, and *Mad City Mystery*) it is still early to point to a body of research literature showing consistent learning gains in mathematics from this approach. Still, for many reasons, both theoretical and pedagogical, this is a promising and rapidly evolving area of study (Barab *et al.*, 2005; Squire and Jenkins, 2003).

Embodied Cognition

One of the traditional criticisms of educational computing is that it distances students from physical, hands-on activities and experiences, and perpetuates a view of the math learner as a disembodied information processor. Within the embodied cognition view (e.g., Barsalou, 2008; Glenberg, 1997; Lakoff and Nuñez, 2001), however, technology should allow people's thoughts and actions to mediate the relationship between real-world phenomena and formal representations.

CamMotion (Boyd and Rubin, 1996) fosters this relationship by providing users ways to extract and analyze data directly from digitized video of objects and events. HyperGami (Eisenberg and Eisenberg, 1998, 1999; Eisenberg and Nishioka 1997) lets students create customized 3-D polyhedral forms on the computer screen that are printed as flat, colorful patterns, but that fold to become tangible models, such as penguins.

Recent advances in new and powerful output devices permit students to design (on the computer) and then print objects in sturdy materials such as wood, acrylic, foam core, wax, and plaster. Advances in materials science and developments in plastics, liquid crystals, and optical fibers also invite new ways of using the hands and body to engage in mathematics, and in so doing, recast the very notion of educational technology (Eisenberg *et al.*, 2005).

This greatly expands the range of mathematical objects and techniques available to students. For example, laser printing can be used to make sliceforms that can be slotted together to form mathematical objects in 3-D that are both educational and esthetic (**Figure 6(a)**). The approach can be used to support proof by construction. It can also be used on fabric, where mathematical patterns can inspire fashion. Laser printing is also instrumental to the MachineShop program (Blauvelt and Eisenberg, 2001, 2006), which directs a laser-cutting device to make toothed gears, cams, and levers from wood based on custom-designed mathematical functions that can then be assembled to make devices, including toys and models of dynamic systems (see **Figure 6(b)** and **6(c)**).

These approaches to tangible mathematics – and there are many more to be reviewed – recall a time when mathematics and art were closer than they are today for many students. Mathematically inspired crafts also encourage a culture of display (Eisenberg *et al.*, 2005). Thus, mathematical craftwork takes the idea of grounding the meaning of mathematics one large step forward, by inviting students to personalize mathematics and enrich our immediate surroundings with beautiful and interesting mathematical entities that we design and construct. Combined with the aims that focus on developing students' spatial reasoning and conceptual and procedural advancement, this is a valuable reminder of what math education can become.

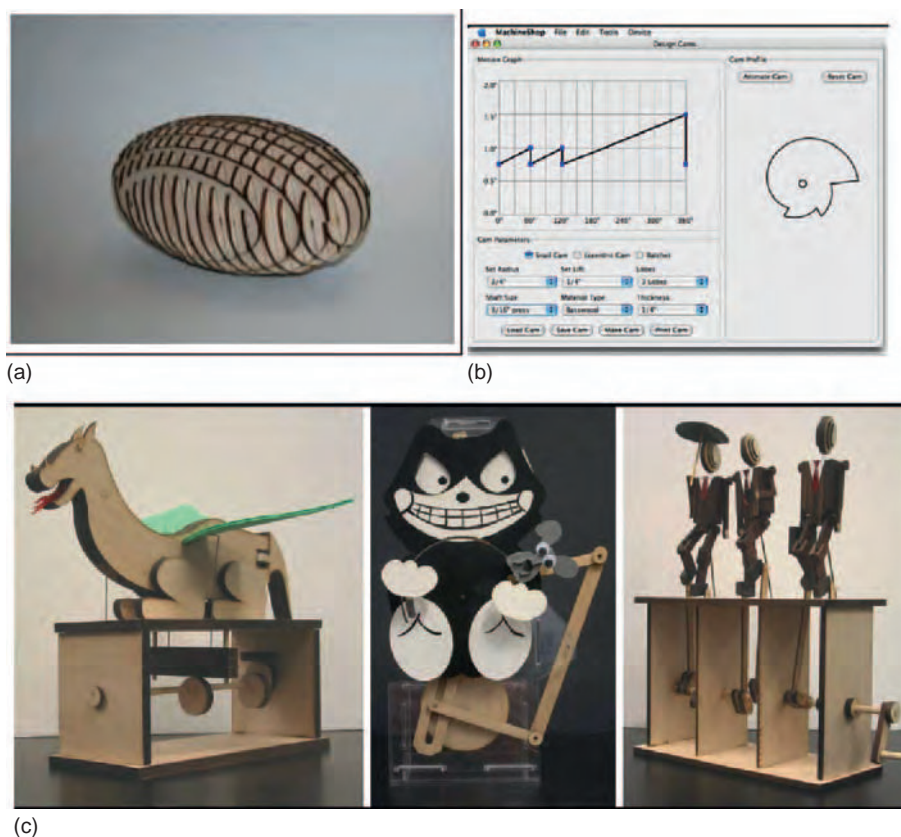


Figure 6 (a) A sliceform ellipsoid constructed from a set of slotted wooden pieces. (b) A custom-designed cam from MachineShop. (c) Some mechanical automata designed and built by students. Adapted from Figure 3 of Eisenberg, M., Eisenberg, A., Blauvelt, G., Hendrix, S., Buechley, L., and Elumeze, N. (2005). Mathematical crafts for children: Beyond scissors and glue. *Proceedings of Art+Math=X Conference*, pp. 61–65, CO: Boulder.

Challenges Facing Technology Supports for Acquiring Mathematics

New views of learning and behavior are contributing to new forms of technology as the interactions between technological tools and tool users are reconceptualized. While *Jasper Woodbury* series, Cognitive Tutors, and SimCalc are notable exceptions, the empirical research base for many of the technological innovations reviewed is still thin. Most evaluations could benefit from qualitative investigations that document implementation fidelity and the learning process alongside more conventional quantitative studies of assessment performance.

Technology enacts a ratchet effect (Tomasello, 1999) on mathematics education, with intellectual advancements supporting the democratization of mathematics for all learners (Kaput, 1994). Yet new technologies introduce new costs to education. First is the cost of the technologies themselves, as well as adequate technical support. Second, new technology calls for new forms of teacher support that must be ongoing and systemic to the educational institutions and it must provide direct connections between the technology and the mathematical

content that it is designed to support. To adequately meet the potentials these new opportunities afford, teachers need additional training. Third, the new technologies present entirely new areas of mathematics (e.g., dynamic systems and inferential statistics) and are shifting prior topics into earlier grade levels (e.g., algebra and calculus). Finally, assessments of student learning, as well as the curriculum and professional standards will need to be revised to keep pace, or fall seriously out of date.

See also: Classroom uses of Technology to Manage Instruction.

Bibliography

- Abelson, H. and diSessa, A. (1981). *Turtle Geometry: The Computer as a Medium for Exploration*. Cambridge, MA: MIT Press.
- Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist* **51**, 355–365.
- Anderson, J. R. (1988). The expert module. In Polson, M. and Richardson, J. (eds.) *Handbook of Intelligent Training Systems*, pp 21–53. Hillsdale, NJ: Erlbaum.
- Barab, S., Thomas, M., Dodge, T., Carteaux, R., and Tuzun, H. (2005). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development* **53**(1), 86–107.

- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology* **59**, 617–645.
- Battista, M. T. and Clements, D. H. (1991). *Logo Geometry*. Morristown, NJ: Silver Burdett and Ginn.
- Ben-Zvi, D. and Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions and challenges. In Ben-Zvi, D. and Garfield, J. (eds.) *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*, pp 3–15. Dordrecht: Springer.
- Blauvelt, G. and Eisenberg, M. (2001). Machine shop: Steps toward exploring novel I/O devices for computational crafts. In *Proceedings of IEEE International Conference on Advanced Learning Technologies (ICALT 2001)*, August, Madison, WI.
- Blauvelt, G. and Eisenberg, M. (2006). Computer-aided design of mechanical automata: Engineering education for children. *Proceedings of ICET 2006, The IASTED International Conference on Education and Technology*, Calgary, Alberta, Canada, July, 2006.
- Bottge, B. A., Heinrichs, M., Chan, S., Mehta, Z. D., and Watson, E. (2003). Effects of video-based and applied problems on the procedural math skills of average- and low-achieving adolescents. *Journal of Special Education Technology* **18**, 5–22.
- Bottge, B. A., Heinrichs, M., Mehta, Z., and Hung, Y. (2002). Weighing the benefits of anchored math instruction for students with disabilities in general education classes. *The Journal of Special Education*, **35**(4), 186–200.
- Bottge, B. A., Rueda, E., LaRoque, P. T., Serlin, R. C., and Kwon, J. (2007). Integrating reform-oriented math instruction in special education settings. *Learning Disabilities Research and Practice* **22**, 96–108.
- Bottge, B. A. and Watson, E. A. (2002). Using video-based math problems to connect the skills and understandings of incarcerated adults with disabilities. *Journal of Special Education Technology*, **17**(2), 25–38.
- Boyd, A. and Rubin, A. (1996). Interactive video: A bridge between motion and math. *International Journal of Computers for Mathematical Learning* **1**(1), 57–93.
- Brown, J. S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* **18**(1), 32–41.
- Burrill, G. (2004). Teaching and learning mathematics using handheld graphing technology. *The Tenth International Congress on Mathematical Education*, Copenhagen, Denmark.
- Burrill, G., Allison, J., Breaux, G., et al. (2002). Handheld graphing technology at the secondary level: Research findings and implications for classroom practice. Dallas, TX: Texas Instruments Corp. <http://education.ti.com/us/global/promo/research.html> (accessed August 2009).
- Burton, R. R. and Brown, J. S. (1979). An investigation of computer coaching for formal learning activities. *International Journal of Man-Machine Studies* **11**, 5–24.
- Chao, S.-J., Stigler, J. W., and Woodward, J. A. (2000). The effects of physical materials on kindergartners' learning of number concepts. *Cognition and Instruction* **18**(3), 285–316.
- Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research* **53**(4), 445–459.
- Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development* **42**(2), 21–29.
- Clements, D. H. and Battista, M. T. (1991). The development of a Logo-based elementary school geometry curriculum. *Final Report: NSF Grant MDR-8651668*. Buffalo, NY: State University of New York at Buffalo and Kent State University.
- Clements, D. H. and Battista, M. T. (1992). Geometry and spatial reasoning. In Grouws, D. A. (ed.) *Handbook of Research on Mathematics Teaching and Learning*, pp 420–464. New York: Macmillan.
- Cobb, G. (1993). Reconsidering statistics education: A national science foundation conference. *Journal of Statistics Education* **1**(1) [Online].
- Cobb, P., Yackel, E., and Wood, T. (1992). A constructivist alternative to the representational view of mind in mathematics education. *Journal of Research in Mathematics Education* **23**(1), 2–33.
- Colella, V., Borovoy, R., and Resnick, M. (1998). Participatory simulations: Using computational objects to learn about dynamic systems. *Proceedings of the Computer Human Interface (CHI) '98 conference*. Association of Computing Machinery, Los Angeles, CA.
- Colella, V., Klopfer, E., and Resnick, M. (1999). StarLogo community of learners workshop. Logo exchange. *Journal of the ISTE Special Interest Group for Logo-Using Educators* **17**(2) [online].
- Confrey, J. and Doerr, H. (1994). Student modelers. *Interactive Learning Environments* **4**(3), 199–217.
- Confrey, J. and Mahoney, A. (1991). Function Probe (Computer program). Santa Barbara, CA, USA.
- Confrey, J. and Mahoney, A. (1996). Function probe. *Communications of the Association for Computing Machinery* **39**(8), 86–87.
- CTGV (Cognition and Technology Group at Vanderbilt University) (1992). The Jasper series as an example of anchored instruction: Theory, program description, and assessment data. *Educational Psychologist* **27**, 291–315.
- CTGV (Cognition and Technology Group at Vanderbilt University) (1997). *The Jasper Project: Lessons in Curriculum, Instruction, Assessment, and Professional Development*. Mahwah, NJ: Erlbaum.
- Dugdale, S. (1982). Green globs: A microcomputer application for graphing of equations. *Mathematics Teacher* **75**(3), 208–214.
- Dugdale, S., Guerrero, S., and Walker, N. (2004). Technology in support of middle grade mathematics: What have we learned? *Journal of Computers in Mathematics and Science Teaching* **23**(1), 5–20.
- Eisenberg, M. (2003). Mindstuff: Educational technology beyond the computer. *Convergence* **9**(2), 29–53.
- Eisenberg, M. and Eisenberg, A. N. (1998). Shop class for the next millennium: Education through computer-enriched handicrafts. *Journal of Interactive Media in Education* **98**(8), 1–30.
- Eisenberg, M., Elumeze, N., Buechley, L., et al. (2005). The homespun museum: Computers, fabrication, and the design of personalized exhibits. In *Proceedings of Creativity and Cognition*, pp 13–21, London UK.
- Eisenberg, M. and Nishioka, A. (1997). Orihedra: Mathematical sculptures in paper. *International Journal of Computers for Mathematical Learning* **1**, 225–261.
- Erickson, T. (2000). *Data in Depth: Exploring Mathematics with Fathom*. Emeryville, CA: Key Curriculum Press.
- Finzer, W. and Erickson, T. (1998). DataSpace – A computer learning environment for data analysis and statistics based on dynamic dragging, visualization, simulation, and networked collaboration. In Pereria-Mendoza, L., Kea, L. S., Kee, T. W., and Wong, W.-K. (eds.) *Statistical Education – Expanding the Network. Proceedings of the Fifth International Conference on Teaching Statistics*, Singapore. Voorburg: The Netherlands: International Statistical Institute.
- Finzer, W. and Jackiw, N. (1998). Dynamic manipulation of mathematical objects. *Paper Presented at the NCTM Standards 2000 Technology Planning Conference*, Washington, DC.
- Frame, M. L. and Mandelbrot, B. B. (2002). *Fractals, Graphics, and Mathematics Education*. Washington, DC: Mathematical Association of America.
- Gee, P. J. (2003/2007). *What Video Games Have to Teach Us about Learning and Literacy*. New York: Palgrave Macmillan.
- Glenberg, A. (1997). What memory is for: Creating meaning in the service of action. *Behavioral and Brain Sciences* **20**, 1–55.
- Hegedus, S. (2005). Dynamic representations: A new perspective on instrumental genesis. *Proceedings of Fourth Congress of the European Society for Research in Mathematics Education*. Sant Feliu de Guixols, Spain: Ramon Llull University.
- Heid, M. K. (1988). Resequencing skills and concepts in applied calculus using the computer as a tool. *Journal for Research in Mathematics Education* **19**(1), 3–25.
- Hiebert, J. (1989). The struggle to link written symbols with understandings: An update. *Arithmetic Teacher* **36**, 38–44.
- Hudson, S., McMahon, K., and Overstreet, C. (2002). *The 2000 National Survey of Science and Mathematics Educators: Compendium of Tables*. Chapel Hill, NC: Horizon Research.
- Jones, K. (2000). Providing a foundation for deductive reasoning: Students' interpretations when using dynamic geometry software and their evolving mathematical explanations. *Educational Studies in Mathematics* **44**(1/2), 55–85.
- Kaput, J. J. (1989). Linking representations in the symbol systems of algebra. In Wagner, S. and Kieran, C. (eds.) *Research Issues in the Learning and Teaching of Algebra*, pp 167–194. Reston, VA: National Council of Teachers of Mathematics.

- Kaput, J. (1994). Democratizing access to calculus: New routes using old roots. In Schoenfeld, A. (ed.) *Mathematical Thinking and Problem Solving*, pp 77–155. Hillsdale, NJ: Erlbaum.
- Kaput, J. (1997). Rethinking calculus: Learning and thinking. *American Mathematical Monthly* **104**(8), 731–737.
- Kaput, J. and Hegedus, S. (2002). Exploiting classroom connectivity by aggregating student constructions to create new learning opportunities. *Twenty-Sixth Conference of the International Group for the Psychology of Mathematics Education*, Norwich, UK.
- Keller, B. A. and Russell, C. A. (1997). Effects of the TI-92 on calculus students solving symbolic problems. *International Journal of Computer Algebra in Mathematics Education* **4**(1), 77–98.
- Koedinger, K. R., Anderson, J. R., Hadley, W. H., and Mark, M. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education* **8**, 30–43.
- Konold, C. (2002). Teaching concepts rather than conventions. *New England Journal of Mathematics* **34**(2), 69–81.
- Konold, C. and Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education* **33**(4), 259–289.
- Kozma, R. B. (1991). Learning with media. *Review of Educational Research* **61**(2), 179–211.
- Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development* **42**(2), 7–19.
- Kuzler, B. (2000). The algebraic calculator as a pedagogical tool for teaching mathematics. *International Journal of Computer Algebra in Mathematics Education* **7**(1), 5–23.
- Laborde, C. (2000). Dynamic geometry environments as a source of rich learning contexts for the complex activity of proving. *Educational Studies in Mathematics* **44**(1), 151–161.
- Lakoff, G. and Núñez, R. (2001). *Where Mathematics Comes from: How the Embodied Mind Brings Mathematics into Being*. New York: Basic Books.
- Lehrer, R. and Littlefield, J. (1993). Relationships among cognitive components in logo learning and transfer. *Journal of Educational Psychology* **85**, 317–330.
- Lehrer, R., Randle, L., and Sancilio, L. (1989). Learning pre-proof geometry with Logo. *Cognition and Instruction* **6**, 159–184.
- Namath, T. (1999). The effects of Cabri geometry for exploring geometry in classroom. In *Proceedings of the Fourth Asian Technology Conference in Mathematics*, Guangzhou, China.
- Nathan, M. J. (1990). Empowering the student: Prospects for an unintelligent tutor for word algebra problem solving. In Carrasco, J. and Whiteside, J. (eds.) *Proceedings of Computer-Human Interaction (CHI)*, pp 407–414. New York: Association of Computing Machinery.
- Nathan, M. J. (1998). The impact of theories of learning on learning environment design. *Interactive Learning Environments* **5**, 135–160.
- Nathan, M. J., Kintsch, W., and Young, E. (1992). A theory of algebra word problem comprehension and its implications for the design of computer learning environments. *Cognition and Instruction* **9**(4), 329–389.
- Nathan, M. J., Stephens, A. C., Masarik, D. K., Alibali, M. W., and Koedinger, K. R. (2002). Representational fluency in middle school: A classroom study. In Mewborn, D. S., Sztajn, P., White, D. Y., et al. (eds.) *Proceedings of the Twenty-Fourth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, vol. 1, pp 462–472. Columbus, OH: ERIC Clearinghouse for Science, Mathematics and Environmental Education.
- Okagaki, L. and Frensch, P. (1994). Effects of video game playing on measures of spatial performance: Gender effects in late adolescence. *Journal of Applied Developmental Psychology* **15**, 33–58.
- Palmiter, J. R. (1991). Effects of computer algebra systems on concept and skill acquisition in calculus. *Journal for Research in Mathematics Education* **22**, 151–156.
- Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. New York: Basic Books.
- Pierce, R. and Stacey, K. (2007). Developing algebraic insight. *Mathematics Teaching* **203**, 12–16.
- Resnick, M., Bruckman, A., and Martin, F. (1996). Planos not stereos: Creating computational construction kits. *Interactions* **3**(6), Sept./Oct.
- Resnick, M. and Wilensky, U. (1998). Diving into complexity: Developing probabilistic decentralized thinking through role-playing activities. *Journal of the Learning Sciences* **7**(2), 153–171.
- Ritter, S., Anderson, J. R., Koedinger, K. R., and Corbett, A. (2007). Cognitive tutor: Applied research in mathematics education. *Psychonomic Bulletin and Review* **14**(2), 249–255.
- Roschelle, J. and Kaput, J. (1996). SimCalc mathworlds for the mathematics of change. *Communications of the ACM* **39**(8), 97–99.
- Roschelle, J., Shechtman, N., and Tatar, D. G. (2008). Mapping “geography of opportunity” in a large-scale randomized experiment on enhancing mathematics with technology. *AERA 2008*, (Session 58.026), Hilton New York.
- Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education* **10**(3) [Online].
- Senk, S., Burrill, G., and Olanoff, D. (unpublished). Knowledge for teaching algebra: What can we learn from examining textbooks? Unpublished manuscript from A Study of Algebra.
- Shaffer, D. W. (2007). *How Computer Games Help Children Learn*. New York: Palgrave Macmillan.
- Sowell, E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education* **20**, 498–505.
- Squire, K. (2006). From content to context: Videogames as designed experience. *Educational Researcher* **35**(8), 19–29.
- Squire, K. and Jenkins, H. (2003). Harnessing the power of games in education. *Insight* **3**(1), 5–33.
- Subrahmanyam, K. and Greenfield, P. (1994). Effect skills in girls and boys. *Journal of Applied Developmental Psychology* **15**, 13–32.
- Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.
- Uttal, D. (2003). On the relation between play and symbolic thought: The case of mathematics manipulatives. In Saracho, O. N. and Spodek, B. (eds.) *Contemporary Perspectives on Play in Early Childhood Education*, pp 97–114. Charlotte, NC: Information Age Publishing.
- Uttal, D. H., Scudder, K. V., and DeLoache, J. S. (1997). Manipulatives as symbols: A new perspective on the use of concrete objects to teach mathematics. *Journal of Applied Developmental Psychology* **18**, 37–54.
- Waits, B. K. and Demana, F. (eds.) (1994). The calculator and computer precalculus project (C2PC): What have we learned in ten years? In Bright, G. W., Waxman, H. C., and Williams, S. E. (eds.) *Impact of Calculators on Mathematics Instruction*. Blue Ridge Summit, PA: University Press of America.
- Waits, B. K. and Demana, F. (2000). Calculators in mathematics teaching and learning: Past, present and future. In Burke, M. J. (ed.) *Learning Mathematics for a New Century*, pp 51–66. Reston, VA: National Council of Teachers of Mathematics.
- Wenglinsky, H. (1998). *Does it Compute? The Relationship between Educational Technology and Student Achievement in Mathematics*. Princeton, NJ: Policy Information Center Research Division.
- Wilensky, U. (1999). NetLogo (Computer program and NetLogo User Manual), Center for connected learning and computer-based modeling, Northwestern University. <http://ccl.northwestern.edu/netlogo/> (accessed August 2009).
- Wilensky, U. and Stroup, W. (1999). Learning through participatory simulations: Network-based design for systems learning in classrooms. *Proceedings of the Computer Supported Collaborative Learning Conference*, Stanford University, Palo Alto, California.
- Yerushalmy, M. and Chazan, D. (1990). Overcoming visual obstacles with the aid of the supposer. *Educational Studies in Mathematics* **21**(3), 199–219.

Further Reading

- Balacheff, N. and Sutherland, R. (1994). Epistemological domain of validity of microworlds, the case of Logo and Cabri-géomètre. In Lewis, R. and Mendelshon, P. (eds.) *Proceedings of the IFIP TC3/WG3.3 : Lessons from Learning*, pp 137–150. North-Holland.
- CTGV (Cognition and Technology Group at Vanderbilt) (1996). Looking at technology in context: A framework for understanding technology

- and education research. In Berliner, D. C. and Calfee, R. (eds.) . *Handbook of Educational Psychology*, pp 807–840. New York: Simon and Schuster Macmillan.
- CTGV (Cognition and Technology Group at Vanderbilt) (2000). Adventures in anchored instruction: Lessons learned from beyond the Ivory tower. In Glaser, R. (ed.) *Advances in Instructional Psychology*, vol. 535–99 Mahwah, NJ: Erlbaum.
- Eisenberg, M. (2002). Output devices, computation, and the future of mathematical crafts. *International Journal of Computers in Mathematical Learning* 7(1), 1–44.
- Goldenberg, E. P., Cuoco, A. A., and Mark, J. (1998). A role for geometry in general education. In Lehrer, R. and Chazan, D. (eds.) *Designing Learning Environments for Developing Understanding of Geometry and Space*, pp 3–44. Mahwah, NJ: Erlbaum.
- Kaput, J., Hegedus, S., and Lesh, R. (2007). Technology becoming infrastructural in mathematics education. In Lesh, R., Hamilton, E., and Kaput, J. (eds.) *Foundations for the Future in Mathematics Education*, pp 173–192. Mahwah, NJ: Erlbaum.
- Lehrer, R. and Chazan, D. (eds.) (1998). *Designing Learning Environments for Developing Understanding of Geometry and Space*. Mahwah, NJ: Erlbaum.
- Lehrer, R., Lee, M., and Jeong, A. (1999). Reflective teaching of logo. *Journal of the Learning Sciences* 8, 245–288.
- Lehrer, R. and Littlefield, J. (1991). Misconceptions in logo: The role of instruction. *Journal of Educational Psychology* 83, 124–133.
- National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Piaget, J. and Inhelder, B. (1974). *The Child's Conception of Quantities*. Boston, MA: Routledge and Kegan Paul.
- Resnick, M. (1994). *Turtles, Termites, and Traffic Jams: Explorations in Massively Parallel Microworlds*. Cambridge, MA: MIT Press.
- Schwartz, J. L., Michal Yerushalmy, M., and Beth Wilson, B. (1993). *The Geometric Supposer What Is It A Case Of?* London: Routledge.
- Squire, K. D. and Jan, M. (2007). Mad city mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. *Journal of Science Education and Technology* 16(1), 5–29.
- Steen, L. (2001). *Mathematics and Democracy: The Case for Quantitative Literacy*. Washington, DC: National Council on Education and the Disciplines.

Relevant Websites

- <http://www.l3d.cs.colorado.edu> – Craft technology.
- <http://www.kaputcenter.umassd.edu> – The James J. Kaput Center for Research and Innovation in Mathematics Education.

Technology Supports for Lifelong Learning

A Collins, Northwestern University, Evanston, IL, USA

R Halverson, University of Wisconsin-Madison, Madison, WI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Community of practice – A group of people participating together to carry out different activities, such as garage bands, ham-radio operators, recovering alcoholics, and research scientists.

Long-tail learning – Learners pursuit of unusual topics of interest to them, such as quasars, Chinese history, Japanese animé, cuneiform writing, or Viking ships.

Introduction

All around us, people are learning with the aid of new technologies: children are playing complex video games, workers are taking online courses to obtain an advanced degree, students are taking courses at commercial learning centers, and adults are consulting *Wikipedia*. New technologies create learning opportunities that challenge traditional schools and colleges. These new learning niches enable people, of all ages, to pursue learning on their own terms. People around the world are taking their education out of the school and into homes, libraries, Internet cafes, and workplaces – where they can decide what they want to learn, when they want to learn, and how they want to learn.

This article describes the emergence of such alternative venues for learning and their implications for schools. It suggests that the seeds of a new education system are forming in the rapid growth of new learning alternatives such as home schooling, learning centers, workplace learning, and distance education. These new learning venues foster self-directed learning for people of all ages. Based upon current trends, this article suggests that public schools will not disappear, but that their dominant role in education will diminish considerably.

New Niches for Learning

The seeds of a new education system can be seen all around us. Many of these seeds will affect the learning of children, but many others will affect people of all ages – as learning becomes a lifelong enterprise. The face of

education is changing rapidly, and it remains to be seen exactly what forms education will take over the next decades.

Home Schooling

Technological advances have fueled the development of home schooling, although they did not give rise to the movement. Internet curriculum materials help organize learning content, leaving parents and others to play the role of facilitators. Virtual schools – such as Florida Virtual Schools – provide some of the administrative and curricular structures to support home schooling through distance education (American Youth Policy Forum, 2002). Most K-12 schools have managed to keep technology at the periphery of their teaching, confined largely to courses such as keyboarding, tech prep, and computer programming (Cuban, 2001). It appears that technology fits much more comfortably into the home environment than into the school environment.

Parents of home-schooled children may, in fact, be more helpful than teachers in a context where the content of what students are learning is embodied in computer and video-based materials. Teachers view themselves as experts – whose role is to convey their knowledge to students. In contrast, parents do not think of themselves as experts and hence take on the role of a coach more naturally (Stevens, 2001). The role of coach puts more responsibility for learning onto the student. The parents are available to help the student figure things out and to encourage the students to work hard, but they usually do not take on the role of explaining everything to the learners. Hence, figuring things out for themselves becomes a major job for home-schooled kids.

Workplace Learning

In recent times, both the military and corporations have been building simulations around popular online video games, using the tools that have been developed for creating these games. The military, in particular, has used game design as a path for recruiting and training a generation that grew up on video games. Companies have also picked up on the potential of games for training. Canon, for example, has developed a simulation where technicians must repair simulated copier machines that have different faults inserted in them. In one Cisco game,

players must put together a computer network on Mars. The companies have found the games are more effective than traditional methods for teaching critical skills.

One can imagine a day when most of the training that workers get for their jobs bypasses traditional educational institutions and takes place in online environments. Sales people might practice their skills with simulations of recalcitrant customers. Doctors might practice their skills trying to diagnose unusual cases. Future travel agents might be challenged to develop cost-effective trip plans using the Web. In fact, almost any job-related skill can be taught by practicing the skill, and computer simulations can create immersive environments where the target skills are necessary for solving engaging problems. It is likely that future workers will have to spend much of their time learning – as workplaces keep introducing new processes, techniques, and equipment. Workers may spend their whole lives learning in order to survive in a changing workplace.

Distance Education

Distance education has been around a long time. Correspondence courses have reached thousands of students over the past century. However, the Internet is pushing new kinds of interactive distance education at both the K-12 and collegiate levels. The University of Phoenix, for example, is famous for having tens of thousands of online students all over the country (Maeroff, 2003). Leading universities – including Stanford, Harvard, Massachusetts Institute of Technology (MIT), and Carnegie-Mellon offer a variety of courses to distant locations. The model of the Open University in the United Kingdom has spread across the globe. In addition, over 24 states have developed virtual high schools to provide courses that are not offered in their own schools to students (Tucker, 2007). Distance education has already spread widely, and it will inevitably keep growing in the coming decades.

As busy people realize they need more education, they increasingly opt to take distance-education courses. Often, people find that – to advance in their careers – they need another degree, and distance education courses provide the easiest means to obtain that degree. Although distance education has a head start in adult education, the recent development of virtual K-12 schools challenges brick-and-mortar public schools by providing affordable alternatives.

Learning Centers

Learning centers – whether nonprofit or for-profit – are spreading widely. They usually offer classes, as well as provide access to technology. They are oriented to providing specific knowledge and skills which the users

want to obtain. Such learning centers are likely to proliferate in the coming years.

Privately owned learning centers are growing to fill gaps in the existing education system. These are currently most common as preparation for taking national tests – such as the Scholastic Aptitude Test (SAT) and American College Testing (ACT) – and to provide tutoring for children who are having problems in school. Learning centers are also beginning to develop for adult education. Princeton Review, Sylvan, Thompson, and Kaplan are only some of the companies that provide learning centers. Some of these companies are setting up centers in towns and cities where people can go to prepare for various types of tests or to get specialized teaching in some area of weakness. As they develop, learning centers are likely to employ more and more network-based education and to spread as a way to obtain specific knowledge and credentials.

Learning centers are also making an impact on career education. In the 1990s, the US Department of Education launched an initiative to support nonprofit community technology centers to serve communities where access to computers and other technologies is very limited. The community technology-center network now has over 1000 centers in many different locales – such as housing projects, storefronts, community organizations, and libraries. Most users come to learn job skills and take classes at the centers, as well as to use the Internet facilities. Many acquire improved English-language skills, get tutoring and homework help, or participate in general educational development (GED) and other adult education programs. They also use computers to get information from the Internet, send and receive e-mail, set-up Web pages, and carry out their own self-directed projects.

Computer Simulations

The advent of home computers has led to a proliferation of learning software – some on compact disks-read-only memory (CD-ROMs) and some on the Web. Among the most famous of these are the Sim series – developed by Maxis: SimCity, the Sims, SimLife, SimEarth, SimAnt, and so on. In SimCity, the player is put in charge of developing and running a city – as a kind of dictator. Depending on the user's decisions, the city and its people may thrive or may deteriorate. People who spend time playing the game are learning some of the tradeoffs involved in the dynamics of a city. SimLife, similarly, teaches the ecology of systems, SimEarth – the dynamics of geology – and SimAnt, the behavior of an ant colony. The Sims games allow people to design a human character to represent them and see how the character's actions play out in a simulated world. It is one of the most popular games on the market, at present. The Sim series are engaging games that teach a variety of subjects, but how much children learn from playing such

games depends on how much they think with regard to the ideas and issues the simulations embody.

An understanding of video games as learning environments is becoming increasingly important as gaming culture rivals schooling for the attention of children and adolescents across the world. Gee (2003) argues that the compelling nature of video game participation is, in part, due to the underlying social, cognitive, and developmental learning principles around which successful games are built. The implication of Gee's perspective is that games and gaming can be a source for inspiration in building more effective learning environments.

Given this popularity and the compelling nature of game-play, software designers should be able to draw on games and the principles of game design to build more compelling learning environments. In the best-selling game *Civilization*, for example, players have the opportunity to relive the development of global social and economic history. Players must plan, choose to negotiate or fight, acquire and allocate resources, and make decisions to advance their civilization. Taken together, these activities point toward how students can integrate theories and practices from across the curriculum in playing a compelling historical simulation. Squire (2004) notes how games such as *Civilization* can provide students – who typically are not engaged deeply in learning history – with an opportunity to replay history. Students can investigate, for example, what would have happened if Africans – rather than Europeans – had colonized the Americas, and can begin to understand what theories of social change look like in action, rather than in books.

Trends, at present, suggest that there will be a steady accumulation of learning software both for children and adults. For a generation that has grown up with sophisticated game simulations, learning with technologies is likely to come more easily. Gaming may help players learn a variety of leadership skills, such as resource allocation, negotiating with friends and adversaries, manipulating situations and environments, actively pursuing their goals, and recovering from failures. As Brown and Thomas (2006) have suggested, the gamers of today may become the leaders of tomorrow.

Technical Certifications

In recent years, a host of companies – such as Microsoft, Cisco, and Novell, as well as technical societies such as the Institute of Electrical and Electronics Engineers (IEEE) Computer Society – have started to offer online learning and examinations that certify the mastery of technical skill in computer-related occupations. The Cisco Networking Academy, for example, provides a comprehensive training program for network administration. The Cisco Academy is based on a tight linkage of curriculum, learning-by-doing, and assessment activities that are

coordinated in a learning environment that blends Web-based and classroom learning. Cisco Academy partners with schools to train their teachers to prepare students for the Cisco exams. They have trained over 400 000 students in 150 countries with a curriculum translated into nine languages. Certification programs such as these have, in turn, led to a number of training programs in commercial colleges and community colleges and on the Web to prepare students for the certification exams. These certification programs provide an alternative to technical degree programs for students – who may otherwise struggle with the academic focus of high school and colleges.

In the long run, these certification programs are a threat to the monopoly held by schools and colleges. Because the technical certifications they provide are more specific than a degree or diploma, they are, in fact, more meaningful to potential employers. They specify precisely what skills a student has acquired in a way that a high school diploma or a college degree cannot do. Furthermore, because the certifications are so specific, it is possible to tailor any educational program directed toward them much more carefully. In fact, the companies and technical societies that are developing certification programs have, in many cases, developed very clear specifications as to what and how courses should be taught to prepare people to take the certification exams.

Internet Cafes

All over the world, Internet cafes are springing up where people can go and log on to the Web for a small fee. These cafes are the libraries of the future. They particularly attract young people who spend hours on the Web – engaging in conversations and games, reading about what is happening in the world, learning how to program, or exploring different sites that relate to their interests.

More and more of the world's accumulated knowledge is spreading to the Web. Hence – with access to the Web in many locations – people can begin to educate themselves. This is the role that public libraries played in the past. As people all over the world see the necessity of an education to prosper in a technological society, they are likely to start teaching themselves through the resources of the Web. Inexpensive access to the Web through cafes provides the whole world with access to knowledge, new social arrangements, and new ways of thinking with regard to learning. Alternatively, of course, it also provides an avenue for wasting an enormous amount of time.

Fostering Self-Directed Learning

The Web gives children and adults the ability to pursue topics they are particularly interested in and feel passionate about – including topics such as quasars, Chinese

history, Japanese *animé*, cuneiform writing, Viking ships, and casino games – to name just a few. These are topics that learners never encounter in school unless they pursue them later in college. Nor are learners likely to find people among their acquaintances who share their interests and would study the topics together with them. Such exotic topics remained as lone pursuits for most people until the Web came along.

Virtual communities make it possible to bring together people, from all over the world, with common passions. Their communication can take place over electronic mail and in computer news groups, social networking sites, or Multi-User Virtual Environments (MUEs), such as Moose Crossing (Bruckman, 2002). In these different computer-based environments, people of different ages are gathering around topics that they are passionate about. This enables them to learn about their passion much more deeply than they can on their own. Often, people cannot find others locally who share their passions, but the reach of the Internet allows them to go beyond their local community to find like-minded communities.

Barron (2006) provides a good example of how students can become self-directed learners through exploring their interests on the Web. She finds that they use a number of different strategies for learning on their own. For example, they may learn from friends or mentors, take courses in school, observe the work of others in online communities such as Xanga, or study the products that others exhibit on different websites. Learning to be a self-directed learner by participating in a variety of learning contexts will help more students to develop the skills necessary to choose learning contexts well.

The phrase “The Long Tail” was first coined by Anderson (2004) in an October *Wired* magazine article to describe how modern culture is increasingly shifting away from a focus on mainstream products at the head of the demand curve and toward a huge number of niches in the tail (as exemplified by Amazon or Netflix – that sell a large number of unique items in relatively small quantities). The Web enables long-tail learning in a variety of ways by supporting participation culture among the members of communities interested in idiosyncratic topics. Novices can lurk in these communities to pick up the issues, techniques, and jargon that sustain the communities. When learners develop expertise, they can display their work or their thoughts to the community and get feedback from the community to guide their further development. Finished products may, later, be posted on YouTube or Epinions – where the world might see their work.

Another way the Web enables long-tail learning derives from the plethora of information available on the Web. When learners go into most libraries, they are likely to find a limited amount of information on any topic; the information they find is very likely out of date for most topics, and there is rarely documentation on the

techniques of the community of practice (Wenger, 1998) around the topic are using. Based on the fact that the Web is both constantly evolving and actively filling up all the long tails of knowledge concerning every conceivable topic, it can support long-tail learning in a way not even the largest library in the world could. In addition, the Web can provide expert audit trails, active simulations, and tutorials on topics that support learning well beyond what learners can glean from books.

Conclusion

The changes that are occurring in education are neither all good nor all bad. There are many benefits to the kinds of education that technology affords – such as the ability of learners to pursue deeply topics of interest to them and to take responsibility for their own education. However, at the same time, some of the effects of technology on education are problematic. In particular, the new technologies can undermine both Thomas Jefferson’s vision of educating citizens who can make sensible public policy decisions, and Horace Mann’s vision of a society where everyone can succeed by obtaining a good education. Increasing the ability to personalize educational opportunities provides a natural advantage to those who can afford the services. Citizenship and equity may be undermined by the fragmentation and commercialization afforded by the technological revolution. There are clear dangers as well as great possibilities when people take learning into their own hands.

When taken together, the cumulative effects of these innovations is to extend learning throughout life and over many venues. With time, these pieces might come to comprise the fragments of a new system of education in which schools are pushed to the peripheral role in learning that they once occupied prior to the industrial revolution. However, these elements supporting technology-based learning outside of school have developed independent of one another. They do not, in any sense, form a coherent system of education. That is where the need for visionaries is most apparent. It will take a new group of energetic visionaries, to once again do the kind of work that Horace Mann and his colleagues did – that is, to figure out how to build an equitable and coherent system from these emergent technological pieces.

Eventually, when people and politicians become worried about what kids are learning or what adults do not know, their automatic reaction may not be “How can we improve the schools?” Instead, they may ask, “How can we develop games to teach mechanical engineering?”, “How can we make new technology resources available to more people?”, or “What kinds of tools can support people to seek out information on their own?” These are all questions that push the envelop for improving education out of the

schools and into new venues. The link between schooling and learning forces our conversation into institutional responses – we do not, yet, know how to ask wider questions when we think about improving education (Collins and Halverson, 2009).

The new view of education is structured around the idea of lifelong learning. Lifelong learning requires moving away from schooling institutions that structure individual learning. Instead, learners need to act as consumers of a wide variety of learning experiences. Learners will need to develop the skills to judge the quality of learning venues and to develop the kinds of social networks that provide guidance and advice.

The recent explosion of social networking points to how technologies can replicate the support and guidance functions of schools. For example, user groups and community sites exist for every known disease and disorder, and doctors across the country know their diagnoses are checked by an increasingly informed patient population. These kinds of social networks are blossoming around topics of particular interest to different groups of people – topics such as poetry, chemistry, digital graphics, and fantasy sports. The networks draw people across all ages from very different backgrounds, some quite expert and others virtual novices. Some learn by lurking in the background and others by asking questions. Groups in the network may jointly investigate topics of interest or argue about issues they think important. The successful sites, however, share the characteristic of providing useful information to guide the interests of users.

If schools cannot change fast enough to keep pace with advances in learning technologies, learning will leave schooling behind. With inexpensive computers, young people in Thailand and Brazil can have access to the same resources for learning that people in the developed world now have. Many may choose to take advantage of these resources to escape from poverty. In some ways, they will be a new kind of twenty-first-century immigrant – instead of moving to a new country, they will use information networks to transform their thinking. They will be able to find like-minded souls to share ideas in cyberspace. This new approach to learning that is developing among people around the world will be highly interactive, very personalized, and give the learner much more control over their own learning.

See also: Technological Supports for Acquiring Twenty-First-Century Skills; Technology Supports for Acquiring Inquiry Skills.

Bibliography

American Youth Policy Forum (2002). *Florida Virtual School: The Future of Learning?* <http://www.aypf.org/forumbriefs/2002/fb101802.htm>.

- Anderson, C. (2004). The long tail. *Wired Magazine* 12(10). <http://www.wired.com/wired/archive/12.10/tail.html> (accessed November, 2009).
- Barron, B. (2006). Interest and self-sustained learning as catalysts of development: A learning ecologies perspective. *Human Development* 49(4), 193–224.
- Brown, J. S. and Thomas, D. (2006). You play Warcraft? You're hired! *Wired* 14(4). <http://www.wired.com/wired/archive/14.04/learn.html>.
- Bruckman, A. (2002). Co-evolution of technological design and pedagogy in an online learning community. In Barab, S. and Gray, J. (eds.) *Designing Virtual Communities in the Service of Learning*, pp 236–298. New York: Cambridge University Press.
- Collins, A. and Halverson, R. (2009). *The Second Educational Revolution: How Technology is Transforming Education Again*. New York: Teachers College Press.
- Cuban, L. (2001). *Oversold and Underused: Computers in the Classroom*. Cambridge, MA: Harvard University Press.
- Gee, J. P. (2003). *What Video Games Have to Teach us about Learning and Literacy*. New York: Palgrave Macmillan.
- Maeroff, G. I. (2003). *A Classroom of One: How Online Learning is Changing our Schools and Colleges*. New York: Palgrave Macmillan.
- Squire, K. D. (2004). Sid Meier's Civilization III. *Simulations and Gaming* 35(1), 135–140.
- Stevens, M. L. (2001). *Kingdom of Children: Culture and Controversy in the Home Schooling Movement*. Princeton, NJ: Princeton University Press.
- Tucker, B. (2007). *Laboratories of Reform: Virtual High Schools and Innovation in Public Education*. <http://www.wired.com/wired/archive/14.04/learn.html>.
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. New York: Cambridge University Press.

Further Reading

- Beck, J. C. and Wade, M. (2004). *Got Game: How the Gamer Generation is Shaping Business Forever*. Cambridge, MA: Harvard Business School Press.
- Black, R. W. (2008). *Adolescents and Online Fan Fiction*. New York: Peter Lang.
- Brown, J. S. and Duguid, P. (2000). *The Social Life of Information*. Cambridge, MA: Harvard Business School Press.
- Friedman, T. L. (2006). *The World is Flat: A Brief History of the Twenty-First Century*. New York: Farrar, Straus and Giroux.
- Ito, M., Horst, H., Bittanti, M., et al. (2008). *Living and Learning with New Media: Summary of Findings from the Digital Youth Project*. Chicago, IL: MacArthur Foundation Digital Media and Learning Project.
- Jenkins, H. (2008). *Convergence Culture: Where Old and New Media Collide*. New York: NYU Press.
- Johnson, S. B. (2005). *Everything Bad is Good for You: How Today's Popular Culture is Actually Making Us Smarter*. New York: Riverhead Press.
- Papert, S. (1993). *The Children's Machine: Rethinking School in the Age of the Computer*. New York: Basic Books.
- Roberts, D. F., Foehr, U. G., and Rideout, V. (2005). *Generation M: Media in the Lives of 8-18 Year-Olds*. Menlo Park, CA: The Henry J. Kaiser Family Foundation. <http://www.kff.org/entmedia/entmedia030905nr.cfm> (accessed August 2009).
- Shaffer, D. W. (2006). *How Computer Games Help Children Learn*. New York: Palgrave.
- Steinkuehler, C. (2006). Virtual worlds, learning, and the new pop cosmopolitanism. *Teachers College Record*, 12843.
- Tapscott, D. (1998). *Growing up Digital: The Rise of the Net Generation*. New York: McGraw-Hill.

Relevant Websites

- <http://cisco.netacad.net> – Cisco Networking Academy: Cisco Systems.
- <http://www.ctcnet.org> – Community Technology Centers' Network (CTCnet).

TECHNOLOGY AND LEARNING – SUPPORTS FOR STUDENT COLLABORATION

Knowledge Communities in the Classroom

J D Slotta and H Najafi, University of Toronto, Toronto, ON, Canada

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Fostering community of learners (FCL) – An educational model advanced by Ann Brown and Joseph Campione, from the University of California, Berkeley. In FCL classrooms, students work collaboratively to investigate research questions, exchange expertise, and develop a deep understanding of curriculum topics.

Intentional learning – A form of learning that is a characteristic of a knowledge-building environment. Students actively define their own cognitive objectives and pursue them through activities of their own design.

Jigsaw – A curriculum structure defined by Elliot Aronson and his students at the University of Texas and the University of California, and employed as a pedagogical structure within numerous educational methods, including the FCL method of Ann Brown and Joseph Campione.

Knowledge base – The aggregated ideas, resources, or semantic connections that result from collaborative learning structures. A common feature of knowledge-community approaches is that students work together, adding knowledge elements to their shared knowledge base.

Knowledge building – An educational model advanced by Marlene Scardamalia and Carl Bereiter, from the University of Toronto. In knowledge-building classrooms, students collaboratively identify their own cognitive goals and advance them through intentional learning. The goal of knowledge building is the improvement of ideas.

Knowledge community – A collection of individuals who share a common learning goal or activity focus. The community typically shares common practices or goals, and employs characteristic media.

Research cycle – A central pedagogical structure to the FCL model, the research cycle includes three stages: students engage in research activities, share their findings with the whole class to create collective

expertise, and use the knowledge that they have acquired to complete a final project.

Scaffold – The notion of scaffolding was introduced by David Wood, Jerome Bruner, and Gail Ross to describe explicit supports provided by teachers or instructional materials that enable learners to achieve intellectual results that would otherwise fall beyond their capacity.

Scaffolded inquiry – The term scaffolded inquiry now refers generally to any method where students are guided through activities and interactions by instructional materials and technology environments.

Web 2.0 – A term coined by Tim O'Reilly and other pioneers of the Internet to refer to a new generation of Internet-based applications that were functionally distinct from those of the previous era. Whereas earlier web applications emphasized individual production and mass consumption (i.e., of a web page), Web 2.0 emphasizes mass production and mass consumption. Interactivity and social networking are typical of Web 2.0 resources (e.g., <http://Facebook.com>), as well as the aggregation of large collections of resources contributed by millions of users (e.g., <http://Flickr.com>, <http://youtube.com>).

Wiki – This refers to a web page or website where the contents can be contributed and edited by many visitors. First developed by Ward Cunningham in 1994 and referred to as the WikiWikiWeb, wikis are now commonly used by businesses, scholars, and social networks around the world. <http://Wikipedia.com> is perhaps the best-known example.

Now more than ever, schools must prepare students as lifelong learners in a society where change is the only constant. New issues, new opportunities, and an increasingly complex range of messages and materials will greet them at every turn. As the demands of the workplace and everyday life become ever more diverse, students must embrace this complexity. Recently, scholars have suggested that we are becoming a knowledge society where

citizens rely increasingly on processes of knowledge creation and advancement over those of labor or industry (Scardamalia, 2002; Hargreaves, 2003; Tapscott, 2008). To serve this vision, our educational systems must shift their focus from one of preparing students for the workforce (i.e., with relevant background knowledge and skills) to preparing them to think critically, develop ideas and arguments, evaluate their own progress, and collaborate constructively with peers.

Of course, researchers have long argued for an emphasis on such an autonomous agency for students (e.g., Dewey, 1938/1997; Driver, 1985; Bruner, 1993; Bransford *et al.*, 1999; Linn and Hsi, 2000). Indeed, two relevant research traditions that are reviewed in other articles elsewhere in the encyclopedia are those of guided discovery and scaffolded inquiry. Here, we focus on a highly regarded body of scholarship that is concerned with fostering knowledge communities within classrooms (Brown and Campione, 1994; Scardamalia and Bereiter, 1999; Bielaczyc and Collins, 2005). Broadly cast, this approach involves giving students a much higher level of agency and responsibility for developing their own questions, exchanging and critiquing ideas with peers, and even evaluating their own progress. Teachers become members of the classroom knowledge community, and participate as knowledgeable mentors. The students in a knowledge community typically create a knowledge base of commonly held resources that are negotiated and continuously revised through their activities conducted in the classroom, at home, or on the playground. Often, this knowledge is situated in a technology-mediated environment that scaffolds students as they add new ideas to the knowledge base, revise materials, or synthesize the contents into new organizations, designs, or arguments (Stahl, 2000; Hoadley and Pea, 2002; Bielaczyc and Collins, 2005).

Young people today are frequently immersed in technology environments that are deeply social and linked to a variety of knowledge media. Facebook, for example, is a social-networking environment that emphasizes networks of friends, but allows for rapid propagation and evolution of ideas about political candidates, music, movies, and philosophical concepts. The notion of the blogosphere, once a fanciful, futuristic term, is now deeply set in the landscape of all media, political organizations, and public institutions. Word gets out about any new idea, as a variety of social networks allow the most popular or resonant ideas to quickly emerge and become interconnected.

There is considerable and justified excitement about the new forms of aggregated information and interaction that characterize the Web 2.0. For example, Wikipedia is a continuously expanding and self-adjusting body of knowledge that literally emerges from the combined contributions of millions of knowledgeable individuals. By virtue of the sheer number of visitors and their shared value for keeping information current and accurate, these

pages are becoming an authoritative reference. In recent years, an array of socially oriented applications has drawn upon patterns of use and semantic metadata to provide new functionality. For example, Amazon.com has made stunning use of recommender systems to suggest books to readers based on their own past preferences as well as the patterns of selections made by millions of other users. The new technology architecture of Web 2.0, which emphasizes the use of semantic metadata and social networks, provides a means for such emergent functionality. After more than a decade of experience with the Internet, we are beginning to think about repositories not as collections of static objects to be searched through, but as dynamic spaces that support the creation, organization, and access of knowledge.

The tremendous intellectual energy and social capital of the Web 2.0 has enhanced the relevance of educational research that has, for several decades now, emphasized the importance of knowledge communities. The spirit of knowledge sharing, collaboration, and knowledge construction that is evoked by these technologies could promote a surge of activity by researchers as well as practitioners in classrooms and online environments. We review some of the principal ideas, models, and projects that have been offered by researchers of knowledge communities. We discuss cross-cutting elements and comment on the challenges involved in implementing these approaches. We close with a suggestion for hybrid models that integrate the core ideas of knowledge communities with those of more conventional inquiry-oriented instruction, relying on scaffolding technologies to help teachers and students gradually develop the practices of a knowledge community.

The Knowledge-Community Approach

While educational researchers have advanced different conceptualizations of knowledge communities, they share at least three elements. First, within a knowledge community, members with diverse expertise collectively and intentionally develop and advance a shared knowledge base (Bielaczyc and Collins, 1999). Second, community members develop metaknowledge about the processes underlying learning and knowledge creation (Garrison *et al.*, 2000; Brown and Campione, 1994). Third, knowledge communities foster a form of discourse that allows for idea sharing, critique, and improvement. Of special interest to this review are the conceptual frameworks of fostering communities of learners (FCL) and knowledge building (KB), which have been widely recognized and researched in K–12 classrooms.

FCL was implemented in urban elementary schools to promote critical thinking and reflective practice rooted in deep content knowledge within a purposeful scientific learning community (Brown and Campione, 1996; Brown, 1997). Introduced in the late 1980s and developed through

a series of design-based research studies (Brown, 1992), FLC emphasized a research cycle with three stages: students engage in research activities, share their findings with the whole class to create collective expertise, and finally use the knowledge that they have acquired to complete a final project.

KB also engages students in identifying questions and developing a cumulative understanding (Scardamalia and Bereiter, 2008). In this approach, individuals' ideas are added to the public domain of the knowledge community and undergo the scrutiny of all members, who share the responsibility of advancing their collective knowledge. A supportive social context is required to allow all students to work as a knowledge community, identifying conceptually important ideas and improving them through constructive review (Scardamalia, 2000). A clear distinction is made between learning, an internal personal process, and knowledge building, an intentional effort to advance public knowledge (Scardamalia and Bereiter, 2006).

We synthesize FCL and KB according to an overarching perspective of knowledge communities, focusing on three cross-cutting dimensions: distribution of cognitive responsibility, a shared community knowledge base, and scaffolding frameworks and technologies.

Distribution of Cognitive Responsibility

Traditional forms of instruction, such as lectures, are not conducive to helping students integrate knowledge for deep understanding of concepts (Brown, 1997; Slotta and Linn, 2009). In a knowledge-community approach, students become involved in articulating their own learning goals and planning the learning activities. Members of a knowledge community are not only responsible for the quality of their individual learning, but are also held accountable to communicate with other community members to scrutinize the quality of collective knowledge (Bielaczyc and Collins, 1999).

Students in FCL classrooms are acknowledged for what they know and are involved in identifying their learning needs, setting goals, and locating necessary resources to achieve those goals. Thus they develop a metacognitive awareness that requires establishing an environment to promote learners' cognitive responsibility (Brown, 1992; Brown and Campione, 1994; Campione *et al.*, 1995). Metacognition and reflection serve to increase students' awareness of how they learn and thereby enable self-reflection, critical inquiry, and flexible application of knowledge in novel situations (Brown, 1992). The responsibility for developing metacognitive knowledge applies to the classroom community as a whole, with individual students sharing responsibility for monitoring and correcting their peers (Brown and Campione, 1994). As the community depends on the quality of its common

knowledge base, students will hold themselves accountable for developing a coherent understanding of the topic assigned to them (Brown and Campione 1994).

Although FCL advocates increased cognitive agency on the part of students, it emphasizes a balance between the accountability of teachers and students. Teachers are not expected to know the answers to all questions; rather, they should become inquiry role models who are competent in identifying problem areas and can effectively locate and utilize sources of knowledge and expertise (Brown, 1992; Brown and Campione, 1994; Brown, 1997). Further, they should distinguish among the situations where the students need help to avoid misconceptions, when they are ready to undertake more advanced topics, or where they should be left on their own (Brown, 1992; Brown and Campione, 1994). Without the support, monitoring, and feedback from teachers, students could develop misconceptions about the topic of inquiry or fail to progress in their understandings.

In KB classrooms, the students also assume responsibility for defining their own learning goals, conducting inquiry to pursue solutions, developing their knowledge as a classroom community and monitoring their own progress (Scardamalia, 2000; Scardamalia and Bereiter, 1999, 2006). They identify relevant problems that will expand their knowledge, develop solutions to those problems, share their understandings with others in the form of knowledge objects and finally attend to inconsistencies in the shared ideas (Scardamalia, 2000, 2002; Scardamalia and Bereiter, 2006).

A higher level of cognitive agency for students differentiates KB from FCL (Scardamalia, 2002; Scardamalia and Bereiter, 2008). FCL takes a rather hierarchical approach to cognitive responsibility. In contrast, KB distributes the cognitive responsibility equally among all members of the community. This collective responsibility takes the form of an ongoing negotiation for each student between personally held ideas and those that are external to them (Scardamalia, 2000). Although teachers may have less control in directing classroom discussions, they are actively developing their own understanding of discussed concepts and also improving their knowledge of processes underlying students' understanding (Scardamalia and Bereiter, 1999).

Shared Community-Knowledge Base

A knowledge-community approach encourages students to pursue their personal interests in the context of shared classroom interests. Knowledge sources are no longer limited to the teacher and curriculum materials, but include knowledge artifacts identified or created by the students (Bielaczyc and Collins, 1999).

Students in FCL classrooms collaboratively build a shared knowledge base that is characterized by integrated and transferable knowledge as opposed to declarative facts (Brown, 1992). During benchmark lessons, carefully selected curricular themes are frequently reintroduced to the students as generative ideas that constitute the foundation of community knowledge base, on which students elaborate during the FCL research cycle (Brown and Campione, 1994; Brown, 1992, 1997). Generative ideas are broad enough to be divided into subtopics that students address during research cycles (Shulman and Sherin, 2004; Sherin *et al.*, 2004). Students and teachers together develop a deep understanding of disciplinary topics, which requires communication channels and collaborative structures for sharing and improving a collective knowledge base.

In FCL classrooms, students formulate research questions about their chosen topics, informed by the seeded ideas, and then refer to selected references to thematically categorize their questions, which are divided into subunits. Students then select a subunit to research, thereby increasing their personal agency in learning. To avoid misconceived ideas or unproductive inquiry efforts, the teacher and, partly, the students should monitor the progress of their community's knowledge development. Structured collaborative activities intentionally distribute expertise among students, requiring them to negotiate their individual knowledge within collaborative groups and organize their new understandings in a common community knowledge base (Brown and Campione, 1994; Brown, 1997). Thus, each member of the learning community is knowledgeable in certain areas, or is achieving expertise through cognitive apprenticeship (c.f., Collins *et al.*, 1989).

In KB classrooms, students take the initiative in solving shared knowledge-building problems that motivate a persistent inquiry geared toward idea improvement. Non-knowledge-building questions test current understandings and can be answered either explicitly or implicitly through available print, digital, or human resources (Scardamalia and Bereiter, 1999; Bereiter and Scardamalia, 2000). Knowledge-building questions, in contrast, engage students in proposing solutions at the level of underlying principles, such as gravity, rather than concrete cases, such as why objects fall (Bereiter and Scardamalia, 2000, 2003). The two kinds of questions can be distinguished by the degree of finality required in responding to them.

Students learn to develop their own theories in response to problems that emerge during ongoing inquiry. These theories become objects of students' inquiry and are subject to continuous improvement, similar to the process of scientific researchers (Scardamalia and Bereiter, 1999, 2006; Bereiter and Scardamalia, 2000). Bringing problems of understanding to the center of inquiry requires abandoning the more familiar topic-centered structure of

classroom where curricular themes dictate the topic of discussions at a given time during the school year.

The community knowledge base in KB is public, requiring all students to share their ideas for critique by peers, comparison with other ideas in the knowledge base, and integration into the shared knowledge (Scardamalia, 2000, 2002). Students' contributions to the community knowledge base will survive and be further built on only if they are deemed important to the community (Scardamalia, 2000). Upon presenting a new idea, students should deliberately negotiate the significance of their contribution by citing its connection to the dominant discourse of the community (Scardamalia and Bereiter, 1999; Scardamalia, 2000). The instructor's attention shifts from engaging students in activities that ultimately showcase their understandings to activities that focus on idea improvement as the goal of classroom discourse (Bereiter and Scardamalia, 2003). Thus any artifact produced during such activities should carry the potential to be improved by others. Moreover, unlike FCL, where ideas are only shared with the whole class at certain intervals, students in KB classes have access to all ideas at all times.

In general, the shared knowledge base should promote diversity of ideas and facilitate access to those ideas in order to encourage the growth of community knowledge (Scardamalia, 2002; Brown and Campione, 1994). Authoritative sources, in the form of print and digital resources or expert opinions, are included, as long as they are treated with a critical stance and not as ultimate truth (Scardamalia, 2002; Bereiter and Scardamalia, 2003; Collins *et al.*, 2004). FCL and KB differ to some extent in where they place the responsibility to assess the congruency of the community knowledge base with that of scientific discourse. In KB, this responsibility is delegated to students, where it is seen as an opportunity for students to practice idea improvement and take corrective action (Scardamalia and Bereiter, 1999; Scardamalia, 2002). In contrast, FCL requires teachers to identify and address misconceptions to prevent an incorrect concept from prevailing within the community discourse.

Pedagogical and Technological Scaffolds

To help teachers and students achieve the nuanced flow of ideas, activities, and the discourse within a knowledge-community approach, researchers often introduce specific forms of guidance, constraints, or scaffolds. These include scripts that delineate the groups that students should form, their goals, and what materials they will access. They can also take the form of technologies that guide students' knowledge-building processes or discourse.

As most of the research was conducted before the arrival of the Web, the scaffolding employed by FCL takes the form of carefully specified collaborative and

pedagogical structures. These structures familiarize students with knowledge sharing and knowledge construction, help them understand what is expected, and guide their transition from one activity to another (Campione *et al.*, 1995). The following structures have served to balance collective and individual learning within the FCL classroom:

- *Reciprocal teaching.* Students work in pairs or small groups to interpret materials or work on a problem, summarizing assigned sections of the text, asking questions about the text, clarifying misconceptions, and predicting how the text would continue (Lehrer and Schauble, 2006). Reciprocal teaching promotes comprehension, cognitive monitoring, and complex reasoning (Brown, 1992; Brown and Campione, 1994).
- *Jigsaw groups.* Students are assigned to a research group that is a subtopic of the overall theme. Although every student in the group is ultimately responsible for learning the whole subtopic, they work independently within their research group, investigating their own questions and engaging in reciprocal teaching with other members of their group. Then new groups are created that include one student from each of the subtopic groups. The new group then works together on some consequential task that is seen as a product of the overall cycle (Brown and Campione, 1994).
- *Cross talk.* Students share their progress across groups and sometimes with the whole class to exchange feedback on their research and, if necessary, modify their investigation approaches. Cross talk often takes place alongside jigsaw groups.

Technological scaffolds designed for FCL were limited, although computers were used from the outset to produce materials. In later phases of the research, the Web was added as a searchable resource, as well as e-mail communications where students consulted with content-knowledge experts (Brown, 1997). These communications served two purposes: to broaden the knowledge pool of the community beyond spatial and temporal limitations of classroom, and to provide students with the opportunity to observe scientific discourse and ways of thinking.

In KB, technology plays an essential role, serving as a repository of ideas as well as a tool for collaborative knowledge construction (Scardamalia and Bereiter, 1999). The development of the technological environment, known as the knowledge forum (KF) – known earlier as the computer-supported intentional learning environment (CSILE,) progressed side by side with development of the KB theory (Scardamalia and Bereiter, 2006). KF presents a visualization of the KB discourse in the form of notes that students create within the environment. In contrast with other electronic discussion forums, where posts are stored chronologically with older post buried

under newer ones, KF represents the notes spatially, with those that have been of greatest interest (e.g., with most links) placed nearer to the center of the screen, no matter when they were created. Teachers benefit from such a visualization (Scardamalia and Bereiter, 1999) by gaining insight into the ideas that students have been most focused upon, allowing them to tailor their feedback to students.

Through the use of shared community knowledge, KB makes the knowledge within the community public so that all members can participate in idea improvement. In KB, idea improvement happens when every member of the community takes actions, such as searching, commenting, synthesizing, critiquing, or modifying, on publicly shared notes (Scardamalia, 2002; Scardamalia and Bereiter, 2006). Through these actions, the whole community takes on the ownership of the ideas and become responsible for the quality of their progress. KF facilitates students' participation in a progressive discourse by providing scaffolds that help clarify the underlying cognitive processes of the composed notes (Scardamalia, 2002).

While FCL employs carefully designed collaborative structures as a scaffold for the exchange of ideas among students, KB uses the technology environment to scaffold continuous collaboration for improving the state of common knowledge (Scardamalia and Bereiter, 2008). Organization and the structure of collaborative groups are less important as long as the students identify opportunities when they can contribute to idea improvement (Bereiter and Scardamalia, 2003). Although not necessarily designed to support small-group collaboration, KF makes the content of small-groups' discourse available to the whole class so that the ideas remain public. In this way, the transition between small-group and whole-class discussions will take place more smoothly.

Challenges for Researchers

We have reviewed two research programs – FCL and KB – within a single overarching category that we refer to as the knowledge-community approach. While the research in FCL and KB may emphasize somewhat different pedagogical processes and epistemological frameworks, both are dedicated to the vision of a classroom where students collaborate in developing ideas, and rely on the contributions of their peers as a primary resource for learning. Both projects require a deep epistemological commitment from students and teachers that is consistent with autonomous lifelong learning, resulting in radical departure from conventional forms of classroom instruction.

While the ideas outlined above offer a compelling vision of knowledge communities in the classroom, and have been acclaimed within research reviews (e.g., Bransford *et al.*, 2000), a number of challenges have restricted the

capability of educational researchers to investigate this approach. Any researcher who wished to investigate using such a model would have to begin by establishing a relationship with a teacher who was willing to embrace the required epistemological commitments and radically revise his or her classroom methods. The researcher would have to work with the teacher for a substantial period of time, perhaps years, as they developed the pedagogical knowledge required to succeed with a knowledge-community approach. Finding such a teacher who is willing to overhaul his or her curriculum and teaching methods would be a daunting challenge, particularly at the secondary level where content expectations are high and high-stakes assessments are often mandated. Still, in order to investigate a knowledge-community model, or make any progress on the theory or methods, a researcher would need first to establish a suitable classroom ecology.

Moreover, the knowledge-community approach is extremely difficult for teachers to implement in advanced topics like secondary science, due to challenging conceptual domains and heavy content expectations. KB, for instance, discourages topic-centered discussions in favor of inquiry problems, and has been successful at the elementary level (Bereiter *et al.*, 1997; Zhang *et al.*, 2009). But with so many required topics in most secondary courses, teachers are not likely to have the freedom to pursue such an unstructured approach. Scardamalia (2002) suggests that it would likely require at least half the school year to establish a knowledge community within a teacher's classroom. Even with such a community running smoothly, it would be quite challenging to help students achieve a deep understanding of some advanced conceptual domains like those in secondary science courses. In one of the very few published studies of KB for secondary science, Van Aalst and Chan (2007) employed a hybrid approach where students in the course (physical geography) received some direct (i.e., conventional) instruction and then held online discussions using KF. The authors of this study acknowledged that they only addressed a portion of the KB principles, and therefore did not technically follow the KB method.

The difficulty of engaging in such research, while understandable, has made it difficult for the field as a whole to adopt the core ideas and contribute new theoretical ideas or practicable models (Slotta and Peters, 2008). Very few replications of the FCL approach have been published, for example, and even fewer theoretical extensions or adaptations. Yet, if there is to be progress in the learning sciences, it is vital that our powerful ideas and theoretical constructs are accessible to researchers in order to support replication and extension of those ideas. The few published studies conducted on FCL-inspired classrooms report challenges with regard to instructional design and implementation (e.g., Sherin *et al.*, 2004). Concerns have also been raised that any new designs could fall

short of the theoretical goals, simply imitating the technological and pedagogical scaffolds and ignoring the important underlying epistemologies (Whitcomb, 2004).

Given the challenges and potential pitfalls, it is important to consider how we can support researchers who are interested in pursuing the knowledge-community approach. What are some of the design principles for establishing a valid knowledge community? How can a new research approach be connected, theoretically, to the work of FCL or KB? Some guidance is provided by Scardamalia (2002), who advances the notion of collective cognitive responsibility, articulating 12 core ideas that are determinants of a learning community. While these clear theoretical descriptors are of value, there remains a need for a more practical set of guidelines or design principles and a wider theoretical space in which new approaches and technology environments could be situated (Lehrer and Schauble, 2006). One important contribution of this nature is provided by Bielaczyc (2006), who offers a framework for defining a social infrastructure, including specific design considerations and examples for each dimension of the framework.

Knowledge Communities for Twenty-First-Century Classrooms

The recent emergence of social networking and knowledge construction technologies, together with new philosophical movements that espouse a twenty-first-century knowledge society (Hargreaves, 2003; Tapscott, 2008) has raised awareness and interest in the knowledge community approach. There is a growing level of excitement about a general shift, in society, away from an industrial-era view of knowledge as commodity (i.e., that is required for the performance of some task) to one of an engaging, collaborative process that transcends specific disciplines (Hargreaves, 2003; Gilbert, 2005). Many authors and researchers argue that the workplace of the twenty-first century is becoming one of dynamic collaboration and real-time learning, requiring an educational system that emphasizes the development of critical thinking, problem solving, collaboration, and creative application of knowledge resources (Bruer, 1993; Scardamalia, 2002; Sawyer, 2007). This characterization of a new knowledge age is due in no small measure to the emergence of the Internet as an ubiquitous phenomenon in the lives of citizens – particularly students (Tapscott, 1998, 2008).

In recent years, the Web itself has become transformed with new genres of knowledge-oriented media that present exciting opportunities for education. Indeed, scholars have begun exploring how Web 2.0 can be used productively for educational purposes (Ullrich *et al.*, 2008; Greenhow *et al.*, in press). Wikis, for example, provide a technology scaffold that enables students to

collaboratively construct knowledge resources. Bryant *et al.* (2005) demonstrated that students involved in a wiki authoring task shift their learning focus from one of personal contribution to one of concern for the shared artifact. Students can also socially tag their wiki entries, resulting in folksonomies that make them accessible to a wider audience, which can encourage more thoughtful and conscientious contributions (Wheeler, *et al.*, 2008). It is evident that such technologies are well suited to scaffolding knowledge communities in the classroom. In particular, the powerful new mechanisms of social networking and data mining (i.e., of semantic metadata) could enable new ways of implementing knowledge bases, connecting students with ideas, and allowing knowledge to grow dynamically within the community (Slotta, in press).

A promising direction for future research would be the application of Web 2.0 technologies to scaffold the complex and emergent processes within a knowledge-community approach (Slotta and Peters, 2008). While KB does employ a technology environment for its community knowledge base, this technology does not digitally interconnect the resource materials with the student-generated ideas, nor does it employ semantic metadata to connect students with peers or reveal emerging themes within the knowledge base. In FCL, the elaborate patterns of jigsaw and cross talk provide a powerful means of ensuring that the knowledge and expertise of the community is distributed, and yet they were orchestrated with no help from scaffolding technologies. Thus, one means of making the knowledge-community approach more accessible to researchers would be to create a new generation of scaffolding technologies that are designed to support the distributed cognitive responsibility, the shared knowledge base, and the pedagogical patterns.

It will be important for researchers who are new to this domain to read carefully from the literature cited above. Establishing a knowledge community within the classroom involves changing the learning practices of all students and teachers, as well as their epistemological perspectives. This cannot be achieved simply by adding a wiki or a blog to a classroom. On the other hand, there is a very wide range of pedagogical applications that satisfy the three dimensions defined above – much wider than the specific terrain prescribed by FCL and KB – which could allow for new models to be investigated. Our own research group, for example, has begun developing a new model called knowledge community and inquiry (KCI) that makes the three dimensions above more accessible to secondary science teachers (Slotta and Peters, 2008; Peters and Slotta, in press).

See also: Computer Assisted Qualitative Data Analysis; Computer-Supported Collaborative Learning: Basic Concepts, Multiple Perspectives, and Emerging Trends; Critical

Thinking; Document Analysis; Evaluating E-Learning; Methods to Evaluate Technology; Peer and Self-assessment; Pragmatism; Social Interaction and Learning; Technology Supports for Acquiring Mathematics.

Bibliography

- Bereiter, C. and Scardamalia, M. (2000). Process and product in problem-based learning (pbl) research. In Evensen, D. H. and Hmelo, C. E. (eds.) *Problem-Based Learning, A Research Perspective on Learning Interactions*, pp 185–195. Mahwah, NJ: Erlbaum.
- Bereiter, C. and Scardamalia, M. (2003). Learning to work creatively with knowledge. In De Corte, E., Verschaffel, L., Entwistle, N., and van Merriënboer, J. (eds.) *Advances in Learning and Instruction Series: Powerful Learning Environments: Unraveling Basic Components and Dimensions*, pp 55–68. Oxford, UK: Elsevier Science.
- Bereiter, C., Scardamalia, M., Cassels, C., and Hewitt, J. (1997). Postmodernism, knowledge building, and elementary science. *Elementary School Journal* 97(4), 329–340.
- Bielaczyc, K. (2006). Designing social infrastructure: Critical issues in creating learning environments with technology. *Journal of the Learning Sciences* 15, 301–329.
- Bielaczyc, K. and Collins, A. (1999). Learning communities in classrooms: A reconceptualization of educational practice. In Reigeluth, C. (ed.) *Instructional Design Theories and Models*, pp 269–292. Mahwah, NJ: Erlbaum.
- Bielaczyc, K. and Collins, A. (2005). Technology as a catalyst for fostering knowledge-creating communities. In O'Donnell, A. M., Hmelo-Silver, C. E., and van der Linden, J. (eds.) *Collaborative Learning, Reasoning, and Technology*, pp 37–60. Mahwah, NJ: Erlbaum.
- Bransford, J. D., Brown, A. L., and Cocking, R. R. (eds.) (1999). *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Research Council.
- Bransford, J. D., Brown, A. L., and Cocking, R. R. (2000). *How people learn: brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences* 2(2), 141–178.
- Brown, A. L. (1997). Transforming schools into communities of thinking and learning about serious matters. *American Psychologist* 52(4), 399–413.
- Brown, A. L. and Campione, J. C. (1994). Guided discovery in a community of learners. In McGilly, K. (ed.) *Classroom Lessons: Integrating Cognitive Theory and Classroom Practice*, pp 229–272. Cambridge, MA: MIT Press/Bradford Books.
- Brown, A. L. and Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In Schauble, L. and Glaser, R. (eds.) *Innovation in Learning: New Environments for Education*, pp 289–325. Mahwah, NJ: Erlbaum.
- Bruer, J. T. (1993). *Schools for thought: A science of learning in the classroom*. MIT Press.
- Bruner, J. T. (1993). *Schools for Thought: A Science of Learning in the Classroom*. Cambridge, MA: MIT Press.
- Bryant, S., Forte, A., and Bruckman, A. (2005). Becoming wikipedia: Transformation of participation in a collaborative online encyclopedia. *Proceedings of GROUP: International Conference on Supporting Group Work*, pp 1–10. FL: Sanibel Island.
- Campione, J. C., Shapiro, A. M., and Brown, A. L. (1995). Forms of transfer in a community of learners: Flexible learning and understanding. In McKeough, A., Lupart, J., and Marini, A. (eds.) *Teaching for Transfer: Fostering Generalization in Learning*, pp 35–68. Hillsdale, NJ: Erlbaum.
- Collins, A., Brown, J. S., and Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In Resnick, L. B. (ed.) *Cognition and Instruction: Issues and Agendas*, pp 453–494. Hillsdale, NJ: Erlbaum.

- Collins, A., Joseph, D., and Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences* **13**(1), 15–42.
- Dewey, J. (1938/1997). *Experience and Education*. New York: Simon and Schuster.
- Driver, R. (1985). Changing perspectives on science lessons. In Bennett, N. and Desforges, C. (eds.) *Recent Advances in Classroom Research. British Journal of Psychology Monograph*.
- Garrison, D. R., Anderson, T., and Archer, W. (2000). Critical thinking in a text-based environment: Computer conferencing in higher education. *Internet and Higher Education* **11**(2), 1–14.
- Gilbert, J. (2005). *Catching the Knowledge Wave? The Knowledge Society and the Future of Education*. Wellington, New Zealand: NZCER.
- Greenhow, C., Robelia, B. and Hughes, J. (2009). Web 2.0 and classroom research: What path should we take now? *Educational Researcher*, **38**(4), 246–259.
- Hargreaves, A. (2003). *Teaching in the Knowledge Society*. New York: Teachers College Press.
- Hoadley, C. and Pea, R. (2002). Finding the ties that bind: Tools in support of a knowledge-building community. In Renninger, K. A. and Shumar, W. (eds.) *Building Virtual Communities: Learning and Change in Cyberspace*, pp 321–354. New York: Cambridge University Press.
- Lehrer, R. and Schauble, L. (2006). Scientific thinking and science literacy. In Damon, W., Lerner, R., Anne Renninger, K., and Sigel, I. E. (eds.) *Handbook of Child Psychology, Volume Four: Child Psychology in Practice*, 6th edn., pp 153–196. Hoboken, NJ: Wiley.
- Linn, M. C. and Hsi, S. (2000). *Computers, Teachers, Peers: Science Learning Partners*. Mahwah, NJ: Erlbaum.
- Peters, V. and Slotta, J. D. (in press). Learning and instruction in the Web 2 era: New opportunities for research. In Jacobson, M. (ed.) *Designs for Learning Environments of the Future: International Learning Sciences Theory and Research Perspectives*. Springer.
- Sawyer, R. K. (2007). *Group Genius: The Creative Power of Collaboration*. New York: Basic Books.
- Scardamalia, M. (2000). Can schools enter a knowledge society? In Selinger, M. and Wynn, J. (eds.) *Educational Technology and the Impact on Teaching and Learning*, pp 6–10. Abingdon: RM.
- Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In Smith, B. (ed.) *Liberal Education in a Knowledge Society*, pp 67–98. Chicago, IL: Open Court.
- Scardamalia, M. and Bereiter, C. (1999). Schools as knowledge building organizations. In Keating, D. and Hertzman, C. (eds.) *Today's Children, Tomorrow's Society: The Developmental Health and Wealth of Nations*, pp 274–289. New York: Guilford.
- Scardamalia, M. and Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In Sawyer, K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 97–118. New York: Cambridge University Press.
- Scardamalia, M. and Bereiter, C. (2008). “Fostering communities of learners” and “knowledge building”: An interrupted dialogue. In Campione, J. C., Metz, K. E., and Palincsar, A. S. (eds.) *Children's learning in the laboratory and in the classroom: Essays in honor of Ann Brown*, pp 197–212. Mahwah, NJ: Erlbaum.
- Sherin, M. G., Mendez, E. P., and Louis, D. A. (2004). A discipline apart: The challenges of ‘fostering a community of learners’ in a mathematics classroom. *Journal of Curriculum Studies* **36**(2), 207–232.
- Shulman, L. S. and Sherin, M. G. (2004). Fostering communities of teachers as learners: Disciplinary perspectives. *Journal of Curriculum Studies* **36**(2), 135–140.
- Slotta, J. D. (in press). Evolving the classroom of the future: The interplay between pedagogy, technology and community. In Fischer, F. and Dillenbourg, P. (eds.) *Classrooms of the Future*. Springer.
- Slotta, J. D. and Linn, M. C. (2009). *WISE Science: Inquiry and the Internet in the Science Classroom*. New York: Teachers College Press.
- Slotta, J. D. and Peters, V. (2008). A Blended model for knowledge communities: Embedding scaffolded inquiry. *Proceedings of the International Conference of the Learning Sciences*. Utrecht.
- Stahl, G. (2000). A model of collaborative knowledge-building. In Fishman, B. and O'Connor-Divelbiss, S. (eds.) *Fourth International Conference of the Learning Sciences*, pp 70–77. Mahwah, NJ: Erlbaum.
- Tapscott, D. (1998). *Growing Up Digital. The Rise of the Net Generation*. New York: McGraw-Hill.
- Tapscott, D. (2008). *Grown Up Digital: How the Net Generation is Changing Your World*. Toronto: McGraw-Hill.
- Ullrich, C., Borau, K., Luo, H., Tan, X., Shen, L., and Shen, R. (2008). Why web 2.0 is good for learning and for research: Principles and prototypes. *Proceeding of the 17th international Conference on World Wide Web*, Beijing, China, April 21–25, 2008. WWW '08. ACM, New York, NY, 705–714. DOI: <http://doi.acm.org/10.1145/1367497.1367593>.
- van Aalst, J. and Chan, C. K. (2007). Student-directed assessment of knowledge building using electronic portfolios. *Journal of the Learning Sciences* **16**(2), 175–220.
- Wheeler, S., Yeomans, P., and Wheeler, D. (2008). The good, the bad and the wiki: Evaluating student-generated content for collaborative learning. *British Journal of Educational Technology* **39**(6), 987–995.
- Whitcomb, J. A. (2004). Dilemmas of design and predicaments of practice: Adapting the ‘fostering a community of learners’ model in the secondary school english language arts classrooms. *Journal of Curriculum Studies* **36**(2), 183–206.
- Zhang, J., Scardamalia, M., Reeve, R., and Messina, R. (2009). Designs for collective cognitive responsibility in knowledge-building communities. *Journal of the Learning Sciences* **18**(1), 7–44.

Further Reading

- Bielaczyc, K. and Collins, A. (2007). Design research: Foundational perspectives, critical tensions, and arenas for action. In Campione, J., Metz, K., and Palincsar, A. M. (eds.) *Children's Learning in Laboratory and Classroom Contexts: Essays in Honor of Ann Brown*, pp 89–112. New York: Routledge.
- Linn, M. C. and Eylon, B.-S. (2006). Science education: Integrating views of learning and instruction. In Alexander, P. A. and Winne, P. H. (eds.) *Handbook of Educational Psychology*, 2nd edn, pp 511–544. Mahwah, NJ: Erlbaum.
- Roschelle, J., Penuel, W. R., and Shechtman, N. (2006). Co-design of innovations with teachers: Definition and dynamics. *Proceedings of the Biennial International Conference of the Learning Sciences*, 606–612. Bloomington, IN.

TECHNOLOGY AND LEARNING – SUPPORTS FOR SUBJECT MATTER LEARNING

Contents

Technology Supports for Science Learning

Technology Supports for Second Language Learning

Technology Supports for Science Learning

C Delgado, University of Texas, Austin, TX, USA

J Krajcik, University of Michigan, Ann Arbor, MI, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Asynchronous – Occurring at different times.

Embedded assessment – Assessment integrated into everyday instructional activities.

Multimedia – Involving several media, for example, video clips, images and text.

Multimodal – Involving several modes, for example, visual, auditory and kinesthetic.

Professional development – Opportunities for training and learning for in-service teachers.

Stereoscopic – Three-dimensional image generated by simultaneously looking at two two-dimensional images.

Titration – Volumetric analysis in which a solution of known concentration is used to determine the unknown concentration of a second solution.

Technology, in the broadest sense, is the application of systematic or scientific knowledge for practical purposes. This article considers only computer-based technological tools for science learning. Technology has the potential to transform science education, by supporting collaborative, inquiry-oriented learning. Technology can increase the range of permissible phenomena to study, enable and promote greater social communication and interaction, provide multimodal and dynamic representations, structure and support student investigations, and support effective instruction and assessment. Technology can also help school science become more authentic – that is, more like the practice of scientists – and can facilitate the learning of science outside of the school setting, thus supporting lifelong learning. Policy reports have proposed that educational technology can provide an unusually high return on investment both in

economic and social terms (e.g., President's Council of Advisors on Science and Technology, 1997). In combination with professional development to support teachers, and innovative curriculum materials developed around clearly defined learning goals, technology may support improvements in student outcomes.

Educational technology can be classified into two categories: amplifiers and transformers (Rubin, 1996). Technology supports for science learning are not neutral, but dependent on the learning theory underlying their design. For instance, a behaviorist, learning-to-mastery paradigm may lead to learning technology that emphasizes memorization and the transmission of knowledge from teacher to student. By increasing the amount of practice with basic mathematical facts, the Math Blaster video game acts as an amplifier. The use of multimedia presentations and handouts for students can amplify the amount of material that can be covered as compared to a traditional class using a blackboard. A constructivist paradigm may, instead, lead to learning technology that engages the user in actively constructing knowledge. Constructivism values students' direct and tool-mediated interaction with their physical and social environment – reflectively building upon their prior knowledge – through the use of engaging and contextualized activities. The tenets of this learning theory are consistent with inquiry-oriented science teaching, in which students pose significant questions, design and conduct experiments to collaboratively address these questions, and communicate their findings. The use of technology in a constructivist manner is more likely to be transformative, and may result in higher student learning gains, even when using the same type of technology (e.g., personal response systems, see Judson and Sawada, 2002).

There are several general categories of educational technology use in science learning that are consistent with constructivism (see Table 1). As with any taxonomy,

Table 1 Some categories of technology use that support constructivist, inquiry-oriented science teaching and learning

<i>Category</i>	<i>Description</i>	<i>Examples</i>
<i>Real-time data acquisition and graphing</i>	Probes for pH, temperature, voltage, pressure, dissolved oxygen level, motion, etc., that connect to a computer, graphing calculator, or handheld device	Vernier probes for use with TI 83 plus calculator-based laboratory; Pasco PASPORT and Xplorer handheld probes
<i>Simulations</i>	Computer-generated versions of real-world objects and phenomena that often allow the user to control settings.	Virtual frog dissection kit (http://froggy.lbl.gov/virtual/); chemical kinetics simulation (http://www.chem.uci.edu/undergrad/applets/sim/simulation.htm); Molecular Workbench (http://workbench.concord.org/)
<i>Virtual communities, social networking sites, and collaborative websites</i>	Websites that support blogging, chat, and posting of user-created materials	http://www.classroom20.com/ ; Project FeederWatch http://www.birds.cornell.edu/pfw/ ; CASES (cases.soe.umich.edu); wikipages; http://teachertube.com/
<i>Virtual laboratories</i>	Internet environment allows students to design and carry out experiments	http://www.chem.ox.ac.uk/vrchemistry , http://learn.arc.nasa.gov/vlab/features.html
<i>Remote access to instruments</i>	Students interact with and manipulate instruments that are not available at their school, over the Internet	University of North Carolina atomic force microscope used by rural and suburban high schools in North Carolina (Jones et al., 2003); weather data (Songer, 1996; Edelson et al., 1999)
<i>Modeling software and microworlds</i>	Computer-programming language or software allowing users to create a world	Logo programming languages; Model-It
<i>In-class interaction technology</i>	Devices that enhance social interaction in the classroom	Interactive whiteboards; student-response systems (aka clickers)
<i>Inquiry-structuring software and sites</i>	Software or websites that help teachers and learners plan and manage long, complex investigations	WISE wise.berkeley.edu ; IdeaKeeper http://biologica.concord.org/
<i>Learning environments</i>	Web-based environments integrating curriculum, multimedia, simulations, support for learners, embedded assessment, etc.	WISE wise.berkeley.edu ; BioLogica http://biologica.concord.org/

these categories are conceptually useful but not absolute. Several types of technology can support each aspect of constructivist, inquiry-oriented science learning and teaching, as described below.

Technology Can Promote Interaction with the Physical Environment

Phenomena

Many scientific phenomena are too dangerous or too expensive to include in science classes at the primary or secondary level. Technology can make these phenomena accessible for students through online simulations, virtual laboratories, or remote access to tools and sensors such as atomic force microscopes or telescopes. Simulations are computer-generated versions of real-world objects and phenomena that often allow the user to control settings. Simulations focus on one aspect of reality and portray it so that it is easier to grasp. For instance, the Molecular Workbench simulations explicitly show charges on particles that facilitate student understanding of the role of forces and interactions in self-assembly. Technology can render visible phenomena that are too small or too fast to observe, such as chemical reactions or a gas at the

molecular level. Many such simulations are available online (see **Table 1**) or featured in companion compact disks (CDs) packaged with textbooks. Likewise, simulations can model phenomena that are too large to observe all at once, such as global weather patterns. Events that occur very slowly, like geological processes, can also be modeled using technology. Simulations can help overcome ethical obstacles, such as those involved in the dissection of live animals.

Technology can also increase the range of actual phenomena that students can experience. The use of video and Internet allows high school teachers to incorporate dangerous chemical reactions into their classroom (see virtual laboratories in **Table 1**). Another way of incorporating phenomena that are too dangerous or expensive for primary or secondary education is through the remote access to sophisticated scientific tools. In one such experience, high school students were able to examine viruses through remote access to an atomic force microscope at the University of North Carolina at Chapel Hill, which provided visual and tactile feedback in real time (Jones et al., 2003). In the Kids as Global Scientists and WorldWatcher curricula, students interacted with actual global and local weather data collected by remote sensors, via software that represented the data in an easy-to-understand format

(e.g., Songer, 1996; Edelson *et al.*, 1999). Thus, the use of technology in education can increase the range of available phenomena with which students can interact.

Supporting Student Investigations

A recent National Research Council (NRC; Singer *et al.*, 2006) report states that laboratory experiences are a fundamental part of science education. Laboratory experiences that involve direct interaction with the physical world or with real data can help students understand the relationship between research and theories, and learn about science as both process and content (Singer *et al.*, 2006). The College Board also supports actual laboratory experiences, offering only conditional authorization for Advanced Placement programs that use virtual laboratories. The use of standalone, computer- or calculator-based data-acquisition tools or probeware simplifies the collection of measurements of pH, gas pressure, dissolved oxygen level, temperature, and other variables. Probeware can help students investigate questions that they otherwise could not address; for instance, collecting data points separated by very short time intervals, or simultaneously collecting data on several variables. Variables, such as dissolved oxygen content or pH, that are traditionally measured using titrations or comparison to standards can more easily be studied experimentally by young students using probeware. Ultrasonic motion-detectors, or rangers, let students generate their own distance, velocity, and acceleration graphs by walking toward or away from a wall, allowing them to study the relationship between these variables without the use of calculus; motion-detectors also open the possibility for experiential, kinesthetic learning experiences that are engaging to many students. The NRC report recommends that laboratory experiences go beyond making observations and collecting and analyzing data, to include posing questions, designing experiments, creating explanatory models, and communicating findings (Singer *et al.*, 2006). Several of these aspects can also be supported by technology, as described below.

Dynamic Representations

Simulations can demonstrate the dynamic nature of processes much more effectively than can static images in books. Technology can make abstract processes concrete through effective visual and multimedia representations. SimCalc is a tool for the learning of the mathematics of change and variation that allows even elementary school students to begin to explore and understand elements of calculus in the context of motion (Roschelle *et al.*, 2000). Dynamically linked multiple representations such as velocity graphs, position graphs, and animations can help learners make connections in building conceptual understanding, when used appropriately (see Ainsworth, 2008).

Simulations for chemical equilibria can graphically represent the forward and reverse reactions that occur at equal rates, combating the common student alternate conception that no process takes place at equilibrium. The simulations can generate accompanying charts displaying the values of variables, helping the learner link between the atomic-molecular and macroscopic levels (see, e.g., Nakhleh and Postek, 2008).

Constructing Models and Representations

While simulations allow the user to manipulate settings, modeling software and programming languages permit the user even greater latitude to model the phenomenon of their choice. The Logo family of constructionist programming languages enables computer environments in which the learner actually builds computer representations of a system of interest (see, e.g., Resnick, 1994). The user determines the rules governing the behavior of one or many agents. Software such as Star Logo enables students to model emergent phenomena such as birds' flocking behavior or traffic jams by defining the rules governing the behavior of a large number of agents. Modeling software such as Model-It (Jackson *et al.*, 2000) allows the user to input variables along with the functional relationships between variables, in order to build – from the ground up – a model of a system such as a river ecosystem. By viewing output graphs, students can explore the behavior of an entire system, and the effects of changing one or more variables.

Technology Can Promote Interaction with the Social Environment

Technology can also facilitate social interactions. In the last decades, science has been envisioned as a socially mediated process (e.g., Kuhn, 1962; Latour and Woolgar, 1986). Similarly, current conceptions of science learning stress the role of communities and social interaction (e.g., Bransford *et al.*, 1999). Innovative curriculum programs such as Kids as Global Scientists and BioKIDS (Songer, 1996; Songer, 2006) have included student–scientist communications via e-mail, student–student collaborations across even widely distant locations, and asynchronous communications using bulletin boards. The anonymity afforded by electronic communications can empower participants who otherwise might not engage in discussions. Other programs – such as Project FeederWatch – use the data from amateur bird-watchers all over the USA to track populations and migrations, and can be incorporated into the science curriculum. Such uses of technology enable the compilation of large-scale databases that allow students to track processes that would otherwise not be possible, while supporting increased collaboration.

While some uses of computer technology can be isolating, experiences involving visualization, modeling, and simulation can provide opportunities for interaction among students and teachers. Technological tools such as interactive whiteboards – which are systems that combine a projector and a large, touch-sensitive screen to display and facilitate manipulation of a computer screen – can enhance classroom interactions. Interactive whiteboards allow teachers to instantly add notes to a presentation, or record students' ideas, and then save them in a digital format. Web pages that support collaborative creation and editing, such as Wiki pages, can also increase social interaction. Wiki pages can pool the expertise of many individuals without the need for face-to-face interaction, allowing for asynchronous and long-term collaborations. Both the process and the product of building a Wiki page can support science learning. Student-response systems, in which students use an infrared remote control or clicker to register a response (e.g., to a multiple-choice question), allow for real-time formative assessment that teachers can use to modify instruction. These systems can also facilitate social interaction even in large college lecture halls, if the instructor encourages small-group discussions about the question. Some computer-based environments for science learning specifically encourage and rely on small-group interaction (e.g., ChemSense – Michalchik *et al.*, 2008).

Many of the above applications depend on student access to a laptop or desktop computer; however, the use of portable, handheld devices such as MP3 players, personal digital assistants (PDAs), and probeware allows for autonomous learning and can increase motivation, collaboration, and communication. One example is the use of handheld PDAs in the BioKIDS curriculum to help students collect biodiversity data in their schoolyards, which were then aggregated over a classroom, a school, or a state. Another example is the use of standalone or calculator-based handheld probeware for data collection in the lab or in the field.

Technology Can Support Reflective Construction of Knowledge

Reducing the Load on Working Memory

The effective use of educational technology can support learning by structuring tasks for students. Humans have a limited working-memory capacity, and technology can help overcome this limitation in several ways. By focusing only on relevant aspects, simulations can simplify a complex natural phenomenon. Multimodal simulations that employ both visual and phonological channels (e.g., text and images alongside narration) can reduce working memory load (Mayer, 2001). In addition, technology can be used to handle routine tasks, such as graphing or simple forms of data

analysis, in order to allow students to concentrate on the relevant concept. This offloading must be used judiciously, however, as it can keep students from learning graphing and analysis processes that are basic to science.

Computer technology can be used also to speed up student workflows and provide immediate feedback. For instance, the use of computer- or calculator-based probeware in the laboratory allows for instantaneous, real-time graphing of data, thus reducing students' need to coordinate the lab experience itself with subsequent data analysis conducted over a long period of time. By allowing students to concentrate on the construction of meaning rather than on recording data in their notes, interactive whiteboards that allow teachers to save and post or reproduce class notes can also reduce the load on working memory. Printouts of computer-based presentations can help students concentrate on taking their notes, to help them make sense of class, rather than on copying the information being delivered.

Managing Time and Complexity

Software tools can help structure a complex task and support planning. Science-learning environments like the Web-based Inquiry Science Environment (WISE) (Williams and Linn, 2002) and software such as Idea Keeper (Quintana *et al.*, 2005) provide structure for students during complex investigations through the use of checklists, prompts and reminders, graphical organizers, and other means of tracking progress. Progress can also be tracked by digital portfolios. Restricting the problem space or simplifying the phenomenon of interest can focus the efforts of students. On the other hand, software tools can problematize student learning – making it more complex but also more productive – by focusing students' attention on unresolved aspects of the investigation or requesting rationales for decisions (Reiser, 2004); for instance, by prompting users to analyze whether parts of the arguments they have constructed are evidence or claims, and whether the evidence supports or disconfirms the claim.

Supporting Teachers

Research has found that quality of instruction is a major determinant of student outcomes (e.g., Wright *et al.*, 1997). Technology can support teachers so that they can help students learn better. Science-learning environments can have embedded formative assessments that teachers can access immediately in order to gauge the effectiveness of their instruction and modify their plans accordingly; student-response systems – used frequently in undergraduate science classes – can achieve the same end. Teachers can search for laboratory lesson plans and safety

data on information-access environments such as the Internet or digital encyclopedias.

Online professional development opportunities for teachers have also been employed, to improve their pedagogical capacity as well as science-content knowledge. CASES (Davis *et al.*, 2004) is an online system for elementary science teachers that includes a unit-plan library, resource library, and descriptions of how to teach, modify, and adapt lesson plans. Virtual communities and social networking sites allow teachers and educators to share ideas, hold discussions, post materials such as videos and lesson plans, and easily search for specific resources that they need.

Technology Can Support Contextualization

Authenticity

The Kids as Global Scientists and World Watcher curricular projects access the actual data of professional scientists, using the Internet and custom software. Other uses of technology involve students in generating genuine scientific data, as in the Project FeederWatch, BioKIDS, and Global Learning and Observations to Benefit the Environment (GLOBE) projects. Remote access to sophisticated tools such as telescopes and microscopes allows students to use the instruments of practicing scientists. All of these factors increase the relevance and authenticity of science learning for students, in comparison to traditional instruction based on books and simple, often decontextualized laboratory experiments. The use of computers, simulations, communications technology, and the Internet is pervasive in professional science; when students learn to use these technologies, they are learning to use some of the tools of scientists. Students can use the Web to locate information, collect data, analyze and display it, and communicate their findings and understanding using technology that is similar to that in use by scientists.

Relevance to Everyday Life

These uses of technology have opened up the possibility of addressing topics that are relevant to students but that were previously too difficult. For instance, the ideas of calculus that were previously taught to talented high school students and undergraduates using complex mathematical notation are now being introduced to much younger children in an intuitive and graphic manner (Roschelle *et al.*, 2000). The behavior of traffic jams is easy to study and comprehend in a computerized environment, but difficult otherwise; this phenomenon is of interest to any city-dweller. Technology has also been

used to support studying the school grounds as an ecosystem (BioKIDS) or a local stream (using Model-It) – also topics of relevance for students.

Technology Supports Increased Access to Information

The explosion of the information available on the Internet is impacting education in multiple ways. The Internet facilitates teachers' searches for factual information (through both general sites like Wikipedia and more targeted sites such as those of National Aeronautics and Space Administration (NASA), Public Broadcasting Service (PBS), or science magazines online) and for educational activities, research, and ideas (e.g., the ERIC database and teacher virtual communities). The Internet lets students access textbook-publishers' companion websites featuring tutorials, practice problems, interactive or stereoscopic figures, simulations, and even the electronic versions of textbooks. Course websites may include podcasts of lectures, notes or multimedia presentations, online quizzes, and links to useful websites. Databases and online journals increase the pace of scientific progress and of science learning by allowing quick and convenient access to journals and papers, along with earlier access to papers published on the Web and posted on the journal website in advance of publication in hard copy.

Integration of Features Can Transform Education

While many of the features mentioned above amplify existing characteristics of science teaching without technology, taken together they can be transformative. Learning environments that combine innovative curriculum and effective use of technology can be qualitatively different from curricula that do not incorporate technology, allowing real-time and asynchronous collaborations among students, teachers, and scientists, focused on otherwise inaccessible phenomena. For instance, the WISE learning environments support students in collecting, synthesizing, and analyzing information from a variety of authentic sources and simulations, to address engaging questions such as, "Does light go on forever?" WISE activities also build students' argumentation and communication abilities. Other modules use multimedia and simulations to address complex questions, with planning support in the form of checklists, prompts, graphic organizers, and so on. Such inquiry-oriented instructional activities are consistent with constructivist learning theory and can lead to deeper, more flexible understanding. Technology affords immediate feedback and multiple opportunities to make students' thinking visible, enabling science teaching that is more responsive to students' needs.

Pitfalls and Challenges to Implementation of Technology

Some challenges to the implementation of technology in the science classroom are general: for example, technology is expensive, requires maintenance and technical support, teachers may need professional development in order to use technology effectively, and so on. Some pitfalls are specific to the use of technology in the science classroom. For instance, simulations that portray relationships between variables as smooth curves may misrepresent the nature of empirical work in science in a way that a laboratory experiment would not. Science teachers may struggle to simultaneously learn to use the technology and learn to teach in inquiry-oriented ways. The opportunity costs of adopting technology for science education must also be considered. Funding for technology may be in competition with funding for laboratory supplies, and time spent on computers may be time not spent in the laboratory or other important activities.

The Role of Technology in Increasing Equity in Education

The costs that a school district incurs in adopting technology are a challenge to equity. Lower-income areas in the United States tend to have poorly funded schools, with smaller percentages of certified teachers teaching in a subject that they are trained to teach, and lower budgets for technology. On the other hand, once in place, technology may offer underprivileged students opportunities that are currently unavailable to them. Small, underfunded, or rural school districts may benefit from online courses in topics that enrich their curriculum at a lower cost than offering the course themselves. Videoconferencing, likewise, can support collaborations or enable access to experts.

Adaptive technology for disabled students can help their access to education, in general. This article has been written at an ergo pod: an accessibility workstation with dictation software, image magnification, screen reader for the visually impaired, adjustable-height work surface, etc. There has been some development of software and equipment specifically for visually impaired students to carry out laboratory chemistry experiments, for instance, colorimeters with an auditory output to help in titrations.

Emerging Applications and Future Directions

The Web 2.0 generation of Internet sites that allow users to create, post, and share content (e.g., YouTube, Facebook,

and Second Life) may affect science learning in interesting ways. Teacher communities allow for discussions, sharing of resources, and collaborative projects. Games such as Spore are not explicitly meant for the education market but embody the principles of evolution and genetics. Recent research at the University of Wisconsin at Madison has found that role-playing games such as World of Warcraft require and build problem-solving skills that may transfer to the science classroom (Steinkuehler and Duncan, 2008). Increasingly powerful and portable smart phones with digital cameras, probes, and Internet connectivity may enable students to pursue science topics in which they are interested outside of the school setting in ways we cannot imagine at present. Cognitive tutors – which track student knowledge and the knowledge space in delivering highly individualized instruction – are currently restricted to well-defined topics such as mathematics or computer programming, and do not, at present, support the contextualized and open-ended exploration typical of inquiry-oriented science teaching. However, the possibilities of individualized instruction delivered via future generations of cognitive tutors that can address more complex domains are tantalizing. Customizing instruction and assessment for groups of students with particular needs (English language learners, the visually impaired, and so on) is another possible future direction for technology in science education.

Large-Scale and Systemic Impact

The scaling up and sustainability of promising educational innovations are an important challenge to the field of science education. The Internet offers a platform that increases access to and the longevity of learning environments. Ongoing projects to digitize books will increase the number of sources of information available to learners, but will also transform the experience of finding information in books, with features such as keyword searching in a single book or across entire databases. As technology allows learners to have access to tools and information independent of their physical location, life-long and self-directed learning will become more feasible. Finally, one can foresee feedback cycles that lead to further leveraging of technology to support science learning. As students routinely use Web 2.0 applications such as YouTube and Facebook, and virtual environments such as Second Life, they construct novel digital literacies. These technological abilities may support the development of future technologies to support science learning in ways as yet unforeseen.

See also: Educational Reform; Formative Assessment; Knowledge Communities in the Classroom; Technology Supports for Acquiring Mathematics; Technology Supports for Lifelong Learning.

Bibliography

- Ainsworth, S. (2008). The educational value of multiple-representations when learning complex scientific concepts. In Gilbert, J. K., Reiner, M., and Nakhleh, M. (eds.) *Visualization: Theory and Practice in Science Education*, pp 191–208. Dordrecht: Springer.
- Bransford, J., Brown, A., and Cocking, R. (eds.) (1999). *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academy Press.
- Davis, E. A., Smithey, J., and Petish, D. (2004). Designing an online learning environment for new elementary science teachers: Supports for learning to teach. In Kafai, Y. B., Sandoval, W. A., Enyedy, N., Nixon, A. S., and Herrera, F. (eds.) *Proceedings of the 6th International Conference of the Learning Sciences, ICLS 2004*, p 594. Mahwah, NJ: Erlbaum.
- Edelson, D., Gordin, D., and Pea, R. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *Journal of the Learning Sciences* 8(3/4), 391–450.
- Jackson, S., Krajcik, J. S., and Soloway, E. (2000). Model-It: A design retrospective. In Jacobson, M. and Kozma, R. (eds.) *Innovations in Science and Mathematics Education: Advanced Designs for the Technologies of Learning*, pp 77–116. Hillsdale, NJ: Erlbaum.
- Jones, M. G., Andre, T., Superfine, R., and Taylor, R. (2003). Learning at the nanoscale: The impact of students' use of remote microscopy on concepts of viruses, scale, and microscopy. *Journal of Research in Science Teaching* 40(3), 303–322.
- Judson, E. and Sawada, D. (2002). Learning from past and present: Electronic response systems in college lecture halls. *Journal of Computers in Mathematics and Science Teaching* 21(2), 167–181.
- Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.
- Latour, B. and Woolgar, S. (1986). An anthropologist visits the laboratory. In Latour, B. and Woolgar, S. (eds.) *Laboratory Life: The Construction of Scientific Facts*, pp 43–90. Princeton, NJ: Princeton University Press.
- Mayer, R. (2001). *Multimedia Learning*. New York: Cambridge University Press.
- Michalchik, V., Rosenquist, A., Kozma, R., Kreikemeier, P., and Schank, P. (2008). Representational resources for constructing shared understandings in the high school chemistry classroom. In Gilbert, J. K., Reiner, M., and Nakhleh, M. (eds.) *Visualization: Theory and Practice in Science Education*, pp 233–282. Dordrecht: Springer.
- Nakhleh, M. B. and Postek, B. (2008). Learning chemistry using multiple external representations. In Gilbert, J. K., Reiner, M., and Nakhleh, M. (eds.) *Visualization: Theory and Practice in Science Education*, pp 209–231. Dordrecht: Springer.
- President's Council of Advisors on Science and Technology (Panel on Educational Technology, Shaw, D.E. (ed.)). (1997). *Report to the President on the Use of Technology to Strengthen K-12 Education in the United States*. Washington, DC: Office of Science and Technology Policy.
- Quintana, C., Zhang, M., and Krajcik, J. S. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist* 40(4), 235–244.
- Reiser, B. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences* 13(3), 273–304.
- Resnick, M. (1994). *Turtles, Termites, and Traffic Jams: Explorations in Massively Parallel Microworlds*. Cambridge, MA: MIT Press.
- Roschelle, J., Kaput, J., and Stroup, W. (2000). SimCalc: Accelerating students' engagement with the mathematics of change. In Jacobson, M. J. and Kozma, R. B. (eds.) *Innovations in Science and Mathematics Education: Advanced Designs for the Technologies of Learning*, pp 47–75. Hillsdale, NJ: Erlbaum.
- Rubin, A. (1996). Educational technology: Support for inquiry-based learning. In Fulton, K., Feldman, A., Wasser, J. D., et al. (eds.) *Technology Infusion and School Change: Perspectives and Practices. Model Schools Partnership Research Monograph*. Cambridge, MA: TERC. <http://rapb.mspnet.org/index.cfm/8353>.
- Singer, R. S., Hilton, K. L., and Schweingruber, H. A. (eds.) (2006). *America's Lab Report: Investigations in High School Science*. Washington, DC: National Academies Press.
- Songer, N. B. (1996). Exploring learning opportunities in coordinated network-enhanced classrooms: A case of kids as global scientists. *Journal of the Learning Sciences* 5(4), 297–327.
- Songer, N. B. (2006). BioKIDS: An animated conversation on the development of curricular activity structures for inquiry science. In Sawyer, R. K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 355–369. New York: Cambridge University Press.
- Steinkuehler, C. and Duncan, S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science Education and Technology* 17(6), 530–543.
- Williams, M. and Linn, M. C. (2002). WISE inquiry in fifth grade biology. *Research in Science Education* 32(4), 415–436.
- Wright, S. P., Horn, S. P., and Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education* 11(1), 57–67.

Further Reading

- Blumenfeld, P., Fishman, B., Krajcik, J. S., Marx, R. W., and Soloway, E. (2000). Creating usable innovations in systemic reform: Scaling up technology-embedded project-based science in urban schools. *Educational Psychologist* 35(3), 149–164.
- Corley, A.-M. (2009). A cell-phone microscope for disease detection. *Technology Review*. Retrieved from <http://www.technologyreview.com/biomedicine/23059/>
- Fagen, A. P., Crouch, C. H., and Mazur, E. (2002). Peer instruction: Results from a range of classrooms. *Physics Teacher* 40(4), 206–209 (personal response systems).
- Finarelli, M. G. (1998). GLOBE: A worldwide environmental science and education partnership. *Journal of Science Education and Technology* 7(1), 77–84.
- Krajcik, J. S. and Blumenfeld, P. (2006). Project-based learning. In Sawyer, R. K. (ed.) *The Cambridge Handbook of the Learning Sciences*, pp 317–334. New York: Cambridge University Press.
- Linn, M. C., Davis, E. A., and Bell, P. (eds.) (2004). *Internet Environments for Science Education*. Mahwah, NJ: Erlbaum.
- NSF Task Force on Cyberlearning (Borgman, C. L., chair). (2008). Fostering learning in the networked world: The cyberlearning opportunity and challenge. <http://www.nsf.gov/pubs/2008/nsf08204/nsf08204.pdf> (accessed August 2009).
- Quintana, C., Reiser, B. J., Davis, E. A., et al. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences* 13(3), 337–386.
- Rieber, L. P. (2005). Multimedia learning in games, simulations, and microworlds. In Mayer, R. (ed.) *The Cambridge Handbook of Multimedia Learning*, pp 549–567. New York: Cambridge University Press.
- Roschelle, J. M., Pea, R. D., Hoadley, C. M., Gordin, D. N., and Means, B. M. (2000). Changing how and what children learn in school with computer-based technologies. *The Future of Children* 10(2), 76–101.

Relevant Websites

- <http://www.biokids.umich.edu> – BioKIDS.
- <http://biologica.concord.org> – BioLogica.
- <http://cases.soe.umich.edu> – CASES.
- <http://www.chem.uci.edu/undergrad/applets/> - Chemistry simulations.
- <http://www.ucmp.berkeley.edu/geology/tectonics.html> – Geology (Plate tectonics).
- <http://mw.concord.org> – Molecular Workbench.
- <http://www.birds.cornell.edu/pfw/> – Project FeederWatch.
- <http://www.globe.gov> – Project GLOBE.
- <http://froggy.lbl.gov> – Virtual frog dissection.
- <http://wise.berkeley.edu> – WISE: Web-based Inquiry Science Environment.
- <http://www.geode.northwestern.edu> – WorldWatcher.

Technology Supports for Second Language Learning

D Nunan, University of Hong Kong, Hong Kong, SAR, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

Glossary

e-learning – Learning through Internet-based instruction rather than in face-to-face instruction.

Foreign language learner – An individual who is learning a language which is not the medium of communication in the country where the learner resides, for example, an individual learning Chinese in the United Kingdom would be said to be learning Chinese as a foreign language.

Globalization – The interlinking of economies and markets and the spread of corporations worldwide.

Internet – An international network linking computers worldwide for the purposes of exchanging electronic messages and documents.

Learning management tool – The use of technology for managing the learning process rather than transmitting content.

Second language learner – An individual who is learning a language other than their first language in a country where the language being learned is widely used for communication, for example, a first language speaker of Punjabi learning English in Canada would be said to be learning English as a second language.

Technology – The application of scientific and technical systems for practical purposes.

World Wide Web – An information system that allows documents to be linked together through the Internet.

Technology has supported second-language pedagogy for many years. In the 1960s and 1970s, language laboratories provided drill-based practice to support audiolingualism – the dominant instructional approach of the day (Richards and Rodgers, 1986). In the 1980s, videotapes became increasingly common, and computers made their first appearance in the second-language classroom (Higgins and Johns, 1984). However, it was in the 1990s, with the explosion in Internet use and the emergence of the World Wide Web, that technology became ubiquitous (Warschauer, 1999; Beatty, 2003).

The first part of this article examines the roles played by technology in second-language learning, focusing in particular on technology as a provider of content, as a learning management tool and as a communication tool. We then examine models of classroom organization, and

suggest that technology has forced us to redefine our concept of the classroom. The final major section of the article enumerates seven areas of second-language pedagogy where technology is demonstrably superior to face-to-face instruction.

The Impetus for Increased Technology Use in Second-Language Classrooms

Technology has become so pervasive that it is not longer appropriate to see it as playing a supporting role to face-to-face instruction. A major stimulus for the rise of technology in second-language classrooms has been the explosion in the demand for language training. This demand is principally, but not exclusively, for English and is a byproduct of globalization. In a recent survey of more than 25 000 employees of global corporations, the percentage of respondents who said that English was either critical or important for success in their current positions increased each quarter over an 18-month period. **Figure 1** shows the growth in the perceived importance and use of English over the period from June 2003 to December 2004. During this time, those who reported that English was critical for their jobs rose from just over 50% to 80%. The percentage of global corporation employees who used English daily in their jobs rose from 35% to 60% (**Figure 2**).

The vast majority of those surveyed reported that their English-language proficiency was insufficient for them to be successful in their current jobs (only 9% said their English-language proficiency was sufficient to do their current jobs). Under these circumstances, the explosion in the demand for English instruction is quite understandable. This demand overwhelmed traditional face-to-face providers, and many workplace human resource managers looked to technology, particularly the Internet, to provide English-language instruction. Private corporations, such as Global English, are meeting a demand that is simply beyond the reach of conventional second-language classrooms.

Roles of Technology in Second-Language Classrooms

The three major roles played by technology in classrooms are as a carrier of content and an instructional tool, as a learning management tool, and as a communication tool.



Figure 1 Percentage of respondents who report that English is critical for their job. From Nunan, D. (2005). The evolution of technology and the value of online English language learning, p. 5, *White Paper*, Global English Corporation, San Francisco, CA, USA.

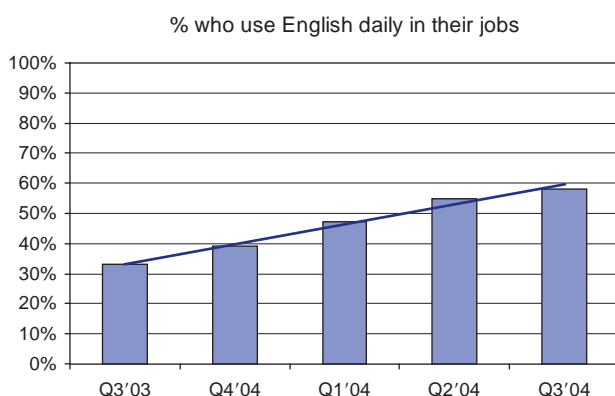


Figure 2 Percentage of respondents who use English daily in their jobs. From Nunan, D. (2005). The evolution of technology and the value of online English language learning, p. 5, *White Paper*, Global English Corporation, San Francisco, CA, USA.

In second-language learning, the dominant role for technology is as a provider of content and as an instructional tool. When the computer presents learners with listening and reading input, and information on pronunciation, vocabulary, and grammar, it is acting as a carrier of content. This content can be specially written for pedagogical purposes, or it can be authentic material that exists on the Web. Web-search activities are becoming increasingly popular in language instruction, and provide foreign-language learners, as opposed to second-language learners, with opportunities to learn through interacting with authentic texts. Such texts were originally created for the purposes of communication, not to teach English (Nunan, 1999). When the computer provides opportunities for learners to practice a language by doing spoken and written drills, completing comprehension questions, carrying out grammar exercises, and so on, it is acting as an instructional tool.

The second major role for technology in classrooms is as a learning management tool. This role is becoming increasingly significant in second-language pedagogy, particularly with the spread of proprietary software, such as Blackboard and Moodle. These learning management software packages enable teachers to carry out many learning management tasks more efficiently and effectively than more traditional face-to-face methods. For example, a good learning management system can:

- administer and collate needs analysis data from students;
- allow teachers to post course information, handouts, and other materials for student to download;
- enable students to submit assignments, and teachers to grade and return assignments, electronically;
- document student achievement and archive learner portfolios containing samples of spoken and written language;
- administer, analyze, collate, and store the results of classroom quizzes;
- administer, collate, and present student evaluations of teachers.

The third major role for technology in language learning is as a communication tool. One of the greatest frustrations for learners attempting to acquire another language in a foreign-rather than second-language context is finding opportunities to activate their language by interacting with other people in that language. One of the greatest benefits of technology is that it can bridge distance and time, enabling learners to interact with native speakers and other learners who are living, working, and learning in a wide range of cultural contexts in different parts of the world. At present, the most common technology-based means of communicating for purposes of language learning are e-mail and text chat, although voice chat and webcam are becoming increasingly common.

The Technology-Supported Second-Language Classroom

As Internet resources began to supplement and, in some cases, replace traditional face-to-face instruction, teachers and instructional designers debated the merits of the virtual classroom (Nunan, 2005). Initially, there was considerable skepticism toward the notion that someone could learn a language through the Internet. Such arguments are rarely heard these days. The issue is no longer whether to use technology, but rather how technology can support classroom instruction, and the extent to which it can replace conventional second-language classrooms. Studies in online second-language learning suggest that this approach has a number of benefits – specifically, online language learning, as it

- promotes a deep rather than surface approach to learning (Biggs and Telfer 1993);
- stimulates active, constructivist learning rather than straight knowledge transfer (Weasenforth *et al.*, 2002);
- fosters students' awareness of discourse-related aspects of communication (Davis and Thiede, 2000); and
- allows students to share perspectives and experiences, to establish relationships, and to seek assistance (Chong, 1998).

Before examining models of classroom organization, a word on terminology is in order. Classrooms are traditionally seen as places in which teachers and learners are gathered together for instructional purposes. "The L2 classroom can be defined as the gathering, for a given period of time, of two or more persons (one of whom generally assumes the role of instructor) for the purposes of language learning" (van Lier, 1988: 47). However, with the development of distance learning, and, in particular, the use of technology, the gathering together may happen in a virtual classroom rather than a physical space.

The ease with which the computer can bring people together across time and space is forcing a redefinition of the classroom. Through the computer, the gathering together for language learning no longer requires the individuals to inhabit the same physical space. The following vignettes illustrate some of the changes wrought by technology and some of the ways in which technology has redefined the second-language classroom.

A teacher educator in Auckland, New Zealand, conducts a graduate class on second-language acquisition through a text chat site with students in Tokyo, Japan; San Diego, California; Bangkok, Thailand; and Buenos Aires, Argentina.

A student in Toronto, Canada, who was unable to make his Spanish class, is able to review a transcript of the lesson which is posted on to the Web several hours after the conclusion of the class.

A secondary school teacher in Hong Kong posts all her assignments and class handouts onto the class website. Students either work with these online as needed or download those that they want in hard copy.

Using voice chat, English as a Foreign Language (EFL) students in China, Korea, and Japan take part in a conversation class with a teacher based in Bogota, Columbia.

A school in Osaka has its students complete an online placement test which automatically assesses and places students into instructional groupings in a fraction of the time it used to take using a pencil-and-paper test.

In a keynote paper, which explored the impact of technology on second-language pedagogy, Legutke commented on the ways in which changes wrought by technology have forced a redefinition of the foreign-language classroom. He notes that "These changes challenge our self-concept as foreign language teachers, because, much

more than in the past, we are now called upon to redefine our roles as educators, since we need to mediate between the world of the classroom and the world of natural language acquisition" (Legutke, 2000: 1).

Several models of classroom organization have been proposed for technology-supported classrooms. All of them involve some form of integration of traditional face-to-face classroom activity and virtual classrooms. Such combinations are referred to as blended or hybrid learning in the online learning literature. Several models have been proposed for blended learning:

- *Model 1: The traditional classroom is supplemented and supported by technology.* In this model, classroom content is delivered in the traditional face-to-face manner, but instruction is supplemented by technology. This supplementation can take the form of additional content or support for the management of learning. The teacher decides what is to be taught, how it is to be taught, and how it is to be assessed, and technology plays a secondary role.
- *Model 2: Technology delivers the content and is supported by web-based live instruction.* In this model, instructional content is delivered through technology, and is supplemented and supported by a teacher. This is the model that is followed by commercial enterprises such as English Town and Global English. For example, Global English consists of a comprehensive ten-level suite of courses which are accompanied by a wide range of additional resources such as online needs analysis instruments, grammar and vocabulary enrichment, and an online magazine. Self-paced learning is supported by a talk-with-the-teacher feature. This feature enables students to interact with a teacher and other learners in a virtual classroom through text and voice chat.
- *Model 3: Technology delivers the content and is supported by supplemental face-to-face instruction.* This model is similar to model 2, except that the live support is provided in traditional face-to-face classrooms rather than through the Internet.
- *Model 4: The fully integrated classroom.* In the fully integrated classroom, technology and live instruction work side by side. Teachers do what they do best, such as facilitating interaction in the target language, while technology does what it does best. (The final section of this article describes seven things that technology does better than teacher-led instruction.)

These models demonstrate that the early dichotomy between face-to-face instruction and virtual, technology-mediated instruction was overly simplistic. At the time that this article was written, the overwhelming consensus among second-language instructors was that some form of blended learning was the preferred option in most pedagogical situations.

Research into Technology and Language learning

A rapidly growing body of research is beginning to provide insights into the ways in which technology is having an impact on, and changing the nature of, second-language instruction. Initially, instructors rather naively wanted to know whether the virtual classroom was more effective than the face-to-face classroom. The question is naive because, as noted in the preceding section, it is not an either-or issue. Technology and teachers each have a unique contribution to make to the learning process. The question is also naive because it is almost impossible to control and assess the relative contributions made by technology and face-to-face instruction. It also represents a return to the old method of comparing studies (is method X superior to method Y?) that proved to be futile over 30 years ago (Scherer and Wertheimer, 1964).

Contemporary research into technology and language learning can be quite sophisticated in terms of the questions and issues it addresses and the range of methods that it employs. This can be seen by looking at the list of contents from recent issues of the top journal in the field, *Language Learning and Technology*. Here, for example, is the contents list from the most recent issue of that journal:

- e-learning and the development of intercultural competence
- ESL students' computer mediated communication practices
- L1 and L2 glosses: their effects on incidental vocabulary learning
- Noticing and text-based chat. (*Language Learning and Technology*, 10(3), September 2006.)

Questions currently preoccupying second-language researchers concerning the impact of technology on second-language pedagogy include:

- What similarities and differences are there between the discourse of online chat and face-to-face classroom interaction?
- What affective factors are at play in technology-supported classrooms?
- What aspects of pedagogy are best delivered through technology and what aspects are best delivered through teacher-led instruction?
- What is the reaction of language learners to technology-enhanced instruction?

The discourse of technology-supported communication has been a particular focus of research in second-language classrooms. This is not surprising, given that language is both the medium and the content of instruction in second-language classrooms. In one of the studies reported in the latest issue of *Language Learning and*

Technology ('Noticing and text-based chat'), it was found that online chat was more effective than face-to-face conversations in prompting learners to notice their own linguistic mistakes (Lai and Zhao, 2006).

In another study into the discourse features of online English as a Second Language (ESL) classes, the researchers drew the following comparisons and contrasts with face-to-face classes:

1. All the pedagogical functions identified in face-to-face classroom interaction are evident in synchronous teacher-hosted chat.
2. Choosing the appropriate responding move is crucial, both to pedagogy and the ongoing interaction.
3. The responding move enables the teacher to build a pedagogical interaction.
4. The multilayering of discourse in chat is even more complex than in face-to-face interaction.
5. Text chat may have pedagogical advantages over voice chat. For example, learners will have a visual record of the interaction. This will provide them with more time to process the message.

Benefits of Technology

There are areas where teacher-led instruction has benefits that clearly outweigh those offered by virtual instruction. These include the ability to organize, monitor, and facilitate interactive learning in small groups of learners, the ability to diagnose and deal with specific learning difficulties (although in some aspects of learning technology is catching up fast), and the ability to offer emotional support and encouragement. The importance of this last benefit should not be underestimated. The attrition rate in virtual classrooms is much higher than that in face-to-face classrooms, and it may be the positive presence of the teacher, along with the support and pressure of peers, that helps to maintain much higher attendance rates in face-to-face classrooms.

Conversely, there are areas where technology offers clear advantages over teacher-led instruction. This section comments on six of these:

- individualized study plans;
- anywhere/anytime instruction;
- patient tutoring;
- a private space to make mistakes;
- immediate, individualized feedback; and
- detailed records of achievement.

Individualized Study Plans

While there is room for some flexibility in face-to-face classrooms, in most, instruction is pitched at the middle ground. Syllabus content as well as course goals and

objectives have to be tailored to the majority, and individual needs and wants, both in terms of learning outcomes and preferred learning pathways, can rarely be satisfactorily catered to.

Technology, however, can facilitate the development of individualized study plans. In courses which are 100% virtual, each learner can have a program tailored to his or her needs, and each learner in the program can thus follow his or her own unique program. In blended classrooms consisting of a mix of teacher-led and virtual instruction, a degree of individualization can also be achieved.

At the beginning of the learning process, a series of menu-driven choices can lead the learner through questionnaires and prompts that elicit a set of terminal learning objectives. Learners can select their own preferred learning styles and strategies, and the program can recommend tasks and activities based on these choices. Learners can also indicate the level of proficiency gain they would like to achieve by a certain date. Given their current level of proficiency, the system can suggest the approximate number of hours that the student will need to study in order to achieve the goals. If the system is linked to an online bank of materials, it can also assemble a tailored set of learning resources for each individual learner.

In online courses, where the content is tagged with lexical, grammatical, and functional information, the software can also perform an important diagnostic function. As students work through the materials, the program can identify those aspects of the linguistic system where students are weak, and can suggest adjustments to the study plan to provide additional remedial practice in these areas.

Anywhere/Anytime Instruction

Unlike conventional classrooms, where the time and place of instruction are established in advance, electronic (e)-learning allows students to engage in learning at the time and place of their choosing. This is possibly the most obvious benefit of e-learning, although it is also a possible disadvantage. Having a set time to attend a teacher-led class provides a learner with a routine that is missing from the virtual classroom. The daily pressures of work and personal life can push the time that was set aside for e-learning to the back of the learner's daily agenda. This is likely one reason for the higher levels of attrition from virtual as opposed to face-to-face courses.

Patient Tutoring

As it is a skill rather than a body of knowledge to be mastered, acquisition of a second language requires time and extensive repetition. It is rarely feasible for adequate

repetition to be provided in conventional classrooms. A natural tendency in a classroom with more than one student is for the teacher to move on when a majority of students demonstrate a skill, leaving the slower learner behind. The technology-driven classroom, however, has infinite patience, and learners can spend as long as they feel they need on a particular exercise or task.

A Private Space to Make Mistakes

Technology enables learners to study in their own private space. Here, they are free to make mistakes in private. They need not be subjected to the embarrassment and even personal humiliation of making mistakes in front of classmates. The fear of making mistakes in public is partly personal and partly cultural (see Ellis, 2009). Many learners are unconcerned at making mistakes in public. Others, however, are paralyzed with fear and embarrassment at the prospect of making mistakes. This attitude to making mistakes in public appears to be a personality factor that is linked to learning style preferences (Willing, 1994).

This is particularly important for Asian learners for whom making mistakes in front of others is a major deterrent to language learning (Biggs and Telfer, 1987).

Immediate, Individualized Feedback

Providing feedback to learners on how successfully they completed a particular task or exercise is a fundamental aspect of the instructional process. In second-language pedagogy, considerable research has been conducted into the dynamics of this process. Researchers have looked at what aspects of a learner's performance should be evaluated by the teacher, when it should be evaluated, and the form that feedback to the learner should take (Chaudron, 1988).

Extensive research in both content classrooms and language classrooms indicates that the timing of the feedback is critical. The closer the feedback is to the actual performance, the more powerful is its impact on subsequent performance as well as learner motivation (Dorneyi, 2001).

In regular classroom instruction, the ability of the teacher to provide individualized feedback is severely constrained. Apart from the occasional asides to particular students during group activities, it is impossible to provide instant feedback to all the students in the class. The best that the teacher can do is to monitor student activity during the lesson, and provide feedback on the most common errors.

Online instructional programs, on the other hand, can provide instant feedback. As soon as the learner has completed a grammar exercise or comprehension task and

submitted his or her responses, the program can provide feedback on how well he or she has done. Most programs can not only tell students which answers are correct and which wrong, but also provide qualitative information on why particular responses are incorrect.

Of course, there are constraints on the types of learner performance that can be evaluated without the intervention of a human being. Questions and tasks that have a clear right or wrong answer are easier for technology to handle than ones in which more than one response is acceptable. Giving feedback on receptive skills (listening and reading) is easier than giving feedback on the productive skills of speaking or writing. At the time that this article was written, considerable time, effort, and money were being spent on developing tools for assessing students' productive language (de Jong and Bernstein, 2001). In terms of speaking, one of the more successful efforts has been the commercial program known as Phonepass (de Jong and Bernstein, 2001). Using speech-recognition technology, this program administers a test of speaking ability on the telephone, and provides assessment takers with a language-proficiency rating. Efforts are currently underway to assess learners' writing skills.

These programs are achieving impressive results. Phonepass assessments correlate highly with standardized proficiency tests such as Test of English for International Communication (TOEIC) and Test of English as a Foreign Language (TOEFL; Bernstein *et al.*, 2000). However, there is considerable resistance from consumers, who are skeptical at the notion that a computer can furnish valid and reliable assessments of second-language speaking and writing (McLure, K. Global English Corporation, Personal Communication).

Detailed Records of Achievement

Allied to the preceding point is the fact that e-learning programs can assemble and provide detailed records of achievement on individual learners. This is especially helpful in a corporate setting where an administrator may be tracking thousands of learners. Global English has a particularly sophisticated set of reports on learners. Corporate administrators can track the progress and status for every user, no matter how large the program, so that they can see the impact that the program is having on individual performance.

Summary

This article began by outlining the three main roles of technology in the second-language classroom: as a

provider of content and instruction, as a learning management tool, and as a communication device. It then described how technology is helping to fill a demand for language instruction that has far outstripped the capacity of the conventional classroom, and outlined four instructional models of the technology-enhanced classroom. The final section of article examined six key aspects of pedagogy where technology offers benefits over conventional teacher-led classroom instruction.

See also: Internet-based Education.

Bibliography

- Beatty, K. (2003). *Teaching and Researching Computer-Assisted Language Learning*. London: Pearson.
- Bernstein, J., de Jong, J., and Pisoni, D. (2000). Two experiments on automatic scoring of spoken language proficiency. In Delcloque, P. (ed.) *Proceedings of InSTIL2000 (Integrating Speech Tech. in Learning)*, August 2000, pp 57–61. Scotland: University of Abertay Dundee.
- Biggs, J. and Telfer, R. (1987). *The Process of Learning*, 3rd edn. Sydney: Prentice Hall.
- Chapelle, C. A. (2003). *English Language Learning and Technology: Lectures on Applied Linguistics in the Age of Information and Communication Technology*. Philadelphia, PA: John Benjamins.
- Chaudron, C. (1988). *Second Language Classrooms: Research on Teaching and Learning*. Cambridge: Cambridge University Press.
- Chong, S.-M. (1998). Models of asynchronous computer conferencing for collaborative learning in large college classes. In Bonk, C. J. and King, K. S. (eds.) *Electronic Collaborators*, pp 157–182. Mahwah, NJ: Erlbaum.
- Christison, M. A. and Nunan, D. (2001). Pedagogical functions in synchronous elearning interaction: The online conversation class. *Paper Presented at the TESOL 2001*, St. Louis, Missouri, USA.
- Davis, B. and Thiede, R. (2000). Writing into change: Style shifting in asynchronous electronic discourse. In Warschauer, M. and Kern, R. (eds.) *Network-Based Language Teaching: Concepts and Practice*, pp 87–120. Cambridge: Cambridge University Press.
- de Jong, J. and Bernstein, J. (2001). Relating PhonePass overall scores to the Council of Europe Framework Level Descriptors. *Technical Report, Language Testing Services*, Velp, The Netherlands.
- Dorney, Z. (2001). *Teaching and Researching Motivation*. London: Longman.
- Egbert, J. (2005). *CALL Essentials: Principles and Practice in CALL Classrooms*. Alexandria, VA: Teachers of English to Speakers of Other Language.
- Ellis, R. (2009). *The Study of Second Language Acquisition*, 2nd edn. Oxford: Oxford University Press.
- Fotos, S. and Browne, C. (eds.) (2004). *New Perspectives on CALL for Second Language Classrooms*. Mahwah, NJ: Erlbaum.
- Higgins, J. and Johns, T. (1984). *Computers in Language Learning*. London: Collins.
- Lai, C. and Zhou, Y. (2006). Noticing and text-based chat. *Language Learning and Technology* 10(3), 102–120.
- Legutke, M. (2000). Redesigning the foreign language classroom: A critical perspective on information technology (IT) and educational change. *Plenary Presentation, International Language in Education Conference*, University of Hong Kong, Hong Kong, December 2000.
- Nunan, D. (1999). *Second Language Teaching and Learning*. Boston: Thomson Learning/Heinle.
- Nunan, D. (2005). The evolution of technology and the value of online English language learning. *White Paper, Global English Corporation*, San Francisco, CA, USA.

- Richards, J. C. and Rodgers, T. (1986). *Approaches and Methods in Language Learning*, 2nd edn. Cambridge: Cambridge University Press.
- Scherer, G. and Wertheimer, M. (1964). *A Psycholinguistic Experiment in Foreign Language Teaching*. New York: McGraw-Hill.
- van Lier, L. (1988). *The Classroom and the Language Learner*. London: Longman.
- Warschauer, M. (1999). *Electronic Literacies: Language, Culture and Power in Online Education*. Mahwah, NJ: Erlbaum.
- Weasenforth, D., Biesenbach-Lucas, S., and Meloni, C. (2002). Realizing constructivist objectives through collaborative

technologies: Threaded discussions. *Language Learning and Technology* **6**(3), 58–86.

Willing, K. (1994). *Learning Styles in Adult Migrant Education*. Sydney: NCELTR.

Further Reading

- Beatty, K. (2003). *Teaching and Researching Computer-Assisted Language Learning*. London: Pearson.

TECHNOLOGY AND LEARNING – TEACHER LEARNING

Contents

Teacher Skills and Knowledge for Technology Integration
Technology in Preservice Teacher Education
Technology Resources for Teacher Learning

Teacher Skills and Knowledge for Technology Integration

N Law, University of Hong Kong, Hong Kong, People's Republic of China

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Information and communication technology (ICT) – The term often used interchangeably with information technology (IT) in the literature on technology in education.

Information technology – It is also used interchangeably in the literature on technology in education with the term ICT and refers both to computer hardware and software as well as networking infrastructure to enable the user to connect to the Internet via the computer hardware.

Teacher learning is crucial to the process and outcome of integrating information technology (IT) in the learning and teaching process in schools. Studies of information and communication technology (ICT) implementation in schools have often found teachers' lack of IT knowledge and skills to be a major obstacle (e.g., Pelgrum and Anderson, 1999). A recent comparative study of pedagogy and ICT use in schools in 22 countries and educational systems found teachers' self-perceived technical and pedagogical ICT competence to be positive, significant predictors for teachers' ICT adoption in their own teaching practice (Law and Chow, 2008). In particular, pedagogical ICT competence, which is the teachers' ability to make appropriate selection and use of ICT tools in different curriculum contexts for different pedagogical purposes, is the most crucial determinant of actual ICT use in instruction.

The core skills and knowledge that teachers need to have differ depending on the perceived purpose and anticipated impact of technology integration in the curriculum. Efforts to integrate ICT in the teaching and

learning process to achieve some general or specific educational goals, rather than as an object of study as in early computer programming classes, started around the early 1980s (e.g., Becker, 1985). Since then, there have been major shifts in the rationale put forward for the role of ICT in the curriculum, both in the academic literature and in educational policy documents around the world, moving from a focus on improving students' learning in the school curriculum to the reforming/transformational role of ICT (Means, 2000; Pelgrum and Law, 2003).

Teacher Learning to Realize the Enabling Potential of ICT in Education

One of the earliest published ICT-related standards for teachers was established in the UK (DES, 1989) for pre-service teachers, which exemplified the kind of knowledge and skills expected of teachers in order that they will be able to select and make appropriate use of IT to promote learning. The standard included competence to make use of technology as a personal productivity tool, such as using a wordprocessor for writing and a spreadsheet for exploring the consequences of different scenarios given a set of mathematical relationships, as well as the ability to critically review and select software and IT devices suitable for the subject matter and student profile concerned, plan, and execute schemes of work that incorporate the use of such resources. This has evolved later into a detailed curriculum for the use of ICT in teaching specific subjects. Standards with this focus can operate within curriculum frameworks that have been developed without specific considerations or demands about what technology may bring to the teaching and learning situation, that is, these are standards for enabling uses of technology to improve students' learning effectiveness.

Related competence includes the knowledge of when, when not, and how to use ICT in different classroom settings, and how to make use of ICT in assessment. The kind of knowledge and the associated skills required of teachers is conceptualized by Mishra and Koehler (2006) within a technological pedagogical content knowledge (TPCK) framework. This framework builds on Shulman's (1987) notion of pedagogical content knowledge as knowledge about how to teach specific concepts within a curriculum. Shulman asserts that a competent teacher possesses not only pedagogical knowledge (general teaching strategies) and content knowledge (knowledge of the subject area), but also pedagogical content knowledge. Mishra and Koehler extend this notion to the appropriate instructional use of technology, arguing that teachers not only need technical knowledge, but also technological content knowledge (i.e., knowledge about how technology and content are reciprocally related to each other, e.g., the use of simulations and modeling tools will allow learners to explore the interrelationships between different entities in a system such as species in an ecological environment without requiring an understanding of calculus (Mellar *et al.*, 1994)), technological pedagogical knowledge (i.e., knowledge of the existence, components, and capabilities of various technologies for use in teaching and learning settings and pedagogical considerations for their selection, e.g., whether e-mail, a discussion forum, or a wiki would be an appropriate means for supporting student collaboration in a specific context), and TPCK (i.e., knowledge of the dynamic, transactional relationship between these three knowledge components to support pedagogical decisions, e.g., a teacher may decide to select a particular concept mapping tool for students to share and discuss ideas about a particular topic which is known to be prone to misconceptions and difficulties for most students, and the choice of the tool is made on the basis of its functionality and ease of use by students).

Many European countries, such as the Scandinavian and East European countries, have adopted the European computer-driving license, a computer-skills-certification program adopted by many professions, as a core part of a teachers' ICT competence for technology integration in their teaching practices. The license comprises seven modules which can be studied and tested in any order: concepts of IT, using the computer and managing files, wordprocessing, spreadsheets, databases, presentations, and information and communication (Internet and e-mail). Some countries have included additional requirements related to pedagogical use of ICT. In the late 1990s, Denmark formalized such requirements into a pedagogical ICT driving license for the integration of ICTs in education. More recently, a European pedagogical ICT license (EPICT) has been established by a consortium of EPICT organizations, involving countries like Greece, Italy, Hungary, Australia, Ireland, Austria, Iceland, Uganda, and UK

for continuous professional development of teachers in the pedagogical integration of information, media, and ICT in education. A teacher needs to complete eight modules (four compulsory and four optional) to qualify for the license within a developmental framework, starting with three compulsory modules on finding information on the web, text and the writing process, and electronic communication and collaboration, followed by a selection of four optional modules among 12 (choices include digital images, spreadsheets, presentations and interactivity, publishing on the web, databases, simulations and models, ICT and learning strategies, reading and ICT, games and ICT, etc.), and finishing with a compulsory module on ICT and school development. The EPICT program emphasizes the blending of pedagogical knowledge of ICT-skills integration with ICT-skills training, and adopts a collaborative and problem-based approach to learning and team-based assessment.

Teacher Learning to Foster Students' Information-Literacy Skills

The advent of personal computers and hence the widespread availability of computers to the general public has brought about significant changes in society at large. Toffler's (1981) argument that the third wave in human history, an information revolution, is already occurring, bringing the world into a new era, the information age, has been widely accepted for at least two decades. Kuhlthau (1987) put forward the concept of information literacy as encompassing library skills and computer literacy, integrating resources and skills into subject-based inquiries across the curriculum, with the school library/media center as the key place for this integration.

With the increasing recognition of the need for schools to prepare students for life in an information society, the definition of what constitutes information literacy has also expanded. The national educational technology standards for students (ISTE, 1998) developed in the US by the International Society for Technology in Education (ISTE) is an example of a comprehensive framework that embodies the use of technology as productivity, communication, research, problem-solving, and decision-making tools as well as an understanding of the social, ethical, and human issues related to technology. On the basis of this set of standards, ISTE (2000) published a corresponding set of teacher standards and performance indicators in order that they would be able to foster the development of information literacy in their students. These standards were designed to be incorporated into subject-matter teaching and hence incorporate some of the technical and pedagogical competences described in the previous section. However, these standards emphasize the ability to use technology to locate, evaluate, and collect information, process data and report results, and to facilitate

higher-order and complex-thinking skills such as problem solving, critical thinking, decision making, and knowledge construction. They also require teachers to be able to use technology in the development of strategies to solve real-world problems and in connecting with peers and experts for collaboration and interaction.

This focus on information literacy skills as a key rationale for integration of ICT in learning and teaching is also evident in education-policy documents in many parts of the world, even if not all of them have developed specific ICT-related teacher education standards. The UNESCO (2002a,b) framework for ICT in teacher education also takes the development of students' ICT literacy skills as the goal for IT integration in the curriculum. The framework comprises four groups of ICT-related competencies: pedagogy, collaboration and networking, social issues, and technical issues. Pedagogy-related competencies are essentially the knowledge and skills necessary for appropriate and effective use of ICT to support teaching and learning in the subject matter concerned. Teachers also need to have adequate understanding of the social issues such as intellectual property and health related to ICT use in the society and be able to guide students on such matters. The other two competencies concern expectations of teachers as members of a professional community. They are expected to be able to make effective use of the communication capability of technology to organize and provide learning opportunities for students outside of the classroom walls as well as new professional-development opportunities for themselves. The technical-issues component relates to whether a teacher is adequately prepared to undertake self-directed learning as a lifelong learner to meet the challenges posed by continuing technological advances.

An important feature of the UNESCO (2002a,b) framework is that it conceptualizes ICT development in schools into four stages of maturity and provides an adaptive guideline for a different implementation approach, depending on the school's specific stage of development. The framework incorporates four stages: emerging, applying, infusing, and transforming uses of technology. At a very preliminary stage, an emerging approach could be adopted whereby a school provides ICT courses to students, taught by individual teachers who have the technical competence. At a later stage, the applying approach can be adopted when the curriculum focus is on providing opportunities for students to apply their ICT skills in some specified learning contexts. The training focus at this stage is still on those teaching the ICT curriculum to ensure that students will have the requisite ICT skills to undertake the designed learning tasks. The school advances to the stage when an infusing approach can be adopted when all teachers share the vision of bringing about new learning opportunities to students through ICT integration. The professional development focus at this stage is on technical

and pedagogical ICT skills in the relevant subject areas as well as on collaborative, cross-curricular uses of ICT. A school reaches the most advanced stage of development according to this framework when the school is ready for and committed to making use of ICT to realize visions for the school of tomorrow; hence the use of ICT in the curriculum is characterized as a transforming approach for this stage. Here, the curriculum emphasis is on providing differentiated and individualized learning opportunities for students, and on learners taking responsibility for their own learning and contributing to solving real-world problems. When the school is ready for a transforming approach to ICT integration, the teacher has to be a lifelong learner to be able to innovate and to contribute to leading the school into a learning community. The framework envisions professional development at this stage as a self-managed, continuous, and critically reflective process.

Teacher Learning to Prepare Students for the Twenty-First Century

The information revolution coupled with an increasing rate of globalization has led to deep changes in many aspects of the social and economic realities in disparate corners of the world. There are many publications describing how the economic order of the world is changing from one heavily reliant on natural resources and industrial production to one based on knowledge (e.g., Drucker, 1993). While different terms such as postindustrial economy, digital economy, service economy, and postcapitalist economy have been used to describe the changing landscape in economic development, knowledge economy is the term that is gaining popularity as it highlights the importance of knowledge as the engine of economic growth and development. Over the last several decades, there is an emergence of businesses that engage in knowledge-intensive activities, exploiting IT to create wealth from knowledge (Chichilnisky, 1998; Schwartz *et al.*, 1999). In this new order, the economic well-being of a country is largely dependent on the level of competence that its people have in accessing, using, managing, and creating knowledge. Hence, the importance of the quality of educational outcomes to national well-being is at the highest in human history, as is evidenced by the main educational reforms and IT in education master plans that have been launched in many different countries around the world (Pelgrum and Law, 2003). In particular, there have been projects that focus on identifying the kind of knowledge and skills needed for functioning effectively in the twenty-first century (e.g., CERI, 2006; European Commission, 2007), with an important focus on preparation for employment. For example, the Partnership for 21st Century Skills (2007), an advocacy organization formed from more than 30 organizations comprising

multinational businesses and professional/educational bodies, put forward a framework for twenty-first century skills for US K-12 student schools that include three sets of skills beyond the core subjects and cross-curricular themes to include: life and career skills; learning and innovation skills; and information, media, and technology skills.

It is within this broader policy context that UNESCO undertook the ICT Competency Standards for Teachers (ICT-CST) project to take further the 2002 curriculum and program guidelines for ICT in teacher education described in the previous section (UNESCO, 2008a). The project is based on the premise that in any nation or community, education plays a central function in promoting personal and societal well-being for sustainable development. The new set of competency standards is constructed around three approaches to ICT integration for educational change on the basis of three factors leading to human-capacity-based growth that have been identified by economists: capital deepening, higher-quality labor, and technological innovation. The technology-literacy approach focuses on increasing the technological uptake of students and citizens by incorporating technology skills in the curriculum, a goal suitable for a capital-deepening economy. The knowledge-deepening approach focuses on building up a high-performance workforce by developing the ability of students and citizens to solve complex, real-world problems by applying the knowledge they have learned. The knowledge-creation approach aims to produce a workforce that can innovate, produce, and benefit from new knowledge for a knowledge economy. The UNESCO (2008b) policy framework provides a set of standards for six aspects of ICT use and teacher professional development for each of these three approaches: policy and vision; curriculum and assessment, pedagogy, ICT, organization and administration; and teacher professional development. The ICT-CST framework and standards was designed to help countries to align their teacher-education program with national-development goals by allowing them to adopt an approach appropriate to their country's current stage of development within the context of the broader developmental trajectory.

Irrespective of the approach to ICT integration adopted, the ICT-CST framework expects that curriculum adaptations and changes may need to be made in order to fulfill the requisite educational goals. For the technology-literacy approach, the curriculum will need to include basic technological literacy skills and the integration of relevant ICT skills into different parts of the curriculum as appropriate. The requisite teacher competencies include technical and pedagogical competence for ICT integration in subject teaching and assessment as well as using ICT for managing classroom data and for supporting their own professional development. The knowledge-deepening approach requires that the curriculum

has emphases on depth of understanding and application of understanding to real-world problems. Teacher competence for undertaking this approach includes the ability to design and implement learning activities that promote deep understanding and application of key concepts to solve complex, real-world problems that make use of open-ended as well as subject-specific tools using student-centered methods and collaborative projects. Teachers should be able to use ICT to monitor students' learning progress as individuals and in groups, use communication and networking technology to support student collaboration, seek help and advice from experts, set up professional collaboration and for their own professional development.

The knowledge-creation approach requires that the curriculum go beyond a focus on knowledge of school subjects to include knowledge creation as an important twenty-first century skill. While achieving this goal entails that students undertake various forms of learning to meet their identified learning needs, knowledge creation (sometimes called knowledge building) is not the same as learning. Learning scientists working in the area of knowledge building generally believe that it has to take place in a social, collaborative process of inquiry. Scardamalia and Bereiter (2003), for example, argue that knowledge building is not the same as guided discovery or other forms of project-based learning where the inquiry is just a form of learning activity for learners to acquire a predefined set of content goals. In contrast, knowledge building has a goal of creating or modifying public knowledge and requires a group of people working intentionally to extend the frontier of knowledge as the community perceives it. The availability and use of a digital environment that supports the interactions among members of the community as they engage in evolving cycles of idea improvement and deepening levels of understanding is an important part of the knowledge-building process.

The most significant goal of the knowledge creation approach is to develop students into lifelong learners, who are capable of determining their own learning goals and plans, identifying learning needs and monitoring progress as they work together in a community of learners. Under this approach, the school becomes a learning organization in which everyone, students and teachers, is involved in learning and knowledge building. To be able to guide students in knowledge creation, the teacher has to model the knowledge-building process by working alongside the students as a co-learner as well as by engaging in a professional community to build knowledge about this new model of schooling. The professional competence required of the teacher in such contexts cannot be developed through conventional learning contexts but has to be nurtured in a setting in which teachers collaborate as a community to undertake curriculum and pedagogical innovations in their own practices (Scardamalia and Bereiter, 2005; Looi *et al.*, 2008).

The shifting emphasis toward transforming schools into learning organizations to prepare students for the twenty-first century as the context of developing a teacher-competence framework and standards for ICT integration is happening in many parts of the world. ISTE (2007) has revised its national education technology standards for students to have a much stronger focus on developing students' ability to make appropriate and effective use of technology in tasks that require twenty-first-century skills, such as creativity and innovation, communication and collaboration, research and information fluency, critical thinking, problem solving and decision making, as well as understanding of issues related to digital citizenship. Competence in technology operations and concepts is a necessary substrate but just a small part of the expected repertoire of competence. A related set of technology standards and performance indicators for teachers was released by ISTE in mid-2008 (ISTE, 2008).

Changes in teacher-competence standards need to be coupled with implementation mechanisms for their realization, which appears to be lagging behind in practice. Within the US, many states have used the ISTE technology standards as the basis for their technology standards for teachers. Few states, however, actually assess teachers' skills in technology integration in any manner other than requiring completion of a course on technology and instruction. Anderson and Plomp (2008) also report, based on results from a survey about national contexts of education and ICT-related education policies from 22 education systems, that most of the surveyed systems do not have any ICT-specific requirements for teacher certification, and none report any requirement related to subject teaching with ICT.

Teacher Learning Beyond Knowledge – ICT Integration as a Lever for Innovation and Transformation

It is clear from the preceding discussion that the knowledge and skills required of the teacher differ depending on the goals and approaches adopted for ICT integration. Even the lowest level of integration – adopting the use of ICT as a tool in the accomplishment of learning tasks in the curriculum – would require some adjustments in curriculum-task designs. In more advanced models of ICT integration, technology is perceived as a lever for curriculum and pedagogical innovation and for institutional transformation. The role of a teacher changes completely as he/she moves from using technology to enhance students' learning effectiveness to setting up and orchestrating communities of learners that make pervasive use of technology. The various research and development projects reviewed earlier identify long lists of teacher competencies that require the mastery of

identifiable knowledge and skills. However, teacher professional development for ICT integration designed to bring about transformative changes in education need to go beyond the development of knowledge and skills for it to be successful.

Webb and Cox (2004), in their review of research on pedagogies associated with the use of ICT in primary and secondary schools, found that the benefits from ICT integration in teaching and learning depend on how the teacher selects and organizes ICT resources and how these are integrated into the learning activities. The teacher's pedagogical approach, and hence the pedagogical reasoning adopted, are crucial determinants of the learning outcomes from ICT integration. There is evidence that teachers need to undertake more complex pedagogical reasoning in their planning and teaching in order to incorporate knowledge of technology affordances with subject-based teaching objectives. Webb and Cox also show that teachers' pedagogical reasoning is strongly influenced by their beliefs about the value of ICT for learning and the nature of successful learning environments.

Kozma (2003) reports on a comparative study of innovative pedagogical practices using technology conducted in 26 countries and education systems. The study found evidence of changing forms of activities as well as changing roles of the teacher and the learner in the 174 case studies collected. Teacher professional development was identified as an essential condition for an innovation to be sustainable beyond the span of a year, and in many of those cases, teachers were provided not only with traditional forms of training but also with collaborative, problem-oriented types of activities such as peer-study groups that allowed teachers to learn and develop new technical and pedagogical skills in the context of the innovation. Building on further analysis of the cases collected in this study, Law (2008) argues that ICT plays a disruptive role in pedagogical innovations as its use is not to sustain but to change existing processes and relationships. A teacher who integrates such uses of technology is in effect an agent of change, one who leverages ICT as a disruptive force. To support educational transformation, teachers need to develop the metacognitive ability as an autonomous learner to identify and tackle problems that arise in the innovation process, the sociometacognitive capacity to work in collaboration with peers (and experts) in a professional community to institute changes at the classroom level and beyond, and the socioemotional capacity to engage in change, take risks, and foster trust. All these functions require courage and motivation. Professional development designed for educational change and innovation needs to foster the development of an epistemological belief aligned with socially grounded, constructivist theories of learning (Oakes, 2002; Law, 2008). Failures to attend to these aspects of teacher learning will undermine efforts

to institute innovation and transformation through the use of technology.

See also: Conceptions of Technology Literacy and Fluency; Relating Technology, Education Reform and Economic Development; Technology in Preservice Teacher Education.

Bibliography

- Anderson, R. E. and Plomp, T. (2008). National contexts. In Law, N., Pelgrum, W. J., and Plomp, T. (eds.) *Pedagogy and ICT Use in Schools around the World: Findings from the SITES 2006 Study*, pp 37–66. Hong Kong: CERC, University of Hong Kong and Springer.
- Becker, H. J. (1985). How schools use microcomputers: Results from a national survey. In Chen, M. and Paisley, W. (eds.) *Children and Microcomputers: Research on the Newest Medium*, pp 87–107. Beverly Hills, CA: Sage.
- CERI (2006). *CERI – Schooling for Tomorrow: Alternative Models of Learning and Innovation*. http://www.oecd.org/document/26/0,3343,en_2649_34487_36271194_1_1_1_1,00.html (accessed August 2009).
- Chichilnisky, G. (1998). The knowledge revolution. *Journal of International Trade and Economic Development* **7**(1), 39–54.
- DES (Department of Education and Science in England) (1989). *Criteria for Accreditation Course of Teacher Training, Circular 24/89*. London: DES.
- Drucker, P. F. (1993). *Post-Capitalist Society*. New York: HarperBusiness.
- European Commission (2007). *Commission Proposes Actions to Foster 21st Century e-Skills*. <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/07/1286&format=HTML&aged=0&language=EN&guiLanguage=en> (accessed August 2009).
- ISTE (1998). *National Educational Technology Standards for Students*. Eugene, OR: International Society for Technology in Education.
- ISTE (2000). *National Educational Technology Standards for Teachers*. Eugene, OR: International Society for Technology in Education.
- ISTE (2007). *National Educational Technology Standards for Students*, 2nd edn. Eugene, OR: International Society for Technology in Education.
- ISTE (2008). *National Educational Technology Standards and Performance Indicators for Teachers*. Eugene, OR: International Society for Technology in Education.
- Kozma, R. (ed.) (2003). *Technology, Innovation, and Educational Change: A Global Perspective*. Eugene, OR: ISTE.
- Kuhlthau, C. C. (1987). Information skills: Tools for learning. *School Library Media Quarterly* Fall, 22–28.
- Law, N. (2008). Teacher learning beyond knowledge for pedagogical innovations with ICT. In Voogt, J. M. and Knezek, G. A. (eds.) *International Handbook of Information Technology in Primary and Secondary Education*, pp 425–430. New York: Springer.
- Law, N. and Chow, A. (2008). Teacher characteristics, contextual factors and how these affect the pedagogical use of ICT. In Law, N., Pelgrum, W. J., and Plomp, T. (eds.) *Pedagogy and ICT Use in Schools around the World: Findings from the SITES 2006 Study*, pp 182–219. Hong Kong: CERC, University of Hong Kong/Springer.
- Looi, C. K., Lim, W. Y., and Chen, W. (2008). Communities of practice for continuing professional development in the 21st century. In Voogt, J. M. and Knezek, G. A. (eds.) *International Handbook of Information Technology in Primary and Secondary Education*, pp 489–505. New York: Springer.
- Means, B. (2000). Technology in America's schools: Before and after Y2K. In Brandt, R. S. (ed.) *Education in a New Era. ASCD Yearbook 2000*, pp 185–210. Alexandria, VA: Association for Supervision and Curriculum and Development.
- Mellar, H., Bliss, J., Boohan, R., Ogborn, J., and Tompsett, C. (eds.) (1994). *Learning with Artificial Worlds: Computer Based Modelling in the Curriculum*. London: Falmer.
- Mishra, P. and Koehler, M. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record* **108**(6), 1017–1054.
- Oakes, J. (2002). *Becoming Good American Schools: The Struggle for Civic Virtue in Education Reform*. San Francisco, CA: Jossey-Bass.
- Partnership for 21st Century Skills (2007). *Framework for 21st Century Learning*. http://www.21stcenturyskills.org/documents/frameworkflyer_072307.pdf (accessed August 2009).
- Pelgrum, W. and Anderson, R. (eds.) (1999). *ICT and the Emerging Paradigm for Life Long Learning*. Amsterdam: IEA.
- Pelgrum, W. and Law, N. (2003). *ICT in Education around the World: Trends, Problems and Prospects*. Paris: UNESCO: International Institute for Educational Planning.
- Scardamalia, M. and Bereiter, C. (2003). Knowledge building. In Guthrie, J. W. (ed.) *Encyclopedia of Education*, 2nd edn., pp 1370–1373. New York: Macmillan.
- Scardamalia, M. and Bereiter, C. (2005). Does education for the knowledge age need a new science? *European Journal of School Psychology* **3**(1), 21–40.
- Schwartz, P., Kelly, E., and Boyer, N. (1999). The emerging global knowledge economy. In OECD (ed.) *The Future of the Global Economy: Towards a Long Boom*, pp 77–113. Paris: OECD.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review* **57**(1), 1–22.
- Toffler, A. (1981). *The Third Wave*, 537p. New York: Bantam Books.
- UNESCO (2002a). *Information and Communication Technologies in Teacher Education: A Planning Guide*. Paris: UNESCO.
- UNESCO (2002b). *Information and Communication Technology in Education: A Curriculum for Schools and Programme of Teacher Development*. Paris: UNESCO.
- UNESCO (2008a). *ICT Competency Standards for Teachers: Policy Framework*. Paris: UNESCO.
- UNESCO (2008b). *ICT Competency Standards for Teachers: Competency Standards Modules*. Paris: UNESCO.
- Webb, M. and Cox, M. (2004). A review of pedagogy related to information and communications technology. *Technology, Pedagogy and Education* **13**(3), 235–286.

Further Reading

- Law, N., Pelgrum, W. J., and Plomp, T. (eds.) (2008). *Pedagogy and ICT in Schools Around the World: Findings from the SITES 2006 Study*, p. 296. Hong Kong: CERC and Springer.

Relevant Websites

- <http://europa.eu> – EUROPA, Actions Proposed by the European Commission to Foster 21st Century e-Skills.
- <http://www.ecdl.com> – European Computer Driving License.
- <http://www.epict.org> – European Pedagogical ICT Licence.
- <http://www.iste.org> – International Society for Technology in Education (ISTE), National Educational Technology Standards.
- <http://www.21stcenturyskills.org> – Partnership for 21st Century Skills, Framework for 21st Century Learning.
- <http://portal.unesco.org> – UNESCO, ICT Competency Standards for Teachers.

Technology in Preservice Teacher Education

N Davis, University of Canterbury, Christchurch, New Zealand

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In the context of our twenty-first-century information society, it is not surprising to find that outcomes relating to information and communication technologies (ICTs) are required for programs preparing tomorrow's teachers. Indeed, the United Nations Educational, Scientific and Cultural Organization (UNESCO) recognizes that teacher training has an important role and recommendations from the world summit on the information society acknowledged that training teachers to use ICTs in education is a critical success factor (UNESCO, 2007).

It is important to appreciate that ICTs, often known simply as technology in the US, is incorporated into preservice teacher education for three main purposes: (1) preparing teachers to use ICTs in educationally effective ways; (2) preparing K-12 teachers to teach ICT-related content; and (3) applying ICTs to serve teacher education. In essence technology is a cross-curricular theme, a content area, and a tool that can be applied to learning and teaching, which includes the use of telecommunications to improve access to education.

Preparing Teachers to Use Technology

This section describes approaches to technology in preservice teacher education, which has developed since its introduction in 1980s in the US and the UK. In Europe, the terms used have been information technology, or ICTs, in initial teacher training. Technology in preservice teacher education expanded from an emphasis on competence with computers to the application of technology to enhance learning, teaching, and professional development. Where the majority of students entering preservice teacher education have grown up with technology, educators have the advantage of assuming basic skills and knowledge and have developed benchmarks or standards for ICTs in teacher education. The goal is similar in other countries but more challenging due to the lack of resources, including basic skills and knowledge about information technology systems. Comprehensive guidance is provided in the planning guide on ICTs and teacher training produced by UNESCO (2002; see online appendices) and UNESCO is in the process of developing ICTs competency standards for teachers (2007).

Standards that have been widely adopted include the International Society of Technology in Education's detailed US national educational technology standards for students,

teachers, and educational administrators (ISTE NETS). The original NETS were organized in six categories (Strudler and Wetzler, 1999). ISTE (2008) revised them into five standards for NETS-T that are also used by preservice programs:

1. Facilitate and inspire student learning and creativity;
2. Design and develop digital-age learning experiences and assessments;
3. Model digital-age working and learning;
4. Promote and model digital citizenship and responsibility; and
5. Engage in professional growth and leadership.

A similar comprehensive set of competencies that had been developed in 1998 in England and Wales was relaxed in 2002 and replaced with extensive non-statutory guidance (Selinger and Austin, 2003). Perhaps most useful here are the seven benchmarks of good practice identified by Kirschner and Davis (2003: 131) in the quick scan study of best practice across Europe, North America, and Australia. All of the 26 programs selected as displaying good practice had the first four benchmarks, and the experts agreed that those programs modeled twenty-first-century constructivist thinking and pedagogy:

- become a competent personal user of ICTs (100%);
- competently make use of ICTs as a tool for teaching (95%);
- competently make use of ICTs as a mind tool (100%);
- master a range of educational paradigms which make use of ICTs (100%);
- master a range of assessment paradigms which make use of ICTs (81%);
- understand the policy dimension of the use of ICTs for teaching/learning (71%); and
- other (often social) aspects of ICTs use in education (91%).

Effective teaching of technology in preservice teacher education recognizes that technology, pedagogic, and content knowledge are all involved, including combinations of all three separately and together. In essence, this concept of technology content pedagogic knowledge (TPCK) builds on the work of Schulman, who clarified the interdependence of content and pedagogic knowledge, and added the domain of technology knowledge. The American Association of Colleges for Teacher Education (AACTE, 2007) *Handbook of TCPK* provides a primer for many content areas including English, Mathematics, and Science.

The major strategies used to incorporate ICTs in pre-service teacher-education programs are: a stand-alone technology course, resource-based learning including mini-workshops, infusion of technology into methods and foundation courses, application during field experience including mentoring, plus combinations of these strategies. Best practice uses a combination of all of these and applies technology to enhance teacher education while also modeling practice for K-12 schools (Kay, 2006; Kirschner and Davis, 2003). As described in section entitled ‘The preparation of teachers to teach ICT’, technology is also offered as a specialist and/or subspecialist area in some programs and is a strong component of preservice programs for business studies teachers.

The first and most widely adopted strategy is the stand-alone technology course. In a national survey of 88 preservice programs in the US (selected because they were part of the national Homes group “committed to making programs of teacher preparation more rigorous and connected to liberal arts education, research on learning and teaching, and wise practice in schools”), 39 of the 53 respondents reported offering a specific introductory course for the preservice program and 36 of them used a three-credit lecture and lab format (mandatory for all students in 27 of the programs). Although a stand-alone course may teach a wide range of basic computer skills, Hargrave and Hsu noted significantly more emphasis was placed on integrating instructional technologies into the curriculum than ICTs use for teacher productivity or personal use. Kay (2006) also noted a range of more innovative approaches that he categorized into content-based, project-based, or process-based use of ICTs. The advantages of a stand-alone course are the provision of a foundation by specialist instructors and improvement of students’ self-efficacy, but the learning of technology skills in isolation is likely to limit later adoption, including application of ICTs in the K-12 classroom. Moursund and Bielefeldt’s (1999) large-scale survey of US programs revealed that a single required technology course did not adequately prepare graduates to use technology in their teaching and Strudler and Wetzel (1999) noted that exemplary programs retained the stand-alone course in addition to the integration of technology within the teacher-education program.

A complementary or alternative strategy is to integrate the use of technology across the program of teacher education. The main advantage is the opportunity for preservice teachers to learn with computers, not about them, in ways that permit meaningful authentic problem solving, and this approach has also been successful in improving confidence and technology skills. Challenges comprise adequate access to resources, including hardware, and limited time and expertise of faculty and supporting staff. The AACTE *Handbook of TPCK* describes technology in content-specific method courses where

teacher educators choose applications, to enhance preservice teachers’ learning, that are powerful because they are transferable to K-12 classrooms. Occasionally, the integration of technology is promoted with other important aspects to stimulate educational renewal. McShay’s (2005) model of double infusion for both technology and multicultural education within preservice programs is notable in this context.

The third strategy, field based, strongly supports standards and benchmarks discussed earlier because it actively supports ICT-based lessons by preservice teachers in K-12 schools, but it remains challenging to adequately prepare and equip students and their cooperating teachers for teaching with ICTs. It has been difficult to place students in K-12 technology-rich classrooms and therefore, innovative programs have worked on the simultaneous renewal of preservice teacher education and K-12 schools in which they place students. One successful strategy has been to provide additional preparation for students so that they become technology mentors for their cooperating K-12 teachers (or faculty) in a way that complements the teacher’s mentoring of the preservice student in K-12 classroom practices. When faculty model transferable practice for student teachers with appropriate technology, such as an interactive white board and mathematical modeling software that are also accessible in the K-12 classroom, it has been possible to provide evidence of the impact of technology in preservice education on the learning of K-12 students. Unfortunately, such tight coupling of resources and practice in both the university and K-12 school is extremely difficult to achieve. Perhaps more examples will be forthcoming in a project that is preparing millennial teachers with practice in schools designed for twenty-first-century learning promoted by the National Commission on Teaching and America’s Future.

The Preparation of Teachers to Teach ICTs

Preservice preparation for teaching of ICTs and the coordination of technology in K-12 schools is an option in some programs in most countries. While the preparation of teachers to teach technology as their major area is not common in the US, ICTs are a discipline in the UK national curriculum, but note that the term technology can be confusing in the UK due to its use for the discipline of design and technology. Many national curricula in Europe include ICTs and it is increasing as an examined subject in Africa and Asia because countries that aim to develop economically have recognized the need to train such teachers. For example, Rwanda has included technology as an optional subject within its national curriculum and the first batch of teachers of computing for Rwandan secondary schools graduated in 2007.

In addition, a minor specialization in technology is available in some preservice programs to prepare elementary and secondary teachers for a future role as a leader of cross-curricular ICTs in K-12 education. Examples of programs with technology minors include Iowa State University in the US (Davis, 2003) and the University of Oulu in Finland (Neimi, 2003). It is useful to note that many programs expect students to make up for gaps in their knowledge and skills with technology through independent learning, which is often problem based. This is successful when accompanied by development of learning resource centers rich in technology and a supportive community, and such centers provide relevant internship opportunities for students in a technology minor.

Technology as a Tool or Medium in Education

This section focuses on application of ICTs to address common challenges in preservice teacher education. Three examples are briefly described: digital images, electronic portfolios, and distance education including telementoring.

Digital images, both still and moving, have a range of applications in preservice teacher education. Video case studies of K-12 practice are deployed successfully in most programs using a constructivist or process-oriented learning approaches to examine teaching in K-12 classrooms. For example, Reading Classroom Explorer is a web-based software environment housing exemplary video clips of literacy in elementary classrooms accompanied by a complementary notepad, database search engine, additional resources, and online discussions (Ferdig *et al.*, 2002). Images of classrooms can also be captured in a virtual reality format so as to permit students to explore a range of classroom layouts. Preservice teachers can also develop their practice by capturing digital images or video of their own teaching. Digital stories with still or moving images, sound, and narration have been used successfully to support reflective practice in many disciplines including mathematics and social studies. The use of presentation software with digital images is a practice that transfers well between preservice teacher education and K-12 classrooms, and thus supports simultaneous renewal in education.

The reflective documentation of professional practice and knowledge in an electronic portfolio is the second example of the application of technology to serve teacher education. As described earlier, most programs of preservice teacher education are assessed in part by a comprehensive set of standards or competencies, including those for ICTs itself. In practice, this requires documentation of assessed evidence for each student and for the program as a whole. A portfolio of evidence from teaching practice and university course work accompanied by reflection on how the artifacts cover the standards is often adopted

during preservice programs and it can provide a foundation for the teacher's career. An electronic format may be used to facilitate the process that moves from a learning portfolio, to an assessment portfolio, and then an employment portfolio to support career progression. Some more-comprehensive software applications also permit multiple portfolios to be used for advising and program evaluation. However, electronic portfolios are a complex innovation that requires a cluster of innovations involving technology, faculty, staff, course, program, and organizational development. At times, the multiple evolving purposes of an electronic portfolio can cause stress on students and faculty where these purposes are in conflict (Strudler and Wetzler, 2005).

ICT are also used to increase access within preservice teacher education, as well as distance education. Telephone, e-mail, and/or videoconference are used by many programs to increase input from the field by providing access to distant guest speakers and virtual visits to K-12 classrooms. Team teaching of complementary classes of preservice and in-service teachers permits mutual benefits such as preservice teachers' development of resources for schools accompanied by feedback from in-service teachers in training. Probably, the most effective collaborations include contrasting cultures, such as inner-city schools linked with a rural preservice program or an international collaboration. Communication technologies, in combination with mentoring by teacher educators, are also valuable in improving the quality of reflection by preservice teachers in K-12 schools. E-mail has been used extensively and online discussion groups with peers have also been found to provide effective support when carefully integrated within the contexts. Such integration includes the development of a community of practice, in which preservice student teachers are legitimate members (Kirschner *et al.*, 2003; Laferriere *et al.*, 2004).

Distance education for preservice teacher education has used technology as the main mode for delivery of a few programs in several countries. The first distance education preservice program for primary and secondary teachers was provided by the UK Open University. Concerns relating to quality of such programs have been addressed by quality-assurance bodies, including the National Council for Accreditation of Teacher Education (NCATE) accreditation of Western Governors University's program in the US. This form of preservice program requires strong collaboration between organizations including K-12 schools with a team to design the program that includes instructors, advisors, designers, and producers who use technology to prepare and publish materials as well as to deliver courses. The scale of this mode of preparation is increasing. For example, in China, the Beijing Foreign Language University has developed a program to prepare thousands of teachers of English each year. Finally, the rapid increase of K-12 students' distance education, particularly in the US, brings with it

an additional mode of education for which preservice teachers must be prepared (Davis and Ferdig, 2009).

Issues in Technology and Preservice Teacher Education

This final section reviews four issues in technology and teacher education: faculty and organizational development and related national initiatives, equitable access to technology and teacher education, and the challenge of researching this field. Faculty and organizational development have been and will remain a major challenge worldwide because technology and education continue to evolve. ICTs are best viewed as clusters of innovation that are adopted and/or rejected by individuals and organizations in stages. For example, as described earlier, best practice in the preparation of preservice teachers to use an interactive whiteboard in a high school classroom includes adoption of new hardware, software, and procedures in the preservice program by faculty with the support of staff and administrators, along with similar adoption in the school where preservice teachers practice. Goodlad (1994) described the link between the development of colleges and K-12 schools as simultaneous renewal to emphasize that one could not come before the other; both must develop together. The adoption of each cluster of innovation takes place in stages and the speed of adoption is related to the concerns of each organization and individual with related development of TCPK and resources (Davis, 2008). In the eclectic range of strategies used to promote effective practice with technology in preservice teacher education, technology mentoring is one of the most successful strategies because it takes individual faculty concerns as a starting point and provides one-on-one support from a student who also gains knowledge of educational practice (Thompson *et al.*, 2007).

The adoption of technology in preservice teacher education has been promoted through national and international projects starting with the project, Integrating new technologies in initial teacher training (INTENT) in the UK in 1992 (Somekh and Davis, 1997), which was followed by preparing tomorrow's teachers to use technology (PT3) in the US. In some other countries, the development of preservice programs has occurred alongside in-service initiatives. For example, the Enlaces initiative in Chile and other South American countries included the majority of preservice institutions (see Hepp *et al.*, 2004) and the telelearning project spanning Canada, which created telelearning communities of practice, included development of preservice teacher education (Laferriere *et al.*, 2004). Finally, the multinational company Intel has worked with educational authorities in many countries and regions to adapt their Intel Teach to the future ICTs teacher-training program and materials to serve national and regional contexts.

Issues of equity related to culture and gender that are common in education and society remain prevalent in preservice teacher education. For example, women and minority students are more likely to enter programs with few technology skills due to lack of access (Bowser-Brown, 2004). Therefore, care is required to provide equitable participation including targeting these populations early. For example, the US PT3 initiative included capacity-building grants for this reason.

The final issue that concludes this article is the challenge of researching ICTs in preservice teacher education. A comprehensive review of research into technology in teacher education by Willis *et al.* (1999) called for multiparadigmatic approaches and it has been answered with research into the current status, descriptions of innovation, instructional design and component studies, and studies that have identified bias of technology. Kay's (2006) review of research concluded that more rigorous and comprehensive research is required into key technology strategies. Additional research in regions beyond the US and Europe is recommended plus a more ecological perspective, which recognizes that ICTs in education requires continuing development of the ecologies of the school, classroom, and of teacher education (Davis, 2008). In addition, there are calls to produce evidence of the impact of technology in preservice teacher education on K-12 student achievement. A longitudinal study of exemplary programs in the US started in 2005 but the findings are likely to be problematic because ICTs change depending on where learning takes place and with whom. Despite that, it is likely that ICTs in preservice teacher education will continue to expand because of the permeation of ICTs in society in the twenty-first century and related links with economic prosperity.

See also: Assessment in Schools – Technology Education and ICT; Continuing Professional Development of Teachers; Curriculum, Digital Resources and Delivery; Equity in Technology Access and Opportunities; Teacher Education and Models of Teacher Reflection.

Bibliography

- AACTE Committee on Technology and Innovation (2007). *Handbook of Technological Pedagogical Content Knowledge. A New Framework for Teacher Knowledge*. Washington, DC: Erlbaum.
- Bowser-Brown, A. (2004). A critical look at the disparities of the digital divide for minority preservice teachers. In Crawford, C., Willis, D. A., Carlsen, R., *et al.* (eds.) *Proceedings of Society for Information Technology and Teacher Education International Conference 2004*, pp 4035–4038. Chesapeake, VA: AACE.
- Davis, N. E. (2003). Technology and teacher education in the USA: What makes for good sustainable practice? *Technology, Pedagogy and Learning* 12(1), 59–84.
- Davis, N. E. (2008). Leadership of teacher learning for educational renewal with IT. In Voogt, J. and Knezek, G. (eds.) *International Handbook of Information Technology in Education*, pp 507–520. Amsterdam, NL: Springer.

- Davis, N. E. and Ferdig, R. (in press) Editorial: What is special about teacher education for Virtual Schooling? *Journal of Technology and Teacher Education*.
- Ferdig, R. E., Roehler, L. R., and Pearson, P. D. (2006). Video and database-driven web environments for pre-service literacy teaching and learning. In McKenna, M. C., Labbo, L. D., Kieffer, R. D., and Reinking, D. (eds.) *International Handbook of Literacy and Technology*, Vol. 2, pp 235–256. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Ferdig, R. E., Roehler, L., and Pearson, P. D. (2002). Scaffolding preservice teacher learning through web-based discussion forums: An examination of online conversations in the reading classroom explorer. *Journal of Computing in Teacher Education* 18(3), 87–94.
- Goodlad, J. (1994). *Educational Renewal: Better Teachers, Better Schools*. San Francisco, CA: Jossey-Bass.
- Hargrave, C. and Hsu, Y. (2000). Survey of instructional technology courses for preservice teachers. *Journal of Technology and Teacher Education* 8(4), 303–314.
- Hepp, P., Hinestrosa, J. E., and Laval, E. (2004). A systemic approach to educational renewal with new technologies: Empowering learning communities in Chile. In Brown, A. and Davis, N. E. (eds.) *The World Yearbook of Education 2004. Digital Technology, Communities and Education*, pp 299–311. London: Routledge/Falmer.
- ISTE (2008). The ISTE National Educational Technology Standards (NETS-T) and Performance Indicators for Teachers. http://www.iste.org/Content/NavigationMenu/NETS/ForTeachers/2008Standards/NETS_T_Standards_Final.pdf (accessed July 2009).
- Kay, R. H. (2006). Evaluating strategies used to incorporate technology into preservice education: A review of the literature. *Journal of Research on Technology in Education* 38(4), 383–408.
- Kirschner, P., Boshuizen, I., and Wooperis, I. (2003). Pedagogy of training in information and communications technology for teachers and beyond ... *Technology, Pedagogy and Learning* 12(1), 149–159.
- Kirschner, P. and Davis, N. E. (2003). Pedagogic benchmarks for information and communications technology in teacher education. *Technology, Pedagogy and Education* 12(1), 125–147.
- Laferriere, T., Breuleux, A., and Erickson, G. (2004). Telecollaborative communities of practice in education within and beyond Canada. In Brown, A. and Davis, N. E. (eds.) *The World Yearbook of Education 2004. Digital Technology, Communities and Education*, pp 264–276. London: Routledge/Falmer.
- McShay, J. (2005). Double infusion: Toward a process of articulation between critical multicultural education and technology education in a teacher preparation program. *Contemporary Issues in Technology and Teacher Education* 4(4), 429–445. <http://www.citejournal.org/vol4/iss4/general/article2.cfm>.
- Moursund, D. and Bielefeldt, T. (1999). *Information Technology in Teacher Education*. Eugene, OR: International Society for Technology in Education.
- Neimi, H. (2003). Towards a learning society in Finland: Information and communications technology in teacher education. *Technology, Pedagogy and Education* 12(1), 85–103.
- Selinger, M. and Austin, R. (2003). A comparison of the influence of government policy on information and communications technology for teacher training in England and Northern Ireland. *Technology, Pedagogy and Education* 12(1), 19–38.
- Somekh, B. and Davis, N. E. (1997). *Using IT Effectively in Teaching and Learning: Studies in Pre-Service and In-Service Teacher Education*. London: Routledge.
- Strudler, N. and Wetzel, K. (1999). Lessons from exemplary colleges of education: Factors affecting technology integration in preservice programs. *Educational Technology Research and Development* 47(4), 63–81.
- Strudler, N. and Wetzler, K. (2005). The diffusion of electronic portfolios in teacher education: Issues of initiation and implementation. *Journal of Research on Technology in Education* 37(4), 411–433.
- Thompson, A. D., Chuang, H. H., and Sahin, I. (eds.) (2007). *Faculty Mentoring: The Power of Students in Developing Technology Expertise*. Charlotte, NC: Information Age.
- UNESCO (2002). *ICT in Teacher Education. A Planning Guide*. Paris: UNESCO. <http://unesdoc.unesco.org/images/0012/001295/129533e.pdf> (accessed Jun. 2009).
- UNESCO (2007). *UNESCO ICT Competency Standards for Teachers*. Paris: UNESCO.
- Willis, J. W., Thompson, A., and Sadara, W. (1999). Research on technology and teacher education: Current status and future directions. *Educational Technology Research and Development* 47(4), 29–45.

Further Reading

- Davis, N., Cho, M. O., and Hagenson, L. (2005). *Special Issue: Intercultural Issues in ICT and Teacher Education. Contemporary Issues in Technology and Teacher Education* 4(4), 384–394.
- Kirschner, P. (ed.) (2003). *Special Issue: Information Technology, Pedagogy and Practice in Education. Journal of Technology, Pedagogy and Learning* 12(1), 1–192.
- McDougall, A. (2008). Teacher education. In Voogt, J. and Knezek, G. (eds.) *International Handbook of Information Technology in Education*, pp 461–474. Amsterdam, NL: Springer.
- Somekh, B. and Davis, N. E. (eds.) (1997). *Using IT Effectively in Teaching and Learning: Studies in Pre-Service and In-Service Teacher Education*. London: Routledge.
- UNESCO Planning Guide for ICT in teacher education: UNESCO (2002). *ICT in Teacher Education. A Planning Guide*. UNESCO, Paris, France. <http://unesdoc.unesco.org/images/0012/001295/129533e.pdf> (accessed June 2009).
- Willis, J. W., Thompson, A., and Sadara, W. (1999). Research on technology and teacher education: Current status and future directions. *Educational Technology Research and Development* 47(4), 29–45.

Relevant Websites

- <http://www.iste.org> – International Society for Technology in Education (ISTE): National Educational Technology Standards (NETS).
- <http://www.intel.com/education> – Intel Teach to the Future.
- <http://www.aace.org/site> – Society for Information Technology in Teacher Education.
- <http://www.public.iastate.edu> – Teacher Education Goes into Virtual Schooling website.
- <http://www.itte.org.uk> – UK Association of Information Technology in Teacher Education.

Technology Resources for Teacher Learning

J E Hinostroza, C Labbé and L López, Universidad de La Frontera, Temuco, Chile

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Learning communities – The conception of teachers' learning and developing within a broader context of community, institution, polity, and profession.

Situated cognition – The theory that suggests that knowledge is situated – being, in part, a product of the activity, context, and culture in which it is developed and used.

Introduction

From a general perspective, many authors claim that teachers are key for the successful implementation of innovations in schools (Barber and Mourshed, 2007), particularly when it deals with the introduction of digital technology in teaching and learning (OECD, 2001; Pelgrum and Law, 2003; Wood, 2002). Calls for teachers to innovate and change their practice entail implicit or explicit demands on teachers. Teachers are expected to prepare their pupils for the Knowledge Society, which requires the ability to create knowledge, apply knowledge to unfamiliar problems, and communicate it effectively to others. To prepare their students for such a society, teachers are called upon to emphasize higher-order thinking skills and metacognition (thinking about thinking), by using constructivist approaches to learning and understanding, cooperative learning strategies, a wide range of assessment techniques, and computer-based and other information technologies that enable pupils to access information independently (Hargreaves, 2003).

Extensive teacher learning is required to respond to such demands. Most teachers were taught the old traditional paradigms of teaching and learning (teacher centered, knowledge reproduction, etc.) in which technology resources, if used at all, helped to reach traditional goals (e.g., to deliver traditional curriculum but in a more effective way) (Lankshear and Knobel, 2003). At present, new paradigms give new opportunities to use technology resources for teaching and learning never seen before.

This article presents an overview of the potential roles of technology resources for teacher learning, particularly focusing on in-service teachers. Against this backdrop, this article addresses the following questions:

1. What are teachers' learning needs?
2. What factors and processes are related to teachers' learning and innovation?
3. What type of information and communication technologies can help these processes?

To address the first question, a review of different theories describing teachers' knowledge and expertise is presented. This section seeks to characterize the types of learning domains that need to be considered. In relation to the second question, a review of the literature describes innovation processes and how teachers can not only learn something new, but also transfer what they have learned to their practice. Finally, to address the last question, approaches for relating particular technology resources to innovation processes are presented.

Teachers' Learning Needs

The body of knowledge that teachers traditionally share has been analyzed from at least four perspectives:

Kinds of Knowledge (or Know-What)

There are many studies that attempt to classify and describe teachers' knowledge as a set of domains of expertise and their intersections (see, e.g., Grossman, 1995; Leach and Moon, 2000; Mishra and Koehler, 2006; Shulman, 1986). Broadly speaking, they propose the following domains:

- (1) Knowledge of subject matter, content knowledge, or subject knowledge.
- (2) Knowledge of general pedagogy or pedagogical knowledge and broad understanding of classroom management and organization.
- (3) Pedagogical content knowledge, that is, ways of representing specific subject knowledge appropriately for learners.
- (4) Knowledge of curriculum, that is, grasp of the materials, resources and tools of the trade.
- (5) Knowledge of learners and their characteristics.
- (6) Knowledge of educational contexts or school knowledge, ranging from groupings, classrooms, schools, education authorities, and national policies to wider communities and cultures.
- (7) Knowledge of educational aims, purposes, and values.

Nature of Knowledge (or What-Knowledge)

Some authors classify teachers' knowledge as formal (university-based knowledge about education) and informal knowledge (developed largely through experience) (Page, 2001). This perspective is similar to Hammond's cognitive continuum theory (Hammond, 1980), which defines analytic and intuitive thinking as poles of a continuum, arguing that most thinking is neither purely intuitive nor purely analytical. This concept is similar to the one described by Olson (1992) as "folkways of teaching," which emphasize the tacit dimension of practice and the process of change. In a similar vein, Schön's theories about reflection-on-action (Schön, 1983), while criticizing the technical, rational models for describing professional expertise, propose to search for "an epistemology of practice implicit in the artistic, intuitive process which some practitioners do bring to situations of uncertainty, instability, uniqueness and value conflict" (Schön, 1983: 49). Further, some classroom-based research indicates that both planning and classroom interaction are responsive, compositional, and situated (i.e., contextualized) (Yinger and Hendriks-Lee, 1995).

Level of Knowledge (or Know-How)

This perspective is based on a different way of describing teachers' knowledge that looks at the process of expertise development or skill acquisition based on the development of "knowing how" rather than "knowing what" (Dreyfus and Dreyfus, 1986). In these models, the pathway to competence is characterized mainly by the ability to recognize features of practical situations and to differentiate among them, to carry out routine procedures under pressure, and to plan ahead. For example, Berliner (1995) describes five levels of expertise that were defined by Dreyfus and Dreyfus (1986): (1) novice, (2) advanced beginner, (3) competent, (4) proficient, and (5) expert. Each level is characterized according to the way that teachers plan their teaching, their behavior in the classroom, and their actions while facing critical situations. In general terms, these authors argue that the way to progress through these levels of expertise can be summarized as improving the management of the classroom and developing strategies to plan the

lessons. In doing this, they implicitly characterize teachers' knowledge as corresponding to classroom management and lesson planning.

Knowledge in Practice (or How-To)

Finally, the fourth perspective – knowledge in practice – is based on analysis of routines and actions in the classroom. For example, Evertson (1995) describes the research about classroom rules and routines that teachers follow while teaching. She defines rules as general norms for expected behaviors and routines as procedures for accomplishing particular classroom tasks (how-to). In particular, she describes the results of a study by Leinhardt *et al.* (1986) that defines routines as systems of exchange that are set up to accomplish tasks, describing three types of routines:

- (1) Management routines, that include housekeeping, discipline maintenance, and people-moving tasks.
- (2) Support routines, that is, specific behaviors and actions necessary for a learning–teaching exchange to take place, for example, "how to pass in papers."
- (3) Exchange routines, that is, the interactive behaviors that permit the teaching–learning exchanges to occur. They govern the language contacts between teachers and students – for example, routines for choral responses.

These routines imply that teachers need to be knowledgeable about these dimensions of their actions.

This short review demonstrates that there is a significant amount of research in the area of teacher knowledge. Summarizing, it is possible to classify the different approaches as trying to characterize teachers' knowledge from the perspective of its kind, nature, level, and practice. **Table 1** presents these different approaches.

In general terms, it would be difficult to contend that only one of these taxonomies of teachers' knowledge is complete enough to characterize what teachers know or need to learn. Hence, the combination of them appears to be more useful to characterize it, also revealing its complexity. In fact, as Mishra and Koehler (2006) argue, apart from looking at each of these types of knowledge in isolation, there is a need to research and understand the type of knowledge resulting from their intersections.

Table 1 Taxonomy of teachers' knowledge taxonomies

<i>Kind (know-what)</i>	<i>Nature (what knowledge)</i>	<i>Level (know-how)</i>	<i>Practice (how-to)</i>
Content	Formal vs. Informal	Novice	Management routines
Pedagogical content	Intuitive vs. analytical	Advanced beginner	Support routines
Curriculum		Competent	Exchange routines
Pedagogy		Proficient	
Learners		Expert	
Educational contexts			
Educational aims, purposes, and values			

Factors and Processes Related to Teachers' Learning and Innovation

Management of change and innovation in schools have been areas of debate for a while, and although in this article we will not engage in the broad discussion, we will draw on this body of knowledge to discuss the factors and processes related to teachers' learning. In particular, we can transfer conclusions drawn by authors with regard to conditions for innovation in education (Scardamalia and Bereiter, 2006) to teachers' learning. There are two key conditions simplifying the arguments: one related to mastering the innovation in the Classroom and the other regarding to sharing it with Colleagues.

Mastering the Innovation in the Classroom and Sharing it with Colleagues

Huberman (1992) describes innovation as a process of grafting the new on to the old, and he comments that every "old" is a distinctive, local context with its own history and configuration (Fullan, 1992; Huberman, 1992). Olson (1988) also defines the process of change not as substituting one practice for another, but rather as subjecting existing practices to challenges posed by another well-conceived practice. Both definitions have an evolutionary approach to change, rather than a revolutionary one. They suppose that the new should be somehow grounded in the old. These arguments can be easily transferred to teachers', learning domain, implying that their learning will be facilitated if the learning process considers their previous knowledge, experiences, and current context.

The Importance of Understanding Teachers and Teachers' Self-Understanding

Fullan and Stiegelbauer (1991) argue that the subjective world – the phenomenology – of the role of incumbents needs to be understood as a necessary precondition for engaging in any change effort with them. These arguments suggest a situated perspective of teachers' professional development similar to the concept of situated cognition described by Brown *et al.* (1989), and aligned with the concept of learning communities, described by Shulman and Shulman (2004), that present a frame for conceptualizing teacher learning and development within communities and contexts. This implies that the professional development is, in part, shaped by, and becomes a product of, the activity, context, and culture in which it is developed and applied.

From a different perspective, the study called Second Information Technology in Education Study: Module 2 (SITES M2) (IEA, 2000) describes a framework for innovation. The study, conducted during 1999–2003, consisted of a series of qualitative studies which identified and described innovative pedagogical practices that used

technology (Kozma, 2003). The study report describes three levels of factors affecting the innovation process:

- (1) Macro-level or system factors, such as cultural norms, social context, educational policy, curriculum standards, etc.
- (2) Meso-level or school factors, such as the Information and Communication Technologies (ICT) infrastructure available, ICT integration plans, school leadership, innovation history, parents, etc.
- (3) Micro-level or individual factors, which, for teachers, include factors such as pedagogical practice, innovation history, educational background, experience with technology, etc.; and, for pupils, include experience with technology, social and cultural background, etc.

In addition, the conceptual framework describes how these factors relate to the characteristics of innovation, namely its complexity, practicality, clarity, relevance, and need. These characteristics of an innovation can be considered to be the expected characteristics of a course for teachers.

A different – yet complementary – perspective on the innovation process focuses on the individual and his/her practice. Arguments supporting this view are given by Fullan (1993), who says that "systems do not change themselves, people change them." Fullan and Stiegelbauer (1991) assert that "educational change depends on what teachers do and think – it's as simple and as complex as that" (p. 117). From this perspective, some authors discuss the factors and conditions necessary for a teacher to innovate. For example, Shulman and Shulman (2004) – based on the idea of developing a community of learners – claim that, in order to innovate, teachers need six key characteristics:

- (1) Vision: readiness to pursue a vision of classrooms or schools that constitutes, for example, communities of learning.
- (2) Motivation: willingness to expend the energy and persistence to sustain such teaching.
- (3) Understanding: understanding the concepts and principles needed for such teaching.
- (4) Practice: ability to engage in the complex forms of pedagogical and organizational practice needed to transform their visions, motives, and understandings into a functioning, pragmatic reality.
- (5) Reflection: capability to learn from their own and others' experiences through active reflection in and on their actions and their consequences.
- (6) Community: capability and experience in working as members of a functioning learning community and/or of forming such communities in the settings where they work (Shulman and Shulman, 2004: 259).

Another perspective for innovation is based on activity theory, which considers activities as collective, systemic

formation that has a complex mediational structure based on instruments, subjects, rules, community, division of labor, and objects (Engeström, 1994). In this framework, innovation occurs as the interplay among these elements. Finally, other theories of teacher innovation and change are provided by Zhao and Frank (2003) who propose an ecological perspective on the introduction of ICT in schools. In this perspective, they characterize schools as ecosystems in which teachers are immersed and defines the introduction of new ways of using ICT as an invasion of an exotic species into the ecosystem. Using this metaphor, they analyse the innovation in schools using similar rules as the ones used to describe the response of ecosystems to invasions of exotic species.

From a general perspective, many authors have related the concepts of ICT and innovation, basically considering ICT as an instrument that helps to bring about change and innovation in schools and/or teachers' practices. In this vein, authors have defined different roles of ICT in helping the innovation process:

- Trojan Horse: the computer is used as an aid to innovate in the teaching strategy (Olson, 2000).
- Catalyst: technology helps to accelerate the process of innovation (Hawkrige *et al.*, 1990; McDonald and Ingvarson, 1997).
- Lever: technology serves as a tool that is applied purposefully to a task of value (Venezky, 2002).

In summary, considering the variety of theories with regard to teachers' innovation, it is possible to argue that this is still an arena of debate and that there is still much to be researched before being able to more precisely define the characteristics of a teaching innovation that ensures its successful adoption and implementation.

Roles of Technology for Teachers' Learning and Innovation

Although there is "very little fundamental research that investigates how teachers might learn with digital technologies" (Fisher *et al.*, 2006: 2), there is a wide range of ways in which digital technologies (computers, Internet, personal digital assistants (PDAs), mobile phones, etc.) can be used. In particular, Fisher *et al.* (2006) propose to "describe the affordances of digital technologies as 'clusters' of purposeful activity in teacher learning" (p. 21), and to classify the activities and possible roles of technology. Table 2 presents their framework.

Based on this classification and drawing on some of the categories described in the previous sections, Table 3 relates teachers' knowledge to innovation characteristics and the role of technology.

Categories presented in this table provide examples of the roles ICT can play in supporting teachers' learning

Table 2 Clusters of activities with digital technology

Clusters	Examples of activities
Knowledge building	<ul style="list-style-type: none"> • adapting and developing ideas • modeling • representing understanding in multimodal and dynamic ways
Distributed cognition	<ul style="list-style-type: none"> • accessing resources • finding things out • writing, composing, and presenting with mediating artifacts and tools
Community and communication	<ul style="list-style-type: none"> • exchanging and sharing communication • extending the context of activity • extending the participating community at local and global levels
Engagement	<ul style="list-style-type: none"> • exploring and playing • acknowledging risk and uncertainty • working with different dimensions of interactivity • responding to immediacy

From Fisher, T., Higgins, C., and Loveless, A. M. (2006). *Teachers Learning with Digital Technologies: A Review of Research and Projects* (No. 14). Bristol: Futurelab, p 20.

process, considering the particular type of knowledge to be taught and relating these processes to specific innovation characteristics. With regard to the examples of ICT tools, it could be argued that many of them could be categorized as helping to develop several of the types of knowledge or innovation processes listed in the table. Indeed, since these technologies were not necessarily designed using a framework such as the one presented here, the classification responds to its potential to contribute to develop each type of knowledge or to support each innovation process.

In order to further investigate the potential adaptation of these technologies, a methodology such as technology mapping (Angeli and Valanides, 2009) could be used. They present it as a methodology for guiding teacher thinking about the ill-defined problem of designing technology-enhanced learning that can help to recognize and define the roles of ICT in developing different types of Technological Pedagogical Content Knowledge (TPCK). Based on the concept of TPCK (Mishra and Koehler, 2006), the method basically defines a structured process to relate software affordances, content representations, their pedagogical uses, and the different types of knowledge.

Conclusion

In the previous sections, we presented an overview of different theories that describe teachers' knowledge, the main trends with regard to educational innovation and

Table 3 Relation between teachers' knowledge, innovation characteristics, and ICT

<i>Types of knowledge</i>	<i>Innovation characteristics</i>	<i>Role of ICT</i>	<i>Example of ICT tools</i>
<i>Kind: know-what.</i> Examples: content, pedagogical content, curriculum pedagogy, learners, educational contexts, educational aims, purposes and values.	This type of knowledge can be associated with the need of understanding the concepts and principles needed for teaching.	Supporting knowledge building, in particular, adapting and developing ideas and modelling and representing understanding in multimodal and dynamic ways.	Encyclopedias, productivity tools (word processor, presentation, etc.), digital content resources, concept mapping software, learning objects, etc.
<i>Nature: what-knowledge</i> Examples: Formal vs. Informal, Intuitive vs. analytical.	This type of knowledge can be associated with the capacity of learning from their own and others' experiences through active reflection in and on their actions and their consequences.	Supporting distributed cognition, that is, accessing resources, finding things out, writing, and composing and presenting with mediating artifacts and tools.	Wikis, blogs, social bookmarking services (e.g., del.icio.us), podcast, simulation, and modeling software, etc.
<i>Level: know-how</i> Examples: Novice, Advanced beginner, Competent, Proficient, Expert.	This type of knowledge can be associated with the one developed through working as members of functioning learning communities and/or of forming such communities in the settings where they work.	Supporting building community and communication, which is, exchanging and sharing communication, extending the context of activity and extending the participating community at local and global levels.	Virtual learning communities (e.g., Tapped In, Teachscape), social networks (e.g., Facebook), immediate response systems (e.g., Twitter), podcast, video, and image sites (e.g., YouTube, Flickr), learning management systems, etc.
<i>Practice: how- to</i> Examples: Management routines, Support routines, Exchange routines.	This type of knowledge can be associated with teachers' practice, being able to engage in the complex forms of pedagogical and organizational practice needed to transform their visions, motives, and understandings into a functioning, pragmatic reality.	Supporting engagement that means, exploring and playing, acknowledging risk and uncertainty, working with different dimensions of interactivity and responding to immediacy.	Personal learning environment (PLE), Virtual Learning Environments (VLE), virtual worlds (e.g., Second life).

proposals concerning the role of ICT in teachers' learning and innovation.

Based on this evidence, we can conclude that – although there are many theories that describe teachers' knowledge – there is still a need to research and understand the type of knowledge resulting from their intersections before being able to more precisely identify teachers' learning needs and, once identified, to define the particular role that ICT will play in supporting the learning process.

With regard to the factors that facilitate teachers' learning and innovation, different frameworks defining characteristics of successful innovation processes have been demonstrated in this article; however, there is very little evidence to show that a particular set of characteristics can ensure that an innovation process will be successfully adopted and implemented (i.e., its predictive potential). In this context, although technology, acting as a lever, could facilitate different types (or stages) of processes, the consequences of these processes cannot be predicted with certainty.

Therefore, we conclude that although technology is available to support different types of learning processes that can be aimed at developing different types of knowledge, if learning needs are not yet clear and the characteristics of the learning and innovation process are not definitive, the particular role of ICT will remain uncertain, since ICT is only a tool that should be purposefully used to support and enhance teachers' learning processes.

See also: Technology in Preservice Teacher Education.

Bibliography

- Angeli, C. and Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT-TPCK: Advances in technological pedagogical content knowledge (TPCK). *Computers and Education* 52(1), 154–168.
- Barber, M. and Mourshed, M. (2007). *How the World's Best-Performing School Systems Come out on Top*. London: McKinsey and Company.

- Berliner, D. C. (1995). Teacher expertise. In Anderson, L. W. (ed.) *International Encyclopedia of Teaching and Teacher Education*, pp 46–52. Pergamon: Oxford.
- Brown, J. S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* **18**(1), 32–42.
- Dreyfus, H. L. and Dreyfus, S. E. (1986). *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. Oxford: Blackwell.
- Engeström, Y. (1994). Teachers as collaborative thinkers: Activity-theoretical study of an innovative teacher team. In Carlgren, I., Handal, G., and Vaage, S. (eds.) *Teachers' Minds and Actions: Research on Teachers' Thinking and Practice*, pp 43–621. London: Falmer.
- Evertson, C. M. (1995). Classroom rules and routines. In Anderson, L. W. (ed.) *International Encyclopedia of Teaching and Teacher Education*, pp 215–218. Oxford: Pergamon.
- Fisher, T., Higgins, C., and Loveless, A. M. (2006). *Teachers Learning with Digital Technologies: A Review of Research and Projects*. (No. 14). Bristol: Futurelab.
- Fullan, M. (1992). *Successful School Improvement: The Implementation Perspective and Beyond*. Buckingham: Open University Press.
- Fullan, M. (1993). *Change Forces: Probing the Depth of Educational Reform*. London: Falmer.
- Fullan, M. and Stiegelbauer, S. (1991). *The New Meaning of Educational Change*, 2nd edn. London: Cassell Educational Limited.
- Grossman, P. L. (1995). Teachers' knowledge. In Anderson, L. W. (ed.) *International Encyclopedia of Teaching and Teacher Education*, pp 20–24. Oxford: Pergamon.
- Hammond, K. R. (1980). *Human Judgement and Decision Making*. New York: Hemisphere.
- Hargreaves, A. (2003). *Teaching in the Knowledge Society: Education in the Age of Insecurity*. Maidenhead: Open University Press.
- Hawkridge, D., Joworosky, J., and McMahan, H. (1990). *Computers into Third-World Schools: Examples, Experiences and Issues*. London: McMillan.
- Huberman, M. (1992). Critical introduction. In Fullan, M. (ed.) *Successful School Improvement: The Implementation Perspective and Beyond*, 1st edn, pp i–vi. Buckingham: Open University Press.
- IEA (2000). SITES module 2: Case studies of innovative pedagogical practices using technology second information technology in education study. www.sitesm2.org (accessed August 2009).
- Kozma, R. B. (ed.) (2003). *Technology, Innovation and Educational Change: A Global Perspective*. Eugene: International Society for Technology in Education/International Association for the Evaluation of Educational Achievement.
- Lankshear, C. and Knobel, M. (2003). *New Literacies: Changing Knowledge and Classroom Learning*. Buckingham: Open University Press.
- Leach, J. and Moon, B. (2000). Pedagogy, information and communications technology and teachers' professional knowledge. *Curriculum Journal* **11**(3), 385–404.
- Leinhardt, G., Weidman, C., and Hammond, K. M. (1986). Introduction and integration of classroom routines by expert teachers. *Curriculum Inquiry* **17**(2), 135–176.
- McDonald, H. and Ingvarson, L. (1997). Technology: A catalyst for educational change. *Journal of Curriculum Studies* **29**(5), 513–527.
- Mishra, P. and Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record* **108**(6), 1017–1054.
- OECD (2001). *Learning to Change: ICT in Schools*. Paris: OECD.
- Olson, J. (1988). *Schoolworlds/Microworlds: Computers and the Culture of the Classroom*. Oxford: Pergamon.
- Olson, J. (1992). *Understanding Teaching*. Philadelphia, PA: Milton Keynes/Open University Press.
- Olson, J. (2000). Trojan horse or teacher's pet? Computer and the culture of the school. *Journal of Curriculum Studies* **32**(1), 1–8.
- Page, R. N. (2001). Common sense: A form of teacher knowledge. *Journal of Curriculum Studies* **33**(5), 523–533.
- Pelgrum, W. J. and Law, N. (2003). *ICT in Education Around the World: Trends, Problems and Prospects*. Paris: UNESCO: International Institute for Educational Planning.
- Scardamalia, M. and Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In Sawyer, K. (ed.) *Cambridge Handbook of the Learning Sciences*, pp 97–118. New York: Cambridge University Press.
- Schön, D. (1983). *The Reflective Practitioner*. New York: Basic Books.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher* **15**(2), 4–14.
- Shulman, L. S. and Shulman, J. H. (2004). How and what teachers learn: A shifting perspective. *Journal of Curriculum Studies* **36**(2), 257–271.
- Venezky, R. L. (2002). Quo Vademus? The transformation of schooling in a networked world. <http://www.oecd.org/dataoecd/48/20/2073054.pdf> (accessed August 2009).
- Wood, D. (2002). *The THINK Report: Technology in Education, Future for Policy*. Brussels: European Schoolnet.
- Yinger, R. J. and Hendriks-Lee, M. S. (1995). Teacher planning. In Anderson, L. W. (ed.) *International Encyclopedia of Teaching and Teacher Education*, pp 188–192. Oxford: Pergamon.
- Zhao, Y. and Frank, K. A. (2003). Factors affecting technology uses in schools: An ecological perspective. *American Educational Research Journal* **40**(4), 807–840.

This page intentionally left blank



VOCATIONAL EDUCATION AND TRAINING

Contents

An Overview of Vocational Education and Training

Informal and Non-Formal Learning in Vocational Education and Training

School-Based Vocational Education and Training

The Economics of Vocational Education and Training

An Overview of Vocational Education and Training

T Karmel, National Centre for Vocational Education Research (NCVER), Adelaide, SA, Australia

© 2010 Elsevier Ltd. All rights reserved.

Vocational education and training is aimed at imparting skills for the labor market, at a subprofessional level. It is the most diverse education sector, delivered in both the classroom and the workplace, and organized in a variety of ways internationally. A cornerstone is the apprenticeship – combining formal education with on-the-job experience – but it is also delivered in schools and tertiary educational institutions. The institutional arrangements of vocational education are complex (with industry playing an important role), its approach to teaching and learning is distinctive (and contested), and expectations by the community demanding, with it being seen as the education sector best positioned to deal with social disadvantage.

The vocational education and training (VET) sector is the least understood and most poorly defined education sector. Whereas early childhood education, primary and secondary schooling, and university education convey immediately to the reader a particular set of content and institutions, this is not the case for vocational education and training. An obvious symptom of this lack of clarity is the multitude of terms used to describe the concept. UNESCO tends to use technical and vocational education and training (TVET), in the United Kingdom people talk about further education, in Australia it has been called technical and further education (TAFE), and vocational and technical education (VTE) although the common term has been vocational education and training.

A second manifestation of the lack of clarity is that its institutional setting is varied. In some countries it is

associated with the latter years of secondary schooling, while in others it tends to be post-school. It can be taught in classrooms but can also occur in the workplace or a combination of the two. It can be at a relatively elementary level or at an advanced level. A further confusion is that much of what is taught at the university level is vocational in nature – what can be more vocational than medicine, for example.

Before going any further, we must make some attempt to define VET. While it is tempting to describe it as being what VET providers do, this is not very helpful. Our definition is that VET refers to education and training that is vocational in nature – that is, with the intent of being useful in the workforce – and is at a subprofessional level. While this is a workable definition it does not do full justice to what VET systems actually do. In particular, VET systems can provide considerable amounts of general education as a preparation for higher levels of education (such as an entry pathway to university education); they address issues of equity in many countries (VET offering second-chance education); and they can be involved in promoting innovation in industry. However, the core of VET is subprofessional vocational education.

In this article, we focus on a range of dimensions of VET that are important to its understanding. Each of these is quite distinctive compared to either school education or higher education. While the survey for this article is relatively self-contained it draws very heavily

on other articles in this encyclopedia that were commissioned under the vocational education and training banner. For convenience, these articles are referenced the first time they are relevant.

For ease of organization we discuss VET under three headings: teaching and learning, the VET system, and VET in context. We end with some comments about current issues and trends.

Teaching and Learning

Teaching and learning is of course a very broad topic. We can think about what is taught or learned, how it is taught or learned, where teaching and learning takes place, and so on.

In terms of what is taught or learned, the competency model has been the centerpiece of VET reform in many countries, for example Scotland, England and Wales, Canada, the United States, Mexico, many South American countries, Australia, New Zealand, Sweden, France, Kuwait, Indonesia, Korea, and Thailand. While the approach has many variants the essence is that the content of courses is based on occupational competency standards. Assessment is in terms of outcomes achieved (competent or not competent) by students or skills demonstrated and, in most cases, not graded. Another feature of this approach is that the competencies are developed by a wide range of interested parties and not left to educationalists. There is input from industry, unions, and bureaucrats as well.

This approach has attracted quite strident criticism from educators who argue that:

- industry needs carry too much weight at the expense of the needs of the students;
- the approach attempts to control teachers' work through national curriculum and assessment; and
- education – even vocational education – should be less instrumental and also include values related to justice, aesthetics, civics, life skills and so on.

These criticisms can be answered by those who advocate the competency approach; it all depends on how VET is constructed. For example, incorporating key competencies can be included in VET to support Dewey's concept of education through rather than specifically for occupations.

Putting aside the tension between those who favor a liberal education and those who are more instrumental, the emphasis on competencies can also be criticized by those who believe that VET is about equipping individuals for the labor market. They argue that narrow technical skills do not equip students for future change in the labor market and therefore generic skills need to be given considerable prominence. In fact, in many countries generic skills have been emphasized upon. One such

example is the Australian Employability Skills Framework as adopted in 2002 which lists the pertinent skills as communication, teamwork, problem-solving, planning and organizing, self-management, and learning skills. It also lists desirable attributes including loyalty, commitment, honesty and integrity, and even a sense of humor. The issue then is how can such skills and attributes be included in VET courses. Should they be integrated within courses containing technical competencies or taught separately? Or should these skills and developments be developed in the workplace rather than in an educational context? Another concern is how can these skills be assessed and recorded.

Back to the concept of competency. Part of the debate around the concept revolves around the lack of clarity of the concept. A commonsense approach is to define competence as the successful performance of a set of tasks; therefore, competency standards are the description of those tasks. But this confuses the tasks and the capacity to perform them. The task view implies that we can have a relatively objective checklist of tasks and omit higher-level competencies such as planning or reacting to contingencies. Another approach views competence in terms of desirable attributes. The trouble with this is that many attributes (e.g., problem-solving) are highly context-dependent. A third approach, labeled as integrated, views competence in terms of knowledge, abilities, skills, and attitudes in the context of a set of realistic occupational tasks. This approach centers on key tasks but emphasizes the attributes (cognitive skills, interpersonal skills, fine motor skills, etc.) which are required for the competent performance of the tasks identified.

A movement that has run parallel to the competency approach is modularization. Modules can be thought of as a group of competencies and can be built into whole qualifications. The argument in favor of modular structures is that they can be more rapidly updated because they are smaller than complete courses. They also provide flexibility for individuals in terms of building their skills. However, the push for modularization has tended to come from the political actors than from educationalists, who have criticized the approach as resulting in fragmented learning, mechanical and instrumental teaching, overassessment and excessive administrative burden for teachers. In trying to assess the validity of the claims for or criticisms of modularization, understanding depends on one key point: the relationship between the qualification as a whole and modules as elements of this whole. Thus there is tension between the idea of modules as being freestanding and being valuable only in the context of the wider structure of the qualification. Naturally, approaches differ. In some countries (e.g., England) the curriculum design process usually starts from the whole, whereas starting from the module is more common in other countries (e.g., in the American system).

While the concept is not straightforward it does appear that modularization remains high on the educational agenda internationally. In some cases, the emphasis has moved to internationally recognized modules, reflecting a trend to an international integration of various aspects of education.

It also needs to be remembered that VET is different from other forms of education and its delivery is not restricted to the classroom. Much of the VET occurs in a workplace setting or on the job. The driving force is the competitive needs of a business. Two shifts in the nature of work have made training in the workplace more important: the movement towards knowledge work and the shortened lifespan of job content. Managers must attend to the skills of employees as well as the management of physical and financial resources.

Learning in the workplace can take a number of forms. It can be planned or unplanned, and formal, nonformal, or informal. Planned and unplanned are self-explanatory, but the distinctions between formal, nonformal, and informal are less obvious. A distinction is drawn between

- formal learning: a program of learning recognized by a qualification or certificate,
- nonformal learning: a program of learning but not usually evaluated, and
- informal learning: learning resulting from daily activities.

There are also a number of modes of instruction that are widely used such as class-room presentation (occurring at a workplace), computer-based training, and distance training. Instruction can also be undertaken by trainers or by experienced employees training novices.

Most structured workplace training occurs in larger organizations. A long-standing policy issue centers on the needs of small businesses which tend to invest relatively little in training. Small businesses provide mostly job-specific training that is directly related to operational needs. Training tends to be reactive originating from immediate needs or, as is often the case, crises. Provision tends to reflect managerial styles and most owners/managers of small firms favor an informal management style. While most owners/managers claim a positive attitude to training, provision is low. In considering the factors affecting provision the most important are directly relevant factors such as market positioning, economic conditions, and availability of relevant training. There is also a whole range of secondary factors such as the cost of training, time constraints, lack of cover for employees who are training, dearth of in-house trainers, and absence of interest in training by employees.

Workplace training by definition has an employer or business perspective. Considerable work has also looked at teaching and learning in VET from a student perspective. There has been much theorizing about how individuals learn, and there are at least 71 models of learning

styles. Much of the research of learning styles in relation to VET has been carried out in Australia and the research suggests that, in general, VET learners are inclined to:

- be more visual than verbal (watching rather than reading or listening),
- prefer to learn by doing and practice, and
- prefer learning in groups and have clear guidance.

Naturally there is variation between different groups. For example, female students tend to be more verbal than males, while trade apprentices prefer a more hands-on approach compared to students in health and community services and business. Of course, the style preferred may not be the most effective.

One particular approach is that of personalized learning. This approach is centered on the needs, aptitudes, and interests of individual learners and has gained popularity in the last 10 years or so in a number of countries including the United Kingdom, the United States of America, Australia, and New Zealand. It should not be confused with sitting alone or self-paced learning. The distinctive features of the approach include:

- teachers' focus on individual students rather than an amorphous group of learners;
- the need for learners to become problem-solvers rather than passive recipients of training;
- assessment that includes analyzing goals, aspirations, and aptitudes at the beginning as well as achievement at the end of the program; and
- information and communications technology seen as key enabler.

We conclude our discussion of teaching and learning with a brief reference to two issues that have received much attention from policy-makers: the recognition of prior learning and noncompletion. Both are related to the efficiency of teaching and learning.

Recognition of prior learning (RPL) – also known as assessment of prior experiential learning and prior learning assessment – has been concentrated in Anglophone countries which tend to separate their initial VET from their secondary schooling, although it is now being widely pushed within the European union. The objective of RPL is to recognize learning that has taken place outside a formal education course. It is seen as a powerful instrument which motivates nonengaged learners, enables existing knowledge and skills to be formally recognized and therefore rewarded, and reduces the cost of training by eliminating time spent on further training. While the public rhetoric has been strident the realization has been a little disappointing with relatively small numbers of students taking advantage of it. This can be attributed to two factors. First, it appears that students often prefer to engage in whole courses of study or training. Second, the assessment of RPL can be expensive with providers

having difficulty relating the unaccredited learning of individuals to the learning outcomes of formal courses.

Completion rates have become key performance indicators for many VET systems. While a completion rate seems to be a sensible indicator of efficiency, noncompletion in VET does not always equate to failure. While a student may not have completed a course he or she may have obtained valuable skills, and the notion of completion itself is problematic when students can return to study shortly after withdrawing from a course. Nevertheless, noncompletion remains a serious problem in many countries, with the Organization for Economic Cooperation and Development reporting considerable variation with rates ranging from 35% in Greece and 42% in New Zealand to 85% in Belgium (OECD, 2007, Table A3.6). Rates vary by course and the personal characteristics of students.

The main reasons for noncompletion can be divided into two groups. The first is institutional factors such as quality of teaching, support for students, flexibility, and monitoring. The second revolves around the student: family background, financial position, intentions, previous educational experience, work situation, motivation, and so on.

A variety of strategies have been mooted to reduce noncompletion. Financial constraints can be addressed by student loan schemes (including income-contingent loans), vouchers, fee subsidies, and paid education leave. The dispositional factors of students whose previous education experience has not been satisfactory can be addressed by ensuring that the education and training is different from what they have experienced in the past. Other strategies tend to stress on a high level of pastoral care and institutional strategies such as the development of curriculum appropriate for intended students, target-setting with formative assessment and improvements to teaching.

The VET System

One thing that is very distinctive about VET is that it is structurally different, at least in many countries, from the school and university sectors. The long-standing apprenticeship model underpins VET in most Anglophone and European countries; this model is unique to VET. Thus, we begin with a discussion of apprenticeships. This is followed by a discussion of two planks that underpin VET systems – qualifications and quality assurance. We then focus on the teaching workforce and, finally, finance.

Apprenticeships date back to medieval Europe when young people went to live in their masters' houses to learn a trade. The essential components of a modern-day apprenticeship are:

- a training regime approved by governments,
- a combination of on- and off-the-job training,

- employers taking responsibility for the development of the apprentice, and
- the award of a qualification or license that enables the occupation to be practiced independently.

This model is not universal with trade training being undertaken in an institutional setting totally in some countries. It should also be noted that occupational licensing appears to be an important element. That is, the trades with the most successful apprenticeship system are often those that are regulated; the apprenticeship is a device for occupational restriction.

The traditional apprenticeship model was based on a model in which the apprentice and the master shared the cost. Under this model apprentices accept low wages in compensation for the costs incurred by the employer in providing the training. This model assumes that the apprenticeship is an indenture. However, this model has broken down with sanctions for breaking an indenture disappearing and apprentices requiring reasonable wages in order to attract them to the apprenticeship. Given that employers in theory will not pay for general training, then it seems that governments have to subsidize apprenticeship to avoid an underprovision of training (although a number of studies have shown that some employers have a strong commitment to apprenticeships despite the economic cost).

The traditional apprenticeship operated in the trades (e.g., plumbing, electrical, carpentry, cookery, and hair-dressing), although in the dual system (which we discuss ahead) it also operates in service occupations. In recent years, as the service sector has become more important, apprenticeship-like models – known as traineeships – have been introduced in a wider range of occupations in a number of countries, notably the United Kingdom and Australia and, to a lesser extent, the United States of America.

The model is facing a number of challenges though. First, it has difficulty coping with economic downturns because employers are less likely to take on an apprentice in a downturn. Structural change is also an issue with the trades declining in importance in developed economies. While it is possible to extend the model to new occupations this is not always appropriate. Many of the new occupations tend to be unregulated and so the model does not fit as comfortably as it does more traditional occupations.

The dual system refers to a specific model of VET that incorporates the apprenticeship model. The dual model uses alternating learning arrangements through partnerships between schools and companies, and is specifically used to describe arrangements in Germany, Switzerland, Austria, Denmark, and the Netherlands. The apprenticeship systems of other countries such as the United Kingdom and Australia have similar features but are not given the dual epithet.

Germany is seen as having the archetypical dual model. Here, the dual system functions as the major nonacademic route for school-leavers and recruits about 60% of 16–19-year-olds, resulting in the German labor market having a very small proportion of unskilled workers. Dual apprenticeships exist in nearly all branches of the economy including the professions and parts of the civil service. The fundamental characteristics of the dual system show that:

- It is a socially accepted pathway into employment. Training is workplace-led and practical. Skill requirements are occupation-based, not task-based.
- The state is involved with the quality of occupational standards as well as legal conditions. The German training culture means that the notion of apprenticeship is set apart from normal work.
- In addition to the government, other social groups have major influence. Public, private, and semiprivate institutions use long-established modes of cooperation, with employers and unions taking the initiative to modernize training regulations.

While not all countries have apprenticeships, virtually all VET systems have qualifications and quality assurance processes. Qualifications can be *ad hoc* or part of a formal framework and in recent years a number of governments have been using qualification framework as a driver of educational reform. The context is that governments have seen qualifications reform as a way to improve flexibility, widen participation, and enhance mobility. The reforms emphasize portability, allowing people to move between qualifications, and transparency, by making explicit what learners have to do to become qualified. Qualifications are seen as motivators for learners but are also very useful instruments for governments which want the VET systems to be more accountable.

Two approaches can be seen. The first is an outcomes approach which incorporates definitions of qualifications, sets of descriptors, and a precise specification of outcomes. This approach has been criticized as invariably leading to overspecification and the trivializing of learning. The second is an institutional approach in which qualifications are not treated as separate instruments of reform but are embedded in the wider education and training system. This approach has been criticized for being slow to adapt to new demands and impeding movement between education sectors. The conclusion, or at least one conclusion, is that qualification systems have a relatively modest role in reform. Successful qualifications are dependent upon establishing communities of trust and these take time to form.

Quality assurance practices vary from one country to another but are an integral part of any system. Quality assurance needs to address three key risk areas. The first is associated with the specification of knowledge and skills the

learner is required to acquire. The second is the risk associated with the training provider. The third relates to the extent to which a VET system meets its wider purpose – meeting the needs of the locality or country.

The first risk is typically managed through the development of systems that provide for expert input. Examples are the Sector Skills Councils in the United Kingdom, Industry Skills Councils in Australia, and Standards Generating Bodies in South Africa. The challenge is to ensure that standards encapsulate all the knowledge and skills relevant to the occupation in question and to ensure that the standards are specified with sufficient clarity to ensure assessors can make reliable judgments. One issue with such systems is the time taken to develop and keep the standards up to date. If there is a trend in respect of such systems, it is a move towards more differentiated, outcome-focused models that can allow both high quality and poor performance to be addressed.

The second risk, relating to providers, can be broken down to a number of elements:

- risks associated with the assessment process;
- resources risk – does the provider have appropriate premises, equipment, and staff; and
- funding risk when public funds are being used – the provider needs to manage the training in accordance with the terms of the funding.

Different countries take different approaches to mitigating the risks. In Australia, for example, quality assurance is largely front-end loaded with an emphasis on approving and auditing providers. By contrast, in the United Kingdom more emphasis is placed on both internal and external inspection.

The third risk in whether the VET system meets its purpose, is typically mitigated through the development of national and regional skill strategies. The aim of such strategies is to provide direction to the VET system, especially with relation to the use of public money. Approaches can be demand-driven, with students or businesses determining what courses are provided, or supply-driven with governments setting national targets. Neither approach guarantees success because of the imperfect information and the impossibility of anticipating the future needs of the labor market.

An important element of any VET system is its workforce. As with many other elements, the process used to train the teachers is quite different from those used in schools and higher education, at least in many countries. Some of the differences in the way VET teachers are trained can be traced to historical influences, with vocational education associated with education for work as distinct from education for culture. The education for work aspect has placed emphasis on occupational experience and many VET teachers (or practitioners) have a background largely through occupational experience rather

than an initial teaching qualification (the craftsman-turned-teacher as distinct from the professional VET teacher). It would seem obvious that a teacher who has never been an electrician, for example, would have no credibility in teaching students to become electricians. On the other hand, proficiency in an occupation may be a necessary condition to being a good teacher but it is not sufficient, given the pressures on teachers to understand new methods of teaching and learning, deal with diverse populations, get on top of new technology, and so on. VET teachers can also be seen as change agents because successful reform can only happen if teachers embed the reforms in their daily work.

Internationally, the professional VET teacher is common in many European countries while the craftsman-turned-teacher is more common in the United Kingdom and Australia, for example. In the United States of America the former is common for the trades/industrial arts while the latter tends to be the model in other vocational areas.

A widespread phenomenon is the use of part-time or casual teachers. This practice is often driven by budgetary pressures and can have negative effects if such teachers are treated as second class and not provided with suitable professional training. On the other hand, such teachers can be very valuable by bringing valuable occupation experience to the classroom. The age distribution of VET teachers is also an issue in many countries, with the VET teacher population being considerably older than other teacher groups. The reason for this is the reliance on teachers with many years of preteaching occupational work experience.

Professional development continues to be a challenge. Elements include support for new teachers (particularly important as a retention strategy), mentoring, opportunities to return to workplace settings to refresh occupational skills, and the development of communities of practice. Professional development is important not only to improve teaching but also to upgrade the status accorded to VET teachers – low status compared to other teachers has been a perennial issue. This is against a background in which VET teachers require a need for a more complex knowledge base than school or university teachers, working in a field that lies between specialist theory and specialist work.

We end our coverage of the VET system with a discussion of finance issues. Training finance has two purposes: it funds the national training system and also has a role in achieving many of the broader goals of national training policy. In conventional training markets there are two components. The private sector training is market-driven, nonsubsidized, and competitive. Firms provide training to their workers and payment is made through initial fees or below-productivity-level wages. Individuals enroll in fee-paying preemployment courses. In most countries there is a parallel public training system, often dominating the supply of structured formal training. Public sector is

predominantly funded by government, and fees, if any, are nominal. Areas covered by public training leave little incentive for private training providers to develop courses. Because of the difficulty of aligning public budgets with outcome measures there is little incentive for the public training sector to align training courses with the needs of the labor market. The very low fees also mean that students are less inclined to see their training as investment in skills from which they expect a return in the labor market. Public training provision remains essentially supply-driven, and this framework has become inadequate to meet skill development needs. Voucher schemes, although largely still experimental, are intended to build an open market and shift providers away from a supply-driven system. Contract financing of government programs is also intended to open the training market to private providers.

We see a trend away from the traditional method of training finance in many countries. Increasing pressures on government budgets have led to a search for alternative sources of funds. At the same time governments have become more interventionist in the training markets in order to counteract the tendency for underprovision of training (because of the presence of market imperfections including externalities – the benefits of skills spilling over to others – and imperfect information).

A number of alternative sources of funding have been pursued: earmarked training taxes or levies, cost-sharing with the beneficiaries (i.e., raising fee levels), additional institutional income through the commercial sale of training services, and an expansion in private training. Various constraints have held back the latter including financial constraints, rigid fee policies, excessive regulation, and information gaps. In addition, enterprise training can be encouraged through levy–grant schemes (incentives to encourage firms to invest in training), direct public subsidy, and through the creation of a training fund.

VET in Context

Arguably, the VET sector has a more complex relationship with society than other educational sectors. This is because it has a multiplicity of roles. It is seen as an agent of social change because in many cases it provides a second chance for disadvantaged groups and a way of integrating such groups into the mainstream labor market. It is also seen, for example, as facilitating the school-to-work transition and a way of spreading innovation and good practice within industry. Indeed, on occasions it seems that VET is being asked to solve all the problems of the world.

In terms of VET's social role, many groups can be identified as being a focus of VET policy including people with a disability, indigenous people, immigrants,

displaced workers, prisoners, older people, and women. In each case the motivation is the same: to provide skills that enable the individuals to lead better and more productive lives. We discuss each of these groups briefly.

People with a Disability

Distinctions are commonly made between intellectual disabilities, learning disabilities, and physical disabilities. For training providers, the primary consideration is the extent to which the disability impinges on the ability to complete a program of study successfully. Apart from the direct effect of a disability on the ability to undertake training, many disabilities can have an effect on the ability to get to training and to participate in work placements (and subsequent capacity to obtain employment). While there has been much support for people with a disability in many VET systems, one matter of concern is that recruitment decisions by employers are made outside the VET systems, and therefore study success does not automatically lead to labor market success.

Support for students with a disability ranges from quite modest to very high levels (e.g., note-takers and sign language interpreters). Funding is typically allocated to students or providers. There is also debate between those who wish to integrate the support for students with a disability within normal training provision and those who think support should be left to experts. A complication is that some students with a disability take longer than their peers to complete their study and funding models often have difficulty accommodating this. There can also be undesirable feedback loops; a provider providing good services to students with a disability will attract more such students, thus exacerbating funding difficulties. One factor that is pointed to in improving support for students with a disability is staff development.

Indigenous People

Definitions of indigenous are contested but a working definition used by the United Nations is along the following lines:

- priority in time, with respect to the occupation of a specific territory;
- the voluntary perpetuation of cultural distinctiveness;
- self-identification; and
- an experience of subjugation, marginalization, dispossession, exclusion, or discrimination.

It is estimated that there are between 300 and 500 million indigenous people worldwide.

For many indigenous people poor schooling means that VET is not only about vocational or technical training but also a means to understanding and engaging with

the outside world. As an education sector, it is arguably the most relevant to capacity development. It is seen as a road to employment and poverty alleviation.

As noted earlier in this article, VET can be delivered in many ways including relatively informal settings. This has the advantage of flexibility but can also lead to fragmented and uncoordinated delivery. Challenges for building effective VET programs are many. Often many indigenous people live in remote and isolated communities where employment options are few and there is low access to even basic levels of education, housing, communication, and health. This means that such people can be locked into being particularly disadvantaged, and this identification with disadvantage can flow over into those living in urban areas. Language and the attitude of many indigenous people who are suspicious of Western education are further challenges.

Key strategies for advancing indigenous engagement with VET include recognition of the indigenous culture, partnership and consultation, blending vocational education with more general education and life skills, using VET to support indigenous knowledge, ensuring that the VET delivered is up-to-standard, reducing barriers to access (e.g., provision of childcare), and time-flexible delivery (not cutting across agricultural seasons and ceremonial cycles, for example).

Immigrants

A central issue in migration policy is the integration of immigrants into the labor market. One reason for immigrants doing poorly in the labor market is their lack of skills relevant to the receiving society. Part of the reason for this is that they often lack qualifications (or the credentials are not recognized) and low-qualified individuals tend in any case to have lowest training participation rates. Similarly, the nature of their employment (e.g., many will be in casual employment) is not conducive to receiving training. Immigrants might also invest less in education and training because of actual or perceived discrimination by employers (which would reduce the financial return to training). Similarly, employers might invest less in the training of immigrants if they believe that they might return to their home country. Cross-country studies suggest that in most countries the native-born undertake more education and training, but this pattern is not universal. However, it needs to be noted that some training, particularly language training, is targeted at immigrants. There is surprisingly little research on the pay-off to VET for immigrants. Israeli research suggests that there is a payoff in Israel, while German research suggests that Turkish vocational school graduates are less able to find appropriate jobs when labor market conditions are poor. Australian data show that strong rates of participation in VET have

not been sufficient to deliver corresponding rates of employment for some immigrant groups.

Displaced Workers

Displaced workers are obvious candidates for VET. It is clearly sensible to provide displaced workers with skills that are in more demand in the labor market; thus they are often targeted with special assistance programs. Evaluations of a number of programs show mixed results. For example, evaluation studies of the Trade Adjustment Assistance program in the United States show moderate gains in the short term and weak gains in the longer term.

One of the challenges in evaluating programs is to ensure that they are scientific, in the sense of making use of experiments or quasi-experiments that provide valid comparison groups. Nonscientific evaluations tend to overestimate the positive effects. However, overall, scientific evaluations have found that there are some improvements in employment prospects and, in some cases, income, particularly in Europe. Of course, clearly the nature of the training must be important and it is likely that the amount of training will have an impact (a week of training is a very different prospect to a year's). Also retraining programs are often evaluated against simpler and cheaper job search assistance programs and costs are rarely taken into account.

Prisoners

Prisoners are typically a very disadvantaged group, with low levels of prior education and often a history of substance abuse. VET has the potential to be of assistance to this group, with the hope that the acquisition of skills will enable an easier integration back into society and lower levels of recidivism. Certainly, VET has the potential in providing marketable skills, although basic literacy and numeracy are probably more important for those with particularly low levels of basic education. There are two broad models of prisoner education. In Nordic countries, prisoners are seen to have rights, one of which is access to education. Another model, prevalent in North America, United Kingdom, and Australia, takes a risk-needs approach in which education and training are seen as a way of reducing reoffending through programs targeting factors such as antisocial attitudes and substance abuse as well as providing skills that would help employment prospects.

Older People

The learning needs of older people have received much attention in many countries, especially in Europe. The motivation comes from two angles. First, the aging population and economic displacement make it important to

increase the proportion of older people who are active in the labor market. Second, learning in later life is seen to have a wide range of positive social effects including life satisfaction, self-esteem, and willingness to take responsibility and increased involvement in community activities. Thus many international institutions emphasize the importance of lifelong learning.

However, participation rates in education and training tend to decline with age. This pattern is consistent to some extent with the argument that front-ended education has a greater payoff; there is a longer period to reap the rewards of greater skills. This argument holds from both employee's and employer's points of view. Another factor is that there are factors other than employer attitudes that reduce training for older workers who often consider themselves unsuited to new learning and lack confidence. Some have negative attitudes to learning and have a disability or health problems. Cost can also be a barrier. More educated workers tend to get more training and this acts against those older workers who are relatively low-skilled. In terms of designing programs for older workers much can be done to take into account older people's needs: a more personal touch, shorter courses, more nonaccredited courses, flexible provision, a consideration of preferred learning styles, a consideration of gender issues, and so on.

Women

Women, as a group, have been the subject of VET policy in both developed and developing countries against a background of occupational segregation and low status. The proportion of women undertaking VET has increased, but this is mostly due to a widening of the occupational coverage of VET rather than success in attracting women into nontraditional occupations. Women have found VET – particularly its provision of courses in general education, business, community services, retail, and hospitality – a useful springboard for re-entry into the labor market. Occupational segregation seems to be very resilient despite attempts to use VET as a change agent.

Moving on, VET is seen as a lever in the wider society in a number of respects. The starting point here is the human capital paradigm in which education is seen as an investment in the skills of individuals which provides a return in terms of more stable employment and higher earnings. The issue of primary interest is which forms of education and training provide the best payoff rather than any argument over the general benefits of education and training. In the allocation of scarce resources VET needs to compete with early childhood education, school education, and higher education. The research suggests that policy needs to ensure that a lot of investment occurs in early life because cognitive skills are less malleable later

in life. By contrast, noncognitive skills can be enhanced into adulthood and this is generally the period targeted by VET. An emphasis on early education does not address the skills deficit of the current generation of adults. Fixing the problem of unskilled adults is hard and costly and it does appear that in a lot of countries the policies being adopted are based more on faith than sound evidence.

We need to go well beyond a simple notion that the role of VET is to provide skills that are valued in the workplace.

First, VET can be a force for innovation. In looking at the role of VET in innovation it is first useful to look at the process of innovation. Innovation today has a lot to do with worker involvement, the idea of sequential innovation – from concept design to refinement and manufacturing – involving many actors. In addition, the distinction between radical and incremental innovation seems to be breaking down, and hence the role of practical knowledge and experience is increasing. Thus it is argued that experience is a very important source of innovation, with new concepts of production giving workers a role as a reflective practitioner meeting the demands of fussy consumers. The question then arises ‘What form of VET trains workers best as innovators?’ It is argued that it is the dual systems and apprenticeship systems that provide the highest level of useful practical experience when compared to institution-based systems. Even a three-year apprenticeship provides limited practical experiences, but this is a much better starting point for augmenting the experience-based than school-based training. Of course, creating innovators does not end there. VET systems need to go beyond simple goals such as employability. They need to aim for graduates who can shape their work: they will need attributes such as expertise in the subject field, virtuosity of performance, social responsibility, and the ability to be reflective. Finally, VET must balance theory and practice; apprenticeship-based systems need to emphasize theory to complement the abundant practice, and school-based systems must provide access to work practice and relate it to theory. This is indeed a challenge.

VET is seen as being an aid to school-to-work transition in an environment which is becoming increasingly difficult. VET has been assigned a number of roles. It has been called on to provide programs to keep students longer at school, and is seen as providing practical skills and work experience to make school-leavers more employable. The apprenticeship path is an obvious success, with its combination of training and work experience. The fact that apprentices have jobs while they train gives them a headstart, particularly if their employer is inclined to keep them on after the apprenticeship is completed. However, despite widespread support for upper secondary VET programs, results are variable. It is not clear if they result in higher school retention. Australian evidence points to a positive impact on postschool employment

(Anlezark *et al.*, 2006, for example). Training provided by labor market programs is more ambiguous in its impact, and questions the effectiveness and efficiency of VET in supporting young people experiencing difficulties with the school-to-work transition.

VET has also been called on to promote the cause of sustainability. It is argued that the key to sustainable development is integrating VET into approaches to development that balance economic and social progress, address cultural difference, and respect ecological values and limits. VET needs to be imbued with values of environmental sustainability, economic sustainability (to supplant the notion of productivities), and social sustainability. Such a push is a challenge since VET is more commonly seen as training for growth or skills for work rather than a vehicle for promoting a broader general education for personal autonomy, citizenship, and sustainability, as advocated by the authors.

Almost all the discussion to date has implicitly assumed that VET is operating in developed or at least stable countries. Protracted conflict leads to widespread destruction of already fragile education and training systems. Young people take up arms instead of learning employment skills. Even when conflicts are over, problems remain. Infrastructure is destroyed, gainful employment is scarce, and often skills available do not suit the job opportunities that do exist. A clean slate provides a window of opportunity, but success in building a VET system depends on political stability and the level of international investment available for reconstruction. New systems tend to address the needs of the formal sector, which invariably is very small, and ignore the development of skills in the informal sector where the majority of people work.

The majority of this article has displayed VET in a positive light although challenges abound. However, there are some issues that need airing. The first is VET’s lower status relative to higher education, primarily because of its association with occupations that have lower status than the professional occupations to which university graduates aspire. Parity of esteem has been an education and training policy goal in many countries, especially in the European Union. The status of VET teachers is also an issue, as has already been noted. Suggestions for raising the status of VET include having education systems that are flexible (e.g., double qualifications and good pathways from VET to higher education) and employing VET to modernize teaching through situation- and work-based learning theories rather than a reliance on academic models. It must be noted that citizens value their career and occupation choices against their aspirations and therefore high status for VET will only come from it being of a high quality and meeting aspirations, not by government decree.

A related policy area is the role of career guidance. Career guidance is seen to be more than advice on

initial training and employment; it is now seen in the context of lifelong learning (lifelong guidance) and is given a wider ambit to cover many aspects of an individual's life (lifewide guidance). Career guidance has traditionally been the purview of educational institutions but now is becoming more common in a number of countries (e.g., Denmark, Ireland, and the United Kingdom) in a workplace setting where it has an emphasis on recurrent VET.

VET is also often called on to solve labor shortages that are often a product of labor market imperfections rather than inadequacies of the training system. While it is easy to blame the training system for shortages, there is a range of factors that need to be considered, notably the slowness of wages to adjust (in a flexible labor market there would be no such thing as a shortage), slowness in adjustment of supply (e.g., length of training, lack of interest by students, employers reluctant to hire qualified but inexperienced workers), and inadequate labor market information to firms, individuals, and education planners.

Concluding Comments

There can be no doubt that VET is an important education sector, and many governments have been giving it increased attention. While general education is very important, and there is no stopping the rise of higher education, there is a widespread view that VET is particularly important because of its focus on skills needed for the workforce. It is also important because it, more than the other education sectors, plays the preeminent role in addressing the disadvantaged.

However, it faces many challenges. The first is, despite much public rhetoric, it has low status relative to higher education. This is unlikely to change while managers and professionals occupy jobs that are better paid and more prestigious than tradespersons, technicians, and para-professionals. It is also unlikely to change unless the education courses leading to professional jobs change to incorporate VET. And this is possible by basing such courses on a sophisticated implementation of competences and contextualizing theory to a much greater extent than is the common practice.

A second challenge is that the direction of structural change does not favor those occupations which VET has traditionally served. Reich (1991) categorizes workers into routine producers, in-person service providers, and symbolic analysts – those who simplify reality into abstract images that can be played with and then transformed back into reality. VET tends to be targeted at the first two of these categories but it is the third, which is largely made up of professionals, that is growing the most rapidly, at least in developed countries. The relative decline of

agriculture and manufacturing, and the growth of services is part of this trend.

VET is also being squeezed by ever-increasing educational levels. We see para-professional groups aspiring to become professionals, and an obvious mechanism to effect this is to move the training from VET to higher education. Thus the VET sector may need to expand the qualifications that it covers or risk being left with a decreasing market share.

The pace of technological change also threatens the underlying orientation of VET. If technology is changing rapidly, then the skills required must also change. But, in general, the possession of general analytic skills makes it easier to update skills compared to the possession of only technical skills. VET tends to focus on the latter rather than the former. Maybe it is time to increase the amount of general education embedded in VET courses, otherwise VET-trained workers will find themselves at a disadvantage in a changing workforce.

VET also risks being assigned a task that is too ambitious. It can be asked to do too much. The skill shortages in the labor market – VET should remedy this. Disadvantaged groups struggle in the labor market – VET is the sector to address this. The economy needs to be more sustainable – let's call on the VET sector to solve this. While education and training are vital for a productive workforce, it needs to be remembered that many skills are learnt on the job and not from formal training and there are many imperfections and distortions in the economy which impinge on the productivity of the workforce. VET is a vital element of creating a skilled workforce and a just society but it is only one element.

Thus VET faces many challenges. To prosper it will need to continually reinvent itself in order to bend itself to changing occupational structures, technology, educational trends, and social conditions.

Acknowledgments

The author is happy to acknowledge that little of this essay is original and makes very liberal use of the material in the individual articles in the encyclopedia. That said, all mistakes and infelicities are mine.

See also: Age Equality in Education and Training; Apprenticeships; Career Guidance in Vocational Education and Training and in the Workplace; Displaced Workers, Unemployed and Vocational Education and Training; Dual System; Immigrant Investment in Host Country Vocational Education and Training; Indigenous People and Vocational Education and Training; Learning Styles in Vocational Education and Training; Learning to Work for the Future; Modularization in Vocational Education and Training; Non-Completion in Vocational

Education and Training: Issues, Patterns and Possibilities; Personalized Learning and Vocational Education and Training; Persons with a Disability and Vocational Education and Training; Prisoner Education and Training; Quality Assurance in Vocational Education and Training; Skills Shortages: Concepts, Measurement and Policy Responses; The Status of Vocational Education and Training; Training and Learning in the Workplace; Training Finance; Vocational Education and Training and the School-to-Work Transition; Vocational Education and Training in Post-Conflict Countries; Vocational Education and Training Workforce.

Bibliography

- Anlezark, A., Karmel, T., and Ong, K. (2006). *Have School Vocational Education and Training Programs Been Successful?* Adelaide: NCVER. <http://www.ncver.edu.au/publications/1654.html> (accessed September 2009).
- Organisation for Economic Co-operation and Development (OECD) (2007). *Education at a Glance: OECD Indicators 2007*. Paris: OECD. <http://www.oecd.org/dataoecd/4/55/39313286.pdf> (accessed September 2009).
- Reich, R. (1991). *The Work of Nations: Preparing Ourselves for 21st Century Capitalism*. London: Simon and Schuster.

Informal and Non-Formal Learning in Vocational Education and Training

F Beddie and S Halliday-Wynes, National Centre for Vocational Education Research, Adelaide, SA, Australia

© 2010 Elsevier Ltd. All rights reserved.

Is watching *The Simpsons* a form of learning? Where does entertainment end and learning begin? Is watching television informal or incidental learning or just relaxing? Does the answer depend on the viewer's intention or on a change in their knowledge about the world? Can labeling an activity as education detract from the benefits of learning, which seems more like fun? Do these distinctions matter?

These questions highlight the complexities faced by educators, employers, and regulators when deciding how to define different forms of learning. Although the idea of informal adult education has been around since at least the 1950s, when the American educator Malcolm Knowles used this category of learning in his attempt to develop a coherent and comprehensive theory of adult learning, it is only more recently that we have seen a preoccupation with the learning that happens outside of the formal education system. This has come about in tandem with an emphasis on learning as an activity no longer confined to the early school-based phases of one's life; but as one which is lifelong and lifewide. We have become ageless learners, moving through various phases of life, being motivated to learn in different ways to obtain new understanding, skills, and attributes (see **Figure 1**).

Informal learning and nonformal learning are the primary foci of this article. We give an overview of how formal, nonformal, and informal learning can contribute to people's lives in various parts of the world and in a variety of work settings. We consider how informal and nonformal learning influence the lives of people in a variety of diverse geographical areas including, the muddy soccer fields of England, the markets of West Africa, and the forests of northeast Thailand. We look at the role of learning in workplaces and reflect on what the increasing focus on informal learning means for vocational education and training (VET) systems. We conclude with some questions about the extent to which less formal sorts of learning should be defined.

Definitions

Educators and trainers are being asked to respond to fast-moving global economic, social, and technological developments; changing labor markets; and constantly evolving skills sets and new types of work and workplaces. Recognizing these challenges, the Organization for Economic Cooperation and Development (OECD) has embarked on

a project to examine the recognition of formal, nonformal, and informal learning. In doing so, it offers some definitions. These are a starting point for a discussion about the forms of learning that education and training systems around the world must embrace or, at least, recognize:

Formal learning: learning through a program of instruction in an educational institution, adult training centre or in the workplace, which is generally recognised in a qualification or a certificate.

Non-formal learning: learning through a program but it is not usually evaluated and does not lead to certification.

Informal learning: learning resulting from daily work-related, family, or leisure activities. (OECD, 2005: 5)

Another way of thinking about these forms of learning is to consider the extent to which they are structured and whether they lead to a credential: a student who enrolls in a computing course, which results in a qualification, has been involved in formal learning. An employee, who attends a day's training to learn a new software application being used in their workplace, is engaged in nonformal learning. And the manager who asks his assistant how to change the look of a PowerPoint presentation has acquired a new skill through informal learning.

These descriptions are not definitive. Indeed, grasping the nature and outcomes of learning, regardless of format, is a matter of debate among researchers, educators, and assessors. According to the Australian Bureau of Statistics (ABS), informal learning constitutes unstructured and noninstitutionalized learning (ABS, 2007) and is more likely than not the product of incidental learning in the exploration of everyday experience (Conlon, 2003). The ABS, along with other collection agencies such as the OECD and *StatsCanada*, has decided, at least for the purpose of measuring learning, it is essential to put some parameters around notions of informality. In its first survey of adult learning, published in December 2007, the ABS posed the following caveat: "Activities may occur on a self-directed basis, but are excluded from scope if there is no specific intention to learn."

While some researchers have questioned this idea of intention (e.g., Baran *et al.*, 2000; Billett, 2001), it does assist in discussing definitional boundaries when these become necessary, for example, for purposes of assessment.

The definitional task is further complicated by the wide-ranging methods of informal learning, which include: reading

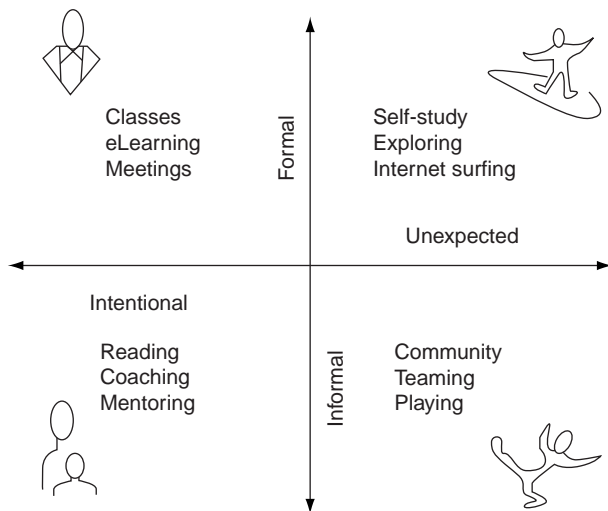


Figure 1 Ageless learners.

journals or articles, participating in workplace mentoring schemes, learning through casual conversation, engaging in on-the-job training, and learning through trial and error. There can be good reasons for people wanting to acquire skills and build on knowledge in these informal ways. Often it will be more efficient – in terms of both time and money – than enrolling in a formal course. For others, education, training, and assessment are terms that conjure up negative memories of school. For them, an informal approach may be more productive, precisely because there are no tests or accreditation processes to encounter.

On the other hand, many adults prefer the discipline and social interaction of the classroom. In terms of vocational training, however, off-the-job training courses may not provide all the necessary content and context for employees, and will be most valuable when reinforced by nonformal and informal learning in the workplace.

In 2007 an OECD policy brief noted that if lifelong learning was to be of value to individuals, member countries would have to reconsider what they meant by the term qualifications. It declared that:

For many, “qualifications” refer to a piece of paper proving someone has successfully completed a specific course of study. But the “qualifications” of use in lifelong learning may be much broader – skills and techniques for performing your job on a day-to-day basis learned from a colleague or workplace mentor, for example. If you stay in the same job, your employer is aware of these added skills, but in a world where jobs are no longer for life, the individual needs a way to communicate this added knowledge to new employers – and they need a way to evaluate these skills. Even paper qualifications tend to lose much, if not all, of their value when someone wants to go and work outside the sector or country that provided the certificate in the first place. (OECD, 2007: 3)

The OECD goes on to argue that the recognition of informal learning is central to creating portable qualifications. If workers learn how to operate a particular machine or software from a more experienced colleague, they may not acquire a formal qualification but have acquired a specific work skill. According to the OECD, recognizing these skills is a cost-effective measure that avoids the need for formal study and ensures that the skills learned are relevant to job requirements. However, the recognition processes are not cost-neutral. They usually involve considerable time in evidence gathering, especially for those with little printed documentation to demonstrate their learning, and a skills assessment, which can be expensive if it involves visits to the workplace (Cameron, 2004). Moreover, recognition is not always necessary, with employers often prepared to recognize specific experience without requiring a formal credential.

Such recognition would also result in a more accurate picture of the extent of adult learning today. Conner (2002) estimates that informal learning accounts for over 75% of the learning taking place in the workplace, although much of that activity is not recorded in official accounts of vocational training. Richardson (2004: 5) suggests that on-the-job training (informal learning) is “the most commonly experienced form of skills development provided by employers.” She goes on to argue that we should pay careful attention to how skills learned informally on the job affect wages and career paths. This type of skills development also needs to be taken into account when measuring the contribution made by employers to workforce training.

Informal learning, according to Livingstone (2000), “is like an iceberg mostly invisible at the surface and immense in its mostly submerged aspects.” Livingstone used this image when discussing the relationship between work and learning, including the knowledge needed not only for paid work, but for volunteering, household- and general-interest activities (Livingstone, 2000: 19), all of which contribute to a person’s vocational skills sets. Becket and Hager (2002) argue that in comparison to formal learning, informal learning is not only more organic but more widespread and effective. But how do we grasp hold of this diffuse and submerged mass of learning? When is it appropriate to do so? (See **Case Study 1** and **Figure 2**.)

Informal Learning across the Globe

That governments around the world are grappling with the difficult task of measuring the extent of informal learning is indicative of its role in helping vocational systems, employers, and individuals cope with contemporary workforce trends such as: greater demand for skilled labor in many industrialized countries; the increasing mobility of workers within and across borders; the growing interest in the

workplace as a learning environment and acknowledgment of the importance of on-the-job learning in skill formation; and the goal of ensuring education for all, including second-chance learners. In the context of equitable development, informal learning should not be identified as just a hook to bring people to more formal avenues of learning and work; it should be recognized as a positive and reliable learning experience in its own right. This is particularly the case in developing countries, where universal formal schooling is not the reality, and where informal learning is sometimes the only option for an individual or a community to gain knowledge and skills (see **Case Study 2**).

The knowledge economy is now placing greater emphasis on creative or more lateral thinking. This presents a challenge to countries that, in the post-World-War-II era, adopted an emphasis on rigid educational

outcomes in order to hasten economic development. These countries are now finding that there is a risk of losing creative ability and are therefore reconsidering the best way to encourage learning. In Singapore, for example, the last decade has seen a shift from an efficiency-driven education system to one where the emphasis is on encouraging ability, encapsulated in the slogan “Thinking Schools, Learning Nation.” Here, the premise is that Singapore’s future depends on the capacity of its people to learn continuously through their lives (Boon and Gopinathan, 2006: 40).

Others, particularly in the developing world, are looking at how informal learning can be used to engage more of their populations. King (1982), whose studies looked at amongst others, Yemen and Senegal, has suggested that in the absence of formal training, skills are often acquired through familial relationships and participation in petty production. For example, someone selling goods in the marketplace becomes numerate by learning from their peers and from experience. This does not, however, detract from the importance of focussing on formal education in basic literacy and numeracy which, King suggests, can increase the gains offered by informal training.

The United Nations Educational, Scientific and Cultural Organization (UNESCO) claims that informal and

Case Study 1 Canada

Guo (2006) estimates that overall Canadians spend longer in informal learning activities than they do enrolled in formal courses. Guo was looking at the impact of this learning on volunteers who were Chinese migrants to Canada. Guo conducted this study through personal interviews and questionnaires, and investigated amongst other themes, motivation and benefits of volunteering. He found that this informal learning assisted individuals to acquire the necessary skills and factual knowledge to adapt to their new home, and also helped increase self-confidence, build networks (both personal and professional), and promote a sense of ownership and connection to their new community. Guo saw that this learning was passed on to family members and neighbours, creating a ripple effect. This suggests that informal and non-formal learning not only has impact that we cannot see (Livingstone’s iceberg) but also has influence across the community.

Gou, S. (2006) Mapping the iceberg of informal learning: Exploring the experience of Chinese volunteers in Vancouver. *Conference Paper, 2006 National Conference On-line Proceedings*, York University, Toronto, ON, May 1006.

Case Study 2 Korea

Traditional Korean pedagogy marries education with wisdom, self-improvement, maturity, and a lifelong appreciation of learning. This holistic approach embraces the concepts of informal and non-formal learning but appears to be in danger of being lost in the development of contemporary Korean emphasis on examined and controlled learning environments.

Pak, M. S. (1997). *Two Wheels for Lifelong Learning in Korea: Credit Banking & Multimedia Technology*. CEDEFOP Discussion Paper.

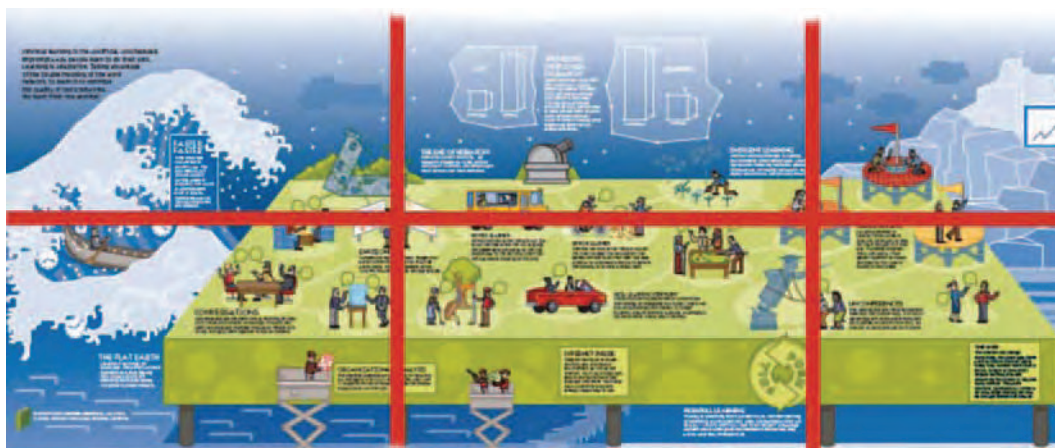


Figure 2 Informal learning by Jay Cross.

nonformal education is a necessary avenue to attain “learning for all” (UNESCO, 2007: 4). This suggests the need to ensure that we understand how informal learning can be harnessed to maximum effect. A research project in Thailand undertook this task and uncovered a process of unstructured learning acquired by active community participation, experience, and practice, which had far-reaching and positive effects on the standard of living of a whole village (see **Case Study 3**) (Thacheen and Lauzon, 2006).

Learning in the Contemporary Workplace

Informal learning is a constant in the contemporary workplace. Such learning takes place as a result of practice and experience and usually goes unrecognized as anything more than getting on with the job. Often there is no need to recognize these processes, which work well as a means of developing expertise and judgment. There are, however, times when it is important to identify informal learning. It may be, for example, that such learning is resulting in bad work practices and needs to be undone, often by instituting formal training. In other instances, the regulatory environment might require certification of skills. And sometimes workers will want to document their skills in order to seek promotion, higher wages, or to move jobs. This requires the individual to articulate what knowledge and abilities he has gained from that experience. Such a process can involve formal assessment, although for many professionals and para-professionals this is the familiar process of compiling and updating curriculum vitae.

Just as in the developing world, where basic literacy and numeracy are a vital foundation skill, accredited training, informal and nonformal courses are important elements in workplace learning. The 1997 and 2001 British Skills Surveys, by Nordman and Hayward (2006) found that workers require both formal and informal training, with the type of learning often dependent on the complexity of the skills required to do the job. Such findings illustrate the value to employers, employees, and trainers of understanding how workplace learning is best conducted.

This is not, however, a simple matter of categorization of the learning. The UK Learning and Skills Research Centre has argued it may be more beneficial to conceive formality and informality as attributes present in all circumstances of learning, with these elements emphasized (or not) according to the context and the learner (Colley *et al.*, 2003: 9). For example, it can be more effective, when trying to bring disengaged learners back to formal training, for vocational trainers and adult educators to decide not to make the pathway to a qualification an explicit goal; but rather to encourage the learner to undertake the course of study at their own pace and without fear of assessment, which might be taking place surreptitiously and which might well include the recognition of existing skills (see **Case Study 4**) (Parker, 2006).

Much discussion about enterprise learning has focused on the apparent lack of training occurring in small- and medium-sized organizations. A study by Ashton *et al.* (2005: 6) suggests this deficit may have been exaggerated by a focus on the formal training courses typically commissioned by larger organizations. The authors argue that if attention is paid to skill formation rather than to the training of individuals, a different picture emerges, one

Case Study 3 Thailand

The people of Bann Na Isarn, in north-eastern Thailand, have improved their quality of life by learning about how to manage the natural environment and achieve a balance between forestry and rice farming. By following the example of a head man in a neighboring area, one member of the Bann Na Isarn community (Liam) persuaded his fellow villagers to drop their dependence on cash crops, which brought in little money and often resulted in the village going hungry, and to concentrate on community sustainability and cooperation.

Liam started by planting a range of new plants and trees that were not only useful for food, but could be used as building materials and for medicinal purposes. Despite initial community resistance, Liam persuaded others to become involved in this practice of agroforestry.

The success in Bann Na Isarn lay in having a meaningful context for the learning and in embedding cooperation. It was reliant on skills and knowledge being passed on by community elders and across generations, as well as on a conscious effort to share practices and techniques through conversations and demonstrations.

www.voced.edu.au/td/tnc_86.564 – 18k

Case Study 4 England

According to Andrew Parker (2006) from the University of Warwick, formal learning is often only a small component of the whole learning experience and imparting knowledge by tangential means is often just as important and effective. This view is borne out of a 2005 study into the experiences of young men participating in an occupational apprenticeship of professional football. Parker suggests that this apprenticeship model provides an example of situated learning, which takes account of the specific, male characteristics of the learners by embracing physical fitness, discipline, and loyalty to teammates and the club, including regulations and procedures. The learning can also be characterized by negative masculine behavior including perpetuating sexist and racist attitudes and entering into abusive dialogs. Nevertheless, the apprentices acquire knowledge through formal means (football skills, game tactics, etc.), and informal interaction, including locker-room banter and peer-group pressure. This combination leads to the boys acquiring not only some certificates in sports-related subjects but also a degree of work-readiness that they may not have achieved outside the particular discipline imposed by the football club.

that includes both informal and formal learning as part of the transfer and acquisition of the knowledge and skills required for performance in the workplace. Many small-business owners and managers are likely to learn informally or incidentally, for example, through discussions with suppliers, customers, trade meetings and other business activities, rather than by attending training courses. Other learning takes place by solving problems with colleagues, observing and working with more experienced colleagues, informal chats, moving between jobs, and participating in team-based activities. The authors suggest that in larger firms much of this informal learning becomes more structured, either through explicit coaching and mentoring programs or specialized training activities (Ashton *et al.*, 2005: 22).

Mentoring is, however, more likely than not to be an informal activity. Indeed, as we see a trend toward the deliberate integration of learning and work, the employee is expected to learn from everyday work activities and through guidance from other workers. In this situation, colleagues become mentors and managers must assume the responsibility for creating a workplace that is conducive to learning.

Trainers, too, have to adapt to this sort of learning environment and to become familiar with the factors that contribute to effective workplace learning. Chappell and Hawke (2005: 21) group these into four categories: the work environment, the context and objectives of the business, the nature of social interaction in the workplace and with external contacts, and managerial attitudes to learning.

The Learner's Perspective

Another important aspect in the consideration of how to define learning is the learner's own perspective. Many people, including employees and their managers equate learning with formal courses – activities which take place in a classroom or in a training center, and which are often associated with certification. As Ashton *et al.* (2005: 31) point out, this experience of learning as synonymous with formal courses is one that echoes the experience of policy-makers and academics. This further complicates the task of devising comprehensive recognition systems and selling the message that sometimes subliminal learning is a better option, for example, when encouraging reluctant learners or when the learning needs to be immediate. In those instances where there is a case for engaging in a recognition process, the policy settings and funding regimes need to be well established. Otherwise informal learning may not be seen as a credible alternative to formal training. For this reason, many OECD countries are exploring how best to recognize informal learning. And in the case of the European Commission, there is a concerted effort in vocational education and training domain to develop a set of common

principles for the validation of nonformal and informal learning. The European Qualifications Framework has adopted an unambiguous focus on outcomes, achieved by whatever means of learning. It is also explicit that the attainment of a qualification can be achieved through a combination of knowledge, skill, and wider personal or professional competence.

Teaching and Assessment

To implement such policies demands flexible teaching and sophisticated assessment practices, whose focus is less on the definitions of types of learning and more on the results of the learning and the capacity for effective action (Smith and Blake, 2005). This approach applies particularly to adults who already have experiences to build on and preferences for the way in which they learn. Some will want to engage in more self-directed learning rather than be guided by a teacher. Others will prefer a hands-on approach rather than learning through listening and reading.

Vocational trainers have to adapt their teaching to the diverse needs of the ageless learner. They must adapt or embrace learner-centric approaches, more traditional knowledge transfer, as well as consider how to facilitate effective workplace learning. They have many avenues open to them, including formal classroom instruction, group discussion, information and communications technology (ICT)-assisted learning, and workplace interactions. In most cases it will be a combination of these that will reap the best results. Assessors, too, have to meet the increasing demand for the recognition of skills and knowledge gained outside the formal education system (i.e., on the job or in the community). This type of assessment may be for accountability purposes or to determine a student's level of performance but can double as a tool to shape the teaching and learning process. This dual role reflects the reality that adults today must be able continuously to adapt to changing circumstances and therefore need to be able to learn new things all the time.

Measuring or assessing informal learning outcomes can pose significant problems. Informal learning is often *ad hoc*. Clarke (2004), who has looked at these issues, notes that because of the difficulties the focus of assessment shifts from measuring outcomes to assessing learning conditions or the opportunities for informal learning to take place. This diversion can mean that a significant issue is overlooked: the recognition of the quality of the learning taking place. It is important that we know people are gaining the right knowledge and are applying it in a safe and effective way. Attention to quality is imperative also if less formal modes of learning are to gain widespread acceptance from employers and assessors. That is why the European Union, in particular through Cedefop, the European agency for the promotion of vocational education

and training (VET), is developing a set of trusted principles for the validation of nonformal and informal learning. In 2007, Cedefop carried out a project on recognizing this learning among VET teachers and trainers. It found it was possible to make informal learning visible and to evaluate it, especially by adopting a portfolio approach. Technology is lending a hand here, allowing learners to demonstrate the richness of their experience, in pictures and film as well as the written word (see **Case Study 5**) (Cotterill, 2007).

Conclusion

This examination of formal and informal learning makes it clear that informal learning is beneficial, but the extent of that benefit is often determined by people's prior educational experience. Those with a sound foundation of formal schooling are likely to gain valuable insights and experience through informal learning on the job and in other extracurricular activities (reading, netsurfing, etc.). Moreover, they are probably more able to articulate such

learning and receive recognition for it, either through gaining promotion, formal qualifications, or receiving credentials. On the other hand, people with little primary education, whose literacy and numeracy may be limited, or whose confidence in their ability to learn in an educational setting has been dented, can find it difficult to articulate the life experience and skills they possess in a way that facilitates a pathway back to training. For them less formal approaches to skills development can be a more effective pathway back to education and training or to building sustainable jobs. This is the experience of much successful community learning, particularly in places where formal training options are limited.

The legitimate attempt to establish boundaries to assist in recognizing informal learning does, however, pose a conundrum: by making it explicit do we change the nature of this learning? And does it consequently lose some of its value? This is one of many questions that the discussion of the nature of learning stimulates. In this article we have put the case not only for the benefits that can accrue from informal learning but also from according it due recognition in those circumstances where this is advantageous to learners, educators, and employers.

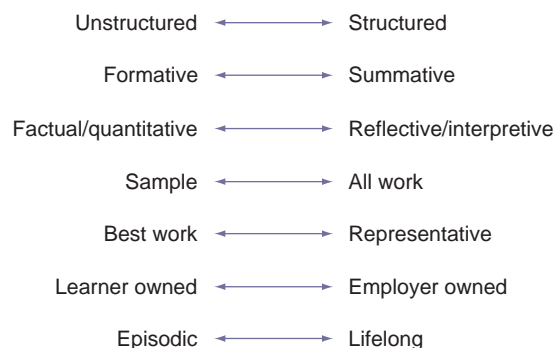
We do so in the hope that this will in turn contribute to a greater impetus to ensuring that all adults are able to engage in the learning they need to live well in the twenty-first century. That may mean being able to watch *The Simpsons* with an easy conscience, having recognized that we all need to keep on learning many things in many ways and knowing we have developed the capacity to sift the learning from the humor.

Case Study 5 Eportfolios: A definition

Simon Cotterill, Senior Research Associate, University of Newcastle upon Tyne, sees the ePortfolio as

a purposeful collection of information and digital artefacts that demonstrates... learning outcomes, skills or competencies. The process of producing an ePortfolio (writing, typing, recording etc.) usually requires the synthesis of ideas, reflection on achievements, self-awareness and forward planning; with the potential for educational, developmental or other benefits. Specific types of ePortfolios can be defined in part by their purpose (such as presentation, application, reflection, assessment and personal development planning), pedagogic design, level of structure (intrinsic or extrinsic), duration (episodic or life-long) and other factors.

Cotterill goes on to explain that within this definition there are many different dimensions, which he presents as follows:



Bibliography

- ABS (Australian Bureau of Statistics) (2007). Adult learning Australia 2006–07. <http://www.abs.gov.au/Ausstats/abs@.nsf/Latestproducts/4229.0Main%20Features22006-07?opendocument&tabname=Summary&prodno=4229.0&issue=2006-07&num=&view=> accessed 3 Dec 2009.
- Ashton, D., Sung, J., and Raddon, A. (2005). A case where size matters: A preliminary investigation into the institutionalisation of skill formation and firm size. *Research Paper No. 60*, Autumn 2005. Centre for Labour Market Studies, University of Leicester. <http://www.skope.ox.ac.uk/WorkingPapers/SKOPERP60.pdf> (accessed October 2009).
- Baran, J., Gilles, B., Richard, R., and Salmon, W. (2000). Applied research branch strategic policy. *Adult Education and Training in Canada: Key Knowledge Gaps*. Canada: Human Resources Development.
- Becket, D. and Hager, P. (2002). *Life, Work and Learning*. London.
- Billet, S. (2001). *Learning in the Workplace: Strategies for Effective Practice*. Sydney: Allen and Unwin.
- Boon, G. C. and Gopinathan, S. (2006). *The Development of Education in Singapore Since 1965 Background Paper Prepared for the Asia Education Study Tour for African Policy Makers*. June 18–30, 2006 National Institute of Education, Nanyang Technological University, Singapore, June 2006 Draft. http://siteresources.worldbank.org/EDUCATION/Resources/278200-1121703274255/1439264-1153425508901/Development_Edu_Singapore_draft.pdf (accessed October 2009).

- Cameron, R. (2004). *RPL and Mature Age Job Seekers*. Adult Learning Australia, Canberra. <http://www.ala.asn.au/public/docs/report/2004-11-CameronFullReport.pdf>.
- Chappell, C. and Hawke, G. (2005). *Investigating Learning and Work*, DRAFT 8 August 2005 – Consortium 2 website. <http://www.consortiumresearchprogram.net.au/html/content/view/22/37/> (accessed October 2009).
- Clarke, N. R. (2004). Organisational factors influencing the assessment of formal and informal learning. *Paper Presented at the Academy of Human Resource Development International Conference Full text from ERIC*. http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1d/85/6c.pdf (accessed October 2009).
- Colley, H., Hodkinson, P., and Malcolm, J. (2003). *Informality and Formality in Learning: A Report for the Learning and Skills Research Centre*. London: Lifelong Learning Institute, University of Leeds.
- Conlon, T. J. (2003). A review of informal learning literature, theory and implications for practice in developing global professional competence. *Journal of European Industrial Training* **28**(2/3/4), 283–295.
- Conner, M. L. (2002). Informal Learning. *Ageless Learner*, 1997–2007. www.marciacconner.com/intros/informal.html (accessed 4 Dec 2009).
- Cotterill, S. (2007). What is an eportfolio? eportfolios 2007. <http://www.eportfolios.ac.uk/definition> (accessed 4 Dec 2009).
- King, K. (1982). Formal, non-formal and informal learning: Some North–South Contrasts. *International Review of Education* **28**(2), 177–187.
- Livingstone, D. (2000). Exploring the icebergs of adult learning: Findings of the first Canadian survey of informal learning practices. <http://www.oise.utoronto.ca/depts/sese/csew/nall/res/10exploring.htm> (accessed October 2009).
- Nordman, C. and Hayward, G. (2006). *Returns to On-the-Job Training: Do Skill Usage, Tasks and Workstation Matter? Evidence from British Workers*. Cardiff: Oxford and Cardiff Universities SKOPE Research Paper.
- OECD (2005). The role of national qualifications systems in promoting lifelong learning. *Report from Thematic Group 2: Standards and Quality Assurance in Qualifications with Special Reference to the Recognition of Non-Formal and Informal Learning*. <http://www.oecd.org/dataoecd/57/27/34376318.pdf> (accessed October 2009).
- OECD (2007). Policy brief: Qualifications and lifelong learning. <http://www.oecd.org/dataoecd/10/2/38500491.pdf> (accessed October 2009).
- Parker, A. (2006). Lifelong learning to labour: apprenticeship, masculinity, and communities of practice *British Educational Research Journal*, 687–701.
- Richardson, S. (2004). *Employers Contribution to Training*. Adelaide: NCVET. <http://www.ncver.edu.au/publications/1506.html> (accessed 7 Dec 2009).
- Smith, P. and Blake, D. (2005). Facilitating learning through effective teaching: At a glance. Adelaide: NCVET. <http://www.ncver.edu.au/publications/1660.html> (accessed October 2009).
- Thacheen, P., and Lauzon (2006). Walking a new path of life: learning tours, agroforestry and the transformation of the village of Bann Na Isran, Thailand. *International Journal of Lifelong Education* **25**(4), 407–431.

Further Reading

- Gou, S. (2006). Mapping the iceberg of informal learning: Exploring the experience of Chinese volunteers in Vancouver. *Conference Paper, 2006 National Conference On-line Proceedings*, York University, Toronto, ON, May 1006.
- International Labor Organization (2003). *Lifelong Learning in Asia and the Pacific: Policies and Practices*, 2003. <http://www.ilo.org/public/english/employment/skills/new/2.htm> (accessed October 2009).
- Pak, M. S. (1997). *Two Wheels for Lifelong Learning in Korea: Credit Banking & Multimedia Technology*. CEDEFOP Discussion Paper.
- Parker, A. (2006). Lifelong learning to labour: Apprenticeship, masculinity and communities of practice. *British Educational Research Journal, Routledge* **32**(5), 687–701.
- Smith, M. K. (2002). Malcolm Knowles, informal adult education, self-direction and andragogy. *The Encyclopedia of Informal Education*. <http://www.infed.org/thinkers/et-knowl.htm> (accessed October 2009).
- The European Commission *erasmus 1987–2007*. http://ec.europa.eu/education/index_en.htm.

Relevant Websites

- <http://www.agelesslearner.com> – Informal Learning, Marcia Conner.
- <http://eportfolios.ac.uk/definition>
- <http://www.voced.edu.au/search/index.php?author=Thacheen>

School-Based Vocational Education and Training

B Knight, National Centre for Vocational Education Research, Adelaide, SA, Australia

R Sweet, University of Melbourne, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Setting the Scene and Defining the Scope

One of the major goals of schooling in practically every country is equipping young people with the knowledge and skills that they need for employment. Vocational education and training (VET) undertaken as part of secondary schooling or its equivalent is one of the major strategies that country education systems use to promote this goal. In many situations, it complements the skills, experience, and work ethics that young people gain from part-time, vacation, or voluntary work. However, the institutional arrangements for school-based VET, the form that it takes, the relationship to other parts of the curriculum, and the stage at which it occurs in a young person's education, all vary considerably among countries and even within countries.

The common thread that allows us to present an overview of school-based VET is that, unlike employment-based apprentices who are legally employees, the participants are students. They may have reached the postcompulsory stage of schooling or still be in the compulsory stage if one is mandated. It should be noted that for some disadvantaged groups and people with limiting disabilities, schooling that includes school-based VET can extend well into their adult years; the relevance of VET to these groups is discussed elsewhere in the encyclopedia.

In many countries, a school-based VET program is a coherent, complete course that leads to a recognized, full qualification, typically at the International Standard Classification of Education (ISCED) level 3 or its equivalent. However, many systems allow school students to undertake single vocational subjects, short courses, or modules as part of a general-education program. Additionally, countries such as Scotland and Australia are trying to use vocational education as a way of broadening the upper-secondary general education-curriculum, increasing participation in VET outside of vocational education institutions, and raising engagement with learning in postcompulsory education. By contrast, countries such as Sweden have intentionally taken an integrated view of general and vocational education in both the higher education and 3-year postcompulsory-schooling sectors and have sought to reduce the distinction between practical and theoretical studies (Cully *et al.*, 2009).

The focus of the discussion that follows is school-based VET as part of the broader set of entry-level education and training arrangements for young people. In the next two sections, we consider two dimensions of school-based

VET: the institutional arrangements that exist and the models that are found in education and training systems around the world. However, there is substantial overlap between these two dimensions. Additionally, much of the information and research evidence about school-based VET is embedded in more general treatments of upper-secondary education, with the two often not treated separately. We note also that the country examples given are just that and are far from complete.

Institutional Arrangements

The considerable international variation in institutional arrangements for providing VET to young people who are still legally students (rather than employees), including the inevitable variation in terminology, makes it necessary to take a broad view of the term school. From around grade 7 onward, countries have a variety of institutions – secondary schools, middle schools, *lycées* (in francophone countries), colleges, education or preparation centers, special-education centers, technical or vocational institutes, etc. – whose functions include providing school-based VET. We use the term school as a matter of convenience to encompass all of these institutions.

This means that for students, school-based VET may be provided through a number of different institutional possibilities, which we classify in four broad groups and illustrate with examples from a few countries:

1. Students attend the same institution for both their general education and VET, as in Finland, Sweden, and Brazil (where some schools offer professional – that is, vocational – training in agriculture starting in the early years of secondary school), and India, where some lower-level apprenticeships start at the beginning of secondary school. Included in this category are centers that provide both general education and school-based VET to young people with special needs, such as those in detention or with a serious long-term disability.

This common-institution arrangement has obvious limitations when the school-based VET requires access to equipment and resources that the institution does not have. Some countries overcome this by establishing special schools that provide both general education and school-based VET to prepare young people for a specific industry or occupation. Examples include defense and civil-service preparation centers in some

countries for students selected at a young age and schools run by state-owned business enterprises such as the national electricity authority in China.

2. The learning that a young person undertakes at school is supplemented by work observation, work experience, or structured learning activities in an actual workplace. In some systems (e.g., France and Australia), structured workplace learning may also be formally accredited and assessed. Many countries use workplace learning of some form to provide VET to school students. Sweden provides an example of a country where employers have shown a strong social commitment to this mode of school-based VET: their motivation may include wanting the opportunity to identify possible future employees in a system that has dispensed with apprenticeships.
3. Students attend one institution for their general education and another for their school-based VET, as in Australia when students do their school-based VET at a technical and further education (TAFE) institute concurrently with their general education at an ordinary secondary school. Such arrangements can give students access to much more appropriate teaching and learning resources for their VET studies but may also increase the total cost and cause considerable logistical difficulties. The arrangements can particularly be complicated in countries that allow young people who are still in school to start an indentured apprenticeship or traineeship that includes paid employment as part of the training program.
4. Students leave the institution at which they started their general education and attend one that provides both VET and general education, either explicitly or embedded within VET studies (this is discussed later under the section 'The nature of the curriculum'). Examples of institutions that offer this option include community colleges in Canada and USA, further education and training (FET) colleges in South Africa, *lycées professionnels* in France, the *collèges d'enseignement général et professionnel* in Quebec, and the *Technikum*, *Kolledz* and *Uchilisbe* in Russia.

From the institutional arrangements for school-based VET, we also observe that many countries have a variety of institutions that deliver school-based VET programs. Examples are secondary schools and further education colleges in the United Kingdom, and secondary schools and TAFE institutes in Australia. In Denmark, school-based VET is provided, among others, by commercial colleges, technical colleges, agricultural schools, and social and healthcare (SOSU) schools, all of which cohabit the school-based VET landscape.

Models of School-Based VET

Some dimensions that can be used to classify or describe school-based VET programs are presented below.

The Nature of the Curriculum

In general, a curriculum covers the goals or objectives of the program, what is to be learnt (the content), how it is taught (pedagogical aspects), how it is assessed, and what standards are applied. In reality, there is also a *de facto* curriculum, which is how it all operates in practice. The curricula for schooling generally cover a number of broad areas, including personal and social development, generic academic or vocational skills, and more specific academic or vocational knowledge and skills. The balance between these broad areas has been the subject of considerable debate for many decades, particularly between the proponents of a broad and balanced liberal education and those that favor a more instrumental view of education. In some contexts, curriculum content that focuses on culture, religion, or behavioral norms may also be included.

For school-based VET, the aspect that is particularly relevant is the extent to which the curriculum focuses only upon vocational content compared to that of general education, cultural studies, personal development, and the like.

At one extreme, there are countries where the curriculum has been imported from a wider VET system that is employer-led and is built around the competency standards developed for particular occupations or industries. This model is found particularly in Anglo-Saxon countries and is arguably too narrow and fragmented for young people who are still in schooling. Australia provides an instructive example. Young people who study full-time at a TAFE institute experience a curriculum that is almost completely vocational with very little general education or personal-development material. The institutional dimension is a factor here, as young people who do their school-based VET in a secondary school setting experience a somewhat broader curriculum. The difference arises because the philosophy and management of the school system are directed by government education ministries whereas TAFE institutes are part of Australia's public VET system, which adopts a largely instrumental approach to vocational learning.

At the other extreme, there are cases such as school-based VET programs in Sweden, Italy, and Japan where about half of the content is general education, personal development, or cultural studies (including sport, and religious studies in the Italian case), and Norway's where about a third is general. This more balanced approach is also found in the curriculum for school-based VET programs developed by the national ministry of education in China. In Norway, the aims cover a person's moral outlook, creative abilities, work understanding and skills, general education, understanding and cooperation in a democratic society, and ecological understanding, and they apply to all primary, secondary, vocational, and adult education (Royal Ministry of Education, Research and Church Affairs, 1997). In Singapore, the Institute of Technical Education

(ITE) is a major public provider of school-based VET and about 15% of the content of the courses is life skills; the curriculum is designed to ensure that the young person's general education continues and is integrated with the more specifically vocational aspects of the courses (its slogan, 'hands-on, minds-on, hearts-on,' reflects this; (Law, 2008).

The Role of Workplace Experience and Learning

We have noted that the learning a young person undertakes at school may be supplemented by work observation, work experience, or structured learning activities in an actual workplace. As with most aspects of school-based VET, this varies among countries.

On the one hand, there are countries such as Hungary (Kis *et al.*, 2008), Japan, and Korea (OECD, 2007a) in which there is almost no workplace contact and where workshop practice and simulations are heavily relied upon. Some countries also use this approach for entry-level training of apprentices. In China and India, for example, the whole or perhaps the first year or two of the apprenticeship program is spent in institution-based training using equipment and facilities that are part of the institution's infrastructure.

At the other extreme are Sweden's upper-secondary programs and the *baccalauréat professionnel* in France where the equivalent of a day a week in the workplace is required in all programs. This is regarded as curriculum time and formally assessed, with resources in place to make this possible.

Between these extremes are systems in which some accredited vocational subjects have a workplace learning requirement, sometimes very structured. Meeting the curriculum requirement for the work placement may be as simple as being there for the required time or it may involve a more demanding formal assessment of the prescribed learning outcomes. Where there is some form of assessment, this may be undertaken by an employer's representative, an assessor from the institution where the young person is enrolled, or even an external assessment authority.

Another instructive example is the social and healthcare schools in Denmark in which a workplace rotation model is used to give young people broad skills within a range of entry-level health and social occupations (OECD, 1998). This has some parallels with the model that occurs in Australia when group-training arrangements are applied to school-based apprentices, with apprentices shared among a group of employers for the on-the-job component of training.

The Level and Focus of School-Based VET Programs

Whether school-based VET programs are segmented by level is another important dimension. The level can also

determine whether students are streamed according to their ability into these programs, either formally or *de facto*; this aspect is closely related to the issues discussed elsewhere in the encyclopedia.

The Nordic countries, and South Korea and Japan to a large extent, provide typical examples of upper-secondary school-based VET programs that are not segmented by level. In Japan, the content is relatively nonspecific, and in the initial years of the programs, in Norway and Sweden, the content is quite broadly built around a family of occupations to encourage occupational exploration. Such approaches are much more difficult if the vocational content is tightly focused upon specific occupations, industries, or levels.

Elsewhere, most school-based VET programs appear to concentrate on a relatively tightly defined occupational or industry area. Moreover, it is common to treat vocational (lower level, generally with a focus on manual trade occupations) and technical (higher level, generally with a focus upon technician-level occupations) programs separately. Austria, Belgium, the Czech Republic, France, and Hungary are all examples of systems where school-based VET programs are segmented by level. School-based VET programs in India also span a wide range of levels, partly as a consequence of the wide range of occupations covered by crafts and apprenticeships in that country (DGET, 2007, 2008).

Country systems which allow several school-based VET models to coexist are better placed to overcome a narrow focus on levels and occupations if government policies allow this. In Australia, the curriculum frameworks could accommodate school-based VET programs at a range of levels, but in practice, the majority are at the relatively low certificate-II level if the student is still attending secondary school or certificate III (the level of apprenticeships in most traditional trades) if the student is undertaking school-based VET at a TAFE institute. A range of political factors accounts for this, plus *de facto* streaming by ability.

Articulation with Tertiary Education Programs

The extent to which school-based VET programs link to courses in tertiary education, particularly higher education, has a significant effect on the VET programs that are available to school students and on take-up rates. Using performance in vocational content as the full or partial basis of admission to higher education is uncommon in the Organization for Economic Cooperation and Development (OECD) countries, as the conceptual and intellectual level of the curriculum is not regarded as sufficient. Some exceptions can be found, usually where there is rigorous, quality-assured, and graded assessment of the VET program or where VET and higher education programs have been designed with articulation in mind (as in some dual-sector

VET/higher education institutions in Canada and Australia). In Singapore, the school-based VET programs at the ITE are at national certificate or higher national certificate level, but there are well-defined pathways to diploma- and even degree-level programs for learners with an acceptable grade-point average.

Strengthening pathways between vocational education (both apprenticeship and school based) and tertiary education has been a common theme in most countries for well over a decade. In the case of school-based VET programs, a number of models can be observed. One of the earliest in Europe is the so-called double-qualifying program found in Austria and Hungary, where young people leave school with both a qualification for work and a qualification for entry into higher education. The programs are normally longer than the standard upper-secondary VET program (4–5 years rather than 3), focus upon higher-level technician occupations, and base the case for entry to in higher education upon the demanding level of the vocational content of the entire program rather than, as in other models, upon performance in general-education subjects that form only part of the school-based program.

In Sweden, there has long been a general provision for all graduates of any upper-secondary school-based VET program to be eligible in principle for higher education entry if they meet specified standards, and in some cases, if they complete supplementary work. However, the decision is based only upon their performance in the general-education subjects within the school-based program. The same type of arrangement can be found in Norway's upper-secondary programs, although there, an arrangement exists to provide supplementary instruction in general subjects. Alternatively, young people can transfer from the vocational to the general track midway through their program. In Finland, 1999 legislation gives all graduates of 3-year upper-secondary vocational programs an entitlement to enter all forms of higher education.

In other countries, pathways between upper-secondary school-based VET programs and higher education are minimal. Where pathways do exist, an issue is whether the pathway is to all forms of higher education, as in Sweden and Norway, or is generally to institutions other than universities, such as the *Fachhochschulen* in Austria and Switzerland.

A related issue is the case of countries such as Canada in which school-based vocational preparation is largely deferred until tertiary education, in this case through the short-cycle programs in community colleges (OECD, 2004), or countries where a substantial number of such programs are found in 2-year institutions as well as in upper-secondary schools: South Korea and China are prime examples. Even though the main focus of this article is on programs that are classified as ISCED 3, these school-based VET programs located within tertiary education deserve mention as they seem to be little different from equivalent programs that

are located in upper-secondary education in other countries. The recent OECD review of tertiary education in China (Gallagher *et al.*, 2009) estimated that somewhere between a fifth and a half of all participation in tertiary education was in fact vocational education at upper-secondary level. The position in South Korea appears to be very similar; in recent years, a seemingly wholesale transfer of programs from upper secondary to tertiary level has occurred. Examples include 4-year degree-level courses in hairdressing, and a 2-year associate-degree program in fields such as party planning.

The Relative Weight of School-Based VET in the Initial Vocational-Preparation System

There are two variables that matter here. The first is the overall size of a country's initial vocational-preparation system; the second is whether the initial vocational-education system consists mostly of apprenticeship programs, a mix of apprenticeship and school-based programs, or mostly school-based programs. In **Table 1**, we endeavor to classify the 30 OECD countries using these two dimensions.

The mixed-model countries such as Norway, Turkey, Austria, and Denmark are particularly revealing. In some cases, there seem to be clear distinctions between the types and levels of qualifications linked to the school-based and apprenticeship programs. For example, in Austria, commercial, clerical, and retail programs are largely delivered using school-based programs; in Denmark, the SOSU schools have a clear separation from the apprenticeship system; and most of Norway's school-based programs are in areas in which apprenticeship does not usually operate. In Turkey, the distinction between apprenticeship and school-based programs does not seem to be very clear in terms of the occupational areas, with much of the difference being a function of the level of education of entrants. In many other Middle Eastern countries with a mixed-mode system, the distinction appears largely in terms of the level of the qualification, with school-based programs generally at a higher level than the work-based programs (Sweet, 2009).

There are also differences between countries in the extent to which school-based initial vocational preparation is growing or declining. **Table 2** shows trends over the 1998–2006 period in the proportion of upper-secondary enrolments in vocational programs for those OECD countries whose vocational preparation system is largely school based.

These programs are fairly stable in size in most countries, but growing in Finland, in Sweden, and in Spain:

- in Finland this is probably because of the new pathway that has been opened up between upper-secondary vocational education and tertiary education.
- in Sweden, the likely cause is the creation in the late 1990s of *kvalificerad yrkesutbildning* (KY), a new type of

Table 1 Initial vocational preparation as a share of upper-secondary enrolments

		<i>Initial vocational preparation as a share of upper-secondary enrolments</i>		
		<i>Large (>50%)</i>	<i>Medium (25–50%)</i>	<i>Small (<25%)</i>
Dominant model(s)	Apprenticeship	Germany		
		Switzerland		
	Mixed	Austria	Denmark	Australia
		France	Ireland	New Zealand?
		Netherlands	Turkey	
		Norway	United Kingdom?	
	School-based	Belgium	Greece	Canada
		Czech Republic	Iceland	Hungary
		Finland	Korea	Japan
		Italy	Poland	Mexico
		Luxembourg	Portugal	United States
		Slovak Republic	Spain	
		Sweden		

Table 2 Proportion (%) of upper-secondary enrollments in vocational programs for selected OECD countries, 1998–2006

<i>Country Year</i>	<i>1998</i>	<i>1999</i>	<i>2001</i>	<i>2002</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>
Belgium	69.0	65.7	69.2	69.7	70.3	68.2	69.6	69.4
Canada	11.2	8.2	15.2	m	m	m	m	5.4
Czech Republic	80.0	80.2	80.7	80.4	79.5	79.4	79.5	79.3
Finland	52.0	53.2	56.7	57.2	58.8	60.1	63.9	65.4
Greece	32.6	25.8	35.2	40.0	36.0	34.0	36.0	33.9
Hungary	67.4	65.5	50.2	49.7	49.8	23.7	24.1	23.7
Iceland	33.2	32.8	36.4	38.3	35.1	38.5	36.8	36.7
Italy	64.8	64.7	64.3	64.8	63.8	62.8	61.5	60.5
Japan	26.8	26.4	25.9	25.7	25.5	24.6	24.7	24.6
Korea	40.0	37.9	34.1	32.1	30.7	29.5	28.5	27.8
Luxembourg	68.0	63.7	63.8	64.0	64.7	63.9	63.4	62.9
Mexico	14.4	14.0	12.2	11.4	10.9	10.5	10.2	9.8
Poland	67.6	66.1	62.1	60.9	54.3	49.5	45.0	44.0
Portugal	25.4	25.0	28.3	28.8	28.5	28.5	31.0	31.5
Slovak Republic	m	79.6	77.6	76.4	75.4	74.1	74.2	73.7
Spain	21.5	31.2	35.6	38.0	37.2	38.7	42.6	42.5
Sweden	40.6	50.1	51.7	49.6	52.9	53.4	53.6	55.1

From OECD, *Education at a Glance*, successive issues.

short-cycle tertiary/postsecondary education program outside of the universities, which as in Finland, has helped to open up new pathways between upper-secondary vocational education and tertiary study.

On the other hand the rates are declining in Hungary, South Korea, and in Poland:

- in the case of Hungary and Poland, the rates were kept artificially high during the communist period, and the loosening of central controls over decisions about resource allocation between different types of schools, combined with high social demand for and recent expansion of higher education, seems to be the explanation.
- in South Korea, the story is much the same, with a loosening of controls over the expansion of higher education institutions, particularly in the private sector, great emphasis on higher education, and a transfer of

many upper-secondary level courses to higher education (not only 2-year programs but also to 4-year institutions) accounting for the changes.

In most of the above examples, both growth and decline, the relationship between tertiary education and upper-secondary vocational education seems to be behind the trend, rather than the changing demand for the skills and qualifications produced by upper-secondary vocational education.

Outcomes and Purposes

The discussion of outcomes from school-based VET programs is closely related to their purposes. In essence, there are two sets of arguments: those about economic

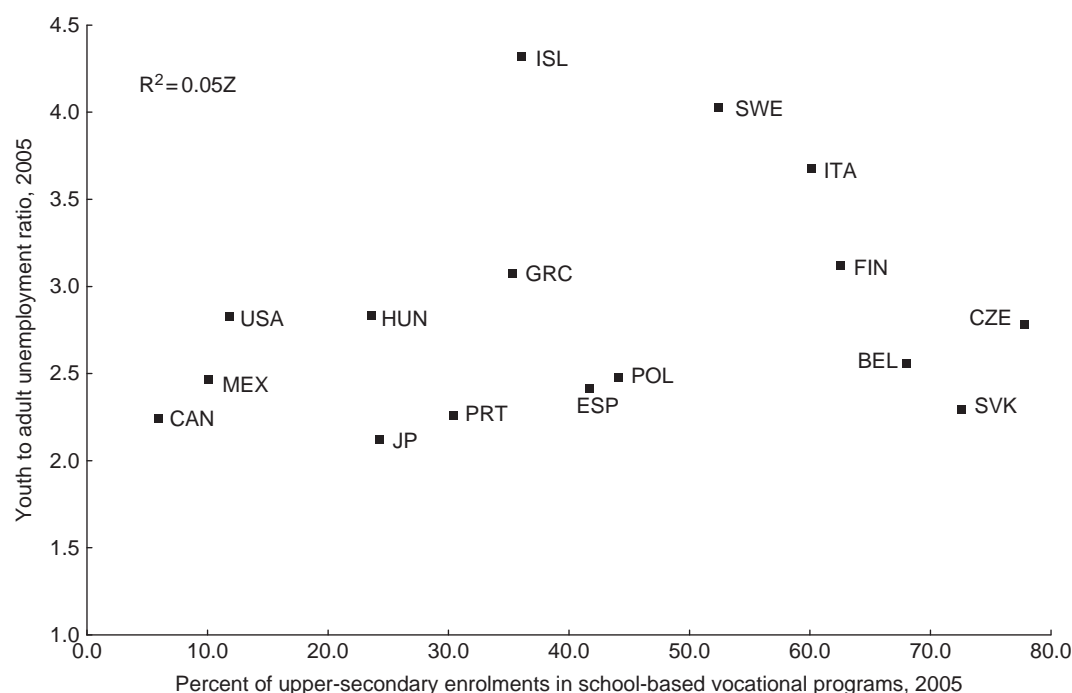


Figure 1 The size of upper-secondary school-based vocational pathways and young people's labor-market competitiveness. From OECD labor force database and *Education at a Glance*, 2007.

purposes and outcomes; and arguments related to participation, pedagogy, and engagement with learning.

Evidence on the economic or skills-based outcomes and purposes of school-based programs are closely linked to the literature on labor market and economic returns from different types of postcompulsory pathways. The impact of pathways is linked to the quality of the students who enter them as well as the labor-market returns of the relevant occupations. Ryan's (1998) paper on apprenticeship returns suggests that there is not much basis for favoring apprenticeship over full-time vocational schooling, although there are some country-specific exceptions. Grubb's (1997) work on sub-baccalaureate returns to vocational education in the USA shows that returns are highly specific to the occupational field. Raffe (2003, 2008) makes similar comments in his work on pathways.

For a number of countries whose upper-secondary vocational pathways are largely school-based, **Figures 1** and **2** show the relationship between the scale of vocational enrollments and two measures of school-to-work transition outcomes: the youth-to-adult unemployment ratio as an indicator of youth's relative labor-market competitiveness; and the number of years expected to be spent inactive (either unemployed or not in the labor market) between the ages of 15 and 29 as an indicator of the smoothness of labor-market insertion. In neither case is there any significant relationship to the scale of upper-secondary vocational programs, suggesting that arguments based upon economic or labor-market purposes need to be treated with some caution as a justification for expanding such programs.

Other arguments about outcomes and benefits focus on the role of school-based VET in raising participation, pedagogy, and engagement with learning. The OECD (2007b) uses such arguments to support the case for VET in schools programs in Scotland. In brief, more active and applied teaching and learning styles are seen as ways to keep young people engaged in education and to build better foundations for lifelong learning. The emphasis upon these sorts of issues seems to be somewhat different in countries such as Scotland and Australia which are trying to raise upper-secondary participation and completion within traditional schooling.

Australia provides possibilities for a case study based on recent experience because the school-based VET options available within the senior secondary certificate were expanded significantly about a decade ago and there has been a significant level of take-up (Polesel, 2001). However, there is little evidence that it has improved retention rates to the end of secondary school and the evidence for improved labor-market outcomes for young people is patchy. School-based apprenticeships were introduced at much the same time, although the take-up has not been great; again, the evidence for positive benefits is patchy (Knight, 2008).

Access and Equity

Evidence for the effect, if any, that school-based VET has on access and equity for young people is difficult to

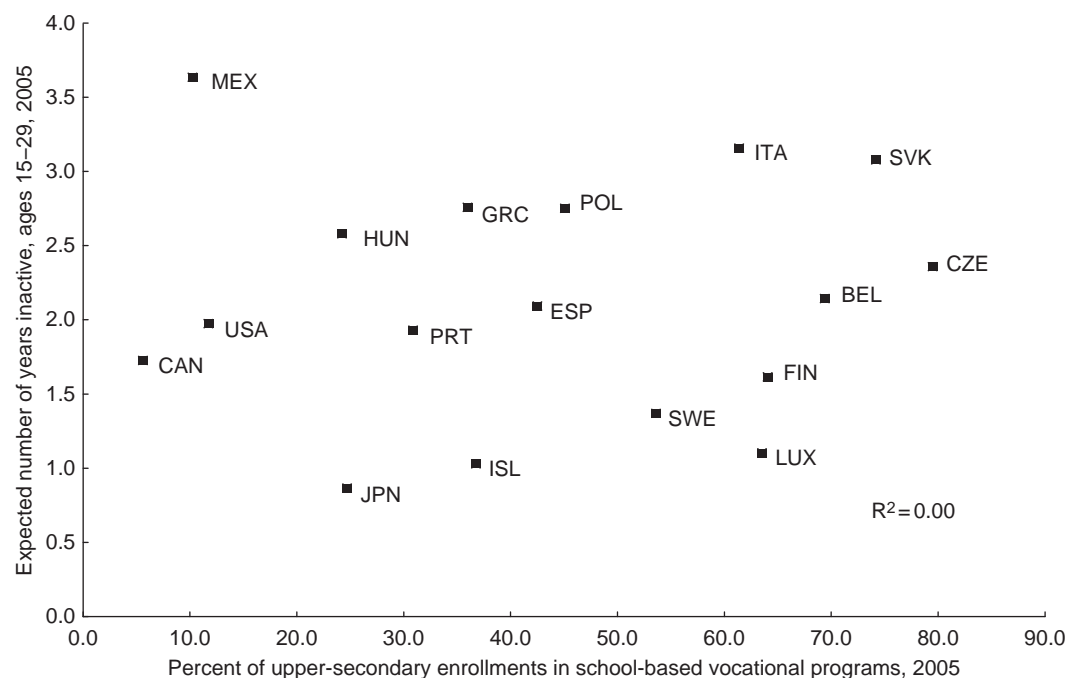


Figure 2 The size of upper-secondary school-based vocational pathways and the smoothness of young people's labor-market insertion. From OECD *Education at a Glance* 2007.

extract from the more general literature on the relationship between access and equity and upper-secondary pathways. As with the question of outcomes, it is hard to disentangle this from issues of the differential inputs to and economic returns from the different pathways. It is only in very specific situations, such as a special school or training center established to cater to a particular disadvantaged group, that a clear link can be observed. Even in these situations, there is little information publicly available about the benefits that might flow from the arrangement apart from the fact that a school-based VET option is available, which is certainly not the case in all countries.

What is clear is that in all countries in which this type of data has been gathered, those young people who enter (school-based) upper-secondary vocational pathways have on average a lower socioeconomic level and lower achievement levels than those who enter general tracks. The OECD's Program for International Student Assessment (PISA) data (OECD, 2008) make this abundantly clear, even if they are confined to 15-year-olds. Furthermore, it clearly shows that where countries segment these vocational-education programs by level, lower-level programs on average attract more of those from lower socioeconomic backgrounds and more with lower-level achievement scores than do the higher-level programs. European research (Müller and Shavit, 1998; Smyth *et al.*, 2001; Müller *et al.*, 2002) leads to the same conclusion.

Less clear are the conclusions about access and equity that should be drawn from this evidence of streaming by socioeconomic status and ability. For example, Müller and

Shavit (1998) point out that this type of tracking can reduce the risk of dropping out of education among such groups. Similarly, in the United States, empirical evidence strongly favors the view that the disproportionate representation of those from lower socioeconomic status backgrounds in community colleges is more likely to be associated with an opening up of educational opportunities with a decrease in the numbers going to university (OECD, 2004). A similar argument is presented by Polesel about upper-secondary vocational education programs in Italy.

Bibliography

- Cully, M., Knight, B., Loveder, P., *et al.* (2009). *Governance and Architecture of Australia's VET Sector: Country Comparisons*. Sydney: Skills Australia.
- Directorate General of Employment and Training (DGET). (2007). *Apprenticeship Training Scheme*. New Delhi: Ministry of Labour and Employment (MLE).
- DGET (2008). *Craftsmen Training Scheme in India*. New Delhi: MLE.
- Gallagher, M., *et al.* (2009). *OECD Reviews of Tertiary Education: China*. Paris.
- Grubb, W. N. (1997). The returns to education in the sub-baccalaureate labor market, 1984–1990'. *Economics of Education Review* 16(3), 231–246.
- Grubb, W. N., Sweet, R., Gallagher, M., and Tuomi, O. (2006). *OECD Thematic Review of Tertiary Education. Country Note: Korea*. Paris.
- Kis, V., Ferreira, M., Field, S., and Zwick, T. (2008). *Learning for Jobs. OECD Reviews of Vocational Education and Training: Hungary*. Paris.
- Knight, B. (2008). Vocational education and training in schools and school-based apprenticeships and traineeships 10 years on. In *VOCAL*, Vol 7, pp 130–142. VETnetwork, Australia.

- Kuczera, M., Field, S., Hoffman, N., and Wolter, S. (2008). *Learning for Jobs. OECD Reviews of Vocational Education and Training: Sweden*. Paris.
- Lamb, S. (2008). Alternative pathways to high school graduation: An international comparison. *California Dropout Research Project Report, No. 7*. Santa Barbara: University of California.
- Law, S. (2008). Vocational technical education and economic development – The Singapore experience. In Lee, S., Goh, C., Fredriksen, B., and Tan, J. (eds.) *Toward a Better Future: Education and Training for Economic Development in Singapore Since 1965*, pp 114–134. Washington D.C: The World Bank.
- Müller, W., et al. (2002). *Indicators on School-to-Work Transitions in Europe. Evaluation and Analyses of the LFS 2000 Ad Hoc Module Data on School-to-Work Transitions: Indicator Report*. Mannheim: MZES. http://www.mzes.uni-mannheim.de/publications/papers/LFS_indicator_report.pdf.
- Müller, W. and Shavit, Y. (1998). The institutional embeddedness of the stratification process: A comparative study of qualifications and occupations in thirteen countries. In Shavit, Y. and Müller, W. (eds.) *From School to Work: A Comparative Study of Educational Qualifications and Occupational Destinations*, pp 1–48. Oxford: Clarendon Press.
- Organisation for Economic Cooperation and Development (OECD). (1998). *Thematic Review of the Transition from School to Work. Country Note: Denmark*. Paris.
- OECD (2004). *Alternatives to universities revisited, Education Policy Analysis*, pp 15–46. Paris.
- OECD (2007a). *Jobs for Youth: Korea*. Paris.
- OECD (2007b). *Reviews of National Policies for Education: Quality and Equity of Schooling in Scotland*. Paris.
- OECD (2008). *VET in PISA: Results from PISA 2003 and 2006*. EDU/EDPC/CERI(2008) 5. Paris.
- Polesel, J. (2006). Reform and reaction: creating new education and training structures in Italy. *Comparative Education* 42(4), 549–562.
- Polesel, J. (2001). Vocational education and training in schools in Victoria: An appraisal six years down the track. *Journal of Vocational Education & Training* 53(2), 325–340.
- Raffe, D. (2003). Pathways linking education and work: A review of concepts, research and policy debates. *Journal of Youth Studies* 6(1), 3–19.
- Raffe, D. (2008). The concept of transition system. *Journal of Education and Work* 24(4), 277–296.
- Royal Ministry of Education, Research and Church Affairs (1997). *Core Curriculum for Primary, Secondary and Adult Education in Norway*. Oslo: Norwegian Board of Education.
- Ryan, P. (1998). Is apprenticeship better? A review of the economic evidence. *Journal of Vocational Education and Training* 50(2), 289–325.
- Smyth, E., Gangl, M., Raffe, D., Hannan, D., and McCoy, S. (2001). *A Comparative Analysis of Transitions from Education to Work in Europe (CATEWE): Final Report*. Dublin: ESRI. http://www.mzes.uni-mannheim.de/projekte/catewe/publ/publ_e.html (accessed October 2009).
- Sweet, R. (2009). *Work-Based Learning Programs for Young People in the Mediterranean Region*. Turin: European Training Foundation.

Further Reading

- Dore, R. and Sako, M. (1998). *How the Japanese Learn to Work*, 2nd edn. New York: Routledge.

Relevant Website

- <http://www.ite.edu.sg> and links – Institute of Technical Education, Singapore.

The Economics of Vocational Education and Training

P Carneiro, University College London, London, UK; Institute for Fiscal Studies, London, UK

L Dearden, University of London, London, UK; Institute for Fiscal Studies, London, UK

A Vignoles, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The dominant paradigm in the economics of education is Human Capital Theory, which suggests that education and training are investments that make individuals genuinely more productive. Individuals who are more productive will, according to this theory, also have higher earnings and be more employable. The private economic return to investing in education or training, that is, the gain to the individuals, can therefore be measured by the net gain in lifetime earnings accruing as a result of their investment in education or training.

Policy-makers around the world tend to accept unquestioningly the premise that investment in education and training is a good thing, with most committed to investment in human capital, including Vocational Education and Training (VET), as a means of securing higher economic growth and national prosperity as well as achieving equity goals (see Wößmann, 2008). However, some economists have argued that individuals who are more able and productive also tend to invest in more education (Spence, 1973; Arrow, 1973). In other words, education does not necessarily make you more productive it simply acts as an expensive sorting device, to enable employers to identify more able individuals. Wolf (2002) claims that the policy emphasis on education and skills as the main driver of economic growth overstates the importance of human capital investments and that a major function of education is as a sorting device and to some extent a social discriminator. In making policy about education and training, it is clearly important to understand these potential economic returns to education and training to the individual, firms, and the wider economy. However we must also recognize that separating out the genuine impact of education on productivity and earnings from its role as a sorting device is very difficult.

Understanding the economic value of education and training, as measured by gains in wages, is important but not in itself enough to inform policy about the optimal type of education and training system. As recent papers such as Carneiro and Heckman (2003), Cunha and Heckman (2009), and Carneiro (2009) argue, it is also crucial that we understand the technology of skill formation over the life cycle. For effective policy formation, we need to know how different types of people can effectively acquire skills at different stages of the life cycle as well as

the economic returns to those skills once acquired. This is particularly important in the area of VET, since international evidence suggests that the economic returns to some VET are low.

Finally, even if one accepts that education and training genuinely do enhance individuals' productivity in the labor market, it is not clear that focusing on earnings is enough. First, it is not clear that wages always accurately reflect individuals' productivity levels, as suggested by economic theory. This is clearly not the case in the public sector, as the wages of nurses and teachers, for example, are state determined (although influenced by prevailing private sector wages). Of course, education is also far more than an economic investment and yields a myriad of noneconomic benefits for the individual. These benefits may be in the form of reduced crime, better parenting skills, and/or better health outcomes. Noneconomic benefits from education are not considered in conventional rates of return analyses. However, looking at these potential wider benefits of education and training is also crucial for sound policy development – particularly in the area of VET.

In the next section, we look at the economic literature on how skills are formed and the implications this has for VET policy. In the subsequent section, we look at the international evidence on the returns to VET. In the last section, we conclude.

The Technology of Skill Formation and Its Implication for VET

The system of vocational education and training and enrolment rates in VET differs markedly by country. This has already been documented elsewhere in this encyclopedia.

But how effective is VET in raising the skill levels of individuals, how does this vary by country, and what are the implications for policy? We argue that in order to look at this question properly from an economic perspective, VET has to be seen as part of a whole lifecycle process of skill formation which starts in the womb and ends in the grave.

The extent to which VET and other education and training raises individuals' skill levels can be measured by estimating the economic return to such qualifications and skills. The work on skills formation shows that these economic returns to skills investment are higher in early

childhood and decline over the life cycle. From an analysis of both theory and evidence, Carneiro and Heckman (2003) argue that, at the current levels of investment, rates of returns to investment in human capital are declining over the life cycle (**Figure 1** taken from Heckman, 2008).

Numerous studies suggest that there should be strong investments in early childhood, both because the sensitive periods for acquiring several capabilities occur early in life, and also because successful learning early in life is the foundation for successful learning later in life (Carneiro and Heckman, 2003; Cunha and Heckman, 2009; Carneiro, 2009; Meghir and Palme, 2005).

However, this literature also argues that it is not enough to just invest in skill formation (both cognitive and noncognitive) in early life. These investments also need to be followed up by further investment in skills, or there is a strong risk that individuals' will not fulfill their potential. This is particularly so for individuals from poorer socioeconomic backgrounds who not only tend to start with lower levels of investment but then also have access to poorer quality schools and thus experience less subsequent investment.

The stock of human capital available to a young adult as he or she enters the labor market and beyond is therefore a function of the history of skill investments that took place up to that moment. This function is much more complex than just the sum of all the investments up to that point, because the timing of investments over the life course is not irrelevant for the final result. One key point emerging from this literature is that if the degree of complementarity between early and late investments is very strong, then it may be very costly or impossible to remediate the lack of early investments.

Second, this body of literature looks at the plasticity of learning (i.e., the ease with which a person's skills can change and develop) and the critical periods for learning. There is increasing evidence that there are sensitive periods of learning, which imply that if the proper investment is not done at the right time (during the sensitive period), then learning opportunities may be permanently lost. This literature suggests that it is very hard to build on cognitive skills if they have not been acquired in the first 10 or so years of life. On the other hand, the same evidence base suggests there is much more scope for enhancing noncognitive skills into adulthood. So there are critical periods of learning. In case they are missed, one does not get another chance.

These two issues have strong implications for the effectiveness of VET provision. Different VET systems will provide a different mix of cognitive and noncognitive skills and VET specialization will occur at different ages in different countries (some countries select children into VET at an early age). In general, the extent to which VET provision can build on prior investments will be critical to the economic value of VET. Any policy initiative in the VET area needs to take this into account this dynamic complementarity of skill investment as well as the plasticity of skill acquisition.

A competing model would say that investments in skill are substitutes over time, and therefore their timing is irrelevant, so all that matters is the sum of all investments one receives during one's life. If that were the case, then education and training could take place at any point over the life course and still prove effective in boosting the skills and earnings of individuals, even those who left initial education with very little in the

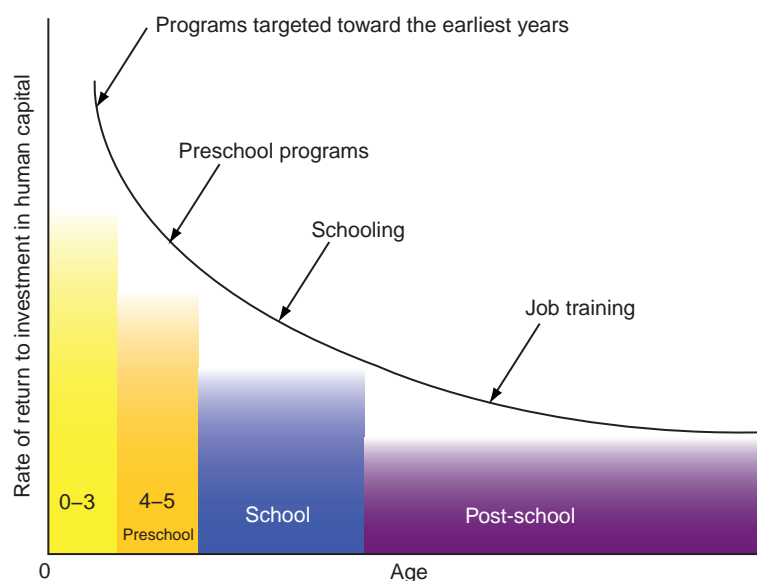


Figure 1 Returns to investment in human capital over the life cycle. Source: Heckman (2008).

way of skill. Our reading of the evidence is that this is not the case and that later investments that are not built on strong initial investments are likely to prove ineffective. Certainly, firms believe this to be so. It is also well known that more educated workers are more likely to be selected for (or select themselves for) company training, suggesting that firm investments in VET are targeted at individuals with higher levels of initial human capital (Blundell *et al.* (1996) for the UK and Carneiro and Heckman (2003) for the USA).

This dynamic complementarity not only means that the returns to later investment in skill are highest for those with strong early investments, but it also means that the returns to early investments in skill will increase with follow-up investments later in life. This is discussed more in the next section – but is important to bear in mind when looking at the rates of return to VET.

Dynamic complementarity is also a potential explanation as to why borrowing constraints in late adolescence play only such a small role in college attendance relative to factors such as family background and ability (Carneiro and Heckman, 2002), and why, as we will see in the next section, the returns to basic skills programs, that is, programs that attempt to improve individuals literacy and numeracy skills (see Torgeson *et al.* (2004) for a summary) and public job training programs, are generally too low across the world, as shown here (e.g., Lalonde, 1995; Cohn and Addison, 1998; Heckman *et al.*, 1999; Carneiro and Heckman, 2003; Dearden *et al.*, 2004; Abramovsky *et al.*, 2005).

International Evidence on the Returns to VET

The wage returns to general education have consistently been found to be strong and significant worldwide. It is widely agreed that the economic return from an extra year of formal full-time education is around 8–10% or even higher for some individuals (see, e.g., the review by Card (1999)).

In investigations that examine the returns to qualifications – rather than years of education – again most studies find significant economic returns. However, it does matter what type of qualification one acquires. In many countries, the wage returns to academic qualifications are significantly higher than the returns to vocational qualifications, government training programs, and adult basic skills training (see, e.g., Blundell *et al.*, 2005; Dearden *et al.*, 2002; Dickerson, 2005; Carneiro and Heckman, 2003).

What are the returns to VET? Why are the returns to some forms of VET so much lower than those for general education? Why do policymakers and governments around the world still invest significant amounts of money in government-sponsored VET programs when the returns are not high?

To address these questions, we start by looking at the international evidence on the returns to basic skills programs, vocational education, government training initiatives, and work-based training and then discuss the rationale for large state-sponsored VET programs.

Basic Skills Training

The value of cognitive skills in the labor market is high in most labor markets, and some studies suggest that there has been an increase in the return to cognitive skill over time, particularly in the USA and the UK (see Tyler (2004a) for an overview; Howell and Wolff, 1991; Murnane *et al.*, 1995, 2000; Bowles *et al.*, 2001). The value of basic skills is also high in some countries, although the evidence base is more limited (see Grinyer (2005) for a summary). In countries such as the USA, Northern Ireland, and the UK, the rate of return to basic literacy and numeracy is substantial (Denny *et al.*, 2003; McIntosh and Vignoles, 2001; Bynner *et al.*, 2001; De Coulon *et al.*, 2007; Tyler, 2004a). However, the returns to basic literacy and numeracy are not high in all countries. In particular, in countries where the supply of basic skills is relatively large compared to demand (e.g., Germany), the rate of return to such skills is much lower (Denny *et al.*, 2003).

Although basic skills have high value in the labor market, it is not the case that the returns to basic skills programs are substantial. Again the evidence base is somewhat limited. A recent international survey by Torgeson *et al.* (2004) suggests first that there is limited robust evidence on the impact of basic skills training. Second, those basic skill interventions that were robustly evaluated (e.g., by Random Control Trial) did not show systematically positive effects. Other UK evidence is consistent with this view: a study using rich cohort data to determine the economic return to basic skills courses and improvements in basic skill in adulthood found limited effects, for example (Dearden *et al.*, 2001). Given that most interventions tend to be small scale and in adulthood, this is perhaps unsurprising. As we discussed in the previous section, later human capital investments need to build on prior investments in the early years, particularly for cognitive skill development. Thus, trying to train illiterate adults is going to be challenging.

This evidence on the effectiveness of basic skills training, or lack thereof, contrasts sharply with the evidence on early interventions that have been found to be successful in improving cognitive skills and basic literacy and numeracy. For example, in the UK context, the *Sure Start* preschool intervention and a primary school intervention called *The Literacy Hour* have been found to be effective in improving children's cognitive and noncognitive skills (NESS, 2008; Machin and McNally, 2004). In the USA, preschool interventions, such as *The Perry Pre-School* random trial which provided 2 years of preschool support for

poor black children in the USA, have had a significant positive impact on children's skills (Nores *et al.*, 2005; Heckman *et al.*, 2008). Thus, it is certainly not the case that one cannot improve individuals' literacy skills, rather that interventions late in adulthood are more difficult.

Vocational Education at the End of Compulsory Schooling

The vocational options available to students at the end of post-compulsory schooling vary hugely across different countries. Some countries specialize in VET options early in middle secondary school (e.g., Germany), while others essentially force students to follow a largely academic curriculum until much later (e.g., the USA). Some countries have been shifting the balance over time, introducing vocational qualifications in secondary schooling that have notional parity with their academic equivalencies (e.g., the UK). Just as the provision of vocational education varies hugely by country, so does the economic value of VET as well. Again, differences in the value of vocational training are linked to the dynamic complementarity argument and the extent to which efficiency and equity considerations are complements, substitutes, or unrelated.

In countries where there are well-developed vocational education systems in schools and well-established apprenticeships systems, returns to vocational education are high (e.g., Steedman, 1993; Acemoglu, 2001; Acemoglu and Pischke, 1999) (or indeed where there is a competitive market for apprentices (e.g., Heckman, 2000). This is not, however, universally the case. In Australia, for example, the returns to VET are positive but vary by qualification level and mode of study (Ryan, 2002).

In countries where there is a less-developed vocational system, the proliferation of vocational qualifications both in the schooling system and beyond has weakened the signal of what the vocational education is providing and returns are less (Wößmann, 2008; Machin and Vignoles, 2005). This is exemplified by the UK experience where vocational specialization generally occurs at the end of compulsory schooling (this is changing somewhat with the introduction of vocationally oriented curricula at 14). Since the 1970s in the UK, there has been a proliferation of vocational qualifications and a dramatic shift away from apprenticeships (the current UK government has tried to reinvigorate the apprenticeship system with some success). Many newer vocational qualifications (e.g., National Vocational Qualifications) have little economic value. Jenkins *et al.* (2007) found, in line with much previous research (Dearden *et al.*, 2004), negative average wage returns to these low-level vocational qualifications and specifically NVQ Level 2 (NVQ2 – supposedly equivalent to the age 16 qualifications acquired in England, namely General Certificates of Secondary Education or GCSEs). Similarly, in the US context, the returns to academic

education are higher (see Kane and Rouse, 1995; Card, 1999). The returns to community college, which tends to be vocationally oriented, are still positive (Kane and Rouse, 1995). However, for those who drop out of high school prior to graduation, the route to obtaining later higher school certification (GED) has been found to have only limited economic returns, and in some cases negative returns (Heckman *et al.*, 2000; Murnane and Willett, 2004; Tyler, 2004b).

Thus, the institutional arrangements for VET provision in the secondary schooling phase and beyond heavily determine the likely outcomes for students who do not proceed down the academic route.

Government Sponsored Adult Training

Public sector job training programs that attempt to provide skills to a population of very low-skilled adults have almost universally been unsuccessful, but despite this such programs remain on many countries' agenda – largely because of equity considerations. Again this comes back to the dynamic-complementarities argument. Such job programs are largely targeted at low-skilled unemployed individuals, and these people accumulated little human capital earlier in their life and therefore they lack strong foundations on which public training can then build on. As a result, they are not able to learn much from the training that is offered to them.

A job program entitled Train to Gain in the UK is classic example of this problem. This program provides subsidized work-related training for low-skilled workers with the subsidy going to firms to encourage them to train. It was introduced on a pilot basis and was fully evaluated in 2003 and 2004. The evaluation suggested only small insignificant effects from the program on the incidence of training being undertaken by firms (Abramovsky *et al.*, 2005). Furthermore, they found that majority of firms undertaking this training would have done so in the absence of the program, that is, the deadweight loss was considerable. Additionally, the training provided in this program at best leads to qualifications which have been found to have extremely low or even nil wage returns. Specifically, the training was designed to lead to lower-level and newer vocational qualifications that have little labor market value (Dearden *et al.*, 2004b; Dickerson, 2005; McIntosh, 2004). Thus, this particular program had the twin challenge of being aimed at adults and also being designed to lead to qualifications that historically have had low labor market value.

Public training programs are not, however, ineffective purely because they are largely targeted at adults. Youth training programs, aimed at unemployed young adults, have been similarly ineffective (e.g., see Fougere (2000) for an analysis of French youth training and Friedlander *et al.* (1997)). Again this may be partially due the modest

amount of training being provided and its low quality. However, it is also the case that such programs are aimed at young adults who lack a good human capital base. They tend to have nil or few qualifications and low levels of cognitive skill. Building on such a weak foundation is undoubtedly more difficult.

Employer-Provided Work-Related Training

There is clear evidence that work-related training, provided by or at least purchased by firms, yields strong positive economic returns across a range of different countries (Almeida and Carneiro, 2009; Bassanini *et al.*, 2005; Blundell *et al.*, 1999; Groot, 1999; Leuven, 2005). This stands in contrast to the evidence on public training programs which have been found to be generally ineffective (Wößmann, 2008). There is also a large body of evidence that convincingly shows that more educated workers receive higher levels of on-and-off-the-job firm-provided training. Furthermore, firms train those who earn the highest returns to VET (Vignoles *et al.*, 2004). Thus, once again we see that there are dynamic complementarities in the VET system and as discussed earlier, firms are more likely to invest in individuals who already have higher levels of skill.

Wider Benefits of VET

Perhaps focusing on just economic returns misses an important part of story (Wößmann, 2008). VET could improve the noncognitive skills of low-skilled adults which will in turn have an impact on the early cognitive and noncognitive skills of their children and as a consequence their lifetime skill acquisition. Alternatively, it might be that training that is not typically valued or supported by government (e.g., that does not provide qualifications or is directly work related) may affect noncognitive skills of adults that will in turn have an impact on the cognitive and noncognitive skills of their children which may have long-term benefits that are typically not evaluated or measured. These potential intergenerational benefits of VET are much more difficult to measure and have not as yet been proven to be important. In fact, some evidence from Australia suggest that while degree and higher-level VET qualifications were associated with better health outcomes, lower-level VET qualifications were not (Stanwick *et al.*, 2006). Only time, and proper long-term evaluations of such qualifications and programs, will tell.

Implications and Conclusions

The economics of VET is complicated and more research, more inventive policy development, and more

rigorous evaluation of VET policies needed worldwide. It is important that these studies not only focus on the economic returns to VET but on the whole process of skill formation. It is clear that policy needs to ensure that lots of investments in skills occur in early life – but this is too late for many adults with low levels of skills. Several current solutions to adults' skills problems are clearly not working and more rigorous testing and evaluation of more innovative policies is what is needed – and these need to take into account the fact that skills beget skills. Furthermore, given the evidence that cognitive skills are potentially less malleable later in life, it may be that adult interventions need to pay closer attention to the development of individuals' noncognitive skill. Focusing on possible intergenerational transmissions of skills may also be important.

The arguments put forth in this article suggest that there should be strong investments in early childhood, both because the sensitive periods for acquiring several capabilities occur early in life, and because successful learning early in life is the foundation for successful learning later in life. These investments also need to be followed up, or there is a strong risk that they will not fulfill their potential. However, this does not address the skills deficit for the current generation of adults in a number of countries in the world. As argued by Wößmann (2008), when policymakers are making decisions about government educational programs, they not only consider economic efficiency (high returns) but also equity (equality of opportunity). He argues that for some educational investments, particularly early investments, efficiency and equity considerations are strong complements – what is good from an efficiency point of view also is good from an equity point of view. However, for other educational investments, in particular, some VET programs, government policymaking often involves a tradeoff between efficiency and equity arguments. This creates a further complexity in the economics of VET.

Fixing the problem of unskilled unqualified adults is certainly hard and costly and at the moment in several countries the solutions being adopted are based on an act of faith rather than sound evidence. We argue that effective, or at least informed, policy involves understanding more fully the nature of dynamic complementarities in the acquisition of skills over the life cycle and a better understanding of the differential sensitive periods in the accumulation of cognitive and noncognitive skills. This is a promising research agenda in which people are starting to make inroads (see Cunha *et al.*, 2008) and this area of research is highly likely to guide education policy, including VET policy in years to come. A more explicit understanding of the tradeoff, or otherwise, between efficiency and equity considerations in government policymaking is also needed. Current research tells us that adult learning builds on learning in adolescence: while equity considerations may

dictate that we need to invest in adults now, only by balancing this with good-quality early investment will we improve outcomes in the future.

Bibliography

- Abramovsky, L., Battistin, E., Fitzsimons, E., Goodman, A., and Simpson, H. (2005). *The Impact of the Employer Training Pilots on the Take-Up of Training Among Employers and Employees*. Institute for Fiscal Studies, DCSF – RR694. <http://www.dcsf.gov.uk/research/data/uploadfiles/ACF8A97.pdf> (accessed October 2009).
- Acemoglu, D. and Pischke, J. (1999). Beyond Becker: Training in imperfect labour markets. *Economic Journal* **109**(453), F112–F142.
- Almeida, R. and Carneiro, P. (2009). The return to firm investments in human capital. *Labour Economics* **16**(1), 97–106.
- Arrow, K. (1973). Higher education as a filter. *Journal of Public Economics* **2**, 193–216.
- Bassanini, A., Booth, A., Brunello, G., De Paola, M., and Leuven, E. (2005). *Workplace Training in Europe (IZA Discussion Paper 1640)*. Bonn: Institute for the Study of Labor.
- Blundell, R., Dearden, L., and Meghir, C. (1996). *The Determinants of Work-Related Training in Britain*. London: Institute for Fiscal Studies.
- Blundell, R., Dearden, L., Meghir, C., and Sianesi, B. (1999). Human capital investment: The returns from education and training to the individual, the firm and the economy. *Fiscal Studies* **20**(1), 1–23.
- Blundell, R., Dearden, L., and Sianesi, B. (2005).
- Bowles, S., Gintis, H., and Osborne, M. (2001). The determinants of individual earnings: Skills, preferences, and schooling. *Journal of Economic Literature* **39**(4), 1137–1176.
- Bynner, J., Dearden, L., McIntosh, S., et al. (2001). *Improving Adult Basic Skills: Benefits to the Individual and to Society*. DfEE Research Brief No. 251.
- Card, D. (1999). The causal effect of education on earnings. In Ashenfelter, O. and Card, D. (eds.) *Handbook of Labour Economics*, pp 1801–1863. Amsterdam: Elsevier.
- Carneiro, P. (2009). Summarizing recent lessons from the technology of skill formation: Implications for education policy. In Carneiro, P. and Spence, M. (eds.) *Education and Growth*. Washington, DC: World Bank.
- Carneiro, P. and Heckman, J. (2002). The evidence on credit constraints in postsecondary schooling. *Economic Journal* **112**(482), 705–734.
- Carneiro, P. and Heckman, J. (2003). In Heckman, J. and Krueger, A. (eds.) *Human Capital Policy, in Inequality in America: What Role for Human Capital Policies*. Cambridge, MA: MIT Press.
- Cohn, E. and Addison, J. (1998). The economic returns to lifelong learning in OECD countries. *Education Economics* **6**(3), 253–307.
- Cunha, F. and Heckman, J. (2009). The economics and psychology of inequality and human development. *NBER Working Paper No 14695*. London: Centre for the Economics of Education.
- Cunha, F., Heckman, J., and Schennach, S. (2008). *Estimating the Technology of Cognitive and Noncognitive Skill Formation*. Unpublished manuscript, The University of Chicago, Chicago, IL.
- Dearden, L., McGranahan, L., and Sianesi, B. (2004). An in-depth analysis of the returns to national vocational qualifications obtained at level 2. *Centre for Economics of Education Discussion Paper No. 46*.
- Dearden, L., McIntosh, S., Myck, M., and Vignoles, A. (2002). The returns to academic and vocational qualifications in Britain. *Bulletin of Economic Research* **54**, 249–274.
- Dearden, L., McIntosh, S., and Vignoles, A. (2001). Basic skills, soft skills and labour market outcomes: Secondary analysis of the NCDS, *DfEE Research Report No. 250 and Research Brief No. 250 and CEE Discussion Papers Nos. 3 and 4*.
- De Coulon, A., Marcenaro-Gutierrez, O., and Vignoles, A. (2007). The value of basic skills in the British labour market. *CEE Discussion Paper, 77*. Centre for the Economics of Education, London School of Economics and Political Science, London.
- Denny, K. J., Harmon, C. P., and O'Sullivan, V. (2003). Education, earnings and skills: A multi-country comparison. *Institute for Fiscal Studies Working Paper W04/08*.
- Dickerson, A. P. (2005). *A Study on Rates of Return to Investment in Level 3 and Higher Qualifications*. Coventry: DTI.
- Fougere, D., Kramarz, F., and Magnac, T. (2000). Youth employment policies in France. *European Economic Review* **44**(4–6), 928–942.
- Friedlander, D., Greenberg, D. H., and Robins, P. K. (1997). Evaluating government training programs for the economically disadvantaged. *Journal of Economic Literature* **35**(4), 1809–1855.
- Grinyer, J. (2005). *Literacy, Numeracy and the Labour Market: Further Analysis of the Skills for Life Survey*. London: Department for Education and Skills Report.
- Groot, W. (1999). Enterprise-related training: A survey. In van Wieringen, F., van Wieringen, A. M. L., and Attwell, G. (eds.) *Vocational and Adult Education in Europe*, pp 357–373. Dordrecht: Springer.
- Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics* **54**(1), 3–56.
- Heckman, J., Hsee, J., and Rubinstein, Y. (2000). The GED is a mixed signal, unpublished working paper, The University of Chicago, Chicago, IL.
- Heckman, J., LaLonde, R., and Smith, J. (1999). The economics and econometrics of active labor market programs. In Ashenfelter, O. and Card, D. (eds.) *Handbook of Labor Economics*, vol. 3A, pp 1865–2085. Amsterdam: Elsevier.
- Heckman, J., Moon, H., Pinto, R., Savelyev, P., and Yavitz, A. (2008). *A Reanalysis of the High/Scope Perry Preschool Program*. Unpublished manuscript, University of Chicago, Department of Economics, Chicago, IL.
- Howell, D. R. and Wolff, E. N. (1991). Trends in the growth and distribution of skills in the US workforce, 1960–1985. *Industrial and Labor Relations Review* **44**(3), 486–502.
- Kane, T. and Rouse, C. (1995). Labor market returns to two- and four-year college. *American Economic Review* **85**(3), 600–614.
- Lalonde, R. (1995). The promise of public sector-sponsored training programs. *Journal of Economic Perspectives* **9**(2), 149–168.
- Leuven, E. (2005). The economics of private sector training: A survey of the literature. *Journal of Economic Surveys* **19**(1), 91–111.
- Machin, S. and McNally, S. (2004). The literacy hour. *Institute for the Study of Labor (IZA) Discussion Paper 1005*, Bonn, Germany.
- Machin, S. and Vignoles, A. (2005). *What's the Good of Education? The Economics of Education in the UK*. Princeton, NJ: Princeton University Press.
- McIntosh, S. (2004). *The Impact of Post-school Vocational Qualifications on the Labour Outcomes of Low-achieving School Leavers in the UK*, mimeo. London: Centre for Economics Performance, London School of Economics.
- McIntosh, S. and Vignoles, A. (2001). Measuring and assessing the impact of basic skills on labour market outcomes. *Oxford Economic Papers* **53**(3), 453–481.
- Meghir, C. and Palme, M. (2005). Educational reform, ability and parental background. *American Economic Review* **95**(1), 414–424.
- Murnane, R. J. and Willett, J. B. (2004). The devil's in the details: Evidence from the GED on the large effects of small differences in high stakes exams. *Economics of Education Review* **23**(4), 336–349.
- Murnane, R. J., Willett, J. B., Duhaldeborde, Y., and Tyler, J. H. (2000). How important are cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management* **19**(4), 547–568.
- Murnane, R. J., Willett, J. B., and Levy, F. (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics* **77**(2), 251–266.
- NESS (National Evaluation of Sure Start) Team (2008). The impact of sure start local programmes on three year olds and their families. *Department for Children, School and Families (DCSF) Research Report NESS/2008/FR/027*, <http://www.dcsf.gov.uk/research/data/uploadfiles/NESS2008FR027.pdf> (accessed October 2009).
- Nores, M., Belfield, C. R., Barnett, W. S., and Schweinhart, L. J. (2005). Updating the economic impacts of the High/Scope Perry Preschool Program. *Educational Evaluation and Policy Analysis* **27**(3), 245–261.
- Ryan, C. (2002). *Individual Returns to Vocational Education and Training Qualifications: Their Implications for Lifelong Learning*. Adelaide: National Centre for Vocational Education Research (NCVER), Australian National Training Authority.
- Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics* **90**, 1–23.

- Stanwick, J., Ong, K., and Karmel, T. (2006). *Vocational Education and Training, Health and Wellbeing: Is There a Relationship?* Adelaide: National Centre for Vocational Education Research.
- Steedman, H. (1993). The Economics of Youth Training in Germany. *Economic Journal* **103**, 1279–1291.
- Torgeson, C., Brooks, G., Porthouse, J., et al. (2004). Adult literacy and numeracy interventions and outcomes: A review of controlled trials. *National Research and Development Centre for Adult Literacy and Numeracy Report*, March. http://www.nrdc.org.uk/uploads/documents/doc_2850.pdf (accessed October 2009).
- Tyler, J. H. (2004a). Basic skills and the earnings of dropouts. *Economics of Education Review* **23**(3), 221–235.
- Tyler, J. H. (2004b). Does the GED improve earnings? Estimates from a sample of both successful and unsuccessful GED candidates. *Industrial and Labor Relations Review* **57**(4), 579–598.
- Vignoles, A., Galindo-Rueda, F., and Feinstein, L. (2004). The labour market impact of adult education and training: A cohort analysis. *Scottish Journal of Political Economy* **51**(2), 266–280.
- Wolf, A. (2002). *Does Education Matter? Myths about Education and Economic Growth*. London: Penguin Books.
- Wößmann, L. (2008). Efficiency and equity in European education and training policies. *International Tax and Public Finance* **1**, 199–230.

Further Reading

- Dearden, L., Reed, H., and Van Reenen, J. (2006). The impact of training on productivity and wages: Evidence from British panel data. *Oxford Bulletin of Economics and Statistics* **68**(4), 397–422.
- Jenkins, A., Vignoles, A., Wolf, A., and Galindo-Rueda, F. (2003). The determinants and labour market effects of lifelong learning. *Applied Economics* **35**(16), 1711–1721.

VOCATIONAL EDUCATION AND TRAINING – INDIVIDUALS AND STUDENTS

Contents

Age Equality in Education and Training
Career Guidance in Vocational Education and Training and in the Workplace
Displaced Workers, Unemployed and Vocational Education and Training
Immigrant Investment in Host Country Vocational Education and Training
Indigenous People and Vocational Education and Training
Persons with a Disability and Vocational Education and Training
Prisoner Education and Training
Vocational Education and Training and the School-to-Work Transition

Age Equality in Education and Training

P Taylor, Swinburne University, Hawthorn, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This article is concerned with older people and vocational education and training (VET). It considers how issues of age and work have received considerable recent attention in many countries, particularly those of Europe, and the reasons for this. Issues covered include lifelong learning, the need for retraining following economic displacement, and pressure to prolong working life as populations age. The article is particularly concerned with the incidence of age barriers in accessing education and training and efforts to overcome them. It begins by considering the importance of participation in learning for older people. Following this, it considers evidence on the extent of participation in European countries, where the issue of extending working lives is currently strongly promoted, before going on to consider factors associated with participation. It then considers the state of policymaking and looks at what is known about overcoming age barriers before offering concluding comments.

The Importance of Social Participation for Active Aging

The World Health Organization (2002) in its Active Ageing Policy Framework defines active aging as “the process of optimizing opportunities for health, participation and security in order to enhance quality of life as people age.” WHO views opportunities for education and lifelong learning as key factors in the social environment that

enhances health, participation, and security: “Loneliness, social isolation, illiteracy and a lack of education. . . greatly increase older people’s risks for disabilities and early death.”

Participation in learning brings major benefits for individuals. The Institute of Employment Studies published a report titled *Learning in Later Life: Motivation and Impact* (2000), key findings of which were as follows:

- 80% of learners reported a positive impact of learning on at least one of the following areas: enjoyment of life, self-confidence, feelings about self, satisfaction areas of life, and coping ability;
- 42% reported an improvement in their ability to stand up and be heard and their willingness to take responsibility;
- 28% reported increased involvement in social, community, and voluntary activities as a result of participation in learning;
- 58% reported that learning improved their enjoyment of life;
- 56% reported an improvement in self-confidence; and
- 36% said that learning had helped them cope better with everyday life.

WHO offers a number of key policy proposals, including the following:

- make basic education available to all across the life course;
- enable the full participation of older people by providing policies and programs in education and training that support lifelong learning; and

- provide older people with opportunities to develop new skills, particularly in areas such as information technologies and new agricultural techniques.

According to Auer and Fortuny (2000), lifelong learning will play a critical role in addressing issues associated with the aging of populations. First, it can help the adjustment of workers' skills and competences to labor-market demand. Second, it can help improve the attachment of older people to the labor market. Third, it can help to overcome productivity declines which may occur with aging.

International institutions emphasize the importance of lifelong learning. For instance, the International Labour Organization (ILO) points to the need for governments to develop policies aimed at avoiding the premature exclusion of older workers from the labor market and stresses the point that workers are less likely to face redundancy in later life if they have benefited from access to lifelong learning opportunities. Lifelong learning is also high on the agenda of the European Commission. The commission communication *Towards a Europe of Knowledge* (1997) places lifelong learning at the center of an integrated approach to future policy action. It stresses the need to promote on a lifelong basis creativity, flexibility, adaptability, the ability to learn to learn and to solve problems which are the conditions we must meet in order to avoid the now-rapid obsolescence of skills. Similarly, at the meeting of the Organization of Economic Cooperation and Development (OECD)'s Education Committee at Ministerial Level: *Making Lifelong Learning a Reality for All* (1996), education ministers emphasized the importance of lifetime learning to deal with the employability of the aging population and agreed on implementation strategies. Additionally, the European Union (EU)'s taskforce on employment led by Wim Kok called for member states to take three key measures to meet European targets:

1. incentives for workers to retire later and for employers to employ older workers;
2. promote access to training for all regardless of age and to develop lifelong learning strategies; and
3. to improve the quality of work to provide attractive, safe, and adaptable work environments throughout the working life.

However, Auer and Fortuny (2000) argue that a shift toward lifelong learning will be a gradual process. Many issues, such as financing remain unanswered. Meanwhile, there is still the imperative need to help the stock of older workers to adapt during the transition. In their view, well-targeted training programs and other labor-market policies can help address specific current labor-market problems of older workers. However, they add that both stock (the current problems facing older workers) and flow (to commence a policy of lifelong learning

for younger-age cohorts) policies must be pursued concurrently.

Levels of Participation in Education and Training by Age

Older people are often considerably less likely than younger ones to participate in training and learning activities. Training tends to be front-loaded but there are important differences between countries. Thus, among the Nordic countries, training appears to be largely age neutral. By contrast, elsewhere, there is a steep decline in the incidence of training by age (OECD, undated).

Table 1 presents figures on job training undertaken in the last 12 months obtained from analysis of the European Working Conditions Survey of 2005. This shows that workers aged 50 or over are generally less likely to receive

Table 1 Incidence of training by age and country for selected European countries (%)

Country	Age group	Under 50	%	50 or older	%
Austria		298	37.1	60	31.1
Belgium		331	43.3	82	39.6
Cyprus		94	22.0	17	11.2
Czech Republic		173	23.3	45	17.9
Germany		194	25.1	39	17.5
Denmark		294	39.9	95	37.8
Estonia		119	29.8	45	25.1
Spain		136	18.6	34	13.4
Finland		366	50.2	162	53.1
France		207	24.4	46	21.7
Greece		102	13.4	11	5.1
Hungary		111	15.7	40	15.0
Ireland		300	40.1	59	24.9
Italy		126	18.6	33	10.6
Lithuania		168	23.2	50	18.7
Luxembourg		179	36.3	29	33.0
Latvia		148	21.7	71	24.0
Netherlands		220	32.5	91	28.5
Malta		136	31.3	36	23.4
Poland		210	26.9	46	25.0
Portugal		119	15.0	17	8.9
Sweden		363	55.1	177	48.0
Slovenia		194	40.1	35	36.1
Slovakia		257	35.0	82	32.8
United Kingdom		307	39.7	83	31.8
Norway		293	44.5	112	35.8
Switzerland		352	50.7	149	46.1
Bulgaria		68	8.9	20	6.1
Croatia		184	22.4	42	26.8
Romania		124	15.7	29	12.7
Turkey		50	6.2	7	4.1
Total		6223		1844	

From European Working Conditions Survey 2005 (author's own analysis).

training than other age-groups, although there are notable exceptions: Finland, Latvia, and Sweden. Notable is the wide variation between countries, with Finland and Sweden having incidences of around 50% for both younger and older workers, with larger countries such as France, Germany, and Italy falling far short of this.

In considering the place of VET in the lives of older people, it is helpful to consider how they fare compared to younger people on specific programs. There is clear evidence older people can benefit from interventions. For instance, in the United Kingdom, Work Based Learning for Adults (WBLA) is a voluntary program which aims to:

- help unemployed people move into sustainable employment;
- help long-term unemployed people to gain skills in areas where there are recognized skill shortages
- enable long-term unemployed people to make a successful transition to self-employment.

In the program, basic employability training is available for people who need extra help before they start job-related training. Those aged 25 or over and have been out of work for 6 months or more and claiming unemployment benefits or another qualifying benefit are eligible.

Findings of a study which aimed to investigate the use and experience of WBLA by people aged over 50 and the factors associated with participation, achievement, and successful provision was published in 2001 (DfEE, 2001). The results were as follows:

- In 1999, people aged 50 and over were underrepresented on WBLA compared to their share of all long-term unemployed people.
- WBLA leavers aged 50 or over were almost as likely as those aged 25–49 to achieve a qualification: 37% of the former compared to 38% of the latter in 1999–2000
- Fewer older leavers found employment: 36% compared to 41% in 1999–2000.

Much of this difference may be explained by the higher proportions of longer-term unemployed and of basic-employability trainees among those aged 50 or over. Fewer trainees of these types in both age groups got jobs. But trainees aged 50 or over who had been unemployed for 3 years or more and nonbasic-employability trainees who had been unemployed for less than 6 months were less likely to get jobs than their younger counterparts. Ageism on the part of employers may have been responsible for this.

Most older trainees interviewed were pleased with their training provision and felt it was meeting their needs in terms of quality of teaching, learning style, and resources. Follow-up survey results showed that about eight out of ten of leavers rated WBLA as useful and helpful in increasing their confidence and in improving their skills/learning new

skills. Leavers aged 50 plus were slightly more satisfied than those aged 25–49 with these aspects.

A more recent evaluation of WBLA undertaken by the Policy Studies Institute and the National Centre for Social Research (2003) and based on survey data shows that short job-focused training (SJFT) accelerated entry to full-time employment. Participation increased prospects of being employed about 5 months after enrolling. The size of this effect was in the region of 5–7 percentage points. However, it was shortlived and no impact was evident by the 10-month mark. However, there did appear to be a sustained effect for clients aged 50 years or over.

More generally, it has been known for a long time that training primarily goes to the most qualified. Research into individual-initiated vocational training among middle-aged workers indicates that it is primarily those with skills already who participate. Those with lower levels of education, who arguably have most to gain, are least likely to self-initiate learning activities (Elman and O'Rand, 2002; Jamieson *et al.*, 1998). Although qualification level is the predominant factor in access to continuing training, age appears to aggravate the situation. Decreasing participation in training with age begins even earlier for those at the bottom of the job ladder. Thus, while a large proportion of managers continue to participate in training after 50 years of age, operatives see their rate of access to training drop significantly around the age of 40. A British Chartered Institute of Personnel and Development survey: *Who Trains at Work*, in 2003, found a significant gap between the training haves and the have-nots with people in lower-grade jobs, those with lower educational achievements, older workers, and part-time workers least likely to be trained by their employer.

However, it is important to note evidence which demonstrates that older people are not disinterested in learning. The UK's National Adult Learning Survey (NALS) found that 67% and 47% of those in the 50–59 and 60–69 age groups respectively were learners. On the other hand, NALS found that 21% of the 50–55 age group and 30% of the 60–69 age group indicated that they were not interested in learning (higher percentages than for any other age group), and the numbers stating that they were too old to learn increased by decade: aged 40–49 (12%), 50–59 (22%), and 60–69-year-olds (30%).

Factors Associated with Participation in Learning Activities

It has been suggested that low levels of training among older workers largely result from a lack of availability (Trinder, 1992). Research would seem to bear this out, providing evidence that the key factor constraining training activities is a lack of opportunities provided by employers rather than disinterest among older workers (Taylor and Urwin, 2001).

This contrasts with the often-expressed view of employers in surveys that training is open to all, and suggests that age discrimination in training often goes unrecognized. Given that the realization of lifelong learning is a major pillar of policymakers' strategies for tackling the economic issues presented by population aging (Auer and Fortuny, 2000: 29), this represents a significant hurdle.

Other research suggests that nevertheless, there may be factors in addition to employer behavior which reduce older workers' training activity. A study of manufacturing workers in a motor-vehicle-manufacturing company (Warr and Birdi, 1998) explored employee-development activities outside formal training. Among the mainly male (95%) workforce questioned:

- older workers were substantially less likely to participate;
- education level, learning motivation, and learning confidence, as well as lower age were found to be predictive of participation;
- support from managers, co-workers, and nonwork sources were positively correlated with activity, while time constraints were found to have a negative association;
- controlling for other factors, age was found to have a negative impact on activity.

Research suggests that older workers often consider themselves unsuited to new learning and lack confidence in such situations. Consideration needs to be given to overcoming such feelings and to the design of programs of training with older people's needs in mind, something that is considered later in this article.

Commentary on Public Policy

In Europe, it is apparent that despite plenty of rhetoric surrounding the social participation of older people in recent years, policymakers are only just beginning to turn their attention from youth education, to education and learning across the life course. In some countries, this is barely the case. In others, rather more has been achieved. According to Withnall (2000) analysis of major policy reports relevant to the development of lifelong learning reveals that strong priority has been given to VET, despite general rhetoric about the noneconomic, personal, and social benefits of lifelong learning generally. She also notes that vague statements concerning the perceived outcomes of policies such as an assured improvement in society and personal fulfillment also tend to permeate recent European policy statements concerned with the promotion of lifelong learning. She adds that while the UK government has acknowledged the need to break down barriers to older people playing a full part in its vision of a learning society, its proposals to date for development and the packages of support available to

adult learners suggest that people who are post work are not genuinely part of the vision. Older people continue to be marginalized in educational policy circles in Europe at a time when later life is primarily characterized as a social problem.

With a new emphasis on extending working lives, policymakers are paying attention to issues of learning for older workers. For instance, in Britain, a potential step forward has come in the form of the Government's Skills Strategy, launched in April 2004, which sets out policies affecting adult learners, including older workers. These include:

- free learning for any adult who does not already have a good foundation of skills necessary for employment;
- opportunities for adults to gain qualifications in technician and higher craft and trade skills;
- a weekly payment for adult learners in priority groups to support them in studying full-time courses in further education;
- removing an age bar on the government's modern apprenticeship scheme so that people over the age of 25 can learn skilled trades; and
- reforming adult information, advice, and guidance services to help adults into learning and ensure that individuals can find out what to learn, where to learn, and what their entitlements are.

On the other hand, policy and practice are apparently at odds in Britain. There is still often an apparent emphasis on younger people, in particular, preparing them for working life. For instance, the British Learning and Skills Council published a consultation document entitled *Funding Adult Learning* in 2003. This set out some potential options and related funding issues for colleges and providers which were being considered. The document argues that the current system is not reaching many of the low skilled and is not leading to a sufficient improvement overall in skills levels. One option, it suggested, would be to focus on adults up to, for example, the age of 30 (19–30 years old), in terms of support. It is stated that:

we recognise that any age cut-off is an artificial mechanism for delimiting support, we do not believe that there are currently sufficient funds in the system to fund fully all adults to achieve a full level 3 qualification. The age limit in this instance would enable young adults who had not completed their initial education to be supported for a longer period. In addition, it would provide an opportunity for those young adults to progress into higher education.

Elsewhere, in France an intersectoral agreement on skills and learning was signed by the social partners in September 2003. Key components are as follows:

- Employees with 20 years experience, and those over the age of 45, regardless of experience, are eligible for a

skills audit after being in post for 12 months, and will be given prioritized access to recognition of their work experience (*Validation des acquis de l'expérience*, (VAE)).

- All employees with 2 years service will be eligible for a career-guidance interview carried out within their company.
- A training passport (passport formation) will be drawn up and updated by the employee. This will set out the skills and knowledge they have acquired.
- An individual 20-h per year training entitlement (*droit individuel à la formation*, (DIF)) is available without age limit. All employees with more than 12 months service are accorded one 20-h credit per year which can be rolled over for 6 years. Employees who lose their jobs will be able to utilize the monetary value of their unused credit to either fund a skills audit, VAE, or a training program.

In Italy, education and lifelong learning have been little developed until recently, with most effort going into equipping younger people with skills. The greatest financial support for lifelong training and employee development at the company level comes from the institution of a training fund under Law No. 236/93, which, from 2003, stipulates that training should be targeted at older workers (45–64 years of age) (OECD, 2004a). In 1999, the Italian government approved the integrated long-term plan on training, education, and research guidelines. As far as lifelong training is concerned, the government has aimed to quadruple the current number of individuals involved in lifelong training activities.

Finally, in the case of Spain, the OECD (2004b) has argued that, despite recent reforms, it lacks a central authoritative body responsible for coordinating adult learning. This is despite evidence, noted above, that it performs poorly compared to other European member states. One could, therefore, conclude then that in Europe, the picture is quite fragmented although the issue of supporting the learning of older people is at least generally on the policy agenda.

Achieving Better Outcomes for Older People

If there is still a gulf between the aspirations and rhetoric of policy makers and the experiences of older people, this is not due to a lack of understanding of how to make training and learning age inclusive. Evans (2002) argues that public policy concerning how to increase and widen participation of mature learners and realize lifelong learning contains many contradictions and paradoxes. He states that the “real harvest of future learners are mature people and yet this obvious and manifest fact is constantly subjected to short-term and often eye catching

headlines, but results in uncertainty in how providers and indeed the potential learners are to be supported on a long term secure basis.” Continuous change to the support mechanisms creates confusion in the potential learner’s mind and changing priorities constantly leads to inevitable initiative overload for providers. Added to this is the front-loading of funding for education.

Recent reviews concerning what works for those aged over 50 and issues for policymakers and providers has been carried out by different bodies (Carlton and Soulsby, 1999, cited in the website of National Institute of Continuing Adult Education; DfEE, 2001; Evans, 2002; Moss and Arrowsmith, 2003; South London Learning Partnership Adult and Community Learning Workshop: The Learning Needs of Older Learners 2003). These are summarized below.

Factors related to job placement and training for older workers:

- Although a range of back-to-work help is available to nonworking older people, and programs are often open to both the economically active and inactive, current take up is mostly among the former, reflecting their greater contact with those able to refer them to programs.
- Having advisers of a similar age to older clients can help overcome barriers and increase rapport.
- For some, experiencing age-specific training makes them feel more comfortable and allows for relevant issues to be addressed, but mixed-age training can help develop a flexible attitude and broaden outlook.
- Clients would welcome in-work support to ensure retention and advancement, particularly at transitional points in their employment.
- Advisers perceive a lack of opportunity (or resources) to refer older clients to specialist providers to address the specific issues they face.
- Both clients and advisers feel that insufficient action has been undertaken to confront and overturn employer age discrimination.
- Older clients need guidance in finding work in a new occupation.
- Skills training before moving into a job is desirable.
- Older people appreciate being consulted about their learning needs.
- Drawing on older peoples’ life experiences and maturity is a useful way of attracting them to learning opportunities.
- Particular barriers include:
 1. lack of and inaccessibility of information;
 2. high cost of courses and loss of earnings for learners;
 3. disabilities and health problems;
 4. inappropriate location and timing of courses;
 5. lack of confidence among older people;

6. negative attitudes toward education, linked to previous experiences and socioeconomic group;
7. gender and cultural issues for both males and females; and
8. availability of transport, mobility problems, and travel costs.

Training provision:

- skills for providers
- lack of recognition and acknowledgment of changing demography in terms of funding and learning needs;
- more effective, specialized training for teachers, trainers, and advisers addressing institutional ageism, low expectations of older learners and assumptions about what older people want to learn;
- use of appropriate teaching strategies and on-course support;
- more collaboration, networking, and sharing of good practice among providers and agencies.
- Communication:
 - more effective and targeted dissemination of information;
 - a comprehensive, centralized source of all relevant information for older learners;
 - more specific targeting of potential older learners;
 - comprehensive and effective guidance systems must be established;
 - establish learning satellites or one-stop shops where people can go for advice on learning opportunities open to them;
 - better use of jargon-free language for induction and course delivery; and
 - more effective targeting of older males.
- Programme design:
 - a more personal touch in recruiting and inducting older learners;
 - better integration of programs;
 - embed older learners' provision in the mainstream;
 - availability of shorter, more focused courses;
 - more nonaccredited courses, work-based and community-based learning opportunities, and social enterprise projects;
 - short work trials for a client group, many of whom may not have worked for a considerable time;
 - offer one-off taster sessions on a noncommitment basis to build confidence and stimulate interest;
 - fewer vocational accredited activities, although for some, accreditation will be relevant and those not desiring a qualification should not be penalized;
 - ensure that ongoing quality evaluation is built into all learning provision;
 - exploration of strategies for improving inclusiveness and widening participation;
 - offer provision which is flexible and tailor-made;

acknowledge that learning is only one aspect of a learning activity;

break away from traditional learning methods and settings and expectations that the learner will come to the learning;

recognize that ethnic minority learners and those who work with them may require additional support;

address physical issues (e.g., hearing and sight problems) faced by older learners;

confidence building activities for older learners, including memory skills;

active creation of a culture of learning and training for older people;

consideration of preferred learning styles;

older tutors, especially male, to act as positive role models;

recognition of gender differences in terms of learning goals and social dimensions of learning;

opportunities for single gender courses and also specifically for older and disabled learners;

providers must develop effective support services for learners on an ongoing and continuous basis;

Individualized and skilled support from a personal adviser and access to specialist provision;

multifunctional use of public premises for one-stop coordinated adult services.

Public policy:

- existing learning provision should be mapped to ensure that it builds on what exists already;
- official bodies should consider how they can make more effective use of the media, both to stimulate interest in learning and to raise awareness;
- the domestic and financial circumstances of learners must be recognized;
- effective and appropriate initial assessment systems must be developed for potential learners followed up with programs of support;
- resources need to be sufficient to respond appropriately to the additional demands made by mature learners; and
- need for more effective partnerships between providers and other agencies.

Concluding Comments

There are marked disparities on the incidence of VET by age. A range of factors are associated with participation among older people. Age is one, but the evidence demonstrates that a better approach would be multifaceted. A need to tackle the issue is acknowledged by European policymakers and international bodies but to date the policy response has been fragmented. There is much

activity in some countries, in others there is little. It is also the case that even among the more active countries, overt institutional barriers to older people's participation remain, although these are being challenged in some cases. At a time when the issue of age and work is the subject of intense interest from public policymakers promoting learning and training remains a major challenge.

Acknowledgment

The research on which this article was based was in part funded by the International Longevity Centre-UK.

See also: Displaced Workers, Unemployed and Vocational Education and Training; Lifelong Learning; Persons with a Disability and Vocational Education and Training.

Bibliography

- Auer, P. and Fortuny, M. (2000). Ageing of the labour force in OECD countries: Economic and social consequences. *Employment Paper*, 2000/2. ILO: Geneva.
- DfEE (Department for Education and Skills) (2001). *Celebrating Older Learners*. <http://www.lifelonglearning.co.uk/cols> (accessed August 2009).
- Elman, C. and O'Rand, A. M. (2002). Perceived job insecurity and entry into work-related training among adult workers. *Social Science Research* 31, 49–76.
- Evans (2002). Mature learners. *t' Magazine*. <http://www.tmag.co.uk/home.html>.
- Jamieson, A., Miller, A., and Stafford, J. (1998). Education in a life course perspective: Continuities and discontinuities. *Education and Ageing* 13(3), 213–228.

- Moss, N. and Arrowsmith, J. (2003). A review of what works for clients aged over 50. *Research Report*. Department for Work and Pensions.
- OECD (undated). Training of adult workers in OECD countries: Measurement and analysis. *Employment Outlook*, p. 149.
- OECD (2004a). *Ageing and Employment Policies: Italy*. Paris: OECD.
- OECD (2004b). *Ageing and Employment Policies: Spain*. Paris: OECD.
- Taylor, P. and Urwin, P. (2001). Age and participation in vocational education and training. *Work, Employment and Society* 15(4), 763–779.
- Trinder, C. (1992). *Present and Future Patterns of Retirement*. London: Public Finance Foundation.
- Warr, P. and Birdi, K. (1998). Employee age and voluntary development activity. *International Journal of Training and Development* 2(3), 190–204.
- Withnall, A. (2000). Older learners – issues and perspectives. *Paper Given at Global on-Line Colloquium*, University of East London Festival of Online Learning, June–October.
- World Health Organization (2002). *Active Ageing, A Policy Framework*. Geneva: WHO.

Further Reading

- European Union Task Force on Employment led by Wim Kok (2003). Jobs jobs jobs: Creating more employment in Europe. *Report to the European Council*, November.
- OECD (2003). Thematic Review of Adult Learning. *Spain Country Note*.

Relevant Websites

- <http://www.eurofound.eu.int> – Eurofound, European Working Conditions Observatory.
- <http://www.eu-employment-observatory.net> – European Employment Observatory.
- <http://www.niace.org.uk> – National Institute of Adult Continuing Education.

Career Guidance in Vocational Education and Training and in the Workplace

P Plant, Danish University of Education, København, Denmark

© 2010 Elsevier Ltd. All rights reserved.

VET and Guidance as Policy Areas

Both vocational education and training (VET) and guidance are high on the policy agenda. The European Union (EU), for example, has set a course for the development of VET: the Copenhagen process for enhancing cooperation regarding VET included four priority areas: (1) European dimension of VET, (2) transparency and guidance, (3) recognition of competences and qualifications, and (4) quality assurance and development. The Copenhagen Declaration mentioned the overall objectives: “to make the EU the most competitive and dynamic knowledge-based economy in the world, capable of sustaining economic growth with more and better jobs and greater social cohesion by 2010” (EU/Commission of the European Communities, 2002). Further, the Copenhagen Declaration reads: “The development of high quality vocational education and training is a crucial and integral part of this strategy notably in terms of promoting social inclusion, cohesion, mobility, employability and competitiveness.”

Further, the pivotal concepts are as follows: dynamic knowledge-based economy; economic growth; social cohesion; social inclusion; and mobility, employability and competitiveness.

Thus, VET is embedded in these policy goals and in the aims of bridging economic and social goals. The same is true in relation to lifelong guidance policies. In 2004, the EU ministers adopted a resolution on lifelong guidance. It describes guidance as a range of activities which enable

citizens of any age and at any point in their lives to identify their capacities, competences and interests, to make educational, training and occupational decisions and to manage their individual life paths in learning, work and other settings in which these capacities and competences are learned and/or used. . . . Guidance provision within the education and training system, and especially in schools or at school level, has an essential role to play in ensuring that individuals' educational and career decisions are firmly based, and in assisting them to develop effective self-management of their learning and career paths. It is also a key instrument for education and training institutions to improve the quality and provision of learning. . . . Guidance throughout life contributes to the achievement of the European Union goals of economic development, labour market efficiency and occupational

and geographical mobility by enhancing the efficiency of investment in education and vocational training, lifelong learning and human capital and workforce development. (EU/Commission of the European Communities, 2004)

Clearly, guidance is seen as an instrument in enhancing lifelong learning, economic development, and social inclusion, all at the same time. This is no an easy task. Policymakers at all levels strive to deal with and to create a balance between these goals. Both the European Training Foundation (ETF) and CEDEFOP, the European Centre for the Development of Vocational Training have issued tools in the form of booklets to facilitate this policymaking process (OECD, 2004a; CEDEFOP, 2005).

A Rationale

An OECD (2002: 121–126) study of human capital development analyzed the role of career guidance in relation to lifelong learning, and active labor market and welfare-to-work strategies. It claimed that in OECD countries, about 40% of the variation in individual earnings is explicable through primary factors, particularly prior educational attainment, literacy levels, and work experience, combined with secondary factors of gender, language background, and parental education. Thus, some of the remaining 60% might be accounted for by motivation and other personal characteristics, including the concept of human meta-capital – that is, people's ability to manage and develop their own human capital – which would include the role of guidance.

Amundson (2003) argues that an individual's engagement with guidance, if it is to make a difference in terms of motivation for lifelong learning, should be personalized through a process that facilitates constructive reflection. This holistic concept of guidance has three dimensions: length, width, and depth. The first dimension implies the need for guidance to take place over a life span, including adulthood and the third age: lifelong guidance. The second dimension considers the wide range of issues that must be addressed in guidance, including personal and social issues, along with career-related ones: life wide guidance. The third dimension concerns the depth of the guidance activities: life deep guidance. How close, how intimate should guidance be? Such an approach has the potential to encourage VET and learning of all types

by individual workers, thereby enhancing their skills as well as promoting their self-confidence and self-esteem.

Such arguments place guidance center-stage although, as Killeen (1996: 84–85) demonstrates, since guidance is often part of a broader package of assistance available for individuals, it is difficult to distinguish learning outcomes arising from guidance from those that are economic. Structural factors explain why there is a greater need for guidance throughout the working lives of adults, including recurrent VET. Fundamental changes in patterns of employment across Western Europe have arisen from a shift from manufacturing to services. A major consequence of this changing pattern has been unemployment and underemployment, distributed unequally across regions, gender, and ethnic groups, that has resulted in social exclusion. One of the aspects of such exclusion is a persistent learning divide. The learning rich with higher levels of qualifications and income are more likely to return to education and benefit from it throughout their working lives. In contrast, the learning poor, who tend to participate the least, work in manual occupations or are unemployed, having left school with few or no qualifications. It is important to note, however, that learning does not need to be formal (EU/Commission of the European Communities, 2001) and, as Colley *et al.* (2002) argue, the boundaries between formal and other learning can be blurred. This also applies to guidance.

Individual employees are increasingly expected to be responsible for managing their own learning and lifelong VET and development throughout their working lives in the context of an economy characterized by increasingly flexible, risky, and uncertain forms of employment (Harrison, 1998: 235). This is where proactive guidance in the workplace comes in: it reaches out to those who would benefit from guidance, but would not approach conventional guidance settings. Access to guidance, physically and mentally, is crucial here (Clayton and McGill, 2000), in order to break the classical issue of the Gospel according to St. Matthew (25:29): “For unto every one that hath shall be given, and he shall have abundance: but from him that hath not, even that which he hath shall be taken away.”

Definitions

Vocational education and training (VET). It can be defined as an activity which prepares learners for careers that are traditionally nonacademic and directly related to a specific trade, occupation, or vocation (http://en.wikipedia.org/wiki/Vocational_education). This definition puts the emphasis on the VET entry points. This article, by contrast, highlights the lifelong and recurrent aspects of VET and guidance.

Career guidance. This, in turn, can be defined in terms of unfolding the width of guidance activities (Ford, 2004).

Signposting. Ensuring that people have accurate information about helping agencies and the guidance services they provide, and are therefore able to select and access the sources of assistance most suited to their requirements.

Informing. Providing information in a range of formats about opportunities available, without any discussion of the relative merits of options for particular individuals.

Advising. Helping individuals and groups to interpret information and choose the most appropriate options.

Counseling. Working with individuals to help them discover, clarify, assess, and understand their own experience, and to explore alternatives and their possible implementation.

Mentoring. Offering individuals and groups appropriate client-focused support to help them overcome personal barriers and realize their potential. Key factors in mentoring include: the skills, personality, and value systems of the mentor; and her/his ability to act as a role model, enter the client's frame of reference, work holistically, and respect the individual's autonomy and independence.

Assessing. Helping individuals, by formal and informal means, to obtain a structured understanding of their personal, educational, and vocational development, in order to enable them to make informed judgments about the appropriateness of particular opportunities.

Teaching. Providing a planned and systematic progression of learner-centered experiences to enable learners to acquire knowledge, skills, and competences related to making personal, educational and career decisions and transitions, and career management.

Sampling. Providing work experience, work trials, learning tasters, and other experiences that enable individuals to gain first-hand experience of opportunities in order to assist and clarify their decisions.

Enabling. Supporting individuals and groups in dealing with organizations providing or influencing employment and learning opportunities.

Advocating. Negotiating directly with organizations on behalf of individuals or groups for whom there may be additional barriers to access.

Following up. Keeping in touch with individuals after main guidance interventions to establish: whether further guidance is required; what forms of guidance and support may be appropriate; and subsequent progress. Follow-up may include the incorporation of tracking procedures where these are considered desirable.

Networking. Establishing specific links with a range of individuals and organizations to support and enhance guidance provision. These links may be formal or informal, but will include regular contact for information exchange, referral, and feedback, and other joint activities such as staff development, monitoring and review, and outreach work.

Feeding back. Gathering and collating information on the unmet needs of individuals and groups (including designated target groups), and encouraging providers of

opportunities to respond by adapting or developing their provision.

Managing. Managing guidance activities into a coherent program, ensuring it is sustainable within its institutional or organizational setting, coordinating and developing its human and physical resources, evaluating its effectiveness, and promoting its services and interests.

Innovating/systems change. Supporting developments and changes in organizational and guidance practice, in order to improve the quality and organization of provision.

Standard Format?

Clearly, from this list it is evident that guidance is so much more than a face-to-face interview: guidance in VET, in particular, includes assessing, mentoring, and sometimes, advocacy, and so on, as depicted in the examples below. However, the standard format of guidance looks like this: two people sit and talk in an office. This is the convention – also in many parts of VET. Yet, some scholars challenge this convention (Amundson, 1998): It is surprising how often guidance closes itself in, where it could open up and reach out to those who actually need guidance. This implies that guidance would have to move out of the closet and out of the office, and enter into new arenas. VET, with its firm links to work, has many options to break away from the conventional mold: learning and guidance are often blended in situated learning (Lave and Wenger, 1991) and in other arenas than the guidance office. Evidently, much VET guidance takes place in educational institutions such as technical schools and colleges, ranging from Australia (VET guidance targeted at indigenous students) to Iceland (guidance to prevent educational dropout). Such approaches are now largely mainstream guidance activities and will not be covered further in this article as they are well known. What is new and interesting is the recognition that much guidance actually takes place outside and alongside formalized institutional realms. In these cases, guidance takes more informal shapes. Workplaces are such arenas to which we will now turn.

In the Workplace

A number of approaches in VET-related guidance in the actual workplace are linked to the concept of mentoring. Mentor, as we recall from the Greek myth, was a friend of Odysseus. When Odysseus left for the Trojan War, he placed Mentor in charge of his son, Telemachus. Thus, the modern use of the word mentor denotes a trusted friend, counsellor, or teacher, usually a more experienced person. In mentoring programs, newcomers are paired with more experienced people in order to obtain good

examples and advice as they advance. Some such approaches are of a general nature, but most have a particular target group in mind such as women or migrant VET students. In these cases, the driving concept is that guidance needs to get out of the offices and into the places where most people spend their working days: in the workplace. This is a proactive approach which includes peer guidance, mentors, shop stewards, and the involvement of the social partners: both employers and trade unions need to benefit from guidance activities, if they are to take place on the shop floor or even on night shifts.

Workplace Guidance in Practice

In many cases, guidance is a remedial activity targeting the unemployed (OECD, 2004b: 24), yet some initiatives involve a more proactive and sometimes outreach-oriented approach (Watt, 1998). One of these is VET-guidance in the actual workplace, taking place at times when workers might require it, within working hours, night shifts included. This approach has been developed in the UK, Denmark, and in Iceland, inspired by Danish initiatives, where the social partners have initiated schemes of workplace guidance in order to improve access to learning and further educational guidance among employees. The following paragraphs draw upon examples from Plant (2006).

Danish trade unions have played an important role in developing voluntary and local initiatives (Villadsen, 1998). These include individual guidance for trade union members on education and training options; targeted information to members on educational and training issues; and group guidance activities. Several trade unions have adopted one or all of these activities. The Danish Confederation of Trade Unions has prepared a tool-box as part of a project known as The Developing Workplace, DUA (Det Udviklende Arbejde, i.e., enriching work). This includes pamphlets on attitudes and barriers toward participation in training and education, intended for discussion groups among workers.

Kvindeligt Arbejderforbund (KAD) has also developed initiatives in peer guidance, similar to those in the UK. KAD was a trade union for women only (now merged with others), mainly for those with few formal qualifications and low pay in different industries and in the service sector. Their attitudes toward participation in continuing VET are mostly utilitarian so that guidance is mainly linked to periods of unemployment (e.g., guidance on unemployment benefit rules and regulations). Trade unions offer personal, social, economic, educational, and vocational guidance to those members requesting it. Although convenient for unemployed members, it was less so for women in hourly paid employment that found it difficult to travel to a KAD office to seek guidance within working hours. KAD therefore attempted several approaches to guidance

activities in order to overcome some of these barriers by reaching out to its members in their workplaces. Some members, known as Spearheads (Spydspidser), attempted peer-guidance activities after being briefed upon the concept of guidance. After a while, this approach failed due to the lack of knowledge of the range of educational and funding options among the Spearheads. Thus, a higher degree of professionalism was necessary for peer-based activities to succeed. KAD then established guidance corners (Vejledningshjørner) in the canteen or resting areas of several major and medium-sized companies. Slightly different guidance corner models are in operation, depending on the initiators (RUE, 2001). A case study from Western Denmark illuminates the main features of this type of initiative (see **Case study 1**).

Jensen (2002: 12), in depicting guidance corners, highlighted the advantages of a broad and impartial approach to workplace guidance. Plant (2004) made a similar point in a report which looked into an EU-funded guidance project in a major company, which was downsizing and moving premises and changing production methods, all at the same time: VET in action. In this case, guidance played a pivotal role in facilitating all these simultaneous changes, making use of both professional guidance staff and a highly engaged trade union representative in the factory. Much of the drive in this guidance project came from the factory floor: a bottom-up approach. This approach, if taken to its limits, may include the users in actually co-designing guidance offers and policies in collaboration with guidance professionals and policymakers (Plant, 2006).

The Witch

In one Icelandic case, a particular method was used, which highlighted the importance of the guidance methods used in the workplace: in this case, metaphors helped the workers to understand and express how they felt about their present work situation and what they might do to change it (see **Case study 2**).

Guidance within Employee Development Schemes

A contrasting top-down approach to workplace guidance initiated by employers was found in terms of the guidance elements of an employee development scheme in a Scottish university, known as Learning Works. Under the scheme, manual, technical, ancillary, clerical, and secretarial staff – all placed on the lower points of the University's pay scale – may apply for an annual learning allowance of £150. A system of site coordinators across 25 different locations in the university exists, and these volunteers, usually previous beneficiaries of the scheme, encourage fellow workers to consider taking up its benefits by

Case study 1: the development of guidance corners

The guidance corner concept is simple: a trade union representative offers person-to-person guidance in a corner of a workplace assembly room, using pamphlets about education or training with a portable computer containing guidance and information programs. This provides members with information on adult education options plus opportunities for discussion, questions, and reflection on their current situation. Initially, such visits were conducted by KAD every 2 weeks in an open consultation mode, with a permanent exhibition of current training and education available, including rules on the funding for educational leave. This, however, did not meet the actual needs of the female membership. First, they did not take pamphlets offered on training or education to any significant degree. Second, few women in work actually requested guidance spontaneously since they did not see how this might benefit them. The concept of guidance, no doubt, was somewhat blurred, and, in their minds, mostly aimed at unemployed people. Third, working in self-governing groups, as is the case in the high-tech company of Bang and Olufsen, puts economic pressure on all members of the group. This limits their willingness to visit the guidance corner since leaving work for guidance would penalize the whole group.

Visits to guidance corners are now conducted every 6 weeks and take place by appointment during working hours, including day or night shifts. In some enterprises, guidance corners are now established permanently, while others are more *ad hoc* and mobile. KAD sees this type of activity as a mainstream member service and part of a long-term strategy to upgrade its members' skills through formal, informal, and nonformal education and training. An important part of this strategy is proactive guidance of an outreach nature that promotes the concept of lifelong learning in workplaces. Interestingly, KAD does not restrict the availability of these guidance corner services to its female membership. Consequently, in practice, some men also benefit from a guidance service that is primarily aimed at women.

Adapted from: Plant and Turner (2005).

offering some initial advice and information. Thus, on a peer basis, they provide a signposting role similar to that of the learning representatives in the UK or the guidance corners in Denmark. In addition, Learning Works employs a part-time learning adviser, who is a qualified career guidance officer with experience of working with adults, to provide impartial guidance to anyone considering taking up a learning allowance (Plant and Turner, 2005).

Conclusion

In Denmark, Iceland, and the UK, accessible guidance in the workplace is an increasingly significant aspect of promoting lifelong VET and learning to develop the skills of those who have not recently participated in education or training: this is recurrent VET. The examples featured here illustrate contrasting approaches to workplace guidance that have been successfully pioneered in the three countries. Employee development schemes in the UK have

Case study 2: Metaphors in guidance

"What is your feeling towards the job in general?" The young worker said he felt frustration toward his work environment because it was disorganized and messy.

Next he was asked to define or judge the environment, to give it a name as if it was a person. He responded: "A witch that does not listen to reason."

"What is your request toward this witch or environment; what would you ask it in order for it to improve things and make them more positive?"

"That she widens the horizon", was his answer.

He was now asked to try to put himself in the shoes of the witch representing his work:

"What is the feeling the witch has towards you?"

"She is feeling wary."

"How does the witch describe you?"

"That I am careless."

"What does it ask of you?"

"That I put order to things in my working environment."

Here we have an individual complaining about his working environment. To help him clarify his attitude he is asked him to judge his environment and to express his wish toward it. He sees himself as passive, a victim. There is no change or improvement of the situation. He may turn the music louder to compensate for his annoyance about his disorderly workplace, but to no avail. A step towards change is taken when he moves his point of view, imagining how the work perceives him and what its request is toward him.

This method brought to light the fact that unskilled workers in Iceland see vocational counseling as something that could be useful for them. Low-paid workers could, in many cases, use advice and support to solve their work-related problems. There is conflict and harassment in many workplaces that need solving. There are also conflicts and grievances regarding the work itself and the working environment and facilities. Often people experience themselves as stuck and not making use of their abilities; they feel that the work contradicts their values, that it does not correspond to their interests and does not give them fulfillment or satisfaction. For such individuals, work is far from fulfilling, they look at work as a necessary evil, as they are just 'doing time' in their workplace, as they cease to live when they enter the workplace and begin a passive existence, without initiative and meaning.

Adapted from <http://www.gla.ac.uk/wg/pilotise.htm>

made professional guidance about learning opportunities more easily available, particularly in larger workplaces. The Icelandic example points to the importance of varied and innovative methods used in guidance. In general, the attitude of the interviewees was sincere, positive, cooperative, and open. The most apparent weakness of the method regarding the group in question was their difficulty to express themselves, about their own experience, interests, abilities, and how they value their work.

The Danish examples of peer guidance and guidance corners also have both strengths and weaknesses. The strengths include the issue of ownership and commitment reported by Plant (1995) in the Danish Eurocounsel studies in which several examples of peer guidance models were

depicted, including an innovative mix of self-governing career development groups of unemployed people and the Public Employment Service (AF).

On the other hand, bringing guidance and information on VET to the workplace creates problems of access. Flexibility in arrangements over location and timing is necessary to allow easier access to guidance for those whose working and domestic arrangements might otherwise create barriers. Guidance must be offered to potential learners in their actual workplace, rather than somewhere else, however central. Breaks at work are often too short to facilitate any in-depth guidance or allow information gathering on a self-help basis to take place. However, those who do not regard themselves as potential learners due to their prior educational history are least likely to seek guidance. This applies particularly to unskilled or manual employees.

Lifelong learning, guidance and counseling, VET, and employment are continuously intersecting cycles and systems in the lives of European citizens. Information and guidance have an essential role to play in facilitating access, progression, and transitions between these cycles and systems over an individual's lifetime (OECD, 2003). A comparative European study by Clayton *et al.* (1999: 31–33) suggests that there is considerable variation in support from policymakers at all levels for guidance to promote lifelong learning. But it is not enough to have a strategy: power issues lie embedded in the introduction of such outreach guidance provision in the workplace – who is to decide whether this sort of activity is feasible or desirable in the workplace? It might well be seen as a disturbance, and moreover, a nuisance which may well inspire workers to leave their present workplace in favor of better offers which were previously unknown to them. The examples in this article, drawing on examples of bottom-up initiatives, suggests that such power issues may hamper workplace VET-guidance, which needs the active cooperation of education providers, employers, trade unions, and guidance bodies, and which has the potential to make lifelong learning a reality if it is complemented by the resources and legal frameworks necessary to sustain it.

Bibliography

- Amundson, N. E. (1998). *Active Engagement: Enhancing the Career Counselling Process*. Vancouver, BC: Ergon Communications.
- Amundson, N. E. (2003). *The Physics of Living*. Vancouver, BC: Ergon Communications.
- CEDEFOP (2005). *Improving Lifelong Guidance Policies and Systems: Using Common European Reference Tools*. Luxembourg: Office for Official Publications of the European Communities. http://www2.trainingvillage.gr/etv/publication/download/panorama/4045_en.pdf (accessed May 2009).
- Colley, H., Hodkinson, P., and Malcolm, J. (2002). *Non-Formal Learning: Mapping the Conceptual Terrain. A Consultation Report*.

- Leeds: University of Leeds Lifelong Learning Institute. http://www.infed.org/archives/e-texts/colley_informal_learning.htm (accessed May 2009).
- Clayton, P. M., Greco, S., Makela, E., and Ward, M. (1999). The provision of vocational guidance and counselling for disadvantaged adults in four European countries: A critical review. *Widening Participation and Lifelong Learning* 1(3), 26–34.
- Clayton, P. M. and McGill, P. (2000). Access issues in adult vocational guidance and counselling for people at risk of social exclusion: Perspectives from two qualitative research projects. *Journal of Access and Credit Studies* 2(1), 4–18.
- EU/Commission of the European Communities (2001). Communication on Lifelong Learning: Making a European Area of Lifelong Learning a Reality (678). Luxembourg: Office of Official Publications.
- EU/Commission of the European Communities (2002). Declaration of the European Ministers of vocational education and training, and the European Commission, convened in Copenhagen on 29 and 30 November 2002, on enhanced European cooperation in vocational education and training. <http://cirius.dynamicweb.dk/Default.aspx?id=473> (accessed May 2009).
- EU/Commission of the European Communities (2004). *Draft Resolution of the Council and of the Representatives of the Member States Meeting within the Council on Strengthening Policies, Systems and Practices in the Field of Guidance throughout Life in Europe*. Luxembourg: Office of Official Publications. http://ec.europa.eu/education/policies/2010/doc/resolution2004_en.pdf (accessed May 2009).
- Ford, G. (2004). *Working Towards All-Age Guidance Partnership. Article from Third Age Guidance (TAG)*. <http://www.gla.ac.uk/tag/taen02.htm%23i> (accessed May 2009).
- Harrison, R. (1998). Guidance in the workplace: Principles under pressure. In Crawford, M., Edwards, R., and Kydd, L. (eds.) *Taking Issue: Debates in Guidance and Counselling in Learning*, pp 235–245. London: Routledge.
- Jensen, P. (2002). *Vejledningshjørner (Guidance Corners) Uddvej 2/2002*, pp 12–15. Odense: Fyns Amt.
- Killeen, J. (1996). The learning and economic outcomes of guidance. In Watts, A. G., Killeen, J., Kidd, J. M., Law, B., and Hawthorn, R. (eds.) *Rethinking Careers Education and Guidance: Theory, Policy and Practice*, pp 72–92. London: Routledge.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Organisation for Economic Co-operation and Development (2002). Education Policy Analysis. Paris: OECD.
- Organisation for Economic Co-operation and Development (2003). *Career Guidance Policy Review*. Paris: OECD. www.oecd.org/els/education/careerguidance.
- OECD (2004a). *Career Guidance: A Handbook for Policy Makers*. Paris: OECD. <http://www1.oecd.org/publications/e-book/9104021E.PDF>.
- Organisation for Economic Co-operation and Development (2004b). *Career Guidance and Public Policy. Bridging the Gap*. Paris: OECD.
- Plant, H. (2006). Involving the Users of Guidance Services in Policy Development. London: The Guidance Council.
- Plant, P. (1995). *Eurocounsel, Phase 1–3*. Dublin: European Foundation for the Improvement of Living and Working Conditions.
- Plant, P. (2004). *Flexibilitet, omstilling og hvidt tøj: Evaluering af vejledningsprojekt på Royal Copenhagen (Flexibility, Change and White Clothes: Evaluation of Guidance Project, Royal Copenhagen)*. Glostrup: Royal Copenhagen.
- Plant, P. and Turner, R. (2005). Getting closer: Workplace guidance for lifelong learning. *International Journal of Lifelong Education* 24(2), 123–135.
- RUE (2001). *Eksisterende vejledningsaktiviteter overfor beskæftigede og virksomheder (Guidance Activities Re Workers and Companies)*. Copenhagen: Raadet for Uddannelses-og Erhvervsvejledning.
- Villadsen, E. (1998). *Uddannelses-og informationscenter: Vejledningshjørner (Education and Information Centre: Guidance Corners)*. Holstebro: Kvindeligt Arbejderforbund.
- Watt, G. (1998). *Supporting Employability. Guides to Good Practice in Employment Counselling and Guidance*. Dublin: European Foundation for the Improvement of Living and Working Conditions.

Further Reading

- Plant, P. (2008). On the shop floor: Guidance in the workplace. In Athanasou, J. and Esbroek, R. V. (eds.) *International Handbook of Career Guidance*, pp 265–283. London: Springer.

Relevant Websites

- <http://www.cedefop.europa.eu> – Cedefop – European Centre for the Development of Vocational Training.
- <http://www.curriculum.edu.au> – Curriculum Corporation.
- <http://www.etf.europa.eu> – ETF (European Training Foundation).
- <http://www.kl.dk> – KL.
- <http://www.kvinfo.dk> – KVINFO Nyt, Mentornetværket.
- <http://www.lo.dk> – LO: Hovedorganisation for fagforeninger.
- <http://www.esoft.gr> – Spiderweb.
- <http://www.gla.ac.uk/wg> – Vocational Guidance for Low-Paid Workers (an EU project on Workplace Guidance).

Displaced Workers, Unemployed and Vocational Education and Training

B Kriechel, ROA – Maastricht University, Maastricht, The Netherlands; IZA, Bonn, Germany

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Workers with education that predestine them to occupations predominantly in declining industries are under severe pressure. There is a higher incidence of job loss in those industries, as the decline usually goes along with plant closures and firm failures. Workers cannot easily change employment out of the industry as they are bound to lose the industry-specific human capital or experience, which is usually only rewarded within the industry.

Particularly in the context of occupations that are linked closely to specific vocational education, occupational mobility is likely to be low. Vocational education is often geared to a specific occupation, teaching specific skills that are not easily transferable. It puts workers with the wrong education at risk of losing their employment. This combined with the additional difficulty of lower reemployment chances given that their education predestines them for occupations in declining industries. The choice is to change occupations, implying the loss of the occupation-specific training and experience, or to try to find employment in a declining sector.

One option is to increase the employability of workers by broadening their skills beyond the occupation-specific requirement to enable them to switch to more general positions. Another option, which is usually only taken after job loss, is to retrain displaced workers such that they can take up different occupations. Many countries offer some sort of training either to specific groups of unemployed workers, for instance, displaced workers, or generally retraining programs to all unemployed.

In this article, we concentrate on the retraining of displaced workers. The preemptive broadening of skills to increase once employability as well as to increase a worker's employment opportunities is not discussed in this context.

In the following, we first discuss the differences between unemployed workers and displaced workers. Next, the different retraining strategies that can be used to aid unemployed workers are discussed. Then we go into programs that countries offer to aid specifically displaced worker, or, in some cases, to general training programs. It will not be possible to completely cover all programs or countries; rather, we have tried to give an overview of approaches from geographically different countries, and to discuss what is known about the effectiveness of these programs.

Displaced Worker

There are considerable differences in the definition of displaced worker. The definitions vary from general layoffs (including firing for cause) to large-scale layoffs due to plant closure or reorganization only. The definition is important, especially in the context of government policies specifically aimed at displaced workers. If displaced workers are to be compensated or aided, they need to be defined properly. In addition, scientific research on displaced workers should also use at least closely related definitions of displaced workers.

Before a special survey was introduced, the Displaced Worker Supplement (DWS) to the US Current Population Survey (CPS), displaced workers were often identified as those workers being fired or unemployed in declining industries, local labor markets, or occupations. This is, however, quite a broad category. With the introduction of the DWS, the questioning used to identify displaced workers in the survey, became a focal point of a definition. According to the DWS, a worker is considered displaced if he/she filled in that he/she lost his/her job due to "a plant closing, an employer going out of business, a layoff without recall, or some similar reason." Most economists use definitions that are very close to the above wording, even if they are not using the DWS.

Displaced workers differ from unemployed workers. These workers are not unemployed for cause, or due to their own actions, but rather as a result of a plant closure or large-scale downsizing. The crucial difference between the groups is that the displaced workers undergo no selection, that is, there is no negative selection of the least-productive workers into the layoff procedure.

The group of unemployed contains a mixture of workers: those who are laid off for cause and those who became unemployed for external reasons. Thus, on average, the productivity of the pool of workers should be lower than that of a comparable group of displaced workers. This is the result of the firm selecting the least-productive workers to be laid off first. It can be shown that displaced workers will do better, in terms of wages earned after their unemployment, than unemployed workers who were not displaced.

Causes of Worker Displacement

There are several reasons that can lead to plant closures and large-scale layoffs. They can be due to new production

technologies, increased foreign competition, changes in taste, and managerial errors. The first two causes can also be seen as external causes of displacement. They are the result of common technological progress, rendering some activities superfluous. Mechanization is one of the examples for the production process that leads to displacement of workers.

The cause of displacement, either through foreign competition or technological progress, actually benefits society as a whole. For example, free trade benefits all members of society, as they can buy more products more cheaply than those that are domestically produced. Opening of trade adversely affects not only import-competing companies, but also workers employed in these firms that directly compete with the foreign firms. Often, a whole range of products will be predominantly imported, rendering many workers unemployed in the present occupations. It is therefore of no surprise that society as a whole tries to redistribute some of the gain it achieves to those who are at a loss due to free trade. Similar arguments can be made for technological progress.

Losses of Displaced Workers

Displacement implies a significant burden on a worker. First of all, workers will have to find new employment, which implies that they have search costs and invest time. During the search process, they will depend on unemployment benefits or severance payments from their former employer. If they find new employment, they are often stripped of rights or protections due to seniority offering them benefits over newcomers. Empirical research has consistently shown that higher-educated workers as well as younger workers have shorter search time. However, these direct costs are, according to research into the effects of displacement, only a minor part of the displacement losses: wage drops for weekly earnings are reported to be around 10–20% on average.

In the long run, these losses can even be greater. First of all, the wage drop is not temporary – average earnings remain consistently lower compared to workers who were not displaced. In the evaluation of long-term earning development of displaced workers, negative income effects could be found even after more than a decade. Many displaced workers are not able to find long-lasting jobs. This makes them prone to dismissal in economic downturns (Eliason and Storrie, 2006).

Explanations of Wage Loss

How can the wage losses be explained? Displaced workers can often be found in companies or industries that pay above the average wage. These wage premiums can have different reasons: they can be attributed to investments in human capital, a reward for staying longer with the firm,

which is also known as deferred-compensation schemes, or it might reflect higher compensation for achieving higher hierarchical positions within a company. Displaced workers with higher-than-average labor-market pay will, after being displaced, fall back to lower wages.

Most of the economic displacement literature concentrates its attention on the human capital explanation, as there is strong empirical evidence for its importance. If the wage premium is due to an investment in human capital, it is important to understand how this is productive for the worker in a specific firm. The literature distinguishes between specific human capital and general human capital. General human capital increases the productivity of a worker in all firms and sectors, whereas specific human capital can be described as knowledge or skills that are specific for a certain firm, occupation, or industry. Wage losses that are incurred have therefore been explained by the loss of firm-specific human capital, and for those changing industries, in addition to the loss of industry-specific human capital. On average, industry switchers incur higher losses than workers who can remain within the same industry.

An alternative explanation is a wage premium that workers earn from working in high-wage industries or occupations, which can be explained by compensating wage differentials. Compensating wage differentials assumes that these wage differentials are the result of preferences of workers. Occupations or industries which are not preferred, because they are, for example, dangerous or unhealthy, warrant a higher wage. A good example is miners, who are often well paid for their work. However, it is dangerous and unhealthy work; the higher wage can be explained by these differences. If workers move from high-wage industries, which involve discomfort, to low-wage industries after displacement, it might simply involve a move to a more agreeable occupation, with lower pay.

The internal labor market with hierarchical positions can offer a further explanation for pay differences. If within a firm or sector, a worker's career can bring him/her across several hierarchical positions, it will increase not only the wage spread, but also the average pay that can be expected. If a worker moves away from such a firm with a hierarchical structure, and if he/she comes from a higher hierarchical position, he/she will have to receive a position that is at par with his/her old hierarchical position. There are firms in which higher positions are filled from within. It is therefore difficult for a worker to be able to recoup a position at the same level of his/her former hierarchical standing. Empirical research shows that higher hierarchical positions increase the chance of lower search time, hence less unemployment, along with lower wage losses. Higher hierarchical position can be seen as a form of general human capital, that is, general management skills, which can be transferred to

new employment, thus diminishing wage losses (Kriechele and Pfann, 2005).

Another important aspect is that of the local labor market. Many workers who are displaced are reluctant to relocate their families to other regions. Therefore, the local labor market and the job opportunities within the local labor market will influence the opportunities that a displaced worker faces. A part of the decline in wages is due to a decline in the local labor market of the occupation a displaced worker held. Workers who restrict themselves to the regional labor market diminish the chances of finding employment at the same level and remuneration as before displacement.

Assisting Displaced Workers

The best way to aid displaced workers is by helping them cope with the job loss and moving on to new employment. Avoiding the job loss and supporting the displacing company just keeps inefficient jobs in place. It is better to move the workers to new, more productive employment opportunities and careers. The problem is, however, that many times the loss of jobs is the result of a structural shift in the employment structure of a region or country. This means that the current employment is no longer viable in the long run. A good example is coal mining in Europe. A combination of technological change, replacing coal by other energy sources, and international trade importing coal from countries in which the extraction was cheaper, made the occupation of coal miner redundant in many European countries. The occupation was often linked to a set of vocational training degrees that allowed a worker to pick up the occupation of a miner. However, once the mines were closed many of the skills were no longer needed.

Retraining is then one of the most feasible options, whose goal is to prepare the displaced workers for a new career in a different sector and occupation. Sometimes a part of the skills can be used in the new occupations; sometimes retraining means shifting the set of needed skills completely.

Retraining can be led by two forces. One is to include the existing set of skills, and try to retrain for occupations that are similar, but are not struck by the downturn; or, and that is the case in many government-sponsored retraining programs aiding unemployed, the training is predominantly demand led, offering retraining to occupations for which there is short supply, either now or in the future.

Assistance programs that offer assistance in job search or retraining to other occupations specifically to displaced workers have a long tradition in the United States. In Europe, the whole focus is on the unemployed. Given the generous welfare system, retraining is seen in Europe

as a way to move unemployed workers out of a costly welfare system into productive occupations. Here, the discussion of the assistance programs is confined to the retraining programs, either devoted to displaced workers or to unemployed workers facing unemployment because of structural change.

Assistance Programs in the United States

The assistance to displaced workers can be traced back to the Manpower Development Training Act of 1962 for those displaced from work due to technology, and the Trade Adjustment Assistance (TAA) Act which targeted workers displaced through import competition. In 1974, the trade act established the current TAA program to assist workers employed by a firm that reorganizes due to increased imports or shifts in production to foreign countries. The related Alternative Trade Adjustment Assistance (ATAA) program does not provide any training, as it is devoted to older workers who are allowed to find alternative employment at lower wage while receiving a wage subsidy from the program fund. The main TAA program, however, entails allowances for job search and relocation, as well as training assistance. The training can be classroom, on the job, or customized to the needs of specific employers. The program offers compensation for the training as well as income support for the duration of the training. Many evaluation studies that are made about these programs show mixed results. Moderate gains can only be found for the short-term (Leigh, 1990; Bloom, 1990). In a longer-term evaluation, only weak evidence can be found that the training programs aimed at the displaced workers support the labor-market situation (Decker and Corson, 1995).

Assistance Programs in Countries of the European Union

Assistance programs in the countries of the European Union (EU) are usually devoted to all unemployed workers. Almost all countries have some provision for retraining unemployed workers. Here, only Austria, France, Germany, Spain, and Sweden are mentioned. The programs mentioned have experienced some form of scientific evaluation.

Austria

Training program is one of the many active labor-market programs in Austria. The aim of such a training program is a qualification enhancement of the participants. Vocational training courses result in degrees that are equivalent to apprenticeship degrees. Next to the vocational training, there is also skill training to enhance, for example, language or computer skills. While scientific evaluation is rare, a cohort study of training participants finds a

positive effect of program participation of a cohort of workers on their employment stability (Zweimüller and Winter-Ebmer, 1996). In an extensive case study of a special training scheme of a large-scale downsizing in the steel industry, a special program of retraining was also evaluated (Winter-Ebmer, 2006). The scheme consisted of a combination of retraining and placement assistance, and resulted in considerable wage gains and improved employment prospects. One should be careful about the generalization of the results of this specific case.

France

France introduced specific legislation to aid displaced workers in 1987. In the *Convention de conversion*, employers are obliged to offer displaced workers retraining schemes. While the participation is voluntary for workers, they are offered to all workers with a tenure of 2 years or more, and who are 57 years or younger. Following an assessment of the worker's skills, job-search assistance is given and extra training is offered.

Germany

Germany has a long tradition of active labor-market policies. Training is one of the most important measures. The workers who are supported by the programs are unemployed as well as those who are threatened by unemployment and reentrants. Through the federal employment agencies, participants can get support to vocational training courses leading to a first professional degree. In addition, further vocational training and retraining programs are offered if deemed useful by the caseworker of the federal employment agency. Since the reforms of 2005 training vouchers that enable competition between training providers are given to eligible participants. Recent evaluations studies, using administrative data, suggest that training has positive effects on the employment probabilities and a positive effect on earnings. These evaluation studies examined the retraining of East-German workers. Both short-term training (lasting up to 6 months) and long-term training were evaluated (Lechner *et al.*, 2007).

Spain

Training schemes in Spain usually combine training with practical work experience. In an evaluation study for one region using a control group to compare the results, training programs were shown to increase the employment chances of the participants. However, average earnings were not improved compared to the unemployed who did not take part in the training program.

Sweden

In Sweden, retraining of unemployed workers is not generally devoted to displaced workers. Rather they

fall under the general labor-market training programs (AMU-Arbetsmarknads utbildning) for all unemployed workers. The objective of the program is to improve job seekers' chances of finding a job. Within the AMU there is a vocational training component, which can take place in educational organizations, universities, and companies. The use of AMU is widespread, and it existed for several decades in Sweden. One of the evaluation studies could show that training participants had a significant higher chance of leaving unemployment if vocational training is followed (Richardson and van den Berg, 2001).

Evaluation of Training Programs

In the evaluation of the assistance programs, it is important to distinguish how an evaluation is made. Most programs involve some form of evaluation. However, not all include a rigorous scientific evaluation. A distinction could be made between evaluation using experiments and quasi-experiments. Scientific evaluation always implies that the success of an assistance program is evaluated using some form of comparison group. In the experimental setup, the workers are divided at random in two groups: one receives assistance, the treatment group, and the other does not receive a treatment, or a simple placebo-type treatment, the control group. In the quasi-experimental setup, a control group is constructed using external factors that influence the participation in training without affecting the variables of interest. One typical example is administrative rules that, for example, allow workers into programs based on age. Comparing workers that just are allowed into the program and those that are just not eligible allows for a comparison of similar groups.

The importance of rigorous scientific evaluations can be seen when one compares the results of informal evaluations that in most cases report positive effects of the program on employment chances and income, with programs that have scientific evaluations. Scientific evaluations in many cases report only small effects on employment chances and on income, if they find any effects at all. Particularly important in this context is to look at the scientific evaluations that have been done for several cases of firm or plant closures in the United States. While the evaluation at the end of each program concluded increased employment chances and positive effects on post-displacement income, the scientific evaluation using quasi-experimental approaches could find no effect on income, and only negligible effect on employment chances. In this light, the retraining programs proved to be ineffective, at least not more effective than simpler job-search assistance programs. These job-search assistance programs tend to be much cheaper than the more elaborate retraining programs. Based on studies that had both a scientific and nonscientific evaluation, one can

thus conclude that nonscientific evaluation studies tend to overestimate the (positive) impact of their programs.

The studies evaluated above reflect only the development over the shorter term. A longer-term scientific evaluation of retraining programs in eastern Germany did find positive effects. In the program, no distinction is made between unemployed and displaced workers. The program information is matched to large-scale administrative data sets to construct control groups. In the evaluation, one can distinguish between participation in short- and long-term training, which are respectively shorter or longer than 6 months – those that enhance the skills within the own profession, and those that use retraining leading to a different vocational degree than the one held by the trainee. In the short-term, the evaluation shows negative effects on earnings and employment prospects. However, in the longer run, about a decade after the training, employment prospects and earnings are increased.

However, especially given the specific circumstances in eastern Germany, it is unclear whether the positive results would also be found in different settings and in other countries. It is crucial to consider careful and scientific evaluation studies to evaluate training programs for displaced workers. Preferably, evaluations should be continued to evaluate long-term effects. Administrative data that have become available in several countries would allow to construct both a control group and a (quasi-)experimentally determined treatment group.

Overall, scientific evaluation of retraining programs has shown that there are some improvements in employment prospects and, in some cases, in income as well. This is especially true for the programs in Europe. Some of the effects can only be found if the evaluations consider longer time frames. However, most evaluations do not take the cost of programs into account. While retraining programs are often evaluated against simpler and cheaper job-search assistance programs, they are not compared in terms of costs and benefits.

See also: An Overview of Vocational Education and Training; Evaluation Research; Globalization and Vocational Education and Training; Human Capital.

Bibliography

Bloom, H. S. (1990). *Back to Work: Testing Reemployment Services for Displaced Workers*. Kalamazoo: W.E. Upjohn Institute.

- Decker, P. T. and Corson, W. (1995). International trade and worker displacement: Evaluation of the Trade Adjustment Assistance program. *Industrial and Labor Relations Review* **48**, 758–774.
- Eliason, M. and Storrie, D. (2006). Lasting or latent scars? Swedish evidence on the long-term effects of job displacement. *Journal of Labor Economics* **24**, 831–856.
- Kriechel, B. and Pfann, G. A. (2005). The role of specific and general human capital after displacement. *Education Economics* **13**, 223–236.
- Lechner, M., Miquel, R., and Wunsch, C. (2007). The curse and blessing of training the unemployed in a changing economy: The case of East Germany after unification. *German Economic Review* **8**, 468–509.
- Leigh, D. E. (1990). *Does Training Work for Displaced Workers? A Survey of Existing Evidence*. Kalamazoo: W.E. Upjohn Institute.
- Richardson, K. and van den Berg, G. J. (2001). The effect of vocational employment training on the individual transition rate from unemployment to work. *Swedish Economic Policy Review* **8**, 175–213.
- Winter-Ebmer, R. (2006). Coping with a structural crisis: Evaluating an innovative redundancy-retraining project. *International Journal of Manpower* **27**, 700–721.
- Zweimueller, J. and Winter-Ebmer, R. (1996). Manpower training programs and employment stability. *Economica* **63**, 113–130.

Further Reading

- Couch, K. A. (2001). Earnings losses and unemployment of displaced workers in Germany. *Industrial and Labor Relations Review* **54**, 559–572.
- Farber, H. S. (1999). Mobility and stability: The dynamics of job change in labor markets. In Ashenfelter, O. C. and Card, D. (eds.) *Handbook of Labor Economics*, vol. 3.1, pp 2439–2483. Amsterdam: Elsevier.
- Fallick, B. C. (1996). A review of the recent empirical literature on displaced workers. *Industrial and Labor Relations Review* **50**, 5–15.
- Gibbons, R. and Katz, L. F. (1991). Layoffs and lemons. *Journal of Labor Economics* **9**, 351–380.
- Hamermesh, D. S. (1989). What do we know about worker displacement in the U.S.? *Industrial Relations* **28**, 51–60.
- Heckman, J. J., Lalonde, R. J., and Smith, J. A. (1999). The economics and econometrics of active labor market programs. In Ashenfelter, O. C. and Card, D. (eds.) *Handbook of Labor Economics*, vol. 3.1, pp 1865–2097. Amsterdam: Elsevier.
- Jacobson, L. (1998). Compensation programs. In Collins, S. M. (ed.) *Imports, Exports and the American Worker*, ch. 11., pp 473–537. Washington, DC: Brookings Institutions Press.
- Jacobson, L. S., LaLonde, R. J., and Sullivan, D. G. (1993). Earnings losses of displaced workers. *American Economic Review* **84**, 685–709.
- Kletzer, L. (1998). Job displacement. *Journal of Economic Perspectives* **12**, 115–136.
- Kluve, J., Card, D., Fertig, M., et al. (2007). *Active Labor Market Policies in Europe – Performance and Perspectives*. Berlin: Springer.
- Kuhn, P. J. (2002). *Losing Work, Moving on: International Perspectives on Worker Displacement*. Kalamazoo: W.E. Upjohn Institute.
- Neal, D. (1995). Industry-specific human capital: Evidence from displaced workers. *Journal of Labor Economics* **13**, 653–677.
- Stevens, A. H. (1997). Persistent effects of job displacement: The importance of multiple job losses. *Journal of Labor Economics* **15**, 165–187.
- Topel, R. (1990). Specific capital and unemployment: Measuring the costs and consequences of job loss. *Carnegie Rochester Conference Series in Public Policy* **33**, 181–214.

Immigrant Investment in Host Country Vocational Education and Training

I Kogan, University of Mannheim, Mannheim, Germany

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Blue-collar occupations – Pertain to the following occupational groups: skilled agricultural and fishery workers; craft and related trades workers; plant and machine operators and assemblers; elementary manual occupations.

Generalized on-the-job training – Training that equips its participants with general skills, for example, literacy, arithmetic, general cognitive skills, basic cultural, and communication abilities. Such skills can be perfectly transferable between occupations.

Specific on-the-job training – Training that delivers occupation specific skills, that is, skills which are more instrumental to particular functional tasks, for example, accounting, computer programming, and mastery of specific crafts, tools, or machinery. The transferability of specific skills is more limited.

White collar occupations – Pertain to the following occupational groups: legislators, senior officials and managers; professionals; technicians and associate professionals; clerks; service workers, shop and market sales workers.

Introduction

The integration of immigrants into the labor market is one of the central issues of migration research. It largely determines the economic impact of immigrants on the receiving country, as well as the social integration of immigrants and their offspring. Whether immigrants soon become economically integrated or whether long-term ethnic inequality emerges has serious consequences for both the immigrants and the receiving society. Research has shown that immigrants' economic difficulties, for example, higher risks of unemployment and their overrepresentation in the lower ranks of the occupational hierarchy, are to some degree explained by their lack of human capital resources relevant for the receiving society.

First, due to the differences in the level of educational opportunity in various countries, in particular, disparities between Western industrialized countries (as host societies) and Third World countries (as sending societies),

immigrants, and above all those coming from the Third World, bring with them on average lower educational qualifications than is the case among the native-born population. In this regard, the issue of immigrant self-selection should also be mentioned. Normally, people voluntarily emigrating from a particular country are not a random sample drawn from the population, but rather a select group. Economic immigrants, according to Chiswick (1978), represent the more ambitious, motivated, risk-taking, and able individuals of source countries. This self-selection argument has, however, been theoretically and empirically challenged by Borjas (1987, 1990, 1994), who has suggested that immigrants' selectivity in both observed and unobserved traits is not always positive, but rather depends upon the relative returns on skills in source and destination countries. Second, some aspects of human capital, particularly language skills and cultural knowledge, are country specific, that is, they are more relevant and productive in some societal contexts than in others (Borjas, 1994; Esser, 1999). Migration makes these aspects less relevant and leads to a certain devaluation of human capital (Chiswick, 1978, 1991; Friedberg, 2000). With the passage of time in the host country, differences in human capital between the native-born population and immigrants should narrow as immigrants learn the host country language, gain knowledge about the functioning of the host country's labor market, and acquire local education and training (Chiswick, 1978). The third point to be mentioned here is that the trend toward assimilation might be disrupted by immigrants' lack of investment in the human capital specific to the host country, including host country language.

Individual Determinants of Investment in Vocational Education and Training

Why might immigrants be investing less in the host country education and training despite the fact that their economic performance is positively affected by host country language acquisition (Chiswick and Miller, 1995, 2002) and by schooling and training obtained in the host country (Borjas, 1994; Friedberg, 2000)? One of the reasons for immigrants' noninvestment is that they often lack human capital or that their sending country's educational credentials are not recognized in the host country.

Further, it is a well-established finding that low-qualified individuals – irrespective of their migration status – have the lowest training participation rates (OECD, 2003; Booth, 1991; Brunello, 2001). Those who are already more educated, be it natives or immigrants, participate in employment that requires high skills and therefore have higher chances to reinforce their training (OECD, 1999; Arulampalam *et al.*, 2004). This pattern holds for all Organization for Economic Co-Operation and Development (OECD) countries, with particularly pronounced exclusion of low-skilled workers from training opportunities in Canada, Norway, Portugal, Spain, Switzerland, and the United Kingdom.

Apart from lower qualification levels, immigrants' possibly lower rates of education and training might be related to their less favorable employment situation. It has been stressed that training incidence depends on whether people work or not, what sector of the employment and the enterprise they work in, their occupation, and the nature of their employment contract (e.g., fixed term or permanent) (OECD, 2003; Arulampalam *et al.*, 2004). In almost all OECD countries, training participation rates are the highest for workers in white-collar high-skilled occupations, those that require higher education attainment levels. Within blue-collar occupations, adult training participation rates are higher for high-skilled than low-skilled jobs. This is true for all OECD countries with the exception for Portugal and Spain. Furthermore, workers in the service sector jobs have higher training probabilities than workers in other sectors. Finally, it appears that the incidence of training decreases with the size of the enterprise.

In almost all Western receiving societies, immigrants are overrepresented in the so-called secondary labor market (Piore, 1971, 1979; Doeringer and Piore, 1971), that is, jobs characterized by low wages, irregular working hours, bad working conditions, and little prospect of mobility. It is shown above that secondary market workers have reduced access to generalized or specific on-the-job training. Thus, immigrants appear unable to considerably, if at all, improve their human capital (Flanagan, 1973; Cain, 1976; Wallace and Kalleberg, 1981). Instead, they become even further de-skilled and unsuited to work in better, primary sector jobs (Taubman and Wachter, 1986).

All the above-mentioned arguments suggest that once one appropriately controls for individual human capital resources as well as employment characteristics, takes into account additional immigrant-specific determinants of investment decisions (e.g., language proficiency), one should not encounter a residual effect of migration status on the probability of acquiring host country education and training. Sometimes, however, this residual effect exists and could be attributed, at least to some degree, to the reluctance on part of immigrants (as well as employers, see below) to invest in their education and training in

the host country. Often, immigrants consider their stay in the host country as temporary, and, reasonably, refuse to make investments that are not certain to pay off, or they are disinterested in further training due to expected lower or absent returns to this quite costly activity (Bonacich, 1972; Heath and Ridge, 1983; Chiswick, 2000; Dustmann, 2000; Duleep and Regets, 1999).

The underinvestment argument can be formally modeled as follows (see Esser, 1993):

$$EU(\text{inv}) = U(\text{squ}) \quad [1a]$$

$$EU(\text{inv}) = pU(\text{inv}) + (1 - p)U(\text{squ}) - C \quad [1b]$$

where *inv* and *inv* are two immediate options faced by an individual, the first meaning the engagement in a risky investing activity, while the second means the waiving of an active investment. *EU* is an expected utility for each of the options. *U(squ)* denotes the (securely) expected gain by keeping the *status quo* without any investment, and *U(inv)* the expected gain for a successful investment. The (subjective) probability of success is *p*, while *C* denotes the (certain) investment costs. Investment occurs when

$$U(\text{inv}) - U(\text{squ}) > C/p \quad [2]$$

The term on the left-hand side stands for the investment motive, the one on the right stands for the investment risk. When *p* decreases, the transition threshold increases disproportionately, and if it approaches zero, then even extremely high incentives for investment have no effect.

Such a model implies three consequences with respect to the investment motives of immigrants. First, immigrants would presumably have higher costs, *C*, as they often lack specific knowledge about the functioning of the educational and training system in the host country and are possibly deficient with regard to social networks needed to gain missing information (Lundberg and Startz, 1998). Furthermore, they might simply have fewer financial resources to fund training or to sustain themselves while on training (Cain, 1976). Second, immigrants may have larger fears of rejection (expecting discrimination) in the application to undertake education and training activities, resulting in a lower subjective probability, *p*. Third, and probably most important with regard to recent or precarious immigrants, their investment utility may be lower due to their often temporary orientation. In the model outlined above, this would result in a lower value for *U(inv)* as the time horizon for the benefits from training activities to be realized is reduced. Even for permanent immigrants *U(inv)* might be lower as the pay-off to education and training might be reduced due to a belief in or in fact actual employer discrimination. This means that immigrants might invest less in further education and training than comparable natives.

It is important to note that employers anticipating immigrants' temporary presence, might also be hesitant

to invest in education or offer on-the-job training for their immigrant employees, since such investments might be lost if immigrants return to their home countries (Offe and Hinrichs, 1977). The employers' decision making could also be modeled as described above and is affected by employment characteristics, such as industry, occupation, and the size of the enterprise.

VET Participation among Immigrants: Empirical Evidence

Most of the empirical research on immigrants' post-migration investment in education and training is conducted by economists who underscore investment in formal education with less, if any, attention paid to the vocational or on-the-job training. Another body of literature deals with host country language acquisition (see Esser, 2006). Studies cited below examine individual determinants of educational investments among immigrants in a single host country with particular focus on pre-migration characteristics (i.e., national origin and pre-migration schooling), conditions at the time of migration (i.e., age at migration, migration motive, and unemployment level) and conditions after migration (i.e., host country language proficiency, and length of stay; see Chiswick and Miller (1994) for the framework).

Thus, using the 1976 Survey of Income and Education (SIE) Borjas (1982) analyzed post-migration years of schooling of male immigrants of Hispanic origin in the United States. Hashmi Khan (1997) replicated Borjas' results complementing the SIE data with the 1980 US Census of population and extending the analyses to various immigrant groups in the US. Another novelty of her study was to expand the set of explanatory variables to measure the cost and quality of US schooling, student status, and citizenship. Along with the dependent variable, current school enrollment was also analyzed together with years of schooling after migration.

For Australia, Chiswick and Miller (1994) examined the acquisition of post-migration qualifications, including the number of years of schooling and the type of education, and their impact on immigrants' labor market performance. A more recent study for Australia by Cobb-Clark *et al.* (2005) utilized longitudinal data to examine school enrollment within the first 4 years after arrival. Furthermore, the role of the partner's education on individual investment decisions is explored in this study. In Europe, empirical research on post-migration educational investments is extremely rare. A single known study is Van Tubergen and van de Werfhorst's (2006) analysis of the determinants of educational participation and attainment in the Netherlands.

All the above-cited studies report similar results for the migration-related characteristics, age at migration,

and length of stay. Age at migration has a negative effect on post-migration investments in education (Hashmi Khan, 1997; Chiswick and Miller, 1994), with the effect being curvilinear, that is, stronger for those who immigrated at a younger age. (Van Tubergen and van de Werfhorst (2006) find a negative linear effect for the Netherlands.) It indeed appears more attractive for younger people to invest in host country education than to those who arrived at an older age because the time horizon to enjoy the benefits of this education is larger. At the same time, the cost of educational investments is lower as younger people are more efficient in learning.

Studies also find that post-migration schooling depends (nonlinearly) on the length of stay in the host country (Cobb-Clark *et al.*, 2005; Hashmi Khan, 1997; Chiswick and Miller, 1994; Van Tubergen and van de Werfhorst, 2006). Immigrants who have been longer in the receiving country have made more investments in education, and this effect is particularly pronounced for the first few years after migration.

The studies report somewhat contradictory results for migration motive and origin groups. Borjas (1982) and Hashmi Khan (1997) put forward the evidence that refugees (approximated by immigrant origin, i.e., immigrants from China, Cuba, and Vietnam) are more likely to invest in post-migration education than non-refugees. Identifying refugees by category of admission instead of country of origin, Chiswick and Miller (1994) find that refugees are slightly less likely to invest in post-migration education. Furthermore, whereas Hashmi Khan (1997) and Cobb-Clark *et al.* (2005) find that non-English origin groups are more likely to invest in education after migration than immigrants more similar to the host country; Chiswick and Miller (1994) do not find differences by language origin. Van Tubergen and van de Werfhorst (2006), on the other hand, report more difficulties for Mediterranean immigrants (Turks and Moroccans) to participate in the Dutch school system than for immigrants from the former Dutch colonies (Suriname and the Dutch Antilles), equipped with the Dutch language upon arrival.

The studies also differ in their conclusions on the relationship between pre- and post-migration investments in training. Borjas (1982) and Hashmi Khan (1997) report a strong negative effect of the number of years of education acquired before arrival and the number of years of schooling in the United States. Chiswick and Miller (1994: 171) criticize this finding, pointing to a possible spurious result of measurement error due to an indirect measure for the years of post-migration schooling. (Hashmi Khan (1997) and Borjas (1982) calculated post-immigration schooling as total education minus pre-immigration schooling, assuming that individuals attend school continuously from age 6.) Measuring post-migration investments in education directly shows that those with an

educational qualification prior to immigration were more likely to obtain one after immigration. Evidence for such a complementary relationship between education obtained in the sending country and qualifications gained in the host country is also found in the studies of Cobb-Clark et al. (2005) and Van Tubergen and van de Werfhorst (2006).

Unfortunately, empirical research has until now failed to produce a systematic cross-national analysis of VET participation among immigrants, despite the fact that some data sources have been collected recently, at least for Europe. Some figures presented below use the information available from the European Union (EU) Labor Force Survey (LFS) 2003 *ad hoc* module on life long learning (LLL) and show the percentage of native-born, immigrants from EU-15, and immigrants from non-EU countries with and without the citizenship of the host country. A significant advantage of the figures below (over the research summarized above) is the possibility to distinguish between various types of VET: (1) formal, that is, acquired in a regular education system; (2) informal (participation in courses, seminars, conferences, etc. outside the regular education system) job related (e.g., on-the-job training); and (3) informal training attained mainly due to personal and social reasons.

Before going in further detail on the type of training, **Figure 1** shows an overall VET participation rate in EU countries for various population groups in the 12 months preceding the survey. Similar to the OECD (2003) findings, countries with higher VET participation rates are Nordic countries and in these countries, both natives and immigrants profit from the variety of training programs. In Southern European countries, on the other hand, VET participation is very low, which also influences immigrants' VET participation rates. As a rule, non-EU immigrants have lower

participation rates than the charter population, except for Denmark, the UK, Ireland, and the Netherlands. In Sweden, Austria, France, Spain, and Portugal naturalized non-EU immigrants (presumably, immigrants with more pronounced permanent orientation) have higher VET rates than those without citizenship of the host country. Finally, in many countries (Finland, Austria, the Netherlands, Portugal, and Greece) immigrants from other EU-15 countries have even higher VET participation than the native-born.

Figure 2 presents participation rates among immigrants and natives in regular education. We observe quite a similar hierarchy of countries as earlier: higher participation in formal education in Nordic countries, the UK and Ireland and lower in continental and Southern European. Overall, it is evident that in the majority of countries, except for Spain, Luxembourg, Austria, France, and Greece, non-EU immigrants (with and without host-country citizenship) are overrepresented among students. Additionally, the rates of participation in formal education for EU-15 immigrants are higher in Finland, the Netherlands, the UK, and Portugal. It is possible to conclude that many immigrants arrive in European countries in order to study.

Immigrants, particularly from outside EU-15, are also overrepresented among those enrolled in VET outside the regular education system for personal and social reasons, which might also include language courses (see **Figure 3**). Their rates of participation are particularly high in Nordic countries and the UK.

Unfortunately, non-EU immigrants are largely underrepresented among the participants in work-related VET in all European countries, which is evident from **Figure 4**. A situation of naturalized non-EU immigrants is somewhat more favorable in this regard: they seem to

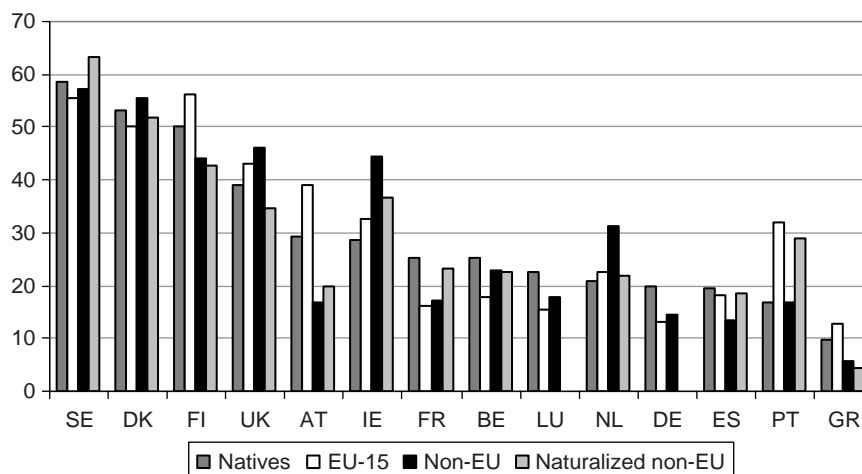


Figure 1 Proportion of individuals, aged 25–64, enrolled in any type of VET in the last 12 months ES – Spain; FI – Finland; IE – Ireland; FR – France; SE – Sweden; GR – Greece; UK – the United Kingdom; BE – Belgium; DK – Denmark; PT – Portugal; NL – the Netherlands; AT – Austria; LU – Luxembourg; DE – Germany. From EULFS LLL 2003, own calculations.

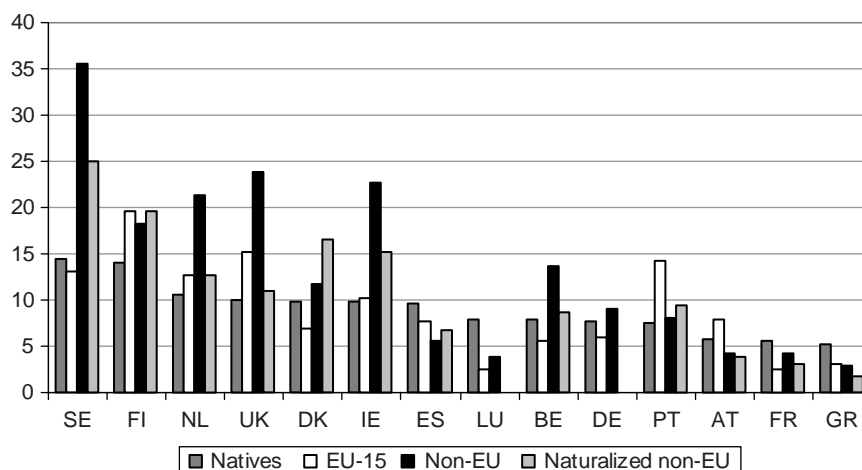


Figure 2 Proportion of individuals, aged 25–64, enrolled in regular education (as student or an apprentice) in the last 12 months. From EULFS LLL 2003, own calculations.

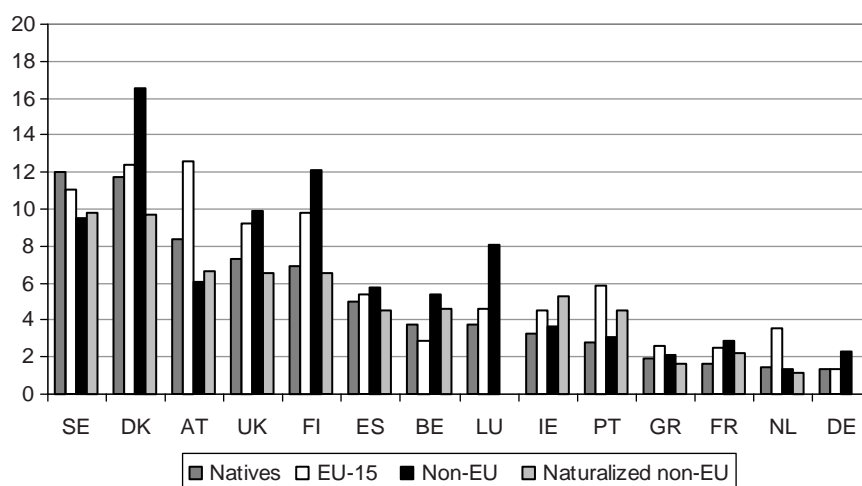


Figure 3 Proportion of individuals, aged 25–64, enrolled in VET outside the regular education due to personal or social reasons in the last 12 months. From EULFS LLL 2003, own calculations.

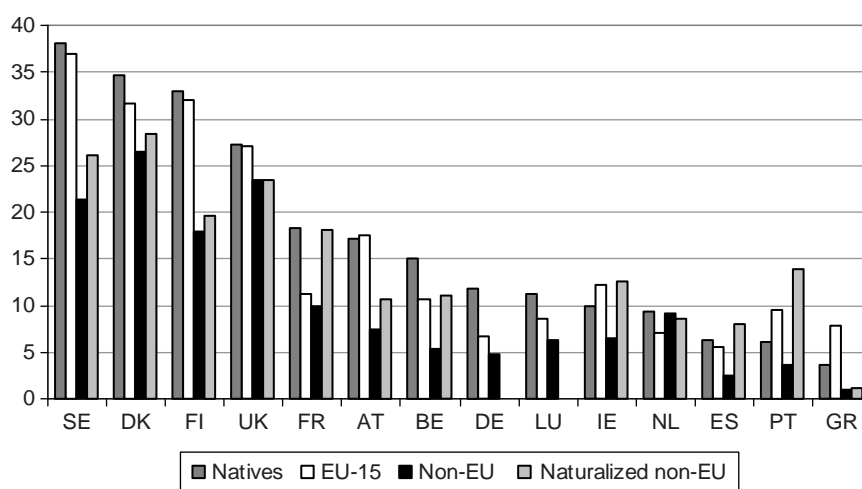


Figure 4 Proportion of individuals, aged 25–64, enrolled in job-related VET outside the regular education in the last 12 months. From EULFS LLL 2003, own calculations.

participate in work-related training more often than their non-naturalized counterparts.

The question arises then, as to whether we can account for the lower levels of work-related VET participation among non-EU immigrants once we control for individual human capital characteristics and for their employment status. A series of logistic regression has been estimated for each country for employed individuals and the results are shown in **Figure 5**. The bars pertain to the log odds of VET participation for non-EU immigrants relative to the group of native-born (the reference category). Log odds below 0 signify that non-EU immigrants have lower propensity for VET participation, other things being equal, and this is true for all EU countries. Since the countries are sorted by the gross effect of work-related VET participation for immigrants outside the EU (a black bar), one could notice quite a similar pattern to the one observed above. The gap between immigrants and the native-born in work-related training is higher in Southern European countries (except for Portugal) and lower in Nordic countries, the UK, and Ireland. Observed cross-national variation in the VET participations rates among immigrants might be related to the institutional characteristics of the analyzed countries with regard to their migration policies and settlement orientations of immigrants, functioning of the educational and training systems as well as the character of their labor markets. Availability of the special migrant-targeting training programs might also serve as an explanation (see the next section).

Controlling for the sociodemographic and human capital characteristics (age, gender, and education) and accounting for the timing of migration (i.e., migrated before and after 1993; a white bar), the gap between natives and immigrants diminishes in Germany, Austria,

France, and the Netherlands (the effect is no longer statistically significant for the latter two). The variable for migration cohort was significant only for Germany and Denmark, signifying higher training participation for immigrants with residence in the host country longer than 10 years.) For the remaining countries, we observe an increase in immigrants' disadvantage with regard to training participation, which signifies that immigrants get less training than comparable (with regard to age, gender, and education) with the native-born. Employment characteristics appear to be a strong predictor for the lack of training among non-EU immigrants, so that the gap between natives and immigrants is further explained in Greece, Spain, Belgium, Finland, Germany, France, Luxembourg, and Denmark (shown by a gray bar). The opposite trend, that is, an increase in the unexplained affect of the immigrant status, is observed in the UK and Ireland. All in all, in most of the countries, the lower propensity for work-related training among non-EU immigrants with socio-economic and employment characteristics similar to those of the native-born remains unexplained. Whether this is due to the factors unaccounted for in the model (e.g., language knowledge or difference in motivation) or due to the reluctance to invest from the side of employers or employees, it needs to be explored in future research.

VET Targeting Immigrants: Qualitative Evidence

It has been shown above that countries differ both with regard to the overall training participation rates as well as the propensity of immigrants to receive VET. Qualitative

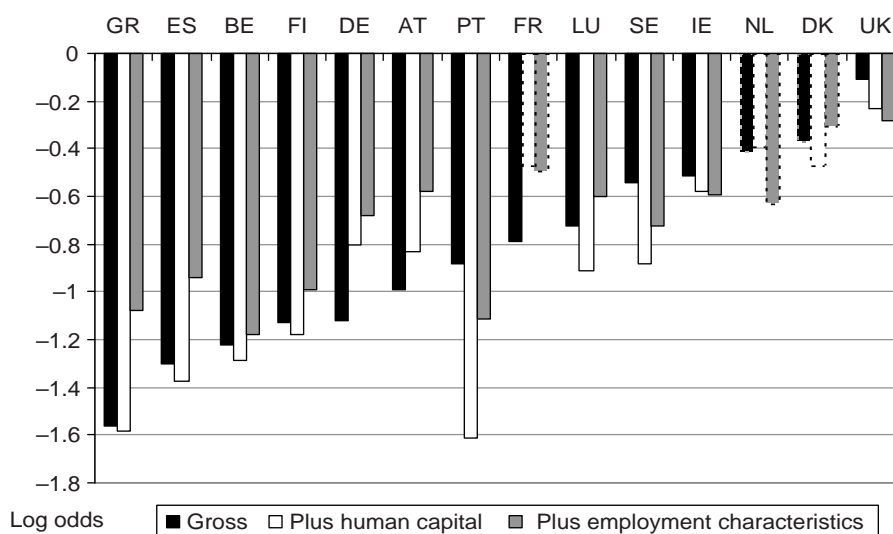


Figure 5 Log odds of enrollment in job-related VET outside the regular education in the last 12 months for employed non-EU immigrants, aged 25–64. The dotted line pertains to the effects that are not statistically significant. From EULFS LLL 2003, own calculations.

studies, some of which are referred to below, give some flavor of VET participation among immigrants but, unfortunately, fail to provide any systematic account for cross-national variation in immigrant training participation.

It is constantly stressed that the lack of the host-country language skills is one of the most important obstacles immigrants face in the new country. Given the difficulties faced by immigrants in successful labor market integration, most OECD countries recently reinforced their integration systems, notably to improve the language skills of newcomers (e.g., France, Germany, and Norway) (OECD, 2003, 2006). Sweden has a strong Swedish for immigrants (Sfi) program, while Denmark also pursues a program of Danish as second language for adult foreigners. The Finnish Broadcasting Company, (YLE), provides Finnish for both Finns and foreigners. In Spain, recent immigration inflows have caused an increase in public and NGO provision of Spanish language and cultural courses. Some countries insist that immigrants participate in classes aimed at improving their integration prospects, for example, in the so-called assimilation and integration training in the Netherlands as a precondition to family reunification and for receipt of social benefits (OECD, 2004).

Historically, in many countries training for newcomers includes not only teaching them the language, but also introducing them to important legal and cultural aspects of the new society or showing them how to operate in the labor market (Wrench, 1998). Some examples include training directed specifically at the needs of a particular organization, where restructuring has affected the existing immigrant workforce. Examples of the firms are ElectroCoat-Genk (Belgium) that provided assistance and training to unemployed miners in the region, which had been in recession since the closure of the local coal training mines. In Austria (iR3 Video International), training was also introduced because insufficient knowledge of the German language in immigrant workers became a potential problem when production began to be restructured. It became no longer sufficient for a production worker to be passively able to understand instructions; instead she/he had to communicate actively within the team, which became the main aim of the course.

Many migrants in the Australian workforce are working below their level of competency due to the fact that their qualifications from the country of origin are not recognized in Australia or because, having been in inappropriate jobs, they have not developed a set of skills that would enable them to become qualified. The lack of recognition of prior learning and inflexibility of learning also contribute to a negative and non-participative view of education and training for migrants (Chandraratna, 1999). In both Canada and New Zealand, the lack of skill recognition was a barrier to gainful employment, though it was not immediately clear, whether recognition of experience

or qualifications was more important. A source of assistance for skilled migrants in this country is the Adult Migrant English Program (AMEP). In Western Australia, its Migrant Pathways Program targeted migrants and refugees with skills in aged care, nursing, engineering, and information technology and provided them with skills recognition and fast-tracked certificate level courses with English language support (The Parliament of the Commonwealth of Australia 2004). CARC (2004) documented some dissatisfaction with programs for English as a second language and the preference of ethnic communities for vocational programs that incorporated English language support. For members of migrant groups, the main purpose of training was to obtain employment and further training is rarely considered once a job has been found. In terms of VET, enrollments and completion were seen to increase if a clear employment pathway was provided, language support was given, cultural issues were addressed, culturally aware teachers were provided, credit for prior learning was given and there was inclusion of work experience programs (Miralles, 2004; Chandraratna, 1999).

VET Pay-Offs

Another body of research deals with the returns to VET and shows that there is a significant and positive relationship between education and training participation and wages (Booth, 1991; Brunello, 2001). Explanations for this are multiple but the issue of causality (i.e., whether training leads to higher wages and not the other way around) has not been yet sufficiently resolved (OECD, 1999). First, it is clear that high-skilled occupations which offer higher training incidence also offer higher wages. Second, people with higher wages are less constrained to purchase training if needed. There are also pieces of evidence for the causal effect of training on wages (e.g., for the UK, Arulampalam and Booth, 1998; for the US, Loewenstein and Spletzer, 1997; Veum, 1995) which pertain to all populations without differentiating individuals by their migration status.

With regard to returns to VET for immigrants, research is also led mainly by economists and much of the empirical evidence stems from Israel, a country with a large immigrant population. Thus, Cohen and Eckstein (2002) analyzing male immigrants from the former Soviet Union arriving in Israel in the 1990s, report a 0.85% growth in wages due to the availability of training in the third year after migration, and 1.4% in the fifth year. They suggest that the effect of training on observed employment and wages is a dynamic phenomenon that takes many years to be realized. Similarly, Cohen-Goldner and Eckstein (2004) confirm these results also for women, in addition pointing to the fact that training is highly valuable for all immigrants but more so for older

and less skilled workers who get the opportunity to search for better jobs. Training, however, seems to have no impact on wages in blue-collar jobs, but has high value for white-collar employment. The above-mentioned studies do not compare immigrants to the native-born population, so it remains unclear whether the two population groups differ with regard to returns to VET.

The study by Seibert (2005) explores whether completion of vocational training in the form of apprenticeship in Germany mitigates disadvantages faced by youth with immigrant backgrounds while entering employment. Results of the occupational registry data from the Employment Study of the Federal Institute for Employment Research (IAB) show that young immigrants holding vocational qualification, with a single exception of Turks, reach similar labor market positions as natives. Turkish men are less able to find appropriate jobs or to enter the labor market in the occupation trained in comparison with Germans, and this is particularly true when labor market conditions are severe. In times of good labor market conditions, Turkish vocational school graduates seem to have similar chances to find appropriate employment.

Evidence from the 2002 Student Outcomes Survey shows that in Australia high rates of participation in VET and successful program completion have not been sufficient to deliver corresponding rates of employment outcomes for some immigrant groups. VET graduates born in non-English speaking countries were more likely to be unemployed than Australia-born graduates or those who arrived from other English-speaking countries. Representing 18% of VET graduates, these youths were only 12.3% among the gainfully employed (Volkoff, 2004). At the same time, those able to find employment were more likely to hold part-time jobs and more likely to have taken longer than Australia-born graduates to find one.

Conclusions

Even though the lack of human capital relevant to the receiving countries has frequently been shown by theoreticians as one of the most serious obstacles to immigrants' labor market integration, empirical evidence concerning educational enrollment and even more so VET participation among immigrants and their children has been scarce until now. Existing studies, largely from the USA and Australia, report underrepresentation of these population groups among students or apprentices and attempt to explain this situation. In Europe, the topic of immigrant education and training has been largely neglected and up until now explored only by Dutch scholars. Even fewer studies (and existing studies tend to deliver unequivocal results) deal with the rate of returns on VET participation among immigrants, compared to countries' charter population. Thus, it is fair to conclude that open questions

remain both with respect to reasons for immigrants' underinvestment into country-specific education and training, their returns to these investments, as well as to the cross-national differences in this regard.

See also: An Overview of Vocational Education and Training; Children of Migrant Populations; Citizenship and Immigrant Education; Race and Ethnicity in the Field of Adult Education; The History of Education: Race and Education; Vocational Education and Training and the School-to-Work Transition.

Bibliography

- Arulampalam, W. and Booth, A. L. (1998). Learning and earning: Do multiple training events pay? – a decade of evidence from a cohort of young British men. Mimeo, Warwick University and Essex University.
- Arulampalam, W., Booth, A. L., and Bryan, M. L. (2004). Training in Europe. *Journal of the European Economic Association* 2(2–3), 346–360.
- Bonacich, E. (1972). A theory of ethnic antagonism: The split labor market. *American Sociological Review* 37, 547–559.
- Booth, A. (1991). Job-related formal training: Who receives it and what is it worth? *Oxford Bulletin of Economics and Statistics* 53(3), 281–294.
- Borjas, G. J. (1982). The earnings of male Hispanic immigrants in the United States. *Industrial and Labour Relations Review* 35, 343–353.
- Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *American Economic Review* 77, 531–553.
- Borjas, G. J. (1990). *Friends or Strangers: The Impact of Immigration on the US Economy*. New York, NY: Basic Books.
- Borjas, G. J. (1994). The economics of immigration. *Journal of Economic Literature* 32, 1667–1717.
- Brunello, G. (2001). On the complementarity between education and training in Europe. *IZA Discussion Paper no. 309*. Bonn: IZA Institute for the Study of Labor.
- Cain, G. G. (1976). The challenge of segmented labor market theories to Orthodox theory: A survey. *Journal of Economic Literature* 14(4), 1215–1257.
- Chandraratna, D. (1999). Taskforce on poverty: Prevalence of poverty among migrant families in Western Australia. *Executive Summary March*.
- Chiswick, B. R. (1978). The effect of Americanisation on the earnings of foreign-born men. *Journal of Political Economy* 86(5), 897–921.
- Chiswick, B. R. (1991). Speaking, reading, and earnings among low-skilled immigrants. *Journal of Labor Economics* 9, 149–170.
- Chiswick, B. R. (2000). Are immigrants favorably self-selected? In Brettell, C. B. and Hollifield, J. F. (eds.) *Migration Theory: Talking across Disciplines*, pp 61–77. New York: Routledge.
- Chiswick, B. R. and Miller, P. W. (1994). The determinants of post-immigration investments in education. *Economics of Education Review* 13, 163–177.
- Chiswick, B. R. and Miller, P. W. (1995). 'The endogeneity between language and earnings: International analyses. *Journal of Labor Economics* 13, 246–288.
- Chiswick, B. R. and Miller, P. W. (2002). Immigrants earnings: Language skilled, linguistic concentration and the business cycle. *Journal of Population Economics* 18, 663–690.
- Cohen, S. and Eckstein, Z. (2002). Labour mobility of immigrants: Training, experience, language and opportunities. *CEPR Discussion Paper no. 3412*. London: Center for Economic Policy Research.
- Cohen-Goldner, S. and Eckstein, Z. (2004). Estimating the return to training and occupational experience: The case of female immigrants. *CEPR Discussion Paper no. 4603*. London: Center for Economic Policy Research.

- Community Affairs Reference Committee (CARC) (2004). A hand up not a hand out: Renewing the fight against poverty. The Senate: Report on poverty and financial hardship. Commonwealth of Australia, Canberra, ACT, Australia.
- Conn-Clark, D., Connolly, M. D., and Worswick, C. (2005). Post-migration investments in education and job search: A family perspective. *Journal of Population Economics* **18**, 663–690.
- Cobb-Clark, D. A., Connolly, M. D., and Worswick, C. (2005). Post-migration investments in education and job search: A family perspective. *Journal of Population Economics* **18**, 663–690.
- Doeringer, P. B. and Piore, M. J. (1971). *Internal Labor Markets and Manpower Analysis*. Lexington, MA: Heath Lexington.
- Duleep, H. O. and Regets, M. C. (1999). Immigrants and human-capital investment. *American Economic Review* **89**, 186–190.
- Dustmann, C. (2000). Temporary migration and economic assimilation. *IZA Discussion Paper* no. 186. IZA, Bonn, Germany.
- Esser, H. (1993). The rationality of everyday behavior. A rational choice reconstruction of the theory of action by Alfred Schütz. *Rationality and Society* **5**, 7–31.
- Esser, H. (1999). *Soziologie. Spezielle Grundlagen. Band 1: Situationslogik und Handeln*. Frankfurt am Main: Campus Verlag.
- Esser, H. (2006). *Sprache und Integration: Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. New York: Campus Verlag.
- Flanagan, R. J. (1973). Segmented market theories and racial discrimination. *Industrial Relations* **12**, 410–431.
- Friedberg, R. M. (2000). You can't take it with you? Immigrant assimilation and the portability of human capital. *Journal of Labor Economics* **18**, 221–251.
- Hashmi Khan, A. (1997). Post-migration investment in education by immigrants in the United States. *Quarterly Review of Economics and Finance* **37**, 285–313.
- Heath, A. and Ridge, J. (1983). Social mobility of ethnic minorities. *Journal of Biosocial Science Supplement* **8**, 169–184.
- Loewenstein, M. A. and Spletzer, J. R. (1997). Belated training: The relationship between training, tenure and wages. Mimeo: US Bureau of Labor Statistics.
- Miralles, J. (2004). *A Fair Go: Factors Impacting on Vocational Education and Training Participation and Completion in Selected Ethnic Communities*. Adelaide, SA: NCVER.
- OECD (1999). *Training of Adult Workers in OECD Countries: Measurement and Analysis*. OECD Employment Outlook. Paris: OECD.
- OECD (2003). *Beyond Rhetoric: Adult Learning Policies and Practices*. Paris: OECD.
- OECD (2006). *International Migration Outlook*. SOPEMI Edition. Paris: OECD.
- Offe, C. and Hinrichs, K. (1977). Sozialökonomie des Arbeitsmarktes und die Lage benachteiligter Gruppen von Arbeitnehmern. In Offe, C. (ed.) *Opfer des Arbeitsmarktes. Zur Theorie der strukturieren Arbeitslosigkeit*, pp S3–S61. Neuwied: Luchterhand.
- Piore, M. J. (1971). The dual labor market. Theory and implications. In Gordon, D. M. (ed.) *Problems in Political Economy: An Urban Perspective*, pp 90–94. Lexington, MA: DC Heath.
- Piore, M. J. (1979). *Birds of Passage: Migrant Labor and Industrial Societies*. Cambridge: Cambridge University Press.
- Seibert, H. (2005). Foreign background still matters. Ethnic differences in the transition process from training to work in Germany. *Paper Presented at the Centre for Public Policy Transitions and Risk: New Directions in Social Policy Conference*. University of Melbourne, Melbourne, VIC, Australia, 23–25 February.
- Taubman, P. and Wachter, M. L. (1986). Segmented labor markets. In Aschenfelter, O. and Layard, R. (eds.) *Handbook of Labor Economics*, vol. II, pp 1183–1217. Amsterdam: Elsevier Science Publishers.
- Van Tubergen, F. and van de Werfhorst, H. (2006). Post-immigration investment in education. *Paper Presented at the Rc-28 Meeting in Nijmegen*, The Netherlands, 11–14 May.
- Veum, J. R. (1995). Sources of training and their impact on wages. *Industrial and Labour Relations Review* **4**, 812–826.
- Volkoff, V. (2004). Tangled threads: Issues faced by non-English speaking background people in vocational education and training. In Bowman, K. (ed.) *Equity in Vocational Education and Training: Research Readings*, pp 120–140. Adelaide, SA: NCVER.
- Wallace, M. and Kalleberg, A. L. (1981). Economic organization of firms and labor market consequences: Toward a specification of dual economy theory. In Berg, I. (ed.) *Sociological Perspectives on Labor Markets*, pp 77–117. London: Academic Press.
- Wrench, J. (1998). *The EU, Ethnic Minorities and Migrants at the Workplace*. London: Kogan Page.

Indigenous People and Vocational Education and Training

R G Schwab, The Australian National University, Canberra, ACT, Australia

© 2010 Elsevier Ltd. All rights reserved.

Indigenous People in the World

Any discussion of Indigenous people and vocational education and training (sometimes referred to as technical and vocational training) must begin with definitions. Though Indigenous might seem a common sense term, it is in fact controversial in the international context. Some countries deny the existence of Indigenous peoples because to acknowledge them would be to open up difficult issues of what defines a nation; instead, they refer to tribal peoples or ethnic minorities. In contrast, some Indigenous groups strongly assert their aboriginal status and refer to themselves as first nations, or by their tribal or language names, claiming or at least implying some measure of sovereignty. Clearly, the official recognition of these groups has political and economic implications that some Indigenous peoples are keen to press and some governments are anxious to avoid. Consequently, a precise definition of Indigenous is somewhat problematic. Still, a working definition of Indigenous is important for understanding the complexities of how Indigenous people engage with programs and policies related to vocational education and training (VET).

A set of characteristics or features of Indigenous peoples underpins United Nations and many other international organizations' and agencies' definitions of Indigenous peoples and they provide a useful framework for a definition:

- priority in time, with respect to the occupation and use of a specific territory;
- the voluntary perpetuation of cultural distinctiveness, which may include the aspects of language, social organization, religion and spiritual values, modes of production, laws, and institutions;
- self-identification, as well as recognition by other groups, or by state authorities, as a distinct collectivity; and
- an experience of subjugation, marginalization, dispossession, exclusion, or discrimination, whether or not these conditions persist (Daes, 1996).

Even with a relatively clear definition, counting the number of Indigenous people in the world is difficult. Not only are government definitions problematic, but Indigenous people themselves also sometimes disagree about who is Indigenous. In some countries, Indigenous groups define membership according to their own criteria, for example, by historical listing on a tribal roll or by blood quantum. Further adding to these difficulties is the fact that many countries with Indigenous populations are developing countries with administrative systems that struggle to locate

and accurately count citizens who live in remote or isolated regions as many Indigenous groups do.

Box 1 The Sámi

The Indigenous people of northern Europe (Norway, Sweden, Finland, and the Kola peninsula of Russia) are known as the Sámi (Saami, Sami, Same). Originally the Sámi were reindeer herders living in the sparsely settled remote north. Sámi language and culture were oppressed by various governments and over time many Sámi hid or denied their identity and language. Consequently, accurate population estimates are difficult though some suggest a population of no more than about 100 000 today (with only about 35 000 language speakers). Most of the Sámi people in Finland live in the far north of the country in an area referred to as the Sámi homeland. Established in 1977, the educational center of the Sámi region provides basic vocational education in various fields, using Finnish and Sámi as languages of instruction. As members of the European Union, Sweden and Finland have received adult education funding that promotes, among other things, a range of Sámi vocational interests including: Sámi handicraft and reindeer husbandry, Sámi film education, and Sámi language reading and writing.

From Lund, S. (2000). *Adult Education and Indigenous Peoples in Norway*. Hamburg: UNESCO Institute for Education; and Ruoppila, S. (2006). *Finland: 'Feeding in' and 'Feeding out', and Integrating Immigrants and Ethnic Minorities. Study of National Policies*, European Commission, DG Employment, Social Affairs and Equal Opportunities.

Given the difficulties, it is only possible to estimate, and the current best estimates of the size of the world's Indigenous population range between 300 and 500 million, comprising over 5000 language and cultural groups spread across more than 70 countries (United Nations Permanent Forum on Indigenous Issues). Most of these people are in Asia and to a lesser extent in Africa and Latin America. It is also important to note that in all these regions, and even in highly industrialized countries like Canada, the United States, and Australia, Indigenous people are typically among the most poor and disadvantaged in their home countries, often with relatively lower levels of education, employment, and health. These demographic features are important for understanding the challenges of providing VET to Indigenous populations.

VET and Socioeconomic Development

In the most simple terms, VET can be defined as education for the world of work. Yet, it is more than this. Indeed,

VET involves manual and technical training – invariably with an aim toward developing skills and knowledge for practical application in various livelihoods – but it also may include integral elements of general education such as literacy and numeracy and basic life skills. In many parts of the world, Indigenous people have had little or no access to basic education and where they have attended school, it has often been only through the primary level. Consequently, VET for Indigenous people in most cases is not only about technical or vocational training, but it is also a means to new insights, knowledge, and understanding about the outside world and the key to participation in the wider society.

Although there is obviously great variation among the circumstances of Indigenous peoples of the world, VET is for many a critically important means of socioeconomic development. As a sector in education, it is arguably the most relevant to capacity development on both the individual and national levels. It is typically flexible, and in terms of delivery and content, can be modified and adapted to varying needs and academic backgrounds of students. In countries with robust and growing economies, or those where economic growth is stagnant or slow, VET is often the means by which Indigenous people gain access to employment and its associated benefits. In many developing countries, especially those where poverty is widespread, VET is often explicitly conceived of as one of the most important avenues to poverty alleviation among Indigenous peoples.

Models of VET Delivery to Indigenous Peoples

VET delivery varies greatly and differs significantly from general education where delivery is typically through a state or other school and later through a tertiary or higher education system. While some countries do fold VET into schooling at the secondary level, it is more commonly offered outside the general-education system. The delivery structure is highly variable. Formal programs might involve specialized technical and vocational colleges or polytechnics, enterprise or industry-based training, training provided or subsidized by nongovernmental organizations (NGOs) or church-based organizations, private providers paid by government, industry, and sometimes through formal on-the-job apprenticeship programs. In many countries and in some circumstances, the provision of VET is relatively informal, and thus more akin to adult education and lifelong learning programs. Such variability has some advantages in terms of flexibility but can also result in fragmented and uncoordinated delivery. For Indigenous people – especially those in remote regions far from educational services in capital or regional centers – the flexible delivery of VET is often critical.

Box 2 First Nations Technical Institute, Canada

The First Nations Technical Institute (FNTI) is an Indigenous owned and controlled postsecondary institute located on Tyendinaga Mohawk Territory in eastern Canada. The Institute was created in 1985 and serves 300 postsecondary students and adult learners from across Canada who may study from a range of courses. The philosophy and approach of the institute emphasizes holistic, culturally based learning. One area of study is Indigenous community health through which students explore Indigenous knowledge and philosophies as they relate to human health and the natural environment. The program is rooted in Indigenous knowledge and exposes First Nations health professionals to cultural teachings which broaden their understanding about health, healthcare, and environmental and human health issues facing First Nations communities.

Challenges for Building Effective VET Programs for Indigenous Peoples

Universal rights to education for Indigenous people are prominent in the United Nations Declaration on the Rights of Indigenous People, adopted in November 2007, and vocational education receives specific emphasis. The declaration states that Indigenous individuals, particularly children, have the right to all levels and forms of education of the state without discrimination (Article 14). Further, the other articles of the declaration refer to the important role that education plays in empowerment (Article 17.2) and promote vocational training and re-training as one of the means to social and economic improvements (Article 21.1). The reality, however, is that realizing these rights is, for most Indigenous people, still a distant promise. The United Nations' Special Rapporteur wrote that “education for Indigenous peoples would seem to be the ‘ugly duckling’ of national education programmes and in general to be assigned low priority and inadequate budgets at the national level” (Stavenhagen, 2005: 16).

Yet, even where governments have the political will to put the UN Declaration into practice, it is a complex challenge. There are enormous variations in the geographic, political, and socioeconomic conditions in which the world's Indigenous peoples live and no single solution to serve their needs. While there certainly are unique challenges for some specific groups, for example, in countries where the existence of Indigenous people is denied, there are also a range of common and interrelated challenges that deserve attention.

Often, as a function of history, many Indigenous peoples live in remote and isolated communities, with limited infrastructure, far from urban or regional centers where mainstream services are provided. This has several implications. Employment options are often few, given the distance from the labor markets. Service provision is sometimes poor, again simply because of the disadvantage of distance, and

Indigenous people may have less access to even basic levels of education, housing, communication, health, and other social services considered essential in the developed world. The result for many is not just a geographic marginalization, but – and this is true for Indigenous people who live in developed countries in urban settings as well – also a political and socioeconomic marginalization that locks Indigenous people into a role as the most economically, politically, and socially disadvantaged citizens within their home countries.

Consequently, many Indigenous face precarious social and economic circumstances that affect their ability to access resources and participate fully in the social and economic development of their countries. Compounding this is the deeply rooted discrimination and racism that many Indigenous people face in their interaction with the dominant society. This marginalized position also correlates with not just higher levels of unemployment and lower levels of education, but also higher levels of alcohol and other substance abuse, social and family dysfunction, increased rates of suicide (particularly among the young), and generally high levels of morbidity and mortality.

In most countries, education is delivered in the official language of the state. There are obvious political reasons why this should be so. For example, a government may promote a single language in order to consolidate a national identity. But there are also quite reasonable economic considerations. Policy and curriculum documents and teaching materials in multiple languages increase costs. In developing countries, which is where most of the world's Indigenous peoples reside, the cost of education is a significant issue that imposes real constraints. Given that Indigenous people by definition speak minority languages, finding qualified bilingual teachers for VET instruction is often difficult; similarly, developing curriculum and instructional materials in those languages is, in most cases, prohibitively expensive. Consequently, it is often the case that either only Indigenous people who are literate in the national language can effectively engage with the VET materials or the materials must be locally adapted so they can be used with individuals who are illiterate in the national language or literate in the local Indigenous language. Further, education in the nonofficial language is likely to limit access to mainstream labor markets, so even where nonofficial materials are available, the economic payoff at the national level may not be adequate to justify the investment.

Another challenge in the provision of VET relates to the potentially negative perceptions of or deep suspicions about education and training many Indigenous peoples hold. Many Indigenous people feel a strong ambivalence to Western education, of which VET is a part. On the one hand, it is universally cast as the single best tool for empowerment and poverty alleviation – education and training provides skills for employability and engagement

with the national economy. Yet, on the other hand, many perceive education as a continuation of colonialism or forced assimilation; in many countries, education was the method of civilizing Indigenous people, diluting their language, undermining their traditional livelihoods, and destabilizing their culture.

Box 3 Safety in the learning environment

Research into factors that facilitated successful vocational education and training among Māori people in New Zealand identified safety in the learning environment as one of the most important.

The “learning environment” refers to the total physical and philosophical ethos of the organisation. It includes the spaces for formal and informal learning as well as sites of learning that extend beyond the traditional classroom setting. While each of the providers interviewed emphasised certain values that underpin their organisation, a paramount goal of all of the providers was “to provide a place of safety” for their learners. “Safety” refers to the physical and psychological elements of well-being, such as pastoral care and cultural safety. (Skill New Zealand-Pūkenga Aotearoa 2001)

In both developed and developing countries, student and institutional funding and support for VET is a major challenge. In many countries, vocational education sits awkwardly outside mainstream primary and secondary education and tertiary studies. Typically, there is a school sector to provide general education to children and a tertiary sector for adults. Vocational education, however, is often far less distinct as a sector with myriad providers and delivery is often uncoordinated, fragmented, and of variable quality. VET delivery may come directly from government, or from NGOs, industry, employers, private providers, local communities, religious institutions, and more. An answer to the question of who is responsible for funding and supporting the VET needs of Indigenous people can vary not only from country to country but from region to region or community to community.

All these challenges make planning for and evaluation of vocational education and training for Indigenous people extremely difficult. To develop policies and programs for the delivery of VET to Indigenous people, data on a wide range of important variables need to be collected. For example, local needs, student readiness, student performance, enrollments, completions, staffing, and facilities are but a few. In addition, for these data to be of value, they need to be accurate and reliable and collected in a systematic fashion. In a developed country with established administrative data systems, this is a challenge; in a developing country where administrative systems are incomplete or nonexistent, the tasks of planning and monitoring the delivery of VET to Indigenous people is exponentially more difficult.

Key Strategies for Advancing Indigenous Engagement with VET

Increasing numbers of countries are articulating commitments to enhance the vocational and life skills of their Indigenous citizens and there are many examples of success. Still, there would be little value in attempting to provide a list of best practice in the provision of VET to Indigenous peoples around the world. The field of VET shifts and changes with new innovations and emphases, governments come in and out of power and reorganize, abolish, or reinvent policies, departments, and institutions, and the political relationships between Indigenous people and their home nations move in and out of the national and international spotlight. Yet, there are some clear strategies that have been implemented in various countries that have resulted in tangible advances for Indigenous peoples:

1. *Recognition of Indigenous culture.* The most fundamental factor in successful VET for Indigenous peoples is recognition by the national (and regional and local) government that Indigenous people have unique histories, traditions, and cultures and that a one-size-fits-all approach to training rarely works. Increasingly, Indigenous people are asking for VET that takes into account Indigenous knowledge systems and intellectual property and incorporates culturally appropriate modes of delivery and instruction.
2. *Partnerships and consultation.* Extending the point above, it is clear that many countries attain gains in participation and outcomes when programs are built in partnership with Indigenous people. That might involve a government- or industry-sponsored Indigenous consultative body at the national level and/or regional or local discussions about what local people feel they want and need in terms of VET. In some countries, Indigenous-controlled VET institutions have been successful, particularly among students who would not otherwise engage in mainstream education. Participatory, bottom-up development of VET enables the delivery of training that reflects local interests and needs and greatly increases the quality of training outcomes. It also appears to be the most effective means of overcoming the ambivalence many Indigenous people feel about education.
3. *Blending vocational education with more general education and life skills.* While employment is almost invariably the desired outcome of VET, for Indigenous people in very remote areas, where labor markets are limited or nonexistent, that will not always be possible. In many countries, VET takes a more informal approach and merges with adult education in the principle of lifelong learning. In some countries, VET is used as a vehicle for delivering basic-skills training (including literacy and numeracy) related to health, nutrition, parenting, communication, governance, and dozens of other critical life skills. In this way VET can also provide life skills and assist people in gaining the confidence to participate in employment and to engage as citizens with the wider society.
4. *VET and Indigenous knowledge.* For many Indigenous people, vocational employment has meant training to work in extractive industries on or near Indigenous lands. This has certainly contributed to the deep ambivalence some feel about VET. In recent years, however, there has been a growing interest in sustainability, and Indigenous lands, often in remote areas, are sometimes among the most ecologically intact and environmentally significant and sensitive within a country. There is increasing awareness across the world of the importance of protecting this natural heritage coupled with expanding recognition of the economic value of Indigenous knowledge of plants and of strategies for land and resource management. Shaping VET to enhance and protect this knowledge and to provide skills to enable local people to work effectively in these areas shows great promise in many parts of the world.
5. *Articulation with national training bodies and national qualification frameworks.* In some countries, Indigenous people express anxiety about the degree to which the VET they are offered is up to standard. This is particularly true in countries with Indigenous populations in remote areas and where training is sporadic, often poorly coordinated, and unmonitored. In some developed countries, Indigenous issues are a major focus of national training bodies and extra efforts are made to ensure Indigenous training is integrated into a national quality framework. The result is typically a more consistent, higher-quality VET system.
6. *Reducing barriers to access.* Many studies of Indigenous participation in education identify relatively simple barriers to access. A lack of child care, for example, often stops women from participating in education or training. Similarly, Indigenous trainees sometimes struggle with structured learning within a foreign system; counselors, tutors, and student-support strategies often make the difference between a person staying in a training program and dropping out. These can be extended to something akin to pastoral support and can be especially effective in the context of workplace training where an employer provides additional training or exposure to knowledge and skills for life (e.g., literacy or health) and not just those needed to perform workplace tasks.
7. *Flexible delivery.* In many Indigenous communities, patterns of pastoral mobility or hunting, agricultural seasons, and ceremonial cycles can disrupt training. Given the relatively higher levels of mortality and

morbidity among Indigenous people in most countries, ill-health and the need to attend funerals often result in students falling behind and eventually withdrawing from training. The most strategic response to these realities is to build VET systems that can accommodate or incorporate such cultural and local needs. Similarly, training that is locally based, hands-on, and culturally sensitive, greatly increases the quality of the outcomes.

Box 4 Factors leading to improved outcomes in vocational education and training for Indigenous people, Australia

A systematic review of research found identified seven factors leading to improved outcomes in vocational education and training for Indigenous Australians:

- Community ownership and involvement.
- The incorporation of Indigenous identities, cultures, knowledge, and values.
- The establishment of true partnerships.
- Flexibility in course design, content, and delivery.
- Quality staff and committed advocacy.
- Extensive student support services.
- Appropriate funding that allows for sustainability.

From Miller, C. (2005). *Aspects of Training That Meet Indigenous Australians' Aspirations: A Systematic Review of Research*. National Centre for Vocational Education and Training, Australian Government, Adelaide.

Conclusion

A vast majority of Indigenous peoples in the world today remain committed to their languages, their cultures, and their futures as distinct peoples. Many struggle for recognition and for equity in law, health, education, economic participation, and social justice. They often face challenges of racism, discrimination, and socioeconomic marginalization. Yet, for Indigenous people everywhere, VET is one vital path to the future, promising not just access to employment and income, but also increased opportunities for participation in the home country and in the wider world.

See also: Education Production Functions: Evidence from Developing Countries; Learning in a Cross-Cultural Perspective; Learning Outside of School; Planning and Policy Development for Technical Vocational Education

and Training Systems; Returns to Education in Developing Countries; Teachers in Developing Countries; The Education of Indigenous Students.

Bibliography

- Daes, E.-I. A. (1996). *Working Paper on the Concept of 'Indigenous People.'* Prepared for the Working Group on Indigenous Populations. UN Doc E/CN.4/Sub.2/AC.4/1996/2.
- Stavenhagen, R. (2005). *Human Rights and Indigenous Issues: Report of the Special Rapporteur on the Situation of Human Rights and Fundamental Freedoms of Indigenous People*. Commission on Human Rights, Economic and Social Council, the United Nations, New York.

Further Reading

- Lund, S. (2000). *Adult Education and Indigenous Peoples in Norway*. Hamburg: UNESCO Institute for Education.
- Miller, C. (2005). *Aspects of Training That Meet Indigenous Australians' Aspirations: A Systematic Review of Research*. National Centre for Vocational Education and Training, Australian Government, Adelaide.
- Noonan, P., Burke, G., and White, P. (2004). Policy developments in VET: Analysis for selected countries. *CEET Working Paper 54*, Monash University – ACER Centre for the Economics of Education and Training, Melbourne.
- Ruoppila, S. (2006). *Finland: 'Feeding in' and 'Feeding out', and Integrating Immigrants and Ethnic Minorities. Study of National Policies*. European Commission, DG Employment, Social Affairs and Equal Opportunities. http://ec.europa.eu/employment_social/spsi/docs/social_inclusion/experts_reports/finland_2_2006.pdf (accessed August 2009).
- Schwab, R. G. (2001). VET-in-school for indigenous students: Success through 'Cultural Fit'. *Research to Reality: Putting VET Research to Work. Proceedings of the Australian Vocational Education and Training Research Association (AVETRA) Conference*, 4th, Adelaide, South Australia, 28–30 March, 2001. http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/19/2e/7d.pdf (accessed August 2009).
- Skill New Zealand-Pūkenga Aotearoa (2001). *Sharing for Success: Good Practice and Issues for Māori Education*, Skill New Zealand - Pūkenga Aotearoa, Wellington.
- UNESCO Institute for Statistics (2006). *Participation in Formal Technical and Vocational Education and Training Programmes Worldwide: An Initial Statistical Study*. Bonn, Germany: International Centre for Technical and Vocational Education and Training.
- United Nations (2008). *United Nations Declaration on the Rights of Indigenous Peoples, Resolution adopted by the General Assembly, 13 September 2007*, The United Nations, New York.

Relevant Website

<http://www.fniti.net> – First Nations Technical Institute.

Persons with a Disability and Vocational Education and Training

F Ferrier, Monash-University-ACER Centre for the Economics of Education and Training, Australia

C S Smith, Chris and Joy Selby Smith Pty Ltd., East Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Vocational education and training (VET) is an important avenue for improving work-related skills. VET contributes to organizational effectiveness, productivity growth, and national competitiveness. It also plays an important role in assisting the less-advantaged members of society, including those who are disadvantaged due to a disability. The balance between these two different emphases is one of the tensions in VET.

There are responsibilities at the societal level for ensuring that VET provides effective assistance for students with a disability. The many significant stakeholders in VET (e.g., students, training institutions, governments, and employers) all have a role to play, but the primary responsibility rests with governments and with VET providers, especially those supported by public funding (Smith and Ferrier, 2004).

In making decisions about whether to participate in, or support, VET (and when, where, and in what form), all VET stakeholders are influenced by the costs they face and the benefits they expect to receive. From a societal perspective, the relationship between the total costs of provision and the total benefits from individuals participating in VET (including individuals with a disability) is crucial for decisions about resource allocation. Overall benefits need to exceed costs, and the more they do so, the stronger becomes the case for societal action. Australian studies suggest that increasing the participation in VET of people with disabilities would result in substantial economic gains, that is, societal benefits would significantly outweigh costs (Dockery *et al.*, 2001). When overall benefits exceed costs, but costs exceed benefits for individual enterprises or providers (e.g., small providers), then public authorities can tilt the balance by subsidizing costs, augmenting the benefits received, or some combination of the two.

Definitions of disability vary across nations and education sectors, but distinctions are commonly made between intellectual disabilities, learning disabilities, and physical disabilities. For educational institutions, a primary consideration is whether, and to what extent, the disability impacts on the ability of students to participate in academic programs and to complete them successfully. Those with multiple disabilities are recognized to be particularly disadvantaged. The adverse implications of a disability can also be cumulative: for example, poor literacy and numeracy outcomes at school contribute to below-average learning prospects in VET and fewer opportunities for satisfactory employment.

The ability of students with disabilities to participate successfully in educational programs can also be affected by the impact of their disabilities on aspects of their day-to-day living, such as housing, transport, health, or personal needs. Difficulties in these areas can have a negative impact on the individual's interest in learning, his/her ability to attend and participate in classes, to access and use available supports, and to participate in work placements. They can also affect the student's subsequent capacity to obtain employment and to advance at work. The broader society shares responsibility with governments and education authorities and providers for providing assistance in such areas. There is also an important role for industrial bodies, including employers and unions, in assisting people with disabilities when changes in the workplace have a disproportionately adverse impact and in making transitions from education to employment.

International Issues and Approaches

Across international settings, there are many common issues and concerns regarding students with disabilities in postsecondary VET. Common goals, ideas, and types of initiatives to address inequalities, inadequacies, and inefficiencies are complemented by some diversity in policy and practical approaches reflecting local contexts and objectives.

A wealth of studies, reports, and discussion papers provides evidence of substantial attention in many settings to the creation of more accessible and inclusive education and training systems; to improving the range, type, and effectiveness of programs and supports for students with disabilities; and to supporting students with disabilities to achieve better outcomes from their studies. Governments and policymakers are wrestling with similar questions concerning how best to support and encourage people with disabilities to participate in the labor market and to provide VET programs that will meet their needs, prepare them for employment, and support their transition from study to work (e.g., Kim and Zirkle, 2006; Zolkowska *et al.*, 2002; LSC, 2005; Kate *et al.*, 2003). In these matters, students with disabilities are considered as a broad group and funding models and cost issues receive special attention.

At the local level (education institutions), attention is centered on how best to support students with disabilities to complete their studies successfully and to obtain the

outcomes they seek. An important consideration is the capacity of staffing and other essential resources, including both general and tailored supports, to assist students with disabilities. In addition, education administrators and teachers are making efforts to gain a greater understanding of the needs of students with disabilities, including those with particular types of disabilities, and to improve learning arrangements and teaching practice. The international literature highlights the evolution of knowledge and practice around the impacts of particular types of disability, such as a learning disability, or a sight or hearing impairment. Many studies explore how best to assist students with specific types of disabilities, as well as students with disabilities in general.

At both government and institution levels, VET is rarely seen as an end in itself, but as a pathway to meaningful employment and greater social participation for people with disabilities. In the UK, for instance, the Learning and Skills Council (LSC) has adopted an approach to further education (FE) for students with disabilities that is guided by the concept of learning for living and work (LSC, 2006). The goal is to drive up participation and to enable people with disabilities to achieve ambitions including improving their lives, strengthening their career opportunities, and developing their skills. The Council notes that this will enable economic participation, the development of self-confidence, and help people of all ages to live autonomously and contribute to the communities in which they live (p. 5).

However, there is also an awareness that recruitment decisions by employers are made outside VET systems and that student success in study does not necessarily lead to labor-market success. Focusing on Europe, Corbett *et al.* (1999) note that people with disabilities comprise a peripheral labor force and tend to be seen as having few marketable skills. Employer attitudes are regarded as a major barrier to employment for people with disabilities. However, overcoming employer attitudes is regarded as only part of a broader struggle by people with disabilities to gain recognition of their potential (e.g., Harrison and Barnett, 2001; Kim and Zirkle, 2006; Waite *et al.*, 2006). A recent longitudinal study in the UK emphasized that while aspirations of young people with disabilities in their mid-teens are very similar to those of their peers without disabilities, by their mid-20s few have been able to reach the levels of education or to gain the employment that they sought and expected. The result can be disaffection and depression. The title of the report 'Frustrated ambition' appears well chosen (Burchardt, 2005).

Costs and Funding

Not all students with a disability require expensive supports to participate successfully in education and training. The extra costs that they and their enrolling institutions can incur vary greatly, depending on the nature and

extent of the disability. In some settings, VET providers have no indication in advance of the number of students who will be seeking disability supports, or of the nature of the supports that students will request. Consequently, they can experience difficulties in predicting the extra costs they will incur from year to year. These difficulties may be especially important where VET providers are small and have very limited resources to draw on in years when the number of students seeking disability support increases beyond budget expectations.

Andrews and Smith (1992) divided the supports required by students with disabilities into three groups: relatively low-cost support requirements, normally provided by educational institutions as student services; more costly supports provided by the institutions, which they consider appropriate for special funding (e.g., large print materials, special furniture); and support requirements related to unique needs for individual students (e.g., note takers or interpreters). They estimated that 6%, 24%, and 70% of total (recurrent) expenditure on supports for students with disabilities was in the three respective groups. High-cost supports were less frequently required, but consumed the largest proportion of the budget. While these estimates imply much higher training costs for VET students with disabilities, they are average costs: marginal costs are likely to be lower (Dockery *et al.*, 2001). Similarly, in relation to workplace training, the (extra) costs of accommodating an employee with a disability may be similar, whether or not the employee is engaged in training.

Two distinct funding models are widely used to allocate support to students with disabilities in different sectors of postsecondary education: funding allocated to students; and funding allocated to institutions (Devlin, 2000). Funding allocated to individual students to cover the costs of additional supports (perhaps on an entitlement basis up to an agreed limit) empowers students to make choices about supports and education providers. However, such funding can be difficult to administer; can place stress on students by requiring them to complete forms and to locate, price, and secure services; and can disadvantage students when negotiating the purchase of equipment or services. Providing funding directly to educational institutions has advantages, particularly for large institutions, which can plan services, adjust flexibly to changing circumstances, and take advantage of their bulk purchasing power.

In Australia, which is a federal country of six states and two territories, current funding arrangements for VET usually comprise a mixture of base funding to institutions and additional funding for special purposes, such as where students have particularly expensive support needs. However, within various VET institutions, a rather different emphasis can be placed on various elements in the funding mix. There are also some differences in the ways in which funds are bid for and allocated (for further details, see Smith and Ferrier, 2004: vol. 2).

Funding arrangements for students with a disability can be assessed using four criteria: portability (i.e., whether funding follows the student from one educational institution to another); whether additional funding reflects the actual cost of providing the support required for each student; administrative efficiency (i.e., whether the funding arrangements limit administrative costs and devolve administration as close to the client as possible); and respect for the autonomy of educational institutions (Devlin, 2000). An additional consideration is whether the funding arrangements offer incentives for VET providers to enrol and to support students with a disability (Buys *et al.*, 1999; Barnett *et al.*, 1996). An institution which gains a reputation for excellence in meeting student needs may encourage increasing enrolments and thus costs: the incentives for providers can be perverse (NBEET, 1994). As governments move from a supply-based to a more demand-based VET system, as in Australia and the UK, in recent years (Smith *et al.*, 2001), it has also become more important to ensure that the interests of those with limited choices, including people with disabilities, are safeguarded.

Some students with disabilities can take longer than other students to reach the required level of knowledge and skills. These students may be disadvantaged where there is little or no flexibility in systems and programs to accommodate their need for additional time (e.g., NEADS, undated; Waite *et al.*, 2006). This is not solely an issue about teaching practice, but also has implications for the methods used to allocate program funding and for the timeframes that are used to evaluate program outcomes and effectiveness. The Australian study (Smith and Ferrier 2004) noted that while students with a disability often require extra time to complete a VET qualification, this was not often recognized in public funding arrangements. Barnett (2004) indicates that resource-allocation processes need to take a longer-term view of outcomes for students with disabilities. Waite *et al.* (2006) suggest a stronger focus on lateral (rather than linear) progress of students with disabilities.

Participation

A common challenge in obtaining a comprehensive picture of participation in VET by people with disabilities is that data can be unreliable. Where self-reporting of a disability is the basis for data collections, enrolments can be understated, for there may be disincentives to disclosure. For instance, students may fail to report a disability for fear that they will be treated differently, offered fewer opportunities, or will encounter negative attitudes about their ability. They may also fail to report a disability because they do not expect to need any accommodations or additional support to complete their study successfully. A further difficulty is inconsistencies in definitions of disabilities

used by different education systems and sectors and by a range of other agencies, for instance, between health and education systems and between schools and post-compulsory-education institutions.

Despite such difficulties, it is widely acknowledged that compared with other students, those with disabilities tend to leave school early and when they enrol in vocational programs, they tend to choose courses leading to lower level qualifications – or to qualifications outside formal-recognition frameworks. For instance, Australian data show that in 2000, only 30% of VET students with disabilities had successfully completed a school-Year-12 qualification, compared with 43% of all VET students (NCVER, 2002); and in 2003 more than half (55%) had left school at or before the end of Year 10, compared with 40% of all VET students (NCVER, 2005).

A study from the UK notes that the scope and level of aspirations of 16-year-olds with and without disabilities are very similar. Differences are apparent only for some small groups: those with mental health problems; with severe impairments or more complex needs; or who became disabled later in childhood (Burchardt, 2005: xi). This represents a considerable change from the past: the study cites a 1958 study of young people which found that the proportion of those with disabilities aspiring to semi-skilled or unskilled jobs was six times the proportion of those without disabilities. Rising aspirations may be one of the reasons the UK's LSC reports growth in participation in FE and work-based learning by students with disabilities. Other contributing factors are the rising incidence of learning disabilities as diagnostic procedures improve and the aging of the population and the associated rise in the rate of acquired disabilities. In 2001/02, 336 537 learners in FE and work-based learning identified themselves as having a learning difficulty and/or disability. In 2004/05, the number had risen substantially to 477 417. The LSC expects growth in participation to continue as the population ages and as more students in compulsory education aspire to access FE and work-based learning (LSC, 2006).

Supporting Students with Disabilities

Addressing the needs of VET students with a disability is complicated by the range and diversity of disabilities. In Australia, “of all disability types, VET students with hearing or vision disabilities have the highest probability of passing assessed subjects, and the highest employment outcomes, whereas VET students with intellectual or learning disabilities have the lowest, and well below the average level for all VET students with a disability.” Thus the type of disability does matter and “different improvement strategies may be necessary for different groups” (NCVER, 2005: 6).

Staff development is a key factor in improving much of the acknowledged uneven quality of education-and-training

provision for students with disabilities. The international literature acknowledges that many staff in institutions offering VET lack understanding of the impact of disabilities and knowledge about effective and appropriate teaching practices and other strategies to support such students. Many studies thus explore ways to respond to the impacts of particular types of disability on students' learning and participation, and to identify the administrative and teaching practices that will be most effective in assisting students with a disability (e.g., Perin, 1990; Ofsted, 2007; Onnela, 2003).

In the UK, a carrot-and-stick approach is being taken to raising the quality of provision for students with disabilities. Innovative delivery, the sharing of ideas, and good practice across the system (including through collaboration) are being promoted, performance measures are being implemented, and intervention occurs where there are no improvements. Centers of excellence are being established as trailblazers to lead system-wide improvements. The LSC also notes that the range and complexity of learning difficulties and disabilities will demand the development of both generalist and highly specific specialist training opportunities (LSC, 2006). Staff-development initiatives include initial training pathways for specialist learning support staff and teachers of learners with learning difficulties and/or disabilities.

The creation of inclusive systems and the adoption of inclusive practices can improve the student experience in VET and the outcomes they achieve (e.g., LSC, 2005). From the US Hart *et al.* (2006) describe a model of inclusive postsecondary education for students with disabilities that is based on collaboration across education sectors and between multiple agencies. It has a strong focus on the individual student's aspirations; and on the provision of services and supports to enable their achievement. Individualized services include an educational coach, tutors, technology, and natural supports. The focus is on establishing a student-identified career goal that directs the course of study and employment experiences (such as internships, apprenticeships, or work-based learning). The authors envisage a collaborative approach between agencies and sharing of costs.

Similarly, a UK study of work-related learning for students with learning difficulties (Waite *et al.*, 2006) notes the need for an increased emphasis on the value of educational diversity and greater opportunities for this group of learners to participate in education in ways that are meaningful and enable them to participate fully as adult citizens. One specific difficulty identified in the study is the current reliance on competence-based testing, which can reinforce students' experience of failure.

Developing an inclusive VET system that "enables individuals to reach their full potential and maximise their abilities" (p. 106) will also lead to the creation of a more inclusive society (Barnett, 2004). However key attitudinal,

resource, knowledge, skill, and systemic barriers affect people with a disability. Good-practice exemplars and learning the lessons of past experience can contribute to the dismantling of such barriers and improve policy and practice (Kate *et al.*, 2003; Barnett, 2004). In Australia, VET providers currently offer a range of services to support students with disabilities, for instance, public providers employ disability officers to liaise with individual students and arrange specific supports such as note-taking and access to adaptive technology. These officers also often have some responsibility for promoting improvements to the physical aspects of the institutional environment (such as access to buildings). However, the nature and extent of the supports that providers can offer is limited by funding and other resources and VET providers can offer only limited assistance with the wider issues that affect a student's capacity to study successfully, such as transport, housing, or activities of daily living (Smith and Ferrier, 2004).

Outcomes from VET

VET systems are strongly linked to employment. They seek to provide the skills sought by industries and employers and enable learners to achieve success and satisfaction in their working lives. However, regardless of differences between VET systems and the structure of economies, VET students with disabilities often experience poorer employment outcomes than other students.

Poorer employment outcomes for VET students with disabilities reflect the experience of people with disabilities in labor markets more generally. People with disabilities remain one of the most marginalized groups, especially when considering access to the mainstream labor market (European Union, 2006). While participation in learning can help to increase employability, action is needed to counter discrimination and employer concerns, such as the efficiency of workers with disabilities and possible additional costs of providing supports. The European Union (EU) is encouraging its member states to offer wage subsidies to employers of workers with disabilities and compensation for additional expenses. It hopes that "such aid may have a significant impact on ... employment opportunities" for people with disabilities.

Conditions that are more conducive to the employment of people with disabilities may be created by the spread and advance of technology. For instance, the EU's Disability Strategy Plan for 2004/05 identified that "new technologies ... empower people with disabilities and therefore facilitate access to employment" (European Union, 2006: 5, footnote). However, difficulties remain. A study from Korea (Kim and Zirkle, 2006) identified that the shift from an agrarian economy to a high-tech one has not met expectations of improved employment opportunities for people with disabilities. Employers remain unwilling to hire them.

In the context of labor-market discrimination and/or reluctance, the ability of VET institutions and systems to support transitions to employment by students with disabilities is limited, but there are things they can do. Studies show that collaborations between education institutions and employers can overcome some of the negative attitudes that influence recruitment practices (e.g., Harrison and Barnett, 2001; DET, 2001). Work placements during educational programs also can help students with disabilities to adapt to workplaces and employment requirements and can help to overcome the negative attitudes of some employers. But “careful preparation (is) required to develop good relationships and to restructure tasks to support work placements” (Waite *et al.*, 2006: 587).

Two other major factors contribute to poorer outcomes from VET for students with disabilities. The first is the problematic nature of transitions for these students – between schools and other education and training institutions; between postsecondary institutions; between programs; and between education, training, and employment. Studies from the UK and the US. identify transition difficulties relating to the lack of transfer of reliable information about learners (Perin, 1990; Ofsted, 2007; Nuehring and Sidlington, 2003). Students can miss out on required supports and costly new diagnostic assessments may also be required – and there is the fact that the necessary expertise might not be available. In Australia, the self-identification of students with disabilities in postsecondary education institutions, and differences in definitions of disability between education sectors are pertinent. Barnett (2004) suggests individuals can be trapped by intersectoral boundaries (or fall between them). There is, she suggests, a primary policy gap at key transition points and thus a need for cross-sectoral approaches.

The second factor is quality, particularly of guidance for students with disabilities and of the student experience in VET. Harrison and Barnett (2001) identify that students with disabilities may have difficulties in navigating and coordinating a complex array of educational opportunities. Appropriate guidance in choosing programs is thus essential. However, studies point to low expectations of students with disabilities, leading to inadequate academic challenges and placement in inappropriate programs. Other quality concerns include: insufficient efforts to integrate students with disabilities into the social life of the educational institution; less-than-required access for students to supports, equipment, and resources; and lack of use of appropriate and effective teaching practices for students with disabilities (Burchardt, 2005; Corbett, *et al.*, 1999; LSC, 2005; Ofsted, 2007). Inclusive approaches may help to counter these.

Conclusions

There is a growing recognition and acknowledgement of the potential of people with disabilities to learn and

to engage as full adult citizens in work and society, and of the important role of VET in assisting and supporting them to fulfill their potential. Typical of such views is Barnett’s statement that “the commitment to building a society which values difference and the need to enable all its citizens to contribute as well as receive, has to be the ultimate reward for building an inclusive training and vocational education system” (Barnett, 2004: 106).

Considerable progress is being made. Governments and policymakers are seeking ways to encourage and support people with disabilities to engage more actively in education and work, and at the institution level, knowledge of the impact of disabilities in general and of particular disabilities is growing and new and more effective teaching practices are emerging alongside initiatives including collaborations with employers. However, considerable scope for improvement remains. The quality of provision, of supports, and of guidance for students with disabilities is uneven. The experiences of students with disabilities in VET can be difficult and unhelpful. Inclusive systems and practices that welcome and support all students are needed. From an efficiency perspective, there is considerable scope for improved use of resources, while from an equity perspective there are significant possibilities for improving access, participation, and outcomes. Improvements can be made in education and training; in employment; and at the points of linkage (e.g., between school and VET or between VET and employment), where those with a disability are especially vulnerable.

See also: Community Integration and Employment of Youth with Special Needs; Educating Students with Special Needs: An Overview; Impact of Assessment on Learner Groups (Disabilities); Postsecondary Participation of Students with Special Needs.

Bibliography

- Andrews, R. and Smith, J. (1992). Additional costs of education and training for people with disabilities. *Final Report to the Department of Employment, Education and Training*, Andrews, R. J. *et al.*, Paradise Point, Australia: DEET.
- Barnett, K. (2004). VET-ability: The capacity of the Australian VET system to be inclusive for people with a disability. In Bowman, K. (ed.) *Equity in Vocational Education and Training: Research Readings*, pp 90–107. Adelaide: National Centre for Vocational Education Research.
- Barnett, K., Jardine, E., and Wilson, S. (1996). *Planning Together: Collaborative Approaches to the Participation of People With Disabilities in Vocational Education and Training*. Adelaide: Department for Employment, Training and Further Education.
- Burchardt, T. (2005). *The Education and Employment of Disabled Young People. Frustrated Ambition*. London: Joseph Rowntree Foundation, Policy Press.
- Buys, N., Kendall, E., and Ramsden, J. (1999). *Vocational Education and Training for People with Disabilities. Review of Research*. Adelaide: NCVER.
- Corbett, J., Kitteringham, J., and McAnespie, B. (1999). Vocational education and training to support the transition of young people with

- learning disabilities to paid employment. *Paper Presented at the European Conference on Educational Research*. September 1999. Lahti, Finland.
- Department of Employment and Training Queensland (DET) (2001). *Equal access? Six Action Research Projects into the Access and Equity of Students with a Disability and Aboriginal and Torres Strait Islander Students in Vocational Education and Training*. Brisbane: DET.
- Devlin, Y. (2000). *Students with Disabilities in Higher Education: At Whose Cost and What Price? Discussion paper for the Equity and Diversity Unit*. Sydney: The University of New South Wales.
- Dockery, M., Birch, E., and Kenyon, P. (2001). The economic and social analysis of increasing opportunities for people with disabilities in VET. *A Report for ANTA*, Perth: Curtin Business School, Curtin University of Technology.
- European Union (2006). Disability mainstreaming in the new streamlined European social protection and inclusion processes. *Disability High Level Group Discussion Paper for Meeting*, 11–12 October 2006. http://ec.europa.eu/employment_social/index/good_practis_en.pdf (accessed October 2009).
- Harrison, C. and Barnett, K. (2001). Collaborative approaches to increasing the participation and outcomes of people with a disability in vocational education and training. *Paper to the 2001 Conference of the Australian VET Research Association*. http://www.avetra.org.au/abstracts_and_papers_2001/Barnett-Harrison_full.pdf (accessed October 2009).
- Hart, D., Grigal, M., Sax, C., Martinez, D., and Will, M. (2006). Postsecondary education options for students with intellectual disabilities. *Research to Practice* 45. www.communityinclusion.org/article.php?article_id=178 (accessed October 2009).
- Hubka, D. and Killean, E. (1999). *Working towards a Coordinated National Approach to Services, Accommodations and Policies for Post-Secondary Students with Disabilities: Ensuring Access to Higher Education and Career Training*. Canada: Ontario.
- Kate, B., et al. (2003). *Mid-Term Review of Bridging Pathways*. Brisbane: Australian National Training Authority.
- Kim, H. and Zirkle, C. (2006). The status of career and technical education and workplace opportunity for the disabled in South Korea: A literature review. *Workforce Education Forum* 33(2), Fall 2006, Ohio State University. <http://voc.ed.psu.edu/projects/publications/books/Fall2006/index.html> (accessed October 2009).
- Learning and Skills Council (LSC) (2005). *Through Inclusion to Excellence. The Report of the Steering Group for the Strategic Review of the LSC's Planning and Funding of Provision for Learners with Learning Difficulties and/or Disabilities Across the Post-16 Learning and Skills Sector*. Coventry: LSC.
- Learning and Skills Council (LSC) (2006). *Learning for Living and Work. Improving Education and Training Opportunities for People with Learning Difficulties and/or Disabilities. Going Forward – Implementing the Vision of Through Inclusion to Excellence. The National Strategy for LSC Funded Provision for Learners with Learning Difficulties and/or Disabilities Across the FE System: 2006/07 to 2009/10*. Coventry: LSC.
- National Board of Employment, Education and Training (NBEET) (1994). Resource implications of the introduction of good strategies in higher education for disadvantaged students. *Commissioned Report No. 30*. Canberra: NBEET.
- National Centre for Vocational Education Research (NCVER) (2002). *Australian Vocational Education and Training Statistics 2000: Students with a Disability in Vocational Education and Training – At a Glance*. Adelaide: NCVER.
- National Centre for Vocational Education Research (NCVER) (2005). *People with a Disability in Vocational Education and Training: A Statistical Compendium*. Adelaide: NCVER.
- Nuehring, M. L. and Sitlington, P. L. (2003). Transition as a vehicle: Moving from high school to an adult vocational service provider. *Journal of Disability Policy Studies* 14(1), 23–35.
- Ofsted (2007). Current provision and outcomes for 16–18 year old learners with learning difficulties and/or disabilities in colleges, Document Reference Number: HMI 2371. London: Ofsted.
- Onnela, J. (2003). Viewpoints on the foundation and future of vocational training for visually impaired people. *Proceedings of the European Employment Conference*, Paris. <http://www.euroblind.org/fichiersGB/rvteconf.htm#0> (accessed October 2009).
- Perin, D. (1990). Instructing students with learning disabilities: Guidelines for community college faculty and support staff. *Conference Proceedings, Institute for Research and Development in Occupational Education, Centre for Advanced Study in Education*, New York: City University of New York.
- Smith, C. S. and Ferrier, F. (2004). *The Funding of Vocational Education and Training for Students with Disabilities*, 2 vols. Adelaide: NCVER.
- Smith, C. S., Ferrier, F., Anderson, D., et al. (2001). *CEET's Stocktake: The Economics of Vocational Education and Training in Australia*. Adelaide: NCVER.
- Waite, S., Lawson, H., and Robertson, C. (2006). Work-related learning for students with significant learning difficulties: relevance and reality. *Cambridge Journal of Education* 36(4), 579–595.
- Zolkowska, T., Kasior-Szerszen, I., and Blaszkiewicz, I. (2002). European Union policy toward people with disabilities. *Disability Studies Quarterly* 22(4), 217–224.

Further Reading

- Department of Education, Science and Training (2002). *Disability Discrimination Act 1992: Disability Standards for Education Draft*. Canberra: DEST. http://www.dest.gov.au/archive/iae/analysis/Draft_Disability_Standards.htm#intro (accessed October 2009).
- Department for Education and Skills (DfES) (2002). *Bridging the Gap: A Guide to the Students with Disabilities' Allowances (DSAs) in Higher Education in 2002/2003*. London: DfES. http://www.dfes.gov.uk/studentsupport/students/stu_students_with_ds.html (accessed October 2009).
- Schwartz, R., McKenzie, P., Hasan, A., and Nexelmann, E. (1997). *Country Note on Australia for the OECD Transition Review*. Paris: OECD.
- Smith, C. S. and Smith, J. S. (2002). Reflections on the impact of research on policy development: A case study of user choice. *Australian and New Zealand Journal of Vocational Education Research* 10(1), 69–94. AVETRA.

Prisoner Education and Training

S Dawe, S H Wynes and J Misko, National Centre for Vocational Education Research Ltd., Adelaide, SA, Australia

© 2010 Elsevier Ltd. All rights reserved.

In the Nordic countries, prisoners participate in mainstream adult education, including VET, while most other European countries emphasize personal development, especially access to basic education. The risk-need' model, identified in Canada in the 1990s, aims to reduce reoffending through programs which target factors directly related to the offence (including antisocial attitudes, gambling, substance abuse, and anger). This model of cognitive-behavioral programs has strongly influenced correctional systems in North America, United Kingdom, and Australia.

The experiences in these countries suggest that VET interventions should aim to meet the needs of individual prisoners, by creating activities that enable inmates to engage with learning. In this way, it can help to support prisoners' personal development as well as equip them to go on to further education and sustainable employment when they return to their community.

Introduction

Prison has traditionally had two major purposes: punishment and rehabilitation. A dramatic rise in the rate of incarceration in the last two decades in many Western countries has seen an increased focus on principles of punishment as the major way of controlling crime (see Crayton and Neusteter, 2008). At the same time, there has also been an effort to provide prisoners with education to aid them in rehabilitation and to prevent recidivism (i.e., reoffending and re-incarceration). Education is also seen as having a major role in ameliorating the behavioral problems that lead people into crime.

Incarceration alone does not prevent repeat offences; for example, around 50% of adult prisoners in the United States return to prison. However, vocational training and general education in prison has been found to produce some of the largest net economic benefits for adult programs in prison. Training programs help prisoners acquire the skills they need to reenter and stay in the community as law-abiding citizens. These programs cover basic education, literacy, language and communication skills, and employability preparation – as well as specific industry job skills – given that prisoners have generally low educational achievement and often lack life skills for coping in mainstream society (e.g., see the Irish Prison Adult Literacy Survey 2003 report by Morgan and Kett, 2003).

In this article, we outline the main characteristics of prisoner education and training programs in Europe, North America, and Australia. We find that whether these programs are underpinned by a human rights approach or a more instrumental one, there are common threads in all successful education and training programs. These include ensuring that the programs are tailored to meet the needs of individuals, create activities that will allow inmates to reconcile themselves with learning, and support prisoners' personal development as well as help to equip them to go on to further education and sustainable employment when they return to their community.

Key Approaches to Prisoner Education and Training

There is no uniform approach to prisoner education within and between prison systems in Europe, North America, and Australia. There are, however, two broadly similar streams of activity. The first is usually mandatory and aimed at changing or managing the behaviors and attitudes of prisoners (including for anger management or reducing drug and alcohol abuse); the other strives to improve the basic, general, and vocational education of individuals. The tenor of both streams of prisoner education is to improve knowledge and behavior so that prisoners have a second chance.

Prisoner Control and Rehabilitation Through Behavior Management Training

The behavior management model dominates those programs that assist offenders to confront and understand their criminal activity, and to develop techniques to control their behavior and avoid situations that may lead to further offending when released from custody or supervision. Such rehabilitation programs are often mandatory, and are also critical to the management of prisoners in prison. Since the 1990s, this approach has been particularly influential in Canada, Australia, and the United Kingdom.

In the Canadian model of the 1990s, referred to as the risk-needs model, risk is assessed in terms of static risk factors that are not amenable to intervention (including age of first offence, and criminal history), and dynamic risk factors that might change over time (such as family and social factors, substance abuse, educational factors,

and nonsevere mental health problems). These dynamic risk factors are also referred to in the literature as criminogenic needs.

Learning interventions are based on psychological theories of behavior management and take account of offender age, ethnicity, gender, disability, and socioeconomic status (Andrews and Bonta, 1998). Andrews (2001) concluded that reductions in recidivism can be maximized for high-risk offenders if they are involved in learning programs which target factors directly related to the reasons for offending (including antisocial attitudes, gambling, substance misuse, and anger). (This title is derived from the often-cited article by Martinson (1974): 'What works? – Questions and answers about prison reform' which reviewed 231 studies of prison rehabilitation programs and concluded that offender treatment programs had been largely ineffective. Critics of the nothing works doctrine actively challenged the assumptions and empirical evidence presented by Martinson (1974) and colleagues. Martinson (1979) also acknowledged errors in his earlier reviews, and cited findings from new studies which demonstrated that some treatments did work.)

In contrast, prisoner behavior management education in the United States of America has been influenced by a sociological perspective. (A Roadmap to Strengthening Public Safety, Correctional Service of Canada (CSC) Review Panel, 2007.) One example of this is the breaking-the-cycle program which has been trialed in three states (Alabama, Florida, and Washington) aimed at asserting the coercive power of the justice system to reduce drug use among offenders (Harrell *et al.*, 2003).

In Australia, prison arrangements vary across state and territory governments, but all have intervention programs that address the factors that impinge on the likelihood of reoffending. Here, the prisoner's sentence may mandate an intervention program aimed at reducing antisocial behavior (including substance abuse and anger) or other criminal conduct. The prisoner may also be involved in prison work.

In the United Kingdom, majority of prisoners are required to work while in prison. These work activities

may involve them in gardening, farming, and animal husbandry to provide food for prisons, as well as laundry, kitchen, bakery, gardening, and maintenance duties. Prisoners also work in industries and workshops supplying goods or services to the outside community on a commercial basis. Such mandatory programs are aimed at controlling and altering the behavior and attributes of prisoners, while also imparting skills to help them integrate back into the community.

Improving Rehabilitation Through Education and Training

In the Nordic countries (Denmark, Sweden, Finland, and Norway), North America, Australia, United Kingdom, and Ireland, prison systems generally recognize the need to provide prisoners with basic education – language, literacy, and numeracy. Prisoner participation in these forms of education is not mandatory. In the Nordic countries, the focus is on ensuring that prisoners have the same rights to education as do citizens outside prison, and that the prisoners have opportunities to participate in all forms of education, including vocational and higher education. In the United States, Canada, Australia, United Kingdom, and Ireland, the focus is on prisoners acquiring basic literacy and numeracy skills to enable them to participate in training that will provide them with the best prospects for employment on their release (see Table 1). In all countries, the aim is to use education and training as a pathway to rehabilitation and, subsequently, to reduce recidivism.

The Nordic Model: Ensuring that Prisoners Have Rights to All Forms of Education and Training

In Denmark, Finland, Norway, and Sweden education and training for prisoners is organized as part of the regular education system, with the same objectives as adult education in the community (Nordic Council of Ministers, 2005).

Table 1 Focus of prison education across various countries

Country	Focus	Provision
Nordic countries	Right to all forms of education and training in the regular education system	Schools in prisons or educational institutions in the community
United States	Focus on basic education and training as a route to further education and employment	Often in partnership with local schools and colleges
Canada	Focus on adult basic education pathways to assist with rehabilitation	
Australia	Focus on basic education and accredited vocational education and training to assist with employment outcomes	Registered training organizations using a learner-centered approach
United Kingdom and Ireland	A focus on providing skills for employment	Partnerships between government training organizations and prison systems

This approach is underpinned by the principle that the only deprivation that prisoners should suffer is the loss of liberty.

In Denmark, most prison schools are designated local adult education centers. The Prisons and Probation Service has traditionally employed prison teachers whose duties include both teaching and case management. When individuals complete a basic adult education program they become eligible for further vocational and educational courses. If a prisoner wishes to pursue a vocational program, while serving his or her sentence, this has to be done in cooperation with an enterprise in an industry outside the prison. Often, it is possible for the prisoner to complete the first parts of the program at the prison workshop and later have day-release privileges to participate in training and to work in the industry in question.

Prison education in Finland is organized in cooperation with and under the supervision of educational institutions in the community, which are responsible for pedagogical guidance and teacher management. Some teachers work both in prison schools and at other schools, some only at prison schools. In prisons, teachers are not normally involved in the general case management of prisoners.

Vocational education and prison work have always gone hand in hand. In recent times, there has also been a shift toward more apprenticeship training. In these programs, the foreman of the prison workshop gives the practical training, while the educational institution provides the theoretical education.

In Norway, prisoners are assured of all their rights as citizens. An upper-secondary school takes the overall responsibility for the education offered in each prison. The prison school operates as a branch of that main school. Formally, the teachers are in the employment of the main school. Many of the workshops in Norwegian prisons have been granted the status of educational workplaces, which means that they can train prisoners as apprentices and trainees. The prison workshop (or the prison) receives a grant of a specific sum per prisoner who has been accepted as an apprentice or trainee. If there is good cooperation between the prison school and the prison workshop, prisoners will receive theoretical education in the classroom and practical training in the workshop.

In Sweden, prisoners are also encouraged to participate in education and training that will help them reintegrate into society. The provision of vocational training in Sweden is based on an agreement between the Prison and Probation Service and the Swedish National Labour Market Administration. This agreement guarantees that prisoners will have the same rights to services and programs as other citizens (Nordic Council of Ministers, 2005).

Despite general support in the Nordic countries for prisoners having access to the same education as other

citizens, a study of prison education (Nordic Council of Ministers, 2005) has found that prisoners' rights to education and training are not clearly defined, and needed to be stipulated in the legislation. (The case in Australia is similar where only one state has legislation providing education as a right for prisoners; see de Graaf, 2007.)

The Nordic study also noted the low level of schooling of most prisoners, and supported enabling individuals to access education and training while in prison, to provide them with the skills and knowledge required for employment. Without these arrangements they would continue to be excluded from the labor market. The report also concluded that education and training in prisons should be considered as investment in crime prevention. It also recognized the need for teachers in prisons to identify and address the different learning needs of the individual prisoners, language, religion, and culture (Nordic Council of Ministers, 2005).

Prison Education in North America: A Focus on Basic Qualifications as a Route to Further Education, Employment, and Rehabilitation

There is no universal approach to prison education across the different states of the United States. The general focus is on ensuring that prisoners who participate in training should be able to complete their basic high school qualification with a view to providing them with better opportunities to move into further education and training or gain employment. This form of education is often provided in partnership with local schools and colleges.

The United States Re-entry Policy Council's Report (Council of State Governments, 2005) concluded that coordinated action by government agencies, nongovernment service providers, and the community is critical to support prisoners and ex-prisoners in getting and retaining a job, obtaining housing and healthcare, and establishing social and family support networks in the community. (The Council of State Governments is a nonpartisan, public, nonprofit organization that provides information, research, and training to state officials in all three branches of government in every state and US territory.)

Since 2004 a range of federal government grants have been available for communities to establish prisoner reentry programs, which focus on the community environment and the ex-offender's employment, education, health support and housing needs in that community. (The Prisoner Re-entry Initiative is supported by the US Department of Justice (DOJ), Office of Justice Programs (OJP), and its federal partners: the US Departments of Education, Health and Human Services, Housing and Urban Development, and Labor. This initiative is a comprehensive effort that addresses both juvenile and adult populations of serious, high-risk offenders.)

The most convincing evidence that education for prisoners has a positive effect on post-release behavior of prisoners in the USA was provided by the Three-State Recidivism Study. (The Three-State Recidivism Study was funded by the US Department of Education, Office of Correctional Education, and conducted in 1999 by the US Correctional Education Association; Steurer *et al.*, 2001.) This study, conducted in the states of Maryland, Minnesota, and Ohio, compared two groups of offenders, those who had participated in education while in prison and those who had not. It found a 30% reduction in recidivism for participants than for nonparticipants. This study was also designed to assess the impact of prison education on employment outcomes after release from prison. The authors concluded that education for prisoners enhances employment opportunities, decreases criminal behavior, and, in so doing, reduces the overall cost of crime to the community.

Another study by Aos *et al.* (2006) set out to measure the economic benefits of prisoner education. It found the marginal cost of vocational training (VT) programs to be \$1182 per prisoner. This was offset by the marginal savings to the taxpayer from lower criminal justice costs of \$6806 per prisoner. (The Prisoner Re-entry Initiative is supported by the US DOJ, OJP, and its federal partners: the US Departments of Education, Health and Human Services, Housing and Urban Development, and Labor. This initiative is a comprehensive effort that addresses both juvenile and adult populations of serious, high-risk offenders.) For general education, the marginal costs were \$962 per person and the taxpayer savings were \$5306. The study went further to add victim savings, finding the net benefit for VT programs was calculated as \$13 738 per inmate and for general education, \$10 669 per inmate. These were for reductions of 9% and 7% in recidivism for VT and general education programs, respectively.

There is also a substantial body of work from the Urban Institute that shows remedial intervention programs are required to help prisoners with low-level skills in reading, writing, mathematics, and oral communication to become law-abiding citizens in the community. (The Urban Institute is a nonpartisan economic and social policy research organization that publishes studies, reports, and books on timely topics worthy of public consideration.) These are also important in reducing recidivism rates. A study by Solomon *et al.* (2006) found that programs which were successful in helping ex-prisoners to find suitable employment and reintegrate into society typically provided intensive job-placement services and ongoing monitoring and support. In one pilot study, a network of employers has demonstrated their willingness to hire former prisoners. (The Returning Home: Understanding the Challenges of Prisoner Re-entry study was piloted in Maryland and was implemented in its entirety in Illinois, Ohio, and Texas.)

Individuals in Canadian prisons may access adult basic education programs at primary, secondary, and post-secondary levels, as well as vocational education programs. This provision was promoted by Porporino and Robinson (1992) who suggested that specific intellectual skills gained through adult basic education could equip offenders to deal more effectively with daily problems encountered in the community. These researchers noted that the sense of achievement and confidence that might result from successfully completing such a program could encourage offenders to make further positive changes in their lives. More recently, the Correctional Services of Canada Review Panel (2007) noted the need to make prison work more meaningful to prisoners to enhance their skills, hence employment opportunities when they are released from prison. The panel recommendations include a more structured workday to allow for the proper allocation between work, education, and correctional programs.

Prisoner Education in Australia: A Focus on Accredited Vocational Education and Training to Improve Employment Outcomes

In Australia, prisoners' participation in education and training and in prison work is voluntary and is encouraged by a payment dependent on good behavior and participation. While there is a major focus on basic education, especially English language and literacy, the aim is also to involve prisoners in training to help improve their labor-market outcomes. This approach was announced in the National Strategy for Vocational Education and Training for Adult Prisoners and Offenders in Australia which aimed "to provide adult prisoners and offenders with educational and vocational pathways which will support their productive contribution to the economic and social life of the community" (Australian National Training Authority, 2001: 3).

In all states and territories, accredited training for inmates is provided by local registered training organizations (which are subject to mainstream independent quality audits). Typically, they will use learner-centered approaches to encourage prisoners to engage in further learning that will help improve their employment prospects on release. Learning may be self-directed, undertaken in small groups, or while on the job when working in prison services or industries.

Connecting vocational education and training (VET) in prisons to the national VET system has a number of benefits. It enables prisoners to obtain nationally recognized, industry-specified qualifications, and in turn can be used to enhance the prospects for employment of ex-prisoners. Callan and Gardner (2005) found in their study of Queensland correctional institutions, 32% of those who did not

participate in VET before their initial release returned to custody within 2 years, while only 23% of VET participants returned to custody within the same period.

Giles *et al.* (2007), in Western Australia, found that significantly more prisoners undertaking VET courses expect better labor-market futures (such as work, more enjoyable work, and more money) than those undertaking nonvocational education courses or solely engaged in prison work alone. (Nonvocational courses are defined as educational activities without a direct link to the labor market, and mostly relate in adult- and community-learning contexts to arts, crafts, and leisure.)

In 2003, a government-commissioned study undertaken by the Australian Institute of Criminology assessed the interventions available for prisoners returning to the community (Borzycki, 2005). The study noted that prisoners were not a homogeneous group and that the needs of certain subgroups needed to be addressed. These included prisoners with mental health problems, those with an intellectual disability, females (especially those with dependent children), Indigenous prisoners, and those who have been incarcerated, or on remand, for very short periods of time. (Offenders charged but not convicted may be held in the custody of correctional services on remand while awaiting trial, when the judge has not granted them bail to live in the community (i.e., provide surety that they will appear at their trial).)

This evaluation also concluded that, if prisoners were to make successful reentry into society, effective through-care strategies were required. (Throughcare describes the process of delivering continuous care – providing consistent services and support to prisoners within and beyond prison in an holistic program of rehabilitation, ideally commencing at first contact between the offender and the justice system; see Borzycki, 2005.) It was essential that government agencies, nongovernment services providers, and the community organizations coordinated activities even before prisoners were released to ensure that ex-prisoners did not fall through the service gaps between agencies post-release (Borzycki, 2005).

Prison Education in the United Kingdom and Ireland: Partnerships between Government Training Agencies and Prison Systems

Findings about the high unemployment rate among ex-prisoners and subsequent higher reconviction rate among those who remain unemployed have influenced the development of major policies and programs dealing with prisoner education in the United Kingdom. In 2006, the government published its three priorities for reducing reoffending through skills and employment. These include engaging employers through a corporate alliance to employ ex-offenders; building on the offender learning and skills

service; and reinforcing the emphasis on skills and jobs (Department for Education and Skills, 2006).

In practice, this means a partnership between the National Offender Management Service and the Learning and Skills Council to design and fund the Integrated Offender Learning and Skills Services program, which caters to offenders in custody and in the community. The program aims to ensure that prison industries and workshops provide more meaningful work and prepare prisoners more effectively for jobs in the community. (Most prisoners are required to work while in prison: prison work usually refers to those activities which supply services to the prison such as laundry, kitchen and bakery duties, cleaning, gardening, and maintenance. This may also include horticulture, e.g., growing fruit and vegetables, or farming and animal husbandry such as dairy to provide food for prisons. Prison industries and workshops usually refer to specific industry work which supply goods or services to the outside community or customers on a commercial basis. These may include metal work production, plastic products, repairing electronic goods, or other manufacturing or catering services depending on the resources available in the prison.) This includes working more closely with employers to meet their needs. The hope is that employers would employ the prisoner on completion of their sentence. Prior to the integrated Offenders Learning and Skills Services program being implemented in August 2005, researchers from the Learning and Skills Development Agency conducted an interim evaluation of prototype activities in the three development regions. (The interim report titled: *Evaluation of Regional Plans for the New Integrated Offender Learning and Skills Service*, highlights findings and lessons learned from evaluation workshops and interviews with key stakeholders in each region (Walker *et al.*, 2005).) The evaluation noted that successful transition to the community can be helped by one-to-one support and motivation for prisoners who move through the gate.

Conclusion

The rights-based approach implemented in the Nordic countries and the more vocationally oriented systems adopted in the Anglophone countries are underpinned by different philosophies. However, their main aims are to engage prisoners in education and training to help them control problem behavior and prepare for a return to the community and to gain employment.

Prisoners are not a homogeneous group; therefore, it is important to tailor education and training to their individual needs, especially when these differ significantly from the mainstream.

Helping prisoners meet the variety of challenges they will face on release is often beyond the scope of

correctional service agencies. In this regard, VET can create opportunities for inmates to re-engage with learning, as well as provide them with opportunities for personal development.

Bibliography

- Andrews, D. (2001). Principles of effective correctional programs. In Motiuk, L. and Serin, R. (eds.) *Compendium 2000 on Effective Correctional Programming*, pp 9–17. Correctional Service of Canada, Ottawa. <http://www.csc-scc.gc.ca/text/rsrch/compendium/2000/index-eng.shtml> (accessed October 2009).
- Andrews, D. and Bonta, J. (1998). *The Psychology of Criminal Conduct*, 2nd edn. Cincinnati, OH: Anderson.
- Aos, S., Miller, M., and Drake, E. (2006). *Evidence-Based Public Policy Options to Reduce Future Prison Construction, Criminal Justice Costs, and Crime Rates*. Olympia: Washington, State Institute for Public Policy.
- Australian National Training Authority (ANTA). (2001). *National Strategy for Vocational Education and Training for Adult Prisoners and Offenders in Australia*. Australian National Training Authority. Brisbane: Department of Education, Employment and Workplace Relations? Formerly on DEST. http://www.acea.org.au/Content/Links/National_Strategy_VET_Corrections.pdf (accessed October 2009).
- Borzycki, M. (2005). Interventions for prisoners returning to the community report prepared by the Australian Institute of Criminology for the Australian Government. Attorney-General's Department, Canberra. <http://www.aic.gov.au/publications/reports/2005-03-prisoners.html>.
- Callan, V. and Gardner, J. (2005). *Vocational Education and Training Provision and Recidivism in Queensland Correctional Institutions*, National Centre for Vocational Education Research (Also covered in Dawe, 2007), Adelaide.
- Correctional Service of Canada Review Panel. (2007). *A Roadmap to Strengthening Public Safety*. Correctional Service of Canada, Ottawa. <http://www.publicsafety.gc.ca/csc-scc/report-eng.aspx> (accessed October 2009).
- Council of State Governments. (2005). *Report of the Re-Entry Policy Council: Charting the Safe and Successful Return of Prisoners to the Community*. Re-Entry Policy Council, Council of State Governments, New York. <http://reentrypolicy.org/Report/About> (accessed October 2009).
- Crayton, A. and Neusteter, S. (2008). The current state of correctional education. *Commissioned Paper for Re-Entry Roundtable on Education held March 31 & April 1 2008*. <http://www.urban.org/projects/reentry-roundtable/upload/Crayton.pdf> (accessed October 2009).
- De Graaf, P. (2007). Should education and vocational training be compulsory in corrections? In Dawe, S. (ed.) *Vocational Educational and Training for Adult Prisoners and Offenders in Australia: Research Readings*. Adelaide: National Centre for Vocational Education Research.
- Department for Education and Skills. (2006). *Reducing Re-Offending Through Skills and Employment: Next Steps*. Nottingham: United Kingdom Department for Education and Skills Publications. <http://www.dfes.gov.uk/offenderlearning/> (accessed October 2009).
- Giles, M., Tram Le, A., Allan, M., et al. (2004). To train or not to train: The role of education and training in prison to work transitions. In Dawe, S. (ed.) *Vocational Educational and Training for Adult Prisoners and Offenders in Australia: Research Readings*, pp 67–77. Adelaide: National Centre for Vocational Education Research.
- Harrell, A., Mitchell, O., Merrill, J., and Marlow, D. (2003). *Evaluation of Breaking the Cycle*. Washington, DC: Urban Institute, Justice Policy Center. <http://www.urban.org/url.cfm?ID=409537> (accessed October 2009).
- Martinson, R. (1974). What works? Questions and answers about prison reform. *Public Interest* 10, 22–54.
- Martinson, R. (1979). New findings, new reviews: A note of caution regarding sentencing reform. *Hofstra Law Review* 7, 243–257.
- Morgan, M. and Kett, M. (2003). *The Prison Adult Literacy Survey: Results and Implications*. Dublin: Irish Prison Service. http://www.irishprisons.ie/documents/Adult_Literacy_Survey.pdf (accessed October 2009).
- Nordic Council of Ministers. (2005). *Nordic Prison Education: A Lifelong Learning Perspective*. Schenck, L. (trans.). Copenhagen: TemaNord. http://www.epea.org/images/pdf/Nordic_Prison_Education_A_lifelong_learning_perspective.pdf (accessed October 2009).
- Porporino, F. and Robinson, D. (1992). *Can Educating Adult Offenders Counteract Recidivism?* Ottawa: Correctional Services Canada. <http://www.csc-scc.gc.ca/text/rsrch/reports/r22/r22e-eng.shtml> (accessed October 2009).
- Solomon, A., Visher, C., La Vigne, N., and Osborne, J. (2006). *Understanding the Challenges of Prisoner Reentry: Research Findings from the Urban Institute's Prisoner Re-Entry Portfolio*. Washington, DC: Urban Institute. <http://www.urban.org/publications/4112889.html> (accessed October 2009).
- Steurer, S., Smith, L., and Tracy, A. (2001). *Three-State Recidivism Study*. Lanham, MD: Correctional Education Association.
- Walker, L., Harrison, D., and Zwart, R. (2005). *Evaluation of Regional Plans for the New Integrated Offender Learning and Skills Service, Research Report*. London: Learning and Skills development Agency. <http://www.lseducation.org.uk> (accessed October 2009).

Relevant Websites

- <http://www.publicsafety-sp-gc.ca> – Executive Summary: A Roadmap to Strengthening Public Safety, Public Safety Canada, Bienvenu.
- <http://www.ojp.usdoj.gov> – Office of Justice Programs.
- <http://www.pesireland.org> – Prison Education Service Ireland.
- <http://www.reentry.gov> – Reentry: Office of Justice Programs.
- <http://www.csg.org> – The Council of State Governments, Lexington.
- <http://www.urban.org> – Urban Institute: Research of Record.

Vocational Education and Training and the School-to-Work Transition

G Hayward, University of Oxford, Oxford, UK

© 2010 Elsevier Ltd. All rights reserved.

It is taken for granted that the path from schooling to employment is an increasingly perilous one for young people compared to the situation only a few years ago. While the idea of a golden age is perhaps exaggerated, both academic researchers and policy commentators seem to agree that school-to-work transitions, defined here as the period between the end of compulsory schooling and the attainment of full-time stable employment (Ryan, 2001), have become more complex and difficult, at least for some young people (cf. Furlong and Cartmel, 1997; OECD, 2006, 2007; Ryan, 2001; Vickerstaff, 2003).

Indicators of School-to-Work Transition Problems

A wide variety of indicators are used to compare and contrast school-to-work transition processes across Organization for Economic Cooperation and Development (OECD) countries (see Quintini and Martin, 2006). On an average between 1995 and 2005:

- Youth unemployment rates fell by 0.8 percentage points.
- The youth to adult employment ratio has risen from 2.4 in 1995 to 2.7 in 2005, indicating that the relative position of youth has deteriorated in more than two-thirds of OECD countries.
- Long-term unemployment declined, although some Eastern European countries and Japan have seen rises.
- Youth employment rates deteriorated. While this may be a good thing if it indicates increased time in education, in many countries the fall was associated with a rise in youth unemployment.
- The proportion of young people Not in Education, Employment or Training (NEET) fell between 1996 and 2003 but remains high at 17% on average for 20–24-year-olds.

The outflow from NEET status is high in some countries, such as France, Finland, and the United States so that this status is of short duration. In some countries, however, a proportion of young people are long-term NEET (more than 5 years) – in Italy 30%, Greece 20%, and more than 10% in several other European countries. A major contributor to these high figures is school failure: on an average

more than 14% of young people in OECD countries leave school without upper-secondary qualifications. These young people are overrepresented in the NEET group.

Even for those who acquire adequate levels of qualification, transition to employment can be long in European countries, an average of 2 years after leaving school. Participation in temporary and part-time work by youth has increased, particularly for initial labor-market entrants. This may not be a problem if such jobs represent a first step to more stable employment, as is the case in many countries. In others, such as Spain, temporary work traps are more prevalent for young workers. Young people are also more likely to experience low pay. However, outflow from these jobs to better paid ones is high, certainly higher than exit from unemployment: it is better, in terms of progressing in the labor market to have a low-paid job than no job at all (Quintini and Martin, 2006).

Rates of over-education, when young people have jobs which require much less skill than that acquired in initial education, have also increased: one in five youth in OECD countries was over educated in 2005. The prevalence of over-education is much higher among young workers in part-time and temporary jobs compared to those in full-time and/or permanent employment. Young people are also more likely to engage in job shopping as they seek to match their skills to employer demands – a process leading to job churn which may be an indicator of healthy labor-market functioning rather than necessarily being a problem.

School-to-work transition rates vary along predictable lines of disadvantage with, in most cases, young people with low educational attainment, from poorer socioeconomic backgrounds, and minority ethnic groups experiencing greater challenges than their more privileged counterparts. In some countries, young women remain at greater risk of unemployment and lower wages than men. Aboriginal young people often face the greatest disadvantage (OECD, 2008a).

Collectively, trends in these various indicators paint a complex picture but they speak to the continuing challenge of school-to-work transition for some young people. However, it would be inaccurate to speak of a crisis. A vast majority do make a successful transition to stable employment, but a small and important minority of disadvantaged and marginalized youth struggle (Marks, 2005; Hayward *et al.*, 2008; Pring *et al.*, 2009).

Policy Responses

It is important to recognize that no country employs a single type of policy response, such as providing improved access to vocational education and training (VET), in order to improve transitions into the labor market for young people. Other policy measures may include macroeconomic measures to stimulate economic growth, labor-market deregulation, subminimum wages for young people, promoting early retirement, and providing subsidized employment in addition to a range of employment and training policy initiatives. Which of these components is employed in the policy mix changes over time and across countries depending upon their particular national congeries of school-to-work transition institutions (such as school employer recruitment networks and mass apprenticeship), macroeconomic climate, stage of the business cycle, labor-market structure, experience-based trends in labor demand, and national pay determination systems. In turn, these can be linked to larger cultural processes such as the model of capitalism and associated forms of the welfare state that have developed in different countries (Hall and Soskice, 2001; Iversen, 2005). However, given weak macroeconomic growth and the historically poor performance of labor-market policies to tackle youth joblessness (Ryan, 2001), attention has increasingly focused on education and training interventions. (Evaluating the effectiveness of any policy measure is beset by a myriad of measurement and econometric analysis problems. These lie beyond the scope of this article but are well reviewed by Grubb and Ryan (1999) and Ryan (2001). Strong claims about the effectiveness of particular interventions need to be treated with caution unless experimental designs have been used or appropriate statistical methods adopted to deal with selection bias, such as Heckman two-stage procedures.) These form the focus of the remainder of this article.

Theories of Action

The theory of action underpinning VET reform is predicated on there being a vocational effect, an improvement in an individual's chances of making a successful school-to-work transition compared to those leaving with lower-secondary education or following an academic upper-secondary education (Ianelli and Raffe, 2007). The nature of the predicted vocational effect varies across theoretical traditions. Human capital theory would predict a positive effect in terms of improved employment and earnings as a result of further investment in education. By contrast, signaling and job queue theories predict a negative effect, as VET participation stigmatizes young people relative to their academic counterparts. Network theories would predict a positive effect if VET participation affords access to networks that support transition to employment. Social reproduction

theorists argue that VET participation simply cools out lower attaining, primarily working-class youth (the diversion hypothesis). Such young people are socialized into economic roles that, while affording transition into employment (the safety-net hypothesis), have little effect on other labor-market outcomes such as wages or occupational status (Ianelli and Raffe, 2007).

Such predictions are highly contingent on the distinction between academic and vocational programs, and employers' preferences for different types of qualifications. Furthermore, the underpinning processes are not mutually exclusive but operate differentially in different national transition systems constituted by the institutional and structural arrangements which shape young people's transitions. Such arrangements include distinctions between internal and occupational labor markets, the degree of labor-market regulation, the degree of stratification of the education system, the linkages between education systems and the labor market, and the occupational specificity of VET.

Schröder (2004) characterizes such transition regimes in a two-dimensional model: one dimension is formed by the strength of linkage between the education system and the labor market, the other by the degree of labor-market regulation. Education systems with strong linkages to the labor market are characterized by a high degree of occupationally specific human capital formation, strong employer/trade union engagement with and influence over the VET system, and good communication of labor-market signals to the VET system. Such linkage should produce a positive vocational effect by enhancing human capital formation in a way that employers value by developing knowledge and skills needed in the labor market, and reducing information asymmetries that stigmatize VET graduates, as employers have direct knowledge of the content of the programs. Regulated labor markets are those where contracts and wages are strictly regulated by law or by collective bargaining, flexible labor markets are then the converse of such arrangements. Taking account of such dimensions is crucial in evaluating the impact of VET programs.

Improving Retention Rates in Full-Time Education

School-to-work transition problems are particularly pronounced for the least well educated in most countries. Data for 2005 indicate that across the OECD, as a whole, completing upper-secondary education reduces the unemployment ratio of nonstudents as a percentage of the 20–24-year-old age cohort by 7.3% and by 7.1% among 25–29-year-olds. In most OECD countries, completing tertiary education reduces this ratio further for 25–29-year-olds, but this is not the case in Greece, Italy, and Turkey, and is only the case for young women in Spain (OECD, 2007). Thus, while more education and training is a usual prescription for problems of school-to-work

transition, it is not a panacea. This is largely because such policy prescription does not pay sufficient attention to structural macroeconomic problems that affect the supply of jobs, and the interaction of macro- and microeconomic factors with the sustainability of certain types of training provision, for example, apprenticeship in Germany.

The variation in indicators of transition success across countries indicates the problems in making direct comparisons of the effectiveness and efficiency of different VET initiatives in supporting young people in their journey from education to the labor market. Nonetheless, one of the strongest findings in the economics of education is that additional years of education is correlated with better transitions into the labor market, both in terms of securing decent, stable employment and improved future earnings. Most OECD countries have witnessed a sustained increase in upper-secondary educational enrolments and attainment over the last three decades. However, the nature of upper-secondary education institutions and curricula may vary radically between countries with historically similar success in promoting successful school-to-work transitions. Contrast, for example, Germany with its tradition of part-time VET grounded in apprenticeship with Japan's full-time upper-secondary education system providing mainly general education (Ryan, 2001).

The last three decades have witnessed both the introduction and the substantial reform of existing VET programs to effect greater postsecondary enrolment and graduation rates by providing more practical and motivating curricula. In addition, some countries are turning to mandates designed to force participation. For example, the Netherlands 2007 Qualification Law (*Kwalificatieplicht Wet*) requires young people who have acquired a basic education qualification (the *Startkwalificatie*, the equivalent of an upper-secondary degree) to follow a full-time education program until the age of 18. For those aged 18–27, who have not successfully completed secondary education, the intention is to introduce mandatory study/work programs by 2009 (OECD, 2008b). Similarly, English policymakers intend to raise the compulsory participation age from 16 to 18 years by 2015. Such mandates speak to the challenge of encouraging voluntary participation in post-secondary education among some early school leavers even when offered more VET provision.

The impact of such vocational curricula and work-based curricula are difficult to evaluate statistically, given the likely effect of selection around unobservables (ability, motivation, and social class), the almost total lack of experimental studies, and the paucity of prior labor-market experience to use in econometric models (Ryan, 2001). In addition, the necessary longitudinal data required to control for selection bias are almost always too time limited to evaluate the longer-term impact of participation in VET on outcomes, such as occupational status and lifetime earnings. Where long-time-run longitudinal data are available, for example, from the UK's National Child Development

Studies, then the evidence indicates larger returns to participation in general education than in VET programs, but positive returns to participation in most upper-secondary VET provision when compared to those not undertaking any upper-secondary education. Such effects are stronger for employment than for other labor-market outcomes, such as occupational status and wages. Further, the sign and magnitude of the effects varies by gender and the type of vocational programs followed and the qualifications obtained (Dearden *et al.*, 2000). Such evidence mirrors that from the US which suggests that traditional vocational curricula are only selectively associated with higher pay, for example, females taking commercial courses and where there is good job matching (Ryan, 2001). Such weak signals from the labor market about the value of vocational qualifications can reduce the incentives to participate (Keep, 2009).

The lack of any general pay gains from participation in vocational education leads to questions about its efficiency given the higher costs associated with vocational as compared to general education programs. Nonetheless, it does appear that participation in upper-secondary VET does produce a positive vocational effect for some young people under some circumstances (given data limitations). In particular, there seems to be robust evidence for strong employment benefits for young people who have participated in postsecondary VET programs. However, evidence from Australian research utilizing the Longitudinal Survey of Australian Youth disputes the generalizability of employment gains from undertaking postsecondary VET. Experience in the labor market, including part-time work, in the first year after school for nonuniversity-bound youth increases the odds of being employed 4 years later compared to participating in nonuniversity study (Marks, 2005). Using the same data, Anlezark *et al.* (2006) found that participation in school VET programs increased retention between years 10 and 11 in Australian secondary schools but reduced retention between years 11 and 12. While transition to employment, at a time of a buoyant labor market, was smoother for those who had undertaken VET in upper-secondary school, the effect diminished with time. The research also found that school VET programs did not align well with out-of-school VET programs that may reflect better real labor-market training needs.

Most research provides very inconclusive evidence about the significance, size, and direction of the vocational effect. In general, the effect is positive for those completing upper-secondary VET programs compared to those with lower-secondary education alone, but negative relative to general upper-secondary education. Transition rates to higher education are lower for those with VET rather than general upper-secondary qualifications (Hölscher *et al.*, 2008). There is often a positive effect in terms of entering employment, but a highly variable effect on wages and occupational status depending upon the type of qualification obtained, mode of study, and gender (Dearden *et al.*, 2000; McIntosh, 2007; Jenkins *et al.*, 2007).

The only possible conclusion to be drawn is that there is no consistent association between participation in upper-secondary VET programs and subsequent transitions to the labor market. This is not to deny the positive motivational effects that can be provided by access to more vocational programs but it does call into question general conclusions of the effectiveness of such reforms in promoting more successful school-to-work transitions (Stanley, 2007). In particular, it should raise questions about their efficiency given the higher costs associated with providing VET.

In part, the different conclusions reached by different studies may be a product of the problems of data limitations and selection bias (Ryan, 2001, 2003; Ryan and Grubb, 1999). In part, they reflect inadequate attention being paid to the characteristics of transition systems in cross-national studies. Ianelli and Raffe (2007) focus on education-labor-market linkages as key determinants of national transition patterns. They distinguish between VET systems which have an employment logic aligned with Schröder's strong linkage and those with an education logic where links to employment and the differentiation between general and vocational education are weaker. Such a framework allows for the development of both characterizations of transition systems in different countries and stronger hypotheses about the relative size of the vocational effect under different transitional arrangements. These hypotheses can then be assessed using econometric techniques, provided appropriate and adequate data are available. It remains too early to assess fully the promise of this type of approach but initial results at least point to interesting new insights (Ianelli and Raffe, 2007; Schröder, 2004).

Apprenticeship

Apprenticeship clearly follows an employment logic but is also often used as a vehicle to support general education for young people. The design of apprenticeship provision varies considerably between countries. For example, in the UK there is no requirement for school-based part-time study as part of an apprenticeship program in contrast to the dual systems of Austria, Denmark, Germany, Norway, and Switzerland. The extent of participation in apprenticeship also varies considerably between countries, providing the majority of VET opportunities in Germany but a minority in Australia, France, the Netherlands, and the UK.

One of the enduring debates about vocationalism is the relative benefits of full-time VET delivered in schools and colleges, and apprenticeship systems in supporting successful transitions to employment. The historical success of apprenticeship in supporting positive school-to-work transitions in countries, such as Germany and Switzerland, is often taken as *prima facie* evidence of its efficacy. Australian evidence also points to the improved employment outcomes is associated with participation in apprenticeship and traineeship compared to participation in upper-secondary VET programs, but only for women (Marks, 2005).

Strong rhetorical messages are used to promote the advantages of apprenticeship both to employers and young people in the UK (Hayward *et al.*, 2004, 2005). Econometric analysis that seeks to directly compare the relative merits of school-based VET and apprenticeship is, however, relatively sparse. It is best where apprenticeship and school-based VET programs lead to the same qualifications, as can be the case in France, the Netherlands, and the UK. However, selection problems remain; for example, negative selection into apprenticeship programs by ability but positive selection by motivation in France. In the UK, there seems to be differential selection effects by ability depending upon the sector and occupation (Hayward, 2005). The alternative comparator group is those who enter the labor market and/or labor-market programs rather than remain in upper-secondary education. Here, apprenticeship performs well, both in terms of employment and improved wages.

Cross-national comparisons indicate a strong employment effect associated with participation in apprenticeship in countries with larger apprenticeship systems, with youth accessing jobs in skilled occupations in high-wage sectors. Micro-econometric evidence suggests that in France, the UK, and the US apprenticeship leads to improvements in early labor-market experiences for some young people relative to full-time VET programs. However, in France, acquisition of a vocational qualification matters more than the way it is obtained for longer-term outcomes. In the USA historically, apprenticeship has provided increased wage premia relative to full-time tertiary education but this may be due to the concentration of apprenticeship in highly unionized manual crafts in the construction sector. With the exception of Germany, in general, apprenticeship does less for young women than men, although in the case of the UK this may be in part the result of not controlling for occupational concentration of female apprenticeship in areas such as hairdressing. In the UK, the positive employment benefits of apprenticeship for young men only apply in some sectors, primarily those with a history of apprenticeship provision (Ryan, 2001; McIntosh, 2007).

Overall, the message about the benefits of apprenticeship is not a general one. It does seem to provide better early career opportunities and experiences but the subsequent impact of participation in apprenticeship on wages and occupational status outcomes compared to those obtaining vocational qualifications through school-based VET provision is less clear and often negative for some groups, such as women, in some countries. Compared to early participation in the labor market and labor-market VET programs the outcomes of apprenticeship are positive (Ryan, 2001).

Labor-Market Policies

To combat joblessness, a range of labor-market policies has been implemented. Passive policies aim at deregulating

the labor market to make it more flexible in order to help people move into work, for example, increasing pay flexibility. Such policy interventions enable employers in offering younger workers lower wages, and weaker employment contracts often on a fixed-term basis. The outcome of such policies may have led to improved employment prospects for young people, as is claimed for Canada (OECD, 2008a). But problems of measurement and selection effects make evaluation results difficult to interpret (Ryan, 2001). Other passive labor-market measures involve changes to rules about access to unemployment and other social security benefits, for instance, in the UK changes in such rules make those aged less than 18 years ineligible for unemployment benefit.

Active Labour Market Programmes (ALMPs) are more voluntary policies, often targeted at specific groups which aim to promote labor-market entry and access to better employment. They include a variable mixture of training and basic skills (language and mathematics) education, work experience, job-search assistance, and direct job creation. Only education and training measures are considered here. Education and training form the active measure which attracted the greatest investment in the European Union 15 (EU-15) countries in 2003 and the second most important in accession states after job creating through subsidized employment and public work opportunities (Descy and Tessaring, 2007).

The shape and role of ALMPs aimed at youth varies depending upon a country's transition structure. For example, there seems to be a greater propensity to intervene in transitions in country's with stricter labor-market regulation unless the barriers to entry are reduced by strong school-employment linkages. Those countries with the highest barriers to labor-market entry and with subsequent large-scale ALMPs tend to target broad groups of those at risk of making poor transitions (Schröder, 2003).

European-training-based ALMPs do seem to yield more positive results in most cases compared to other types, especially job-creation programs (Hujer *et al.*, 2004). However, Martin and Grubb (2001) report that although training measures are the most expensive types of ALMPs, their efficacy is far from clear. Descy and Tessaring's (2005) review of selected studies carried out in North European countries suggest that it is difficult to draw firm conclusions about the effectiveness of training measures from evaluation results. These are contingent on external labor-market conditions and the type of target group.

Conclusions

Some young people still find it difficult to make the transition from schooling to work. The magnitude of this effect varies between countries and may be linked to the regulatory frameworks surrounding their labor markets and other aspects of the design of their national transition

systems. For most young people this is not a crisis but for disadvantaged youth the transitions can be especially problematic. Furthermore, we still have much to understand about the reasons for cross-national variation in school-to-work transition indicators, and this requires further research working with better-quality longitudinal data sets.

Despite widespread support for the further development of upper-secondary VET programs to retain young people for longer in the education system, the impact of such initiatives is highly variable, often zero and sometimes negative. Apprenticeship does seem to provide a more certain route to employment, but it still has certain drawbacks and is proving difficult to sustain in the face of global economic pressures. VET provided through ALMPs is the most variable and ambiguous in its impact. Collectively, and given the expense of implementing good-quality VET programs, the research calls into question the effectiveness and efficiency of VET in supporting young people experiencing difficulties with school-to-work transition to enter the labor market. Nonetheless, providing more VET remains a popular policy choice because of the political need to do something about youth joblessness among certain segments of the population, and the undoubted motivational impact of more practical learning opportunities for many young people.

Part of the problem with supporting policy learning in this area is the way that VET initiatives are often designed, implemented, and evaluated. Often such initiatives are conceived as entities that can be developed quickly, implemented alongside other social institutions, and then evaluated independently of those other social institutions. This is what Grubb and Ryan (1999) term as program view, whereby VET programs are developed and evaluated as self-contained and independent entities. Energy is put into the continual development of new initiatives and their frequent abandonment after relatively short periods of time when program evaluation indicates their relative ineffectiveness.

This can be contrasted with a systems perspective, which views VET programs developing within a larger system that evolves synergistically over time. Energy is now devoted to institution building over a longer period of time that focuses on the linkage between learning opportunities that enables individuals to develop the knowledge, skills, and competences needed to enter and be successful in the labor market over the course of their working lives. Systemic evaluations then focus on a program's success (and the reasons for its success) in terms of articulation with other VET programs, linkage to employers' recruitment and hiring practices, its relationship with other established institutions and practices, and its sustainability and potential for replication.

See also: Apprenticeship Approach to Learning; Apprenticeships; Assessment in Vocational Education; Career

Guidance in Vocational Education and Training and in the Workplace; Displaced Workers, Unemployed and Vocational Education and Training; Dual System; Globalization and Vocational Education and Training; Industry Involvement in the Vocational Education and Training System; Modularization in Vocational Education and Training; Non-Completion in Vocational Education and Training; Issues, Patterns and Possibilities; Personalized Learning and Vocational Education and Training; Planning and Policy Development for Technical Vocational Education and Training Systems; Public Policy and Inequality in Postsecondary Opportunity: Educational Statistics and the Failure of Education Reform; Qualifications Frameworks and their Role in the Reform of Education and Training; Quality Assurance in Vocational Education and Training; Skills Shortages: Concepts, Measurement and Policy Responses; The Competency Model; The Status of Vocational Education and Training; Training Markets; Vocational Teacher Education.

Bibliography

- Anlezark, A., Karmel, T., and Ong, K. (2006). *Have School Vocational Education and Training Programs been Successful?*, Adelaide: NCVER. <http://www.ncver.edu.au/publications/1654.html> (accessed August 2009).
- Dearden, L., McIntosh, S., Myck, M., and Vignoles, A. (2000). *The Returns to Academic and Vocational Qualifications in Britain*. London: Centre for the Economics of Education.
- Descy, P. and Tessaring, M. (2007). Combating labour market exclusion: Does training work? *European Journal of Vocational Training* 41, 64–83.
- Furlong, A. and Cartmel, F. (1997). *Young People and Social Change: Individualization and Risk in Late Modernity*. Buckingham: Open University Press.
- Grubb, W. N. and Ryan, P. (1999). *The Roles of Evaluation for Education: Plain Talk to the Field of Dreams*. London: Kogan Page/ILO.
- Hall, P. A. and Soskice, D. (eds.) (2001). *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. Oxford: Oxford University Press.
- Hayward, G. (2005). *A Systems Analysis of the English VET System*. London: Learning and Skills Development Agency.
- Hayward, G., Hodgson, A., Johnson, J., et al. (2005). *Nuffield 14–19 Review Annual Report*. Oxford: OUDES.
- Hayward, G., Hodgson, A., Johnson, J., et al. (2006). *Nuffield 14–19 Review Annual Report*. Oxford: OUDES.
- Hayward, G., Wilde, S., and Williams, R. (2008). *The Engaging Youth Enquiry Final Report*. <http://www.nuffield14-19review.org.uk/files/documents196-1.pdf> (accessed August 2009).
- Hölscher, M., Hayward, G., Ertl, H., and Dunbar-Godet, H. (2008). The transition from vocational education and training to higher education: A successful pathway? *Research Papers in Education* 23(2), 139–152.
- Hujer, R., Callendo, M., and Radi, D. (2004). Methods and limitations of evaluation and impact research. In Descy, P. and Tessaring, M. (eds.) *The Foundations of Evaluation and Impact Research: Third Report on Vocational Training Research in Europe: Background Report*. Luxembourg: Office for Official Publications of the European Communities.
- Ianelli, C. and Raffae, D. (2007). Vocational upper-secondary education and the transition from school. *European Sociological Review* 23(4), 49–63.
- Iversen, T. (2005). *Capitalism, Democracy and Welfare*. Cambridge: Cambridge University Press.
- Jenkins, A., Greenwood, C., and Vignoles, A. (2007). *The Return to Qualifications in England: Updating the Evidence Base on Level 2 and Level 3 Vocational Qualifications*. London: Centre for the Economics of Education.
- Keep, E. (2009). *Internal and External Incentives to Engage in Education and Training – A Framework for Analysing the Forces Acting on Individuals?* SKOPE Monograph 12: Oxford and Cardiff University.
- Marks, G. N. (2005). Issues in the school-to-work transition. *Journal of Sociology* 41(4), 363–385.
- Martin, J. D. and Grubb, N. (2001). What works and for whom: A review of OECD's countries experiences with active labour market policies. *Swedish Economic Policy Review* 8(2), 9–56.
- McIntosh, S. (2007). Apprenticeship and other vocational qualifications: A cost-benefit analysis. *Department for Education and Skills Research Report 834*. Nottingham: DFES.
- OECD (2002). *Employment Outlook*. Paris: OECD.
- OECD (2007). *Education at a Glance*. Paris: OECD.
- OECD (2008a). *Jobs for Youth: Canada*. Paris: OECD.
- OECD (2008b). *Jobs for Youth: Netherlands*. Paris: OECD.
- Pring, R., Hayward, G., Hodgson, A., et al. (2009). *Education for All: The Future of Education and Training for 14–19 Year Olds*. London: Routledge.
- Quintini, G. and Martin, S. (2006). Starting well or losing their way? The position of youth in the labour market. *OECD Social, Employment and Migration Working Papers*. Paris: OECD.
- Ryan, P. (2001). The school to work transition: A cross-national perspective. *Journal of Economic Literature* 39, 34–92.
- Ryan, P. (2003). Evaluating vocationalism. *European Journal of Education* 38, 147–162.
- Schröder, L. (2004). The role of youth programmes in the transition from school to work. *Working Paper 2004:5*. Uppsala: Institute for Labour Market Policy Evaluation.
- Stanley, G. (2007). Education for work: The current dilemma of post-compulsory education. *Australian Educational Researcher* 34(3), 91–100.
- Vickerstaff, S. A. (2003). Apprenticeship in the 'golden age': Were youth transitions really smooth and unproblematic back then? *Work, Employment and Society* 17(2), 269–287.

Further Reading

- Grubb, N. and Lazerson, M. (2004). *The Education Gospel: The Economic Power of Schooling*. Cambridge, MA: Harvard University Press.
- Jung, T.-H., Misko, J., Lee, K., et al. (2004). *Effective Measures for School-to-Work Transition in the Vocational Education System: Lessons from Australia and Korea*. Adelaide: NCVER.
- Lafer, G. (2002). *The Job Training Charade*. Ithaca, NY: Cornell University Press.
- Muller, W. and Ganl, M. (eds.) (2003). *Transitions from Education to Work in Europe: The Integration of Youth into EU Labour Markets*. Oxford: Oxford University Press.
- OECD (2000). *Thematic Review of the Transition from Initial Education to Working Life*. Paris: OECD.
- OECD (2005). *From Education to Work: A Difficult Transition for Young Adult with Low Levels of Education*. Paris: OECD.

Relevant Websites

- <http://cedefop.europa.eu> – CEDEFOP – European Centre for the Development of Vocational Training.
- <http://cee.lse.ac.uk> – Centre for the Economics of Education.
- <http://www.skoep.ox.ac.uk> – ESRC Research Centre on Skills, Knowledge and Organisational Performance (SKOPE).
- <http://www.ilo.org> – International Labour Organization.
- <http://vocserve.berkeley.edu> – National Centre for Research in Vocational Education 1988–199.
- <http://www.ncver.edu.au> – National Centre for Vocational Education Research.
- <http://www.nrccte.org> – National Research Centre for Career and Technical Education.
- <http://www.nuffield14-19review.org.uk> – Nuffield 14–19 Review.
- <http://www.oecd.org> – Organization for Economic Cooperation and Development (OECD).

VOCATIONAL EDUCATION AND TRAINING – INDUSTRY AND EMPLOYERS

Contents

Apprenticeships

Skills Shortages: Concepts, Measurement and Policy Responses

Training and Learning in the Workplace

Training in Small Businesses

Apprenticeships

E Smith, University of Ballarat, Ballarat, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Apprenticeship is an institution that, for centuries, has successfully effected entry into working life for young people, and has also been responsible for the maintenance of the skills base of many national economies. Apprenticeships began in medieval Europe when young people went to live in their masters' houses to learn trades, over a period of up to 7 years. Although apprenticeships have become less demanding of both master and apprentice, they have survived in many countries over the centuries (Lane, 1996). A typical dictionary definition of an apprentice (*Chambers Twentieth Century Dictionary*: 6) is "one bound to another to learn a craft," an apprenticeship is "the state of an apprentice: a time of training for a trade, or for any activity." There are two key points in these definitions: one is the employment relationship, and the closeness of that relationship. The apprentice is bound to the master, often by a formal contract and by unwritten understandings as well. The other key point in the above definitions is time: the focus is not only on what is learned, but also on the time period over which it is learned.

This article examines the meaning of apprenticeship, perspectives in the apprenticeship literature, the different nature of apprenticeship in different countries, and some challenges with which apprenticeship systems are grappling.

Formal and Informal Meanings of Apprenticeships

The term apprenticeship is used in many contexts. It is often used informally to describe any process of learning a job or a skill from another person, generally in the context of an older or more experienced person mentoring or coaching a younger and more junior person. It is also

used to describe processes of learning at work which are primarily work-based but include elements of formal off-the-job learning. For example, professional occupations such as nursing, accountancy, and law formerly followed this model, although they tend to be university-based these days. However for the purposes of this article we confine the discussion of apprenticeship to formal systems that aim to develop skilled workers, and that are in occupations serviced by the vocational education and training system rather than by university education.

The essential components of a formal apprenticeship are generally understood to be:

- a training regime set up by, or with the approval of, governments;
- a combination of off- and on-the-job training;
- the assumption of responsibility by the employer for the development of the apprentice; and
- the award of a qualification and/or licence and/or some other recognition that enables an occupation to be practiced independently once the apprenticeship is successfully completed.

Apprenticeships are often, but not always, intended for young people rather than older people, and often, but not always, incorporate a close relationship between a novice and a particular expert worker. They normally, but not always, involve the apprentice being actually employed in the enterprise where on-the-job training is carried out.

Perspectives on Apprenticeships

The literature reveals six major ways of looking at apprenticeships. These different viewpoints to some extent reflect the presence of different stakeholders in the

apprenticeship system, such as trade unions, educationalists, training policymakers, and youth commentators, and policymakers. These viewpoints are discussed briefly below.

Apprenticeship as a Passage to Adulthood

Many writers (e.g., Hamilton, 1990, in the United States and Sweet, 1995, in Australia) have been concerned with the uncertain transition from school to work for many young people. For writers with these concerns, apprenticeship is seen as being a safe way for young people to enter employment, providing job security and guaranteed development of skills. The notion of passage to adulthood incorporates the key point of time served. Worries about young people's employment future have become less prominent in some economies (e.g., UK, Australia) that have experienced economic boom during the twenty-first century, but in some countries that still have high unemployment, such as Germany, this issue is still of importance (Deissinger and Hellwig, 2005).

Apprenticeship as a Means of Industry and National Skill Formation

Gospel (1994: 51) maintains that "apprenticeship has constituted a collective source of skill formation supplying recognised skills to firms of all sizes." Many firms have long-established apprenticeship systems; in Australia and the UK, these systems became less important during the 1980s, partly as the result of economic recession but such systems have become re-established in many major Australian companies over the past decade. Governments have a considerable interest in monitoring and encouraging apprenticeships to assure the maintenance of the skill base of economies. However, not all countries use apprenticeship as a means of national skill formation. In the US, for example, the apprenticeship system is rudimentary and confined to only a few trades; there are fewer than half-a-million apprentices and they comprise only 0.3% of the workforce (Glover *et al.*, 2007: 477–478). Some authors regard the institution of apprenticeship as outdated and wasteful as a means of skill formation, because of the long time taken for apprentices to become fully skilled (e.g., Sweet, 1987). It has been argued that apprenticeship training does not coincide with the demand for skills: apprenticeship commencements always fall during periods of recession, but 3 or 4 years later, when apprentices finish their time, the recession may have ended and there is thus a skills shortage. This argument is, of course based on the premise that apprentices do not start doing useful work until the day they qualify, whereas many studies (e.g., the classic English study undertaken by Venables, 1967 and an Australian study by Smith, 1998) indicate that apprentices perform useful work from their early days.

Occupational Identity and Being a Craftsman

The completion of an apprenticeship is felt to confer a status in society (Unwin, 1996). An important part of apprenticeship is assuming this status. Shields (1992) maintains that, in the nineteenth century, "membership of the trade carried strong ideological and moral overtones ... 'the time-served man (was) set apart from 'inferior' workers.'" Therefore, apprentices have traditionally been taught not only the skills involved in the trade but also how to become a member of that trade (Venables, 1967; Smith, 1992). As McIntyre (1996: 44) puts it, "A person assumes the identity of (a carpenter, for example) and learns both to carry out activities (the practice) and learns the meaning of the practice." Becoming a tradesperson is signified by such rites of passage as acquiring the tools of one's trade and wearing particular types of clothing (Riemer, 1977). Winning (1993) goes further, suggesting that choosing to become a tradesperson involves the choice of a particular way of life. Brown (1997), however, has argued against assuming that occupational identity is central to all apprentices and tradespeople.

Apprenticeship as a Device for Occupational Restriction

Apprenticeships are often described, particularly in the industrial relations literature, as being a device for restricting the entry to certain trades, to ensure that the price of adult or skilled labor is kept artificially high. To assist this process, the number of apprentices is restricted, the employment of non-apprentice junior labor is prohibited, and there is a lengthy time needed to become skilled (Shields, 1995). It is often maintained that the skill in apprenticed occupations is socially constructed. Apprenticeship has been seen as a form of ritual servitude that legitimated an illusory division between skilled and unskilled work (Shields, 1995: 239). Feminist critiques focus on the fact that in Anglophone countries it is mainly male trades which have had the power to construct their trades as skilled while the skill involved in female work is undervalued (Korczynski, 2005). Littler (1982: 10–11) has provided a seminal discussion of this topic from a labor-process standpoint. During World War II, in Australia as in many other countries, women undertook many of the work roles previously undertaken by men who had been apprenticed. After the War, women were forced out of these roles and those people who had learned some of the job roles without undertaking a full apprenticeship were dismissively described as dilutees (Ray, 2001).

Economic Arguments about Apprenticeship

According to human capital theory (e.g., Becker, 1964), apprenticeship is a form of general training, and so the

cost should primarily be borne by the employee not the employer, in the form of low wage rates while in training. Historically, apprentices were often not paid (Lane, 1996), since the master provided all living expenses. Indeed, parents in medieval England often paid substantial sums of money to employers who were willing to train their children in the more desirable trades (Lane, 1996: 19). However, in order to attract people into apprenticeships, apprentices in the modern day generally have to be over-paid in terms of what human capital theory sees as their worth to employers. Without governmental subsidies, it is argued, employers would not employ apprentices. It is also argued that it is for this reason that apprenticeships are so long; employers require a long period of indentured service to recoup some of the loss in the early years. Empirical studies have been undertaken to investigate economic reasons for hiring of apprentices; for example, Dockery *et al.* (1998). Such studies generally conclude in bewilderment that, since there is little economic benefit to an employer in taking on an apprentice, other, non-economic factors must be involved. And, indeed, studies have shown that employers who take on apprentices tend to have a strong normative commitment to the notion of apprenticeships and the maintenance of a supply of skilled workers in the appropriate industries (e.g., Smith, 1998). Thus, it is doubtful therefore whether the level of wages of apprentices is really a serious issue either one way or the other. However, there seems to be more concern about the cost-benefits of apprenticeships in the dual-system countries (Walther *et al.*, 2005; Grollmann and Rauner, 2007).

Pedagogical Issues

Although learning is the basis of apprenticeship, there is relatively little literature on this issue compared with some of the other perspectives. Much of the pedagogical literature on apprenticeship focuses on the relationship between off-the-job and on-the-job learning. It is generally held that on-the-job training provides practical learning while theoretical learning is best undertaken off the job (Uwameiye and Iyamu, 2002). It is sometimes said that it is ideal for on-the-job experiences to enable apprentices to practice the exact skills that are concurrently learned at the training provider, but Australian studies of on- and off-the-job learning (e.g., Harris *et al.*, 1998; Smith, 2002) tend to conclude that apprentices are well able to cope with learning that takes place in different arenas and are able to integrate these into their own practice of the trade. There have been some studies specifically of off-the-job learning in apprenticeships; one interesting but not altogether surprising conclusion is that apprentices tend to prefer hands-on experiential learning rather than theory-based learning (Smith, 2003). Studies of on-the-job learning emphasize the progressive skilling of

apprentices (Smith, 1998), the introduction to a community of practice (Unwin and Fuller, 2003), and the importance of the planning of the work apprentices should do so that they gain a wide range of experiences. Such arrangements are described as expansive rather than restrictive learning environments by Unwin and Fuller (2003).

Different Models of Apprenticeships

Different countries have different expectations of apprenticeships and therefore regulate and manage their apprenticeship systems in different ways. This section provides a discussion of five issues: qualifications, employment status, examination of proficiency, coverage of occupations, and funding regimes. A case study of Australia follows which illustrates how these different issues interrelate to form a system of great complexity.

Qualifications

Earlier in this article, we have included the award of a qualification and/or licence as a common feature of formal apprenticeships in the modern era although in some countries such as Nigeria there is still no qualification involved (Evawoma-Enuku and Mgbor, 2005). In Australia, all apprenticeships and traineeships (see section entitled 'A case study of apprenticeship systems: Australia', for a discussion of traineeships) provide a formal qualification, usually at Certificate III level or higher. The curriculum for qualifications for apprenticeships and traineeships consists of units of competency taken from the sets of competency standards in national training packages (Smith and Keating, 2003). In general, apprentices attend a technical and further education (TAFE) college (TAFE – the public providers) or a private training provider on 1-day-a-week basis or for block periods, for 2 or 3 years. Trainees may also attend college in this way, but it is becoming increasingly common for trainees to be trained 100% on the job. However, even in the latter case, a training provider (known as a Registered Training Organisation (RTO)) must oversee the training and is responsible for the assessment and the award of the qualification. Similar diverse systems apply in the UK, with various ways of gaining the required qualification of a national vocational qualifications (NVQ) (Fuller and Unwin, 2007). In Germany, by contrast, the availability of qualifications is more regulated. Under the dual system, all apprenticeships involve off-the-job training and such training is only provided by public schools. In Australia and the UK, there is not usually any regulation associated with the on-the-job training provided by the employer, but in Germany there is regulation of the on-the-job training (Grollmann and Rauner, 2007).

Employment Status

As mentioned previously, most apprenticeships involve a contract of employment so that an apprentice is primarily a worker rather than a student. In some countries, training providers take a more central role. In the Netherlands, for example, apprenticeships may be training-provider-based or work-based; in training-provider-based apprenticeships, on-the-job training takes place in work placements rather than as a formal employment contract (Onstenk and Blokhuis, 2007). In Australia, would-be apprentices may commence part of the apprentice qualification through a pre-apprenticeship course at a training provider, (Dumbrell and Smith, 2007) but must gain employment before proceeding very far toward the qualification. It is thus assumed in Australia that the qualification is only of full utility when combined with considerable time spent on the job learning at work. In Germany, high unemployment has led to the growth of off-the-job training in full-time vocational schools alongside the dual system route, but the former suffer from a perception of low status (Deissinger *et al.*, 2006).

Examination of proficiency

In Australia, the completion of an apprenticeship formerly involved a trade test administered by the state training authority, along the lines of the final examinations managed by the Chambers in Germany (Grollmann and Rauner, 2007) but generally these days, in Australia, it is assumed that completion of the appropriate qualification negates the need for an additional trade test, although this assumption is not unchallenged. The award of a license to practice a trade is currently, in Australia, a separate process from the award of the qualification, with licensing bodies guarding their prerogative to decide who can practice a trade, although there have been attempts by national governments over the past 20 years to align the qualification-awarding and the licensing processes. In the UK, there has been some controversy over whether an NVQ is sufficient to pronounce an apprentice proficient.

Coverage of Occupations

As Western economies continue to move away from primary and secondary industries toward the service sector (Triplett and Bosworth, 2004; Barnes and Kennard, 2002), the apprenticeship system in some countries has struggled to meet these changing times. In the dual-system countries, apprenticeships have always covered a range of occupations including the service sector, and there is a system in place to incorporate new occupations into the apprenticeship system. But in other countries such as the UK, Australia, and the United States, systems have

needed to adapt and change. In Australia, the introduction of traineeships has addressed the need for apprentice-like arrangements in a wider range of occupational areas. Traineeships are discussed in detail in the Australian case study below. A similar process was undertaken in the UK, where modern apprenticeships (Fuller *et al.*, 2005) were introduced in the early 1990s by the UK government to broaden the reach of apprentice-like arrangements. As in Australia, the UK modern apprenticeships also had roots in the youth training schemes introduced at times of high youth unemployment, particularly in the early 1980s. This made them initially unpopular (Fuller *et al.*, 2005). In both of these countries, the newer apprentice-like arrangements have subsequently been moved under a broad umbrella simply called apprenticeships. In the US, only a limited range of occupations is covered (Glover *et al.*, 2007).

Funding Regimes

In order to support their apprenticeship systems, countries have many methods of funding apprenticeships. These funding arrangements are designed variously to encourage employers to employ apprentices, to provide funds for training providers to undertake the off-the-job training, and (for individual apprentices) to provide a living allowance for apprentices where there are no wages, or to provide supplementation of low wages. For example, in Nigeria, apprentices receive a monthly stipend equivalent to about 15% of the national minimum wage) and employers receive a payment for training an apprentice (Evawoma-Enuku and Mgbor, 2005). In the UK, apprentices who do not have employed status receive a training allowance (Fuller and Unwin, 2007) and funding is given to training providers under contract to the government. The Australian government provides funding to employers and training providers, as described in the detailed case study below, and also provides minor allowances such as tools allowances and living-away-from-home allowances for apprentices who must move to find employment as an apprentice. By contrast, in the US there is little government investment in apprenticeships. The federal government sponsors some minor programs and some state governments provide some funding for off-the-job training; otherwise, all costs are borne by the employer and sometimes trust funds created from a levy of employers (Glover *et al.*, 2007: 478).

A Case Study of Apprenticeship Systems: Australia

This section provides a description of the apprenticeship system in Australia, illustrating the interrelationship

between the six issues discussed in sections 'Apprenticeship as a passage to adulthood' through 'Pedagogical issues' and the additional points raised in sections 'Qualifications' through 'Funding regimes' that mark the differences among countries' systems. In Australia, the institution of apprenticeship is currently very strong. Twenty years ago apprenticeships in Australia were confined to a defined number of occupations, mainly male manual workers, but the advent of traineeships (which are included with traditional apprenticeships under the broad umbrella term Australian apprenticeships) has expanded both the numbers of apprentices and the types of jobs which have contracted training associated with them. This success story has been the product of very conscious planning by the federal government including the introduction of new agencies to promote apprenticeships and manage their quality. These agencies sit alongside preexisting organizations and mechanisms at the federal and state level.

In Australia, the apprenticeship system involving 3- or 4-year contracts of training in the traditional trades has existed since first settlement by Europeans. In 1985, short, 1- and 2-year traineeships (Kirby, 1985) were introduced. Traineeships expanded into many occupational areas that had not previously supported contracted training such as retail, tourism, and hospitality (Robinson, 2001). In 1997, the traditional apprenticeship and the traineeship systems were brought together under the umbrella of the new apprenticeship, now called Australian apprenticeship, system, although in common usage they are usually referred to separately (Dumbrell and Smith, 2007). The numbers of Australian apprenticeships escalated dramatically from about 120 000 in 1995 to over 400 000 by 2003, fueled mainly by traineeship growth. Around 35% are 4-year apprentices in traditional trade areas while the remainder are trainees (NCVER, 2004). The development of training packages – national sets of competency standards – for these occupational areas also stimulated growth, providing the basis for apprenticeship and traineeship qualifications. Training packages are developed by national industry skills councils, each overseeing a range of industry and occupational areas.

The proportion of workers in Australian apprenticeships represents 3.5% of the working-age population, one of the highest rates of contracted training in the developed world (Walters, 2003). This favorable picture is in part related to the strong Australian economy. However, the high proportion of workers in apprenticeships is also the result of very deliberate government policies over the past 20 years. These have included the widening of apprenticeship opportunities to part-time and mature, aged workers, and the availability of state government funding for off-the-job training by private RTOs as well as by the public provider, TAFE (Smith and Keating, 2003). This process, whereby employers, in conjunction supposedly with the apprentice himself or herself, are able to

select the RTO of their choice, is known as user choice. The availability of such funds to private training providers has enabled massive expansion in areas such as retail and aged care, where TAFE would not have been able to meet the demand. In addition to these policies, the growth of new bodies to manage apprenticeships has been important.

There are a number of regulatory arrangements associated with Australian apprenticeships. Contracts of training must be signed by employers, by employees (and by parents where the employees are aged under 18), and by the training provider (RTO). The contracts are registered with the state or territory training authority. Employment incentives are supplied by the federal government on commencement and completion, and off-the-job training is funded by the state training authority. State training authorities and the federal government alike maintain regional and local offices where staff work to promote apprenticeships and to manage the quality of apprenticeships. In addition to these long-established processes, apprenticeships are now promoted through school education systems (which in Australia are managed by state governments) because apprenticeships can be commenced on a part-time basis while students are still at school.

There have been a number of quality problems associated with the rapid growth of the apprenticeship and traineeship system in Australia (Schofield, 1999; Snell and Hart, 2007). These have been partially addressed by new policies, such as the introduction of the Australian quality training framework (AQTF) which aims to ensure good quality training in TAFE and RTOs by regulating the registration of training providers and the delivery of training (Smith and Keating, 2003). The AQTF, through state training authorities, is responsible for the quality of all vocational qualifications, not just those associated with Australian apprenticeships. The weighting of employment incentives toward completion of apprenticeships, instead of equal payments on commencement and completion, has also tended to improve quality. There is still a perception, however, among some commentators that employers of trainees may be overly influenced by the availability of employment subsidies (Snell and Hart, 2007).

There are two sets of agencies that directly contribute to the apprentice system and a number of others that make an indirect contribution. The two direct contributors are group training organizations (GTOs) formerly known as group training companies (GTCs) and Australian apprenticeship centers (AACs) formerly known as new apprenticeship Centers (NACs).

GTOs act as employers of apprentices, leasing them out to companies and thereby relieving companies both of the risk of taking on an apprentice for a lengthy period and of the paperwork associated with employing an apprentice (Dumbrell and Smith, 2007). There are 180 GTOs in Australia and they receive government funding through the joint group training program scheme whereby funding is

allocated primarily on the basis on numbers of apprenticeships and traineeships commenced and completed. JGTP funding is provided equally by federal and state governments, and around \$20 million a year flows to GTOs through JGTP (Hood *et al.*, 2007). A GTO, as the employer of the apprentice, also receives the normal government employment incentive. GTOs receive payments from the host employers, but this is usually only just enough to cover the wages that GTOs pay to the apprentices.

AACs are newer than GTOs and were set up in the mid-1990s to increase the number of people entering apprenticeships. AACs, acting under contract to the federal government, market apprenticeships to potential employers and apprentices, manage the signing-up process, and make sure that appropriate employment and completion incentives are paid. They also make employers aware of special incentives that may be available for employing apprentices from disadvantaged groups, for example, indigenous or disabled people. AACs also have a role in making sure that the employer-apprentice relationship proceeds smoothly and to report any problems to the appropriate authority, normally the local office of the state training authority (Smith *et al.*, in progress).

Besides these agencies, other agencies have some role in promoting apprenticeships. These are funded by either state or federal government and may also earn income through commercial activities. They include:

- RTOs. They have an interest in employers recruiting apprentices, because they can then access user choice funding by providing the training for the apprentices.
- Job network providers. These agencies provide an employment brokerage service. Often they place their clients in jobs that include a contract of training.
- Industry skills councils (ISCs). There are ten national industry skills councils covering the range of Australian industry, and in some states there are state counterparts. They promote apprenticeships and traineeships to industry because then there will be greater take-up of the training packages which ISCs oversee.

Challenges for Apprenticeship Systems

While only a relatively small number of countries have been mentioned in this article, the broad range of apprenticeship models discussed enables some general points to be made. Apprenticeship systems have been designed to serve many overt purposes: to transition young people into working life, to provide skills to the workforce, to maintain quality of skills, and to reduce unemployment rates. They also perform more covert purposes such as valorizing the status of some occupations above others and restricting access to some occupations. These many purposes are the

reason why there are many stakeholders in apprenticeships. This is a good thing because it means that apprenticeships are supported by diverse groups and are very firmly rooted into economies and societies. However, it is also a bad thing because change to apprenticeship systems can be fiercely resisted for a range of reasons. The number of stakeholders also means that apprenticeships are culturally specific and that it is difficult to read across from one country to another (Deissinger *et al.*, 2006).

Some challenges faced by apprenticeship systems include:

- Their ability to cope with fluctuations in economic prosperity and/or in the rate of unemployment. While apprenticeships are to some extent designed to soften peaks and troughs in employment, their role in this respect tends to be reactive rather than proactive. It is difficult in periods of economic boom for apprenticeships to provide qualified tradespeople at a high-enough rate to meet labour-market demands; and because apprenticeships are usually employment-based, numbers tend to drop when economic times are hard because employers cannot afford to hire them. Fairly robust government intervention is required to address these issues and proffered solutions may not be attractive to all stakeholders. In the end, apprenticeships rely, in most countries, on employers' willingness to employ apprentices, and so governments cannot effect rapid change autonomously.
- The question of 'Who pays?' While in some countries governments accept a considerable amount of funding responsibility for apprenticeship, in other countries this is not so. The Australian experience represents the results of heavy financial investment by the national government, but some argue that employers should take more financial responsibility.
- The extent to which systems embrace changing occupational patterns. While some countries have straightforward means of adding new occupations to the list of apprenticed trades, in others, the addition of occupations, especially where they compete for existing funding, may be resisted by interest groups representing traditional occupations.
- The extent of regulation. Regulatory requirements may apply to qualifications, contracts of employment, requirements for on-the-job training, wage rates, and so on. The high degree of regulation in some systems can be seen as problematic by employers and training providers.
- The invisible ingredient: pedagogy. The regulatory panoply associated with apprenticeships and the range of stakeholders mean that it is all too easy to see apprenticeship as a public policy artifact rather than a pedagogical process. In the end, though, it is the quality of learning that is important, and due attention needs to be paid to pedagogy.

Apprenticeships have shown themselves able to adapt to change over the centuries but some countries have managed change more effectively and with less conflict than others. Flexibility is difficult where stakeholders have firmly entrenched interests. There remains the issue of how much apprenticeships can change while still retaining their essential nature, and this core will always vary from country to country.

See also: Apprenticeship Approach to Learning; Australia; Dual System; Industry Involvement in the Vocational Education and Training System; Planning and Policy Development for Technical Vocational Education and Training Systems.

Bibliography

- Barnes, P. and Kennard, S. (2002). *Skill and Australia's Productivity Surge*. Canberra: Productivity Commission.
- Becker, G. S. (1964). *Human Capital: A Theoretical Analysis with Special Reference to Education*. New York: Columbia University Press.
- Brown, A. (1997). A dynamic model in occupational identity formation. In Brown, A. (ed.) *Promoting Vocational Education and Training: European Perspectives*, pp 59–67. Tampere: EUROPROF.
- Deissinger, T. and Hellwig, S. (2005). Apprenticeships in Germany: Modernising the dual system. *Special Issue: Apprenticeships. Education and Training* 47(4/5), 312–324.
- Deissinger, T., Smith, E., and Pickersgill, R. (2006). Models of full-time and part-time vocational training for school-leavers: A comparison between Germany and Australia. *International Journal of Training Research* 4(1), 30–50.
- Dockery, A. M., Koshy, P., Stromback, T., and Ying, W. (1998). The cost of training apprentices in Australian firms. *Australian Bulletin of Labour* 23(4), 255–274.
- Dumbrell, T. and Smith, E. (2007). *Pre-Apprenticeships in Three Key Trades*. Adelaide: National Centre for Vocational Education Research (NCVER).
- Evawoma-Enuku, U. and Mgbor, M. (2005). The National Directorate of Employment's Open Apprenticeship Scheme in Nigeria: New wine in old wineskins? *Special Issue: Apprenticeships Education and Training* 47(4/5), 325–336.
- Fuller, A. and Unwin, L. (2007). What counts as good practice in contemporary apprenticeships? Evidence from two contrasting sectors in England. *Special Issue: Quality in Apprenticeships. Education and Training* 49(6), 447–458.
- Fuller, A., Beck, V., and Unwin, L. (2005). The gendered nature of apprenticeship: Employers' and young people's perspectives. *Special Issue: Apprenticeships. Education and Training* 47(4/5), 298–311.
- Glover, R. W., Clopton, L., McCollum, M., and Xinge, W. (2007). Building an apprenticeship and traineeship system for maintenance occupations in the American transit industry. *Special Issue: Quality in Apprenticeships. Education and Training* 49(6), 474–488.
- Gospel, H. (1994). The survival of apprenticeship training in Australia. *Journal of Industrial Relations* 36(1), 37–56.
- Grollmann, P. and Rauner, F. (2007). Exploring innovative apprenticeship: Quality and costs. *Special Issue: Quality in Apprenticeships. Education and Training* 49(6), 431–446.
- Hamilton, S. F. (1990). *Apprenticeship for Adulthood*. New York: Free Press.
- Harris, R., Willis, P., Simons, M., and Underwood, F. (1998). *Learning the Job: Juggling the Messages in on- and off-the-Job Training*. Adelaide: NCVER.
- Hood, S., Fung, H., Smith, E., Bush, A., and Ride, L. (2007). *Review of the Joint Group Training Program*. Canberra: Department of Education, Science and Training.
- Kirby, P. (1985). *Report of the Committee of Inquiry into Labour Market Programs*. Canberra: AGPS.
- Korczynski, M. (2005). Skills in service work: An overview. *Human Resource Management Journal* 15(2), 3–14.
- Lane, J. (1996). *Apprenticeship in England, 1600–1914*. London: UCL Press.
- Littler, C. (1982). *The Development of the Labour Process in Capitalist Societies*. London: Heinemann.
- McIntyre, J. (1996). On becoming a mediator: Reflections on adult learning and social context. In Willis, P. and Neville, B. (eds.) *Qualitative Research Practice in Adult Education*. Melbourne: Lovell Publishing.
- NCVER (2004). *Australian Vocational Education and Training Statistics: Apprentices and Trainees, December 2003*. Adelaide: NCVER.
- Nechvoglod, L., Karmel, T., and Saunders, J. (2009). *The Cost of Training Apprentices*. Adelaide: NCVER.
- Onstenk, J. and Blokhuis, F. (2007). Apprenticeship in the Netherlands: Connecting school- and work-based learning. *Special Issue: Quality in Apprenticeships. Education and Training* 49(6), 489–499.
- Ray, J. (2001). Apprenticeship in Australia: A concise history. In Smart, N. (ed.) *Australian Apprenticeships: Research Readings*, pp 15–43. Adelaide: NCVER.
- Riemer, J. W. (1977). Becoming a journeyman electrician: Some implicit indicators in the apprenticeship process. *Sociology of Work and Occupations* 4(1), 87–99.
- Robinson, C. (2001). *Australian Apprenticeships: Facts, Fiction and Future*. Adelaide: NCVER.
- Schofield, K. (1999). *Independent Investigation into the Quality of Training in Queensland's Traineeship System: Final Report: 2 July 1999*. Brisbane: Queensland Department of Employment, Training and Industrial Relations.
- Shields, J. (1992). Craftsmen in the making: The memory and meaning of apprenticeship in Sydney between the Great War and the Great Depression. In Shields, J. (ed.) *All Our Labours: Oral Histories of Working Life in Twentieth Century Sydney*, pp 97. Kensington, NSW: New South Wales University Press.
- Shields, J. (1995). A matter of skill: The revival of apprenticeship in early twentieth-century New South Wales. *Journal of Industrial Relations* June, 236–262.
- Smith, D. (1992). The psychology of apprenticeship: A discussion paper. In Goranzon, B. and Florin, M. (eds.) *Skills and Education: Reflection and Experience*, pp 83–97. London: Springer.
- Smith, E. (1998). How apprentices learn to work. *Australian Bulletin of Labour* 24(2), 127–140.
- Smith, E. (2002). Theory and practice: The contribution of off-the-job training to the development of apprentices and trainees. *Journal of Vocational Education and Training* 54(3), 431–453.
- Smith, E. and Keating, J. (2003). *From Training Reform to Training Packages*. Tuggerah Lakes, NSW: Social Science Press.
- Smith, E., Comyn, P., Brennen Kemmis, R., and Smith, A. (forthcoming). *High Quality Traineeships: Identifying What Works*. Adelaide: NCVER.
- Smith, P. J. (2003). Learning strategies used by apprentices in flexible delivery. *Journal of Vocational Education and Training* 55(3), 369–383.
- Snell, D. and Hart, A. (2007). Vocational training in Australia: Is there a link between attrition and quality? *Special Issue: Quality in Apprenticeships. Education and Training* 49(6), 500–512.
- Sweet, R. (1987). Industry restructuring and reskilling: Apprenticeship and beyond. *Paper Presented at a H.V. Evatt Memorial Foundation Seminar on Industry Restructuring*, 7 August.
- Sweet, R. (1995). *All of Their Talents? Policies and Programs for Fragmented and Interrupted Transitions*. Sydney: Dusseldorp Skills Forum.
- Triplett, J. and Bosworth, P. (2004). *Productivity in the U.S. Services Sector: New Sources of Economic Growth*. Washington, DC: Brookings Institution Press.

- Unwin, L. (1996). Employer-led realities: Apprenticeship past and present. *Journal of Vocational Education and Training* **48**(1), 57–68.
- Unwin, L. and Fuller, A. (2003). *Expanding Learning in the Workplace*. Leicester: National Institute of Adult Continuing Education.
- Uwameiyiye, R. and Iyamu, E. (2002). Training methodology used by the Nigerian indigenous apprenticeship system. *Adult Education and Development IIZ/DVV* **59**, 169–180.
- Venables, E. (1967). *The Young Worker at College: A Study of a Local Tech*. London: Faber and Faber.
- Walters, C. (2003). (Keynote address). *Changing Face of VET: 6th Annual Conference of the Australian VET Research Association*. Sydney, 9–11 April.
- Walther, B., Scheri, J., and Wolter, S. C. (2005). Shall I train your apprentice? An empirical investigation of outsourcing of apprenticeship training in Switzerland. *Special Issue: Apprenticeships. Education and Training* **47**(4/5), 251–269.
- Winning, A. (1993). Apprenticeships and RPL: Re-cognizing the learning of a trade. *Education and Training* **35**(7), 18–21.

Skills Shortages: Concepts, Measurement and Policy Responses

C Shah and G Burke, Monash University, Clayton, VIC 3800, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

A level of anxiety currently exists about skills shortages globally, and its impact on the economy. In the developed countries, aging of the population is often cited as a cause of looming shortages.

Shortages could result in underutilization of capacity and, in turn, reduced production. If the response of employers is to bid up wages in ways that do not match productivity, then it could result in higher inflation, lower retention (due to higher job-to-job turnover), and less productivity.

Employers, employees, and policy analysts have varying perspectives on skills shortages. The lack of a common understanding of the concept often obscures analyses of problems, their causes, and possible solutions. Nearly half a century ago Arrow and Capron (1959) wrote: "some proposals for solving shortage problems stem from a misunderstanding of the causes of shortages as well as from an exaggeration of the evidence."

This article provides a discussion of the varying views on skill shortages and some of its causes. Issues of measurement and indicators of skills shortages are also discussed as are public policy responses to skills shortages.

Meanings, Perspectives on, and Causes of, Skills Shortages

What are Skills?

A skill is an ability to perform a productive task at a certain level of competence. Skills are often associated with a formal qualification, although an individual can acquire skills in other ways, including various forms of informal learning and on-the-job experience. As a job involves doing a number of tasks, a person must possess a bundle of skills to perform in a job.

Becker (1962) distinguished between general and specific skills. He defined general skills as those that are useful to many firms and specific skills as those that are generally useful to the firm providing the training.

The literature often distinguishes between basic (literacy, numeracy, and, increasingly, computer literacy), generic (problem solving, team working, and ability to improve personal learning and performance), and vocational/occupational-specific skills. Personal attributes related to motivation, attitude, leadership, and initiative are also considered desirable skills by some employers who put a high

value on them when hiring new employees (Green *et al.*, 1998), although Keep and Mahew (1999) argue that these are not really skills as such.

Differentiating Meanings of Skills Shortages

The three common meanings attached to skills shortages are as follows:

1. *Hard-to-fill vacancies.* These exist when employers have considerable difficulty filling vacancies in an occupation, within a reasonable period, and at current levels of pay and conditions of employment. They can coexist with relatively high overall unemployment in the occupation.
2. *Skills gap.* These occur where existing employees lack the required qualifications, experience, and/or specialized skills to meet the firm's skill needs to perform in a job. They may apply to new employees, where employers are unable to find suitable applicants for a job and recruit workers who need further training and/or experience to meet the firm's skill needs. Skills gaps do not necessarily relate to formal vocational or occupation-specific skills as sometimes the complaint is about generic skills or attitudes. When employers are not fully aware of the skills needed for optimal production, skills gaps can be latent.
3. *Recruitment difficulties.* These occur when employers have some difficulty filling vacancies in spite of adequate supply of skilled workers. Relatively low remuneration, poor working conditions, location hard to commute to, and ineffective recruitment or firm-specific and highly specialized skills needs are some of the causes of recruitment difficulties. Thus, recruitment problems generally do not constitute a market-wide skills shortage because with appropriate offers of wage compensation workers could be enticed to work in many of the situations described above.

Perspectives on Skills Shortages

Employer perspective

The employer perspective is usually in terms of recruitment difficulties experienced by individual employers. These do not necessarily lead to unfilled vacancies in the short run but can result in other labor-market difficulties or rising labor costs that the employer may regard as a shortage.

Green *et al.* (1998) show that although many employers equate skills shortages with hard-to-fill vacancies, a substantial number do not. The analysis showed that some employers viewed internal skills deficiencies, or gaps, in the same light as skills shortages. Others were declaring the existence of skills shortages when in fact they were unwilling to offer competitive wages.

In the short run, instead of recruiting new workers, some firms react to an increase in the product demand with a range of coping mechanisms that could involve the reallocation of resources within the internal labor market, increasing hours of work per employee, and changing the incentive system to increase worker effort and, thus, improve efficiency (Akerlof and Yellen, 1986; Solow, 1979). When a firm attempts to adjust employment to meet new product demand, then its response shifts to interaction with the external labor market.

Irrespective of whether the firm relies on the internal or external labor market to solve the problem of increased product demand, if costs increase at the margin then some firms might view this as a skills shortage. With continuing problems, some firms substitute skilled labor in short supply with differently qualified labor. Sometimes this necessitates additional training, either internally provided or purchased from an external provider. In such situations, even though the employer might feel a skills shortage exists, from a market perspective, the positions were filled and hence no shortage exists.

Union perspective

Skill is central to the bargaining between labor and management over wages (Rainbird, 1992). Unions generally want to improve the skill level of their members because higher skills increase their bargaining power in wage negotiations and help deliver better services.

In occupations with strong union representation, such as nursing and teaching, where demand is sometimes set against some desired staffing standards, it would appear to be in the interest of the union to overstate the extent of current and future skills shortages because it helps push up the wages of new entrants and increase union membership.

The monopoly powers of some unions enable them to restrict supply and assist in creating scarcity of skills. For example, some medical specialist groups have significant powers over who will be admitted for training and the numbers that will be trained (Borland, 2002; Seldon *et al.*, 1998; Friedman and Kuznets, 1945).

Causes of Skills Shortages

Slowness in the adjustment of wages

Much of the theory about defining and identifying occupational shortages developed from research that focused on wage movements (Blank and Stigler, 1957; Arrow and

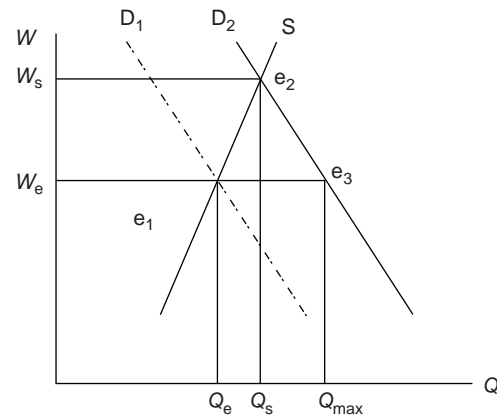


Figure 1 Market with increasing demand for a particular skill. From Roy, R., Henson, H., and Lavoie, C. (1996). *A Primer on Skill Shortages in Canada, R-96-8E*. Hull: Human Resources Development Canada.

Capron, 1959). (See Roy *et al.* (1996) and Veneri (1999) for a review.)

The price–quantity graph in **Figure 1** provides a way to explain the role of wage adjustment in the dynamics of skills imbalance. It shows the initial demand curve for a particular skill at D_1 and the supply curve at S . The market is at equilibrium at point e_1 with wages W_e and supply of skills at Q_e . If the demand for the skill increases from D_1 to D_2 , then unless wages adjust instantaneously to the new equilibrium value of W_s , the quantity of skills supplied will be less than the quantity demanded by the amount, that is, the difference between Q_{max} and Q_e . Hence, under current market conditions, employers will be unable to hire workers with these skills and a shortage will occur.

In practice, however, wage adjustments are more gradual. Impediments include delays by employers in accepting the need for wage adjustment and the reluctance to disturb existing wage structures in order to raise the rates for new employees with the required skills (Arrow and Capron, 1959). In some situations, the firm might provide in-house training either to lesser-skilled employees in the firm or to new hires without the required skills to prevent wage inflation.

Arrow and Capron (1959) introduced the concept of a dynamic shortage in which demand for a skill grows faster than supply. The causes of a dynamic shortage included a rapid and persistent rise in demand and supply responding only slowly and less than proportionately to wage changes.

In sectors, such as health and education, the government influence on the supply of, and demand for, skills is significant, and the management of the market is more direct. Market equilibrium will be difficult to achieve in such instances because wage adjustments might respond very slowly to changes in demand or supply.

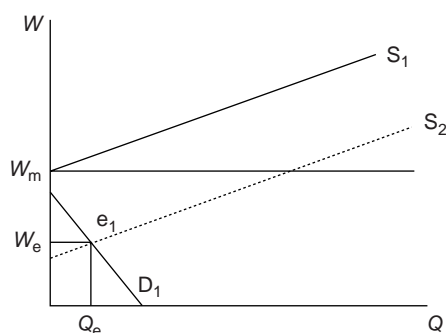


Figure 2 Market with equilibrium wage below a certain minimum. From Roy, R., Henson, H., and Lavoie, C. (1996). *A Primer on Skill Shortages in Canada, R-96-8E*. Hull: Human Resources Development Canada.

Skills shortages can occur when the equilibrium wage for a particular skill is below some minimum level. Here, the minimum wage could correspond to the wage equivalent of public welfare benefits payments (Roy *et al.*, 1996). **Figure 2** depicts one of these situations, typically encountered in low-skilled occupations. Initially, the equilibrium wage is W_e with demand D_1 , but no worker is willing to work at this wage because it is below the socially acceptable minimum wage of W_m . Under this situation, employers will report skills shortages or recruitment difficulties. If employers are allowed to hire workers from outside the country's labor market (e.g., through temporary or guest worker programs), the supply curve shifts to S_2 , and equilibrium is achieved at e_1 , but only if the socially acceptable minimum wage is inapplicable to the outside workers.

Slowness in the adjustment of supply

The main factors affecting the speed at which supply adjusts are the length of training required and the responses of the training institutions to changes in demand. In some occupations, education and training requirements can take a long time to complete, although refresher courses for those who are returning to the same occupation after an extended period of absence from it can accelerate the process.

Educational institutions adjust only slowly in offering places for particular courses, although this may vary across sectors. The time taken for training authorities to recognize a particular need, the internal budgetary processes in public institutions, and the current staffing patterns indicate that the number of places provided in areas that are in shortage adjust slowly over time. Making public funds for training contestable between the public and private providers of training has the potential to improve response times; however, against this is a risk of poor quality if providers cut corners to win contracts.

The expansion of publicly funded places is not the only factor affecting the numbers. Students must be

attracted to take up the new places, and in the case of apprentices, secure a contract with an employer which can be difficult if employers view training as a cost rather than an investment.

Student demand for places in a particular program can decrease if the program traditionally attracted particular types of students who now have other more attractive employment alternatives available to them. For example, nursing traditionally mainly attracted females but nowadays wider career opportunities have opened up for females, thus making it harder to fill training places in nursing in some developed countries.

Apparent shortages can arise because some employers may prefer employees from a certain demographic or ethnic group. They can also arise from the 2-year experience syndrome among some employers who are reluctant to hire employees without experience.

Workers leaving an occupation are a factor that reduces supply. The rate of turnover of workers varies by age and sex and across occupational groups. The net rate – the number leaving an occupation net of reentrants – tends to be lower in higher-skilled occupations than in lower-skilled ones (Shah and Burke, 2003).

Net immigration affects supply too and can vary over time with changes in migration policies. Governments generally have more control over immigration and less over emigration. The time lag from the moment a particular skills shortage is identified and when the first migrants with those skills arrive can be, however, long.

Lack of labor market information

The availability of good-quality labor-market information to firms, individuals, and education planners is the life-blood of an efficient market. Lack of reliable information can hamper the speed of the market-adjustment process and, therefore, the duration of shortages. Not only the time taken for information flow to workers regarding new opportunities but also the time taken by them to take advantage of those opportunities adds to the time lag. Since some workers might need training, inefficiencies arise when there is lack of transparency in the links between education and training courses and labor-market opportunities (Heijke and Borghans, 1998).

Identifying and Measuring Skills Shortages

To develop policy responses to skills shortages, it is necessary to have an indication of its size, its causes, how it affects production, and whether, on current trends and policy settings, it is likely to be a short-term phenomenon. As it is rarely possible to provide an unambiguous single measure of skills shortages, a range of measures should be used to indicate its existence and possible size. Approaches

to identifying shortages and estimating their extent are varied but generally fall in two broad classes (Roy *et al.*, 1996).

The first class of measures includes economic indicators, such as vacancy, hiring and separation rates, relative wage movements, and employment and unemployment changes, to infer imbalances for particular occupation groups. These measures provide a market-wide perspective and identify shortages only if there are insufficient number of appropriately skilled people in the market to fill vacancies at the going wage rates and employment conditions.

The second class utilizes employer-based surveys, interviews, and focus groups in the main, to make inference on imbalances. Typically, this involves collecting and collating employers' experiences in hiring workers with specific skills in particular occupations.

Initial Issues

Irrespective of the approach taken – economic indicators or employer-based surveys – to find evidence of skills shortages, the results critically depend on the practical definition of skills shortage adopted, the aggregation level, and the time and geographical dimensions. In practice, however, two factors will more or less determine the approach taken – availability and reliability of the data.

Level of aggregation of occupational and skills data

A high degree of aggregation is generally sufficient when producing broad information on the labor market or for vocational and educational guidance. In contrast, for specific publicly supported programs aimed at increasing the supply of particular skills in shortage, or the writing of curriculum for education and training programs, too high a level of aggregation of the data on skills imbalance may not be of much practical use (Roy *et al.*, 1996). It is worth noting that the likelihood of finding market imbalances is higher when the data are highly disaggregated. This is because imbalances tend to offset each other in aggregated data.

Time dimension

The period chosen for the analysis is likely to affect the results because in some cases no imbalance may be evident over a season or a business cycle. The characteristic of the occupational group and whether one is interested in the short-run or the long-run fluctuations, fundamental structural imbalances should be the main guide for determining the period for analysis.

Spatial dimension

The variable performance of labor markets across regions signifies that while some regions may have shortages of

particular skills, other regions may have surpluses. Inclusion of spatial dimension to the analysis is, therefore, important, as internal migration sometimes becomes part of the solution to the problem of skills imbalances. Since financial costs of moving relative to expected earnings are lower for a higher-skilled worker than a lower-skilled one, and training costs increase with skill level, relocation would be a more viable solution for higher-skilled workers (Roy *et al.*, 1996).

Spatial analysis of imbalances adds further pressure on the detail required in the data, especially if it is to be at a disaggregated level.

Indicators of Skills Shortages

Hard-to-fill vacancies and vacancy rates

Unfilled vacancy rates are sometimes used to assess the tightness in the labor market, although it should be noted that because of job-to-job turnover, vacancy statistics tend to overestimate market needs. A situation with a large number of unfilled vacancies that are hard to fill is indicative of a skills shortage.

There are, however, problems in interpreting unfilled vacancy statistics from employer-based surveys. These problems are mainly related to the inconsistent interpretation of various aspects of skills shortages by employers in self-reported surveys (Green *et al.*, 1998). For example, of the 570 000 vacancies estimated in England from the *Employers Skills Survey* in 2005, 204 000 were hard to fill; however, on closer analysis, only 143 000 of these were genuine market shortages, the rest were due to other recruitment difficulties (Learning Skills Council, 2006).

Unemployment rates

Occupational unemployment rates can provide indications of skills imbalances, but it would be naive to think that any positive unemployment rate is an indication of a surplus. There are numerous reasons why the observed and equilibrium occupational unemployment rates are never zero. The challenge with this type of analysis is in determining, at the occupational level, what is the normal (equilibrium) unemployment rate, above which would be considered a surplus situation. Furthermore, among the unemployed may be persons who are not qualified to work in the occupation and, therefore, the number of unemployed overestimates supply. Conversely, if there are persons qualified to work in the occupation but who are employed elsewhere, then unemployment numbers may underestimate supply.

Net vacancies

If both vacancy and unemployment statistics are available at the occupational level, which they are often not, then the unemployment–vacancy (UV) technique can be used to assess skills shortages (Roy *et al.*, 1996). A simple

interpretation would be that if the number of vacancies is larger (smaller) than the number of unemployed persons, then a shortage (surplus) could be inferred. At the aggregate level, the UV analysis shows us that at any point in time, unemployment and presence of vacancies coexist largely because of labor turnover.

A number of assumptions underlie this technique. First, it has to be assumed that vacancies and the unemployment level represent unmet demand and supply under current market conditions. This assumption does not always hold because, as discussed above, the unemployment rate can somewhat exaggerate supply. Therefore, if vacancies reflected true unmet demand, then the UV analysis is likely to bias results toward an absence of a shortage.

Wages

In a market economy, where wages and prices moved freely, occupational labor imbalance would translate to changes in relative wages over time. Changes in relative wages have, therefore, been used to study occupational labor-market imbalances.

There are, however, practical problems in interpreting the wage differential data because a proper analysis requires knowing the supply and demand elasticities, that is, how much the supply and demand will change in response to a change in wages.

Another practical problem with interpreting wages data arises when institutional factors, rather than increased demand, drive wages up. For example, anti-discrimination or minimum wage legislation can drive up wages in some occupational groups. Changes in wages could also reflect the power of the union representing the workers rather than a skills shortage in the occupation. Data on wages need to be of the recently hired because the normally available average or median occupational wage data can vary due to quality or quantity shifts within the occupational group. Wage data at this level of detail are rarely available.

Cross-country evidence shows wages are relatively inflexible and do not strongly respond to current market conditions (OECD, 1994). This is suggested to be the main reason for wide dispersion in unemployment rates across occupational groups. The result is persistent imbalances that will tend to clear through interoccupational mobility rather than wage adjustments (Roy *et al.*, 1996).

Other labor-market indicators

Other indicators, such as hours and intensity of work (i.e., overtime); production levels; changes in employment; flows of new entrants and leavers; training expenditure by firms; levels of subcontracting; hiring standards; and levels of immigration and emigration, have proved useful in analyzing skills imbalances. Their potential usefulness once again depends on data availability, reliability, and its careful interpretation.

Employer-based surveys

Employer-based surveys, to a great extent, rely on employers' perceptions of skills shortages. Laslett (1992) concluded that skill shortages identified from *ad hoc* employer surveys often disappeared when the problem was followed up. Furthermore, he concluded that it is always possible to elicit positive responses to questions about skills shortages.

As mentioned earlier, research conducted in England showed that employers' understanding of a skills shortage was broader than a hard-to-fill vacancy or a skills gap (Green *et al.*, 1998). This study underscores the importance of careful design, implementation, and interpretation of any employer-based survey of shortages.

Structural models

Structural models of firms adjusting to internal labor-market imbalance and of the economy-wide effects of supply and demand shocks have been used to assess current and future labor-market shortages (Wilson, 1992; Hughes, 1994; Heijke, 1994; Boothby *et al.*, 1995; Adams *et al.*, 1994).

An advantage of structural models is that they provide some initial baseline data for more rigorous study of imbalances. There are, however, drawbacks limiting their use in practice. First, in general, the models have very large requirements of quality data to calibrate and maintain them. Consequently, many of these models are highly aggregated and simplified. Second, most models project occupational demand requirements rather than current or future market imbalances, which somewhat limits their use unless ancillary information from other sources is available on turnover and supply.

In-depth studies

The above discussion suggests that finding evidence of apparent skills shortages is relatively easy, whereas assessing genuine market shortages is more difficult. A holistic, or in-depth, approach that considers a raft of measures for a group of interrelated occupations is more likely to provide robust information on imbalances (Roy *et al.*, 1996).

An in-depth study would ideally provide a framework within which to situate the labor market under investigation together with its structure in terms of its demographic, educational, and jobs characteristics as well as include analysis of all other relevant data such as on jobs growth, turnover, and job openings. The study would ascertain whether the imbalances had a geographic dimension or not and would provide indications of how, and how fast, the occupational market resolves imbalances.

In-depth studies are costly if they were to be done at an economy-wide level. Since these types of studies are time consuming, it may only be feasible to use the technique to investigate markets in which imbalances are persistent and adjustment periods are long.

Public Policy Responses to Shortages

In a competitive market, there always will be unfilled vacancies, and a proportion of these could be considered hard-to-fill vacancies. When the number of unfilled vacancies reaches some intolerable level, there are calls for intervention in the market. In some cases, however, the observed imbalance may be socially optimal for the particular labor market, and public intervention would simply be adding to costs of adjustment that are being borne by the market (Roy *et al.*, 1996). Therefore, when is it appropriate for government intervention? Equally important are the questions of why and what of government interventions.

Labor markets may be inefficient due to a range of factors – externalities, legislative or institutional. Externalities occur when an activity (e.g., training) undertaken by individuals or firms leads to benefits (or costs) to other individuals or firms. If the externalities are significant, then the result could be low level of training activity and a likelihood of skills shortages or gaps. Such factors as wages policy, subsidies, and lack of good market information can reduce the market-clearing role of prices and wages.

Once market imbalances, including their severity and causes, have been identified, the next question as to what government action, in terms of its effectiveness and cost, is to be taken arises. The course of action will depend on the specific labor market that is in imbalance and its geographical distribution, as each market may need a tailored solution.

Publicly Provided Training

Provision of publicly funded training places is one of the ways of overcoming skills shortages and gaps. Public provision is usually rationalized in terms of efficiency and equity arguments. Since economic performance is negatively affected by skills gaps, it is argued that more training will result in an adaptable workforce. The training has, however, to be appropriate to be effective.

Publicly funded training to satisfy employer needs requires a thorough and extensive knowledge of the workings of the particular labor market because not all such needs are most effectively satisfied through such means. While initially it may appear that government action is called for to do something about the shortage, it may not always be the best first option. Often, the market will take sufficient corrective actions to alleviate shortages. Government intervention may be counterproductive if its effect is too late or it overcorrects the problem.

The nature of the intervention will depend on whether the problem is a skills shortage, recruitment difficulties, or a skills gap. Recruitment difficulties may not need publicly provided training as a remedy, although raising employers' awareness to the causes of recruitment difficulties and possible solutions for them are important issues.

If it is concluded that there is a need for a training response, there are still important questions about its cost, how it should be financed, how will it be provided, and how will students or workers be attracted to it. The vocational education and training sector has an important role in any such solution. In some countries, such as Australia, the sector is generally considered to be responsive as it is able to tailor programs to meet particular needs fairly quickly. The policy choices need to be seen in the context of a commitment to lift the overall skill level of the workforce. It should also be remembered that the public system has a major role in promotion of equity and education for citizenship or personal development.

Encouraging Training in Industry

The greater the employers' awareness of the importance of training and the greater their involvement in planning and provision, the greater will be the employers' capacity to respond to shortages. Engendering a training culture among employers takes time, as it requires attitudinal changes and long-term perspective on training and its benefits for innovation and profits.

A way of increasing employer involvement and responsibility for training tried by various governments has been various types of training levies. Levies have a longer history in European countries than in the United States or Australia. It has been suggested that a training-levy system that is owned and controlled by industry groups, including unions, may have a better chance of success than a system that is nationally imposed, particularly for trade and low-skill occupations (Government of South Australia, 2003; Gasskov, 2002). Any training-levy system has to ensure equity of access to training for all workers, particularly those who are least qualified and who have the least bargaining power.

Employer training can also be encouraged by tax relief or subsidies to firms commensurate with the amount of training they provide their employees.

Labor Market Information

Transparency is essential for the efficient functioning of a market. Pertinent, reliable information about various aspects of the market that is up to date and available to all provides this transparency, and helps clear imbalances in the market. The information must relate to not just a particular market but also across all markets and must include current wage rates, rates of return, unemployment rate, job openings, job turnover, and labor supply.

Producing quality information is expensive. If this function is left to the free market, the provision is likely to be less than what would be regarded as socially adequate because of the public good nature of the information.

Hence, governments have an important role to play in providing such information.

Immigration

Immigration can be used as a policy instrument to solve skilled and unskilled labor shortage problems. Many countries, such as Australia and Canada, have permanent skilled migration programs to meet expected shortfall in particular types of skills. Generally, the programs are not open ended; annual quotas impose a ceiling to the numbers who are allowed in and occupations in the priority list change as well. One of the problems with the system is that arrival of migrants, with appropriate skills, often lags by a considerable time the actual occurrence of the shortage. This means that sometimes the market has already corrected itself by the time the migrants arrive. The result is that migrants often find themselves having to work in alternative, and sometimes lower-skill, occupations. Migrants' general preference for settling in large metropolitan areas means shortages in regional or remote areas often remain unaddressed.

More recent trends suggest temporary migration programs becoming increasingly popular, including in traditional immigration countries, such as Australia, that previously eschewed such programs. Temporary migration schemes usually begin as a short-term solution to labor shortages in particular sectors or geographical areas; however, they often become a permanent feature of the labor supply as the sector becomes dependent on them and native labor flees the sector because of lower wages and inferior working conditions. This is unlikely to be conducive to sustainable education and training programs or competitive advantage.

Immigration may solve the skills shortage problems of receiving, usually developed, countries but it has the potential to severely disrupt the labor market of the donor, usually developing, countries. When this happens in sectors such as health and education, which are crucial for the development of the human capital of the country, the impact can be quite detrimental. Generally, unless there is a bilateral government-to-government agreement, the donor country has no recourse to recoup the investment it had made in training the individuals who leave.

Encouraging Mobility

Assistance to encourage mobility may help alleviate imbalances in geographical skills, but an important issue is that qualifications are recognized across the different regions.

Internal migration programs are more likely to succeed if they are backed with comprehensive information systems, not only about the labor market but also about other services, such as schooling, health, and housing. Large

disparities in housing costs between different regions could inhibit worker flows. Whether governments should provide financial assistance for internal migration or not is a difficult policy issue and would depend on how strategic the problem is.

Concluding Remarks

This article has provided a review of the concepts, measurement issues about, and responses to, skills shortages. Economic theory suggests that skills imbalances are a permanent feature of a competitive market. Without major hindrances to the market-adjustment process and without externalities, these imbalances resolve over time and do not require intervention. There will always be unfilled vacancies coexisting with unemployment because of continual adjustment of the market toward equilibrium. Adjustment usually takes time because of the speed at which the information about the changes in the market is dissipated and the time it takes an individual to acquire new skills.

Real-world labor markets are, however, seldom perfect. Government action to correct the market may be justified if there is strong evidence showing sub-optimality. Before any effective policy is designed and implemented, it is important to understand the operation of the current market and to investigate reasons for its failure. Lack of good empirical knowledge of how markets adjust, and the time lag from when an imbalance is observed to the impact of a policy makes effective intervention difficult in practice.

With rapid technological changes, there will always be a skills gap in the workforce. The rate at which the gaps are bridged depends on the training system in place. If employers rely on the public training system to finance and bridge the gap, then the adjustment process will be longer because of the bureaucratic necessities and the time it takes a third party to gear up with new curriculum and delivery strategies. If the firms were to take greater responsibility to bridge the employee skills gap, then the adjustment will be much faster and the training could be tailored specific to the firm, thus reducing problems of externalities.

Bibliography

- Adams, P., Dixon, P., McDonald, D., Meagher, G., and Parmenter, B. (1994). Forecasts for the Australian economy using the MONASH model. *International Journal of Forecasting* **10**, 557–571.
- Akerlof, G. A. and Yellen, J. L. (1986). *Efficiency Wage Models of the Labour Market*. Cambridge: Cambridge University Press.
- Arrow, K. J. and Capron, W. M. (1959). Dynamic shortages and price rises: The engineer–scientist case. *Quarterly Journal of Economics* **73**, 292–308.

- Becker, G. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy* **70**(supplement), 9–49.
- Blank, D. and Stigler, G. (1957). *The Demand and Supply of Scientific Personnel*. New York: National Bureau of Economic Research.
- Boothby, D., Roth, W., and Roy, R. (1995). The Canadian occupational projection system: A methodological enhancement. *Research Paper No. R-95-2*. Human Resources Development, Canada, Hull, Quebec.
- Borland, J. (2002). The market for medical specialists in Australia. *Proceedings of Health Policy Roundtable*, 245–292 Melbourne: Productivity Commission.
- Friedman, M. and Kuznets, S. (1945). *Income from Independent Professional Practice*. New York: National Bureau of Economic Research.
- Gasskov, V. (2002). Industry support for training, an international perspective. *Paper Presented to CEET National Conference 2002, Melbourne*, 26 August.
- Government of South Australia (2003). Skills for the future. Schofield, K. (reviewer) *Final Report of the Ministerial Inquiry*. Adelaide.
- Green, F., Machin, S., and Wilkinson, D. (1998). The meaning and determinants of skills shortages. *Oxford Bulletin of Economics and Statistics* **60**(2), 165–187.
- Heijke, H. (ed.) (1994). *Forecasting the Labour Market by Occupation and Education*. Boston, MA: Kluwer.
- Heijke, H. and Borghans, L. (eds.) (1998). *Towards a Transparent Labour Market for Educational Decisions*. Aldershot: Ashgate.
- Hughes, G. (1994). An overview of occupational forecasting in OECD countries. *International Contributions to Labour Studies* **4**, 129–144.
- Keep, E. and Mahew, K. (1999). The assessment: Knowledge, skills and competitiveness. *Oxford Review of Economic Policy* **15**, 1–15.
- Laslett, R. (1992). Surveys of skill shortages in engineering labour markets. In Bosworth, D., Dutton, P., and Lewis, J. (eds.) *Skill Shortages: Causes and Consequences*, pp 111–117. Aldershot: Avebury.
- Learning Skills Council (LSC) (2006). *National Employers Skills Survey 2005: Main Report*. LSC. <http://readingroom.lsc.gov.uk/Lsc/2006/research/commissioned/nat-nationalemployersskillsurvey2005keyfindings-re-june2006.pdf> (accessed June 2009).
- OECD (1994). *OECD Jobs Study: Evidence and Explanations*. Paris: OECD.
- Rainbird, H. (1992). Union perspectives on skill shortages. In Bosworth, D., Dutton, P., and Lewis, J. (eds.) *Skill Shortages: Causes and Consequences*, pp 196–206. Aldershot: Avebury.
- Roy, R., Henson, H., and Lavoie, C. (1996). *A Primer on Skill Shortages in Canada, R-96-8E*. Hull: Human Resources Development Canada.
- Seldon, B., Jung, C., and Cavazos, R. (1998). Market power among physicians in the U.S., 1983–1991. *Quarterly Review of Economics and Finance* **38**(4), 799–824.
- Shah, C. and Burke, G. (2003). Job turnover: Replacement needs and vacancies by occupation. *Report for the Department of Employment and Workplace Relations*. Canberra.
- Solow, R. M. (1979). Another possible source of wage stickiness. *Journal of Macro-Economics* **1**, 79–82.
- Veneri, C. (1999). Can occupational labor shortages be identified using available data? *Monthly Labor Review* **1999**, 15–21.
- Wilson, R. (1992). Forecasting occupational change and skill imbalances. In Bosworth, D., Dutton, P., and Lewis, J. (eds.) *Skill Shortages: Causes and Consequences*, pp 49–68. Aldershot: Avebury.
- Borghans, L., de Grip, A., and van Smoorenburg, M. (1998). *Measuring Skill Shortages, ROA-R-1998/4E*. Maastricht: Research Centre for Education and the Labour Market.
- Long, M., Ryan, R., Burke, G., and Hopkins, S. (2000). *Enterprise-Based Education and Training*. Wellington: Ministry of Education.
- Stern, D. (1995). Market failure in firm-based education and training. In Carnoy, M. (ed.) *International Encyclopaedia of Economics of Education*, 2nd edn., pp 172–175. Oxford: Elsevier Science.

Further Reading

Training and Learning in the Workplace

R L Jacobs, The Ohio State University, Columbus, OH, USA

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Deductive learning – An approach to learning by which the training content is presented in a systematic manner involving definitions, steps, attributes, example, and non-examples. This approach is often thought of as expository learning.

Far transfer of training – The process by which information to be transferred to another setting is related to the information presented during the training.

Financial benefits of training – Training programs should be linked to performance issues, which can be equated in financial terms. When the performance issue is addressed through training, then the financial benefits can be calculated by determining the cost of the training, the value of the performance that results from the training investment, and the financial benefits that result from subtracting the training costs from the performance value.

Formal learning – The array of learning experiences that occur in settings that are explicitly intended for learning, such as a training classroom.

Human competence – The relative level of a person's ability – from novice to master – in respect to a particular unit of work.

Inductive learning – An approach to learning by which trainees acquire the content through the learning of the critical attributes first, and then synthesizing this information into a formal concept or principle. This approach is often thought of as discovery learning.

Informal learning – The array of learning experiences that occur in the context of other activities, such as working.

Knowledge work – The type of work that features the manipulation and dissemination of information as part of their work.

Near transfer of training – The process by which the information to be transferred to another setting is identical to the information presented during the training.

Planned training – A training program that has been developed using a systems approach, which requires that the training content be based on a set of training objectives and that the effectiveness of the training

be determined by comparing the training objectives with the training outcomes.

Structured on-the-job training – The planned process of using experienced employees to train novice employees in the work setting.

Training delivery – The sequence of events taken by which the training content is presented to learners. In terms of structured on-the-job training, there are five train events: (1) prepare the trainee, (2) present the training, (3) require a response, (4) provide feedback, and (5) evaluate performance.

Unplanned training – A learning experience that has not been developed using a systems approach.

The purpose of this article is to discuss the current practice, theory, and research related to training and learning in the workplace. There is a growing awareness of the strategic role of training and learning for organizational effectiveness. This article gives particular emphasis to the planned training and learning programs that occur in the context of the work setting or on the job. On-the-job training (OJT) and learning has garnered increased interest among researchers and practitioners because it addresses a number of practical and theoretical issues.

The article has the following sections: the first section discusses the context of workplace training and learning as it relates to the knowledge economy and implications for employee competence; the next section seeks to define workplace training and learning in terms of the nature of the setting – formal or informal – and the degree of planning that the learning experiences receive; the third describes structured OJT (S-OJT) as one example of workplace training and learning that has received considerable interest in organizations and the implementation of S-OJT; and the fourth section reviews the research on S-OJT and raise issues for future research.

Context of Training and Learning in the Workplace

Training and learning in the workplace can be best understood from the context of the competitive needs of organizations. That organizations have been subject to unique challenges as a result of the global economy should now be accepted without much surprise or dispute. Perhaps the

most well-known spokesperson for this perspective is Friedman (2005), who introduced the notion that instead of traditional markets comprising barriers across national boundaries, markets have now become more flat in a metaphorical sense. That is, because of advanced communications technology, an organization, say, in Taiwan can now compete on an equal basis with an organization in some other country to reach customers in a third country. By using its unique resources to best advantage, each company can ensure its distinctiveness in the marketplace to meet the demands of customers. A constant factor in this scenario is the complex relationship between suppliers and customers, in which each is driven to lower costs, provide higher service and product quality, and demonstrate flexibility in responding to changing market demands.

The emergence of the global economy has brought about two fundamental shifts in the work that people do. First, the nature of work has increasingly moved toward the inclusion of knowledge (Ackerman, 1998). Knowledge work is defined as when employees take in information from a variety of sources, use that information to derive a set of solutions, and, as a result, generate new sets of information (Mohrman, 2003). Thus, the inputs, processes, and outputs constitute the manipulation of information rather than objects or things. The emergence of knowledge work has shown to be relatively pervasive across jobs. For instance, Lee (2004) showed that among bank employees, knowledge work is no longer the monopoly of professionals or senior managers. Rather, movement toward knowledge work occurred across all levels of employees, though these changes seem to have affected frontline employees and middle managers more than senior managers.

The second shift regarding work is that the life cycle of job content has become increasingly shorter, with changes occurring on a more frequent basis. Job change is often a second-order result when organizations change in the way they operate. Osman-Gani and Jacobs (2001) reported that technology was the major driver for prompting job change among companies in Singapore across several business sectors. The introduction of technology, whether to enhance management or production capabilities, invariably causes a redefinition of the organization and roles. Lean manufacturing, for one major organizational change, often serves as one primary example of how job content dramatically changes as a result of changes in organizational structure.

These two fundamental shifts – the movement of jobs toward knowledge work and the shortened life span of job content itself – have heightened the need to better understand employee competence and how to find ways of responding to changing competence needs both effectively and efficiently. As shown in **Figure 1**, Jacobs (2002a) has proposed five levels of a taxonomy of employee competence, starting at the lowest level of knowledge and skill,

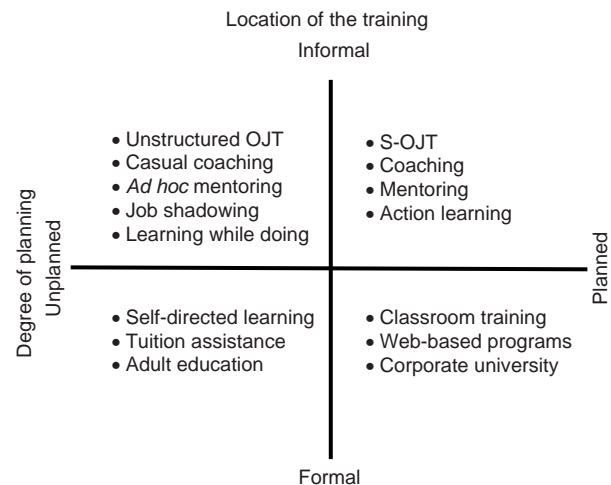


Figure 1 Training and learning in the workplace.

the novice level, and then moving upward to the highest level, the master level.

The five categories follow how individuals might progress in their learning of a specific unit of work. In this framework, a person's ability on a task should be conceived as a relatively narrow construct, not as an overarching ability. Thus, the taxonomy illustrates how individuals could experience a change in their competence when the organization undergoes change. For instance, an experienced engineer might be viewed as an expert in troubleshooting equipment problems in field settings. However, when a remote sensor technology was installed on the equipment, the troubleshooting process could be conducted in an office location instead, requiring an understanding of how to use the software and how to interpret the results to others. In effect, the experienced engineer's level of competence on this troubleshooting task dropped from being expert to perhaps a novice.

How organizations respond to such fluctuations in employee competence reflects how well the organization can remain competitive in the long term. While managers understand the importance of recruiting a qualified and skilled workforce at the outset, equally important is having the systems to respond to changes in job requirements over time. Beyond the absolute need to capably manage their physical and financial resources, organizations must also respond to their current and anticipated employee competence needs, mostly through wise investments in workplace training and learning programs, which have become the primary means to transmit this information.

Training and Learning in the Workplace

Training and learning in the workplace encompasses the various activities and processes used by organizations to impart job knowledge and skills. Previous descriptions of this information have focused on the various training

methods and media that have been used and their potential benefits, such as classroom presentation, computer-based training, or distance training. The most recent annual survey of the state of the workplace learning and performance industry from American Society for Training and Development (ASTD; Rivera and Paradise, 2006) reports the estimated average percentage of training hours per category of delivery method: live instructor led, self-paced, and noncomputer based. The report also compares the percentage of learning hours of instructor-led and technology-based training methods. While this information has importance for organizations seeking to benchmark their human resource development practices, it remains discrete without any conceptual framework to assist understanding.

Table 1 presents two dimensions for understanding training and learning in the workplace: (1) unplanned and planned and (2) informal and formal. The resulting four cells represent the training and learning approaches from the interaction of the dimensions. Unplanned training is often characterized by an *ad hoc* approach to the training. That is, there is uncertainty or lack of concern about the outcomes of the learning and the methods used for the learning. In a sense, unplanned training might be viewed as being phenomenological in that whatever occurs during and after the training might be viewed as being acceptable. The experience may itself be sufficient to warrant the individuals perceiving value from the training. Planned training suggests the use of a systems approach to develop the training. That is, a systems approach requires that the outcomes be specified first; then, the methods to achieve the outcomes are specified afterward. The relationship between means and ends is a prominent aspect of planned training.

Current perspectives about using a planned approach for designing training programs comes from a quality-management perspective. Increasingly, training is being viewed as a critical process that affects the organization's

ability to deliver products and services that meet customer requirements. In this sense, the International Standards Organization (ISO) 10015 Quality of Training standard has come to signify the importance of controlling training as a process. The standard has the following phases:

1. analyze performance problems;
2. identify training needs;
3. develop and plan the training;
4. provide for the training;
5. evaluate the training; and
6. monitor and improve the training (Liu *et al.*, 2005).

To be most effective, the phases of the standard should be connected to the quality-management system of the organization.

The dimension of informal–formal describes the location of the training. The distinction between informal and formal learning environments has been proposed by several authors, including MacLean and van der Pol (2006). Informal locations are environments that have not necessarily been designated for learning, but in which individuals learn naturally nevertheless (Beckett and Hager, 2002). Barnett (1999) proposes that learning and work need to be worked out at different levels, both organizational and personal, and in different modes, formal and informal. The key aspect of an informal learning environment is that it represents what occurs in an activity- or experienced-based situation. The workplace is the most commonly used example of an informal learning setting. Lohman (2005) described informal learning as involving those learning activities that employees initiate in the workplace, involving the use of physical or cognitive abilities, and result in the development of professional knowledge and skills. Informal learning settings typically engage individuals in a more holistic manner, in which entities such as peer learning and communities of practice might likely occur as a result (Hager, 2004).

In contrast, formal learning refers to the locations and processes that are designated as places for learning to occur. Sambrook (2005) differentiated work-related learning as learning at work and learning in work. Learning at work is associated with the more formal provision of education and training courses such as induction, mandatory health and safety, and a range of accredited and nonaccredited in-house courses. Learning in work is associated more with informal processes embedded in work activities, such as observing, asking questions, problem solving, project work, attachments, coaching, and being part of multidisciplinary teams. Instead of being a holistic experience, formal learning is often made up of discrete planned learning experiences that might have a more narrow and circumscribed boundary. Formal learning is usually institutionally sponsored and may occur in a classroom or similar setting in which the control of the learning is in the hands of the organization, and not the individual.

Table 1 Taxonomy of employee competence

Category	Description
Master	The real expert among experts. This person sets the standards for others
Expert	One who can do both the routine and nonroutine cases of the work
Experienced specialist	One who has performed the work repeatedly and can do it with ease
Specialist	One who can reliably perform most work; however, the range is limited
Novice	One who is new to the work and lacks the ability to meet requirements

Source: Jacobs, R. L. (2003). *Structured On-the-Job Training: Unleashing Employee Expertise in the Workplace*. San Francisco, CA Berrett-Koehler.

As stated, the interaction between the two dimensions – degree of planning and location of the training and learning – provide a framework for understanding different training approaches. The unplanned/informal cell represents what constitutes much of the employee learning that occurs in organizations, the most prominent of which might be called unstructured OJT, or follow-Joe training. Clearly, there is a role for these training approaches in terms of allowing some degree of exploration and discovery. However, the consequences of such training can be harmful to organizations and the individuals involved. The planned/informal cell represents an opportunity to have a planned program with predictable outcomes, but to have the training in the work setting. The following section provides more in-depth information about S-OJT, perhaps the most prominent training approach in this cell.

The unplanned/formal cell represents those training experiences that may have a clear intent, but it is uncertain how the information relates to the employee's expectations. The approaches in this cell might be considered more educational in nature, not training *per se*. Finally, the planned/formal cell represents the types of off-the-job training approaches that have characterized organizational training programs, such as classroom training and web-based training programs. Taken together, the four cells provide a relatively comprehensive view on how to understand the various training and learning programs used in organizations.

Structured OJT

Of all the training and learning approaches conducted in the workplace, the most attention has been paid to S-OJT. S-OJT has been defined as the planned process of having experienced employees train novice employees in the workplace or a setting similar to the work setting (Jacobs, 2003). From this definition, S-OJT benefits from providing a greater degree of predictability in the training outcomes, and from being conducted mostly by experienced employees.

Rationale for Using S-OJT

From a practical perspective, S-OJT offers some apparent advantages for managers because the training might be conducted when the employee need arises and without demanding any special resources from the organization, such as a classroom Internet connection. Beyond these reasons, the rationale for using S-OJT comes from an understanding of its potential to address three basic issues related to workplace training: (1) the extent of instructional contiguity, (2) the potential for the transfer of

training, and (3) the relevance of the content during the learning.

Instructional contiguity was first introduced by Gagne (19) and describes the temporal relationship among the three fundamental components of any instructional sequence: the presentation of the content, the opportunity for the learner to respond or make use of the content, and the feedback provided to the learner about the adequacy of the response. Contiguity relates to the association of phenomena because of the degree of proximity. Regardless of the theoretical perspective taken, however, contiguity remains an issue for all training approaches in how to reduce the spacing across the events as much as possible. The principle states that the closer the spacing among the events of the instructional sequence is, the more effective the learning. S-OJT has the potential of addressing this issue as much as any other training approach, if not more so, since the task is presented in the same setting in which it is practiced, with little delay between the two.

Transfer of training concerns the extent to which a person is able to take information learned in one setting and use that information in another setting (Broad, 2005). Perhaps, no other training issue is as important in organizations, that is, how to ensure that trainees actually use what they are asked to learn in the training. The literature suggests a combination of perspectives for viewing this issue. For one thing, the planned process used to design the training ensures that the training objectives are specified in a way to meet the needs of the organization and the interests of the individuals.

Another perspective suggests that transfer of training needs to be understood by the nature of the work environment or climate, in which individuals will be asked to use the learned information. This may be done through the extent of management support and knowledge of the training content. Interacting with these various perspectives is the underlying motivation of the individuals to learn and to perform in the first place.

Of particular relevance to S-OJT is the understanding that the match between the learning setting and the work setting facilitates the transfer of training (Jacobs, 2003). Since S-OJT is conducted in the work setting, or a setting that approximates the features of the work setting, and managers or supervisors often serve as trainers, the need to address transfer of training as a specific issue appears to be less critical. The characteristics of S-OJT inherently address transfer of training issues that would otherwise require attention if some other training approach might be used. For instance, how to provide concrete learning experiences and opportunities for meaningful responses continues to be an issue of concern for researchers studying online learning environments (e.g., Srinivasan *et al.*, 2006).

Finally, the rationale for S-OJT is based on the increased relevance of the training because of the location

in which it is conducted. There is an emerging understanding that cognition and its context are interdependent on each other in all learning situations (Stein, 2001). Situated learning posits that interactions between the physical and social contexts are integral aspects that determine what is learned. What is taken from the learning depends on how individuals construct meaning from the situation. The use of authentic learning experiences is preferred as these provide individuals the opportunity to manipulate things and symbols that have meaning beyond the immediate training session.

S-OJT as a System

As shown in **Figure 2**, facilitating the understanding of S-OJT is to view it as a type of training system (Jacobs, 2003). Similar to all training systems, a system view includes the inputs, processes, outputs, and the context of the system in relation to other systems. The S-OJT system shows that the inputs comprise the novice employee, or trainee, who is expected to have the motivation and readiness for the training. The experienced employee serves as the trainer and one who is expected to have sufficient knowledge of the task and is qualified by the organization to deliver the training. The training location identifies the criteria for selecting the exact location of the training. Oftentimes, S-OJT is intended to be conducted in the work setting, but there are constraints that do not allow this to occur as planned; therefore, an alternative setting needs to be identified without sacrificing the critical attributes of the work setting. The work to be learned is the unit of work that is being presented in the training, as represented by a training module format. S-OJT modules do not present an entire job *per se*, but only a defined task, assignment, or project that the trainee is expected to perform. Finally, the inputs include whatever

communications technology that may be necessary to manage or even help deliver the training. More and more S-OJT programs are relying upon some use of technology because the trainee and trainer may not be in the same location.

The processes include the activities that are done for the trainer to get ready to train, the training events used by the trainer when actually delivering the training, and the means used, such as performance ratings, follow-up observations, and periodic feedback, to ensure that the trainee has learned the content. The literature has differed somewhat on the training steps to be used by trainers. The original Training Within Industry (TWI) materials from the war production efforts of World War II prescribes four steps (Dooley, 1945). Jacobs (2003) suggests five steps to be followed by trainers:

1. prepare the trainee;
2. present the training;
3. require a response;
4. provide feedback about the response; and
5. evaluate the trainee.

The outputs of the S-OJT system are the results that occur in terms of accomplishing the training objectives, the impacts on the work, and the contribution that the training makes to the individual's own development progress. Finally, the organizational context includes the factors in the work setting that facilitate or constrain the system's components, such as the resources – time, people, and equipment – allotted to the training.

Viewing S-OJT as a system has benefits for both practice and research. For practice, the system view shows what issues need to be addressed when developing such a system. In addition, the components guide troubleshooting when seeking to improve the system once it is implemented. For research, the system view identifies the set of

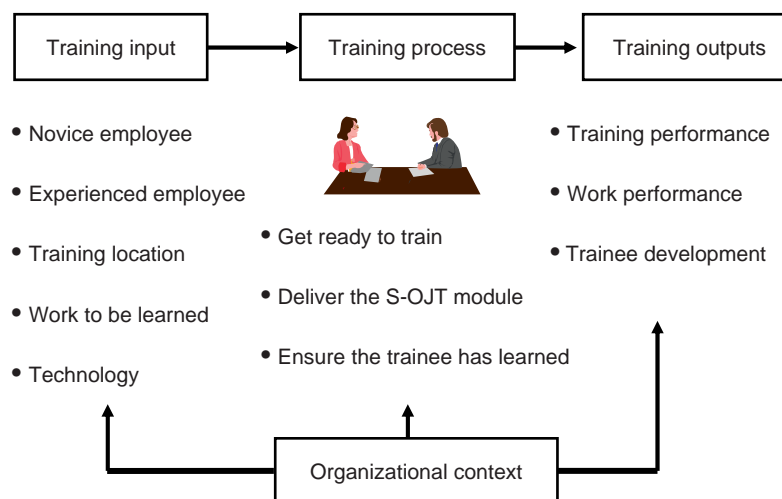


Figure 2 Structured on-the-job training represented as a system.

variables that together form a theoretical framework to guide what questions might be asked. For instance, one topic of continuing interest is research questions related to the background and behavior of the experienced employees who serve as S-OJT trainers (Johnson and Leach, 2001). The research suggests the difficulty of having individuals who are relative experts on the topic, but who are relatively ineffective in their ability to articulate the information to others.

Examples of S-OJT

Numerous examples of S-OJT programs have been reported in the literature, both from scholarly and practitioner sources. With some exceptions, most of the case studies have described the use of S-OJT for technical training, rather than it being for managerial or awareness training. Mafi (2001) proposed a model on how S-OJT might be used for managerial training for supervisors and managers. The model suggests that S-OJT used for managerial training must be connected to specific work expectations and the training should have the following sequence of events:

1. Present why the current approach is not working.
2. Present a model or process to follow.
3. Present techniques on how to use the model or process.
4. Ask the trainee to respond in a role play situation.
5. Evaluate the trainee.

Additional information about the use of managerial training, which used S-OJT to help newly promoted supervisors understand the organization's mission and core values, was reported by Jacobs (2003). Experienced supervisors served as the trainers, though their role was more of a coach than a trainer *per se*. The sessions between the new and experienced supervisors were conducted in office or conference room settings and used an inductive questioning approach.

Table 2 presents a listing of published S-OJT case studies. The table identifies the organization, the positions involved, and the reported results. The cases are taken from a variety of refereed and nonrefereed sources.

Jacobs and Osman-Gani (2005) collected case studies with an emphasis on how organizations in international situations have used forms of S-OJT. The cases illustrate both the similarities and differences based on cross-cultural considerations. Cases include organizations from Japan, Taiwan, the United Kingdom, the Netherlands, and Singapore. Perhaps the largest use of S-OJT comes from a reported project to develop new-hire engineers in a refinery setting in Kuwait (Al-Muzaini *et al.*, 2002). The project is also unique because its focus is on employees who are at the professional–technical level, rather than a frontline employee, such as a call center representative or production worker.

Regardless of the context, the various S-OJT case studies have some aspects in common. The training is conducted in the work setting, not a classroom or some other formal setting. It is delivered either by an experienced peer employee or manager. The training is accompanied by a set of materials that document the unit of work being presented during the training. Finally, it connects, to a large extent, with specific work expectations.

Research on S-OJT

Research on S-OJT is just in an emerging status. Arguably, the first research study related to the effectiveness of OJT was the Lens Grinder Study (Dooley, 1945; reprinted in Jacobs, 2003) that was conducted in 1941. The study sought to establish the effectiveness of a planned approach to training on the job. The results showed that the length of time to train apprentice lens grinders could be reduced from 5 years to 6 months. The Lens Grinder Study was reported as part of the final report authored by Channing Rice Dooley, as he was closing the TWI project during the last months of World War II. Most observers suggest that the TWI project influenced much of current understandings of S-OJT.

Much of the early research on S-OJT focused on determining its financial benefits (Jacobs and McGiffin, 1987; Jacobs *et al.*, 1992; Jacobs and Hruby-Moore, 1998; Jacobs 2002b, 2005). Two areas of focus came from these various studies: training efficiency and training effectiveness. Training efficiency addresses the question whether

Table 2 S-OJT case examples

Organization	Job	Results
GM – US	Operators Supervisors	Reduced the number of wiring errors and implement JIT system
Truck manufacturer – US	Production Technicians	Reduced the number of leaky windshields
KLM	Cabin attendants	Provided more relevant customer service training experience
Regional Hospital – US	New supervisors	Provided information about the mission and vision faster
Large electric utility	Supervisors	Reduced the number of back injuries among production employees
Apple Computer-iBook	Skilled assemblers	Reduced training time and fewer assembly errors
Seagate – Singapore	Production	Reduced number of inspection errors
Pharmaceutical	New-hire lab techs	Reduced time required to complete job rotations
Abbott – Diagnostics Div	Lab technicians	Met ISO and FDA requirements

S-OJT took less time to deliver and whether the investment made to reduce the training time was more or less than the value of the outcome of the training. In general, the results related to training efficiency suggest that S-OJT takes less time to conduct and achieve the objectives when compared to unstructured OJT, classroom training, and blended versions of the training. In addition, the results suggest that the reduction in training time is accompanied by greater financial benefits. The proportion of the time savings and the financial benefits depend on the individual situations.

Training effectiveness addresses the question whether S-OJT leads to better training outcomes when compared to other training approaches and whether the investment to achieve those training outcomes results in financial benefits in terms of improved work outcomes. An illustrative study in this regard was reported by Jacobs (2002c), who compared the value of the work defects (leaks found in windshields in a truck assembly plant) when employees had received unstructured buddy training and S-OJT. The results are reported in the cost of defects reduced when using S-OJT, while also taking into consideration the cost of the training.

Additional research has been conducted to understand how S-OJT might be better understood and extended. For example, Lohman (1994) investigated the effectiveness of inductive training strategies compared to the more commonly used deductive training strategies. The research suggested that using an inductive strategy might be possible, but this changes, to a large extent, the role of the trainer and trainee. Cushnie (2000) conducted a qualitative study to investigate the extent to which mentor-mentee relationships naturally occur along expected norms as established in the literature. The results suggest that mentors and mentees may not have similar backgrounds in order for the relationship to be considered a success. Cho (2004) studied the unanticipated benefits of being an S-OJT trainer in a large Korean organization.

Barnard (2005) differentiated between near and far transfer as it relates to training outcomes. S-OJT has long been considered effective in achieving near transfer of training, or enabling the trainee to perform exactly what was presented during the training. On the other hand, far transfer of training proposes that if an individual learns more general principles, they might be able to extrapolate that information and use it in related, but not identical, situations later on (Kim and Lee). Barnard found some differences in the way supervisors viewed their expected level of performance in performing a coaching task.

Additional survey research has been conducted to understand the impact of organizational change on the training and learning decisions made in organizations. Osman-Gani and Jacobs reported that technological advancements were the most frequent driver of change

among organizations in Singapore, and that S-OJT was the most frequently used training approach used to respond to these changes. Quazi and Jacobs surveyed organizations involved in ISO 9001 and found that their training and development activities had dramatically increased during and continued on after this effort. Learning on the job was identified as the most frequently used training approach for frontline employees.

Implications of Workplace Training and Learning

It is clear that the workplace training and learning will have growing importance in organizations. The emergence of flexible work systems, flatter organizational structures, and a downsized workforce will be ever more dependent on developing employees to perform a wider range of skills. How managers respond to this landscape will have implications on workplace training and learning. Three related implications are presented here for discussion.

First, while the article has emphasized S-OJT, there is also a need to explore the possibilities of other informal training approaches, regardless of whether they are planned in nature. Watkins and Marsick (1990) continue to espouse the value of learning in informal settings based on the notion that learning and knowledge generation oftentimes go hand in hand (Watkins, 2005). Learning in informal work settings has been shown to be a social process (Enos *et al.*, 2003). Planned learning assumes that the knowledge has been previously defined and ready to be imparted to the trainee. An alternate perspective is to have the learner become involved in the learning experience through the sharing of stories, action learning, collaborative inquiry, and questioning strategies. Such processes bring a unique richness to the learning setting that may not be possible otherwise.

The second implication proposes that, given the value of informal learning settings, there should be a strategic blend of both planned and unplanned approaches used within situations. How to view planned and unplanned approaches as being complementary is yet to be accomplished in practice. However, given the demonstrated value of each, it makes sense to view them as each having potential to contribute. For instance, when developing new-hire engineers, it seems appropriate to use S-OJT as a means to provide known information about the position. However, similar to many positions that require making decisions in sometimes-ambiguous situations, an action learning approach might be used to help the engineers untangle especially complicated and uncertain incidents they might encounter.

The final implication is that whatever approach might be taken, there needs to be continued awareness that the

learning experience should meet both the needs of the organization and the individuals alike. Past discussions about the foundational purpose of the human resource development field – whether it is intended for improving organizational performance or to promote individual learning – appear moot at this point of time. In truth, one goal cannot likely be accomplished without considering the other as well. That is, the goals are both distinct and complementary at the same time. That seems the fundamental challenge of ensuring the most effective use of workplace training and learning programs.

See also: Apprenticeship Approach to Learning; Organizational Learning; Trends in Workplace Learning Research; Workplace Learning Frameworks.

Bibliography

- Ackerman, F. (1998). *The Changing Nature of Work*. Washington, DC: Island.
- Al-Muzaini, B., Al-Keane, A., and Al-Awadi, S. (2002). Transitioning from unstructured to structured on-the-job training. In Jacobs, R. (ed.) *Implementing on-the-Job Learning*, pp 41–43. Alexandria, VA: ASTD.
- Barnard, J. K. (2005). *The Effects of a Near Versus Far Transfer of Training Approach on Trainees' Confidence to Coach Related and Unrelated Tasks*. Dissertation: THE:EDN2005PHDB376. Columbus, OH: Ohio State University.
- Barnett, R. (1999). Learning to work and working to learn. In Boud, D. and Garrick, J. (eds.) *Understanding Learning at Work*, pp 29–44. London: Routledge.
- Beckett, D. and Hager, P. (2002). *Life, Work and Learning: Practice in Postmodernity*. London: Routledge.
- Broad, M. L. (2005). *Beyond Transfer of Training: Engaging Systems to Improve Performance*. San Francisco, CA: Pfeiffer.
- Cho, D. Y. (2004). *Unintended consequences among trainers delivering structured on-the-job training in a Korean organization*. Unpublished doctoral dissertation, Columbus, OH: The Ohio State University.
- Cushnie, M. E. (2000). *African-American women first-line supervisors : A qualitative study of their career development process*. Unpublished doctoral dissertation, Columbus, OH: The Ohio State University.
- Dooley, C. R. (1945). *The Training Within Industry Report (1940–1945): A Record of the Development of Management Techniques for Improvement of Supervision: Their Use and the Results*. Washington, DC: War Manpower Commission, Bureau of Training, Training Within Industry Service.
- Enos, M. D., Kehrhahn, M. T., and Bell, A. (2003). Informal learning and the transfer of learning: How managers develop proficiency. *Human Resource Development Quarterly* 14(4), 369–387.
- Friedman, T. L. (2005). *The World is Flat: A Brief History of the Twenty-First Century*. New York: Farrar, Straus and Giroux.
- Hager, P. (2004). Conceptions of learning and understanding at work. *Studies in Continuing Education* 26(1), 3–17.
- Jacobs, R. (2001). Managing employee competence and human intelligence in global organizations. In Richter, F. (ed.) *Maximizing Human Intelligence in Asia Business: The Sixth Generation Project*, pp 125–137. New York: Prentice-Hall.
- Jacobs, R. (2002a). Implementing structured on-the-job learning. In Jacobs, R. (ed.) *Implementing Structured On-the-job Learning: A Casebook*, pp 1–8. Alexandria, VA: American Society for Training and Development.
- Jacobs, R. (2002b). Using structured on-the-job training to inform new supervisors of the organization's core values. In Jacobs, R. (ed.) *Implementing Structured On-the-job Learning: A Casebook*, pp 97–110. Alexandria, VA: American Society for Training and Development.
- Jacobs, R. (2002c). Implementing structured on-the-job learning. In Jacobs, R. (ed.) *Implementing Structured on-the-Job Learning: A Casebook*. Alexandria, VA: American Society for Training and Development.
- Jacobs, R. (2005). Comparing the forecasted financial benefits of blended training, classroom training, and structured on-the-job training. *Paper Presented at the ASTD, Research to Practice Conference*, Orlando, FL, USA.
- Jacobs, R. and Hruby-Moore, M. (1998). Comparing the forecasted and actual financial benefits of human resource development programs: Learning from failure. *Performance Improvement Quarterly* 11(2), 93–100.
- Jacobs, R., Jones, M., and Neil, S. (1992). A case study in forecasting the financial benefits of unstructured and structured on-the-job training. *Human Resource Development Quarterly* 3(2), 133–139.
- Jacobs, R. and McGiffin, T. (1987). A human performance system using a structured on-the-job training approach. *Performance and Instruction* 26(5), 8–11.
- Jacobs, R. and Osman-Gani, A. (2005). *Case Studies in Workplace Training and Learning: A Cross-Cultural Perspective*. Singapore: Pearson Prentice Hall.
- Jacobs, R. L. (2003). *Structured On-the-Job Training: Unleashing Employee Expertise in the Workplace*. San Francisco, CA: Berrett-Koehler.
- Johnson, S. and Leach, J. (2001). Using expert employees to train on the job. In Jacobs, R. (ed.) *Advances in Developing Human Resources* 3(4), 425–434.
- Kim, J. H. and Lee, C. (2001). Implications of near and far transfer of training on structured on-the-job training. *Advances in Developing Human Resources* 3(4), 442–451.
- Lee, C. (2004). *Perceived Job Change toward Dimensions of Knowledge Work among Three Levels of Employees in a Korean Bank*. Doctoral Dissertation: THE:EDS2004PHDL432. Columbus, OH: Ohio State University.
- Liu, L., Wang, B., and Jacobs, R. (2005). ISO 10015 – Quality of training as a standard for the human resource development field. In Osman Gani, A. A. (ed.) *Proceedings of the annual conference of the Asian Chapter of the Academy of Human Resource Development*. Taipei, Taiwan.
- Lohman, M. C. (1994). *The effects of a guided discovery versus a deductive training method on supervisors' ability to solve problems and facilitate problem-solving skill of subordinates*. Unpublished doctoral dissertation, Columbus, OH: The Ohio State University.
- MacLean, R. and van der Pol, H. (2006). *Participation in Formal Technical and Vocational Education and Training Programmes Worldwide: An Initial Statistical Study*. Bonn, Germany: UNESCO (UNEVOC International Centre for Technical and Vocational Education and Training).
- Mafi, S. (2001). Planned on-the-job managerial training. In Jacobs, R. (ed.) *Advances in Developing Human Resources*, 3(4), pp 488–495.
- Mohrman, S. A. (2003). Designing work for knowledge-based competition. In Jackson, S. E., Hitt, M. A., and DeNisi, A. S. (eds.) *Managing Knowledge for Sustained Competitive Advantage: Designing Strategies for Effective Human Resource Management*, pp 94–123. San Francisco, CA: Jossey-Bass.
- Rivera, R. and Paradise (2006). *2006 State of the Industry in Leading Enterprises: ASTD's Annual Review of Trends in Workplace Learning and Performance*. Alexandria, VA: ASTD.
- Sambrook, S. (2005). Factors influencing the context and process of work-related learning: Synthesizing findings from two research projects. *Human Resource Development International* 8(1), 101–119.
- Srinivasan, S., Lewis, D., and Crooks, S. (2006). Effects of concreteness and contiguity on learning from computer-based reference maps. *Journal of Interactive Online Learning* 5(3), 243–253.
- Stein, D. (2001). Situated learning and planned training on the job. *Advances in Developing Human Resources* 3(4), 415–424.
- Watkins, K. (2005). Rethinking workplace learning in a knowledge economy. *International Journal of Adult and Continuing Education* 8(4), 147–164.
- Watkins, K. and Marsick, V. (1990). *Informal and Incidental Learning in the Workplace*. New York: Routledge.

Further Reading

- De Jong, J., Thijssen, J., and Versloot, B. (2001). Planned training on the job: A typology. In Jacobs, R. (ed.) *Advances in Developing Human Resources*, **3**(4), 408–414.
- Jacobs, R. (1994). Case studies that compare the training efficiency and product quality of unstructured and structured OJT. In Phillips, J. (ed.) *The Return on Investment in Human Resource Development: Cases on the Economic Benefits of HRD*, pp 123–132. Alexandria, VA: American Society for Training and Development.
- Jacobs, R. and Hruby, M. (1998). Comparing the forecasted and actual financial benefits of human resource development programs: Learning from failure. *Performance Improvement Quarterly* **11**(2), 93–100.
- Jones, M. and Jacobs, R. L. (1995). Developing frontline employees: New challenge for achieving organizational effectiveness. In Kaufman, R., Thiagarajan, S., and MacGillis, P. (eds.) *The guidebook for performance improvement: working with individuals and organizations*, pp 241–260. San Diego, CA: Pfeiffer.
- Kim, J. and Lee, C. (2001). Implications of near and far transfer of training on structured on-the-job training.
- In Jacobs, R. (ed.) *Advances in Developing Human Resources*, **3**(4), 442–451.
- Lohman, M. (1997). Effects of an inductive versus deductive instructional approach on the constructive feedback and problem-solving skills of supervisors. *Performance Improvement Quarterly* **10**(3), 37–55.
- Lohman, M. (2001). Deductive and inductive on-the-job training strategies. In Jacobs, R. (ed.) *Advances in Developing Human Resources*, **3**(4), 435–441.
- Mortenson, K., Julin, B., and Wang-Chavez, J. (2005). The real deal reaching high levels of employee competency and productivity with structured on-the-job training (SOJT). *Presentation at the Annual Conference of the International Society for Performance Improvement*. Vancouver, BC, USA.
- Osman-Gani, A. and Jacobs, R. L. (2005). Technological changes and human resource development practices in Asia: A study of Singapore-based companies. *International Journal of Training and Development* **9**(4), 271–280.
- Quazi, H. A. and Jacobs, R. L. (2004). Nature and impacts of ISO certification on training and development activities on Singapore organizations. *International Journal of Quality and Reliability Management* **21**(5), 497–517.

Training in Small Businesses

H Matlay, Birmingham City University, Birmingham, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

As a topic of research, training in small businesses has progressed a long way from its humble beginnings as a minor subtopic of human resource management (HRM). Interestingly, during its long infancy, researchers and practitioners have invested a great deal of time and resources in an attempt to mold large business training solutions into small firm employee-development programs. It was not until the 1970s, when the importance of the small business sector began to be acknowledged, that training in small businesses broke free from its conceptual and contextual straightjacket and evolved into a research topic of its own. Currently, this topic occupies a central position within the wider training debate and relevant dissemination represents a major contribution to the fast-growing small business and enterprise development field of research (Matlay, 2004).

There are a number of important reasons why training in small businesses is considered important to the modern economy. First, small businesses represent a vast proportion of the modern economy, amounting to about 99% of all economic units in both developed and developing countries. Second, they make a significant contribution to job creation and mitigate the unemployment caused by the restructuring and downscaling tendencies of large businesses and multinational corporations. Third, due to their flatter managerial structure and inherent flexibility, small firms can offer fast and focused responses to rapid changes in local, regional, and global competitive conditions (Westhead and Matlay, 2006). Thus, a healthy and growing small business sector is widely recognized as an important contributor to the socioeconomic and political infrastructure of a nation.

In many countries, due to the importance attributed to the small business sector, governments invest considerable resources in their drive to increase the uptake of training and improve productivity and competitiveness. Small business owner/managers are encouraged to evaluate their employees' training needs and to develop specific programs that would upskill, motivate, and enhance their workforce (Matlay, 1999a). In turn, small business employees are offered incentives to develop and enhance their job-specific skills. Researchers in this area, however, have noticed that, as compared to larger organizations, training uptake in small businesses is relatively low and that there remain critical factors that act as barriers to employee development in smaller-sized economic units.

In this article we undertake a critical review of the most important aspects of training in small businesses and offer some constructive recommendations to mitigate persistent barriers and difficulties (Matlay, 1999b).

Conceptual and Contextual Considerations

There exists a considerable body of empirical as well as practitioner knowledge that emphasizes the importance of training in small businesses. Conventional wisdom attributes considerable benefits to all the parties involved in the training contract. The rhetoric relating to training in smaller firms states that providing training for employee development and upskilling confers a wide range of benefits to employers, employees, and society as a whole. In theory, employers could take advantage of free or subsidized training, expect to recover their investment in employee development, and benefit from increased productivity and competitiveness. Similarly, to better-trained and upskilled employees, there would accrue benefits in terms of improved job security, commensurable pay increments, and increased job satisfaction. In practice, however, small business owner/managers distrust government intervention, have difficulty in locating and securing relevant subsidies, and tend to operate mostly on short-termist strategies that invariably construe training as a cost to the firm rather than an investment in employee skills and/or productivity. Most small business employees cannot find the time or cover the cost of personal development and resent investing in their training for the benefit of employers who are reluctant to provide job-specific programs.

From an academic and practitioner perspective, there is recognition of, and convergence on, the importance of the skill levels and training needs of individuals employed in smaller business. Unfortunately, there appears to be little agreement on how much and what type of training should be provided in small businesses. Similarly, various studies on this topic have generated a mass of contradictory results, ranging from praise to condemnation of small business owner/managers for their efforts, level of investment, and consistency of training provision in their firms. Much of the confusion in this field originates in conceptual and contextual difficulties that are inherent in small business research. For instance, there are no standard or widely accepted definitions of training and human

resource development. Similarly, there is no general agreement as to what constitutes a small business. To complicate matters, there are also considerable methodological disparities across research studies that focus specifically upon the training needs of small business owner/managers and their workforce. Cumulatively, these problems make generalization and cross-country comparisons difficult and render most emergent results as situational and therefore only relevant to a narrow context.

To overcome some of the problems inherent in this research topic, it is decided to adopt a wide and convergent approach and avoid, as far as possible, definitional debate. Thus, a number of working definitions are used in order to minimize conceptual and contextual differences and maximize the generalization potential of emerging results. Training and employee development are used interchangeably to denote a wide variety of training opportunities provided and/or financed by employers. A small business is defined as an economically active and wholly owned independent unit, employing in addition to the owner(s) between one and 50 employees or full-time equivalent (FTE). References have been kept to a minimum and readers who are interested in the empirical aspects of the training debate are referred to the author's extensive publications in academic and practitioner journals as well as relevant chapters in textbooks and training monographs.

Motivation to Train

There are considerable similarities and differences in training between small and large organizations. Similarities involve issues relating to the motivation to train and the rates of return on investment in HRM. Differences tend to relate to outlook and expectations as well as to resource allocation and approaches to what are basically owner/manager controlled employee-development tasks. The motivation to train emerges as a crucial and overriding aspect of the training function, which affects most if not all of the aspects relating to management learning, employee training, and personal development in a small business context.

From a narrow perspective, Kitching and Blackburn (2002) argue that workforce training in small businesses has two purposes. First, training provides new employees with sufficient skills and job-specific knowledge to enable them to perform their tasks and roles to a minimum standard. Second, training is aimed to secure new employees' cooperation in relation to the small business owner/manager's aims and objectives. Cumulatively, these two motives accounted for 71% of training reasons in smaller firms. Other, less important reasons to train included: the introduction of new hardware and software, compliance

with legal obligations, and the provision of qualifications for employees.

In contrast, the author (see Matlay, 2002) takes a wider, interrelated approach to motivation to train in small businesses. This view includes a systemic as well as organizational perspective and highlights the complexities involved in training provision within a varied and heterogeneous small business sector. First, there is a considerable mismatch between the supply of, and the demand for, small business training on offer in an economy. This is mainly due to a combination of government interference in the training market and owner/manager reluctance or inability to purchase firm-specific provision at an economic price. Second, training needs in small businesses are increasingly characterized by multi-skilling and diversity of application. This is a reflection of inherent advantages of scope and scale as well as flexibility to fulfill short-term niche-market requirements. Third, the importance of the owner/managers to small business operation and their reluctance to relinquish overall control tend to add credence to the ignorance hypothesis in relation to their training decision-making processes. The author found that the main motive to train in small businesses relates to the pressures of gaining and sustainable competitive advantage under constrained resource allocation conditions, followed by legal compliance and delegation considerations.

Types of Training Provision

Small business owner/managers tend to provide mostly job-specific training that, in their perception, is directly relevant to their immediate operational needs. Unfortunately, the author found that few owner/managers in this category carry out regular skill audits or measure the training needs of their workforce. Thus, the vast majority of training provision in small businesses is reactive in nature and originates from immediate skills shortages or acute production crises. Consequently, there is seldom time to evaluate the specific training needs of employees or choose the best, most appropriate, and cost-effective training program or staff-development route. The author found only a handful of cases involving proactive training provision and these were invariably part of formal staff-development initiatives in the larger of the small firms in the research sample. Moreover, reactive training was mainly driven by the quality assurance or technological upgrading of productive or service capabilities requested by their main customers.

It should be noted that the incidence and intensity of both reactive and proactive training provision tends to increase in direct proportion with the size of the firm and can vary considerably across the widely heterogeneous small business sector. Importantly, training

provision closely reflects the preferred managerial styles of these owner/managers, which is also a function of size. Almost all the owner/managers of firms employing fewer than ten employees or FTE (i.e., micro-businesses) preferred informal managerial styles. Formative, mixed formal, and informal managerial styles were reported by respondents who employed between 11 and 30 individuals or FTE. In the larger of these small firms (e.g., those with 31 to 50 individuals or FTE) formal managerial styles were well established.

The three different types of training can be explained as:

1. **Formal training** – It involves structured and prescribed training courses and programs, directly relevant to the specific needs of small businesses. General managerial courses geared specifically for management development are also included in this category. This type of training course is formally structured and scheduled around well-defined objectives and in pursuit of specific outcomes. Formal training courses and programs in small businesses usually take place away from the immediate operational base of a firm and are delivered by qualified and/or designated trainers. Relevant feedback is collected from both the trainer and the trainee. Training outcomes are routinely discussed with the trainee and formalized. A detailed report and feedback comments are also sent to the owner/managers who chose and funded the training course or program.
2. **Informal training** – It includes a wide variety of casual, incidental, or accidental and largely unplanned training inputs, aimed narrowly at imparting knowledge and/or experience of isolated task or problem-solving solutions. These are mostly unstructured, of relatively short length, and in pursuit of a set of narrow outcomes. Typically, informal training takes place within the immediate operational base of a firm and can be delivered by owner/managers or better-trained or experienced employees. Usually, the training feedback stage is bypassed or ignored. This makes the measurement of outcomes difficult to collect, record, or analyze. Furthermore, a large number of informal training incidents are part of routine on-the-job supervision and management and not acknowledged or logged.
3. **Mixed formal and informal training** – This type of training incorporates elements of both formal and informal training intervention. This type of training often forms part of the personal development of managers or shadow managers in the more sizeable of small businesses, where owner/managers often coach or mentor some of the more capable individuals to take on managerial roles. In order to motivate and reward such individuals, owner/managers choose a number of formal training courses of various lengths and send them to widen their general and/or specialist managerial skills. Interestingly, formal training takes place

away from the firm, while informal aspects are undertaken on the premises and delivered by owner/managers. Incidents of relevant training feedback generally fall within the formal and informal patterns and stereotypes are observed in the relevant categories.

The Paradox of Training in Small Businesses

The specialist literature on training in small businesses claims that owner/manager attitudes determine their preferences in relation to this important organizational function. Our initial pilot study found a direct link between owner/managers attitudes to and actual provision of training (Hyland and Matlay, 1997). The wider, main study however established a paradox of training in small businesses, which is detailed and explained below:

1. **Attitudes to training:** It is generally acknowledged that small business owner/managers make most, if not all, the important decisions in their firms. As such, they are involved in all the important operational aspects of running their businesses, including the training function. The issue of who is or should be responsible for training in small and medium enterprises (SMEs) is a decisive aspect of training provision in smaller firms and tends to impact considerably upon related outcomes. Previous research on this topic has highlighted some attitudinal variations in smaller businesses as well as owner/manager ambivalence in relation to the degree of responsibility that they are prepared to accept for training and personal development within their firms. Longitudinal, cross-national research carried out by the author has established that just over 90% of small business owner/managers claim to have a positive or very positive attitude to training. Fewer than 10% of respondents claimed to be either indifferent or have negative attitudes to this important organizational function.
2. **Actual provision of training:** Research focusing upon actual provision of training small businesses has provided contradictory results (Matlay, 2000). Depending upon the measurement tools and definitions adopted, some studies established that about one-thirds of the respondents in their sample have provided some training. In contrast, the more optimistic results claimed that as much as two-thirds of the small business owner/managers interviewed had provided training for their workforce. Using consistent and comparable measurement tools and definitions, our research has established that just over 90% of owner/managers in the sample failed to provide any training during the calendar year preceding the interviews. Considering that a similar

proportion of respondents returned positive and very positive attitudes to training, the small percentage of training in small businesses is both surprising and worrying.

3. The small business training paradox: The emerging results of an empirically rigorous research study has established the existence of a small business training paradox: about 90% of owner/managers interviewed claimed to have a positive and very positive attitude to training while a similar proportion admitted that they have not provided any courses or programs to their workforce (Matlay, 1998). This compares unfavorably with the training provision in large organizations and multinational corporations. Considering the literature on employee recruitment, retention, and development, it becomes clear that a lack of training can have a detrimental effect upon the motivation and productivity of small business employees in general and firm competitiveness in particular. The next section outlines some of the more important factors that affect training in small businesses and their impact upon owner/manager provision of training.

Factors Affecting Training Provision

Owner/managers in our research sample highlighted two main categories of factors that influenced their decision-making processes in relation to the provision of training in their small businesses:

1. Directly relevant factors – These factors, such as the market positioning of a firm, prevailing economic conditions, and the availability of relevant training, were perceived as being of primary importance to the training-related, decision-making process of small business owner/managers.
2. Indirectly relevant factors – They were considered to be of secondary importance to the training strategies of these owner/managers, but still exerted a considerable influence upon the quantity and quality of the training provision in small businesses.

Individually and cumulatively, the effects of these factors can explain, to some extent, the gap that exists between the small business owner/managers' attitudes to, and the actual provision of, training in their firms. A brief analysis of both directly and indirectly relevant factors to affect training provision in small businesses is provided below:

1. Directly relevant factors
 - Market positioning – The market positioning of a small firm has emerged as the most important of the directly relevant factors to affect the provision

of training within the research sample. Market-positioning strategies were used by owner/managers in all sectors of economic activity to promote and discriminate their own products or services from those of their competitors. Thus, product- or service-related quality issues (arguably, the main discriminant factor at the disposal of an owner/manager) emerged as the most important determinant of training and human resource strategies in the small business sector.

- Prevailing economic conditions – The economic conditions prevailing at the time that training-related decisions were made was the second most important factor to directly affect the provision of training in smaller firms. Interestingly, periods of boom and bust tend to influence owner/manager perceptions. Thus, the relative growth in the demand for goods and services which occur during boom periods can have a positive effect on levels of recruitment and incrementally increase training needs. Conversely, recessionary conditions have an adverse effect upon training, when small business owner/managers shed some staff and become reluctant to spend on staff development.
 - Availability of relevant training – The availability of relevant training was the third of the directly relevant factors that small business owner/managers claimed to have affected the provision of training in their firms. In the majority of cases, the decision to train or retrain the workforce was implicitly or explicitly included in market-positioning strategies. During favorable economic conditions, owner/managers tend to scan the market in an attempt to identify the best sources of training. A lack of relevant, firm-specific training can significantly handicap small business human resource development strategies.
2. Indirectly relevant factors
 - Cost of training – Due to a lack of relevant and firm-specific training experienced by most of the respondents, the majority of owner/managers have been forced to pay the going market rate for training their workforce. It appears that they consider these rates as very expensive and inflated. Furthermore, almost all of them seem to have encountered difficulties in identifying and evaluating the cost of training schemes. It should be noted, however, that attempts to compute the full extent of actual, marginal, or incidental costs relating to training were rarely encountered.
 - Time constraints – Time constraints which resulted from short staffing appear to have imposed considerable constraints upon some small business

owner/managers. This affected the provision of training at various key stages in the process, including the assessment of training possibilities and choices as well as the cost effectiveness and feedback analysis of actual training schemes. Some owner/managers were forced by time constraints to delegate the training function to other managers or members of their family who, arguably, did not fully realize its importance to the overall business strategy.

- Lack of trainee cover – The lack of in-house trainers mainly reflected a respondent's preference for internal methods of training upon which s/he could have full control. In the smaller businesses in the sample, even one employee away from his/her usual workplace amounted to a significant loss in productive output.
- Lack of in-house trainers – Few small businesses can afford the luxury of in-house trainers or training managers. Arguably, a dedicated employee could take responsibility for training and relieve the pressure on the owner/manager. A small number of the larger firms in the sample employed part-time training managers or self-employed trainers.
- Lack of trainee motivation – Interestingly, a lack of motivation for undergoing training among employees sometime restricts training efforts in small businesses. Owner/managers are often reluctant to reward or promote trained employees in case they leave or are poached by competitors.
- Lack of trainee interest – This aspect is linked to staff motivation and can also have a negative effect on training strategies in small businesses. Employees can and often question the motives that owner/managers have for training their staff. In some cases, their suspicions have been confirmed when production line speeds were increased in accordance with post-training productivity expectations.

In terms of the training paradox identified in the small business sector, the directly relevant factors identified by owner/managers explain, at least partially, the gap between attitudes and actual provision. Most of these factors can act to restrict both reactive and proactive training strategies in small businesses. Arguably, by taking a more proactive approach toward the training function in their firms, small business owner/managers can mitigate or even avoid some of the financial costs associated with of the negative factors outlined above.

Concluding Remarks

The training function in small businesses differs significantly from training solutions adopted by large organization

and multinational corporations. Although there is a great deal of debate about all aspects of training, certain trends emerge that set this function in small businesses apart from training and human resource development in larger organizations. The main motive to train in small businesses relates to gaining and sustainable competitive advantage under constrained resource allocation conditions. In addition, legal compliance and delegation considerations as well as multi-skilling needs also affect owner/managers' motivation to provide training to their workforce. Much of the training provided is reactive rather than proactive in nature and reflects largely the informal management style preferred by most small business owner/managers.

The training paradox identified by the author relates to the wide gap that continues to exist between the owner/managers' overwhelmingly positive attitudes to training and their apparent reluctance or inability to provide it adequately in their firms. Two type of factors influence owner/managers' decision-making processes in relation to the provision of training in their small businesses: directly relevant and indirectly relevant factors. Directly relevant factors involve the market positioning of a firm, prevailing economic conditions, and the availability of relevant training. These were perceived to be of primary importance to the training-related, decision-making process of small business owner/managers. Indirectly relevant factors included the cost of training, time constraints, lack of trainee cover and in-house trainers, as well as a lack of trainee motivation and/or interest. Indirectly relevant factors were considered to be of secondary importance but were still influential to the training strategies of small business owner/managers.

Most small business owner/managers feel that the training market is supply oriented and that government initiatives tend to support training providers rather than their prospective customers. Similarly, much of the information related to training availability and relevant subsidies are directed toward training advisors and providers. Therefore, when seeking advice on training, owner/managers claim to notice a marked bias toward off-the-shelf training solutions that do not match their needs or expectations. Furthermore, researchers have noticed that most training initiatives are reviewed by commercial organizations or government-sponsored institutions. These reviews lack empirical rigor and credibility and invariably claim success, even when some initiatives blatantly failed to address their explicit small business training agenda. It is therefore suggested that training initiative should be aimed directly at specific needs of small business owner/managers and their workforce and should support the demand side of the training provision equation. Training information and relevant subsidies should be made available to owner/managers either through direct mail or via independent small business advisors or intermediaries.

Bibliography

- Hyland, T. and Matlay, H. (1997). Small businesses, training needs and VET provision. *Journal of Education and Work* **10**(2), 129–139.
- Kitching, J. and Blackburn, R. (2002). *The Nature of Training and Motivation to Train in Small Firms*, Research Report RR330. London: Department for Education and Skills.
- Matlay, H. (1998). The paradox of training in the small business sector of the British economy. *Journal of Vocational Education and Training* **49**(4), 573–589.
- Matlay, H. (1999a). Vocational education and training in Britain: A small business perspective. *Education and Training* **41**(1), 6–13.
- Matlay, H. (1999b). Vocational education, training and organisational change: A small business perspective. *Strategic Change* **8**(5), 277–286.
- Matlay, H. (2000). The future of workplace learning and knowledge management: A small business perspective. *Journal of Management Research* **23**(11), 110–121.

- Matlay, H. (2002). Training and HRD strategies in family and non-family owned small businesses: A comparative approach. *Education and Training* **44**(8–9), 357–369.
- Matlay, H. (2004). Contemporary training initiatives in Britain: A small business perspective. *Journal of Small Business and Enterprise Development* **11**(4), 504–513.
- Westhead, P. and Matlay, H. (2006). Skills associated with employment positions in SMEs and favourable attitudes towards self-employment: Longitudinal evidence from students who participated in the Shell technology enterprise programme. *Technology Analysis and Strategic Management* **18**(1), 93–124.

Further Reading

- Matlay, H. (2001). New labour, new deal? Employment relations in the small business sector of the British economy. *Journal of Management Research* **24**(10/11), 43–61.

VOCATIONAL EDUCATION AND TRAINING – RESEARCH ON VET

Contents

Language of Vocational Education and Training

Research on Vocational Education and Training

Vocational Education and Training Information Resources

Language of Vocational Education and Training

J Misko, J MacKenzie, and R-A Polvere, National Centre for Vocational Education Research, Adelaide, SA, Australia

© 2010 J Misko, J MacKenzie, and R-A Polvere. Published by Elsevier Ltd. All rights reserved.

When we speak or write about vocational education and training (VET) in any culture, we often use the language borrowed from government policy statements and diverse academic disciplines, occupations, and industry sectors.

We note that there are differences in language and terminology between countries and regions, for example, in Australia and in much of Europe, the term vocational education and training (VET) is used. The United States of America uses the term career and technical education (CTE) and vocational education. The United Kingdom uses the term further education (FE). South Africa uses the term further education and training (FET), while most of the Asian countries tend to employ the term technical and vocational education and training (TVET). The term TVET has also been favored by the Organization for Economic Cooperation and Development (OECD). Although different cultures have their own preferred names for VET, the sector generally aims to prepare people to enter specific trade-and-technician occupations, or make the transition from school or unemployment to work. It also provides opportunities for existing workers to upgrade their skills and qualifications, marginalized groups to have a second chance at education and training, and adults to engage in personal development. In the last 20 years or so, VET has been associated with the provision of lifelong learning.

In this article, we argue that there is no unique language for VET and provide some examples of how explanations of VET arrangements are influenced by the languages of different fields. We use the acronym VET (favored in this encyclopedia) for ease of discussion.

The article is divided into themes dealing with how terminology associated with particular fields has influenced the language of VET. We look in turn at government policy, economics, globalization and technological

innovation, skills, useful metaphors for VET structures and strategies, and teaching, learning, and assessment. We end the article by drawing some conclusions.

VET Uses the Language of Government Policy

During the last 20 years or so, many countries have used the terminology favored by governments to announce training-reform strategies aimed at furthering economic prosperity and competitiveness at home and abroad. From the mid-1980s onward, we see industrialized countries (including Australia, New Zealand, and the United Kingdom) reforming traditional VET systems to address increased competition from low-wage countries (including China, Indonesia, Malaysia, and Thailand) for low-skilled routine manufacturing jobs, and other OECD countries with higher levels of education (including Japan, Korea, and Germany) and innovative work practices (including Japan and Sweden) for high or intermediate-level skilled manufacturing jobs producing value-added goods. Increasingly, customer service and administrative jobs in a variety of industries (including telecommunications, information technology, financial services, tourism, and hospitality) are also being exported offshore to other low-wage countries including India and Thailand.

Across the world, the desire for a skilled and flexible workforce to enable domestic industry to compete in international markets has driven training reform. Terms like manpower development, human-resource development, workforce development and skills formation have been used as umbrella terms for government initiatives to increase the national skills pool. In Australia, with industry expected

to play a key role in developing these reforms, language describing VET as an industry-led system begins to emerge.

Due to its tight connection to the world of work, VET is the sector considered most able to help students move from the schoolyard to the workplace, especially for those not immediately bound for university. Terms like school-to-work transition become popular terminology for describing this movement.

Governments must also be concerned with addressing the needs of all citizens, whatever their socioeconomic or cultural background, and often espouse and provide funding to support goals which aim to provide citizens with fair and equitable opportunity and treatment in education, employment, and social life. Terms and concepts like access and equity, cultural diversity, social inclusion, and second-chance education are regularly used to describe government initiatives for addressing the needs of marginalized groups via VET.

Explanations which appeal to the need for VET to prepare workers for what has come to be called the knowledge economy, information economy, or knowledge society have also emerged in the VET literature. This type of language is used to help make a distinction between economies and occupations requiring complex cognitive processes and those requiring the lower-level cognitive processes needed for low-skilled routine jobs.

Rapid advances in information and telecommunications technology, workplace change, and workforce aging are regularly used to justify government policies aimed at encouraging workers to maintain their competitive edge in changing labor markets by upgrading their skills and knowledge. In countries with an aging workforce and declining proportions of younger people (like many OECD countries including China) governments have also considered the need to provide incentives to encourage workers, who might otherwise go into retirement, to continue working. As a result, we see increased use of terms like existing workers, mature-age workers, and older workers used in VET as it tries to customize training arrangements and teaching methodologies to the needs of these groups.

Perhaps the most pervasive of reform terminology that has moved from policy documents into the language of VET worldwide is lifelong learning. Across the globe, this refers to ongoing education and training aimed at improving current and future employment prospects or life in general. As VET has historically had more flexible approaches to admission and delivery than other sectors, it is especially equipped to help citizens participate in training or personal development throughout life.

In recent years, cross-border employability and transportability of skills and qualifications have been major preoccupations of governments as they deal with increased levels of cross-border mobility of labor. These issues have mainly dealt with improving the transparency and

portability of qualifications. The European concepts of the skills passport and the vocational certificate supplement and the Australian and New Zealand concept of mutual recognition have been introduced to help migrant workers gain appropriate work.

The concept of VET promoted by the European Union in its Copenhagen Declaration provides us with further examples of how terminology and rhetoric favored by governments and their representatives in international forums capture the way we think and talk about VET:

... Strategies for life-long learning and mobility are essential to promote employability, active citizenship, social inclusion and personal development. Developing a knowledge-based Europe and ensuring that the European labour market is open to all is a major challenge to the vocational educational and training systems in Europe and to all actors involved. The same is true of the need for these systems to continuously adapt to new developments and changing demands of society. (Copenhagen Declaration, 2002)

This vision might be used to describe the aspirations that many countries have for their VET system (Table 1).

Useful to VET Are Economist Explanations and Predictions of Labor-Market Environments and Activity

The language of economics and economists plays a central role in explaining the VET environment. This increasing foray into a field traditionally occupied by the language of educators is buttressed by a perceived need for explanations and measurements used by economists to gauge the effectiveness of government reforms. Some of these measures are terms like returns on investment in training, student retention and attrition, program completions, reductions in unemployment rates (and especially long-term unemployment), improvement in human capital, and achievement of access and equity for marginalized groups. Another concern of economists which is reflected in the language of VET is unmet demand, that is, when the demand for training outstrips the supply of training.

The analytical tools generally applied by economists are also used to predict the supply and demand for training. Increasingly, attention has focused on responding to the impact of demographic change and responding to the needs of an aging workforce. The concept of the market and the terminology associated with it has also been used extensively to describe the VET environment (Table 2). Terms like training market and labor market (including these for adults and youth) have been used in turn to explain the arenas in which training providers compete for students and courses, and individuals compete for jobs.

Table 1 Terms used in policymaking and governance which have entered the language of VET

Access and equity	An umbrella term for enabling equality of access to education and training and employment for marginalized groups
Cultural diversity	A recognition that VET should address the range of individuals from non-mainstream or marginalized groups
Employability skills	A set of personal skills and attributes that enable an individual to compete successfully for jobs
Information economy, Knowledge economy	An economy concerned with the production and trading of knowledge and information services rather than material goods and services (Porat, 1977).
Knowledge society, Information society	A society 'whose processes and practices are based on the production, distribution and use of knowledge' (CEDEFOP, 2005)
Knowledge workers	These are workers whose main activities are concerned with the transformation of information into commodities rather than the production of material goods. (Drucker, 1992)
Life-long learning	On-going education and training throughout one's lifetime
Manpower development Workforce development Human resource development	National or provincial arrangements for increasing the national skills knowledge pool
Marginalized groups	Groups who are disadvantaged in education, employment, and life in general
Portability/Transportability of qualifications	The ability of the skills and qualifications gained locally to be accepted across provincial and national borders.
Skilled and flexible workforce	Workers who have the technical skills and personal skills to adapt themselves to the changing requirements of workplaces
Social inclusion	An umbrella term for ensuring that VET addresses the needs of all marginalized groups so that they can participate effectively in society
Technological innovation	New developments in technology used in different industries and, always supported by electronic hardware and software
Training reform	The process of introducing new training arrangements or modifying existing ones to meet national and provincial goals for economic competitiveness at home and abroad
Transition from school to work	The process of leaving school with the appropriate skills and qualifications to gain employment
Transparency	An attribute which reveals the true meanings of a process, practice, or other piece of information; often used for accountability purposes

Table 2 Terms reflecting explanations of environments and measures used by economists

Completion rates	The proportion of students who commence a program of training and stay on until they have successfully finished the training
Demand for training	The desire for training programs by students and workers
Human capital	The stock of knowledge and skills embodied in people. Includes education and training, health, intelligence, expertise, access to networks, and ability to work and learn
Labor market	An arena in which employers compete for labor and workers compete for jobs
Labor mobility	Ability of individual workers to move within and between labor markets at home and abroad
Long-term unemployment	The proportion of the working age population that has not been in employment for a specific amount of time (generally 12 months or over)
Marginalized groups	Groups who are economically and culturally disadvantaged because of low socioeconomic backgrounds, remoteness, ethnicity, and race
Open training market	A market in which public and private providers of training compete for students and funding
Returns on investment in training	Financial and nonfinancial benefits received by individual workers, employers, and governments who pay for or subsidize training
Student retention	Students remaining in school or programs of training until completion
Tight labor market	A labor market where jobs are plentiful
Training market	The arena in which training is bought and sold
Training supply	The training programs provided or made available by training providers
Unemployment rates	The proportion of the working-age population that is not in work and is looking for work
Unmet demand	Occurs when the number of students wanting to do a particular course outstrips the training places available
Weak labor market	A labor market where jobs are scarce
Youth labor market	An arena in which employers compete for young workers (generally between the ages of 15 and 19) and young workers compete for jobs

The teenage or youth labor market is a special concern for VET. In an open training market, all providers are able to compete for government resources and students; in a closed training market only certain providers have access to specific sorts of government funds. In a tight labor market, jobs are plentiful while in a weak labor market, jobs are scarce. In addition, the phrase *making smooth transitions to the labor market* is used regularly to describe what are considered to be key objectives for VET. The concept of labor mobility has also been used to explain reasons for developing more transparent and portable VET qualifications.

The Effects of Globalization and Technological Innovation have been a Special Focus for VET

Globalization, the term for the multiplicity of linkages and interconnections between the states and societies that make up the current world system (VOCED Thesaurus), has also moved into the vocabulary of VET. It is now widely used to explain increased labor mobility. It is also quoted as a key reason for advanced economies wishing to compete with low-wage countries to focus on what are called value-added goods and services. For many advanced economies, like Australia, this includes education and training, health services, insurance and superannuation, information technology, and telecommunications, and bio-medical and environmental expertise. For countries like Japan and South Korea, it includes manufacturing expertise and products in electronics and computer technology. Many countries, including Germany, Australia, and Britain, are also involved in exporting their technical and vocational education offshore and importing student clients to their onshore programs. Terms like international education,

transnational education, borderless education, and cross-border education have also been used to describe training arrangements for nondomestic students.

Accelerated take-up of technological innovation in the developed countries has also produced a range of concepts and terms which have generally dealt with training responses to help workers become and remain more competitive in the labor market (Table 3). New metaphors like digital divide and information rich and information poor have been used to explain divisions between those who have or do not have access to the infrastructure and telecommunications technology that is making domestic and international communication easier and faster.

VET is About Developing Technical, Technological, and Personal Skills

The language of VET has been infused by the concept of skills training. Although associated in great part with the language of psychology, the idea of skills training (also called skills formation, skills development, and skill acquisition) in VET incorporates training required for acquiring specialist technical skills for specific occupations and industries (content-specific skills), and generic and/or transferable skills which can be easily transported across workplace, community, and regional contexts. There is a major focus on practical skills appropriate and relevant to industry needs. As VET aims to develop skilled and flexible workers, the concepts and terminology related to multiskilling, reskilling, and upskilling have also gained currency. Bodies which provide governments and industry with information on training needs and market information have also adopted the term skills.

Table 3 Terms associated with globalization and its effects that have influenced the language of VET

Borderless education	Education where students, teachers, providers, programs, curricula, and course materials cross national borders
Cross-border education	
Transnational education	
Digital divide	Refers to the division between those who have and do not have equitable access to both the technology (i.e., the hardware and software) and the skills to use it
Information poor	Refers to individuals and countries (generally in the developing economies or among marginalized groups in advanced economies) that do not have equitable access to information technology and resources
Information rich	Refers to individuals and countries (generally advanced economies) who have abundant access to information technology and resources
International cooperation	Used to indicate the internationalization of the worlds of education and work
International education	Education which is provided to students from different countries either within country or at host country
International educational exchange	Exchange of students and programs between domestic and international institutions
International market	A market for international students
International program	A program provided for international students at home or abroad
International student	A student who studies in an institution which is not a domestic institution
Internationalization	The increasing geographical dispersion of economic activities across national borders
Offshore training	Training that is provided abroad by domestic institutions
Onshore training	Training for overseas clients that is provided by domestic institutions at home

Australia has industry skills councils, the United Kingdom has sector skills councils, and New Zealand has industry training organizations. Although in Germany such bodies are called chambers, they are also involved in providing governments with industry advice about training for occupational skills. Societies concerned with increasing the level of high-order skills among citizens are referred to as high-skill societies.

The language of the environmentalist and biologist has also infiltrated VET and we note a preoccupation with the development of sustainable skills and identification and maintenance of skills ecosystems. There is a concern for ensuring that we have appropriate and relevant skills for the twenty-first century and many countries have developed policy statements identifying the types of skills that VET must aim to deliver. The hope shared by countries internationally is for VET to develop skilled workers for a more internationally competitive economy and a more equitable society.

In recent years, there has also been a debate about whether traditional apprenticeship training should provide broad-based occupational skills or should just prepare tradespersons to perform a subset of such skills also called skill sets. Employee associations mainly want to preserve the marketability of individuals by ensuring that they have the full repertoire of skills belonging to an occupation, while the majority of employers may want to support a system of skill sets, which address their immediate and particular skill needs.

Key personal skills and attributes critical for obtaining, maintaining, and progressing through jobs are also a focus for VET. Although configurations of such skills may differ across cultures, they are about competencies that underpin all work tasks and can be transported between jobs

(Table 4). These are called key competencies, generic skills, and employability skills in Australia; core skills in England and Germany; essential skills in New Zealand; workplace know-how in the United States; and social skills in Sweden.

Systems, Frameworks, Pathways, and Cultures Are Useful Metaphors for VET Structures and Strategies

Concepts and terminology used in organizational theory, engineering, building, and anthropology have also appeared in the language of VET, providing useful metaphors for the organization, delivery, and recognition of training.

Systems Are Useful Metaphors for Describing VET Arrangements

There has been a dependence on the term system (often used by organizational theorists) as an umbrella term which brings together a number of related components. What is referred to as the VET system in Australia is called the FE system in the United Kingdom. The training of apprentices in Germany is called the dual system, in the United Kingdom and New Zealand it goes by the name modern apprenticeship system, while in Canada it is called the apprenticeship system. In Australia, training reforms have brought about a number of name changes (often with the change of governments) to what is generally known as the apprenticeship and traineeship system. Apprentices and trainees undertake what are called contracts of training (formerly indentures) in Australia, training contracts in Germany, and apprenticeship contracts in Canada.

Table 4 The types of skills and economic and social environments associated with VET

Broad-based occupational skills	Technical skills and personal attributes associated with all aspects of an occupation.
Generic (basic, core, essential or employability) skills	Skills and personal attributes that underpin all task focused skills and are portable from job to job (often includes communication, self organization, literacy and numeracy, working with teams, problem-solving, initiative independence, resilience, job flexibility, willingness to learn)
High skills economy High skills society	Refers to economic and social environments in which 'high levels of workforce skills ... are combined with ... high wages and high labour productivity' (Brown <i>et al.</i> , 2001)
Multiskilling	This is applied to ensure that all workers are able to develop and use a variety of skills to get work done. In Australia it was originally applied to breaking down traditional demarcations between what different occupations were allowed to do according to industrial awards.
Reskilling/upskilling	Although available for all workers these are particularly focused on existing and older workers who need to gain more or better skills to maintain their employability
Skill sets	Skill sets provide a clearly defined statement of the skills and knowledge required by an individual to meet either industry needs or a licensing or regulatory requirement.
Skill ecosystems	A skill ecosystem is a concentration of workforce skills and knowledge in an industry or region (for example, cleaning industry, family support services, wine district) (Finegold, 1999)
Specialist skills	Technical skills that are identified with a specific occupation or parts of an occupation
Sustainable skills	Skills which enable individuals to compete effectively in labor markets which are constantly changing
Transferable skills	Skills that transfer from one workplace to another, from one occupation to another, from one country or region to another

Arrangements for the recognition of qualifications are also explained in terms of systems, that is, qualification-recognition systems. For example, the mutual recognition of qualifications refers to the process of recognizing the qualifications awarded by a similarly accredited training organization. This can apply to qualifications awarded by domestic or international organizations. In Australia, providers registered to provide national qualifications (registered training organizations (RTOs)) must recognize qualifications and statements of attainment (detailing any subjects or credits gained in a course) awarded by any other RTO. Australia has the mutual-recognition agreement, between the Commonwealth government and the governments of the states and territories, intended to establish a scheme for the implementation of mutual-recognition principles for goods and occupations for the purpose of promoting the goal of freedom of movement of goods and service providers in a national market. Australia and New Zealand have signed the Trans-Tasman Mutual Recognition Agreement which states that a person registered to practise an occupation in one country is entitled to practise an equivalent occupation in the other country without the need to undergo further testing or examination. In Europe, the European Credit System for Vocational Education and Training (ECVET) describes arrangements for the recognition of cross-border vocational qualifications.

Frameworks Are Useful Conceptual Structures Which Explain Key Arrangements in VET

The term framework to describe fundamental structures has also crept into the language of VET. Generally used in the fields of architecture and building and construction to refer to an external scaffolding, it is used in VET to explain how a nation, province, or state will arrange and deliver training and assessment services, award qualifications, and ensure the quality of VET.

Australia has the National Training Framework (NTF) which depicts the national approach to VET. This broad umbrella framework comprises the Australian Qualifications Framework (AQF) and the Australian Quality Training Framework (AQTF). The first describes 15 nationally recognized qualifications awarded in secondary education, VET, or university sectors. The second comprises two sets of quality-assurance standards: one is for the acquisition and maintenance of RTO status, the other is for agencies which have the responsibility of issuing such qualifications. New Zealand and South Africa have also established a national qualifications framework (NQF, and South Africa Qualifications Framework (SAQF) respectively). Scotland has developed a credit accumulation and transfer framework. The European Commission is in the process of developing the European Qualifications Framework (EQF). Besides, in Europe, the 'Tuning of educational structures in Europe' project is

intended to aid the development of the European Credit Transfer System (ECTS) into an overarching pan-European credit accumulation and transfer framework. Also being developed in Europe is the Common Quality Assurance Framework for VET in Europe.

Pathways Help to Describe Routes to Work and Further Training

Recently, the main objective of VET systems in most countries is to open up access to training to enable easier movement between different qualifications or levels of courses and to ensure that access is not blocked as a result of earlier decisions or constraints. The concept of pathways has been adopted in VET as a useful metaphor for describing routes to qualifications and employment.

Associated with the concepts of merging or diverging routes to qualifications (often called flexible pathways) are mechanisms to allow this movement to occur, including credit transfer (the transferring of results from one course or institution to another) and formal recognition of learning, experience, and skill that an individual has acquired either at work, in community activities, or in life. Australia and New Zealand use the term recognition of prior learning (RPL); the accreditation of prior experiential learning (APEL) is favored in the United Kingdom. A variant of these processes dealing in assessments of the currency of a competency or skill is called recognition of current competency (RCC) in Australia.

Learning Cultures Are Environments Committed to Learning

The term culture (as in learning culture) has been applied to describe a holistic approach to learning in VET and in other sectors. It is based on concepts of acculturation favored by anthropologists in describing how individuals acquire understanding of the norms and values of the societies in which they are born. When it is applied to discussions of VET, it highlights the importance of organizations displaying a commitment to learning by providing workers with incentives and resources to continuously upgrade their skill and knowledge (**Table 5**). Terms like learning organization, learning town, city, or community are also used in the language of VET to describe environments where all component parts are committed to learning.

The concept is especially pertinent to the education of Indigenous Australians in VET. The Australian strategy 'Partners in a Learning Culture: Australia's National Aboriginal and Torres Strait Islander Strategy for Vocational Education and Training 2000–2005' aimed at identifying strategies to use a learning culture approach to ensure the provision of culturally appropriate VET to

Table 5 Terminology associated with systems, frameworks, and learning cultures, organizations, and cities

Accreditation and registration systems	Arrangements for determining the eligibility of providers to provide training leading to nationally recognized qualifications
Apprenticeship system	Arrangements for the employment and training of apprentices
Competency-based training system	Arrangements for the delivery of training based on industry-competency standards
Dual system	Arrangements for the training of apprentices in Germany and other European countries using similar approaches
Further education system	Arrangements for the provision of postschool VET in the United Kingdom
Learning cities, towns, regions and communities	Cities, towns, regions, and communities which promote the development of economic, cultural, and social networks and activities for developing the skills, knowledge, and potential of individuals and enterprises
Learning culture	An environment of shared beliefs, values, and principles which support learning
Learning organization	Organizations committed to learning new and better ways to achieve corporate strategic objectives, improve organizational practices, and develop the talents and potential of all staff
Mutual recognition systems	Arrangements for reciprocal recognition of qualifications between countries and VET institutions
National training frameworks	Conceptual structures describing key elements of national approaches to VET
Qualification frameworks	Conceptual structures for describing national qualifications
Quality assurance systems	Arrangements for assuring the quality of VET training, assessment, and certification
Recognition of current competency	Formal assessments of the currency of skill and experience gained through prior learning or experience
Recognition of prior learning (RPL), Accreditation of prior experiential learning (APEL)	Formal assessments of learning, skill, and experience gained in work, in the community, or life in general

improve education, training, and employment outcomes for Indigenous Australians.

In the United Kingdom, there has been a concerted attempt to develop a culture of lifelong learning for all. Here, the minister for lifelong learning recently declared that what was required was the creation of a special culture to underscore the need for people to make their own decisions about what, when, where, and how they learn. That is, “a culture where everyone knows the value of learning and expects to learn throughout life, from pre-school years to post-retirement” (BBC news report).

The Language of Teaching, Learning, and Assessment Continues to Be Integral to the Way We Understand VET

The language of teaching, learning, and assessment is necessarily woven into the way we talk about VET. Concepts of formal, nonformal, and informal learning make it easy for us to understand how different arrangements, structures, and environments contribute to the development of skills and knowledge that are required in employment and life.

VET also aims to increase learner independence and self-management in preparation for the modern workplace. In doing so, it has implemented flexible models of training which give learners the responsibility of determining how, when, what, and where they learn. Terminology like flexible delivery, self-paced learning, or individual learning are used to describe these models of

training. More recently, terms like personalized learning have been introduced to describe training arrangements which tailor training objectives to take into account the personal histories and aspirations of individual learners.

All have required a change in the way VET practitioners deliver training. Increasingly, they trade the role of conveyors of information for the role of facilitators of learning. As facilitators, they are required to advise, motivate, and assist students who are working independently or in groups through learning programs, which may be delivered in classrooms, in the workplace, online, or a combination of these.

When institution-based staff or external consultants go into workplaces to assess the work of students or workers, or conduct RPL assessments, they are referred to as workplace assessors. When suitably qualified or experienced, workplace tradespersons and other employees are given the responsibility of the training and assessment of apprentices and trainees, and they are generally referred to as workplace supervisors. Those who provide in-house training that is not of a technical nature may be called personnel trainers, staff-development officers, professional-development officers, or more recently human-resource consultants.

We also find concepts of coaching (borrowed from the language of sports) and mentoring (borrowed from the language of business) in the language of VET. These have been increasingly used to describe the provision of assistance and guidance to new or existing workers by colleagues, supervisors, or executives (including external consultants).

One of the major reforms of the mid-1980s and the early 1990s in Australia, New Zealand, and the United Kingdom was the introduction of the competency-based training (CBT) approach to training. Integral to this approach was the development of what were called industry competency standards to provide guidelines for the type of performance that was to be demonstrated in a practical way by students hoping to acquire a certain competency. RPL, APEL, and RCC assessments and challenge tests were used to avoid the need to duplicate already acquired skills and knowledge.

It was common to speak about criterion-based assessment, which made judgments of demonstrated performance of individuals against a defined standard. This had taken the place of normative assessment where such performance was judged against the performance of others. In such a system, students who were successful were assessed as competent, those who were not were deemed to be not competent or not-yet competent. As time went by, VET practitioners and employers began to be frustrated with drawbacks of the criterion-based assessments. The principal drawback of these was the inability of such an approach to provide sufficient information about varying

levels of performance. The term ‘graded competency’ then entered the language of VET. Key assessment issues continued to be reliability and validity of assessments in VET while in Australia registered providers (called RTOs) also ensured that their assessments were flexible and fair.

Thinking and speaking of students and employers as clients or customers has also become increasingly common as has measuring training needs and client satisfaction with training. This is where VET terminology has borrowed from the language of business and human resources. Providers identify what their clients want from the training system by undertaking the processes called needs analyses, which are used to compile individual training plans and measure the extent to which such needs are met via client-satisfaction surveys.

There is also terminology borrowed from the language favored by human-resource professionals, like on-the-job training to denote training which happens during the production process, and off-the-job training which refers to training away from the process of work (in company training rooms or at training institutions) (Table 6). These distinctions are especially pertinent in apprenticeship training which in many cases combines both these forms.

Table 6 Terminology associated with teaching, learning, and assessment that has crept into the language of VET

Challenge test	A test used by individuals to demonstrate that they have acquired a certain competency and so do not have to repeat the training
Competency-based assessment (criterion-based assessment)	Assessment in which candidates must demonstrate the ability to perform a certain skill to a pre-determined standard. Differs from normative-based assessment which compares the performance of individuals to others taking the assessment
Competency-based training	A form of training which aims to prepare students to acquire competencies identified by industry to prepare students for the workplace
Facilitator	A VET teacher or trainer who provides learning support and assessment services for students engaged in self-paced learning
Flexible delivery	Training arrangements which are customized to the needs of individual students and employers
Formal learning	Learning which occurs in schools, VET institutions, and universities leading to recognized qualifications
Graded competencies	Assessments which recognize the performance of individuals according to the quality of the performance
Informal learning	Learning which occurs as a result of experience in the workplace, community, and in the family
Mentor	An experienced and knowledgeable individual who provides advice and assistance to those with less experience
Nonformal learning	Learning which occurs in structured training programs outside recognized educational institutions
Recognition of current competency	Formal assessments which take into account evidence of an individual's current ability to perform a given skill to a desired standard
Recognition of prior learning (RPL), Accreditation of prior learning, Accreditation of prior experiential learning (APEL)	Formal assessments which take into account evidence of learning gained through experience in the workplace, in the community, or at home
Self-paced learning	A model of learning where individuals decide how, when, where, and what they will learn and progress through learning at their own pace
VET practitioner	Individuals who are engaged in delivering training and assessment in VET institutions and enterprises
Workplace assessor	Individuals who have the responsibility of assessing the workplace competencies of students and existing workers
Workplace supervisor	Suitably qualified and experienced individuals with the responsibility of on-the-job training of apprentices and trainees

Conclusion

Identifying the language of VET is not always a straightforward task in view of the variety of ways that different countries (and sometimes local states and provinces) have adapted for describing their concepts and structures for VET and in view of the difficulty of ascribing a particular influence to a specific event or catalyst. However, it is clear that even within and sometimes across cultures, a broadly common language is emerging to describe the aims and objectives, policy initiatives, organizational infrastructure and training delivery, assessment mechanisms, and quality-assurance procedures. This has been shaped by the economic environment in which VET operates and the sharing, exporting, or importing of knowledge and expertise across cultures. It has also been influenced by the wish for VET qualifications (wherever they are acquired) to have currency within and across provincial and national borders.

Bibliography

- Brown, P., Green, A., Lauder, H. (2001). *High Skills*. Oxford University Press: Oxford.
- Drucker, P. F. (1992). *Managing for the Future*. Butterworth Heinemann: Oxford.
- Finegold, D. (1999). Creating self-sustaining, high-skill ecosystems. *Oxford Review of Economic Policy* 15(1).
- Porat, M. U. (1977). *The Information Economy*, 9 vols. Washington, DC: Department of Commerce/Office of Telecommunications.
- Adam, S. (2005). *Principles of a Pan-European Credit Accumulation Framework: Good Practice Guidelines*, p 1. Bilbao: Tuning Educational Structures in Europe Management Committee.
- ATED (2003). *Australian Thesaurus of Educational Descriptors*, 3rd edn. Melbourne: Australian Council for Educational Research.
- Australian National Training Authority (ANTA) (2000). *Partners in a Learning Culture: Australia's National Aboriginal and Torres Strait Islander Strategy for Vocational Education and Training 2000–2005*. <http://www.dest.gov.au/archive/iae/analysis/learning/1/learning.htm> (accessed October 2009).
- BBC News (2000). *BBC News Report*. 30 June 2000. <http://news.bbc.co.uk/1/hi/education/811723.stm> (accessed October 2009).
- Cedefop (2004). *Terminology of Vocational Training Policy: A Multilingual Glossary for an Enlarged Europe*. Luxembourg: Office for Official Publications of the European Communities.
- Common Quality Assurance Framework. http://ec.europa.eu/education/policies/2010/doc/vetquality_en.pdf
- European Commission (2002). Copenhagen Declaration. http://www.ec.europa.eu/education/pdf/doc125_en.pdf (accessed October 2009).
- European Commission (2009). *European Credit System for Vocational Education and Training (ECVET)*. http://ec.europa.eu/education/ecvt/index_en.html (accessed October 2009).
- ILO (2005). *ILO Thesaurus*, 2005. Geneva: International Labour Organisation.
- Knight, A. and Nestor, M. (2000). *A Glossary of Australian Vocational Education and Training Terms*. Leabrook: NCVER.
- Ministry of Economic Development (1997). Trans-Tasman Mutual Recognition Agreement. http://www.coag.gov.au/mutual_recognition/tt_mutual_recog_agreement.cfm
- Mutual Recognition Agreement. http://www.coag.gov.au/mutual_recognition/mutual_recog_agreement.cfm
- UIE (1999). *Glossary of Adult Learning in Europe*. UNESCO Institute for Education, Hamburg.
- VOCED (2006). *VOCED Thesaurus*, NCVER. <http://www.voced.edu.au/thesaurus/index.php> (accessed October 2009).

Research on Vocational Education and Training

J Robson, London Metropolitan University, London, UK

B Bailey, University of Greenwich, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Vocational education and training (VET) takes place in a huge variety of contexts and equivalence in the meaning of terms across national and sector boundaries cannot be assumed. In particular, there is debate about whether VET refers to formal programs and qualifications, and structures and systems, or whether the term might also encompass a wider range of formal and informal experiences that relate to work (Stasz and Wright, 2004). In many countries, VET takes place in a range of settings, including schools, colleges, workplaces, adult education, and training institutes and universities. This fragmentation, as well as the range of levels involved (from basic instruction to higher professional training) and its potential to embrace all occupations and trades, makes it more difficult to present a systematic account of research in the area than may be possible for primary, secondary, or higher education. Similarly by its very nature, VET is susceptible to a wider range of outside social and economic interests and to a wider range of research approaches and disciplinary perspectives. The research agenda may be more likely to be set in line with the needs of industry, for example, than is the case for schools research. In policy and practice, there is also confusion about the purposes of VET (Lovat, 2003; Stanton and Bailey, 2005).

This article focuses on research into the VET that takes place outside university systems. Since there is also debate about what constitutes research in this context (see Stanton, 2000), some clarification is needed here, too.

Although the two may be linked (and some would argue should be linked more often than they are), a distinction needs to be made between research enquiry for development (with its emphasis on practical outcomes and knowledge, e.g., relating to new curricula or changes to specific practices in local settings) and pure research (with its emphasis on new theoretical understandings, on methodological rigor and reflexivity, and the capacity for generalization to or the drawing of analogies with other contexts). Much development work in VET now takes place in response to policy or funding initiatives (e.g., those requiring the development of new programs, or the audit and routine evaluation of provision, or the close monitoring of institutional performance) and there is no doubt that such data have value and are frequently underexploited. However, the article mainly focuses on investigations and

activities that primarily aim at the enhancement of theoretical understandings and knowledge which, to a greater or lesser extent, may have general relevance for learners, teachers, trainers, policymakers, funders, or managers of VET in a range of contexts. We are principally concerned with VET research published in English since 1994, and relating to provision in the UK, Canada, the USA, Australia, New Zealand, and Europe.

An Emerging Field

It has become a commonplace in recent years to observe that in most countries research in VET is limited and sparse by comparison with research in other educational sectors. In part, this can be attributed to the fragmentation of VET provision (mentioned above), to its complexity, to a lack of understanding about the nature of the research that may be relevant, as well as to the difficulty of identifying sector boundaries (Stanton, 2000). A tendency in Anglo-Saxon culture to favor academic over vocational interests (DES, 1991; Raggatt and Williams, 1999), and an associated preoccupation with the royal route to higher education that has dominated debate about 14–19 education and training at least in the UK (Brown *et al.*, 2004) may also help to explain the relative paucity of research in vocational areas.

The lack of specialist networks and journals is also a contributing factor.

However, in recent decades, governments of Western capitalist democracies have placed greater political emphasis on the importance of the development of VET provision and so have actively sought to promote research into aspects of VET. This has led to some limited allocation of resources for research. A concern with competitiveness, productivity, and economic performance in Australia and Britain, for example, has prompted the development of policies based on raising participation in VET, on standards and quality, the expansion of user choice, and the encouragement of a significant input from industry (Bailey, 2003).

Calls for such policies to be based on evidence and for the strengthening of links between research communities and government departments have been accompanied as often as not by an insistence that what works is what counts. The nature of the desired evidence has been in contention (with a clear preference in some official circles for quantitative data, longitudinal studies, and large-scale

surveys). In the UK, a report by Rendall (2003) was commissioned in response to “widespread concern about the state of quantitative studies in post-16 education and training.” The document identified a number of issues, including difficulties in recruiting researchers with quantitative data analysis skills. Bates *et al.* (1997) argued for research that focused on democratic and social justice rather than on economic imperatives; even after a decade, it still appears significantly easier to obtain funding for the evaluation and review of official VET policies and initiatives relating directly to national economic interests rather than to broader and more critical social questions.

There is a lack of capacity for VET research in many institutions with a research focus, and within relevant government departments, fewer staff are likely to have knowledge and experience of vocational programs and to be familiar with the complexities and issues requiring investigation. However, a number of private contractors are emerging, keen to pick up funded consultations and projects. Many have fewer institutional constraints (than those, e.g., in university departments may have), but they may also be more isolated from the VET community, from its various teaching, training, and learning contexts. Even within universities, feelings of isolation among VET researchers may be prevalent since most colleagues are likely to be focused on the work of schools. In universities, increasingly, some staff are employed as career researchers, with research-only roles, and they tend to lack background expertise in VET and familiarity with the post-compulsory sector generally (Stanton, 2000). Those who do have VET backgrounds may lack formal research training (Smith, 2004). Concern with research capacity in the UK led to the establishment of a dedicated network for capacity building in 2002, as part of the Economic and Social Science Research Council (ESRC)-funded Teaching and Learning Research Programme (TLRP, 2002). Recently, and also in the UK, funding has been available from the Learning and Skills Council (LSC) for the secondment of teachers in further education to research partnerships and collaborations with university staff or external consultants. Such initiatives are increasingly seen by some managers as providing important opportunities for the professional development of teachers (Robson, 2006).

Funding and Organization

Funding for VET research comes from a variety of sources. These include international organizations such as the World Bank, the United Nations Educational, Scientific and Cultural Organization (UNESCO), the Organization for Economic Cooperation and Development (OECD), the European Commission, and large corporations such as RAND. National governments also fund VET research,

through their various departments and agencies (e.g., the Learning and Skills Research Council in England) and funding is also available through research councils, charitable and independent trusts, and publicly funded educational organizations such as universities, colleges, and some awarding and inspection bodies.

Huge variations exist, of course, in the research priorities of these organizations, the amount of funding they have available at any point in time, the degree of competitiveness that may be involved in accessing it, and the constraints associated with it. Most of these funders are more focused on other sectors of education and likely to fund VET research less often. As noted above, it is likely to be significantly easier to obtain official funding for the evaluation of specific policy initiatives and such funding is also likely to bring constraints with regard to time allocation and the dissemination of findings, for example (Bates *et al.*, 1997). Some research carried out to official specifications never reaches a wider audience, and funders may be particularly inclined to constrain dissemination if the findings are discomforting or unwelcome. Some sources of funding (such as research council funding) may carry greater academic prestige and wide dissemination is usually a requirement of such awards.

The official focus on VET that is increasingly evident in many countries has resulted in the establishment of some national centers for the promotion of VET research activity, such as the National Centre for Vocational Education Research (NCVER) in Australia (a nonprofit-making company owned by government) and in the USA, the National Research Centres for Career and Technical Education, funded by the US Department of Education. Funding from the European Commission has led to the establishment in all European Union (EU) and European Economic Area (EEA) member states of National Resource Centres for Vocational Guidance (NRCVG), but these are focused predominantly on promoting occupational mobility throughout Europe. They undertake activities or enquiries aimed at the development of training modules or the identification of innovative practices, for instance. In Europe, a number of universities host centers for vocational education research (such as the Research Centre for Vocational Education located at the University of Tampere in Finland, and the *Institut d'Economie Scientifique et de Gestion* (IESG) in France) and their funding comes both from their host universities and from a range of external sources.

Next, we elaborate some selected and indicative examples of VET research undertaken in a range of countries.

The ESRC-Funded TLRP, UK

The TLRP in the UK was set up in response to public criticism of the quality of much educational research

(Pollard, 2005). The ESRC manages the program on behalf of its funders (the largest of which is the Higher Education Funding Council). The program is the largest research program currently being managed by the ESRC and represents a significant investment in educational research. The first projects began in 2000 and publications will continue into 2010.

Over 60 research projects cover all phases of education, from early years to compulsory schooling, further education, higher education, and lifelong learning. Several research projects focus on the teaching and learning that take place following (or, in some cases, alongside) compulsory schooling in vocational and work-related contexts. The project titled Transforming Learning Cultures in Further Education took the view that learning and teaching could not be decontextualized from broader social, economic, historical, and political forces and that study of the complexity of relationships between teachers, teaching, learners, learning, and, the wider contexts of learning, was key to understanding what might be most useful for the purposes of informing policy and practice (Hodkinson and James, 2003). When researchers examined a range of VET courses within 14–19 provision, on a variety of sites, they found that the most effective ones were likely to have close and synergistic links with employers. However, this could make it difficult to challenge existing workplace practices and values, such as gender stereotyping, and tensions between learning sites sometimes arose. VET courses without such strong links with employers were less likely to provide effective progression routes into employment. Where students could commit themselves to a specific vocational area, the courses appeared to work well, but for students who were or became unsure of their future direction, VET provision was much less well suited. Researchers recommended, among other things, that VET provision should recognize and support career uncertainty by adopting more flexible structures (Hodkinson and Colley, 2006).

The TLRP project titled Literacies for Learning in Further Education in England and Scotland aimed to find out how reading and writing enhanced learning in vocational and academic subjects in further education (Ivanic, 2006). Researchers studied literacy involved in learning across 11 further-education curriculum areas and found a vast range of texts in use, including booklets, graphical user interfaces, websites, letters, handouts, overhead presentations, textbooks, logbooks, file notes, labels, maps, diagrams, board writing, measurements, and lists. Some of the literacies involved in demonstrating learning (e.g., in hospitality and catering) bore little relation to what was being learned or to the futures for which courses were preparing their students. Further, the assessment of literacy capabilities tended to be decontextualized or at best simulated. Project researchers recommended, among other things, that the demands made by the specialist texts and

literacy practices of each curriculum area should be recognized and that specialist teachers of communication work alongside subject specialists.

The Community College Research Centre at the University of Columbia, USA

The Community College Research Centre (CCRC) was established by the Alfred P. Sloan Foundation in 1996 and is located in Teachers' College, Columbia University, New York. Its work is funded through contributions from the Sloan and other foundations, and federal agencies such as the Office of Vocational and Adult Education (OVAE). CCRC has a research mission to establish a basis of research data and findings which contributes to the development of practice and policy and to students' improved achievements. It undertakes national surveys and studies with particular regard to equity in access and participation in postsecondary education and claims status as the leading independent authority on the community colleges in USA.

The CCRC's research issues include: distance education, student retention and progression patterns, minority and low-income students' access and success, and workforce education.

A recent national field study involved detailed case studies of 15 colleges including interviews with students and staff, which provided the findings for the publication, *Defending the Community College Agenda* (Bailey and Morest, 2006). This report investigates the challenges faced by the 2-year colleges in their mission to provide access and success for their students who constitute nearly 50% of the college students but a disproportionately high number of those from poorer, ethnic minority and academically underprepared students. Chief among the challenges is the loss of state funding which has led to increases in tuition fees, and to difficulties for colleges as they try to maintain the teaching and support services needed by the growing and diverse student body. The changing context requires the colleges to address the barriers to learning faced by students after their admission to college and to improve their on-course success and progression into employment.

The NCVER in Australia

The Technical and Further Education National Centre for Research and Development was established in 1981. The name was changed to NCVER in 1992. NCVER is the national body which undertakes research in-house and manages a continuous program of VET research. This strategy takes the form of priorities and aspects for investigation for which expressions of interest are invited

as part of a competitive grants program. Projects which contribute to the research strategy are commissioned to be undertaken by researchers in universities, by staff in VET providers, and in private organizations. Current priorities for investigation include: basic skills, access, equity, participation, e-learning, and higher education/vocational linkages. Rather than a small number of large projects, NCVER funds a range of smaller projects on aspects of, for example, students and individuals (student satisfaction and destinations), the VET system (quality framework criteria), industry, and employers (apprenticeships and traineeships). One theme of research in recent years has been the changing role and identity of the teacher professional in VET. One such project examined the effects of changes in the VET system in Australia (e.g., the introduction of a competitive market for VET provision) on practitioners at a variety of VET sites across the country (Chappell and Johnston, 2003). Through interviews it was established that in the new commercial environment different roles were required of teachers and other staff. Although the practitioners interviewed maintained their educational identity, they were faced with business discourses which tested their understanding of VET as an educational activity and public good. This led them to place great importance on the teacher–learner relationship and on the educational values and norms of conduct on which that relationship is grounded.

NCVER also collects and makes available the findings of VET research from around the world on the (UNESCO-funded) Vocational Education and Training Research Database (VOCED) and is the center for the collection and analysis of Australian national statistics on VET.

Federal Institute for Vocational Education and Training, Germany

The Federal Institute for Vocational Education and Training (*Bundesinstitut für Berufsbildung* BIBB), under the Federal Ministry of Education and Research, is a national center for research on VET and its progressive development. It was established in 1970 and its research, development, and advisory work are aimed at identifying future trends, promoting innovation, and developing practice-oriented solutions in VET. Current research work, reflecting the circumstances of postunification Germany and EU policies, includes projects on the training market and the world of employment, VET for particular groups including, and the internationality of VET. The wider European agenda includes moves toward a European VET area through agreement on a credit transfer and qualifications which develop a European Qualifications Framework and BiBB is currently involved in a cross-national project on this. Another collaborative task has been that with the social research center in Halle on VET in the eastern states ('Between

market forces and subsidization: The effectiveness and future of VET structures in Eastern Germany'). This research investigation showed that only 4% of trainees in the western states of Germany learn their occupation through an external training provider rather than at a company which provides in-house training alongside practical work experience. This compared with the 30% of trainees in the eastern states who are trained in government-funded schemes organized by external providers of education and training in which the trainees have work placements in local firms. The study examines these innovative collaborative arrangements which develop and add to Germany's dual system of VET, and considers the implications of demographic trends in the east which will lead to a sharp reduction in the demand for training places in that part of the country.

In order to make VET research more readily accessible to a wider circle of users, the BiBB has developed a knowledge map which takes the form of a thematic catalog of issues, activities, and results of German VET research. Since the beginning of the 1990s, the BiBB has operated the Literature Database for VET (LDBB) which currently comprises references to VET and VET research since 1988.

Dissemination and Impact

It is not easy to estimate the impact of research on VET policies and practice. Clearly, the processes of any influence on the making and implementation of policies with regard to VET are not linear and immediate, since events can be expected to be ahead of the agenda and the output of the research community. Despite the absence of a one-to-one relation between research findings and practice, the interest of governments and their agencies in funding and making use of research studies of aspects of VET, as is shown in the exemplars outlined above, has increased in recent decades. This reflects the greater official emphasis on VET in many countries as a contributor to the development or the modernization of national economies in the global economic competition. It may also, of course, reflect governmental interests in shifting traditional attitudes and practices in the workplace and workforce.

Whatever the drivers behind it, the increase in VET research activity in many countries in recent years, mainly on the part of governments and their agencies, has been striking and has led to important changes in the nature of research in VET and in other sectors of education systems. The first of these is the centralization of the setting of the research agenda and of control over the ownership of research findings. As has been shown, different structures and centers for research have been established in different countries, and standards of research practice appear to have improved as a result of greater transparency in the setting up of projects and the publication of

results. As already noted, however, some funders insert copyright clauses into their contracts with researchers effectively to constrain the publication of research findings, particularly when they are seen to run counter to a policy decision, choice, or preference.

Much VET research is susceptible to the criticisms leveled against educational research in general. That is, it is not cumulative in building on earlier work, it is sometimes opaque and inaccessible to the nonspecialist reader, and has little connection with the practice of education and training. One way of meeting the first of these objections is the availability of databases which provide comprehensive and up-to-date information about research publications in report, article, and book forms. VOCED in the NCVER in Adelaide is perhaps the best known of these; others include the Current Educational Research in the UK (CERUK) and the European Centre for the Development of Vocational Training (CEDEFOP) and those maintained by the national centers, described above.

In addition, in many countries, the dissemination of research is being improved through the planned provision of opportunities for researchers to discuss their work with practitioners and other users of research. While there will be debate as to what or who a user or user group is, the exchanges between the researchers and the other stakeholders in formal contexts or in informal settings, such as conferences and seminars, can lead to better understanding on both sides of the purposes and possibilities of research. The Australian Vocational Educational Research Association (AVETRA) is an example of this, while the Nuffield 14–9 Review in England (2003 to date), as part of its way of working, has invited participants from all sectors to discuss the agenda of the review and the contributions of researchers and others. There is an increasing trend to write into research briefs, as a requirement, the place and the form of dissemination as part of the researchers' duties (as is the case with projects funded by NCVER in Australia and the ESRC-TLRP in the UK).

Conclusion

Since 1994, as indicated above, VET research has undergone a number of changes. Alongside the greater volume of published research, there is the wider range of methodological approaches. Traditionally, VET research was predominantly quantitative in its approach, appropriate perhaps when interest focused on often large-scale analyses of age participation rates, school-to-employment transitions, student numbers recruited, qualifications gained, etc. Studies referred to above include studies of learners' and teachers' experiences in VET which inevitably use qualitative methods like action research and semi-structured interviews. Further, the range of issues investigated is wider and it is probably accurate to say that the quality of much research

is better than was the case earlier. These developments show that there have been real advances in VET research in recent years.

Bibliography

- Bailey, B. (2003). Research strategies for vocational education and training in England and Australia. *Research in Post Compulsory Education* 8(3), 379–392.
- Bailey, T. R. and Morest, V. S. (2006). *Defending the Community College Agenda*. Baltimore, MD: John Hopkins University Press.
- Bates, I., Hodgkinson, P., Unwin, L., and Young, M. (1997). Towards a new research agenda for post-compulsory education and training. *Research in Post-Compulsory Education* 2(3), 313–317.
- Brown, N., Comey, M., and Stanton, G. (2004). *Breaking out of the Silos: 14-30 Education and Skills Policy*. Nigel Brown Associates. NBrown2349@aol.com (accessed August 2009).
- Chappell, C. and Johnston, R. (2003). *Changing Work: Changing Roles for Vocational Education and Training Teachers and Trainers*. Adelaide: NCVER.
- DES (1991). *Education and Training for the 21st Century*. London: HMSO.
- Hodkinson, P. and Colley, H. (2006). Transforming learning cultures in FE. In: *Teaching and Learning Research Programme, 14–19. Education and Training: A Commentary by the Teaching and Learning Research Programme*. www.tlrp.org (accessed August 2009).
- Hodkinson, P. and James, D. (2003). Introduction: Transforming learning cultures in further education. *Journal of Vocational Education and Training* 55(4), 389–406.
- Ivanic, R. (2006). Literacies for learning in further education in England and Scotland. In: *Teaching and Learning Research Programme, 14–19. Education and Training: A Commentary by the Teaching and Learning Research Programme*. www.tlrp.org (accessed August 2009).
- Lovat, T. (2003). The relationships between research and decision-making in education: An empirical investigation. *Australian Educational Researcher* 30(2), 43–55.
- Pollard, A. (2005). Taking the initiative? TLRP and educational research, Educational Review Guest Lecture. www.tlrp.org/pub/acadpub/APollard_Ed_Review_Lecture.pdf (accessed August 2009).
- Raggatt, P. and Williams, S. (1999). *Government, Markets and Vocational Qualifications: An Anatomy of Policy*. London: Falmer.
- Rendall, M. (2003). *Quantitative Research: A Scoping Study for the Learning and Skills Research Centre*. London: LSRC.
- Robson, J. (2006). *Teacher Professionalism in Further and Higher Education: Challenges to Culture and Practice*. London: Routledge.
- Smith, A. (2004). Never mind the width, feel the quality: Improving VET research in Australia. *Research in Post-compulsory Education* 9(2), 203–221.
- Stanton, G. (2000). Research. In Smithers, A. and Robinson, P. (eds.) *Further Education Reformed*, pp 167–182. London: Falmer.
- Stanton, G. and Bailey, B. (2005). In Search of VET. SKOPE Research Paper No. 62. <http://www.skope.ox.ac.uk> (accessed August 2009).
- Stasz, C. and Wright, S. (2004). Emerging policy for vocational learning in England, *LSRC Research Report*. London: LSRC.
- Teaching and Learning Research Programme (TLRP) (2002). Available at: www.tlrp.org (accessed August 2009).

Relevant Websites

- <http://www.cccr.tc.columbia.edu> – Community College Research Center, New York, USA.
- www.tlrp.org – ESRC Teaching and Learning Research Programme, UK.

<http://www.cedefop.europa.eu> – European Centre for the Development of Vocational Education and Training.

<http://ieseg.fr> – Institut d'Economie Scientifique et de Gestion, Catholic University of Lille, France.

www.kibb.de – *Kommunikations- und Informationssystem Berufliche Bildung*.

<http://www.nccte.org> – National Research Centre University of Minnesota, USA.

<http://nuffield14-19review.org.uk> – Nuffield Foundation Review of 14-19 education and training, Oxford University and Institute of Education, UK.

<http://www.skope.ox.ac.uk> – Skills, Knowledge, Organisational Performance at Oxford and Cardiff Universities, UK.

<http://www.voced.edu.au> – VOCED database, Australia.

Vocational Education and Training Information Resources

L-A Harris, R Naidu, and M Saunders, National Centre for Vocational Education Research, Station Arcade, SA, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

There are numerous sources of information available to researchers, both in electronic and online formats as well as the traditional, print-based media. This article describes information resources related to research in the field of vocational education and training (VET) and offers the reader a road map for finding information about VET. Information sources covered comprise databases, directories/search engines/subject portals, citation indices, books and library catalogs, journals, newspapers, organizations/government departments/associations/networks, conferences, statistics, glossaries, dictionaries and thesauri, research in progress, and social software tools.

Throughout the course of identifying relevant English-language information sources and tools to include in this article, it quickly became evident that VET-specific resources are small in number and not always easily identified. Those resources that are not VET specific in coverage generally have the topic of VET listed within the social sciences and humanities discipline. These resources have been included in this article as they provide access to valuable VET-related research information that could otherwise be overlooked.

Databases

Databases, which can be both bibliographic and full text, contain information in electronic format. Bibliographic databases include details about authors, title, date, and, often, an abstract describing the document. Full-text databases, as the name suggests, provide the full document. Databases may index specific types of information, such as books, journals, conferences, statistics, or a combination of resources. They may be general in coverage such as library catalogs that index a range of resources held in a library's collection or they may be subject specific.

The two comprehensive VET subject-specific databases accessible online are VOCED, the VET research database, produced by Australia's National Centre for Vocational Education Research (NCVER) and VET-Bib, database of VET literature in Europe, the European Centre for the Development of Vocational Training (CEDEFOP). A smaller VET-focused database detailing information about European national research, research papers, and researchers is the European Research

Overview (ERO). For further information about these and other key online bibliographic databases covering the subject of VET, see **Table 1**.

Online databases may be accessed freely on the Internet or by subscription to commercial services, where bibliographic information produced by publishers about materials and/or the full text of articles, reports, etc. is provided. Examples of these subscription services include EBSCO, ProQuest, and Dialog. The most common way to access and search the databases made available by subscription is via a library (university, college, public, or private), although the library's selection of database services may be largely dependent upon the institution's focus. Even though access to full-text information may be restricted to subscribers only, the ability to search the content of the databases for free makes them useful resources in identifying material on a particular topic.

Directories/Search Engines/Subject Portals

Directories and search engines are in essence large databases containing links to information across a broad range of topics, while subject portals provide a pathway or guide to information and related services that can be tailored to the needs of users, usually on specific topics. Search engines are useful starting points to provide a broad coverage of information on a particular research topic. Caution should be used with search results as each search engine produces its listing of results using different ranking methods and, more often than not, the more academic or scholarly results are listed deeper in the search results. Directories, on the other hand, provide a narrower focus on the research topic because their structure is organized by subject. As subject portals provide access to information under one topic, for example, education or government, any information identified will generally be pertinent to the research topic.

Directories

Directories are topical lists of resources, arranged hierarchically, usually browsed but can be searched as well. These directories differ from search engines in one major way – the human element involved in collecting and updating the information. A directory is a useful resource

Table 1 Key online bibliographic databases

<i>Name</i>	<i>Description</i>	<i>Subject coverage</i>	<i>Availability</i>
VOCED	Research database comprehensive for Australia, with key international English-language resources	Vocational education	Free to search www.voced.edu.au
VET-Bib	Research database of vocational education and training (VET) literature in Europe	Vocational education	Free to search www.cedefop.europa.eu
ERO base	Information on European experts/researchers, projects (European Union and national) and papers	Vocational education. Human resource development	Free to search www.trainingvillage.gr/etv/Projects_Networks/ERO
edna	Online resources for education, training, and research Australian with links to international resources	Education broadly and vocational education component	Free to search www.edna.edu.au
Educational Resources Information Center (ERIC)	Predominately North American and increasingly from other countries	Education broadly and vocational education component	Free web search Fee-based via commercial vendors www.eric.ed.gov
Australian Education Index (AEI)	Australian education research and information	Education broadly and vocational education component	Fee based www.acer.edu.au
British Education Index (BEI)	UK literature which supports educational research, policy, and practice	Education broadly and vocational education component	Fee based Free to search latest additions www.leeds.ac.uk
Education Research Complete	Worldwide, predominately full-text education journals.	Education broadly and vocational education component	Fee based www.ebsco.com
ProQuest Career and Technical Information	Worldwide, predominately full text	Vocational, technical, and career education	Fee based www.proquest.com

Table 2 Key directories

<i>Name</i>	<i>Description</i>	<i>Subject coverage</i>	<i>Availability</i>
Intute: social sciences	Social sciences is one of a collection of key education subject directories. Provides access to key resources in the subject area	Further education; adult education; vocational education; and educational policy	www.intute.ac.uk
Infomine	Organized collection of annotated and indexed links to university level research and educational tools. Social sciences and humanities constitute the subcategory listing VET-related information	Vocational education; vocational guidance; career development; and technical education	http://infomine.ucr.edu
Librarians' Internet Index	Directory of websites selected, evaluated, described, and organized by librarians. Society and social science is the subcategory listing VET-related information	Adult education; career education; education research; and research institutes	http://lii.org
Vocational Information Center: career and technical education resources	Comprehensive coverage of US resources with some international links	Career and technical education; vocational education; and adult education	www.khake.com

when wanting to identify specific information about a particular topic. Vocational education usually appears under the broad subject category of social sciences and humanities. Examples of directories that provide access to information on the topic of vocational education or to related topics, such as adult education and career education, are provided in **Table 2**.

Search Engines

No search engine exists that is dedicated to the topic of vocational education and no one search engine is comprehensive in its coverage. It is therefore important to consider using more than one when searching for information. To assist in identifying which search engine or

directory best suits the information topic, websites such as Beaucoup and Search Engines 2 list search engines, directories, and indices in subject categories (such as education under which vocational education is generally listed), or by scope to help narrow the choice. Access details of these are given in **Table 3**.

Subject Portals

A portal is a website containing an aggregation of subject-specific links leading to other websites. It can be a useful resource as it offers high-quality service targeted to the identified needs of its audience and presents a one-stop-shop approach to finding relevant information. While undertaking the scoping exercise to identify relevant VET-specific resources for this article, two VET-subject-specific portals were identified – CEDEFOP and WIFO the gateway to research on education in Europe. There are also numerous education subject portals that provide avenues to VET-related links and examples of these along with further details on the CEDEFOP and WIFO portals are provided in **Table 4**.

Citation Indices

A citation index shows where a piece of work has been cited, how often, and provides a research trail forward from the year a piece of work was first published.

Their main purpose is to provide information about who is citing whom, to find articles that are related to each other, and to indicate the impact of individual articles.

As VET generally falls under the subject category of social science, the relevant fee-based citation index is the Social Science Citation Index (SSCI). It provides access to current and retrospective (from 1970 onwards) bibliographic information, author abstracts, and cited references from over 1700 social science journals. Google Scholar's cited-by feature is a free alternative to fee-based citation indices which allows researchers to determine the frequency with which a specific article is cited by others, the relationships between articles, and authors citing articles on the same topic. While Google Scholar searches across many of the same databases as SSCI, it does not provide access to as many succinct results as the fee-based citation indices do (especially to full-text academic articles). It does, however, search across a broad range of open-access repositories and a range of other materials deemed as scholarly or academic. **Table 5** illustrates several key citation indices that can be used to identify VET-related information.

Books and Library Catalogs

Books (monographs) usually contain a broad overview or an in-depth exploration of a topic and range from popular to scholarly publications. Books on a broad topic may contain chapters or essays on a topic which are often not

Table 3 Search engine resources

<i>Name</i>	<i>Description</i>	<i>Availability</i>
Beaucoup	Provides links to 2000+ search engines, directories, and indices	www.beaucoup.com
Search Engines 2	Over 12 500 links to local, regional, national, foreign, and meta-Internet search engines and web directories	www.search-engines-2.com
Choose the best search engine for your information need	A question-and-answer cheat sheet to help match the information need to the most appropriate resource. Located under 21st Century Literacies – Information Literacy links	www.noodletools.com

Table 4 Key subject portals

<i>Name</i>	<i>Description</i>	<i>Availability</i>
CEDEFOP	Provides access to the European Union reference centre for vocational education and training	www.cedefop.europa.eu
edna.au	Online resource collection and collaborative network for the education and training community. Predominantly Australian but with international information	www.edna.edu.au
British Education Internet Resource Catalogue	Database of information about professionally evaluated and described Internet sites which support British educational research, policy, and practice (includes further/ vocational education)	http://brs.leeds.ac.uk/~beiwwww/beirc.htm
WIFO	Provides access to resources in the field of European education, with a focus on vocational education and human resource development.	www.b.shuttle.de/wifo

Table 5 Citation indices

<i>Name</i>	<i>Description</i>	<i>Availability</i>
Arts and Humanities Citation Index (A&HCI)	Provides access to cited references from arts and humanities journals and relevant items from science and social sciences journals	Fee based http://scientific.thomson.com
Google Scholar	Provides access to cited references to papers, theses, books, and articles from academic publishers, professional societies, preprint repositories, universities, and other scholarly organizations	Free to search http://scholar.google.com
Scopus	Provides access to cited references to peer-reviewed titles and includes open-access journals, conference proceedings, trade publications, and book series	Fee based www.info.scopus.com
Social Sciences Citation Index (SSCI)	Provides access to cited references found in scholarly social sciences journals covering a range of disciplines	Fee based http://scientific.thomson.com

indicated by the book's title. The bibliography in a book is a useful place to find additional sources, while the index at the back of a book provides a guide to the coverage of a specific topic.

Books and out-of-print material on VET can be bought from specialized bookstores and online bookstores. There are thousands of online general bookstores that sell new, used, and out-of-print books and textbooks. Some of the well-known online stores include AbeBooks.com; Amazon.com, and BarnesandNoble.com. There are also commercial publishers including Springer and Routledge (an imprint of the Francis and Taylor Group); university presses such as Oxford University Press; and organizations such as the Office for Official Publications of the European Communities, Organization for Economic Cooperation and Development (OECD), and the National Centre for Vocational Education Research that specifically deal with vocational education; or one can use a one-stop ecommerce search engine such as BookFinder.com that searches for over 125 million books for sale. Many publishers also offer regular e-mail alerts to keep one updated with the latest educational publications.

Books can be borrowed or ordered on inter-library loan from libraries. Most libraries use the Dewey Decimal Classification System (DDC) to organize library collections. The broad Dewey numbers to use when browsing the library shelves are 370 for education and 370.113 for vocational education. Some of the most comprehensive online library catalogs are either national in focus or combine several catalogs in the one search. The most common searches are by author, title, keyword, or subject heading. Library of Congress Subject Headings are used in the majority of library catalogs. The base subject heading for VET is vocational education. Some of the more comprehensive catalogs are shown in **Table 6**.

Book reviews can be found in traditional sources such as journals and newspapers and also from reader reviews posted to online bookstores and catalogs. Commercial resources include BooksinPrint.com, Book Review Index,

and H-Net Review, which review academic books in the social sciences.

In recent times, there has been a push to convert books into digital form to create a large library of e-books (electronic books) available from the Web. The DigitalBookIndex is an online meta-index for most major e-book sites, including the Gutenberg Project, the Library of Congress, the British Library, and hundreds of university libraries, university presses, and commercial publishers.

For an in-depth analysis of current international developments concerning VET, some useful books include the UNESCO-UNEVOC series *Technical and Vocational Education and Training: Issues, Concerns and Prospects*, the *International Handbook on Education for the World of Work* published by Springer, and *Vocational Education: International Perspectives and Development* published by Routledge.

Journals

Journals and periodicals are continuing resources, issued in a succession of discrete parts that contain current information. While to some extent the decision as to what constitutes a scholarly journal is subjective, they do have certain characteristics – sometimes refereed or peer-reviewed and have undergone a rigorous approval and editing process by experts in the field before acceptance for publication; appearance is serious in format and generally include an abstract, literature review, methodology, results, conclusion, and bibliography; there is little advertising; the authors are professionals and researchers in the discipline; and the language is technical and the purpose is to inform, report, and make research available to the rest of the community. The two reference books Ulrich's International Periodical Directory and Magazines for Libraries contain detailed information on journals published throughout the world and also designate refereed titles.

Table 6 Library catalogs

<i>Name</i>	<i>Description</i>	<i>Availability</i>
Library of Congress online catalog	Catalog of the collections of the Library of Congress. Contains records of approximately 14 million items	http://catalog.loc.gov
Libraries Australia	Database of holdings of over 850 major libraries around Australia. Contains more than 40 million items.	http://librariesaustralia.nla.gov.au Enhanced version available on subscription and at some libraries
British Library integrated catalogue	Catalog of the collections of the British Library. Contains more than 13 million volumes	www.bl.uk
WorldCat	A combined catalog of thousands of international libraries affiliated with OCLC (Online Computer Library Center). It includes more than 52 million records in 400 languages	www.worldcat.org Fee based product with additional features available through institutional subscription. at www.oclc.org/worldcat
Libweb	Connects directly to the online catalogs of libraries in 135 countries	http://lists.webjunction.org/libweb
Google Book Search	Identifies and catalogs books provided by publishers and library partners	www.google.com.au/books
Copac academic and national library catalogue	Provides access to the merged online catalogs of major university and national libraries in the UK and Ireland, including the British Library	http://copac.ac.uk
European library	Offers access to the combined resources of the 43 national libraries of Europe, including the British Library. You can search across libraries or make a specific selection	www.theeuropeanlibrary.org
Libdex	Index to the library catalogs of over 18,000 libraries world wide	www.libdex.com

Journal impact factors and article citation counts are methods used to evaluate an author's published research. Article citation counts can be determined by using citation indices, while journal impact is used to determine the expected influence and worth of individual papers within a core set of the highest-quality scholarly journals determined by ISI Thomson Scientific. ISI's Journal Citation Reports social sciences edition evaluates over 1700 of the world's leading social science journals and their impact and influence in the international research community.

The advent of the open-access movement, which brings free and unrestricted online access to articles in scholarly journals, has raised the issue of how to distinguish between web impact – readership or downloading – with actual citation impact. Recent studies on the impact of educational research indicates that there is little relationship between an author's most highly cited articles and articles with the most website hits (Martin, 2007), although publishing articles on the Web is having an effect on the immediacy of articles being cited (Shin, 2004).

The major journal publishers in vocational education are SAGE Publications, Routledge – Taylor & Francis Group, and Blackwell Publishing. To find articles, publishers' websites can be searched or the numerous educational databases and citation indices can assist in identifying and finding relevant journal articles. Many of these sources provide online access to selected journals in full text and also offer services to alert users to the most recent issue of a

particular journal. Some additional journal resources with specific vocational education content include Highwire Press which contains over 4000 VET-related articles and ProQuest's commercial databases – Sociological Abstracts and Education Abstracts.

There are many educational journals which provide some coverage of VET-related topics. The journals currently in print that consistently produce high-quality articles on VET, based on an analysis of the content in the VOCED and Education Research Complete databases, are included in **Table 7**.

Newspapers

Newspapers can provide very up-to-date information on topics relevant to VET, including changes in government policy and current trends affecting the sector. All the major search engines, including Google and Yahoo, have electronic news services that provide the ability to keyword search across thousands of news sources found on the Web. Some also allow alerts to automatically track regular queries. Two commercial databases that supply news services of material not available free on the Web are Factiva and LexisNexis New & Business. Factiva provides full-text coverage from nearly 9000 sources in 22 languages, including local, national, and international newspapers, business magazines, trade publications, and newswires. LexisNexis News & Business

Table 7 Key VET journals

<i>Name</i>	<i>Description</i>	<i>Publisher</i>
<i>Career and Technical Education Research (CTER)</i>	Three issues per year. Covers research and development activities related to career and technical education	Association for Career and Technical Education Research, National Research Center for Career and Technical Education
<i>Education + Training</i>	Nine issues per year. Addresses vocational learning and highlights the changing nature of the partnership between the worlds of work and education	Emerald Publishing Group Limited
<i>Human Resource Development International</i>	Four issues per year. Looks at the practice and research issues of learning and performance by individuals, groups and organizations	Routledge–Taylor & Francis Group
<i>International Journal of Training and Development</i>	Four issues per year. Scholarly research on the determinants of training, training and development practice and policy	Blackwell Publishing
<i>International Journal of Vocational Education and Training</i>	Bi-annual. Features articles on research, theory, and practice broadly related to international VET	International Vocational Education and Training Association
<i>Journal of Education and Work</i>	Five issues per year. International academic research that focuses on the interplay of education and economic systems	Routledge–Taylor & Francis Group
<i>Journal of Further and Higher Education</i>	Four issues per year. Represents the entire field of post-16 education and training	Routledge–Taylor & Francis Group
<i>Journal of Vocational Education and Training</i>	Four issues per year. Looks at the development of practice and theory in work-related education and in particular the study of all aspects of vocational and pre-vocational education throughout the world	Routledge–Taylor & Francis Group
<i>Journal of Workplace Learning</i>	Eight issues per year. Focuses on research into individual, group and team learning in, from and for the workplace	Emerald Group Publishing Limited
<i>Studies in Continuing Education</i>	Three issues per year. Covers all aspects of adult, continuing and professional education and lifelong learning	Routledge–Taylor & Francis Group

provides the full text of thousands of newspapers, newswires, and magazines worldwide. LexisNexis also offers a free to search, pay per view AlaCarte! service.

There are also websites that help to find newspapers from around the world by title, or by region, country, or city, and link to their online equivalents. General news portals include ABYZ News Links, NewsVoyager, Newspapers.com, online newspapers.com, and world-newspapers.com. Only a few news services are on education and their coverage of VET is specific to either the United States or the United Kingdom as shown in **Table 8**.

Organizations/Government Departments/Associations/Networks

National organizations, government departments, member/industry associations, and networks (both formal and informal), whose primary focus is VET, provide useful sources of information. A primary tool for disseminating information about and by these bodies are their websites through the provision of up-to-date information about policy, practice, research, trends, and key topics. Some

of these minor bodies may not have a web presence, although information about these may generally be sourced by other means such as subject directories or via a search engine.

Organizations

National research organizations that have a specific focus on the broad topic of VET or a more specialized interest such as the economics of education, provide access to the most recent research information produced by that organization. The links area of an organization's website can provide access to other related organizations on the same topic/subject. **Table 9** highlights some key vocational education organizations from various regions.

Government Departments

Many national governments have departments whose responsibilities may range from policy development and implementation and administration of the VET system in that country to providing funding for specific or topical programs or areas of research. In some instances, the

Table 8 News services

<i>Name</i>	<i>Description</i>	<i>Availability</i>
Education Week on the Web	A daily update of education news from US newspapers, magazines, and journals	www.edweek.org
Education News	Claims to be the number one source of education news; emphasis on the United States	www.ednews.org
ED.gov	The United States Education Department compiles department related news on its website	www.ed.gov
News Centre	The United Kingdom Department of Innovation, Universities and Skills press releases	www.dius.gov.uk
EducationGuardian	Daily news updates from the UK Guardian newspaper	http://education.guardian.co.uk

Table 9 Key vocational education research organizations

<i>Name</i>	<i>Description</i>	<i>Availability</i>
UNEVOC International Centre for Technical and Vocational Education and Training (TVET)	The Centre assists UNESCO's 192 member states to strengthen and upgrade their TVET systems. Is a specialized center for TVET and the hub of a worldwide Network of TVET institutions	www.unevoc.unesco.org
European Training Foundation (ETF)	Assists countries outside of the European Union to develop quality education and training systems	www.etf.europa.eu
National Centre for Vocational Education Research (NCVER)	Australia's principal provider of VET research and statistical information	www.ncver.edu.au
SEAMEO VOTTECH	Aims to strengthen the VTET systems in SEAMEO Member Countries (Southeast Asia) through collaborative efforts incorporating research and development, program initiatives, networks, partnerships, and information services	www.vottech.org.bn
Bundesinstitut für Berufsbildung (BiBB) (Federal Institute for Vocational Education and Training)	Germany's national center for research into and development of initial and continuing VET	www.bibb.de English language version www.bibb.de/en
Centre d'études et de recherches sur les qualifications (Center for Research on Education, Training and Employment)	French center undertaking research and statistical surveys on the training-employment relationship and the acquisition and certification of vocational skills and job mobility	www.cereq.fr English language version http://mimosa.cereq.fr/angl
Korea Research Institute for Vocational Education and Training (KRIVET)	KRIVET is involved in conducting policy-oriented research on a national basis in the areas of VET, industrial labor policy, VET and the qualification system, and the development, provision and evaluation of VET programs	www.krivet.re.kr
National Center for Career and Technical Education (NCCTE)	Focus is on career and technical education at local, state and national levels in the United States	www.nccte.org

primary responsibility may be for the education sector as a whole of which VET is one component. An example of this, in Australia, is the Department of Education, Science and Training (DEST) whose responsibility is to develop and implement policies in school education, higher education (universities), and VET sectors. Unlike some national government departments, the United States Department of Education has an office specifically focused on vocational and adult education (OVAE). This office acts as the principle adviser to the US Education Department Secretary on departmental matters related to high school, career, technical, adult education, and lifelong learning as well as community colleges, workforce, and economic

development. In Europe, member states of the European Commission are responsible for their education and training systems; however, The European Commission, Directorate-General for Education and Culture works with member states to help achieve common goals by providing policy coordination and funding of programs such as the Lifelong Learning Programme.

As the names and focus of government departments often vary with changes in national government, keeping track of relevant and up-to-date government information can be challenging. To help alleviate this problem, many search engines, such as Google, Ask.com, and MSNSearch, provide an option to limit a search query by a particular

generic domain name. For example, to identify US government materials on the Web, add the site limit syntax to the query, that is, add site:gov in the search box along with the search term. For other countries, include the country code with the site name, for example, site:gov.au (Australian government websites). Alternatively, an online resource such as the University of Michigan Library's Foreign Government Resources on the Web, which provides links to official government websites and documents organized by country, may be used.

Associations and Networks

Vocational education industry and member associations and networks are a useful source of current information on an international, national, or regional basis. An example of an international VET association whose worldwide membership consists of practitioners, researchers, and students in the field of VET as well as institutions, organizations, and companies engaged in human resource development is the International Vocational Education and Training Association (IVETA). IVETA hosts an annual conference with a theme representing current issues in the VET field and publishes the refereed *International Journal of Vocational Education and Training*. Examples of key associations and networks are included in **Table 10**.

Table 10 Key associations and networks

Name	Description	Availability
International Vocational Education and Training Association (IVETA)	Members include vocational educators, vocational skills training organizations, business and industrial firms, and other interested individuals/groups involved in VET worldwide	Restricted free access Membership fee to join www.iveta.org
Vocational Education and Training Network (VETNET)	Provides news and information about VETNET activities and forms the basis for a community of practice in VET research in Europe	Free to access Members only area www.vet-research.net
Australian Vocational Education and Training Research Association (AVETRA)	A national, independent association of research stakeholders and researchers from the TAFE, university, industry and government sectors	Free to access Membership fee to join www.avetra.org.au
Further Education Research Association (FERA)	Aims to disseminate advances in good practice, research findings and policy evaluation relevant to the further education and training sector	Free to access www.fera.uk.net

Table 11 Online conference directories

Name	Description	Availability
AllConferences.com	Provides information about conferences, conventions, exhibits, seminars, workshops, events, trade shows and business meetings. The advanced search allows searching by keyword (e.g., vocational education), date, country or category (e.g., education)	www.allconferences.com
edna.edu.au	The Events link lists conferences, scholarships, celebrations and competitions held for education and training, categorized by events in Australia and outside Australia	www.edna.edu.au
International Congress Calendar	Details of scheduled key future meetings of international organizations	Fee based subscription at www.uia.org

Conferences

Conference papers are listed, along with books and journal articles in the main education indexing services and are increasingly made available on the Web simply by conducting a search engine query. **Table 11** lists online directories which are international in scope.

Statistics

Statistical information relevant to VET can be found in a number of ways, through a variety of sources, and in both raw data formats and processed formats. Most countries have a bureau that publishes official statistical data that include education and training statistics. International organizations like OECD, International Labour Organization (ILO), the United Nations Educational, Scientific and Cultural Organization (UNESCO), and the World Bank also have data banks that include education and training statistics. The National Centre for Vocational Education Research in Australia and the National Center for Education Statistics in the US are examples of organizations offering VET statistics specific to their respective countries. Other key statistical agencies and organizations are listed in **Table 12**.

Table 12 Statistical agencies and organizations

<i>Name</i>	<i>Description</i>	<i>Availability</i>
Career/Technical Education Statistics (CTES)	CTES derives its data from existing National Center for Education Statistics' (NCES) surveys. NCES is the primary entity for collecting education-related data in the U.S. and other surveys it conducts include those relating to post-secondary education and adult literacy	http://nces.ed.gov
Eurostat	The European Commission's statistical portal serving the statistical needs of the European Union. Included under its theme Population and social conditions is the sub-theme Education and training. Available in German, English, and French	http://epp.eurostat.ec.europa.eu
International Labour Organization (ILO), Bureau of Statistics	Provides core official labor statistics, surveys and key indicators for different countries. Products include LABORSTA – database of labor statistics – and Key Indicators of the Labour Market (KILM)	www.ilo.org
Library of Statistics	A unit of Statistics Finland and has a large collection of national and international statistics. It provides links to public agencies and research institutions that produce statistics and its Webstat database can be used to search for statistics by topic and country	www.stat.fi/index_en.html
OECD Statistics Portal	Provides free access to some OECD databases and extracts from other databases. It covers 26 topics that include Education and training and labor	www.oecd.org/
UNESCO Institute for Statistics	Provides global and internationally comparable statistics on education, science, technology, culture, and communication	www.uis.unesco.org
World Bank Data & Statistics	National and international data are provided. Specific topics, including education and labor and employment have their own online databases. Country-specific information is also provided	www.worldbank.org

Glossaries, Dictionaries, and Thesauri

Glossaries contain technical/specialized terms related to a specific topic/subject with accompanying definitions which are usually more detailed than that provided in a dictionary, although subject dictionaries do provide detailed definitions/explanations. Glossaries are usually published as part of a publication where they cover terms used within that publication, but there are also subject-specific stand-alone glossaries. Thesauri contain headings or descriptors used in a particular database, catalog, or index. A thesaurus is used to search for synonyms, related terms, and preferred terms. Its purpose is to provide a controlled vocabulary list of terms from which subject headings are assigned and which aid in information retrieval. A selection of key resources are listed in **Table 13**.

Research in Progress

There is no easy way to stay in touch with what new research in VET is underway or for checking for early dissemination of findings, which are sometimes produced as preprints or as working papers prior to official publication. To find current research in the VET field a search of government education and educational research organizations and centers' web pages will often lead to a research-in-progress section or an online institutional

repository which may include preprint material. These repositories are usually also searchable via Google Scholar. Many institutions are still in the early stages of populating their repositories with data from their own institutions but over time these repositories should become a valuable resource for both research in progress and research publications.

Social Software Tools

The natural evolution of the Web has seen the emergence of social software tools such as wikis, blogs, RSS feeds, and podcasts that facilitate collaboration and sharing between users. The use of these tools in the delivery of VET is still in its infancy, although recent research suggests that using social software can be an effective strategy for knowledge sharing, capability development, and for enhancing teaching and learning in VET. Two tools that are of particular interest are social bookmarking and mailing lists. Social bookmarking involves saving the URLs of websites to a public website and sharing these bookmarks with other users. Tagging them with keywords allows them to be classified and organized into categories. Examples of social bookmarking services are Delicio.us and Furl. Mailing lists continue to play an important role in providing forums for discussions. JISCmail, the National Academic Mailing List Service based in the UK, provides a forum for users from higher and further education and research communities worldwide.

Table 13 Glossaries, dictionaries, and thesauri

Name	Description	Availability
Australian VET Online Glossary	Covers Australian VET terms and includes VET organizations and acronyms in Australia. Also includes some key international terms, organizations and acronyms	www.ncver.edu.au
The Jargon Buster	Covers common terms relating to adult learning in England and Wales. Published in 2005, the print copy is available from the National Institute of Adult and Continuing Education (NIACE) online bookshop	www.niace.org.uk
Jargon Buster (Learning and Skills Network)	Covers terms relating to learning and skill development in the UK	www.lsc.gov.uk/Jargonbuster
Terminology of vocational training policy: a multilingual glossary for an enlarged Europe	Covers key terms used in VET policy in Europe in six languages: English, French, German, Czech, Hungarian, and Polish. Published in 2004, it is available from the European Training Village Bookshop	www.trainingvillage.gr
ERIC Thesaurus	Contains descriptors (controlled-vocabulary) of education-related terms	www.eric.ed.gov
Multilingual Thesaurus of Vocational Training	Contains terms in six languages (English, French, German, Italian, Portuguese, and Spanish). Published in 1992, it is available from the European Training Village Bookshop	www.trainingvillage.gr
VOCED thesaurus	Contains descriptors used in searching the VOCED database. Based on the original APSDEP Thesaurus produced by the International Labour Organization's Asian and Pacific Skills Development Programme (ILO/APSDEP)	www.voced.edu.au

Conclusion

The advent of the Internet and the emergence of new technologies have made information much more accessible to a wider audience of users. At the same time, the Internet has created a tangled maze for the researcher trying to find specific data, particularly, in the area of VET which is often only found after delving deep into the disciplines of the social sciences and education. It is further predicted that more data will be created in the next 5 years than has been collected in the whole of human history resulting in a deluge of information and the challenge will remain to find and extract relevant information. This article has attempted to guide the researcher through the maze to the key information resources pertaining to VET.

Bibliography

- Martin, B. (2007). Scholarly recognition or social impact: Not both? *Campus Review* 17(16), 4.
- Shin, E. (2004). Measuring the impact of electronic publishing on citation indicators of education journals. *LIBRI* 54(40), 221–227.

Further Reading

- Dochartaigh, N. (2007). *Internet Research Skills: How to do Your Literature Search and Find Research Information Online*. London: Sage.
- Evans, V. and Larri, L. (2006). *Networks, Connections and Community: Learning with Social Software: A Report on the Use of Social Software for Knowledge Sharing and Capability Development of Vocational Education and Training Practitioners and in Vocational Education and Training Delivery*. Brisbane: Australian Flexible Learning Framework. http://www.flexiblelearning.net.au/fix/webdav/site/fixsite/users/cpickles/public/Final_Report_Social_Software_for_Learning17April.pdf (accessed August 2009).
- Garfield, E. (2006). The agony and the ecstasy – the history and meaning of the journal impact factor. *Journal of the American Medical Association* 295(1), 9–93.
- Golian-Lui, L. M. (2006). Internet resources for education reference. In Miller, W. and Pellen, R. M. (eds.) *Evolving Internet Reference Resources*, pp 195–208. Binghamton, NY: Haworth Information.
- Lonsdale, M. (2003). *Global Gateways: A Guide to Online Knowledge Networks*. Camberwell, VIC: Australian Council for Educational Research.
- Noruzi, A. (2005). Google scholar: The new generation of citation indexes. *LIBRI* 55(4), 170–180. <http://eprints.rclis.org/archive/00005595/> (accessed August 2009).
- Thomson, I. S. I. (2005). History of citation indexing. <http://scientific.thomson.com/knowtrend/essays/citationindexing/history/> (accessed August 2009).

VOCATIONAL EDUCATION AND TRAINING – TEACHING AND LEARNING

Contents

Employability Skills

Learning Styles in Vocational Education and Training

Modularization in Vocational Education and Training

Personalized Learning and Vocational Education and Training

Recognition of Prior Learning and Vocational Education and Training

The Competency Model

Employability Skills

E Smith, University of Ballarat, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

What Are Employability Skills?

The Concept of Generic Skills: International Developments

As Brown *et al.* (2003: 107) state, “political and business leaders consistently tell us that efficiency and justice depend on people acquiring the knowledge, skills, and capabilities that employers need in an increasingly knowledge-driven economy.” It is generally accepted that such knowledge, skills, and capabilities encompass not only technical skills but also generic skills. Generic skills, sometimes known as key competencies or key skills, have been on the international education policy agenda for some time. International differences exist relating to which competencies should be designated as key (Weinert, 2001). The US SCANS list, produced in 1991, for example, has three foundation areas: basic skills, thinking skills, and personal qualities, that underpin five competencies that relate to resources, infraction, interpersonal systems, and technology (Nash and Korte, 1997). The Australian key competencies list (Australian Education Council (AEC)/MOVEET, 1993), which reigned for approximately 10 years until the introduction of the Employability Skills Framework, had seven key competencies:

- collecting, analyzing, and organizing information;
- communicating ideas and information;
- planning and organizing activities;
- working with others and in teams;
- using mathematical ideas and techniques;
- solving problems; and
- using technology.

Much to the regret of many stakeholders, an eighth key competency – cultural understanding – was considered but not eventually included. The Russian list of key competencies for school education is broader and involves specific application. It comprises competencies in the spheres of: cognitive activities, civil–social activities, socio-working activities, household, and culture/leisure activities (Pavlova, 2005: 348). Despite these differences among countries, it is clear that the notions of a desired list of generic skills, and of their development at least partly through formal education, is well known and well accepted.

Skills for Work or Skills for Life?

In the past decade, there has been a shifting of nomenclature from key skills or key competencies to employability skills. The official definition of employability skills that has been adopted in Australia is as follows:

The skills required to gain employment or establish an enterprise, (and) also to progress within an enterprise or expand employment capability, so as to achieve one’s potential and contribute successfully to enterprise strategic directions. (Department of Education, Science and Training (DEST), 2002)

A similar definition, albeit with a more critical edge, is proposed by Brown *et al.*:

Employability can be defined as the relative chances of acquiring and maintain different kinds of employment. (Brown *et al.*, 2003: 111)

In this definition, relative means relative to other employees or would-be employees.

Employability Skills Framework

The Australian Employability Skills Framework as adopted in 2002 is discussed here as follows.

Skills

- Communication skills that contribute to productive and harmonious relations across employees and customers;
- Team work skills that contribute to productive working relationships and outcomes;
- Problem-solving skills that contribute to productive outcomes;
- Initiative and enterprise skills that contribute to innovative outcomes;
- Planning and organizing skills that contribute to long- and short-term strategic planning;
- Self-management skills that contribute to employee satisfaction and growth;
- Learning skills that contribute to ongoing improvement and expansion in employee and company operations and outcomes; and
- Technology skills that contribute to effective execution of tasks.

Attributes

Loyalty, commitment, honesty and integrity, enthusiasm, reliability, personal presentation, common sense, positive self-esteem, a sense of humor, a balanced attitude to work and home life, an ability to deal with pressure, motivation, and adaptability are the attributes of the prospective employees (DEST, 2002).

The major difference in the skills list from the earlier Australian list of key competencies is the addition of initiative and enterprise skills, self-management skills and learning skills, and the removal of mathematics skills and collecting information. Another key difference is the addition of a list of desired personal attributes. Some controversy raged about this inclusion, and after quite a short time it was agreed that these attributes were not to be taught, but that learners may be familiarized with them as an awareness-raising activity (Australian Chamber of Commerce and Industry, 2007: 59). The final key difference is that each competency is given a descriptor that relates to workplaces; for example, "Communication that contributes to productive and harmonious relations across employees and customers." Thus, it has been said that the employability skills framework "is focused directly on the requirements of an employer and his/her workplace, whereas the key competencies relate to the individual's

capacity to learn and develop key skills that can be utilized in any workplace" (Association of Independent Schools of Victoria (AISV), 2003: 7).

The shift toward a workplace focus in generic skills development is part of a more general tension between general education and vocational education, and is mirrored in differing conceptions of lifelong learning policies (Chapman *et al.*, 2005) who identify three elements of lifelong learning in OECD documents: "learning for a more highly skilled workforce, learning for a better democracy and an inclusive society, and learning for a more personally rewarding life." Thus, to return to the quote from Brown *et al.* (2003) which opened this section, employability skills are said to be necessary for both efficiency and justice.

While this article focuses on the vocational education and training (VET) setting, employability skills are also in the foreground in school education and in university education (e.g., Precision Consultancy, 2007). In school education in Australia, the VET-sector employability skills are addressed in school curriculum in each state and territory in various ways, having replaced key competencies in this role. In university education, employability skills are incorporated within lists of graduate qualities or graduate attributes, which have been developed by all Australian universities. A typical list of graduate attributes is compared with VET-sector employability skills in **Table 1**.

While the university list is broader and at a higher level, conceptually, the lists are fairly similar, partly reflecting the fact that there has been involvement of business groups in setting the employability skills agenda in the university sector as there has been in schools and VET.

The Australian VET sector has thus moved very firmly toward a skills-for-work rather than a skills-for-life approach to generic skills. The prominent role of industry is in line with international trends that have been noted by Trier (2001). The set of employability skills described above, that has become firmly embedded in Australian VET, was developed by the two largest employer peak bodies, the Australian Chamber of Commerce and Industry and the Business Council of Australia. This capture of the employability skills agenda by employer peak bodies has not been without its critics. Sheldon and Thornthwaite (2005) argue that "employer associations in Australia have forged a VET policy agenda that highlights preferred employee attitudes and behaviors."

They suggest that this is a response to a tighter labor market. A focus on employability skills was originally a response to high unemployment because it makes prospective employees more acceptable to employers, but in many Western countries unemployment is now very low and there are labor shortages. In such situations, argue Sheldon and Thornthwaite (2005: 405), the purpose of a focus on employability skills, from an employer's perspective, is to align workers' attitudes to those of management, thereby increasing commitment and reducing labor turnover.

Table 1 List of graduate attributes and VET-sector employability skills

<i>Key competencies</i>	<i>ATN graduate attributes</i>
Initiative and enterprise skills that contribute to innovative outcomes	
Communication skills that contribute to productive and harmonious relations across employees and customers	Communicates effectively in professional practice and as a member of the community
Planning and organizing skills that contribute to long- and short-term strategic planning	
Team work skills that contribute to productive working relationships and outcomes	Can work both autonomously and collaboratively as a professional
Self-management skills that contribute to employee satisfaction and growth	
Problem-solving skills that contribute to productive outcomes	Is an effective problem solver, capable of applying logical, critical, and creative thinking to a range of problems
Technology skills that contribute to effective execution of tasks	
	Operates effectively with and upon a body of knowledge of sufficient depth to begin professional practice
	Demonstrates international perspectives as a professional and as a citizen
	Is committed to ethical action and social responsibility as a professional and citizen
Learning skills that contribute to ongoing improvement and expansion in employee and company operations and outcomes	Is prepared for life-long learning in pursuit of personal development and excellence in professional practice

Data from Ainley *et al.* (2006). Australian Report, MCEETYA.

are simply taken for granted. The perceived preference for employability skills is the stated reason why employer associations lobby for the inclusion of employability skills training in educational institutions. Similarly, in the university context, in a South African study, a government officer was reported as saying “The product that walks out does not meet the needs of industry,” thus leading to a wish to change the university curriculum to improve the product (i.e., graduates) (Kruss, 2004: 679).

However, there is another group of learners not in work: those who are temporarily out of work. Virgona and Waterhouse (2004) studied workers who had become unemployed and noted that the shift from generic to employability skills had been accompanied by a shift of responsibility for career management from employer to employee. They noted that displaced workers had a need to learn how to recognize and articulate the employability skills that they possessed and how to apply them in new situations (Virgona and Waterhouse, 2004: 119).

In disciplines outside education, such as labor market studies, the major focus in discussions of employability skills is on unemployed people and on the factors beyond the individual that create difficulties for people in finding or progressing within employment; these factors are becoming increasingly important as governments attempt to move people from welfare to work (Gore, 2005: 342). Debates in this literature revolve around supply-side approaches (relating to the need for individuals to make themselves more employable) versus demand-side theories (including employer attitudes, the availability of suitable education and training, and assistance for job-seekers). Hillage and Pollard (1998), in McQuaid and Lindsay, (2005: 207–208), discuss employability skills within a framework of employability assets (belonging to the individual), presentation (the use of these assets to get a job), deployment (i.e., use of assets within an organization and the labor market more generally), and context factors (the interaction of personal circumstances and the labor market).

Who Needs Employability Skills?

As implied by the definitions given above, everybody who wishes to work needs employability skills: those in work and those not yet in work. Employability skills are needed to get jobs, to keep them, and to get on at work. As Virgona and Waterhouse (2004: 120) put it, “generic skills develop throughout life and are not simply ‘acquired’ in a single ‘dose’ which can then be applied everywhere.”

Within VET, most attention has been focused on developing employability skills among learners who are not yet in full-time work, that is, young people. It is said that most employers value employability skills above technical competence in new entrants (e.g., ACCI, 2003: 3), although such responses may indicate that technical skills

How Can Employability Skills be Developed?

Inclusion in Qualifications

In Australia, generic skills are required to be included in training packages, which are the industry and occupational-based documents that collect together units of competency and package them into qualifications (Smith and Keating, 2003). The way in which the former key competencies were addressed varied considerably. Sometimes, there were discrete units of competency addressing, for example, communication skills; sometimes, the key competencies were directly mentioned within units, and sometimes key competencies formed a separate part of the unit which

explained how they might be developed; and sometimes with examples from the relevant industry area (Julian, 2004). Employability skills are now replacing key competencies in training packages, with Industry Skills Councils being required to review their training packages and make the required changes. The preferred method of including employability skills in training packages is for advice to be given in each qualification about how employability skills might be developed, rather than including advice in each unit of competency. For example, the Training and Assessment Training Package contains the following advice about problem-solving skills in the Certificate IV in Training and Assessment:

- Identifying hazards and assessing risks in the learning environment;
- Using time management skills in designing learning programs;
- Calculating cost of programs, logistics of delivery, and accessing appropriate resources; and
- Generating a range of options to meet client needs.

Source: National Training Information Service, <http://www.ntis.gov.au/Default.aspx?/trainingpackage/TAA04/qualification/TAA40104/eskills>.

A similar list is provided for each of the other employability skills.

Training packages are also required to address the level at which the relevant employability skill is to be developed. There are three levels, with level 3 being the most sophisticated (Down, 2003).

In Australian schools, employability skills are required to be taught in VET-in-schools subjects, which are now undertaken by over one-third of senior secondary students. The employability skills as indicated in the training package are transferred to the relevant school curriculum document. In addition, though, employability skills are recognized as being important in general school curriculum and are addressed, for example, through teaching programs designed to prepare students in general terms for entry into the workforce.

University curriculum is decided at a local rather than of the national level. Australian universities generally require some attention to be paid to graduate qualities or graduate attributes at the level of formal curriculum documents. In some cases, this is translated into documentation for individual subjects; for example, the University of South Australia requires documents for each subject to show clearly which graduate qualities are addressed at what level. Academic teaching staff may also use their own versions of employability skills; for example, lecturers in an English information technology course developed employability skill sets that they embedded in course design (Cox and King, 2006); an Irish university used VET-sector key skills in a mapping exercise (Tariq *et al.*, 2004).

Teaching

It is all very well to include employability skills in qualifications and curriculum documents, but these inclusions must be accompanied by teaching and training strategies. It has been noted that for this to happen, implementation strategies must be devised that assist teachers to teach the skills, and professional development for teachers should be provided (Down, 2003).

While employability skills can be successfully developed in workplaces, as will be discussed later, there are sound arguments, and strong political lobbying, for employability skills development to take place within educational institutions. As we have seen, within an educational institution context, employability skills may be developed within individual subjects, across subjects or within specific nonsubject programs. The latter is easier to achieve in a school than a VET or university context, because within schools, extra-curricular activities are offered that do not relate to specific units of study. For example, communication skills can be developed through debating competitions, drama productions, or leadership training; teamwork may be developed through programs such as the Duke of Edinburgh award (AISV, 2003: 9).

Another and perhaps more authentic way to develop employability skills is through the introduction of a workplace component into an education course. Periods of work experience, work placements, and internships have often been used for this purpose, in all sectors of education (e.g., Cox and King, 2006). While such placements often have a focus on the development of technical skills, some are more generic. In Australia, for example, short work-experience programs undertaken by school students in Year 10 are likely to be generic in nature, and employers are often asked explicitly to focus on key competencies or employability skills when a student is placed with them (Smith and Green, 2001).

In the absence of the opportunity for such programs, simulated or real enterprises may be set up within, or attached to, the educational institution. These are usually designed to develop particular vocational skills as well as employability skills. For example, a mock office may be set up for students to practice business administration skills (Deissinger *et al.*, 2006). Realistic work flows are set up to simulate the real world and institutions may link with each other to increase verisimilitude. Sometimes, these enterprises take on real work and earn money. These enterprises may be as simple as, for example, the production of greetings cards (Erebus Consulting Partners, 2003) or may be sophisticated enterprises that service large numbers of customers, such as TAFE training restaurants.

While the above suggestions link the development of employability skills to workplaces or simulated workplaces, it is of course possible to develop these generic

skills in contexts that do not relate to workplaces. It could well be argued that much general education aims at the development of such skills and that teachers naturally seek to develop them in their students. However, in the current context of competency-based VET curriculum that focuses greatly on the development of specific technical skills, it could be that this process is almost, by definition, under-developed in VET. Brogan (2006) suggests that employability skills are best taught when the intent to teach them is made clear to learners and learners are required to reflect on their achievements. He suggests that they are appropriately developed within a problem-based learning framework. Reflection may also assist learners transfer their generic skills into workplaces by assisting them to recontextualize and repackaging them (Yashin-Shaw *et al.*, 2004: 401).

A great deal of employability skills development takes place in workplaces. In many countries, young people start work part-time at a young age while still at school (Smith and Green, 2001) and hence are already working before much formal employability skills development is included in their formal education curriculum. A project investigating the ways in which employers managed and developed such very young workers found that employers had a range of strategies to develop employability skills (Smith and Comyn, 2003). These included strong induction programs, buddying arrangements, the allocation of increasing responsibility, and task rotation. The research also produced recommendations for other methods, including a work experience model, individual induction programs, project learning, and training or information for supervisors and co-workers.

A work experience model. Preparing employers for a novice worker in much the same way as they are prepared for work experience would assist the employer in understanding what a novice worker's needs were and how employability skills could most effectively and most speedily be developed.

Individual induction programs. Encouraging employers of novice workers to develop 6-month plans for their new staff would enable a range of experiences and opportunities for feedback to be planned.

Project learning. Engaging novice workers in authentic but small projects, which are of immediate use to the workplace could improve novice workers' confidence.

Training or information for supervisors and co-workers. Structured training for supervisors, buddies, and co-workers in dealing with novice workers and developing their employability skills would be valuable.

Assessing

Assessment of employability skills has long been recognized as a problematic issue (AISV, 2003). As Curtis (2004)

points out, generic skills constitute a complex construct and assessment cannot be straightforward. In assessing generic skills, we are to some extent assessing the learner as a person, and this is not only complex, it is also problematic (Williams, 2005). As Williams (2005: 45) says

an assessor must necessarily position themselves (sic) within the regime of dominant cultural norms which defines competency standards and informs the assessment ... (generic skills) are necessarily an attempt at normalization to the values, beliefs and expectations of the dominant culture which defines them.

Curtis (2004) suggests that the assessment of generic skills may be undertaken in one of four ways: holistic judgment by a teacher or panel of teachers; portfolio assessment; assessment in the workplace; and a standardized assessment instrument such as that used for university entry purposes in some countries. In universities, capstone assessment tasks may draw together learning from a whole course to illustrate the development of graduate attributes. This is the case, for example, in Charles Sturt University's VET teacher-training degree where a final subject includes an assignment where students are required to map and reflect on their progress in each of the graduate qualities during the lifetime of the course.

Clayton *et al.* (2004) suggest that whichever assessment method is selected, transparency and the provision of multiple opportunities to demonstrate achievement are key factors in good assessment. A well-known exemplar of the portfolios system within the Australian VET system is LINKup, a system utilized in the engineering discipline within TAFE in South Australia. In this system, students are asked to enter their own evidence to show their achievement of generic skills (Denton, 2005). This online system allows students to view the descriptors for the employability skills in detail and provides examples of evidence that might be provided.

Recording

Recording of assessment of generic skills is a key issue and one tied intimately to the assessment process, particularly in a portfolio approach. While schools in Australia, in some jurisdictions, report on employability skills as part of school-to-work portfolios, in VET institutions, there is no requirement to report on the attainment of employability skills. In fact, there is no place for such recording on official transcripts for qualifications. Universities, at least in Australia, have not adopted standard methods for the recording of achievement. In the light of such confusion, an agreement was reached at a national policy level in Australia that the best method of recording was to make each individual person responsible for recording his or her own

achievement of employability skills over a working lifetime (DEST, 2004). This method is only possible due to the advent of online technologies allowing each person's record to be continuously updated. As a consequence, funding was provided to an organization to develop a website for this purpose with the idea that individuals would record their development of employability skills through education, employment, community, and recreational activities.

Emerging Trends

Where Does the Responsibility Lie?

Underlying the whole employability skills debate – and also the methods of teaching, assessment, and recording – is the question of whose is the responsibility for developing employability skills in individual people. While industry often advocates for the development of these skills in educational institutions, and it has been stated that employers are “resistant to the idea that they have a positive contribution to make to improving employability” (Gore, 2005), it is also the case, as shown above, that they can be most appropriately developed in context, at work. There are reasons why workplaces are appropriate, and even vitally important, sites for their development. Employability skills are developed throughout one's working life, and hence employers need to view the process of employability skills development as a whole-of-workforce issue, not just one that relates to new entrants in their first jobs. The range of employability skills possessed by young workers starting their first jobs varies greatly. Some may have well-developed skills and others, sometimes through no fault of their own, poorly developed skills. Employers need to be prepared for the full range of employability skills, particularly when they are recruiting younger teenagers and in the current tight labor market where they cannot pick and choose. Employability skills are context-bound in that different industries and employers value and weight the skills and attributes quite differently. Hence, there are sound economic reasons for employers contributing to the cost of their development. Moreover, the worth of employability skills can only be fully appreciated in the workplace where the consequences of such skills can be seen.

However, not everyone has access to workplaces, and hence development within institutions must always take place to some extent. Further, many of the employability skills are those which underpin a sound general education and would therefore naturally be addressed in schools, VET providers, and universities. Individuals must also bear some of the responsibility, and the portfolio methods of assessment and reporting reflect this assumption.

Mismatch Between Learners' and Employers' Viewpoint

Employability skills are mentioned by employers as being vitally important in recruiting workers (e.g., ACCI, 2007). Yet, it is not easy for learners to understand this. There is certainly some evidence that learners tend to undervalue generic skills compared with technical skills (e.g., Saunders *et al.*, 2004), and this is perhaps an inevitable consequence of what some perceive as an overly vocational curriculum. Does this mean that VET courses should be restructured to focus much more explicitly on employability skills? Or merely that individuals need to take more responsibility for their own employability skills? But would an increased focus on employability skills reduce and/or devalue necessary technical content?

What Skills Do Teachers and Others Need to Develop Employability Skills in Learners?

It has been recognized that teachers need specific skills to develop these skills in their learners, but there has been little work done to identify what particular skills are needed (AISV, 2003: 6). In the current Australian minimum qualification for VET teachers, the Certificate IV in Training and Assessment, there is no explicit mention of employability skills in units or elements of competency. A study in Australian TAFE Institutes in 2003 (Callan, 2004) found that teachers were not well prepared to teach generic skills. Students also reported that teachers left them “to guess or infer the actual importance of many (generic) skills” (Callan, 2004: 61).

Smith and Comyn (2003) found evidence, based on their study of companies that recruited large numbers of part-time teenage workers (e.g., fast food outlets) that attention was paid to developing skills in those supervising and training such workers. It was suggested that training for the supervisors and trainers needs to include an understanding of what it is like to begin working life, as well as suggested processes for assisting the development of employability skills. There was little standardization of learning procedures except in large companies which routinely recruited large numbers of teenagers. In one large fast food company, part of the performance management of managers depended on their ability to develop employability skills in young workers. It is possible that this model could be transferred to the education sector, so that performance appraisal of VET teachers includes their ability to develop employability skills in their learners.

Conclusion

It is generally agreed that employability skills are important for everybody, but there remain important areas of

disagreement that will probably never be resolved. Some of these areas of disagreement are as follows:

- What is the nature of employability skills: should they be weighted toward what employers are seen to want, or more broadly conceptualized?
- How much responsibility should the education system (as opposed to employers and/or individuals) take for their development?
- In particular, what should be the responsibility of the sectors of education that provide education for young people before they start their first full-time jobs?
- How can awareness of a desired set of employability skills be raised among learners, teachers/trainers, and employers?
- How can employability skills be best taught, assessed, and reported?
- How can teachers, trainers, and others responsible for the management of learning in students and workers be better trained to develop employability skills in those for whom they have responsibility?

These areas are all interrelated, and so it is likely that there will always be tension around the concept of employability skills. This tension is productive rather than otherwise.

See also: Australia; Economic Outcomes of Adult Education and Training; Employability of University Graduates and Graduate Outcomes; Industry Involvement in the Vocational Education and Training System; Lifelong Learning; The Capabilities Approach; Training and Learning in the Workplace; Vocational Education and Training and the School-to-Work Transition.

Bibliography

- AISV (Association of Independent Schools of Victoria) (2003). *Employability Skills*. Melbourne, VIC: AISV.
- ACCI (Australian Chamber of Commerce and Industry) (2003). *Giving Australian Youth Relevant Skills – Meeting Employers' Expectations*. Canberra: ACCI.
- ACCI (2007). *Skills for a Nation: A Blueprint for Improving Educational and Training*. Canberra: ACCI.
- Australian Education Council (AEC)/MOVEET (1993). *Putting General Education to Work: The Key Competencies Report (The 'Mayer Report')*. Canberra: AGPS.
- Brogan, M. (2006). 'What you do first is get them into groups': Project-based learning and teaching of employability skills. *Fine Print* 29(2), 11–16.
- Brown, P., Hesketh, A., and Williams, S. (2003). Employability in a knowledge-driven economy. *Journal of Education and Work* 16(2), 107–126.
- Callan, V. (2004). VET teacher and student attitudes about generic skills. In Gibb, J. (ed.) *Generic Skills in Vocational Education and Training: Research Readings*, pp 53–68. Adelaide, SA: National Centre for Vocational Education Research (NCVER).
- Chapman, J., Gaff, J., Toomey, R., and Aspin, D. (2005). Policy on lifelong learning in Australia. *International Journal of Lifelong Education* 24(2), 99–122.
- Clayton, B., Blom, K., Meyers, D., and Bateman, A. (2004). Assessing and certifying generic skills. In Gibb, J. (ed.) *Generic Skills in Vocational Education and Training: Research Readings*, pp 157–172. Adelaide, SA: NCVER.
- Cox, S. and King, D. (2006). Skill sets: An approach to embed employability in course design. *Education + Training* 48(4), 262–274.
- Curtis, D. (2004). The assessment of generic skills. In Gibb, J. (ed.) *Generic Skills in Vocational Education and Training: Research Readings*, pp 136–156. Adelaide, SA: NCVER.
- Deissinger, T., Smith, E., and Pickersgill, R. (2006). Models of full-time and part-time vocational training for school-leavers: A comparison between Germany and Australia. *International Journal of Training Research* 4(1), 30–50.
- Denton, R. (2005). Employability skills assessment: Not just designed...but engineered to work! *14th National Vocational Education and Training Research Conference*. Wodonga, VIC, Canada, 8 July.
- DEST (Department of Education, Science and Training) (2002). *Employability Skills for the Future*. Canberra: DEST.
- DEST (2004). *Final Report: Development of a Strategy to Support the Universal Recognition and Recording of Employability Skills*. Canberra: DEST.
- Down, C. (2003). Employability skills: Key competencies recycled or a step forward? *12th National Vocational Education and Training Research Conference*. Perth, 8–11 July.
- Erebus Consulting Partners (2003). *Enterprise Education*. Canberra: Department of Education, Science and Training.
- Gore, T. (2005). Extending employability skills or solving employers' recruitment problems? Demand led approaches as an instrument of labour market policy. *Urban Studies* 42(2), 341–353.
- Julian, T. (2004). Employability skills: Balancing the equation. In Gibb, J. (ed.) *Generic Skills in Vocational Education and Training: Research Readings*, pp 84–94. Adelaide, SA: NCVER.
- Kruss, G. (2004). Employment and employability: Expectations of higher education responsiveness in South Africa. *Journal of Education Policy* 19(6), 673–689.
- McQuaid, R. and Lindsay, C. (2005). The concept of employability. *Urban Studies* 42(2), 197–219.
- Nash, B. E. and Korte, R. C. (1997). Validation of SCANS competencies by a national job analysis study. In O'Neil, H. F. (ed.) *Workforce Readiness: Competencies and Assessments*, pp 77–102. Mahwah, NJ: Erlbaum.
- Pavlova, M. (2005). Life skills for employability, citizenship and sustainable development: A case study of vocationalization in Russia. *Prospects* XXXV(3), 343–354.
- Precision Consultancy (2007). *Graduate Employability Skills: Paper prepared for the Business, Industry and Higher Education Collaboration Council*. Melbourne, VIC: Precision Consultancy.
- Saunders, M. N. K., Skinner, D., and Beresford, R. (2004). Mismatched perceptions and expectations: An exploration of stakeholders' view of key and technical skills in vocational education and training. *Journal of European Industrial Training* 29(5), 369–382.
- Sheldon, P. and Thornthwaite, L. (2005). Employability skills and vocational education and training policy in Australia: An analysis of employer association agendas. *Asia Pacific Journal of Human Resources* 43(3), 404–425.
- Smith, E. and Comyn, P. (2003). *The Development of Employability Skills in Novice Workers through Employment*. Adelaide, SA: NCVER.
- Smith, E. and Green, A. (2001). *School Students' Learning from Their Paid and Unpaid Work*. Adelaide, SA: NCVER.
- Smith, E. and Keating, J. (2003). *From Training Reform to Training Packages*. Tuggerah Lakes, NSW: Social Science Press.
- Tariq, V. N., Scott, E. M., Cochrane, A. C., Lee, M., and Ryles, L. (2004). Auditing and mapping key skills within university curricula. *Quality Assurance in Education* 12(2), 70–81.
- Trier, U. (2001). *12 Countries Contributing to DeSeCo – A Summary Report*. Neuchâtel: University of Neuchâtel on behalf of the Swiss Federal Statistics Office.
- Virgona, C. and Waterhouse, P. (2004). Making experience work: Displaced workers provide new insights into generic skills. In Gibb, J. (ed.) *Generic Skills in Vocational Education and Training: Research Readings*, pp 109–123. Adelaide, SA: NCVER.

- Weinert, F. (2001). Concept of competence: A conceptual clarification. In Rychen, D. and Salganik, L. (eds.) *Defining and Selecting Key Competencies*. Gottingen: Hogrefe and Huber.
- Williams, C. (2005). The discursive construction of the 'competent' learner-worker: From key competencies to 'employability skills'. *Studies in Continuing Education* 27(1), 33–49.
- Yashin-Shaw, I., Buckridge, M., Buckridge, P., and Ferres, K. (2004). Enhancing graduate attributes: Mechanisms for facilitating the transfer of academically acquired generic skills into the workplace. In Danaher, P. A., Macpherson, C., Nouwens, F., and Orr, D. (eds.). *Lifelong Learning: Whose Responsibility and What is Your Contribution? Refereed Papers from the 3rd International Lifelong Learning Conference, Yeppoon, Central Queensland, Australia, 13–16 June 2004*, pp 396–402. Rockhampton, QLD: Central Queensland University Press.

Learning Styles in Vocational Education and Training

J Vorhaus, University of London, London, UK

© 2010 Elsevier Ltd. All rights reserved.

This article provides an overview of research on learning styles, in the context of vocational education and training (VET). Since there are over 70 models of learning styles, attention is largely given over to the secondary literature – to reviews and analyses of learning-styles research. These tend to draw attention to the complexity of the field, and the uncertainty that exists about the accuracy and reliability of many of the models now widely in use. We identify the characteristics of the best-known models, the implications for teaching and training, and the empirical evidence of their impact on learning and performance. We review what is known about the learning styles of vocational learners, and how far teachers and trainers take account of these. This requires acknowledgment of the distinguishing features of VET as a teaching and learning environment. Research has recently focused on situated and informal learning, communities of practice, and legitimate peripheral participation, and we explore how learning styles theory and practice are related to these themes.

Definitions

Any review of definitions of learning styles would take up most of this encyclopedia; no account is beyond dispute, but the following is probably less controversial than most.

In a review of research, Adey *et al.* (1999) distinguished between learning styles, learning strategies, and learning skills. A learning style is an individual's deep-rooted preference for a particular type of learning; the example is given of folding one's arms – each person has a preferred way of doing this, and while we can all learn to fold our arms in different ways this would be less natural and harder to achieve as compared with our preferred way of going on. Equally, if some theories of learning styles are correct, there are people who are imagers – they prefer to learn from pictures and diagrams – and there are people who are verbalizers, preferring to learn from words. Imagers can learn from words and verbalizers can learn from images, but in both cases this would represent the harder option.

Learning skills are likened to tricks – they are specific, teachable, and designed for individual tasks. An example of a learning skill is the use of a mnemonic as an aid to memory – as when we recite “Richard of York Gave Battle in Vain” to help remember the seven colors of the rainbow.

Between these two options are learning strategies, which denote the groups of skills acquired by a learner

for particular purposes, such as setting objectives, and selecting and formulating questions (*ibid.*: 2).

In their review, Entwistle and Peterson offer a similar account. Learning styles are defined as “consistent preferences for adopting learning processes, irrespective of the task or problem presented”; learning strategies, on the other hand, are “specific to education and training, being more context-specific ways of tackling learning tasks that involve characteristic combinations of intentions, learning processes and study activities” (Entwistle and Peterson, p. 2).

Learning-Style Models

As with definitions, there are numerous models of learning styles; this article offers only a brief summary, identifying a small sample that has influence in the context of VET.

Witkin (1977) explored individual differences in perceptual processing, and developed a theory of perception that drew attention to field dependence and independence. While some people find it easy to see a figure embedded in a background display, others find that hard to do. Witkin suggested that field-independent people see the figure easily when they are not confused by what surrounds it. This was extended to the notion of a learning style: some people are able to analyze something in isolation from surrounding issues, while others need to take account of the surround throughout the learning process. In a similar vein, Pask (1976) suggested that there are those who learn by taking individual items in turn, before putting these together to form a whole; others prefer to grasp the whole right from the start. In other words, some learners tend to be serialists and others tend to be holists.

Marton and Säljö (1976) argued that some learners are more typically ‘deep processors,’ while others are ‘surface processors.’ Deep processors generally seek to understand underlying concepts and theories, and aim to make connections between these and their existing knowledge. Surface processors are more likely to settle for knowing the facts or techniques without, at the same time, acquiring an understanding of the underlying theory.

A widely used model of learning styles was developed by Kolb (1976). He suggested that individuals learn and solve problems by progressing through a four-stage cycle: concrete experience, reflective observation, abstract concepts, and active experimentation. Concrete experience and abstract concepts represent two ends of one spectrum,

and active experimentation and reflective observation represent two ends of a second spectrum. Kolb identified four learning styles: accommodator, assimilator, diverger, and converger – and these are usefully summarized in Coffield (2004a: 62):

- The accommodating style (concrete, active): emphasises concrete experience and active experimentation; likes doing things, carrying out plans; good at adapting to changing circumstances; solves problems in an intuitive trial-and-error manner; at ease with people but sometime seen as impatient and 'pushy'.
- The assimilating style (abstract, reflective): prefers abstract conceptualisation and reflective observation; likes to reason inductively and to create theoretical models; is more concerned with ideas and abstract concepts than with people; thinks it more important that ideas be logically sound than practical.
- The diverging style (concrete, reflective) emphasises concrete experience and reflective observation; is imaginative and aware of meanings and values; views concrete situations from many perspectives; adapts by observation rather than action; interested in people and tends to be feeling-oriented.
- The converging style (abstract, active) relies primary on abstract conceptualisation and active experimentation; good at problem solving, decision making and the practical application of ideas; does best in situations like conventional intelligence tests; is controlled in the expression of emotion and prefers dealing with technical problems rather than interpersonal issues.

According to Kolb, while someone may prefer one learning style in one situation, and another style in another situation, the preferred style is likely to remain the same within the same learning context.

Honey and Mumford (1992: 1) define a learning style as "a description of the attitudes and behaviour which determine an individual's preferred way of learning," and they identify four learning styles under the headings of activists, reflectors, theorists, and pragmatists (Honey and Mumford, 2000).

Activists' strengths include a tendency to be flexible, open minded, and prepared for new situations. Weaknesses include a tendency to take the immediately obvious action without thinking through the consequences and taking unnecessary risks. Reflectors' strengths include a tendency to be careful, thorough, and methodical; weaknesses include a tendency to hold back from direct participation, being slow to reach decisions, and being too cautious and risk averse. Theorists' strengths include a tendency to be logical, rational, and objective, and to ask probing questions. Weaknesses include a low tolerance for uncertainty and ambiguity, and impatience with the intuitive and subjective. Pragmatists strengths include an eagerness to test things out in practice, being practical,

down-to-earth, and realistic. Weaknesses include a tendency to reject anything without an obvious application, impatience with indecision, and being more oriented toward tasks than people (Coffield, 2004a: 81).

The two primary uses of the Honey and Mumford Learning Styles Questionnaire (Honey and Mumford, 2000) are to devise personal development plans and to show managers how to assist staff by, for example, choosing activities that are suited to the learning style of the learners. However, they point out (Honey and Mumford, 2000: 52): "managers, if they encourage learning at all, will tend to do so in ways consistent with their own learning styles."

Cognition-Centered Approaches

Allinson and Hayes developed the Cognitive Styles Index (CSI) for use in adult organizational contexts. The items of the CSI reflect their view that: "intuition, characteristic of right-brain orientation, refers to immediate judgement based on feeling and the adoption of a global perspective. Analysis, characteristic of left-brain orientation, refers to judgement based on mental reasoning and a focus on detail" (Allinson and Hayes, 1996: 122).

At one end of the wholist-analytical dimension of cognitive style are those who break down information into its component parts (the analytics), while others prefer a global or overall view of information (the wholists) (Riding, 1997: 30). Among holists, there is a tendency not to pay sufficient attention to the distinctions between the parts of a problem; among analytics, there is a tendency to focus on the parts at the expense of other parts or the whole.

While acknowledging that cognitive style can be shaped by culture and altered by experience, Allinson and Hayes argue that it may prove to be useful in work settings – through fitting people to jobs and, where feasible, adjusting job demands to what best suits the individual. Intuitive managers were found to be more liked, irrespective of the style of their subordinates, and matched styles were found often to be effective in mentoring relationships.

According to Riding (1997), verbalizers "consider the information they read, see or listen to, in words or verbal associations"; while imagers, when they read, listen to or consider information, experience "fluent spontaneous and frequent pictorial mental pictures" (Riding, 1994: 48). Riding (1991) has devised a computer-administered assessment of the verbal-imagery dimension of cognitive style (the Cognitive Styles Analysis (CSA)), and he suggests that learning is affected by the interaction of verbal-imagery cognitive style and the form in which material is presented. For example, Riding and Watts (1997) presented participants with a choice of three instructional formats (unstructured verbal, structured verbal, and structured pictorial) and found that most verbalizers selected the verbal version while most imagers selected

the pictorial version. Riding (2001: 61) concludes: “imagers generally learn best from pictorial presentations whereas verbalisers learn best from verbal presentations.”

In research carried out with high school students and trainees in industry, Riding and Pearson (1994) suggest that the wholist-analyst dimension interacts with the structure of teaching material to affect learning performance. Riding and Al-Sanabani (1998) observed an interaction between gender and wholist-analyst style in their effect on learning from materials that varied in their use of headings and subdivisions of textual content. Sadler-Smith and Riding (1999) explored the learning preferences of business and management undergraduates and observed that analytics preferred to have control themselves whereas the wholists had no preference (Riding *et al.*, 1997; Glass and Riding, 2000).

The State of the Field

There are now at least 71 models of learning styles. These have been reviewed by Coffield *et al.* (2004a). Despite the proliferation of models, Coffield concluded that the field, as a whole, could not be characterized as healthy: “the field of learning styles research as a whole is characterised by a very large number of small-scale applications of particular models to small samples of students in specific contexts” (ibid. 1). This makes it difficult to assess the impact of learning styles on learning and teaching, including learning and teaching in vocational contexts: “there are very few robust studies which offer reliable and valid evidence and clear implications for practice based on empirical findings.” Nor does the field exhibit the conditions necessary for the development of a robust evidence base: “Leading theorists, and developers of instruments tend to ignore, rather than engage with each other. The result is fragmentation, with little cumulative knowledge and cooperative research” (ibid.: with omissions). Besides this, there are large commercial interests at stake in respect of some of the better-known models, and that is not always conducive to encouraging robust and independent evaluation.

Learning Styles and VET

Much of the research on learning styles in relation to VET has been carried out in Australia. In general terms, research indicates that VET learners are inclined to:

- be more visual than verbal, preferring to watch and see rather than read and listen,
- prefer to learn by doing and by practicing, and
- prefer learning in groups, to have instructor guidance, and to have a clear understanding of what is required of them.

For example, Christie and Choy (1998) showed that VET learners in Queensland were averse to textual presentations, and that they tended not to be independent learners. A study by Smith (2000) in Victoria produced similar findings. Brennan (2003), in an online context, showed up the importance of social contexts for learning among VET students, and the need to develop a lower reliance on texts, and greater self-direction.

However, VET students are by no means a homogeneous group. The quoted studies found that female students were more verbal than males and were likely to be self-directed. Students in areas such as health and community studies and business were found to be more verbal and less hands-on; while apprentices were more hands-on and less verbal.

Smith and Dalton (2005) examined learning styles in the VET context. They showed how practitioners view style differences between students, and how they take account of those differences in designing and delivering teaching. Sites included five technical and further education (TAFE) institutes and one professional network of trainers in private and public-registered training organizations. Participants included teaching staff, students, and management staff, and were largely self-selected. The major findings were as follows:

- VET providers deliver programs in new and flexible ways, using a wide range of media and approaches, in a variety of locations.
- The diversity of students requires that teaching staff are able to adapt to different learner groups and individuals.
- VET teachers generally recognize group and individual learning style differences among their learners.
- Teachers see the identification of style and their responses to it as part of good professional practice.
- Teachers rely on previous experience: they have a range of methods for identifying individual and group learning styles, and for responding to these.
- Teachers rely on practice rather than theory; they observe learner reactions to the learning context and to the media used, and they adapt their teaching practices in the light of those observations.
- Professional development should focus more on practical examples of good practice and practical teaching settings, and less on theories of learning styles.
- Students had a limited knowledge of learning style but thought that teachers did take account of their learning characteristics in their teaching.
- There is need for the development of learning-to-learn training for students.

Buch and Bartley (2002) have explored the hypothesis that learning style influences training preference. A total of 165 employees from a large US financial institution completed the Kolb Learning Style Instrument and

a survey that measured preferences for styles of training delivery. Results found evidence of a relationship between the two, with convergers showing a stronger preference for computer-based delivery and assimilators showing a stronger preference for print-based delivery. However, results also revealed an overall preference for classroom-based delivery for adults in the study, irrespective of their learning style.

Flexible Learning in the Workplace

Sadler-Smith and Smith (2004) reviewed the literature on flexible learning programs and learning styles. There is a growing use of flexible forms of delivery for workplace learning and development (Calder and McCollum, 1998). Distance learning, programmed instruction, technology-based training, telematics, and e-learning enable learners to acquire job-related knowledge and skills in ways that match their learning preferences.

Nikolova and Collis (1998) examined flexible instruction from the point of view of designers' and learners, and argue for a role for telematics. Morgan *et al.* (2000) explored the instructional features of learning environments as a source of understanding the relationship between the learner, the environment, and learning outcomes. The interaction of learners' style and instructional design has already been discussed; in this context, Ford and Chen (2001) provide evidence that matching and mismatching can have significant effects upon learning outcomes.

Learning preferences

Researchers have explored the impact of preferences on work-based learning (Allinson and Hayes, 1996; Riding and Sadler-Smith, 1992; Smith, 2000a, 2000b). Smith (2000a, 2000b) examined the styles and learning preferences of over a thousand VET learners in Australia. His evidence suggests that the majority of VET learners tend to be dependent learners, preferring visual means of presenting and absorbing information, while resource-based flexible delivery tends to assume that VET learners are independent, and prefer text- or oral-based means of presenting and absorbing information.

Through interview-based research with apprentices and trainers, Brooker and Butler (1997) suggested that apprentices rated highly the learning methods that involved structured approaches with assistance from other experts in the workplace, and that learning or practicing alone were not favored approaches. In addition, Smith's (2000a) research with apprentices provided further evidence for more structured, rather than self-directed, approaches to learning.

Billett (1996) used observational approaches in the workplace to explore the use of resource-based learning materials and engagement in everyday practice. He found that everyday practice was consistently viewed as more

effective than instructional materials. In the UK, Sadler-Smith *et al.* (2000) compared managers' perceptions of the effectiveness of a range of learning methods. Work-based methods, such as on-the-job learning, were perceived as being more effective – and were implemented more frequently – than other methods such as taught courses, while the method perceived as least effective (and used least frequently) was distance learning.

Sadler-Smith *et al.* (1999) investigated the learning preferences of personnel practitioners in the UK and found that work-based and traditional learning methods were preferred over self-directed methods. This preference is not confined to individuals in the workplace; Sadler-Smith and Riding (1999) found that business and management students preferred dependent and collaborative learning methods to autonomous methods.

Readiness for flexible learning

Smith (2001) developed workplace strategies for improving the readiness for flexible learning. Success requires that enterprises accept the need to develop learning to learn and task-related skills of workplace learners, through the use of instructional design, learner-centered, and workplace support strategies.

Riding and Sadler-Smith (1997: 203) identified strategies for trainees to make learning easier by playing to strengths and overcoming weaknesses. This requires that learners are trained to be self-reliant (see Gropper, 1983: 124). It has been argued that learners' understanding of their own styles promotes their performance in independent learning programs (Boote, 1998), and Smith (2000a, 2000b) argues for the need for strategies that enable vocational learners to become self-directed learners in a flexible learning context.

In their research, Warner *et al.* (1998) found an un-readiness among vocational learners for the self-directed learning required by flexible delivery. Smith (2000a, 2000b) and Sadler-Smith *et al.* (2000) suggest that workplaces often lack effective learner support for flexible delivery, and that includes an unresponsiveness to differences in individual learning styles. Learner support would (1) identify learners' styles and give appropriate feedback; (2) enable learners to distinguish among styles, strategies, and preferences; and (3) use learning contracts and integrate an awareness of learning styles into these. Workplace trainers will themselves need training in how to identify learning styles, and incorporate that knowledge into their work (Smith, 2001).

Related Theories of Learning

The understanding and use of models of learning styles must go hand in hand with some knowledge of related models of learning in the workplace. We can only briefly

note these here. Eraut has emphasized the nature and importance of nonformal learning, implicit learning, and tacit knowledge in the workplace (Eraut, 2000). The distinguishing feature of learning and knowledge of these kinds is that the learner is either (largely) unaware of much of it, or it remains far from explicit. However, explicitness is a distinguishing feature of learning-style theory and practice – learners being encouraged to become aware of their learning preferences and dispositions. There is, then, a question of which combination of attention to the explicit and implicit works best in vocational learning environments.

Situated learning theory provides resources for analyzing how learners become participants in a community of practice. The concept of “legitimate peripheral participation,” developed by Lave and Wenger (1991), is used in connection with their account of how learners participate in communities of practitioners, and, in so doing, acquire the knowledge and skill to allow for full participation in the practices of the community. A “community of practice” describes how people learn through mutual engagement in common tasks. The work by Fuller and Unwin (2003) on apprenticeship in the UK illustrates the detail and application of these ideas.

Theories of learning styles are sometimes criticized for taking too little account of the social and contextual detail of learning, giving insufficient attention to the nature of the workplace and the distinguishing features of the relationships and communities that develop in them. Profiles of learners in the workforce provide one example: UK evidence shows that these differ from learner profiles found in other sectors in post-16 education. A longitudinal study of literacy and numeracy interventions in the workplace found, from a sample of 564, that almost two-thirds were male, the average age was just over 40, and that English for Speakers of Other Languages (ESOL) learners were heavily represented – 35% of the sample. This compares with 3% ESOL in the UK employee workforce as a whole, while in many other post-16 education contexts (including the largest, Further Education) we find a lower average age and a higher representation of women (Wolf and Evans, *in press*). These distinguishing characteristics of workplace learners will have a bearing on how they learn and on how they prefer to be trained and mentored. The use and development of learning-style practices in the workplace would profit from more extensive use of situated and contextualized learning theory.

Concluding Remarks

Trainers are busy people: there are many elements of pedagogy to think about besides learning styles, and there are numerous learners making demands on their time. Nor can we realistically expect practitioners to

assess each of the many models of learning styles competing for their attention. How are trainers expected to manage? Smith (2006: 265) found that teachers often revealed an understanding of learning styles in the absence of any knowledge of theory: “functional and effective approaches are taken by teachers in the absence of theoretical understandings of style, but nevertheless developed to quite high levels of sophistication” (Smith, 2006: 265). He also found that teachers did not, on the whole, distinguish between different forms of learning style, learning preference, or cognitive style (*ibid.*). It would, in any event, make a large demand on time if teachers were to assess all the different theories of learning styles, since there so many, and of uneven quality, and there is no clear and agreed-upon body of knowledge. In addition, there are numerous other practices to take account of that promise significant benefits to learners – formative assessment, for example (Black and Wiliam, 1998).

The question of which theory of learning styles to adopt will also depend on the context in which it is to be applied – the institution, culture, curriculum, and so on. Unwin *et al.* have explored the implications of this for learning at work. They offer the example of a ‘tutor pack’ in a component-manufacturing plant, which might be thought of as a standard training manual for use with new shop-floor operatives, but that would be quite misleading: “all those who use the pack on a daily basis (operative, shopfloor coaches, trainers and shift supervisors) regard it as a ‘live’ artefact to which they regularly contribute and, by doing so, engage in a collaborative process where learning crosses role boundaries and takes place on several levels. Thus the pack is not limited to being a mechanism for training new recruits but has become a vehicle for the consideration of how work is organised, how knowledge can be shared, and how new knowledge might be created” (Unwin, 2005).

Two points follow for research on learning styles and VET: first, any theory has to acknowledge context in its account of learning styles. (For a fierce expression of this view, see Reynolds, 1997.) Second, the application of theory will itself be subject to the influence of context, and that will require of those using it – trainers, mentors, or learners – a detailed understanding of the workplace, its working culture and practices, and so on.

This article points to a final, as yet unanswered, question. Much of the research suggests a preference among workplace learners for traditional methods of teaching and learning, in a formal learning environment (such as a classroom) that includes a teacher, trainer, or mentor, and that allows for a fair degree of dependency. However, many of our hopes for improving learning in the context of VET lie in the direction of supporting learners to become more autonomous – able to direct their own learning using various media, and often on their own or away from the presence of an expert. In this same spirit,

the priority among many researchers and policymakers is personalized learning and learning to learn, enabling learners to become self-directed, independent, and autonomous. The question remains: do we know enough about the actual preferences and dispositions of VET learners, and the conditions and environments in which they learn, to know how these aspirations are to be achieved in practice?

See also: Cultural–Historical Activity Theory; Learning Strategies; Personalized Learning and Vocational Education and Training; Self-Regulated Learning and Socio-Cognitive Theory.

Bibliography

- Adey, P., Fairbrother, R., and William, D. (1999). *Learning Styles and Strategies: A Review of Research*. London: King's College, School of Education.
- Allinson, C. W. and Hayes, J. (1996). The cognitive style index. *Journal of Management Studies* **33**, 119–135.
- Allinson, C. W. and Hayes, J. (1996). The cognitive style index: A measure of intuition-analysis for organisational research. *Journal of Management Studies* **33**(1), 119–135.
- Billett, S. R. (1996). Situated learning: Bridging sociocultural and cognitive theorizing. *Learning and Instruction* **6**, 263–280.
- Black, P. J. and William, D. (1998). *Inside the Black Box: Raising Standards through Classroom Attainment*. London: King's College London.
- Boote, J. (1998). Learning to learn in vocational education and training: Are students and teachers ready for it? *Australian and New Zealand Journal of Vocational Education Research* **6**(2), 59–86.
- Brennan, R. (2003). *One Size Doesn't Fit All: Pedagogy in the Online Environment*. Adelaide: NCVER.
- Brooker, R. and Butler, J. (1997). The learning context within the workplace: As perceived by apprentices and their workplace trainers. *Journal of Vocational Education and Training* **49**(4), 487–510.
- Buch, K. and Bartley, S. (2002). Learning style and training delivery mode preference. *Journal of Workplace Learning* **14**(1), 5–10.
- Calder, J. and McCollum, A. (1998). *Open and Flexible Learning in Vocational Education and Training*. London: Kogan Page.
- Coffield, F. J., Moseley, D. V., Hall, E., and Ecclestone, K. (2004a). *Learning Styles and Pedagogy in Post 16 Learning*. London: Learning and Skills Research Centre/University of Newcastle upon Tyne.
- Eraut, M. (2000). Non formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology* **70**, 113–136.
- Ford, N. and Chen, S. Y. (2001). Matching/mismatching revisited: An empirical study of learning and teaching styles. *British Journal of Educational Technology* **32**(1), 5–22.
- Fuller, A. and Unwin, L. (2003). Learning as Apprentices in the Contemporary UK Workplace: Creating and managing expansive and restrictive participation. *Journal of Education and Work* **16**(4), 407–426.
- Gropper, G. L. (1983). A behavioral approach to instructional prescription. In Riegeluth, C. M. (ed.) *Instructional-Design Theories and Models: An Overview of Their Current Status*, pp 101–161. Hillsdale, NJ: Erlbaum.
- Honey, P. and Mumford, A. (1992). *The Manual of Learning Styles*. Maidenhead: Peter Honey.
- Kolb, D. A. (1984). *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ: Prentice Hall.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. New York: Cambridge University Press.
- Marton, F. and Säljö, R. (1976). On qualitative differences in learning: I – outcome and process. *British Journal of Educational Psychology* **46**, 4–11.
- Nikolova, I. and Collis, B. (1998). Flexible learning and the design of instruction. *British Journal of Educational Technology* **29**, 59–72.
- Pask, G. (1976). Styles and strategies of learning. *British Journal of Educational Psychology* **46**, 128–148.
- Reynolds, M. (1997). Learning styles, a critique. *Management Learning* **28**(2), 115–133.
- Riding, R. J. and Sadler-Smith, E. (1992). Type of instructional material, cognitive style and learning performance. *Educational Studies* **18**, 323–340.
- Riding, R. J. and Sadler-Smith, E. (1997). Cognitive style and learning strategies: Some implications for training design. *International Journal of Training and Development* **1**(3), 199–208.
- Riding, R. J. and Watts, M. (1997). The effect of cognitive style on the preferred format of instructional material. *Educational Psychology* **17**, 179–183.
- Sadler-Smith, E., Allinson, C. W., and Hayes, J. (2000). Cognitive style and learning preferences: Some implications for CPD. *Management Learning* **31**, 239–256.
- Sadler-Smith, E. and Riding, R. (1999). Cognitive style and instructional preferences. *Instructional Science* **27**(5), 355–371.
- Sadler-Smith, E. and Smith, P. J. (2004). Strategies for accommodating individual's styles and preferences in flexible learning programmes. *British Journal of Educational Technology* **35**(4), 395–412.
- Smith, P. J. (2000a). Flexible delivery and apprentice training: Preferences problems and challenges. *Journal of Vocational Education and Training* **52**(2), 483–502.
- Smith, P. J. (2000b). Preparedness for flexible delivery among vocational learners. *Distance Education* **21**(1), 29–48.
- Smith, P. J. (2000c). *Preparing for Flexible Delivery in Industry: Learners and Their Workplaces*. Unpublished doctoral dissertation, Deakin University, Victoria, Australia.
- Smith, P. J. (2001). Learners and their workplaces: Towards a strategic model of flexible delivery of training in the workplace. *Journal of Vocational Education and Training* **53**(4), 609–628.
- Smith, P. J. (2003). Learning strategies used by apprentices in flexible delivery. *Journal of Vocational Education and Training* **55**(3), 369–383.
- Smith, P. J. and Dalton, J. (2005). *Accommodating Learning Styles: Relevance and Good Practice in VET*. Adelaide: NCVER.
- Unwin, L., et al. (2005). *Learning as Work: Teaching and Learning Processes in the Contemporary Work Organisation*, University of Leicester, The Centre for Labour Market Studies, Learning as Work Research Paper, no. 4.
- Warner, D., Christie, G., and Choy, S. (1998). *The Readiness of the VET Sector for Flexible Delivery Including On-Line Learning*. Brisbane: Australian National Training Authority.
- Wolf, A. and Evans, K. (in press). Workplace learning and adult basic skills. National Research and Development Centre for Adult Literacy and Numeracy, Economic and Social Research Council Teaching and Learning Research Programme.

Further Reading

- Allinson, C. W., Armstrong, S. J., and Hayes, J. (2001). The effect of cognitive style on leader–member exchange: A study of manager-subordinate dyads. *Journal of Occupational and Organizational Psychology* **74**, 201–220.
- Armstrong, S. J., Allinson, C. W., and Hayes, C. (2002). Formal mentoring systems: An examination of the effects of mentor/protégé cognitive styles on the mentoring process. *Journal of Management Studies* **39**(8), 1111–1137.
- Canfield, A. (1980). *Learning Styles Inventory Manual*. Ann Arbor, MI: Humanics Media.
- Coffield, F. J., Moseley, D. V., Hall, E., and Ecclestone, K. (2004b). *Should We Be Using Learning Styles? What Research Has to Say to Practice*. London: Learning and Skills Research Centre/University of Newcastle upon Tyne.

- Desmedt, E. and Valcke, M. (2003). Learning style awareness: Why would it work? In Armstrong, S., Graff, M., Lashley, C., et al. (eds.) *Bridging Theory and Practice. Proceedings of the Eighth Annual European Learning Styles Information Network Conference, University of Hull*, pp 139–150. Hull: University of Hull.
- Dunn, R., Dunn, K., and Price, G. E. (1989). *Learning Styles Inventory*. Lawrence, KS: Price Systems.
- Entwistle, N. J. (1988). *Styles of Learning and Teaching*. London: David Fulton.
- Fuller, A. and Unwin, L. (2002). Developing pedagogies for the contemporary workplace. In Evans, K., Hodgkinson, P., and Unwin, L. (eds.) *Working to Learn: Transforming Learning in the Workplace*, pp 95–111. London: Kogan Page.
- Gardner, H. (1993). *Frames of Mind: The Theory of Multiple Intelligence*. New York: Basic Books.
- Gerber, R., Lankshear, C., Larsson, S., and Larsson, L. (1995). Self-directed learning in a work context. *Education Training* 8, 26–32.
- Gombair, D. (eds.) Learning styles: Reliability and validity. *Proceedings of the 7th Annual European Learning Styles Information Network Conference*, pp 295–301. 26–28 June, Ghent. Ghent: University of Ghent.
- Gregorc, A. F. (1982). *ORGANON: An Adult's Guide to Style*. Columbia, CT: Gregorc Associates.
- Gregorc, A. F. (1982). *Gregorc Style Delineator*. Maynard, MA: Gabriel Systems.
- Hayes, J. and Allinson, C. W. (1996). The implications of learning styles for training and development: A discussion of the matching hypothesis. *British Journal of Management* 7, 63–73.
- Hayes, J. and Allinson, C. W. (1997). Learning styles and training and development in work settings: Lessons from educational research. *Educational Psychology* 17(1/2), 185–193.
- Hills, H. (2003). *Individual Preferences in e-Learning*. Aldershot, UK: Gower.
- Jung, I. (2001). Building a theoretical framework of web-based instruction in the context of distance education. *British Journal of Educational Technology* 32(5), 525–534.
- Löfström, E. (2002). Person–situation interactions in SMEs: A study of cognitive style and sources of job satisfaction. In Valcke, M. and Gombair, D. (eds.) *Learning Styles: Reliability and Validity. Proceedings of the 7th Annual European Learning Styles Information Network Conference*, 26–28 June, Ghent. Ghent: University of Ghent.
- McCarthy, B. (1979). *The 4MAT System: Teaching to Learning Styles with Left/Right Mode Techniques*. Oak Brook, IL: Excel Inc.
- Paivio, A. (1971). Styles and strategies of learning. *British Journal of Educational Psychology* 46, 128–148.
- Rayner, S. and Riding, R. (1997). Towards a categorisation of cognitive styles and learning styles. *Educational Psychology* 17(1/2), 5–28.
- Riding, R. and Cheema, I. (1991). Cognitive styles – an overview and integration. *Educational Psychology* 11, 193–216.
- Riding, R. and Douglas, G. (1993). The effect of cognitive style and mode of presentation on learning performance. *British Journal of Educational Psychology* 63, 297–307.
- Riding, R. and Rayner, S. (1998). *Cognitive Styles and Learning Strategies: Understanding Style Differences in Learning Behaviour*. London: David Fulton.
- Riding, R. J. and Rayner, S. (1998). *Cognitive Styles and Learning Strategies*. London: David Fulton.
- Ross, J. and Schulz, R. (1999). Can computer-aided instruction accommodate all learners equally? *British Journal of Educational Technology* 30(1), 5–24.
- Sadler-Smith, E. (1996). Learning styles and instructional design. *Innovations in Education and Training International* 33, 185–193.
- Sadler-Smith, E. (1996). Learning styles: A holistic approach. *Journal of European Industrial Training* 20, 29–36.
- Sadler-Smith, E. (2001). The relationship between learning style and cognitive style. *Personality and Individual Differences* 30, 609–616.
- Smith, P. J., Robertson, I., and Wakefield, L. (2002). Developing preparedness for flexible delivery of training in enterprises. *Journal of Workplace Learning* 14(6), 222–232.
- Smith, S. M. and Woody, P. C. (2000). Interactive effect of multimedia instruction and learning styles. *Teaching of Psychology* 27(3), 220–223.
- Sternberg, R. J. and Grigorenko, E. L. (2001). A capsule history of theory and research on styles. In Sternberg, R. J. and Zhang, L. -F. (eds.) *Perspectives on Thinking, Learning and Cognitive Styles*. Mahwah, NJ: Lawrence Erlbaum.
- Vermunt, J. D. and Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and Instruction* 9, 257–280.
- Witkin, H. A., Moore, C. A., Goodenough, D. R., and Cox, P. W. (1977). Field dependent and field independent cognitive styles and their educational implications. *Review of Educational Research* 47, 1–64.

Modularization in Vocational Education and Training

H Ertl and G Hayward, University of Oxford, Oxford, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Modularization is regarded as one of the main approaches to create increased flexibility in vocational education and training (VET). Flexibility of structures, contents, aims, and outcomes in VET is, in turn, a major objective of current modernization processes in many countries globally. The main reasons for this period of change in VET are the economic and social megatrends that are influencing economies and consequentially training systems in similar ways worldwide. While these megatrends have certainly been influential in terms of the relevant policy discourse, the extent to which they have actually lead to convergence in training provisions remains to be seen. However, most commentators would agree that megatrends include globalization, internationalization, individualization, new information technology, and the advent of the knowledge economy (Flitner *et al.*, 1999).

Particularly in the 1990s, a far-ranging discussion on more flexible training provision as a consequence of the debate on megatrends ensued. Flexibility in the context of this discussion primarily meant the following:

- The responsiveness of training provision to the changing work environment. This responsiveness is necessary in order to meet the latest skill demands which emphasize comprehensive skills and knowledge structures.
- The responsiveness of training provisions to the varying degrees of personal potential of trainees in the form of individualized training pathways. This individualization is also concerned with increasingly individualized pedagogical approaches and assessment procedures (Sloane, 1997: 231; de Bruijn *et al.*, 1993: 1).

The main argument in favor of modularized structures is that they can potentially be more rapidly updated than the conventional training provision. This curriculum flexibility is also linked to short-term assessment goals and enhanced motivation of learners due to manageable short-term goals and regular feedback. Further, proponents of modularity argue that modules can structure a course of learning to allow for appropriate and progressive development of knowledge and skills.

Despite these pedagogical aims, the debate regarding modularization in VET in the 1990s seemed to be largely driven by the arguments of political actors and, often, conflicting assessments by the social partners, rather than by arguments concerning pedagogical aspects or the processes of teaching and learning (for the German

discourse, see, e.g., the comments by Pütz (1997) and Münk (1995); for a discussion of the English context, see Hayward and McNicoll, 2007). It is, therefore, perhaps not surprising that evaluations of modular qualification structures often come to rather negative conclusions regarding the potential benefits of modular approaches. These negative assessments range from modularized programs resulting in fragmented and commodified learning to mechanical and instrumental approaches to teaching (teaching to the test) to overassessment resulting in examination burnout (learners) and excessive administrative burdens (teachers) (see, e.g., Hayward and McNicholl, 2007).

An analysis of the debate on modularization also shows that the unclear and ambiguous use of the terms module and modularization is widely acknowledged, but still has detrimental effects on both the debate on modularization and its impact on structuring learning processes in VET (cf. Wiegand, 1996a; Heidegger and Schreier, 2006).

Definitions of Modules in Education

On Using a Comparative Approach

Defining the nature of modules in education is no straightforward task. A comparison of selected definitions (Ertl, 2001) suggests that there is no generally accepted explanation of what is meant by organizing the teaching/learning process in modules (Postlethwait, 1985; Schmidt, 1997a; Kloas, 1996, 1997a; European Commission, 2006). Nevertheless, it seems to be generally accepted that any definition can only be valid and appropriate in the normative context within which it has been developed. Therefore, this article focuses on the concepts of modularization developed in two specific national contexts: Germany and England.

An important premise for the comparative approach underlying this article is that systems of education and training are embedded in the cultural environments and traditions of their respective societies. As the discourse in the field of comparative education has demonstrated, it is not advisable to take over arbitrarily selected parts of educational practice from foreign systems to inform the debate in other countries (see, e.g., Noah, 1984; Ochs and Phillips, 2002).

The English and the German contexts were chosen for the development of the conceptualizations because they have developed divergent modular structures in the training sector. It has been argued elsewhere that this is due to

the different origins of modular approaches in the two countries. (It would go beyond the scope of this article to attempt to trace the origins of modularity in VET in the two countries. For details, refer to, e.g., Ertl (2002a), Evans and Kersh (2006), and Hayward and McNicholl (2007).) Research in these two contexts has also substantially contributed to the debate in other countries.

The motives for research in this field differ considerably: Whereas the research in the English context has been mainly aimed at the further development of existing modularized qualifications (such as National Vocational Qualifications (NVQs) and General NVQs (GNVQs)), research in Germany has been mainly preoccupied with the potential of modularity for the modernization of the dual system of training. (GNVQs are currently being phased out and replaced by applied General Certificates of Secondary Education (GCSEs) and A levels. As a result, the uptake of other vocational qualifications has increased.) Modular structures exist in comparatively small areas of VET in Germany, for example, in the qualification of specific target groups, in certain programs for further training and retraining, schemes for additional qualifications in initial training (see Ertl, 2002b), and prevocational training (see Seyfried, 2002, 2003; BIBB, 2006). For instance, the German crafts sector has developed a structure of building blocs aimed at integrating prevocational, initial, and further training (Kloas, 2003; ZHD, 2004).

On Different Definitions of Modules in VET

A description of the way a module in VET is to be constructed and embedded in a wider curricular framework, its intended function, and its pedagogical structure cannot be separated from the wider context of VET.

The terminology in English and in German is confusing. It is unclear in many cases whether the words unit and module are used interchangeably (Watkins, 1987) or whether they mark a difference. A widely accepted English definition of these terms explains a unit as “a coherent set of learning outcomes” and a module as “a sub set of a learning programme” (FEU, 1995). In line with this definition, Stanton (1997: 122) specifies unitization as “the analysis of existing qualifications into units of achievement” and – in contrast – modularization as the process of “[...] subdividing courses into modules of delivery.” In Raffé’s definition, however, this distinction is not as clear: “a module is a unit of delivery and a unit of assessment/credit” (Raffé, 1994).

In the discussion in Germany, both of these meanings are normally ascribed to the term *Modul*. In most cases, a word for the above-mentioned meaning of unit is not even needed as German vocational qualifications are not defined in outcomes; however, the qualifications are regulated by defining the relevant input factors, such as the

duration of the training course or the qualification of the trainer. Nevertheless, a German term for unit could be necessary in the future if the connection between modularity and the outcomes approach (Young, 1998) follows the patterns established in the post-16 education and training sector in England and Wales. The development of multilingual training profiles as a result of the need for more transparent vocational certificates in the single European market toward the end of the 1990s may be regarded as an initial step to supplement the German provision with outcome criteria (Benner, 1997; Herz and Jäger, 1998).

Following the technical origins of the term, there seems to be agreement that a module is a part of a larger entity or system. Modules are distinguished from conventional courses by their shorter duration (Watkins, 1987; Young, 1995; Lauterbach and Grollman, 1998; European Commission, 2006). In this context, a modularized system is a whole composed of a number of self-contained elements. The comparison of over 20 definitions drawn from the relevant literature in England and Germany (see Ertl, 2001) shows that the understanding of what constitutes modules varies decisively in one key point: the relationship between the qualification system as a whole and modules as the elements of this whole.

Some definitions emphasize the self-contained character of modules: they cover “a single conceptual unit of subject matter” (Postlethwait, 1985), “are complete in themselves and examinable as such” (Wiegand, 1996b), and independent from other curricular elements (Deissinger, 1996).

The other group of definitions emphasize that educational modules can only fulfill their envisaged purpose within the context of other parts of the system: they are primarily “part of a whole” (Schmidt, 1997a). The whole in the context of VET may be a qualification in the broadest sense which itself is embedded in “the horizontal and vertical structure of a qualification system” (Reuling and Sauter, 1996). For the attainment of a qualification “a designated number of modules is required” (Theodossin, 1986). In this line of argument “modules only have a logic in terms of the broader qualifications and set of awards of which they form a part” (Nasta, 1994). On their own, modules only provide a qualification “below the level of occupational and professional needs” (Lauterbach and Grollman, 1998). In terms of the curriculum design process, the latter definitions of modules entail a process which starts with considering the qualification as a whole before considering its elements. The learner’s choice is essentially between one course of study or another. If the former approach toward modules is taken, the building of a curriculum starts from the design of the individual modules and the learner chooses a certain combination of modules instead of a discrete qualification.

The contradiction of modules as basically free standing and as only valuable in the context of the wider structure

of a qualification may be overcome by the built-up effect as suggested by Ainley (1990): “a module can stand alone or form a part of a route picked through the various units on offer,” or by Schmidt (1997b): “Combining them [modules] like building blocks, they can form an overall qualification.” This notion seems to be in line with the twofold character attributed to educational modules by Sloane (1997), who not only regards modules as wholes in themselves, but also requires them “to be embedded into a ‘bigger whole.’” In Young’s (1995, 1998) definition of modularization, an existing curriculum seems to represent such a bigger whole and, therefore, the starting point for the module design process. For instance, Nasta (1994) identifies starting a curriculum design process from the whole as the predominant practice in England, whereas starting from the module is widespread in the American system.

Characteristics of Modules in VET

The relationship between modules and qualifications seems to be a crucial aspect for the setup of a modular structure in VET. In the following, this relationship is clarified by deriving the characteristics of modules from a technical example.

The method of induction for defining the functions of modules in education has been applied several times before. For instance, Postlethwait (1985) derives the definition of modules in education from the use of the term in biology and Kloas (1997a) illustrates his definition of modules in VET with the relationship of the elements of a building (modules) and the building as a whole. The method seems to be appropriate as the term module itself originates from technical and biological contexts. In the work by Ertl (2001), the analogy of modules or components in the motor manufacturing industry is utilized to clarify the application of modularization in VET. The following list of characteristics of modules in VET generalizes the analogy:

- *Integrative function.* An overall qualification consists of a combination of modules or part-qualifications. This combination of modules generates the qualification’s function (which might be to develop vocational competence of the individual). This overall function is more than the sum of the single functions of the part-qualifications. Some modules may be utilizable independently from the overall function. This characteristic cannot, however, substitute the overall function of a qualification.
- *Standard definition.* The capability of working competently in an occupational environment is defined by social and economic standards. Some modules are indispensable for basic standards in VET and others improve the individual’s capability to work in a highly sophisticated environment. Accepted standards are the precondition for the combination of modules leading to

recognized qualifications. The standard of modules needs to be broad enough to ensure relevance of modules for more than only one or a few companies and providers of qualifications.

- *Multiple relevance.* Modular systems in VET are effective if modules are relevant to several areas of occupation. However, the variability of the temporal sequence of modules is restricted by pedagogic criteria and inherent content structures.
- *Outcome orientation.* Modules are the result of a qualification process and are only of value if they contribute to the overall function of the qualification. Therefore, modules represent categories of outcome or competence. When new competences are necessary, single modules can be modified, replaced, and adapted without changing the overall qualification.
- *Additional competences.* Modules required for a certain vocation can be augmented by additional modules in order to improve the vocational perspectives of the learner. Accepted standards of modules are the precondition for supplementing the overall function of a qualification reasonably (see also Kloas, 1997a, 1997b).

The underlying assumption for a number of these characteristics is that modules in education can fulfill their envisaged purpose only in the context of an overarching system. This assumption follows the second line of argumentation in the previous subsection. Following the German notion of *Berufskonzept*, the ability of the individual to act and work competently in the occupational environment is the superordinate aim of VET (see Ertl, 2000). As we will see in the following section, there are concepts of modularization in which this aim plays a minor or no role. The next section attempts to identify more general principles of modular strategies by analyzing combinations of module characteristics as applied in different modular approaches. The result of this process is a number of different concepts of modularization.

Before a number of conceptualizations are introduced, it needs to be reiterated that the sociocultural context in which modules are developed seems to determine the status of and relationship between modules and full qualifications. For instance, the notion of what constitutes the occupational environment differs between and, often, also within national contexts. In England, for instance, the primary aim of new modules is often to develop a specific, narrow piece of occupational competence. Recent examples for this are new modules amending a variety of existing qualifications to fulfill new health and safety requirements. This has resulted in the creation of comparatively small health and safety modules which were then conceptualized as independent certificates. This exemplifies how the distinction between modules and qualifications can become blurred due to the way an occupational environment is structured.

Concepts of Modularization

Conceptualization of Modularization in the German Context

In many respects, the German debate about modularity seems to be in the early stages. The assessment by Richardson *et al.* (1995a: 22) that “[...] education and training systems which are relatively stable have shown little interest to date in modularity and credit” still applies to the case of Germany. However, there is an increasingly intense debate regarding the modernization of the dual system of initial training. This debate is based on the perceived inadequacies of the dual system for meeting the demands of a rapidly changing economic environment. Modularization is regarded as a promising way of reforming the initial training system by increasing the flexibility and responsiveness of training provisions.

In the debate, it often remains unclear as to what concept of modularization is under scrutiny. The following three concepts are a synopsis of different conceptualizations developed in the German context:

- *Expansion concept.* Modules supplement initial training qualifications to generate additional competences that are typically the subject of further education and training. The overall functions of initial qualifications are expanded. There are two types of organizational implementation: in the consecutive model, the contents of initial and further training remain separated; in the integrative model, the contents of further training are integrated into the initial qualification.
- *Differentiation concept.* Modules are self-contained and can be assessed and credited individually; however, they are only marketable as part of an overall qualification. The framework of the overall qualification regulates the combination of modules. Modules can be the result of restructuring the curricula of existing qualifications. The differentiation concept facilitates the accreditation of prior learning. Modules can be accredited not only toward different vocational qualifications, but also across the academic/vocational divide.
- *Fragmentation concept.* Modules are credited and marketable without the framework of an overall qualification. By combining modules freely, trainees create individualized qualifications that mirror the requirements of a rapidly changing occupational environment. In comparison to the expansion and the differentiation concept, the fragmentation concept offers the greatest opportunities for flexibility and individuality in VET. Unlike the other two concepts, it is not consistent with the German concept of the vocation.

This attempt to conceptualize modularization in VET is based on accounts by Deissinger (1996), Zedler (1996), Sloane (1997), Rützel (1997), and Kloas (1997a, 1997c).

Conceptualization of Modularization in the English Context

In England, the system of NVQs represents a modularized structure of qualifications that has been created to reform an entire existing training and prevocational sector. This task is similar to the role that modularization could play in other countries that have not yet experimented with modularity to a great extent.

In the 1990s, research at the London Institute of Education has distinguished three forms of modularization in England: internal, external, and connective. (The conceptualization presented here draws on the work of the Learning for the Future and Unified Learning projects and, in particular, on the work of Michael Young within these projects. The following accounts provide further details: Richardson *et al.* (1995b), Hodgson and Spours (1997), and Young (1994, 1995, 1998).)

For all three forms, modularization means the breaking up of the curriculum of a qualification into discrete and relatively short phases of learning. Each module can be assessed independently from other modules at the end of the learning experience. In order to facilitate the assessment process, the curriculum is expressed in terms of measurable learning outcomes. This outcome's approach contrasts with conventional curricula which are expressed in terms of inputs (teachers' and trainers' qualifications, class contact hours, training contents, etc.; Koch and Reuling, 1998). Young (1994: 11) identifies three kinds of outcome approaches and provides examples derived from the British education and training system:

1. In the unitized approaches to outcomes, modules (which are mostly termed as units in these approaches) can be assessed on their own and there is no overall curriculum goal. The most important examples of this approach are units of NVQs.
2. In the integrated outcome approaches, outcomes are assessed with regard to an overall curriculum goal, that is, modules are grouped together to form a qualification. These approaches are exemplified in GNVQs (which are currently phased out), and many other (academic and vocational) qualifications in England.
3. In connective outcome approaches, groups of outcomes are related not only to specific qualifications, but also to the individual aspiration of the learner with regard to the curriculum as a whole. Connectivity in this context is regarded by Hodgson and Spours (1997) and Young (1998) as a stepping stone for a “curriculum of the future” that would not only bridge the vocational/academic divide, but also create the basis for the lifelong continuation of the learner's educational pathway.

In theory, all three of these outcome approaches can be combined with the three concepts of modularization in

VET as identified by the research at the London Institute of Education:

- *Internal modularization.* Courses and qualifications are subdivided into modules of delivery. Modularity is only adopted within a particular qualification. Internal modularization is typically combined with integrated outcome approaches to provide a step-by-step approach to qualifications that allows regular assessment of the trainees' progress. There is the opportunity for regular feedback for trainees so that they can build on their strengths over time. Each module is only valid for one specific qualification and not marketable outside this qualification.
- *External modularization.* Modularity is adopted across different qualifications, that is, a particular module can be part of several qualifications. The choice of a particular qualification can be made during the training process as training modules can be taken and accredited toward several qualifications. Credit transfer schemes make this kind of credit accumulation possible, particularly if they are combined with appropriate rules of combination to assemble modules into qualifications. Typically, existing qualifications are restructured, resulting in outcome-oriented provisions. Unitized as well as integrated approaches of outcome-oriented assessment are possible in this concept of modularization. The number and characteristics of overall qualifications are determined by competent bodies. Individual modules are not marketable outside this system of modularized qualifications.
- *Connective modularization.* The disadvantages of restricted forms of modularity are overcome by the adoption of a holistic curriculum approach. Connective modularity combined with connective outcome approaches provides a tool through which trainees can relate specialist vocational knowledge, broad-based general knowledge and skills, and generic skills. This would entail the restructuring of qualifications into a single system based on modules of delivery and on units of assessment. The restructuring would result in a unified curriculum for all young people in the postcompulsory sector. Connective modularization, therefore, would bridge the entrenched division between vocational and academic curricula. Individual modules in such a unified curriculum would be connected not only by an overarching credit framework, but also by value-added models of provisions designed to promote progression and attainment.

Concepts of Modularization: Similarities and Differences

Concepts of modularization are deeply embedded in the relevant national context. Both conceptualizations are

aiming for a more integrated system of training and education. Integration in this context can be regarded in two different ways that determine the debates on and conceptualizations of modularization in England and Wales on the one hand and in Germany on the other.

First, there is the integration of initial and further training. In both contexts, initial training is no longer regarded as a separate phase in young people's lives, but as the first part of a lifelong learning and training process. However, the concept of initial training as a separate educational phase seems to be particularly strong in Germany (Sengenberger, 1987). The traditional, chronological pattern of training comprising vocational orientation and preparation, initial vocational training, and further training and education which has determined the careers of the majority of working people in Germany for decades is rapidly dissolving (Kloas, 2003). The expansion concept in particular recognizes these developments and integrates initial and further training in a systematic way.

Second, there is the integration of vocational training and general education in the postcompulsory sector. The debates in both national contexts reveal the need for the integration of these two sectors; in addition, in both national contexts, certain concepts of modularization are regarded as particularly suitable reform options. Both the differentiation concept in the German context and external modularization in the English context recognize this type of integration as a central aim of modularity. However, in England the academic/vocational divide is certainly more pronounced than in Germany. Whereas in Germany vocational education always includes subjects providing a certain degree of general education, training provisions in England seem to be more focused on occupation-specific knowledge and skills, in some cases only relevant for certain companies. As a consequence, the concept of a unified curriculum determines the discussion on the further development of modular structures in the training provisions (see Hodgson and Spours, 1997; Young, 1998; Tomlinson, 2004). Connective modularization is mainly regarded as a stepping stone on the way to the ultimate goal of a unified curriculum. As the reaction to the Tomlinson report (Tomlinson, 2004) demonstrated, it has proved to be politically impossible in the English context to achieve this goal.

Trends in Modularity in VET

From the investigation of modular approaches in VET in six European countries (in the work by Ertl (2001), the conceptualizations derived from developments in England and Germany (see previous section) are used as an analytical framework for the investigation of modular approaches in France, Scotland, Spain, and The Netherlands) and an

analysis of two international projects on module development (see European Commission (2006) and RRBK (2007)) the following development lines can be identified.

Integration of Educational Sectors

Modular structures are being increasingly adopted in order to integrate formerly separate sectors of education and training. Modular frameworks can cover vocational preparation, initial, and further training. These frameworks connect the training and education of people in work with training provision for full-time students. This means that modularization is being used to bring school-based and work-based training routes closer together. Further, the divide between general and vocational education in the postcompulsory sector can be bridged by establishing modular banks which include both academic and vocational modules. A cohesive system for all post-compulsory education and training may be the outcome of modularization.

Integration Rather Than Fragmentation

In a number of national contexts, a move away from the fragmentation concept can be identified. Increasingly, overall qualifications that regulate combinations of modules have been developed. This represents a change to the modular strategy: free-standing modules are no longer perceived as marketable part-qualifications. Instead, the attainment of more broadly based combinations of modules in the sense of comprehensive overall qualifications becomes the aim of training. Often, trainees' choice of modules has been restricted in order to reduce the number of final overall qualifications.

Introduction of Modularity in Mainstream Training Provisions

In countries in which modularization has not yet been integrated into regular training provision, modularity is increasingly being adopted in sectors outside initial training. In these countries, modular courses are provided for vocational preparation and orientation, for qualifying specific target groups, and in certain further training schemes. It seems that the experience with modularization at the edges of the mainstream VET system has paved the way for a wider application of modularity.

Supranational Development of Modules

Currently, projects funded by the European Union aim at developing modules through international cooperation. The emphasis of these projects is to establish common criteria for the format and structure of modules in VET and thereby developing a framework that assures transparency and quality of modular systems (European Commission, 2006; RRBK, 2007).

These recent developments indicate that modularization remains high on the educational agenda internationally. It needs to be stressed that neither the political nor the academic discourse does necessarily use the term module explicitly to refer to self-contained elements of a learning program. However, the potential of relatively small and units or blocs of learning has had an impact not only in all areas of vocational education, but also in other educational sectors, such as academically oriented secondary education and higher education. The emphasis of the discussion has moved towards internationally recognized modules, often developed by consortia of partners from a number of countries, and the development of a quality assurance framework for modular systems.

See also: Curriculum and Structuralist Sociology: The Theory of Codes and Knowledge Structures; Early Childhood Curriculum and Developmental Theory.

Bibliography

- Ainley, P. (1990). *Vocational Education and Training*. London: Cassell Educational Limited.
- Benner, H. (1997). Entwicklung anerkannter Ausbildungsberufe – Fortschreibung überkommener Regelungen oder Definition zukunftsbezogener Ausbildungsgänge? In Euler, D. and Sloane, P. F. E. (eds.) *Duales System im Umbruch. Eine Bestandsaufnahme der Modernisierungsdebatte*, pp 53–69. Pfaffenweiler, Germany: Centaurus-Verl.-Ges.
- BIBB (Bundesinstitut für Berufsbildung) (2004). *Berufsausbildungsvorbereitung – Entwicklung von Qualifizierungsbausteinen: Beispiele für die Praxis*. Bonn, Germany: BIBB.
- de Bruijn, E., Froissart, C., Tirados, R. M. G., et al. (1993). *Current Issues in Modular Training: An Interview Study with Trainers in Six European Countries*. Edinburgh, UK: Centre for Educational Sociology at the University of Edinburgh.
- Deissinger, T. (1996). Modularisierung der Berufsausbildung. – Eine didaktisch-curriculare Alternative zum "Berufsprinzip"? In Beck, K. and Achtenhagen, F. (eds.) *Berufserziehung im Umbruch. Didaktische Herausforderungen und Ansätze zu ihrer Bewältigung*, pp 189–207. Weinheim, Germany: Beltz.
- Ertl, H. (2000). *Modularisation of Vocational Education in Europe: NVQs and GNVQs as a Model for the Reform of Initial Training Provisions in Germany? Series: Monographs in International Education*. Wallingford: Symposium Books.
- Ertl, H. (2001). *Modularisation in Vocational Education and Training: Conceptualisations and Approaches in European Union Countries. Series: Cadernos PRESTIGE*. Lisbon: Educa.
- Ertl, H. (2002a). *The Role of European Union Programmes and Approaches to Modularisation in Vocational Education: Fragmentation or Integration? Series: Münchner Beiträge zur Wirtschaftspädagogik*. München: Utz Verlag.
- Ertl, H. (2002b). The concept of modularisation in vocational education and training: The debate in Germany and its implication. *Oxford Review of Education* 28(1), 53–73.
- European Commission (2006). Modules for Europe: Approval of modules in prevocational education and training. <http://www.modules-for-europe.eu/index.html> (accessed May 2009).
- FEU (1995). *A Framework for Credit. A Common Framework for Post-14 Education and Training for the Twenty-First Century*. London: FEU.
- Flitner, A., Petry, C., and Richter, I. (eds.) (1999). *Wege aus der Ausbildungskrise. Memorandum des Forums "Jugend – Bildung – Arbeit" mit Untersuchungsergebnissen des Instituts für*

- Arbeitsmarkt-und Berufsforschung der Bundesanstalt für Arbeit. Opladen, Germany: Leske + Budrich.
- Hayward, G. and McNicholl, J. (2007). Modular mayhem? A case study of the development of the A level science curriculum in England. *Assessment and Education* **14**, 335–351.
- Herz, G. and Jäger, A. (1998). Module in der Berufsbildung oder des Kaisers neue Kleider? *Berufsbildung in Wissenschaft und Praxis* **27**(1), 14–20.
- Hodgson, A. and Spours, K. (1997). Modularization and the 14–19 qualifications system. In Hodgson, A. and Spours, K. (eds.) *Daring and Beyond. 14–19 Qualifications, Frameworks and Systems*, pp 105–120. London: Kogan Page.
- Kloas, P-W. (1996). Modulare Weiterbildung im Verbund mit Beschäftigung – Arbeitsmarkt-und bildungspolitische Aspekte eines strittigen Ansatzes. *Berufsbildung in Wissenschaft und Praxis* **25**(1), 39–46.
- Kloas, P-W. (1997a). *Modularisierung in der beruflichen Bildung. Modebegriff, Streitthema oder konstruktiver Ansatz zur Lösung von Zukunftsproblemen*. Berlin: Bundesinstitut für Berufsbildung.
- Kloas, P-W. (1997b). Modulare Berufsausbildung in Deutschland: Streitthema ohne Wirkung oder Perspektive mit Zukunft? *Gewerkschaftliche Bildungspolitik* **11/12-97**, 18–25.
- Kloas, P-W. (1997c). Berufskonzept und Modularisierung in der deutschen Berufsbildung. *Modularisierung der Berufsbildung – Gleichwertigkeit von allgemeiner und beruflicher Bildung. Deutsch-Britisches Seminar zur Berufsbildungspolitik 3–5. Februar 1997 in Berlin*, pp 15–36. Bonn: BMBF.
- Kloas, P-W. (2003). Qualifizierungsbausteine – Eine Chance für Jugendliche mit besonderem Förderungsbedarf. *Handwerk Magazin* 3.
- Koch, R. and Reuling, J. (1998). Public quality control of vocational training in Germany, France and the United Kingdom. *Vocational Education European Journal* **15**(3), 7–12.
- Lauterbach, U. and Grollman, P. (1998). Berufliche Bildung und allgemeine Bildung im Spannungsverhältnis, oder: Die Krise des Dualen Systems. In *Vocational Education and Training in Germany and Sweden. Strategies of Control and Movements of Resistance and Opposition. Series: TNTEE Publications*, vol. 1. Osnabrück: Thematic Network on Teacher Education in Europe.
- Münk, D. (1995). Kein Grund zur Eu(ro)phorie. Anmerkungen zu zentralen berufsbildungspolitischen Kontroversen des “Memorandums der Kommission über die Berufsausbildungspolitik der Gemeinschaft für die 90er Jahre”. *Zeitschrift für Berufs-und Wirtschaftspädagogik* **91**(1), 28–45.
- Nasta, T. (1994). *How to Design a Vocational Curriculum. A Practical Guide for Schools and Colleges*. London: Kogan Page.
- Noah, H. J. (1984). The use and abuse of comparative education. *Comparative Education Review* **28**(4), 550–562.
- Ochs, K. and Phillips, D. (2002). *Towards a Structural Typology of Cross-National Attraction in Education*. Lisbon: Educa.
- Postlethwait, S. N. (1985). Module approach. In Hussen, T. and Postlethwaite, N. T. (eds.) *The International Encyclopaedia of Education*, pp 3398–3400. Oxford: Pergamon Press.
- Pütz, H. (1997). “Modularisierung” – das falsche Thema, ungenügend bearbeitet, zur ungeeigneten Zeit. In Kloas, P-W. (ed.) *Modularisierung in der beruflichen Bildung. Modebegriff, Streitthema oder konstruktiver Ansatz zur Lösung von Zukunftsproblemen*, pp 62–75. Berlin: Bundesinstitut für Berufsbildung.
- Raffe, D. (1994). Notes on modular strategies. In Hodgson, A., Hayton, A., Bell, R., Melliss, N., and Young, M. (eds.) *Modular Strategies and the Hamlyn/CILNTEC Post-16 Unified Curriculum Project. Series: Unified 16+ Curriculum*, vol. 6, pp 33–36. London: Institute of Education.
- Reuling, J. and Sauter, E. (1996). BIBB-Positionen zu aktuellen Herausforderungen in der beruflichen Bildung: V. Modularisierung in der beruflichen Bildung. *Berufsbildung in Wissenschaft und Praxis* **25**(3), 6f.
- Richardson, W., Spours, K., Woolhouse, J., and Young, M. (1995a). *Learning for the Future: Current Developments in Modularity and Credit. Working Paper 5*. London/Warwick: Institute of Education, Post-16 Education Centre, University of London/Centre for Education and Industry, University of Warwick.
- Richardson, W., Spours, K., Woolhouse, J., and Young, M. (1995b). *Learning for the Future: Current 14–19 Education and Training. Working Paper 1*. London/Warwick: Institute of Education, Post-16 Education Centre, University of London/Centre for Education and Industry, University of Warwick.
- RRBK (Rudolf Rempel Business College) (2007). Connect – cultural and organisational network for new European contents in training. <http://www.connect-web.com/rrbk/english/indexe.html> (accessed May 2009).
- Rützel, J. (1997). Reform der beruflichen Bildung durch Modularisierung. *Berufsbildung* **43**, 5–9.
- Schmidt, H. (1997a). ‘Vorwort’ In Kloas, P-W. (ed.) *Modularisierung in der beruflichen Bildung. Modebegriff, Streitthema oder konstruktiver Ansatz zur Lösung von Zukunftsproblemen*, pp 3f. Berlin: Bundesinstitut für Berufsbildung.
- Schmidt, H. (1997b). Module in der Berufsbildung – Teufelszeug oder Ausweg aus der Krise? *Berufsbildung* **43**, 42.
- Sengenberger, W. (1987). *Struktur und Funktionsweise von Arbeitsmärkten. Die Bundesrepublik im internationalen Vergleich*. Frankfurt am: Campus.
- Seyfried, B. (2002). *Qualifizierungsbausteine in der Berufsvorbereitung*. Bonn: BIBB.
- Seyfried, B. (2003). *Special Issue: Berufsausbildungsvorbereitung und Qualifizierungsbausteine. Berufsbildung in Wissenschaft und Praxis*, 21–23.
- Sloane, P. F. E. (1997). Modularisierung in der beruflichen Ausbildung – oder: Die Suche nach dem Ganzen. In Euler, D. and Sloane, P. F. E. (eds.) *Duales System im Umbruch. Eine Bestandsaufnahme der Modernisierungsdebatte*, pp 223–245. Pfaffenweiler: Centaurus-Verl.-Ges.
- Stanton, G. (1997). Unitization: Developing a common language for describing achievement. In Hodgson, A. and Spours, K. (eds.) *Daring and Beyond. 14–19 Qualifications, Frameworks and Systems*, pp 121–134. London: Kogan Page.
- Theodossin, E. (1986). *The Modular Market. Studies in Further Education*. Bristol: Further Education Staff College.
- Tomlinson, M. (2004). *14–19 Curriculum and Qualifications Reform*. London: DfES Publications.
- Watkins, P. (1987). *Modular Approaches to the Secondary Curriculum. Series: SCDC Publications*. York, UK: Longman.
- Wiegand, U. (1996a). Reform des Ordnungsrahmens: Modulsysteme oder Flexibilisierung der Ausbildungsordnungen? In Schlaffke, W. and Weiß, R. (eds.) *Das duale System der Berufsausbildung: Leistung, Qualität und Reformbedarf. Series: Kölner Texte & Thesen*, pp 260–276. Köln, Germany: Hundt Druck.
- Wiegand, U. (1996b). Modules in vocational training. In CEDEFOP (ed.) *Pedagogic Innovation. Vocational Training. European Journal*, No. 7, pp 28–30. Thessaloniki: CEDEFOP.
- Young, M. (1994). Modularisation as a unifying strategy. In Hodgson, A., Hayton, A., Bell, R., Melliss, N., and Young, M. (eds.) *Modular Strategies and the Hamlyn/CILNTEC Post-16 Unified Curriculum Project. Series: Unified 16+ Curriculum*, vol. 6, pp 5–20. London: Institute of Education.
- Young, M. (1995). Modularization and the outcomes approach: Towards a strategy for a curriculum of the future. In Burke, J. W. (ed.) *Outcomes, Learning and the Curriculum. Implications for NVQ’s, GNVQ’s and Other Qualifications*, pp 169–181. London: Falmer.
- Young, M. (1998). *The Curriculum of the Future*. London: Routledge.
- ZDH (Zentralstelle für die Weiterbildung im Handwerk) (2004). *Qualifizierungsbausteine im Handwerk – Grundkonzeption*. Düsseldorf, Germany: ZDH.
- Zedler, R. (1997). Modularisierung: keine Perspektive für die Berufsausbildung. *Berufsbildung* **43**, 42.

Further Reading

- Cornford, I. R. (1997). Ensuring effective learning from modular courses: A cognitive psychology-skill learning perspective. *Journal of Vocational Education and Training* **49**(2), 237–251.
- Dearing, R. (1995). Key Note Address, in Towards the 21st Century with Modular A levels. October 19th 1995, University of London Examinations and Assessment Centre.

- de Bruijn, E. and Howieson, C. (1995). Modular vocational education and training in Scotland and The Netherlands: Between Specificity and Coherence. *Comparative Education* **31**(1), 83–99.
- Gonon, P. (1998). Modularisierung als reflexive Modernisierung. In Euler, D. (ed.) *Berufliches Lernen im Wandel – Konsequenzen für die Lernorte? Dokumentation des 3. Forums Berufsbildungsforschung 1997 an der Friedrich-Alexander-Universität Erlangen-Nürnberg. Series: Beiträge zur Arbeitsmarkt-und Berufsforschung, BeitrAB 214*, pp 305–321. Nürnberg, Germany: Institut für Arbeitsmarkt-und Berufsforschung der Bundesanstalt für Arbeit.
- Gonon, P. (2001). Modularisierung als länderübergreifende und (berufsbildungs-) systemunabhängige Modernisierungsstrategie. In Reinisch, H., Bader, R., and Straka, G. (eds.) *Modernisierung der Berufsbildung in Europa. Neue Befunde der berufs-und wirtschaftspädagogischen Forschung*, pp 183–193. Opladen, Germany: Leske + Budrich.
- Hayton, A. (1994). The role of the module writers' group. In Hodgson, et al. (ed.) pp 29–32.
- Hodgson, A., Hayton, A., Bell, R., Melliss, N., and Young, M. (eds.) (1994). *Modular Strategies and the Hamlyn/CILNTEC Post-16 Unified Curriculum Project. Series: Unified 16+ Curriculum* (vol. 6. London: Institute of Education.
- Hodgson, A. and Spours, K. (2001). *Evaluating Stage 1 of the Hargreaves Review of Curriculum 2000: An Analysis of Teachers' and Students' Views and the Future of the Reform Process*. London: IOE.
- McClune, B. (2001). Modular A-levels – who are the winners and losers? A comparison of lower-sixth and upper-sixth students' performance in linear and modular A-level physics examinations. *Educational Research* **43**(1), 79–89.
- Moon, B. (ed.) (1988). Introducing the modular curriculum to teachers. In *Modular Curriculum*, London: Paul Chapman.
- Pitz, M. (1999). *Modulare Strukturen in der Beruflichen Bildung – Eine Alternative für Deutschland? Eine explorative Studie am Beispiel des schottischen Modulsystems. Series: Wirtschaftspädagogisches Forum*, vol. 9. Markt Schwaben, Germany: Eusl.
- Raffe, D. (1994). Modular strategies for overcoming academic/vocational divisions: Issues arising from the Scottish experience. *Journal of Educational Policy* **9**(2), 141–154.
- Raffe, D., Spours, K., Young, M., and Howieson, C. (1997). *The Unification of Post-Compulsory Education: Towards a Conceptual Framework. Series: Unified Learning Project. Unifying Academic and Vocational Learning: Scottish and English/Welsh Approaches, Working Paper Two*. London: Post-16 Education Centre, Institute of Education and Centre for Educational Sociology, University of Edinburgh.
- Stobart, G. (1995). Modularity in GNVQ and A level, in towards the 21st century with modular A levels. 19th October 1995, University of London Examinations and Assessment Centre.
- Taverner, S. and Wright, M. (1997). Why go modular? A review of modular A-level mathematics. *Educational Research* **39**(1), 104–112.
- Theodossin, E. (1981). The modularisation of English higher education. *Research in Education* **26**, 89–103.
- Warwick, D. (ed.) (1988). *Teaching and Learning through Modules*. Oxford: Basil Blackwell.

Personalized Learning and Vocational Education and Training

I Duckett, Learning and Skills Network, London, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Personalized learning is a teaching and learning approach which is centered on the needs, aptitudes, and interests of individual learners. While it is not a new concept, it has gathered increased popularity in the last 10 years or so in a number of countries, including United Kingdom, United States of America, Australia, and New Zealand. It has been promoted as a key learning approach to prepare young people for the demands of the twenty-first century and the expectations placed on them by society (Bentley, 2005). Nowhere is this more true than in the area of vocational learning (Duckett and Jones, 2006).

In the United Kingdom, personalized learning has been an offshoot of a broader set of reforms aimed at customizing public services to respond specifically to the diverse needs of individuals. In other places personalized learning has evolved from policies aimed at ensuring equality of opportunity for all students. Here it has focused on helping students from minority groups reach their maximum potential.

In this article, we explore the nature and philosophical underpinnings of personalized learning, its evolution as a mainstream learning approach, and its implementation in a number of different countries, mainly the United Kingdom, Australia, United States of America, and New Zealand.

What Is Personalized Learning?

Personalized learning is a distinct student-centered approach to learning that is increasingly being used to ensure that students are able to meet their goals and their potential. It has been a key element of recent and planned educational reform in the United Kingdom, United States, Australia, and New Zealand. The philosophy underpinning this approach is a belief in the ability of every learner to succeed, communicating this to each individual learner to unlock their potential, and providing them with the resources that will enable them to succeed. Support for its application is based on the findings of research studies which indicate that personalization of learning and assessment results in improved scholastic attainment, thinking skills, personal development, and self-confidence for learners.

Although by definition personalized learning is necessarily student centered, we should not confuse personalized learning with the child-centered learning approaches

of the past which catered mainly to individual interests (OECD, 2006). Personalized learning is also about raising standards by taking account of both personal interests and aptitudes. It is about providing a safe environment for learning and coordinating learning support structures to help students maximize achievement toward a set of negotiated targets. Nevertheless, knowing learners' personal interests continues to be of vital importance in engaging the learner and improving motivation.

Personalized learning is not about abandoning curriculum. "At the core there would still be a common script – the basic curriculum – but that script could branch out in many different ways, to have many styles and endings" (Leadbeater, 2004: 14). A commitment to providing opportunities for students to participate in making decisions about their learning is another key component. This is aimed at equalizing opportunity for learning. However, giving learners (and their parents) a voice in deciding the nature and delivery of education will only equalize opportunity if learners from minority groups have the resources to be able to make such choices. This means that for some groups, the more the "services become personalized" the more the public resources will have to be skewed toward the least well-off (Leadbeater, 2004: 22).

A Culture Shift for the Educational Establishment

"Personalized learning is a highly structured approach that places the needs, interests and learning styles of students at the centre" (Kearney *et al.*, 2007). It requires a change in how educational establishments view and behave toward learners. It requires them to develop a professional ethos that recognizes the importance of listening to learners to ensure that they understand and take account of what learners are saying. They must also accept that learners have unique aspirations, different types and levels of skills, and learn in different ways. Personalized learning requires a commitment to working jointly with learners to identify their personal learning needs. Identifying the way a particular learner learns is crucial to the process. If personalized learning is to be effectively implemented in schools, then educational establishments need also to be committed to supporting teachers through appropriate and effective continuing professional development. Personalizing education also requires education systems to have a commitment to lifelong learning and to provide learners with flexible

learning environments. It is not about allowing learners to do their own thing, or permitting them to coast at their own pace.

Personalized learning should not be confused with individuals sitting alone learning and working through assignments or self-paced programs – a mode that may suit some learners but certainly not all. Personalized learning is about schools promoting a “community of learning” approach which “cultivates strong relationships . . . between home, community, local institutions, businesses and services” (Kearney *et al.*, 2007: 3).

A changed role for teachers

Personalized learning changes the traditional role of teachers. They move their focus from teaching students as an amorphous group of learners toward teaching students according to their individual aptitudes and interests. They become facilitators and learning brokers helping students make the most suitable choices. In doing so, they will also need to respond to the different ways learners achieve their best. An initial step is for teachers to identify individual needs, aptitudes, and preferred learning styles. This will enable them to tailor their teaching and assessment practice so that every learner has the opportunity to achieve the highest possible standard. In personalized learning, teachers set individual learning targets and regularly monitor and review progress toward these. They provide learners with structured feedback and teach them how to use it to improve their work. They also help students learn how to learn by providing them with advice on how to organize their learning and apply study skills which have been found to be effective. They encourage learners to engage in critical self-assessment. In personalized learning, it is crucial for teachers to have high expectations for all learners regardless of background and ability levels and monitor learning needs and progress through on-going dialog and formative assessment. Personalized learning requires students to use information on student progress to modify their learning.

It is sometimes a challenge to keep students interested and motivated (Duckett and Jones, 2006) even when learners have chosen a course of study. This is where personal knowledge of individual learners comes in to play. Knowing which buttons to press for each individual can provide the ignition to motivate and inspire any learner. Once teachers know their learners they are 90% toward raising achievement levels. Which button will differ from learner to learner as the very uniqueness of each individual will form the basis for their personal determination and motivation. Learners crave realism. Making learning real, by linking it to the outside world, can inspire and increase motivation.

A changed role for learners

Personalized learning also changes the role of learners from being passive recipients of predetermined programs

to becoming active agents and problem solvers. Negotiating learning which suits their particular talents or shortcomings not only improves motivation for learners but also helps them to maximize their potential. This gives them a sense of power and fulfillment which in turn helps to increase confidence and self-esteem. However, not all learners will be able to take the same advantage of this increased choice and these learners will have to be supported by knowledgeable and sensitive teachers and counsellors.

Different players in the learning process may share the same or similar goals for learning, but the support individual learners require to achieve these goals may vary. What is fundamentally different in personalized learning is that learning becomes a shared process with learners actively in control. This has clear implications for how students learn how to learn.

A changed meaning for assessment

Assessment in personalized learning is not only about giving learners a test of their achievement at the end of a program of study. It is also about ensuring that learner's goals, aspirations, and aptitudes are analyzed at the beginning. At this time, it is important to diagnose precise learning difficulties that may require additional learning support. Other initial assessments include specific standardized diagnostic tests which in some countries (e.g., the United Kingdom) are used to identify and address any physical disabilities (dyslexia, and vision and hearing impairments) in order to comply with legislation which aims to ensure that these learners have access to required support or modifications. For all learners, the results of such initial assessments are used to help teachers develop appropriate learning approaches, materials, and assignments. Formative assessment processes are used by teachers and students to help students review and understand what they have been learning and how they can improve. Summative assessments are used to determine the extent to which negotiated targets are met.

A key role for information and communications technology

Information and communications technology is a key enabler of personalized learning. It can help systems respond to the needs of learners by providing a personalized space for students to do their learning. This type of learning is not confined to the classroom and allows the learner “to live locally whilst learning globally” (Kearney *et al.*, 2007: 2). The use of technology to undertake learning at home, at work, and at school has also been promoted in the Australian VET system as a key mechanism for ensuring that training remains client-focused and makes use of modern technologies. This generally goes under the banner of flexible delivery.

Personalized Learning in the United Kingdom

Personalization and personalized learning are much in the UK education news. In a conference of educators in January 2004, then Minister of State for School Standards, David Miliband, MP, spoke about the government's aim to put individuals of this country at the center of all public services, including education. This would enable each individual to have a say in the design and improvement of the organization he serves. At the heart of Miliband's speech was his belief in the ability of personalized learning to promote equality and social justice. Since Miliband's key speech, the term personalized learning or the concept of personalization has been repeated in many education reports and in educational discourse.

The Department for Education and Skills (DfES) supports the introduction of the personalized learning approach by appealing to the findings of research on teaching and learning. This includes studies which indicate a link between key principles and improved student performance in scholastic attainment, thinking skills, and self-esteem.

However, implementation is still in its early stages. For instance, in the report 14–19 curriculum and qualifications reform (Tomlinson, 2004) the working group on 14–19 reform highlights observations mirrored in many Office for Standards in Education (Ofsted) and Adult Learning Inspectorate (ALI) inspection reports and puts forward recommendations for personalized learning including:

- stretching and challenging able learners;
- raising learners' personal awareness;
- increasing understanding of learners' strengths; and
- identifying learning and development needs.

A key driver for change toward personalized learning in the United Kingdom is the need to ensure that schooling provides learners with the skills and attributes they will require for the twenty-first century and to meet community expectations (Bentley, 2005). The introduction of student profiles in the United Kingdom schools has gone some way to personalize the data that schools have about individual students. Riel and Bentley (2006) are of the view that this could be extended by treating school facilities as "entry points to much wider networks of flexible, specialist, provision and participation" (p. 122).

Recently, the Learning Skills Council has endorsed the Recognising and Recording Progress and Achievement (RARPA) project. Although RARPA was initially designed for staff working in nonaccredited learning, it has many lessons for practitioners working across the range of accredited and non-accredited programs (Duckett and Brooke, 2006).

The five stages of the RARPA process are:

1. setting the aim of the program of learning, which may define its general purpose, or may be more specific (curriculum design);
2. initial assessment of the learner, to find out the learner's needs and aspirations; this is important as it enables learners to express their desires, interests, motivations and support needs; if undertaken during induction, this stage can promote a greater understanding of the purpose of the program and enable learners and teachers to get to know each other; importantly, in the process of recording progression, it can establish where the starting-point is for the learner and identify what he or she wants to achieve by undertaking learning;
3. identification of learning objectives that meet the stated learning outcomes of the program of learning for the learner group and the individual learner;
4. formative assessment, or the ways in which teachers identify and record progress made by the learners; this may be a very creative process with suggestions of video recordings, learner diaries, exhibitions, as well as more traditional assignments and tests; and
5. end-of-program review of progress and achievement, which may involve learner or peer-group assessment, teacher's record of assessment, and a whole range of artifacts as appropriate to the program of study (summative assessment).

Australia

The term personalized learning is also gradually gaining currency in Australia. However, the concept is often seen as an extension of existing processes for placing students at the center of teaching and learning. Always, the main aim is to enable students to develop the skills, knowledge, and personal attributes required for acquiring qualifications, job skills, and employment.

For example, personalized learning is promoted in the South Australian Government's strategy for reforming senior secondary education, and the South Australian Certificate of Education (SACE). Here, the key aims are to provide students with engaging curriculum and to help them make successful school-to-work transitions. Teachers are asked to work with individual students to take a realistic account of their aptitudes, interests, career goals, and other opportunities and to develop personal learning and transition plans. Schools are asked to develop strong partnerships with the local community, business, industry, and education and training providers to enable students to gain valuable and challenging experience with work, training, or study. A student mentoring program has been established to provide students (generally those

at risk of dropping out) with support and advice from teacher mentors who have been trained to support the learner's well-being, achievement, and pathway planning. In this context, the personalized learning plan which comprises the setting of targets against key learning outcomes is seen as especially important. Currently, about 29 secondary schools are engaged in pilot projects to trial these reforms.

Personalized learning has been especially promoted in the education of indigenous students who have historically not engaged with or benefited from schooling to the same extent as other Australians (Ministerial Council for Education, Employment, Training and Youth Affairs, 2006).

The personalized learning approach is also currently being applied in the Australian Science and Mathematics School, a senior secondary school located within Flinders University in South Australia. Here, it is characterized by the development of personalized learning plans for individual students, and a belief in challenging and rigorous programs using inquiry-based and experiential learning strategies. Personalized learning plans (gradually being developed for all students) involve tutors and students in discussions about what students believe are their strengths and weaknesses and their short and longer-term goals. These are then used to identify strategies to help students achieve these goals. A variety of online tools are also used to help students think about learning preferences and career options. Some students have used electronic portfolios to document their work and their achievements and their plans for the future. These have been found to be successful thus far.

It is also important that we are aware that many of the aspects of personalized learning are exemplified in other approaches to customizing education to the needs of individuals. The concept is also reflected in the flexible delivery initiatives first adopted in Australia to reform the public vocational education and training (VET) system in the early 1990s. The aim of these approaches is to ensure that the training system remains client-focused by enabling students to choose how, when, where, and what they learn. Self-paced learning programs customized the pace of learning and assessments to student interests and aptitudes, to assist them in becoming self-directed and independent learners. This aims to prepare students for workplace of the future. Teachers and trainers adopt a learning facilitator role and work with students individually and in groups to respond to questions, introduce topics, or discuss issues. Computer-assisted learning and other information and communications technologies play key roles in the development of technical skills and in monitoring the progress of students.

A continuing belief in every student's capacity for learning is also exemplified in the 2007 statement on the

future of schools in Australia. Endorsed by the Commonwealth and State and Territory government ministers who have portfolios in the areas of education, employment training, and youth affairs, this statement sets out broad directions for how schools, parents, and communities can assist learners to develop fully their intellectual, physical, moral, and aesthetic talents and capacities in school and throughout life. A commitment to students is a key goal.

Personalized Learning in the United States of America (USA)

The use of the term personalized learning is still in its infancy in the USA. Kearny *et al.*, (2006) are of the view that the term was first used in the 1999 Breaking Ranks report which highlighted the need for personalization through the establishment of smaller schools and the use of a variety of instructional methods. Today a follow-up report Breaking Ranks II developed at the Education Alliance at Brown University stresses the need for personalization. It recommends that schools ensure that personal learning plans and a personal adult advocate for each student, recognize that each student learns differently, and provide support and guidance for each student to set, review, and achieve his or her goals. A personal learning plan for each education is also recommended so that they continue to improve their own skills and knowledge in helping students in the learning process.

Personalized learning as a learning philosophy has also been especially adopted by 35 schools established by the Big Picture Company in conjunction with state educational authorities in a number of different states. The motto one student at a time has been adopted by the 35 schools involved and exemplified by the following mission statement (Big Picture Learning, <http://www.bigpicture.org>).

Each student's learning plan should grow out of his or her unique needs, interests, and passions. We believe that the education system must ensure that students and families are active participants in the design and authentic assessment of each child's learning. Schools must be small enough to encourage the development of a community of learners, and to allow for each child to be known well by at least one adult. School staff and leaders must be visionaries and lifelong learners. Schools must connect students, and the school, to the community—both by sending students out to learn from mentors in the real world, and by allowing the school itself to serve as an asset to the local community and its needs. Finally, schools must allow for admission to, and success in, college to be a reality for every student, and work closely with students, families, and colleges throughout – and beyond—the application process.

Historically, the need to customize learning to the needs and aptitudes of students in mainstream American schools has generally been associated with equalizing educational opportunity for all students and especially those from disadvantaged groups. The No Child Left Behind initiative of the Bush administration which resulted in amendments to the 1965 Elementary and Secondary Education Act aims to “ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging State academic achievement standards and state academic assessments” (Sec. 101, Title 1, amendment to Elementary and Secondary Education Act 1965 (20 U.S.C 6301 *et seq.*).

Other initiatives to improve the lot of disadvantaged students include early childhood programs like the Head-start program which aims to improve the school readiness of pre-schoolers from disadvantaged backgrounds. A key aim is to develop reading and math skills of these pre-schoolers so that they can experience success in school. The program provides children and their parents with educational, health, nutritional, and other services. In 1995, the program was extended to ensure that children from birth to 3 years of age were also engaged. Parents are involved in the administration of their local program and in their children’s learning. They are also assisted to develop and meet their own educational, literacy, and employment goals.

New Zealand

The Ministry of Education of New Zealand has launched its vision for personalizing learning (Ministry of Education, 2007). It notes the need for the country to transform its education system to meet the challenges of the twenty-first century. In doing so it emphasizes its mission to no longer tolerate failure for any student.

Personalized learning in New Zealand is viewed not only as a strategy to place students at the heart of the learning system, but also about making learning more meaningful. It is felt to succeed best when students know what they know, how they know it, and what they need to learn next. Two key components of the New Zealand approach are effective teaching and appropriate assessments. Teachers will be expected to appreciate the capacity for learning of every student and to monitor progress for future learning. They are also expected to help students develop their skills of working independently and in groups, and support each other in learning. To do this, they will develop a wide range of effective learning strategies including using new technologies to support student learning. Initial needs assessments will be conducted and the results used to determine the strengths and weaknesses of each student and to develop suitable

learning programs. Timely and effective feedback that improves learning outcomes and engages students in learning is also required. The “Assess to Learn” project and assessment tools for teaching and learning (aTTle) have been developed to help teachers, students, and parents with information which accurately records student achievement and progress.

In turn, the Ministry of Education promises to play its part in providing the resources, conditions, and policies to support these changes. Personalized learning will also be supported by the New Zealand Curriculum which will provide direction for what is to be taught and opportunities for teachers to better structure learning so that it meets the needs of individual learners. It will require strong partnerships between schools and parents to provide teachers with the information they require to better structure learning to suit the individual needs and aspirations of students. The “Team Up” program established to facilitate this partnership will also help parents understand how they can help in their children’s learning. The professional leadership skills of principals are also critical in implementing personalized learning. A First Time Principals program has been established to train new principals for their roles. A Kiwi Leadership Framework is also being drawn up to reflect the country’s unique educational culture and environment.

The Te kotahitanga program is an example of a program which implements a key principle of personalized learning, (that is, the need for teachers to understand the backgrounds and strengths of students and build supportive relationships with students), to improve school retention and achievement of indigenous students.

Issues around Personalization

Personalized learning locates students at the heart of the learning process by not only giving them a voice in determining their own learning, but also ensuring that they are provided with suitable guidance and targets to meet their goals. Research conducted by the Quality Improvement Agency (QIA) in the United Kingdom has highlighted some issues of controversy, which largely focus on learner voice and provider guidance (LSN, 2006). It is clear that the issue of learner voice, and the extent to which it plays a role in the delivery of personalization, needs to be resolved.

The major issue is that of learner voice, that is, how much learner control is appropriate (see e.g., QIA, 2006). A better formulation is how we enable the learner to take more control. It is the transition from an inexperienced to a self-directed learner which is key and which coincides with the concept of the expert learner. Clearly, becoming an expert learner, and being able to take control of their learning, are two very similar skills which come with age

and experience. In short, although the issue of learner participation seems to be a significant point of debate and potential conflict amongst the sector, this can be resolved by taking a more flexible, developmental approach. We know that expert learner skills can be developed over time – so it is perfectly feasible that the (very similar) skills required for a learner to make responsible and informed choices about his or her learning can be developed too.

Conclusion: Personalized and Individualized Learning

Individualizing learning is something which learning and skills providers have been carrying out for many years. It is much more than offering a broad range of curriculum choices and flexibilities – it is also about adapting teaching and learning strategies to meet the needs and abilities of different learners. This may include (and often does) individualized learning plans, individual targets, group and one-to-one work, and different forms of pedagogy to suit different learners. We have used the term personalized learning to distinguish it from the student-centered learning approaches of the past; however, we must also be careful to make sure that we do not invent another term to describe activities that are essentially similar in scope, intent, and application.

Bibliography

- Bentley, T. (2005). *Everyday Democracy*. London: Demos.
- Duckett, I. and Brooke, D. (2006). *New Measures of Success: Recognising and Recording Progress and Achievement*. London: Learning and Skills Network.
- Duckett, I. and Jones, C. A. (2006). *Personalized Learning: Meeting Individual Learner Needs*. London: Learning and Skills Network.
- Kearney, K., Nicholas, H., Mahar, S., and Herrick, C. (2007). Personalizing education: From research to policy and practice. *Paper No. 11*. September, Office for Education Policy and Innovation, State Government of Victoria, Department of Education and Early Childhood Development, Melbourne.
- Leadbeater, C. (2004). *Learning about Personalisation: How Can We Put the Learner at the Heart of an Educational System?* London: Demos.
- Miller, R. and Bentley, T. (2006). *Unique Creations: Possible Futures – Four Scenarios for 21st Century Schooling*. Nottingham: NCSL.
- Ministerial Council on Education, Employment, Training and Youth (2006). *National Protocols for Higher Education*. <http://www.curriculum.edu.au/mceetya/> (accessed 2 June 2009).
- Ministry of Education (2007). *Personalizing Learning*. http://www.tki.org.nz/r/personalizing_learning/pdf/personalizing-learning.pdf (accessed May 2009).
- OECD (2006). *Education at a Glance*. Paris: OECD.
- Tomlinson, M. (2004). *14–19 Curriculum and Qualifications Reform: Final Report of the Working Group on 14–19 Reform*. <http://www.dfes.gov.uk/14-19/documents/Final%20Report.pdf> (accessed May 2009).

Further Reading

- Coffield, F., Moseley, D., Hall, E., and Ecclestone, K. (2004a). *Learning Styles and Pedagogy in Post-16 Learning: A Systematic and Critical Review*. London: Learning and Skills Research Centre.
- Coffield, F., Moseley, D., Hall, E., and Ecclestone, K. (2004b). *Should We Be Using Learning Styles? What Research Has to Say to Practice*. London: Learning and Skills Research Centre.
- DES (1978). *Special educational needs: Report of the Committee of Enquiry into the Education of Handicapped Children and Young People. Warnock Report. Cmnd 7212*. London: Department of Education and Science.
- Duckett, I. (2006). What is personalized learning? In *14–19 News. Autumn 2006*, Learning and Skills Network.
- Duckett, I. (2007). “Literacy, employability and vocational learning” *Literacy and Numeracy – Meeting Targets, Raising Standards*, pp 59–61. London: Westminster Education.
- Duckett, I., Goulbourne, A., Hull, L., and Mattick, J. (2005). *Partnerships that Work*. London: Department for Education and Skills.
- Duckett, I., Vickers, D., and Baddeley, S. (2007). *14–19: An Overview*. London: Learning and Skills Network.
- Jones, C. A. (2005). *Assessment for Learning*. London: Learning and Skills Development Agency.
- Jones, C. A. and Duckett, I. (2004). *Tutorials and Target-Setting in the Effective Delivery of Vocational A-Levels*. London: Learning and Skills Development Agency.
- Jones, C. A. and Duckett, I. (2005). *Putting Learning at the Centre of the New A-Levels*. London: Learning and Skills Development Agency.
- Learning and Skills Network ([Duckett, I. and Jones, C.] 2006). *Personalized Learning Quick Guides*. (Personalized learning: Tutorials and peer mentoring; Personalized learning: Learner empowerment; Personalized learning: Planning for learning; Personalized learning: Information, advice and guidance; and Personalized learning: Assessment for learning). London: Learning and Skills Network. Coventry: Quality Improvement Agency.
- McLean, N. (2005). *Personalized Learning: ICT Enabling Universal Access*. London: Qualifications and Curriculum Authority. <http://www.qca.org.uk/futures> (accessed May 2009).
- Ofsted (2005). *HMI 2434: The Common Inspection Framework for Inspecting Education and Training*. London: Ofsted.
- Tomlinson, M. (2003). *Working Group on 14–19 Reform: Principles for Reform of 14–19 Learning Programs and Qualifications*. <http://www.14-19reform.gov.uk> (accessed May 2009).
- <http://www.bigpicture.org> – Big Picture Learning.
- <http://www.decs.sa.gov.au> – Department of Education and Children’s Services, Government of Australia.
- <http://www.standards.dfes.gov.uk> – Department for Education and Skills, Standards Site, Government of UK.
- <http://www.nassp.org> – National Association of Secondary School Principals.
- <http://edlinked.soe.waikato.ac.nz> – School of Education, University of Waikato, New Zealand.
- <http://www.acf.hhs.gov> – US Department for Health and Human Services, Administration for Children and Families.
- <http://www.youthengagement.sa.edu.au> – Youth Mentoring Networks, Australia’s Mentoring Programs.

Recognition of Prior Learning and Vocational Education and Training

J Keating, University of Melbourne, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The idea that learning that takes place outside of formal education courses can and should be recognized within formal qualifications or their components is relatively widespread and has been around for some time. The received history of the first example of the recognition of prior learning (RPL) is that of US servicemen returning at the end of World War II who sought formal recognition of their learning gained from their service experiences (Harris, 2006). (RPL is practiced under different names across different countries. The various forms of assessment and recognition of informal and prior learning are all referred to as RPL in this article.)

The second iteration in the form of formal systems and more widespread innovations in the implementation of RPL was concentrated in Anglophone countries, notably the United Kingdom, Australia, New Zealand, Canada, parts of the USA, and South Africa. These countries have had some common features, apart from language. All separate their initial vocational education and training (VET) systems from their secondary school sectors and none have had the strong social partner participation in the steering and management of their VET sectors that is common in Europe. In most of them, liberal market orientation economic reforms were introduced in the 1980s or 1990s, together with the devolution and decentralization of governance in education and training. In several of these countries, forms of competency-based training (CBT) and associated forms of outcomes-based qualifications evolved with the spread of RPL.

It now seems that RPL in its latest iteration of recognition of informal learning is entering a third age as the concept is championed by international organizations including the Organization for Economic Cooperation and Development (OECD) and the International Labor Organization (ILO) and spreads to other countries, especially those within the European Union (EU).

The RPL Claim

From the basic objective of the US servicemen to have their learning formally recognized, the RPL journey has been a mixture of robust, ambitious and at times extravagant claims, multiple innovations, and diverse practice. The RPL journey has been most intense over the past two

decades and recently has traveled beside major institutional changes in the structures and processes for the formal recognition of learning.

The period of the evolution of RPL has been that of the massification of tertiary education, with a number of countries achieving levels of 50% or more of school leavers entering tertiary education and the EU adopting its Lisbon target of 50% participation. The period has also seen the evolution of the human capital agenda across all developed countries and an increase in public and private investment in postschool education and training.

The RPL story is immersed within these developments. Across its phases it is an innovation that has expressed democratic ideals of education, challenged the construct of formal knowledge and teaching practices within education systems, initiated major systemic and cross-national dialog and formed a foundation element of systemic changes in the characteristics of formal education and training systems and especially qualifications systems.

The basic principles, let alone the more extravagant rhetoric of its champions, have not gone unchallenged, nor have its most conducive curriculum forms of CBT and outcomes-based education and training and their epistemological underpinnings. Given its now considerable history in the Anglophone countries, the realization of RPL could be judged as disappointing (Cameron and Millar, 2004; Pokorny, 2006).

Yet, RPL in its latest iteration of the recognition of informal learning has attracted even more ambitious claims. The EU has endorsed the "principle of learning throughout life, the identification and validation of non-formal and informal learning aim to make visible and to value the full range of knowledge and competences held by an individual, irrespective of where or how these have been acquired" (EU, 2004:1). The ILO has several conventions that endorse RPL and the right of workers to have their skills formally recognized. The OECD has launched a new activity on the recognition of nonformal and informal learning.

By the late 1990s, RPL had traveled to the Scandinavian countries, which, perhaps not coincidentally, had adopted more liberal and decentralized governance arrangements for education and training earlier in the decade. It is now being addressed by countries such as Germany where stakeholder organizations have until quite recently taken a cautious view toward the recognition of informal learning. The extraordinary spread of national qualifications frameworks

(NQFs) across Asian, Middle Eastern, Southern African, and Latin American countries, as well as Europe, is likely to signal a further spread of RPL and its underpinning principles.

So what is the extent of the RPL claim, and what has been realized through the concept and the myriad systemic and localized innovations and practices? Is the RPL claim legitimate, and does it have legitimacy within systems of qualifications that have been designed for formal learning systems? Does it enhance or threaten the practices of education, especially within its construct through the subject disciplines?

Drivers and Constituents

These are large questions, and it is perhaps unfair to suggest that the RPL movement and its champions have made such extravagant challenges. However, the RPL claim is significant. For example, a recent EURYDICE (2007) document describes RPL in the following terms:

“RPL ... enables access, transfer and progression. It can be instrumental in motivating non-engaged adults to resume or continue learning as it identifies and makes visible knowledge, skills and competence of which the adult may not have been aware and, if a national framework of qualifications exists, it places the individual at the appropriate level on the framework. It enables existing knowledge, skills and competence to be formally recognised, rewarded and signalled to different stakeholders, including employers, thus increasing economic returns. It reduces opportunity costs through eliminating or reducing the need to spend time and money relearning what has already been learned” (p. 1).

These various perceived benefits of RPL relate to the rights of workers to have their skills recognized, mechanisms that encourage individuals to access and engage with formal learning, the capacity to align the levels of knowledge and skills for the purpose of learning progression, increased productivity and economic returns, and resource efficiencies within education and training systems. RPL therefore brings a range of constituencies.

In countries such as Australia, much of the early impetus came from unions that saw RPL as a means of securing formal recognition of their members' skills within industry-based wage systems. In South Africa this extended to the wider non-white population that had been largely excluded from both the formal waged economy and the postelementary education system (Cooper, 1998). Post-apartheid, the position of the Congress of South African Trade Unions on RPL, was readily adopted by the African National Congress as a means for the non-white population to help overcome and redress exclusion. The strength of this constituency and its expectations of the principle of RPL extended to the

decision to move the entire education and training systems to an outcomes-based approach, where sets of unit standards were developed independently of established qualifications and educational institutions (Lugg *et al.*, 1998).

At the same time, decisions were taken by the New Zealand government to establish an NQF and a qualifications system based upon unit standards. These developments occurred in a different political context to that of South Africa. By the early 1990s, New Zealand, arguably more than any other OECD country, had embraced neoliberal economic reforms with widespread economic deregulation and privatization and the injection of market principles into most areas of the economy. The idea of a demand-led education and training system and more open education and especially training markets connected with the idea that learning can take place in contexts outside of formal education and training institutions. The newly established New Zealand Qualifications Authority (NZQA) began a program of building a system of standards-based qualifications, which in principle were intended to include all education sectors. Thus, in 1990, the government passed legislation that empowered the NZQA to “develop a framework for national qualifications in secondary schools and in post-school education and training in which:

- (i) All qualifications have a purpose and a relationship to each other that students and the public can understand; and
- (ii) There is a flexible system for the gaining of qualifications, with recognition of competency already achieved.” (cited in Blackmur, 2003).

The common experiences of the two countries juxtaposed against radically different political landscapes were mirrored in Australia and the UK, albeit less starkly. A Labor Government in Australia initiated a national training reform based upon competency standards. The reform program was informed by developments in the UK, but much of the push had come from unions, and notably the metals industry union (as was the case in South Africa). The VET qualifications that have evolved are packages of units of competencies, which, in principle, can be gained by individuals from a range of learning experiences and contexts and thus can be assessed and recognized independently from a formal training course. In the UK, a competency-based VET system had been introduced more or less in parallel with the Australian system under the Conservative Government that had introduced deregulation and privatization to wide areas of the economy, including the public sector. The national vocational qualifications (NVQs) that have evolved are based upon the principle that their specified knowledge and skills can be gained and therefore should be able to be recognized outside of formal learning programs.

Thus, constituents for RPL included a spread of unions, employers, and governments of different political hues. However, RPL's philosophical underpinnings also are diverse. On the one hand, it is seen as a democratic principle of the right of people who have been disadvantaged within or excluded from formal education and training systems to have their knowledge and skilled formally recognized for the purposes of re-accessing educational programs and for employment and social recognition. On the other, RPL and its base in learning outcomes that can be independent from formal education programs is a means of breaking down supply-side education and training systems and reducing provider capture of the education market. Furthermore, outcomes-based education and training systems, when linked to funding systems, can reduce the need for central administration but increase the accountability of education and training systems and increase the capacity for policy steerage.

The assembly of these constituents, principles, and reforms in RPL took place mainly in the early 1990s. They were confined largely to the VET sector, and across the Anglophone countries the drawbridges into the higher education and school sectors were drawn. Allais (2007) has shown that the unit standards of the SAQF have had almost no impact on the higher education and schools sectors. The resistance of the universities contributed to the near demise of the NZQA (Strathdee, 2003). The universities in Australia rejected outcomes- or competency-based curriculum and qualifications (Bowden and Masters, 1993). In the UK, there was little attempt to assault the bastions of the universities and the prestigious senior secondary qualifications, the English (and Welsh) A Levels and the Scottish Highers. The RPL movement lacked both the constituents and the cultural and political capital to tackle the universities and the senior secondary certificates that have a symbiotic relationship with university education.

Through what we have described as a second phase of RPL, there have been a series of settlements within the Anglophone education systems that largely confine RPL to the VET sector. The confinement has not been total, as many, if not most, universities have formal policies in support of RPL, allow some access to postgraduate courses on the basis of professional experience, and have given *de facto* RPL through the establishment of professional higher degrees that can be completed over shorter time period through reduced credit points. However, in most countries where RPL has been such a deliberate movement, endorsed by national and regional governments – the UK, South Africa, New Zealand, and Australia – and institutionalized within outcomes-based VET systems, the institutional divides between the VET sector, on the one hand, and the university and senior secondary education sectors, on the other, have been significant.

Philosophical Underpinnings

To an extent the literature on RPL carries the assumption that the benefits and moral underpinnings are self-evident truths. The literature on equity is voluminous, and there is little doubt that access to education and training is not evenly distributed across social groups, and that opportunities for success within educational systems are uneven. The idea, therefore, that learning that is achieved outside of formal instruction should be recognized does have a moral base in democratic rights. Systems that allow for the recognition of informal learning and thus avoid unnecessary time and effort being invested in the duplication of this learning also have clear benefits. Furthermore, if recognition systems can encourage individuals to engage in formal learning, this should increase the level of participation in education, the skills base, and productivity across a nation.

The RPL claim, however, must go beyond the benefits and address the foundations of formal knowledge and the processes of formalization. The claim appears to recognize that formal knowledge has an elevated status. Otherwise, there is a potential for a contradiction in the claim: that is, if informal learning is as legitimate as formal learning, then why should it be formalized through the RPL processes?

Of course, a claim of equal legitimacy is not an inherent part of the recognition of the informal learning argument. In its basic form, the RPL claim is that learning can take place in multiple ways, through multiple settings, and through multiple experiences and that the processes for the recognition of learning should take account of these multiplicities. This claim would thus acknowledge the value in the formalization of learning through accreditation and qualifications, as in Young's (2001) terms, there are communities of trust that bestow this value. However, the claim should also confront the nature of formalized learning, including its knowledge base.

The traditional construct of a qualification has been through disciplines that in most cases constitute a hierarchy of learning that is bounded by codes and with systems and procedures for ordering and relating knowledge. Subject disciplines have been fundamental to university learning. On the other hand, qualifications within the VET sector have had their base in sets of knowledge and skills drawn from occupational modes and conventions. The training and induction procedures, and the recognition procedures, for new entrants into the occupations, replicate these modes and procedures.

RPL has evolved with two broadly associated movements that in some versions challenge both of these traditions. Social constructivist theories of learning underpin much of the rhetoric on informal learning and either directly or by association with RPL as well. Learner-centered

approaches that are based upon the learners' own construction of understanding and knowledge have an obvious link with the idea of informal learning and give it both a claim for legitimacy and a platform for its recognition. On the other hand, the qualifications within the VET sector in the Anglophone countries are now based upon empirically derived sets of outcomes or standards that are packaged into qualifications.

The constructivist view of learning has to adopt a theory of knowledge that deals with the relationship of personally derived knowledge with socially derived forms, especially in the form of subject disciplines. A constructivist concept of learning can take a poststructural position where legitimate knowledge is constructed by the individual and where the codified versions of knowledge in the form of mainstream qualifications represent the constructions of powerful or dominant groups within a society. This interpretation thus leads to the RPL dilemma that if informally derived learning is as valuable and legitimate as formally derived learning, why should people bother to then gain recognition for informal learning?

It should be stressed that a constructivist view of learning does not have to equate with an individualistic view of knowledge – that is, informal and formal learning are not mutually exclusive. Dewey (1966), who is often cited as one of the founders of constructivism, rejected the individualistic view of knowledge and argued that an individual's construction of knowledge needed to take place in social settings. Indeed, this was the very essence of his notion of democracy in education.

On the other hand, the idea that knowledge can be expressed and recognized as outcomes is more central to the RPL case. Young (2001) located outcomes-based qualifications as being a mainly Anglophone countries approach, with process-based qualifications being more common in Europe. His classifications are not perfect. However, RPL has been most robustly associated with qualifications systems and subsystems that are outcomes or competency based. The high form of an outcomes-based qualifications system is that of South Africa where the objective has been to construct all qualifications through unit standards developed by user rather than provider groups. Allias (2007) is severe in her critique of the South African qualifications system. She argues that its construct of qualifications lacks an epistemological base, and for this reason the vast majority of qualifications that are awarded continue to be those that were developed outside of the unit standard system.

Keep (2006) makes the point that the VET sector in the UK has evolved without the active participation of the 'social partners' in the ongoing system. The same point could be made for Australia and New Zealand and this compares with the active roles of employers and unions in the European countries. Thus, the more formalized processes of the European countries, to which the social partners are strongly attached, contrast with the centrality of

validity and reliability of the outcomes in the Anglophone countries. Thus, it is not surprising that there has been (and continues to be) resistance to the idea of RPL within the VET sectors from some of the social partners in a number of European countries, and especially Germany (based on consultations with VET officials in the Netherlands, Flanders, Germany, Norway, Sweden, and Finland (April, 2007)) where governance of VET is very much in their hands (Attwell, 2006).

This does not necessarily challenge the educational or epistemological legitimacy of RPL. However, it does raise the question of the nature of a qualification. The idea of a qualification is deep within the educational psyche. Outcomes-based education has made only sporadic tilts at the higher education windmill and with minimal impact. Correspondingly, RPL has a weak presence within the sector. Pokorny, 2006 noted that APEL has "failed to establish itself as part of the mainstream" (p. 261) within higher education in the UK. Universities have mostly ignored the unit standard system in South Africa (Mathais, 2007) and have resisted a similar system in New Zealand. In Australian universities, RPL appears to be beset with problems of inconsistency (Fox, 2005).

In the UK and Australia, a high proportion of universities have RPL policies in place, but with little evidence that they are being accessed by students. This possibly relates to Wheelahan's (2006) discussion about the nature of a qualification, which she argues is more than the sum of its parts, and the inherent attributes of gradueness mean that a qualification should not be awarded solely upon the basis of RPL. The concept is useful because it approaches the idea of a qualification from both the provision and participation sides. The reluctance of the education sector to issue an award solely upon the basis of assessments without some formal engagement in learning is matched by the reluctance of the potential graduate to gain an award without some engagement. The numerous studies of RPL participation have frequently identified the lack of student applications as a key factor in low rates of realization.

This argument then leads to the question of whether there are and should be differences between RPL for course entry and for credit. It seems likely that there is a widespread acceptance of the principle that informal learning should be a basis for course entry.

Practice and Outcomes

The point of the above discussion is that much of the official rhetoric on RPL from national education and training agencies and international organizations is about the RPL claim and the procedures that need to be taken to achieve recognition. However, behind this rhetoric the very idea of prior and informal learning, there are fundamental

questions about what constitutes informal learning and its association with formal learning Colley *et al.*, 2006) and by implication the sociology of knowledge.

Such questions may not consciously confront most RPL practitioners and participants, but they are inherent in their practice, motivations, and experiences. Much of the case study and other literature on credit-based RPL has typically located student preference for engaging in whole courses of study or training and provider and provider personnel behaviors that do not readily relate the unaccredited learning of individuals to the learning outcomes of formal qualifications and their elements.

Where the outcomes are stated in behavioral terms, as they are in competency based training (CBT), which, in turn, can be gained through experience of apprenticeship types of learning, the capacity for RPL should be greater because the individual and providers can more readily associate experience with the nature of the outcomes and processes for achieving it. So it should be expected that the realization of RPL would be greater within CBT and outcomes-based qualifications than in the more traditional course- or process-based qualifications. In Australia, one of the countries that have led the charge on RPL and which has had a CBT-based VET system for well over a decade, the percentage of VET students with an RPL was 3.6% in 2004 (Hargreaves, 2007). The percentage had grown through the 1990s, but now appears to have stabilized. Although the percentage is not insignificant and is certainly greater than in the higher education sector, it would appear to be relatively modest. Total recurrent VET expenditure in Australia in 2004 was \$4887.1 million. Therefore, the level of financial savings through RPL could be \$176 million, but is discounted by the fact that not all students gained all of their awards through RPL and also by the costs of the RPL procedures. There is some speculation that the savings and costs are nearly the same.

On the other hand, 10% of students at the diploma levels and above had an RPL. RPL for course entry is also likely to be higher at the diploma level. As Hargreaves (2007) points out, this suggests that it does remain an important principle and instrument for adult learners, and its potential role in continuing education and training should not be under-valued.

Lifelong Learning: A Third Age of RPL?

Despite the mainly disappointing levels of realization of RPL, the international rhetorical and policy investments in it have risen in recent years. This appears to be associated with two interrelated sets of developments: the policy goal of lifelong learning and the reform and integration of qualification systems.

In 1996, OECD education ministers adopted the principle of lifelong learning for all as the guiding motif for

education and training policies and activities. In 2001, the European Commission adopted a Communication on *Making a European area of lifelong learning a reality* (EC, 2001). Subsequently, the Council of the EU has passed a series of resolutions endorsing principles and practices of RPL, and under the banner of making learning visible, the EC has initiated a series of activities including the establishment of an inventory on tools and systems for recognition (Otero *et al.*, 2006).

Parallel with these developments has been the spread of NQFs beyond the Anglophone countries to most of the European countries (CEDEFOP, 2006) and to most other parts of the world. The OECD has drawn the link between lifelong learning, qualifications systems, and RPL in both its study on national qualifications systems (OECD, 2006a) and its recently launched activity on the recognition of non-formal and informal learning (OECD, 2006b).

There is an observably new interest in RPL across countries and there are frequent link between RPL initiatives and NQFs. Countries with strong apprenticeship systems with high levels of social partner steerage or management of qualifications systems, such as Germany and the Netherlands, are now developing NQFs. There is a common assumption that these developments will go hand in hand with the development and spread of different recognition practices, and in each of these countries there are reported tensions with some stakeholder groups and organizations over these principles and practices.

Furthermore, these developments appear to have revived interest in extending recognition of prior and informal learning to younger age groups. For example, the Council of the EU has resolved that "The value and visibility of non-formal and informal learning for young people should be enhanced by recognising the work and achievements of young people and those active in youth work and youth organisations" (EUROPA, 2006: 1). In Australia, two states have recently included the recognition of community-based learning within their senior secondary certificates.

Of course, the use of NQFs as a means of spreading, deepening, and validating the practices of RPL has yet to be realized in any measurable forms. In this third stage and iteration of RPL, there are questions about the characteristics of NQFs. If Allias' (2007) and Strathdee's (2003) critiques of the South African and New Zealand NQFs were to be applied across all NQFs, it is likely that their evolution will be troubled and their capacity to advance the RPL case will be limited. However, NQFs are a variable concept and have very different institutional forms. By and large, the South African and New Zealand examples were regulatory frameworks: all qualifications were to be designed within common formats. The alternative is enabling frameworks that establish a common language for qualifications and their learning elements to be compared and articulated. For example, the Scottish

example (the Scottish Credit and Qualifications Framework) is described as a credit framework that is designed to enable and encourage articulation between courses and to support the wider recognition of prior informal learning (Scottish Executive, 2004).

The hopes and expectations of RPL have grown and waned and now grown again over the last two decades. The returns to the investments in rhetoric, system governance, and innovations in qualifications, assessment and RPL procedures to this stage have been weak. The latest surge in investment is considerable and comes with the rhetorical force of lifelong learning and major developments in qualifications systems. Its realization is much anticipated.

See also: Assessment in the Workplace of Performance, Developing Expertise and Competence; Assessment in Vocational Education; Provision of Prior Learning Assessment; Qualifications Frameworks and their Role in the Reform of Education and Training.

Bibliography

- Allais, S. (2007). *The Rise and Fall of the NQF*. PhD Thesis, unpublished, Witwatersrand University, South Africa.
- Attwell, G. (2006). *Recognition of Informal Learning in Germany and the UK. Thema 2006/08: ICQVET – validating informal competencies*. . <http://www.dji.de/cgi-bin/projekte/output.php?projekt=613&Jump1=LINKS&Jump2=31> (accessed May 2009) .
- Bowden, J. A. and Masters, G. N. (1993). *Implications for Higher Education of a Competency-Based Approach to Education and Training*. Canberra, ACT: Australian Govt. Publishing Service.
- Blackmur, D. (2003). Policy challenges in educational quality assurance: The New Zealand Qualifications Authority, 1997–1999 University of the Western Cape. *Paper Presented at the South African Qualifications' Authority Conference: 2002, Durban, South Africa*, unpublished.
- Cameron, R. and Millar, P. (2004). *RPL. Why Has It Failed to Act as a Mechanism for Social Change? Social Change in the 21st century*. Brisbane: Centre for Social Change, Queensland University of Technology.
- CEDEFOP (2006). *Summary of Responses Received to the Commission's Consultation on the EQF during the 2nd Half of 2005, CEDEFOP Info 1*.
- Colley, H., Hodgkinson, P., and Malcolm, J. (2007). *Non-Formal Learning: Mapping the Conceptual Terrain. A Consultation Report, infed*. http://www.infed.org/archives/e-texts/colley_informal_learning.htm (accessed May 2009).
- Cooper, L. (1998). From "Rolling Mass Action" to "RPL": The changing discourse of experience and learning in the South African labour movement. *Studies in Continuing Education* **20**(3), 143–157.
- Dewey, J. (1966). *Democracy and Education: An Introduction to the Philosophy of Education*. New York: Free Press.
- EC (European Commission) (2001). Communication on "Making a European Area of Lifelong Learning a Reality", European Commission. http://ec.europa.eu/education/policies/lll/life/communication/com_en.pdf (accessed May 2009).
- EUROPA (2006). *Recognition of Non-Formal and Informal Learning (in the Field of Youth)*. <http://europa.eu/scadplus/leg/en/cha/c11096.htm> (accessed May 2009).
- EU (2004). *Conclusions of the Council and of the Representatives of the Governments of the Member States Meeting within the Council on Common European Principles for the Identification and Validation of Non-Formal and Informal Learning*. http://ec.europa.eu/education/policies/2010/doc/validation2004_en.pdf (accessed May 2009).
- EURYDICE (2007). *Non-Vocational Adult Education in Europe. Executive Summary of National Information in Eurydice*. <http://www.eurydice.org/portal/page/portal/Eurydice/PubPage?pubid=083EN&fragment=52&page=1> (accessed May 2009).
- Fox, T. (2005). Adult learning and recognition of prior learning: The 'white elephant' in Australian universities. *Australian Journal of Adult Learning* **45**(3), 352–370.
- Hargreaves, J. (2007). *Recognition of Prior Learning. At a Glance*. Adelaide: National Centre for Vocational Education Research.
- Harris, J. (2006). *Introduction and Overview of Chapters*, In Andersson and Harris.
- Keep, E. (2006). State control of the English education and training system – playing with the biggest train set in the world. *Journal of Vocational Education and Training* **58**(1), 47–64.
- Lugg, R., Mabitla, A., Louw, G., and Angelis, D. (1998). Workers' experience of RPL in South Africa: Some implications for redress, equity and transformation. *Studies in Continuing Education* **20**(2), 201–215.
- OECD (2006a). *Qualifications System. Bridges to the Future*. Paris: OECD.
- OECD (2006b). *New OECD Activity on the Recognition of Non-Formal and Informal Learning. Guidelines for Country Participation*. Paris (unpublished).
- Otero, M. S., McCoshan, A., Junge, K., and Winter, J. (2006). *A European Inventory on Validation of Non-Formal and Informal Learning. Overview of Findings: Needs and Initiatives*. Birmingham: ECOTEC Research & Consulting Limited.
- Pokorny, H. (2006). *Recognising Prior Learning: What Do We Know?* In Andersson and Harris (eds.).
- Scottish Executive (2004). Scotland's lifelong learning framework, Edinburgh. <http://www.scotland.gov.uk/Publications/2004/04/19207/35629> (accessed May 2009).
- Strathdee, R. (2003). The qualifications framework in New Zealand: Reproducing existing inequities or disrupting the positional conflict for credentials. *Journal of Education and Work* **16**(2), 147–164.
- Wheelahan, L. (2006). *Vocations, 'Graduateness' and the Recognition of Prior Learning*. In Andersson and Harris (eds.).
- Young, M. (2001). *The Role of National Qualifications Systems in Promoting Lifelong Learning*. Background paper for the OECD. Paris, unpublished.

Further Reading

- Andersson, P. and Harris, J. (eds.) (2006). *Re-Theorising the Recognition of Prior Learning*. Gosport: Ashford Color Press.
- Gallacher, J., Toman, N., Caldwell, J., Edwards, R., and Raffae, D. (2005). *Evaluation of the Impact of the Scottish Credit and Qualifications Framework (SCQF)*, Edinburgh Scottish Executive. <http://www.scotland.gov.uk/Publications/2005/11/30173432/34323> (accessed May 2009).
- LSC (Learning and Skills Council) (2007). *LSC Position Paper on Recognising and Recording Progress and Achievement in Non-accredited Learning*. <http://readingroom.lsc.gov.uk/lsc/National/rarpa-PositionPaper.doc> (accessed May 2009).
- NCVER (National Centre or Vocational Education Research) (2006). *Students outcomes survey 2005*. http://www.ncver.edu.au/statistics/surveys/sos05/sos05_counts_TAFE_MC.xls#T104!A1 (accessed May 2009).
- OECD (2005). *International Assessment of Adult Skills: Proposed Strategy*, COM/DELSA/EDU(2005)4/REV1. Paris: OECD.
- Young, M. (2006). *Endword*. In Andersson and Harris (eds.).

The Competency Model

A Gonczi, University of Sydney, Sydney, NSW, Australia

P Hager, University of Technology, Sydney, NSW, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

This article first outlines the recent origins of the competency-based training and learning model. It then outlines briefly the various main understandings of competence. These differences help to account for variation in the ways that the model has been deployed in different countries. Next, the international impact of agencies such as the Organization for Economic Cooperation and Development (OECD) on the spread of the competency model is outlined. Some examples of the implementation of competency-based training and learning are then briefly described. The article concludes with a discussion of various important ongoing issues surrounding the competency model. These include:

- limitations of verbal descriptors of competence;
- role of learning from work in competence development;
- contextuality of generic competencies;
- relation of competencies to curriculum;
- relation of competencies to assessment; and
- competency and lifelong learning.

The Recent Origins of Competency-Based Training and Learning

Over recent years, there has been an increased interest internationally in the relationship between education and the workplace. Countries in almost every part of the world, including Scotland, England, and Wales, Canada, the United States, Mexico, many South American countries, Australia, New Zealand, Sweden, France, Kuwait, Indonesia, Korea, and Thailand, have all undertaken substantial reforms of their vocational education and training systems. A number of them are also making substantial changes to higher education and to their school systems. But it is in the vocational education training (VET) area where the changes have been most significant. These reforms include: attempts to link traditional vocational education more closely to industry requirements; attempts to increase VET student numbers; introduction or expansion of vocational options in schools; attempts to develop partnerships between vocational and other educational sectors, partly to raise the status of vocational education; the development, by industry bodies with government support or participation, of occupational and employment-related competency standards which influence the content of

the curriculum in both general and vocational education; schemes to facilitate student transfer from one level of education to another, and to facilitate the recognition of prior and experiential learning; blurring of the distinction between vocational and university education; attempts to increase the quantity and quality of industry training and to develop alternatives to what is perceived as a government monopoly in the delivery of VET.

In addition to these systemic reforms, virtually all the countries mentioned above are committed to a particular approach to VET – competency-based education/training – though what is meant by this differs substantially between countries. What the reforms have in common is that the content of VET courses is based on occupational-competency standards, that it is expressed in terms of outcomes to be achieved by students, and that the assessment of students is (at least partly) based on the criteria expressed in the competency standards. Another common feature is that influential people from outside of the education profession have been involved in the formulation of the aims and content of education. This includes industrialists, governments, unionists, lobbyists, and bureaucrats. What is different between countries is how these competency standards are conceptualized, how and by whom they are developed, and the degree to which the standards shape the curriculum and assessment of VET.

The international interest in VET and in more relevant professional education and in employment-related competencies in general education in schools, is partly a product of a complex range of factors associated with national and international economies. Technological change, industry restructuring due to changes in global trade arrangements, changes in the capital and labor markets, and so on, have contributed to general unease and uncertainty about the aims of education. The reforms in higher education, the upgrading of VET, and the introduction of employment competencies in the school sector, have all been justified by their capacity to increase the levels of skills and flexibility in the economy, assumed to be a prerequisite to increased international competitiveness in the global economy. The relative cultural stability of the period from 1950 to 1980 has been deposed by the globalization of the last two decades and by the vast technological changes which, among other things, have led to a phenomenal decrease in the half-life of knowledge. This has led to a far more complex world and a widespread cultural anxiety which is at the root of much of the

educational policy of the last decade. It is certainly at the heart of the currently perceived need to develop skills-formation policies and to enlist the education system at all levels in this process. But it also helps explain the desire of a variety of groups to influence and even determine the content of educational provision.

In the schools sector, governments have attempted, among other things, to develop and incorporate employment-related competencies into the general-education curriculum. The aim has been to develop in young people competencies which would make them more effective participants in work, more flexible and adaptable, and more able to move into higher education, since schooling has failed to develop in students skills that are needed for the high-skills workplace of the 1990s. For instance, the Finn report in Australia in the late 1980s argued that Australia had inherited a tradition of education which accepted the divisions between general and vocational education and that the two needed to be brought together both for the sake of individuals and for the economy generally. It argued that schools needed to be concerned with employability and that the vocational sector needed to be concerned with more general, broader vocational education.

The Finn Report recommended that schools needed to be concerned with competency areas that all young people needed for employment: language and communication; mathematical and technological understanding; cultural understanding; interpersonal; and problem solving. In 1991, a new committee which has become known as the Mayer committee was given the task of developing the areas outlined in the Finn report into a number of more specific competencies. In the Mayer report, *Putting General Education to Work* (1993), seven key competencies were developed: collecting, analyzing, and organizing information; communicating ideas and information; planning and organizing activities; working with others and in teams; using mathematical ideas and techniques; solving problems; and using technology. Later, another was added – displaying cultural understanding.

These reforms which were introduced in many countries in the 1990s resulted in some fierce debates among educationalists. One view was that they have been part of a conservative agenda to refashion education in the interests of financial capital. Another, however, has been that introducing employment-related competencies in the curriculum is essential to equity policy – the desire to increase life chances of individuals (most often those from working-class backgrounds) failed by the traditional education system.

Most academic educators have opposed these reforms on the grounds that:

- they give too much weight to industry needs at the expense of the needs of students – that the reforms are too economically driven and are overly instrumental;

- they attempt to control teachers'/trainers' work through national curriculum and assessment; and
- the outcomes of education should be far wider than vocational outcomes and must include values and orientations related to justice, esthetics, democracy life skills, and so on.

They also claim that these reforms are inimical to the interests of students. Any vocational emphasis in education, it is argued, is technocratic, specific, practical, and managerial. It supports a view of the function of education that it is concerned with the transmission of exploitable knowledge and participation in the market through the development of skills which possess an exchange value (Feinberg, 1983). General academic education on the other hand is democratic, egalitarian, critical, and collaborative (Wellington, 1993).

Interestingly, similar arguments are used by at least some conservative politicians. They claim that such developments are merely trendy, that they deemphasize the traditional canon and thus lead to a reduction in educational rigor. In the United Kingdom, this has led to demands from some strands of the conservative hierarchy to strengthen academic education (the so-called gold standard) at the expense of education designed to appeal to a wider cohort of students.

However, it can be argued that these criticisms of the education reforms in schools (which include criticisms of the role of key competencies in general-education settings) are ill-founded. They rest on a view of vocational education which is both elitist and conceptually impoverished. Furthermore, they ignore the potential of key competencies to liberalize vocational education as well as to liberalize, even transform the curriculum and pedagogy of general education. Incorporating the key competencies into VET will ensure that VET becomes more liberal, supporting Dewey's concept of education through, rather than specifically for, occupations.

In addition, incorporating the key competencies into general education may provide the impetus for the reform of traditional subjects and traditional modes of teaching (see Lohrey, 1995). A key competency such as problem solving, for example, can be developed through traditional academic subjects but lends itself better to holistic, interdisciplinary knowledge and its application and the use, for example, of problem-based teaching methods.

Taken as a whole, these developments mark a significant change in educational policy. For the first time, education has been openly debated in instrumental terms and is being judged, at least partly, by the extent to which it is succeeding in advancing national goals related to economic restructuring and competitiveness. It is hardly surprising, given the politically charged atmosphere in which it was first advocated, that there has been little real analysis of the nature of such competency-based education and the concepts on which it is based. It is to this subject that this article now turns.

Nature of Different Approaches to Competency-Based Education

Competence can be thought of in several quite different ways. How we view competence significantly shapes the ways competency standards are used and assessed. As performance of tasks is something that can be readily described, the common sense but naive view of competence equates it with successful performance of a set of discrete tasks. Competency standards are then thought of as objective descriptions of the set of tasks. The basic error here is the confusion of tasks with the capacity to perform the tasks. According to the Concise Oxford Dictionary, competence (or competency) denotes the ability to do something or the ability for a task. Here the emphasis is on competent people having ability or capability that will enable them to complete tasks. But abilities or capabilities involve concepts such as knowledge, skills, and attitudes, that is, relevant personal characteristics that underlie competent performance. But these personal characteristics are not so readily describable as are sets of tasks. Our understanding of human capacities and abilities remains limited, in the sense that attempts to specify them are invariably debatable and contested. The task view of competence skates over these matters by pretending that a relatively objective checklist of tasks describes competence. It does no such thing. One result is that the task view of competence omits vital higher-level competencies, such as planning or reacting to contingencies. Hence training programs based on this approach are flawed.

Another approach views competence as having a series of desirable attributes including appropriate kinds of knowledge, skills, abilities, and attitudes. This is a generic approach that centers on training and assessing learners' performance on each of these separate attributes. At least this approach promises to address the less predictable, nonroutine aspects of work. But its main limitation is that training and assessing attributes in isolation from actual work practice bears little relation to future occupational performance. It turns out that attributes such as problem solving, analysis, pattern recognition, etc. are highly context dependent, thereby vitiating the effectiveness of attempts to train and assess them out of context. Another problem with this approach is that when the generic attributes of various occupations are compared, it can be difficult to tell one occupation from another. This is another way of saying that competency-based theory (CBT) needs to be grounded in context.

The two approaches considered so far have focused exclusively on either tasks or attributes. A third approach, the integrated conception, views competence in terms of knowledge, abilities, skills, and attitudes displayed in the context of a carefully chosen set of realistic professional tasks which are of an appropriate level of generality

(Gonczi *et al.*, 1990; Hager, 1994; Biggs, 1994). This approach centers on key tasks involved in the practice of the occupation. Major attributes, such as cognitive skills (knowledge, critical thinking, and problem-solving strategies), interpersonal skills, affective attributes, and technical/psychomotor skills that are required for competent performance of these key tasks are identified. Competency standards that integrate these attributes and tasks provide a richer representation of practice than was possible with the other two approaches. This approach accords with the dictionary definition of competence that links ability or capability with tasks. The integrated view of competence is holistic in that it situates attributes in the kinds of contexts in which they are employed in the practice of the occupation. This is also suggestive of training and assessment situations suitable for effective learning and assessment.

The integrated approach highlights an important asymmetry surrounding the concept of competence. We have seen that competence links abilities and capabilities with tasks. But though performance of tasks is directly observable, the abilities or capabilities that underlie the performance are not directly observable. We infer their presence or otherwise on the basis of the performance. So assessment of competence is necessarily based on inference from samples of performance. A literature has also developed around the notion of capability as a replacement concept for competence. However, there appears to be little discernible difference between the integrated approach to competence and the notion of capability.

The Role of the OECD in the Debate about a Competency Approach to Education

Over the last decade, there has been an increasing emphasis in OECD publications on the impact of social and economic change on education and a desire to investigate whether the citizens of developed countries are learning what is appropriate for the new age. What are these skills understandings and dispositions (what we define as competencies in our integrated approach to competence)? Are students well prepared to meet the challenges of the future? Are they able to analyze, reason, and communicate their ideas effectively? Do they have the capacity to continue learning throughout life? These are questions that parents, students, the public, educational bureaucrats, and increasingly, politicians, in both developed and developing countries, are asking. One project in particular, Definition and Selection of Competencies: Theoretical and Conceptual Foundations (De Se Co), commenced in 1997 under the OECD's education indicators program International Indicators of Education Systems (INES) has recently attempted to answer this question. The project's

aim was to clarify what competencies apart from reading, writing, and computing are necessary for an individual to lead an overall successful life and for society to face the challenges of the present and the future. The program of work was the most extensive ever undertaken to answer these and other questions. A number of activities were undertaken: an analysis of international studies on educational indicators, a study reviewing scholarly work on the concept of competence, expert papers on key competencies by academics from different disciplines and different countries (economics, psychology, anthropology, sociology, and philosophy from Switzerland, France, Germany, UK, and the USA) and comments on these from practitioners and policymakers. These were all discussed at a symposium in 1999 and the papers and comments later published in a book in 2001 (Rychen and Salganik). Following this, an analysis of country-specific key competencies was undertaken and a second symposium convened in 2002. A strategy paper followed and then a second book appeared by the same authors (Rychen and Salganick, 2003). The project concluded that the notion of competence is an important one and goes well beyond the notion of skill. Competent performance implies the mobilizing of knowledge, cognitive, and practical skills, as well as attitudes, emotions, values, and motivations. Competencies become key when they: contribute to highly valued outcomes at the individual and societal level, are instrumental for meeting important complex demands and challenges in a wide variety of contexts, and are important for all individuals.

They identified three main categories of competencies:

- interacting in socially heterogeneous groups,
- acting autonomously, and
- using tools interactively.

The specific competencies in the first category – the ability to relate well to others, to cooperate, and to manage and resolve conflict were seen to be vital to life in a globalized world and increasingly in individual multicultural societies. In the second category – the ability to act within the big picture, to form and conduct life plans, the ability to defend and assert one's rights, interests, needs, and limits – are all vital in various aspects of life: work, family, and civic society. The third category included the ability to use, interactively, language, symbols and texts, knowledge and information, and technology. This category had in mind reading, mathematical, and scientific literacy among other things.

Underpinning all these is a critical and reflective stance. The conclusion of the project was that if we are to understand the idea of key competencies, it is necessary to go beyond the outcomes of the typical curriculum of formal institutions, which stress on variously recalling knowledge and sometimes thinking abstractly, to the kind of capacity that is developed through the integration

of formal and informal knowledge. These are capacities which can only be developed gradually over a lifespan.

Clearly, this approach to the nature of competency provides a persuasive justification for lifelong learning and in particular the need to provide opportunities for adults to develop and learn throughout their lives. In addition though, this approach also obviously has potential implications for the curriculum and structures of schools and tertiary institutions.

The project also supported the principle that the key competencies were interrelated and that these constellations of key competencies (DESeCo summary p. 4) will vary with the contexts in which they are applied.

The implications of this work for future international assessments are significant. The project also enables those involved in school-based assessment to understand the work on assessing key competencies in adults. This should have the effect of improving the assessments in both contexts. The International program for Student Assessment (PISA) is the most prominent example of these assessments. The PISA program is a collaborative program which measures how well students at age 15, the end of compulsory schooling in most OECD countries, are prepared to meet the challenges of contemporary society. The assessment reflects the curriculum in the OECD but is focused on the use of knowledge in the real world. The assessment reflects the capacity of students to engage in lifelong learning by using their school-developed knowledge in a variety of nonschool settings. Thus while PISA does assess curriculum knowledge, it also examines the capacity to reflect and apply knowledge. The assessment takes place every 3 years, through the use of pen-and-paper tests. The survey was implemented in 43 countries in 2000/2 and 42 countries in 2003 and again in 2006 (results available in December 2007). Between 5000 and 10 000 students from 150 plus schools are typically tested in each country. It is administered over a 3-year cycle covering reading literacy (2000), mathematical literacy (2003), and scientific literacy (2006).

The reasons for adopting this approach are outlined in the PISA 2003 Framework (OECD, 2003). In summary, the OECD believes that certain broad general skills are essential to success in life beyond school. These include communication, adaptability, flexibility, problem solving, and the use of information technologies. These competencies are developed across the curriculum and assessment of them requires a cross-curriculum focus. In addition, in the specific domains measured in PISA, the application of knowledge in adult life depends more on the acquisition of broader concepts than knowledge acquisition in schools. In mathematics, for example, the capacity to reason quantitatively is more important than answering text-book questions.

These OECD initiatives reflect increasing international attention to so-called generic skills, core skills, or

basic skills, or, more recently, employability skills. (Sometimes the term competencies replaces skills.) These terms refer to diverse qualities and capacities that are argued to be important in many life situations, though work and education as a preparation for work are usually the main focus. As noted earlier, examples include thinking (logical and analytical reasoning), problem solving, communication skills, teamwork skills, knowledge and information skills, together with values such as persistence, integrity, and tolerance. These so-called generic skills are seen as different from the knowledge that characterizes traditional disciplines or the physical skills associated technical fields.

As the list makes clear, generic skills covers a very diverse range of things – skill, attitudes, values, and dispositions. Some of these are clearly not skills in the usual sense, such as attitudes and dispositions. Even those that most resemble genuine skills range from the physical (e.g., body language in interpersonal communication) to the mental (e.g., reasoning). There may be significant differences here that are masked by the blanket term skills. The more attitudinal and dispositional qualities can be more accurately viewed as products of cultural, ethical, and social circumstances that may be refined and modified by knowledge and reflection. Rather than being unitary things, they are complex relations that connect persons and particular contexts.

Adoption of the Competency Model in Various Countries

Competency-based approaches to education have been taken up in many countries over the past decade. However diversity is evident in these initiatives, reflecting differing views of purpose, as well as somewhat different understandings of competence. The United Kingdom for instance has set up a system of national vocational qualifications (NVQs) and general national vocational qualifications that cover many occupations and work roles, including many that had no previous formal qualifications. The NVQs in particular have tended toward a task approach and have been criticized for atomizing work into checklists of discrete observable skills. In Australia, a more integrated understanding of competence has been influential. This has tended to distinguish competency standards from course curricula. This separation has reached the stage where so-called training packages which outline competency standards and ways to develop them effectively have effectively replaced formal curricula. It is left to the training providers to develop their own curriculum in order to deliver a given training package. New Zealand, on the other hand, specifically equates units of competency standards with units of curriculum. Scotland also has a system that bears similarities to the one in New Zealand.

Many countries in Latin America have adopted competency-based approaches to VET over the last decade or so. The impact of globalization on countries such as Mexico, Costa Rica, Ecuador, and Chile, for example, has resulted in significant economic and educational reform to both vocational and in some cases higher education. In Mexico, for example, with the support of the World Bank, technical education was reformed, between 1995 and the early years of this century through the development of a national system of labor competencies and a system of colleges *El Colegio Nacional de Educación Profesional Técnica* (CONALEP) which developed competency-based curriculum to enable students to meet these national labor competencies. In the early stages of this modernization process, there was little industry involvement, though this is slowly changing. However, unlike Britain and Australia, the implementation of these reforms has been government rather than industry led. Developments in Mexico have been influential in many Latin countries (Morales, 2000).

Ongoing Issues Regarding Competency-Based Training and Learning

Limitations of Verbal Descriptors of Competence

Competencies are commonly assumed to be easily and unequivocally describable in language. So it is thought to be a straightforward matter to describe occupational competence in written form. But actually, it is tasks to be performed that can be accurately stated in language. As already suggested, specifying the precise attributes that underlie performance is less well founded. Some aspects of performance are tacit or implicit. Consider, for instance, the case of face recognition. Humans can easily pick out a familiar face from thousands of unfamiliar ones. But to state in words exactly how this is done is something that remains elusive. As Lum (1999: 410) succinctly puts it, the assumption that human capabilities can be unequivocally described and accurately communicated by means of language is unfounded. So, at best, written competency standards are rough and ready, though useful, guides. They have inherent limitations and disadvantages. We should be wary of the tendency to assume that reality is exhausted by attempted descriptions of it.

Role of Learning from Work in Competence Development

The competence approach to vocational education can be seen as an attempt to render it fail-safe. It was thought that traditional curricula reflected what teachers regarded as important rather than what industry really needed. CBT remedies this by being based on explicit descriptions of competent performance in the given occupation. So, it was

assumed that if CBT is implemented effectively, it will produce graduates who are ready to function competently in any workplace. However, this optimism proved to be misplaced. It seems that a gap between skill-development outcomes and workplace competence is inescapable. A prime reason for this is that contextual factors, which are often unique to particular workplaces, shape and partly constitute competence in any given situation. This means that

... the learning required for competent workplace performance is normally much greater than the learning that can occur in formal pre-service courses based on standardised training outcomes. It seems that some context-specific learning, that can only occur from the actual practice of an occupation, is a vital part of competence. Nor is this learning necessarily directly transferable to practice of the same occupation in a different context. (Hager and Smith, 2004: 43)

This crucial feature of CBT was widely overlooked leading to inevitable disappointment as the new approach failed to produce graduates able to immediately perform competently in the workplace. This gap can be reduced if suitable varied workplace learning and experience is organized in parallel with standards-based learning and achievement of skill-development outcomes. This is what happens in well-run apprenticeships. However, even here, if the novice worker or apprentice moves to a job in a different organization, he/she will very likely need a further period of workplace learning and experience to become competent in the new situation.

Contextuality of Generic Competencies

Not only is competence significantly contextual as outlined in the previous section, but generic attributes and the ways in which they cluster are strongly shaped by the particular features of the context in which work is carried out (Hager, 1997: 13–15). The influence of the context is such that it is unhelpful to identify the generic attributes of an occupation or profession. The relative importance of these generic attributes and the ways in which they cluster will very much change with the workplace context. The notion of context is itself complex and includes a multiplicity of workplace-related factors such as:

- the specific history of a workplace or company;
- its particular culture and norms;
- its institutions and practices, for example, work organization and career structure;
- its economic and social environment;
- its strategic needs;
- its deployment of technology; and
- the extent and intensity of change to which it is subject.

Research on the workplace role of generic attributes shows that there is significant variation in competence requirements across work sites within one occupation. Such variation means that prespecified skill-development outcomes, including standardized generic attributes, cannot meet all of the requirements of particular work sites. For example, in the United States Stasz *et al.* found differences not only in generic attributes needed across occupations, but also in the same occupation practiced in different organizations and work sites. They concluded (Stasz *et al.*, 1996: 102) that

... whereas generic skills and dispositions are identifiable in all jobs, their specific characteristics and importance vary among jobs. The characteristics of problem solving, teamwork, communication ... are related to job demands, which in turn depend on the purpose of the work, the tasks that constitute the job, the organization of the work, and other aspects of the work context.

Thus, even within the same occupation, job demands can vary so much between different companies or work sites that it makes little sense to try to specify the exact generic-attributes mix for a particular occupation. The high contextual sensitivity of generic-attribute requirements of work is further illustrated by the later research findings of Stasz and Brewer (1999).

Relation of Competencies to Curriculum

We can start this section with some basic logical points (Hager, 1995). Performance descriptors are outcomes. They describe what the end products are like at particular stages of development. Curricula are about what needs to be gone through to attain outcomes, that is, they are essentially processes. So performance descriptors as outcomes are logically distinct from curricula as descriptions of processes. Moreover, learning outcomes are mini-stages in these course processes, and so are not to be confused with the performance outcomes of a workplace. To the extent that they have any connection with courses, performance descriptors describe course outcomes (states), whereas curricula describe a process to be gone through (literally a course to be run). So, a typical curriculum describes what is to be learnt and some processes of how it is to be learnt, as well a statement of what people should be able to do at the end of the course. As performance descriptors are summative rather than formative, they will hinder curriculum development if they are thought of as being a curriculum. In other words, a curriculum describes a developmental process by which certain outcomes will be achieved, whereas, at best, performance descriptors describe some of these outcomes without specifying how they are to be learnt. The temptation to confuse performance descriptors with curricula presumably comes from viewing learning as bite-sized discrete chunks and performance

as a series of discrete skills that might be equated with these chunks. But these ideas spring from viewing learning as a product together with its other accompanying assumptions. As is argued later, a better view of learning is available.

Similar arguments apply to teaching. Teaching is an important part of the process that aims to arrive at, among other things, some of the outcomes specified in the performance descriptors. However, various attributes, such as knowledge and skills are typically prerequisites for performances specified in the performance descriptors. Thus it will be a waste of effort to try to teach the performance descriptors bit by bit unless the groundwork of requisite knowledge and skills has already been laid down. But as we have seen, the constituents of competence are themselves contested, so there is room for various views about the capabilities, abilities, skills, etc. that a given course should be fostering. Besides, in most cases, this requires a significant developmental process. Thus the performance descriptors, as a guide to actual teaching activities, are of very limited value. Certainly, if useful at all, it will be in the later, rather than the earlier, stages of a course. This is when vocational courses are usually more concerned with actual workplace conditions anyway. The futility of teaching the performance descriptors bit by bit should be obvious except to those with the most naive learning-as-product views. Rather than raising the skill levels and capacities of the workforce, the opposite would result from such an approach.

As always, there is more than one way to teach effectively to achieve some specified outcome. Hence, though the performance descriptors at best describe some major expected outcomes of successful completion of the course, they do not prescribe the teaching strategies that will best lead to these outcomes. Thus, for teachers there is as much responsibility as ever to decide what to teach, and when and how to teach it.

Relation of Competencies to Assessment

Atomistic performance descriptors are attractive to some because they offer the possibility of simple piece-by-piece assessment by direct observation of the performance of tasks against a check list. Once again, this springs from viewing learning as a product together with its accompanying assumptions. However, such assessment raises invalidity problems of several kinds. In general, it assesses superficial aspects of an occupation and ignores the holistic character of quality performance. Experience has also shown that, despite its initial attraction for some, this approach leads to unacceptably large amounts of time being spent on assessment of a myriad of discrete tasks.

In the early stages of courses, foundational knowledge and a range of enabling skills, which provide an essential basis for the future development of occupational competence,

need to be acquired. So the occupational performance descriptors will not be immediately relevant to course-assessment practices. When curriculum is construed as a process, attainment of appropriate developmental stages, rather than of full occupational competence, will be most typically the concern of assessment. Only in later stages of the courses will assessment against occupational performance descriptors be likely to be of any relevance.

It has been argued at length that we need to clearly distinguish:

- performance and its outcomes;
- the underpinning constituents of competence (capabilities, abilities, and skills); and
- the education, training, or development of people to be competent performers.

The idea that learning is a product encourages these three categories to be confused one with another. More holistic and process-oriented approaches to teaching avoid this trap.

Competency and Lifelong Learning

The preceding considerations suggest one way of thinking about lifelong learning. From the early years of schooling and before, learners can be expected to be in situations in which they would be acquiring some basic proficiency in deploying at least some common generic attributes, for example, using household microelectronic technology. One outcome of a sound education would be a growing capacity to deploy successfully generic attributes in an increasingly diverse range of situations and contexts. This suggests that the development of generic attributes should become gradually more integrated and holistic as young people move through schooling. The idea is that sound performance in very many of life's situations centers on successful deployment of suitable combinations of generic attributes. Such a staged development of generic attributes would facilitate students' transition to vocational training, higher education, work, and other postschool activities.

In the Australian vocational sector, the recent agenda to embed generic attributes in the training courses is an attempt to achieve a balance between the capacity for lifelong learning in the longer term and employability in the short term. The greater articulation between vocational-education courses and university courses, as well as more clearly delineated curriculum pathways are, of course, an identifiable form of lifelong learning.

The development of graduate attributes by universities is closely linked to their role in fostering graduates with a capacity for lifelong learning. Various graduate attributes have been recognized as important for lifelong learning (see, e.g., Candy *et al.*, 1994). Developments at the University of Otago provide an example of the close links

between these two notions. Viewing lifelong learning as an element of gradueness, a concept elaborated by the English Higher Education Quality Council (HEQC, 1996), the university has instituted an iterative process in which links are strengthened in its courses between the fostering of desirable attributes and the deployment of innovative teaching and instructional-design strategies. Part of this iterative process is the obtaining of ongoing feedback from both employers and recent graduates on the generic attributes required of graduates.

It is also crucial that generic attributes should be thought of more broadly than in terms of just university and work. These attributes represent a basis for lifelong learning in all kinds of life situations. Rather than being viewed as discrete attributes that people learn to transfer, generic attributes should be seen as learnt capacities to handle an increasing variety of diverse situations. Thus transfer becomes more a growth in confidence and adaptability as learners experience ever more success in their deployment of generic skills in a range of situations. To put it another way, perhaps it is not so much generic attributes that transfer, as growing understanding of how to deal with different contexts. In this way, nonwork experiences can benefit workplace performance and vice versa.

Conclusion

The competency model, despite the diversity of interpretations and practices associated with it, continues to flourish. Whether it will evolve further and continue to maintain its current position of preeminence in VET remains to be seen. Certainly, there is little evidence of a different model of VET emerging to replace it in the near future.

See also: Assessment in the Workplace of Performance, Developing Expertise and Competence; Assessment in Vocational Education; Impact of Assessment on Classroom Practice; Modularization in Vocational Education and Training; Training and Learning in the Workplace.

Bibliography

- Feinberg, W. (1983). *Understanding Education: Towards the Reconstruction of Educational Enquiry*. Cambridge: Cambridge University Press.
- Gonczi, A., Hager, P., and Oliver, L. (1990). Establishing competency-based standards in the professions. *Research Paper No. 1, National Office of Overseas Skills Recognition, DEET*. Canberra: Australian Government Publishing Service.
- Hager, P. (1994). Conceptions of competence. *Proceedings of Philosophy of Education* **49**, 355–64.
- Hager, P. (1995). Competency standards – A help or a hindrance?: An Australian perspective. *Journal of Vocational Education and Training* **47**(2), 141–151.
- Hager, P. and Beckett, D. (1995). Philosophical underpinnings of the integrated conception of competence. *Educational Philosophy and Theory* **27**(1), 1–24.
- Lum, G. (1999). Where's the competence in competence-based education and training? *Journal of Philosophy of Education* **33**(3), 403–418.
- Rychen, D. S. and Salganick, L. H. (2003). *Key Competencies for a Successful Life in a Well Functioning Society*. Cambridge, MA: Hogrefe and Huber.
- Stasz, C. and Brewer, D. J. (1999). *Academic Skills at Work: Two Perspectives* (Reprinted from the National Center for Research in Vocational Education). Rand Education: Santa Monica.
- Stasz, C., Ramsey, K., Eden, R., Melamid, E., and Kaganoff, T. (1996). *Workplace Skills in Practice, Rand/National Center for Research in Vocational Education*. Santa Monica: CA.
- Wellington, J. J. (1993). *The Work Related Curriculum*. London: Kogan Page.

Further Reading

- Arguelles, A. and Gonczi, A. (eds.) (2000). *Competency Based Education: A World Perspective*. Mexico: Noriega Limusa.
- Gonczi, A. (2006). The OECD and its role in the key competencies debate and the development of lifelong learning. In Hager, P. and Holland, S. (eds.) *Graduate Attributes: Learning and Employability*, pp 105–124. Dordrecht: Springer.
- Hager, P. (2004). The competence affair, or why vocational education and training urgently needs a new understanding of learning. *Journal of Vocational Education and Training* **56**(3), 409–433.
- Hager, P. and Smith, E. (2004). The inescapability of significant contextual learning in work performance. *London Review of Education* **2**(1), 33–46.
- Hyland, T. (1997). Reconsidering competence. *Journal of Philosophy of Education* **31**(3), 491–503.
- Lum, G. (2003). Towards a richer conception of vocational knowledge. *Journal of Philosophy of Education* **37**(1), 1–15.

VOCATIONAL EDUCATION AND TRAINING – VET IN CONTEXT

Contents

Globalization and Vocational Education and Training

Learning to Work for the Future

The Relationship between Technical and Vocational Education and Training and Innovation

The Status of Vocational Education and Training

Vocational Education and Training in Post-Conflict Countries

Globalization and Vocational Education and Training

P Loveder, National Centre for Vocational Education Research, Adelaide, SA, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Globalization has many adherents and detractors (e.g., Friedman, 2005; Bhagwati, 2004; Stiglitz, 2003), including large groups of people opposed to the effects they believe it has on traditional ways of living, especially in developing countries (Klein, 1999).

Defining globalization is a problematic task and usually involves economic, technological, sociocultural, and political contexts. As early as 1994, Friedman came up with a useful literal definition of globalization as the “process of transformation of local or regional phenomena into global ones. It can be described as a process by which the people of the world are unified into a single society and function together” (Friedman, 1994).

Since that time, understanding of globalization has become more sophisticated, such that commentators now characterize globalization as consisting of three phases:

- Phase 1: the mobility of capital.
- Phase 2: the mobility of goods and services.
- Phase 3: the mobility of people.

Each of these phases poses distinct challenges to vocational education and training (VET), in terms of adapting to changes brought about through economic restructuring, the entry and exit of skills through migration flows, and a focus on new types of skills in an increasingly knowledge-based society.

This article seeks to establish the causes for the growing interdependence of nations and the degree to which VET is being impacted by the developments within the global economy. In particular, it addresses issues around patterns of labor mobility and the development of mechanisms to facilitate this phenomenon, including

international education, the growth of higher-level skills, the transparency of qualifications systems, and the transferability of skills.

Economic Restructuring

The free flow of capital and goods and services allows employers the option to relocate investment from one region of the world to another, in anticipation of greater returns. In many Western countries, the most recent examples of this are the closure of factories, with operations typically moving to China or Southeast Asia (low-cost manufacturing), and the establishment of call centers and back-room processing for financial services in India (service-sector work). Workers who are made redundant may be unable to find work in the same industry and may have to acquire new skills to regain employment. This has been the fate of many workers in the textiles, clothing, and footwear industry in some developed nations (Webber and Weller, 2001).

Finegold (2009) suggests there is a process of ‘bifurcation’ of skills and jobs occurring in the global labor market – where there is demand at both the ‘low-end’ and ‘high-end’ of the job market creating a form of ‘barbell’ market. With, presumably, most of the jobs at the low-end concentrated in developing countries, the corollary is that those at the high-end are concentrating in more developed economies that have access to pools of highly educated and skilled people. Many jobs in the middle are simply disappearing.

Further, some estimates suggest that, by 2030, up to 60% of all jobs could be automated, and new industries are beginning to emerge (such as biotechnology, renewable energy, and nanotechnology) which are all requiring

new skills to be developed – and it is these (and other) industries which will provide the high-skill jobs of the future (Elliott, 2007).

At the macro level, governments worldwide have adopted different responses to the economic and social consequences of globalization. Often, assistance is provided to firms and workers in vulnerable industries to enable them to adjust – with the scale and type of assistance provided varying according to the industries affected. The political consequences of allowing adjustment to occur with or without assistance are also an influencing factor.

Another significant strategy adopted by governments is the development/enhancement of workforce skills to maintain or improve the competitiveness of industries – relative to equivalent industries in other countries. Industry bodies – including unions and employer associations – are willing supporters of government policy and initiatives in this area. Examples of these types of initiatives are occupationally linked literacy, language, and numeracy programs.

A further response (which has usually come much later) has been to develop overarching national skills strategies – examples being the Leitch Review in the United Kingdom which resulted in a national skills strategy for Britain through to 2020 and the Lisbon Strategy in the European Union. In Australia, the National Skills Shortages Strategy (NSSS) – and, more recently, the establishment of Skills Australia – are examples of national skills planning in response to global economic factors.

Alongside the restructuring occurring as a result of more open economies has been the development of the new information and communication technologies (ICT), which have been increasingly deployed across all areas of industry and which have had a profound impact in many industries, from computer-aided design and just-in-time management of inventories, to a range of new and enhanced telecommunication services requiring workers to gain entirely new skill-sets.

During the 1990s, the term knowledge economy became a popular way of describing the economies now underpinned by the information and communications technologies. The term had originally been coined by the well-known management expert, Peter Drucker (1968), in his book, *The Age of Discontinuity*. Governments in many countries adopted the notion of a knowledge economy as an overarching goal through which to frame overall policy direction – including in the area of education and training. The European Union, for example, set itself a goal in its 2000 council meeting in Lisbon to become, by 2010, “the most competitive and dynamic knowledge-based economy in the world,” with lifelong learning one of the means of realizing this goal.

An early commentary which anticipates the impact the knowledge economy will exert on people’s working lives

is Reich’s (1991) *The Work of Nations*, in which he categorizes the workforce according to three groups: routine producers, in-person service providers, and symbolic analysts. The last category was defined as those who “solve, identify and broker problems by manipulating symbols. They simplify reality into abstract images that can be rearranged, juggled, experimented with, communicated to other specialists, and then, eventually, transformed back into reality” (p.178). Examples include engineers, lawyers, journalists, academics, and most professionals.

The VET system needs to reflect the general societal changes from an industrial society to a knowledge society as well, and be able to meet the skills needs of a changing labor market. New occupations are arising within sectors such as ICT, the services, and in new technologies such as biotechnology, and these have to be integrated as new training provisions within the VET system.

In addition, the process of economic restructuring has highlighted the disjuncture of VET systems which, in the main, have been set up primarily to deliver skills to entry-level students when, in reality, a component of the existing workforce requires re-skilling. Therefore, it is imperative to explore a range of mechanisms that provide workers, who have already completed their formal training, with the opportunity to acquire new skills (Ball *et al.*, 2001).

Movement of Skills through Migration Flows

Another facet of globalization is the mass movement of people. The United Nations (2003) estimated that some 175 million people lived outside of their country of birth at the beginning of the twenty-first century.

An aspect combined with this phenomenon has been skilled migration. Comparative data (from 2000) suggest that some 20 million highly skilled workers from developing countries (i.e., highly skilled foreign-born workers with tertiary qualifications) now live in Organisation of Economic Co-operation and Development (OECD) member countries alone; a 70% increase in merely 10 years (Beine *et al.*, 2006).

The global trend toward mass movement of skills has increased significantly following World War II. This phenomenon has seen a brain drain to receiving nations – and a perception that more developed nations have been ‘poaching’ the best and brightest (Grecic, 2007).

Recent theoretical literature suggests, however, that emigration prospects can raise the expected return to human capital and foster investment in education at home – although there are clearly winners and losers in this (Beine *et al.*, 2006). In fact, some economists have established that there is an overall increase in the standard

of living for both donor and recipient country through the migration of skilled people.

There are a range of consequences for VET arising from labor mobility, especially in supporting migrants to work to their full capacity. This may include the need for training in language, literacy, and numeracy skills; country-specific health and safety training; and particular vocational skills training. In addition, there could be a need for more practical training on living in a new culture (Learning and Skills Council, 2007).

VET also has a role in ensuring the integration of immigrants and weak learners, and thereby contributing to social cohesion. As informal learning is the dominant form of skills acquisition in some countries and for many learners, these experiences will increasingly need to be incorporated into the more formal system of skills development – especially if these learners are to be fully effective in the modern labor market. In this respect, VET is starting to play an important role in the integration and social policies of many countries.

To ensure good learning outcomes for new migrants, educators and trainers need to be aware of alternative classroom-management strategies. These may include smaller class sizes, different types of institutional support, and placing learners from similar backgrounds together (Burgoyne and Hull, 2007). There are other considerations such as integrating education, training, and employment pathways more effectively. The main barriers to success for many migrants relate to a lack of recognition of the qualifications and experiences they bring from their former countries and restrictions on using overseas-acquired skills through regulation (Phillimore *et al.*, 2007).

International Education Markets

International education markets, according to the World Trade Organization (WTO), are defined as the cross-border trade or ‘consumption abroad’ of education and education-related services. This can be in the form of students attending study in a country other than their own (onshore education), through distance learning, or by undertaking courses offered in a student’s own country by an education provider of a second country (offshore or transnational education). Many countries, such as the United Kingdom, America, and Australia, have been quick to grasp the potential for this – especially in the domain of higher education.

The twentieth century saw a surge in higher education; in the early twenty-first century, the idea of going abroad to study has become thinkable for ordinary students. In 2006, the most recent year for which figures are available, nearly 3 million were enrolled in higher education institutions outside their own countries – a rise of more than 50% since 2000 (Economist, 2008).

(Australia – by comparison with other OECD countries, for example – has by far the highest proportion of overseas students among its higher education intake, at 17.8%, well in excess of the OECD average of 6.9% (OECD 2008, Table C3.1). International education services are Australia’s third largest export industry worth 12.6 billion dollars per annum (ABS, 2007).

Finegold (2009) suggests there are other benefits beyond the influx of dollars that can be generated by global education services, especially through the access to new innovation networks provided through the graduates. In fact, graduates are a major source of potential business incubation and start-up companies, especially in high-technology areas. Benefits also derive to the recipients who take home credentials that are considered world-class and which help local business.

International education has had a direct impact on patterns of mobility as well. For the qualification holder, having Western qualifications often makes migration (especially for skilled-migration purposes) to a new country much easier. Access to Western-country qualifications is also seen to benefit the growth in trade and the rise of the middle class in many countries (e.g., China and India). In countries such as Australia, international education has been a direct route to permanent residency.

The growth of international education has resulted in a range of issues, however, including how to assure quality, the accreditation process where multiple countries’ contexts need to be considered, possible diminished research capacity, and crowding out of native-born populations in the intensified efforts to recruit international students (Lim, 2007; Finegold, 2009).

The move toward internationalizing curriculum which takes into account a wider view of the world – and is aimed at preparing all students for performing professionally and socially in a global context – is becoming increasingly important both in terms of good global citizenship and plain good business! Although difficult to quantify the actual benefits, the value of international education in promoting intercultural development should not be underestimated.

Transparency and Transferability

Another element of globalization and its links to both education and the labor market has been the rush to develop a range of mechanisms to facilitate mobility. In this, the education and training sector has a major role to play.

People increasingly require skills that are portable and transferrable and which, in some circumstances, can be recognized across national borders. In fact, the development of qualifications systems that can accommodate skills and qualifications acquired in other countries are being

seen as a key part of trade diplomacy. In some regions of the world (particularly Europe through the Copenhagen and Bologna processes), efforts are afoot to realign qualifications frameworks to become regional in nature.

The need then for harmonizing country-specific VET systems (such as what is currently occurring through the new European Qualifications Framework (EQF)) has significant consequences for the cultural patterns underlying both the practice, the theory, as well as the policy of VET in these countries, and in addressing incompatibility between systems (Deissinger, 2008). For example, the development of country-specific VET systems happen over a significant time span and are fashioned to meet local and national imperatives. The question of how to then balance these national needs with the wider regional or international agenda often produces conflicting goals.

A strategy for dealing with perceived incompatibility may lie in a nation-specific strategy which encourages and enables change without dumping the benefits and the functionality of the established system (Deissinger, 2008).

Another recent strategy from Europe has been the introduction of the European Credit System for Vocational Education and Training (ECVET – a device which will facilitate the transfer, validation, and recognition of learning outcomes acquired by individuals moving from one learning context to another or from one qualification system to another – in particular, during a mobility period – and who wish to obtain a qualification (European Commission, 2009).

Issues around the recognition of qualifications and transnational quality assurance strategies which require local, national, and regional approaches need to be considered. This will become increasingly pressing as the world continues to divide into major trading blocs with the erosion of traditional borders and trade barriers. These need to be balanced with other issues to do with quality assurance and ensuring training standards are maintained, however.

Innovation and the Growth of Higher-Level Skills

It has been widely argued that global economic competitiveness rests in the main on the knowledge and skills of the workforce (see, in particular, Brown *et al.*, 2001). But the link between global competitiveness and skill formation is complex – especially when other factors are taken into account, such as country context, social justice, and the role of government. In some countries, especially those in Europe, there is a far stronger role for government in driving skill formation – irrespective of global economic factors – than in others (such as the United States).

Bosworth *et al.* (2008) identified globalization as a major factor in placing pressure on less skilled jobs and driving the

policy in Western nations, in particular, toward concentrating on raising education and skills levels in response. This has especially been a feature of many European countries – through the European Lisbon Strategy. This has also been the experience in Australia as well, where successive governments have targeted skill development as a way of maintaining the country's economic performance.

Several recent government reports ('Rising above the storm' from the United States, 'Race to the top' from the United Kingdom, and 'Benchmarking against the global best' from Canada) note the importance of innovation in driving growth in national productivity and competitiveness and that this plays an equally important role in achieving social-inclusion objectives (Finegold, 2009).

Globalization is perceived as a factor in stimulating growth in higher skills as the nature of production becomes more sophisticated (Curtain, 2004). This dictates that as methods of production become more complex, so the need for people with higher level technical and problem-solving skills also increases. Despite there being an accepted view that there is a strong and tangible link between university education and the process of innovation, there is anecdotal evidence that there also needs to be a pool of highly trained technical expertise at the para-professional level – and that this, in itself, can foster the development of entrepreneurial skills in workers; essential for success in an increasingly globalized free economy.

Changes in production, in particular, have resulted in the need for the workforce to acquire higher-level skills; and VET clearly needs to have a focus on the development of higher skills if it is to contribute to this phenomenon. However, the nature and numbers of these higher-level skills required for the economy are difficult to determine, and can only be predicted at a highly aggregated level for a few years ahead. This clearly has implications for training-program development and in a wider national skills-planning sense.

Examples of higher-level qualifications within the rubric of VET include foundation degrees in Britain, professional degrees and Advanced VET in Scandinavia, and vocational graduate certificates and diplomas in Australia.

Conclusion

This article has attempted to demonstrate that the effect of globalization on national VET systems is substantial and complex. A range of international factors such as economic restructuring – coupled with technological development and the growth of ICT and the internationalization of education and skills development – are all playing a part in the move toward the growth of higher-level skills. To date, this has probably impacted higher education more than VET, but this is changing as work practices and occupational skill-sets become more sophisticated.

Second, the trends described in this article all have a significant professional development needs attached to them; and VET practitioners need to be provided with ample opportunity to develop skills for coping with a rapidly changing world.

Finally, it is important to note that it is not correct to attribute all structural change to the process of globalization. There are many factors – including the specific country's economic and social development as well as the prevailing economic, political, and educational institutions – fashioning structural change within the economy. Nevertheless, globalization is a substantial force, and it has direct consequences for the VET sector.

Bibliography

- Ball, K., Lee, Y. -H., Phan, O., and Ra, Y. -S. (2001). *Adult Retraining and Re-Skilling in Australia and South Korea*. Adelaide: NCVER.
- Beine, M., Docquier, F., and Rapoport, H. (2006). *Brain Drain and Human Capital Formation in Developing Countries – Winners and Losers*. University of Luxembourg.
- Bhagwati, J. (2004). *In Defence of Globalization*. New York: Oxford University Press.
- Bosworth, D., Jones, P., and Wilson, R. (2008). The transition to a highly qualified workforce. *Education Economics* 16(2), 127–147.
- Brown, P., Green, A., and Lauder, H. (2001). *High Skills: Globalisation, Competitiveness and Skill Formation*. London: Oxford University Press.
- Burgoyne, U. and Hull, O. (2007). Classroom management strategies to address the needs of Sudanese refugee learners, NCVER, Adelaide.
- Curtain, R. (2004). *Vocational Education and Training, Innovation and Globalisation*. Adelaide: NCVER.
- Deissinger, T. (2008). Institutional and cultural determinants of national VET systems: Problems arising with the European Qualifications Framework (EQF). *Paper Presented at the Australian Vocational Education and Training Research Association Conference*, 11th, 2008, Adelaide, Australia.
- Drucker, P. (1968). *The Age of Discontinuity: Guidelines to Our Changing Society*. New York: Harper & Row.
- Economist (2008). The future is another country, in *The Economist*, December, 2008.
- Elliott, S. (2007). Projecting the impact of computers on work in 2030. *Workshop on Research Evidence Related to Future Skill Demands*, Centre for Education, National Research Council, May 31–June 1, 2007.
- European Commission (2009). *The European Credit System for Vocational Education and Training*, http://ec.europa.eu/education/ecvt/index_en.html (accessed August 2009).
- Finegold, D. (2009). Skills, jobs and the global economy: A comparative perspective. *Paper Presented to the BIG SKILLS Conference*, 3–5 March, 2009, Sydney.
- Friedman, T. (1994). *Cultural Identity and Global Process*. London: Sage.
- Friedman, T. (2005). *The World is flat: A Brief History of the Globalized World in the Twenty-First Century*. London: Allen Lane.
- Klein, N. (1999). *No Logo: Taking Aim at the Brand Bullies*. New York: Picador.
- Learning and Skills Council (2007). Employer perceptions of migrant workers. *Research report, Learning Skills Council*, Coventry.
- Lim, F. (2007). Quality assurance of Australian offshore education: The complexity and frameworks for understanding the problem. *Postgraduate Journal of Education Research* 8(1), 19–36.
- OECD (Organisation for Economic Co-operation and Development) (2008). *Education at a Glance: OECD Indicators 2008*. Paris: OECD.
- Phillimore, J., Goodson, L., Hennesy, D., et al. (2007). Employability pathways: An integrated approach to recognising the skills and experiences of new migrants, National Institute of Adult Continuing Education, Leicester.
- Reich, R. (1991). *The Work of Nations: Preparing Ourselves for 21st-Century Capitalism*. London: Simon & Schuster.
- Stiglitz, J. (2003). *Globalization and Its Discontents*. New York: W.W. Norton.
- United Nations (2002). *International Migration Report - 2002*. Sales No. ST/ESA/SER.A/220.
- Webber, M. and Weller, S. (2001). *Refashioning the Rag Trade: Internationalising Australia's Textiles, Clothing and Footwear Industries*. Sydney: UNSW Press.

Further Reading

- Gregic, V. (2007). The economic impact of migration flows: Sustainable development of the Western Balkans and emigration. *Paper Presented at the Annual Meeting of the International Studies Association 48th Annual Convention*, Chicago.
- Moran, L. and Ryan, Y. (2004). *The Australian Quality Training Framework and Globalisation of Training Markets*. Canberra: ANTA.

Learning to Work for the Future*

J Fien, RMIT University, Melbourne, VIC, Australia

R Maclean, UNESCO-UNEVOC International Centre, Bonn, Germany

D N Wilson[†]

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Citizenship education – A learning programme designed to promote participation in democratic societies by equipping learners with the knowledge, skills, values, and dispositions for active informed civic life.

Economic literacy – The ability to understand and apply basic economic concepts in situations relevant to one's life.

Economic sustainability – An aspect of sustainable development that seek to maintain and enhance economic capital to promote human prosperity and ecosystem health.

Education for sustainable development – A dynamic and expansive undertaking that envisions a world where every person has the chance to benefit from educational opportunities and to learn the skills, knowledge, and values that are necessary to create a sustainable future.

Environmental sustainability – An aspect of sustainable development that seeks to improve the total quality of life, both now and in the future, in a way that maintains the ecological processes on which all human and nonhuman life depends.

Social sustainability – An aspect of sustainable development that encompasses peace, international understanding, gender equality, human rights, labor rights, and corporate governance. In common with environmental and economic sustainability, social sustainability is underpinned by the idea that future generations should have the same or access to resources as the current generation.

Sustainable development – A process by which the needs of present generations can be satisfied without compromising the ability of future generations to satisfy their needs.

Sustainable production – An approach to the manufacturing and delivery of goods and services in ways that respond to basic human needs, promote a better quality of life, and minimize environmental impacts, so as not to jeopardize the needs of future generations.

A Call to Humanity

This article considers some of the implications for technical and vocational education and training (TVET) that flow from the processes of environmental decline that are characteristic of the modern world. The crises of unsustainable development that have followed economic modernization, and which interacts with local, regional, and global patterns of culture, geography, and population, have not been sufficiently analyzed in TVET or, indeed, in broader educational circles. This is despite the fact that, scientists around the globe, including 102 Nobel laureates, signed the World Scientists' Warning to Humanity, which read in part:

Human beings and the natural environment are on a collision course. Human activities inflict harsh and often irreversible damage on the environment and on critical resources. If not checked, many of our current practices put at serious risk the future that we wish for human society and the plant and animal kingdoms, and may so alter the living world that it will be unable to sustain life in the manner that we know. Fundamental changes are urgent if we are to avoid the collision our present course will bring about.

WARNING – We the undersigned, senior members of the world's scientific community, hereby warn all humanity of what lies ahead. A great change in our stewardship of the earth and the life on it is required, if vast human misery is to be avoided and our global home on this planet is not to be irretrievably mutilated. (see <<http://www.secondnature.org/vision/vision.nsf>>)

Building the skills base required understandings to help meet this challenge is a key task of TVET. Integrating TVET into approaches to development that balance economic and social progress, address cultural differences, and respect ecological values and limits is the key to sustainable development. However, efforts to define exactly what sustainable development is must reflect the varying conditions in different parts of the world and their impact upon national and cultural priorities and values. For example, "to an individual living in rural poverty in the developing world, 'sustainable development', if it is to make any sense, must mean increased consumption and a higher living standard. By contrast, to an individual in a wealthy country, with a closet full of clothes, a pantry full of food and a garage full of cars, 'sustainable development' could mean

* Adapted from the discussion paper prepared by the authors for the UNESCO-UNEVOC International Centre, Bonn (UNESCO-UNEVOC, 2006).

[†] Deceased.

more modest and carefully considered consumption” (UNESCO, 1997: para 25).

Thus, sustainable development is not a fixed concept; rather, it is a culturally directed search for a dynamic balance in the relationships between social, economic, and natural systems, a balance that seeks to promote equity between the present and the future, and equity between countries, races, social classes, and genders. The interdependence of people and the environment requires that no single development or environmental objective be pursued to the detriment of others. The environment cannot be protected in a way that leaves half of humanity in poverty. Likewise, there can be no long-term development on a depleted planet.

This makes sustainable development more a moral precept than a scientific concept, and links it as much with notions of peace, human rights, and fairness as it does with theories of ecology or global warning. Indeed, while sustainable development involves the natural sciences, policy, and economics, it is primarily a matter of culture: it is concerned with the values people cherish and with the ways in which we perceive our relationship with others and with the natural world (UNESCO, 2002).

The Crucial Role of TVET

Business and industry today often speak of the triple bottom line of their operations. This is a recognition of the integration of social, economic, and environmental concerns and a move away from business as usual, when the priority was only on economic growth and profit. In integrating social, economic, and environmental concerns, sustainable development involves new ways of thinking about how we use resources to provide the goods and services we need. Thus, Salim (2002) from Indonesia, who was the chair of the United Nations World Summit on Sustainable Development, argues that:

... there are serious shortcomings in the way development has taken place in the 20th century ... development has followed only the economic track and has left behind social and environmental stability, resulting in rising poverty, inequality in income and development and natural disasters through rising flood levels affected by sea rise due to global warming. Development as implemented in the 20th century was not sustainable. (p. 16)

Yet, these problems were not caused intentionally. Rather, they were the results of well-meaning development initiatives that, unfortunately, were implemented without a full understanding of the interlocking nature of social, economic, and environmental issues. Education and training for sustainable development is a process of learning how to make decisions that consider the long-term future of the economy, ecology, and equity in the

workplace and wider community. Building the competencies and commitments needed for such future-oriented thinking is a key task of TVET for sustainable development. This requires attention to the environmental, economic, and social pillars of sustainability. This article explores the implications for TVET of each of these three.

TVET and Environmental Sustainability

Using resources wisely and minimizing waste and pollution are central to ensuring that the natural environment will be able to continually supply business and industry with the natural resources and energy needed for economic activities. Indeed, there can be no long-term economic growth on a planet depleted on natural resources, too infertile to support the production of the plant and animal products upon which people and industry depend, and too polluted for humans to enjoy a healthy and productive life. Environmental sustainability requires a change from business as usual to the responsible use of raw materials, energy, water, etc., awareness of the impacts of production processes, and careful management to minimize any unintended results of production. A number of approaches have been developed to support sustainable production, including: cleaner production, environmental auditing through ISO14000 quality assurance, life-cycle analysis, ecodesign, industrial ecology, the Natural Step, natural capitalism, Factor 4, and so on (see International Institute for Sustainable Development, 2001; UNEP, 2002). However, all share several common principles as illustrated in **Figure 1**.

However, it must be constantly borne in mind that, despite the growing adoption of such principles and associated tools by business and industry, the resource efficiencies that result may not be enough to ensure long-term environmental sustainability. Drawing upon a United Nations Environment Programme (UNEP) report on the future of the global environment, Yencken (2000) argues that:

Given the assumed growth of the world economy between 1990 and 2050, a 4 to 5 fold increase in overall ‘efficiency’ is needed just to keep environmental pressure at its current level. In many cases this will not be sufficient to prevent environmental degradation of environmental resources. Depending on assumptions made with respect to population growth, economic growth and levels of sustainable use of various natural resources, the actual efficiency improvements required within the next half century are estimated to be in the range of five to twenty fold. This can only be achieved by increase in technological efficiency and the ‘dematerialisation’ of production and consumption. (p. 10)

Principle	Application
Nature is not subject to systematically increasing concentrations of substances extracted from the Earth's crust, for example, fossil fuels, heavy metals, and minerals.	Substituting scarce minerals with ones that are more abundant; using all mined materials efficiently; and systematically reducing dependence on fossil fuels.
Nature is not subject to systematically increasing concentrations of substances produced by society, for example, PCBs, CFCs, HCFCs, and PVC.	Substituting persistent and unnatural compounds with ones that are normally more abundant and/or biodegradable.
Nature is not subject to systematically increasing degradation by physical means, for example, dams to change river flows, clearing rainforests, intensive monoculture agriculture, and overfishing.	Drawing resources only from well-managed ecosystems; systematically pursuing the most productive and efficient use both of those resources; and exercising caution in all modifications of nature.
Human needs are met worldwide, for example, access to food, water, healthcare, shelter, and education.	Using all our resources efficiently, fairly, and responsibly so that the needs of all people on whom we have an impact, and the future needs of people who are not yet born, stand the best chance of being met.

Figure 1 Principles and applications for environmental sustainability. From Smith (2002).

Environmental imperatives such as these call for the integration of environmental sustainability into all aspects of TVET. This will involve designing programs and courses that not only provide technical expertise for a student's chosen industry but also include the following objectives:

- developing an understanding of a relevant range of environmental concepts (see **Figure 2**);
- developing an ethic of care and responsibility, and encouraging reflection on the effects of personal values and lifestyle choice; and
- promoting skills for critical thinking and practical action.

TVET and Economic Sustainability

Unfortunately, TVET in many countries remains locked into the role of being a supplier of skilled labor to industry. Anderson attributes this to the culture of productivism in TVET “which presupposes that economic growth is a permanent and necessary feature of human existence, regardless of its environmental impact and consequences” (Anderson, 2000, 2003). Giving precedence to economic interests, productivism subordinates the needs of individual learners to those of industry and prioritizes work and employability over the noneconomic outcomes of TVET. This has resulted in TVET being seen as training for growth and skills for work, two goals for TVET that are antithetical to the broader general education for personal autonomy, citizenship, and sustainability that TVET also needs to serve.

Nevertheless, the economic aspects of sustainability are very important but, instead of productivism, they look to TVET systems, institutions, and instructors to ensure that students and workers develop a different,

wider, set of economically related knowledge, skills, and attitudes. These include:

- *Economic literacy.* This involves using appropriate economic ways of thinking and problem solving that lead to sound and informed economic choices as consumers, producers, savers, and investors and as effective participants in the local, national, and global economy. It involves understanding how changes in government policies, in taxation, interest and exchange rates, and in demographic and market trends will impact upon decisions to be made by individuals, families, communities, and enterprises. In relation to the world of work, economic literacy involves an understanding of sustainable production and ways in which resources can be conserved, waste managed through recycling and reuse, and toxic waste and pollution minimized and controlled (National Council for Economic Education, 2002).
- *Sustainable production.* This is an approach to the manufacturing and delivery of “goods and services in ways that respond to basic human needs and bring a better quality of life, while minimizing the use of natural resources, toxic materials and emissions of waste and pollutants over the life cycle, so as not to jeopardize the needs of future generations” (Norwegian Ministry of the Environment, 1994).
- *Sustainable consumption.* This is the necessary corollary of sustainable production. The people of the developed countries are among the world's largest consumers of natural resources and their production and consumption patterns have major environmental, social, and economic impacts around the world. As indicated in the 2004 *State of the World Report*, the 20% of the world's population living in Organization for Economic Cooperation and Development (OECD) countries earn 85% of the world's annual income; consume 75% of global energy and over 80% of other resources annually; and generate

- *Sustainable development*: A process by which the needs of present generations can be satisfied without compromising the ability of future generations to satisfy their needs.
- *Carrying capacity*: The capacity of ecosystems to support continued growth in population numbers, resource consumption, and waste production.
- *Ecospace*: The total amount of energy, land, water, and other resources that can be used regionally or globally without environmental damage, disadvantaging the capacities of others to meet their basic needs or impinging on the rights of future generations.
- *Ecological footprint*: The area of land and water needed to support the total flow of energy and materials consumed by a person, household, community, or work place.
- *Natural capitalism*: An approach to managing workplace processes in ways that restore, conserve, and expand natural resources (stocks of natural capital), use, recycle, and reuse resource inputs as efficiently as possible, and assume responsibility for making products last longer and easier to dismantle for reuse or recycling.
- *Eco-efficiency*: A strategy for maximizing the productivity of material and energy inputs to a production process while also reducing resource consumption and waste production and generating cost savings and competitive advantage.
- *Life-cycle analysis*: A management tool for identifying the net flows of resource and energy used in the production, consumption, and disposal of a product or service in order to leverage eco-efficiency gains.
- *Triple Bottom Line reporting*: An approach to corporate accounting that reports not only on financial matters but also the outcomes of a firm's environmental and social activities.
- *Environmental management system*: A coordinated approach to ensuring that all environmental issues are taken into account in the workplace and regularly monitored and improved to ensure compliance.
- *The 5 Rs*: Reduce, reuse, renew, recycle, and rethink!
- *Local-global links*: The recognition that the production and consumption of a product or service in one part of the world is dependent on flows of energy and materials in other parts of the world and that this creates potential opportunities and losses economically, socially, and environmentally at all points in the local-global chain. The brain drain from developing to developed nations also affects sustainability in TVET and education.

Figure 2 Concepts of environmental sustainability relevant to TVET. From UNESCO (2002). *Education for Sustainability: Lessons Learnt from a Decade of Commitment from Rio to Johannesburg*. Report to World Summit on Sustainable Development. Paris: UNESCO.

75% of annual global pollution (Worldwatch Institute, 2004). Too much consumption also affects health with over 50% of people in countries such as the US and the UK reported as being overweight (Barber, 2001). While these problems are disturbing, they pale beside those of the 2.8 billion around the world who do not have adequate access to food, water, and shelter. Consumption provides one of the most potent examples of the inequalities that exist today. Thus, a key aspect of education and training for sustainable consumption is the integration of concepts of sustainable consumption into the planning of workplace production processes and in the daily lives and expectations of what it means to live a good life in the minds of workers and their families.

Social Sustainability and TVET

Social sustainability is the third pillar of sustainable development. On both the global and local scale, social sustainability involves ensuring that the basic needs of all

people are satisfied and that all, regardless of gender, ethnicity, or geography, have an opportunity to develop and utilize their talents in ways that enable them to live happy, healthy, and fulfilling lives.

Sustainable livelihoods are central to social sustainability and meaningful work plays an important role in this. The concept of sustainable livelihoods embraces existing concepts of work and employment but widens them to include the multiple forms of economic and noneconomic activities through which people create opportunities to sustain themselves, their families, and their communities. The United Nations Development Programme (UNDP) defines livelihoods as “the assets, activities and entitlements which people utilize in order to make a living” – with assets including local natural resources (i.e., land, water, common-property resources, flora, and fauna), as well as social (i.e., community, family, and social networks), political (i.e., participation and empowerment), human (i.e., education, labor, health, and nutrition), physical (i.e., roads, clinics, markets, schools, and bridges), and economic resources (i.e., jobs, savings, and credit).

The wide view of resources and abilities in the concept of sustainable livelihoods raises questions about traditional person–job relationship that forms the foundation of many approaches to TVET. While education, particularly TVET, has important roles in developing the social, human, and physical capital needed for a sustainable livelihood, it is perhaps a very different form of TVET that is required (Lawrence, 1997). Nevertheless, it is important to ensure that young people receive the best education possible to prepare them for a life of productive employment and to have the entrepreneurial skills not only to develop work opportunities for themselves and others but also to have the commitment and initiative to contribute to the social, economic, and environmental well-being of their communities.

Thus, basic education is central to effective TVET. Literacy and numeracy are vital here. The health and safety of workers often depends upon their ability to read instructions (e.g., on fertilizer bags) and to make accurate calculations (e.g., of mixing and application levels). The wider skills of scientific and social literacy are also important for, for example, equipment maintenance and repair and understanding technological change (scientific literacy) and for group work, dialog and negotiation with colleagues and supervisors, gender and ethnic tolerance, and other skills needed for harmonious relations in the workplace (social literacy). The application of such literacies to the world of work and active citizenship needs to become core dimensions of TVET if it is to respond to the imperatives of social sustainability (Lawrence and Tate, 1997).

Thus, Quisumbing argues for an “holistic and integrated human resource development program for TVET” that “aims to prepare the individual to become a responsible, free and mature person, equipped not only with the appropriate skills and know how of the latest technologies, but also with deep human and spiritual values and attitudes – a sense of self worth, self esteem and dignity” (Quisumbing, 2001). Central to the development of knowledge, skills, and attitudes for social sustainability, she argues, are the abilities:

- To work by oneself and with others in teams, with integrity and honour, with honesty, punctuality, and responsibility;
- To adapt to varying situations; to know and understand problems and issues; to work out solutions creatively;
- To resolve conflicts peacefully;
- To have a good grasp of the reality of the world, of oneself and of others;
- To possess some general knowledge with specialization in some field or area of work; and
- To continue learning and pursue lifelong education in a learning society (Quisumbing, 2001).

A focus on the knowledge, skills, and attitudes for social sustainability can develop all the powers and faculties of the individual – cognitive, affective, and behavioral – and

from them can flow such “work values and attitudes as creativity and adaptability, productivity, quality and efficiency, patience and perseverance, loyalty and commitment, freedom and responsibility, accountability, the spirit of service, a futures orientation, and a genuine love for work itself be developed” (Quisumbing, 2001). This view places ethics at the heart of developing social sustainability through TVET. Wonacott notes that the literature on ethical issues in TVET is most often concerned with dilemmas in teaching and the use of technology related to questions of power, access, control, intellectual property rights, privacy, equity, speech, etc. Ethical and legal issues for specific occupations are also often addressed (Wonacott, 2001). There are also definite ethical and moral implications associated with social sustainability. Some of these include:

- *Respect for cultural diversity.* This is a core value in social sustainability. All people have the right to employment regardless of their ethnic or racial heritage and their religious beliefs. The rights to employment of indigenous peoples are especially important. This also applies to opportunities for further training and promotion. The internationalization of the workforce through globalization and labor migration also emphasizes the importance of developing respect for cultural diversity in all TVET programs (Pegg, 1997; Brown, 2002).
- *Gender equality.* This is also a core value in social sustainability. The rights of women to equality of outcomes from education and training (as well as access) and to equality of employment opportunities, working conditions, and access to further training and promotion are important human rights that need to be enshrined in TVET programs. The vital importance of freedom from discrimination and sexual harassment, associated monitoring, reporting, and disciplinary processes also needs to be taught. These are matters for both male and female students and workers: women need training in ways of protecting their rights and freedom in the workplace, while men need training in their obligations to respect and honor all their work colleagues (Scott, 2003).
- *Inclusion.* The inclusion of other excluded groups – disabled, ethnic minorities, etc. – is essential for the attainment of sustainability in TVET and workplaces. Without the inclusion of all groups in a society sustainability will not be achieved.
- *Workplace relations.* One positive result of the reduction in levels of management and the increase in workers’ levels of educational attainment has been the empowerment of workers to advise management of better ways to operate or produce finished goods. This reduction from as many as eight to as few as three levels of management has improved communications between labor and management. Historically, communications between employers and employees has been mainly

top-down. Increasingly it has become the practice of enlightened employers to elicit – and utilize – ideas from their employees that improve production and lessen waste. The same holds true for the creation of sustainability, both in TVET and in the workplace. Relations between coworkers also benefit from improved communication, as well as from tolerance of others' differences. It goes without saying that a contentious workplace is not likely to be a sustainable one.

- *Teamwork at the workplace.* A harmonious workplace is one at which teamwork is both valued and practiced. The restructuring of assembly lines during the industrial age, and the maintenance of many assembly-line principles and practices in the emerging information age – in particular, the assembly of electronic equipment of all types – necessitate the enhancement of teamwork principles to ensure sustainability. Productivity measures, for example, the failure rate of assembled equipment, highlight the importance of teamwork. Therefore, it is incumbent upon TVET institutions to foster the necessary climate and/or culture of teamwork right from the initial entry of students and trainees into TVET institutions. It is also imperative that TVET teachers and instructors set a correct example by functioning as a team.
- *Relations between employers and employees.* Conflict between labor and management has been a long-standing impediment to harmonious relationships between employers and employees. However, in some countries, enlightened employers recognize that harmony is directly related to improved productivity, reduced spoilage, and even innovations suggested by employees.
- *Safety.* Considerations of safety are of prime importance in TVET and at the workplace. Employers bear responsibility for the working conditions and well-being of their employees. Employees are responsible for actions that might place their peers in peril, produce dangerous or substandard goods, or damage property. This suggests that another aspect of safety is the protection of TVET students/trainees and employees at the workplace.
- *Citizenship.* Social sustainability depends upon the willingness of people to cooperate in building and safeguarding a fair and democratic society. Reciprocal rights and responsibilities are important in a democracy, where the collective voice of citizens is the source of all legitimate authority. These rights include: equality before the law and the freedom to vote, to speak freely on public issues, and to participate in public interest groups. The duties of responsible citizenship include: paying taxes, obeying laws, demonstrating commitment and loyalty to democratic ideals, constructively criticizing the conditions of political and civic life, and participating to improve the quality of national and community life (Klusmeyer, 1996).

The rights and responsibilities of citizens extend to the workplace as well. This is why respect for gender and

cultural differences and skills for developing harmonious workplace relations, teamwork, and negotiating improvements in work practices are so important to social sustainability. TVET has key responsibilities to ensure that these civic disposition and participation skills are developed with experience, suggesting that this can perhaps best be done through the following kinds of learning experiences:

- student participation in democratically conducted student organizations;
- college-facilitated community service that is connected directly to the curriculum and classroom instruction; and
- cooperative learning activities in which groups of students cooperate to pursue a common goal, such as inquiring about a public issue or responding to a community problem (Patrick, 1999; Battersby, 1998).

Conclusion

This article has examined the role of TVET in addressing three pillars of sustainability. Although all aspects of sustainability have been analyzed separately, there are very close links and interrelationships between them. It is proposed here that an integration of environmental, economic, and social sustainability in TVET can only be undertaken – and achieved – through the inclusion of a relevant range of environmental, economic, and social concepts into the content of teaching (by identifying eco-footprints, life-cycle analysis, toxic materials, etc.). There is also a pressing need to orient students toward sustainable production; through the appropriate development of students' responsibility, their personal values, consumption patterns, and appreciation of cultural diversity; and through development of generic skills of critical thinking, team work, workplace relationships, and safety, and their ability to adapt to varying, often very demanding situations.

It is also highlighted in this article that sustainable development is not a fixed concept, but culturally and contextually directed activities for achieving an appropriate balance in the relationships between social, economic, and ecological systems.

Therefore, TVET's role is considered not only as building the skills base relevant to the sustainable practices of particular occupations, but also in developing understanding and appreciation of all aspects of sustainability.

Bibliography

- Anderson, D. (2000). VET and ecologism: Charting the terrain. *Enriching Learning Cultures. Proceedings of the 11th Annual International Conference on Post-compulsory Education and Training*, vol. 1, pp 17–23. Centre for Learning Research, Griffith University.

- Anderson, D. (2003). From productivism to ecologism: Dilemmas, issues and strategies for VET. *Australian Vocational Education Review* 10(1), 1–14.
- Barber, J. (2001). *The Sustainable Production and Consumption of Food: An NGO Perspective*. Washington, DC: Integrative Strategies Forum.
- Battersby, M. (1998). Education for citizenship: Service-learning and the reflective citizen. *Learning Quarterly* 2, 3–6.
- Brown, B. L. (2002). *Trends and Issues Alert No. 35: Global Mobility of Workers*. Columbus, OH: Clearinghouse on Adult, Career, and Vocational Education.
- International Institute for Sustainable Development (2001). *Instruments for Change: Definitions and Concepts*. <http://www.iisd.org/susprod/principles.htm> (accessed June 2009).
- Klusmeyer, D. B. (1996). *Between Consent and Descent: Conceptions of Democratic Citizenship*. Washington, DC: Carnegie Endowment for International Peace.
- Lawrence, J. (1997). Adult education and jobs, or sustainable livelihoods? *Presentation at UNESCO Panel on Changes in the World of Work*. CONFINTA V: Hamburg.
- Lawrence, J. and Tate, S. (1997). Basic education for sustainable livelihoods: The right questions. *Discussion Paper Presented to the United Nations Development Programme International Working Group on Sustainable Livelihoods*. New York: Pearl River.
- National Council for Economic Education (2002). *Campaign for Economic Literacy: Challenge and Response*. <http://www.ncoee.com/cel.html> (accessed June 2009).
- Norwegian Ministry of the Environment (1994). *Report of the Symposium on Sustainable Consumption*. Oslo: OECD.
- Patrick, J. J. (1999). *The Concept of Citizenship in Education for Democracy*. <http://library.educationworld.net/a10/a10-145.html>.
- Pegg, L. C. (1997). Diversity training and education in the work place. *Journal for Vocational Special Needs Education* 19(2), 62–66.
- Quisumbing, L. R. (2001). The importance of values education for TVET and its economic and human resource development program. *Paper Presented at the UNESCO Asia Pacific Conference*. Adelaide.
- Salim, E. (2002). The journey of hope: From Rio to Johannesburg. *Words into Action. For the UN World Summit on Sustainable Development, Johannesburg*. 26 August–4 September 2000. London: International Institute for Environment and Development (IIED).
- Scott, M. L. (ed.) (2003). *Equity Issues in Career and Technical Education. Information Series, No. 390*. Columbus, OH: Clearinghouse on Adult, Career, and Vocational Education.
- Smith, J. (2003). *Redesign of Government Sustainability Education Programs for Business Personnel – From Awareness Raising to Changing Behaviour*. Unpublished PhD thesis, University of New England, Armidale.
- UNEP (2002). *Industry as a Partner for Sustainable Development, 10 Years after Rio: The UNEP Assessment, Extended Executive Summary*. Paris: UNEP.
- UNESCO (1997). Environment and society: Education and public awareness for sustainability. *Background Paper Prepared for UNESCO International Conference*, Thessaloniki.
- UNESCO (2002). *Education for Sustainability: Lessons Learnt from a Decade of Commitment from Rio to Johannesburg*. Report to World Summit on Sustainable Development. Paris: UNESCO.
- UNESCO-UNEVOC (2006). *Orienting Technical and Vocational Education and Training for Sustainable Development*. Available at <http://www.unevoc.unesco.org/pubsdir.php?akt=id&st=&id=ISSN+1817-0374&lg=en>.
- United Nations Development Programme (UNDP) (nd). Sustainable Livelihoods. <http://www.eldis.org/go/topics/dossiers/livelihoods-connect/what-are-livelihoods-approaches/evolution>.
- Wonacott, M. (2001). *Trends and Issues Alert, No. 24: Ethics: The Role of Adult and Vocational Education*. Columbus, OH: ERIC Clearinghouse on Adult, Career and Vocational Education.
- Worldwatch Institute (2004). *State of the World 2004: The Consumer Society*. <http://www.worldwatch.org/pubs/sow/2004>.
- Yencken, D. (2000). Sustainable Australia: Refocusing government. *TELA Paper, Issue 3*. Melbourne: Australian Conservation Foundation.

Further Reading

- Baines, J., Cohen, J., and Martin, S. (2005). United Kingdom: Skills development for the twenty-first century. *Prospects* 35(3), 355–365.
- Blewitt, J. and Cullingford, C. (eds.) (2004). *The Sustainability Curriculum: The Challenge for Higher Education*, pp 1–9. London: Earthscan.
- Chinien, C. (2003). Skills to last: Broadly transferable sustainable development skills for the Canadian workforce. *A Technical Research Report for Human Resources Development Canada, Human Resources Partnerships, National Centre for Workforce Development*. <http://www.unevoc.unesco.org/sustainable/docweb/SkillsToLast.pdf>.
- Doppelt, B. (2003). *Leading Change Toward Sustainability. A Change – Management Guide for Business, Government and Civil Society*. Sheffield: Greenleaf Publishing.
- Egan Review (2004). *Skills for Sustainable Communities*. London: Office of the Deputy Prime Minister, UK Government.
- Falk, I. and Kilpatrick, S. (2002). Lifecycles and leadership: Partnerships for VET. Making a world of difference. *5th Australian Vocational Education and Training Research Association. National Conference 20–22 March 2002*. <http://www.avetra.org.au> (accessed June 2009).
- Gagliardi, R. and Alftan, T. (1993). *Environment Training: Policy and Practice for Sustainable Development*. Geneva: ILO.
- Hilton, M. and Smith, D. (2001). Professional education and training for sustainable development in SME'S. *European Foundation for the Improvement of Living and Working Conditions*. Luxembourg: Office of the Official Publications of the European Union.
- National Centre for Sustainability (2005). *Guideline Standards for Sustainability, Swinburne University*. http://www.swinburne.edu.au/ncs/researchprojects/knowledge/Guideline_Standards_for_Sustainability.pdf (accessed June 2009).
- Organisation for Economic Cooperation and Development (2000). *Human and Social Capital and Sustained Growth and Development*. Paris: OECD.
- Plane, K. (2006). Towards a national sustainable development framework for VET in Australia: Greening small business and communities, or greening the marketplace? http://www.avetra.org.au/publications/documents/PA0003_000.pdf (accessed June 2009).
- Quisumbing, L. R. (2005). Education for the world and citizenship: Towards sustainable societies. *Prospects* 35(3), 289–301.
- Scott, W. A. H. and Gough, S. R. (2004). *Key Issues in Sustainable Development and Learning*. London: RoutledgeFalmer.

Relevant Websites

- <http://www.secondnature.org> – Environmental Responsibility in Academic Institutions.
- <http://www.unevoc.net> – International Centre for Technical and Vocational Education and Training, UNESCO-UNEVOC Education for Sustainability.

The Relationship between Technical and Vocational Education and Training and Innovation

K Ruth and L Deitmer, University of Bremen, Bremen, Germany

© 2010 Elsevier Ltd. All rights reserved.

Understanding of Innovation

The notion of innovation is rather complex and multifaceted, because it is not only describing purely economic change in companies' policy and business strategies but also referring to organizational and technical change. Thus, in a very broad sense, innovations could also be thought of quite simply as changes – such as the practical implementation of alternative concepts for work organization (e.g., group work). The example of group work quite clearly illustrates that innovation can be assessed from different perspectives: from the perspective of workers, for example, group work may signify an improvement when compared to their previous working conditions; for some companies, however, the expected benefits fail to materialize – that is, no increase in productivity is generated (Manske *et al.*, 2002).

It is for these reasons that we would like to use innovation in a broader understanding to find out whether or not there are relationships between innovation and the vocational education of workers. The notion of innovation as a social process, which includes many different actors, is widely accepted in innovation research (e.g., Lundvall, 1992; Morgan, 1997). In the classical understanding of innovation as companies' process to develop and commercialize new products or to launch new processes in order to improve product quality or lower the production cost, the view was very strictly limited to the boundaries of companies where engineers were the experts for innovation (or more precisely, for new product development or the design of processes). Among others, Morgan (1997) expresses the concerns with this limited understanding of innovation by pointing out that "... the weakness, which is still prevalent in the western countries today, stems from an elitist conception of knowledge in which scientific knowledge is extolled, while lower forms of knowledge are undervalued" (Morgan, 1997: 493).

It was the work of von Hippel (1988) on the sources of innovation that drew the attention to the hitherto unrecognized innovation actors – users and suppliers. He opened the view to network-based innovation processes by considering other companies (suppliers or customers) as important actors in the innovation arena. By his work, innovation experts who were not recognized up to then were brought into the game. These new experts also introduced new knowledge domains into the innovation processes: in von

Hippel's case studies, key users of the machinery or aircraft industry had the opportunity to mainly bring in their practice- and experience-based expertise into the development process, and similarly supplier companies were given the chance to include manufacturing and material know-how into the overall innovation process (von Hippel, 1988). These tendencies have nowadays developed toward innovation networks (Duschek, 2004), which today are a widespread form of organizing innovation processes. Innovation networks can have different forms: either intra-organizational or inter-organizational; they can either be local and regional or global; and their communication and cooperation can range between direct contacts (meetings) and virtual/electronic communication and collaboration or a mix of these forms. Finally, the innovation networks do gather a wide range of people with different educational and professional backgrounds. These different actors contribute different types of knowledge, skills, and expertise, which add new dimensions to innovation processes.

An important step in opening the innovation arena was targeting design to manufacturability: this meant to leave the sequential track, which saw the manufacturability of a new product as the final problem to be solved (if at all targeted). It was the Japanese who taught the world that manufacturability is a primary goal of product innovation and that in order to ensure a high level of manufacturability, the production planning and shop-floor people need to be included in very early stages of the innovation cycle (Imai, 1991). These changes widened up the formerly closed monastery of design monks – that is, the engineers as key innovation actors – to a lively bazaar of many actors with diverse backgrounds and a diversity of design and innovation goals. Since knowledge creation is widely accepted and studied as a core activity of innovation (Nonaka and Takeuchi, 1995; Nonaka and Nishiguchi, 2001), the focus of the following paragraphs is on knowledge capacity development, particularly on knowledge sharing.

Shared knowledge is one of the important resources for successful innovation. Therefore, those organizations or those networks or teams, which have established the best structures and methods to develop a common understanding among individuals representing different functions (organizations or disciplines) are the ones with the best innovation performance and thus with competitive

advantages (Hong *et al.*, 2004). The synergy of shared knowledge cannot easily be achieved, because knowledge does not have a unique nature. There can be different cognitive universes between organizations or departments; there can be incompatible knowledge domains between departments or teams; knowledge can be tacit in its nature and thus difficult to codify or to make explicit (Senker, 1993); and, eventually, a lot of knowledge is sticky, that is, held by individuals and cannot be separated or objectified. Thus, we observe plenty of obstacles to knowledge sharing, which impede knowledge sharing as a planned process. But nevertheless, good practice with regard to knowledge sharing is reported (e.g., Nonaka and Teece, 2001; Harryson, 1998).

Morgan (1997) pointed out the interactiveness and the multiactor-based character of innovation processes: “There is now growing support for the view that innovation is an interactive process between firms and the basic science infrastructure, between the different functions within the firms, between users and producers on inter-firm level and between the firm and the wider institutional milieu and that this process should be conceived as a process of interactive learning in which a wide array of institutional mechanisms can play a role” (Morgan, 1997: 493).

Taking into account the above-sketched circumstances, we need to consider two influential aspects of contemporary innovation:

1. The idea of sequential innovation steps – concept design, system design, detail design, prototype testing and refinement, and manufacturing – has made room for the conception of concurrent innovation which is including many actors.
2. Particularly, the hitherto well-accepted distinction between radical and incremental innovation is difficult to maintain in the future, if trying to figure out the role of practical knowledge and experience in innovation processes.

The latter refers mainly to the insight that many radical or breakthrough innovations in history have not been made by scientists or engineers, but rather through technicians or highly skilled craftsmen. As an example we might think of the technological progress before 1850, particularly in the case of the steam engine that was developed and produced by craftsmen, skilled workers, or technicians who deployed a knowledge that originated from practice rather than from science (Mokyr, 2002; Rolt, 1986). Similarly, practitioners developed the transistor by advancing through trial-and-error methods, which eventually rendered a result (Braun and MacDonald, 1978; Rosenberg, 1982; Senker, 1993). Thus, the assumption that radical innovation is based on (theoretical and scientific) knowledge and incremental innovation thrives on skills/experiential knowledge is nowadays increasingly

obsolete – not to mention the difficulty to distinguish radical from incremental innovation.

The aforementioned evolution from sequential to concurrent (or at least overlapping) innovation phases has an important impact on the necessity to share knowledge: knowledge and experiences that are available in functional areas at the end of the formerly sequential innovation process (e.g., not only manufacturing expertise, but also suppliers’ capacities) are now required at the early stages (concept design) and during the ongoing innovation process. The particular case of knowledge sharing between manufacturing and early design stages is alike with the need to bring workers’ knowledge into the design process.

The following section picks up the findings of this special case and feeds them into a more general concept of experience as an innovation factor.

Workers’ Experience as an Important Source of Innovation

The contact point of technical and vocational education and training (TVET) and innovation is the quality of production and products. Industries that are still in the field of large-scale production with little variants and a rather fixed mode of production will need workers who exactly just operate the handlings on the product they are obliged to do. There is little maneuver for any change and the impulses from the production line onto changes or incremental improvements are very limited. Eventually, quality control departments are needed to assure the quality of the products and the work done by the production workers. However, in a post-Fordist era of production, there are more complex and saturated markets to be dealt with. Industrial and private costumers get a bigger say in the final function of the product and wish to receive tailor-made solutions. This requires new production concepts, which assign workers an important role as reflective practitioners and which take their experiences and knowledge as an important source for improving production. Industries that follow this one-of-a-kind product philosophy will increasingly need highly skilled workers who are competent to act on comprehensive work principles. These principles imply that the workers are able to do the planning and control of their own work – beside the factual production itself. Therefore, these workers must hold planning and control competencies. Besides, they must also be able to undertake production process improvements, which require innovative skills that enable them to shape the production processes. These skills are needed to avoid pitfalls and dysfunction in the production process which will occur when new products are to be introduced – a situation that comes up in a one-of-a-kind production model quite often. Equipped with shaping skills (Rauner, 2007) they must be

able to react on challenges for improving the product and its efficient production. Technical and social communication skills with planners or engineers are needed, concerning improvements of the manufacturability or the design-to-assembly quality of the different products (Ruth, 2006).

Looking closer at the relationship of workers' abilities and innovation, there is a gap between the designated engineers' models of machine systems design and that of the actual operation mode of these systems at the side of production and at the side of the workers. The engineering and the production knowledge represent different knowledge domains, one representing the planning knowledge and the second representing the practical knowledge in producing the products. This accumulation of experience and knowledge is very often used to improve the machines or to make the operation more efficient. The direct machine workers as well as the maintenance workers are in that respect some very important innovation actors who might contribute to both, process and product innovation.

In the case of process innovation, there are production systems, which are more tolerant toward different operation modes and that are in every stage under the immediate control of workers (e.g., short assembly lines for simple products). Such easygoing systems are an ideal playground for influencing (or in other words: improving the processes) by production line workers. Here, workers can apply their expertise which is to a large extent experience based – experience gained through daily operation in the production line. This is somewhat different in the case of large technological systems (such as nuclear power plants, or chemical processes), where controlling and maintaining the systems is a complex task, which additionally is characterized by a separation of the real processes and the actual operators. In other words, the processes are of reach of the immediate senses, but nevertheless, their basic source of building expertise (which contributes to the construction of innovation capability) is experience. But, experience itself – though a potential wellspring of expertise – always covers the risk of blinding people and narrowing their minds by lulling them into a false sense of security about their hypotheses (on the operation, their tasks), their skills, and their expertise and mastery (e.g., the case of a nuclear power plant accident in Japan is analyzed in Ruth, 2001: 177). In order to avoid wrong or at least misleading hypotheses on the operation mode of complex production systems, it is essential that the operators do also hold a certain level of theoretical knowledge, which ideally is complementary to the experiences – or rather, theory is the backing into which experience is embedded. At this stage of the argument, it can be stated that education and training systems and respective qualification frameworks at best need to cover and address both, theoretical and practical knowledge. A closer look at different types of vocational education and training

(VET) systems will provide insights into the relationships between favorable (or disadvantageous) combinations of theoretical and experience-based knowledge, and the level of innovativeness. Before approaching this issue, the field of product innovation is touched upon briefly.

Similar to the sketched field of process innovation, the production workers can play an important role in product innovation. If we consider the manufacturing of machinery as the arena, the actual production experts are workers of manufacturing and assembly – in so far as it concerns improvements in the manufacturability or the design-to-assembly quality of products. In the very particular case of machine tool manufacturing, the products manufactured (and assembled) are at the same time the (technical) means of production. In particular this typical bivalent situation enables the workers to use their experience, knowledge, and imagination for product improvements (Ruth, 1995, 2003). Even if these improvements are somehow incremental steps in the continuous innovation process, their impact can be enormous, for example, regarding the increase of a company's competitiveness (Imai, 1991).

The outlined relationships between work process experience (knowledge) and innovation obviously depend on various factors, which are supporting the innovativeness of companies' product and process improvements. Considering not the wafer development in the semiconductor industries as an example of a strongly science-driven high-tech innovation field, but rather the product innovation in the more traditional mechanical and electrical machinery as the source and subject of the following reasoning, we can assume the qualification structure and qualification level on the shop floor as a crucial component of innovation capabilities. The composition of qualification and skill is covered by the concept of competences, which can be defined as:

... an ability that extends beyond the possession of knowledge and skills. It includes: i) cognitive competence involving the use of theory and concepts, as well as informal tacit knowledge gained experientially; ii) functional competence (skills or know-how), those things that a person should be able to do when they work in a given area; iii) personal competence involving knowing how to conduct oneself in a specific situation; and iv) ethical competence involving the possession of certain personal and professional value. (Leney, 2004: 94)

The important point, besides the aspects of recognizing different forms of learning (theoretical and practical learning) originating in different learning spheres (shop floor, research and development (R&D) laboratories, and also outside organizations in everyday life), is the factor of providing learning opportunities at work while working. We can assume that the more learning opportunities workplaces (especially on shop floor) do provide, the better are the opportunities to actively involve the

people in innovation processes. However, evidently this involvement is highly dependent on the individual's competence level, which we can assume as being made up of the general education level, the knowledge level of the subject field, and of experiences during work at the workplace. A systematic and international comparative view on the relationship between qualification and innovation, thus requires one to consider the interaction between the systems of VET, and the innovation systems – the latter not on highly aggregated levels (such as national or regional innovation systems) but rather on the company or network level.

Institutionalizing Innovation Capacities in VET Systems by Implementing a Shaping Orientation

If the assumption of an existing relationship between skills, experience, and knowledge on the shop-floor personnel side, and the level of companies' innovativeness is acceptable, then we can assume that the underlying vocational learning and training system can also exert an important impact on innovation.

In an international comparative perspective, the differences between national VET systems need to be investigated if the possible relationships with innovation and innovativeness is to be stressed. It is particularly necessary to consider possible differences in the structure, the processes and the outcomes of the VET systems. **Table 1** gives a general overview of the structure of VET systems (focusing on the initial VET). A structural analysis shows one cluster of countries with a predominantly apprenticeship system, which is labeled dual system, because it combines theoretical and practical learning at different learning sites (school and workplaces). Among others, not only Germany, Austria, Denmark, and Switzerland can be mentioned as examples but also internationally some other Asian countries follow this dual model, such as Malaysia, Thailand, and Taiwan.

The second logical cluster of VET schemes comprises school-based systems, which – according to the international classification system (ISCED) – can be differentiated into one subcluster that follows a vocational path leading to a qualification that is recognized by the labor market (e.g., Finland, Sweden, France, and China), and another subsector that follows a prevocational track directing toward a general educational level with no recognition at the labor market (e.g., Ireland and Greece). The fourth cluster is characterized by a mix of school-based and apprenticeship/dual systems (e.g., the UK, Netherlands, and Italy).

This classification, on the first hand, gives us a description of the routes or pathways followed to achieve a certain goal – employability. It also allows for some

Table 1 Structures of IVET in Europe

<i>Predominantly dual system/ apprenticeship routes</i> Examples: Austria, Germany, Swiss, and Denmark	<i>Mix of school-based and apprenticeship routes</i> Examples: UK, the Netherlands, and Italy
<i>School-based (mainly vocational)</i> Examples: Finland, France, Sweden, and China	<i>School-based (mainly prevocational)</i> Examples: Ireland and Greece

From Leney, T., et al. (2004). *Achieving the Lisbon Goal: The Contribution of VET*, p. 32. London: QCA; and own sources.

considerations regarding the outcomes of the different pathways: as a rough estimate we can assume that those individuals running through a dual system/apprenticeship program will most likely hold a competence that comprises theoretical and practical experiential knowledge and can support innovation processes. The former is acquired during schooling, the latter earned at the workplace during real production processes. Compared with this type of outcome the school-based pathways lead to somewhat deficient competence structures. The prevocational school-based track supplies individuals with almost no practical job-relevant experiences to the workplaces. Evidently, these workers at their early entrance stages into the workplaces cannot be expected to hold significant practical, experiential knowledge that might contribute to innovation processes. Similarly, the school-based vocational path does not provide the learners with sufficient practical knowledge. This classification is challenged by the perpetual trend of thinking on the further improvement of their VET systems, which is observable in almost all countries. Even countries with little history in dual VET think of following a work-oriented change in the didactics of the VET curricula which try to identify significant vocational work situations and the associated work process knowledge as the key factor to redesign their vocational curricula as well as their teaching and learning processes. What is interesting in this respect is the change of perspective from merely academic discipline-based teaching methods toward a reformulation of action-oriented vocational teaching methods for VET practice that are building on development theories (Rauner and Spöttl, 2002; Bremer and Haasler, 2004; Dittrich and Deitmer, 2003; Deitmer and Heinemann, 2007; Rauner, 2007). Development theories reflect the idea of a VET student being confronted with developmental tasks (Havighurst, 1972). These are paradigmatic work and learning tasks derived from the occupational field the student intends to work in; the tasks' complexity is increasing to allow competence development that enables the learner's aptitude to shape work and technology.

Relevance of Practical and Theoretical Learning for Developing Competent Workers

The highest level of useful practical experience is acquired through dual systems with a systematic theory–practice relationship, but since apprenticeship programs have a limited duration (usually around 3 years), the practical experiences are limited anyway. But it is a much better starting point for augmenting the experience basis – as compared with school-based systems. Even if the latter include some temporary work placements, these practical experiences often do not offer sufficient learning opportunities, nor do they necessarily have a close relationship to the subjects taught in school. However, we have to consider that even the graduates from school-based systems have the chance to gain experiences during their work. This evidently depends upon the opportunities for learning offered in work, once they made the transition from school to work. Thus, we can say that the type of VET can offer an initial advantage (or disadvantage) regarding the acquisition and composition of theoretical and experiential knowledge. Depending on learning opportunities offered in and during work, this deficiency can be compensated after the school-to-work transition.

If the relationship between the VET systems and their outcomes, and the innovation capabilities of individuals are to be strengthened, we have to go beyond simply offering learning opportunities in work and rather demand the shaping orientation of working and learning (Rauner *et al.*, 1988). The shaping orientation can be understood as a goal that targets at enabling the workers to proactively shape their immediate working environment, form their processes, and be able to understand the work processes and the products they are dealing with, in order to develop improvements and put them into innovative action. This can be learned by developmental tasks. These developmental tasks have two learning functions. First, they are used as an evaluation instrument to demonstrate the formation of occupational competence and identity at the (as-yet-unidentified) critical thresholds of occupational competence development (Bremer and Haasler, 2004). Second, developmental tasks are at the same time a didactic tool for establishing and designing vocational curricula as well as learning and working tasks for structurally oriented VET (cf. Howe *et al.*, 2001).

Prospects: Can TVET Systems Be Transformed into a Pillar of Innovation?

Even though the arguments developed and the examples given in the preceding section focused on the

manufacturing sector and its particular constitution of TVET, it can nevertheless be concluded that the findings also apply to the service sector – at least those fields of the service sector that have similar structures of VET as the manufacturing sector. Besides the service fields that are closely linked to production (such as repair and maintenance of machinery and equipment) we might consider the field of healthcare as such a field, where skilled persons (e.g., at the level of nurses) can become a valuable source for improvements of machinery used for diagnosis and treatment. Furthermore, their experience can be valuable for innovating processes within hospitals, etc. Although more systematic research in the service sector is needed, we assume that basically the role of TVET, the contribution of practical experience held by the directly involved workers in service sectors to product and particularly process innovation is similar to what was developed for the manufacturing sector.

In conclusion, in trying to estimate the future role of experiential knowledge, there are two dimensions to be considered: first, the potential role of work-based practical knowledge is a huge and widely ignored store of knowledge; therefore, second, it is a matter of utilizing this source by creating favorable circumstances. As favorable conditions we might count (1) production processes that are simultaneously learning environments, to which the provision of learning opportunities at work is inherent; furthermore, (2) a corporate culture that facilitates cooperation, communication (and thus knowledge sharing) among all relevant innovation actors – particularly including those players who hold practical experience, that is, shop-floor people; and another beneficial condition is (3) a pronounced shaping orientation being expressed at all levels of vocational training and learning (in initial VET, continuous VET, and all informal and nonformal learning).

As far as VET systems are concerned, the most favorable to contribute to a high level of innovativeness are those that target beyond such simple goals as employability in the meaning of creating functional skills and qualification. Even if these functional qualifications are acquired at a high level of proficiency, their graduates lack those core capacities that make up for shaping orientation: expertise in the subject field, virtuosity of performance, (social) responsibility, transcendence, and reflexivity. Evidently these goals are idealistic to some degree, but they are highly useful as role models, and the better existing VET systems are approaching these goals, the more likely they can achieve shaping-oriented working and learning, which eventually would sustain innovation and innovativeness.

There is only *la petitesse* remaining: how to implement the shaping orientation in VET? Beyond the above-given orientations, at this point, only one general remark can be

made: theoretical and practical learning need to mutually sustain each other. In every phase of vocational training and learning, there must be a balancing of theory and practice. For the school-based VET systems, this means providing access to work practice and relate practice systematically with theory learning. For dual systems, the abundance of practice has to become seamlessly interlocked with theoretical knowledge.

If this balance of theoretical and experiential/practical knowledge can be achieved, a leapfrog has been done; however, it requires other important organizational and sociocultural factors to be met, one aforementioned is the culture of knowledge sharing, which is the transmission belt under which innovation competence is thriving.

Bibliography

- Braun, E. and MacDonald, S. (1978). *Revolution in Miniature*. Cambridge: Cambridge University Press.
- Bremer, R. and Haassler, B. (2004). Analyse der Entwicklung fachlicher Kompetenz und beruflicher Identität in der beruflichen Erstausbildung. (Analysing the development of technical competence and vocational identity in initial vocational education and training) *Bildung im Medium beruflicher Bildung. Zeitschrift für Pädagogik (ZfPäd)* **50**, 162–181.
- Deitmer, L. and Heinemann, L. (2009). Evaluation approaches for workplace learning partnerships in VET: How to investigate the learning dimension? In Marja-Leena Stenström, M. -L. and Tynjälä, P. (eds.) *Towards Integration of Work and Learning. Strategies for Connectivity and Transformation*. Doordrecht: Springer.
- Dittrich, J. and Deitmer, L. (2003). The challenges in teaching about intelligent building technology. *Journal of Technology Studies* **29**(2), 6.
- Duschek, S. (2004). Inter-firm resources and sustained competitive advantage: Management revue. *The International Review of Management Studies* **14**, 53–73.
- Harryson, S. (1998). *Japanese Technology and Innovation Management: From Know-How to Know-Who*. Cheltenham, UK: Edward Elgar.
- Havighurst, R. J. (1972). *Developmental Tasks and Education*. New York: David McKay.
- Hong, P., et al. (2004). Knowledge sharing in integrated product development. *European Journal of Innovation Management* **7**, 102–112.
- Howe, F., Heermeyer, R., Heuermann, H., Höpfner, H., and Rauner, F. (2001). *Lern- und Arbeitsaufgaben für eine gestaltungsorientierte Berufsbildung*. Konstanz, Christiani.
- Imai, M. (1991). *Kaizen – The Key to Japan's Competitive Success*. Singapore: McGraw-Hill.
- Leney, T., et al. (2004). *Achieving the Lisbon Goal: The Contribution of VET*. London: QCA.
- Lundvall, B. A. (ed.) (1992). *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. London: Pinter.
- Manske, F., Moon, Y., Ruth, K., and Deitmer, L. (2002). Ein prozess- und akteurorientiertes Evaluationsverfahren als Reflexionsmedium und Selbststeuerungsinstrument für Innovationsprozesse. ('The process and actors oriented evaluation approach as a medium for reflection and decision making'). *Zeitschrift für Evaluation* **2**, 245 ff.
- Mokyr, J. (2002). *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton, NJ: Princeton University Press.
- Morgan, K. (1997). The learning region: Institutions, innovation and regional renewal. *Regional Studies* **31**, 491–503.
- Nonaka, I. and Nishiguchi, T. (2001). *Knowledge Emergence. Social, Technical, and Evolutionary Dimensions of Knowledge Creation*. New York: Oxford University Press.
- Nonaka, I. and Takeuchi, H. (1995). *The Knowledge-Creating Company. How Japanese Companies Create the Dynamics of Innovation*. New York: Oxford University Press.
- Nonaka, I. and Teece, D. (2001). *Managing Industrial Knowledge. Creation, Transfer and Utilization*. London: Sage.
- Rauner, F. (2007). Competence development in vocational curricula and work situations. *European Journal of Vocational Training* **40**(1), 197–2019.
- Rauner, F., et al. (1988). The social shaping of technology and work: Human centred computer integrated manufacturing systems. *A I and Society* **2**, 47–61.
- Rauner, F. and Spöttl, G. (2001). *Der Kfz-Mechatroniker - Vom Neuling zum Experten. ('The car mechatronic - from novice to expert')*. Bielefeld: W. Bertelsmann.
- Rolt, L. T. C. (1986). *Tools for the Job. A History of Machine Tools to 1950*. London: Her Majesty's Stationery Office.
- Rosenberg, N. (1982). *Inside the Black Box. Technology and Economics*. Cambridge, MA: Cambridge University Press.
- Ruth, K. (1995). *Industriekultur als Determinante des CNC-Werkzeugmaschinenbaus. (Industrial Culture as a Determinant for the CNC tool making industry)*. Berlin: edition sigma.
- Ruth, K. (2001). Innovation competence: Intangibles in cooperative innovation processes. In Banerjee, P. and Richter, F. -J. (eds.) *Intangibles in Competition and Cooperation*, pp 177–197. Euro-Asian Perspectives Hampshire: Palgrave.
- Ruth, K. (2003). Manufacturing culture: Enginology or socioneering? In Ito, Y., Moritz, E. F., and Ruth, K. (eds.) *Synergy of Culture and Production Vol. 2 – Localized Engineering for Globalized Manufacturing?* pp 13–39. Sottrum: artefact Verlag.
- Ruth, K. (2006). Innovation policy for SME in Japan: The case of technology transfer centres. In Storz, C. (ed.) *Small Firms and Innovation Policy in Japan*, pp 56–81. London: Routledge.
- Senker, J. (1993). The contribution of tacit knowledge to innovation. *A I and Society* **7**, 208–224.
- von Hippel, E. (1988). *The Sources of Innovation*. New York: Oxford University Press.

Further Reading

- Boreham, N., Fischer, M., and Samurçay, R. (2002). *Work Process Knowledge*. London: Routledge.
- Corbett, J. M., Rasmussen, L. B., and Rauner, F. (1991). *Crossing the Border – The Social and Engineering Design of Computer Integrated Manufacturing Systems*. London: Springer.
- Deitmer, L. and Attwell, G. (1999). Partnerships and networks: A approach to learning in regions. In Nyhan, B., Attwell, G., and Deitmer, L. (eds.) *Education and Regional Innovation in the European Union and the United States*. Thessaloniki: CEDEFOP.
- Ruth, K. (2000). *Cooperative product innovation: How small and medium enterprises can improve their innovation competences*, pp 15. Tokyo: Bremen.

The Status of Vocational Education and Training

J Lasonen, University of Jyväskylä, Jyväskylä, Finland

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Candidate countries – They refer to Croatia (HR), Macedonia (FYROM), and Turkey (TR) in the European Union (EU).

EEA – This is the European Economic Area including Iceland (IS), Liechtenstein (LI), and Norway (NO).

EU – It refers to the European Union. EU-27 countries are the 27 member states of the European Union. The EU countries (abbreviations and the year of joining to EU) are as follows: Belgium (BE) (1952), Bulgaria (BG) (2007), Czech Republic (CZ) (2004), Denmark (DK) (1973), Germany (DE) (1952), Estonia (EE) (2004), Greece (EL) (1981), Spain (ES) (1986), France (FR) (1952), Ireland (IE) (1973), Italy (IT) (1952), Cyprus (CY) (2004), Latvia (LV) (2004), Lithuania (LT) (2004), Luxembourg (LU) (1952), Hungary (HU) (2004), Malta (MT) (2004), Netherlands (NL) (1952), Austria (AT) (1995), Poland (PL) (2004), Portugal (PT) (1986), Slovenia (SI) (2004), Slovakia (SK) (2004), Finland (FI) (1995), Romania (RO) (2007), Sweden (SE) (1995), and United Kingdom (UK) (1973).

Formal learning – It includes learning taking place in an organized and structured environment (formal education, staff training at workplace) that is designed for learning. It may lead to official recognition (certificate). From the learner's point of view, formal learning is intentional.

Informal learning – It takes place in everyday contexts at work, at home, and in one's free time. It is often referred to as experiential learning, and to a certain degree it can also be considered as unintentional learning. It is not structured in terms of goals, time, or guidance, and is normally not attested by any certificates. It may be purposeful but is typically unintentional and takes place more or less accidentally.

Nonformal learning – It covers planned action that is not actually designed for the purpose of learning, but still includes important elements of learning. From the learner's point of view, however, learning is intentional.

Occupational status – It corresponds to prestige, the socioeconomic status of an occupation, and to class measures in the past.

Social status – It is understood as ranking or hierarchy of prestige as perceived by the respective

groups and as seen in economic structure and division of labor.

Status of VET – It is related to the social and occupational status that is yielded by vocational qualifications such as recognition, prestige, and autonomy attached to one's position in society. In terms of a type of education and training, measured as progression chances and income returns later in young people's lives, the meaning of status (or standing or prestige) of VET is associated with parity of esteem between VET and other education tracks while comparing their standing of equality.

Validation – It refers to identification, evaluation, and recognition of the broad range of skills and competencies people have acquired in various contexts throughout their lives.

Introduction

For some 20 years, improving the attractiveness and status of vocational education and training (VET) has been one of the education and training policy priorities and objectives set by the European Commission, recently by the Lisbon–Copenhagen processes, and related follow-ups such as the Maastricht and Helsinki Communiqués. Attractiveness is strongly associated to the observed status of VET in terms of employability, salary, and chances for education and occupational mobility, as well as to young people's and their families' attitudes and preferences.

Status of VET is associated to social and occupational status that is yielded by vocational qualifications as recognition, prestige, and autonomy attached to one's position in society. In terms of a type of education and training measured as progression chances and income returns later in young people's life, the meaning of standing and prestige are close to status referring to the social status. Social status is understood as ranking or hierarchy of prestige as perceived by the respective groups and as seen in economic structure and division of labor. Occupational status corresponds to prestige, the socioeconomic status of an occupation, and to class measures in the past.

The status of and esteem for VET has been discussed in comparison with general/academic education, which has traditionally been assumed to be a more attractive pathway than VET to students and their parents. However,

the status of VET varies between countries and social groups measured as students' choices of educational pathways. Measures of esteem for VET refer to (1) improving access and entry to employment, career development, and progression; (2) the diversity of quality learning environments (meeting the needs and aspirations of different learners); (c) self-development choices as a citizen and individual; and (d) willingness to invest in VET. Attractiveness is observed as preferences, attitudes, and related behavior of individuals, groups, and their families.

Improving the attractiveness of VET is an educational policy goal in the European Union (EU). Attractiveness is a policy term rather than a theoretical concept. Some variables, such as equal access to progression through different education and training pathways that determine esteem for VET, are also assumed to improve the attractiveness of VET. The disparity of esteem between vocational and academic education may have its roots in the occupational status based on a traditional conception of social class structure that does not function in modern societies.

Parity of esteem is a theoretical concept that has fairly recently come out in academic literature. The first references are dated in the early 1990s. Charles Taylor did some pioneer work followed by Fraser, Habermas, Honneth, and Tully. They all share the premise that a just and stable political order guarantees all citizens a possibility to enjoy due recognition, which can be significantly promoted by education policy, as well. Thompson (2002) attaches the concepts of parity of esteem and the politics of recognition. The dual concept of parity of esteem and, especially, the theoretical and philosophical concept of recognition were introduced in the context of analyzing the diversity of communities and distribution of resources. The diversity of communities and nations and its lacking recognition have even led to conflicts and wars.

Parity of esteem describes the vertical relationship between a state and a citizen, whereas due recognition refers to horizontal relationships between citizens. Parity of esteem and recognition are based on the identities of different people and groups; for example, minorities and indigenous peoples. The notion of parity of esteem is parallel to the principles of equal opportunities and rights as well as nondiscrimination. However, it is not synonymous with equal rights or equal treatment. Parity of esteem refers to measures of equal recognition enjoyed by two or more communities.

Socioeconomic status is often measured in parity of income, education, and employment, while the lack or unequal distribution of these is often perceived as seeds of community conflicts. Du Toit (2004) points out, however, that the real causes are connected with regional autonomy and national independence, which at the individual level have to do with self-respect, dignity, recognition, and sense of worthiness. In analyzing parity of esteem, he highlights the concept of self-esteem, stressing

its importance for personal identity: "... a positive sense of self-esteem, along with a sense of self-efficiency (a sense of personal competence) and self-consistency (a sense of personal coherence), allow the individual to maintain an identity ... " (Du Toit, 2004: 197). Self-esteem develops on interaction with other people within communities attached to certain cultural backgrounds, values, and norms. A person's social identity consists of the social categories in which he or she builds his/her self-image and identifies with. Du Toit (2004) argues that groups with different identities compete with each other. Parity of esteem is the ultimate goal of such competition and its results. "Perceiving one's self as being an object of primary value within a meaningful universe, within a competitive context where culture and ethnicity affect both identities and competitive abilities, is the key to maintaining parity of esteem" (Du Toit, 2004: 200).

In the context of education and training systems, parity of esteem concerns

- the flexible and transparent education and training systems and job markets that provide access for all individuals in learning, career chances, and in life histories and
- relevant lifelong learning visions and options for all.

Parity of esteem has been analyzed between different pathways and/or different groups of people (e.g., gender, ethnicity, and social class) to examine the results of possible streaming and allocation of resources. The focus can thus be on the future and looking for learning opportunities for all with relevance to individual's quality of life.

The status of VET is looked at in the European education and training context where the concern has been one of the policy targets for several decades. A history of the status of VET can be traced to research on social stratification. Instead, the idea of parity of esteem has emerged in the early 1990s to the political theory, philosophy, and education research though some roots derive from Hegel's thoughts. This article, first, casts an overview to the main terms that are associated with the status of VET after which an origin and background of social status is sought. Third, the recent VET policy context in Europe is surveyed. Finally, parity of esteem between academic and general education is elaborated as reform strategies.

Parity of Esteem between Academic and Vocational Education and Training

Parity of esteem refers to a founding principle of recognition and a just and stable political order that allow all citizens to enjoy dignity by having equal access to resources available in their communities with cultural, social, and ethnic diversity. Parity of esteem between vocational

and general education has been an education and training policy goal followed by some reforms in the member states of the EU.

Improving the status of VET has been a European education policy goal since the 1990s. The first cycle of the Leonardo da Vinci program attracted several research and development projects that evaluated European education and training systems and upper secondary education reforms about the extent to which education and training policies promote parity of esteem between academic and vocational education. The following section elaborates the updates of the original strategies.

Reform Strategies

Parity of esteem between vocational and general/academic education is related to the societal rewards resulting from education and training whereby the concept is linked to the status of VET. Such rewards may be social status, the level of salary, prestige, and chances for further education and career development. In addition to societal and economic rewards of gaining occupational qualifications, having access to personal and intrinsic values such as self-fulfillment and meaningfulness provides esteem for vocations among individuals.

The synthesis of the results of four European collaborative projects with some 20 European partner institutions, which focused on the upper secondary education reform strategies and schemes, showed potential to reduce the disparity of VET. The conclusion drawn from the strategies of improving the attractiveness of VET and of promoting parity of esteem between vocational and general education is based on research results provided by four major partnership projects (Lasonen and Manning, 2000).

Upper secondary education systems can be characterized in several ways on the basis of how they have managed to offer opportunities to move either to higher education or to work for young people with diversified talents and needs. A study of European upper secondary education system identified four hypothetical strategies (unification, linkages, enrichment, and enhancement) to indicate the esteem for vocational education.

Unification

Unification is about merging vocational and general education into a single system of post-16 education. The educational systems in Sweden and Scotland were based on this strategy, where the goal has been to offer all students a core program of common general subjects and abolish the distinction between vocational and general learning. Their unification strategies were not identical. Sweden emphasized uniformity of treatment and outcomes for all students and joint administration for vocational and academic courses; in Scotland, the emphasis

was, instead, on providing choices among a flexible range of opportunities through modularized study programs. The distinction between vocational and academic students and vocational and academic teachers, however, remains even after the administrative unification and modularization of study programs. The Higher Still Reform of post-16 education in Scotland, consisting of seven levels of studies, replaces all upper secondary education and partly further education and follows the principles of flexibility in progress and inclusiveness in access. Looking back to a decade, this strategy has not worked out in its ideal way. There have been signs of academic drift, referring to favor academic tracks rather than vocational tracks, especially in Scotland. In Sweden, under the same institutional administration some 15 different vocational and academic education pathways, that have quite distinct cultures, exist. Raffe *et al.* (2001) found that there is no clear association between unification and parity of esteem.

Linkages

Linkage seeks to make vocational and general education formally more equal by connecting them through measures such as a common certification framework, arrangements for credit recognition and transfer, and common curricular elements. The English, French, and Spanish educational systems, seeking to make their vocational education more attractive to parents and students, have adopted the linkage approach. In France, the emphasis of the *Baccalaureat professionnelle* was created to improve employment prospects. In England, again, the main focus has been on enabling students on the new vocational programs through General National Vocational Qualifications (GNVQ) to have access to university on similar terms that have been available to A-levels students. Traditionally, in the English and French systems, academic education has tended to prepare upper- and middle-class children and young people for high-status jobs. Academic education has been popular to aspire after as it carried a prospect of social status and prestige (**Case study 1**).

Especially as the enhancement and linkages strategies are not exclusive but their features can be found across the countries, in the British context they have been combined as a linked strategy. The three countries of Great Britain illustrate tracked, linked, and unified strategies in academic and vocational education (Howieson and Raffe, 1999). England has developed distinct vocational diploma. The Welsh Baccalaureate represents an overarching qualification linking academic and vocational qualifications. Scotland has the Scottish National Qualifications standing for a partially unified system. However, in many other European countries' context, the enrichment strategy, including the features of the linkage strategy, is more relevant from the aspects of joint programs and organizing students' choices.

Case study 1

Is there a parity of esteem between vocational and academic qualifications?

However, Robinson (1997) indicated that there is no parity of esteem between academic and vocational qualifications in the labor market. The conclusion was drawn from an empirical study based the British labor force survey showing that on average male and females working full-time with academic qualifications at one level in the British qualifications framework earn about the same as men and women with vocational qualifications set notionally one level higher. For example, those with A levels have earnings similar to those with higher or level 4 vocational qualifications. However, the length of work experience and learning is much higher for persons who have attained vocational qualifications. Academic qualifications are more successful in providing access to more highly paid occupations. Further, academic qualifications are associated with higher earnings leading to the highly paid managerial, professional and technical occupations.

Enrichment

Mutual enrichment bring various types of schools, such as vocational and academic institutions, closer by encouraging cooperation between vocational education establishments and general upper secondary schools and between vocational schools and enterprises with the aim of providing students with a wider range of options and offering them stimulating learning methods and environments. The educational systems of Finland and Norway were examples of the enrichment strategy. It has enriched both vocational and general education through measures that allow them to draw on each other's best features in students' choices. The administrators and career counselors from both the types of education providers have to cooperate to design a set of integrated courses based on students' choices. In Finland, new types of qualifications and certificates have been established. The strategy has lowered the traditional barriers between these two types of education, vocational and academic upper secondary education, while maintaining a distinct identity for each. On the one hand, general subjects have been added to the curriculum in order to ensure preparedness for continuing education and training. On the other hand, curriculum reform has led to a restructuring of study programs by improved opportunities during on-the-job training, with the result that work-based learning has renewed teaching approaches. Cooperation between vocational education and enterprises has enriched both of them in terms of work-based learning and lifelong learning. In Norway, general and vocational education teachers are educated in the same teachers training colleges. Both countries of the enrichment strategy have a strong full-time vocational education system and have proved to improve esteem for VET and parity and to avoid academic drift (unrealistic desire for academic education). On the contrary, by diversification and differentiation of higher education to vocational dimensions, a phenomenon of vocational drift can be observed.

In addition to the macro- and system-level approaches, curriculum-level reform schemes may show new kind of qualifications produced. The schemes of dual qualifications in Europe revealed a holistic model of teaching qualifications that provide choices and chances for students. The schemes of dual qualification prepare for both

employment and higher education (Manning, 1997). The starting point of dual qualifications is a varying degree of integration between general/academic and vocational education and work- and school-based learning in students' learning. The most successful examples of such qualifications represent school-based full-time education (e.g., Finland). Most dual qualification systems have emerged within vocational education as a part of the qualification process (e.g., Austria and Germany). A strong tradition of vocational programs offering a dual qualification has existed in Czech Republic.

The implementation of dual qualifications is assessed in terms of enrolment, the integration of academic and vocational education, and the students' success in their studies and in finding employment and/or gaining admission to further or higher education. Depending on the country, the proportion of the age cohort taking dual qualifications ranges from 1% to 45%. Those completing a dual qualification either enter working life or continue their studies, in most cases, in higher vocational education. However, in most countries the stress is on employment. The employment rate of holders of dual qualifications is high (**Case study 2**).

Enhancement

Vocational enhancement stresses the distinctive nature of vocational education on the basis of its characteristic content and links between employers and the providers of education of this kind. The German, Austrian, Czech, Spanish, Dutch, Danish, and Swiss educational systems have been founded mostly on this strategy. They have enhanced vocational education and made it more attractive to potential students through measures that have maintained and strengthened its distinctive ethos and its separate character alongside general education. Esteem for vocational education is linked with the high standard of the content and pedagogy offered in VET that provides a path to employment and to higher education. In addition to reforming the vocational upper secondary education, the countries have invested in higher vocational education. Parity of esteem between general subject teachers and vocational teachers has been enhanced with equalizing measures. As a strong work-based system,

Case study 2

An evaluation of co-operation between general and vocational upper secondary education and training

In Finland, although general and vocational upper secondary education (ISCED 3) provide parallel opportunities to progress to higher education, academic upper secondary schools and vocational institutions function separately. The reform in the 1990s focused on enrichment strategy by establishing cooperation between these two sectors in order to increase student choice of courses/modules, for students' benefit. This offered secondary students new options and flexible study paths. A recent large-scale evaluation of the Finnish upper secondary education revealed that 63% of VET providers cooperated with the providers of general upper secondary education during the academic year of 2005–6 (Mäensivu *et al.*, 2007). Correspondingly, 83% of the providers of general upper secondary education cooperated with those of vocational education and training. The providers in the study formed 102 cooperation networks that include, in minimum, one provider from each type of secondary education.

Based on cooperation, students have three options to utilize the course work across general education and vocational programs for their certification and qualification: completing one or more separate courses or a whole module either in vocational institutions or upper secondary schools, or completing a double or triple qualification. About 5% of all upper-secondary graduates in the whole country and 6.9% of the cooperation networks' graduates of the evaluation study have completed double or triple qualification studies.

Vocational students are more eager than general education students to take courses across the education institutions. On average, 2.2% of general upper secondary students have completed at least one course offered by vocational schools in the cooperation networks. Accordingly, 4.1% of the vocational students have completed at least one course offered by upper secondary schools in the network in 2004–5. Similarly, vocational students are more goal oriented in their course choices from the academic side than the general education students. On average, 5% of the students who completed a vocational degree were pursuing a double or triple degree, and 10% of the students completing their vocational degree also took and passed the matriculation examinations.

The cooperation between general and vocational education providers may be restricted by various factors, such as distance between the schools, incompatible schedules, differences in working environments and administration, as well as mutual relations and attitudes. The systems of funding and administration by results do not support cooperation, although it is binding according to the upper secondary education legislation.

countries with vocational enhancement have high rate of participation in education and training and inclusiveness.

A further study on the strategies focused on the vocational-enhancement strategy as European countries are faced with the challenge of improving the quality of VET. Additionally, most of European upper secondary students enrol in vocational programs. The sub-strategies of the enhancement strategy for improving upper secondary education and the quality of VET were identified: (1) promoting access to higher education; (2) increasing cooperation with employers and enhancing quality of curriculum; and (3) raising the status and qualifications of VET teachers.

Access to Higher Education

The expected length of time spent in education has increased since 1995 along with greater participation in preschool, secondary, and higher education. In the majority of European countries, the expected duration of studies for the current five-year-olds is more than 16 years. The highest expectancy figures are found in Belgium, Finland, Iceland, Sweden, and the UK – ranging from 18.5 to 20 years. The mean of the years of educational attainment in formal education was 11.9 years in the OECD countries in 2004 (OECD, 2006).

On average, 53% of the youth of industrialized countries participate in some sort of higher education in universities or polytechnics. In 2003, some 16% of higher education students were enrolled in educational institutions other than universities (OECD, 2005). The increasing demand

for higher education also calls for a broader and better educational choice, diverse learning environments of good quality, and opportunities for further studies.

Direct access from initial vocational education to higher education has improved in many European countries, helping to promote parity of esteem between vocational and academic education. On the one hand, there are countries which are trying to improve access for those in vocational courses through a unified system of higher education (such as in the UK and Sweden). On the other hand, some countries are seeking to diversify their system of higher education (or to expand an already diversified system) by developing vocational higher education for graduates of initial vocational education (e.g., Finland, Germany, and Austria). In Finland, the upgraded polytechnics are also designed to attract those graduating from general upper secondary schools. (Although the Finnish polytechnic system does not have a particular international model, it resembles the German Fachhochschule and the Dutch HBO [*middlebar beroepsonderwijs*], now BOL4 [*beropleidende leerweg*] system.) This can be seen as another way of promoting the parity of esteem of vocational and general upper secondary education. In Europe, four different systemic patterns of organizing higher vocational education can be identified:

- gradual integration has led to a unification pattern and to a single higher education system (e.g., UK);
- the co-existence of university and non-university sectors with different status refers to a diversifying pattern (e.g., Germany);

- upgrading existing tertiary VET for separate but equal non-university sector indicates two parallel higher education systems, vocational and academic higher education (e.g., Finland); and
- establishing a new vocational non-university sector again represents a diversifying pattern (e.g., Austria).

About 17% of all higher education students enrolled in vocational higher education programs in EU-25 countries in 2004 (Eurostat, 2006). The proportion varies among European countries. For instance, in Finland more students ($n = 20\,821$) graduated from vocational higher education colleges than from universities ($n = 18\,293$) in 2004.

Co-Operation between Workplaces and Education Institutions

When compared to school-based vocational education with apprenticeship training system, cooperation with enterprises has a different nature. For instance, in Germany workplace learning is company-led whereas in Finland it is led by educational institutions. In Finland, the upper-secondary-level workplace learning scheme has the unanimous support not only of educational decision makers but also of the organizations of working life. The social partners have taken active part in planning and implementing workplace learning periods as well as in preparing subsequent legislation for integrating on-the-job learning periods to initial vocational education on regular bases. Behind the employers' willingness to contribute toward a smoother transition from education to working life lies their worries about the aging of the population and about recruiting newly qualified people. The government and the pivotal labor market organizations have issued a recommendation in favor of work-based learning, and the social partners and education providers signed joint agreements committing all the parties to the implementation of the scheme. The wording of the Vocational Education Act (1998) emphasizes the workplace experience component of learning. The trainees are called workplace learners (previously trainee), and workplace-based practice is called work-based learning (previously work practice). This undertaking also affects the way in which curricula are reformed to promote work-based learning. The guidance involved in, foci defined for, and assessment of work-based learning are based on educational objectives set down in a curriculum planned and designed mainly by the teachers.

In school-driven education systems, learning through work has been traditionally considered an informal activity, that is, an activity without links with formal study requirements and gaining a certified qualification. In the 1990s, a central objective of Finnish educational policy was to bridge the gap between educational and employment policies. The place of learning at work in upper

secondary education (ages 16–19) has been underpinned by schemes for developing competence-based qualifications and other examinations for adults as well as by a reform of apprenticeship training.

Raising the Status and Qualifications of VET Teachers

The position of VET teachers varies widely across different European countries. Some countries have detailed selection criteria including both university-based qualifications and specific periods of relevant employment. In others, colleges are free to recruit as they wish. However, in terms of qualifications, work experience and pedagogic skills as well as closer equivalence with general education teachers, there is greater emphasis on developing and broadening the skills of vocational teachers so as to also raise their status. The prestige of VET is reflected in the status of vocational teachers and their education, and vice versa. Prestigious VET is essentially based on broadly competent vocational teachers who respect their own work. The demands set for their expertise have evolved from vocational pedagogics into broad professional mastery of relevant sections of the education system and economic sector. The curricular development of vocational teacher education with respect to education strategic competence and international educational knowledge are some of the challenges of quality assurance. New demands set for education, such as transparency, cooperation, and internationalism are not field specific, although they will take on different forms and emphases in different VET programs. Teachers need to have a broad international perspective in order to achieve the present and future educational goals. To be successful, vocational counseling calls for an overview of the options of further studies abroad, as well.

Vocational teacher education curricula can be criticized for the way they build up systemic thinking in teachers: Do the student teachers obtain a holistic overview of the role, goals, and criteria for the successful functioning of the whole education system, or do these matters end up scattered at the margins of separate sections?

Beyond the Strategies

The results of this project on European reform strategies were exploited in the UK as Home International comparisons of post-16 education and training (Croxford, 2001). The education and training systems of the four UK home countries were compared in data on a single cohort. As a result of the policy learning from abroad was found higher levels of social segregation and wider social inequalities in attainment in England than in Scotland and Wales. The English education policy has emphasized diversity whereas Scotland and Wales have more uniform

comprehensive education systems. The British practice of exploitation of the reform strategies showed a possibility to view the four strategies as representing a continuum between strategies based on the distinctiveness of academic and vocational education and those based on their full integration, with linkages and mutual enrichment as intermediate strategies between the two poles. This continuum can be described for several different dimensions of change, such as curriculum, certification, and institutions, and a country's strategy can vary across these dimensions.

Each country's reform programs have included elements of different strategies, and the emphasis of a country's policy might have changed over time. No one strategy was found to be superior or more effective than the others; the effects of each strategy must be judged in relation to the educational system and the context in which it was introduced. The analyses of the strategies have pinpointed many practical lessons which the European countries have learnt from experience elsewhere through which they have reviewed their home systems from a different perspective as a result of policy learning.

The current challenges of vocational education systems in many countries emphasize recognition of prior learning, especially among immigrants, and flexible training modes in order to get employed and have a chance to contribute to a nation's economy as citizens. The most recent developments of education and training policies in Europe show that the most important objectives for which the strategies strove have been achieved. A new kind of thinking has taken over in the development of a European vocational education area.

Different types of educational systems provide different starting points for promoting social equality through education and training. When looking back to European history and the development of their educational systems, some countries have been more advanced than the others in terms of a rate of inclusiveness. Upper secondary education systems can be characterized in several ways on the basis of how they have managed to offer opportunities to move either to higher education or to work for young people with diversified talents and needs. In dual-systems (examples from Austria, Denmark, Germany, the Czech Republic, the Netherlands), VET has had high status and equal hold in educational systems representing professional technical and VET cultures. Also, strong relationship between education system and labor market exists, including a tracked system of education and a qualification structure which has direct relevance for occupational entry. Advantages of the systems have been that they allow many kinds of opportunities for learning and inclusiveness by meeting diverse learners' various needs.

In school-based systems, such as in the Nordic countries (Finland, Norway, and Sweden), vocational

education has been integrated within educational systems representing educationalist culture. The aim is to promote parity of esteem between academic training and VET. Upper secondary education offers alternative progression routes and prepares citizens to higher educated society, especially in Finland and Norway. Relationship between education system and labor market has been quite loose, allowing for predominant school-based, broad vocational, and subsequent firm-based training. Advantages of the systems have been that relatively low drop-put and exclusion rates exist.

In so-called mixed systems (for example England and France), educational systems have been highly selective and divisive in the past. The function of education has been to prepare students for academic studies and ensure education for the elites. The education system has been used as a mean of social class tracking. Vocational education has had a low status and exclusiveness has been relatively high. However, the countries have had many development projects to raise the quality and status of technical and VET.

VET Policy Context in Europe

The Bologna, Lisbon, and Copenhagen agreements and the associated processes that are underway have given a new drive for education and training policy formation and implementation in Europe. The development of high-quality VET has been a crucial and integral part of the Lisbon strategy (2000) in terms of promoting social inclusion, cohesion, mobility, employability, and competitiveness. The Copenhagen Declaration (November 2002) and the Council Resolution on Enhanced European Cooperation in Vocational Education and Training (December 2002) identified concrete outputs in the fields of quality assurance, transparency, and recognition in order to improve the overall performance and attractiveness of VET. The first joint review of the process took place at a ministerial meeting in Maastricht. Based on the Education Council Conclusions adopted on 15 November 2004, the Maastricht Communiqué set out priorities for the next phase of the process.

The Maastricht Communiqué (2004), linked with the Copenhagen Process and Education and Training 2010 Work Programme, introduced national priorities. The necessary reforms and investment should focus on:

- the image and attractiveness of the vocational route for employers and individuals in order to increase participation in VET;
- achieving high levels of quality and innovation in VET systems in order to benefit all learners and make European VET globally competitive;

- linking VET with the labor market requirements of the knowledge economy for a highly skilled workforce, and especially, due to the strong impact of demographic change, the upgrading and competence development of older workers; and
- the needs of low-skilled (about 80 million persons aged between 25 and 64 years in the EU) and disadvantaged groups for the purpose of achieving social cohesion and increasing labor market participation.

European countries are challenged to make further efforts to cope with demographic trends and rapid structural and technological changes. Aging of the European workforce in numbers means that by 2009, there will be more people aged between 55 and 64 than young people aged between 15 and 24. By 2030, the number of young people will decrease by 9 million and the number of those aged between 55 and 64 will increase by 14 million (European Commission, 2005b). There is also a trend toward polarization of jobs. Compared to several non-European countries, the EU-27 countries score relatively low in high skills and high in low skills. Some 80 million European citizens, that is, some 30% of the European working-age population are formally low skilled. Meanwhile, two-thirds of the jobs in the EU-27 countries are found in skilled and higher skilled occupations. Increasing demand is expected for higher-skilled people across all occupations due to technological changes and innovations.

VET is increasingly taking place at all educational levels and, therefore, the parity of esteem and links between VET and general education, particularly higher education, need to be fostered by innovative strategies and instruments at the national and European levels. This should include designing VET systems that attract more students to higher qualifications, which is assumed to contribute to innovation and competitiveness.

The Helsinki Communiqué set four priorities as of 2007: (1) policy focused on improving the attractiveness and quality of VET, again; (2) development of common instruments and tools to enhance a European area of VET and a European labor market; (3) strengthening learning from others; and (4) taking all stakeholders on board.

Particularly, priority 1 refers to the objective that focuses on improving the attractiveness and quality of VET. This calls for:

- strong links between VET and working life, both in formal initial VET and continuing workplace learning for employees;
- better counseling and information in preparation for working life and improved guidance throughout life;
- permeable VET systems offering access to flexible, individualized pathways and progression to further education and training, for example, higher education; and

- highlighting excellence in skills – for example, by greater use of skills competitions.

European Commission (2006) in its Education and Training Programme 2010 has identified five priority benchmarks that are monitored in European countries. The agreed priorities and benchmarks to be attained by 2010 are:

- All member states should halve the rate of early school leavers by 2010 in order to achieve an EU average of 9% or less. The progress report in 2006 showed that EU-25 average was as high as 14.9%. Instead, Poland (5.5%), Slovakia (5.8%), and Czech Republic (6.4%) have already reached the European benchmark in taking care of the youngsters who are under the threat of dropping out.
- Members should ensure that the average percentage of 25–29-year-olds with at least an upper secondary education reaches 80% or more. The EU-25 average was 77.3% in 2006. Slovakia (91.5%), Slovenia (90.6%), and Czech Republic (90.3%) had the highest upper-secondary competition rates.
- The EU average level of participation in lifelong learning should be at least 15% of the adult population aged between 25 and 65, and no country should have it lower than 10%. The EU countries' average is 10.8% challenging to reach the benchmark of 12.5%. Sweden (34.7%), UK (29.1%), and Denmark (27.6%) were the best performers in adult participation in lifelong learning according to the 2006 progress report.
- The percentage of low-achieving 15-year-olds in reading, mathematics, and science, as measured by PISA tests, should be halved in each member state between the years 2000 and 2010.
- Member states are expected at least to halve the level of gender imbalance among graduates in mathematics, science, and technology while securing a significant overall increase in the total number of graduates from 2000 to 2010.

A comprehensive European VET agenda has been launched since 2002 within the framework of the Copenhagen process. The common objectives to be gained were agreed upon by the participating countries (33 countries include EU-27, EEA, and candidate countries), the social partners, and the European Commission. The common objectives were, to prepare a European area of VET, to support competitiveness of European labor market, and to facilitate mobility of workers and students. The following instruments and tools have been established to strive for the objectives:

- European Qualification Framework (EQF);
- European credit transfer system for VET (ECVET);
- Europass;

- Common quality assurance framework (CQAF) and the European network of quality assurance in VET (ENQA-VET);
- common principles on validation of non-formal and informal learning; and
- lifelong guidance.

At the national levels, the member states of the European Union are expected to continue to modernize their VET systems and to improve the image, attractiveness, quality, relevance, and efficiency of VET.

Are European citizens satisfied with VET? The Eurobarometer survey was conducted in 25 EU countries at the end of 2004 to find out how well the current VET system meets the expectations of EU citizens and what reforms should be made. Instead of general/academic education, most EU citizens recommended VET or apprenticeship training to young people leaving school. In the VET Eurobarometer 39% of the respondents recommended VET as the first option, whereas 32% were in favor of academic studies. VET is most popular in France and Great Britain, where 57% of the respondents would recommend it as the first option. In addition, Finland is ranked above average with 48% in favor of VET and 29% recommending academic studies.

As many as 83% of the respondents were very or fairly happy with their education. The highest rates of satisfaction in this respect were found in Luxembourg (92%) and Malta (91%), followed by Finland (90%).

Education alongside work is particularly popular in Europe. A large majority of the respondents (73%) were willing to study in their free time. Also employers support continuing vocational education alongside work; 52% of the respondents stated that their employer supports and encourages continuing education.

More negative responses concerned study counseling and guidance. Only 25% of the respondents had got help or guidance in educational matters during the past 12 months. In Finland, the situation was much better, as 55% had received such help and guidance, while Sweden ranked second with 41%.

European Qualifications Framework

As part of the Lisbon Programme and subsequent Education and Training Programme, the Commission issued a proposal on 5 September 2006, for a Parliament and Council recommendation in order to establish an EQF (KOM (2006) 479 Final). The EQF has been constructed from the viewpoint of lifelong learning and it covers general education, VET, higher education, and adult education. The proposed recommendation aims at improving EU citizens' international mobility and promoting lifelong learning, for example, by making it easier to compare and determine the equivalence of qualifications against a

common frame of reference in the cooperation between member states. The member states should relate their national qualification levels to the EQF levels and create a national qualifications framework where applicable by national legislation and practice. According to the proposal, by the year 2011, all new qualification certificates granted should state their EQF level.

The EQF is based on eight levels that would describe the learning achievements of the person having the qualification. The learning achievements are defined in terms of skills, knowledge, and competence. Levels 5–8 correspond to the cycle descriptions of the framework for the European Higher Education Area (EHEA) in line with the Bologna Process. Placing a qualification on an EQF level would not affect the rights migrants have by the directive on professional qualification.

Vocational Qualifications and Status

For individuals, studying for vocational qualifications means a long-term, goal-oriented learning process so as to develop necessary competence and assets for the labor market. A qualification certificate tells outsiders something about the person's competence with regard to certain standard requirements. From the employment point of view, a qualification may largely determine placement in certain jobs as well as the income level, although ultimate success in working life does call for competence rather than diplomas and certificates. A recognized certificate is nevertheless an easy way to show and attest one's competence to a potential employer. Qualifications are also important for professional esteem. For society, qualification systems and related quest for competence contribute to more highly skilled workforce and greater productivity. International classification frameworks enable evaluation and comparison of educational attainment levels in relation to other countries.

The status of VET is linked to the qualities and properties of the vocational qualification system, such as follows:

- Flexibility and efficiency – referring to how effectively the vocational qualification system functions as such and as part of the larger education and qualification system. The continuum of initial, specialist, and further vocational qualifications promotes lifelong learning. Study programs of different fields should be designed to cater for a wide range of options and cross-disciplinary contents accounting for eligibility for further studies and possibilities for career changes later in life.
- Education that takes into account possible changes of occupations during one's work career, developing and sustaining working-life skills so that the person is able to flexibly respond to changing skills demands.

- Employment indicators that give education providers feedback about young people's placement in the labor market.
- Initial, specialist, and further vocational qualifications that provide eligibility for polytechnic and university studies, which needs to be accounted for in curricular planning (Case study 3).

Recognition of Informal Learning

In 2002, European ministers of education issued a recommendation in Copenhagen, according to which common policies need to be formulated for the validation and recognition of prior learning, so that lifelong learning can be accounted for in practice and give citizens due credit for it. Validation refers to identification, evaluation, and recognition of the broad range of skills and competencies people have acquired in various formal, informal, and non-formal learning contexts throughout their lives. According to Colardyn and Bjørnavold (2004) most EU countries already have such national policies defined at some level, but permanent arrangements to this effect have been endorsed only in Finland and the UK.

Anyway, the development of systems to recognize informal learning has commenced in Europe.

Recognition of informal learning calls for nationally defined standards that serve as reference scales. When validating informal learning, vocational qualification requirements and related evaluation criteria must be defined (Case study 4).

Different parties (states, regions, local communities, social partners, professional associations, voluntary organizations) are developing their own validation systems. This approach of various parallel methods is accepted in many countries (e.g., Denmark, the Netherlands, Norway, and Sweden). A current debate concerns the issue of whether or not validation should be bound to official systems, and what kind of a European approach should be developed based on different national solutions.

Summary

Three criteria – personal competence, educational mobility, and occupational mobility – underpin the assumption that status and esteem correlate to the quality of

Case study 3

A Finnish case: Changing occupations during one's work career

Increasing changes in the occupational structure makes people change their occupations. People also want to secure their professional future by anticipating the change and seeking new posts. According to the statistics on occupational transfers in 1995–2000, more than half a million people, ~27% of the target population, had changed their occupational group during the 5-year period (Statistics Finland, 2002). Here, occupational transfers refer to career changes from one occupational group to another, and concerning people being employed in 1995 and still in 2000. The greatest in-flow rates were seen in the groups of artistic and cultural managers and producers (55.8%), office workers (44.7%), business managers and experts (42.6%), and experts of mathematics and science (42.3%). Correspondingly, the lowest in-flow rates were found among doctors and nurses (about 9%), police, firemen and prison guards (11.9%), military personnel (12.7%), and teachers (13.4%). In general, occupations that typically call for a secondary-level qualification receive people from other groups at an average rate, although there is considerable variation between specific occupations. For example, among mechanics the in-flow from other occupations was 34.9%, machinists 33.3%, and storage workers 30.9%, whereas the corresponding figures were much lower in beauty services (15%), cleaning work, and basic health services (18.3%). Shop-floor level occupations also attract frequently people from other occupations. Hence, labor market dynamics does not concern solely higher-level occupations but affects all occupational groups. Occupations with highly regulated qualification requirements are less likely to attract large numbers of job switchers.

Case study 4

How do the Nordic Countries set the standards for validation?

Sweden seems to have a different view on the need for standards, however. There everybody has a statutory right to get an assessment, validering, of their skills and knowledge attained in different ways, but the Swedish assessment just describes the person's competencies without comparison to any qualifications. In Norway, school curricula serve as standards, even if there is also criticism against such practice. For example, tacit knowledge and related competence gained through experience may be overlooked when using school-based standards. Denmark, then again, has introduced a new adult education system, which offers adults shorter routes to qualifications and further vocational education through recognition of prior learning. To get admission to an adult education program, an applicant must have relevant professional background and work experience of at least two years. In Finland, for example, validation methods have been developed in a national context without European models. Candidates demonstrate their competence in authentic work situations. Such authenticity seems to be missing from the assessment methods used in other European countries, albeit assessment of working is sometimes mentioned. Hence, labor market dynamics, when no other evidence is available about a person's professional competence.

vocational education and to access to further studies, through which VET is seen to be as attractive and prestigious as the other educational pathways. Improving the status of VET has been one of the key education policy objectives in the EU. Rising educational attainment level is supposed to give the best preconditions for active citizenship and employment. By the same token, European countries have adopted educational reform strategies for improving the progression from vocational secondary education to higher education and for enhancing the status and quality of vocational education.

Traditional research on occupational status seems to be based on an assumption that classification of occupations follows a pattern of social stratification. However, modern citizens value their career and occupation choices against their aspirations, identity formation, and cultural communities. Family background is related to young people's education and career choices to varying degrees in different European countries. In addition to equity policy and its implementation, relevant career guidance, enhancing progression and quality of VET can improve the status of VET.

Progression opportunities including choices for double qualifications and access to higher education, prestigious occupations, and interesting learning environments are assumed to attract young people to VET. From the perspective of lifelong learning, it does not matter which type of education was chosen by young and older people, but it matters if education and training produce transparent and flexible competences for continuous education and training at any point in individual lifespan. This is the future focus of the European education policy implementation.

Nearly all the European countries have achieved the goal of universal upper secondary education (ISCED 3). The mass of students in formal education challenges the education and training systems in terms of diversification of secondary and tertiary education. The learning environments of VET are different from those of traditional academic/liberal education. Teaching and learning approaches of VET draw on strategies of constructivist, and situated, work-based, and expanding learning theories. Thus, VET can contribute to modernizing teaching and learning methods and curricula in education systems.

VET seems to hold quite a high status among young people in most of the European countries, especially among male students. In 2004, enrolment in VET programs varied, depending on the country, from 9.2% to 79.2% of all upper secondary education students. In 20 countries out of 35 countries, more than 50% of the upper secondary education students enrolled in VET in 2004. In all EU-27 countries, VET is more popular among male students compared to female students. Exceptions are the UK and Sweden. Ireland has only academically oriented programs at the ISCED 3 level. Instead, she has

pre-vocational programs and vocational programs at the ISCED 4 level (post-secondary non-tertiary programs).

Statistics show that the average salary level rises and the risk of unemployment declines along with the level of education and training, both in Europe and abroad. VET is appreciated if this route is also seen to enable progress in educational and professional careers. With respect to professional socialization and structural reproduction in society, family plays a central role in passing on values and attitudes. Some exceptions exist only in the Nordic countries (Finland and Sweden) where the result of social reproduction between social groups is not strong. Parents' educational and occupational background as well as their attitudes toward different options have significant influence when students are choosing between pathways after primary or lower secondary education.

Vocational and general education and training programs and courses offer different learning environments and experiences. Learning environments which include options meet the needs of a diverse student population and of their different learning styles better than one kind of program, whether general or vocational. Due to the cooperation between different educational institutions at the local level, some countries have made it possible for students to study both general and vocational courses, either for interest and variety or in order to obtain two diplomas. Most countries offer either general or vocational programs to their students, without more options.

Four alternative reform strategies were identified on parity of esteem between vocational and general/academic education as tools for analyzing the differences and similarities of the reform approaches in European upper secondary education systems. The strategies of vocational enhancement, mutual enrichment, linkages, and unification have been largely used for developing national and European education policies. Progression indicates the chances to continue from VET to higher education at any point in the lifespan. The organization of higher education programs varies in Europe. In certain countries, such as Austria, Finland, and Germany there are parallel vocational and academic higher education tracks. In these countries, progression chances are better for vocational students graduating from vocational upper secondary education than in the countries that have only one academic higher education degree.

See also: Dual System; Industry Involvement in the Vocational Education and Training System.

Bibliography

- Colardyn, D. and Bjørnavold, J. (2004). Validation of formal, non-formal and informal learning: Policy and practices in the EU Member States. *European Journal of Education* 39(1), 69–89.

- Croxford, L. (2001). School differences and school education: Comparison between England, Wales and Scotland. *Education Review* 15(1), 68–73.
- du Toit, P. (2004). "Parity of esteem": A conceptual framework for assessing peace processes, with a South African case study. *International Journal of Comparative Sociology* 45(3–4), 195–212.
- European Commission (2005a). Keydata 2005. Brussels: Eurydice.
- European Commission (2005b). Modernizing education and training: A vital contribution to prosperity and social cohesion in Europe. COM(2005) 549 final/2. Brussels, 30.11.2005. http://ec.europa.eu/education/policies/2010/doc/progressreport06_en.pdf.
- European Commission (2005c). *Vocational Training. Special Eurobarometer 216*. Brussels: Author.
- Eurostat (2006). *Database in Population and Social Conditions: Education and Training*. http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1996,45323734&_dad=portal&_schema=PORTAL&screen=welcomeref&close=/popul/edtr/educ&language=en&product=EU_MAIN_TREE&root=EU_MAIN_TREE&scrollto=342 (accessed May 2009).
- European Commission (2006a). *Quality Assurance in Teacher Education in Europe*. Brussels: Eurydice.
- European Commission (2006b). Progress towards the Lisbon objectives in education and training. *Report based on indicators and benchmarks*. SEC(2006) 639. Brussels 16.5.2006.
- Howieson, C. and Raffae, D. (1999). *The 'Unification' of Post-16 Education, CES Briefing 15*. Edinburgh: Centre of Education Sociology, University of Edinburgh.
- Lasonen, J. and Manning, S. (2000). Transnational comparisons of parity of esteem between vocational and general education. *International Journal of Vocational Education and Training* 8(1–2), 7–21.
- Manning, S. (ed.) (1997). *Qualifications with a Dual Orientation towards Employment and Higher Education: A Collaborative Investigation of Selected Issues in Seven European Countries. INTEQUAL Report II*. Berlin: Wissenschaftsforum Bildung und Gesellschaft e.V. (WIFO).
- OECD (2005). *Education at Glance*. Paris: OECD.
- OECD (2006). *Education at Glance*. Paris: OECD.
- Raffae, D. (1996). European strategies for parity of esteem. In Lasonen, J. (ed.) *Theory into Practice No 92: Reforming Upper Secondary Education in Europe. The Leonardo da Vinci Project Post-16 Strategies*, pp 273–278. Jyväskylä: Institute for Educational Research, University of Jyväskylä.
- Robinson, P. (1997). *Discussion Paper No. 354: The Myth of Parity of Esteem: Earnings and Qualifications*. London, UK: Centre for Economic Performance.
- Thompson, S. (2002). Parity of esteem and the politics of recognition. *Contemporary Political Theory* 1(2), 203–220.

Further Reading

- Azizi, N. and Lasonen, J. (2006). *Education, Training and the Economy. Preparing Young People for a Changing Labor Market*. Jyväskylä: Institute for Educational Research, University of Jyväskylä.
- Cedefop (European Centre for the Development of Vocational Training) (2009). *Modernising Vocational Education and Training*. Fourth report on vocational training research in Europe: background report. Volume 1. (Cedefop Reference Series, 69–I). Luxembourg: Office for Official Publications of the European Commission. http://www.cedefop.europa.eu/etv/Projects_Networks/ResearchLab/published.asp (accessed July 2009).
- European Commission (2007). *Indicators for Monitoring the Employment Guidelines. 2006 Compendium*. Brussels: European Commission, DG Employment, Social Affairs and Equal Opportunities.
- Iannelli, C. (2002). Parental education and young people's educational and labour market outcomes. In Kogan, I. and Müller, W. (eds.) *School-to-Work Transitions in Europe: Analyses of the EULFS 2000 Ad hoc Module*, pp 5–29. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung.
- ILO (2005). *Key Indicators of the Labour Market*, 4th edn. Geneva: ILO.
- Lamont, M. (2000). *The Dignity of Working Men: Morality and the Boundaries of Race, Class, and Immigration*. Cambridge: Harvard University Press/Russell Sage Foundation.
- Lasonen, J. (ed.) (1996). *Theory into Practice No 92: Reforming Upper Secondary Education in Europe. The Leonardo da Vinci Project Post-16 Strategies*. Jyväskylä: Institute for Educational Research, University of Jyväskylä.
- Lasonen, J. (ed.) (1999). *Ammatillisen koulutuksen arvostus Euroopassa ja Suomessa (Esteem for and Quality of Initial Vocational Education in Finland and in Europe: Pathways from Upper Secondary Education to Employment and Further Training)*. Jyväskylä: University of Jyväskylä, Institute for Educational Research.
- Lasonen, J. and Young, M. (eds.) (1998). *Strategies for Achieving Parity of Esteem in European Upper Secondary Education*. Jyväskylä: Institute for Educational Research, University of Jyväskylä.
- Leney, T. and Green, A. (2005). Achieving the Lisbon goal: The contribution of vocational education and training. *European Journal of Education* 40(3), 261–278.
- Rauner, F. (ed.) (2005). *Handbuch Berufsbildungsforschung (Handbook of Vocational Education Research)*. Bielefeld: W. Bertelsmann Verlag GmH.
- Smyth, E. (2002). Gender differentiation and early labour market integration across Europe. In Kogan, I. and Müller, W. (eds.) *School-to-Work Transitions in Europe: Analyses of the EULFS 2000 Ad hoc Module*, pp 31–60. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung.
- Trant, A., Branson, J., Frangos, C., et al. (1999). *Reconciling Liberal and Vocational Education*. Dublin: Curriculum Development Unit.

Vocational Education and Training in Post-Conflict Countries

E Lyby, Jacaranda Consult, Denmark

© 2010 Elsevier Ltd. All rights reserved.

Conflict forms part of the human condition. Since time immemorial, people have been fighting over power, resources, and ideas. Thomas Hobbes, in his monumental and controversial 1651 work, *The Leviathan* considered that

“...during the time men live without a common Power to keep them all in awe, they are in that condition which is called Warre; and such is Warre, as is of every man, against every man.” (Hobbes, 1651:)

However, to counteract the atrocious consequences of eternal Warre, Hobbes arrived at what he called the “fundamental law of nature,” which is “to seek peace, and follow it.” This can be done through a commonwealth – a government or other central power to which people have delegated the right to conduct violence on their behalf through a social contract. Hobbes’ books were burnt in public when they came out, but his ideas inspired later thinkers such as John Locke, and were reflected in the drafting of the American Constitution.

This article addresses conditions for vocational education and training (VET) in post-conflict countries. In the early twenty-first century, this refers by and large to countries in the developing world, roughly Africa, Asia/Pacific and Latin America, as all recent conflicts have taken place there – the industrialized or developed countries of the West have been involved in many of these conflicts but not on their own territory, hence the focus on developing countries.

The use of the term developing countries for nations in Africa, Asia, and Latin America tends to imply that these are in the process of developing their economies and institutions on the road toward modernity as known in the Western world. In many countries, protracted armed conflict has severely disrupted this development. A new paradigm is therefore underway, by which the way to development is no longer seen as following (more or less) the same path. New structures emerge in the globalized age that tend to weaken state power and disrupt traditional economies. In addition, rising levels of violence and misery are becoming accepted as normal, and new forms of humanitarian aid intervention, far from solving the problem, accommodate and coexist with this instability and inequality.

The Nature of Conflict

Conflict takes place both within and between social actors – the family, village, clan, tribe, nation; or the economic,

cultural, or religious group. Its nature, extent, and level of intensity vary in time and place. The first half of the twentieth century saw two world wars, in which the major powers of the time confronted each other head-on, concluding with the development and use of the nuclear bomb. Many conflicts were conducted by proxy in the developing countries during the Cold War, with the major powers supporting their junior allies directly and indirectly. In addition, an increasing number of conflicts became internal, civil wars, rather than conflicts between nations.

A total of 228 armed conflicts have been recorded after World War II and 118 after the end of the Cold War. In 2004, there were 30 active armed conflicts; this number includes seven conflicts that had become inactive since the previous year, a new one that had started, and seven that had restarted of which three had action taken by new rebel groups and four by previously recorded actors.

When is a conflict over? When is it really post-conflict time, when some 50% of conflicts resume within a year after the peace treaty has been signed? The term post-conflict itself therefore becomes ambiguous: is the conflict really over, or are we in-between two phases of it? Key to this issue is the level of security that can be maintained, and security and development are increasingly seen as interdependent – post-conflict reconstruction and development are impossible without a reasonable degree of security – and security cannot be sustained in the longer term without reconstruction and development taking place. The very different cultures of the military forces and the development-aid community grapple with finding a suitable mode of coexistence in the post-conflict situation with focus on reconstruction.

Armed conflict inevitably adds to the already existing problems of developing countries, causing great social and economic destruction and often leading to the displacement of millions of people. As peace is gradually restored, large numbers of refugees, internally displaced persons (IDPs), and ex-combatants wish to return to their communities, thereby placing an additional burden on overstretched resources. Gainful employment is in short supply (see **Case Study 1**).

New Opportunities?

Post-conflict reconstruction can offer new opportunities to reinvent VET systems from a clean slate, but such development is most often hampered by political competition

Case Study 1 Vocational education and training (VET) in developing countries

Some developing countries draw benefits from the increasing globalization of markets, using the comparatively cheap costs of human labor in mass production for exports to achieve high economic growth rates. The need to produce goods of high quality for the world market gradually calls for the upgrading of skills of the workforce, and human resource development in such situations becomes a key factor for companies that want to be competitive – something they must invest in.

However, many developing countries still only participate in the global markets to a very limited degree, and often only in export of one or two primary products such as oil, minerals, or agricultural produce. Even without having been exposed to violent conflict, the VET systems in such countries is often one of neglect. The following characteristics apply in many countries:

- Modern-sector economy is small and often based on one or two products. The labor market is dominated by an informal (microenterprise) sector, in which informal apprenticeship is the dominant mode of skills transfer. The informal sector is facing harsh competition from cheap imported products, hence has limited potential for growth on its own. The skills development of the informal apprentice remains at the level of what the master craftsman already knows.
- Government-owned and -run VET systems are inflexible and try to cater to a formal sector industry of yesterday. VET standards set 40–50 years ago were perhaps suited to conditions during colonial times but fall short of the requirements of the globalized era.
- Formal VET institutions are starved of resources and driven by supply rather than demand. Their management and instructors are neither interested in, nor capable of orienting themselves to the informal sector on the one hand, or to keep up with technological development on the other. They produce poorly qualified graduates that are rejected by employers in the formal, as well as informal sectors.
- Motivation to join VET programmes is often low among youths and their parents, who prefer secondary and tertiary education rather than vocational training. If a training course is chosen as a last resort, the motivation is likely to be low, the learning may be ineffective, and the proportion of VET candidates actually putting their acquired skills to use is often low.

for resources and power bases, shortage of skilled trainers, no reliable labor-market information, as well as very weak administrative support systems. Add to this security concerns including private armies, landmines, and return of refugees in large numbers with high expectations for a better future.

The potential to seize such opportunities obviously vary greatly between post-conflict countries due to factors such as the level and intensity of the conflict; whether it has actually stopped; geographical, political, and cultural characteristics; and the level of external military and economic support available to win and secure a sustainable peace. The differences between the post-conflict situations are more conspicuous than the similarities, making generalizations hard to arrive at. The examples

described in the sections below serve as illustrations of the diversity more than as indicators of general trends.

The two countries to be described are South Africa after the long but relatively low-intensity struggle against apartheid, and Afghanistan after 23 years of occupation and civil war, most of it with high-intensity violence. In addition to the different types of conflict, the two countries also represent dissimilar sociogeographic features, with South Africa as a unique combination of the two worlds – the industrialized and the developing world – within its own borders at the southern tip of Africa, and Afghanistan, a tradition-bound Asian country first caught in the crossfire of the Cold War and then in self-destructive internal fighting.

South Africa

Industrial development in South Africa took off with the discovery of diamonds in Kimberley in 1869 and gold on the Witwatersrand in 1886. Control over these resources was at the heart of the Anglo-Boer War 1899–1902, ending with the establishment in 1910 of the Union of South Africa, a *de facto* self-governing entity within the British Commonwealth. Economic and political power was in the hands of the white minority, consisting of Afrikaans-speaking descendants of Dutch settlers and more recent immigrants who were mostly English-speaking. A constitution based on political parties was put in place, however for whites only.

The mining industry and, increasingly, a manufacturing industry, provided job opportunities for many workers. Mining especially attracted young Afrikaans-speaking men from the farming areas, most of them with little education. The nature of the work in deep-pit gold mining called for large numbers of unskilled or low-skilled workers, beyond what the white community could supply. This gave rise to the migrant labor system, by which Africans were given time-bound contracts to fill the lower ranks of the workforce under the supervision of white workers. In theory, Africans were regarded as agricultural tribesmen outside the modern world, who did stints of work for a while in order to earn money for a specific purpose such as marriage or acquisition of land. In reality, the government applied a head tax on Africans that compelled them to work for money and hence become part of the industrial world.

The jobs of white workers came gradually under threat from pressure from the Africans who were paid much less for the same work. This led to a general strike by white workers in 1920 and the setting up of a Job Reservation Act by which a specified number of occupations were reserved for whites. From the point of view of South African (white) industrialists this was an inefficient use of the available workforce, and from the point of view of the Africans an unacceptable discrimination against them and their livelihoods.

The migrant labor system continued and became the foundation of the apartheid (separateness) policy that was implemented after the victory of the Afrikaans-speaking National Party in 1948, which aimed at total separation of the ethnic groups in space and time, each group living in its designated area and only visiting other areas for specific reasons such as temporary employment.

The country was divided into areas for whites, coloreds (people of mixed race), Indians (descendants of indentured workers from India who were imported to work in the sugarcane fields in Natal in the 1860s), and Africans, with the latter required to carry and show a passbook to the police upon request when visiting areas of other groups. The result was a country that combined First- and Third-World institutions in parallel systems, one for whites – well resourced and run – and one for blacks – of poor quality and starved of resources. This dichotomy applied in full force to the educational and training systems.

Resistance against racial discrimination took shape with the formation of the African National Congress (ANC) in 1910, initially based on nonviolent principles of the South African Indian Congress formed earlier by Mahatma Gandhi. With the introduction of apartheid in 1948, peaceful resistance increased until the 1960 Sharpeville massacre in which police forces killed 69 unarmed black participants in a lawful demonstration against the detested passbooks. Sharpeville led to the ANC under Nelson Mandela to declare the armed struggle, initially against government installations and not against human targets.

ANC was banned as an organization and Mandela and most of ANC's top leadership were jailed on Robben Island in 1963, after which ANC for many years primarily operated from exile elsewhere in Africa and in Europe. The pressure on the South African governments grew with international isolation through boycott of goods and participation in sports events, as well as with the ANC's increasing capability to upset the functioning of the system through mass action. The pressure from below was met with police brutality, torture, and inexplicable deaths in detention. However, it became gradually more and more obvious that the system was unsustainable in the long run, and apartheid finally died in April 1994 with the first elections based on universal suffrage, which gave power to the ANC with Mandela as the first black president.

From Apartheid to National Qualifications Framework

Recent development in the South African VET sector can arguably be divided into three phases.

1989–94: From resistance to policy thinking

This period was driven by the process toward abolition of apartheid, a political objective that claimed all the

attention of the involved actors. Important was the role of the ANC-affiliated Congress of South African trade Unions (COSATU), which had strong links to unions abroad and was able to benefit from their experience. As primarily freedom fighters with limited experience in the workings of a normal labor market, this support was useful to the South African unionists in the process when ideas of competency-based modular training and recognition of prior learning became prominent. It was in this period that COSATU initiated the National Training Strategy Initiative in which many items were negotiated with other stakeholders in the National Training Board. It was also in this period that employers and unions agreed in principle on some form of a National Qualifications Framework (NQF) that would integrate education and training. An important point for the unions in this debate was to arrive at a system that would include even those with little or no education at all.

1994–99: Turning policy ideas into legislation

Mandela's presidency saw the birth of major development programs, notably the Reconstruction and Development Program (RDP) and the subsequent Growth, Employment and Redistribution (GEAR) program, both with ambitious growth targets in employment, housing, and other areas. In reality however, South Africa experienced a significant decline in formal sector employment in this period – 600 000 formal jobs were lost between 1996 and 2000 – partly due to hesitance by foreign investors, and against an annual growth of the labor force of more than 400 000.

The NQF was elaborated in 1995–96 in order to reconstruct and develop the education and training system into an integrated approach which could enhance access to, and mobility and quality within education and training. To oversee the quality of the NQF, the South African Qualifications Authority (SAQA) was established in 1995. This process was supported in 1997 by the adoption of a Green Paper that proposed a transformed skills development system based on three mechanisms: (1) a levy–grant system, aimed at increasing the involvement of, and contributions by, employers in training of their workforce; (2) establishment of a series of Sector Education and Training Authorities (SETAs); and (3) introduction of a new notion of learnerships.

The NQF consists of eight levels divided into three bands: General, Further, and Higher Education and Training, with basic adult education (literacy training) at the bottom but still as part of General Education, to postdoctoral research degrees at the top of the third band. While this level of integration is clearly a very ambitious agenda for any country, its implementation was actually divided between two agencies, the Department of Education and the Department of Labour, an arrangement that has led to many difficulties.

1999 and later: Implementing the vision

Nelson Mandela was succeeded as president by Thabo Mbeki in 1999. Mbeki's presidency has been tasked with transforming the ideas into practice. The skills levy has been introduced but the lack of administrative capacity to make it a major financing source of skills development and VET has reduced its efficiency, at least in the early years. The GEAR program has helped establish macro-economic stability but has not produced the number of formal sector jobs that were intended.

The VET is placed in the Further Education and Training (FET) band of the NQF. Legislation in 2000 on the FET Certificate stipulates that public FET colleges will be merged into larger entities with 2000 full-time FET learners as the minimum institutional size. Similar mergers take place within higher educational institutions. In the case of VET, this means merging former white VET colleges with former black ones into much larger units. The extent of this challenge on the (white and black) instructors and students, as well as the amount of redistribution of resources between them is daunting.

In conclusion, it can be said that South Africa has succeeded to a considerable degree in overcoming the legacy of decades of conflict. Access to education and training has become much more equal, NQF is undergoing continuous development, and results are beginning to show. On the reverse side, job growth is grossly insufficient and much talent has been lost through emigration of whites. Human Immunodeficiency Virus (HIV)/Acquired Immune Deficiency Syndrome (AIDS) takes a hard toll on qualified people who fall victim to the disease, and the need for replacements puts extra stress on the education and training systems. Add to this continued complaints that the distribution of resources and opportunities remains as skewed as ever. While South Africa had managed to emerge from conflict and become a stable country without major fallbacks to violence, the country is still faced with huge challenges.

Afghanistan

The *coup d'état* of Daoud Khan against his cousin the king, Zahir Shah, in 1973 concluded 40 years of royal governance, and Afghanistan was declared a republic. Daoud claimed to promote modernization of the country and turned toward the West, something that did not please his neighbors to the North, the Soviet Union; he was already at odds with the southern neighbor, Pakistan. In April 1978, he was himself killed in a communist coup.

This was followed by the invasion of Afghanistan by the Union of Soviet Socialist Republics (USSR) in December 1979 and 10 years of Soviet occupation, a period known as the *jihād* (holy war) where private armies under *mujahedeen* (holy warrior) commanders, were supported by the

Central Intelligence Agency (CIA) and Saudi Arabia through the Pakistani secret service Inter-Services Intelligence (ISI). This support was intensified in the second half of the 1980s and became a key factor in the dissolution of the USSR. Militarily, the development and supply of the shoulder-fired heat-seeking Stinger missile to the *mujahedeen* became a turning point with its capacity to destroy the Soviet Hind gunships of which almost 200 were lost in 1988.

The exit of the Red Army in 1988 did not bring peace. Instead, the many armed groups fought each other in a series of constellations where alliances changed constantly. This period was even bloodier than the occupation; in 1994 alone, half of Kabul was destroyed and 25 000 people killed, and the fighting continued in most of the country. The Taliban (*talib*: scholar in Islam) fundamentalist movement entered the fighting in 1994 and gradually advanced throughout the country, taking Kabul in 1996 and most other major towns thereafter. The Taliban introduced a particular strict version of *sharia* (Islam: the road) legislation; banning women from work; girls from schooling; making music, TV, films and photos, sports, games, etc., illegal and subject to severe punishment.

Following the September 2001 attacks on the World Trade Center in New York and the Pentagon in Washington DC, the US army linked up with the Northern Alliance consisting of anti-Taliban *mujahedeen* commanders in the pursuit of Osama Bin Laden and the overthrow of the Taliban. The operation succeeded, and an Interim government under Hamid Karzai was established in Kabul.

The Karzai government had its mandate extended through general elections, but the security situation remained fragile, as Taliban regrouped and prepared new attacks. The International Security Assistance Force (ISAF) under American command was supplemented with North Atlantic Treaty Organization (NATO) troops from several countries, as a national Afghan army was built.

The Aftermath

Throughout the twentieth century, Afghanistan has been among the poorest countries in the world. Twenty-three years of bloody conflict did not improve things – 1.5 million Afghans had died, more were injured, many had been subjected to physical or psychological torture. Afghanistan had become the country in the world with the highest infestation of landmines, and many had lost one or both legs as a consequence of these. Four million had become refugees mostly in Pakistan and Iran, and more than a million became internally displaced. A special problem related to the ex-combatants in the private armies who had to be disarmed and reintegrated into society. In 2002, the big return started. Millions came back to look for their land, their property, or a job.

Most returnees had spent many years abroad and some of them had acquired new citizenship in the country of asylum, as well as knowledge, skills, and even some capital that could be invested. However, the majority returned with low levels of human capital. In 2004, Afghanistan's overall adult literacy rate was estimated at 36%, about the worst in the world; of this, the literacy rate for women was only 21% while the rate for men was 51% – however, these figures hide huge urban–rural discrepancies with over 90% illiteracy for women in some rural areas. Other alarming social indicators include a fertility rate of 6.3 children per woman; an under-5 mortality rate of 257, equivalent to one in every four children dying before their 5-year birthday; and a maternal mortality rate of 1600 per 100 000 live births.

Located on the Silk Route at the crossroads of Asia, Afghans for many centuries had trading links with countries far away. While trade had been disturbed throughout the years of conflict, some of the returnees brought with them new contacts to other, often remote, parts of the world, hence opening new opportunities for trade and production. However, this took place in a setting in which the infrastructure was destroyed and poverty and lawlessness were widespread. The first major economic activity to recover was the cultivation of opium poppy. In spite of government-led and internationally sponsored campaigns to reduce or eliminate poppy, Afghanistan became the world's largest exporter of raw opium, accounting for three-quarters of Western Europe's heroin market with a street value of USD 30 billion in 2004. Opium became Afghanistan's largest industry and is estimated to have grown from 76% to 87% of the world's production between 2002 and 2004.

A New Beginning?

Robust data on the economy and labor market in Afghanistan are few and unreliable. The population was estimated at 24 million in 2004, of which only 20% lives in urban areas, but it was a widespread assumption that the real size was close to 30 million. 80%–90% of enterprises are considered to be informal (unregistered and largely untaxed) family businesses. The labor force was estimated at 8 million with 20%–30 % open unemployment, and high underemployment in the labor-intensive informal sector due to low productivity. Investment in the reconstruction effort by international donors has led to a boom in the construction industry, reinforced by wealthier Afghan returnees who build houses and shops for themselves, especially in Kabul. Most of the construction work is carried out by foreign companies.

Given the challenges ahead, the demand for skills development is extensive and covers a wide range of different types of skills. The country needs masons, road builders, farmers, legal scholars, teachers, carpenters, electricians, hair dressers, accountants, governance experts, office

managers, painters, engineers, and persons in many other professions. However, asking Afghan formal-sector companies about the constraints they are faced with, human capital and skills development is low on the list, which is otherwise dominated by infrastructure issues (electricity, accessibility), administration (corruption), and lack of finance capital.

Responsibility for and delivery of the post-conflict VET sector in Afghanistan can be divided into five categories:

1. Public provision by three ministries:
 - a. Technical and Vocational Education and Training (TVET) by the Ministry of Higher Education (MoHE). Part of tertiary education tier.
 - b. Forty-two VET institutions under the Ministry of Education (MoE). Infrastructure in poor state. Courses 2–5 years with Grade 9 as entry criterion for most courses.
 - c. Seventeen training centers under the Ministry of Labour and Social Affairs (MoLSA) offer nonformal training courses – a variety of mostly practical courses with no common entry criteria.
2. Vocational training by nongovernmental organizations (NGOs). Important especially during the war, but later squeezed out by the government ministers who favor public provision. NGO training is mostly urban based. Most trainers are recruited among craftsmen from the locality. Trainees receive a stipend. NGOs carry out training for donors as well as for MoLSA.
3. On-the-job-training (OJT) by employers takes place in some enterprises and larger reconstruction projects.
4. Private training institutions. Urban based and new, many of these cater to a great demand in the post-conflict situation for computer training and English-language courses for adults.
5. Informal apprenticeships. Impossible to measure but undoubtedly a major mode of skills transfer taking place in the informal micro- and small enterprises (MSEs).

Skills development is seen by the government as important and some VET programs are included in the so-called Afghan Compact which sets the framework for the government's main development programs. However, the emphasis is clearly on public provision where the responsibilities of three ministries often overlap. Streamlining of public VET provision is needed and the setting up of a National VET Authority planned for the medium-to long term, in recognition of the difficulties involved in reorganizing the mandates of the involved ministries.

The central program to restart the VET sector is the National Skills Development and Market Linkages Programme (NSDP), which aims to tackle the following problems:

- Many school dropouts and future high-school graduates are seeking work.

- There is a lack of consistency, quality, and coordination within and among various training providers.
- There is a lack of standardized curricula and qualifications, as well as an overall national VET policy.
- There is no systematic collection of labor-market information.
- There are wide divisions in approaches of MoLSA and MoE training departments.
- There is a lack of qualified staff at MoLSA and MoE with little experience of modern technological development.

The overall goal of the NSDP is to contribute to the socioeconomic recovery of Afghanistan through the provision of a national vocational and training system that is responsive to labor-market needs and provides Afghan women and men with the knowledge and skills for decent work. To achieve this goal the NSDP pursues three objectives, thereby combining short- and long-term tracks in one strategy:

1. Direct provision of training through government and private training providers and NGOs: in the short term, to provide skills for wage or self-employment to 150 000 unemployed Afghans, of whom at least 35% are women.
2. Capacity building of training providers and trainers in pedagogical methods and curricula development.
3. Establishment of an enabling environment:
 - a. In the short term through conducting labor-market surveys, mapping, and registering training providers.
 - b. In the long term by setting up a NQF, developing a VET policy, and establishing a National Vocational Education and Training Authority.

Afghanistan is facing a steep agenda in the VET sector and elsewhere. Security remains a problem with the re-emergence of the Taliban. Poverty is endemic. The infrastructure is dilapidated. The external funding for reconstruction has not come forward in the amounts and with the speed that was hoped for; corruption in the civil service together with the high cost of the military efforts take up a large share of the investments that should go into reconstruction and development.

The VET and skills development programs in Afghanistan are overwhelmingly targeting the small formal sector. This is surprising in a country where the vast majority of enterprises – 80%–90% – are thought to be informal MSEs, and where the majority of skills transfer takes place in the form of informal apprenticeship. The limitations here often lie in the skills of the master craftsman himself – he can only teach what he himself knows already. Some programs have been successful when focusing on the upgrading of skills of master craftsmen, or in the form of sponsored apprenticeships, as shown in (Case Study 2). Such programs have managed to combine the

Case Study 2 Employment of ex-trainees

A study of the employment of Afghan youth who had attended conventional vocational training centers in Baluchistan, Pakistan, in the 1980s, showed that a very small percentage were able to gain employment or become self-employed using their skills. Studies of a sponsored apprenticeship program for Afghan refugees in North-West Frontier Province, Pakistan, in the early 1990s showed that a majority of the ex-trainees were in employment that used their skills.

From: Sinclair, M. (2002). *Planning Education in and after Emergencies*. Paris: UNESCO IIEP.

practical experience of being an apprentice in a real (small) enterprise rather than in a training institution with a low training budget.

Conclusion

Countries emerging from armed conflict need accelerated economic growth and development to bring about internal stability and a sustainable peace. Destroyed or dilapidated infrastructure is a major obstacle to investment and growth, as in Afghanistan. Human capital is typically low due to disturbance of the schooling and VET systems, with many youths having become freedom fighters or combatants in private armies with no marketable skills in the post-conflict era. The apartheid system in South Africa intentionally kept the black majority from developing their talents and skills. The continued presence of firearms and animosity between former enemies calls for special programs for disarmament and integration of the ex-combatants into the national army, or into civilian life.

The two cases presented here illustrate the diversity of challenges that post-conflict countries can be faced with. Does the destruction of old systems through conflict offer new opportunities for reinventing VET from a clean slate? There is obviously no one answer to that question – the most common feature seems to be the great obstacles that lie ahead for such an enterprise. Another one is the recognition of the informal sector, which seems to be very difficult to deal with for governments, in spite of their dominance in terms of employment in many developing countries. When considering the high need for jobs in post-conflict countries, it would be advisable for governments to focus at least part of the VET effort on development of skills and productivity in the informal sector through assisted apprenticeships, upgrading of master craftsmen's skills, and access to credit and business development services, rather than exclusively on the formal sector which typically is small and capital intensive with few job opportunities.

Further Reading

- Ahmad, I. (2004). *Gulbuddin Hekmatyar—An Afghan Trail from Jihad to Terrorism*. Islamabad: Society for Tolerance and Education.
- Byrd, W. and Ward, C. (2004). Afghanistan's Drug Economy. World Bank Report no. SASPR-5.
- Duffield, M. (2001). *Global Governance and the New Wars*. New York: Zed Books.
- Harbom, L. and Wallensteen, P. (2005). Armed conflict and its international dimensions, 1946–2004. *Journal of Peace Research* 42(5), 623–635.
- Hobbes, T. and MacPherson, C. B. (1982). *Leviathan*. Penguin Classics (paperback), New edition.
- Lyby, E. (2001). Vocational training for refugees: A case study from Tanzania. In Crisp, J., Talbot, Ch., and Cipollone, D. (eds.) *Learning for a Future: Refugee Education in Developing Countries*, pp 217–259. Geneva: UNHCR.
- Marshall, M. G. Major Episodes of Political Violence 1946–2006, Center for Systemic Peace. <http://members.aol.com/CSPmgm/warlist.htm>.
- McGrath, S. and Badroodien, A. (2005). *International Influences on the Evolution of South Africa's National Skills Development Strategy, 1989-2004*. Eschborn: Crystal Working Paper, GTZ.
- Rashid, A. (2004). *Taliban: Islam, Oil and the New Great Game in Central Asia*. London: Yale University Press.
- Sinclair, M. (2002). *Planning Education in and after Emergencies*. Paris: UNESCO IIEP.
- UNODC (2003). *Global Illicit Drug Trends*. Vienna: UNDOC.

VOCATIONAL EDUCATION AND TRAINING – VET SYSTEM

Contents

Dual System

Industry Involvement in the Vocational Education and Training System

Non-Completion in Vocational Education and Training: Issues, Patterns and Possibilities

Planning and Policy Development for Technical Vocational Education and Training Systems

Qualifications Frameworks and their Role in the Reform of Education and Training

Quality Assurance in Vocational Education and Training

Training Finance

Training Markets

Vocational Education and Training Teacher Training

Vocational Education and Training Workforce

Dual System

T Deissinger, University of Konstanz, Konstanz, Germany

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Dual systems are specific structural and didactical patterns of vocational learning. Normally, they are referred to as systems of initial training, which can basically happen in three ways: (1) by attending a full-time vocational education and training (VET) course in a (vocational) school, college, or higher education institution with neither a training or an employment contract involved; (2) by gaining specific skills or competences in a company based on contractual employment (trainee or employee) – and therefore in a job-specific environment (learning on the job); or (3) by entering a VET program, for example, an apprenticeship, which uses (part-time) school-based and company-based modes of learning and therefore works as a dual system in a wider understanding of the term. In most cases, learners in the latter case are in some kind of contractual employment which can be, but does not need to be, an apprenticeship contract. However, it is the apprenticeship system which offers opportunities of vocational learning within a predominantly occupational context. This means that apprenticeships are not necessarily part of the formal secondary school system.

Looking at the modeling of VET pathways, it has to be observed that VET systems and their typical features are more than mere constructions, mainly triggered by political motivations or economic interests. Instead, in order to understand how they function and how they are capable

of reacting to external demands, they have to be looked at as historical entities. This is especially true for the dual system which at the same time, in some countries at least, functions as a more or less traditional apprenticeship training system. Examples for this subtype are Germany, Switzerland, Austria, Denmark, or the Netherlands. Even in Anglo-Saxon countries, such as Australia or the United Kingdom, apprenticeships, mostly with the additional label modern or new, have been reinvented in the last two decades, following reform purposes to establish alternative routes into employment besides nonformal on-the-job training, traditional school-based VET programs, and higher education courses meant to produce labor-market-relevant qualifications. However, structural and conceptual differences with countries using apprenticeship as a traditional model on a wider scale still remain.

In the following, the major focus will lie on the dual system of Germany. VET in the German-speaking countries of Europe mostly can be described as using alternating learning arrangements, above all as partnerships between schools and companies. Both in Switzerland, Austria, and in Germany, dual systems are specific historical derivatives of apprenticeships. However, despite industrialization being one of the major common features in modern history, differences between the German-speaking systems and most other countries in Europe in terms of a specific apprenticeship culture or learning culture, respectively, are obvious. Their relevance for the current debate

on VET and its modernization may be associated with globalization and various developments at the European policy level which address national VET systems in a specific way. These challenges get a national dimension insofar as solutions appear only possible against the background of what may be called the system reference of a given VET system. This includes different notions of what countries perceive as educational or pedagogical with respect to their VET systems and therefore as realization modes applied to link the idea of training for an occupation or a job to the notion of personality development of individuals.

The idea to bridge the historically and mentally separated worlds of education and training by establishing a specific alternating learning environment for school leavers and at the same time to define it as an apprenticeship system building up on a specific cultural heritage, is very typical for Germany, and also for its two German-speaking neighboring countries. The share of school leavers entering an apprenticeship in the dual system in these countries lies between 40% and 60%, although alternative routes, such as full-time VET or special integration programs, partly using the same setting, partly offering other kinds of training measures, have increased in importance. In Germany, this has been due to what has been called training-market crisis while in Austria full-time vocational schools have traditionally served specific sectors of the labor market quite successfully. Both in Switzerland and in Austria, initial VET can be combined with qualifications leading into the tertiary sector. Denmark, on the other hand, whose dual system is also discussed in this article, has a specific understanding of complementing apprenticeship with school-based VET. This is not possible in the case of Germany which refers to the German VET system as probably the most traditional type of a dual-apprenticeship system.

Structural Features and Working Principles of the Dual System of Initial Vocational Training in Germany

Key Features and Working Principles

In a comparative perspective, German firms employ a high proportion of the workforce with intermediate-level qualifications. The reason for this is that vocational training mostly occurs in the dual system (*Duales System*) which functions as the major nonacademic route for German school leavers by giving them formal access to the labor market as skilled workers, craftsmen, or clerks. The system has traditionally recruited between 50 and 60% of 16–19-year-olds and contributes to limiting the number of unskilled employees to a constantly low proportion in the

German labor market. Unlike in the UK or France, where they form a marginal sector within the vocational training systems, dual apprenticeships exist in nearly all branches of the German economy including the professions and parts of the civil service.

Hereby, the function of the dual system unequivocally is to impart initial training to school leavers in a given range of declared trades or recognized training occupations (*Ausbildungsberufe*) by using two sites of learning: the training company (*Ausbildungsbetrieb*) and the part-time vocational school (*Berufsschule*), with compulsory school attendance for all young people under the age of 18 not attending a higher or a full-time vocational school, hence covering virtually all who have entered an apprenticeship. Instruction can be part time during the week or on a block-release basis. Compulsory instruction in the part-time vocational school is both laid down in the various federal state school acts and indirectly regulated in the Vocational Training Act (*Berufsbildungsgesetz*), which defines the duties of both parties out of the training contract. When it comes to working time and working conditions, the Youth Employment Protection Act (*Jugendarbeitsschutzgesetz*) obliges employers to release young people to attend the vocational school during normal working hours.

The two learning sites correspond with two legal responsibilities due to the German federal political system. The various levels or dimensions of this dualism can be seen in **Table 1**.

Although the dual institutional principle certainly is the striking feature, the working principles of Germany's dual system also comprise at least three more dimensions which render it the character of a special kind of dual system:

Table 1 Levels of dualism in the dual system

	<i>Berufsschule</i>	<i>Ausbildungsbetrieb</i>
Legal status	• Public	• Private
Supervision	• School administration	• Competent authorities (chamber system)
Legal basis	• Education law (federal state)	• Vocational training law (central state)
Young person's status	• Student	• Apprentice
Training personnel	• vocational teachers	• Masters, trainers
Didactical instrument	• Vocational syllabus	• Training ordinance
Form of learning	• Classroom instruction	• Workplace or workshop instruction
Contents of learning	• Theoretical	• Practical
Kind of award	• School certificate	• Chamber-skilled award

- The dual system is a well understood and socially accepted pathway into employment as it follows a traditional pattern of skill formation recurring to the medieval mode of apprenticeship training. This means that training is workplace-led and predominantly practical and that work experience during the training period is seen as an essential asset of this kind of VET. It also implies the notion that the system works in accordance with skill requirements defined around the workplace since it is not task based but occupation based.
- Although traditional at its core, the modern German dual system is also subject to the involvement of the state with regard to the nature and quality of occupational standards as well as to legal conditions underlying apprenticeship training. The German training culture is based on the notion that an apprenticeship should be based on an underpinning pedagogical understanding which sets it apart from normal work.
- As the government sets quality standards with respect to in-company training only in a formal manner, other social groups have a major influence on the dual system. This means that public, private, and semi-private institutions use long-established modes of cooperation within the system and that employers and unions normally take the initiative with respect to training regulations and their revision or modernization.

The specific occupational character of training (*Beruf/Berufsprinzip*) is interlinked with the structural composition of the dual system. This orientation in fact can be traced back to the legal restitution of the master apprenticeship and the development of the vocational character of the continuation schools some 100 years ago. The historical reinvention of the principle of self-administration turned out as the starting point of a consolidation and universalization process which at the beginning of the twentieth century also incorporated industrial and commercial training, thereby creating a general institutional principle for the division of labor and the assignment of competences.

Besides its didactical principles, and its legal and institutional characteristics, the German system relies on a functioning training market which has the character of a suppliers' market. Once a training contract has been signed, the principal financial responsibility for the training process that includes, besides training allowances, all direct and indirect costs such as training personnel, machinery, training administration, and social insurance contributions, lies with the companies. The fact that the system is supposed to be financed mainly by employers reflects the principle of self-government reaffirmed by law in the late nineteenth century. Therefore, companies provide training opportunities on a voluntary basis. Training in the craft sector has a particularly strong tradition (although with a decreasing tendency), and the chambers (*Kammern*) execute public

functions such as the organization of examinations and the supervision of training companies.

Against this background, the German dual-apprenticeship system may be viewed as a system of training rather than a system of employment in which the wages of apprentices reflect this emphasis, with German apprentices typically receiving wages that are far lower than adult rates and apprentice rates in Australia or in the UK. Training allowances are the result of collective bargaining but keep attached to the purpose of giving young people a basic start into their working lives without putting too much burden on employers. As the apprenticeship system is neither part of the school or education system nor a normal sphere of work the system reference is clearly training and recruitment for skilled work. Such a clear separation of pathways or subsystems implies that expectations which rest on the dual system and frictions on the training market can hardly be compensated without additional activities on the side of the state. Among these, the promotion of external training options and the introduction of incentives for employers have been the most important ones in recent years.

Historical Background and Cultural Underpinnings

The German dual system in its present form goes back to the corporatist framework established by legal sanction in the late nineteenth century. It has remained virtually unchanged in its crucial features up to the establishment of the Vocational Training Act (*Berufsbildungsgesetz*) passed in 1969 and recently (April 2005) revised. Around the turn of the century, compulsory attendance at the part-time vocational school emerged as the second pillar, along with workplace training, underlying formalized vocational training. Georg Kerschensteiner (1854–1932), widely known as the father of the German vocational school, was successful in bridging the gap between elementary school and the beginning of military service by establishing vocational schools for school leavers. Although the continuation schools hereby followed the ideal of *Menschenbildung* (education of the individual) Kerschensteiner saw the individual also as a social being, both with respect to his occupation and to his citizenship within the community. This meant a complete break with traditional educational thinking. Kerschensteiner's prize essay delivered to the Erfurt Academy of Sciences in 1901 must be seen to comprise thoughts that were to be of revolutionary significance for the German education system and eventually emerged in the dual system which we have today.

On the company side of the dual system, the historical development appears ambivalent. In 1845, a Prussian Trade Act, though not repealing freedom of entrepreneurship as a whole, reimposed restrictions on free craftsmanship by making a distinction between small workshops and

industrial premises confining the right to take and train apprentices in handicraft occupations to examined journeymen. The definition of apprentice, however, was extended to everybody who entered an employment with a master to learn a trade notwithstanding whether he paid a premium or became a wage earner. Freedom of enterprise was finally incorporated in the Trade Act of the North German Federation in 1869. Three decades later, this was counterbalanced again with the restitution of the old apprenticeship model. In the 1880s, the *Mittelstandsbewegung* (small-business movement) constituted an influential pressure group. The 1897 Trade Act (*Handwerkerschutzgesetz*) was one of the outcomes of this policy. It is now seen as an oblique predecessor of the Vocational Training Act. Although it did not prescribe the *Meisterbrief* (master's certificate) as a training prerequisite, the Act of 1897 revived some of the old apprenticeship regulations. The newly established chambers and guilds became systematically involved with training matters and were given the right to hold examinations for journeymen and master craftsmen. This Act also made provision for the technical qualification required for the training of apprentices by confining it to skilled journeymen of at least 24 years of age who had either served a 3-year apprenticeship or pursued their trades for at least 5 years as independent artisans. Contracts became general practice in the craft sector as well as during the 3-year training period, at the end of which the apprentice should have the opportunity to take his examination. The *Handwerkerschutzgesetz* laid the foundations of the corporatist framework typical of the dual system. Ironically, the new system of industrial training emerged in a climate of anti-industrial agitation holding up the idea of *Beruf* (vocation/occupation) in a world of accelerated change. The pivotal amendment which modern apprenticeship legislation later added to the foundations laid in the late nineteenth century is the involvement of both the state and the trade unions within a system which has remained corporatist in its basic features.

The Critical Relationship Between the Dual System and Alternative Forms of VET in Germany

Recently, in Germany, initiatives that have been taken to assist school-leavers' access to VET qualifications have included several strategies. These strategies comprise, among others, stronger support of sectoral, regional, and in-company training-place initiatives and collaborative vocational-training ventures as well as training schemes organized by external providers in close cooperation with local enterprises with a view to developing sustainable VET structures. New stipulations in the revised Vocational Training Act (2005) are meant to link up vocational preparation programs and full-time VET with the formal apprenticeship system more reliably. Hereby, coping with

the rising number of participants in school-based vocational preparation courses as well as full-time students in VET, of whom only some 50% attend courses leading to vocational qualifications that are nationally portable on the labor market, is a major challenge. To improve the status of these courses, it is necessary to strengthen the work-related features of VET carried out in full-time vocational schools in order to open up genuine alternative pathways and opportunities for young people outside the dual system.

Traditionally, in the German debate on VET, there has always been an understanding that company-based and school-based training represent different pedagogical logics based on diverging paradigms of learning. This may be seen as the other side of the coin when it comes to the dual system. Whereas VET in schools has been associated with a more or less unambiguous pedagogical ethos and therefore not purely with socialization and utilitarian principles, training in an enterprise is bound to occur within an economic environment where normally a strong bias on non-educational purposes prevails. This difference in character is underlined in the dual system of Germany by the fact that even the part-time vocational schools use syllabuses which make provision for the core of the occupational curriculum as well as for additional general education.

Apart from the parking function or buffer function of (full-time) vocational schools due to training-market restraints, the relationship between the dual system and the various subtypes within the system of school-based VET under the auspices of the federal states has to be described as ambivalent. This means that vocational schools basically serve three functions which coexist depending on the specific course and the institution offering it:

- vocational preparation (mostly 1–2 years) which helps young people to apply for an apprenticeship by improving their stakes on the training market;
- further education (mostly 2–3 years) which means leading young people to achieve a higher school-qualification level (including the university entrance qualification); and
- vocational training (mostly 2–3 years) which means leading young people to achieve a portable labor-market-relevant occupational qualification outside the dual system.

With respect to the vocational-training function, school-based VET is considerably complex since full-time vocational schools offer courses leading to qualifications either within or without the scope of the Vocational Training Act. Besides, some of the schools deliver entry-level training based on specialized federal regulations, such as in the area of health occupations. Especially, the ordinary full-time vocational schools accommodate a range of different students and aspirations. Among the major subtypes are both schools leading to a full occupational qualification and institutions which only partly

focus on occupation-relevant competences, as they deliver either school qualifications (such as the intermediate secondary school leaving certificate) or concentrate on vocational preparation. One of the biggest problems certainly is the lack of acceptance in the labor market of most vocational qualifications obtained in school-based full-time courses against the background of a dominant dual system.

In the federal state of Baden-Württemberg the implementation of practice firms (*Übungsfirmen*) in full-time VET, above all in vocational colleges (*Berufskollegs*) is one strategy to tackle the apprenticeship problem. The vocational college is a postcompulsory institution, and as such is an alternative to the later years of the grammar school that academic students attend. It is open to students, normally aged between 16 and 18, with a medium-level school-leaving qualification. Recent research shows that the recognition problem of qualifications gained in a vocational college still persists. Companies have strong reservations concerning learning in a completely theoretical environment and normally rate a dual system (apprenticeship) qualification higher than anything else offered in the VET system. Even practice firms, from their viewpoint, fail to come up with a reliable quality of skill provision as they indeed offer practical and realistic learning opportunities, but lack the benefits of real workplaces and company-specific socialization effects. It is not surprising against this background that both the government and the social partners continue to stick to the apprenticeship system as the still best-practice approach in the area of initial VET.

Features and Working Principles of the Dual System in Denmark

Denmark comes quite close to the German model of the dual system, which can be described as a neocorporatist form of regulation, where the social partners play a central role in the VET sector. However, in Denmark, just one-third of all young people undertake training in the dual system, which uses a block-release structure in the area of cooperation of learning sites. In 1956, the foundations of the modern Danish VET system were laid when theoretical instruction was taken away from evening schools and transferred to technical schools. Modern VET in Denmark now starts with a full-time VET foundation course (*grundforløb*), which can last between 20 and 60 weeks, depending on prior learning achievements or occupational fields. For the following main part of training, young people have to enter a training contract with a company. In contrast to Germany, the vocational college has a distinct supporting function and can even act as a contractor in the apprenticeship itself. In the real dual or cooperative system, apprentices spend two-thirds of their training time in the company. Training ends with a final

examination before a regional branch committee representing the chosen occupation.

The system in Denmark is much less self-focused than the German dual system as it offers additional qualification modules and transition routes into the tertiary sector. Starting with the early 1990s, VET reform in Denmark has been directed toward more parity of esteem between general and vocational education and toward establishing a uniform model out of an amalgamation of the traditional school-based forms of VET and the apprenticeship system. On the other hand, this does not mean inflexibility since programs are individually tailored within a given range of opportunities.

Cooperation between learning sites is an important issue in Danish VET policy. The purpose of this does not only aim at optimizing learning processes and learning quality but also at coping with individualization of learning pathways and institutional flexibility. Although this is also a feature of the VET systems in England and Australia, the Danish system probably comes much closer to the German dual-apprenticeship model where alternating periods of training in two different contexts represent the basic philosophy of the VET system.

Features and Working Principles of Apprenticeships in England and Australia

In England and Australia, recent political endeavors have been directed toward a systematic revival of the apprenticeship system. The schemes created have become known as modern apprenticeships (England) and new apprenticeships (Australia). In both cases, training follows the overarching principles of CBT (competence-based education and training).

The history of apprenticeship in Australia has predominantly been one of industrial relations issues and has been associated with a form of learning for the lower strata of society. This status given to apprenticeships was reinforced when some of its occupational areas were transferred to higher education, such as pharmacy. Today, it is part of the open training market where more or less structured subtypes of vocational learning exist, making use both of full-time and part-time dual models. The new apprenticeship arrangements have also introduced further elements such as training packages and user choice, and they cover school-based and part-time apprenticeships and traineeships as well as formal training that can be wholly on the job or off the job at a training provider. Moreover, they now have been made available to employees, which is a typical characteristic of the English modern apprenticeship scheme, too. In Australia, VET qualifications have now even been made available for study during secondary schooling. Hereby, students can go for a full or a part of a national qualification. A more clearly dual mode has

been created by establishing school-based new apprenticeships and traineeships which involve a part-time apprenticeship or traineeship-employment contract and studying at a college.

However, in contrast with the German type of dual training, apprenticeships in the Anglo-Saxon world are organized in a much more open, volatile way: In England, for example, the apprentice is expected to keep to normal working hours with an employer while undergoing on-the-job training. This allows the apprentice to achieve national vocational qualifications (NVQs). At the same time, the learner is supposed to spend time with a learning provider to gain key skills, and to study for a technical certificate. Hereby, the weekly balance of work and study depends on the type of apprenticeship and on the individual employer. There are models where apprentices are given time off work on certain days to go to a learning provider, often a local further education college. The dual system therefore is much more employment based and not regulated in a uniform way, let alone based on vocational-training law. Many researchers of apprenticeship training in the UK have consistently pointed to the structural weaknesses of modern apprenticeships lacking minimum training periods, compulsory part-time technical education, and other quality standards.

The absence of process regulation in the UK or Australia obviously corresponds with the competence-based approach in the area of skill certification. What matters here is demonstrated competence in the performance of work tasks and no substantial educational attainments. The Anglo-Saxon model of apprenticeship therefore is an interesting mixture of old and new without the political or institutional buttressing of the system typical of Germany, and without locating apprenticeships in senior secondary education, which is the case in Denmark. As a matter of fact, apprenticeship numbers have stagnated on a level which is seen by policy as too low for stabilizing the company-based entry-level training system. Under these circumstances, learning on the job in a more or less formalized manner, is still the dominant way of acquiring skills outside the system of further education. Skills remain job specific as they are not based on a broadly designed initial training program.

Conclusion

As already mentioned above, one of the crucial traits of the German apprenticeship system is its dual character. Whereas in other European countries, including the UK, on-the-job training – even under modern apprenticeship programs – is complemented by off-the-job training on a more or less voluntary basis, in Germany, it is mandatory. While there has been an ongoing discussion about the process character of vocational training in the

UK – including the scope for expansive participation of companies in workplace-related training, in Germany, the law provides the framework for dual apprenticeships by making sure that school leavers are kept within the educational system – although there is now a growing number of school leavers in school-based VET and various vocational preparation and transition programs. For each training occupation, the state education ministries, in line with training regulations under the federal law, work out syllabuses for the vocational and general subjects within a given occupation taught at the part-time vocational schools.

In other countries, dual systems in this narrow sense of the word, do not exist, with the exception of Switzerland and Austria which resemble the German model. In Denmark, the dual system has undergone a number of flexibility changes, including single-venue training in a vocational college and modularization. Against this background, the most traditional model of a dual system exists in Germany and may be described as comparatively resilient to change, which certainly poses problems in the context of European VET policy which tries to integrate obviously different cultural concepts of organizing labor, labor markets, and modes of skill formation.

See also: Apprenticeships; Characteristics, Scholarship and Research of Teacher Educators; The Status of Vocational Education and Training; Training and Learning in the Workplace.

Further Reading

- Broeder, C. (1995). *The Match Between Education and Work: What Can We Learn from the German Apprenticeship System?* Den Haag: Central Planbureau.
- Clarke, L. and Winch, C. (2007). Vocational education: Conceptual differences between Britain and Germany. In Beichel, J. J. and Fees, K. (eds.) *Bildung Oder Outcome? Leitideen der standardisierten Schule im Diskurs*, pp 75–90. Herbolzheim: Centaurus.
- Deissinger, T. (1996). Germany's Vocational Training Act: Its function as an instrument of quality control within a tradition-based vocational training system. *Oxford Review of Education* 22, 317–336.
- Deissinger, T. (1997). The German dual system – a model for Europe? *Education and Training* 39, 297–302.
- Deissinger, T. (1998). *Beruflichkeit als "organisierendes Prinzip" der deutschen Berufsausbildung*. Markt Schwaben: Eusl.
- Deissinger, T. (2000). The German "philosophy" of linking academic and work-based learning in higher education – the case of the "vocational academies". *Journal of Vocational Education and Training* 52, 609–630.
- Deissinger, T. (2004). Apprenticeship systems in England and Germany: Decline and survival. In Greinert, W.-D. and Hanf, G. (eds.) *Towards a History of Vocational Education and Training (VET) in Europe in a Comparative Perspective. Proceedings of the First International Conference, October 2002, Florence, vol. 1*, pp 28–45. Luxembourg: Office for Official Publications of the European Communities.
- Deissinger, T. and Hellwig, S. (2005). *Structures and Functions of Competence-Based Education and Training (CBET): A Comparative Perspective*. Mannheim: InWEnt.
- Deissinger, T., Smith, E., and Pickersgill, R. (2006). Models of full-time and part-time vocational training for school-leavers: A comparison

- between Germany and Australia. *International Journal of Training Research* **4**, 30–50.
- Géhin, J. P. and Méhaut, P. (1995). The German dual system: A model for Europe. *Industrielle Beziehungen* **2**, 64–81.
- Greiner, W.-D. (1994). *The "German System" of Vocational Training. History, Organization, Prospects*. Baden-Baden: Nomos.
- Grollmann, P., Gottlieb, S., and Kurz, S. (2003). *Berufsbildung in Dänemark: dual und kooperativ?* Bremen: ITB.
- Harris, R. (2001). Training reform in Australia – implications of a shift from a supply to a demand-driven VET system. In Deissinger, T. (ed.) *Berufliche Bildung zwischen nationaler Tradition und globaler Entwicklung*, pp 231–254. Baden-Baden: Nomos.
- Harris, R. and Deissinger, T. (2003). Learning cultures for apprenticeships: A comparison of Germany and Australia. In Searle, J., Yashin-Shaw, I., and Roebuck, D. (eds.) *Enriching Learning Cultures. Proceedings of the 11th Annual International Conference on Post-Compulsory Education and Training*, vol. 2, pp 23–33. Brisbane: Australian Academic Press.
- Idriss, C. M. (2002). Challenge and change in the German vocational system since 1990. *Oxford Review of Education* **28**, 473–490.
- Jorgensen, C. H. (2004). Connecting work and education: Should learning be useful, correct or meaningful? *Journal of Workplace Learning* **16**, 455–465.
- Keating, J. (2002). *Comparative Study of Vocational Education and Training Systems*. Adelaide: NCVER.
- Marsden, D. and Ryan, P. (1991). Initial training, labour market structure and public policy: Intermediate skills in British and German industry. In Ryan, P. (ed.) *International Comparisons of Vocational Education and Training for Intermediate Skills*, pp 251–285. London: Falmer Press.
- Maurice, M. (1993). The link between the firm and the educational system in vocational training: The cases of France, Germany, and Japan. In University of Laval and International Labour Office (eds.) *Vocational Training. International Perspectives*, pp 43–69. Laval: Department of Industrial Relations.
- Misko, J. (2006). *Vocational Education and Training in Australia, the United Kingdom and Germany*. Adelaide: NCVER.
- Raggatt, P. (1988). Quality control in the dual system of West Germany. *Oxford Review of Education* **14**, 163–186.
- Ryan, P. (2001). Apprenticeship in Britain – tradition and innovation. In Deissinger, T. (ed.) *Berufliche Bildung zwischen nationaler Tradition und globaler Entwicklung*, pp 133–157. Baden-Baden: Nomos.
- Weissshuhn, G. and Büchel, F. (1998). Germany. In Organisation for Economic Co-operation and Development (ed.) *Pathways and Participation in Vocational and Technical Education and Training*, pp 195–239. Paris: OECD Publications.

Relevant Websites

- <http://www.bmbf.de> – BMBF: Federal Ministry of Education and Research.
- <http://www.bibb.de> – Federal Institute for Vocational Training (BIBB).
- <http://www.bildungsserver.de> – Deutscher Bildungsserver: Neuigkeiten.

Industry Involvement in the Vocational Education and Training System

M Hughes and C Hughes, Aylesford Mews, Sunderland, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Brokers – Government-appointed agents who direct employers to appropriate further education and training providers, and provide continuing support for company development.

Employer engagement – A term introduced in the UK White Paper: Learning to Succeed (1999) to describe the active involvement of employers in the VET process.

Levy – A charge on employers, usually statutory, to finance training for themselves and other companies in their sector.

LSC – The learning and skills council, a government agency responsible for funding and regulation of the learning and skills sector within England.

OECD – Organization for Economic Cooperation and Development.

Preservice or initial VET – The education or training provision which prepares individuals for work in a general rather than occupational specific sense.

Responsiveness to employers – A term used to describe an open and flexible attitude to working with employers in order to meet their training and development needs.

Select committee – A cross-party committee of UK members of parliament responsible for the scrutiny of government departments' plans, policy, and spending.

TAFE – Technical and further education (in Australia).

Workforce development – The training and development of current employees.

and Northern Ireland and, to a lesser degree, in Scotland. The discussion will also refer to VET systems in other countries facing similar issues in relation to skills development and examine how supporting structures facilitate or impede industry's influence.

Industry is a major user of VET systems, both in terms of vocational qualifications and skills development. There are clearly benefits to employers from their involvement in VET. Securing industry's influence and interest in post-16 learning and workforce development may secure greater relevance in VET provision and qualifications. This in turn should support competitiveness and productivity and develop capacity to cope with changing demands. From a demand-side perspective, industry's engagement in learning could increase employers' commitment to learning and demand for skills. The workplace also provides a focal point from which to reach individual employees and to encourage their participation in learning.

Engaging Industry in VET

The Policy Context

Post-16 learning and skills policy worldwide is increasingly concerned with developing the skills and learning of the workforce to increase economic productivity. Common economic issues, such as globalization, technological advancement, and the increasing pace of change in products and working practice, together with the emergence of new and strongly competing economies, feature strongly in national and international discussions about the role of VET. Restructuring of industry to meet these demands – typified in descriptions of a flexible and adaptable workforce – must be matched with VET systems that are sufficiently nimble to provide the learning required by employers and their workforce – being both responsive to emerging needs and proactive in anticipating future needs.

The ongoing policy debate within the United Kingdom culminated recently in the Leitch Review of Skills, which described the skills profile that should be achieved by 2020 to maximize growth, productivity, and social justice. The development of the skills of the current workforce is seen as a priority because of the demographic profile of the UK workforce – 70% of the workforce of 2020 has already left compulsory education and the flow of young recruits into the workplace will reduce. While school standards have

Introduction

This article considers the influence of business and industry on the vocational education and training (VET) system and the different forms this may take. Its primary focus is on the United Kingdom, as this system has been subject to substantial changes which have then often been mirrored in the systems of other countries. While many of the initiatives described relate to England, most will have been adopted in the devolved administrations of Wales

risen, more than one in six young people leave school with basic skills needs, and the proportion of people remaining in education and training post-16 is below the average in the Organization for Economic Cooperation and Development (OECD) countries. The United Kingdom will need double attainment at most skills levels if it is to be within the upper quartile of the OECD countries.

The review places a significant responsibility on employers to develop skills and proposes a commission for employment and skills as a medium for increasing employer engagement and investment in skills. It also suggests that legislation should be introduced if employers do not voluntarily rise to the challenge of developing skills. These proposals have largely been accepted by the government, but the extent to which the subsequent action is strong enough to bring about change remains to be seen.

Incentivizing the Demand for Skills Training

Many VET systems are looking at the role of industry and attempting to clarify the contribution that it can make in helping people to become more employable and developing the relevance of skills supply. France, for example, has in place a train-or-pay requirement with entitlement to employees' continuing development funded by an employer levy and worker entitlement to learning. In the Netherlands, there is a powerful use of fiscal incentives, such as tax relief on trainee wages. In South Africa, there is a 1% wage-bill levy in the private sector to pay for worker training. Within the United Kingdom, free training for basic skills and for a vocational qualification at level 2 is available for employees in England through the Train to Gain program. Train to Gain also provides a brokerage system, to enable employers to identify their training needs and source training provision.

The Train to Gain program was preceded by the employer training pilots (ETPs), which were established in September 2002, in certain parts of England. The pilots aimed to test the effectiveness of an offer of free or subsidized training to employees without a level-2 qualification, wage compensation (of various levels) to their employers for giving time off to train, plus access to information, advice, and guidance. An evaluation of these pilots estimated that only around 10–15% of the training provided by the ETP would not have occurred without the initiative, suggesting that it had a small effect on the incidence of training among eligible employers. Most of the employers taking part also said they would have provided similar training in any event, although only a minority said that this would have been for the same people trained in the pilots.

However, a strong message emerging from the ETPs was that employers were more prepared to take part in training if this was provided in or near the workplace with minimum disruption to work patterns. This message needs to be heeded by providers of skills, who need to

develop more flexible learning opportunities, tailored both to the needs of individuals and employers as well.

The wage compensation within the ETP was not generally seen by employers as a major incentive to train, although this may have been due to the high proportion of large organizations taking part in the pilots. The Train to Gain program has dropped the wage compensation element of the pilots and initially restricted the offer of free training to full qualifications at level 2, with some exceptions where it may be available for level 3.

Some countries have introduced forms of compulsion to increase training of employees, often in the form of the right to time off for training. Takeup has varied considerably. In the United Kingdom, the scheme for 16–17-year-olds has a takeup rate of only 3%. In Denmark, time off for employees has an estimated annual takeup of 9% of the eligible population. In contrast, levies which require firms to spend more on training, can significantly increase the volume of training, as has happened in France in recent years. International statistics suggest that compulsion only works if the training is delivered flexibly and produces clear business benefits.

Relationship to the Supply of Skills Training

Incentives to raise demand for training need to be underpinned by strong links to the providers and qualifications within the VET system. To understand the rationale for industry's involvement in VET in the United Kingdom and the structural arrangements for its delivery requires an appreciation of the changing economic climate, public perceptions of the VET service, and the political debate surrounding this service.

As in most advanced countries, the United Kingdom experienced a significant economic recession during the 1970s and early 1980s. The recession forced industries to downsize and restructure. Similarly, the institutional, organizational, and funding arrangements for VET provision have undergone substantial change and reform, including the measures by which business and industry can exert influence.

Rising unemployment, particularly youth unemployment and the severe long-term and generational unemployment in the United Kingdom during the 1970s and 1980s led to the creation of the manpower services commission (MSC). The focus of the MSC was on job creation, rather than skills development. The trades unions were still in a position of strength, and the agreement to develop work experience and work preparation provision under the youth opportunities program was dependent on these not replacing the traditional apprenticeships or foundation training that employers already provided.

In effect, the diminution of the United Kingdom's key strategic industries, such as shipbuilding and mining, reduced the need for a mass of skilled workers trained collaboratively with large employers. The recession also

weakened the collective bargaining power of the unions and the MSC were later to develop traineeships, which eventually became modern apprenticeships (now apprenticeships). However, the takeup of apprenticeships is only now reaching expectations and perceptions of their quality remain low, reinforced by completion rates that have only recently averaged 56%. The legacy of the reduction in VET during this period is manifest in the low level of skills within the current workforce.

Sectoral Approaches

One of the most commonly found approaches to promoting industry's involvement in VET is through occupational-sector bodies. These bodies are intended to facilitate employers' engagement in skills development and to create a culture of learning across a sector. In Canada, for example, a national system of 30 sector councils has been in place since the 1980s. The councils have mandatory equal representation from employers, trade unions, and government. They were originally intended to be self-financing (by employers), but this was later abandoned and they remain government funded. As a consequence, they are seen as quasi-governmental organizations.

The sector councils have a strong industrial-relations role and are seen to be effective in bringing the government, employers, and unions together. Their focus has shifted from labor-market issues, such as the impact of declining markets, to training. They are increasingly concerned with training standards and promoting a training culture. Standard setting is said to have enhanced employer involvement. Provincial sector councils and provincial colleges form affinity groups to translate sector skills needs into training provision. This is seen to be very successful, but there are tensions between employers' priorities and those of federal government.

Within the Netherlands, sector bodies are at the heart of the VET framework in relation to securing occupational competence. They accredit work-based trainers, provide quality control of examinations, and accredit company training schemes.

In France, there is a highly consensual view on the benefits of a levy system for all companies and all sectors. This system covers initial vocational training for young people and continuing vocational training for those in work or retiring. From 2004, employees have a right to up to 20 h of training per year of service. Employers meet all costs including an allowance for out-of-hours training. There is no national sector network and collection of the levy can be sectoral, multisectoral, or for specific employers.

A sectoral approach is well established in Australia though there is continuing federal-government impatience with the effectiveness of the system. The most recent manifestation of this has been the creation of 24 employer-led Australian technical colleges, operating outside the states'

technical and further education (TAFE) system. Prior to this, in 2004, the industry training advisory boards, which were established in 1992, were restructured into ten industry skills councils (ISCs). These new bodies were formed on a tripartite basis with employers and unions, and were intended to be more strategic and facilitate employer demand for skills as the main driver for the supply side. However, their most significant contribution continues to be the system of training packages, which set the frameworks for the recognition and assessment of sector skills.

Interestingly, the German VET system, often seen as the gold standard of work-related education and training, is nonsectoral, although employers are part of the institutional and regulatory arrangements that develop VET frameworks. This highly centralized, corporatist approach has been under great strain since the reunification of Germany in 1990. The economy can no longer sustain the required number of apprenticeships, and many firms are now poaching skilled workers rather than developing them through training.

The United Kingdom previously had a network of industrial training boards (ITBs), which collected a training levy and secured a ready supply of suitably trained people for their industries. However, many companies, especially smaller firms, perceived the levy to be more beneficial to large firms and the ITB's influence on the relevance and quality of the supply of training to be weak. Most ITBs declined after legislation in 1964, removed their legal entitlement to collect the levy. Only two ITBs remain, the construction industry training board recently re-named construction skills, and the engineering construction industry training board (ECITB). Some became industrial training organizations (ITOs), which had responsibility for setting national occupational standards in the national vocational qualifications.

The UK's sector skills councils (SSCs) now have responsibility for promoting employer engagement, and for describing the learning content of vocational qualifications, but they have little power to impose compliance on learning providers or to secure employers' commitment to take part in the training and development required for the sector. The reluctance to impose a levy system on industry has resulted in a voluntarist approach to employer engagement, although the Leitch review strongly suggested that if industry does not respond to their role in developing skills by 2010, legislation will be introduced to secure a legal compulsion for employers to train.

More recently, the UK's SSCs have led to the development of sector skills agreements (SSAs), which will articulate the skills needed by the workforce and how these skills should be supplied. The SSAs are being created in partnership with the SSCs, trade associations, employer bodies, and organizations that supply and fund education and training. SSAs will describe the training needed by their sector.

This has already resulted in the development of new qualifications or apprenticeship frameworks.

A general message from sectoral approaches is the need to secure industry's commitment to the remit of the sector bodies and to establish the added value of their activities. It appears that sectoral approaches need to focus on a limited number of strategic cross-sector issues. Sector bodies also need clarity about their leadership role and their relationship to other representative bodies.

In practice, VET systems operate in an environment of interrelated funding regimes and regulations, regulatory and assessment bodies, sector-specific and regional development bodies, and so on. The complexity of the interrelationship of these bodies, and their potentially overlapping responsibilities can be confusing to employers, trainees, and providers. There is a need for a multi-stakeholder approach, which is aware of and appreciates the role of other players, and the direct involvement between employers and providers needs to be strongly promoted to secure their active engagement.

Reforming Vocational Qualifications

The changes to the industrial landscape after the World War II have had a major impact on VET systems. The United Kingdom was at the forefront of the development of competence-based qualifications, following the DeVille review, which led to the development of national vocational qualifications. In part, the review was undertaken because of industry's dissatisfaction with the time-served apprenticeships framework and qualifications, and their perceived lack of quality and relevance. The national qualifications framework now includes national vocational qualifications from entry level to level 5, and includes all qualifications accredited by the qualifications and curriculum authority (QCA), including general and academic as well as vocational qualifications. In theory, this enables learners, training providers, and employers to assess the broad equivalence of qualifications in terms of their type and level. However, the level of a qualification does not indicate:

- which jobs that qualification is suitable for;
- equivalence to other qualifications in terms of content or duration;
- equivalence to other qualifications in terms of size; and
- equivalence to any qualifications QCA has not accredited, some of which are highly regarded professional qualifications.

England has been slow to adopt credit accumulation and transfer (CAT) frameworks, although systems are in place in Wales, and higher education CAT systems operate across the United Kingdom. The lack of such a framework has hampered the development of more flexible VET qualifications. QCA is currently testing a unit-based qualification framework that will recognize a wider range of learner

achievements than the national qualifications framework. This will be more responsive to employer and learner needs, but is unlikely to be in place before 2010. Meanwhile, it is very difficult to accredit the bespoke training that employers need.

The UK's National Skills Strategy

The national skills task force further refined UK policy on skills and the resulting skills strategy (2003) asserted that skills and business innovation drive productivity, and that employers should see engagement in training as benefiting their businesses. The confederation of British industry (CBI) and individual employers influenced and endorsed the skills strategy, but difficulties remain in getting beyond a general awareness of the need to engage in VET to genuine participation, especially in the expanding small and medium enterprise (SME) sector. The lack of active involvement bedevils attempts to clarify expectations and needs other than at a very superficial level.

The impact of the decline of the trades unions, and the removal of the levy, is still evident in the UK's response to the engagement of industry in skills development and VET. Government action has focused on making the business case for learning to employers, rather than on imposing industry's involvement in VET through legislation. In contrast to arrangements to develop occupational-sector ownership of skills development in other countries, the UK SSCs have few links to the Trades Unions.

Paying for Training

An underlying theme within this debate is the importance of employers' buy-in to learning in its literal sense in paying for VET. A fully publicly funded VET system addressing the needs of both entrants to the workforce and current employees would require significant shifts in emphasis in the distribution of public funds at a time when demands on other services, notably health, are also increasing.

The United Kingdom has a large further-education system, comprising over 500 further-education and sixth-form colleges and over 1000 work-based training providers. The colleges and providers in England draw most of their funding from a national learning and skills council (LSC) and similar arrangements are in place in the devolved administrations of Wales and Scotland, and in Northern Ireland. The LSC commissions a specified volume of education and training programs for which there is a national funding formula. Most of this funding is now for general education programs for young people and apprenticeships.

Public funding for adult learning is now almost entirely devoted to the Train to Gain program, which initially restricted public funding for workforce development to the achievement of a first full level-2 qualification for employees and any associated basic skills needs.

The rationale for the public subsidy up to level 2 derives from the view of the CBI that employers are less willing to train their employees in basic skills and competences that they expect the public-education system to have already developed. By providing publicly funded education and training for adults in the workforce up to level 2, the government expects to increase employers' contribution to the development of higher-level skills. As qualification strategies are developed for individual sectors, funding for Train to Gain is now extended to encompass further level 2 and level 3 program.

While the Train to Gain program aims to attract greater participation by employers in fee-for-service training, previous national funding formulae and numerous European Social Fund (ESF) and European Union (EU) initiatives enabled colleges and providers to offer training at no cost to employers. This created a dependency culture, which has reduced employers' propensity to pay for training and limits the market for fee-for-service work. Bespoke provision of this nature may well be more immediately relevant to employers' business needs than full qualifications, especially for SMEs, but many employers are reluctant to pay for this.

The Train to Gain program also provides skills brokers, who act as intermediaries to help employers to identify their skills needs and source the supply of skills training. However, their success in attracting employers who would not otherwise have taken part in training has been limited. There are also concerns that too much emphasis on brokerage may hinder the development of direct and long-term relationships between VET providers and employers. Brokers may be particularly useful in securing the involvement of hard-to-reach small firms, but this is a difficult market and does not lend itself to the achievement of short-term targets set for the brokerage service.

A demonstration of the onus placed on employers to drive up demand for skills is the skills pledge initiative, launched in 2007 as part of the Train to Gain program. Employers are exhorted to make a voluntary, public commitment by the leadership of a company or organization to support all its employees to develop their basic skills, including literacy and numeracy, and work toward relevant, valuable qualifications to at least level 2.

Despite the offer of free training, the brokerage service and incentives such as the pledge, take-up of the Train to Gain offer has been lower than anticipated. The emphasis on full qualifications, often at an inappropriate level, is thought to be a major reason for this. Flexibilities introduced in 2008, plus the economic downturn, have strengthened employer take-up of Train to Gain.

Reforming the Supply of Skills

The UK government's voluntarist approach to industry involvement in VET is in stark contrast to the numerous quality initiatives and funding measures used to urge

providers to be more responsive to the demands of industry. One of these introduced the requirement for 3-year development plans for colleges and providers, which set targets to improve in a number of key activities, including employer engagement.

A review of the employer-engagement targets revealed a diverse and uncoordinated activity. A broad range of views on what constitutes employer engagement was evident, suggesting the lack of a shared definition. The review proposed a typology for employer engagement, based on a number of key purposes for engaging employers in post-16 learning, which may help to clarify the part played by employers and ensure that their roles are appropriate in different circumstances. The typology and how it relates to targets set by providers and more general applications is shown in **Table 1**.

Since then, efforts to reform the supply of education and training post-16 have continued. LSC is to introduce measures of success that describe key dimensions of colleges' or providers' performance, which will include responsiveness to employers. A standard for employer engagement and vocational excellence the Training Quality Standard (TQS) is also now available, which will provide accreditation for providers' responsiveness to employers.

Alongside these activities, a major program of staff and management training to develop more responsive approaches to employer engagement is now in place, led by the learning and skills improvement service (LSIS). While this has been well received, it is evident that there is a substantial culture shift required to move VET providers' focus from the needs of individual learners to the needs of employers.

The combination of funding measures, quality requirements, and quality improvement may bring about the changes required to secure a VET system that is in tune with industry's requirements, but this is a long-term proposition. There are also tensions in reconciling colleges' and training providers' mission to develop individuals and communities, for which government and funding bodies have also set targets, while serving the needs of industry.

Relevance of the VET Curriculum

The education and training sector is often seen by employers to be remote from the realities of the world of work and unable to keep abreast of changing needs. Another tactic to increase industry's involvement in VET has been to develop more specialist provision, geared up to meet industry's specific needs. In Australia, this is demonstrated in the new technical colleges. In England, the centers of vocational excellence (CoVEs) program set out to improve radically the capacity of the further education (FE) sector to deliver high-quality learning in specific vocational contexts by developing specialist resources and staff with strong links to employers. There are now over 400 CoVEs,

Table 1 A typology for employer engagement

<i>Type of engagement</i>	<i>Examples of targets set</i>	<i>More general examples</i>
<i>Employers as stakeholders</i>	<ul style="list-style-type: none"> ● Increase by 10% students taking part in industrial/business visits 	<ul style="list-style-type: none"> ● Providing work experience places
Providing leadership through involvement in the design, development, management, delivery, and assessment of learning	<ul style="list-style-type: none"> ● Benefiting from speakers from the world of industry taking part in work experience ● Increase numbers of employers involved in advisory boards from 50 to 80 ● Increase employer involvement in assignments from 10 to 20 ● All vocational sectors to have employer advisory panels ● Increase the number of employers involved in the design of learning programs within the college's Center of Vocational Excellence (CoVE) 	<ul style="list-style-type: none"> ● Acting as a visiting speaker ● Advising on curriculum and its assessment ● Participating in college governance
<i>Employers as consumers</i>	<ul style="list-style-type: none"> ● Increase the number of sponsoring employers by 12% 	<ul style="list-style-type: none"> ● Using training needs analysis (TNA) services to identify workforce-development needs
Purchasing diagnostic services and skills development from LSC-funded providers	<ul style="list-style-type: none"> ● Increased income targets, increasing the range of employer contacts and networking ● Customized provision through an enterprise unit full cost increased by 5% ● Over the 3 years of the plan (name of college) will work with at least 200 SMEs, and the college will increase its revenue framework with employers by 10% 	<ul style="list-style-type: none"> ● Taking advantage of information, advice, and guidance (IAG) services to source training provision ● Using customized day release or regular provision as part of company training ● Buying bespoke training for updating ● Using WBL provision (e.g., national vocational qualification (NVQ) training and assessment) ● Using national qualifications to select new staff and upskill current staff
<i>Employers as strategic partners</i>	<ul style="list-style-type: none"> ● Expand links with providers and develop an employers' forum through links with trade associations to provide the college with LMI on the needs of employers 	<ul style="list-style-type: none"> ● Using providers as sources of support for business development
Sustained interaction between employers and the planners and providers of learning	<ul style="list-style-type: none"> ● To establish a structured workforce-development agreement with at least 25 employers in the priority sectors of manufacturing, care, retail, and construction ● Set up in partnership with an employer a dedicated training facility for industrial updating/training courses 	<ul style="list-style-type: none"> ● Collaborating with planners in developing new provision for the benefit of own company and wider sector ● Contributing in cash and kind to new or updated resources for learning-joint ventures and sharing training facilities ● Sharing or subsidizing specialist staff

Adapted from Hughes, M. (2004). *Reviewing the Impact of Employer Engagement Targets*. London: LSDA.

but the extent to which they have developed stronger links with industry varied enormously. An evaluation of the CoVE program concluded that:

Good progress has been made in the identification of skills priorities and improvements in arrangements to consult with employers in the planning of provision, particularly in engineering, construction and care. There is clear evidence of a positive change in employers' attitudes towards training and their willingness to become more closely involved in its delivery, most notably in CoVEs in engineering, care, and hospitality and catering. However, many employers remain reluctant to meet the costs of training. Most CoVEs in care, construction, and engineering have developed flexible modes of delivery to meet employers'

needs. This is less evident in CoVEs in ICT and in hospitality and catering.

Also that:

Most CoVEs have invested well in developing industry-standard learning environments. Construction and engineering CoVEs in particular have benefited from employers' donations of specialist equipment. Employer sponsorship in hospitality and catering CoVEs has been more limited.

CoVEs received premium funding for the first 3 years of operation, which enabled them to invest in equipment and staff development. However, few CoVEs have put in place plans to replace this funding from other sources and the

reluctance of employers to pay for training will severely restrain their ability to maintain their excellent provision.

More recently, the government proposed the development of a new network of national skills academies (NSAs). These are intended to be employer led, forming a strong network in each sector linking CoVEs with universities, training providers, and specialist schools. The aim is to establish 12 NSAs by 2007/2008 and over time to support one in every major sector.

The first of the NSAs to be established – the Fashion Retail Academy – is supported by major retailers including The Arcadia Group Ltd, which includes the popular high-street brands such as Next, and Marks & Spencer. The academy combines training in retail business and fashion education to introduce the practical and vocational skills needed for the world of fashion retail. Students have access to state-of-the-art technology at the resources of the retail partners' organizations. Teaching is provided by both college and industry staff. Students are given work placements and industry master classes, delivered by key people from the fashion retail industry.

The curriculum has been specifically developed for the Fashion Retail Academy and is fully accredited by nationally recognized qualifications. Students are encouraged to retain or take up paid part-time employment in fashion-retail companies, to expand their skills and put the theory taught at the academy into practice.

The establishment of more academies is proving to be difficult, mainly in relation to securing the required level of funding from employers, and in sustaining the long-term commitment to skills development that the academies require.

Conclusions and Issues

VET systems are facing a huge challenge in re-orientating their mission, purpose, and ways of working to meet the requirements of employer responsiveness. Greater clarity in articulating the role and purpose of employer engagement would support a more coherent approach to target setting and make better use of employers as a strategic resource for learning.

A major policy goal for many governments in reforming their VET systems is to make these systems more employer facing. In the United Kingdom, the current preoccupation is to ensure VET providers are demanded, that is, responsive to employers' requirements rather than continuing to provide their established inventory of courses. The quest to reposition employers as the principal players in skills systems rests on three elements.

The first of these is the creation of employer-led sector bodies responsible for shaping training strategies for their occupational areas. The strongest of such systems can be found in the Netherlands where sector bodies

(knowledge centers) are at the heart of the VET framework, with responsibility for occupational-competency standards, accreditation of work-based trainers, quality control of examinations, and approval of company training schemes.

Very few countries have achieved such a powerful sector-based framework as the Netherlands, though many have sector-skills bodies, which exercise considerable influence, for example, the Australian industry skills councils and their training packages, which create the framework for the recognition and assessment of job skills. In Canada, the tripartite sector councils have led the development of a culture of learning across their sectors. But often, and certainly in the United Kingdom, sector bodies have faced real challenges in representing the views and needs of very diverse sectors, and of SMEs in particular. As such, they are unlikely to serve as alternatives to direct engagement between providers and businesses at local level. If they are to achieve credibility, sector-skills bodies need appropriate levels of resource. Canadian sector councils are fully funded by government, which is unusual, but these were originally intended to become self-financing. In South Africa, the 27 sector education and training authorities are financed by a 1% of wage-bill levy across the private sector.

When such clear revenue streams do not exist, there is a danger that the sector-skills organizations will lack authority and dissipate their efforts in chasing money, projects, and short-term initiatives. This undermines their ability to achieve clarity of role and take a strong leadership position on strategic cross-sector-skills issues. Seeking further funding from employers may cloud the issue and be perceived as paying twice for the same service.

The second mechanism for employers to gain real purchase on the skills system is their relationship with providers and the qualifications system. In the United Kingdom, the Leitch review has proposed that the SSCs be given the power to approve vocational qualifications for eligibility for public funding. This approach has been challenged by the education and skills select committee of the House of Commons as being potentially bureaucratic and less responsive to employer needs. There is also evidence from Finland and Sweden that highly valued VET systems require well-articulated, flexible, progressive qualification routes, which almost certainly means a multiple-stakeholder route. Too much emphasis on employers securing a consensus from employers on their needs and how they should be met beyond a superficial level is difficult for a number of reasons:

- Employers are not a homogeneous group—even within the same occupational sector. The size, success, and maturity of the company and its market position may be a stronger influence on skills needs than the specific demands of the sector. Establishing long-term groupings based on these characteristics would be very difficult, given their dynamic nature.

- There are difficulties in distinguishing between employers' wants and needs, and in establishing long-term priorities for development that will benefit the economy and society in general as well as individual organizations.

Reliance on employers strategically to drive the supply of skills may then be dangerous, producing short-term, instrumental solutions that do not add up to an effective vocational-qualifications framework.

Finally, employers can be central to the VET system through their role in raising demand for skills, usually via financial-incentive schemes. In France, there is a clear train-or-pay (the levy) culture resulting in a strong continuous professional-development program funded by levy and backed up by a statutory worker entitlement to learning. Interestingly, despite this legislative imposition on industry, there is a high level of consensus in France about this investment in skills updating. The Netherlands has a powerful fiscal incentive scheme that provides firms with tax relief on trainees' wages. The key to the effectiveness of these measures is strong linkage through the entire VET system to achieve supply equilibrium with the incentivized demand. There is also a danger that highly centralized schemes cannot cope with internally complex sectors producing homogeneous solutions to a heterogeneity of demand.

See also: Apprenticeships; Dual System; Qualifications Frameworks and their Role in the Reform of Education and Training.

Further Reading

- Aitken, L., Burgess, C., and Hughes, M. (2006). *Sustaining CoVE Development*. London: LSDA.
- CBI (2007). *Shaping up for the Future: The Business Vision for Education and Skills*. London: CBI.
- DfES (1988). *Towards a National Skills Agenda: First Report of the National Skills Task Force*. London: HMSO.
- DfES (1999). *Learning to Succeed: A New Framework for Post 16 Learning*. Cm 4392. London: HMSO.
- DfES (2001). Modern apprenticeships: The way to work. *The Report of the Modern Apprenticeship Advisory Committee*. London: DfES.
- DfES (2002). *Success for All: Reforming Further Education and Training*. London: HMSO.
- DfES (2003). *21st Century Skills – Realising Our Potential: Individual, Employers, Nation*. Cm5810. London: HMSO.

- DfES (2005a). *Realising the Potential: A Review of the Future Role of Further Education Colleges*. London: The Stationery Office.
- DfES (2005b). *Skills: Getting on in Business, Getting on at Work*. Cm6483. London: Department for Education and Skills.
- Doyle, L. and Hughes, M. (2004). *Learning without Lessons*. London: LSDA.
- Fuller, A. and Unwin, L. (2004). Does apprenticeship still have meaning in the UK? The consequences of voluntarism and sectoral change. In Hayward, G. and James, S. (eds.) *Balancing the Skills Equation Key Issues and Challenges for Policy and Practice*, pp 101–115. Bristol: The Policy Press.
- Hillage, J., Loukas, G., Newton, B., and Tomkin, P. (2006). Employer training pilots: Final evaluation report. *DfES Research Reports RR 774*. London: DfES.
- HM Treasury (2006). *Leitch Review of Skills: Prosperity for All in the Global Economy – World Class Skills*. London: HMSO.
- House of Commons Education and Skills Committee (2007). *Post 16 Skills*. London: The Stationery Office.
- Hughes, M. (2003). *Successful Engagement: Guidance for Colleges and Providers on Effective Engagement in Post 16 Learning*. London: LSDA.
- Hughes, M. (2004). *Reviewing the Impact of Employer Engagement Targets*. London: LSDA.
- Hughes, M. and Mager, C. (2000). *The Skills Agenda: Issues for Post-16 Providers*. London: LSDA.
- Hughes, M. and Smeaton, B. (2006). *Exploring Employers' Perceptions of Employer Engagement*. London: LSDA.
- Macleod, D. and Hughes, M. (2006a). *Apprenticeships: A Review of Recent Policy and Practice*. London: LSDA.
- Macleod, D. and Hughes, M. (2006b). *What We Know about Working with Employers: A Synthesis of LSDA Work on Employer Engagement*. London: LSDA.
- Manpower Services Commission (1981). *A New Training Initiative: An Agenda for Action*. Sheffield: HMSO.
- Mukherjee, S. (1974). *There's Work to Be Done: Unemployment and Manpower Policies*. Sheffield: Manpower Services Commission, HMSO.
- OECD (2005). *Promoting Adult Learning*. Paris: OECD.
- Ofsted and Adult Learning Inspectorate (ALI) (2005). *Centres of Vocational Excellence in Practice: A Survey Report on 40 CoVEs in Five Skills Sectors*. Coventry: ALI.
- Stasz, C. and Wright, S. (2004). *Emerging Policy for Vocational Learning in England: Will It Lead to a Better System?* London: Learning and Skills Research Centre.
- University of Leicester (2006). *Skills abroad: A comparative assessment of international policy approaches to skills*. *SSDA Research Report 16*. Yorkshire: Sector Skills Development Agency.

Relevant Websites

- <http://www.lsc.gov.uk> – The Learning and Skills Council.
- <http://www.trainingqualitystandard.co.uk> – The New Standard for Employer Responsiveness and Vocational Excellence.
- <http://www.qca.org.uk> – The Qualifications and Curriculum Authority.
- <http://www.isis.org.uk> – The Quality Improvement Agency.
- <http://www.traintogain.gov.uk> – The Train to Gain Programme.

Non-Completion in Vocational Education and Training: Issues, Patterns and Possibilities

S Lamb and A Walstab, University of Melbourne, Melbourne, VIC, Australia

© 2010 Elsevier Ltd. All rights reserved.

Research into noncompletion in some sectors of education, such as in schools and higher education, is well developed. However, comparable research into noncompletion in vocational education and training (VET) is more difficult to find and certainly less systematic. Despite this, in many countries, there has been a growing interest in the monitoring of levels of noncompletion and investigation of the causes and consequences. Reducing noncompletion rates is an important goal for both educators and policymakers. Yet, there is a good deal of controversy about how best to measure completion and noncompletion rates. This is partly due to the fact that completion and noncompletion have become less easy to define. One reason for this is that patterns of education and work have become more varied. Some people are taking more time to study and spending varied periods of time in and out of education and training courses. These patterns make it more difficult to establish cut-off points for estimating noncompletion. Another issue is related to intentions. Some participants enrol in courses but only intend to complete part of the course, such as a unit or module. While such participants are course or qualification non-completers, they are completers in terms of their original intentions.

In addition to these matters, noncompletion in VET does not always equate with failure. Noncompletion may signify the achievement of desired goals, either in the sense that skills have been gained, employment outcomes realized, or articulation to further or higher studies successfully negotiated. Given that many students return to study fairly soon after withdrawing from a course and that a substantial number return some time later, the notion of noncompletion from a lifelong learning perspective is less meaningful than it once was. However, for the most part, noncompletion remains a serious problem, especially for students from disadvantaged circumstances and for students in particular fields of study and types of courses.

The main aim of this article is to examine the issue of noncompletion in VET, to look at how it is defined, at some of the different ways it is measured, and to provide some estimates of noncompletion using recent data. In addition, it also examines some of the reasons for non-completion and, following this, presents a brief outline of the sorts of strategies employed to reduce noncompletion. A brief list of further readings of key publications on the topic is also provided.

Defining and Measuring Noncompletion

Internationally, the drive for accountability now makes for stronger emphasis on completion data as a key performance measure in VET. Completion and noncompletion, though, are not straightforward to define or measure. In an environment where lifelong learning is promoted and encouraged, with increasing emphasis on course transfers, recognition of prior learning, and articulation between courses and sectors, traditional measurements and conventional indicators may be inadequate. Added to these factors is the issue that VET provision itself is becoming increasingly complex. There is a broader range of users, increasingly diverse across different age groups and educational backgrounds, a broader range of courses and programs and more varied arrangements for delivery.

International comparisons of VET noncompletion are made more complex because of some of these issues. Some of the difficulties are related to institutional and organizational differences: there are variations between countries in how qualification structures are organized, the range of VET offerings, and the modes of delivery. There are also differences in levels of access to vocational programs, as well as in local labor-market contexts and in the demand for skills.

Other difficulties relate to how noncompletion is measured. The Organization for Economic Cooperation and Development (OECD), for example, publishes survival rates for tertiary type-B programs. Tertiary type-B programs are classified at the International Standard of Classification of Education (ISCED) level 5, the first stage of tertiary education, and focus on vocational and occupation skills for direct entry to the workforce. While the programs have a minimum duration of 2 years full-time enrolment, in some countries they can extend to 5 years of full-time study. They are, however, typically shorter programs than those of tertiary-type A, which are largely theoretical and generally lead to a traditional university-degree qualification (OECD, 2007).

The survival rates in tertiary-type B education for 2004, published for 17 OECD countries, are given in **Table 1**. The completion rates represent the number of graduates in 2004 divided by the number of new entrants n years before, where n is the number of years required to complete the course. While on average two-thirds of students successfully complete tertiary-type B qualifications,

Table 1 Survival rates in tertiary type-B education, 2004

OECD countries	Percent
Belgium (Fl.)	85
Czech Republic	61
Germany	79
Greece	35
Hungary	48
Iceland	76
Ireland	69
Japan	87
Mexico	63
New Zealand	42
Poland	74
Portugal	58
Slovak Republic	77
Spain	79
Sweden	68
Turkey	79
United Kingdom	53
OECD average	67

From OECD (2007). *Education at a Glance 2006*. Paris: OECD, Table A3.6

this figure conceals considerable variation between countries. Rates range from 35% in Greece and 42% in New Zealand to 87% in Japan and 85% in Belgium.

This method for calculating course-completion rates – taking the ratio of the number of students at the point of entry to exit over the length of the course – commonly used in the higher education and schools sectors, has a number of limitations in the VET context. By assuming constant student flows, these survival rates do not account for those who do not complete within the normal time but are still eligible to complete, and fully intend to do so, which includes students enrolled part-time and those taking a sanctioned break from study. Nor do they account for students who have transferred to another course at or above the same qualification level. The growing flexibility in the delivery of VET, whether this be online or in the workplace, along with the modularization of the curriculum, mean that this methodology increasingly excludes a larger proportion of learners in VET. Variations in any of these factors will impact on the reported rate of completion.

Countries with centrally collated administrative data collections are well placed to measure and report on non-completion in vocational programs. England, for example, uses such a collection for the annual reporting of benchmarking data. These data, based on individual student record files and published by the learning skills council, include measures of success, retention, and achievement. Success rates are calculated as the number of learning aims achieved, divided by the number of commencing students, excluding those who have transferred to another course within the same institution (LSC, 2007). The overall success rate in further education in the 2005–06 academic year was 76%. Retention rates on the other hand are calculated by

dividing the number of completed qualifications by the number commencing, excluding those who transferred. The figure for 2005–06 was 87%. The achievement rate represents the ratio of the number of students who achieved qualifications to the number that completed qualifications (88% in 2005–06). The success rate is the product of achievement and retention rates.

Where there is no centralized collection of VET administrative data, national reporting of noncompletion must rely on other sources. In the United States, the National Centre for Educational Statistics draws on its cohort-based longitudinal studies, based on national samples of students who are tracked over time to map pathways and outcomes, including measures of course completion. The 2004–06 beginning postsecondary students longitudinal study (BPS) tracks a nationally representative sample of students who began postsecondary education for the first time in 2003–04, surveying in 2004, and again in 2006 (NCES, 2007).

Longitudinal studies can allow for more flexibility in the measuring and reporting of completion rates, with the ability to take into account student movement across institutions and courses. Continuing students taking longer than usual to complete can also be identified. Findings from the BPS follow-up survey in 2006 report these figures as degree attainment and persistence for students enrolled or with degree plans in vocational certificate programs (which usually take less than 2 years to complete) and associate's degree programs (2–3 years duration) in public 2-year institutions (usually community colleges). For those with certificate degree plans in 2003–04, 31.8% had completed certificate or associate's degree, 21.2% were still continuing, and 47% had discontinued without completing their qualification. Those with initial associate-degree plans had a completion rate of 16.1% (attainment of a certificate or associate degree), a continuation rate of 39%, and a noncompletion rate of 44.9% (NCES, 2007:11).

Levels of Noncompletion Among Different Groups

Aggregate completion rates mask wide variations in completion and noncompletion for different groups. VET has a diverse clientele; reporting completion rates for different groups helps providers to target programs to where there is the most need. Examples of noncompletion and completion rates for different groups by selected course and student characteristics are given in **Table 2**.

Partial Completion

Traditional measures of VET course completion, like the ones discussed above, are based on the assumption that all students enrolling in the course intend to complete the full qualification. Notions of lifelong learning, on the

Table 2 Completion rates by course and student characteristics

<i>Course characteristics</i>	
Length of course	England (LSC 2007): success rates for short courses (up to 24 weeks, usually 12 weeks) in 2005–06 were 84.6% and 69.9% for long courses, thus rates decreasing with increased length of course.
Level of course	Australia (Shah & Burke 2003): withdrawal rates tend to increase with Australian Qualification Framework (AQF) level of the course. The withdrawal rate for Certificate I (lowest level) courses was 25.9% and 27.4% for Certificate II courses. This increases to 36.7% for diploma level and 35.1% for advanced diploma. These figures exclude partial completers.
Field of study	Ireland (Kinsella and Roe, 2005): percentage completion rates by ISCED area of study show large variations across course types, ranging from 60.3% for engineering, manufacturing, and construction to 87.4% for agricultural and veterinary (certificate and diploma courses reported together).
<i>Student characteristics</i>	
Age	United States (NCES 2007): tables reporting persistence show that the proportion of students who discontinued from all education and training in 2006 without obtaining a certificate or associate's degree increased with age of enrolment. For example, 33.9% of those aged 18 years or younger on enrolment had withdrawn compared to 39.3% of those aged 19 years. This increased to 53.7% of those aged 20–23 years and 55.3% aged 24–29 years, and 57.5% of students aged 30 or older.
Gender	New Zealand (Scott 2005): 35% of females who enrolled in certificate courses in 1998, and 34% of females in diplomas had successfully completed by 2002 compared to 26% of males enrolled in certificate courses and 28% of males in diplomas.
Ethnicity	England (LSC 2007): success rates reported for further education institutions by ethnicity showed a range from 69% for students from a Black Caribbean and other background to 76% for Indian students and 77% for White students.

From LSC (Learning Skills Council) (2007). *Further Education and Work-Based Learning for Young People – Learner Outcomes in England 2005/06*. Coventry: Learning Skills Council; Shah, C. and Burke, G. (2003). Completion and partial completion in courses in TAFE, Australia. *Working Paper No. 51*. Melbourne: Centre for the Economics of Education and Training, Monash University; Kinsella, E. and Roe, J. (2006). *Completion Rates for Students Undertaking Full-Time Programmes of Study in Institutes of Technology, A Study Carried out for the Council of Directors of Institutes of Technology and The Dublin Institute of Technology*. Dublin: CIRCA Group Europe; NCES (National Centre for Education Statistics) (2007). *Persistence and Attainment of 2003–04 Beginning Postsecondary Students: After Three Years*. Washington, DC: NCES; Scott, D. (2005). Retention, completion and progression in tertiary education in New Zealand. *Journal of Higher Education Policy and Management* 27(1), 3–17.

other hand, emphasize the acquisition of skills and qualifications over the longer term, often in smaller, modular components. Indeed, the flexibility of VET systems encourages self-paced learning. In Australia, for example, researchers have estimated that nearly half (49%) of all technical and further education (TAFE) student enrolments result in partial completion of a course (Foyster *et al.*, 2000). This study of student flows through TAFE institutes in the mid-1990s defines partial completion as completing one or more modules within a course, not failing any other modules and having exited the system (Foyster *et al.*, xii). The completion rate (completion of a full course) was estimated to be 24%, which was lower than the noncompletion rate (representing academic failure) of 27%. Partial completion in this context is viewed by the researchers as a positive outcome. However, as Davies (2001) argues, without data on student intention, it is not possible to isolate partial completers who intended to complete the full course from those wanting to complete modules only.

While administrative data collections in the United Kingdom and Canada record student intentions and plans on enrolment (in the form of, e.g., qualification aims), this is not common practice (Lamb, 2005). In a discussion regarding the value of this measure, Lamb (2005) points to methodological implications, including the need to collect this data on enrolment rather than on exit, although it needs to be acknowledged that motivations and intentions are subjective and can change over time.

Information regarding intent may not routinely be collected in many countries; however, it is often possible to compare the characteristics of those who complete a full qualification (graduates) against those who complete only part of a course before exiting the VET system (partial completers). In Australia, such comparisons reveal that the age profile of graduates is younger than for partial completers, with 40.4% of graduates aged between 15 and 24 years, compared to 22.9% of partial completers. Conversely, 22.9% of graduates are aged 45 years or over, compared to 36.7% of partial completers. Partial completers are more likely to come to VET with higher levels of educational attainment, with 26.5% having completed a Certificate IV or higher prior to enrolment, compared to 22.9% of graduates. This could suggest that partial completers are building on qualifications they have already achieved with specific skills and competencies, and therefore pursuing a full course qualification is less important than for the younger graduates who are more likely to be building on school-level study.

Reasons for Noncompletion

Understanding why some learners tend not to complete their study is the key to developing policies or interventions

to help improve chances of completion. Yet, identifying the causes of noncompletion is a difficult task because the process is influenced by an assortment of factors related to individual students, their circumstances, backgrounds and motives for study, and to the college and community settings in which they are located.

Some of the complexity is illustrated by the variety of reasons that learners report for not completing. In a study of 2180 learners in the United Kingdom who withdrew from further education and work-based learning courses without completing, course-related reasons for withdrawal were mentioned by 32%, changes in circumstances were mentioned by 28%, and work-related reasons were mentioned by 20% (Simm *et al.*, 2007). The most specific reasons were that the course was not what they wanted to do or they changed their mind about the course (14%), poor quality of teaching (11%), the course not being at the right level (11%), and work commitments (8%).

But these reasons do not necessarily get at the underlying causes behind why learners withdraw from study, particularly the causes that have shaped from earlier-in-life attitudes, behaviors, and experiences of study and learning. Moreover, as many factors may contribute, it is difficult to identify the various factors and the interconnected nature of their impact on the decision to withdraw from study. The approach of some researchers to this issue has been to develop theories and conceptual models of noncompletion using a variety of qualitative and quantitative research methods, most often elaborated by researchers in the United States. Spady (1971), for example, proposed a model mapping the influence of five predictors (achievement, academic integration, course choice, socioeconomic status (SES), and gender) relayed through a fifth variable (social integration) on decisions by individuals about remaining in study. Tinto (1975), building on this work, developed a more elaborate model, but maintained a focus on the importance of student integration into the social and academic life of college and the sets of factors that shaped these aspects of student engagement. Other models have also been developed, varying in their emphasis, in the different sets of factors, and the importance placed on different precursors (see, e.g., Chen and Thomas, 2001; Sandler, 2000; Bean and Metzner, 1985). Basically, the models treat noncompletion as a result of the student's interaction and engagement with the institution (as a learner and a social participant), with engagement mediated by a wide range of background or input factors such as social, educational, and economic background.

Drawing on this work, it is possible to identify a couple of broad perspectives for understanding the reasons for noncompletion. The first is what could be described as sets of barriers associated with institutions themselves and the way they operate and are organized as places of learning, including quality of teaching, support for

learners, flexibility, and monitoring. The second is the range of factors associated with what individuals bring with them to the place of learning, including their family context, their intentions, previous educational experiences and attainment, work situation, perceptions, dispositions, interests, and motivations. Both perspectives reflect on the way contexts of learning are constructed and operate.

Research on institutional barriers focuses on the way individual's experiences are shaped by institutional settings. Empirical research on noncompletion has identified a number of factors associated with VET providers that influence the likelihood of completion. For brevity, only a few are mentioned here. One is the quality of teaching. Research in several countries has demonstrated a relationship between quality of teaching and attrition rates. In Australia, for example, a study of VET noncompleters who had attended TAFE colleges found that students rated poor quality of the teaching staff and teachers not having relevant industry experience as key reasons for not continuing their programs (Callan, 2005). Similarly, Polesel *et al.* (2004) reported that students dissatisfied with the quality of teaching were far more likely to leave without completion than students who were satisfied. In the United Kingdom, several studies have found that satisfaction with teachers and teacher support correlates with perseverance and completion in ways which suggest that for many, the quality of teaching and support are very important (Davies, 1999). Studies in Canada and the United States have also identified quality of teaching as a predictor of noncompletion (see, e.g., Mueller, 2007).

Factors associated with quality of academic and social interaction are also important influences on student experiences and decisions. These include features such as student-staff contact, student interaction, active learning, prompt feedback, time on task, high expectations, respect for diverse talents and ways of learning, and inclusive, and affirming environments where expectations for performance are clearly communicated (Wetzel *et al.*, 1999; Simm *et al.*, 2007).

Research focusing more on the backgrounds and attributes of students – such as their values, attitudes, behaviors, family circumstances, work, educational history – has identified a range of factors which influence decisions to complete study or not. Important among these are negative attitudes toward formal study often associated with low educational attainment. Leaving school without a qualification, now more prevalent in many countries among older age groups, can often follow low school achievement or academic failure and promotes a lack of confidence and low self-esteem in relation to study and training. These experiences can predispose learners to noncompletion, particularly if courses are not quite as learners expect, or they are difficult, or formal study renews feelings of uncertainty and undermines confidence.

Financial considerations are often reported among factors that lead to noncompletion. The impact of costs may be felt unevenly across different age and social groups. Although lack of income support is often cited as a deterrent to continued participation in study, it tends to be associated with other barriers associated with the structure of provision and student dispositions toward education and training. Dispositional factors are a powerful deterrent to participating in education and training among people with low levels of attainment, and together with other obstacles, such as cost and place, compound the barriers to participation.

Linked with costs are issues with work and family responsibilities, and childcare. This can be a major issue for females and for particular age groups (such as 25–39-year olds). Accompanying these issues is the problem of time constraints as well as other factors including location and distance from place of study.

It is very important to recognize that not all noncompletion is negative or needs to be addressed. Students can leave courses of study for a variety of positive reasons such as changes in their personal circumstances and the attraction of alternative employment and training opportunities. Some learners may never have intended to complete all of the components leading to a certificate, or they obtain from the parts of the course they undertake all that they want without fulfilling the requirements for formal completion. Intentions, plans, and circumstances may change during a course of study. Noncompletion, therefore, is not necessarily associated with failure or negative outcomes, since many learners can achieve what they want from their study even though they may not complete it in the formal sense.

Strategies to Address Noncompletion

The broad perspectives underpinning research on reasons for noncompletion in VET suggest that there are different sorts of strategies needed to improve outcomes. Learners who tend not to complete their courses of study have diverse characteristics and contexts which call for initiatives that can respond to their varied circumstances and needs. As there are interrelated and multiple causes associated with noncompletion, successful interventions need to involve comprehensive, multiservice approaches to providing support.

In the body of research on the effects of the backgrounds and attributes of students, initiatives focus on the physical, material, and dispositional barriers to learning among students, particularly those with low educational attainment and for those from low SES backgrounds.

In terms of financial constraints, various schemes have been developed. Individual learning accounts, student loan schemes, vouchers, reductions in fees for the poor,

paid education leave, and systems of deferred fee payment have been implemented in different countries as mechanisms to support student learning in vocational education (Haukka *et al.*, 2004). Income support schemes are another strategy. In Australia, for example, youth allowance (YA) was introduced to encourage young people from poor families to participate in education or training if they lack the skills to find full-time employment. It is a means-tested income support scheme providing participants with funds while they learn. Evaluations suggest that it has been influential in encouraging greater participation and persistence (Lamb and Robinson, 2001; DFaCS, 2002). Similar schemes exist for different target groups, such as single parents. An example is the jobs, education, and training scheme in Australia which is designed to assist single parents meet the costs of education and training.

Structural or systemic changes are also important in supporting learners more likely not to complete. The curriculum content and institutional delivery mechanisms of education and training can be significant barriers. Negative attitudes to education and training are often the product of an individual's early experience of failure in formal education or training. One way of overcoming such dispositional barriers is to ensure that the content and structure of education and training provision targeted to reluctant participants is different from what they have experienced in the past.

Summaries of initiatives proposed to address noncompletion in VET, based on the broad individual and institutional perspectives on influential factors, suggest the following types of strategies:

- improving and extending advice and guidance services;
- paying particular attention to the early stages of programs of learning (student induction, initial assessment);
- closer monitoring and follow-up of poor attendance;
- early identification of underperforming students or students who are at risk;
- the early diagnosis of student requirements for basic skills and additional learning support and the provision of such support as far as possible within student learning programs;
- the development of a curriculum framework which is appropriate for intended students;
- target setting combined with formative assessment and feedback; and
- improvements to teaching (Davies, 2006; Martinez, 2001; McGivney, 2003).

The focus of strategies and policy effort should be on creating the conditions for effective learning and personal growth that underpin quality education and training programs. This calls for a combination of measures responsive to differences in individual learning needs, on the one

hand, and to the creation of quality programs and learning environments, on the other.

Bibliography

- Bean, J. and Metzner, B. (1985). A conceptual model of non-traditional undergraduate students attrition. *Review of Educational Research* 55(4), 485–540.
- Callan, V. (2005). *Why Students Leave? Leaving Vocational Education and Training with No Recorded Achievement*. Adelaide, NCVER.
- Chen, S. and Thomas, H. (2001). Constructing vocational and technical college student persistence models. *Journal of Vocational Education Research* 26(1), 26–55.
- Davies, M. (2001). *Completion, Non-Completion and Re-Orientation*. Melbourne: Educational Outcomes Research Unit, University of Melbourne.
- Davies, P. (1999). *Learning and Earning*. London: Learning and Skills Development Agency.
- Davies, P. (2006). *Learner Retention and Achievement: Summary of Research Findings*. London: Learning and Skills Network.
- DFaCS (Department of Family and Community Services) (2002). *Youth Allowance Evaluation: Final Report*. Canberra, Commonwealth of Australia.
- Foyster, J., Hon, K. F., and Shah, C. (2000). *Student Flows through Australian TAFE Courses*. NCVER: Adelaide.
- Haukka, S., Keating, J., and Lamb, S. (2004). *Alternative Mechanisms to Encourage Individual Contributions to VET*. Adelaide, NCVER.
- Kinsella, E. and Roe, J. (2006). *Completion Rates for Students Undertaking Full-Time Programmes of Study in Institutes of Technology, A Study Carried out for the Council of Directors of Institutes of Technology and The Dublin Institute of Technology*. Dublin: CIRCA Group Europe.
- Lamb, S. (2005). *Technical Report: Survey Methodological Options for a) Non-Completers of VET and b) Apprentices and Trainees over Time*. Adelaide, NCVER.
- Lamb, S. and Robinson, L. (2001). *Youth Allowance and Participation in Education, Training and Work: Results from the Second Wave of the Youth Allowance Longitudinal Survey*. Canberra: Department of Family and Community Services.
- LSC (Learning Skills Council) (2007). *Further Education and Work-Based Learning for Young People – Learner Outcomes in England 2005/06*. Coventry, Learning Skills Council.
- Martinez, P. (2001). *Improving Student Retention and Achievement: What Do We Know and What Do We Need to Find Out?* London: Learning and Skills Development Agency.
- McGivney, V. (2003). *Staying or Leaving the Course: Non-Completion and Retention of Mature Students in Further and Higher Education*. Coventry, NIACE.
- Mueller, R. (2007). *Access and Persistence of Students from Low-Income Backgrounds in Canadian Post-Secondary Education: A Review of the Literature*. Montreal: University of Canadian Millennium Scholarship Foundation.
- NCES (National Centre for Education Statistics) (2007). *Persistence and Attainment of 2003–04 Beginning Postsecondary Students: After Three Years*. Washington, DC, NCES.
- OECD (2007). *Education at a Glance 2006*. Paris: OECD.
- Polesel, J., Davies, M., and Teese, R. (2004). *Course Completion and Instructional Experience in TAFE*. Adelaide, NCVER.
- Sandler, M. (2000). Career decision-making self-efficacy, perceived stress, and an integrated model of student persistence: A structural model of finances, attitudes, behavior, and career development. *Research in Higher Education* 41(5), 537–580.
- Scott, D. (2005). Retention, completion and progression in tertiary education in New Zealand. *Journal of Higher Education Policy and Management* 27(1), 3–17.
- Shah, C. and Burke, G. (2003). Completion and partial completion in courses in TAFE, Australia. *Working Paper No. 51*. Melbourne: Centre for the Economics of Education and Training, Monash University.
- Simm, C., Page, R., and Miller, L. (2007). Reasons for early leaving from FE and work-based learning courses. *Research Report RR849*. Sussex: Institute for Employment Studies, University of Sussex.
- Spady, W. G. (1971). Dropouts from higher education: Toward an empirical model. *Interchange* 2(3), 38–62.
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research* 45(1), 89–125.
- Wetzel, J. N., O'Toole, D., and Peterson, S. (1999). Factors affecting student retention probabilities: A case study. *Journal of Economics and Finance* 23(1), 45–55.

Further Reading

- Martinez, P. and Munday, F. (1998). *9,000 Voices: Student Persistence and Drop-Out in Further Education*. London: Further Education Development Agency.

Planning and Policy Development for Technical Vocational Education and Training Systems

R Maclean, Hong Kong Institute of Education, Hong Kong, People's Republic of China

M Pavlova, Griffith University, Brisbane, QLD, Australia

© 2010 Elsevier Ltd. All rights reserved.

Introduction

Planning and policy development in any organization can be analyzed and understood through a number of approaches. Traditionally, policy has been conceptualized as a program of action. Any policy has usually been presented as a “rational plan, consciously articulated by an authoritative body, usually a government or government agency, codified in text such as law or regulation which articulates clear expectations for behaviour” (Bascia, *et al.*, 2005: xii). This tidy linear process in which policy progresses from one stage to another and where policy is defined as whatever governments choose to do, or not to do (Dye, 1992) has been criticized by policy researchers from different theoretical standpoints (e.g., liberal, neoliberal, positivist, or post-structural approaches). There is now a considerable body of theoretical and empirical work that views the nature of planning and policy development as “far more complex, dynamic, and interactive than any of the traditional linear or staged models suggest” (Walford, 2003: 2). The complexity of the process has become an important feature of a number of well-known frameworks for policy analysis, for example Ball’s (1993, 1994) and the Taylor *et al.* (1997) models. Taylor *et al.* (1997) focus on three aspects of policy: context (economic, social, and political factors that give rise to an issue emerging on the policy agenda), text (broadly refer to the content of the policy), and consequences (differences in implementation). Ball’s (1993, 1994) model of policy as text and policy as a discourse provides another way to understand the complexity of policymaking. Both frameworks are widely used in the literature (e.g., Bell and Stevenson, 2006; Olssen *et al.*, 2004). For example, Olssen, *et al.* (2004) in developing Ball’s approach argue that: “Education policy must be contextualised both nationally and globally as a transformative discourse that can have real social effects in response to contemporary crises of survival and sustainability” (p. 3).

The purpose of this article is not to analyze the differences between the various approaches to policy development and policy analysis but to explore a number of important considerations in terms of technical and vocational education and training (TVET) policy development, such as scope, scale, principles, and theoretical justifications. The point that we are making is that

TVET policy is multidimensional, is value-laden, exists in particular contexts, interacts with policies in other fields, its implementation is never straightforward, and it results in unintended and intended consequences. Policy-making for TVET is mainly a state activity and the state is regarded as a complex and nonunitary entity of competing parts that adds an additional aspect to any analysis.

Complexity of TVET

Scope

The great diversity of TVET systems worldwide illustrate a range of policy options that have been used and are available to shape and reform them. In trying to classify existing TVET systems, researchers have considered a number of factors, including place of learning (typically, vocational schools and the workplace/enterprise workshops); type of regulation (government, market, and tradition, see e.g., Greinert, 1995); or nonformal/informal learning as a leading criterion for classifications.

In reality, combinations of these patterns across the above dimensions can be found in particular countries and all these dimensions should be accounted for when undertaking planning and policy development in particular countries.

Scale

Another component of complexity in policy formulation for TVET can be related to the scale of the policy within (any) TVET system. Improvements within a TVET system can be carried out in terms of a particular element of the system, such as qualification standards and specialization; articulation with general education; assessment and certification for achievements; methods of curriculum development; governance; labor-market analysis; financing; legislation; access and admission; provision of training places; special target groups; institutional arrangements for delivery; and, information and guidance. Alternatively, improvements can be made to the system as a whole which is known as a systematic reform. Systematic reform is built upon the assumption that changing most or all elements of the system is more likely to lead to TVET improvement.

Coherence

In planning TVET systems and in developing policies, coherence across sectors such as the interests of different institutions involved and different policies should be taken into consideration. TVET is situated at the border between education and work. To contribute successfully to economic development TVET policy must outline: “coherent and integrated changes in a range of related institutions, including the economic, human resources and [TV]ET agencies of the state, the labour market, the social organisation of work and in the forms of employer and trade union organisation” (Kraak, 2006: 3).

A coherent approach to TVET policy development needs to relate it to a range of other policies and issues, including general education policy (both school and university); youth policy; adult education policy; vocational rehabilitation; public service employment and remuneration conditions; economic policy; private-sector and enterprise-promotion policies; industrial development policies; and labor-market and employment policies (Krönner, 2006). Thus, TVET policy should be closely linked with other policy areas and institutions, in particular to those dealing with education, work, employment, and economic development.

Globalization

The strategic direction of much of education policy is frequently justified as a means to enhanced economic development leading to a more competitive economy, greater productivity, and increased wealth. Investment in education and training is believed to provide the key to national competitiveness. Globalization processes and the pressures for global uniformity have the tendency to strengthen the Anglo-American influence on policy development and raise concerns as to the appropriateness of such an occurrence:

Anglo-American scholars continue to exert a disproportionate influence on theory, policy and practice. Thus a relatively small number of scholars and policy makers representing less than 8% of the world's population purport to speak for the rest. (Walker and Dimmock, 2002: 15)

However, many recent studies have shown that globalization-driven policy developments are sometimes so sharply modified by national contexts that the output they promise would vary greatly (Rhoten, 2000; Welmond, 2002; Astiz *et al.*, 2002) if implemented in another context. Therefore, for the policy formulation, the specific context of a particular country should be taken into account to identify an adequate response to immediate and future labor-market needs. It is important to “identify that way in which global pressures have driven state restructuring. . . and the particular way in which restructuring at the

micro-level shapes policy development at an institutional level” (Bell and Stevenson, 2006: 33). What is required are the analyses of the state, economy, and models of policy development that recognize different cultural contexts. “Generally, the more ‘loosely-coupled’ state structures are, the more opportunity there is for policy variation at the local level” (Bell and Stevenson, 2006: 30).

Principles of Policy Planning

At the level of global policy, principles for policy planning and administration have been established in the joint document by the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the International Labour Organization (ILO): *Technical and Vocational Education and Training for the Twenty-First Century* (UNESCO and ILO, 2002). The document states that policy should be directed to both the structural and the qualitative improvement of TVET. The partnership between the government and the other interested partners must “create a coherent legislative framework to enable the launching of a national strategy for change” (para 9 a, p. 13). The UNESCO and ILO document presents five principles:

- “Planning should respond to national and, if possible, regional, economic and social trends, to project changes in demand for different classes of goods and services, and for different types of skills and knowledge in such a way that technical and vocational education may easily adapt to the evolving scientific, technological and socio-economic changes. This planning should also be coordinated with current and projected training action and the evolution of the world of work in both urban and rural areas.” (para 11, p. 14)
- “Policies for the structural improvement of technical and vocational education should be established within the framework of broad policies designed to implement the principle of lifelong education through creation of open, flexible and complementary structures of education, training and education and vocational guidance, considering the provisions of modern information technology in education regardless of whether these activities take place within the system of formal education or outside it” (para 13, p. 15)
- “Policy should be directed to ensuring high quality so as to exclude discrimination between the different educational streams. In this respect, special efforts should be made to ensure that national technical and vocational education seeks to meet international standards” (para 14, p. 16)
- “In order to ensure quality, responsible national authorities should establish criteria and standards, subject to periodic review and evaluation, applying to all aspects of technical and vocational education, including to the

greatest extent possible, non-formal education” (para 15, p. 16)

- “National policy should foster research related to technical and vocational education, with particular emphasis on its potential within lifelong learning, and directed to its improvement and relevance to the prevailing socio-economic context. This research should be carried out at national and institutional levels, as well as through individual initiative” (para16, p. 16).

These principles highlight the importance of national policies aimed at the quality of training, which can be achieved through structural change. These principles and approaches represent a traditional way of interpreting policy as a rational plan of action; therefore, the level of effectiveness of their application will be different for different countries. For example, with a state-guided labor market, Singapore might find these principles to be more useful than other countries where the labor market is less regulated. These principles provide a useful list of general rules that should be used creatively by governments.

Theoretical Frameworks

This section discusses a possible basis for theoretical frameworks that could provide a foundation for TVET policy formulation in both developed and developing countries. The first set of theories relates to the structural reform/policy formulation and the second set to the theories that shape our understanding of how humans learn.

The idea of an alignment between education, training, labor market, and economic policies can be underpinned by the theoretical arguments of the high-skills thesis, combined with what is described as the joined-up approach (Brown, *et al.*, 2001). The high-skills thesis has caught the imagination of policymakers around the world (p. 31). It was developed through the work of a UK team in the late 1980s and 1990s (see Finegold and Soskice, 1988; Finegold, 1991; Ashton and Green, 1996; Brown *et al.*, 2001). They argued that diversity and variability in economic performance are underpinned by social, cultural, and historical factors. Finegold (1991) explores the concept of equilibrium in terms of the self-reinforcing nature of the institutional networks that support the continuation of a given skills-formation system and economic development path. Any change of one variable without corresponding shifts in the other institutional variables is unlikely to lead to long-term changes in the social and economic system as a whole. Using this concept, Finegold identified two ideal types of economic and education/training (ET) systems: an institutional framework based on low-skills equilibrium (LSE) and one based on high-skills equilibrium (HSE). Finegold *et al.* viewed Britain as being typical of an LSE

society where “the majority of enterprises are staffed by poorly trained managers and workers produce low quality goods and services” (Finegold and Soskice, 1988: 22). LSE institutional factors identified by Finegold discourage and constrain any movement toward a high-skill alternative. They included uncoordinated state policies, incoherent TVET policies, low educational level of the majority of workers, minimal state intervention in TVET, and labor markets. Another analysis (see Brown, *et al.*, 2001) demonstrates that there is some presence of high skills in Britain. Thus, joined-up or cross-sectoral policy coordination and complementarity are associated with a high-skills thesis. Joining up is essentially an argument about the necessity for educational reforms to interlock with macroeconomic, industrial, and labor-market reforms so that the combined impact would influence the outcome in a significant and positive way.

A high-skills policy has been the center of economic, social, and educational policies in many Western countries. However, due to differences in context “there is no one recipe or blueprint for developing a high-skills economy” (Lauder and Brown, 2006: 37). Lauder and Brown (2006) identified four different types of labor market structures. An occupational labor market where education and training are closely related to the labor market; an internal labor market, where high general skills are provided by TVET and specific skilling is undertaken by corporations; and the state-guided labor market which is a demand-led (e.g., specific to Singapore where the state encourages high-end production and innovation and matches skills to the demand of the multinational corporations) market structure. All three types of labor markets have in common a close integration of TVET and the labor market. The fourth type of market, a flexible labor market, is characterized by a high degree of individualism, “in which the fit between education, training and the labour market is often problematic; there is considerable time spent in job search and work is insecure; and the flexibility is numerical, so that employers can hire and fire as economic conditions change” (Lauder and Brown, 2006: 39). Some authors are less critical of flexible labor market; however, there is agreement that more emphasis on the development of generic skills and good credentialing systems are important for this type of labor market to be successful. Each type of labor market influences policy development in TVET and the type of skills that are emphasized in the policy. If one accepts that the long-term aim of the economy is to achieve an HSE and capacity building, then contributing to achieving such a goal should be seen as an important function of TVET policy development.

Each country can be conceptualized within this framework, for example, high-skills society (Germany), developmental high-skills society (Singapore), bipolar high-skills/low-skills society (UK), and hybrid and

differentiated low-, intermediate-, and high-skills society (South Africa) (Brown, *et al.*, 2001; Kraak, 2006). TVET policy developers could use this analysis as a starting point for policy formulation.

Theories about human learning have shaped our views of the conditions necessary to support curriculum development and instruction. However, it should be noted that learning in TVET as well as learning at work, as such, are seldom the central focus of policy. “Rather, the typical situation is that policies that are targeted at something else include learning at work within their scope. In the process, learning at work is usually assumed to be, in an uncritical and ‘common sense’ way, simply the acquisition of discrete items of knowledge and skills” (Hager, 2005: 830).

Recent research on TVET (see, e.g., Staron *et al.*, 2006) argued that life-based learning (not work-based learning) was required for vocational education focusing on capability development and consideration of the learner as a whole person. According to Staron *et al.*, life-based learning is a win-win situation with benefits to both employee and employer, based on employees taking personal responsibility for their learning, based on the provision of “rich learning environments” (p. 49), which provides, according to Staron *et al.*: “performance, growth and opportunity. It is adaptive, self-facilitated, based on reflexive practice and uses any strategy appropriate to the task” (p. 49). This broad interpretation of learning views TVET as “a knowledge-based industry, where knowledge is its core business” (Staron *et al.*, 2006: 24). As argued in the *Life Based Learning* report (Staron *et al.*, 2006): the current period is one of rapid change where new ways of working and living are required. Staron *et al.* argue that a learning ecologies metaphor can be used to understand how we might work and live successfully. Staron *et al.* focus on the tendency of ecological environments to move to equilibrium as the key to the metaphor. Thus, when there is change (to work or life) this is regarded as a disturbance and we need to work to achieve equilibrium, and we do that by adapting. Rather than a model or a set of procedures, the focus of an ecological approach is orientation and thus “it offers a complex, diverse, dynamic and adaptive framework that gives us a fresh perspective on working and learning in contemporary environment” (Staron *et al.*, 2006: 27).

These two types of theories, one that relates to the economic aim of achieving an HSE and the other that focuses on the learner in coherent ways could provide a useful framework for policy development.

Cases Around the World

Policy development as a complex and multifaceted process can be identified through the product, the policy document, and the process in which implementation is as

important as formulation. According to Bell and Stevenson (2006), “It is important to understand the context in which policy development takes place – how policies emerge, how they form and take shape, and how they become lived through the actions of those engaged in the policy-development process” (p. 25). Smyth (1993) has emphasized the need to recognize the pivotal role of central governments in shaping policy.

This section examines four policies across developed and developing countries to illustrate the points discussed above. Two examples from England and Australia relate to policy as text or the content of the policy itself. They illustrate a capacity-building thesis within the lifelong learning paradigm. Two examples from developing countries relate to policy as discourse that gradually builds over time and sets boundaries to what actors are allowed to think and do. Discourses, then, as Ball (1990) summarizes them, “embody meaning and social relationships, they constitute both subjectivity and power relations” (p. 2).

The first example is the English participation policy (2007). The English document: *Raising Expectations: Staying in Education and Training Post-16* presented to Parliament in March 2007 (Department for Education and Skills, 2007) argues for the new policy on the basis that “the structure of the economy has changed dramatically, with an ever-growing dependence on more highly skilled workers” (1.5, p. 9). The goal of the policy is: “to stop young people dropping out of education or training before 18, because of the benefits that higher attainment and longer participation bring, to the individual young person, to the economy and to society” (p. 11). Continuing in education or training is seen as important to the economy, “which will increasingly demand more highly skilled employees” (2.2, p. 11) and the aim is to reach 90% participation in education or training among 17-year-olds by 2015. Measures across different areas already being implemented in England include: reforms to the 14–19 curriculum; changes to Key Stage 3; a major expansion of capacity in work-based learning; the Education Maintenance Allowance; the Youth Matters reforms, and others. Employers are playing an important part in providing training or releasing young people from work to undertake training. This policy is closely related to a high-skills thesis. It is also an example of a policy that addresses one issue across a number of sectors and positions it within some existing policies. The pressure of globalization is presented through the request to increase competitiveness of the economy.

Since the 1990s there has been a similar concern across Australia that young people are not gaining the right skills to compete in the global economy. Although the different state governments (who have responsibility for education) have responded in different ways, *Australia’s National Strategy for Vocational Education and Training 2004–2010*

(ANTA, 2003) developed at the national level, highlighted an attempt at a systematic approach toward TVET development in the country. Key objectives for vocational education and training (VET) that are formulated in the National Strategy 2004–10 highlighted:

1. the global economic competition as the reason for skills development (industry will have a highly skilled workforce to support strong performance in the global economy);
2. the end-user as a focus of the policy (employers and individuals will be at the center of VET); and
3. benefits for communities (communities and regions will be strengthened economically and socially through learning and employment and Indigenous Australians will have skills for viable jobs and their learning and culture will be shared).

In November 2003, state ministers endorsed the key performance measures for the VET system as part of the 2004–10 National Strategy. The key performance measures cover the following critical areas:

- Student participation and achievement in VET.
- Student employment outcomes and satisfaction from VET.
- Employer engagement and satisfaction with VET.
- VET outcomes for Indigenous Australians.
- Community engagement and satisfaction with VET.
- VET system efficiency (ANTA and NTSC, 2005).

Some of the strategies identified include: increase participation and achievement in TVET, particularly by existing workers; help clients navigate and interact with VET; improve the value, brand, language, and image of VET and public recognition of its employment outcomes; strengthen industry's role in anticipating skill requirements and develop products and services to meet them; make learning pathways seamless; improve quality and consistency; and facilitate access to international markets. This policy is an example of systematic reform, where participation (as in the English case) is just one component of the policy.

Both examples from developed countries represent competitiveness-driven TVET policies as governments sought to defend their global competitiveness by enhancing the productivity of the domestic labor force.

Two examples from developing countries come from Africa and Cambodia. Although education policies in developing countries can be closely related to donor priorities and perspectives, these cases provide illustrations of how policy can be used in different ways. On the one hand, international funds can be used to implement measures that are episodic and lack essential coherence. On the other hand, they can be used to provide partial financial support for a large-scale systematic educational reform.

Between the 1980s and 1990s, international donor funding in Africa was directed away from TVET and toward support for basic education. TVET provision was marginalized as the attention of policymakers was directed elsewhere (McLean and Kamau, 1999). However, during the late 1990s, two factors emerged: a growing crisis of youth unemployment and low levels of economic competitiveness. These factors became the main policy drivers in bringing skills back into the policy spotlight, particularly for the Southern African countries (Roberts, 2005: 50). One of the challenges facing educational planners is the need to develop a coherent policy on skills development that articulates well with the economic development plans.

This example can be analyzed within Estevez-Abe *et al.* (2001) approach where they distinguish three types of skills: firm-specific (Japan and Korea), industry-specific (Germany), and general skills. In any economy all three types of skills will be utilized; however, in some structured economies, the development of one of these skills types will predominate over the others. Firm-specific skills are the least portable, industry-specific skills are portable across an industry, while general skills can be applied across a range of firms and industries. Each type requires different type of training and assumes particular kinds of economic development strategies. As argued by Lauder *et al.* (2006), “for firm-specific skills a high level of general education is desirable. This is then used as the foundation for the in-house development of skills that firms demand. Industry-specific skills require some interaction between the education and training system and industry for the training to be appropriate and up to date in supplying the skills required. . . . For general skills, education at only a minimum level is required for the lower end of the flexible labour market.” (p. 48). Thus, the assumption that an improvement in general education in Africa would provide the opportunity for firm-specific training proved to be wrong. This was due to the limited number of firms able to provide such training. The conclusion to be drawn from this is that more emphasis on the development of industry-specific skills is required so that the economic development plans can be achieved. “This requires government commitment in the form of policy development, sound skills and human resource development planning, inter-ministerial co-operation – primarily between education and labour ministries – and financial commitment to developing a skilled population” (Roberts, 2005: 52).

Cambodia is an example to demonstrate the effectiveness of a sector-wide education approach presented through a developed national program for education in a developing country. In Cambodia's case the approach comprised: The Education Sector Support Program (ESSP) 2001–06. The overall goal of the ESSP was to contribute to the “achievement of the Government's poverty reduction strategy through pro-poor systematic and

targeted interventions” (Forsberg and Ratcliffe, 2002). TVET is included in this policy. Government flexibility in negotiating a mix of donor-financing modalities helped maintain inclusive partnerships. High-level national leadership and authority of the ESSP reform process has been critical and it was provided directly through the minister. This is an example to illustrate when policy development is not a simple case of setting up the priorities. Policy had been seen as a dialectic process in which the government and the donors were involved in shaping its development. “Policy development is therefore both a continuous and contested process in which those with competing values and different access to power seek to form and shape policy in their own interests” (Bell and Stevenson, 2006: 2).

These four cases demonstrate the different ways the governments approach policy development: through the high-skills thesis in Australia and England; and through the combination approach (high skills/low skills) in Africa and Cambodia. Systematic approach at different levels has been seen important for the three cases examined.

Conclusions

This article analyzes issues, such as the scope, scale, and coherence within TVET and across different sectors within particular countries to demonstrate the complexity of TVET policy formulation. It is argued in the article that tensions between the pressures for global uniformity and regional specificity should be considered by policy-makers. It is also argued that any attempt to develop a universal set of principles for policy development should be adopted with caution, particularly due to differences in the labor-market structures across the countries. Two types of theories are proposed to provide the basis for understanding how educational policy could be planned. They are high-skills thesis and life-based learning. They provide the basis for macro- and micro-level planning.

The issues and theories discussed in the opening section of the article provide the basis for the analysis of policy examples from developed and developing worlds presented in the next section. The conclusion is advanced that TVET systems policies and planning should exist as part of a system of lifelong learning adapted to the needs of each particular country and to worldwide economic development. It is important to be aware of the complexity of TVET and its interdependence with other policy areas. In a modern market economy, governments carry primary responsibility for TVET policy design. However, a partnership between government, employers, professional associations, industry, employees and their representatives, the local communities, and nongovernment organizations should play a crucial role in policy design (UNESCO/ILO, 2002: 9a). The overall conclusion to be advanced in this article is that policy planning and

development should be based on research that is directed toward TVET improvement and its relevance to particular socioeconomic contexts.

Bibliography

- ANTA (Australian National Training Authority) (2003). *Shaping our Future: Australia's National Strategy for Vocational Education and Training 2004–2010: Final Report*. Australian National Training Authority, Brisbane.
- ANTA (Australian National Training Authority) and NTSC (National Training Statistics Committee) (2005). *Shaping our Future: Measuring the Future: Key Performance Measures for Australian Vocational Education and Training 2004 to 2010: An Implementation Plan / [National Training Statistics Committee]*. Brisbane: ANTA.
- Ashton, D. and Green, F. (1996). *Education, Training and Global Economy*. Cheltenham: Edward Elgar.
- Astiz, M., Wiseman, A., and Baker, D. (2002). Slouching towards decentralization: Consequences of globalisation for curriculum control in national education systems. *Comparative Education Review* 46(1), 66–88.
- Ball, S. (ed.) (1990). *Foucault and Education: Disciplines and Knowledge*. London: Routledge.
- Ball, S. (1993). What is policy? Texts, trajectories and toolboxes. *Discourse* 13(2), 10–17.
- Ball, S. (1994). *Education Reform*. Buckingham: Open University Press.
- Bascia, N., Cumming, A., Datnow, A., Leithwood, K., and Livingstone, D. (eds.) (2005). Introduction. *International Handbook of Educational Policy*, pp XIII–XXXVII. Springer: Dordrecht.
- Bell, L. and Stevenson, H. (2006). *Education Policy: Process, Themes and Impact*. Routledge: London.
- Brown, P., Green, A., and Lauder, H. (2001). *High Skills, Globalisation, Competitiveness and Skill Formation*. Oxford: Oxford University Press.
- Department for Education and Skills (2007). *Raising Expectations: Staying in Education and Training Post-16*. Norwich: TSO.
- Dye, T. (1992). *Understanding Public Policy*. Englewood Cliffs, NJ: Prentice Hall.
- Estevez-Abe, M., Iversen, T., and Soskice, D. (2001). Social protection and the formation of skills: A reinterpretation of the welfare state. In Hall, P. and Soskice, D. (eds.) *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. Oxford: University Press.
- Finegold, D. (1991). Education, training and economic performance in comparative perspective. *Oxford Studies in Comparative Education* 1, 57–68.
- Finegold, D. and Soskice, D. (1988). The failure of training in Britain: Analysis and prescriptions. *Oxford Review of Economic Policy* 4(3), 21–51.
- Forsberg, G. and Ratcliffe, M. (2002). Cambodia education case study: Education sector support program. *Presented at the IIEP Conference on Sector Wide Approaches in Education*, Paris, December 2002.
- Greiner, W.-D. (1995). *The Dual System of Vocational Education and Training in the Federal Republic of Germany. Structure and Function*, 2nd rev. edn. Stuttgart: Holland + Josenhans.
- Hager, P. (2005). Current theories of workplace learning: A critical assessment. In Bascia, N., Cumming, A., Datnow, A., Leithwood, K., and Livingstone, D. (eds.) *International Handbook of Educational Policy*, pp 829–846. Dordrecht: Springer.
- Kraak, A. (2006). High skills and joined-up policy: An introduction to the debate. In Kraak, A., Lauder, H., Brown, P., and Ashton, D. (eds.) *Debating High Skills and Joined-Up Policy*, pp 1–30. Cape Town: HSRC (Human Science Research Council) Press.
- Krönner, H. (2006). Scope of TVET Policy. *Summaries of the National Workshop on Technical and Vocational Education and Training. Pyongyang, Democratic People's Republic of Korea, 30 August–1 September 2006*, .
- Lauder, H. and Brown, P. (2006). The high-skills thesis. In Kraak, A., Lauder, H., Brown, P., and Ashton, D. (eds.) *Debating High Skills and*

- Joined-Up Policy*, pp 31–43. Cape Town: HSRC (Human Science Research Council) Press.
- Lauder, H., Brown, P., and Ashton, D. (2006). Globalisation, skills formation and the dilemmas of integrated policy: The case of South Africa. In Kraak, A., Lauder, H., Brown, P., and Ashton, D. (eds.) *Debating High Skills and Joined-Up Policy*, pp 31–43. Cape Town: HSRC (Human Science Research Council) Press.
- McLean, G. N. and Kamau, D. G. (1999). Human resource development and vocational and technical education in Kenyatta University, Kenya. Paper presented at the *European Conference on Educational Research*, Lahti, Finland. <http://www.leeds.ac.uk/educol/documents/00001203.htm> (accessed September 2009).
- Olssen, M., Codd, J., and O'Neill, A.-M. (2004). *Education Policy: Globalization, Citizenship and Democracy*. London: Sage.
- Rhoten, D. (2000). Education decentralisation in Argentina: The global local conditions of possibility approach to state, market and society change. *Journal of Education Policy* **15**, 593–619.
- Roberts, J. A. (2005). Locating skills development in the framework. *OpenSpace* **1**(2), 47–52.
- Smyth, J. (ed.) (1993). *A Socially Critical View of the Self-Managing School*. London: Falmer.
- Staron, M., Jasinski, M., and Weatherley, R. (2006). Life based learning: A strength based approach for capability development in vocational and technical education. *A Report on the Research Project "Designing Professional Development for the Knowledge Era"* Darlinghurst, NSW: TAFE NSW.
- Taylor, S., Rizvi, F., Lingard, B., and Henry, M. (1997). *Educational Policy and the Politics of Change*. London: Routledge.
- UNESCO and ILO (International Labour Organisation) (2002). *Technical and vocational education and training for the twenty-first century*. UNESCO/OIT: UNESCO and ILO Recommendations Paris.
- Walford, G. (2003). Introduction: Investigating education policy. In Walford, G. (ed.) *Investigating Educational Policy through Ethnography*, pp 1–15. Oxford: Elsevier Science.
- Walker, A. and Dimmock, C. (eds.) (2002). *School Leadership and Administration: Adopting a Cultural Perspective*. London: Routledge.
- Welmond, M. (2002). Globalisation viewed from the periphery: The dynamics of teaching identity in the Republic of Benin. *Comparative Education Review* **46**(1), 10–65.

Qualifications Frameworks and their Role in the Reform of Education and Training

M Young, University of London, London, UK; University of Bath, London, UK

© 2010 M Young. Published by Elsevier Ltd. All rights reserved.

Introduction

This article focuses on the recent trend for governments (in the UK and elsewhere) to use qualifications as a main driver of educational reform, especially, but not only, in the field of vocational education and training. (Qualifications are increasingly linked closely to funding arrangements. These links are important but beyond the scope of this article.) Qualifications are an element of the education systems of all societies; they do, to an extent, motivate learners, they are used by employers to screen applicants for jobs, and increasingly, as discussed in this article, they are used by governments in controlling educational institutions. However, what is questioned is whether, as is increasingly assumed, qualifications expressed as outcomes or standards can be the basis for improving education and, more specifically, can provide an adequate basis for teachers to develop the curriculum. This can only lead, it is suggested, to a collapse of educational standards and a loss of credibility of qualifications.

Outcomes-based qualifications, if seen as a basis for replacing syllabi, in effect, do away the idea of a curriculum at all. The consequences, in the case of South Africa, of introducing such an approach to educational reform are made very clear by Allais (2003) and by a number of contributors to the book by Young and Gamble (2006).

Emerging Trends and Issues in Qualification Reform

The issues discussed in this article arise from the growing interest of governments in the idea of a national qualification framework, based on outcomes defined according to a common set of criteria, and including all levels and types of qualification, whether they are school or work based, general, vocational, or professional in purpose.

The origin of the idea of a qualification framework based on outcomes can be traced back to the National Vocational Qualifications (NVQs) framework introduced in the UK in 1987. (NVQs drew on the method of functional analysis developed within occupational psychology (Wolf, 1995).) To the best of our knowledge, this was the first attempt to establish a framework that was based on outcomes that were defined independently of any specific learning program, although the NVQ's framework was

restricted to vocational qualifications. However, Jessup (1991), a leading figure in its design, envisaged that learning outcomes could be the basis for all education.

What was not fully recognized at the time of the launch of NVQs in 1987 was that the concept of qualifications based on learning outcomes implies very specific ideas about learning. If there is to be formal comparability between different qualifications within a framework, the criteria for learning must be precisely specifiable in advance of the learning itself. It follows that such qualifications disregard any learning that is not specifiable in advance or may not be specifiable at all (e.g., apprentices acquire integrity and a sense of responsibility which could be assessed by professional judgment but not on the basis of criteria alone).

Some kind of broad prespecification of outcomes is, of course, a feature of all qualifications. However, in the case of NVQs, not only did the prespecification of outcomes become more detailed, but also the other features of traditional qualifications, such as a recognized syllabus and an assumed and the requirement to attend a course of study, were excluded. The precise specification of outcomes was designed to replace the kind of tacit assumptions which, in more traditional qualifications, were used to compare and rank learners. There was also an assumption in the NVQ approach that precisely defined outcomes would not only free qualifications from any reference to where they were achieved, but also minimize the human judgment required for assessing learning. It was this claim to independence from both professional judgment and required attendance at an institution that gave outcomes-based qualifications their claim to be radical and egalitarian.

The advantages of an approach to qualifications based on outcomes apply most readily to vocational qualifications, where competences can be specified. However, the implications are far more problematic in relation to the goals of general education which can never be adequately expressed in terms of precisely specified outcomes. The detailed prespecification of outcomes of the kind attempted for NVQs downgrades both the knowledge that may be required for competent performance and the role of the professional judgment of teachers. (The requirement, insisted by the National Council for Vocational Qualifications (NCVQ) in the UK, was that teachers and trainers had to learn to interpret criteria in the correct way. It was

not surprising that this resulted in box ticking and was experienced as de-skilling.) It also underemphasizes the forms of tacit learning that cannot be assessed or expressed in terms of outcomes but which all learners rely on.

The second significant development that this article takes into account is the growing body of research and experience that suggests that outcomes-based frameworks have a number of unfortunate consequences (Young, 2004). It will be argued that these consequences are not just teething problems associated with any radical innovation, but that they reflect fundamental flaws in qualifications-driven approaches to educational reform.

The third issue that is important to bear in mind is that a number of countries, most notably those associated with the Germanic and Nordic traditions of education and training, have remained largely immune from the pressures to develop outcome-based qualification frameworks (Young, 2004). In this article, we contrast the Anglophone approach that sees a qualification framework as the best way of responding to global economic change with one that sees qualifications as part of the institutional structure of a country's education and training system.

The Rest of the Article

The remaining sections of this article explore in more detail the three trends already referred to. It makes a distinction between countries adopting an outcomes-based approach to reform and those relying on an institutional approach. Finally, it draws on a distinction between the intrinsic and institutional logics (Raffe, 1992) of any educational policy. Is Raffe's analysis a step further by distinguishing between the macro and micro aspects of the two types of logic?

Intrinsic logics refer to the claims made for a reform that are independent of the actual contexts in which it might be implemented. They express the political rationale of a reform and are invariably the basis of any borrowing of ideas and strategies between countries. Institutional logics, on the other hand, refer to the social, political, and institutional contexts, and the role that they are likely to play in how (and whether) any reform is implemented.

Whereas macro aspects of institutional logics refer to forms of stratification, power relations, and institutional hierarchies, micro aspects refer to the specific practices of teachers, learners, and assessors. It is the macro aspects of institutional logics that are the most visible constraints on educational reform but it is the less-visible micro aspects that provide the conditions for processes such as acquiring a qualification to take place and gain public credibility.

Those countries that have developed outcomes-based approaches share common notions of their intrinsic logic.

However, they appear to have given very little attention to either aspects of the institutional logic of such reforms. Contrasting the two approaches can shed light on: (1) the limitations of outcomes-based approaches; (2) the strengths and weaknesses of institutional models; and (3) the dangers of assuming that qualification reforms on their own can play more than a modest role in improving education and training.

Qualification Reforms: The Context

The main goals of recent qualification reforms have been to improve flexibility, widen participation, and enhance mobility. In furthering these goals, most reforms emphasize portability – allowing people to move between different types of qualifications and transparency – making explicit what learners have to do to become qualified.

A number of questions about the link between these aims and the proposed qualification reforms need to be asked. First, why have governments focused on qualifications as their main instrument of reform? Second, why has the idea of outcomes-based frameworks proved so attractive to governments? Third, is it realistic to envisage common criteria that could be the basis of a framework for all qualifications? Fourth, given that some of the most effective education and training systems up to now (e.g., those in East Asia and parts of continental Europe) rely on a quite different approach, why are outcomes-based frameworks so popular in Anglophone countries?

Why Qualifications?

It seems likely that one reason why governments have become so enthusiastic about emphasizing qualifications is that not only are they assumed to motivate learners, but they can also serve quite other roles that are just as important but frequently less-explicit aspects of government policy. For example, a greater emphasis on qualifications enables central governments to:

- increase their control of educational provision;
- provide measurable criteria for allocating funds;
- make local and regional bodies more accountable; and
- provide evidence of the success of government policies.

Two continuing parallels are worth noting. The first is that congruence between the growing interest in qualifications and neoliberal, market-oriented approaches both to the economy and the public sector. The second is that qualification reform from the 1980s has been given most support in those countries which had weaker public systems of education and training and most readily endorsed the original Reagan–Thatcher neo-Liberal policies (e.g., the UK, Australia, and New Zealand).

There are parallels between new qualifications, authorities, and the bodies that have been created to regulate the new privatized monopolies such as water, gas, and electricity. Qualifications offer an ideal regulatory instrument for a reforming government. They provide incentives for individual learners and can be used to make educational institutions accountable. The educational problem, however, is that these purposes can be in conflict with each other. More emphasis on accountability leads to tighter specification of outcomes. Promoting learning, however, especially among slow learners, requires teachers the space to take risks. Tackling the problems faced by slow learners will require qualifications that rely more on professional judgment and are less specified in advance.

Why an Outcomes-Based Qualification Framework?

Qualification arrangements in England and Wales have in the past had a number of features in common with those found in other European countries. For example:

- General and vocational qualifications have developed separately with limited possibilities of progression between them. The latter are organized by sectors with few opportunities for transfer.
- Most qualifications are linked to programs in educational institutions.
- Qualifications have traditionally been underpinned by specialist communities – trades, crafts, and professional organizations and disciplinary associations. Historically, these specialist communities have been socially exclusive, both generally and in relation to specific disadvantaged groups distinguished by race, class, and gender.
- Traditional qualifications have offered only limited forms of access to adult learners except by retracing the steps of young learners.
- Many occupations and sectors have been characterized by very few of those employed having any qualifications at all. In the UK, a license to practice has only been applied in the case of a small number of occupations.

Since the late 1980s, all these features have been seen by as barriers to learning, and at odds with the fast-changing skill and knowledge demands of the global economy. A single framework with a single set of levels in which all qualifications are located appeared in principle to be a logical way of overcoming these barriers.

The shift that has been argued for, since 1987 in the UK, is from a qualification system based on shared practices and judgments to one based on formally explicit criteria defined independently of any specific experience or practice.

Support for a change from a qualification system based on shared practices to one based on criteria was political as much as educational. The old provider culture of further education (FE) colleges, local employers, and the examining bodies was widely discredited among civil servants and within national employer organizations. This view fitted in with the marketizing zeal of both Conservative and Labour governments. Initial support for reform was tied to the government's determination to break the power of trade unions and their out-of-date and restrictive practices (Raggatt and Williams, 1999). Attempts to break with the past have been a familiar feature of qualification reforms, and not only associated with the political Right. Examples are Wales and Scotland, both with left-of-center coalitions and the first post-apartheid government in South Africa (Jansen and Christie, 1997).

Lastly, it has been claimed (Jessup, 1991) that a qualification framework, freed from any links with educational institutions, can provide access to groups who have in the past been excluded. The idea of a qualification framework has been closely linked to the accreditation of informal learning (DfEE, 1998). However, it is hard to find evidence that substantial numbers of workers have had their informally acquired skills recognized as qualifications.

The Typology: Outcomes-Based and Institution-Based Approaches to Qualifications

The typology distinguishes between outcomes-based and institution-based approaches. Qualification reforms, however, should not be seen in isolation from other elements of the their wider social context which will shape the way in which qualifications operate. Examples are:

- funding systems,
- devolution of government,
- the role of the private sector as provider and funder, and
- the location of assessment (in public or private bodies)

We shall use the two models to highlight changes as well as some of the problems they give rise to, not as descriptions of particular countries.

Outcomes-Based Approaches

Outcomes-based approaches to qualifications assume:

- a single definition of qualifications,
- a single set of level descriptors, and
- the precise specification of outcomes

As a result of relying on precise specification, outcomes-based approaches invariably lead to over-specification and the trivializing of learning (Wolf, 1995). There is evidence from both New Zealand and South Africa that

outcomes-based frameworks have a series of other serious but unintended consequences. For example:

1. they encourage resistance rather than support from high-status institutions, such as universities (Ensor, 2004) and professional bodies;
2. their emphasis on flexibility neglects the positive role of boundaries between qualifications in promoting particular types of learning (especially at higher levels);
3. in emphasizing generic criteria for all qualifications, single frameworks tend to neglect the importance of specific content in both vocational and general subject areas;
4. the leadership role of qualifications is overemphasized; and
5. there is as yet very little evidence that the quality and amount of learning have been enhanced in countries that have introduced national frameworks.

In the case of NVQs in the UK, the gross numbers achieving vocational qualifications have not changed significantly since they were introduced.

Institutional Approaches

Institutional approaches to qualifications are less easy to characterize, partly because in many countries they are taken very much for granted and have a long history. Qualifications are not treated as separate instruments of reform but as embedded in the wider education and training system; outcomes are closely linked to institutional inputs or programs of study. Qualifications are not used by governments as an independent lever for change.

However, in the context of global economic changes, institutional approaches are not without their weaknesses. For example:

- they exhibit inertia and resistance to change;
- change is difficult because it involves the whole system;
- they are slow to adapt to new learning demands; and
- progression between different sectors is difficult.

Discussion

Most industrial countries still largely have institution-based systems of qualifications, although a growing number are exploring the scope of outcomes-based approaches. The exceptions appear to be the Anglophone countries. However, England, like New Zealand before it and more recently South Africa, is stepping back from an extreme version of the outcomes position initially adopted by the NCVQ (Raggatt and Williams, 1999). The proposed European Qualifications Framework has a different approach as a voluntary register. This evidence of doubts about outcomes-based approaches points to a middle way. The next section draws

on the distinction between intrinsic and institutional logics to suggest future possibilities.

The Intrinsic and Institutional Logics of Outcomes-Based Frameworks

The rationale for introducing an outcomes-based qualifications framework is as follows. First, it is consistent with the increased flexibility of successful modern economies. Second, it offers scope for transferring qualifications between sectors as learners accumulate credit toward further qualifications throughout their careers. Third, the clear specification of outcomes means that higher-level qualifications are open to all, even those without access to a university – nurses can become doctors. This is the intrinsic logic of an outcomes-based qualification framework. However, it comes up against what was earlier referred to as the macro and the micro aspects of the institutional logic of modern societies.

Macro Aspects

First, the global economic changes have not turned out to be as progressive as predicted by political economists such as Piore and Sabel (1984) and Reich (1991). Majority of people in work still stay in the same field or sector for most of their working lives. The new industries of the e-economy create remarkably few jobs. Furthermore, some of the most characteristic jobs of the new economy are in call centers and the fast food and security industries, none of which requires many higher-qualified knowledge workers. In other words, the increasingly mobile and qualified society on which the claims for qualification frameworks are based bears little relationship to the realities of modern economies.

Second, a seamless framework of qualification levels does not by itself guarantee progression or provide a basis for overcoming deep-seated divisions between different types of qualifications. In England, there is a five-level vocational qualifications framework. However (1) the majority of vocational qualifications is awarded at levels one and two, (2) there are extremely few vocational qualifications at level five, and (3) of the few who obtain vocational qualifications at levels four and five, most have entered the framework as graduates from university (levels 4 and 5). These are examples that reflect the influence of what we referred to as the institutional logic of modern societies – in particular, the continued significance of academic/vocational and professional/vocational divisions.

The solution to overcoming these divisions is not primarily to be found through a qualification framework. It will depend on expanding the range of higher-level professional occupations and improving the institutional opportunities for people to progress from low-level

qualifications. It seems likely that such developments will be inhibited by extending the role of outcomes-based approaches.

The older progression routes from vocational to professional qualifications were largely restricted to engineering. A more recent example of progression to professional occupations is through the links between vocational qualifications for accountancy technicians and professional accountancy qualifications. The key factor in this case is the role of the professional body and its involvement in supporting and guaranteeing the quality of both professional and vocational qualifications.

The future policy tasks are not just to create new qualification levels within a single framework. They are: (1) to create genuine occupational progression pathways in sectors that have had few qualified people in the past, such as care, retailing, and hospitality; and (2) to create new access routes for people without initial qualifications. An outcome-based framework could support such developments was in place; it cannot lead them.

Micro Aspects

The micro aspects of the institutional logic of qualifications are less visible but no less fundamental in the problems they pose for reformers. They are expressed in the extent to the credibility and currency of a qualification is only partly based on what it records that the person who is qualified can do or knows (the outcomes). Of far greater significance is the trust that society in general and specific users in particular have in the qualification. This trust, as mentioned earlier, has in the past been embedded in various forms of specialized and usually elite community. These communities of trust have taken time to establish and have developed their own forms of exclusiveness and resistance to change. If these traditional communities of trust are weakened or destroyed (or in the case of many new sectors, do not exist), their function cannot be fulfilled by the specification of outcomes alone. If new communities of trust or some equivalents are not created to give substance and add practical experience to criteria, new outcomes-based qualifications will not gain credibility among users. This analysis may go some way to explaining (1) the persistence of old non-outcome-based qualifications and (2) the phenomenon of academic drift. Academic qualifications still rely on traditional communities of trust such as those linking specialist subject teachers with colleagues in universities and syllabi and have therefore retained their trust and currency value beyond their particular specialist communities.

Relying on qualifications based on old communities of trust is not an adequate solution in the fast-changing and increasingly competitive economic environment that we face today. As has already been noted, they have been exclusive and elitist and are rarely the basis for responding

to new skill and knowledge demands. It was partly these features that led to the development of alternatives based on transparent criteria. However, if interpreted in terms of criteria alone, transparency is a chimera. Learners always, at least to some extent, have to take on trust what they will learn in a program of study. Similarly, those who recruit and select draw on their experience as well as on someone's qualifications in making their judgments about someone's suitability; qualifications are never more than part of the data that they draw on. The traditional communities of trust were not planned – they emerged. New communities of trust in, for example, care, retail, and customer service occupations will have to be created, but not by employer-led organizations with the responsibility for National Occupational Standards. In sectors with a long tradition of apprenticeships linked to college-based programs, employer-led sectoral organizations can continue to take the lead because the old communities of trust can be built on. However, in sectors that recruit largely unskilled labor and where profits can be made by selling low-quality services or low-quality products (Keep, 1997), a quite new and more complex approach is needed that does not rely on employers generating demands for more qualified people. It is likely to involve a new type of leadership role for colleges and universities.

The implication of this analysis is to recognize the relatively modest role of outcomes-based frameworks in educational reform. The major tasks will be creating new forms of production and service that require new qualifications and establishing the new communities of trust that will be needed to underpin them. It is far from clear what the basis of new communities of trust will be or the roles of the various partners involved.

Conclusions

The two approaches to qualification reform that we have discussed in this article have origins in different historical times, and the very different roles of the state, employer organizations, and trade unions in the countries where they have been adopted or replaced. The new circumstances that most countries face today are, however, increasingly common. A qualifications strategy for the future cannot be based on the shared practices of occupational communities and their links with institutions because in many cases neither any longer exists. Conversely, outcomes-based approaches, which break with the traditional occupational communities and their links with educational institutions, take no account of how the trust on which qualifications rely is actually created and sustained.

The specification of criteria or outcomes is designed to make qualifications portable across different contexts. Many of the critiques of this process have shown how such specification is endless and trivializing. There is no

alternative to embedding criteria in shared practices as the condition for the trust on which qualifications rely for their credibility. The question becomes, what form might this embedding take when the identifiable occupational communities within which it took place in traditional qualifications no longer exist, or at least in only a very few sectors? We conclude this article with some suggestions from Gamble's research on cabinet makers (Gamble, 2006) in South Africa, as a way of linking the discussion of qualification reform in this article to the wider debates about the role of knowledge in education (Young, 2007).

Gamble (2006) makes two kinds of distinction between types of knowledge. One is between context-independent and context-dependent knowledge and the other is between types of context independence, principled and procedural. Traditional vocational qualifications relied on context-dependent knowledge located within occupational communities of trust. Gamble points out that in the specific case of cabinet makers, the context-dependent knowledge associated with the different tasks involved in making a cabinet are held together by a vision of the whole – the making of a complete cabinet. The vision of the whole is however context independent – it does not refer to a specific cabinet. At the same time, this vision is tacit, and neither codified nor codifiable; it is in the head of the master cabinet maker. Being un-codifiable, its theory implicit, it cannot be a model for future qualifications. It is, however, important, in pointing to the conditions for the development of future qualifications. Because the outcomes-based model rejects the idea of theory (codifiable context-independent knowledge embedded in specialist communities), the relationship between the qualification and its parts (the outcomes) is arbitrary. The learner acquires the procedural knowledge to make certain types of routine links between tasks. However, he or she has no basis for understanding the relationship between the whole and the parts or for progressing by taking his or her procedural knowledge to new levels of generality. Outcomes-based qualifications claim to be portable – a qualification can be used beyond the context in which it is acquired. However, if such claims are to be realized, learners need access to principled knowledge – concepts that are systematically related to each other (as in a theory) and embedded in contexts which give meaning to the theory. Separate criteria on their own are not adequate to define a qualification. In policy terms, the implications of these suggestions would of course involve a complete reorganization of existing relationships between professions, universities, colleges, employers, and qualification bodies.

Bibliography

Allais, S. (2003). The national qualifications framework in South Africa: A democratic project trapped in a neo-liberal paradigm? *Journal of Education and Work* 16(3), 305–323.

- DfEE (1998). *The Learning Age: A Renaissance for a New Britain*. London: Department for Education and Employment.
- Ensor, P. (2003). The national qualifications framework and higher education in South Africa: Some epistemological issues. *Journal of Education and Work* 16(3), 325–346.
- Gamble, J. (2006). What kind of knowledge for the vocational curriculum? In Mjelde, L. and Daly, R. (eds.) *Working Knowledge in a Globalizing World*, pp 61–81. Bern/Switzerland: Peter Lang AG/International Academic Pub.
- Jansen, J. and Christie, P. (1999). *Changing Curriculum: Studies on Outcomes-based Education in South Africa*. Cape Town: Juta.
- Jessup, G. (1991). *Outcomes: NVQs and the Emerging Model of Vocational Education and Training*. London: Falmer Press.
- Keep, E. (1998). Changes in the economy and the labour market: We are all knowledge workers now. in *Work and Education*, London: University of London Institute of Education (Post 16 Education Centre).
- Piore, M. and Sabel, C. (1984). *The Second Industrial Divide*. New York: Basic Books.
- Raggatt, P. and Williams, C. (1999). *Governments, Markets and Vocational Qualifications: An Anatomy of Policy*. London: Routledge.
- Raffe, D. (1992). *Modularisation in Initial Vocational Training: Recent Developments in Six European Countries*. Edinburgh: University of Edinburgh (Centre for Educational Sociology).
- Reich, R. (1991). *The Work of Nations*. London: Heineman.
- Wolf, A. (1995). *Competence-Based Assessment*. Buckingham: Open University Press.
- Young, M. (2004). National qualifications frameworks as a global phenomenon: A comparative perspective editorial introduction. *Journal of Education and Work* 16(3), 223–237.
- Young, M. (2007). *Bringing Knowledge Back in*. London: Routledge.
- Young, M. and Gamble, J. (eds.) (2006). *Knowledge, Qualifications and the Curriculum for South African Further Education*. Pretoria: HSRC Press.

Further Reading

- Ertl, H. (2002). The concept of modularization in vocational education and training: The debate in Germany and its implications. *Oxford Review of Education* 28(1), 53–73.
- Glanville, G. (2003). "Stop making Sense": Chaos and coherence in the formulation of the Irish Qualifications Framework. *Journal of Education and Work* 16(3), 259–270.
- Grugulis, I. (2003). *Skill and Qualification: The Contribution of NVQs to Raising Skill Levels, SKOPE Research Paper 36*. Cardiff: University of Cardiff.
- Keating, J. (2003). Qualifications frameworks in Australia. *Journal of Education and Work* 16(3), 277–288.
- Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Mansfield, B. and Mitchell, L. (1996). *Towards a Competent Workforce*. Aldershot, UK: Gower.
- Philips, D. (2003). Lessons from the New Zealand qualifications framework. *Journal of Education and Work* 16(3), 289–304.
- Raffe, D. (2003). 'Simplicity itself': The creation of the Scottish credit and qualifications framework. *Journal of Education and Work* 16(3), 94–109.
- Strathdee, R. (2003). The qualifications framework in New Zealand: Reproducing existing inequalities or disrupting the positional conflict for credentials. *Journal of Education and Work* 16(2), 147–164.
- Unwin, L., Fuller, A., Turbin, J., and Young, M. (2004). What determines the impact of vocational qualifications: A literature review. *Research Report 522*. London: Department for Education and Skills.
- Young, M. (1998). *The Curriculum of the Future*. London: Falmer.
- Young, M. (2003). National qualifications frameworks as a global phenomenon: A comparative perspective editorial introduction. *Journal of Education and Work* 16(3), 223–237.
- Young, M. (2004). National qualifications frameworks: Their feasibility and effective implementation in developing countries. *A Report prepared for the International Labour Office*. Geneva: ILO.

Quality Assurance in Vocational Education and Training

D Gunning, Welsh Assembly Government, Cardiff, UK

© 2010 Elsevier Ltd. All rights reserved.

Introduction

The article discusses the main risks associated with vocational education and training (VET) and the associated quality-assurance strategies that can be applied to each, pointing to examples of national and international developments. Generally, since the terms used in quality assurance vary considerably from one country to another and may also be used differently in technical and lay contexts (e.g., the word accreditation) this article has avoided technical terms wherever possible.

The scopes of VET and of quality assurance are worthy of a long and detailed consideration, particularly in an international context, since each is strongly linked to national cultural issues, but there is insufficient space in this article for such detail. For the purposes of this article, the scope of VET is considered to be that which prepares the learner for the world of work, but excludes general education of the kind associated with the school sector and excludes preparation for entry to work that requires a professional qualification obtained from an institute of higher education.

Although several chapters could be filled in the discussion of distinctions between quality management, quality control, and quality assurance, this article takes the scope of quality assurance to cover systems and procedures that are designed to ensure that the VET is fit for its defined purpose and that risks associated with VET are managed and minimized.

The scope of VET adopted here includes learning taking place at colleges such as the post-school institutes of technical and further education, at for-profit training organizations, at in-company training facilities, and in the workplace itself, using methods that range from traditional 'learn by doing' through to more recent e-learning. In this article, the term VET provider has been used generically to describe those organizations.

Key Risks

The risks associated with VET could be said to be located at three main levels. Each is discussed further in the succeeding sections.

The first level of risk is associated with the specification of the knowledge and skills that the learner is required to acquire. Whatever methodology is used to define the knowledge and skills – that is, the outcomes

of the learning – the outcomes must be appropriate to the purpose and must be defined in such a way that they can be reliably assessed.

The second level of risk is associated with the person or organization – referred to as the provider – whose responsibility it is to assist the learner to achieve the outcomes of the learning and to assess that the learner has achieved the outcomes. Although there is considerable literature on the issue of learner retention and success rates as a quality issue for providers, such as Hall (2001) and Martinez (2001), there is relatively little literature on other quality issues for providers, particularly compared to the corresponding volume for higher education and for schools.

The third level of risk operates beyond the boundaries of the learners and the providers and relates to the extent to which the VET system meets its purpose of providing a locality, region, state, or country with the knowledge and skills required to sustain a successful economy and vibrant communities and cultures. There is an associated transnational element where, as for example in the European Union, there are agreed transnational policies relating to mutual recognition of skills or free mobility of labor; in such circumstances, there is a risk that lack of compatibility of national VET systems prevent these transnational policies from being successfully implemented.

Key Risk 1 – Specification of Knowledge and Skills

The risk associated with the specification of the knowledge and skills can be managed through the development of systems that provide for expert input into the development of the specification, for example, through the use of employers to specify required job knowledge and skills; through the development of standardized approaches to the specification of knowledge and skills, for example, using a standard template, terminology; and through the use of level descriptors in national qualifications frameworks. Employer, and in some cases employee, involvement in the development of specifications of knowledge and skills is usually exercised through standard-setting bodies with similar functions – for example, industry skills councils in Australia, sector skills councils in the UK, and standards generating bodies in South Africa.

The way in which the knowledge and skills required to complete a course of VET is specified and assessed has

changed over the years. Specifications in the form of a syllabus – and assessments based on traditional written and practical examinations have given way in many countries to a modular structure, where each module (also called units and units of competence in some national systems) represents a discrete occupational element. In each module, the successful completion by the learner is assessed against specified outcomes, each with its associated assessment criteria. In many countries, which use this type of specification, the specifications are publicly available; the requirements for successful completion are therefore clear to both learners and assessors.

An early example of such a modular, outcomes-based system was developed in Scotland in 1983; the modular approach in Scotland is described in Gunning (2000). The approach was taken further in England in response to the Manpower Services Commission report (1986), which recommended changes to the English system to reduce the plethora of vocational courses available at the time and to increase the role of employers. The latter imperative was because of the view that the system had become dominated by what providers wanted to provide, and not by the skills that employers wanted. The English reforms proposed to adopt a single, employer-led specification of knowledge and skills for each occupation described as national occupational standards and controlled by a new regulatory body that would be responsible for the quality of these standards; the approach is described in Jessup (1991).

Many countries have since redeveloped their VET systems along similar lines to those initiated in England. National occupational standards are defined by bodies that represent employers, as the customers of the VET system, and are developed using a variety of approaches such as Developing a Curriculum (DACUM), which was developed in the US, and functional analysis, developed in the UK. In some countries, employers and trade unions are both involved in the specification of standards as social partners.

One consequence of the processes for developing nationally recognized, portable certificates based on standards set by employers was that the time taken to develop and quality assure the standards could be protracted. Such protracted development times have been seen, in recent reviews of systems of national occupational standards, to be a problem because they impair the systems' capacity to respond quickly to changing industry requirements, particularly in areas where technological change is rapid. See, for example, the reviews of the UK system by Beaumont (1996) and, more recently, by Leitch (2006) and of the Australian system by McDonald and Schofield (2004). Despite the identification of the problem, a solution has been difficult to find to date.

The quality assurance of national occupational standards – and other specifications of vocational knowledge and skills – should take account of two issues. First, do the standards encapsulate all the knowledge and skills

relevant to the occupation in question? Second, are the standards specified in the appropriate technical format and are they written with sufficient clarity for assessors to be able to make reliable assessment judgments on learners? Examples of guidance on standard setting can be found on the websites of national qualifications bodies or regulatory systems such as, for example, that for the Australian Quality Training Framework (AQTF 2007).

Where countries have developed a national qualifications framework that covers VET courses, these are usually defined by parameters of level (a measure of the complexity of the knowledge and skills acquired through a module or course) and the volume of learning (a measure of the notional time involved in acquiring the knowledge and skills in the module or course). In such frameworks, the level descriptors form a useful benchmark for the quality assurance of the specification of knowledge and skills; the quality assurance can check whether the knowledge and skills are at the right level and whether they are comparable to those in other modules or courses in other disciplines. Examples of such level descriptors can be found in the Victorian Registration and Qualification Authority's Credit Matrix and in the proposed European Qualifications Framework.

Where the system is managed by a national regulatory body, a further quality-assurance issue relates to the totality of vocational qualifications and national occupational standards in the national system. Is there coverage of all required occupations across the system of qualifications and standards and has unnecessary duplication been avoided so that the system is as simple as possible for employers, learners, and providers to navigate? Recent reviews of the system of vocational qualifications and standards, such as that led by Leitch for the UK Treasury (2006), continue to highlight problems of proliferation and of slow development times – these are issues which, despite the use of national occupational standards as the basis for VET for nearly two decades, remain to be resolved.

The specification of the outcomes of VET in the form of national occupational standards has also been adopted as a near-universal norm, especially in Anglophone countries that were redeveloping their VET systems. Besides the development of forms of national occupational standards as the basis for reform of VET and vocational qualifications in the UK, Australia, and New Zealand, for example, similar developments have taken place in southern Africa, as described by McGrath (2005).

Key Risk 2 – Providers

The prime risk associated with providers of VET relates to the assessment process. Has the assessment of the learner been carried out accurately against the requirements for success, however these are specified? If so, the

learner will have the knowledge and skills listed on his or her record of achievement or certificate; if not, the learner will either have had the knowledge and skills but been inaccurately assessed as not having them, or, perhaps more dangerously, will have been inaccurately assessed as having the knowledge and skills when the learner does not have them. That final scenario is particularly problematic when the certification of the body of knowledge and skills is seen as a license to practice in an occupation that has health and safety implications.

There are further risks associated with providers, one of which relates to resources. Particularly where the provider operates in a competitive environment, the provider may claim to be able to provide learning that leads to the knowledge and skills for an occupation but where the provider does not have the appropriate premises, equipment, or staff to deliver the learning and/or conduct assessment against the standards.

A further risk associated with providers relates to situations where governments allocate public funding to learning – for example, where the learners are unemployed youth. Has the provider managed the learning and assessment of learners in accordance with the terms under which public funding was provided?

These three risks – assessment risk, resources risk, and funding risk – are discussed in the sections that follow. For reasons of logical sequencing, the resources risk is discussed first because the first stage of quality assurance applied to providers is usually an approval process.

Some providers had pursued internal means to testify to the quality of the education and training they offered, drawing on practice in other industries. For example, in the 1990s, there was interest in the UK among some providers in seeking accreditation against the International Standards for Organization (ISO) 9000 standard, particularly since some businesses required that their supply-chain companies become certified against the standard. This was partly because providers thought that employers might see them as part of their supply chain and partly because providers saw ISO 9000 accreditation as giving them a competitive edge at a time when providers were being encouraged to operate in a competitive environment. The application of such quality standards in a UK context is discussed in Hammond (2005).

Resources Risk

One way to ensure that a provider has the required resources to enable learners to achieve the required knowledge and skill standards is for a regulatory body responsible for VET to operate a process of approval. An approval process will specify criteria that the provider has to be able to meet before any learners can be enrolled. Such criteria may be a combination of generic criteria related, for example, to management practice, and specific

criteria related to the particular occupations to which the learning is intended to lead. Examples of such criteria can be found on the websites of national qualifications or regulatory bodies such as the Scottish Qualifications Authority or the website of the AQTF 2007.

Such approval processes will measure the potential of the provider to comply with the approval criteria. In the case of a new provider, there may be little or no evidence of operation and the evidence available will therefore tend to be of documented management systems and records of staff qualifications and experience. In the case of an already-approved provider which is seeking approval to extend the range of learning it offers, there will be evidence of track record in other areas of learning and the approval process should therefore be able to concentrate on capacity to offer in the new area of learning.

To ensure that the provider's potential capacity is turned into practice, approval systems may be accompanied by a system of follow-up audits. These audits may be based on an assessment of the provider's experience and track record; thus, for example, priority for audit would be assigned to recently approved providers, those who have been found in previous audits to have ongoing compliance difficulties, or providers which had been the subject of learner complaints. Noncompliance discovered during such audits would lead to an action plan to recover compliance or, in the event of repeated noncompliance, to removal of the provider's approved status, either for specific areas of learning or in total.

The body responsible for approval of providers may be a national agency or an arm of government. Where it is not part of the national government, issues of mutual recognition can arise. In the UK, for example, a system had to be put in place to ensure that a provider approved by one awarding body had its approval status recognized by other awarding bodies; this system is described in QCA (2006). In the Australian federal system, where approval is the responsibility of the eight states and territories, a mutual recognition system is in place so that providers that operate in more than one state or territory do not have to undergo multiple approvals.

The approaches to providers adopted by regulators in the UK and Australia were, perhaps inevitably, over-detailed and based on a one-size-fits-all approach. More recently, following system reviews and the influence of regulatory trends promoted by writers such as Sparrow (2000) and by bodies such as the UK's Better Regulation Task Force (2005), which lists five principles of good regulation under a general title of 'less is more', there has been a move away from overburdensome, compliance-driven, process-orientated approval and audit systems by regulatory bodies toward a more constructive approach focusing on key outcomes. The basis of such a move is that previous systems had become orientated toward 'what can be measured is important', rather than 'measure what is important'.

In Australia, the new national quality system for providers includes an optional measure of excellence on the basis that it will influence learner and employer choice. The new Australian system is radically different from its predecessor; providers have welcomed moves to focus audit processes on key outcome measures such as student success rates and student and employer satisfaction rates as well as to reduce audit load on consistently compliant providers.

Funding Risk

Providers that are in receipt of public funding for the delivery of VET may have to satisfy quality criteria that focus on accountability for the use of that funding. Approval and audit processes will therefore concentrate on aspects such as financial and record-keeping systems and on issues related to public policy such as equity and inclusion. In Australia, public funding that supports the New Apprenticeship system is distributed by state and territory governments and these governments have quality-assurance processes that protect the use of such funds. This might involve audits of the accounts of providers and inspection of the delivery of specific elements of the New Apprenticeship system such as the setting up of individual training plans for each apprentice. A summary of the Australian regulatory and financing system can be found in Burke (2005).

In the UK, in the 1990s, the responsibility for distribution of such public funding was initially devolved to local arms of government; national providers and employers were again frustrated by the differences between systems adopted at local level. In addition, there was no mutual recognition system between the approval systems of the awarding bodies and the bodies dealing with accountability for public funding, even though the criteria often overlapped. Providers were therefore dealing with multiple audit systems, each with their own documentation.

In Scotland, an attempt was made to bring together the approval requirements of the awarding body; the quality systems operated by the two national enterprise bodies and their 22 local agencies responsible for public funds for training; and the requirements of the British Standard (BS)5750 and Investors in People. The resulting Scottish Quality Management System (SQMS) standards were applied to training providers seeking to offer courses which attracted public funding. An evaluation of SQMS by Marshall (2002) suggested that it had not met all of its ambitious aims and that providers' main justification for using it was because they had to do so, rather than because SQMS was a particularly effective quality scheme.

Assessment Risk

The provider is responsible for the assessment of whether or not a learner has provided sufficient evidence of

knowledge and skills against the outcomes specified. This requires that an appropriate method of assessment has been chosen and appropriate instruments of assessment have been designed; it also requires that the assessment instruments have been applied fairly and reliably to the learner's evidence of achievement. The competence of the provider's employees in the vocational area being assessed and in the practice of assessment is therefore a criterion applied in approval processes for providers. In the UK, these competences are defined by the national body responsible for standards in education and training.

In many systems, providers have assistance available to them in the form of guidance on assessment. For example, Australian national training packages, which provide the specifications for national occupational standards, also supply guidance for providers on assessment. Quality assurance of the design and appropriateness of assessment instruments might take the form of a peer-review process, for example, where the provider sets up a panel to conduct the review, perhaps made up of the provider's staff and local employers.

Further quality-assurance processes are required to ensure that the assessment judgments made by assessors are consistent, fair, and accurate. This can be through (1) the use of independent assessors (i.e., assessors who have no prior knowledge of the learner), or (2) the use of multiple assessors, or (3) a system of peer verification of the accuracy of the assessor's judgments of the learner's evidence of achievement. A further extension of independence would be introduced by having the verification of assessment decisions done by a verifier external to the provider; that externality would also raise the possibility that consistency of assessment judgments between providers can be carried out if the external verifier verifies a range of providers' assessments.

The assessment risk may be increased by a trend toward providers, even those which are public institutions, operating in a way that sees other providers as competitors. The risk is still higher if the provision of public funding to providers is linked to outputs rather than on throughputs – that is, where the provider receives all or most of the funding only if the learner is assessed as having successfully acquired the specified knowledge and skills.

In the UK systems of VET, providers are required to have a verification system in place through which to review the design of assessments and to review the assessment judgments of assessors. A review of such internal verification systems by Warmington and Wilmot (2004) examined the benefits to organizations whose employees undertook this role. There is also a further layer of external verification, carried out on a risk-based sampling approach by external verifiers employed by the awarding bodies which ultimately issue the record of achievement to a successful learner. Forms of internal and external verification require that the verifiers have specialist

knowledge of the vocational subject being assessed – the verifiers need that knowledge to be able to review learner evidence of achievement against the standards set.

By contrast, the Australian national quality system for VET is largely front loaded and does not involve a system of external verification. The absence of this quality-assurance process places greater importance on the process for approving and auditing providers. In the Australian system, it is assumed that all the specialist knowledge required of those who quality assure assessment rests in the provider – the approval and audit processes are conducted by external auditors appointed by the state or territory regulator for their audit expertise, not because they are specialists in the vocational subject being offered by the provider. Clayton *et al.* (2004) provide an analysis of the implications of the Australian approach to quality assurance for assessment decision making.

Key Risk 3 – The VET System

The third level of risk relates to the extent to which a VET system meets its purpose of providing a locality, region, state, or country with the knowledge and skills required to sustain a successful economy and vibrant communities and cultures. There is an associated risk, where there are transnational policies relating to mutual recognition of skills or free mobility of labor, that lack of compatibility of national VET systems prevents these policies from being successfully implemented.

The management of the fitness for purpose risk may involve formal links between the VET system and the wider skills or economic development strategies adopted by a locality, region, state, or country. The risk associated with transnational policies, especially in a context such as that of the European Union, is usually managed in an environment where individual governments have autonomy; hence approaches to risk management will tend to emphasize voluntary or peer-pressure-driven compliance with transnational frameworks rather than the use of regulatory instruments.

VET serves many different purposes. It can provide generic education and training that allows people to develop their employability skills for entry to the workforce – for example, school leavers and people who are unemployed for a long term. It can provide job-specific education and training that prepares an individual for a specific trade or profession – for example, through apprenticeships or technicianships. In addition, it can provide education and training to those already employed – for example to up-skill to keep pace with changes in technology.

It is beyond the scope of this article to discuss this issue further or to discuss the quality-assurance consequences other than to note that national and regional skills strategies and the setting of conditions on the funding of VET are

forms of quality assurance to ensure that the supply of VET is well matched to customer demand. For examples of such strategies, see Victorian Government (2006) and Welsh Assembly Government (2007).

The aim of such strategies is typically to provide a structure and direction to the VET system, especially in relation to the use of public money to buy training. Such strategies may provide an overall approach, such as moving toward a demand-led rather than provider-led VET system, and may determine ways in which the use of public money is directed, for example, toward defined local and national skills priorities or toward particular training schemes such as apprenticeships. The extent to which VET systems can be demand driven is largely dependent on the extent to which labor market intelligence can accurately predict current and future skills needs and on the extent to which learner choice can be directed toward courses that meet those needs. Keep (2007) argues that there may be inherent tensions between demand-driven systems and the tendency of governments to set national targets and between short-term fluctuations in the labor market and the lead time required by providers to change the courses they offer.

A further large-scale quality-assurance issue arises when individual nations work together to create transnational economic communities and to encourage mobility of labor across national borders. Another form of transnational collaboration occurs when one nation assists another less-well-developed country to build up its national infrastructure, perhaps including its system of VET.

In the case of infrastructure building, there have been international collaborations on VET at various levels. International organizations, such as the United Nations Educational, Scientific and Cultural Organization (UNESCO)/UNIVOC, have identified the importance of quality assurance in VET; UNESCO (2006) provides guidance on the setting up of quality-assurance systems within a VET system.

The development of an economic community of nations and labor mobility is furthestmost advanced in Europe, where the free movement of labor has been seen as a key driver of the development of a genuinely borderless European community. An important prerequisite to such mobility is the recognition of qualifications so that, for example, an electrician trained and qualified in Italy can practice in Ireland without the need to retrain.

Recent work in Europe has been concentrated at a systems level with the aim of making those systems more compatible, within the overall European policy of national subsidiarity. At the meeting of the European Council in Lisbon, the European Union recognized the importance of a radical transformation of its education systems, with 2010 as the target date. This led to a declaration by its ministers of education and training (the Copenhagen Declaration) that a set of priorities would be pursued which

included the promotion of cooperation in quality assurance with particular focus on exchange of models and methods, as well as common criteria and principles for quality in VET.

More recently, the European Union has established the European Forum on Quality in Vocational Education and Training as part of the actions arising from the Copenhagen Declaration; the Forum has drawn up detailed work programs. One of the early aims is to work toward a common quality-assurance framework for VET, including a core set of quality criteria, a set of quality indicators, and a guide to self-assessment for providers. However, initial progress toward the Lisbon 2010 objectives was slow; a further declaration in 2006 by ministers of education and training in Helsinki included renewed emphasis on VET, including the development of common tools for qualifications, credit transfer, and quality assurance.

These developments will have maximum impact on regulatory and qualifications bodies in the individual countries rather than directly on providers. Those bodies may be expected through time to ensure that their quality processes and procedures are compatible with the European framework rather than being expected to adopt the common framework. Over time, this will lead to harmonized quality-assurance systems and, therefore, to greater mutual trust and recognition between countries.

Conclusion

Quality-assurance systems play an important part in the lives of most people who work in, or are in contact with, the VET system – employers as customers or as standard setters, provider employees as teachers and assessors and, last but not least, learners whose pathway to employment depends on the VET system working fairly and effectively. The scale of the VET system and the complexity of the interactions within it are a far cry from the largely local interactions of the traditional apprenticeship systems whose origins date back hundreds of years.

Given the importance of the VET system to those people and the very large sums of public money invested in the system by many governments, it is perhaps surprising that the body of research literature on quality assurance in VET is relatively sparse, especially when compared to that for school and higher education.

Another striking feature of the VET system is the way in which a single model, based on national occupational standards and national qualifications frameworks, has been adopted across a wide variety of countries. If there is a trend in quality assurance within such VET systems, it is perhaps toward more differentiated, outcomes-focused models because these allow high quality to be rewarded and resources targeted on poorer performance.

Bibliography

- Beaumont, G. (1996). *Review of 100 NVQs and SVQs: A Report to the Department for Education and Employment*. London: National Council for Vocational Qualifications (NCVQ).
- Better Regulation Task Force (2005). *Regulation – Less is More: Reducing Burdens, Improving Outcomes*. London: Cabinet Office.
- Burke, G. (2005). Financing and regulating vocational education and training (VET) in Australia. *Paper Presented at International Seminar on Employment / Unemployment Insurance*, Seoul 2005, Korean Labour Institute in cooperation with the Korean Ministry of Labour.
- Clayton, B., Roy, S., Booth, R., and House, R. (2004). *Maximising Confidence in Assessment Decision-Making*. Adelaide: National Council for Vocational Education Research (NCVER).
- Gunning, D. (2000). *A Modular, Outcomes-Based Qualifications System – the Scottish Experience in Comparability of Vocational Qualifications Systems*. Eschborn: Deutsche Gesellschaft für Technische Zusammenarbeit.
- Hall, J. C. (2001). *Retention and Wastage in FE and HE*. Edinburgh: Scottish Council for Research in Education (SCRE).
- Hammond, M. (2005). Quality management or quality confusion? Are quality management techniques in FE actually preventing the development of transformational leadership? *Paper Presented at Conference on Rethinking Leadership – New Directions in the Learning and Skills Sector*, Lancaster University. <http://www.leeds.ac.uk/educol/documents/141488.htm> (accessed July 2009).
- Jessup, G. (1991). *Outcomes, NVQs and the Emerging Model of Education and Training*. London: Falmer.
- Keep, E. (2007). Skill shortages in the UK – issues, problems and ways forward. *Paper Presented at a Seminar Organised by the Centre for the Economics of Education and Training*, Melbourne. <http://www.education.monash.edu.au/centres/ceet/docs/seminars/skill-shortage-uk.ppt>.
- Manpower Services Commission / Department of Education and Science (1986). *Review of Vocational Qualifications in England and Wales*. London: HMSO.
- Marshall, G. H. (2002). The Scottish quality management system: An interim evaluation. *Managerial Auditing Journal* **17**(5), 251–260.
- Martinez, P. (2001). Effectiveness and improvement: School and college research compared. *Paper Presented at the Annual Conference of the British Educational Research Association*, Leeds University.
- McDonald, R. and Schofield, K. (2004). *Moving on – Report of the High-Level Review of Training Packages*. Brisbane: Australian National Training Authority (ANTA).
- McGrath, S. (2005). *The Multiple Contexts of Vocational Education and Training in Southern Africa in Vocational Education and Training in Southern Africa – A Comparative Study*. Cape Town: HSRC Press.
- QCA (Qualifications and Curriculum Authority) / Welsh Assembly Government (2006). *NVQ Code of Practice*. London: QCA.
- Sparrow, M. K. (2000). *The Regulatory Craft*. Washington, DC: Brookings Institution Press.
- Victorian Government (2006). *Maintaining the Advantage: Skilled Victorians*. Melbourne: Victorian Government.
- Warrington, P. and Wilmot, J. (2004). Re-inventing the wheel? Divergent conceptions of the role of internal verifiers within organisations. *Journal of Vocational Education and Training* **56**(2), 307–328.
- Welsh Assembly Government (2007). *Skills that Work for Wales: A Skills and Employment Strategy*. Cardiff: Welsh Assembly Government.

Further Reading

- Blum, K. and Meyers, D. (2003). *Quality Indicators in Vocational Education and Training*. Adelaide: NCVER.
- Research Association. London: Learning and Skills Development Agency (LSDA).
- SAQA (2000). *The National Qualifications Framework and the Standards Setting*. Pretoria: South African Qualifications Authority (SAQA).

Treasury, H. M. (2006). *Prosperity for all in the Global Economy – World Class Skills*. Norwich: HMSO.

UNEVOC and the Scottish Qualifications Authority (2006). The development of a national system of vocational qualifications. A discussion paper. Bonn: UNEVOC.

http://europa.eu/index_en.htm – Europa: Gateway to European Union Information.

<http://www.lluk.org> – Lifelong Learning UK.

<http://www.ntis.gov.au> – National Training Information Service (Australia).

<http://www.sqa.org.uk> – Scottish Qualifications Authority.

<http://www.vrqa.vic.gov.au> – Victorian Registration and Qualifications Authority (VRQA): The Credit Matrix (Australia).

Relevant Websites

<http://www.training.com.au> – Australian Quality Training Framework.

Training Finance

A Ziderman, Bar-Ilan University, Ramat Gan, Israel

© 2010 Elsevier Ltd. All rights reserved.

Introduction

A central theme of this article is that the system of training finance (and, in particular, its constituent financing mechanisms) has a twofold purpose: not only does it supply funding for the national training system, but it also has a central role to play in achieving many of the broader goals of national training policy. The article opens with a schematic account of training-finance flows as a whole in conventional training markets. This is followed by a discussion of innovative financing mechanisms that have been adopted in many national training systems aimed both at correcting shortcomings of conventional training finance systems in terms of meeting labor-market needs and containing training costs and also to better meet the challenges of competition, technological change, and globalization. Directions of change include a greater diversification of funding sources for skills development (including enhanced cost recovery, cost sharing, and training taxes), encouragement of more and higher-quality enterprise training, the development of private training markets, and increased competition between public and private training providers. The final section presents a schematic account of finance flows that integrates these innovative financing mechanisms, within a unified system, to facilitate the desired development of integrated, competitive, demand-driven training markets.

There is relatively little discussion in this article of the financing of vocational education provided within the secondary schooling system. While vocational education is a major component of Vocational Education and Training (VET) in many countries, with some exceptions (notably, some Latin American countries), the central financing issues relating to vocational education are common to the schooling system generally and, therefore, no special discussion is merited here.

This article draws extensively on earlier writings of the author, particularly Ziderman (2003) to which the reader is referred for a more extensive discussion of the major issues treated in this article.

Finance Flows

Conventional, Fragmented Training Markets

We turn first to conventional, fragmented training markets. A schematic overview is presented in **Figure 1**; institutions that provide training are indicated by the shaded boxes.

Training may be provided through private training markets either by firms or in proprietary training institutions; it may also be provided within the public sector at public training institutions. However, the training market is fragmented into two distinctly differing sectors – private and public.

The major training providers in the private sector are enterprises and proprietary training institutions. The private training sector is market driven, nonsubsidized, and (usually) competitive. Firms (in both the informal and formal employment sectors) provide training to trainees/workers in their employment; payment is made by the employee/trainee through initial lump-sum fee or implicitly in the form of low, below-productivity wage levels. Individuals enrol in pre-employment courses, at full fees, at private training institutions; training fees for continuing training may be borne partially by formal sector employers.

Usually, public training systems have been established, in parallel; they constitute the leading supplier of structured, pre-employment training, frequently dominating the market as a provider of formal sector training. Public sector training institutions are, predominantly, financed from government budgetary allocations. Where fees are current, they are set at purely nominal levels and often accrue to the government rather than remaining with institutions; thus, there is little incentive for training institutions to develop market-demanded courses that could generate fee income. Firms may enrol their workers for continuing training courses, provided at full cost but more usually at subsidized fee levels. Because budget allocations to public training providers are usually unrelated to objective, outcome measures – such as success in placing trainees in productive employment – there is little incentive for institutional training providers to align training courses offered with the needs of the labor market. Linkages between public training centers and formal sector employers remain poor; training provision is dominantly supply driven. Moreover, training centers do not develop training programs focusing on the particular needs of informal sector employment nor do they cater well to the special needs of minority and disadvantaged groups.

A number of emerging trends have rendered this conventional financing framework increasingly outmoded. Ongoing technological change, structural adjustment policies, new and changing patterns of trade and competition, and globalization have combined in many countries to create the need for a much more flexible and responsive

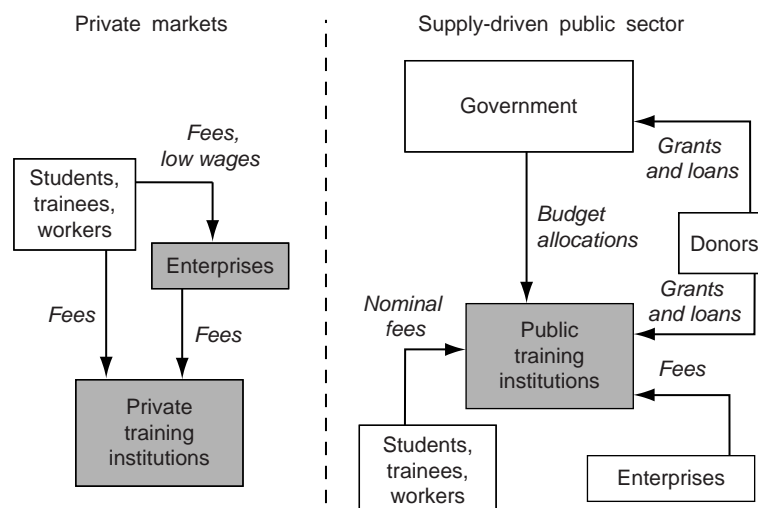


Figure 1 Finance flows: fragmented training markets. Adapted from Ziderman, A. (2003). *Financing Vocational Training in Sub-Saharan Africa*. Washington, DC: The World Bank.

training system than has been manifest in more protected regimes in the past. An increasing problem in many training systems is a tendency to market failure, with firms undertraining in transferable skills, both in terms of the amount of training provided and its quality. The consequent shortages of well-trained, skilled workers in the formal sector, may stunt productivity growth, competitiveness, and industrial development.

This conventional financing framework has become inadequate to meet society's skill development needs. Public training provision remains essentially supply driven. It is not subject to the discipline of competition with other training providers, nor are guidance mechanisms in place to match the skills supplied by public training institutions with the skill needs of the market. In addition, the system is fragmented; in particular, private training institutions do not operate within the same financing framework as public sector training providers.

Innovative Financing Mechanisms

Thus, in many national training systems a strong trend away from these traditional forms of training finance and provision is in evidence. The driving force behind these moves has been the increased intervention of the state in training markets; paradoxically, this process has involved a retreat by the state in the financing of training. Increasing pressures on government budgets in general, and on public training budgets in particular, have led to a search for additional or alternative sources of funding for training. In addition to the need to tap nongovernment sources of funding, governments have intervened more strongly in training markets in order to counter shortcomings of conventional private training markets and, notably, a tendency for enterprises to undertrain.

Augmenting funding for training

A central feature, common to virtually all training systems, is the pressing need to augment the total amount of funding for public sector training, in the light of a paucity of government funding – in part the result of the adoption of structural adjustment policies and increasing calls on government funding from competing sectors. The response is greater funding diversification: seeking alternative or additional funding for public training from other sources.

Funding diversification can take many various forms; in particular, four different avenues may be pursued, separately or in combination.

Fund augmentation

Public sector funds available for the support of training institutions through subventions may be augmented from other sources. Earmarked training taxes, usually levied on the payrolls of enterprises, have emerged as the most widely adopted alternative to central government budgetary allocations for training. Payroll levies of this type were first introduced in Brazil in the 1940s, spread widely to other countries in Latin America (a recent discussion is provided in Galhardi, 2002) and has been adopted by training systems in many other countries. Levies are usually set at between 1% and 2% of the total wages bill of the enterprise; proceeds are used mainly to support public sector training provision, with the emphasis on initial training at formal public training institutions. Training levies can constitute a stable and protected source of funding for national training provision. However, the expectation that levy income would complement existing government financing, thus providing an additional source of funding, has not always realized in practice, and levy income has displaced government subventions for training. There are also notable cases of the opposite tendency,

where earmarked training taxes are absorbed into general government revenues rather than being used for the financing of public training. A fuller treatment of payroll levies is provided in Ziderman (2008).

Governments in developing countries may turn to donor institutions to provide funding, either to the government or directly to individual training institutions; in some country settings, donor funding constitutes a very important finance source for public sector training institutions.

Cost sharing

The weight of training finance falling on public funds may be lightened through the introduction of cost sharing with the beneficiaries of training. Unlike fund augmentation, which results in a larger funding pool, cost sharing aims at reducing the size of allocations to individual training institutions. The best-known and most widely used method is the imposition, or raising the level, of user fees to trainees or students enrolled in training courses. These measures allow a reduction in public subsidies for training or the provision of more or better training services with given levels of public support.

A central issue in fee policy is whether a regime of standard, nationwide compulsory fees should be instituted or freedom should be accorded to individual training institutions to fix the level of fees, overall and differing by type of training course. Institutional autonomy in the setting of fees represents the more desirable approach; it will encourage training providers to develop a more dynamic, even aggressive, approach to exploiting the potential of the local market environment. It is in this manner that institutional fee policy becomes more than a device for cost recovery and cost sharing: in providing a mechanism for varying fee levels across courses and client groups it serves as a tool for moving the training system toward an environment characterized by open, demand-oriented training. However, the voluntary setting of user fees may not be feasible in the otherwise centralized training systems; standard, compulsory fee setting is generally acceptable as the second-best measure for reducing pressures on public budgets.

The positive financial benefits from greater cost recovery need to be examined alongside the potentially adverse effects on equity. There is a clear trade-off here. Higher, realistic fees will exclude from training those who are unable to pay; fees set at comfortably low levels will fail to make a sizable contribution to cost recovery. In particular, negative impacts on the access to training opportunities of the poor, minorities, rural populations, and other disadvantaged groups are likely to ensue. This risk points to the widely recognized need to introduce targeted subsidies directed to these at-risk groups, in the form of scholarships and reduced fees.

Institutional income generation

Income generated from the sale of production and service activities of trainees can constitute a useful form of additional institutional income. Income may be derived as a byproduct of the training process itself. But it is possible, more purposefully, to utilize available skills and facilities to produce output for sale in the local market; indeed, exposure to local markets may lead to more relevant, market-oriented training. Here, the issue is one of maintaining a healthy balance between these two activities. As more weight is given to instruction, the income potential from production declines; alternatively, quality of training will suffer as emphasis is placed on production rather than on instruction. Training institutions may also generate income from the sale of services, including the renting out of underused facilities and providing consulting services to local enterprises.

Private training provision

The growth of private training institutions, with trainees paying full costs, represents a pathway for expanding the national training system without heavy commitments of public funds. Indeed, the encouragement by government of private training institutions' development, through subsidies and nonmonetary means, may represent an effective way of both generating additional funding for training and, in parallel, reducing the call on public funds. Thus, reduced public training provision could be possible (and concomitant budgetary reductions) with the reduction in public training supply made up by compensating expansion of private training institutions. In many countries, the lack of private training provision is the result of various constraints that hold back the development of private training institutions. These are: financial constraints, issues of rigid fees policy, excessive regulation of private institutions, and information gaps, and these are detailed below:

- To offset a lack of capital resources, especially for high-cost industrial and technical courses, governments may offer development loans or subsidies, particularly in strategic skill areas, to assist these firms in their start-up phases.
- Imposed tuition fee ceilings, while aimed at protecting trainees from exploitive activities by private training institutions, may too rigidly limit the ability of these institutions to enter new training markets, especially those with high investment and recurrent costs.
- Private training institutions are unlikely to flourish in an overly strict regulatory environment. Regulation and enforcement should be sparing; while sufficiently robust to counter dishonest practices and low-quality training, they should be designed to encourage private training institutions to operate fairly and efficiently within a facilitating, regulatory environment.

- Without reliable information, consumers are unable to make wise and informed choices. Information on both the quality and stability of private training institutions is often lacking; this may be provided by government in the form of updated information on the relevance of courses to labor-market demands and job opportunities.

These various approaches are set out in **Figure 2**. The first three approaches act directly in bringing in additional revenues to the training sector, while the fourth one affects training budgets only indirectly. Of the methods of direct funding augmentation, the first one increases the size of the funding pool available for distribution to training institutions but there is no immediate effect on the income of individual training institutions. Diversification options are not alternatives and all four avenues are often explored simultaneously.

Encouraging enterprise training

A second major reason for government intervention in conventional training markets is a corrective one: to encourage formal sector enterprises to provide more and better training. Governments subsidize enterprise training, either directly from central government budget appropriations or, less usually, from specially designated training funds, also financed (fully or in part) by government (see below). However, tight public budgets may limit the government's ability to subsidize enterprise training from public funds.

Levy-Grant Schemes

Levy-grant schemes, based on payroll taxes, have provided governments in many countries with an alternative mechanism for promoting company training. Unlike the

revenue-generating rationale for payroll levies discussed above – where the revenues from training levies are earmarked to finance public sector training institutions – levy-grant schemes are directed toward training provided by enterprises. While many variants are found in terms of actual practice, the common feature of levy-grant schemes is the provision of incentives for firms to invest in more and better in-service training. Thus, payroll levies are often linked to reimbursement mechanisms, whereby firms receive payments related to the amount of designated forms of training they provide. Firms are encouraged to invest more in the skills development of its workforce, be it in the sphere of on-the-job training (setting up or extending and improving existing company training) or by sending workers to train externally. The need for government intervention, by the introduction of levy-grant arrangements, arises because of shortcomings in the amount and/or quality of enterprise training.

While there are numerous variants, a threefold classification of levy-grant schemes (Gasskov, 1994) has been widely adopted – cost reimbursement, cost redistribution, and levy exemption, which are explained below:

- Under cost reimbursement, firms receive grants on a cost-incurred basis, for certain designated forms of training (both on and off the job). The purpose of these schemes is often misunderstood, particularly among employers; the scheme aims not at reimbursement of the levy as such but rather reimbursement of training expenditures incurred (to encourage firms to train more or better). Thus, a training expenditure reimbursement ceiling (for firms that train to acceptable standards) is usually set, up to a given percentage of the levy paid.
- Designed in particular to deal with the ill-effects on training supply of the poaching of skilled workers by nontraining firms, a cost redistribution scheme redistributes the burden of training expenditures among enterprises away from companies who do not train, toward those who do. Training companies may receive grants far in excess of the amount of levy paid, providing strong incentives for firms to train.
- Levy exemption is usually employed as part of broader cost-reimbursement scheme. This exemption allows firms, adequately meeting their training needs, to withdraw from the levy-grant system or at least to benefit from reduced levy assessments. A major advantage is freeing firms from the bureaucratic fatigues of levy payment and subsequent grant claim – potential cash-flow problems are avoided. This mechanism is found more typically in industrialized economies.

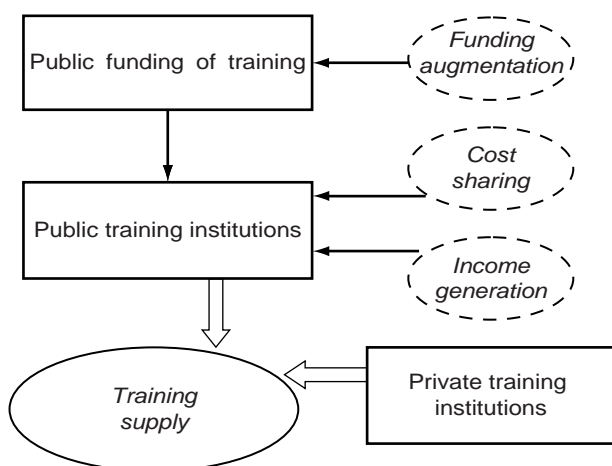


Figure 2 Funding diversification. Adapted from Zideman, A. (2003). *Financing Vocational Training in Sub-Saharan Africa*. Washington, DC: The World Bank.

While national payroll levies are the best-known, and most widespread, form of training taxation, a number of countries have preferred to employ forms of training

taxes. These training taxes tend either to be sector specific or are introduced to finance a fairly narrow training outcome. Sectoral training levies may be based on value of turnover, output, and value of contracts or employment, rather than on company payrolls. The main advantage of sectoral levies is that they offer a means of tailoring the levy format to the specific characteristics and needs of the sector in question. The disadvantage of sectoral levies is their narrow focus obviating an integrated, national approach to the finance and planning of skills development. Thus, most countries have preferred to employ standard, national-level levy schemes, based on enterprise payrolls

There are some limitations of levy-grant schemes. Given particular training needs, many firms, particularly small ones, do not benefit from the levy-grant scheme; this breeds resentment, opposition, and compromises the status of training levies as a form of benefit taxation. Firms may come to regard the scheme as just another tax. For these reasons, stakeholder ownership in the operation of levy-grant schemes – particularly with regard to disbursement policy – should be assured by employer representation levy-grant scheme management boards.

Earmarked taxes in general may be seen not to conform well with the principles of sound public finance and weaken attempts to unify the national tax system. Payroll levies raise the cost of labor to the employer, possibly discouraging employment. Alternatively, employers may shift the incidence of the levy on to workers in the form of lowered wages; in this case, workers, not the employers, bear the burden of the tax.

We have noted the usual classification of national payroll tax schemes into two distinct groups, reflecting very different underlying objectives: revenue generation schemes (where levy proceeds are used to finance training provided by public sector institutions) and levy-grant schemes (aimed at encouraging training investment by firms themselves). However, this traditional dichotomy is becoming somewhat outdated as evolving levy schemes begin to take on a broader range of tasks, particularly in the context of the development of national training funds and training authorities.

Alternative Forms of Subsidy

Where firms undertrain, training incentives may be justified to encourage firms to train. Apart from training-cost reimbursement that is part of a levy-grant system (as discussed above), direct subsidies may be provided out of public funds to encourage enterprise training; or indirect subsidies may be offered through concessions on company tax obligations, for firms that train.

Levy-grant systems have some clear advantages over the two alternative incentive systems: direct government

subsidy payments and concessions on enterprise tax obligations. A major advantage of levy-grant systems is that they do not draw on public funds, a point of some importance in times of parsimonious government budgets; in addition, they can lead to a more systematic, structured approach to training, rather than a more *ad hoc* one. The central lacunae in training under-provision is not only that amount of training provided is too low, but rather that it is often piecemeal and not sufficiently well integrated.

The response to direct and indirect subsidy schemes may be low, if they are insufficiently focused to catch the attention of senior management. However, in the case of levy-grant schemes, involvement is assured automatically by the compulsory payment of the levy. The disadvantages of tax concession schemes have militated against their adoption barring a very few countries: they require a well-developed and broadly based system of corporate taxation, often lacking in developing countries and responsiveness of firms may be low where few firms earn sufficient profits to benefit from tax exemptions.

Training funds

One outcome of the introduction and spread of earmarked training levies has been the development of a relatively new type of financing mechanism: the national training fund. These training funds usually constitute not only the depository of collected training levies but also the mechanism for their distribution. Government budgetary allocations may supplement levy income to the Fund, or represent its major income source; donor support is important in some cases. The intention is to provide a sheltered funding source for national training development, including the financing of public sector training, the provision of incentives for enterprise training, and meeting the skill needs of special groups. Training funds usually operate outside normal government budgetary channels; thus, they are more readily accessed and may be utilized more flexibly than would be the case normally with direct government-financed training programs. However, operating as they do under varying degrees of autonomy from government control, a fund's freedom of maneuver may be constrained. Thus, in cases where the degree of independence from Ministry of Labor and Treasury control is limited, the funds may often emerge as conservative, reactive bodies, rather than adopting a proactive, independent stance in fund policy and management.

National training authorities

In many countries national training agencies or authorities (NTAs) form the linchpin of the financing system. NTAs may be attached, with varying degrees of autonomy, to a government department (usually the labor ministry); however, they are likely to operate more effectively

as largely autonomous bodies forming a buffer between government and the training system. They are usually run by boards representing the training system's major stakeholders. While most NTAs receive general government funding, a large number are financed solely or in addition, by payroll levies. NTAs are much broader in scope than training funds; they are usually empowered with a wide range of national training responsibilities. In addition to managing the system of enterprise training subsidies and, where levy-grant systems are in place, levy reimbursements, they may be charged with responsibilities for developing national training policies and standards, planning the national training system, accreditation of institutions, and generating and disseminating relevant labor-market information.

NTAs may be better placed than environmentally constrained government departments to operate payment mechanisms for training institutions in ways which promote efficiency and competitiveness in training markets. This is discussed below.

Funding allocation to training institutions

Formula funding of public training institutions (such as output-related funding) provides one example of such measures. The allocation among institutional training providers of the total government budget for training (or of the national training fund) is a major component of the training financing system in most countries. However, and particularly in developing countries, a clearly formulated, objective disbursement policy is lacking. A much needed reform is the adoption of moves toward the gradual dismantling of the arbitrary, *ad hoc* institutional core funding

arrangements in place, and their gradual replacement by objective funding formula, such as those related to inputs, outputs, and outcomes. Based on competitive tender that is open to public and private training institutions, contracted training (particularly the needs of disadvantaged groups) can, through the bidding process, both integrate training markets and pressure public training institutions to be more efficient and to operate at lower cost. In line with this approach, donor agencies would fund the NTA only, not individual training institutions.

These reforms are important because the mechanism through which government transfers funds to training institutions has an important effect on the way in which this funding is used and on institutional behavior more generally. An inherent shortcoming in the transfer mechanisms currently used is that they promote low internal efficiency of training institutions and a strengthening of supply-driven training provision. An important task of funding disbursement policies is to provide an appropriate mix of regulation and incentives to ensure that public training can hold its own in an environment of competitive training markets

Integrated, Competitive, Demand-Driven Training Markets

Most industrialized economies have adopted some or all of the financing mechanisms surveyed in the previous section; strong moves to reform the training financing system in this direction are evident in many developing and transition economies. However, the value of these

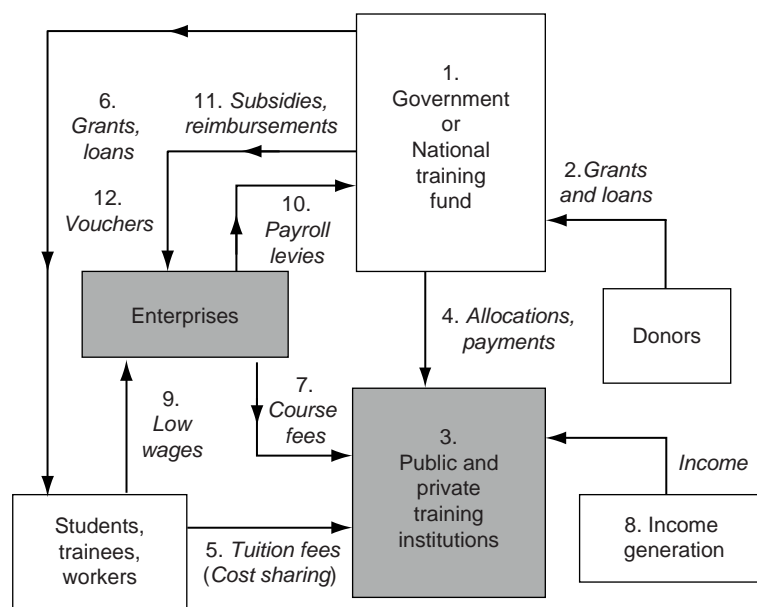


Figure 3 Finance flows: integrated, demand-driven training markets. Adapted from Zideman, A. (2003). *Financing Vocational Training in Sub-Saharan Africa*. Washington, DC: The World Bank.

reforms lies in their adoption in combination as an integrated system, rather than piecemeal. Innovative training finance mechanisms should be fashioned as an integrated system whose central role is to encourage and facilitate the transformation of fragmented, inefficient training systems (with underfunded, supply-driven public provision) into an integrated competitive, demand-oriented training system. The financing flows, in such an integrated framework, is set out in schematic form in **Figure 3**.

Government funding of training institutions is made either direct or through a national training fund (or NTA), as shown in (1). Where available, donor funding is supplied centrally to the government or training fund and not to individual institutions (2). Core financing of public training institutions is based on objective formula funding (3); contract financing of designated government programs (such as for the unemployed) are made available to both public and private and training institutions on a competitive basis (4). Greater cost sharing is introduced in public training institutions, through augmented course fees closer toward competitive levels (5); this is facilitated by the availability of selective scholarships for the poor or student/trainee loans (6). Additional income for public training institutions derives from fee payments for tailor-made courses for firms (7) and from income-generation activities (8). On-the-job training within firms is partially financed by workers (Becker, 1964) through low wages (9). Payroll levies on firms (10) may be used to either augment national funding for training or may constitute part of a levy-grant scheme to encourage enterprises to train more (11). Vouchers schemes, although still largely experimental, may enable potential trainees to purchase training in the open market, with public and private training providers competing for trainee enrolments (12). Vouchers, like grants, typically do not lighten the financing

burden falling on the funding body; training remains subsidized and cost recovery is not an integral part of a voucher scheme.

Bibliography

- Becker, G. S. (1964). *Human Capital*. New York: National Bureau of Economic Research.
- Galhardi, R. M. A. A. (2002). *Financing Training: Innovative Approaches in Latin America*. Geneva: International Labour Office.
- Gasskov, V. (ed.) (1994). *Alternative Schemes for Financing Training*. Geneva: International Labour Office.
- Ziderman, A. (2003). *Financing Vocational Training in Sub-Saharan Africa*. Washington, DC: The World Bank.
- Ziderman, A. (2008). Financing training by payroll levies. In Maclean, R. and Wilson, D. N. (eds.) *Handbook of Technical and Vocational Education and Training*. Dordrecht: Kluwer Academic Publishers.

Further Reading

- Frank, J. (2005). *Financing Technical and Vocational Education and Training*. Bonn: InWEnt.
- Middleton, J., Adrian, Z., and Adams, A. V. (1993). *Skills for Productivity: Vocational Education and Training in Developing Countries*. New York: Oxford University Press.
- Tsang, M. (1997). The cost of vocational training. *International Journal of Manpower* 18(1/2), 63–89.
- Whalley, J. and Ziderman, A. (1990). Financing training in developing countries: The role of payroll taxes. *Economics of Education Review* 9 (4), 377–387.
- World Bank (1991). *Vocational and Technical Education and Training: A World Bank Policy Paper*. Washington, DC: The World Bank.

Relevant Website

<http://www.cedefop.gr> – European Centre for the Development of Vocational Training (CEDEFOP), European Training Village.

Training Markets

P Ryan, King's College, Cambridge, UK

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Asymmetric information – A situation in which the buyer of the service has less information about its quality than does the seller.

Commercialization – The introduction of contractual relationships and for-profit private provision into the organization of a public service.

Contract – A legally enforceable agreement to buy and sell specified goods or services.

Cost disease – The tendency for the unit cost of a service to rise relative to its counterpart in other sectors, as a result of its low relative growth rate of labor productivity.

Efficiency – The extent to which the allocation of resources to the production of different goods and services maximizes the total value of output, including the extent to which the cost of producing any particular mix of goods and services is minimized.

Motivation – The considerations that elicit effort from the providers of training services, including economic (profit, pay, and effort) and noneconomic (professionalism and altruism) components.

Quasi-market – An allocative mechanism in which a public agency sets the demand for and the price of a service and a range of providers, public as well as private, for profit as well as nonprofit, compete for contracts to provide it.

Reputation effect – When buyers in a market obtain sufficient information about low-quality training programs to avoid purchasing from their sellers.

Training provider – A producer of training services, including public colleges, training companies, training charities, employers training their own staff, and employer groups.

The training market is a mechanism through which many governments nowadays organize the training services that they choose to subsidize. A public agency contracts with training providers on behalf of potential trainees for the delivery of specified training services. The contractors may be public or private organizations, for-profit or nonprofit; the trainees may be employed by them, by another employer, or out of work. The transactions typically comprise skills enhancement, job search, and other services for disadvantaged and unemployed workers. In some

countries, training markets also play a role in the vocational preparation of young people.

This definition of the training market is narrower than the literal one, which would include training that is financed entirely privately, by individuals or employers – including training in foreign languages and software applications, as advertised in cities around the world. This article focuses on the subset of training for which a public agency is the dominant purchaser.

A prominent contemporary example of the training market is provided by youth training in the UK (i.e., strictly speaking, England). Under the Apprenticeships program, a public agency, the Learning and Skills Council (LSC), contracts with a range of training providers, including further education colleges, specialist training organizations, including companies and nonprofit private associations, and employers themselves, whether singly or jointly. These training providers, in turn, subcontract extensively among themselves for specific services, including off-the-job training and the assessment of trainee attainments (Ryan and Unwin, 2001; Lewis and Ryan, 2009).

The training market is part of the wider commercialization of the public services (Pollitt and Bouckaert, 2004). Commercial relationships and market-like allocations replace, in whole or in part, the combination of political representation and public administration that previously governed and administered services, such as compulsory education and vocational training. Commercialization nowadays dominates such public services as rubbish collection and information technology. In the US and the UK in particular, it has made substantial inroads into the prison, security, defense (frontline military), welfare to work, television, education, and health services.

In both countries, the training market has operated more extensively and for longer than have its better-known counterparts in education and health services. The training market plays a greater role in the vocational development of young people in the liberal market economies, including the US and the UK, than in the coordinated ones, including much of continental Europe (Hall and Soskice, 2001).

The training market may be termed a quasi-market, as it both resembles and differs from a regular market in the private sector (LeGrand and Bartlett, 1993). It is similar to a regular market in that contractual exchange governs the allocation of resources, potential training suppliers compete for contracts, and a market price exists for specified services. It differs from a regular market on the demand

side, where a public body dominates purchasing decisions. This means that political priorities rather than trainees' preferences determine the content of training. It also means market power: the market price is set by the buyer rather than by market clearing, governed by fluctuations of supply and demand across time and place, as in commodity markets.

On the supply side, even when providers compete for public business, economic incentives and competition are potentially weakened when the providers are public bodies, as are vocational colleges in apprenticeship training in Germany. Market forces may then be replaced by administrative decisions. However, the relationship between buyers and sellers within the public sector can still be contractual, and public sellers be required to compete for business with private providers, as are further education colleges in the UK.

Quasi-markets have been used to organize training programs for disadvantaged young people in the US since the advent of the Job Corps program in the 1960s. Their role expanded rapidly in the 1980s, under the Job Training and Partnership Act in the US and the Youth Training Scheme in the UK. A striking attribute of training markets in both the US and the UK has been employer leadership. Employers' representatives – in principle at least – supervise the allocation of public funds to training programs or determine training standards for their sectors (Heinrich, 2000; Burghardt *et al.*, 2001; MAAC, 2001).

Adult training programs are organized through quasi-markets or regular markets in most developed countries, including the ones as institutionally different from the US and the UK as Sweden and Germany. However, market-based organization has had only limited effect on youth (apprentice) training in Germany and its smaller neighbors, where the governance and administration of vocational preparation are shared between representatives of employers, employees, and vocational educators (Ryan, 2000).

A salient concern in the UK training market has been low service quality. A low point was reached in 2002, when the public inspectorate failed a majority of the commercial training providers that held public contracts and a majority of young trainees did not complete their training programs (ALI, 2002).

Two aspects of the training market are discussed here: its economic functioning and merits, with particular reference to service quality; and the political constraints on its performance. The examples used are taken primarily from youth training in the UK.

The Economics of the Training Market

The economic potential of the training market starts with private ownership and competition. Just as a

private-ownership market economy is shown by welfare economics – under particular conditions – to maximize efficiency, so commercialization, that is, the introduction of contracting and the profit motive, might be expected to increase efficiency in training activities. Certainly, public administration, which has often provided the alternative, is characterized by weak incentives to cost reduction and innovation (Shleifer, 1998; Dixit, 2002).

A key ingredient is private ownership and profit seeking. The presence of for-profit providers, that is, not just public colleges and training centers, increases incentives to eliminate waste and improve products and processes. Under public provision, the performance of the service tends to be reduced by ineffective management, by high pay for and low effort by employees, and by weak pedagogy. Training companies have the incentive to increase profit by making improvements in all three areas: to improve management structure, to obtain from staff as much effort for as little pay as the external labor market permits, and to develop new modes of service delivery, such as replacing classroom-based instruction with information technology (IT)-based self-instruction.

In principle, the incentive to providers to cut costs can be intensified by including performance-based bonuses in both the provider's overall contractual payments and its employees' pay. Thus, under output-related funding, one-fifth of the LSC's payment for providing a place in the UK's apprenticeship program has, in recent years, depended on the trainee's completion of training.

The training market also offers the prospect of economies of scale. Training providers are encouraged to specialize in their own areas of expertise. Subcontracting between themselves according to expertise is expected to reduce costs further.

Potential market-related benefits also include increased choice for trainees and increased responsiveness of the content of training to the preferences of its beneficiaries, both trainees and employers. However, the quasi-market has less to offer here than would a full market. The dominance of the demand side by a public authority means that the preferences of trainees and employers play only an indirect role in purchasing decisions.

The benefits of commercialization also depend on the extent of competition for public contracts. If many providers compete for contracts, none enjoys market power, and the price that the public purchaser pays is pushed down to the minimum cost of the most efficient competitors. Public providers, such as post-compulsory colleges, that wish to win contracts must then reorganize their operations in order to compete with commercial providers.

The competitive potential of the training market appears to be broadly realized in practice. In the UK, the ten largest noncollege providers account for only one-fourth of apprenticeship training places (Lewis and Ryan, forthcoming). Seller concentration is undoubtedly

higher in particular submarkets, such as that for a particular occupation in a particular town. Nonetheless, given low barriers to entry by new suppliers, a dominant purchaser, and fixed-price contracting, even large sellers, should have limited scope to profit directly from market power. Indeed, commercial providers face particular risks as part of short-term contracting with a buyer whose priorities and requirements are subject to sharp change at short notice. An extreme instance was the UK's Individual Learning Accounts program, set up in 2000 and canceled abruptly a year later, in response to evidence of lax administration and contractual abuse by some providers.

Economic Difficulties

Nevertheless, the effects of the training market on efficiency are problematic. The principal problem is asymmetric information. Training services may be taken to have both quantity and quality dimensions: for example, the number of people trained, and the amount learned by each, respectively. Information on quantity is typically better than that on quality. One problem in measuring quality is disagreement on the meaning of learning and skill: for example, the types of each that matter in a particular program (Kirkpatrick and Martinez Lucio, 1995). Thus, the UK's training programs have seen disputes as to whether technical (underpinning) knowledge should be valued separately from practical skills within vocational competence.

A further problem is that, even were a particular definition of quality to be widely accepted, the degree to which it is attained in practice would still be difficult to measure. Were the learning in question to be certifiable, the qualifications attained by trainees would in principle capture both quantity and quality. However, some learning is invariably not certified, and both the validity and the reliability of assessments of the learning that is certified are often questionable. Therefore, training quality tends to be proxied in practice by various attributes of provision, including the qualifications of trainers, the use of appropriate materials and equipment, and the completion rates of trainees.

Assuming for simplicity that quality is an agreed, unidimensional but not directly measurable attribute of a training program, contracts for training are then intrinsically incomplete: that is, any quality-related clauses cannot be enforced at law. Assume further that: suppliers are motivated entirely by profit; producing higher quality means higher cost; up to some nontrivial level of quality, the economic benefit of higher quality exceeds its cost; and training providers are better informed about the quality of their services than is the public purchaser. An efficient contract must then motivate providers to offer the optimal level of quality.

Then, the quality that they supply depends on the type of contract on offer. Three possibilities stand out. The first bases their payment on some proxy for training quality – such as output-related funding in the UK's apprenticeship program, under which the LSC's payment to a provider is partly contingent on the qualifications attained by its trainees. The problem here is that, as quality is not fully measurable, any proxy measure is prone to distortion, to an extent that increases with the intensity of the incentive. Thus, when the UK government, in the early 1990s, increased the performance component from one-fifth to one-half, it soon reversed its decision. The new payment system was seen to further undermine what was already a fragile assessment system for work-based competence under the National Vocational Qualifications system, under which training providers enjoyed considerable scope to certify inadequate learning as competence (Felstead and Unwin, 2001).

A similar problem potentially affects the labor contracts used by the training provider. A strong incentive component potentially increases employee motivation. However, if the quality dimension of performance is not readily measurable, dysfunctional results are to be anticipated. These include overproduction of the quantitative aspects, which can be measured and rewarded (e.g., numbers trained and formal qualifications), and underproduction of the qualitative ones (e.g., trainee learning).

The second contractual option avoids these snags: a fixed-price contract, in which payments to the supplier depend only on quantity, with no reward for raising quality. Producers now maximize profit by offering minimum quality – a clearly inefficient outcome.

The third option is a cost-plus contract, which compensates the provider in full for the additional cost of providing higher quality. The provider is now induced to offer high quality – indeed, to gold plate, raising quality beyond its optimal level. Such contracts have been widely used in defense procurement and health services, when quality is seen as more important than cost reduction (Chalkey and Malcomson, 2000).

The best contractual choice for the public buyer is then a second-best one, combining all three approaches. Partial cost sharing compensates providers partly but not wholly for providing more expensive services, with higher cost assumed to result from higher quality. Small performance bonuses encourage higher effort without providing strong incentives to reduce quality.

The contracts used by the LSC for its apprenticeships program involve all three principles, but the fixed-price component dominates. The price paid to training providers varies by occupation, skill level, and trainee disadvantage – all potentially associated with differences in the quality and cost of training. One-fifth of the total payment depends on the attainment of qualifications by trainees. Otherwise, the price is fixed: within a particular

occupation and skill level, providers who provide higher-quality programs at a higher cost are not compensated for doing so, and have therefore little incentive to do so – in contrast to, for example, an automobile producer considering the introduction of a high-quality model (LSC, 2003).

The exception is training provided by an employer for its own staff. Here, the benefit of an increased future supply of skilled labor provides an incentive to high quality that is absent for a specialist training provider, whose training budget is limited to revenues received from the public purchaser of training (MAAC, 2001; Ryan *et al.*, 2007).

Commercial contracting threatens to result in not only low service quality but also positive selection into training. If the cost of training varies across individuals and if the training provider knows more about individual circumstances than does the public purchaser, then fixed-price and incentive contracts encourage the selection of individuals with low training costs – that is, cream skimming. When the goals of the training program include equity as well as efficiency, for example, seeking to serve disadvantaged low educational achievers, their attainment may therefore be impeded by commercialization.

The prospects for high quality in the training market are therefore limited, in terms of contractual enforcement at least. They are improved by a further economic mechanism: reputation effects. It may more plausibly be assumed that training quality, while not strictly measurable and contractually enforceable, is observable: that is, buyers have some, albeit only partial, information about the quality of a particular provider's training. The provider then has an incentive to raise quality, insofar as low quality fosters a reputation that reduces demand for its services. The feedback from quality to demand may involve both search and experience: buyers may obtain information before purchasing, and buyers may learn from their earlier purchases, which affects the demand for repeated purchases (MacLeod, 2007).

The attributes of both providers and trainees might suggest only weak reputation effects in the training market. The business is one in which the priorities of the public buyer can change rapidly, economies of scale are limited, and new providers can enter easily, as capital requirements are low, and proprietary knowledge is not important. The benefits to a training provider of investing in a reputation for quality would therefore be low, even were the price received for the service to respond favorably.

On the demand side of the market, most trainees are young, and unlikely to be either well-informed first buyers; and, as only a minority participate more than once in a public program, repeat purchases are rare. Nevertheless, as the purchasers are not trainees themselves but the public body that buys services on their behalf, reputation effects could still prove strong. Thus, the LSC has bought apprenticeship training for vast

numbers of trainees, at both local and national levels. It is intrinsically a repeat purchaser, buying training repeatedly from many providers over the years. It should possess enough information about particular providers' service quality to steer its purchases to those that provide acceptable quality. The LSC claims to have withdrawn contracts from particular providers, but the extent to which it uses information on quality to reallocate contracts more generally remains unclear.

The information on which reputation effects depend is potentially enhanced by the external inspection of training provision and the publication of inspection findings. In the UK, the Adult Learning Inspectorate (ALI, now part of the schools inspectorate – the Office for Standards in Education, Children's Services and Skills (OfSTED)) has reported on a 4-year cycle on the quality of the services of all contractors for public training programs, and published the reports individually on its website. The quality of the provider's operations is assessed in detail and as a whole. Providers whose services were rated inadequate overall were in principle debarred from public contracts if they failed to make the changes deemed necessary by the inspectors.

Nevertheless, the informational contribution of external inspection proves disappointing, in the UK at least. Inspection practice focuses more on the provider's organization and procedures than on the substantive attributes of its services: for example, the creation of individual training plans for trainees receives more attention than their content and implementation. The primary role of inspection proves to be sociopolitical rather than economic. Inspection concentrates on the attributes of providers that are easily inspected and readily improved, thereby providing reassurance that public money is being well spent even when all is far from well on the ground. Thus, the ALI's inspections have emphasized procedural attributes, notably leadership and management, rather substantive ones, such as the content of training and the amount learned. In particular, the extent to which trainees complete training has had little influence on inspection verdicts: providers with very low completion rates have been rated adequate as long as their management procedures are deemed satisfactory. Providers who have been deemed inadequate at first inspection have typically managed within a year to implement the procedural changes needed to pass on reinspection, without being required to show evidence of substantive improvement. The informational value of inspection reports to purchasers is correspondingly weak (Power, 1999; Lewis and Ryan, 2009).

A further potential buttress for training quality under market-based contracting is the presence of noneconomic motivation, including both altruistic and professional dimensions. The owners and employees of a service provider may be less concerned with maximizing their

incomes and more concerned with the needs of trainees or the standard of service provided. As such, they may opt to provide higher service quality than is strictly in their own economic interests.

Such outcomes are particularly likely under nonprofit provision, including both private bodies, such as youth charities, and public colleges. It potentially mobilizes gifts, whether of labor, from employees and volunteers, or capital, from philanthropists. For-profit providers have negligible access to such gifts, because they can be expected to expropriate them – that is, use them to increase profit rather than to improve the service. Similarly, professional commitment on the part of training staff is intrinsically more compatible with nonprofit than with for-profit ownership. Commercial contracting weakens noneconomic motivation, by undermining professionalism and weakening the willingness of employees to work hard for substandard pay (Rose-Ackerman, 1996).

The supply of altruistic and professional motivation to training services may, however, be limited. The traditional view of the public services, in which employees were seen as ‘knights’ and entrepreneurs as ‘knaves’, ignored evidence of self-interest and inefficient practices, in terms of weak management and slack employee effort, in the public provision of many public services. To the extent that the supply of altruism and professionalism is limited, the problem of service quality remains to be solved, but nothing much is lost by commercialization (LeGrand, 2003).

More generally, the effects of commercializing a public service depend on its specific attributes. The principal ones are: the socioeconomic importance of service quality itself; the scope for cost reduction, both through reorganization and innovation, under for-profit contracting; how much collateral damage is done to service quality by cost reduction; and the strength of reputation effects, external inspection, and noneconomic motivation as antidotes to low quality. Service-specific combinations of these attributes favor commercialization more strongly, for example, in rubbish collection than in education or defense services (Hart *et al.*, 1997).

What about training services? First, in principle, high service quality is accepted as important by all contributors, including governments. That should apply to the vocational preparation of young people, for which educational and personal development, and not just economic performance, is at stake.

Second, the scope for cost reduction under traditional public provision was undoubtedly considerable. Undemanding employment conditions for vocational teachers in UK further education restricted labor inputs and protected traditional pedagogical methods. The scope for innovation in training is, however, disputed between those favoring classroom-based group teaching and those who look instead to self-guided, IT-based learning. This difference of opinion

reflects a difference in objectives: the extent to which youth training should comprise educational development, not just occupational learning. The assessment of the third attribute follows directly. If youth training is viewed in terms of personal development as well as of job skills, serious damage to quality can be anticipated from the cost cutting that is encouraged by commercialization. Finally, as already discussed, the scope for reputation effects and inspection to curb quality losses under commercial contracting appears to be modest, while the contribution of professional and altruistic motivation is weakened by for-profit provision in particular.

The balance of these considerations is not readily determined. The evidence on them is limited, as is to be expected from the low visibility of training quality. Comparing the performance of for-profit and nonprofit specialist training organizations in the US and the UK, no systematic differences are found between their completion rates and their trainees’ outcomes after leaving, (Heinrich, 2000; Lewis and Ryan, 2009). Therefore, commercialization does not appear to reduce quality.

However, such evidence is of limited value, as it covers only specialist training providers and narrowly defined program objectives. In the UK, employers who train their own employees have substantially higher completion rates than do specialist training organizations – a difference consistent with other evidence, such as additional investment and greater staff expertise, of higher quality in many employers’ programs (*ibid.*).

Second, the learning aspirations of public training programs are themselves limited. This is most clear in comparison to the national apprenticeship systems that pursue clear educational objectives. Commercial methods are less suited to the diverse objectives of vocational education than to the narrower objectives of public training programs. By the same token, the commercialization of public training programs in Britain has reoriented youth learning from vocational education toward job training.

The Cost Disease, Public Spending and Politics

The quality problems of the training market may have additional causes. The leading candidate is the cost disease of the services, combined with fiscal politics.

Education and training services are labor intensive. They are produced with little physical capital, with little scope for technical change and productivity growth. Their technological stagnancy means that they exhibit the cost disease of the services: cost per unit of output – and, in a market allocation, the price – of their output increases over time, relative to that in other sectors. The lower the sector’s relative productivity growth, the faster is the rise in its relative unit cost (Inman, 1985). The trend increase

in real tuition fees for private education in the UK and the US can be understood in such terms.

If the demand for a technologically stagnant service is income elastic and price inelastic, then in a mainstream market system its share of national income at current prices will rise over time: increasing real income stimulates demand, while increasing relative price does little to reduce it. The private education market shows such attributes, in terms of trend increases in both relative price and the demand for places. The same is taken here to apply to skills and training generally.

In a quasi-market, the response of demand to rising relative unit cost depends on the government's budgetary policy. If the government chooses to reduce the real growth of public spending on the service, its output grows more slowly and it may even fall. The effect on output may involve quantity, quality, or both. If the public buyer gives priority to quantity, then quality suffers (Ryan, 1992).

The politics of training in liberal democracies encourage the prioritization of quantity over quality. Four reasons may be distinguished. First, the electoral politics of public spending and taxation encourage governments to restrict public spending and reduce personal tax rates. Second, the greater visibility of quantity than quality in training programs means that a government can anticipate greater political rewards, for a given training budget, from making the program available to more eligible citizens than from offering a better service to fewer people. Third, in Europe at least, the goals of public programs involve social justice as well as efficiency, and entitlement-based participation means quantitatively large programs. Thus, the targets that have dominated training programs in the UK in the past decade have been primarily quantitative: for example, the target that all eligible young people have access to the Apprenticeships program by 2013 (DIUS, 2008). Finally, the career prospects of ministers tend to benefit from highly publicized policy innovations, whose presentation matters politically more than their content and implementation (Pierson, 1996; Keep, 2006).

Evidence of political constraints on training quality can be found in the decisions governments actually make when a choice must be made between quantity and quality. Both, the replacement of the Youth Training Scheme by the Modern Apprenticeship in the UK in the 1990s, and the subsequent introduction of requirements for training toward a technical certificate and basic skills in all MA programs, suggest the reverse prioritization: that is, quality over quantity. However, the detailed content and the implementation of such changes suggest otherwise, in terms of: the marked continuity in training methods and outcomes between successive programs; the lack of requirements for the skills and qualifications of training staff; the educational weakness of many of the technical certificates recognized for public support; and

the substantive failings of external inspection (Ryan and Unwin, 2001; Ryan *et al.*, 2007).

The quality problems of the training market may therefore reflect the wider difficulty of raising productivity and reducing costs in a labor-intensive public service, combined with fiscal pressure to restrict spending on it, and the political incentive to give quantity priority over quality when doing so.

Summary

The training market sees a range of providers, public and private, for profit and nonprofit, compete for contracts to provide training as part of publicly funded programs. It is characterized by a dominant public purchaser and fixed-price contracts.

Relative to traditional public administration, the training market strengthens economic incentives, with the introduction of the profit motive and competition. It promises increased efficiency, as the result of improved management, reduced labor costs, and increased innovation. Its potential to improve options for trainees and to make training more responsive to their needs is, however, limited.

A weakness of training markets is the difficulty, given asymmetric information, of ensuring acceptable quality under contractual relationships and economic motivation. In principle, the problem can be countered by reputation effects, external inspection, and nonprofit provision, but the power of all three mechanisms is limited.

Low service quality is also promoted by the technological stagnancy of training services, in conjunction with limits on public spending and political incentives to give priority to quantity over quality. Low quality may result more from political choices than from any technical impossibility of ensuring acceptable service quality under commercial contracting.

The implication is not, however, that traditional public administration should be preferred to quasi-markets. For vocational education at least, better results can be anticipated from joint administration by employers and the representatives of employees and educators, as in the national apprenticeship systems of continental Europe.

See also: Educational Privatization.

Bibliography

- ALI (2002). *Annual Report of the Chief Inspector, 2001–2002*. Coventry: Adult Learning Inspectorate.
- Burghardt, J., Schochet, P., McConnell, S., *et al.* (2001). *Does Job Corps Work? Summary of National Job Corps Study*. Princeton, NJ: Mathematica Policy Research.

- Chalkey, M. and Malcomson, J. M. (2000). Government purchasing of health services. Culyer, A. J. and Newhouse, J. P. (eds.) *Handbook of Health Economics*, vol. 1a, pp 847–890. Amsterdam: Elsevier.
- DIUS (2008). *World Class Apprenticeships*. London: Department of Innovation, Universities and Skills.
- Dixit, A. (2002). Incentives and organizations in the public sector. *Journal of Human Resources* **37**, 696–727.
- Felstead, A. and Unwin, L. (2001). Funding post-compulsory education and training: A retrospective analysis of the TEC and FEFC systems and their impact on skills. *Journal of Education and Work* **14**, 91–111.
- Hall, P. and Soskice, D. (eds.) (2001). *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. Oxford: Oxford University Press.
- Hart, O., Shleifer, A., and Vishny, R. (1997). The proper scope of government: Theory and an application to prisons. *Quarterly Journal of Economics* **112**, 1127–1161.
- Heinrich, C. J. (2000). Organizational form and performance: An empirical investigation of nonprofit and for-profit job-training service providers. *Journal of Policy Analysis and Management* **19**, 233–261.
- Inman, R. P. (1985). *Managing the Service Economy*. Cambridge: Cambridge University Press.
- Keep, E. (2006). State control of the English education and training—playing with the biggest train set in the world. *Journal of Vocational Education and Training* **58**, 47–64.
- Kirkpatrick, I. and Martinez Lucio, M. (eds.) (1995). Introduction. In *The Politics of Quality in the Public Sector*, pp 1–25. London: Routledge.
- LeGrand, J. (2003). *Motivation, Agency and Public Policy*. Oxford: Oxford University Press.
- LeGrand, J. and Bartlett, W. (eds.) (1993). *Quasi-Markets and Social Policy*. London: Macmillan.
- Lewis, P. and Ryan, P. (2009). The role of external inspection in the public services: The case of the UK training market. *Public Administration*. (in press).
- LSC (2003). *Requirements for Funding Work-Based Learning for Young People 2003/04*. Coventry: Learning and Skills Council.
- MAAC (2001). Modern apprenticeships: The way to work. *Report of the Modern Apprenticeship Advisory Committee*. London: Department for Education and Skills.
- Macleod, W. B. (2007). Reputations, relationships and contract enforcement. *Journal of Economic Literature* **45**, 595–628.
- Pierson, P. (1996). The new politics of the welfare state. *World Politics* **48**, 143–179.
- Pollitt, C. and Bouckaert, G. (2004). *Public Management Reform: A Comparative Analysis*. Princeton, NJ: Princeton University Press.
- Power, M. (1999). *The Audit Society: Rituals of Verification*, 2nd edn. Oxford: Oxford University Press.
- Rose-Ackerman, S. (1996). Altruism, non-profits and economic theory. *Journal of Economic Literature* **34**, 701–728.
- Ryan, P. (1992). Unbalanced growth and fiscal restriction: Public spending on higher education in advanced economies since 1970. *Structural Change and Economic Dynamics* **3**, 261–288.
- Ryan, P. (2000). The institutional requirements of apprenticeship: Evidence from smaller EU countries. *International Journal of Training and Development* **4**, 42–65.
- Ryan, P., Gospel, H., and Lewis, P. (2007). Large employers and apprenticeship training in Britain. *British Journal of Industrial Relations* **45**, 127–153.
- Ryan, P. and Unwin, L. (2001). Apprenticeship in the British “training market”. *National Institute Economic Review* **178**, 99–114.
- Shleifer, A. (1998). State versus private ownership. *Journal of Economic Perspectives* **12**, 133–150.

Relevant Websites

- <http://www.dius.gov.uk> – Department of Innovation, Universities and Skills, UK.
- <http://www.lsc.gov.uk> – Learning and Skills Council, UK.
- <http://www.ofsted.gov.uk> – Office for Standards in Education, UK.

Vocational Education and Training Teacher Training

S Nielsen, The European Training Foundation, Torino, Italy

© 2010 Elsevier Ltd. All rights reserved.

Glossary

Communities of practice – A term especially introduced in Lave and Wenger's *Situated Learning. Legitimate Peripheral Participation* (1991) meaning an activity system about which participants share understandings concerning what they are doing and what that means in their lives and for their communities.

Competence – The ability to apply learning outcomes adequately in a defined context (education, work, personal), or professional development.

Continuing professional development – Any kind of activity (education, training, development work, nonformal learning) which after initial education contributes to improve knowledge and skills and continuing personal or professional development.

Learning facilitator – Anyone who promotes the acquisition of knowledge and skills by establishing a favorable learning environment, including anyone exercising a teaching, training, supervision, or guidance function. The facilitator helps the learner develop knowledge and skills by providing guidelines, feedback, and advice throughout the learning process.

Learning outcomes – The set of knowledge, skills, and/or competences an individual has acquired and/or is able to demonstrate after completion of a learning process, either formal, nonformal, or informal.

Professionalization strategy – The concept or professionalization refers to two related processes. The first, internal, aspect concerns the development and consolidation, by a group of individuals, of the competences required to practice a profession. The second, external, refers to the claiming of a distinct social status as part of the division of labor in society. Both sides of the process are interrelated. The transmission of codified practices during initial training is part of the professionalization of the occupation and contributes toward its social recognition.

Teacher – A person whose function is to impart knowledge, know-how or skills to learner in an education or training institution. A teacher may fulfill several tasks such as organizing and carrying out training programs/courses and transmitting knowledge, whether generic or specific, theoretical

or practical. A teachers in a vocationally oriented institution may be referred to as a “trainer.”

Trainer – Anyone who fulfills one or more activities linked to the (theoretical or practical) training function, either in an institution for education and training, or at the workplace. Two categories of trainers exist: (i) professional trainers are training specialists whose job may coincide with that of the teacher in a VET establishment and (ii) part-time or occasional trainers are professionals in various fields who take on, in their normal duties, part-time training activity, either in-company or externally.

Vocational education and training (VET) – Education and training which aims to equip people with knowledge, skills, and competences required in particular occupations or more broadly on the labor market.

Introduction

The recruitment, training, and work of vocational education and training (VET) school teachers are core topics of vocational and business education and training. However, until today there is neither a requirement profile substantiated by theory and supported by empirical investigations nor a common job description for VET school teachers. What we know is that there is a great variety in the structure, status, and situation of the initial and continuing training and development systems for VET teachers and trainers. In this article, we first define the teachers in VET and then analyze major contextual factors and drivers of change influencing the educational and professional development needs of VET teachers and their new roles. We then examine initial teacher training for VET teachers and activities and trends in continuing professional development. In the final section, we consider future challenges for the VET teaching profession.

It is a complex task to give a global overview of the situation of VET teachers, the training of teachers, and the institutions in which they are educated and are working now. Studies of teachers from recent years do not make particular reference to the specific problems of teachers in vocational education. (These include the recent monitoring activities of the Organization for Economic Cooperation and Development (OECD) with regard to attracting,

developing and retaining teachers; the Working Group of the European Union (EU) on the qualifications of Teachers within the process Education 2010; the activities in the Asian context or the work of International Centre for Technical and Vocational Education and Training (UNEVOC) the comparative studies carried out by United Nations Educational, Scientific and Cultural Organization European Centre for Higher Education (UNESCO-CEPES) on institutional approaches to teacher education in Europe and in the Asia-Pacific Region (Battezzati *et al.*, 2004; Central Institute for Vocational Technical Education of the People's Republic of China, 2000; European Commission, Directorate-General for Education and Culture, and Working Group, Improving Education of Teachers and Trainers, 2003; Hopkins and Stern, 1996; Leney and The Lisbon-to-Copenhagen-to-Maastricht Consortium Partners, 2005; OECD, 2002; Morris and Williamson, 2000; Moon *et al.*, 2003). The significance of vocational learning is often overshadowed by the greater emphasis societies place on academic education and credentials. Not so much is, in fact, known about VET teachers; the last comprehensive international studies on VET teachers occurred more than 30 years ago (International Labour Organization, 1964; UNESCO, 1973; UNESCO's section for technical and vocational education, 1997). However, a study has been published (Grollmann and Rauner, 2007).

Another barrier is the difficulty related to comparative research. A general lesson from the history of comparative education is that “every system of education is shaped by its local, historical, economic, cultural and social context. . . . Education, in short, cannot be de-contextualized from its local culture . . .” (Crosley and Watson, 2003: 39). While we have overviews of VET TT systems and practices in a number of regions in the world (CEDEFOP, 2004–2007; ETF, 2005; etc.) – quite rich in data – they do not describe the societal background in any depth or detail. The main problem is the lack of contextualization and in-depth analysis of trends from a national perspective.

The low status of VET and the general problem of increasing the status of the teaching profession are a third factor. While vocational teachers and trainers are essential to supporting skill development in the workforce, they are not granted high status in this role. In industrialized countries, some two-thirds of the workforce that constitutes the backbone of each economy are intermediate-level workers and employees who have learned a substantial part of their occupational skills and knowledge through the support of teachers, trainers, and instructors from the domains of nonacademic technical vocational education and human resources development (cf. CEDEFOP, 1998). Given the basic importance of vocational training for economic success, it is significant that in many countries VET has failed to achieve the level of social recognition that is needed to establish VET teaching and training as a well-regarded profession that attracts

societal support, including attracting appropriate individuals to practice as vocational teachers. This is reinforced by the fact that teaching generally has always had problems in gaining professional recognition, and has even been referred to as a ‘semi-profession’ (Etzioni, 1969). The low professional status of teachers in VET is accompanied by a fragmentation of the profession through the variety of existing profiles and multiple ways of teacher training and recruitment.

With the increasing emphasis on lifelong learning, VET teachers (and trainers) can now be regarded as learning facilitators and VET teaching can gain acceptance as a core profession in the knowledge society. Improving the standing of teachers is, therefore, a significant lever for increasing the quality of vocational education, as acknowledged by many international and national organizations. The awareness of the roles of teachers and trainers as learning facilitators has markedly increased.

It is even becoming ‘high politics.’ In Helsinki on 5 December 2006, European Ministers of Education stated that “highly qualified teachers and trainers who undertake continuous professional development are essential to improve the quality of education and training systems.” However, VET teachers and trainers everywhere are getting older, their tasks more varied and complex, and they receive little support to cope with the changes and new demands made of them. In **Figure 1**, the European Centre for the Development of Vocational Training (CEDEFOP) captures well many of the new challenges for VET teachers and trainers in today's changing societies.

Who Are the VET Teachers?

In an analysis of the international landscape of VET, Rauner (1999) identified roughly four models of the school-to-work transition, based on the socioeconomic contexts, and producing a huge variety of functions of the educational day-to-day work distributed between different profiles of teaching and training. Job specialization has grown in all countries, albeit it is not well defined everywhere. Rauner identifies six clusters of teaching profiles within VET and, given the four models of school-to-work transition, this already leads to 24 different profiles of VET teachers, lecturers, and instructors. In reality, the picture will be even more complex as country in-depth studies would show. The global reality will be much more complex; given the vast number of different profiles and functions of VET teachers, this is indeed a very fragmented profession. (Grollmann (2007) identifies five different basic professional profiles of teachers, trainers and instructors in VET.)

The term VET teacher generally designates personnel in secondary level schools and VET colleges, regardless of the level of education.

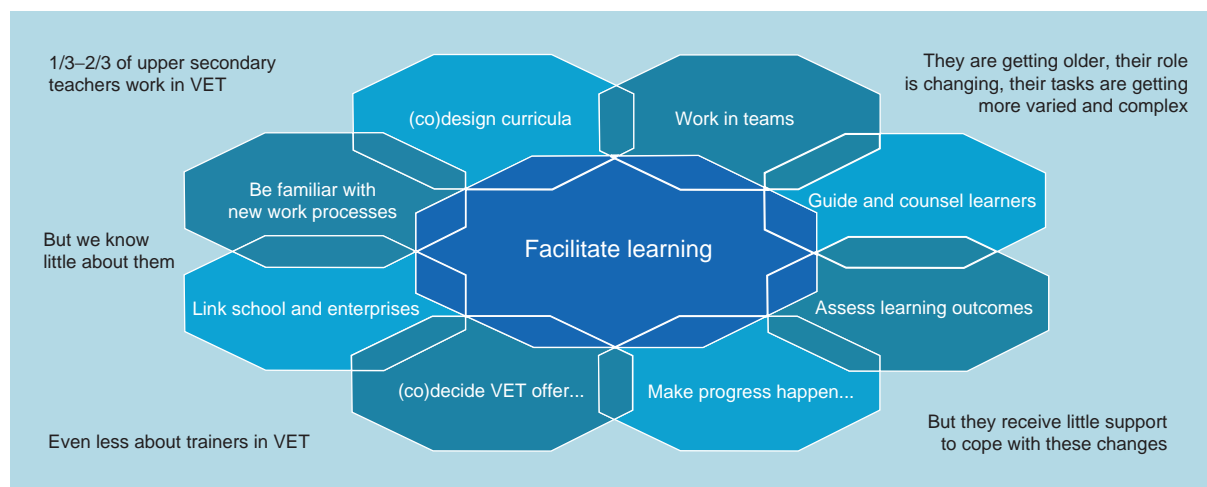


Figure 1 VET teachers and trainers: the forgotten change agents. From CEDEFOP, Conference outcomes: VET teachers and trainers: the forgotten change agents? CEDEFOP TTnet 9th Annual conference, Thessaloniki, 7–8 December 2006.

VET teachers:

1. Working context: public system, education system, Initial Vocational Education and Training (IVET).
2. Teaching content: theoretical part of VET.
3. Functions: general subject teachers, vocational subject teachers.

The term trainer normally refers to professionals involved in apprenticeship systems, on-the-job and off-the-job training and, more generally, in private sector training markets.

VET trainers:

1. Working context: private system, industry, and commerce sectors, Continuing Vocational Education and Training (CVET).
2. Teaching content: practical part of VET.
3. Functions: trainers, other learning facilitators.

Contextual Factors and Drivers of Change Influencing VET Teachers' Work

Educational reforms have been on the political agenda for many years. Earlier, several years passed between reforms but in this century each day seems to bring a new educational reform so there is no reason for calling it reforms anymore; what is going on can be called continuous change.

In education and VET there is a global movement toward standards. This approach assumes that the quality of education and productivity of labor can best be achieved by setting high performance standards for teaching and learning, and then measure to what extent the standards have been met. The Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA) are good examples of

some of the ingredients of this concept. The EU Open Method of Coordination in many ways reflects this.

Standardization seems (currently) to be the preferred approach, described by Hargreaves *et al.* (2001) as “a new official orthodoxy of educational reform,” very much inspired by the Anglo-Saxon community.

As Sahlberg (2004) points out, recent research in England and Wales, New Zealand, Australia, Japan, Singapore, and the USA indicates that standardized reforms may have drawbacks, for instance, de-professionalization of teachers' work and a more narrow focus on basic skills in core academic subjects, and less on how to seek information and use that information in problem-solving, etc. Over-standardization, therefore, does not seem to be the answer to challenges from the knowledge society, where people need to be able to acquire knowledge, use it, develop new knowledge, apply it to new situations, and share it.

Andreas Schleicher (2006) writes in The Lisbon Council Policy Brief that the most successful countries have something in common:

... they have all shifted policy away from control over the resources and content of education toward a focus on obtaining better outcomes. They have moved from “hit and miss” teaching practices to establishing universal high standards. They have shifted from uniformity in the system to embracing diversity and individualising learning. They have changed from a focus on provision to a focus on choice, and they have moved from a bureaucratic approach towards devolving responsibilities and enabling outcomes, from talking about equity to delivering equity. Most important, they have put the emphasis on creating a “knowledge-rich” (knowledge-rich means (in this context) that both national prescriptions and professional judgments are informed) education system, in which teachers and school principals act as partners and have the authority to act, the

necessary information to do so, and access to effective support systems to assist them in implementing change (Schleicher, 2004).

Such changes involve radical challenges for teachers and other professionals in VET. Two major policy elements (Ball, 1999) can be found in almost all countries. One is the insertion of the market form which is intended to subject education to the dynamics and culture of competition and business; the other is performativity. This seems to lead to a control system based on a so-called noninterventionary form of governance – the market decides what is necessary, and standards are set up in order to tell the institutions about good behavior. Governmental steering from a distance has been ever more common. In most cases, it has been combined with giving educational institutions more autonomy, authority, and accountability.

Some of the main drivers behind educational reforms are:

- Challenges from the knowledge society.
- The idea behind new public management and the efforts to change the public sector to make it more market oriented, customer related, performance based, competitive, flexible, and accountable.
- The constructivist approach, meaning that we all actually construct our own knowledge, and that the truth is what we believe in.
- The pressure from new groups of students with new and different expectations as to how to be treated in school – and how to learn.
- The never-ending story about information and communication technology (ICT) development and what it means to both teachers' and students' competence.
- Challenges from living in a global world and the subsequent need for understanding different languages, different cultures, etc.

It can be illustrated as shown in **Figure 2**.



Figure 2 Main drivers for education reforms.

These issues are shared globally. The discussions all over the world are more or less the same. In a new study of globalization Grewal (2008) analyzes the drivers for global change and catches these in his concept of network power. Networks are the means by which globalization proceeds. All networks have standards embedded in them. In theory we can choose among standards and become free. In practice, Grewal shows, our choices tend to narrow over time so that standards are imposed on us. Driving forces include the spread of the internet with its worldwide fiber-optic network and the growth of outsourcing. This generates the process of worldwide industrialization and intensifies worldwide interdependencies; the continuous flow of people, ideas, capital and goods ensures that a global knowledge-based economy emerges. With modern communications and the aggressive marketing of Western culture, the poor are now much better informed about how the other half lives. (In Giddens (2002) *Runaway World*, he gives an illustrative example of how an anthropologist friend of his studied village life in a remote area in central Africa. The day she arrived she was invited to a local home for an evening's entertainment. She expected to find out about the traditional pastimes of this isolated community. Instead, the occasion turned out to be a viewing of *Basic Instinct* on video, the film which at that point had not even reached the cinemas in London (p. 6).)

The challenge is how to cope with these changes in the specific national and cultural context. In the end, all educational reforms touch the teachers (and the students). At this crucial interface, they may have opinions differing from those of politicians and school leaders.

Today, more focus is put on the teacher as a facilitator, a person who will be able to organize learning for students, recognize students' prior learning, and cooperate with both students and colleagues (in teams?) and is supposed to work for both, students and country. Experience from countries already down that road shows that it is a very difficult process. At least two challenges are present. Teachers consider themselves (1) experts in their professional field, and (2) professionals, in the sense that they know what is best for the students, and therefore cannot be commanded and/or controlled by others (the management).

A teacher (as an expert) with a lot of experience in classroom teaching is able to act appropriately if something unexpected happens – immediate, adequate, and fast. But what happens if conditions change? If, for instance, teachers are now supposed to work as facilitators and in environments, based on principles from open learning or resource-based learning approaches? What will happen if teachers, due to a new reform, must cooperate with each other (in teams and across subjects), focus more on learning than on teaching, and accept that students should have much more influence on both form and content? Will they continue to use their original expertise

(from classroom-teaching)? And will that still be appropriate? In most cases the teachers' response to (the new) challenges will be inadequate. The teachers' options then would be to either (1) claim that the new conditions are bad for both teaching and learning or (2) accept to start as beginners in this new game, and to drop the idea of being experts. The first option is very difficult in relation to school leaders and policymakers, and the second one would be very difficult for the teachers themselves.

What is obvious everywhere is that the challenge needs to be solved by creating a new teacher professionalism involving new forms of relationship with colleagues, students, and society. This requires new and more comprehensive competences for both teachers and school leaders.

Initial Training for VET Teachers

Qualified VET teachers and trainers are in short supply in many countries. There are many problems related to the human resource base for VET and the VET teacher profession are almost everywhere, except maybe Germany, a young country. How could we best institutionalize initial teacher training and how do we recruit the right people into such programs? In small countries it is difficult to provide training in the many vocational specialties without underutilizing training capacity; this is today the case in the new republics of the former Soviet Union and former Yugoslavia. In most countries, we see challenges to ensure that the pedagogy part of VET teacher training is sufficiently relevant for the instruction of practical skills in the vocational specialty. Sufficient and relevant work experience from industry is also difficult to ensure when recruiting teacher candidates.

In general two traditions in VET teacher education have dominated: the school tradition where teacher training takes place in designated teacher-training institutions, and the academic tradition where teacher training takes place at universities.

Two main VET teacher types dominate:

The craftsman-turned-teacher. These teachers view themselves as skilled craftsmen; to be a teacher is only a secondary and, often, lately acquired part of

their professional identity. These teachers teach the practical or technical subjects in VET. In some countries, for example, the UK, the requirement of professional teacher training has been introduced only recently.

The professional VET teacher. Few countries have special professional educational courses aimed at becoming a VET teacher (e.g., Belgium, Germany, Hungary, the Netherlands, Norway, Poland, and Sweden). Students choose VET teaching as their career at the beginning of higher education itself, and to be a VET teacher is the primary and maybe only aspect of these teachers' professional identity.

There are many models of programs which give access to becoming a VET teacher. We will concentrate here on initial training for teachers of VET subjects; general subject teachers are educated in ordinary teacher education programs. To illustrate the scope of programs we will here compare highly different VET TT models in three Northern European neighboring, and in many ways, isomorphic countries, Denmark, Germany, and Norway which all have successful dual VET systems (Table 1).

This illuminates the scope of models. The Danish in-service tradition with a short pedagogical training course delivered as a sandwich program with a considerable part taking part in VET schools where candidates are employed (the craftsman-turned-teacher) versus the long university programs in Germany (the professional VET teacher). Norway has recently introduced 3-year programs in order to cope with challenges of the knowledge society.

Grollmann and Rauner (2007) in their comparative study describe four types of program content:

- A purely methodological training on teaching methods, for example, the teaching certificate in UK as a preparatory measure or the in-service courses in Denmark, the USA, Brazil, Japan, and (as one variety) in France. This is often connected to a situation based on the recruitment of practitioners of a certain field of occupational work.
- A consecutive concept based on a sequence of studying the subject matter (e.g., at the BA level) and obtaining an appropriate entry qualification to the education sector through acquiring general teaching skills in a

Table 1 Initial VET teacher training in Denmark, Norway, and Germany

	Where institutions	When pre/in-service	Whom access	How structure	How long duration	What curriculum
DE	University (many)	Preservice (young)	Abitur Work experience	University Practice ('Referendariat')	5 years 11/2 years	VET subject General subject Pedagogy
DK	Semi-academic institution (one)	In-service (mature)	Skilled worker 5 years of work experience	Sandwich course With Mentoring in schools	14 weeks (660 h)	Pedagogy Psychology Didactics
NO	Semi-academic institution (10)	Preservice (young)	Abitur Skilled worker	Study with School practice	3 years	VET subject General subjects Pedagogy

designated course program (typical in the USA). This concept can also be found in the Western Balkans and Turkey leading to a 4-year BA level degree.

- A model based on the concurrent study of a subject matter and educational sciences leading to a BA or an MA degree. Often the subject matter study is a reduced part of the ordinary business or engineering degree. Sometimes special vocational didactics are added. This is the model in the former Soviet Union and former Yugoslavia.
- A model based on an integrated conception of vocational disciplines where the subject matter is derived from the world of work and a model of competence development within this domain. This paradigm of vocational learning turned into teacher education is found in Northern German teacher education institutions and, to some extent, in Norway and some Chinese VET teacher education programs.

The UK has around 30% of VET delivered by providers without public institutional status; here teacher qualification requirements are limited to those who assess learner competence and with no parallel requirements for VET teachers. Also in the UK, with some parallels in Australia, there is a work-based alternative to school-based education; a host school or the teacher's own workplace can provide this training. In Albania (and in many other countries) neither pedagogical requirements nor courses to qualify VET teachers exist. In most of the southern European countries (e.g., Greece, Spain, France, Italy, Cyprus, Portugal, etc.), a state exam (*concours*) is the decisive point for becoming a teacher. Available courses are structured as specific preparation for this exam.

In Europe (Parsons *et al.*, forthcoming) a very complex patchwork of different stages of development exists, and there is no common tradition for VET teacher training. Structures diverge but there are common and enduring problems: aging, status, deregulation, limited career paths, and limited durability of employment. Entry requirements are rising across Europe on VET teachers who are increasingly becoming a graduatized workforce. A widespread emphasis is put on building and extending pedagogical skills but through a general education model which is rarely VET centered.

For Europe the common conclusion is one of convergence and divergence. The EU is pushing for member states to converge and this process of convergence is evident throughout Europe. At the same time, we see an adaptation process where the individual member states adapt the EU policies to national practices or interpret and transform them to suit national practices.

However, there are still a couple of unsolved problems which are found in most countries of the world.

Almost all studies identify two critical findings about VET teacher training (Rauner and Maclean, 2009):

- A relevance to practice of the specialized programs that is felt to be deficient. In almost all cases the practical experience as a skilled worker or engineer was seen as being of greater value for working as a teacher than studies in pedagogical sciences.
- The lack of any interlinking of the educational/sociological part of programs with the concrete professional and teaching situations, so that hardly any pedagogical competences relevant to working as a teacher were acquired during the studies to become a VET teacher.

Continuing Professional Development of VET Teachers

Throughout the world there is a shortage of in-service continuing development of VET teachers to keep updated on technological change. The concept of professionalization is closely linked to updating, upgrading and developing VET teachers' and trainers' competences so that they can meet the challenges facing their profession and act professionally in their daily work. In general, the continuing professional development (CPD) and in-service dimension of how VET practitioners stay qualified has received little attention in comparative analysis. In-service continuing training is generally a large and heterogeneous area with numerous offers of great variety and usually many different private/public profit/nonprofit providers.

Substantial differences exist between EU countries concerning in-service and continuing training and development (continuing professional development) for VET teachers. (CEDEFOP study on Training of VET teachers and trainers in Europe: a comparative perspective.) In many cases (e.g., Cyprus, Iceland, Ireland, Italy, Poland, and Slovenia), CPD is in principle voluntary, but teachers are expected to participate. In most of the new Eastern European EU countries CPD is laid down by law stipulating the minimum requirements to teachers' CPD within a certain period of time. CPD categories vary from obligatory in-service training, arrangements stemming from collective bargaining (Finland, Hungary), and CPD structures relying substantially on self-motivated in-service development. Where common themes are apparent, they seem to stress enhanced pedagogical training, teamworking, leadership, enhanced/updated subject knowledge, and preparation for advancement and on wider nonspecific professional skills. One component of CPD emerging in a number of EU countries is for post-qualification training in workplace-based technology and organization. Also in Germany and the UK among others, recent company-initiated initiatives have attempted to bridge the gap between theoretical vocational knowledge with skills use and new processes in industry and commerce.

In the CIS countries (of the former Soviet Union), the former Yugoslav countries, and among others. Turkey, continuing vocational training (CVT) is highly centralized and offered through a national catalog of training courses on offer. In these countries that are in transition, the CVT activity is extremely low; a foreign donor-led reform is ongoing in several countries to create a free market in CVT and establish accreditation systems for CVT providers.

The types of CVT span from on-the-job competence development in connection with implementing a reform to Master programs. A general trend is toward CPD in coaching, mentoring, and supervision as more student-oriented approaches have been introduced (e.g., in Austria, Belgium, Denmark, Norway, and the UK). This emphasizes the need for a diverse multifaceted provision for CPD that is focused on developing new competencies. In-service training for trainers in the on-the-job part of dual VET systems are regulated in countries like Austria, Germany, Hungary, and Slovenia; in Denmark there is a belief that learning organized in accordance with a production logic instead of a pedagogical logic is highly effective in a combination with formalized learning processes in school.

Continuing training of VET teachers is often criticized for not preparing teachers adequately to update their skills, and for not equipping them sufficiently to master new teaching methodologies. Much more than is usually apparent in studies of VET TT, it is the concrete conditions of work that primarily influence the understanding of vocational pedagogical tasks, the professional self-image, cooperation of teachers – in short, the essential dimensions of the process knowledge of vocational teachers. The impact of the learning environment within concrete working conditions of teachers in VET institutions is often overlooked. As contributors to CPD, comparative studies of those concrete working conditions and practices could be of great interest for understanding the reflected practitioners in VET.

An interesting response to this lacuna is found in Denmark which has a long tradition of local innovations seen as central for modernizing curricula and systems in VET. The vehicle is the Forsøgs- og Udviklingsarbejde (FoU)-scheme (innovation and development grant program) financing school-based projects led by teachers, and at the provider and cross-provider level. The Ministry of Education lays down overall reform priority areas stimulating and financing teacher-led school projects, with expertise to guide these projects from VET teaching expert consultants funded directly by government. These development projects are seen as important contributors in the continuing development of teachers in Denmark where the innovation of content, methods, and teacher competence development go hand in hand.

This is one of the few examples of practitioner-led change vehicles focusing on teachers as change agents, while in other countries comparable programs in the VET modernization agenda are fully policy led.

Funding of CPD is constrained everywhere. New and more cost-effective delivery methods are tested in a number of countries. In England a largely cascade-model for subject-based new learning methods has been implemented. In the Netherlands and Denmark, group and peer learning methods and communities of practice are starting to emerge as methods for reprofessionalization of VET teachers with some positive evidence of being both cost-effective and well received by teachers as a focus for reflective practice and for knowledge sharing. In Australia, CPD for VET teachers has become an important policy lever through a collaborative model to provide for a bottom-up approach to CVT (Dickie *et al.*, 2004).

Future Challenges

The Missing VET Didactics

There is no distinctive professional knowledge base in the pedagogy of VET learning and delivery. Many national systems which require a pedagogical foundation for VET practitioners continue to derive this from teacher training rooted in general education contexts. Also the demand for more learner-centered approaches in a vocational context calls for more specific and applied VET knowledge which is not always served well by current approaches. Universities in most cases lack the application and workplace experience to respond effectively.

The challenge is to combine the worlds of work and education and consequently different fields of science. VET teachers work in a more complex field compared to teachers in general education because their subjects, and thus their knowledge base, are characterized by a double theory/practice problem. They work in a field that lies between specialist theory and specialist work as well as between school practice and work and/or professional practice. This is a domain which is not really there but needs to be better conceptualized and developed. A research context which could help uncover the domain may be inspired by US research in connection with pedagogical content knowledge (PCK) (Shulman, 1986; Bauer and Grollmann, 2009).

The PCK approach represents a unique knowledge base of the teachers; it forms an amalgam from content-related (field-specific) and pedagogical knowledge and is interdisciplinary. The PCK strategy has concerned itself especially with how subject-related and curricular contents are transformed by teachers in the teaching activity itself. This inquiry into the professional knowhow of VET teachers would be worthwhile pursuing in the coming years.

Peer Learning and Communities of Practice

Communities of practice have started to emerge as vehicles in the reprofessionalization of VET teachers. There is some narrow but positive evidence that these are proving to be both cost effective and also well received and regarded by practitioners as a focus for reflective practice, a foundation for professionalism, and for sharing this with others in the professional field. This merits more attention through analysis of the various examples of innovation, both at policy, provider, and practitioner levels.

In most countries, still, the concept of CPD of teachers is only seen as continuing training of teachers provided by external delivery systems. CPD is a much more promising strategy; it is much cheaper and it reestablishes a social recognition of teachers as professionals and stakeholders of reform.

With the aim to establish another teacher professionalization strategy in Southeast Europe, the European Training Foundation (ETF) has launched a new teaching and learning project from 2007 to 2009. The 3-year ETF project aims at strengthening the national institutions which have a potential leverage for carrying forward ideas and projects in this field. The project will support VETCenters or other relevant institutions in order to enhance their capacity:

- to carry out continuous innovation and adaptations to changing conditions and increase the responsiveness of VET to local needs;
- to cope with the challenges of new policies through school-based development work, innovation of teaching and learning, and enhancing capabilities for international network learning and project cooperation; and
- to enhance their own and local school's expertise by actively taking part in knowledge-sharing based on horizontal learning processes.

The project is grounded in horizontal learning processes which are important tools for policy learning, and makes use of a principle of community of practice as a vehicle for network learning where participants (1) share a given practice, (2) are able and willing to learn together and (3) actually work together on improvements of practice, inventing new procedures, models, tools, etc. and share the results of their mutual work. Communities of practice are seen as useful tools both as an instrument for national learning activities and as a knowledge-sharing tool across countries. Shared learning activities will be based on the exchange of participants' own experience and the improvement of practice and competence will go hand in hand.

VET Teachers as Change Agents

In modern vocational education and training systems, teachers are, at the same time, professional educators

and key change agents. Continuing innovation and development has become a core task of a modern professional teacher. The professional expertise of teachers committed to change and modernization is also an important source of knowledge for policymakers leading the development and implementation of national policies. A key question therefore is: How to involve VET teachers actively in VET reform so that ownership will be better translated into quality learning in the classroom and professional expertise from teaching and learning processes can guide system reform?

Successful reform can only happen with engagement and commitment of teachers, if only because in the end, they will be the ones who will have to make it happen in their daily work with students and others. However, this is no longer simply a matter of establishing broad ownership and acceptance.

What makes the situation different today is that the new professional profile of teachers includes innovation and development as a key competence. They are no longer the executors of educational programs decided in detail by others but have to adapt learning processes and outcomes to the specific changing needs of their students and local labor market situations. Teachers are stakeholders in their capacity of educational professionals. This, in turn, reflects the fact that the current reforms in VET are very complex development processes, especially true for reforms in transition countries. Such reforms are not one-off events designed by external experts but ongoing change processes set within a broadly agreed reform agenda, which can be quite radical but that requires further operational detailing based on local innovation processes. It is because of this that teachers who are actively engaged in local innovation and experimentation are an important source of expertise for national policymakers; reform strategies have to build on engaging teachers working inside their school organizations. Such an understanding of reform puts policy learning, capacity building, and policy advice at both national and school levels in a new perspective and at the same time with considerable more urgency than before. Traditional top-down or bottom-up strategies are insufficient to make reforms work. Reform processes require a continuous interaction and dialog between national and local partners. There are therefore strong pressures to include teachers among the principal stakeholders in reform.

Changing Frameworks or Changing Actors?

There is an intense international debate about the lack of real substance per se in standards and national qualification frameworks. The focus only on learning outcomes is now perceived as too simplistic and has led to the belief that it really does not matter at all how people arrive at certain learning outcomes. It is questioned whether this

model sufficiently reflects the premise that rapidly changing and unstable labor markets ask for new kinds of competences enabling people to cope with increasing uncertainties. As a result there is a growing interest – also at the policy level – to pay more attention to the quality of learning processes, including the role that teachers play in helping people to learn. Learners and learning processes are back in the spotlight. VET schools and teachers and trainers have become key factors. Increasingly, the challenge to balance learning inputs, outcomes, and processes has international dimensions as well. The ETF Yearbook 2007 is placed within this broader context and the focus of the debate raised here is: How do we rebalance the tensions in VET systems between input and outcomes and how can a renewed focus on substance and learning processes be fed into VET reform policy?

We still need to find answers to this question.

Bibliography

- Ball, S. J. (1999). Global trends in educational reform and the struggle for the soul of the teacher. *Paper Presented at the British Educational Research Association Annual Conference*. University of Sussex at Brighton, 2–5 September 1999.
- Bauer, W. and Grollmann, P. (2009). Vocational training research for the professionalisation of vocational school teachers. In *UNESCO-UNEVOC Book Series on TVET*. Dordrecht: Springer.
- Beijaard, D., Verloop, N., and Vermunt, J. D. (2000). Teachers' perceptions of professional identity: And exploratory study from a personal knowledge perspective. *Teaching and Teacher Education* **16**, 749–764.
- Carnoy, M. and DeAngelis, K. (2002). The teaching workforce: Concerns and policy challenges. In OECD (ed.) *Education Policy Analysis*, ch. 3. Paris: OECD.
- CEDEFOP (2006). Conference outcomes: VET teachers and trainers: The forgotten change agents? *CEDEFOP TTnet 9th Annual Conference*, Thessaloniki, 7–8 December 2006.
- CEDEFOP (2004–2007). National VET systems database: Detailed thematic analysis. http://www.trainingvillage.gr/etv/Information_resources/National/Vet/thematic/analysis.asp (accessed August 2009).
- Crosley, M. and Watson, K. (2003). *Comparative and International Research in Education*. London: RoutledgeFalmer.
- Dickie, M., Eccles, C., Fitzgerald, I., and McDonald, R. (2004). *Enhancing the Capability of VET Professionals*. Brisbane: Australian National Training Authority.
- Etzioni, A. (ed.) (1969). *The Semi-Professionals and their Organization*. New York: Free Press.
- European Commission Working Group (2003). Improving education and training of teachers and trainers. *Progress Report*, November 2003.
- European Training Foundation (ETF) (2005). *Compendium of Good Practice in VET Teacher and Trainer Training in South East Europe*, Nielsen, S. and Oldroyd, D. (eds.). http://www.europa.eu/web.nsf/pages/EmbedPub_EN?OpenDocument (accessed August 2009).
- Giddens, A. (2002). *Runaway World*. London: Profile Books.
- Grewal, D. S. (2008). *Network Power: The Social Dynamics of Globalisation*. New Haven, CT: Yale University Press.
- Grollmann, P. and Rauner, F. (eds.) (2007). *International Perspectives on Teachers and Lecturers in Technical and Vocational Education: Volume 7. UNESCO-UNEVOC Book Series on TVET*. Dordrecht: Springer.
- Hargreaves, A., Earl, L., More, S., and Manning, S. (2001). *Learning to Change. Teaching beyond Subjects and Standards*. San Francisco, CA: Jossey-Bass.
- International Labour Organisation (1964). *Training Vocational Teachers*. Geneva: ILO.
- Moon, B., Lazar, V., and Leland, C. B. (eds.) (2003). *International Approaches to Teacher Education within Higher Education in Europe: Current Models and New Developments*. Bucharest: UNESCO/Cepes.
- OECD (2002). *Education Policy Analysis*. Paris: OECD.
- Parsons, D., et al. (forthcoming). The training and development of VET teachers and trainers in Europe. In CEDEFOP (ed.) *Modernising Vocational Education and Training. Fourth Report on Vocational Education and Training Research in Europe: Background Report*. Luxembourg: Publication Office.
- Sahlberg, P. (2004). Teaching and globalization. *Managing Global Transition* **2**(1), 65–83.
- Schleicher, A. (2006). The economics of knowledge: Why education is key for Europe's success. *Lisbon Council Policy Brief*.
- Shulman, L. S. (1986). Those who understands: Knowledge growth in teaching. *Educational Researcher* **15**(2), 3–13.
- UNESCO (1973). *Technical and Vocational Teacher Education and Training*. Bristol: John Wright.
- CEDEFOP (2007). *Training of VET Teachers and Trainers in Europe: A Comparative Perspective*. Thessaloniki: CEDEFOP.
- Cort, P., Härkönen, A., and Volmar, K. (2004). PROFF – professionalisation of VET teachers for the future. *CEDEFOP Panorama Series*; **104**.
- Cunliffe, J. (2004). The future of initial teacher education for the learning and skills sector. An agenda for reform. *A Consultative Paper*. Staff Qualifications and Development Team, Standards Unit, Department for Education and Skills, Sheffield.
- European Commission (2005). *Common European Principles for Teacher Competences and Qualifications*. http://europa.eu.int/comm/dgs/education_culture (accessed August 2009).
- Green, A., Wolf, A., and Leney, T. (1999). *Convergence and Divergence in European Education and Training Systems*. London: Bedford Way Papers, Institute of Education, University of London.
- Grollmann, P. (2005). *Berufspädagogen im internationalen Vergleich. Eine empirische Studie anhand ausgewählter Fälle aus den USA, Dänemark und Deutschland*. Bielefeld: Bertelsmann.
- Grootings, P. and Nielsen, S. (eds.) (2005). *The Role of Teachers in VET Reforms: Professionals and Stakeholders. ETF Yearbook 2005*. Turin: ETF.
- Hargreaves, A. (2003). *Teaching in the Knowledge Society: Education in the Age of Insecurity*. New York: Teachers College.
- Hargreaves, D. (1994). *Changing Teachers, Changing Times*. New York: Teachers College Press.
- Lynch, R. L. and Ruhland, S. K. (2007). Career and technical teaching and teacher education in the United States of America. In Grollmann, P. and Rauner, F. (eds.) *International Perspectives on Teachers and Lecturers in Technical and Vocational Education: Volume 7. UNESCO-UNEVOC Book Series on TVET*, pp 277–306. Dordrecht: Springer.
- Mitchell, J., Chappel, C., Bateman, A., and Roy, S. (2006). *Quality Is the Key: Critical Issues in Teaching, Learning and Assessment in Vocational Education and Training*. National Centre for Vocational Education Research (NCVER) <http://www.ncver.edu.au/publications/1710.html> (accessed August 2009).
- Nielsen, S. (1999). Reshaping the focus and structure of VET teaching personnel training in the partner countries. *A Cross Country Review on Needs, Achievements and Obstacles*. ETF.
- Nielsen, S. (2002). The experience of the Nordic countries and Germany. *Conference Paper, Joint CEDEFOP-ETF Conference*, Aalborg, 21–22 November 2002.
- Nielsen, S. (2007). The professional situation and training of vocational teachers in Denmark. In Grollmann, P. and Rauner, F. (eds.) *International Perspectives on Teachers and Lecturers in Technical*

- and Vocational Education: Volume 7. UNESCO-UNEVOC Book Series on TVET*, pp 76–98. Dordrecht: Springer.
- Nielsen, S. and Angeloska-Galevska, N. (eds.) (2006). *Catalogue of VET Teacher Training Systems in South East Europe*. Skopje: Faculty of Philosophy.
- Nielsen, S., Angeloska-Galevska, N., and Ciolan, L. (eds.) (2007). *Bringing Teachers Back to the Scene. Transformation of VET Teachers and Teaching in South East Europe*. Turin: ETF.
- OECD (2006). *Attracting, Developing and Retaining Effective Teachers*. Paris: OECD.
- Smylie, M. A. (1995). Teacher learning in the workplace: Implications for school reform. In Guskey, T. R. and Huberman, M. (eds.) *New Paradigms and Practices in Professional Development*, pp 93–113. New York: Teachers College Press.
- Terada, M. (2007). The development and present situation of vocational and technical teachers' professions in Japan. In Grollmann, P. and Rauner, F. (eds.) *International Perspectives on Teachers and Lecturers in Technical and Vocational Education: Volume 7. UNESCO-UNEVOC Book Series on TVET*, pp 160–184. Dordrecht: Springer.
- Villegas-Reimers, E. (2003). *Teacher Professional Development. An International Review of the literature*. Paris: UNESCO, IIEP.
- Zhao, Z. and Lu, L. (2007). China's TVET teachers and their professionalization. In Grollmann, P. and Rauner, F. (eds.) *International Perspectives on Teachers and Lecturers in Technical and Vocational Education: Volume 7. UNESCO-UNEVOC Book Series on TVET*, pp 54–76. Dordrecht: Springer.

Relevant Websites

<http://www.minedu.fi> – The European Ministers of Vocational Education and Training, the European Social Partners and the Commission adopted in their meeting on 5 December 2006 the Helsinki Communiqué.

Vocational Education and Training Workforce

D L Wallin, University of Georgia, Athens, GA, USA

© 2010 Elsevier Ltd. All rights reserved.

VET System: VET Workforce

Vocational education and training (VET) is known by many names including vocational education, workforce education, industrial education, manual education, occupational education, further education, training, career education, and technical education. It is a large and diverse educational enterprise that spans both secondary and post-secondary education. Nevertheless, there are certain characteristics that are typically used to define VET and set it apart from traditional liberal arts education. Perhaps most significant is its purpose. Vocational education, by whatever name, has as its purpose the preparation of individuals for work. Historically, the purpose of vocational education has been to prepare individuals for entry-level jobs in occupations requiring less than a baccalaureate degree. More recently, however, vocational education has been reconceptualized into a broader preparation that includes academic as well as vocational and technical skills. Thus, vocational education has morphed into an enterprise that integrates occupational and academic knowledge, not only for entry-level work, but also for fully functioning in the world of work, for advancement, and for continuing education to keep up with the changing requirements of the workplace.

This article is divided into four distinct but interrelated sections: (1) historical roots of VET and the VET workforce; (2) VET faculty characteristics, qualifications, and credentials; (3) permanent/casual VET faculty; and (4) professional development for VET faculty.

Historical Roots of VET Workforce

One cannot understand the nature of the VET workforce without first understanding the history and competing philosophies that underpin contemporary vocational education. For centuries, there have been two types of education – education for work and education for culture. These two types of education have been separated by social class. Education for work was typically delivered through imitation and practical experience with tools, materials, and machines in the field, at home, or in the factory. Education for culture, on the other hand, was delivered through the medium of books, formal instruction in the home through tutors, and later through private and then public schools (Scott and Sarkees-Wircenski, 2001). It is only recently that educators and policymakers have begun to see that these two types of education are not so

distinct: indeed, except for the very elite and independently wealthy, all education is vocational education in that it prepares individuals for work. Nevertheless, there are still significant distinctions in the type of work implied by VET.

Perhaps the earliest vocational education occurred from father to son, mother to daughter as parents passed along to their children the skills by which they earned their living. However, vocational education became formalized when organized apprenticeships began. These apprenticeships were usually sponsored by guilds, particularly in England and France, and later in America. These guilds were composed of organized groups of craftsman who were among the earliest vocational instructors.

During the 1800s, schools were strictly divided by social class and the purpose of education for the wealthy classes was much different than it was for the working or indigent classes. Manual training, or handwork, became the centerpiece of vocational training for the lower classes. Germany was the center of the manual-arts movement and maintained elementary free and compulsory schooling, thus providing a secure foundation for industrial training. Rousseau was an early supporter of manual arts and believing that the manual arts could be seen as a means of mental training and an expanded view of vocational education. Pestalozzi built on Rousseau's thinking and insisted that children should be taught not only to think but also to do and that education should consist largely of manual labor. His ideas spread through Europe and into the United States throughout the nineteenth century (Gordon, 2003).

In the United States, by the late 1800s, industrialization had drastically changed the agrarian landscape of the previous century. While universal education was available up to the eighth grade for most students, the school curriculum was largely focused on preparing individuals for higher education in colleges and universities despite the fact that only a very small number even graduated high school. Thus, the manual-arts movement gained ascendancy as a social and economic issue: the belief that people who are trained to work for wages purchase goods and services and become contributing members of society. Pragmatism as a philosophical ideal became a part of faculty thinking, emphasizing the concrete and practical over the abstract. Education was seen more as a science, and the faculty were free to engage students in problem solving and concrete applications of their learning. John Dewey, a strong advocate of vocational education, came to symbolize early organized vocational

education in America, encouraging faculty to employ the project method of instruction which was being employed in some manual-training schools. He advocated methods of instruction that involved student interests, activity, or learning by doing, as well as group work and cooperation. These tenets were precursors of the emphasis in modern VET on collaboration, teamwork, and experiential learning for both children and adults. Innovative European VET recognizes that the development of professional competencies both for faculty and for students involve active experience. “In this context, learning is not committing to memory or fragmented and de-contextualized knowledge, but learning by doing” (Leney and Green, 2005: 269).

VET Faculty Characteristics, Qualifications, and Credentials

Viable VET programs require teachers who are competent in their occupational fields and competent in their pedagogy and teaching skills. While the list varies depending upon the compiler, the following disciplines are typically considered to be representative of vocational education at both the secondary and postsecondary level: agriculture, business, family and consumer sciences, health occupations, marketing, technology (including engineering technologies), trade, and industrial education.

Traditionally, a two-tier method has been utilized for the preparation of vocational faculty. Teachers in agriculture, business, and family and consumer sciences, for example, are primarily trained through baccalaureate degree programs. Public and private universities and colleges offer professional teacher-education programs, along with discipline-specific coursework. Most students are required to complete a student-teaching internship under the direction of an experienced teacher and mentor. Trade and industrial teachers, on the other hand, have been prepared largely through years of occupational experience coupled with minimal academic or professional-education courses. Teacher-certification requirements at the secondary level are changing to require a baccalaureate degree and certification of all teachers; however, at the postsecondary level there is still much more variance and dispute in the relative importance of experience versus formal education. While trade and industry faculty, health-occupations faculty, and those teaching in more than one specialty are less likely to have baccalaureate degrees, those whose teaching specialties included agriculture, business, technology, and family and consumer sciences, are more like their academic peers in terms of education and degree attainment.

The fact remains that overall VET faculty typically have fewer academic qualifications and more work experience than liberal-arts faculty. Historically, vocational faculty have maintained that real-world practical experience is a necessary precursor to teaching in some areas.

Educational attainment varies markedly depending upon the teaching specialty of the faculty member. This tendency for limited formal education is most apparent in the trades and industrial areas. For example, it would seem obvious that teachers who have never been auto mechanics, or welders, or plumbers, or electricians would have difficulty teaching these skills to students with any degree of credibility. However, studies over many years have indicated that there is no relationship between the number of years of experience and the effectiveness of teaching. The studies also suggest that extensive occupational experience does not confer extensive benefits on vocational teaching; nevertheless teaching is enhanced with some practical experience. Thus, VET faculty in the trades and industrial areas would do well to combine both formal education and knowledge of pedagogy with some practical, on-the-job experience for maximum effectiveness in the classroom.

Due to the wide variance in education and certification, as well as teacher shortages in secondary schools, a variety of alternative certification programs have been developed to encourage VET faculty into the classroom. Alternative routes to the profession include measures of occupation competency, professional experience, and completion of a baccalaureate degree in a relevant subject area without any formal professional-education courses. While some critics insist that alternatively certified teachers are inferior to traditionally certified teachers, others believe that providing alternative certification will attract competent and diverse candidates, as well as those equipped and willing to teach in urban schools, and will help alleviate the teacher shortage.

A large study involving 31 European nations revealed similar issues in the fragmented situation with regard to the education, training, and recruitment of VET professionals. The study suggested that closer cooperation between institution of VET teacher and trainer education and training across Europe, along with a closer connection between the community of practice of VET professionals, researchers, and policymakers would do a great deal to improve the professionalism and skills of VET teachers. Several countries indicated specific measures and goals to strengthen training for VET teachers including:

re-establishing the prestige of the teacher's profession (Bulgaria); enterprise experience for VET teachers during their careers (France); . . . introduction of the VET Teacher Training Certificate (Malta); a project of making teaching more attractive (Sweden); and special teacher training for teachers lacking qualifications (Sweden)” (Leney and Green, 2005: 267–268).

That same study emphasized the importance of faculty being able to provide an experience-based route into work-life hinged on the acquisition of key competencies which complement the more traditional teaching and

learning of technical skills. The study also indicates the centrality of being both a part of the labor market and having close ties and access to general and higher education. All these issues require constant vigilance and attention of VET faculty.

Understanding the challenges faced by contemporary VET faculty helps in developing curricular components for viable and useful teacher-preparation programs. Some of the areas in which future VET faculty will need strong competencies, regardless of their teaching specialties, include contextualized teaching and learning, articulation of secondary/postsecondary programs, alternative methods of student assessment, work-based learning which incorporates service learning, internships, cooperative education, and apprenticeships. They need to understand the importance of interpersonal relationships and so-called soft skills. Entrepreneurship education occupies an increasingly important niche in vocational education in as much as that is where many new jobs are created. Education for entrepreneurship is also seen as a key competence which merits a more prominent role in VET in Europe, Asia, and Africa as well. Understanding and teaching entrepreneurship involves not only discipline-based skills, but also skills in understanding capital acquisition, business plan development, and marketing. Perhaps most importantly, the contemporary VET faculty must be able to integrate academic and vocational education and understand the importance of their synergy.

VET faculty have to be well aware of the role of globalization and international business in their teaching. As Shaw (1999) maintains, there is no doubt that change and reform in European Union (EU) training and education systems will ultimately be driven by changes in the marketplace. Many of these changes in the workforce are driven by increasing globalization and have a direct impact on education and training systems and subsequently on VET faculty. Thus the need is strong to "ensure that larger portions of the workforce hold vocational qualification; ensure that VET processes are relevant to the needs of industry and commerce; and allow for and build in continuing education in order to meet changing skills and technology requirements" (Shaw, 1999: 140).

Similarly, in Finland, through the globalizing of industry, vocational education changed from an occupation focus into an organization focus. Occupational work became subsumed under the skills and competencies needed for flexible employability. Teachers became preoccupied with networking with various stakeholders and organizations rather than spending time with students. "Vocational pedagogy was substituted by training of skills, organizational and personnel development. Thus, the characteristic features of Finnish vocational education seem to lose their function. Vocational teachers are no more primarily proponents of occupation, but didactic-managerial experts" (Heikkinen, 2001: 242).

VET reform in Central Asia as a part of the post-Soviet economic transition has seen old, specialized training for declining industries replaced by a multiskilled labor force that is more able to respond to the emerging market economy. Initiatives to raise the status of teachers are necessarily coupled with an attempt to reduce teacher numbers and curriculum hours and develop more dynamic, student-centered learning. There is opportunity for VET if existing institutions will take the lead in responding to market forces and new business and industry (Howse, 2001).

In America, as in Europe and Asia, vocational education is in a transition period in which there is a perceptible shift from job-specific vocational preparation to a more generic approach, frequently labeled the new vocationalism. The new vocationalism posits that both secondary and postsecondary curricula should be more general and less job specific, allowing students more possible career options. It places strong emphasis on three core curriculum issues: (1) integration of academic and vocational education, (2) articulation of secondary and postsecondary programs, and (3) connections between school and the world of work (Rojewski, 2002). This connection between school and work is exemplified in the model of VET being delivered in the workplace with an increased emphasis on assessment rather than on teaching.

The new vocationalism also implies new demands on VET faculty, requiring them to become well versed in curriculum integration while assuming leadership in providing work-based learning activities such as job shadowing, internships, cooperative education, and school-based enterprises. They are expected to acquire new methods of pedagogy, new ways to assess student learning, new understanding of diverse populations, and a command of instructional technology. Faculty are expected to be fully engaged in these general practices while still maintaining content-area competencies. Further, many faculty have faced uncertainty about the future of their subject areas as their industrial area has either changed or contracted rapidly. New teaching areas have been introduced, including beauty (cosmetology), hospitality, and information technology (IT). Teachers have been required to change the way they taught because of the introduction of competency-based training and the need to develop courses that suit particular employers. Similar to what has transpired in the US, in Australia, alterations in the nature, status, and qualification requirements of the teaching workforce in the VET sector have led to changes in the settings in which VET is delivered, expansion in the market, and new initiatives in the qualifications for VET teachers. In short, the role of a vocational faculty member has become much more demanding and challenging than it once was. It is not sufficient today, as it once was, only to be proficient in a trade or profession.

As is the case with faculty in general, vocational instructors are aging. Many joined educational institutions

in the early boom years for vocational education in the 1970s and are nearing retirement. Particularly troublesome is the fact that vocational education teachers are in general older than their liberal-arts colleagues, though they are similar in years of teaching experience. This may be explained by the fact that many vocational-education teachers have substantial business or industry experience prior to entering the educational arena. The pattern is not unique to America; in Australia as well, the VET workforce is relatively old, with many teachers in their 40s and 50s (Smith, 2005). Whatever the reasons, the aging of vocational faculty and administrators is a policy issue for political and educational leaders. Consideration should be given to the types of incentives that might be employed to encourage young educators to engage in experiences and training that would lead them to become part of the secondary or postsecondary educational system as VET faculty.

Permanent/Casual Faculty in the VET Workforce

Part-time faculty, adjunct faculty, contingent workforce, casual faculty – however they are labeled, part-time faculty have become increasingly important, particularly in postsecondary education. Research has established that in the United States, part-time faculty comprise a substantial portion of higher education faculty and are a majority – nearly two-thirds – among community-college instructors. The community and technical college sector in America is the home of most VET faculty. Thus, many of the VET faculty are involved in education as a secondary pursuit to their full-time work in a trade or profession. In Australia, a majority of full-time teachers were used in VET programs, with a large complement of part-time staff who taught in addition to their regular jobs in trades and professions. Additionally, there are increasingly more part-timers and an extension of casualization and temporary-contract staff (Smith, 2005). Some of the same issues are present in the Central Asian republics of the former Soviet Union as they struggle with a reduction in the number of institutions and fewer full-time VET teachers. Financial realities dictate an emphasis on part-time or short courses and the expansion of private training outside VET, presenting yet additional uncertainty (Howse, 2001).

Given the budget pressures that postsecondary education confronts throughout the world, it is likely that casual faculty are a permanent part of the postsecondary educational landscape. Adjuncts are valued for their specialized knowledge and real-world experience and the close connection to business and industry. They also bring obvious economic benefits to an institution in as much as they are usually paid much less than permanent faculty. In most cases, they do not receive benefits, nor do they require office space or secretarial support. Adjunct faculty can be hired and fired at will and are usually only contracted for

one or two courses and for a single semester at a time. While they may not have as strong a pedagogical background as permanent faculty, many part-time faculty who are in the work-world bring a fresh and reality-based perspective to the classroom. They add diversity, enrichment, and scheduling flexibility to the institution. Those who currently work in a VET field – a business person, a realtor, a carpenter, or a nurse – can bring to the classroom contemporary experiences that are often lacking in full-time faculty. Further, contrary to conventional wisdom, most studies have shown that part-time faculty are as effective as full-time faculty, if student-learning outcomes are the measure. Students learn as much, perform as well, and are as likely to be retained when taught by part-time faculty as they do when taught by their full-time counterparts. Adjunct faculty have no more discipline problems; further, their student evaluations and the grades they give students are comparable to permanent faculty.

Nevertheless, it is true that adjunct faculty may not be as involved or knowledgeable about other important functions (such as financial aid, student services, library services, etc.). Thus, it is imperative that administrators be sensitive to the needs of adjunct faculty and be sure they have access to the same information and resources as do full-time faculty. That is particularly important if the part-time faculty teach in the evening or on weekends or at sites other than the main campus, when the full complement of services may not be available. It is the responsibility of management to recruit, evaluate, and retain successful part-time faculty. One way to support casual faculty is to provide a strong professional-development program tailored to the specific needs of part-time faculty. They need to understand what the institution expects of them through thorough orientation programs and ongoing support. Realizing that part-time faculty choose to teach for a variety of reasons, administrators need to be cognizant of the best ways to integrate them into the culture of the institution.

Professional Development for VET Faculty

Retention strategies are critical in maintaining a strong VET teaching corps at both the secondary and postsecondary levels. Research suggests the importance of retention strategies, particularly in the first 5 years of a new teacher's career. One strategy with documented success is the assignment of a support team to a new teacher. The team includes a subject-area peer teacher, an administrator, a staff person, and a teacher educator. Together, the team sees that the new teacher has the necessary resources and support to be successful.

Another strategy for supporting novice teachers at either the secondary or postsecondary level is the use of a mentoring system. Ideally, the mentor is a master

teacher in the discipline with a track record of good classroom management and positive relationships with students and peers. While mentoring programs are often looked at as a costly expense, considering the cost of recruiting and rehiring new teachers on a frequent basis, mentoring might be considered as a wise investment and fiscally efficient and effective.

When queried as to their needs regarding professional development, VET faculty responded that classroom management and working with special populations were areas in which they felt inadequately prepared (Ruhland and Bremer, 2003). As the demographics of the student population, particularly in Europe and in North America, are changing rapidly, with more variety in ethnicity, language, age, and academic preparation, VET faculty require additional support and assistance in meeting the needs of their new students.

VET faculty, more so than liberal arts faculty, need opportunities to return to the workplace for visits, internships, shadowing experiences, and industrial training. Particularly in fast-changing technology-based disciplines, it is imperative that faculty remain current within their teaching area and within the larger field of vocational and technical education. Thus, short-term leaves of absence should be made available, as well as opportunities to attend professional-association meetings and conferences. Where appropriate, back-to-industry experiences should be sought out and supported. Inservice training is often required for continuing licensure or certification in many fields, particularly in healthcare and in engineering technologies. Administration must be mindful of the time, cost, and commitment involved in maintaining vocational competency and provide appropriate support.

Administrators who are committed to providing professional-development opportunities for both permanent and adjunct faculty are often frustrated by the increasingly difficult challenge of bringing faculty together in one place at one time. The growth of the Internet has provided new avenues for bringing professional development, continuing education, and other educational opportunities to geographically dispersed and busy VET faculty. There is no need to tune in at a particular time or to drive to a specific location. Instead, learning can be asynchronous and accessed on the schedule of the faculty member. Professional-development opportunities are vital for VET faculty to continue to grow with their respective disciplines and to be well-prepared to provide competent, up-to-date leadership in the classroom, in the laboratories, and in the field.

Professional development can also play a significant role in upgrading the status and respect accorded VET faculty. In a study of 13 European partners, reform strategies were developed with the purpose of improving vocational education and its status relative to general education. A key element for reform was the improvement

of the status and qualifications of vocational teachers. Of the four strategies adopted, the one regarding the status and qualifications of vocational teachers was the most difficult to implement largely because of the great salary differentials between the private and public sectors. Nevertheless, opportunities for professional development and the standardization of qualifications for students and teachers surfaced as a common trans-European trend (Stenstrom and Lasonen 2004).

Summary

The roots of VET are both deep and broad. While the earliest vocational training took place within the family, vocational education was soon formalized through the guilds and apprenticeships. VET faculty have always been challenged to be not only competent, but also forward-thinking and on the cutting edge of their respective disciplines. While some faculty have been prepared for their roles through traditional college and university courses, others have brought extensive work experience. The increasing use of casual or part-time faculty has had both positive and negative effects on the profession. Often the part-timers bring valuable current experience to the classroom. However, with the budget pressures endemic in education, casual faculty are sometimes hired in place of, rather than as adjunct to, full-time faculty. Finally, if the VET workforce is to preserve its importance in the education and training of both young people and adults, it will be increasingly necessary for faculty to have continuing professional development experiences in order to maintain their currency and their credibility in the classroom, within the institution, and within the workplace. As the requirements of the workplace change and as academic and vocational enterprises become less distinct, the VET workforce will be constantly challenged to meet the needs of an increasingly diverse and globalized work environment.

Bibliography

- Gordon, H. R. D. (2003). *The History and Growth of Vocational Education in America*, 2nd edn. Prospect Heights, IL: Waveland Press.
- Heikkinen, A. (2001). The transforming peripheries of vocational education: Reflections from the case of Finland. *Journal of Education and Work* **14**, 227–250.
- Howse, G. (2001). VET under review: The challenges in Central Asia. *European Journal of Education* **36**, 35–43.
- Leney, T. and Green, A. (2005). Achieving the Lisbon goal: The contribution of vocational education and training. *European Journal of Education* **40**, 261–278.
- Rojewski, J. W. (2002). Preparing the workforce of tomorrow: A conceptual framework for career and technical education. *Journal of Vocational Education Research* **27**, 7–35.
- Ruhland, S. K. and Bremer, C. D. (2003). Perceptions of traditionally and alternatively certified career and technical education teachers. *Journal of Vocational Education Research* **28**, 285–302.

- Scott, J. L. and Sarkees-Wircenski, M. (2001). *Overview of Career and Technical Education*, 2nd edn. Homewood, IL, American Technical Publishers.
- Shaw, G. (1999). European standards in vocational education and training (VET): What are they and who wants them? *European Journal of Education* **34**, 137–152.
- Smith, E. (2005). A rich tapestry: Changing views of teaching and teaching qualifications in the vocational education and training sector. *Asia-Pacific Journal of Teacher Education* **33**, 339–351.
- Stenstrom, M. L. and Lasonen, J. (2004). Strategies for reforming workforce preparation programs in Europe. *Journal of Technology Studies* **30**, 38–43.
- Cohen, A. M. and Brawer, F. B. (2003). *The American Community College*, 4th edn. San Francisco, CA, Jossey-Bass.
- Dewey, J. (1916). *Democracy and Education*. New York: Macmillan.
- Dougherty, K. J. (1994). *The Contradictory College*. Albany, NY: State University of New York press.
- Grubb, W. N. (1997). Not there yet: Prospects and problems for “Education through occupations.” *Journal of Vocational Education Research* **22**, 77–94.
- Lynch, R. L. (1996). Vocational teacher education: At a crossroads. *Vocational Education Journal* **71**, 22–24.
- Roberts, K. (2001). The new East European model of education, training and youth employment. *Journal of Education and Work* **14**, 315–328.
- Wallin, D. L. (2004). Valuing professional colleagues: Adjunct faculty in community and technical colleges. *Community College Journal of Research and Practice* **28**, 1–19.
- Wallin, D. L. (ed.) (2005). *Adjunct Faculty in Community Colleges: An Academic Administrator’s Guide to Recruiting, Supporting, and Retaining Great Teachers*. Bolton, MA: Anker Press.
- Wallin, D. L. and Smith, C. L. (2005). Professional development needs of full-time faculty in technical colleges. *Community College Journal of Research and Practice* **29**, 87–108.

Further Reading

- Carnevale, A. and Fry, R. (2002). *The Economic and Demographic Roots of Education and Training*. Washington, DC: National Association of Manufacturers, Center for Workforce Success.

Subject Index

Page numbers suffixed by T and F refer to Tables and Figures respectively. vs. indicates a comparison.

A

- abductive research methods 3:77–82
 - exploratory factor analysis 3:77
 - general discussion 3:77, 3:81
 - grounded theory 3:80
 - theory of explanatory coherence 3:78
- Abecedarian Early Childhood Intervention Program
 - characteristics 2:40
 - cost–benefit analysis 2:41*t*
 - economic returns 2:41, 2:42*t*, 2:45*t*
 - funding estimates 2:58*t*
 - outcomes analyses 2:452
 - public investment benefits 2:64
 - research reviews 2:40*t*
 - simulation studies 2:42
 - transition programs 2:44, 2:45*t*
- ABET accreditations 4:211
- ability testing 4:1–6
 - classical test theory 4:2
 - closed versus open tests 4:3
 - consequences 4:5
 - developed ability 4:3
 - dyslexia 4:4
 - general discussion 4:2, 4:5
 - historical background 3:493, 4:2
 - measurement scales 4:2
 - performance monitoring 4:4
 - racial groups 4:3
 - school effectiveness 4:5
- absolute error variance 4:63*t*, 4:64, 4:64*f*
- Abstract Reasoning Test (ART) 7:89, 7:90*f*, 7:90*t*, 7:91*f*, 7:91*t*
- abusive behaviors *see* workplace mistreatment/abuse
- academic achievement outcomes
 - Catholic schools 2:357
 - cost–benefit analysis 2:409
 - cross-cultural learning
 - Asian students
 - cultural values 5:452
 - general discussion 5:452
 - Markus–Kitayama concept of self-construal 5:453
 - Sue–Okazaki concept of relative functionalism 5:454
 - Black/Hispanic students
 - general discussion 5:453
 - institutional practices 5:454
 - Ogbu's two types of minorities theory 5:454
 - social class and wealth 5:453
 - future directions 5:455
- desegregation
 - advantages/disadvantages 2:398
 - court-ordered plans 2:435–436
 - methodological challenges 2:398
 - research results 2:399, 2:399*t*
- early childhood interventions 2:452
- educational production function model
 - cost–benefit analysis 2:409
 - developed countries
 - cost–benefit analysis 2:409
 - general discussion 2:407, 2:410
 - quality assurance 2:408
 - school resources 2:408
 - teachers and schools versus family background 2:409
 - general discussion 2:407, 2:410
 - quality assurance 2:408
 - school resources 2:408
 - teachers and schools versus family background 2:409
- health effects 2:58
- inequality–education relationship
 - fathers' educational background 2:274*f*
 - gender differences 2:275, 2:276*f*, 2:277*t*
 - race/ethnicity/immigrant status 2:274, 2:275*f*, 2:432
 - social background 2:273, 2:273*f*
- neighborhood characteristics 2:431, 2:432*f*
- neighborhoods 2:431–437
 - general discussion 2:431
 - influencing factors 2:431, 2:432*f*
 - moving-to-opportunity (MTO) program 2:432, 2:434*t*
 - research evidence 2:432
 - research results 2:436
- peer relationships 2:431–437
 - general discussion 2:431
 - influencing factors 2:431, 2:432*f*
 - research results 2:436
 - schoolmate/classmate characteristics 2:435
- poverty 2:273, 2:432
- Programme for International Student Assessment (PISA)
 - gender differences 2:275, 2:277*t*
 - immigrant status 2:275*f*
- race/racism
 - neighborhood characteristics 2:432
 - schoolmate/classmate characteristics 2:435
- segregation 2:432
- self-efficacy beliefs 6:671
- socioeconomic status (SES) 2:258
- teacher training and licensing 2:499
- Trends in International Mathematics and Science Study (TIMSS)
 - developing countries 2:416, 2:417*t*
 - gender differences 2:275
 - social background 2:273–274
- academic capitalism 4:300–306
 - background information 4:300
 - copyright 4:303
 - criticisms 4:304
 - economic development
 - general discussion 4:302
 - incubators 4:303
 - research parks 4:302
 - start-up companies 4:302
 - technology transfer 4:303
 - government policies and programs 4:300
 - infrastructure 4:303
 - patents
 - background information 4:301
 - exceptions 4:301
 - personnel coverage 4:301
 - royalties 4:301
- academic development *see* education development
- academic freedom 4:328–333
 - basic concepts 4:328, 4:434
 - definitions 4:341
 - freedom of academics 4:330, 4:332
 - historical background
 - freedom of academics 4:330
 - general discussion 4:329
 - institutional autonomy 4:329
 - restrictions 4:330
 - student freedom 4:331
 - institutional autonomy 4:341–346
 - background information 4:341
 - basic concepts 4:331
 - civic engagement 4:399
 - definitions 4:341
 - external boundaries 4:342
 - general discussion 4:345
 - historical background 4:329
 - internal boundaries 4:343
 - postconflict societies 4:519
 - restrictions 4:330, 4:331
 - student freedom 4:331, 4:332
- academic judgment challenges 3:382
- academic language
 - bilingualism 5:357
 - characteristics 5:352
 - cognitive perspective 5:353
 - sociocultural perspective 5:353
- academic managers 4:229–237
 - complex roles 4:230
 - contractual obligations 4:230, 4:231*t*, 4:232*t*
 - effective leadership behaviors 4:234*t*
 - excellence criteria 4:233*t*

- academic managers (*continued*)
 future challenges 4:235
 heads-of-school framework 4:232*t*
 importance 4:229
 priority duties 4:231*t*
 specific leadership needs 4:234*t*
 staff expectations 4:231
 training and development 4:234, 4:235*t*, 4:236*t*
 transitions to leadership 4:233*t*
- academic mobility 4:559–565
 future directions 4:564
 globalization effects 4:209
 government policies and programs 4:559
 historical background 4:559
 human resource issues 4:531
 internationalization processes 4:533, 4:542
 professional impacts 4:563
 statistical data
 faculty mobility 4:562
 general discussion 4:560
 student mobility 4:561, 4:598
 student perspectives 4:562
- academic self-concept (ASC) 6:660–667
 academic self-concept (ASC)–achievement correlations
 multidimensional aspects 6:661
 reciprocal effects model (REM)
 age generalizability 6:664
 Baumeister–Marsh and Craven debate 6:664
 causality considerations 6:662, 6:663*f*
 cross-cultural generalizability 6:663
 physical domain generalizability 6:664
 positive-effects debate 6:664
 research results 6:665
 background information 6:660
 multidimensional aspects
 academic self-concept (ASC)–achievement correlations 6:661
 definitions 6:660
 hierarchical models 6:660, 6:661*f*
- academic values and expectations 4:433–440
 background information 4:433
 core cultural components 4:434
 external pressures
 accountability 4:437
 economic imperatives 4:436
 emerging skill sets 4:436
 learning societies 4:437
 massification and diversification 4:436
 political impacts 4:435
 technological advances 4:436
 general discussion 4:438
 new initiatives
 funding considerations 4:438
 learning centers 4:437–438
 policies and programs 4:438
 quality assurance 4:438
 teaching impacts
 deep learning versus surface learning 4:435
 general discussion 4:434
 learning paradigm shift 4:435
 research-led teaching and learning 4:434
- academic work habits 6:726
- acalculia 2:649
- accommodation learning 1:37
- accountability
 academic values and expectations 4:437
 accreditation processes 4:388
 adult literacy education 1:110, 1:111*t*
 assessments and evaluations 3:289, 3:379, 3:398
 early childhood education and care (ECEC) evaluations 2:29
 educational assessments
 characteristics 3:794*t*
 functional role 4:46
 importance 3:792
 interventions 3:793, 3:794*t*
 organizational level 3:793, 3:794*t*
 performance measurement 3:792
 summative assessments 3:792, 3:793*f*, 3:794*t*
 educational evaluation 3:600–606
 appropriateness 3:601
 definitions 3:600
 design considerations
 context-sensitive outcomes 3:603
 data analysis and interpretation 3:604
 data sources 3:603
 general discussion 3:602
 multiple measures 3:603
 quality assurance 3:603
 reliability 3:604
 reporting considerations 3:605
 responsibility assignments 3:604
 target audience 3:605
 evaluation reporting and communicating 3:746–756
 audiences 3:753*t*, 3:754
 cartoons 3:751, 3:751*f*
 communicating and reporting plan 3:754*f*, 3:755*f*, 3:756
 communication strategies 3:750
 creative formats 3:750, 3:751*t*
 cross-cultural considerations 3:753
 definitions 3:746
 design considerations 3:754
 design, graphics, and layout 3:748
 e-mail and postcards 3:750
 final evaluation report 3:747, 3:748*t*
 interactive formats 3:749
 newsletters and brochures 3:750
 online collaborative tools 3:750, 3:750*t*
 photography 3:748
 poetry 3:751, 3:752*f*
 purposes 3:747
 stakeholders 3:753
 strategies/formats and resources 3:755
 theatrical presentations 3:752, 3:752*t*
 timing 3:755
 videos 3:749
 visual appeal 3:747
 general discussion 3:605
 outcome-based accountability 3:520
 stakeholder involvement 3:602
 validity and utility framework
 general discussion 3:600
 purpose and scope 3:601
 educational production function (EPF) 2:405
 history/social studies/civics assessments 3:327
 information and communication technology (ICT) 8:6
 lateral accountability 6:279
 mathematics education 5:404
 regulatory accountability 3:795
 school-based accountability programs 2:374–379
 contemporary assessments 5:892
 data sharing ability 8:128
 design approaches 2:375
 educational effectiveness 2:802
 educational evaluation 3:520
 general discussion 2:374
 No Child Left Behind Act (2001) 2:374
 rationale 2:374
 student achievement impacts 2:376
 unintended consequences 2:377
 school–higher education institution (HEI) partnerships 7:761
 science-education programs 3:660
 standards-based accountability 2:342
 test-based accountability 4:153–158
 general discussion 4:153
 growth models
 average gain scores 4:154
 challenges 4:156
 general discussion 4:154
 random-effects models 4:155
 student fixed-effects models 4:155
 value tables 4:155, 4:155*f*
 measurement issues
 alignment analysis 4:157
 general discussion 4:156
 performance-level reporting 4:157
 student inclusion 4:156
 test preparation/score inflation 4:156
 performance indices
 general discussion 4:153
 growth models 4:154
 quasi-longitudinal measures 4:154
 status measures 4:153
 successive cohort measures 4:154
 suggested guidelines 4:157
see also assessment strategies

- accountability talk 5:352
- accounting education *see* business/accounting education
- accreditation 3:1–6, 7:495–502
 - allied health professions 3:52, 3:53*t*
 - benefits 3:2
 - business/accounting education 3:42
 - challenges 7:497
 - characteristics 3:1
 - clergy education 3:35
 - development guidelines 3:3
 - effectiveness evaluations 3:3
 - elder learning organizations 1:180
 - elementary education 7:728
 - engineering profession
 - Accreditation Board for Engineering and Technology (ABET) 3:47
 - Bologna Accord (1999) 3:47
 - United Kingdom 3:46
 - Washington Accord 3:47, 4:209–215
 - ABET accreditations 4:211
 - assessment impacts 4:213
 - curriculum impacts 4:212
 - general discussion 4:214
 - graduate attributes 4:210
 - historical background 4:210
 - impacts 4:211
 - outcome-based accreditation 4:211
 - pedagogy impacts 4:212
 - professional competencies 4:210
 - related international partnerships 4:210, 4:210*r*
 - research initiatives 4:213
 - signatories 4:210*r*
 - future directions 7:499
 - general discussion 3:5, 7:495
 - governance processes 7:495
 - higher education institutions 4:384–389
 - access expansion
 - accreditation impact 4:386
 - distance learning programs 4:385
 - for-profit programs 4:386
 - general discussion 4:385
 - basic concepts 4:384
 - change impacts
 - access expansion 4:385
 - accountability 4:388
 - internationalization processes 4:387
 - core values 4:384
 - functional roles 4:385
 - general discussion 4:389
 - international accreditation 4:533–539
 - agency responses 4:535
 - barriers 4:534
 - capacity development 4:537
 - cross-border education 4:537
 - current developments 4:536
 - future challenges 4:537
 - general discussion 4:538
 - improved dialog and cooperation 4:537
 - intergovernmental bodies 4:537
 - internationalization processes 4:533
 - mobility issues 4:533
 - professional accrediting bodies 4:535
 - quality assurance 4:533
 - university associations and networks 4:534
 - quality assurance 4:533–539
 - agency responses 4:535
 - background information 4:533
 - barriers 4:534
 - capacity development 4:537
 - cross-border education 4:537
 - current developments 4:536
 - future challenges 4:537
 - general discussion 4:538
 - improved dialog and cooperation 4:537
 - intergovernmental bodies 4:537
 - internationalization processes 4:533
 - professional accrediting bodies 4:535
 - university associations and networks 4:534
 - jurisdictional functions 7:496
 - legal education 3:25
 - medical education 3:23
 - models 3:2
 - professionalization process 7:497
 - professional standards 7:495
 - public professional initiatives 7:500
 - recognition processes 3:1
 - standards-based accreditation
 - challenges 7:498
 - evidence weaknesses 7:498
 - mixed professional support 7:498
 - weak professional consensus 7:498
 - systemic effectiveness 7:499
 - veterinary education 3:72
- Accreditation Board for Engineering and Technology (ABET) 3:47
- ACCUPLACER 8:48
- achievement assessments
 - background information 4:186
 - score inflation 4:186–192
 - evidence 4:186, 4:187*f*, 4:187*r*
 - historical background 4:187
 - inadequate tradition approaches 4:187
 - validity of inferences under high-stakes conditions (VIHS)
 - behavioral responses 4:189
 - conceptual elements 4:190, 4:190*f*
 - effective test weight 4:189
 - evaluation frameworks 4:188
 - future directions 4:191
 - general discussion 4:188
 - target of inference 4:189, 4:189*f*
- test-based accountability 4:153–158
 - general discussion 4:153
 - growth models
 - average gain scores 4:154
 - challenges 4:156
 - general discussion 4:154
 - random-effects models 4:155
 - student fixed-effects models 4:155
 - value tables 4:155, 4:155*f*
 - measurement issues
 - alignment analysis 4:157
 - general discussion 4:156
 - performance-level reporting 4:157
 - student inclusion 4:156
 - test preparation/score inflation 4:156
 - performance indices
 - general discussion 4:153
 - growth models 4:154
 - quasi-longitudinal measures 4:154
 - status measures 4:153
 - successive cohort measures 4:154
 - suggested guidelines 4:157
- achievement goal theory 6:537–543
 - classroom goal structures 6:539
 - correlates 6:538
 - definitions 6:537–538
 - historical background 6:538
 - outcomes analyses 6:541
 - performance–avoidance/performance–approach goals 6:538
 - research areas 6:540
- achievement motivation
 - cultural influences 6:577
 - ethnic minorities 1:753–759
 - ability beliefs
 - intelligence theories 1:755
 - racial stereotypes 1:754, 2:776
 - achievement values/effort
 - basic concepts 1:756
 - oppositional identity 1:756, 2:776
 - self-affirmation theories 1:757
 - valued peers 1:757
 - general discussion 1:753–759
 - intelligence theories
 - discrimination 1:756
 - entity versus incremental theories 1:755
 - racial stereotypes
 - African Americans 1:754, 2:776
 - Asian Americans 1:755, 2:776
 - definitions 1:754
 - research areas 1:758
- acquired immune deficiency syndrome (AIDS) *see* HIV/AIDS
- ACT Assessment 4:8, 8:51
- action research 6:311–319, 7:540–547
 - beginning versus experienced teachers 7:545
 - benefits 6:317
 - characteristics 7:540

- action research (*continued*)
 - classifications 7:542
 - collaborative action research 7:759
 - critical theory perspective 1:497
 - definitions 6:313
 - educational applications
 - classroom environments 6:315
 - higher education 6:317
 - theater/arts/media representations 6:315
 - website resources 6:316
 - general discussion 6:312
 - historical background 1:355–361, 6:312, 7:541
 - international programs 6:317
 - as a learning tool 7:540
 - participant observation 6:442–448
 - boundaries 6:445
 - characteristics 6:443
 - direct participation 6:443
 - dramatis personae* 6:444–445
 - ethical practices 6:444
 - field experiences 6:446
 - intentional networks 6:444, 6:444f, 6:446f
 - mirror groups 6:446f
 - ontological commitment 6:445
 - qualitative inquiry 6:493
 - significance 6:447
 - stranger sensitization 6:445
 - symbolic interactionism 6:445
 - participatory action research (PAR) 3:597, 6:338
 - practical approaches
 - active research projects 7:544
 - basic concepts 6:313
 - general discussion 7:543
 - interactive learning strategies
 - academic and professional knowledge 7:544
 - application and construction of professional knowledge 7:544
 - educational and methodological knowledge 7:544
 - ideological, instrumental, and empirical knowledge 7:544
 - individual and collective knowledge 7:544
 - learning conditions 7:545
 - learning improvements 7:544
 - school-based action research project 7:541
 - teacher-based action research 1:355–361
 - future directions and challenges 1:359
 - general discussion 1:355
 - historical background
 - Collier's approach 1:357
 - continuity approach 1:358
 - democratic approach 1:358–359
 - Dewey's approach 1:356
 - Lewin's approach 1:357
 - practitioner research approach 1:359
 - Progressive Era 1:356
 - Wann review 1:355
 - see also* Teaching for Understanding (TfU)
 - action–research curriculum evaluation model 1:338
 - active aging 8:262
 - Active Labour Market Programmes (ALMPs) 8:309
 - active participation 6:443
 - active rehearsal 5:324
 - activism-based learning 1:194
 - activity theory
 - background information 5:330
 - basic concepts 5:494, 5:494f
 - cultural–historical activity theory (CHAT) 5:491–497, 6:360–366
 - activity concepts 5:494, 5:494f
 - activity systems 5:494, 5:495f
 - background information 5:496
 - benefits 6:366
 - developmental background
 - cultural contexts 5:493
 - developmental levels 5:492
 - higher mental functions 5:492
 - theoretical perspectives 5:491, 6:360
 - tool-mediated action 5:493, 5:494f
 - educational approaches
 - development theories 6:361
 - learning 6:361
 - learning–development relationship 6:361
 - research examples 6:362
 - general discussion 6:360
 - and Lev Vygotsky 5:491, 5:492, 6:360
 - research examples
 - general discussion 6:362
 - long division/gazzinta process 6:362, 6:365f
 - puzzles 6:362, 6:365f
 - reading remediation 6:363, 6:365f
 - research results and analysis 6:364
 - weaving instruction 6:363, 6:365f
 - semiotic mediation 5:494
 - theoretical perspectives 6:360
 - third generation systems 5:495, 5:495f
 - workplace learning 1:85
 - e-learning pedagogies 4:423f
 - Teaching for Understanding (TfU) 7:637, 7:638f
 - acute versus chronic stressors 6:570
 - adaptable learning model 5:504
 - adapted feedback 3:226
 - adapted physical activity programs 2:519–529
 - basic concepts
 - historical background 2:521
 - individualized education plans (IEPs) 2:520
 - theoretical/philosophical perspectives 2:522
 - class organization and management strategies 2:525, 2:526f
 - effective physical education pedagogy 2:522
 - general discussion 2:527
 - government policies and programs 2:520
 - importance 2:520
 - instructional strategies
 - active teaching 2:523
 - adventure education 2:523
 - cooperative learning 2:524
 - flexible groups 2:524
 - peer tutoring/reciprocal reaching 2:524
 - self-instructional formats 2:525
 - standards 2:523, 2:524f
 - task teaching 2:524
 - teaching through questions/Socratic method 2:524
 - professional preparation and development 2:526
 - adaptive evaluations 3:638
 - adaptive university model 4:249
 - Addams, Jane 1:55, 6:481, 6:485
 - additive bilingualism 1:294, 1:296f, 5:356
 - adequate education 2:204–209
 - background information 2:204
 - cost–function estimation
 - data requirements 2:207
 - efficiency measures 2:206
 - estimation methods 2:207
 - functional form 2:206
 - general discussion 2:206
 - variables 2:206
 - cost–function results 2:207, 2:208f
 - cost measurement methodologies 2:204
 - '*adl* (justice) 6:69
 - admissions tests 1:845–849, 4:7–14
 - background information 4:7
 - changing contexts 1:845
 - international institutions 4:12
 - noncognitive variables
 - characteristics 1:845, 1:846f
 - community involvement 1:846f, 1:847
 - long-range goals 1:846f, 1:847
 - Noncognitive Questionnaire (NCQ) 1:847, 4:11–12
 - nontraditional knowledge acquisition 1:846f, 1:847
 - positive self-concept 1:846, 1:846f
 - racial discrimination 1:846, 1:846f
 - realistic self-appraisal 1:846, 1:846f
 - successful leadership experiences 1:846f, 1:847
 - support services 1:846f, 1:847
 - Psychometric Entrance Test (PET) 4:13
 - purpose 1:845
 - SAT examinations 1:845
 - SweSAT 4:12–13
 - United States
 - fairness 4:12
 - predictive validity
 - grade point average (GPA) 4:10
 - personal characteristics measures 4:11
 - primary admissions tests
 - ACT Assessment 4:8
 - general discussion 4:7

- Graduate Management Admission Test (GMAT) 4:9
- Graduate Record Examinations (GRE) 4:8
- Law School Admission Test (LSAT) 1:695, 4:9
- Medical College Admission Test (MCAT) 4:9
- SAT examinations 4:8
- usage
 - graduate/professional school programs 4:10
 - undergraduate institutions 4:9
- adolescent literacies 1:438
- Adorno, Theodor 6:133, 6:349
- adult basic education (ABE) 1:102–106
 - future directions 1:105
 - general discussion 1:102
 - national developments 1:102
 - Swedish educational model/Komvux
 - background information 1:103
 - labor-market training programs 1:103
 - vocational-based learning (VBL) 1:104, 1:104*f*, 1:105*f*
- adult cross-cultural kids (ACCKs) 4:639
- adult education and training (AET) 1:1–11
 - adult basic education (ABE) 1:102–106
 - future directions 1:105
 - general discussion 1:102
 - national developments 1:102
 - Swedish educational model/Komvux
 - background information 1:103
 - labor-market training programs 1:103
 - vocational-based learning (VBL) 1:104, 1:104*f*, 1:105*f*
- adult learning 1:5
- Australia 5:513
- barriers 1:234–239
 - conceptual perspectives 1:234
 - large-scale surveys
 - background information 1:234
 - research findings 1:235, 1:238
 - limitations 1:238
 - qualitative studies 1:236
 - theoretical perspectives
 - bounded-agency model 1:238
 - human capital theory 1:237
 - participation research 1:237
- basic concepts 1:2
- Black Book 1:9
- characteristics 1:240, 1:241*r*
- citizenship education 1:113–119
 - assessment strategies 3:326–327
 - background information 1:113
 - basic concepts
 - agency 1:115
 - civic virtues 1:114–115
 - general discussion 1:113
 - identity 1:114
 - status 1:114
 - challenges 7:720
 - educational programs 1:116
 - future directions 1:118
 - intercultural education 1:117
 - moral values 7:651
 - multicultural education 1:117
 - social capital 6:677
 - social studies curriculum 1:411
- civil society 1:191–197
 - functional role 1:191
 - global civil societies 1:195
 - theoretical perspectives
 - communitarian perspective 1:192
 - Gramscian perspective 1:194
 - Habermasian perspective 1:193
- educational sectors 1:4
- Estonia 5:559
- Fiji 5:566
- Finland 5:572
- formal learning 1:207
- Gambia 5:860
- Ghana 5:594
- government-sponsored training 1:206
- Greece 5:601
- historical background 1:1
- historical narratives 1:96–101
 - characteristics 1:96
 - Enlightenment period 1:97–98
 - individualization 1:99
 - industrialization period 1:98
 - institutional histories 1:97
 - national independence movements 1:99
 - post-World War II period 1:99
 - Protestant Reformation period 1:97
 - radical political movements 1:97–98, 1:99
 - reform movements 1:99
 - research background 1:96
 - social movements 1:96
 - theoretical perspectives 1:100
 - vocational education and training (VET) 1:98
 - women's movement 1:98
 - workplace learning 1:98
- human capital theory 1:7, 1:217, 1:223, 8:255, 8:256*f*
- Iceland 5:608
- immigrant education 1:113–119
 - background information 1:113
 - basic concepts 1:115
 - educational programs 1:116
 - future directions 1:118
 - intercultural education 1:117
 - multicultural education 1:117
- importance 1:234
- informal learning
 - characteristics 1:242
 - demographic data 1:243
 - private-sector training 1:207
- Latvia 5:633
- Lithuania 5:661
- Malawi 5:670
- Malta 5:679
- Mongolia 5:683
- Namibia 5:687
- nation building 1:198–203
 - basic concepts 1:198
 - general discussion 1:198
 - importance 1:202
 - Southern African Development Community (SADC)
 - economic development 1:201
 - HIV/AIDS education 1:200
 - lifelong learning 1:201
 - political and cultural development 1:200
 - social purposes 1:199
 - socioeconomic/sociopolitical characteristics 1:199
 - survival considerations 1:200
- Norway 5:713
- outcomes analyses 1:6, 1:204–210
 - benefits 1:229, 1:230*r*
 - civic and social engagement effects 1:231
 - context-related effects 1:231
 - economic effects 1:230
 - general discussion 1:204
 - government-sponsored training 1:206
 - health effects 1:231
 - implications 1:232
 - methodologies 1:205
 - participation research 1:229
 - private-sector training 1:207
 - self-awareness 1:230
 - social and emotional outcomes 6:673–684
 - analytical framework 6:673
 - childhood/adolescence 6:674
 - community-level outcomes 6:679
 - educational inequality 6:679
 - family-level outcomes 6:678
 - general discussion 6:673, 6:677, 6:679
 - individual-level effects 6:674, 6:677
 - national-level outcomes 6:679
 - school-level effects 6:676
 - social status 1:230
- Pakistan 5:732
- participation research 1:240–250
 - barriers 1:237
 - definitions 1:242
 - demographic data 1:6, 1:243, 1:244*r*
 - economic perspectives
 - cost limitations/credit restraints 1:248
 - financial support 1:244*r*, 1:248
 - incentive-to-invest factors 1:248
 - job-related reasons 1:248

- adult education and training (AET) (*continued*)
 - motivation determinants 1:248
 - future research areas 1:249
 - general discussion 1:240
 - informal learning
 - characteristics 1:242
 - demographic data 1:243
 - motivation determinants
 - disciplinary perspectives 1:243
 - economic perspectives 1:248
 - general discussion 1:243
 - individual–social interaction model 1:249
 - psychological perspectives 1:243
 - psychosocial models 1:249
 - social perspectives 1:247
 - participation rates 1:226*t*, 1:240, 1:241*t*, 1:242*t*
 - psychological perspectives
 - attitudes 1:247
 - job-related reasons 1:247
 - motivational orientations 1:247
 - motivation determinants 1:243
 - personality traits and abilities 1:243
 - recent trends 1:240
- Poland 5:739
- political economy model 1:223–228
 - background information 1:223
 - government policies and programs
 - educational strategies 1:227
 - general discussion 1:226
 - lifelong learning 1:226
 - neoliberal ideologies and strategies 1:226
 - human capital theory 1:223
 - knowledge-based economies 1:7, 1:224
 - occupational impacts 1:223
 - skill shortages and demands 1:225, 1:226*t*
 - social constructivism 1:224
- popular adult education 1:162–168
 - basic concepts 1:162
 - common dimensions
 - action-oriented dimension 1:166
 - pedagogical–participatory dimension 1:166
 - political–ideological dimension 1:166
 - current trends 1:165
 - general discussion 1:162, 1:167
 - historical background 1:163
 - regional differences
 - general discussion 1:163
 - Latin America 1:164
 - Nordic countries 1:163
 - North America 1:165
 - socioeconomic modernization 1:218
- private-sector training 1:207
- program evaluation 3:548–554
 - design considerations 3:549
 - developmental changes and trends 3:548
 - effectiveness 3:553
 - general discussion 3:548
 - modeling approaches
 - formative assessments 3:550
 - general discussion 3:550
 - Jacobs' model 3:551, 3:552*f*
 - key questions 3:550
 - Kirkpatrick's model 3:551, 3:551*f*
 - self-evaluation 3:552
 - summative assessments 3:550
 - purpose 3:548
- Qatar 5:747
- race and ethnicity 1:77–82
 - background information 1:77
 - historical perspective
 - civil rights movement 1:77–78
 - general discussion 1:77
 - multiculturalism 1:79
 - research background 1:79
 - social justice movement 1:79
 - regional perspectives 1:81
 - Southern hemisphere 1:80
 - Western societies 1:77
- Rwanda 5:762
- Singapore 5:779
- Slovenia 5:786
- social and emotional outcomes 6:673–684
 - analytical framework
 - evidential organization 6:674
 - general discussion 6:673
 - multi-level contexts 6:673
 - community-level outcomes
 - crime rate 6:679
 - general discussion 6:679
 - educational inequality 6:679
 - family-level outcomes
 - intergenerational effects 6:678
 - post-16 participation 6:678
 - general discussion 6:673, 6:679
 - individual-level effects
 - ability grouping 6:675
 - childhood/adolescence 6:674
 - civic engagement 6:677
 - classroom management 6:675
 - confidence levels 6:678
 - curriculum 6:675
 - general discussion 6:677
 - health and well-being 6:677
 - literacy and numeracy data 6:678, 8:257
 - mentoring programs 6:675
 - peer collaboration 6:675
 - school-age signals 6:677
 - self-efficacy beliefs 6:678
 - teacher expectations 6:675
 - tutoring programs 6:675
 - national-level outcomes 6:679
 - school-level effects
 - childhood/adolescence 6:676
 - social climate 6:676
 - teaching practices 6:676
 - transition considerations 6:676
- social class 1:64–69
 - characteristics 1:64
 - educational practices 1:66
 - education–health association 1:231
 - general discussion 1:68
 - theoretical perspectives 1:65
- socioeconomic modernization 1:217–222
 - characteristics 1:217
 - community education 1:218
 - historical perspectives 1:220
 - human resource development 1:219
 - literacy education 1:217
 - popular adult education 1:218
 - workplace learning 1:219
- socioeconomic perspectives 1:7
- Solomon Islands 5:794
- South Africa 5:801, 5:802*t*
- Spain 5:810
- Sultanate of Oman 5:724
- Suriname 5:829
- Taiwan 5:836*f*, 5:837*t*, 5:838
- tension and conflict 1:9
- Thailand 5:854
- training programs 1:7
- United States 5:893
- value considerations 1:227
- Zanzibar 5:899
- adult learning 1:12–17
 - adult basic education (ABE) 1:102–106
 - future directions 1:105
 - general discussion 1:102
 - national developments 1:102
 - Swedish educational model/Komvux
 - background information 1:103
 - labor–market training programs 1:103
 - vocational-based learning (VBL) 1:104, 1:104*f*, 1:105*f*
 - background information 1:12
 - barriers 1:234–239
 - conceptual perspectives 1:234
 - large-scale surveys
 - background information 1:234
 - research findings 1:235, 1:238
 - limitations 1:238
 - qualitative studies 1:236
 - theoretical perspectives
 - bounded-agency model 1:238

- human capital theory 1:237
- participation research 1:237
- biographic perspectives 1:25–30
 - early versus adult experiences 1:29
 - general discussion 1:25, 1:29
 - historical background 1:25
 - interdisciplinary learning approaches 1:28
 - learning processes 1:26
 - paradigm approach 1:27
 - psychosocial approaches 1:28
 - reflexivity concept 1:27
 - research areas 1:26
- characteristics **1:36–41**
 - adult–child learning comparisons 1:38
 - background information 1:36
 - content learning dimension 1:38
 - general discussion 1:41
 - incentive learning dimension 1:39
 - interactive learning dimension
 - computer-assisted instruction 1:40
 - everyday learning 1:40
 - school-based learning 1:40
 - special interests 1:40
 - workplace learning 1:40
 - key features
 - barriers 1:37
 - dimensional characteristics 1:36, 1:37f
 - interaction processes 1:36, 1:37f
 - learning process types 1:37
 - psychological processes 1:36–37
 - participation rates 1:240, 1:241r
 - transformative learning theory 1:18
- context-based learning 1:15
- early education–cognition relationships 5:228
- formal learning 1:207
- funding considerations **1:133–138**
 - distribution mechanisms 1:135
 - economic principles 1:133
 - efficiency principles 1:134
 - financiers
 - general discussion 1:133
 - private financing 1:133
 - public financing 1:134
 - statistical data 1:133
 - financing mechanisms
 - direct public funding 1:135
 - employers 1:136
 - individuals 1:136
 - partnerships/co-financing 1:135
 - infrastructure 1:135
 - policies and programs
 - affordability considerations 1:137
 - co-financing programs 1:137
 - gap-closing programs 1:136
 - Organisation for Economic Cooperation and Development (OECD)
 - countries 1:136
 - promotion strategies 1:138
- general discussion 1:16
- government-sponsored training 1:206
- health considerations **1:211–216**
 - continuing medical education (CME) 1:213
 - current trends 1:211
 - general discussion 1:215
 - health literacy 1:211
- historical narratives **1:96–101**
 - characteristics 1:96
 - Enlightenment period 1:97–98
 - individualization 1:99
 - industrialization period 1:98
 - institutional histories 1:97
 - national independence movements 1:99
 - post-World War II period 1:99
 - Protestant Reformation period 1:97
 - radical political movements 1:97–98, 1:99
 - reform movements 1:99
 - research background 1:96
 - social movements 1:96
 - theoretical perspectives 1:100
 - vocational education and training (VET) 1:98
 - women's movement 1:98
 - workplace learning 1:98
- importance 1:234
- individual learning characteristics
 - andragogy 1:13
 - general discussion 1:12
 - self-directed learning 1:13
 - transformational learning 1:14
- informal learning
 - characteristics 1:242
 - demographic data 1:243
 - private-sector training 1:207
- institutional models 1:92, 1:227
- instructional strategies 1:5, 1:18
- learning to learn (L2) **3:268–272**
 - current trends
 - collective and institutional weaving concept 3:270
 - EU pilot project 3:270
 - general discussion 3:269
 - individual differences 3:269
 - research areas 3:270
 - developmental psychological traditions 3:269
 - educational impacts 3:271
 - historical background 3:268
 - motivational/volitional traditions 3:269
 - sociocultural traditions 3:269
 - theoretical perspectives 3:268
 - transfer processes 3:271
- mind–body perspectives
 - emotional experiences 1:16
 - general discussion 1:15
 - somatic knowing 1:16
 - spirituality 1:16
- museums
 - free-choice learning 1:153
 - general discussion 1:152, 1:154
 - nonformal education 1:153
 - theoretical perspectives 1:151
- outcomes analyses **1:204–210, 1:229–233**
 - benefits 1:229, 1:230r
 - civic and social engagement effects 1:231
 - context-related effects 1:231
 - economic effects 1:230
 - general discussion 1:204
 - government-sponsored training 1:206
 - health effects 1:231
 - implications 1:232
 - methodologies 1:205
 - participation research 1:229
 - private-sector training 1:207
 - self-awareness 1:230
 - social and emotional outcomes **6:673–684**
 - analytical framework 6:673
 - childhood/adolescence 6:674
 - community-level outcomes 6:679
 - educational inequality 6:679
 - family-level outcomes 6:678
 - general discussion 6:673, 6:677, 6:679
 - individual-level effects 6:674, 6:677
 - national-level outcomes 6:679
 - school-level effects 6:676
 - social status 1:230
- participation research **1:240–250**
 - barriers 1:237
 - definitions 1:242
 - demographic data
 - adult education and training (AET) 1:6, 1:243, 1:244r
 - informal learning 1:243
 - economic perspectives
 - cost limitations/credit restraints 1:248
 - financial support 1:244r, 1:248
 - incentive-to-invest factors 1:248
 - job-related reasons 1:248
 - motivation determinants 1:248
 - future research areas 1:249
 - general discussion 1:240
 - informal learning
 - characteristics 1:242
 - demographic data 1:243
 - motivation determinants
 - disciplinary perspectives 1:243
 - economic perspectives 1:248
 - general discussion 1:243

- adult learning (*continued*)
 - individual–social interaction model 1:249
 - psychological perspectives 1:243
 - psychosocial models 1:249
 - social perspectives 1:247
- outcomes analyses 1:229
- participation rates 1:226*t*, 1:240, 1:241*t*, 1:242*t*
- psychological perspectives
 - attitudes 1:247
 - job-related reasons 1:247
 - motivational orientations 1:247
 - motivation determinants 1:243
 - personality traits and abilities 1:243
- recent trends 1:240
- philosophical perspectives 6:114–120
 - andragogy 6:114
 - background information 6:114
 - competency-based learning methods 6:117
 - diversity considerations 6:116
 - experience 6:114
 - historical foundations 6:115
 - know-how knowledge 6:116
 - learning purpose 6:117
 - new model worker concept 6:117
 - practices and identities 6:118
 - propositional knowledge 6:116
 - Rylean theory 6:116
 - self-direction 6:118
- plasticity 5:229
- portfolios 3:236
- preservice education programs
 - criteria design guidelines
 - co-construction 3:413
 - evidentiary support 3:414
 - feedback 3:413–414
 - goal setting 3:414
 - general discussion 3:410
 - research areas
 - adult learning 3:412
 - feedback 3:411
 - motivation 3:411
 - professional learning 3:413
 - self-assessments 3:411
 - summative assessments 3:415
- social and emotional outcomes 6:673–684
 - analytical framework
 - evidential organization 6:674
 - general discussion 6:673
 - multi-level contexts 6:673
 - community-level outcomes
 - crime rate 6:679
 - general discussion 6:679
 - educational inequality 6:679
 - family-level outcomes
 - intergenerational effects 6:678
 - post-16 participation 6:678
 - general discussion 6:673, 6:679
 - individual-level effects
 - ability grouping 6:675
 - childhood/adolescence 6:674
 - civic engagement 6:677
 - classroom management 6:675
 - confidence levels 6:678
 - curriculum 6:675
 - general discussion 6:677
 - health and well-being 6:677
 - literacy and numeracy data 6:678, 8:257
 - mentoring programs 6:675
 - peer collaboration 6:675
 - school-age signals 6:677
 - self-efficacy beliefs 6:678
 - teacher expectations 6:675
 - tutoring programs 6:675
 - national-level outcomes 6:679
 - school-level effects
 - childhood/adolescence 6:676
 - social climate 6:676
 - teaching practices 6:676
 - transition considerations 6:676
- socioeconomic modernization 1:217–222
 - characteristics 1:217
 - community education 1:218
 - historical perspectives 1:220
 - human resource development 1:219
 - literacy education 1:217
 - popular adult education 1:218
 - workplace learning 1:219
- transformative learning theory 1:18–24
 - basic concepts 1:19
 - general discussion 1:18, 1:23
 - instructional strategies
 - critical self-reflection 1:22
 - empowerment 1:22
 - general discussion 1:21
 - support factors 1:22
 - theoretical perspectives
 - connected knowing 1:20
 - ecological view 1:21
 - general discussion 1:20
 - group and organizational transformation 1:21
 - intuition, imagination, and nurturing soul 1:21
 - social change 1:20
- see also* lifelong learning; university adult and continuing education (ACE)
- Adult Literacy and Lifeskills Survey (ALL) 7:415
- adult literacy education 1:107–112
 - assessment and accountability systems 1:110, 1:111*t*
 - background information 1:107
 - basic concepts 1:107, 1:108*t*
 - educational role 1:109
 - functional literacy
 - critical literacy 1:107
 - humanist approach 1:107
 - skill-based approach 1:107
 - general discussion 1:111
 - social-practices model
 - basic concepts 1:108
 - reconciliation methods 1:108, 1:109*f*
 - workforce influences 1:111
- adult third culture kids (ATCKs) 4:637
- Advanced Placement (AP) Program Examinations 8:51
- The Adventures of Jasper Woodbury* 8:177
- Advisory Committee on Veterinary Training (ACVT) 3:72
- affect 3:256–261, 6:544–548
 - assessment strategies
 - child behavior check list (CBCL) 3:259
 - early intervention services 3:260
 - emotional-knowledge tests 3:260
 - self-assessments 3:259
 - situational-judgment tests (SJTs) 3:260
 - social and emotional learning (SEL) 3:260
 - cognition–emotion relationship 6:547
 - components
 - cognitive aspects 6:547
 - embodied aspects 6:546
 - emotional intelligence 6:547
 - functional roles 6:544
 - definitions 6:544
 - effortful control 6:725
 - emotional intelligence 3:258, 5:186, 6:547
 - general discussion 3:256, 6:547
 - personality traits
 - basic concepts 3:256
 - coping behaviors 3:257
 - extraversion 3:258
 - neuroticism 3:256
 - test anxiety 3:257
 - positive affect negative affect scale (PANAS) 3:258
 - state–trait measures of mood 3:258
 - see also* empathy
- affiliated networks 4:512
- affirmative action
 - Brazil 1:760–766
 - academic achievement outcomes 1:764, 1:764*t*, 1:765*t*
 - black students 1:762*t*
 - course performance 1:764, 1:766*t*
 - demographic data 1:762*t*
 - ethnographic studies 1:762
 - impacts 1:763, 1:764*t*
 - public survey results 1:760
 - quota systems 1:760, 1:762
 - quota versus nonquota student performance 1:764, 1:764*t*, 1:765*t*, 1:766*t*
 - student distribution by school 1:764*t*

- white students 1:762*t*
- legal education 1:691–697
 - Educational Diversity Project (EDP) 1:694
 - general discussion 1:696
 - historical perspectives 1:691
 - social science research 1:693
- Afghanistan
 - conflict background 8:444
 - early childhood education and care (ECEC) 2:111*t*
 - higher education 4:518
 - post-war society 8:444
 - skills development programs 8:445
- Africa
 - adult education 1:80
 - classroom assessment 3:449–456
 - assessment types
 - continuous assessments 3:450, 3:451*t*
 - international assessment programs 3:449
 - national assessment programs 3:449
 - public (external) examinations 3:450
 - challenges 3:454
 - community education 1:123
 - educational perspectives 6:50–55
 - background information 6:50
 - educational priorities 6:54
 - precolonial versus postcolonial philosophies 6:50, 6:51*t*
 - traditional education
 - basic principles 6:51
 - colonialism legacy 6:54
 - communalism and consensus 6:53
 - current philosophical trends 6:52
 - ethical responsiveness 6:53
 - indigenous knowledge systems 6:52
 - ubuntu* 6:53
 - worldview versus critical activity philosophy 6:51, 6:51*t*
- enrollment rates
 - gender disparities 2:416*t*
 - primary schools 2:415*t*
 - secondary schools 2:416*t*
- higher education
 - national technical talent 1:855–859
 - institutional investments 1:858
 - internationalization processes 1:857
 - philanthropic assistance 1:857
 - recruitment and retention 1:855
 - South African Network of Skills Abroad (SANSA) 1:857–858
 - Southern African Norwegian Training, Education and Development (SANTED) initiative 1:855–856
 - study abroad programs 1:855
 - private higher education (PHE) 4:622
 - student mobility 4:561, 4:598
- labor education 1:141
- literacy probability 2:417*f*
- music teacher education 7:706
- postcolonial theory 1:302
- returns to education—cognitive skills 2:310*t*
- school choice 6:285
- school fees 2:419
- Southern African Development Community (SADC)
 - social purposes
 - background information 1:199
 - economic development 1:201
 - HIV/AIDS education 1:200
 - lifelong learning 1:201
 - political and cultural development 1:200
 - survival considerations 1:200
 - socioeconomic/sociopolitical characteristics 1:199
- student achievement test performance 2:247*f*, 2:417*t*, 2:418*t*
- sub-Saharan Africa
 - background information 3:449
 - classroom assessment 3:449–456
 - assessment types 3:449
 - background information 3:449
 - challenges 3:454
 - curriculum structure 3:450*t*
 - enrollment 3:449
 - pupil–teacher ratios (PTRs) 2:506*f*
 - student mobility 4:561
- textbook publishing 1:370
- vocational education and training (VET) 8:473
- see also* South Africa
- African Americans
 - access and equity patterns 5:52
 - achievement motivation
 - oppositional identity 1:756, 2:776
 - self-affirmation theories 1:757
 - stereotype threats 1:754, 2:776
 - valued peers 1:757
 - assimilation policies 1:575
 - cultural competence 2:565–570
 - basic concepts 2:566
 - community involvement 2:568
 - importance 2:569
 - overrepresentation 2:565
 - Pedersen's model 2:566
 - research background 2:565
 - cultural difference theory model 2:776
 - culturally and linguistically diverse (CLD) learners 2:656
 - deaf-blind population 2:623*t*
 - disadvantaged urban youth 1:806–813
 - academic achievement outcomes 1:808
 - community involvement 1:811
 - educational preparedness 1:810
 - emancipatory education 1:812
 - knowledge as power 1:811
 - male underachievement 1:808
 - phenomenological variant of ecological systems theory (PVEST) 1:807*f*, 1:808–809, 1:809*f*
 - research background 1:806
 - resilience 1:810
 - self-assessments 1:812
 - self-efficacy beliefs 1:810
 - sociocultural factors 1:806
 - support systems 1:806
 - vulnerability level/resiliency prediction model 1:806, 1:807*f*, 1:809*f*
 - diversity research 1:693, 1:825
 - early childhood education and care (ECEC) 2:64
 - educational history 1:881–885
 - Anglo-Protestant culture 1:882
 - cultural alienation and annihilation 1:575
 - culture of exclusion 1:577
 - expanded research agenda 1:578
 - general discussion 1:881, 1:884
 - globalization effects 1:573
 - internal expulsion 1:577
 - reform efforts
 - desegregation 1:692, 1:883
 - market-based reforms 1:884
 - post-civil rights era 1:884
 - post-Civil War era 1:691
 - schools as raced spaces 1:881
 - state-sanctioned racism
 - general discussion 1:882
 - intelligence tests 1:883
 - racial segregation 1:882
 - underutilization impacts 1:578
 - United States 1:691
- educational opportunities 1:881
- educational statistics 1:581–588
 - academic achievement outcomes 1:808
 - choice effects 1:585
 - continuing disparities 1:584
 - demographic diversity 1:582
 - digital technology implications 1:584
 - disparity measures 1:581, 1:874
 - dropout rates 1:874, 1:875
 - education beyond borders 1:586
 - global village concept 1:587
 - graduation rates 1:874
 - incarceration rates 1:874
 - twenty-first century perspectives 1:581, 1:587
 - United States 1:582
- education–health association 1:668–669
- higher education 1:767–771
 - background information 1:767
 - consequences 1:770
 - demographic data 1:687
 - economic returns 1:542–546
 - corporate environments 1:544
 - earnings differentials 1:543
 - economic well-being 1:545
 - income gaps 1:543

- African Americans (*continued*)
 research background 1:542
 social mobility 1:542
 stratification 1:542
 educational attainment **1:799–805**
 black males 1:800, 1:839
 completion rates 1:800, 1:803, 1:839
 dropout rates 1:803
 earnings differentials 1:804
 18–24-year-olds demographics 1:802, 1:803*f*
 enrollment rates 1:800, 1:801*f*, 1:802*f*, 1:803*f*
 female–male participation rates 1:799, 1:804
 gender gap research 1:799, 1:804
 United States 1:799
 educational decision-making 1:769
 educational experiences 1:770
 enrollment opportunities 1:656, 1:658*f*
 historically black colleges and universities (HBCUs) 1:767–768, 1:769, 4:479–480
 predominantly white universities 1:768, 4:479–480
 student learning outcomes **4:479–484**
 benefits 4:483
 cognitive complexity 4:479, 4:480*f*
 humanitarianism 4:480*f*, 4:481
 interpersonal and intrapersonal competence 4:480*f*, 4:482
 knowledge acquisition 4:480, 4:480*f*
 practical competence 4:480*f*, 4:482
 research background 4:479
 systemic racism
 consequences 1:770
 historical background 1:767
 historically black colleges and universities (HBCUs) 1:769
 predominantly white universities 1:768
 identity politics 6:39, 6:152
 information and communication technology (ICT) access and equity
 home access 8:95, 8:96*t*
 home use 8:95
 outcomes analyses 8:95, 8:97
 school access 8:96, 8:96*f*
 school use 8:96
 legal education **1:691–697**
 Educational Diversity Project (EDP) 1:694
 general discussion 1:696
 historical perspectives 1:691
 social science research 1:693
 male underachievement 1:808, 1:839
 race-specific consequences 1:876
 scientific racism **1:554–564**
 background information 1:554
 contemporary science education
 communication norms 1:559
 educational impacts 1:559
 funding and training considerations 1:560
 craniometry 1:557
 eugenics 1:558
 historical perspectives 1:555
 medical research abuse 1:557
 monogenism/polygenism 1:558
 physical anthropology 1:557
 psychometric theories 1:558
 racial evolution theories 1:555
 sociohistorical constructs 1:562
 theoretical retrofitting 1:558
 segregation policies 2:775
 slave trade 1:574
 standardized tests 2:577
 traumatic brain injury (TBI) 2:695
 Africana Studies **1:565–572**
 African-centered thought
 classification scheme 1:567, 1:568*t*
 comparison studies 1:569*t*
 framework 1:568
 anthropological and geographical considerations 1:566
 Black Studies 1:567, 1:623
 central tendencies 1:565
 conceptual perspectives 1:566
 creative works 1:569
 disciplinary perspectives 1:565
 gender research 1:568
 historical background 1:566
 instructional strategies 1:570
 methodologies 1:566
 mission 1:570
 National Council for Black Studies (NCBS) 1:568
 research publications 1:568
 study abroad programs 1:570
 African-descendant populations **1:573–580**
 Black Nationalist model 1:573
 cultural alienation and annihilation 1:575
 culture of exclusion 1:577
 expanded research agenda 1:578
 globalization effects 1:573
 international historical perspectives 1:573
 underutilization impacts 1:578
 African educational philosophy **6:50–55**
 background information 6:50
 educational priorities 6:54
 precolonial versus postcolonial philosophies 6:50, 6:51*t*
 traditional education
 basic principles 6:51
 current philosophical trends
 colonialism legacy 6:54
 communalism and consensus 6:53
 ethical responsiveness 6:53
 general discussion 6:52
 indigenous knowledge systems 6:52
 ubuntu 6:53
 worldview versus critical activity philosophy 6:51, 6:51*t*
 African National Congress (ANC) 8:443
 Afro-Brazilians **1:760–766**
 after-action review (AAR) feedback 3:228
 age- and grade-based norms 4:97
 ageless learners 8:241*f*
 aggregation bias 7:210
 aggressive behaviors 8:115
 aggressive children 6:692
 aging and cognitive abilities **5:340–344**
 brain aging 5:341
 cognition–brain relationship 5:343
 long-term prognoses 5:343
 memory performance
 age-related effects 5:342
 cognitive decline 5:341
 episodic memory 5:342
 executive functions 5:342
 familiarity 5:342
 long-term memory 5:341
 recollection 5:342
 semantic memory 5:342
 short-term memory 5:341
 working memory (WM) capacity 5:342
 neurocognitive plasticity 5:343
 neuroimaging studies 5:340
 aphasia 2:649
 Akaike Information Criterion (AIC) 7:9, 7:191, 7:193, 7:478, 7:487, 7:487*t*
 Alabama 1:654*f*
 Alaska 1:654*f*
 Alaska Native students
 deaf-blind population 2:623*t*
 educational attainment 1:802*f*, 1:803*f*
 indigenous education 1:664
see also indigenous peoples
 Albania
 comparative education societies 1:270
 early childhood education and care (ECEC) 2:111*t*
 higher education 4:517
 Alberta, Canada 8:32*t*
 alexia 2:649
 Algeria
 early childhood education and care (ECEC) 2:111*t*
 information and communication technology (ICT) 8:25*t*
 Allgemeine Hochschulreife 5:585*f*
 allied health professions **3:51–57**
 accreditation 3:52, 3:53*t*
 characteristics 3:51, 3:52*t*
 educational trends
 assessment strategies 3:56
 clinical specialization 3:56
 curriculum
 curriculum shift 3:55
 professional identity 3:55
 faculty 3:55

- institutions and degrees 3:55
- interprofessional education 3:56
- pedagogy 3:56
- historical background 3:52
- practice domains 3:53
- professional responsibilities
 - access and referral practices 3:54
 - ethical practices 3:54
 - professional turf wars 3:54
 - technical paraprofessionals 3:54
- alternate form reliability 4:28, 4:29*t*
- alternate form test–retest reliability 4:28, 4:29*t*
- altruism 2:358
- alumni involvement 2:231
- American Association for Adult Education (AAAE) 1:1–2
- American Association of Museums (AAM) 1:150
- American Association of University Professors (AAUP) 4:330
- American Bar Association (ABA) 3:25–26
- American Educational Research Association (AERA) 3:380
- American Federation of Teachers (AFT) 2:512
- American Indians *see* indigenous peoples; Native Americans
- American Society for Engineering Education (ASEE) 3:47
- American Society of Civil Engineers (ASCE) 3:47
- American Speech–Language–Hearing Association (ASHA) 2:654, 2:746, 2:835
- Americans with Disabilities Act (ADA, 1990) 2:634–635, 2:689, 2:797, 2:808
- American Veterinary Medical Association (AVMA) 3:71
- America's Choice 1:371
- amygdala 6:546
- Analogical Reasoning Learning Test (ARLT) 3:222
- analysis of covariance (ANCOVA) 7:18–24
 - basic concepts 7:5, 7:18, 7:19*f*, 7:20*f*
 - educational application 7:22, 7:23*f*
 - experimental research design 7:129, 7:129*f*
 - fixed groups 7:20
 - general discussion 7:24
 - generalizations of ANOVA 7:21
 - hypothesis testing 7:21
 - internal validity 3:94
 - multiway ANCOVA 7:21
 - potential outcomes framework 7:20
 - random groups 7:20
 - regression patterns 7:20*f*
 - statistical paradoxes 7:429
- analysis of variance (ANOVA) 7:32–36
 - basic concepts 7:5, 7:32
 - experimental research design 7:129, 7:129*f*
 - F* ratio
 - calculation 7:33
 - least significant difference (LSD) procedure 7:313
 - significance interpretation 7:34
 - generalizability theory 4:61
 - internal validity 3:94
 - least significant difference (LSD) procedure 7:313
 - meta-analysis 7:281*t*
 - multivariate linear regression 7:326
 - one-way ANOVA 7:313
 - planned comparisons 7:36
 - Rasch model 4:116*f*, 4:117, 4:117*t*
 - repeated-measures design 7:34
 - two-way ANOVA 7:34, 7:35*f*
 - type I errors 7:32, 7:313
 - variance-accounted-for measures 7:445, 7:447*t*
 - variance estimation 7:33
- analytical tradition 6:108–113
 - philosophical perspectives 6:108
 - philosophy of education
 - contemporary perspectives 6:112
 - criticisms 6:111
 - and Peters 6:109
 - postmodern/poststructural perspectives 6:111
- Anchored Instruction 8:177
- Andorra 2:111*t*
- andragogy 1:13, 6:114
- Anglo-Protestant culture 1:882
- Angola
 - socioeconomic/sociopolitical characteristics 1:199
 - student achievement test performance 2:418*t*
 - see also* Africa
- Anguilla 2:111*t*
- ANIMATE system 8:175, 8:176*f*
- Antigonish movement 1:55, 1:165
- Antigua 1:702*f*, 1:703*f*
- anti-racist education 1:739
- antisocial behaviors
 - challenges
 - demographic changes 2:820
 - fragmented initiatives 2:821
 - school failure–juvenile delinquency relationship 2:821
- neuroticism 3:256
- occurrences 2:820
- school-based prevention programs 2:820–828
 - administrative leadership/faculty buy-in
 - academic and behavioral screening 2:823
 - data systems/data-based decision making 2:824
 - evaluation practices 2:826
 - evidence-based strategies 2:825, 2:825*f*
 - general discussion 2:822
 - integration strategies 2:826
 - needs-based professional development 2:823
 - general discussion 2:826
 - goals and objectives 2:822, 2:822*t*
 - social networks 6:709
- anxiety 6:549–557
 - assessment measures 6:553*t*
 - background information 6:549
 - cognition–emotion relationship 5:240
 - cognitive performance measures 6:553
 - conceptualizations 6:550
 - evaluative anxiety
 - general discussion 6:550
 - math/computer anxieties 6:551
 - test anxiety 3:257, 6:550, 6:552*t*
 - general discussion 6:554
 - interventions 6:554, 6:555*t*
 - school leadership/student voice 4:806
 - subjective self-report measures 6:552
 - test anxiety 3:498–505
 - characteristics 3:498
 - comparison studies
 - alternate assessments 3:503*t*
 - cooperative learning activities 3:502
 - general discussion 3:501
 - online assessments 3:502
 - oral examinations 3:502
 - portfolios 3:502
 - product/project assessments 3:503
 - written tests 3:502
 - conservation of resources (COR) model
 - basic concepts 3:499, 3:500*f*
 - educational applications 3:499
 - future directions 3:503
 - general discussion 3:498
 - resource categories 3:499
 - resource loss and gain spirals 3:501, 3:501*f*
 - future directions 3:503
 - general discussion 3:504
 - implications 3:503
 - Yerkes–Dodson law 3:498, 3:499*f*
- Apel, Carl-Otto 6:65
- aphasia 2:649
- apprenticeships 5:469–475, 8:312–319
 - Australia 8:452
 - Australian case study 8:315
 - background information 8:312
 - basic concepts 8:232, 8:312
 - business/accounting education 3:41
 - challenges 8:317
 - characteristics 5:470
 - cognitive apprenticeships 3:26
 - definitions 5:470
 - Denmark 8:452
 - dual apprenticeship system 8:448–454
 - Australia 8:452
 - background information 8:448
 - Denmark 8:452
 - England 8:452
 - Germany
 - characteristics 8:449, 8:449*t*
 - dual system–vocational training relationship 8:451
 - historical background 8:450
 - working principles 8:449
 - terminology and jargon sources 8:347, 8:349*t*

- apprenticeships (*continued*)
 England 8:452
 Finland 5:570*f*, 5:572
 France 5:578
 future directions 5:474
 general discussion 5:469
 key learning features
 bodily learning 5:473
 general discussion 5:471
 imitation 5:473
 participatory learning 5:472
 quality assessments 5:474
 responsibility-based learning 5:472, 5:473*f*
 tool-based learning 5:473
 workplace learning 5:472
 learning processes
 decentered approaches 5:471
 person-centered approaches 5:471
 nursing education 3:65
 philosophical perspectives
 as a device for occupational restriction 8:313
 human capital theory 8:313
 as a means of skill formation 8:313
 occupational identity 8:313
 as a passage to adulthood 8:313
 pedagogical issues 8:314
 professions education 3:37
 school-to-work transitions 8:309
 self-study 7:689
 situated learning 5:336
 system characteristics
 employment status 8:315
 funding considerations 8:315
 general discussion 8:314
 occupation range 8:315
 proficiency examinations 8:315
 qualifications 8:314
 terminology and jargon sources 8:347, 8:349*t*
 theoretical perspectives
 anthropological approaches 5:470
 philosophical approaches 5:470
 psychological/educational approaches 5:471
 see also school-based vocational education and training
 Architect Registration Examination (ARE) 3:18
 architecture education 3:13–18
 background information 3:13
 continuing education requirements 3:18
 internships and training requirements 3:17
 licensing examinations 3:18
 professional architecture degrees
 characteristics
 admission requirements 3:16
 assessment strategies 3:17
 coursework 3:16
 design and graphics 3:16
 general education and electives 3:17
 history/theory 3:17
 practice courses 3:17
 students and faculty 3:15
 technology 3:17
 degree programs 3:14
 general discussion 3:13
 program classifications 3:13
 Argentina
 assessment system development 3:479
 early childhood education and care (ECEC) 2:111*t*
 earnings–education level relationship 2:306*t*
 higher education
 collective bargaining 4:496
 enrollment rates 1:702*f*, 1:703*f*
 information and communication technology (ICT) 8:23*t*
 public subsidization 2:308*t*
 returns to education–education level 2:308*t*
 student achievement test performance 2:247*f*, 2:417*t*
 Argyris, Chris 1:357
 Aristotle 6:3, 6:108, 6:349
 Arizona
 geographical information systems (GIS) research
 data sources and methods 1:638
 illustrative examples
 college knowledge 1:645*f*, 1:646*f*
 racial demographics 1:643*f*, 1:644*f*
 research methodologies 1:638
 Tucson elementary schools 1:639*f*, 1:643*f*, 1:644*f*
 Tucson high schools 1:641*f*, 1:645*f*, 1:646*f*
 Tucson middle schools 1:640*f*
 research implications 1:647
 tuition and needs-based grants 1:654*f*
 Arkansas 1:654*f*
 armed conflicts *see* postconflict societies; warfare
 armed forces qualifying test (AFQT) 2:294
 Armenia
 early childhood education and care (ECEC) 2:111*t*
 information and communication technology (ICT) 8:25*t*
 arousal 6:544
 articulation
 advanced writing levels 5:428
 higher education
 admittance pathways 4:612
 cross-border higher education (CBHE) 4:511
 vocational education and training (VET) 8:249
 language articulation 7:589
 speech and language impairments 2:654
 artifacts 4:784, 5:493, 5:494*f*, 6:360
 artificial intelligence (AI) 5:246
 artificial neural networks (ANNs)
 connectionism 5:477, 5:478*f*, 5:478*t*, 5:479*f*
 decision-making models 1:485, 1:485*f*, 1:486*f*
 arts curricula 1:423–427
 accountability considerations 1:425–426
 arts integration 1:426
 discipline-based arts education (DBAE) 1:424, 3:265
 general discussion 1:423
 globalization effects 1:423
 subject matter development 1:423–424
 technology-supported learning environments 1:423
 Arts Education Partnership (AEP) 1:425
 arts teacher education 7:695–701
 content knowledge 7:696
 emerging pedagogical practices
 methodologies 7:698
 pedagogical strategies 7:698
 international perspective 7:695
 new technologies 7:697
 preservice education programs 7:697
 professional and institutional concerns
 faculty backgrounds 7:699
 general discussion 7:699, 7:700
 international alliances 7:700
 recruitment 7:699
 shared knowledge potentials 7:700
 see also music teacher education
 Arturo Toscanini Complex (ATC) 6:256
 Aruba 2:111*t*
 Asia
 class size 6:200
 enrollment rates
 gender disparities 2:416*t*
 primary schools 2:415*t*
 secondary schools 2:416*t*
 faculty mobility 4:562
 higher education 4:205
 inequality–education relationship 2:273*f*
 leadership development programs 4:753*t*, 4:754*t*, 4:755*t*, 4:758*t*
 post-migration education 8:282
 private higher education (PHE) 4:623
 pupil–teacher ratios (PTRs) 2:506*f*
 research-quality assessments 4:308
 school choice 6:285
 student achievement test performance 2:247*f*, 2:417*t*
 student mobility 4:561, 4:598
 teacher education programs 7:527, 7:529
 test score–social background relationship 2:273*f*
 textbook publishing 1:370
 vocational education and training (VET) 8:515
 Asian Americans
 access and equity patterns 5:52
 achievement motivation
 oppositional identity 1:756, 2:776
 stereotype threats 1:755, 2:776
 cultural competence 2:565–570
 basic concepts 2:566

- community involvement 2:568
- importance 2:569
- overrepresentation 2:565
- Pedersen's model 2:566
- research background 2:565
- cultural difference theory model 2:776
- deaf-blind population 2:623*t*
- educational opportunities 1:881
- educational statistics 1:581–588
 - choice effects 1:585
 - continuing disparities 1:584
 - demographic diversity 1:582
 - digital technology implications 1:584
 - disparity measures 1:581
 - education beyond borders 1:586
 - 18–24-year-olds demographics 1:802, 1:803*f*
 - enrollment rates 1:596, 1:596*f*, 1:597*f*, 1:598*f*, 1:800, 1:801*f*, 1:802*f*, 1:803*f*
 - global village concept 1:587
 - twenty-first century perspectives 1:581, 1:587
 - United States 1:582
- higher education 1:595–601
 - demographic data 1:596*f*
 - educational attainment 1:598
 - enrollment rates 1:596, 1:596*f*, 1:597*f*, 1:598*f*
 - general discussion 1:595
 - participation background 1:596
 - racial and ethnic distinctions 1:598, 1:599*f*
- information and communication technology (ICT) access and equity
 - home access 8:95, 8:96*t*
 - home use 8:95
 - outcomes analyses 8:95, 8:97
 - school access 8:96, 8:96*f*
 - school use 8:96
 - research areas 1:600
- Asian Network of Training and Research Institutions in Educational Planning (ANTRIEP) 4:691
- Asian Regional Centre for Educational Planners and Administrators (ARCEPA) 4:691
- Asperger syndrome 2:579, 2:662*f*, 2:663
- Assessment Reform Group (ARG) 3:359
- assessment strategies
 - academic judgment challenges 3:382
 - accountability 3:379
 - adult literacy education 1:110, 1:111*t*
 - affect
 - child behavior check list (CBCL) 3:259
 - early intervention services 3:260
 - emotional-knowledge tests 3:260
 - self-assessments 3:259
 - situational-judgment tests (SJTs) 3:260
 - social and emotional learning (SEL) 3:260
 - allied health professions 3:56
- alternative assessments 3:202–208
 - alternative uses
 - formative assessments 3:206
 - general discussion 3:206
 - definitions 3:202
 - educational assessments
 - extended assignments/projects 3:205
 - formative assessments 3:205
 - general discussion 3:206
 - philosophical perspectives 3:204
 - portfolios 3:204
 - standards 3:204
 - summative assessments 3:205
 - teacher-based external assessments 3:205
 - formative assessments
 - diagnostic assessments 3:207, 3:276, 3:397
 - dynamic assessments 3:207
 - educational assessments 3:205
 - feedback 3:207
 - purpose 3:206
 - formats 3:203
 - modeling approaches
 - educational assessments 3:204
 - general discussion 3:203
 - performance assessments 3:203
 - performance assessments
 - background information 3:203
 - realistic simulations 3:203
 - workplace-related competency assessments 3:203
- assessment accommodations 2:530–536
 - background information 2:530
 - future directions 2:535
 - government legislation 2:531
 - individualized education plans (IEPs) 2:534
 - research areas 2:534
 - typologies 2:531, 2:532*t*
 - validity measures 2:532, 2:533*f*
- assessment conferences 3:403–409
 - basic concepts 3:404
 - constructive responses 3:275
 - definitions 3:403
 - historical background 3:404
 - informal assessment systems 3:405
 - parental involvement 3:406
 - portfolio assessments 3:406
 - pre-referral intervention strategies 2:817
 - purpose 3:403
 - recent developments
 - general discussion 3:407
 - Learning Record 3:407
 - New Zealand Education Monitoring Project (NZEMP) 3:408
 - self-assessments 3:406
 - validity 3:408
- automated essay scoring (AES) 4:20–26
 - available programs 4:21, 8:51
 - basic concepts 4:20
 - educational applications
 - automated speech-recognition scoring programs 4:24
 - Criterion online essay evaluation service 4:22
 - English language learners (ELLs) 4:23
 - general discussion 4:22
 - Text Adaptor 4:23
 - general discussion 4:25
 - historical foundations 4:20
 - modeling approaches 4:20–21
 - reliability and validity 4:22
- classroom assessment 3:209–219
 - assessment categories 3:211
 - assessment impacts 3:506–511
 - consistent results/potential benefits 3:508
 - consistent results/problematic effects 3:508
 - formative assessments 3:464, 3:510
 - general discussion 3:506
 - high-visibility external tests 3:508
 - influencing factors 3:509
 - overlapping purposes 3:506
 - research areas 3:510
 - assessment methods
 - informal classroom observations 3:217
 - paper-and-pencil tests 3:212
 - performance assessments 3:214, 3:215*f*, 3:216*f*
 - portfolio assessments 3:215
- Australia 3:417–424
 - basic concepts 3:209
 - challenges
 - classroom environments 3:464
 - fidelity and flexibility considerations 3:465
 - teachers' and students' roles 3:465
 - criteria design guidelines
 - co-construction 3:413
 - evidentiary support 3:414
 - feedback 3:413–414
 - goal setting 3:414
- curriculum-based assessment (CBA) 4:102–110
 - curriculum-based measurement (CBM) 2:572, 2:771, 4:103
 - future directions 4:109
 - mastery measurement 4:102
 - progress monitoring 4:102
- design requirements
 - assessment tasks 3:210
 - evidence interpretation 3:210
 - general discussion 3:210
 - learning goals 3:210, 3:211*f*
- Eastern Europe 3:476
 - educational impacts 4:69–74
 - formative assessments 4:73
 - general discussion 4:69
 - high-visibility external tests 4:71
 - historical perspectives 4:69
 - measurement purposes 4:69

- assessment strategies (*continued*)
 research areas 4:73
 fairness 4:56–60
 fine arts education
 criteria 3:264
 dance 3:266
 drama/theater 3:266
 music 3:265
 visual arts 3:265
 formative assessments 3:413
 French-language sources 3:432–437
 general discussion 3:209, 3:218
 historical perspectives 3:506
 Hong Kong 3:438–442
 Learning How to Learn (LHTL) project
 background information 3:270, 3:464
 classroom environments 3:464
 collaborative classroom learning 3:466
 fidelity and flexibility considerations 3:465
 importance 3:469
 supportive school policies and practices 3:466*t*, 3:467, 3:469*f*
 teachers' and students' roles 3:465
 teachers' assessment practices 3:466*t*
 teachers' professional learning practices 3:466, 3:466*t*
 New Zealand 3:443–448
 preservice education programs 3:410
 purpose 3:209
 reading skills 3:189–195
 affective aspects 3:190
 authentic assessments 3:190
 computer-assisted assessments 3:193
 fluency 3:193
 general discussion 3:194
 group discussions 3:190
 group tests 3:189
 individual tests 3:190
 informal reading inventories (IRIs) 3:191
 motivational aspects 3:190
 norm-referenced tests 3:189
 peer assessments 3:194
 phonics skills 3:191
 portfolios 3:193
 pre-reading sub-skills 3:191
 reading miscue inventories (RMIs) 3:191, 3:192*t*, 3:193*t*
 retellings 3:193
 self-assessments 3:194, 3:194*t*
 systematic observation 3:191
 research areas
 adult learning 3:412
 feedback 3:411
 motivation 3:411
 professional learning 3:413
 self-assessments 3:411
 summative assessments 3:411
 second-language learning 4:144
 student learning impacts 3:198
 summative assessments 3:211, 3:411, 3:415, 3:451*t*
 supportive school policies and practices 3:469
 technology supports 8:42–47
 basic concepts 8:42
 educational implications 8:45
 functional roles 8:43*t*
 general discussion 8:42
 large-scale commercial-based systems 8:44, 8:45*t*
 small-scale research-based systems 8:45, 8:45*t*
 validity measures 3:217
 writing instruction 3:276
 credentialing purposes 3:380
 culturally responsive evaluations 3:592–599
 critical pedagogy perspective 3:595
 cultural psychology theory 3:594
 ecological systems theory 3:593
 intervention considerations 3:593
 Kamehameha Early Education Program (KEEP) 3:596
 methodologies
 objectivist approaches 3:595
 transformative participatory approaches 3:596
 participatory action research (PAR) 3:597
 Photovoice technique 3:597–598
 research areas 3:599
 subtractive schooling 3:594
 Youth Empowerment Strategies (YES!) Program 3:597
 curriculum-based assessment (CBA) 2:571–576, 4:102–110
 curriculum-based measurement (CBM)
 basic concepts 4:103
 individualized instructional programs 4:108
 instructional decision making 4:104, 4:107*f*
 mathematics education 2:771, 4:108
 reading instruction 4:108
 Response to Intervention (RTI) 4:104
 secondary prevention strategies 4:108
 special needs students 2:572, 2:771
 student focus 4:106
 definitions 2:571
 design considerations 2:573
 effectiveness evaluations 2:574
 future directions 4:109
 general curriculum impacts 2:573
 general discussion 2:571, 2:575
 historical background 2:572
 importance 2:571
 mastery measurement 4:102
 motivational strategies 2:574
 progress monitoring 4:102
 relevant measurement models 2:573
 usage guidelines 2:575
 developmental assessment 3:373
 digital technology 8:91
 dispositions 3:181–188
 assessment modes 3:185
 assessment practices and purposes 3:186, 3:186*f*
 challenges 3:183
 characteristics 3:182–183, 3:183*f*
 competence 3:182
 conceptual perspectives 3:181
 desirable dispositions 3:183
 general discussion 3:181, 3:187
 as learning outcomes 3:185
 learning styles 3:182
 reliability 3:184
 sociohistorical background 3:182
 early childhood education 3:316–322
 background information 3:316
 general discussion 3:321
 value-driven choices
 childhood perspectives 3:316
 learning theories 3:317, 3:318*t*
 methodologies and outcomes 3:320
 purposes 3:319
 Eastern Europe 3:472–478
 classroom assessment 3:476
 examination reforms 3:473
 general discussion 3:477
 standards
 international comparative studies 3:475
 sample-based assessments 3:475
 educational assessments 3:161–171, 3:792–797
 accountability
 characteristics 3:794*t*
 importance 3:792
 interventions 3:793, 3:794*t*
 organizational level 3:793, 3:794*t*
 performance measurement 3:792
 summative assessments 3:792, 3:793*f*, 3:794*t*
 alternative assessments
 extended assignments/projects 3:205
 formative assessments 3:205
 general discussion 3:206
 philosophical perspectives 3:204
 portfolios 3:204
 standards 3:204
 summative assessments 3:205
 teacher-based external assessments 3:205
 amelioration
 characteristics 3:794*t*
 evaluation capacity building (ECB) 3:795
 formative assessments 3:793*f*, 3:794
 importance 3:793
 needs assessment 3:794
 policy-shaping contributions 3:795
 assessment activities 3:162
 assessment impacts 3:162

- cross-cutting issues
 - assessment distinctions 3:166
 - developmental support 3:169
 - formative assessments 3:165
 - models of assessment versus models of learning 3:166
 - purposes 3:165
- design guidelines 8:56–65
 - automated task generation 8:63
 - conceptual assessment frameworks (CAFs) 8:58, 8:58f, 8:60
 - delivery supports 8:63
 - domain analysis supports 8:58
 - domain modeling supports 8:59
 - evidence-centered design (ECD) model 8:58, 8:58f, 8:59t
 - evidence model 8:60
 - general discussion 8:58, 8:64
 - implementation supports 8:62
 - student model 8:60
 - task-authoring support tools 8:62
 - task model 8:61, 8:62f
- domain specificity 3:163
- future directions 3:796
- general discussion 3:170, 3:792
- key concepts and issues 3:164
- large-scale assessments 3:163
- policies and practices 3:164
- regulatory accountability 3:795
- sociopolitical contexts 3:795
- terminology 3:161
- educational impacts 4:69–74
 - formative assessments 4:73
 - general discussion 4:69
 - high-visibility external tests
 - influencing factors 4:72
 - potential benefits 4:71
 - problematic effects 4:71
 - historical perspectives 4:69
 - measurement purposes 4:69
 - research areas 4:73
- effective leaders 4:748
- fairness 4:56–60
 - assessment impacts
 - bias 4:57
 - equity 4:57
 - general discussion 4:56
 - intelligence quotients (IQs) 4:56–57
 - legal challenges 4:57
 - equity approaches
 - fairness principles 4:59t
 - important questions 4:58, 4:58t
 - test development 4:58, 4:59t
 - general discussion 4:56, 4:60
 - informal evaluations/assessments 4:59
 - large-scale assessments 4:59
- feedback 3:226–230
 - adapted feedback 3:226
 - after-action review (AAR) feedback 3:228
 - characteristics 3:226, 3:227t
 - cognitive load theory 3:227
 - computer-based systems 3:228
 - delayed feedback 3:226–227
 - dual-coding theory 3:228
 - dynamic testing 3:229
 - elaboration feedback 3:226
 - formative assessments 3:229
 - general discussion 3:226
 - immediate feedback 3:226–227
 - performance feedback 3:226–227, 3:227t, 4:747
 - timing 3:226–227
 - verification feedback 3:226
 - visual and verbal representations 3:227
- fine arts education 3:262–267
 - authentic assessments 3:262–263
 - challenges 3:262
 - classroom-based assessments
 - criteria 3:264
 - dance 3:266
 - drama/theater 3:266
 - music 3:265
 - visual arts 3:265
 - general discussion 3:266
 - performance assessments 3:262–263
- standardized assessments
 - general discussion 3:263
 - National Assessment of Educational Progress (NAEP) 3:264
 - standards movement 3:263
- formative assessments 3:359–364
 - adult education and training (AET) 3:550
 - alternative assessments
 - diagnostic assessments 3:207, 3:276, 3:397
 - dynamic assessments 3:207
 - educational assessments 3:205
 - feedback 3:207
 - purpose 3:206
 - amelioration 3:793f, 3:794
 - basic concepts 3:360, 3:378
 - challenges 3:363
 - classroom assessment 3:211, 3:229, 3:413, 3:451t, 3:464, 3:510
 - cross-cutting issues 3:165
 - current perspectives 3:359
 - early childhood education and care (ECEC) 2:89
 - educational impacts 4:69–74
 - functional role 4:73
 - general discussion 4:69
 - high-visibility external tests 4:71
 - historical perspectives 4:69
 - measurement purposes 4:69
 - research areas 4:73
 - feedback 3:229, 3:563
 - formative assessment–summative assessment continuity 3:351
 - French-language-based classroom assessments 3:434
 - functional role 4:46
 - general discussion 3:312, 7:151
 - higher education 3:249
 - implementation considerations 3:353–358
 - case-study approach 3:353
 - general discussion 3:358
 - Neartown High School case study 3:355
 - Ourtown Elementary School case study 3:353
 - principle versus practice 3:353
 - school change–system change relationship 3:358
 - instructional planning 3:365–368
 - current perspectives 3:365
 - Dewey's perspective 3:365
 - dialog 3:367
 - general discussion 3:368
 - learning regulation 3:366
 - practical guidelines 3:365
 - zone of proximal development (ZPD) 3:367
 - King's Medway Oxford Formative Assessment Project (KMOFAP) 3:360
 - learning regulation impacts 3:349, 3:366
 - legal education 3:29
 - mathematics education 2:771, 3:282, 3:653
 - Neartown High School case study
 - assessment for learning principles 3:356
 - background information 3:355
 - roadblocks 3:357
 - summative assessments 3:355
 - oral language skills 3:288
 - Ourtown Elementary School case study
 - assessment development guidelines 3:354
 - background information 3:353
 - classroom routine changes 3:355
 - school self-review 3:354
 - summative assessments 3:353
 - test results analysis 3:353
 - triangulated evidence 3:354
 - portfolios 3:245
 - practical applications
 - further learning areas 3:362
 - oral and written feedback 3:361
 - peer- and self-assessment 3:362
 - primary school science
 - evidence collection 3:294
 - evidence interpretation 3:295, 3:295t
 - feedback 3:296
 - goals 3:294
 - peer assessments 3:296
 - policies and practices 3:294
 - student self-assessments 3:296
 - records of achievement 3:245
 - reflective practices 3:332
 - research background 3:359

- assessment strategies (*continued*)
 - research-quality assessments 4:312
 - school-based prevention programs 2:826
 - secondary school science 3:300
 - self-assessments 3:232
 - Task Group on Assessment and Testing (TGAT) report 3:359
 - teacher education 3:563–571
 - assessment instruments 3:563
 - deployment stages 3:563, 3:564*f*
 - design guidelines 3:564, 3:565*t*
 - guiding principles 3:564*t*
 - informative feedback 3:563
 - intervention strategies 3:564
 - peer assessments 3:566
 - portfolios 3:565
 - purpose 3:563
 - research-based principles 3:567, 3:568*t*
 - self-assessments 3:567
 - teacher evaluations 3:675
 - technology education 3:313
 - technology supports 8:42–47
 - basic concepts 8:42
 - educational implications 8:45
 - functional roles 8:43*t*
 - general discussion 8:42
 - large-scale commercial-based systems 8:44, 8:45*t*
 - small-scale research-based systems 8:45, 8:45*t*
 - vocational education and training (VET) 3:339
 - workplace-based assessments 3:332
 - writing instruction 3:274
- gender-based learner groups 3:486–491
 - future directions 3:490
 - general discussion 3:486, 3:490
 - learning–assessment relationship
 - general discussion 3:486
 - theoretical perspectives 3:486
- sex-group differences
 - difference explanations 3:488
 - fixed-notion perspectives 3:487
 - international/regional assessment programs 3:487
 - research limitations 3:488
- sociocultural perspectives
 - assessment techniques 3:489
 - performance assessments 3:489
 - research background 3:488
 - research limitations 3:489
- group work
 - assessment criteria 3:345
 - assessment focus 3:343
 - group contributions 3:344
 - guiding principles 3:346
 - individual contributions 3:343
- higher education 3:249–255
 - basic concepts 3:249
 - general discussion 3:254
 - group assessment 3:253
 - information and communication technology (ICT) 3:253
 - quality assurance 3:252, 4:381
 - reference frameworks 3:251
 - scope 3:249
- historical background 3:512
- history/social studies/civics 3:323–329
 - Center on Organization and Restructuring Schools (CORS) standards 3:325, 3:326*t*
 - challenges 3:326
 - classroom-based assessments 3:325, 3:325*t*
 - document-based question (DBQ) assessment 3:324
 - general discussion 3:323
 - large-scale assessments 3:323
- history/social studies/civics assessments
 - achievement tests 4:175–180
 - research background 4:175, 4:657
 - standards movement 4:177
 - testing challenges 4:179
- informal evaluations/assessments 4:59
- information and communication technology (ICT) 8:6
- institutional effectiveness 3:172–180
 - basic concepts 3:173
 - challenges 3:179
 - data collection
 - educational and school process information 3:175
 - general discussion 3:174
 - student background characteristics 3:174
 - student baselines 3:174
 - student outcomes 3:174
- measurement approaches
 - data requirements 3:175, 3:175*t*
 - statistical analysis techniques 3:175, 8:92
 - value-added models (VAMs) 3:177, 3:177*f*
- raw league tables 3:172
- reform efforts 3:172
- value-added models (VAMs)
 - applications 3:178
 - basic concepts 3:173
 - research results 3:177, 3:177*f*
- instructional assessment 3:397–402
 - accountability assessments 3:398
 - diagnostic assessments 3:207, 3:276, 3:397
 - embedded assessments
 - associationist assessments 3:400
 - catechism form 3:400
 - general discussion 3:400
 - interpretive assessments 3:401, 3:401*t*
 - I–R–E sequence 3:400
 - structured talk format 3:401, 3:401*t*
 - epistemological perspectives 3:399
 - general discussion 3:397, 3:402
- intelligent systems 4:75–80
- large-scale assessments
 - computer-assisted assessments 8:48–55
 - conceptual framework 8:48
 - first-generation computer-based tests 8:48, 8:49*t*
 - general discussion 8:54
 - Generation R tests 8:49*t*, 8:52
 - next-generation electronic tests 8:49*t*, 8:50
 - technology-rich environments scenarios 8:52, 8:53*f*
 - fairness 4:59
 - history/social studies/civics 3:323
 - Latin America 3:479
 - portfolios 3:237
 - research methodologies 3:744
 - special needs students 2:752–758
 - accountability 2:755
 - assessment accommodations 2:530
 - assessment options 2:753
 - background information 2:752–753
 - general discussion 2:757
 - purpose 2:753
 - United States education system
 - historical perspectives 6:303
 - research findings 6:305
- Latin America 3:479–485
 - Argentina 3:479
 - background information 3:479
 - Bolivia 3:480
 - Brazil 3:480
 - challenges 3:484
 - Chile 3:480
 - Colombia 3:480
 - Costa Rica 3:481
 - Cuba 3:481
 - Dominican Republic 3:481
 - Ecuador 3:481
 - El Salvador 3:481
 - Guatemala 3:482
 - Honduras 3:482
 - international assessment participation 3:483
 - large-scale assessments 3:479
 - Mexico 3:482
 - Nicaragua 3:482
 - Panama 3:483
 - Paraguay 3:483
 - Peru 3:483
 - Uruguay 3:483
 - Venezuela 3:483
- learner group assessment strategies 3:486–491
 - future directions 3:490
 - general discussion 3:486, 3:490
 - learning–assessment relationship
 - general discussion 3:486
 - theoretical perspectives 3:486
 - sex-group differences

- difference explanations 3:488
 - fixed-notion perspectives 3:487
 - international/regional assessment programs 3:487
 - research limitations 3:488
- sociocultural perspectives
 - assessment techniques 3:489
 - performance assessments 3:489
 - research background 3:488
 - research limitations 3:489
- learning cities and regions 1:148
- learning regulation 3:348–352
 - basic concepts 3:348
 - educational impacts
 - feedback 3:350
 - formative assessments 3:349
 - general discussion 3:349
 - self-assessment/student involvement 3:351
 - formative assessment–summative assessment continuity 3:351
 - instructional planning 3:366
 - self-regulation models 3:348
- learning to learn (L2) 3:268–272
 - current trends
 - collective and institutional weaving concept 3:270
 - EU pilot project 3:270
 - general discussion 3:269
 - individual differences 3:269
 - research areas 3:270
 - developmental psychological traditions 3:269
 - educational impacts 3:271
 - historical background 3:268
 - motivational/volitional traditions 3:269
 - sociocultural traditions 3:269
 - theoretical perspectives 3:268
 - transfer processes 3:271
- legal education 3:29
- mathematics education 3:279–284
 - accountability 3:280, 5:404
 - future directions 3:283
 - trends
 - 1970–1990 3:279
 - 1990–2008 3:280
 - accountability testing 3:280
 - computer-based testing 3:281
 - formative assessments 3:282
 - information and communication technology (ICT) 3:281
 - summative assessments 3:281
- moderation processes 3:457–463
 - applications
 - general discussion 3:457
 - high stakes school-based certification 3:458
 - low stakes school-based certification 3:458
 - university course assessments 3:459
 - vocational training assessments 3:458
 - challenges 3:462
 - characteristics 3:457
 - consistency goals 3:460
 - design considerations 3:460
 - participative strategies 3:459
 - proactive strategies 3:459
 - research areas 3:462
 - social moderation versus statistical moderation 3:459
 - strong quality control (high stakes)
 - expert moderation 3:461
 - group/consensus moderation 3:461
 - panel moderation 3:461
 - weak quality control (low stakes)
 - assessor meetings 3:461
 - assessor partnerships and conversations 3:461
 - self-audits 3:461
- national assessment programs 4:89–94
 - see also* National Assessment of Educational Progress (NAEP)
 - background information 3:690
 - design considerations
 - across-time comparisons 4:92
 - age levels 4:91
 - data analysis and reporting 4:91
 - inclusion criteria 4:91
 - outcome selection 4:90
 - participation maximization 4:91
 - functional role 4:89
 - general discussion 4:89
 - modeling approaches
 - international/regional assessment programs 4:89
 - national samples of students 4:90
 - national student cohort testing 4:90
 - school inspections 4:90
 - national systems
 - England 4:92
 - New Zealand 4:93
 - United States 4:94
 - special needs students 3:495
- oral language skills 3:285–292
 - background information 3:285
 - characteristics
 - discourse and pragmatic skills 3:288
 - domains 3:287
 - general discussion 3:286
 - grammatical skills 3:288
 - key features 3:287*t*
 - lexical skills 3:288
 - oral-language registers 3:287
 - phonological skills 3:288
 - development and usage guidelines
 - accountability purposes 3:289
 - diagnostic assessments 3:289
 - instructional purposes 3:288
 - operationalization 3:288
 - test development 3:289
 - diverse learners 3:291
 - future directions 3:291
 - importance 3:285
 - test development
 - general discussion 3:289
 - standards-setting procedures 3:290
 - validity establishment 3:289
 - validity threats 3:290
- personal epistemology 5:214
- portfolios 3:243–248
 - adult learning 3:236
 - alternative assessments 3:204
 - assessment conferences 3:406
 - background information 3:243
 - benefits 3:241, 3:246
 - classroom assessment 3:215, 3:451*t*
 - development processes
 - general discussion 3:238
 - reflective practices 3:238
 - selection and collection 3:238
 - electronic portfolios 3:239, 3:246
 - fine arts education assessments 3:262–263, 3:264
 - future directions
 - general discussion 3:247
 - practical considerations 3:247
 - tensions and contradictions 3:248
 - general discussion 3:236
 - group work assessments 3:346
 - history/social studies/civics assessments 3:325, 3:327
 - importance 3:244
 - large-scale assessments 3:237
 - National Record of Achievement (NRA) 3:237
 - purposes 3:237
 - summative assessments 3:238
 - teacher judgment 3:238
 - test anxiety 3:502
 - United Kingdom
 - background information 3:244
 - formative assessments 3:245
 - lessons learned 3:246
 - National Record of Achievement (NRA) 3:245
 - personal development planning and progress file 3:245
 - writing instruction 3:274
- pre-referral intervention strategies 2:815–819
 - basic concepts 2:815
 - effective strategies 2:817
 - general discussion 2:818
 - implications 2:818
- primary school science 3:293–299
 - formative assessments
 - evidence collection 3:294
 - evidence interpretation 3:295, 3:295*t*
 - feedback 3:296
 - goals 3:294

- assessment strategies (*continued*)
 peer assessments 3:296
 policies and practices 3:294
 student self-assessments 3:296
 future directions 3:298
 group performance reports 3:297
 individual performance reports 3:296
 inquiry-based learning 3:293
 purpose 3:293
 summative assessments
 combination assessments 3:297
 evidence summaries 3:297
 policies and practices 3:296
 special tasks or tests 3:296
 professional architecture degrees 3:17
 purpose **3:392–396**
 basic concepts 3:392
 classification schemes 3:393
 design criteria 3:394
 fit-for-purpose 3:392
 general discussion 3:378
 multiple assessment purposes 3:395
 range and quantity 3:392
 single assessment systems 3:395
 records of achievement **3:243–248**
 background information 3:243
 future directions
 general discussion 3:247
 practical considerations 3:247
 tensions and contradictions 3:248
 importance 3:244
 portfolios
 background information 3:243
 benefits 3:246
 electronic portfolios 3:246
 formative assessments 3:245
 importance 3:244
 lessons learned 3:246
 National Record of Achievement (NRA) 3:245
 personal development planning and progress file 3:245
 United Kingdom 3:244
 transcripts 3:247
 United Kingdom
 background information 3:244
 formative assessments 3:245
 lessons learned 3:246
 National Record of Achievement (NRA) 3:237, 3:245
 personal development planning and progress file 3:245
 school-based prevention programs 2:826
 secondary school science **3:300–310**
 concept maps 3:306, 3:308*f*
 conceptual inventories 3:305, 3:306*f*, 3:307*f*
 formative versus summative assessments 3:300
 general discussion 3:309
 information technologies 3:304, 3:305*f*
 item clusters 3:305
 performance assessments 3:301, 3:303*f*
 predict–observe–explain (POE) format 3:301, 3:302*f*
 question–demonstration assessments 3:301, 3:302*f*
 science notebooks 3:309, 3:309*f*
 second-language learning **4:140–147**
 construct research 4:142
 ethical issues
 general discussion 4:143
 professionalism 4:143
 use consequences 4:143
 use justifications 4:143
 future challenges
 classroom assessment 4:144
 foreign language proficiency testing 4:144
 general discussion 4:144
 historical perspectives
 language ability as multicomponential, communicative competence 4:141
 language ability as performance 4:141
 language ability as pragmatic expectancy grammar 4:140
 skills and elements 4:140
 purpose 4:140
 self-assessments
 affect measures 3:259
 assessment conferences 3:406
 classroom assessment 3:411
 cognitions and decision-making processes 7:512
 definition 3:713
 disadvantaged urban youth 1:812
 formative assessments 3:362
 learning regulation impacts 3:351
 reading skills 3:194, 3:194*t*
 teacher education 3:567, 7:512
 writing assessments 3:276
 social influences **3:512–517**
 assessment models 3:513, 3:515*f*, 3:517*f*
 Elizabeth (case study) 3:515
 significance 3:516
 social studies/civic education programs 3:668
 special needs students
 adequate yearly progress (AYP) targets 3:495
 assessment accommodations **2:530–536**
 background information 2:530
 future directions 2:535
 government legislation 2:531
 individualized education plans (IEPs) 2:534
 research areas 2:534
 typologies 2:531, 2:532*t*
 validity measures 2:532, 2:533*f*
 challenges 3:496
 diagnostic and classification processes 2:577, 3:492
 educational impacts 3:495
 large-scale assessments **2:752–758**
 accountability 2:755
 assessment accommodations 2:530
 assessment options 2:753
 background information 2:752–753
 general discussion 2:757
 purpose 2:753
 national assessment accommodations 3:495
 pedagogical assessments 3:495
 reform impacts 3:492
 speech and language impaired students 2:658
 student learning impacts **3:196–201**
 classroom assessment 3:198
 post-assessments 3:198
 pre-assessments 3:197
 pure assessments 3:198
 quality indicators 3:199
 research areas 3:200
 research background 3:196
 students' rights **3:378–383**
 academic judgment challenges 3:382
 assessment accountability 3:379
 Code of Fair Testing Practices in Education 3:378, 3:380
 fair practice standards 3:380
 general discussion 3:382
 Standards for Educational and Psychological Testing 3:378, 3:380
 United Nations Convention on the Rights of the Child 3:378, 3:381
 summative assessments
 accountability 3:792, 3:793*f*, 3:794*t*
 adult education and training (AET) 3:550
 alternative assessments 3:205
 basic concepts 3:378
 classroom assessment 3:211, 3:411, 3:415, 3:451*t*
 cross-cutting issues 3:166
 design-based research 6:371
 diagnostic assessments 3:397
 educational impacts **4:69–74**
 general discussion 4:69
 high-visibility external tests 4:71
 historical perspectives 4:69
 measurement purposes 4:69
 research areas 4:73
 epistemological perspectives 3:399
 formative assessment–summative assessment continuity 3:351
 general discussion 3:312, 7:151
 higher education 3:249
 illustrative example 3:353, 3:355
 mathematics education 3:281, 3:653
 portfolios 3:238
 primary school science
 combination assessments 3:297
 evidence summaries 3:297
 policies and practices 3:296
 special tasks or tests 3:296
 research-quality assessments 4:312

- rubrics and scoring guides 4:124
- school-based prevention programs 2:826
- science-education programs 3:663
- secondary school science 3:300
- teacher evaluations 3:675
- teacher's role 3:384–391
 - characteristics 3:384
 - contextual perspectives 3:390
 - coursework assessment case study 3:386
 - General Certificate of Secondary Education (GCSE) 3:386
 - inference estimations 3:389
 - judgment and clarity 3:388
 - New Zealand Education Monitoring Project (NZEMP) 3:386
 - performance standard criteria 3:388
 - process model 3:387
 - purpose 3:384
 - quality assurance 3:389
 - state-wide portfolio assessment case study 3:385
 - systematic approaches 3:389
 - task conditions 3:388
 - task specification 3:388
- technology education 3:314
- technology evaluations 3:579
- vocational education and training (VET) 3:339
- technological tools 7:619
- test anxiety 3:498–505
 - characteristics 3:498
 - comparison studies
 - alternate assessments 3:503*t*
 - cooperative learning activities 3:502
 - general discussion 3:501
 - online assessments 3:502
 - oral examinations 3:502
 - portfolios 3:502
 - product/project assessments 3:503
 - written tests 3:502
 - conservation of resources (COR) model
 - basic concepts 3:499, 3:500*f*
 - educational applications 3:499
 - future directions 3:503
 - general discussion 3:498
 - resource categories 3:499
 - resource loss and gain spirals 3:501, 3:501*f*
 - future directions 3:503
 - general discussion 3:504
 - implications 3:503
 - Yerkes–Dodson law 3:498, 3:499*f*
- test-based accountability 4:153–158
 - general discussion 4:153
 - growth models
 - average gain scores 4:154
 - challenges 4:156
 - general discussion 4:154
 - random-effects models 4:155
 - student fixed-effects models 4:155
 - value tables 4:155, 4:155*f*
 - measurement issues
 - alignment analysis 4:157
 - general discussion 4:156
 - performance-level reporting 4:157
 - student inclusion 4:156
 - test preparation/score inflation 4:156
 - performance indices
 - general discussion 4:153
 - growth models 4:154
 - quasi-longitudinal measures 4:154
 - status measures 4:153
 - successive cohort measures 4:154
 - suggested guidelines 4:157
- test equating and linking 4:50–55
 - basic concepts 4:50
 - general discussion 4:50, 4:54
 - item response theory (IRT) linking
 - basic concepts 4:52
 - NEAT design 4:53
 - observed-score equating methods
 - basic concepts 4:51
 - Kernel equating method 4:52
 - Rasch model
 - general discussion 4:114
 - person–item location distributions 4:115*f*, 4:115*t*, 4:116*f*
 - scaling methods 4:51
 - United Nations Convention on the Rights of the Child 3:378, 3:381
 - vocational education and training (VET) 3:337–341
 - assessment impacts 3:340
 - criterion-referenced assessments 3:338
 - formative assessments 3:339
 - general discussion 3:337, 3:341
 - general vocational education 3:337
 - hybrid assessment model 3:338
 - norm-referenced assessments 3:338
 - purpose 3:338
 - summative assessments 3:339
 - uncertainties 3:340
 - workplace-based assessments 3:330–336
 - active knowledge transformation 3:334
 - challenges 3:335
 - competence-based assessments 3:332
 - general discussion 3:330
 - initial education and training 3:330
 - knowledge development and transformation 3:333
 - knowledge-intensive workplaces 3:331
 - lifelong learning 3:334
 - reflective formative assessments 3:332
 - writing instruction 3:273–278
 - background information 3:273
 - classroom assessment 3:276
 - constructive responses
 - assessment conferences 3:275
 - general discussion 3:275
 - peer assessments 3:276
 - self-assessments 3:276
 - written feedback 3:275
 - formative assessments 3:274
 - general discussion 3:277
 - portfolio assessments 3:274
 - testing evaluations 3:273
- assimilation learning 1:37
- assisted self-adaptive testing (ASAT) 4:149–150
- assistive technology (AT) 2:537–543
 - augmentative–alternative communication (AAC) 2:545
 - characteristics 2:538*f*, 2:538*t*
 - cognitive disabled students 2:673, 2:674*t*, 2:675*t*
 - functional role 2:537
 - general discussion 2:542
 - implementation considerations
 - activities and embedded tasks 2:539
 - framework 2:540*f*
 - general discussion 2:539
 - human demands
 - cognitive demands 2:541
 - demand responses 2:541
 - physical demands 2:540
 - sociolinguistic demands 2:541
 - tool use 2:541
 - physically disabled students 2:642
 - research areas 2:542
 - student performance 2:538*f*
- Association for Childhood Education International (ACEI) 7:728–729
- Association for Medical Education in Europe (AMEE) 3:23
- Association of Faculties of Medicine of Canada (AFMC) 3:23
- Association of MBAs (AMBA) 3:42
- Association of Southeast Asian Nations (ASEAN) 4:514
- association reversal phenomena 7:60, 7:61*t*
- asthma 1:547, 2:257, 2:426, 2:635
- athletics
 - academic achievement outcomes 1:860
 - Coach Carter* film screening 1:865
 - critical intervention programs 1:861
 - incentive and reward systems 1:864
 - popular culture 1:863
 - Scholar–Baller® program 1:860–865
 - background information 1:860
 - basic concepts 1:860
 - Coach Carter* film screening 1:865
 - critical intervention programs 1:861
 - implementation case study 1:862
 - implications 1:864
 - logos 1:863
 - ThinkMan™ 1:860, 1:861*f*, 1:863
 - ThinkWoman™ 1:860, 1:861*f*, 1:863
 - status symbols 1:861, 1:863

- attention 5:165–171
 - general discussion 5:165
 - infants
 - attention–learning models
 - compensatory systems 5:170
 - directed-attention model 5:169
 - dyadic attention 5:169
 - intentional understanding 5:170
 - plasticity 5:170
 - triadic attention/joint attention 5:170
 - behavioral phases 5:166, 5:167*f*
 - brain networks 5:167
 - face processing 5:168
 - general discussion 5:170
 - newborn systems 5:165
 - novelty factors 5:168
 - selective attention development 5:166, 5:167*f*
 - visual recognition memory 5:168
 - neuroscience research 5:314
- attention deficit/hyperactivity disorders 2:533–534
- attention deficit/hyperactivity disorders (ADHD) 2:581, 2:590, 2:636, 2:854
- attention termination 5:166
- attribution theory 6:558–563
 - background information 6:558
 - causality considerations
 - causal antecedents
 - biases 6:559
 - general discussion 6:559
 - impression management techniques 6:560
 - teacher communications 6:560
 - causal consequences
 - behaviors 6:561
 - emotions 6:561
 - success expectancies 6:561
 - causal properties 6:559
 - success–failure attributes 6:558
 - culturally influenced motivation 6:579
 - general discussion 6:562
 - historical background 6:538
- attrition
 - higher education 4:467–472
 - cost estimation 4:471
 - estimation methods 4:467
 - general discussion 4:467
 - impacts 4:471
 - retention/intervention strategies 4:470
 - theoretical perspectives 4:468
 - see also* dropouts
- atypical autism 2:663
- auditory readers 2:883
- audits 3:713
- augmentative–alternative communication (AAC) 2:544–549
 - characteristics
 - general discussion 2:545
 - manual communication 2:545, 2:624
 - technological devices 2:545
 - communication initiation
 - challenges 2:547
 - communication partners 2:548
 - use settings 2:548
 - vocabulary selection 2:548
 - vowel-based communication chart 2:549*f*
 - communication partners 2:547, 2:548
 - goals and objectives 2:544
 - needs assessment 2:546
- augmented reality (AR) interfaces 8:161
- Ausbildungsbetrieb* 8:449, 8:449*t*
- Australia 5:510–516
 - aboriginal populations 1:575
 - accreditation processes 4:384–389
 - adult basic education (ABE) 1:102–103
 - apprenticeship case study 8:315
 - Australian Employability Skills Framework
 - attributes 8:369, 8:370*t*
 - skills 8:369
 - VET-sector employability skills 8:370*t*
 - background information 5:510
 - Catholic schools 2:356, 2:358
 - childcare programs 2:49
 - childhood brain development 2:57*f*
 - classroom assessment policies 3:417–424
 - external versus classroom assessments 3:421
 - future challenges
 - increased diversity 3:422
 - lifelong learning goals 3:422
 - special needs students 3:422
 - technological tools 3:422
 - general discussion 3:423
 - influential factors
 - curriculum 3:418
 - diversity and inclusivity 3:417
 - federalization increases 3:420
 - professional teaching standards 3:418
 - school control mechanisms 3:418
 - purpose 3:417
 - school control mechanisms
 - general discussion 3:418
 - K–10 syllabi 3:419
 - outcomes-based syllabi 3:419
 - professional development support 3:420
 - senior schooling syllabi 3:419
 - student work verification 3:420
 - syllabus frameworks 3:419
- class size 2:492*f*
- community education 1:123
- comparative education societies 1:268
- crime–education relationship 1:879
- dual apprenticeship system 8:347, 8:452
- early childhood education and care (ECEC) 2:4*t*, 2:13, 2:56*t*
- earnings–education level relationship 2:307*t*
- education levels 2:279*f*
- education system
 - administrative and supervisory structure 5:514
 - challenges 5:515
 - educational finance 5:514
 - enrollment rates 2:493*f*
 - goals 5:510
 - government policies and programs 5:515
 - monitoring and evaluation programs 5:515
 - research programs 5:515
 - school effectiveness research (SER) 6:269
- structural and operational characteristics
 - adult education 5:513
 - alternative education 5:512
 - early childhood education and care (ECEC) 2:105, 2:111*t*, 2:114*t*, 5:512
 - higher education 5:511*f*, 5:513
 - nonformal education 5:513
 - preschool education 5:512
 - primary education 5:511*f*, 5:512
 - schematic diagram 5:511*f*
 - secondary education 5:511*f*, 5:512
 - special education 5:512
 - staff and students 5:512*t*
 - vocational education and training (VET) 5:512
- teaching profession 5:513
- ethical practices 4:717–718, 4:719, 4:720
- faith-based education
 - administrative and supervisory structure 5:514
 - background information 1:272
 - recent trends and conflicts 1:272
- higher education
 - binary systems 4:618
 - collective bargaining 4:496
 - enrollment rates 4:521*t*
 - federal political systems 4:606
 - foreign doctoral students 4:292*t*
 - income contingent loans (ICLs) 4:504
 - institutional expenditures 1:733*t*
 - literacy rates 4:521*t*
 - loans and grants 4:501*f*, 4:502*f*
 - parental expenditures 1:734*t*
 - professional development 4:411
 - public expenditures 4:521*t*
 - structural and operational characteristics 5:511*f*, 5:513
 - student and researcher mobility 4:287*f*
 - student experiences 4:475
 - student mobility 4:561, 4:598
 - tertiary education expenditures 4:500*f*
 - tuition fees 4:500*f*
- indigenous peoples
 - demographic data 1:664*t*
 - educational status 1:664

- historical policies and practices 1:662
 - self-determination 1:663
 - tribal colleges 1:663
 - inequality–education relationship
 - earnings differentials 2:277*t*
 - employment rate–educational attainment relationship 2:278*f*
 - gender differences 2:277*t*
 - immigrant status 2:275*f*
 - test score–social background relationship 2:273*f*
 - information and communication technology (ICT)
 - computer access at home 8:25*t*
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - mathematics and science instruction 8:32*t*
 - student proficiency and confidence 8:31*t*
 - inquiry-oriented teacher education 7:606
 - leadership development programs 4:753*t*, 4:754*t*, 4:755*t*
 - lifelong learning model 1:157*t*, 1:160
 - music teacher education 7:707
 - National Assessment Program (NAP) 3:323
 - national assessment programs 3:387
 - personalized learning 8:393
 - pharmacy education 3:59
 - physical education and sports 1:459
 - post-migration education 8:282
 - prison education programs 8:301, 8:301*t*, 8:303
 - research-quality assessments 4:311
 - school choice 6:285
 - school–higher education institution (HEI) partnerships 7:758
 - self-managing schools 5:73
 - student achievement test performance 2:247*f*
 - student attrition 4:467
 - teaching hours 2:493*f*
 - vocational education and training (VET)
 - dual apprenticeship system 8:347, 8:452
 - employability skills 8:368–375
 - employer involvement 8:457
 - initial vocational preparation system 8:251*t*
 - learning styles 8:378
 - lifelong learning model 1:157*t*, 1:160
 - National Centre for Vocational Education Research (NCVER) 8:354
 - partial completion research 8:464
 - planning and policy development 8:472
 - prison education programs 8:303
 - recognition of prior learning (RPL) 8:397, 8:398
 - school-based programs 8:247
 - special needs students 8:294–299
 - women in employment 2:56*f*
 - working mothers 2:151
 - Australian Qualifications Framework (AQF) 8:348
 - Australian Quality Training Framework (AQTF) 8:348
 - Austria
 - age-related policies and programs 8:263*t*
 - childhood brain development 2:57*f*
 - class size 2:492*f*
 - displaced worker assistance programs 8:277
 - early childhood education and care (ECEC) 2:4*t*, 2:111*t*, 2:114*t*
 - earnings–education level relationship 2:307*t*
 - education levels 2:279*f*
 - education system 5:517–523
 - administrative and supervisory structure 5:517
 - compulsory education
 - admission criteria 5:517
 - assessments 5:519
 - class size 5:518
 - curriculum 5:518
 - length of school day/week/year 5:518
 - phases 5:517, 5:519*t*
 - progression 5:519
 - special needs students 5:521, 5:521*t*
 - demographic data 5:517
 - educational finance 5:517
 - higher education
 - admission requirements 5:520
 - institution types 5:520
 - qualifications 5:520
 - reform efforts 5:522
 - language of instruction 5:517
 - primary and secondary education reforms
 - afternoon childcare 5:522
 - class heads upgrades 5:522
 - early language learning 5:522
 - five-day school week 5:522
 - gifted and talented students 5:522
 - Monday start days 5:522
 - teaching unit blocks 5:522
 - reform efforts
 - higher education 5:522
 - primary and secondary education 5:522
 - special needs students
 - assessments 5:521
 - class size 5:521
 - compulsory education 5:521, 5:521*t*
 - curriculum 5:521
 - general discussion 5:521
 - progression 5:521
 - vocational education and training (VET) 5:521
 - structural and operational characteristics
 - compulsory education 5:517, 5:519*t*
 - higher education 5:520
 - post-secondary education 5:519
 - preprimary education 5:517, 5:518*t*
 - schematic diagram 5:518*f*
 - special needs students 5:521
 - upper secondary education 5:519
 - vocational education and training (VET) 5:519, 5:521
 - teaching profession 5:521
 - upper and post-secondary education
 - admission criteria 5:519
 - assessments 5:519
 - curriculum 5:519
 - education types 5:519, 5:519*t*
 - progression 5:519
 - higher education
 - binary systems 4:617
 - foreign doctoral students 4:292*t*
 - loans and grants 4:501*f*, 4:502*f*
 - student and researcher mobility 4:287*f*
 - tertiary education expenditures 4:500*f*
 - tuition fees 4:500*f*
 - university associations 4:485
- inequality–education relationship
 - earnings differentials 2:277*t*
 - employment rate–educational attainment relationship 2:278*f*
 - fathers' educational background 2:274*f*
 - immigrant status 2:275*f*
 - test score–social background relationship 2:273*f*
- information and communication technology (ICT)
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - student proficiency and confidence 8:31*t*
- leadership development programs 4:753*t*, 4:754*t*, 4:755*t*
- participation research
 - regular education 8:284*f*
 - vocational education and training (VET) 8:283*f*, 8:284*f*
 - work-related vocational education and training (VET) 8:284*f*, 8:285*f*
- research-quality assessments 4:309
- student achievement test performance 2:247*f*
- teacher age distribution 2:492*f*
- teacher wages and salaries 2:496*f*
- teaching hours 2:493*f*
- vocational education and training (VET)
 - age-related policies and programs 8:263*t*
 - dual apprenticeship system 8:232
 - initial vocational preparation system 8:251*t*
 - participation research 8:283*f*, 8:284*f*, 8:285*f*
 - school-based programs 8:249
 - vocational–academic education parity 8:435
- women in employment 2:56*f*
- Austrian School 6:12
- authentic leadership 5:7–11
 - arenas of influence 5:9, 5:9*f*
 - characteristics 5:7
 - complications
 - cross-cultural issues 5:10
 - inauthentic/false leadership 5:10
 - general discussion 5:10
 - individual valuation processes 5:7, 5:8*f*
 - valuation processes 5:7
- authentic tests 1:171
- authoritative discourse 1:433

autism spectrum disorders (ASD) 2:661–668
 augmentative–alternative communication (AAC) 2:544–549
 characteristics
 general discussion 2:545
 manual communication 2:545
 technological devices 2:545
 communication initiation
 challenges 2:547
 communication partners 2:548
 use settings 2:548
 vocabulary selection 2:548
 vowel-based communication chart 2:549f
 communication partners 2:547, 2:548
 goals and objectives 2:544
 needs assessment 2:546
 characteristics
 Asperger syndrome 2:662f, 2:663
 autism 2:662, 2:662f
 childhood disintegrative disorder 2:662f
 general discussion 2:662
 pervasive developmental disorder—not otherwise specified 2:662f, 2:663
 Rett syndrome 2:662f
 symptoms continuum 2:662f
 historical perspectives 2:661
 research areas 2:667
 school-based services 2:854
 screening and diagnosis
 diagnostic and classification processes 2:579
 etiology 2:664
 general discussion 2:663
 prevalence 2:664
 treatment/intervention approaches 2:664, 2:666f
 autobiographical memory 5:201
 autobiographical narrative inquiry 6:437
 autobiographies, biographic perspectives 1:26
 autocratic curriculum evaluation model 1:338
 autocratic leadership 5:4
 automated essay scoring (AES) 4:20–26
 available programs 4:21, 8:51
 basic concepts 4:20
 educational applications
 automated speech-recognition scoring programs 4:24
 Criterion online essay evaluation service 4:22
 English language learners (ELLs) 4:23
 general discussion 4:22
 Text Adaptor 4:23
 general discussion 4:25
 historical foundations 4:20
 modeling approaches 4:20–21
 reliability and validity 4:22
 automated speech-recognition scoring programs 4:24
 autonomy 6:121–126
 basic concepts 6:121, 6:618, 6:621
 benefits 6:124
 criticisms 6:123
 educational perspectives 6:7, 6:121
 ethnic and women's studies programs 1:625
 freedom and responsibility 6:84
 Hobbesian perspective 6:121
 Kantian perspective 6:121
 personal autonomy 6:107
 philosophical perspectives 6:123
 autoregressive moving average (ARMA)/autoregressive integrated moving
 average (ARIMA) processes 7:477
 average calculations *see* measures of central tendency
 Ayer, AJ 6:5
 Azerbaijan
 early childhood education and care (ECEC) 2:111t
 examination systems 3:474
 information and communication technology (ICT) 8:23t

B

Bachelier, Louis 7:451
Bacillerato 5:809
 backwash effect 3:197, 3:473, 3:506
 bagging technique 7:387–388
 Bahamas
 early childhood education and care (ECEC) 2:111t
 higher education enrollment 1:702f, 1:703f

Bahrain
 early childhood education and care (ECEC) 2:111t
 information and communication technology (ICT) 8:25t
 secondary education initiatives 6:252
 Trends in International Mathematics and Science Study (TIMSS) 5:645t
 balanced-repeated replication (BRR) 7:394
 Baltic States 3:472, 3:473t
 Bandura, Albert 6:668
 Bangladesh 5:524–533
 background information 5:524
 early childhood education and care (ECEC) 2:111t
 education system
 administrative structure
 higher education 5:530
 primary level 5:529
 secondary level 5:530
 universities 5:531
 background information 5:525
 challenges 5:532
 educational finance 5:531
 educational reforms 5:532
 goals and objectives 5:525
 legal framework 5:525
 monitoring and evaluation programs 5:531
 research programs 5:531
 structural and operational characteristics
 enrollment ratios 5:528
 higher education 5:527, 5:528t
 major streams 5:526
 nonformal education 5:528
 primary education 5:526, 5:527t
 religious and spiritual education 5:527
 secondary education 5:527, 5:527t
 structural characteristics 5:526, 5:526f
 vocational education and training (VET) 5:527
 supervisory structure
 primary level 5:531
 secondary level 5:531
 teaching profession
 recruitment and retention 5:529
 training 5:528
 free school meals 2:420
 higher education
 enrollment rates 4:522t
 expansion effects 4:523
 tuition and fees 4:523
 school choice 6:285–286
 Barbados 5:534–539
 background information 5:534
 early childhood education and care (ECEC) 2:111t
 education system
 early childhood education and care (ECEC) 5:536
 educational finance 5:539, 5:539t
 Education Sector Enhancement Project (ESEP) 5:538
 government-aided programs
 Schools Lunch Pilot Programme 5:538
 school uniform grants 5:538
 TextBook Loan Scheme 5:538
 historical background 5:534
 Marriott–Mayhew Commission 5:535
 Ministry of Education administration
 general discussion 5:535
 Schools Division 5:536
 supervision and management unit 5:536
 Technical Division 5:536
 primary education
 assessment modes 5:536
 general discussion 5:536
 secondary education 5:537
 teaching profession 5:538
 tertiary education
 Barbados Community College 5:537
 Erdiston Teachers' College 5:537
 general discussion 5:537
 Samuel Jackman Prescod Polytechnic 5:537
 University of the West Indies 5:537
 higher education enrollment 1:702f, 1:703f
 historical background 1:681
 secondary education 1:679
 Barbuda 1:702f, 1:703f
 basic education for all (BEFA) 6:250

- basic interpersonal communicative skills (BICS) 5:352
 basis expansion methods 7:138, 7:138f
 Bass, Bernard 5:158
 Bateson, Gregory 6:499
 Bayesian statistical analysis 7:37–45
 basic concepts 7:7, 7:37
 Bayesian estimators 7:369, 7:456
 Bayesian inference
 basic concepts 7:37
 conjugate priors 7:39
 decision theory 7:120
 exchangeability 7:38
 exponential family 7:39
 model selection 7:292
 Bayesian networks 7:199
 Bayesian sequential decision theory
 appropriate action determinations 7:404, 7:405t
 derivation calculations 7:404
 fixed-length mastery problem 7:403
 framework 7:401
 notation 7:402
 optimal rules 7:403
 posterior predictive (Bayesian) calculations 7:404
 simulation models
 comparison strategies 7:405
 general discussion 7:404
 item response generation 7:405
 research results 7:405, 7:406t
 test types 7:405
 threshold loss
 basic concepts 7:402
 costs of testing 7:402–403
 measurement model 7:403
 parameters 7:402t
 Bayes Information Criterion (BIC) 7:9, 7:191, 7:193, 7:487
 computational statistics 7:94
 empirical Bayes methods 7:7, 7:142–149
 beta-binomial model 7:145
 educational applications 7:148
 extensions 7:143f, 7:145
 gamma-Poisson model 7:144
 National Assessment of Educational Progress (NAEP) 7:146, 7:147f, 7:148t
 normal-normal model 7:142, 7:143f, 7:144f
 school performance example 7:142, 7:143t, 7:144, 7:144f
 shrinkage effects 7:144f, 7:146, 7:147f
 small-area estimation 7:144
 extensions
 hierarchical structures 7:43
 mixed models/model selection 7:43
 random hyperparameters 7:43
 hypothesis testing 7:42
 interval estimation 7:42
 missing data handling techniques 7:286
 multiple comparison procedures 7:315
 parameter estimation
 basic concepts 7:40
 educational applications 7:41, 7:41t, 7:42f
 multinomial parameters 7:40
 statistical decision theory 7:39
 Bayh–Dole Act (1980) 1:732, 4:300, 4:362
 Beante 4:493
 Beauvoir, Simone de 6:57
 Becker, Howard 6:484
 beginning-reading programs
 combined programs 6:222
 computer-assisted instruction 6:222
 cooperative learning 6:222
 cost effectiveness 6:223
 curricula 6:221
 instructional-process programs 6:222
 kindergarten-only programs 6:223
 research background 6:221
 research results 6:223
 beginning teachers
 belief systems
 change strategies 7:625
 educational innovations 7:627
 mathematics 7:626
 research background 7:624
 teacher induction programs 7:563–568
 behavioral disorders 2:678–687
 affective traits
 basic concepts 3:256
 coping behaviors 3:257
 neuroticism 3:256
 test anxiety 3:257
 assessment strategies 2:681
 causal factors
 academic performance 2:680
 cultural influences 2:680
 family issues 2:680
 general discussion 2:680
 individual child 2:680
 characteristics
 behavioral characteristics
 externalizing behaviors 2:679
 general discussion 2:679
 internalizing behaviors 2:679
 cognitive characteristics 2:679
 general discussion 2:679
 gifted and talented students 2:590
 cognitive-behavioral interventions
 behavioral principles 2:683
 general discussion 2:683
 punishments 2:683
 reinforcers 2:683
 teacher-appropriate replacement behaviors
 general discussion 2:684
 problem solving skills 2:684
 self-regulation 2:684
 social skills training 2:684
 definitions 2:678
 early intervention services 3:260
 future directions
 alternative educational settings 2:685
 general discussion 2:685
 minority students 2:685
 suspension or expulsion 2:685
 general discussion 2:685
 Individuals with Disabilities Education Act (IDEA, 1990) 2:678, 3:260
 placement strategies 2:682
 prereferral strategies 2:681
 prevalence 2:680
 prevention strategies
 classroom discipline 2:683
 general discussion 2:682
 individual-student-centered interventions 2:683
 school-wide discipline 2:683, 2:820–828
 referral strategies
 behavioral intervention plans 2:682
 eligibility determination 2:681
 functional behavioral assessments 2:682
 behavior intervention plans (BIPs) 2:708–709
 behaviorism
 cognitive revolution 5:245
 curriculum development 1:307
 language learning theories 5:346
 learning theory 1:522
 behavior management
 early childhood education and care (ECEC) 2:89
 pre-referral intervention strategies 2:817
 prison education programs 8:300
 Belarus 2:111t
 Belgium
 age-related policies and programs 8:263t
 childhood brain development 2:57f
 classroom assessment 3:432–437
 background information 3:432
 competency-based programs 3:433
 effective communication strategies 3:434
 external testing 3:435
 formative assessments 3:434
 general discussion 3:436
 multilingual sources 3:432
 report cards 3:434
 school reform 3:432
 standards 3:435
 translated meanings 3:432
 class size 2:492f
 early childhood education and care (ECEC) 2:4t, 2:56t, 2:111t, 2:114t
 earnings-education level relationship 2:307t

- Belgium (*continued*)
 education levels 2:279*f*
 higher education
 foreign doctoral students 4:292*t*
 loans and grants 4:501*f*; 4:502*f*
 student and researcher mobility 4:287*f*
 tertiary education expenditures 4:500*f*
 tertiary type-B program survival rates 8:464*t*
 tuition fees 4:500*f*
 inequality–education relationship
 earnings differentials 2:277*t*
 employment rate–educational attainment relationship 2:278*f*
 immigrant status 2:275*f*
 test score–social background relationship 2:273*f*
 information and communication technology (ICT)
 computer access in school 8:23*t*
 frequency of computer use in school and at home 8:26*t*
 student proficiency and confidence 8:31*t*
 participation research
 regular education 8:284*f*
 vocational education and training (VET) 8:283*f*; 8:284*f*
 work-related vocational education and training (VET) 8:284*f*; 8:285*f*
 returns to education–education level 2:308*t*
 student achievement test performance 2:247*f*
 teacher shortages 2:494*f*
 teacher wages and salaries 2:496*f*
 teaching hours 2:493*f*
 vocational education and training (VET) 8:249; 8:251*t*; 8:252*f*; 8:253*f*; 8:263*t*
- belief systems 7:622–628
 characteristics 7:623
 definitions 7:622
 general discussion 7:626
 importance 7:622
 measurement approaches 7:624
 prospective teachers
 change strategies 7:625
 educational innovations 7:627
 mathematics 7:626
 research background 7:624
 relation to motivational variables and processes 7:623; 7:623*f*
- Belize
 early childhood education and care (ECEC) 2:111*t*
 higher education enrollment 1:702*f*; 1:703*f*
 student achievement test performance 2:417*t*
- Bell, Alexander Graham 2:714
- bell curve theory 1:558; 1:738; 7:471; 7:472*f*
- benchmark assessments 3:288
- Bénin
 early childhood education and care (ECEC) 2:111*t*
 student achievement test performance 2:418*t*
- Benjamini–Hochberg test 7:314
- Benjamin, Walter 6:349
- Bennis, Warren 3:43
- Bentham, Jeremy 6:76
- Bereiter–Engelmann instruction model 1:516
- Bergen Communiqué 1:773*f*; 4:567
- Berlin Communiqué 1:773*f*; 4:567
- Bernoulli distribution 7:132
- Bernstein, Basil 1:495; 1:505–506
- Berry–Esséen theorem 7:235
- Berry, Thomas 1:747–748
- Berufsakademie* 5:585*f*
Berufsfachschulen 5:585*f*
Berufsoberschule 5:585*f*
Berufsschule 8:449; 8:449*t*
- best linear unbiased estimates (BLUEs) 7:483
- beta distribution 7:98
- Betti, Emilio 6:64
- Bhutan 2:111*t*
- bias
 assessment impacts 4:57
 creativity assessment 4:172
 educational statistics 1:581–588
 choice effects 1:585
 continuing disparities 1:584
 demographic diversity 1:582
 digital technology implications 1:584
 disparity measures 1:581
 education beyond borders 1:586
 global village concept 1:587
 twenty-first century perspectives 1:581; 1:587
- United States 1:582
 standardized tests 4:36
 test adaptations 4:167
- big-fish–little-pond effect (BFLPE) 3:500
- Bildung* 6:127–132
 anthropological matrix 6:129
 basic concepts 6:127
 contemporary usage 6:130
 contextual perspectives 6:7; 6:128
 general discussion 1:163
 historical background 6:127
 reflective considerations 6:131
 semantic characteristics 6:128
- Bildungsinländer* 5:585*f*; 5:587
- bilingualism 1:475–480; 5:355–360
 academic environments
 education program types
 bilingual immersion 1:293; 1:296*f*; 1:475; 5:358
 dual bilingual programs 1:475; 5:358
 educational effectiveness 1:297
 enrichment programs 1:295; 1:296*f*
 general discussion 1:298
 heritage language programs 1:295; 1:296*f*; 1:477
 local funds of knowledge 1:477
 maintenance bilingual programs 1:294–295; 1:296*f*; 5:358–359
 Second Language (SL) programs 5:359
 sheltered content instruction 1:295–296; 5:359
 transitional bilingual programs 1:294; 1:296*f*; 5:359
 efficacy assessments 5:359
 general discussion 5:357
 theoretical constructs 5:357
 age of acquisition 5:355
 bilingual/immersion education 1:293; 1:296*f*
 biliteracy
 continua model 1:477
 general discussion 1:477
 integrated instructional strategies 1:477
 characteristics 5:355
 cognitive development effects 1:479
 culturally and linguistically diverse (CLD) learners 2:656
 dual-language programs 1:476; 5:358
 educational models 1:475
 family and community involvement
 general discussion 1:476
 heritage language programs 1:295; 1:296*f*; 1:477
 language acquisition 1:476
 local funds of knowledge 1:477
 general discussion 1:475; 5:355
 hearing impaired students 2:631
 language processing methods 5:356
 learning concerns
 codeswitching 1:478
 general discussion 1:478
 language transfer 1:478
 research areas 1:479
- second-language learning 5:419–425
 assessment strategies 4:140–147
 classroom assessment 4:144
 construct research 4:142
 ethical issues 4:143
 foreign language proficiency testing 4:144
 future challenges 4:144
 historical perspectives 4:140
 purpose 4:140
 computer-assisted instruction 8:204–210
 background information 8:204
 benefits 8:207
 classroom instruction 8:205
 English-language training demand 8:204; 8:205*f*
 functional role 8:204
 research results and analysis 8:207
 subject-specific uses 8:4
 theoretical perspectives
 cognitive approach 5:419
 contrastive linguistics 5:419
 critical-age hypothesis 5:422
 foreign language learning 5:384; 5:422
 interlanguage 5:421
 motivation 5:420
 school learning impacts 5:423
 social theory approach 5:420

- variation analysis 5:421
- speech and language impairments 2:656
- see also* foreign language learning
- Binet, Alfred
 - ability testing 3:493, 4:2
 - early childhood curriculum 1:516
 - intelligence tests 2:688, 2:870, 3:220, 4:81, 4:95, 5:188, 7:471
- binomial distributions
 - basic concepts 7:178
 - deviance measures 7:179^r
 - discrete probability distributions 7:132
 - illustrative example 7:180
 - inference estimations 7:179
 - multilevel modeling 7:213
 - negative binomial distribution 7:134
- biofeedback 6:555^r
- biographical accounts *see* narrative practices
- Biographical Inventory of Creativity 4:173
- BioKIDS 8:199
- Birmingham Centre for Cultural Studies (BCCS) 6:355
- black Americans *see* African Americans
- black and white doll study 6:499–500
- Black Book 1:9
- black Brazilians 1:760–766
- Black Nationalist model 1:573
- Black Power Movement 1:623
- Black Studies 1:567, 1:623
- Black–White test-score gap 2:274, 2:294, 2:435
- blended learning 3:534, 4:445, 8:103
- blindness and low vision 2:880–885
 - augmentative–alternative communication (AAC) 2:544–549
 - characteristics
 - general discussion 2:545
 - manual communication 2:545, 2:624
 - technological devices 2:545
 - communication initiation
 - challenges 2:547
 - communication partners 2:548
 - use settings 2:548
 - vocabulary selection 2:548
 - vowel-based communication chart 2:549^f
 - communication partners 2:547, 2:548
 - goals and objectives 2:544
 - needs assessment 2:546
- deaf-blind students 2:621–627
 - classroom intervention strategies 2:625
 - communication challenges 2:624
 - definitions 2:621
 - demographic data 2:623^r
 - developmental challenges 2:622
 - early intervention services 2:622, 2:625
 - etiology 2:624, 2:624^r
 - general discussion 2:626
 - hearing loss criteria 2:623
 - historical background 2:621
 - legal blindness criteria 2:623
 - transition processes 2:626
- educational needs 2:881
- etiologies
 - definitions 2:880
 - functional implications 2:881
 - prevalence 2:880
 - general discussion 2:884
 - government policies and programs 2:880
 - instructional needs 2:883
 - learning materials accessibility
 - auditory readers 2:883
 - Braille systems 2:883
 - closed circuit televisions (CCTVs) 2:882–883
 - general discussion 2:882
 - optical devices 2:882–883
 - occurrences 2:880
 - school-based services 2:854
 - special education background 2:715
- blocked designs 7:128
- block grant research funding 4:281
- blogs 3:534, 3:536, 4:431, 8:190, 8:366
- blood oxygenation level-dependent (BOLD) response 5:317
- Blow, Susan 1:514
- Blumer, Herbert 6:481, 6:485
- Boal, Augusto 1:164–165
- Board of Education of Oklahoma v. Dowell 2:394
- bodily gestures 5:249
- Boekaerts' model of adaptable learning 5:504
- Bolivia
 - assessment system development 3:480
 - early childhood education and care (ECEC) 2:111^r
 - earnings–education level relationship 2:306^r
 - higher education enrollment 1:702^f, 1:703^f
 - public subsidization 2:308^r
 - returns to education–education level 2:308^r
 - teacher incentives 2:482, 2:485^r
- Bologna Accord (1999) 1:772–778, 4:566–572
 - accreditation programs 3:47, 4:388
 - action lines 4:568
 - basic concepts 1:772, 4:202
 - Bologna Follow-up Group (BFUG) 1:774, 4:567
 - challenges 1:777
 - development process 1:773, 1:773^f
 - economic impacts 4:569
 - future directions 1:777, 4:571
 - goals 4:566
 - higher education
 - academic mobility 4:560
 - background information 1:772
 - binary systems 4:617–618
 - common frameworks 1:774, 4:568
 - future directions 1:777
 - Germany 5:587
 - globalization effects 1:734–735, 1:775
 - implementation challenges 1:776
 - institutional autonomy 4:342
 - internationalization processes 1:776, 4:544
 - international programs 4:570
 - quality assurance 4:569
 - state steering models 4:557–558
 - student achievement reporting 4:569
 - study program levels 4:349
 - transfer systems 4:202
 - historical background 1:772, 4:566
 - international programs 4:570
 - legal education 3:25
 - medical education 3:23
 - participating countries 1:774^t, 4:567^r
 - structural organization 4:567
 - supervision practices 7:615–616
 - teacher education 7:536
 - vocational education and training (VET) 8:435
- Bologna University 1:728, 4:201
- bonding networks 5:114
- bonding social capital 6:98
- Bonferroni inequality test 3:148, 7:313, 7:315^r
- bootstrap method 7:46–51
 - applications
 - bias corrections 7:47
 - confidence intervals
 - basic concepts 7:48
 - bootstrap percentile method 7:48
 - bootstrap-*t* method 7:48
 - centered bootstrap percentile method 7:48
 - data analysis examples 7:48, 7:49^f, 7:50^f
 - standard error approximations 7:47
 - basic concepts 7:7
 - bootstrap repeated replication 7:394
 - central limit theorem 7:47
 - computational statistics 7:96
 - cross-validation methods 7:387–388
 - jackknife versus bootstrap method 7:231
 - limitations 7:49
 - sample size considerations ($m/n \rightarrow 0$) 7:50
 - sampling distributions 7:46
- borderless education 4:509
 - see also* cross-border higher education (CBHE)
- Bosnia-Herzegovina
 - higher education 4:517
 - information and communication technology (ICT) 8:25^r
- Botswana
 - education system
 - challenges 3:454
 - classroom assessment 3:450, 3:451^t, 3:454
 - curriculum structure 3:450^r
 - exit points/continuation/exit skills 3:452^r

- Botswana (*continued*)
 indigenous curricula 1:309
 information and communication technology (ICT) 8:25*t*
 music teacher education 7:706
 socioeconomic/sociopolitical characteristics 1:199
 student achievement test performance 2:418*t*
 see also Africa
- bounded-agency model 1:238
- Bourdieu, Pierre 1:495, 1:505–506
- BRAC (Building Resources across Countries) 6:285–286, 6:533
- Braille, Louis 2:715
- Braille systems 2:624, 2:883
- brain drain 4:286
- brain research 5:310–316
 affect and emotion 6:546
 aging and cognitive abilities 5:340–344
 brain aging 5:341
 cognition–brain relationship 5:343
 long-term prognoses 5:343
 memory performance
 age-related effects 5:342
 cognitive decline 5:341
 episodic memory 5:342
 executive functions 5:342
 familiarity 5:342
 long-term memory 5:341
 recollection 5:342
 semantic memory 5:342
 short-term memory 5:341
 working memory (WM) capacity 5:342
 neurocognitive plasticity 5:343
 neuroimaging studies 5:340
 background information 5:310
 children 2:51, 2:57*f*
 creativity development 5:252
 educational effectiveness 5:53
 emotion-related processes
 cognitive stimulation 5:313, 5:314*f*
 domain specificity 5:313
 memory and attention 5:314
 social interactions 5:313, 5:314*f*
 empathy 6:602
 intelligence theories 5:187
 nature versus nurture theory 5:310
 neural networks
 developmental dyslexia 5:319
 educational skills 5:312
 mathematics skills 5:312
 reading skills 5:312
 neuroimaging technologies 5:311, 5:340
 neuroscience–education connection 5:315
 reading skills 5:317–322
 current technologies 5:317
 developmental dyslexia
 neural activation time course 5:320
 neural networks 5:319
 remediation effects 5:320
 longitudinal studies 5:321
 neuroimaging studies
 developmental dyslexia 5:319
 learning to read 5:318
 neural activation time course 5:320
 neural networks 5:319
 remediation effects 5:320
 phonological awareness 5:317
 reading development models 5:317
- branch campuses 4:512
- Brazil
 arts teacher education 7:697
 assessment system development 3:480
 class size 2:492*f*
 conditional cash transfer (CCT) programs 2:419
 early childhood education and care (ECEC) 2:105, 2:111*t*
 earnings differentials 2:296, 2:306*t*
 education levels 2:279*f*, 2:306*t*
 higher education
 academic values and expectations 4:438
 collective bargaining 4:496
 enrollment rates 1:702*f*, 1:703*f*, 4:521*t*
 institutional expenditures 1:733*t*
 literacy rates 4:521*t*
 private higher education (PHE) 4:623
 public expenditures 4:521*t*
 quota systems 1:760–766
 academic achievement outcomes 1:764, 1:764*t*, 1:765*t*
 background information 1:760
 black students 1:762*t*
 course performance 1:764, 1:766*t*
 demographic data 1:762*t*
 ethnographic studies 1:762
 impacts 1:763, 1:764*t*
 public survey results 1:760
 quota versus nonquota student performance 1:764, 1:764*t*, 1:765*t*, 1:766*t*
 student distribution by school 1:764*t*
 white students 1:762*t*
 information and communication technology (ICT) 8:23*t*
 inquiry-oriented teacher education 7:608
 moral and social reasoning 5:282
 nonformal education 1:42, 5:458, 5:458*f*, 5:459*t*
 public subsidization 2:308*t*
 returns to education–education level 2:308*t*
 school supply shortages 2:506*f*
 sociopolitical factors 1:310
 State School Isaac Schraiber/reading and writing project 8:76
 student achievement test performance 2:247*f*
 teacher incentives 2:507
 vocational education and training (VET) 8:247
- bridging networks 5:114
- bridging social capital 6:98
- Bridgman, Laura 2:621
- Bristol Achievement Test 3:189
- British Association for International and Comparative Education (BAICE) 1:252
- British Virgin Islands 2:111*t*
- Bronfenbrenner, Urie 2:72
- Broudy, Harry 1:424
- Brown & Williamson (B&W) 1:551
- Brownian motion
 financial mathematics 7:452
 fractional Brownian motion 7:452
 historical perspectives 7:451
 two-dimensional Brownian motion 7:452
- Brown, Robert 7:451
- Brown v Board of Education* (1954) 1:692, 1:883, 2:394, 2:688, 2:716, 6:36, 6:499
- Brunei Darussalam 2:111*t*
- Bruner, J 3:367
- Buber, Martin 6:598
- Buddy Project 8:128
- Bulgaria 5:540–549
 age-related policies and programs 8:263*t*
 background information 5:540
 early childhood education and care (ECEC) 2:111*t*
 education system
 educational finance 5:546
 evaluation policies and practices 5:547
 higher education
 Academic Degrees and Titles Act (1972) 5:542
 educational finance 5:546
 evaluation policies and practices 5:547
 Higher Education Act (1995) 5:542
 management policies 5:545
 Promotion of Scientific Research Act (2003) 5:542
- legislative foundations
 higher education 5:542
 public education 5:541
- management policies 5:545
- public education
 educational finance 5:546
 evaluation policies and practices 5:547
 management policies 5:545
 Professional Education and Training Act (1999) 5:542
 Public Education Act (1991) 5:541
 Public Education Act Amendments (1998) 5:541
 Public Education Act Amendments (2002) 5:541
 Public Education Act Amendments (2006) 5:541
 Public Education Act Amendments (2006) 5:541
 Study Curriculum Act (1999) 5:541
- statistical data
 higher education 5:548*t*
 higher education students 5:548*t*
 higher education teachers 5:548*t*
 number of schools 5:547*t*
 number of students 5:547*t*

- number of teachers 5:547*t*
 - private professional schools 5:548*t*
 - private professional school students 5:548*t*
 - private schools 5:548*t*
 - private school students 5:548*t*
 - private universities 5:548*t*
 - professional schools 5:548*t*
 - professional school students 5:548*t*
 - special schools 5:548*t*
 - state universities 5:548*t*
 - structural and operational characteristics
 - higher education 5:543
 - preschool education 5:542
 - primary schools 5:543
 - schematic diagram 5:543*f*
 - secondary schools 5:543
 - teacher training 5:544
 - inequality–education relationship 2:273*f*
 - information and communication technology (ICT)
 - computer access at home 8:25*t*
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - student proficiency and confidence 8:31*t*
 - information and communication technology (ICT) access and equity 8:97
 - test score–social background relationship 2:273*f*
 - vocational education and training (VET) 8:263*t*
 - bulletin boards 4:431
 - bullying 6:693, 8:116
 - Bundesinstitut für Berufsbildung* (BIBB) 8:355
 - bureaucratic curriculum evaluation model 1:338
 - Burkina Faso
 - early childhood education and care (ECEC) 2:111*t*
 - student achievement test performance 2:418*t*
 - Burns, James McGregor 5:158
 - Burundi
 - early childhood education and care (ECEC) 2:111*t*
 - student achievement test performance 2:418*t*
 - bu-shi-ban* 5:838
 - business/accounting education 3:39–44
 - background information
 - academic training 3:41
 - apprenticeships 3:41
 - educational programs 3:39
 - ethical and moral values and attitudes 3:41, 3:43
 - historical background 3:40
 - current issues
 - accreditation organizations 3:42
 - certification and licensing
 - certificate in investment performance measurement (CIPM) 3:43
 - certified public accountant (CPA) 3:42
 - chartered financial analyst (CFA) 3:43
 - financial risk management (FRM) 3:43
 - general discussion 3:42
 - globalization effects 3:43
 - professional code of conduct 3:43
 - signature pedagogy 3:43
- Butler, Judith 6:58, 6:176
- C**
- calculators 8:4, 8:136
 - California
 - Asian American/Pacific Islander student enrollment 1:597*f*
 - bilingual education 2:656
 - binary higher education systems 4:618–619
 - class size 6:201
 - cost-function analysis 2:208*t*
 - education system 1:586
 - history/civics assessments 4:177
 - leadership development programs 4:753*t*
 - Learning Record 3:407
 - San Diego City School District study 6:296
 - standards movement 4:177
 - state and school district data systems 2:183
 - tuition and needs-based grants 1:654*f*
 - California High School Exit Exam (CAHSEE) 8:51
 - California Standards Tests (CSTs) 4:177–178
 - Cambodia
 - early childhood education and care (ECEC) 2:111*t*
 - higher education
 - medium of instruction (MOI) 4:205
 - private higher education (PHE) 4:623
 - vocational education and training (VET) 8:473–474
 - Cameroun
 - early childhood education and care (ECEC) 2:111*t*
 - higher education collective bargaining 4:494
 - student achievement test performance 2:418*t*
 - CamMotion 8:179
 - Campbell's law 4:187
 - Camus, Albert 6:2–3, 6:83
 - Canada
 - Catholic schools 2:356
 - classroom assessment 3:432–437
 - background information 3:432
 - competency-based programs 3:433
 - effective communication strategies 3:434
 - external testing 3:435
 - formative assessments 3:434
 - general discussion 3:436
 - multilingual sources 3:432
 - report cards 3:434
 - school reform 3:432
 - standards 3:435
 - translated meanings 3:432
 - community education 1:123
 - community-focused schools 6:257
 - comparative education societies 1:266
 - competitive schooling 2:334
 - early childhood education and care (ECEC) 2:4*t*, 2:66–67, 2:105, 2:114*t*
 - earnings–education level relationship 2:307*t*
 - education levels 2:279*f*
 - enrollment rates 2:493*f*
 - ethical practices 4:717, 4:720
 - faith-based education 1:272
 - higher education
 - federal political systems 4:604
 - foreign doctoral students 4:292*t*
 - gender equity 1:689
 - institutional expenditures 1:733*t*
 - loans and grants 4:501*f*
 - parental expenditures 1:734*t*
 - professional development 4:411
 - student and researcher mobility 4:287*f*
 - tertiary education expenditures 4:500*f*
 - tuition fees 4:500*f*
 - university associations 4:485
 - immigrant education 1:116
 - indigenous peoples
 - demographic data 1:664*t*
 - educational status 1:664
 - historical policies and practices 1:662
 - self-determination 1:663
 - tribal colleges 1:663
 - inequality–education relationship
 - employment rate–educational attainment relationship 2:278*f*
 - immigrant status 2:275*f*
 - test score–social background relationship 2:273*f*
 - informal learning 8:241
 - information and communication technology (ICT)
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - student proficiency and confidence 8:31*t*
 - information and communication technology (ICT) access and equity 8:97
 - inquiry-oriented teacher education 7:606
 - Internet-based education 8:103
 - leadership development programs 4:753*t*, 4:754*t*, 4:755*t*, 4:758*t*
 - moral and social reasoning 5:282
 - music teacher education 7:707
 - pharmacy education 3:59
 - physical education and sports 1:459
 - popular adult education 1:165
 - prison education programs 8:300, 8:301*t*, 8:303
 - research-quality assessments 4:309
 - school effectiveness research (SER) 6:269
 - self-managing schools 5:72
 - student achievement test performance 2:247*f*
 - vocational education and training (VET)
 - employer involvement 8:457
 - initial vocational preparation system 8:251*t*
 - outcomes analyses 8:252*f*, 8:253*f*

- Canada (*continued*)
 partial completion research 8:464
 recognition of prior learning (RPL) 8:397
 school-based programs 8:247
 women in employment 2:56f
- Canadian Architectural Certification Board (CACB) 3:13
- Canadian Association of Speech–Language Pathologists and Audiologists (CASLPA) 2:654
- cancer 1:602, 1:603f, 2:637
see also health considerations
- canonical correlation analysis (CCA) 7:52–57
 background information 7:52
 basic concepts 7:52
 beef and pork consumption data example 7:54, 7:55f, 7:55t, 7:56f
 effect size measures 7:447t
 general discussion 7:330
 general use 7:56
 interpretation guidelines 7:54
- Cantelli's inequality 7:427
- capabilities approach 6:194–199
 educational contributions
 general discussion 6:196
 justice 6:197
 pedagogy and curriculum design 6:198
 value and functions 6:197
 future research areas 6:198
 normative framework
 Nussbaum's perspective 6:195
 Sen's perspective 6:194
 philosophical perspectives 6:194
- Cape Verde 2:111t
- capitalist societies 1:65, 1:220
- cardiovascular diseases *see* health considerations
- career guidance 8:269–274
 challenges 8:237–238
 Denmark 8:271
 employee development models 8:272
 government policies and programs 8:269
 human capital theory 8:269
 Iceland 8:272
 importance 8:272
 mentoring strategies 8:271
 outreach guidance 8:271
 Scotland 8:272
 standard formats 8:271
 terminology 8:270
 workplace guidance 8:271
- career stages 7:582–586
 background information 7:582
 emerging environmental changes
 external environmental factors 7:583
 general discussion 7:583
 organizational factors 7:583
 personal environmental factors 7:585
 teacher shortages 7:584
 future directions 7:585
 modeling approaches
 early models 7:582
 nonlinear dynamic models 7:583
 research-based linear models 7:582
- Caribbean countries
 access and equity patterns 5:52
 decolonization process 1:301
 earnings–education level relationship 2:306t
 higher education enrollment 1:698, 1:702f, 1:703f
 HIV/AIDS education 1:814
 pupil–teacher ratios (PTRs) 2:506f
 racialized masculinity 1:840
 secondary education 1:679–685
 Barbados 1:681
 historical background 1:679
 Jamaica 1:684
 Trinidad and Tobago 1:682
- Caribbean Examination Council (CXC) 1:680
- Caribbean Secondary Education Certificate (CSEC) 1:680
- Carl D. Perkins Vocational and Technical Education Act 2:876
- Carnegie Classification 4:619
- Carnegie Foundation for the Advancement of Teaching 3:37
- case-based education methods 7:592–597
 characteristics 7:592
 craft knowledge 7:647
 empirical bases 7:593
 purpose 7:592
 theoretical bases
 cognitive flexibility theory 7:594
 general discussion 7:594
 reasoning and reflective practice 7:595
 situated cognition theory 7:594
 social cognitive theory 7:595
 theory–practice gap 7:596
- Cases in Educational Leadership* (CEL) 4:732
- CASES (online system) 8:201
- case studies
 characteristics
 general discussion 6:456
 qualitative case studies 6:456
 typology 6:457
 general discussion 6:456, 6:461
 qualitative case studies 6:456–462
 characteristics 6:456
 design and implementation
 data collection and analysis 6:459
 general discussion 6:458
 selection criteria 6:458
 research benefits 6:461
 write-ups 6:460
- categorical context 4:704–705
- categorical data analysis 7:58–65
 characteristics 7:5
 Cochran–Mantel–Haenszel (CMH) test
 association reversal phenomena 7:60, 7:61t
 common odds ratio 7:61, 7:262, 7:262t
 cross-classification table 7:61t
 differential item functioning (DIF) 7:61
 Simpson's paradox 7:60, 7:61t, 7:262
- contingency tables
 basic concepts
 general discussion 7:58
 hierarchical log-linear models 7:59, 7:60f
 log-linear model–logistic regression analysis comparisons 7:62
 log-linear models 7:58, 7:59t
 odds ratio 7:58, 7:59t, 7:261, 7:262t
 relationship tests 7:58, 7:59t
 discrete nominal variables 7:261, 7:262t
 effect size measures 7:447t
 general discussion 7:58
 large-sample statistical methods 7:236
- latent variable models
 basic concepts 7:62
 cognitive diagnosis models (CDMs) 7:63
 differential item functioning (DIF) 7:63
 grade of membership (GoM) model 7:63
 item response theory (IRT) 7:63
 quasi-symmetry analysis 7:63
 Rasch model 7:63
 logistic regression (LR) analysis
 basic concepts 7:61
 differential item functioning (DIF) 7:62
 log-linear model–logistic regression analysis comparisons 7:62
 missing data handling techniques 7:285
- categorical responses 3:154
- Catholic schools 2:354–359
 Catholic high schools 2:237, 2:357
 civic engagement 2:237, 2:358
 competition 2:333
 demand factors 2:356
 educational market 2:355
 education effects
 academic achievement outcomes 2:357
 altruism 2:358
 civic engagement 2:237, 2:358
 competition 2:358
 economic outcomes 2:358
 educational attainment 2:357
 international networks 2:358
 religion 2:358
 risky behaviors 2:358
 enrollment 2:354
 future research areas 2:358
 geographic distribution 2:355
 historical background 2:354
 international networks 2:356

- school practices 2:355
 - student population 2:354
 - teachers 2:354
 - tuition and fees 2:355
 - vouchers 2:355
 - Cauchy distribution 7:99, 7:99^f
 - Cauchy–Schwarz inequality 7:53–54
 - causal inference 7:66–71
 - assignment mechanisms
 - formal notation 7:68
 - illustrative example 7:67, 7:67^t
 - basic concepts 7:6, 7:66
 - complications 7:70
 - external validity 6:522
 - instrumental variables 7:6, 7:223
 - modes
 - Fisher approach 7:69
 - Neyman approach 7:69
 - posterior predictive (Bayesian) approach 7:69–70
 - stable unit treatment value assumption (SUTVA) 7:66
 - Cayman Islands 2:111^t
 - Center for Research on Evaluation, Standards, and Student Testing (CRESST) 3:210, 3:228
 - Center on Organization and Restructuring Schools (CORS) standards 3:325, 3:326^t
 - Central African Republic
 - early childhood education and care (ECEC) 2:111^t
 - student achievement test performance 2:418^t
 - Central and Eastern Europe (CEE) 3:472, 3:473^t
 - Central Asia *see* Asia; Eastern Europe
 - Central Directorate Council (CODICEN) 5:875, 5:881^t
 - central limit theorem 7:47, 7:218, 7:218^f, 7:234, 7:270, 7:334, 7:469
 - central tendency measures 7:264–266
 - basic concepts
 - general discussion 7:264
 - mean 7:264
 - median 7:264
 - mode 7:264
 - data distribution 7:264
 - dispersion measures 7:265
 - general discussion 7:264
 - mean
 - alternative versions
 - general discussion 7:265
 - geometric mean 7:265
 - harmonic mean 7:265
 - robust estimation 7:266
 - trimmed mean 7:266
 - weighted mean 7:266
 - Winsorized mean 7:266
 - basic concepts 7:264
 - characteristics 7:265
 - median
 - basic concepts 7:264
 - characteristics 7:265
 - mode
 - basic concepts 7:264
 - characteristics 7:265
- Centre for Higher Education Development (CHE) rankings 4:382, 4:547
- Centre for Scholarship in Engineering Education (CSEE) 3:47
- cerebrovascular disease 1:602, 1:603^f
- certificate in investment performance measurement (CIPM) 3:43
- certification and licensing 3:1–6
 - architecture education 3:18
 - benefits 3:4
 - business/accounting education
 - certificate in investment performance measurement (CIPM) 3:43
 - certified public accountant (CPA) 3:42
 - chartered financial analyst (CFA) 3:43
 - financial risk management (FRM) 3:43
 - general discussion 3:42
 - characteristics 3:3
 - clergy education 3:35
 - computer-assisted instruction 8:186
 - effectiveness evaluations 3:5
 - general discussion 3:5
 - information and communication technology (ICT) 8:212
 - legal education 3:25
 - lifelong learning 8:186
 - medical education 3:23
 - models 3:4
 - pharmacy education 3:59
 - recognition processes 3:1
 - recruitment and retention 2:475
 - requirements 3:5
 - sociocultural factors 7:529
 - technical certifications 8:186
 - veterinary education 3:72
- certified public accountant (CPA) 3:42
- Chad
 - early childhood education and care (ECEC) 2:111^t
 - student achievement test performance 2:418^t
- chain graph models 7:199–200
- chain migration theory 1:613
- chain-of-response model 1:249
- charisma 4:784
- chartered engineers 3:46
- chartered financial analyst (CFA) 3:43
- Charter of Rights and Freedoms (Canada) 2:880
- charter schools 2:360–366
 - background information 2:360
 - funding considerations 2:337
 - general discussion 2:364
 - outcomes analyses
 - academic performance 2:362
 - competitive effects 2:363
 - funding considerations 2:361
 - general discussion 2:360
 - student characteristics 2:361
 - principal's role 5:68
- cheating 2:377
- Chebyshev's inequality 7:234, 7:426
- Chernoff's inequality 7:428
- Chicago Child–Parent Centers
 - characteristics 2:40
 - cost–benefit analysis 2:41^t
 - economic returns 2:41, 2:42^t, 2:45^t
 - outcomes analyses 2:451–452
 - public investment benefits 2:64
 - research reviews 2:40^t
 - simulation studies 2:42
 - transition programs 2:44, 2:45^t
- Chicago School
 - grounded theory 6:406
 - historical background 1:25, 6:12, 6:90–91
 - participation research 6:443
 - qualitative inquiry 6:489
 - symbolic interactionism 6:481
- Chicano students *see* Latino students
- child art/child craft 1:424
- child behavior check list (CBCL) 3:259
- childcare programs 2:100
- childhood apraxia of speech (CAS) 2:654
- childhood development
 - attention 5:165–171
 - general discussion 5:165
 - infants
 - behavioral phases 5:166, 5:167^f
 - brain networks 5:167
 - compensatory systems 5:170
 - directed-attention model 5:169
 - dyadic attention 5:169
 - face processing 5:168
 - general discussion 5:170
 - intentional understanding 5:170
 - newborn systems 5:165
 - novelty factors 5:168
 - plasticity 5:170
 - selective attention development 5:166, 5:167^f
 - triadic attention/joint attention 5:170
 - visual recognition memory 5:168
- brain development 2:51, 2:57^f
- early childhood education and care (ECEC) 2:98–103
 - basic concepts 2:98
 - center-based programs 2:100, 2:600
 - combination programs 2:101, 2:600
 - effectiveness 2:98, 5:53
 - general discussion 2:101
 - home-based programs 2:99, 2:599
 - limitations 2:102
 - New Zealand 5:694
 - parenting support and education 2:126–131
 - Early Head Start program (US) 2:128

- childhood development (*continued*)
 - Early Start program (New Zealand) 2:126
 - general discussion 2:126
 - Healthy Families America (HFA) program 2:126
 - Home Instruction Program for Preschool Youngsters (HIPPY) 2:129
 - home-visiting programs (Jamaica) 2:127
 - Incredible Years program 2:129
 - lessons learned 2:130
 - Nurse–Family Partnership (NFP) program 2:127
 - Sure Start program (UK) 2:57, 2:128, 2:158, 2:159
 - Turkish Early Enrichment Program (TEEP) 2:130
- empathic responses 6:599
- moral and social reasoning 5:279–284
 - adolescents 5:281
 - cultural influences 5:282
 - definitions 5:279
 - educational implications 5:283
 - general discussion 5:283
 - historical background 5:279
 - middle childhood 5:280
 - preschool-aged children 5:280
 - stereotypes 5:283
- parent–child relationships 5:450–456
 - autonomy support 5:451
 - early interactions 5:450
 - learning and language participation 5:451
 - out-of-home learning 5:453
 - warmth and sensitivity 5:451
- postmodern societies 2:49–54
 - childhood brain development 2:51
 - demographic effects 2:50
 - ecological perspectives 2:49
 - family structure 2:50
 - future directions 2:54
 - government policies and programs 2:52
 - technology and media effects 2:51
 - well-being and happiness 2:52
- well-being and happiness 2:32–37
 - environmental factors
 - climate change 2:34
 - environmental toxins 2:34
 - food consumption changes 2:35
 - home living environment changes 2:35
 - urban environment changes 2:35
 - familial factors
 - family structure 2:33
 - maternal employment 2:33
 - parental education 2:33
 - parenting styles 2:33
 - poverty 2:33
 - general discussion 2:37
 - influencing factors
 - community factors 2:33
 - environmental factors 2:34
 - familial factors 2:33
 - general discussion 2:32
 - genetic factors 2:32
 - social support 2:33
 - societal effects 2:33
 - long-term effects
 - cumulative effects 2:36
 - early experiences 2:35
 - latency/sleeper effects 2:35
 - pathway effects 2:36
 - postmodern societies 2:52
 - social change impacts 2:32
 - societal effects
 - health considerations 2:33
 - social capital 2:34
 - social gradients 2:34
- working mothers 2:33, 2:55, 2:151
- children
 - brain development 2:51, 2:57*f*
 - child abuse 2:17–24
 - abuse prevention 2:22
 - adverse effects 2:20
 - general discussion 2:17
 - maltreatment characteristics
 - emotional abuse 2:18
 - neglect 2:18
 - parental substance abuse 2:18
 - physical abuse 2:17
 - sexual abuse/exploitation 2:18
 - occurrences 2:17
 - protective responses
 - child welfare 2:21
 - early childhood education and care (ECEC) 2:21
 - educational responses 2:22
 - interagency cooperation 2:21
 - legislative responsibilities 2:21
 - mandatory reporting 2:22
 - risk factors
 - ecologically-based needs assessments 2:19, 2:20*f*
 - key indicators 2:18, 2:18*t*
 - screening approaches 2:18
 - structured risk assessment 2:19
 - childhood brain development 2:51
 - children's rights 1:290, 2:73, 6:17
 - constructivist learning theories 5:488
 - early childhood education assessment strategies 3:316–322
 - background information 3:316
 - general discussion 3:321
 - value-driven choices
 - childhood perspectives 3:316
 - learning theories 3:317, 3:318*t*
 - methodologies and outcomes 3:320
- childhood disintegrative disorder 2:662*f*
- childhood obesity 8:113
- Child Language Data Exchange System (CHILDES) 5:376, 6:501
- child-rearing practices 2:151–156
 - caring perspectives 2:152
 - decision-making tensions 2:155
 - diversity considerations 2:151
 - general discussion 2:155
 - marketized childcare 2:155
 - multidisciplinary/multiprofessional work 2:154
 - parental involvement 2:154
- parenting style–educational outcome relationship 2:153
- psychoanalytical research 6:92–96
 - background information 6:92
 - educational implications 6:94
 - Freudian perspective 6:92
 - Lacanian perspective 6:93
- social class effects
 - academic achievement 2:425
 - after-school activities 2:424
 - literacy gaps 2:423
 - role models 2:424–425
- training programs 2:153
- well-being and happiness 2:32–37
 - environmental factors
 - climate change 2:34
 - environmental toxins 2:34
 - food consumption changes 2:35
 - home living environment changes 2:35
 - urban environment changes 2:35
 - familial factors
 - family structure 2:33
 - maternal employment 2:33
 - parental education 2:33
 - parenting styles 2:33
 - poverty 2:33
 - general discussion 2:37
 - influencing factors
 - community factors 2:33
 - environmental factors 2:34
 - familial factors 2:33
 - general discussion 2:32
 - genetic factors 2:32
 - social support 2:33
 - societal effects 2:33
 - long-term effects
 - cumulative effects 2:36
 - early experiences 2:35
 - latency/sleeper effects 2:35
 - pathway effects 2:36
 - postmodern societies 2:52
 - social change impacts 2:32
 - societal effects
 - health considerations 2:33
 - social capital 2:34
 - social gradients 2:34

- purposes 3:319
 - explanatory talk 5:464–468
 - causal explanations 5:465
 - definitions 5:464
 - diachronic explanations 5:467, 5:467f
 - discourse genres 5:464
 - extended discourse 5:465
 - general discussion 5:468
 - narrative thinking 5:464
 - social interaction theory 5:466
 - synchronic explanations 5:467, 5:467f
 - theoretical perspectives 5:464
 - feeding and eating disorders 2:581
 - friendship 6:564–569
 - characteristics 6:564
 - intervention strategies 6:568
 - peer acceptance 6:565
 - school adjustment effects 6:565
 - social networks 6:706
 - social tasks
 - conflict management skills 6:567
 - formation processes 6:566
 - friendship transgressions 6:567
 - general discussion 6:566
 - help-giving skills 6:567
 - reliable alliances 6:567–568
 - self-disclosure 6:568
 - sociometric measures 6:564
 - indigenous curricula 1:307–311
 - Convention on the Rights of the Child 1:307
 - educational traditions 1:307
 - Gandhian model 1:307
 - national aims 1:310
 - national–local educational dilemma 1:308
 - relevance–quality–efficiency factors 1:309
 - sociopolitical factors 1:310
 - vocational education and training (VET) 1:309
 - language development 2:745–751
 - developmental milestones 2:747
 - effective communication skills 2:745, 2:746r
 - family involvement 2:749
 - general discussion 2:749
 - language difference versus language disorder 2:747
 - proactive strategies 2:749
 - serious emotional disturbances (SEDs) 2:748
 - specific disability impacts 2:748
 - moral and social reasoning 5:279–284
 - adolescents 5:281
 - cultural influences 5:282
 - definitions 5:279
 - educational implications 5:283
 - general discussion 5:283
 - historical background 5:279
 - middle childhood 5:280
 - preschool-aged children 5:280
 - stereotypes 5:283
 - music learning 5:407
 - narrative socialization
 - conarration 5:304
 - early narrative stages 5:303
 - genres 5:303
 - peer relationships 2:794–801
 - chronological age 2:795
 - cultural influences 2:797
 - gender-related differences 2:795
 - general discussion 2:794, 2:798
 - inclusive interventions 2:798
 - program integration effects 2:797
 - research perspectives 2:794
 - social networks 6:705–711
 - academic performance 6:709
 - adolescents 6:707
 - characteristics 6:705
 - deviant behaviors 6:709
 - research approaches 6:706
 - young children 6:706
 - postmodern societies 2:49–54
 - childhood brain development 2:51
 - demographic effects 2:50
 - ecological perspectives 2:49
 - family structure 2:50
 - future directions 2:54
 - government policies and programs 2:52
 - technology and media effects 2:51
 - well-being and happiness 2:52
 - social development 6:585–590
 - basic concepts 6:585
 - characteristics
 - conflict management skills 6:587–588
 - friendship 6:587
 - joint attention 6:586
 - peer relationships 6:587, 6:706
 - social competence 6:589
 - social play 6:588
 - task behavior 6:589
 - temperament 6:588
 - general discussion 6:589
 - research background 6:586
 - speech and language impairments 2:654–660
 - culturally and linguistically diverse (CLD) learners
 - African-American English dialect speakers 2:656
 - bilingual students 2:656
 - demographic data 2:656
 - individualized education plans (IEPs) 2:658, 2:658f
 - Individuals with Disabilities Education Act (IDEA, 1990) 2:657
 - disorder characteristics
 - articulation 2:654
 - childhood apraxia of speech (CAS) 2:654
 - general discussion 2:654
 - language-based learning disabilities (LBLD) 2:655
 - language disorders 2:655
 - orofacial myofunctional disorders (OMD) 2:654
 - phonological disorders 2:654
 - stuttering 2:655
 - voice disorders 2:655
 - future directions
 - evidence-based practices 2:659
 - general discussion 2:658
 - technology-supported learning environments 2:658
 - testing and assessments 2:658
 - general discussion 2:654, 2:659
 - traumatic brain injury (TBI) 2:695
 - well-being and happiness 2:32–37
 - environmental factors
 - climate change 2:34
 - environmental toxins 2:34
 - food consumption changes 2:35
 - home living environment changes 2:35
 - urban environment changes 2:35
 - familial factors
 - family structure 2:33
 - maternal employment 2:33
 - parental education 2:33
 - parenting styles 2:33
 - poverty 2:33
 - general discussion 2:37
 - influencing factors
 - community factors 2:33
 - environmental factors 2:34
 - familial factors 2:33
 - general discussion 2:32
 - genetic factors 2:32
 - social support 2:33
 - societal effects 2:33
 - long-term effects
 - cumulative effects 2:36
 - early experiences 2:35
 - latency/sleeper effects 2:35
 - pathway effects 2:36
 - postmodern societies 2:52
 - social change impacts 2:32
 - societal effects
 - health considerations 2:33
 - social capital 2:34
 - social gradients 2:34
- Children Bill and Children Act (2004) 2:158
- Children's Scholarship Fund (CSF) 2:383
- Children's Workforce Development Council (CWDC) 2:159
- Children with Specific Learning Disabilities Act 2:647
- Chile
 - assessment system development 3:480
 - Catholic schools 2:356

- Chile (*continued*)
 class size 2:492*f*
 competency-based learning model 8:407
 competitive schooling 2:335
 early childhood education and care (ECEC) 2:111*t*
 earnings–education level relationship 2:306*t*
 educational vouchers 2:338, 2:382*t*, 2:384
 education levels 2:279*f*
 education system 7:526
 higher education
 collective bargaining 4:496
 enrollment rates 1:702*f*, 1:703*f*
 institutional expenditures 1:733*t*
 parental expenditures 1:734*t*
 private higher education (PHE) 4:623
 inequality–education relationship 2:273*f*
 information and communication technology (ICT)
 computer access in school 8:23*t*
 frequency of computer use in school and at home 8:26*t*
 mathematics and science instruction 8:32*t*
 student proficiency and confidence 8:31*t*
 neoliberal ideologies and strategies 6:12
 private schools 2:337
 public subsidization 2:308*t*
 returns to education–cognitive skills 2:310*t*
 returns to education–education level 2:308*t*
 school choice 6:285
 sociopolitical factors 1:310
 student achievement test performance 2:247*f*, 2:417*t*
 teacher education programs 7:526–527
 teacher incentives 2:484, 2:485*t*, 2:507
 teacher training 2:509
 teaching hours 2:493*f*
 test score–social background relationship 2:273*f*
 voucher systems 2:338, 2:370, 2:382*t*, 2:384, 2:419–420
- China
 accreditation processes 4:384–389
 Chinese in Japan 1:789
 class size 6:200
 comparative education societies 1:269
 digital technology 1:390
 early childhood education and care (ECEC) 2:93–97
 enrollment rates 2:111*t*
 future directions 2:96
 historical background 2:93
 reform effects
 consolidation processes 2:94
 funding considerations 2:95, 2:95*t*
 private providers 2:95, 2:95*t*
 transitions and uncertainties 2:96
 elder learning organizations 1:179, 1:180
 ethnic minorities
 cultural perspectives 1:616
 educational policies 1:616–622
 case studies 1:617
 challenges 1:620
 ethnic pluralism 1:619
 historical perspectives 1:619
 Koreans 1:619
 Mongols 1:618
 Naxi 1:618–619
 preferential treatment 1:616
 sociocultural factors 1:617
 Tibetans 1:618
 Uyghurs 1:618
 faith-based education 1:273
 higher education
 collective bargaining 4:495
 English-delivered instruction 4:205
 enrollment rates 4:521*t*, 4:522*t*
 expansion effects 4:520
 faculty mobility 4:562
 gender equity 1:689
 income sources 4:521*t*
 literacy rates 4:521*t*
 partnerships 4:521
 private higher education (PHE) 4:623
 public expenditures 4:521*t*
 quality assurance 4:525
 student mobility 4:521, 4:561, 4:598
 inequality–education relationship 2:275*f*
 information and communication technology (ICT) 8:23*t*
 information and communication technology (ICT) access and equity 8:98
 inquiry-oriented teacher education 7:606
 leadership development programs 4:753*t*, 4:754*t*, 4:755*t*
 mathematics curricula 1:454
 moral and social reasoning 5:282
 motivational concepts 6:581
 music teacher education 7:706
 pharmacy education 3:59
 post-migration education 8:282
 research-quality assessments 4:308
 school choice 6:285
 secondary education initiatives 6:253
 STEM (science, technology, engineering, and mathematics) 1:838
 student achievement test performance 2:247*f*
 teacher education programs 7:527, 7:529
 teacher evaluations 3:677
 textbook publishing 1:370
 vocational education and training (VET) 8:247
- chi-squared (χ^2) distribution 7:99, 7:104, 7:290
 Chladenius, Johan Martin 6:414
 Chomsky, Noam 5:246, 5:382–383
 Christian homeschooling 6:216
 chronic diseases 1:602
 chronic motor disorders 2:582
 citizenship education 1:113–119
 assessment strategies 3:326–327
 background information 1:113
 basic concepts
 agency 1:115
 civic virtues 1:114–115
 general discussion 1:113
 identity 1:114
 status 1:114
 challenges 7:720
 economic sustainability 8:418
 educational programs 1:116
 future directions 1:118
 higher education 1:866–873
 background information 1:866
 democratic competency development
 campus experiences 1:868
 long-term effects 1:870
 meaningful initiatives 1:870
 multicultural competence 1:869
 student attitudes, values, and behaviors 1:869
 young adult participation 1:868
 diversity exposure 1:867
 general discussion 1:872
 service opportunity recommendations 1:870
 intercultural education 1:117
 moral values 7:651
 multicultural education 1:117
 philosophical perspectives 6:7
 rational autonomy 6:122
 social capital 6:677
 social studies curriculum 1:411
 social sustainability 8:419
 city–regions *see* learning cities and regions
 Civic Education Study (CIVED) 4:656–662
 characteristics 3:687*t*
 design and implementation 4:657, 4:658*t*
 educational impacts 4:667
 findings implications and utilization 4:660
 general discussion 4:662
 historical foundations 4:657
 participating countries 4:658*t*
 results dissemination 4:660
 survey results and analysis 4:658
 theoretical framework 4:659*f*
 civic engagement 2:235–238
 adult learning outcomes 1:231
 economic perspectives 2:235
 educational attainment correlations 2:235
 future directions 2:238
 higher education 1:866–873, 4:398–403
 background information 1:866
 challenges 4:402
 community partnerships 4:390–397
 background information 4:390
 higher-education drivers 4:392

- higher-education governance barriers 4:395
- national policy barriers 4:393
- regional development drivers 4:390
- regional governance barriers 4:395
- science-driven models 4:392–393
- democratic competency development
 - campus experiences 1:868
 - long-term effects 1:870
 - meaningful initiatives 1:870
 - multicultural competence 1:869
 - student attitudes, values, and behaviors 1:869
 - young adult participation 1:868
- diversity exposure 1:867
- first-order engagement 4:400
- general discussion 1:872, 4:398
- governance structures 4:401
- Hegelian perspective 4:399
- historical background 4:398
- second-order engagement 4:400
- service opportunity recommendations 1:870
- stakeholders 4:401
- third-order engagement 4:400
- intensive field experiences 1:714
- public versus private schooling 2:236
- years of schooling 2:235
- civics education programs
 - assessment strategies 3:323–329
 - Center on Organization and Restructuring Schools (CORS) standards 3:325, 3:326*t*
 - challenges 3:326
 - classroom-based assessments 3:325, 3:325*t*
 - document-based question (DBQ) assessment 3:324
 - general discussion 3:323, 4:657
 - large-scale assessments 3:323
 - Civic Education Study (CIVED) 4:656–662
 - characteristics 3:687*t*
 - design and implementation 4:657, 4:658*t*
 - educational impacts 4:667
 - findings implications and utilization 4:660
 - general discussion 4:662
 - historical foundations 4:657
 - participating countries 4:658*t*
 - results dissemination 4:660
 - survey results and analysis 4:658
 - theoretical framework 4:659*f*
 - educational program evaluation 3:667–673
 - assessment validity and reliability 3:668
 - challenges 3:667
 - classroom implementation
 - assessment practices 3:670
 - content decisions 3:670
 - general discussion 3:670
 - instructional impacts 3:671
 - curriculum 3:667, 3:668*t*
 - general discussion 3:523, 3:672
 - recommendations 3:672
 - research background 3:671
- Civil Rights Act (1964) 2:689
- civil rights movement 1:77, 1:884, 2:688, 2:716, 6:263
 - see also* critical race theory
- civil society
 - adult education 1:191–197
 - functional role 1:191
 - global civil societies 1:195
 - theoretical perspectives
 - communitarian perspective 1:192
 - Gramscian perspective 1:194
 - Habermasian perspective 1:193
 - basic concepts 7:533
 - characteristics 1:191
 - functional role 1:191
 - higher education 4:371
- claim-rights 6:17
- Clark, Burton 1:728
- Clark, Kenneth and Mamie Phipps 6:499–500
- classical test theory 4:2, 4:27
- classification and discriminant analysis *see* discriminant analysis
- classification-tree method 7:139, 7:140*f*, 7:141*f*
- classroom assessment 3:209–219
 - assessment categories 3:211
 - assessment impacts 3:506–511
 - formative assessments 3:464, 3:510
 - general discussion 3:506
 - high-visibility external tests
 - consistent results/potential benefits 3:508
 - consistent results/problematic effects 3:508
 - influencing factors 3:509
 - overlapping purposes 3:506
 - research areas 3:510
- assessment methods
 - informal classroom observations 3:217
 - paper-and-pencil tests
 - basic concepts 3:212
 - constructed response items 3:213
 - selected response items 3:212
 - performance assessments 3:214, 3:215*f*, 3:216*f*
 - portfolio assessments 3:215
- Australia 3:417–424
 - external versus classroom assessments 3:421
 - future challenges
 - increased diversity 3:422
 - lifelong learning goals 3:422
 - special needs students 3:422
 - technological tools 3:422
 - general discussion 3:423
 - influential factors
 - curriculum 3:418
 - diversity and inclusivity 3:417
 - federalization increases 3:420
 - professional teaching standards 3:418
 - school control mechanisms 3:418
 - purpose 3:417
 - school control mechanisms
 - general discussion 3:418
 - K–10 syllabi 3:419
 - outcomes-based syllabi 3:419
 - professional development support 3:420
 - senior schooling syllabi 3:419
 - student work verification 3:420
 - syllabus frameworks 3:419
- basic concepts 3:209
- challenges
 - classroom environments 3:464
 - fidelity and flexibility considerations 3:465
 - teachers' and students' roles 3:465
- criteria design guidelines
 - co-construction 3:413
 - evidentiary support 3:414
 - feedback 3:413–414
 - goal setting 3:414
- curriculum-based assessment (CBA) 2:571–576, 4:102–110
- curriculum-based measurement (CBM)
 - basic concepts 4:103
 - individualized instructional programs 4:108
 - instructional decision making 4:104, 4:107*f*
 - intervention responsiveness 4:104
 - mathematics education 2:771, 4:108
 - reading instruction 4:108
 - Response to Intervention (RTI) 4:104
 - secondary prevention strategies 4:108
 - special needs students 2:572, 2:771
 - student focus 4:106
- definitions 2:571
- design considerations 2:573
- effectiveness evaluations 2:574
- future directions 4:109
- general curriculum impacts 2:573
- general discussion 2:571, 2:575
- historical background 2:572
- importance 2:571
- mastery measurement 4:102
- motivational strategies 2:574
- progress monitoring 4:102
- relevant measurement models 2:573
- usage guidelines 2:575
- design requirements
 - assessment tasks 3:210
 - evidence interpretation 3:210
 - general discussion 3:210
 - learning goals 3:210, 3:211*f*
- Eastern Europe 3:476
- educational impacts 4:69–74

- classroom assessment (*continued*)
 formative assessments 4:73
 general discussion 4:69
 high-visibility external tests
 influencing factors 4:72
 potential benefits 4:71
 problematic effects 4:71
 historical perspectives 4:69
 measurement purposes 4:69
 research areas 4:73
- educational innovations 4:450
- fairness 4:56–60
 assessment impacts
 bias 4:57
 equity 4:57
 general discussion 4:56
 intelligence quotients (IQs) 4:56–57
 legal challenges 4:57
 equity approaches
 fairness principles 4:59*t*
 important questions 4:58, 4:58*t*
 test development 4:58, 4:59*t*
 general discussion 4:56, 4:60
 informal evaluations/assessments 4:59
 large-scale assessments 4:59
- fine arts education
 criteria 3:264
 dance 3:266
 drama/theater 3:266
 music 3:265
 visual arts 3:265
- formative assessments 3:211, 3:229, 3:413, 3:464, 3:510
- French-language sources 3:432–437
 background information 3:432
 competency-based programs 3:433
 effective communication strategies 3:434
 external testing 3:435
 formative assessments 3:434
 general discussion 3:436
 multilingual sources 3:432
 report cards 3:434
 school reform 3:432
 standards 3:435
 translated meanings 3:432
 general discussion 3:209, 3:218
 historical perspectives 3:506
- Hong Kong 3:438–442
 background information 3:438
 general discussion 3:441
 primary schools
 basic competency assessments 3:439, 3:440*t*
 characteristics 3:438
 learning-to-learn reforms 3:439
 target-oriented curriculum (TOC) reform 3:438
 secondary schools
 characteristics 3:440
 school-based assessments 3:441
 washback studies 3:440
- Learning How to Learn (LHTL) project
 background information 3:270, 3:464
 challenges
 classroom environments 3:464
 collaborative classroom learning 3:466
 fidelity and flexibility considerations 3:465
 teachers' and students' roles 3:465
 teachers' assessment practices 3:466*t*
 importance 3:469
 supportive school policies and practices 3:466*t*, 3:467, 3:469*f*
 teachers' professional learning practices 3:466, 3:466*t*
- New Zealand 3:443–448
 assessment policies 3:444
 background information 3:443
 primary schools 3:444, 5:694
 school curriculum policy 3:443
 secondary schools 3:446
- preservice education programs 3:410
- purpose 3:209
- reading skills 3:189–195
 affective aspects 3:190
 authentic assessments 3:190
 computer-assisted assessments 3:193
- fluency 3:193
 general discussion 3:194
 group discussions 3:190
 group tests 3:189
 individual tests 3:190
 informal reading inventories (IRIs) 3:191
 motivational aspects 3:190
 norm-referenced tests 3:189
 peer assessments 3:194
 phonics skills 3:191
 portfolios 3:193
 pre-reading sub-skills 3:191
 reading miscue inventories (RMIs) 3:191, 3:192*t*, 3:193*t*
 retellings 3:193
 self-assessments 3:194, 3:194*t*
 systematic observation 3:191
- research areas
 adult learning 3:412
 feedback 3:411
 motivation 3:411
 professional learning 3:413
 self-assessments 3:411
 summative assessments 3:411
- second-language learning 4:144
- student learning impacts 3:198
- sub-Saharan Africa 3:449–456
 assessment types
 continuous assessments 3:450, 3:451*t*
 international assessment programs 3:449
 national assessment programs 3:449
 public (external) examinations 3:450
 background information 3:449
 challenges 3:454
 summative assessments 3:211, 3:411, 3:415, 3:451*t*
- supportive school policies and practices 3:469
- technology supports 8:42–47
 basic concepts 8:42
 educational implications 8:45
 functional roles 8:43*t*
 general discussion 8:42
 large-scale commercial-based systems 8:44, 8:45*t*
 small-scale research-based systems 8:45, 8:45*t*
- validity measures 3:217
- writing instruction 3:276
- classroom discourse 5:361–366
 academic socialization
 adult-child communication 5:363
 classroom speech events 5:363
 modeling strategies 5:364
 recasts 5:364
 scaffolding strategies 5:364
 specialized academic registers 5:364
- definitions 5:361
- gender analysis 5:365
- general discussion 5:366
- nonstandard dialects 5:365
- second-language learners 5:365
- teacher-student communication
 asymmetric roles 5:362
 characteristics 5:361
 student participation opportunities 5:363
 three-part question-answer sequence 5:362
- classroom diversity 1:867
- classroom ethnography 6:320–325
 background information 6:320
 basic concepts 6:321
 challenges 6:323
 illustrative examples 6:323
- class size 2:443–449, 6:200–206
 classroom process impacts
 curriculum effects 6:204
 discipline management 6:203
 general discussion 6:203
 group work effects 6:204
 individualized teaching 6:203
 morale 6:203
 peer relationships 6:204
 pupil attentiveness/on-task behaviors 6:203
 pupil interactions 6:203
 stress factors 6:203
 teacher-student relationships 6:203

- class size–academic outcome relationship 2:443
- class-size reduction research
 - benefits 2:449
 - empirical analyses 2:443
 - experimental research
 - Hawthorne effects 2:448
 - nonrandom movement effects 2:447
 - Project STAR (Student Teacher Achievement Ratio) 2:43, 2:45*t*, 2:444, 5:54, 7:227
 - randomization checks 2:445, 2:446*t*
 - test score results 2:446, 2:446*t*, 2:447*t*
 - nonexperimental research 2:444
 - policy-induced reductions 2:449
 - quasi-experimental research 2:448
- educational performance
 - California 6:201
 - Class Size and Pupil–Adult Ratio (CSPAR) project 6:201
 - correlational designs 6:201
 - general discussion 6:200
 - pupil–teacher ratios (PTRs) 6:201
 - research results
 - alternative programs 6:202
 - average class sizes 6:202
 - cultural effects 6:202
 - school performance 6:202
 - student age 6:202
 - student characteristics 6:202
 - threshold effects 6:202
 - Student Achievement Guarantee in Education (SAGE) program 6:201
 - Student/Teacher Achievement Ratio (STAR) project 6:201
- general discussion 2:443
- government policies and programs 6:200
- practical implications 6:204
- Class Size and Pupil–Adult Ratio (CSPAR) project 6:201
- clergy education 3:31–36
 - certification and licensing 3:35
 - characteristics
 - accreditation
 - baccalaureate and certificate programs 3:35
 - graduate professional theological education 3:35
 - educational culture 3:34
 - education models
 - baccalaureate and certificate programs 3:32
 - general discussion 3:32
 - graduate professional theological education 3:32
 - international programs 3:32
 - signature pedagogy
 - contextualization 3:34
 - formational pedagogy 3:34
 - general discussion 3:34
 - interpretation 3:34
 - performance 3:35
 - theological curriculum
 - general discussion 3:33
 - ministry contexts 3:33
 - moral sensibilities 3:33
 - professional skills and abilities 3:33
 - spiritual awareness 3:33
 - texts and traditions 3:33
 - theological understanding 3:33
 - continuing education 3:35
 - educational practices 3:32
 - general discussion 3:31
 - learned ministries 3:31
 - qualifications and authorizations 3:31
- Cleveland voucher program 2:338–339, 2:381, 2:382*t*
- clever cities theory 4:373
- CLIA (competence, learning, intervention, and assessment) model 5:401, 5:405
- clinical laboratory science 3:52*t*
- Clinical Legal Education Association (CLEA) 3:28
- cliques 6:566, 6:709
- closed circuit televisions (CCTVs) 2:882–883
- closure (social networks) 6:705
- cluster analysis 7:72–83
 - characteristics 7:5, 7:72
 - design guidelines 7:82
 - hierarchical clustering technique 7:73, 7:74*t*, 7:75*t*, 7:76*f*, 7:77*f*
 - linkage clustering 7:77*f*
 - methodologies 7:14
 - model-based clustering technique 7:76, 7:80*f*, 7:81*t*
 - number of clusters 7:80, 7:81*f*
 - optimization clustering technique 7:74
 - partitioning around medoids (PAM) technique 7:74, 7:78*t*, 7:79*f*, 7:79*t*
 - proximity measures
 - basic concepts 7:72
 - binary data/similarity measures 7:73*t*
 - distance measures 7:73*t*
 - protein consumption data 7:74*t*
 - protein consumption data/distance measures 7:75*t*
- cluster sampling 3:144, 3:156, 7:392
- Coach Carter* film screening 1:865
- Coady, Moses 1:55
- Cochran–Mantel–Haenszel (CMH) test
 - association reversal phenomena 7:60, 7:61*t*
 - common odds ratio 7:61, 7:262, 7:262*t*
 - differential item functioning (DIF) 7:61
 - Simpson's paradox 7:60, 7:61*t*, 7:262
- Cochran theorem 7:235
- Code of Fair Testing Practices in Education 3:378, 3:380
- codes of conduct
 - assessments and evaluations 3:378, 3:380
 - business/accounting education 3:43
 - engineering profession 3:48
 - pharmacy education 3:60
- code-switching 5:356
- cognition
 - adult development 5:225–230
 - aging 5:227
 - general discussion 5:225, 5:229
 - lifelong learning effects
 - early education–cognition relationships 5:228
 - plasticity 5:229
 - life-span development theories
 - general discussion 5:225
 - multidirectionality and intellectual function 5:225
 - selectivity 5:226
 - self-regulated learning 5:226
 - memory considerations 5:227
 - text-based learning
 - discourse structures 5:228
 - general discussion 5:227
 - situation model processing 5:228
 - textbase processing 5:227
 - vocabulary 5:227
 - adult learning 1:15, 1:38
 - aging and cognitive abilities 5:340–344
 - brain aging 5:341
 - cognition–brain relationship 5:343
 - long-term prognoses 5:343
 - memory performance
 - age-related effects 5:342
 - cognitive decline 5:341
 - episodic memory 5:342
 - executive functions 5:342
 - familiarity 5:342
 - long-term memory 5:341
 - recollection 5:342
 - semantic memory 5:342
 - short-term memory 5:341
 - working memory (WM) capacity 5:342
 - neurocognitive plasticity 5:343
 - neuroimaging studies 5:340
- attention 5:165–171
 - general discussion 5:165
 - infants
 - behavioral phases 5:166, 5:167*f*
 - brain networks 5:167
 - compensatory systems 5:170
 - directed-attention model 5:169
 - dyadic attention 5:169
 - face processing 5:168
 - general discussion 5:170
 - intentional understanding 5:170
 - newborn systems 5:165
 - novelty factors 5:168
 - plasticity 5:170
 - selective attention development 5:166, 5:167*f*
 - triadic attention/joint attention 5:170
 - visual recognition memory 5:168
- bilingualism 1:479
- cognition–emotion relationship 5:237–244
 - affect 6:547

- cognition (*continued*)
- causal factors
 - cognitive-appraisal hypothesis 5:239
 - emotional arousal and valence 5:240
 - pedagogical challenges 5:242
 - results analysis 5:242
 - definitions 5:238
 - research background 5:237
 - definitions 5:238
 - embodied cognition 8:179, 8:180^f
 - emotional feelings 6:547
 - insight 5:276
 - intelligence theories 5:185
 - learning theories 1:522
 - learning to learn (L2) 3:268–272
 - current trends
 - collective and institutional weaving concept 3:270
 - EU pilot project 3:270
 - general discussion 3:269
 - individual differences 3:269
 - research areas 3:270
 - developmental psychological traditions 3:269
 - educational impacts 3:271
 - historical background 3:268
 - motivational/volitional traditions 3:269
 - sociocultural traditions 3:269
 - theoretical perspectives 3:268
 - transfer processes 3:271
 - narrative socialization
 - narrative thinking 5:304, 5:464, 6:438
 - theories of the mind 5:305
 - personal epistemology 5:213
 - recent research trends
 - bodily gestures 5:249
 - collaboration 5:249
 - cultural values and beliefs 5:249
 - spatial thought 5:249
 - transfer processes 5:249
 - research areas
 - comprehension and conceptual understanding 5:248
 - general discussion 5:247
 - language acquisition and reading 5:247
 - memory 5:248
 - metacognition 5:248
 - perception and attention 5:247
 - problem solving and reasoning 5:248
 - research background 5:245–250
 - behaviorist perspective 5:245
 - cognitive psychology 5:246
 - cognitive revolution 5:245
 - general discussion 5:245
 - returns to education 2:310, 2:310^r
 - situated cognition
 - adult learning 1:15
 - case-based education methods 7:594
 - characteristics 7:750^r
 - ethnomethodology research 6:396
 - nursing education 3:66
 - science learning 5:391
 - situated learning 5:335–339
 - apprenticeships 5:336
 - educational applications 5:337
 - prior learning assessment (PLA) 1:172
 - professional development 7:549, 7:549^r
 - sense of identity 5:336
 - theoretical perspectives 5:335
 - social cognitive theory
 - basic concepts 6:668
 - case-based education methods 7:595
 - self-regulated learning 5:504
 - see also* metacognition
- Cognitive Academic Language Proficiency (CALP) 5:352
- cognitive apprenticeships 3:26
- Cognitive Assessment System 5:188
- cognitive behavioral modification 6:555^r
- cognitive conflict 1:491
- cognitive diagnosis models (CDMs) 7:63
- cognitive disabilities 2:669–677
 - assistive technology (AT) 2:673, 2:674^r, 2:675^r
 - characteristics 2:669, 2:670^r
 - effective teaching behaviors
 - general discussion 2:670
 - observational learning 2:670–671
 - reinforcer learning 2:671–672, 2:672^r
 - response–prompt hierarchy 2:671, 2:671^r, 2:672^f
 - time–delay learning 2:671, 2:672^r
 - verbal learning mediation 2:670- general discussion 2:676
- instructional strategies 2:669
- positive adult outcomes 2:675
- school-based services 2:853
- strategic educational programming 2:672, 2:674^r

cognitive flexibility theory 7:594

cognitive load theory 3:227

cognitive psychology 7:84–92

 - background information 7:84
 - characteristics 7:750^r
 - construct validity 7:84
 - general discussion 7:91
 - research background 4:2, 5:246
 - statistical models
 - applications
 - Abstract Reasoning Test (ART) 7:89, 7:90^f, 7:90^r, 7:91^f, 7:91^r
 - spatial learning ability test (SLAT) 7:88, 7:88^f, 7:89^r
 - general component latent trait model (GLTM) 7:87
 - linear logistic test model (LLTM) 7:85, 7:89^r
 - linear partial credit model (LPCM) 7:86
 - multicomponent latent trait model (MLTM) 7:87, 7:87^f
 - two-parameter logistic (2PL) model 7:86, 7:91^r- test design and development
 - cognitive design system (CDS) 7:84–85, 7:85^r
 - item development 7:85
 - procedural framework 7:84, 7:85^r

cognitive strategy instruction (CSI)

 - literacy instruction 2:759, 2:760^r
 - mathematics instruction 2:770

Cognitive Styles Index (CSI) 8:377

cognitive theory of multimedia learning (CTML) 8:145, 8:146^f

Cognitive Tutors 8:52, 8:173, 8:174^f

Cohen's *d* 7:445

Coleman Report 2:403, 2:409, 2:438

collaborative action research 7:759

collaborative leadership 5:4

collaborative learning

 - basic concepts 6:685

computer-supported collaborative learning (CSCL) 5:290–296

 - background information 5:291
 - design-based research
 - pedagogical design 5:293
 - technology design 5:293
 - future directions 5:294
 - research approaches
 - dialogic approach 5:292
 - systemic approach 5:291
 - sociocultural perspectives 5:290

co-teaching 2:560–564

 - benefits 2:561
 - characteristics 2:560
 - general discussion 2:563
 - limitations 2:562
 - need requirements 2:562
 - professional standards 2:562, 2:563^r
 - rationale 2:560
 - teaching approaches 2:561

organizational learning characteristics 7:558–559

social characteristics 6:685–690

 - friendship 6:686
 - general discussion 6:688
 - peer relationships 5:445, 6:686
 - status relationships 6:687

social interactions

 - experimental research 6:701
 - general discussion 6:700
 - learning characteristics 6:702
 - peer discussions 6:701
 - productive interactions
 - effective talk 6:703
 - general discussion 6:702
 - relationship quality 6:702
 - task design 6:702
 - surveying classroom activity 6:701

technology-supported learning environments 6:719

- collaborative networks 5:92–99
 educational leadership 4:699
 Networked Learning Communities (NLC) program
 basic concepts 5:93, 5:94^f
 challenges 5:98
 incentivization and implementation 5:95
 outcomes analyses 5:97
 research findings 5:96
 policies and programs 5:93
 school improvement systems 5:92
 school-to-school collaboration 6:274–280
 benefits 6:274, 6:275^t
 collaborative advantages 6:276
 collaborative innovation 6:277, 6:278^f
 effectiveness 6:275^t
 England 6:274
 general discussion 6:274
 knowledge transfer 6:277
 leadership
 capacity-building improvements 6:278
 lateral accountability 6:279
 modeling approaches 6:279, 6:279^f
 partnership ethos 6:279
 Leading Edge Partnership Program (LEPP) 6:274, 6:275
 school improvements 6:280
- collective bargaining
 higher education 4:491–498
 basic concepts 4:491
 challenges 4:496
 country analysis
 Argentina 4:496
 Australia 4:496
 Brazil 4:496
 Cameroun 4:494
 Chile 4:496
 China 4:495
 East Africa 4:493
 France 4:493
 general discussion 4:492
 Germany 4:493
 India 4:494
 Indonesia 4:495
 Japan 4:495
 Malaysia 4:495
 Mexico 4:496
 Nigeria 4:494
 Russia 4:493
 South Africa 4:494
 South Korea 4:494
 Sweden 4:493
 United Kingdom 4:493
 United States 4:492
 West Africa 4:493
 international standards 4:491
 teachers' unions 2:512, 2:513
- collective case studies 6:458
 collective learning 1:492, 5:156
 collectivism 4:828
 College Entrance Examination Board 4:7
 colleges
 access and equity patterns 1:636–648
 academic preparation 1:637
 college knowledge 1:637
 demographic changes 1:636
 financial considerations 1:637
 geographical information systems (GIS)
 advantages 1:638
 characteristics 1:637
 college knowledge 1:645^f, 1:646^f
 data sources and methods 1:638
 illustrative examples 1:638
 racial demographics 1:643^f, 1:644^f
 research implications 1:647
 Tucson elementary schools 1:639^f, 1:643^f, 1:644^f
 Tucson high schools 1:641^f, 1:645^f, 1:646^f
 Tucson middle schools 1:640^f
 international access patterns 4:334
 labor market 1:636
 social stratification 1:637
 Tucson, Arizona
 college knowledge 1:645^f, 1:646^f
 elementary schools 1:639^f, 1:643^f, 1:644^f
 geographical information systems (GIS) research 1:638
 high schools 1:641^f, 1:645^f, 1:646^f
 middle schools 1:640^f
 racial demographics 1:643^f, 1:644^f
 research implications 1:647
 research methodologies 1:638
 Latino students 1:610–615
 college-choice model
 basic concepts 1:611, 1:612^f
 chain migration theory 1:613
 decision-making models 1:613^f
 quantitative methods 1:612
 social capital theory 1:613
 demographic data 1:610
 educational attainment 1:610
 learning opportunities 1:610
 Philippines 5:868^t, 5:869
 Scotland 5:772
 student learning outcomes 4:479–484
 benefits 4:483
 cognitive complexity 4:479, 4:480^f
 humanitarianism 4:480^f, 4:481
 interpersonal and intrapersonal competence 4:480^f, 4:482
 knowledge acquisition 4:480, 4:480^f
 practical competence 4:480^f, 4:482
 research background 4:479
 see also higher education; universities
- collegiality 4:252, 4:266, 4:434
 Collier, John 1:357, 6:500
 Colombia
 assessment system development 3:480
 early childhood education and care (ECEC) 2:111^t
 earnings–education level relationship 2:306^t
 educational vouchers 2:382^t, 2:384
 higher education enrollment 1:702^f, 1:703^f
 inequality–education relationship 2:273^f
 information and communication technology (ICT)
 computer access at home 8:25^t
 computer access in school 8:23^t
 frequency of computer use in school and at home 8:26^t
 student proficiency and confidence 8:31^t
 moral and social reasoning 5:282
 private higher education (PHE) 4:623
 public subsidization 2:308^t
 returns to education–education level 2:308^t
 student achievement test performance 2:247^f, 2:417^t
 test score–social background relationship 2:273^f
 voucher systems 2:419–420
- Colorado 1:654^f
 colorblindness 6:40–41
 Columbia University (New York) 8:354
 Comenius programme 4:673
 comic books 8:109
 Commission for the Coordination of Education (CCE) 5:874
 Committee of Canadian Architectural Councils (CCAC) 3:13
 Common Code of Data (CCD) 2:178, 2:181^t
 Common Entrance Examination (CEE) 1:680
 Common European Framework (CEF) 1:442
 Common European Higher Education Area (CEHEA) 1:772, 4:566
 common odds ratio 7:61, 7:262, 7:262^t
 common school 6:44–49
 basic concepts 6:44
 challenges 6:48–49
 general discussion 6:44
 as object of choice 6:48
 school choice 6:46
 values analyses 6:47
- Commonwealth Council for Educational Administration and Management (CCEA/M) 4:681
 Commonwealth of Independent States (CIS) 3:472, 3:473^t
 communication disorders 2:654–660
 allied health professions 3:52^t
 augmentative–alternative communication (AAC) 2:544–549
 characteristics
 general discussion 2:545
 manual communication 2:545, 2:624
 technological devices 2:545
 communication initiation
 challenges 2:547
 communication partners 2:548

- communication disorders (*continued*)
 - use settings 2:548
 - vocabulary selection 2:548
 - vowel-based communication chart 2:549^f
- communication partners 2:547, 2:548
- goals and objectives 2:544
- needs assessment 2:546
- culturally and linguistically diverse (CLD) learners
 - African-American English dialect speakers 2:656
 - bilingual students 2:656
 - community-based experiences 2:550–554
 - best practices strategies 2:552
 - general discussion 2:553
 - legal mandates 2:552
 - research background 2:550
 - transition frameworks 2:550
- cultural difference theory model
 - cultural incongruity model 2:776
 - general discussion 2:776
 - oppositional cultural framework 2:776
 - stereotype threats 2:776
- demographic data 2:656
- individualized education plans (IEPs) 2:658, 2:658^f
- Individuals with Disabilities Education Act (IDEA, 1990) 2:657
- oral language skills assessments 3:291
- overrepresentation 2:774–779
 - challenges 2:778
 - cultural deficits 2:775
 - cultural difference theory model 2:776
 - deficit thinking effects 2:775
 - ecological/contextual perspectives 2:777
 - historical perspectives 2:565, 2:690, 2:774
 - school failure explanations 2:775
- deaf-blind students 2:621–627
 - classroom intervention strategies 2:625
 - communication challenges 2:624
 - definitions 2:621
 - demographic data 2:623^r
 - developmental challenges 2:622
 - early intervention services 2:622, 2:625
 - etiology 2:624, 2:624^r
 - general discussion 2:626
 - hearing loss criteria 2:623
 - historical background 2:621
 - legal blindness criteria 2:623
 - transition processes 2:626
- diagnostic and classification processes 2:579
- disorder characteristics
 - articulation 2:654
 - childhood apraxia of speech (CAS) 2:654
 - general discussion 2:654
 - language-based learning disabilities (LBLD) 2:655
 - language disorders 2:655
 - orofacial myofunctional disorders (OMD) 2:654
 - phonological disorders 2:654
 - stuttering 2:655
 - voice disorders 2:655
- future directions
 - evidence-based practices 2:659
 - general discussion 2:658
 - technology-supported learning environments 2:658
 - testing and assessments 2:658
- general discussion 2:654, 2:659
- language difference versus language disorder 2:747
- school-based services 2:854
- communication literacy 8:37
- communicative competence 5:382, 5:383^r, 6:398
- communicative language teaching (CLT) model
 - characteristics 1:429
 - criticisms 1:429
 - curricular advantages 1:432, 1:432^f
 - educational conflicts 1:431
 - key assumption and mission 1:430
 - persuasive discourses 1:433
 - student resistance 1:432
- communitarian perspective 1:192, 6:7, 6:123
- Community College Research Center (CCRC) 8:354
- community colleges 4:610
- community education 1:120–126
 - advantages/disadvantages 1:124
 - basic principles 1:121, 1:122^f
 - characteristics 1:120
 - community connotations 1:125
 - general discussion 1:125
 - historical foundations 1:120
 - international perspectives 1:122
 - socioeconomic modernization 1:218
- community-focused schools 6:255–260
 - characteristics 6:255
 - educational impacts 6:258
 - functional role 6:256
 - future research areas 6:259
 - outcomes analyses 6:258
 - rationales 6:256
- community involvement *see* civic engagement
- community learning and development (CLD) programs 5:772–773
- community of inquirers 6:87
- community of practice (CoP) model 1:84
- Comoros
 - early childhood education and care (ECEC) 2:111^r
 - student achievement test performance 2:418^r
- Comparative and International Education Society (CIES) 1:266
- Comparative and International Education Society of Canada (CIESC) 1:266
- Comparative Education* 1:252
- comparative education 1:257–265
 - basic concepts 1:257, 1:258^f
 - comparison studies
 - educational development 1:261
 - foreign education/area studies 1:261
 - international education 1:251–256, 1:260
- cross-cultural analyses
 - anthropological perspectives 1:262
 - cultural perspectives 1:262
 - globalization effects 1:264
 - national cultural identity 1:263
 - theoretical perspectives 1:261
- English-language scholarly publications 1:251–256
 - editorial processes 1:253
 - electronic publications 1:255
 - English-language dominance 1:255
 - future directions 1:256
 - general discussion 1:251
 - historical background 1:251
 - manuscript submission and selection 1:254
 - monolingualism 1:255
 - publishing incentives 1:254
 - research trends and challenges 1:253
- professional societies 1:266–271
 - challenges 1:270
 - oldest societies 1:266
 - World Council of Comparative Education Societies (WCCES)
 - foreign education versus comparative education 1:269
 - international focus 1:268
 - member societies 1:267^r
 - non-member societies 1:270
 - origins 1:266
 - society characteristics 1:267
 - world congresses 1:268^r
- research framework 1:258^f
- theoretical perspectives 1:257
- Comparative Education Review* 1:252
- Comparative Education Society (CES) 1:252, 1:266
- Comparative Education Society in Europe (CESE) 1:266
- comparative fit index (CFI) 7:193
- Compare* 1:252
- COMPASS 8:48
- CompEd study
 - characteristics 3:687^r
 - Functional Information Technology Test (FITT) 4:645
 - importance 4:649
 - purpose 4:644
 - Stage 1 survey results 4:644
 - Stage 2 survey results 4:645
- compensating differentials models 2:465–472
 - background information 2:465
 - displaced workers 8:276
 - general discussion 2:470
 - hedonic wage model
 - controllable versus uncontrollable cost factors 2:466, 2:467^r
 - geographic cost-of-education index 2:467, 2:468^f
 - limitations
 - comparable wage index (CWI) 2:468

- cost factor interpretation 2:469
- fixed-effects model 2:469
- flexible wages 2:469
- general discussion 2:468
- geographic considerations 2:470
- regional factors 2:468
- teacher quality measures 2:469
- public versus private schools 2:467
- time correlations 2:467
- historical research 2:465
- teacher labor markets 2:466
- competence 6:618, 6:621
- competence-based assessments 3:332
- competency-based learning model **8:403–410**
 - adult learning 6:117
 - basic principles 8:405
 - challenges
 - assessment relationships 8:409
 - curriculum relationships 8:408
 - generic competencies 8:408
 - lifelong learning 8:409
 - limited verbal descriptions 8:407
 - skill-development–workplace competence gap 8:407
 - competence, learning, intervention, and assessment (CLIA) model 5:401, 5:405
 - general discussion 8:403
 - historical perspectives 8:403
 - international acceptance 8:407
 - mathematics education 5:401, 5:405
 - medical education 3:22
 - Organisation for Economic Cooperation and Development (OECD) initiatives 8:405
 - vocational education and training (VET)
 - general discussion 8:230
 - recognition of prior learning (RPL) 8:400
 - reform efforts 8:403
 - terminology and jargon sources 8:350, 8:350*t*
- competency-based measurements 4:32
- competition **2:331–336**
 - basic concepts 2:331
 - Catholic schools 2:358
 - charter schools 2:363
 - contiguous public schools 2:333
 - empirical analyses 2:331
 - general discussion 2:335
 - non-US research 2:334
 - public choice 2:331
 - public versus private schools 2:333
 - Tiebout choice 2:332
- complete participation 6:443
- complex systems theory **1:481–487**
 - catastrophe theory 1:482, 1:482*f*
 - chaos theory 1:484, 1:485*f*, 5:130
 - decision-making models 1:485, 1:485*f*, 1:486*f*
 - educational innovation 5:130
 - limitations 1:486
 - mathematical models 1:481
 - metaphorical and analogical models 1:486
 - morphodynamical models 1:484
 - recursive models 1:483, 1:483*f*
 - social interactions 1:481
- component analysis (CA) 7:162
- comprehension monitoring 5:207, 5:326
- Comprehensive School Reform (CSR) models 5:56
- comprehensive school reform models 6:286, 7:559
- computational statistics **7:93–97**
 - computationally intensive methods 7:95
 - data mining/machine learning 7:96
 - educational programs 7:94
 - general discussion 7:93, 7:96
 - historical background 7:93
 - linear regression 7:95
 - Monte Carlo analysis 7:94, 7:95
 - optimization problems 7:95
 - random number generation 7:94
 - resampling methods 7:96
 - research publications 7:93–94
 - simulation methods 7:94
 - statistical computing 7:94
- computer adaptive testing (CAT) 4:18
- computer algebra systems (CASs) 8:173
- computer anxiety 6:551
- Computer-Assisted Qualitative Data Analysis Software (CAQDAS) **6:326–334**
 - background information 6:326
 - data exploration methods
 - annotation tools 6:330
 - general discussion 6:330
 - text search tools 6:330
 - word frequency tools 6:330
 - data integration methods
 - general discussion 6:328
 - incorporated or referenced data 6:328
 - multimedia data 6:329
 - qualitative and quantitative analyses 6:329
 - data organization methods
 - data grouping and filtering 6:331
 - data linking/hyperlinking 6:332
 - general discussion 6:331
 - interpretive coding 6:331
 - general discussion 6:333
 - management practices
 - connection-making methods 6:332
 - data interrogation 6:332
 - data retrieval 6:333
 - general discussion 6:332
 - pattern and relationship identification 6:333
 - subset comparisons 6:333
 - writing methods 6:332
 - project planning/project teamwork 6:327, 6:328*t*
 - supported qualitative tasks 6:326, 6:327*f*
- computer games 4:422, 8:7
- computerized adaptive testing (CAT)
 - background information 7:396
 - basic concepts 7:397
 - dynamic assessments 3:222–223
 - item response theory (IRT) applications 4:87
 - purpose 4:148
 - sampling methods 7:395
 - sequential probability ratio test (SPRT) 7:397, 7:398*t*
- computer technologies **8:1–10**
 - academic achievement impacts 8:109–110
 - academic values and expectations 4:436
 - access and equity patterns **8:95–101**
 - Canada 8:97
 - China 8:98
 - European Union (EU) 8:97
 - general discussion 8:95
 - global access and use 8:98, 8:98*t*
 - improvement efforts 8:99
 - India 8:98
 - international trends 8:21–33
 - challenges 8:30
 - computer access at home 8:25, 8:25*t*
 - computer access in school 8:22, 8:23*t*
 - frequency of use at home 8:25, 8:26*t*
 - frequency of use in school 8:25, 8:26*t*
 - large-scale assessments 8:22
 - mathematics and science instruction 8:30, 8:32*t*
 - patterns of use 8:26, 8:27*f*
 - student access 8:21
 - student proficiency and confidence 8:27, 8:31*t*
 - theoretical perspectives 8:21
 - United States
 - home access 8:95, 8:96*t*
 - home use 8:95
 - outcomes analyses 8:95, 8:97
 - school access 8:96, 8:96*f*
 - school use 8:96
- automated essay scoring (AES) **4:20–26**
 - available programs 4:21
 - basic concepts 4:20
 - educational applications
 - automated speech-recognition scoring programs 4:24
 - Criterion online essay evaluation service 4:22
 - English language learners (ELLs) 4:23
 - general discussion 4:22
 - Text Adaptor 4:23
 - general discussion 4:25
 - historical foundations 4:20
 - modeling approaches 4:20–21
 - reliability and validity 4:22
- CompEd study
 - characteristics 3:687*t*

- computer technologies (*continued*)
 - Functional Information Technology Test (FITT) 4:645
 - importance 4:649
 - purpose 4:644
 - Stage 1 survey results 4:644
 - Stage 2 survey results 4:645
- computer-based testing
 - educational assessments 4:49
 - mathematics education 3:281
 - survival data analysis 7:459
- computerized adaptive testing (CAT)
 - background information 7:396
 - basic concepts 7:397
 - dynamic assessments 3:222–223
 - item response theory (IRT) applications 4:87
 - purpose 4:148
 - sampling methods 7:395
 - sequential probability ratio test (SPRT) 7:397, 7:398*t*
- computer-supported collaborative learning (CSCL) 5:290–296
 - background information 5:291
 - design-based research
 - pedagogical design 5:293
 - technology design 5:293
 - future directions 5:294
 - research approaches
 - dialogic approach 5:292
 - systemic approach 5:291
 - sociocultural perspectives 5:290
- educational applications
 - adult learning 1:40
 - educational impacts 8:2
 - effectiveness 8:7
 - emerging uses
 - general discussion 8:6
 - immersive environments and games 4:422, 8:7
 - interactive classroom communication systems 8:7
 - social networks 4:422, 5:120, 8:6
 - Web 2.0 4:431, 8:6
 - general discussion 8:9
- higher education 4:419–426
 - assessment strategies 3:253
 - background information 4:419
 - curriculum 4:420
 - effective learning 4:421
 - e-learning pedagogies 1:522, 4:423, 4:423*t*
 - flexible learning 4:423, 4:441–447
 - future directions 4:425
 - teaching approaches 4:424
 - technology enhanced learning (TEL) 4:423
- implementation strategies 8:8
- out-of-school learning 8:9
- rapidly growing uses
 - assessment and accountability systems 8:6
 - online courses 8:5
 - open-source online resources 8:5
 - virtual schools 8:5
- subject-specific uses
 - language instruction 8:4
 - mathematics instruction 8:4
 - reading and literacy 8:3
 - science instruction 8:4
 - writing instruction 8:3
- typical uses 8:2
- educational impacts 8:88–94
 - goals and objectives 8:88
 - growth factors 8:88
 - implementation considerations 8:89
 - importance 8:93
 - modeling approaches 8:92
 - outcomes analyses 8:91
 - usage measurements 8:90
- educational leadership 5:120–127
 - educational impacts
 - community relations 5:121, 5:121*f*
 - general discussion 5:120
 - instructional strategies 5:121, 5:121*f*
 - literacy instruction 5:121, 5:121*f*
 - social networks 5:120, 5:121*f*
 - influencing factors 5:125, 5:125*f*
 - integration framework 5:123*f*
 - requirements and standards 5:121
 - social and legal challenges 5:124
 - teacher leadership 5:122, 5:123*f*
 - technology-mediated leadership development programs 5:123, 5:123*t*
- elementary/middle-grades reading curricula 1:466
- general discussion 8:1–2
- growth factors 8:2
- help seeking processes 6:657
- image-based education research 6:500
- instructional management 8:133–139
 - background information 8:133
 - classroom network technologies
 - effectiveness 8:137
 - general discussion 8:136
 - instructional strategies 8:137
 - key features 8:136
 - future directions 8:137
 - student-response systems
 - effectiveness 8:135
 - general discussion 8:134
 - key features 8:134
 - research areas 8:134
 - teaching models 8:134
- international access patterns 8:21–33
 - challenges 8:30
 - computer access at home 8:25, 8:25*t*
 - computer access in school 8:22, 8:23*t*
 - frequency of use at home 8:25, 8:26*t*
 - frequency of use in school 8:25, 8:26*t*
 - large-scale assessments 8:22
 - mathematics and science instruction 8:30, 8:32*t*
 - patterns of use 8:26, 8:27*f*
 - student access 8:21
 - student proficiency and confidence 8:27, 8:31*t*
 - theoretical perspectives 8:21
- large-scale assessments 8:48–55
 - conceptual framework 8:48
 - first-generation computer-based tests 8:48, 8:49*t*
 - general discussion 8:54
 - Generation R tests 8:49*t*, 8:52
 - next-generation electronic tests 8:49*t*, 8:50
 - technology-rich environments scenarios 8:52, 8:53*f*
- lifelong learning 8:185
- mathematics education 8:172–183
 - challenges 8:180
 - collaborative problem solving
 - The Adventures of Jasper Woodbury* 8:177
 - Anchored Instruction 8:177
 - Enhanced Anchored Mathematics 8:177, 8:178*f*
 - MathWorlds 8:178
 - networked devices 8:177
 - participatory simulations 8:177–178
 - serious game systems 8:178
 - SimCalc 8:178, 8:179*f*, 8:199
 - embodied cognition 8:179, 8:180*f*
 - importance 8:172
 - mathematical discovery
 - ANIMATE system 8:175, 8:176*f*
 - Dynamic Geometry systems 8:176
 - dynamic statistics packages 8:176
 - Function Probe 8:175
 - general discussion 8:174
 - Logo 8:175, 8:199
 - skills and concept learning
 - background information 8:172
 - Cognitive Tutors 8:173, 8:174*f*
 - computer algebra systems (CASs) 8:173
 - Geometric Supposer 8:173
 - hand-held graphing calculators 8:172, 8:173*f*
 - intelligent tutoring systems (ITSs) 8:173
- multiple information sources 5:195
- new media technologies 8:140–149
 - animation
 - general discussion 8:143
 - picture comprehension 8:144
 - picture design 8:144
 - realistic and dynamic pictures 8:143, 8:143*f*
 - general discussion 8:141
 - learning models
 - cognitive theory of multimedia learning (CTML) 8:145, 8:146*f*
 - design guidelines 8:148
 - dual-coding theory 8:145

- integrated model of text and picture comprehension (IMTPC) 8:145, 8:146*f*
- research findings 8:147
- theoretical models 8:145
- multimedia learning environments
 - animation 8:143
 - future directions 8:148
 - general discussion 8:141
 - learning models 8:145
 - linear and nonlinear text 8:142
 - text design guidelines 8:142
- research findings
 - coherence and contiguity 8:147
 - general discussion 8:147
 - modality 8:147
 - redundancy 8:147
- text design guidelines
 - additional stimulation 8:142
 - conciseness 8:142
 - general discussion 8:142
 - linguistic simplicity 8:142
 - structure and organization 8:142
- physically disabled students 2:642
- reading skills 8:150–157
 - beginning-reading programs 6:222
 - comprehension 8:152
 - emergent literacy 8:150
 - fluency 8:151
 - general discussion 8:150, 8:155
 - Internet access 8:154
 - phonics and phonemic awareness 8:151
 - text-to-speech applications 8:153
 - theoretical perspectives 8:154
 - upper elementary grades 6:224
 - vocabulary 8:152
- science learning 8:197–203
 - access and equity promotion 8:202
 - challenges 8:202
 - constructivist perspectives 8:197, 8:198*r*
 - contextualization
 - authenticity 8:201
 - everyday experiences 8:201
 - future directions 8:202
 - information access 8:201
 - learning environment transformation 8:201
 - physical environment interactions
 - model construction 8:199
 - scientific phenomena 8:198, 8:198*r*
 - simulations/dynamic representations 8:199
 - student investigations 8:199
 - reflective construction of knowledge
 - program management 8:200
 - teacher support and professional development 8:200
 - working memory (WM) capacity 8:200
 - social environment interactions 8:199
 - systemic impacts 8:202
- Second Information Technology in Education Studies (SITES)
 - basic concepts 4:646
 - educational impacts 4:667
 - importance 4:649
 - Lithuania 5:665
 - Module 1 (SITES M1) survey results 4:646
 - Module 2 (SITES M2) survey results 4:647
 - SITES 2006 4:648
- second-language learning 8:204–210
 - background information 8:204
 - benefits
 - anywhere/anytime instruction 8:208
 - general discussion 8:207
 - immediate feedback 8:208
 - individualized study plans 8:207
 - mistakes in private concept 8:208
 - patient tutoring 8:208
 - records of achievement 8:209
 - classroom instruction 8:205
 - English-language training demand 8:204, 8:205*f*
 - functional role 8:204
 - research results and analysis 8:207
 - subject-specific uses 8:4
- self-adaptive testing 4:148–152
 - assisted self-adaptive testing (ASAT) 4:149–150
 - basic concepts 4:148, 4:149*f*
 - benefits 4:148
 - dynamic assessments 3:222–223
 - general discussion 4:151
 - item pool 4:149
 - item-selection strategies 4:150, 4:150*f*
 - perceived control/perceived performance hypotheses 4:148
 - restricted self-adaptive testing (RSAT) 4:149
 - theoretical investigations 4:150
- speech and language impaired students 2:658
- technological impacts 8:112–118
 - health-related effects
 - direct and indirect physical effects 8:113
 - health information accessibility 8:113
 - treatment/intervention approaches 8:114
 - social development
 - aggressive behaviors 8:115
 - bullying 8:116
 - computer-mediated communication 8:116
 - general discussion 8:115
 - social relationships 8:116
- technology literacy and fluency 8:34–41
 - basic concepts 8:36
 - educational implications
 - curriculum changes 8:39
 - digital divides 8:39
 - literate lives 8:38
 - performance assessments 8:39
 - framework development 8:37, 8:38*f*, 8:39*f*
 - future directions 8:40
 - historical perspectives 8:35
 - multiple literacies 8:37
 - social implications 8:34
 - see also* Internet-based education
- concept learning 5:172–176
 - conceptual change perspective
 - basic concepts 5:389
 - change processes 5:389
 - definitions 5:172
 - general discussion 5:176
 - Piagetian perspective 5:387, 5:465
 - science learning 5:387
 - theoretical perspectives
 - classical view 5:172
 - domain specificity
 - biological concepts 5:174
 - constraints 5:174
 - cultural influences 5:174
 - mathematical concepts 5:175
 - probabilistic view
 - basic concepts 5:172
 - central tendency measures 5:172
 - exemplar theories 5:173
 - family resemblance principle 5:173
 - prototypes 5:173
 - typicality effects 5:172
 - theory view 5:173
- concept maps 3:306, 3:308*f*
- conceptual inventories 3:305, 3:306*f*, 3:307*f*
- concurrent articulation 4:612
- concurrent validity 4:34, 4:181–182
- conditional cash transfer (CCT) programs 2:419
- conditional maximum likelihood (CML) method 4:53, 7:371
- conditioning 5:198, 5:198*f*
- Conduct Curriculum 1:515
- conduct disorders 2:581
- Co-NECT model 8:74
- conferences
 - assessment conferences 3:403–409
 - basic concepts 3:404
 - constructive responses 3:275
 - definitions 3:403
 - historical background 3:404
 - informal assessment systems 3:405
 - parental involvement 3:406
 - portfolio assessments 3:406
 - purpose 3:403
 - recent developments
 - general discussion 3:407
 - Learning Record 3:407
 - New Zealand Education Monitoring Project (NZEMP) 3:408

- conferences (*continued*)
 self-assessments 3:406
 validity 3:408
 pre-referral intervention strategies 2:817
- confidence intervals 7:220, 7:221*f*
- confidence levels 6:678
- confidentiality 3:623, 6:33
- Confidential Measurement Based Self Evaluation (COMBSE) project 4:4
- confirmatory factor analysis (CFA)
 characteristics 7:164*t*
 general discussion 7:164
 linear structural equation models 7:453
 multilevel confirmatory factor analysis (MCFA)
 basic concepts 7:165
 estimation process 7:166
 illustrative application
 analysis and results 7:166
 methodology 7:166
 multitrait-multimethod (MTMM) designs
 confirmatory factor analysis with correlated uniqueness (CFA-CU) 3:115, 3:115*f*, 3:116*t*, 3:118, 3:118*t*
 confirmatory factor analysis with trait and method factors (CFA-TM) 3:114, 3:114*f*, 3:116, 3:116*t*, 3:117*t*
 post-analysis decisions
 model fit statistics 7:165
 model misspecification statistics 7:165
 pre-analysis decisions
 covariance matrix 7:164
 model identification 7:164
 multivariate normality 7:164
 parameter estimation theory 7:165
 rival models 7:164
- conflict management skills 6:567, 6:587–588
- confounded assignment mechanisms 7:69
- Confucianism 1:273
- congenital heart defects 2:636
- Congo
 early childhood education and care (ECEC) 2:111*t*
 student achievement test performance 2:418*t*
- Congress of South African Trade Unions (COSATU) 8:443
- connected knowing 1:20
- Connecticut 1:654*f*
- connectionism 5:476–484
 artificial neural networks (ANNs) 5:477, 5:478*f*, 5:478*t*, 5:479*f*
 characteristics 5:477
 educational relevance 5:483
 learning theories 5:347
 modeling approaches
 importance 5:477
 mathematics models
 general discussion 5:481
 knowledge-based cascade-correlation (KBCC) 5:482, 5:482*f*
 multiplication 5:481, 5:481*f*, 5:482*f*, 5:482*t*
 primality 5:482
 modern connectionism 5:477, 5:478*f*, 5:478*t*, 5:479*f*
 old versus modern connectionism 5:479, 5:479*t*
 reading models 5:480, 5:480*f*
- connective modularization 8:386
- connotation 6:476
- consequential validity 6:510, 7:515
- conservation of resources (COR) model
 basic concepts 3:499, 3:500*f*
 educational applications 3:499
 general discussion 3:498
 resource categories 3:499
 resource loss and gain spirals 3:501, 3:501*f*
- conservative hermeneutics 6:64
- constitutive phenomenology 6:450
- Constructing Knowledge Societies* (World Bank) 4:591, 4:593*t*
- constructivism 5:485–490
 basic concepts 5:485
 characteristics 7:750*t*
 curriculum 1:488–493
 cognitive conflict 1:491
 collective learning 1:492
 error evaluations 1:491
 general discussion 1:488
 metaphorical models 1:489, 1:490*f*, 5:390
 participatory curricula 1:492
 theoretical perspectives 1:488
 trans-level learning 1:490
 curriculum use 1:350
- early childhood education and care (ECEC) 2:107*t*
- e-learning pedagogies 1:522, 4:423*t*
- learning theories
 children's ideas 5:488
 core ideas 5:486
 general discussion 5:485
 intellectual development stages 5:487
 Piagetian perspective 1:488–489, 5:486
 Vygotskian perspective 1:489, 5:488
- museums 1:151–152
- recognition of prior learning (RPL) 8:399–400
- research program perspective 5:489
- tecnology-supported science learning 8:197, 8:198*t*
- theoretical perspectives 1:488, 5:486
- construct validity 2:578, 3:159, 4:170, 4:181–182, 7:84
- content validity 2:578, 3:158, 4:32, 4:181
- contingent leadership 5:45
- continuing education *see* continuing professional development
- continuing medical education (CME) 1:213
- continuing professional development 6:249–254
 basic concepts 6:249
 clergy education 3:35
 effectiveness 6:252
 entitlements 6:250
 general discussion 6:249
 impact evaluations 6:252
 professional knowledge and skills 2:889
 purpose 6:253
 school evaluations 3:540
 stakeholders and agendas 1:127–132
 background information 1:127
 educational agendas 1:130
 implications 1:131
 institutional agendas 1:128, 1:129*f*
 professionalization effects 1:127
 school and professional agendas 1:128
 theoretical orientations 1:129*f*
 vocational teacher education 8:508
 workplace-based assessments 3:333
 see also professional learning communities
- continuous probability distributions 7:98–102
 basic concepts 7:98
 beta distribution 7:98
 Cauchy distribution 7:99, 7:99*f*
 chi-squared (χ^2) distribution 7:99
 Dirichlet distribution 7:40, 7:99
 exponential distribution 7:100
 exponential family of distributions 7:100
 F distribution 7:100
 gamma distribution 7:100
 general discussion 7:102
 lognormal distribution 7:101
 Student's *t* test 7:101
 uniform distribution 7:101
 Wishart distribution 7:102
- contrastive linguistics 5:419
- control group designs 7:127
- controllability 6:559
- convenience sampling 3:145
- conventional planning theory 1:53
- Convention on the Rights of the Child 1:307
- convergent thinking problems 4:171, 5:273
- convergent validity 3:111
- conversational analysis (CA) 6:335–336, 6:394
- conversational language 5:357
- Cook Islands 2:111*t*
- Cook's distance 7:486
- cooperative learning 5:177–183
 background information 5:177
 beginning-reading programs 6:222
 goal reconciliation effects 5:181
 peer learning 5:445
 pre-referral intervention strategies 2:817
 schematic diagram 5:178*f*
 special needs students 2:555–559
 advantages 2:558
 basic concepts
 characteristics 2:556, 2:556*t*
 face-to-face promotive interaction 2:557
 group processing 2:557
 individual accountability 2:557
 positive goal interdependence 2:556

- social skills development 2:557
- research results and analysis 2:557
- theoretical perspectives 2:555
- structured group interaction 5:181
- theoretical perspectives
 - cognitive perspectives
 - cognitive elaboration perspective 5:181
 - developmental perspectives 5:180
 - general discussion 5:180
 - general discussion 5:177
 - group interactions 5:180
 - group rewards 5:179
 - motivational perspectives 5:178
 - social cohesion perspective 5:179
 - upper elementary grades 6:224
- Cooper, David 6:110–111
- co-participation/co-emergence model 1:85
- Copenhagen Accord 4:570, 8:435
- coping strategies 6:570–575
 - acute versus chronic stressors 6:570
 - affective traits 3:257
 - classroom interventions 6:574
 - coping–volition relationship
 - dual processing self-regulation model 6:572, 6:573f
 - emotional arousal regulation 6:573
 - volitional control 6:573
- effectiveness
 - general discussion 6:570
 - new stressor situations 6:571
 - situation assessments 6:571
 - strategy selection 6:571
- modeling approaches 6:574
- test anxiety 3:498–505
 - characteristics 3:498
 - comparison studies
 - alternate assessments 3:503t
 - cooperative learning activities 3:502
 - general discussion 3:501
 - online assessments 3:502
 - oral examinations 3:502
 - portfolios 3:502
 - product/project assessments 3:503
 - written tests 3:502
 - conservation of resources (COR) model
 - basic concepts 3:499, 3:500f
 - educational applications 3:499
 - future directions 3:503
 - general discussion 3:498
 - resource categories 3:499
 - resource loss and gain spirals 3:501, 3:501f
 - future directions 3:503
 - general discussion 3:504
 - implications 3:503
 - Yerkes–Dodson law 3:498, 3:499f
- copyright 4:303
- Core-40 End-of-Course Assessments 8:50
- Corey, Stephen 1:357–358, 7:542
- corrected ordinary least-squares analysis (C-OLS) 2:211
- correlation analysis *see* canonical correlation analysis (CCA); multivariate data analysis
- correspondence analysis (CA) 7:103–111
 - analysis of square asymmetric tables 7:111
 - background information 7:103
 - basic concepts 7:16
 - asymmetric maps 7:105–106, 7:106f
 - general discussion 7:103
 - illustrative example 7:106, 7:106f, 7:107f, 7:107t
 - inertia 7:104, 7:105f
 - triangular coordination system 7:104f
 - canonical CA 7:111
 - contributions (proportions) 7:108
 - data coding schemes 7:108
 - multiple correspondence analysis (MCA) 7:106, 7:109t, 7:110f
 - statistical inference 7:111
 - subset CA/subset MCA 7:111
 - supplementary points 7:108, 7:110f
- cortex 6:546
- Costa Rica
 - assessment system development 3:481
 - competency-based learning model 8:407
 - early childhood education and care (ECEC) 2:111t
 - earnings–education level relationship 2:306t
 - higher education enrollment 1:702f, 1:703f
 - public subsidization 2:308t
 - returns to education–education level 2:308t
- cost–benefit analysis 2:199–203
 - academic achievement outcomes 2:409
 - definitions 2:199
 - educational evaluation 3:587
 - illustrative example 2:201
 - methodologies
 - benefits measurements 2:201
 - cost estimation 2:200
 - general discussion 2:200
 - purpose 2:199
 - signaling model 2:322f
- cost–effectiveness analysis 2:199–203
 - definitions 2:199
 - educational evaluation 3:588
 - illustrative example 2:202
 - methodologies
 - cost estimation 2:200
 - effects measurements 2:201
 - general discussion 2:200
 - purpose 2:199
- co-teaching 2:560–564
 - benefits 2:561
 - characteristics 2:560
 - general discussion 2:563
 - limitations 2:562
 - need requirements 2:562
 - professional standards 2:562, 2:563t
 - rationale 2:560
 - teaching approaches 2:561
- Côte d'Ivoire
 - early childhood education and care (ECEC) 2:111t
 - student achievement test performance 2:418t
- Council for Exceptional Children (CEC) 2:562, 2:563t
- Council of Engineering Institutions (CEI) 3:46
- countenance curriculum evaluation model 1:337
- Cox–Small test 7:337
- Cox's proportional hazards (PH) regression model 7:462, 7:463f, 7:463t
- craft knowledge 7:642–649
 - accessibility
 - case-based education 7:647
 - general discussion 7:648
 - incorporation strategies 7:647
 - knowing-in-action 7:645
 - stimulated recall interview 7:647
 - characteristics 7:750t
 - importance 7:648
 - knowledge base considerations 7:642
 - research background
 - characteristics 7:644
 - content characteristics 7:644, 7:646t
 - research-generated knowledge–teaching practice relationship 7:643, 7:644f
- Cramer's *V* 7:447t
- Cramér–von Mises statistics 7:236
- Cramér–Wold theorem 7:235
- cram schools 5:838
- creativity 5:251–254
 - creative potential versus creative performance 5:251
 - developmental processes
 - influencing factors
 - behavior reinforcement 5:252
 - family structure 5:253
 - motivation 5:253
 - role models 5:253
 - nature versus nurture theory 5:252
 - fulfillment of potentials perspective 5:254
 - problem solving skills 5:251
- creativity assessment 4:170–174
 - bias 4:172
 - biographical indicators 4:173
 - challenges 4:174
 - criterion problem 4:173
 - divergent thinking tests 4:171
 - domain-specific assessments 4:172
 - general discussion 4:170
 - objective raters 4:173
 - originality measures 4:170
 - personality assessments 4:173
 - problem solving skills 4:172
 - tests as samples of behavior 4:170

- Credit Bank System (South Korea) 1:157*t*, 1:159
 crime and safety surveys 2:180, 2:181*t*
 crime–education relationship 2:239–244
 alternative/probationary schools
 educational performance indicators 1:878*t*
 effectiveness 1:877
 general discussion 1:876
 causal mechanisms 1:878
 criminal justice system 1:874–880
 alternative/probationary schools
 educational performance indicators 1:878*t*
 effectiveness 1:877
 general discussion 1:876
 causal mechanisms 1:878
 incarceration impacts
 dropout rates 1:875
 first arrests 1:875
 race-specific consequences 1:876
 international perspectives 1:879
 research background 1:879
 dropout rates 1:874
 economic factors
 general discussion 2:239
 wage and unemployment rates 2:239
 youth crime 2:240
 empirical analyses
 arrest and incarceration impacts 2:243
 educational attainment correlations 2:240
 general discussion 2:240
 prison education programs 2:243
 general discussion 2:244
 incarceration rates 1:874
 Criterion online essay evaluation service 4:22
 criterion-referenced measurements 4:31–35
 characteristics 4:31
 definitions 4:31
 educational applications 4:34
 emerging issues 4:35
 higher education 3:251
 science-education programs 3:663
 test development
 concurrent validity 4:34
 content validity 4:32
 domain specifications 4:32
 general discussion 4:32
 item analysis and selection 4:33
 item/assessment task writing 4:32
 results reporting 4:34
 standard setting 4:33
 test reliability
 decision consistency 4:33
 general discussion 4:33
 interrater reliability 4:34
 single-test administration 4:34
 vocational education and training (VET) 3:338
 criterion validity 2:578, 3:159, 4:181–182
 critical-age hypothesis 5:422
 critical ethnography 6:337–340
 future directions 6:339
 research background
 action anthropology 6:338
 cultural critiques 6:338
 educational anthropology 6:338
 feminist perspectives 6:337
 investigative research 6:338
 politicized anthropology 6:338–339
 positivist perspectives 6:337
 critical hermeneutics 6:65
 critical literacy 1:107
 critical media literacy 6:357
 critical pedagogy 6:26
 critical planning theory 1:55
 critical race theory 6:341–347
 background information 6:341
 basic concepts
 civil rights progress 6:343
 critique of liberalism 6:343
 experiential knowledge 1:781, 6:343
 interest convergence principle 6:343
 racism 6:343
 storytelling 1:782, 6:343
 White supremacy 6:342
 education research 1:779–786
 background information 1:779
 centrality of experiential knowledge 1:781, 6:343
 centrality of race 1:780
 critical legal theory 1:779
 educational applications 1:783
 educational theory reconceptualizations 1:783
 empirical educational research 1:783
 future implications 1:784
 narrative practices 1:782
 policy and practice research 1:783
 related areas of study 1:782
 general discussion 6:346
 misunderstandings 6:344
 origins 6:341
 qualitative research 6:490
 unresolved issues 6:345
 see also pragmatism
 critical self-reflection
 diversity considerations 7:522
 transformative learning theory 1:22
 critical theory perspective 6:133–140, 6:348–353
 characteristics 6:348
 critical pedagogy 6:23, 6:136, 6:352
 educational reform 6:352
 Frankfurt School 6:133, 6:349
 general discussion 6:133
 global perspectives
 African American tradition 6:351
 colonial and postcolonial theories 6:176, 6:351
 Latin America 6:351–352
 historical background
 classical philosophers 6:349
 Frankfurt School 6:349
 modern foundations 6:349
 and Jurgen Habermas 6:133
 and Paulo Freire 1:55, 1:495, 6:137, 6:351–352
 critical thinking 6:141–145
 background information 6:141
 basic concepts
 critical disposition 6:142
 normativity 6:141
 skills and abilities
 general discussion 6:141
 reason assessment component 6:142
 as a fundamental educational ideal
 criticisms 6:144
 general discussion 6:143
 justification 6:143
 nursing education 3:66
 Croatia
 age-related policies and programs 8:263*t*
 early childhood education and care (ECEC) 2:111*t*
 higher education 4:517
 information and communication technology (ICT)
 computer access in school 8:23*t*
 frequency of computer use in school and at home 8:26*t*
 student proficiency and confidence 8:31*t*
 vocational education and training (VET) 8:263*t*
 Cronbach's alpha 4:28
 cross-border higher education (CBHE) 4:507–513
 academic mobility 4:559–565
 future directions 4:564
 government policies and programs 4:559
 historical background 4:559
 internationalization processes 4:533, 4:542
 professional impacts 4:563
 statistical data
 faculty mobility 4:562
 general discussion 4:560
 student mobility 4:561
 student perspectives 4:562
 affiliated networks 4:512
 basic concepts
 borderless education 4:509
 cross-border education 4:509
 general discussion 4:508, 4:587
 offshore education 4:508
 transnational education 4:508
 branch campuses 4:512

- capacity-building/quality assurance initiatives 4:600
- challenges 4:512
- current trends 4:408
- distance education 4:511
- double/joint degree programs 4:511
- frameworks 4:510, 4:510*r*
- franchises 4:511
- growth trends 4:507
- independent institutions 4:512
- international educational terminology 4:507, 4:508*r*
- joint degree programs 4:511
- mergers and acquisitions 4:512
- program mobility typology 4:511
- providers
 - alternative providers 4:511
 - mobility typology 4:512
 - traditional higher education institutions 4:510
- quality assurance 4:537
- study centers 4:512
- twining programs 4:511
- validation arrangements 4:511
- virtual schools 4:511
- cross-cultural kids (CCKs) 4:639, 4:640*f*
- cross-cultural learning 5:450–456
 - academic achievement outcomes
 - Asian students
 - cultural values 5:452
 - general discussion 5:452
 - Markus–Kitayama concept of self-construal 5:453
 - Sue–Okazaki concept of relative functionalism 5:453
 - Black/Hispanic students
 - general discussion 5:453
 - institutional practices 5:454
 - Ogbu's two types of minorities theory 5:454
 - social class and wealth 5:453
 - future directions 5:455
 - future directions 5:455
 - general discussion 5:450
 - intellectual styles
 - characteristics 5:451
 - cultural impacts
 - characteristics 5:451
 - field dependence/independence 5:451
 - learning approaches 5:451
 - thinking styles 5:452
 - future directions 5:455
- cross-lag panel design 3:95, 3:95*f*
- cross-racial interactions 1:867
- cross talk methods 8:192
- crowds 6:566
- crystallized intelligence 4:2, 5:185
- Csikszentmihalyi, Mihaly 6:605
- Cuba
 - assessment system development 3:481
 - childcare programs 2:49
 - early childhood education and care (ECEC) 2:111*r*
 - higher education enrollment 1:702*f*, 1:703*f*
 - indigenous curricula 1:309
 - postcolonial perspectives 1:300–301
 - post-migration education 8:282
 - sociopolitical factors 1:310
- Cubberly, Ellwood 4:724
- cultural capital
 - basic concepts 6:100
 - challenges 6:101
 - educational applications 6:101
 - educational interventions 3:595
 - typology 6:100
- cultural globalization 6:529
- cultural–historical activity theory (CHAT) 5:491–497, 6:360–366
 - activity concepts 5:494, 5:494*r*
 - activity systems 5:494, 5:495*f*
 - background information 5:496
 - benefits 6:366
 - developmental background
 - cultural contexts 5:493
 - developmental levels 5:492
 - higher mental functions 5:492
 - theoretical perspectives 5:491, 6:360
 - tool-mediated action 5:493, 5:494*f*
 - educational approaches
 - development theories 6:361
 - learning 6:361
 - learning–development relationship 6:361
 - research examples
 - general discussion 6:362
 - long division/gazzinta process 6:362, 6:365*r*
 - puzzles 6:362, 6:365*r*
 - reading remediation 6:363, 6:365*r*
 - research results and analysis 6:364
 - weaving instruction 6:363, 6:365*r*
 - general discussion 6:360
 - and Lev Vygotsky 5:491, 5:492, 5:498, 6:360
 - semiotic mediation 5:494
 - theoretical perspectives 6:360
 - third generation systems 5:495, 5:495*f*
 - workplace learning 1:85
- culturally and linguistically diverse (CLD) learners
 - community-based experiences 2:550–554
 - best practices strategies 2:552
 - general discussion 2:553
 - legal mandates 2:552
 - research background 2:550
 - transition frameworks 2:550
 - dropouts 2:846
 - gifted and talented students 2:872
 - hearing impaired students 2:631
 - immigrant populations 4:631
 - language development considerations 2:746
 - oral language skills assessments 3:291
 - overrepresentation 2:774–779
 - challenges 2:778
 - cultural deficits 2:775
 - cultural difference theory model
 - cultural incongruity model 2:776
 - general discussion 2:776
 - oppositional cultural framework 2:776
 - stereotype threats 2:776
 - deficit thinking effects 2:775
 - ecological/contextual perspectives 2:777
 - historical perspectives 2:565, 2:690, 2:774
 - school failure explanations 2:775
 - special needs students
 - community-based experiences 2:550–554
 - best practices strategies 2:552
 - general discussion 2:553
 - legal mandates 2:552
 - research background 2:550
 - transition frameworks 2:550
 - cultural competence 2:565–570
 - basic concepts 2:566
 - community involvement 2:568
 - importance 2:569
 - overrepresentation 2:565
 - Pedersen's model 2:566
 - research background 2:565
 - cultural difference theory model
 - cultural incongruity model 2:776
 - general discussion 2:776
 - oppositional cultural framework 2:776
 - stereotype threats 2:776
 - dropouts 2:846
 - gifted and talented students 2:872
 - hearing impaired students 2:631
 - language development considerations 2:746
 - overrepresentation 2:774–779
 - challenges 2:778
 - cultural deficits 2:775
 - cultural difference theory model 2:776
 - deficit thinking effects 2:775
 - ecological/contextual perspectives 2:777
 - historical perspectives 2:690, 2:774
 - school failure explanations 2:775
- speech and language impairments
 - African-American English dialect speakers 2:656
 - bilingual students 2:656
 - demographic data 2:656
 - individualized education plans (IEPs) 2:658, 2:658*f*
 - Individuals with Disabilities Education Act (IDEA, 1990) 2:657
- speech and language impairments
 - African-American English dialect speakers 2:656
 - bilingual students 2:656

- culturally and linguistically diverse (CLD) learners (*continued*)
 - demographic data 2:656
 - individualized education plans (IEPs) 2:658, 2:658*f*
 - Individuals with Disabilities Education Act (IDEA, 1990) 2:657
- culturally responsive educational evaluations 3:592–599
 - critical pedagogy perspective 3:595
 - cultural psychology theory 3:594
 - ecological systems theory 3:593
 - intervention considerations 3:593
 - Kamehameha Early Education Program (KEEP) 3:596
 - methodologies
 - objectivist approaches 3:595
 - transformative participatory approaches 3:596
 - participatory action research (PAR) 3:597
 - Photovoice technique 3:597–598
 - research areas 3:599
 - subtractive schooling 3:594
 - Youth Empowerment Strategies (YES!) Program 3:597
- cultural racism 6:38
- cultural studies 6:354–359
 - adolescent behaviors 6:356
 - background information 6:355
 - characteristics 6:354
 - education systems 6:355
 - importance 6:358
 - popular culture 6:356
 - sociopolitical impacts 6:358
- cultural tools 5:499
- culture of peace movement 1:852
- curriculum
 - allied health professions
 - curriculum shift 3:55
 - professional identity 3:55
 - arts curricula 1:423–427
 - accountability considerations 1:425–426
 - arts integration 1:426
 - discipline-based arts education (DBAE) 1:424, 3:265
 - general discussion 1:423
 - globalization effects 1:423
 - subject matter development 1:423–424
 - technology-supported learning environments 1:423
 - Australia 3:418
 - bilingual/immersion education 1:293, 1:296*f*
 - business/accounting education 3:43
 - civics education programs 3:667, 3:668*t*
 - class size 6:204
 - complex systems theory 1:481–487
 - catastrophe theory 1:482, 1:482*f*
 - chaos theory 1:484, 1:485*f*
 - decision-making models 1:485, 1:485*f*, 1:486*f*
 - limitations 1:486
 - mathematical models 1:481
 - metaphorical and analogical models 1:486
 - morphodynamical models 1:484
 - recursive models 1:483, 1:483*f*
 - social interactions 1:481
 - constructivism 1:488–493
 - cognitive conflict 1:491
 - collective learning 1:492
 - error evaluations 1:491
 - general discussion 1:488
 - metaphorical models 1:489, 1:490*f*, 5:390
 - participatory curricula 1:492
 - theoretical perspectives 1:488
 - trans-level learning 1:490
 - critical theory perspective 1:494–498
 - action research 1:497
 - Bernstein's theory 1:495
 - Bourdieu's theory 1:495
 - critical race theory 1:495
 - feminist theories 1:495
 - political factors 1:497
 - poststructural theories 1:496
 - power factors 1:497
 - research areas 1:497
 - social control theory 1:494
 - sociological research background 1:494
 - curricular diversity 1:867
 - curriculum evaluations 1:335–341, 3:636–644
 - challenges 3:636
 - general discussion 1:340, 3:643
 - historical background 1:336
 - influential levels 3:642
 - international developments 1:339
 - methodologies
 - cultural perspectives 1:339
 - democratic perspectives 1:339
 - general discussion 1:339, 3:642
 - political perspectives 1:339
 - technological perspectives 1:339
 - modeling approaches 1:337
 - modeling orientations
 - adaptive evaluations 3:638
 - emergent/emancipatory evaluations 3:639
 - general discussion 3:637
 - instrumental evaluations 3:637
 - situated evaluations 3:638
 - multiple objects/multiple representations 3:642
 - purpose 1:335, 3:636, 3:641
 - recent trends 1:339
 - curriculum reform 1:510–513
 - basic concepts 1:510
 - efficacy assessments 1:512
 - influencing factors 1:510
 - modeling approaches 1:512
 - outcomes-based education (OBE) movement 1:511
 - political aspects 1:512
 - standards-based curriculum (SBC) 1:511
 - curriculum use 1:348–354
 - curriculum potential 1:349
 - definitions 1:348
 - as economic text 1:351
 - integrated analytical model 1:352, 1:352*f*
 - parental involvement 1:351
 - as political text 1:351
 - social class stakeholders 1:350
 - society as user 1:350
 - students as users 1:349
 - teachers
 - as choice makers 1:348
 - as curriculum adaptors 1:348
 - as curriculum creators 1:349
 - teacher-proof curricula 1:348
 - definitions 1:362, 1:514
 - discourse analysis 1:342–347
 - digital technology 1:345–346
 - historical perspectives 1:342
 - international perspectives 1:344
 - postmodern perspectives 1:343
 - early childhood curriculum 1:514–519
 - developmentally appropriate practices (DAP) 1:516
 - developmental theory
 - influential factors 1:517
 - limitations 1:518
 - modeling approaches
 - Bereiter–Engelmann model 1:516
 - child as active thinker 1:516
 - child as creative player 1:515
 - child as subject 1:515, 1:516
 - child unfolding over time 1:514
 - economically poor children 1:516
 - Head Start program 1:516
 - sociocultural perspectives 1:517
 - early childhood education and care (ECEC) 2:72
 - Estonia 5:560
 - gender research 1:277–282
 - background information 1:277
 - feminist perspectives 1:277
 - gendered access 1:279
 - general discussion 1:281
 - male–female continuum 1:280
 - queer studies 1:281
 - twentieth-century perspective 1:277
 - globalization 1:312–317, 1:403–410
 - background information 1:403
 - challenges 1:316
 - curriculum design 1:404
 - epistemological perspectives
 - altered conditions/state of affairs theory 1:314
 - discourse theory 1:314
 - process theory 1:315
 - project theory 1:315

- relationship theories 1:313
- general discussion 1:312, 1:409
- knowledge-based economies 1:312–313, 1:403
- lifelong learning model 1:404, 1:405*t*
- reading curricula 1:421
- reform efforts
 - assessment methods 1:407
 - classroom dynamics 1:408
 - course content 1:406
 - critical perspectives 1:408
 - institutional change 1:406
 - instructional strategies 1:407
- research and development expenditure levels 1:404*t*
- traditional learning model 1:405*t*
- government policies and programs 1:379–383
 - effectiveness 1:382
 - general discussion 1:379, 1:383
 - political perspectives
 - characteristics 1:379
 - decision-making processes 1:381
 - influential factors 1:382
 - participants 1:380
- hidden curricula 1:631
- higher education
 - computer-assisted instruction 4:420
 - educational innovations
 - design considerations 4:452
 - general discussion 4:448
 - outcomes analyses 4:451
 - gender equity 1:631
 - globalization effects 1:699
- human rights-based approach 1:287–292
 - educational rights
 - abuse prevention 1:290
 - acceptability 1:289
 - accessibility 1:289
 - adaptability 1:289
 - availability 1:289
 - children's rights 1:290
 - guiding principles 1:291
 - parent's rights 1:290
 - teacher's rights 1:290
 - Universal Declaration of Human Rights (UDHR) 1:288
 - guiding principles
 - learning about human rights 1:291
 - learning cycle 1:291, 1:291*f*
 - learning for human rights 1:291
 - learning through human rights 1:291
- indigenous curricula 1:307–311
 - Convention on the Rights of the Child 1:307
 - educational traditions 1:307
 - Gandhian model 1:307
 - national aims 1:310
 - national–local educational dilemma 1:308
 - relevance–quality–efficiency factors 1:309
 - sociopolitical factors 1:310
 - vocational education and training (VET) 1:309
- informal learning 1:44
- inquiry-oriented teacher education 7:605
- language and literacy instruction
 - beginning-reading programs 6:221
 - continua model 1:477
 - general discussion 1:435, 1:439, 1:477
 - integrated instructional strategies 1:477
 - primary schools
 - language curricula 1:436
 - literature teaching 1:437
 - reading curricula 1:436
 - writing instruction 1:437
 - secondary schools
 - adolescent literacies 1:438
 - digital technology 1:439
 - language curricula 1:438
 - literature teaching 1:438
- legal education 3:29
- literature curricula 1:447–452
 - colonial and postcolonial periods 1:449
 - critical theory-based literature 1:448
 - economic factors 1:450
 - historical background
 - general discussion 1:447
 - literary texts 1:447
 - primers 1:447
 - reader-centered approaches 1:447
 - text-based approaches 1:447
 - international perspectives 1:448
 - political factors 1:451
 - political stability 1:449
 - wartime effects 1:450
- local teacher-based curriculum development 1:283–286
 - general discussion 1:283, 1:286
 - historical background 1:283
 - illustrative example 1:284
 - professionalization process 1:283–284
- mathematics education 1:453–458
 - attained curriculum evaluation
 - assessment tasks 3:657
 - conceptual diagram 3:654*f*
 - importance 3:657
 - performance analysis 3:657
 - characteristics 3:654
 - curriculum development and revision
 - general discussion 1:454
 - Singapore 1:455, 1:455*f*, 1:456*f*
 - United States 1:456
 - curriculum reform 1:457
 - historical perspectives 1:453
 - implementation considerations 1:457
 - implemented curriculum evaluation 3:654*f*, 3:655, 3:656*t*, 3:657*f*
 - intended curriculum evaluation 3:654, 3:654*f*, 3:655*f*, 3:656*t*
 - selection guidelines 1:453
- medical education
 - characteristics 3:20
 - clinical education 3:20
 - learning processes
 - competency-based learning methods 3:22
 - problem-based learning methods 3:22
 - simulation methods 3:21
 - reform efforts 3:20
- New Zealand 3:443
- personalized learning 6:229
- pharmacy education 3:62
- physical education and sports 1:459–463
 - background information 1:459
 - formal examinations and matriculation 1:460
 - general discussion 1:463
 - models-based instruction 1:461
 - national and state curricula 1:459
 - sports education 1:461
 - Teaching Games for Understanding (TGfU) 1:461, 1:462
- popular culture 1:318–323
 - educational applications 6:356
 - research areas
 - background information 1:318
 - future directions 1:321
 - general discussion 1:320
 - popular culture as curriculum 1:320
 - popular culture–curriculum interactions 1:320
 - popular culture–curriculum relationship 1:319
 - Scholar–Baller[®] program 1:863
- postcolonial theory 1:299–306
 - background information 1:299
 - cultural interfaces 1:304
 - decolonized curriculum 1:303
 - Eurocentric educational approaches
 - cultural identity factors 1:301
 - feminism 1:303
 - implications 1:301
 - language of instruction 1:302
 - legacy insights 1:302
 - racial and ethnic concepts 1:303
 - global educational inequities 1:300
 - post-postcolonial perspectives 1:304
- poststructuralist theory 1:499–504
 - feminist poststructuralist perspectives 1:502
 - historical background
 - deconstructionism 1:501
 - general discussion 1:500
 - reconceptualization 1:501
 - structuralism 1:500
 - theoretical perspectives 1:499
 - twenty-first century perspectives 1:503

- curriculum (*continued*)
- publishing industry 1:368–373
 - America's Choice 1:371
 - current trends 1:372
 - global context 1:369
 - knowledge control and influence 1:369
 - textbook production 1:368
 - United States 1:370
 - reading curricula 1:415–422
 - curricula globalization 1:421
 - curriculum reform
 - fragmented curriculum 1:420, 1:420f
 - school literacy curriculum construction 1:421
 - schoolwide report institutionalization 1:421
 - staircase curriculum 1:419–420, 1:420f
 - standards-based change (SBC) process 1:419, 1:419f, 1:420
 - student and family ownership 1:421
 - definitions 1:415
 - primary schools 1:436
 - traditional instructional strategies 1:415, 1:416f
 - twenty-first century considerations
 - changing contexts 1:418
 - changing readers 1:418
 - changing texts and literacies 1:418
 - general discussion 1:417
 - religious and spiritual education 1:272–276
 - background information 1:272
 - recent trends and conflicts 1:272
 - theoretical perspectives
 - Charlemagne's curriculum 1:274
 - Christian education 1:274
 - Enlightenment period 1:275
 - ethical responsibilities 1:275–276
 - Greek pedagogy 1:273
 - medieval curriculum 1:274
 - Roman education 1:273
 - school-based vocational education and training 8:248
 - science 1:468–474
 - design-based research
 - curriculum development 1:468
 - curriculum use 1:350
 - design guidelines 1:472
 - design patterns approach 1:469, 1:470f, 1:471f
 - design principles approach 1:469, 1:471f
 - thermodynamics curriculum unit 1:470, 1:471f
 - future research areas 1:472
 - general discussion 1:468
 - second-language learning 1:441–446
 - background information 1:441
 - basic concepts
 - oral competency 1:441
 - proficiency orientation 1:441
 - challenges 1:445
 - Common European Framework (CEF) 1:442
 - content-based instruction 1:443
 - general discussion 1:445
 - immersion programs 1:443
 - language resource centers 1:444
 - standards 1:442
 - teacher preparation 1:444
 - social and emotional outcomes 6:675
 - social capital
 - challenges 6:100
 - formal curriculum 6:99
 - informal curriculum 6:99
 - social studies 1:411–414
 - challenges 1:411, 3:667
 - characteristics 1:411, 1:412f, 3:668f
 - competing approaches 1:411, 1:412f
 - general discussion 1:413
 - NCSS standards 1:413
 - subject-specific disciplines 1:412
 - textbooks 1:413
 - structuralist sociology 1:505–509
 - code theory
 - Bernstein's theory 1:506
 - ESSA group studies 1:507
 - pedagogic communication modalities 1:506
 - general discussion 1:509
 - historical background 1:505
 - knowledge–curriculum relationship 1:508
 - knowledge structure theory 1:508
 - Sultanate of Oman
 - basic education 5:721, 5:721f
 - postbasic education 5:722
 - preschool education 5:721
 - syllabus design 1:362–367
 - curriculum models
 - critical model 1:364
 - general discussion 1:364
 - outcomes-based model 1:364
 - process-based model 1:364
 - traditional content model 1:365
 - definitions 1:362
 - general discussion 1:366
 - high-equity/high-quality systems 1:365
 - published research
 - English-language teaching 1:363
 - general discussion 1:363
 - higher education teaching 1:364
 - technical form 1:365
 - teacher change 1:374–378
 - content knowledge 1:375
 - curriculum development connection 1:374
 - instructional techniques 1:375
 - organizational change 1:376
 - pedagogical content knowledge 1:376
 - professional learning communities 1:376
 - teacher beliefs 1:375
 - teachers' dispositions 1:376
 - unintended consequences 1:377
 - technology literacy and fluency 8:39
 - twenty-first century skills 8:158–166
 - challenges 8:164
 - framework development 8:159
 - innovative instructional strategies
 - basic concepts 8:160
 - emerging media 8:161
 - historical documents 8:164f
 - multi-user virtual environment (MUVE) interfaces 8:161
 - River City MUVE 8:162, 8:162f, 8:163f, 8:164f, 8:165f
 - situated learning 8:161
 - virtual microscopy 8:164f, 8:165f
 - knowledge-based economies 8:158
 - selected new literacies 8:159, 8:160f
 - Tyler rationale 1:499
 - Washington Accord 4:212
 - curriculum-based assessment (CBA) 2:571–576, 4:102–110
 - curriculum-based measurement (CBM)
 - basic concepts 4:103
 - individualized instructional programs 4:108
 - instructional decision making 4:104, 4:107f
 - mathematics education 2:771, 4:108
 - reading instruction 4:108
 - Response to Intervention (RTI) 4:104
 - secondary prevention strategies 4:108
 - special needs students 2:572, 2:771
 - student focus 4:106
 - definitions 2:571
 - design considerations 2:573
 - effectiveness evaluations 2:574
 - future directions 4:109
 - general curriculum impacts 2:573
 - general discussion 2:571, 2:575
 - historical background 2:572
 - importance 2:571
 - mastery measurement 4:102
 - motivational strategies 2:574
 - progress monitoring 4:102
 - relevant measurement models 2:573
 - usage guidelines 2:575
 - curriculum inquiry 1:324–334
 - challenges 1:333
 - characteristics 1:325
 - current developments
 - context expression and expansion 1:329
 - general discussion 1:329
 - internationalization processes 1:331
 - practical focus approach 1:332
 - subject matter development 1:331
 - theoretical literature development 1:332
 - curriculum theory 1:328
 - definitions and terminology 1:324
 - practical–contextual–theoretical considerations 1:325

- research trends 1:324
 - subject matter 1:326
 - teacher-based action research 1:355–361
 - future directions and challenges 1:359
 - general discussion 1:355
 - historical background
 - Collier's approach 1:357
 - continuity approach 1:358
 - democratic approach 1:358–359
 - Dewey's approach 1:356
 - Lewin's approach 1:357
 - practitioner research approach 1:359
 - Progressive Era 1:356
 - Wann review 1:355
 - topics and preoccupations 1:326
 - working structures 1:325
 - curriculum planning 1:384–389
 - curriculum domains
 - classroom curriculum planning 1:385
 - general discussion 1:384
 - institutional curriculum planning 1:384
 - interdependence factors 1:385
 - programmatic curriculum planning 1:384
 - digital technology 1:390–396
 - challenges
 - relationship considerations 1:394
 - unproductive attitudes 1:395
 - usability 1:394
 - characteristics
 - general discussion 1:390
 - interactivity 1:391
 - multimedia formats 1:390
 - socialization aspects 1:391
 - recent developments
 - expanded access 1:390
 - expanded capacity 1:390
 - personalized publishing 1:390
 - student resources
 - educational software 1:393
 - e-textbooks 1:393
 - general discussion 1:393
 - online courses 1:394
 - online learning systems 1:394
 - teaching resources
 - assessment tools 1:393
 - curriculum resources 1:392
 - digital planning tools 1:392
 - general discussion 1:392
 - lesson plans 1:392
 - online curriculum standards 1:392
 - online discussion forums 1:393
 - general discussion 1:384
 - systems theory
 - bottom-up model 1:387
 - combination model 1:388
 - general discussion 1:386, 1:388
 - top-down model 1:387
 - teacher change 1:374–378
 - content knowledge 1:375
 - curriculum development connection 1:374
 - instructional techniques 1:375
 - organizational change 1:376
 - pedagogical content knowledge 1:376
 - professional learning communities 1:376
 - teacher beliefs 1:375
 - teachers' dispositions 1:376
 - unintended consequences 1:377
 - cyber bullies 8:116
 - cyber schools 6:217
 - see also* Internet-based education
 - Cyprus
 - age-related policies and programs 8:263*t*
 - early childhood education and care (ECEC) 2:111*t*
 - inequality–education relationship 2:273*f*
 - information and communication technology (ICT) 8:25*t*
 - student achievement test performance 2:247*f*
 - test score–social background relationship 2:273*f*
 - vocational education and training (VET) 8:263*t*
 - Czech Republic
 - age-related policies and programs 8:263*t*
 - class size 2:492*f*
 - early childhood education and care (ECEC) 2:4*t*, 2:111*t*, 2:114*t*
 - earnings–education level relationship 2:307*t*
 - education levels 2:279*f*
 - higher education
 - foreign doctoral students 4:292*t*
 - institutional expenditures 1:733*t*
 - loans and grants 4:501*f*
 - parental expenditures 1:734*t*
 - student and researcher mobility 4:287*f*
 - tertiary education expenditures 4:500*f*
 - tertiary type-B program survival rates 8:464*t*
 - tuition fees 4:500*f*
 - inequality–education relationship
 - employment rate–educational attainment relationship 2:278*f*
 - test score–social background relationship 2:273*f*
 - information and communication technology (ICT)
 - computer access at home 8:25*t*
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - student proficiency and confidence 8:31*t*
 - school choice 6:285
 - self-managing schools 5:72–73
 - teaching hours 2:493*f*
 - vocational education and training (VET) 8:249, 8:251*t*, 8:252*f*, 8:253*f*, 8:263*t*, 8:435
- ## D
- Dakar World Education Forum 4:690, 6:251
 - Damasio, Antonio 1:36–37, 6:602
 - dance education
 - classroom-based assessments 3:266
 - globalization effects 1:423
 - technology-supported learning environments 1:423
 - dance teachers
 - content knowledge 7:696
 - pedagogical strategies 7:698
 - Daniel R. R. v State Board of Education* (1989) 2:717
 - Dark Green Schools (DGSS) 1:751
 - Darwin, Charles 2:870, 5:187, 6:87, 7:471
 - data envelopment analysis (DEA) 2:210
 - Data Explorer 8:51
 - data mining 7:112–118
 - advantages 7:113
 - computational statistics 7:96
 - general discussion 7:10, 7:112, 7:156
 - methodologies
 - causal data mining 7:115
 - clustering algorithms 7:114, 7:114*t*
 - data distillation for human judgment 7:114*t*, 7:116
 - discovery with models 7:114*t*, 7:115
 - domain models 7:116
 - educational applications 7:116
 - general discussion 7:113
 - illustrative example 7:117
 - prediction 7:113, 7:114*t*
 - relationship mining 7:114*t*, 7:115
 - data systems 2:178–186
 - background information 2:178
 - general discussion 2:178
 - international data systems 2:184, 2:184*t*
 - national data systems
 - accessibility 2:180
 - characteristics 2:181*t*
 - Common Code of Data (CCD) 2:178, 2:181*t*
 - crime and safety surveys 2:180, 2:181*t*
 - general discussion 2:178
 - National Assessment of Educational Progress (NAEP) 2:179, 2:181*t*
 - national longitudinal studies 2:179, 2:181*t*
 - School and Staffing Survey (SASS) 2:178, 2:181*t*
 - state and school district data systems
 - Alaska 2:184
 - Arizona Department of Education 2:183
 - California 2:183
 - Center for Analysis of Longitudinal Data in Education Research (CALDER)
 - California 2:183
 - Florida 2:182
 - general discussion 2:182
 - Missouri 2:183
 - New York 2:182
 - North Carolina 2:183
 - Texas 2:182

- data systems (*continued*)
 - Chicago Public Schools 2:183
 - general discussion 2:181
 - Nevada 2:184
 - student data systems 8:14–20
 - background information 8:14
 - educational applications 8:16
 - future directions 8:17
 - general discussion 8:19
 - key features 8:15
 - system types 8:14
- data warehousing systems 8:15
- Davis, Michael 3:48
- deafness *see* hearing impairments/loss
- Dearden, Robert 6:6, 6:109, 6:187
- Decennial Programme of Development of Education (PDDE)
 - educational policies 5:700
 - operational objectives 5:701
 - strategic objectives 5:701
- decision theory 7:119–124
 - axiomatic decision theory 7:121
 - current developments 7:123
 - decision-making process 7:119
 - decision problems
 - decision tree 7:121^f
 - definitions 7:119, 7:120
 - illustrative example 7:121, 7:123
 - intuitive solutions
 - maximum expected utility 7:121
 - minimax 7:121
 - educational applications 7:120, 7:123
 - historical background 7:120
 - normative decision theory 7:119
 - probability theory 7:378
- declarative memory 5:198, 5:198^f
- decontextualized language 5:352
- deductive reasoning 5:273
- de Finetti's theorem 7:38
- degree mills 4:600
- Delaware 1:654^f
- delayed feedback 3:226–227
- delayed gratification 6:725
- deleted-*d* jackknife 7:230
- deliberative planning theory 1:54
- delinquent behaviors
 - challenges
 - demographic changes 2:820
 - fragmented initiatives 2:821
 - school failure–juvenile delinquency relationship 2:821
- neuroticism 3:256
- occurrences 2:820
- school-based prevention programs 2:820–828
 - administrative leadership/faculty buy-in
 - academic and behavioral screening 2:823
 - data systems/data-based decision making 2:824
 - evaluation practices 2:826
 - evidence-based strategies 2:825, 2:825^f
 - general discussion 2:822
 - integration strategies 2:826
 - needs-based professional development 2:823
 - general discussion 2:826
 - goals and objectives 2:822, 2:822^t
 - social networks 6:709
- Delors Report 6:250
- Delphi technique 3:727
- Delta statistical method 7:235
- democratic curriculum evaluation model 1:338
- democratic deliberation 1:193
- democratic leadership 5:12–17
 - characteristics 5:12
 - comparison studies 5:4
 - conceptualizations 5:13
 - distributed versus democratic leadership 5:13, 5:14^f
 - empirical research approaches 5:14
 - future directions 5:16
 - theoretical perspectives 5:12
- Democratic Republic of the Congo
 - early childhood education and care (ECEC) 2:111^t
 - socioeconomic/sociopolitical characteristics 1:199
 - student achievement test performance 2:418^t
 - see also* Africa
- de Moivre, Abraham 7:467
- de Moivre distribution *see* normal distributions
- de Moivre–Laplace central limit theorem 7:234
- dendrograms *see* cluster analysis
- Denmark
 - age-related policies and programs 8:263^t
 - childhood brain development 2:57^f
 - class size 2:492^f
 - dual apprenticeship system 8:232, 8:435, 8:452
 - early childhood education and care (ECEC) 2:4^t, 2:13, 2:56^t, 2:58^t, 2:111^t, 2:114^t
 - earnings–education level relationship 2:307^t
 - education levels 2:279^f
 - gender-equity gap 2:56^t
 - higher education
 - access and equity patterns 1:689
 - foreign doctoral students 4:292^t
 - institutional expenditures 1:733^t
 - loans and grants 4:501^f, 4:502^f
 - parental expenditures 1:734^t
 - student and researcher mobility 4:287^f
 - tertiary education expenditures 4:500^f
 - tuition fees 4:500^f
 - inequality–education relationship
 - earnings differentials 2:277^t
 - employment rate–educational attainment relationship 2:278^f
 - immigrant status 2:275^f
 - test score–social background relationship 2:273^f
 - information and communication technology (ICT)
 - computer access in school 8:23^t
 - frequency of computer use in school and at home 8:26^t
 - mathematics and science instruction 8:32^t
 - student proficiency and confidence 8:31^t
 - information and communication technology (ICT) access and equity 8:97
 - leadership development programs 4:753^t, 4:754^t, 4:755^t
 - participation research
 - regular education 8:284^f
 - vocational education and training (VET) 8:283^f, 8:284^f
 - work-related vocational education and training (VET) 8:284^f, 8:285^f
 - popular adult education 1:163
 - prison education programs 8:301, 8:301^t
 - research-quality assessments 4:308
 - returns to education–education level 2:308^t
 - student achievement test performance 2:247^f
 - teacher shortages 2:494^f
 - teacher wages and salaries 2:496^f
 - teaching hours 2:493^f
 - vocational education and training (VET)
 - age-related policies and programs 8:263^t
 - career guidance 8:269, 8:271
 - dual apprenticeship system 8:232, 8:435, 8:452
 - initial vocational preparation system 8:251^t
 - participation research 8:283^f, 8:284^f, 8:285^f
 - school-based programs 8:247
 - teacher training 8:507^t
 - vocational–academic education parity 8:435
 - women in employment 2:56^f
- denotation 6:476
- dense social networks 6:705
- density-based classification methods
 - basic concepts 7:138
 - normal populations with equal variance 7:139
 - normal populations with unequal variances 7:139
 - practical applications 7:139
- dental examinations *see* school-based health clinics
- Department for Education and Science (DFES) 3:539
- depressed children 6:692
- Derrida, Jacques
 - critical theory philosophy 6:351
 - linguistic phenomenology 6:451
 - neoliberal ideologies 6:14
 - phenomenological hermeneutics 6:417
 - poststructuralist theory 1:501, 6:111, 6:175, 6:478
 - radical hermeneutics 6:65
 - reason and rationality 6:185–186
- Descartes, René 1:275, 6:80, 6:87, 6:184–185
- descriptive feedback 3:411, 3:413–414
- desegregation 2:394–401
 - academic achievement outcomes
 - advantages/disadvantages 2:398
 - court-ordered plans 2:435–436
 - methodological challenges 2:398

- research results 2:399, 2:399*t*
- desegregation techniques 2:394
- educational reform 1:692, 1:823, 1:883, 2:394, 2:688
- enrollment pattern trends
 - demographic data 2:394
 - exposure index 2:395*t*
 - racial composition 2:395*t*
 - school-level dissimilarity index 2:396*t*
- Supreme Court decisions 1:692, 1:823, 1:883, 2:394, 2:688
- white enrollment decline 2:396, 2:397*t*
- design-based research 6:367–372
- computer-supported collaborative learning (CSCL)
 - pedagogical design 5:293
 - technology design 5:293
- methodologies
 - dependent variable characterizations
 - climate variables 6:369–370
 - general discussion 6:369
 - outcome variables 6:370
 - systemic variables 6:370
 - design modifications 6:368
 - general discussion 6:368
 - implementation considerations 6:368
 - independent variable characterizations 6:370
 - multilevel analyses 6:369
 - reporting considerations 6:370
- modern experimentation conditions 6:367–368
- psychological experimentation 6:367
- research areas 6:371
- science education
 - curriculum development 1:468
 - curriculum use 1:350
 - design guidelines 1:472
 - design patterns approach 1:469, 1:470*t*, 1:471*f*
 - design principles approach 1:469, 1:471*f*
 - thermodynamics curriculum unit 1:470, 1:471*f*
- statistical power analysis 7:436–443
- summative evaluations 6:371
- design of experiments
 - basic concepts 7:6
 - general discussion 7:125
 - quality indicators 7:125
 - quasi-experimental designs
 - general discussion 7:128
 - regression discontinuity (RD) designs 7:128
 - time series designs 7:128
- randomized experimental designs
 - blocked designs 7:128
 - control group designs 7:127
 - factorial designs 7:127
 - general discussion 7:127
- single-case research designs 7:128
- statistical analysis techniques 7:129, 7:129*f*, 7:130*f*
- developed countries
 - educational production function model 2:407–411
 - academic achievement outcomes
 - cost-benefit analysis 2:409
 - general discussion 2:407, 2:410
 - quality assurance 2:408
 - school resources 2:408
 - teachers and schools versus family background 2:409
 - teacher quality 2:409
- school effectiveness 6:269–273
 - ability testing 4:5
 - contemporary research developments 6:272
 - future research areas 6:273
 - historical background 6:269
 - influencing factors 6:271
 - International Congress for School Effectiveness and Improvement (ICSEI) 4:683, 6:269
 - methodological characteristics 6:270
 - research findings 6:270
 - school effectiveness research (SER) program 5:51–52, 5:54, 6:269
 - school improvement impacts 6:272
- developing countries
 - educational leadership 4:826
 - higher education 4:590–596
 - background information 4:590
 - future directions 4:595
 - globalization effects 1:698
 - lending statistics 4:594*t*
 - new commitments 4:594*t*
 - policy frameworks 4:590, 4:593*t*
 - recent activities 4:592
 - research studies 4:594*t*
- returns to education 2:305–312, 2:412–422
 - demand interventions 2:417
 - earnings differentials 2:305
 - age-earnings profiles 2:306*f*
 - basic concepts 2:305
 - causality considerations 2:309
 - cognitive ability 2:310, 2:310*t*
 - gender differences 2:306–307, 2:309*f*
 - OECD countries 2:307*t*, 2:308*t*
 - private returns to education 2:309*f*, 2:309*t*
 - public subsidization 2:308*t*
 - years of schooling 2:305, 2:306*t*, 2:307*f*, 2:307*t*, 2:308*t*
- enrollment
 - benefits 2:419*t*
 - gender disparities 2:416*t*
 - general discussion 2:415
 - income effects 2:417
 - primary schools 2:415*t*
 - secondary schools 2:416*t*
 - TIMSS data 2:416, 2:417*t*
- future research areas 2:421
- general discussion 2:412
- government policies 2:420
- literacy probability 2:417*f*
- policy considerations 2:310
- student achievement test performance 2:417*t*, 2:418*t*
- school improvements 6:281–288
 - capacity-building improvements 6:281
 - challenges 6:286
 - comprehensive school reform models 6:286
 - curriculum and instructional practices 6:281
 - decentralization trends 6:284
 - general discussion 6:281
 - public versus private schools 6:285
 - school-based management approaches 6:284
 - school choice 6:285
 - support systems 6:282
- secondary education 6:249–254
 - Education for All (EFA) 6:250
 - education initiatives
 - access and expansion 6:252
 - delivery challenges 6:253
 - marketization 6:253
 - enrollment 2:415, 2:416*t*
 - future directions 6:253
- teachers 2:504–511
 - educational expenditures 2:504, 2:505*t*
 - government policies
 - general discussion 2:507
 - hiring alternatives 2:509
 - incentive programs 2:507
 - lessons learned 2:510
 - salaries 2:507
 - school-based management approaches 2:508
 - teacher shortages 2:509
 - teacher training 2:509
- labor markets
 - accountability and management practices 2:506
 - general discussion 2:504
 - pupil-teacher ratios (PTRs) 2:506*f*
 - recruitment and requirements 2:505
 - salary characteristics 2:505
 - salary levels 2:504
 - supply shortages 2:506*f*
 - working conditions 2:505
- Millennium Development Goals (MDGs) 2:504
- vocational education and training (VET) 8:473
- see also* postconflict societies
- developmental assessment
 - assessment improvements 3:371
 - assessment strategies 3:373
 - progress maps 3:369–377
 - achievement assessments
 - general discussion 3:373
 - learning research advances 3:375
 - reporting and analyses 3:374, 3:374*f*, 3:375*f*, 3:376*f*
 - assessment improvements 3:371

- developmental assessment (*continued*)
 - basic concepts 3:369, 3:370f
 - general discussion 3:369, 3:376
 - map construction 3:372, 3:373f
 - outcomes analyses 3:372
 - speaking progress map 3:370f
 - school learning perspectives 3:371
- developmental coordination disorders 2:579
- developmental dyslexia
 - neuroimaging studies
 - neural activation time course 5:320
 - neural networks 5:319
 - remediation effects 5:320
- developmentally delayed children 2:854
- developmental psychology 4:2
- developmental scales 4:96
- deviance information criterion (DIC) 7:456
- deviant behaviors 6:709
- Dewey, John
 - action research 1:355, 7:542
 - administrative theories 4:725
 - art education 1:424
 - contemporary liberalism 6:76
 - democratic leadership concept 5:12
 - early childhood curriculum 1:515
 - educational theory 6:90
 - Ethical–Cultural School 6:499
 - inquiry-based learning 5:297, 6:90, 7:604
 - instructional planning 3:365
 - learning theories 3:167, 5:258, 6:115
 - peace education 1:850
 - philosophical perspectives 6:3, 6:89, 6:109
 - reason and rationality 6:184–185, 6:186–187
 - reflective practices theory 6:90, 7:629
 - supervision theory 7:615
 - symbolic interactionism 6:481
 - theory–practice relationship 7:669
- diabetes *see* health considerations
- Diagnostic and Statistical Manual of Mental Disorders (DSM–IV–TR)* 2:577
- attention deficit/hyperactivity disorders (ADHD) 2:581
- child behavior check list (CBCL) 3:259
- communication disorders 2:579
- conduct disorders 2:581
- diagnostic and classification processes 3:494
- disability categories 2:578
- elimination disorders 2:582
- emotional and behavioral disorders (EBD) 2:581
- feeding and eating disorders 2:581
- learning disabilities 2:579
- mental retardation (cognitive disability) 2:578
- motor skills disorders 2:579
- oppositional defiant disorders 2:581
- pervasive developmental disorders 2:579
- reactive attachment disorder 2:582
- selective mutism 2:582
- separation anxiety disorder 2:582
- stereotypic movement disorder 2:582
- tic disorders 2:582
- diagnostic assessments 3:207, 3:276, 3:397
- dialogical evaluation 3:575
- dialogic teaching 6:700
- Diana v California State Board of Education* (1970) 2:716–717
- Dickens, Charles 5:285
- Didaktik* 1:345
- difference-in-differences (DD) 2:190
- differential geometric view model 7:344
- differential item functioning (DIF) 4:36–44
 - admissions test fairness 4:12
 - bias 4:36, 4:167
 - characteristics 4:37
 - Cochran–Mantel–Haenszel (CMH) test 7:61
 - definitions 4:37
 - detection guidelines 4:41
 - latent variable models 7:63
 - logistic regression (LR) analysis 7:62
 - purification procedures 4:40
 - Rasch model 4:116, 4:116f, 4:117f
 - research areas 4:41
 - standardized tests 4:12, 4:36
 - technique classifications
 - background information 4:37
 - item response theory (IRT) 4:38f, 4:39
 - logistic regression (LR) analysis 4:38f, 4:39
 - Mantel–Haenszel (MH) statistic 4:38
 - nonparametric DIF techniques 4:37
 - parametric DIF techniques 4:39
 - SIBTEST analysis 4:38f, 4:39
 - structural equations modeling 4:40
 - technique criteria 4:38f
 - test adaptations 4:167
- differential psychology 4:2
- differentiated evaluations 3:676
- differentiated instruction
 - gifted and talented students 2:832
 - special needs students 2:728
- diffusion tensor imaging (DTI) 5:340
- Digital Millennium Copyright Act (1998) 4:303
- digital technology 1:390–396
 - academic achievement impacts 8:109–110
 - academic values and expectations 4:436
 - access and equity patterns 8:95–101
 - Canada 8:97
 - China 8:98
 - European Union (EU) 8:97
 - general discussion 8:95
 - global access and use 8:98, 8:98f
 - improvement efforts 8:99
 - India 8:98
 - international trends 8:21–33
 - challenges 8:30
 - computer access at home 8:25, 8:25f
 - computer access in school 8:22, 8:23f
 - frequency of use at home 8:25, 8:26f
 - frequency of use in school 8:25, 8:26f
 - large-scale assessments 8:22
 - mathematics and science instruction 8:30, 8:32f
 - patterns of use 8:26, 8:27f
 - student access 8:21
 - student proficiency and confidence 8:27, 8:31f
 - theoretical perspectives 8:21
 - United States
 - home access 8:95, 8:96f
 - home use 8:95
 - outcomes analyses 8:95, 8:97
 - school access 8:96, 8:96f
 - school use 8:96
- challenges
 - adolescent literacies 1:439
 - relationship considerations 1:394
 - unproductive attitudes 1:395
 - usability 1:394
- characteristics
 - general discussion 1:390
 - interactivity 1:391
 - multimedia formats 1:390
 - socialization aspects 1:391
- computer-supported collaborative learning (CSCL) 5:290–296
 - background information 5:291
 - design-based research
 - pedagogical design 5:293
 - technology design 5:293
 - future directions 5:294
 - research approaches
 - dialogic approach 5:292
 - systemic approach 5:291
 - sociocultural perspectives 5:290
- curriculum discourse analysis 1:345–346
- digitized speech 8:153
- educational impacts 8:88–94
 - goals and objectives 8:88
 - growth factors 8:88
 - implementation considerations 8:89
 - importance 8:93
 - modeling approaches 8:92
 - outcomes analyses 8:91
 - usage measurements 8:90
- electronic portfolios 3:239, 3:246, 3:346, 3:422, 8:219
- higher education 4:419–426
 - assessment strategies 3:253
 - background information 4:419
 - curriculum 4:420
 - effective learning

- collaborative learning 4:422
- communication-based learning 4:422
- construction-based learning 4:421
- instructional strategies 4:421
- e-learning pedagogies
 - average comparative ratings 1:531*t*
 - basic concepts 1:522
 - content appropriateness and relevancy 1:522
 - course goals 1:523
 - dimensions analysis 1:532
 - diversity considerations 1:523
 - effectiveness 4:423, 4:423*t*
 - interactive learning strategies 1:524
 - learners' needs 1:523
 - learning styles 1:522
 - multimedia 1:524
- flexible study 4:423
- future directions 4:425
- teaching approaches 4:424
- image-based education research 6:500
- instructional management 8:133–139
 - background information 8:133
 - classroom network technologies
 - effectiveness 8:137
 - general discussion 8:136
 - instructional strategies 8:137
 - key features 8:136
 - future directions 8:137
 - student-response systems
 - effectiveness 8:135
 - general discussion 8:134
 - key features 8:134
 - research areas 8:134
 - teaching models 8:134
- international access patterns 8:21–33
 - challenges 8:30
 - computer access at home 8:25, 8:25*t*
 - computer access in school 8:22, 8:23*t*
 - frequency of use at home 8:25, 8:26*t*
 - frequency of use in school 8:25, 8:26*t*
 - large-scale assessments 8:22
 - mathematics and science instruction 8:30, 8:32*t*
 - patterns of use 8:26, 8:27*f*
 - student access 8:21
 - student proficiency and confidence 8:27, 8:31*t*
 - theoretical perspectives 8:21
- lifelong learning 8:185
- mathematics education 8:172–183
 - challenges 8:180
 - collaborative problem solving
 - The Adventures of Jasper Woodbury* 8:177
 - Anchored Instruction 8:177
 - Enhanced Anchored Mathematics 8:177, 8:178*f*
 - MathWorlds 8:178
 - networked devices 8:177
 - participatory simulations 8:177–178
 - serious game systems 8:178
 - SimCalc 8:178, 8:179*f*, 8:199
 - embodied cognition 8:179, 8:180*f*
 - importance 8:172
 - mathematical discovery
 - ANIMATE system 8:175, 8:176*f*
 - Dynamic Geometry systems 8:176
 - dynamic statistics packages 8:176
 - Function Probe 8:175
 - general discussion 8:174
 - Logo 8:175, 8:199
 - skills and concept learning
 - background information 8:172
 - Cognitive Tutors 8:173, 8:174*f*
 - computer algebra systems (CASs) 8:173
 - Geometric Supposer 8:173
 - hand-held graphing calculators 8:172, 8:173*f*
 - intelligent tutoring systems (ITSs) 8:173
- new media technologies 8:140–149
 - animation
 - general discussion 8:143
 - picture comprehension 8:144
 - picture design 8:144
 - realistic and dynamic pictures 8:143, 8:143*f*
 - general discussion 8:141
 - learning models
 - cognitive theory of multimedia learning (CTML) 8:145, 8:146*f*
 - design guidelines 8:148
 - dual-coding theory 8:145
 - integrated model of text and picture comprehension (IMTPC) 8:145, 8:146*f*
 - research findings 8:147
 - theoretical models 8:145
 - multimedia learning environments
 - animation 8:143
 - future directions 8:148
 - general discussion 8:141
 - learning models 8:145
 - linear and nonlinear text 8:142
 - text design guidelines 8:142
 - research findings
 - coherence and contiguity 8:147
 - general discussion 8:147
 - modality 8:147
 - redundancy 8:147
 - text design guidelines
 - additional stimulation 8:142
 - conciseness 8:142
 - general discussion 8:142
 - linguistic simplicity 8:142
 - structure and organization 8:142
 - physically disabled students 2:642
 - recent developments
 - expanded access 1:390
 - expanded capacity 1:390
 - personalized publishing 1:390
- science learning 8:197–203
 - access and equity promotion 8:202
 - challenges 8:202
 - constructivist perspectives 8:197, 8:198*t*
 - contextualization
 - authenticity 8:201
 - everyday experiences 8:201
 - future directions 8:202
 - information access 8:201
 - learning environment transformation 8:201
 - physical environment interactions
 - model construction 8:199
 - scientific phenomena 8:198, 8:198*t*
 - simulations/dynamic representations 8:199
 - student investigations 8:199
 - reflective construction of knowledge
 - program management 8:200
 - teacher support and professional development 8:200
 - working memory (WM) capacity 8:200
 - social environment interactions 8:199
 - systemic impacts 8:202
- second-language learning 8:204–210
 - background information 8:204
 - benefits
 - anywhere/anytime instruction 8:208
 - general discussion 8:207
 - immediate feedback 8:208
 - individualized study plans 8:207
 - mistakes in private concept 8:208
 - patient tutoring 8:208
 - records of achievement 8:209
 - classroom instruction 8:205
 - English-language training demand 8:204, 8:205*f*
 - functional role 8:204
 - research results and analysis 8:207
 - subject-specific uses 8:4
- student resources
 - educational software 1:393
 - e-textbooks 1:393
 - general discussion 1:393
 - online courses 1:394
 - online learning systems 1:394
- teacher learning 8:211–216, 8:222–227
 - certification and licensing 8:212
 - change management
 - educational innovations 8:224
 - general discussion 8:224
 - Second Information Technology in Education Studies (SITES) 8:224
 - understanding/self-understanding factors 8:224
 - competency standards 8:213

- digital technology (*continued*)
 curriculum frameworks 8:211
 functional role
 activity clusters 8:225*t*
 general discussion 8:225
 knowledge–innovation relationships 8:226*t*
 general discussion 8:225
 importance 8:211
 integration strategies 8:215
 knowledge-deepening versus knowledge-creation approach 8:214
 knowledge demands 8:222
 literacy skills development 8:212
 need requirements
 formal versus informal knowledge 8:223, 8:223*t*
 knowledge domains 8:222, 8:223*t*
 knowledge in practice 8:223, 8:223*t*
 knowledge levels 8:223, 8:223*t*
 socioeconomic skills 8:213
 standards 8:211
 Teaching for Understanding (TfU) 7:637
 teaching resources
 assessment tools 1:393
 curriculum resources 1:392
 digital planning tools 1:392
 as an educational tool 8:219
 general discussion 1:392
 lesson plans 1:392
 online curriculum standards 1:392
 online discussion forums 1:393
 technological impacts 8:112–118
 health-related effects
 direct and indirect physical effects 8:113
 health information accessibility 8:113
 treatment/intervention approaches 8:114
 social development
 aggressive behaviors 8:115
 bullying 8:116
 computer-mediated communication 8:116
 general discussion 8:115
 social relationships 8:116
 technology literacy and fluency 8:34
 basic concepts 8:36
 educational implications
 curriculum changes 8:39
 digital divides 8:39
 literate lives 8:38
 performance assessments 8:39
 framework development 8:37, 8:38*f*, 8:39*f*
 future directions 8:40
 historical perspectives 8:35
 multiple literacies 8:37
 social implications 8:34
see also Internet-based education
 diglossia 5:355
 Dilthey, Wilhelm 6:63, 6:415
 directed evaluation 3:574
 direct instruction (DI)
 basic assumptions 2:584
 benefits 2:586
 future directions 2:585
 literacy instruction
 beginning-reading programs 6:222
 special needs students 2:759, 2:760*t*
 special needs students 2:584–587
 basic assumptions 2:584
 future directions 2:585
 literacy instruction 2:759, 2:760*t*
 theoretical/philosophical perspectives 2:584
 Directorate for Science, Technology and Industry (DSTI) 4:588
 direct participation 6:443
 Dirichlet distribution 7:40, 7:99
 Disability Discrimination Act (UK) 2:880
 disadvantaged schools 5:58–65
 challenges
 deficit views 5:58
 effective leadership 4:777
 policies and programs 5:59
 reform impacts 5:60
 teacher turnover 5:60
 leadership 4:777–782
 challenges 4:777
 context-specific improvement 4:780
 effective leadership
 community engagement 4:779
 distributed leadership 4:778
 general discussion 4:781
 improvement strategies 4:779
 learner-centered leadership 4:779
 research background 4:778
 shared values and vision 4:778
 staff development 4:779
 improvement interventions 4:777
 leadership impacts 5:61
 urban leadership 4:809–816
 challenges 4:810
 defining characteristics 4:811
 demographic background 4:809
 general discussion 4:809
 leadership strategies 4:815
 local perspectives
 estate schooling 4:814, 4:814*f*
 extended area communities 4:814
 Leadership on the Front-Line project 4:811
 multiple communities 4:812, 4:813*f*
 single/homogenous communities 4:812, 4:812*f*, 4:813*f*
 disciplinary knowledge 6:67
 discipline-based arts education (DBAE) 1:424, 3:265
 discontinuity-based assignment 2:189
 discourse analysis 6:373–379
 analytical traditions and foci 6:373, 6:374*t*
 critical-discourse analysis (CDA) 6:373, 6:374*t*
 diversity discourse 1:823–826
 discrimination 1:826–826
 feminist perspectives 1:824
 higher education institutions 1:823
 interdisciplinary studies 1:824
 legal contexts 1:823
 social contexts 1:823
 Supreme Court decisions 1:823
 document analysis 6:381*t*, 6:382
 educational perspectives
 identities, memberships, and differences 6:376
 linguistic and ethnographic analyses of communication 6:374, 6:374*t*
 pedagogical applications 6:376
 research background 6:374
 sociocultural factors 6:376
 future challenges 6:377
 micro-level analyses 6:374, 6:374*t*
 discrete probability distributions 7:132–134
 basic concepts 7:132
 Bernoulli distribution 7:132
 binomial distribution 7:132
 general discussion 7:134
 geometric distribution 7:133
 hypergeometric distribution 7:133
 multinomial distribution 7:133
 negative binomial distribution 7:134
 Poisson distribution 7:134
 uniform distribution 7:134
 discrete-time survival model 7:213
 discriminant analysis 7:135–141
 basic concepts 7:5, 7:16
 classification-tree method 7:139, 7:140*f*, 7:141*f*
 definitions 7:135
 density-based classification methods
 basic concepts 7:138
 normal populations with equal variance 7:139
 normal populations with unequal variances 7:139
 practical applications 7:139
 illustrative example 7:136, 7:136*f*
 k-nearest-neighbor classification method 7:141
 linear regression
 basis expansion methods 7:138, 7:138*f*
 Fisher discrimination
 general discussion 7:136
 K-population example 7:137
 two-population example 7:137
 general discussion 7:135
 least-squares regression 7:135, 7:136*f*
 logistic regression (LR) analysis 4:38*t*, 4:39, 4:167, 7:136, 7:136*f*, 7:138*f*
 support vector classifiers 7:137
 two-group example 7:136*f*

- discriminant function analysis 7:447*r*
 discrimination 1:606, 1:737, 1:756, 1:824, 6:37
see also ethnicity; gender research; race/racism
 disordered behaviors 2:678
 displaced workers 8:275–279
 assistance programs
 Austria 8:277
 evaluation practices 8:278
 France 8:278
 Germany 8:278
 retraining 8:277
 Spain 8:278
 Sweden 8:278
 United States 8:277
 causal factors 8:275
 characteristics 8:275
 losses 8:276
 occupational mobility 8:275
 vocational education and training (VET) 8:236
 wages and salaries 8:276
 dispositional barriers 1:234
 dispositions 3:181–188
 assessment modes 3:185
 assessment practices and purposes 3:186, 3:186*f*
 challenges 3:183
 characteristics 3:182–183, 3:183*f*
 competence 3:182
 conceptual perspectives 3:181
 desirable dispositions 3:183
 general discussion 3:181, 3:187
 as learning outcomes 3:185
 learning styles 3:182
 reliability 3:184
 sociohistorical background 3:182
 disruptive behaviors *see* emotional and behavioral disorders (EBD)
 Dissemination Efforts Supporting School Improvement (DESSI) 6:304
 distance education 4:404–410
 accreditation 4:385
 age-participation rates (APRs) 4:404
 background information 4:404
 basic concepts 4:404
 challenges 4:409
 characteristics 4:407
 computer-assisted instruction 8:185
 cross-border higher education (CBHE) 4:511
 current trends 4:408
 dual-mode institutions 4:406
 evolutionary background 4:405
 expansion history 4:405
 general discussion 4:410
 home schooling 6:217
 India 4:522–523
 Indonesia 5:612*f*, 5:613
 lifelong learning 8:185
 Malawi 5:669
 mega-universities 4:406
 Spain 5:810
 student characteristics 4:408
 student support services 4:430
 technology-supported learning environments 8:129, 8:219
see also e-learning; Internet-based education
 distributed leadership 5:1–6
 basic concepts 4:741, 4:818
 comparison studies 5:4
 disadvantaged schools 4:778
 effectiveness 5:4
 general discussion 5:1, 5:5
 heroic versus distributive leadership model 4:744
 leader plus aspect model
 conceptual perspective 5:2
 empirical perspective 5:2
 practice aspect model
 conceptual perspective 5:3
 empirical perspective 5:3
 principal's role 5:67
 professional learning communities 5:155
 teacher leadership 5:40
 theoretical perspectives 4:764
 district-level reform efforts 6:295–302
 Boston Public Schools (BPS) 6:299
 Cross City Campaign for Urban School Reform 6:296–297
 Duval County, Florida 6:297–298
 general discussion 6:295
 key components 6:295–296
 Knowsley Local Education Authority, England 6:300, 6:301*r*
 Literacy Collaborative (LC) model 6:298
 San Diego City School District study 6:296
 state's role 6:301
 stuck versus moving districts 6:295
 District of Colombia 1:654*f*
 divergent thinking problems 5:273
 divergent thinking tests 4:171
 diversity discourse 1:823–826
 discrimination 1:826–826
 feminist perspectives 1:824
 higher education institutions 1:823
 interdisciplinary studies 1:824
 legal contexts 1:823
 social contexts 1:823
 Supreme Court decisions 1:823
 diversity exposure 1:867
 Diversity in Early Childhood Education and Training (DECET) network 2:82
 diversity research 1:693, 1:825
 divinity degrees *see* clergy education
 Djibouti 2:111*r*
 doctoral student attrition 1:827–837
 background information 1:827
 data sources 1:828, 1:829*r*
 methodology
 data-analysis strategy 1:830
 dependent variable 1:830
 descriptive statistics 1:830
 discrete-time hazard analysis 1:828
 independent variables 1:830, 1:831*r*
 research background 1:827
 research questions 1:828
 research results
 descriptive statistics
 hazard rate 1:832*f*
 life table 1:832*r*
 survival curve 1:832*f*, 1:833*f*
 general discussion 1:833
 multivariate model 1:834–835*r*, 1:831
 policy and practice implications 1:835
 theoretical frameworks 1:828
 document analysis 6:380–385
 analytical approaches
 characteristics 6:380
 discourse analysis 6:381*r*, 6:382
 ethnographic approach 6:381*r*, 6:383
 ethnomethodological ethnographic approach 6:381*r*, 6:384
 meta-synthesis approach 6:381, 6:381*r*
 general discussion 6:380, 6:384
 document-based question (DBQ) assessment 3:324
 Dodo *see* memory
 domain-referenced measurements 4:31
 Dominica
 early childhood education and care (ECEC) 2:111*r*
 higher education enrollment 1:702*f*, 1:703*f*
 Dominican Republic
 assessment system development 3:481
 early childhood education and care (ECEC) 2:111*r*
 earnings–education level relationship 2:306*r*
 higher education enrollment 1:702*f*, 1:703*f*
 private higher education (PHE) 4:623
 Doob, Joseph 7:451
 double/joint degree programs 4:511
 double sampling 3:145
 doubling coding 7:108
 drama/theater education
 classroom-based assessments 3:266
 content knowledge 7:696–697
 globalization effects 1:423
 pedagogical strategies 7:698
 technology-supported learning environments 1:423
 drawings *see* image-based education research
 dropouts
 causal factors 2:845
 college completion rates 1:803
 culturally and linguistically diverse (CLD) learners 2:846
 ethnic minorities 1:803
 high school dropouts 2:325–330

- dropouts (*continued*)
 background information 2:325
 causal factors
 community influences 2:328
 family issues 2:327
 general discussion 2:326
 individual factors 2:327
 institutional factors 2:327
 peer influences 2:328
 school characteristics 2:327
 consequences 2:326
 dropout rates 1:874, 1:875, 2:325
 general discussion 2:329
 intervention strategies
 challenges 2:328
 programmatic approaches 2:329
 systemic solutions 2:329
 racial and ethnic demographics 1:874
 India 1:592, 1:593*t*
 prevention strategies 2:845–849
 community's role 2:848
 Comprehensive Support Model (CSM) 2:846
 family's role 2:847
 general discussion 2:848
 government's role 2:848
 real-life problems 2:845
 school completion goals 2:846
 school's role 2:847
 student's role 2:847
 Uruguay 5:876, 5:879*f*
 see also student attrition
 Drucker, Peter 3:43, 7:534, 8:412
 Drumcondra Verbal Reasoning Test (DVRT)
 classification trees 7:384*f*, 7:384*t*, 7:385*t*, 7:386, 7:386*f*
 inference trees 7:388*f*
 standard deviation 7:269, 7:269*f*
 Dryfoos, Joy 6:257
 dual apprenticeship system 8:448–454
 Australia 8:452
 Austria 8:232
 background information 8:448
 characteristics 8:232
 Denmark 8:232, 8:435, 8:452
 England 8:452
 general discussion 8:453
 Germany
 characteristics 8:449, 8:449*t*
 dual system–vocational training relationship 5:585, 5:585*f*, 5:586–587, 8:435, 8:451
 general discussion 8:453
 historical background 8:450
 working principles 8:449
 Netherlands 8:232
 status considerations 8:435
 Switzerland 8:232
 terminology and jargon sources 8:347, 8:349*t*
 dual-coding theory 3:228, 8:145
 dual credit articulation 4:612–613
 dual language immersion programs 1:296, 1:296*f*
 dual-mode institutions 4:406
 dual processing self-regulation model 6:572, 6:573*f*
 Dublin Accord 4:210, 4:210*t*
 DuBois, W.E.B. 1:542, 6:351
 Dummett, Michael 6:111–112
 Duncan's multiple-range test 7:34, 7:313
 Duquesne School 6:451
 Durkheimian theory 1:505–506
 dyadic attention 5:169
 dynamic assessments 3:220–225
 adaptive testing 3:222–223
 advantages 3:220
 basic concepts 3:207
 feedback 3:221, 3:229
 functional role 3:223
 future directions 3:224
 graduated-prompts approach 3:222
 inductive reasoning 3:222
 intervention strategies 3:221
 short-term dynamic tests 3:222
 standardized versus clinical approaches 3:221
 Dynamic Geometry systems 8:176
 dynamic indicators of basic early literacy skills (DIBELS) 2:823–824
 dyscalculia 2:649
 dysgraphia 2:649
 dyslexia 2:649, 4:4, 5:319
 Dyson, Anne 6:501
 dysphasia 2:649
 dyspraxia 2:642
- ## E
- early childhood curriculum 1:514–519
 developmentally appropriate practices (DAP) 1:516
 developmental theory
 influential factors 1:517
 limitations 1:518
 modeling approaches
 Bereiter–Engelmann model 1:516
 child as active thinker 1:516
 child as creative player 1:515
 child as subject 1:515, 1:516
 child unfolding over time 1:514
 economically poor children 1:516
 general discussion 1:514
 Head Start program 1:516
 Norway 5:710
 sociocultural perspectives 1:517
 early childhood education and care (ECEC) 2:1–10, 2:98–103
 accessibility 2:110–119
 administrative issues 2:113
 enrollment rates 2:111, 2:111*t*, 2:114*t*
 governance 2:113
 growth rates 2:110
 international programs 2:118
 quality assurance 2:116
 universal versus targeted programs 2:117
 assessment strategies 3:316–322
 background information 3:316
 general discussion 3:321
 value-driven choices
 childhood perspectives 3:316
 learning theories 3:317, 3:318*t*
 methodologies and outcomes 3:320
 purposes 3:319
 Australia 2:105, 2:111*t*, 2:114*t*, 5:512
 basic concepts 2:98
 brain research 2:57*f*
 center-based programs 2:100, 2:600
 child abuse prevention 2:21, 2:22
 child-rearing practices 2:151–156
 caring perspectives 2:152
 decision-making tensions 2:155
 diversity considerations 2:151
 general discussion 2:155
 marketized childcare 2:155
 multidisciplinary/multiprofessional work 2:154
 parental involvement 2:154
 parenting style–educational outcome relationship 2:153
 social class effects
 academic achievement 2:425
 after-school activities 2:424
 literacy gaps 2:423
 role models 2:424–425
 training programs 2:153
 working mothers 2:33, 2:55, 2:151
 China 2:93–97
 future directions 2:96
 historical background 2:93
 reform effects
 consolidation processes 2:94
 funding considerations 2:95, 2:95*t*
 private providers 2:95, 2:95*t*
 transitions and uncertainties 2:96
 combination programs 2:101, 2:600
 cost-effective programs 2:38–48
 Abecedarian Early Childhood Intervention Program 2:40, 2:40*t*, 2:41*t*, 2:42*t*, 2:45*t*
 Chicago Child–Parent Centers 2:40, 2:40*t*, 2:41*t*, 2:42*t*, 2:45*t*
 class-size reduction research 2:43, 2:45*t*
 cost–benefit analysis 2:40, 2:41*t*
 cumulative effects analyses 2:39, 2:40*t*

- economic returns 2:41, 2:42*t*, 2:44, 2:45*t*, 2:46*f*
- full-day kindergarten (FDK) effects 2:43, 2:45*t*
- general discussion 2:38
- key principles of effectiveness
 - comprehensive family services 2:47
 - educational content 2:46–47
 - general discussion 2:45
 - ongoing evaluation studies 2:47
 - PK–3 programs 2:46
 - teacher preparedness 2:46
- outcomes analyses 2:38, 2:39*f*
- Perry Preschool Program study 2:40, 2:40*t*, 2:41*t*, 2:42*t*, 2:45*t*
- PK–3 programs 2:44, 2:45*t*
- research focus 2:38
- simulation studies 2:42
- supplemental programs 2:43, 2:45*t*
- transition programs 2:44, 2:45*t*
- early childhood special education (ECSE) 2:594–607
 - basic concepts 2:595
 - birth to 8 years 2:594
 - demographic data 2:598*t*
 - family involvement 2:601
 - future directions 2:604
 - government legislation 2:596, 2:597*t*
 - intervention recipients 2:597
 - intervention services 2:595
 - itinerant professionals 2:602
 - personnel preparedness 2:601
 - service delivery
 - center-based programs 2:600
 - characteristics 2:598
 - combination programs 2:600
 - home-based programs 2:599
 - service locations 2:599*t*
 - service types 2:599*t*
 - transition processes
 - early intervention to preschool transition 2:602
 - general discussion 2:602
 - preschool to kindergarten transition 2:603
- Early Learning and Development Standards (ELDS) 2:138–143
 - conceptual background 2:138
 - development and implementation guidelines
 - general discussion 2:139
 - implementation processes 2:141
 - initial decision-making stage 2:139
 - standards documentation 2:140, 2:140*t*
 - validation processes 2:140
 - general discussion 2:138
 - lessons learned
 - conceptual insights 2:141
 - implementation insights 2:142
 - procedural insights 2:142
 - status reports 2:141
- Early Learning Standards (ELS) 2:132–137
 - document content
 - general discussion 2:134
 - infant–toddler content 2:135
 - preschool content 2:135
 - general discussion 2:132
 - historical perspectives 2:132
 - implementation processes 2:135
 - importance 2:136
 - United States
 - background information 2:133
 - content resources 2:134
 - development process 2:133
- effectiveness 2:98, 5:53
- Effective Provision of Preschool Education (EPPE) 2:86–92
 - background information 2:86
 - effectiveness evaluations 2:87
 - general discussion 3:320–321
 - OECD curriculum guidelines 2:90, 2:91*t*
 - pedagogical practices
 - adult–child interactions 2:88
 - curriculum differentiation 2:89
 - discipline/behavior management policies 2:89
 - formative assessments 2:89
 - parental involvement 2:89
 - qualitative case studies 2:87
 - quality and diversity evaluations 2:87
 - results analysis 2:90
 - success factors
 - leadership and staff 2:89
 - pedagogical content knowledge 2:89
 - enrollment rates 2:4*t*, 2:56*t*
 - evaluation practices 2:25–31
 - background information 2:25
 - challenges
 - accountability controversies 2:29
 - applied research perspectives 2:28
 - contextual information 2:28
 - general discussion 2:28
 - policy ramifications 2:28
 - program quality 2:29
 - characteristics
 - confirmatory program evaluations 2:26
 - data sources 2:26
 - evaluators 2:26
 - formative evaluations 2:26
 - general discussion 2:25
 - impact evaluations 2:25
 - process evaluations 2:25
 - programmatic evaluations 2:26
 - quality evaluations 2:25
 - summative evaluations 2:26
 - international perspectives 2:29
 - methodologies
 - general discussion 2:27
 - qualitative methods 2:28
 - quantitative methods 2:27
 - quantitative methodologies
 - characteristics 2:27
 - comparison groups 2:27
 - comparison to standards approach 2:27
 - group randomization 2:27
 - pre-test/post-test methodology 2:27
 - family involvement
 - child allowances 2:2
 - general discussion 2:2
 - leave policies 2:3
 - tax benefits 2:2–3
 - Fiji 5:564
 - Finland 5:570
 - funding considerations 2:55–62
 - direct-funding models 2:60
 - government policies and programs
 - childcare programs 2:56
 - children over 3 years of age 2:59
 - children under 3 years of age 2:60
 - cost controls 2:60
 - effective funding models 2:60
 - funding extent 2:57
 - gross domestic product (GDP) percentage 2:58, 2:59*f*
 - investment levels 2:58, 2:59*f*
 - targeted programs 2:57, 2:58*t*
 - parent-subsidy models 2:60
 - Gambia 5:859
 - gender research 2:11–16
 - care regimes 2:11
 - gender equity
 - adult worker model 2:14
 - care-related social rights 2:14
 - domestic work and care patterns 2:14
 - general discussion 2:13
 - international care perspectives 2:15
 - paid parental leave 2:12, 2:14
 - public policy responses 2:15
 - male-breadwinner model 2:12
 - policies and programs 2:12
 - provision arrangements
 - Australia 2:13
 - balance of care providers 2:12
 - Denmark 2:13
 - Germany 2:13
 - Iceland 2:13
 - women's liberation movement 2:11
 - general discussion 2:101
 - Ghana 5:592, 5:592*t*
 - governance 2:6
 - growth rates 2:110
 - home-based programs 2:99, 2:599
 - Iceland 5:606, 5:607*f*

- early childhood education and care (ECEC) (*continued*)
- importance 2:110
 - Indonesia 5:612*f*, 5:613
 - infants and toddlers 2:69–74
 - child development research 2:69
 - Early Learning Standards (ELS) 2:132–137
 - content resources 2:134
 - development process 2:133
 - document content 2:134
 - general discussion 2:132
 - historical perspectives 2:132
 - implementation processes 2:135
 - importance 2:136
 - infant–toddler content 2:135
 - preschool content 2:135
 - United States 2:133
 - family pattern changes 2:70–71
 - general discussion 2:73
 - government policies and programs 2:71–72
 - grandparent care 2:71
 - health-focused models 2:70
 - human capital theory 2:70
 - participatory models 2:70
 - Piagetian perspective 2:72
 - program criticisms 2:70
 - research trends 2:71
 - theoretical perspectives 2:72
 - Vygotskian perspective 2:72
 - integrated services 2:104–109
 - Australia 2:106
 - basic concepts 2:104
 - Canada 2:106
 - challenges 2:108
 - curricular and pedagogical frameworks 2:107, 2:107*t*
 - government policies and programs 2:104
 - research background 2:104
 - structural elements 2:106
 - United States 2:105
 - workforce unification 2:106
 - interdisciplinary collaboration 2:739–744
 - collaborative models
 - basic concepts 2:739
 - effectiveness 2:742
 - practical perspectives 2:741
 - role change 2:741
 - future directions 2:743
 - general discussion 2:739, 2:743
 - professional standards 2:740
 - recommended practices 2:740
 - Jamaica 5:627, 5:628*f*
 - labor market impacts 2:66
 - Latvia 5:631
 - limitations 2:102
 - Malawi 5:667
 - Malta 5:675, 5:676*f*
 - mentoring strategies 2:162–168
 - contemporary definitions 2:163
 - definitions 2:163
 - e-mentoring 2:164, 2:166
 - functional role 2:167
 - group and peer mentoring 2:163
 - models
 - curriculum frameworks 2:165
 - general discussion 2:164
 - group-based e-mentoring 2:166
 - professional development 2:163
 - reflective practices 2:164
 - Mongolia 5:682
 - moral and social reasoning 5:280
 - New Zealand 5:694
 - nonfamily resources 2:8
 - optimal outcomes 2:144–150
 - basic concepts
 - accessibility 2:146
 - benefit measurements 2:146
 - long-term effects 2:147
 - wide effects 2:147
 - comprehensive service orientation 2:145, 2:145*f*
 - differentiated service orientation 2:144, 2:145*f*
 - early experiences 2:144
 - program evaluation
 - alternative delivery and access modes 2:148
 - general discussion 2:147
 - quality models 2:147
 - regression discontinuity (RD) designs designs 2:148
 - research areas 2:149
 - targeted provision 2:145
 - universal provision 2:146
 - Pakistan 5:729
 - parent–child relationships 5:450–456
 - autonomy support 5:451
 - early interactions 5:450
 - learning and language participation 5:451
 - out-of-home learning 5:453
 - warmth and sensitivity 5:451
 - parenting support and education 2:126–131
 - Early Head Start program (US) 2:128
 - Early Start program (New Zealand) 2:126
 - general discussion 2:2, 2:126
 - Healthy Families America (HFA) program 2:126
 - Home Instruction Program for Preschool Youngsters (HIPPY) 2:129
 - home-visiting programs (Jamaica) 2:127
 - Incredible Years program
 - basic concepts 2:129
 - United States 2:129
 - Wales 2:129
 - lessons learned 2:130
 - Nurse–Family Partnership (NFP) program 2:127
 - Sure Start program (UK) 2:57, 2:128, 2:158, 2:159
 - Turkish Early Enrichment Program (TEEP) 2:130
 - participation research 2:81–85
 - achievement gaps 2:81
 - cultural diversity 2:82
 - government policies and programs 2:82
 - inequality–education relationship 2:81
 - parent involvement 2:83
 - quality criteria 2:82
 - service accessibility 2:81–82
 - Philippines 5:866
 - Poland 5:738
 - policies and programs 2:1
 - postmodern societies 2:49–54
 - childhood brain development 2:51
 - demographic effects 2:50
 - ecological perspectives 2:49
 - family structure 2:50
 - future directions 2:54
 - government policies and programs 2:52
 - technology and media effects 2:51
 - well-being and happiness 2:52
 - private sector services
 - accessibility 2:6
 - advantages/disadvantages 2:5
 - characteristics 2:4
 - quality concerns 2:5
 - public investment benefits 2:63–68
 - advantages 2:63
 - characteristics 2:63
 - child development effects 2:64
 - cost–benefit analysis 2:67
 - government policies and programs 2:67
 - market-based perspectives 2:63
 - parental employment 2:66
 - Qatar 5:745
 - Researching Effective Pedagogy in the Early Years (REPEY) project 2:86, 3:320–321
 - Rwanda 5:760
 - social change impacts 2:55
 - Solomon Islands 5:792, 5:797*t*
 - South Africa 5:800, 5:800*t*
 - Spain 5:808, 5:809*t*
 - special needs students
 - early childhood special education (ECSE) 2:594–607
 - basic concepts 2:595
 - birth to 8 years 2:594
 - center-based programs 2:600
 - combination programs 2:600
 - demographic data 2:598*t*
 - early intervention to preschool transition 2:602
 - family involvement 2:601
 - future directions 2:604
 - government legislation 2:596, 2:597*t*
 - home-based programs 2:599
 - intervention recipients 2:597

- intervention services 2:595
- itinerant professionals 2:602
- personnel preparedness 2:601
- preschool to kindergarten transition 2:603
- service delivery 2:598
- service locations 2:599*t*
- service types 2:599*t*
- transition processes 2:602
- special needs transition processes 2:613–620
 - age 3 transition
 - child and family factors 2:616
 - community factors 2:617, 2:617*t*
 - legal considerations 2:616
 - outcomes analyses 2:617
 - state and local plans 2:617
 - conceptual framework 2:614, 2:615*f*
 - early service transitions 2:615
 - general discussion 2:619
 - implications 2:613–614
 - individualized family service plans (IFSPs) 2:615, 2:850–851
 - influencing factors
 - age 3 transition 2:616
 - kindergarten transition 2:617*t*, 2:618, 2:618*t*
 - state and local plans 2:616*f*, 2:617
 - kindergarten education 2:617*t*, 2:618, 2:618*t*
- Sri Lanka 5:816
- staff support requirements 2:162
- Sultanate of Oman 5:721
- Taiwan 5:835, 5:836*f*, 5:837*t*
- transition-to-school concepts 2:120–125
 - background information 2:120
 - continuity considerations 2:122
 - dynamic, relationships-based approach 2:123
 - general discussion 2:124
 - multilayered, multiyear approaches 2:122
 - readiness and preparedness 2:120
 - set school induction practices 2:121
 - stakeholder involvement 2:124
 - time-limited change events 2:121
- United Kingdom 2:157–161
 - Effective Provision of Preschool Education (EPPE)
 - 2:78, 2:86–92, 3:320–321, 5:53
 - enrollment rates 2:111*t*, 2:114*t*
 - government policies and programs
 - contextual perspectives 2:157
 - implications 2:161
 - integrated services 2:159
 - legislative actions 2:158
 - multiagency and transdisciplinary teams 2:159
 - support services and initiatives 2:157
 - legislative actions
 - childcare programs 2:158
 - educational entitlements 2:158
 - reform efforts 2:158
 - multiagency and transdisciplinary teams
 - effectiveness challenges 2:160
 - provisions and settings 2:159
 - workforce reform 2:159
- United States 2:2*f*, 2:132–137, 5:889, 5:890*f*
- women in employment
 - child allowances 2:2
 - childcare programs 2:55, 2:56*t*
 - economic benefits 2:55
 - gender-equity gap 2:56*t*
 - general discussion 2:2
 - leave policies 2:3
 - percentage employed 2:56*f*
 - social change impacts 2:55
 - tax benefits 2:2–3
- Zanzibar 5:896
- Early Childhood Environment Rating Scale–Revised (ECERS–R) 2:87
- early childhood interventions 2:450–455
 - academic achievement outcomes 2:64, 2:452
 - background information 2:450
 - child abuse 2:17–24
 - abuse prevention 2:22
 - adverse effects 2:20
 - general discussion 2:17
 - maltreatment characteristics
 - emotional abuse 2:18
 - neglect 2:18
 - parental substance abuse 2:18
 - physical abuse 2:17
 - sexual abuse/exploitation 2:18
 - occurrences 2:17
 - protective responses
 - child welfare 2:21
 - early childhood education and care (ECEC) 2:21
 - educational responses 2:22
 - interagency cooperation 2:21
 - legislative responsibilities 2:21
 - mandatory reporting 2:22
 - risk factors
 - ecologically-based needs assessments 2:19, 2:20*f*
 - key indicators 2:18, 2:18*t*
 - screening approaches 2:18
 - structured risk assessment 2:19*f*
 - cognitive outcomes 2:452
 - economic rationale
 - general discussion 2:450
 - market failures 2:450
 - spending capacities 2:450
 - educational effectiveness 5:53
 - feeding and eating disorders 2:581
 - general discussion 2:454
 - indirect benefits
 - female labor supply 2:454
 - general discussion 2:453
 - intergenerational effects 2:454
 - intervention characteristics 2:451
 - noncognitive outcomes 2:453
 - sociobehavioral changes 2:453
 - special needs students 2:608–612
 - benefits 2:608
 - collaborative environments 2:611
 - deaf-blind students 2:622, 2:625
 - early childhood special education (ECSE) 2:594–607
 - basic concepts 2:595
 - birth to 8 years 2:594
 - center-based programs 2:600
 - combination programs 2:600
 - demographic data 2:598*t*
 - early intervention to preschool transition 2:602
 - family involvement 2:601
 - future directions 2:604
 - government legislation 2:596, 2:597*t*
 - home-based programs 2:599
 - intervention recipients 2:597
 - intervention services 2:595
 - itinerant professionals 2:602
 - personnel preparedness 2:601
 - preschool to kindergarten transition 2:603
 - service delivery 2:598
 - service locations 2:599*t*
 - service types 2:599*t*
 - transition processes 2:602
 - general discussion 2:608
 - importance 2:608
 - limitations 2:609
 - Milwaukee Project 2:609
 - teacher preparedness 2:610
- Early Childhood Longitudinal Study (ECLS) 2:501
- Early Head Start program (US) 2:128
- Early Learning and Development Standards (ELDS) 2:138–143
 - conceptual background 2:138
 - development and implementation guidelines
 - general discussion 2:139
 - implementation processes 2:141
 - initial decision-making stage
 - age range 2:139
 - domains 2:139
 - general discussion 2:139
 - guiding principles 2:139
 - uses 2:140
 - standards documentation 2:140, 2:140*t*
 - validation processes 2:140
 - general discussion 2:138
 - lessons learned
 - conceptual insights 2:141
 - implementation insights 2:142
 - procedural insights 2:142
 - status reports 2:141
- Early Start program (New Zealand) 2:126

- early writing development 5:367–374
 - developmental stages
 - conventional orthographic writing 5:371
 - general discussion 5:368
 - phonemic writing 5:370
 - prephonetic writing
 - characteristics 5:369
 - figurative writing 5:369, 5:369f
 - global writing 5:370
 - letters in rows 5:369, 5:370f
 - scribbling 5:369, 5:369f
 - semiphonemic writing 5:370, 5:370f
 - general discussion 5:373
 - individual variations
 - cultural variations 5:371, 5:372f, 5:373f
 - language variations 5:373
 - universalities 5:371, 5:372f
 - research background 5:367
 - theoretical perspectives 5:367
- Early Years Foundation Stage (EYFS) 2:91, 2:159
- earnings differentials 2:288–297
 - ability/educational quality differences 2:293
 - Black–White test-score gap 2:294, 2:435
 - Brazil 2:296
 - change effects 2:295, 2:295t
 - cross-sectional studies 2:291, 2:291t
 - educational attainment 1:804
 - education equalization effects 2:277t, 2:291t
 - general discussion 2:288
 - human capital model 2:290
 - Israel 2:296
 - minority–White income gap 2:292f
 - minority women 2:292
 - school resources 2:313–320
 - district-level studies 2:314, 2:315f, 2:317f
 - estimation methods 2:313
 - future research areas 2:319
 - general discussion 2:313
 - methodological weaknesses 2:314
 - research challenges 2:318
 - school-level studies 2:314, 2:315f, 2:318
 - state-level studies 2:314, 2:315f, 2:317f, 2:318
 - statistical significance findings
 - rate of return estimates 2:316, 2:317f
 - United Kingdom 2:315
 - United States 2:314, 2:315f
 - women 2:315
 - variation explanations
 - age dependence 2:318
 - structural changes 2:318
 - United States 2:289, 2:289t
- East Africa 4:493
- East Asia *see* Asia
- Eastern Europe
 - background information 3:472
 - educational assessment strategies 3:472–478
 - classroom assessment 3:476
 - examination reforms 3:473
 - general discussion 3:477
 - standards
 - international comparative studies 3:475
 - sample-based assessments 3:475
 - enrollment rates
 - gender disparities 2:416t
 - primary schools 2:415t
 - secondary schools 2:416t
 - pupil–teacher ratios (PTRs) 2:506f
 - regional groupings 3:472, 3:473t
- ecclesiastic education 5:852
- Economic and Social Science Research Council (ESRC)-funded Teaching and Learning Research Programme (TLRP) 8:353
- economic growth 2:245–252
 - education quality
 - cognitive skills development 2:248
 - importance 2:251
 - institutional interactions 2:250
 - student achievement test performance 2:246, 2:247f
 - test-score measures 2:248, 2:249f
 - years of schooling 2:249f
 - general discussion 2:245
 - reform impact simulations 2:250, 2:251f
 - student achievement test performance 2:246, 2:247f
 - years of schooling 2:245, 2:246f
- economic returns *see* returns to education
- economics of education 2:193–198
 - adequate education 2:204–209
 - adult learning outcomes 1:230
 - cost–benefit analysis 2:199–203
 - academic achievement outcomes 2:409
 - definitions 2:199
 - illustrative example 2:201
 - methodologies
 - benefits measurements 2:201
 - cost estimation 2:200
 - general discussion 2:200
 - purpose 2:199
 - signaling model 2:322f
- cost–effectiveness analysis 2:199–203
 - definitions 2:199
 - illustrative example 2:202
 - methodologies
 - cost estimation 2:200
 - effects measurements 2:201
 - general discussion 2:200
 - purpose 2:199
 - definitions 2:193
- educational reform 1:737–745
- educational vouchers 2:380–387
 - achievement assessments 2:382t, 2:385
 - advantages/disadvantages 2:380
 - domestic voucher programs
 - general discussion 2:381
 - privately funded programs 2:382t, 2:383
 - publicly funded programs 2:381
 - future directions 2:385
 - international voucher programs
 - Chile 2:338, 2:370, 2:382t, 2:384, 2:419–420
 - Colombia PACES program 2:382t, 2:384
 - general discussion 2:384
 - program structure 2:380
 - publicly funded programs
 - Cleveland voucher program 2:338–339, 2:381, 2:382t
 - Milwaukee voucher program 2:338–339, 2:370, 2:381, 2:382t
 - United States 2:338–339, 2:381, 2:382t
- education production 2:196
- empirical research methods 2:187–192
 - causal effects
 - combination methods 2:191
 - difference-in-differences (DD) 2:190
 - discontinuity-based assignment 2:189
 - instrumental variables 2:190
 - propensity score matching 2:188
 - randomized assignment 2:189
 - regression analysis 2:188
 - statistical controls 2:187
 - general discussion 2:191
 - terminology and meaning 2:187
- general discussion 2:193, 2:197
- higher education 4:573–583
 - economic impacts
 - externalities 4:575
 - general discussion 4:573
 - rate of return estimates 4:574
 - educational production function model 4:576
 - funding strategies 4:579
 - general discussion 4:573
 - information access and use 4:577
 - organizational structures 4:576
- human capital theory 2:194
- markets/market failure 2:194
- parental choice 2:367–373
 - adequate information resources 2:371
 - economic perspectives 2:367
 - general discussion 2:367, 2:371
 - income and ability sorting 2:370
 - parental preferences
 - academic performance 2:369
 - general discussion 2:368
 - location 2:370
 - school atmosphere 2:370
 - school demographics 2:368
 - parents' rights 6:19

- public versus private schools 1:585, 2:368
- public investment benefits
 - early childhood education and care (ECEC) 2:63–68
 - advantages 2:63
 - characteristics 2:63
 - child development effects 2:64
 - cost–benefit analysis 2:67
 - government policies and programs 2:67
 - market-based perspectives 2:63
 - parental employment 2:66
 - state's role 2:347–353
 - benefits 2:351
 - criticisms 2:350
 - educational reform 6:301
 - education system construction 2:347
 - equal opportunities 2:350
 - general discussion 2:347
 - government failure
 - corruption considerations 2:351
 - enforced mediocrity 2:351
 - general discussion 2:350
 - inequities 2:351
 - production inefficiencies 2:350
 - rent-seeking behaviors 2:351
 - standardization 2:351
 - market failure
 - asymmetric information 2:349
 - economies of scale 2:350
 - external costs and benefits 2:348
 - general discussion 2:348
 - uncertainty analyses 2:349
 - public goods 2:348
 - school finance 2:223
 - school reform 1:737–745
 - shrinking role 2:352
 - twenty-first century skills 8:158–166
 - challenges 8:164
 - framework development 8:159
 - innovative instructional strategies
 - basic concepts 8:160
 - emerging media 8:161
 - historical documents 8:164f
 - multi-user virtual environment (MUVE) interfaces 8:161
 - River City MUVE 8:162, 8:162f, 8:163f, 8:164f, 8:165f
 - situated learning 8:161
 - virtual microscopy 8:164f, 8:165f
 - knowledge-based economies 8:158
 - selected new literacies 8:159, 8:160t
- Ecuador
 - assessment system development 3:481
 - competency-based learning model 8:407
 - early childhood education and care (ECEC) 2:111t
 - higher education enrollment 1:702f, 1:703f
 - public subsidization 2:308t
 - returns to education–education level 2:308t
- Edel, A 6:6
- Education Act (Canada) 2:880
- education action zones (EAZs) 5:93, 5:107–108
- Education Act (UK) 6:158
- Educational Administration Quarterly* (EAQ)
 - article submission data 4:733t
 - article themes 4:735, 4:736t, 4:737t
 - internationality 4:734
 - publication data 4:731, 4:732t
- educational assessments 3:161–171, 3:792–797, 4:45–49
 - alternative assessments
 - extended assignments/projects 3:205
 - formative assessments 3:205
 - general discussion 3:206
 - philosophical perspectives 3:204
 - portfolios 3:204
 - standards 3:204
 - summative assessments 3:205
 - teacher-based external assessments 3:205
 - assessment activities 3:162
 - assessment impacts 3:162
 - cross-cutting issues
 - assessment distinctions 3:166
 - developmental support 3:169
 - formative assessments 3:165
 - models of assessment versus models of learning 3:166
 - purposes 3:165
 - design guidelines 8:56–65
 - conceptual assessment frameworks (CAFs)
 - evidence-centered design (ECD) model 8:58
 - evidence model 8:60
 - schematic diagram 8:58f
 - student model 8:60
 - task model 8:61, 8:62f
 - technology supports 8:60
 - delivery supports 8:63
 - domain analysis supports 8:58
 - domain modeling supports 8:59
 - evidence-centered design (ECD) model 8:58, 8:58f, 8:59t
 - general discussion 8:58, 8:64
 - implementation supports
 - automated task generation 8:63
 - task-authoring support tools 8:62
 - domain specificity 3:163
 - educational impacts 4:69–74
 - formative assessments 4:73
 - general discussion 4:69
 - high-visibility external tests
 - influencing factors 4:72
 - potential benefits 4:71
 - problematic effects 4:71
 - historical perspectives 4:69
 - measurement purposes 4:69
 - research areas 4:73
 - functional role
 - accountability 4:46
 - achievement assessments 4:45
 - formative assessments 4:46, 4:69, 4:73
 - general discussion 4:45
 - monitoring and evaluation programs 4:45
 - student decision-making processes 4:45
 - future directions 3:796
 - general discussion 3:170
 - importance 4:49
 - key concepts and issues 3:164
 - large-scale assessments 3:163
 - measurement theory
 - general discussion 4:47
 - item response theory (IRT) 4:48
 - reliability 4:48
 - validity 4:47
 - missing data modeling techniques 7:286
 - policies and practices 3:164
 - purposes
 - accountability
 - characteristics 3:794t
 - importance 3:792
 - interventions 3:793, 3:794t
 - organizational level 3:793, 3:794t
 - performance measurement 3:792
 - summative assessments 3:792, 3:793f, 3:794t
 - amelioration
 - characteristics 3:794t
 - evaluation capacity building (ECB) 3:795
 - formative assessments 3:793f, 3:794
 - importance 3:793
 - needs assessment 3:794
 - policy-shaping contributions 3:795
 - general discussion 3:792
 - regulatory accountability 3:795
 - sociopolitical contexts 3:795
 - terminology 3:161
 - testing standards 4:45
 - testing technologies 4:49
- Educational Commission for Foreign Veterinary Graduates (ECFVG) 3:72
- educational data mining (EDM) 7:112–118
 - advantages 7:113
 - general discussion 7:112
 - methodologies
 - causal data mining 7:115
 - clustering algorithms 7:114, 7:114t
 - data distillation for human judgment 7:114t, 7:116
 - discovery with models 7:114t, 7:115
 - domain models 7:116
 - educational applications 7:116
 - general discussion 7:113
 - illustrative example 7:117

- educational data mining (EDM) (*continued*)
 - prediction 7:113, 7:114*r*
 - relationship mining 7:114*r*, 7:115
- educational data modeling 8:11–13
 - basic concepts
 - learning scenarios 8:12
 - observational scenarios 8:12
 - feedback 8:13
 - technology-enhanced learning (TEL) 8:11
- Educational Diversity Project (EDP) 1:694
- educational effectiveness 5:51–57
 - access and equity patterns 5:52
 - educational attainment 5:51
 - educational disadvantage
 - definitions 5:52
 - early childhood interventions 5:53
 - effective schools
 - characteristics 5:55, 5:55*r*
 - influencing factors 5:56
 - leadership 5:55
 - research background 5:53
 - ineffective schools 5:55
 - school effectiveness research (SER) program 5:51–52, 5:54
- educational efficiency/inefficiency 2:210–215
 - econometric approaches
 - analytical techniques 2:210
 - corrected ordinary least-squares analysis (C-OLS) 2:211
 - data envelopment analysis (DEA) 2:210
 - stochastic frontier analysis (SFA) 2:211
 - general discussion 2:214
 - inefficiency determinants 2:214
 - measurement approaches
 - environmental factors 2:213
 - school inputs 2:212
 - school outputs 2:212
 - standardized tests 2:212
 - research background 2:210
 - research findings 2:213
 - specification estimation strategies 2:212
- educational evaluation 3:518–529
 - accountability 3:600–606
 - appropriateness 3:601
 - definitions 3:600
 - design considerations
 - context-sensitive outcomes 3:603
 - data analysis and interpretation 3:604
 - data sources 3:603
 - general discussion 3:602
 - multiple measures 3:603
 - quality assurance 3:603
 - reliability 3:604
 - reporting considerations 3:605
 - responsibility assignments 3:604
 - target audience 3:605
 - general discussion 3:605
 - stakeholder involvement 3:602
 - validity and utility framework
 - general discussion 3:600
 - purpose and scope 3:601
 - conceptual frameworks 3:798–807
 - alternative-approaches framework
 - characteristics 3:800, 3:801*r*
 - consumer-oriented approaches 3:801, 3:801*r*
 - expertise-oriented approaches 3:801, 3:801*r*
 - management-oriented approaches 3:801, 3:801*r*
 - objectives-oriented approaches 3:800, 3:801*r*
 - participant-oriented approaches 3:801*r*, 3:802
 - chronological framework 3:799, 3:799*f*, 3:800*r*
 - forms-and-approaches framework
 - approaches linked to forms 3:803, 3:803*r*
 - characteristics 3:802
 - evaluation forms 3:802
 - general discussion 3:798, 3:806
 - metaevaluation approach
 - characteristics 3:804
 - pseudoevaluations 3:804, 3:805*r*
 - quasi-evaluations 3:804, 3:805*r*
 - true evaluations 3:805, 3:805*r*
 - cost analysis 3:585–591
 - basic principles 3:585
 - historical perspectives 3:585
 - importance 3:590
 - investment costs 3:586
 - main issues
 - cost-benefit analysis 3:587
 - cost-effectiveness analysis 3:588
 - cost model 3:586, 3:587*f*
 - general discussion 3:586
 - performance guidelines
 - cost-effectiveness analysis 3:589
 - cost model 3:589
 - data requirements 3:590
 - design requirements 3:589
 - objectives 3:589
 - results reporting 3:590
 - scale requirements 3:589
 - sensitivity analyses 3:590
- culturally responsive evaluations 3:592–599
 - critical pedagogy perspective 3:595
 - cultural psychology theory 3:594
 - ecological systems theory 3:593
 - intervention considerations 3:593
 - Kamehameha Early Education Program (KEEP) 3:596
 - methodologies
 - objectivist approaches 3:595
 - transformative participatory approaches 3:596
 - participatory action research (PAR) 3:597
 - Photovoice technique 3:597–598
 - research areas 3:599
 - subtractive schooling 3:594
 - Youth Empowerment Strategies (YES!) Program 3:597
- developmental frameworks
 - diversity considerations 3:524
 - general discussion 3:520
 - generalization theory 3:526
 - globalization effects 3:523
 - outcome-based accountability 3:520
 - stakeholder involvement 3:522
 - theory-driven evaluations 3:525
- evaluation definitions 3:776
- evaluation reporting and communicating 3:746–756
 - communication strategies
 - e-mail and postcards 3:750
 - newsletters and brochures 3:750
 - online collaborative tools 3:750, 3:750*r*
 - creative formats
 - audience engagement 3:750, 3:751*r*
 - cartoons 3:751, 3:751*f*
 - poetry 3:751, 3:752*f*
 - theatrical presentations 3:752, 3:752*r*
 - cross-cultural considerations 3:753
 - definitions 3:746
 - design considerations
 - audiences 3:753*r*, 3:754
 - communicating and reporting plan 3:754*f*, 3:755*f*, 3:756
 - strategies/formats and resources 3:755
 - timing 3:755
 - final evaluation report 3:747, 3:748*r*
 - interactive formats 3:749
 - visual appeal 3:747
- interactive formats
 - general discussion 3:749
 - web conferences 3:749
 - working sessions 3:749, 3:749*r*
- purposes 3:747
- stakeholders 3:753
- visual appeal
 - design, graphics, and layout 3:748
 - general discussion 3:747
 - photography 3:748
 - videos 3:749
- future directions 3:526
- general discussion 3:519, 3:806
- governance 3:607–613
 - basic concepts 3:607, 3:613
 - best practices strategies
 - design requirements 3:608
 - ethical practices 3:609
 - general discussion 3:607
 - independence 3:608
 - monitoring strategies 3:609
 - participation 3:609

- policy 3:607
- quality assurance 3:608
- transparency 3:609
- evaluation planning
 - design requirements 3:610
 - methodologies 3:611
 - resource allocation 3:610
 - responsiveness 3:610
 - scope 3:611
 - timing 3:611
- team formation and management 3:612
- historical policy evaluation 3:645–652
 - Keynesian welfare state model 3:645, 3:646*t*
 - knowledge capitalist model 3:646*t*, 3:649
 - neoliberal era 3:646*t*, 3:647
 - purpose and scope 3:645
- informal education 3:572–577
 - background information 3:572
 - challenges 3:575
 - changing focus 3:574
 - evaluation approaches
 - dialogical evaluation 3:575
 - directed evaluation 3:574
 - negotiated evaluation 3:574
 - general discussion 3:576
 - research background 3:573
- international evaluations 3:681–689
 - assessment programs 3:687*t*
 - design guidelines
 - cross-country validity 3:684
 - general discussion 3:684
 - survey implementation comparisons 3:688
 - target population comparisons 3:686
 - general discussion 3:681, 3:688
 - historical background 3:681
 - policies and programs
 - educational reform 3:684
 - importance 3:682
 - pace of change 3:683
 - performance targets 3:683
 - success factors 3:682
 - research frameworks 3:681, 3:682*t*
- knowledge organization concepts 3:519
- occurrences 3:798
- quality measures 3:775–780
 - anti-criteriological approaches 3:778
 - criteriological approaches
 - process-oriented approaches 3:777
 - product-oriented approaches 3:776
 - user-oriented approaches 3:777
 - utility-oriented approaches 3:777
 - definitions 3:775
 - educational evaluation 3:776
 - enhancement initiatives
 - checklists 3:779
 - educational training 3:780
 - evaluation capacity building (ECB) 3:779
 - general discussion 3:779
 - guidelines 3:779
 - professionalization 3:780
 - standards 3:779
 - evaluation definitions 3:776
 - situational approaches 3:778
- educational gerontology 1:175–182
 - characteristics
 - age 1:176
 - demographic data 1:176
 - learning and intelligence 1:178
 - preretirement education 1:177
 - retirement 1:176
 - stereotypes 1:177
 - educational institutions
 - accreditation 1:180
 - characteristics 1:178
 - general discussion 1:180
 - human capital theory 1:175
 - lifelong learning 1:175
 - educational goals 6:103–107
 - knowledge for knowledge's sake concept 6:103
 - rational autonomy 6:121
 - and Richard Stanley Peters 6:103–107
 - vocational training 6:104
 - well-being and happiness 6:105
- educational indicators 4:193–199
 - basic concepts 4:193
 - criteria guidelines 4:197
 - policy and research implications 4:198
 - validation processes
 - criteria guidelines 4:197
 - methodologies 4:194
 - purpose 4:196
- educational innovation 5:128–134
 - higher education 4:448–454
 - classroom assessment 4:450
 - contextual perspectives 4:448
 - curriculum
 - design considerations 4:452
 - general discussion 4:448
 - outcomes analyses 4:451
 - individual versus systemic approaches 4:449
 - information and communication technology (ICT) 4:453
 - student approaches to learning 4:451
 - student learning impacts 4:450
 - teaching and curriculum changes 4:448
 - teaching practices 4:450
 - typologies 4:449
 - whole-degree program 4:452
 - information and communication technology (ICT) 8:224
 - organizational change
 - challenges 5:131
 - complex systems theory 5:130
 - Fullan's concepts 5:132, 5:133*t*
 - future directions 5:133
 - historical perspectives 5:128
 - sociopolitical perspectives 5:131
 - teachers' beliefs 7:627
- educational leadership 4:695–703, 4:852–856
 - authentic leadership 5:7–11
 - arenas of influence 5:9, 5:9*f*
 - characteristics 5:7
 - complications
 - cross-cultural issues 5:10
 - inauthentic/false leadership 5:10
 - general discussion 5:10
 - individual valuation processes 5:7, 5:8*f*
 - valuation processes 5:7
 - basic concepts 4:741
 - contextual literacy
 - achievement measures 4:697
 - competing forces
 - continuity versus constant change 4:696
 - dependence versus independence 4:696
 - homogeneity versus heterogeneity 4:697
 - individualism versus community 4:696
 - general discussion 4:695
 - contextual perspectives 4:704–709
 - basic concepts 4:704
 - context levels
 - macrolevels 4:706
 - mesolevels 4:705
 - microlevels 4:705
 - general discussion 4:704, 4:708
 - global context
 - demographic globalization 4:708
 - economic globalization 4:707
 - general discussion 4:706
 - technological globalization 4:707
 - contingent leadership 5:45
- democratic leadership 5:12–17
 - characteristics 5:12
 - conceptualizations 5:13
 - distributed versus democratic leadership 5:13, 5:14*f*
 - empirical research approaches 5:14
 - future directions 5:16
 - theoretical perspectives 5:12
- disadvantaged schools 5:58–65
 - challenges 4:777
 - deficit views 5:58
 - policies and programs 5:59
 - reform impacts 5:60
 - teacher turnover 5:60
 - context-specific improvement 4:780

- educational leadership (*continued*)
 - effective leadership 4:777–782
 - community engagement 4:779
 - distributed leadership 4:778
 - general discussion 4:781
 - improvement strategies 4:779
 - learner-centered leadership 4:779
 - research background 4:778
 - shared values and vision 4:778
 - staff development 4:779
 - improvement interventions 4:777
 - leadership impacts 5:61
 - urban environments 4:809–816
- dissatisfaction measures 4:855
- distributed leadership 5:1–6
 - basic concepts 4:741, 4:818
 - comparison studies 5:4
 - disadvantaged schools 4:778
 - effectiveness 5:4
 - general discussion 5:1, 5:5
 - heroic versus distributive leadership model 4:744
 - leader plus aspect model
 - conceptual perspective 5:2
 - empirical perspective 5:2
 - practice aspect model
 - conceptual perspective 5:3
 - empirical perspective 5:3
 - professional learning communities 5:155
 - school-to-school collaboration 6:279, 6:279f
 - teacher leadership 5:40
 - theoretical perspectives 4:764
- diverse cultures 4:824–831
 - future research areas 4:829
 - influencing factors 4:824
 - key themes
 - change agents 4:825
 - cultural influences 4:828
 - developing countries 4:826
 - general discussion 4:825
 - leader development programs 4:827
 - successful leadership 4:826
 - research background 4:824
- diversity considerations 4:832–838
 - demographic data 4:832
 - general discussion 4:832, 4:836
 - inclusion and equity strategies 4:834
 - leadership approach characteristics 4:833
 - school community participation 4:834
 - student experiences 4:833
- early childhood education and care (ECEC) 2:89
- educational effectiveness 5:55
- esthetics 4:783–789
 - general discussion 4:783
 - organizational artifacts 4:784
 - organizational capital 4:787
 - organizational contexts 4:783
 - organizational culture 4:786
 - organizational leadership 4:784
 - organizational power 4:786
 - research areas 4:788
- ethical practices 4:717–723
 - empirical analyses 4:720
 - future research areas 4:721
 - international scholarly networks 4:681, 4:721
 - societal context 4:717
 - theoretical perspectives
 - administrators and school leaders 4:719
 - multiple perspective approach 4:719
 - recent developments 4:718
- expectations
 - characteristics 4:852
 - commitment 4:854
 - determinants 4:853
 - general discussion 4:852
 - improvement strategies 4:853
 - performance standards 4:853
- general discussion 4:695, 4:855
- higher education
 - leadership structures 4:246
 - university leaders 4:238–244
 - diversity considerations 4:238
 - future research areas 4:242
 - historical perspectives 4:238
 - literature reviews 4:238
 - literature types 4:239
 - presidential leadership 4:239, 4:240
 - senior management teams 4:239, 4:241
- historical background 4:724–730
 - administrative theories 4:724
 - civil rights movement 4:726
 - efficiency improvements 4:724
 - female leadership 4:725
 - postpositivist perspectives 4:727
 - professional managers 4:725
 - scientific management approach 4:724
 - Theory Movement 4:725
 - twentieth-century perspectives 4:727
 - twenty-first century perspectives 4:728
 - United Kingdom 4:726
- information and communication technology (ICT) 5:120–127
 - educational impacts
 - community relations 5:121, 5:121f
 - general discussion 5:120
 - instructional strategies 5:121, 5:121f
 - literacy instruction 5:121, 5:121f
 - social networks 5:120, 5:121f
 - influencing factors 5:125, 5:125f
 - integration framework 5:123f
 - requirements and standards 5:121
 - social and legal challenges 5:124
 - teacher leadership 5:122, 5:123f
 - technology-mediated leadership development programs 5:123, 5:123r
- instructional leadership 5:18–25
 - basic concepts 5:19, 5:20f
 - empirical analyses
 - general discussion 5:20
 - research findings 5:21
 - trends 4:742, 5:21
 - historical background 5:18
 - international trends 4:759
 - research trends 4:742
 - shared leadership models 5:22
 - student learning impacts 5:45
- leadership development programs 4:752–761
 - diverse cultures 4:827
 - general discussion 4:752, 4:759
 - importance 4:752
 - international approaches 4:752, 4:753r
 - international trends
 - andragogic principles 4:757
 - centralized and decentralized programs 4:753, 4:754r
 - communication and cooperation 4:756
 - course-based learning 4:758r
 - experience-oriented/application-oriented methods 4:757, 4:758r
 - explicit goals and objectives 4:758
 - instructional leadership 4:759
 - integral leadership 4:759
 - integrative leadership 4:759
 - knowledge production processes 4:757
 - leadership for improvement 4:757
 - modular and multiphase designs 4:756, 4:756r
 - moral leadership 4:759
 - partnership models 4:754
 - personal development 4:756
 - posttransformational leadership 4:759
 - preparatory qualifications 4:755, 4:755r
 - problem-based learning methods 4:758
 - program duration 4:755, 4:755r
 - quality assurance 4:753
 - school's core purpose 4:759
 - team leadership capacity 4:757
 - theory–practice relationships 4:754
 - transformational leadership 4:759
 - workplace learning 4:758
 - technology-mediated leadership development programs 5:123, 5:123r
- leadership for learning 4:746–751
 - assessment programs 4:748
 - belief and value systems 4:751
 - characteristics 4:747r
 - curricular programs 4:748
 - instructional program 4:747
 - learning communities 4:748

- organizational culture 4:750
- performance feedback 4:747
- resource allocation and use 4:749
- reward systems 4:750
- school–community relations 4:751
- school-to-school collaboration 6:279, 6:279f
- social advocacy 4:750
- vision and mission 4:746
- managerial leadership 4:759, 5:45
- moral leadership 4:759, 5:45
- National College for School Leadership (NCSL) 3:539, 4:762, 5:93, 6:243
- organizational change
 - challenges 5:131
 - complex systems theory 5:130
 - educational innovation 5:131
 - Fullan's concepts 5:132, 5:133t
 - future directions 5:133
 - historical perspectives 5:128
 - sociopolitical perspectives 5:131
- organizational savvy
 - collaborative networks 4:699
 - general discussion 4:698
 - mechanistic–living system transitions 4:698
 - personalization considerations 4:699
 - professional learning communities 4:700
 - social capital 4:700
 - thin–deep democracy transitions 4:698
- participative leadership 5:45
- philosophical perspectives 4:710–716
 - empiricism 4:710
 - epistemology 4:714
 - general discussion 4:710
 - ontology 4:712
 - theoretical perspectives
 - general discussion 4:711
 - scientific approach 4:710
 - values 4:714
- principal's role 5:67
- productive schools 4:746–751
 - assessment programs 4:748
 - belief and value systems 4:751
 - characteristics 4:747t
 - curricular programs 4:748
 - instructional program 4:747
 - learning communities 4:748
 - organizational culture 4:750
 - performance feedback 4:747
 - resource allocation and use 4:749
 - reward systems 4:750
 - school–community relations 4:751
 - social advocacy 4:750
 - vision and mission 4:746
- professional research journals 4:731–739
 - article themes 4:735, 4:736t, 4:737t
 - background information 4:738
 - Cases in Educational Leadership* (CEL) 4:732
 - editorial processes 4:732, 4:733t
 - Educational Administration Quarterly* (EAQ)
 - article submission data 4:733t
 - article themes 4:735, 4:736t, 4:737t
 - internationality 4:734
 - publication data 4:731, 4:732t
 - Educational Management Administration & Leadership* (EMAL)
 - article submission data 4:733t
 - article themes 4:735, 4:736t, 4:737t
 - internationality 4:734
 - publication data 4:731, 4:732t
 - electronic journals 4:732
 - internationality 4:734
 - International Journal of Educational Management* (IJEM) 4:734
 - Journal of Educational Administration* (JEA)
 - article submission data 4:733t
 - article themes 4:735, 4:736t, 4:737t
 - internationality 4:734
 - publication data 4:731, 4:732t
 - Journal of Leadership in Education* (JLE) 4:734
 - publication data 4:731
 - refereed articles 4:733
 - research methodologies 4:733
 - selected journals 4:732t
 - submission data 4:733t
- research areas 4:702
- research results analyses 4:766
- research trends 4:740–745
 - background information 4:740
 - backward mapping 4:742, 4:743f
 - empirical analyses 5:21
 - general discussion 4:744
 - heroic versus distributive leadership model 4:744
 - impacts on students 4:741, 5:21
 - pedagogical content knowledge 4:742, 4:743f
 - professional learning communities 4:743
 - transformational–instructional leadership relationship 4:742
- satisfaction measures 4:854
- school–higher education institution (HEI) partnerships 7:761
- school improvement 5:135–142
 - challenges 5:138
 - collaborative networks 5:92
 - conceptualizations 5:135
 - leadership conceptualizations 5:136
 - research areas 5:138
 - research perspectives
 - cultural perspectives 5:137
 - general discussion 5:136
 - political perspectives 5:137
 - rational–technical model 5:136
- school restructuring role 5:66–71
 - changing roles
 - cultural role 5:69
 - environmental role 5:69
 - instructional role 5:69
 - political role 5:68
 - conflicts and tensions
 - accountability 5:67
 - organizational restructuring 5:68
 - professionalism 5:67
 - structural coupling expectations 5:68
 - distributed leadership 5:67
 - general discussion 5:66
 - implications 5:70
 - professional learning communities 5:66
 - reculturing processes 5:66
 - results analysis 5:69
 - rule changes 5:69
 - social justice 5:67
 - stakeholder relationships 5:69
- smart practices 4:701
- social capital 5:85–91
 - background information 5:85
 - basic concepts 5:87
 - impact on educational success 5:88
 - impact on leadership roles and responsibilities 5:89
 - as an intellectual construct 5:86
 - social trust–child well-being relationship 5:87t
- special education 2:691
- strategic leadership 5:34–39, 5:45
 - basic concepts 5:34
 - characteristics
 - activities and functions 5:34, 5:38t
 - culture and shared values development 5:36
 - distinctive competencies development 5:37
 - network development 5:37
 - organizational process development 5:35
 - personnel development 5:35
 - general discussion 5:38
 - social intelligence 5:37
 - student learning impacts 5:45
- student learning impacts 5:45–50
 - achievement assessments 5:47
 - general framework 5:45, 5:46f
 - leadership models 5:45
 - limitations 5:46
 - mediating variables 5:46f, 5:47
 - moderating variables 5:48
 - research trends 4:741, 5:21
 - unit of analysis challenge 5:48
- student voice 4:803–808
 - basic concepts
 - co-leadership role 4:805
 - consulting role 4:804
 - general discussion 4:804
 - leadership roles 4:805

- educational leadership (*continued*)
- benefits 4:806
 - challenges
 - anxiety 4:806
 - credibility 4:807
 - general discussion 4:806
 - inclusivity 4:806
 - innovation issues 4:807
 - contemporary opportunities 4:803
 - general discussion 4:807
 - historical perspectives 4:803
 - sustainable development reform 6:292
 - system leadership 5:78–84
 - activities and capabilities 5:80, 5:81*f*, 5:82*f*
 - basic concepts 5:78
 - challenges 5:83
 - development opportunities 5:81
 - functional roles 5:79
 - general discussion 5:78
 - school-to-school collaboration 6:279, 6:279*f*
 - teacher leadership 5:40–44
 - basic concepts 5:40
 - distributed leadership 5:40
 - educational impacts
 - school level 5:41
 - student level 5:42
 - teacher level 5:42
 - information and communication technology (ICT) 5:122, 5:123*f*
 - leadership for learning 4:818
 - organizational development considerations 5:40
 - professional development considerations 5:43
 - theoretical perspectives
 - context matters 4:763
 - core practices 4:764
 - distributed leadership 4:764
 - learning-centered leadership 4:763
 - polyphonic work theory 4:765
 - research background 4:762
 - succession planning 4:764
 - sustainable leadership 4:765
 - transformational leadership 5:158–164
 - characteristics
 - general discussion 5:159
 - instructional program management 5:161
 - organization design 5:160
 - people development practices 5:159
 - setting directions 5:159
 - comparison studies 5:4
 - criticisms 5:162
 - historical background 5:158
 - influencing effects 5:158
 - instructional program management
 - buffering functions 5:162
 - general discussion 5:161
 - instructional support 5:162
 - monitoring operations 5:162
 - staffing considerations 5:161
 - international trends 4:759
 - organization design
 - collaborative cultures 5:160
 - community engagement 5:161
 - general discussion 5:160
 - restructuring considerations 5:161
 - school networks 5:161
 - people development practices
 - general discussion 5:159
 - individualized support 5:160
 - intellectual stimulation 5:160
 - leadership models 5:160
 - research trends 4:742
 - school improvement 5:138
 - setting directions
 - general discussion 5:159
 - group goal acceptance 5:159
 - high-performance expectations 5:159
 - shared vision 5:159
 - student learning impacts 5:45
 - transformative leadership 5:26–33
 - benefits 5:31
 - criticisms 5:31
 - ethical leadership 5:26
 - evolutionary background 5:29
 - general discussion 5:26
 - modern perspectives 5:30
 - origins 5:26
 - theoretical perspectives 5:27*t*
 - transactional leadership 5:27, 5:27*t*
 - transformative leadership 5:27*t*, 5:28
 - transformative learning theory 5:28
 - transforming leadership 5:27, 5:27*t*
 - twenty-first century perspectives 4:728
 - university leaders 4:238–244
 - diversity considerations 4:238
 - future research areas 4:242
 - historical perspectives 4:238
 - literature reviews 4:238
 - literature types 4:239
 - presidential leadership 4:239, 4:240
 - senior management teams 4:239, 4:241
 - urban environments 4:809–816
 - challenges 4:810
 - defining characteristics 4:811
 - demographic background 4:809
 - general discussion 4:809
 - leadership strategies 4:815
 - local perspectives
 - estate schooling 4:814, 4:814*f*
 - extended area communities 4:814
 - Leadership on the Front-Line project 4:811
 - multiple communities 4:812, 4:813*f*
 - single/homogenous communities 4:812, 4:812*f*, 4:813*f*
 - workforce reform 6:242–248
 - basic principles
 - performance leadership emphasis 6:243
 - qualified teacher status (QTS) 6:242
 - remodelling project 6:242
 - standards 6:242
 - current trends 6:242
 - justification
 - functionalist position 6:244
 - general discussion 6:244
 - socially critical position 6:245
 - outcomes analyses 6:246
 - workplace mistreatment/abuse 4:790–796
 - abuse effects 4:793
 - abusive behavior typologies 4:791, 4:792
 - basic concepts 4:790
 - general discussion 4:790
 - modeling approaches 4:791
 - research approaches
 - future directions 4:794
 - methodologies 4:794
 - published results 4:793
 - theoretical perspectives 4:791
 - educational longitudinal studies 2:179, 2:181*t*
 - Educational Management Administration & Leadership* (EMAL)
 - article submission data 4:733*t*
 - article themes 4:735, 4:736*t*, 4:737*t*
 - internationality 4:734
 - publication data 4:731, 4:732*t*
 - educational management organizations (EMOs) 4:681–687
 - Commonwealth Council for Educational Administration and Management (CCEA/M) 4:681
 - International Congress for School Effectiveness and Improvement (ICSEI) 4:683, 6:269
 - international scholarly networks 4:686
 - support services 2:338
 - educational measurement *see* educational assessments
 - educational privatization 2:337–341
 - background information 2:337
 - educational management organizations (EMOs) 2:338
 - evaluation processes 2:339
 - financial approaches
 - educational vouchers 2:338
 - evaluation processes 2:341
 - fees 2:339
 - philanthropic resources 2:339
 - tuition tax credits 2:339
 - general discussion 2:341
 - privatization criteria evaluations 2:340
 - regulation considerations 2:341
 - school characteristics

- charter schools 2:337
 - home schooling 2:337, 6:214
 - private schools 2:337
 - public schools 2:337–338
 - support services 2:338, 2:341
- educational production function model 2:402–406, 2:407–411
 - academic achievement outcomes
 - cost–benefit analysis 2:409
 - general discussion 2:407, 2:410
 - quality assurance 2:408
 - school resources 2:408
 - teachers and schools versus family background 2:409
 - basic concepts 2:402, 2:412
 - developed countries 2:407–411
 - academic achievement outcomes
 - cost–benefit analysis 2:409
 - general discussion 2:407, 2:410
 - quality assurance 2:408
 - school resources 2:408
 - teachers and schools versus family background 2:409
 - teacher quality 2:409
 - developing countries 2:412–422
 - demand interventions 2:417
 - enrollment
 - benefits 2:419*t*
 - gender disparities 2:416*t*
 - general discussion 2:415
 - income effects 2:417
 - primary schools 2:415*t*
 - secondary schools 2:416*t*
 - TIMSS data 2:416, 2:417*t*
 - future research areas 2:421
 - general discussion 2:412
 - government policies 2:420
 - literacy probability 2:417*f*
 - student achievement test performance 2:417*t*, 2:418*t*
 - estimation methods
 - attrition bias 2:414
 - basic concepts 2:413
 - bias reduction methods 2:415
 - endogenous program placement bias 2:414
 - measurement error bias 2:414
 - omitted variable bias 2:414
 - selection bias 2:414
 - higher education funding 4:576
 - models
 - accountability 2:405
 - basic assumptions 2:403
 - evidential research 2:405
 - general assumptions 2:403
 - general discussion 2:402
 - multilevel models 2:405
 - student outcomes 2:403
 - value-added models (VAMs) 2:405
 - student incentives 2:342
 - teacher quality 2:409, 2:438
- educational program evaluation 3:706–712
 - basic concepts 3:707, 7:151, 7:151*f*
 - comparison studies 3:707
 - evaluation frameworks
 - basic concepts 3:707
 - credible evidence 3:710
 - data collection strategies 3:710
 - evaluation focus 3:710
 - lessons learned 3:711
 - logic modeling 3:709, 3:709*f*
 - program descriptions 3:709
 - results analysis and interpretation 3:710
 - stakeholder involvement 3:708
 - usage facilitation 3:708*f*, 3:711
 - general discussion 3:711
 - impact evaluations 3:706
 - outcome evaluations 3:706
 - process evaluations 3:706
 - purpose 3:706
 - science education 3:660–666
 - accountability considerations 3:660
 - background characteristics 3:664
 - contextual evaluations 3:661
 - data analysis and reporting 3:664
 - formative evaluation tools 3:663*t*
 - outcomes evaluations 3:663
 - process model 3:660, 3:661*f*
 - process skills 3:662*t*
 - program implementation 3:662, 3:663*t*
 - variable measures 3:664
- social studies/civic education programs 3:667–673
 - assessment validity and reliability 3:668
 - challenges 3:667
 - classroom implementation
 - assessment practices 3:670
 - content decisions 3:670
 - general discussion 3:670
 - instructional impacts 3:671
 - curriculum 3:667, 3:668*t*
 - general discussion 3:672
 - recommendations 3:672
 - research background 3:671
- stakeholders 3:629–635
 - empirical research results 3:631
 - evaluation activities 3:629
 - evaluation utilization impacts 3:633
 - importance 3:633
 - participation guidelines 3:522, 3:708
 - Program Evaluation Standards*
 - participation guidelines 3:629
 - theoretical perspectives 3:630
 - theoretical perspectives
 - broad involvement 3:631
 - deep involvement 3:631
 - organizational learning perspective 3:630
 - Program Evaluation Standards* 3:630
 - social justice perspective 3:630
 - summative evaluations 3:706
- educational reform
 - comprehensive school reform models 6:286, 7:559
 - critical theory perspective 6:352
 - desegregation 1:883
 - district-level reform efforts 6:295–302
 - Boston Public Schools (BPS) 6:299
 - Cross City Campaign for Urban School Reform 6:296–297
 - Duval County, Florida 6:297–298
 - general discussion 6:295
 - key components 6:295–296
 - Knowsley Local Education Authority, England 6:300, 6:301*t*
 - Literacy Collaborative (LC) model 6:298
 - San Diego City School District study 6:296
 - state's role 6:301
 - stuck versus moving districts 6:295
 - elementary education 7:729
- ethnic minorities 1:737–745
 - alternative educational settings 1:742
 - anti-racist education 1:739
 - assimilation policies 1:737
 - background information 1:737
 - core issues
 - differential performance outcomes 1:740
 - general discussion 1:740
 - school-leaving impacts 1:742
 - suspensions and exclusions 1:742
 - underachievement 1:740
 - cultural pluralism 1:737
 - government policies and programs 1:743
 - multicultural education 1:738
 - national curriculum development 1:740
 - post-war migration impacts 1:737
 - French-language-based classroom assessments 3:432
- higher education 1:698–704
 - diversification guidelines 1:700
 - globalization effects 1:698
 - goals and objectives 1:701–702
 - lifelong learning 1:699
 - public university funding cuts 1:701, 1:707
 - World Conference on Higher Education 1:698, 4:601
- music teacher education 7:707
- Organisation for Economic Cooperation and Development (OECD) 5:72
- organizational change 5:128
- sustainable development 6:289–294
 - educational practices
 - Finland 6:292
 - London Borough of Tower Hamlets 6:293
 - importance 6:289

- educational reform (*continued*)
 - integrated and evidence-based theories 6:291
 - theoretical perspectives 6:293
 - unsustainability era
 - age of innovations 6:290
 - large-scale reform (LSR) efforts 6:290–291
 - reculturing efforts 6:290
 - restructuring efforts 6:290
- United States 6:261–268
 - civil rights movement 6:263
 - funding considerations 6:266
 - future policies and programs 6:265
 - general discussion 6:261
 - global impact 6:267
 - historical perspectives 6:261, 6:303
 - No Child Left Behind Act (2001) 6:265
 - organizational and cultural barriers 6:263
- vocational teacher education 8:505, 8:506*f*
- whole-school designs 6:303–309
 - middle schooling 6:236, 6:236*f*, 6:237*t*
 - social change impacts 6:303
- United States
 - educational implications 6:307
 - education system characteristics 6:303
 - implementation fidelity/reliability 6:306
 - large-scale assessments 6:303
 - reform efforts 6:303
 - research findings 6:305
 - research implications 6:308
 - sustainability 6:307
- educational statistics 1:581–588
 - choice effects 1:585
 - continuing disparities 1:584
 - demographic diversity 1:582
 - digital technology implications 1:584
 - disparity measures 1:581
 - education beyond borders 1:586
 - global village concept 1:587
 - twenty-first century perspectives 1:581, 1:587
 - United States 1:582
- Educational Testing Service (ETS)
 - admissions tests 4:7
 - automated essay scoring (AES) 4:21, 4:175, 8:51
 - background information 3:681
 - technology literacy and fluency 8:36, 8:37
 - test development and fairness 4:58, 4:59*t*
- educational vouchers 2:380–387
 - achievement assessments 2:382*t*, 2:385
 - advantages/disadvantages 2:380
 - Catholic schools 2:355
 - Chile 2:338, 2:370, 2:382*t*, 2:384
 - domestic voucher programs
 - general discussion 2:381
 - privately funded programs 2:382*t*, 2:383
 - publicly funded programs
 - Cleveland voucher program 2:338–339, 2:381, 2:382*t*
 - Milwaukee voucher program 2:338–339, 2:370, 2:381, 2:382*t*
 - other voucher programs 2:383
 - United States 2:338–339, 2:381, 2:382*t*
 - educational privatization 2:338
 - future directions 2:385
 - international voucher programs
 - Chile 2:370, 2:382*t*, 2:384
 - Colombia PACES program 2:382*t*, 2:384
 - general discussion 2:384
 - program structure 2:380
- Education and Training 2010 programme 4:672
- education development 4:411–418
 - challenges
 - accredited programs 4:415
 - aging demographics 4:414
 - curricula vocationalization 4:414
 - new technologies 4:414
 - structural instabilities 4:415
 - theoretical perspectives 4:415
 - contextual perspective and background 4:411
 - educational approaches
 - background information 4:412
 - learner-focused approaches 4:413
 - organization-focused approaches 4:413
 - sector-focused approaches 4:413
 - teacher-focused approaches 4:413
 - educational models 4:412
 - future directions
 - distributed roles 4:416
 - general discussion 4:415
 - international and national collaboration 4:416
 - key performance indicators 4:415
 - professionalization 4:416
 - general discussion 4:416
 - terminology and meaning 4:411
- education feminism 6:58
- Education for All (EFA)
 - expansion efforts 6:250
 - goals 6:249
 - indigenous curricula 1:307
- International Institute for Educational Planning (IIEP) 4:690
- Pakistan 5:728
- Solomon Islands 5:797
- Suriname 5:826–827
- Uganda 1:795
- universal primary education (UPE) 6:250
- Education for All Handicapped Children Act (EAHCA, 1975)
 - adapted physical activity programs 2:521
 - autism spectrum disorders (ASD) 2:661–662
 - culturally and linguistically diverse (CLD) learners 2:657
 - deaf-blind students 2:621
 - historical background 2:596, 2:689, 2:717
 - paraprofessionals 2:780
 - school-based services
 - government policies and programs 2:850
 - health clinics 3:555
 - health impaired students 2:634–635
 - specific learning disabilities (SLDs) 2:647
 - see also* Individuals with Disabilities Education Act (IDEA, 1990); Individuals with Disabilities Education Improvement Act (IDEIA, 2004)
- education for sustainable development 1:746–752
 - background information 1:750
 - basic concepts 1:750
 - educational implications 1:747
 - educational role 1:749
 - future directions 1:751
 - general discussion 6:148
 - importance 1:751
 - Judeo-Christian traditions 1:747–748
 - modern society values and cultural perspectives 1:747
- Education for Sustainable Development (ESD) 6:289
- education–health association 1:667–672, 2:169–177
 - alcohol consumption 1:670
 - background information 1:667
 - causal relationship
 - behavioral risk factors 1:547, 1:669–670, 2:174
 - enhanced cognitive abilities 2:176
 - environmental factors 2:170, 2:257
 - genetic risk factors 2:174
 - health insurance 1:669, 2:175
 - life expectancy effects 1:602, 1:603*f*, 1:603*t*, 1:668–669, 2:173*f*
 - pathway diagram 2:171*f*
 - research background 2:169
 - social deprivation 2:173
 - social standing 1:231, 2:172, 2:253–259
 - stress factors 2:171
 - general discussion 2:176
 - health-care services 1:669
 - international perspectives 1:670
 - lifestyle behaviors 1:669–670
 - measurement approaches 1:667
 - mortality rates 1:602–609
 - causal factors
 - discrimination 1:606
 - inequality 1:604
 - lifestyle behaviors 1:604, 1:605*t*
 - preventable deaths 1:605, 1:605*f*
 - sense of control 1:606
 - social psychological resources 1:606
 - social structural conditions 1:604
 - social support 1:606
 - stratification 1:604
 - working conditions 1:604
 - demographic characteristics
 - context of residence 1:607, 1:608*f*
 - family conditions 1:607
 - gender 1:606
 - infant/childhood mortality 1:607*t*

- marital status 1:607
 - race and ethnicity 1:606, 1:607*r*
 - rural environments 1:608*f*
 - socioeconomic status (SES) 1:607
 - suburban environments 1:608*f*
 - urban environments 1:608*f*
 - life expectancy 1:602, 1:603*f*, 1:603*r*
 - policy implications 1:608
 - socioeconomic status (SES) 1:668–669, 2:173*f*
 - typology
 - chronic diseases 1:602
 - general discussion 1:602
 - HIV/AIDS 1:602
 - infant/childhood mortality 1:602
 - suicide 1:602
 - research areas 1:671
 - research results 1:547, 1:668
 - self-rated health 1:668
 - social and psychological components 1:669
 - socioeconomic status (SES)
 - causal relationships 1:231, 2:175
 - educational attainment 1:667
 - external benefits 2:253
 - mortality rates 1:607, 1:668–669, 2:173*f*
 - student achievement effects 2:253, 2:258
 - Education Leader Development Programs (ELDPs) 4:827
 - Education Reform Act (1988) 1:740, 3:263, 6:242
 - education research
 - critical race theory 1:779–786
 - background information 1:779
 - centrality of experiential knowledge 1:781, 6:343
 - centrality of race 1:780
 - critical legal theory 1:779
 - educational applications 1:783
 - educational theory reconceptualizations 1:783
 - empirical educational research 1:783
 - future implications 1:784
 - narrative practices 1:782
 - policy and practice research 1:783
 - related areas of study 1:782
 - ethnomethodology 6:392–400
 - conversation analysis (CA) 6:394
 - formal analysis concept 6:393–394
 - invisible colleges/college of education
 - classroom instruction studies 6:397
 - local study–situated action research 6:396
 - research background 6:395
 - research background
 - Garfinkel's model 6:393
 - general discussion 6:392
 - invisible colleges/college of education 6:395
 - mid-twentieth century developments 6:393
 - sequential analysis 6:394
 - significance 6:398
 - grounded theory 6:409
 - naturalistic inquiry 6:485
 - philosophical perspectives 6:29–35
 - educational policies 6:30
 - educational practices 6:30
 - ethical and political significance 6:33
 - general discussion 6:30, 6:34
 - objectivity and subjectivity 6:32
 - philosophical contributions 6:29
 - qualitative/quantitative debate 6:31
 - research perspectives 6:29
 - research methodologies 7:125–131
 - evidence-based practices 7:125, 7:130
 - experimental design
 - general discussion 7:125
 - quality indicators 7:125
 - quasi-experimental designs 7:128
 - randomized experimental designs 7:127
 - single-case research designs 7:128
 - statistical analysis techniques 7:129, 7:129*f*, 7:130*f*
 - historical background 7:125
 - quasi-experimental designs
 - general discussion 7:128
 - regression discontinuity (RD) designs 7:128
 - time series designs 7:128
 - randomized experimental designs
 - blocked designs 7:128
 - control group designs 7:127
 - factorial designs 7:127
 - general discussion 7:127
 - research design
 - causal studies 7:126
 - descriptive studies 7:126
 - general discussion 7:126
 - quasi-experimental designs 7:128
 - randomized experimental designs 7:127
 - relational studies 7:126
 - single-case research designs 7:128
 - statistical analysis techniques 7:129, 7:129*f*, 7:130*f*
 - symbolic interactionism 6:481–487
 - basic assumptions 6:481
 - basic concepts 6:482
 - exemplary models
 - George Herbert Mead 6:483
 - Howard Becker 6:484
 - Peter Woods 6:484
 - focus and scope 6:481
 - historical background 6:481
 - participant observation 6:445
 - relevancy 6:486
 - scholarly associations 6:483
 - test score measurements 7:267–273
 - central moments 7:267
 - dispersion
 - general discussion 7:267, 7:272
 - interquartile range (IQR) 7:268, 7:268*f*
 - median absolute deviation (MAD) 7:269
 - range 7:268, 7:268*f*
 - standard deviation 7:268, 7:269*f*
 - variance 7:268, 7:269*f*
 - distribution shapes
 - bimodal distributions 7:271*f*
 - kurtosis 7:271*f*, 7:272
 - normal distributions 7:270, 7:270*f*
 - skewness 7:271, 7:271*f*
 - kurtosis 7:267–268, 7:269*f*, 7:271*f*, 7:272
 - normal distributions 7:267, 7:270, 7:270*f*
 - skewness 7:267–268, 7:269*f*, 7:271, 7:271*f*
 - visual data 6:498–504
 - background information 6:498
 - current research contributions 6:501
 - research background
 - digital technology 6:500
 - field studies and experiments 6:499
 - Hine's photography 6:499
 - Houston's contribution 6:499
 - image-elicitation techniques 6:500
 - microethnography studies 6:500
 - significance 6:503
- Education Sector Strategy Update* (World Bank) 4:592, 4:593*r*
- EduSpeak 4:24
- effective leaders 4:746–751
 - assessment programs 4:748
 - belief and value systems 4:751
 - characteristics 4:747*r*
 - curricular programs 4:748
 - disadvantaged schools
 - community engagement 4:779
 - distributed leadership 4:778
 - general discussion 4:781
 - improvement strategies 4:779
 - learner-centered leadership 4:779
 - research background 4:778
 - shared values and vision 4:778
 - staff development 4:779
 - instructional programs 4:747
 - learning communities 4:748
 - organizational culture 4:750
 - performance feedback 4:747
 - resource allocation and use 4:749
 - reward systems 4:750
 - school–community relations 4:751
 - social advocacy 4:750
 - vision and mission 4:746
- effective learning 4:461–466
 - background information 4:461
 - computer-assisted instruction
 - collaborative learning 4:422
 - communication-based learning 4:422
 - construction-based learning 4:421

- effective learning (*continued*)
 - instructional strategies 4:421
 - Self-Regulated Learning (SRL) 4:461, 4:462
 - Student Approaches to Learning (SAL) 4:461, 4:462
 - student engagement 4:464
 - teaching guidelines 4:465
- Effective Provision of Preschool Education (EPPE) 2:86–92
 - effectiveness evaluations 2:87, 5:53
 - general discussion 3:320–321
 - OECD curriculum guidelines 2:90, 2:91*t*
 - pedagogical practices
 - adult–child interactions 2:88
 - curriculum differentiation 2:89
 - discipline/behavior management policies 2:89
 - formative assessments 2:89
 - parental involvement 2:89
 - preschool education 2:78
 - qualitative case studies 2:87
 - quality and diversity evaluations 2:87
 - results analysis 2:90
 - success factors
 - leadership and staff 2:89
 - pedagogical content knowledge 2:89
- effect size
 - common measures 7:447*t*
 - functional role
 - background information 7:448
 - confidence intervals 7:448
 - interpretive framework 7:448
 - hierarchical linear models 7:446, 7:447*t*
 - measurement approaches 7:445
 - research background 7:445
 - standardized mean difference
 - multivariate D_M 7:445
 - univariate d (Cohen's d) 7:445
 - usage guidelines 7:449*f*
 - variance-accounted-for measures
 - basic concepts (η^2 , R^2) 7:445
 - corrected variance-accounted-for measures 7:446, 7:447*t*
 - η^2 versus partial η^2 7:446, 7:447*t*
 - general discussion 7:445
- Egypt
 - academic values and expectations 4:438
 - early childhood education and care (ECEC) 2:111*t*
 - information and communication technology (ICT) 8:25*t*
 - student achievement test performance 2:247*f*
 - Trends in International Mathematics and Science Study (TIMSS) 5:645*t*
- Eight Year Study 3:706, 6:304
- Einstein, Albert 7:451
- elaboration feedback 3:226
- elaboration learning strategies 5:325
- elaboration strategies 5:325
- Elderhostel 1:178–179
- elder learning 1:175–182
 - characteristics
 - age 1:176
 - demographic data 1:176
 - learning and intelligence 1:178
 - preretirement education 1:177
 - retirement 1:176
 - stereotypes 1:177
 - educational institutions
 - accreditation 1:180
 - characteristics 1:178
 - general discussion 1:180
 - human capital theory 1:175
 - lifelong learning 1:175
- e-learning 8:102–108
 - Arab world 1:521–535
 - basic concepts 8:102
 - current trends 4:408
 - educational data modeling 8:11–13
 - basic concepts
 - learning scenarios 8:12
 - observational scenarios 8:12
 - feedback 8:13
 - technology-enhanced learning (TEL) 8:11
 - effectiveness research
 - student learning outcomes 8:106
 - student retention 8:106
 - e-learning pedagogies
 - average comparative ratings 1:531*t*
 - basic concepts 1:522
 - content appropriateness and relevancy 1:522
 - course goals 1:523
 - dimensions analysis 1:532
 - diversity considerations 1:523
 - effectiveness 4:423, 4:423*t*
 - interactive learning strategies 1:524
 - learners' needs 1:523
 - learning styles 1:522
 - multimedia 1:524
 - electronic portfolios 3:239
 - evaluation policies and practices 3:530–538
 - challenges 3:536
 - clarification processes
 - context 3:531, 3:531*t*, 3:532*t*, 3:535
 - criteria 3:531*t*, 3:533, 3:533*t*, 3:535
 - data collection and analysis 3:534, 3:536
 - evaluands 3:531*t*, 3:532, 3:533*t*, 3:535
 - methodology selection 3:533, 3:536
 - questions 3:533, 3:535
 - stakeholders 3:531*t*, 3:532, 3:535
 - completion processes
 - comparison studies 3:534, 3:536
 - metaevaluation 3:534, 3:536
 - recommendations 3:534, 3:536
 - reporting 3:534, 3:536
 - flexible frameworks 3:530
 - illustrative example
 - comparison studies 3:536
 - completion processes 3:536
 - context clarification 3:535
 - criteria clarification 3:535
 - data collection and analysis 3:536
 - evaluand clarification 3:535
 - general discussion 3:534
 - metaevaluation 3:536
 - methodology selection 3:536
 - question clarification 3:535
 - recommendations 3:536
 - reporting 3:536
 - stakeholder clarification 3:535
 - key components 3:531*f*
 - opportunities 3:536
- flexible learning 4:441–447
 - background information 4:441
 - basic concepts 4:423
 - blended learning 4:445
 - criticisms 4:445
 - definitions 4:441
 - delivery challenges 4:444
 - delivery effectiveness 4:443
 - democratization of learning 4:444
 - flexible delivery 4:442
 - general discussion 4:441, 4:446
 - information and communication technology (ICT) 4:442
 - philosophical perspectives 4:444
 - responsiveness and availability 4:445
- general discussion 8:107
- higher education
 - background information 1:521
 - case-study analysis
 - average comparative ratings 1:530, 1:531*t*
 - top-rated factors 1:532
 - critical factors 1:525*f*, 1:533
 - enrollment growth 4:598
 - faculty
 - administrative skills 1:526
 - average comparative ratings 1:531*t*
 - competency standards 1:525
 - dimensions analysis 1:532
 - faculty engagement 1:526
 - innovativeness 1:526
 - pedagogical proficiency 1:526
 - recognition 1:527
 - technical skills 1:526
 - framework dimensions
 - critical factors 1:525*f*
 - faculty 1:525
 - general discussion 1:522
 - institutions 1:529

- learner support 1:528
 - pedagogy 1:522
 - technological infrastructure 1:527
- institution
 - appropriate processes 1:529
 - average comparative ratings 1:531*r*
 - dimensions analysis 1:533
 - financial feasibility and sustainability 1:529
 - leadership commitment 1:529
 - quality commitment 1:529
- interactive learning strategies
 - basic concepts 1:524
 - course goals 1:524
 - feedback 1:524
 - flexibility 1:524
- learner support
 - appropriate support services 1:528
 - average comparative ratings 1:531*r*
 - dimensions analysis 1:533
 - faculty responsiveness 1:528
 - future directions 4:431
 - general discussion 1:528
- multimedia
 - diversity considerations 1:524
 - relevancy 1:524
- pedagogy
 - average comparative ratings 1:531*r*
 - basic concepts 1:522
 - content appropriateness and relevancy 1:522
 - course goals 1:523
 - dimensions analysis 1:532
 - diversity considerations 1:523
 - interactive learning strategies 1:524
 - learners' needs 1:523
 - learning styles 1:522
 - multimedia 1:524
- research objectives and methodology 1:530
- technological infrastructure
 - accessibility 1:528
 - appropriateness 1:527
 - average comparative ratings 1:531*r*
 - dimensions analysis 1:532
 - reliability 1:527
- image-based education research 6:501
- Internet cafes 8:186
- lifelong learning 8:186
- literacy instruction 8:154
- new media technologies 8:140–149
 - animation
 - general discussion 8:143
 - picture comprehension 8:144
 - picture design 8:144
 - realistic and dynamic pictures 8:143, 8:143*f*
 - general discussion 8:141
- learning models
 - cognitive theory of multimedia learning (CTML)
 - 8:145, 8:146*f*
 - design guidelines 8:148
 - dual-coding theory 8:145
 - integrated model of text and picture comprehension (IMTPC) 8:145, 8:146*f*
 - research findings 8:147
 - theoretical models 8:145
- multimedia learning environments
 - animation 8:143
 - future directions 8:148
 - general discussion 8:141
 - learning models 8:145
 - linear and nonlinear text 8:142
 - text design guidelines 8:142
- research findings
 - coherence and contiguity 8:147
 - general discussion 8:147
 - modality 8:147
 - redundancy 8:147
- text design guidelines
 - additional stimulation 8:142
 - conciseness 8:142
 - general discussion 8:142
 - linguistic simplicity 8:142
 - structure and organization 8:142
- open and distance learning (ODL) 4:408, 8:129
 - purposes
 - at-risk students 8:105
 - educational enhancement 8:105
 - general discussion 8:104
 - replacement courses
 - advanced levels 8:105
 - at-risk students 8:105
 - rural areas 8:105
 - Rwanda 5:764
 - school–community relations 8:126, 8:129
 - secondary education
 - international programs 8:104
 - North America 8:103
 - see also* Internet-based education; technology-supported learning environments
- electroencephalography (EEG) 5:317
 - see also* neuroscience research
- electronic bulletin boards 4:431
- electronic communication 8:126
- electronic game systems 8:112–113
- electronic portfolios
 - assessment strategies
 - classroom assessment 3:422
 - group work assessments 3:346
 - informal learning 8:244–245
 - records of achievement 3:246
 - teacher education 3:565
 - basic concepts 3:239
 - as an educational tool 8:219
- elementary education 7:725–731
 - accreditation 7:728
 - background information 7:725
 - challenges
 - educational reform 7:729
 - professionalism 7:729
 - teacher shortages 7:729
 - general discussion 7:730
 - India 1:593*r*
 - in-service training 7:728
 - instructional accommodations 2:731
 - Internet-based education 8:102
 - Philippines 5:866
 - preservice education programs
 - academic and professional coursework 7:727
 - characteristics 7:726
 - examinations 7:727
 - practicum programs 7:728
 - tertiary degree requirements 7:727
- reading instruction
 - combined programs 6:225
 - computer-assisted instruction 6:224
 - cooperative learning 6:224
 - curricula 1:465, 6:223
 - instructional-process programs 6:224
 - requirements 6:223
 - social studies curriculum 1:411
 - system characteristics 7:726
 - teacher education programs 7:526
- elimination disorders 2:582
- Elizabeth (case study) 3:515
- El Salvador
 - assessment system development 3:481
 - early childhood education and care (ECEC) 2:111*r*
 - earnings–education level relationship 2:306*r*
 - higher education enrollment 1:702*f*, 1:703*f*
 - information and communication technology (ICT) 8:25*r*
 - private higher education (PHE) 4:623
 - public subsidization 2:308*r*
 - returns to education–education level 2:308*r*
 - school-based management approaches 2:508
- e-mail communication 4:431, 8:126
- emancipation 6:134
- Emancipation Proclamation (1863) 1:691
- embezzlement 2:239
- embodied cognition 8:179, 8:180*f*
- embodied cultural capital 6:100
- embodied learning 1:16
- e-mentoring 2:164, 2:166
- emergent/emancipatory evaluations 3:639
- emergent literacy 8:150

- emotion
 - cognition–emotion relationship 5:237–244
 - affect 6:547
 - causal factors
 - cognitive-appraisal hypothesis 5:239
 - emotional arousal and valence 5:240
 - pedagogical challenges 5:242
 - results analysis 5:242
 - definitions 5:238
 - research background 5:237
 - components
 - cognitive aspects 6:547
 - embodied aspects 6:546
 - emotional intelligence 6:547
 - functional roles 6:544
 - definitions 5:238, 6:544
 - general discussion 6:547
 - volitional control 6:724–731
 - background information 6:724
 - characteristics 6:724
 - effortful control
 - academic work habits 6:726
 - emotional regulation 6:725
 - learning task management strategies 6:726
 - research background 6:725
 - future research areas 6:728
 - intervention programs
 - general discussion 6:727
 - middle schools 6:727
 - primary schools 6:727
 - secondary schools 6:728
 - undergraduate settings 6:728
 - emotional abuse *see* workplace mistreatment/abuse
 - emotional and behavioral disorders (EBD) 2:678–687
 - affective traits
 - basic concepts 3:256
 - coping behaviors 3:257
 - neuroticism 3:256
 - test anxiety 3:257
 - assessment strategies 2:681
 - causal factors
 - academic performance 2:680
 - cultural influences 2:680
 - family issues 2:680
 - general discussion 2:680
 - individual child 2:680
 - videogames 8:115
 - characteristics
 - behavioral characteristics
 - externalizing behaviors 2:679
 - general discussion 2:679
 - internalizing behaviors 2:679
 - cognitive characteristics 2:679
 - general discussion 2:679
 - gifted and talented students 2:590
 - cognitive–behavioral interventions
 - behavioral principles 2:683
 - general discussion 2:683
 - punishments 2:683
 - reinforcers 2:683
 - teacher-appropriate replacement behaviors
 - general discussion 2:684
 - problem solving skills 2:684
 - self-regulation 2:684
 - social skills training 2:684
 - definitions 2:678
 - diagnostic and classification processes 2:581
 - early intervention services 3:260
 - future directions
 - alternative educational settings 2:685
 - general discussion 2:685
 - minority students 2:685
 - suspension or expulsion 2:685
 - general discussion 2:685
 - Individuals with Disabilities Education Act (IDEA, 1990) 2:678, 3:260
 - placement strategies 2:682
 - prereferral strategies 2:681
 - prevalence 2:680
 - prevention strategies
 - classroom discipline 2:683
 - general discussion 2:682
 - individual-student-centered interventions 2:683
 - school-wide discipline 2:683, 2:820–828
 - referral strategies
 - behavioral intervention plans 2:682
 - eligibility determination 2:681
 - functional behavioral assessments 2:682
 - school-based services 2:853
- emotional arousal regulation 6:573
- emotional experiences 6:591–596
 - adult learning 1:16
 - basic concepts
 - affective tendency 6:591
 - core effect 6:591
 - general discussion 6:591
 - influencing factors
 - appraisals and judgments 6:591
 - social construction 6:592
 - educational settings 6:591
 - future research areas 6:594
 - general discussion 6:595
 - student experiences
 - learning and achievement 6:593
 - self-regulation and motivation 6:592
 - teacher experiences
 - emotional rules and labor 6:594
 - unpleasant emotional experiences 6:594
- emotional intelligence 2:691, 3:258, 5:186, 6:547
- emotional regulation
 - coping strategies 6:570–575
 - acute versus chronic stressors 6:570
 - classroom interventions 6:574
 - coping–volition relationship
 - dual processing self-regulation model 6:572, 6:573f
 - emotional arousal regulation 6:573
 - volitional control 6:573
 - effectiveness
 - general discussion 6:570
 - new stressor situations 6:571
 - situation assessments 6:571
 - strategy selection 6:571
 - modeling approaches 6:574
- empathy 6:597–604
 - basic concepts 6:597
 - brain research 6:602
 - empathic intelligence 6:600
 - engagement 6:601
 - enthusiasm 6:601
 - expertise 6:601
 - historical research 6:598
 - infant studies 6:599
 - intelligent caring 6:601
 - narrative practices 6:600
 - psychoanalytical research 6:598
 - see also* affect
- empirical Bayes methods 7:7, 7:142–149
 - beta–binomial model 7:145
 - educational applications 7:148
 - extensions 7:143f, 7:145
 - gamma–Poisson model 7:144
 - National Assessment of Educational Progress (NAEP) 7:146, 7:147f, 7:148t
 - normal–normal model 7:142, 7:143f, 7:144f
 - school performance example 7:142, 7:143t, 7:144, 7:144f
 - shrinkage effects 7:144f, 7:146, 7:147f
 - small-area estimation 7:414
- empirical–rational change strategies 7:557
- employability skills 8:368–375
 - Australian Employability Skills Framework
 - attributes 8:369, 8:370t
 - skills 8:369
 - VET-sector employability skills 8:370t
 - definitions 8:368
 - development guidelines
 - assessment strategies 8:372
 - recording methods 8:372
 - skill inclusion qualifications 8:370
 - teaching strategies 8:371
 - emerging trends
 - employer–learner mismatch 8:373
 - responsibility factors 8:373
 - skill requirements 8:373
 - general discussion 8:373

- generic skills 8:368
- target learners 8:370
- employee training models 1:157*t*, 1:159
- empiricism 4:710, 6:87–88
- encopresis 2:582
- endurance 6:559
- engaged education 1:711–718
 - engagement concepts 1:711, 1:712*f*
 - experiential learning 1:713, 1:714*f*
 - importance 1:716
 - intensive field experiences
 - basic concepts 1:714
 - curriculum development and logistics 1:716, 1:717*t*
 - goals and objectives 1:715
 - social and cultural change 1:714
- Engeström's activity theory 7:637, 7:638*f*
- engineering profession 3:45–50
 - accreditation programs
 - Accreditation Board for Engineering and Technology (ABET) 3:47
 - Bologna Accord (1999) 3:47
 - United Kingdom 3:46
 - Washington Accord 4:209–215
 - ABET accreditations 4:211
 - assessment impacts 4:213
 - background information 3:47
 - curriculum impacts 4:212
 - general discussion 4:214
 - graduate attributes 4:210
 - historical background 4:210
 - impacts 4:211
 - outcome-based accreditation 4:211
 - pedagogy impacts 4:212
 - professional competencies 4:210
 - related international partnerships 4:210, 4:210*t*
 - research initiatives 4:213
 - signatories 4:210*t*
 - American Society for Engineering Education (ASEE) 3:47
 - American Society of Civil Engineers (ASCE) 3:47
 - background information 3:45
 - chartered engineers 3:46
 - codes of conduct 3:48
 - engineering qualifications 3:46
 - female participation 1:687
 - France 3:47
 - future directions 3:49
 - gender equity 1:630, 1:686
 - globalization effects 4:209
 - learned societies 3:45
 - mergers 3:46
 - professional associations 3:45, 3:47
 - professional autonomy 3:48
 - professionalization process 3:48
 - United Kingdom 3:46
 - United States 3:47
 - universities 3:48
 - Washington Accord 4:209–215
 - background information 3:47
 - general discussion 4:214
 - graduate attributes 4:210
 - historical background 4:210
 - impacts
 - ABET accreditations 4:211
 - assessments 4:213
 - curriculum 4:212
 - outcome-based accreditation 4:211
 - pedagogy 4:212
 - research initiatives 4:213
 - professional competencies 4:210
 - related international partnerships 4:210, 4:210*t*
 - signatories 4:210*t*
 - women of color 1:688
- England 5:550–554
 - accreditation processes 4:384–389
 - Catholic schools 2:356
 - classroom assessment policies 3:425–431
 - assessment beliefs and practices 3:430
 - current practices 3:425
 - France–England national assessment comparison studies 3:427, 3:427*t*, 3:429*t*
 - general discussion 3:425
 - PACE (primary assessment, curriculum, and experience) study 3:425
 - QUEST (quality of education systems transnationally) project 3:428
 - class size 6:200
 - common schools versus separate schools 6:44
 - competitive schooling 2:335
 - dual apprenticeship system 8:452
 - early childhood education and care (ECEC)
 - Effective Provision of Preschool Education (EPPE) 2:86–92, 3:320–321
 - integrated services 2:105
 - education system
 - background information 5:550
 - district-level reform efforts 6:300, 6:301*t*
 - educational finance 5:554
 - 14–19 education and skills reform 5:552
 - higher education 5:550
 - bachelor's degrees 5:553
 - doctorates 5:554
 - loans and grants 5:554
 - master's degrees 5:554
 - professional qualifications 5:554
 - quality assurance 5:550
 - reporting and regulation practices 4:275*t*
 - structural and operational characteristics 5:553
 - reform efforts 6:266
 - school governors 5:550
 - school inspections 5:550
 - school workforce reform 6:242–248
 - current trends 6:242
 - functionalist position 6:244
 - justification 6:244
 - outcomes analyses 6:246
 - performance leadership emphasis 6:243
 - qualified teacher status (QTS) 6:242
 - remodelling project 6:242
 - socially critical position 6:245
 - standards 6:242
 - structural and operational characteristics
 - further education and training (FET) 5:553
 - higher education 5:553
 - preprimary education 5:551
 - primary education 5:551
 - private education 5:552
 - secondary education 5:551
 - sixth forms/sixth-form colleges 5:552
 - teacher training 5:554
- General Certificate of Secondary Education (GCSE) 3:386
- inequality–education relationship 2:273*f*
- informal learning 8:243
- information and communication technology (ICT) 8:25*t*
- leadership development programs 4:753*t*, 4:754*t*, 4:755*t*, 4:758*t*
- national assessment programs 4:92
- physical education and sports 1:459
- school choice 6:285
- school inspections
 - historical background 3:720
 - purpose 5:550
 - research results 3:722
- school-to-school collaboration 6:274
- self-managing schools 5:72–73
- special education background 2:714
- teacher education programs
 - program curriculum 7:528
 - sociocultural factors 7:527
- teacher evaluations 3:677
- test score–social background relationship 2:273*f*
- vocational education and training (VET) 8:383, 8:386, 8:434, 8:472
 - see also* United Kingdom
- English 11 assessment 8:50
- English as a foreign language (EFL) 1:428–434
 - communicative language teaching (CLT) model
 - characteristics 1:429
 - criticisms 1:429
 - curricular advantages 1:432, 1:432*f*
 - educational conflicts 1:431
 - key assumption and mission 1:430
 - persuasive discourses 1:433
 - student resistance 1:432
 - definitions 1:428
 - future directions 1:433
 - knowledge production and circulation 1:428
 - notational/functional linguistics 1:429
 - structural linguistics 1:428
- English-speaking Caribbean *see* Caribbean countries

- Enhanced Anchored Instruction (EAI) 2:771
 Enhanced Anchored Mathematics 8:177, 8:178*f*
 Enlightenment period
 critical theory perspective 6:134, 6:176, 6:349
 historical narratives 1:97–98
 rational autonomy 6:121
 reason and rationality 6:80, 6:183
 religious and spiritual education 1:275
 special education 2:714
 entrepreneurial university model 1:734, 4:249
 enuresis 2:582
 environmental education 6:146–151
 basic concepts 6:147
 importance 6:146
 knowledge requirements 6:149
 primary concerns 6:146
 sociopolitical factors 6:150
 sustainable development considerations 1:746–752
 background information 1:750
 basic concepts 1:750
 educational implications 1:747
 educational role 1:749
 future directions 1:751
 general discussion 6:148
 importance 1:751
 Judeo-Christian traditions 1:747–748
 modern society values and cultural perspectives 1:747
 epileptic seizures 2:635, 8:113
 episodic memory 5:198, 5:198*f*, 5:199, 5:200*f*, 5:342
 episteme 7:671
 e-portfolios
 assessment strategies
 classroom assessment 3:422
 group work assessments 3:346
 informal learning 8:244–245
 records of achievement 3:246
 teacher education 3:565
 basic concepts 3:239
 as an educational tool 8:219
 equal partnerships 7:760*r*
 equating methods *see* test equating and linking
 Equatorial Guinea
 early childhood education and care (ECEC) 2:111*r*
 student achievement test performance 2:418*r*
 equity 4:57
 ERASMUS *see* European Region Action Scheme for the Mobility of University Students (ERASMUS)
 e-rater scoring program 4:21, 8:51
 Erikson, Erik 6:154
 Eritrea 2:111*r*
 e-SOL (Standards of Learning) assessment 8:49
 ESSA group studies 1:507
 essay questions
 automated essay scoring (AES) 4:20–26
 available programs 4:21
 basic concepts 4:20
 educational applications
 automated speech-recognition scoring programs 4:24
 Criterion online essay evaluation service 4:22
 English language learners (ELLs) 4:23
 general discussion 4:22
 Text Adaptor 4:23
 general discussion 4:25
 historical foundations 4:20
 modeling approaches 4:20–21
 reliability and validity 4:22
 paper-and-pencil tests 3:214, 3:215*f*, 3:216*f*, 3:503*r*
 estate schooling 4:814, 4:814*f*
 esthetics 4:783–789
 general discussion 4:783
 organizational artifacts 4:784
 organizational capital 4:787
 organizational contexts 4:783
 organizational culture 4:786
 organizational leadership 4:784
 organizational power 4:786
 research areas 4:788
 Estonia 5:555–562
 age-related policies and programs 8:263*r*
 background information 5:555
 early childhood education and care (ECEC) 2:111*r*
 education levels 2:279*f*
 education system
 achievement assessments 5:560
 administrative and supervisory structure 5:559
 challenges 5:561
 curriculum development 5:560
 educational finance 5:559
 examination systems 5:560
 goals 5:555
 monitoring and evaluation programs 5:561
 research programs 5:561
 structural and operational characteristics
 adult education and training (AET) 5:559
 formal structure 5:557*f*
 general discussion 5:556
 nonformal education 5:559
 preschool education 5:558
 primary schools 5:556
 secondary schools 5:556
 special needs students 5:558
 tertiary education 5:556
 teaching profession 5:559
 employment rate–educational attainment relationship 2:278*f*
 inequality–education relationship 2:278*f*
 information and communication technology (ICT)
 computer access in school 8:23*r*
 mathematics and science instruction 8:32*r*
 vocational education and training (VET) 8:263*r*
 e-textbooks 1:393
 Ethical–Cultural School 6:499
 ethical phenomenology 6:451
 ethical practices 4:717–723
 business/accounting education 3:41, 3:43
 educational research 6:33
 empirical analyses 4:720
 evaluation policies and practices 3:622–628
 challenge responses
 colleague-based feedback 3:627
 cultural contexts 3:627
 entry/contracting stage 3:626
 personal values 3:627
 professional guidelines and standards 3:626
 Guiding Principles
 competency 3:623
 confidentiality 3:623
 informed consent 3:623
 integrity and honesty 3:623
 respect 3:623
 responsibility 3:624
 systematic inquiry 3:622
 Program Evaluation Standards 3:624
 research areas
 data analysis and interpretation 3:625
 data collection 3:625
 design considerations 3:625
 entry/contracting stage 3:624
 findings utilization 3:626
 general discussion 3:624
 results reporting 3:625
 significant events 3:622
 future research areas 4:721
 gifted programs 7:734
 international scholarly networks 4:681, 4:721
 justice and care 6:164–169
 caring perspectives 6:166
 conflicting perspectives 6:167
 educational applications 6:168
 female perspective 6:166
 justice as fairness 6:165
 principles of justice 6:165
 theoretical perspectives 6:164
 traditional justification approaches 6:168
 military officer education 3:10
 pharmacy education 3:60
 poststructuralist theory 6:177
 societal context 4:717
 theoretical perspectives
 administrators and school leaders 4:719
 multiple perspective approach 4:719
 recent developments 4:718
 veterinary education 3:72–73

- Ethiopia
 early childhood education and care (ECEC) 2:111*t*
 school fees 2:419
 student achievement test performance 2:418*t*
- ethnicity
 achievement motivation 1:753–759
 ability beliefs
 intelligence theories 1:755
 racial stereotypes 1:754, 2:776
 achievement values/effort
 basic concepts 1:756
 oppositional identity 1:756, 2:776
 self-affirmation theories 1:757
 valued peers 1:757
 general discussion 1:754–759
 intelligence theories
 discrimination 1:756
 entity versus incremental theories 1:755
 racial stereotypes
 African Americans 1:754, 2:776
 Asian Americans 1:755, 2:776
 definitions 1:754
 research areas 1:758
 admissions tests 1:846, 4:12
 adult education 1:77–82
 background information 1:77
 historical perspective
 civil rights movement 1:77–78
 general discussion 1:77
 multiculturalism 1:79
 research background 1:79
 social justice movement 1:79
 regional perspectives 1:81
 Southern hemisphere 1:80
 Western societies 1:77
- China
 cultural perspectives 1:616
 educational policies 1:616–622
 case studies 1:617
 challenges 1:620
 ethnic pluralism 1:619
 historical perspectives 1:619
 Koreans 1:619
 Mongols 1:618
 Naxi 1:618–619
 preferential treatment 1:616
 sociocultural factors 1:617
 Tibetans 1:618
 Uyghurs 1:618
 cultural competence 2:565–570
 basic concepts 2:566
 community involvement 2:568
 importance 2:569
 overrepresentation 2:565
 Pedersen's model 2:566
 research background 2:565
 demographic data 1:687
 disadvantaged urban youth 1:806–813
 academic achievement outcomes 1:808
 community involvement 1:811
 educational preparedness 1:810
 emancipatory education 1:812
 knowledge as power 1:811
 male underachievement 1:808
 phenomenological variant of ecological systems theory
 (PVEST) 1:807*f*, 1:808–809, 1:809*f*
 research background 1:806
 resilience 1:810
 self-assessments 1:812
 self-efficacy beliefs 1:810
 sociocultural factors 1:806
 support systems 1:806
 vulnerability level/resiliency prediction
 model 1:806, 1:807*f*, 1:809*f*
 diversity discourse 1:823
 diversity research 1:693
 earnings differentials 2:288–297
 ability/educational quality differences 2:293
 Black–White test-score gap 2:294, 2:435
 Brazil 2:296
 change effects 2:295, 2:295*t*
 cross-sectional studies 2:291, 2:291*t*
 educational attainment 1:804
 education equalization effects 2:291*t*
 general discussion 2:288
 human capital model 2:290
 Israel 2:296
 minority–White income gap 2:292*f*
 minority women 2:292
 United States 2:289, 2:289*t*
- educational attainment 1:799–805
 access and equity patterns 5:52
 black males 1:800, 1:839
 completion rates 1:800, 1:803, 1:839
 dropout rates 1:803
 earnings differentials 1:804
 18–24-year-olds demographics 1:802, 1:803*f*
 enrollment rates 1:800, 1:801*f*, 1:802*f*, 1:803*f*
 female–male participation rates 1:799, 1:804
 gender gap research 1:799, 1:804
 United States 1:799
- educational statistics 1:581–588
 academic achievement outcomes 1:808
 choice effects 1:585
 continuing disparities 1:584
 demographic diversity 1:582
 digital technology implications 1:584
 disparity measures 1:581, 1:874
 dropout rates 1:874, 1:875
 education beyond borders 1:586
 18–24-year-olds demographics 1:802, 1:803*f*
 enrollment rates 1:800, 1:801*f*, 1:802*f*, 1:803*f*
 global village concept 1:587
 graduation rates 1:874
 incarceration rates 1:874
 twenty-first century perspectives 1:581, 1:587
 United States 1:582
- education–health association 1:668–669
- ethnic studies 1:623–628
 autonomy concerns 1:625
 general discussion 1:623, 1:627
 historical background 1:623
 institutional racism 1:626
 interdisciplinary collaboration 1:625
 scholarship 1:624
- gifted and talented students 2:872
- higher education
 admissions test fairness 4:12
 demographic data 1:687
 economic returns 1:542–546
 corporate environments 1:544
 earnings differentials 1:543
 economic well-being 1:545
 income gaps 1:543
 research background 1:542
 social mobility 1:542
 stratification 1:542
- educational attainment 1:799–805
 black males 1:800, 1:839
 completion rates 1:800, 1:803, 1:839
 dropout rates 1:803
 earnings differentials 1:804
 18–24-year-olds demographics 1:802, 1:803*f*
 enrollment rates 1:800, 1:801*f*, 1:802*f*, 1:803*f*
 female–male participation rates 1:799, 1:804
 gender gap research 1:799, 1:804
 United States 1:799
- enrollment opportunities 1:656, 1:658*f*
- student learning outcomes 4:479–484
 benefits 4:483
 cognitive complexity 4:479, 4:480*f*
 humanitarianism 4:480*f*, 4:481
 interpersonal and intrapersonal competence 4:480*f*, 4:482
 knowledge acquisition 4:480, 4:480*f*
 practical competence 4:480*f*, 4:482
 research background 4:479
- United Kingdom 1:673–678
 academic preparation 1:674
 campus experiences 1:676
 diversity considerations 1:673
 institutional dynamics 1:675
 parental involvement 1:676

- ethnicity (*continued*)
 peer influences 1:676
 poverty concerns 1:675
 primary and secondary education 1:674
 inequality–education relationship 2:274
 Japanese newcomers 1:787–792
 foreign nationals 1:788^f
 internationalization processes 1:790
 Japanese returnees 1:788
 Korean nationals 1:787
 mother tongue survey 1:790^f
 new foreigners 1:789, 1:790^f
 sociocultural portrayals 1:791
 legal education 1:691–697
 Educational Diversity Project (EDP) 1:694
 general discussion 1:696
 historical perspectives 1:691
 social science research 1:693
 mortality rates 1:606, 1:607^t
 race-specific consequences 1:876
 racial and ethnic concepts 6:36
 segregation policies 2:775
 United Kingdom
 higher education 1:673–678
 academic preparation 1:674
 campus experiences 1:676
 diversity considerations 1:673
 institutional dynamics 1:675
 parental involvement 1:676
 peer influences 1:676
 poverty concerns 1:675
 primary and secondary education 1:674
see also civic engagement
 ethnographic studies 6:386–391
 background information 6:386
 Brazil 1:762
 classroom ethnography 6:320–325
 background information 6:320
 basic concepts 6:321
 challenges 6:323
 illustrative examples 6:323
 computer technologies 6:390
 etymological concepts 6:386
 historical background 6:320
 interpretive research 6:419–423
 background information 6:419
 basic concepts 6:419
 big tent group 6:421
 chronotopes 6:422
 constructivism 6:421
 general discussion 6:423
 hybrid identities group 6:422
 interpretive turn 6:420
 positivist perspectives 6:422
 postmodernism 6:421–422
 qualitative inquiry 6:421
 signpost group 6:420
 language learning theories 5:349
 methodological perspectives 6:387
 phenomenological approaches 6:451
 qualitative inquiry 6:492
 research developments 6:389
 symbolic interactionism 6:482
 tensions and criticisms 6:388
 ethnomethodology 6:392–400
 conversational analysis (CA) 6:335–336, 6:394
 document analysis 6:381^t, 6:384
 formal analysis concept 6:393–394
 invisible colleges/college of education
 classroom instruction studies 6:397
 local study–situated action research 6:396
 research background 6:395
 research background
 Garfinkel's model 6:393
 general discussion 6:392
 invisible colleges/college of education 6:395
 mid-twentieth century developments 6:393
 sequential analysis 6:394
 significance 6:398
 Euclidian distance model 7:305, 7:305^f
 eugenics 1:558, 1:883, 7:431
 Europe
 displaced worker assistance programs 8:277
 earnings–education level relationship 2:307^t
 education levels 2:279^f
 enrollment rates
 gender disparities 2:416^t
 primary schools 2:415^t
 secondary schools 2:416^t
 higher education
 admissions tests 4:12
 Bologna Accord (1999)
 academic mobility 4:560
 background information 1:772
 binary systems 4:617–618
 common frameworks 1:774, 4:568
 future directions 1:777
 Germany 5:587
 globalization effects 1:775
 implementation challenges 1:776
 institutional autonomy 4:342
 internationalization processes 1:776, 4:544
 international programs 4:570
 quality assurance 4:569
 state steering models 4:557–558
 student achievement reporting 4:569
 study program levels 4:349
 faculty mobility 4:562
 historical development 1:728
 private higher education (PHE) 4:622
 student mobility 4:561
 inequality–education relationship 2:273^f
 Internet-based education 8:104
 labor education 1:141
 leadership development programs 4:753^t, 4:754^t, 4:755^t
 school choice 6:285
 student achievement test performance 2:247^f
 student experiences 4:473
 teacher education programs 7:528
 test score–social background relationship 2:273^f
 vocational education and training (VET) 8:514
 policy considerations
 European Qualification Framework (EQF) 8:413, 8:437, 8:479
 informal learning recognition 8:438
 official agreements 8:435
 vocational qualifications and status 8:437
 vocational–academic education parity 8:434
 European Association of Establishments for Veterinary
 Education (EAEVE) 3:72
 European Commission (EC) 3:716, 4:669
 European Credit System for Vocational Education and Training
 (ECVET) 8:414
 European Credit Transfer and Accumulation System (ECTS) 4:569, 8:348
 European Foundation for Management Development (EFMD) 3:42
 European Network for Accreditation of Engineering Education
 (ENAAE) 3:47
 European Qualification Framework (EQF) 8:413, 8:437, 8:479
 European Quality Improvement System (EQUIS) 3:42
 European Region Action Scheme for the Mobility of University Students
 (ERASMUS) 4:534, 4:540, 4:542, 4:559, 4:673
see also academic mobility
 European School Survey Project on Alcohol and Other Drugs (ESPAD) 5:665
 European Trade Union Confederation (ETUCE) 7:533
 European Union (EU) 4:669–675
 background information 4:669
 early childhood education and care (ECEC) 2:3
 educational cooperation
 contemporary programs 4:670
 historical background 4:669
 Lisbon Programme 4:671
 tensions and challenges 4:674
 European Commission (EC) 3:716, 4:669
 higher education 4:557–558
 information and communication technology (ICT) access and equity 8:97
 learning cities and regions 1:147
 Lifelong Learning Programme
 background information 4:670
 Comenius programme 4:673
 ERASMUS 4:673
 general discussion 4:673
 Grundtvig programme 4:673
 Jean Monnet programme 4:674

- Leonardo da Vinci programme 4:673
 - transversal programme 4:673
 - postconflict societies 4:514
 - research-quality assessments 4:307
 - Eurydice 4:670
 - evaluation policies and practices
 - accountability 3:600–606
 - appropriateness 3:601
 - definitions 3:600
 - design considerations
 - context-sensitive outcomes 3:603
 - data analysis and interpretation 3:604
 - data sources 3:603
 - general discussion 3:602
 - multiple measures 3:603
 - quality assurance 3:603
 - reliability 3:604
 - reporting considerations 3:605
 - responsibility assignments 3:604
 - target audience 3:605
 - general discussion 3:605
 - stakeholder involvement 3:602
 - validity and utility framework
 - general discussion 3:600
 - purpose and scope 3:601
 - culturally responsive evaluations 3:592–599
 - critical pedagogy perspective 3:595
 - cultural psychology theory 3:594
 - ecological systems theory 3:593
 - intervention considerations 3:593
 - Kamehameha Early Education Program (KEEP) 3:596
 - methodologies
 - objectivist approaches 3:595
 - transformative participatory approaches 3:596
 - participatory action research (PAR) 3:597
 - Photovoice technique 3:597–598
 - research areas 3:599
 - subtractive schooling 3:594
 - Youth Empowerment Strategies (YES!) Program 3:597
 - education system
 - Bulgaria
 - general discussion 5:547
 - higher education 5:547
 - public education 5:547
 - Lithuania 5:664
 - e-learning 3:530–538
 - challenges 3:536
 - clarification processes
 - context 3:531, 3:531*t*, 3:532*t*, 3:535
 - criteria 3:531*t*, 3:533, 3:533*t*, 3:535
 - data collection and analysis 3:534, 3:536
 - evaluands 3:531*t*, 3:532, 3:533*t*, 3:535
 - methodology selection 3:533, 3:536
 - questions 3:533, 3:535
 - stakeholders 3:531*t*, 3:532, 3:535
 - completion processes
 - comparison studies 3:534, 3:536
 - metaevaluation 3:534, 3:536
 - recommendations 3:534, 3:536
 - reporting 3:534, 3:536
 - flexible frameworks 3:530
 - illustrative example
 - comparison studies 3:536
 - completion processes 3:536
 - context clarification 3:535
 - criteria clarification 3:535
 - data collection and analysis 3:536
 - evaluand clarification 3:535
 - general discussion 3:534
 - metaevaluation 3:536
 - methodology selection 3:536
 - question clarification 3:535
 - recommendations 3:536
 - reporting 3:536
 - stakeholder clarification 3:535
 - key components 3:531*f*
 - opportunities 3:536
 - ethical practices 3:622–628
 - challenge responses
 - colleague-based feedback 3:627
 - cultural contexts 3:627
 - entry/contracting stage 3:626
 - personal values 3:627
 - professional guidelines and standards 3:626
 - Guiding Principles
 - competency 3:623
 - confidentiality 3:623
 - informed consent 3:623
 - integrity and honesty 3:623
 - respect 3:623
 - responsibility 3:624
 - systematic inquiry 3:622
 - Program Evaluation Standards 3:624
 - research areas
 - data analysis and interpretation 3:625
 - data collection 3:625
 - design considerations 3:625
 - entry/contracting stage 3:624
 - findings utilization 3:626
 - general discussion 3:624
 - results reporting 3:625
 - significant events 3:622
 - evaluation fundamentals 3:698
 - evaluation reporting and communicating 3:746–756
 - communication strategies
 - e-mail and postcards 3:750
 - newsletters and brochures 3:750
 - online collaborative tools 3:750, 3:750*r*
 - creative formats
 - audience engagement 3:750, 3:751*r*
 - cartoons 3:751, 3:751*f*
 - poetry 3:751, 3:752*f*
 - theatrical presentations 3:752, 3:752*t*
 - cross-cultural considerations 3:753
 - definitions 3:746
 - design considerations
 - audiences 3:753*t*, 3:754
 - communicating and reporting plan 3:754*f*, 3:755*f*, 3:756
 - strategies/formats and resources 3:755
 - timing 3:755
 - final evaluation report 3:747, 3:748*t*
 - interactive formats 3:749
 - visual appeal 3:747
 - interactive formats
 - general discussion 3:749
 - web conferences 3:749
 - working sessions 3:749, 3:749*r*
 - purposes 3:747
 - stakeholders 3:753
 - visual appeal
 - design, graphics, and layout 3:748
 - general discussion 3:747
 - photography 3:748
 - videos 3:749
- evaluation use/utilization 3:614–621
 - diversity considerations 3:619
 - educational applications 3:618
 - ethical practices 3:626
 - evidence-based interventions 3:615
 - external evaluation utilization strategies 3:784
 - internal evaluation utilization strategies 3:782
 - professional organizations 3:615
 - research background
 - centrality of use 3:616
 - historical perspectives 3:615
 - influence 3:617
 - use terminology criticisms 3:617
 - standards 3:614–615, 3:615*t*
 - theoretical/philosophical perspectives 3:614
 - use-oriented effects 3:620
- historical perspectives 3:645–652
 - Keynesian welfare state model 3:645, 3:646*t*
 - knowledge capitalist model 3:646*t*, 3:649
 - neoliberal era 3:646*t*, 3:647
 - purpose and scope 3:645
- internal–external evaluation 3:781–785
 - coexistence considerations 3:784
 - definitions 3:781
 - educational applications 3:781
 - general discussion 3:784
 - mutual benefits
 - external evaluation utilization strategies 3:784

- evaluation policies and practices (*continued*)
 - general discussion 3:782
 - internal evaluation utilization strategies 3:782
 - interpretation improvements 3:783
 - scope expansion 3:782, 3:783
 - validity legitimization 3:783
 - teacher evaluations 3:675
- personnel administration 4:842
- portfolios 7:512, 7:619
- principal evaluation 3:698–705
 - definitions
 - stakeholder impacts 3:700
 - working definition 3:701
 - evaluated objects
 - basic concepts 3:701
 - information collection criteria 3:703
 - leadership mega-skills 3:702
 - evaluation fundamentals 3:698
 - published results
 - accountability-based evaluation (2003) 3:699
 - general discussion 3:700
 - role-based evaluation (1996) 3:698
 - research guidelines 3:704
- prospective teachers 7:511–517
 - cognitions and decision-making processes
 - background information 7:512
 - peer assessments 7:512
 - portfolios 7:512
 - self-assessments 7:512
 - simulation methods 7:512
 - general discussion 7:516
 - good teaching views 7:511
 - instructional behaviors 7:511
 - learning-oriented assessments 7:512, 7:513*t*
 - purpose 7:511
 - quality assessments 7:514
 - standards-based assessments 7:513
 - validity measures
 - characteristics 7:514
 - consequential validity 7:515
 - constructs and assessment task criteria 7:514
 - extended interpretations 7:514
 - reliability 7:515
- research methodologies 3:735–745
 - design considerations
 - mixed-methods designs 3:742
 - qualitative methodologies 3:741
 - quantitative methodologies 3:739
 - stakeholders 3:743
 - value interdependence 3:743
 - evidence-based education (EBE) 3:744
 - general discussion 3:736
 - globalization effects 3:744
 - inquiry characteristics
 - evaluands 3:736, 3:737*t*
 - question alignment 3:737*t*, 3:738
 - unit of analysis 3:737*t*, 3:738
 - large-scale assessments 3:744
 - purpose 3:736
 - quantitative methodologies
 - characteristics 3:739
 - observational (nonexperimental) designs 3:741
 - quasi-experimental designs 3:740
 - randomized field experiments 3:739
- science-education programs 3:660–666
 - accountability considerations 3:660
 - background characteristics 3:664
 - contextual evaluations 3:661
 - data analysis and reporting 3:664
 - formative evaluation tools 3:663*t*
 - outcomes evaluations 3:663
 - process model 3:660, 3:661*f*
 - process skills 3:662*t*
 - program implementation 3:662, 3:663*t*
 - variable measures 3:664
- stakeholders 3:629–635
 - basic concepts 3:629
 - empirical research results 3:631
 - evaluation activities 3:629
 - evaluation utilization impacts 3:633
 - importance 3:633
 - Program Evaluation Standards*
 - participation guidelines 3:629
 - theoretical perspectives 3:630
 - theoretical perspectives
 - broad involvement 3:631
 - deep involvement 3:631
 - organizational learning perspective 3:630
 - Program Evaluation Standards* 3:630
 - social justice perspective 3:630
- teacher evaluations 3:674–680
 - challenges 3:678
 - evaluation forms 3:675
 - future research areas 3:679
 - policy frameworks 3:676
 - professional autonomy 3:677
 - professional standards 3:677
 - purposes 3:674
 - research background 3:674
 - typologies 3:678
- technology 3:578–584
 - evaluation method selection process 3:582, 3:583*f*
 - importance 3:578
 - key issues
 - implementation considerations 3:580
 - measurement characteristics and control 3:581
 - treatment control 3:580
 - methodologies
 - combined methods 3:580
 - qualitative methods 3:580
 - quasi-experiments 3:579
 - random assignment approximations 3:579
- evaluation research 7:150–155
 - case studies 6:458
 - challenges 7:155
 - definitions 7:150, 7:151*f*
 - general discussion 7:10
 - methodologies
 - mixed-methods research 7:152
 - qualitative methods 7:152
 - quantitative methods 7:151
 - monitoring procedures
 - general discussion 7:154
 - meta-evaluation 7:154
 - stakeholder involvement
 - design considerations 7:154
 - ethical conduct 7:154
 - standards 7:154
 - professionalization 7:153
 - program evaluation 3:706–712
 - basic concepts 3:707, 7:151, 7:151*f*
 - comparison studies 3:707
 - evaluation frameworks
 - basic concepts 3:707
 - credible evidence 3:710
 - data collection strategies 3:710
 - evaluation focus 3:710
 - lessons learned 3:711
 - logic modeling 3:709, 3:709*f*
 - program descriptions 3:709
 - results analysis and interpretation 3:710
 - stakeholder involvement 3:522, 3:708
 - usage facilitation 3:708*f*, 3:711
 - general discussion 3:711
 - impact evaluations 3:706
 - outcome evaluations 3:706
 - process evaluations 3:706
 - purpose 3:706
 - summative evaluations 3:706
 - research background 7:152
 - evaluative feedback 3:411
 - event-related potentials (ERPs) 5:247
 - Evergreen School 6:256
 - everyday memories 5:200
 - examination systems
 - Bulgaria 5:541
 - Caribbean countries 1:680
 - Lithuania 5:660
 - Malta 5:677
 - reform efforts 3:172
 - excellence in cities (EiC) programs 5:93
 - exceptional learners *see* gifted and talented students

- excessive help seeking 6:653
exchange students 4:542, 4:559
executive help seeking 6:653
executive professional leadership (EPL) model 5:18
existential phenomenology 6:82, 6:111, 6:451
expectation–maximization (EM) 7:242
expedient help seeking 6:653
experiential bias 4:172
experiential learning
 e-learning pedagogies 4:423*t*
 engaged education 1:713, 1:714*f*
 nursing education 3:65
 prior learning assessment (PLA) 1:172
experimental design
 basic concepts 7:6
 general discussion 7:125
 quality indicators 7:125
 quasi-experimental designs
 general discussion 7:128
 regression discontinuity (RD) designs 7:128
 time series designs 7:128
 randomized experimental designs
 blocked designs 7:128
 control group designs 7:127
 factorial designs 7:127
 general discussion 7:127
 single-case research designs 7:128
 statistical analysis techniques 7:129, 7:129*f*, 7:130*f*
explanatory talk 5:464–468
 children's conceptions 5:465
 definitions 5:464
 diachronic explanations 5:467, 5:467*f*
 discourse genres 5:464
 extended discourse 5:465
 general discussion 5:468
 narrative thinking 5:464
 social interaction theory 5:466
 synchronic explanations 5:467, 5:467*f*
 theoretical perspectives 5:464
explicit teacher knowledge 7:750*r*
exploratory data analysis (EDA) 7:156–161
 basic concepts 7:156
 computational statistics 7:96
 computer software 7:160
 educational applications 7:157
 exam marks example 7:158*f*, 7:159, 7:160*f*, 7:161*f*
 functional role 7:156
 future directions 7:161
 general discussion 7:156
 graphical techniques 7:157, 7:158*f*
 methodologies 7:157
 research publications 7:159
 response plots 7:3, 7:3*f*
exploratory factor analysis 3:77
 characteristics 7:164*t*
 conceptual/theoretical considerations 7:162
 design considerations 7:163
 general discussion 7:162
 reporting considerations 7:163
 statistical considerations 7:163
exponential distribution 7:100
exponential family of distributions 7:100
exponential model 7:464
extended area communities 4:814
extended discourse 5:352, 5:465
extended-term mixed-method (ETMM) research 3:742
external benefits 2:260–271
 aggregate externalities studies 2:263*t*, 2:269
 characteristics 2:260
 health influences 2:253, 2:258
 implications 2:270
 indirect benefits 2:261*f*, 2:262
 market benefits 2:260, 2:261*f*, 2:348
 nonmarket benefits 2:260, 2:261*f*
 quantification studies
 crime rate reductions 2:263*t*, 2:267
 democratization 2:263*t*, 2:265
 direct social benefits 2:263*t*, 2:264
 environmental benefits 2:263*t*, 2:268
 estimation methods 2:262, 2:263*t*
 happiness measures 2:263*t*, 2:268
 human rights 2:263*t*, 2:266
 indirect benefits 2:263*t*, 2:268
 inequality reductions 2:263*t*, 2:267
 life expectancy 2:263*t*, 2:267
 political stability 2:263*t*, 2:266
 poverty reductions 2:263*t*, 2:267
 prison cost reductions 2:263*t*, 2:267
 public welfare cost reductions 2:263*t*, 2:267
 social capital 2:263*t*, 2:268
 tax receipt increases 2:263*t*, 2:267
 technology dissemination 2:263*t*, 2:268
 social benefits 2:261, 2:261*f*
 total value 2:263*t*, 2:269
externally-designed partnerships 7:760*r*
external modularization 8:386
external validity 6:521–527
 causal effects measurements 2:187
 future outlook 6:525
 generalizability theory
 basic concepts 6:522
 main questions 6:522
 pragmatic justification principles
 general discussion 6:524
 principle of causal explanation 6:525
 principle of discriminant validity 6:525
 principle of empirical interpolation and extrapolation 6:525
 principle of heterogeneous irrelevancies 6:524
 principle of proximal similarity 6:524
 research background 6:521
 traditional justification approaches
 analog models 6:523
 causal inference 6:522
 meta-analysis 6:524
 purposive sampling 6:523
 statistical sampling theory 6:522
 technology transfer models 6:523
extraversion 3:258
extreme value theory 7:25–31
 educational applications 7:25, 7:26*f*
 extensions 7:31
 modeling approach
 basic model
 extreme value index (EVI) 7:25
 peaks-over-threshold–generalized Pareto distribution (POT–GPD)
 assumption 7:26, 7:27*f*, 7:28*f*
 Pareto-type distributions 7:28
 quantile plotting 7:29, 7:29*f*, 7:30*f*
 tail estimation 7:26, 7:27*f*, 7:28*f*
 regression analysis 7:31, 7:31*f*
 research background 7:31
extrinsic motivation 6:618–623
 basic concepts 6:619, 6:619*f*
 behavioral consequences 6:621
 classroom environments 6:626
 creativity 4:170
 external regulation 6:619
 general discussion 6:618, 6:622
 identified regulation 6:620
 integrated regulation 6:620
 internalization processes 6:620
 introjected regulation 6:620
 reward systems 6:621
 second-language learning 5:420
 socioenvironmental contexts 6:620
- ## F
- Facebook 8:190, 8:202
face validity 2:578, 3:158
Fachdidaktik 7:656
Fachhochschulen 4:348–349, 5:585*f*, 5:587
Fachoberschule 5:585*f*
facial expressions 6:544
factor analysis 7:162–170
 categorical data 7:165
 confirmatory factor analysis (CFA)
 characteristics 7:164*t*
 general discussion 7:164
 linear structural equation models 7:453
 multilevel confirmatory factor analysis (MCFA)

- factor analysis (*continued*)
 - basic concepts 7:165
 - estimation process 7:166
 - illustrative application 7:162
- post-analysis decisions
 - model fit statistics 7:165
 - model misspecification statistics 7:165
- pre-analysis decisions
 - covariance matrix 7:164
 - model identification 7:164
 - multivariate normality 7:164
 - parameter estimation theory 7:165
 - rival models 7:164
- educational applications 7:4
- exploratory factor analysis
 - characteristics 7:164*r*
 - conceptual/theoretical considerations 7:162
 - design considerations 7:163
 - general discussion 7:162
 - reporting considerations 7:163
 - statistical considerations 7:163
- general discussion 7:15, 7:169
- principal component analysis (PCA) comparison 7:377
- factorial designs 7:127
- faculty development *see* education development
- faculty mobility *see* academic mobility
- faculty values **4:433–440**
 - background information 4:433
 - core cultural components 4:434
 - external pressures
 - accountability 4:437
 - economic imperatives 4:436
 - emerging skill sets 4:436
 - learning societies 4:437
 - massification and diversification 4:436
 - political impacts 4:435
 - technological advances 4:436
 - general discussion 4:438
 - new initiatives
 - funding considerations 4:438
 - learning centers 4:437–438
 - policies and programs 4:438
 - quality assurance 4:438
 - teaching impacts
 - deep learning versus surface learning 4:435
 - general discussion 4:434
 - learning paradigm shift 4:435
 - research-led teaching and learning 4:434
- failed states 4:514
- fairness **4:56–60**
 - admissions tests 4:12
 - assessment impacts
 - bias 4:57
 - equity 4:57
 - general discussion 4:56
 - intelligence quotients (IQs) 4:56–57
 - legal challenges 4:57
 - general discussion 4:56, 4:60
 - justice as fairness 6:165
 - process approaches
 - equity
 - fairness principles 4:59*r*
 - important questions 4:58, 4:58*r*
 - test development 4:58, 4:59*r*
 - informal evaluations/assessments 4:59
 - large-scale assessments 4:59
- fair practice standards 3:380
- fallibilism 6:87
- false autobiographical memories 5:201
- false detection rate (FDR) measure 7:313
- false leadership 5:10
- familiarity 5:342
- familywise error rate (FWE) measures 7:312
- Fanon, Frantz 6:351
- F* distribution 7:100
- fear of failure (FOF) theory 6:538
- Federal Bureau of Investigation (FBI) 2:242
- Federal Institute for Vocational Education and Training (BIBB) 8:355
- Fédération Européenne d'Associations Nationales d'Ingénieurs (FEANI) 3:47
- feedback **3:226–230**
 - adapted feedback 3:226
 - after-action review (AAR) feedback 3:228
 - characteristics 3:226, 3:227*r*
 - classroom assessment 3:411, 3:413–414
 - cognitive load theory 3:227
 - computer-based systems 3:228, 8:208
 - delayed feedback 3:226–227
 - descriptive feedback 3:411, 3:413–414
 - diagnostic assessments 3:397
 - dual-coding theory 3:228
 - dynamic assessments 3:221, 3:229
 - educational data modeling 8:13
 - elaboration feedback 3:226
 - evaluative feedback 3:411
 - formative assessments 3:207, 3:229
 - gender research 5:434
 - general discussion 3:226
 - immediate feedback 3:226–227
 - learning regulation impacts 3:350
 - peer assessments 3:231
 - performance feedback 3:226–227, 3:227*r*, 4:747
 - primary school science 3:296
 - self-regulated learning 5:507
 - terminology 8:270
 - timing 3:226–227
 - verification feedback 3:226
 - visual and verbal representations 3:227
 - writing assessments 3:275
- feminism
 - diversity discourse 1:824
 - early childhood education and care (ECEC) 2:11
 - Eurocentric educational approaches 1:303
 - feminist theories **6:56–62**
 - critical theory perspective 1:495
 - current perspectives 6:60
 - education feminism 6:58
 - feminist philosophies
 - analytic feminism 6:58
 - continental feminism 6:58
 - exemplars 6:58
 - general discussion 6:58
 - pragmatist feminism 6:58
 - wave feminist theory 6:57
 - Western perspectives 6:56
 - historical background 1:25
 - leadership research 4:798
 - lifelong learning 1:91
 - philosophy of education 6:112
 - poststructuralist theory 1:502, 4:799, 6:178
 - rational autonomy 6:123
 - women's studies **1:623–628**
 - autonomy concerns 1:625
 - general discussion 1:623, 1:627
 - historical background 1:623
 - institutional racism 1:626
 - interdisciplinary collaboration 1:625
 - scholarship 1:624
 - see also* pragmatism
- fetal alcohol syndrome 2:426
- field experiences
 - basic concepts 1:714
 - curriculum development and logistics 1:716, 1:717*r*
 - goals and objectives 1:715
 - mother tongue teacher education 7:741
 - participant observation 6:446
 - qualitative inquiry 6:494
 - supervision 7:619
 - teacher education 7:522
- field experimentation **3:83–89**
 - benefits 3:89
 - design considerations 3:83
 - importance 1:716
 - inference estimations 3:84
 - random assignment approximations
 - basic concepts 3:84
 - benefits 3:89
 - implementation considerations
 - blocking tactics 3:88
 - cluster sampling 3:88
 - general discussion 3:87
 - intervention fidelity 3:88
 - intraclass cluster correlation (ICC) 3:88

- unit congruence 3:88
- internal validity threats 3:85, 3:86*t*
- planning guidelines 3:87
- Fife Mathematics Project (Scotland) 1:457
- Fifteenth Amendment (U.S. Constitution) 1:691
- Fiji 5:563–568
 - background information 5:563
 - demographic data 5:563
 - early childhood education and care (ECEC) 2:111*t*
 - economy 5:564
 - education system
 - administrative and supervisory structure 5:566
 - challenges 5:567
 - educational finance 5:566
 - future directions 5:567
 - goals 5:564
 - historical background 5:563
 - monitoring and evaluation programs 5:567
 - research programs 5:567
 - structural and operational characteristics
 - adult education 5:566
 - early childhood education and care (ECEC) 5:564
 - general discussion 5:564
 - higher education 5:566
 - nonformal education 5:566
 - primary education 5:564
 - secondary education 5:565
 - special needs students 5:565
 - vocational education and training (VET) 5:566
 - teaching profession 5:566
 - international programs 5:567
 - language 5:563
 - political background 5:563
- financial aid in higher education 2:229
- financial incentives 2:344
- financial mathematics 7:452
- financial returns *see* returns to education
- financial risk management (FRM) 3:43
- fine arts education
 - assessment strategies 3:262–267
 - authentic assessments 3:262–263
 - challenges 3:262
 - classroom-based assessments
 - criteria 3:264
 - dance 3:266
 - drama/theater 3:266
 - music 3:265
 - visual arts 3:265
 - general discussion 3:266
 - performance assessments 3:262–263
 - standardized assessments
 - general discussion 3:263
 - National Assessment of Educational Progress (NAEP) 3:264
 - standards movement 3:263
 - subject matter development 1:424
- Finland 5:569–575
 - adult basic education (ABE) 1:103
 - age-related policies and programs 8:263*t*
 - background information 5:569
 - childhood brain development 2:57*f*
 - early childhood education and care (ECEC)
 - enrollment rates 2:4*t*, 2:56*t*, 2:111*t*, 2:114*t*
 - funding estimates 2:58*t*
 - mentoring models
 - curriculum frameworks 2:165
 - general discussion 2:164
 - group-based e-mentoring 2:166
 - earnings–education level relationship 2:307*t*
 - economic growth 8:84
 - education levels 2:279*f*
 - education system
 - administrative and supervisory structure 5:573
 - challenges 5:574
 - educational finance 5:573
 - goals 5:569
 - information and communication technology (ICT) 8:84
 - monitoring and evaluation programs 5:574
 - research programs 5:574, 6:292
 - structural and operational characteristics
 - adult education and training (AET) 5:572
 - apprenticeships 5:570*f*, 5:572
 - basic education 5:570*f*, 5:571
 - characteristics 5:570
 - early childhood education and care (ECEC) 5:570
 - educational institutions 5:571*t*
 - polytechnical institutions 5:570*f*, 5:572
 - preschool education 5:570, 5:570*f*
 - schematic diagram 5:570*f*
 - universities 5:570*f*, 5:572
 - upper secondary schools 5:570*f*, 5:571
 - vocational education and training (VET) 5:570*f*, 5:572
 - teaching profession 5:572
 - gender-equity gap 2:56*t*
 - higher education
 - access and equity patterns 1:689, 4:336–337
 - binary systems 4:617–618
 - foreign doctoral students 4:292*t*
 - institutional expenditures 1:733*t*
 - loans and grants 4:501*f*, 4:502*f*
 - student and researcher mobility 4:287*f*
 - tertiary education expenditures 4:500*f*
 - tuition fees 4:500*f*
 - inequality–education relationship
 - earnings differentials 2:277*t*
 - employment rate–educational attainment relationship 2:278*f*
 - fathers' educational background 2:274*f*
 - test score–social background relationship 2:273*f*
 - information and communication technology (ICT)
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - mathematics and science instruction 8:32*t*
 - student proficiency and confidence 8:31*t*
 - inquiry-oriented teacher education 7:606
 - lifelong learning model 1:157*t*, 1:158
 - mother tongue teacher education 7:739
 - participation research
 - regular education 8:284*f*
 - vocational education and training (VET) 8:283*f*, 8:284*f*
 - work-related vocational education and training (VET) 8:284*f*, 8:285*f*
 - prison education programs 8:301, 8:301*t*
 - research-quality assessments 4:309
 - returns to education–education level 2:308*t*
 - student achievement test performance 2:247*f*
 - sustainable development efforts 6:292
 - teacher age distribution 2:492*f*
 - teacher education programs 7:526–527
 - teacher shortages 2:494*f*
 - teacher wages and salaries 2:496*f*
 - teaching hours 2:493*f*
 - vocational education and training (VET)
 - age-related policies and programs 8:263*t*
 - curriculum development 8:515
 - initial vocational preparation system 8:251*t*
 - outcomes analyses 8:252*f*, 8:253*f*
 - participation research 8:283*f*, 8:284*f*, 8:285*f*
 - school-based programs 8:247
 - vocational–academic education parity 8:435
 - women in employment 2:56*f*
- first-generation computer-based tests 8:48, 8:49*t*
- First International Mathematics Study (FIMS) 3:687*t*, 4:650, 4:664
- First International Science Study (FISS) 4:651
- first language learning 1:435–440
 - acquisition processes 5:375–381
 - cognitive research 5:247
 - cross-linguistic variations 5:379
 - developmental disorders 5:380
 - developmental milestones
 - early word production and comprehension 5:377, 5:378*f*
 - general discussion 5:376
 - infant babble 5:377
 - morpheme acquisition order 5:378*t*
 - pre-linguistic achievements 5:376
 - sentence formation 5:378
 - skill refinements 5:379
 - Wug test 5:378*f*
 - developmental research 5:376
 - individual variations 5:380
 - play 5:333
 - theoretical perspectives 5:380
 - curricula
 - general discussion 1:435
 - primary schools

- first language learning (*continued*)
 language curricula 1:436
 literature teaching 1:437
 reading curricula 1:436
 writing instruction 1:437
 secondary schools
 adolescent literacies 1:438
 digital technology 1:439
 language curricula 1:438
 literature teaching 1:438
 general discussion 1:435, 1:439
 human language characteristics 5:375
 influential factors 1:435
- First Nations Technical Institute (Canada) 8:290*r*
- Fisher discrimination
 general discussion 7:136
K-population example 7:137
 two-population example 7:137
- Fisher-scoring algorithm 7:368–369
- Fisher's *F*-test 7:337
- Fisher's *Z* 7:276
- five-factor model (FFM) of personality
 coping behaviors 3:257
 extraversion 3:258
 general discussion 3:256
 neuroticism 3:256
 test anxiety 3:257
- fixed-effects model
 average effect-size calculations 7:277, 7:280*f*; 7:280*r*
 hedonic wage model 2:469
 peer relationships 2:435
 random assignment approximations 3:101
 test-based accountability 4:155
- flexible learning **4:441–447**
 background information 4:441
 basic concepts 4:423
 blended learning 4:445
 criticisms 4:445
 definitions 4:441
 delivery challenges 4:444
 delivery effectiveness 4:443
 democratization of learning 4:444
 flexible delivery 4:442
 general discussion 4:441, 4:446
 information and communication technology (ICT) 4:442
 philosophical perspectives 4:444
 responsiveness and availability 4:445
 workplace learning
 general discussion 8:379
 learning preferences 8:379
 readiness and responsiveness 8:379
- Florida 1:654*f*; 6:297–298
- Florida Price Level Index (FPLI) 2:466
- flow theory **6:605–611**
 basic concepts 6:605
 flow experience
 activity characteristics 6:605
 educational applications
 classroom paradox 6:607
 Key School 6:608
 general discussion 6:605
 model diagram 6:606*f*
 optimal state 6:605
 future research areas 6:609
- fluid intelligence 4:2, 5:185
- Flynn effect 4:3
- focal tests 4:186
- focus groups **6:401–405**
 characteristics 6:401
 data collection and analysis 6:405
 functional role 6:402
 research design
 general discussion 6:402
 interview question guidelines 6:403
 moderating approaches 6:404
 participant selection 6:403
 usefulness 6:405
- folkblidning* 1:163
- folkeop(p)lysning* 1:163
- Follow Through Classroom Observation Evaluation (FTCOE) study 6:304
- Forbes top-ranked companies study 7:386, 7:386*t*; 7:387*f*; 7:387*t*; 7:388*f*
- foreign language learning **1:428–434, 1:441–446, 5:382–386**
 background information 1:441
 basic concepts
 oral competency 1:441
 proficiency orientation 1:441
 challenges 1:445
 characteristics 5:422
 Common European Framework (CEF) 1:442
 communicative competence 5:382, 5:383*t*
 communicative language teaching (CLT) model
 characteristics 1:429
 criticisms 1:429
 curricular advantages 1:432, 1:432*f*
 educational conflicts 1:431
 key assumption and mission 1:430
 persuasive discourses 1:433
 student resistance 1:432
 content-based instruction 1:443
 English as a foreign language (EFL) 1:428
 communicative language teaching (CLT) model
 characteristics 1:429
 criticisms 1:429
 curricular advantages 1:432, 1:432*f*
 educational conflicts 1:431
 key assumption and mission 1:430
 persuasive discourses 1:433
 student resistance 1:432
 future directions 1:433
 knowledge production and circulation 1:428
 notational/functional linguistics 1:429
 structural linguistics 1:428
 English-language instruction 5:382
 foreign-language ability model 5:382
 future directions 1:433
 general discussion 1:445
 immersion programs 1:443
 knowledge production and circulation 1:428
 language resource centers 1:444
 learning process theories 5:384
 notational/functional linguistics 1:429
 research challenges 5:382
 socialization aspects 5:384
 standards 1:442
 structural linguistics 1:428
 teacher preparation 1:444
see also bilingualism
- Foresight Report (2007) 3:73–74
- forgery 2:239
- formal knowledge 7:671
- formal learning
 ageless learners 8:241*f*
 basic concepts 1:3–4
 definitions 5:457, 8:240
 elder learning 1:176
 formal–informal learning distinction 1:43
 idealized characteristics 5:462*t*
- Lithuania
 special needs students 5:661
 vocational education and training (VET) 5:660
- private-sector training 1:207
- Spain 5:810
- terminology and jargon sources 8:350*t*
- workplace learning 1:45, 8:329, 8:329*f*
- formal reasoning 5:248
- formative assessments **3:359–364**
 adult education and training (AET) 3:550
 alternative assessments
 diagnostic assessments 3:207, 3:276, 3:397
 dynamic assessments 3:207
 educational assessments 3:205
 feedback 3:207
 purpose 3:206
 amelioration 3:793*f*; 3:794
 basic concepts 3:360, 3:378
 challenges 3:363
 classroom assessment 3:211, 3:229, 3:413, 3:451*t*; 3:464, 3:510
 cross-cutting issues 3:165
 current perspectives 3:359
 early childhood education and care (ECEC) 2:89
 educational impacts **4:69–74**
 functional role 4:73

- general discussion 4:69
 high-visibility external tests
 influencing factors 4:72
 potential benefits 4:71
 problematic effects 4:71
 historical perspectives 4:69
 measurement purposes 4:69
 research areas 4:73
 feedback 3:229, 3:563
 formative assessment–summative assessment continuity 3:351
 French-language-based classroom assessments 3:434
 functional role 4:46
 general discussion 3:312, 7:151
 higher education 3:249
 implementation considerations 3:353–358
 case-study approach 3:353
 general discussion 3:358
 Neartown High School case study
 assessment for learning principles 3:356
 background information 3:355
 roadblocks 3:357
 summative assessments 3:355
 Ourtown Elementary School case study
 assessment development guidelines 3:354
 background information 3:353
 classroom routine changes 3:355
 school self-review 3:354
 summative assessments 3:353
 test results analysis 3:353
 triangulated evidence 3:354
 principle versus practice 3:353
 school change–system change relationship 3:358
 instructional planning 3:365–368
 current perspectives 3:365
 Dewey's perspective 3:365
 dialog 3:367
 general discussion 3:368
 learning regulation 3:366
 practical guidelines 3:365
 zone of proximal development (ZPD) 3:367
 King's Medway Oxford Formative Assessment Project (KMOFAP) 3:360
 learning regulation impacts 3:349, 3:366
 legal education 3:29
 mathematics education 2:771, 3:282, 3:653
 oral language skills 3:288
 portfolios 3:245
 practical applications
 further learning areas 3:362
 oral and written feedback 3:361
 peer- and self-assessment 3:362
 primary school science
 evidence collection 3:294
 evidence interpretation 3:295, 3:295*t*
 feedback 3:296
 goals 3:294
 peer assessments 3:296
 policies and practices 3:294
 student self-assessments 3:296
 records of achievement 3:245
 reflective practices 3:332
 research background 3:359
 research-quality assessments 4:312
 school-based prevention programs 2:826
 secondary school science 3:300
 self-assessments 3:232
 Task Group on Assessment and Testing (TGAT) report 3:359
 teacher education 3:563–571
 assessment instruments 3:563
 assessment tools
 peer assessments 3:566
 portfolios 3:565
 self-assessments 3:567
 deployment stages 3:563, 3:564*f*
 design guidelines 3:564, 3:565*t*
 guiding principles 3:564*t*
 informative feedback 3:563
 intervention strategies 3:564
 purpose 3:563
 research-based principles 3:567, 3:568*t*
 teacher evaluations 3:675
 technology education 3:313
 technology supports 8:42–47
 basic concepts 8:42
 educational implications 8:45
 functional roles 8:43*t*
 general discussion 8:42
 large-scale commercial-based systems 8:44, 8:45*t*
 small-scale research-based systems 8:45, 8:45*t*
 vocational education and training (VET) 3:339
 workplace-based assessments 3:332
 writing instruction 3:274
 formative curriculum evaluation model 1:337
 forward-and-backward-translation test designs 4:166
 forward-only-translation test designs 4:166
 Foucault, Michel
 critical theory philosophy 6:176, 6:351
 governance 7:532
 neoliberal ideologies 6:14
 philosophical perspectives 6:3, 6:172
 postmodernism 1:91
 poststructuralist theory 1:497, 1:500–501, 6:111, 6:175, 6:478
 radical hermeneutics 6:65
 reason and rationality 6:185–186
 Fourteenth Amendment (U.S. Constitution) 1:691
 fractional Brownian motion 7:452
 France 5:576–581
 accreditation processes 4:384–389
 age-related policies and programs 8:263*t*, 8:265–266
 background information 5:576
 childhood brain development 2:57*f*
 classroom assessment 3:432–437
 background information 3:432
 competency-based programs 3:433
 effective communication strategies 3:434
 external testing 3:435
 formative assessments 3:434
 general discussion 3:436
 multilingual sources 3:432
 policies and practices 3:425–431
 assessment beliefs and practices 3:430
 current practices 3:426
 France–England national assessment comparison studies 3:427, 3:427*t*, 3:429*t*
 general discussion 3:425
 QUEST (quality of education systems transnationally) project 3:428
 reform efforts 3:426*t*
 STEP (systems, teachers, and education policy) studies 3:427
 report cards 3:434
 school reform 3:432
 standards 3:435
 translated meanings 3:432
 class size 2:492*f*
 displaced worker assistance programs 8:278
 early childhood education and care (ECEC) 2:4*t*, 2:56*t*, 2:58*t*, 2:111*t*, 2:114*t*, 2:154
 earnings–education level relationship 2:307*t*
 education levels 2:279*f*
 education system
 achievement assessments 5:580
 administrative and supervisory structure 5:579
 challenges 5:580
 economic factors 2:347
 educational finance 5:580
 goals and objectives 5:577
 secondary education
 demographic data 5:578
 first cycle 5:578
 lycées d'enseignement général/nlet technologique (LEGT) 5:578
 lycées professionnels (LP) 5:578
 second cycle 5:578
 vocational education and training (VET) 5:578
 structural and operational characteristics
 apprenticeships 5:578
 continuous adult education 5:578
 higher education 5:578
 primary education 5:577
 secondary education 5:578
 teaching staff 5:578
 higher education
 collective bargaining 4:493
 faculty mobility 4:562
 historical development 1:728
 institutional expenditures 1:733*t*

- France (*continued*)
- loans and grants 4:501*f*
 - parental expenditures 1:734*t*
 - structural and operational characteristics 5:578
 - student and researcher mobility 4:287*f*
 - student experiences 4:473
 - student mobility 4:561
 - tertiary education expenditures 4:500*f*
 - tuition fees 4:500*f*
 - university associations and subgroups 4:486
- inequality–education relationship
- earnings differentials 2:277*t*
 - employment rate–educational attainment relationship 2:278*f*
 - fathers' educational background 2:274*f*
 - gender differences 2:277*t*
 - immigrant status 2:275*f*
 - test score–social background relationship 2:273*f*
- information and communication technology (ICT) 8:32*t*
- leadership development programs 4:753*t*, 4:754*t*, 4:755*t*, 4:758*t*
- lifelong learning model 1:157*t*, 1:158
- participation research
- regular education 8:284*f*
 - vocational education and training (VET) 8:283*f*, 8:284*f*
 - work-related vocational education and training (VET) 8:284*f*, 8:285*f*
- pharmacy education 3:59
- special education background 2:714
- student achievement test performance 2:247*f*, 2:417*t*
- teacher age distribution 2:492*f*
- teacher wages and salaries 2:496*f*
- teaching hours 2:493*f*
- textbook publishing 1:370
- vocational education and training (VET)
- age-related policies and programs 8:263*t*, 8:265–266
 - employer involvement 8:457
 - initial vocational preparation system 8:251*t*
 - lifelong learning model 1:157*t*, 1:158
 - participation research 8:283*f*, 8:284*f*, 8:285*f*
 - school-based programs 8:247
 - vocational–academic education parity 8:435
- franchises 4:511
- Frankfurt School 6:65, 6:133, 6:176, 6:349, 6:358
- Frank, Lawrence 1:515
- Fraport AG (Germany) 1:157*t*, 1:159
- F* ratio
- calculation 7:33
 - least significant difference (LSD) procedure 7:313
 - significance interpretation 7:34
- fraud 2:239
- free-choice learning 1:153
- Frege, Gottlob 6:108
- Freire, Paulo
- critical theory perspective 1:55, 1:495, 6:137, 6:351–352
 - educational theory 1:31–35
 - background information 1:31, 1:814
 - learning and instructional strategies
 - action–reflection processes 1:34
 - dynamic knowledge 1:34
 - learners' existential realities 1:34
 - pedagogy of empowerment 1:817
 - planning processes 1:34
 - praxis
 - authority versus authoritarianism 1:33
 - codification 1:31
 - collective dimensions of learning 1:34
 - community knowledge 1:31
 - dialectical relations 1:32
 - directive approach 1:33
 - empty theorizing 1:32
 - exile as praxis 1:31
 - learner as subject 1:32
 - learner–educator inequality 1:32
 - tact and prudence 1:33
 - teacher student versus student teachers 1:32
- image-elicitation techniques 6:500
- nonformal education 1:42
- peace education 6:23
- popular adult education 1:164
- French-language-based classroom assessments 3:432–437
- background information 3:432
 - competency-based programs 3:433
 - effective communication strategies 3:434
 - external testing 3:435
 - formative assessments 3:434
 - general discussion 3:436
 - multilingual sources 3:432
 - report cards 3:434
 - school reform 3:432
 - standards 3:435
 - translated meanings 3:432
- frequency viewpoint 7:479
- Freud, Sigmund 6:92, 6:133, 6:598
- Friedman, Milton
- contemporary liberalism 6:77
 - state's role in education 2:331, 2:350
- Friedman–Rafsky test 7:337
- friendship 6:564–569
- characteristics 6:564, 6:587
 - childhood schooling 6:693
 - collaborative learning 6:686
 - help seeking behaviors 6:656
 - intervention strategies 6:568
 - peer acceptance 6:565
 - school adjustment effects 6:565
- social networks 6:705–711
- academic performance 6:709
 - adolescents 6:707
 - characteristics 6:705
 - computer-mediated communication 8:116
 - deviant behaviors 6:709
 - research approaches 6:706
 - young children 6:706
- social tasks
- conflict management skills 6:567
 - formation processes 6:566
 - friendship transgressions 6:567
 - general discussion 6:566
 - help-giving skills 6:567
 - reliable alliances 6:567–568
 - self-disclosure 6:568
 - sociometric measures 6:564
- Froebel, Friedrich W. 1:514
- Fromm, Erich 6:349
- Fullan, Michael 5:128, 5:132, 6:289, 6:290–291
- full-service schools *see* community-focused schools
- functional behavior assessment (FBA) 2:708–713
- assessment questions 2:709*t*
 - assessment strategies 2:710
 - basic concepts 2:708
 - behavior intervention plans (BIPs) 2:708–709
 - data analysis 2:710, 2:711*t*
 - data collection 2:710, 2:710*t*, 2:711*t*
 - hypothesis generation and testing 2:709
 - individualized education plans (IEPs) 2:708–709
 - intervention strategies 2:711
 - problem behaviors 2:709
 - referral strategies 2:682
 - replacement behaviors 2:709
- functional data analysis (FDA) 7:417–425
- applications
 - El Niño time series 7:420*f*, 7:423, 7:424*f*
 - phoneme data 7:423
 - spectrometric data 7:418*f*, 7:421*f*, 7:422, 7:423*f*, 7:424*f*
 - characteristics 7:417
 - general discussion 7:10, 7:424
 - illustrative examples
 - basic concepts 7:417
 - discretized curves 7:418*f*
 - log-periodograms 7:419*f*
 - monthly sea surface temperature 7:420*f*
 - spectrometric curves 7:418*f*
 - year-by-year temperatures 7:420*f*
 - modeling approaches 7:420
 - multivariate data analysis 7:419, 7:421*f*
- Functional Information Technology Test (FITT) 4:645
- functional literacy
- critical literacy 1:107
 - humanist approach 1:107
 - skill-based approach 1:107
- functional magnetic resonance imaging (fMRI) 2:171, 5:247, 5:311, 5:317, 5:340
- see also* neuroscience research
- Function Probe 8:175

further education and training (FET)
 Mongolia 5:682*t*, 5:683
 South Africa 3:451–452, 3:451*t*, 3:452*t*, 3:453, 5:801, 8:444
 Furth, Hans 1:36–37
 fuzzy coding 7:108

G

Gadamer, H.-G. 6:63, 6:65, 6:185, 6:416, 6:450–451

Gage, Phineas 6:546–547

Gallaudet, Thomas 2:714

Galton, Francis 2:870, 5:187, 7:431, 7:471

Galtung, Johan 6:24

Gambia 5:858–864

background information

demographic characteristics 5:858

economic characteristics 5:858

general discussion 5:858

physical and climatic characteristics 5:858

sociocultural characteristics 5:858

early childhood education and care (ECEC) 2:111*t*

education system

challenges 5:863

educational finance

general discussion 5:862

government expenditures 5:862

recurrent expenditures 5:862

goals 5:859

monitoring and evaluation programs 5:863

sector management 5:861

structural and operational characteristics

adult education 5:860

basic education 5:859

early childhood education and care (ECEC) 5:859

general discussion 5:859

lower-basic education 5:860

Madrassa education 5:860

nonformal education 5:860

secondary education 5:860

tertiary education 5:861

upper-basic education 5:860

vocational education and training (VET) 5:861

teacher education 5:861

student achievement test performance 2:418*t*

game systems 8:112–113

gamma distribution 7:100

Gandhi, Mahatma 1:307

Gardner's Multiple Intelligences (MI) model 2:831

Garfinkel, Harold 6:393

Gates–MacGinitie Reading Tests 3:189

Gates Millennium Scholars program 1:847–848

Gauss, Carl Friedrich 7:467

Gaussian distribution *see* normal distributions

Gauss' inequality 7:427

Gauss–Laplace distribution *see* normal distributions

Gauss–Newton method 7:341, 7:343

Gautreaux residential mobility program 2:432

gazzinta process 6:362, 6:365*t*

Geertz, Clifford 6:419

gender research

achievement motivation 1:757

Africana Studies 1:568

child-rearing practices 2:151

curriculum 1:277–282

background information 1:277

feminist perspectives 1:277

gendered access 1:279

general discussion 1:281

indigenous curricula 1:309

male–female continuum 1:280

queer studies 1:281

twentieth-century perspective 1:277

diversity discourse 1:823

early childhood education and care (ECEC) 2:11–16

care regimes 2:11

gender equity

adult worker model 2:14

care-related social rights 2:14

domestic work and care patterns 2:14

general discussion 2:13

international care perspectives 2:15

paid parental leave 2:12, 2:14

public policy responses 2:15

male-breadwinner model 2:12

policies and programs 2:12

provision arrangements

Australia 2:13

balance of care providers 2:12

Denmark 2:13

Germany 2:13

Iceland 2:13

women's liberation movement 2:11

early writing development 5:371

educational statistics 1:581–588

choice effects 1:585

continuing disparities 1:584

demographic diversity 1:582

digital technology implications 1:584

disparity measures 1:581

education beyond borders 1:586

global village concept 1:587

twenty-first century perspectives 1:581, 1:587

United States 1:582

gender analysis 1:70–76

background information 1:70

benefits 1:74

classroom discourse 5:365

doctoral student attrition 1:827

earnings differentials

ability/educational quality differences 2:293

Black–White test-score gap 2:294, 2:435

Brazil 2:296

change effects 2:295, 2:295*t*

cross-sectional studies 2:291, 2:291*t*

educational attainment 1:804

education equalization effects 2:291*t*

human capital model 2:290

Israel 2:296

minority women 2:292

private returns to education 2:306–307, 2:309*f*

United States 2:289, 2:289*t*

educational leadership 4:797–802

diversity challenges 4:800

feminist perspectives 4:798

general discussion 4:797

mainstream research 4:797

engineering profession 3:48

future challenges 1:75

general discussion 1:75

growth-curve models 7:206*t*, 7:207*f*, 7:208*t*

inequality–education relationship 2:275, 2:276*f*, 2:277*t*

life expectancy 1:602, 1:603*f*, 1:603*t*

mortality rates 1:606

narrative socialization 5:305

publication productivity

gender differences 4:459

motherhood 4:459

networking and collaboration 4:459

research approaches

data collection and analysis 1:73

general discussion 1:72

strategic initiatives 1:74

social relationships 2:795

Sri Lanka 5:814

theoretical perspectives

adult education agenda 1:71

diverse research approach 1:70

general discussion 1:70

hybrid research approach 1:71

oppositional research approach 1:70

sexual minorities 1:72

higher education 1:629–635

academic identities 1:630

access and equity patterns 4:336, 4:336*t*

challenges 1:629

demographic data 1:687

educational attainment 1:799–805

black males 1:800, 1:839

completion rates 1:800, 1:803, 1:839

dropout rates 1:803

earnings differentials 1:804

- gender research (*continued*)
 18–24-year-olds demographics 1:802, 1:803*f*
 enrollment rates 1:800, 1:801*f*, 1:802*f*, 1:803*f*
 ethnic minorities 1:799, 1:804
 female–male participation rates 1:799, 1:804
 United States 1:799
 female participation 1:630
 future directions 1:633
 global gender parity index (GPI) 1:630
 government policies and programs 1:633
 humanitarianism 4:481
 identity politics 1:86, 6:39, 6:152
 learner group assessment strategies 3:486–491
 future directions 3:490
 general discussion 3:486, 3:490
 learning–assessment relationship
 general discussion 3:486
 theoretical perspectives 3:486
 sex-group differences
 difference explanations 3:488
 fixed–notion perspectives 3:487
 international/regional assessment programs 3:487
 research limitations 3:488
 sociocultural perspectives
 assessment techniques 3:489
 performance assessments 3:489
 research background 3:488
 research limitations 3:489
 learning experiences 5:432–437
 emerging issues 5:436
 gender-based interactions 5:436
 gender-equity reforms 5:435
 general discussion 5:432
 research focus
 boys and learning 5:434–435
 feedback 5:434
 hidden agendas 5:434
 research flaws 5:433
 women's voice 5:434
 sex-differences research 5:432
 single sex–coeducation comparison studies 5:435
 lifelong learning 1:91
 male underachievement 1:838–844
 future research areas 1:842
 general discussion 1:838
 influencing factors
 classroom behavior 1:841
 cool and dumb behaviors 1:841
 curriculum and class structure 1:842
 family composition 1:840
 hegemonic masculinity 1:840
 identity development 1:840
 racialized masculinity 1:840
 student engagement 1:841
 teachers' perceptions 1:841
 research focus 1:839
 research perspectives 1:839
 moral and social reasoning 5:282–283
 mortality rates 1:606
 racial and ethnic demographics 1:687
 school resources–earnings relationship 2:315
 traumatic brain injury (TBI) 2:695
 working mothers 2:33, 2:55, 2:151
 General Agreement on Trade in Services (GATS) 1:537, 1:734–735, 4:598, 6:532, 7:535
 General Certificate of Secondary Education (GCSE) 3:386
 general component latent trait model (GLTM) 7:87
 general equivalency diploma (GED)
 Catholic schools 2:357
 prison education programs 2:243
 signaling model 2:323
 general factor (*g*) 4:2, 5:185
 generalizability theory 4:61–68
 advantages/disadvantages 4:67
 background information 4:61
 conceptual framework 4:61
 external validity
 basic concepts 6:522
 main questions 6:522
 International System for Teacher Observation and Feedback (ISTOF) project 3:728
 multivariate generalizability theory 4:66, 4:66*r*
 reliability 4:27
 standardized tests 2:577
 univariate generalizability theory
 D studies/infinite universe of generalization
 basic concepts 4:63
 coefficients 4:65
 D study variance components 4:63, 4:63*r*
 error variances 4:64, 4:64*f*
 universe scores 4:63
 D studies/restricted universes 4:65
 general discussion 4:61
 G studies 4:62
 universe of admissible observations 4:62
 generalized linear models (GLMs) 7:171–177, 7:178–183
 background information 7:178
 basic concepts 7:178
 categorical data analysis 7:61
 conditional and marginal relationships 7:175, 7:175*f*
 deviance measures 7:179*r*
 educational applications 7:8, 7:212
 estimation methods 7:175
 general discussion 7:164, 7:182
 generalized linear mixed models (GLMMs) 7:8
 illustrative example 7:180, 7:180*r*, 7:181*r*, 7:182*r*
 inference estimations 7:179
 latent class (LC) models 7:240
 logistic random-intercept model 7:171, 7:172*f*, 7:172*r*
 missing data handling techniques 7:285
 multilevel modeling 7:176, 7:212
 random coefficient models 7:172
 random-effects model 7:175
 research areas 7:344
 residuals 7:180
 response types
 binary responses 7:173
 continuous responses 7:173
 counts 7:174
 general discussion 7:173
 latent response formulations 7:174, 7:174*f*
 ordinal responses 7:174
 software packages 7:175
 variance-accounted-for measures 7:445
 General National Vocational Qualifications (GNVQs) 8:384, 8:386
 General Teaching Council (GTC) 3:539
 Generation R tests 8:49*r*, 8:52
 Generation X 7:585
 generic evaluations 3:676
 generic skills 8:368
 geometric distribution 7:133
 geometric mean 7:265
 Geometric Supposer 8:173
 Georgia
 early childhood education and care (ECEC) 2:111*r*
 information and communication technology (ICT) 8:25*r*
 tuition and needs-based grants 1:654*f*
 Germany 5:582–589
 accreditation processes 4:384–389
 African-descendant populations 1:573
 age-related policies and programs 8:263*r*
 background information 5:582
 childhood brain development 2:57*f*
 class size 2:492*f*
 comparative education societies 1:269
 displaced worker assistance programs 8:278
 early childhood education and care (ECEC) 2:4*r*, 2:13, 2:56*r*, 2:111*r*, 2:114*r*
 earnings–education level relationship 2:307*r*
 education levels 2:279*f*
 education system
 educational finance 5:588
 future directions
 general discussion 5:587
 higher education 5:588
 preschool education 5:587, 5:588*r*
 primary education 5:587, 5:588*r*
 quality assurance 5:588
 secondary education 5:587, 5:588*r*
 national level institutions 5:582
 structural and operational characteristics
 higher education 5:585*f*, 5:587, 5:589*r*
 instructional language 5:582

- preprimary education 5:585
- primary education 5:585, 5:585*f*
- schematic diagram 5:585*f*
- secondary education 5:585*f*, 5:586, 5:586*t*
- enrollment rates 2:493*f*
- gender-equity gap 2:56*t*
- higher education
 - binary systems 4:617
 - characteristics 5:587
 - collective bargaining 4:493
 - faculty mobility 4:562
 - federal political systems 4:605
 - future directions 5:588
 - historical development 1:730
 - institutional expenditures 1:733*t*
 - loans and grants 4:501*f*, 4:502*f*
 - number of institutions 5:589*t*
 - private higher education (PHE) 4:622
 - student and researcher mobility 4:287*f*
 - student mobility 4:561
 - tertiary education expenditures 4:500*f*
 - tertiary type-B program survival rates 8:464*t*
 - university associations 4:485
- immigrant populations 5:582
- inequality–education relationship
 - earnings differentials 2:277*t*
 - employment rate–educational attainment relationship 2:278*f*
 - fathers' educational background 2:274*f*
 - gender differences 2:276*f*, 2:277*t*
 - immigrant status 2:275*f*
 - test score–social background relationship 2:273*f*
- information and communication technology (ICT)
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - student proficiency and confidence 8:31*t*
- information and communication technology (ICT) access and equity 8:97
- Länder* governments 4:605, 5:582
- leadership development programs 4:753*t*, 4:754*t*, 4:755*t*, 4:758*t*
- lifelong learning model
 - market-led model 1:157*t*, 1:159
 - social partnership model 1:157*t*, 1:159
- participation research
 - regular education 8:284*f*
 - vocational education and training (VET) 8:283*f*, 8:284*f*
 - work-related vocational education and training (VET) 8:284*f*, 8:285*f*
- research-quality assessments 4:309
- teacher age distribution 2:492*f*
- teacher education programs 7:528
- teacher wages and salaries 2:496*f*
- teaching hours 2:493*f*
- vocational education and training (VET)
 - age-related policies and programs 8:263*t*
 - dual apprenticeship system 8:448–454
 - background information 8:448
 - characteristics 8:232, 8:449, 8:449*t*
 - dual system–vocational training relationship 5:585, 5:585*f*, 5:586–587, 8:435, 8:451
 - general discussion 8:453
 - historical background 8:450
 - status considerations 8:435
 - terminology and jargon sources 8:347
 - working principles 8:449
 - employer involvement 8:457
 - Federal Institute for Vocational Education and Training (BIBB) 8:355
 - initial vocational preparation system 8:251*t*
 - modularization 8:383, 8:386
 - participation research 8:283*f*, 8:284*f*, 8:285*f*
 - teacher training 8:507*t*
 - vocational–academic education parity 8:435
- women in employment 2:56*f*
- Gesamtschulen* 5:585*f*, 5:586
- Gesell, Arnold
 - child development theories 6:499
 - early childhood curriculum 1:515
- gestures 5:249
- Getty Institute 1:424, 3:263
- Ghana 5:590–598
 - background information 5:590
 - early childhood education and care (ECEC) 2:111*t*
 - education system
 - administrative and supervisory structure 5:594
 - challenges 3:454, 5:597
 - classroom assessment 3:450, 3:451*t*, 3:454
 - curriculum structure 3:450*t*
 - educational finance 5:595, 5:596*t*
 - exit points/continuation/exit skills 3:452*t*
 - goals 5:591
 - indigenous curricula 1:309
 - monitoring and evaluation programs 5:597
 - research programs 5:597
 - structural and operational characteristics
 - adult education and training (AET) 5:594
 - early childhood education and care (ECEC) 5:592, 5:592*t*
 - enrollment 5:592*t*
 - higher education 5:592*t*, 5:594
 - junior secondary schools 5:592*t*, 5:593
 - nonformal education 5:592*t*, 5:594
 - preschool education 5:592, 5:592*t*
 - primary schools 5:592, 5:592*t*
 - senior secondary schools 5:592*t*, 5:593
 - teaching profession 5:592*t*, 5:594
 - vocational education and training (VET) 5:590, 5:592*t*, 5:593
- higher education gender equity 1:632
- information and communication technology (ICT) 8:25*t*
- physically disabled students 2:520
- returns to education–cognitive skills 2:310*t*
- school choice 6:285–286
- school fees 2:419
- student achievement test performance 2:247*f*, 2:418*t*
- Gibbs–Markov equivalence theorem 7:199
- Gibbs sampling 7:246, 7:247*f*
- gifted and talented students 2:870–875
 - alternative educational settings 2:873
 - characteristics 2:830, 2:872
 - definitions 2:871
 - educational placement alternatives 2:873
 - ethnic minorities 2:872
 - etiology 2:872
 - general discussion 2:874
 - historical perspectives 2:870
 - identification strategies 2:872
- instructional strategies 2:829–833
 - curricular models
 - general discussion 2:831
 - Multiple Intelligences (MI) model 2:831
 - Schlichter Model for Talents Unlimited Inc. (TU) 2:831
 - Schoolwide Enrichment Triad Model (SEM) 2:831
 - Stanley model 2:831
 - Sternberg's Triarchic Componential Model (STTCM) 2:831
 - VanTassel–Baska Integrated Curriculum Model 2:832
 - differentiated instruction 2:832
 - general discussion 2:832
 - intelligence theories 2:830
- Mongolia 5:683
- prevalence 2:872
- research background 2:829, 2:870
- special needs students 2:588–593
 - characteristics
 - attention deficit/hyperactivity disorders (ADHD) 2:590
 - emotional and behavioral disorders (EBD) 2:590
 - general discussion 2:589, 2:592
 - learning disabilities 2:590
 - educational programs 2:591
 - identification strategies 2:588
 - occurrences 2:588
- teacher education programs 7:732–737
 - challenges 7:735
 - desirable characteristics 7:735, 7:735*t*
 - diversity considerations 7:736
 - general discussion 7:736
 - philosophical perspectives 7:732
 - research methodologies and findings 7:734
 - standards-based programs
 - assessment strategies 7:733
 - collaboration 7:734
 - communication skills 7:733
 - ethical practices 7:734
 - foundations 7:733
 - general discussion 7:733
 - individual learning differences 7:733
 - instructional planning 7:733
 - instructional strategies 7:733

- gifted and talented students (*continued*)
 learning capabilities 7:733
 learning environments 7:733
 social interactions 7:733
 teacher training 7:732
 Gini index of diversity 7:384, 7:384f
 Giroux, Henry 6:133
 Glass Ceiling Index 1:689
 Glass's Δ 7:445
 Global Campaign for Education (GCE)
 EFA Class of 2015 6:533
 general discussion 6:533
 Global Action Week 6:533
 global advocacy 6:534
 year-round campaigns 6:533
 global citizenship education (GCE) 1:116, 7:722
 global civil societies 1:195
 Global Forum on International Quality Assurance, Accreditation and the
 Recognition of Qualifications (UNESCO) 4:599
 global information societies 4:363
 Global Initiative for Quality Assurance Capacity (GIQAC) 4:600
 globalization
 academic mobility 4:209
 African-descendant populations 1:573
 arts curricula 1:423
 business/accounting education 3:43
 comparative education 1:264
 cultural globalization 6:529
 curriculum 1:312–317, 1:403–410
 background information 1:403
 challenges 1:316
 curriculum design 1:404
 epistemological perspectives
 altered conditions/state of affairs theory 1:314
 discourse theory 1:314
 process theory 1:315
 project theory 1:315
 relationship theories 1:313
 general discussion 1:312, 1:409
 knowledge-based economies 1:312–313, 1:403
 lifelong learning model 1:404, 1:405t
 reading curricula 1:421
 reform efforts
 assessment methods 1:407
 classroom dynamics 1:408
 course content 1:406
 critical perspectives 1:408
 institutional change 1:406
 instructional strategies 1:407
 research and development expenditure levels 1:404t
 traditional learning model 1:405t
 curriculum discourse analysis 1:344
 dance education 1:423
 drama/theater education 1:423
 educational evaluation 3:523
 educational globalization 6:529–535
 African-descendant populations 1:573
 BRAC (Building Resources across Countries) 6:533
 commodification 6:531
 cultural globalization 6:529
 educational paradigm change 6:531
 Europe 1:775
 Global Campaign for Education (GCE)
 EFA Class of 2015 6:533
 general discussion 6:533
 Global Action Week 6:533
 global advocacy 6:534
 year-round campaigns 6:533
 higher education 1:734
 neoliberal perspectives 6:531
 nongovernment organizations (NGOs) 6:532
 recent developments 6:534
 United Nations involvement 6:529
 educational leadership
 demographic globalization 4:708
 economic globalization 4:707
 technological globalization 4:707
 educational paradigm change 6:531
 engineering profession 4:209
 evaluation research methodologies 3:744
 for-profit sector 1:705
 graduate employment 4:358
 higher education 1:536–541, 4:201–208
 Cambodia 4:205
 China 4:205
 contemporary globalization 4:202
 contextual perspectives 4:597
 educational impacts 4:206, 4:601
 enrollment growth 4:597
 global English 4:203
 globalization-from-above
 general discussion 1:537
 globalization-from-above versus globalization-from-below 1:536
 homogenizing effects 1:538
 intensified hierarchies 1:537
 privatization 1:537
 globalization-from-below
 collaborative relationships 1:538
 general discussion 1:538
 globalization-from-above versus globalization-from-below 1:536
 massification 1:539
 socially responsible practices 1:539
 historical background 4:201
 India 4:205
 medium of instruction (MOI) 4:202, 4:205
 neoliberal perspectives 1:536
 private higher education (PHE) 4:597
 research background 1:536
 research results and analysis 1:540
 institutional management 4:265
 labor market impacts 8:411
 lifelong learning 1:404, 1:405t, 1:699, 6:170–171
 mathematics education 3:658
 music learning 1:423
 national assessment programs 3:696
 neoliberal perspectives 6:531
 philosophy of education 6:8
 student experiences 4:477
 teacher education 7:534
 United Nations involvement 6:529
 universities 4:201–208
 Cambodia 4:205
 China 4:205
 contemporary globalization 4:202
 educational impacts 4:206
 global English 4:203
 historical background 4:201
 India 4:205
 medium of instruction (MOI) 4:202, 4:205
 vocational education and training (VET) 8:411–415
 characteristics 8:411
 economic restructuring 8:411
 educational impacts 8:414
 higher-level skills development 8:414
 international education markets 8:413
 knowledge-based economies 8:412
 mass movement/skills migration 8:412
 planning and policy development 8:470
 terminology and jargon sources 8:346, 8:346t
 transferable skills 8:413
 workforce issues 8:412
 Global Leadership and Organizational Behavior Effectiveness (GLOBE) project
 8:201
 global nomads 4:637
 GLOBE study 4:827
 Going Global initiative *see* Early Learning and Development Standards (ELDS)
 good faith bargaining 2:513
 goodness-of-fit (GOF) models 7:190–196
 characteristics 7:190
 continuous observed data
 GOF indices 7:193
 numerical examples 7:194, 7:195t
 overall fit assessment statistics 7:192
 residuals/piecewise assessments of fit 7:193
 structural equation models (SEMs) 7:192
 discrete observed data
 basic concepts 7:190
 GOF indices 7:191
 limited information methods
 characteristics 7:191
 marginal residuals/piecewise assessments of fit 7:192
 overall fit assessment statistics 7:192

- misfit/piecewise assessments of fit 7:191
- numerical examples 7:194, 7:194_t
- overall fit assessment statistics 7:190
- general discussion 7:193
- research areas 7:195
- Good Start, Grow Smart program 2:132–133
- Gore, Albert 1:748
- governance
 - accreditation 7:495
 - characteristics 7:532
 - teacher education 7:532–539
 - background information 7:532
 - general discussion 7:537
 - modeling approaches
 - basic concepts 7:532, 7:533_f
 - Bologna Accord (1999) 7:536
 - decentralized centralism 7:534
 - globalization effects 7:534
 - management by objectives (MBO) 7:534
 - neoliberal ideologies and strategies 7:534
 - new institutionalism 7:535
 - policy tools 7:535
- government-sponsored training 1:206
- grade of membership (GoM) model 7:63
- grade point average (GPA) 3:250, 4:10, 7:148
- grading systems
 - automated essay scoring (AES) 4:20–26
 - available programs 4:21, 8:51
 - basic concepts 4:20
 - educational applications
 - automated speech-recognition scoring programs 4:24
 - Criterion online essay evaluation service 4:22
 - English language learners (ELLs) 4:23
 - general discussion 4:22
 - Text Adaptor 4:23
 - general discussion 4:25
 - historical foundations 4:20
 - modeling approaches 4:20–21
 - reliability and validity 4:22
 - higher education 3:249
 - vocational teacher education 7:747
- graduate employment 4:354–359
 - differentiation considerations 4:356
 - division of labor 4:357
 - employability enhancements 4:356
 - employment rates and earnings 4:354
 - expansion effects 4:354, 4:365
 - globalization 4:358
 - job quality 4:355
 - knowledge society theory 4:358
 - labor market 4:366
 - research biases 4:359
 - student demands 4:354
 - transition periods 4:357
 - worker productivity 4:355
- Graduate Management Admission Test (GMAT) 4:9, 8:49
- Graduate Record Examinations (GRE) 4:8, 7:148, 8:49
- Gramscian perspective 1:194, 6:352, 6:355
- grant aid programs in higher education 2:229
- graphical models 7:197–202
 - analysis of deviance 7:200
 - Bayesian networks 7:199
 - chain graph models 7:199–200
 - general discussion 7:10, 7:201
 - Gibbs–Markov equivalence theorem 7:199
 - graph theory 7:197
 - joint probability distributions 7:198
 - Markov properties 7:199
 - math grade example
 - covariance matrix 7:198_t
 - general discussion 7:197
 - partial correlation matrix 7:198_t
 - test scores 7:198_f
 - model fitting and selection 7:200
 - path diagrams 7:200
 - Simpson's paradox 7:199, 7:199_f
- graphic organizers 2:769, 2:769_f
- graphing calculators 8:4, 8:136, 8:172, 8:173_f
- Grassroots Alliance for Community Education (GRACE) 1:123
- Greece 5:599–604
 - age-related policies and programs 8:263_t
 - background information 5:599
 - class size 2:492_f
 - early childhood education and care (ECEC) 2:4_t, 2:111_t
 - education levels 2:279_f
 - education system
 - administrative and supervisory structure 5:602
 - educational finance 5:602
 - goals 5:599
 - performance assessments 5:603
 - reform efforts 5:603
 - structural and operational characteristics
 - adult education 5:601
 - general discussion 5:600
 - postsecondary nontertiary education/vocational training 5:601
 - preschool education 5:600
 - primary education 5:600
 - secondary education 5:600
 - tertiary education 5:601
 - teacher training 5:602
- enrollment rates 2:493_f
- higher education
 - institutional expenditures 1:733_t
 - loans and grants 4:501_f
 - parental expenditures 1:734_t
 - student and researcher mobility 4:287_f
 - tertiary education expenditures 4:500_f
 - tertiary type-B program survival rates 8:464_t
- inequality–education relationship
 - employment rate–educational attainment relationship 2:278_f
 - test score–social background relationship 2:273_f
- information and communication technology (ICT)
 - computer access in school 8:23_t
 - frequency of computer use in school and at home 8:26_t
 - student proficiency and confidence 8:31_t
- information and communication technology (ICT) access and equity 8:97
- participation research
 - regular education 8:284_f
 - vocational education and training (VET) 8:283_f, 8:284_f
 - work-related vocational education and training (VET) 8:284_f, 8:285_f
- school supply shortages 2:506_f
- student achievement test performance 2:247_f
- teacher wages and salaries 2:496_f
- teaching hours 2:493_f
- vocational education and training (VET) 8:251_t, 8:252_f, 8:253_f, 8:263_t
- Greene, Maxine 6:59
- Grenada
 - early childhood education and care (ECEC) 2:111_t
 - higher education enrollment 1:702_f, 1:703_f
- grounded theory 6:406–412
 - abductive research methods 3:80
 - coding methods 6:409–410
 - constructivist grounded theory–positivist grounded theory comparisons 6:408
 - definitions 6:406
 - future directions 6:411
 - historical foundations 6:406
 - memo-writing 6:410
 - qualitative research 6:406
 - research strategies 6:409
 - theoretical sampling 6:410–411
- grouped jackknife 7:230–231
- group-level norms 4:97
- Group Reading Test 3:189
- group work 3:342–347
 - assessment strategies
 - assessment criteria 3:345
 - assessment focus 3:343
 - group contributions 3:344
 - guiding principles 3:346
 - individual contributions 3:343
 - background information 3:342
 - class size 6:204
 - disadvantages 3:342
 - peer learning 5:445
 - pre-referral intervention strategies 2:817
- growth-curve models 7:203–209
 - basic concepts
 - higher-level models 7:207
 - linear growth 7:205
 - multilevel/hierarchical linear modeling 7:204, 7:205_f
 - multivariate growth-curve models 7:207, 7:208_t
 - nonlinear growth

- growth-curve models (*continued*)
 gender analysis 7:206*t*, 7:207*f*
 general discussion 7:206
 individual growth curves 7:206*f*
 random effects 7:207*t*
 standard errors 7:206*t*, 7:207*t*
 birth cohort study example 7:204, 7:204*t*, 7:205*t*
 modeling approaches 7:208
 research background 7:203
- Growth, Employment and Redistribution (GEAR) program 8:443
- growth models 7:8
Grundschule 5:585*f*, 5:586
 Grundtvig programme 4:673
Gutter v Bollinger (2003) 1:691, 1:692, 1:823
- Guatemala
 assessment system development 3:482
 early childhood education and care (ECEC) 2:111*t*
 earnings–education level relationship 2:306*t*
 higher education enrollment 1:702*f*, 1:703*f*
- guided participation 5:501
- Guinea 2:111*t*
- Guinée 2:418*t*
- Guinée Bissau 2:418*t*
- Guttman errors 4:16
- Guyana
 early childhood education and care (ECEC) 2:111*t*
 higher education enrollment 1:702*f*, 1:703*f*
- Gymnasium* 5:585*f*, 5:586
- H**
- Habermasian perspective 1:193
- Habermas, Jurgen 6:65, 6:133, 6:171, 6:176, 6:178, 6:350, 6:416–417
- habit 6:88
- Hájek–Šidak central limit theorem 7:234
- Hall, G. Stanley 1:515
- Hamel, Gary 3:43
- Hamlyn, David 6:110–111
- handicapped students *see* special needs students
- hand injuries 8:113
- Handy, Charles 3:43
- Hanushek meta-analyses study 2:444
- happiness and well-being
 capabilities approach 6:194–199
 educational contributions
 general discussion 6:196
 justice 6:197
 pedagogy and curriculum design 6:198
 value and functions 6:197
 future research areas 6:198
 normative framework
 Nussbaum's perspective 6:195
 Sen's perspective 6:194
 philosophical perspectives 6:194
- children 2:32–37
 environmental factors
 climate change 2:34
 environmental toxins 2:34
 food consumption changes 2:35
 home living environment changes 2:35
 urban environment changes 2:35
 familial factors
 family structure 2:33
 maternal employment 2:33
 parental education 2:33
 parenting styles 2:33
 poverty 2:33
 general discussion 2:37
 influencing factors
 community factors 2:33
 environmental factors 2:34
 familial factors 2:33
 general discussion 2:32
 genetic factors 2:32
 social support 2:33
 societal effects 2:33
 long-term effects
 cumulative effects 2:36
 early experiences 2:35
 latency/sleeper effects 2:35
 pathway effects 2:36
 postmodern societies 2:52
 social change impacts 2:32
 societal effects
 health considerations 2:33
 social capital 2:34
 social gradients 2:34
 educational goals 6:105
 harassment *see* workplace mistreatment/abuse
 hard-to-fill vacancies 8:320, 8:323
 harmonic mean 7:265
 Harrison, C. Keith 1:860
Hauptschule 5:585*f*, 5:586
- Hawaii
 Asian American/Pacific Islander student enrollment 1:597*f*
 tuition and needs-based grants 1:654*f*
- Hawthorne effects 2:448
- Hayek, Friedrich von 6:12
- heads of schools *see* academic managers
- Head Start program
 academic achievement outcomes 2:451–452
 center-based programs 2:100
 economically poor children 1:516
 funding estimates 2:58*t*
 government policies and programs 2:57
 public investment benefits 2:65
 research reviews 2:40*t*
 school-based health clinics 3:555
 transition programs 2:44
- health considerations 1:211–216
 adult learning outcomes 1:231
 asthma 2:635
 attention deficit/hyperactivity disorders (ADHD) 2:581, 2:590, 2:636, 2:854
 congenital heart defects 2:636
 continuing medical education (CME) 1:213
 current trends 1:211
 education–health association 1:667–672, 2:169–177
 alcohol consumption 1:670
 background information 1:667
 causal relationship
 behavioral risk factors 1:547, 1:669–670, 2:174
 enhanced cognitive abilities 2:176
 environmental factors 2:170, 2:257
 genetic risk factors 2:174
 health insurance 1:669, 2:175
 life expectancy effects 1:602, 1:603*f*, 1:603*t*, 1:668–669, 2:173*f*
 pathway diagram 2:171*f*
 research background 2:169
 social deprivation 2:173
 social standing 1:231, 2:172, 2:253–259
 stress factors 2:171
 general discussion 2:176
 health-care services 1:669, 2:175
 international perspectives 1:670
 lifestyle behaviors 1:547, 1:604, 1:669–670, 2:174
 measurement approaches 1:667
 mortality rates 1:602–609
 causal factors 1:604
 chronic diseases 1:602
 context of residence 1:607, 1:608*f*
 demographic characteristics 1:606
 discrimination 1:606
 family conditions 1:607
 gender 1:606
 HIV/AIDS 1:602
 inequality 1:604
 infant/childhood mortality 1:602, 1:607*t*
 life expectancy 1:602, 1:603*f*, 1:603*t*
 lifestyle behaviors 1:604, 1:605*t*
 marital status 1:607
 policy implications 1:608
 preventable deaths 1:605, 1:605*f*
 race and ethnicity 1:606, 1:607*t*
 rural environments 1:608*f*
 sense of control 1:606
 social support 1:606
 socioeconomic status (SES) 1:607, 1:668–669, 2:173*f*
 stratification 1:604
 suburban environments 1:608*f*
 suicide 1:602
 typology 1:602

- urban environments 1:608*f*
- working conditions 1:604
- research areas 1:671
- research results 1:547, 1:668
- self-rated health 1:668
- social and psychological components 1:669
- socioeconomic status (SES)
 - causal relationships 1:231, 2:175
 - educational attainment 1:667
 - mortality rates 1:607, 1:668–669, 2:173*f*
 - student achievement effects 2:253, 2:258
- general discussion 1:215
- health literacy 1:211
- school-based health clinics 3:555–562
 - future directions 3:560
 - general discussion 3:555
 - historical background 3:555
 - research evaluations
 - clinic effectiveness/costs 3:559
 - lessons learned 3:560
 - mental health access/services 3:557
 - reproductive health programs 3:558
 - utilization 3:556
- school health education 1:547–553
 - characteristics 1:547
 - coordinated school health programs 1:549, 1:550*f*
 - education content areas 1:548
 - general discussion 1:552
 - National Health Education Standards (NHES) 1:548
 - program frameworks 1:548
 - tobacco-use prevention
 - adolescent smoking consequences 1:549
 - educational interventions 1:549
 - tobacco industry involvement 1:551
- seizures and epilepsy 2:635
- social networks 6:709
- see also* special needs students
- health impairments 2:854
 - see also* physically disabled students; special needs students
- health informatics 3:52*t*
- health insurance 2:426
- health-related professions 3:51–57
 - accreditation 3:52, 3:53*t*
 - characteristics 3:51, 3:52*t*
 - educational trends
 - assessment strategies 3:56
 - clinical specialization 3:56
 - curriculum
 - curriculum shift 3:55
 - professional identity 3:55
 - faculty 3:55
 - institutions and degrees 3:55
 - interprofessional education 3:56
 - pedagogy 3:56
 - historical background 3:52
 - practice domains 3:53
 - professional responsibilities
 - access and referral practices 3:54
 - ethical practices 3:54
 - professional turf wars 3:54
 - technical paraprofessionals 3:54
- Healthy Babies, Healthy Children (HBHC) program 2:106
- Healthy Families America (HFA) program 2:126
- hearing impairments/loss 2:628–633
 - augmentative–alternative communication (AAC) 2:544–549
 - characteristics
 - general discussion 2:545
 - manual communication 2:545, 2:624
 - technological devices 2:545
 - communication initiation
 - challenges 2:547
 - communication partners 2:548
 - use settings 2:548
 - vocabulary selection 2:548
 - vowel-based communication chart 2:549*f*
 - communication partners 2:547, 2:548
 - goals and objectives 2:544
 - needs assessment 2:546
 - challenges 2:632
 - deaf-blind students 2:621–627
 - classroom intervention strategies 2:625
 - communication challenges 2:624
 - definitions 2:621
 - demographic data 2:623*t*
 - developmental challenges 2:622
 - early intervention services 2:622, 2:625
 - etiology 2:624, 2:624*t*
 - general discussion 2:626
 - hearing loss criteria 2:623
 - historical background 2:621
 - legal blindness criteria 2:623
 - transition processes 2:626
 - hearing loss criteria 2:623
 - instructional strategies
 - culturally and linguistically diverse (CLD) learners 2:631
 - early intervention 2:629
 - family support 2:628
 - general discussion 2:628
 - language/communication skills development 2:630
 - literacy skills development 2:630
 - sign language 2:545, 2:624, 2:630
 - school-based services 2:854
 - special education background 2:714
- heart disease *see* health considerations
- hedonic wage model
 - controllable versus uncontrollable cost factors 2:466, 2:467*t*
 - geographic cost-of-education index 2:467, 2:468*f*
 - limitations
 - comparable wage index (CWI) 2:468
 - cost factor interpretation 2:469
 - fixed-effects model 2:469
 - flexible wages 2:469
 - general discussion 2:468
 - geographic considerations 2:470
 - regional factors 2:468
 - teacher quality measures 2:469
 - public versus private schools 2:467
 - time correlations 2:467
- Hegel, Georg Wilhelm Friedrich 6:3, 6:111, 6:349
- hegemonic masculinity 1:840
- Heidegger, Martin 6:63–64, 6:82, 6:111, 6:184–185, 6:415, 6:450–451
- Heinz dilemma task 5:279–280
- help seeking 6:653–659
 - characteristics 6:653
 - computer-assisted processes 6:657
 - contextual influences 6:654
 - cultural influences 6:654
 - encouragement strategies
 - peer influences 6:656
 - teacher influences 6:655
 - importance 6:657
 - student achievement goals 6:654
- Helsinki Communiqué (2007) 8:436
- Herfindahl index of enrollment 2:333
- heritage language programs 1:295, 1:296*f*
- hermeneutics 6:63–68, 6:413–418
 - ancient historical background 6:63, 6:413
 - definition 6:63
 - general discussion 6:67, 6:413
 - Humanist perspectives 6:414
 - major themes
 - conservative hermeneutics 6:64
 - critical hermeneutics 6:65
 - educational thought
 - disciplinary knowledge 6:67
 - general discussion 6:65
 - language 6:66
 - questioning 6:66
 - world-historical situation 6:66
 - general discussion 6:64
 - moderate hermeneutics 6:65
 - radical hermeneutics 6:65
 - medieval practices 6:413
 - modern perspectives 6:63, 6:414
 - phenomenological hermeneutics 6:417
 - post-Reformation practices 6:413
 - pre-Romantic perspectives 6:414
 - reason and rationality 6:186
 - twentieth-century perspectives 6:7, 6:63, 6:415
- heroic leadership model 4:744
- heteronomy 6:121
- hidden curricula 1:631

- hierarchical linear models 7:210–215
 - basic concepts 7:8
 - effect size 7:446, 7:447*t*
 - general discussion 7:210, 7:214
 - generalized linear models (GLMs) 7:212
 - longitudinal/repeated measures model 3:120, 3:121*t*, 7:212
 - meta-analysis 7:214
 - multivariate longitudinal data analysis
 - basic concepts 3:120
 - self-concept study 3:121, 3:121*t*, 3:122*t*
 - variable measures 3:121*t*, 3:122*t*
 - nested data structures 7:210
 - recommended reading 7:214
 - shrinkage estimates 7:213
 - two-level linear model 7:210
 - value-added models (VAMs) 7:491
- high blood pressure *see* health considerations
- higher education 4:217–228
 - academic freedom 4:328–333
 - basic concepts 4:328, 4:434
 - freedom of academics 4:330, 4:332
 - historical background
 - freedom of academics 4:330
 - general discussion 4:329
 - institutional autonomy 4:329
 - restrictions 4:330
 - student freedom 4:331
 - institutional autonomy 4:341–346
 - background information 4:341
 - basic concepts 4:331
 - civic engagement 4:399
 - definitions 4:341
 - external boundaries 4:342
 - general discussion 4:345
 - historical background 4:329
 - internal boundaries 4:343
 - postconflict societies 4:519
 - restrictions 4:330, 4:331
 - student freedom 4:331, 4:332
 - academic mobility 4:559–565
 - future directions 4:564
 - globalization effects 4:209
 - government policies and programs 4:559
 - historical background 4:559
 - human resource issues 4:531
 - internationalization processes 4:533, 4:542
 - professional impacts 4:563
 - statistical data
 - faculty mobility 4:562
 - general discussion 4:560
 - student mobility 4:561, 4:598
 - student perspectives 4:562
 - academic values and expectations 4:433–440
 - background information 4:433
 - core cultural components 4:434
 - external pressures
 - accountability 4:437
 - economic imperatives 4:436
 - emerging skill sets 4:436
 - learning societies 4:437
 - massification and diversification 4:436
 - political impacts 4:435
 - technological advances 4:436
 - general discussion 4:438
 - new initiatives
 - funding considerations 4:438
 - learning centers 4:437–438
 - policies and programs 4:438
 - quality assurance 4:438
 - teaching impacts
 - deep learning versus surface learning 4:435
 - general discussion 4:434
 - learning paradigm shift 4:435
 - research-led teaching and learning 4:434
 - access and equity patterns 1:636–648
 - academic preparation 1:637, 1:652, 1:652*t*
 - college knowledge 1:637
 - demographic changes 1:636
 - financial considerations 1:637, 1:650, 1:652, 1:654*f*, 1:655*f*, 1:659
 - geographical information systems (GIS)
 - advantages 1:638
 - characteristics 1:637
 - college knowledge 1:645*f*, 1:646*f*
 - data sources and methods 1:638
 - illustrative examples 1:638
 - racial demographics 1:643*f*, 1:644*f*
 - research implications 1:647
 - Tucson elementary schools 1:639*f*, 1:643*f*, 1:644*f*
 - Tucson high schools 1:641*f*, 1:645*f*, 1:646*f*
 - Tucson middle schools 1:640*f*
 - government policies and programs 1:649–660
 - academic preparation 1:652, 1:652*t*
 - access trends 1:653, 1:655*f*
 - attendance costs 1:651*f*, 1:659
 - background information 1:650
 - continuation rates 1:655*f*
 - enrollment opportunities 1:656, 1:656*f*
 - enrollment rates 1:653, 1:655*f*
 - enrollment trends 1:657, 1:658*f*
 - grant and loan programs 1:650, 1:652, 1:654*f*, 1:655*f*
 - influencing factors 1:659
 - minority students 1:657, 1:658*f*
 - Pell grants 1:650, 1:651*f*, 1:656
 - post-secondary education opportunities 1:653
 - tuition and needs-based grants 1:652, 1:654*f*, 1:655*f*
 - international access patterns 4:334
 - labor market 1:636
 - social stratification 1:637
 - Tucson, Arizona
 - college knowledge 1:645*f*, 1:646*f*
 - elementary schools 1:639*f*, 1:643*f*, 1:644*f*
 - geographical information systems (GIS) research 1:638
 - high schools 1:641*f*, 1:645*f*, 1:646*f*
 - middle schools 1:640*f*
 - racial demographics 1:643*f*, 1:644*f*
 - research implications 1:647
 - research methodologies 1:638
- accreditation process 4:384–389
 - access expansion
 - accreditation impact 4:386
 - distance learning programs 4:385
 - for-profit programs 4:386
 - general discussion 4:385
 - basic concepts 4:384
 - change impacts
 - access expansion 4:385
 - accountability 4:388
 - internationalization processes 4:387
 - core values 4:384
 - functional roles 4:385
 - general discussion 4:389
 - international accreditation 4:533–539
 - agency responses 4:535
 - barriers 4:534
 - capacity development 4:537
 - cross-border education 4:537
 - current developments 4:536
 - future challenges 4:537
 - general discussion 4:538
 - improved dialog and cooperation 4:537
 - intergovernmental bodies 4:537
 - internationalization processes 4:533
 - mobility issues 4:533
 - professional accrediting bodies 4:535
 - quality assurance 4:533
 - university associations and networks 4:534
 - quality assurance 4:533–539
 - agency responses 4:535
 - background information 4:533
 - barriers 4:534
 - capacity development 4:537
 - cross-border education 4:537
 - current developments 4:536
 - future challenges 4:537
 - general discussion 4:538
 - improved dialog and cooperation 4:537
 - intergovernmental bodies 4:537
 - internationalization processes 4:533
 - professional accrediting bodies 4:535
 - university associations and networks 4:534
- action research 6:317
- admissions tests 1:845–849, 4:7–14

- background information 4:7
 changing contexts 1:845
 international institutions 4:12
 noncognitive variables
 characteristics 1:845, 1:846*t*
 community involvement 1:846*t*, 1:847
 long-range goals 1:846*t*, 1:847
 Noncognitive Questionnaire (NCQ) 1:847, 4:11–12
 nontraditional knowledge acquisition 1:846*t*, 1:847
 positive self-concept 1:846, 1:846*t*
 racial discrimination 1:846, 1:846*t*
 realistic self-appraisal 1:846, 1:846*t*
 successful leadership experiences 1:846*t*, 1:847
 support services 1:846*t*, 1:847
 Psychometric Entrance Test (PET) 4:13
 purpose 1:845
 SAT examinations 1:845
 SweSAT 4:12–13
 United States
 ACT Assessment 4:8
 fairness 4:12
 grade point average (GPA) 4:10
 Graduate Management Admission Test (GMAT) 4:9
 graduate/professional school programs 4:10
 Graduate Record Examinations (GRE) 4:8
 Law School Admission Test (LSAT) 1:695, 4:9
 Medical College Admission Test (MCAT) 4:9
 personal characteristics measures 4:11
 predictive validity 4:10
 primary admissions tests 4:7
 SAT examinations 4:8
 undergraduate institutions 4:9
 usage 4:9
 admittance pathways 4:609–615
 articulation 4:612
 bachelor's programs 4:609
 entry qualifications 4:610*t*, 4:611*t*
 postgraduate programs 4:611
 promotion strategies 4:613
 second chance students 4:611
 student transfers 4:610
 African Americans 1:767–771
 background information 1:767
 consequences 1:770
 demographic data 1:687
 educational decision-making 1:769
 educational experiences 1:770
 historically black colleges and universities (HBCUs) 1:767–768, 1:769, 4:479–480
 predominantly white universities 1:768, 4:479–480
 systemic racism
 consequences 1:770
 historical background 1:767
 historically black colleges and universities (HBCUs) 1:769
 predominantly white universities 1:768
 age-participation rates (APRs) 4:404
 Argentina 4:496
 Asian American/Pacific Islander student enrollment 1:596, 1:597*f*, 1:598*f*
 assessment strategies 3:249–255
 basic concepts 3:249
 general discussion 3:254
 group assessment 3:253
 information and communication technology (ICT) 3:253
 quality assurance 3:252, 4:381
 reference frameworks 3:251
 scope 3:249
 Australia
 collective bargaining 4:496
 federal political systems 4:606
 foreign doctoral students 4:292*t*
 income contingent loans (ICLs) 4:504
 loans and grants 4:501*f*, 4:502*f*
 structural and operational characteristics 5:511*f*, 5:513
 student and researcher mobility 4:287*f*
 student mobility 4:561
 tertiary education expenditures 4:500*f*
 tuition fees 4:500*f*
 Austria 5:520
 Bangladesh 5:527, 5:528*t*
 Belgium
 foreign doctoral students 4:292*t*
 loans and grants 4:501*f*, 4:502*f*
 student and researcher mobility 4:287*f*
 tertiary education expenditures 4:500*f*
 tuition fees 4:500*f*
 Bologna Accord (1999)
 academic mobility 4:560
 background information 1:772
 binary systems 4:617–618
 common frameworks 1:774, 4:568
 future directions 1:777
 Germany 5:587
 globalization effects 1:734–735, 1:775
 implementation challenges 1:776
 institutional autonomy 4:342
 internationalization processes 1:776, 4:544
 international programs 4:570
 quality assurance 4:569
 state steering models 4:557–558
 student achievement reporting 4:569
 study program levels 4:349
 Brazil
 collective bargaining 4:496
 quota systems 1:760–766
 academic achievement outcomes 1:764, 1:764*t*, 1:765*t*
 background information 1:760
 black students 1:762*t*
 course performance 1:764, 1:766*t*
 demographic data 1:762*t*
 ethnographic studies 1:762
 impacts 1:763, 1:764*t*
 public survey results 1:760
 quota versus nonquota student performance 1:764, 1:764*t*, 1:765*t*, 1:766*t*
 student distribution by school 1:764*t*
 white students 1:762*t*
 Bulgaria
 Academic Degrees and Titles Act (1972) 5:542
 Higher Education Act (1995) 5:542
 number of students 5:548*t*
 number of teachers 5:548*t*
 private universities 5:548*t*
 Promotion of Scientific Research Act (2003) 5:542
 state universities 5:548*t*
 structural and operational characteristics 5:543
 Cameroun 4:494
 Canada
 federal political systems 4:604
 foreign doctoral students 4:292*t*
 gender equity 1:689
 loans and grants 4:501*f*
 student and researcher mobility 4:287*f*
 tertiary education expenditures 4:500*f*
 tuition fees 4:500*f*
 Chile 4:496
 China
 collective bargaining 4:495
 English-delivered instruction 4:205
 enrollment rates 4:521*t*, 4:522*t*
 expansion effects 4:520
 faculty mobility 4:562
 gender equity 1:689
 income sources 4:521*t*
 literacy rates 4:521*t*
 partnerships 4:521
 private higher education (PHE) 4:623
 public expenditures 4:521*t*
 quality assurance 4:525
 student mobility 4:521, 4:561
 citizenship education 1:866–873
 background information 1:866
 democratic competency development
 campus experiences 1:868
 long-term effects 1:870
 meaningful initiatives 1:870
 multicultural competence 1:869
 student attitudes, values, and behaviors 1:869
 young adult participation 1:868
 diversity exposure 1:867
 general discussion 1:872
 service opportunity recommendations 1:870
 civic engagement 4:398–403

- higher education (*continued*)
 - challenges 4:402
 - community partnerships 4:390–397
 - background information 4:390
 - higher-education drivers 4:392
 - higher-education governance barriers 4:395
 - national policy barriers 4:393
 - regional development drivers 4:390
 - regional governance barriers 4:395
 - science-driven models 4:392–393
 - first-order engagement 4:400
 - general discussion 4:398
 - governance structures 4:401
 - Hegelian perspective 4:399
 - historical background 4:398
 - second-order engagement 4:400
 - stakeholders 4:401
 - third-order engagement 4:400
 - collective bargaining 4:491–498
 - basic concepts 4:491
 - challenges 4:496
 - country analysis
 - Argentina 4:496
 - Australia 4:496
 - Brazil 4:496
 - Cameroun 4:494
 - Chile 4:496
 - China 4:495
 - East Africa 4:493
 - France 4:493
 - general discussion 4:492
 - Germany 4:493
 - India 4:494
 - Indonesia 4:495
 - Japan 4:495
 - Malaysia 4:495
 - Mexico 4:496
 - Nigeria 4:494
 - Russia 4:493
 - South Africa 4:494
 - South Korea 4:494
 - Sweden 4:493
 - United Kingdom 4:493
 - United States 4:492
 - West Africa 4:493
 - international standards 4:491
 - collegiality 4:252, 4:266, 4:434
 - community partnerships 4:390–397
 - background information 4:390
 - higher-education drivers 4:392
 - higher-education governance barriers 4:395
 - national policy barriers 4:393
 - regional development drivers 4:390
 - regional governance barriers 4:395
 - science-driven models 4:392–393
 - computer-assisted instruction 4:419–426
 - assessment strategies 3:253
 - background information 4:419
 - curriculum 4:420
 - effective learning
 - collaborative learning 4:422
 - communication-based learning 4:422
 - construction-based learning 4:421
 - instructional strategies 4:421
 - e-learning pedagogies
 - average comparative ratings 1:531*t*
 - basic concepts 1:522
 - content appropriateness and relevancy 1:522
 - course goals 1:523
 - dimensions analysis 1:532
 - diversity considerations 1:523
 - effectiveness 4:423, 4:423*t*
 - interactive learning strategies 1:524
 - learners' needs 1:523
 - learning styles 1:522
 - multimedia 1:524
 - flexible learning
 - background information 4:441
 - basic concepts 4:423, 4:441–447
 - blended learning 4:445
 - criticisms 4:445
 - definitions 4:441
 - delivery challenges 4:444
 - delivery effectiveness 4:443
 - democratization of learning 4:444
 - flexible delivery 4:442
 - general discussion 4:441, 4:446
 - information and communication technology (ICT) 4:442
 - philosophical perspectives 4:444
 - responsiveness and availability 4:445
 - future directions 4:425
 - teaching approaches 4:424
 - technology enhanced learning (TEL) 4:423
- cross-border higher education (CBHE) 4:507–513
 - academic mobility 4:559–565
 - faculty mobility 4:562
 - future directions 4:564
 - government policies and programs 4:559
 - historical background 4:559
 - internationalization processes 4:533, 4:542
 - professional impacts 4:563
 - statistical data 4:560
 - student mobility 4:561
 - student perspectives 4:562
 - basic concepts
 - borderless education 4:509
 - cross-border education 4:509
 - general discussion 4:508, 4:587
 - offshore education 4:508
 - transnational education 4:508
 - challenges 4:512
 - current trends 4:408
 - frameworks 4:510, 4:510*r*
 - growth trends 4:507
 - international educational terminology 4:507, 4:508*r*
 - program mobility typology 4:511
 - providers
 - alternative providers 4:511
 - mobility typology 4:512
 - traditional higher education institutions 4:510
 - quality assurance 4:537
 - curricula globalization 1:403–410
 - background information 1:403
 - curriculum design 1:404
 - general discussion 1:409
 - knowledge-based economies 1:403
 - lifelong learning model 1:404, 1:405*r*
 - reform efforts
 - assessment methods 1:407
 - classroom dynamics 1:408
 - course content 1:406
 - critical perspectives 1:408
 - institutional change 1:406
 - instructional strategies 1:407
 - research and development expenditure levels 1:404*r*
 - traditional learning model 1:405*r*
 - curriculum innovations
 - design considerations 4:452
 - general discussion 4:448
 - outcomes analyses 4:451
 - degree mills 4:600
 - diversification considerations 4:347–353
 - classifications 4:347
 - diversity discourse 1:823
 - influencing factors 4:352
 - institution types 4:348
 - key features 4:347
 - rankings systems 4:350
 - reputational diversification 4:350
 - study program levels 4:349
 - theoretical perspectives 4:350
 - diversity discourse 1:823–826
 - background information 1:823
 - discrimination 1:826–826
 - feminist perspectives 1:824
 - interdisciplinary studies 1:824
 - legal contexts 1:823
 - social contexts 1:823
 - Supreme Court decisions 1:823
 - doctoral student attrition 1:827–837
 - background information 1:827
 - data sources 1:828, 1:829*t*

- methodology
 - data-analysis strategy 1:830
 - dependent variable 1:830
 - discrete-time hazard analysis 1:828
 - independent variables 1:830, 1:831*t*
- research background 1:827
- research questions 1:828
- research results
 - descriptive statistics 1:830
 - general discussion 1:833
 - hazard rate 1:832*f*
 - life table statistics 1:832*t*
 - multivariate model 1:834–835*t*, 1:831
 - policy and practice implications 1:835
 - survival curve 1:832*f*, 1:833*f*
- theoretical frameworks 1:828
- East Africa 4:493
- educational innovations 4:448–454
 - classroom assessment 4:450
 - contextual perspectives 4:448
 - curriculum
 - design considerations 4:452
 - general discussion 4:448
 - outcomes analyses 4:451
 - individual versus systemic approaches 4:449
 - information and communication technology (ICT) 4:453
 - student approaches to learning 4:451
 - student learning impacts 4:450
 - teaching and curriculum changes 4:448
 - teaching practices 4:450
 - typologies 4:449
 - whole-degree program 4:452
- educational reform 1:698–704
 - diversification guidelines 1:700
 - globalization effects 1:698
 - goals and objectives 1:701–702
 - lifelong learning 1:699
 - public university funding cuts 1:701, 1:707
 - World Conference on Higher Education 1:698, 4:601
- effective learning 4:461–466
 - background information 4:461
 - computer-assisted instruction
 - collaborative learning 4:422
 - communication-based learning 4:422
 - construction-based learning 4:421
 - instructional strategies 4:421
 - Self-Regulated Learning (SRL) 4:461, 4:462
 - Student Approaches to Learning (SAL) 4:461, 4:462
 - student engagement 4:464
 - teaching guidelines 4:465
- e-learning
 - background information 1:521
 - basic concepts 8:102
 - case-study analysis
 - average comparative ratings 1:530, 1:531*t*
 - top-rated factors 1:532
 - critical factors 1:525*f*, 1:533
 - enrollment growth 4:598
 - faculty
 - administrative skills 1:526
 - average comparative ratings 1:531*t*
 - competency standards 1:525
 - dimensions analysis 1:532
 - faculty engagement 1:526
 - innovativeness 1:526
 - pedagogical proficiency 1:526
 - recognition 1:527
 - technical skills 1:526
 - framework dimensions
 - critical factors 1:525*f*
 - faculty 1:525
 - general discussion 1:522
 - institutions 1:529
 - learner support 1:528
 - pedagogy 1:522
 - technological infrastructure 1:527
 - institution
 - appropriate processes 1:529
 - average comparative ratings 1:531*t*
 - dimensions analysis 1:533
 - financial feasibility and sustainability 1:529
 - leadership commitment 1:529
 - quality commitment 1:529
 - interactive learning strategies
 - basic concepts 1:524
 - course goals 1:524
 - feedback 1:524
 - flexibility 1:524
 - learner support
 - appropriate support services 1:528
 - average comparative ratings 1:531*t*
 - dimensions analysis 1:533
 - faculty responsiveness 1:528
 - future directions 4:431
 - general discussion 1:528
 - multimedia
 - diversity considerations 1:524
 - relevancy 1:524
 - pedagogy
 - average comparative ratings 1:531*t*
 - basic concepts 1:522
 - content appropriateness and relevancy 1:522
 - course goals 1:523
 - dimensions analysis 1:532
 - diversity considerations 1:523
 - interactive learning strategies 1:524
 - learners' needs 1:523
 - learning styles 1:522
 - multimedia 1:524
 - research objectives and methodology 1:530
 - technological infrastructure
 - accessibility 1:528
 - appropriateness 1:527
 - average comparative ratings 1:531*t*
 - dimensions analysis 1:532
 - reliability 1:527
- emerging economies 4:520–526
 - Bangladesh 4:523
 - challenges 4:525
 - China 4:520, 4:521*t*
 - cross-cutting issues
 - employability 4:525
 - expansion financing 4:523
 - private sector services 4:524, 4:524*t*
 - quality assurance 4:525
 - relevancy 4:525
 - enrollment rates 4:521*t*
 - India 4:521, 4:521*t*, 4:522*t*
 - literacy rates 4:521*t*
 - Pakistan 4:523
 - public expenditures 4:521*t*
 - Sri Lanka 4:523
- engaged education 1:711–718
 - engagement concepts 1:711, 1:712*f*
 - experiential learning 1:713, 1:714*f*
 - importance 1:716
 - intensive field experiences
 - basic concepts 1:714
 - curriculum development and logistics 1:716, 1:717*t*
 - goals and objectives 1:715
 - social and cultural change 1:714
- England
 - bachelor's degrees 5:553
 - doctorates 5:554
 - loans and grants 5:554
 - master's degrees 5:554
 - professional qualifications 5:554
 - quality assurance 5:550
 - structural and operational characteristics 5:553
- English-delivered instruction 4:203
- Estonia 5:556
- ethnic minorities
 - admissions test fairness 4:12
 - demographic data 1:687
 - enrollment opportunities 1:656, 1:658*f*
 - United Kingdom 1:673–678
 - academic preparation 1:674
 - campus experiences 1:676
 - diversity considerations 1:673
 - institutional dynamics 1:675
 - parental involvement 1:676
 - peer influences 1:676

- higher education (*continued*)
 poverty concerns 1:675
 primary and secondary education 1:674
 federal political systems **4:603–608**
 Australia 4:606
 Canada 4:604
 characteristics 4:603
 general discussion 4:608
 Germany 4:605
 Mexico 4:607
 Switzerland 4:606
 United States 4:604
 Fiji 5:566
 financial considerations **4:499–506, 4:573–583**
 characteristics
 living costs 4:499
 loans and grants 4:500, 4:502f
 public support 4:500, 4:501f
 tertiary education expenditures 4:499, 4:500f
 tuition fees 4:499, 4:500f
 economic impacts
 externalities 4:575
 general discussion 4:573
 rate of return estimates 4:574
 educational production function model 4:576
 funding strategies 4:579
 general discussion 4:499, 4:506, 4:573
 historical development 4:221
 income contingent loans (ICLs)
 Australia 4:504
 background information 4:504
 New Zealand 4:504
 South Africa 4:505
 Sweden 4:504
 United Kingdom 4:505
 United States 4:505
 Yale Plan 4:504
 information access and use 4:577
 organizational structures 4:576
 student loans
 conventional loans 4:503
 cost–benefit analysis 4:503
 importance 4:501
 income contingent loans (ICLs) 4:503
 international reforms 4:504
 Finland 5:570f, 5:572
 for-profit sector **1:705–710**
 comparisons with nonprofit institutions 1:708
 driving forces
 commercialization 1:706
 globalization 1:705
 labor market 1:708, 1:710f
 public good versus private good issues 1:706
 public university funding cuts 1:701, 1:707
 technology impacts 1:707
 future directions 1:709
 growth factors
 degree-granting institutions 1:706t
 enrollment 1:706t
 general discussion 1:705
 non-degree-granting institutions 1:706t
 France
 collective bargaining 4:493
 faculty mobility 4:562
 loans and grants 4:501f
 structural and operational characteristics 5:578
 student and researcher mobility 4:287f
 student mobility 4:561
 tertiary education expenditures 4:500f
 tuition fees 4:500f
 future directions
 expansion effects 4:223, 4:377
 knowledge-based economies 4:223, 4:265
 political economy 4:225
 sociocultural factors 4:226
 Gambia 5:861
 gender equity **1:629–635**
 academic identities 1:630
 access and equity patterns 4:336, 4:336t
 challenges 1:629
 demographic data 1:687
 female participation 1:630
 future directions 1:633
 global gender parity index (GPI) 1:630
 government policies and programs 1:633
 Germany
 collective bargaining 4:493
 federal political systems 4:605
 future directions 5:588
 structural and operational characteristics 5:585f, 5:587, 5:589t
 Ghana 5:592t, 5:594
 global impact 4:209, 4:217, 4:377, 4:540–541
 globalization **1:536–541, 4:201–208**
 Cambodia 4:205
 China 4:205
 contemporary globalization 4:202
 contextual perspectives 4:597
 educational impacts 4:206, 4:601
 enrollment growth 4:597
 global English 4:203
 globalization-from-above
 general discussion 1:537
 globalization-from-above versus globalization-from-below 1:536
 homogenizing effects 1:538
 intensified hierarchies 1:537
 privatization 1:537
 globalization-from-below
 collaborative relationships 1:538
 general discussion 1:538
 globalization-from-above versus globalization-from-below 1:536
 massification 1:539
 socially responsible practices 1:539
 historical background 4:201
 India 4:205
 medium of instruction (MOI) 4:202, 4:205
 neoliberal perspectives 1:536
 private higher education (PHE) 4:597
 research background 1:536
 research results and analysis 1:540
 graduate employment **4:354–359**
 differentiation considerations 4:356
 division of labor 4:357
 employability enhancements 4:356
 employment rates and earnings 4:354
 expansion effects 4:354, 4:365
 globalization 4:358
 job quality 4:355
 knowledge society theory 4:358
 labor market 4:366
 research biases 4:359
 student demands 4:354
 transition periods 4:357
 worker productivity 4:355
 Greece 5:601
 historical development
 current trends
 funding considerations 4:221
 policies and programs 4:219
 push–pull process 4:219, 4:371
 structural reforms 4:220
 student demands 4:219, 4:354, 4:377
 expansion effects 4:218, 4:354, 4:365, 4:370, 4:377
 globalization 4:201
 governance **1:728–736**
 Catholic Church 1:728
 entrepreneurial university model 1:734, 4:249
 evaluative-based transformations 1:732
 globalization 1:734
 institutional expenditures 1:733t
 massification (1950–1980) 1:731
 medieval university 1:728
 modern universities (1800–1950) 1:730
 parental expenditures 1:734t
 Protestant Reformation period 1:729
 quality assurance 1:728–729, 1:731, 1:734
 triangle of coordination 1:728, 1:729f
 tuition and fees 1:733–734
 universal enrollment 1:735
 theoretical perspectives 4:222
 historically black colleges and universities (HBCUs) 1:767–768, 1:769, 4:479–480
 human resource issues **4:527–532**

- academic mobility 4:531
- advancement and promotion opportunities 4:529
- challenges 4:527
- general discussion 4:527, 4:531
- part-time and temporary appointments 4:528
- recruitment and hiring 4:528
- workplace and employment trends 4:530
- Iceland 5:607*f*, 5:607*t*, 5:608
- India
 - collective bargaining 4:494
 - English-delivered instruction 4:205
 - enrollment rates 4:521*t*, 4:522*t*
 - faculty mobility 4:562
 - government policies and programs 4:521
 - institutional expenditures 1:733*t*
 - literacy rates 4:521*t*
 - parental expenditures 1:734*t*
 - private higher education (PHE) 4:623
 - public expenditures 4:521*t*
 - quality assurance 4:525
 - statistical data 4:522*t*
 - student mobility 4:561
 - tuition and fees 4:523
- indigenous education 1:663
- Indonesia
 - characteristics 5:613, 5:615
 - collective bargaining 4:495
 - number of lecturers 5:615*t*
 - structural and operational characteristics 5:612*f*
- institutional autonomy 4:341–346
 - background information 4:341
 - civic engagement 4:399
 - definitions 4:341
 - external boundaries 4:342
 - general discussion 4:345
 - internal boundaries 4:343
- institutional management 4:264–272
 - administration 4:247
 - changing contexts
 - general discussion 4:264
 - market environment 4:265
 - new public management (NPM) 4:264
 - policy environment 4:264
 - councils or boards
 - chairperson 4:277
 - changing roles 4:275
 - composition 4:276
 - department heads 4:229–237
 - complex roles 4:230
 - contractual obligations 4:230, 4:231*t*, 4:232*t*
 - effective leadership behaviors 4:234*t*
 - excellence criteria 4:233*t*
 - future challenges 4:235
 - heads-of-school framework 4:232*t*
 - importance 4:229
 - priority duties 4:231*t*
 - specific leadership needs 4:234*t*
 - staff expectations 4:231
 - training and development 4:234, 4:235*t*, 4:236*t*
 - transitions to leadership 4:233*t*
 - general discussion 4:264, 4:270
 - governance 4:273–278
 - changing contexts 4:273
 - collegiality 4:252, 4:266
 - councils or boards 4:275
 - emerging trends 4:273
 - enrollment growth 4:273
 - entrepreneurial models 4:266
 - financial control models 4:274*t*
 - general discussion 4:277
 - governing boards 4:267
 - institutional control models 4:274*t*
 - institutional structures 4:246, 4:267, 4:401
 - new public management (NPM) 4:265
 - reporting and regulation practices 4:275*t*
 - standards/codes of practice 4:277
 - state steering models 4:554–558
- institutional head
 - as change leader 4:270
 - functional role 4:267
 - influencing factors 4:268
- leadership structures 4:246
- market environment
 - competition 4:265
 - globalization 4:265
 - knowledge-based economies 4:223, 4:265
- international access patterns 4:334–340
 - gender parity index 4:336*t*
 - general discussion 4:339
 - global participation 4:335, 4:335*f*
 - research areas
 - data collection 4:338
 - expansion limits 4:338
 - funding considerations 4:339
 - meritocratic entry concepts 4:338
 - open access policies 4:338
 - privatization impacts 4:339
 - research data collection 4:334
 - social equity studies 4:334
 - underrepresented groups 4:336, 4:336*t*
- internationalization processes 4:540–545
 - academic mobility 4:559–565
 - background information 4:540
 - basic concepts 4:541
 - challenges 1:776, 4:544
 - faculty mobility 4:562
 - globalization 4:541, 4:597
 - governance 1:734
 - institutional strategies 4:543
 - student mobility 4:542, 4:561, 4:598
- Israel 5:621, 5:622*t*
- Jamaica 5:627, 5:628*f*
- Japan 4:495
- knowledge society 4:360–364
 - general discussion 4:364
 - graduate employment 4:358
 - intellectual property rights 4:362
 - internationalization processes 4:540–541
 - knowledge production processes 4:361
 - political goals
 - general discussion 4:362
 - global information societies 4:363
 - nation-states level 4:363
 - regional level 4:363
 - social transformation 4:358, 4:373
- labor market relationship 4:365–369
 - access and equity patterns 1:636
 - expansion effects 4:365
 - general discussion 4:365
 - graduate employment 4:366
 - institutional characteristics 4:367
- Mexico
 - demographic structure 1:720, 1:721*t*
 - economic dynamics 1:720
 - employment trends 1:722–724, 1:722*t*
 - future directions 1:726
 - industrial structure 1:722, 1:722*t*, 1:723*t*
 - social change impacts 1:719
 - university–industry collaboration 1:724
- qualitative aspects 4:366
- quantitative aspects 4:365
- research background 4:368
- Latino students 1:610–615
 - college-choice model
 - basic concepts 1:611, 1:612*f*
 - chain migration theory 1:613
 - decision-making models 1:613*f*
 - quantitative methods 1:612
 - social capital theory 1:613
 - demographic data 1:610
 - educational attainment 1:610
 - learning opportunities 1:610
- Latvia 5:632, 5:633*t*
- learning cities and regions 1:147
- Lebanon 5:641
- Lithuania 5:660
- Malawi 5:671
- Malaysia 4:495
- Malta 5:676*f*, 5:678
- management practices 4:245–250, 4:251–255
 - adaptive university model 4:249
 - background information 4:245

- higher education (*continued*)
- business-like practices 4:253, 4:266
 - changing contexts
 - competition 4:252, 4:265
 - financial considerations 4:252
 - institutional growth 4:251–252
 - collegiality 4:252, 4:266
 - entrepreneurial university model 1:734, 4:249
 - institutional management 4:264–272
 - administration 4:247
 - changing contexts 4:264
 - general discussion 4:264, 4:270
 - governance 4:246, 4:265, 4:273–278, 4:401
 - institutional head 4:267
 - leadership structures 4:246
 - key features
 - administration 4:247
 - governance structures 4:246, 4:401
 - leadership structures 4:246
 - learning university model 4:249
 - new public management (NPM)
 - background information 4:251
 - basic concepts 4:248
 - collegiality 4:266
 - current trends 4:264
 - entrepreneurial models 4:266
 - evaluation policies and practices 3:648
 - governance 4:265
 - governing boards 4:267
 - ideological perspectives 4:251
 - institutional autonomy 4:342
 - institutional structures 4:267
 - strategic planning 4:256, 4:257*f*
 - organizational culture 4:253
 - professionalization 4:248–249
 - strategic management 4:248
 - university leaders 4:238–244
 - diversity considerations 4:238
 - future research areas 4:242
 - historical perspectives 4:238
 - literature reviews 4:238
 - literature types 4:239
 - presidential leadership 4:239, 4:240
 - senior management teams 4:239, 4:241
- Mexico
- collective bargaining 4:496
 - federal political systems 4:607
 - loans and grants 4:501*f*, 4:502*f*
 - social change 1:719–727
 - demographic structure 1:720, 1:721*t*
 - economic dynamics 1:720
 - employment trends 1:722–724, 1:722*t*
 - future directions 1:726
 - industrial structure 1:722, 1:722*t*, 1:723*t*
 - labor-market perspective 1:719
 - university–industry collaboration 1:724
 - tertiary education expenditures 4:500*f*
 - tertiary type-B program survival rates 8:464*t*
- minority students
- admissions test fairness 4:12
 - demographic data 1:687
 - enrollment opportunities 1:656, 1:658*f*
- United Kingdom 1:673–678
- academic preparation 1:674
 - campus experiences 1:676
 - diversity considerations 1:673
 - institutional dynamics 1:675
 - parental involvement 1:676
 - peer influences 1:676
 - poverty concerns 1:675
 - primary and secondary education 1:674
- Mongolia 5:682*t*, 5:683
- Namibia 5:687
- national science policy 4:294–299
- general discussion 4:299
 - governance processes 4:295
 - historical background 4:294
 - institutional structures 4:297
 - policies and programs 4:295
 - research challenges 4:298
 - state–university relations 4:296
- New Zealand 5:695
- foreign doctoral students 4:292*t*
 - income contingent loans (ICLs) 4:504
 - loans and grants 4:501*f*, 4:502*f*
 - structural and operational characteristics 5:695
 - student and researcher mobility 4:287*f*
 - tertiary education expenditures 4:500*f*
 - tertiary type-B program survival rates 8:464*t*
 - tuition fees 4:500*f*
- Niger 5:702, 5:704*t*
- Nigeria 4:494
- Norway 5:711*f*, 5:712*t*, 5:713
- open and distance learning (ODL) 4:598
- Organisation for Economic Cooperation and Development (OECD)
- access and equity patterns 4:334
 - cross-border higher education (CBHE) 4:509, 4:587
 - Directorate for Science, Technology and Industry (DSTI) 4:588
 - future scenarios 4:587
 - global dimensions 4:588
 - Institutional Management in Higher Education (IMHE) 4:588
 - policies and programs 4:349, 4:586
 - policy debates 4:598
- Pakistan 5:731
- Philippines
- admission requirements 5:870
 - Commission on Higher Education (CHED) 5:868
 - degree programs 5:870
 - grading systems 5:870
 - institutional distribution 5:868*t*
 - private higher education (PHE) 5:870
 - quality assurance 5:870
 - research activities 5:871
 - state universities and colleges (SUCs) 5:869
 - structural and operational characteristics 5:866
- Poland 5:739
- postconflict societies 4:514–519
- basic concepts 4:514
 - functional role 4:515
 - future directions 4:519
 - intervention strategies
 - Afghanistan 4:518
 - general discussion 4:516
 - Iraq 4:518
 - South Africa 4:516
 - south-southeast Europe (SEE) 4:517
 - postconflict requirements 4:516
- predominantly white universities 1:768, 4:479–480
- private higher education (PHE) 4:622–627
- accountability 4:625
 - characteristics
 - general discussion 4:623
 - group identity 4:623
 - nonelite subsector 4:624
 - semi-elite institutions 4:623
 - emergence and growth 4:597, 4:622
 - funding considerations 4:625
 - general discussion 4:622, 4:626
 - governance 4:625
- professional autonomy 3:48, 4:434
- professional education development 4:411–418
- challenges
 - accredited programs 4:415
 - aging demographics 4:414
 - curricula vocationalization 4:414
 - new technologies 4:414
 - structural instabilities 4:415
 - theoretical perspectives 4:415
 - contextual perspective and background 4:411
 - educational approaches
 - background information 4:412
 - learner-focused approaches 4:413
 - organization-focused approaches 4:413
 - sector-focused approaches 4:413
 - teacher-focused approaches 4:413
 - educational models 4:412
 - future directions
 - distributed roles 4:416
 - general discussion 4:415
 - international and national collaboration 4:416
 - key performance indicators 4:415
 - professionalization 4:416

- general discussion 4:416
- terminology and meaning 4:411
- publication productivity 4:455–460
 - age-related considerations
 - age decrement hypothesis 4:458
 - cumulative disadvantage hypothesis 4:458
 - general discussion 4:458
 - intellectual deadlock hypothesis 4:458
 - obsolescence hypothesis 4:458
 - seniority burden hypothesis 4:458
 - utility-maximizing hypothesis 4:458
 - current trends 4:455
 - gender differences
 - basic concepts 4:459
 - motherhood 4:459
 - networking and collaboration 4:459
 - general discussion 4:459
 - measurement approaches 4:455
 - output analyses
 - ability measures 4:456
 - collaboration 4:457
 - communication networks 4:457
 - cumulative advantages and disadvantages 4:457
 - location 4:457
 - resource availability 4:457
 - statistical data 4:456
 - time and energy 4:456
- Qatar 5:747
- quality assurance 4:377–383
 - external quality assurance practices
 - basic concepts 4:377, 4:378*t*
 - examination systems 4:381
 - information availability 4:381
 - national qualifications frameworks 4:379
 - quality assessments 4:380
 - importance 4:382
 - racial and ethnic demographics 1:687
- rankings systems 4:546–553
 - background information 4:546
 - general discussion 4:552
 - global university rankings
 - general discussion 4:548
 - Newsweek* rankings 4:551
 - performance factors 4:549, 4:550*t*
 - Shanghai Jiao Tong University Institute of Higher Education (SJTJHE)
 - rankings 4:204, 4:548, 4:548*t*, 4:550*t*
 - Times Higher Education Supplement (THES) rankings 4:548*t*, 4:549
 - holistic reputational rankings 4:551
 - national rankings
 - Centre for Higher Education Development (CHE) rankings 4:547
 - geographic distribution 4:546
 - U.S. News & World Report (USNWR)* rankings 2:231, 4:547
 - research methodologies 4:551
- research commercialization 4:300–306
 - background information 4:300
 - copyright 4:303
 - criticisms 4:304
 - economic development
 - general discussion 4:302
 - incubators 4:303
 - research parks 4:302
 - start-up companies 4:302
 - technology transfer 4:303
 - government policies and programs 4:300
 - infrastructure 4:303
 - patents
 - background information 4:301
 - exceptions 4:301
 - personnel coverage 4:301
 - royalties 4:301
- research funding 4:279–285
 - background information 4:279
 - commercialization efforts 4:284
 - government policies and programs 4:281
 - priority setting 4:282
 - public funding 4:281
 - quality assessments 4:283
 - research characteristics and functions 4:280
 - university–industry collaboration 4:283
- research management 4:316–321
 - background information 4:316
 - emerging professional networks 4:318
 - functional role 4:318
 - general discussion 4:320
 - management changes
 - collaborative research 4:317
 - competition and selectivity 4:316–317
 - external research contracts 4:317–318
 - financial transparency 4:317
 - general discussion 4:316
 - increased accountability 4:316
 - knowledge transfer 4:317
- Russia 5:754
 - collective bargaining 4:493
- Rwanda 5:762, 5:762*t*
- school–higher education institution (HEI) partnerships 7:757–763
 - characteristics 7:757
 - collaborative action research 7:759
 - general discussion 7:762
 - Innovative Links Project 7:758
 - National Schools Network 7:758
 - networked learning communities (NLC) 7:759
 - partnership models
 - accountability 7:761
 - change processes 7:762
 - cooperation and collaboration 7:761
 - features and consequences 7:760*t*
 - general discussion 7:760
 - improvement strategies 7:762
 - leadership considerations 7:761
 - sustainability 7:761
 - preservice education programs 7:758
 - professional development schools 7:758
 - school-level development 7:758
 - supervision 7:619
 - vocational education and training (VET) 8:249
- Scotland 5:772
- Singapore 5:778
- Slovenia 5:786
- social transformation 4:370–376
 - civil society 4:371
 - clever cities theory 4:373
 - domestic participation 4:371
 - expansion effects 4:218, 4:354, 4:365, 4:370, 4:377
 - future directions 4:375
 - general discussion 4:227
 - knowledge production processes 4:374
 - knowledge society theory 4:358, 4:360–364, 4:373
 - push–pull process 4:219, 4:371
- Solomon Islands 5:794
- South Africa
 - collective bargaining 4:494
 - gender equity 1:630–631
 - income contingent loans (ICLs) 4:505
 - national technical talent 1:855–859
 - institutional investments 1:858
 - internationalization processes 1:857
 - philanthropic assistance 1:857
 - recruitment and retention 1:855
 - South African Network of Skills Abroad (SANSA) 1:857–858
 - Southern African Norwegian Training, Education and Development (SANTED) initiative 1:855–856
 - study abroad programs 1:855
- postconflict societies 4:516
 - structural and operational characteristics 5:802, 5:802*t*
- South Korea 4:494
- Spain 5:809*t*, 5:810
- Sri Lanka 5:816, 5:822
- state's role 4:554–558
 - Clark's model 4:554
 - multilateral models
 - corporate-pluralist model 4:556
 - institutional model 4:556
 - sovereign, rationality-bounded model 4:556
 - supermarket model 4:556
 - neoliberal ideologies and strategies 4:557
 - policy instruments 4:554
 - state steering models
 - bilateral relationships 4:555
 - contextual relationships 4:556
 - general discussion 4:555
 - multilateral relationships 4:556

higher education (*continued*)

strategic planning 4:256–263

basic model

deliberate versus emergent strategies 4:258, 4:258*f*

empirical analyses 4:257

general discussion 4:256

new public management (NPM) approach 4:257*f*planning cycle 4:257*f*plan outline 4:258*f*

current trends 4:256

goal and performance cards 4:262, 4:262*f*goals and objectives 4:256, 4:257*f*

internal contract management 4:261

methodologies and instrumentation

general discussion 4:260

portfolio analyses 4:261, 4:262*f*

strengths, weaknesses, opportunities, and threats (SWOT) analyses 4:261

performance areas 4:257

strategic perspectives 4:259, 4:260*f*

student athletes 1:860

academic achievement outcomes 1:860

Coach Carter film screening 1:865

critical intervention programs 1:861

incentive and reward systems 1:864

popular culture 1:863

Scholar–Baller® program 1:862

status symbols 1:861, 1:863

student attrition 4:467–472

cost estimation 4:471

estimation methods 4:467

general discussion 4:467

impacts 4:471

retention/intervention strategies 4:470

theoretical perspectives 4:468

student experiences 4:473–478

convergence patterns 4:477

diversity considerations 4:476

first-year experiences 4:475

general discussion 4:473

globalization effects 4:477

program reviews 4:473

psychosocial models 4:474

social dimensions 4:475

student engagement 4:476

study characteristics and purpose 4:473

student learning outcomes 4:479–484

benefits 4:483

cognitive complexity 4:479, 4:480*f*humanitarianism 4:480*f*, 4:481interpersonal and intrapersonal competence 4:480*f*, 4:482knowledge acquisition 4:480, 4:480*f*practical competence 4:480*f*, 4:482

research background 4:479

student mobility 4:598

student support services 4:427–432

academic support

current perspectives 4:428

delivery approaches 4:428

historical background 4:427

basic concepts 4:427

contextual perspectives 4:427

distance education 4:430

future directions 4:431

information literacy

delivery approaches 4:429

historical background 4:428

importance 4:428

libraries/learning spaces

current trends 4:429

historical background 4:430

Learning Commons 4:430

Sultanate of Oman 5:723

Suriname 5:830

Sweden

collective bargaining 4:493

foreign doctoral students 4:292*t*

income contingent loans (ICLs) 4:504

loans and grants 4:501*f*, 4:502*f*student and researcher mobility 4:287*f*tertiary education expenditures 4:500*f*tertiary type-B program survival rates 8:464*t*tuition fees 4:500*f*

Switzerland 4:606

system structure 4:616–621

binary systems 4:617

classification schemes 4:619

general discussion 4:616

stratified systems 4:616

structural typologies 4:616

tension and challenges 4:619

unified systems 4:616

Taiwan 5:836*f*, 5:837, 5:837*t*

Tanzania 5:843

Thailand 5:853

tuition and fees 2:229–234

cost research 2:229

future directions 2:233

graduate and professional programs 2:232

private institutions 2:230

public institutions 2:232

UNESCO (United Nations Educational, Scientific and Cultural Organization)

access and equity patterns 4:334

accreditation processes 4:537

collective bargaining 4:491

cross-border education 4:509

enrollment predictions 1:700, 1:702*f*, 1:703*f*

globalization effects 1:698, 4:601

global mission 4:597–602

background information 4:597

capacity-building/quality assurance initiatives 4:600

cross-border higher education (CBHE) 4:600

degree mills 4:600

Global Forum on International Quality Assurance, Accreditation and the Recognition of Qualifications 4:599

Global Initiative for Quality Assurance Capacity (GIQAC) 4:600

policy debates 4:598

standard setting 4:599

UNESCO Portal 4:600

graduate employment 4:359

massification (1950–1980) 1:731

state steering models 4:557–558

universal enrollment 1:735

university associations 4:487

World Conference on Higher Education 1:698, 4:601

United Kingdom

bachelor's degrees 5:553

binary systems 4:618

collective bargaining 4:493

doctorates 5:554

enrollment rates 1:673

ethnic minorities 1:673–678

academic preparation 1:674

campus experiences 1:676

diversity considerations 1:673

institutional dynamics 1:675

parental involvement 1:676

peer influences 1:676

poverty concerns 1:675

primary and secondary education 1:674

foreign doctoral students 4:292*t*

gender equity 1:689

historical development 1:731

income contingent loans (ICLs) 4:505

loans and grants 4:501*f*, 4:502*f*, 5:554

master's degrees 5:554

professional qualifications 5:554

quality assurance 5:550

structural and operational characteristics 5:553

student and researcher mobility 4:287*f*, 4:561tertiary education expenditures 4:500*f*tertiary type-B program survival rates 8:464*t*tuition fees 4:500*f*

United States

collective bargaining 4:492

degree-granting institutions 5:892*t*demographic data 5:893*t*enrollment 5:892*t*

faculty mobility 4:562

federal political systems 4:604

foreign doctoral students 4:292*t*

gender equity 1:689

general discussion 5:892

- historical development 1:730
- income contingent loans (ICLs) 4:505
- loans and grants 4:501*f*
- student and researcher mobility 4:287*f*, 4:561
- tertiary education expenditures 4:500*f*
- tuition fees 4:500*f*
- universal enrollment 1:735
- university associations 4:485–490
 - consortia 4:487
 - future directions 4:489
 - international accreditation issues 4:534
 - international associations 4:487
 - national and regional subgroups 4:486
 - national associations 4:485
 - scope 4:485
 - theoretical analysis 4:489
- university leaders 4:238–244
 - diversity considerations 4:238
 - future research areas 4:242
 - historical perspectives 4:238
 - literature reviews 4:238
 - literature types 4:239
 - presidential leadership 4:239, 4:240
 - senior management teams 4:239, 4:241
- Uruguay 5:875*f*
- vocational–academic education parity 8:433
- West Africa 4:493
- World Bank (WB)
 - access and equity patterns 4:334
 - developing countries 4:590–596
 - background information 4:590
 - future directions 4:595
 - lending statistics 4:594*t*
 - new commitments 4:594*t*
 - policy frameworks 4:590, 4:593*t*
 - recent activities 4:592
 - research studies 4:594*t*
 - funding considerations 4:221
 - importance 4:590
 - state steering models 4:557–558
- Zanzibar 5:898, 5:898*t*
 - see also* distance education; tertiary education; universities
- Higher Education Act (1965) 1:650
- Higher Education Act (1995) 5:542
- Higher Education in Developing Countries* (World Bank) 4:591, 4:593*t*
- Higher Education: Lessons of Experience* (World Bank) 4:590, 4:593*t*
- Higher Education Opportunity Act (2008) 8:119
- Highlander Folk School 1:55, 1:77–78, 1:165
- high school dropouts 2:325–330
 - background information 2:325
 - causal factors
 - general discussion 2:326
 - individual factors 2:327
 - institutional factors
 - community influences 2:328
 - family issues 2:327
 - general discussion 2:327
 - peer influences 2:328
 - school characteristics 2:327
 - consequences 2:326
 - dropout rates 1:874, 1:875, 2:325
 - general discussion 2:329
 - intervention strategies
 - challenges 2:328
 - programmatic approaches 2:329
 - Scholar–Baller[®] program 1:863
 - systemic solutions 2:329
 - racial and ethnic demographics 1:874
- High/Scope program *see* Perry Preschool Program study
- high-stakes evaluations 3:675
- high-stakes testing
 - background information 4:186
 - historical background 4:187
 - records of achievement 3:248
 - score inflation 4:186–192
 - evidence 4:186, 4:187*f*, 4:187*t*
 - historical background 4:187
 - inadequate tradition approaches 4:187
 - validity of inferences under high-stakes conditions (VIHS)
 - behavioral responses 4:189
 - conceptual elements 4:190, 4:190*f*
 - effective test weight 4:189
 - evaluation frameworks 4:188
 - future directions 4:191
 - general discussion 4:188
 - target of inference 4:189, 4:189*f*
- high-technology assistive devices 2:644
- Hill, Patty Smith 1:515
- Hine, Lewis 6:499
- Hirsch, ED, Jr 6:64
- Hirst, Paul 6:5, 6:103, 6:109, 6:187
- Hispanic population 1:610–615, 1:687
 - college-choice model
 - basic concepts 1:611, 1:612*f*
 - chain migration theory 1:613
 - decision-making models 1:613*f*
 - quantitative methods 1:612
 - social capital theory 1:613
- cultural competence 2:565–570
 - basic concepts 2:566
 - community involvement 2:568
 - importance 2:569
 - overrepresentation 2:565
 - Pedersen's model 2:566
 - research background 2:565
- cultural difference theory model 2:776
- deaf-blind population 2:623*t*
- demographic data 1:610
- educational attainment 1:610
- educational statistics 1:581–588
 - choice effects 1:585
 - college enrollment opportunities 1:656, 1:658*f*
 - continuing disparities 1:584
 - demographic diversity 1:582
 - digital technology implications 1:584
 - disparity measures 1:581, 1:874
 - dropout rates 1:874, 1:875
 - education beyond borders 1:586
 - 18–24-year-olds demographics 1:802, 1:803*f*
 - enrollment rates 1:800, 1:801*f*, 1:802*f*, 1:803*f*
 - global village concept 1:587
 - graduation rates 1:874
 - incarceration rates 1:874
 - twenty-first century perspectives 1:581, 1:587
 - United States 1:582
- information and communication technology (ICT) access and equity
 - home access 8:95, 8:96*t*
 - home use 8:95
 - outcomes analyses 8:95, 8:97
 - school access 8:96, 8:96*f*
 - school use 8:96
- learning opportunities 1:610
- oppositional identity 1:756, 2:776
- race-specific consequences 1:876
- traumatic brain injury (TBI) 2:695
- valued peers 1:757
- historically black colleges and universities (HBCUs) 1:767–768, 1:769, 4:479–480
- historical organizational case studies 6:457
- history education
 - assessment strategies 3:323–329
 - achievement tests 4:175–180
 - research background 4:175, 4:657
 - standards movement 4:177
 - testing challenges 4:179
 - Center on Organization and Restructuring Schools (CORS) standards 3:325, 3:326*t*
 - challenges 3:326
 - classroom-based assessments 3:325, 3:325*t*
 - document-based question (DBQ) assessment 3:324
 - general discussion 3:323, 4:657
 - large-scale assessments 3:323
- HIV/AIDS
 - community education 1:123, 1:199
 - education–health association 1:602
 - Ghana 5:592*t*
 - International Institute for Educational Planning (IIEP) 4:690
 - low-literacy teaching module
 - AIDS module 1:816
 - background information 1:814
 - HIV module
 - AIDS module 1:816
 - FACTS (prevention) 1:816

- HIV/AIDS (*continued*)
 LUCK (stigma reduction) 1:817
 science lesson 1:815, 1:815^f
 transmission pathways 1:815
 teach-back methodology
 community activism 1:819
 educational impacts 1:821
 pedagogy of action 1:819, 1:821
 pedagogy of empowerment 1:817
 preparation for the field 1:818
 student transformations 1:819
 University of Zululand 1:820
 University of Zululand
 new teachers 1:820
 pedagogy of action 1:819
 student perspectives 1:820
- Malawi 5:666
 postconflict societies 8:444
 prevention programs **1:814–822**
 background information 1:814
 low-literacy teaching module 1:814
 teach-back methodology
 community activism 1:819
 educational impacts 1:821
 pedagogy of action 1:819, 1:821
 pedagogy of empowerment 1:817
 preparation for the field 1:818
 student transformations 1:819
 University of Zululand 1:820
 University of Zululand
 new teachers 1:820
 pedagogy of action 1:819
 student perspectives 1:820
- Rwanda 5:765
 school-based services 2:637
 school health education programs 1:548
 Southern African Development Community (SADC) 1:200
- Hobbes, Thomas 6:74, 6:121, 8:441
 Hoeffding's inequality 7:428
 Hölder's inequality 7:428
 Hollingworth, Leta 2:870
 Holt, John 6:214, 6:215–216
 Home Instruction Program for Preschool Youngsters (HIPPY) 2:98, 2:99, 2:129
 home schooling **6:214–220**
 characteristics 8:184
 funding considerations 2:337
 general discussion 6:214, 6:219
 growth factors 6:214, 6:215^t
 justifications 6:215^t
 parental choice 6:214
 political perspectives 6:214
 research areas 6:218
 teaching and learning strategies
 age considerations 6:217
 independent learning 6:218
 learning styles 6:217
 parental involvement 6:217
 social interactions 6:218
 traditional versus nontraditional approaches 2:337, 6:214
 home-visiting programs (Jamaica) 2:127
 homework 1:350
- Honduras
 assessment system development 3:482
 conditional cash transfer (CCT) programs 2:419
 early childhood education and care (ECEC) 2:111^t
 earnings–education level relationship 2:306^t
 higher education enrollment 1:702^f, 1:703^f
 public subsidization 2:308^t
 returns to education–education level 2:308^t
 school-based management approaches 2:508
- Hong Kong
 classroom assessment **3:438–442**
 background information 3:438
 general discussion 3:441
 primary schools
 basic competency assessments 3:439, 3:440^t
 characteristics 3:438
 learning-to-learn reforms 3:439
 target-oriented curriculum (TOC) reform 3:438
 secondary schools
 characteristics 3:440
 school-based assessments 3:441
 washback studies 3:440
 class size 6:200
 ethical practices 4:720
 inequality–education relationship
 immigrant status 2:275^f
 test score–social background relationship 2:273^f
 information and communication technology (ICT)
 computer access at home 8:25^t
 computer access in school 8:23^t
 mathematics and science instruction 8:32^t
 leadership development programs 4:753^t, 4:754^t, 4:755^t, 4:758^t
 mathematics curricula 1:454
 moral and social reasoning 5:282
 research-quality assessments 4:310
 school supply shortages 2:506^f
 student achievement test performance 2:247^f
 teacher education programs 7:529
 teacher evaluations 3:677
- Hong Kong Certificate of Education Examination (HKCEE) 3:440
 Hong Kong Examinations and Assessment Authority (HKEAA) 3:440
 Hooks, Bell 1:824, 6:58
 Horace Mann–Lincoln Institute for School Experimentation 1:357–358
 Horkheimer, Max 6:133, 6:349
 Horton, Myles 1:55, 1:165
 Horvitz–Thompson (HT) estimator 7:393
 Hosmer–Lemeshow GOF statistic 7:195
 hot-deck imputation procedures 7:284
 Hotelling–Lawley trace 7:327
 Hotelling's T^2 -test 7:337
 Houston, Charles Hamilton 6:499
 Hout, Michael 1:543
 Hull House 1:55
 human capital theory **2:282–287**
 academic mobility 4:564
 adult education and training (AET) 1:7, 1:217, 1:223, 8:255, 8:256^f
 apprenticeships 8:313
 basic concepts 2:282
 career guidance 8:269
 displaced workers 8:276
 early childhood education and care (ECEC) 2:70
 earnings differentials 2:277, 2:290
 economic perspectives 2:194
 educational barriers 1:237
 empirical analyses
 methodologies 2:283
 nonmonetary returns 2:285
 on-the-job training 2:285
 post-secondary education 2:284
 primary and secondary education 2:284
 work experience 2:285
 general discussion 2:286
 globalization effects 4:203
 graduate employment **4:354–359**
 differentiation considerations 4:356
 division of labor 4:357
 employability enhancements 4:356
 employment rates and earnings 4:354
 expansion effects 4:354
 globalization 4:358
 job quality 4:355
 knowledge society theory 4:358
 labor market 4:366
 research biases 4:359
 student demands 4:354
 transition periods 4:357
 worker productivity 4:355
 immigrant education **8:280–288**
 immigrant characteristics 8:280
 investment/noninvestment determinants 8:280
 pay-offs 8:286
 regular education participation 8:284^f
 research areas 8:287
 targeting programs 8:285
 VET participation research 8:282, 8:283^f, 8:284^f, 8:285^f
 knowledge-based economies 1:175, 1:223
 older learners 1:175
 post-secondary education 2:284
 research background 2:282
 returns to education 2:282
 school-to-work transitions 8:307

- Singapore 8:66, 8:71, 8:71*t*
 social returns 2:285
 vocational education and training (VET) 8:236–237
see also signaling model
- humanist teaching perspective 1:18, 1:307
- human language characteristics 5:375
- human rights
 background information 1:287
 children's rights 1:290, 2:73, 6:17
 curriculum development 1:287–292
 educational rights
 abuse prevention 1:290
 acceptability 1:289
 accessibility 1:289
 adaptability 1:289
 availability 1:289
 children's rights 1:290, 6:17
 guiding principles 1:291
 parent's rights 1:290, 6:19
 teacher's rights 1:290
 Universal Declaration of Human Rights (UDHR) 1:288
 guiding principles
 learning about human rights 1:291
 learning cycle 1:291, 1:291*f*
 learning for human rights 1:291
 learning through human rights 1:291
 moral considerations 6:20
 parents' rights 1:290, 6:19
 peace education 1:852
 special needs students 2:688
- Hume, D. 6:123
- Hungary
 age-related policies and programs 8:263*t*
 childhood brain development 2:57*f*
 class size 2:492*f*
 early childhood education and care (ECEC) 2:4*t*, 2:56*t*, 2:111*t*, 2:114*t*
 earnings–education level relationship 2:307*t*
 education levels 2:279*f*
 higher education
 foreign doctoral students 4:292*t*
 institutional expenditures 1:733*t*
 loans and grants 4:501*f*, 4:502*f*
 parental expenditures 1:734*t*
 student and researcher mobility 4:287*f*
 tertiary education expenditures 4:500*f*
 tertiary type-B program survival rates 8:464*t*
 tuition fees 4:500*f*
 inequality–education relationship
 employment rate–educational attainment relationship 2:278*f*
 test score–social background relationship 2:273*f*
 information and communication technology (ICT)
 computer access at home 8:25*t*
 computer access in school 8:23*t*
 frequency of computer use in school and at home 8:26*t*
 student proficiency and confidence 8:31*t*
 information and communication technology (ICT) access and equity 8:97
 returns to education–education level 2:308*t*
 school choice 6:285
 self-managing schools 5:72–73
 teacher shortages 2:494*f*
 teaching hours 2:493*f*
 vocational education and training (VET) 8:249, 8:251*t*, 8:252*f*, 8:253*f*, 8:263*t*
- Husserl, Edmund 6:80, 6:184–185, 6:450
- Hymes, Dell 5:382–383
- hypergeometric distribution 7:133
- hypertension *see* health considerations
- hypothesis testing 7:216–222
 analysis of covariance (ANCOVA) 7:21
 Bayesian statistical analysis 7:42
 confidence intervals 7:220, 7:221*f*
 general discussion 7:216
 inferential statistics 7:217
 multivariate analysis of variance (MANOVA) 7:319
 multivariate linear regression 7:327
 order statistics 7:360, 7:362
 random assignment approximations 3:106
 research questions 7:216
 samples 7:217
 sampling distribution
 basic concepts 7:217
 central limit theorem 7:218, 7:218*f*
 probability distributions 7:218
 standard error 7:217
 statistical hypothesis
 directional and nondirectional alternative hypotheses 7:216
 null hypothesis 7:216, 7:436
 statistical population 7:217
 statistical power analysis 7:436
 test statistics
 decision rules 7:219, 7:220*f*
 probability estimates 7:218
t-distributions 7:218, 7:219*f*
 Type I/Type II statistical errors 7:221, 7:221*f*
- I**
- Iceland 5:605–610
 background information 5:605
 career guidance 8:272
 childhood brain development 2:57*f*
 class size 2:492*f*
 early childhood education and care (ECEC) 2:4*t*, 2:13, 2:56*t*, 2:105, 2:111*t*
 education levels 2:279*f*
 education system
 administrative and supervisory structure 5:609
 challenges 5:610
 educational finance 5:609
 goals 5:606
 monitoring and evaluation programs 5:609
 research programs 5:609
 structural and operational characteristics
 adult education 5:608
 compulsory education 5:606, 5:607*f*
 continuing education 5:608
 early childhood education and care (ECEC) 5:606, 5:607*f*
 higher education 5:607*f*, 5:608
 lower secondary education 5:606, 5:607*f*
 pre-primary school 5:606, 5:607*f*
 primary education 5:606, 5:607*f*
 schematic diagram 5:607*f*
 students and faculty 5:607*t*
 upper secondary schools 5:606, 5:607*f*
 teaching profession 5:608
 enrollment rates 2:493*f*
 foreign doctoral students 4:292*t*
 gender-equity gap 2:56*t*
 higher education
 institutional expenditures 1:733*t*
 loans and grants 4:501*f*, 4:502*f*
 parental expenditures 1:734*t*
 tertiary education expenditures 4:500*f*
 tertiary type-B program survival rates 8:464*t*
 tuition fees 4:500*f*
 inequality–education relationship
 employment rate–educational attainment relationship 2:278*f*
 test score–social background relationship 2:273*f*
 information and communication technology (ICT)
 computer access in school 8:23*t*
 frequency of computer use in school and at home 8:26*t*
 student proficiency and confidence 8:31*t*
 student achievement test performance 2:247*f*
 teaching hours 2:493*f*
 vocational education and training (VET) 8:251*t*, 8:252*f*, 8:253*f*, 8:272
 women in employment 2:56*f*
- Idaho 1:654*f*
- identified regulation 6:620
- identity 6:152–157
 basic concepts 6:153
 cultural identity 6:155
 essentialist perspectives 6:152
 historical perspectives 6:152
 identity formation 6:154
 identity politics 1:86, 6:39, 6:152
 philosophical perspectives 6:155
 postmodern perspectives 6:152
 workplace learning 1:85
 ignorable assignment mechanisms 7:69
 ill-defined problems 5:273
- Illinois
 Asian American/Pacific Islander student enrollment 1:597*f*
 tuition and needs-based grants 1:654*f*

- illiteracy in India 1:592, 1:592*t*
- image-based education research **6:498–504**
 - background information **6:498**
 - current research contributions **6:501**
 - research background
 - digital technology **6:500**
 - field studies and experiments **6:499**
 - Hine's photography **6:499**
 - Houston's contribution **6:499**
 - image-elicitation techniques **6:500**
 - microethnography studies **6:500**
 - significance **6:503**
- immediate feedback **3:226–227**
- immersive environments and games **4:422, 8:7**
- immigrant education **1:113–119**
 - background information **1:113**
 - basic concepts **1:115**
 - cultural difference theory model **2:776**
 - educational history **1:882**
 - educational programs **1:116**
 - future directions **1:118**
- host country vocational education and training **8:280–288**
 - immigrant characteristics **8:280**
 - investment/noninvestment determinants **8:280**
 - pay-offs **8:286**
 - regular education participation **8:284*f***
 - research areas **8:287**
 - targeting programs **8:285**
 - VET participation research **8:282, 8:283*f*, 8:284*f*, 8:285*f***
- intercultural education **1:117**
- multicultural education **1:117**
- sustainable development **6:293**
- see also* adult education and training (AET)
- immigrant equity policy **1:157*t*, 1:158**
- immigrant populations
 - children and young adults **4:629–635**
 - family background
 - parental education **4:630**
 - poverty **4:630**
 - seasonal migrant work **4:631**
 - undocumented status **4:631**
 - global migration **4:629**
 - individual-level factors **4:632**
 - school considerations
 - second-language learning **4:631**
 - segregation **4:631**
 - social support
 - community support **4:633**
 - general discussion **4:633**
 - mentoring relationships **4:633**
 - motivation **4:633**
 - peer relationships **4:634**
 - inequality–education relationship **2:274, 2:275*f***
 - vocational education and training (VET) **8:235, 8:412**
- impasses **2:513**
- imperfect states **4:514**
- implicit cognition **7:750*t***
- inauthentic leadership **5:10**
- incentive programs **2:481–488**
 - advantages
 - group incentives
 - efficiency and productivity gains **2:483**
 - research findings **2:485*t***
 - individual incentives
 - efficiency and productivity gains **2:482**
 - recruitment and retention **2:483**
 - research findings **2:485*t***
 - background information **2:481**
 - developing countries **2:507**
 - disadvantages **2:484**
 - general discussion **2:486**
 - rationale **2:481**
 - unintended consequences **2:486**
- inclusive education **2:721–726, 4:846–851, 6:158–163**
 - advantages/disadvantages **2:722**
 - cultural influences **4:848**
 - definitions **2:721**
 - educational impacts
 - family members **2:724**
 - general discussion **2:723**
 - students with and without disabilities **2:723**
 - teachers and service providers **2:724**
 - future directions **4:850**
 - general discussion **2:725, 4:846**
 - historical perspectives
 - feeling included versus feeling excluded **6:160**
 - general discussion **2:721**
 - inclusion versus integration **4:846, 6:158**
 - mainstream provisions versus special provisions **6:160**
 - medical versus social model of disability **6:159**
 - segregation versus integration **6:158**
 - implementation strategies **2:723, 4:846**
 - inclusive classrooms **2:728**
 - international perspectives **2:724**
 - leadership strategies **4:849**
 - least restrictive environments **2:718, 2:721**
 - networking and collaboration **4:849**
 - organizational conditions **4:847**
 - philosophical perspectives **6:161**
 - policies and programs **4:846**
- income contingent loans (ICLs)
 - Australia **4:504**
 - background information **4:504**
 - New Zealand **4:504**
 - South Africa **4:505**
 - Sweden **4:504**
 - United Kingdom **4:505**
 - United States **4:505**
 - Yale Plan **4:504**
- income taxes **2:224**
- Incredible Years program
 - basic concepts **2:129**
 - United States **2:129**
 - Wales **2:129**
- incubators **4:303**
- independent institutions **4:512**
- index of dependability (Φ) **4:65**
- India **1:589–594**
 - accreditation processes **4:384–389**
 - adult education **1:80**
 - caste system
 - basic concepts **1:589**
 - categories **1:590*t***
 - dropout rates **1:592, 1:593*t***
 - educational policies **1:591**
 - enrollment **1:593*t***
 - future directions **1:593**
 - literacy rates **1:592, 1:592*t***
 - origins **1:589**
 - demographic data **1:591**
- early childhood education and care (ECEC) **2:111*t***
- higher education
 - collective bargaining **4:494**
 - English-delivered instruction **4:205**
 - enrollment rates **4:521*t*, 4:522*t***
 - faculty mobility **4:562**
 - government policies and programs **4:521**
 - institutional expenditures **1:733*t***
 - literacy rates **4:521*t***
 - parental expenditures **1:734*t***
 - private higher education (PHE) **4:623**
 - public expenditures **4:521*t***
 - quality assurance **4:525**
 - statistical data **4:522*t***
 - student mobility **4:561**
 - tuition and fees **4:523**
- indigenous curricula
 - Gandhian model **1:307**
 - general discussion **1:307**
 - national aims **1:310**
 - national–local educational dilemma **1:308**
 - relevance–quality–efficiency factors **1:309**
 - sociopolitical factors **1:310**
 - vocational education and training (VET) **1:309**
- information and communication technology (ICT) access and equity **8:98**
- moral and social reasoning **5:282**
- National University of Educational Planning and Administration (NUEPA, India) **4:676–680**
 - challenges **4:679**
 - characteristics **4:677**
 - functional role **4:677–678**
 - funding sources **4:678**

- future directions 4:679
- governance 4:678
- historical background 4:676
- International Institute for Educational Planning (IIEP) 4:691
- UNESCO (United Nations Educational, Scientific and Cultural Organization) 4:676
- preservice education programs 8:75
- school choice 6:285–286
- STEM (science, technology, engineering, and mathematics) 1:838
- student achievement test performance 2:247^f
- teacher incentives 2:482, 2:485^t, 2:508
- vocational education and training (VET) 8:247
- Indiana 1:654^f
- indigenous education 1:661–666
 - challenges 1:665
 - current status 1:663
 - curriculum 1:307–311
 - Convention on the Rights of the Child 1:307
 - educational traditions 1:307
 - Gandhian model 1:307
 - national aims 1:310
 - national–local educational dilemma 1:308
 - relevance–quality–efficiency factors 1:309
 - sociopolitical factors 1:310
 - vocational education and training (VET) 1:309
 - future directions 1:665
 - historical policies and practices 1:662
 - indigenous knowledge 1:661
 - indigenous peoples 1:661
 - language programs 1:295
 - post-secondary education 1:663
 - self-determination 1:663
- indigenous peoples
 - assimilation policies 1:575
 - characteristics 1:661, 8:289
 - cultural difference theory model 2:776
 - self-determination 1:663
 - vocational education and training (VET) 8:289–293
 - challenges 8:235, 8:290
 - First Nations Technical Institute (Canada) 8:290^t
 - instructional strategies 8:292
 - Māori culture 8:291^t
 - outcomes analyses 8:293^t
 - program models 8:290
 - Sámi culture 8:289^t
 - socioeconomic development 8:289
- individual bilingualism 5:355
- individual differences (ID) MDS model 7:307, 7:308^t, 7:309^f
- individual interest 6:613
- individualism 6:88
- individualized education plans (IEPs)
 - adapted physical activity programs 2:520
 - adolescence to adulthood transition services 2:876
 - assessment accommodations 2:534
 - early intervention to preschool transition 2:602
 - functional behavior assessment (FBA) 2:708–709
 - historical background 2:689
 - kindergarten transitions 2:618
 - language development 2:748
 - large-scale assessments 2:753
 - related services 2:834
 - school-based services 2:634–635, 2:850–851
 - speech and language impairments 2:658, 2:658^f
 - traumatic brain injury (TBI) 2:697
- individualized family service plans (IFSPs) 2:597, 2:615, 2:634–635, 2:689, 2:850–851
- individualized instructional programs
 - curriculum-based assessment (CBA) 4:108
 - pre-referral intervention strategies 2:817
- individualized learning 8:396
- Individuals with Disabilities Education Act (IDEA, 1990)
 - adolescence to adulthood transition services 2:876
 - assessment accommodations 2:530
 - assistive technology (AT) 2:673
 - autism spectrum disorders (ASD) 2:661–662
 - community-based experiences 2:552
 - culturally and linguistically diverse (CLD) learners 2:552, 2:657
 - diagnostic and classification processes 3:493
 - direct instruction (DI) 2:585
 - emotional and behavioral disorders (EBD) 2:678, 3:260
 - family involvement 2:790
 - historical background 2:596, 2:689, 2:717
 - inclusion practices 2:721
 - instructional accommodations 2:727
 - intervention recipients 2:597
 - key components 2:597^t
 - language development 2:747
 - paraprofessionals 2:780
 - physically disabled students 2:520
 - post-secondary education 2:808
 - pre-referral intervention strategies 2:816
 - school-based services 2:634, 2:851–852
 - speech and language impairments 2:657
 - student-to-teacher ratio reductions 2:475
 - traumatic brain injury (TBI) 2:695
 - twice-exceptional students 2:591
 - visual impairment 2:880
- Individuals with Disabilities Education Improvement Act (IDEIA, 2004)
 - adolescence to adulthood transition services 2:876
 - assistive technology (AT) 2:537–538
 - community-based experiences 2:552
 - educational effectiveness 2:802
 - historical background 2:689
 - inclusion practices 2:722
 - instructional accommodations 2:727
 - paraprofessionals 2:780
 - post-secondary education 2:808
 - purpose 2:845
 - related services 2:834–839
 - general discussion 2:839
 - physically disabled students 2:640
 - related services providers
 - characteristics 2:834
 - consultative services 2:835
 - direct services 2:835
 - integrated services 2:835
 - service provision recommendations 2:838
 - service types
 - art therapy 2:641
 - audiology services 2:835
 - counseling services 2:836
 - early identification and assessment services 2:836
 - general discussion 2:835
 - medical services 2:640, 2:836
 - occupational therapy 2:640, 2:836
 - orientation and mobility services 2:640, 2:836
 - parent counseling services 2:837
 - physical therapy 2:640, 2:837
 - psychological services 2:837
 - recreational therapy 2:837
 - rehabilitation counseling 2:837
 - school health/school nurse services 2:837
 - social work services 2:838
 - speech/language pathology services 2:640, 2:838
 - transportation services 2:640, 2:838
 - school-based services 2:634–635, 2:850–851
 - social inclusion 2:797
- Indonesia 5:611–617
 - background information 5:611
 - early childhood education and care (ECEC) 2:111^t
 - education system
 - challenges 5:616
 - demographic data 5:612^t
 - educational finance 5:615
 - educational streaming 5:613
 - Education Law No. 20/2003 5:611, 5:617
 - goals 5:611
 - higher education 5:615, 5:615^t
 - Ministry of National Education (MONE) 5:611
 - national curriculum guidelines 5:614
 - reform efforts 5:617
 - structural and operational characteristics 5:612^f
 - teacher education 5:614, 5:614^t
 - higher education
 - characteristics 5:613, 5:615
 - collective bargaining 4:495
 - institutional expenditures 1:733^t
 - number of lecturers 5:615^t
 - parental expenditures 1:734^t
 - structural and operational characteristics 5:612^f
 - inequality–education relationship 2:273^f
 - information and communication technology (ICT)

- Indonesia (*continued*)
 computer access at home 8:25*t*
 computer access in school 8:23*t*
 student achievement test performance 2:247*f*, 2:417*t*
 test score–social background relationship 2:273*f*
- induction programs 7:563–568
 challenges 7:566
 characteristics 7:565
 critical elements 7:566
 general discussion 7:563
 goals 7:564
 historical background 7:563
 importance 7:567
 mentoring programs 2:477, 7:566
 recruitment and retention 2:477
 supervision 7:616
- inductive reasoning 3:222, 5:273
- ineffective schools 5:55
- inequality–education relationship 2:272–281
 economic outcomes
 causal impacts 2:277, 2:279*t*
 earnings differentials 2:276, 2:277*t*
 education-related benefits 2:278
 employment rate–educational attainment relationship 2:278*f*
 human capital theory 2:277
 international education levels 2:279*f*
 skill-biased technical change (SBTC) theory 2:280
 supply-and-demand effects 2:279*f*, 2:280
- educational outcomes
 fathers' educational background 2:274*f*
 gender differences 2:275, 2:276*f*, 2:277*t*
 parenting styles 2:153
 race/ethnicity/immigrant status 2:274, 2:275*f*, 2:432
 social background 2:273, 2:273*f*
- general discussion 2:272, 2:280
- infants and toddlers
 attention 5:165–171
 attention–learning models
 compensatory systems 5:170
 directed-attention model 5:169
 dyadic attention 5:169
 intentional understanding 5:170
 plasticity 5:170
 triadic attention/joint attention 5:170
 behavioral phases 5:166, 5:167*f*
 brain networks 5:167
 face processing 5:168
 general discussion 5:165, 5:170
 newborn systems 5:165
 novelty factors 5:168
 selective attention development 5:166, 5:167*f*
 visual recognition memory 5:168
- early childhood education and care (ECEC) 2:69–74
 child development research 2:69
 Early Learning Standards (ELS) 2:132–137
 content resources 2:134
 development process 2:133
 document content 2:134
 general discussion 2:132
 historical perspectives 2:132
 implementation processes 2:135
 importance 2:136
 infant–toddler content 2:135
 preschool content 2:135
 United States 2:133
 family pattern changes 2:70–71
 general discussion 2:73
 government policies and programs 2:71–72
 grandparent care 2:71
 health-focused models 2:70
 human capital theory 2:70
 participatory models 2:70
 Piagetian perspective 2:72
 program criticisms 2:70
 research trends 2:71
 theoretical perspectives 2:72
 Vygotskian perspective 2:72
- early intervention services 2:608–612
 benefits 2:608
 collaborative environments 2:611
 deaf-blind children 2:622, 2:625
 early childhood special education (ECSE) 2:594–607
 basic concepts 2:595
 birth to 8 years 2:594
 center-based programs 2:600
 combination programs 2:600
 demographic data 2:598*t*
 early intervention to preschool transition 2:602
 family involvement 2:601
 future directions 2:604
 government legislation 2:596, 2:597*t*
 home-based programs 2:599
 intervention recipients 2:597
 intervention services 2:595
 itinerant professionals 2:602
 personnel preparedness 2:601
 preschool to kindergarten transition 2:603
 service delivery 2:598
 service locations 2:599*t*
 service types 2:599*t*
 transition processes 2:602
 general discussion 2:608
 importance 2:608
 limitations 2:609
 Milwaukee Project 2:609
 teacher preparedness 2:610
- empathic responses 6:599
- feeding and eating disorders 2:581
- first language learning
 early word production and comprehension 5:377, 5:378*f*
 infant babble 5:377
 morpheme acquisition order 5:378*t*
 pre-linguistic achievements 5:376
 Wug test 5:378*f*
- language development 2:747
- mortality rates 1:602, 1:607*t*
- music learning 5:407
- parent–child relationships 5:450–456
 autonomy support 5:451
 early interactions 5:450
 learning and language participation 5:451
 out-of-home learning 5:453
 warmth and sensitivity 5:451
- infectious diseases 1:602
- inferential statistics 7:217
- informal learning 1:42–46, 8:240–246
 ageless learners 8:241*f*
 assessment strategies 8:244
 basic concepts 1:3–4
 benefits 8:245
 characteristics 1:242
 contestation issues 1:46
 definitions 1:42, 5:457, 8:240
 demographic data 1:243
 educational environments 1:44
 elder learning 1:176
 evaluation policies and practices 3:572–577
 background information 3:572
 challenges 3:575
 changing focus 3:574
 evaluation approaches
 dialogical evaluation 3:575
 directed evaluation 3:574
 negotiated evaluation 3:574
 general discussion 3:576
 research background 3:573
- formal–informal learning distinction 1:43
- formalization processes 1:45
- general discussion 8:240
- idealized characteristics 5:462*t*
- Indonesia 5:612*f*, 5:613
- international trends 8:241
- learner's perspective considerations 8:244
- Lithuania 5:661
- nonformal learning comparisons 1:42
- play 5:330–334
 first language learning 5:333
 intent participation 5:331
 participation research 5:332
 peer learning 5:331
 Piagetian perspective 5:330
 play communities 5:332

- research background 5:330
- threats/teasing/joking/play politics 5:332
- Vygotskian perspective 5:331
- private-sector training 1:207
- research background 1:42
- schematic diagram 8:242*f*
- theoretical perspectives 1:43
- vocational education and training (VET) 8:244, 8:438
- workplace learning 1:45, 8:243, 8:329, 8:329*f*
- see also* lifelong learning
- informal reading inventories (IRIs) 3:191
- informal reasoning 5:248
- informational barriers 1:234
- information and communication technology (ICT) 8:1–10
 - academic values and expectations 4:436
 - access and equity patterns 8:95–101
 - Canada 8:97
 - China 8:98
 - European Union (EU) 8:97
 - general discussion 8:95
 - global access and use 8:98, 8:98*t*
 - improvement efforts 8:99
 - India 8:98
 - international trends 8:21–33
 - challenges 8:30
 - computer access at home 8:25, 8:25*t*
 - computer access in school 8:22, 8:23*t*
 - frequency of use at home 8:25, 8:26*t*
 - frequency of use in school 8:25, 8:26*t*
 - large-scale assessments 8:22
 - mathematics and science instruction 8:30, 8:32*t*
 - patterns of use 8:26, 8:27*f*
 - student access 8:21
 - student proficiency and confidence 8:27, 8:31*t*
 - theoretical perspectives 8:21
 - United States
 - home access 8:95, 8:96*t*
 - home use 8:95
 - outcomes analyses 8:95, 8:97
 - school access 8:96, 8:96*f*
 - school use 8:96
- belief systems 7:623, 7:627
- characteristics 2:537, 2:538*t*
- computerized adaptive testing (CAT)
 - background information 7:396
 - basic concepts 7:397
 - dynamic assessments 3:222–223
 - item response theory (IRT) applications 4:87
 - purpose 4:148
 - sampling methods 7:395
 - sequential probability ratio test (SPRT) 7:397, 7:398*t*
- computer-supported collaborative learning (CSCL) 5:290–296
 - background information 5:291
 - design-based research
 - pedagogical design 5:293
 - technology design 5:293
 - future directions 5:294
 - research approaches
 - dialogic approach 5:292
 - systemic approach 5:291
 - sociocultural perspectives 5:290
- economic returns 8:81–87
 - economic growth
 - knowledge–acquisition model 8:82
 - knowledge–creation model 8:83
 - knowledge–deepening model 8:82
 - research perspectives 8:81
 - sources 8:82
 - economic rationale 8:81
 - Finland 8:84
 - policy implications 8:85
 - Singapore 8:81, 8:83
- educational applications
 - educational impacts 8:2
 - effectiveness 8:7
 - emerging uses
 - general discussion 8:6
 - immersive environments and games 4:422, 8:7
 - interactive classroom communication systems 8:7
 - social networks 4:422, 5:120, 5:121*f*, 8:6
 - Web 2.0 4:431, 8:6
 - general discussion 8:9
 - implementation strategies 8:8
 - out-of-school learning 8:9
 - rapidly growing uses
 - assessment and accountability systems 8:6
 - online courses 8:5
 - open-source online resources 8:5
 - virtual schools 8:5
 - subject-specific uses
 - language instruction 8:4
 - mathematics instruction 8:4
 - reading and literacy 8:3
 - science instruction 8:4
 - writing instruction 8:3
 - typical uses 8:2
- educational leadership 5:120–127
- educational impacts
 - community relations 5:121, 5:121*f*
 - general discussion 5:120
 - instructional strategies 5:121, 5:121*f*
 - literacy instruction 5:121, 5:121*f*
 - social networks 5:120, 5:121*f*
- influencing factors 5:125, 5:125*f*
- integration framework 5:123*f*
- requirements and standards 5:121
- social and legal challenges 5:124
- teacher leadership 5:122, 5:123*f*
- technology-mediated leadership development programs 5:123, 5:123*t*
- elementary/middle-grades reading curricula 1:466
- flexible learning
 - basic concepts 4:442
 - delivery challenges 4:444
 - delivery effectiveness 4:443
- general discussion 8:1–2
- group work assessments 3:346
- growth factors 8:2
- higher education 4:419–426
 - assessment strategies 3:253
 - background information 4:419
 - curriculum 4:420
 - educational innovations 4:453
- effective learning
 - collaborative learning 4:422
 - communication-based learning 4:422
 - construction-based learning 4:421
 - instructional strategies 4:421
- e-learning pedagogies
 - average comparative ratings 1:531*t*
 - basic concepts 1:522
 - content appropriateness and relevancy 1:522
 - course goals 1:523
 - dimensions analysis 1:532
 - diversity considerations 1:523
 - effectiveness 4:423, 4:423*t*
 - interactive learning strategies 1:524
 - learners' needs 1:523
 - learning styles 1:522
 - multimedia 1:524
- flexible study 4:423
- future directions 4:425
- teaching approaches 4:424
- international access patterns 8:21–33
 - challenges 8:30
 - computer access at home 8:25, 8:25*t*
 - computer access in school 8:22, 8:23*t*
 - frequency of use at home 8:25, 8:26*t*
 - frequency of use in school 8:25, 8:26*t*
 - large-scale assessments 8:22
 - mathematics and science instruction 8:30, 8:32*t*
 - patterns of use 8:26, 8:27*f*
 - student access 8:21
 - student proficiency and confidence 8:27, 8:31*t*
 - theoretical perspectives 8:21
- International Association for the Evaluation of Educational Achievement (IEA)
 - studies 4:644–649
 - background information 4:644
- CompEd study
 - characteristics 3:687*t*
 - Functional Information Technology Test (FITT) 4:645
 - importance 4:649
 - purpose 4:644

- information and communication technology (ICT) (*continued*)
 Stage 1 survey results 4:644
 Stage 2 survey results 4:645
 importance 4:649
 Second Information Technology in Education Studies (SITES)
 basic concepts 4:646
 characteristics 3:687*t*
 educational impacts 4:667
 importance 4:649
 Lithuania 5:665
 Module 1 (SITES M1) survey results 4:646
 Module 2 (SITES M2) survey results 4:647, 8:224
 SITES 2006 4:648, 8:22
 knowledge-based economies 1:224
 mathematics education 3:281
 open and distance learning (ODL) 4:408, 8:129
 personalized learning 8:392
 preservice education programs 8:217–221
 challenges 8:220
 purpose 8:217
 teaching preparation strategies 8:218
 technology as an educational tool
 digital technology 8:219
 distance education 8:219
 electronic portfolios 8:219
 usage preparation strategies 8:217
 public–private partnerships (PPPs)
 advantages 8:74
 case studies
 critical analysis 8:78
 preservice teacher education programs 8:75
 School District of Philadelphia/School of the Future project 8:77
 State School Isaac Schraiber/reading and writing project 8:76
 critical analysis
 accountability 8:79
 holistic approaches 8:78
 pilot programs 8:78
 situation-and-needs-analysis phase 8:78
 stakeholders 8:79
 sustainability 8:79
 future directions 8:79
 Partners-In-Learning (PiL) initiative (Microsoft) 8:73, 8:74
 returns to education 8:81–87
 economic growth
 knowledge–acquisition model 8:82
 knowledge–creation model 8:83
 knowledge–deepening model 8:82
 research perspectives 8:81
 sources 8:82
 economic rationale 8:81
 Finland 8:84
 policy implications 8:85
 Singapore 8:81, 8:83
 school–community relations 8:125–132
 alternative schooling approaches
 accountability data sharing 8:128
 current trends 8:130
 distance education 8:129
 general discussion 8:128
 one-to-one computing 8:128
 virtual schools 8:129
 challenges 8:125
 importance 8:125
 research areas 8:130
 research background 8:126
 technology initiatives
 e-mail communication 8:126
 growth factors 8:128
 Internet-based approaches 8:126
 websites 8:127
 school-to-school collaboration 6:276
 secondary school science assessments 3:304, 3:305*f*
 self-adaptive testing 4:148–152
 assisted self-adaptive testing (ASAT) 4:149–150
 basic concepts 4:148, 4:149*f*
 benefits 4:148
 dynamic assessments 3:222–223
 general discussion 4:151
 item pool 4:149
 item-selection strategies 4:150, 4:150*f*
 perceived control/perceived performance hypotheses 4:148
 restricted self-adaptive testing (RSAT) 4:149
 theoretical investigations 4:150
 Singapore 8:66, 8:67*f*
 speech and language impaired students 2:658
 student leadership for learning 4:819
 Suriname 5:832
 teacher learning 8:211–216, 8:222–227
 certification and licensing 8:212
 change management
 educational innovations 8:224
 general discussion 8:224
 Second Information Technology in Education Studies (SITES) 8:224
 understanding/self-understanding factors 8:224
 competency standards 8:213
 curriculum frameworks 8:211
 functional role
 activity clusters 8:225*t*
 general discussion 8:225
 knowledge–innovation relationships 8:226*t*
 general discussion 8:225
 importance 8:211
 integration strategies 8:215
 knowledge–deepening versus knowledge–creation approach 8:214
 knowledge demands 8:222
 literacy skills development 8:212
 need requirements
 formal versus informal knowledge 8:223, 8:223*t*
 knowledge domains 8:222, 8:223*t*
 knowledge in practice 8:223, 8:223*t*
 knowledge levels 8:223, 8:223*t*
 socioeconomic skills 8:213
 standards 8:211
 technology education 3:311–315
 assessment models
 classroom interactions and feedback 3:313
 formative assessments 3:312
 functional role 3:314
 general discussion 3:311
 sociocultural perspectives 3:311
 summative assessments 3:314
 teacher's role 3:312
 characteristics 3:311
 general discussion 3:315
 technology literacy and fluency 8:34–41
 basic concepts 8:36
 educational implications
 curriculum changes 8:39
 digital divides 8:39
 literate lives 8:38
 performance assessments 8:39
 framework development 8:37, 8:38*f*, 8:39*f*
 future directions 8:40
 historical perspectives 8:35
 multiple literacies 8:37
 social implications 8:34
 transversal programme 4:673
 twenty-first century skills 8:158–166
 challenges 8:164
 framework development 8:159
 innovative instructional strategies
 basic concepts 8:160
 emerging media 8:161
 historical documents 8:164*f*
 multi-user virtual environment (MUVE) interfaces 8:161
 River City MUVE 8:162, 8:162*f*, 8:163*f*, 8:164*f*, 8:165*f*
 situated learning 8:161
 virtual microscopy 8:164*f*, 8:165*f*
 knowledge-based economies 8:158
 selected new literacies 8:159, 8:160*t*
 Information and Communication Technology (ICT) Assessment Tasks 8:52
 Information Commons 4:429
 information index 7:384
 information literacy
 Information Commons 4:429
 Learning Commons 4:430
 student support services
 delivery approaches 4:429
 historical background 4:428
 importance 4:428
 technology literacy and fluency 8:37
 informative feedback 3:563, 6:626

- informed consent 3:623, 6:33
 initiation–response–feedback (IRF) discourse 6:698
 innovation adoption theory 8:21
 Innovative Links Project 7:758
 inquiry-based learning 5:297–302
 background information 5:297, 6:90
 basic concepts 7:605
 challenges 5:300
 characteristics 7:604
 contemporary classroom practices 5:298
 inquiry cycles 5:299, 8:169–170
 professional inquiry–classroom inquiry comparisons 5:298
 professional learning communities 5:155
 science education 8:167–171
 background information 8:167
 basic principles 8:168
 hypothesis generation 8:168, 8:169
 inquiry cycles 8:169–170
 primary schools 3:293
 research areas 8:170
 scaffolding processes 8:169
 scientific literacy 8:167
 skill acquisition 8:169
 technology-enhanced learning (TEL) 8:168
 inquiry-oriented teacher education 7:604–609
 basic concepts 7:605
 challenges 7:608
 characteristics 7:604
 historical background 7:604
 program characteristics 7:605
 insight 5:276
 inspection time 5:186
 Institute for Retired Professionals 1:178
 institutional autonomy 4:341–346
 background information 4:341
 basic concepts 4:331
 definitions 4:341
 external boundaries 4:342
 general discussion 4:345
 historical background 4:329
 internal boundaries 4:343
 institutional barriers 1:234
 institutional effectiveness 3:172–180
 basic concepts 3:173
 challenges 3:179
 data collection
 educational and school process information 3:175
 general discussion 3:174
 student background characteristics 3:174
 student baselines 3:174
 student outcomes 3:174
 measurement approaches
 data requirements 3:175, 3:175^r
 statistical analysis techniques 3:175, 8:92
 value-added models (VAMs) 3:177, 3:177^f
 raw league tables 3:172
 reform efforts 3:172
 value-added models (VAMs)
 applications 3:178
 basic concepts 3:173
 research results 3:177, 3:177^f
 institutionalized cultural capital 6:100
 Institutional Management in Higher Education (IMHE) 4:588
 institutional racism 1:626, 1:741, 6:38
 institutional research funding 4:281
 Institution of Civil Engineers 3:45–46
 Institution of Electrical Engineers (IEE) 3:46
 Institution of Engineering and Technology 3:46
 Institution of Incorporated Engineers (IIE) 3:46
Instituts Universitaire de Technologie (IUTs) 4:348–349
 instructional assessment 3:397–402
 accountability assessments 3:398
 diagnostic assessments 3:207, 3:276, 3:397
 embedded assessments
 associationist assessments 3:400
 catechism form 3:400
 general discussion 3:400
 interpretive assessments 3:401, 3:401^r
 I–R–E sequence 3:400
 structured talk format 3:401, 3:401^r
 epistemological perspectives 3:399
 general discussion 3:397, 3:402
 instructional design 4:423^r
 instructional leadership 5:18–25
 basic concepts 5:19, 5:20^f
 empirical analyses
 general discussion 5:20
 research findings 5:21
 trends 4:742, 5:21
 historical background 5:18
 international trends 4:759
 shared leadership models 5:22
 student learning impacts 5:45
 instructional management systems (IMS) 8:133–139
 background information 8:133
 classroom network technologies
 effectiveness 8:137
 general discussion 8:136
 instructional strategies 8:137
 key features 8:136
 future directions 8:137
 student–response systems
 effectiveness 8:135
 general discussion 8:134
 key features 8:134
 research areas 8:134
 teaching models 8:134
 instructional planning 3:365–368
 current perspectives 3:365
 Dewey's perspective 3:365
 dialog 3:367
 general discussion 3:368
 gifted and talented students 7:733
 learning regulation 3:366
 practical guidelines 3:365
 zone of proximal development (ZPD) 3:367
 instructional technology (IT) 2:537, 2:538^f, 2:538^r
 instrumental case studies 6:458
 instrumental evaluations 3:637
 instrumental variables 7:223–228
 background information 7:223
 causal inference 7:6, 7:223
 economics of education 2:190
 educational applications
 class size
 demographic variations 7:227
 general discussion 7:227
 Maimonides' rule 7:227
 Project STAR (Student Teacher Achievement Ratio) 7:227
 returns to education
 basic concepts 7:226
 birth statistics 7:226
 compulsory school laws 7:227
 distance to school 7:227
 military service or benefits 7:227
 school system changes 7:227
 estimation methods
 conditional requirements 7:224, 7:225^f
 local average treatment effect (LATE) interpretation 7:226
 ordinary least squares (OLS) method 7:223
 specification tests 7:226
 state of identification 7:223
 weak instruments 7:225
 general discussion 7:227
 random assignment approximations 3:100
 integrated articulation 4:612
 integrated model of text and picture comprehension (IMTPC) 8:145, 8:146^f
 integrated regulation 6:620
 integrative teaching 3:67
 intellectual property rights 4:362
 intelligence 5:184–190
 ability testing 4:1–6
 classical test theory 4:2
 closed versus open tests 4:3
 consequences 4:5
 developed ability 4:3
 dyslexia 4:4
 general discussion 4:2, 4:5
 historical background 4:2
 measurement scales 4:2
 performance monitoring 4:4
 racial groups 4:3

- intelligence (*continued*)
 - school effectiveness 4:5
- developmental processes
 - background information 5:187
 - Piagetian perspective 5:187
 - Vygotskian perspective 5:187
- dynamic assessments 3:220–225
 - adaptive testing 3:222–223
 - advantages 3:220
 - basic concepts 3:207
 - feedback 3:221, 3:229
 - functional role 3:223
 - future directions 3:224
 - graduated-prompts approach 3:222
 - inductive reasoning 3:222
 - intervention strategies 3:221
 - short-term dynamic tests 3:222
 - standardized versus clinical approaches 3:221
- emotional intelligence 3:258, 5:186, 6:547
- ethnic minorities
 - discrimination 1:737, 1:756
 - entity versus incremental theories 1:755
- Flynn effect 4:3
- heritability 4:3, 5:188
- intelligence quotients (IQs)
 - assessment impacts 4:56–57
 - cognitive measures 2:578, 2:679
 - historical background 5:188
 - racially-based tests 1:558, 4:3
 - standard deviation 7:269
 - traumatic brain injury (TBI) 2:698
- intelligence tests
 - bell curve theory 1:558, 1:738, 7:471, 7:472*f*
 - convergent thinking problems 4:171
 - criticisms 3:220
 - diagnostic assessments 2:651, 3:397
 - experiential bias 4:172
 - giftedness 2:588, 2:870
 - historical background 4:7, 4:95, 5:188
 - Milwaukee Project 2:609
 - racially-based tests 1:558, 1:883
 - special needs students 2:588, 2:688, 2:716, 2:775
 - standardized versus clinical approaches 3:221
 - student achievement effects 5:52
- malleability 5:189
- measurement approaches
 - historical background 5:187
 - intelligence tests 5:188
 - statistical distribution 5:188
- multiple intelligences 4:2, 5:186
- older learners 1:178
- regression to the mean
 - body height analyses 7:433*f*, 7:434*f*
 - convergence trends 7:434
 - Galton's analyses 7:431, 7:433*f*
 - test results analysis 7:431, 7:432*f*
- socioeconomic status (SES) 5:189
- theoretical perspectives
 - biological theories 5:187
 - cognitive-contextual theories 5:186
 - implicit theories 5:184
 - psychometric theories 4:2, 5:185
- triarchic theory of successful intelligence 4:2, 5:186
- Intelligent Essay Assessor (IEA) 4:21, 8:51
- intelligent systems 4:75–80
- Intellimetric scoring program 4:21, 8:51
- intensive field experiences
 - basic concepts 1:714
 - curriculum development and logistics 1:716, 1:717*r*
 - goals and objectives 1:715
- intent participation 5:331
- interactional diversity 1:867
- Interactive Strategy Training for Active Reading and Thinking (iSTART) 8:3, 8:153
- Inter-Agency Network for Education in Emergencies (INEE) 4:691
- intercultural education 1:117
- interest 6:612–617
 - characteristics
 - general discussion 6:612
 - individual interest 6:613
 - situational interest 6:613
 - state of interest concept 6:613
 - vocational interests 6:614
- developmental models 6:614
- educational technologies 6:616
- measurement approaches 6:616
- motivational concepts 6:615
- research areas 6:616
- research background 6:612
- self-regulated learning 6:615
- interest convergence principle 6:343
- interlanguage 5:421
- internal consistency/homogeneity 3:158, 4:28, 4:29*r*
- internally-designed partnerships 7:760*r*
- internally persuasive discourse 1:433
- internal modularization 8:386
- internal validity 3:90–96
 - background information 3:91
 - causal effects measurements 2:187
 - causal studies 7:126
 - design considerations
 - analysis of covariance (ANCOVA) 3:94
 - analysis of variance (ANOVA) 3:94
 - cross-lag panel design 3:95, 3:95*f*
 - key features 3:93
 - pretest–posttest control group designs 3:94
 - regression continuity design 3:95
 - Solomon four-group design 3:94
 - random assignment approximations 3:97, 3:107
 - relational studies 7:126
 - statistical adjustments 3:95
- threats
 - additive/interactive threats 3:93
 - ambiguous temporal precedence 3:91
 - attrition 3:92
 - field experimentation 3:85, 3:86*r*
 - general discussion 3:91
 - history 3:92
 - instrumentation 3:93
 - maturation 3:92
 - oral language skill tests 3:290
 - regression to the mean 3:92
 - selection 3:91
 - testing 3:93
- International Adult Literacy Survey (IALS) 5:694
- International Association for the Evaluation of Educational Achievement (IEA) 4:663–668
 - background information 3:681, 4:45, 4:650, 4:663
- Civic Education Study (CIVED) 4:656–662
 - characteristics 3:687*r*
 - design and implementation 4:657, 4:658*r*
 - educational impacts 4:667
 - findings implications and utilization 4:660
 - general discussion 4:662
 - historical foundations 4:657
 - participating countries 4:658*r*
 - results dissemination 4:660
 - survey results and analysis 4:658
 - theoretical framework 4:659*f*
- educational evaluation systems 3:523
- educational impacts 4:666
- growth factors 4:668
- history/civics assessments 3:323
- information and communication technology (ICT) studies 4:644–649
 - background information 4:644
- CompEd study
 - characteristics 3:687*r*
 - Functional Information Technology Test (FITT) 4:645
 - importance 4:649
 - purpose 4:644
 - Stage 1 survey results 4:644
 - Stage 2 survey results 4:645
- importance 4:649
- Second Information Technology in Education Studies (SITES)
 - basic concepts 4:646
 - characteristics 3:687*r*
 - educational impacts 4:667
 - importance 4:649
 - Lithuania 5:665
 - Module 1 (SITES M1) survey results 4:646
 - Module 2 (SITES M2) survey results 4:647, 8:224
 - SITES 2006 4:648, 8:22

- international/regional assessment programs 3:475, 3:687*t*
 literature curricula 1:448
 mathematics and science studies 4:650–655
 background information 4:650
 early studies 4:650, 4:664
 First International Mathematics Study (FIMS) 3:687*t*, 4:650, 4:664
 First International Science Study (FISS) 4:651
 general discussion 4:654
 Second International Mathematics Study (SIMS) 3:687*t*, 4:651
 Second International Science Study (SISS) 3:687*t*, 4:651
 survey results and analysis 4:653
 Third International Mathematics and Science Study 1995 (TIMSS) 3:687*t*, 4:652
 Third International Mathematics and Science Study 2003 (TIMSS) 3:687*t*, 4:653
 Third International Mathematics and Science Study 2007 (TIMSS) 4:653
 Third International Mathematics and Science Study Repeat (TIMSS-R) 3:687*t*, 4:652
 research background 4:656
 research studies
 chronology 4:665*t*
 funding considerations 4:664
 organization and development 4:663
 transformation and expansion 4:665
 student achievement test performance 2:246, 3:726
 valid educational indicators 4:198
 International Atomic Energy Agency (IAEA) 4:514
 International Civic and Citizenship Education Study (ICCS) 5:665
 International Classification of Disease (ICD) 3:494
 International Commission on Mathematical Instruction (ICMI) 3:653
 International Conference on Adult Education (CONFINTEA) 1:2, 1:195
 International Congress for School Effectiveness and Improvement (ICSEI) 4:683, 6:269
 International Congress on Mathematical Education (ICME) 3:653
 International Council of Adult Education 1:195
 International Council of Museums (ICOM) 1:150
 international evaluations 3:681–689
 assessment programs 3:687*t*
 design guidelines
 cross-country validity 3:684
 general discussion 3:684
 survey implementation comparisons 3:688
 target population comparisons 3:686
 general discussion 3:681, 3:688
 historical background 3:681
 policies and programs
 educational reform 3:684
 importance 3:682
 pace of change 3:683
 performance targets 3:683
 success factors 3:682
 research frameworks 3:681, 3:682*t*
 International Federation of Adapted Physical Activity (IFAPA) 2:521
 International Institute for Educational Planning (IIEP) 4:688–693
 background information
 general discussion 4:688
 institutional autonomy 4:689
 origins 4:689
 training and research functions
 Advanced Training Program (ATP) 4:690, 4:691
 general discussion 4:689
 Regional Training Course (RTC) 4:691
 research programs 4:690
 contextual perspectives 4:688
 general discussion 4:692
 geopolitical events
 knowledge production 4:691
 responsiveness 4:690
 shared knowledge mechanisms 4:691
 training and organizational support 4:691
 mandates 4:688
International Journal of Educational Development (IJED) 1:252–253
International Journal of Educational Management (IJEM) 4:734
 International Labor Organization (ILO) 4:491, 8:262, 8:397, 8:470
 internationally mobile children *see* third culture kids (TCKs)
 International Military Education and Training (IMET) program 3:11
 International Monetary Fund (IMF)
 Finland 6:292
 higher education 1:536
 institutional autonomy 4:342
 neoliberal ideologies and strategies 6:12
 policies and programs 4:707
 International Ranking Expert Group (IREG) 4:552
International Review of Education (IRE) 1:251
 international schools
 Sri Lanka 5:817, 5:817*t*
 Sultanate of Oman 5:723
 International Society for Technology in Education (ISTE) 8:212–213
 International Successful School Principals Project (ISSPP) 4:826
 International System for Teacher Observation and Feedback (ISTOF) project 3:726–734
 advantages/disadvantages 3:732
 background information 3:726
 design rationale
 generalizability theory 3:728
 inductive and deductive approaches 3:727
 mixed-methods approaches 3:727
 modified Delphi technique 3:727
 key components
 active learning practices 3:733
 assessment and evaluation techniques 3:733
 classroom climate 3:733
 classroom management 3:733
 differentiation considerations 3:733
 inclusion considerations 3:733
 instructional clarity 3:732
 instructional skills 3:732
 lesson plans 3:733
 long-term planning 3:733
 measurement approaches 3:730*t*
 metacognition 3:733
 pedagogical content knowledge 3:733
 six steps/queries 3:728
 survey results and analysis 3:729
 International Test of Developed Abilities (ITDA) 4:4
 international voucher programs
 Chile 2:338, 2:370, 2:382*t*, 2:384, 2:419–420
 Colombia PACES program 2:382*t*, 2:384
 general discussion 2:384
 International Working Group on Education (IWGE) 4:691
 Internet-based education 8:102–108
 access and equity patterns 8:95–101
 Canada 8:97
 China 8:98
 European Union (EU) 8:97
 general discussion 8:95
 global access and use 8:98, 8:98*t*
 improvement efforts 8:99
 India 8:98
 international trends 8:21–33
 challenges 8:30
 computer access at home 8:25, 8:25*t*
 computer access in school 8:22, 8:23*t*
 frequency of use at home 8:25, 8:26*t*
 frequency of use in school 8:25, 8:26*t*
 large-scale assessments 8:22
 mathematics and science instruction 8:30, 8:32*t*
 patterns of use 8:26, 8:27*f*
 student access 8:21
 student proficiency and confidence 8:27, 8:31*t*
 theoretical perspectives 8:21
 United States
 home access 8:95, 8:96*t*
 home use 8:95
 outcomes analyses 8:95, 8:97
 school access 8:96, 8:96*f*
 school use 8:96
 basic concepts 8:102
 educational impacts 8:88–94
 goals and objectives 8:88
 growth factors 8:88
 implementation considerations 8:89
 importance 8:93
 modeling approaches 8:92
 outcomes analyses 8:91
 usage measurements 8:90
 effectiveness research
 student learning outcomes 8:106
 student retention 8:106
 electronic portfolios 3:239
 evaluation policies and practices 3:530–538
 challenges 3:536

- Internet-based education (*continued*)
- clarification processes
 - context 3:531, 3:531*t*, 3:532*t*, 3:535
 - criteria 3:531*t*, 3:533, 3:533*t*, 3:535
 - data collection and analysis 3:534, 3:536
 - evaluands 3:531*t*, 3:532, 3:533*t*, 3:535
 - methodology selection 3:533, 3:536
 - questions 3:533, 3:535
 - stakeholders 3:531*t*, 3:532, 3:535
 - completion processes
 - comparison studies 3:534, 3:536
 - metaevaluation 3:534, 3:536
 - recommendations 3:534, 3:536
 - reporting 3:534, 3:536
 - flexible frameworks 3:530
 - illustrative example
 - comparison studies 3:536
 - completion processes 3:536
 - context clarification 3:535
 - criteria clarification 3:535
 - data collection and analysis 3:536
 - evaluand clarification 3:535
 - general discussion 3:534
 - metaevaluation 3:536
 - methodology selection 3:536
 - question clarification 3:535
 - recommendations 3:536
 - reporting 3:536
 - stakeholder clarification 3:535
 - key components 3:531*f*
 - opportunities 3:536
- flexible learning **4:441–447**
- background information 4:441
 - basic concepts 4:423
 - blended learning 4:445
 - criticisms 4:445
 - definitions 4:441
 - delivery challenges 4:444
 - delivery effectiveness 4:443
 - democratization of learning 4:444
 - flexible delivery 4:442
 - general discussion 4:441, 4:446
 - information and communication technology (ICT) 4:442
 - philosophical perspectives 4:444
 - responsiveness and availability 4:445
- general discussion 8:107
- home schooling 6:217
- image-based education research 6:501
- international access patterns 8:21–33
- challenges 8:30
 - computer access at home 8:25, 8:25*t*
 - computer access in school 8:22, 8:23*t*
 - frequency of use at home 8:25, 8:26*t*
 - frequency of use in school 8:25, 8:26*t*
 - large-scale assessments 8:22
 - mathematics and science instruction 8:30, 8:32*t*
 - patterns of use 8:26, 8:27*f*
 - student access 8:21
 - student proficiency and confidence 8:27, 8:31*t*
 - theoretical perspectives 8:21
- Internet cafes 8:186
- knowledge communities 8:190
- lifelong learning 8:186
- literacy instruction 8:154
- mathematics education **8:172–183**
- challenges 8:180
 - collaborative problem solving
 - The Adventures of Jasper Woodbury* 8:177
 - Anchored Instruction 8:177
 - Enhanced Anchored Mathematics 8:177, 8:178*f*
 - MathWorlds 8:178
 - networked devices 8:177
 - participatory simulations 8:177–178
 - serious game systems 8:178
 - SimCalc 8:178, 8:179*f*, 8:199
 - embodied cognition 8:179, 8:180*f*
 - importance 8:172
 - mathematical discovery
 - ANIMATE system 8:175, 8:176*f*
 - Dynamic Geometry systems 8:176
 - dynamic statistics packages 8:176
 - Function Probe 8:175
 - general discussion 8:174
 - Logo 8:175, 8:199
 - skills and concept learning
 - background information 8:172
 - Cognitive Tutors 8:173, 8:174*f*
 - computer algebra systems (CASs) 8:173
 - Geometric Supposer 8:173
 - hand-held graphing calculators 8:172, 8:173*f*
 - intelligent tutoring systems (ITSs) 8:173
- new media technologies **8:140–149**
- animation
 - general discussion 8:143
 - picture comprehension 8:144
 - picture design 8:144
 - realistic and dynamic pictures 8:143, 8:143*f*
 - general discussion 8:141
 - learning models
 - cognitive theory of multimedia learning (CTML) 8:145, 8:146*f*
 - design guidelines 8:148
 - dual-coding theory 8:145
 - integrated model of text and picture comprehension (IMTPC) 8:145, 8:146*f*
 - research findings 8:147
 - theoretical models 8:145
 - multimedia learning environments
 - animation 8:143
 - future directions 8:148
 - general discussion 8:141
 - learning models 8:145
 - linear and nonlinear text 8:142
 - text design guidelines 8:142
 - research findings
 - coherence and contiguity 8:147
 - general discussion 8:147
 - modality 8:147
 - redundancy 8:147
 - text design guidelines
 - additional stimulation 8:142
 - conciseness 8:142
 - general discussion 8:142
 - linguistic simplicity 8:142
 - structure and organization 8:142
- open and distance learning (ODL) **4:408, 8:129**
- purposes
- educational enhancement 8:105
 - general discussion 8:104
 - replacement courses
 - advanced levels 8:105
 - at-risk students 8:105
 - rural areas 8:105
- school–community relations 8:126
- science learning **8:197–203**
- access and equity promotion 8:202
 - challenges 8:202
 - constructivist perspectives 8:197, 8:198*t*
 - contextualization
 - authenticity 8:201
 - everyday experiences 8:201
 - future directions 8:202
 - information access 8:201
 - learning environment transformation 8:201
 - physical environment interactions
 - model construction 8:199
 - scientific phenomena 8:198, 8:198*t*
 - simulations/dynamic representations 8:199
 - student investigations 8:199
 - reflective construction of knowledge
 - program management 8:200
 - teacher support and professional development 8:200
 - working memory (WM) capacity 8:200
 - social environment interactions 8:199
 - systemic impacts 8:202
- secondary education
- international programs 8:104
 - North America 8:103
- second-language learning **8:204–210**
- background information 8:204
 - benefits
 - anywhere/anytime instruction 8:208
 - general discussion 8:207
 - immediate feedback 8:208

- individualized study plans 8:207
- mistakes in private concept 8:208
- patient tutoring 8:208
- records of achievement 8:209
- classroom instruction 8:205
- English-language training demand 8:204, 8:205^f
- functional role 8:204
- research results and analysis 8:207
- subject-specific uses 8:4
- student support services 4:431
- see also* e-learning; technology-supported learning environments
- interorganizational learning 1:48
- interpretive context 4:704–705
- interpretive research 6:419–423
 - background information 6:419
 - basic concepts 6:419
 - big tent group 6:421
 - chronotopes 6:422
 - constructivism 6:421
 - general discussion 6:423
 - hybrid identities group 6:422
 - interpretive turn 6:420
 - positivist perspectives 6:422
 - postmodernism 6:421–422
 - qualitative inquiry 6:421
 - signpost group 6:420
- intrarater/intrarater reliability 3:158, 4:28, 4:29^t, 4:172
- Interstate New Teacher Assessment and Support Consortium (INTASC) 2:562, 2:563^t, 7:513
- Interstate School Leaders Licensure Consortium (ISLLC) 4:768
- interviews
 - personal epistemology assessments 5:214
 - prior learning assessment (PLA) 1:171
- intrinsic case studies 6:458
- intrinsic motivation 6:618–623
 - basic concepts 6:618, 6:619^f
 - behavioral consequences 6:621
 - classroom environments 6:627
 - creativity 4:170
 - general discussion 6:618, 6:622
 - internalization processes 6:620
 - reward systems 6:621
 - second-language learning 5:420
 - socioenvironmental contexts 6:620
- introjected regulation 6:620
- Iowa 1:654^f, 2:483
- Iowa School 6:481
- Iowa Test of Basic Skills (ITBS) 3:189
- Iran
 - early childhood education and care (ECEC) 2:111^t
 - inequality–education relationship 2:273^f
 - information and communication technology (ICT) 8:25^t
 - student achievement test performance 2:247^f, 2:417^t
 - test score–social background relationship 2:273^f
 - Trends in International Mathematics and Science Study (TIMSS) 5:645^t
- Iraq
 - early childhood education and care (ECEC) 2:111^t
 - higher education 4:518
 - warfare effects 1:797
- Ireland
 - age-related policies and programs 8:263^t
 - Catholic schools 2:356
 - childhood brain development 2:57^f
 - class size 2:492^f
 - community education 1:120
 - early childhood education and care (ECEC) 2:4^t, 2:114^t
 - earnings–education level relationship 2:307^t
 - education levels 2:279^f
 - enrollment rates 2:493^f
 - gender-equity gap 2:56^t
 - higher education
 - admissions tests 4:12
 - loans and grants 4:501^f
 - student and researcher mobility 4:287^f
 - tertiary education expenditures 4:500^f
 - tertiary type-B program survival rates 8:464^t
 - inequality–education relationship
 - earnings differentials 2:277^t
 - employment rate–educational attainment relationship 2:278^f
 - fathers' educational background 2:274^f
 - test score–social background relationship 2:273^f
- information and communication technology (ICT)
 - computer access in school 8:23^t
 - frequency of computer use in school and at home 8:26^t
 - student proficiency and confidence 8:31^t
- participation research
 - regular education 8:284^f
 - vocational education and training (VET) 8:283^f, 8:284^f
 - work-related vocational education and training (VET) 8:284^f, 8:285^f
- physical education and sports 1:459
- prison education programs 8:301, 8:301^t, 8:304
- teacher age distribution 2:492^f
- teacher shortages 2:494^f
- teacher wages and salaries 2:496^f
- teaching hours 2:493^f
- vocational education and training (VET) 8:251^t, 8:263^t
- Isaacs, Susan 3:317
- Islamic educational philosophy 6:69–73
 - acts of justice 6:70
 - communal engagement 6:70
 - conceptual perspectives 6:71
 - good action concept 6:71
 - individuation 6:71
 - just striving concept 6:71
 - public deliberation 6:70
 - socialization 6:71
 - universal justice 6:69
- Islamic schools 5:860
- isolated children 6:692
- Israel 5:618–626
 - accreditation processes 4:384–389
 - background information 5:618
 - class size 2:492^f
 - competitive schooling 2:335
 - early childhood education and care (ECEC) 2:111^t
 - earnings differentials 2:296
 - education levels 2:279^f
 - education system
 - challenges 5:624
 - compulsory education 5:619
 - early-childhood education 5:619
 - higher education 5:621, 5:622^t
 - primary education 5:619
 - secondary education 5:620, 5:621^t
 - special issues
 - Arab sector 5:623
 - Haredi* (Ultra-Orthodox) sector 5:622
 - teaching profession 5:624, 5:625^t
 - higher education
 - admissions tests 4:13
 - institutional expenditures 1:733^t
 - parental expenditures 1:734^t
 - private higher education (PHE) 4:622
 - inequality–education relationship
 - employment rate–educational attainment relationship 2:278^f
 - test score–social background relationship 2:273^f
 - information and communication technology (ICT)
 - computer access at home 8:25^t
 - computer access in school 8:23^t
 - mathematics and science instruction 8:32^t
 - moral and social reasoning 5:282
 - student achievement test performance 2:247^f
 - teacher education programs 7:528
 - teacher incentives 2:485^t
 - teaching hours 2:493^f
- Italy
 - age-related policies and programs 8:263^t, 8:266
 - Catholic schools 2:356, 2:358
 - childhood brain development 2:57^f
 - class size 2:492^f
 - early childhood education and care (ECEC) 2:4^t, 2:56^t, 2:111^t, 2:114^t
 - earnings–education level relationship 2:307^t
 - education levels 2:279^f
 - enrollment rates 2:493^f
 - higher education
 - faculty mobility 4:562
 - foreign doctoral students 4:292^t
 - historical development 1:728
 - institutional expenditures 1:733^t
 - loans and grants 4:501^f, 4:502^f
 - parental expenditures 1:734^t
 - private higher education (PHE) 4:622

- Italy (*continued*)
 tertiary education expenditures 4:500*f*
 tuition fees 4:500*f*
 inequality–education relationship
 earnings differentials 2:277*t*
 employment rate–educational attainment relationship 2:278*f*
 fathers' educational background 2:274*f*
 gender differences 2:277*t*
 test score–social background relationship 2:273*f*
 information and communication technology (ICT)
 computer access at home 8:25*t*
 computer access in school 8:23*t*
 frequency of computer use in school and at home 8:26*t*
 mathematics and science instruction 8:32*t*
 student proficiency and confidence 8:31*t*
 leadership development programs 4:753*t*, 4:754*t*, 4:755*t*
 student achievement test performance 2:247*f*
 teacher age distribution 2:492*f*
 teacher wages and salaries 2:496*f*
 teaching hours 2:493*f*
 vocational education and training (VET) 8:251*t*, 8:252*f*, 8:253*f*, 8:263*t*, 8:266
 women in employment 2:56*f*
 item- and skill-level norms 4:97
 item banking 4:86, 4:164
 item clusters 3:305
 item response theory (IRT) 4:81–88
 basic concepts 4:15, 7:63, 7:369
 beta distribution 7:99
 computerized adaptive testing (CAT) 7:397
 differential item functioning (DIF) 4:38*t*, 4:39
 educational applications
 computerized adaptive testing (CAT) 4:87
 educational assessments 4:48
 general discussion 7:9
 item banking 4:86
 item calibration and measurement 4:86
 missing data 4:86, 7:283
 optimal test assembly 4:87, 4:87*f*
 empirical Bayes methods 7:146, 7:147*f*
 general discussion 4:15
 individual item-score patterns 4:15, 4:16*f*
 item response function (IRF) 4:15
 latent class (LC) models 7:240
 Markov chain Monte Carlo (MCMC) analysis 7:251
 maximum likelihood estimate (MLE)
 conditional maximum likelihood (CML) method 4:53, 7:371
 empirical Bayes methods 7:146, 7:147*f*
 illustrative example 7:372, 7:372*t*, 7:373*t*
 joint maximum likelihood (JML) method 4:53, 7:370
 marginal maximum likelihood (MML) method 4:53, 7:351, 7:371
 missing data 7:283
 test equating and linking 4:53
 model fit statistics 7:9
 national assessment programs 4:92
 person-fit analysis 4:15–19
 applications
 computer adaptive testing (CAT) 4:18
 general discussion 4:18
 misfitting patterns 4:18
 respondent characteristics 4:18
 Guttman errors 4:16
 limitations 4:19
 nonparametric IRT-based/group-based person-fit statistics 4:16
 parametric IRT-based person-fit statistics 4:16
 person response function (PRF) 4:17
 Rasch model 4:19
 Rasch model
 general discussion 4:117
 individual item-score patterns 4:15, 4:16*f*
 linear logistic test model (LLTM) 7:85, 7:89*t*
 maximum likelihood estimation (MLE) 7:371, 7:372*t*, 7:373*t*
 research background 4:2, 4:81
 response models
 cognitive-component models 4:84
 dichotomous items 4:82
 hierarchical models 4:85
 multidimensional models 4:84, 4:84*f*
 nonparametric models 4:85
 polytomous items
 basic concepts 4:83, 4:83*f*
 graded-response model 4:83
 nominal-response model 4:83
 partial-credit models 4:84
 three-parameter logistic (3PL) model 4:82, 4:82*f*
 two-parameter logistic (2PL) model 4:82, 7:86, 7:91*t*
 self-adaptive testing 4:148–152
 assisted self-adaptive testing (ASAT) 4:149–150
 basic concepts 4:148, 4:149*f*
 benefits 4:148
 dynamic assessments 3:222–223
 general discussion 4:151
 item pool 4:149
 item-selection strategies 4:150, 4:150*f*
 perceived control/perceived performance hypotheses 4:148
 restricted self-adaptive testing (RSAT) 4:149
 theoretical investigations 4:150
 test adaptations 4:167
 test equating and linking
 basic concepts 4:52
 NEAT design
 basic concepts 4:53
 concurrent calibration 4:53
 estimation methods 4:53
 fixed parameters calibration 4:53
 Haebara linking 4:54
 mean–mean/mean–var calibrations 4:53
 Stocking–Lord linking 4:54
 score equating methods 4:54
 true score equating methods 4:54
 test scores 2:577
 iterative purification 4:40
 Itô, Kiyosi 7:451
 Ivy League 4:486–487
- ## J
- jackknife method 7:229–231
 basic concepts 7:7, 7:229
 computational statistics 7:96
 deleted-*d* jackknife 7:230
 general discussion 7:231
 grouped jackknife 7:230–231
 illustrative example
 correlation and regression 7:230, 7:230*t*
 population mean 7:230
 population variance 7:230
 jackknife versus bootstrap method 7:231
 multiple variables 7:230
 National Assessment of Educational Progress (NAEP) application 7:231
 usage 7:229
 jackknife repeated replication (JRR) 7:394
 Jacob K. Javits Gifted and Talented Student Education Act (1988) 2:829–830, 2:870
 Jamaica 5:627–629
 early childhood education and care (ECEC) 2:111*t*
 earnings–education level relationship 2:306*t*
 education system
 background information 5:627
 early childhood education and care (ECEC) 5:627, 5:628*f*
 primary schools 5:627, 5:628*f*
 secondary schools 5:627, 5:628*f*
 tertiary education 5:627, 5:628*f*
 higher education enrollment 1:702*f*, 1:703*f*
 historical background 1:684
 home-visiting programs 2:127
 secondary education 1:679
 James, William 3:167, 5:341, 6:86, 6:109, 6:546, 6:660, 6:672
 Japan
 accreditation processes 4:384–389
 childhood brain development 2:57*f*
 class size 2:492*f*, 6:200
 comparative education societies 1:266
 early childhood education and care (ECEC) 2:4*t*, 2:56*t*, 2:111*t*
 education levels 2:279*f*
 education system construction 2:347
 enrollment rates 2:493*f*
 higher education
 collective bargaining 4:495
 historical development 1:731
 institutional expenditures 1:733*t*
 loans and grants 4:501*f*, 4:502*f*
 parental expenditures 1:734*t*

- student mobility 4:561
 tertiary education expenditures 4:500f
 tertiary type-B program survival rates 8:464t
 tuition fees 4:500f
 inequality–education relationship
 employment rate–educational attainment relationship 2:278f
 gender differences 2:277t
 test score–social background relationship 2:273f
 information and communication technology (ICT)
 computer access at home 8:25t
 computer access in school 8:23t
 frequency of computer use in school and at home 8:26t
 mathematics and science instruction 8:32t
 student proficiency and confidence 8:31t
 inquiry-oriented teacher education 7:606–607
 lesson study 7:575–581
 lifelong learning model 1:157
 mathematics curricula 1:454
 music teacher education 7:706
 newcomers 1:787–792
 foreign nationals 1:788f
 internationalization processes 1:790
 Japanese returnees 1:788
 Korean nationals 1:787
 mother tongue survey 1:790f
 new foreigners 1:789, 1:790f
 sociocultural portrayals 1:791
 physically disabled students 2:520
 private higher education (PHE) 4:622
 school supply shortages 2:506f
 student achievement test performance 2:247f, 2:417t
 teacher education programs 7:529
 teaching hours 2:493f
 vocational education and training (VET) 8:249, 8:251t, 8:252f, 8:253f
 Japan Comparative Education Society (JCES) 1:266
 Jarra Report 1:732
 Jean Monnet programme 4:674
 Jensen's inequality 7:428
 jigsaw groups 8:192
jibad (just striving) 6:71
 joint attention 5:170, 6:586
 Joint Committee on Testing Practices (JCTP) 3:380
 joint degree programs 4:511
 jointly focused attention 6:685
 joint maximum likelihood (JML) method 4:53, 7:370
 Joint Negotiating Committee for Higher Education Staff (JINCHES) 4:493
 Joint Professional Military Education (JPME) 3:8
 Jordan
 academic values and expectations 4:438
 accreditation processes 4:384–389
 early childhood education and care (ECEC) 2:111t
 inequality–education relationship 2:273f
 information and communication technology (ICT)
 computer access at home 8:25t
 computer access in school 8:23t
 frequency of computer use in school and at home 8:26t
 student proficiency and confidence 8:31t
 pharmacy education 3:59
 student achievement test performance 2:247f, 2:417t
 test score–social background relationship 2:273f
 Trends in International Mathematics and Science Study (TIMSS) 5:645t
Journal of Educational Administration (JEA)
 article submission data 4:733t
 article themes 4:735, 4:736t, 4:737t
 internationality 4:734
 publication data 4:731, 4:732t
Journal of Leadership in Education (JLE) 4:734
Journal of Research on Educational Effectiveness (JREE) 3:84
 junior colleges 4:610
 junior highs 6:235
 justice and care 6:164–169
 caring perspectives 6:166
 conflicting perspectives 6:167
 educational applications 6:168
 female perspective 6:166
 justice as fairness 6:165
 principles of justice 6:165
 theoretical perspectives 6:164
 traditional justification approaches 6:168
 juvenile delinquency *see* antisocial behaviors; delinquent behaviors
- K**
 Kamehameha Early Education Program (KEEP) 3:596
 Kameoka Lifelong Learning Foundation 1:157, 1:157t
 Kansas 1:654f
 Kant, Immanuel 5:221, 6:3, 6:121, 6:176, 6:183–184, 6:186, 6:349
 Kaplan–Meier (K–M) Product-limited Method 7:460, 7:461f, 7:462f
 Kappa (κ) statistic 7:262–263, 7:263t
 Kazakhstan
 comparative education societies 1:268
 early childhood education and care (ECEC) 2:111t
 private higher education (PHE) 4:623
 Keller, Helen 2:621
 Kendall's tau (τ) rank correlation coefficient 7:261
 Kentucky 1:654f
 Kentucky Instructional Results Information System (KIRIS) 4:186, 4:187t
 Kenya
 accreditation processes 4:384–389
 early childhood education and care (ECEC) 2:111t
 free school meals 2:420
 private sector services 4:524t
 returns to education–cognitive skills 2:310t
 school choice 6:285–286
 school fees 2:419
 secondary education initiatives 6:252
 student achievement test performance 2:418t
 teacher hiring practices 2:509
 teacher incentives 2:483–484, 2:485t, 2:508
 Kernel equating method 4:52
 Kerschensteiner, Georg 8:450
 Keynes, John Maynard 3:645, 6:12–13, 6:76
 Key School (US) 6:608
 Kids as Global Scientists curricula 8:198
 Kierkegaard, S. 6:111
kikoku-shijo 1:788, 4:637
 kindergarten education
 beginning-reading programs 6:223
 center-based programs 2:100
 cost-effective programs 2:43, 2:45t
 curriculum-based measurement (CBM)
 mathematics education 4:108
 reading instruction 4:108
 Germany 5:585, 5:587, 5:588t
 Indonesia 5:612f
 Malta 5:675, 5:676f
 modeling approaches 1:514
 social relationships 6:692
 special needs students 2:603, 2:617t, 2:618, 2:618t
 see also early childhood education and care (ECEC)
 King's Medway Oxford Formative Assessment Project (KMOFAP) 3:360
 Kiribati 2:111t
k-means algorithms 7:76, 7:79t, 7:81f
k-nearest-neighbor classification method 7:141
 know-how knowledge 6:116
 knowledge-based cascade-correlation (KBCC) 5:482, 5:482f
 knowledge-based economies
 academic values and expectations 4:436
 adult education and training (AET) 1:7, 1:224
 curriculum globalization 1:312–313, 1:403
 evaluation policies and practices 3:646t, 3:649
 globalization effects 8:412
 government policies and programs
 educational strategies 1:227
 general discussion 1:226
 lifelong learning 1:226
 neoliberal ideologies and strategies 1:226
 graduate employment 4:358
 higher education
 future directions 4:223, 4:265
 globalization effects 4:206
 graduate employment 4:358
 intellectual property rights 4:362
 internationalization processes 4:540–541
 knowledge society theory 4:360–364
 social transformation 4:373
 World Bank (WB) 4:590, 4:593t
 human capital theory 1:175, 1:223
 older learners 1:175
 Singapore 8:83
 twenty-first century skills 8:158
 workplace learning 8:328

- knowledge communities 8:189–196
 - challenges 8:193
 - cognitive responsibility 8:191
 - fostering communities of learners (FCL) model
 - characteristics 8:190
 - cognitive responsibility 8:191
 - pedagogical and technological scaffolds 8:192
 - shared knowledge base 8:191
 - general discussion 8:189–190
 - knowledge-building (KB) model
 - characteristics 8:190
 - cognitive responsibility 8:191
 - pedagogical and technological scaffolds 8:192
 - shared knowledge base 8:191
 - pedagogical and technological scaffolds 8:192
 - shared knowledge base 8:191
 - twenty-first century classrooms 8:194
 - knowledge discovery in databases (KDD) 7:112–118
 - advantages 7:113
 - general discussion 7:112
 - methodologies
 - causal data mining 7:115
 - clustering algorithms 7:114, 7:114*t*
 - data distillation for human judgment 7:114*t*, 7:116
 - discovery with models 7:114*t*, 7:115
 - domain models 7:116
 - educational applications 7:116
 - general discussion 7:113
 - illustrative example 7:117
 - prediction 7:113, 7:114*t*
 - relationship mining 7:114*t*, 7:115
 - knowledge domains 5:255–264
 - disciplinary domain development 5:256
 - general discussion 5:255, 5:262
 - historical background 5:256
 - learning theories
 - background information 5:258
 - history research 5:260
 - mathematics research 5:260
 - reading research 5:259
 - science research 5:261
 - writing research 5:259
 - knowledge networks and communities 5:100–105
 - background information 5:100
 - high bonding levels 5:102
 - high bridging levels 5:103
 - importance 5:104
 - knowledge production processes
 - basic concepts 4:361, 5:100
 - higher education 4:374
 - intellectual property rights 4:362
 - new technologies 5:101
 - open source communities 5:103
 - social capital theory 5:102
 - system level networks 5:104
 - knowledge society
 - basic concepts
 - as a discourse 4:360
 - as an intellectual construct 4:360
 - as a learning society 4:361
 - higher education 4:360–364
 - general discussion 4:364
 - graduate employment 4:358
 - intellectual property rights 4:362
 - internationalization processes 4:540–541
 - knowledge production processes 4:361
 - political goals
 - general discussion 4:362
 - global information societies 4:363
 - nation-states level 4:363
 - regional level 4:363
 - social transformation 4:358, 4:373
 - Knowles, M 6:114
 - Kohlberg, Lawrence 5:279–280
 - Kohlberg tradition of moral development 5:279–280, 7:650
 - Kolmogorov, Andrei 7:451
 - Kolmogorov's inequality 7:428
 - Kolmogorov–Smirnov test 7:236, 7:337, 7:348
 - Kommunen* 5:585, 5:585*f*
 - Komvux
 - background information 1:103
 - labor-market training programs 1:103
 - Korea
 - accreditation processes 4:384–389
 - childhood brain development 2:57*f*
 - class size 2:492*f*, 6:200
 - comparative education societies 1:266
 - early childhood education and care (ECEC) 2:4*t*, 2:56*t*, 2:111*t*, 2:114*t*
 - earnings–education level relationship 2:307*t*
 - education levels 2:279*f*
 - higher education
 - faculty mobility 4:562
 - foreign doctoral students 4:292*t*
 - institutional expenditures 1:733*t*
 - loans and grants 4:501*f*, 4:502*f*
 - parental expenditures 1:734*t*
 - student and researcher mobility 4:287*f*
 - student mobility 4:561
 - tertiary education expenditures 4:500*f*
 - tuition fees 4:500*f*
 - indigenous curricula 1:309
 - inequality–education relationship
 - earnings differentials 2:277*t*
 - employment rate–educational attainment relationship 2:278*f*
 - test score–social background relationship 2:273*f*
 - informal learning 8:241–242
 - information and communication technology (ICT)
 - computer access at home 8:25*t*
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - student proficiency and confidence 8:31*t*
 - Koreans in Japan 1:787
 - mathematics curricula 1:454
 - music teacher education 7:706
 - research-quality assessments 4:308
 - returns to education–education level 2:308*t*
 - school supply shortages 2:506*f*
 - student achievement test performance 2:247*f*, 2:417*t*
 - teacher age distribution 2:492*f*
 - teacher education programs 7:527, 7:529
 - teacher shortages 2:494*f*
 - teaching hours 2:493*f*
 - vocational education and training (VET) 8:249, 8:251*t*
 - Korean Comparative Education Society (KCES) 1:266
 - Kosovo 4:517
 - KPMG Study (1999) 3:73
 - Kruskal–Wallis test 7:349
 - Kuder–Richardson 20 (KR20) formula 4:28
 - kurtosis 7:269*f*, 7:271*f*, 7:272
 - Kuwait
 - early childhood education and care (ECEC) 2:111*t*
 - inequality–education relationship 2:273*f*
 - information and communication technology (ICT) 8:25*t*
 - student achievement test performance 2:417*t*
 - test score–social background relationship 2:273*f*
 - Kvindeligt Arbejderforbund (KAD) 8:271–272
 - Kyrgyzstan
 - early childhood education and care (ECEC) 2:111*t*
 - information and communication technology (ICT) 8:23*t*
- L**
- laboratory schools 1:355, 1:357–358
 - labor education 1:139–143
 - background information 1:139
 - characteristics 1:139
 - general discussion 1:143
 - knowledge-based economies 1:142
 - learning conditions 1:141
 - labor market
 - basic skills training 8:257
 - compensating differentials models 2:465–472
 - background information 2:465
 - displaced workers 8:276
 - general discussion 2:470
 - hedonic wage model 2:466
 - displaced workers 8:275–279
 - assistance programs
 - Austria 8:277
 - evaluation practices 8:278
 - France 8:278
 - Germany 8:278
 - retraining 8:277

- Spain 8:278
- Sweden 8:278
- United States 8:277
- causal factors 8:275
- characteristics 8:275
- losses 8:276
- occupational mobility 8:275
- wages and salaries 8:276
- early childhood education and care (ECEC) effects 2:66
- globalization effects 8:411
- graduate employment 4:354–359
 - differentiation considerations 4:356
 - division of labor 4:357
 - employability enhancements 4:356
 - employment rates and earnings 4:354
 - expansion effects 4:354
 - globalization 4:358
 - job quality 4:355
 - knowledge society theory 4:358
 - research biases 4:359
 - student demands 4:354
 - transition periods 4:357
 - worker productivity 4:355
- hedonic wage model
 - controllable versus uncontrollable cost factors 2:466, 2:467*t*
 - geographic cost-of-education index 2:467, 2:468*f*
 - limitations
 - comparable wage index (CWI) 2:468
 - cost factor interpretation 2:469
 - fixed-effects model 2:469
 - flexible wages 2:469
 - general discussion 2:468
 - geographic considerations 2:470
 - regional factors 2:468
 - teacher quality measures 2:469
 - public versus private schools 2:467
 - time correlations 2:467
- higher education 4:365–369
 - access and equity patterns 1:636
 - expansion effects 4:365
 - for-profit sector 1:708, 1:710*f*
 - general discussion 4:365
 - graduate employment 4:366
 - institutional characteristics 4:367
 - Mexico
 - demographic structure 1:720, 1:721*t*
 - economic dynamics 1:720
 - employment trends 1:722–724, 1:722*t*
 - future directions 1:726
 - industrial structure 1:722, 1:722*t*, 1:723*t*
 - social change impacts 1:719
 - university–industry collaboration 1:724
 - qualitative aspects 4:366
 - quantitative aspects 4:365
 - racial and ethnic groups 1:542–546
 - corporate environments 1:544
 - earnings differentials 1:543
 - economic well-being 1:545
 - income gaps 1:543
 - research background 1:542
 - social mobility 1:542
 - stratification 1:542
 - research background 4:368
- immigrant education 8:280–288
 - immigrant characteristics 8:280
 - investment/noninvestment determinants 8:280
 - pay-offs 8:286
 - regular education participation 8:284*f*
 - research areas 8:287
 - targeting programs 8:285
 - VET participation research 8:282, 8:283*f*, 8:284*f*, 8:285*f*
- Komvux 1:103
- mass movement/skills migration 8:412
- post-war migration impacts 1:737
- school-to-work transitions 8:309
- skill shortages 8:322, 8:325
- teacher labor markets 2:456–464
 - compensating differentials models 2:466
 - developing countries
 - accountability and management practices 2:506
 - general discussion 2:504
 - pupil–teacher ratios (PTRs) 2:506*f*
 - recruitment and requirements 2:505
 - salary characteristics 2:505
 - salary levels 2:504
 - supply shortages 2:506*f*
 - working conditions 2:505
- supply-and-demand model
 - allocation policies 2:458
 - background information 2:456
 - causal relationships 2:457
 - education–production function model 2:457
 - future directions 2:461
 - methodologies 2:459
 - qualified teachers 2:456, 2:458*f*
 - recruitment and retention 2:457
- workforce reform 6:242–248
 - current trends 6:242
 - functionalist position 6:244
 - justification 6:244
 - outcomes analyses 6:246
 - performance leadership emphasis 6:243
 - qualified teacher status (QTS) 6:242
 - remodelling project 6:242
 - socially critical position 6:245
 - standards 6:242
- teacher supply 2:473, 2:489, 2:490*f*
- twenty-first century skills 8:158
- working mothers 2:33, 2:55, 2:151
- see also* workplace learning
- labor unions
 - career guidance 8:271
 - higher education collective bargaining 4:491
 - skill shortages 8:321
 - teachers' unions 2:512–517
 - collective bargaining laws 2:513
 - district-wide impacts
 - general discussion 2:515
 - union–school district spending relationship 2:516
 - union–student achievement relationship 2:516
 - union–student/teacher ratio relationship 2:515
 - union–wage relationship 2:515
 - union–working conditions relationship 2:515
 - general discussion 2:516
 - historical background 2:512
 - influential factors 2:512
 - negative impacts
 - community welfare reductions 2:514
 - democratic process threats 2:515
 - general discussion 2:514
 - lack of trust considerations 2:514
 - performance incentives reductions 2:514
 - resource allocation reductions 2:514
 - positive impacts
 - efficiency-enhancing resources 2:514
 - information flow improvements 2:514
 - market failure reduction 2:514
 - stable school districts 2:514
 - teacher quality 2:513
 - time management 2:514
 - union impacts
 - district-wide impacts 2:515
 - general discussion 2:513
 - negative impacts 2:514
 - positive impacts 2:513
- Lacan, Jacques 6:93, 6:111, 6:351
- Ladders to Literacy program 6:223
- Länder governments 4:605, 5:582
- language 5:345–354
 - academic language
 - characteristics 5:352
 - cognitive perspective 5:353
 - sociocultural perspective 5:353
 - acquisition processes 5:375–381
 - cross-linguistic variations 5:379
 - developmental disorders 5:380
 - developmental milestones
 - early word production and comprehension 5:377, 5:378*f*
 - general discussion 5:376
 - infant babble 5:377
 - morpheme acquisition order 5:378*t*
 - pre-linguistic achievements 5:376
 - sentence formation 5:378
 - skill refinements 5:379

- language (*continued*)
 Wug test 5:378f
 developmental research 5:376
 individual variations 5:380
 theoretical perspectives 5:380, 5:384
 computer-assisted instruction 8:4
 culturally and linguistically diverse (CLD) learners
 community-based experiences 2:550–554
 best practices strategies 2:552
 general discussion 2:553
 legal mandates 2:552
 research background 2:550
 transition frameworks 2:550
 cultural difference theory model
 cultural incongruity model 2:776
 general discussion 2:776
 oppositional cultural framework 2:776
 stereotype threats 2:776
 dropouts 2:846
 gifted and talented students 2:872
 hearing impaired students 2:631
 immigrant populations 4:631
 language development considerations 2:746
 oral language skills assessments 3:291
 overrepresentation 2:774–779
 challenges 2:778
 cultural deficits 2:775
 cultural difference theory model 2:776
 deficit thinking effects 2:775
 ecological/contextual perspectives 2:777
 historical perspectives 2:665, 2:690, 2:774
 school failure explanations 2:775
 speech and language impairments
 African-American English dialect speakers 2:656
 bilingual students 2:656
 demographic data 2:656
 individualized education plans (IEPs) 2:658, 2:658f
 Individuals with Disabilities Education Act (IDEA, 1990) 2:657
 definitions 2:746
 developmental processes 2:746
 functional role 5:345–346
 general discussion 5:353
 human language characteristics 5:375
 language and literacy instruction 1:435–440
 computer-assisted instruction 8:150–157
 comprehension 8:152
 educational impacts 5:121, 5:121f
 emergent literacy 8:150
 fluency 8:151
 general discussion 8:150, 8:155
 Internet access 8:154
 phonics and phonemic awareness 8:151
 software systems 8:3
 text-to-speech applications 8:153
 theoretical perspectives 8:154
 vocabulary 8:152
 curricula
 adolescent literacies 1:438
 digital technology 1:439
 general discussion 1:435
 primary language curricula 1:436
 primary literature teaching 1:437
 primary reading curricula 1:436
 primary writing instruction 1:437
 secondary language curricula 1:438
 secondary literature teaching 1:438
 general discussion 1:435, 1:439
 influential factors 1:435
 primary schools
 language curricula 1:436
 literature teaching 1:437
 reading curricula 1:436
 writing instruction 1:437
 secondary schools
 adolescent literacies 1:438
 digital technology 1:439
 language curricula 1:438
 literature teaching 1:438
 socioeconomic modernization 1:217
 special needs students 2:759–766
 workplace learning 1:86
 language articulation 2:654, 7:589
 language-based learning disabilities (LBD) 2:655
 language disorders 2:655
 learning theories
 behaviorist perspective 5:346
 cognitive perspective
 academic language 5:353
 comparison studies 5:349, 5:352
 connectionism 5:347
 constructivist approaches 5:347
 first language acquisition 5:247
 general discussion 5:347
 literacy acquisition 5:350
 sociocognitive approaches 5:347
 comparison studies
 general discussion 5:349
 individual versus social learning approaches 5:350
 innate versus acquired aspects 5:349
 learning focus 5:350
 literacy acquisition 5:352
 task conceptualization 5:349
 general discussion 5:346
 sociocultural perspective 5:384
 academic language 5:353
 comparison studies 5:349, 5:352
 ethnographic studies 5:349
 general discussion 5:348
 literacy acquisition 5:351
 sociolinguistic/socialization approaches 5:348
 Vygotskian tradition 5:232, 5:348, 5:499, 5:501
 literacy acquisition
 bilitery
 continua model 1:477
 general discussion 1:477
 integrated instructional strategies 1:477
 cognitive perspective 5:350
 comparison studies 5:352
 general discussion 5:350
 hearing impaired students 2:630
 sociocultural perspective 5:351
 moderate hermeneutics 6:66
 mother tongue teacher education 7:738–743
 challenges 7:741
 field experiences 7:741
 general discussion 7:738
 pedagogical content knowledge 7:739
 postcolonial theory 1:302
 professional development
 current issues 7:739
 demographic shifts 7:740
 literacy considerations 7:741
 methods courses 7:740
 subject-matter knowledge 7:738
 narrative practices 7:612
 nonstandard dialects 5:365
 official language of schooling 1:435–440
 curricula
 adolescent literacies 1:438
 digital technology 1:439
 general discussion 1:435
 primary language curricula 1:436
 primary literature teaching 1:437
 primary reading curricula 1:436
 primary writing instruction 1:437
 secondary language curricula 1:438
 secondary literature teaching 1:438
 general discussion 1:435, 1:439
 influential factors 1:435
 primary schools
 language curricula 1:436
 literature teaching 1:437
 reading curricula 1:436
 writing instruction 1:437
 secondary schools
 language curricula 1:438
 literature teaching 1:438
 second-language learning 5:419–425
 ability models 5:382, 5:383f
 assessment strategies 4:140–147
 classroom assessment 4:144
 construct research 4:142

- ethical issues 4:143
- foreign language proficiency testing 4:144
- future challenges 4:144
- historical perspectives 4:140
- purpose 4:140
- classroom discourse 5:365
- computer-assisted instruction 8:4, 8:204–210
 - background information 8:204
 - benefits 8:207
 - classroom instruction 8:205
 - English-language training demand 8:204, 8:205^f
 - functional role 8:204
 - research results and analysis 8:207
 - subject-specific uses 8:4
- first language environments 1:441–446
 - background information 1:441
 - basic concepts 1:441
 - challenges 1:445
 - Common European Framework (CEF) 1:442
 - content-based instruction 1:443
 - general discussion 1:445
 - immersion programs 1:443
 - language resource centers 1:444
 - oral competency 1:441
 - proficiency orientation 1:441
 - standards 1:442
 - teacher preparation 1:444
- immigrant populations 4:631
- oral language skills assessments 3:291
- research challenges 5:382
- theoretical perspectives
 - cognitive approach 5:419
 - contrastive linguistics 5:419
 - critical-age hypothesis 5:422
 - foreign language learning 5:384, 5:422
 - interlanguage 5:421
 - motivation 5:420
 - school learning impacts 5:423
 - social theory approach 5:420
 - variation analysis 5:421
- sign language 2:545, 2:624, 2:630
- special needs students 2:745–751
 - developmental milestones 2:747
 - effective communication skills 2:745, 2:746^t
 - family involvement 2:749
 - general discussion 2:749
 - language difference versus language disorder 2:747
 - proactive strategies 2:749
 - serious emotional disturbances (SEDs) 2:748
 - specific disability impacts 2:748
 - speech and language impairments 2:654–660
- speech and language impairments 2:654–660
 - culturally and linguistically diverse (CLD) learners
 - African-American English dialect speakers 2:656
 - bilingual students 2:656
 - demographic data 2:656
 - individualized education plans (IEPs) 2:658, 2:658^f
 - Individuals with Disabilities Education Act (IDEA, 1990) 2:657
- diagnostic and classification processes 2:579
- disorder characteristics
 - articulation 2:654
 - childhood apraxia of speech (CAS) 2:654
 - general discussion 2:654
 - language-based learning disabilities (LBLD) 2:655
 - language disorders 2:655
 - orofacial myofunctional disorders (OMD) 2:654
 - phonological disorders 2:654
 - stuttering 2:655
 - voice disorders 2:655
- future directions
 - evidence-based practices 2:659
 - general discussion 2:658
 - technology-supported learning environments 2:658
 - testing and assessments 2:658
 - general discussion 2:654, 2:659
 - school-based services 2:854
- see also* bilingualism; first language learning
- languages other than English (LOTE) 1:441
- Laos 2:111^t
- Laplace, Pierre-Simon 7:467
- laptop computers 8:128
- large-sample statistical methods 7:232–237
 - asymptotic theory 7:232
 - categorical data analysis 7:236
 - central limit theorem 7:234
 - empirical distribution functions 7:236
 - estimator properties 7:235
 - general discussion 7:237
 - linear regression models 7:236
 - Monte Carlo analysis 7:10
 - order statistics 7:236
 - statistical inference 7:232
 - stochastic convergence 7:233
 - test statistics 7:236
- large-scale assessments
 - computer-assisted assessments 8:48–55
 - conceptual framework 8:48
 - first-generation computer-based tests 8:48, 8:49^t
 - general discussion 8:54
 - Generation R tests 8:49^t, 8:52
 - next-generation electronic tests 8:49^t, 8:50
 - technology-rich environments scenarios 8:52, 8:53^f
 - fairness 4:59
 - history/social studies/civics 3:323
 - information and communication technology (ICT) surveys 8:22
 - Latin America 3:479
 - portfolios 3:237
 - research methodologies 3:744
 - special needs students 2:752–758
 - accountability 2:755
 - assessment accommodations 2:530
 - assessment options 2:753
 - background information 2:752–753
 - general discussion 2:757
 - purpose 2:753
 - United States education system
 - historical perspectives 6:303
 - research findings 6:305
- Larry P. v Riles (1979) 2:716–717
- l'Association pour le développement des méthodologies en éducation* (ADMÉE) 3:432
- latent class (LC) models 7:238–244
 - background information 7:238
 - basic concepts
 - data sets 7:239^t
 - general discussion 7:238
 - proportion and probability parameters 7:239^t
 - educational applications 7:9
 - explanatory variables 7:241
 - extensions 7:241
 - generalized linear models (GLMs) 7:240
 - item response theory (IRT) 7:240
 - maximum likelihood estimation (MLE) 7:242
 - model selection considerations 7:242
 - probabilistic Guttman models 7:240
 - Proctor model 7:240
 - proportion and probability parameters 7:239^t, 7:240^t
 - restricted models 7:240
 - software packages 7:243
- latent growth curves model (LGCM)
 - basic concepts 3:122, 3:122^f, 3:123^f
 - self-concept study 3:123, 3:123^t, 3:124^t
- latent variable models 7:15, 7:15^t
- lateral accountability 6:279
- Latin America
 - accreditation processes 4:384–389
 - assessment system development 3:479–485
 - Argentina 3:479
 - background information 3:479
 - Bolivia 3:480
 - Brazil 3:480
 - challenges 3:484
 - Chile 3:480
 - Colombia 3:480
 - Costa Rica 3:481
 - Cuba 3:481
 - Dominican Republic 3:481
 - Ecuador 3:481
 - El Salvador 3:481
 - Guatemala 3:482
 - Honduras 3:482
 - international assessment participation 3:483
 - large-scale assessments 3:479

Latin America (*continued*)

- Mexico 3:482
- Nicaragua 3:482
- Panama 3:483
- Paraguay 3:483
- Peru 3:483
- Uruguay 3:483
- Venezuela 3:483
- community education 1:123
- conditional cash transfer (CCT) programs 2:419
- critical theory perspectives 6:351–352
- earnings–education level relationship 2:306*t*, 2:307*f*
- enrollment rates
 - gender disparities 2:416*t*
 - primary schools 2:415*t*
 - secondary schools 2:416*t*
- higher education enrollment 1:698, 1:702*f*, 1:703*f*
- inequality–education relationship 2:273*f*
- Japanese nationals 1:789
- private higher education (PHE) 4:623
- pupil–teacher ratios (PTRs) 2:506*f*
- returns to education–education level 2:308*t*
- school choice 6:285
- student achievement test performance 2:247*f*, 2:417*t*
- teacher training 2:509
- test score–social background relationship 2:273*f*
- Latino students 1:610–615
 - college-choice model
 - basic concepts 1:611, 1:612*f*
 - chain migration theory 1:613
 - decision-making models 1:613*f*
 - quantitative methods 1:612
 - social capital theory 1:613
 - cultural competence 2:565–570
 - basic concepts 2:566
 - community involvement 2:568
 - importance 2:569
 - overrepresentation 2:565
 - Pedersen's model 2:566
 - research background 2:565
 - cultural difference theory model 2:776
 - deaf-blind population 2:623*t*
 - demographic data 1:610
 - educational attainment 1:610
 - educational statistics 1:581–588
 - choice effects 1:585
 - college enrollment opportunities 1:656, 1:658*f*
 - continuing disparities 1:584
 - demographic diversity 1:582
 - digital technology implications 1:584
 - disparity measures 1:581, 1:874
 - dropout rates 1:874, 1:875
 - education beyond borders 1:586
 - 18–24-year-olds demographics 1:802, 1:803*f*
 - enrollment rates 1:800, 1:801*f*, 1:802*f*, 1:803*f*
 - global village concept 1:587
 - graduation rates 1:874
 - incarceration rates 1:874
 - twenty-first century perspectives 1:581, 1:587
 - United States 1:582
 - information and communication technology (ICT) access and equity
 - home access 8:95, 8:96*t*
 - home use 8:95
 - outcomes analyses 8:95, 8:97
 - school access 8:96, 8:96*f*
 - school use 8:96
 - learning opportunities 1:610
 - oppositional identity 1:756, 2:776
 - race-specific consequences 1:876
 - standardized tests 2:577
 - traumatic brain injury (TBI) 2:695
 - valued peers 1:757
- Latvia 5:630–635
 - age-related policies and programs 8:263*t*
 - background information 5:630
 - early childhood education and care (ECEC) 2:111*t*
 - education system
 - administrative and supervisory structure 5:633
 - challenges 5:635
 - educational finance 5:634
 - goals 5:630

- monitoring and evaluation programs 5:634
- research programs 5:634
- structural and operational characteristics
 - adult education 5:633
 - alternative educational settings 5:631
 - basic education 5:631
 - early childhood education and care (ECEC) 5:631
 - higher education 5:632, 5:633*t*
 - nonformal education 5:633
 - preschool education 5:631
 - primary education 5:631
 - secondary education 5:631
 - special needs students 5:632
 - vocational education and training (VET) 5:632
- teaching profession 5:633
- inequality–education relationship 2:273*f*
- information and communication technology (ICT)
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - student proficiency and confidence 8:31*t*
- information and communication technology (ICT) access and equity 8:97
- test score–social background relationship 2:273*f*
- vocational education and training (VET) 8:263*t*
- Lau v Nichols* (1974) 2:716–717
- Law School Admission Test (LSAT) 1:695, 4:9
- laws of large numbers (LLNs) 7:234
- Lazarus's transactional theory of stress and coping 6:550
- leadership 4:695–703, 4:852–856
 - authentic leadership 5:7–11
 - arenas of influence 5:9, 5:9*f*
 - characteristics 5:7
 - complications
 - cross-cultural issues 5:10
 - inauthentic/false leadership 5:10
 - general discussion 5:10
 - individual valuation processes 5:7, 5:8*f*
 - valuation processes 5:7
 - basic concepts 4:741
 - contextual literacy
 - achievement measures 4:697
 - competing forces
 - continuity versus constant change 4:696
 - dependence versus independence 4:696
 - homogeneity versus heterogeneity 4:697
 - individualism versus community 4:696
 - general discussion 4:695
 - contextual perspectives 4:704–709
 - basic concepts 4:704
 - context levels
 - macrolevels 4:706
 - mesolevels 4:705
 - microlevels 4:705
 - general discussion 4:704, 4:708
 - global context
 - demographic globalization 4:708
 - economic globalization 4:707
 - general discussion 4:706
 - technological globalization 4:707
 - contingent leadership 5:45
 - democratic leadership 5:12–17
 - characteristics 5:12
 - conceptualizations 5:13
 - distributed versus democratic leadership 5:13, 5:14*f*
 - empirical research approaches 5:14
 - future directions 5:16
 - theoretical perspectives 5:12
- disadvantaged schools 4:777–782, 5:58–65
 - challenges 4:777
 - deficit views 5:58
 - policies and programs 5:59
 - reform impacts 5:60
 - teacher turnover 5:60
 - context-specific improvement 4:780
- effective leadership
 - community engagement 4:779
 - distributed leadership 4:778
 - general discussion 4:781
 - improvement strategies 4:779
 - learner-centered leadership 4:779
 - research background 4:778
 - shared values and vision 4:778

- staff development 4:779
 - improvement interventions 4:777
 - leadership impacts 5:61
 - urban environments 4:809–816
- dissatisfaction measures 4:855
- distributed leadership 5:1–6
 - basic concepts 4:741, 4:818
 - comparison studies 5:4
 - disadvantaged schools 4:778
 - effectiveness 5:4
 - general discussion 5:1, 5:5
 - heroic versus distributive leadership model 4:744
 - leader plus aspect model
 - conceptual perspective 5:2
 - empirical perspective 5:2
 - practice aspect model
 - conceptual perspective 5:3
 - empirical perspective 5:3
 - professional learning communities 5:155
 - school-to-school collaboration 6:279, 6:279f
 - teacher leadership 5:40
 - theoretical perspectives 4:764
- diverse cultures 4:824–831
 - future research areas 4:829
 - influencing factors 4:824
 - key themes
 - change agents 4:825
 - cultural influences 4:828
 - developing countries 4:826
 - general discussion 4:825
 - leader development programs 4:827
 - successful leadership 4:826
 - research background 4:824
- diversity considerations 4:832–838
 - demographic data 4:832
 - general discussion 4:832, 4:836
 - inclusion and equity strategies 4:834
 - leadership approach characteristics 4:833
 - school community participation 4:834
 - student experiences 4:833
- early childhood education and care (ECEC) 2:89
- educational effectiveness 5:55
- esthetics 4:783–789
 - general discussion 4:783
 - organizational artifacts 4:784
 - organizational capital 4:787
 - organizational contexts 4:783
 - organizational culture 4:786
 - organizational leadership 4:784
 - organizational power 4:786
 - research areas 4:788
- ethical practices 4:717–723
 - empirical analyses 4:720
 - future research areas 4:721
 - international scholarly networks 4:681, 4:721
 - societal context 4:717
 - theoretical perspectives
 - administrators and school leaders 4:719
 - multiple perspective approach 4:719
 - recent developments 4:718
- expectations
 - characteristics 4:852
 - commitment 4:854
 - determinants 4:853
 - general discussion 4:852
 - improvement strategies 4:853
 - performance standards 4:853
- gender analysis 4:797–802
 - diversity challenges 4:800
 - general discussion 4:797
 - research approaches
 - feminist perspectives 4:798
 - mainstream studies 4:797
- general discussion 4:695, 4:855
- higher education
 - leadership structures 4:246
 - university leaders 4:238–244
 - diversity considerations 4:238
 - future research areas 4:242
 - historical perspectives 4:238
 - literature reviews 4:238
 - literature types 4:239
 - presidential leadership 4:239, 4:240
 - senior management teams 4:239, 4:241
- historical background 4:724–730
 - administrative theories 4:724
 - civil rights movement 4:726
 - efficiency improvements 4:724
 - female leadership 4:725
 - postpositivist perspectives 4:727
 - professional managers 4:725
 - scientific management approach 4:724
 - Theory Movement 4:725
 - twentieth-century perspectives 4:727
 - twenty-first century perspectives 4:728
 - United Kingdom 4:726
- information and communication technology (ICT) 5:120–127
 - educational impacts
 - community relations 5:121, 5:121f
 - general discussion 5:120
 - instructional strategies 5:121, 5:121f
 - literacy instruction 5:121, 5:121f
 - social networks 5:120, 5:121f
 - influencing factors 5:125, 5:125f
 - integration framework 5:123f
 - requirements and standards 5:121
 - social and legal challenges 5:124
 - teacher leadership 5:122, 5:123f
 - technology-mediated leadership development programs 5:123, 5:123r
- instructional leadership 5:18–25
 - basic concepts 5:19, 5:20f
 - empirical analyses
 - general discussion 5:20
 - research findings 5:21
 - trends 4:742, 5:21
 - historical background 5:18
 - international trends 4:759
 - shared leadership models 5:22
 - student learning impacts 5:45
- leadership development programs 4:752–761
 - diverse cultures 4:827
 - general discussion 4:752, 4:759
 - importance 4:752
 - international approaches 4:752, 4:753r
 - international trends
 - andragogic principles 4:757
 - centralized and decentralized programs 4:753, 4:754r
 - communication and cooperation 4:756
 - course-based learning 4:758r
 - experience-oriented/application-oriented methods 4:757, 4:758r
 - explicit goals and objectives 4:758
 - instructional leadership 4:759
 - integral leadership 4:759
 - integrative leadership 4:759
 - knowledge production processes 4:757
 - leadership for improvement 4:757
 - modular and multiphase designs 4:756, 4:756r
 - moral leadership 4:759
 - partnership models 4:754
 - personal development 4:756
 - posttransformational leadership 4:759
 - preparatory qualifications 4:755, 4:755r
 - problem-based learning methods 4:758
 - program duration 4:755, 4:755r
 - quality assurance 4:753
 - school's core purpose 4:759
 - team leadership capacity 4:757
 - theory–practice relationships 4:754
 - transformational leadership 4:759
 - workplace learning 4:758
 - technology-mediated leadership development programs 5:123, 5:123r
- leadership for learning 4:746–751
 - assessment programs 4:748
 - belief and value systems 4:751
 - characteristics 4:747r
 - curricular programs 4:748
 - instructional program 4:747
 - learning communities 4:748
 - organizational culture 4:750
 - performance feedback 4:747
 - resource allocation and use 4:749
 - reward systems 4:750

- leadership (*continued*)
- school–community relations 4:751
 - school-to-school collaboration 6:279, 6:279f
 - social advocacy 4:750
 - vision and mission 4:746
 - managerial leadership 4:759, 5:45
 - moral leadership 4:759, 5:45
 - National College for School Leadership (NCSL) 3:539, 4:762, 5:93, 6:243
 - organizational change
 - challenges 5:131
 - complex systems theory 5:130
 - educational innovation 5:131
 - Fullan's concepts 5:132, 5:133t
 - future directions 5:133
 - historical perspectives 5:128
 - sociopolitical perspectives 5:131
 - organizational savvy
 - collaborative networks 4:699
 - general discussion 4:698
 - mechanistic–living system transitions 4:698
 - personalization considerations 4:699
 - professional learning communities 4:700
 - social capital 4:700
 - thin–deep democracy transitions 4:698
 - participative leadership 5:45
 - philosophical perspectives 4:710–716
 - empiricism 4:710
 - epistemology 4:714
 - general discussion 4:710
 - ontology 4:712
 - theoretical perspectives
 - general discussion 4:711
 - scientific approach 4:710
 - values 4:714
 - principal's role 5:67
 - productive schools 4:746–751
 - assessment programs 4:748
 - belief and value systems 4:751
 - characteristics 4:747t
 - curricular programs 4:748
 - instructional program 4:747
 - learning communities 4:748
 - organizational culture 4:750
 - performance feedback 4:747
 - resource allocation and use 4:749
 - reward systems 4:750
 - school–community relations 4:751
 - social advocacy 4:750
 - vision and mission 4:746
 - professional research journals 4:731–739
 - article themes 4:735, 4:736t, 4:737t
 - background information 4:738
 - Cases in Educational Leadership* (CEL) 4:732
 - editorial processes 4:732, 4:733t
 - Educational Administration Quarterly* (EAQ)
 - article submission data 4:733t
 - article themes 4:735, 4:736t, 4:737t
 - internationality 4:734
 - publication data 4:731, 4:732t
 - Educational Management Administration & Leadership* (EMAL)
 - article submission data 4:733t
 - article themes 4:735, 4:736t, 4:737t
 - internationality 4:734
 - publication data 4:731, 4:732t
 - electronic journals 4:732
 - internationality 4:734
 - International Journal of Educational Management* (IJEM) 4:734
 - Journal of Educational Administration* (JEA)
 - article submission data 4:733t
 - article themes 4:735, 4:736t, 4:737t
 - internationality 4:734
 - publication data 4:731, 4:732t
 - Journal of Leadership in Education* (JLE) 4:734
 - publication data 4:731
 - refereed articles 4:733
 - research methodologies 4:733
 - selected journals 4:732t
 - submission data 4:733t
 - research areas 4:702
 - research results analyses 4:766
 - research trends 4:740–745
 - background information 4:740
 - backward mapping 4:742, 4:743f
 - empirical analyses 5:21
 - general discussion 4:744
 - heroic versus distributive leadership model 4:744
 - impacts on students 4:741, 5:21
 - pedagogical content knowledge 4:742, 4:743f
 - professional learning communities 4:743
 - transformational–instructional leadership relationship 4:742
 - satisfaction measures 4:854
 - school–higher education institution (HEI) partnerships 7:761
 - school improvement 5:135–142
 - challenges 5:138
 - collaborative networks 5:92
 - conceptualizations 5:135
 - leadership conceptualizations 5:136
 - research areas 5:138
 - research perspectives
 - cultural perspectives 5:137
 - general discussion 5:136
 - political perspectives 5:137
 - rational–technical model 5:136
 - school restructuring role 5:66–71
 - changing roles
 - cultural role 5:69
 - environmental role 5:69
 - instructional role 5:69
 - political role 5:68
 - conflicts and tensions
 - accountability 5:67
 - organizational restructuring 5:68
 - professionalism 5:67
 - structural coupling expectations 5:68
 - distributed leadership 5:67
 - general discussion 5:66
 - implications 5:70
 - professional learning communities 5:66
 - reculturing processes 5:66
 - results analysis 5:69
 - rule changes 5:69
 - social justice 5:67
 - stakeholder relationships 5:69
 - smart practices 4:701
 - social capital 5:85–91
 - background information 5:85
 - basic concepts 5:87
 - impact on educational success 5:88
 - impact on leadership roles and responsibilities 5:89
 - as an intellectual construct 5:86
 - social trust–child well-being relationship 5:87t
 - special education 2:691
 - strategic leadership 5:34–39
 - basic concepts 5:34
 - characteristics
 - activities and functions 5:34, 5:38t
 - culture and shared values development 5:36
 - distinctive competencies development 5:37
 - network development 5:37
 - organizational process development 5:35
 - personnel development 5:35
 - general discussion 5:38
 - social intelligence 5:37
 - student learning impacts 5:45
 - student learning impacts 5:45–50
 - achievement assessments 5:47
 - general framework 5:45, 5:46f
 - leadership models 5:45
 - limitations 5:46
 - mediating variables 5:46f, 5:47
 - moderating variables 5:48
 - research trends 4:741, 5:21
 - unit of analysis challenge 5:48
 - student voice 4:803–808
 - basic concepts
 - co-leadership role 4:805
 - consulting role 4:804
 - general discussion 4:804
 - leadership roles 4:805
 - benefits 4:806
 - challenges
 - anxiety 4:806

- credibility 4:807
- general discussion 4:806
- inclusivity 4:806
- innovation issues 4:807
- contemporary opportunities 4:803
- general discussion 4:807
- historical perspectives 4:803
- sustainable development reform 6:292
- system leadership 5:78–84
 - activities and capabilities 5:80, 5:81*f*, 5:82*f*
 - basic concepts 5:78
 - challenges 5:83
 - development opportunities 5:81
 - functional roles 5:79
 - general discussion 5:78
 - school-to-school collaboration 6:279, 6:279*f*
- teacher leadership 5:40–44
 - basic concepts 5:40
 - distributed leadership 5:40
 - educational impacts
 - school level 5:41
 - student level 5:42
 - teacher level 5:42
 - information and communication technology (ICT) 5:122, 5:123*f*
 - leadership for learning 4:818
 - organizational development considerations 5:40
 - professional development considerations 5:43
- theoretical perspectives
 - context matters 4:763
 - core practices 4:764
 - distributed leadership 4:764
 - learning-centered leadership 4:763
 - polyphonic work theory 4:765
 - research background 4:762
 - succession planning 4:764
 - sustainable leadership 4:765
- transformational leadership 5:158–164
 - characteristics
 - general discussion 5:159
 - instructional program management 5:161
 - organization design 5:160
 - people development practices 5:159
 - setting directions 5:159
 - comparison studies 5:4
 - criticisms 5:162
 - historical background 5:158
 - influencing effects 5:158
 - instructional program management
 - buffering functions 5:162
 - general discussion 5:161
 - instructional support 5:162
 - monitoring operations 5:162
 - staffing considerations 5:161
 - international trends 4:759
 - organization design
 - collaborative cultures 5:160
 - community engagement 5:161
 - general discussion 5:160
 - restructuring considerations 5:161
 - school networks 5:161
 - people development practices
 - general discussion 5:159
 - individualized support 5:160
 - intellectual stimulation 5:160
 - leadership models 5:160
 - research trends 4:742
 - school improvement 5:138
 - setting directions
 - general discussion 5:159
 - group goal acceptance 5:159
 - high-performance expectations 5:159
 - shared vision 5:159
 - student learning impacts 5:45
- transformative leadership 5:26–33
 - benefits 5:31
 - criticisms 5:31
 - ethical leadership 5:26
 - evolutionary background 5:29
 - general discussion 5:26
 - modern perspectives 5:30
 - origins 5:26
 - theoretical perspectives 5:27*r*
 - transactional leadership 5:27, 5:27*r*
 - transformative leadership 5:27*r*, 5:28
 - transformative learning theory 5:28
 - transforming leadership 5:27, 5:27*r*
- twenty-first century perspectives 4:704–709
 - contextual perspectives 4:704–709
- urban environments 4:809–816
 - challenges 4:810
 - defining characteristics 4:811
 - demographic background 4:809
 - general discussion 4:809
 - leadership strategies 4:815
 - local perspectives
 - estate schooling 4:814, 4:814*f*
 - extended area communities 4:814
 - Leadership on the Front-Line project 4:811
 - multiple communities 4:812, 4:813*f*
 - single/homogenous communities 4:812, 4:812*f*, 4:813*f*
- workforce reform 6:242–248
 - basic principles
 - performance leadership emphasis 6:243
 - qualified teacher status (QTS) 6:242
 - remodelling project 6:242
 - standards 6:242
 - current trends 6:242
 - justification
 - functionalist position 6:244
 - general discussion 6:244
 - socially critical position 6:245
 - outcomes analyses 6:246
- workplace mistreatment/abuse 4:790–796
 - abuse effects 4:793
 - abusive behavior typologies 4:791, 4:792
 - basic concepts 4:790
 - general discussion 4:790
 - modeling approaches 4:791
 - research approaches
 - future directions 4:794
 - methodologies 4:794
 - published results 4:793
 - theoretical perspectives 4:791
- Leadership Development Programme 4:236*r*
- leadership for learning 4:746–751, 4:817–823
 - assessment programs 4:748
 - basic principles 4:820
 - belief and value systems 4:751
 - characteristics 4:747*r*
 - curricular programs 4:748
 - distributed leadership 4:818
 - instructional program 4:747
 - international perspectives 4:820
 - learning communities 4:748
 - nonschool networks 4:820
 - organizational culture 4:750
 - organizational learning 4:817–818, 4:818*f*
 - performance feedback 4:747
 - professional learning 4:817, 4:818*f*
 - research background 4:817
 - resource allocation and use 4:749
 - reward systems 4:750
 - schematic diagram 4:818*f*
 - school–community relations 4:751
 - school-to-school collaboration 6:279, 6:279*f*
 - sense of agency 4:821
 - social advocacy 4:750
 - student leadership for learning 4:819
 - student learning 4:817, 4:818*f*
 - teacher leadership 4:818
 - vision and mission 4:746
- Leadership on the Front-Line project 4:811
- Leading Edge Partnership Program (LEPP) 6:274, 6:275
- lead poisoning 2:425
- Learning Activity Management System (LAMS) 4:424–425
- learning centers 4:437–438, 8:185
- learning cities and regions 1:144–149
 - assessment strategies 1:148
 - basic concepts 1:144
 - current initiatives 1:146
 - future directions 1:148
 - governance 1:145

- learning cities and regions (*continued*)
 - higher education 1:147
 - Pascal virtual community 1:146
 - political factors 1:145
 - research areas 1:146
 - as a social construct 1:144
- Learning Commons 4:430
- learning continuum *see* progress maps
- Learning How to Learn (LHTL) project 3:464
 - background information 3:270, 3:464
 - challenges
 - classroom environments 3:464
 - collaborative classroom learning 3:466
 - fidelity and flexibility considerations 3:465
 - teachers' and students' roles 3:465
 - teachers' assessment practices 3:466*t*
 - importance 3:469
 - supportive school policies and practices 3:466*t*, 3:467, 3:469*f*
 - teachers' professional learning practices 3:466, 3:466*t*
- learning outside school 5:457–463
 - characteristics 5:462
 - formal versus informal learning 5:457
 - general discussion 5:457
 - general versus specific learning 5:459
 - information and communication technology (ICT) 8:9
 - pedagogical–participatory dimension 5:462, 5:462*t*
 - representations versus meaning 5:459, 5:460*t*
 - thought versus action 5:461
 - written versus oral practices 5:458, 5:458*f*, 5:459*t*
- Learning Potential Test for Ethnic Minorities (LEM) 3:222
- Learning Record 3:407
- Learning Stories 3:318, 3:320
- learning strategies 5:323–329
 - assessment approaches
 - instruments 5:328
 - purposes 5:328
 - teaching methods 5:328
 - domain-independent versus domain-dependent strategies 5:326
 - general discussion 5:323
 - historical background 5:324
 - mnemonics 5:324
 - model of strategic learning (MSL) 5:327, 5:327*f*
 - repertoire development 5:326
 - typology
 - comprehension monitoring 5:326
 - elaboration strategies 5:325
 - general discussion 5:324
 - metacognitive strategies 5:326
 - organizational strategies 5:325
 - rehearsal strategies 5:324
 - self-regulation learning strategies 5:325
- learning to learn (L2) 3:268–272
 - current trends
 - collective and institutional weaving concept 3:270
 - EU pilot project 3:270
 - general discussion 3:269
 - individual differences 3:269
 - research areas 3:270
 - developmental psychological traditions 3:269
 - educational impacts 3:271
 - historical background 3:268
 - motivational/volitional traditions 3:269
 - sociocultural traditions 3:269
 - theoretical perspectives 3:268
 - transfer processes 3:271
- learning to read 5:394–400
 - formal reading instruction 5:395, 5:397*f*
 - foundations
 - general discussion 5:394
 - listening comprehension 5:394
 - motivation 5:395
 - oral language skills 5:394
 - print concepts 5:394
 - schematic diagram 5:395*f*
 - semantic development 5:394
 - written language skills 5:394
 - written symbol system 5:394
 - general discussion 5:399
 - neuroimaging studies 5:318
 - preschool education 5:395
 - printed word identification 5:395
 - reading comprehension 5:397
 - reading fluency 5:398, 5:399*f*
 - reading levels 5:398, 5:399*t*
 - see also* reading comprehension
- learning university model 4:249
- Learning Works 8:272
- least squares (LS) criterion 7:306, 7:326, 7:340, 7:483
- least-squares regression 7:135, 7:136*f*
- Lebanon 5:636–655
 - background information 5:636
 - early childhood education and care (ECEC) 2:111*t*
 - economic sectors 5:637*t*
 - education system
 - administrative and supervisory structure 5:642
 - challenges
 - baseline–current–target values 5:654*t*
 - Education Sector reforms 5:646*t*
 - key performance indicators 5:654*t*
 - Paris III reforms 5:645, 5:646*t*
 - pre-Paris III reforms 5:645, 5:646*t*
 - educational finance 5:643, 5:644*t*
 - first foreign language-based student distributions 5:637*t*
 - goals 5:637
 - monitoring and evaluation programs 5:644, 5:645*t*, 5:646*t*
 - National Education Strategy 5:637
 - research programs 5:644
 - structural and operational characteristics
 - currently studying resident distribution by age and educational level 5:640*t*
 - general discussion 5:638
 - higher education 5:641
 - pre-elementary kindergarten 5:638
 - primary education 5:639
 - resident distribution by age 5:639*t*
 - secondary education 5:639
 - student distribution by cycle 5:638*t*
 - student distribution by registration situation 5:639*t*
 - vocational education and training (VET) 5:640, 5:640*t*, 5:641*t*
 - teaching profession
 - characteristics 5:641
 - distribution by age 5:642*t*
 - distribution by category 5:642*t*
 - distribution by degree and number of schools 5:642*t*, 5:643*t*
 - information and communication technology (ICT) 8:25*t*
 - pharmacy education 3:59
 - Trends in International Mathematics and Science Study (TIMSS) 5:645*t*
- legal blindness criteria 2:623
 - see also* blindness and low vision; visual impairment
- legal education 3:25–30
 - accreditation 3:25
 - assessment strategies 3:29
 - case–dialog method 3:26
 - certification and licensing 3:25
 - cognitive apprenticeships 3:26
 - curriculum 3:29
 - educational diversity 1:691–697
 - general discussion 1:696
 - historical perspectives 1:691, 1:823
 - law school students 1:694
 - social science research 1:693
 - educational institutions 3:26
 - instructional strategies
 - clinical programs 3:27
 - delivery strategies 3:29
 - externships 3:27
 - legal writing 3:28
 - practicum courses 3:28
 - simulation courses 3:27–28
 - lawyers 3:25
 - professional identity and values 3:28
 - professional skills and abilities 3:27
 - signature pedagogy 3:26
 - student preparation 3:25
- legal rights 6:17–21
 - children's rights 1:290, 6:17
 - general discussion 6:17
 - moral considerations 6:20
 - parents' rights 1:290, 2:73, 6:19
- Lehrfreiheit* 4:329
- Leipzig Learning Test (LLT) 3:222
- Leithwood, Ken 5:4
- Leonardo da Vinci programme 4:673

- Leontiev, Alexei N. 5:491, 5:494, 5:498, 6:360
- Lesotho
- early childhood education and care (ECEC) 2:111*t*
 - socioeconomic/sociopolitical characteristics 1:199
 - student achievement test performance 2:418*t*
 - see also* Africa
- lesson plans *see* instructional planning
- lesson study 7:575–581
- characteristics 7:575, 7:576*f*
 - future directions 7:579
 - general discussion 7:575
 - historical background 7:575
 - modeling approaches
 - challenges 7:578
 - student learning 7:578
 - teacher learning 7:576
 - three-part teacher learning model 7:577, 7:577*f* - research lessons 7:579
- Leuven Communiqué 4:567
- Levenberg–Marquardt algorithm 7:344
- Levinas, Emmanuel 6:111, 6:451
- Levi-Strauss, Claude 1:500, 6:111
- Lévy, Paul 7:451
- Lewin, Kurt 1:357, 6:312, 7:542
- Liaison Committee on Medical Education (LCME) 3:23
- Liapounov central limit theorem 7:234
- liberalism 6:74–79
- classical liberalism 6:74
 - contemporary liberalism
 - basic concepts 6:76
 - educational perspectives 6:76
 - and Friedman 6:77
 - historical background 6:75
 - and Rawls 6:77–78 - criticisms 6:78
 - economic liberalism 6:75
 - general discussion 6:78
 - historical foundations 6:74
- Libya 2:111*t*
- licensure *see* certification and licensing
- Liechtenstein
- information and communication technology (ICT)
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - student proficiency and confidence 8:31*t*
- life-course perspective 1:585, 5:303
- life-history research 6:430–435
- basic concepts 6:430
 - biographic perspectives 1:26
 - challenges 6:433
 - characteristics 6:431, 6:458
 - data collection and analysis 6:432
 - general discussion 6:434
 - historical foundations 6:431
 - unique contributions 6:430
- lifelong learning 1:89–95, 1:156–161
- academic values and expectations 4:437
 - adult education and training (AET) 1:201, 1:217
 - ageless learners 8:241*f*
 - age-related policies and programs 8:262
 - alternative learning environments 8:184–188
 - computer software 8:185
 - distance education 8:185
 - general discussion 8:184
 - home schooling 6:214–220, 8:184
 - Internet cafes 8:186
 - learning centers 8:185
 - self-directed learning 8:186
 - technical certifications 8:186
 - workplace learning 8:184
- Australia 3:422
- background information 1:183, 6:170
- barriers 1:234–239
- conceptual perspectives 1:234
 - large-scale surveys
 - background information 1:234
 - research findings 1:235, 1:238 - limitations 1:238
 - qualitative studies 1:236
 - theoretical perspectives
 - bounded-agency model 1:238
 - human capital theory 1:237
 - participation research 1:237
- basic concepts 1:2, 1:175
- career guidance 8:269
- competency-based learning model 8:409
- continuing professional development 6:253
- early childhood education and care (ECEC) 2:72–73
- funding considerations 1:133–138
- distribution mechanisms 1:135
 - economic principles 1:133
 - efficiency principles 1:134
- financiers
- general discussion 1:133
 - private financing 1:133
 - public financing 1:134
 - statistical data 1:133
- financing mechanisms
- direct public funding 1:135
 - employers 1:136
 - individuals 1:136
 - partnerships/co-financing 1:135
- infrastructure 1:135
- policies and programs
- affordability considerations 1:137
 - co-financing programs 1:137
 - gap-closing programs 1:136
- Organisation for Economic Cooperation and Development (OECD)
- countries 1:136
 - promotion strategies 1:138
- general discussion 1:160
- globalization effects 1:404, 1:405*t*, 1:699, 6:170–171
- global trends 1:184
- government policies and programs 1:89, 1:93, 1:226, 6:170
- health considerations 1:211–216
- continuing medical education (CME) 1:213
 - current trends 1:211
 - general discussion 1:215
 - health literacy 1:211
- importance 1:234
- initial education 1:90
- institutional models 1:92, 1:227
- learning cities and regions 1:144–149
- assessment strategies 1:148
 - basic concepts 1:144
 - current initiatives 1:146
 - future directions 1:148
 - governance 1:145
 - higher education 1:147
 - Pascal virtual community 1:146
 - political factors 1:145
 - research areas 1:146
 - as a social construct 1:144
- learning to learn (L2) 3:268–272
- current trends
 - collective and institutional weaving concept 3:270
 - EU pilot project 3:270
 - general discussion 3:269
 - individual differences 3:269
 - research areas 3:270 - developmental psychological traditions 3:269
 - educational impacts 3:271
 - historical background 3:268
 - motivational/volitional traditions 3:269
 - sociocultural traditions 3:269
 - theoretical perspectives 3:268
 - transfer processes 3:271
- market-led model
- benefits 1:158
 - characteristics 1:156, 1:157*t*
 - Germany 1:157*t*, 1:159
 - South Korea 1:157*t*, 1:159
- neoliberal ideologies and strategies 1:219, 1:226, 1:699
- older learners 1:175
- partial completion research 8:464
- policy models
- characteristics 1:156, 1:157*t*
 - market-led model 1:156, 1:157*t*
 - social partnership model 1:156, 1:157*t*
 - state-led model 1:156, 1:157*t*
- reflective formative assessments 3:332
- Second Information Technology in Education Studies (SITES)

- lifelong learning (*continued*)
 basic concepts 4:646
 educational impacts 4:667
 importance 4:649
 Lithuania 5:665
 Module 1 (SITES M1) survey results 4:646
 Module 2 (SITES M2) survey results 4:647
 SITES 2006 4:648
 self-assessments 3:233
 social capital 5:113
 social partnership model
 Australia 1:160
 benefits 1:159
 characteristics 1:156, 1:157*t*
 Germany 1:157*t*, 1:159
 socioeconomic modernization 1:219
 Southern African Development Community (SADC) 1:201
 state-led model
 benefits 1:157
 characteristics 1:156, 1:157*t*
 Finland 1:157*t*, 1:158
 France 1:157*t*, 1:158
 Japan 1:157
 Thailand 5:854
 theoretical perspectives 6:170–174
 Habermas's perspective 6:172
 learning theories 6:173
 MacIntyre's perspective 6:171
 panopticon society 6:172
 postmodernism 1:90, 1:220, 6:115
 UNESCO (United Nations Educational, Scientific and Cultural Organization)
 1:89, 6:170
 vocational education and training (VET) 1:157*t*, 1:160, 8:401, 8:470, 8:504
 workplace-based assessments 3:334
 World Bank (WB) 1:89, 1:93, 1:227
see also adult learning; informal learning; nonformal education; university adult
 and continuing education (ACE); vocational education and training (VET)
- Lifelong Learning Institute 1:179
- Lifelong Learning Programme
 background information 4:670
 Comenius programme 4:673
 ERASMUS 4:673
 general discussion 4:673
 Grundtvig programme 4:673
 Jean Monnet programme 4:674
 Leonardo da Vinci programme 4:673
 transversal programme 4:673
- Life Skills Training (LST) program 1:551
- likelihood-ratio tests 7:21–22, 7:190, 7:290
- Lindeberg–Feller central limit theorem 7:234
- linear logistic test model (LLTM) 7:85, 7:89*t*
- linear partial credit model (LPCM) 7:86
- linear regression
 Akaike Information Criterion (AIC) 7:487, 7:487*t*
 basic concepts 7:482
 basis expansion methods 7:138, 7:138*f*
 computational statistics 7:95
 diagnostic tools
 general discussion 7:485
 residuals versus fitted values 7:485*f*
 residuals versus leverage 7:486*f*
 standardized residuals 7:486*f*
- Fisher discrimination
 general discussion 7:136
K-population example 7:137
 two-population example 7:137
- fitted regression models 7:483
- general discussion 7:135, 7:482
- inference estimations 7:484, 7:484*t*
- information criteria 7:487
- large-sample statistical methods 7:236
- least-squares regression 7:135, 7:136*f*
- logistic regression (LR) analysis
 categorical data analysis
 basic concepts 7:61
 differential item functioning (DIF) 7:62
 log-linear model–logistic regression analysis comparisons 7:62
 differential item functioning (DIF) 4:38*t*, 4:39, 7:62
 methodologies 4:167, 7:136, 7:136*f*, 7:138*f*
 test adaptations 4:167
 model comparisons 7:486, 7:487*t*
- multiple linear regression models 7:7
- multivariate linear regression 7:324–331
 basic concepts 7:324
 general discussion 7:330
 hypothesis testing 7:327
 illustrative example 7:328, 7:329*t*, 7:330*t*
 least squares (LS) criterion 7:326
 mathematical formulation 7:325
 statistical inference 7:326
 regression coefficients 7:484*t*
 software packages 7:487
 support vector classifiers 7:137
 univariate linear regression 7:482–488
- linear structural equation models 7:453
- linguistic phenomenology 6:451
- linking social capital 6:98
- linking ties 5:114
- Lipps, Theodore 6:598
- Lisbon Programme 4:537, 4:569, 4:671, 8:435
- listening comprehension 3:285, 5:394
- literacy conferences 3:403–409
 basic concepts 3:404
 constructive responses 3:275
 definitions 3:403
 historical background 3:404
 informal assessment systems 3:405
 parental involvement 3:406
 portfolio assessments 3:406
 purpose 3:403
 recent developments
 general discussion 3:407
 Learning Record 3:407
 New Zealand Education Monitoring Project (NZEMP) 3:408
 self-assessments 3:406
 validity 3:408
- literacy instruction
 beginning-reading programs
 combined programs 6:222
 computer-assisted instruction 6:222
 cooperative learning 6:222
 cost effectiveness 6:223
 curricula 6:221
 instructional-process programs 6:222
 kindergarten-only programs 6:223
 research background 6:221
 research results 6:223
- biliteracy
 continua model 1:477
 general discussion 1:477
 integrated instructional strategies 1:477
- computer-assisted instruction 8:150–157
 beginning-reading programs 6:222
 comprehension 8:152
 educational impacts 5:121, 5:121*f*
 emergent literacy 8:150
 fluency 8:151
 general discussion 8:150, 8:155
 Internet access 8:154
 phonics and phonemic awareness 8:151
 software systems 8:3
 text-to-speech applications 8:153
 theoretical perspectives 8:154
 upper elementary grades 6:224
 vocabulary 8:152
- general discussion 6:225
- primary schools 1:437
- Scholar–Baller® program 1:861
- secondary schools 1:438
- socioeconomic modernization 1:217
- special needs students 2:759–766
 importance 2:759
 instructional strategies
 cognitive strategy instruction (CSI) 2:759, 2:760*t*
 directed instruction (DI) 2:759, 2:760*t*
 reading skills 2:759
 vocabulary 2:763
 writing skills 2:764
- reading skills
 beginning reading 2:759, 2:761*t*
 fluency 2:761
 instructional strategies 2:759

- peer-assisted learning strategies (PALS) 2:761–762, 2:762*t*
- reading comprehension 2:762
- think before reading, think while reading, think after reading (TWA)
 - strategy 2:762, 2:763*t*
- writing skills
 - instructional strategies 2:764
 - transcription skills 2:765
 - written expression 2:764, 2:764*t*
- upper elementary grades
 - combined programs 6:225
 - computer-assisted instruction 6:224
 - cooperative learning 6:224
 - curricula 6:223
 - instructional-process programs 6:224
 - requirements 6:223
 - workplace learning 1:86
- literary canon 1:438
- literature curricula 1:447–452
 - colonial and postcolonial periods 1:449
 - critical theory-based literature 1:448
 - economic factors 1:450
 - historical background
 - general discussion 1:447
 - literary texts 1:447
 - primers 1:447
 - reader-centered approaches 1:447
 - text-based approaches 1:447
 - international perspectives 1:448
 - political factors 1:451
 - political stability 1:449
 - wartime effects 1:450
- Lithuania 5:656–665
 - age-related policies and programs 8:263*t*
 - background information
 - demographic data 5:657
 - economic factors 5:657
 - general discussion 5:656
 - government 5:657
 - historical background 5:656
 - nationalities 5:657
 - religion 5:657
 - early childhood education and care (ECEC) 2:111*t*
 - educational policies 5:657
 - education system
 - administrative and supervisory structure 5:662
 - assessment programs 5:661
 - availability 5:661
 - educational finance
 - financing sources 5:663
 - investment programs 5:663
 - material support 5:664
 - tuition and fees 5:664
 - wages and salaries 5:664
 - evaluation policies and practices 5:664
 - goals 5:658
 - monitoring and evaluation programs 5:663
 - program selection options 5:661
 - quality assessments 5:662
 - research programs 5:665
 - school self-governance 5:663
 - structural and operational characteristics
 - adult education and training (AET) 5:661
 - adult participation 5:660*t*
 - basic education 5:660
 - education levels 5:658*f*, 5:659*f*
 - general discussion 5:658
 - higher education 5:660
 - number of learners 5:659*t*
 - number of schools 5:659*t*
 - number of teachers 5:659*t*
 - preprimary education 5:658
 - preschool education 5:658
 - primary schools 5:658
 - secondary schools 5:660
 - special education 5:661
 - vocational education and training (VET) 5:660
- inequality–education relationship 2:273*f*
- information and communication technology (ICT)
 - computer access at home 8:25*t*
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - mathematics and science instruction 8:32*t*
 - student proficiency and confidence 8:31*t*
- information and communication technology (ICT) access and equity 8:97
- teachers and pedagogical staff
 - professional development 5:662
 - retraining 5:662
 - training 5:662
 - work hours/work load 5:662
- test score–social background relationship 2:273*f*
- vocational education and training (VET) 8:263*t*
- local average treatment effect (LATE) 7:226
- localism *see* new localism
- local norms 4:97
- local teacher-based curriculum development 1:283–286
 - general discussion 1:283, 1:286
 - historical background 1:283
 - illustrative example 1:284
 - professionalization process 1:283–284
- Locke, Alain 1:78
- Locke, John 3:316–317, 6:74, 8:441
- locus of control 6:559
- logic models 3:737
- logistic models 7:343
- logistic random-intercept model 7:171, 7:172*f*, 7:172*t*
- logistic regression (LR) analysis
 - categorical data analysis
 - basic concepts 7:61
 - differential item functioning (DIF) 7:62
 - log-linear model–logistic regression analysis comparisons 7:62
 - differential item functioning (DIF) 4:38*t*, 4:39, 7:62
 - methodologies 7:136, 7:136*f*, 7:138*f*
 - test adaptations 4:167
- logit model 7:465
- log-logistic model 7:464
- lognormal distribution 7:101
- Logo (computer software) 8:175, 8:199
- London Borough of Tower Hamlets 6:293
- London Communiqué 1:773*f*, 4:567
- London School of Economics 6:12
- London Times* rankings 4:379
- loneliness 6:693
- longitudinal studies 3:757–764
 - academic achievement outcomes 2:501
 - educational barriers 1:234
 - Educational Diversity Project (EDP) 1:694
 - educational evaluation project 3:706
 - Effective Provision of Preschool Education (EPPE) 2:86–92
 - effectiveness evaluations 2:87
 - general discussion 3:320–321
 - OECD curriculum guidelines 2:90, 2:91*t*
 - pedagogical practices
 - adult–child interactions 2:88
 - curriculum differentiation 2:89
 - discipline/behavior management policies 2:89
 - formative assessments 2:89
 - parental involvement 2:89
 - preschool education 2:78
 - qualitative case studies 2:87
 - quality and diversity evaluations 2:87
 - results analysis 2:90
 - success factors
 - leadership and staff 2:89
 - pedagogical content knowledge 2:89
 - general discussion 3:757
- Learning How to Learn (LHTL) project 3:270, 3:464
- longitudinal/repeated measures model 3:120, 3:121*t*, 7:212
- mathematics instruction 2:767
- multivariate longitudinal data analysis 3:120–127
 - general discussion 3:126
 - hierarchical linear models
 - basic concepts 3:120
 - self-concept study 3:121, 3:121*t*, 3:122*t*
 - variable measures 3:121*t*, 3:122*t*
 - latent growth curves model (LGCM)
 - basic concepts 3:122, 3:122*f*, 3:123*f*
 - self-concept study 3:123, 3:123*t*, 3:124*t*
 - research background 3:120
 - structural cross-lagged regression model
 - basic concepts 3:124, 3:124*f*, 3:125*f*
 - self-concept study 3:125, 3:126*t*
- national data systems 2:179, 2:181*t*

- longitudinal studies (*continued*)
 - neuroscience research 5:321
 - research collection designs
 - cohort designs 3:758
 - cross-sectional designs 3:758
 - general discussion 3:757
 - panel designs 3:758
 - time-series designs 3:759
 - trend designs 3:758
 - social networks 6:706
 - special needs students 2:716
 - statistical analysis techniques
 - cross-lagged model 3:762, 3:763f, 3:764f
 - general discussion 3:759
 - multilevel growth model 3:760, 3:761t
 - survival model 3:761, 3:762t
 - student attrition rates 4:468
 - survival data analysis 7:459
 - teacher incentives 2:483
 - test-based accountability 4:154
 - long-term memory 5:198–199, 5:200f, 5:246, 5:341
 - Lord chi-squared (χ^2) statistic 4:38, 4:38t
 - Lord's paradox 7:429
 - lotteries 2:224
 - Louisiana 1:654f
 - low birth weight effects 2:427
 - low-incidence disabilities 2:855
 - low-technology assistive devices 2:644
 - low vision **2:880–885**
 - educational needs 2:881
 - etiologies
 - definitions 2:880
 - functional implications 2:881
 - prevalence 2:880
 - general discussion 2:884
 - government policies and programs 2:880
 - instructional needs 2:883
 - learning materials accessibility
 - auditory readers 2:883
 - Braille systems 2:883
 - closed circuit televisions (CCTVs) 2:882–883
 - general discussion 2:882
 - optical devices 2:882–883
 - occurrences 2:880
 - school-based services 2:854
 - special education background 2:715
 - Luria, AR 5:498, 6:360
 - Luther, Martin 6:413
 - Luxembourg
 - age-related policies and programs 8:263t
 - class size 2:492f
 - early childhood education and care (ECEC) 2:4t, 2:111t
 - earnings–education level relationship 2:307t
 - education levels 2:279f
 - enrollment rates 2:493f
 - inequality–education relationship
 - employment rate–educational attainment relationship 2:278f
 - immigrant status 2:275f
 - information and communication technology (ICT) 8:23t
 - information and communication technology (ICT) access and equity 8:97
 - participation research
 - regular education 8:284f
 - vocational education and training (VET) 8:283f, 8:284f
 - work-related vocational education and training (VET) 8:284f, 8:285f
 - student and researcher mobility 4:287f
 - teaching hours 2:493f
 - vocational education and training (VET) 8:251t, 8:253f, 8:263t
 - Lyapunov's inequality 7:428
 - lycées d'enseignement général et technologique* (LEGT) 5:578
 - lycées professionnels* (LP) 5:578
 - Lyotard, Jean-François 6:14, 6:111, 6:185–186
- M**
- Maastricht Communiqué (2004) 8:435
 - Maastricht Treaty (1992) 4:671
 - Macao
 - early childhood education and care (ECEC) 2:111t
 - inequality–education relationship 2:275f
 - information and communication technology (ICT)
 - computer access in school 8:23t
 - frequency of computer use in school and at home 8:26t
 - student proficiency and confidence 8:31t
 - Macedonia
 - early childhood education and care (ECEC) 2:111t
 - inequality–education relationship 2:273f
 - test score–social background relationship 2:273f
 - MacGregor, Frances 6:499
 - MacIntyre, Alasdair 6:111, 6:171, 6:183, 6:187
 - Madagascar
 - early childhood education and care (ECEC) 2:111t
 - student achievement test performance 2:418t
 - Magna Charta Universitatum* 4:201
 - magnetic resonance imaging (MRI) 5:317, 6:501
 - see also* neuroscience research
 - magnetic source imaging (MSI) 5:317
 - see also* neuroscience research
 - magnetoencephalography (MEG) 5:317
 - see also* neuroscience research
 - Mahalanobis D_M 7:337, 7:445
 - Maine 1:654f
 - Maine Learning Technology Initiative 8:128
 - mainstreaming 2:721, 6:160
 - malaria 1:602
 - Malawi **5:666–672**
 - background information 5:666
 - education system
 - administrative and supervisory structure 5:671
 - challenges 5:671
 - educational finance 5:671
 - policies and reforms 5:666
 - secondary education
 - access and equity patterns 5:668
 - structural and operational characteristics 5:668
 - teaching quality 5:668
 - structural and operational characteristics
 - adult education 5:670
 - basic education 5:667
 - continuing education 5:669
 - distance education 5:669
 - early childhood education and care (ECEC) 5:667
 - general discussion 5:667
 - higher education 5:671
 - Ministry of Education, Science and Technology (MOEST) 5:667
 - nonformal education 5:670
 - primary education 5:667
 - secondary education 5:668
 - special education 5:670
 - vocational education and training (VET) 5:669
 - teaching profession 5:670
 - school fees 2:419
 - socioeconomic/sociopolitical characteristics 1:199
 - student achievement test performance 2:418t
 - see also* Africa
 - Malaysia
 - accreditation processes **4:384–389**
 - early childhood education and care (ECEC) 2:111t
 - higher education
 - collective bargaining 4:495
 - enrollment rates 4:522t
 - medium of instruction (MOI) 4:205
 - private higher education (PHE) 4:623
 - private sector services 4:524t
 - quality assurance 4:525
 - student mobility 4:561
 - inequality–education relationship 2:273f
 - information and communication technology (ICT) 8:25t
 - mathematics curricula 1:454
 - music teacher education 7:706
 - student achievement test performance 2:247f, 2:417t
 - test score–social background relationship 2:273f
 - Maldives 2:111t
 - male underachievement **1:838–844**
 - future research areas 1:842
 - general discussion 1:838
 - influencing factors
 - classroom behavior
 - cool and dumb behaviors 1:841
 - student engagement 1:841
 - curriculum and class structure 1:842
 - family composition 1:840

- identity development
 - general discussion 1:840
 - hegemonic masculinity 1:840
 - racialized masculinity 1:840
 - teachers' perceptions 1:841
- research focus 1:839
- research perspectives 1:839
- Mali
 - early childhood education and care (ECEC) 2:111*t*
 - student achievement test performance 2:418*t*
- Malta 5:673–680
 - age-related policies and programs 8:263*t*
 - background information 5:673
 - early childhood education and care (ECEC) 2:111*t*
 - education system
 - challenges 5:679
 - enrollment 5:674*t*
 - historical background 5:673
 - Malta College of Arts, Science and Technology (MCAST) 5:678
 - postcolonial theory 1:300
 - structural and operational characteristics
 - adult education 5:679
 - continuing education 5:679
 - directorates responsibilities 5:675*f*
 - early childhood education and care (ECEC) 5:675, 5:676*f*
 - examination systems 5:677
 - general discussion 5:674
 - kindergarten education 5:675, 5:676*f*
 - postsecondary institutions 5:676*f*, 5:678
 - primary schools 5:676, 5:676*f*
 - schematic diagram 5:676*f*
 - secondary schools 5:676*f*, 5:677
 - special needs students 5:676, 5:676*f*
 - students and faculty 5:674*t*
 - teaching profession 5:679
 - University of Malta 5:678
- management by objectives (MBO) 7:534
- managerial leadership 4:759, 5:45
- Mandela, Nelson 8:443
- Mann–Whitney *U* test 7:348
- Manpower Development Training Act (1962) 8:277
- Mantel–Haenszel (MH) statistic 4:38, 4:38*t*, 4:167
- manual communication 2:545, 2:624
- Māori culture 8:291*t*
- Māori students *see* New Zealand
- Marcuse, Herbert 6:133, 6:349
- marginal maximum likelihood (MML) method 4:53, 7:286, 7:351, 7:371
- market economy 7:533
- marketized childcare 2:155
- Markov chain Monte Carlo (MCMC) 7:245–252
 - applications
 - item response theory (IRT) 7:251
 - sensitivity analyses 7:251
 - background information 7:245
 - beta distribution 7:99
 - characteristics 7:302
 - general discussion 7:10
 - Gibbs sampling 7:246, 7:247*f*
 - implementation
 - convergence assessments 7:250
 - model selection 7:250
 - sample quality 7:250
 - sample usage 7:250
 - software packages 7:250
- Markov chains 7:246
- Metropolis–Hastings algorithm 7:247, 7:249*f*
- missing data handling techniques 7:286
- order statistics 7:361
- recent developments 7:250
- simulated annealing (SA) algorithm 7:248
- Markov chains 7:451
- Markov's inequality 7:426
- Markus–Kitayama concept of self-construal 5:453
- Marland Report 2:870, 2:871
- Marshall, Alfred 2:194, 2:282
- Marshall Islands 2:111*t*
- Martin, Jane Roland 6:59
- Marxism 6:133
- Marx, Karl 6:111, 6:349
- Maryland 1:654*f*
- Massachusetts
 - Asian American/Pacific Islander student enrollment 1:597*f*
 - Boston Public Schools (BPS) 6:299
 - tuition and needs-based grants 1:654*f*
- mass media
 - academic achievement outcomes 8:109–111
 - challenges 8:110
 - general discussion 8:109
 - historical background 8:109
 - research background 8:110
 - theoretical perspectives
 - displacement theory 8:110
 - information-processing theory 8:110
 - interest-stimulation theory 8:110
 - short-term gratifications theory 8:110
 - music teacher education 7:709
- master of business administration (MBA) 3:39–40
- mastery goals 6:537–538
- mastery testing 4:32
- math anxiety 6:551
- mathematics education 5:401–406
 - ability testing 4:4
 - assessment strategies 3:279–284
 - accountability 3:280, 5:404
 - curriculum-based measurement (CBM) 2:771, 4:108
 - future directions 3:283
 - high-stakes test score inflation 4:187*f*, 4:187*t*
 - trends
 - 1970–1990 3:279
 - 1990–2008 3:280
 - accountability testing 3:280
 - computer-based testing 3:281
 - formative assessments 3:282
 - information and communication technology (ICT) 3:281
 - summative assessments 3:281
- CompEd study
 - characteristics 3:687*t*
 - Functional Information Technology Test (FITT) 4:645
 - importance 4:649
 - purpose 4:644
 - Stage 1 survey results 4:644
 - Stage 2 survey results 4:645
- competence, learning, intervention, and assessment (CLIA) model 5:405
- competency characteristics 5:401
- computer-assisted instruction 8:172–183
 - challenges 8:180
 - classroom network technologies 8:136
 - collaborative problem solving
 - The Adventures of Jasper Woodbury* 8:177
 - Anchored Instruction 8:177
 - Enhanced Anchored Mathematics 8:177, 8:178*f*
 - MathWorlds 8:178
 - networked devices 8:177
 - participatory simulations 8:177–178
 - serious game systems 8:178
 - SimCalc 8:178, 8:179*f*, 8:199
 - embodied cognition 8:179, 8:180*f*
 - importance 8:172
 - mathematical discovery
 - ANIMATE system 8:175, 8:176*f*
 - Dynamic Geometry systems 8:176
 - dynamic statistics packages 8:176
 - Function Probe 8:175
 - general discussion 8:174
 - Logo 8:175, 8:199
 - skills and concept learning
 - background information 8:172
 - Cognitive Tutors 8:173, 8:174*f*
 - computer algebra systems (CASs) 8:173
 - Geometric Supposer 8:173
 - hand-held graphing calculators 8:172, 8:173*f*
 - intelligent tutoring systems (ITSs) 8:173
 - subject-specific uses 8:4
- constructivist learning theories
 - basic concepts 5:485
 - characteristics 7:750*t*
 - children's ideas 5:488
 - core ideas 5:486
 - general discussion 5:485
 - intellectual development stages 5:487
 - Piagetian perspective 1:488–489, 5:486

- mathematics education (*continued*)
 Vygotskian perspective 1:489, 5:488
 curriculum 1:453–458
 attained curriculum evaluation
 assessment tasks 3:657
 conceptual diagram 3:654*f*
 importance 3:657
 performance analysis 3:657
 characteristics 3:654
 curriculum development and revision
 general discussion 1:454
 Singapore 1:455, 1:455*f*, 1:456*f*
 United States 1:456
 curriculum reform 1:457
 historical perspectives 1:453
 implementation considerations 1:457
 implemented curriculum evaluation 3:654*f*, 3:655, 3:656*t*, 3:657*f*
 intended curriculum evaluation 3:654, 3:654*f*, 3:655*f*, 3:656*t*
 selection guidelines 1:453
 female participation 1:687
 gender equity 1:630, 1:686
 general discussion 5:401
 instructional design processes 5:403
 International Association for the Evaluation of Educational Achievement (IEA)
 mathematics and science studies 4:650–655
 background information 4:650
 early studies 4:650, 4:664
 First International Mathematics Study (FIMS) 4:650, 4:664
 First International Science Study (FISS) 4:651
 general discussion 4:654
 Second International Mathematics Study (SIMS) 4:651
 Second International Science Study (SISS) 4:651
 survey results and analysis 4:653
 Third International Mathematics and Science Study 1995 (TIMSS) 4:652
 Third International Mathematics and Science Study 2003 (TIMSS) 4:653
 Third International Mathematics and Science Study 2007 (TIMSS) 4:653
 Third International Mathematics and Science Study Repeat (TIMSS-R) 4:652
 long division/gazzinta process 6:362, 6:365*t*
 mathematics learning processes 5:402
 neural networks 5:312
 out-of-school learning
 representations versus meaning 5:459, 5:460*t*
 thought versus action 5:461
 written versus oral practices 5:458, 5:458*f*, 5:459*t*
 program evaluation 3:653–659
 attained curriculum
 assessment tasks 3:657
 conceptual diagram 3:654*f*
 importance 3:657
 performance analysis 3:657
 curriculum
 attained curriculum 3:654*f*, 3:657
 characteristics 3:654
 content analysis guidelines 3:656*t*
 implemented curriculum 3:654*f*, 3:655, 3:656*t*, 3:657*f*
 intended curriculum 3:654, 3:654*f*, 3:655*f*, 3:656*t*
 formative and summative evaluations 3:653
 globalization effects 3:658
 research background 3:653
 Second Information Technology in Education Studies (SITES)
 basic concepts 4:646
 educational impacts 4:667
 importance 4:649
 Lithuania 5:665
 Module 1 (SITES M1) survey results 4:646
 Module 2 (SITES M2) survey results 4:647
 SITES 2006 4:648
 special needs students 2:767–773
 background information 2:767
 general discussion 2:772
 historical research 2:768
 instructional strategies
 cognitive strategy instruction (CSI) 2:770
 concrete-to-representational-to-abstract (CRA) sequence 2:769
 Enhanced Anchored Instruction (EAI) 2:771
 formative assessments 2:771
 general discussion 2:768
 graduated guidance sequence 2:769
 graphic organizers 2:769, 2:769*f*
 peer-mediated instruction 2:770
 schema-based instruction 2:769
 self-regulation 2:770
 women of color 1:688
 mathematics models
 general discussion 5:481
 knowledge-based cascade-correlation (KBCC) 5:482, 5:482*f*
 multiplication 5:481, 5:481*f*, 5:482*f*, 5:482*t*
 primality 5:482
 mathematics teacher education 7:702–705
 future directions 7:704
 general discussion 7:702
 learning trends
 general discussion 7:702
 reflective practices 7:703
 social dimensions 7:702
 teaching conditions 7:704
 professional development 7:552
 research areas 7:702
 Mathematics Test Creation Assistant 8:51
 MathWorlds 8:178
 matrix algebra 7:253–259
 basic concepts
 determinants 7:254
 inverse matrix 7:254
 operations 7:253, 7:255*t*
 positive-definite/positive-semidefinite matrices 7:255
 rank 7:253
 vector spaces 7:254
 basic types 7:254*t*
 definitions 7:253
 eigenvalues/eigenvectors 7:256
 inverse matrix
 basic concepts 7:254
 generalized inverse matrix
 least-squares g-inverse matrix 7:257
 minimum norm g-inverse matrix 7:257
 reflexive g-inverse matrix 7:257
 weaker g-inverse matrix 7:257
 Moore–Penrose inverse matrix 7:258
 Kronecker product 7:258
 methodologies
 correlation matrix 7:258, 7:258*t*, 7:375, 7:376*t*
 principal component analysis (PCA) 7:258, 7:258*t*, 7:259*t*
 orthogonal projector matrix 7:256
 singular-value decomposition 7:256
 spectral decomposition 7:256
 Vec operator 7:258
 Matura examination system 3:474
 maturity examinations 5:541, 5:660
 Mauritania
 early childhood education and care (ECEC) 2:111*t*
 student achievement test performance 2:418*t*
 Mauritius
 early childhood education and care (ECEC) 2:111*t*
 socioeconomic/sociopolitical characteristics 1:199
 student achievement test performance 2:418*t*
 see also Africa
 maximum-information criterions 4:87
 maximum likelihood (ML) theory
 computerized adaptive testing (CAT) 7:397
 conditional maximum likelihood (CML) method 4:53, 7:371
 expectation–maximization (EM) 7:242
 general discussion 7:165, 7:306
 joint maximum likelihood (JML) method 4:53, 7:370
 marginal maximum likelihood (MML) method 4:53, 7:286, 7:351, 7:371
 maximum likelihood estimation (MLE)
 basic concepts 7:341
 empirical Bayes methods 7:147, 7:147*f*
 extreme value theory 7:26
 generalized linear models (GLMs) 7:175
 inference estimations 7:179, 7:368, 7:483
 item response theory (IRT)
 conditional maximum likelihood (CML) method 4:53, 7:371
 empirical Bayes methods 7:146, 7:147*f*
 illustrative example 7:372, 7:372*t*, 7:373*t*
 joint maximum likelihood (JML) method 4:53, 7:370
 marginal maximum likelihood (MML) method 4:53, 7:351, 7:371
 missing data 7:283
 test equating and linking 4:53
 latent class (LC) models 7:242
 logistic random-intercept model 7:171, 7:172*f*, 7:172*t*

- multivariate linear regression 7:326
- Newton–Raphson algorithm 7:242
- shrinkage effects 7:144^f
- structural equation models (SEMs) 7:455
- survival data analysis 7:464
- Mead, George Herbert 3:167, 6:88, 6:481, 6:483
- Mead, Margaret 6:499
- mean
 - alternative versions
 - general discussion 7:265
 - geometric mean 7:265
 - harmonic mean 7:265
 - robust estimation 7:266
 - trimmed mean 7:266
 - weighted mean 7:266
 - Winsorized mean 7:266
 - basic concepts 7:264
 - characteristics 7:265
- mean squared error (MSE) 7:367
- Measures of Academic Progress (MAP) assessment 8:49
- measures of association 7:260–263
 - continuous variables 7:260
 - discrete nominal variables
 - basic concepts 7:261
 - Cochran–Mantel–Haenszel (CMH) test 7:262, 7:262^t
 - Kappa (κ) statistic 7:263^t
 - odds ratio 7:262^t
 - Simpson’s paradox 7:262
 - discrete ordinal variables 7:261
 - educational applications 7:3
 - general discussion 7:260
- measures of central tendency 7:264–266
 - basic concepts
 - general discussion 7:264
 - mean 7:264
 - median 7:264
 - mode 7:264
 - data distribution 7:264
 - dispersion measures 7:265
 - general discussion 7:264
 - mean
 - alternative versions
 - general discussion 7:265
 - geometric mean 7:265
 - harmonic mean 7:265
 - robust estimation 7:266
 - trimmed mean 7:266
 - weighted mean 7:266
 - Winsorized mean 7:266
 - basic concepts 7:264
 - characteristics 7:265
- median
 - basic concepts 7:264
 - characteristics 7:265
- mode
 - basic concepts 7:264
 - characteristics 7:265
- media communication
 - academic achievement outcomes 8:109–111
 - challenges 8:110
 - general discussion 8:109
 - historical background 8:109
 - research background 8:110
 - theoretical perspectives
 - displacement theory 8:110
 - information-processing theory 8:110
 - interest-stimulation theory 8:110
 - short-term gratifications theory 8:110
 - music teacher education 7:709
- median
 - basic concepts 7:264
 - characteristics 7:265
- mediated actions 5:232, 5:499, 5:501
- Medical College Admission Test (MCAT) 4:9
- medical education 3:19–24
 - accreditation organizations 3:23
 - admission requirements 3:19
 - certification and licensing 3:23
 - curriculum
 - characteristics 3:20
 - clinical education 3:20
- learning processes
 - competency-based learning methods 3:22
 - problem-based learning methods 3:22
 - simulation methods 3:21
- reform efforts 3:20
- general discussion 3:23
- goals 3:19
- residency education 3:22
- medical insurance 2:426
- medieval universities 1:728
- mega-universities 4:406
- Meier, G. F. 6:414
- membership-based learning 1:192
- memory 5:197–203, 5:285–289
 - age-related effects 5:342
 - background information 5:285
 - classroom learning 5:201, 5:202^f
 - cognition–emotion relationship 5:240
 - cognitive decline 5:341
 - declarative memory 5:198, 5:198^f
 - discursive perspective
 - background information 5:285
 - conversational interactions 5:286
 - general discussion 5:288
 - identity membership 5:287
 - inference estimations 5:287
 - reminiscences 5:287
 - sequential organization 5:286
 - socially organized settings 5:287
 - episodic memory 5:198, 5:198^f, 5:199, 5:200^f, 5:342
 - everyday memories 5:200
 - executive functions 5:342
 - false autobiographical memories 5:201
 - familiarity 5:342
 - future research areas 5:202
 - general discussion 5:197
 - human memory types 5:198, 5:198^f
 - intelligence theories 5:185
 - learning strategies 5:324
 - long-term memory 5:198–199, 5:200^f, 5:341
 - metaphorical models 5:199
 - neuroscience research 5:314
 - nondeclarative memory 5:198, 5:198^f, 5:342
 - procedural memory 5:198, 5:198^f
 - recollection 5:342
 - self-regulated learning 5:505
 - semantic memory 5:198, 5:198^f, 5:342
 - short-term memory 5:198–199, 5:341
 - thought versus action 5:461
 - working memory (WM) capacity 5:186, 5:198–199, 5:225, 5:246, 5:342, 8:200
- mental age 5:188
- mental health services 3:557
- mental retardation (cognitive disability) 2:578, 2:853
- mentoring strategies 7:749–756
 - career guidance 8:271
 - characteristics 7:750, 7:750^t
 - contemporary definitions 2:163
 - definitions 2:163
 - e-mentoring 2:164, 2:166
 - functional role 2:167
 - general discussion 7:749, 7:756
 - group and peer mentoring 2:163
 - immigrant populations 4:633
 - informal learning 8:243–244
 - learning themes 7:751^t
 - limitations 7:755
 - mentoring variations 7:750^t, 7:751
- models
 - curriculum frameworks
 - general discussion 2:165
 - group and peer mentoring 2:165^f, 2:166
 - group mentoring 2:165, 2:165^f
 - mentor characteristics 2:166
 - single mentors 2:165, 2:165^f
 - structural characteristics 2:165, 2:165^f
 - web-based support 2:166
 - general discussion 2:164
 - group-based e-mentoring
 - characteristics 2:166
 - working communities 2:166
- origins 7:749

- mentoring strategies (*continued*)
 - professional development 2:163
 - recruitment and retention 2:477
 - reflective practices 2:164
 - research areas
 - disadvantages
 - effects on beginning teachers 7:755
 - effects on mentors 7:755
 - general discussion 7:755
 - effectiveness
 - educational system benefits 7:753
 - general discussion 7:752
 - mentee benefits 7:752
 - mentor benefits 7:752
 - general discussion 7:752
 - mentoring conditions
 - contextual support 7:753
 - mentoring strategies 7:754
 - preparation and support 7:754
 - selection and pairing processes 7:753
 - social and emotional outcomes 6:675
 - teacher induction programs 2:477, 7:566
 - terminology and jargon sources 8:270, 8:350*r*
 - mergers and acquisitions 4:512
 - merit pay 2:481
 - Merleau-Ponty, Maurice 6:81, 6:449, 6:451
 - Mesure et évaluation en éducation* (MÉE) 3:432
 - meta-analysis 7:274–282
 - background information 7:274
 - effect-size indices
 - d* family 7:276
 - general discussion 7:276
 - r* family 7:276
 - external validity 6:524
 - general discussion 7:281
 - hierarchical linear models 7:214
 - illustrative example
 - analysis of variance (ANOVA) 7:281*t*
 - effect-size plot 7:280*f*
 - fixed-effect/random-effect models 7:280*r*
 - general discussion 7:279
 - moderator variables 7:280*r*
 - regression models 7:281*t*
 - process steps 7:275
 - statistical procedures
 - average effect-size calculations 7:277
 - effect-size heterogeneity assessments 7:278
 - general discussion 7:277
 - moderator variable influence analysis 7:279
 - statistical conclusion validity 3:150
 - metacognition 5:204–210
 - academic domains
 - general discussion 5:206
 - mathematics research 5:208
 - reading research
 - comprehension monitoring 5:207
 - general discussion 5:206
 - intervention studies 5:207
 - metacognitive knowledge 5:206
 - science research 5:208
 - writing research 5:207
 - assessment methods
 - general discussion 5:205
 - judgements of learning (JOL) tasks 5:206
 - knowledge monitoring 5:206
 - online processing 5:206
 - verbal reports 5:205
 - characteristics 5:204
 - general discussion 5:208
 - historical background 5:205
 - International System for Teacher Observation and Feedback (ISTOF) project 3:733
 - memory 5:286
 - motivational regulation 6:633
 - music learning 5:409
 - personalized learning 6:230
 - research background 5:248
 - self-regulated learning 5:503–508
 - metaevaluation 3:765–774
 - conceptual frameworks
 - characteristics 3:804
 - pseudoevaluations 3:804, 3:805*r*
 - quasi-evaluations 3:804, 3:805*r*
 - true evaluations 3:805, 3:805*r*
 - e-learning 3:534, 3:536
 - general discussion 3:765, 3:773
 - practice background 3:770
 - prescriptive perspective
 - general discussion 3:767
 - methodological approaches 3:769
 - Program Evaluation Standards 3:767, 3:768*r*
 - questions and criteria 3:767
 - purpose
 - formative metaevaluation 3:765
 - general discussion 3:765
 - research evaluations 3:766
 - summative metaevaluation 3:766
 - syntheses 3:766
 - standard practice guidelines
 - conflict of interest 3:772
 - cost effectiveness 3:772
 - disclosure of findings 3:772
 - evaluator credibility 3:771
 - human interactions 3:772
 - political viability 3:772
 - practical procedures 3:772
 - values identification 3:771
 - Metroplitan Achievement Test 3:189
 - Metropolis–Hastings algorithm 7:247, 7:249*f*
 - Mexican Americans
 - cultural difference theory model 2:776
 - deaf-blind population 2:623*r*
 - educational opportunities 1:881
 - educational statistics 1:581–588
 - choice effects 1:585
 - continuing disparities 1:584
 - demographic diversity 1:582
 - digital technology implications 1:584
 - disparity measures 1:581, 1:874
 - dropout rates 1:874, 1:875
 - education beyond borders 1:586
 - global village concept 1:587
 - graduation rates 1:874
 - incarceration rates 1:874
 - twenty-first century perspectives 1:581, 1:587
 - United States 1:582
 - identity politics 6:39
 - oppositional identity 1:756, 2:776
 - race-specific consequences 1:876
 - valued peers 1:757
 - Mexico
 - age-earnings profiles 2:306*f*
 - assessment system development 3:482
 - childhood brain development 2:57*f*
 - class size 2:492*f*
 - competency-based learning model 8:407
 - conditional cash transfer (CCT) programs 2:419
 - early childhood education and care (ECEC) 2:4*t*, 2:111*t*, 2:114*r*
 - earnings–education level relationship 2:306*r*
 - education levels 2:279*f*
 - enrollment rates 2:493*f*
 - foreign doctoral students 4:292*r*
 - higher education
 - academic values and expectations 4:438
 - collective bargaining 4:496
 - enrollment rates 1:702*f*, 1:703*f*
 - federal political systems 4:607
 - institutional expenditures 1:733*r*
 - loans and grants 4:501*f*, 4:502*f*
 - parental expenditures 1:734*r*
 - social change 1:719–727
 - demographic structure 1:720, 1:721*t*
 - economic dynamics 1:720
 - employment trends 1:722–724, 1:722*r*
 - future directions 1:726
 - industrial structure 1:722, 1:722*t*, 1:723*r*
 - labor–market perspective 1:719
 - university–industry collaboration 1:724
 - tertiary education expenditures 4:500*f*
 - tertiary type-B program survival rates 8:464*r*
 - indigenous peoples 1:664*r*
 - inequality–education relationship 2:278*f*

- information and communication technology (ICT) 8:23*t*
 public subsidization 2:308*t*
 returns to education—education level 2:308*t*
 school supply shortages 2:506*f*
 student achievement test performance 2:247*f*
 student and researcher mobility 4:287*f*
 teacher education programs 7:528
 teacher incentives 2:482, 2:485*t*, 2:507
 teacher shortages 2:494*f*
 teaching hours 2:493*f*
 vocational education and training (VET) 8:251*t*, 8:252*f*, 8:253*f*, 8:407
 women in employment 2:56*f*
- Mezirow, J 1:19, 1:37
 Michaelis–Menten regression function 7:341, 7:342*f*
 Michigan 1:654*f*, 2:65
 Michigan School Readiness Program 2:451–452
 microethnography studies 6:500
 middle college high schools 5:68
 Middle East
 pharmacy education 3:59
 private higher education (PHE) 4:622
 pupil–teacher ratios (PTRs) 2:506*f*
 middle schooling 6:233–241
 background information 6:233
 middle schools
 reading curricula 1:465
 self-regulation intervention programs 6:727
 structural characteristics 6:235
 Uruguay 5:875*f*
 research evidence
 effectiveness 6:239
 general discussion 6:238
 implementation extent 6:238
 research perspectives
 academic achievement declines 6:233
 adolescence needs and challenges 6:234
 changing contexts and goals 6:234
 educational practices 6:235
 engagement declines 6:233
 general discussion 6:233
 middle school structures 6:235
 schooling approaches 6:234
 student learning processes 6:234
 research results and analysis 6:239
 whole-school design concept 6:236, 6:236*f*, 6:237*t*
 Middle-school Mathematics through Applications Project (MMAP) 5:338
 migrant populations *see* immigrant populations
 military officer education 3:7–12
 educational continuum
 intermediate education 3:8
 pre-commissioning 3:8
 primary education 3:8
 senior education 3:8
 general/flag officer (G/FO) 3:8
 International Military Education and Training (IMET) program 3:11
 leadership development and education 3:8
 military service academies
 core curriculum
 ethics education 3:10
 general discussion 3:9
 law education 3:10
 leadership education 3:9
 philosophy 3:10
 officer/leader development 3:9
 professional military education (PME) 3:10
 professional training background 3:7
 military service academies 3:7, 3:9
 Millennium Cohort Study (UK) 2:276, 2:277*t*
 Millennium Development Goals (MDGs) 2:504, 4:690, 5:728, 6:249–250
Milliken v Bradley 2:394
 Mill, John Stuart
 contemporary liberalism 6:76
 human capital theory 2:194
 rational autonomy 6:123
 Mills, C. W. 6:338
Mills v Board of Education (1972) 2:689
 Milwaukee Project 2:609
 Milwaukee voucher program 2:338–339, 2:381, 2:382*t*
 mind, brain, and education (MBE) 5:310–316
 background information 5:310
 emotion-related processes
 cognitive stimulation 5:313, 5:314*f*
 domain specificity 5:313
 memory and attention 5:314
 social interactions 5:313, 5:314*f*
 nature versus nurture theory 5:310
 neural networks
 educational skills 5:312
 mathematics skills 5:312
 reading skills 5:312
 neuroimaging technologies 5:311
 neuroscience—education connection 5:315
 Ministry for Education and Culture (MEC) 5:874
 Minkowski's inequality 7:428
 Minnesota 1:654*f*
 minority students
 access and equity patterns 5:52
 achievement motivation 1:753–759
 ability beliefs
 intelligence theories 1:755
 racial stereotypes 1:754, 2:776
 achievement values/effort
 basic concepts 1:756
 oppositional identity 1:756, 2:776
 self-affirmation theories 1:757
 valued peers 1:757
 general discussion 1:754–759
 intelligence theories
 discrimination 1:756
 entity versus incremental theories 1:755
 racial stereotypes
 African Americans 1:754, 2:776
 Asian Americans 1:755, 2:776
 definitions 1:754
 research areas 1:758
 admissions tests 1:846, 4:12
 Brazil 1:760–766
 Catholic schools 2:356
 China
 cultural perspectives 1:616
 educational policies 1:616–622
 case studies 1:617
 challenges 1:620
 ethnic pluralism 1:619
 historical perspectives 1:619
 Koreans 1:619
 Mongols 1:618
 Naxis 1:618–619
 preferential treatment 1:616
 sociocultural factors 1:617
 Tibetans 1:618
 Uyghurs 1:618
 cultural competence 2:565–570
 basic concepts 2:566
 community involvement 2:568
 importance 2:569
 overrepresentation 2:565
 Pedersen's model 2:566
 research background 2:565
 demographic data 1:610
 disadvantaged urban youth 1:806–813
 academic achievement outcomes 1:808
 community involvement 1:811
 educational preparedness 1:810
 emancipatory education 1:812
 knowledge as power 1:811
 male underachievement 1:808
 phenomenological variant of ecological systems theory (PVEST) 1:807*f*, 1:808–809, 1:809*f*
 research background 1:806
 resilience 1:810
 self-assessments 1:812
 self-efficacy beliefs 1:810
 sociocultural factors 1:806
 support systems 1:806
 vulnerability level/resiliency prediction model 1:806, 1:807*f*, 1:809*f*
 diversity discourse 1:823
 diversity research 1:693
 early childhood education and care (ECEC) 2:64, 2:81
 educational statistics 1:581–588
 academic achievement outcomes 1:808
 choice effects 1:585

- minority students (*continued*)
 continuing disparities 1:584
 demographic diversity 1:582
 digital technology implications 1:584
 disparity measures 1:581, 1:874
 dropout rates 1:874, 1:875
 education beyond borders 1:586
 global village concept 1:587
 graduation rates 1:874
 incarceration rates 1:874
 twenty-first century perspectives 1:581, 1:587
 United States 1:582
- educational vouchers 2:383
- gifted and talented students 2:872
- higher education
 admissions test fairness 4:12
 demographic data 1:687
 economic returns 1:542–546
 corporate environments 1:544
 earnings differentials 1:543
 economic well-being 1:545
 income gaps 1:543
 research background 1:542
 social mobility 1:542
 stratification 1:542
 educational attainment 1:799–805
 black males 1:800, 1:839
 completion rates 1:800, 1:803, 1:839
 dropout rates 1:803
 earnings differentials 1:804
 18–24-year-olds demographics 1:802, 1:803*f*
 enrollment rates 1:800, 1:801*f*, 1:802*f*, 1:803*f*
 female–male participation rates 1:799, 1:804
 gender gap research 1:799, 1:804
 United States 1:799
 enrollment opportunities 1:656, 1:658*f*
 student learning outcomes 4:479–484
 benefits 4:483
 cognitive complexity 4:479, 4:480*f*
 humanitarianism 4:480*f*, 4:481
 interpersonal and intrapersonal competence 4:480*f*, 4:482
 knowledge acquisition 4:480, 4:480*f*
 practical competence 4:480*f*, 4:482
 research background 4:479
 United Kingdom 1:673–678
 academic preparation 1:674
 campus experiences 1:676
 diversity considerations 1:673
 institutional dynamics 1:675
 parental involvement 1:676
 peer influences 1:676
 poverty concerns 1:675
 primary and secondary education 1:674
 information and communication technology (ICT) access and equity
 home access 8:95, 8:96*t*
 home use 8:95
 outcomes analyses 8:95, 8:97
 school access 8:96, 8:96*f*
 school use 8:96
- legal education 1:691–697
 Educational Diversity Project (EDP) 1:694
 general discussion 1:696
 historical perspectives 1:691
 social science research 1:693
- male underachievement 1:808, 1:839
- race-specific consequences 1:876
- racial and ethnic demographics 1:687, 1:874
- scientific racism 1:554–564
 background information 1:554
 contemporary science education
 communication norms 1:559
 educational impacts 1:559
 funding and training considerations 1:560
 craniometry 1:557
 eugenics 1:558
 historical perspectives 1:555
 medical research abuse 1:557
 monogenism/polygenism 1:558
 physical anthropology 1:557
 psychometric theories 1:558
 racial evolution theories 1:555
 sociohistorical constructs 1:562
 theoretical retrofitting 1:558
 segregation policies 2:775
 standardized tests 2:577
- United Kingdom
 African-Caribbean population
 alternative educational settings 1:742
 anti-racist education 1:739–740
 assimilation policies 1:575, 1:737
 school-leaving impacts 1:742
 suspensions and exclusions 1:577, 1:742
 underachievement 1:575, 1:740, 1:839
 educational reform 1:737–745
 alternative educational settings 1:742
 anti-racist education 1:739
 assimilation policies 1:737
 background information 1:737
 core issues 1:740
 cultural pluralism 1:737
 differential performance outcomes 1:740
 government policies and programs 1:743
 multicultural education 1:738
 national curriculum development 1:740
 post-war migration impacts 1:737
 school-leaving impacts 1:742
 suspensions and exclusions 1:742
 underachievement 1:740
 higher education 1:673–678
 academic preparation 1:674
 campus experiences 1:676
 diversity considerations 1:673
 institutional dynamics 1:675
 parental involvement 1:676
 peer influences 1:676
 poverty concerns 1:675
 primary and secondary education 1:674
 race/racism 1:737
- South Asian population
 alternative educational settings 1:742
 anti-racist education 1:739–740
 assimilation policies 1:737
 multicultural education 1:738
 suspensions and exclusions 1:742
 sustainable development reform 6:293
 underachievement 1:740
- Mischel, Walter 6:725
- missing data 7:283–288
 educational assessments
 item nonresponse models 7:286
 mean absolute error (MAE) 7:287, 7:287*t*
 missing at random (MAR) 7:287*f*, 7:287*t*
 item response theory (IRT) 4:86, 7:283
 missing data handling techniques
 deletion procedures 7:284
 general discussion 7:284
 model-based estimation 7:285
 modeling approaches 7:286
 multiple imputations 7:285
 single imputation procedures 7:284
 missing data mechanisms
 general discussion 7:283
 ignorable missingness 7:283
 missing at random (MAR) 7:283, 7:287*f*, 7:287*t*
 missing completely at random (MCAR) 7:283
 nonignorable missingness 7:284
 threats 7:283
 missing value plot 7:158*f*
- Mississippi 1:654*f*
- Missouri v. Jenkins* 2:394
- Mittlerer Schulabschluss* 5:585*f*
- mixed-methods research 6:514–520
 advantages/disadvantages 6:519
 data analysis strategies 6:518
 data collection strategies 6:518
 design considerations
 basic concepts 6:516
 conversion designs 6:517*f*
 fully-integrated designs 6:518*f*
 parallel designs 6:516*f*
 sequential designs 6:517*f*
 evaluation research 7:152

- evaluation research methodologies 3:742
- inference transferability 6:519
- integrated inferences 6:518
- International System for Teacher Observation and Feedback (ISTOF) project 3:727
- purposes 6:515
- research background 6:514
- research questions 6:515
- teacher education 7:506
- mixture models 7:238–244
 - background information 7:238
 - basic concepts
 - data sets 7:239^r
 - general discussion 7:238
 - proportion and probability parameters 7:239^r
 - explanatory variables 7:241
 - extensions 7:241
 - generalized linear models (GLMs) 7:240
 - item response theory (IRT) 7:240
 - maximum likelihood estimation (MLE) 7:242
 - model selection considerations 7:242
 - probabilistic Guttman models 7:240
 - Proctor model 7:240
 - proportion and probability parameters 7:239^r, 7:240^r
 - restricted models 7:240
 - software packages 7:243
- Mizrahi Democratic Rainbow Coalition 5:620–621
- mnemonics 5:324
- mobility 4:286–293
 - benefits 4:290, 4:290^f
 - displaced workers 8:275
 - foreign doctoral student enrollment 4:292^r
 - global effects 4:292
 - international mobility 4:291, 4:291^f
 - migration statistics 4:286–287, 4:287^f
 - mobility indicators and measurements 4:289
 - PhD students
 - challenges 4:288
 - exchange structure and organization 4:288
 - foreign student enrollment 4:292^r
 - general discussion 4:288
 - scientists 4:289
 - skill shortages 8:326
- mode
 - basic concepts 7:264
 - characteristics 7:265
- model fit statistics 7:9, 7:165, 7:200
- Model-It 8:199
- model misspecification statistics 7:165
- model selection 7:289–295
 - Bayesian inference 7:292
 - characteristics 7:9
 - continuous data approach 7:291
 - general discussion 7:289, 7:294
 - latent class (LC) models 7:242
 - likelihood-ratio tests 7:190, 7:290
 - linear models 7:293
 - Markov chain Monte Carlo (MCMC) 7:250
 - mixture models 7:242
 - one-parameter logistic (1PL) model 7:286, 7:289–290
 - Pearson chi-squared (χ^2) tests 7:190, 7:291
 - penalty criteria approach 7:291
 - time series analysis 7:478
- moderate hermeneutics 6:65
- moderate participation 6:443
- moderation processes 3:457–463
 - applications
 - general discussion 3:457
 - high stakes school-based certification 3:458
 - low stakes school-based certification 3:458
 - university course assessments 3:459
 - vocational training assessments 3:458
 - challenges 3:462
 - characteristics 3:457
 - consistency goals 3:460
 - design considerations 3:460
 - participative strategies 3:459
 - proactive strategies 3:459
 - research areas 3:462
 - social moderation versus statistical moderation 3:459
 - strong quality control (high stakes)
 - expert moderation 3:461
 - group/consensus moderation 3:461
 - panel moderation 3:461
 - weak quality control (low stakes)
 - assessor meetings 3:461
 - assessor partnerships and conversations 3:461
 - self-audits 3:461
- modern apprenticeship (MA) programs 5:772
- modified Delphi technique 3:727
- modularization 8:383–390
 - background information 8:230, 8:383
 - basic concepts
 - English model 8:386
 - German model 8:386
 - similarity–difference comparisons 8:387
 - characteristics 8:385
 - current trends
 - educational sector integration 8:388
 - general discussion 8:387
 - mainstream training 8:388
 - reduced fragmentation 8:388
 - supranatural development 8:388
 - English versus German module comparisons 8:383
- Mofet Institute (Israel) 7:683
- Moivre, Abraham de 7:467
- Moldova
 - early childhood education and care (ECEC) 2:111^r
 - inequality–education relationship 2:273^f
- Monaco 2:111^r
- Mongolia 5:681–686
 - background information 5:681
 - early childhood education and care (ECEC) 2:111^r
 - education system
 - administrative and supervisory structure 5:684
 - challenges 5:685
 - educational finance 5:684
 - goals 5:681
 - monitoring and evaluation programs 5:685
 - research programs 5:685
 - structural and operational characteristics
 - adult education 5:683
 - early childhood education and care (ECEC) 5:682
 - further education and training (FET) 5:682^r, 5:683
 - general discussion 5:682
 - gifted and talented students 5:683
 - higher education 5:682^r, 5:683
 - nonformal education 5:683
 - preschool education 5:682, 5:682^r
 - primary education 5:682, 5:682^r
 - secondary education 5:682, 5:682^r
 - special needs students 5:683
 - staff and students 5:682^r
 - vocational education and training (VET) 5:682^r, 5:683
 - teaching profession 5:684
- monomolecular growth model 7:343
- Montana 1:654^f
- Monte Carlo analysis 7:296–303
 - advanced techniques
 - Markov chain Monte Carlo (MCMC) 7:302
 - Quasi-Monte Carlo (QMC) 7:302
 - sequential Monte Carlo (SMC) 7:302
 - background information 7:245, 7:296
 - basic concepts
 - elementary methods 7:298
 - fundamental theorem of sampling 7:299, 7:300^f
 - illustrative example 7:296, 7:297^f
 - importance sampling 7:300, 7:301^f
 - inversion sampling 7:299, 7:299^f
 - rejection sampling 7:299, 7:300^f
 - Buffon's needles 7:297^f
 - computational statistics 7:94, 7:95
 - historical background 7:297, 7:297^f
 - Markov chain Monte Carlo (MCMC) 7:245–252
 - applications
 - item response theory (IRT) 7:251
 - sensitivity analyses 7:251
 - background information 7:245
 - beta distribution 7:99
 - characteristics 7:302
 - general discussion 7:10
 - Gibbs sampling 7:246, 7:247^f

- Monte Carlo analysis (*continued*)
 implementation
 convergence assessments 7:250
 model selection 7:250
 sample quality 7:250
 sample usage 7:250
 software packages 7:250
 Markov chains 7:246
 Metropolis–Hastings algorithm 7:247, 7:249*f*
 missing data handling techniques 7:286
 order statistics 7:361
 recent developments 7:250
 simulated annealing (SA) algorithm 7:248
 Monte Carlo Expectation–Maximization (MCEM) algorithm 7:455
 variance reduction techniques
 antithetic variables 7:302
 control variates 7:302
 general discussion 7:301
 Rao–Blackwellization 7:301
 Montenegro 8:23*t*
 Montessori, Maria
 early childhood curriculum 1:515
 peace education 1:850
 Mont Pelerin Society 6:12
 Monserrat 2:111*t*
 mood
 cognition–emotion relationship 5:237–244
 affect 6:547
 causal factors
 cognitive-appraisal hypothesis 5:239
 emotional arousal and valence 5:240
 pedagogical challenges 5:242
 results analysis 5:242
 definitions 5:238
 research background 5:237
 definitions 6:544
 moral and social reasoning 5:279–284
 adolescents 5:281
 cultural influences 5:282
 definitions 5:279
 educational implications 5:283
 general discussion 5:283
 historical background 5:279
 middle childhood 5:280
 preschool-aged children 5:280
 stereotypes 5:283
 moral autonomy 6:121
 moral conduct 6:123
 moral justice 4:57
 moral leadership 4:759, 5:45
 moral values 7:650–655
 business/accounting education 3:41, 3:43
 characteristics 7:650
 citizenship development 7:651
 educational purpose 6:105
 justice and care 6:164–169
 caring perspectives 6:166
 conflicting perspectives 6:167
 educational applications 6:168
 female perspective 6:166
 justice as fairness 6:165
 principles of justice 6:165
 theoretical perspectives 6:164
 traditional justification approaches 6:168
 moral development concepts 7:650
 pharmacy education 3:60
 teacher–student perspectives 7:651
 theoretical perspectives
 general discussion 7:652
 moral politics 7:653, 7:654*t*
 moral sensitivity 7:653, 7:654*t*
 participation and dialog activities 7:653, 7:654*t*
 reflective–practitioner approach 7:653, 7:654*t*
 value transfer 7:652, 7:654*t*
Morgan v Kerrigan 2:394
 Morocco
 early childhood education and care (ECEC) 2:111*t*
 inequality–education relationship 2:273*f*
 information and communication technology (ICT) 8:25*t*
 student achievement test performance 2:247*f*, 2:417*t*, 2:418*t*
 test score–social background relationship 2:273*f*
 Trends in International Mathematics and Science Study (TIMSS) 5:645*t*
 morphemes 5:378*t*, 5:396
 Morris, Henry 6:257
 mother tongue teacher education 7:738–743
 challenges 7:741
 field experiences 7:741
 general discussion 7:738
 pedagogical content knowledge 7:739
 postcolonial theory 1:302
 professional development
 current issues 7:739
 demographic shifts 7:740
 literacy considerations 7:741
 methods courses 7:740
 subject-matter knowledge 7:738
 motivation
 achievement goal theory 6:537–543
 classroom goal structures 6:539
 correlates 6:538
 definitions 6:537–538
 historical background 6:538
 outcomes analyses 6:541
 performance–avoidance/performance–approach goals 6:538
 research areas 6:540
 active music making skills 5:410
 background information 6:712
 belief system interactions 7:623, 7:623*f*
 classroom assessment 3:411
 classroom environments 6:624–630
 challenges 6:624
 expectancy considerations
 curricula 6:625
 goal setting 6:626
 informative feedback 6:626
 success–failure attributes 6:625
 expectancy–value model 6:624
 value considerations
 extrinsic motivation approaches 6:626
 general discussion 6:626
 intrinsic motivation approaches 6:627
 motivational strategies 6:628
 self-determination theory 6:627
 conservation of resources (COR) model
 basic concepts 3:499, 3:500*f*
 educational applications 3:499
 future directions 3:503
 general discussion 3:498
 resource categories 3:499
 resource loss and gain spirals 3:501, 3:501*f*
 creativity 4:170
 cultural influences 6:576–584
 background research 6:576
 indigenous perspectives 6:581
 methodological shortcomings 6:582
 research areas 6:582
 theoretical perspectives
 achievement motivation 6:577
 agentic modes 6:580
 attribution theory 6:579
 educational practices 6:579
 extrinsic motivation approaches 6:578
 general discussion 6:577
 goal orientation 6:579
 intrinsic motivation approaches 6:578
 self-beliefs 6:579
 self-determination theory 6:578
 social diversity 6:579
 curriculum-based assessment (CBA) 2:574
 ethnic minorities 1:753–759
 ability beliefs
 intelligence theories 1:755
 racial stereotypes 1:754, 2:776
 achievement values/effort
 basic concepts 1:756
 oppositional identity 1:756, 2:776
 self-affirmation theories 1:757
 valued peers 1:757
 general discussion 1:754–759
 intelligence theories
 discrimination 1:756
 entity versus incremental theories 1:755
 racial stereotypes
 African Americans 1:754, 2:776

- Asian Americans 1:755, 2:776
- definitions 1:754
- research areas 1:758
- extrinsic motivation **6:618–623**
 - basic concepts 6:619, 6:619f
 - behavioral consequences 6:621
 - classroom environments 6:626
 - creativity 4:170
 - external regulation 6:619
 - general discussion 6:618, 6:622
 - identified regulation 6:620
 - integrated regulation 6:620
 - internalization processes 6:620
 - introjected regulation 6:620
 - reward systems 6:621
 - second-language learning 5:420
 - socioenvironmental contexts 6:620
- immigrant populations 4:633
- individual valuation processes 5:7, 5:8f
- interest concepts 6:615
- intrinsic motivation **6:618–623**
 - basic concepts 6:618, 6:619f
 - behavioral consequences 6:621
 - classroom environments 6:627
 - creativity 4:170
 - general discussion 6:618, 6:622
 - internalization processes 6:620
 - reward systems 6:621
 - second-language learning 5:420
 - socioenvironmental contexts 6:620
- motivational regulation **6:631–635**
 - basic concepts 6:631
 - dimensional characteristics 6:633
 - general discussion 6:634
 - influencing factors 6:633
 - sociocultural factors 6:633
 - student learning impacts 6:631
 - theoretical perspectives 6:632
 - volitional theories 6:632
- personal epistemology 5:213
- reading skills 5:395
- second-language learning 5:420
- self-efficacy beliefs 6:581, 6:625, 6:671
- small business training 8:338
- sociocultural issues **6:712–717**
 - general discussion 6:716
 - research areas 6:714
 - theoretical perspectives
 - culture and cultural practices 6:713
 - general discussion 6:712
 - interpersonal relationships 6:714
 - transformative internalization and externalization 6:714
 - zone of proximal development (ZPD) 1:517, 6:713
- teacher educators 7:682, 7:686
- technology-supported learning environments 6:719
- volitional control **6:724–731**
 - background information 6:724
 - characteristics 6:724
 - effortful control
 - academic work habits 6:726
 - emotional regulation 6:725
 - learning task management strategies 6:726
 - research background 6:725
 - future research areas 6:728
 - intervention programs
 - general discussion 6:727
 - middle schools 6:727
 - primary schools 6:727
 - secondary schools 6:728
 - undergraduate settings 6:728
- motor skills disorders 2:579
- movies 8:109
- moving-to-opportunity (MTO) program 2:432, 2:434f
- Mozambique
 - school fees 2:419
 - student achievement test performance 2:418f
- multicollinearity 7:343
- multicomponent latent trait model (MLTM) 7:87, 7:87f
- multicultural education 1:117, 1:738, 6:356, 7:723
- multidimensional scaling (MDS) **7:304–311**
 - basic concepts 7:15
 - characteristics 7:5
 - data collection 7:305
 - dimensionality selection 7:307
 - Euclidian distance model 7:305, 7:305f
 - fitting criteria 7:306
 - individual differences (ID) MDS model 7:307, 7:308f, 7:309f
 - scale levels 7:306
 - similarity–dissimilarity analyses 7:304, 7:305f, 7:305f
 - software packages 7:310
 - unfolding analysis 7:309, 7:310f, 7:310f
- multilevel confirmatory factor analysis (MCFA)
 - basic concepts 7:165
 - estimation process 7:166
 - illustrative application
 - analysis and results
 - assessment of fit 7:166, 7:167f
 - model estimation 7:166, 7:167f
 - preliminary analysis 7:166
 - standardized factor loadings and errors 7:168f
 - methodology
 - data source and sample 7:166
 - evaluation questionnaire 7:166
 - hypothesized factor model 7:166, 7:167f
- multilevel models **7:210–215**
 - basic concepts 7:8
 - educational production function (EPF) 2:405
 - general discussion 7:210, 7:214
 - generalized linear models (GLMs) 7:176, 7:212
 - hierarchical linear models **7:210–215**
 - basic concepts 7:8
 - effect size 7:446, 7:447f
 - general discussion 7:210, 7:214
 - generalized linear models (GLMs) 7:212
 - longitudinal/repeated measures model 3:120, 3:121f, 7:212
 - meta-analysis 7:214
 - multivariate longitudinal data analysis
 - basic concepts 3:120
 - self-concept study 3:121, 3:121f, 3:122f
 - variable measures 3:121f, 3:122f
 - nested data structures 7:210
 - recommended reading 7:214
 - shrinkage estimates 7:213
 - two-level linear model 7:210
 - value-added models (VAMs) 7:491
 - institutional effectiveness measures 3:175, 8:92
 - longitudinal/repeated measures model 7:212
 - meta-analysis 7:214
 - multivariate data analysis 7:16
 - nested data structures 7:210
 - recommended reading 7:214
 - school effectiveness research (SER) 6:270
 - shrinkage estimates 7:213
 - two-level linear model 7:210
- multimedia *see* e-learning; Internet-based education
- Multimedia Educational Resource for Learning and Online Teaching (MERLOT) 4:425
- Multimodal Behavior Assessment (MMBA) 2:577
- multimodal literacy 8:37
- multinomial distribution 7:133
- multiple comparison procedures **7:312–318**
 - Bayesian multiple comparisons 7:315
 - correlation analysis 7:316, 7:317f
 - directional error rates 7:315
 - error rates 7:312
 - illustrative example 7:314, 7:315f
 - least significant difference (LSD) procedure 7:313, 7:315f
 - multiple confidence intervals 7:315f, 7:316
 - research background 7:312
 - Type I statistical errors 7:312
- multiple correspondence analysis (MCA) 7:106, 7:109f, 7:110f
- multiple-indicator multiple-cause (MIMIC) model 4:38f, 4:40
- multiple information sources **5:191–196**
 - educational implications 5:195
 - information integration strategies 5:191
 - learning processes
 - expertise considerations 5:193
 - general discussion 5:193
 - personal epistemology 5:194
 - reading tasks 5:194
 - mental representation processes 5:192
 - research areas 5:195
 - text-based learning 5:191
 - World Wide Web (WWW) 5:194

- Multiple Intelligences (MI) model 2:831, 4:2, 5:186
- multiple linear regression models 7:7
- multiresponse models 7:344
- multistage sampling 3:145, 7:392
- multitrait-multimethod (MTMM) designs 3:111–119
- alternative models
 - confirmatory factor analysis (CFA)
 - confirmatory factor analysis with correlated uniqueness (CFA-CU) 3:115, 3:115f, 3:116t, 3:118, 3:118t
 - confirmatory factor analysis with trait and method factors (CFA-TM) 3:114, 3:114f, 3:116, 3:116t, 3:117t
 - direct product (DP) model 3:115, 3:116t, 3:118, 3:118t
 - basic concepts 3:111, 3:112f
 - Campbell-Fiske model
 - basic concepts 3:111
 - data analysis examples 3:116, 3:116t
 - heterotrait-heteromethod (HTHM) 3:112, 3:112f
 - heterotrait-monomethod (HTMM) 3:112, 3:112f
 - monotrait-heteromethod (MTHM) 3:111–112, 3:112f
 - monotrait-monomethod (MTMM) 3:111–112, 3:112f
- multi-user virtual environment (MUVE) interfaces 8:161
- see also* River City MUVE
- multivariate analysis of variance (MANOVA) 7:319–323
- basic concepts 7:5, 7:319
 - educational applications 7:322
 - estimation methods 7:321
 - factors 7:321
 - Fisher discrimination 7:137
 - fixed effects 7:320
 - growth-curve models 7:204–205
 - hypothesis testing 7:319
 - inference estimations 7:322
 - multivariate linear regression 7:326
 - multiway MANOVA 7:321
 - nonparametric statistical methods 7:352
 - random effects 7:320
- multivariate data analysis 7:12–17
- abductive research methods 3:77–82
 - exploratory factor analysis 3:77
 - general discussion 3:77, 3:81
 - grounded theory 3:80
 - theory of explanatory coherence 3:78
 - correlation matrix 7:13
 - data matrix 7:12
 - data types 7:13
 - dependence versus independence 7:14
 - experimental research design 7:129
 - functional data analysis (FDA) 7:419, 7:421f
 - general discussion 7:12
 - graphical models 7:197–202
 - analysis of deviance 7:200
 - Bayesian networks 7:199
 - chain graph models 7:199–200
 - general discussion 7:10, 7:201
 - Gibbs-Markov equivalence theorem 7:199
 - graph theory 7:197
 - joint probability distributions 7:198
 - Markov properties 7:199
 - math grade example
 - covariance matrix 7:198t
 - general discussion 7:197
 - partial correlation matrix 7:198t
 - test scores 7:198f
 - model fitting and selection 7:200
 - path diagrams 7:200
 - Simpson's paradox 7:199, 7:199f
 - inference estimations 7:14
 - methodologies
 - classification 7:135–141
 - basis expansion methods 7:138, 7:138f
 - classification-tree method 7:139, 7:140f, 7:141f
 - definitions 7:135
 - density-based classification methods 7:138
 - Fisher discrimination 7:136
 - illustrative example 7:136, 7:136f
 - k*-nearest-neighbor classification method 7:141
 - least-squares regression 7:135, 7:136f
 - linear regression 7:135
 - logistic regression (LR) analysis 4:38t, 4:39, 4:167, 7:136, 7:136f, 7:138f
 - support vector classifiers 7:137
 - two-group example 7:136f
 - cluster analysis 7:14
 - correspondence analysis 7:16
 - discriminant analysis 7:135–141
 - basic concepts 7:5, 7:16
 - basis expansion methods 7:138, 7:138f
 - classification-tree method 7:139, 7:140f, 7:141f
 - definitions 7:135
 - density-based classification methods 7:138
 - Fisher discrimination 7:136
 - illustrative example 7:136, 7:136f
 - k*-nearest-neighbor classification method 7:141
 - least-squares regression 7:135, 7:136f
 - linear regression 7:135
 - logistic regression (LR) analysis 4:38t, 4:39, 4:167, 7:136, 7:136f, 7:138f
 - support vector classifiers 7:137
 - two-group example 7:136f
 - general discussion 7:14
 - latent variable models 7:15, 7:15t
 - multidimensional scaling (MDS) 7:15
 - multilevel modeling 7:16
 - path analysis 7:16
 - principal component analysis (PCA) 7:15
 - regression analysis 7:16
 - structural equation models (SEMs) 7:16
- multivariate linear regression 7:324–331
- basic concepts 7:324
 - general discussion 7:330
 - hypothesis testing 7:327
 - illustrative example 7:328, 7:329t, 7:330t
 - least squares (LS) criterion 7:326
 - mathematical formulation 7:325
 - statistical inference 7:326
- multivariate longitudinal data analysis 3:120–127
- general discussion 3:126
 - hierarchical linear models
 - basic concepts 3:120
 - self-concept study 3:121, 3:121t, 3:122t
 - variable measures 3:121t, 3:122t
 - latent growth curves model (LGCM)
 - basic concepts 3:122, 3:122f, 3:123f
 - self-concept study 3:123, 3:123t, 3:124t
 - research background 3:120
 - structural cross-lagged regression model
 - basic concepts 3:124, 3:124f, 3:125f
 - self-concept study 3:125, 3:126t
- multivariate normal distribution 7:332–338
- bivariate normal distribution 7:332, 7:333f, 7:335f
 - calculation example 7:336
 - central limit theorem 7:334
 - conditional distribution 7:332, 7:334f
 - equal-density contours 7:332, 7:335f
 - Hotelling's T^2 -test 7:337
 - joint density function 7:333f
 - key properties 7:334
 - Mahalanobis D_M 7:337
 - marginal distribution 7:332, 7:333f
 - standardized distributions 7:336
 - testing approaches 7:337
- structural equation models (SEMs) 7:453–458
- Bayesian estimators 7:456
 - differential item functioning (DIF) 4:40
 - health-related quality of life (HRQOL) example 7:456, 7:457f
 - linear structural equation models 7:453
 - maximum likelihood (ML) approach 7:455
 - methodologies 7:16
 - Monte Carlo Expectation-Maximization (MCEM) algorithm 7:455–456
 - nonlinear structural equation models 7:454
 - ordered categorical data 7:454
 - purpose 7:453
 - reciprocal effects model (REM) 6:664
 - two-level structural equation models 7:455
- summarization methods 7:14
- multivariate normality 7:164
- multivariate time series 7:479
- Municipal Reform Movement 4:724
- museums 1:150–155
- adult learning
 - free-choice learning 1:153
 - general discussion 1:152, 1:154
 - nonformal education 1:153
 - theoretical perspectives 1:151

educational role 1:150
 educational theories 1:151
 general discussion 1:150
 music learning **5:407–412**
 active music making skills
 acquisition processes 5:408
 group participation 5:410
 motivation 5:410
 appraisal and evaluation 5:408
 classroom-based assessments 3:265
 general discussion 5:407, 5:411
 globalization effects 1:423
 infants and young children 5:407
 listening processes 5:408
 musical skills 5:407, 5:408*t*
 response processes 5:408
 Slovenia 5:786
 technology-supported learning environments 1:423
 music teacher education **7:706–711**
 challenges
 computer technologies 7:709
 cultural diversity 7:708
 educational reform 7:707
 mass media influences 7:709
 professional partnerships 7:708
 teacher shortages 7:709
 curricula 7:707
 education systems
 Africa 7:706
 Asia 7:706
 characteristics 7:706
 Europe 7:706
 North America 7:707
 southern hemisphere 7:707
 future directions 7:710
 see also arts teacher education
 Muslim education 1:273, 1:743
 see also Islamic educational philosophy
 mutuality 6:685
 Myrdal, Gunnar 1:542

N

naive teaching 7:662
 Namibia **5:687–691**
 background information 5:687
 early childhood education and care (ECEC) 2:111*t*
 education system
 administrative and supervisory structure 5:689
 challenges 5:690
 educational finance 5:689
 goals 5:687
 monitoring and evaluation programs 5:689
 research programs 5:689
 structural and operational characteristics 5:687
 teaching profession 5:688
 socioeconomic/sociopolitical characteristics 1:199
 Southern African Norwegian Training, Education and Development (SANTED) initiative 1:856
 student achievement test performance 2:418*t*
 see also Africa
 Nanyang Technical University (NTU) 8:69
 narrative inquiry **6:436–441**
 autobiographical narrative inquiry 6:437
 basic concepts
 general discussion 6:436
 place 6:436
 sociality 6:436
 temporality 6:436
 change dimensions 6:440
 definitions 6:436
 design considerations
 ethical practices 6:440
 form and audience issues 6:440
 general discussion 6:437
 justification
 personal justification 6:438
 practical justification 6:438
 social justification 6:438
 living the experience
 field–field text continuum 6:438
 field text–interim research text continuum 6:439
 general discussion 6:438
 interim research text–research text continuum 6:439
 naming practices
 narrative thinking 6:438
 research puzzles 6:438
 positioning techniques 6:439
 general discussion 6:436
 life-history research 6:431
 starting points
 general discussion 6:437
 living stories 6:437
 telling stories 6:437
 narrative practices **7:610–614**
 characteristics 7:610
 critical incident vignette 7:610
 critical race theory 1:782, 6:343
 language 7:612
 professional importance 7:613
 social context 7:612–613
 student teachers 7:611
 voice 7:613
 narrative socialization **5:303–309**
 child-based discourse
 conarration 5:304
 early narrative stages 5:303
 genres 5:303
 cognition
 narrative thinking 5:304, 5:464, 6:438
 theories of the mind 5:305
 life-course perspective 5:303
 research areas 5:307
 sociocultural groups
 cultural communities 5:306
 gender analysis 5:305
 general discussion 5:305
 social class 5:305
 National Academy of Engineering (NAE) 3:45
 National Administration for Public Education (ANEP) 5:874, 5:881*t*, 5:883*t*
 National Adult Learning Survey (NALS) 8:264
 National Architectural Accrediting Board (NAAB) 3:13
 National Assessment Governing Board (NAGB) 3:690
 National Assessment of Educational Progress (NAEP) **3:690–697**
 achievement levels 3:692, 6:303
 affective assessments 3:261
 assessment design 8:63
 background information 2:178, 3:523–524, 3:690, 4:45
 basic concepts 2:179, 2:181*t*
 Catholic schools 2:357
 challenges 3:695, 3:696*f*
 discipline-based arts education (DBAE) 1:425
 empirical Bayes methods 7:146, 7:147*f*
 fine arts education 3:263
 future research areas 6:265
 globalization effects 3:696
 high-stakes test score inflation 4:186, 4:187*t*
 history/civics assessments 3:323, 3:325*t*, 4:175–176, 4:657
 jackknife method 7:231
 large-scale assessments 7:4, 8:50
 long-term trend (LTT) assessments 3:691, 3:691*f*, 3:692*f*
 main NAEP
 basic concepts 3:693
 special educational studies 3:695
 state NAEP 3:694, 3:694*f*
 TUDA programs 3:695
 national assessment programs 4:94
 norm-referenced measurement
 criterion-referenced reading achievement levels 4:99*t*
 general discussion 4:97
 labeling comparisons 4:98*t*
 percent of labeled students 4:99*t*
 student reading profile report 4:100*f*
 progress maps 3:373
 random number generation 7:184
 sampling methods 7:6, 7:390
 scale scores 3:692
 school finance 2:223–224
 science-education programs 3:664
 social studies/civic education programs 3:669
 student achievement 2:246, 2:376–377, 3:693

- National Assessment of Educational Progress (NAEP) (*continued*)
 technology-rich environments scenarios 8:52, 8:53*f*
 United States education system 5:891
 National Assessment Program (NAP) 3:323
 national assessment programs 4:89–94
 background information 3:690
 design considerations
 across-time comparisons 4:92
 age levels 4:91
 data analysis and reporting 4:91
 inclusion criteria 4:91
 outcome selection 4:90
 participation maximization 4:91
 functional role 4:89
 general discussion 4:89
 modeling approaches
 international/regional assessment programs 4:89
 national samples of students 4:90
 national student cohort testing 4:90
 school inspections 4:90
 national systems
 England 4:92
 New Zealand 4:93
 United States 4:94
 reform efforts 3:172
 special needs students 3:495
see also National Assessment of Educational Progress (NAEP)
 National Association for Early Childhood Specialists in State Departments of Education (NAECS/SDE) 2:133
 National Association for Sport and Physical Education (NASPE) 2:523, 2:524*t*
 National Association for the Advancement of Colored People (NAACP) 6:499
 National Association for the Education of Young Children (NAEYC) 1:516, 2:133, 2:740
 National Board for Professional Teacher Standards (NBPTS) 2:562, 2:563*t*, 6:501, 7:536–537
 National Center for Education Statistics (NCES) 2:178, 3:690
 National Center for Gifted Children (NCGC) 2:871
 National Center for Research on Evaluation, Standards, and Student Testing (CRESST) 3:210, 3:228
 National Center of Education Statistics (NCES) 6:706
 National Center on Education and the Economy (NCEE) 1:371
 National Centre for Vocational Education Research (NCVER) 8:353, 8:354
 National Certificates of Educational Achievement (NCEA) 3:447, 5:694–695
 National College for School Leadership (NCSL) 3:539, 4:762, 5:93, 6:243
 National Collegiate Athletic Association (NCAA) 1:860
see also Scholar-Baller[®] program
 National Commission on Teaching and America's Future (NCTAF) 2:498
 National Council for Accreditation of Teacher Education (NCATE)
 elementary education 7:728–729
 literature curricula 1:450
 preservice education programs 8:219–220
 professionalization process 7:497
 science teacher education 7:716
 teacher education standards 2:499, 7:536–537
 National Council for Black Studies (NCBS) 1:568
 National Council for Technological Awards (NCTA) 3:46
 National Council for the Social Studies (NCSS) 1:411, 1:413, 3:326
 National Council of Architectural Registration Boards (NCARB) 3:13
 National Council of Examiners for Engineering and Surveying (NCEES) 3:46
 National Council of Teachers of Mathematics (NCTM) 1:456, 2:767, 4:768, 5:208, 7:626–627
 National Educational Longitudinal Study (NELS) 2:483, 6:706
 National Education Association (NEA) 2:512
 National Education Monitoring Project (NEMP) 4:93
 National Health and Nutrition Examination Survey (NHANES) 3:788
 National Health Education Standards (NHES) 1:548
 National Information Center for Children and Youth with Disabilities (NICYD) 2:654
 National Institute of Child Health and Human Development (NICHD) 2:148
 National Joint Committee on Learning Disabilities (NJCLD) 2:647
 National Labor Relations Act (1970) 4:492
 national longitudinal studies 2:179, 2:181*t*
 National Longitudinal Survey of Youth (NLSY) 1:207, 2:178, 2:241, 2:293
 national norms 4:97
 National Qualifications Framework (NQF) 8:443
 National Reading Panel (NRP) report 2:759
 National Record of Achievement (NRA) 3:237, 3:245
 National Research Centers for Career and Technical Education 8:353
 National Resource Centres for Vocational Guidance (NRCVG) 8:353
 National Schools Network 7:758
 National Science Foundation (NSF) 3:45
 national science policy 4:294–299
 general discussion 4:299
 governance processes 4:295
 historical background 4:294
 institutional structures 4:297
 policies and programs 4:295
 research challenges 4:298
 state–university relations 4:296
 National Survey of Student Engagement (NSSE) 4:381, 4:461, 4:464, 4:476, 4:726
 National Training Framework (NTF) 8:348
 National University of Educational Planning and Administration (NUEPA, India) 4:676–680
 challenges 4:679
 characteristics 4:677
 functional role 4:677–678
 funding sources 4:678
 future directions 4:679
 governance 4:678
 historical background 4:676
 International Institute for Educational Planning (IIEP) 4:691
 UNESCO (United Nations Educational, Scientific and Cultural Organization) 4:676
 national victimization survey (NVS) 2:180, 2:181*t*
 National Vocational Qualifications (NVQs) 2:159, 8:384, 8:386, 8:398, 8:407, 8:476
 National Writing Project (NWP) 7:553
A Nation at Risk (1983) 1:283–284, 1:652
 nation–state concept 1:730, 1:775, 4:514, 4:557–558, 4:707, 7:532–533
 Native Americans
 access and equity patterns 1:658*f*
 cultural competence 2:565–570
 basic concepts 2:566
 community involvement 2:568
 importance 2:569
 overrepresentation 2:565
 Pedersen's model 2:566
 research background 2:565
 cultural difference theory model 2:776
 deaf-blind population 2:623*t*
 educational attainment 1:800, 1:801*f*
 educational opportunities 1:881
 historical policies and practices 1:662
 information and communication technology (ICT) access and equity
 home access 8:95, 8:96*t*
 home use 8:95
 outcomes analyses 8:95, 8:97
 school access 8:96, 8:96*f*
 school use 8:96
 segregation policies 2:775
 naturalistic inquiry 6:485
 nature *see* environmental education
 Nauru 2:111*t*
 NEAT (nonequivalent groups with an anchor test) design 4:53
 Nebraska 1:654*f*
 need for achievement (nAch) theory 6:538
 needs assessments 3:786–791
 augmentative–alternative communication (AAC) 2:546
 definitions
 formal process 3:787
 levels of needs 3:786
 needs 3:786, 3:787*t*
 early childhood interventions 2:19, 2:20*f*
 educational applications 3:790
 educational assessments 3:794
 general discussion 3:786
 modeling approaches 3:787, 3:789*t*
 Trends in International Mathematics and Science Study (TIMSS) 3:790
 negative binomial distribution 7:134
 negative emotions 6:570
 negotiated evaluation 3:574
 Negro Education Act (1844) 1:679
 neighborhoods
 educational outcomes 2:431–437
 general discussion 2:431
 influencing factors 2:431, 2:432*f*
 moving-to-opportunity (MTO) program 2:432, 2:434*t*
 research evidence 2:432
 research results 2:436
 socioemotional development 6:648
 neoliberal ideologies and strategies
 educational globalization 6:531
 evaluation policies and practices 3:646*t*, 3:647

- higher education 1:536, 1:699, 4:557
 historical background 6:11–16
 Friedrich von Hayek 6:12
 New Right 6:13
 origins 6:12
 performance management 6:15
 performativity 6:14
 postmodernism 6:14
 indigenous curricula 1:309
 institutional autonomy 4:342
 lifelong learning 1:219, 1:226, 1:699
 philosophical perspectives 6:2, 6:5
 research commercialization 4:300
 teacher education 7:534
 Nepal 2:111*t*
 nested articulation 4:613
 nested data structures 7:210
 Netherlands
 age-related policies and programs 8:263*t*
 Catholic schools 2:356, 2:358
 childhood brain development 2:57*f*
 class size 2:492*f*, 6:200
 early childhood education and care (ECEC) 2:4*t*, 2:56*t*, 2:111*t*, 2:114*t*
 earnings–education level relationship 2:307*t*
 education levels 2:279*f*
 enrollment rates 2:493*f*
 higher education
 binary systems 4:617
 loans and grants 4:501*f*, 4:502*f*
 student and researcher mobility 4:287*f*
 tertiary education expenditures 4:500*f*
 tuition fees 4:500*f*
 university associations 4:485
 inequality–education relationship
 earnings differentials 2:277*t*
 employment rate–educational attainment relationship 2:278*f*
 fathers' educational background 2:274*f*
 immigrant status 2:275*f*
 test score–social background relationship 2:273*f*
 information and communication technology (ICT)
 computer access in school 8:23*t*
 frequency of computer use in school and at home 8:26*t*
 student proficiency and confidence 8:31*t*
 information and communication technology (ICT) access and equity 8:97
 inquiry-oriented teacher education 7:606
 leadership development programs 4:753*t*, 4:754*t*, 4:755*t*, 4:758*t*
 mathematics education 3:283
 participation research
 regular education 8:284*f*
 vocational education and training (VET) 8:283*f*, 8:284*f*
 work-related vocational education and training (VET) 8:284*f*, 8:285*f*
 private schools 2:337
 research-quality assessments 4:309
 school effectiveness research (SER) 6:269
 school inspections 3:723
 student achievement test performance 2:247*f*
 teacher age distribution 2:492*f*
 teacher evaluations 3:677
 teacher shortages 2:494*f*
 teacher wages and salaries 2:496*f*
 teaching hours 2:493*f*
 vocational education and training (VET)
 age-related policies and programs 8:263*t*
 dual apprenticeship system 8:232, 8:435
 employer involvement 8:457, 8:461
 initial vocational preparation system 8:251*t*
 participation research 8:283*f*, 8:284*f*, 8:285*f*
 vocational–academic education parity 8:435
 women in employment 2:56*f*
 Netherlands Antilles 2:111*t*
 networked learning communities (NLC)
 Networked Learning Communities (NLC) program
 basic concepts 5:93, 5:94*f*
 challenges 5:98
 incentivization and implementation 5:95
 outcomes analyses 5:97
 research findings 5:96
 professional learning communities 5:152
 school–higher education institution (HEI) partnerships 5:94, 7:759
 neurocognitive plasticity 5:343
 neuroscience research 5:310–316
 affect and emotion 6:546
 aging and cognitive abilities 5:340–344
 brain aging 5:341
 cognition–brain relationship 5:343
 long-term prognoses 5:343
 memory performance
 age-related effects 5:342
 cognitive decline 5:341
 episodic memory 5:342
 executive functions 5:342
 familiarity 5:342
 long-term memory 5:341
 recollection 5:342
 semantic memory 5:342
 short-term memory 5:341
 working memory (WM) capacity 5:342
 neurocognitive plasticity 5:343
 neuroimaging studies 5:340
 background information 5:310
 children 2:51, 2:57*f*
 creativity development 5:252
 educational effectiveness 5:53
 emotion-related processes
 cognitive stimulation 5:313, 5:314*f*
 domain specificity 5:313
 memory and attention 5:314
 social interactions 5:313, 5:314*f*
 empathy 6:602
 intelligence theories 5:187
 nature versus nurture theory 5:310
 neural networks
 developmental dyslexia 5:319
 educational skills 5:312
 mathematics skills 5:312
 reading skills 5:312
 neuroimaging technologies 5:311, 5:340
 neuroscience–education connection 5:315
 reading skills 5:317–322
 current technologies 5:317
 developmental dyslexia
 neural activation time course 5:320
 neural networks 5:319
 remediation effects 5:320
 longitudinal studies 5:321
 neuroimaging studies
 developmental dyslexia 5:319
 learning to read 5:318
 neural activation time course 5:320
 neural networks 5:319
 remediation effects 5:320
 phonological awareness 5:317
 reading development models 5:317
 neuroticism 3:256
 Nevada 1:654*f*
 New American Schools initiative 6:303
 newcomers in Japan 1:787–792
 ethnic minorities 1:787
 foreign nationals 1:788*f*
 internationalization processes 1:790
 Japanese returnees 1:788
 mother tongue survey 1:790*f*
 new foreigners 1:789, 1:790*f*
 sociocultural portrayals 1:791
 New Criticism movement 1:447
 New Hampshire 1:654*f*
 New Jersey
 Asian American/Pacific Islander student enrollment 1:597*f*
 early childhood education and care (ECEC) 2:65
 leadership development programs 4:753*t*, 4:755*t*, 4:758*t*
 tuition and needs-based grants 1:654*f*
 new literacy studies (NLS) 1:108
 new localism
 background information 5:106
 community capacity 5:107
 community involvement 5:109
 education action zones (EAZs) 5:107–108
 general discussion 5:110
 parental involvement 5:109
 place-based learning 5:106
 political alliances 5:109

- new localism (*continued*)
 social capital 5:108
 Newman–Keuls test 7:34, 7:313
 Newmann, Fred 3:325
 new media technologies 8:140–149
 general discussion 8:141
 multimedia learning environments
 animation
 general discussion 8:143
 picture comprehension 8:144
 picture design 8:144
 realistic and dynamic pictures 8:143, 8:143f
 future directions 8:148
 general discussion 8:141
 learning models
 cognitive theory of multimedia learning (CTML) 8:145, 8:146f
 design guidelines 8:148
 dual-coding theory 8:145
 integrated model of text and picture comprehension (IMTPC) 8:145, 8:146f
 research findings 8:147
 theoretical models 8:145
 linear and nonlinear text 8:142
 research findings
 coherence and contiguity 8:147
 general discussion 8:147
 modality 8:147
 redundancy 8:147
 text design guidelines
 additional stimulation 8:142
 conciseness 8:142
 general discussion 8:142
 linguistic simplicity 8:142
 structure and organization 8:142
 New Mexico 1:654f
 new public management (NPM)
 background information 4:251
 basic concepts 4:248
 curricula globalization 1:315
 evaluation policies and practices 3:648
 higher education
 background information 4:251
 collegiality 4:266
 current trends 4:264
 entrepreneurial models 4:266
 governance 4:265
 governing boards 4:267
 ideological perspectives 4:251
 institutional autonomy 4:342
 institutional structures 4:267
 strategic planning 4:256, 4:257f
 ideological perspectives 4:251, 7:534
 New Right 6:13
 new scientist training 4:322–327
 role of science 4:322
 scientific disciplines 4:322
 scientist's role 4:323
 training programs
 career paths 4:326
 doctoral (PhD) training
 characteristics 4:324
 diversified knowledge production modes 4:325
 organizational structures 4:325
 postdoctoral training 4:326
 professionalism 4:326
 program duration 4:324
 research arrangements 4:325
 specialization–breadth of knowledge balance 4:325
 early education and training 4:323
 general discussion 4:323
Newsweek rankings 4:551
 Newton–Raphson algorithm 7:242, 7:368–369
 New York
 Asian American/Pacific Islander student enrollment 1:597f
 tuition and needs-based grants 1:654f
 New Zealand 5:692–698
 background information 5:692
 childhood brain development 2:57f
 classroom assessment 3:443–448
 assessment policies 3:444
 background information 3:443
 primary schools 3:444, 5:694
 school curriculum policy 3:443
 secondary schools 3:446
 class size 6:200
 comparative education societies 1:268
 crime–education relationship 1:879
 early childhood education and care (ECEC) 2:4t, 2:70, 2:105, 2:108, 2:111t
 Early Start program 2:126
 earnings–education level relationship 2:307t
 education levels 2:279f
 education system
 challenges
 compulsory education 5:697
 noncompulsory education 5:698
 characteristics 5:692
 early childhood education 3:318, 3:320, 5:694
 educational finance 5:696
 government policies and programs 5:693
 monitoring and evaluation programs 5:696
 primary schools
 assessment strategies 5:694
 characteristics 5:694
 curriculum 5:694, 5:695f
 research programs 5:696
 schematic diagram 5:693f
 secondary schools 5:694
 teacher education 5:696
 tertiary education 5:695
 enrollment rates 2:493f
 gender-equity gap 2:56t
 higher education
 foreign doctoral students 4:292t
 income contingent loans (ICLs) 4:504
 institutional expenditures 1:733t
 loans and grants 4:501f, 4:502f
 parental expenditures 1:734t
 structural and operational characteristics 5:695
 student and researcher mobility 4:287f
 tertiary education expenditures 4:500f
 tertiary type-B program survival rates 8:464t
 tuition fees 4:500f
 university associations 4:485
 indigenous peoples
 demographic data 1:664t
 educational status 1:664
 self-determination 1:663
 tribal colleges 1:663
 inequality–education relationship
 employment rate–educational attainment relationship 2:278f
 immigrant status 2:275f
 test score–social background relationship 2:273f
 information and communication technology (ICT)
 computer access in school 8:23t
 frequency of computer use in school and at home 8:26t
 student proficiency and confidence 8:31t
 leadership development programs 4:753t, 4:754t, 4:755t
 music teacher education 7:707
 national assessment programs 4:93
 New Zealand Education Monitoring Project (NZEMP) 3:386, 3:407
 personalized learning 8:395
 physical education and sports 1:459
 research-quality assessments 4:308
 returns to education–education level 2:308t
 school choice 6:285
 self-managing schools 5:72–73, 5:697
 student achievement test performance 2:247f
 teacher age distribution 2:492f
 teacher wages and salaries 2:496f
 teaching hours 2:493f
 vocational education and training (VET) 8:251t, 8:397, 8:398
 women in employment 2:56f
 New Zealand Education Monitoring Project (NZEMP) 3:386, 3:407
 New Zealand Qualifications Authority (NZQA) 8:398
 New Zealand Qualifications Framework (NZQF) 8:348
 next-generation electronic tests 8:49t, 8:50
 Nicaragua
 assessment system development 3:482
 conditional cash transfer (CCT) programs 2:419
 early childhood education and care (ECEC) 2:111t
 higher education enrollment 1:702f, 1:703f
 Nietzsche, Friedrich 6:65, 6:111, 6:175

- Niger 5:699–708
 background information 5:699
 basic needs indicators 5:700*t*
 early childhood education and care (ECEC) 2:111*t*
 education system
 admission rates 5:703*t*
 Decennial Programme of Development of Education (PDDE)
 educational policies 5:700
 operational objectives 5:701
 strategic objectives 5:701
 economic development 5:704, 5:707*t*
 educational cycles 5:702*t*
 educational expenditures 5:707*t*
 educational finance 5:707*t*, 5:708
 enrollment disparities 5:703, 5:705*t*, 5:706*t*
 examination success rate 5:703, 5:705*t*
 faculty data 5:703, 5:704*t*, 5:706*t*, 5:707*t*
 general discussion 5:708
 gross enrollment 5:705*t*
 higher education enrollment 5:702, 5:704*t*
 nonformal education 5:706
 primary teachers 5:706*t*
 public and private school enrollment 5:703*t*, 5:704*t*
 secondary school enrollment 5:703*t*
 secondary teachers 5:706*t*
 sector analysis 5:702
 sociocultural development 5:706
 structural and operational characteristics
 characteristics 5:699
 educational policies 5:700
 management policies 5:701
 student/staff ratios 5:707*t*
 vocational education and training (VET) 5:702, 5:704*t*
 principal macroeconomic indicators 5:702*t*
 student achievement test performance 2:418*t*
- Nigeria
 early childhood education and care (ECEC) 2:111*t*
 higher education
 collective bargaining 4:494
 gender equity 1:630–631
 private sector services 4:524*t*
 school choice 6:285–286
 secondary education initiatives 6:253
 student achievement test performance 2:418*t*
- Niue 2:111*t*
- No Child Left Behind Act (2001)
 art education 1:425, 3:263
 assessment accountability 2:531, 3:379, 8:6, 8:14
 background information 1:583, 5:888, 7:125
 direct instruction (DI) 2:585
 elementary/middle-grades reading curricula 1:466
 gifted and talented students 2:829–830, 2:870–871
 historical background 2:689, 6:264
 importance 6:265
 inclusion practices 2:722
 infant and toddler programs 2:71
 literature curricula 1:450
 mathematics education 3:280, 5:404
 National Assessment of Educational Progress (NAEP) 2:179
 oral language skills assessments 3:291
 outcome-based accountability 3:520
 paraprofessionals
 background information 2:780
 instructional assignments 2:784
 training programs 2:781
 parental choice 2:371
 post-secondary education 2:808
 private sector services 2:338
 publishing industry 1:370
 purpose 2:845
 racial equality 1:884
 school-based accountability programs 2:374–379
 contemporary assessments 5:892
 data sharing ability 8:128
 design approaches 2:375
 educational effectiveness 2:802
 educational evaluation 3:520
 general discussion 2:374
 rationale 2:374
 student achievement impacts 2:376
 unintended consequences 2:377
 school-based services 2:634–635, 2:851–852
 school finance 2:223–224
 standardized assessments 3:263, 3:324
 sustainable development 6:290–291
 teacher credentials 2:438
 teacher training 2:498
- Noddings, Nel 6:59
 nominal responses 3:154
 Noncognitive Questionnaire (NCQ) 1:847, 4:11–12
 nonconsequential validity 6:510
 nondeclarative memory 5:198, 5:198*f*, 5:342
 nonformal education 8:240–246
 adult learning 1:153
 ageless learners 8:241*f*
 assessment strategies 8:244
 Australia 5:513
 Bangladesh 5:528
 basic concepts 1:3–4
 benefits 8:245
 definitions 8:240
 elder learning 1:176
 Estonia 5:559
 Fiji 5:566
 Gambia 5:860
 general discussion 8:240
 Ghana 5:592*t*, 5:594
 Indonesia 5:612*f*, 5:613
 informal learning comparisons 1:42
 international trends 8:241
 Latvia 5:633
 learner's perspective considerations 8:244
 Lithuania
 special needs students 5:661
 vocational education and training (VET) 5:660
 Malawi 5:670
 Mongolia 5:683
 museums 1:153
 Niger 5:706
 Pakistan 5:732
 Philippines 5:868
 Poland 5:739
 Qatar 5:747
 Rwanda 5:762
 schematic diagram 8:242*f*
 Slovenia 5:786
 Solomon Islands 5:794
 South Korea 1:157*t*, 1:159
 Sri Lanka 5:817
 Sultanate of Oman 5:724
 terminology and jargon sources 8:350*t*
 Thailand 5:854
 United States 5:893
 vocational education and training (VET) 8:244, 8:379
 workplace learning 8:243
 Zanzibar 5:899
see also lifelong learning
- nongovernment organizations (NGOs)
 adult learning funding 1:133
 Afghanistan 8:445
 early childhood education and care (ECEC) 2:4
 educational globalization 6:529–535
 BRAC (Building Resources across Countries) 6:285–286, 6:533
 commodification 6:531
 cultural globalization 6:529
 educational paradigm change 6:531
 functional role 6:532
 Global Campaign for Education (GCE)
 EFA Class of 2015 6:533
 general discussion 6:533
 Global Action Week 6:533
 global advocacy 6:534
 year-round campaigns 6:533
 neoliberal perspectives 6:531
 recent developments 6:534
 school reform 6:286
 United Nations involvement 6:529
 elder learning 1:178
 Fiji 5:566
 immigrant populations 4:633
 indigenous curricula 1:309, 8:290
 International Institute for Educational Planning (IIEP) 4:691

- nongovernment organizations (NGOs) (*continued*)
 - international scholarly networks 4:681–682
 - peace education 6:22
 - Poland 5:740
 - postconflict societies 4:517
 - Rwanda 5:760
 - Suriname 5:829
 - United Nations involvement 6:529
- nonlinear regression analysis 7:339–346
 - basic concepts 7:8, 7:339
 - diagnostic studies 7:343
 - estimation methods
 - data sets 7:339
 - Gauss–Newton method 7:341
 - least-squares estimation 7:340
 - maximum likelihood estimation (MLE) 7:341
 - Michaelis–Menten regression function 7:341, 7:342*f*
 - factorization methods 7:344
 - intrinsic curvature 7:346
 - model building
 - research areas 7:343
 - yield–density models 7:343
 - nonlinearity assessments
 - confidence regions 7:342
 - coordinate system 7:343
 - intrinsic nonlinearity 7:342, 7:342*f*, 7:345
 - parameterization 7:341
 - parameter-effects nonlinearity 7:346
- nonlinear structural equation models 7:454
- nonparametric statistical methods 7:347–353
 - advantages 7:347
 - basic concepts 7:7
 - modeling approaches
 - bootstrap tests 7:349
 - degree of association determinations 7:349
 - distribution estimations 7:348, 7:348*f*
 - general discussion 7:348
 - permutation tests 7:349
 - multivariate analysis of variance (MANOVA) 7:137, 7:204–205, 7:352
 - nonparametric IRT-based/group-based person-fit statistics 4:16
 - nonparametric regression analysis
 - basic concepts 7:350
 - bins 7:350
 - general discussion 7:349
 - kernel smoothing 7:351, 7:351*f*, 7:352*f*
 - marginal maximum likelihood (MML) method 7:351
 - research areas 7:344
 - properties 7:347
 - recursive partitioning 7:383–389
 - classification trees 7:384*f*, 7:384*t*, 7:385*t*, 7:386, 7:386*f*
 - extensions 7:387
 - inference trees 7:387, 7:388*f*
 - prediction methods 7:383, 7:388
 - regression trees 7:386, 7:386*t*, 7:387*f*, 7:387*t*, 7:388*f*
- nonparticipation 6:443
- nonprobability sampling 3:145, 3:155
- nonprofessional graduate architecture degrees 3:13
- nonroutine problems 5:273
- nonuniform differential item functioning (DIF) 4:37
- nonwhites *see* African Americans; ethnicity; minority students
- normal distributions 7:467–473
 - basic concepts 7:178
 - basic properties 7:469
 - bell curve theory 7:471, 7:472*f*
 - bell image 7:468*f*
 - binomial to normal distribution 7:469, 7:470*f*, 7:471*f*
 - central limit theorem 7:469
 - cumulative distribution function (CDF) 7:468, 7:468*f*
 - empirical rule 7:468*f*
 - general discussion 7:467
 - historical background 7:467
 - illustrative example 7:470, 7:471*f*
 - inference estimations 7:179
 - mean and variance 7:468*f*
 - occurrences 7:470
 - probability density function (PDF) 7:467, 7:468*f*
 - standard normal distribution 7:468, 7:469*f*
 - statistical applications 7:473
- normative–reeducative change strategies 7:557
- norm-referenced measurement 4:95–101
 - challenges 4:101
 - definitions 4:31
 - higher education 3:251
 - historical background 4:95
 - normative scales
 - characteristics 4:95
 - developmental scales 4:96
 - status scales 4:96
 - norm types
 - age- and grade-based norms 4:97
 - characteristics 4:96
 - group-level norms 4:97
 - item- and skill-level norms 4:97
 - local norms 4:97
 - national norms 4:97
 - user norms 4:97
 - reading assessments 3:189
 - science-education programs 3:663
 - score interpretation
 - criterion-referenced reading achievement levels 4:99*t*
 - general discussion 4:97
 - labeling comparisons 4:98*t*
 - percent of labeled students 4:99*t*
 - student reading profile report 4:100*f*
 - testing theories 2:577
- North Africa 2:506*f*
 - see also* Africa
- North American Veterinary Licensing Exam (NAVLE) 3:72
- North Atlantic Treaty Organization (NATO) 4:514
- North Carolina
 - early childhood education and care (ECEC) 2:105
 - state and school district data systems 2:183
 - tuition and needs-based grants 1:654*f*
- North Dakota 1:654*f*
- Norway 5:709–717
 - age-related policies and programs 8:263*t*
 - background information 5:709
 - childhood brain development 2:57*f*
 - early childhood education and care (ECEC) 2:4*t*, 2:56*t*, 2:58*t*, 2:65–66, 2:105, 2:111*t*, 2:114*t*
 - earnings–education level relationship 2:307*t*
 - education levels 2:279*f*
 - education system
 - administrative and supervisory structure 5:714, 5:714*f*, 5:715*t*
 - basic concepts 5:710
 - challenges 5:716
 - educational finance 5:714
 - enrollment 5:712*t*
 - goals 5:710
 - monitoring and evaluation programs 5:715
 - reform efforts 5:716
 - research programs 5:715
 - structural and operational characteristics
 - adult education and training (AET) 5:713
 - general discussion 5:710
 - higher education 5:711*f*, 5:712*t*, 5:713
 - postsecondary vocational education and training (VET) 5:713
 - preschool education 5:710, 5:711*f*, 5:712*t*
 - primary schools 5:710, 5:711*f*, 5:712*t*
 - secondary schools 5:710, 5:711*f*, 5:712*t*
 - tertiary education 5:711*f*, 5:712*t*, 5:713
 - upper secondary schools 5:712, 5:712*t*
 - vocational education and training (VET) 5:712, 5:712*t*, 5:715*t*
- enrollment rates 2:493*f*
- gender-equity gap 2:56*t*
- higher education
 - admissions tests 4:12
 - enrollment rates 4:521*t*
 - foreign doctoral students 4:292*t*
 - literacy rates 4:521*t*
 - loans and grants 4:501*f*, 4:502*f*
 - public expenditures 4:521*t*
 - Southern African Norwegian Training, Education and Development (SANTED) initiative 1:855–856
 - student and researcher mobility 4:287*f*
 - tertiary education expenditures 4:500*f*
 - tuition fees 4:500*f*
- inequality–education relationship
 - employment rate–educational attainment relationship 2:278*f*
 - immigrant status 2:275*f*
 - test score–social background relationship 2:273*f*
- information and communication technology (ICT)
 - computer access at home 8:25*t*
 - computer access in school 8:23*t*

frequency of computer use in school and at home 8:26*t*
 mathematics and science instruction 8:32*t*
 student proficiency and confidence 8:31*t*
 information and communication technology (ICT) access and equity 8:97
 popular adult education 1:163
 prison education programs 8:301, 8:301*t*
 research-quality assessments 4:308
 returns to education—education level 2:308*t*
 student achievement test performance 2:247*f*
 teacher age distribution 2:492*f*
 teacher shortages 2:494*f*
 teacher wages and salaries 2:496*f*
 teaching hours 2:493*f*
 vocational education and training (VET)
 age-related policies and programs 8:263*t*
 initial vocational preparation system 8:251*t*
 school-based programs 8:249
 structural and operational characteristics 5:711*f*, 5:712, 5:712*t*, 5:715*t*
 teacher training 8:507*t*
 vocational–academic education parity 8:435
 women in employment 2:56*f*
 nuclear imaging professions 3:52*t*
 null hypothesis 7:216, 7:313, 7:436, 7:444
 numerical responses 3:154
 numerosity 5:175
 Nurse–Family Partnership (NFP) program 2:44–45, 2:45*t*, 2:127
 nursing education 3:65–70
 critical thinking skills 3:67
 domain-specific teaching and learning 3:66
 formation experiences 3:68
 instructional strategies 3:65
 integrative teaching 3:67
 practical reasoning skills 3:66
 professional responsibilities 3:68
 Nussbaum, Martha 6:194, 6:195
 nutrition and dietetics 3:52*t*
 nutrition effects 2:427
 Nyerere, Julius 5:841

O

Oberti v Board of Education of the Borough of Clementon School District (1993) 2:717
 obesity 8:113
 objectified cultural capital 6:100
 objectives achievement curriculum model 1:337
 objectives-based measurements 4:32
 observational learning 2:670–671
 observational studies 7:354–359
 basic concepts 7:6
 challenges
 biases 7:355
 randomization 7:354
 variable measures 7:355
 characteristics 6:458
 definitions 7:354
 general discussion 7:358
 hidden bias
 challenges 7:355
 detection methods 7:356
 sensitivity analyses
 causality 7:358
 circumstance choices 7:357
 coherent hypotheses 7:358
 design considerations 7:357
 heterogeneity 7:358
 methodologies 7:357
 illustrative example
 bereavement effects 7:354
 lead exposure effects 7:354
 overt bias
 adjustment factors
 matched sampling 7:355
 model-based adjustments 7:356
 stratification 7:356
 challenges 7:355
 matched sampling
 algorithm matching 7:356
 goals 7:355
 optimal matching 7:356
 propensity score 7:355
 selection guidelines 7:355

structural controls 7:356
 thick descriptions 7:356
 observed-score equating methods
 basic concepts 4:51
 Kernel equating method 4:52
 occupational therapy 3:52*t*
 O'Connor, DJ 6:5
 odds ratio 7:58, 7:59*t*, 7:261, 7:262*t*
 Oedipus complex 6:92
 Office for Standards in Education (Ofsted)
 district-level reform efforts 6:300
 school-based accountability programs 3:520
 school inspections
 classroom assessment policies 3:429–430
 purpose 5:550
 research results 3:722
 self-inspection versus self-evaluation 3:717, 3:720
 officer candidate schools (OCSs) 3:7
 officer training schools (OTSs) 3:7
 official language of schooling 1:435–440
 curricula
 general discussion 1:435
 primary schools
 language curricula 1:436
 literature teaching 1:437
 reading curricula 1:436
 writing instruction 1:437
 secondary schools
 adolescent literacies 1:438
 digital technology 1:439
 language curricula 1:438
 literature teaching 1:438
 general discussion 1:435, 1:439
 influential factors 1:435
 offshore education 4:508
 Ogbu's two types of minorities theory 5:454
 Ohio 1:654*f*
 Oklahoma 1:654*f*, 2:65
 older learners 1:175–182
 characteristics
 age 1:176
 demographic data 1:176
 learning and intelligence 1:178
 preretirement education 1:177
 retirement 1:176
 stereotypes 1:177
 educational institutions
 accreditation 1:180
 characteristics 1:178
 general discussion 1:180
 human capital theory 1:175
 lifelong learning 1:175
 vocational education and training (VET) 8:236
 Oman 5:718–726
 background information 5:718
 early childhood education and care (ECEC) 2:111*t*
 education system
 administrative and supervisory structure 5:724
 challenges 5:725
 demographic data 5:718–719, 5:719*t*
 educational finance 5:724, 5:725*t*
 educational status 5:719*t*
 enrollment 5:719*t*
 goals 5:719
 manpower training 5:720
 monitoring and evaluation programs 5:724
 Omanization 5:720
 research programs 5:724
 structural and operational characteristics
 adult education 5:724
 basic education 5:721, 5:721*t*
 career guidance 5:722
 general discussion 5:720
 higher education 5:723
 international schools 5:723
 nonformal education 5:724
 postbasic education 5:722
 preschool education 5:721
 private schools 5:722
 special education 5:722
 teaching profession 5:724
 technical education 5:723

- Oman (*continued*)
 illiteracy rate 5:724, 5:724*t*
 information and communication technology (ICT) 8:25*t*
 secondary education initiatives 6:252
- One Laptop Per Child (OLPC) project 8:99, 8:128
- one-parameter logistic (1PL) model 7:286, 7:289–290
- one-way language learning programs 1:295–296
- online courses *see* e-learning; Internet-based education
- ontological phenomenology 6:450–451
- open and distance learning (ODL) **4:404–410**
 age-participation rates (APRs) 4:404
 background information 4:398–399, 4:404
 basic concepts 4:404
 challenges 4:409
 characteristics 4:407
 current trends 4:408
 dual-mode institutions 4:406
 evolutionary background 4:405
 expansion history 4:405
 general discussion 4:410
 higher education 4:598
 mega-universities 4:406
 Rwanda 5:764
 student characteristics 4:408
 virtual schools 8:129
- open educational resources (OERs) 4:408, 4:420, 5:103, 8:102
- open method of coordination (OMC) 7:537
- open-source online resources 8:5
- Open University 1:176, 1:178, 4:398–399, 4:405, 4:420, 4:425, 4:430
- oppositional defiant disorders 2:581
- oppression 6:38
- oral histories 1:25
- oralism 2:714
- oral language skills
 assessment strategies **3:285–292**
 background information 3:285
 characteristics
 discourse and pragmatic skills 3:288
 domains 3:287
 general discussion 3:286
 grammatical skills 3:288
 key features 3:287*t*
 lexical skills 3:288
 oral-language registers 3:287
 phonological skills 3:288
 development and usage guidelines
 accountability purposes 3:289
 diagnostic assessments 3:289
 instructional purposes 3:288
 operationalization 3:288
 test development 3:289
 diverse learners 3:291
 future directions 3:291
 importance 3:285
 test development
 general discussion 3:289
 standards-setting procedures 3:290
 validity establishment 3:289
 validity threats 3:290
 reading foundations 5:394
- order statistics **7:360–365**
 basic concepts
 basic distribution theory 7:361
 random variables 7:360
 statistical inference 7:362
 educational applications 7:362
 illustrative example 7:363
 importance 7:364
 large-sample statistical methods 7:236
- ordinal responses 3:154, 7:174
- ordinary least squares (OLS) method 2:188, 7:223, 7:483
- Oregon 1:654*f*
- Oregon State University (OSU) 1:848
- Organisation for Economic Cooperation and Development (OECD) **4:584–589**
 accreditation processes **4:384–389**, 4:537
 age-related policies and programs 8:262
 benefits 4:589
 childhood development 2:49
 competency-based learning model 8:405
 computer access in school 8:23*t*
 curriculum reform analyses 1:512
 early childhood education and care (ECEC)
 center-based programs 2:100
 Effective Provision of Preschool Education (EPPE) 2:90, 2:91*t*
 gender research 2:12
 governance 2:6
 integrated services 2:104
 leave policies 2:3
 parenting style–educational outcome relationship 2:153
 policies and programs 2:1
 staff support 2:162
 educational evaluation policy models **3:645–652**
 Keynesian welfare state model 3:645, 3:646*t*
 knowledge capitalist model 3:646*t*, 3:649
 neoliberal era 3:646*t*, 3:647
 purpose and scope 3:645
 educational evaluation systems 3:523
 educational policy role 4:586
 enrollment rates
 gender disparities 2:416*t*
 primary schools 2:415*t*
 secondary schools 2:416*t*
 external educational benefits 2:261
 functional role 4:584
 funding sources 4:585
 gender research 5:432
 graduation rates 2:325–326
 higher education
 access and equity patterns 4:334
 accreditation processes 4:537
 community partnerships **4:390–397**
 background information 4:390
 higher-education drivers 4:392
 higher-education governance barriers 4:395
 national policy barriers 4:393
 regional development drivers 4:390
 regional governance barriers 4:395
 science-driven models 4:392–393
 cross-border higher education (CBHE) 4:509, 4:587
 Directorate for Science, Technology and Industry (DSTI) 4:588
 funding considerations 4:221, 4:574
 future scenarios 4:587
 global dimensions 4:588
 graduate employment 4:354
 institutional expenditures 1:733*t*
 Institutional Management in Higher Education (IMHE) 4:588
 internationalization processes 4:540–541
 learning cities and regions 1:147
 parental expenditures 1:734*t*
 policies and programs 4:349, 4:586
 policy debates 4:598
 polytechnical institutions 4:349
 school-based vocational education and training 8:249
 state steering models 4:557–558
 student mobility 4:542
 tertiary education expenditures 4:499, 4:500*f*
 informal learning 8:240
 international evaluations
 assessment programs 3:687*t*
 design guidelines
 cross-country validity 3:684
 general discussion 3:684
 survey implementation comparisons 3:688
 target population comparisons 3:686
 general discussion 3:681, 3:688
 historical background 3:681
 policies and programs
 educational reform 3:684
 importance 3:682
 pace of change 3:683
 performance targets 3:683
 success factors 3:682
 research frameworks 3:681, 3:682*t*
 small-area estimation 7:415
 learning cities and regions 1:146
 lifelong learning
 funding considerations 1:133–138
 economic principles 1:133
 financiers 1:133
 financing mechanisms 1:134
 policies and programs 1:136

- policies and programs 1:89, 1:227
 - membership growth 4:585
 - national science policy 4:295
 - nonmember outreach 4:585
 - origins 4:584
 - professional recognition processes 3:1
 - pupil-teacher ratios (PTRs) 2:506*f*
 - purpose 4:584
 - recognition of prior learning (RPL) 8:397, 8:401
 - research activity 4:280
 - research-quality assessments 4:311
 - school-based evaluation 3:716
 - school reform 5:72
 - school supply shortages 2:506*f*
 - science and engineering university degrees 4:322, 4:323
 - Scottish education system 5:774–775
 - social capital impacts 5:89, 6:97
 - student achievement test performance 2:246, 2:247*f*, 3:726
 - student and researcher mobility 4:286–287, 4:287*f*
 - student attrition rates 4:467
 - sustainable development efforts 6:292
 - teacher shortages 2:494, 2:494*f*
 - tertiary type-B program survival rates 4:349, 8:463, 8:464*r*
 - United States education system 5:891
 - university adult and continuing education (ACE) 1:183
 - valid educational indicators 4:198
 - vocational education and training (VET) 8:353, 8:401
 - welfare state models 3:645, 3:646*r*
 - see also* International Association for the Evaluation of Educational Achievement (IEA); Programme for International Student Assessment (PISA)
 - organizational learning 1:47–52, 5:143–150
 - basic concepts 1:47, 5:143
 - challenges 5:148
 - constraints 5:147
 - dimensional characteristics
 - change processes 1:48
 - cognitive-behavioral approaches 1:49
 - content and learning processes 1:49
 - general discussion 1:47
 - knowledge-creating approaches 1:50
 - levels of analysis 1:48
 - sociocultural approaches 1:50
 - educational implications 5:145
 - importance 5:144
 - intervention strategies 1:50
 - power and politics 1:51
 - research background 1:47, 4:817–818, 5:143
 - schematic diagram 4:818*f*
 - organizational strategies 5:325
 - organized labor *see* labor unions; unions
 - orienting attention 5:166
 - orofacial myofunctional disorders (OMD) 2:654
 - other backward classes (OBCs) 1:590, 1:590*r*
 - Otherness 6:7, 6:94
 - outcome-based accreditation 4:211
 - out-of-school learning 5:457–463
 - characteristics 5:462
 - formal versus informal learning 5:457
 - general discussion 5:457
 - general versus specific learning 5:459
 - information and communication technology (ICT) 8:9
 - pedagogical-participatory dimension 5:462, 5:462*r*
 - representations versus meaning 5:459, 5:460*r*
 - thought versus action 5:461
 - written versus oral practices 5:458, 5:458*f*, 5:459*r*
 - racial and ethnic distinctions 1:598, 1:599*f*
 - research areas 1:600
 - Pacific students *see* New Zealand
 - Pädagogische Hochschulen* 5:585*f*, 5:587
 - Pakistan 5:727–736
 - background information 5:728
 - demographic characteristics 5:728
 - early childhood education and care (ECEC) 2:111*r*
 - education system
 - administrative structure 5:729
 - challenges 5:735
 - educational finance 5:735
 - examination system 5:734
 - gender disparities 5:735
 - goals 5:728
 - structural and operational characteristics
 - adult education 5:732
 - curriculum 5:733
 - early childhood education and care (ECEC) 5:729
 - enrollment rates 5:731*f*, 5:731*r*
 - general discussion 5:729
 - higher education 5:731
 - higher-secondary education 5:731
 - literacy rates 5:729
 - madrassab* system 5:732
 - medium of instruction (MOI) 5:733
 - middle-level education 5:731
 - nonformal education 5:732
 - number of institutions 5:731*r*
 - number of teachers 5:731*r*
 - primary education 5:729
 - private schools 5:732
 - religious schools 5:732
 - schematic diagram 5:730*f*
 - secondary education 5:731
 - vocational education and training (VET) 5:732
 - teaching profession 5:733, 5:733*r*, 5:734*r*
 - higher education
 - enrollment rates 4:522*r*
 - expansion effects 4:523
 - returns to education–cognitive skills 2:310*r*
 - Palau 2:111*r*
 - Palestinian Autonomous Territories
 - early childhood education and care (ECEC) 2:111*r*
 - information and communication technology (ICT) 8:25*r*
 - Trends in International Mathematics and Science Study (TIMSS) 5:645*r*
 - Panama
 - assessment system development 3:483
 - early childhood education and care (ECEC) 2:111*r*
 - earnings–education level relationship 2:306*r*
 - higher education enrollment 1:702*f*, 1:703*f*
 - Pancasila* 5:611
 - Panel Study of Income Dynamics (PSID) 2:178
 - panopticon society 6:172
 - paper-and-pencil tests
 - basic concepts 3:212
 - constructed response items
 - essay questions 3:214, 3:215*f*, 3:216*f*, 3:503*r*
 - general discussion 3:213
 - short-answer items 3:213, 3:503*r*
 - selected response items
 - fill-in-the-blank tests 3:503*r*
 - general discussion 3:212
 - matching items 3:212, 3:213*f*, 3:503*r*
 - multiple-choice items 3:212, 3:213*f*, 3:503*r*
 - true–false items 3:213, 3:503*r*
 - Papua New Guinea
 - early childhood education and care (ECEC) 2:111*r*
 - Eurocentric educational approaches 1:303
 - international programs 5:567
 - parables 7:592–593
 - Paraguay
 - assessment system development 3:483
 - higher education enrollment 1:702*f*, 1:703*f*
 - public subsidization 2:308*r*
 - returns to education–education level 2:308*r*
 - parameter estimation theory 7:40, 7:165
 - parametric IRT-based person-fit statistics 4:16
 - paraprofessionals 2:780–787
 - background information 2:780
 - general discussion 2:786
- P**
- PACE (primary assessment, curriculum, and experience) study 3:425
- Pacific Islander students
 - deaf-blind population 2:623*r*
 - educational attainment 1:802*f*, 1:803*f*
 - educational statistics 1:597*f*, 1:598*f*
 - higher education 1:595–601
 - demographic data 1:596*f*
 - educational attainment 1:598
 - enrollment rates 1:596, 1:596*f*, 1:597*f*, 1:598*f*
 - general discussion 1:595
 - participation background 1:596

- paraprofessionals (*continued*)
 instructional assignments 2:784, 2:785*t*
 training programs
 general discussion 2:781
 skill requirements 2:782
 United Kingdom 2:783*t*
 United Kingdom
 general discussion 2:781, 2:786
 higher-level teaching assistant (HLTA) standards 2:781, 2:782*t*
 instructional assignments 2:784, 2:785*t*
 national occupational standards (NOS) 2:781*t*
 number of teachers and teaching assistants 2:781*t*
 training programs 2:782, 2:783*t*
- parental choice 2:367–373
 adequate information resources 2:371
 economic perspectives 2:367
 general discussion 2:367, 2:371
 income and ability sorting 2:370
 parental preferences
 academic performance 2:369
 general discussion 2:368
 location 2:370
 school atmosphere 2:370
 school demographics 2:368
 parents' rights 6:19
 public versus private schools 1:585, 2:368
- parental distancing behavior 5:465
- parent–child relationships 5:450–456
 autonomy support 5:451
 child-rearing practices 2:151–156
 caring perspectives 2:152
 decision-making tensions 2:155
 diversity considerations 2:151
 general discussion 2:155
 marketized childcare 2:155
 multidisciplinary/multiprofessional work 2:154
 parental involvement 2:154
 parenting style–educational outcome relationship 2:153
 social class effects
 academic achievement 2:425
 after-school activities 2:424
 literacy gaps 2:423
 role models 2:424–425
 training programs 2:153
 working mothers 2:151
 early interactions 5:450
 learning and language participation 5:451
 out-of-home learning 5:453
 warmth and sensitivity 5:451
- parenting support and education 2:126–131
 early childhood education and care (ECEC) 2:83
 Early Head Start program (US) 2:128
 Early Start program (New Zealand) 2:126
 general discussion 2:126
 Healthy Families America (HFA) program 2:126
 Home Instruction Program for Preschool Youngsters (HIPPY) 2:129
 home-visiting programs (Jamaica) 2:127
 Incredible Years program
 basic concepts 2:129
 United States 2:129
 Wales 2:129
 lessons learned 2:130
 Nurse–Family Partnership (NFP) program 2:127
 Sure Start program (UK) 2:57, 2:128, 2:158, 2:159
 Turkish Early Enrichment Program (TEEP) 2:130
- parents' rights 1:290, 6:19
- parent–teacher conferences 3:406
- parity 3:341
- participation research
 adult learning 1:240–250
 barriers 1:237
 definitions 1:242
 demographic data
 adult education and training (AET) 1:6, 1:243, 1:244*t*
 informal learning 1:243
 economic perspectives
 cost limitations/credit restraints 1:248
 financial support 1:244*t*, 1:248
 incentive-to-invest factors 1:248
 job-related reasons 1:248
 motivation determinants 1:248
 future research areas 1:249
 general discussion 1:240
 informal learning
 characteristics 1:242
 demographic data 1:243
 motivation determinants
 disciplinary perspectives 1:243
 economic perspectives 1:248
 general discussion 1:243
 individual–social interaction model 1:249
 psychological perspectives 1:243
 psychosocial models 1:249
 social perspectives 1:247
 outcomes analyses 1:229
 participation rates 1:229, 1:240, 1:241*t*, 1:242*t*
 psychological perspectives
 attitudes 1:247
 job-related reasons 1:247
 motivational orientations 1:247
 motivation determinants 1:243
 personality traits and abilities 1:243
 recent trends 1:240
 educational barriers 1:237
 focus groups 6:401–405
 characteristics 6:401
 data collection and analysis 6:405
 functional role 6:402
 research design
 general discussion 6:402
 interview question guidelines 6:403
 moderating approaches 6:404
 participant selection 6:403
 usefulness 6:405
 immigrant education
 regular education 8:284*f*
 vocational education and training (VET) 8:282, 8:283*f*, 8:284*f*
 work-related vocational education and training (VET) 8:284*f*, 8:285*f*
- participant observation 6:442–448
 boundaries 6:445
 characteristics 6:443
 direct participation 6:443
dramatis personae 6:444–445
 ethical practices 6:444
 field experiences 6:446
 intentional networks 6:444, 6:444*f*, 6:446*f*
 mirror groups 6:446*f*
 ontological commitment 6:445
 qualitative inquiry 6:493
 significance 6:447
 stranger sensitization 6:445
 symbolic interactionism 6:445
- play 5:332
- situated learning 5:335–339
 apprenticeships 5:336
 educational applications 5:337
 prior learning assessment (PLA) 1:172
 professional development 7:549, 7:549*t*
 sense of identity 5:336
 theoretical perspectives 5:335
- participative leadership 5:45
- participatory action research (PAR) 3:597, 6:338
- Partners-In-Learning (PiL) initiative (Microsoft) 8:73, 8:74
- Pascal virtual community 1:146
- passive participation 6:443
- passive rehearsal 5:324
- Passmore, John 6:110–111
- patents
 background information 4:301
 exceptions 4:301
 personnel coverage 4:301
 royalties 4:301
- path analysis 7:16
- path diagrams 7:200
- Pavlovian theory 5:498
- Peabody, Elizabeth 1:514
- peace education 1:850–854, 6:22–28
 central trends
 critical education trends 6:23
 liberal trends 6:24
 modern, positivistic trends 6:22
 multicultural trends 6:23–24

- positive versus negative peace 6:24
- postcolonialist trends 6:24
- postmodern trends 6:23
- challenges 6:26
- concept of peace 6:24
- counter-education 6:26
- culture of peace movement 1:852
- definitions 1:850
- disarmament education 1:851
- historical background 1:850
- human rights 1:852
- international law 1:851
- maximalist versus minimalist perspectives 1:850
- methodologies 1:853
- peace advocacy 1:852
- peace research 1:852
- philosophical perspectives 1:852
- theoretical perspectives 6:22
- UNESCO (United Nations Educational, Scientific and Cultural Organization) 1:851, 6:22
- Pearson chi-squared (χ^2) tests 7:179, 7:190, 7:291
- Pearson, Karl 7:429
- Pearson product-moment correlation coefficient 7:260
- pedagogical content knowledge 7:656–661
 - basic concepts 7:600, 7:656
 - criticisms 7:659
 - developmental process
 - design considerations 7:659
 - general discussion 7:658
 - student learning focus 7:658
 - subject-matter knowledge 7:658
 - teaching experience 7:658
 - discourse analysis 6:376
 - early childhood education and care (ECEC) 2:89
 - International System for Teacher Observation and Feedback (ISTOF) project 3:733
 - leadership research trends 4:742, 4:743f
 - mother tongue teacher education 7:739
 - research background 7:656
 - science teacher education
 - assessment methods 7:714
 - curricula 7:714
 - instructional strategies 7:713
 - student learning 7:713
 - teacher change 1:376
 - vocational teacher education 8:509
- Pedersen's model of cultural competence 2:566
- peer assessments 3:231–235
 - benefits 3:231
 - characteristics 3:231
 - cognitions and decision-making processes 7:512
 - formative assessments 3:362
 - general discussion 3:234
 - learning processes 3:232
 - learning regulation impacts 3:351
 - primary school science 3:296
 - reading skills 3:194
 - teacher education 3:566, 7:512
 - writing assessments 3:276
- Peer-Assisted Learning Strategies (PALS) 2:805
- peer learning 5:444–449, 6:636–642
 - beneficial promotion strategies
 - collaborative groups
 - altered expectations 6:639
 - reasoning and explaining skills 6:639
 - status relationships 6:639
 - structured group interaction
 - cognitive role specialization 6:640
 - explanation prompts 6:640
 - general discussion 6:639
 - group-work task manipulation 6:640
 - guided reciprocal questioning 6:640
 - reciprocal teaching 6:639
 - structured controversy 6:640
 - cognitive-developmental perspectives
 - cognitive elaboration perspective 6:637
 - knowledge co-construction 6:637
 - Piagetian perspective 5:331, 6:636
 - Vygotskian perspective 5:331, 6:637
 - cognitive theories 5:445
 - debilitating interpersonal processes
 - conflict considerations 6:638
 - help-seeking failures 6:638
 - negative socioemotional processes 6:639
 - uncoordinated communication 6:639
 - unequal participation 6:638
 - general discussion 5:447, 6:636
 - peer-assisted learning strategies (PALS) 2:761–762, 2:762t, 2:770
 - research areas 6:641
 - research background 5:444
 - social-behavioral perspectives
 - general discussion 6:636
 - motivational strategies 6:636
 - socially cohesive groups 6:636
 - sociocultural theories 5:446
 - teacher influence
 - discourse modeling 6:641
 - goal structures and norms 6:640
 - theoretical perspectives
 - group work 5:445
 - Piagetian perspective 5:444
 - situated learning 5:445
 - Vygotskian perspective 5:445
- peer relationships
 - acceptance and rejection 6:565, 6:693
 - class size 6:204
 - collaborative learning 6:686
 - educational outcomes 2:431–437
 - general discussion 2:431
 - influencing factors 2:431, 2:432f
 - research results 2:436
 - schoolmate/classmate characteristics 2:435
 - friendship 6:565
 - help seeking behaviors 6:656
 - immigrant populations 4:634
 - social development 6:587
 - social networks 6:705–711
 - academic performance 6:709
 - adolescents 6:707
 - characteristics 6:705
 - computer-mediated communication 8:116
 - deviant behaviors 6:709
 - research approaches 6:706
 - young children 6:706
- special needs students 2:794–801
 - chronological age 2:795
 - cultural influences 2:797
 - gender-related differences 2:795
 - general discussion 2:794, 2:798
 - inclusive interventions 2:798
 - program integration effects 2:797
 - research perspectives 2:794
- peer review processes 4:384
- peer tutoring/reciprocal reaching 2:802–807
 - adapted physical activity programs 2:524
 - basic concepts 2:802
 - benefits 2:805
 - design guidelines 2:802
 - importance 2:806
 - knowledge communities 8:192
 - pre-referral intervention strategies 2:817
 - processes and procedures
 - active presentation-response skills 2:805, 2:805t
 - general discussion 2:803
 - material content 2:805
 - preparation processes 2:804
 - specific-tasks structure 2:803
 - research background 2:802
- Peircean semiotics 5:222
- Peirce, Charles Sanders 6:86, 6:109, 6:475–476
- Pell grants 1:650, 1:651f, 1:656
- Pennsylvania 1:654f
- Pennsylvania Association for Retarded Children (PARC) v Commonwealth of Pennsylvania* (1972) 2:689, 2:716
- People Project 1:104
- Peoples' Republic of China *see* China
- perceptual representation system (PRS) 5:198, 5:198f
- per comparison error rate (PCE) measures 7:312
- per family error rate (PFE) measures 7:312
- performance assessments
 - alternative assessments
 - background information 3:203

- performance assessments (*continued*)
 realistic simulations 3:203
 workplace-related competency assessments 3:203
classroom assessment 3:214, 3:215*f*, 3:216*f*
fine arts education 3:262–263
gender-based learner groups 3:489
Greece 5:603
rubrics and scoring guides 4:124
secondary school science 3:301, 3:303*f*
Slovenia 5:787
Teaching for Understanding (TfU) 7:636
technology literacy and fluency 8:39
Uruguay
 education system
 attendance rate 5:878*f*
 dropout rates 5:876, 5:879*f*
 enrollment 5:876, 5:877*t*
 net schooling rate 5:878*f*
 population literacy and schooling 5:876, 5:876*f*
 Programme for International Student Assessment (PISA) 5:879, 5:880*t*, 5:881*t*
 reading and mathematics trends 5:881*t*
 repeating rates 5:876, 5:878*f*, 5:879*f*
 schooling rate by educational cycle 5:876, 5:877*f*
 young adults 5:876
performance–avoidance/performance–approach goals 6:538
performance feedback 3:226–227, 3:227*t*, 4:747
Peril and Promise (World Bank) 4:591
Perkins School for the Blind 2:621
Permanent Forum on Indigenous Issues 1:663
Perry Preschool Program study
 characteristics 2:40
 cost–benefit analysis 2:41*t*
 economic returns 2:41, 2:42*t*, 2:45*t*
 education–health association 2:169–170, 2:173
 effectiveness 2:101, 5:53, 8:257–258
 infant and toddler programs 2:71
 outcomes analyses 2:147, 2:452
 public investment benefits 2:64
 research reviews 2:40*t*
 simulation studies 2:42
personal autonomy 6:107
personal digital assistants (PDAs) 8:200
personal epistemology 5:211–217
 assessment strategies 5:214
 definitions 5:211
 development processes 5:211
 dimensionality 5:211
 domain-general versus domain-specific beliefs 5:212, 5:214
 educational implications 5:215
 learning impacts 5:213
 multiple information sources 5:194
personalized learning 6:227–232, 8:391–396
 Australia 8:393
 background information 8:231, 8:391
 basic concepts 6:227, 6:228*f*, 8:391
 challenges
 assessment changes 8:392
 cultural shifts 8:391
 information and communication technology (ICT) 8:392
 learners' role 8:392
 learner voice 8:395
 teachers' role 8:392
 curriculum development 6:229
 general discussion 6:227
 implementation strategies 6:231, 6:231*t*
 individualized learning 8:396
 learning skills framework 6:230
 metacognition 6:230
 New Zealand 8:395
 United Kingdom 6:231, 6:231*t*, 8:393
 United States 8:394
personal racism 6:38
person-fit analysis 4:15–19
 applications
 computer adaptive testing (CAT) 4:18
 general discussion 4:18
 misfitting patterns 4:18
 respondent characteristics 4:18
 Guttman errors 4:16
 limitations 4:19
 nonparametric IRT-based/group-based person-fit statistics 4:16
 parametric IRT-based person-fit statistics 4:16
 person response function (PRF) 4:17
 Rasch model 4:19
personnel administration 4:839–845
 challenges
 business-based accountability 4:844
 decentralized leadership 4:844
 general discussion 4:843
 quality instruction 4:844
 teacher shortages 4:843
 functional role
 compensation policies and practices 4:843
 evaluation policies and practices 4:842
 general discussion 4:841
 induction programs 4:841
 professional development 4:842
 recruitment 4:841
 selection methods 4:841
 supervision strategies 4:842
 general discussion 4:839, 4:844
 historical background 4:840
 policy considerations 4:840
 purpose 4:839
 person response function (PRF) 4:17
Peru
 assessment system development 3:483
 early childhood education and care (ECEC) 2:111*t*
 higher education enrollment 1:702*f*, 1:703*f*
 student achievement test performance 2:247*f*
pervasive developmental disorder—not otherwise specified 2:579, 2:662*f*, 2:663
Pestalozzi, Johann Heinrich 1:514
Peters, Richard Stanley 6:2, 6:5, 6:103, 6:109, 6:187
Pew Report (1989) 3:73
pharmacy education 3:58–64
 global-level education 3:58
 professional focus 3:58
 United States
 certification and licensing 3:59
 current issues
 curricular delivery methods 3:63
 experiential coursework 3:62
 outcomes assessments 3:63
 preprofessional coursework 3:62
 professionalism 3:63
 residency education 3:63
 general discussion 3:59
 medication therapy management (MTM) 3:61
 professional code of conduct 3:60
 responsibilities and services 3:60
 signature pedagogy 3:62
phenomenological variant of ecological systems theory (PVEST) 1:807*f*, 1:808–809, 1:809*f*
phenomenology 6:80–85, 6:449–455
 anthropological perspectives 6:82
 autonomy 6:84
 basic concepts 6:449
 corporeality perspective 6:81
 empirical and reflective methodologies 6:453
 epistemological perspectives 6:80
 ethnographic studies 6:451
 existential phenomenology 6:82, 6:111, 6:451
 freedom and responsibility 6:84
 as a human science 6:451
 I–world relationship 6:81
 life experiences 6:453
 pedagogical phenomenology 6:451
 reduction methodology
 eidetic reduction 6:82, 6:453
 experiential reduction 6:453
 general discussion 6:452
 hermeneutic reduction 6:452–453
 heuristic reduction 6:452
 Sartre's perspective 6:82
 subjectivity versus intersubjectivity 6:82
 traditional perspectives
 ethical phenomenology 6:451
 existential phenomenology 6:451
 general discussion 6:450
 linguistic phenomenology 6:451
 ontological phenomenology 6:450–451
 transcendental phenomenology 6:450
Philip Morris (PM) 1:551

- Philippines 5:865–872
 background information 5:865
 early childhood education and care (ECEC) 2:111*t*
 education for sustainable development 1:746–752
 background information 1:750
 basic concepts 1:750
 educational implications 1:747
 educational role 1:749
 future directions 1:751
 general discussion 6:148
 importance 1:751
 Judeo-Christian traditions 1:747–748
 modern society values and cultural perspectives 1:747
- education system
 challenges
 improved access policies 5:872
 international education programs 5:871
 ladderization programs 5:871
 language-in-education policies 5:872
 mutual recognition programs 5:871
 quality assurance 5:871
 trifocalization reform 5:871
 educational finance 5:871
 goals 5:865
 governance 5:866
 higher education
 admission requirements 5:870
 Commission on Higher Education (CHED) 5:868
 degree programs 5:870
 grading systems 5:870
 institutional distribution 5:868*t*
 private higher education (PHE) 5:870
 quality assurance 5:870
 research activities 5:871
 state universities and colleges (SUCs) 5:869
 historical background 5:865
 structural and operational characteristics
 alternative education 5:868
 basic education 5:866
 early childhood education and care (ECEC) 5:866
 educational ladder 5:866, 5:867*f*
 elementary education 5:866
 higher education 5:866, 5:868, 5:868*t*
 nonformal education 5:868
 secondary education 5:867
 technical education 5:866
 teacher education 5:871
 gender-equity gap 2:56*t*
 indigenous population 1:748–749
 inequality–education relationship 2:273*f*
 mathematics curricula 1:454
 music teacher education 7:706
 regional map 5:869*f*
 student achievement test performance 2:247*f*, 2:417*t*
 test score–social background relationship 2:273*f*
- philosophy
 analytical tradition 6:108
 education research 6:29–35
 critical race theory 1:779–786
 background information 1:779
 centrality of experiential knowledge 1:781, 6:343
 centrality of race 1:780
 critical legal theory 1:779
 educational applications 1:783
 educational theory reconceptualizations 1:783
 empirical educational research 1:783
 future implications 1:784
 narrative practices 1:782
 policy and practice research 1:783
 related areas of study 1:782
 educational policies 6:30
 educational practices 6:30
 ethical and political significance 6:33
 general discussion 6:30, 6:34
 objectivity and subjectivity 6:32
 philosophical contributions 6:8, 6:29
 qualitative/quantitative debate 6:31
 research perspectives 6:29
- feminist philosophies
 analytic feminism 6:58
 continental feminism 6:58
 exemplars 6:58
 general discussion 6:58
 pragmatist feminism 6:58
 philosophy of education 6:103–107
 analytical tradition 6:6, 6:8
 contemporary perspectives 6:112
 criticisms 6:111
 and Peters 6:109
 postmodern/poststructural perspectives 6:111
 continental tradition 6:8
 distinguishing features 6:3
 diversity considerations 6:7
 educational purpose 6:103–107
 knowledge for knowledge's sake concept 6:103
 rational autonomy 6:7, 6:121
 and Richard Stanley Peters 6:103–107
 vocational training 6:104
 well-being and happiness 6:105
 education research 6:8
 feminist theories 6:56–62
 analytic feminism 6:58
 continental feminism 6:58
 critical theory perspective 1:495
 current perspectives 6:60
 education feminism 6:58
 exemplars 6:58, 6:59
 feminist philosophies 6:58
 pragmatist feminism 6:58
 wave feminist theory 6:57
 Western perspectives 6:56
 globalization effects 6:8
 great thinker traditions 6:3
 group beliefs 6:2
 identity formation 6:156
 justice and care 6:164–169
 caring perspectives 6:166
 conflicting perspectives 6:167
 educational applications 6:168
 female perspective 6:166
 justice as fairness 6:165
 principles of justice 6:165
 theoretical perspectives 6:164
 traditional justification approaches 6:168
 problem definitions 6:1
 purpose 6:1
 theoretical perspectives 6:2
 twentieth-century perspectives
 analytical movement 6:6
 influential trends 6:4
 and Richard Stanley Peters 6:5
 rational autonomy 6:121
 phonics and phonemic awareness 1:464, 3:191, 4:108, 8:151
 phonological awareness 2:654, 2:759, 2:761*t*, 5:317
 photography *see* image-based education research
 Photovoice technique 3:597–598
 phronesis 7:671
 physical education and sports 1:459–463
 background information 1:459
 formal examinations and matriculation 1:460
 general discussion 1:463
 models-based instruction 1:461
 national and state curricula 1:459
 sports education 1:461
 Teaching Games for Understanding (TGfU) 1:461, 1:462
 physically disabled students 2:640–645
 adapted physical activity programs 2:519–529
 basic concepts
 historical background 2:521
 individualized education plans (IEPs) 2:520
 theoretical/philosophical perspectives 2:522
 class organization and management strategies 2:525, 2:526*t*
 effective physical education pedagogy 2:522
 general discussion 2:527
 government policies and programs 2:520
 importance 2:520
 instructional strategies
 active teaching 2:523
 adventure education 2:523
 cooperative learning 2:524
 flexible groups 2:524
 peer tutoring/reciprocal reaching 2:524
 self-instructional formats 2:525
 standards 2:523, 2:524*t*

- physically disabled students (*continued*)
 task teaching 2:524
 teaching through questions/Socratic method 2:524
 professional preparation and development 2:526
 background information 2:640
 diagnostic and classification processes 2:579
 early intervention services 2:608–612, 2:641
 benefits 2:608
 collaborative environments 2:611
 general discussion 2:608
 importance 2:608
 limitations 2:609
 Milwaukee Project 2:609
 teacher preparedness 2:610
 learning potential maximization 2:642
 related services 2:640
 school-based services 2:854
 physical therapy 2:640, 2:837, 3:52*t*
 Piaget, Jean 5:218–224
 concept learning 5:387, 5:465
 constructivist theories 1:488–489, 5:221, 5:486
 current research areas
 assessments and interventions 5:220
 future directions 5:223
 general discussion 5:219
 pedagogical perspectives 5:219
 development epistemology
 actions 5:222
 frameworks 5:222, 5:222*f*
 future directions 5:223
 general discussion 5:221
 knowledge formations 5:221
 norms
 adults' norms 5:221*f*
 basic concepts 5:221
 children's norms 5:222*f*
 reasons 5:222, 5:223*t*
 early childhood cognition 2:72, 5:330
 early childhood curriculum 1:516
 human development theories 5:219
 intellectual development stages 5:487
 intelligence theories 4:2, 5:187
 learning process types 1:37
 learning theories 5:219
 moral and social reasoning 5:279
 peer learning 2:794, 5:205, 5:331, 5:444, 6:636
 psychological development 5:245–246
 Pica 2:581
 picture exchange communication system (PECS) 2:545
 Pillai's trace 7:327
 pilot testing 3:159
 Pilot Twelve-Country Study 3:687*t*, 4:650, 4:664
 Pirivena education 5:817, 5:817*t*
 PK-3 programs 2:44, 2:45*t*
 plague 1:602
 Plato 6:3, 6:349
 play 5:330–334
 first language learning 5:333
 intent participation 5:331
 participation research 5:332
 peer learning 5:331
 Piagetian perspective 5:330
 play communities 5:332
 research background 5:330
 threats/teasing/joking/play politics 5:332
 Vygotskian perspective 5:331
Plessy v. Ferguson (1896) 1:692, 2:394, 6:499
 pluralism 6:87–88
 podcasts 3:534, 4:431
 point estimation methods 7:366–373
 estimators
 Bayesian estimators 7:369
 definition 7:367
 evaluation methods
 basic concepts 7:367
 basic criteria 7:367
 mean squared error (MSE) 7:367
 unbiased estimators 7:367
 finding methods 7:368
 maximum likelihood estimate (MLE) 7:368
 method of moments 7:368
 item response theory (IRT)
 basic concepts 7:369
 maximum likelihood estimate (MLE)
 conditional maximum likelihood (CML) method 7:371
 illustrative example 7:372, 7:372*t*, 7:373*t*
 joint maximum likelihood (JML) method 7:370
 marginal maximum likelihood (MML) method 7:371
 statistical inference 7:366
 Poisson distribution
 basic concepts 7:178
 deviance measures 7:179*t*
 discrete probability distributions 7:134
 illustrative example 7:180
 inference estimations 7:179
 Poland 5:737–743
 age-related policies and programs 8:263*t*
 background information 5:737
 class size 2:492*f*
 early childhood education and care (ECEC) 2:4*t*, 2:111*t*
 earnings–education level relationship 2:307*t*
 education levels 2:279*f*
 education system
 administrative and supervisory structure 5:740
 challenges 5:741
 educational finance 5:741
 goals 5:737
 monitoring and evaluation programs 5:741
 nongovernment organizations (NGOs) 5:740
 research programs 5:741
 structural and operational characteristics
 adult education 5:739
 alternative schooling approaches 5:739
 early childhood education and care (ECEC) 5:738
 education levels 5:738*f*
 higher education 5:739
 nonformal education 5:739
 post-compulsory education 5:739
 preschool education 5:738
 primary education 5:739
 secondary education 5:739
 technical education 5:739
 teaching profession 5:740
 enrollment rates 2:493*f*
 higher education
 loans and grants 4:501*f*, 4:502*f*
 structural and operational characteristics 5:739
 student and researcher mobility 4:287*f*
 tertiary education expenditures 4:500*f*
 tertiary type-B program survival rates 8:464*t*
 inequality–education relationship 2:278*f*
 information and communication technology (ICT)
 computer access in school 8:23*t*
 frequency of computer use in school and at home 8:26*t*
 student proficiency and confidence 8:31*t*
 teaching hours 2:493*f*
 vocational education and training (VET) 8:251*t*, 8:252*f*, 8:253*f*, 8:263*t*
 political engagement 1:868
 polytechnical institutions 4:348–349, 5:570*f*, 5:572
 Popper, Karl 3:106, 6:13, 6:185
 popular adult education 1:162–168
 basic concepts 1:162
 common dimensions
 action-oriented dimension 1:166
 pedagogical–participatory dimension 1:166
 political–ideological dimension 1:166
 current trends 1:165
 general discussion 1:162, 1:167
 historical background 1:163
 regional differences
 general discussion 1:163
 Latin America 1:164
 Nordic countries 1:163
 North America 1:165
 socioeconomic modernization 1:218
 popular culture 1:318–323
 educational applications 6:356
 influential factors 1:318
 research areas
 background information 1:318
 future directions 1:321
 general discussion 1:320

- popular culture as curriculum 1:320
- popular culture–curriculum interactions 1:320
- popular culture–curriculum relationship 1:319
- Scholar–Baller® program 1:863
- portfolios **3:236–242**
 - assessment strategies **3:243–248**
 - adult learning 3:236
 - alternative assessments 3:204
 - assessment conferences 3:406
 - background information 3:243
 - benefits 3:241, 3:246
 - classroom assessment 3:215, 3:451*t*
 - development processes
 - general discussion 3:238
 - reflective practices 3:238
 - selection and collection 3:238
 - electronic portfolios 3:239, 3:246
 - fine arts education assessments 3:262–263, 3:264
 - future directions
 - general discussion 3:247
 - practical considerations 3:247
 - tensions and contradictions 3:248
 - general discussion 3:236
 - group work assessments 3:346
 - history/social studies/civics assessments 3:325, 3:327
 - importance 3:244
 - large-scale assessments 3:237
 - National Record of Achievement (NRA) 3:237
 - purposes 3:237
 - reading skills 3:193
 - student learning impacts 3:198
 - summative assessments 3:238
 - teacher education 3:565
 - teacher judgment 3:238
 - test anxiety 3:502
 - United Kingdom
 - adult learning 3:236
 - background information 3:244
 - formative assessments 3:245
 - lessons learned 3:246
 - National Record of Achievement (NRA) 3:237, 3:245
 - personal development planning and progress file 3:245
 - Vermont case study 3:385
 - writing instruction 3:274
- definitions 3:236
- electronic portfolios
 - assessment strategies
 - classroom assessment 3:422
 - group work assessments 3:346
 - informal learning 8:244–245
 - records of achievement 3:246
 - teacher education 3:565
 - basic concepts 3:239
 - as an educational tool 8:219
 - evaluation policies and practices 3:565, 7:512, 7:619
 - general discussion 3:236
 - inquiry-oriented teacher education 7:607
 - oral language skills 3:288
 - prior learning assessment (PLA) 1:171
 - professional development 3:565, 7:686
 - strategic planning methods 4:261, 4:262*f*
- Portugal
 - accreditation processes **4:384–389**
 - African-descendant populations 1:573
 - age-related policies and programs 8:263*t*
 - childhood brain development 2:57*f*
 - class size 2:492*f*
 - early childhood education and care (ECEC) 2:4*t*, 2:56*t*, 2:111*t*, 2:114*t*
 - earnings–education level relationship 2:307*t*
 - education levels 2:279*f*
 - higher education
 - binary systems 4:617
 - foreign doctoral students 4:292*t*
 - institutional expenditures 1:733*t*
 - loans and grants 4:501*f*
 - parental expenditures 1:734*t*
 - private higher education (PHE) 4:622
 - student and researcher mobility 4:287*f*
 - tertiary education expenditures 4:500*f*
 - tertiary type-B program survival rates 8:464*t*
 - tuition fees 4:500*f*
- inequality–education relationship
 - employment rate–educational attainment relationship 2:278*f*
 - fathers' educational background 2:274*f*
 - test score–social background relationship 2:273*f*
- information and communication technology (ICT)
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - student proficiency and confidence 8:31*t*
- participation research
 - regular education 8:284*f*
 - vocational education and training (VET) 8:283*f*, 8:284*f*
 - work-related vocational education and training (VET) 8:284*f*, 8:285*f*
- school supply shortages 2:506*f*
- student achievement test performance 2:247*f*
- teacher age distribution 2:492*f*
- teacher shortages 2:494*f*
- teacher wages and salaries 2:496*f*
- teaching hours 2:493*f*
- vocational education and training (VET) 8:251*t*, 8:252*f*, 8:253*f*, 8:263*t*
- women in employment 2:56*f*
- positive affect negative affect scale (PANAS) 3:258
- positive behavior supports (PBS) 2:825, 2:825*f*
- positron emission tomography (PET) 5:340
- postcolonial theory **1:299–306**
 - adult education and training (AET) 1:99, 1:220
 - background information 1:299
 - cultural interfaces 1:304
 - decolonized curriculum 1:303
 - Eurocentric educational approaches
 - cultural identity factors 1:301
 - feminism 1:303
 - implications 1:301
 - language of instruction 1:302
 - legacy insights 1:302
 - racial and ethnic concepts 1:303
 - global educational inequities 1:300
 - post-postcolonial perspectives 1:304
- postconflict societies **4:514–519**
 - armed conflicts 8:441
 - basic concepts 4:514
 - higher education
 - functional role 4:515
 - future directions 4:519
 - intervention strategies
 - Afghanistan 4:518
 - general discussion 4:516
 - Iraq 4:518
 - South Africa 4:516
 - south-southeast Europe (SEE) 4:517
 - postconflict requirements 4:516
 - vocational education and training (VET) **8:441–447**
 - Afghanistan
 - conflict background 8:444
 - post-war society 8:444
 - skills development programs 8:445
 - background information 8:441
 - challenges 8:446
 - educational opportunities 8:441
 - South Africa
 - apartheid–National Qualifications Framework (NQF) transition 8:443
 - conflict background 8:442
- post hoc* comparisons 7:34
- postmodernism
 - curriculum studies analysis 1:343, 1:518
 - early childhood education and care (ECEC) **2:49–54**
 - childhood brain development 2:51
 - curriculum 1:518
 - demographic effects 2:50
 - ecological perspectives 2:49
 - family structure 2:50
 - future directions 2:54
 - government policies and programs 2:52
 - technology and media effects 2:51
 - well-being and happiness 2:52
 - historical background 6:14
 - identity theory 6:152
 - lifelong learning 1:91, 1:220, 6:115
 - philosophy of education 6:111
 - and poststructuralist theory 6:175

- postmodernism (*continued*)
 - rational autonomy 6:123
 - reason and rationality 6:186
 - see also* pragmatism
- post-professional architecture degrees 3:13
- poststructuralist theory 6:175–182
 - critical challenges 6:178
 - critical pedagogy 6:178
 - critical theory perspective 1:497, 6:176, 6:177
 - curriculum 1:499–504
 - critical theory perspective 1:497
 - feminist poststructuralist perspectives 1:502
 - historical background
 - deconstructionism 1:501
 - general discussion 1:500
 - reconceptualization 1:501
 - structuralism 1:500
 - theoretical perspectives 1:499
 - twenty-first century perspectives 1:503
 - educational implications 6:180
 - ethical practices 6:177
 - feminist perspectives 1:502, 4:799, 6:178
 - identity politics 6:39, 6:152
 - knowledge and power 6:177
 - linguistics 6:176
 - philosophy of education 6:3, 6:111
 - semiotics 6:478
 - subjectivity 6:177
 - theoretical perspectives 6:175
 - truth and knowledge 6:176
- poverty 6:643–652
 - Asian Americans 1:599
 - background information 5:58
 - capabilities approach 6:194
 - child development effects 2:33, 2:81, 6:643, 6:644f
 - compensating differentials models 2:469
 - disadvantaged schools 5:58–65
 - challenges 4:777
 - deficit views 5:58
 - policies and programs 5:59
 - reform impacts 5:60
 - teacher turnover 5:60
 - context-specific improvement 4:780
 - effective leadership 4:777–782
 - community engagement 4:779
 - distributed leadership 4:778
 - general discussion 4:781
 - improvement strategies 4:779
 - learner-centered leadership 4:779
 - research background 4:778
 - shared values and vision 4:778
 - staff development 4:779
 - improvement interventions 4:777
 - leadership impacts 5:61
 - urban environments 4:809–816
 - educational effectiveness 5:51–57
 - access and equity patterns 5:52
 - educational attainment 5:51
 - educational disadvantage
 - definitions 5:52
 - early childhood interventions 5:53
 - effective schools
 - characteristics 5:55, 5:55t
 - influencing factors 5:56
 - leadership 5:55
 - research background 5:53
 - ineffective schools 5:55
 - school effectiveness research (SER) program 5:51–52, 5:54
 - educational outcomes 1:675, 2:153, 2:273, 2:432
 - external social benefits 2:263t, 2:267
 - Latino students 1:610
 - parenting styles 2:153
 - policy implications 6:649
 - socioemotional development
 - characteristics 6:645
 - depth of poverty 6:645
 - mechanisms
 - early childhood education and care (ECEC) 6:647
 - family stress model 6:646
 - home environment 6:647
 - neighborhood characteristics 6:648
 - parenting behavior 6:646
 - persistence of poverty 6:646
 - timing of poverty 6:646
 - urban environments 4:809–816
 - challenges 4:810
 - defining characteristics 4:811
 - demographic background 4:809
 - general discussion 4:809
 - leadership strategies 4:815
 - local perspectives
 - estate schooling 4:814, 4:814f
 - extended area communities 4:814
 - Leadership on the Front-Line project 4:811
 - multiple communities 4:812, 4:813f
 - single/homogenous communities 4:812, 4:812f, 4:813f
 - World Bank (WB) activities 4:590
- power analysis 7:436–443
- power-load–margin model 1:249
- POWERSOURCE intervention 3:229
- practical knowledge 7:671, 7:750t
- practical reasoning 6:184, 7:750t
- practicum programs 7:607
- pragmatism 6:86–91
 - characteristics 6:91
 - classical pragmatism 6:86
 - philosophical perspectives 6:86
 - pragmatist feminism 6:58
- Prague Circle of linguists 6:477
- Prague Communiqué 1:773f, 4:567
- Praxis I 8:48–49
- precedents 7:592–593
- precision teaching 2:573
- prediction 7:479, 7:479f
- predictive validity 4:10, 4:170, 4:173, 4:181
- predict–observe–explain (POE) format 3:301, 3:302f
- predominantly white universities 1:768, 4:479–480
- prefrontal cortex 6:546–547
- preprimary education
 - Austria 5:517
 - England 5:551
 - Germany 5:585, 5:587, 5:588t
 - Lithuania 5:658
 - Suriname 5:827
- pre-professional architecture degrees 3:13
- pre-referral intervention strategies 2:815–819
 - basic concepts 2:815
 - effective strategies 2:817
 - general discussion 2:818
 - implications 2:818
- preschool education
 - accessibility 2:110–119
 - administrative issues 2:113
 - enrollment rates 2:111, 2:111t, 2:114t
 - governance 2:113
 - growth rates 2:110
 - international programs 2:118
 - quality assurance 2:116
 - universal versus targeted programs 2:117
 - Australia 5:512
 - Austria 5:517
 - Bulgaria 5:542
- cost-effective programs 2:38–48
 - Abecedarian Early Childhood Intervention Program 2:40, 2:40t, 2:41t, 2:42t, 2:45t
 - Chicago Child–Parent Centers 2:40, 2:40t, 2:41t, 2:42t, 2:45t
 - class-size reduction research 2:43, 2:45t
 - cost–benefit analysis 2:40, 2:41t
 - cumulative effects analyses 2:39, 2:40t
 - economic returns 2:41, 2:42t, 2:44, 2:45t, 2:46f
 - full-day kindergarten (FDK) effects 2:43, 2:45t
 - general discussion 2:38
 - key principles of effectiveness
 - comprehensive family services 2:47
 - educational content 2:46–47
 - general discussion 2:45
 - ongoing evaluation studies 2:47
 - PK-3 programs 2:46
 - teacher preparedness 2:46
 - outcomes analyses 2:38, 2:39f
 - Perry Preschool Program study 2:40, 2:40t, 2:41t, 2:42t, 2:45t
 - PK-3 programs 2:44, 2:45t

- research focus 2:38
- simulation studies 2:42
- supplemental programs 2:43, 2:45*t*
- transition programs 2:44, 2:45*t*
- England 5:551
- Estonia 5:558
- Finland 5:570, 5:570*f*
- France 5:577
- Germany 5:585, 5:587, 5:588*t*
- Ghana 5:592, 5:592*t*
- Greece 5:600
- growth rates 2:110
- Iceland 5:606, 5:607*f*, 5:607*t*
- importance 2:110
- Latvia 5:631
- Lebanon 5:638
- literacy and numeracy development 2:75–80
 - adult models 2:76
 - concept formation 2:75
 - early experiences 2:76
 - general discussion 2:75
 - research background 2:78
 - teaching practices
 - appropriate pedagogy 2:78
 - interest and awareness programs 2:77
 - reading materials 2:78
- Lithuania 5:658
- Mongolia 5:682, 5:682*t*
- moral and social reasoning 5:280
- Norway 5:710, 5:711*f*, 5:712*t*
- Poland 5:738
- Qatar 5:745
- Russia 5:751
- Rwanda 5:760
- Scotland 5:771
- Singapore 5:777
- Slovenia 5:783, 5:784*f*
- special needs students 2:602, 2:617*t*
- Sri Lanka 5:816
- Sultanate of Oman 5:721
- Suriname 5:827
- Taiwan 5:835, 5:836*f*, 5:837*t*
- Tanzania 5:843
- Thailand 5:850, 5:850*f*
- United States 5:889, 5:890*f*
- Uruguay 5:875*f*
- see also* early childhood education and care (ECEC)
- preservice education programs
 - adult learning
 - criteria design guidelines
 - co-construction 3:413
 - evidentiary support 3:414
 - feedback 3:413–414
 - goal setting 3:414
 - general discussion 3:410
 - research areas
 - adult learning 3:412
 - feedback 3:411
 - motivation 3:411
 - professional learning 3:413
 - self-assessments 3:411
 - summative assessments 3:415
 - arts teacher education 7:697
 - belief systems
 - change strategies 7:625
 - educational innovations 7:627
 - mathematics 7:626
 - research background 7:624
 - classroom assessment research and theory
 - criteria design guidelines
 - co-construction 3:413
 - evidentiary support 3:414
 - feedback 3:413–414
 - goal setting 3:414
 - general discussion 3:410
 - research areas
 - adult learning 3:412
 - feedback 3:411
 - motivation 3:411
 - professional learning 3:413
 - self-assessments 3:411
 - summative assessments 3:415
 - elementary education
 - academic and professional coursework 7:727
 - characteristics 7:726
 - examinations 7:727
 - practicum programs 7:728
 - tertiary degree requirements 7:727
 - information and communication technology (ICT) 8:217–221
 - challenges 8:220
 - India 8:75
 - public–private partnerships (PPPs) 8:75
 - purpose 8:217
 - teaching preparation strategies 8:218
 - technology as an educational tool
 - digital technology 8:219
 - distance education 8:219
 - electronic portfolios 8:219
 - usage preparation strategies 8:217
 - mentoring processes 7:749–756
 - mother tongue teacher education 7:738–743
 - challenges 7:741
 - field experiences 7:741
 - general discussion 7:738
 - pedagogical content knowledge 7:739
 - professional development
 - current issues 7:739
 - demographic shifts 7:740
 - literacy considerations 7:741
 - methods courses 7:740
 - subject-matter knowledge 7:738
 - pedagogical content knowledge
 - design considerations 7:659
 - student learning focus 7:658
 - subject-matter knowledge 7:658
 - teaching experience 7:658
 - school–higher education institution (HEI) partnerships 7:758
 - self-study 7:689
 - sociocultural factors 7:525–531
 - career attractiveness 7:529
 - certification and licensing 7:529
 - entry requirements 7:526
 - general discussion 7:525
 - globalization–regionalism tensions 7:525, 7:529
 - program curriculum 7:528
 - structural organization 7:527
- Primary Reading Test 3:189
- primary schools
 - arts teacher education 7:695–701
 - Australia 5:511*f*, 5:512
 - Austria 5:517
 - Bangladesh 5:526, 5:527*t*
 - Bulgaria 5:543
 - developing countries 2:415, 2:415*t*
 - Eastern Europe 3:476
 - England 5:551
 - Estonia 5:556
 - Fiji 5:564
 - Finland 5:570*f*, 5:571
 - France 5:577
 - Germany 5:585, 5:585*f*, 5:587, 5:588*t*
 - Ghana 5:592, 5:592*t*
 - Greece 5:600
 - Hong Kong
 - basic competency assessments 3:439, 3:440*t*
 - characteristics 3:438
 - learning-to-learn reforms 3:439
 - target-oriented curriculum (TOC) reform 3:438
 - Iceland 5:606, 5:607*f*, 5:607*t*
 - India 1:593*t*
 - Indonesia 5:612*f*, 5:613
 - instructional accommodations 2:731
 - Israel 5:619
 - Jamaica 5:627, 5:628*f*
 - Latvia 5:631
 - Lebanon 5:639
 - Lithuania 5:658
 - Malawi 5:667
 - Malta 5:676, 5:676*f*
 - Mongolia 5:682, 5:682*t*
 - Namibia 5:687
 - New Zealand

- primary schools (*continued*)
 characteristics 5:694
 classroom assessment 3:444, 5:694
 curriculum 5:694, 5:695*f*
 Niger 5:702, 5:706*t*
 Norway 5:710, 5:711*f*, 5:712*t*
 Pakistan 5:729
 Poland 5:739
 Qatar 5:746
 reading curricula
 government policies and programs 1:465
 official language of schooling 1:436
 phonics and phonemic awareness 1:464
 Russia 5:751
 Rwanda 5:760, 5:761*t*
 science education assessments **3:293–299**
 formative assessments
 evidence collection 3:294
 evidence interpretation 3:295, 3:295*t*
 feedback 3:296
 goals 3:294
 peer assessments 3:296
 policies and practices 3:294
 student self-assessments 3:296
 future directions 3:298
 group performance reports 3:297
 individual performance reports 3:296
 inquiry-based learning 3:293
 purpose 3:293
 summative assessments
 combination assessments 3:297
 evidence summaries 3:297
 policies and practices 3:296
 special tasks or tests 3:296
 Scotland 5:771
 self-regulation intervention programs 6:727
 Singapore 5:777
 Slovenia 5:783, 5:784*f*
 social studies curriculum 1:411
 Solomon Islands 5:793, 5:795*t*, 5:797*t*
 South Africa 5:801, 5:801*t*
 Spain 5:808, 5:809*t*
 Sri Lanka 5:816
 Suriname 5:827
 Taiwan 5:835, 5:836*f*, 5:837*t*
 Tanzania 5:843
 teacher education programs 7:526
 Thailand 5:850*f*, 5:852
 United Kingdom 1:674
 United States 5:889, 5:890*f*
 universal primary education (UPE) 6:250
 Uruguay
 repeating and dropout rates 5:876, 5:878*f*, 5:879*f*
 schooling rates 5:876
 structural and operational characteristics 5:875*f*
 Zanzibar 5:896, 5:897*t*, 5:898*t*
principal axis factoring (PAF) 7:163
principal component analysis (PCA) **7:374–377**
 background information 7:374
 basic concepts
 component loading 7:376, 7:377*t*
 correlation matrix 7:15, 7:258, 7:258*t*, 7:259*t*, 7:375, 7:376*t*
 general discussion 7:374, 7:377
 sample dataset 7:375*t*
 two-variable analysis 7:375, 7:375*f*
 variance 7:376, 7:376*f*, 7:376*t*
correspondence analysis (CA) **7:103–111**
 analysis of square asymmetric tables 7:111
 background information 7:103
 basic concepts 7:16
 asymmetric maps 7:105–106, 7:106*f*
 general discussion 7:103
 illustrative example 7:106, 7:106*f*, 7:107*f*, 7:107*t*
 inertia 7:104, 7:105*f*
 triangular coordination system 7:104*f*
 canonical CA 7:111
 contributions (proportions) 7:108
 data coding schemes 7:108
 multiple correspondence analysis (MCA) 7:106, 7:109*t*, 7:110*f*
 statistical inference 7:111
 subset CA/subset MCA 7:111
 supplementary points 7:108, 7:110*f*
 educational applications 7:4
 factor analysis comparison 7:377
 general discussion 7:163
 sampling considerations 7:377
principal evaluation **3:698–705**
 definitions
 stakeholder impacts 3:700
 working definition 3:701
 evaluated objects
 basic concepts 3:701
 information collection criteria 3:703
 leadership mega-skills 3:702
 evaluation fundamentals 3:698
 published results
 accountability-based evaluation (2003) 3:699
 general discussion 3:700
 role-based evaluation (1996) 3:698
 research guidelines 3:704
principals
 school restructuring role **5:66–71**
 changing roles
 cultural role 5:69
 environmental role 5:69
 instructional role 5:69
 political role 5:68
 conflicts and tensions
 accountability 5:67
 organizational restructuring 5:68
 professionalism 5:67
 structural coupling expectations 5:68
 distributed leadership 5:67
 general discussion 5:66
 implications 5:70
 professional learning communities 5:66
 reculturing processes 5:66
 results analysis 5:69
 rule changes 5:69
 social justice 5:67
 stakeholder relationships 5:69
prior learning assessment (PLA) **1:169–174**
 aims and approaches 1:170
 assessment approaches 1:170
 basic concepts 1:169
 benefits 1:172
 critical perspectives 1:172
 definitions and terminology 1:169
 general discussion 1:169, 1:173
 historical background 1:169
 learning theories 1:171
 practical approaches 1:171
prison education programs **8:300–305**
 educational models
 Australia 8:301*t*, 8:303
 behavior management 8:300
 general discussion 8:300
 Nordic model 8:301, 8:301*t*
 North America 8:301*t*, 8:302
 rehabilitation improvement 8:301, 8:301*t*
 risk-needs model 8:300
 United Kingdom 8:301*t*, 8:304
 general discussion 8:300, 8:304
 general equivalency diploma (GED) 2:243
 incarceration impacts 8:300
 vocational education and training (VET) 8:236
private higher education (PHE) **4:622–627**
 accountability 4:625
 Bulgaria 5:548*t*
 characteristics
 general discussion 4:623
 group identity 4:623
 nonelite subsector 4:624
 semi-elite institutions 4:623
 emergence and growth 4:597, 4:622
 funding considerations 4:625
 general discussion 4:622, 4:626
 governance 4:625
 Philippines 5:868*t*, 5:870
private schools
 Australia 5:514
 developing countries 6:285

- funding considerations 2:337
- No Child Left Behind Act (2001) 2:338
- Pakistan 5:732
- public versus private schools 2:333
- Rwanda 5:765
- Spain 5:809, 5:809*f*, 5:810
- Sri Lanka 5:817, 5:817*f*
- Sultanate of Oman 5:722
- Tiebout sorting and competition theory 2:390, 2:391
- private-sector training 1:207
- privatization 2:337–341
 - background information 2:337
 - educational management organizations (EMOs) 2:338
 - evaluation processes 2:339
 - financial approaches
 - educational vouchers 2:338
 - evaluation processes 2:341
 - fees 2:339
 - philanthropic resources 2:339
 - tuition tax credits 2:339
 - general discussion 2:341
 - privatization criteria evaluations 2:340
 - regulation considerations 2:341
 - school characteristics
 - charter schools 2:337
 - home schooling 2:337
 - private schools 2:337
 - public schools 2:337–338
 - support services 2:338, 2:341
- probability density function (PDF) 7:467, 7:468*f*
- probability sampling
 - characteristics 3:155
 - cluster sampling 3:144, 7:392
 - design considerations 7:391
 - double sampling 3:145
 - general discussion 3:143
 - multistage sampling 3:145, 7:392
 - proportionate sampling 3:144
 - simple random sampling (SRS) 3:143, 7:391
 - stratified random sampling 3:144
 - stratified sampling 7:392
 - systematic sampling 3:144, 7:393
- probability theory 7:378–382
 - asymptotic theory 7:382
 - axioms 7:379
 - basic concepts 7:4
 - conceptual perspectives 7:378
 - conditional probability 7:380
 - mathematical theory 7:379
 - normality 7:381
 - random variables
 - basic concepts 7:380
 - independence concepts 7:381
 - uncertainty analyses 7:378
- Probeware 8:199
- problem solving and reasoning 5:265–272, 5:273–278
 - basic concepts 5:265
 - cognitive processes 5:274
 - cognitive research 5:248
 - creative processes 5:251
 - creativity assessment 4:172
 - current research areas 5:271
 - definitions 5:273
 - expertise acquisition 5:269, 5:270*f*
 - expert problem-solving stages
 - construction of a representation 5:266*f*, 5:268, 5:268*f*
 - general discussion 5:266
 - problem categorization 5:266*f*, 5:267, 5:267*f*
 - schematic diagram 5:266*f*
 - solution evaluation and storage 5:266*f*, 5:269
 - strategies and procedural applications 5:266*f*, 5:269
 - summation 5:269
 - general discussion 5:271
 - insight 5:276
 - knowledge domains 5:274
 - methodologies 5:265
 - nursing education 3:66
 - problem-solving transfer 5:275
 - problem space 5:276
 - productive versus reproductive thinking 5:275
 - real-life problems 5:277
 - rigid thinking 5:275
 - teacher-appropriate replacement behaviors 2:684
 - typology 5:273
- procedural memory 5:198, 5:198*f*
- professional architecture degrees
 - characteristics
 - admission requirements 3:16
 - assessment strategies 3:17
 - coursework 3:16
 - design and graphics 3:16
 - general education and electives 3:17
 - history/theory 3:17
 - practice courses 3:17
 - students and faculty 3:15
 - technology 3:17
 - degree programs 3:14
 - general discussion 3:13
- professional autonomy
 - basic concepts 4:434
 - engineering profession 3:48
 - teacher evaluations 3:677
- professional development 7:548–556
 - action research 7:540–547
 - beginning versus experienced teachers 7:545
 - characteristics 7:540
 - classifications 7:542
 - historical background 7:541
 - as a learning tool 7:540
 - practical approaches
 - active research projects 7:544
 - general discussion 7:543
 - interactive learning strategies 7:544
 - learning conditions 7:545
 - learning improvements 7:544
 - school-based action research project 7:541
 - adapted physical activity programs 2:526
 - career stages 7:582–586
 - background information 7:582
 - emerging environmental changes
 - external environmental factors 7:583
 - general discussion 7:583
 - organizational factors 7:583
 - personal environmental factors 7:585
 - teacher shortages 7:584
 - future directions 7:585
 - modeling approaches
 - early models 7:582
 - nonlinear dynamic models 7:583
 - research-based linear models 7:582
- continuing professional development 6:249–254
 - basic concepts 6:249
 - clergy education 3:35
 - effectiveness 6:252
 - entitlements 6:250
 - general discussion 6:249
 - impact evaluations 6:252
 - professional knowledge and skills 2:889
 - purpose 6:253
 - stakeholders and agendas 1:127–132
 - background information 1:127
 - educational agendas 1:130
 - implications 1:131
 - institutional agendas 1:128, 1:129*f*
 - professionalization effects 1:127
 - school and professional agendas 1:128
 - theoretical orientations 1:129*f*
 - vocational teacher education 8:508
 - workplace-based assessments 3:333
- effectiveness 7:569
- empirical research
 - general discussion 7:550
 - large-scale analyses 7:553
 - National Writing Project (NWP) 7:553
 - phase 1: mathematics teachers 7:552
 - phase 2: writing teachers 7:553
 - problem-solving cycle (PSC) model 7:552
 - general discussion 7:555
- higher education 4:411–418
 - challenges
 - accredited programs 4:415
 - aging demographics 4:414

- professional development (*continued*)
 - curricula vocationalization 4:414
 - new technologies 4:414
 - structural instabilities 4:415
 - theoretical perspectives 4:415
- contextual perspective and background 4:411
- educational approaches
 - background information 4:412
 - learner-focused approaches 4:413
 - organization-focused approaches 4:413
 - sector-focused approaches 4:413
 - teacher-focused approaches 4:413
- educational models 4:412
- future directions
 - distributed roles 4:416
 - general discussion 4:415
 - international and national collaboration 4:416
 - key performance indicators 4:415
 - professionalization 4:416
- general discussion 4:416
- terminology and meaning 4:411
- high-quality professional development
 - characteristics
 - content characteristics 7:549
 - general discussion 7:549
 - structural and operational characteristics 7:550, 7:551*t*
- lesson study 7:575–581
 - characteristics 7:575, 7:576*f*
 - future directions 7:579
 - general discussion 7:575
 - historical background 7:575
 - modeling approaches
 - challenges 7:578
 - student learning 7:578
 - teacher learning 7:576
 - three-part teacher learning model 7:577, 7:577*f*
 - research lessons 7:579
- mother tongue teacher education
 - current issues 7:739
 - demographic shifts 7:740
 - literacy considerations 7:741
 - methods courses 7:740
- professional development schools
 - research areas 7:672
 - school–higher education institution (HEI) partnerships 7:619, 7:758
- school-based prevention programs 2:823
- school organization 7:557–562
- special education teachers 2:889
- supportive school policies and practices 3:466*t*, 3:467, 3:469*f*
- teacher educators 7:681–688
 - characteristics 7:681, 7:682*f*
 - common activities
 - advanced academic degrees 7:684
 - feedback opportunities 7:685
 - portfolios 3:565, 7:686
 - professional organizations and conferences 7:684
 - research and scholarship 7:684
 - self-study 7:685
 - short courses and lectures 7:684
- constraints
 - motivation and confidence 7:686
 - support 7:686
 - time and resource availability 7:686
- context
 - off-job learning 7:683
 - on-job learning 7:683
- feedback opportunities
 - colleague-based feedback 7:685
 - student-based feedback 7:685
 - superior-based feedback 3:563, 7:685
- formative assessments 3:563–571
 - assessment instruments 3:563
 - deployment stages 3:563, 3:564*f*
 - design guidelines 3:564, 3:565*t*
 - guiding principles 3:564*t*
 - informative feedback 3:563
 - intervention strategies 3:564
 - peer assessments 3:566
 - portfolios 3:565
 - purpose 3:563
 - research-based principles 3:567, 3:568*t*
 - self-assessments 3:567
 - general discussion 7:686
 - purposes
 - beginning teachers 7:682
 - competency development 7:682
 - educational improvements 7:681
 - external standards 7:681
 - professional status 7:682
 - teacher leadership 5:43
 - technological trends 7:554
 - technology-mediated leadership development programs 5:123, 5:123*t*
 - traditional versus contemporary approaches 7:548, 7:549*t*
 - vocational teacher education 8:508, 8:516
- workplace-based assessments 3:330–336
 - active knowledge transformation 3:334
 - challenges 3:335
 - competence-based assessments 3:332
 - general discussion 3:330
 - initial education and training 3:330
 - knowledge development and transformation 3:333
 - knowledge-intensive workplaces 3:331
 - lifelong learning 3:334
 - reflective formative assessments 3:332
- workplace learning 7:569–574
 - contextual levels 7:571
 - effectiveness 7:569
 - general discussion 7:569, 7:573
 - theoretical perspectives 7:570
 - workplace conditions
 - critical factors 7:571
 - feedback opportunities 7:572
 - learning resources 7:573
 - shared norms 7:572
 - task complexity 7:571
 - teacher participation and collaboration 7:572
 - work process formalization 7:572
- professional learning communities 3:539–547, 5:151–157
 - characteristics
 - community 5:151
 - key features 3:539, 5:153
 - learning 5:152
 - membership 5:152
 - professionalism 5:151
 - development processes
 - evolutionary development 5:156
 - general discussion 5:154
 - leadership support
 - distributed leadership 5:155
 - general discussion 5:154
 - principal leadership 5:155
 - promotion of learning
 - collective learning 5:156
 - general discussion 5:155
 - inquiry-based learning 5:155
 - work-based learning 5:155
 - social resources 5:155
 - structural resources 5:155
 - diverse cultures 4:828
 - educational leadership 4:700, 4:743, 5:66, 5:154
 - effective leadership 4:748
 - general discussion 5:151, 5:156
 - impacts
 - capacity-building improvements 5:154
 - general discussion 5:153
 - learning improvements 5:153
 - student learning outcomes 5:154
 - teacher morale 5:153
 - reading curricula reform
 - school literacy curriculum construction 1:421
 - schoolwide report institutionalization 1:421
 - standards-based change (SBC) process 1:420
 - student and family ownership 1:421
 - research background 3:539
 - research results 3:545
 - school-based prevention programs 2:826
 - school environments 3:540, 3:541*t*, 3:542*t*
 - social capital 5:117, 5:117*f*
 - summary indicators
 - outcomes analyses
 - raw school performance 3:544
 - school performance–PLC factual characteristics correlations 3:545

- school performance–PLC indicator correlations 3:544
- value-added models (VAMs) 3:543
- professional and pupil learning ethos 3:542, 3:542*t*
- staff involvement 3:543, 3:543*t*
- teacher change 1:376
- professional research journals 4:731–739
 - article themes 4:735, 4:736*t*, 4:737*t*
 - background information 4:738
 - Cases in Educational Leadership* (CEL) 4:732
 - editorial processes 4:732, 4:733*t*
 - Educational Administration Quarterly* (EAQ)
 - article submission data 4:733*t*
 - article themes 4:735, 4:736*t*, 4:737*t*
 - internationality 4:734
 - publication data 4:731, 4:732*t*
 - Educational Management Administration & Leadership* (EMAL)
 - article submission data 4:733*t*
 - article themes 4:735, 4:736*t*, 4:737*t*
 - internationality 4:734
 - publication data 4:731, 4:732*t*
 - electronic journals 4:732
 - internationality 4:734
 - International Journal of Educational Management* (IJEM) 4:734
 - Journal of Educational Administration* (JEA)
 - article submission data 4:733*t*
 - article themes 4:735, 4:736*t*, 4:737*t*
 - internationality 4:734
 - publication data 4:731, 4:732*t*
 - Journal of Leadership in Education* (JLE) 4:734
 - publication data 4:731
 - refereed articles 4:733
 - research methodologies 4:733
 - selected journals 4:732*t*
 - submission data 4:733*t*
- professional societies of comparative education 1:266–271
 - challenges 1:270
 - oldest societies 1:266
 - World Council of Comparative Education Societies (WCCES)
 - foreign education versus comparative education 1:269
 - international focus 1:268
 - member societies 1:267*t*
 - non-member societies 1:270
 - origins 1:266
 - society characteristics 1:267
 - world congresses 1:268*t*
- professions education 3:37–38
 - definitions 3:37
 - historical perspectives 3:37
 - research background 3:37
 - signature pedagogies 3:37
- program evaluation 3:706–712
 - adult education and training (AET) 3:548–554
 - design considerations 3:549
 - developmental changes and trends 3:548
 - effectiveness 3:553
 - general discussion 3:548
 - modeling approaches
 - formative assessments 3:550
 - general discussion 3:550
 - Jacobs' model 3:551, 3:552*f*
 - key questions 3:550
 - Kirkpatrick's model 3:551, 3:551*f*
 - self-evaluation 3:552
 - summative assessments 3:550
 - purpose 3:548
 - basic concepts 3:707, 7:151, 7:151*f*
 - comparison studies 3:707
 - cost analysis 3:585–591
 - basic principles 3:585
 - historical perspectives 3:585
 - importance 3:590
 - investment costs 3:586
 - main issues
 - cost–benefit analysis 3:587
 - cost–effectiveness analysis 3:588
 - cost model 3:586, 3:587*f*
 - general discussion 3:586
 - performance guidelines
 - cost–effectiveness analysis 3:589
 - cost model 3:589
 - data requirements 3:590
 - design requirements 3:589
 - objectives 3:589
 - results reporting 3:590
 - scale requirements 3:589
 - sensitivity analyses 3:590
 - evaluation frameworks
 - basic concepts 3:707
 - credible evidence 3:710
 - data collection strategies 3:710
 - evaluation focus 3:710
 - lessons learned 3:711
 - logic modeling 3:709, 3:709*f*
 - program descriptions 3:709
 - results analysis and interpretation 3:710
 - stakeholder involvement 3:522, 3:708
 - usage facilitation 3:708*f*, 3:711
 - general discussion 3:711
 - impact evaluations 3:706
 - mathematics education 3:653–659
 - attained curriculum
 - assessment tasks 3:657
 - conceptual diagram 3:654*f*
 - importance 3:657
 - performance analysis 3:657
 - curriculum
 - attained curriculum 3:654*f*, 3:657
 - characteristics 3:654
 - content analysis guidelines 3:656*t*
 - implemented curriculum 3:654*f*, 3:655, 3:656*t*, 3:657*f*
 - intended curriculum 3:654, 3:654*f*, 3:655*f*, 3:656*t*
 - formative and summative evaluations 3:653
 - globalization effects 3:658
 - research background 3:653
 - outcome evaluations 3:706
 - process evaluations 3:706
 - purpose 3:706
 - science education 3:660–666
 - accountability considerations 3:660
 - background characteristics 3:664
 - contextual evaluations 3:661
 - data analysis and reporting 3:664
 - formative evaluation tools 3:663*t*
 - outcomes evaluations 3:663
 - process model 3:660, 3:661*f*
 - process skills 3:662*t*
 - program implementation 3:662, 3:663*t*
 - variable measures 3:664
 - social studies/civic education programs 3:667–673
 - stakeholders 3:629–635
 - empirical research results 3:631
 - evaluation activities 3:629
 - evaluation utilization impacts 3:633
 - importance 3:633
 - Program Evaluation Standards
 - participation guidelines 3:629
 - theoretical perspectives 3:630
 - theoretical perspectives
 - broad involvement 3:631
 - deep involvement 3:631
 - organizational learning perspective 3:630
 - Program Evaluation Standards 3:630
 - social justice perspective 3:630
 - summative evaluations 3:706
- Program Evaluation Standards 3:629, 3:767, 3:768*t*
- Program for the Assessment of Veterinary Education Equivalence (PAVE) 3:72
- Programme d'analyse des systèmes de la CONFEMEN (PASEC) 4:89
- Programme for International Student Assessment (PISA)
 - achievement gaps 2:81
 - affective assessments 3:261
 - assessment accountability 3:379
 - background information 3:681, 4:45, 7:171
 - competency-based learning model 8:406
 - decentralized decision-making 5:74
 - educational attainment 5:51–52
 - educational evaluation systems 3:523
 - educational outcomes
 - gender differences 2:275, 2:277*t*
 - general discussion 1:315
 - immigrant status 2:275*f*
 - exploratory data analysis (EDA) 7:157
 - Finland 5:574, 6:292, 8:85

- Programme for International Student Assessment (PISA) (*continued*)
 gender research 5:432
 Germany 5:587
 governance models 7:535
 information and communication technology (ICT) surveys 8:22
 international/regional assessment programs
 Eastern Europe 3:475–476
 educational evaluation systems 3:523
 Hong Kong 3:438
 international data systems 2:184, 2:184*t*
 Latin America 3:483
 modeling approaches 4:89
 Latin America 3:483
 Latvia 5:635
 Lithuania 5:665
 mathematics education 3:280, 3:282
 and the NAEP 3:696
 new scientist training 4:323
 New Zealand 5:694
 progress maps 3:373, 3:373*f*
 reading curricula 1:466
 science-education programs 3:664
 small-area estimation 7:415
 social capital impacts 5:89
 student achievement test performance 2:246, 3:693, 3:726
 supply shortages 2:505, 2:506*f*
 test adaptations 4:166
 United States education system 5:891, 6:262
 Uruguay 5:879, 5:880*t*, 5:881*t*
 valid educational indicators 4:198
 program planning theory 1:53–57
 background information 1:53
 challenges 1:56
 conventional planning theory 1:53
 critical planning theory 1:55
 deliberative planning theory 1:54
 theory–practice gap 1:56
 Progress in International Reading Literacy Study (PIRLS)
 background information 3:681, 4:45, 4:656
 characteristics 3:687*t*
 educational attainment 5:51–52
 educational evaluation systems 3:523
 educational impacts 4:666
 funding considerations 4:664
 international/regional assessment programs
 Eastern Europe 3:475
 educational evaluation systems 3:523
 international data systems 2:184, 2:184*t*
 Latin America 3:483
 modeling approaches 4:89
 sub-Saharan Africa 3:449
 Latin America 3:483
 Lithuania 5:665
 and the NAEP 3:696
 New Zealand 5:694
 progress maps 3:373
 research studies 4:644
 small-area estimation 7:415
 student achievement test performance 2:416, 2:417*t*, 3:726
 Progressive Era 1:356, 1:515
 progress maps 3:369–377
 achievement assessments
 general discussion 3:373
 learning research advances 3:375
 reading skills analyses 3:374, 3:374*f*, 3:375*f*, 3:376*f*
 assessment improvements 3:371
 basic concepts 3:369, 3:370*f*
 general discussion 3:369, 3:376
 map construction 3:372, 3:373*f*
 outcomes analyses 3:372
 speaking progress map 3:370*f*
 progress monitoring *see* curriculum-based assessment (CBA)
 Project Essay Grade (PEG) 4:20, 8:51
 project research funding 4:281
 Project STAR (Student Teacher Achievement Ratio) 2:43, 2:45*t*, 2:444, 5:54, 7:227
 prolepsis 3:595
 propensity score matching 2:188, 3:100, 3:108
 property crimes 2:239
 proportionate sampling 3:144
 prosocial children 6:692
 prosody 5:398
 prospective teachers
 belief systems
 change strategies 7:625
 educational innovations 7:627
 mathematics 7:626
 research background 7:624
 evaluation policies and practices 7:511–517
 cognitions and decision-making processes
 background information 7:512
 peer assessments 7:512
 portfolios 7:512
 self-assessments 7:512
 simulation methods 7:512
 general discussion 7:516
 good teaching views 7:511
 instructional behaviors 7:511
 learning-oriented assessments 7:512, 7:513*t*
 purpose 7:511
 quality assessments 7:514
 standards-based assessments 7:513
 validity measures
 characteristics 7:514
 consequential validity 7:515
 constructs and assessment task criteria 7:514
 extended interpretations 7:514
 reliability 7:515
 science teacher education
 pedagogical content knowledge
 assessment methods 7:714
 curricula 7:714
 instructional strategies 7:713
 student learning 7:713
 philosophical orientations 7:713
 prior knowledge analyses 7:712
 subject-matter knowledge 7:712
 self-study 7:689
 Protestant Reformation period 1:97, 1:729
 prototypes 7:592–593
 psychoanalysis 6:92–96
 background information 6:92
 educational implications 6:94
 Freudian perspective 6:92
 Lacanian perspective 6:93
 Psychometric Entrance Test (PET) 4:13
 publication productivity 4:455–460
 age-related considerations
 age decrement hypothesis 4:458
 cumulative disadvantage hypothesis 4:458
 general discussion 4:458
 intellectual deadlock hypothesis 4:458
 obsolescence hypothesis 4:458
 seniority burden hypothesis 4:458
 utility-maximizing hypothesis 4:458
 current trends 4:455
 gender differences
 basic concepts 4:459
 motherhood 4:459
 networking and collaboration 4:459
 general discussion 4:459
 measurement approaches 4:455
 output analyses
 ability measures 4:456
 collaboration 4:457
 communication networks 4:457
 cumulative advantages and disadvantages 4:457
 location 4:457
 resource availability 4:457
 statistical data 4:456
 time and energy 4:456
 public–private partnerships (PPPs) 8:73–80
 background information 8:73
 future directions 8:79
 information and communication technology (ICT)
 advantages 8:74
 case studies
 critical analysis 8:78
 preservice teacher education programs 8:75
 School District of Philadelphia/School of the Future project 8:77
 State School Isaac Schraiber/reading and writing project 8:76
 critical analysis
 accountability 8:79

- holistic approaches 8:78
- pilot programs 8:78
- situation-and-needs-analysis phase 8:78
- stakeholders 8:79
- sustainability 8:79
- future directions 8:79
- Partners-In-Learning (PiL) initiative (Microsoft) 8:73, 8:74
- model guidelines 8:73
- Partners-In-Learning (PiL) initiative (Microsoft) 8:73
- public schools
 - developing countries 6:285
 - funding considerations 2:337–338
 - public versus private schools 2:333
- publishing industry 1:368–373
 - America's Choice 1:371
 - current trends 1:372
 - global context 1:369
 - knowledge control and influence 1:369
 - textbook production 1:368
 - United States 1:370
- purposive sampling 3:145, 6:523

Q

Qatar 5:744–749

- background information 5:744
- early childhood education and care (ECEC) 2:111*t*
- education system
 - administrative and supervisory structure 5:748
 - challenges 5:749
 - educational finance 5:748
 - goals 5:744
 - government policies and programs 5:744
 - literacy rates 5:745
 - monitoring and evaluation programs 5:748
 - research programs 5:748
 - structural and operational characteristics
 - adult education 5:747
 - early childhood education and care (ECEC) 5:745
 - further education and training (FET) 5:746
 - higher education 5:747
 - nonformal education 5:747
 - policies and programs 5:745
 - preschool education 5:745
 - primary education 5:746
 - secondary education 5:746
 - special needs students 5:746
 - staff and students 5:746*t*
 - teaching profession 5:746*t*, 5:748
 - vocational education and training (VET) 5:744
- information and communication technology (ICT)
 - computer access at home 8:25*t*
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - student proficiency and confidence 8:31*t*
- population 5:744, 5:745*t*
- secondary education initiatives 6:252

Q-Q plots 7:338

Qualification Card (Q-Card) 1:157*t*, 1:159

qualitative case studies 6:456–462

- characteristics 6:456
- design and implementation
 - data collection and analysis 6:459
 - general discussion 6:458
 - selection criteria 6:458
- research benefits 6:461
- write-ups 6:460

qualitative data management 6:463–474

- coding methods 6:470
- computer software
 - challenges
 - database size 6:473
 - data de-contextualization 6:473
 - interpretation processes 6:473
 - program structures 6:473
 - quantitative analysis 6:473
 - theoretical underpinnings 6:473
 - content analysis 6:472
 - historical background 6:471
 - systematization 6:471
- theory generation 6:472
- data collection
 - data summaries 6:466, 6:468*f*
 - early theory generation 6:466
 - face sheets 6:465, 6:467*f*
 - focus group data 6:464
 - general discussion 6:464
 - interview data 6:464, 6:465*f*
 - observational data 6:464, 6:466*f*
 - preliminary data analysis 6:464, 6:465*f*, 6:466*f*, 6:467*f*
 - textual data 6:464, 6:466*f*
 - visual documentation 6:467*f*
- general discussion 6:474
- philosophical perspectives
 - audience 6:464
 - researcher characteristics and beliefs 6:463
 - researcher position 6:463
 - research requirements 6:463
 - research styles 6:463
- post-collection management
 - block-and-file approach
 - data fragmentation 6:469, 6:470*f*
 - intact data 6:469, 6:469*f*
 - conceptual mapping 6:470, 6:471*f*
 - thematic analysis 6:468
- theory generation
 - computer software 6:472
 - data collection 6:466
 - methodologies 6:474

qualitative inquiry 6:505–513

- evidence 6:488–497
 - basic concepts 6:489
 - contextual perspectives
 - data availability and reliability 6:489
 - narrative authority 6:493
 - participant observation 6:493
 - research characteristics and methods 6:489
 - traditional versus postpositive approach 6:491
 - distance 6:495
 - field experiences 6:494
 - quality assessments 6:488
 - questioning skills 6:494
 - social location 6:494
 - validity 6:496
- general discussion 6:505, 6:512
- historical perspectives 6:506
- postmodern/poststructural perspectives 6:510
- postpositivist perspectives
 - connoisseurship/critique 6:509
 - ethical probity 6:509
 - evidence 6:491
 - general discussion 6:507
 - realist postpositivism 6:508
 - social constructivism 6:508
- social context 6:506
- technology evaluations 3:580

qualitative research methods

Computer-Assisted Qualitative Data Analysis Software (CAQDAS) 6:326–334

- background information 6:326
- data exploration methods
 - annotation tools 6:330
 - general discussion 6:330
 - text search tools 6:330
 - word frequency tools 6:330
- data integration methods
 - general discussion 6:328
 - incorporated or referenced data 6:328
 - multimedia data 6:329
 - qualitative and quantitative analyses 6:329
- data organization methods
 - data grouping and filtering 6:331
 - data linking/hyperlinking 6:332
 - general discussion 6:331
 - interpretive coding 6:331
- general discussion 6:333
- management practices
 - connection-making methods 6:332
 - data interrogation 6:332
 - data retrieval 6:333
 - general discussion 6:332
 - pattern and relationship identification 6:333

- qualitative research methods (*continued*)
 - subset comparisons 6:333
 - writing methods 6:332
 - project planning/project teamwork 6:327, 6:328*r*
 - supported qualitative tasks 6:326, 6:327*f*
- cultural–historical activity theory (CHAT) 6:360–366
 - activity concepts 5:494, 5:494*r*
 - activity systems 5:494, 5:495*f*
 - background information 5:496
 - benefits 6:366
 - developmental background 6:360
 - educational approaches
 - development theories 6:361
 - learning 6:361
 - learning–development relationship 6:361
 - research examples 6:362
 - general discussion 6:360
 - and Lev Vygotsky 6:360
 - research examples
 - general discussion 6:362
 - long division/gazzinta process 6:362, 6:365*r*
 - puzzles 6:362, 6:365*r*
 - reading remediation 6:363, 6:365*r*
 - research results and analysis 6:364
 - weaving instruction 6:363, 6:365*r*
 - theoretical perspectives 6:360
 - third generation systems 5:495, 5:495*f*
 - workplace learning 1:85
- focus groups 6:401–405
 - characteristics 6:401
 - data collection and analysis 6:405
 - functional role 6:402
 - research design
 - general discussion 6:402
 - interview question guidelines 6:403
 - moderating approaches 6:404
 - participant selection 6:403
 - usefulness 6:405
- quality assurance 3:457–463
 - academic achievement outcomes 2:408
 - academic values and expectations 4:438
 - applications
 - general discussion 3:457
 - high stakes school-based certification 3:458
 - low stakes school-based certification 3:458
 - university course assessments 3:459
 - vocational training assessments 3:458
 - challenges 3:462
 - characteristics 3:457
 - consistency goals 3:460
 - cross-border higher education (CBHE) 4:537
 - definition 3:713
 - design considerations 3:460
 - early childhood education and care (ECEC) 2:116
 - education system
 - Germany 5:588
 - Philippines 5:870, 5:871
 - Thailand 5:856
 - higher education 4:377–383
 - accreditation process 4:533–539
 - agency responses 4:535
 - background information 4:533
 - barriers 4:534
 - capacity development 4:537
 - cross-border education 4:537
 - current developments 4:536
 - future challenges 4:537
 - general discussion 4:538
 - improved dialog and cooperation 4:537
 - intergovernmental bodies 4:537
 - international accreditation 4:533
 - internationalization processes 4:533
 - professional accrediting bodies 4:535
 - university associations and networks 4:534
 - assessment strategies 3:252
 - Bologna Accord (1999) 4:569
 - China 4:525
 - external quality assurance practices
 - basic concepts 4:377, 4:378*r*
 - examination systems 4:381
 - information availability 4:381
 - national qualifications frameworks 4:379
 - quality assessments 4:380
 - governance 1:728–729, 1:731, 1:734
 - importance 4:382
 - India 4:525
 - Malaysia 4:525
 - Philippines 5:870
 - leadership development programs 4:753
 - participative strategies 3:459
 - proactive strategies 3:459
 - research areas 3:462
 - social moderation versus statistical moderation 3:459
 - strong quality control (high stakes)
 - expert moderation 3:461
 - group/consensus moderation 3:461
 - panel moderation 3:461
 - teacher's role 3:389
 - vocational education and training (VET) 8:482–488
 - general discussion 8:482, 8:487
 - knowledge and skills specifications 8:482
 - national qualifications frameworks 8:483
 - provider risks
 - assessment risks 8:485
 - characteristics 8:483
 - funding risks 8:485
 - resources risks 8:484
 - VET system 8:486
 - weak quality control (low stakes)
 - assessor meetings 3:461
 - assessor partnerships and conversations 3:461
 - self-audits 3:461
- quasi-experimental research
 - class-size reduction research 2:448
 - controllability 3:128
 - evaluation research methodologies 3:740
 - general discussion 2:187, 3:133, 7:128
 - posttest-only designs
 - nonequivalent groups
 - basic concepts 3:129
 - matching approach 3:130
 - multiple control groups 3:130
 - proxy pretests 3:129
 - regression discontinuity (RD) designs 3:134–141
 - basic concepts
 - general discussion 3:134
 - model specification 3:134
 - pre-post design 3:135*r*
 - with treatment and interaction effect 3:136*f*
 - with treatment effect 3:136*f*
 - without treatment effect 3:135*f*
 - bias 3:134
 - characteristics 3:140
 - design requirements
 - assignment variables 3:135
 - cutoff point selection 3:136
 - extensions 3:137
 - empirical research methods 2:189
 - evaluation practices 2:27
 - general discussion 7:128
 - program evaluation 2:148
 - random assignment approximations 3:98, 3:109
 - RD design–quasi-experiment combination 3:140
 - RD design–randomized experiment combination 3:139, 3:140*f*
 - theoretical perspectives
 - randomized experiments 3:137, 3:138*f*
 - selection process 3:138
 - threats 3:139
 - technology evaluations 3:579
 - time series designs 7:128
 - two-group designs 3:128–133
 - general discussion 3:133
 - posttest-only designs
 - general discussion 3:129
 - nonequivalent groups 3:129
 - pretest–posttest control group designs
 - general discussion 3:131
 - internal validity 3:94
 - multiple pretests 3:132
 - reversed-treatment control groups 3:133
 - selection–maturation threats 3:131
 - switching replications 3:132

quasi-markets 8:496–497
 see also training markets
 Quasi-Monte Carlo (QMC) 7:302
 quasi-symmetry analysis 7:63
 queer studies curriculum 1:281
 question–demonstration assessments 3:301, 3:302^f
 questionnaires 5:214
 QUEST (quality of education systems transnationally) project 3:428
 quota sampling 3:145, 3:157
 QWERTY keyboards 5:232

R

race/racism 6:36–43
 academic achievement outcomes
 neighborhood characteristics 2:432
 schoolmate/classmate characteristics 2:435
 admissions tests 1:846
 adult education 1:77–82
 background information 1:77
 historical perspective
 civil rights movement 1:77–78
 general discussion 1:77
 multiculturalism 1:79
 research background 1:79
 social justice movement 1:79
 regional perspectives 1:81
 Southern hemisphere 1:80
 Western societies 1:77
 anti-racist education 1:739
 basic concepts 1:881
 Brazil 1:760–766
 critical race theory 6:341–347
 background information 6:341
 basic concepts
 civil rights progress 6:343
 critique of liberalism 6:343
 experiential knowledge 6:343
 interest convergence principle 6:343
 racism 6:343
 storytelling 6:343
 White supremacy 6:342
 general discussions 6:346
 misunderstandings 6:344
 origins 6:341
 qualitative research 6:490
 unresolved issues 6:345
 definitions 6:37
 demographic data 1:687
 desegregation 2:394–401
 diversity discourse 1:823
 diversity research 1:693
 earnings differentials 2:288–297
 ability/educational quality differences 2:293
 Black–White test-score gap 2:294, 2:435
 Brazil 2:296
 change effects 2:295, 2:295^t
 cross-sectional studies 2:291, 2:291^t
 educational attainment 1:804
 education equalization effects 2:291^t
 general discussion 2:288
 human capital model 2:290
 Israel 2:296
 minority–White income gap 2:292^f
 minority women 2:292
 United States 2:289, 2:289^t
 educational attainment 1:799–805
 black males 1:800, 1:839
 completion rates 1:800, 1:803, 1:839
 dropout rates 1:803
 earnings differentials 1:804
 18–24-year-olds demographics 1:802, 1:803^f
 enrollment rates 1:800, 1:801^f, 1:802^f, 1:803^f
 female–male participation rates 1:799, 1:804
 gender gap research 1:799, 1:804
 United States 1:799
 educational history 1:881–885
 Anglo-Protestant culture 1:882
 general discussion 1:881, 1:884
 reform efforts
 desegregation 1:692, 1:823, 1:883, 2:688
 market-based reforms 1:884
 post-civil rights era 1:884
 post-Civil War era 1:691
 schools as raced spaces 1:881
 state-sanctioned racism
 general discussion 1:882
 intelligence tests 1:883
 racial segregation 1:882
 United States 1:691
 educational statistics 1:581–588
 academic achievement outcomes 1:808
 choice effects 1:585
 continuing disparities 1:584
 demographic diversity 1:582
 digital technology implications 1:584
 disparity measures 1:581, 1:874
 dropout rates 1:874, 1:875
 education beyond borders 1:586
 18–24-year-olds demographics 1:802, 1:803^f
 enrollment rates 1:800, 1:801^f, 1:802^f, 1:803^f
 global village concept 1:587
 graduation rates 1:874
 incarceration rates 1:874
 male underachievement 1:808
 twenty-first century perspectives 1:581, 1:587
 United States 1:582
 essentialist perspectives 6:39
 ethnic and women's studies programs 1:626
 general discussion 1:77
 historical perspectives 1:554
 identity politics 6:39, 6:152
 inequality–education relationship 2:274, 2:432
 life expectancy 1:602, 1:603^f, 1:603^t
 mortality rates 1:606, 1:607^t
 moving-to-opportunity (MTO) program 2:434^t
 philosophical perspectives 6:36
 Plessy v Ferguson (1896) 6:499
 race-specific consequences 1:876
 reform efforts 6:36
 scientific racism 1:554–564
 background information 1:554
 contemporary science education
 communication norms 1:559
 educational impacts 1:559
 funding and training considerations 1:560
 craniometry 1:557
 eugenics 1:558
 historical perspectives 1:555
 medical research abuse 1:557
 monogenism/polygenism 1:558
 physical anthropology 1:557
 psychometric theories 1:558
 racial evolution theories 1:555
 segregation policies 2:775
 sociohistorical constructs 1:562
 theoretical retrofitting 1:558
 as a social construct 1:562, 6:36
 whiteness 6:36, 6:40
 see also civic engagement
 racial harassment *see* workplace mistreatment/abuse
 racialized masculinity 1:840
 radical empiricism 6:87–88
 radical hermeneutics 6:65
 radio-frequency identification (RFID) 1:154
 RAND Change Agent study 6:304
 random assignment approximations 3:97–103, 3:104–110
 advantages 3:102, 3:109
 causal inference 3:97, 3:105
 challenges 3:98, 3:107
 economics of education 2:189
 field experimentation
 basic concepts 3:84
 benefits 3:89
 implementation considerations
 blocking tactics 3:88
 cluster sampling 3:88
 general discussion 3:87
 intervention fidelity 3:88
 intraclass cluster correlation (ICC) 3:88
 unit congruence 3:88

- random assignment approximations (*continued*)
 internal validity threats 3:85, 3:86*t*
 functional role 3:97, 3:106
 internal validity 3:97
 methodologies
 fixed-effects model 3:101
 general discussion 3:107
 instrumental variables 3:100
 propensity score matching 3:100, 3:108
 regression discontinuity (RD) designs 3:98, 3:109
 randomized experimental designs
 blocked designs 7:128
 control group designs 7:127
 factorial designs 7:127
 general discussion 7:127
 technology evaluations 3:579
 variable relationship measures 3:97
 random coefficient models 7:172
 random-effects model 4:155, 7:8, 7:175, 7:277, 7:280*t*
 randomized control trials (RCTs) 2:27, 2:187
 random number generation 7:184–189
 background information 7:184
 computational statistics 7:94
 deterministic generation methods 7:185, 7:186*f*
 efficiency 7:187–188
 general discussion 7:10
 periodicity 7:186
 portability 7:187
 prominence 7:188
 random number distributions 7:185
 reproducibility 7:185
 research areas 7:188
 tests of randomness 7:187
 true random numbers 7:184
 random sampling 3:156
 rankings systems 4:546–553
 background information 4:546
 general discussion 4:552
 global university rankings
 general discussion 4:548
 Newsweek rankings 4:551
 performance factors 4:549, 4:550*t*
 Shanghai Jiao Tong University Institute of Higher Education (SJTJHE)
 rankings 4:204, 4:548, 4:548*t*, 4:550*t*
 Times Higher Education Supplement (THES) rankings 4:548*t*, 4:549
 holistic reputational rankings 4:551
 national rankings
 Centre for Higher Education Development (CHE) rankings 4:382, 4:547
 geographic distribution 4:546
 national survey of student engagement (NSSE) 4:381
 U.S. News & World Report (USNIWR) rankings 2:231, 4:379, 4:547
 reputational diversification 4:350
 research methodologies 4:551
 rapid automatized naming (RAN) 5:318–319
 Rasch model 4:111–122
 ability testing 4:2
 basic concepts 7:63
 contiguous intervals for ordered categories 4:112*f*
 dimensionality 4:112
 general discussion 4:111
 item characteristic curves 4:111, 4:112*f*, 4:116*f*, 4:118*f*
 item response theory (IRT)
 general discussion 4:117
 individual item-score patterns 4:15, 4:16*f*
 linear logistic test model (LLTM) 7:85, 7:89*t*
 maximum likelihood estimation (MLE) 7:371, 7:372*t*, 7:373*t*
 model controversies 4:119
 person-fit analysis 4:19
 polytomous ordered categories
 basic concepts 4:119
 latent structure 4:119, 4:120*f*
 latent thresholds 4:120, 4:120*f*
 reversed thresholds 4:121, 4:121*f*
 requirements 4:111
 response independence 4:112
 standard response formats 4:111, 4:112*t*
 sufficiency statistic
 basic concepts 4:112
 conditional estimates 4:113
 item parameter estimates 4:113, 4:113*t*
 personal parameter estimates 4:114
 relative item difficulties 4:113
 standard errors 4:114
 test equating and linking
 general discussion 4:114
 person-item location distributions 4:115*f*, 4:115*t*, 4:116*f*
 test of fit
 basic concepts 4:115
 differential item functioning (DIF) 4:116, 4:116*f*, 4:117*t*
 general test of fit 4:116
 item characteristic curves 4:116, 4:116*f*
 traditional test theory (TTT)
 attenuation resolution 4:118
 general discussion 4:117
 reliability 4:117
 standard error paradoxes 4:118, 4:118*f*
 ratio IQ 5:188
 rationality *see* reason and rationality
 Rawls, John 6:5–6, 6:77–78, 6:164, 6:185
 reactive attachment disorder 2:582
 reading comprehension 5:413–418
 accountability talk 5:352
 basic concepts 5:397
 cognitive perspective 5:352
 computer-assisted instruction 8:152
 definition 5:413
 general discussion 5:417
 instructional strategies 5:417
 intervention strategies 5:417
 key features
 reader development 5:413
 task considerations 5:414
 text–reader–task interactions 5:414, 5:414*f*
 texts 5:413
 multiple information sources 5:191–196
 educational implications 5:195
 information integration strategies 5:191
 learning processes
 expertise considerations 5:193
 general discussion 5:193
 personal epistemology 5:194
 reading tasks 5:194
 mental representation processes 5:192
 research areas 5:195
 text-based learning 5:191
 World Wide Web (WWW) 5:194
 neuroimaging studies 5:318
 predictors 5:351
 primary schools 1:436
 taxonomic perspective 5:415, 5:416*f*
 text-to-speech applications 8:153
 theoretical perspectives 5:416
 traditional instructional strategies 1:415, 1:416*t*
 reading curricula 1:415–422
 beginning-reading programs 6:221
 curricula globalization 1:421
 curriculum reform
 fragmented curriculum 1:420, 1:420*f*
 school literacy curriculum construction 1:421
 schoolwide report institutionalization 1:421
 staircase curriculum 1:419–420, 1:420*f*
 standards-based change (SBC) process 1:419, 1:419*f*, 1:420
 student and family ownership 1:421
 definitions 1:415
 elementary/middle-grades 1:465, 6:223
 government policies and programs 1:465
 primary schools
 government policies and programs 1:465
 official language of schooling 1:436
 phonics and phonemic awareness 1:464
 traditional instructional strategies 1:415, 1:416*t*
 twenty-first century considerations
 changing contexts 1:418
 changing readers 1:418
 changing texts and literacies 1:418
 general discussion 1:417
 Reading Literacy Study (RLS) 3:687*t*
 reading miscue inventories (RMIs) 3:191, 3:192*t*, 3:193*t*
 Reading Recovery Program 2:44, 2:45*t*
 reading skills 5:394–400
 acquisition theories 5:247
 assessment conferences 3:403–409

- basic concepts 3:404
- definitions 3:403
- historical background 3:404
- informal assessment systems 3:405
- parental involvement 3:406
- portfolio assessments 3:406
- purpose 3:403
- recent developments
 - general discussion 3:407
 - Learning Record 3:407
 - New Zealand Education Monitoring Project (NZEMP) 3:408
- self-assessments 3:406
- validity 3:408
- assessment strategies 3:189–195
 - affective aspects 3:190
 - authentic assessments 3:190
 - computer-assisted assessments 3:193
 - fluency 3:193
 - general discussion 3:194
 - group discussions 3:190
 - group tests 3:189
 - individual tests 3:190
 - informal reading inventories (IRIs) 3:191
 - motivational aspects 3:190
 - norm-referenced tests 3:189
 - peer assessments 3:194
 - phonics skills 3:191
 - portfolios 3:193
 - pre-reading sub-skills 3:191
 - reading miscue inventories (RMIs) 3:191, 3:192*t*, 3:193*t*
 - retellings 3:193
 - self-assessments 3:194, 3:194*t*
 - systematic observation 3:191
- computer-assisted instruction 8:150–157
 - beginning-reading programs 6:222
 - comprehension 8:152
 - educational applications 8:3
 - emergent literacy 8:150
 - fluency 8:151
 - general discussion 8:150, 8:155
 - Internet access 8:154
 - phonics and phonemic awareness 8:151
 - software systems 8:3
 - text-to-speech applications 8:153
 - theoretical perspectives 8:154
 - upper elementary grades 6:224
 - vocabulary 8:152
- curriculum-based measurement (CBM) 4:108
- formal reading instruction 5:395, 5:397*f*
- foundations
 - general discussion 5:394
 - listening comprehension 5:394
 - motivation 5:395
 - oral language skills 5:394
 - print concepts 5:394
 - schematic diagram 5:395*f*
 - semantic development 5:394
 - written language skills 5:394
 - written symbol system 5:394
- general discussion 5:399
- hearing impaired students 2:630
- information and communication technology (ICT) 8:76
- instructional strategies 6:221–226
 - beginning-reading programs
 - combined programs 6:222
 - computer-assisted instruction 6:222
 - cooperative learning 6:222
 - cost effectiveness 6:223
 - curricula 6:221
 - instructional-process programs 6:222
 - kindergarten-only programs 6:223
 - research background 6:221
 - research results 6:223
 - general discussion 6:225
 - struggling readers 6:221, 6:225
 - upper elementary grades
 - combined programs 6:225
 - computer-assisted instruction 6:224
 - cooperative learning 6:224
 - curricula 6:223
 - instructional-process programs 6:224
 - requirements 6:223
- neural networks 5:312
- neuroscience research 5:317–322
 - current technologies 5:317
 - developmental dyslexia
 - neural activation time course 5:320
 - neural networks 5:319
 - remediation effects 5:320
 - longitudinal studies 5:321
 - neuroimaging studies
 - developmental dyslexia 5:319
 - learning to read 5:318
 - neural activation time course 5:320
 - neural networks 5:319
 - remediation effects 5:320
 - phonological awareness 5:317
 - reading development models 5:317
- peer-assisted learning strategies (PALS) 2:761–762, 2:762*t*
- phonics and phonemic awareness 1:464, 2:759, 2:761*t*, 3:191, 4:108, 8:151
- preschool education 5:395
- printed word identification 5:395
- progress maps 3:374, 3:374*f*, 3:375*f*, 3:376*f*
- reading comprehension 5:397
- reading fluency 1:464, 3:193, 4:108, 5:398, 5:399*f*
- reading levels 5:398, 5:399*t*
- remediation approaches 6:363, 6:365*t*
- repeated reading 2:761
- special needs students
 - beginning reading 2:759, 2:761*t*
 - fluency 2:761
 - instructional strategies 2:759
 - peer-assisted learning strategies (PALS) 2:761–762, 2:762*t*
 - reading comprehension 2:762
 - think before reading, think while reading, think after reading (TWA) strategy 2:762, 2:763*t*
 - struggling readers 6:221, 6:225
- Realistic Mathematics Education (RME) 1:457
- real-life problems 2:845
- Realschule* 5:585*f*, 5:586
- reason and rationality 6:183–188
 - educational rationality 6:186
 - Enlightenment period 6:80, 6:183
 - philosophical perspectives 6:183
 - practical reasoning 6:184
- reasoning and reflective practice
 - case-based education methods 7:595
- cognitive research 5:248
- definitions 5:273
- moral and social reasoning 5:279–284
 - adolescents 5:281
 - cultural influences 5:282
 - definitions 5:279
 - educational implications 5:283
 - general discussion 5:283
 - historical background 5:279
 - middle childhood 5:280
 - preschool-aged children 5:280
 - stereotypes 5:283
- secondary school science 3:300
- self-regulated learning 5:505
- reciprocal effects model (REM)
 - academic self-concept (ASC)—achievement correlations
 - age generalizability 6:664
 - Baumeister–Marsh and Craven debate 6:664
 - causality considerations 6:662, 6:663*f*
 - cross-cultural generalizability 6:663
 - physical domain generalizability 6:664
 - positive-effects debate 6:664
- Recognising and Recording Progress and Achievement (RARPA)
 - project 8:393
- recognition of prior learning (RPL) 8:397–402
 - background information 8:231–232, 8:397
 - benefits 8:397
 - competency-based learning methods 8:400
 - constituents and drivers 8:398
 - lifelong learning 8:401
 - outcomes analyses 8:400
 - philosophical perspectives 8:399
 - terminology and jargon sources 8:349*t*, 8:350*t*
- recollection 5:342
- Reconstruction and Development Program (RDP) 8:443

- records of achievement 3:243–248
 - background information 3:243
 - future directions
 - general discussion 3:247
 - practical considerations 3:247
 - tensions and contradictions 3:248
 - importance 3:244
 - portfolios
 - background information 3:243
 - benefits 3:246
 - electronic portfolios 3:246
 - formative assessments 3:245
 - importance 3:244
 - lessons learned 3:246
 - National Record of Achievement (NRA) 3:245
 - personal development planning and progress file 3:245
 - United Kingdom 3:244
 - second-language learning 8:209
 - transcripts 3:247
 - United Kingdom
 - background information 3:244
 - formative assessments 3:245
 - lessons learned 3:246
 - National Record of Achievement (NRA) 3:237, 3:245
 - personal development planning and progress file 3:245
- recruitment difficulties 8:320
- recursive models
 - complex systems theory 1:483, 1:483^f
 - recursive partitioning 7:383–389
 - classification trees 7:384^f, 7:384^t, 7:385^t, 7:386, 7:386^f
 - extensions 7:387
 - inference trees 7:387, 7:388^f
 - prediction methods 7:383, 7:388
 - regression trees 7:386, 7:386^t, 7:387^f, 7:387^t, 7:388^f
 - recursive self-regulation model 5:504
- referent cognitive theory (RCT) 4:791–792
- reflective judgment interviews 5:214
- reflective practices 7:629–634
 - case-based education methods 7:595
 - characteristics 7:629, 7:750^t
 - current perspectives 7:630
 - diversity considerations 7:522
 - formative assessments 3:332
 - historical background 6:90, 7:629
 - inquiry-oriented teacher education 7:607
 - mentoring strategies 2:164
 - model continua
 - content 7:632
 - general discussion 7:631
 - means 7:632
 - purpose 7:631
 - narrative practices 7:611, 7:614
 - pedagogical concepts 7:590
 - significance 7:633
 - teacher education 3:567
 - teacher induction programs 7:565
 - Teaching for Understanding (TfU) 7:637
 - technological tools 7:619
 - transformative learning theory 1:22
 - workplace learning 1:84, 7:572
- reflexology 5:498
- Regents of University of California v Bakke* (1978) 1:692
- Reggio Emilia, Italy 2:70, 2:153–154, 3:317, 3:320
- registered training organizations (RTOs) 8:348
- regression analysis
 - basic concepts 2:188
 - educational efficiency/inefficiency analyses 2:210
 - effect size measures 7:447^t
 - experimental research design 7:129
 - extreme value theory 7:31, 7:31^f
 - institutional effectiveness measures 3:175, 3:176^f
 - least-squares regression 7:135
 - limitations 7:459
 - linear regression
 - basis expansion methods 7:138, 7:138^f
 - computational statistics 7:95
 - Fisher discrimination
 - general discussion 7:136
 - K-population example 7:137
 - two-population example 7:137
 - general discussion 7:135
 - large-sample statistical methods 7:236
 - least-squares regression 7:135, 7:136^f
 - logistic regression (LR) analysis 4:38^t, 4:39, 4:167, 7:136, 7:136^f, 7:138^f
 - multiple linear regression models 7:7
 - multivariate linear regression 7:324–331
 - basic concepts 7:324
 - general discussion 7:330
 - hypothesis testing 7:327
 - illustrative example 7:328, 7:329^t, 7:330^t
 - least squares (LS) criterion 7:326
 - mathematical formulation 7:325
 - statistical inference 7:326
 - support vector classifiers 7:137
- multiple linear regression models 7:7
- nonlinear regression analysis 7:339–346
 - basic concepts 7:8, 7:339
 - diagnostic studies 7:343
 - estimation methods
 - data sets 7:339
 - Gauss–Newton method 7:341
 - least-squares estimation 7:340
 - maximum likelihood estimation (MLE) 7:341
 - Michaelis–Menten regression function 7:341, 7:342^f
 - factorization methods 7:344
 - intrinsic curvature 7:346
 - model building
 - research areas 7:344
 - yield–density models 7:343
 - nonlinearity assessments
 - confidence regions 7:342
 - coordinate system 7:343
 - intrinsic nonlinearity 7:342, 7:342^f, 7:345
 - parametrization 7:341
 - parameter-effects nonlinearity 7:346
- recursive partitioning 7:383–389
 - classification trees 7:384^f, 7:384^t, 7:385^t, 7:386, 7:386^f
 - extensions 7:387
 - inference trees 7:387, 7:388^f
 - prediction methods 7:383, 7:388
 - regression trees 7:386, 7:386^t, 7:387^f, 7:387^t, 7:388^f
- regression continuity design 3:95
- regression discontinuity (RD) designs 3:134–141
 - basic concepts
 - general discussion 3:134
 - model specification 3:134
 - pre-post design 3:135^t
 - with treatment and interaction effect 3:136^f
 - with treatment effect 3:136^f
 - without treatment effect 3:135^f
 - bias 3:134
 - characteristics 3:140
 - design requirements
 - assignment variables 3:135
 - cutoff point selection 3:136
 - extensions 3:137
 - empirical research methods 2:189
 - evaluation practices 2:27
 - program evaluation 2:148
 - quasi-experimental designs 7:128
 - random assignment approximations 3:98, 3:109
 - RD design–quasi-experiment combination 3:140
 - RD design–randomized experiment combination 3:139, 3:140^f
 - theoretical perspectives
 - randomized experiments 3:137, 3:138^f
 - selection process 3:138
 - threats 3:139
 - see also* multivariate data analysis
- regression imputation procedures 7:284
- regulatory accountability 3:795
- Rehabilitation Act (1973) 2:596, 2:721, 2:727, 2:808
- rehearsal learning strategies 5:324
- reinforcer learning 2:671–672, 2:672^t
- relatedness 6:621
- relational context 4:704–705
- relative error variance 4:63^t, 4:64, 4:64^f
- relative functionalism 5:453
- relativism 6:87–88
- relaxation training 6:555^t
- reliability 4:27–30
 - aggregation effects 4:29
 - alternate form reliability 4:28, 4:29^t

- alternate form test–retest reliability 4:28, 4:29*r*
- automated essay scoring (AES) 4:22
- basic concepts 3:157, 4:27
- classical test theory 4:27
- educational assessments 4:48
- growth scores 4:29
- increase methods 4:29
- internal consistency/homogeneity 3:158, 4:28, 4:29*r*
- interrater/intrarater reliability 3:158, 4:28, 4:29*r*, 4:172
- magnitude factors 4:29
- peer assessments 3:232
- reliability coefficients 4:28, 4:29*r*
- reliability–validity relationship 3:157, 4:30
- science-education programs 3:664
- social studies/civic education program assessments 3:668
- split-half reliability 4:28, 4:29*r*
- standard error of measurement 4:27, 4:29, 4:29*r*
- test–retest reliability 3:157, 4:28, 4:29*r*
- religious and spiritual education 6:189–193
 - Australia 5:514
 - Bangladesh 5:527
 - curriculum 1:272–276
 - background information 1:272
 - recent trends and conflicts 1:272
 - theoretical perspectives
 - Charlemagne's curriculum 1:274
 - Christian education 1:274
 - Enlightenment period 1:275
 - ethical responsibilities 1:275–276
 - Greek pedagogy 1:273
 - medieval curriculum 1:274
 - Roman education 1:273
 - educational models
 - confessional approach 6:190
 - general discussion 6:190
 - interpretive approach 6:191
 - phenomenological approach 6:191
 - importance 6:193
 - Indonesia 5:612*f*, 5:613
 - Pakistan 5:732
 - philosophical and controversial perspectives 6:189
 - twenty-first century issues 6:192
- Renzulli Schoolwide Enrichment Triad Model (SEM) 2:831
- repeated reading 2:761
- repetitive-based injuries 8:113
- reproductive health programs 3:558
- resampling methods *see* jackknife method
- research and development curriculum evaluation model 1:338
- research commercialization 4:300–306
 - background information 4:300
 - copyright 4:303
 - criticisms 4:304
 - economic development
 - general discussion 4:302
 - incubators 4:303
 - research parks 4:302
 - start-up companies 4:302
 - technology transfer 4:303
 - government policies and programs 4:300
 - infrastructure 4:303
 - patents
 - background information 4:301
 - exceptions 4:301
 - personnel coverage 4:301
 - royalties 4:301
- Researching Effective Pedagogy in the Early Years (REPEY) project 2:86, 3:320–321
 - see also* Effective Provision of Preschool Education (EPPE)
- research lessons 7:579
- research parks 4:302
- research-quality assessments 4:307–315
 - background information 4:307
 - general discussion 4:314
 - international initiatives 4:311
 - key issues
 - assessment levels 4:312
 - autonomy 4:312
 - expert/peer review approaches
 - advantages 4:313
 - disadvantages 4:313
 - formative versus summative assessments 4:312
 - frameworks 4:311–312
 - general discussion 4:311
 - metrics/performance indicators
 - advantages 4:313
 - disadvantages 4:313
- national models
 - Australia 4:311
 - benchmark assessments 4:309
 - bibliometric analysis 4:307–308
 - driving forces 4:307
 - Hong Kong 4:310
 - Netherlands 4:309
 - performance-based funding 4:309
 - United Kingdom 4:310
- Reserve Officer Training (ROTC) programs 3:7
- resilience
 - disadvantaged urban youth 1:810
 - special needs students 2:840–844
 - background information 2:840
 - benefits 2:843
 - protective factors
 - family factors 2:842
 - general discussion 2:841
 - individual factors 2:841
 - school and community factors 2:842
 - third culture kids (TCKs) 4:638
- respiratory diseases 1:602, 1:603*f*
- response–prompt hierarchy 2:671, 2:671*t*, 2:672*f*
- Response to Intervention (RTI) 2:596, 2:681, 2:771, 2:816, 2:825, 2:825*f*, 4:104
- responsive curriculum evaluation model 1:337
- restricted self-adaptive testing (RSAT) 4:149
- Rett syndrome 2:579, 2:662*f*
- returns to education 2:298–304
 - causal relationships 2:298
 - civic engagement 2:235–238
 - economic perspectives 2:235
 - educational attainment correlations 2:235
 - future directions 2:238
 - public versus private schooling 2:236
 - years of schooling 2:235
 - current trends 2:302
- developing countries 2:305–312, 2:412–422
 - demand interventions 2:417
 - earnings differentials 2:305
 - age–earnings profiles 2:306*f*
 - basic concepts 2:305
 - causality considerations 2:309
 - cognitive ability 2:310, 2:310*r*
 - gender differences 2:306–307, 2:309*f*
 - OECD countries 2:307*t*, 2:308*r*
 - private returns to education 2:309*f*, 2:309*r*
 - public subsidization 2:308*r*
 - years of schooling 2:305, 2:306*t*, 2:307*f*, 2:307*t*, 2:308*r*
- enrollment
 - benefits 2:419*r*
 - gender disparities 2:416*r*
 - general discussion 2:415
 - income effects 2:417
 - primary schools 2:415*r*
 - secondary schools 2:416*r*
 - TIMSS data 2:416, 2:417*r*
 - future research areas 2:421
 - general discussion 2:412
 - government policies 2:420
 - literacy probability 2:417*f*
 - policy considerations 2:310
 - student achievement test performance 2:417*t*, 2:418*r*
- estimation by schooling equations
 - annual returns 2:302
 - basic calculations 2:299
 - bias research
 - ability bias 2:300
 - causal relationships 2:300
 - education system characteristics 2:300
 - twin studies 2:300
 - error measurements 2:299
 - general discussion 2:298
 - heterogeneous returns 2:301
 - hourly wages versus hours of work 2:299
 - signaling/sheepskin effects 2:302
 - human capital theory 2:282

- returns to education (*continued*)
 - illustrative estimates 2:302, 2:303*t*
 - information and communication technology (ICT) 8:81–87
 - economic growth
 - knowledge–acquisition model 8:82
 - knowledge–creation model 8:83
 - knowledge–deepening model 8:82
 - research perspectives 8:81
 - sources 8:82
 - economic rationale 8:81
 - Finland 8:84
 - policy implications 8:85
 - Singapore 8:81, 8:83
 - instrumental variables estimation
 - basic concepts 7:226
 - birth statistics 7:226
 - compulsory school laws 7:227
 - distance to school 7:227
 - military service or benefits 7:227
 - school system changes 7:227
 - international comparisons 2:302
 - racial and ethnic groups 1:542–546
 - corporate environments 1:544
 - earnings differentials 1:543
 - economic well-being 1:545
 - income gaps 1:543
 - research background 1:542
 - social mobility 1:542
 - stratification 1:542
 - school finance 2:222–228
 - educational benefits 2:222, 2:223*t*
 - general discussion 2:222
 - neighborhood characteristics 2:431, 2:432*f*
 - resource distribution
 - adjustment factors 2:225
 - efficiency factors 2:226
 - equity and adequacy litigation 2:225
 - flat grants 2:225
 - foundation programs 2:225
 - general discussion 2:224
 - guaranteed tax base (GTB) programs 2:225
 - resource utilization
 - educational services expansion 2:227
 - general discussion 2:226
 - special education 2:226
 - teacher compensation 2:226
 - revenue sources 2:223
 - vocational education and training (VET) 8:255–261
 - human capital theory 8:255, 8:256*f*
 - international evidence
 - basic skills training 8:257
 - general discussion 8:257
 - post-secondary options 8:258
 - public training programs 8:258
 - wider benefits 8:259
 - work-related vocational education and training (VET) 8:259
 - policy formation 8:255
 - policy implications 8:259
 - skill formation implications 8:255, 8:256*f*
- reversal paradox 7:429–430, 7:431*f*
- Rhode Island 1:654*f*
- rhyming skills 2:759, 2:761*t*
- Richards growth model 7:343
- Ricoeur, Paul 6:65, 6:185, 6:417
- rights of children and parents 6:17–21
 - children's rights 1:290, 2:73, 6:17
 - general discussion 6:17
 - moral considerations 6:20
 - parents' rights 1:290, 6:19
- right-to-work laws 2:513
- rigid thinking 5:275
- risky behaviors 2:358, 6:709
- River City MUVE 8:162, 8:162*f*, 8:163*f*, 8:164*f*, 8:165*f*
- Robeson, Paul 1:860
- robust estimation 7:266
- rogue states 4:514
- role of the state in education *see* economics of education
- role play 5:330, 5:333
 - see also* play
- Romania
 - age-related policies and programs 8:263*t*
 - comparative education societies 1:270
 - early childhood education and care (ECEC) 2:111*t*
 - examination systems 3:473
 - inequality–education relationship 2:273*f*
 - information and communication technology (ICT)
 - computer access at home 8:25*t*
 - computer access in school 8:23*t*
 - information and communication technology (ICT) access and equity 8:97
 - student achievement test performance 2:247*f*
 - test score–social background relationship 2:273*f*
 - vocational education and training (VET) 8:263*t*
- Ronald W. Reagan National Defense Authorization Act (2005) 3:10
- root mean square error of approximation (RMSEA) index 7:194
- Rorschach Test 6:499–500
- Rorty, Richard 6:185–186
- Rousseau, Jean Jacques 3:316–317, 6:186
- routine problems 5:273
- Royal Society 3:45
- Roy's maximum root 7:327
- Rubin Causal Model 7:66
- rubrics and scoring guides 4:123–139
 - functional role 4:124
 - general discussion 4:139
 - holistic versus analytical rubrics 4:132, 4:132*t*, 4:134*t*
 - rubrics for rubrics
 - characteristics 4:134, 4:135*t*
 - quality analysis
 - book report rubric 4:138*t*
 - clarity criteria 4:138
 - coverage/organization criteria 4:137
 - general discussion 4:136
 - research report rubric 4:137*t*
 - task-specific versus general scoring guides
 - analytical trait rubric for mathematics 4:130*f*
 - analytical trait rubric for oral presentations 4:131*t*
 - comparisons 4:124, 4:133*t*
 - holistic rubric 4:132*t*
 - 6+1–trait writing-assessment rubric 4:129*f*
 - task-specific scoring guide 4:125*f*
- rumination disorder 2:581
- Russell, Bertrand 6:108
- Russia 5:750–757
 - background information 5:750
 - class size 2:492*f*
 - early childhood education and care (ECEC) 2:111*t*
 - education levels 2:279*f*
 - education system
 - challenges 5:756
 - educational finance 5:755
 - goals 5:750
 - governance 5:755
 - monitoring and evaluation programs 5:756
 - research programs 5:756
 - structural and operational characteristics
 - characteristics 5:750
 - higher education 5:754
 - nonuniversity institutes 5:753
 - postgraduate professional programs 5:754
 - preschool education 5:751
 - primary education 5:751
 - schematic diagram 5:752*f*
 - secondary education 5:751
 - staff and students 5:753*t*
 - supplementary education 5:754
 - vocational education and training (VET) 5:753
 - teaching profession 5:754
- higher education
 - collective bargaining 4:493
 - enrollment rates 4:521*t*
 - institutional expenditures 1:733*t*
 - literacy rates 4:521*t*
 - public expenditures 4:521*t*
- inequality–education relationship
 - immigrant status 2:275*f*
 - test score–social background relationship 2:273*f*
- information and communication technology (ICT)
 - computer access at home 8:25*t*
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - mathematics and science instruction 8:32*t*
 - student proficiency and confidence 8:31*t*
- Rwanda 5:758–769
 - background information 5:758

- education system
 - general discussion 5:758
 - goals and policies
 - challenges 5:759
 - educational goals 5:759
 - general discussion 5:758
 - guiding principles 5:760
 - mission statement 5:758
 - objectives 5:759
 - international compliance 5:760
 - monitoring and evaluation programs 5:766, 5:767*t*, 5:769*f*
 - new policies and programs
 - cross-cutting issues 5:765
 - educational finance 5:763, 5:763*t*
 - girls' education 5:765
 - nine-year basic education program 5:763
 - open, distance, and e-learning (Ode-L) education 5:764
 - peace and reconciliation 5:765
 - post-basic education and training 5:764
 - private sector services 5:765
 - science and technological education 5:764
 - special needs students 5:765
 - structural and operational characteristics
 - adult education 5:762
 - higher education 5:762, 5:762*t*
 - nonformal education 5:762
 - preschool education 5:760
 - primary education 5:760, 5:761*t*
 - secondary education 5:760, 5:761*t*
 - technical and professional education 5:761
 - vocational training 5:761
 - teaching profession 5:766
 - student achievement test performance 2:418*t*
 - Rylean theory 6:116
- S**
- Sacramento City Unified School District Board of Education v. Rachel H.* (1994) 2:717
 - Saint Kitts and Nevis
 - early childhood education and care (ECEC) 2:111*t*
 - higher education enrollment 1:702*f*, 1:703*f*
 - Saint Lucia 2:111*t*
 - Saint Vincent and the Grenadines
 - early childhood education and care (ECEC) 2:111*t*
 - higher education enrollment 1:702*f*, 1:703*f*
 - Salamanca Statement (1994) 4:846, 6:159
 - salaries 2:473, 2:481, 2:495*f*, 2:496, 2:496*f*
 - salience 3:66
 - Sámi culture 8:289*t*
 - Samoa 2:111*t*
 - sampling methods 3:142–146, 7:390–395
 - basic concepts 7:390
 - estimation process 3:143
 - population 3:142
 - sample units 3:142
 - sampling frame 3:143
 - selection process 3:143
 - cluster sampling 3:156
 - computerized adaptive testing application 7:395
 - convenience sampling 3:145
 - design considerations 7:391
 - educational applications 7:6
 - general discussion 3:146
 - Horvitz–Thompson (HT) estimator 7:393
 - nonprobability sampling 3:145, 3:155
 - probability sampling
 - characteristics 3:155
 - cluster sampling 3:144, 7:392
 - design considerations 7:391
 - double sampling 3:145
 - general discussion 3:143
 - multistage sampling 3:145, 7:392
 - proportionate sampling 3:144
 - simple random sampling (SRS) 3:143, 7:391
 - stratified random sampling 3:144
 - stratified sampling 7:392
 - systematic sampling 3:144, 7:393
 - purpose 3:142, 7:390
 - purposive sampling 3:145, 6:523
 - quota sampling 3:145, 3:157
 - random sampling 3:156
 - sample size 3:146, 3:157, 7:393
 - simple random sampling (SRS) 3:143, 3:156, 7:391, 7:437
 - snowball sampling 3:157
 - statistical power 3:146, 7:436
 - stratified sampling 3:156
 - survey research methods 3:155
 - systematic sampling 3:156
 - variance estimation 7:394
 - Sao Tomé and Príncipe
 - early childhood education and care (ECEC) 2:111*t*
 - student achievement test performance 2:418*t*
 - Sartre, Jean Paul 6:2–3, 6:82, 6:111
 - SAT examinations
 - background information 4:7
 - general discussion 1:845
 - large-scale assessments 8:51
 - multivariate linear regression analysis 7:324
 - primary admissions tests 4:8
 - probability theory 7:4
 - standard deviation 7:269
 - satisfaction measures 4:854, 6:105
 - Saudi Arabia
 - early childhood education and care (ECEC) 2:111*t*
 - information and communication technology (ICT) 8:25*t*
 - pharmacy education 3:59
 - Trends in International Mathematics and Science Study (TIMSS) 5:645*t*
 - Saussure, Ferdinand de 1:500, 6:475
 - scaffolding concept 3:367, 5:507
 - scheduled castes 1:590, 1:590*t*
 - scheduled tribes 1:590, 1:590*t*
 - Scheffé test 7:34, 7:313–314, 7:315*t*
 - Scheler, Max 6:451
 - schema-based instruction 2:769
 - Schleiermacher, Friedrich 6:63
 - Schleiermacher, G. W. F. 6:414
 - Schlichter Model for Talents Unlimited Inc. (TU) 2:831
 - Scholar–Baller[®] program 1:860–865
 - background information 1:860
 - basic concepts 1:860
 - Coach Carter* film screening 1:865
 - critical intervention programs 1:861
 - implementation case study 1:862
 - implications 1:864
 - logos 1:863
 - ThinkMan[™] 1:860, 1:861*f*, 1:863
 - ThinkWoman[™] 1:860, 1:861*f*, 1:863
 - Schön, Donald
 - action research 1:357
 - reflective practices theory 7:629–630
 - School and Staffing Survey (SASS) 2:178, 2:181*t*
 - school-based accountability programs 2:374–379
 - contemporary assessments 5:892
 - data sharing ability 8:128
 - design approaches 2:375
 - educational evaluation 3:520
 - general discussion 2:374
 - No Child Left Behind Act (2001) 2:374, 3:520
 - rationale 2:374
 - student achievement impacts 2:376
 - unintended consequences 2:377
 - school-based evaluation 3:713–718
 - advantages/disadvantages 3:721
 - background information 3:713
 - balance and reciprocity 3:716
 - critical friendships 3:714
 - external evaluations 3:714
 - modeling approaches
 - cooperative systems 3:716
 - parallel systems 3:715
 - sequential systems 3:715–716
 - supporting systems 3:715
 - motivating factors 3:713
 - self-inspection versus self-evaluation 3:717, 3:717*f*
 - school-based health clinics 3:555–562
 - future directions 3:560
 - general discussion 3:555
 - historical background 3:555
 - research evaluations
 - clinic effectiveness/costs 3:559
 - lessons learned 3:560
 - mental health access/services 3:557
 - reproductive health programs 3:558

- school-based health clinics (*continued*)
 - utilization 3:556
- school-based vocational education and training 8:247–254
 - access and equity patterns 8:252
 - characteristics
 - curriculum 8:248
 - higher education articulation 8:249
 - initial vocational preparation system 8:250, 8:251*r*
 - level and focus 8:249
 - work experience 8:249
 - focus and scope 8:247
 - institutional arrangements 8:247
 - outcomes analyses 8:251, 8:252*f*, 8:253*f*
- school choice 6:44–49
 - basic concepts 6:46
 - common schools versus separate schools 6:44
 - developing countries 6:285
 - general discussion 6:44
 - home schooling 6:214
 - parents' rights 6:19
 - values analyses 6:47
- school–community relations 5:106–112
 - effective leaders 4:751
 - leadership for learning 4:751
 - new localism
 - background information 5:106
 - community capacity 5:107
 - community involvement 5:109
 - education action zones (EAZs) 5:107–108
 - general discussion 5:110
 - parental involvement 5:109
 - place-based learning 5:106
 - political alliances 5:109
 - social capital 5:108
 - standards movement 5:106
 - technology-supported learning environments 8:125–132
 - alternative schooling approaches
 - accountability data sharing 8:128
 - current trends 8:130
 - distance education 8:129
 - general discussion 8:128
 - one-to-one computing 8:128
 - virtual schools 8:129
 - challenges 8:125
 - importance 8:125
 - research areas 8:130
 - research background 8:126
 - technology initiatives
 - e-mail communication 8:126
 - growth factors 8:128
 - Internet-based approaches 8:126
 - websites 8:127
- school crime supplement (SCS) 2:180, 2:181*r*
- School District of Philadelphia/School of the Future project 8:77
- school effectiveness 6:269–273
 - ability testing 4:5
 - contemporary research developments 6:272
 - future research areas 6:273
 - historical background 6:269
 - influencing factors 6:271
 - International Congress for School Effectiveness and Improvement (ICSEI) 4:683, 6:269
 - methodological characteristics 6:270
 - research findings 6:270
 - school effectiveness research (SER) program 5:51–52, 5:54, 6:269
 - school improvement impacts 6:272
- school efficiency/inefficiency 2:210–215
 - econometric approaches
 - analytical techniques 2:210
 - corrected ordinary least-squares analysis (C-OLS) 2:211
 - data envelopment analysis (DEA) 2:210
 - stochastic frontier analysis (SFA) 2:211
 - general discussion 2:214
 - inefficiency determinants 2:214
 - measurement approaches
 - environmental factors 2:213
 - school inputs 2:212
 - school outputs 2:212
 - standardized tests 2:212
 - research background 2:210
 - research findings 2:213
 - specification estimation strategies 2:212
- school finance 2:222–228
 - educational benefits 2:222, 2:223*r*
 - general discussion 2:222
 - neighborhood characteristics 2:431, 2:432*f*
 - resource distribution
 - adjustment factors 2:225
 - efficiency factors 2:226
 - equity and adequacy litigation 2:225
 - flat grants 2:225
 - foundation programs 2:225
 - general discussion 2:224
 - guaranteed tax base (GTB) programs 2:225
 - resource utilization
 - educational services expansion 2:227
 - general discussion 2:226
 - special education 2:226
 - teacher compensation 2:226
 - revenue sources 2:223
- school finance reform 2:216–221
 - equalization efforts 2:216
 - reform impacts
 - community composition 2:218
 - general discussion 2:218, 2:220
 - housing prices 2:219
 - private monetary contributions 2:219
 - private school enrollments 2:219
 - public services 2:218
 - spending levels and distribution 2:216
 - student performance 2:217
- school health education 1:547–553
 - characteristics 1:547
 - coordinated school health programs 1:549, 1:550*f*
 - education content areas 1:548
 - general discussion 1:552
 - National Health Education Standards (NHES) 1:548
 - program frameworks 1:548
 - tobacco-use prevention
 - adolescent smoking consequences 1:549
 - educational interventions 1:549
 - tobacco industry involvement 1:551
- school–higher education institution (HEI) partnerships 7:757–763
 - characteristics 7:757
 - collaborative action research 7:759
 - general discussion 7:762
 - Innovative Links Project 7:758
 - National Schools Network 7:758
 - networked learning communities (NLC) 5:94, 7:759
 - partnership models
 - accountability 7:761
 - change processes 7:762
 - cooperation and collaboration 7:761
 - features and consequences 7:760*r*
 - general discussion 7:760
 - improvement strategies 7:762
 - leadership considerations 7:761
 - sustainability 7:761
 - preservice education programs 7:758
 - professional development schools 7:758
 - school-level development 7:758
 - supervision 7:619
- school improvement
 - collaborative networks 5:92, 6:280
 - developing countries 6:281–288
 - capacity-building improvements 6:281
 - challenges 6:286
 - comprehensive school reform models 6:286
 - curriculum and instructional practices 6:281
 - decentralization trends 6:284
 - general discussion 6:281
 - public versus private schools 6:285
 - school-based management approaches 6:284
 - school choice 6:285
 - support systems 6:282
 - district-level reform efforts 6:295–302
 - Boston Public Schools (BPS) 6:299
 - Cross City Campaign for Urban School Reform 6:296–297
 - Duval County, Florida 6:297–298
 - general discussion 6:295
 - key components 6:295–296
 - Knowsley Local Education Authority, England 6:300, 6:301*r*
 - Literacy Collaborative (LC) model 6:298
 - San Diego City School District study 6:296

- state's role 6:301
- stuck versus moving districts 6:295
- educational leadership 5:135–142
 - challenges 5:138
 - collaborative networks 5:92
 - conceptualizations 5:135
 - leadership conceptualizations 5:136
 - research areas 5:138
 - research perspectives
 - cultural perspectives 5:137
 - general discussion 5:136
 - political perspectives 5:137
 - rational–technical model 5:136
- inspections/external reviews 3:719–725
 - advantages/disadvantages 3:723
 - England
 - historical background 3:720
 - research results 3:722
 - general discussion 3:719
 - Netherlands 3:723
 - purposes 3:720
 - self-evaluations 3:714, 3:721
- school effectiveness research (SER) 6:272
- whole-school designs 6:303–309
 - middle schooling 6:236, 6:236f, 6:237t
 - social change impacts 6:303
 - United States
 - educational implications 6:307
 - education system characteristics 6:303
 - implementation fidelity/reliability 6:306
 - large-scale assessments 6:303
 - reform efforts 6:303
 - research findings 6:305
 - research implications 6:308
 - sustainability 6:307
- school inspections 3:719–725
 - advantages/disadvantages 3:723
 - educational effectiveness 5:53
 - England
 - historical background 3:720
 - research results 3:722
 - general discussion 3:719
 - national assessment programs 4:90
 - Netherlands 3:723
 - purposes 3:720
 - self-evaluation 3:713–718
 - background information 3:713
 - balance and reciprocity 3:716
 - critical friendships 3:714
 - external evaluations 3:714
 - modeling approaches
 - cooperative systems 3:716
 - parallel systems 3:715
 - sequential systems 3:715–716
 - supporting systems 3:715
 - motivating factors 3:713
 - self-inspection versus self-evaluation 3:717, 3:717f
 - self-evaluations 3:721
- school leadership 4:852–856
 - basic concepts 4:741
 - contextual perspectives 4:704–709
 - basic concepts 4:704
 - context levels
 - macrolevels 4:706
 - mesolevels 4:705
 - microlevels 4:705
 - general discussion 4:704, 4:708
 - global context
 - demographic globalization 4:708
 - economic globalization 4:707
 - general discussion 4:706
 - technological globalization 4:707
- contingent leadership 5:45
- disadvantaged schools 5:58–65
 - challenges
 - deficit views 5:58
 - policies and programs 5:59
 - reform impacts 5:60
 - teacher turnover 5:60
 - leadership impacts 5:61
 - urban environments 4:809–816
- dissatisfaction measures 4:855
- distributed leadership 5:1–6
 - basic concepts 4:741, 4:818
 - comparison studies 5:4
 - disadvantaged schools 4:778
 - effectiveness 5:4
 - general discussion 5:1, 5:5
 - heroic versus distributive leadership model 4:744
 - leader plus aspect model
 - conceptual perspective 5:2
 - empirical perspective 5:2
 - practice aspect model
 - conceptual perspective 5:3
 - empirical perspective 5:3
 - professional learning communities 5:155
 - school-to-school collaboration 6:279, 6:279f
 - teacher leadership 5:40
 - theoretical perspectives 4:764
- diverse cultures 4:824–831
 - future research areas 4:829
 - influencing factors 4:824
 - key themes
 - change agents 4:825
 - cultural influences 4:828
 - developing countries 4:826
 - general discussion 4:825
 - leader development programs 4:827
 - successful leadership 4:826
 - research background 4:824
- diversity considerations 4:832–838
 - demographic data 4:832
 - general discussion 4:832, 4:836
 - inclusion and equity strategies 4:834
 - leadership approach characteristics 4:833
 - school community participation 4:834
 - student experiences 4:833
- early childhood education and care (ECEC) 2:89
- educational effectiveness 5:55
- expectations
 - characteristics 4:852
 - commitment 4:854
 - determinants 4:853
 - general discussion 4:852
 - improvement strategies 4:853
 - performance standards 4:853
- general discussion 4:855
- historical background 4:724–730
 - administrative theories 4:724
 - civil rights movement 4:726
 - efficiency improvements 4:724
 - female leadership 4:725
 - postpositivist perspectives 4:727
 - professional managers 4:725
 - scientific management approach 4:724
 - Theory Movement 4:725
 - twentieth-century perspectives 4:727
 - twenty-first century perspectives 4:728
 - United Kingdom 4:726
- information and communication technology (ICT) 5:120–127
 - educational impacts
 - community relations 5:121, 5:121f
 - general discussion 5:120
 - instructional strategies 5:121, 5:121f
 - literacy instruction 5:121, 5:121f
 - social networks 5:120, 5:121f
 - influencing factors 5:125, 5:125f
 - integration framework 5:123f
 - requirements and standards 5:121
 - social and legal challenges 5:124
 - teacher leadership 5:122, 5:123f
 - technology-mediated leadership development programs 5:123, 5:123t
- instructional leadership 5:18–25
 - basic concepts 5:19, 5:20f
 - empirical analyses
 - general discussion 5:20
 - research findings 5:21
 - trends 4:742, 5:21
 - historical background 5:18
 - international trends 4:759
 - research trends 4:742
 - shared leadership models 5:22
 - student learning impacts 5:45
- leadership development programs 4:752–761

- school leadership (*continued*)
- diverse cultures 4:827
 - general discussion 4:752, 4:759
 - importance 4:752
 - international approaches 4:752, 4:753*t*
 - international trends
 - andragogic principles 4:757
 - centralized and decentralized programs 4:753, 4:754*t*
 - communication and cooperation 4:756
 - course-based learning 4:758*t*
 - experience-oriented/application-oriented methods 4:757, 4:758*t*
 - explicit goals and objectives 4:758
 - instructional leadership 4:759
 - integral leadership 4:759
 - integrative leadership 4:759
 - knowledge production processes 4:757
 - leadership for improvement 4:757
 - modular and multiphase designs 4:756, 4:756*t*
 - moral leadership 4:759
 - partnership models 4:754
 - personal development 4:756
 - posttransformational leadership 4:759
 - preparatory qualifications 4:755, 4:755*t*
 - problem-based learning methods 4:758
 - program duration 4:755, 4:755*t*
 - quality assurance 4:753
 - school's core purpose 4:759
 - team leadership capacity 4:757
 - theory–practice relationships 4:754
 - transformational leadership 4:759
 - workplace learning 4:758
 - technology-mediated leadership development programs 5:123, 5:123*t*
 - leadership for learning 4:746–751
 - assessment programs 4:748
 - belief and value systems 4:751
 - characteristics 4:747*t*
 - curricular programs 4:748
 - instructional program 4:747
 - learning communities 4:748
 - organizational culture 4:750
 - performance feedback 4:747
 - resource allocation and use 4:749
 - reward systems 4:750
 - school–community relations 4:751
 - school-to-school collaboration 6:279, 6:279*f*
 - social advocacy 4:750
 - vision and mission 4:746
 - leadership standards 4:768–776
 - definitions
 - general discussion 4:770
 - standards as measures 4:770
 - standards as principles 4:770
 - development guidelines
 - content standards 4:772
 - definitions 4:770
 - guiding conception of leadership 4:771
 - performance assessment measures 4:773
 - performance standards implementation 4:774
 - principal components 4:771
 - well-written characteristics 4:772
 - importance 4:768, 4:774
 - international developments 4:768
 - professional certification 4:769
 - professional learning 4:769
 - managerial leadership 4:759, 5:45
 - moral leadership 4:759, 5:45
 - National College for School Leadership (NCSL) 3:539, 4:762, 5:93, 6:243
 - organizational change
 - challenges 5:131
 - complex systems theory 5:130
 - educational innovation 5:131
 - Fullan's concepts 5:132, 5:133*t*
 - future directions 5:133
 - historical perspectives 5:128
 - sociopolitical perspectives 5:131
 - participative leadership 5:45
 - philosophical perspectives 4:710–716
 - empiricism 4:710
 - epistemology 4:714
 - general discussion 4:710
 - ontology 4:712
 - theoretical perspectives
 - general discussion 4:711
 - scientific approach 4:710
 - values 4:714
 - principal's role 5:67
 - productive schools 4:746–751
 - assessment programs 4:748
 - belief and value systems 4:751
 - characteristics 4:747*t*
 - curricular programs 4:748
 - instructional program 4:747
 - learning communities 4:748
 - organizational culture 4:750
 - performance feedback 4:747
 - resource allocation and use 4:749
 - reward systems 4:750
 - school–community relations 4:751
 - social advocacy 4:750
 - vision and mission 4:746
 - professional research journals 4:731–739
 - article themes 4:735, 4:736*t*, 4:737*t*
 - background information 4:738
 - Cases in Educational Leadership* (CEL) 4:732
 - editorial processes 4:732, 4:733*t*
 - Educational Administration Quarterly* (EAQ)
 - article submission data 4:733*t*
 - article themes 4:735, 4:736*t*, 4:737*t*
 - internationality 4:734
 - publication data 4:731, 4:732*t*
 - Educational Management Administration & Leadership* (EMAL)
 - article submission data 4:733*t*
 - article themes 4:735, 4:736*t*, 4:737*t*
 - internationality 4:734
 - publication data 4:731, 4:732*t*
 - electronic journals 4:732
 - internationality 4:734
 - International Journal of Educational Management* (IJEM) 4:734
 - Journal of Educational Administration* (JEA)
 - article submission data 4:733*t*
 - article themes 4:735, 4:736*t*, 4:737*t*
 - internationality 4:734
 - publication data 4:731, 4:732*t*
 - Journal of Leadership in Education* (JLE) 4:734
 - publication data 4:731
 - refereed articles 4:733
 - research methodologies 4:733
 - selected journals 4:732*t*
 - submission data 4:733*t*
 - research results analyses 4:766
 - research trends 4:740–745
 - background information 4:740
 - backward mapping 4:742, 4:743*f*
 - empirical analyses 5:21
 - general discussion 4:744
 - heroic versus distributive leadership model 4:744
 - impacts on students 4:741, 5:21
 - pedagogical content knowledge 4:742, 4:743*f*
 - professional learning communities 4:743
 - transformational–instructional leadership relationship 4:742
 - satisfaction measures 4:854
 - school improvement 5:135–142
 - challenges 5:138
 - collaborative networks 5:92
 - conceptualizations 5:135
 - leadership conceptualizations 5:136
 - research areas 5:138
 - research perspectives
 - cultural perspectives 5:137
 - general discussion 5:136
 - political perspectives 5:137
 - rational–technical model 5:136
 - school policies and practices 3:466*t*, 3:467, 3:469*f*
 - school restructuring role 5:66–71
 - changing roles
 - cultural role 5:69
 - environmental role 5:69
 - instructional role 5:69
 - political role 5:68
 - conflicts and tensions
 - accountability 5:67
 - organizational restructuring 5:68

- professionalism 5:67
- structural coupling expectations 5:68
- distributed leadership 5:67
- general discussion 5:66
- implications 5:70
- professional learning communities 5:66
- reculturing processes 5:66
- results analysis 5:69
- rule changes 5:69
- social justice 5:67
- stakeholder relationships 5:69
- social capital 5:85–91
 - background information 5:85
 - basic concepts 5:87
 - impact on educational success 5:88
 - impact on leadership roles and responsibilities 5:89
 - as an intellectual construct 5:86
 - social trust–child well-being relationship 5:87*t*
- special education 2:691
- strategic leadership 5:45
- student learning impacts 5:45–50
 - achievement assessments 5:47
 - general framework 5:45, 5:46*f*
 - leadership models 5:45
 - limitations 5:46
 - mediating variables 5:46*f*, 5:47
 - moderating variables 5:48
 - research trends 4:741, 5:21
 - unit of analysis challenge 5:48
- student voice 4:803–808
 - basic concepts
 - co-leadership role 4:805
 - consulting role 4:804
 - general discussion 4:804
 - leadership roles 4:805
 - benefits 4:806
 - challenges
 - anxiety 4:806
 - credibility 4:807
 - general discussion 4:806
 - inclusivity 4:806
 - innovation issues 4:807
 - contemporary opportunities 4:803
 - general discussion 4:807
 - historical perspectives 4:803
- sustainable development reform 6:292
- system leadership 5:78–84
 - activities and capabilities 5:80, 5:81*f*, 5:82*f*
 - basic concepts 5:78
 - challenges 5:83
 - development opportunities 5:81
 - functional roles 5:79
 - general discussion 5:78
 - school-to-school collaboration 6:279, 6:279*f*
- teacher leadership 5:40–44
 - basic concepts 5:40
 - distributed leadership 5:40
 - educational impacts
 - school level 5:41
 - student level 5:42
 - teacher level 5:42
 - information and communication technology (ICT) 5:122, 5:123*f*
 - leadership for learning 4:818
 - organizational development considerations 5:40
 - professional development considerations 5:43
- theoretical perspectives
 - context matters 4:763
 - core practices 4:764
 - distributed leadership 4:764
 - learning-centered leadership 4:763
 - polyphonic work theory 4:765
 - research background 4:762
 - succession planning 4:764
 - sustainable leadership 4:765
- transformational leadership 5:158–164
 - characteristics
 - general discussion 5:159
 - instructional program management 5:161
 - organization design 5:160
 - people development practices 5:159
 - setting directions 5:159
 - comparison studies 5:4
 - criticisms 5:162
 - historical background 5:158
 - influencing effects 5:158
 - instructional program management
 - buffering functions 5:162
 - general discussion 5:161
 - instructional support 5:162
 - monitoring operations 5:162
 - staffing considerations 5:161
 - international trends 4:759
 - organization design
 - collaborative cultures 5:160
 - community engagement 5:161
 - general discussion 5:160
 - restructuring considerations 5:161
 - school networks 5:161
 - people development practices
 - general discussion 5:159
 - individualized support 5:160
 - intellectual stimulation 5:160
 - leadership models 5:160
 - research trends 4:742
 - school improvement 5:138
 - setting directions
 - general discussion 5:159
 - group goal acceptance 5:159
 - high-performance expectations 5:159
 - shared vision 5:159
 - student learning impacts 5:45
- transformational leadership 5:26–33
 - benefits 5:31
 - criticisms 5:31
 - ethical leadership 5:26
 - evolutionary background 5:29
 - general discussion 5:26
 - modern perspectives 5:30
 - origins 5:26
 - theoretical perspectives 5:27*t*
 - transactional leadership 5:27, 5:27*t*
 - transformational leadership 5:27*t*, 5:28
 - transformational learning theory 5:28
 - transforming leadership 5:27, 5:27*t*
- urban environments 4:809–816
 - challenges 4:810
 - defining characteristics 4:811
 - demographic background 4:809
 - general discussion 4:809
 - leadership strategies 4:815
 - local perspectives
 - estate schooling 4:814, 4:814*f*
 - extended area communities 4:814
 - Leadership on the Front-Line project 4:811
 - multiple communities 4:812, 4:813*f*
 - single/homogenous communities 4:812, 4:812*f*, 4:813*f*
- workforce reform 6:242–248
 - basic principles
 - performance leadership emphasis 6:243
 - qualified teacher status (QTS) 6:242
 - remodelling project 6:242
 - standards 6:242
 - current trends 6:242
 - justification
 - functionalist position 6:244
 - general discussion 6:244
 - socially critical position 6:245
 - outcomes analyses 6:246
- workplace mistreatment/abuse 4:790–796
 - abuse effects 4:793
 - abusive behavior typologies 4:791, 4:792
 - basic concepts 4:790
 - general discussion 4:790
 - modeling approaches 4:791
 - research approaches
 - future directions 4:794
 - methodologies 4:794
 - published results 4:793
 - theoretical perspectives 4:791
- school organization 7:557–562
 - change strategies 7:557
 - comprehensive school reform models 7:559

- school organization (*continued*)
 - general discussion 7:557, 7:561
 - organizational learning characteristics 7:558
- school reform *see* educational reform
- school resources—earnings relationship 2:313–320
 - district-level studies 2:314, 2:315f, 2:317f
 - estimation methods 2:313
 - future research areas 2:319
 - general discussion 2:313
 - methodological weaknesses 2:314
 - research challenges 2:318
 - school-level studies 2:314, 2:315f, 2:318
 - state-level studies 2:314, 2:315f, 2:317f, 2:318
 - statistical significance findings
 - rate of return estimates 2:316, 2:317f
 - United Kingdom 2:315
 - United States 2:314, 2:315f
 - women 2:315
 - variation explanations
 - age dependence 2:318
 - structural changes 2:318
- school survey on crime and safety (SSOCS) 2:180, 2:181t
- school-to-school collaboration 6:274–280
 - benefits 6:274, 6:275t
 - collaborative advantages 6:276
 - collaborative innovation 6:277, 6:278f
 - effectiveness 6:275t
 - England 6:274
 - general discussion 6:274
 - knowledge transfer 6:277
 - leadership
 - capacity-building improvements 6:278
 - lateral accountability 6:279
 - modeling approaches 6:279, 6:279f
 - partnership ethos 6:279
 - Leading Edge Partnership Program (LEPP) 6:274, 6:275
 - school improvements 6:280
- school-to-work transitions 8:306–311
 - challenges 8:237
 - key indicators 8:306
 - policy responses and reforms
 - apprenticeships 8:309
 - education retention rate improvements 8:307
 - general discussion 8:307
 - human capital theory 8:307
 - labor market policies 8:309
 - young adults 8:310
- Schoolwide Enrichment Triad Model (SEM) 2:831
- Schulkindergarten* 5:585, 5:585f
- science
 - curriculum design 1:468–474
 - design-based research
 - curriculum development 1:468
 - curriculum use 1:350
 - design guidelines 1:472
 - design patterns approach 1:469, 1:470t, 1:471f
 - design principles approach 1:469, 1:471f
 - thermodynamics curriculum unit 1:470, 1:471f
 - future research areas 1:472
 - general discussion 1:468
- science education
 - CompEd study
 - characteristics 3:687t
 - Functional Information Technology Test (FITT) 4:645
 - importance 4:649
 - purpose 4:644
 - Stage 1 survey results 4:644
 - Stage 2 survey results 4:645
- computer-assisted instruction 8:197–203
 - access and equity promotion 8:202
 - challenges 8:202
 - classroom network technologies 8:136
 - constructivist perspectives 8:197, 8:198t
 - contextualization
 - authenticity 8:201
 - everyday experiences 8:201
 - future directions 8:202
 - information access 8:201
 - learning environment transformation 8:201
 - physical environment interactions
 - model construction 8:199
 - scientific phenomena 8:198, 8:198t
 - simulations/dynamic representations 8:199
 - student investigations 8:199
 - reflective construction of knowledge
 - program management 8:200
 - teacher support and professional development 8:200
 - working memory (WM) capacity 8:200
 - social environment interactions 8:199
 - subject-specific uses 8:4
 - systemic impacts 8:202
- constructivist learning theories
 - basic concepts 5:485
 - characteristics 7:750t
 - children's ideas 5:488
 - core ideas 5:486
 - general discussion 5:485
 - intellectual development stages 5:487
 - Piagetian perspective 1:488–489, 5:486
 - Vygotskian perspective 1:489, 5:488
- educational program evaluation 3:660–666
 - accountability considerations 3:660
 - background characteristics 3:664
 - contextual evaluations 3:661
 - data analysis and reporting 3:664
 - formative evaluation tools 3:663t
 - outcomes evaluations 3:663
 - process model 3:660, 3:661f
 - process skills 3:662t
 - program implementation 3:662, 3:663t
 - variable measures 3:664
- female participation 1:687
- gender equity 1:630, 1:686
- inquiry skills 8:167–171
 - background information 8:167
 - basic principles 8:168
 - hypothesis generation 8:168, 8:169
 - inquiry cycles 8:169–170
 - primary schools 3:293
 - research areas 8:170
 - scaffolding processes 8:169
 - scientific literacy 8:167
 - skill acquisition 8:169
 - technology-enhanced learning (TEL) 8:168
- International Association for the Evaluation of Educational Achievement (IEA)
 - mathematics and science studies 4:650–655
 - background information 4:650
 - early studies 4:650, 4:664
 - First International Mathematics Study (FIMS) 4:650, 4:664
 - First International Science Study (FISS) 4:651
 - general discussion 4:654
 - Second International Mathematics Study (SIMS) 4:651
 - Second International Science Study (SISS) 4:651
 - survey results and analysis 4:653
 - Third International Mathematics and Science Study 1995 (TIMSS) 4:652
 - Third International Mathematics and Science Study 2003 (TIMSS) 4:653
 - Third International Mathematics and Science Study 2007 (TIMSS) 4:653
 - Third International Mathematics and Science Study Repeat (TIMSS-R) 4:652
- new scientist training 4:322–327
 - doctoral (PhD) training
 - characteristics 4:324
 - diversified knowledge production modes 4:325
 - organizational structures 4:325
 - postdoctoral training 4:326
 - professionalism 4:326
 - program duration 4:324
 - research arrangements 4:325
 - specialization—breadth of knowledge balance 4:325
 - role of science 4:322
 - scientific disciplines 4:322
 - scientist's role 4:323
 - training programs
 - career paths 4:326
 - doctoral (PhD) training 4:324
 - early education and training 4:323
 - general discussion 4:323
 - postdoctoral training 4:326
- primary school assessments 3:293–299
 - formative assessments
 - evidence collection 3:294
 - evidence interpretation 3:295, 3:295t

- feedback 3:296
- goals 3:294
- peer assessments 3:296
- policies and practices 3:294
- student self-assessments 3:296
- future directions 3:298
- group performance reports 3:297
- individual performance reports 3:296
- inquiry-based learning 3:293
- purpose 3:293
- summative assessments
 - combination assessments 3:297
 - evidence summaries 3:297
 - policies and practices 3:296
 - special tasks or tests 3:296
- race/racism **1:554–564**
 - contemporary science education
 - communication norms 1:559
 - educational impacts 1:559
 - funding and training considerations 1:560
 - craniometry 1:557
 - eugenics 1:558
 - historical perspectives 1:555
 - medical research abuse 1:557
 - monogenism/polygenism 1:558
 - physical anthropology 1:557
 - psychometric theories 1:558
 - racial evolution theories 1:555
 - scientific racism 1:554
 - segregation policies 2:775
 - sociohistorical constructs 1:562
 - theoretical retrofitting 1:558
- Rwanda 5:764
- secondary school assessments **3:300–310**
 - concept maps 3:306, 3:308^f
 - conceptual inventories 3:305, 3:306^f, 3:307^f
 - formative versus summative assessments 3:300
 - general discussion 3:309
 - information technologies 3:304, 3:305^f
 - item clusters 3:305
 - performance assessments 3:301, 3:303^f
 - predict–observe–explain (POE) format 3:301, 3:302^f
 - question–demonstration assessments 3:301, 3:302^f
 - science notebooks 3:309, 3:309^f
- Second Information Technology in Education Studies (SITES)
 - basic concepts **4:646**
 - educational impacts **4:667**
 - importance **4:649**
 - Lithuania 5:665
 - Module 1 (SITES M1) survey results **4:646**
 - Module 2 (SITES M2) survey results **4:647**
 - SITES 2006 **4:648**
- women of color 1:688
- science learning **5:387–393**
 - acquisition learning perspective
 - concept learning
 - change processes 5:389
 - conceptual change 5:389
 - implications 5:390
 - Piagetian perspective 5:387
 - social constructivism 5:390
 - sociocultural perspective 5:390
 - Vygotskian perspective 5:390
 - background information 5:387
 - participation learning perspective
 - implications 5:392
 - situated cognition 5:391
 - technology supports **8:197–203**
 - access and equity promotion 8:202
 - challenges 8:202
 - constructivist perspectives 8:197, 8:198^t
 - contextualization
 - authenticity 8:201
 - everyday experiences 8:201
 - future directions 8:202
 - information access 8:201
 - learning environment transformation 8:201
 - physical environment interactions
 - model construction 8:199
 - scientific phenomena 8:198, 8:198^t
 - simulations/dynamic representations 8:199
 - student investigations 8:199
 - reflective construction of knowledge
 - program management 8:200
 - teacher support and professional development 8:200
 - working memory (WM) capacity 8:200
 - social environment interactions 8:199
 - systemic impacts 8:202
- science notebooks 3:309, 3:309^f
- science teacher education **7:712–718**
 - curricula design 7:714
 - instructional strategies 7:715
 - preparation assessments 7:716
 - professional development 7:716
- prospective teachers
 - pedagogical content knowledge
 - assessment methods 7:714
 - curricula 7:714
 - instructional strategies 7:713
 - student learning 7:713
 - philosophical orientations 7:713
 - prior knowledge analyses 7:712
 - subject-matter knowledge 7:712
- research background 7:712
- score inflation **4:186–192**
 - evidence **4:186, 4:187^f, 4:187^t**
 - historical background **4:187**
 - inadequate tradition approaches **4:187**
 - validity of inferences under high-stakes conditions (VIHS)
 - behavioral responses **4:189**
 - conceptual elements **4:190, 4:190^f**
 - effective test weight **4:189**
 - evaluation frameworks **4:188**
 - future directions **4:191**
 - general discussion **4:188**
 - target of inference **4:189, 4:189^f**
- scoring guides *see* rubrics and scoring guides
- Scotland **5:770–775**
 - background information 5:770
 - Catholic schools 2:356
 - class size **6:200**
 - community education 1:120
 - early childhood education and care (ECEC) **2:105**
 - education system 7:526
 - administrative and supervisory structure 5:773
 - challenges 5:774
 - educational finance 5:773
 - goals 5:771
 - monitoring and evaluation programs 5:774
 - research programs 5:774
 - structural and operational characteristics
 - colleges 5:772
 - community learning and development (CLD) programs 5:772–773
 - higher education institutions 5:772
 - modern apprenticeship (MA) programs 5:772
 - preschool education 5:771
 - primary schools 5:771
 - Scottish Qualifications Authority (SQA) 5:773
 - secondary schools 5:771
 - teaching profession 5:773
 - inequality–education relationship 2:273^f
 - information and communication technology (ICT) **8:25^t**
 - physical education and sports 1:459
 - teaching hours 2:493^f
 - test score–social background relationship 2:273^f
 - vocational education and training (VET) **8:272, 8:434, 8:483**
 - see also* United Kingdom
- Scottish Qualifications Authority (SQA) 5:773
- screening tests 3:288
- secondary attention-deficit hyperactivity disorder (SADHD) **2:698**
- secondary schools
 - arts teacher education **7:695–701**
 - Australia 5:511^f, 5:512
 - Austria 5:519
 - Bangladesh 5:527, 5:527^t
 - Bulgaria 5:543
 - Caribbean countries **1:679–685**
 - Barbados 1:681
 - historical background 1:679
 - Jamaica 1:684
 - Trinidad and Tobago 1:682
 - developing countries **6:249–254**

- secondary schools (*continued*)
 Education for All (EFA) 6:250
 education initiatives
 access and expansion 6:252
 delivery challenges 6:253
 marketization 6:253
 enrollment 2:415, 2:416*t*
 future directions 6:253
 Eastern Europe 3:476
 England 5:551
 Estonia 5:556
 Fiji 5:565
 Finland 5:570*f*, 5:571
 France
 demographic data 5:578
 first cycle 5:578
 lycées d'enseignement général/nlet technologique (LEGT) 5:578
 lycées professionnels (LP) 5:578
 second cycle 5:578
 vocational education and training (VET) 5:578
 Gambia 5:860
 Germany 5:585*f*, 5:586, 5:586*t*, 5:587, 5:588*t*
 Ghana 5:592*t*, 5:593
 Greece 5:600
 Hong Kong
 characteristics 3:440
 school-based assessments 3:441
 washback studies 3:440
 Iceland 5:606, 5:607*f*, 5:607*t*
 India 1:593*t*
 Indonesia 5:612*f*, 5:613
 instructional accommodations 2:731
 Internet-based education
 background information 8:102
 international programs 8:104
 North America 8:103
 Israel 5:620, 5:621*t*
 Jamaica 5:627, 5:628*f*
 Latvia 5:631
 Lebanon 5:639
 Lithuania 5:660
 Malawi 5:668
 Malta 5:676*f*, 5:677
 Mongolia 5:682, 5:682*t*
 Namibia 5:687
 New Zealand 3:446, 5:694
 Niger 5:702, 5:703*t*, 5:706*t*
 Norway 5:710, 5:711*f*, 5:712*t*
 Pakistan 5:731
 Philippines 5:867
 Poland 5:739
 Qatar 5:746
 Russia 5:751
 Rwanda 5:760, 5:761*t*
 Scholar-Baller[®] program 1:863
 school-based vocational education and training 8:249
 science education assessments 3:300–310
 concept maps 3:306, 3:308*f*
 conceptual inventories 3:305, 3:306*f*, 3:307*f*
 formative versus summative assessments 3:300
 general discussion 3:309
 information technologies 3:304, 3:305*f*
 item clusters 3:305
 performance assessments 3:301, 3:303*f*
 predict–observe–explain (POE) format 3:301, 3:302*f*
 question–demonstration assessments 3:301, 3:302*f*
 science notebooks 3:309, 3:309*f*
 Scotland 5:771
 self-regulation intervention programs 6:728
 Singapore 5:778
 Slovenia 5:783, 5:784*f*
 social studies curriculum 1:411, 1:412*t*
 Solomon Islands 5:793, 5:793*t*, 5:795*t*, 5:797*t*
 South Africa 5:801, 5:801*t*
 Spain 5:808, 5:809*t*
 Sri Lanka 5:816
 Suriname 5:828
 Taiwan 5:835, 5:836*f*, 5:837*t*
 Tanzania 5:843
 teacher education programs 7:526
 Thailand 5:850*f*, 5:852
 United Kingdom 1:674
 United States 5:889, 5:890*f*
 Uruguay
 repeating and dropout rates 5:876, 5:879*f*
 schooling rates 5:876
 structural and operational characteristics 5:875*f*
 Zanzibar 5:897, 5:898*t*
 Second Information Technology in Education Studies (SITES)
 basic concepts 4:646
 characteristics 3:687*t*
 educational impacts 4:667
 importance 4:649
 Lithuania 5:665
 Module 1 (SITES M1) survey results 4:646
 Module 2 (SITES M2) survey results 4:647, 8:224
 SITES 2006 4:648, 8:22
 Second International Mathematics Study (SIMS) 3:687*t*, 4:651
 Second International Science Study (SISS) 3:687*t*, 4:651
 second-language learning 5:419–425
 ability models 5:382, 5:383*t*
 assessment strategies 4:140–147
 construct research 4:142
 ethical issues
 general discussion 4:143
 professionalism 4:143
 use consequences 4:143
 use justifications 4:143
 future challenges
 classroom assessment 4:144
 foreign language proficiency testing 4:144
 general discussion 4:144
 historical perspectives
 language ability as multicomponential, communicative competence 4:141
 language ability as performance 4:141
 language ability as pragmatic expectancy grammar 4:140
 skills and elements 4:140
 purpose 4:140
 classroom discourse 5:365
 computer-assisted instruction 8:204–210
 background information 8:204
 benefits
 anywhere/anytime instruction 8:208
 general discussion 8:207
 immediate feedback 8:208
 individualized study plans 8:207
 mistakes in private concept 8:208
 patient tutoring 8:208
 records of achievement 8:209
 classroom instruction 8:205
 English-language training demand 8:204, 8:205*f*
 functional role 8:204
 research results and analysis 8:207
 subject-specific uses 8:4
 first language environments 1:441–446
 background information 1:441
 basic concepts
 oral competency 1:441
 proficiency orientation 1:441
 challenges 1:445
 Common European Framework (CEF) 1:442
 content-based instruction 1:443
 general discussion 1:445
 immersion programs 1:443
 language resource centers 1:444
 standards 1:442
 teacher preparation 1:444
 immigrant populations 4:631
 research challenges 5:382
 theoretical perspectives
 cognitive approach 5:419
 contrastive linguistics 5:419
 critical-age hypothesis 5:422
 foreign language learning 5:384, 5:422
 interlanguage 5:421
 motivation 5:420
 school learning impacts 5:423
 social theory approach 5:420
 variation analysis 5:421
 Second Life 8:202
 segregation 2:432
 seizures 2:635, 8:113

- selective attention 5:166, 5:167f
- selective mutism 2:582
- self-adaptive testing 4:148–152
- assisted self-adaptive testing (ASAT) 4:149–150
 - basic concepts 4:148, 4:149f
 - benefits 4:148
 - dynamic assessments 3:222–223
 - general discussion 4:151
 - item pool 4:149
 - item-selection strategies 4:150, 4:150f
 - perceived control/perceived performance hypotheses 4:148
 - restricted self-adaptive testing (RSAT) 4:149
 - theoretical investigations 4:150
- self-advocacy 2:861, 2:878
- self-assessments 3:231–235
- affect measures 3:259
 - assessment conferences 3:406
 - benefits 3:231
 - characteristics 3:231
 - classroom assessment 3:411
 - cognitions and decision-making processes 7:512
 - definition 3:713
 - disadvantaged urban youth 1:812
 - formative assessments 3:362
 - general discussion 3:234
 - learning processes 3:232
 - learning regulation impacts 3:351
 - lifelong learning 3:233
 - primary school science 3:296
 - reading skills 3:194, 3:194f
 - teacher education 3:567, 7:512
 - technological tools 7:619
 - writing assessments 3:276
- self-awareness
- adult learning outcomes 1:230
 - curriculum-based assessment (CBA) 2:574
- self-concept 6:660–667
- academic self-concept (ASC)—achievement correlations
 - multidimensional aspects 6:661
 - reciprocal effects model (REM)
 - age generalizability 6:664
 - Baumeister–Marsh and Craven debate 6:664
 - causality considerations 6:662, 6:663f
 - cross-cultural generalizability 6:663
 - physical domain generalizability 6:664
 - positive-effects debate 6:664
 - research results 6:665 - multidimensional aspects
 - academic self-concept (ASC)—achievement correlations 6:661
 - definitions 6:660
 - hierarchical models 6:660, 6:661f
- self-construal concept 5:453
- Self Description Questionnaires (SDQs) 6:660–661
- self-determination 2:857–864
- assessment strategies 2:862–863
 - basic concepts 2:857, 6:618, 6:619f
 - behavioral consequences 6:621
 - curricular materials 2:862
 - general discussion 2:857, 2:863, 6:622
 - indigenous peoples 1:663
 - influential factors
 - adaptive behaviors 2:858
 - environmental factors 2:858
 - intelligence level 2:858–859
 - intra-individual factors 2:858
 - quality of life 2:675–676, 2:859
 - instructional models 2:862
 - internalization processes 6:620
 - motivation 6:578, 6:627
 - promotion strategies
 - choice making 2:860
 - decision making 2:860
 - general discussion 2:860
 - goal setting 2:860
 - problem solving 2:860
 - self-advocacy 2:861
 - self-awareness 2:861
 - self-knowledge 2:861
 - self-regulation 2:861
 - reward systems 6:621
 - socioenvironmental contexts 6:620
- special needs students
- adolescence to adulthood transition 2:878
 - general discussion 2:857
 - importance
 - general education access 2:859
 - positive adult outcomes 2:675, 2:859
 - promotion strategies 2:860
 - choice making 2:860
 - decision making 2:860
 - general discussion 2:860
 - goal setting 2:860
 - problem solving 2:860
 - self-advocacy 2:861
 - self-awareness 2:861
 - self-knowledge 2:861
 - self-regulation 2:861
- Self-Determined Learning Model of Instruction (SDLMI) 2:862
- self-development theory 6:89
- self-directed learning
- adult learning 1:13
 - biographic perspectives 1:26
 - computer-assisted instruction 8:186
 - curriculum use 1:350
 - workplace learning 1:84
- self-disclosure 6:568
- self-efficacy beliefs 6:668–672
- academic achievement outcomes 6:671
 - African Americans 1:810
 - basic concepts 6:668
 - classroom motivation 6:581, 6:625
 - conservation of resources (COR) model
 - basic concepts 3:499, 3:500f
 - educational applications 3:499
 - future directions 3:503
 - general discussion 3:498
 - resource categories 3:499
 - resource loss and gain spirals 3:501, 3:501f
 - developmental considerations 6:670
 - disadvantaged urban youth 1:810
 - importance 6:672
 - influencing effects 6:670
 - physically disabled students 2:643
 - self-assessments 3:233
 - social and emotional outcomes 6:678
 - social cognitive theory 6:668
 - sources 6:669
- self-evaluation
- balance and reciprocity 3:716
 - critical friendships 3:714
 - definition 3:713
 - external evaluations 3:714
 - modeling approaches
 - adult education and training (AET) 3:552
 - cooperative systems 3:716
 - parallel systems 3:715
 - sequential systems 3:715–716
 - supporting systems 3:715
 - motivating factors 3:713
- school-based evaluation 3:713–718
- advantages/disadvantages 3:721
 - background information 3:713
 - balance and reciprocity 3:716
 - critical friendships 3:714
 - external evaluations 3:714
 - modeling approaches
 - cooperative systems 3:716
 - parallel systems 3:715
 - sequential systems 3:715–716
 - supporting systems 3:715
 - motivating factors 3:713
 - self-inspection versus self-evaluation 3:717, 3:717f
 - self-inspection versus self-evaluation 3:717, 3:717f
 - teacher education 3:567
 - teacher evaluations 3:675
- self-knowledge 2:574
- self-managing schools 5:72–77
- challenges
 - corruption considerations 5:75
 - decentralized decision-making 5:74
 - leadership workload 5:75
 - learning impacts 5:74

- self-managing schools (*continued*)
 - needs-based funding 5:75
 - definitions 5:72
 - driving forces 5:73
 - future directions 5:76
 - global trends 5:72
 - New Zealand 5:72–73, 5:697
 - self-motivation
 - curriculum-based assessment (CBA) 2:574
 - music learning 5:409–410
 - self-reflection
 - diversity considerations 7:522
 - teacher education 3:567
 - transformative learning theory 1:22
 - self-regulated learning 5:503–508
 - basic concepts 5:503
 - coregulation 5:506
 - curriculum use 1:350
 - life-span development theories 5:226
 - modeling approaches
 - Boekaerts' model of adaptable learning 5:504
 - commonalities
 - agency 5:505
 - goals 5:505
 - memory and reasoning 5:505
 - monitoring processes 5:505
 - task environment 5:505
 - research areas 5:507
 - Winne and Hadwin's model of recursive self-regulation 5:504
 - Zimmerman's social-cognitive model 5:504
 - motivational concepts 6:615
 - motivational regulation 6:631–635
 - basic concepts 6:631
 - dimensional characteristics 6:633
 - general discussion 6:634
 - influencing factors 6:633
 - sociocultural factors 6:633
 - student learning impacts 6:631
 - theoretical perspectives 6:632
 - volitional theories 6:632
 - promotion strategies 2:861
 - research areas
 - feedback 5:507
 - modeling approaches 5:507
 - scaffolding 5:507
 - sociocognitive learning processes 5:506
 - sociocognitive perspectives 5:505
 - sociocultural influences 5:506
 - technology-supported learning environments 6:719
 - terminology and jargon sources 8:350*t*
- Self-Regulated Learning (SRL) 4:461, 4:462
- self-regulated strategy development (SRSD) instructional model 2:764
- self-regulation
- coping strategies 6:570–575
 - acute versus chronic stressors 6:570
 - classroom interventions 6:574
 - coping-volition relationship
 - dual processing self-regulation model 6:572, 6:573*f*
 - emotional arousal regulation 6:573
 - volitional control 6:573
 - effectiveness
 - general discussion 6:570
 - new stressor situations 6:571
 - situation assessments 6:571
 - strategy selection 6:571
 - modeling approaches 6:574
 - effective learning 4:461, 4:462
 - emotional feelings 6:547
 - help seeking 6:653–659
 - characteristics 6:653
 - computer-assisted processes 6:657
 - contextual influences 6:654
 - cultural influences 6:654
 - encouragement strategies
 - peer influences 6:656
 - teacher influences 6:655
 - importance 6:657
 - student achievement goals 6:654
 - mathematics instruction 2:770
 - modeling approaches 3:348
 - parent-child relationships 5:451
 - promotion strategies 2:861
 - self-assessments 3:233
 - volitional control 6:724–731
 - background information 6:724
 - characteristics 6:724
 - coping-volition relationship
 - control mechanisms 6:573
 - dual processing self-regulation model 6:572, 6:573*f*
 - emotional arousal regulation 6:573
 - effortful control
 - academic work habits 6:726
 - emotional regulation 6:725
 - learning task management strategies 6:726
 - research background 6:725
 - future research areas 6:728
 - intervention programs
 - general discussion 6:727
 - middle schools 6:727
 - primary schools 6:727
 - secondary schools 6:728
 - undergraduate settings 6:728
- self-reliance 2:675
- self-review processes 3:713, 4:384
- self-study 7:689–694
 - background information 7:689
 - basic concepts 7:690
 - general discussion 7:693
 - inquiry-oriented teacher education 7:607
 - pedagogical concepts 7:588
 - research focal points 7:692
 - research methodologies 7:692
 - significance 7:689
 - teacher educators 7:677
- semantic memory 5:198, 5:198*f*, 5:342
- semiosis 6:476
- semiotics 6:475–480
 - characteristics 6:479
 - Peircean semiotics 5:222
 - poststructuralist theory 6:478
 - Prague Circle of linguists 6:477
 - sign classification systems 6:475
 - structure theory 6:476
 - terminology and meaning 6:476
 - Vygotskian perspective 5:499, 5:501
- semiparametric regression analysis 7:344
- Sen, Amartya 6:194
- Sénégal
 - curriculum structure 3:450*r*
 - early childhood education and care (ECEC) 2:111*t*
 - student achievement test performance 2:418*t*
- Senge, Peter 3:43
- seniors' learning 1:175–182
 - characteristics
 - age 1:176
 - demographic data 1:176
 - learning and intelligence 1:178
 - preretirement education 1:177
 - retirement 1:176
 - stereotypes 1:177
 - educational institutions
 - accreditation 1:180
 - characteristics 1:178
 - general discussion 1:180
 - human capital theory 1:175
 - lifelong learning 1:175
- separate schools
 - common schools versus separate schools 6:44
 - school choice 6:46
 - values analyses 6:47
- separation anxiety disorder 2:582
- sequential articulation 4:612
- sequential bilinguals 5:355–356
- sequential mastery test (SMT) 7:401–406
 - Bayesian sequential decision theory
 - appropriate action determinations 7:404, 7:405*t*
 - derivation calculations 7:404
 - fixed-length mastery problem 7:403
 - framework 7:401
 - notation 7:402
 - optimal rules 7:403
 - posterior predictive (Bayesian) calculations 7:404

- simulation models
 - comparison strategies 7:405
 - general discussion 7:404
 - item response generation 7:405
 - research results 7:405, 7:406*t*
 - test types 7:405
- threshold loss
 - basic concepts 7:402
 - costs of testing 7:402–403
 - measurement model 7:403
 - parameters 7:402*t*
- general discussion 7:406
- limitations 7:401
- research background 7:401
- sequential Monte Carlo (SMC) 7:302
- sequential probability ratio test (SPRT) 7:396–400
 - background information 7:396
 - basic concepts 7:396
 - computerized adaptive testing (CAT) 7:397, 7:398*t*
 - general discussion 7:10, 7:400
 - stochastic curtailment 7:399
 - three-category classification problem 7:399, 7:399*f*, 7:399*t*
- Serbia
 - examination systems 3:473
 - higher education 4:517
 - information and communication technology (ICT)
 - computer access at home 8:25*t*
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - student proficiency and confidence 8:31*t*
- serious emotional disturbances (SEDs) 2:748
- serious game systems 8:178
- Service University 4:342
- Sesame Street Effect 8:99–100
- SET (science, engineering, and technology) 1:687
- severe/multiple disabilities 2:855
- sexual harassment *see* workplace mistreatment/abuse
- Seychelles
 - early childhood education and care (ECEC) 2:111*t*
 - student achievement test performance 2:418*t*
- Sfard, Anna 5:387
- Shanghai Jiao Tong University Institute of Higher Education (SJTIHE) rankings 4:204, 4:548, 4:548*t*, 4:550*t*
- Shapiro–Wilk test 7:337, 7:348
- shared task alignment 6:685
- sharing attention 6:586
- sheepskin effects 2:302, 2:323
- Sheff v O'Neill* 2:394
- short-term memory 5:198–199, 5:341
- shura* (public deliberation) 6:70
- SIBTEST analysis 4:38*t*, 4:39
- sickle cell anemia 2:637
- Sierra Leone 2:418*t*
- signage 6:475–480
 - characteristics 6:479
 - hermeneutic theory 6:414
 - post-structuralist theory 6:478
 - Prague Circle of linguists 6:477
 - sign classification systems 6:475
 - structure theory 6:476
 - terminology and meaning 6:476
- signal detection theory (SDT) 7:407–410
 - basic concepts
 - model 7:407, 7:408*f*
 - response types 7:407, 7:408*t*
 - general discussion 7:10, 7:407
 - hits, misses, false alarms, and correction rejection 7:407, 7:408*t*
 - practical applications 7:409, 7:409*t*
- signaling model 2:321–324
 - basic concepts
 - cost–benefit analysis 2:322*f*
 - lifetime benefits 2:322*f*
 - years of schooling 2:321
 - empirical analyses
 - comparison studies 2:323
 - dropout rates 2:323
 - earnings differentials 2:322
 - enrollment rates 2:323
 - sheepskin effects 2:302, 2:323
 - general discussion 2:324
 - see also* human capital theory
- Significant Multicultural Community Resources (SMCR) 2:551–552
- signification 6:476
- sign language 2:545, 2:624, 2:630
- signposting 8:270
- SimCalc 8:178, 8:179*f*, 8:199
- similarity–dissimilarity analyses 7:304, 7:305*f*, 7:305*t*
- Simon, Theodore 2:870
- simple random sampling (SRS) 3:143, 3:156, 7:391, 7:437
- simple summary measures 7:3
- Simpson's paradox 7:60, 7:61*t*, 7:199, 7:199*f*, 7:210, 7:262, 7:429
- simulated annealing (SA) algorithm 7:248
- simultaneous bilinguals 5:355
- simultaneous item bias test (SIBTEST) 4:167
- Singapore 5:776–781
 - background information 5:776
 - economic growth 8:83
 - education system
 - administrative and supervisory structure 5:779
 - challenges 5:780
 - educational finance 5:780
 - goals 5:776
 - information and communication technology (ICT) 8:66, 8:67*f*
 - monitoring and evaluation programs 5:780
 - quality assurance 5:780
 - research programs 5:780
 - structural and operational characteristics
 - adult continuing education 5:779
 - characteristics 5:777, 8:68
 - Institute of Technical Education 5:778, 8:69
 - polytechnical institutions 5:778, 8:69
 - preschool education 5:777
 - preuniversity education 5:778
 - primary education 5:777
 - schematic diagram 5:777*f*
 - secondary education 5:778
 - special education 5:779
 - universities 5:778
 - vocational education and training (VET) 8:69
 - teaching profession 5:779
 - Thinking Schools, Learning Nation (TSLN) program 5:781
- human capital development 8:66, 8:71, 8:71*t*
- inequality–education relationship 2:273*f*
- information and communication technology (ICT)
 - computer access at home 8:25*t*
 - economic returns 8:81, 8:83
 - mathematics and science instruction 8:32*t*
- inquiry-oriented teacher education 7:606
- Internet-based education 8:104
- leadership development programs 4:753*t*, 4:754*t*, 4:755*t*, 4:758*t*
- mathematics curricula
 - curriculum development and revision 1:455, 1:455*f*, 1:456*f*
 - curriculum reform 1:457
 - selection guidelines 1:454
- medium of instruction (MOI) 4:205
- music teacher education 7:706
- postcolonial perspectives 1:301
- student achievement test performance 2:247*f*
- teacher education programs 7:527
- technology workforce development 8:66–72
 - historical perspectives 8:66, 8:67*f*
 - human capital development 8:66, 8:71, 8:71*t*
 - information and communication technology (ICT) programs
 - education master plans 8:70, 8:70*t*
 - historical perspectives 8:66
 - IT plans/interactive and digital media 8:69
 - test score–social background relationship 2:273*f*
 - vocational education and training (VET) 8:249
- single-case research designs 7:128
- single imputation procedures 7:284
- situated cognition
 - adult learning 1:15
 - case-based education methods 7:594
 - characteristics 7:750*t*
 - ethnomethodology research 6:396
 - nursing education 3:66
 - science learning 5:391
- situated evaluations 3:638
- situated learning 5:335–339
 - apprenticeships 5:336
 - educational applications 5:337
 - e-learning pedagogies 4:423*t*

- situated learning (*continued*)
 legal education 3:26
 museums 1:152
 nursing education 3:65
 preschool education 2:75
 prior learning assessment (PLA) 1:172
 professional development 7:549, 7:549*t*
 sense of identity 5:336
 theoretical perspectives 1:43, 5:335, 5:445
 twenty-first century skills 8:158–166
 challenges 8:164
 framework development 8:159
 innovative instructional strategies
 basic concepts 8:160
 emerging media 8:161
 historical documents 8:164*f*
 legitimate peripheral participation 8:161
 multi-user virtual environment (MUVE) interfaces 8:161
 River City MUVE 8:162, 8:162*f*, 8:163*f*, 8:164*f*, 8:165*f*
 virtual microscopy 8:164*f*, 8:165*f*
 knowledge-based economies 8:158
 selected new literacies 8:159, 8:160*t*
 vocational-based learning (VBL) 1:104, 8:380
 situational barriers 1:234
 situational interest 6:613
 situational–judgment tests (SJTs) 3:260
 skewness 7:269*f*, 7:271, 7:271*f*
 skill-biased technical change (SBTC) theory 2:280
 skills gap 8:320
 skill shortages 8:320–327
 causes
 labor market information availability 8:322
 supply adjustment slowness 8:322
 wage adjustment slowness 8:321, 8:321*f*, 8:322*f*
 general discussion 8:320, 8:326
 general perspectives
 employers 8:320
 unions 8:321
 meaning differentiations 8:320
 measurement approaches
 general discussion 8:322
 initial issues
 data availability and reliability 8:323
 general discussion 8:323
 spatial dimension 8:323
 time dimension 8:323
 shortage indicators
 employer-based surveys 8:324
 hard-to-fill vacancies 8:323
 in-depth studies 8:324
 miscellaneous labor-market indicators 8:324
 net vacancies 8:323
 structural models 8:324
 unemployment rates 8:323
 vacancy rates 8:323
 wages 8:324
 public policy responses
 appropriate action 8:325
 immigration 8:326
 labor market information availability 8:325
 mobility encouragement 8:326
 publicly funded training 8:325
 training encouragement 8:325
 skill characteristics 8:320
 Skills, Opportunities and Recognition (SOAR) Program 2:43, 2:45*t*
 Skinner, B. F. 5:246
 slave trade 1:574
 sleep deprivation 1:547
 Slovak Republic
 age-related policies and programs 8:263*t*
 class size 2:492*f*
 early childhood education and care (ECEC) 2:4*t*, 2:111*t*
 education levels 2:279*f*
 higher education
 foreign doctoral students 4:292*t*
 loans and grants 4:501*f*, 4:502*f*
 student and researcher mobility 4:287*f*
 tertiary education expenditures 4:500*f*
 tertiary type-B program survival rates 8:464*t*
 tuition fees 4:500*f*
 inequality–education relationship
 employment rate–educational attainment relationship 2:278*f*
 test score–social background relationship 2:273*f*
 information and communication technology (ICT)
 computer access in school 8:23*t*
 frequency of computer use in school and at home 8:26*t*
 mathematics and science instruction 8:32*t*
 student proficiency and confidence 8:31*t*
 self-managing schools 5:72–73
 vocational education and training (VET) 8:251*t*, 8:252*f*, 8:253*f*, 8:263*t*
 Slovenia 5:782–789
 age-related policies and programs 8:263*t*
 background information 5:782
 early childhood education and care (ECEC) 2:111*t*
 education levels 2:279*f*
 education system
 administrative and supervisory structure 5:787
 educational finance 5:787
 examination reforms 3:474
 goals 5:782
 performance assessments 5:787
 policies and programs 5:788
 structural and operational characteristics
 adult education 5:786
 basic concepts 5:784*f*
 compulsory education 5:783, 5:784*f*
 educational providers, students, and teaching staff 5:785*t*
 higher education 5:786
 music education 5:786
 nonformal education 5:786
 preschool education 5:783, 5:784*f*
 special needs students 5:785–786
 vocational education and training (VET) 5:786
 teaching profession 5:787
 higher education 4:517
 inequality–education relationship
 employment rate–educational attainment relationship 2:278*f*
 test score–social background relationship 2:273*f*
 information and communication technology (ICT)
 computer access at home 8:25*t*
 computer access in school 8:23*t*
 frequency of computer use in school and at home 8:26*t*
 mathematics and science instruction 8:32*t*
 student proficiency and confidence 8:31*t*
 vocational education and training (VET) 8:263*t*
 women in employment 2:56*f*
 Slutsky theorem 7:234
 small-area estimation 7:411–416
 educational applications 7:415
 general discussion 7:10
 methodologies
 best linear unbiased prediction (BLUP) estimation 7:414
 composite estimation 7:413
 conditional autoregressive (CAR) model 7:415
 demographic methods 7:411
 empirical Bayes methods 7:414
 hierarchical Bayes methods 7:414
 models with area-specific effects 7:414
 spatial statistics 7:415
 structure-preserving estimation (SPREE) 7:413, 7:413*f*
 synthetic and related methods
 general discussion 7:411
 ratio correlation 7:412
 sample regression method 7:412
 symptomatic regression 7:412
 synthetic estimation 7:412
 variance regression methods 7:413
 sample surveys 7:411, 7:412*f*
 small business training 8:337–342
 basic concepts 8:337
 directly and indirectly relevant influencing factors 8:340
 general discussion 8:341
 importance 8:337
 motivation 8:338
 owner/manager attitudes 8:339
 training program characteristics 8:338
 small-group instruction 2:817
 small pox 1:602
 small-schools movement 5:68
 Smith, Adam
 compensating differentials models 2:465
 economic liberalism 6:75

- educational benefits 2:348–349
 human capital theory 2:194, 2:282
 smoking effects 1:549, 1:604, 1:605*t*, 2:174, 2:426
 snowball sampling 3:157
 social and emotional learning (SEL) 3:260
 social behaviorism 6:88
 social capital 5:113–119, 6:97–102
 appeal 6:98
 basic concepts 5:113, 6:97
 benefits 1:229, 5:115, 6:674, 6:679
 building and utilization processes 5:114, 5:114*f*
 civic engagement 6:677
 community engagement
 development processes 5:116, 5:116*f*
 key indicators 5:117
 modeling approaches 5:115
 educational leadership 5:85–91
 background information 5:85
 basic concepts 5:87
 impact on educational success 5:88
 impact on leadership roles and responsibilities 5:89
 as an intellectual construct 5:86
 key indicators 5:117
 organizational savvy 4:700
 social trust–child well-being relationship 5:87*t*
 educational perspectives 6:99
 external social benefits 2:263*t*, 2:268
 future directions 5:117, 5:117*f*
 general discussion 5:113, 6:97
 historical background 6:97
 knowledge networks and communities 5:102
 Latino students 1:613
 learning cities and regions 1:146
 new localism 5:108
 Pascal virtual community 1:146
 professional learning communities 5:117, 5:117*f*
 school–community relations 5:108
 school curriculum
 challenges 6:100
 formal curriculum 6:99
 general discussion 6:99
 informal curriculum 6:99
 student leadership for learning 4:819
 supportive school policies and practices 3:466*t*, 3:467, 3:469*f*
 theoretical perspectives 1:192
 typology 6:98
 well-being and happiness 2:34
 social class 1:64–69
 characteristics 1:64
 curriculum use 1:350
 diversity discourse 1:823
 educational practices 1:66
 education–health association 1:231
 general discussion 1:68
 narrative socialization 5:305
 student achievement effects 2:423–430
 child-rearing practices
 academic achievement 2:425
 after-school activities 2:424
 literacy gaps 2:423
 role models 2:424–425
 general discussion 2:428
 health influences 2:253, 2:258, 2:425
 housing instability 2:427
 income and wealth differences 2:427
 Milwaukee Project 2:609
 research background 2:423
 theoretical perspectives 1:65
 see also socioeconomic status (SES)
 social cognitive theory
 basic concepts 6:668
 case-based education methods 7:595
 self-regulated learning 5:504
 social constructivism
 adult education and training (AET) 1:224
 curriculum development 1:307, 1:489, 5:390
 early childhood education and care (ECEC) 2:107*t*
 e-learning pedagogies 4:423*t*
 recognition of prior learning (RPL) 8:399–400
 social depravity *see* poverty
 social development 6:585–590
 basic concepts 6:585
 characteristics
 conflict management skills 6:587–588
 friendship 6:587
 joint attention 6:586
 peer relationships 6:587
 social competence 6:589
 social play 6:588
 task behavior 6:589
 temperament 6:588
 general discussion 6:589
 research background 6:586
 social interactions
 emotion-related processes 5:313, 5:314*f*
 gifted and talented students 7:733
 learning and development 6:698–704
 collaborative learning
 experimental research 6:701
 general discussion 6:700
 learning characteristics 6:702
 peer discussions 6:701
 productive interactions 6:702
 relationship quality 6:702
 surveying classroom activity 6:701
 task design 6:702
 effective talk 6:703
 general discussion 6:698, 6:703
 relationship quality 6:703
 teaching process
 dialogic teaching 6:700
 knowledge elicitation and questioning 6:698
 responsiveness 6:699
 shared experiences 6:699
 teacher–student communication 6:698
 vocational education and training (VET) 8:466
 social justice
 principal's role 5:67
 race and ethnicity 1:79
 supervision 7:618
 teacher education 7:598, 7:599*t*
 social moderation 3:459
 social networks 6:705–711
 academic performance 6:709
 adolescents 6:707
 characteristics 6:705
 computer-assisted applications 4:422, 4:443, 5:120, 5:121*f*, 8:6
 computer-mediated communication 8:116
 deviant behaviors 6:709
 general discussion 8:112–113
 knowledge communities 8:190, 8:194
 research approaches 6:706
 young children 6:706
 social reasoning *see* moral and social reasoning
 social relationships
 adult learning outcomes 1:231
 childhood schooling 6:585–590, 6:691–697
 adjustment factors
 antecedent behaviors 6:694
 behavioral precursors 6:695
 interpersonal relationships 6:693
 relational precursors 6:695
 basic concepts 6:585
 challenges 6:691
 characteristics
 conflict management skills 6:587–588
 friendship 6:587
 joint attention 6:586
 peer relationships 6:587
 social competence 6:589
 social play 6:588
 task behavior 6:589
 temperament 6:588
 future directions 6:696
 general discussion 6:589
 interpersonal relationships
 adjustment factors 6:693
 aggressive children 6:692
 classroom friendships 6:693
 general discussion 6:692
 multiple classroom relationships 6:694
 peer acceptance and rejection 6:693

- social relationships (*continued*)
 - peer victimization 6:693
 - prosocial children 6:692
 - teacher–child relationships 6:694
 - withdrawn children 6:692
 - research background 6:586
 - student performance 6:691
- collaborative learning 6:685–690
 - friendship 6:686
 - general discussion 6:688
 - peer relationships 5:445, 6:686
 - status relationships 6:687
- computer-mediated communication 8:116
- help seeking behaviors 6:656
- social networks 6:705–711
 - academic performance 6:709
 - adolescents 6:707
 - characteristics 6:705
 - computer-mediated communication 8:116
 - deviant behaviors 6:709
 - research approaches 6:706
 - young children 6:706
- special needs students 2:794–801
 - chronological age 2:795
 - cultural influences 2:797
 - gender-related differences 2:795
 - general discussion 2:794, 2:798
 - inclusive interventions 2:798
 - program integration effects 2:797
 - research perspectives 2:794
- student athletes 1:860
 - academic achievement outcomes 1:860
 - Coach Carter* film screening 1:865
 - critical intervention programs 1:861
 - incentive and reward systems 1:864
 - popular culture 1:863
 - Scholar–Baller[®] program 1:862
 - status symbols 1:861, 1:863
 - technological impacts 8:116
 - technology-supported learning environments 6:719
- social science research
 - diversity research 1:693
 - nonprobability sampling 3:145
- social studies education
 - assessment strategies 3:323–329
 - Center on Organization and Restructuring Schools (CORS) standards 3:325, 3:326*r*
 - challenges 3:326
 - classroom-based assessments 3:325, 3:325*r*
 - document-based question (DBQ) assessment 3:324
 - general discussion 3:323, 4:657
 - large-scale assessments 3:323
- curriculum 1:411–414
 - challenges 1:411
 - characteristics 1:411, 1:412*r*
 - competing approaches 1:411, 1:412*r*
 - general discussion 1:413
 - NCSS standards 1:413
 - subject-specific disciplines 1:412
 - textbooks 1:413
- educational program evaluation 3:667–673
 - assessment validity and reliability 3:668
 - challenges 3:667
 - classroom implementation
 - assessment practices 3:670
 - content decisions 3:670
 - general discussion 3:670
 - instructional impacts 3:671
 - curriculum 3:667, 3:668*r*
 - general discussion 3:672
 - recommendations 3:672
 - research background 3:671
- social studies teacher education 7:719–724
 - challenges
 - citizenship development 7:720
 - content considerations 7:719–720
 - disciplinary content versus method/pedagogy 7:720
 - general discussion 7:719
 - purpose 7:719
 - research limitations 7:721
 - educational trends
 - authentic pedagogy 7:722
 - culturally relevant pedagogy (CRP) 7:723
 - general discussion 7:722
 - global citizenship education (GCE) 7:722
 - multicultural education 7:723
 - future directions 7:723
 - societal bilingualism 5:355
 - Society of Civil Engineers (SCE) 3:45–46
 - sociocultural theory 5:498–502
 - socioeconomic status (SES) 2:253–259
 - adult learning outcomes 1:230
 - diversity discourse 1:823
 - education–health association
 - causal relationships 1:231, 2:175
 - educational attainment 1:667
 - educational outcomes 2:258
 - external benefits 2:253
 - mortality rates 1:607, 1:668–669, 2:173*f*
 - general discussion 2:253
 - health effects
 - educational attainment 1:667–672
 - alcohol consumption 1:670
 - background information 1:667
 - health-care services 1:669, 2:175
 - international perspectives 1:670
 - lifestyle behaviors 1:547, 1:604, 1:669–670, 2:174
 - measurement approaches 1:667
 - mortality rates 1:668–669, 2:173*f*
 - research areas 1:671
 - research evidence 2:255
 - research results 1:547, 1:668
 - self-rated health 1:668
 - social and psychological components 1:669
 - educational outcomes 2:258
 - external benefits 2:253
 - future outcomes 2:255
 - research evidence
 - acute illnesses 2:257
 - adult health 2:255
 - asthma 2:257
 - birth weight 2:256
 - causal factors 2:254
 - correlation studies 2:254
 - educational attainment correlations 1:667–672, 2:255
 - environmental toxins 2:257
 - future outcomes 2:255
 - mental health 2:257
 - nutrition effects 2:256
 - in utero* conditions 2:256
 - intelligence 5:189
 - Israeli education system 5:620
 - logistic random-intercept model 7:171, 7:172*f*, 7:172*r*
 - mortality rates 1:607
 - narrative socialization 5:305
 - student achievement effects 2:423–430
 - achievement gaps 2:81
 - child-rearing practices
 - academic achievement 2:425
 - after-school activities 2:424
 - literacy gaps 2:423
 - role models 2:424–425
 - early childhood education and care (ECEC) 2:81
 - general discussion 2:428, 5:52
 - health influences 2:253, 2:258, 2:425
 - housing instability 2:427
 - importance 2:436
 - income and wealth differences 2:427
 - Milwaukee Project 2:609
 - neighborhood characteristics 2:432
 - progress maps 3:374, 3:375*f*
 - research background 2:423
 - teacher credentials 2:500
 - traumatic brain injury (TBI) 2:695
 - vocational education and training (VET) 8:429, 8:466
 - see also* social class
 - Socrates 6:108, 6:349
 - Socratic method 2:524, 5:297
 - Solomon four-group design 3:94
 - Solomon Islands 5:790–798
 - background information 5:790
 - early childhood education and care (ECEC) 2:111*r*

- education system
 - administrative and supervisory structure 5:796
 - background information 5:791
 - challenges 5:797
 - educational finance 5:796
 - enrollment 5:797*t*
 - goals 5:791
 - monitoring and evaluation programs 5:796
 - research programs 5:796
 - structural and operational characteristics
 - adult education 5:792*f*, 5:794
 - early childhood education and care (ECEC) 5:792, 5:792*f*, 5:797*t*
 - nonformal education 5:792*f*, 5:794
 - primary education 5:792*f*, 5:793, 5:795*t*, 5:797*t*
 - schematic diagram 5:792*f*
 - secondary education 5:792*f*, 5:793, 5:793*t*, 5:795*t*, 5:797*t*
 - tertiary education 5:792*f*, 5:794
 - vocational education and training (VET) 5:792*f*, 5:794, 5:796*t*
 - teaching profession
 - gender analysis 5:795*t*
 - general discussion 5:795
 - primary school teachers 5:795*t*
 - secondary school teachers 5:795*t*
 - vocational education and training (VET) teachers 5:796*t*
- somatic knowing 1:16
- Sorbonne Declaration 4:566
- Sourcer's Apprentice (computer application) 5:195
- South Africa 5:799–805
 - accreditation processes 4:384–389
 - background information 5:799, 8:442
 - classroom assessment 3:449–456
 - assessment types
 - continuous assessments 3:450, 3:451*t*
 - international assessment programs 3:449
 - national assessment programs 3:449
 - public (external) examinations 3:450
 - challenges 3:454
 - policies and programs
 - current national assessment policies 3:452
 - Curriculum 2005 3:451, 3:451*t*
 - exit points/continuation/exit skills 3:452*t*
 - formal assessments 3:453*t*
 - further education and training (FET) 3:451–452, 3:451*t*, 3:452*t*, 3:453
 - general discussion 3:451
 - general education and training (GET) 3:451–452, 3:451*t*, 3:452*t*, 3:453
 - outcomes-based education (OBE) movement 3:451
 - recording guidelines 3:455*t*
 - revised national curriculum statement (RNCS) 3:452, 3:453
 - community-focused schools 6:257
 - conflict background 8:442
 - curriculum structure 3:450*t*
 - early childhood education and care (ECEC) 2:111*t*
 - education system
 - enrollment 5:800, 5:801*t*
 - general discussion 5:799, 5:804
 - monitoring and evaluation programs 5:803
 - percentage of students by degree level 5:802*t*
 - post-Apartheid reform 5:799, 8:443
 - race and gender data 5:802*t*
 - structural and operational characteristics
 - adult education and training (AET) 5:801, 5:802*t*
 - early childhood education and care (ECEC) 5:800, 5:800*t*
 - further education and training (FET) 3:451–452, 3:451*t*, 3:452*t*, 3:453, 5:801, 8:444
 - higher education 5:802, 5:802*t*
 - primary education 5:801, 5:801*t*
 - secondary education 5:801, 5:801*t*
 - teaching profession 5:803, 5:803*t*
- ethical practices 4:718
- higher education
 - collective bargaining 4:494
 - enrollment rates 4:521*t*
 - gender equity 1:630–631
 - income contingent loans (ICLS) 4:505
 - literacy rates 4:521*t*
 - national technical talent 1:855–859
 - institutional investments 1:858
 - internationalization processes 1:857
 - philanthropic assistance 1:857
 - recruitment and retention 1:855
 - South African Network of Skills Abroad (SANSA) 1:857–858
 - Southern African Norwegian Training, Education and Development (SANTED) initiative 1:855–856
 - study abroad programs 1:855
 - postconflict societies 4:516
 - public expenditures 4:521*t*
 - university associations 4:485
 - HIV/AIDS education 1:819
 - inequality–education relationship 2:273*f*
 - information and communication technology (ICT) 8:32*t*
 - music teacher education 7:706
 - Open University 4:405
 - returns to education–cognitive skills 2:310*t*
 - school choice 6:285
 - socioeconomic/sociopolitical characteristics 1:199
 - student achievement test performance 2:247*f*, 2:417*t*, 2:418*t*
 - test score–social background relationship 2:273*f*
 - vocational education and training (VET) 8:247, 8:397, 8:398, 8:444
 - see also* Africa
- South African Network of Skills Abroad (SANSA) 1:857–858
- South African Qualifications Authority (SAQA) 8:348, 8:443
- South America
 - accreditation processes 4:384–389
 - earnings–education level relationship 2:306*t*
 - enrollment rates
 - gender disparities 2:416*t*
 - primary schools 2:415*t*
 - secondary schools 2:416*t*
 - inequality–education relationship 2:273*f*
 - private higher education (PHE) 4:623
 - returns to education–education level 2:308*t*
 - student achievement test performance 2:247*f*, 2:417*t*
 - test score–social background relationship 2:273*f*
- South Asia *see* Asia
- South Carolina 1:654*f*, 2:65
- South Dakota 1:654*f*
- Southern African Consortium for Monitoring Educational Quality (SACMEQ) 3:449, 4:89, 4:691
- Southern African Development Community (SADC)
 - social purposes
 - background information 1:199
 - economic development 1:201
 - HIV/AIDS education 1:200
 - lifelong learning 1:201
 - political and cultural development 1:200
 - survival considerations 1:200
 - socioeconomic/sociopolitical characteristics 1:199
- Southern African Norwegian Training, Education and Development (SANTED) initiative 1:855–856
- South Korea
 - higher education collective bargaining 4:494
 - lifelong learning model 1:157*t*, 1:159
 - mathematics curricula 1:454
 - private higher education (PHE) 4:622
 - research-quality assessments 4:308
 - see also* Korea
- south-southeast Europe (SEE) 4:517
- Spain 5:806–812
 - age-related policies and programs 8:263*t*, 8:266
 - background information 5:806
 - childhood brain development 2:57*f*
 - class size 2:492*f*
 - displaced worker assistance programs 8:278
 - early childhood education and care (ECEC) 2:4*t*, 2:56*t*, 2:105, 2:111*t*
 - earnings–education level relationship 2:307*t*
 - education levels 2:279*f*
 - education system
 - administrative and supervisory structure 5:811
 - challenges 5:812
 - educational finance 5:811
 - goals 5:807
 - monitoring and evaluation programs 5:811
 - research programs 5:811
 - structural and operational characteristics
 - adult education 5:810
 - basic education 5:808
 - early childhood education and care (ECEC) 5:808, 5:809*t*
 - educational indicators 5:809*t*
 - higher education 5:809*t*, 5:810
 - lower secondary education 5:808
 - primary education 5:808, 5:809*t*
 - schematic diagram 5:808*f*

- Spain (*continued*)
- secondary education 5:809*t*
 - staff and students 5:809*t*
 - upper secondary education 5:809
 - vocational education and training (VET) 5:809, 5:809*t*
 - teaching profession 5:810
 - gender-equity gap 2:56*t*
 - higher education
 - faculty mobility 4:562
 - foreign doctoral students 4:292*t*
 - historical development 1:728
 - institutional expenditures 1:733*t*
 - loans and grants 4:501*f*; 4:502*f*
 - parental expenditures 1:734*t*
 - student and researcher mobility 4:287*f*
 - student mobility 4:561
 - tertiary education expenditures 4:500*f*
 - tertiary type-B program survival rates 8:464*t*
 - tuition fees 4:500*f*
 - inequality–education relationship
 - earnings differentials 2:277*t*
 - employment rate–educational attainment relationship 2:278*f*
 - fathers' educational background 2:274*f*
 - test score–social background relationship 2:273*f*
 - information and communication technology (ICT)
 - computer access in school 8:23*t*
 - frequency of computer use in school and at home 8:26*t*
 - mathematics and science instruction 8:32*t*
 - student proficiency and confidence 8:31*t*
 - participation research
 - regular education 8:284*f*
 - vocational education and training (VET) 8:283*f*; 8:284*f*
 - work-related vocational education and training (VET) 8:284*f*; 8:285*f*
 - research–quality assessments 4:308
 - school supply shortages 2:506*f*
 - special education background 2:714
 - student achievement test performance 2:247*f*
 - teacher education programs 7:528
 - teacher shortages 2:494*f*
 - teacher wages and salaries 2:496*f*
 - teaching hours 2:493*f*
 - vocational education and training (VET) 8:251*t*; 8:252*f*; 8:253*f*; 8:263*t*; 8:266
- spatial learning ability test (SLAT) 7:88, 7:88*f*; 7:89*t*
- spatial thought 5:249
- Spearman–Brown prophecy formula 4:28
- Spearman rank coefficient 7:261
- Spearman's ρ 7:349, 7:447*t*
- special education
 - cost–function analysis 2:208*t*
 - early childhood special education (ECSE) 2:594–607
 - basic concepts 2:595
 - birth to 8 years 2:594
 - demographic data 2:598*t*
 - family involvement 2:601
 - future directions 2:604
 - government legislation 2:596, 2:597*t*
 - intervention recipients 2:597
 - intervention services 2:595
 - itinerant professionals 2:602
 - personnel preparedness 2:601
 - service delivery
 - center-based programs 2:600
 - characteristics 2:598
 - combination programs 2:600
 - home-based programs 2:599
 - service locations 2:599*t*
 - service types 2:599*t*
 - transition processes
 - early intervention to preschool transition 2:602
 - general discussion 2:602
 - preschool to kindergarten transition 2:603- general discussion 2:693
- goals and objectives 2:689
- historical background 2:688–694, 2:714–720
 - contemporary period
 - advocacy groups 2:717
 - civil rights movement 2:688, 2:716
 - court decisions 2:688, 2:716
 - disability rights movement 2:717
 - distinctive and empirically-based pedagogies 2:718
 - general discussion 2:716
 - international actions 2:717
 - least restrictive environments 2:718, 2:721
 - legislative actions 2:717
 - overidentification concerns 2:690, 2:718
 - socially constructed disability categories 2:718
- future research areas 2:718
- general discussion 2:714
- origins
 - blind individuals 2:715
 - deaf individuals 2:714
 - general discussion 2:714
 - philosophical perspectives 2:714
- research perspectives 2:688
- specialized services
 - advocacy groups 2:716
 - general discussion 2:715
 - institutionalization 2:715
 - intelligence tests 2:716
 - interventions, schools, and classes 2:715
 - Skeels' and Dye's research 2:716
- inclusion practices 2:721–726, 6:158–163
 - advantages/disadvantages 2:722
 - definitions 2:721
 - educational impacts
 - family members 2:724
 - general discussion 2:723
 - students with and without disabilities 2:723
 - teachers and service providers 2:724
 - general discussion 2:725
 - historical perspectives 2:721
 - feeling included versus feeling excluded 6:160
 - inclusion versus integration 6:158
 - mainstream provisions versus special provisions 6:160
 - medical versus social model of disability 6:159
 - segregation versus integration 6:158
 - implementation strategies 2:723
 - international perspectives 2:724
 - least restrictive environments 2:718, 2:721
 - philosophical perspectives 6:161
- leadership skills 2:691
- Lithuania 5:661
- school finances 2:226
- special education professionals
 - professional development 2:889
 - school-based services 2:852
 - teacher induction programs 2:889
- Sri Lanka 5:817, 5:817*t*
- Sultanate of Oman 5:722
- teacher preparedness 2:691
- special needs students
 - adapted physical activity programs 2:519–529
 - basic concepts
 - historical background 2:521
 - individualized education plans (IEPs) 2:520
 - theoretical/philosophical perspectives 2:522
 - class organization and management strategies 2:525, 2:526*t*
 - effective physical education pedagogy 2:522
 - general discussion 2:527
 - government policies and programs 2:520
 - importance 2:520
 - instructional strategies
 - active teaching 2:523
 - adventure education 2:523
 - cooperative learning 2:524
 - flexible groups 2:524
 - peer tutoring/reciprocal reaching 2:524
 - self-instructional formats 2:525
 - standards 2:523, 2:524*t*
 - task teaching 2:524
 - teaching through questions/Socratic method 2:524
 - professional preparation and development 2:526
 - adolescence to adulthood transition 2:876–879
 - challenges 2:876
 - general discussion 2:878
 - government policies and programs 2:876
 - planning guidelines
 - collaboration and communication strategies 2:877
 - effectiveness 2:877
 - parental and family involvement 2:877
 - self-determination 2:878
 - assessment strategies

- adequate yearly progress (AYP) targets 3:495
- assessment accommodations 2:530–536
 - background information 2:530
 - future directions 2:535
 - government legislation 2:531
 - individualized education plans (IEPs) 2:534
 - research areas 2:534
 - typologies 2:531, 2:532*t*
 - validity measures 2:532, 2:533*f*
- challenges 3:496
- diagnostic and classification processes 2:577, 3:492
- educational impacts 3:495
- large-scale assessments 2:752–758
 - accountability 2:755
 - assessment accommodations 2:530
 - assessment options 2:753
 - background information 2:752–753
 - general discussion 2:757
 - purpose 2:753
- national assessment accommodations 3:495
- pedagogical assessments 3:495
- reform impacts 3:492
- augmentative–alternative communication (AAC) 2:544–549
 - characteristics
 - general discussion 2:545
 - manual communication 2:545, 2:624
 - technological devices 2:545
 - communication initiation
 - challenges 2:547
 - communication partners 2:548
 - use settings 2:548
 - vocabulary selection 2:548
 - vowel-based communication chart 2:549*f*
 - communication partners 2:547, 2:548
 - goals and objectives 2:544
 - needs assessment 2:546
- Australia 3:422, 5:512
- Austria 5:521
- autism spectrum disorders (ASD) 2:661–668
 - characteristics
 - Asperger syndrome 2:662*f*, 2:663
 - autism 2:662, 2:662*f*
 - childhood disintegrative disorder 2:662*f*
 - general discussion 2:662
 - pervasive developmental disorder—not otherwise specified 2:662*f*, 2:663
 - Rett syndrome 2:662*f*
 - symptoms continuum 2:662*f*
 - historical perspectives 2:661
 - research areas 2:667
 - school-based services 2:854
 - screening and diagnosis
 - etiology 2:664
 - general discussion 2:663
 - prevalence 2:664
 - treatment/intervention approaches 2:664, 2:666*f*
- characteristics 2:834
- cognitive disabilities 2:669–677
 - assistive technology (AT) 2:673, 2:674*t*, 2:675*t*
 - characteristics 2:669, 2:670*t*
 - effective teaching behaviors
 - general discussion 2:670
 - observational learning 2:670–671
 - reinforcer learning 2:671–672, 2:672*t*
 - response–prompt hierarchy 2:671, 2:671*t*, 2:672*f*
 - time–delay learning 2:671, 2:672*t*
 - verbal learning mediation 2:670
 - general discussion 2:676
 - instructional strategies 2:669
 - positive adult outcomes 2:675
 - strategic educational programming 2:672, 2:674*t*
- cooperative learning 2:555–559
 - advantages 2:558
 - basic concepts
 - characteristics 2:556, 2:556*t*
 - face-to-face promotive interaction 2:557
 - group processing 2:557
 - individual accountability 2:557
 - positive goal interdependence 2:556
 - social skills development 2:557
 - research results and analysis 2:557
 - theoretical perspectives 2:555
- culturally and linguistically diverse (CLD) learners
 - community-based experiences 2:550–554
 - best practices strategies 2:552
 - general discussion 2:553
 - legal mandates 2:552
 - research background 2:550
 - transition frameworks 2:550
- cultural competence 2:565–570
 - basic concepts 2:566
 - community involvement 2:568
 - importance 2:569
 - overrepresentation 2:565
 - Pedersen's model 2:566
 - research background 2:565
- cultural difference theory model
 - cultural incongruity model 2:776
 - general discussion 2:776
 - oppositional cultural framework 2:776
 - stereotype threats 2:776
- dropouts 2:846
- gifted and talented students 2:872
- hearing impaired students 2:631
- language development considerations 2:746
- overrepresentation 2:774–779
 - challenges 2:778
 - cultural deficits 2:775
 - cultural difference theory model 2:776
 - deficit thinking effects 2:775
 - ecological/contextual perspectives 2:777
 - historical perspectives 2:565, 2:690, 2:774
 - school failure explanations 2:775
- speech and language impairments
 - African-American English dialect speakers 2:656
 - bilingual students 2:656
 - demographic data 2:656
 - individualized education plans (IEPs) 2:658, 2:658*f*
 - Individuals with Disabilities Education Act (IDEA, 1990) 2:657
- curriculum-based assessment (CBA) 2:571–576
- curriculum-based measurement (CBM) 2:572, 2:771
 - definitions 2:571
 - design considerations 2:573
 - effectiveness evaluations 2:574
 - general curriculum impacts 2:573
 - general discussion 2:571, 2:575
 - historical background 2:572
 - importance 2:571
 - motivational strategies 2:574
 - relevant measurement models 2:573
 - usage guidelines 2:575
- deaf–blind students 2:621–627
 - classroom intervention strategies 2:625
 - communication challenges 2:624
 - definitions 2:621
 - demographic data 2:623*t*
 - developmental challenges 2:622
 - early intervention services 2:622, 2:625
 - etiology 2:624, 2:624*t*
 - general discussion 2:626
 - hearing loss criteria 2:623
 - historical background 2:621
 - legal blindness criteria 2:623
 - transition processes 2:626
- diagnostic and classification processes 2:577–583
 - assessment strategies 2:577, 3:492
 - basic principles 3:493
 - challenges 3:496
 - Diagnostic and Statistical Manual of Mental Disorders* (DSM–IV–TR) 2:577
 - attention deficit/hyperactivity disorders (ADHD) 2:581
 - communication disorders 2:579
 - conduct disorders 2:581
 - diagnostic and classification processes 2:581, 3:494
 - disability categories 2:578
 - elimination disorders 2:582
 - emotional and behavioral disorders (EBD) 2:581
 - feeding and eating disorders 2:581
 - learning disabilities 2:579
 - mental retardation (cognitive disability) 2:578
 - motor skills disorders 2:579
 - pervasive developmental disorders 2:579
 - reactive attachment disorder 2:582
 - selective mutism 2:582

- special needs students (*continued*)
 - separation anxiety disorder 2:582
 - stereotypic movement disorder 2:582
 - tic disorders 2:582
 - educational impacts 3:495
 - general discussion 2:582
 - historical background 3:493
 - national assessment accommodations 3:495
 - pedagogical assessments 3:495
 - purposes 3:492
 - reform impacts 3:492
 - testing theories 2:577
- direct instruction (DI) 2:584–587
 - basic assumptions 2:584
 - benefits 2:586
 - future directions 2:585
 - literacy instruction 2:759, 2:760*t*
 - theoretical/philosophical perspectives 2:584
- early childhood special education (ECSE) 2:594–607
 - basic concepts 2:595
 - birth to 8 years 2:594
 - demographic data 2:598*t*
 - family involvement 2:601
 - future directions 2:604
 - government legislation 2:596, 2:597*t*
 - intervention recipients 2:597
 - intervention services 2:595
 - itinerant professionals 2:602
 - personnel preparedness 2:601
 - service delivery
 - center-based programs 2:600
 - characteristics 2:598
 - combination programs 2:600
 - home-based programs 2:599
 - service locations 2:599*t*
 - service types 2:599*t*
 - transition processes
 - early intervention to preschool transition 2:602
 - general discussion 2:602
 - preschool to kindergarten transition 2:603
- early intervention services 2:608–612
 - benefits 2:608
 - collaborative environments 2:611
 - deaf-blind students 2:622, 2:625
- early childhood special education (ECSE) 2:594–607
 - basic concepts 2:595
 - birth to 8 years 2:594
 - center-based programs 2:600
 - combination programs 2:600
 - demographic data 2:598*t*
 - early intervention to preschool transition 2:602
 - family involvement 2:601
 - future directions 2:604
 - government legislation 2:596, 2:597*t*
 - home-based programs 2:599
 - intervention recipients 2:597
 - intervention services 2:595
 - itinerant professionals 2:602
 - personnel preparedness 2:601
 - preschool to kindergarten transition 2:603
 - service delivery 2:598
 - service locations 2:599*t*
 - service types 2:599*t*
 - transition processes 2:602
 - general discussion 2:608
 - importance 2:608
 - limitations 2:609
 - Milwaukee Project 2:609
 - teacher preparedness 2:610
- Estonia 5:558
- family-school relationships 2:788–793
 - family characteristics 2:788
 - family-school interactions 2:790
 - general discussion 2:792
 - participation impacts 2:789
 - partnership benefits 2:791
 - school environment impacts 2:791
- Fiji 5:565
- France 5:577
- gifted and talented students 2:588–593
 - characteristics
 - attention deficit/hyperactivity disorders (ADHD) 2:590
 - emotional and behavioral disorders (EBD) 2:590
 - general discussion 2:589, 2:592
 - learning disabilities 2:590
 - educational programs 2:591
 - identification strategies 2:588
 - occurrences 2:588
- inclusion
 - see also* No Child Left Behind Act (2001)
- inclusion practices 2:721–726, 6:158–163
 - advantages/disadvantages 2:722
 - definitions 2:721
 - educational impacts
 - family members 2:724
 - general discussion 2:723
 - students with and without disabilities 2:723
 - teachers and service providers 2:724
 - general discussion 2:725
 - historical perspectives
 - feeling included versus feeling excluded 6:160
 - general discussion 2:721
 - inclusion versus integration 6:158
 - mainstream provisions versus special provisions 6:160
 - medical versus social model of disability 6:159
 - segregation versus integration 6:158
 - implementation strategies 2:723
 - inclusive classrooms 2:728
 - international perspectives 2:724
 - least restrictive environments 2:718, 2:721
 - philosophical perspectives 6:161
- individualized education plans (IEPs) 2:834
- Indonesia 5:612*f*, 5:613
- instructional accommodations 2:727–732
 - background information 2:727
 - differentiated instruction 2:728
 - elementary school levels 2:731
 - general discussion 2:732
 - inclusive classrooms 2:728
 - modeling approaches
 - accommodation selection decision tree 2:730*f*
 - general discussion 2:729
 - model types 2:729*t*
 - setting, methods, motivation, and materials (SM3) model 2:729
- secondary school levels 2:731
- setting, methods, motivation, and materials (SM3) model
 - general discussion 2:729
 - materials 2:730
 - methods 2:730
 - motivation 2:730
 - setting 2:729
- interdisciplinary collaboration 2:741
- labeling effects 2:688–689, 2:788
- language development 2:745–751
 - developmental milestones 2:747
 - effective communication skills 2:745, 2:746*t*
 - family involvement 2:749
 - general discussion 2:749
 - language difference versus language disorder 2:747
 - proactive strategies 2:749
 - serious emotional disturbances (SEDs) 2:748
 - specific disability impacts 2:748
- large-scale assessments 2:752–758
 - accountability
 - basic concepts 2:755
 - educator accountability 2:756
 - modeling approaches 2:756
 - student accountability 2:756
 - system accountability 2:755
 - assessment options
 - accommodations/modifications 2:753
 - alternate assessments 2:754
 - general assessments 2:753
 - general discussion 2:753
 - background information 2:752–753
 - general discussion 2:757
 - purpose 2:753
- Latvia 5:632
- literacy instruction 2:759–766
 - importance 2:759
 - instructional strategies
 - cognitive strategy instruction (CSI) 2:759, 2:760*t*

- directed instruction (DI) 2:759, 2:760*t*
- reading skills 2:759
- vocabulary 2:763
- writing skills 2:764
- reading skills
 - beginning reading 2:759, 2:761*t*
 - fluency 2:761
 - instructional strategies 2:759
 - peer-assisted learning strategies (PALS) 2:761–762, 2:762*t*
 - reading comprehension 2:762
 - think before reading, think while reading, think after reading (TWA) strategy 2:762, 2:763*t*
- writing skills
 - instructional strategies 2:764
 - transcription skills 2:765
 - written expression 2:764, 2:764*t*
- Lithuania 5:661
- Malawi 5:670
- Malta 5:676, 5:676*f*
- mathematics instruction 2:767–773
 - background information 2:767
 - general discussion 2:772
 - historical research 2:768
 - instructional strategies
 - cognitive strategy instruction (CSI) 2:770
 - concrete-to-representational-to-abstract (CRA) sequence 2:769
 - Enhanced Anchored Instruction (EAI) 2:771
 - formative assessments 2:771
 - general discussion 2:768
 - graduated guidance sequence 2:769
 - graphic organizers 2:769, 2:769*f*
 - peer-mediated instruction 2:770
 - schema-based instruction 2:769
 - self-regulation 2:770
- Mongolia 5:683
- oral language skills assessments 3:291
- paraprofessionals 2:780–787
 - background information 2:780
 - general discussion 2:786
 - instructional assignments 2:784, 2:785*t*
 - training programs
 - general discussion 2:781
 - skill requirements 2:782
 - United Kingdom 2:783*t*
 - United Kingdom
 - general discussion 2:781, 2:786
 - higher-level teaching assistant (HLTA) standards 2:781, 2:782*t*
 - instructional assignments 2:784, 2:785*t*
 - national occupational standards (NOS) 2:781*t*
 - number of teachers and teaching assistants 2:781*t*
 - training programs 2:782, 2:783*t*
- peer/social relationships 2:794–801
 - chronological age 2:795
 - cultural influences 2:797
 - gender-related differences 2:795
 - general discussion 2:794, 2:798
 - inclusive interventions 2:798
 - program integration effects 2:797
 - research perspectives 2:794
- physically disabled students 2:640–645
 - background information 2:640
 - early intervention services 2:608–612, 2:641
 - benefits 2:608
 - collaborative environments 2:611
 - general discussion 2:608
 - importance 2:608
 - limitations 2:609
 - Milwaukee Project 2:609
 - teacher preparedness 2:610
 - learning potential maximization 2:642
 - related services 2:640
 - school-based services 2:854
- post-secondary education 2:808–814
 - future directions 2:813
 - importance 2:808
 - participation, retention, and persistence considerations
 - reform strategies 2:812
 - support services 2:811
 - preparation and support 2:808
 - reform strategies 2:810
 - transition processes 2:808
- pre-referral intervention strategies 2:815–819
 - basic concepts 2:815
 - effective strategies 2:817
 - general discussion 2:818
 - implications 2:818
- Qatar 5:746
- related services 2:834–839
 - general discussion 2:839
 - physically disabled students 2:640
 - related services providers
 - characteristics 2:834
 - consultative services 2:835
 - direct services 2:835
 - integrated services 2:835
 - service provision recommendations 2:838
 - service types
 - art therapy 2:641
 - audiology services 2:835
 - counseling services 2:836
 - early identification and assessment services 2:836
 - general discussion 2:835
 - medical services 2:640, 2:836
 - occupational therapy 2:640, 2:836
 - orientation and mobility services 2:640, 2:836
 - parent counseling services 2:837
 - physical therapy 2:640, 2:837
 - psychological services 2:837
 - recreational therapy 2:837
 - rehabilitation counseling 2:837
 - school health/school nurse services 2:837
 - social work services 2:838
 - speech/language pathology services 2:640, 2:838
 - transportation services 2:640, 2:838
- risk–resilience frameworks 2:840–844
 - background information 2:840
 - benefits 2:843
 - protective factors
 - family factors 2:842
 - general discussion 2:841
 - individual factors 2:841
 - school and community factors 2:842
- Rwanda 5:765
- school-based services 2:850–856
 - disability categories
 - attention deficit/hyperactivity disorders (ADHD) 2:581, 2:636, 2:854
 - autism 2:854
 - blindness and low vision 2:854
 - communication disorders 2:579, 2:854
 - deafness and hearing loss 2:854
 - developmentally delayed children 2:854
 - diagnostic and classification processes 2:577–583
 - emotional and behavioral disorders (EBD) 2:581, 2:853
 - general discussion 2:853
 - health impairments 2:634–639, 2:854
 - learning disabilities 2:579, 2:853
 - low-incidence disabilities 2:855
 - mental retardation (cognitive disability) 2:578, 2:853
 - physical disabilities 2:579, 2:854
 - severe/multiple disabilities 2:855
 - educational services options 2:850, 2:851*t*
 - general discussion 2:638, 2:855
 - government policies and programs 2:850
 - health impaired students 2:634–639
 - asthma 2:635
 - attention deficit/hyperactivity disorders (ADHD) 2:636
 - cancer 2:637
 - congenital heart defects 2:636
 - general discussion 2:638
 - historical perspectives 2:634
 - HIV/AIDS 2:637
 - seizures and epilepsy 2:635
 - sickle cell anemia 2:637
 - teacher preparedness 2:638
 - historical perspectives 2:634
 - Individuals with Disabilities Education Act (IDEA, 1990) 2:634, 2:851–852
 - special education professionals 2:852
 - teacher preparedness 2:638
 - transition-to-school concepts 2:855
- school finance 2:226
- self-determination
 - adolescence to adulthood transition 2:878

- special needs students (*continued*)
 - general discussion 2:857
 - importance
 - general education access 2:859
 - positive adult outcomes 2:675, 2:859
 - promotion strategies
 - choice making 2:860
 - decision making 2:860
 - general discussion 2:860
 - goal setting 2:860
 - problem solving 2:860
 - self-advocacy 2:861
 - self-awareness 2:861
 - self-knowledge 2:861
 - self-regulation 2:861
- Singapore 5:779
- Slovenia 5:785–786
- specific learning disabilities (SLDs) **2:646–653**
 - background information 2:646
 - characteristics
 - early childhood 2:648
 - elementary grades 2:648
 - definitions 2:646
 - diagnostic assessments
 - achievement tests 2:651
 - clinical interviews 2:650
 - ecological assessments 2:650
 - general discussion 2:650
 - intelligence tests 2:651
 - parent interviews 2:650
 - reports and records reviews 2:651
 - teacher interviews 2:651
 - instructional strategies 2:651
 - prevalence 2:648
 - research areas 2:652
- speech and language impairments **2:654–660**
 - culturally and linguistically diverse (CLD) learners
 - African-American English dialect speakers 2:656
 - bilingual students 2:656
 - demographic data 2:656
 - individualized education plans (IEPs) 2:658, 2:658f
 - Individuals with Disabilities Education Act (IDEA, 1990) 2:657
 - disorder characteristics
 - articulation 2:654
 - childhood apraxia of speech (CAS) 2:654
 - general discussion 2:654
 - language-based learning disabilities (LBLD) 2:655
 - language disorders 2:655
 - orofacial myofunctional disorders (OMD) 2:654
 - phonological disorders 2:654
 - stuttering 2:655
 - voice disorders 2:655
 - future directions
 - evidence-based practices 2:659
 - general discussion 2:658
 - technology-supported learning environments 2:658
 - testing and assessments 2:658
 - general discussion 2:654, 2:659
- Sri Lanka 5:817, 5:817r
- Sultanate of Oman 5:722
- Suriname 5:828
- Taiwan 5:836f, 5:837r, 5:838
- Taiwan 5:852
- transition processes **2:613–620**
 - adolescence to adulthood **2:876–879**
 - challenges 2:876
 - general discussion 2:878
 - government policies and programs 2:876
 - planning guidelines 2:877
 - self-determination 2:878
 - age 3 transition
 - child and family factors 2:616
 - community factors 2:617, 2:617r
 - legal considerations 2:616
 - outcomes analyses 2:617
 - state and local plans 2:617
 - conceptual framework 2:614, 2:615f
 - deaf-blind students 2:626
 - early service transitions 2:615
 - general discussion 2:619
 - implications 2:613–614
 - individualized family service plans (IFSPs) 2:615, 2:850–851
 - influencing factors
 - age 3 transition 2:616
 - kindergarten transition 2:617r, 2:618, 2:618r
 - state and local plans 2:616f, 2:617
 - kindergarten education 2:617r, 2:618, 2:618r
 - traumatic brain injury (TBI) **2:695–700**
 - affected mental functions 2:696r
 - categories
 - coordination of services 2:697
 - general discussion 2:695
 - injury factors 2:695
 - recovery factors 2:698
 - school readiness factors 2:696
 - characteristics 2:695
 - general discussion 2:699
 - school-based services 2:855
- visual impairment **2:880–885**
 - educational needs 2:881
 - etiologies
 - definitions 2:880
 - functional implications 2:881
 - prevalence 2:880
 - general discussion 2:884
 - government policies and programs 2:880
 - instructional needs 2:883
 - learning materials accessibility
 - auditory readers 2:883
 - Braille systems 2:883
 - closed circuit televisions (CCTVs) 2:882–883
 - general discussion 2:882
 - optical devices 2:882–883
 - occurrences 2:880
 - school-based services 2:854
- vocational education and training (VET) **8:294–299**
 - challenges 8:235, 8:294
 - future directions 8:298
 - international approaches
 - funding considerations 8:295
 - government policies and programs 8:294
 - outcomes analyses 8:297
 - participation research 8:296
 - support systems 8:296
- workforce issues **2:886–891**
 - challenges 2:886
 - conditions of service
 - basic concepts 2:889
 - compensation 2:890
 - work context 2:890
 - general discussion 2:890
 - personnel shortage/surplus
 - general discussion 2:886
 - workforce demand 2:887
 - workforce retention 2:887
 - workforce supply 2:886
 - professional knowledge and skills
 - continuing professional development 2:889
 - importance 2:888
 - induction programs 2:889
 - initial preparation 2:888
 - qualified workforce 2:886, 2:887f
- Zanzibar 5:898
- see also* No Child Left Behind Act (2001)
- special program research funding 4:281
- Special Strategies study 6:304
- specific learning disabilities (SLDs) **2:646–653**
 - background information 2:646
 - characteristics
 - early childhood 2:648
 - elementary grades 2:648
 - definitions 2:646
 - diagnostic assessments
 - achievement tests 2:651
 - clinical interviews 2:650
 - ecological assessments 2:650
 - general discussion 2:650
 - intelligence tests 2:651
 - parent interviews 2:650
 - reports and records reviews 2:651
 - teacher interviews 2:651
 - instructional strategies 2:651

- prevalence 2:648
 - research areas 2:652
 - speech and language impairments **2:654–660**
 - allied health professions 3:52*r*
 - augmentative–alternative communication (AAC) **2:544–549**
 - characteristics
 - general discussion 2:545
 - manual communication 2:545, 2:624
 - technological devices 2:545
 - communication initiation
 - challenges 2:547
 - communication partners 2:548
 - use settings 2:548
 - vocabulary selection 2:548
 - vowel-based communication chart 2:549*f*
 - communication partners 2:547, 2:548
 - goals and objectives 2:544
 - needs assessment 2:546
 - culturally and linguistically diverse (CLD) learners
 - African-American English dialect speakers 2:656
 - bilingual students 2:656
 - community-based experiences **2:550–554**
 - best practices strategies 2:552
 - general discussion 2:553
 - legal mandates 2:552
 - research background 2:550
 - transition frameworks 2:550
 - demographic data 2:656
 - individualized education plans (IEPs) 2:658, 2:658*f*
 - Individuals with Disabilities Education Act (IDEA, 1990) 2:657
 - oral language skills assessments 3:291
 - diagnostic and classification processes 2:579
 - disorder characteristics
 - articulation 2:654
 - childhood apraxia of speech (CAS) 2:654
 - general discussion 2:654
 - language-based learning disabilities (LBLD) 2:655
 - language disorders 2:655
 - orofacial myofunctional disorders (OMD) 2:654
 - phonological disorders 2:654
 - stuttering 2:655
 - voice disorders 2:655
 - future directions
 - evidence-based practices 2:659
 - general discussion 2:658
 - technology-supported learning environments 2:658
 - testing and assessments 2:658
 - general discussion 2:654, 2:659
 - oral language skills assessments 3:291
 - school-based services 2:854
- SpeechRater™ 4:24
- Spence, Michael 2:321
- Spinoza, Benedict de 6:414
- spiritual education *see* religious and spiritual education
- spirituality 1:16
- Spivak, Gayatri Chakravorty 6:58
- split-half reliability 4:28, 4:29*r*
- spontaneous concepts 5:501
- sports *see* physical education and sports
- Sri Lanka **5:813–825**
 - background information 5:813
 - demographic data 5:813
 - economic factors 5:814
 - education system
 - administrative and supervisory structure
 - governance 5:818
 - implementation structure 5:820*f*
 - national and provincial schools 5:819
 - national level 5:819, 5:820*f*
 - provincial level 5:819, 5:820*f*
 - educational finance
 - government expenditures 5:818*f*, 5:820, 5:821*r*
 - household expenditures 5:821
 - medium-term budgetary framework (MTBF) 5:821
 - school revenues 5:819*f*, 5:821
 - education sector development framework and program (ESDFP)
 - challenges 5:821, 5:823*r*
 - education quality 5:822, 5:823*r*
 - equitable access 5:822, 5:823*r*
 - governance and service delivery 5:822, 5:823*r*
 - monitoring and evaluation programs 5:822, 5:823*r*
 - general discussion 5:822
 - historical background
 - ancient period 5:814
 - British period 5:814
 - Dutch period 5:814
 - modern era 5:815
 - Portuguese period 5:814
 - postindependence era 5:814
 - pre-independence era 5:814
 - public school system 5:816, 5:816*r*
 - semi- and nongovernment educational institutions
 - international schools 5:817, 5:817*r*
 - Pirivena education 5:817, 5:817*r*
 - private schools 5:817, 5:817*r*
 - special education schools 5:817, 5:817*r*
 - structural and operational characteristics
 - early childhood education and care (ECEC) 5:816
 - general discussion 5:815
 - higher education 5:816
 - nonformal education 5:817
 - preschool education 5:816
 - primary education 5:816
 - public school system 5:816, 5:816*r*
 - schematic diagram 5:815*f*
 - secondary education 5:816
 - semi- and nongovernment educational institutions 5:817, 5:817*r*
 - teacher education 5:818
 - vocational education and training (VET) 5:816
 - tertiary education
 - higher education institutions 5:822
 - postgraduate programs 5:822
 - public university education 5:822
 - technical education 5:822
 - vocational education and training (VET) 5:822
- gender analysis 5:814
- higher education
 - expansion effects 4:523
 - medium of instruction (MOI) 4:205
 - tuition and fees 4:523
- higher education gender equity 1:630–631
- literacy rates 5:813–814
- stability 6:559
- stable unit treatment value assumption (SUTVA) 7:66
- stakeholders
 - basic concepts 3:629
 - educational program evaluation **3:629–635**
 - empirical research results 3:631
 - evaluation activities 3:629
 - evaluation utilization impacts 3:633
 - importance 3:633
 - participation guidelines 3:522, 3:708
 - Program Evaluation Standards*
 - participation guidelines 3:629
 - theoretical perspectives 3:630
 - theoretical perspectives
 - broad involvement 3:631
 - deep involvement 3:631
 - organizational learning perspective 3:630
 - Program Evaluation Standards* 3:630
 - social justice perspective 3:630
- ethical evaluation practices **3:622–628**
 - challenge responses
 - colleague-based feedback 3:627
 - cultural contexts 3:627
 - entry/contracting stage 3:626
 - personal values 3:627
 - professional guidelines and standards 3:626
 - Guiding Principles
 - competency 3:623
 - confidentiality 3:623
 - informed consent 3:623
 - integrity and honesty 3:623
 - respect 3:623
 - responsibility 3:624
 - systematic inquiry 3:622
 - Program Evaluation Standards 3:624
- research areas
 - data analysis and interpretation 3:625
 - data collection 3:625
 - design considerations 3:625
 - entry/contracting stage 3:624

- stakeholders (*continued*)
 - findings utilization 3:626
 - general discussion 3:624
 - results reporting 3:625
 - significant events 3:622
 - evaluation research methodologies 3:602, 3:743
 - International System for Teacher Observation and Feedback (ISTOF) project 3:729, 3:730*r*
- Standardized Testing and Reporting (STAR) Program 4:177–178
- standardized tests
 - admissions tests 1:845–849, 4:7–14
 - background information 4:7
 - changing contexts 1:845
 - fairness 4:12
 - international institutions 4:12
 - noncognitive variables
 - characteristics 1:845, 1:846*t*
 - community involvement 1:846*t*, 1:847
 - long-range goals 1:846*t*, 1:847
 - Noncognitive Questionnaire (NCQ) 1:847, 4:11–12
 - nontraditional knowledge acquisition 1:846*t*, 1:847
 - positive self-concept 1:846, 1:846*t*
 - racial discrimination 1:846, 1:846*t*
 - realistic self-appraisal 1:846, 1:846*t*
 - successful leadership experiences 1:846*t*, 1:847
 - support services 1:846*t*, 1:847
 - Psychometric Entrance Test (PET) 4:13
 - purpose 1:845
 - SAT examinations 1:845
 - SweSAT 4:12–13
 - background information 4:186
 - computer-assisted assessments
 - educational assessments 4:49
 - first-generation computer-based tests 8:48, 8:49*t*
 - Generation R tests 8:49*t*, 8:52
 - next-generation electronic tests 8:49*t*, 8:50
 - technology-rich environments scenarios 8:52, 8:53*f*
 - criticisms 3:220
 - differential item functioning (DIF) 4:12, 4:36–44
 - bias 4:36
 - characteristics 4:37
 - definitions 4:37
 - detection guidelines 4:41
 - purification procedures 4:40
 - research areas 4:41
 - technique classifications
 - background information 4:37
 - item response theory (IRT) 4:38*t*, 4:39
 - logistic regression (LR) analysis 4:38*t*, 4:39
 - Mantel–Haenszel (MH) statistic 4:38
 - nonparametric DIF techniques 4:37
 - parametric DIF techniques 4:39
 - SIBTEST analysis 4:38*t*, 4:39
 - structural equations modeling 4:40
 - technique criteria 4:38*t*
 - digital technology 8:91
 - fairness 4:12
 - institutional effectiveness 3:172–180
 - basic concepts 3:173
 - challenges 3:179
 - data collection
 - educational and school process information 3:175
 - general discussion 3:174
 - student background characteristics 3:174
 - student baselines 3:174
 - student outcomes 3:174
 - measurement approaches
 - data requirements 3:175, 3:175*t*
 - statistical analysis techniques 3:175, 8:92
 - value-added models (VAMs) 3:177, 3:177*f*
 - raw league tables 3:172
 - reform efforts 3:172
 - value-added models (VAMs)
 - applications 3:178
 - basic concepts 3:173
 - research results 3:177, 3:177*f*
 - minority students 2:577
 - norm-referenced measurement 4:95–101
 - challenges 4:101
 - higher education 3:251
 - historical background 4:95
 - normative scales
 - characteristics 4:95
 - developmental scales 4:96
 - status scales 4:96
 - norm types
 - age- and grade-based norms 4:97
 - characteristics 4:96
 - group-level norms 4:97
 - item- and skill-level norms 4:97
 - local norms 4:97
 - national norms 4:97
 - user norms 4:97
 - score interpretation
 - criterion-referenced reading achievement levels 4:99*t*
 - general discussion 4:97
 - labeling comparisons 4:98*t*
 - percent of labeled students 4:99*t*
 - student reading profile report 4:100*f*
 - testing theories 2:577
 - paper-and-pencil tests
 - basic concepts 3:212
 - constructed response items
 - essay questions 3:214, 3:215*f*, 3:216*f*, 3:503*t*
 - general discussion 3:213
 - short-answer items 3:213, 3:503*t*
 - selected response items
 - fill-in-the-blank tests 3:503*t*
 - general discussion 3:212
 - matching items 3:212, 3:213*f*, 3:503*t*
 - multiple-choice items 3:212, 3:213*f*, 3:503*t*
 - true-false items 3:213, 3:503*t*
 - prior learning assessment (PLA) 1:171
 - score inflation 4:186–192
 - evidence 4:186, 4:187*f*, 4:187*t*
 - historical background 4:187
 - inadequate tradition approaches 4:187
 - validity of inferences under high-stakes conditions (VIHS)
 - behavioral responses 4:189
 - conceptual elements 4:190, 4:190*f*
 - effective test weight 4:189
 - evaluation frameworks 4:188
 - future directions 4:191
 - general discussion 4:188
 - target of inference 4:189, 4:189*f*
 - test anxiety 3:498–505
 - characteristics 3:498
 - comparison studies
 - alternate assessments 3:503*t*
 - cooperative learning activities 3:502
 - general discussion 3:501
 - online assessments 3:502
 - oral examinations 3:502
 - portfolios 3:502
 - product/project assessments 3:503
 - written tests 3:502
 - conservation of resources (COR) model
 - basic concepts 3:499, 3:500*f*
 - educational applications 3:499
 - future directions 3:503
 - general discussion 3:498
 - resource categories 3:499
 - resource loss and gain spirals 3:501, 3:501*f*
 - future directions 3:503
 - general discussion 3:504
 - implications 3:503
 - Yerkes–Dodson law 3:498, 3:499*f*
 - test-based accountability 4:153–158
 - general discussion 4:153
 - growth models
 - average gain scores 4:154
 - challenges 4:156
 - general discussion 4:154
 - random-effects models 4:155
 - student fixed-effects models 4:155
 - value tables 4:155, 4:155*f*
 - measurement issues
 - alignment analysis 4:157
 - general discussion 4:156
 - performance-level reporting 4:157
 - student inclusion 4:156
 - test preparation/score inflation 4:156

- performance indices
 - general discussion 4:153
 - growth models 4:154
 - quasi-longitudinal measures 4:154
 - status measures 4:153
 - successive cohort measures 4:154
- suggested guidelines 4:157
- United States
 - fairness 4:12
 - predictive validity
 - grade point average (GPA) 4:10
 - personal characteristics measures 4:11
 - primary admissions tests
 - ACT Assessment 4:8
 - general discussion 4:7
 - Graduate Management Admission Test (GMAT) 4:9
 - Graduate Record Examinations (GRE) 4:8
 - Law School Admission Test (LSAT) 1:695, 4:9
 - Medical College Admission Test (MCAT) 4:9
 - SAT examinations 4:8
 - usage
 - graduate/professional school programs 4:10
 - undergraduate institutions 4:9
- standards
 - alternative assessments 3:204
 - assessments and evaluations 3:380
 - early childhood education and care (ECEC)
 - Early Learning and Development Standards (ELDS) 2:138–143
 - conceptual background 2:138
 - development and implementation guidelines 2:139
 - general discussion 2:138
 - lessons learned 2:141
 - status reports 2:141
 - National Association for the Education of Young Children (NAEYC) 2:740
 - evaluation research
 - ethical practices 3:624
 - evaluation use/utilization 3:614–615, 3:615*t*
 - monitoring procedures 7:154
 - information and communication technology (ICT) 8:211
 - institutional management 4:277
 - leadership standards 4:768–776
 - definitions
 - general discussion 4:770
 - standards as measures 4:770
 - standards as principles 4:770
 - development guidelines
 - content standards 4:772
 - definitions 4:770
 - guiding conception of leadership 4:771
 - performance assessment measures 4:773
 - performance standards implementation 4:774
 - principal components 4:771
 - well-written characteristics 4:772
 - importance 4:768, 4:774
 - information and communication technology (ICT) 5:121
 - international developments 4:768
 - professional certification 4:769
 - professional learning 4:769
 - purpose 4:853
 - National Council for the Social Studies (NCSS) 1:413
 - professional teaching standards 3:418
 - second-language learning 1:442
 - social studies curriculum 1:413
 - teacher education 7:536
 - teacher evaluations 3:677
 - Thailand
 - educational professionals 5:854
 - educational quality 5:856
- standards-based accountability 2:342
- Standards for Educational and Psychological Testing 3:378, 3:380
- standards movement
 - fine arts education 3:263
 - history/civics assessments 4:177
 - school–community relations 5:106
- Stanford–Binet test 3:493, 4:2, 5:188
- Stanley model of talent identification 2:831
- start-up companies 4:302
- State Education Commission (SEC) 7:729
- State School Isaac Schraiber 8:76
- state–space models 7:479
- state's role in education *see* economics of education
- state–trait measures of mood 3:258
- statistical conclusion validity 3:147–151
 - definitions 3:147
 - importance 3:150
 - meta-analysis 3:150
 - threats
 - assumption violations 3:148
 - effect size estimation inaccuracies 3:150
 - error rates 3:148
 - extraneous variance 3:149
 - fishing 3:148
 - general discussion 3:147
 - heterogeneity of respondents 3:150
 - low statistical power 3:147
 - measurement unreliability 3:149
 - range restrictions 3:149
 - treatment implementation unreliability 3:149
- statistical inequalities 7:426–428
 - Cantelli's inequality 7:427
 - Chebyshev's inequality 7:426
 - Gauss' inequality 7:427
 - Markov's inequality 7:426
 - moment inequalities
 - general discussion 7:427
 - Hölder's inequality 7:428
 - Jensen's inequality 7:428
 - Lyapunov's inequality 7:428
 - Minkowski's inequality 7:428
- statistical inference
 - correspondence analysis (CA) 7:111
 - hypothesis testing 7:217
 - large-sample statistical methods 7:232–237
 - multiple comparison procedures 7:312
 - multivariate linear regression 7:326
 - order statistics 7:362
 - point estimation methods 7:366
 - probability theory 7:4
- statistical moderation 3:459
- statistical paradoxes 7:429–435
 - general discussion 7:429, 7:435
 - Lord's paradox 7:429
 - regression to the mean
 - body height analyses 7:433*f*, 7:434*f*
 - convergence trends 7:434
 - Galton's analyses 7:431, 7:433*f*
 - test results analysis 7:431, 7:432*f*
 - reversal paradox 7:429–430, 7:431*f*
 - Simpson's paradox 7:199, 7:199*f*, 7:210, 7:429
- statistical power analysis 7:436–443
 - analytical procedures
 - fixed power and effect size 7:437
 - fixed sample size and effect size 7:437, 7:438*t*
 - general discussion 7:437
 - minimum detectable effect size 7:437
 - basic concepts 7:436
 - general discussion 7:443
 - hypothesis testing 7:436
 - independent groups
 - covariate designs
 - computing power example 7:438*t*, 7:440, 7:442
 - multilevel (clustered) designs 7:442
 - operational effect size and sample size 7:440, 7:442
 - simple random sampling (SRS) 7:440
 - multilevel (clustered) designs
 - computing power example 7:438*t*, 7:441, 7:442
 - covariate designs 7:442
 - general discussion 7:441
 - operational effect size and sample size 7:441, 7:442
 - simple random sampling (SRS)
 - computing power example 7:438*t*, 7:440
 - covariate designs 7:440
 - effect size 7:437
 - general discussion 7:437
 - null hypothesis 7:436
 - sampling methods 3:146, 7:436
- statistical significance 7:444–450
 - effect size
 - common measures 7:447*t*
 - functional role
 - background information 7:448
 - confidence intervals 7:448

- statistical significance (*continued*)
 - interpretive framework 7:448
 - hierarchical linear models 7:446, 7:447*t*
 - measurement approaches 7:445
 - research background 7:445
 - standardized mean difference
 - multivariate D_M 7:445
 - univariate d (Cohen's d) 7:445
 - usage guidelines 7:449*f*
 - variance-accounted-for measures
 - basic concepts (η^2 , R^2) 7:445
 - corrected variance-accounted-for measures 7:446, 7:447*t*
 - η^2 versus partial η^2 7:446, 7:447*t*
 - general discussion 7:445
- statistical significance testing
 - basic concepts 7:444
 - criticisms 7:444
 - usage guidelines 7:449*f*
- status scales 4:96
- STEM (science, technology, engineering, and mathematics)
 - basic concepts 1:687
 - female participation 1:687
 - Gates Millennium Scholars program 1:847–848
 - gender equity 1:630, 1:686
 - needs assessments 3:791
 - student achievement 1:838
 - women of color 1:688
- Stenhouse, L. 7:542, 7:604
- STEP (systems, teachers, and education policy) studies 3:427
- stereotypic movement disorder 2:582
- Sternberg's Triarchic Componential Model (STTCM) 2:831
- stochastic frontier analysis (SFA) 2:211
- stochastic processes 7:451–452
 - Brownian motion
 - financial mathematics 7:452
 - historical perspectives 7:451
 - two-dimensional Brownian motion 7:452
 - fractional Brownian motion 7:452
 - historical perspectives 7:451
 - large-sample statistical methods 7:233
 - Markov chains 7:451
- storehouse metaphor 5:199
- storytelling *see* narrative practices
- strategic leadership 5:34–39
 - basic concepts 5:34
 - characteristics
 - activities and functions 5:34, 5:38*t*
 - culture and shared values development 5:36
 - distinctive competencies development 5:37
 - network development 5:37
 - organizational process development 5:35
 - personnel development 5:35
 - general discussion 5:38
 - social intelligence 5:37
 - student learning impacts 5:45
- stratified random sampling 3:144
- stratified sampling 3:156, 7:392
- stressful situations
 - coping strategies 6:570–575
 - acute versus chronic stressors 6:570
 - affective traits 3:257
 - classroom interventions 6:574
 - coping–volition relationship
 - dual processing self-regulation model 6:572, 6:573*f*
 - emotional arousal regulation 6:573
 - volitional control 6:573
 - effectiveness
 - general discussion 6:570
 - new stressor situations 6:571
 - situation assessments 6:571
 - strategy selection 6:571
 - modeling approaches 6:574
- strikes 2:513
- stroke 1:602, 1:603*f*
- structural cross-lagged regression model
 - basic concepts 3:124, 3:124*f*, 3:125*f*
 - self-concept study 3:125, 3:126*t*
- structural diversity 1:867
- structural equation models (SEMs) 7:453–458
 - basic concepts 7:5
 - Bayesian estimators 7:456
 - differential item functioning (DIF) 4:40
 - goodness-of-fit (GOF) models 7:192
 - health-related quality of life (HRQOL) example 7:456, 7:457*f*
 - linear structural equation models 7:453
 - maximum likelihood (ML) approach 7:455
 - methodologies 7:16
 - Monte Carlo Expectation-Maximization (MCEM) algorithm 7:455–456
 - nonlinear structural equation models 7:454
 - ordered categorical data 7:454
 - purpose 7:453
 - reciprocal effects model (REM) 6:664
 - school effectiveness research (SER) 6:270
 - two-level structural equation models 7:455
- structuralism 6:175
- structuralist sociology 1:505–509
 - code theory
 - Bernstein's theory 1:506
 - ESSA group studies 1:507
 - pedagogic communication modalities 1:506
 - gender research 4:797
 - general discussion 1:509
 - historical background 1:505
 - knowledge–curriculum relationship 1:508
 - knowledge structure theory 1:508
- structure-of-intellect theory 5:185
- struggling readers 6:221, 6:225
- student achievement
 - class size 2:443–449
 - class size–academic outcome relationship 2:443
 - class-size reduction research
 - benefits 2:449
 - empirical analyses 2:443
 - experimental research 2:444
 - Hawthorne effects 2:448
 - nonexperimental research 2:444
 - nonrandom movement effects 2:447
 - policy-induced reductions 2:449
 - Project STAR (Student Teacher Achievement Ratio) 2:43, 2:45*t*, 2:444, 5:54, 7:227
 - quasi-experimental research 2:448
 - randomization checks 2:445, 2:446*t*
 - test score results 2:446, 2:446*t*, 2:447*t*
 - general discussion 2:443
 - male underachievement 1:838–844
 - future research areas 1:842
 - general discussion 1:838
 - influencing factors
 - classroom behavior 1:841
 - cool and dumb behaviors 1:841
 - curriculum and class structure 1:842
 - family composition 1:840
 - hegemonic masculinity 1:840
 - identity development 1:840
 - racialized masculinity 1:840
 - student engagement 1:841
 - teachers' perceptions 1:841
 - research focus 1:839
 - research perspectives 1:839
- social class 2:423–430
 - child-rearing practices
 - academic achievement 2:425
 - after-school activities 2:424
 - literacy gaps 2:423
 - role models 2:424–425
 - general discussion 2:428
 - health influences 2:253, 2:258, 2:425
 - housing instability 2:427
 - income and wealth differences 2:427
 - Milwaukee Project 2:609
 - research background 2:423
- teacher quality 2:438–442
 - characteristics 2:438
 - educational production function model 2:409, 2:438
 - future research areas 2:441
 - indicators
 - degree level 2:440
 - subject-matter knowledge 2:440
 - teacher ability 2:440
 - teacher certification 2:439
 - teacher experience 2:439
- Student Achievement Guarantee in Education (SAGE) program 6:201

- student achievement test performance
 - background information 4:186
 - educational indicators 4:194
 - score inflation 4:186–192
 - evidence 4:186, 4:187f, 4:187t
 - historical background 4:187
 - inadequate tradition approaches 4:187
 - validity of inferences under high-stakes conditions (VIHS)
 - behavioral responses 4:189
 - conceptual elements 4:190, 4:190f
 - effective test weight 4:189
 - evaluation frameworks 4:188
 - future directions 4:191
 - general discussion 4:188
 - target of inference 4:189, 4:189f
- test-based accountability 4:153–158
 - general discussion 4:153
 - growth models
 - average gain scores 4:154
 - challenges 4:156
 - general discussion 4:154
 - random-effects models 4:155
 - student fixed-effects models 4:155
 - value tables 4:155, 4:155f
 - measurement issues
 - alignment analysis 4:157
 - general discussion 4:156
 - performance-level reporting 4:157
 - student inclusion 4:156
 - test preparation/score inflation 4:156
 - performance indices
 - general discussion 4:153
 - growth models 4:154
 - quasi-longitudinal measures 4:154
 - status measures 4:153
 - successive cohort measures 4:154
 - suggested guidelines 4:157
- Student Approaches to Learning (SAL) 4:461, 4:462
- student assessment systems 8:15
- student athletics
 - Scholar–Baller® program 1:860–865
 - background information 1:860
 - basic concepts 1:860
 - Coach Carter film screening 1:865
 - critical intervention programs 1:861
 - implementation case study 1:862
 - implications 1:864
 - logos 1:863
 - ThinkMan™ 1:860, 1:861f, 1:863
 - ThinkWoman™ 1:860, 1:861f, 1:863
- student attrition
 - higher education 4:467–472
 - cost estimation 4:471
 - estimation methods 4:467
 - general discussion 4:467
 - impacts 4:471
 - retention/intervention strategies 4:470
 - theoretical perspectives 4:468
 - see also* dropouts
- student-centered learning 8:391
- student data systems 8:14–20
 - background information 8:14
 - educational applications 8:16
 - future directions 8:17
 - general discussion 8:19
 - key features 8:15
 - system types 8:14
- student engagement 4:464
- student financial incentives 2:344
- student incentives 2:342–346
 - educational policies
 - standards-based accountability 2:342
 - student financial incentives 2:344
 - teacher financial incentives 2:344
 - educational production function model 2:342
 - general discussion 2:342, 2:345
- student information systems 8:15
- student leadership for learning 4:819
- student learning
 - assessment impacts 3:196–201
 - classroom assessment 3:198
 - post-assessments 3:198
 - pre-assessments 3:197
 - pure assessments 3:198
 - quality indicators 3:199
 - research areas 3:200
 - research background 3:196
- educational leadership 5:45–50
 - achievement assessments 5:47
 - general framework 5:45, 5:46f
 - leadership models 5:45
 - limitations 5:46
 - mediating variables 5:46f, 5:47
 - moderating variables 5:48
 - research trends 4:741, 5:21
 - unit of analysis challenge 5:48
- higher education
 - educational innovations 4:450
 - learning outcomes 4:479–484
 - benefits 4:483
 - cognitive complexity 4:479, 4:480f
 - humanitarianism 4:480f, 4:481
 - interpersonal and intrapersonal competence 4:480f, 4:482
 - knowledge acquisition 4:480, 4:480f
 - practical competence 4:480f, 4:482
 - research background 4:479
- Internet-based education 8:106
- leadership for learning 4:817, 4:818f
- lesson study 7:578
- middle schooling 6:234
- motivational regulation 6:631
- pedagogical content knowledge 7:658, 7:713
- professional learning communities 5:154
- science teacher education 7:713
- student loans
 - conventional loans 4:503
 - cost–benefit analysis 4:503
 - importance 4:501
 - income contingent loans (ICLs) 4:503
 - international reforms 4:504
- student mobility *see* academic mobility
- students' rights 3:378–383
 - assessments and evaluations
 - academic judgment challenges 3:382
 - assessment accountability 3:379
 - Code of Fair Testing Practices in Education 3:378, 3:380
 - fair practice standards 3:380
 - general discussion 3:382
 - Standards for Educational and Psychological Testing 3:378, 3:380
 - United Nations Convention on the Rights of the Child 3:378, 3:381
 - background information 3:378
- Student's *t* test 7:101, 7:313, 7:337, 7:349
- student support services 4:427–432
 - academic support
 - current perspectives 4:428
 - delivery approaches 4:428
 - historical background 4:427
 - basic concepts 4:427
 - contextual perspectives 4:427
 - distance education 4:430
 - future directions 4:431
 - information literacy
 - delivery approaches 4:429
 - historical background 4:428
 - importance 4:428
 - libraries/learning spaces
 - current trends 4:429
 - historical background 4:430
 - Learning Commons 4:430
- students with disabilities *see* special needs students
- Student/Teacher Achievement Ratio (STAR) project 6:201
- student voice 4:803–808
 - basic concepts
 - co-leadership role 4:805
 - consulting role 4:804
 - general discussion 4:804
 - leadership roles 4:805
 - benefits 4:806
 - challenges
 - anxiety 4:806
 - credibility 4:807
 - general discussion 4:806

- student voice (*continued*)
 - inclusivity 4:806
 - innovation issues 4:807
 - contemporary opportunities 4:803
 - general discussion 4:807
 - historical perspectives 4:803
- Studienakademie* 5:585*f*
- study abroad programs 1:570, 4:542, 4:559
- study centers 4:512
- stuttering 2:655
- subject-centered constructivism 1:488–489
- subjective well-being 6:732–738
 - characteristics 6:732
 - educational environments
 - general discussion 6:733
 - student well-being
 - general discussion 6:733
 - predictors 6:735, 6:735*f*
 - research approaches 6:734
 - response discrepancies 6:734
 - teacher well-being 6:736
 - general discussion 6:732
 - predictors 6:733
- submersion language programs 1:295–296, 1:296*f*
- sub-Saharan Africa
 - background information 3:449
 - classroom assessment 3:449–456
 - assessment types
 - continuous assessments 3:450, 3:451*t*
 - international assessment programs 3:449
 - national assessment programs 3:449
 - public (external) examinations 3:450
 - background information 3:449
 - challenges 3:454
 - curriculum structure 3:450*t*
 - enrollment 3:449
 - pupil–teacher ratios (PTRs) 2:506*f*
 - student mobility 4:561
 - see also* South Africa
- subtractive bilingualism 1:294, 1:296*f*, 5:356
- subtractive schooling 3:594
- Success for All program 2:100, 6:222
- Sudan 2:111*t*
- Sue–Okazaki concept of relative functionalism 5:453
- Sullivan, Anne 2:621
- Sultanate of Oman 5:718–726
 - background information 5:718
 - education system
 - administrative and supervisory structure 5:724
 - challenges 5:725
 - demographic data 5:718–719, 5:719*t*
 - educational finance 5:724, 5:725*t*
 - educational status 5:719*t*
 - enrollment 5:719*t*
 - goals 5:719
 - manpower training 5:720
 - monitoring and evaluation programs 5:724
 - Omanization 5:720
 - research programs 5:724
 - structural and operational characteristics
 - adult education 5:724
 - basic education 5:721, 5:721*t*
 - career guidance 5:722
 - general discussion 5:720
 - higher education 5:723
 - international schools 5:723
 - nonformal education 5:724
 - postbasic education 5:722
 - preschool education 5:721
 - private schools 5:722
 - special education 5:722
 - teaching profession 5:724
 - technical education 5:723
 - illiteracy rate 5:724, 5:724*t*
- summative assessments
 - accountability 3:792, 3:793*f*, 3:794*t*
 - adult education and training (AET) 3:550
 - alternative assessments 3:205
 - basic concepts 3:378
 - classroom assessment 3:211, 3:411, 3:415, 3:451*t*
 - cross-cutting issues 3:166
 - design-based research 6:371
 - diagnostic assessments 3:397
 - educational impacts 4:69–74
 - general discussion 4:69
 - high-visibility external tests
 - influencing factors 4:72
 - potential benefits 4:71
 - problematic effects 4:71
 - historical perspectives 4:69
 - measurement purposes 4:69
 - research areas 4:73
 - epistemological perspectives 3:399
 - formative assessment–summative assessment continuity 3:351
 - general discussion 3:312, 7:151
 - higher education 3:249
 - illustrative example 3:353, 3:355
 - mathematics education 3:281, 3:653
 - portfolios 3:238
 - primary school science
 - combination assessments 3:297
 - evidence summaries 3:297
 - policies and practices 3:296
 - special tasks or tests 3:296
 - research–quality assessments 4:312
 - rubrics and scoring guides 4:124
 - school-based prevention programs 2:826
 - science–education programs 3:663
 - secondary school science 3:300
 - teacher evaluations 3:675
 - teacher's role 3:384–391
 - case studies
 - coursework assessments 3:386
 - General Certificate of Secondary Education (GCSE) 3:386
 - New Zealand Education Monitoring Project (NZEMP) 3:386
 - state-wide portfolio assessments 3:385
 - characteristics 3:384
 - contextual perspectives 3:390
 - process model
 - inference estimations 3:389
 - judgment and clarity 3:388
 - performance standard criteria 3:388
 - task conditions 3:388
 - task specification 3:388
 - purpose 3:384
 - quality assurance 3:389
 - systematic approaches 3:389
 - technology education 3:314
 - technology evaluations 3:579
 - vocational education and training (VET) 3:339
- summative curriculum evaluation model 1:337
- sum of squares calculations 7:33
- supervision 7:615–621
 - current perspectives
 - diversity considerations 7:618
 - licensure alternatives 7:618
 - school–university partnerships 7:619
 - social justice considerations 7:618
 - field experiences 7:619
 - future directions 7:620
 - general discussion 7:620
 - Greece 5:602
 - historical background 7:615
 - induction programs 7:616
 - modeling approaches 7:619
 - personal identity construction 7:616
 - purpose 7:615
 - research background 7:616
 - Slovenia 5:787
 - supervision triad
 - characteristics 7:617
 - cooperating teacher 7:617
 - teacher education candidate 7:617
 - university supervisor 7:617
 - technological tools
 - general discussion 7:619
 - online supervision 7:620
 - reflection/self-analysis 7:619
 - Thailand 5:854, 5:855*f*
- supervision administration 4:839–845
 - challenges
 - business-based accountability 4:844

- decentralized leadership 4:844
- general discussion 4:843
- quality instruction 4:844
- teacher shortages 4:843
- functional role
 - compensation policies and practices 4:843
 - evaluation policies and practices 4:842
 - general discussion 4:841
 - induction programs 4:841
 - professional development 4:842
 - recruitment 4:841
 - selection methods 4:841
 - supervision strategies 4:842
- general discussion 4:839, 4:844
- historical background 4:840
- policy considerations 4:840
- purpose 4:839
- support vector classifiers 7:137
- Supreme Court decisions 1:692, 1:823, 1:883, 2:394, 2:688
- Sure Start program (UK) 2:57, 2:70, 2:128, 2:158, 2:159, 8:257–258
- Suriname 5:826–833
 - background information 5:826
 - early childhood education and care (ECEC) 2:111*t*
 - education system
 - administrative and supervisory structure 5:831
 - challenges
 - curricular policies 5:833
 - decentralization 5:832
 - information and communication technology (ICT) 5:832
 - monitoring and evaluation programs 5:831
 - educational finance 5:831, 5:831*t*
 - goals 5:826
 - objectives 5:827
 - performance assessments 5:831
 - staff–student ratio 5:832*t*
 - structural and operational characteristics
 - adult education 5:829
 - general discussion 5:827
 - higher education 5:830
 - interior sector 5:828
 - junior secondary education 5:828
 - preprimary education 5:827
 - primary education 5:827
 - schematic diagram 5:829*f*, 5:830*f*
 - second-chance education 5:830
 - senior secondary education 5:829
 - special education 5:828
 - higher education
 - enrollment rates 1:702*f*, 1:703*f*
 - structural and operational characteristics 5:830
- survey research methods 5:455
 - applications 3:152
 - data quality 3:159
 - definitions 3:152
 - objectives 3:153
 - personal epistemology assessments 5:214
 - questions and responses
 - open questions versus closed questions 3:153
 - responses 3:154
 - survey question stems 3:154
- reliability
 - basic concepts 3:157
 - internal consistency/homogeneity 3:158
 - interrater/intrarater reliability 3:158
 - test–retest reliability 3:157
- response rates 3:157
- sampling methods
 - characteristics 3:155
 - sample size 3:157
- survey instruments
 - interviews 3:154
 - reliable/valid survey instruments 3:157
 - self-administered questionnaires 3:154
 - structured observations 3:155
 - structured record reviews 3:154
- survey sampling
 - characteristics 3:155
 - eligibility criteria 3:155
- validity
 - basic concepts 3:158
 - construct validity 3:159
 - content validity 3:158
 - criterion validity 3:159
 - face validity 3:158
- survival data analysis 7:459–466
 - computer-based testing 7:459
 - computer software 7:465
 - discrete time events 7:465
 - general discussion 7:10
 - historical background 7:460
 - limitations 7:459
 - logit model 7:465
 - nonparametric/semiparametric models
 - Cox's proportional hazards (PH) regression model 7:462, 7:463*f*, 7:463*t*
 - Kaplan–Meier (K–M) Product-limited Method 7:460, 7:461*f*, 7:462*f*
 - parametric models 7:464
 - terminology and meaning 7:460
 - Weibull model 7:464, 7:465*f*, 7:465*t*
- sustainable development 1:746–752
 - background information 1:750
 - basic concepts 1:750
 - educational implications 1:747
 - educational reform 6:289–294
 - educational practices
 - Finland 6:292
 - London Borough of Tower Hamlets 6:293
 - importance 6:289
 - integrated and evidence-based theories 6:291
 - theoretical perspectives 6:293
 - unsustainability era
 - age of innovations 6:290
 - large-scale reform (LSR) efforts 6:290–291
 - reculturing efforts 6:290
 - restructuring efforts 6:290
 - educational role 1:749
 - future directions 1:751
 - general discussion 6:148
 - importance 1:751
 - Judeo-Christian traditions 1:747–748
 - modern society values and cultural perspectives 1:747
 - vocational education and training (VET) 8:416–422
 - economic sustainability 8:418
 - educational role 8:417
 - environmental sustainability 8:417, 8:418*f*, 8:419*f*
 - functional role 8:421
 - implications 8:416
 - social sustainability 8:419
- sustained attention 5:166
- Swaziland
 - early childhood education and care (ECEC) 2:111*t*
 - socioeconomic/sociopolitical characteristics 1:199
 - student achievement test performance 2:418*t*
 - see also* Africa
- Sweatt v Painter* (1950) 1:693
- Sweden
 - age-related policies and programs 8:263*t*
 - childhood brain development 2:57*f*
 - competitive schooling 2:335
 - displaced worker assistance programs 8:278
 - early childhood education and care (ECEC)
 - enrollment rates 2:4*t*, 2:56*t*, 2:111*t*, 2:114*t*
 - funding estimates 2:58*t*
 - integrated services 2:104, 2:107*t*
 - leave policies 2:3
 - earnings–education level relationship 2:307*t*
 - education levels 2:279*f*
 - enrollment rates 2:493*f*
 - ethical practices 4:718
 - gender-equity gap 2:56*t*
 - higher education
 - admissions tests 4:12–13
 - collective bargaining 4:493
 - foreign doctoral students 4:292*t*
 - income contingent loans (ICLs) 4:504
 - institutional expenditures 1:733*t*
 - loans and grants 4:501*f*, 4:502*f*
 - student and researcher mobility 4:287*f*
 - tertiary education expenditures 4:500*f*
 - tertiary type-B program survival rates 8:464*t*
 - tuition fees 4:500*f*
 - inequality–education relationship
 - earnings differentials 2:277*t*

- Sweden (*continued*)
 employment rate—educational attainment relationship 2:278f
 gender differences 2:277t
 immigrant status 2:275f
 test score—social background relationship 2:273f
 information and communication technology (ICT)
 computer access at home 8:25t
 computer access in school 8:23t
 frequency of computer use in school and at home 8:26t
 student proficiency and confidence 8:31t
 Komvux
 background information 1:103
 labor-market training programs 1:103
 leadership development programs 4:753t, 4:754t, 4:755t
 participation research
 regular education 8:284f
 vocational education and training (VET) 8:283f, 8:284f
 work-related vocational education and training (VET) 8:284f, 8:285f
 popular adult education 1:163
 prison education programs 8:301, 8:301t
 research-quality assessments 4:309
 returns to education—education level 2:308t
 school choice 6:285
 student achievement test performance 2:247f
 teacher age distribution 2:492f
 teacher shortages 2:494f
 teacher wages and salaries 2:496f
 vocational education and training (VET)
 age-related policies and programs 8:263t
 initial vocational preparation system 8:251t
 outcomes analyses 8:252f, 8:253f
 participation research 8:283f, 8:284f, 8:285f
 school-based programs 8:247
 vocational—academic education parity 8:435
 women in employment 2:56f
 SweSAT 4:12–13
 Switzerland
 age-related policies and programs 8:263t
 childhood brain development 2:57f
 classroom assessment 3:432–437
 background information 3:432
 competency-based programs 3:433
 effective communication strategies 3:434
 external testing 3:435
 formative assessments 3:434
 general discussion 3:436
 multilingual sources 3:432
 report cards 3:434
 school reform 3:432
 standards 3:435
 translated meanings 3:432
 class size 2:492f
 early childhood education and care (ECEC) 2:4t, 2:111t
 earnings—education level relationship 2:307t
 education levels 2:279f
 enrollment rates 2:493f
 higher education
 binary systems 4:617
 federal political systems 4:606
 foreign doctoral students 4:292t
 institutional expenditures 1:733t
 loans and grants 4:501f, 4:502f
 student and researcher mobility 4:287f
 tertiary education expenditures 4:500f
 tuition fees 4:500f
 inequality—education relationship
 employment rate—educational attainment relationship 2:278f
 immigrant status 2:275f
 test score—social background relationship 2:273f
 information and communication technology (ICT)
 computer access in school 8:23t
 frequency of computer use in school and at home 8:26t
 student proficiency and confidence 8:31t
 leadership development programs 4:753t, 4:754t, 4:755t
 research-quality assessments 4:309
 returns to education—education level 2:308t
 student achievement test performance 2:247f
 teacher age distribution 2:492f
 teacher shortages 2:494f
 teacher wages and salaries 2:496f
 vocational education and training (VET) 8:232, 8:251t, 8:263t
 women in employment 2:56f
 Sydney Accord 4:210, 4:210t
 syllabus design 1:362–367
 curriculum models
 critical model 1:364
 general discussion 1:364
 outcomes-based model 1:364
 process-based model 1:364
 traditional content model 1:365
 definitions 1:362
 general discussion 1:366
 high-equity/high-quality systems 1:365
 published research
 English-language teaching 1:363
 general discussion 1:363
 higher education teaching 1:364
 technical form 1:365
 symbolic interactionism 6:481–487
 basic assumptions 6:481
 basic concepts 6:482
 exemplary models
 George Herbert Mead 6:483
 Howard Becker 6:484
 Peter Woods 6:484
 focus and scope 6:481
 historical background 6:481
 participant observation 6:445
 relevancy 6:486
 scholarly associations 6:483
 symptomatic accounting techniques (SAT) 7:411
 Syria
 early childhood education and care (ECEC) 2:111t
 information and communication technology (ICT) 8:25t
 Trends in International Mathematics and Science Study (TIMSS) 5:645t
 systematic desensitization 6:555t
 systematic sampling 3:144, 3:156, 7:393
 systematic screening for behavior disorders (SSBD) 2:824
 systemic racism
 African Americans
 higher education
 consequences 1:770
 historical background 1:767
 historically black colleges and universities (HBCUs) 1:769
 predominantly white universities 1:768
 characteristics 6:37
 whiteness 6:40
 system leadership 5:78–84
 activities and capabilities 5:80, 5:81f, 5:82f
 basic concepts 5:78
 challenges 5:83
 development opportunities 5:81
 functional roles 5:79
 general discussion 5:78
 school-to-school collaboration 6:279, 6:279f
 System of Intelligent Evaluation using Tests (SIETTE) 8:63
- T**
 tactile communication cards 2:624
ta'dy'b (good action) 6:71
 Taiwan 5:834–840
 background information 5:834
 education system
 administrative and supervisory structure 5:839
 challenges 5:840
 educational finance 5:839
 goals 5:834
 monitoring and evaluation programs 5:839
 research programs 5:839
 structural and operational characteristics
 adult education 5:836f, 5:837t, 5:838
 continuing education 5:836f, 5:837t, 5:838
 early childhood education and care (ECEC) 5:835, 5:836f, 5:837t
 general discussion 5:835
 higher education 5:836f, 5:837, 5:837t
 preschool education 5:835, 5:836f, 5:837t
 primary education 5:835, 5:836f, 5:837t
 schematic diagram 5:836f
 secondary education 5:835, 5:836f, 5:837t
 special education 5:836f, 5:837t, 5:838
 staff and students 5:837t
 supplementary education 5:838

- vocational education and training (VET) 5:836*f*, 5:837, 5:837*t*
- teaching profession 5:838
- inequality–education relationship 2:273*f*
- information and communication technology (ICT)
 - computer access at home 8:25*t*
 - computer access in school 8:23*t*
 - mathematics and science instruction 8:32*t*
- private higher education (PHE) 4:622
- student achievement test performance 2:247*f*
- test score–social background relationship 2:273*f*
- Tajikistan 2:111*t*
- talented students *see* gifted and talented students
- Talents Unlimited Inc. (TU) 2:831
- Talloires Declaration (2005) 4:402
- ta'ly'm* (socialization) 6:71
- Tanzania 5:841–846
 - background information 5:841
 - early childhood education and care (ECEC) 2:111*t*
 - education system
 - administrative and supervisory structure 5:844
 - background information 5:841
 - challenges 3:454
 - classroom assessment 3:450, 3:451*t*, 3:454
 - curriculum structure 3:450*t*
 - educational finance
 - donor support 5:845
 - state funding 5:845
 - exit points/continuation/exit skills 3:452*t*
 - goals 5:842
 - monitoring and evaluation programs 5:845
 - postcolonial theory 1:300–301
 - reform efforts 5:845
 - research programs 5:845
 - structural and operational characteristics
 - enrollment 5:843*t*
 - general discussion 5:842
 - higher education 5:843
 - preschool education 5:843
 - primary schools 5:843
 - secondary schools 5:843
 - vocational education and training (VET) 5:844
 - teaching profession 5:844
 - higher education
 - gender equity 1:630–631
 - structural and operational characteristics 5:843
 - nonformal education 1:42
 - returns to education–cognitive skills 2:310*t*
 - school fees 2:419
 - socioeconomic/sociopolitical characteristics 1:199
 - student achievement test performance 2:418*t*
 - see also* Africa
- tarbiyyah* (individuation) 6:71
- task goals 6:537–538
- Task Group on Assessment and Testing (TGAT) 3:359
- taxes 2:224
- teacher-based action research 1:355–361
 - future directions and challenges 1:359
 - general discussion 1:355
 - historical background
 - Collier's approach 1:357
 - continuity approach 1:358
 - democratic approach 1:358–359
 - Dewey's approach 1:356
 - Lewin's approach 1:357
 - practitioner research approach 1:359
 - Progressive Era 1:356
 - Wann review 1:355
- Teacher Career Cycle Model 7:583
- teacher career stages 7:582–586
 - background information 7:582
 - emerging environmental changes
 - external environmental factors 7:583
 - general discussion 7:583
 - organizational factors 7:583
 - personal environmental factors 7:585
 - teacher shortages 7:584
 - future directions 7:585
 - modeling approaches
 - early models 7:582
 - nonlinear dynamic models 7:583
 - research-based linear models 7:582
- teacher education
 - accreditation 7:495–502
 - challenges 7:497
 - elementary education 7:728
 - future directions 7:499
 - general discussion 7:495
 - governance processes 7:495
 - jurisdictional functions 7:496
 - professionalization process 7:497
 - professional standards 7:495
 - public professional initiatives 7:500
 - standards-based accreditation
 - challenges 7:498
 - evidence weaknesses 7:498
 - mixed professional support 7:498
 - weak professional consensus 7:498
 - systemic effectiveness 7:499
 - adapted physical activity programs 2:526
 - Australia 5:513
 - Austria 5:521
 - Bangladesh
 - recruitment and retention 5:529
 - training 5:528
 - Barbados 5:538
 - case-based education methods 7:592–597
 - characteristics 7:592
 - craft knowledge 7:647
 - empirical bases 7:593
 - purpose 7:592
 - theoretical bases
 - cognitive flexibility theory 7:594
 - general discussion 7:594
 - reasoning and reflective practice 7:595
 - situated cognition theory 7:594
 - social cognitive theory 7:595
 - theory–practice gap 7:596
 - conceptual approaches 7:598–603
 - clinician–professional approach 7:599*t*, 7:601
 - contemplative approach 7:599*t*, 7:600
 - general discussion 7:598, 7:602
 - social justice approach 7:598, 7:599*t*
 - teaching for understanding approach 7:599, 7:599*t*
 - craft knowledge 7:642–649
 - accessibility
 - case-based education 7:647
 - general discussion 7:648
 - incorporation strategies 7:647
 - knowing-in-action 7:645
 - stimulated recall interview 7:647
 - characteristics 7:750*t*
 - importance 7:648
 - knowledge base considerations 7:642
 - research background
 - characteristics 7:644
 - content characteristics 7:644, 7:646*t*
 - research-generated knowledge–teaching practice relationship 7:643, 7:644*f*
 - diversity considerations 7:518–524
 - characteristics 7:518
 - disabilities 7:519
 - educational preparedness
 - general discussion 7:522
 - goals 7:522
 - methodologies 7:523
 - pedagogical strategies 7:522
 - general discussion 7:523
 - occurrences 7:519*t*
 - pedagogical strategies
 - courses and programs 7:522
 - critical self-reflection 7:522
 - field experiences 7:522
 - general discussion 7:522
 - good practice models 7:523
 - societal issues 7:522
 - religious diversity 7:519
 - sexual diversity 7:519
 - student teachers 7:520
 - teaching competencies 7:521
 - workforce composition 7:519, 7:520*t*
 - elementary education 7:725–731
 - accreditation 7:728
 - background information 7:725

- teacher education (*continued*)
- challenges
 - educational reform 7:729
 - professionalism 7:729
 - teacher shortages 7:729
 - general discussion 7:730
 - in-service training 7:728
 - preservice education programs
 - academic and professional coursework 7:727
 - characteristics 7:726
 - examinations 7:727
 - practicum programs 7:728
 - tertiary degree requirements 7:727
 - system characteristics 7:726
- England 5:554
- Fiji 5:566
- formative assessments 3:563–571
- assessment instruments 3:563
 - assessment tools
 - peer assessments 3:566
 - portfolios 3:565
 - self-assessments 3:567
 - deployment stages 3:563, 3:564f
 - design guidelines 3:564, 3:565t
 - guiding principles 3:564t
 - informative feedback 3:563
 - intervention strategies 3:564
 - purpose 3:563
 - research-based principles 3:567, 3:568t
- Gambia 5:861
- Ghana 5:592t, 5:594
- gifted and talented students 7:732–737
- challenges 7:735
 - desirable characteristics 7:735, 7:735t
 - diversity considerations 7:736
 - general discussion 7:736
 - philosophical perspectives 7:732
 - research methodologies and findings 7:734
 - standards-based programs
 - assessment strategies 7:733
 - collaboration 7:734
 - communication skills 7:733
 - ethical practices 7:734
 - foundations 7:733
 - general discussion 7:733
 - individual learning differences 7:733
 - instructional planning 7:733
 - instructional strategies 7:733
 - learning capabilities 7:733
 - learning environments 7:733
 - social interactions 7:733
 - teacher training 7:732
- governance role 7:532–539
- accreditation 7:495
 - background information 7:532
 - general discussion 7:537
 - modeling approaches
 - basic concepts 7:532, 7:533f
 - Bologna Accord (1999) 7:536
 - decentralized centralism 7:534
 - globalization effects 7:534
 - management by objectives (MBO) 7:534
 - neoliberal ideologies and strategies 7:534
 - new institutionalism 7:535
 - policy tools 7:535
- Iceland 5:608
- Indonesia 5:614, 5:614t
- knowledge base
 - craft knowledge 7:642
 - educational foundations 7:662–668
 - good theory concept 7:666
 - naive teaching 7:662
- Latvia 5:633
- Lebanon
 - characteristics 5:641
 - distribution by age 5:642t
 - distribution by category 5:642t
 - distribution by degree and number of schools 5:642t, 5:643t
- long-term effects 7:503–510
- definitions 7:503
 - empirical evidence 7:503
 - general discussion 7:508
 - research approaches
 - analytical methodology 7:508
 - data collection 7:508
 - decision-oriented evaluation models 7:505
 - design requirements 7:507
 - instrumentation 7:507
 - limitations 7:505
 - study methodologies 7:505
 - research goals 7:503
 - research relevance 7:503
 - state of the art research
 - research reviews 7:504
 - seminal studies 7:504
 - study methodologies
 - general discussion 7:505
 - large-scale quantitative studies 7:506
 - mixed-methods research 7:506
 - phenomenographic studies 7:506
 - validity requirements 7:507
- Malawi 5:670
- Malta 5:679
- mathematics teacher education 7:702–705
- future directions 7:704
 - general discussion 7:702
 - learning trends
 - general discussion 7:702
 - reflective practices 7:703
 - social dimensions 7:702
 - teaching conditions 7:704
 - research areas 7:702
- mentoring processes 7:749–756
- Mongolia 5:684
- mother tongue teacher education 7:738–743
- challenges 7:741
 - field experiences 7:741
 - general discussion 7:738
 - pedagogical content knowledge 7:739
 - professional development
 - current issues 7:739
 - demographic shifts 7:740
 - literacy considerations 7:741
 - methods courses 7:740
 - subject-matter knowledge 7:738
- Namibia 5:688
- narrative practices 7:610–614
- characteristics 7:610
 - critical incident vignette 7:610
 - language 7:612
 - professional importance 7:613
 - social context 7:612–613
 - student teachers 7:611
 - voice 7:613
- New Zealand 5:696
- Pakistan 5:733, 5:733t, 5:734t
- pedagogical concepts 7:587–591
- general discussion 7:590
 - historical background 7:587
 - learning about teaching
 - general discussion 7:589
 - reflective practices 7:590
 - student-teacher-as-researcher approach 7:590
 - student to teacher transition 7:590
 - practical considerations 7:587
 - teacher educators 7:678–679
 - teaching about teaching
 - characteristics 7:589
 - language articulation 7:589
 - tacit aspects 7:589
- Philippines 5:871
- prospective teacher evaluation 7:511–517
- cognitions and decision-making processes
 - background information 7:512
 - peer assessments 7:512
 - portfolios 7:512
 - self-assessments 7:512
 - simulation methods 7:512
 - general discussion 7:516
 - good teaching views 7:511
 - instructional behaviors 7:511
 - learning-oriented assessments 7:512, 7:513t

- purpose 7:511
- quality assessments 7:514
- standards-based assessments 7:513
- validity measures
 - characteristics 7:514
 - consequential validity 7:515
 - constructs and assessment task criteria 7:514
 - extended interpretations 7:514
 - reliability 7:515
- Qatar 5:746t, 5:748**
- reflective practices 7:629–634**
 - characteristics 7:629
 - current perspectives 7:630
 - historical background 7:629
 - model continua
 - content 7:632
 - general discussion 7:631
 - means 7:632
 - purpose 7:631
 - significance 7:633
- Russia 5:754**
- science teacher education 7:712–718**
 - curricula design 7:714
 - instructional strategies 7:715
 - pedagogical content knowledge
 - assessment methods 7:714
 - curricula 7:714
 - instructional strategies 7:713
 - student learning 7:713
 - philosophical orientations 7:713
 - preparation assessments 7:716
 - prospective teachers 7:712
 - research background 7:712
 - subject-matter knowledge 7:712
- Singapore 5:779**
- social studies teacher education 7:719–724**
 - challenges
 - citizenship development 7:720
 - content considerations 7:719–720
 - disciplinary content versus method/pedagogy 7:720
 - general discussion 7:719
 - purpose 7:719
 - research limitations 7:721
 - educational trends
 - authentic pedagogy 7:722
 - culturally relevant pedagogy (CRP) 7:723
 - general discussion 7:722
 - global citizenship education (GCE) 7:722
 - multicultural education 7:723
 - future directions 7:723
- sociocultural factors 7:525–531**
 - career attractiveness 7:529
 - certification and licensing 7:529
 - entry requirements 7:526
 - general discussion 7:525
 - globalization–regionalism tensions 7:525, 7:529
 - program curriculum 7:528
 - structural organization 7:527
- Solomon Islands**
 - gender analysis 5:795t
 - general discussion 5:795
 - primary school teachers 5:795t
 - secondary school teachers 5:795t
 - vocational education and training (VET) teachers 5:796t
- Spain 5:810**
- special education 2:691**
- Sri Lanka 5:818**
- standards 7:536**
- supervision 7:615–621**
 - current perspectives
 - diversity considerations 7:618
 - licensure alternatives 7:618
 - school–university partnerships 7:619
 - social justice considerations 7:618
 - field experiences 7:619
 - future directions 7:620
 - general discussion 7:620
 - historical background 7:615
 - induction programs 7:616
 - modeling approaches 7:619
 - personal identity construction 7:616
 - purpose 7:615
 - research background 7:616
 - supervision triad
 - characteristics 7:617
 - cooperating teacher 7:617
 - teacher education candidate 7:617
 - university supervisor 7:617
 - technological tools
 - general discussion 7:619
 - online supervision 7:620
 - reflection/self-analysis 7:619
- Taiwan 5:838**
- Teaching for Understanding (TfU) 7:635–641**
 - activity theory 7:637, 7:638f
 - future directions 7:640
 - general discussion 7:635, 7:640
 - importance 7:640
 - key features
 - achievement assessments 7:636
 - collaborative practices 7:636
 - generative topics 7:636
 - goals 7:636
 - performance assessments 7:636
 - learning contexts
 - continuous inquiry 7:639
 - formal learning 7:638
 - general discussion 7:638
 - organizational learning 7:639
 - WIDE World online courses 7:639
 - mastery teaching programs 7:635
 - networked/online technologies 7:637
- theory–practice relationship 7:669–675**
 - background information 7:669
 - gap-bridging strategies 7:672
 - impact limitations 7:669
 - integrated approaches 7:673
 - practical versus formal knowledge 7:671
 - research areas 7:673
 - research background 7:669
 - research–practice gap 7:670
 - technical–rationality model 7:669
- Zanzibar 5:899**
- Teacher Education Accreditation Council (TEAC) 7:497–498, 7:728–729**
- Teacher Education Development Study–Mathematics (TEDS–M) 4:664**
- teacher educators 7:676–680**
 - characteristics 7:676
 - future directions 7:679
- professional development 7:681–688**
 - characteristics 7:681, 7:682f
 - common activities
 - advanced academic degrees 7:684
 - feedback opportunities 7:685
 - portfolios 3:565, 7:686
 - professional organizations and conferences 7:684
 - research and scholarship 7:684
 - self-study 7:685
 - short courses and lectures 7:684
 - constraints
 - motivation and confidence 7:686
 - support 7:686
 - time and resource availability 7:686
 - context
 - off-job learning 7:683
 - on-job learning 7:683
 - feedback opportunities
 - colleague-based feedback 7:685
 - student-based feedback 7:685
 - superior-based feedback 3:563, 7:685
 - formative assessments 3:563–571
 - assessment instruments 3:563
 - deployment stages 3:563, 3:564f
 - design guidelines 3:564, 3:565t
 - guiding principles 3:564t
 - informative feedback 3:563
 - intervention strategies 3:564
 - peer assessments 3:566
 - portfolios 3:565
 - purpose 3:563
 - research-based principles 3:567, 3:568t
 - self-assessments 3:567
 - general discussion 7:686

- teacher educators (*continued*)
- purposes
 - beginning teachers 7:682
 - competency development 7:682
 - educational improvements 7:681
 - external standards 7:681
 - professional status 7:682
 - research approaches 7:677
 - scholarship 7:677
- teacher financial incentives 2:344
- teacher incentives **2:481–488**
- advantages
 - group incentives
 - efficiency and productivity gains 2:483
 - research findings 2:485*t*
 - individual incentives
 - efficiency and productivity gains 2:482
 - recruitment and retention 2:483
 - research findings 2:485*t*
 - background information 2:481
 - developing countries 2:507
 - disadvantages 2:484
 - general discussion 2:486
 - rationale 2:481
 - unintended consequences 2:486
- teacher induction programs **7:563–568**
- challenges 7:566
 - characteristics 7:565
 - critical elements 7:566
 - general discussion 7:563
 - goals 7:564
 - historical background 7:563
 - importance 7:567
 - mentoring programs 2:477, 7:566
 - recruitment and retention 2:477
 - special education teachers 2:889
 - supervision 7:616
- teacher leadership **5:40–44**
- basic concepts 5:40
 - distributed leadership 5:40
 - educational impacts
 - school level 5:41
 - student level 5:42
 - teacher level 5:42
 - information and communication technology (ICT) 5:122, 5:123*f*
 - leadership for learning 4:818
 - organizational development considerations 5:40
 - professional development considerations 5:43
- teacher learning
- information and communication technology (ICT) **8:211–216, 8:222–227**
 - certification and licensing 8:212
 - change management
 - educational innovations 8:224
 - general discussion 8:224
 - Second Information Technology in Education Studies (SITES) 8:224
 - understanding/self-understanding factors 8:224
 - competency standards 8:213
 - curriculum frameworks 8:211
 - functional role
 - activity clusters 8:225*t*
 - general discussion 8:225
 - knowledge–innovation relationships 8:226*t*
 - general discussion 8:225
 - importance 8:211
 - integration strategies 8:215
 - knowledge-deepening versus knowledge-creation approach 8:214
 - knowledge demands 8:222
 - literacy skills development 8:212
 - need requirements
 - formal versus informal knowledge 8:223, 8:223*t*
 - knowledge domains 8:222, 8:223*t*
 - knowledge in practice 8:223, 8:223*t*
 - knowledge levels 8:223, 8:223*t*
 - socioeconomic skills 8:213
 - standards 8:211
 - professional development
 - lesson study 7:576
 - three-part teacher learning model 7:577, 7:577*f*
 - teacher mistreatment/abuse **4:790–796**
 - abuse effects 4:793
 - abusive behavior typologies 4:791, 4:792
 - basic concepts 4:790
 - general discussion 4:790
 - modeling approaches 4:791
 - research approaches
 - future directions 4:794
 - methodologies 4:794
 - published results 4:793
 - theoretical perspectives 4:791
 - teacher quality **2:438–442**
 - characteristics 2:438
 - educational production function model 2:409, 2:438
 - future research areas 2:441
 - indicators
 - degree level 2:440
 - subject-matter knowledge 2:440
 - teacher ability 2:440
 - teacher certification 2:439
 - teacher experience 2:439
 - student achievement impacts 2:438
 - vocational teacher education 7:747
 - teacher–researcher curriculum evaluation model 1:337–338
 - teachers' beliefs **7:622–628**
 - characteristics 7:623
 - definitions 7:622
 - general discussion 7:626
 - importance 7:622
 - measurement approaches 7:624
 - prospective teachers
 - change strategies 7:625
 - educational innovations 7:627
 - mathematics 7:626
 - research background 7:624
 - relation to motivational variables and processes 7:623, 7:623*f*
 - teachers' unions **2:512–517**
 - general discussion 2:516
 - historical background 2:512
 - influential factors 2:512
 - regulatory laws 2:513
 - school resources–earnings relationship 2:318
 - teachers' unions 2:513
 - union impacts
 - district-wide impacts
 - general discussion 2:515
 - union–school district spending relationship 2:516
 - union–student achievement relationship 2:516
 - union–student/teacher ratio relationship 2:515
 - union–wage relationship 2:515
 - union–working conditions relationship 2:515
 - general discussion 2:513
 - negative impacts
 - community welfare reductions 2:514
 - democratic process threats 2:515
 - general discussion 2:514
 - lack of trust considerations 2:514
 - performance incentives reductions 2:514
 - resource allocation reductions 2:514
 - positive impacts
 - efficiency-enhancing resources 2:514
 - information flow improvements 2:514
 - market failure reduction 2:514
 - stable school districts 2:514
 - teacher quality 2:513
 - time management 2:514
 - teacher supply **2:489–497**
 - demand elements
 - age distribution 2:490, 2:492*f*
 - class size 2:490, 2:492*f*
 - enrollment rates 2:493*f*
 - pupil demographics 2:490
 - teacher shortages 2:491–492, 2:493*f*
 - teaching hours 2:490–491, 2:493*f*
 - future directions 2:496
 - general discussion 2:489
 - geographic location 2:474
 - influencing factors 2:492
 - labor market 2:473, 2:489, 2:490*f*
 - psychological benefits and costs 2:474
 - recruitment and retention **2:473–480**
 - barriers 2:475
 - certification and licensing 2:475
 - demand factors

- general discussion 2:475
- hiring processes 2:475
- institutional constraints 2:476
- student enrollment 2:475
- student-to-teacher ratio reductions 2:475
- teacher retirement 2:475
- general discussion 2:479
- geographic location 2:474
- policies and programs
 - career ladders 2:478
 - due process reform 2:478
 - entry requirement changes 2:477
 - general discussion 2:476
 - hiring practice improvements 2:478
 - induction programs 2:477
 - mentoring programs 2:477
 - monetary incentives 2:476
 - partnership models 2:476
 - performance-based pay 2:477
 - tenure policies 2:478
- psychological benefits and costs 2:474
- supply-and-demand model 2:473
- wages and salaries 2:473
- working conditions 2:473
- school responses 2:494, 2:494^f
- supply-and-demand model 2:473–480, 2:490, 2:491^f
- wages and salaries 2:473, 2:481, 2:495^f, 2:496, 2:496^f
- working conditions 2:473
- teacher training and licensing
 - Greece 5:602
 - United States 2:498–503
 - academic achievement outcomes 2:499
 - background information 2:498
 - current licensing requirements 2:499
 - effectiveness variations 2:502
 - empirical research 2:499
- Teach First 6:266
- Teach for America (TFA) 2:498, 6:264
- Teaching and Learning International Survey (TALIS) 5:665
- Teaching and Learning Research Programme (TLRP) 8:353
- teaching assistants *see* paraprofessionals
- Teaching for Understanding (TFU) 7:635–641
 - activity theory 7:637, 7:638^f
 - future directions 7:640
 - general discussion 7:635, 7:640
 - importance 7:640
 - key features
 - achievement assessments 7:636
 - collaborative practices 7:636
 - generative topics 7:636
 - goals 7:636
 - performance assessments 7:636
 - learning contexts
 - continuous inquiry 7:639
 - formal learning 7:638
 - general discussion 7:638
 - organizational learning 7:639
 - WIDE World online courses 7:639
 - mastery teaching programs 7:635
 - networked/online technologies 7:637
- Teaching Games for Understanding (TGfU) 1:461, 1:462
- technology acceptance model (TAM) 8:21
- technology education 3:311–315
 - assessment models
 - classroom interactions and feedback 3:313
 - formative assessments 3:312
 - general discussion 3:311
 - information and communication technology (ICT) 3:314
 - sociocultural perspectives 3:311
 - summative assessments 3:314
 - teacher's role 3:312
 - characteristics 3:311
 - female participation 1:687
 - gender equity 1:630, 1:686
 - general discussion 3:315
 - Poland 5:739
 - Sri Lanka 5:822
 - Sultanate of Oman 5:723
 - women of color 1:688
- technology enhanced learning (TEL) 4:423
- Technology Enhanced Student Assessment (TESA) 8:49
- technology evaluations 3:578–584
 - evaluation method selection process 3:582, 3:583^f
 - importance 3:578
 - key issues
 - implementation considerations 3:580
 - measurement characteristics and control 3:581
 - treatment control 3:580
 - methodologies
 - combined methods 3:580
 - qualitative methods 3:580
 - quasi-experiments 3:579
 - random assignment approximations 3:579
- technology literacy and fluency 8:34–41
 - basic concepts 8:36
 - educational implications
 - curriculum changes 8:39
 - digital divides 8:39
 - literate lives 8:38
 - performance assessments 8:39
 - framework development 8:37, 8:38^f, 8:39^f
 - future directions 8:40
 - historical perspectives 8:35
 - multiple literacies 8:37
 - social implications 8:34
- Technology-Related Assistance Act (1988) 2:673
- technology-supported learning environments 8:1–10
 - academic values and expectations 4:436
 - access and equity patterns 8:95–101
 - Canada 8:97
 - China 8:98
 - European Union (EU) 8:97
 - general discussion 8:95
 - global access and use 8:98, 8:98^t
 - improvement efforts 8:99
 - India 8:98
 - international trends 8:21–33
 - challenges 8:30
 - computer access at home 8:25, 8:25^t
 - computer access in school 8:22, 8:23^t
 - frequency of use at home 8:25, 8:26^t
 - frequency of use in school 8:25, 8:26^t
 - large-scale assessments 8:22
 - mathematics and science instruction 8:30, 8:32^t
 - patterns of use 8:26, 8:27^f
 - student access 8:21
 - student proficiency and confidence 8:27, 8:31^t
 - theoretical perspectives 8:21
 - United States
 - home access 8:95, 8:96^t
 - home use 8:95
 - outcomes analyses 8:95, 8:97
 - school access 8:96, 8:96^f
 - school use 8:96
- arts curricula 1:423
- Australia 3:422
- automated essay scoring (AES) 4:20–26
 - available programs 4:21
 - basic concepts 4:20
 - educational applications
 - automated speech-recognition scoring programs 4:24
 - Criterion online essay evaluation service 4:22
 - English language learners (ELLs) 4:23
 - general discussion 4:22
 - Text Adaptor 4:23
 - general discussion 4:25
 - historical foundations 4:20
 - modeling approaches 4:20–21
 - reliability and validity 4:22
- CompEd study
 - characteristics 3:687^t
 - Functional Information Technology Test (FITT) 4:645
 - importance 4:649
 - purpose 4:644
 - Stage 1 survey results 4:644
 - Stage 2 survey results 4:645
- computerized adaptive testing (CAT)
 - background information 7:396
 - basic concepts 7:397
 - dynamic assessments 3:222–223
 - item response theory (IRT) applications 4:87
 - purpose 4:148

- technology-supported learning environments (*continued*)
- sampling methods 7:395
 - sequential probability ratio test (SPRT) 7:397, 7:398*t*
 - educational applications
 - adult learning 1:40
 - educational impacts 8:2
 - effectiveness 8:7
 - emerging uses
 - general discussion 8:6
 - immersive environments and games 4:422, 8:7
 - interactive classroom communication systems 8:7
 - social networks 4:422, 5:120, 8:6
 - Web 2.0 4:431, 8:6
 - general discussion 8:9
 - implementation strategies 8:8
 - out-of-school learning 8:9
 - rapidly growing uses
 - assessment and accountability systems 8:6
 - online courses 8:5
 - open-source online resources 8:5
 - virtual schools 8:5
 - subject-specific uses
 - language instruction 8:4
 - mathematics instruction 8:4
 - reading and literacy 8:3
 - science instruction 8:4
 - writing instruction 8:3
 - typical uses 8:2
 - educational leadership 5:120–127
 - educational impacts
 - community relations 5:121, 5:121*f*
 - general discussion 5:120
 - instructional strategies 5:121, 5:121*f*
 - literacy instruction 5:121, 5:121*f*
 - social networks 5:120, 5:121*f*
 - influencing factors 5:125, 5:125*f*
 - integration framework 5:123*f*
 - requirements and standards 5:121
 - social and legal challenges 5:124
 - teacher leadership 5:122, 5:123*f*
 - technology-mediated leadership development programs 5:123, 5:123*t*
 - elementary/middle-grades reading curricula 1:466
 - flexible learning 4:441–447
 - background information 4:441
 - basic concepts 4:423
 - blended learning 4:445
 - criticisms 4:445
 - definitions 4:441
 - delivery challenges 4:444
 - delivery effectiveness 4:443
 - democratization of learning 4:444
 - flexible delivery 4:442
 - general discussion 4:441, 4:446
 - information and communication technology (ICT) 4:442
 - philosophical perspectives 4:444
 - responsiveness and availability 4:445
 - general discussion 8:1–2
 - growth factors 8:2
 - help seeking processes 6:657
 - higher education 4:419–426
 - assessment strategies 3:253
 - background information 4:419
 - curriculum 4:420
 - effective learning
 - collaborative learning 4:422
 - communication-based learning 4:422
 - construction-based learning 4:421
 - instructional strategies 4:421
 - e-learning pedagogies
 - average comparative ratings 1:531*t*
 - basic concepts 1:522
 - content appropriateness and relevancy 1:522
 - course goals 1:523
 - dimensions analysis 1:532
 - diversity considerations 1:523
 - effectiveness 4:423, 4:423*t*
 - interactive learning strategies 1:524
 - learners' needs 1:523
 - learning styles 1:522
 - multimedia 1:524
 - flexible learning
 - background information 4:441
 - basic concepts 4:423, 4:441–447
 - blended learning 4:445
 - criticisms 4:445
 - definitions 4:441
 - delivery challenges 4:444
 - delivery effectiveness 4:443
 - democratization of learning 4:444
 - flexible delivery 4:442
 - general discussion 4:441, 4:446
 - information and communication technology (ICT) 4:442
 - philosophical perspectives 4:444
 - responsiveness and availability 4:445
 - future directions 4:425
 - teaching approaches 4:424
 - technology enhanced learning (TEL) 4:423
 - instructional management 8:133–139
 - background information 8:133
 - classroom network technologies
 - effectiveness 8:137
 - general discussion 8:136
 - instructional strategies 8:137
 - key features 8:136
 - future directions 8:137
 - student-response systems
 - effectiveness 8:135
 - general discussion 8:134
 - key features 8:134
 - research areas 8:134
 - teaching models 8:134
 - international access patterns 8:21–33
 - challenges 8:30
 - computer access at home 8:25, 8:25*t*
 - computer access in school 8:22, 8:23*t*
 - frequency of use at home 8:25, 8:26*t*
 - frequency of use in school 8:25, 8:26*t*
 - large-scale assessments 8:22
 - mathematics and science instruction 8:30, 8:32*t*
 - patterns of use 8:26, 8:27*f*
 - student access 8:21
 - student proficiency and confidence 8:27, 8:31*t*
 - theoretical perspectives 8:21
 - lifelong learning 8:184–188
 - computer software 8:185
 - distance education 8:185
 - general discussion 8:184
 - home schooling 8:184
 - Internet cafes 8:186
 - learning centers 8:185
 - self-directed learning 8:186
 - technical certifications 8:186
 - workplace learning 8:184
 - mathematics education 8:172–183
 - challenges 8:180
 - collaborative problem solving
 - The Adventures of Jasper Woodbury* 8:177
 - Anchored Instruction 8:177
 - Enhanced Anchored Mathematics 8:177, 8:178*f*
 - MathWorlds 8:178
 - networked devices 8:177
 - participatory simulations 8:177–178
 - serious game systems 8:178
 - SimCalc 8:178, 8:179*f*; 8:199
 - embodied cognition 8:179, 8:180*f*
 - importance 8:172
 - mathematical discovery
 - ANIMATE system 8:175, 8:176*f*
 - Dynamic Geometry systems 8:176
 - dynamic statistics packages 8:176
 - Function Probe 8:175
 - general discussion 8:174
 - Logo 8:175, 8:199
 - skills and concept learning
 - background information 8:172
 - Cognitive Tutors 8:173, 8:174*f*
 - computer algebra systems (CASs) 8:173
 - Geometric Supposer 8:173
 - hand-held graphing calculators 8:172, 8:173*f*
 - intelligent tutoring systems (ITSs) 8:173
 - new media technologies 8:140–149
 - animation

- general discussion 8:143
- picture comprehension 8:144
- picture design 8:144
- realistic and dynamic pictures 8:143, 8:143f
- general discussion 8:141
- learning models
 - cognitive theory of multimedia learning (CTML) 8:145, 8:146f
 - design guidelines 8:148
 - dual-coding theory 8:145
 - integrated model of text and picture comprehension (IMTPC) 8:145, 8:146f
 - research findings 8:147
 - theoretical models 8:145
- multimedia learning environments
 - animation 8:143
 - future directions 8:148
 - general discussion 8:141
 - learning models 8:145
 - linear and nonlinear text 8:142
 - text design guidelines 8:142
- research findings
 - coherence and contiguity 8:147
 - general discussion 8:147
 - modality 8:147
 - redundancy 8:147
- text design guidelines
 - additional stimulation 8:142
 - conciseness 8:142
 - general discussion 8:142
 - linguistic simplicity 8:142
 - structure and organization 8:142
- personal epistemology 5:213
- physically disabled students 2:642
- preservice education programs 8:217–221
 - challenges 8:220
 - purpose 8:217
 - teaching preparation strategies 8:218
- technology as an educational tool
 - digital technology 8:219
 - distance education 8:219
 - electronic portfolios 8:219
 - usage preparation strategies 8:217
- professional development 7:554
- reading skills 8:150–157
 - beginning-reading programs 6:222
 - comprehension 8:152
 - emergent literacy 8:150
 - fluency 8:151
 - general discussion 8:150, 8:155
 - Internet access 8:154
 - phonics and phonemic awareness 8:151
 - software systems 8:3
 - text-to-speech applications 8:153
 - theoretical perspectives 8:154
 - upper elementary grades 6:224
 - vocabulary 8:152
- school–community relations 8:125–132
 - alternative schooling approaches
 - accountability data sharing 8:128
 - current trends 8:130
 - distance education 8:129
 - general discussion 8:128
 - one-to-one computing 8:128
 - virtual schools 8:129
 - challenges 8:125
 - importance 8:125
 - research areas 8:130
 - research background 8:126
 - technology initiatives
 - e-mail communication 8:126
 - growth factors 8:128
 - Internet-based approaches 8:126
 - websites 8:127
- science 1:468–474, 8:197–203
 - access and equity promotion 8:202
 - challenges 8:202
 - constructivist perspectives 8:197, 8:198t
 - contextualization
 - authenticity 8:201
 - everyday experiences 8:201
 - design-based research
 - curriculum development 1:468
 - design guidelines 1:472
 - design patterns approach 1:469, 1:470t, 1:471f
 - design principles approach 1:469, 1:471f
 - thermodynamics curriculum unit 1:470, 1:471f
 - future directions 8:202
 - future research areas 1:472
 - general discussion 1:468
 - information access 8:201
 - learning environment transformation 8:201
 - physical environment interactions
 - model construction 8:199
 - scientific phenomena 8:198, 8:198t
 - simulations/dynamic representations 8:199
 - student investigations 8:199
 - reflective construction of knowledge
 - program management 8:200
 - teacher support and professional development 8:200
 - working memory (WM) capacity 8:200
 - social environment interactions 8:199
 - systemic impacts 8:202
- Second Information Technology in Education Studies (SITES)
 - basic concepts 4:646
 - educational impacts 4:667
 - importance 4:649
 - Lithuania 5:665
 - Module 1 (SITES M1) survey results 4:646
 - Module 2 (SITES M2) survey results 4:647
 - SITES 2006 4:648
- second-language learning 8:204–210
 - background information 8:204
 - benefits
 - anywhere/anytime instruction 8:208
 - general discussion 8:207
 - immediate feedback 8:208
 - individualized study plans 8:207
 - mistakes in private concept 8:208
 - patient tutoring 8:208
 - records of achievement 8:209
 - classroom instruction 8:205
 - English-language training demand 8:204, 8:205f
 - functional role 8:204
 - research results and analysis 8:207
 - subject-specific uses 8:4
- self-adaptive testing 4:148–152
 - assisted self-adaptive testing (ASAT) 4:149–150
 - basic concepts 4:148, 4:149f
 - benefits 4:148
 - dynamic assessments 3:222–223
 - general discussion 4:151
 - item pool 4:149
 - item-selection strategies 4:150, 4:150f
 - perceived control/perceived performance hypotheses 4:148
 - restricted self-adaptive testing (RSAT) 4:149
 - theoretical investigations 4:150
- situated learning 5:337
- socioemotional aspects 6:718–723
 - background information 6:718
 - general discussion 6:718, 6:721
 - methodological challenges 6:721
 - motivation 6:719
 - pedagogical models 6:718
 - research approaches 6:720
- speech and language impaired students 2:658
- teacher learning 8:211–216, 8:222–227
 - certification and licensing 8:212
 - change management
 - educational innovations 8:224
 - general discussion 8:224
 - Second Information Technology in Education Studies (SITES) 8:224
 - understanding/self-understanding factors 8:224
 - competency standards 8:213
 - curriculum frameworks 8:211
 - functional role
 - activity clusters 8:225t
 - general discussion 8:225
 - knowledge–innovation relationships 8:226t
 - general discussion 8:225
 - importance 8:211
 - integration strategies 8:215
 - knowledge-deepening versus knowledge-creation approach 8:214

- technology-supported learning environments (*continued*)
 knowledge demands 8:222
 literacy skills development 8:212
 need requirements
 formal versus informal knowledge 8:223, 8:223*r*
 knowledge domains 8:222, 8:223*r*
 knowledge in practice 8:223, 8:223*r*
 knowledge levels 8:223, 8:223*r*
 socioeconomic skills 8:213
 standards 8:211
- Teaching for Understanding (TfU) 7:635–641
 activity theory 7:637, 7:638*f*
 future directions 7:640
 general discussion 7:635, 7:640
 importance 7:640
 key features
 achievement assessments 7:636
 collaborative practices 7:636
 generative topics 7:636
 goals 7:636
 performance assessments 7:636
 learning contexts
 continuous inquiry 7:639
 formal learning 7:638
 general discussion 7:638
 organizational learning 7:639
 WIDE World online courses 7:639
 mastery teaching programs 7:635
 networked/online technologies 7:637
- technology education 3:311–315
 assessment models
 classroom interactions and feedback 3:313
 formative assessments 3:312
 general discussion 3:311
 information and communication technology (ICT) 3:314
 sociocultural perspectives 3:311
 summative assessments 3:314
 teacher's role 3:312
 characteristics 3:311
 general discussion 3:315
see also assistive technology (AT)
- technology transfer 4:303, 6:523
 Telecommunications Act (1996) 4:303
 television captioning 8:119
 television's impact 8:109
 temperament 6:588
 tendinitis 8:113
- Tennessee
 Project STAR (Student Teacher Achievement Ratio) 2:43, 2:45*t*, 2:444, 5:54, 7:227
 teacher incentives 2:483
 tuition and needs-based grants 1:654*f*
- Tennessee Value-Added Assessment System (TVAAS) 4:155
- Terman, Lewis 2:870
- tertiary education
 academic mobility 4:559–565
 future directions 4:564
 government policies and programs 4:559
 historical background 4:559
 internationalization processes 4:542
 professional impacts 4:563
 statistical data
 faculty mobility 4:562
 general discussion 4:560
 student mobility 4:561
 student perspectives 4:562
- access and equity patterns 1:636–648
 academic preparation 1:637
 college knowledge 1:637
 demographic changes 1:636
 financial considerations 1:637
 geographical information systems (GIS)
 advantages 1:638
 characteristics 1:637
 college knowledge 1:645*f*, 1:646*f*
 data sources and methods 1:638
 illustrative examples 1:638
 racial demographics 1:643*f*, 1:644*f*
 research implications 1:647
 Tucson elementary schools 1:639*f*, 1:643*f*, 1:644*f*
 Tucson high schools 1:641*f*, 1:645*f*, 1:646*f*
 Tucson middle schools 1:640*f*
 international access patterns 4:334
 labor market 1:636
 social stratification 1:637
- Tucson, Arizona
 college knowledge 1:645*f*, 1:646*f*
 elementary schools 1:639*f*, 1:643*f*, 1:644*f*
 geographical information systems (GIS) research 1:638
 high schools 1:641*f*, 1:645*f*, 1:646*f*
 middle schools 1:640*f*
 racial demographics 1:643*f*, 1:644*f*
 research implications 1:647
 research methodologies 1:638
- admittance pathways 4:609–615
 articulation 4:612
 bachelor's programs 4:609
 entry qualifications 4:610*t*, 4:611*r*
 postgraduate programs 4:611
 promotion strategies 4:613
 second chance students 4:611
 student transfers 4:610
- Australia 5:511*f*, 5:513
- Bulgaria
 Academic Degrees and Titles Act (1972) 5:542
 Higher Education Act (1995) 5:542
 number of students 5:548*t*
 number of teachers 5:548*t*
 private universities 5:548*t*
 Promotion of Scientific Research Act (2003) 5:542
 state universities 5:548*t*
 structural and operational characteristics 5:543
- collective bargaining 4:491–498
 basic concepts 4:491
 challenges 4:496
 country analysis
 Argentina 4:496
 Australia 4:496
 Brazil 4:496
 Cameroun 4:494
 Chile 4:496
 China 4:495
 East Africa 4:493
 France 4:493
 general discussion 4:492
 Germany 4:493
 India 4:494
 Indonesia 4:495
 Japan 4:495
 Malaysia 4:495
 Mexico 4:496
 Nigeria 4:494
 Russia 4:493
 South Africa 4:494
 South Korea 4:494
 Sweden 4:493
 United Kingdom 4:493
 United States 4:492
 West Africa 4:493
 international standards 4:491
- emerging economies 4:520–526
 Bangladesh 4:523
 challenges 4:525
 China 4:520, 4:521*t*
 cross-cutting issues
 employability 4:525
 expansion financing 4:523
 private sector services 4:524, 4:524*t*
 quality assurance 4:525
 relevancy 4:525
 enrollment rates 4:521*t*
 India 4:521, 4:521*t*, 4:522*t*
 literacy rates 4:521*t*
 Pakistan 4:523
 public expenditures 4:521*t*
 Sri Lanka 4:523
- Estonia 5:556
- federal political systems 4:603–608
 Australia 4:606
 Canada 4:604
 characteristics 4:603
 general discussion 4:608

- Germany 4:605
- Mexico 4:607
- Switzerland 4:606
- United States 4:604
- Fiji 5:566
- Finland 5:570*f*, 5:572
- France 5:578
- Gambia 5:861
- Germany 5:585*f*, 5:587, 5:588, 5:589*t*
- Ghana 5:592*t*, 5:594
- Greece 5:601
- Iceland 5:607*f*, 5:608
- indigenous education 1:663
- Indonesia 5:612*f*, 5:613, 5:615, 5:615*t*
- institutional autonomy 4:341–346
 - background information 4:341
 - definitions 4:341
 - external boundaries 4:342
 - general discussion 4:345
 - internal boundaries 4:343
- international access patterns 4:334–340
- gender parity index 4:336*t*
- general discussion 4:339
- global participation 4:335, 4:335*f*
- research areas
 - data collection 4:338
 - expansion limits 4:338
 - funding considerations 4:339
 - meritocratic entry concepts 4:338
 - open access policies 4:338
 - privatization impacts 4:339
- research data collection 4:334
- social equity studies 4:334
- underrepresented groups 4:336, 4:336*t*
- Israel 5:621, 5:622*t*
- Jamaica 5:627, 5:628*f*
- Latvia 5:632, 5:633*t*
- learning cities and regions 1:147
- Lebanon 5:641
- Lithuania 5:660
- Malawi 5:671
- Malta 5:676*f*, 5:678
- management practices 4:251–255
 - business-like practices 4:253
 - changing contexts
 - competition 4:252, 4:265
 - financial considerations 4:252
 - institutional growth 4:251–252
 - collegiality 4:252
 - new public management (NPM)
 - background information 4:251
 - ideological perspectives 4:251
 - strategic planning 4:256, 4:257*f*
 - organizational culture 4:253
- Mongolia 5:682*t*, 5:683
- Namibia 5:687
- national science policy 4:294–299
 - general discussion 4:299
 - governance processes 4:295
 - historical background 4:294
 - institutional structures 4:297
 - policies and programs 4:295
 - research challenges 4:298
 - state–university relations 4:296
- New Zealand 5:695
- Niger 5:702
- Norway 5:711*f*, 5:712*t*, 5:713
- Organisation for Economic Cooperation and Development (OECD)
 - access and equity patterns 4:334
 - cross-border higher education (CBHE) 4:587
 - Directorate for Science, Technology and Industry (DSTI) 4:588
 - future scenarios 4:587
 - global dimensions 4:588
 - Institutional Management in Higher Education (IMHE) 4:588
 - policies and programs 4:349, 4:586
- Philippines 5:868, 5:868*t*
- Poland 5:739
- postconflict societies 4:514–519
 - basic concepts 4:514
 - functional role 4:515
 - future directions 4:519
 - intervention strategies
 - Afghanistan 4:518
 - general discussion 4:516
 - Iraq 4:518
 - South Africa 4:516
 - south-southeast Europe (SEE) 4:517
 - postconflict requirements 4:516
- Qatar 5:747
- research funding 4:279–285
 - background information 4:279
 - commercialization efforts 4:284
 - government policies and programs 4:281
 - priority setting 4:282
 - public funding 4:281
 - quality assessments 4:283
 - research characteristics and functions 4:280
 - university–industry collaboration 4:283
- Rwanda 5:762, 5:762*t*
- school-based vocational education and training 8:249
- Scotland 5:772
- Slovenia 5:786
- Solomon Islands 5:794
- South Africa 5:802, 5:802*t*
- Spain 5:809*t*, 5:810
- Sri Lanka 5:816
 - higher education institutions 5:822
 - postgraduate programs 5:822
 - public university education 5:822
 - technical education 5:822
 - vocational education and training (VET) 5:822
- strategic planning 4:256–263
 - basic model
 - deliberate versus emergent strategies 4:258, 4:258*f*
 - empirical analyses 4:257
 - general discussion 4:256
 - new public management (NPM) approach 4:257*f*
 - planning cycle 4:257*f*
 - plan outline 4:258*f*
 - current trends 4:256
 - goal and performance cards 4:262, 4:262*f*
 - goals and objectives 4:256, 4:257*f*
 - internal contract management 4:261
 - methodologies and instrumentation
 - general discussion 4:260
 - portfolio analyses 4:261, 4:262*f*
 - strengths, weaknesses, opportunities, and threats (SWOT) analyses 4:261
 - performance areas 4:257
 - strategic perspectives 4:259, 4:260*f*
- student attrition 4:467–472
 - cost estimation 4:471
 - estimation methods 4:467
 - general discussion 4:467
 - impacts 4:471
 - retention/intervention strategies 4:470
 - theoretical perspectives 4:468
- Sultanate of Oman 5:723
- Suriname 5:830
- Taiwan 5:836*f*, 5:837, 5:837*t*
- Tanzania 5:843
- Thailand 5:853
- tuition and fees 2:229–234
 - cost research 2:229
 - future directions 2:233
 - graduate and professional programs 2:232
 - private institutions 2:230
 - public institutions 2:232
- United States 5:892, 5:892*t*, 5:893*t*
- Uruguay 5:875*f*
- vocational–academic education parity 8:433
- Zanzibar 5:898, 5:898*t*
- see also* higher education
- test adaptations 4:166–169
 - background information 4:166
 - differential item functioning (DIF) analysis 4:167
 - future directions 4:168
 - score equivalency assessments 4:167
 - transadaptation process 4:166
- test anxiety 3:498–505
 - affective traits 3:257
 - assessment comparison studies
 - alternate assessments 3:503*t*

- test anxiety (*continued*)
 - cooperative learning activities 3:502
 - general discussion 3:501
 - online assessments 3:502
 - oral examinations 3:502
 - portfolios 3:502
 - product/project assessments 3:503
 - written tests 3:502
- characteristics 3:498
- conservation of resources (COR) model
 - basic concepts 3:499, 3:500^f
 - educational applications 3:499
 - future directions 3:503
 - general discussion 3:498
 - resource categories 3:499
 - resource loss and gain spirals 3:501, 3:501^f
- evaluative anxiety 6:550, 6:552^r
- future directions 3:503
- general discussion 3:504
- implications 3:503
- Yerkes–Dodson law 3:498, 3:499^f
- test-based accountability 4:153–158
 - general discussion 4:153
 - measurement issues
 - alignment analysis 4:157
 - general discussion 4:156
 - performance-level reporting 4:157
 - student inclusion 4:156
 - test preparation/score inflation 4:156
- performance indices
 - general discussion 4:153
 - growth models
 - average gain scores 4:154
 - challenges 4:156
 - general discussion 4:154
 - random-effects models 4:155
 - student fixed-effects models 4:155
 - value tables 4:155, 4:155^f
 - quasi-longitudinal measures 4:154
 - status measures 4:153
 - successive cohort measures 4:154
- suggested guidelines 4:157
- test development 4:159–165
 - process model
 - content definition 4:160
 - effective item principles 4:161
 - general discussion 4:159
 - item banking 4:164
 - item flaws 4:162
 - overall test plan 4:159
 - passing score standards 4:163, 4:163^r
 - prompt development 4:161
 - response scoring
 - general discussion 4:163
 - item analysis 4:163
 - key validation 4:163
 - summary documentation 4:164
 - test administration 4:162
 - test design and assembly 4:162
 - test form production 4:162
 - test results reporting 4:164
 - test specifications
 - characteristics 4:161
 - rationale 4:161
 - test blueprint 4:161
 - twelve-step plan 4:160^r
- test equating and linking 4:50–55
 - basic concepts 4:50
 - general discussion 4:50, 4:54
 - item response theory (IRT) linking
 - basic concepts 4:52
 - NEAT design
 - basic concepts 4:53
 - concurrent calibration 4:53
 - estimation methods 4:53
 - fixed parameters calibration 4:53
 - Haebara linking 4:54
 - mean–mean/mean–var calibrations 4:53
 - Stocking–Lord linking 4:54
 - score equating methods 4:54
 - true score equating methods 4:54
 - observed-score equating methods
 - basic concepts 4:51
 - Kernel equating method 4:52
 - Rasch model
 - general discussion 4:114
 - person–item location distributions 4:115^f, 4:115^t, 4:116^f
 - scaling methods 4:51
- Test of English as a Foreign Language (TOEFL) 4:24, 8:49
- test–retest reliability 3:157, 4:28, 4:29^r
- test score measurements 7:267–273
 - central moments 7:267
 - dispersion
 - general discussion 7:267, 7:272
 - interquartile range (IQR) 7:268, 7:268^f
 - median absolute deviation (MAD) 7:269
 - range 7:268, 7:268^f
 - standard deviation 7:268, 7:269^f
 - variance 7:268, 7:269^f
 - distribution shapes
 - bimodal distributions 7:271^f
 - kurtosis 7:271^f, 7:272
 - normal distributions 7:270, 7:270^f
 - skewness 7:271, 7:271^f
 - kurtosis 7:267–268, 7:269^f, 7:271^f, 7:272
 - normal distributions 7:267, 7:270, 7:270^f
 - skewness 7:267–268, 7:269^f, 7:271, 7:271^f
- Texas
 - Asian American/Pacific Islander student enrollment 1:597^f
 - cost-function analysis 2:208^r
 - state and school district data systems 2:182
 - tuition and needs-based grants 1:654^f
- Text Adaptor 4:23
- text-based learning
 - adult development
 - discourse structures 5:228
 - general discussion 5:227
 - situation model processing 5:228
 - textbase processing 5:227
 - vocabulary 5:227
 - multiple information sources 5:191
- textbooks 1:397–402
 - current perspectives 1:401
 - e-textbooks 1:393
 - functional role 1:397
 - interaction knowledge 1:400
 - power knowledge 1:400
 - publishing industry 1:368–373
 - America's Choice 1:371
 - current trends 1:372
 - global context 1:369
 - knowledge control and influence 1:369
 - textbook production 1:368
 - United States 1:370
 - research perspectives 1:398
 - social studies curriculum 1:413
 - UNESCO report and recommendations 1:397
 - vocational teacher education 7:746
 - work knowledge 1:399
- text-to-speech applications 8:153
- Thailand 5:847–857
 - background information 5:847
 - comparative education societies 1:270
 - early childhood education and care (ECEC) 2:111^r
- education system
 - administrative and supervisory structure 5:854, 5:855^f
 - educational finance 5:855
 - framework
 - Constitution BE 2540 (1997) 5:848
 - goals 5:848
 - National Education Act (1999) 5:848
 - National Education Plan (2002–16) 5:849, 5:849^f
 - general discussion 5:856
 - standards and quality assurance 5:856
- structural and operational characteristics
 - adult education 5:854
 - characteristics 5:850, 5:850^f
 - early childhood care 5:850
 - ecclesiastic education 5:852
 - educational institutions 5:851^r
 - higher education 5:853
 - nonformal education 5:854

- preschool education 5:850, 5:850^f
- primary education 5:850^f, 5:852
- secondary education 5:850^f, 5:852
- special needs students 5:852
- teaching profession 5:854
- vocational education and training (VET) 5:853
- higher education
 - enrollment rates 4:522^t
 - private higher education (PHE) 4:623
- inequality–education relationship 2:273^f
- informal learning 8:242–243
- information and communication technology (ICT)
 - computer access at home 8:25^t
 - computer access in school 8:23^t
 - frequency of computer use in school and at home 8:26^t
 - mathematics and science instruction 8:32^t
 - student proficiency and confidence 8:31^t
- school supply shortages 2:506^f
- student achievement test performance 2:247^f, 2:417^t
- teacher education programs 7:526–527
- test score–social background relationship 2:273^f
- Thematic Apperception Test (TAT) 6:499–500
- theological schools *see* clergy education
- Theory Movement 4:725
- theory of explanatory coherence 3:78
- theory–practice gap 1:56, 7:596
- thermodynamics curriculum unit 1:470, 1:471^f
- think before reading, think while reading, think after reading (TWA) strategy
 - 2:762, 2:763^t
- ThinkMan™ 1:860, 1:861^f, 1:863
- ThinkWoman™ 1:860, 1:861^f, 1:863
- Third Age Learning International Studies (TALIS) 1:179
- third culture kids (TCKs) 4:636–643
 - adult third culture kids (ATCKs) 4:637
 - basic concepts 4:636, 4:637^f, 4:639^f
 - cross-cultural kids (CCKs) 4:639, 4:640^f
 - general discussion 4:642
 - globalization effects 4:636
 - hidden diversity issues
 - basic concepts 4:640, 4:640^f
 - cultural identity matrix 4:641^f
 - spheres of influence 4:641^f
 - identity challenge responses 4:641
 - shared experiences
 - benefits and challenges
 - adaptability 4:638
 - cultural imbalances 4:638
 - general discussion 4:637
 - identity challenges 4:638
 - multilingualism 4:638
 - relationship considerations 4:638
 - resilience 4:638
 - rootlessness versus restlessness 4:638
 - unresolved grief 4:639
 - worldwide versus local perspectives 4:637
 - characteristics 4:637
 - general discussion 4:637
- Third International Mathematics and Science Study 1995 (TIMSS) 3:687^t, 4:652
- Third International Mathematics and Science Study 2003 (TIMSS) 3:687^t, 4:653
- Third International Mathematics and Science Study 2007 (TIMSS) 4:653
- Third International Mathematics and Science Study Repeat (TIMSS–R) 3:687^t, 4:652
- Thirteenth Amendment (U.S. Constitution) 1:691
- Thomists 6:112
- Thorndike, Edward L.
 - connectionism 5:479
 - early childhood curriculum 1:515
 - educational assessments 4:69
 - learning theories 5:258, 6:361
 - standardized tests 3:399
- three-parameter logistic (3PL) model 4:82, 4:82^f
- three-stratum model of intelligence 5:185
- tic disorders 2:582
- Tiebout sorting and competition theory 2:388–393
 - background information 2:388
 - capitalization 2:388
 - financial and nonfinancial inputs 2:389
 - general discussion 2:332
 - private school markets 2:390, 2:391
 - productive efficiency enhancements 2:391
 - residential location–school choice disconnection 2:391
 - residential mobility 2:388
- time–delay learning 2:671, 2:672^t
- time series analysis 7:474–481
 - autoregressive moving average (ARMA)/autoregressive integrated moving average (ARIMA) processes 7:9, 7:477
 - basic concepts
 - birth rate analysis example 7:474, 7:475^f, 7:477^f
 - enrollment rate analysis example 7:474, 7:475^f, 7:477^f
 - general discussion 7:474
 - bootstrap methods 7:49
 - educational applications 7:9
 - frequency viewpoint 7:479
 - longitudinal evaluations 3:759
 - modeling approaches 7:475, 7:477^f
 - model selection and estimation 7:478
 - multivariate time series 7:479
 - nonlinear models 7:480
 - objectives 7:474
 - prediction 7:479, 7:479^f
 - quasi-experimental research 7:128
 - state–space models 7:479
- Times Higher Education Supplement (THES) rankings 4:548^t, 4:549
- Timor-Leste 2:111^t
- TinkerPlots 8:177
- tobacco-use prevention
 - adolescent smoking consequences 1:549
 - educational interventions 1:549
 - tobacco industry involvement 1:551
- toddlers *see* infants and toddlers
- Togo
 - early childhood education and care (ECEC) 2:111^t
 - student achievement test performance 2:418^t
 - teacher hiring practices 2:509
- Tokelau 2:111^t
- Tonga 2:111^t
- tool-mediated action 5:493, 5:494^f
- Tourette's disorder 2:582
- Tower Hamlets (London) 6:293
- Trade Act (1897) 8:450–451
- Trade Adjustment Assistance Act 8:277
- trade union education 1:139–143
 - background information 1:139
 - career guidance 8:271
 - characteristics 1:139
 - general discussion 1:143
 - knowledge-based economies 1:142
 - learning conditions 1:141
 - see also* unions
- traditional learning model 1:405^t
- traditional tests 1:171
- training finance systems 8:489–495
 - background information 8:489
 - finance flows
 - conventional, fragmented training markets 8:489, 8:490^f
 - funding diversification
 - cost sharing 8:491, 8:492^f
 - funding augmentation 8:490, 8:492^f
 - general discussion 8:490
 - institutional income generation 8:491, 8:492^f
 - private training provisions 8:491, 8:492^f
 - innovative financing mechanisms
 - enterprise training 8:492
 - funding diversification 8:490, 8:492^f
 - general discussion 8:490
 - integrated, demand-driven training markets 8:494, 8:494^f
 - levy–grant schemes 8:492
 - payroll taxes 8:492
 - subsidy schemes
 - advantages/disadvantages 8:493
 - funding allocations 8:494
 - national training authorities (NTAs) 8:493
 - training funds 8:493
- training markets 8:496–502
 - basic concepts 8:496
 - cost–benefit factors 8:500
 - economic factors 8:497
 - general discussion 8:501
 - political factors 8:501
 - quality challenges
 - contracts 8:498
 - external inspections 8:499
 - noneconomic motivation 8:499–500
 - reputation effects 8:499

- Train to Gain program 8:456, 8:459
 transcripts 3:247
 transformational leadership 5:158–164
 characteristics
 general discussion 5:159
 instructional program management
 buffering functions 5:162
 general discussion 5:161
 instructional support 5:162
 monitoring operations 5:162
 staffing considerations 5:161
 organization design
 collaborative cultures 5:160
 community engagement 5:161
 general discussion 5:160
 restructuring considerations 5:161
 school networks 5:161
 people development practices
 general discussion 5:159
 individualized support 5:160
 intellectual stimulation 5:160
 leadership models 5:160
 setting directions
 general discussion 5:159
 group goal acceptance 5:159
 high-performance expectations 5:159
 shared vision 5:159
 comparison studies 5:4
 criticisms 5:162
 historical background 5:158
 influencing effects 5:158
 international trends 4:759
 research trends 4:742
 school improvement 5:138
 student learning impacts 5:45
 transformational learning 1:14
 transformative internalization and externalization 6:714
 transformative leadership 5:26–33
 benefits 5:31
 criticisms 5:31
 ethical leadership 5:26
 evolutionary background 5:29
 general discussion 5:26
 modern perspectives 5:30
 origins 5:26
 theoretical perspectives 5:27*t*
 transactional leadership 5:27, 5:27*t*
 transformative leadership 5:27*t*, 5:28
 transformative learning theory 5:28
 transforming leadership 5:27, 5:27*t*
 transformative learning theory 1:18–24
 basic concepts 1:19, 1:37, 5:28
 general discussion 1:18, 1:23
 instructional strategies
 critical self-reflection 1:22
 empowerment 1:22
 general discussion 1:21
 support factors 1:22
 theoretical perspectives
 connected knowing 1:20
 ecological view 1:21
 general discussion 1:20
 group and organizational transformation 1:21
 intuition, imagination, and nurturing soul 1:21
 social change 1:20
 transformed item difficulty (TID) technique 4:167
 transient tic disorder 2:582
 transition shock 7:504, 7:669
 transition-to-school concepts 2:120–125
 background information 2:120
 continuity considerations 2:122
 dynamic, relationships-based approach 2:123
 general discussion 2:124
 multilayered, multiyear approaches 2:122
 readiness and preparedness 2:120
 set school induction practices 2:121
 special needs students 2:855
 stakeholder involvement 2:124
 time-limited change events 2:121
 translated tests 4:166
 trans-level learning 1:490
 transnational education 4:508
 see also cross-border higher education (CBHE)
 transversal programme 4:673
 traumatic brain injury (TBI) 2:695–700
 affected mental functions 2:696*t*
 categories
 coordination of services
 individualized education plans (IEPs) 2:697
 multidisciplinary planning 2:697
 general discussion 2:695
 injury factors 2:695
 recovery factors
 continuous positive behavior supports 2:698
 general discussion 2:698
 post-injury development 2:698
 secondary attention-deficit hyperactivity disorder (SADHD) 2:698
 school readiness factors
 re-entry considerations 2:696
 teacher preparedness 2:697
 characteristics 2:695
 general discussion 2:699
 school-based services 2:855
 Trends in International Mathematics and Science Study (TIMSS)
 assessment accountability 3:379
 background information 1:453, 3:681, 4:45, 4:656
 Bahrain 5:645*t*
 decentralized decision-making 5:74
 educational attainment 5:51–52
 educational evaluation systems 3:523
 educational impacts 4:666
 educational outcomes
 developing countries 2:416, 2:417*t*
 gender differences 2:275
 social background 2:273–274
 Egypt 5:645*t*
 Finland 5:574
 funding considerations 4:664
 governance models 7:535
 image-based education research 6:501
 information and communication technology (ICT) surveys 8:22
 international/regional assessment programs
 Eastern Europe 3:475
 educational evaluation systems 3:523
 Hong Kong 3:438
 international data systems 2:184, 2:184*t*
 Latin America 3:483
 modeling approaches 4:89
 sub-Saharan Africa 3:449
 Iran 5:645*t*
 Jordan 5:645*t*
 Latin America 3:483
 Lebanon 5:644, 5:645*t*
 Lithuania 5:665
 mathematics education 3:280
 Morocco 5:645*t*
 and the NAEP 3:696
 needs assessments 3:790
 New Zealand 5:694
 Palestine 5:645*t*
 progress maps 3:373
 research studies 4:644
 Saudi Arabia 5:645*t*
 school effectiveness research (SER) 6:269–270
 science-education programs 3:664
 Singapore 8:83
 small-area estimation 7:415
 student achievement test performance 2:246, 3:693, 3:726
 Syria 5:645*t*
 Taiwan 5:839–840
 teachers' unions 2:516
 test adaptations 4:166
 Third International Mathematics and Science Study 1995 (TIMSS) 3:687*t*, 4:652
 Third International Mathematics and Science Study 2003 (TIMSS) 3:687*t*, 4:653
 Third International Mathematics and Science Study 2007 (TIMSS) 4:653
 Third International Mathematics and Science Study Repeat (TIMSS-R) 3:687*t*, 4:652
 Tunisia 5:645*t*
 United States 6:262, 6:303
 valid educational indicators 4:198
 Yemen 5:645*t*
 triadic attention 5:170

- triarchic theory of successful intelligence 4:2, 5:186
 tribal colleges 1:663
 trimmed mean 7:266
 Trinidad and Tobago
 early childhood education and care (ECEC) 2:111*t*
 higher education enrollment 1:702*f*, 1:703*f*
 historical background 1:682
 secondary education 1:679
 triple-helix systems 4:361, 5:104
 Troops to Teachers 2:498
 true-score theory 2:577
 trustworthiness 6:505–513
 general discussion 6:505, 6:512
 historical perspectives 6:506
 postmodern/poststructural perspectives 6:510
 postpositivist perspectives
 connoisseurship/critique 6:509
 ethical probity 6:509
 evidence 6:491
 general discussion 6:507
 realist postpositivism 6:508
 social constructivism 6:508
 social context 6:506
 social networks 6:706
 Truth 6:87
t-tests 7:101, 7:129, 7:337, 7:349, 7:429, 7:447*t*
 Tucker–Lewis index (TLI) 7:193
 Tukey box plot 7:157, 7:158*f*
 Tukey–Kramer test 7:314
 Tukey’s HSD test 7:34, 7:314, 7:315*t*
 Tunisia
 early childhood education and care (ECEC) 2:111*t*
 inequality–education relationship 2:273*f*
 information and communication technology (ICT)
 computer access at home 8:25*t*
 computer access in school 8:23*t*
 student achievement test performance 2:247*f*, 2:417*t*, 2:418*t*
 test score–social background relationship 2:273*f*
 Trends in International Mathematics and Science Study (TIMSS) 5:645*t*
 Turiel, Elliot 5:280
 Turkey
 accreditation processes 4:384–389
 age-related policies and programs 8:263*t*
 class size 2:492*f*
 early childhood education and care (ECEC) 2:4*t*, 2:111*t*
 earnings–education level relationship 2:307*t*
 education levels 2:279*f*
 higher education
 admissions tests 4:12
 foreign doctoral students 4:292*t*
 historical development 1:731
 institutional expenditures 1:733*t*
 loans and grants 4:501*f*, 4:502*f*
 parental expenditures 1:734*t*
 student and researcher mobility 4:287*f*
 tertiary education expenditures 4:500*f*
 tertiary type-B program survival rates 8:464*t*
 tuition fees 4:500*f*
 inequality–education relationship
 employment rate–educational attainment relationship 2:278*f*
 test score–social background relationship 2:273*f*
 information and communication technology (ICT)
 computer access at home 8:25*t*
 computer access in school 8:23*t*
 frequency of computer use in school and at home 8:26*t*
 student proficiency and confidence 8:31*t*
 information and communication technology (ICT) access and equity 8:97
 school supply shortages 2:506*f*
 student achievement test performance 2:247*f*, 2:417*t*
 teacher wages and salaries 2:496*f*
 teaching hours 2:493*f*
 Turkish Early Enrichment Program (TEEP) 2:130
 vocational education and training (VET) 8:251*t*, 8:263*t*
 women in employment 2:56*f*
 Turks and Caicos Islands 2:111*t*
 Tuskegee Experiment 1:557
 Tuvalu 2:111*t*
 twenty-first century skills 8:158–166
 challenges 8:164
 framework development 8:159
 innovative instructional strategies
 basic concepts 8:160
 emerging media 8:161
 historical documents 8:164*f*
 multi-user virtual environment (MUVE) interfaces 8:161
 River City MUVE 8:162, 8:162*f*, 8:163*f*, 8:164*f*, 8:165*f*
 situated learning 8:161
 virtual microscopy 8:164*f*, 8:165*f*
 knowledge-based economies 8:158
 selected new literacies 8:159, 8:160*t*
 twice-exceptional students 2:588–593
 characteristics
 attention deficit/hyperactivity disorders (ADHD) 2:590
 emotional and behavioral disorders (EBD) 2:590
 general discussion 2:589, 2:592
 learning disabilities 2:590
 educational programs 2:591
 identification strategies 2:588
 occurrences 2:588
 twinning programs 4:511
 twin studies 2:174, 2:300
 two-dimensional Brownian motion 7:452
 two-level linear model 7:210
 two-level structural equation models 7:455
 two-parameter logistic (2PL) model 4:82, 7:86, 7:91*t*
 two-stage purification 4:40
 two-way language learning programs 1:296, 1:296*f*
 two-year colleges 4:610
 Tyler rationale 1:53, 1:499, 3:506, 3:706, 4:69
 Type I/Type II statistical errors 7:32, 7:221, 7:221*f*, 7:312
 Type III statistical errors 7:315
- ## U
- ubuntu* 6:53
 Uganda
 academic performance 1:796*t*
 early childhood education and care (ECEC) 2:111*t*
 enrollment rates 1:796*f*, 1:796*t*
 higher education gender equity 1:630–631
 school fees 2:419
 student achievement test performance 2:418*t*
 warfare effects 1:795
 Ukraine
 early childhood education and care (ECEC) 2:111*t*
 information and communication technology (ICT) 8:25*t*
ummab (communal engagement) 6:70
 unbalanced bilingualism 5:356
 uncertainty analyses *see* decision theory
 unconfounded assignment mechanisms 7:68
 undergraduate settings 6:728
 unemployed workers 8:275
 see also displaced workers
 UNESCO (United Nations Educational, Scientific and Cultural Organization)
 academic freedom 4:328
 accreditation processes 4:384–389, 4:537
 adult education 1:2, 1:81
 background information 4:689, 6:249
 community education 1:122
 Cuban education system 1:300–301
 early childhood education and care (ECEC) 2:1, 2:98, 2:111, 2:111*t*, 2:114*t*
 educational evaluation policy models 3:645–652
 Keynesian welfare state model 3:645, 3:646*t*
 knowledge capitalist model 3:646*t*, 3:649
 neoliberal era 3:646*t*, 3:647
 purpose and scope 3:645
 Education for All (EFA)
 expansion efforts 6:250
 goals 6:249
 indigenous curricula 1:307
 International Institute for Educational Planning (IIEP) 4:690
 Solomon Islands 5:797
 Suriname 5:826–827
 Uganda 1:795
 universal primary education (UPE) 6:250
 Education for Sustainable Development (ESD) 6:289
 elementary education 7:725–726
 global information societies 4:363
 higher education
 access and equity patterns 4:334
 accreditation processes 4:537

UNESCO (United Nations Educational, Scientific and Cultural Organization)

(continued)

- collective bargaining 4:491
- cross-border education 4:509
- enrollment predictions 1:700, 1:702*f*, 1:703*f*
- globalization effects 1:698, 4:601
- global mission 4:597–602
 - background information 4:597
 - capacity-building/quality assurance initiatives 4:600
 - cross-border higher education (CBHE) 4:600
 - degree mills 4:600
 - Global Forum on International Quality Assurance, Accreditation and the Recognition of Qualifications 4:599
 - Global Initiative for Quality Assurance Capacity (GIQAC) 4:600
 - policy debates 4:598
 - standard setting 4:599
 - UNESCO Portal 4:600
- graduate employment 4:359
- massification (1950–1980) 1:731
- state steering models 4:557–558
- universal enrollment 1:735
- university associations 4:487
- World Conference on Higher Education 1:698, 4:601
- indigenous knowledge 1:661
- informal learning 8:242–243
- information and communication technology (ICT)
 - ICT Competency Standards for Teachers (ICT–CST) 8:213
 - literacy skills development 8:213
- International Association for the Evaluation of Educational Achievement (IEA) 4:663
- international evaluations 3:681
- Iraq 1:797
- lifelong learning 1:89, 6:170
- National University of Educational Planning and Administration (NUEPA, India) 4:676
- nonformal education 1:42
- older learners 1:217
- Organisation for Economic Cooperation and Development (OECD) 4:588
- origins 6:530
- peace education 1:851, 6:22
- preservice education programs 8:217
- research-quality assessments 4:308, 4:311
- scholarly publications 1:251
- school-based evaluation 3:716
- science and engineering university degrees 4:322
- secondary education initiatives 6:250
- self-managing schools 5:75
- student mobility 4:561
- textbooks 1:397
- university adult and continuing education (ACE) 1:183
- vocational education and training (VET)
 - planning and policy development 8:470
 - research areas 8:353
- World Council of Comparative Education Societies (WCCES) 1:266
 - see also* International Institute for Educational Planning (IIEP)
- unfolding analysis 7:309, 7:310*f*, 7:310*t*
- Uniform Crime Reports (UCR) 2:242
- uniform differential item functioning (DIF) 4:37
- uniform distribution
 - continuous probability distributions 7:101
 - discrete probability distributions 7:134
- uninformed professionalism 1:365–366
- unions
 - career guidance 8:271
 - higher education collective bargaining 4:491
 - skill shortages 8:321
 - teachers' unions 2:512–517
 - collective bargaining laws 2:513
 - district-wide impacts
 - general discussion 2:515
 - union–school district spending relationship 2:516
 - union–student achievement relationship 2:516
 - union–student/teacher ratio relationship 2:515
 - union–wage relationship 2:515
 - union–working conditions relationship 2:515
 - general discussion 2:516
 - historical background 2:512
 - influential factors 2:512
 - negative impacts
 - community welfare reductions 2:514
 - democratic process threats 2:515
 - general discussion 2:514
 - lack of trust considerations 2:514
 - performance incentives reductions 2:514
 - resource allocation reductions 2:514
 - positive impacts
 - efficiency-enhancing resources 2:514
 - information flow improvements 2:514
 - market failure reduction 2:514
 - stable school districts 2:514
 - teacher quality 2:513
 - time management 2:514
 - union impacts
 - district-wide impacts 2:515
 - general discussion 2:513
 - negative impacts 2:514
 - positive impacts 2:513
- union education 1:139–143
 - background information 1:139
 - characteristics 1:139
 - general discussion 1:143
 - knowledge-based economies 1:142
 - learning conditions 1:141
- union representation 2:513
 - see also* labor unions
- United Arab Emirates 2:111*t*
- United Kingdom
 - accreditation processes 4:384–389
- African-descendant populations
 - cultural alienation and annihilation 1:575, 1:737
 - culture of exclusion 1:577, 1:742
 - historical perspectives 1:573
- anti-racist education 1:739
- Catholic schools 2:356
- childhood brain development 2:57*f*
- class size 2:492*f*, 6:200
- common schools versus separate schools 6:44
- comparative education societies 1:268
- crime–education relationship 1:879
- early childhood education and care (ECEC) 2:157–161
 - Effective Provision of Preschool Education (EPPE) 2:78, 2:86–92, 3:320–321, 5:53
- enrollment rates 2:4*t*, 2:56*t*, 2:111*t*, 2:114*t*
- government policies and programs
 - contextual perspectives 2:157
 - implications 2:161
 - integrated services 2:159
 - legislative actions 2:158
 - multiagency and transdisciplinary teams 2:159
 - support services and initiatives 2:157
- legislative actions
 - childcare programs 2:158
 - educational entitlements 2:158
 - reform efforts 2:158
- multiagency and transdisciplinary teams
 - effectiveness challenges 2:160
 - provisions and settings 2:159
 - workforce reform 2:159
- parental involvement 2:154
- Sure Start program 2:57, 2:128
- earnings–education level relationship 2:307*t*
- education levels 2:279*f*
- enrollment rates 2:493*f*
- ethnic minorities
 - access and equity patterns 5:52
 - African-Caribbean population
 - access and equity patterns 5:52
 - alternative educational settings 1:742
 - anti-racist education 1:739–740
 - assimilation policies 1:575, 1:737
 - school-leaving impacts 1:742
 - suspensions and exclusions 1:577, 1:742
 - underachievement 1:575, 1:740, 1:839
 - educational reform 1:737–745
 - alternative educational settings 1:742
 - anti-racist education 1:739
 - assimilation policies 1:737
 - background information 1:737
 - core issues 1:740
 - cultural pluralism 1:737
 - differential performance outcomes 1:740
 - government policies and programs 1:743

- multicultural education 1:738
- national curriculum development 1:740
- post-war migration impacts 1:737
- school-leaving impacts 1:742
- suspensions and exclusions 1:742
- underachievement 1:740
- race/racism 1:737
- South Asian population
 - access and equity patterns 5:52
 - alternative educational settings 1:742
 - anti-racist education 1:739–740
 - assimilation policies 1:737
 - multicultural education 1:738
 - suspensions and exclusions 1:742
 - sustainable development reform 6:293
 - underachievement 1:740
- sustainable development reform 6:293
- faith-based education
 - background information 1:272
 - ethnic minorities 1:742
 - recent trends and conflicts 1:272
- General Certificate of Secondary Education (GCSE) 3:386
- higher education
 - admissions tests 4:12
 - bachelor's degrees 5:553
 - binary systems 4:618
 - collective bargaining 4:493
 - doctorates 5:554
 - enrollment rates 1:673
 - ethnic minorities 1:673–678
 - academic preparation 1:674
 - campus experiences 1:676
 - diversity considerations 1:673
 - institutional dynamics 1:675
 - parental involvement 1:676
 - peer influences 1:676
 - poverty concerns 1:675
 - primary and secondary education 1:674
 - foreign doctoral students 4:292*t*
 - gender equity 1:689
 - historical development 1:731
 - income contingent loans (ICLs) 4:505
 - institutional expenditures 1:733*t*
 - loans and grants 4:501*f*; 4:502*f*; 5:554
 - master's degrees 5:554
 - parental expenditures 1:734*t*
 - professional development 4:411
 - professional qualifications 5:554
 - quality assurance 5:550
 - reporting and regulation practices 4:275*t*
 - structural and operational characteristics 5:553
 - student and researcher mobility 4:287*f*; 4:561
 - student experiences 4:475
 - tertiary education expenditures 4:500*f*
 - tertiary type-B program survival rates 8:464*t*
 - tuition fees 4:500*f*
 - university associations and subgroups 4:486
- inequality—education relationship
 - earnings differentials 2:277*t*
 - employment rate—educational attainment relationship 2:278*f*
 - fathers' educational background 2:274*f*
 - gender differences 2:276*f*; 2:277*t*
 - test score—social background relationship 2:273*f*
- information and communication technology (ICT)
 - computer access at home 8:25*t*
 - computer access in school 8:23*t*
- leadership development programs 4:753*t*; 4:754*t*; 4:755*t*; 4:758*t*
- Learning How to Learn (LHTL) project
 - background information 3:464
 - challenges
 - classroom environments 3:464
 - collaborative classroom learning 3:466
 - fidelity and flexibility considerations 3:465
 - teachers' and students' roles 3:465
 - teachers' assessment practices 3:466*t*
 - importance 3:469
 - supportive school policies and practices 3:466*t*; 3:467; 3:469*f*
 - teachers' professional learning practices 3:466; 3:466*t*
- London Borough of Tower Hamlets 6:293
- mathematics education 3:282
- national assessment programs 4:92
- Open University 1:176; 1:178; 4:398–399; 4:405; 4:420; 4:425; 4:430
- paraprofessionals
 - general discussion 2:781; 2:786
 - higher-level teaching assistant (HLTA) standards 2:781; 2:782*t*
 - instructional assignments 2:784; 2:785*t*
 - national occupational standards (NOS) 2:781*t*
 - number of teachers and teaching assistants 2:781*t*
 - training programs 2:782; 2:783*t*
- participation research
 - regular education 8:284*f*
 - vocational education and training (VET) 8:283*f*; 8:284*f*
 - work-related vocational education and training (VET) 8:284*f*; 8:285*f*
- personalized learning 6:231; 6:231*t*; 8:393
- pharmacy education 3:59
- physically disabled students 2:520
- prison education programs 8:301; 8:301*t*; 8:304
- records of achievement
 - background information 3:244
 - formative assessments 3:245
 - lessons learned 3:246
 - National Record of Achievement (NRA) 3:237; 3:245
 - personal development planning and progress file 3:245
- research-quality assessments 4:307; 4:309; 4:310
- returns to education—education level 2:308*t*
- school effectiveness research (SER) 6:269
- school resources—earnings relationship 2:315
- school workforce reform 6:242–248
 - basic principles
 - performance leadership emphasis 6:243
 - qualified teacher status (QTS) 6:242
 - remodelling project 6:242
 - standards 6:242
 - current trends 6:242
 - justification
 - functionalist position 6:244
 - general discussion 6:244
 - socially critical position 6:245
 - outcomes analyses 6:246
- student achievement test performance 2:247*f*; 2:417*t*
- student attrition 4:467
- teacher age distribution 2:492*f*
- teacher education programs 7:527
- teacher evaluations 3:677
- textbook publishing 1:370
- training markets 8:496
- vocational education and training (VET)
 - age-related policies and programs 8:263*t*; 8:265
 - apprenticeships 8:309
 - career guidance 8:272
 - Economic and Social Science Research Council (ESSRC)-funded Teaching and Learning Research Programme (TLRP) 8:353
 - employer involvement 8:455–462
 - centers of vocational excellence (CoVEs) 8:459
 - challenges 8:461
 - curriculum development 8:459
 - general discussion 8:455
 - government policies 8:455
 - industrial training boards (ITBs) 8:457
 - national skills academies (NSAs) 8:461
 - national skills strategy 8:458
 - pilot programs 8:456
 - public funding 8:458
 - sectoral approaches 8:457
 - sector skills agreements (SSAs) 8:457–458
 - sector skills councils (SSCs) 8:457
 - skills training demand incentives 8:456
 - skills training supply 8:456
 - skill supply reform 8:459; 8:460*t*
 - vocational qualification reform efforts 8:458
 - initial vocational preparation system 8:251*t*
 - knowledge and skills specifications 8:482
 - modularization 8:383; 8:386
 - National Vocational Qualifications (NVQs) 2:159; 8:386; 8:398; 8:407; 8:476
 - partial completion research 8:464
 - recognition of prior learning (RPL) 8:397; 8:398
 - school-based programs 8:247
 - school-to-work transitions 8:309
 - special needs students 8:294–299
 - teacher training 8:507
 - terminology and jargon sources 8:347

- United Kingdom (*continued*)
 - vocational–academic education parity 8:434
 - women in employment 2:56^f
- United Nations
 - background information 4:688
 - education initiatives 6:249
 - globalization efforts 6:529
 - peace education
 - culture of peace movement 1:852
 - historical background 1:850
 - human rights 1:852
 - international law 1:851
 - UNESCO (United Nations Educational, Scientific and Cultural Organization) 1:851, 6:22
- United Nations Children's Fund (UNICEF) 4:691
- United Nations Conference on Environment and Development (UNCED) 1:750
- United Nations Convention on the Rights of Indigenous People 8:290
- United Nations Convention on the Rights of the Child 1:852, 3:378, 3:381, 6:530
- United Nations International Covenant on Economic, Social, and Cultural Rights 2:880
- United Nations Permanent Forum on Indigenous Issues 1:663
- United Nations Universal Declaration of Human Rights 1:288, 1:851, 2:880, 4:688, 6:249, 6:529
- United States 5:888–895
 - accreditation processes 4:384–389
 - admissions tests
 - fairness 4:12
 - predictive validity
 - grade point average (GPA) 4:10
 - personal characteristics measures 4:11
 - primary admissions tests
 - ACT Assessment 4:8
 - general discussion 4:7
 - Graduate Management Admission Test (GMAT) 4:9
 - Graduate Record Examinations (GRE) 4:8
 - Law School Admission Test (LSAT) 1:695, 4:9
 - Medical College Admission Test (MCAT) 4:9
 - SAT examinations 4:8
 - usage
 - graduate/professional school programs 4:10
 - undergraduate institutions 4:9
 - adult basic education (ABE) 1:102
 - African-descendant populations 1:573
 - background information 5:888
 - childhood brain development 2:57^f
 - class size 2:492^f, 6:200
 - comparative education societies 1:266
 - demographic diversity 1:582
 - desegregation 2:394–401
 - displaced worker assistance programs 8:277
 - early childhood education and care (ECEC)
 - Early Head Start program 2:128
 - Early Learning Standards (ELS) 2:132–137
 - background information 2:133
 - content resources 2:134
 - development process 2:133
 - document content 2:134
 - general discussion 2:132
 - historical perspectives 2:132
 - implementation processes 2:135
 - importance 2:136
 - infant–toddler content 2:135
 - preschool content 2:135
 - enrollment rates 2:4_t, 2:56_t, 2:111_t, 2:114_t
 - entitlements 2:114_t
 - Healthy Families America (HFA) program 2:126
 - Incredible Years program 2:129
 - integrated services 2:105
 - Nurse–Family Partnership (NFP) program 2:127
- earnings differentials 2:289
- earnings–education level relationship 2:307_t
- educational vouchers 2:338–339, 2:370, 2:381, 2:382_t
- education levels 2:279^f
- education system
 - administrative and supervisory structure 5:891
 - civil rights movement 6:263
 - contemporary assessments 5:891
 - degree-granting institutions 5:892_t
 - demographic data 5:893_t
 - demographic diversity 1:582
 - educational finance 5:891
 - emerging issues 5:894
 - enrollment 5:892_t
 - goals 5:888
 - higher education 5:892, 5:892_t, 5:893_t
 - historical shifts 5:888, 6:261
 - large-scale assessments
 - historical perspectives 6:303
 - research findings 6:305
 - quality concerns 6:303
 - reform efforts 6:261–268
 - civil rights movement 6:263
 - funding considerations 6:266
 - future policies and programs 6:265
 - general discussion 6:261
 - global impact 6:267
 - historical perspectives 6:261, 6:303
 - No Child Left Behind Act (2001) 6:265
 - organizational and cultural barriers 6:263
 - school effectiveness research (SER) 6:269
 - structural and operational characteristics
 - adult education 5:893
 - alternative schooling arrangements 5:889
 - early childhood education and care (ECEC) 5:889
 - elementary education 5:889
 - nonformal education 5:893
 - preschool education 5:889
 - schematic diagram 5:890^f
 - secondary education 5:889
 - teaching and professional staff 5:889
 - whole-school designs
 - educational implications 6:307
 - implementation fidelity/reliability 6:306
 - large-scale assessments 6:303
 - research findings 6:305
 - research implications 6:308
 - sustainability 6:307
- enrollment rates 2:493^f
- ethical practices 4:717–718, 4:720
- faith-based education
 - background information 1:272
 - recent trends and conflicts 1:272
- higher education
 - binary systems 4:618–619
 - collective bargaining 4:492
 - ethnic minorities 1:799
 - faculty mobility 4:562
 - federal political systems 4:604
 - foreign doctoral students 4:292_t
 - gender equity 1:689
 - gender gap research 1:799
 - historical development 1:730
 - income contingent loans (ICLs) 4:505
 - institutional expenditures 1:733_t
 - loans and grants 1:654^f, 4:501^f
 - parental expenditures 1:734_t
 - professional development 4:411
 - student and researcher mobility 4:287^f, 4:561
 - tertiary education expenditures 4:500^f
 - tuition fees 1:654^f, 4:500^f
 - university associations and subgroups 4:486
- indigenous peoples
 - demographic data 1:664_t
 - educational status 1:664
 - historical policies and practices 1:662
 - self-determination 1:663
 - tribal colleges 1:663
- inequality–education relationship
 - earnings differentials 2:277_t
 - employment rate–educational attainment relationship 2:278^f
 - gender differences 2:276^f, 2:277_t
 - immigrant status 2:275^f
 - test score–social background relationship 2:273^f
- information and communication technology (ICT)
 - computer access at home 8:25_t
 - computer access in school 8:23_t
 - home access 8:95, 8:96_t
 - home use 8:95
 - outcomes analyses 8:95, 8:97
 - school access 8:96, 8:96^f
 - school use 8:96
- inquiry-oriented teacher education 7:606

- Internet-based education 8:103
- leadership development programs 4:753*t*, 4:754*t*, 4:755*t*, 4:758*t*
- mathematics curriculum development 1:456
- moral and social reasoning 5:282
- music teacher education 7:707
- national assessment programs 4:94
- personalized learning 8:394
- pharmacy education
 - certification and licensing 3:59
 - current issues
 - curricular delivery methods 3:63
 - experiential coursework 3:62
 - outcomes assessments 3:63
 - preprofessional coursework 3:62
 - professionalism 3:63
 - residency education 3:63
 - general discussion 3:59
 - medication therapy management (MTM) 3:61
 - professional code of conduct 3:60
 - responsibilities and services 3:60
 - signature pedagogy 3:62
- physical education and sports 1:459
- physically disabled students 2:520
- popular adult education 1:165
- principal evaluation 3:699
- prison education programs 8:301, 8:301*t*, 8:302
- research-quality assessments 4:307–308
- returns to education–education level 2:308*t*
- school resources–earnings relationship 2:314, 2:315*f*
- self-managing schools 5:72
- student achievement test performance 2:247*f*, 2:417*t*
- student attrition 4:467–468
- teacher education programs
 - program curriculum 7:528
 - requirements 7:527
- teacher incentives 2:483
- teacher training 2:498–503
 - academic achievement outcomes 2:499
 - background information 2:498
 - current licensing requirements 2:499
 - effectiveness variations 2:502
 - empirical research 2:499
- teacher wages and salaries 2:496*f*
- teaching hours 2:493*f*
- textbook publishing 1:370
- training markets 8:496
- vocational education and training (VET) 8:247, 8:251*t*, 8:252*f*, 8:253*f*, 8:515
- women in employment 2:56*f*
- United States Agency for International Development (USAID) 1:796
- univariate *d* 7:445
- univariate linear regression 7:482–488
 - Akaike Information Criterion (AIC) 7:487, 7:487*t*
 - basic concepts 7:482
 - diagnostic tools
 - general discussion 7:485
 - residuals versus fitted values 7:485*f*
 - residuals versus leverage 7:486*f*
 - standardized residuals 7:486*f*
 - fitted regression models 7:483
 - inference estimations 7:484, 7:484*t*
 - information criteria 7:487
 - model comparisons 7:486, 7:487*t*
 - regression coefficients 7:484*t*
 - software packages 7:487
- univariate normal distributions 7:467–473
 - basic concepts 7:178
 - basic properties 7:469
 - bell curve theory 7:471, 7:472*f*
 - bell image 7:468*f*
 - binomial to normal distribution 7:469, 7:470*f*, 7:471*f*
 - central limit theorem 7:469
 - cumulative distribution function (CDF) 7:468, 7:468*f*
 - empirical rule 7:468*f*
 - general discussion 7:467
 - historical background 7:467
 - illustrative example 7:470, 7:471*f*
 - inference estimations 7:179
 - mean and variance 7:468*f*
 - occurrences 7:470
 - probability density function (PDF) 7:467, 7:468*f*
 - standard normal distribution 7:468, 7:469*f*
 - statistical applications 7:473
- Universal Declaration of Human Rights (UDHR) 1:288, 1:851, 2:880, 4:688, 6:249, 6:529
- universal design for learning (UDL) 8:119–124
 - background information 8:119
 - basic principles 8:120
 - future directions 8:123
 - UDL guidelines
 - general discussion 8:120
 - multiple means of action and expression
 - executive function options 8:122
 - expressive skills and fluency options 8:121
 - physical action options 8:121
 - multiple means of engagement
 - recruiting interest options 8:122
 - self-regulation options 8:123
 - sustained attention and effort options 8:122
 - multiple means of representation
 - comprehension options 8:121
 - language and symbols options 8:121
 - perception options 8:120
 - universal design practices 8:119
- universal primary education (UPE) 6:250
- universal Truth 6:87
- universities
 - academic freedom 4:328–333
 - basic concepts 4:328
 - freedom of academics 4:330, 4:332
 - historical background
 - freedom of academics 4:330
 - general discussion 4:329
 - institutional autonomy 4:329
 - restrictions 4:330
 - student freedom 4:331
 - institutional autonomy
 - basic concepts 4:331
 - historical background 4:329
 - postconflict societies 4:519
 - restrictions 4:330, 4:331
 - student freedom 4:331, 4:332
 - academic mobility 4:559–565
 - future directions 4:564
 - government policies and programs 4:559
 - historical background 4:559
 - internationalization processes 4:542
 - professional impacts 4:563
 - statistical data
 - faculty mobility 4:562
 - general discussion 4:560
 - student mobility 4:561
 - student perspectives 4:562
 - access and equity patterns 1:636–648
 - academic preparation 1:637
 - college knowledge 1:637
 - demographic changes 1:636
 - financial considerations 1:637
 - geographical information systems (GIS)
 - advantages 1:638
 - characteristics 1:637
 - college knowledge 1:645*f*, 1:646*f*
 - data sources and methods 1:638
 - illustrative examples 1:638
 - racial demographics 1:643*f*, 1:644*f*
 - research implications 1:647
 - Tucson elementary schools 1:639*f*, 1:643*f*, 1:644*f*
 - Tucson high schools 1:641*f*, 1:645*f*, 1:646*f*
 - Tucson middle schools 1:640*f*
 - international access patterns 4:334
 - labor market 1:636
 - social stratification 1:637
 - Tucson, Arizona
 - college knowledge 1:645*f*, 1:646*f*
 - elementary schools 1:639*f*, 1:643*f*, 1:644*f*
 - geographical information systems (GIS) research 1:638
 - high schools 1:641*f*, 1:645*f*, 1:646*f*
 - middle schools 1:640*f*
 - racial demographics 1:643*f*, 1:644*f*
 - research implications 1:647
 - research methodologies 1:638
- admissions tests 1:845–849, 4:7–14
 - background information 4:7

- universities (*continued*)
 - changing contexts 1:845
 - international institutions 4:12
 - noncognitive variables
 - characteristics 1:845, 1:846*r*
 - community involvement 1:846*r*, 1:847
 - long-range goals 1:846*r*, 1:847
 - Noncognitive Questionnaire (NCQ) 1:847, 4:11–12
 - nontraditional knowledge acquisition 1:846*r*, 1:847
 - positive self-concept 1:846, 1:846*r*
 - racial discrimination 1:846, 1:846*r*
 - realistic self-appraisal 1:846, 1:846*r*
 - successful leadership experiences 1:846*r*, 1:847
 - support services 1:846*r*, 1:847
 - Psychometric Entrance Test (PET) 4:13
 - purpose 1:845
 - SAT examinations 1:845
 - SweSAT 4:12–13
 - United States
 - ACT Assessment 4:8
 - fairness 4:12
 - grade point average (GPA) 4:10
 - Graduate Management Admission Test (GMAT) 4:9
 - graduate/professional school programs 4:10
 - Graduate Record Examinations (GRE) 4:8
 - Law School Admission Test (LSAT) 4:9
 - Medical College Admission Test (MCAT) 4:9
 - personal characteristics measures 4:11
 - predictive validity 4:10
 - primary admissions tests 4:7
 - SAT examinations 4:8
 - undergraduate institutions 4:9
 - usage 4:9
- admittance pathways 4:609–615
 - articulation 4:612
 - bachelor's programs 4:609
 - entry qualifications 4:610*r*, 4:611*r*
 - postgraduate programs 4:611
 - promotion strategies 4:613
 - second chance students 4:611
 - student transfers 4:610
- age-participation rates (APRs) 4:404
- Asian American/Pacific Islander student enrollment 1:596, 1:597*f*, 1:598*f*
- associations 4:485–490
 - consortia 4:487
 - future directions 4:489
 - international accreditation issues 4:534
 - international associations 4:487
 - national and regional subgroups 4:486
 - national associations 4:485
 - scope 4:485
 - theoretical analysis 4:489
- Australia 5:511*f*, 5:513
- Austria 5:520
- Bangladesh 5:531
- civic engagement 4:398–403
 - challenges 4:402
 - community partnerships 4:390–397
 - background information 4:390
 - higher-education drivers 4:392
 - higher-education governance barriers 4:395
 - national policy barriers 4:393
 - regional development drivers 4:390
 - regional governance barriers 4:395
 - science-driven models 4:392–393
 - first-order engagement 4:400
 - general discussion 4:398
 - governance structures 4:401
 - Hegelian perspective 4:399
 - historical background 4:398
 - second-order engagement 4:400
 - stakeholders 4:401
 - third-order engagement 4:400
- collective bargaining 4:491–498
 - basic concepts 4:491
 - challenges 4:496
 - country analysis
 - Argentina 4:496
 - Australia 4:496
 - Brazil 4:496
 - Cameroon 4:494
 - Chile 4:496
 - China 4:495
 - East Africa 4:493
 - France 4:493
 - general discussion 4:492
 - Germany 4:493
 - India 4:494
 - Indonesia 4:495
 - Japan 4:495
 - Malaysia 4:495
 - Mexico 4:496
 - Nigeria 4:494
 - Russia 4:493
 - South Africa 4:494
 - South Korea 4:494
 - Sweden 4:493
 - United Kingdom 4:493
 - United States 4:492
 - West Africa 4:493
 - international standards 4:491
 - collegiality 4:252, 4:266, 4:434
 - community partnerships 4:390–397
 - background information 4:390
 - higher-education drivers 4:392
 - higher-education governance barriers 4:395
 - national policy barriers 4:393
 - regional development drivers 4:390
 - regional governance barriers 4:395
 - science-driven models 4:392–393
 - computer-assisted instruction 4:419–426
 - background information 4:419
 - curriculum 4:420
 - effective learning
 - collaborative learning 4:422
 - communication-based learning 4:422
 - construction-based learning 4:421
 - instructional strategies 4:421
 - e-learning pedagogies 4:423, 4:423*r*
 - flexible study 4:423
 - future directions 4:425
 - teaching approaches 4:424
 - technology enhanced learning (TEL) 4:423
 - cross-border higher education (CBHE) 4:408
 - diversification considerations 4:347–353
 - classifications 4:347
 - diversity discourse 1:823
 - influencing factors 4:352
 - institution types 4:348
 - key features 4:347
 - rankings systems 4:350
 - reputational diversification 4:350
 - study program levels 4:349
 - theoretical perspectives 4:350
- diversity discourse 1:823–826
 - background information 1:823
 - discrimination 1:826–826
 - feminist perspectives 1:824
 - interdisciplinary studies 1:824
 - legal contexts 1:823
 - social contexts 1:823
 - Supreme Court decisions 1:823
- dual-mode institutions 4:406
- educational innovations 4:448–454
 - classroom assessment 4:450
 - contextual perspectives 4:448
 - curriculum
 - design considerations 4:452
 - general discussion 4:448
 - outcomes analyses 4:451
 - individual versus systemic approaches 4:449
 - information and communication technology (ICT) 4:453
 - student approaches to learning 4:451
 - student learning impacts 4:450
 - teaching and curriculum changes 4:448
 - teaching practices 4:450
 - typologies 4:449
 - whole-degree program 4:452
- educational reform 1:698–704
 - diversification guidelines 1:700
 - globalization effects 1:698
 - goals and objectives 1:701–702

- lifelong learning 1:699
- public university funding cuts 1:701, 1:707
- World Conference on Higher Education 1:698
- effective learning 4:461–466
 - background information 4:461
 - computer-assisted instruction
 - collaborative learning 4:422
 - communication-based learning 4:422
 - construction-based learning 4:421
 - instructional strategies 4:421
 - Self-Regulated Learning (SRL) 4:461, 4:462
 - Student Approaches to Learning (SAL) 4:461, 4:462
 - student engagement 4:464
 - teaching guidelines 4:465
- engaged education 1:711–718
 - engagement concepts 1:711, 1:712f
 - experiential learning 1:713, 1:714f
 - importance 1:716
 - intensive field experiences
 - basic concepts 1:714
 - curriculum development and logistics 1:716, 1:717r
 - goals and objectives 1:715
 - social and cultural change 1:714
- engineering profession 3:48
- English-delivered instruction 4:203
- Estonia 5:556
- federal political systems 4:603–608
 - Australia 4:606
 - Canada 4:604
 - characteristics 4:603
 - general discussion 4:608
 - Germany 4:605
 - Mexico 4:607
 - Switzerland 4:606
 - United States 4:604
- Fiji 5:566
- financial considerations 4:499–506, 4:573–583
 - characteristics
 - living costs 4:499
 - loans and grants 4:500, 4:502f
 - public support 4:500, 4:501f
 - tertiary education expenditures 4:499, 4:500f
 - tuition fees 4:499, 4:500f
 - economic impacts
 - externalities 4:575
 - general discussion 4:573
 - rate of return estimates 4:574
 - educational production function model 4:576
 - funding strategies 4:579
 - general discussion 4:499, 4:506, 4:573
 - historical development 4:221
 - income contingent loans (ICLs)
 - Australia 4:504
 - background information 4:504
 - New Zealand 4:504
 - South Africa 4:505
 - Sweden 4:504
 - United Kingdom 4:505
 - United States 4:505
 - Yale Plan 4:504
 - information access and use 4:577
 - organizational structures 4:576
 - student loans
 - conventional loans 4:503
 - cost-benefit analysis 4:503
 - importance 4:501
 - income contingent loans (ICLs) 4:503
 - international reforms 4:504
- Finland 5:570f, 5:572
- France 5:578
- Gambia 5:861
- Germany 5:585f, 5:587, 5:588, 5:589r
- Ghana 5:592r, 5:594
- globalization 4:201–208
 - Cambodia 4:205
 - China 4:205
 - contemporary globalization 4:202
 - educational impacts 4:206
 - global English 4:203
 - historical background 4:201
 - India 4:205
 - medium of instruction (MOI) 4:202, 4:205
 - graduate employment 4:354–359
 - differentiation considerations 4:356
 - division of labor 4:357
 - employability enhancements 4:356
 - employment rates and earnings 4:354
 - expansion effects 4:354, 4:365
 - globalization 4:358
 - job quality 4:355
 - knowledge society theory 4:358
 - labor market 4:366
 - research biases 4:359
 - student demands 4:354
 - transition periods 4:357
 - worker productivity 4:355
 - historical development
 - globalization 4:201
 - governance 1:728–736
 - Catholic Church 1:728
 - entrepreneurial university model 1:734, 4:249
 - evaluative-based transformations 1:732
 - globalization 1:734
 - institutional expenditures 1:733r
 - massification (1950–1980) 1:731
 - medieval university 1:728
 - modern universities (1800–1950) 1:730
 - parental expenditures 1:734r
 - Protestant Reformation period 1:729
 - quality assurance 1:728–729, 1:731, 1:734
 - triangle of coordination 1:728, 1:729f
 - tuition and fees 1:733–734
 - universal enrollment 1:735
 - theoretical perspectives 4:222
 - human resource issues 4:527–532
 - academic mobility 4:531
 - advancement and promotion opportunities 4:529
 - challenges 4:527
 - general discussion 4:527, 4:531
 - part-time and temporary appointments 4:528
 - recruitment and hiring 4:528
 - workplace and employment trends 4:530
 - Iceland 5:607f, 5:608
 - indigenous education 1:663
 - Indonesia 5:612f, 5:613, 5:615, 5:615r
 - institutional autonomy 4:341–346
 - background information 4:341
 - civic engagement 4:399
 - definitions 4:341
 - external boundaries 4:342
 - general discussion 4:345
 - internal boundaries 4:343
 - institutional management 4:264–272
 - administration 4:247
 - changing contexts
 - general discussion 4:264
 - market environment 4:265
 - new public management (NPM) 4:264
 - policy environment 4:264
 - councils or boards
 - chairperson 4:277
 - changing roles 4:275
 - composition 4:276
 - department heads 4:229–237
 - complex roles 4:230
 - contractual obligations 4:230, 4:231r, 4:232r
 - effective leadership behaviors 4:234r
 - excellence criteria 4:233r
 - future challenges 4:235
 - heads-of-school framework 4:232r
 - importance 4:229
 - priority duties 4:231r
 - specific leadership needs 4:234r
 - staff expectations 4:231
 - training and development 4:234, 4:235r, 4:236r
 - transitions to leadership 4:233r
 - general discussion 4:264, 4:270
 - governance 4:273–278
 - changing contexts 4:273
 - collegiality 4:252, 4:266
 - councils or boards 4:275
 - emerging trends 4:273

- universities (*continued*)
 - enrollment growth 4:273
 - entrepreneurial models 4:266
 - financial control models 4:274*t*
 - general discussion 4:277
 - governing boards 4:267
 - institutional control models 4:274*t*
 - institutional structures 4:246, 4:267, 4:401
 - new public management (NPM) 4:265
 - reporting and regulation practices 4:275*t*
 - standards/codes of practice 4:277
 - state steering models 4:554–558
 - institutional head
 - as change leader 4:270
 - functional role 4:267
 - influencing factors 4:268
 - leadership structures 4:246
 - market environment
 - competition 4:265
 - globalization 4:265
 - knowledge-based economies 4:223, 4:265
 - Israel 5:621, 5:622*t*
 - Jamaica 5:627, 5:628*f*
 - knowledge society 4:360–364
 - general discussion 4:364
 - graduate employment 4:358
 - intellectual property rights 4:362
 - internationalization processes 4:540–541
 - knowledge production processes 4:361
 - political goals
 - general discussion 4:362
 - global information societies 4:363
 - nation-states level 4:363
 - regional level 4:363
 - social transformation 4:358, 4:373
 - labor market relationship 4:365–369
 - access and equity patterns 1:636
 - expansion effects 4:365
 - general discussion 4:365
 - graduate employment 4:366
 - institutional characteristics 4:367
 - Mexico
 - demographic structure 1:720, 1:721*t*
 - economic dynamics 1:720
 - employment trends 1:722–724, 1:722*t*
 - future directions 1:726
 - industrial structure 1:722, 1:722*t*, 1:723*t*
 - social change impacts 1:719
 - university–industry collaboration 1:724
 - qualitative aspects 4:366
 - quantitative aspects 4:365
 - research background 4:368
 - Latvia 5:632, 5:633*t*
 - learning cities and regions 1:147
 - Lebanon 5:641
 - Lithuania 5:660
 - Malawi 5:671
 - Malta 5:676*f*, 5:678
 - management practices 4:245–250, 4:251–255
 - adaptive university model 4:249
 - background information 4:245
 - business-like practices 4:253, 4:266
 - changing contexts
 - competition 4:252, 4:265
 - financial considerations 4:252
 - institutional growth 4:251–252
 - collegiality 4:252, 4:266
 - entrepreneurial university model 4:249
 - institutional management 4:264–272
 - administration 4:247
 - changing contexts 4:264
 - general discussion 4:264, 4:270
 - governance 4:246, 4:265, 4:273–278, 4:401
 - institutional head 4:267
 - leadership structures 4:246
 - key features
 - administration 4:247
 - governance structures 4:246, 4:401
 - leadership structures 4:246
 - learning university model 4:249
 - new public management (NPM)
 - background information 4:251
 - basic concepts 4:248
 - collegiality 4:266
 - current trends 4:264
 - entrepreneurial models 4:266
 - evaluation policies and practices 3:648
 - governance 4:265
 - governing boards 4:267
 - ideological perspectives 4:251
 - institutional autonomy 4:342
 - institutional structures 4:267
 - strategic planning 4:256, 4:257*f*
 - organizational culture 4:253
 - professionalization 4:248–249
 - strategic management 4:248
 - university leaders 4:238–244
 - diversity considerations 4:238
 - future research areas 4:242
 - historical perspectives 4:238
 - literature reviews 4:238
 - literature types 4:239
 - presidential leadership 4:239, 4:240
 - senior management teams 4:239, 4:241
- mega-universities 4:406
- Mexico
 - collective bargaining 4:496
 - federal political systems 4:607
 - loans and grants 4:501*f*, 4:502*f*
 - social change 1:719–727
 - demographic structure 1:720, 1:721*t*
 - economic dynamics 1:720
 - employment trends 1:722–724, 1:722*t*
 - future directions 1:726
 - industrial structure 1:722, 1:722*t*, 1:723*t*
 - labor-market perspective 1:719
 - university–industry collaboration 1:724
 - tertiary education expenditures 4:500*f*
 - tertiary type-B program survival rates 8:464*t*
- Mongolia 5:682*t*, 5:683
- Namibia 5:687
- national science policy 4:294–299
 - general discussion 4:299
 - governance processes 4:295
 - historical background 4:294
 - institutional structures 4:297
 - policies and programs 4:295
 - research challenges 4:298
 - state–university relations 4:296
- New Zealand 5:695
- Norway 5:711*f*, 5:712*t*
- Pakistan 5:731
- Philippines 5:868*t*, 5:869
- Poland 5:739
- postconflict societies 4:514–519
 - basic concepts 4:514
 - functional role 4:515
 - future directions 4:519
 - intervention strategies
 - Afghanistan 4:518
 - general discussion 4:516
 - Iraq 4:518
 - South Africa 4:516
 - south-southeast Europe (SEE) 4:517
 - postconflict requirements 4:516
- professional autonomy 3:48, 4:434
- professional education development 4:411–418
 - challenges
 - accredited programs 4:415
 - aging demographics 4:414
 - curricula vocationalization 4:414
 - new technologies 4:414
 - structural instabilities 4:415
 - theoretical perspectives 4:415
 - contextual perspective and background 4:411
 - educational approaches
 - background information 4:412
 - learner-focused approaches 4:413
 - organization-focused approaches 4:413
 - sector-focused approaches 4:413
 - teacher-focused approaches 4:413
 - educational models 4:412

- future directions
 - distributed roles 4:416
 - general discussion 4:415
 - international and national collaboration 4:416
 - key performance indicators 4:415
 - professionalization 4:416
- general discussion 4:416
- terminology and meaning 4:411
- publication productivity 4:455–460
- age-related considerations
 - age decrement hypothesis 4:458
 - cumulative disadvantage hypothesis 4:458
 - general discussion 4:458
 - intellectual deadlock hypothesis 4:458
 - obsolescence hypothesis 4:458
 - seniority burden hypothesis 4:458
 - utility-maximizing hypothesis 4:458
- current trends 4:455
- gender differences
 - basic concepts 4:459
 - motherhood 4:459
 - networking and collaboration 4:459
- general discussion 4:459
- measurement approaches 4:455
- output analyses
 - ability measures 4:456
 - collaboration 4:457
 - communication networks 4:457
 - cumulative advantages and disadvantages 4:457
 - location 4:457
 - resource availability 4:457
 - statistical data 4:456
 - time and energy 4:456
- Qatar 5:747
- rankings systems 4:546–553
 - background information 4:546
 - general discussion 4:552
 - global university rankings
 - general discussion 4:548
 - Newsweek* rankings 4:551
 - performance factors 4:549, 4:550^t
 - Shanghai Jiao Tong University Institute of Higher Education (SJTJTHE) rankings 4:204, 4:548, 4:548^t, 4:550^t
 - Times Higher Education Supplement (THES) rankings 4:548^t, 4:549
 - holistic reputational rankings 4:551
 - national rankings
 - Centre for Higher Education Development (CHE) rankings 4:547
 - geographic distribution 4:546
 - U.S. News & World Report (USNWR)* rankings 2:231, 4:547
- research methodologies 4:551
- research commercialization 4:300–306
 - background information 4:300
 - copyright 4:303
 - criticisms 4:304
 - economic development
 - general discussion 4:302
 - incubators 4:303
 - research parks 4:302
 - start-up companies 4:302
 - technology transfer 4:303
 - government policies and programs 4:300
 - infrastructure 4:303
 - patents
 - background information 4:301
 - exceptions 4:301
 - personnel coverage 4:301
 - royalties 4:301
- research funding 4:279–285
 - background information 4:279
 - commercialization efforts 4:284
 - government policies and programs 4:281
 - priority setting 4:282
 - public funding 4:281
 - quality assessments 4:283
 - research characteristics and functions 4:280
 - university–industry collaboration 4:283
- research management 4:316–321
 - background information 4:316
 - emerging professional networks 4:318
 - functional role 4:318
 - general discussion 4:320
- management changes
 - collaborative research 4:317
 - competition and selectivity 4:316–317
 - external research contracts 4:317–318
 - financial transparency 4:317
 - general discussion 4:316
 - increased accountability 4:316
 - knowledge transfer 4:317
- Russia 5:754
- Rwanda 5:762, 5:762^t
- school–higher education institution (HEI) partnerships 7:757–763
 - characteristics 7:757
 - collaborative action research 7:759
 - general discussion 7:762
 - Innovative Links Project 7:758
 - National Schools Network 7:758
 - networked learning communities (NLC) 7:759
- partnership models
 - accountability 7:761
 - change processes 7:762
 - cooperation and collaboration 7:761
 - features and consequences 7:760^t
 - general discussion 7:760
 - improvement strategies 7:762
 - leadership considerations 7:761
 - sustainability 7:761
- preservice education programs 7:758
- professional development schools 7:758
- school-level development 7:758
- supervision 7:619
- vocational education and training (VET) 8:249
- Scotland 5:772
- Singapore 5:778
- Solomon Islands 5:794
- South Africa 5:802, 5:802^t
- Spain 5:809^t, 5:810
- Sri Lanka 5:816, 5:822
- strategic planning 4:256–263
 - basic model
 - deliberate versus emergent strategies 4:258, 4:258^f
 - empirical analyses 4:257
 - general discussion 4:256
 - new public management (NPM) approach 4:257^f
 - planning cycle 4:257^f
 - plan outline 4:258^f
 - current trends 4:256
 - goal and performance cards 4:262, 4:262^f
 - goals and objectives 4:256, 4:257^f
 - internal contract management 4:261
 - methodologies and instrumentation
 - general discussion 4:260
 - portfolio analyses 4:261, 4:262^f
 - strengths, weaknesses, opportunities, and threats (SWOT) analyses 4:261
 - performance areas 4:257
 - strategic perspectives 4:259, 4:260^f
- student athletes 1:860
 - academic achievement outcomes 1:860
 - Coach Carter* film screening 1:865
 - critical intervention programs 1:861
 - incentive and reward systems 1:864
 - Scholar–Baller[®] program 1:862
 - status symbols 1:861, 1:863
- student attrition 4:467–472
 - cost estimation 4:471
 - estimation methods 4:467
 - general discussion 4:467
 - impacts 4:471
 - retention/intervention strategies 4:470
 - theoretical perspectives 4:468
- student learning outcomes 4:479–484
 - benefits 4:483
 - cognitive complexity 4:479, 4:480^f
 - humanitarianism 4:480^f, 4:481
 - interpersonal and intrapersonal competence 4:480^f, 4:482
 - knowledge acquisition 4:480, 4:480^f
 - practical competence 4:480^f, 4:482
 - research background 4:479
- student support services 4:427–432
 - academic support
 - current perspectives 4:428
 - delivery approaches 4:428

- universities (*continued*)
- historical background 4:427
 - basic concepts 4:427
 - contextual perspectives 4:427
 - distance education 4:430
 - future directions 4:431
 - information literacy
 - delivery approaches 4:429
 - historical background 4:428
 - importance 4:428
 - libraries/learning spaces
 - current trends 4:429
 - historical background 4:430
 - Learning Commons 4:430
 - Sultanate of Oman 5:723
 - Suriname 5:830
 - Taiwan 5:836f, 5:837, 5:837t
 - Tanzania 5:843
 - United States 5:892, 5:892t, 5:893t
 - vocational-academic education parity 8:433
 - Zanzibar 5:898, 5:898t
 - see also* distance education; higher education; tertiary education
- Universities Council for Adult Education (UCAE) 1:186
- university adult and continuing education (ACE) 1:183–190
- background information 1:183
 - extramural model 1:185
 - global trends 1:184, 1:186
 - international patterns 1:185
 - structural tensions 1:186
 - student demographics 1:183
 - see also* lifelong learning
- University Correspondence College 4:430
- University Council for Educational Administration (UCEA) 4:681, 4:721
- university curricula
- globalization 1:403–410
 - background information 1:403
 - curriculum design 1:404
 - general discussion 1:409
 - knowledge-based economies 1:403
 - lifelong learning model 1:404, 1:405t
 - reform efforts
 - assessment methods 1:407
 - classroom dynamics 1:408
 - course content 1:406
 - critical perspectives 1:408
 - institutional change 1:406
 - instructional strategies 1:407
 - research and development expenditure levels 1:404t
 - traditional learning model 1:405t
- University Mobility in Asia and the Pacific (UMAP) program 4:540
- University of Chester management development guidelines 4:235t, 4:236t
- University of Chicago Laboratory School 1:355
- University of Newcastle 4:232t
- University of New England 4:431
- University of the Republic (UDELAR) 5:874, 5:882t, 5:883t, 5:884
- University of the Third Age (U3A) 1:178
- University of Utrecht School 6:451–452
- unscheduled castes 1:590, 1:590t
- unschooling *see* home schooling
- untouchables 1:590, 1:590t
- Urban and Suburban/Rural Special Strategies for Educating Disadvantaged Students 6:304
- urban environments 4:809–816
- challenges 4:810
 - defining characteristics 4:811
 - demographic background 4:809
 - general discussion 4:809
 - leadership strategies 4:815
 - local perspectives
 - estate schooling 4:814, 4:814f
 - extended area communities 4:814
 - Leadership on the Front-Line project 4:811
 - multiple communities 4:812, 4:813f
 - single/homogenous communities 4:812, 4:812f, 4:813f
- Urban Teacher Academy Program (UTAP) 2:476
- Uruguay 5:873–887
- background information 5:873
 - early childhood education and care (ECEC) 2:111t
 - earnings-education level relationship 2:306t
 - education system
 - assessment system development 3:483
 - building infrastructure 5:884, 5:886f
 - goals and objectives 5:873, 5:874t
 - information and communication technology (ICT) 5:884, 5:885f
 - monitoring and evaluation programs 5:882
 - new institutions 5:884, 5:886f
 - performance assessments
 - attendance rate 5:878f
 - dropout rates 5:876, 5:879f
 - enrollment 5:876, 5:877t
 - net schooling rate 5:878f
 - population literacy and schooling 5:876, 5:876f
 - Programme for International Student Assessment (PISA) 5:879, 5:880t, 5:881t
 - reading and mathematics trends 5:881t
 - repeating rates 5:876, 5:878f, 5:879f
 - schooling rate by educational cycle 5:876, 5:877f
 - young adults 5:876
 - public expenditures
 - financing sources 5:884t
 - general discussion 5:881
 - total expenditures 5:883f, 5:884f
 - reform efforts 5:884
 - research programs 5:881
 - structural and operational characteristics 5:874, 5:875f
 - teaching profession
 - general discussion 5:880
 - nonuniversity institutes 5:882t
 - teacher characteristics 5:883t
 - teacher education centers 5:881t
 - university programs 5:882t
 - geography 5:874f
 - higher education enrollment 1:702f, 1:703f
 - information and communication technology (ICT)
 - computer access in school 8:23t
 - frequency of computer use in school and at home 8:26t
 - student proficiency and confidence 8:31t
 - population distribution 5:874f
 - public subsidization 2:308t
 - returns to education-education level 2:308t
 - school supply shortages 2:506f
 - student achievement test performance 2:247f
 - teacher training 2:509
- US Air Force Academy (Colorado Springs, CO) 3:7
- see also* military officer education
- user norms 4:97
- US Military Academy (Westpoint, NY) 3:7
- see also* military officer education
- US Naval Academy (Annapolis, MD) 3:7
- see also* military officer education
- U.S. News & World Report (USNWR) rankings 2:231, 4:379, 4:547
- Utah 1:654f
- utilitarianism 6:105, 6:510
- Uzbekistan 2:111t

V

- valence 6:544
- validation arrangements 4:511
- validity 4:181–185, 6:505–513
- alternative conceptualizations 4:184
 - assessment accommodations 2:532, 2:533f
 - assessment conferences 3:408
 - automated essay scoring (AES) 4:22
 - basic concepts 3:158
 - Campbell-Fiske multitrait-multimethod model 3:111
 - causal effects measurements 2:187
 - classroom assessment 3:217
 - concurrent validity 4:34, 4:181–182
 - consequential validity 6:510, 7:515
 - construct validity 2:578, 3:159, 4:170, 4:181–182, 7:84
 - contemporary conceptualizations 4:182
 - content validity 2:578, 3:158, 4:32, 4:181
 - convergent validity 3:111
 - criterion validity 2:578, 3:159, 4:181–182
 - educational assessments 4:47
 - evidence
 - content-based evidence 4:183
 - internal structure 4:183
 - relations to other variables 4:183
 - response processes 4:183

- testing consequences 4:183
- evolutionary background 4:181
- external validity 6:521–527
 - causal effects measurements 2:187
 - future outlook 6:525
 - generalizability theory
 - basic concepts 6:522
 - main questions 6:522
 - pragmatic justification principles
 - general discussion 6:524
 - principle of causal explanation 6:525
 - principle of discriminant validity 6:525
 - principle of empirical interpolation and extrapolation 6:525
 - principle of heterogeneous irrelevancies 6:524
 - principle of proximal similarity 6:524
 - research background 6:521
 - traditional justification approaches
 - analog models 6:523
 - causal inference 6:522
 - meta-analysis 6:524
 - purposive sampling 6:523
 - statistical sampling theory 6:522
 - technology transfer models 6:523
- face validity 2:578, 3:158
- general discussion 4:181, 4:184, 6:505, 6:512
- high-stakes test score inflation 4:186–192
 - evidence 4:186, 4:187f, 4:187r
 - historical background 4:187
 - inadequate tradition approaches 4:187
 - validity of inferences under high-stakes conditions (VIHS)
 - behavioral responses 4:189
 - conceptual elements 4:190, 4:190f
 - effective test weight 4:189
 - evaluation frameworks 4:188
 - future directions 4:191
 - general discussion 4:188
 - target of inference 4:189, 4:189f
- historical perspectives 6:506
- internal validity 3:90–96
 - background information 3:91
 - causal effects measurements 2:187
 - causal studies 7:126
 - design considerations
 - analysis of covariance (ANCOVA) 3:94
 - analysis of variance (ANOVA) 3:94
 - cross-lag panel design 3:95, 3:95f
 - key features 3:93
 - pretest–posttest control group designs 3:94
 - regression continuity design 3:95
 - Solomon four-group design 3:94
 - random assignment approximations 3:97, 3:107
 - relational studies 7:126
 - statistical adjustments 3:95
 - threats
 - additive/interactive threats 3:93
 - ambiguous temporal precedence 3:91
 - attrition 3:92
 - field experimentation 3:85, 3:86r
 - general discussion 3:91
 - history 3:92
 - instrumentation 3:93
 - maturational 3:92
 - oral language skill tests 3:290
 - regression to the mean 3:92
 - selection 3:91
 - testing 3:93
- oral language skill tests 3:289
- postmodern/poststructural perspectives 6:510
- postpositivist perspectives
 - connoisseurship/critique 6:509
 - ethical probity 6:509
 - evidence 6:491
 - general discussion 6:507
 - realist postpositivism 6:508
 - social constructivism 6:508
- predictive validity 4:10, 4:170, 4:173, 4:181
- prospective teacher evaluation
 - characteristics 7:514
 - consequential validity 7:515
 - constructs and assessment task criteria 7:514
 - extended interpretations 7:514
 - reliability 7:515
 - qualitative inquiry 6:496
 - reliability–validity relationship 3:157, 4:30
 - science-education programs 3:664
 - social context 6:506
 - social studies/civic education program assessments 3:668
 - standards 4:181
 - statistical conclusion validity 3:147–151
 - definitions 3:147
 - importance 3:150
 - meta-analysis 3:150
 - threats
 - assumption violations 3:148
 - effect size estimation inaccuracies 3:150
 - error rates 3:148
 - extraneous variance 3:149
 - fixing 3:148
 - general discussion 3:147
 - heterogeneity of respondents 3:150
 - low statistical power 3:147
 - measurement unreliability 3:149
 - range restrictions 3:149
 - treatment implementation unreliability 3:149
 - Technical Recommendations for Psychological Tests and Diagnostic Techniques* 4:181–182
 - testing theories 2:578
 - Test Standards* 4:181
 - value-added models (VAMs) 7:489–494
 - advantages 7:490
 - background information 7:489
 - bias
 - achievement-related persistence/fadeout 7:493
 - achievement-related scaling 7:493
 - general discussion 7:492
 - nonrandom assignment 7:492
 - causal effects 7:489
 - characteristics 7:8
 - educational production function (EPF) 2:405
 - general discussion 7:493
 - institutional effectiveness measures
 - applications 3:178
 - basic concepts 3:173
 - research results 3:177, 3:177f
 - modeling approaches
 - ad hoc* models 7:491
 - econometric models 7:490
 - hierarchical linear models 7:491
 - statistical models 7:491
 - precision concerns 7:493
 - value systems
 - authentic leadership
 - individual valuation processes 5:7, 5:8f
 - valuation processes 5:7
 - characteristics 7:750r
 - Van Tassel–Baska Integrated Curriculum Model 2:832
 - Venezuela
 - academic values and expectations 4:438
 - assessment system development 3:483
 - comparative education societies 1:270
 - early childhood education and care (ECEC) 2:111r
 - earnings–education level relationship 2:306r
 - higher education enrollment 1:702f, 1:703f
 - public subsidization 2:308r
 - returns to education–education level 2:308r
 - verbal learning mediation 2:670
 - verification feedback 3:226
 - Vermont 1:654f, 3:385
 - veterinary education 3:71–76
 - academic requirements 3:71
 - accreditation organizations 3:72
 - certification and licensing 3:72
 - challenges 3:75
 - current issues
 - animals for teaching 3:75
 - clinical competency 3:74–75
 - theory–practice relationship 3:74
 - curriculum reform efforts 3:73
 - ethical practices 3:72–73
 - veterinary medicine profession 3:71
 - Vico, Giambattista 6:414
 - video conferences 3:534

- videogames 8:113, 8:115
- video taping 2:644
- Vietnam
 - academic values and expectations 4:438
 - early childhood education and care (ECEC) 2:111*r*
 - indigenous curricula 1:309
 - post-migration education 8:282
 - private higher education (PHE) 4:623
- violent crimes 2:239
- Virginia 1:654*f*
- virtual learning environments (VLEs) 4:424
- virtual networks 6:708
- virtual schools 4:511, 8:5, 8:129
- virtue ethicists 6:123
- Vischer, Friedrich 6:598
- visual art education
 - classroom-based assessments 3:265
 - content knowledge 7:697
 - globalization effects 1:423
 - pedagogical strategies 7:698
 - technology-supported learning environments 1:423
- visual-culture approach to art education (VCAE) 3:265
- visual data **6:498–504**
 - background information 6:498
 - current research contributions 6:501
 - research background
 - digital technology 6:500
 - field studies and experiments 6:499
 - Hine's photography 6:499
 - Houston's contribution 6:499
 - image-elicitation techniques 6:500
 - microethnography studies 6:500
 - significance 6:503
- visual impairment **2:880–885**
 - augmentative–alternative communication (AAC) **2:544–549**
 - characteristics
 - general discussion 2:545
 - manual communication 2:545, 2:624
 - technological devices 2:545
 - communication initiation
 - challenges 2:547
 - communication partners 2:548
 - use settings 2:548
 - vocabulary selection 2:548
 - vowel-based communication chart 2:549*f*
 - communication partners 2:547, 2:548
 - goals and objectives 2:544
 - needs assessment 2:546
 - deaf–blind students **2:621–627**
 - classroom intervention strategies 2:625
 - communication challenges 2:624
 - definitions 2:621
 - demographic data 2:623*r*
 - developmental challenges 2:622
 - early intervention services 2:622, 2:625
 - etiology 2:624, 2:624*r*
 - general discussion 2:626
 - hearing loss criteria 2:623
 - historical background 2:621
 - legal blindness criteria 2:623
 - transition processes 2:626
- educational needs 2:881
- etiologies
 - definitions 2:880
 - functional implications 2:881
 - prevalence 2:880
- general discussion 2:884
- government policies and programs 2:880
- instructional needs 2:883
- learning materials accessibility
 - auditory readers 2:883
 - Braille systems 2:883
 - closed circuit televisions (CCTVs) 2:882–883
 - general discussion 2:882
 - optical devices 2:882–883
- legal blindness criteria 2:623
- occurrences 2:880
- school-based services 2:854
- special education background 2:715
- visual literacy 8:37
- visual recognition memory 5:168
- vocabulary
 - adult learning 5:227
 - computer-assisted instruction **8:150–157**
 - comprehension 8:152
 - emergent literacy 8:150
 - fluency 8:151
 - general discussion 8:150, 8:155
 - instructional strategies 8:152
 - Internet access 8:154
 - phonics and phonemic awareness 8:151
 - software systems 8:3
 - text-to-speech applications 8:153
 - theoretical perspectives 8:154
 - predictors 5:351
 - primary schools 1:465
 - reading comprehension 5:397
 - special needs students 2:763
- vocal tic disorder 2:582
- Vocational Education Amendments (1976) 2:876
- vocational education and training (VET) **8:229–239, 8:513–518**
 - admittance pathways 4:610, 4:611*r*
 - adult basic education (ABE) **1:102–106**
 - future directions 1:105
 - general discussion 1:102
 - national developments 1:102
 - Swedish educational model/Komvux
 - background information 1:103
 - labor-market training programs 1:103
 - vocational-based learning (VBL) 1:104, 1:104*f*, 1:105*f*
 - age equality **8:262–268**
 - benefits 8:262
 - general discussion 8:262
 - government policies and programs 8:265
 - influencing factors 8:264
 - outcomes analyses 8:266
 - participation rates 8:263, 8:263*r*
 - policies and programs 8:267
 - assessment strategies **3:337–341**
 - assessment impacts 3:340
 - criterion-referenced assessments 3:338
 - formative assessments 3:339
 - general discussion 3:337, 3:341
 - general vocational education 3:337
 - hybrid assessment model 3:338
 - norm-referenced assessments 3:338
 - purpose 3:338
 - summative assessments 3:339
 - uncertainties 3:340
- Australia
 - dual apprenticeship system 8:452
 - employability skills **8:368–375**
 - employer involvement 8:457
 - lifelong learning model 1:157*r*, 1:160
 - National Centre for Vocational Education Research (NCVER) 8:354
 - partial completion research 8:464
 - prison education programs 8:303
 - recognition of prior learning (RPL) 8:397, 8:398
 - school-based programs 8:247
 - structural and operational characteristics 5:512
- Austria 5:519, 8:232
- Bangladesh 5:527
- Canada 8:397, 8:457, 8:464
- career guidance **8:269–274**
 - challenges 8:237–238
 - Denmark 8:271
 - employee development models 8:272
 - government policies and programs 8:269
 - human capital theory 8:269
 - Iceland 8:272
 - importance 8:272
 - mentoring strategies 8:271
 - outreach guidance 8:271
 - Scotland 8:272
 - standard formats 8:271
 - terminology 8:270
 - workplace guidance 8:271
- challenges 8:238
- Denmark 8:232, 8:435, 8:452
- displaced workers **8:275–279**
 - assistance programs
 - Austria 8:277

- evaluation practices 8:278
- France 8:278
- Germany 8:278
- retraining 8:277
- Spain 8:278
- Sweden 8:278
- United States 8:277
- causal factors 8:275
- characteristics 8:275
- losses 8:276
- occupational mobility 8:275
- wages and salaries 8:276
- dual apprenticeship system 8:448–454
 - Australia 8:452
 - background information 8:448
 - characteristics 8:232
 - Denmark 8:232, 8:435, 8:452
 - England 8:452
 - general discussion 8:453
 - Germany
 - characteristics 8:449, 8:449*t*
 - dual system–vocational training relationship 5:585, 5:585*f*, 5:586–587, 8:435, 8:451
 - general discussion 8:453
 - historical background 8:450
 - working principles 8:449
 - status considerations 8:435
 - terminology and jargon sources 8:347, 8:349*t*
- economic returns 8:255–261
 - human capital theory 8:255, 8:256*f*
 - international evidence
 - basic skills training 8:257
 - general discussion 8:257
 - post-secondary options 8:258
 - public training programs 8:258
 - wider benefits 8:259
 - work-related vocational education and training (VET) 8:259
- policy formation 8:255
- policy implications 8:259
- skill formation implications 8:255, 8:256*f*
- educational purpose 6:104
- employability skills 8:368–375
 - Australian Employability Skills Framework
 - attributes 8:369, 8:370*t*
 - skills 8:369
 - VET-sector employability skills 8:370*t*
- definitions 8:368
- development guidelines
 - assessment strategies 8:372
 - recording methods 8:372
 - skill inclusion qualifications 8:370
 - teaching strategies 8:371
- emerging trends
 - employer–learner mismatch 8:373
 - responsibility factors 8:373
 - skill requirements 8:373
 - general discussion 8:373
 - generic skills 8:368
 - target learners 8:370
- employer involvement 8:455–462
- England 5:553
- faculty
 - curriculum development 8:515
 - permanent versus casual faculty 8:516
 - professional development 8:508, 8:516
 - qualifications and certification 8:514
- Fiji 5:566
- financial considerations 8:234
- Finland 5:570*f*, 5:572
- France
 - age-related policies and programs 8:263*t*, 8:265–266
 - employer involvement 8:457
 - general discussion 5:578
 - lifelong learning model 1:157*t*, 1:158
 - participation research 8:283*f*, 8:284*f*, 8:285*f*
 - vocational–academic education parity 8:435
- Gambia 5:861
- gendered access 1:280
- general discussion 3:337, 8:482, 8:513, 8:517
- general vocational education 3:337
- Germany
 - age-related policies and programs 8:263*t*
 - dual apprenticeship system 8:448–454
 - background information 8:448
 - characteristics 8:232, 8:449, 8:449*t*
 - dual system–vocational training relationship 5:585, 5:585*f*, 5:586–587, 8:435, 8:451
 - general discussion 8:453
 - historical background 8:450
 - status considerations 8:435
 - working principles 8:449
 - employer involvement 8:457
 - Federal Institute for Vocational Education and Training (BIBB) 8:355
 - modularization 8:383, 8:386
 - social partnership model 1:157*t*, 1:159
- Ghana 5:590, 5:592*t*, 5:593
- globalization effects 8:411–415
 - characteristics 8:411
 - economic restructuring 8:411
 - educational impacts 8:414
 - higher-level skills development 8:414
 - international education markets 8:413
 - knowledge-based economies 8:412
 - mass movement/skills migration 8:412
 - transferable skills 8:413
- Greece 5:601
- historical background 8:513
- human capital theory 8:236–237
- Iceland 5:606
- immigrant education 8:280–288
 - immigrant characteristics 8:280
 - investment/noninvestment determinants 8:280
 - pay-offs 8:286
 - regular education participation 8:284*f*
 - research areas 8:287
 - targeting programs 8:285
 - VET participation research 8:282, 8:283*f*, 8:284*f*, 8:285*f*
- importance 7:744
- indigenous curricula 1:309
- indigenous peoples 8:289–293
 - challenges 8:235, 8:290
 - First Nations Technical Institute (Canada) 8:290*t*
 - instructional strategies 8:292
 - Māori culture 8:291*t*
 - outcomes analyses 8:293*t*
 - program models 8:290
 - Sámi culture 8:289*t*
 - socioeconomic development 8:289
- Indonesia 5:612*f*, 5:613
- informal learning 8:244, 8:438
- information resources 8:358–367
 - associations and networks 8:365, 8:365*t*
 - bibliographic databases 8:358, 8:359*t*
 - books 8:360
 - citation indices 8:360, 8:361*t*
 - conferences/conference papers 8:365, 8:365*t*
 - directories 8:358, 8:359*t*
 - general discussion 8:358, 8:367
 - glossaries/dictionaries/thesauri 8:366, 8:367*t*
 - government departments 8:363
 - journals 8:361, 8:363*t*
 - library catalogs 8:360, 8:362*t*
 - national research organizations 8:363, 8:364*t*
 - newspapers 8:362, 8:364*t*
 - research in progress 8:366
 - search engines 8:359, 8:360*t*
 - social software tools 8:366
 - statistical data sources 8:365, 8:366*t*
 - subject portals 8:360, 8:360*t*
- innovation 8:423–428
 - basic concepts 8:423
 - challenges 8:237
 - future directions 8:427
 - higher-level skills development 8:414
 - manufacturability 8:423
 - networking and collaboration 8:423
 - practical and theoretical learning approaches 8:427
 - shared knowledge 8:423–424
 - worked-based innovation 8:424
- international perspectives 8:229
- labor shortages 8:238
- Latvia 5:632

vocational education and training (VET) (*continued*)

learning styles **8:376–382**

definitions **8:376**

flexible learning

learning preferences **8:379**

readiness and responsiveness **8:379**

workplace learning **8:379**

models

cognition-centered models **8:377**

Honey and Mumford Learning Styles Questionnaire **8:377**

Kolb Learning Style **8:376–377**

research background **8:376**

research limitations **8:378**

research challenges **8:380**

VET learners

flexible learning **8:379**

implicit learning **8:379**

learner profiles **8:380**

nonformal learning **8:379**

research results **8:378**

situated learning **8:380**

tacit knowledge **8:379**

Lebanon **5:640, 5:640*t*, 5:641*t***

lifelong learning **1:157*t*, 1:160, 8:401, 8:470, 8:504**

Lithuania **5:660**

Malawi **5:669**

models

apprenticeships **8:232**

dual systems **8:232**

outcome-focused models **8:233**

qualifications **8:233**

quality assurance **8:233**

risk levels **8:233**

modularization **8:383–390**

background information **8:230, 8:383**

basic concepts

English model **8:386**

German model **8:386**

similarity–difference comparisons **8:387**

characteristics **8:385**

current trends

educational sector integration **8:388**

general discussion **8:387**

mainstream training **8:388**

reduced fragmentation **8:388**

supranatural development **8:388**

English versus German module comparisons **8:383**

Mongolia **5:682*t*, 5:683**

Namibia **5:687**

Netherlands

age-related policies and programs **8:263*t***

dual apprenticeship system **8:232, 8:435**

employer involvement **8:457, 8:461**

participation research **8:283*f*, 8:284*f*, 8:285*f***

vocational–academic education parity **8:435**

New Zealand **8:397, 8:398**

Niger **5:702, 5:704*t***

noncompletion **8:463–468**

causes

academic–social interactions **8:466**

family issues **8:467**

financial considerations **8:467**

general discussion **8:465**

student characteristics **8:466**

teaching quality **8:466**

definitions **8:463**

general discussion **8:232, 8:463**

measurement approaches

course characteristics **8:465*t***

group characteristics **8:464**

partial completion **8:464**

student characteristics **8:465*t***

tertiary type-B program survival rates **8:463, 8:464*t***

outcome improvement strategies **8:467**

Norway **5:711*f*, 5:712, 5:712*t*, 5:715*t***

Pakistan **5:732**

personalized learning **8:394**

planning and policy development **8:469–475**

basic principles **8:470**

development approaches

background information **8:469**

coherence **8:470**

globalization **8:470**

scale **8:469**

scope **8:469**

international perspectives **8:472**

research areas **8:474**

skills development programs **8:473**

theoretical frameworks **8:471**

postconflict societies **8:441–447**

Afghanistan

conflict background **8:444**

post-war society **8:444**

skills development programs **8:445**

background information **8:441**

challenges **8:446**

educational opportunities **8:441**

South Africa

apartheid–National Qualifications Framework (NQF) transition **8:443**

conflict background **8:442**

purpose **3:338**

Qatar **5:744**

qualifications **8:476–481**

contextual perspectives **8:480**

European Qualification Framework (EQF) **8:413, 8:479**

general discussion **8:476**

institutional approaches **8:479**

outcomes-based qualification framework

general discussion **8:233**

logical perspectives **8:479**

macro aspects **8:479**

micro aspects **8:480**

reform efforts **8:478**

typology **8:478**

reform efforts

general discussion **8:477**

government policies **8:477**

National Vocational Qualifications (NVQs) **2:159, 8:384, 8:386, 8:398, 8:476**

outcomes-based qualification framework **8:478**

transferable skills **8:413**

recognition of prior learning (RPL) **8:397–402**

background information **8:231–232, 8:397**

benefits **8:397**

competency-based learning methods **8:400**

constituents and drivers **8:398**

lifelong learning **8:401**

outcomes analyses **8:400**

philosophical perspectives **8:399**

terminology and jargon sources **8:349*t*, 8:350*t***

reform efforts **8:403**

research areas **8:352–357**

background information **8:352**

Community College Research Center (CCRC) **8:354**

current issues **8:352**

Economic and Social Science Research Council (ESRC)-funded Teaching and Learning Research Programme (TLRP) **8:353**

Federal Institute for Vocational Education and Training (BIBB) **8:355**

funding considerations **8:353**

general discussion **8:356**

National Centre for Vocational Education Research (NCVER) **8:354**

research impacts **8:355**

risk levels/quality assurance systems **8:482–488**

general discussion **8:482, 8:487**

knowledge and skills specifications **8:482**

national qualifications frameworks **8:483**

provider risks

assessment risks **8:485**

characteristics **8:483**

funding risks **8:485**

resources risks **8:484**

VET system **8:233, 8:486**

Russia **5:753**

Rwanda **5:761**

school-based programs **8:247–254**

access and equity patterns **8:252**

characteristics

curriculum **8:248**

higher education articulation **8:249**

initial vocational preparation system **8:250, 8:251*t***

level and focus **8:249**

work experience **8:249**

focus and scope **8:247**

innovation capacities **8:426, 8:426*t***

institutional arrangements **8:247**

- outcomes analyses 8:251, 8:252*f* 8:253*f*
- school-to-work transitions 8:306–311
 - challenges 8:237
 - key indicators 8:306
 - policy responses and reforms
 - apprenticeships 8:309
 - education retention rate improvements 8:307
 - general discussion 8:307
 - human capital theory 8:307
 - labor market policies 8:309
 - young adults 8:310
- Singapore 8:69
- Slovenia 5:786
- Solomon Islands 5:794, 5:796*t*
- South Africa 8:247, 8:397, 8:398, 8:444
- Spain 5:809, 5:809*t*
- special needs students 2:876, 8:294–299
 - challenges 8:294
 - future directions 8:298
 - international approaches
 - funding considerations 8:295
 - government policies and programs 8:294
 - outcomes analyses 8:297
 - participation research 8:296
 - support systems 8:296
- Sri Lanka 5:816, 5:822
- status considerations 8:429–440
 - basic concepts 8:429
 - basic criteria 8:438
 - European policy considerations
 - European Qualification Framework (EQF) 8:413, 8:437, 8:479
 - informal learning recognition 8:438
 - official agreements 8:435
 - vocational qualifications and status 8:437
 - reform strategies
 - enhancement 8:432
 - enrichment programs 8:432
 - linkages 8:431
 - qualifications frameworks 8:476–481
 - unification 8:431
 - vocational–academic education parity 8:431
 - research areas 8:438
 - socioeconomic status (SES) 8:429, 8:466
 - vocational–academic education parity
 - assessment strategies 3:338
 - European educational policies 8:434
 - general discussion 3:337, 8:430
 - higher education accessibility 8:433
 - reform strategies 8:431
 - teacher status and qualifications 3:337–338, 8:434, 8:504
 - workplace–educational institution cooperation 8:434
- structural characteristics 8:426, 8:426*t*
- sustainable development 8:416–422
 - challenges 8:237
 - economic sustainability 8:418
 - educational role 8:417
 - environmental sustainability 8:417, 8:418*f*, 8:419*f*
 - functional role 8:421
 - implications 8:416
 - social sustainability 8:419
- Switzerland 8:232
- Taiwan 5:836*f*, 5:837, 5:837*t*
- Tanzania 5:844
- target populations
 - displaced workers 8:236
 - general discussion 8:234
 - immigrant populations 8:235
 - indigenous peoples 8:235
 - older learners 8:236
 - prisoners 8:236
 - students with disabilities 8:235
 - women 8:236
- teaching and learning practices
 - competency model 8:230
 - completion rates 8:232
 - general discussion 8:230
 - modularization 8:230
 - personalized learning 8:231
 - recognition of prior learning (RPL) 8:231–232
 - workplace learning 8:231
- terminology and jargon sources 8:343–351
 - economics/labor markets 8:344, 8:345*t*
 - framework concepts 8:348, 8:349*t*
 - general discussion 8:343, 8:351
 - globalization 8:346, 8:346*t*
 - government policies and programs 8:343, 8:345*t*
 - learning culture concepts 8:348, 8:349*t*
 - pathway concepts 8:348, 8:349*t*
 - registered training organizations (RTOs) 8:348
 - skills development 8:346, 8:347*t*
 - systems theory 8:347, 8:349*t*
 - teaching–learning–assessment concepts 8:349, 8:350*t*
 - Thailand 5:853
 - United Kingdom
 - age-related policies and programs 8:263*t*, 8:265
 - Economic and Social Science Research Council (ESRC)-funded Teaching and Learning Research Programme (TLRP) 8:353
 - employer involvement 8:455–462
 - centers of vocational excellence (CoVEs) 8:459
 - challenges 8:461
 - curriculum development 8:459
 - general discussion 8:455
 - government policies 8:455
 - industrial training boards (ITBs) 8:457
 - national skills academies (NSAs) 8:461
 - national skills strategy 8:458
 - pilot programs 8:456
 - public funding 8:458
 - sectoral approaches 8:457
 - sector skills agreements (SSAs) 8:457–458
 - sector skills councils (SSCs) 8:457
 - skills training demand incentives 8:456
 - skills training supply 8:456
 - skill supply reform 8:459, 8:460*t*
 - vocational qualification reform efforts 8:458
 - modularization 8:383, 8:386
 - National Vocational Qualifications (NVQs) 2:159, 8:386, 8:398, 8:407, 8:476
 - partial completion research 8:464
 - recognition of prior learning (RPL) 8:397, 8:398
 - workforce issues 8:233–234
 - workplace-based assessments 3:330–336
 - active knowledge transformation 3:334
 - challenges 3:335
 - competence-based assessments 3:332
 - general discussion 3:330
 - initial education and training 3:330
 - knowledge development and transformation 3:333
 - knowledge-intensive workplaces 3:331
 - lifelong learning 3:334
 - reflective formative assessments 3:332
- Zanzibar 5:898
 - see also* adult education and training (AET); lifelong learning
- vocational interests 6:614
- Vocational Rehabilitation Act (1973) 2:634–635, 2:689
- vocational teacher education 7:744–748, 8:503–512
 - activity- and problem-oriented tasks 7:746
 - admission requirements 7:745
 - barriers 8:503
 - change agents 8:505*f*
 - characteristics 7:745, 8:514
 - continuing professional development 8:508
 - future challenges
 - change agents 8:505*f*, 8:510
 - learning communities 8:510
 - professional knowledge base 8:509
 - future directions 8:510
 - grading systems 7:747
 - importance 7:744
 - initial teacher training 8:507, 8:507*t*
 - laws and regulations 7:745
 - modeling approaches 7:744
 - multidisciplinary considerations 7:746
 - professional development 8:508, 8:516
 - recent trends 7:747
 - standardization/educational reforms 8:505, 8:506*f*
 - student teachers 7:747
 - teacher profiles 8:504
 - teacher quality 7:747
 - teacher status and qualifications 3:337–338, 8:233–234, 8:434, 8:504
 - teaching materials 7:746
 - training and curriculum 7:745
 - training locations 7:746

- vocational teacher education (*continued*)
 work culture transference 7:746
- voice disorders 2:655
- voice output communication devices (VOCAs) 2:545
- volitional control 6:724–731
 background information 6:724
 characteristics 6:724
 coping strategies 6:573
 effortful control
 academic work habits 6:726
 emotional regulation 6:725
 learning task management strategies 6:726
 research background 6:725
 future research areas 6:728
 intervention programs
 general discussion 6:727
 middle schools 6:727
 primary schools 6:727
 secondary schools 6:728
 undergraduate settings 6:728
 motivational regulation 6:632
- Volkshildung* 1:2, 1:162–163
- volunteer activities *see* civic engagement
- voting *see* civic engagement
- vouchers
 Catholic schools 2:355
 educational privatization 2:338
 income and ability sorting 2:370
 Latin America 2:370, 2:419–420
- Vygotsky, Lev 5:231–236
 activity theory 5:330, 7:637
 background information 5:231, 5:498
 cognitive development theory
 genetic analysis 5:231
 human mental functioning
 mediated actions 5:232, 5:499, 5:501
 social origins 5:233
 impacts 5:235
 infants and toddlers 2:72
 intermental processes 5:233
 research background 5:231
 constructivist theories 1:489, 5:390, 5:488
 cultural–historical activity theory (CHAT) 5:491–497, 5:498, 6:360
 cultural tools 5:499
 development theories 6:361
 early childhood cognition 2:72, 2:75, 5:331
 educational impacts 5:501
 instructional planning 3:367, 6:361
 intelligence theories 5:187
 language learning theory 5:232, 5:348, 5:499, 5:501
 learning theories 6:361
 motivation research 5:205, 6:712
 peer learning 5:331, 5:445, 6:637
 psychological development 5:245–246
 sociocultural theory 1:517, 3:167, 5:231, 5:390, 5:498–502, 6:360
 zone of proximal development (ZPD)
 cognitive development theory 5:234
 dynamic assessments 3:220–221
 early childhood curriculum 1:517
 instructional planning 3:367, 6:361
 language learning theory 5:500
 mastery theories 3:168
 motivation 6:713
 science learning 5:390
- W**
- wages and salaries 2:473, 2:481, 2:495*f*, 2:496, 2:496*f*
- Wales
 Catholic schools 2:356
 class size 6:200
 Incredible Years program 2:129
 leadership development programs 4:753*t*, 4:754*t*, 4:755*t*, 4:758*t*
 physical education and sports 1:459
 vocational education and training (VET) 8:434
see also United Kingdom
- Walton Family Foundation 2:383
- Wann, Kenneth 1:355
- warfare 1:793–798
 causal mechanisms
 biological theories 1:794
 general discussion 1:793
 psychological theories 1:793
 social scientific theories 1:794
 sociological theories 1:794
 stability and peace theories 1:794
 characteristics 8:441
 educational impacts
 future directions 1:797
 general discussion 1:795
 Iraq 1:797
 Uganda 1:795, 1:796*f*, 1:796*t*
 empowerment 1:793
 Warnock Report (1978) 3:493, 6:158
 washback studies 3:197, 3:440, 3:506
 Washington
 Asian American/Pacific Islander student enrollment 1:597*f*
 leadership development programs 4:753*t*, 4:755*t*, 4:758*t*
 tuition and needs-based grants 1:654*f*
- Washington Accord 4:209–215
 background information 3:47
 general discussion 4:214
 graduate attributes 4:210
 historical background 4:210
 impacts
 ABET accreditations 4:211
 assessments 4:213
 curriculum 4:212
 outcome-based accreditation 4:211
 pedagogy 4:212
 research initiatives 4:213
 professional competencies 4:210
 related international partnerships 4:210, 4:210*r*
 signatories 4:210*r*
- Washington Consensus 6:13–14
- wave feminist theory 6:57
- ways-of-coping questionnaire (WCQ) 3:257
- weak instruments 7:225
- Web 2.0 4:431, 8:6, 8:190, 8:194, 8:202
- Web-based Inquiry Science Environment (WISE) 1:470, 8:200
- Weber, Max 6:184
- webfolios *see* electronic portfolios
- websites 8:127
- Weibull model 7:343, 7:464, 7:465*f*, 7:465*t*
- weighted least squares (WLS) analysis 7:165
- weighted maximum likelihood (WML) method 7:397
- weighted mean 7:266
- welfare reform 1:91
- welfare states 1:238, 3:645
- well-being and happiness
 capabilities approach 6:194–199
 educational contributions
 general discussion 6:196
 justice 6:197
 pedagogy and curriculum design 6:198
 value and functions 6:197
 future research areas 6:198
 normative framework
 Nussbaum's perspective 6:195
 Sen's perspective 6:194
 philosophical perspectives 6:194
- children 2:32–37
 environmental factors
 climate change 2:34
 environmental toxins 2:34
 food consumption changes 2:35
 home living environment changes 2:35
 urban environment changes 2:35
- familial factors
 family structure 2:33
 maternal employment 2:33
 parental education 2:33
 parenting styles 2:33
 poverty 2:33
- general discussion 2:37
- influencing factors
 community factors 2:33
 environmental factors 2:34
 familial factors 2:33
 general discussion 2:32
 genetic factors 2:32
 social support 2:33
 societal effects 2:33

- long-term effects
 - cumulative effects 2:36
 - early experiences 2:35
 - latency/sleeper effects 2:35
 - pathway effects 2:36
- postmodern societies 2:52
- social change impacts 2:32
- societal effects
 - health considerations 2:33
 - social capital 2:34
 - social gradients 2:34
- educational goals 6:105
- well-defined problems 5:273
- Weschler Intelligence Scale for Children (WISC) 4:3
- West Africa 4:493
- West Virginia 1:654*f*, 2:65
- Wharton, Joseph 3:39
- wheelchair-bound students 2:643
- white-collar crimes 2:239
- White, John 6:109
- whiteness 6:36, 6:40
- white privilege 6:40
- White supremacy 6:342
- white universities 1:768, 4:479–480
- whole-school designs 6:303–309
 - middle schooling 6:236, 6:236*f*, 6:237*t*
 - social change impacts 6:303
 - United States
 - educational implications 6:307
 - education system characteristics 6:303
 - implementation fidelity/reliability 6:306
 - large-scale assessments
 - historical perspectives 6:303
 - research findings 6:305
 - reform efforts 6:303
 - research implications 6:308
 - sustainability 6:307
- WIDE World online courses 7:639
- wikis
 - curriculum inquiry 1:331
 - educational applications
 - science education 8:199
 - student support services 4:431
 - evaluation policies and practices 3:535
 - knowledge communities 8:190, 8:194
 - vocational education and training (VET) resources 8:366
- Wilcoxon rank-sum test 7:348
- Wilk's Λ 7:320, 7:327, 7:337, 7:447*t*
- Wilson, John 6:6
- WinBUGS model 7:456
- Winne and Hadwin's model of recursive self-regulation 5:504
- Winsorized mean 7:266
- Wisconsin 1:654*f*
- Wishart distribution 7:102, 7:320
- withdrawn children 6:692
- Wittgenstein, Ludwig 6:3, 6:14, 6:92, 6:110, 6:184–185
- Wollstonecraft, Mary 2:11, 6:57
- women in employment
 - childcare programs 2:55, 2:56*t*
 - early childhood education and care (ECEC)
 - child allowances 2:2
 - general discussion 2:2
 - leave policies 2:3
 - tax benefits 2:2–3
 - economic benefits 2:55
 - gender-equity gap 2:56*t*
 - percentage employed 2:56*f*
 - social change impacts 2:55
 - vocational education and training (VET) 8:236
 - working mothers 2:33, 2:55, 2:151
- Women, Infants, and Children (WIC) program 2:44, 2:45*t*
- women of color
 - admissions test fairness 4:12
 - childcare programs 2:15
 - educational attainment 1:799–805
 - completion rates 1:800, 1:803
 - dropout rates 1:803
 - earnings differentials 1:804
 - 18–24-year-olds demographics 1:802, 1:803*f*
 - enrollment rates 1:800, 1:801*f*, 1:802*f*, 1:803*f*
 - female–male participation rates 1:799, 1:804
 - gender gap research 1:799, 1:804
 - United States 1:799
 - ethnic and women's studies programs 1:625
 - racial and ethnic demographics 1:687
 - STEM (science, technology, engineering, and mathematics) 1:688
 - women's liberation movement 2:11
 - women's studies 1:623–628
 - autonomy concerns 1:625
 - general discussion 1:623, 1:627
 - historical background 1:623
 - institutional racism 1:626
 - interdisciplinary collaboration 1:625
 - scholarship 1:624
 - Woodcock Reading Mastery Tests 3:189
 - Woodson, Carter G. 6:351
 - Woods, Peter 6:484
 - word learning 5:350
 - work-based learning 1:404, 5:155, 7:683
 - Work Based Learning for Adults (WBLA) 8:264
 - Workers' Educational Association (WEA) 1:2
 - workers' rights movement 1:77
 - work habits 6:726
 - working-class populations 6:355
 - working memory (WM) capacity 5:186, 5:198–199, 5:225, 5:246, 5:342, 8:200
 - workplace-based assessments 3:330–336
 - active knowledge transformation 3:334
 - challenges 3:335
 - competence-based assessments 3:332
 - general discussion 3:330
 - initial education and training 3:330
 - knowledge development and transformation 3:333
 - knowledge-intensive workplaces 3:331
 - lifelong learning 3:334
 - reflective formative assessments 3:332
 - workplace learning 1:58–63, 7:569–574, 8:328–336
 - adult learning processes 1:40
 - apprenticeships 5:472
 - assessment strategies 3:334
 - background information 1:58, 8:231
 - basic concepts 1:59
 - career guidance 8:271
 - computer-assisted instruction 8:184
 - conceptual/procedural advances
 - general discussion 1:61
 - pedagogic aspects 1:61
 - self and agency learning 1:61
 - contextual levels 7:571
 - definitions 1:83
 - flexible learning
 - general discussion 8:379
 - learning preferences 8:379
 - readiness and responsiveness 8:379
 - formal–informal learning distinction 1:45
 - general discussion 7:569, 7:573
 - implications 8:334
 - informal learning 1:45, 8:243, 8:329, 8:329*f*
 - informal versus formal training and learning 8:329, 8:329*f*
 - knowledge-based economies 8:328
 - learning frameworks 1:62
 - on-the-job training
 - structured programs
 - advantages 8:331
 - basic concepts 8:332, 8:332*f*
 - general discussion 8:331
 - illustrative examples 8:333, 8:333*t*
 - research areas 8:333
 - unstructured programs 8:331
 - professional development programs 7:569
 - purpose 1:59
 - research trends 1:83–88
 - future directions 1:87
 - general discussion 1:83
 - identity and learning 1:85
 - language and literacy instruction 1:86
 - learning processes
 - community of practice (CoP) model 1:84
 - co-participation/co-emergence model 1:85
 - cultural–historical activity theory 1:85
 - general discussion 1:83
 - individual knowledge acquisition 1:83
 - levels of learning 1:84
 - network utility models 1:84
 - reflective practices 1:84

- workplace learning (*continued*)
 self-directed learning 1:84
 power and politics 1:86
 socioeconomic modernization 1:219
 taxonomic categories 8:330*t*
 theoretical perspectives 7:570
 workplace conditions
 critical factors 7:571
 feedback opportunities 7:572
 learning resources 7:573
 shared norms 7:572
 task complexity 7:571
 teacher participation and collaboration 7:572
 work process formalization 7:572
workplace mistreatment/abuse 4:790–796
 abuse effects 4:793
 abusive behavior typologies 4:791, 4:792
 basic concepts 4:790
 general discussion 4:790
 modeling approaches 4:791
 research approaches
 future directions 4:794
 methodologies 4:794
 published results 4:793
 theoretical perspectives 4:791
World Bank (WB)
 Caribbean countries 1:680–681
 early childhood education and care (ECEC) 2:98
 educational evaluation policy models 3:645–652
 Keynesian welfare state model 3:645, 3:646*t*
 knowledge capitalist model 3:646*t*, 3:649
 neoliberal era 3:646*t*, 3:647
 purpose and scope 3:645
 higher education
 access and equity patterns 4:334
 developing countries 4:590–596
 background information 4:590
 future directions 4:595
 lending statistics 4:594*t*
 new commitments 4:594*t*
 policy frameworks 4:590, 4:593*t*
 recent activities 4:592
 research studies 4:594*t*
 funding considerations 4:221
 globalization effects 1:536, 1:698
 importance 4:590
 policies and programs 4:438
 state steering models 4:557–558
 lifelong learning 1:89, 1:93, 1:227
 Organisation for Economic Cooperation and Development (OECD) 4:588
 policies and programs 4:707, 6:12
 secondary education 6:251
 social capital 6:97
 vocational education and training (VET) research 8:353
World Conference on Higher Education 1:698
World Council of Comparative Education Societies (WCCES)
 foreign education versus comparative education 1:269
 international focus 1:268
 member societies 1:267*t*
 non-member societies 1:270
 origins 1:266
 society characteristics 1:267
 world congresses 1:268*t*
World Disarmament Campaign 1:851
World Federation for Medical Education (WFME) 3:23
World Health Organization (WHO) 2:520, 8:262
world music 7:708
World Trade Organization (WTO) 1:536, 4:598, 4:707, 6:532, 7:535, 8:413
WorldWatcher (computer software) 1:469, 8:198
World Wide Web (WWW)
 curriculum discourse analysis 1:345–346
 electronic portfolios 3:239
 flexible learning 4:442
 growth factors 8:2
 image-based education research 6:501
 knowledge communities 8:194
 learning strategies 8:103
 multiple information sources 5:194
 open-source online resources 8:5
 self-directed learning 8:186
 vocational education and training (VET) resources 8:366
 see also Internet-based education
worry *see* anxiety
writing instruction
 advanced levels 5:426–431
 expression and articulation 5:428
 future directions 5:430
 multimodal communication 5:428
 narrative genres 5:426
 1950s and 1960s 5:426
 1960s to 1980s 5:426
 1990s to present 5:428
 speaking–writing connection 5:430
 transition stages
 argumentative mode 5:428
 primary to secondary school level 5:428
 real-world discourses 5:427
assessment conferences 3:403–409
 basic concepts 3:404
 constructive responses 3:275
 definitions 3:403
 historical background 3:404
 informal assessment systems 3:405
 parental involvement 3:406
 portfolio assessments 3:406
 purpose 3:403
 recent developments
 general discussion 3:407
 Learning Record 3:407
 New Zealand Education Monitoring Project (NZEMP) 3:408
 self-assessments 3:406
 validity 3:408
assessment strategies 3:273–278
 background information 3:273
 classroom assessment 3:276
 constructive responses
 assessment conferences 3:275
 general discussion 3:275
 peer assessments 3:276
 self-assessments 3:276
 written feedback 3:275
 formative assessments 3:274
 general discussion 3:277
 portfolio assessments 3:274
 testing evaluations 3:273
computer-assisted instruction 8:3
early writing development 5:367–374
information and communication technology (ICT) 8:76
special needs students
 instructional strategies 2:764
 transcription skills 2:765
 written expression 2:764, 2:764*t*
Wug test 5:378*f*
Wyoming 1:654*f*
- ## Y
- Yemen
 early childhood education and care (ECEC) 2:111*t*
 Trends in International Mathematics and Science Study (TIMSS) 5:645*t*
Yerkes–Dodson law 3:498, 3:499*f*
yield–density models 7:343
youth crime 2:240
Youth Empowerment Strategies (YES!) Program 3:597
YouTube 8:202
Yule's paradox *see* Simpson's paradox
- ## Z
- Zambia
 socioeconomic/sociopolitical characteristics 1:199
 student achievement test performance 2:418*t*
 see also Africa
Zanzibar 5:896–903
 background information 5:896
 education system
 challenges 5:902
 educational finance 5:901, 5:902*t*
 enrollment 5:897*t*, 5:898*t*

- ministry of education 5:899
- monitoring and evaluation programs 5:900
- research programs 5:900
- structural and operational characteristics
 - adult education and training (AET) 5:899
 - alternative programs 5:899
 - early childhood education and care (ECEC) 5:896
 - general discussion 5:896
 - higher education 5:898, 5:898*t*
 - nonformal education 5:899
 - postbasic education 5:897
 - primary education 5:896, 5:897*t*, 5:898*t*
 - pupil–teacher ratios (PTRs) 5:898*t*
 - secondary education 5:897, 5:898*t*
 - special needs students 5:898
 - vocational education and training (VET) 5:898
 - teaching profession 5:899
 - vision and mission 5:896
 - student achievement test performance 2:418*t*
- Zimbabwe
 - early childhood education and care (ECEC) 2:111*t*
 - student achievement test performance 2:247*f*, 2:418*t*
- Zimmerman's social–cognitive model 5:504
- zone of proximal development (ZPD)
 - cognitive development theory 5:234
 - dynamic assessments 3:220–221
 - early childhood curriculum 1:517
 - instructional planning 3:367, 6:361
 - language learning theory 5:500
 - mastery theories 3:168
 - motivation 6:713
 - science learning 5:390

CONTRIBUTORS

H Abdi

The University of Texas at Dallas, Richardson, TX, USA

S K Abell

University of Missouri, Columbia, MO, USA

W AbuSneineh

University of Bradford, Bradford, UK

T W Acee

Texas State University – San Marcos, San Marcos, TX, USA

M Acker-Hocevar

Washington State University, Pullman, WA, USA

C A Adams

University of Alberta, Edmonton, AB, Canada

M H Adamshick

US Naval Academy, Annapolis, MD, USA

M Ahmed

Brac University, Dhaka, Bangladesh

J Ainley

Australian Council for Educational Research, Camberwell, VIC, Australia

M Ainley

University of Melbourne, Melbourne, VIC, Australia

M Ainscow

University of Manchester, Manchester, UK

C A Albers

University of Wisconsin–Madison, Madison, WI, USA

D P Aldridge

Emory University, Atlanta, GA, USA

D O Aleshire

The Association of Theological Schools in the United States and Canada, Pittsburgh, PA, USA

P A Alexander

University of Maryland, College Park, MD, USA

S Alexander

University of Technology Sydney, Sydney, NSW, Australia

F A N Alhija

Tel Aviv University, Tel Aviv, Israel

M C Alkin

University of California, Los Angeles, Los Angeles, CA, USA

L Allal

University of Geneva, Geneva, Switzerland

A Allalouf

National Institute for Testing and Evaluation, Jerusalem, Israel

J Alleman

Michigan State University, East Lansing, MI, USA

A-R Allen

University of Rochester, Rochester, NY, USA

W R Allen

University of California, Los Angeles, Los Angeles, CA, USA

C Allport

National Tertiary Education Union (Australia), South Melbourne, VIC, Australia

R G Almond

ETS, Princeton, NJ, USA

J W Altschuld

The Ohio State University, Columbus, OH, USA

D E Alvermann

University of Georgia, Athens, GA, USA

A C Alwasilah

Indonesia University of Education, Bandung, West Java, Indonesia

J-A Amadeo

University of Maryland, College Park, MD, USA

A Amodeo

Law School Admission Council, Newtown, PA, USA

M B Anderson

Association of American Medical Colleges, Washington, DC, USA

P Anderson

University of the West Indies, Kingston, Jamaica

R Anderson

University of Minnesota, Wayzata, MN, USA

S E Anderson

University of Toronto, Toronto, ON, Canada

P Andersson

Linköping University, Linköping, Sweden

D F Andrade

Universidade Federal de Santa Catarina, Florianópolis, Brazil

R Andrews

University of London, London, UK

D Andrich

The University of Western Australia, Crawley, WA, Australia

A Anning

University of Leeds, Leeds, UK

R Apodaca,

University of Pittsburgh, Pittsburgh, PA, USA

B Applebaum

Syracuse University, Syracuse, NY, USA

J S Arnfast

University of Copenhagen, Copenhagen S, Denmark

R Arnold

University of Sydney, Sydney, NSW, Australia

K Aronsson

Stockholm University, Stockholm, Sweden

J A Arter

Educational Testing Service, Portland, OR, USA

M Arvaja

University of Jyväskylä, Jyväskylä, Finland

B Y Ashbaker

Brigham Young University, Provo, UT, USA

S R Asher

Duke University, Durham, NC, USA

V G Aukrust

University of Oslo, Oslo, Norway

P G Avery

University of Minnesota, Minneapolis, MN, USA

S Avveduto

National Research Council, Institute for Research on Population and Social Policies, Rome, Italy

J B Axe

Simmons College, Boston, MA, USA

L F Bachman

University of California, Los Angeles, Los Angeles, CA, USA

E Bahizi

Rwandan National Commission for UNESCO, Kigali, Republic of Rwanda

A L Bailey

University of California, Los Angeles, Los Angeles, CA, USA

B Bailey

University of Greenwich, London, UK

I Bairamov

Izmir University of Economics, Izmir, Turkey

L Baker

University of Maryland, Baltimore County, Baltimore, MD, USA

R S J d Baker

Worcester Polytechnic Institute, Worcester, MA, USA

R Baker

New Zealand Council for Educational Research, Wellington, New Zealand

M Bakia

SRI International, Arlington, VA, USA

J P Bakken

Illinois State University, Normal, IL, USA

B M J Balasooriya

Ministry of Education, Battaramulla, Sri Lanka

A Y Baldwin

University of Connecticut, Storrs, CT, USA

G Baldwin

The University of Melbourne, Melbourne, VIC, Australia

M Barber**H B Barker**

Regis University, Denver, CO, USA

C Barnes

University of New England, Armidale, NSW, Australia

W S Barnett

Rutgers University, New Brunswick, NJ, USA

M Barrs

Centre for Literacy in Primary Education, London, UK

S Barthel

GlaxoSmithKline, London

D J Bartholomew

The London School of Economics and Political Science, London, UK

F Bartolucci

Università degli Studi di Perugia, Perugia, Italy

R M Bassett

The World Bank, Washington, DC, USA

A Basu

Indian Statistical Institute, Kolkata, India

R Bates

University of Deakin, Waurn Ponds, VIC, Australia

F Baudoin

Purdue University, West Lafayette, IN, USA

D E Beals

The University of Tulsa, Tulsa, OK, USA

D E Beck

University of Florida, Gainesville, FL, USA

D Beckett

The University of Melbourne, Melbourne, VIC, Australia

K A Beckmann

Columbia University, New York, NY, USA

F Beddie

National Centre for Vocational Education Research,
Adelaide, SA, Australia

P T Begley

Pennsylvania State University, University Park, PA, USA

D Beijjaard

Eindhoven University of Technology, Eindhoven,
The Netherlands

J Beirlant

Katholieke Universiteit Leuven, Heverlee, Belgium

I I Bejar

ETS, Princeton, NJ, USA

C Belfield

Queens College, City University of New York, Flushing,
NY, USA

P Benner

Carnegie Foundation for the Advancement of Teaching,
Berkeley, CA, USA

C Bennett

US Census Bureau, Suitland, MD, USA

J Bennett

OECD, Paris, France

R E Bennett

Educational Testing Service, Princeton, NJ, USA

M Ben-Peretz

University of Haifa, Haifa, Israel

G Berberoğlu

Middle East Technical University, Ankara, Turkey

E Bernhardt

Stanford University, Stanford, CA, USA

A Berry

Monash University, Clayton, VIC, Australia

B Bessette-Symons

Syracuse University, Syracuse, NY, USA

G Bethell

Anglia Assessment Ltd., Stowmarket, UK

E P Bettinger

Stanford University, Stanford, CA, USA

J R Betts

University of California, San Diego, La Jolla, CA, USA

L Beudert

University of Arizona, Tucson, AZ, USA

W L Bewley

University of California at Los Angeles, Los Angeles,
CA, USA

J Lo Bianco

University of Melbourne, Melbourne, VIC, Australia

S K Biklen

Syracuse University, Syracuse, NY, USA

S Billett

Griffith University, Brisbane, QLD, Australia

B Billingsley

Virginia Tech, Blackburg, VA, USA

C Bingham

Simon Fraser University, Burnaby, BC, Canada

P Black

King's College London, London, UK

J Blackmore

Deakin University, Burwood, VIC, Australia

J Blanden

University of Surrey, Guildford, UK

P Bélanger

UNESCO Institute for Education, Hamburg, Germany

J Blase

University of Georgia, Athens, GA, USA

P Blatchford

Institute of Education, University of London, London, UK

H Blok

University of Amsterdam, Amsterdam, The Netherlands

S Blum-Kulka

Hebrew University, Jerusalem, Israel

M Boekaerts

Leiden University, Leiden, The Netherlands

I E Bogotch

Florida Atlantic University, Boca Raton, FL, USA

M Bong

Korea University, Seoul, South Korea

F B Bonner

Howard University, Washington, DC, USA

M Bonnett

University of Bath, Bath, UK

H Borko

Stanford University, Stanford, CA, USA

M Bottery

University of Hull, Hull, UK

B A Bottge

University of Kentucky, Lexington, KY, USA

J Boyd

Arizona State University, Tempe, AZ, USA

M Bray

International Institute for Educational Planning, Paris, France

M Bray

The University of Hong Kong, China

J Brennan

The Open University, London, UK

R L Brennan

University of Iowa, Iowa City, IA, USA

T Brennen

University of Oslo, Oslo, Norway

L Bresler

University of Illinois at Urbana-Champaign, Champaign, IL, USA

A Brewer

San Diego State University, San Diego, CA, USA

D J Brewer

University of Southern California, Los Angeles, CA, USA

B K Bridges

American Council on Education, Washington, DC, USA

D Bridges

University of Cambridge, Cambridge, UK

S Brinkman

University of Adelaide, Adelaide, SA, Australia

R Britez

University of Illinois at Urbana-Champaign, Champaign, IL, USA

P R Britto

Yale University, New Haven, CT, USA

P Broadfoot

University of Gloucestershire, Cheltenham, UK

P J Brockwell

Colorado State University, Fort Collins, Co, USA

C Brook

Curtin University of Technology, Perth, WA, Australia

J Brooks-Gunn

Columbia University, New York, NY, USA

J Brophy[†]**C N Brouwer**

Radboud University Nijmegen, Nijmegen, The Netherlands

A J Brown

University of Warwick, Coventry, UK

B A Brown

Stanford University School of Education, Stanford, CA, USA

C A Brown**K D Brown**

University of South Carolina, Columbia, SC, USA

M Brown

King's College London, London, UK

R S Brown

University of Southern California, Los Angeles, CA, USA

I Bråten

University of Oslo, Oslo, Norway

P Brug

St. Mary's University College, Twickenham, UK

A Bryant

Leeds Metropolitan University, Leeds, UK

E Bryant

University of California, Los Angeles, Los Angeles, CA, USA

P Buchwald

Heinrich-Heine University Dusseldorf, Dusseldorf, Germany

R Buddin

RAND Corporation, Santa Monica, CA, USA

J Buitink

University of Groningen, Groningen, The Netherlands

S Bukstein

University of Central Florida, Orlando, FL, USA

L M Bullock

University of North Texas, Denton, TX, USA

G Burke

Monash University, Clayton, VIC 3800, Australia

S Burkhardt

Saint Xavier University, Chicago, IL, USA

M S Burns

George Mason University, Fairfax, VA, USA

J Burstein

Educational Testing Service, Princeton, NJ, USA

P L Busk

University of San Francisco, San Francisco, CA, USA

J Cai

University of Delaware, Newark, DE, USA

K Cairns

Ontario Institute for Studies in Education, Toronto, ON, Canada

[†] Deceased

P Caldarella

Brigham Young University, Provo, UT, USA

P Calder

London South Bank University, London, UK

B J Caldwell

Educational Transformations Pty Ltd, Melbourne, VIC, Australia

J Campbell

University of Warwick, Coventry, UK

B Cantwell

University of Arizona, Tucson, AZ, USA

D Carless

University of Hong Kong, Hong Kong, People's Republic of China

E Carlson

Westat, Rockville, MD, USA

P Carneiro

University College London, London, UK; Institute for Fiscal Studies, London, UK

G N Carnes

University of South Carolina, Columbia, SC, USA

M Carnoy

Stanford University, Stanford, CA, USA

D Carr

University of Edinburgh, Edinburgh, UK

G Cartledge

The Ohio State University, Columbus, OH, USA

M A Ceja

California State University, Northridge, Northridge, CA, USA

R M Cervero

The University of Georgia, Athens, GA, USA

R Chakrabarti

Federal Reserve Bank of New York, New York, NY, USA

J Chambers

American Institutes for Research, Palo Alto, CA, USA

L Chang

The Chinese University of Hong Kong, Hong Kong People's Republic of China

B Chapman

Australian National University, Canberra, ACT, Australia

C Chapman

University of Manchester, Manchester, UK

K Charmaz

Sonoma State University, Rohnert Park, CA, USA

M Chatterji

Columbia University, New York, NY, USA

A Chavdarova

Sofia University "St. Kliment Ohridski", Sofia, Bulgaria

H-M Cheah

Nanyang Technological University, Singapore

M T H Chi

Arizona State University, Tempe, AZ, USA

K M Chinn

New Mexico State University, Las Cruces, NM, USA

I Chis Ster

University College London, London, UK

V Cho

The University of Texas at Austin, Austin, TX, USA

L Christensen

Brigham Young University, Provo, UT, USA

C A Christie

University of California, Los Angeles, California, USA

S Chuang

National Taiwan Normal University, Taiwan, Republic of China

R Cigman**D J Clandinin**

University of Alberta, Edmonton, AB, Canada

M H Clark

Southern Illinois University, Carbondale, Carbondale, IL, USA

D Clark-Carter

Staffordshire University, Stoke-on-Trent, UK

M Clarke

University of New South Wales, Sydney, NSW, Australia

M Clayton

University of Warwick, Coventry, UK

G Cleveland

University of Toronto Scarborough, Toronto, ON, Canada

D Cloney

University of Melbourne, Melbourne, VIC, Australia

D Cohen-Zada

Ben-Gurion University, Beer-Sheva, Israel

S Colley**A Collins**

Northwestern University, Evanston, IL, USA

E Comeaux

University of Kentucky, Lexington, KY, USA

F M Connelly

University of Toronto, Toronto, ON, Canada

B Connolly

National University of Ireland, Maynooth, Republic of Ireland

J C Conroy

University of Glasgow, Glasgow, UK

M A Conostas

Cornell University, Ithaca, NY, USA

K J Copeland

Missouri State University, Springfield, MO, USA

L Corno

Columbia University, Sherborn, MA, USA

M A M Cortés

University of Kassel, Kassel, Germany

W C Coscarelli

Southern Illinois University, Carbondale, IL, USA

T E Costantino

University of Georgia, Athens, GA, USA

B Cowie

University of Waikato, Hamilton, New Zealand

E D Cramer

Florida International University, Miami, FL, USA

P Cranton

Penn State University at Harrisburg, Middletown, PA, USA

B P M Creemers

University of Groningen, Groningen, The Netherlands

R D Crick

University of Bristol, Bristol, UK

T J Crooks

University of Otago, Dunedin, New Zealand

G M Crow

Indiana University, Bloomington, IN, USA

R L Crowson

Vanderbilt University, Nashville, TN, USA

J Crowther

University of Edinburgh, Edinburgh, UK

F Cuisinier

University of Paris Ouest, Nanterre, France

K M Culp

Education Development Center, Inc., New York, NY, USA

J J Cumming

Griffith University, Brisbane, QLD, Australia

J Currie

Columbia University, New York, NY, USA

C H O Daeyeon

Sookmyung Women's University, Seoul, Republic of Korea

P Dahler-Larsen

University of Southern Denmark, Odense, Denmark

R Dale

University of Bristol, Bristol, UK

B J Daley

University of Wisconsin – Milwaukee, Milwaukee, WI, USA

M Danesi

University of Toronto, Toronto, ON, Canada

J Daniel

Commonwealth of Learning, Vancouver, BC, Canada

K L Daniel

University of Cincinnati, Cincinnati, OH, USA

V I Daniels

Southern University and A&M College, Baton Rouge, LA, USA

A A I Daouda

University Abdou Moumouni de Niamey, Niamey, Republic of Niger

R Daugherty

Cardiff University, Cardiff, Wales, UK

A Davies

Classroom Connections International, Courtenay, BC, Canada

B J Davies

University of Hull, Hull, UK

B Davies

University of Hull, Hull, UK

B Davis

University of British Columbia, Vancouver, BC, Canada

N Davis

University of Canterbury, Christchurch, New Zealand

R Davis

Stanford University, Stanford, CA, USA

T E Davis

University of Maryland, College Park, MD, USA

A W Daviso III

The University of Akron, Akron, OH, USA

S Dawe

National Centre for Vocational Education Research Ltd., Adelaide, SA, Australia

M Daxner

Carl-von-Ossietzky Universitaet Oldenburg, Oldenburg, Germany

C Day

University of Nottingham, Nottingham, UK

C E Daye

University of North Carolina, Chapel Hill, Chapel Hill, NC, USA

N Day-Vines

Virginia Tech, Falls Church, VA, USA

A F De Champlain

National Board of Medical Examiners, Philadelphia, PA, USA

E De Corte

Katholieke Universiteit Leuven, Leuven, Belgium

T de Jong

University of Twente, Enschede, The Netherlands

M de Rosnay

University of Sydney, Sydney, NSW, Australia

S de Vries

University of Texas at San Antonio, San Antonio, TX, USA

L Dearden

University of London, London, UK; Institute for Fiscal Studies, London, UK

J M Dearn

University of Canberra, Canberra, ACT, Australia

E L Deci

University of Rochester, Rochester, NY, USA

C Dede

Harvard Graduate School of Education, Cambridge, MA, USA

T S Dee

Swarthmore College, Swarthmore, PA, USA

N DeFrance

Grand Valley State University, Grand Rapids, MI, USA

T Deissinger

University of Konstanz, Konstanz, Germany

L Deitmer

University of Bremen, Bremen, Germany

C Delgado

University of Texas, Austin, TX, USA

Z Deng

Nanyang Technological University, Singapore

M E Deo

University of California, Los Angeles, Los Angeles, CA, USA

R Desjardins

Danish University of Education, Copenhagen, Denmark

J Devonish

University of the West Indies, Kingston, Jamaica

R Z Dewan

Director (Research and Documentation) (Retired), National Academy for Educational Management (NAEM), Ministry of Education, Government of Bangladesh, Dhaka, Bangladesh

G Dierckx

Hogeschool-Universiteit Brussel, Brussel, Belgium

D Dill

The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

G Y Dinwiddie

University of Maryland College Park, College Park, MD, USA

F Dochy

University of Leuven, Leuven, Belgium

P Dolton

University of London, Surrey, UK

M L Donahue

University of Illinois at Chicago, Chicago, IL, USA

E A Dooley

West Virginia University, Morgantown, WV, USA

J T dos Santos

Federal University of Bahia, Salvador, Brazil

T Downes

Tufts University, Medford, MA, USA

S M Downing

University of Illinois at Chicago, Chicago, IL, USA

D Drake-Clark

The University of Georgia, Athens, GA, USA

K Draper

University of Pretoria, Pretoria, South Africa

M J Drummond

Formerly of the Faculty of Education, University of Cambridge, Cambridge, UK

I Duckett

Learning and Skills Network, London, UK

L M Garrick Duhaney

State University of New York at New Paltz, New Paltz, NY, USA

P Duignan

Australian Catholic University, Sydney, NSW, Australia

C Duke

University of Leicester, Leicester, UK; University of Stirling, Scotland, UK; RMIT University, Melbourne, VIC, Australia

S C Dunn

Brigham Young University, Provo, UT, USA

R P Durán

University of California Santa Barbara, Santa Barbara, CA, USA

S Dvir-Gvirsman

Hebrew University, Jerusalem, Israel

T T Dyches

Brigham Young University, Provo, UT, USA

J Dyer

Memorial University of Newfoundland, St. John's, NL, Canada

A Dyson

University of Manchester, Manchester, UK

L Earl

Aporia Consulting, Toronto, ON, Canada

P Earley

University of London, London, UK

J S Eaton

Council for Higher Education Accreditation, Washington, DC, USA

K Ecclestone

Oxford Brookes University, Oxford, UK

P D Eckel

American Council on Education, Washington, DC, USA

K M Edmondson

Cornell University, Ithaca, NY, USA

C P Edwards

University of Nebraska–Lincoln, Lincoln, NE, USA

T J H M Eggen

Cito, Psychometric Research Center, Arnhem, The Netherlands

R G Ehrenberg

Cornell University, Ithaca, NY, USA

E R Eide

Brigham Young University, Provo, UT, USA

B Eilam

University of Haifa, Haifa, Israel

C J Eleweke

Radford University, Radford, VA, USA

E El-Khawas

George Washington University, Washington, DC, USA

J G Elliott

Durham University, Durham, UK

S N Elliott

Vanderbilt University, Nashville, TN, USA

J Ellis

Ministry of Education, Windhoek, Republic of Namibia

P-E Ellström

Linköping University, Linköping, Sweden

J Elwood

Queen's University Belfast, Belfast, Northern Ireland, UK

S Embretson

Georgia Institute of Technology, Atlanta, GA, USA

W H M Emons

Tilburg University, Tilburg, The Netherlands

J Enders

University of Twente, Enschede, The Netherlands

P Enslin

University of Glasgow, Glasgow, UK

F Erickson

University of California, Los Angeles, Los Angeles, CA, USA

O Erstad

University of Oslo, Oslo, Norway

H Ertl

University of Oxford, Oxford, UK

G Estrella

Universidad Autónoma de Baja California, Tijuana, Mexico

L Evans

Texas A&M University, College Station, TX, USA

C W Evers

The University of New South Wales, Sydney, NSW, Australia

S C Faircloth

Pennsylvania State University, University Park, PA, USA

X Fan

University of Virginia, Charlottesville, VA, USA

J J Faraway

University of Bath, Bath, UK

P Farenga

Medford, MA, USA

C Farrugia

University Ombudsman, Valletta, Malta

H Farstad

Ministry of Education and Research, Oslo, Norway

A Fasulo

University of Portsmouth, UK

J R Feagin

Texas A&M University, College Station, TX, USA

M Fedri

The Pennsylvania State University, University Park, PA, USA

W Feinberg

University of Illinois at Urbana–Champaign, Champaign, IL, USA

T Fenwick

University of Alberta, Edmonton, AB, Canada

F Ferraty

Toulouse University, Toulouse, France

A Ferrer

University of Calgary, Calgary, AB, Canada

F Ferrier

Monash-University-ACER Centre for the Economics of Education and Training, Australia

R Fessler

Johns Hopkins University, Baltimore, MD, USA

J Field

University of Stirling, Stirling, UK

J Fielden

CHEMS Consulting, Hampshire, UK

J Fien

RMIT University, Melbourne, VIC, Australia

S E Fienberg

Carnegie Mellon University, Pittsburgh, PA, USA

L A Fierro

Claremont Graduate University, Claremont, CA, USA

D N Figlio

Northwestern University, Evanston, IL, USA

M Figueroa

Stanford University, Stanford, CA, USA

N File

University of Wisconsin-Milwaukee, Milwaukee, WI, USA

A Filer

University of Cambridge, Cambridge, UK

V M Filippov

Peoples' Friendship University of Russia, Moscow, Russian Federation

A Fink

University of California, Los Angeles, Los Angeles, CA, USA; The Langley Research Institute, Pacific Palisades, CA, USA

W A Firestone

Rutgers University, New Brunswick, NJ, USA

K W Fischer

Harvard University Graduate School of Education, Cambridge, MA, USA

M Fleer

Monash University, Melbourne, Australia

J D Fletcher

Institute for Defense Analyses, Alexandria, VA, USA

B Fogo

Stanford University, Stanford, CA, USA

D Foley

The University of Texas at Austin, Austin, TX, USA

B A Ford

The University of Akron, Akron, OH, USA

T G Ford

Michigan State University, East Lansing, MI, USA

G Forrester

Liverpool John Moores University, Liverpool, UK

K Forrester

University of Leeds, Leeds, UK

M Forster

Australian Council for Educational Research, Melbourne, VIC, Australia

J A Forte

Salisbury University, Salisbury, MD, USA

K J Fosnacht

University of California, Los Angeles, Los Angeles, CA, USA

S A Fowler

University of Illinois, Champaign, IL, USA

T M Franke

University of California at Los Angeles, Los Angeles, CA, USA

K Fraser

Australian Catholic University, Melbourne, Victoria, Australia

K Freeman

Southern University and A & M College System, Baton Rouge, LA, USA

A Fu

Stanford University, Stanford, CA, USA

D Fuchs

Vanderbilt University, Nashville, TN, USA

L S Fuchs

Vanderbilt University, Nashville, TN, USA

R Fukkink

University of Amsterdam, Amsterdam, The Netherlands

M Fullan

University of Toronto, Toronto, ON, Canada

Furqon

Indonesia University of Education, Bandung, West Java, Indonesia

R A Gable

Old Dominion University, Norfolk, VA, USA

M J Gaertner**M Galton**

University of Cambridge, Cambridge, UK

J Gamble

University of Cape Town, Cape Town, South Africa

P Gammage**S Ganesh**

Massey University, Palmerston North, New Zealand

C García-Forero

University of Barcelona, Barcelona, Spain

J Gardner

Queen's University Belfast, Belfast, Northern Ireland

R Gardner III

Ohio State University, Columbus, OH, USA

J L Gastwirth

The George Washington University, Washington, DC, USA

R Gaztambide-Fernández

Ontario Institute for Studies in Education, Toronto, ON, Canada

C Genishi

Columbia University, New York, NY, USA

J E Gentle

George Mason University, Fairfax, VA, USA

D Gijbels

University of Antwerp, Antwerpen, Belgium

W H Gijsselaers

Maastricht University, Maastricht, The Netherlands

R Gilbert

The University of Queensland, Brisbane, QLD, Australia

D Gillborn

Institute of Education, London, UK

C Gipps

University of Wolverhampton, Wolverhampton, UK

C A W Glas

University of Twente, Enschede, The Netherlands

L D Glasman

University of California, Berkeley, CA, USA

N S Glasman

University of California, Berkeley, CA, USA

P Glewwe

University of Minnesota, St. Paul, MN, USA

J Goddard

Newcastle University, Newcastle upon Tyne, UK

L Goedegebuure

University of New England, Armidale, NSW, Australia

S R Goldman

University of Illinois, Chicago, IL, USA

E B Goldring

Vanderbilt University, Nashville, TN, USA

J Gómez-Benito

University of Barcelona, Barcelona, Spain

A Gonczi

University of Sydney, Sydney, NSW, Australia

J Goodman

Harvard University, Cambridge, MA, USA

U Goswami

University of Cambridge, Cambridge, UK

C R Graham

Brigham Young University, Provo, UT, USA

P Graham

Queensland University of Technology, Kelvin Grove, QLD, Australia

S Graham

University of California, Los Angeles, Los Angeles, CA, USA

P A Grant

Slippery Rock University of Pennsylvania, Slippery Rock, PA, USA

P B Grant**S G Grant**

Binghamton University, State University of New York, Binghamton, NY, USA

M E Graue

University of Wisconsin–Madison, Madison, WI, USA

J W Gravel

Center for Applied Special Technology, Wakefield, MA, USA

C Grbich

Flinders University, Adelaide, SA, Australia

J L Green

University of California, Santa Barbara, Santa Barbara, CA, USA

M F Green

American Council on Education, Washington, DC, USA

S L Green

State University of New York College at Buffalo, Buffalo, New York, USA

J G Greeno

University of Pittsburgh, Pittsburgh, PA, USA

M Greenacre

Universitat Pompeu Fabra, Barcelona, Spain

B Greer

Portland State University, Portland, OR, USA

R S Grenier

University of Connecticut, Storrs, CT, USA

E L Grigorenko

Yale University, New Haven, CT, USA

P Gronn

University of Cambridge, Cambridge, UK

K Gürüz

Retired Professor, Middle East Technical University, Ankara, Turkey

B Guinovart

National Administration for Public Education (ANEP), Montevideo, Uruguay

M Gunderson

University of Toronto, Toronto, ON, Canada

D Gunning

Welsh Assembly Government, Cardiff, UK

M Gunnlaugsdóttir

Ministry of Education, Science and Culture, Reykjavik, Iceland

H M Gunter

University of Manchester, Manchester, UK

I Gur-Ze'ev

University of Haifa, Haifa, Israel

E Gutiérrez-Peña

Universidad Nacional Autónoma de México, México D.F., Mexico

S J Haberman

Educational Testing Service, Princeton, NJ, USA

A F Hadwin

University of Victoria, Victoria, BC, Canada

G D Haertel

SRI International, Menlo Park, CA, USA

P Hager

University of Technology, Sydney, NSW, Australia

B E Hagtvet

University of Oslo, Oslo, Norway

B D Haig

University of Canterbury, Christchurch, New Zealand

B J Hake

University of Leiden, Leiden, The Netherlands

P Häkkinen

University of Jyväskylä, Jyväskylä, Finland

P R Hallam

Brigham Young University, Provo, UT, USA

S Hallam

Institute of Education, London, UK

J Halliday

University of Strathclyde, Glasgow, UK

S Halliday-Wynes

National Centre for Vocational Education Research, Adelaide, SA, Australia

P Hallinger

Asia Pacific Centre for Leadership and Change, Hong Kong Institute of Education, Hong Kong

P Hallinger

Mahidol University, Bangkok, Thailand

R Halverson

University of Wisconsin-Madison, Madison, WI, USA

A-L Halvorsen

Michigan State University, East Lansing, MI, USA

L S Hamilton

RAND Corporation, Pittsburgh, PA, USA

M Hammersley

The Open University, Milton Keynes, UK

P Hanani

National Institute for Testing and Evaluation, Jerusalem, Israel

F Haniff

Wits International Office, University of the Witwatersrand, Johannesburg, South Africa

N Z Haniff

University of Michigan, Ann Arbor, MI, USA

E Hannam

Jamaican National Commission for UNESCO, Kingston, Jamaica

E A Hanushek

Stanford University, Stanford, CA, USA

M Hardardottir

Ministry of Education, Science and Culture, Reykjavik, Iceland

B R Hare

Syracuse University, Syracuse, NY, USA

A Hargreaves

Boston College, Chestnut Hill, MA, USA

W Harlen

University of Bristol, Bristol, UK

G Harman

University of New England, Armidale, NSW, Australia

K Harman

University of New England, Armidale, NSW, Australia

A Harris

Institute of Education, London, UK

A Harris

University of Warwick, Coventry, UK

A Harris

University of Washington, Seattle, WA, USA

D N Harris

University of Wisconsin-Madison, Madison, WI, USA

M K Harris

University of Wisconsin-Milwaukee, Milwaukee, WI, USA

L-A Harris

National Centre for Vocational Education Research, Station Arcade, SA, Australia

P Harris-Obiakor

University of Wisconsin-Milwaukee, Milwaukee, WI, USA

C K Harrison

University of Central Florida, Orlando, FL, USA

C Harrison

King's College, London, UK

R J Harrison

Howard University, Washington, DC, USA

T Hascher

University of Salzburg, Salzburg, Austria

J A Hatch

University of Tennessee, Knoxville, TN, USA

A Hautamäki

University of Helsinki, Helsinki, Finland

J Hautamäki

University of Helsinki, Helsinki, Finland

A Havnes

University of Bergen, Bergen, Norway

M Hayden

Southern Cross University, Lismore, NSW, Australia

G Haydon

Institute of Education, London, UK

G Hayward

University of Oxford, Oxford, UK

M L Hazelton

Massey University, Palmerston North, New Zealand

R H Heck

College of Education, Honolulu, HI, USA

L V Hedges

Northwestern University, Evanston, IL, USA

S Hegarty

International Association for the Evaluation of Educational Achievement (IEA)

D Heikkinen

Eastern Michigan University, Ypsilanti, MI, USA

M Heimann

Linköping University, Linköping, Sweden; The Swedish Institute for Disability Research, Linköping, Sweden; The Norwegian Network for Infant Mental Health, Oslo, Norway

L Hemphill

Wheelock College, Boston, MA, USA

M Henkel

Brunel University, London, UK

G C Hentschke

University of Southern California, Los Angeles, CA, USA

J L Herman

University of California, Los Angeles, Los Angeles, CA, USA

R Hermans

Ghent University, Ghent, Belgium

J Heywood

Trinity College Dublin, Dublin, Republic of Ireland

A Hickling-Hudson

Queensland University of Technology, Brisbane, QLD, Australia

M D Hidalgo

University of Murcia, Murcia, Spain

D Higgins,

Educational Testing Service, Princeton, NJ, USA

R Higham

University of London, London, UK

I M Hija

Ministry of Education and Vocational Training, Zanzibar, Tanzania

J E Hinostroza

Universidad de La Frontera, Temuco, Chile

S J Hite

Brigham Young University, Provo, UT, USA

L Hixson

University at Albany-SUNY, Albany, NY, USA

A J Hobson

University of Nottingham, Nottingham, UK

V Hodapp

Goethe University Frankfurt am Main, Frankfurt, Germany

S R Hodge

The Ohio State University, Columbus, OH, USA

J Hodgen

King's College London, London, UK

P Hodgkinson

University of Leeds, Leeds, UK

K Hoffman**P Hogan**

National University of Ireland, Maynooth, Ireland

R Höghjelm

Stockholm University, Stockholm, Sweden

A Holmen

Aarhus University, Copenhagen NV, Denmark

B Holmes

Hamdan Bin Mohammed e-University, Dubai, United Arab Emirates

E Hong

University of Nevada, Las Vegas, NV, USA

S Hong

Korea University, Seoul, South Korea

H D Hoover

University of Iowa, Iowa City, IA, USA

D Hopkins

University of London, London, UK

K Horsthemke

University of the Witwatersrand, Johannesburg, South Africa

H D Horton

University at Albany-SUNY, Albany, NY, USA

H Horz

University of Applied Sciences Northwestern Switzerland, Olten, Switzerland

K S Howard

Marymount Manhattan College, New York, NY, USA

S J Howie

University of Pretoria, Pretoria, South Africa

W J Hoyer

Syracuse University, Syracuse, NY, USA

C K Hsiao

National Taiwan University, Taipei, Taiwan

W-Y Hsieh

University of Illinois at Chicago, Chicago, IL, USA

G-H Huang

National Chiao Tung University, Hsinchu, Taiwan

H-H Huang

Academia Sinica, Taipei, Taiwan

S-Y Huang

Academia Sinica, Taipei, Taiwan

T Huang

Wenzao Ursuline College of Languages, Taiwan, Republic of China

J Huber

University of Regina, Regina, SK, Canada

S G Huber

University of Teacher Education Central Switzerland (PHZ) Zug, Zug, Switzerland

A Hüfner

Berlin, Germany

K Hüfner

Berlin, Germany

C Hughes

Aylesford Mews, Sunderland, UK

M Hughes

Aylesford Mews, Sunderland, UK

P Hughes

Australian National University, Canberra, ACT, Australia

H Hulshof

ICLON, Leiden University Graduate School of Teaching, Leiden, The Netherlands

J Husman

Arizona State University, Tempe, AZ, USA

J T Hustedt

Rutgers University, New Brunswick, NJ, USA

R Hutton

Peabody College of Vanderbilt University, Nashville, TN, USA

S Hynds

Syracuse, NY, USA

S Iksal

LIUM (Computer Sciences Laboratory of the University of Maine), Laval, France

K Illeris

Danish University of Education, Copenhagen, Denmark

J Imants

Radboud University, Nijmegen, The Netherlands

J Imazeki

San Diego State University, San Diego, CA, USA

M H Immordino-Yang

University of Southern California, Los Angeles, CA, USA

L Ingvarson

Australian Council for Educational Research, Camberwell, VIC, Australia

E Ip

Wake Forest University, Winston-Salem, NC, USA

H Ishii

Nagoya University, Nagoya, Japan

S A Issan

Sultan Qaboos University, Muscat, Sultanate of Oman

D Jackson

National College for School Leadership, Bedford, UK

J F L Jackson

University of Wisconsin–Madison, Madison, WI, USA

J Jacobs

University of Colorado, Boulder, CO, USA

R L Jacobs

The Ohio State University, Columbus, OH, USA

A James

Loyola Marymount University, USA

M James

University of Cambridge, Cambridge, UK

R James

The University of Melbourne, Melbourne, VIC, Australia

N S Jampol

University of Maryland, College Park, MD, USA

A Janowski

Academy of Humanities and Economics in Lodz, Poland

M Janus

McMaster University, Hamilton, ON, Canada

P Jarvis

University of Surrey, Guildford, UK

S Järvelä

University of Oulu, Finland

B D Jee

Northwestern University, Evanston, IL, USA

T Jeffs

Durham University, Durham, UK and University of Bedfordshire, Luton, UK

K N Jenkins

Newcastle University, Newcastle upon Tyne, UK

G M Jensen

Creighton University, Omaha, NE, USA

J O Jewell

Loyola Marymount University, Los Angeles, CA, USA

M-C Jipguep-Akhtar

Howard University, Washington, DC, USA

Y A-M Jobe

Gambia National Commission for UNESCO, Banjul, Gambia

A M Johansen

University of Warwick, Coventry, UK

A D Johnson

Columbia University, New York, NY, USA

S Johansson

University of Umeå, Umeå, Sweden

S Johns

University of Tasmania, Launceston, TAS, Australia

M S Johnson

Columbia University, New York, NY, USA

J Johnson-Bailey

The University of Georgia, Athens, GA, USA

S Johnston-Rodriguez

Northern Illinois University, DeKalb, IL, USA

A Joma

State Education Development Agency, Riga, Latvia

A Jones

University of Waikato, Hamilton, New Zealand

J N Jørgensen

University of Copenhagen, Copenhagen S, Denmark

P Joyce,

Dublin City University, Dublin, Republic of Ireland

J Jung

The University of Texas at Austin, Austin, TX, USA

B W Junker

Carnegie Mellon University, Pittsburgh, PA, USA

S L Kagan

Columbia University, New York, NY, USA

Y Kali

Israel Institute of Technology, Haifa, Israel

S B Kamerman

Columbia University, New York, NY, USA

A Kanwar

Commonwealth of Learning, Vancouver, BC, Canada

C Kapitzke

Queensland University of Technology, Kelvin Grove, QLD, Australia

S A Karabenick

University of Michigan, Ann Arbor, MI, USA

A Karagiannis

Greek Ministry of National Education and Religious Affairs, Athens, Greece

G E Karlsen

South Trondelag University College, Trondheim, Norway

T Karmel

National Centre for Vocational Education Research (NCVER), Adelaide, SA, Australia

J D Katsafanas

O'Shea Learning Consultants, Media, PA, USA

B Kaur

National Institute of Education, Singapore

J Keating

University of Melbourne, Melbourne, VIC, Australia

B M Kehm

University of Kassel, Kassel, Germany

A Kelava

Technical University Darmstadt, Darmstadt, Germany

G Kelchtermans

Katholieke Universiteit Leuven, Leuven, Belgium

T Kellaghan

St. Patrick's College, Dublin, Ireland

T Kennie

Ranmore Consulting, Ranmore, UK

M Keränen

Ministry of Education and Research, Tartu, Estonia

C Kessels

Leiden University, Leiden, The Netherlands

R J Kettler

Vanderbilt University, Nashville, TN, USA

P Kgobe

Centre for Education Policy Development, Johannesburg, South Africa

D Khasnabis

University of Michigan, Ann Arbor, MI, USA

E Khoo

Umeå University, Umeå, Sweden

J K Kidd

George Mason University, Fairfax, VA, USA

M Killen

University of Maryland, College Park, MD, USA

S Kilpatrick

University of Tasmania, Launceston, TAS, Australia

H S Kim

Georgia Institute of Technology, Atlanta, GA, USA

K Kimmel

Murdoch University, Murdoch, WA, Australia

E Kimura-Walsh

Santa Clara University, Santa Clara, CA, USA

B M King

Clemson University, Clemson, SC, USA

C Kinginger

Pennsylvania State University, University Park, PA, USA

D Kirk

University of Bedfordshire, Bedford, UK

J Kirkland

The Association of Commonwealth Universities,
London, UK

V Klenowski

Queensland University of Technology, Brisbane, QLD,
Australia

B Knight

National Centre for Vocational Education Research,
Adelaide, SA, Australia

J Knight

University of Toronto, Toronto, ON, Canada

L L Knoche

University of Nebraska–Lincoln, Lincoln, NE, USA

E Ko

The Hong Kong University of Science and Technology,
Hong Kong, People's Republic of China

K Koellner

University of Colorado, Denver, CO, USA

J Kõrgesaar

Ministry of Education and Research, Tartu, Estonia

V Kožemiakina

Education Development Centre, Vilnius, Lithuania

I Kogan

University of Mannheim, Mannheim, Germany

U Kogan[†]**G Kokolakis**

National Technical University of Athens, Athens, Greece

T R Konold

University of Virginia, Charlottesville, VA, USA

M Kooy

University of Toronto, Toronto, ON, Canada

D Koretz

Harvard Graduate School of Education, Cambridge,
MA, USA

F A J Korthagen

VU University, Amsterdam, The Netherlands

K R Kossar

California University of Pennsylvania, California, PA, USA

R B Kozma

SRI International, California, San Francisco, CA, USA

K Krainer

Universität Klagenfurt, Klagenfurt, Austria

J Krajcik

University of Michigan, Ann Arbor, MI, USA

M Krashinsky

University of Toronto Scarborough, Toronto, ON, Canada

T R Kratochwill

University of Wisconsin–Madison, Madison, WI, USA

B Kriechel

ROA – Maastricht University, Maastricht,
The Netherlands; IZA, Bonn, Germany

G D Kuh

Indiana University Bloomington, Bloomington, IN, USA

A M Kuntz

University of Alabama, Tuscaloosa, AL, USA

S Kupiainen

University of Helsinki, Helsinki, Finland

D-B Kwon

Sookmyung Women's University, Seoul, Republic of Korea

L Kyriakides

University of Cyprus, Nicosia, Cyprus

S Kyvik

NIFU STEP, Oslo, Norway

C Labbé

Universidad de La Frontera, Temuco, Chile

V K LaBoskey

Mills College, Oakland, CA, USA

S A Lachtman

University of Oxford, Oxford, UK

G W Ladd

Arizona State University, Tempe, AZ, USA

H F Ladd

Duke University, Durham, NC, USA

G Ladson-Billings

University of Wisconsin–Madison, Madison, WI, USA

J G Ladwig

University of Newcastle, Newcastle, NSW, Australia

K C Lai

The Hong Kong Institute of Education, Hong Kong,
People's Republic of China

C Lai

Michigan State University, East Lansing, MI, USA

A B Lakota

Ministry of Education and Sport, Ljubljana, Slovenia

S Lamb

University of Melbourne, Melbourne, VIC, Australia

[†] Deceased

S Lambert

Paris School of Economics, Paris, France

S Landau

King's College London, De Crespigny Park, London, UK

L Langlois

Laval University, Sainte-Foy, QC, Canada

H Lankford

University at Albany, Albany, NY, USA

C Lanskey

University of Cambridge, Cambridge, UK

U Lanzendorf

University of Kassel, Kassel, Germany

J Larson

University of Rochester, Rochester, NY, USA

J Lasonen

University of Jyväskylä, Jyväskylä, Finland

J Laspina

Independent Scholar, Los Angeles, CA, USA

D Laurillard

Institute of Education, London, UK

D Laveault

University of Ottawa, Ottawa, ON, Canada

N Law

University of Hong Kong, Hong Kong, People's Republic of China

J Leach

University of Leeds, Leeds, UK

C-M Lee

National Taiwan Normal University, Taiwan, Republic of China

S-Y Lee

The Chinese University of Hong Kong, Hong Kong, People's Republic of China

C B Leggon

Georgia Institute of Technology, Atlanta, GA, USA

K Leithwood

University of Toronto, Toronto, ON, Canada

R A Leon

University of Wisconsin-Madison, Madison, WI, USA

T L Lepicki

The Ohio State University, Columbus, OH, USA

F K S Leung**B Levering**

Utrecht University Utrecht, The Netherlands; Fontys Professional University, Tilburg, The Netherlands

C Levesque

Missouri State University, Springfield, MO, USA

B Levin

University of Toronto, Toronto, ON, Canada

H M Levin

Teachers College, Columbia University, New York, NY, USA

T Levin

Tel Aviv University, Tel Aviv, Israel

H G Levine

University of California, Davis, CA, USA

D C Levy

University at Albany (SUNY), Albany, NY, USA

A Lewins

University of Surrey, Guildford, UK

C Lewis

Fordham University, New York City, NY, USA

X Li

University of California, Los Angeles (UCLA), Los Angeles, CA, USA

C P Lim

Edith Cowan University, Perth, WA, Australia

S M-Y Lim

Nanyang Technological University, Singapore

Angel M Y Lin

City University of Hong Kong, Hong Kong SAR, People's Republic of China

P Lindberg

National Institute for Health and Welfare, Helsinki, Finland

M Linn

University of California, Berkeley, CA, USA

R L Linn

University of Colorado at Boulder, Ouray, CO, USA

K Littleton

Open University, Milton Keynes, UK

H Liu

Beijing Normal University, Beijing, People's Republic of China

S Llinares

Universidad de Alicante, Alicante, Spain

M E R Lloyd

College of Charleston, Charleston, SC, USA

L Lochner

University of Western Ontario, London, ON, Canada

S Loeb

Stanford University, Stanford, CA, USA

C Lohrenscheit

German Institute for Human Rights, Berlin, Germany

N T Longford

Universitat Pompeu Fabra, Barcelona, Spain

N T Longford

UPF, Barcelona, Spain

L López

Universidad de La Frontera, Temuco, Chile

F Lopez-Segrera

Technical University of Catalonia, Barcelona, Spain

R Lotan

Stanford University, Stanford, CA, USA

J T Lott

Tamayo Lott Associates, Colesville, MD, USA

J Loughran

Monash University, Clayton, VIC, Australia

A J Love

Independent Evaluation Consultant, Toronto, ON, Canada

P Loveder

National Centre for Vocational Education Research, Adelaide, SA, Australia

T S Lovelace

Duquesne University, Pittsburgh, PA, USA

S R Ludvigsen

University of Oslo, Oslo, Norway

J Ludwig

University of Chicago, Chicago, IL, USA

A Luke

Queensland University of Technology, Kelvin Grove, QLD, Australia

M Lunenberg

VU University, Amsterdam, The Netherlands

K Luwel

Katholieke Universiteit Leuven, Heverlee, Belgium

E Lyby

Jacaranda Consult, Denmark

M Lynde

University of Texas at San Antonio, San Antonio, TX, USA

X Ma

University of Kentucky, Lexington, KY, USA

J MacBeath

University of Cambridge, Cambridge, UK

D Macbeth

The Ohio State University, Columbus, OH, USA

S Machin

University College London, London, UK

S F Macintyre

The University of Melbourne, Parkville, VIC, Australia

J MacKenzie

National Centre for Vocational Education Research, Adelaide, SA, Australia

R Maclean

Hong Kong Institute of Education, Hong Kong, People's Republic of China

R Maclean

UNESCO-UNEVOC International Centre, Bonn, Germany

C L Madda

University of Illinois, Chicago, IL, USA

L Maggioni

University of Maryland, College Park, MD, USA

S Magnussen

University of Oslo, Oslo, Norway

N Mahmood

University of the Punjab, Lahore, Pakistan

H M Mahmoud

The George Washington University, Washington, DC, USA

D P Malasa

Ministry of Education and Human Resources Development, Honiara, Solomon Islands

A Malderez

University of Leeds, Leeds, UK

R Malik

University of the Punjab, Lahore, Pakistan

C A Mallozzi

University of Kentucky, Athens, GA, USA

A Mandal

Indian Statistical Institute, Kolkata, India

S Marginson

University of Melbourne, Melbourne, VIC, Australia

F Marín-Martínez

University of Murcia, Murcia, Spain

R Marples

Roehampton University, London, UK

B Marshall

Kings College London, London, UK

S Marshall

Leadership Foundation for Higher Education, London, UK

H W Marsh

Oxford University, Oxford, UK

J Marsh

University of Sheffield, Sheffield, UK

I Martin

University of Edinburgh, Edinburgh, UK

A Marvasti

Penn State Altoona, Altoona, PA, USA

C Mason

New Mexico State University, Las Cruces, NM, USA

L H Mason

The Pennsylvania State University, University Park,
PA, USA

M Mason

Hong Kong Institute of Education, Hong Kong, People's
Republic of China

J Masschelein**G Masters**

Australian Council for Educational Research, Melbourne,
VIC, Australia

S Mathison

University of British Columbia, Vancouver, BC, Canada

H Matlay

Birmingham City University, Birmingham, UK

G E Matt

San Diego State University, San Diego, CA, USA

T Matuchniak

University of California, Irvine, Irvine, CA, USA

G S Maxwell

Brisbane, QLD, Australia

S May

University of Waikato, Hamilton, New Zealand

A Maydeu-Olivares

University of Barcelona, Barcelona, Spain

R E Mayer

University of California, Santa Barbara, CA, USA

P Mayo

University of Malta, Msida, Malta

K McAlpine,

National Tertiary Education Union (Australia), South
Melbourne, VIC, Australia

D F McCaffrey

The RAND Corporation, Pittsburgh, PA, USA

J McCarthy

University of Nevada, Las Vegas, NV, USA

M J McCollin

Slippery Rock University of Pennsylvania, Slippery Rock,
PA, USA

K L McDonald

University of Maryland, College Park, MD, USA

S-K McDonald

NORC at the University of Chicago, Chicago, IL, USA

P J McEwan

Wellesley College, Wellesley, MA, USA

C McInnis

PhillipsKPA, Richmond, Australia

S J McKinney

University of Glasgow, Glasgow, UK

P McLaren

University of California, Los Angeles, CA, USA

J McLeod

The University of Melbourne, Melbourne, VIC, Australia

W W McMahan

University of Illinois at Urbana-Champaign, Urbana,
IL, USA

G McNamara

Dublin City University, Dublin, Republic of Ireland

I McNay

University of Greenwich, London, UK

A McSwain

Norfolk State University, Williamsburg, VA, USA

B Means

SRI International, Menlo Park, CA, USA

D L Medin

Northwestern University, Evanston, IL, USA

V L Meek

University of Melbourne, Parkville, Victoria, Australia

T Mehring

Emporia State University, Emporia, KS, USA

P C Meijer

Utrecht University, Utrecht, the Netherlands

R R Meijer

University of Groningen, Groningen, The Netherlands

M Mendoza

Instituto Tecnológico Autónomo de México, México D.F.,
Mexico

S B Merriam

University of Georgia, Athens, GA, USA

M S Merry

University of Amsterdam, Amsterdam, The Netherlands;
Nieuwe Prinsengracht 130, 1018 VZ, Amsterdam,
The Netherlands

R Middlehurst

Leadership Foundation for Higher Education, London, UK

D Middleton

Loughborough University, Loughborough, UK

S C Middleton

Southern Illinois University, Carbondale, Carbondale,
IL, USA

J F Milem

University of Arizona, Tucson, AZ, USA

J L Miller

Columbia University, New York, NY, USA

M D Miller

University of Florida, Gainesville, FL, USA

R G Milder

Johns Hopkins University, Baltimore, MD, USA

J Misko

National Centre for Vocational Education Research,
Adelaide, SA, Australia

R J Mislevy

University of Maryland, College Park, MD, USA

F R Mkandawire

Malawi National Commission for UNESCO, Lilongwe,
Malawi

N A Mneimneh

Ministry of Education and Higher Education, Beirut,
Lebanon

P Moctezuma

Universidad Autónoma de Baja California, Tijuana,
Mexico

L Moffatt

University of British Columbia, Vancouver, BC, Canada

K H Mok

The Hong Kong Institute of Education, Hong Kong,
People's Republic of China

G Molenberghs

Universiteit Hasselt, Diepenbeek, Belgium

J Møller

University of Oslo, Oslo, Norway

D H Monk

The Pennsylvania State University, University Park,
PA, USA

S M Monnat

University of Nevada, Las Vegas, Las Vegas, NV, USA;
University at Albany-SUNY Albany, NY, USA

G F Moodie

Griffith University, Brisbane, QLD, Australia

J L Moore III

The Ohio State University, Columbus, OH, USA

T Moore

The Royal Children's Hospital, Melbourne, Australia

A I Mørch

University of Oslo, Oslo, Norway

J Moreland

University of Waikato, Hamilton, New Zealand

D L Morgan**J Morgan**

Swansea Metropolitan University, Wales, UK

L Morley

University of Sussex, Falmer, Brighton, UK

K Moronski

University of Michigan, Ann Arbor, MI, USA

M Morris

University of New Haven, West Haven, CT, USA

J Mottley

University of Central Florida, Orlando, FL, USA

S A Mueller

Pflugerville ISD, Austin, TX, USA

P Muennig

Columbia University, New York, NY, USA

D Muijs

University of Manchester, Manchester, UK

G M Mukuria

Slippery Rock University, Slippery Rock, PA, USA

D Mulera

Malawi National Commission for UNESCO, Lilongwe,
Malawi

B Mulford

University of Tasmania, Hobart, TAS, Australia

J Muller

University of Cape Town, Cape Town, South Africa

A Mungaray

Universidad Autónoma de Baja California, Tijuana,
Mexico

A Murata

Stanford University School of Education, Stanford,
CA, USA

J Murphy

Vanderbilt University, Nashville, TN, USA

F B Murray

University of Delaware, Newark, DE, USA

S Musti-Rao

The University of Cincinnati, Cincinnati, OH, USA

J W Mutegi

Indiana University, Indianapolis, IN, USA

J Myung

Stanford University, Stanford, CA, USA

R Naidu

National Centre for Vocational Education Research,
Station Arcade, SA, Australia

H Najafi

University of Toronto, Toronto, ON, Canada

M J Nathan

University of Wisconsin-Madison, Madison, WI, USA

J Navickaitė

Ministry of Education and Science, Vilnius, Lithuania

R C Neath

City University of New York, New York, NY, USA

T J Nechyba

Duke University, Durham, NC, USA; National Bureau of
Economic Research, Cambridge, MA, USA

T Nesbit

Simon Fraser University, Vancouver, BC, Canada

M J Neuman**R Neumann**

Macquarie University, Sydney, NSW, Australia

D Nevo

Tel Aviv University, Tel Aviv, Israel

C B Newman

University of California, Los Angeles, CA, USA

I Newman

Florida International University, Miami, FL, USA

R S Newman

University of California – Riverside, Riverside, CA, USA

S K Newsom

University of Tennessee, Knoxville, TN, USA

P E Newton

Cambridge Assessment, Cambridge, UK

P T Ng

Nanyang Technological University, Singapore

Y J Ni

The Chinese University of Hong Kong, Hong Kong SAR, People's Republic of China

K Nielsen

University of Aarhus, Aarhus, Denmark

S Nielsen

The European Training Foundation, Torino, Italy

P Nikander

University of Jyväskylä, Jyväskylä, Finland

Z P Nkabinde

New Jersey City University, Jersey City, NJ, USA

J L Nkata

Makerere University, Kampala, Uganda

A Noble

Massey University, Palmerston North, New Zealand

T J Nokes

University of Pittsburgh, Pittsburgh, PA, USA

J J Norcini

Foundation for Advancement of International Medical Education and Research (FAIMER), Philadelphia, PA, USA

M Nores

NIEER, Rutgers University, New Brunswick, NJ, USA

D Nunan

University of Hong Kong, Hong Kong, SAR, People's Republic of China

T Nunes

University of Oxford, Oxford, UK

J Nunnery

Old Dominion University, Norfolk, Virginia, USA

B Nyland

RMIT University, Bundoora, VIC, Australia

F Oberklaid

The Royal Children's Hospital, Parkville, VIC, Australia

S O Obi

Kentucky State University, Frankfort, KY, USA

F E Obiakor

University of Wisconsin-Milwaukee, Milwaukee, WI, USA

M T Offor

University of Wisconsin-Milwaukee, Milwaukee, WI, USA

J O'Hara

Dublin City University, Dublin, Republic of Ireland

H S Olesen

Roskilde University, Roskilde, Denmark

T L Oliver

St. Mary's University College, Twickenham, UK

H F O'Neil

University of Southern California, Los Angeles, CA, USA

P Onghena

Katholieke Universiteit Leuven, Heverlee, Belgium

P Oreopoulos

University of Toronto, Toronto, ON, Canada

H R Ori

Ministry of Education and Community Development, Paramaribo, Republic of Suriname

D Osborn

University of Rochester, Rochester, NY, USA

M Osborn

University of Bristol, Bristol, UK

M D Osborne

University of Illinois at Urbana-Champaign, Champaign, USA

L Oseguera

University of California, Irvine, Irvine, CA, USA

D J O'Shea

O'Shea Learning Consultants, Media, PA, USA

L J O'Shea

Delaware County Intermediate Unit, Morton, PA, USA

P S O'Sullivan

UCSF School of Medicine, San Francisco, CA, USA

L A Owens

University of Wisconsin – Milwaukee, Milwaukee, WI, USA

J Page

Southern Cross University, Lismore, NSW, Australia

M E Page

University of California-Davis, Davis, CA, USA

F Pajares[†]**A S Palincsar**

University of Michigan, Ann Arbor, MI, USA

R Palmer

Network for Policy Research, Review and Advice on Education and Training, Amman, Jordan

J Pampallis

Centre for Education Policy Development, Johannesburg, South Africa

A T Panter

University of North Carolina, Chapel Hill, Chapel Hill, NC, USA

H P Parette

Illinois State University, Normal, IL, USA

J M Parisi

University of Illinois at Urbana-Champaign, Champaign, IL, USA

C Parks

Long Beach Unified School District, Long Beach, CA, USA

J M Parr

University of Auckland, Auckland, New Zealand

P J Pashley

Law School Admission Council, Newtown, PA, USA

H A Patrinos

World Bank, Washington, DC, USA

M D Pattie

Missouri State University, Springfield, MO, USA

J M Patton

The College of William and Mary, Williamsburg, VA, USA

L Paurienė

Education Development Centre, Vilnius, Lithuania

M Pavlova

Griffith University, Brisbane, QLD, Australia

A A Payne

McMaster University, Hamilton, ON, Canada

P A Pérez

California State University, Fullerton, Fullerton, CA, USA

D Pedder

University of Cambridge, Cambridge, UK

I Pelayo

University of Southern California, Los Angeles, CA, USA

J W Pellegrino

University of Illinois at Chicago, Chicago, IL, USA

H Penn

University of East London, London, UK

W R Penuel

Center for Technology in Learning, CA, USA

M Perez

Stanford University, Stanford, CA, USA

L Perez-Felkner

University of Chicago, Chicago, IL, USA

M A Peters

University of Illinois at Urbana-Champaign, Champaign, IL, USA

K D Peterson

University of Wisconsin-Madison, Madison, WI, USA

G R Peterson-Karlan

Illinois State University, Normal, IL, USA

A Petriwskyj

Queensland University of Technology, Brisbane, QLD, Australia

J Phelan

University of California, Los Angeles, Los Angeles, CA, USA

N Philipsen Hetzner

Columbia University, New York, NY, USA

R Picciotto

King's College, London, UK

D N Plank

Stanford University, Stanford, CA, USA

P Plant

Danish University of Education, København, Denmark

T Plevnik

Ministry of Education and Sport, Ljubljana, Slovenia

I Plewis

University of Manchester, Manchester, UK

B Pocock

University of South Australia, Adelaide, SA, Australia

M Podgursky

University of Missouri – Columbia, Columbia, MO, USA

A Pollard

University of London, London, UK

R-A Polvere

National Centre for Vocational Education Research, Adelaide, SA, Australia

F Pons

University of Oslo, Oslo, Norway

P Ponte

Leiden University, Leiden, The Netherlands

A C Porter

Vanderbilt University, Nashville, TN, USA

D Post

The Pennsylvania State University, University Park, PA, USA

[†] Deceased

G A Postiglione

The University of Hong Kong, Hong Kong, People's Republic of China

D G Pounder

University of Utah, Salt Lake City, UT, USA

C N Power

University of Queensland, Brisbane, QLD, Australia

L P Prater

Southeast Missouri State University, Cape Girardeau, MO, USA

M A Prater

Brigham Young University, Provo, UT, USA

M Print

University of Sydney, Sydney, NSW, Australia

G Psacharopoulos

London School of Economics, World Bank, Washington, DC, USA

B Pusser

University of Virginia, Charlottesville, VA, USA

L G Putney

University of Nevada, Las Vegas, Las Vegas, NV, USA

J Puukka

OECD Directorate for Education, Paris, France

J Qian

Educational Testing Service, Princeton, NJ, USA

V Quah

Singapore

C D Qualls

State University of New York College at Buffalo, Buffalo, New York, USA

D M Queiroz

Federal University of Bahia, Salvador, Brazil

P D Quijada

University of Texas at San Antonio, San Antonio, TX, USA

L F Quinn

University of Nevada, Las Vegas, NV, USA

B Raban

The University of Melbourne, Melbourne, VIC, Australia

S Rabe-Hesketh

University of California, Berkeley, CA, USA; University of London, London, UK

S N Rabinowitz

WestEd, San Francisco, CA, USA

J Radinsky

University of Illinois, Chicago, IL, USA

D Raffae

University of Edinburgh, Edinburgh, Scotland

G Ragusa

University of Southern California, Los Angeles, CA, USA

S F Rallis

University of Massachusetts Amherst, Amherst, MA, USA

S Ramirez

University of Washington, Seattle, WA, USA

A Rampal

Delhi University, Delhi, India

M Ramírez-Urquidy

Universidad Autónoma de Baja California, Tijuana, Mexico

E S Rangel

University of Pittsburgh, Pittsburgh, PA, USA

T E Raphael

University of Illinois, Chicago, IL, USA

T Rapley

Newcastle University, Newcastle upon Tyne, UK

P Ratcliffe

University of Warwick, Coventry, UK

N B Ratner

University of Maryland, College Park, MD, USA

E Rauscher

University of Wisconsin–Madison, Madison, WI, USA

M Raveaud

Université du Maine, Le Mans, France

G Rees

Cardiff University, Cardiff, UK

W C M Resing

Leiden University, Leiden, The Netherlands

L B Resnick

University of Pittsburgh, Pittsburgh, PA, USA

J Revuelta

Autonoma University of Madrid, Madrid, Spain

A J Reynolds

University of Minnesota, Minneapolis, MN, USA

D Reynolds

University of Plymouth, Plymouth, UK

A V Rezaev

St. Petersburg State University, St. Petersburg, Russian Federation

C Rhoads

Northwestern University, Evanston, IL, USA

E Rice

Johns Hopkins University, Baltimore, MD, USA

J K Rice

University of Maryland, College Park, MD, USA

M P Richards

The University of Texas at Austin, Austin, TX, USA

C B Richardson

University of Maryland, College Park, MD, USA

D Rich

University of Tasmania, Launceston, TAS, Australia

N Ricken

Katholieke Universiteit Leuven, Leuven, Belgium;
Universität Bremen, Bremen, Germany

W C Riddell

University of British Columbia, Vancouver, BC, Canada

K A Riley

Institute of Education, University of London, London, UK

D Rindskopf

CUNY Graduate Center, New York, NY, USA

S Rivkin

Amherst College, Amherst, MA, USA

F M Rizo

Universidad Autónoma de Aguascalientes,
Aguascalientes, Mexico

K D Roberts

University of Hawaii at Manoa, Honolulu, HI, USA

R D Roberts

Educational Testing Service, Princeton, NJ, USA

S Roberts

National Tertiary Education Union (Australia), South
Melbourne, VIC, Australia

V M J Robinson

The University of Auckland, Auckland, New Zealand

W Robinson

University of Exeter, Exeter, UK

M Robitaille

University of Quebec in Montreal, Montreal, QC, Canada

J Robson

London Metropolitan University, London, UK

K Roe

Katholieke Universiteit Leuven, Leuven, Belgium

C Rogers

University of Minnesota, Twin Cities, MN, USA

F H Rogers

The World Bank, Washington, DC, USA

T Rogers

University of British Columbia, Vancouver, BC, Canada

S Rohrmann

Technical University Darmstadt, Darmstadt, Germany

T R Rone

Morgan State University, Baltimore, MD, USA

J Roschelle

SRI International, Menlo Park, CA, USA

D H Rose

Center for Applied Special Technology, Wakefield,
MA, USA

P R Rosenbaum

University of Pennsylvania, Philadelphia, PA, USA

G B Rossman

University of Massachusetts Amherst, Amherst, MA, USA

A Rotatori

Saint Xavier University, Chicago, IL, USA

W-M Roth

University of Victoria, Victoria, BC, Canada

R Rothstein

Economic Policy Institute, Washington, DC, USA

I Rots

Ghent University, Ghent, Belgium

M Rouse

University of Aberdeen, Aberdeen, UK

J Roy

Georgetown University, Washington, DC, USA

C B Royeen

Saint Louis University, St. Louis, MO, USA

K Rubenson

University of British Columbia, Vancouver, BC, Canada

D B Rubin

Harvard University, Cambridge, MA, USA

T Rudas

Eötvös Loránd University, Budapest, Hungary

R Rudder

Ministry of Education, St. Michael, Barbados

J Rudduck[†]**R Rueda**

University of Southern California, Los Angeles, CA, USA

C Rukspollmuang

Chulalongkorn University, Bangkok, Thailand

R W Rumberger

University of California, Santa Barbara, Santa Barbara,
CA, USA

M A Runco

California State University, Fullerton, Fullerton, CA, USA;
Norwegian School of Economics and Business
Administration, Bergen, Norway

M A Runco

University of Georgia, Athens, GA, USA

M Russell

Boston College, Chestnut Hill, MA, USA

T Russell

Queen's University, Kingston, ON, Canada

V J Russell

The University of Melbourne, Melbourne, VIC, Australia

[†] Deceased

K Ruth

University of Bremen, Bremen, Germany

J Ryan

The Ontario Institute for Studies in Education, Toronto, ON, Canada

P Ryan

King's College, Cambridge, UK

Y Ryan

Australian Catholic University, Virginia, QLD, Australia

J J Ryoo

University of California, Los Angeles, CA, USA

D R Sadler

Griffith University, Mt Gravatt, QLD, Australia

D M Sainato

Ohio State University, Columbus, OH, USA

S J Salend

State University of New York at New Paltz, New Paltz, NY, USA

R Säljö

Göteborg University, Göteborg, Sweden

S R Salman

Consultant, Doha, Qatar

J Salmi

The World Bank, Washington, DC, USA

P Sammons

University of Nottingham, Nottingham, UK

W Sander

DePaul University, Chicago, IL, USA

A Sandill

University of Manchester, Manchester, UK

G Sang

Ghent University, Ghent, Belgium

L Santibañez

Fundación IDEA, Mexico City, Mexico

M Sapp

University of Wisconsin–Milwaukee, Milwaukee, WI, USA

W Sapp

Visual Impairment Education Services, Cohutta, GA, USA

P Sarda

Toulouse University, Toulouse, France

E Sato

WestEd, San Francisco, CA, USA

M Sato

University of Minnesota, Twin Cities, MN, USA

M Saunders

National Centre for Vocational Education Research, Station Arcade, SA, Australia

L F Scalas

Oxford University, Oxford, UK

D W Schanzenbach

University of Chicago, Chicago, IL, USA

L M Scheier

LARS Research Institute, Las Vegas, NV, USA;
Washington University, School of Medicine, St. Louis, MO, USA

A Schleicher

OECD, Paris, France

J A Schmidt

Northern Illinois University, DeKalb, IL, USA

E Schmitt

Southern Connecticut State University, New Haven, CT, USA

S Schnall

University of Plymouth, Plymouth, UK

B Schneider

Michigan State University, East Lansing, MI, USA

M Schneider

American Institutes for Research, Washington, DC, USA

W Schnotz

University of Koblenz-Landau, Landau, Germany

J Schostak

Manchester Metropolitan University, Manchester, UK

S Schreiber

University of Massachusetts, Amherst, MA, USA

D Schugurensky

University of Toronto, Toronto, ON, Canada

T Schuller

OECD, Paris, France

K Schultz

University of Pennsylvania, Philadelphia, PA, USA

R Schulz

University of Manitoba, Winnipeg, MB, Canada

D H Schunk

The University of North Carolina at Greensboro, Greensboro, NC, USA

C D Schunn

University of Pittsburgh, Pittsburgh, PA, USA

P A Schutz

University of Texas at San Antonio, San Antonio, TX, USA

R G Schwab

The Australian National University, Canberra, ACT, Australia

C Schwarzer

Heinrich-Heine University Dusseldorf, Dusseldorf, Germany

J Schuille

Michigan State University, East Lansing, MI, USA

D Schwinge

Adelphi University, Garden City, NY, USA

P Scott

Kingston University, Kingston upon Thames, UK

P Scott

University of Leeds, Leeds, UK

C Scott-Little

University of North Carolina at Greensboro, Greensboro, NC, USA

L Scrucca

Università degli Studi di Perugia, Perugia, Italy

W E Sedlacek

University of Maryland, College Park, MD, USA

M Segers

University of Leiden, Leiden, The Netherlands

V M Segovia

Miriam College – Environmental Studies Institute (ESI), Quezon City, Philippines

M Sellier

Paris 75013, France

I Senibulu

Ministry of Education, Suva, Fiji

C Shah

Monash University, Clayton, VIC 3800, Australia

P Shannon

Pennsylvania State University, University Park, PA, USA

M Shattock

University of London, London, UK

R Shavelson

Stanford University, Stanford, CA, USA

J Shemwell

Stanford University, Stanford, CA, USA

S M Sheridan

University of Nebraska–Lincoln, Lincoln, NE, USA

M D Shermis

The University of Akron, Akron, OH, USA

C M Shields

The University of Illinois at Urbana–Champaign, Champaign, IL, USA

M Shiyko

CUNY Graduate Center, New York, NY, USA

M H Showalter

Brigham Young University, Provo, UT, USA

S A Shrock

Southern Illinois University, Carbondale, IL, USA

T R Shultz

McGill University, Montreal, QC, Canada

V J Shute

Florida State University, Tallahassee, FL, USA

H Siegel

University of Miami, Coral Gables, FL, USA

K Sijtsma

Tilburg University, Tilburg, The Netherlands

H Silins

Flinders University of South Australia, Adelaide, SA, Australia

C Silver

University of Surrey, Guildford, UK

C G Simpson

Sam Houston State University, Huntsville, TX, USA

J M Singer

Universidade de São Paulo, São Paulo, Brazil

K Singh

Rutgers University, Piscataway, NJ, USA

S Sinharay

ETS, Princeton, NJ, USA

I Siraj-Blatchford

Institute of Education, University of London, London, UK

S Sjøberg

University of Oslo, Oslo, Norway

M Sklar

San Diego State University, San Diego, CA, USA

A Skrondal

Norwegian Institute of Public Health and University of Oslo, Oslo, Norway

S Slaughter

University of Georgia, Athens, GA, USA

R E Slavin

Johns Hopkins University, Baltimore, MD, USA

P Sleepers

University of Twente, Enschede, The Netherlands

J D Slotta

University of Toronto, Toronto, ON, Canada

M Slowey

Dublin City University, Dublin, Republic of Ireland

L Smethem

University of Nottingham, Nottingham, UK

P Smeyers

Ghent University, Gent, Belgium

P K Smith

University of Bergen, Bergen, Norway

C S Smith

Chris and Joy Selby Smith Pty Ltd., East Melbourne, VIC, Australia

D J Smith

Jackson State University, Jackson, MS, USA

E Smith

University of Ballarat, Ballarat, VIC, Australia

L Smith

Freelance Researcher, Lake District, UK

S W Smith

University of Florida, Gainesville, FL, USA

J Sánchez-Meca

University of Murcia, Murcia, Spain

C E Snow

Harvard Graduate School of Education, Cambridge, MA, USA

M Socias

American Institutes for Research, Palo Alto, CA, USA

Å Sohlman

Ministry of Employment, Stockholm, Sweden

X-Y Song

The Chinese University of Hong Kong, Hong Kong, People's Republic of China

G Southworth

National College for School Leadership, Nottingham, UK

C N Spence

Spelman College, Atlanta, GA, USA

M B Spencer

University of Pennsylvania, PA, USA

D E Spicer

Harvard Graduate School of Education, Cambridge, MA, USA

J P Spillane

Northwestern University, Evanston, IL, USA

B Sporn

WU (Vienna University of Economics and Business), Vienna, Austria

M G Springer

Peabody College of Vanderbilt University, Nashville, TN, USA

E P St. John

University of Michigan, Ann Arbor, MI, USA

J Stalker

University of Waikato, Hamilton, New Zealand

R J Starratt

Boston College, Chestnut Hill, MA, USA

R St.Clair

University of Glasgow, Glasgow, UK

A Stella

Australian Universities Quality Agency, Melbourne, VIC, Australia

R J Sternberg

Tufts University, Boston, MA, USA

J B Stewart

Pennsylvania State University, Pennsylvania, PA, USA

M S L Stewart

University of Central Florida, Orlando, FL, USA

E A L Stine-Morrow

University of Illinois at Urbana-Champaign, Champaign, IL, USA

G Stobart

Institute of Education, University of London, London, UK

R A Stodden

University of Hawaii at Manoa, Honolulu, HI, USA

L Stoll

University of London, London, UK

E B Stolzenberg

University of California, Los Angeles, Los Angeles, CA, USA

L Stone

University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

K Strid

University of Gothenburg, Gothenburg, Sweden

E Stringer

Curtin University of Technology, Perth, WA, Australia

S Stringfield

University of Louisville, Louisville, KY, USA

H I Strømso

University of Oslo, Oslo, Norway

K Strunk

University of Southern California, Los Angeles, CA, USA

M M Suárez-Orozco

Immigration Studies at New York University, New York, NY, USA

C Suárez-Orozco

Immigration Studies at New York University, New York, NY, USA

K Subrahmanyam

California State University, Los Angeles, Los Angeles, CA, USA

D Sumara

University of British Columbia, Vancouver, BC, Canada

M Sutphen

Carnegie Foundation for the Advancement of Teaching, San Francisco, CA, USA

R Sutton

Education Consultant, Cumbria, UK

R Sweet

University of Melbourne, Melbourne, VIC, Australia

J Tacq

Catholic University of Brussels, Brussels, Belgium

Y Takane

McGill University, Montreal, QC, Canada

A Tashakkori

Florida International University, Miami, FL, USA

S Taut

Pontificia Universidad Católica de Chile, Santiago, Chile

C Tayler

University of Melbourne, Melbourne, VIC, Australia

G G Taylor

University of Florida, Gainesville, FL, USA

L L Taylor

Texas A&M University, College Station, TX, USA

P Taylor

Swinburne University, Hawthorn, VIC, Australia

C Teddlie

Louisiana State University, Baton Rouge, LA, USA

D J Tedick

University of Minnesota, Minneapolis, MN, USA

U Teichler

University of Kassel, Kassel, Germany

J A Temple

University of Minnesota, Minneapolis, MN, USA

R T Teranishi

New York University, New York, NY, USA

L Terzi

Roehampton University, London, London, UK

L Tett

University of Edinburgh, Edinburgh, UK

B Thayer-Bacon

University of Tennessee, Knoxville, TN, USA

H Theisens

Centre for Educational Research and Innovation, OECD, Paris, France

D Thirumalai

University of California, Los Angeles, Los Angeles, CA, USA

A R Thomas

University of Wollongong, Wollongong, NSW, Australia

S M Thomas

University of Bristol, Bristol, UK

V G Thomas

Howard University, Washington, DC, USA

J Thom

University of Victoria, Victoria, BC, Canada

C M Thompson

The Pennsylvania State University, University Park, PA, USA

P Thomson

Nottingham University, Nottingham, UK

K Thorpe

Queensland University of Technology, Brisbane, QLD, Australia

R Throop

University of Pennsylvania, Philadelphia, PA, USA

M L Thurlow

University of Minnesota, Minneapolis, MN, USA

A Tiana

Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain

D E H Tigelaar

ICLON Leiden University School of Training, Leiden, The Netherlands

A Tijiūnelytė

Education Development Centre, Vilnius, Lithuania

J B G Tilak

National University of Educational Planning and Administration, New Delhi, India

H Tillema

Leiden University, Leiden, The Netherlands

B Tinsley

University of Pennsylvania, PA, USA

J W Tippeconnic

Pennsylvania State University, University Park, PA, USA

M Tjiattas

North Carolina State University, Raleigh, NC, USA

T Tjus

University of Gothenburg, Gothenburg, Sweden

G F Tkach

Peoples' Friendship University of Russia, Moscow, Russian Federation

P D Tomlinson

University of Leeds, Leeds, UK

K J Topping

University of Dundee, Dundee, UK

J Torney-Purta

University of Maryland, College Park, MD, USA

R T Torres

Torres Consulting Group, New Orleans, LA, USA

K Toura

Greek Ministry of National Education and Religious Affairs, Athens, Greece

T Townsend

The University of Glasgow, Glasgow, UK

S Tozer

University of Illinois, Chicago, IL, USA

S C Trent

University of Virginia, Charlottesville, VA, USA

K Trigwell

University of Sydney, Sydney, Australia

C Troncoso

National Administration for Public Education (ANEP),
Montevideo, Uruguay

S Tøsse

Norwegian University of Science and Technology,
Trondheim, Norway

R Tsuneyoshi

The University of Tokyo, Tokyo, Japan

Y-K Tu

University of Leeds, Leeds, UK

P D Tucker

University of Virginia, Charlottesville, VA, USA

P Tulip

Board of Governors of the Federal Reserve System,
Washington, DC, USA

J E Tuovinen

University of the Sunshine Coast, Maroochydore DC,
QLD, Australia

P B Tymms

Durham University, Durham, UK

S J Unsworth

San Diego State University, San Diego, CA, USA

A Unwin

University of Augsburg, Augsburg, Germany

T Urdan

Santa Clara University, Santa Clara, CA, USA

S Uvalić-Trumbić

UNESCO, Paris, France

M Valcke

Ghent University, Ghent, Belgium

E A P Valenzuela

Southeast Asian Ministers of Education Innovation and
Technology, Quezon City, The Philippines

J Välimaa

University of Jyväskylä, Jyväskylä, Finland

G M van der Aalsvoort

Utrecht University, Utrecht, The Netherlands

W J van der Linden

CTB/McGraw-Hill, Monterey, CA, USA

M van der Wende

University of Twente, Enschede, The Netherlands

J H Van Driel

ICLON-Leiden University Graduate School of Teaching,
Leiden, The Netherlands

M van Manen

University of Alberta, Edmonton, AB, Canada

R E Van Reken

Families in Global Transition, Indianapolis, IN, USA

J van Tartwijk

ICLON Leiden University School of Training, Leiden,
The Netherlands

K van Veen

Leiden University, The Netherlands

M van Zanten

Foundation for Advancement of International Medical
Education and Research (FAIMER), Philadelphia,
PA, USA

M Vandenbroeck

Gent University, Gent, Belgium

E Vegas

The World Bank, Washington, DC, USA

S Vehviläinen

University of Helsinki, Helsinki, Finland

M Ventura

Educational Testing Service, Princeton, NJ, USA

J K Vermunt

Tilburg University, Tilburg, The Netherlands

L S Verplaetse

Southern Connecticut State University, New Haven,
CT, USA

L Verschaffel

Katholieke Universiteit Leuven, Leuven, Belgium

D Vesselinov

Sofia University "St. Kliment Ohridski", Sofia, Bulgaria

W Veugelers

University for Humanistics, Utrecht, The Netherlands

P Vieu

Toulouse University, Toulouse, France

J Vigdor

Duke University, Durham, NC, USA

A Vignoles

University of London, London, UK

C P Vistro-Yu

Ateneo de Manila University, Philippines

A Volansky

Tel-Aviv University, Tel Aviv, Israel

S Volet

Murdoch University, Murdoch, WA, Australia

A A von Davier

Educational Testing Service, Princeton, NJ, USA

J Vorhaus

University of London, London, UK

H J Vos

University of Twente, Enschede, The Netherlands

R Vue

University of California, Los Angeles, CA, USA

H Wagemaker

International Association for the Evaluation of Educational Achievement, Amsterdam, The Netherlands

Y Waghid

Stellenbosch University, Matieland, South Africa

J Wagner

University of California, Davis, Davis, CA, USA

A Walker

Hong Kong Institute of Education, Hong Kong SAR, People's Republic of China

R A Walker

University of Sydney, Sydney, NSW, Australia

D G Wall,

Brigham Young University, Provo, UT, USA

D L Wallin

University of Georgia, Athens, GA, USA

A Walstab

University of Melbourne, Melbourne, VIC, Australia

S Walters

University of Western Cape, Bellville, South Africa

G G Wankhede

Tata Institute of Social Sciences, Mumbai, India

M Warschauer

University of California, Irvine, Irvine, CA, USA

D Watson

University of London, London, UK

J C Wayman

The University of Texas at Austin, Austin, TX, USA

N M Webb

University of California, Los Angeles, Los Angeles, CA, USA

C F Webber

University of Calgary, Calgary, AB, Canada

S M Weber

Marburg University, Marburg, Germany

R Wedgwood**A M Weekers**

University of Twente, Enschede, The Netherlands

J W Wegner

University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

M M Wehella

Ministry of Education, Battaramulla, Sri Lanka

M L Wehmeyer

University of Kansas, Lawrence, KS, USA

B Weiner

University of California, Los Angeles, Los Angeles, CA, USA

C E Weinstein

The University of Texas at Austin, Austin, TX, USA

K Weir

Queensland University of Technology, Kelvin Grove, QLD, Australia

J V Wertsch

Washington University in St. Louis, St. Louis, MO, USA

L West

Canterbury Christ Church University, Canterbury, UK

J West-Burnham

St Mary's University College, London, UK

C Whatford

UK National Commission for UNESCO, London, UK

J A Whitcomb

University of Colorado at Boulder, Boulder, CO, USA

B White

University of Minnesota, Minneapolis, MN, USA

L F Wightman

University of North Carolina, Greensboro, NC, USA

S U Wijeratne

Ministry of Education, Battaramulla, Sri Lanka

J Wiley

University of Illinois at Chicago, Chicago, IL, USA

D Wiliam

University of Pittsburgh, Pittsburgh, PA, USA

D D Williams

Brigham Young University, Provo, UT, USA

K Williams

University of Michigan, Ann Arbor, MI, USA

A L Wilson

Cornell University, Ithaca, NY, USA

D N Wilson[†]**S M Wilson**

Michigan State University, East Lansing, MI, USA

[†] Deceased

S Wineburg

Stanford University, Stanford, CA, USA

L A Wingate

Western Michigan University, Kalamazoo, MI, USA

D Wink

University of Illinois, Chicago, IL, USA

P H Winne

Simon Fraser University, Burnaby, BC, Canada

F I Winters

University of Maryland, College Park, MD, USA

M S Wiske

Harvard Graduate School of Education, Cambridge, MA, USA

A Wolf

King's College, London, UK

C A Wolters

University of Houston, Houston, TX, USA

M W-y Wong

Hong Kong Baptist University, Hong Kong, People's Republic of China

M Wong-Lo

University of North Texas, Denton, TX, USA

J Wood

New Mexico State University, Las Cruces, NM, USA

E Woodberry

University of New England, Armidale, NSW, Australia

S Woodfield

University of Surrey, Guildford, UK

D Woodhouse

Australian Universities Quality Agency, Melbourne, VIC, Australia

A Woods

Queensland University of Technology, Kelvin Grove, QLD, Australia

L Wößmann

University of Munich, Munich, Germany

C A Wringe

Keele University, Keele, UK

T Wubbels

Utrecht University, Utrecht, The Netherlands

J Wyckoff

University of Virginia, Charlottesville, VA, USA

C Wylie

New Zealand Council for Educational Research, Wellington, New Zealand

S H Wynes

National Centre for Vocational Education Research Ltd., Adelaide, SA, Australia

Zeng Xiaodong

Beijing Normal University, Beijing, Peoples Republic of China

M Xie

Rutgers University, Piscataway, NJ, USA

S J Xu

University of Windsor, Windsor, ON, Canada; National Research Center for Foreign Language Education, Beijing Foreign Studies University, China

H Yanai

St. Luke's College of Nursing, Tokyo, Japan

J S Yang

California State University, Fullerton, Fullerton, CA, USA

L Yates

The University of Melbourne, Melbourne, VIC, Australia

R Yelland

OECD Directorate of Education, Paris, France

B Yembuu

National Centre for Non Formal and Distance education, Ulaanbaatar, Mongolia

R J Yinger

University of Cincinnati, Cincinnati, OH, USA

K Yokoyama

University of California, Berkeley, California, USA

A Young

Georgia Southern University, Statesboro, GA, USA

K R Young

Brigham Young University, Provo, UT, USA

M Young

University of London, London, UK; University of Bath, London, UK

P K Yu-kwong

The Chinese University of Hong Kong, Hong Kong, People's Republic of China

A L Yurick

Cleveland State University, Cleveland, OH, USA

R Zablacké

Education Development Centre, Vilnius, Lithuania

M Zairi

University of Bradford, Bradford, UK

D Zapata-Rivera

Educational Testing Service, Princeton, NJ, USA

L Zechlin

University of Duisburg-Essen, Duisburg, Germany

K Zechner

Educational Testing Service, Princeton, NJ, USA

M Zeidner

University of Haifa, Haifa, Israel

J Zhang

University of Maryland, College Park, MD, USA

L-F Zhang

The University of Hong Kong, Hong Kong

Z Zhang

McGill University, Montreal, QC, Canada

G Zhang

Michigan State University, East Lansing, MI, USA

Y Zhao

Michigan State University, East Lansing, MI, USA

A Ziderman

Bar-Ilan University, Ramat Gan, Israel

R Zimmer

Michigan State University, East Lansing, MI, USA

L Zimmerman

University of Pretoria, Pretoria, South Africa

R Zwick

University of California, Santa Barbara, Santa Barbara, CA, USA